

BASKETBOL MİLLİ TAKIM ADAY OYUNCU HAVUZUNUN  
MAKİNE ÖĞRENİMİ ALGORİTMALARIYLA TAHMİNİ

T.C. KOCAELİ SAĞLIK VE TEKNOLOJİ ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

CANDİDE ÖZTÜRK

YÜKSEK LİSANS TEZİ  
BİLİŞİM SİSTEMLERİ MÜHENDİSLİĞİ YÜKSEK LİSANS  
PROGRAMI

KOCAELİ, 2025

## ONAY SAYFASI

Bu tez Kocaeli Sağlık ve Teknoloji Üniversitesi Lisansüstü Eğitim Enstitüsü tarafından onaylanmıştır.

---

Prof. Dr. Yurdanur DİKMEN  
Enstitü Müdürü

Bu tezin **Kocaeli Sağlık ve Teknoloji Üniversitesi, Bilişim Sistemleri Mühendisliği Bölümü Anabilim Dalında Yüksek Lisans** derecesinin tüm gerekliliklerini karşıladığını onaylıyorum.

---

Dr. Öğr. Üyesi Mehmet KARA  
Bölüm Başkanı

CANDİDE ÖZTÜRK tarafından teslim edilen BASKETBOL MİLLİ TAKIM ADAY OYUNCU HAVUZUNUN MAKİNE ÖĞRENİMİ ALGORİTMALARIYLA TAHMİNİ başlıklı bu tezin kapsam ve kalite bakımından Yüksek Lisans derecesi için yeterli olduğunu değerlendirmektedir.

---

Dr. Öğr. Üyesi Ulaş VURAL  
Danışman

**Tez Jürisi Üyeleri: (1. isim: Jüri başkanı; 2. isim: Danışman)**

Dr. Öğr. Üyesi Hikmetcan ÖZCAN  
Bilgisayar Mühendisliği Bölümü, Kocaeli Üniversitesi

---

Dr. Öğr. Üyesi Ulaş VURAL  
Mühendislik ve Doğa Bilimleri Fakültesi, Kocaeli Sağlık ve Teknoloji Üniversitesi

---

Prof. Dr. Nevcihan DURU  
Mühendislik ve Doğa Bilimleri Fakültesi, Kocaeli Sağlık ve Teknoloji Üniversitesi

---

**Tarih: 20/01/2025**

## ORİJİNALLİK SAYFASI

İşbu belge ile çalışmamda yer alan tüm bilgilerin akademik ve etik kurallara uygun olarak elde edildiğini ve sunulduğunu beyan ederim. Ayrıca, kurallar gereği bu çalışmada özgün olmayan tüm materyal ve sonuçlar için ilgili kaynakların verildiğini beyan ederim.

Ad, Soyad: Candide ÖZTÜRK

İmza:

## **ABSTRACT**

### **PREDICTION OF THE NATIONAL BASKETBALL TEAM CANDIDATE PLAYER ROSTER USING MACHINE LEARNING ALGORITHMS**

ÖZTÜRK, Candide

Master's Thesis, Graduate Education Institute Information Systems Engineering

Supervisor: Dr.Ulaş VURAL

January 2025, 181 pages

The present study aims to provide a solution to the problem of selecting players who will exhibit high performance in the national team according to season or match statistical attributes among players in the Turkish Women's Basketball Super League. This study is important to evaluate all players according to their statistical attributes while determining the national team candidate rosters.

To this end, a range of classification algorithms – namely Naïve Bayes, Random Forest, Support Vector Machines, Extreme Gradient Boosting and Decision Trees – were applied to a data set comprising basketball statistics. The model with the highest accuracy and F1 Score was then created. The study revealed an imbalance between the number of Turkish women basketball players in the Women's Basketball League and the number of players who are candidates for the national team. While there are many players, the candidate roster represents a narrower group. The sampling algorithms applied in this study are designed to address this imbalance and insufficiency of data. In the study, firstly, a Principal Component Analysis was performed on the attributes consisting of statistical data. This analysis was performed with four different variance values, and the important ones of the attributes were preserved. Subsequently, before proceeding to the classification stage, the challenges posed by the small and imbalanced data set were re-evaluated using a range of oversampling methods, namely Synthetic Minority Oversampling, Random Over Sampling, Borderline SMOTE, Random Under Sampling and Random Under Sampling 5% + SMOTE. The modelling

phase was then undertaken by implementing classification algorithms. The modelling process yielded the conclusion that the XGBoost algorithm functions with 86% accuracy and an F1 Score when employing the random oversampling method. The reliability of the models was validated through 10-fold cross-validation.

It is hypothesized that the findings of this thesis study will contribute to studies based on player selection, such as determining the national team candidate roster. In future studies, it is anticipated that these models will be developed as software and will contribute to the evaluation of the performance of players while forming the rosters of both national teams and teams. The evaluation of the individual and interconnected situations of the players is also expected to contribute to the country's sporting success, facilitated by a performance-oriented modelling approach that employs numerical values.

**Keywords:** Machine learning, sports analytics, basketball, national team candidate roster, XGBoost

## ÖZ

### **BASKETBOL MİLLİ TAKIM ADAY OYUNCU HAVUZUNUN MAKİNE ÖĞRENİMİ ALGORİTMALARIYLA TAHMİNİ**

ÖZTÜRK, Candide

Yüksek Lisans, Lisansüstü Eğitim Enstitüsü Bilişim Sistemleri Mühendisliği

Tez Yöneticisi: Dr. Öğr. Üyesi Ulaş VURAL

Ocak 2025, 181 sayfa

Bu çalışma, Türkiye Kadın Basketbol Süper Ligi'nde oynayan oyuncular arasında milli takımda yüksek performans sergileyecek oyuncuların sezon veya maç istatistiksel özniteliklerine göre seçimi problemine çözüm sunmasını amaçlamaktadır. Milli takım aday kadroları belirlenirken bütün oyuncuların istatistiksel özniteliklerine göre değerlendirilebilmesi adına bu çalışma önem arz etmektedir.

Basketbol istatistiklerinden oluşan veri setine makine öğrenmesinin sınıflandırma algoritmalarından Naive Bayes, Rastgele Orman, Destek Vektör Makineleri, Ekstrem Gradyan Güçlendirme, Karar Ağaçları sınıflandırma algoritmaları uygulanarak en yüksek doğruluk ve F1 Skora sahip modelin oluşturulması sağlanmıştır. Kadın Basketbol Ligi'nde bulunan Türk kadın basketbolcuların sayısı ve milli takıma aday olan oyuncu sayısı arasında bir dengesizlik mevcuttur çünkü çok fazla oyuncu varken aday havuzu daha dar bir grubu temsil eder. Bu çalışmada uygulanan örnekleme algoritmalarıyla bu dengesizlik ve az veri olması problemine çözüm bulunması hedeflenmiştir.

Çalışmada öncelikle istatistiksel verilerden oluşan özniteliklere Birincil Bileşen Analizi yapılmış ve bu analiz 4 farklı varyans değeri ile yapılarak özniteliklerin önemli olanları korunmuştur. Sonrasında sınıflandırma yapılmadan önce az ve dengesiz veri setine sahip olması problemleri Sentetik Azınlık Aşırı Örnekleme, Rastgele Aşırı Örnekleme, Borderline SMOTE, Rastgele Alt Örnekleme ve Rastgele Alt Örnekleme %5 + SMOTE örnekleme yöntemleriyle yeniden değerlendirilmiş ve sınıflandırma algoritmaları uygulanarak modelleme yapılmıştır. Bu modellemeler sonucunda

Rastgele Aşırı Örnekleme yöntemi ile XGBoost algoritmasının %86 doğruluk ve F1 Skoru ile çalıştığı sonucu elde edilmiştir. 10 katlı çapraz doğrulama sonucu modellerin kararlılığı sağlanmıştır.

Bu tez çalışmasının sonuçlarının milli takım aday havuzunun belirlenmesi gibi oyuncu seçimine dayalı çalışmalara katkı sağlayacağı düşünülmektedir. İleride yapılacak olan çalışmalarda bu modeller yazılım olarak geliştirilerek hem milli takım hem de takımların kadrolarını oluştururken oyuncuların performansını değerlendirmesine katkı sağlayacağı düşünülmektedir. Sayısal değerlerle tamamen performans odaklı çalışan modellemeler sayesinde oyuncuların bireysel ve birbirleriyle bağlantılı durumlarının değerlendirilmesinin ülke sporuna da katkı sağlayacağı düşünülmektedir.

**Anahtar kelimeler:** Makine öğrenmesi, spor analitiği, basketbol, milli takım aday kadrosu, XGBoost

## TEŞEKKÜR

Araştırmamın öncelikle başlangıcında fikir sağlamasıyla, uygulamaların oluşturulması ve makine öğrenmesi kısmında yardımlarıyla nihai sonuca ulaşmamı sağlayan danışman hocam Dr. Öğr. Üyesi Ulaş VURAL’ a teşekkür ederim.

Yüksek lisansa başladığım zamanda benimle ders alan değerli arkadaşım Özgür SAĞIR’ a da teşekkür etmezsem haksızlık etmiş olurum. Birlikte ders aldığım bütün sınıf arkadaşlarım da her zaman destek olup bana yardım ettiler, kendilerine teşekkürü borç bilirim.

Okula başladığım ilk dönemden son anıma kadar bana her şekilde yardım eden, ders çalışmam için ortam sağlayan ve kapılarımı açıp, destek olan “Afife Kahve & Kruvasan” ailesine teşekkür ederim.

Ayrıca eğitimim boyunca bana farklı perspektifler açan, çalışmalarımda beni destekleyen, takıldığım her konuda bana el uzatan ve yol gösteren Kocaeli Sağlık ve Teknoloji Üniversitesi Mühendislik ve Doğa Bilimleri Fakültesi’nin çok değerli öğretim üyeleri hocalarıma sonsuz teşekkürlerimi sunarım.

En büyük teşekkürüm aileme. Bu tezi yazacak konuma gelmemi sağlayan, yaptığım kariyer değişikliğinde, aldığım her kararda arkamda duran annem Özlem ÖZTÜRK ve babam Tuncay ÖZTÜRK’ e minnetle teşekkürlerimi sunuyorum

## İÇİNDEKİLER

|                                                                  |      |
|------------------------------------------------------------------|------|
| ONAY SAYFASI.....                                                | ii   |
| ORİJİNALLIK SAYFASI.....                                         | iii  |
| ABSTRACT.....                                                    | iv   |
| ÖZ .....                                                         | vi   |
| TEŞEKKÜR.....                                                    | viii |
| İÇİNDEKİLER .....                                                | ix   |
| ŞEKİLLER DİZİNİ.....                                             | xi   |
| TABLolar DİZİNİ .....                                            | xii  |
| KISALTMALAR VE SEMBOLLER DİZİNİ .....                            | xiii |
| 1.GİRİŞ .....                                                    | 1    |
| 1.1 Tezin Amacı ve Kapsamı.....                                  | 2    |
| 1.2 Geçmiş Çalışmalar .....                                      | 3    |
| 1.2.1 NBA Liginde Yapılmış Çalışmalar.....                       | 3    |
| 1.2.2 Euroleague ve Avrupa Ligleri İçin Yapılmış Çalışmalar..... | 4    |
| 1.2.3 Türkiye Basketbol Ligi İçin Yapılmış Çalışmalar.....       | 4    |
| 1.3 Kavramsal Tanımlar .....                                     | 4    |
| 1.3.1 Basketbol .....                                            | 4    |
| 1.3.2 Makine Öğrenmesi.....                                      | 5    |
| 2.BASKETBOL VE İSTATİSTİKSEL VERİLER .....                       | 7    |
| 2.1 Basketbolda Kullanılan İstatistiksel Veriler .....           | 7    |
| 2.2 Basketbolda İstatistiksel Verilerin Önemi .....              | 9    |
| 3.YÖNTEM.....                                                    | 10   |

|       |                                                                                                          |     |
|-------|----------------------------------------------------------------------------------------------------------|-----|
| 3.1   | Veri Setinin Elde Edilmesi.....                                                                          | 10  |
| 3.2   | Veri Setinin Temizlenmesi ve Normalizasyonu.....                                                         | 12  |
| 3.3   | Veri Setine PCA Uygulanması.....                                                                         | 13  |
| 3.4   | İstatistiksel Verilerin Değerlendirilmesi İçin Kullanılacak Sınıflandırma Modellerinin Belirlenmesi..... | 14  |
| 3.4.1 | Rastgele Orman.....                                                                                      | 14  |
| 3.4.2 | Naive Bayes .....                                                                                        | 15  |
| 3.4.3 | Karar Ağacı.....                                                                                         | 15  |
| 3.4.4 | Destek Vektör Makineleri .....                                                                           | 16  |
| 3.4.5 | Ekstrem Gradyan Güçlendirme .....                                                                        | 16  |
| 3.5   | Örnekleme Modellerinin Belirlenmesi .....                                                                | 17  |
| 3.5.1 | Sentetik Azınlık Aşırı Örnekleme .....                                                                   | 18  |
| 3.5.2 | Rastgele Aşırı Örnekleme .....                                                                           | 18  |
| 3.5.3 | Borderline SMOTE.....                                                                                    | 18  |
| 3.5.4 | Rastgele Alt Örnekleme .....                                                                             | 18  |
| 3.5.5 | RUS%5 + SMOTE .....                                                                                      | 18  |
| 3.6   | Değerlendirme Ölçütlerinin Belirlenmesi .....                                                            | 19  |
| 4.    | BULGULAR.....                                                                                            | 22  |
| 5.    | TARTIŞMA.....                                                                                            | 34  |
| 6.    | SONUÇ VE ÖNERİLER .....                                                                                  | 40  |
|       | KAYNAKÇA .....                                                                                           | 42  |
|       | EK A.....                                                                                                | 45  |
|       | EK B.....                                                                                                | 146 |
|       | KİŞİSEL YAYINLAR VE ESERLER .....                                                                        | 167 |
|       | ÖZGEÇMİŞ .....                                                                                           | 168 |

## ŞEKİLLER DİZİNİ

|                                                                                   |    |
|-----------------------------------------------------------------------------------|----|
| Şekil 1.1 Basketbolda oyuncu mevkileri.....                                       | 5  |
| Şekil 2.1 Örnek bir istatistik belgesi [10].....                                  | 7  |
| Şekil 3.1 Rastgele Orman Sınıflandırıcı [15].....                                 | 14 |
| Şekil 3.2 Naive Bayes Sınıflandırma [16].....                                     | 15 |
| Şekil 3.3 Karar Ağacı Sınıflandırma [17] .....                                    | 16 |
| Şekil 3.4 XGBoost Sınıflandırma [18].....                                         | 17 |
| Şekil 3.5 Çalışmanın karmaşıklık matrislerinden örnek .....                       | 20 |
| Şekil 3.6 10 katlı çapraz doğrulama[19] .....                                     | 21 |
| Şekil 4.1 En iyi kat grafiği (PCA Yok).....                                       | 22 |
| Şekil 4.2 En iyi kat grafiği (PCA_0,90).....                                      | 23 |
| Şekil 4.3 En iyi kat grafiği (PCA_0,95).....                                      | 24 |
| Şekil 4.4 En iyi kat grafiği (PCA_0,99).....                                      | 24 |
| Şekil 4.5 En kötü kat doğruluk oranı grafiği (PCA Yok) .....                      | 25 |
| Şekil 4.6 En kötü kat doğruluk oranı grafiği (PCA_0,90) .....                     | 26 |
| Şekil 4.7 En kötü kat doğruluk oranı grafiği (PCA_0,95) .....                     | 26 |
| Şekil 4.8 En kötü kat doğruluk oranı grafiği (PCA_0,99) .....                     | 27 |
| Şekil 4.9 Modellerin örnekleme yöntemlerine göre F1 Skoru grafiği (PCA Yok).....  | 28 |
| Şekil 4.10 Modellerin örnekleme yöntemlerine göre F1 Skoru grafiği (PCA 0,90) ..  | 28 |
| Şekil 4.11 Modellerin örnekleme yöntemlerine göre F1 Skoru grafiği (PCA 0,95) ..  | 29 |
| Şekil 4.12 Modellerin örnekleme yöntemlerine göre F1 Skoru grafiği (PCA 0,99) ..  | 30 |
| Şekil 4.13 PCA yapılmadan model doğruluk grafiği .....                            | 30 |
| Şekil 4.14 Kümülatif varyans 0,90 ile PCA uygulanmış model doğruluk grafiği ..... | 31 |
| Şekil 4.15 Kümülatif varyans 0,95 ile PCA uygulanmış model doğruluk grafiği ..... | 31 |
| Şekil 4.16 Kümülatif varyans 0,99 ile PCA uygulanmış model doğruluk grafiği ..... | 32 |
| Şekil 5.1 Model Karşılaştırma Grafiği .....                                       | 38 |

## TABLÖLAR DİZİNİ

|                                                                                                 |    |
|-------------------------------------------------------------------------------------------------|----|
| Tablo 3.1 Öznitelikler ve Açıklamaları .....                                                    | 11 |
| Tablo 3.2 Sabit istatistikler harici oluşturulan öznitelikler ve açıklamaları.....              | 12 |
| Tablo 3.3 PCA ile elde edilen bileşen sayıları.....                                             | 13 |
| Tablo 3.4 Karmaşıklık matrisi .....                                                             | 20 |
| Tablo 4.1 Modellerin PCA yaklaşımlarına göre doğruluk ve F1 Skorları .....                      | 33 |
| Tablo 5.1 PCA yapılmadan örnekleme ve modellere göre F1 Skoru ve Doğruluk Değerleri .....       | 34 |
| Tablo 5.2 PCA 0,90 alındığında örnekleme ve modellere göre F1 Skoru ve Doğruluk Değerleri ..... | 35 |
| Tablo 5.3 PCA 0,95 alındığında örnekleme ve modellere göre F1 Skoru ve Doğruluk Değerleri ..... | 36 |
| Tablo 5.4 PCA 0,99 alındığında örnekleme ve modellere göre F1 Skoru ve Doğruluk Değerleri ..... | 37 |

## KISALTMALAR VE SEMBOLLER DİZİNİ

|         |                                                                              |
|---------|------------------------------------------------------------------------------|
| DT      | Decision Tree (Karar Ağacı)                                                  |
| FN      | False Negative (Yanlış negatifler)                                           |
| FP      | False Positive (Yanlış pozitifler)                                           |
| KBSL    | Türkiye Kadınlar Basketbol Süper Ligi                                        |
| MT      | Milli Takım Durumu                                                           |
| MVP     | Most Valuable Player (En Değerli Oyuncu)                                     |
| NB      | Naive Bayes                                                                  |
| NBA     | National Basketball Association (Amerikan Basketbol Ligi)                    |
| PCA     | Principal Component Analysis (Temel Bileşen Analizi)                         |
| RF      | Random Forest (Rastgele Orman)                                               |
| ROS     | Random Over Sampling (Rastgele Aşırı Örnekleme)                              |
| RUS     | Random Under Sampling (Rastgele Alt Örnekleme)                               |
| SMOTE   | Synthetic Minority Oversampling Technique (Sentetik Azınlık Aşırı Örnekleme) |
| SVM     | Support Vector Machines (Destek Vektör Makineleri)                           |
| TBF     | Türkiye Basketbol Federasyonu                                                |
| TN      | True Negative (Doğru negatifler)                                             |
| TP      | True Positive (Doğru pozitifler)                                             |
| XGBoost | Extreme Gradient Boosting (Ekstrem Gradyan Güçlendirme)                      |

## 1. GİRİŞ

Basketbol her geçen gün izlenirliđi artan bir spor dalıdır. Günümüzde Türk takımlarının, Avrupa liglerinde kadın ve erkek takımlarıyla aldıkları başarılarla izlenme oranları çok artmıştır. Voleybol Kadın Milli Takımı'nın Avrupa kupalarında aldığı başarılar, 2024 Paris Olimpiyatları'nda elde ettikleri derece ayrıca futbol milli takımımızın Euro2024 başarılarıyla, milli takımlar düzeyinde de spor ön plana çıkmaktadır. Verilen örneklerden görüldüğü gibi basketbol milli takımlarımız altyapılarda başarılar elde etse de şu anda A takım düzeyinde iyi sonuçlar alamamaktadır. Oysa ki ligler düzeyinde bakıldığında basketbol liglerimiz Euroleague, EuroCup gibi Avrupa liglerinde hem erkek hem kadın takımları olarak şampiyonluklar ve dereceler elde etmektedir. Kulüpler düzeyinde alınan bu başarıların milli takım düzeyinde de sürdürülebilmesi adına makine öğrenimi yöntemlerinden faydalanmak basketbolda başarıya ulaşma için önemli bir adım olacaktır. Bu sebeple kulüpler düzeyinde oynanan maçlardan elde edilen istatistikleri modelleme yöntemlerinde veri olarak kullanılarak milli takım için bir aday havuzu oluşturulabilir.

Spor günümüzde sadece eğlenmek için yapılan ve izlenen bir aktivite olmaktan ziyade artık şirketlerin ve şahısların ticari ortaklığı, reklam ortaklığı gibi ticari etkileri de olan bir alan haline gelmiştir. Bununla birlikte bahis şirketlerinin de gözünde popülaritesi artmıştır. Bundan dolayı da maç sonuçlarının tahmini önemli bir konu olmuştur. Ayrıca performans değerlendirmeleri açısından istatistik verilerinin incelenmesi gelecek maçları etkileyecek bir durumdur. Makine öğrenmesinin yöntemleriyle incelenen ve modellenen veriler daha efektif olarak kullanılabilir ve birçok alanda fayda sağlayacaktır.

Bahis şirketlerinin bu verileri incelemesi kendi kazançlarına bir durumu doğurduğundan ötürü bu konuya özellikle yatırım yapmaktadırlar. Hatta sadece basketbol değil futbol, beyzbol, Amerikan futbolu gibi branşlarda da veriler incelenmektedir. İnternetin gelişmesiyle istatistik verilerine ulaşmak kolaylaşmıştır. Sürekli bir akış içerisinde veriler elde edilmekte ve veri tabanlarında saklanmaktadır. Geçmiş çalışmalar futbol üzerine odaklansa da günümüzde çok yönlü çalışmalar

yapılmaktadır. Futbol verilerinin incelenmesine dair çalışmalar istatistiksel modeller ve makine öğrenmesi modelleri olarak ayrılmaktadır[1].

Bir takımın istatistiksel metotlarla bir sezon içinde attığı ve yediği sayılar, maçlardaki kazanıp kazanamama durumuna bakılarak yapılan çalışmalar daha yüzeysel kalırken makine öğrenmesi yöntemleriyle hücum, savunma, bireysel oyuncu istatistikleri gibi öznitelikler dahil edilerek daha kapsamlı çalışmalar yapılabilir. Dean Oliver kitabında basketbol başarısına etki eden faktörleri belirlemiştir ayrıca Horvat ve Job'da NBA (National Basketball Association) maçları için bir formül oluşturmuşlardır [2], [3].

Bu çalışmayı yaparken bu zamana kadar maç sonuçları, NBA oyuncu seçimleri ve En Değerli Oyuncu (MVP) seçimi için yapılmış çalışmaları incelenerek bunların üzerinden Kadınlar Basketbol Süper Ligi istatistiklerinden oluşan verilerle elde edilen veri setiyle kadın basketbol milli takımı için makine öğrenimi ile bir modelleme çalışması yapılacak ve milli takım için olabilecek en iyi oyuncu havuzunu belirleyebilen modellerin belirlenmesi sağlanacaktır.

### **1.1 Tezin Amacı ve Kapsamı**

Spor verilerinin analizi, sporun parçaları olan takımlar, oyuncular, antrenörler, menajerler ve taraftarlar için gerekli verilerin elde edilmesine olanak sağlar. Oyuncular kendi performans gelişimini görebilirken antrenörler ve takımlar kadro oluşturmada ve maça hazırlık için veri elde eder. Ayrıca günümüzde sporun popülaritesinin artmasıyla taraftarlar da artık takımın durumunu, transfer edilen ya da edilecek olan oyuncunun istatistiksel verilerini de inceliyorlar. Buna bağlı olarak bu analizlerin makine öğrenimi yöntemleriyle yapılması verilerden elde edilebileceklerin artmasını ve gelişmesini sağlar. Bu analizleri sadece takım bazında değil milli takımlar düzeyinde de yapmak ülke sporunun dünya çapında yükselmesine katkı sağlar. Bu çalışmanın amacı da ülkemizin Kadın Basketbol Milli Takımı için bir aday havuzu belirlemektir. Bu sayede sezon boyunca oyuncuların elde ettiğim istatistikler makine öğrenmesi algoritmalarıyla analiz edilerek hangi oyuncunun milli takım düzeyinde olduğunu ve aday kadroda bulunabileceğini tespit ederek antrenörlere ve yöneticilere fayda sağlanacaktır.

Çalışmamın kapsamı Türkiye Basketbol Federasyonu'nun (TBF) sitesinden ulaşılabilen Türkiye Kadınlar Basketbol Süper Ligi (KBSL) istatistiklerinden oluşturulan veri setleri üzerinden olacaktır. KBSL'de oynayan Türk oyuncuların son 4

sezonu kapsayan istatistikleri ve bu sezonlarda milli takıma seçilmiş oyuncu listelerinden veri setimiz oluşturulmuştur.

## **1.2 Geçmiş Çalışmalar**

Spor müsabakalarının popülaritesi arttıkça bu konuda yapılan çalışmalar da artmıştır. İstatistiksel veriler, maç görüntüleri, oyuncuların daha önceki istatistikleri, takım istatistikleri ve bunlar gibi birçok verinin de kayıt altına alınması ile makine öğrenmesi algoritmalarının kullanımı da spor alanında yaygınlaşmıştır. Makine öğrenimi algoritmaları sakatlıkların tespiti, NBA’ye seçilecek oyuncu, yıldız olma kapasitesine sahip oyuncu gibi alanlarda da kullanılmaya başlamıştır. Birçok spor dalında veri analizi çalışmaları yapılmaktadır. Basketbolda popüler bir spor olduğu için her lig özelinde geniş çalışmalar yapılmaktadır. Ayrıca bir takım gözetmeksizin basketbolun genel çerçeveleri içinde makine öğrenmesi kullanılarak yapılan çalışmalar mevcuttur. Ancak milli takımlar düzeyinde oyuncu belirlenmesi için yapılmış çalışmalar henüz çok yaygınlaşmış bir alan olmamaktadır. Bu konuda örnek olarak benzerlik sağlayabilecek çalışmalara NBA seçimleri için yapılmış oyuncu seçimi ya da yine NBA’de MVP seçimi için yapılan çalışmalar örnek gösterilebilir.

### **1.2.1 NBA Liginde Yapılmış Çalışmalar**

NBA ligine yönelik maç sonucu tahmini konusunda yapılan çalışmaların önemli bir kısmı makine öğrenmesi tekniklerinden yararlanmaktadır. Houde çalışmasında, NBA maç sonuçlarının tahmin edilmesi için kullanılan makine öğrenmesi yöntemlerini incelemektedir [4].

Horvat ve arkadaşları Symmetry dergisinde yayınlanan makalelerinde NBA maç sonuçlarının tahmin edilmesi için kullanılan bir veri odaklı makine öğrenimi algoritmasını sunmaktadır. Çalışmalarında gelişmiş NBA istatistiklerinden takım verimlilik değeri baz alınmıştır [3].

2022 yılında Wang ve arkadaşları, NBA oyun sisteminin yapay zekâ temelli uygulamalar ile nasıl tahmin edileceğini açıklamaktadır. Bu çalışmada özellikle “Açıklanabilir yapay zekâ (Explainable AI)” tekniklerinin, maç analizlerinde ve tahminlerinde nasıl fayda sağlayabileceğine vurgu yapılmaktadır [5].

### **1.2.2 Euroleague ve Avrupa Ligleri İçin Yapılmış Çalışmalar**

Avrupa basketbol ligleri ve özellikle EuroLeague üzerine yapılan çalışmalar da benzer biçimde makine öğrenmesi yaklaşımlarına odaklanmaktadır. Ballı ve Özdemir, hibrit özellik çıkarma ve makine öğrenmesi tekniklerini bir arada kullanarak EuroLeague maç sonuçlarını tahmin etmeye yönelik yeni bir yöntem sunmaktadır [6].

Lampis ve arkadaşları makalelerinde, Avrupa basketbol maç sonuçlarını öngörmek amacıyla kullanılan makine öğrenimi algoritmalarını incelemekte ve bu algoritmaların performanslarını değerlendirmektedir [7].

### **1.2.3 Türkiye Basketbol Ligi İçin Yapılmış Çalışmalar**

Özdemir ve Ballı 2022 yılında yaptıkları çalışmalarında, Türkiye Erkekler Basketbol Süper Ligi maç sonuçlarının tahmin edilmesi için kullanılan makine öğrenmesi yöntemlerini incelemektedir. Çalışmalarında 2017- 2018 sezonuna ait verileri kullanmışlardır ve lojistik regresyon modeli ile %96,5 doğruluk oranına ulaşmışlardır [8].

Çetin ve arkadaşları, EuroBasket 15 için Türkiye Erkek Milli Takım'ı için oyuncu seçimine yönelik çalışma yapmışlardır. Bu çalışma, millî takım için oyuncu seçimini çok sayıda kriteri dikkate alarak değerlendirmekte, ancak makine öğrenmesi algoritmalarına yer vermemektedir [9].

## **1.3 Kavramsal Tanımlar**

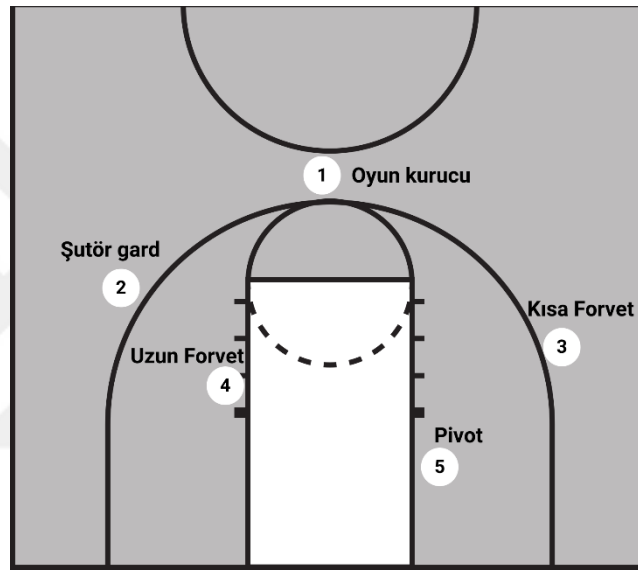
### **1.3.1 Basketbol**

Basketbol, temelinde bir topu çemberden geçirme amacı yatan 2 takım ve 5'er oyuncuyla oynanan bir spor türüdür. Olimpiyat kategorisinde bir spor dalıdır. Dünya'nın her yerinde oynanan hatta son yıllarda başka spor türleriyle birleştirilerek yeni türevleri (korfbol gibi) geliştirilmiş bir oyun türüdür. Savunma ve hücum mantığıyla hareket edilir. 10'ar dakikalık 4 periyot süren bu sporda takımlar 24 saniye içinde hem hücum etmeli hem de kendi potalarını savunmaya, sayı yememeye çalışmaktadırlar.

Basketbolda her takımdan sahada 5 oyuncu bulunur. Sahada bulunan oyuncular genellikle "Oyun kurucu (1), Şutör gard (2), Kısa forvet (3), Uzun forvet (4) ve Pivot

(5)” pozisyonlarında oynayan oyunculardır (Şekil 1.1). Oyuncuların pozisyonları oyunda bulundukları konum ve görevlerine göre sınıflandırılır. Pivot oyuncusu pota altında görev alıp potaya yakın atışlar yapıp, ribaund alma görevini üstlenirken oyun kurucu tepe noktada bulunur ve topun dağılımı, hücum setinin başlaması gibi görevler üstlenir. Farklı pozisyonlardaki oyuncular için farklı oyun metrikleri daha önemlidir.

Basketbolda çok yönlü bir oyuncuya sahip olmak takıma değer katan bir özelliktir. Eğer sizin pivot oyuncunuz 3 sayı kullanabilen de bir oyuncuysa bu takıma değer katar, ayrıca kısa forvet oyuncunuzun da içeriden sırtı dönük şekilde hücum edebilmesi savunmayı zorlayan bir durum ortaya çıkartır.



Şekil 1.1 Basketbolda oyuncu mevkileri

### 1.3.2 Makine Öğrenmesi

Makine öğrenmesi, yapay zekâ ve bilgisayar bilimi alanlarının bir dalıdır. Makine öğrenmesi, insanların öğrenme şeklini taklit etmek için verileri ve algoritmaları kullanır. Makine öğrenmesi, doğruluğunu zamanla artırmaya odaklanır. Makine öğrenmesi, çeşitli paradigmalara sahiptir, örneğin denetimli, denetimsiz, çevrimiçi, toplu, meta-öğrenme ve güçlendirilmiş öğrenmedir. Ayrıca sınıflandırma, regresyon, kümeleme, anormallik tespiti ve veri temizleme gibi çeşitli alanlarda uygulamaları vardır.

Makine öğrenmesinin birçok türü vardır ancak bu çalışmada iki türü üzerinden modelleme yöntemlerine bakacağız. Bunlar denetimli öğrenme ve denetimsiz öğrenmedir [8]. Denetimli öğrenmede verilerin sınıfları bellidir ve tahmin için verilen

verileri ait olduđu k melerde deęerlendirir. Yani algoritmalar belirli kriterlere g re performans g sterirler. Denetimsiz  ęrenmede ise veriler  zerinde eęitilen algoritmalar bir kriter g zetmeksizin performans g sterirler.

Bu  alıřmada, Rastgele Orman (RF), Naive Bayes (NB), Karar Aęacı (DT), Destek Vekt r Makineleri (SVM) ve Ekstrem Gradyan G  lendirme (XGBoost) modelleme algoritmaları kullanılmıřtır.



## 2. BASKETBOL VE İSTATİSTİKSEL VERİLER

Basketbol aslında bir olasılıklar oyunudur ve bu yüzden matematiksel hesaplamalara çok elverişlidir. Her sporda olduğu gibi bu spor dalında da önemli tek sonuç kazanıp kazanamamaktır. Ancak nihai sonuca ulaşırken ve hatta rakip takıma hazırlanırken takımlara ait özellikler, istatistiksel verilerde gizlidir.

### 2.1 Basketbolda Kullanılan İstatistiksel Veriler

Bir basketbol maçı süresince canlı olarak istatistik tutulmaktadır. Bu istatistiklerin içinde maç sayısı, oyunda kaldığı süre, 2 sayı atışı, 2 sayı atış yüzdesi, 3 sayı atış sayısı, 3 sayı atış yüzdesi, toplam atış sayısı, toplam atış yüzdesi, serbest atış sayısı, serbest atış yüzdesi, hücum ribaundu sayısı, savunma ribaundu sayısı, toplam ribaund sayısı, asist sayısı, top çalma sayısı, blok sayısı, top kaybı sayısı, yaptığı faul sayısı, +/- ve verimlilik puanları yer alır.

| No | Basketbolcu                                                                                                 | Maç | SÜ       | SY  | AG                 | ZS                 | 3S                 | SA                | SR | HR | TR  | AS  | BL | TÇ | TK | FA | +/- | VP  |
|----|-------------------------------------------------------------------------------------------------------------|-----|----------|-----|--------------------|--------------------|--------------------|-------------------|----|----|-----|-----|----|----|----|----|-----|-----|
| 1  |  GAMZE TAKMAZ            | 14  | 03:56:12 | 42  | 14 / 32<br>43,75%  | 4 / 9<br>44,44%    | 10 / 23<br>43,48%  | 4 / 8<br>50,00%   | 15 | 3  | 18  | 12  | 2  | 4  | 10 | 37 | 36  | 46  |
| 3  |  JESSICA LAUREN THOMAS   | 14  | 08:55:33 | 253 | 86 / 199<br>43,22% | 48 / 97<br>49,48%  | 38 / 102<br>37,25% | 43 / 50<br>86,00% | 51 | 7  | 58  | 112 | 0  | 10 | 40 | 30 | 63  | 273 |
| 5  |  CEREN AKPINAR           | 5   | 00:05:20 | 0   | 0 / 1<br>0,00%     | 0 / 1<br>0,00%     | 0 / 0<br>0,00%     | 0 / 0<br>0,00%    | 0  | 0  | 0   | 1   | 0  | 0  | 1  | 2  | -1  | -1  |
| 7  |  SEHERNAZ ÇİDAL          | 14  | 05:53:28 | 83  | 28 / 80<br>35,00%  | 11 / 23<br>47,83%  | 17 / 57<br>29,82%  | 10 / 15<br>66,67% | 20 | 10 | 30  | 21  | 2  | 9  | 12 | 32 | 43  | 76  |
| 12 |  İNCİ GÜÇLÜ              | 10  | 01:22:54 | 27  | 12 / 18<br>66,67%  | 12 / 18<br>66,67%  | 0 / 0<br>0,00%     | 3 / 8<br>37,50%   | 15 | 8  | 23  | 2   | 0  | 1  | 9  | 11 | -49 | 33  |
| 14 |  EDA BALKAN              | 2   | 00:04:00 | 2   | 1 / 4<br>25,00%    | 1 / 4<br>25,00%    | 0 / 0<br>0,00%     | 0 / 0<br>0,00%    | 1  | 0  | 1   | 0   | 0  | 1  | 0  | 0  | 1   | 1   |
| 17 |  FERAY LAİÇ              | 4   | 00:07:28 | 4   | 1 / 1<br>100,00%   | 1 / 1<br>100,00%   | 0 / 0<br>0,00%     | 2 / 2<br>100,00%  | 2  | 1  | 3   | 0   | 1  | 0  | 0  | 2  | -1  | 8   |
| 20 |  VICTORIA AYO MACAULAY   | 14  | 07:56:49 | 227 | 87 / 180<br>48,33% | 68 / 128<br>53,13% | 19 / 52<br>36,54%  | 34 / 47<br>72,34% | 88 | 16 | 104 | 27  | 23 | 26 | 47 | 51 | 89  | 254 |
| 21 |  NİMET ÇALIŞKAN          | 1   | 00:01:18 | 0   | 0 / 0<br>NaN%      | 0 / 0<br>NaN%      | 0 / 0<br>0,00%     | 0 / 0<br>0,00%    | 0  | 0  | 0   | 0   | 0  | 0  | 1  | 0  | -3  | -1  |
| 22 |  BERFİN SERTOĞLU         | 12  | 04:39:25 | 84  | 31 / 75<br>41,33%  | 24 / 46<br>52,17%  | 7 / 29<br>24,14%   | 15 / 23<br>65,22% | 35 | 12 | 47  | 36  | 1  | 13 | 18 | 34 | 66  | 111 |
| 24 |  ROBYN LASHAE PARKS      | 14  | 08:18:43 | 195 | 75 / 179<br>41,90% | 49 / 100<br>49,00% | 26 / 79<br>32,91%  | 19 / 25<br>76,00% | 81 | 21 | 102 | 42  | 2  | 23 | 20 | 21 | 47  | 234 |
| 45 |  ASTOU BARRO NDOUR GUEYE | 11  | 05:53:16 | 189 | 76 / 133<br>57,14% | 69 / 111<br>62,16% | 7 / 22<br>31,82%   | 30 / 39<br>76,92% | 67 | 35 | 102 | 16  | 13 | 10 | 12 | 33 | 84  | 252 |
| 88 |  NUR NİHAN DABAKOĞLU     | 2   | 00:15:34 | 10  | 3 / 6<br>50,00%    | 0 / 1<br>0,00%     | 3 / 5<br>60,00%    | 1 / 4<br>25,00%   | 4  | 1  | 5   | 1   | 0  | 1  | 1  | 1  | -5  | 10  |

Şekil 2.1 Örnek bir istatistik belgesi [10]

Şekil 2.1’de verilen istatistik örneğinde bulunan verilerinin ne anlama geldiği aşağıda detaylı olarak açıklanmıştır.

**Maç:** Oyuncunun süre aldığı maç sayısı

**SÜ:** Oyuncunun oynadığı toplam süre

**SY:** Oyuncunun attığı toplam sayı

**AG:** Oyuncunun attığı toplam atış sayısı

**İki sayılık atış sayısı:** Bir oyuncunun kaç tane iki sayılık atış kullandığını ve kaç tanesini sayıya çevirdiğini gösterir.

**İki sayılık şut yüzdesi:** Bir oyuncunun iki sayılık şutlardan kaç tanesini yaptığına ve kaç tanesini kaçırdığına bakarak hesaplanır.

**Üç sayılık atış sayısı:** Bir oyuncunun kaç tane üç sayılık atış kullandığını ve kaç tanesini sayıya çevirdiğini gösterir.

**Üç sayılık şut yüzdesi:** Bir oyuncunun üç sayılık şutlardan kaç tanesini yaptığına ve kaç tanesini kaçırdığına bakarak hesaplanır.

**Şut yüzdesi:** Bir oyuncunun toplam şutlardan kaç tanesini sayı yaptığına ve kaç tanesini kaçırdığına bakarak hesaplanır.

**Serbest atış sayısı:** Bir oyuncunun kaç tane serbest atış kullandığını ve kaç tanesini sayıya çevirdiğini gösterir.

**Serbest atış yüzdesi:** Bir oyuncunun serbest atışlardan kaç tanesini yaptığına ve kaç tanesini kaçırdığına bakarak hesaplanır.

**Toplam ribaund:** Bir oyuncunun veya takımın toplam ribaund sayısına denir.

**Savunma ribaundu:** Kaçan şutun savunma tarafından sahip olunmasına denir.

**Hücum ribaundu:** Sayı olmayan bir atış sonrası çemberden seken topa hücum oyuncusunun sahip olmasına denir.

**Asist:** Bir oyuncunun, bir takım arkadaşının sayı olan bir hücumda, sayı atması için verdiği son pasa denir.

**Top çalma:** Bir oyuncunun rakip takımın topunu çalmasıdır.

**Blok:** Bir oyuncunun rakip takımın şutunu engellemesidir.

**Top kaybı:** Bir oyuncunun sahip olduğu bir topu kaybetmesidir bu fazla adım olabilir, dışarı atma olabilir ya da rakip oyuncunun elinden alması şeklinde olabilir.

**Faul:** Bir oyuncunun rakip takımın oyuncusuna karşı kural ihlali yapmasıdır.

**+/- :** Oyuncunun oynadığı dakika boyunca oyuna sağladığı katkı

**VP:** Oyuncunun oynadığı maçlarda elde ettiği verimlilik puanı

## 2.2 Basketbolda İstatistiksel Verilerin Önemi

Basketbol, karmaşık bir takım sporu olup, oyuncuların bireysel ve takım performanslarını değerlendirmek için istatistiksel verilerden yararlanmak önemlidir. Bu veriler, oyuncuların ve takımların güçlü ve zayıf yönlerini belirlemek, oyun stratejilerini geliştirmek ve oyuncu gelişimini izlemek için kullanılabilir.

Basketbolda kullanılan istatistiksel veriler, bireysel oyuncu istatistikleri ve genel takım istatistikleri olarak tutulabilmektedir.

**Kişisel İstatistikler:** Bir oyuncunun bireysel performansını değerlendiren verilerdir. Örneğin, sayı, ribaund, asist, top çalma, blok, faul ve iç saha atışı yüzdesi gibi veriler bu kategoriye girer.

**Takım İstatistikleri:** Bir takımın performansını değerlendiren verilerdir. Örneğin, sayı farkı, ribaund farkı, asist farkı, top çalma farkı, blok farkı ve serbest atış yüzdesi gibi veriler bu kategoriye girer.

Basketbolda istatistiksel verilerin önemi şu şekilde özetlenebilir:

- Oyuncuların ve takımların güçlü ve zayıf yönlerini belirlemek için kullanılabilir. Örneğin, bir oyuncunun sayı ve ribaund ortalamaları yüksekse, onun bir skorere ve rebounder olduğunu söyleyebiliriz. Bir takımın top çalma ve blok ortalamaları yüksekse, onun savunmada etkili olduğunu söyleyebiliriz.
- Oyun stratejilerini geliştirmek için kullanılabilir. Örneğin, bir takım top çalma konusunda başarılı ise, savunmada pres taktiği uygulayabilir. Bir takım ribaund konusunda başarılı ise, hızlı hücum taktiği uygulayabilir.
- Oyuncu gelişimini izlemek için kullanılabilir. Örneğin, bir oyuncunun sayı ortalamasının artması, onun gelişiminin olumlu bir göstergesidir. Bir takımın ribaund ortalamasının düşmesi, savunmada bir sorun olabileceğini gösterir.

Basketbolda istatistiksel verilerin kullanımı, son yıllarda giderek artmaktadır. Bu durum, basketbolun daha profesyonel bir spor haline gelmesi ve istatistiklerin öneminin daha iyi anlaşılması ile açıklanabilir.

### 3. YÖNTEM

#### 3.1 Veri Setinin Elde Edilmesi

Bu çalışmada kullanılan veri seti Türkiye Basketbol Federasyonu resmi internet sitesinden alınmıştır [10]. Veriler Python'ın BeautifulSoup kütüphanesi kullanılarak internet sitesi üzerinden çekilmiştir. KBSL takımlarının son 4 yıla ait istatistikleri takımlar bazında yabancı oyuncular, oyuncu görselleri çıkartılarak Excel dosyasına kaydedilmiştir. Çekilen veriler 18 tane öznitelik içermektedir. Bu özniteliklerden tek sütunda bulunan toplam başarılı atış, toplam atış girişimi, toplam başarılı 2 sayılık atış sayısı, toplam 2 sayılık atış sayısı girişimi, toplam başarılı 3 sayılık atış sayısı, toplam 3 sayılık atış sayısı girişimi, toplam başarılı serbest atış sayısı, toplam serbest atış sayısı girişimi olarak sütunlara ayrılmıştır. Sezon, Pozisyon ve Milli Takım Durumu (MT) sütunları da verilerin ait olduğu sezon, oyuncunun oynadığı pozisyon ve oyuncunun o sezonda milli takım aday kadrosuna çağırılıp çağırılmadığı bilgisini içermektedir. Bu özniteliklere ait veriler yine Türkiye Basketbol Federasyonu'nun internet sitesinden alınmıştır. Ayrıca 25 tane daha yeni öznitelik oluşturulmuştur. Bu öznitelikler sayı, ribaund istatistiklerinin süre ve oynanan maç sayısına oranına ve sayı istatistiklerinin toplam sayıya oranına göre elde edilmiştir. Bu veriler de Excel dosyasına eklenmiştir. Bütün verilerin eklenmesi sonucu 534 satır ve 50 sütundan oluşan bir veri seti oluşturulmuştur. Veri setinde bulunan bütün öznitelikler ve açıklamaları Tablo 3.1'de verilmiştir. "MT" sütunu hedef sütun olarak belirlenmiştir. Veri setinde kategorik, tam sayı ve ondalıklı sayı veri türlerinde değerler bulunmaktadır.

Tablo 3.1 Öznitelikler ve Açıklamaları

| ÖZNİTELİK                  | AÇIKLAMA                         | VERİ TÜRÜ     |
|----------------------------|----------------------------------|---------------|
| <b>Basketbolcu</b>         | Basketbolcunun Adı-Soyadı        | Kategorik     |
| <b>Pozisyon</b>            | Oyuncunun oynadığı pozisyon      | Tam sayı(int) |
| <b>MT (Hedef Değişken)</b> | Milli takımda oynayıp oynamadığı | Kategorik     |
| <b>Maç</b>                 | Oynadığı maç sayısı              | Tam sayı(int) |
| <b>SÜ</b>                  | Oynadığı Süre                    | Tam sayı(int) |
| <b>SY</b>                  | Attığı toplam sayı               | Tam sayı(int) |
| <b>AG-S</b>                | Sayı olan toplam atış            | Tam sayı(int) |
| <b>AG-A</b>                | Toplam atış sayısı               | Tam sayı(int) |
| <b>2S-S</b>                | Sayı olan 2 sayılık atış         | Tam sayı(int) |
| <b>2S-A</b>                | Toplam 2 sayılık Atış sayısı     | Tam sayı(int) |
| <b>3S-S</b>                | Sayı olan 3 sayılık atış         | Tam sayı(int) |
| <b>3S-A</b>                | Toplam 3 sayılık Atış sayısı     | Tam sayı(int) |
| <b>SA-S</b>                | Sayı olan serbest atış sayısı    | Tam sayı(int) |
| <b>SA-A</b>                | Toplam serbest Atış sayısı       | Tam sayı(int) |
| <b>SR</b>                  | Savunma ribaundu                 | Tam sayı(int) |
| <b>HR</b>                  | Hücum ribaundu                   | Tam sayı(int) |
| <b>TR</b>                  | Toplam ribaund                   | Tam sayı(int) |
| <b>AS</b>                  | Asist                            | Tam sayı(int) |
| <b>BL</b>                  | Blok                             | Tam sayı(int) |
| <b>TÇ</b>                  | Top çalma                        | Tam sayı(int) |
| <b>TK</b>                  | Top Kaybı                        | Tam sayı(int) |
| <b>FA</b>                  | Oyuncunun Yaptığı Faul           | Tam sayı(int) |
| <b>+/-</b>                 | Oynadığı dakikada oyuna katkısı  | Tam sayı(int) |
| <b>VP</b>                  | Toplam verimlilik puanı          | Tam sayı(int) |
| <b>Sezon</b>               | İstatistiğin ait olduğu sezon    | Tam sayı(int) |

Basketbola ait istatistiklere ait öznitelikler haricinde oluşturulan 25 özniteliğin açıklamaları Tablo 3.2’de verilmiştir. Tablo 2’de yer alan öznitelikler orantısal değerler içermektedir. Bu oranlar süre ve maç sayısına göre ribaund ve sayı değerlerinin oranlarıdır.

Tablo 3.2 Sabit istatistikler harici oluşturulan öznitelikler ve açıklamaları

| ÖZNİTELİK | AÇIKLAMA                                               | VERİ TÜRÜ             |
|-----------|--------------------------------------------------------|-----------------------|
| AG-S/SÜ   | Sayı olan toplam atış sayısının süreye oranı           | Ondalıklı sayı(float) |
| AG-A/SÜ   | Toplam atış sayısının süreye oranı                     | Ondalıklı sayı(float) |
| 2S-S/SÜ   | Sayı olan toplam 2 sayılık atış sayısının süreye oranı | Ondalıklı sayı(float) |
| 3S-S/SÜ   | Sayı olan toplam 3 sayılık atış sayısının süreye oranı | Ondalıklı sayı(float) |
| SA-S/SÜ   | Sayı olan toplam serbest atış sayısının süreye oranı   | Ondalıklı sayı(float) |
| SA-A/SÜ   | Toplam serbest atış sayısının süreye oranı             | Ondalıklı sayı(float) |
| SR/SÜ     | Savunma ribaundu sayısının süreye oranı                | Ondalıklı sayı(float) |
| HR/SÜ     | Hücum ribaundu sayısının süreye oranı                  | Ondalıklı sayı(float) |
| TR/SÜ     | Toplam ribaund sayısının süreye oranı                  | Ondalıklı sayı(float) |
| SY/SÜ     | Toplam sayının süreye oranı                            | Ondalıklı sayı(float) |
| AG-S/Maç  | Sayı olan toplam atış sayısının maç sayısına oranı     | Ondalıklı sayı(float) |
| 2S-S/Maç  | Sayı olan 2 sayılık atış sayısının maç sayısına oranı  | Ondalıklı sayı(float) |
| 3S-S/Maç  | Sayı olan 3 sayılık atış sayısının maç sayısına oranı  | Ondalıklı sayı(float) |
| SA-S/Maç  | Sayı olan serbest atış sayısının maç sayısına oranı    | Ondalıklı sayı(float) |
| SA-A/Maç  | Toplam serbest atış sayısının maç sayısına oranı       | Ondalıklı sayı(float) |
| SY/Maç    | Toplam sayının maç sayısına oranı                      | Ondalıklı sayı(float) |
| AG-A/Maç  | Toplam atış sayısının maç sayısına oranı               | Ondalıklı sayı(float) |
| 3S-A/Maç  | Toplam 3 sayılık atış sayısının maç sayısına oranı     | Ondalıklı sayı(float) |
| 2S-A/Maç  | Toplam 2 sayılık atış sayısının maç sayısına oranı     | Ondalıklı sayı(float) |
| TR/Maç    | Toplam ribaund sayısının maç sayısına oranı            | Ondalıklı sayı(float) |
| HR/Maç    | Hücum ribaundu sayısının maç sayısına oranı            | Ondalıklı sayı(float) |
| SR/Maç    | Savunma ribaundu sayısının maç sayısına oranı          | Ondalıklı sayı(float) |
| AG-A/SY   | Toplam atış sayısının toplam sayıya oranı              | Ondalıklı sayı(float) |
| 3S-A/SY   | Toplam 3 sayılık atış sayısının toplam sayıya oranı    | Ondalıklı sayı(float) |
| 2S-A/SY   | Toplam 2 sayılık atış sayısının toplam sayıya oranı    | Ondalıklı sayı(float) |

### 3.2 Veri Setinin Temizlenmesi ve Normalizasyonu

Veri seti genel hatlarıyla internet çekilerek ve yeni öznitelikler ile oluşturulduktan sonra bu veriler üzerinde temizleme ve normalizasyon işlemleri yapılmıştır. Oluşturulan veri setinde hedef sütun “MT” sütunudur ve kategorik değer içermektedir. “MT” sütunun değerleri “EVET” ve “HAYIR” olarak tutulmaktadır. Toplam 533 satır bulunmaktadır bunların 425 tanesi “HAYIR”, 108 tanesi “EVET” olarak

işaretlenmiştir. Veri setinde bulunan 50 öznitelik sütununun özellikleri ve açıklamaları Tablo 3.1 ve Tablo 3.2’de verilmiştir.

Tablo 3.1’de belirtilen özniteliklerden hedef sütun olan “MT” sütunu kategorik veri içerdiği için “Label Encoder” kullanılarak 0 ve 1 olarak dönüştürülmüştür. Özniteliklerden “MT”, “Basketbolcu” ve “Sezon” sütunları veri setinden çıkartılarak kodun hatalı çalışmasına sebep olacak durumlar önlenmiştir. Veri setinde eksik bilgi olmadığı için boş değerler için bir temizleme işlemi yapılmamıştır. Veri setinde bulunan tam sayı (int) ve ondalıklı sayılar (float) sistemin doğru çalışması için “MinMaxScaler” kullanılarak [0,1] aralığında normalize edilmiştir. Normalizasyon, veri setimizde bulunan değerlerin belirli bir aralıkta ölçeklenmesini sağlar. Bu şekilde model çalışırken çok büyük değerlere sahip özniteliklerin baskın çıkması engellenmiş olur. Çalışmanın devamında Principal Component Analysis (PCA) kullanıldığı için normalizasyon uygulanması değeri büyük öznitelik bileşenine olan yönelimi engeller.

### 3.3 Veri Setine PCA Uygulanması

PCA, veri setinde bulunan özniteliklerin hangisinin modellemede ağırlığı daha fazla olan özniteliklerin belirlenmesi ve mümkün olan az sayıda bileşenle fazla bilgiyi korumak için kullanılır. Ayrıca makine öğrenmesi algoritmaları da daha hızlı ve kolay çalışabilir. PCA ile boyut indirgemesi yapıldığında veri setinde bulunan değişkenler arasında bulunan kovaryans ve korelasyon yapısı incelenir, en fazla varyansa sahip değişkenler “temel bileşenler” olarak belirlenir [11]. Bu çalışmada PCA kovaryansları olarak “0,90, 0,95, 0,99 ve PCA yok” tanımlanmıştır. Bu eşiğin anlamı, “veri setinin toplam varyansının %90, %95 veya %99’unu açıklayacak kadar bileşen” seçilmesidir. Farklı varyans değerlerinin seçilmesi, bilgi kaybını, hesaplama hızını ve modelin performansını doğrudan etkilemektedir. Tablo 3.3’te her bir varyans eşiği ile elde edilen temel bileşen sayıları bulunmaktadır.

Tablo 3.3 PCA ile elde edilen bileşen sayıları

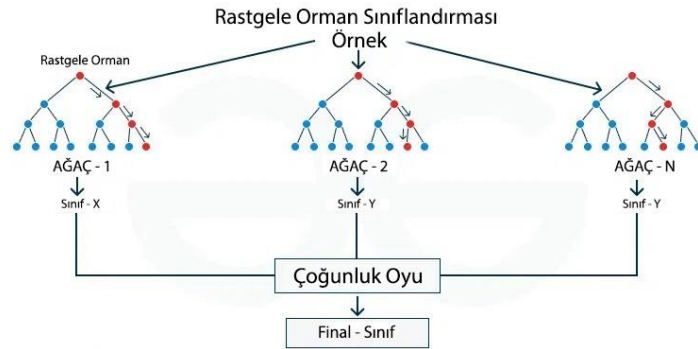
| PCA KÜMÜLATİF VARYANS DEĞERİ | ELDE EDİLEN TEMEL BİLEŞEN SAYISI |
|------------------------------|----------------------------------|
| PCA yok                      | 50                               |
| %90                          | 9                                |
| %95                          | 13                               |
| %99                          | 22                               |

### 3.4 İstatistiksel Verilerin Değerlendirilmesi İçin Kullanılacak Sınıflandırma Modellerinin Belirlenmesi

Milli takım aday havuzunun belirlenmesinde sınıflandırma modelleri kullanılmıştır. Bu sınıflandırma modellerinin performanslarının karşılaştırılması ile en iyi model belirlenmiştir. Modeller doğruluk ve F1 skor değerine göre değerlendirilmiştir. Çalışmada yer alan sınıflandırma modelleri; Rastgele Orman (Random Forest), Naive Bayes, Karar Ağacı (Decision Tree), Destek Vektör Makineleri (Support Vector Machines) ve Ekstrem Gradyan Güçlendirme (XGBoost) olarak seçilmiştir. Modeller belirlenirken yapılmış olan diğer çalışmalar da incelenmiştir. Spor analitiğinde yaygın olarak kullanılan modeller seçilmiştir [7], [12], [13], [14].

#### 3.4.1 Rastgele Orman

Rastgele orman makine öğrenmesinde hem regresyon problemlerinin hem de sınıflandırma problemlerinin çözümünde kullanılan bir makine öğrenmesi algoritmasıdır. Birden çok karar ağacı oluşturur, her bir karar ağacı rastgele bir sınıfla eğitilir ve sınıflandırma için bir tahminde bulunur ve temelinde en yüksek puanı alan değer seçilir (Şekil 3.1). Bu rastgelelik, modelin aşırı öğrenmesini azaltır ve genelleme yeteneğini artırır.



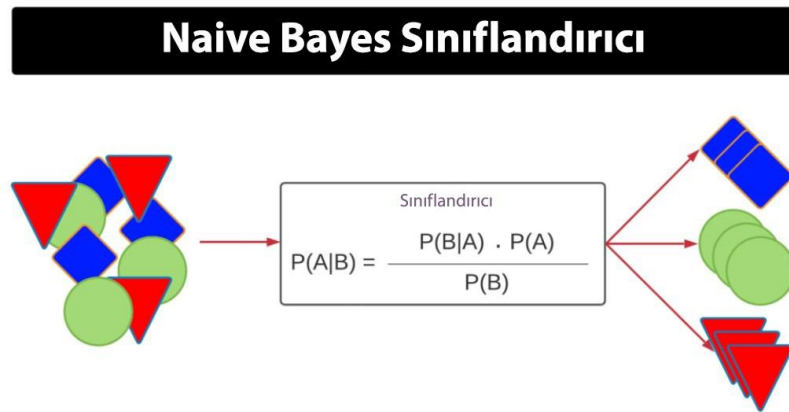
Şekil 3.1 Rastgele Orman Sınıflandırıcısı [15]

RF hem doğrusal hem de doğrusal olmayan problemler için etkili bir şekilde kullanılabilir, ancak özellikle doğrusal olmayan ilişkilerin olduğu durumlarda daha büyük avantaj sağlar. Karar ağaçları, verileri özyinelemeli olarak bölerek doğrusal olmayan ilişkileri modelleyebilir, özellikler arasındaki etkileşimleri otomatik olarak

yakalayabilir ve aşırı öğrenmeye dirençli olmasından kaynaklı olarak doğrusal olmayan problemler daha iyi başa çıkabilir.

### 3.4.2 Naive Bayes

Naive Bayes algoritması Bayes Teoremi'ne dayanan bir sınıflandırma algoritmasıdır. Algoritmanın çalışma temelinde her eleman için olasılığını Bayes Teoremi kullanarak hesaplar ve sonunda en yüksek çıkan değere göre sınıflandırmayı gerçekleştirir. Naive Bayes Algoritması, özellikle büyük veri kümeleri ve metin sınıflandırma problemleri için uygundur. Bu algoritmanın esnekliği ve performansı, çeşitli alanlarda geniş bir uygulama yelpazesi sunar. Naive Bayes Algoritması, genellikle doğrusal olmayan problemler yerine doğrusal problemlerde daha iyi performans gösterir. Bu algoritmanın temel varsayımı, özelliklerin birbirinden bağımsız olduğudur. Her bir özelliğin diğerlerinden bağımsız olarak sınıf etiketini belirlemesine olanak tanır. Bu varsayım, doğrusal ilişkilerin bulunduğu veri setlerinde daha doğru sonuçlar üretir.



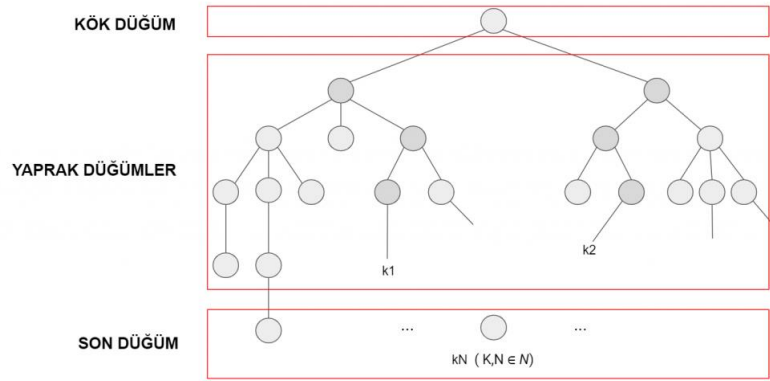
Şekil 3.2 Naive Bayes Sınıflandırma [16]

Naive Bayes algoritmasının birden fazla çeşidi bulunmaktadır. Yaygın kullanılan NB türleri arasında Bernouli Naive Bayes, Gaussian Naive Bayes, Çok terimli Naive Bayes bulunmaktadır. Bu çalışmada Gaussian Naive Bayes türü seçilmiştir.

### 3.4.3 Karar Ağacı

Karar ağaçları isminden de belli olduğu gibi bir ağaç yapısı gibi çalışırlar. Bu ağaç yapısının en tepesinde kök düğüm bulunur buradan itibaren her bir durumu bir ağaç yaprağında belirtecek şekilde dallara ayrılır. Bütün durumlar için bir yaprak oluşturulur, bu işlem tüm örnekler sınıflandırılana kadar devam eder. En sonunda artık

bölünme olmayan düğüm bulunur buna da son düğüm denir (Şekil 3.3). Karmaşık problemlerin hızlı ve kolay bir şekilde sınıflandırılabilmesi açısından önemlidir. İnsan mantığına yakın bir algoritma olduğu için sonuçların yorumlanabilmesini de kolaylaştırır.



Şekil 3.3 Karar Ağacı Sınıflandırma [17]

#### 3.4.4 Destek Vektör Makineleri

Sınıflandırma ve regresyon problemlerinde yaygın olarak kullanılan bu makine öğrenimi algoritmasıdır. Farklı sınıflara ait örneklerin arasında bir hiper düzlem oluşturur. Bu boşluğu mümkün olan en geniş boyutta tutmaya çalışır ve bu düzleme en yakın noktalara da destek vektörleri ismini verir.

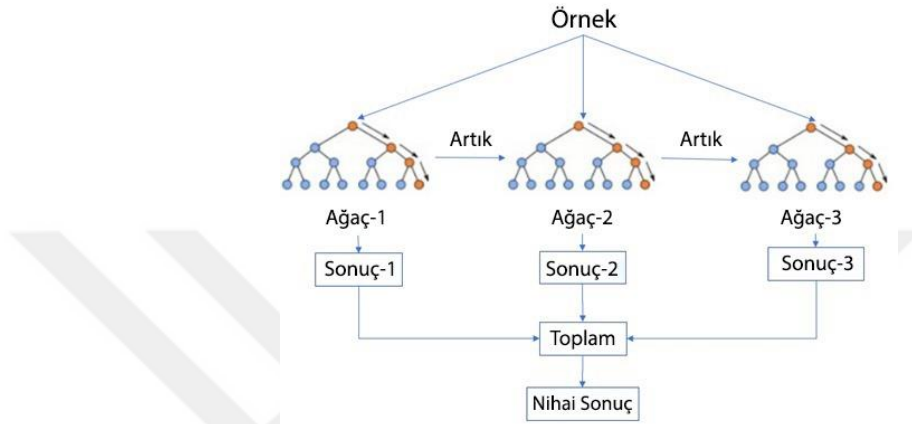
Destek Vektör Makineleri (SVM), birden çok özelliğin birbirleriyle olan bağımlılık durumlarını analiz etmede genellikle doğrudan etkili değildir. SVM'nin temel varsayımı, verinin yüksek boyutlu bir uzaya projekte edilerek doğrusal bir hiper düzlemle ayrılabilceğidir. Bu nedenle, SVM'nin amacı, veriyi sınıflandırmak için en iyi hiper düzlemi bulmaktır, özellikler arasındaki bağımlılıkları doğrudan modellemek konusunda iyi sonuçlar vermeyebilir.

#### 3.4.5 Ekstrem Gradyan Güçlendirme

Temelinde karar ağaçları modelini kullanan ancak bunları birleştirerek daha güçlü bir algoritma geliştiren makine öğrenimi algoritmasıdır. Bu algorithmada oluşan her yeni ağaç bir önceki ağaçta bulunan hataları düzeltir ve bu şekilde modelin performansını

arttırır. Aşırı öğrenme problemini kendi içinde çözen bir algoritma olduğu için yaygın olarak kullanılır ve yüksek doğrulukla sonuçlar verilir [13].

XGBoost, doğrusal olmayan problemlerle daha iyi çalışır. Karar ağaçları temelinde çalıştığı için, özellikler arasındaki doğrusal olmayan ilişkileri başarılı bir şekilde modelleyebilir. XGBoost, gradient boosting yöntemiyle bu yeteneği daha da geliştirir ve karmaşık veri setlerinde yüksek performans gösterir.



Şekil 3.4 XGBoost Sınıflandırma [18]

### 3.5 Örnekleme Modellerinin Belirlenmesi

Veri setlerinde bir sınıfa ait veriler diğer sınıflara göre daha az sayıda bulunuyorsa bu veri setlerine dengesiz veri setleri adı verilir. Bu şekilde azınlık verilerin bulunduğu dengesiz veri setlerinde modellerin doğru şekilde eğitilebilmesi için çeşitli örnekleme yöntemleri kullanılır. Bu yöntemler sentetik veri kullanarak, azınlık veriden rastgele arttırım yapma ya da çoğunluk sınıfından rastgele azaltmalar yapılmasıdır. Bu yöntemler ile model azınlık olan veriyi de iyi şekilde öğrenecek ve daha doğru sınıflandırmalar yapması sağlanacaktır. Bu çalışmada, Sentetik Azınlık Aşırı Örnekleme (SMOTE), Borderline SMOTE, Rastgele Aşırı Örnekleme (ROS), Rastgele Alt Örnekleme (RUS) ve RUS%5 + SMOTE örnekleme yöntemleri seçilmiştir. Bu yöntemlerin seçilmesindeki sebep dengesiz bir veriye sahip olmasıdır. Hedef sütun olan “MT” sütununda 425 adet çoğunluk (HAYIR), 108 tane azınlık (EVET) verisi bulunmaktadır. Örnekleme yöntemlerinin her biri ile modeller eğitilmiş ve sonuçları kaydedilmiştir.

### 3.5.1 Sentetik Azınlık Aşırı Örneklemeye

Sentetik Azınlık Aşırı Örneklemeye (SMOTE), örneklemeye yönteminde amaç azınlık sınıfında bulunan verilere benzer sentetik veriler oluşturmaktır. Oluşturulan sentetik veriler ile veri setindeki dengesizlik ortadan kaldırılır ve model eşit sayıda sınıf örneği ile eğitilir. Sentetik veriler, azınlık sınıfının örneklerini alır ve her örneğin en yakın komşusunu bulur ve bu örnek ile azınlık sınıfındaki örnek arasındaki farkı rastgele bir değer ile çarpar ve yeni sentetik bir örnek oluşturur. Oluşan yeni örnekler var olan örneklerin arasında bir değer olur.

### 3.5.2 Rastgele Aşırı Örneklemeye

Azınlık sınıfında bulunan değerler rastgele olarak seçilir. Bu verilere ait değerler veri setine tekrar eklenir. Bu ekleme işlemi aradaki dengesizlik en aza indirgenene kadar devam eder. Kullanımı kolay bir yöntem olsa da aşırı öğrenme riski en yüksek yöntemlerden birisidir. Çünkü sürekli aynı verilerin kullanılması durumu oluşur.

### 3.5.3 Borderline SMOTE

SMOTE yönteminin bir türüdür. Bu örneklemeye yönteminde önce azınlık sınıfındaki veriler kendi içinde riskli, gürültülü ve güvenli olarak sınıflandırılır. Bu sınıflandırmada riskli olan veriler çoğunluk sınıfı verilerine yakın, yanlış sınıflandırmaya sebep olabilecek verilerdir. Bu yöntem bu riskli veriler üzerinden sentetik veri üretir ve böylece modelin aşırı öğrenme riski azaltılmış olur.

### 3.5.4 Rastgele Alt Örneklemeye

RUS örneklemeye yöntemi rastgele olarak çoğunluk verilerinden veri azaltma işlemi yapar. ROS yönteminde olduğu gibi bu yöntemde de azaltma işlemi çoğunluk verilerinin sayısı azınlık verilerinin sayısına yakın olana kadar devam eder. Bu yöntemde veri kaybı çok olduğu için değerli verilerin kaybolma riski çoktur.

### 3.5.5 RUS%5 + SMOTE

Hem RUS hem SMOTE yöntemlerinin kombine olarak uygulandığı bir yöntemdir. Öncelikle RUS ile çoğunluk verileri %5 oranında azaltılır daha sonra SMOTE yöntemi ile azınlık verilerinden sentetik veri üretimi gerçekleştirilir. Bu kombine yöntem ile

veri seti dengelenmiş olur. Veri kaybını azaltıp sentetik veriler ile de model daha iyi eğitilir.

### 3.6 Değerlendirme Ölçütlerinin Belirlenmesi

Sınıflandırma algoritmaları üzerinden yapılan tahminlerde sadece doğruluk oranına bakmak her zaman yeterli olmayabilir. Bunun için Accuracy (doğruluk), F-measure (F-Skoru) ölçütleri kullanılmıştır. Tabakalı K-Katlı (10 k katlı çapraz doğrulama) ile model kararlılığı artırılmıştır.

**Doğruluk:** Doğruluk, bir sınıflandırma modelinin performansını değerlendirmek için en yaygın kullanılan ölçütlerden biridir. Modelin doğru sınıflandırdığı örnek sayısının toplam örnek sayısına oranı şeklinde tanımlanır. Örnekle açıklamak gerekirse, modelin tahmin ettiği sınıf etiketlerinin gerçek sınıf etiketleriyle ne kadar uyumlu olduğunu gösterir.

$$\text{Doğruluk} = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

TP (True Positive): Doğru pozitifler

TN (True Negative): Doğru negatifler

FP (False Positive): Yanlış pozitifler

FN (False Negative): Yanlış negatifler

**F-Skoru:** F1 skoru, sınıflandırma modellerindeki başarıyı değerlendirirken dengeyi yakalayan, doğrulukla birlikte geri çağırma'yı da göz önünde bulunduran bir ölçüttür. Hassasiyet ve duyarlılık değerlerinin armonik ortalaması olarak tanımlanır. Hassasiyet, modelin pozitif olarak sınıflandırdığı örneklerin gerçekte pozitif olan örnekler oranını, hatırlama ise pozitif olan tüm örneklerin model tarafından pozitif olarak sınıflandırılan örnekler oranını ifade eder. F-Skoru, hassasiyet ve hatırlama arasındaki dengeyi ölçerek, modelin hem yanlış pozitifleri hem de yanlış negatifleri minimize etmedeki başarısını değerlendirir. Özellikle, sınıf dengesizliği olan veri setlerinde F-Skoru, doğruluk ölçütüne göre daha iyi bir performans göstergesi olabilir.

$$F_1 = 2 \times \frac{\text{Hassasiyet (Precision)} \times \text{Duyarlılık (Recall)}}{\text{Hassasiyet (Precision)} + \text{Duyarlılık (Recall)}}$$

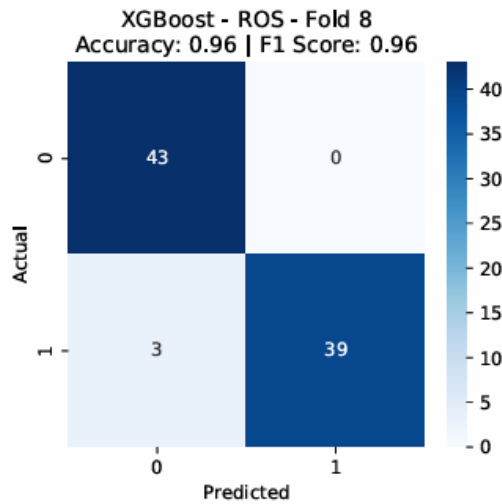
$$\text{Hassasiyet (Precision)} = \frac{TP}{TP+FP}$$

$$\text{Duyarlılık (Recall)} = \frac{TP}{TP+FN}$$

**Karmaşıklık Matrisi:** Karmaşıklık matrisi, bir sınıflandırma modelinin performansını detaylı bir şekilde analiz etmek için kullanılan bir tablodur. Matris, modelin gerçek sınıf etiketlerini tahmin ettiği sınıf etiketlerine göre sınıflandırdığı örneklerin sayısını gösterir. Karmaşıklık matrisi, doğru pozitifler, yanlış pozitifler, doğru negatifler ve yanlış negatifler gibi bilgileri içerir. Bu bilgiler sayesinde, modelin hangi sınıfları karıştırdığı, hangi sınıfları doğru bir şekilde sınıflandırdığı ve hangi sınıfları sınıflandırmakta zorlandığı gibi sorulara cevap bulunabilir.

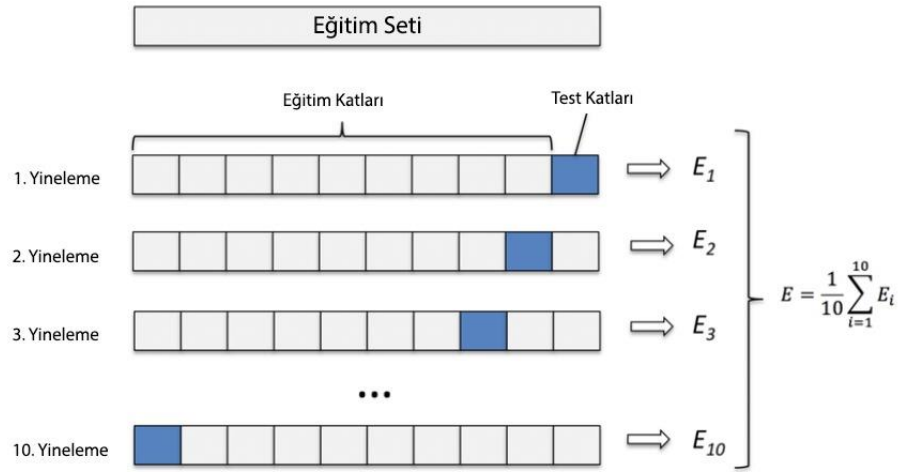
Tablo 3.4 Karmaşıklık matrisi

|        |   | Tahmin |    |
|--------|---|--------|----|
|        |   | 0      | 1  |
| Gerçek | 0 | TN     | FP |
|        | 1 | FN     | TP |



Şekil 3.5 Çalışmanın karmaşıklık matrislerinden örnek

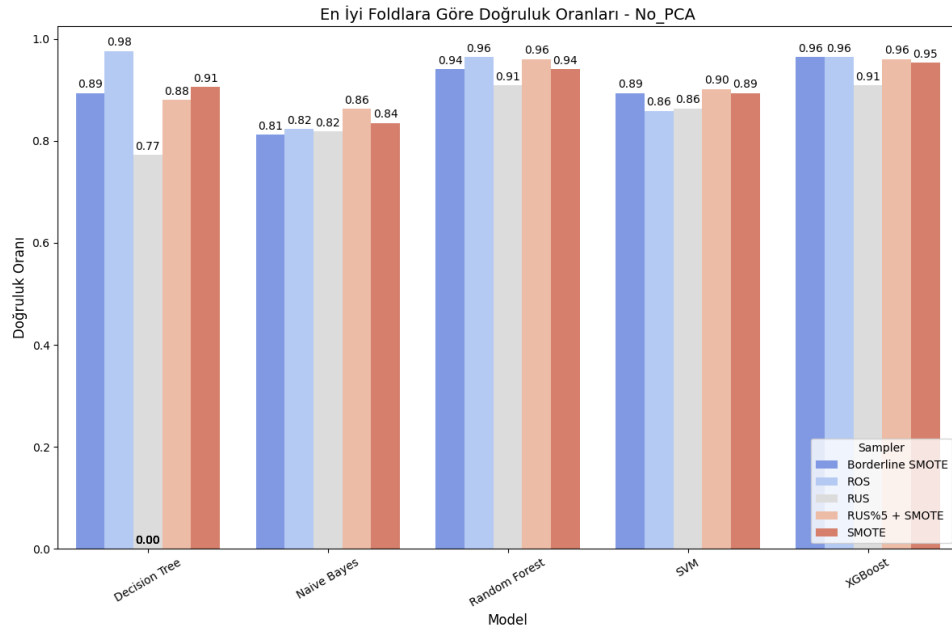
**Tabakalı K-Katlı (10 katlı çapraz doğrulama):** Tabakalı K-Katlı çapraz doğrulama, bir makine öğrenmesi modelinin genelleme yeteneğini değerlendirmek için kullanılan bir yöntemdir. Bu yöntemde, veri seti öncelikle hedef değişken (sınıf etiketi) dikkate alınarak eşit sayıda örnek içeren K adet alt kata (fold) ayrılır. Daha sonra, her bir kat sırayla test kümesi olarak seçilirken, kalan K-1 kat eğitim kümesi olarak kullanılır. Bu işlem K kez tekrarlanır ve her tekrar sonunda modelin performansı hesaplanır. Elde edilen performans değerlerinin ortalaması, modelin genel performansı olarak kabul edilir. Tabakalı K-Katlı çapraz doğrulama, veri setindeki sınıf dengesizliğini koruyarak daha güvenilir sonuçlar elde etmeye yardımcı olur. Bu çalışmada K değeri olarak 10 alınmıştır.



Şekil 3.6 10 katlı çapraz doğrulama [19]

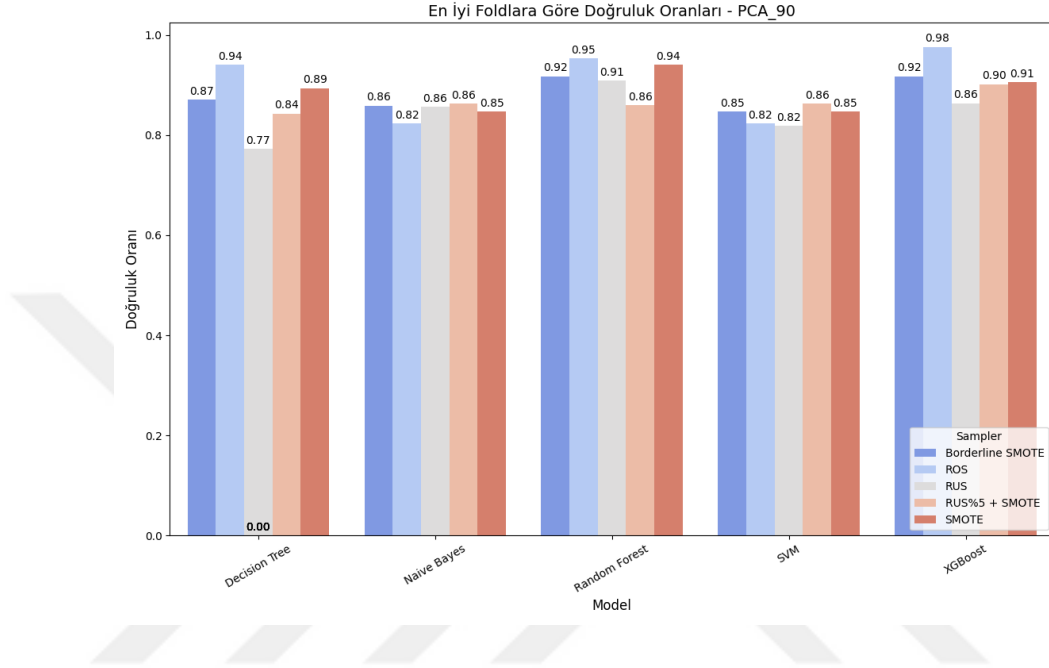
#### 4. BULGULAR

Bu çalışmada KBSL istatistik verilerinden elde edilen 533 satır ve 50 sütun (öznitelik) bulunan veri setine öncelikle 4 farklı PCA yaklaşımı yapılmıştır. Daha sonra 5 farklı örnekleme yöntemi, 5 farklı sınıflandırma algoritması ile çalıştırılmıştır. Modeller geliştirilirken Tabakalı 10 Katlı Çapraz Doğrulama yapılarak modelin genel performansı bir den fazla kez test edilmiş ve daha kararlı sonuçlar ortaya koyması amaçlanmıştır. Elde edilen veriler karmaşıklık matrisi, doğruluk ve F1-Skor değerlerine bakılarak karşılaştırılmıştır. Elde edilen bulgular Excel olarak kaydedilmiş (EK-A), her model ve örnekleme yöntemine ait karmaşıklık matrisleri çizdirilerek pdf olarak oluşturulmuştur (EK-B). Algoritmalar değerlendirilirken 10 kat çapraz doğrulamaya ait her bir kata ait en iyi katlar ve en kötü katlar örnekleme yöntemi, model ve PCA yaklaşımına göre ayrı ayrı görselleştirilmiştir (Şekil 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8). Modellere ve örnekleme yöntemlerine ait her bir PCA yaklaşımı için F1 skorlarına ait grafikler çizilmiştir (Şekil 4.9, 4.10, 4.11, 4.12). Ayrıca modellerin doğruluk oranlarına bağlı karşılaştırma grafiği çizilmiştir (Şekil 4.13, 4.14, 4.15, 4.16).



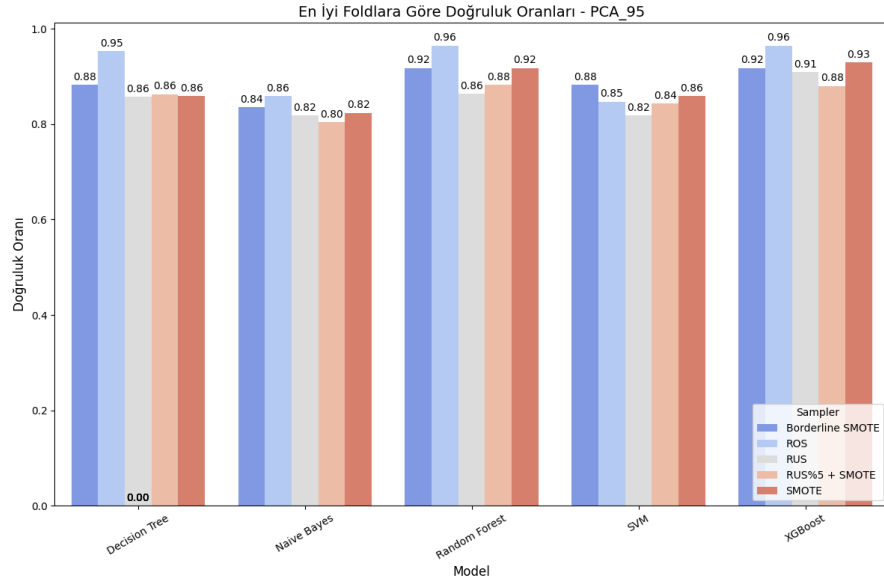
Şekil 4.1 En iyi kat grafiği (PCA Yok)

PCA yaklaşımı uygulanmayan modelde DT algoritması ROS ile uygulandığında en iyi katı 0,98 doğruluk oranına sahip olmuştur. Ancak algoritmaların her bir örnekleme yöntemine ait genel değerlendirmesine baktığımızda PCA olmadığı durumda XGBoost algoritması en iyi katlarında 0,95 doğruluğa ulaşmıştır.



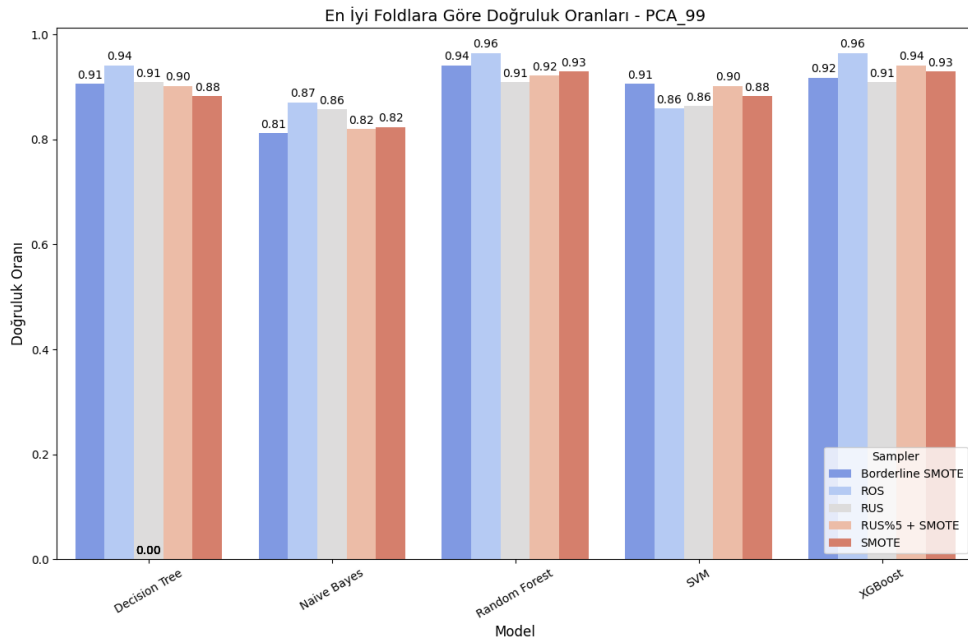
Şekil 4.2 En iyi kat grafiği (PCA\_0,90)

Kümülatif varyans 0,90 alınan PCA durumunda en iyi kat XGBoost modelinin ROS örnekleme yöntemiyle elde edilmiştir, 0,98 doğruluk oranına sahiptir. En iyi folda ait tüm örnekleme yöntemlerine göre model değerlendirilmesi sonucu 0,92 ile RF modeline aittir.



Şekil 4.3 En iyi kat grafiği (PCA\_0,95)

0,95 kümülatif varyans ile PCA uygulandığı durumda RF ve XGBoost algoritmaları en iyi katı ROS örneklemesinde 0,96 doğruluk oranıyla elde etmiştir. Ancak modelin değerlendirmesi tüm örnekleme yöntemlerine ait en iyi kat doğruluk oranı olarak yapıldığında 0,93 ile XGBoost öne çıkmaktadır.

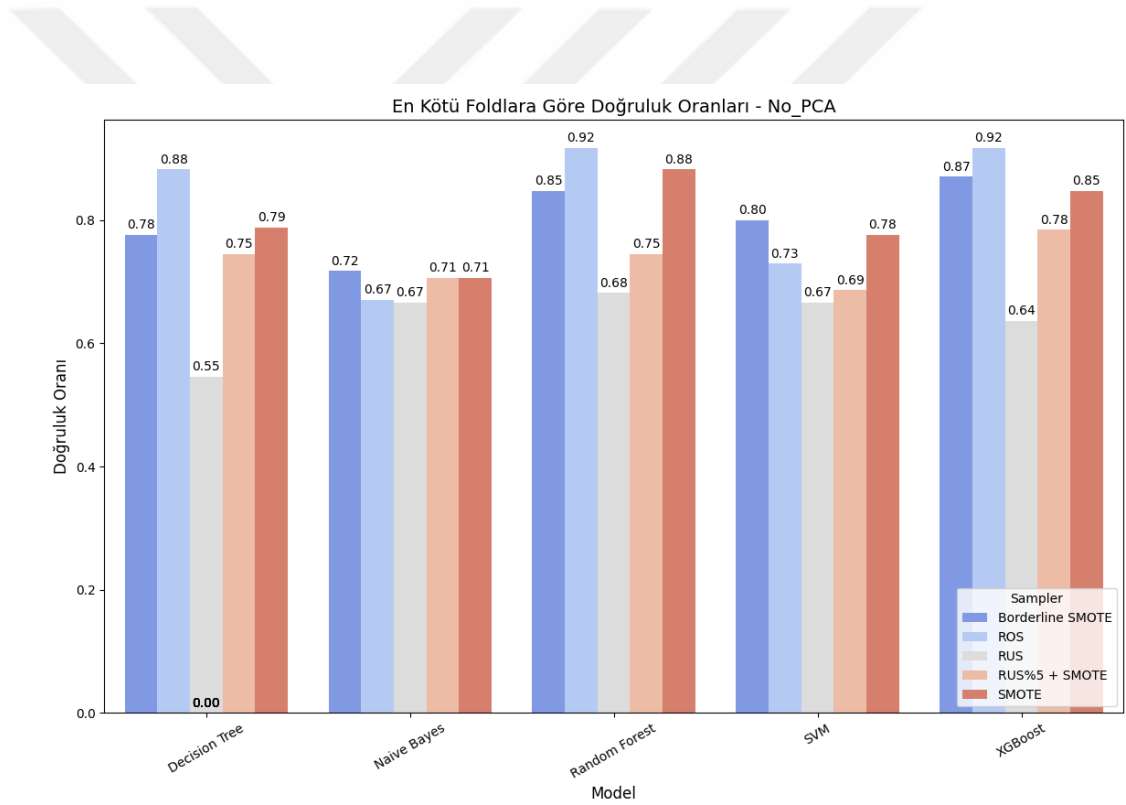


Şekil 4.4 En iyi kat grafiği (PCA\_0,99)

PCA yaklaşımında kümülatif varyans değeri 0,99 alındığında en iyi kat yine RF ve XGBoost modellerinde ROS örnekleme yönteminde elde edilmiştir, %96 doğruluk oranı ile model çalışmıştır. Modelin bütün örnekleme yöntemlerinde de bu iki model 0,93 ile en iyi katlara sahip olduğu gözlemlenmiştir.

En iyi kat performansları genel olarak değerlendirildiğinde her bir PCA yaklaşımı için ROS örnekleme en iyi değerlere sahiptir. Model olarak ise XGBoost modelinin performansının en iyi olduğu yorumlanabilir.

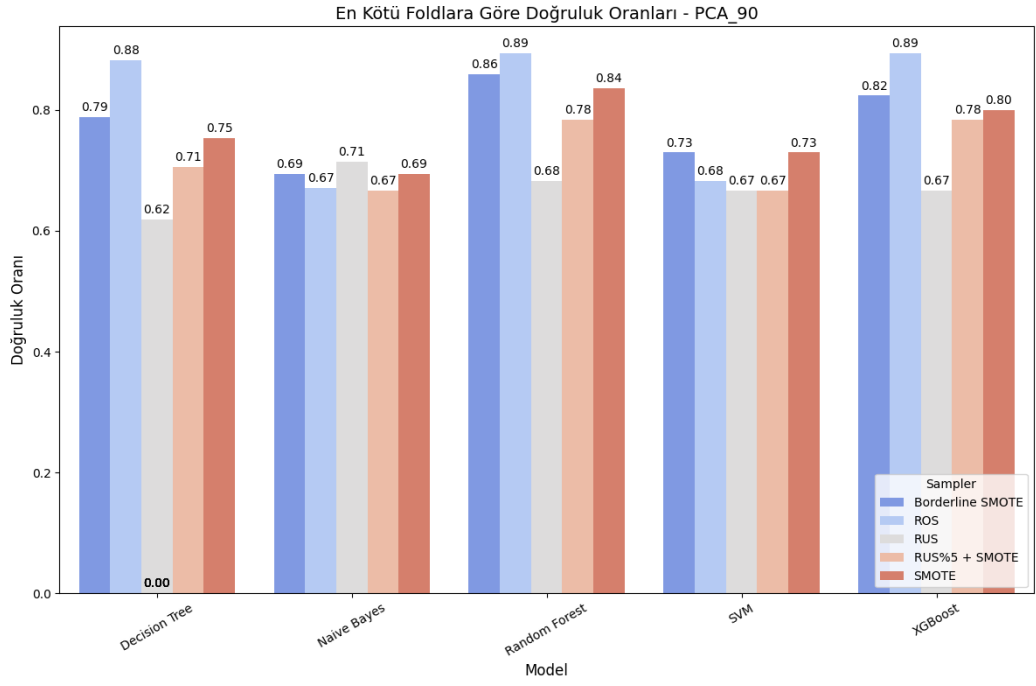
Modellere ait katlar incelenirken en kötü katlara ait grafiklerde çizdirilmiştir. Bu grafiklerden elde edilen sonuçlara göre de RUS örnekleme yöntemi ve Naive Bayes modelleme algoritması en kötü değerlere sahiptir.



Şekil 4.5 En kötü kat doğruluk oranı grafiği (PCA Yok)

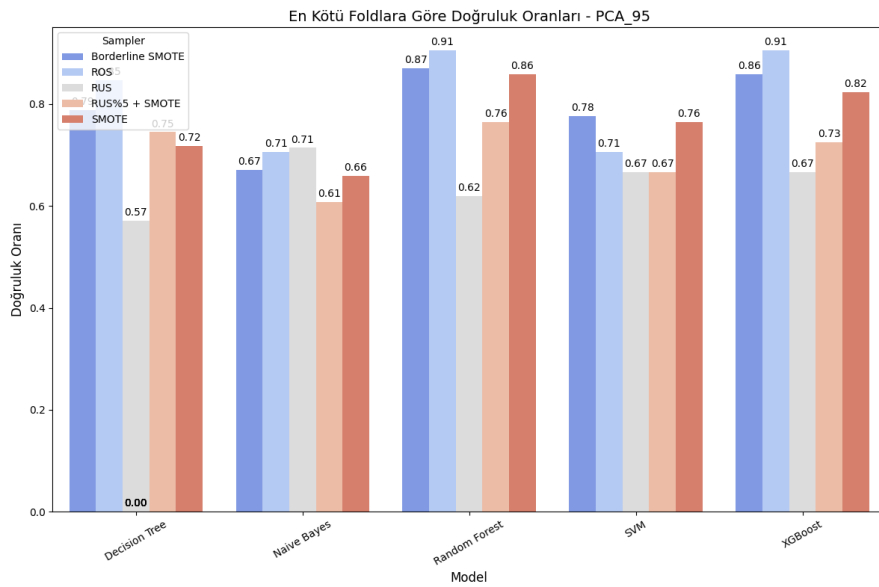
PCA yaklaşımı uygulanmayan modelde DT algoritması RUS örnekleme yöntemi ile uygulandığında en kötü katı 0,55 doğruluk oranına sahip olmuştur.

Modellerin en kötü kat için genel doğruluk oranları incelendiğinde 0,70 doğruluk oranı ile Naive Bayes en kötü kat sonucuna sahiptir.



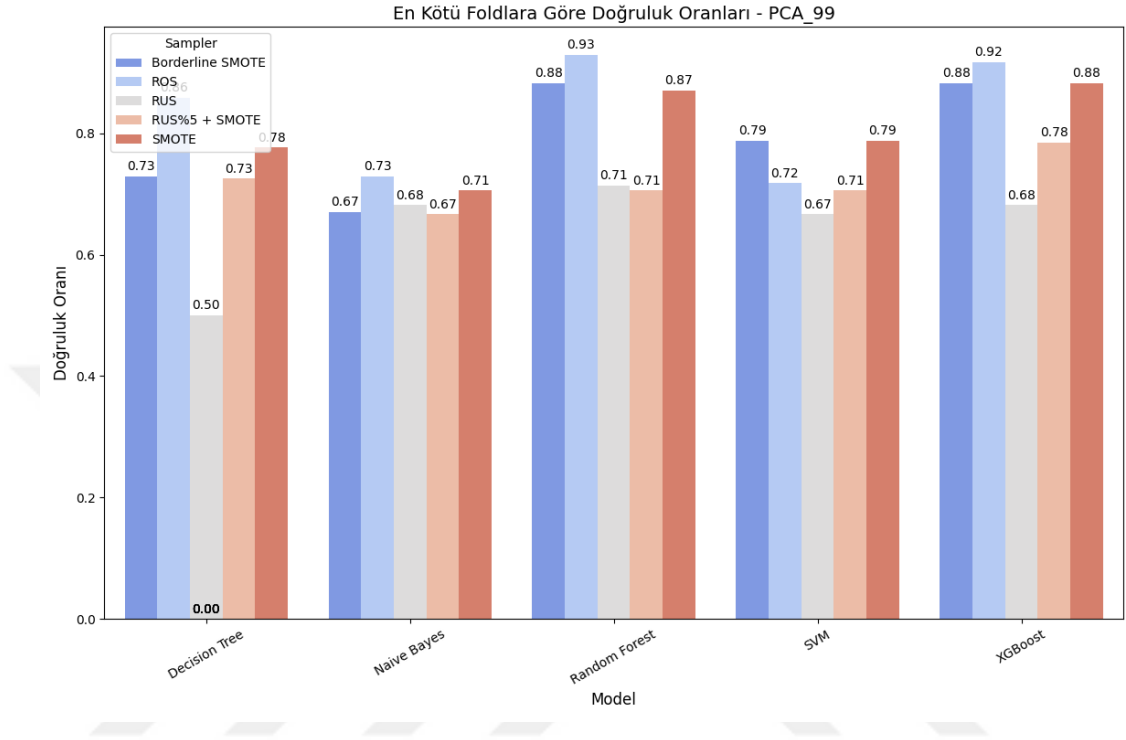
Şekil 4.6 En kötü kat doğruluk oranı grafiği (PCA\_0,90)

Kümülatif varyans değeri 0,90 alınan PCA durumunda en kötü kat DT modelinin RUS örnekleme yöntemiyle elde edilmiştir. 0,62 doğruluk oranına sahiptir. En kötü kata ait tüm örnekleme yöntemlerine göre model değerlendirmesi sonucu 0,69 doğruluk oranı ile Naive Bayes modeline aittir.



Şekil 4.7 En kötü kat doğruluk oranı grafiği (PCA\_0,95)

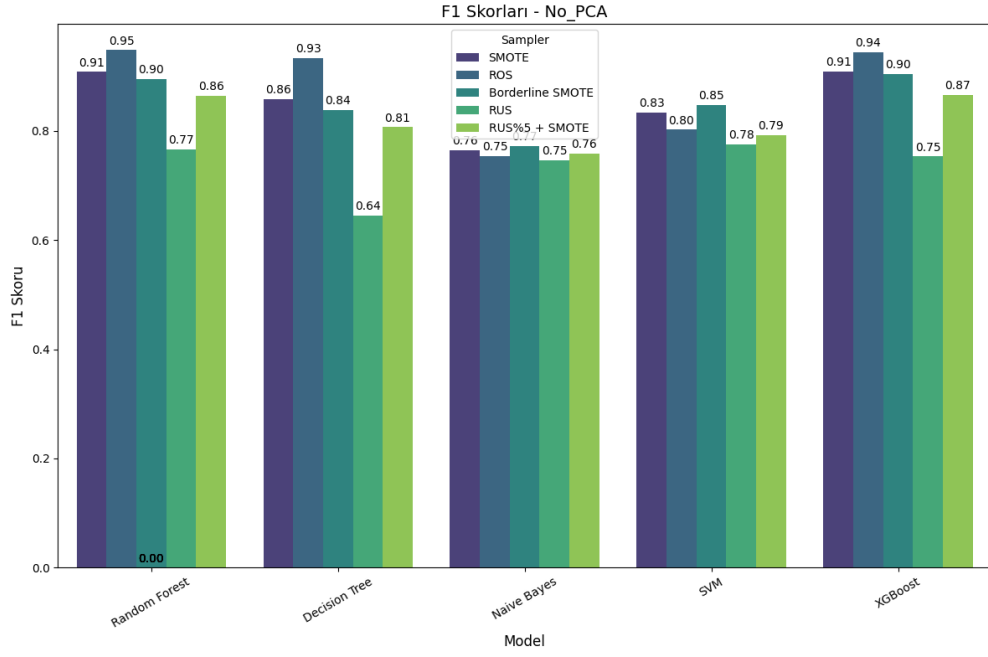
0,95 kümülatif varyans değeri ile çalışılan PCA durumunda en kötü kat RUS örnekleme yöntemiyle DT modelinde elde edilmiştir ve 0,57 doğruluk oranına sahiptir. En kötü kata ait tüm örnekleme yöntemlerine göre model değerlendirmesi sonucu 0,67 ile Naive Bayes modeline aittir.



Şekil 4.8 En kötü kat doğruluk oranı grafiği (PCA\_0,99)

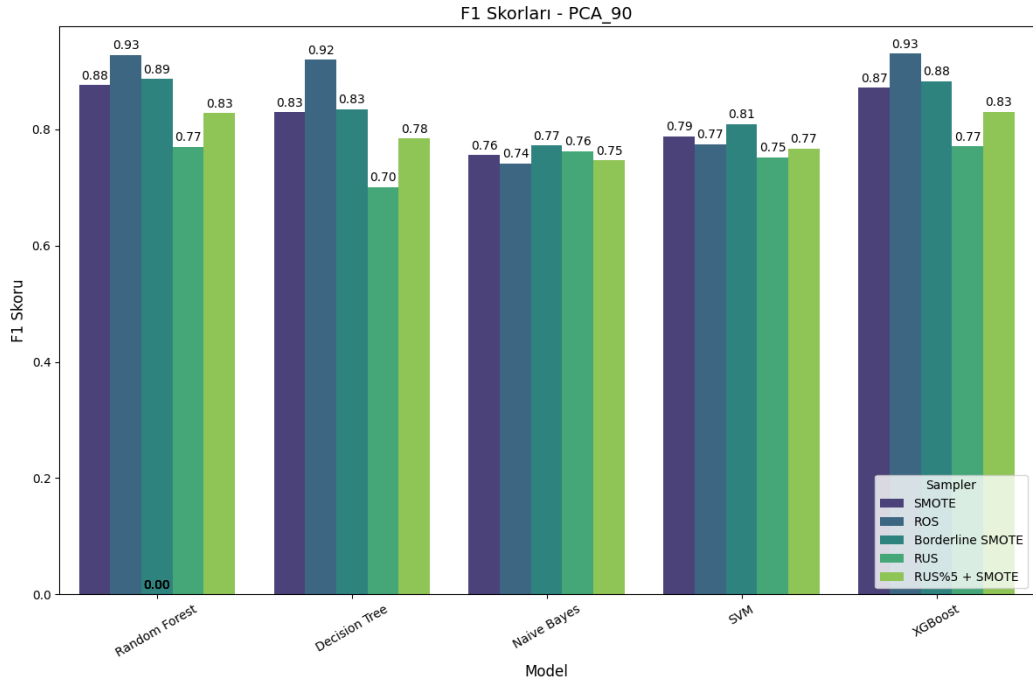
Kümülatif varyans 0,99 alınan PCA yaklaşımında en kötü kat RUS örnekleme yöntemi ile uygulanan DT modelinde 0,50 ile elde edilmiştir. Genel model değerlendirmelerinde en kötü kat sonucunda en düşük doğruluk oranına Naive Bayes 0,69 ile sahiptir.

Modellerin değerlendirilmesinde kullanılan F1 skoru için her PCA yaklaşımına ait grafikler her model ve her örnekleme yöntemi için oluşturulmuştur. Oluşturulan grafiklerin sonuçları genel olarak karşılaştırıldığında XGBoost ve RF modelleri arasında 0,004'lük bir fark bulunmakta olduğu görülmüştür. Yalnızca kümülatif varyansın 0,99 alındığı PCA yaklaşımında fark 0,12'dir.



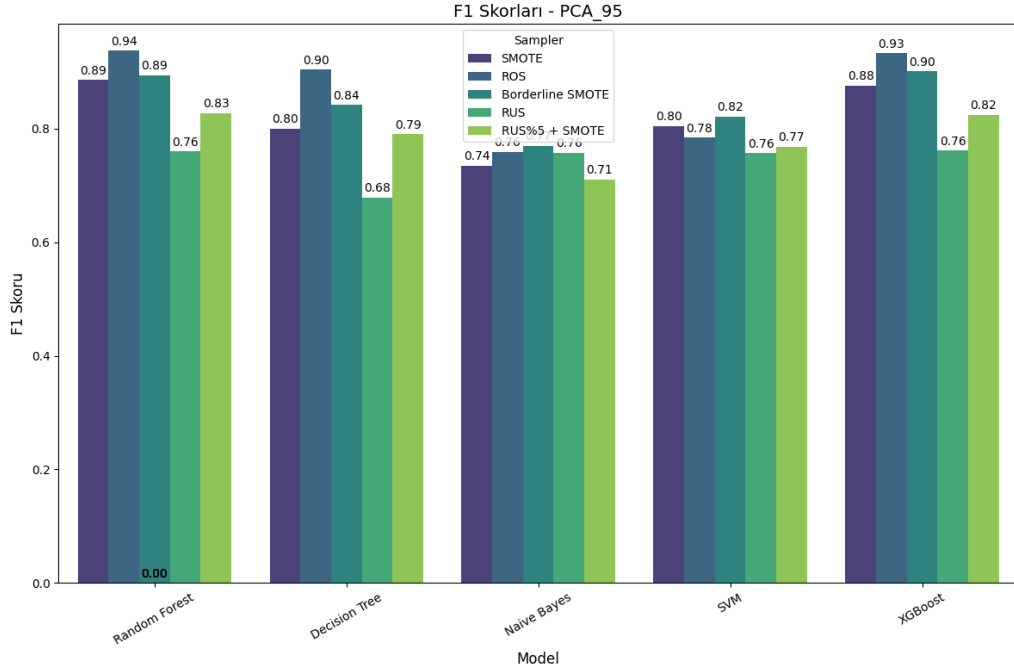
Şekil 4.9 Modellerin örnekleme yöntemlerine göre F1 Skoru grafiği (PCA Yok)

PCA uygulanmadan modellemeler yapıldığında ROS örnekleme yöntemiyle RF algoritması 0,95’lik bir F1 Skoru elde etmiştir. Genel ortalamalara bakıldığında da RF ile XGBoost algoritması arasında 0,004’lük bir fark olduğu ve RF algoritmasının 0,878 bir oranla çalıştığı gözlemlenmiştir.



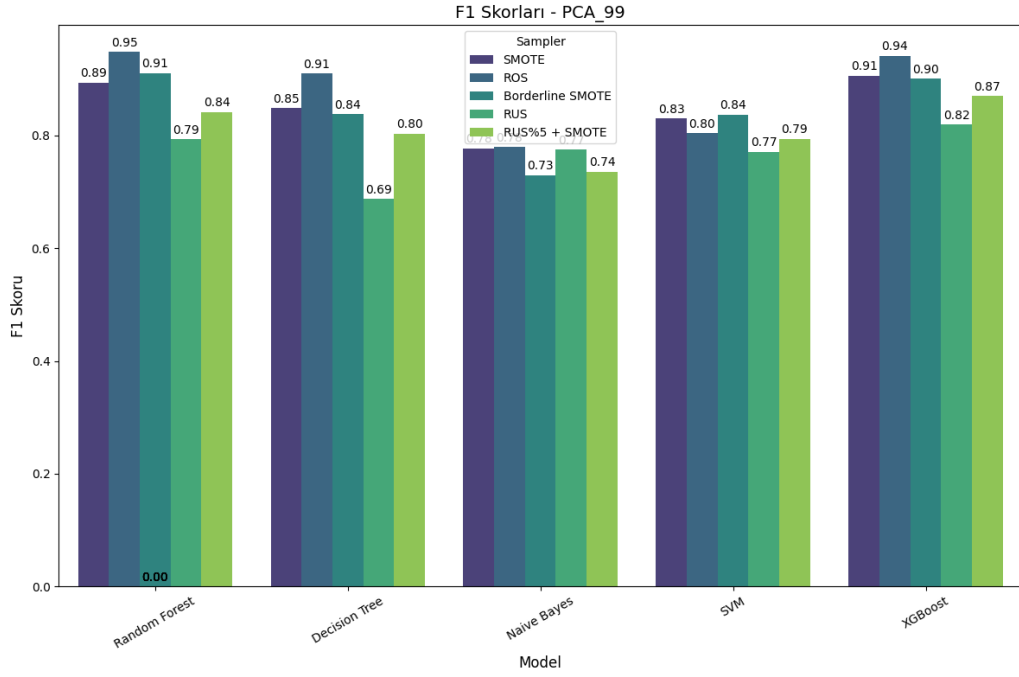
Şekil 4.10 Modellerin örnekleme yöntemlerine göre F1 Skoru grafiği (PCA 0,90)

Kümülatif varyans değeri 0,90 alınan PCA yaklaşımında modellere ait algoritma F1 Skorları şu şekilde kaydedilmiştir: Tek bir örnekleme yöntemi ve model bakıldığında ROS örnekleme ile RF ve XGBoost modelleri 0,93 skorunu yakalamışlardır, genel model değerlendirmesinde ise 0,86 ile RF algoritması daha iyi bir F1 Skoru elde etmiştir.



Şekil 4.11 Modellerin örnekleme yöntemlerine göre F1 Skoru grafiği (PCA 0,95)

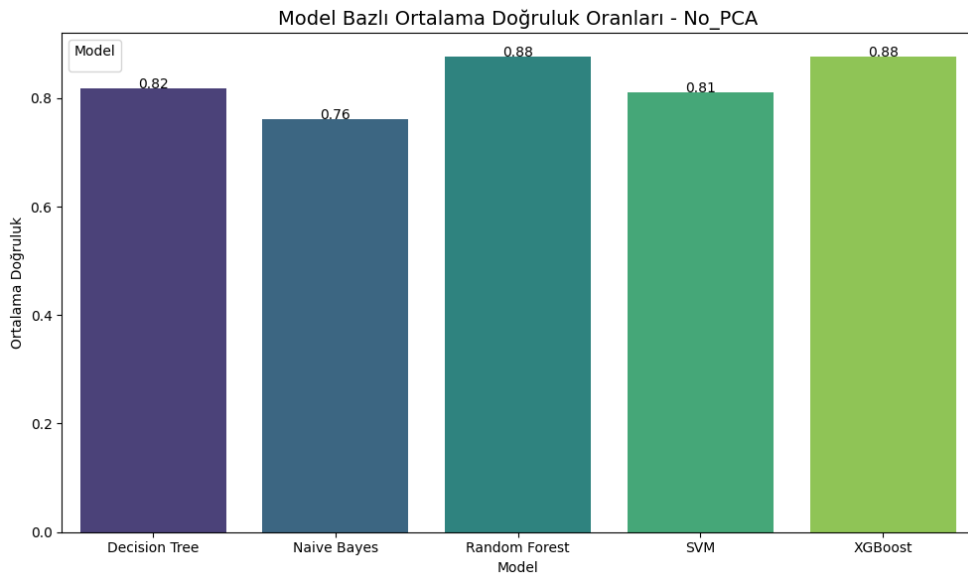
PCA yaklaşımında kümülatif varyansı 0,95 aldığımızda elde edilen F1 Skorlarında ROS örnekleme yöntemiyle RF tekil değerlendirmede 0,94 skorunu elde etmiştir. Modellerin genel değerlendirmesi incelendiğinde RF ve XGBoost modelleri arasında 0,004'lük bir skor farkı olmuştur. RF 0,862 skoru ile öne geçmiştir.



Şekil 4.12 Modellerin örnekleme yöntemlerine göre F1 Skoru grafiği (PCA 0,99)

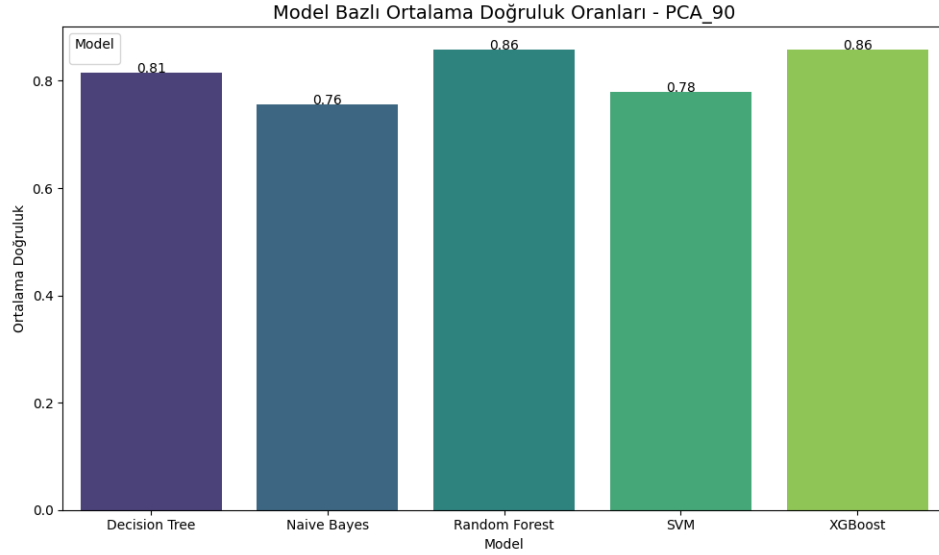
F1 skorlarının 0,99 kümülatif varyans ile PCA uygulandığı durumlarda ROS örnekleme ile RF modeli 0,95 skorunu vermiştir. Bunun yanında genel değerlendirme XGBoost 0,888 F1 Skoru ile diğer modellerden daha iyi sonuç vermiştir.

Son olarak oluşturduğumuz genel model doğrulukları grafikleri incelenmiştir. Bu doğruluk oranı değerlendirmesi her PCA yaklaşımı için ayrı ayrı yapılmış daha sonra genel bir model belirlenmiştir.

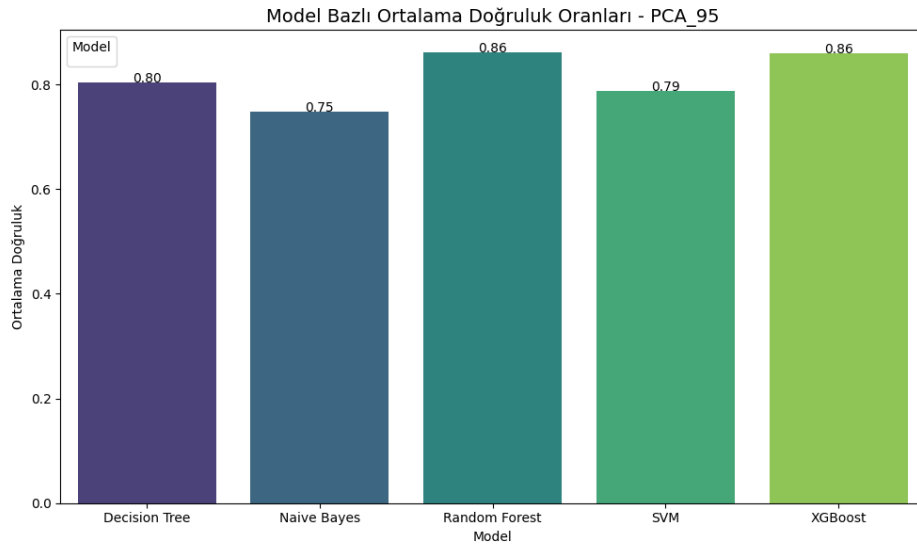


Şekil 4.13 PCA yapılmadan model doğruluk grafiği

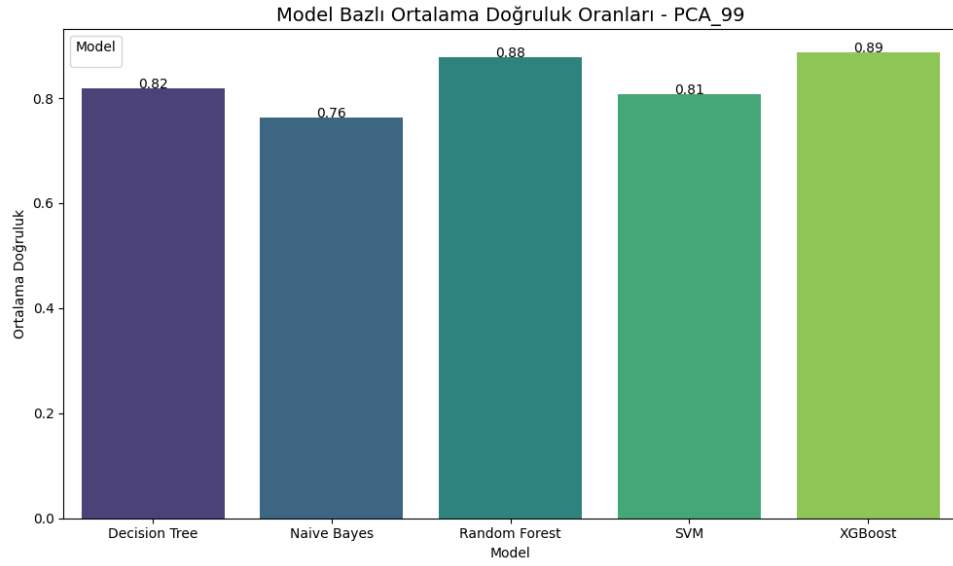
PCA yaklaşımının olmadığı durumda XGBoost ve RF modelleri %88 doğruluk değeri ile sonuç üretmiştir.



Şekil 4.14 Kümülatif varyans 0,90 ile PCA uygulanmış model doğruluk grafiği  
Kümülatif varyansın 0,90 alındığı yaklaşımda yine RF ve XGBoost modelleri en iyi sonuca ulaşmış ve %86 doğruluk oranı elde etmişlerdir.



Şekil 4.15 Kümülatif varyans 0,95 ile PCA uygulanmış model doğruluk grafiği  
Kümülatif varyansın 0,95'e çekilmesi bileşen sayısını 0,90'a göre 4 arttırmış olsa da modellerin genel doğruluk oranında iki varyans değeri arasında bir fark olmamıştır. XGBoost ve RF modelleri %86 doğruluk ile çalışmaktadır.



Şekil 4.16 Kümülatif varyans 0,99 ile PCA uygulanmış model doğruluk grafiği

Kümülatif varyans değeri 0,99'a çekildiğinde hem bileşen sayısı artmış hem de model doğruluğu bir modelde daha yüksek sonuçlar vermiştir. XGBoost modeli %89 doğruluk ile tüm PCA yaklaşımları içinde en yüksek doğruluk oranına sahip model olarak gözlemlenmiştir.

Tablo 4.1'e bakıldığında her PCA sonucu için en etkili sonuçların RF ve XGBoost modelleri için alındığı görülmektedir. RF modeli en yüksek %86 doğruluk oranını yakalarken XGBoost modeli %89 ile en yüksek doğruluk oranına sahip olmuştur.

Tablo 4.1 Modellerin PCA yaklaşımlarına göre doğruluk ve F1 Skorları

| PCA      | MODEL         | F1 SKORU | DOĞRULUK |
|----------|---------------|----------|----------|
| PCA_0,99 | Random Forest | 0,88     | 0,88     |
|          | XGBoost       | 0,89     | 0,89     |
|          | Decision Tree | 0,82     | 0,82     |
|          | SVM           | 0,81     | 0,81     |
|          | Naive Bayes   | 0,76     | 0,76     |
| PCA_0,95 | Random Forest | 0,86     | 0,86     |
|          | XGBoost       | 0,86     | 0,86     |
|          | Decision Tree | 0,81     | 0,81     |
|          | SVM           | 0,79     | 0,79     |
|          | Naive Bayes   | 0,75     | 0,75     |
| PCA_0,90 | Random Forest | 0,86     | 0,86     |
|          | XGBoost       | 0,86     | 0,86     |
|          | Decision Tree | 0,81     | 0,81     |
|          | SVM           | 0,78     | 0,78     |
|          | Naive Bayes   | 0,75     | 0,76     |
| PCA YOK  | Random Forest | 0,87     | 0,87     |
|          | XGBoost       | 0,87     | 0,87     |
|          | Decision Tree | 0,81     | 0,81     |
|          | SVM           | 0,80     | 0,80     |
|          | Naive Bayes   | 0,76     | 0,76     |

## 5. TARTIŞMA

Tablo 5.1 PCA yapılmadan örnekleme ve modellere göre F1 Skoru ve Doğruluk Değerleri

| PCA    | MODEL         | ÖRNEKLEME        | ORT.F1 SKORU | ORT.DOĞRULUK |
|--------|---------------|------------------|--------------|--------------|
| No PCA | Decision Tree | Borderline SMOTE | 0,8381       | 0,8385       |
| No PCA | Decision Tree | ROS              | 0,9167       | 0,9174       |
| No PCA | Decision Tree | RUS              | 0,6777       | 0,6817       |
| No PCA | Decision Tree | RUS%5 + SMOTE    | 0,7959       | 0,7969       |
| No PCA | Decision Tree | SMOTE            | 0,8338       | 0,8341       |
|        |               |                  | 0,8124       | 0,8137       |
| No PCA | Naive Bayes   | Borderline SMOTE | 0,7609       | 0,7638       |
| No PCA | Naive Bayes   | ROS              | 0,7584       | 0,7594       |
| No PCA | Naive Bayes   | RUS              | 0,7601       | 0,7630       |
| No PCA | Naive Bayes   | RUS%5 + SMOTE    | 0,7379       | 0,7408       |
| No PCA | Naive Bayes   | SMOTE            | 0,7582       | 0,7603       |
|        |               |                  | 0,7551       | 0,7575       |
| No PCA | Random Forest | Borderline SMOTE | 0,8964       | 0,8968       |
| No PCA | Random Forest | ROS              | 0,9403       | 0,9406       |
| No PCA | Random Forest | RUS              | 0,7725       | 0,7753       |
| No PCA | Random Forest | RUS%5 + SMOTE    | 0,8402       | 0,8411       |
| No PCA | Random Forest | SMOTE            | 0,8908       | 0,8912       |
|        |               |                  | 0,8681       | 0,8690       |
| No PCA | SVM           | Borderline SMOTE | 0,8285       | 0,8297       |
| No PCA | SVM           | ROS              | 0,7910       | 0,7924       |
| No PCA | SVM           | RUS              | 0,7640       | 0,7665       |
| No PCA | SVM           | RUS%5 + SMOTE    | 0,7801       | 0,7817       |
| No PCA | SVM           | SMOTE            | 0,8139       | 0,8144       |
|        |               |                  | 0,7955       | 0,7969       |
| No PCA | XGBoost       | Borderline SMOTE | 0,8973       | 0,8976       |
| No PCA | XGBoost       | ROS              | 0,9370       | 0,9374       |
| No PCA | XGBoost       | RUS              | 0,7759       | 0,7774       |
| No PCA | XGBoost       | RUS%5 + SMOTE    | 0,8472       | 0,8480       |
| No PCA | XGBoost       | SMOTE            | 0,8905       | 0,8909       |
|        |               |                  | 0,8696       | 0,8703       |

Tablo 5.2 PCA 0,90 alındığında örnekleme ve modellere göre F1 Skoru ve Doğruluk Değerleri

| PCA                     | MODEL         | ÖRNEKLEME        | ORT.F1 SKORU | ORT. DOĞRULUK |
|-------------------------|---------------|------------------|--------------|---------------|
| 90% Varyans (9 Bileşen) | Decision Tree | SMOTE            | 0,8253       | 0,8259        |
| 90% Varyans (9 Bileşen) | Decision Tree | ROS              | 0,9041       | 0,9047        |
| 90% Varyans (9 Bileşen) | Decision Tree | Borderline SMOTE | 0,8232       | 0,8235        |
| 90% Varyans (9 Bileşen) | Decision Tree | RUS              | 0,7001       | 0,7035        |
| 90% Varyans (9 Bileşen) | Decision Tree | RUS%5 + SMOTE    | 0,7840       | 0,7856        |
|                         |               |                  | 0,8073       | 0,8086        |
| 90% Varyans (9 Bileşen) | Naive Bayes   | SMOTE            | 0,7570       | 0,7576        |
| 90% Varyans (9 Bileşen) | Naive Bayes   | ROS              | 0,7427       | 0,7435        |
| 90% Varyans (9 Bileşen) | Naive Bayes   | Borderline SMOTE | 0,7651       | 0,7659        |
| 90% Varyans (9 Bileşen) | Naive Bayes   | RUS              | 0,7619       | 0,7641        |
| 90% Varyans (9 Bileşen) | Naive Bayes   | RUS%5 + SMOTE    | 0,7467       | 0,7482        |
|                         |               |                  | 0,7547       | 0,7559        |
| 90% Varyans (9 Bileşen) | Random Forest | SMOTE            | 0,8805       | 0,8812        |
| 90% Varyans (9 Bileşen) | Random Forest | ROS              | 0,9326       | 0,9329        |
| 90% Varyans (9 Bileşen) | Random Forest | Borderline SMOTE | 0,8792       | 0,8800        |
| 90% Varyans (9 Bileşen) | Random Forest | RUS              | 0,7701       | 0,7727        |
| 90% Varyans (9 Bileşen) | Random Forest | RUS%5 + SMOTE    | 0,8279       | 0,8288        |
|                         |               |                  | 0,8581       | 0,8591        |
| 90% Varyans (9 Bileşen) | SVM           | SMOTE            | 0,7946       | 0,7953        |
| 90% Varyans (9 Bileşen) | SVM           | ROS              | 0,7712       | 0,7729        |
| 90% Varyans (9 Bileşen) | SVM           | Borderline SMOTE | 0,8099       | 0,8118        |
| 90% Varyans (9 Bileşen) | SVM           | RUS              | 0,7519       | 0,7548        |
| 90% Varyans (9 Bileşen) | SVM           | RUS%5 + SMOTE    | 0,7665       | 0,7679        |
|                         |               |                  | 0,7788       | 0,7805        |
| 90% Varyans (9 Bileşen) | XGBoost       | SMOTE            | 0,8807       | 0,8812        |
| 90% Varyans (9 Bileşen) | XGBoost       | ROS              | 0,9339       | 0,9341        |
| 90% Varyans (9 Bileşen) | XGBoost       | Borderline SMOTE | 0,8783       | 0,8788        |
| 90% Varyans (9 Bileşen) | XGBoost       | RUS              | 0,7704       | 0,7723        |
| 90% Varyans (9 Bileşen) | XGBoost       | RUS%5 + SMOTE    | 0,8295       | 0,8308        |
|                         |               |                  | 0,8586       | 0,8594        |

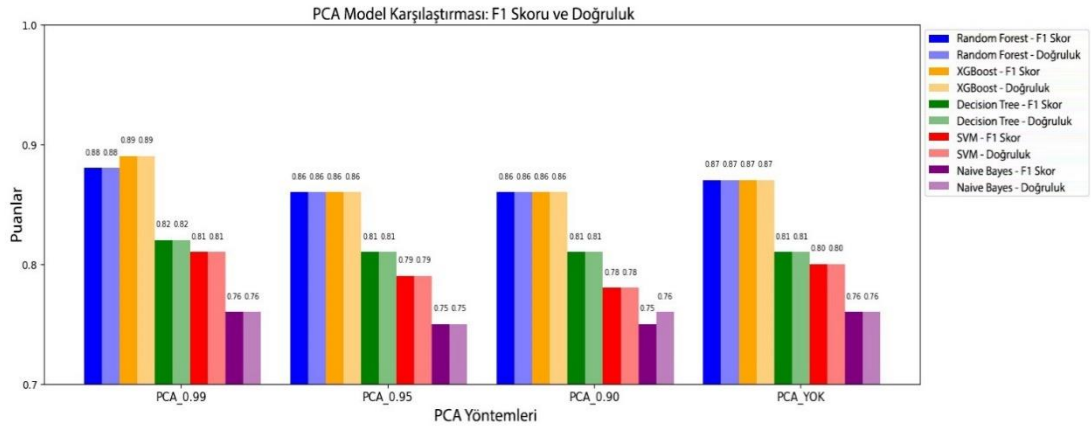
Tablo 5.3 PCA 0,95 alındığında örnekleme ve modellere göre F1 Skoru ve Doğruluk Değerleri

| PCA                      | MODEL         | ÖRNEKLEME        | ORT.F1 SKORU | ORT. DOĞRULUK |
|--------------------------|---------------|------------------|--------------|---------------|
| 95% Varyans (13 Bileşen) | Decision Tree | SMOTE            | 0,8136       | 0,8141        |
| 95% Varyans (13 Bileşen) | Decision Tree | ROS              | 0,9026       | 0,9035        |
| 95% Varyans (13 Bileşen) | Decision Tree | Borderline SMOTE | 0,8407       | 0,8412        |
| 95% Varyans (13 Bileşen) | Decision Tree | RUS              | 0,6790       | 0,6857        |
| 95% Varyans (13 Bileşen) | Decision Tree | RUS%5 + SMOTE    | 0,7902       | 0,7915        |
|                          |               |                  | 0,8052       | 0,8072        |
| 95% Varyans (13 Bileşen) | Naive Bayes   | SMOTE            | 0,7352       | 0,7376        |
| 95% Varyans (13 Bileşen) | Naive Bayes   | ROS              | 0,7618       | 0,7624        |
| 95% Varyans (13 Bileşen) | Naive Bayes   | Borderline SMOTE | 0,7648       | 0,7682        |
| 95% Varyans (13 Bileşen) | Naive Bayes   | RUS              | 0,7577       | 0,7595        |
| 95% Varyans (13 Bileşen) | Naive Bayes   | RUS%5 + SMOTE    | 0,7110       | 0,7147        |
|                          |               |                  | 0,7461       | 0,7485        |
| 95% Varyans (13 Bileşen) | Random Forest | SMOTE            | 0,8781       | 0,8788        |
| 95% Varyans (13 Bileşen) | Random Forest | ROS              | 0,9386       | 0,9388        |
| 95% Varyans (13 Bileşen) | Random Forest | Borderline SMOTE | 0,8829       | 0,8835        |
| 95% Varyans (13 Bileşen) | Random Forest | RUS              | 0,7601       | 0,7634        |
| 95% Varyans (13 Bileşen) | Random Forest | RUS%5 + SMOTE    | 0,8275       | 0,8288        |
|                          |               |                  | 0,8574       | 0,8587        |
| 95% Varyans (13 Bileşen) | SVM           | SMOTE            | 0,8053       | 0,8059        |
| 95% Varyans (13 Bileşen) | SVM           | ROS              | 0,7900       | 0,7918        |
| 95% Varyans (13 Bileşen) | SVM           | Borderline SMOTE | 0,8198       | 0,8212        |
| 95% Varyans (13 Bileşen) | SVM           | RUS              | 0,7570       | 0,7595        |
| 95% Varyans (13 Bileşen) | SVM           | RUS%5 + SMOTE    | 0,7678       | 0,7699        |
|                          |               |                  | 0,7880       | 0,7896        |
| 95% Varyans (13 Bileşen) | XGBoost       | SMOTE            | 0,8843       | 0,8847        |
| 95% Varyans (13 Bileşen) | XGBoost       | ROS              | 0,9302       | 0,9306        |
| 95% Varyans (13 Bileşen) | XGBoost       | Borderline SMOTE | 0,8890       | 0,8894        |
| 95% Varyans (13 Bileşen) | XGBoost       | RUS              | 0,7610       | 0,7630        |
| 95% Varyans (13 Bileşen) | XGBoost       | RUS%5 + SMOTE    | 0,8238       | 0,8249        |
|                          |               |                  | 0,8577       | 0,8585        |

Tablo 5.4 PCA 0,99 alındığında örnekleme ve modellere göre F1 Skoru ve Doğruluk Değerleri

| PCA                      | MODEL         | ÖRNEKLEME        | ORT.F1 SKORU | ORT. DOĞRULUK |
|--------------------------|---------------|------------------|--------------|---------------|
| 99% Varyans (22 Bileşen) | Decision Tree | SMOTE            | 0,8344       | 0,8353        |
| 99% Varyans (22 Bileşen) | Decision Tree | ROS              | 0,9111       | 0,9118        |
| 99% Varyans (22 Bileşen) | Decision Tree | Borderline SMOTE | 0,8537       | 0,8541        |
| 99% Varyans (22 Bileşen) | Decision Tree | RUS              | 0,6869       | 0,6890        |
| 99% Varyans (22 Bileşen) | Decision Tree | RUS%5 + SMOTE    | 0,8027       | 0,8033        |
|                          |               |                  | 0,8178       | 0,8187        |
| 99% Varyans (22 Bileşen) | Naive Bayes   | SMOTE            | 0,7736       | 0,7776        |
| 99% Varyans (22 Bileşen) | Naive Bayes   | ROS              | 0,7795       | 0,7800        |
| 99% Varyans (22 Bileşen) | Naive Bayes   | Borderline SMOTE | 0,7297       | 0,7365        |
| 99% Varyans (22 Bileşen) | Naive Bayes   | RUS              | 0,7744       | 0,7781        |
| 99% Varyans (22 Bileşen) | Naive Bayes   | RUS%5 + SMOTE    | 0,7357       | 0,7403        |
|                          |               |                  | 0,7586       | 0,7625        |
| 99% Varyans (22 Bileşen) | Random Forest | SMOTE            | 0,8961       | 0,8965        |
| 99% Varyans (22 Bileşen) | Random Forest | ROS              | 0,9409       | 0,9412        |
| 99% Varyans (22 Bileşen) | Random Forest | Borderline SMOTE | 0,9079       | 0,9082        |
| 99% Varyans (22 Bileşen) | Random Forest | RUS              | 0,7932       | 0,7963        |
| 99% Varyans (22 Bileşen) | Random Forest | RUS%5 + SMOTE    | 0,8416       | 0,8425        |
|                          |               |                  | 0,8759       | 0,8769        |
| 99% Varyans (22 Bileşen) | SVM           | SMOTE            | 0,8231       | 0,8235        |
| 99% Varyans (22 Bileşen) | SVM           | ROS              | 0,8020       | 0,8035        |
| 99% Varyans (22 Bileşen) | SVM           | Borderline SMOTE | 0,8391       | 0,8400        |
| 99% Varyans (22 Bileşen) | SVM           | RUS              | 0,7712       | 0,7736        |
| 99% Varyans (22 Bileşen) | SVM           | RUS%5 + SMOTE    | 0,7939       | 0,7955        |
|                          |               |                  | 0,8059       | 0,8072        |
| 99% Varyans (22 Bileşen) | XGBoost       | SMOTE            | 0,9128       | 0,9129        |
| 99% Varyans (22 Bileşen) | XGBoost       | ROS              | 0,9338       | 0,9341        |
| 99% Varyans (22 Bileşen) | XGBoost       | Borderline SMOTE | 0,9030       | 0,9035        |
| 99% Varyans (22 Bileşen) | XGBoost       | RUS              | 0,8190       | 0,8197        |
| 99% Varyans (22 Bileşen) | XGBoost       | RUS%5 + SMOTE    | 0,8698       | 0,8702        |
|                          |               |                  | 0,8877       | 0,8881        |

Bu çalışmada elde edilen sonuçlara göre XGBoost modeli genel sonuçlarda 0,87 doğruluk ve F1 Skor değerleriyle başarılı sonuçlar vermiştir. Tablo 5.1, 5.2, 5.3 ve 5.4 incelendiğinde RF ve XGBoost modellerinin çok yakın sonuçlar verdiği görülmektedir. RF modeli 0,867 gibi 0,003 daha az doğruluk ve F1 Skoruna sahiptir ancak yine de başarılı olduğu söylenebilir. Tablo 5.1, 5.2, 5.3'te işaretli olan en iyi örnekleme ve modelleme sonuçlarına bakıldığında 3 tabloda da aynı algoritmalar aynı örnekleme yöntemleriyle en iyi sonuçları vermekteydi ancak Tablo 5.4'te NB algoritması da ROS örnekleme ile iyi sonuçlar vermeye başlamıştır. Her iki modelde milli takım aday havuzuna oyuncu belirlemek için yapılan makine öğrenmesi modellemelerinde kullanılabilir olduğunu göstermiştir.



Şekil 5.1 Model Karşılaştırma Grafiği

Srinivasan ve arkadaşları çalışmalarında Lojistik Regresyon + SVM ile hibrit bir model geliştirmiş ayrıca yine SVM ile model doğruluğu çalışmışlardır. Onların modellerinde SVM, hibrit modelden daha kötü bir performans sergilemiştir [20]. Bu çalışmada yapılan çalışmalar ile benzer bir hibrit modelin XGBoost ile oluşturulması model doğruluğunu daha da arttırabilir.

Han ve Yu çalışmalarında RF algoritmasını en değerli oyuncu seçimi için basketbola ait istatistiksel verileri kullanmışlardır [21]. Bu çalışmanın sonuçlarında da RF algoritmasının başarılı sonuçlar verdiği gözlenmiştir ancak XGBoost algoritması ile hem doğruluk hem de F1 Skor metriklerinde daha iyi sonuçlar vermiştir.

Muniz ve Flamand oyuncuların sinerji değerlerini hesaplayarak, oyuncuların takım kimyasına uygunluğunu kümeleyerek takım kadrolarını oluşturmada kullanılacak model geliştirmişlerdir [22]. Oyuncuların pozisyonlarına uygun olarak kurulan takımlar daha başarılı sonuçlar elde edebilir. Bu çalışmanın geliştirilmesinde bu

şekilde sinerjik değ r hesaplaması ve modellerin bu şekilde optimizasyonu daha iyi sonuçlar i in kullanılabilir.

Takım kadrosunun kurulumuna dair futbolda yapılmıř  alıřmalar bulunmaktadır. Al-Shboul ve arkadaşları, bir futbol takımı i in sinir ađları modelleri kullanarak oyuncu se imine odaklanmıřlardır.  alıřmalarının sonucunda bilin li oyuncu se iminin takımın performansına katkısı olduđunu ve sporda veri analitiđi kullanılmasının takımlara ve spora katkısını vurgulamıřlardır [23].



## 6. SONUÇ VE ÖNERİLER

Çalışmanın sonuçlarına göre, XGBoost, basketbol maç sonuçlarını tahmin etmek için en etkili yöntemdir. Bu yöntem, %86 başarı oranı ile önceki çalışmalarda elde edilen başarı oranlarından daha yüksek bir başarı oranı elde etmiştir. Çalışmada kullanılan değişkenler, oyuncuların ve takımların bireysel ve takım performanslarını değerlendiren değişkenlerdir. Veri setinde bulunan “Pozisyon” özneliği çalışmaya doğrusal olmayan bir sınıflandırma olması özelliği katmaktadır. Oyuncuların mevkilerine göre özelliklerinin sınıflandırılması sonucu XGBoost algoritmasının daha iyi sonuç verdiği görülmektedir.

Bu sonuçlar, basketbol milli takımı aday havuzunun belirlenmesi için makine öğrenmesi yöntemlerinin etkili bir şekilde kullanılabileceğini göstermektedir. Bu yöntemler, oyuncu seçimi, performans analizi ve oyun stratejilerinin geliştirilmesi gibi alanlarda kullanılabilir.

Çalışmanın bazı sınırlamaları da vardır. Çalışmada kullanılan veri seti, Türkiye Kadınlar Basketbol Süper Ligi 2021-2025 sezonuna ait verilerdir. Bu nedenle, çalışmanın sonuçlarının diğer ligler için de geçerli olup olmadığı bilinmemektedir.

Bu çalışmanın önerileri aşağıda sunulmuştur:

- Gelecekteki çalışmalarda, farklı liglerden veriler kullanılarak ve farklı değişkenler incelenerek çalışmanın sonuçlarının geçerliliği ve güvenilirliği artırılabilir.
- Elde edilen sonuçlara göre makine öğrenmesi yöntemleri entegre edilmiş yazılım geliştirilerek oyuncu istatistikleri anlık olarak değerlendirilmesi sağlanabilir.
- Makine öğrenmesi yöntemlerinin performansını artırmak için görüntü işleme algoritmaları da dahil edilmiş yeni yöntemler geliştirilebilir.
- Daha büyük veri setleri kullanılabilir. Daha büyük veri setleri, makine öğrenmesi modellerinin daha doğru tahminler yapmasına yardımcı olabilir.
- Hibrit makine öğrenmesi yöntemleri geliştirilebilir.

Bu öneriler, basketbol milli takımı aday havuzu için makine öğrenmesi algoritmalarıyla tahmin yaparak milli takımlar düzeyinde hem ülke sporunun başarısının artmasına katkı, hem de oyuncuların kendi performanslarını milli takım düzeyine çıkarmasına olanak sağlayabilir.



## KAYNAKÇA

- [1] B. Karaoğlu, “Makine Öğrenmesi ile Spor Karşılaşmalarının Modellenmesi,” *Makine Öğrenmesi ile Spor Karşılaşmalarının Modellenmesi, Cilt*, vol. 5, pp. 1–6, Jun. 2015.
- [2] D. Oliver, *Basketball on Paper: Rules and Tools for Performance Analysis*. 2004. [Online]. Available: <https://www.ebsco.com/terms-of-use>
- [3] T. Horvat, J. Job, R. Logoza, and Č. Livada, “A Data-Driven Machine Learning Algorithm for Predicting the Outcomes of NBA Games,” *Symmetry (Basel)*, vol. 15, no. 4, Mar. 2023, doi: 10.3390/sym15040798.
- [4] M. Houde, “Predicting the Outcome of NBA Games,” Honors Projects, Bryant University, 2021. [Online]. Available: [https://digitalcommons.bryant.edu/honors\\_data\\_science](https://digitalcommons.bryant.edu/honors_data_science)
- [5] Y. Wang, W. Liu, and X. Liu, “Explainable AI techniques with application to NBA gameplay prediction,” *Neurocomputing*, vol. 483, pp. 59–71, Feb. 2022, doi: 10.1016/j.neucom.2022.01.098.
- [6] S. Ballı and E. Özdemir, “A novel method for prediction of EuroLeague game results using hybrid feature extraction and machine learning techniques,” *Chaos Solitons Fractals*, vol. 150, Sep. 2021, doi: 10.1016/j.chaos.2021.111119.
- [7] T. Lampis, N. Ioannis, V. Vasilios, and D. Stavrianna, “Predictions of european basketball match results with machine learning algorithms,” *Journal of Sports Analytics*, vol. 9, no. 2, pp. 171–190, Jul. 2023, doi: 10.3233/jsa-220639.
- [8] E. Özdemir and S. Ballı, “Türkiye Erkekler Basketbol Ligi Maç Sonuçlarının Makine Öğrenmesi Yöntemleri İle Tahmini,” *Mühendislik Bilimleri ve Tasarım Dergisi*, Sep. 2020, doi: 10.21923/jesd.723109.
- [9] B. ÇETİN and T. EREN, “Türkiye Erkek Milli Basketbol Takımının Eurobasket 2015 İçin Oyun Kurucu Seçiminin Çok Ölçütlü Karar

Verme Yöntemleriyle Yapılması,” *Kafkas Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 2016, doi: 10,9775/kauibfd.2016.011.

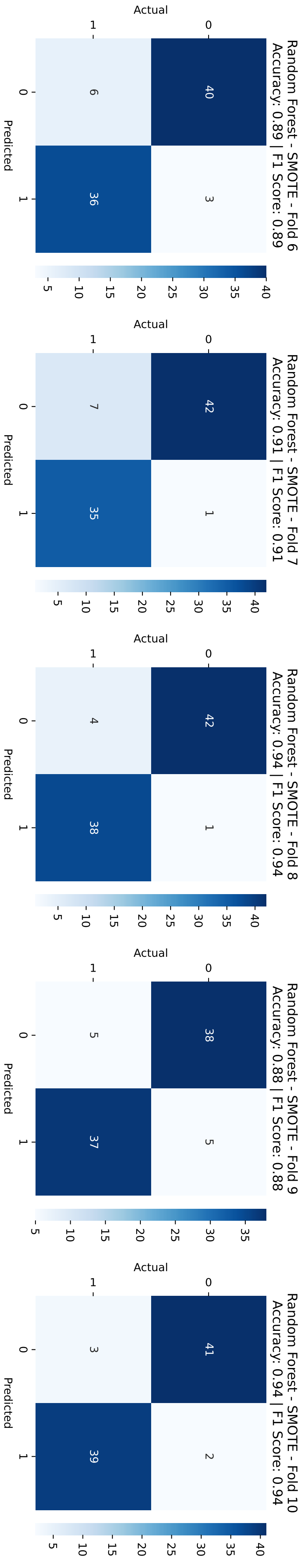
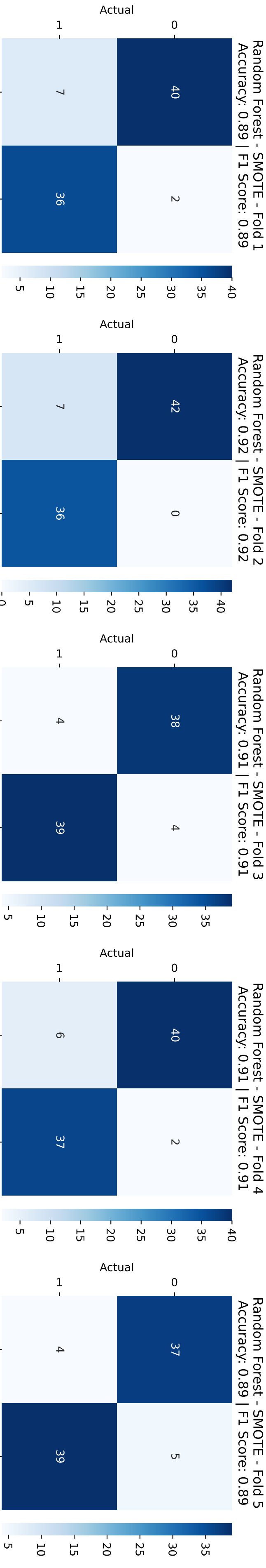
- [10] Türkiye Basketbol Federasyonu, “TBF.” Accessed: Jan. 12, 2025. [Online]. Available: <https://www.tbf.org.tr/ligler/kbsl-2024-2025>
- [11] Y. Ke, R. Bian, and R. Chandra, “A unified machine learning framework for basketball team roster construction: NBA and WNBA,” 2024, doi: 10,24433/CO.7570439.v1.
- [12] E. ÇENE, “Makine Öğrenmesi Yöntemleriyle Euroleague Basketbol Maç Sonuçlarının Tahmin Edilmesi ve Maç Sonuçları Üzerinde En Etkili Değişkenlerin Bulunması,” *Spor ve Performans Araştırmaları Dergisi*, vol. 13, no. 1, pp. 31–54, Apr. 2022, doi: 10,17155/omuspd.963235.
- [13] Y. Ouyang *et al.*, “Integration of machine learning XGBoost and SHAP models for NBA game outcome prediction and quantitative analysis methodology,” *PLoS One*, vol. 19, no. 7 July, Jul. 2024, doi: 10,1371/journal.pone.0307478.
- [14] W. J. Chen, M. J. Jhou, T. S. Lee, and C. J. Lu, “Hybrid basketball game outcome prediction model by integrating data mining methods for the national basketball association,” *Entropy*, vol. 23, no. 4, Apr. 2021, doi: 10,3390/e23040477.
- [15] GeekforGeeks, “Scikit-learn kullanarak Rastgele Orman Sınıflandırıcısı.” Accessed: Jan. 14, 2025. [Online]. Available: <https://www.geeksforgeeks.org/random-forest-classifier-using-scikit-learn/>
- [16] Mahmoud Elzeiny, “The ultimate guide to Naive Bayes.” Accessed: Jan. 14, 2025. [Online]. Available: <https://mlarchive.com/machine-learning/the-ultimate-guide-to-naive-bayes/>
- [17] Miraç Öztürk, “Python ile Sınıflandırma Analizleri – Karar Ağaçları.” Accessed: Jan. 14, 2025. [Online]. Available: <https://miracozturk.com/python-ile-siniflandirma-analizleri-karar-agaclari/>
- [18] W. Wang, G. Chakraborty, and B. Chakraborty, “Predicting the risk of chronic kidney disease (Ckd) using machine learning algorithm,” *Applied Sciences (Switzerland)*, vol. 11, no. 1, pp. 1–17, Jan. 2021, doi: 10,3390/app11010202.

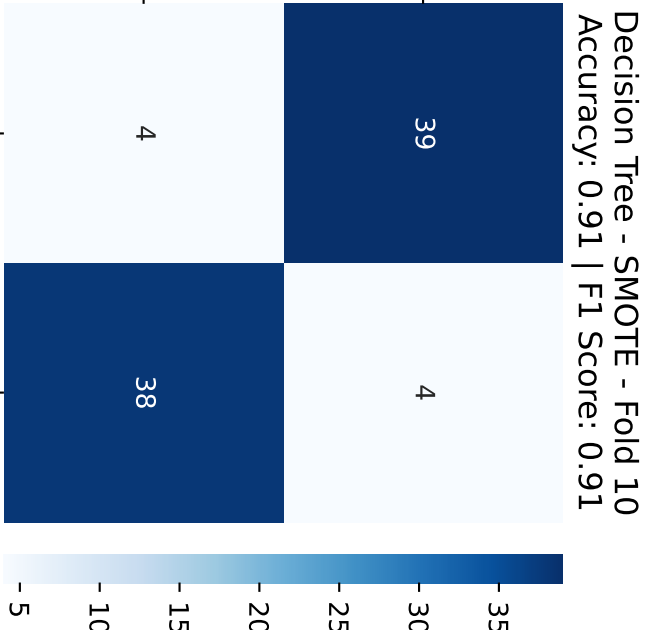
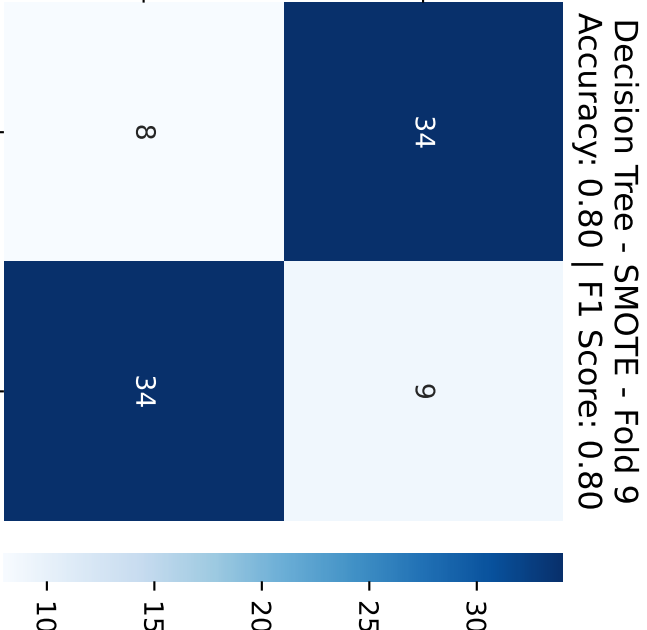
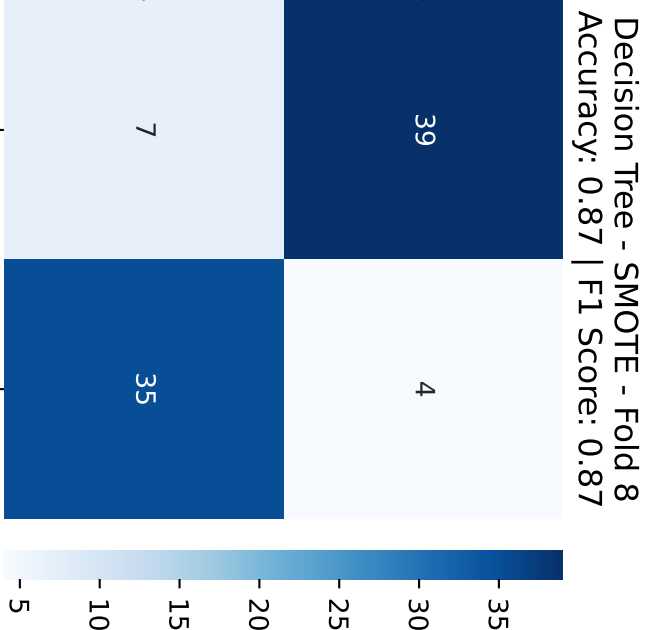
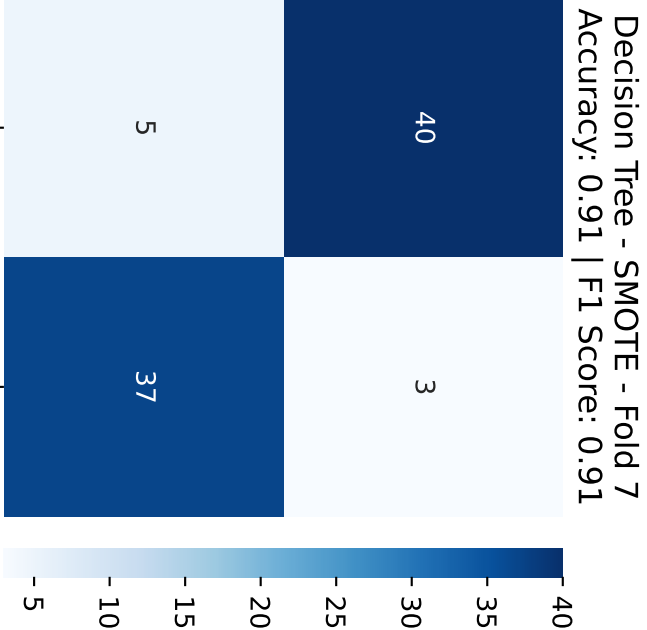
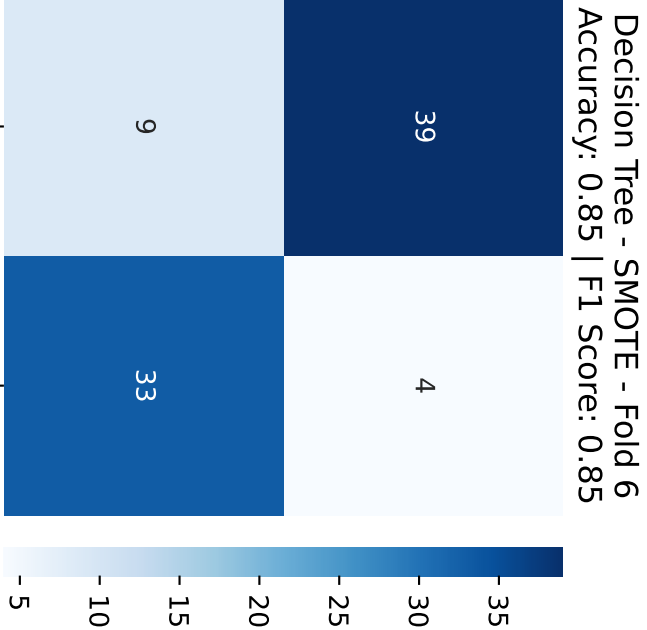
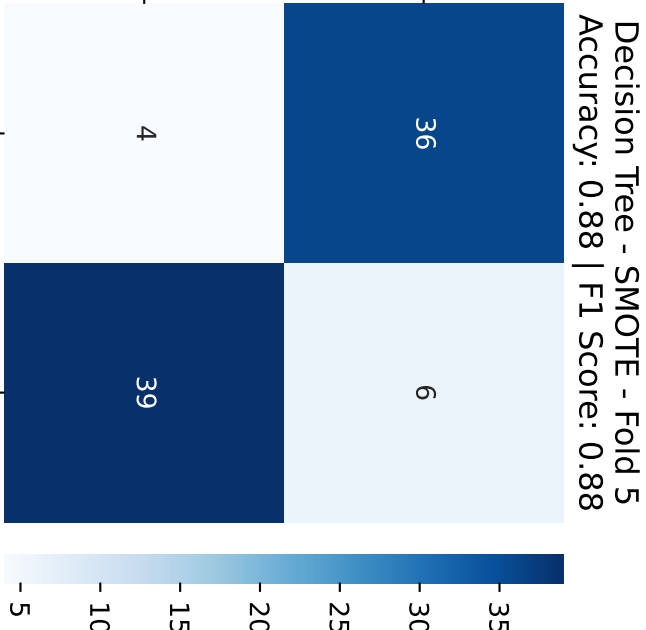
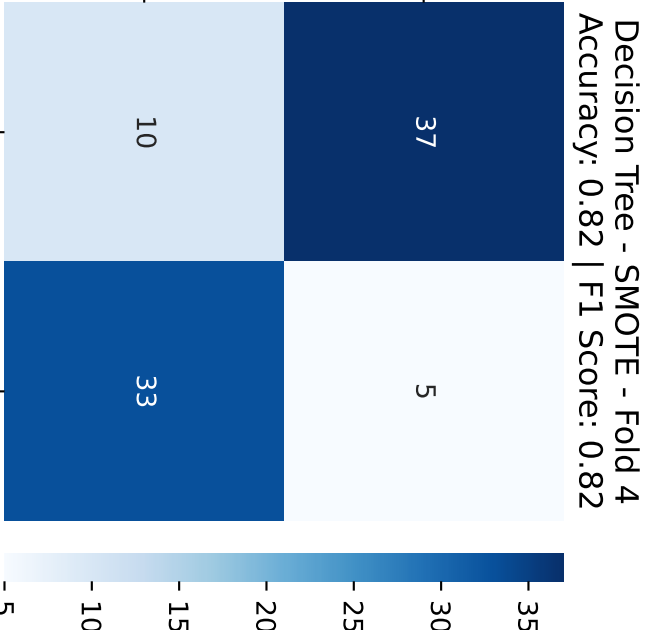
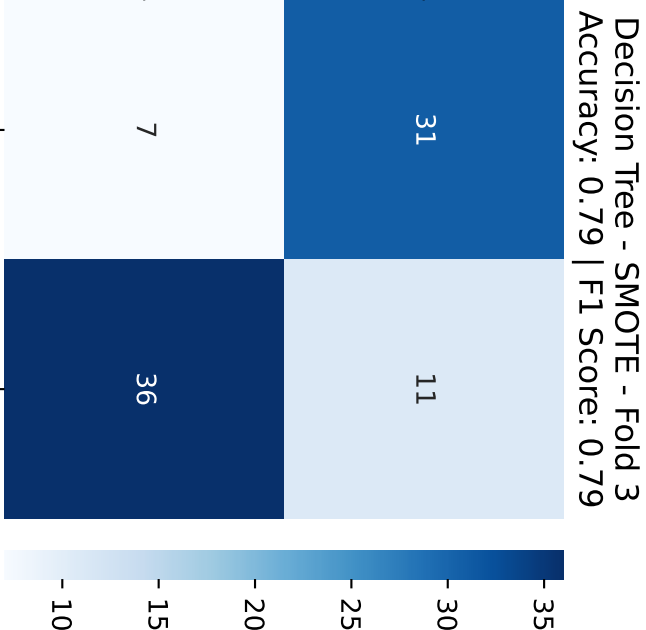
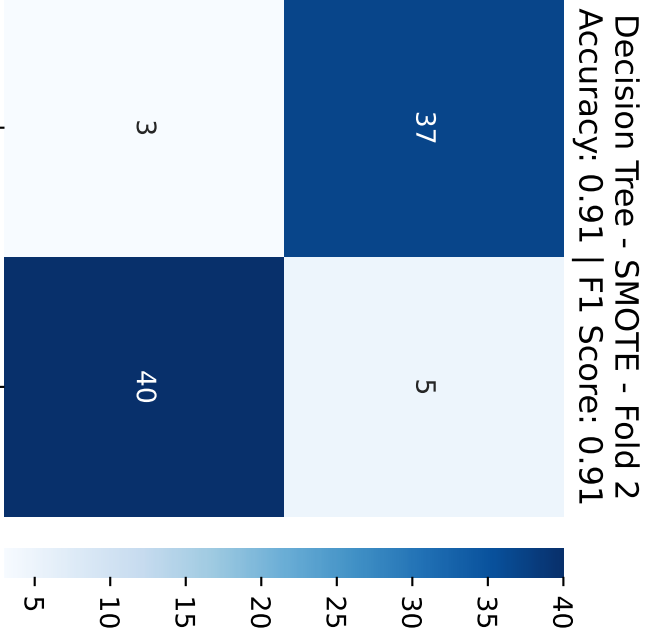
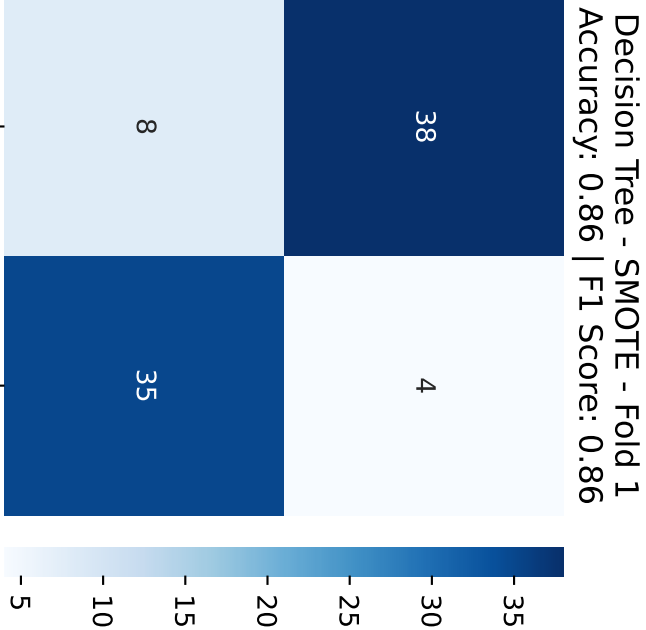
- [19] R. P. Bunker and F. Thabtah, "A machine learning framework for sport result prediction," *Applied Computing and Informatics*, vol. 15, no. 1, pp. 27–33, Jan. 2019, doi: 10.1016/j.aci.2017.09.005.
- [20] R. Srinivasan, V. Balasubramanian, and A. Vidyasagar, "A supervised learning model to identify the star potential of a basketball player," *Expert Syst*, vol. 39, no. 5, Jun. 2022, doi: 10.1111/exsy.12772.
- [21] J. Han and Z. Yu, "Random Forest Prediction of NBA Regular Season MVP Winners Based on Metrics Optimization," *Information and Knowledge Management*, vol. 4, no. 4, 2023, doi: 10.23977/infkm.2023.040409.
- [22] M. Muniz and T. Flamand, "Sports analytics for balanced team-building decisions," *Journal of the Operational Research Society*, vol. 74, no. 8, pp. 1892–1909, 2023, doi: 10.1080/01605682.2022.2118634.
- [23] R. Al-Shboul, T. Syed, J. Memon, and F. Khan, "Automated Player Selection for a Sports Team using Competitive Neural Networks," 2017. [Online]. Available: [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)

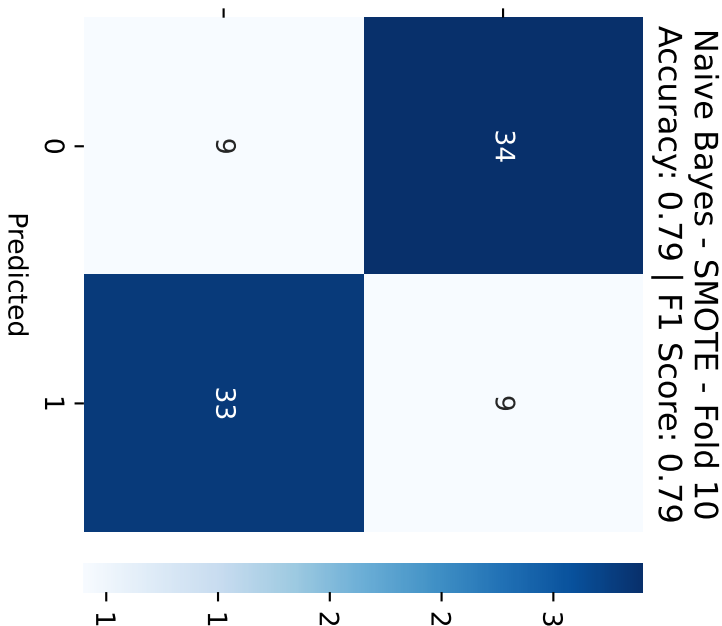
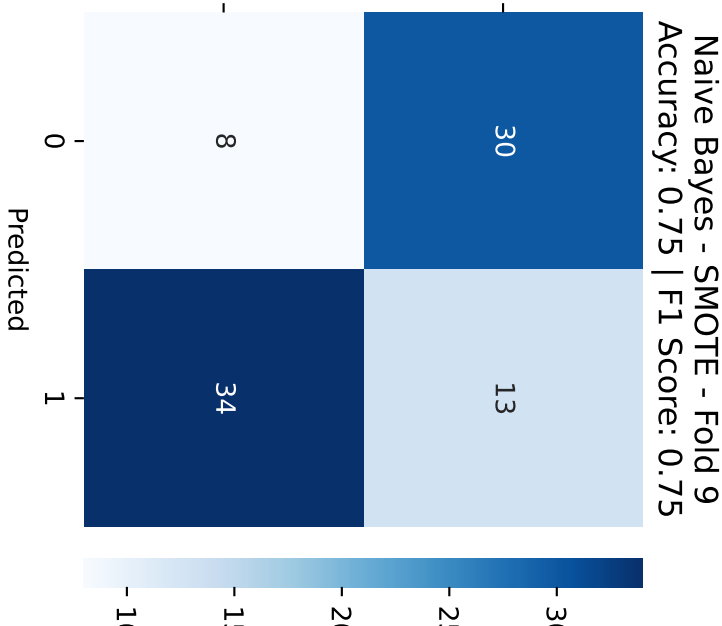
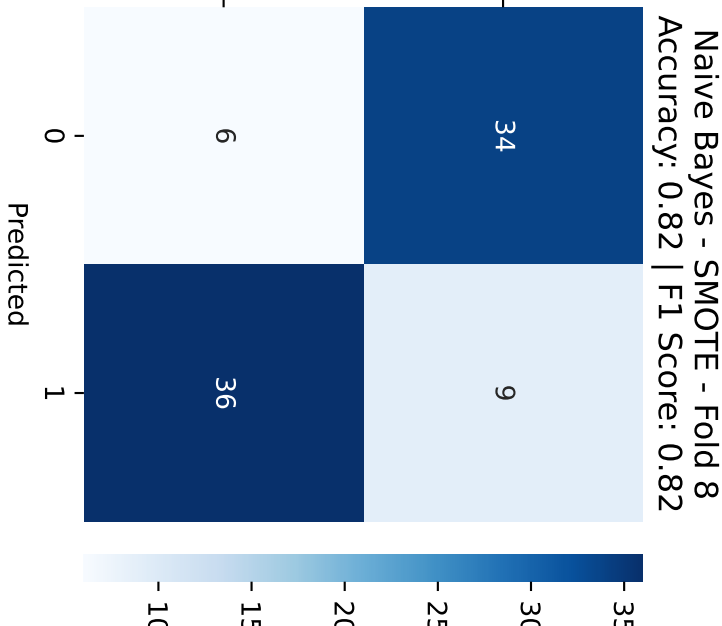
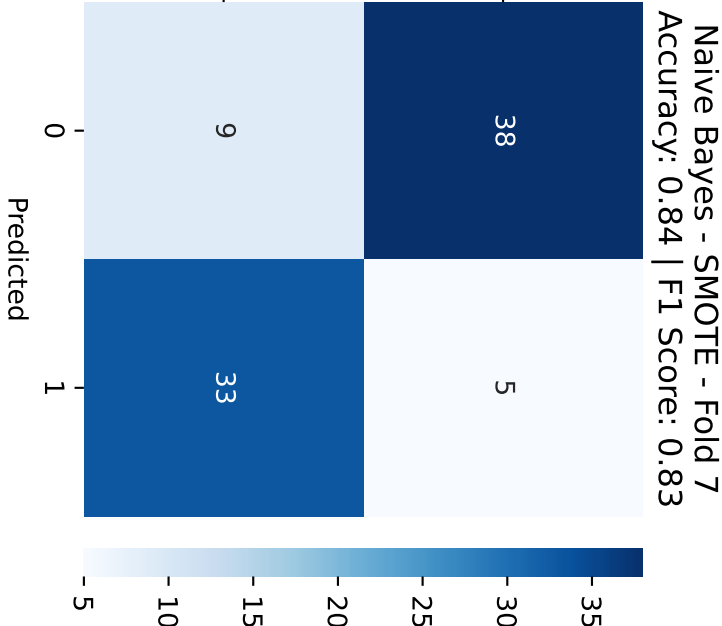
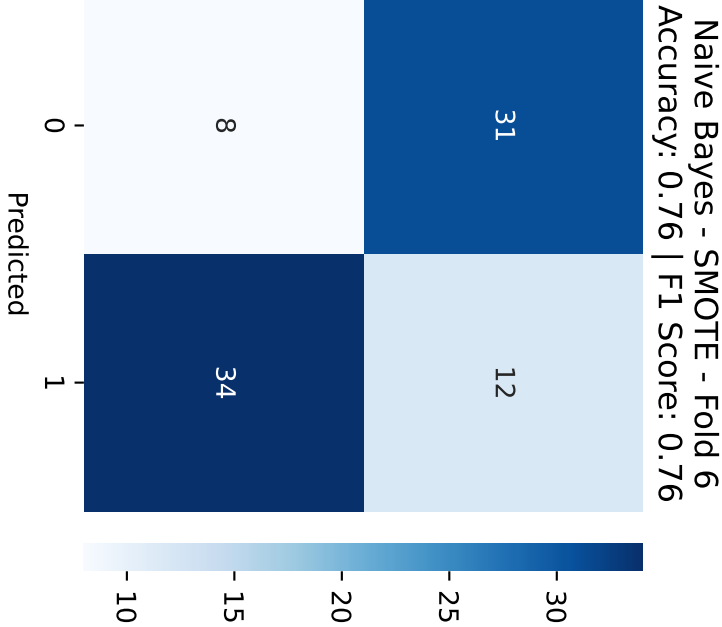
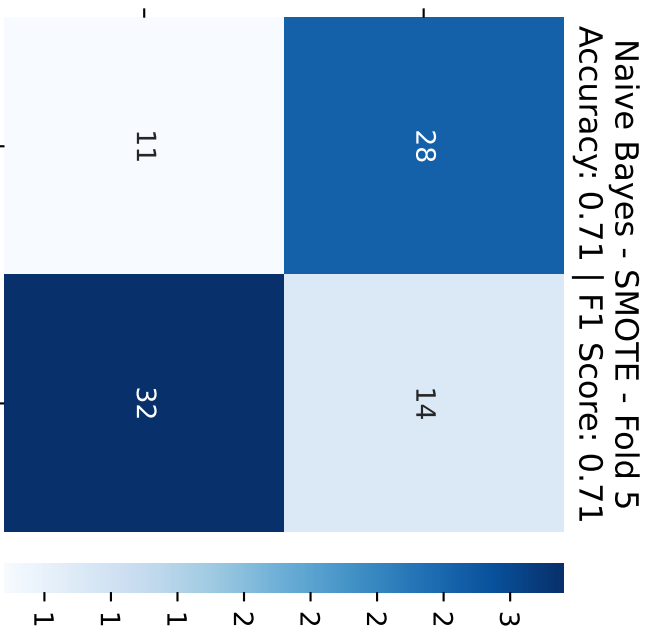
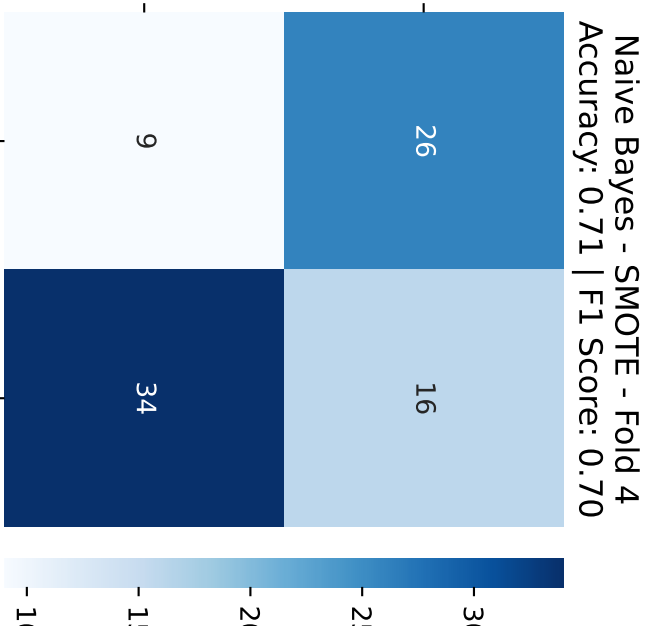
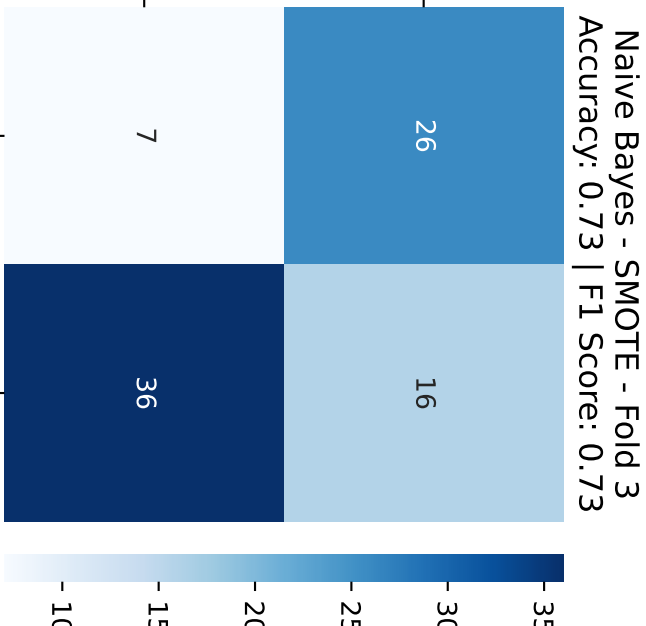
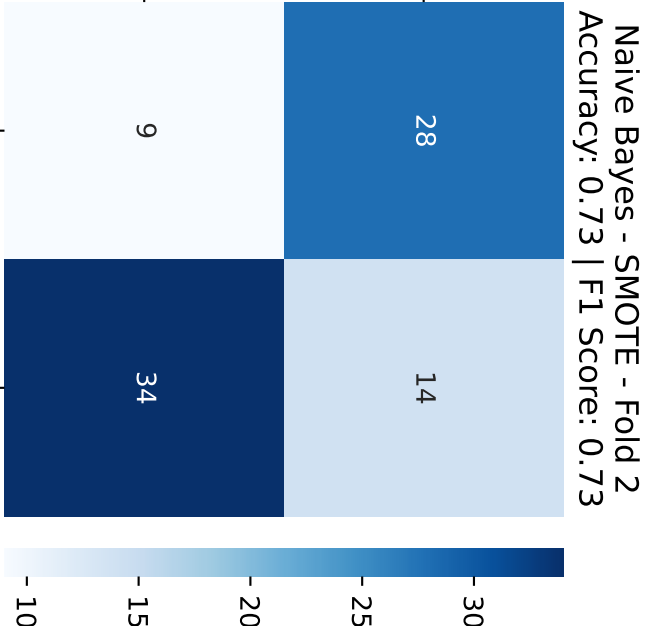
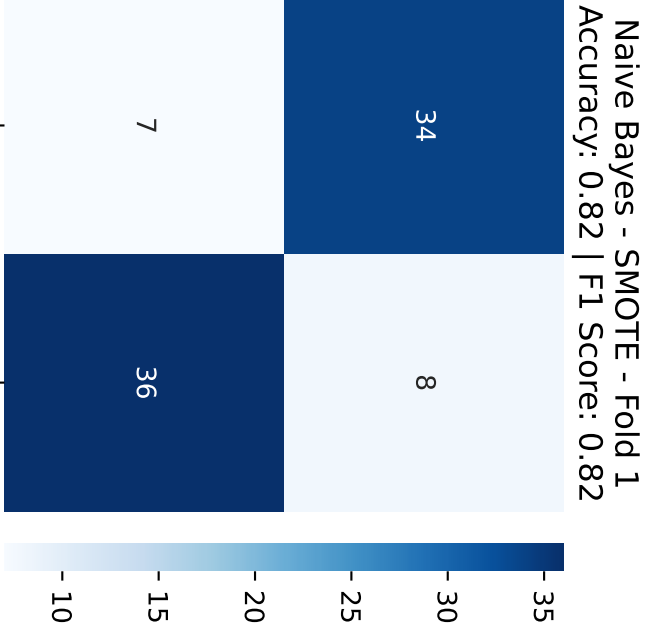
## EKA

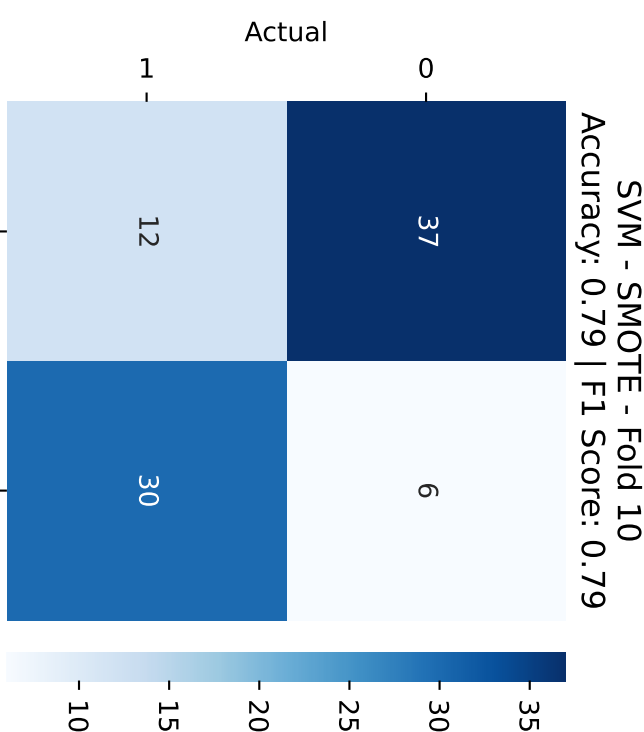
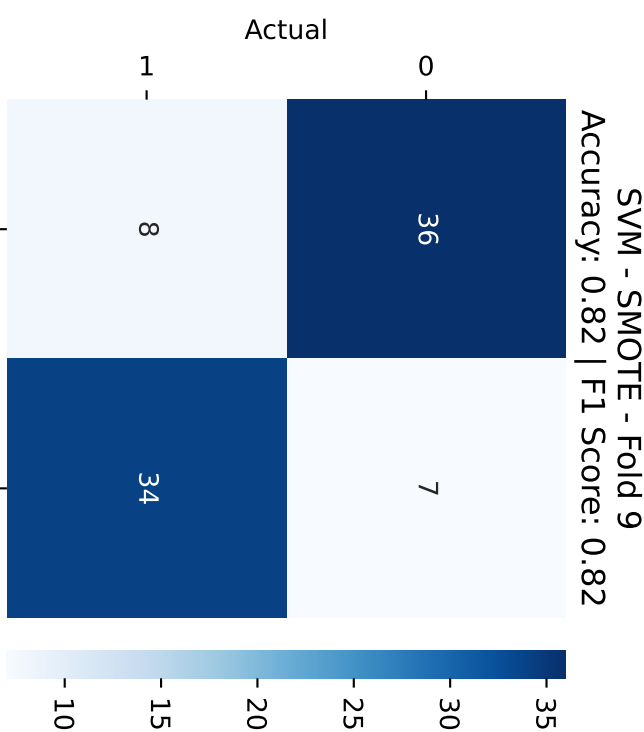
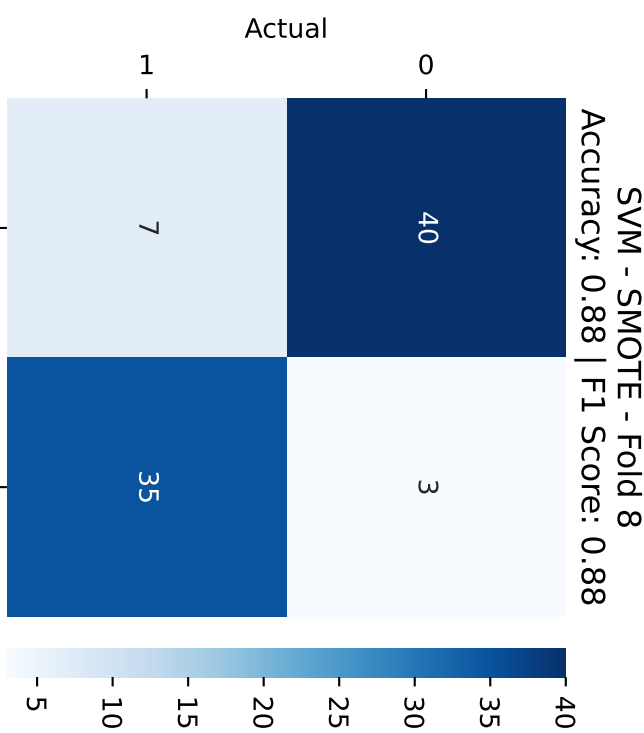
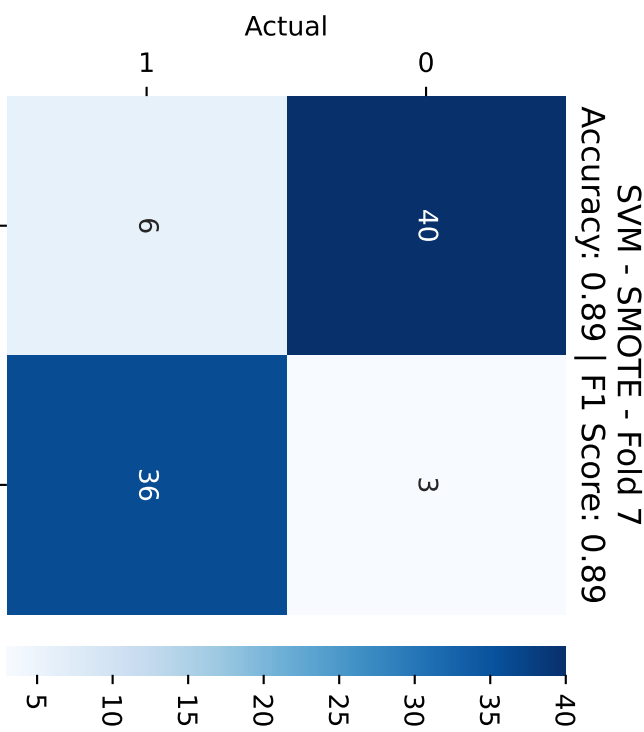
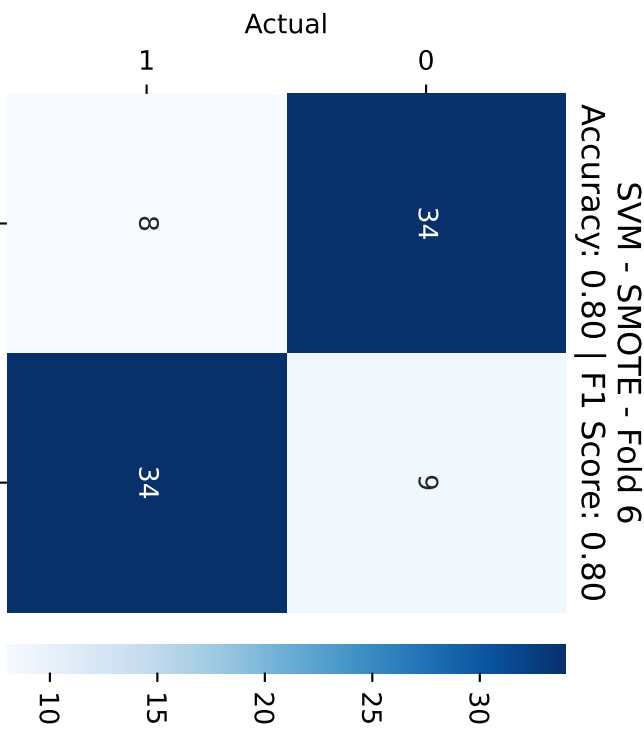
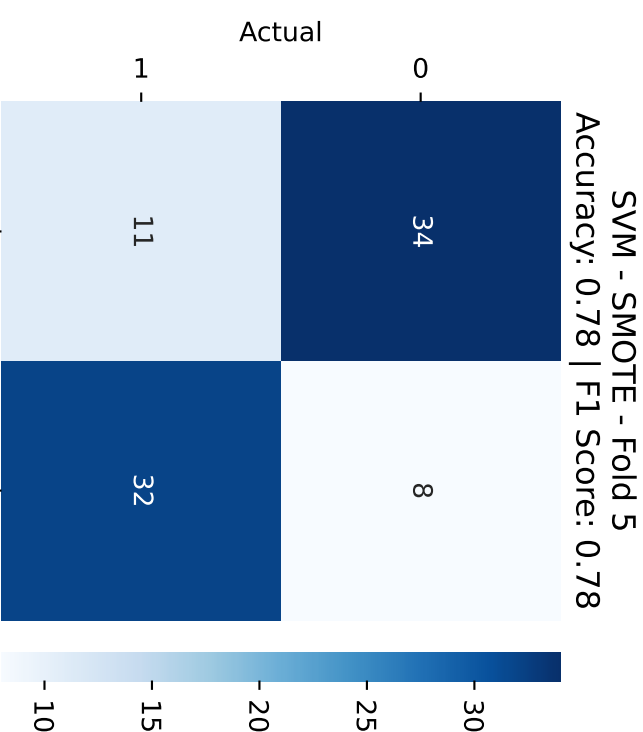
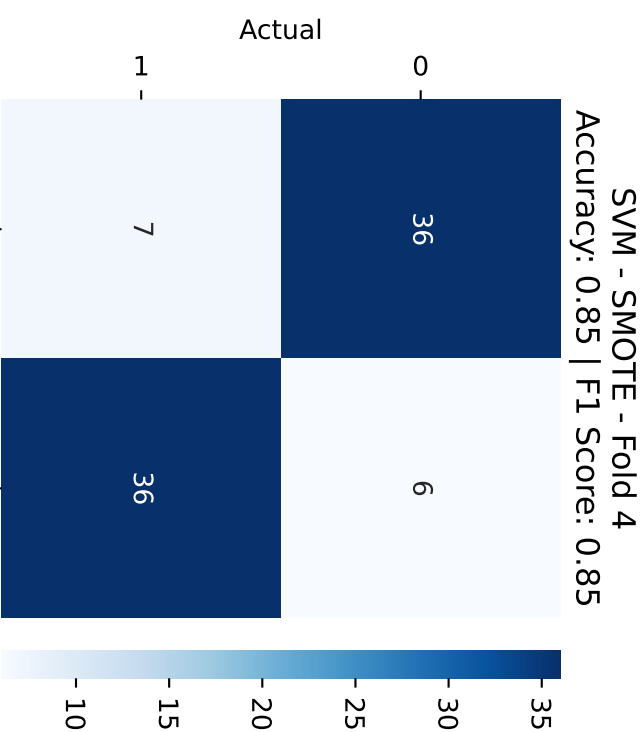
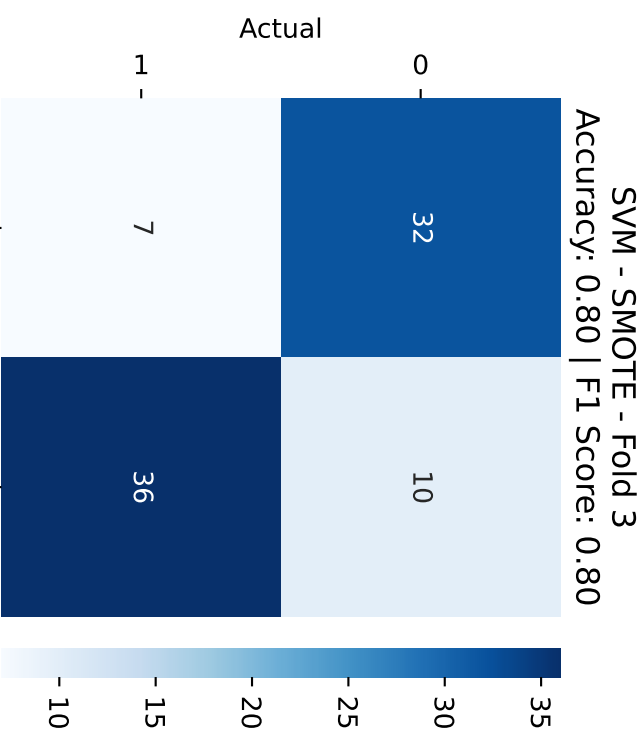
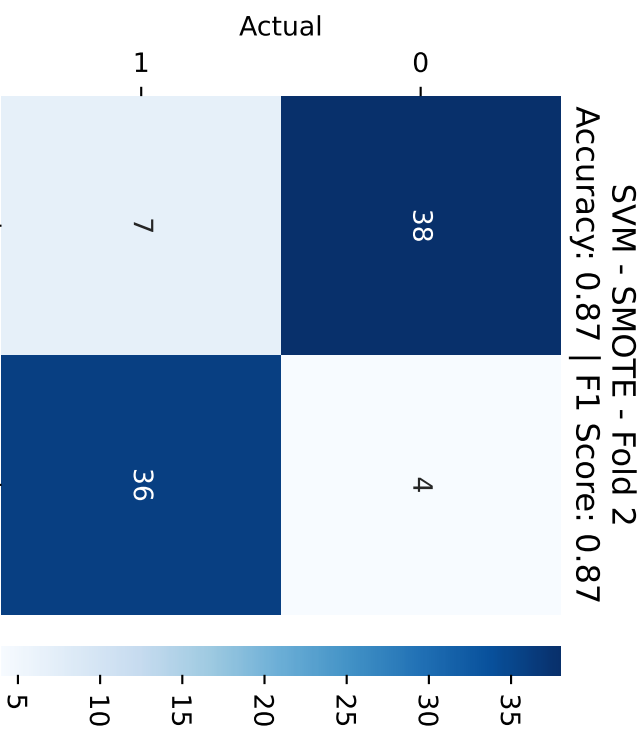
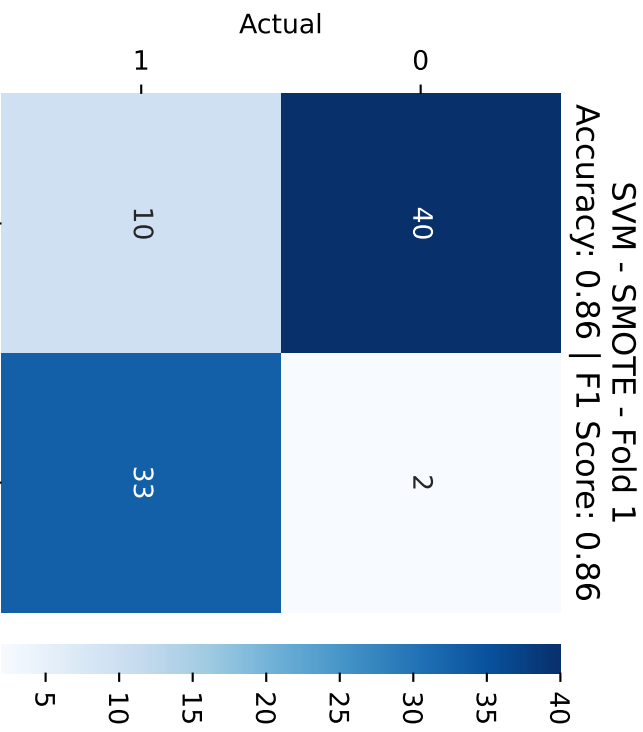
Çalışma sonucunda her bir PCA yaklaşımına ait örnekleme ve modelleme yöntemlerine ait doğruluk, F1 Skoru ve karmaşıklık matrislerini içerir.

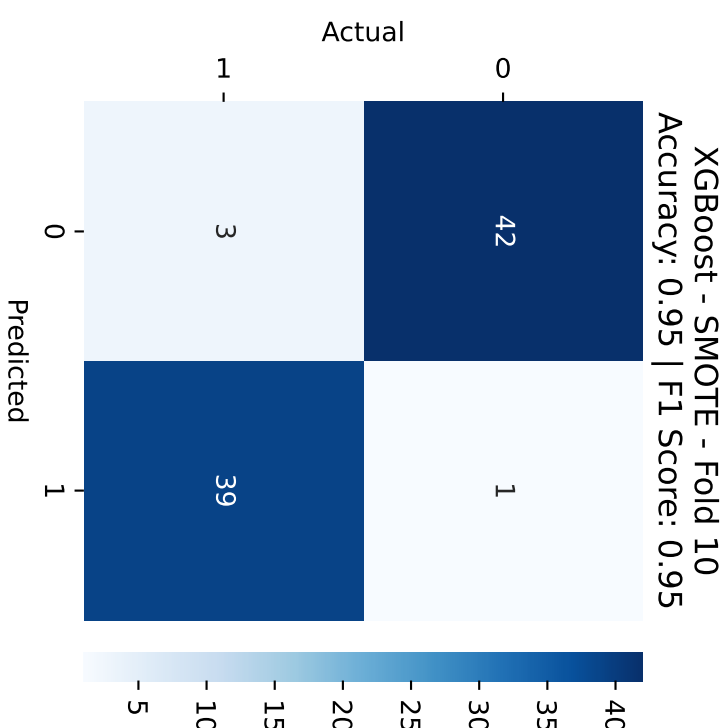
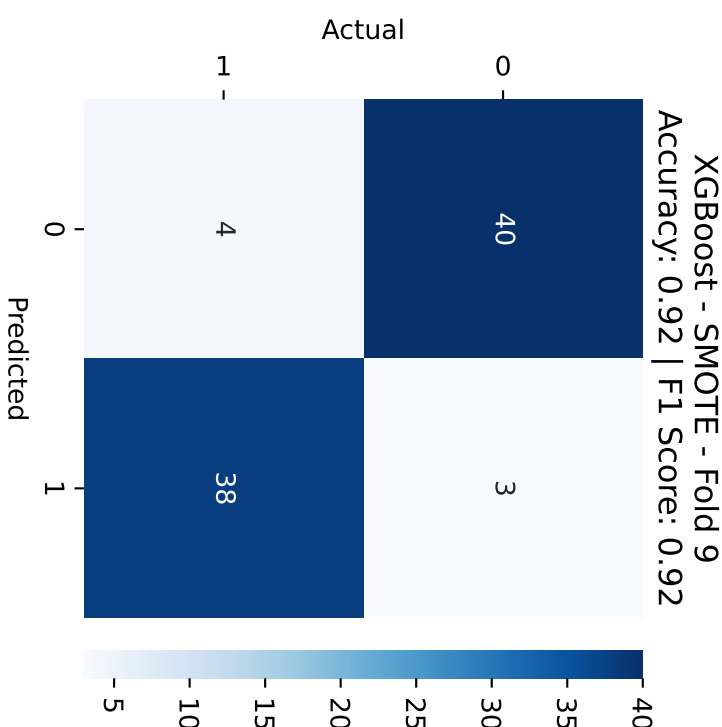
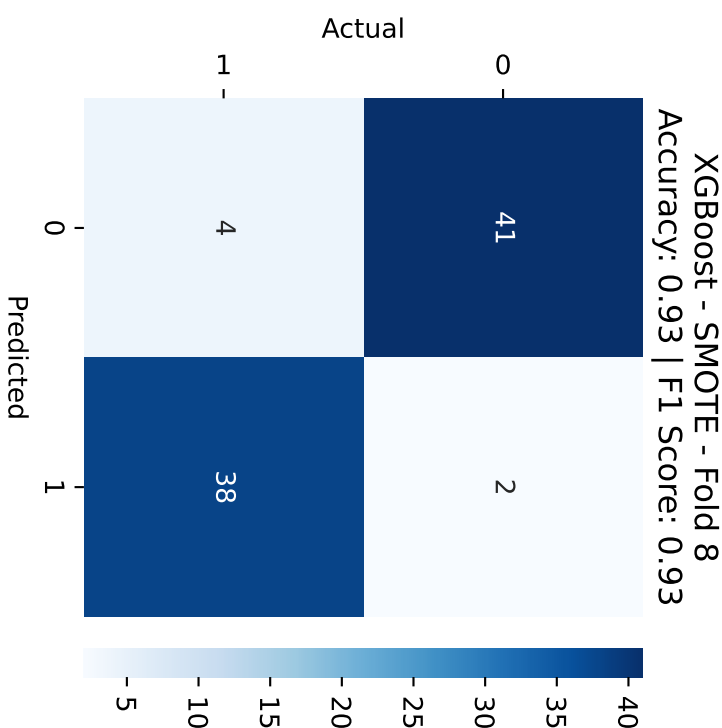
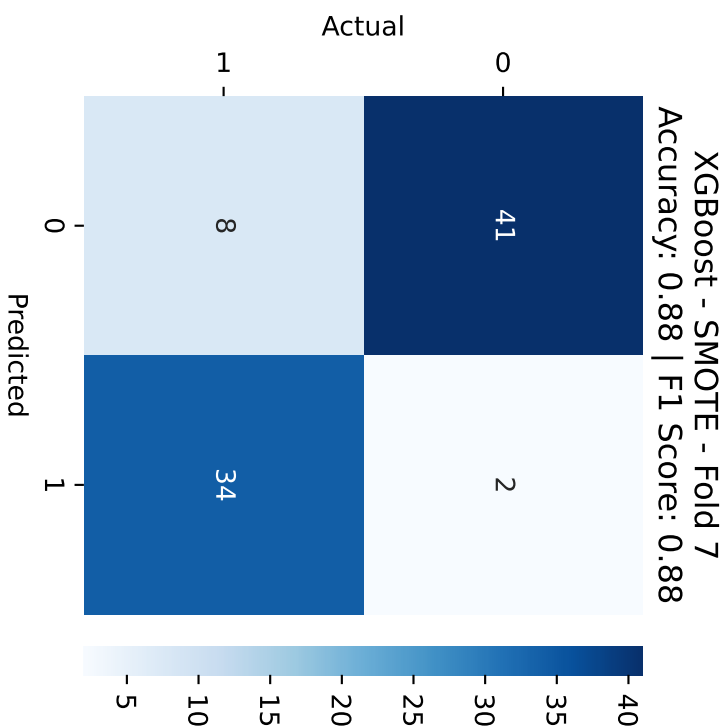
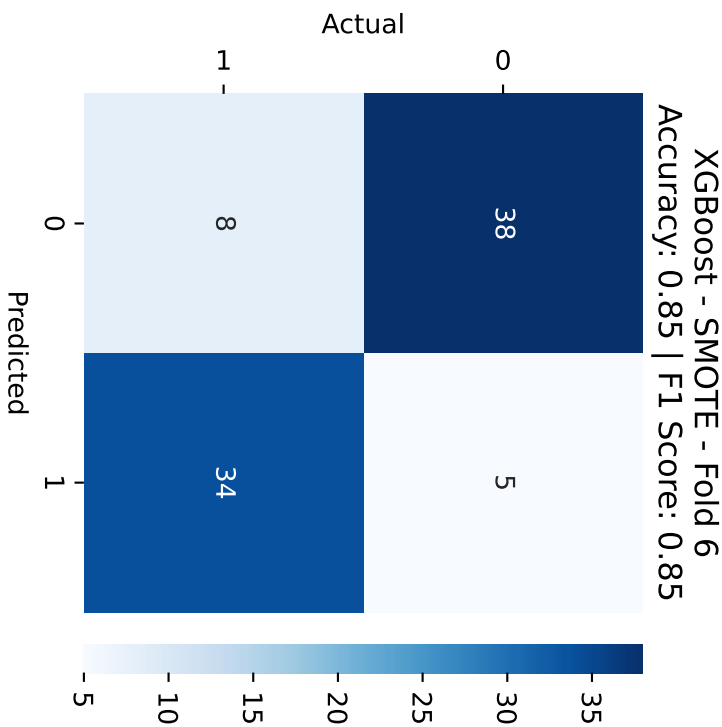
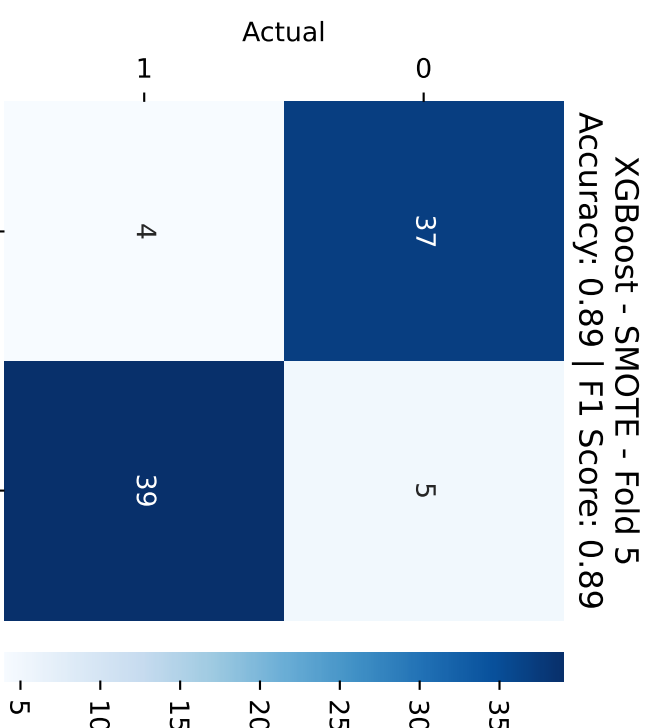
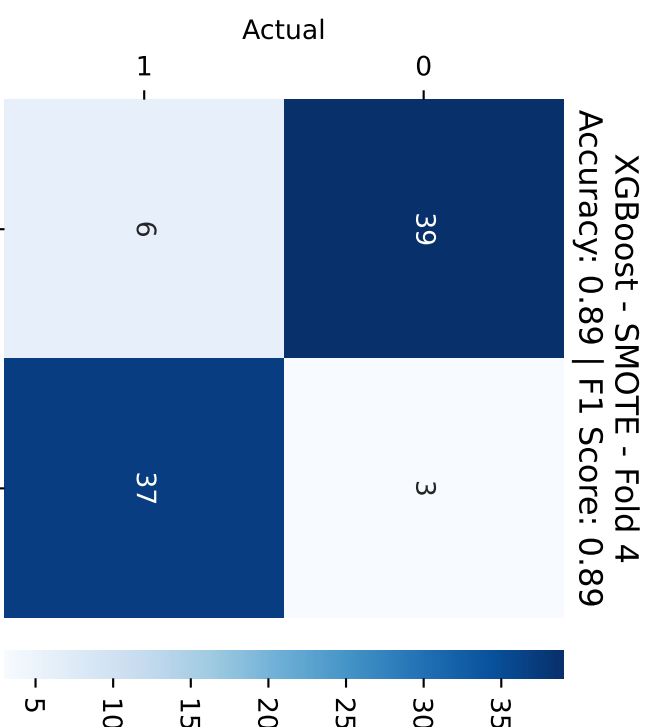
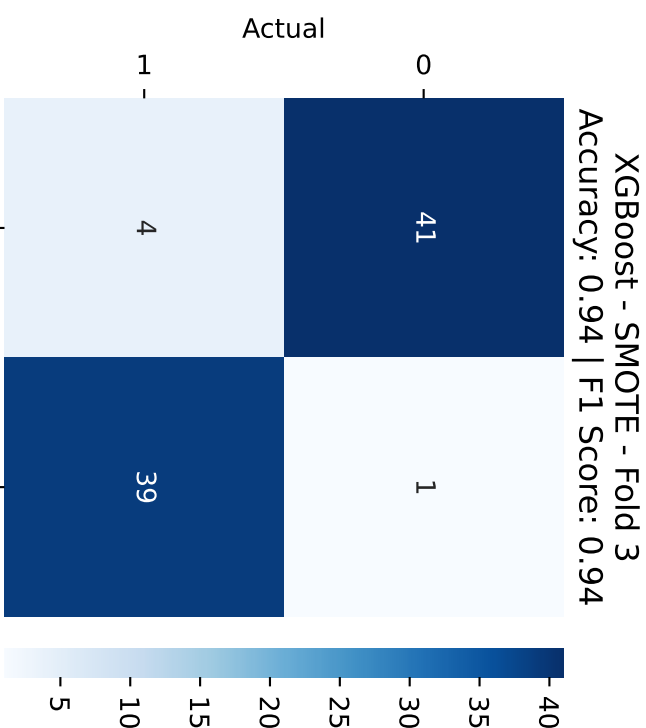
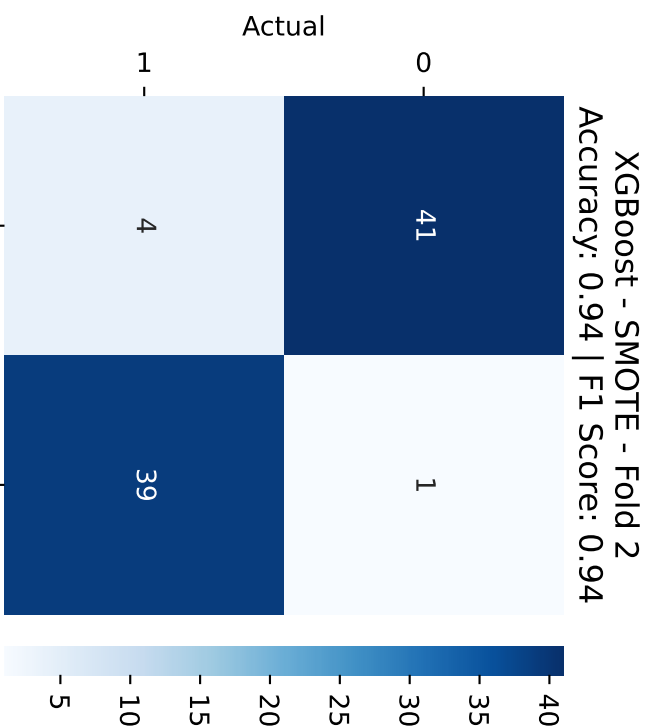
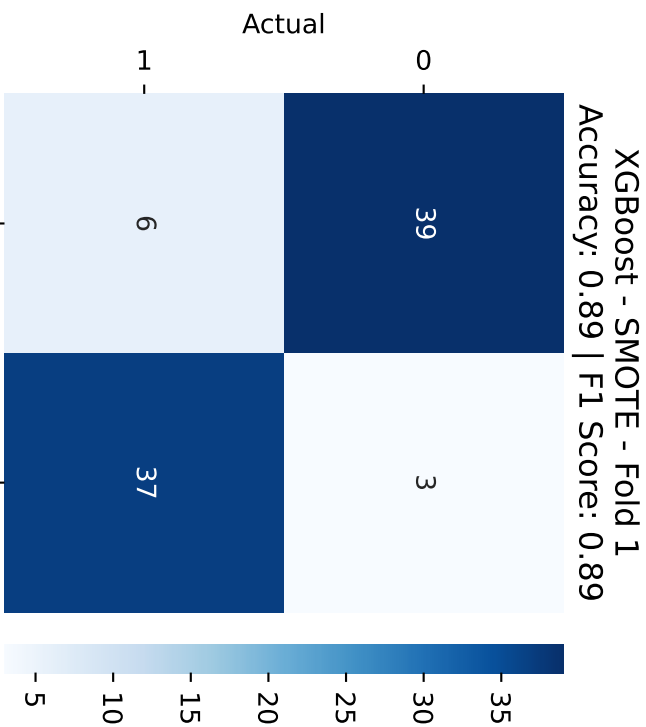


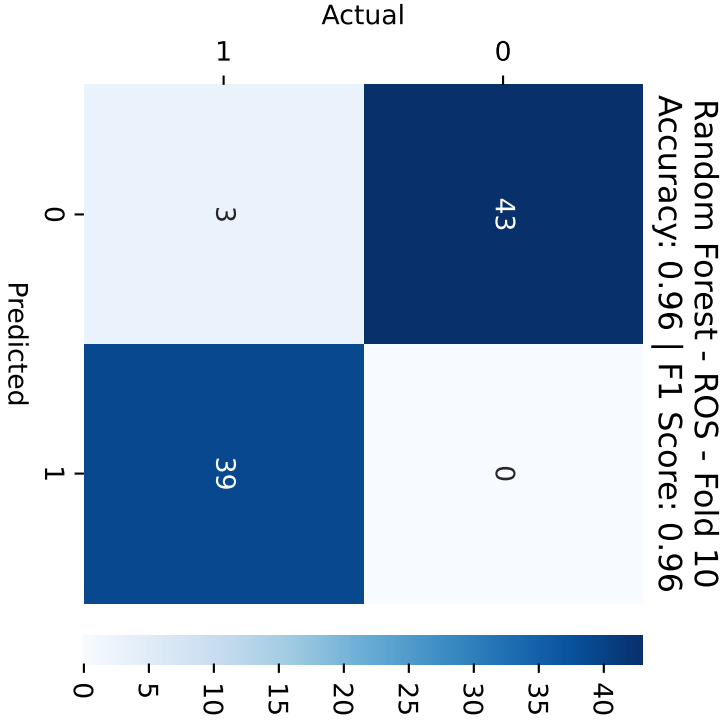
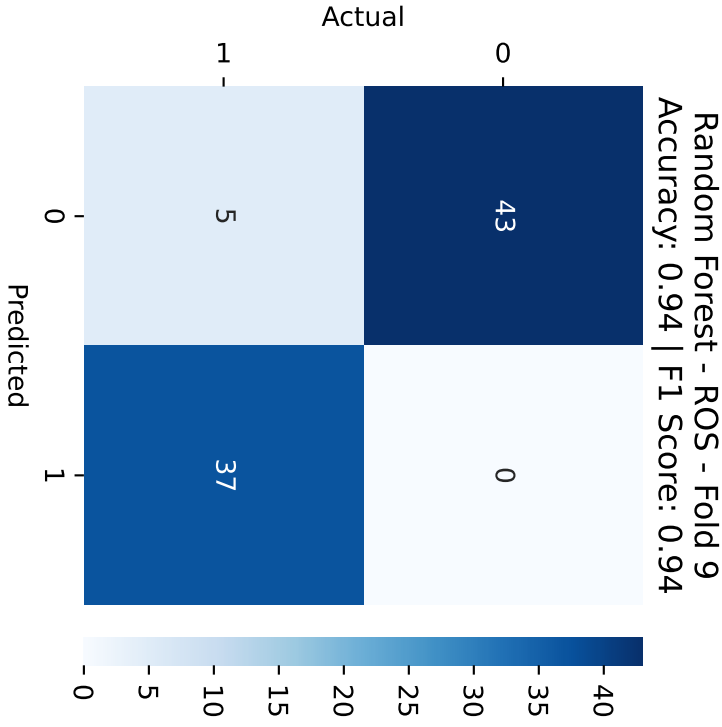
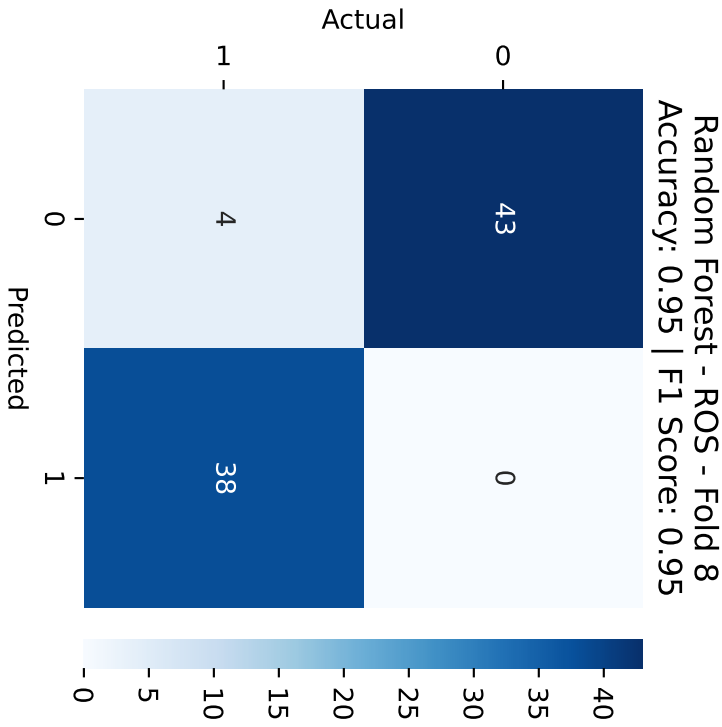
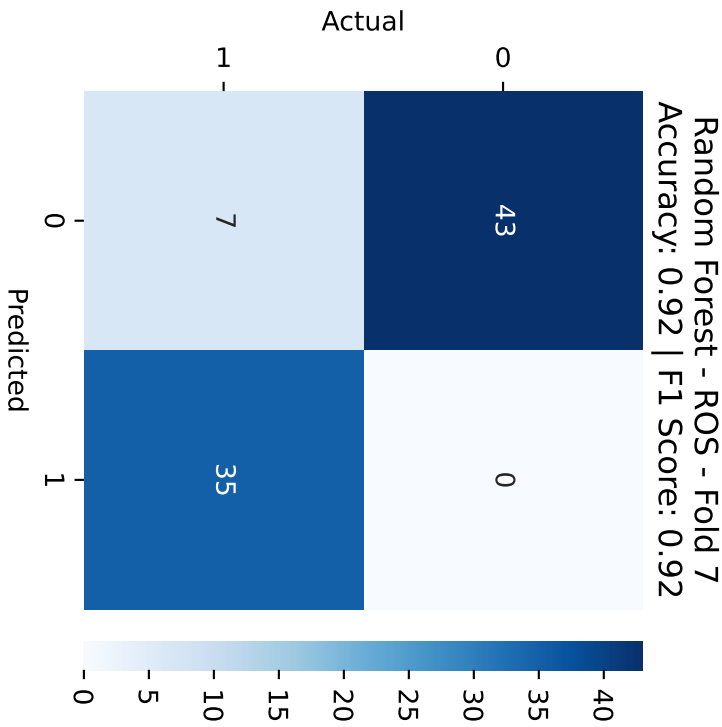
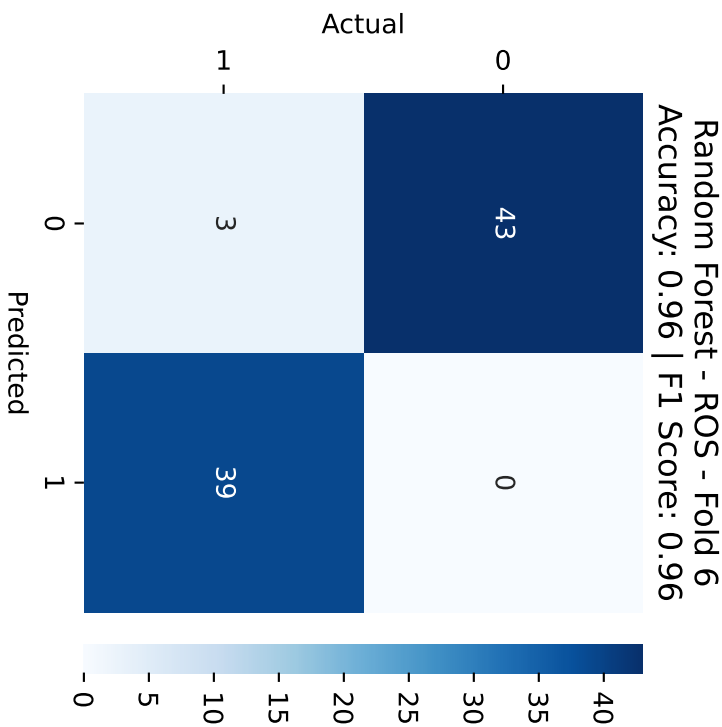
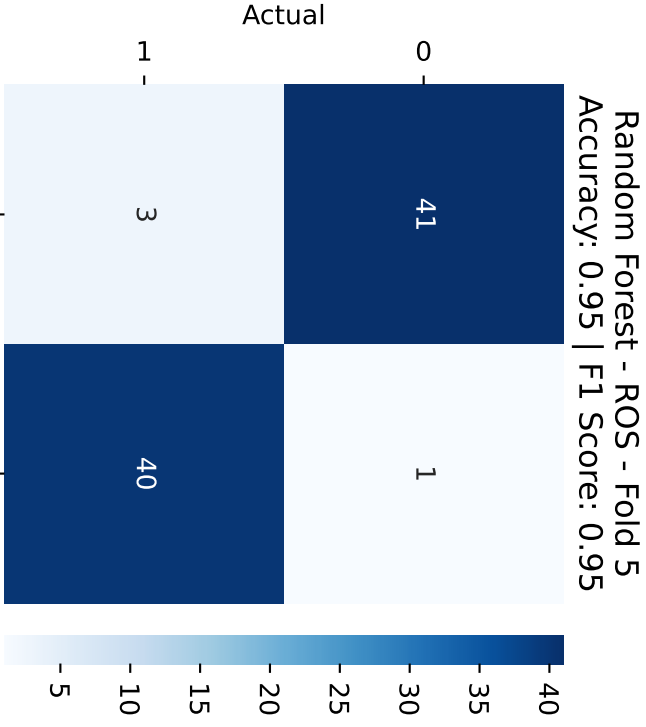
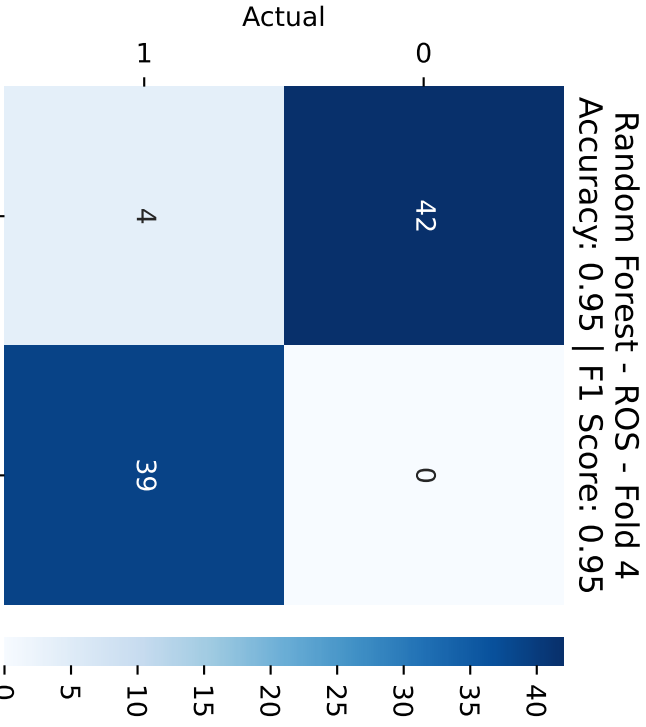
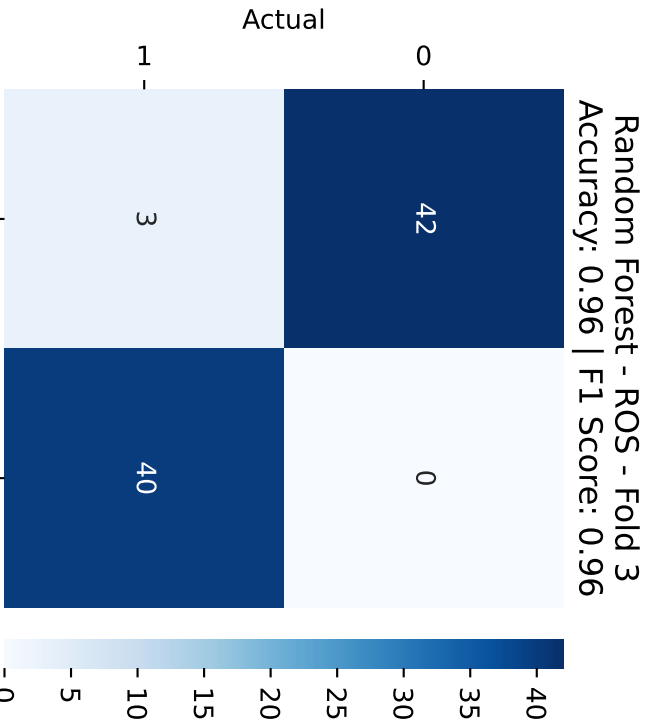
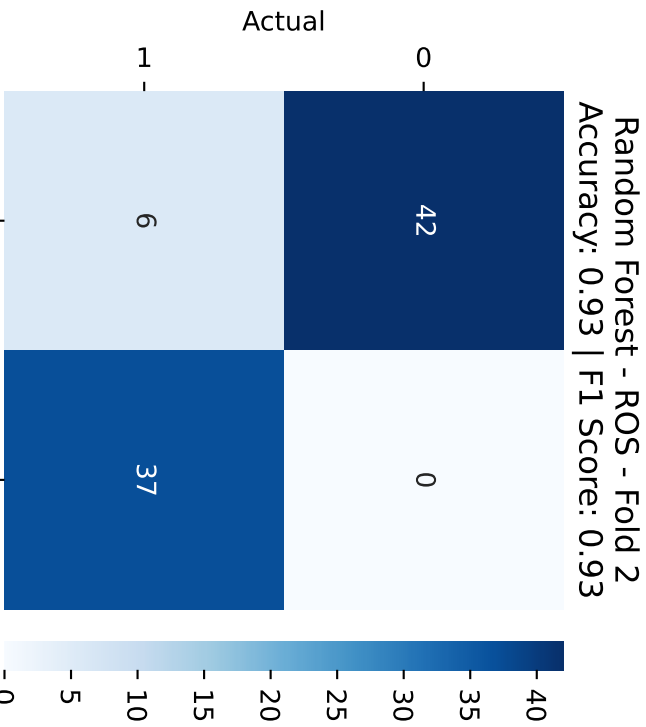
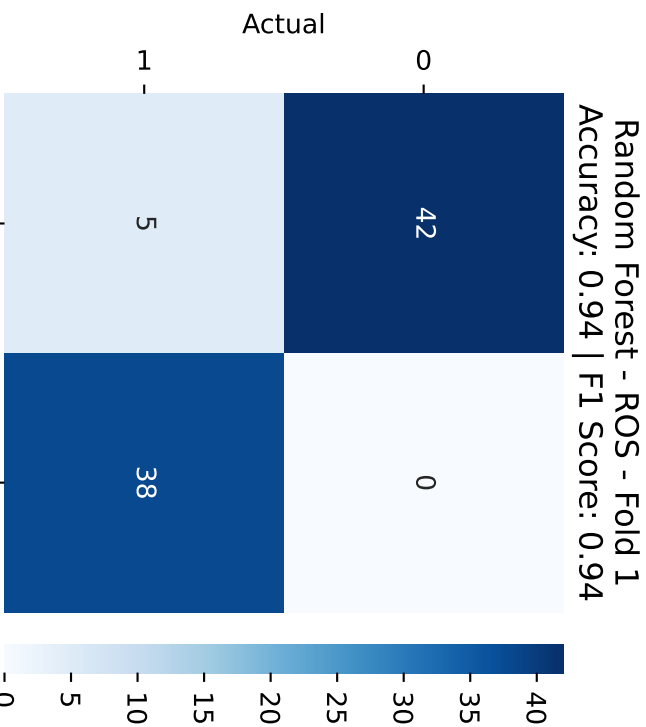


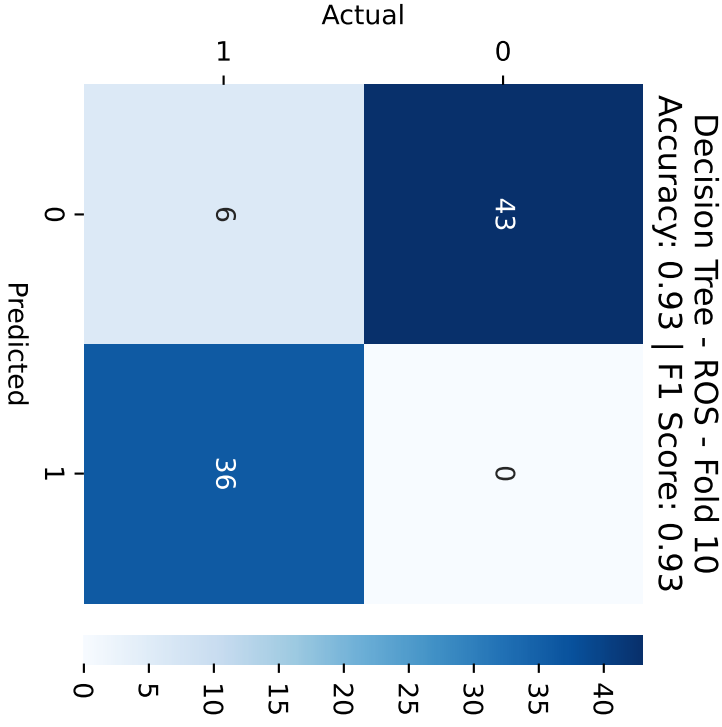
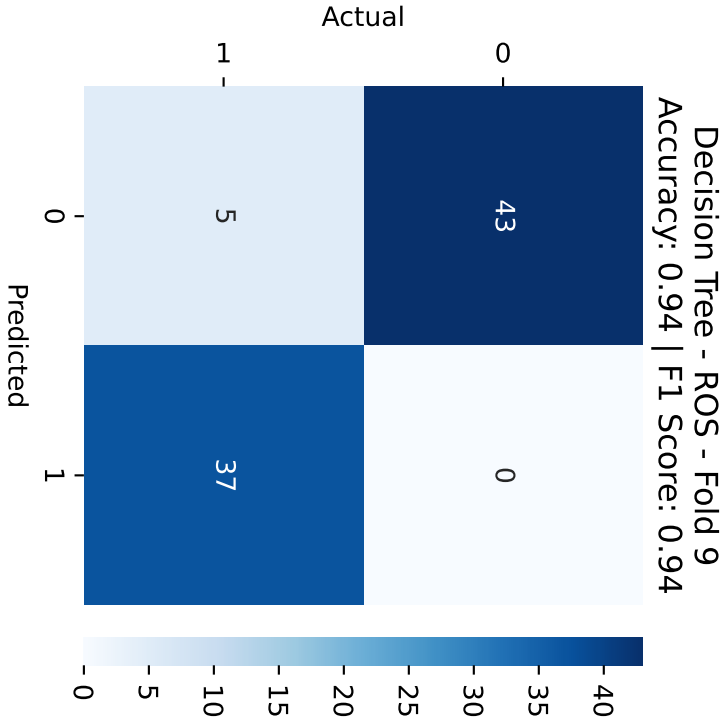
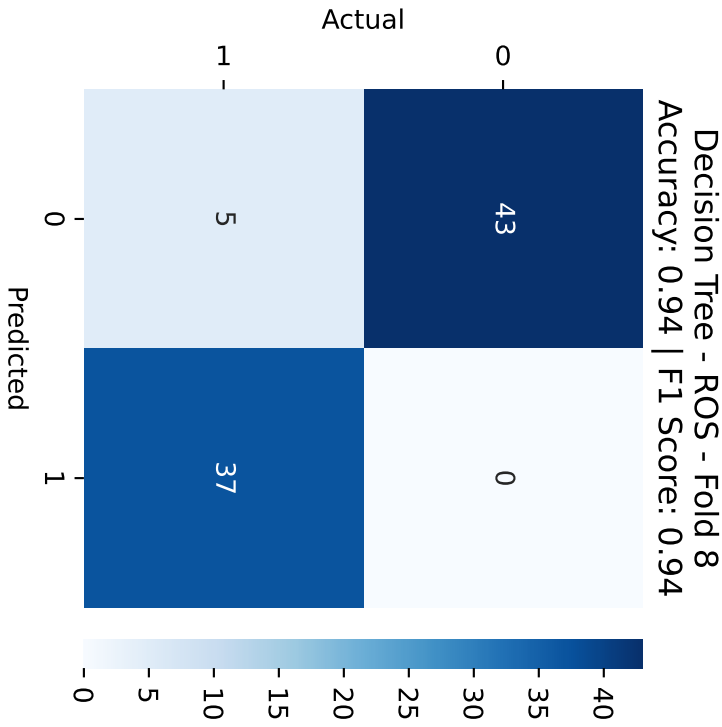
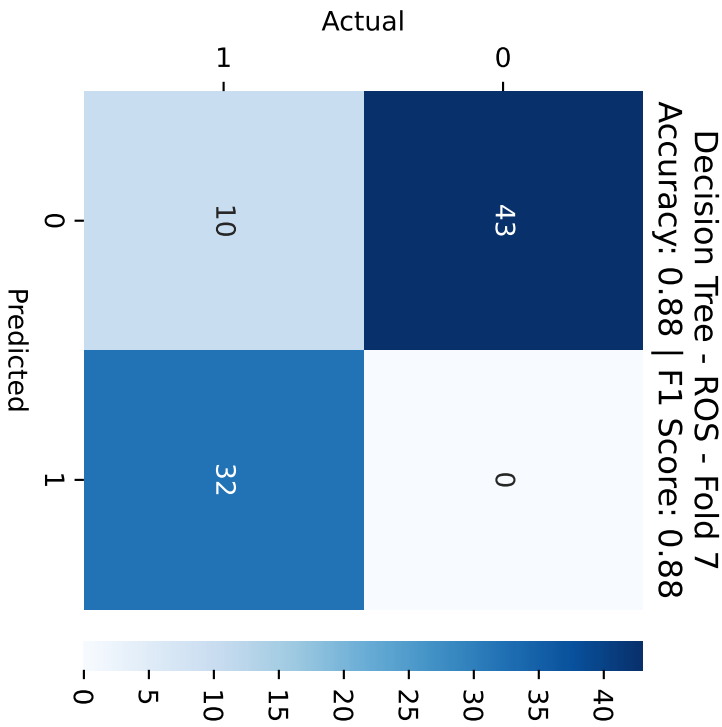
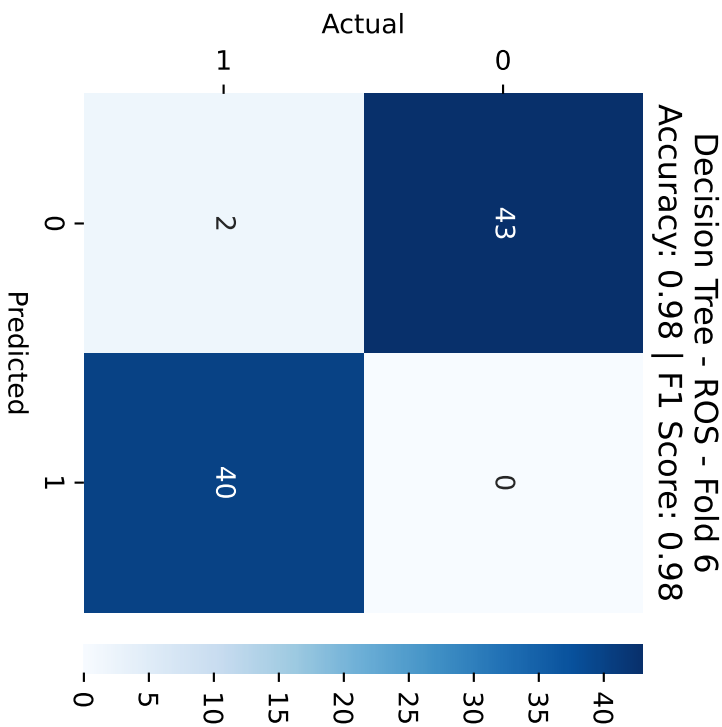
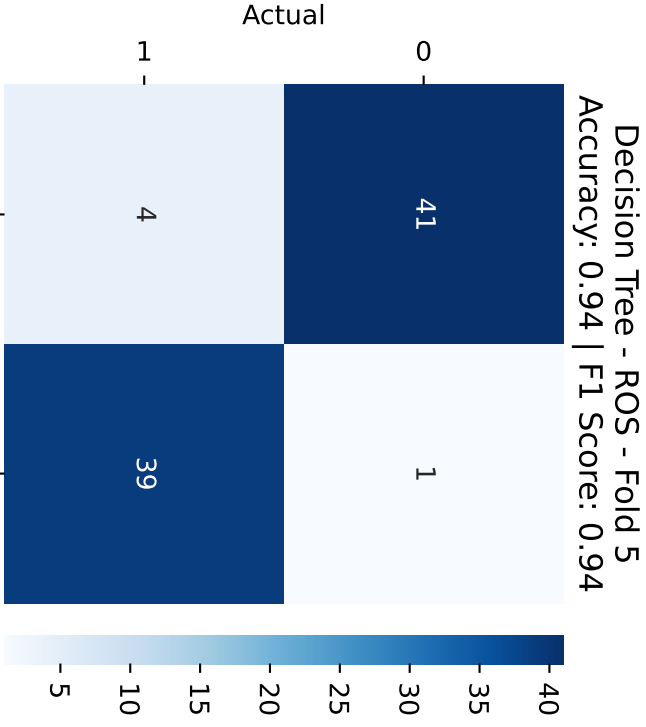
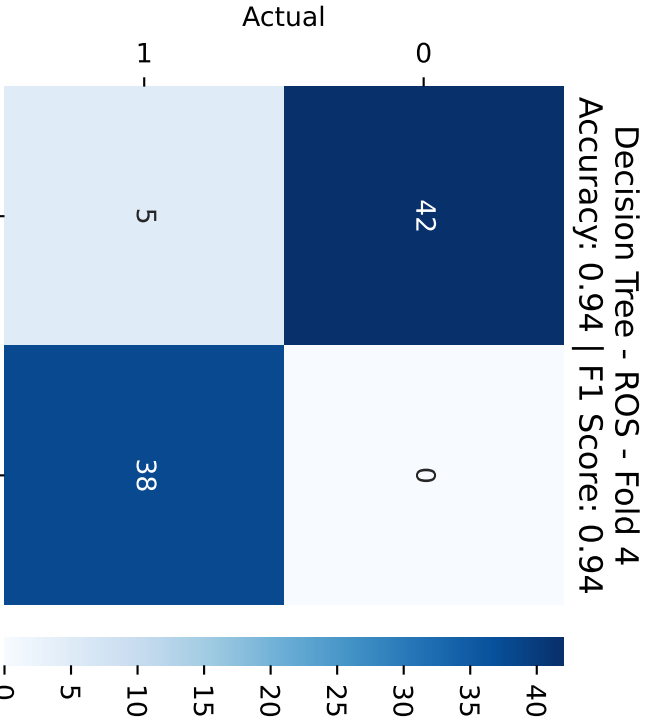
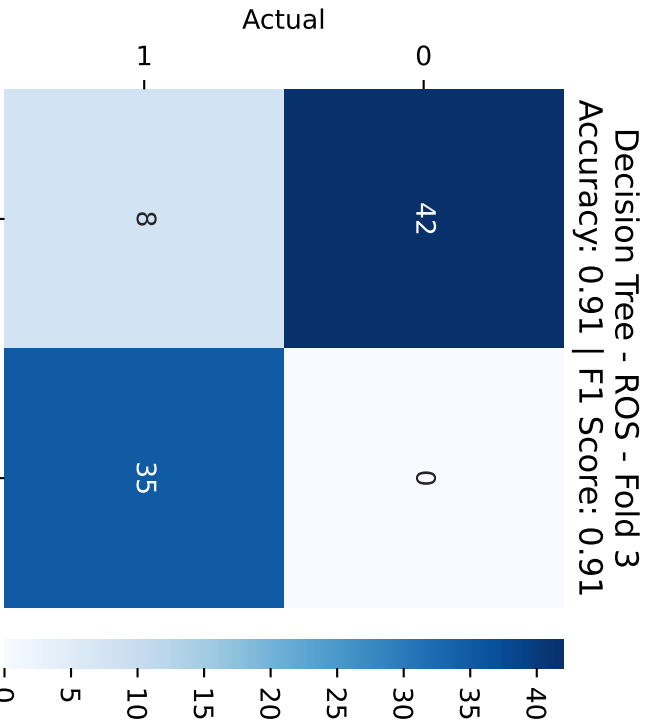
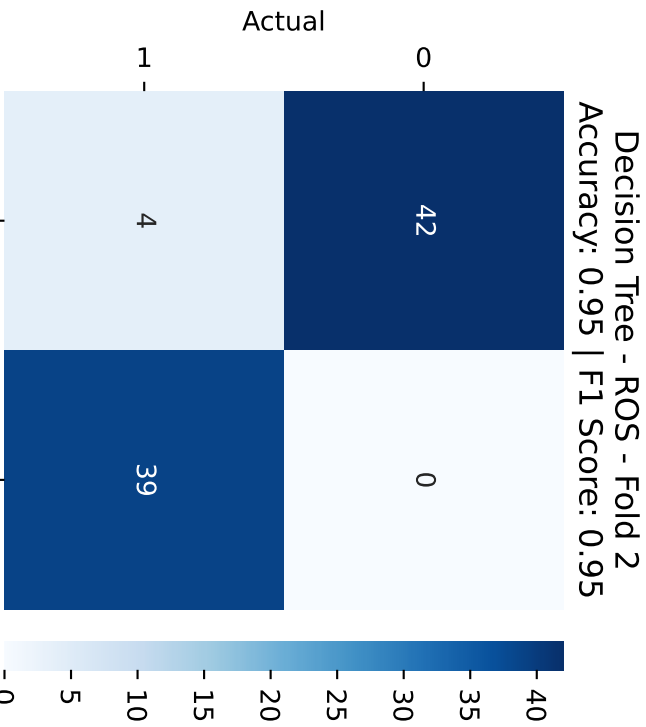
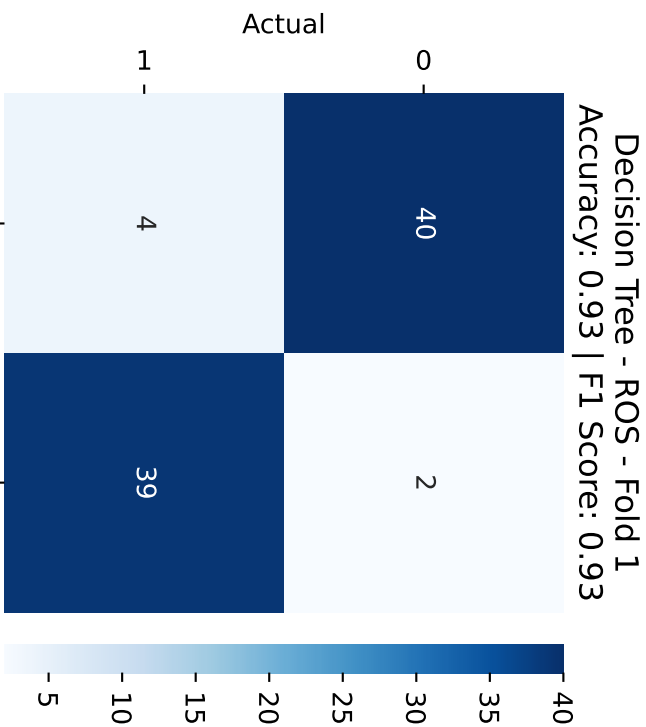


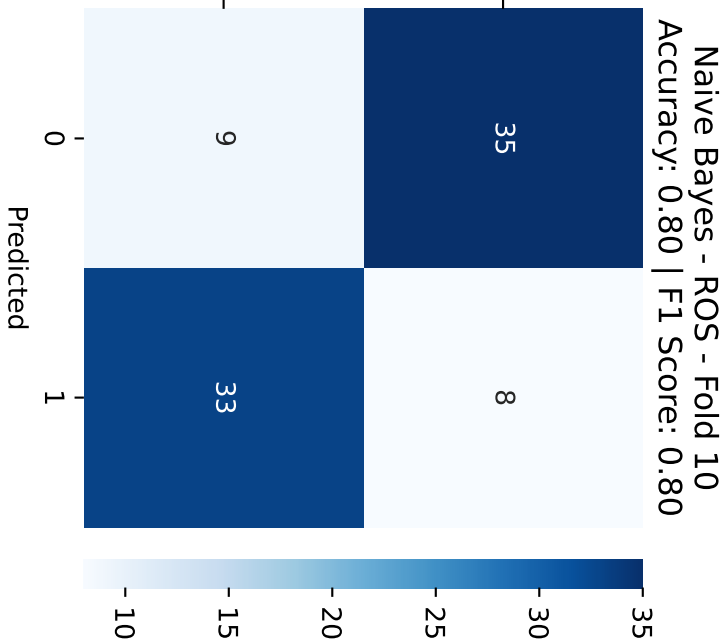
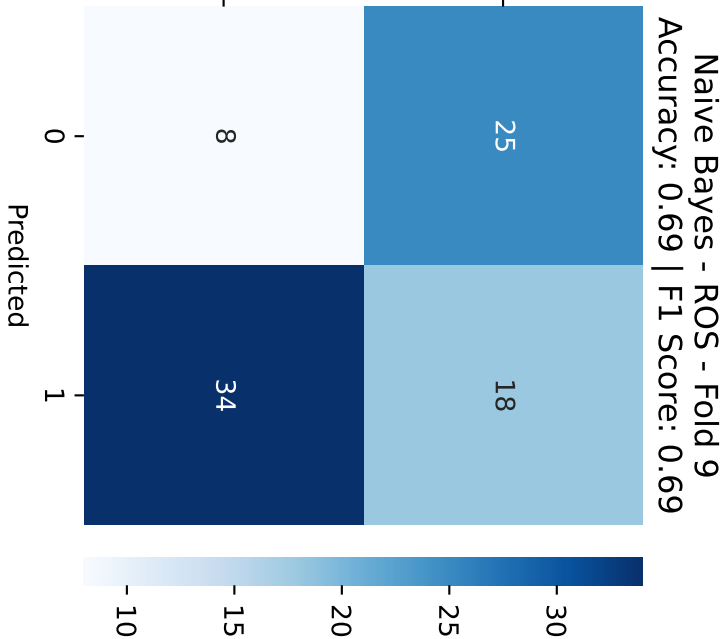
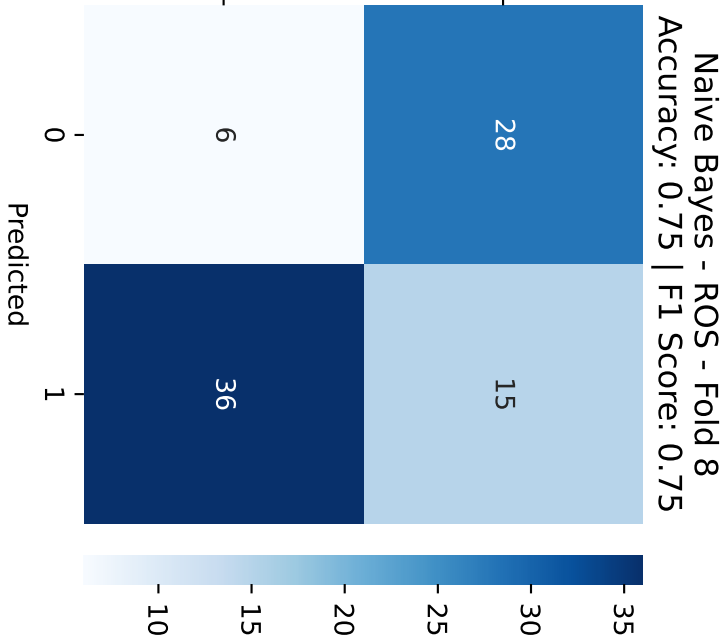
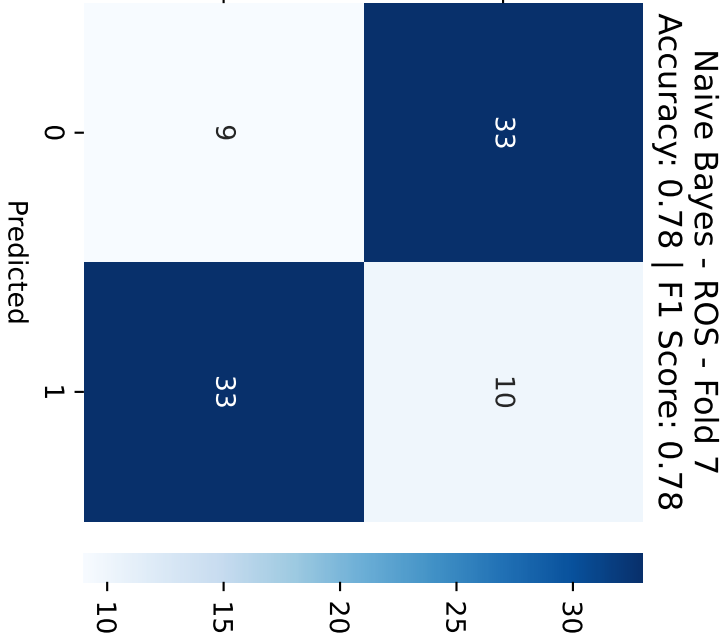
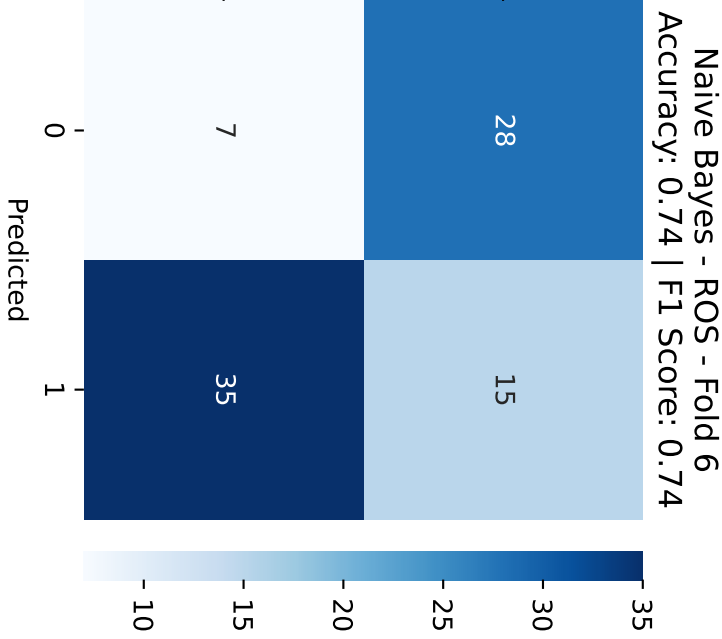
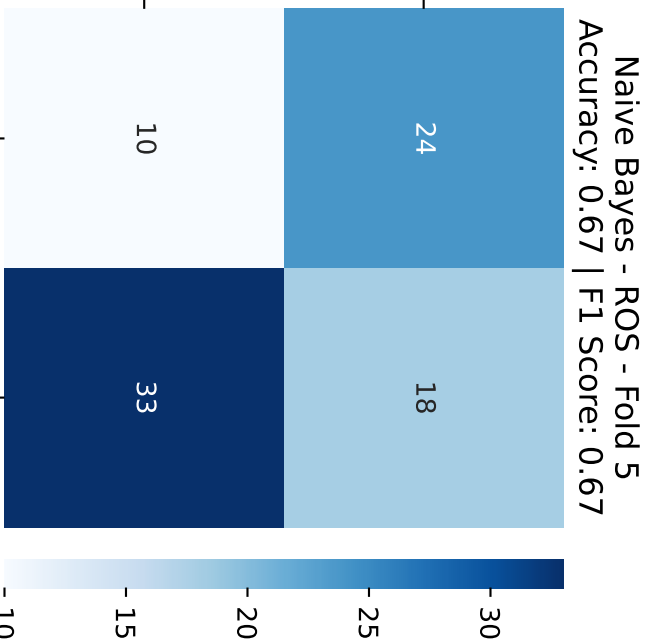
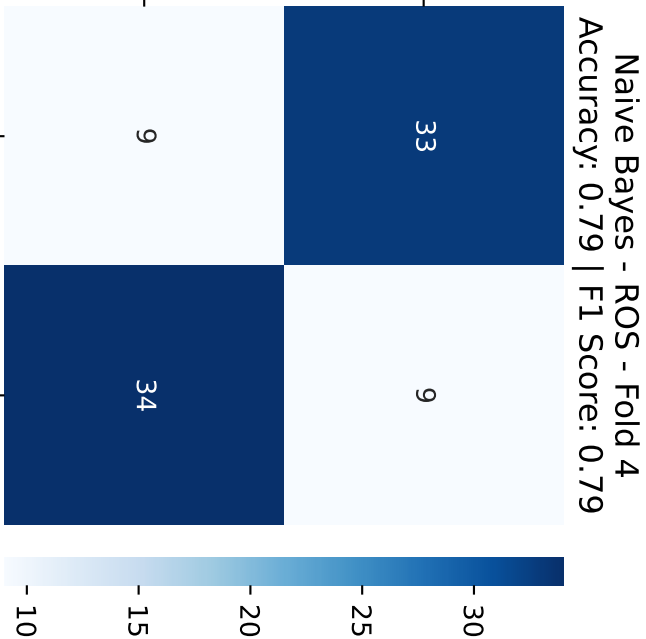
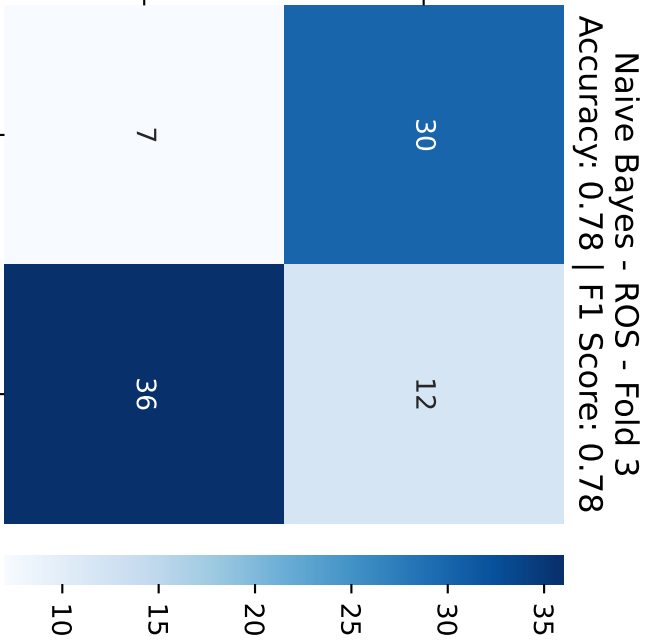
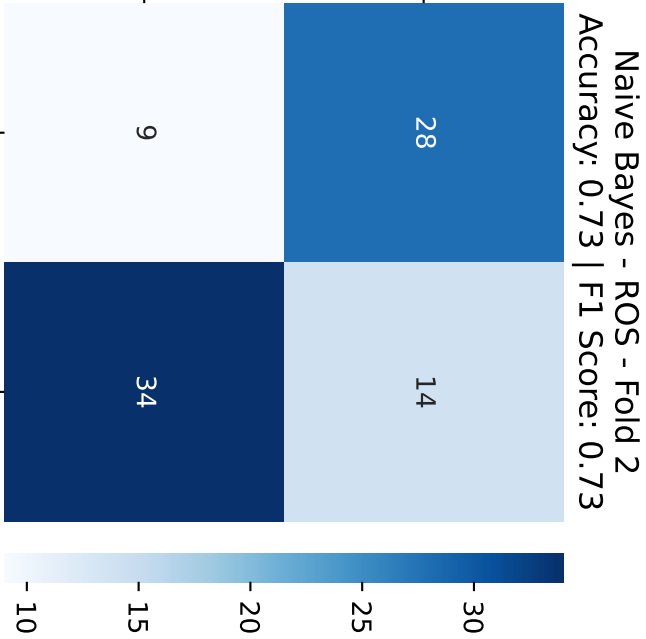
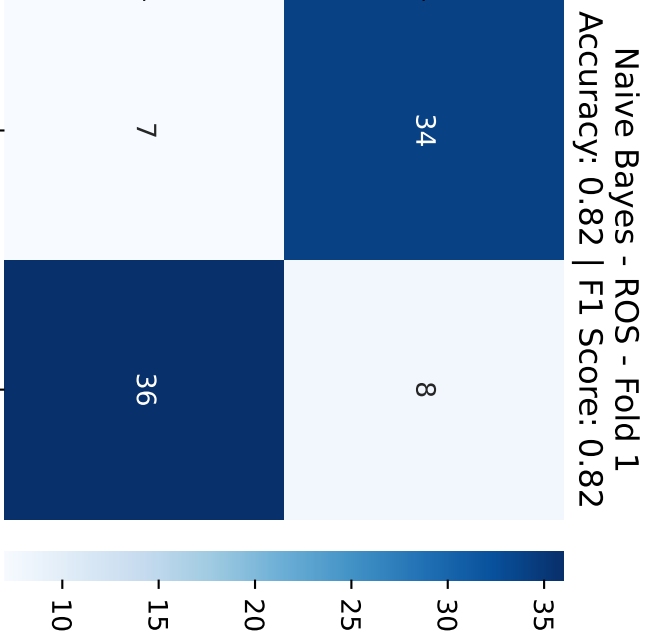


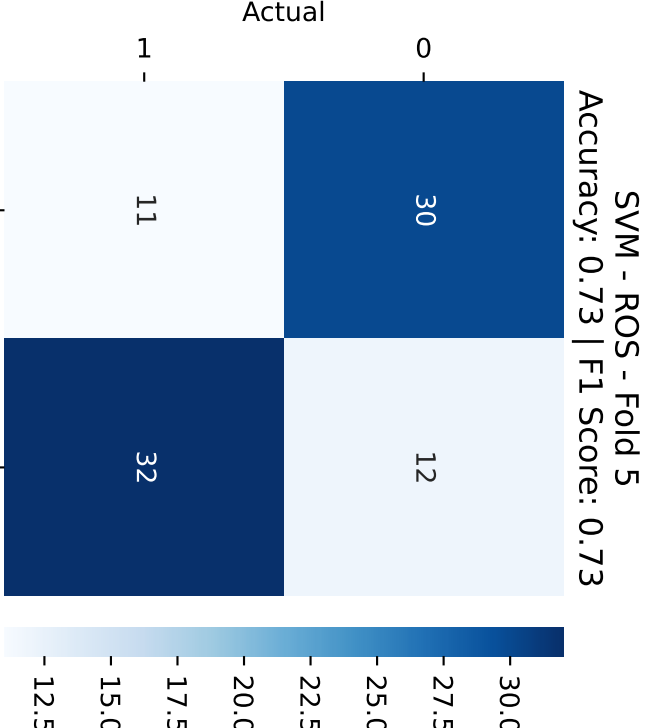
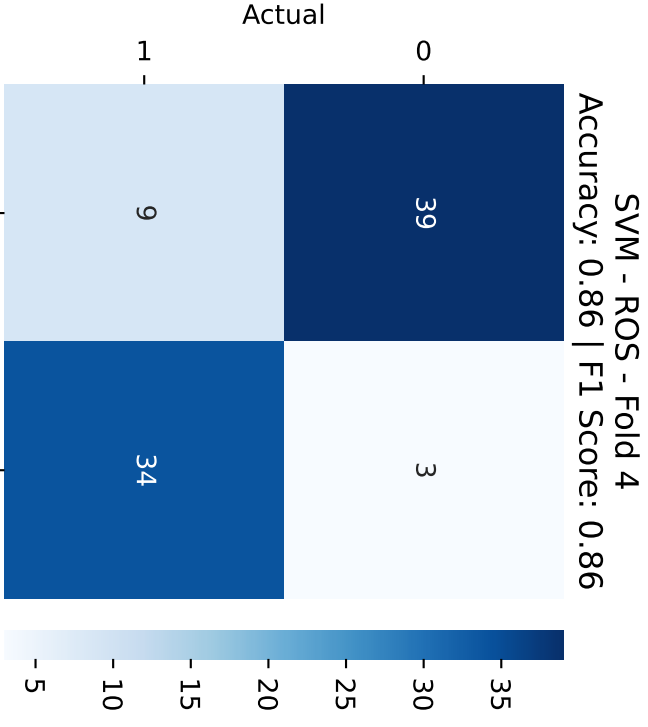
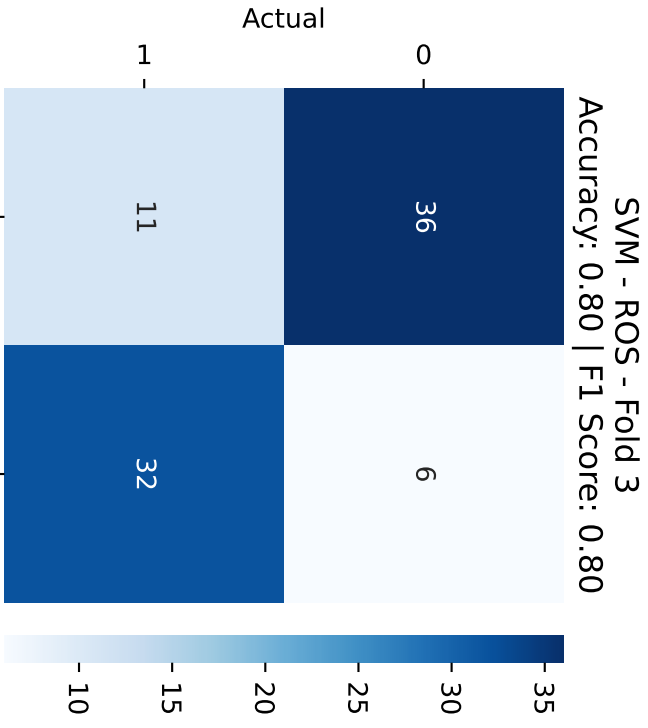
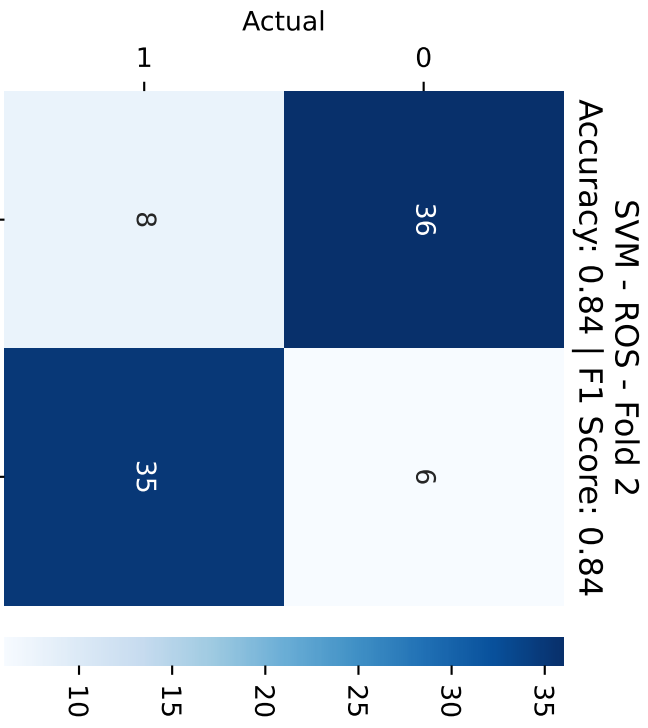
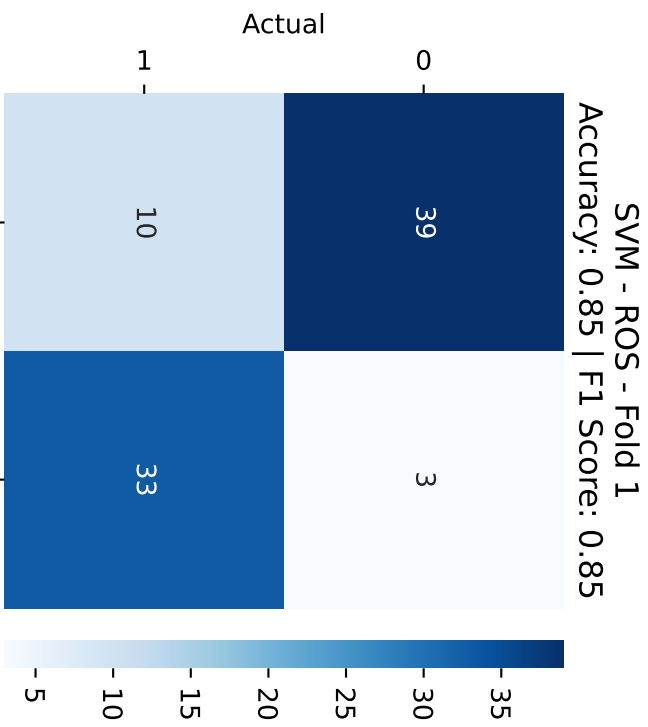




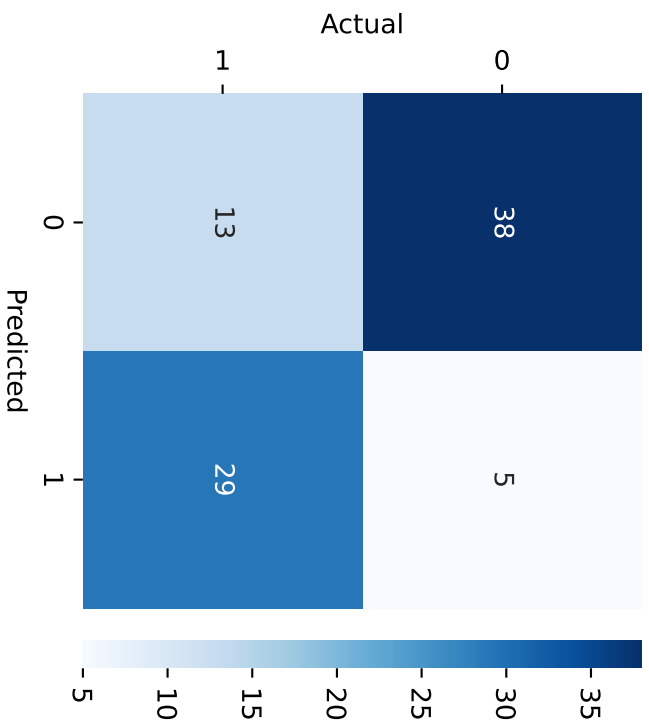




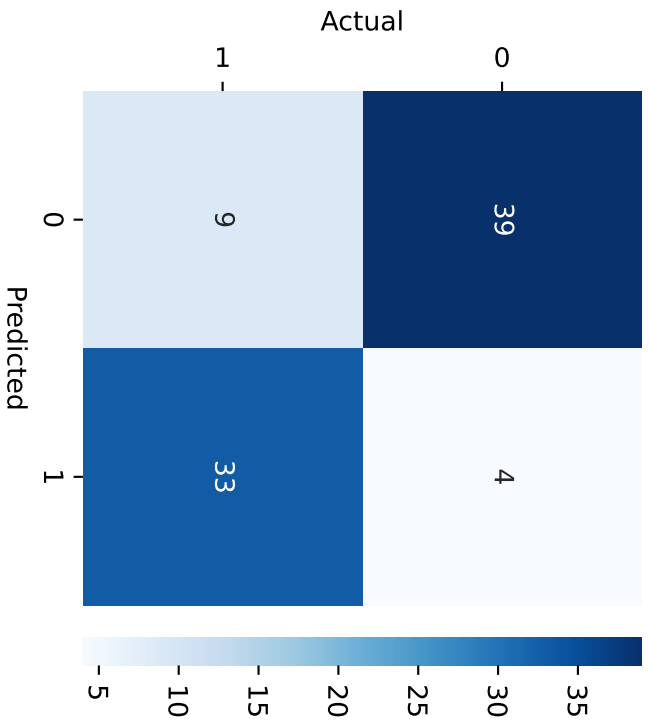




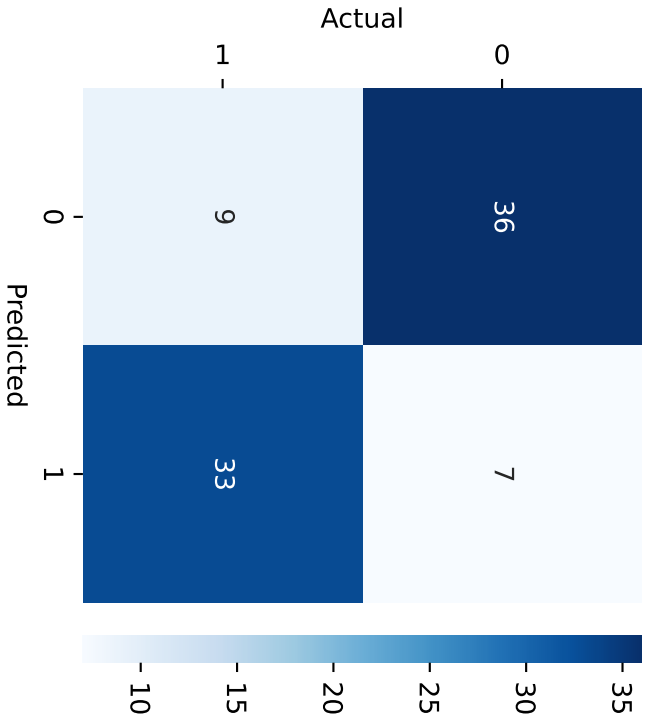
SVM - ROS - Fold 6  
Accuracy: 0.79 | F1 Score: 0.79



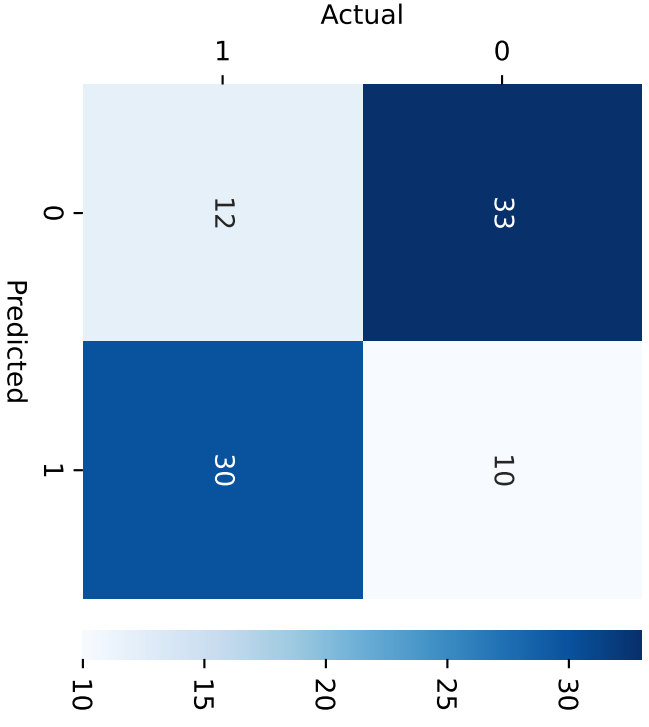
SVM - ROS - Fold 7  
Accuracy: 0.85 | F1 Score: 0.85



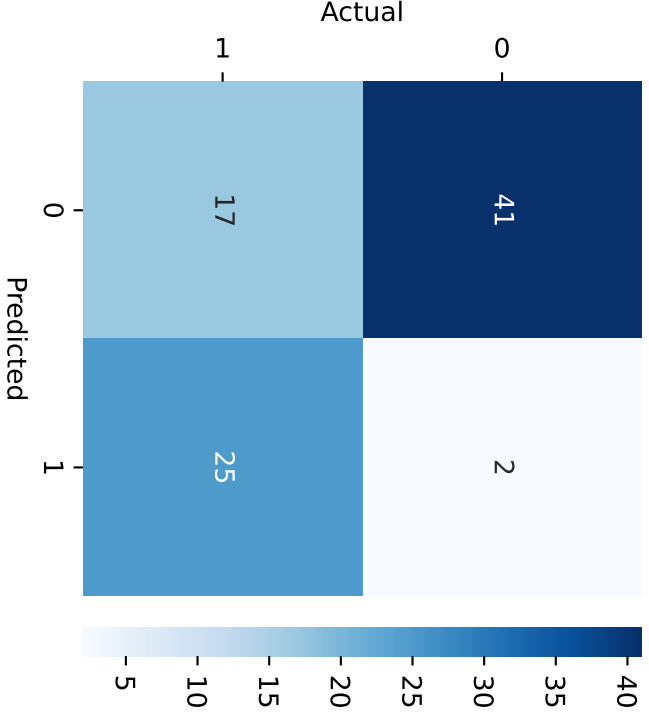
SVM - ROS - Fold 8  
Accuracy: 0.81 | F1 Score: 0.81

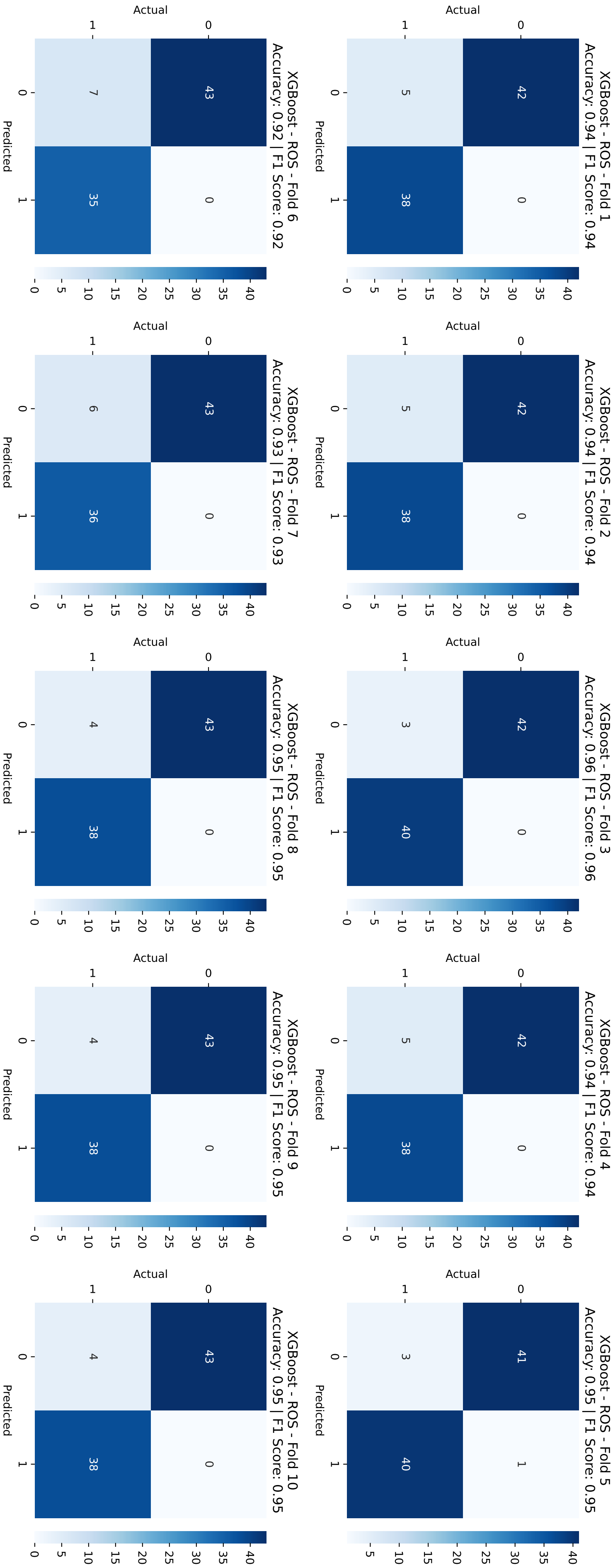


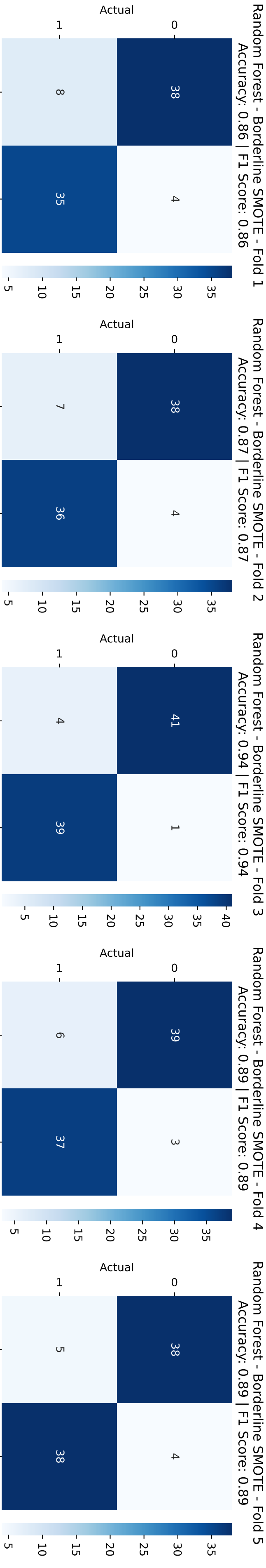
SVM - ROS - Fold 9  
Accuracy: 0.74 | F1 Score: 0.74



SVM - ROS - Fold 10  
Accuracy: 0.78 | F1 Score: 0.77







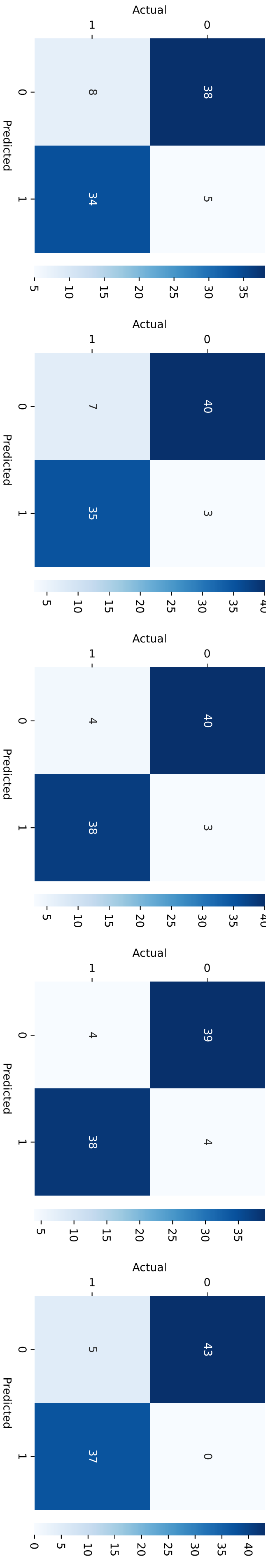
Random Forest - Borderline SMOTE - Fold 6  
Accuracy: 0.85 | F1 Score: 0.85

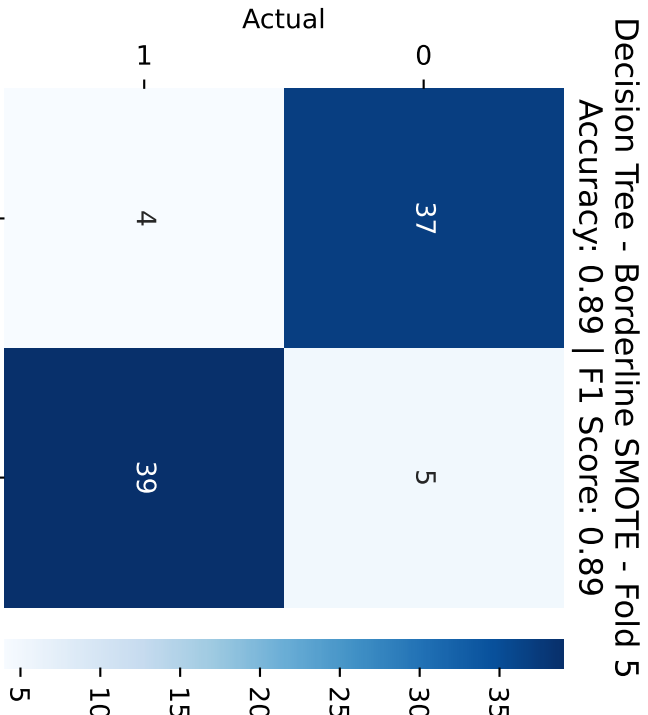
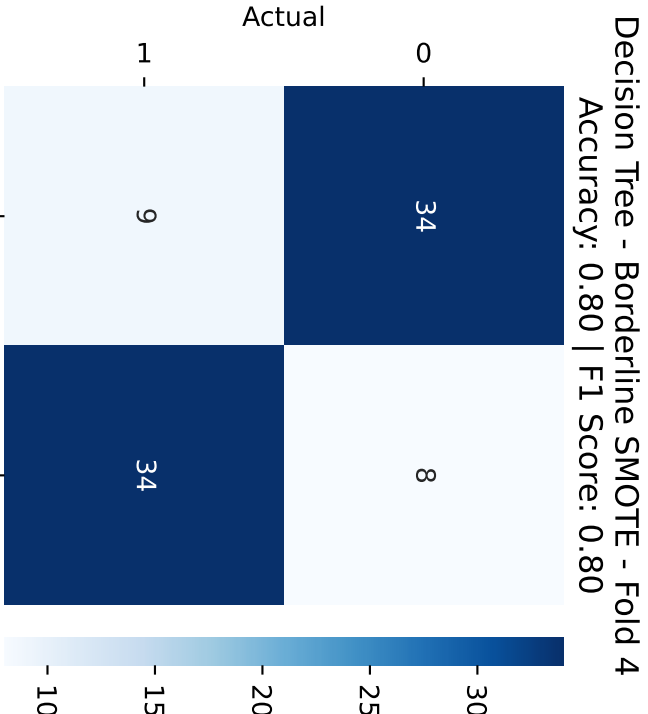
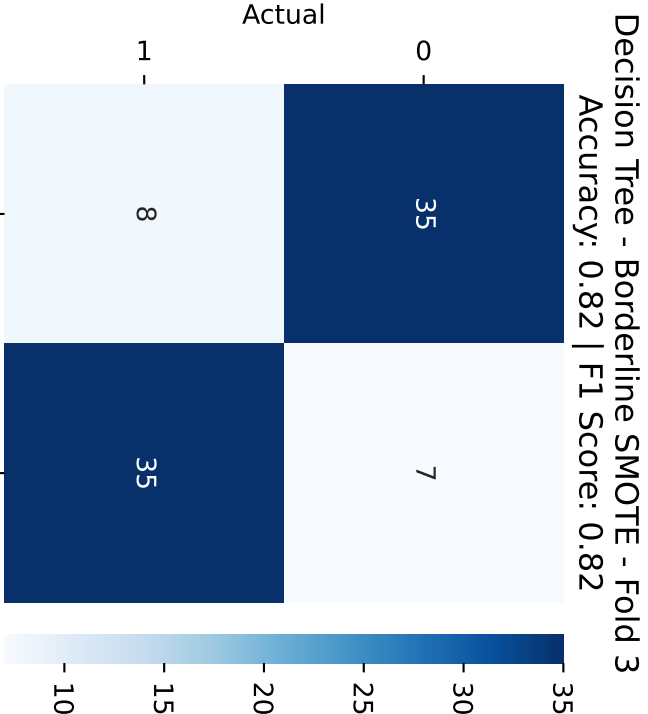
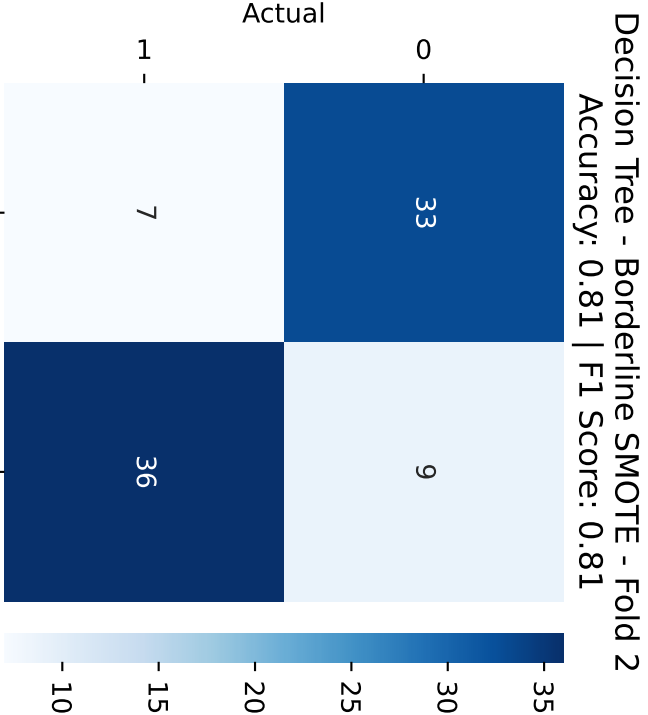
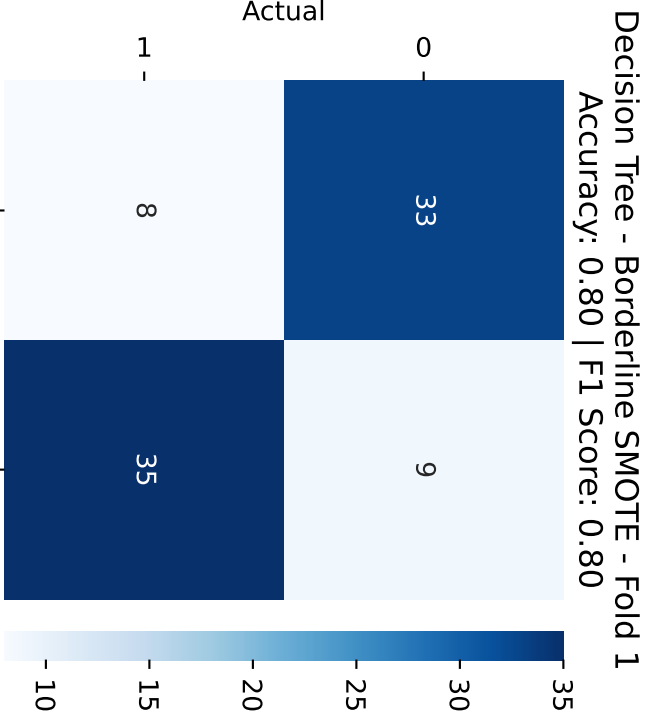
Random Forest - Borderline SMOTE - Fold 7  
Accuracy: 0.88 | F1 Score: 0.88

Random Forest - Borderline SMOTE - Fold 8  
Accuracy: 0.92 | F1 Score: 0.92

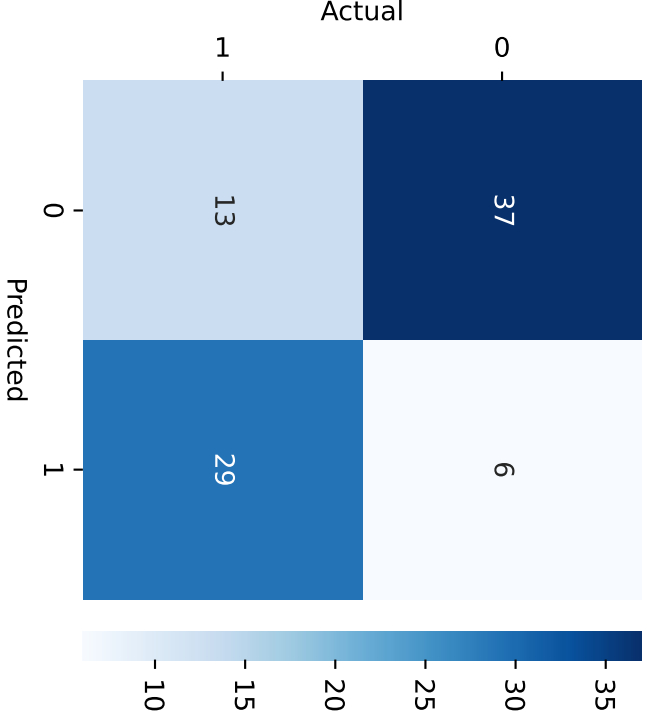
Random Forest - Borderline SMOTE - Fold 9  
Accuracy: 0.91 | F1 Score: 0.91

Random Forest - Borderline SMOTE - Fold 10  
Accuracy: 0.94 | F1 Score: 0.94

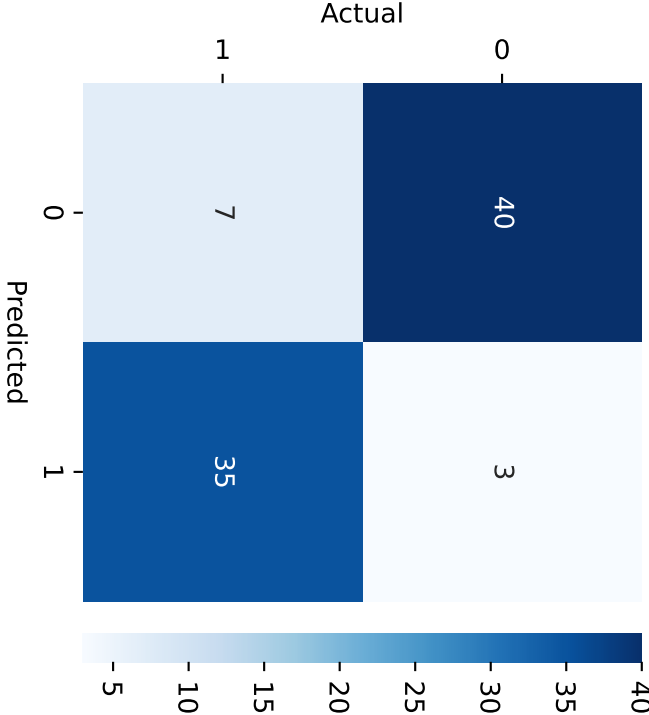




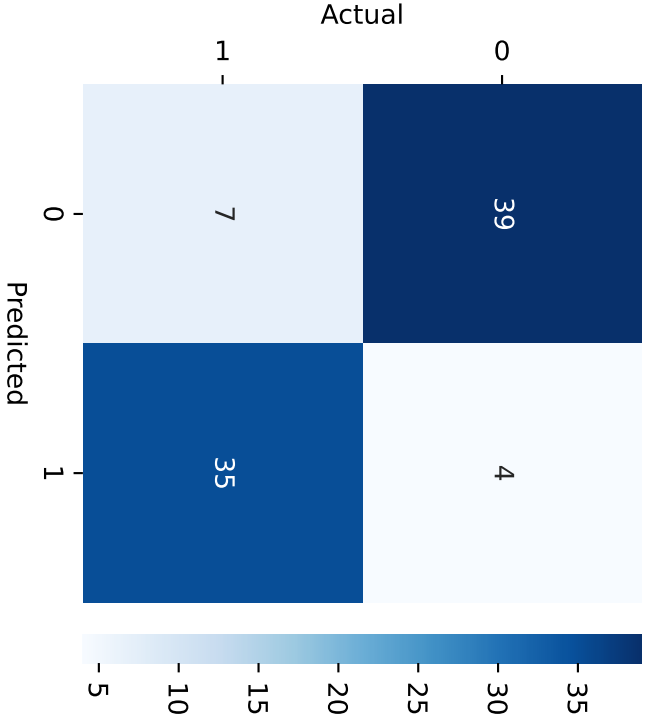
Decision Tree - Borderline SMOTE - Fold 6  
Accuracy: 0.78 | F1 Score: 0.77



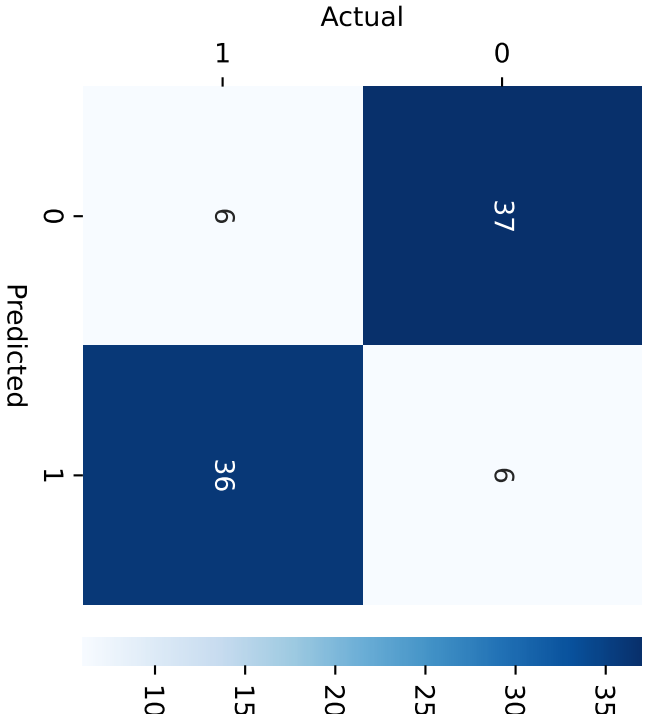
Decision Tree - Borderline SMOTE - Fold 7  
Accuracy: 0.88 | F1 Score: 0.88



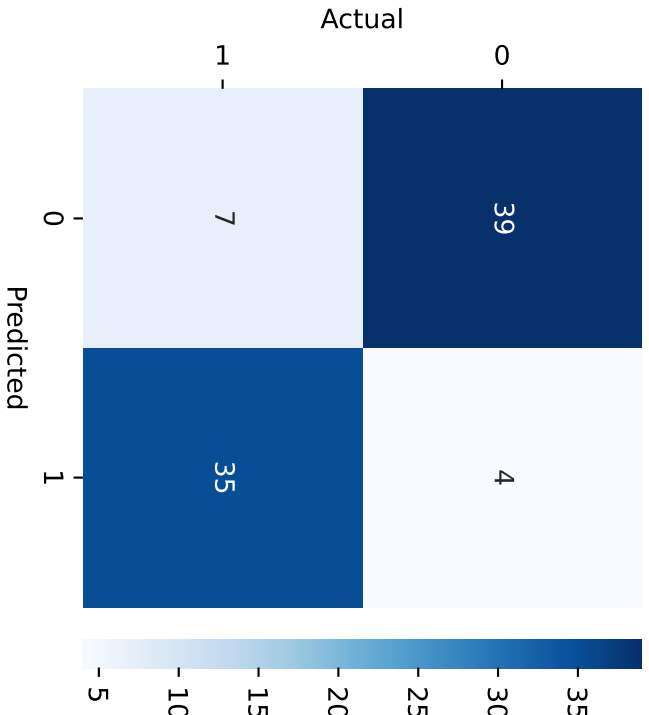
Decision Tree - Borderline SMOTE - Fold 8  
Accuracy: 0.87 | F1 Score: 0.87

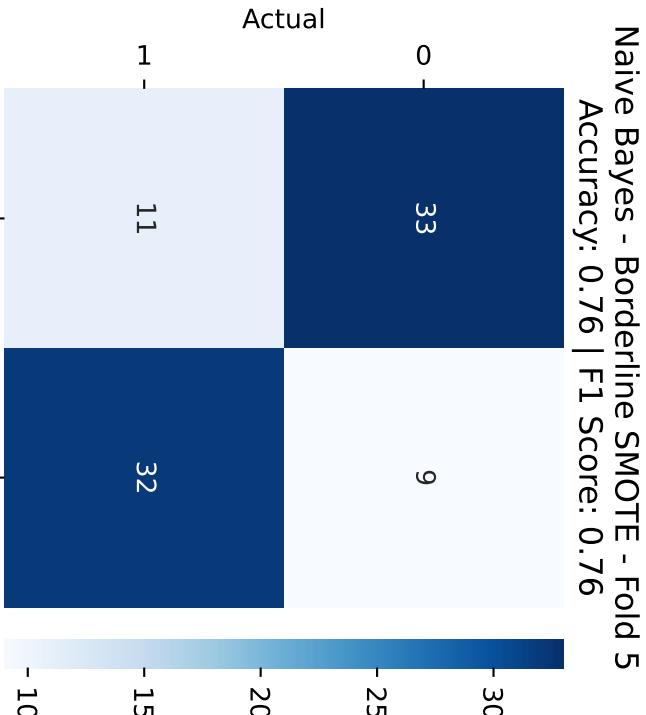
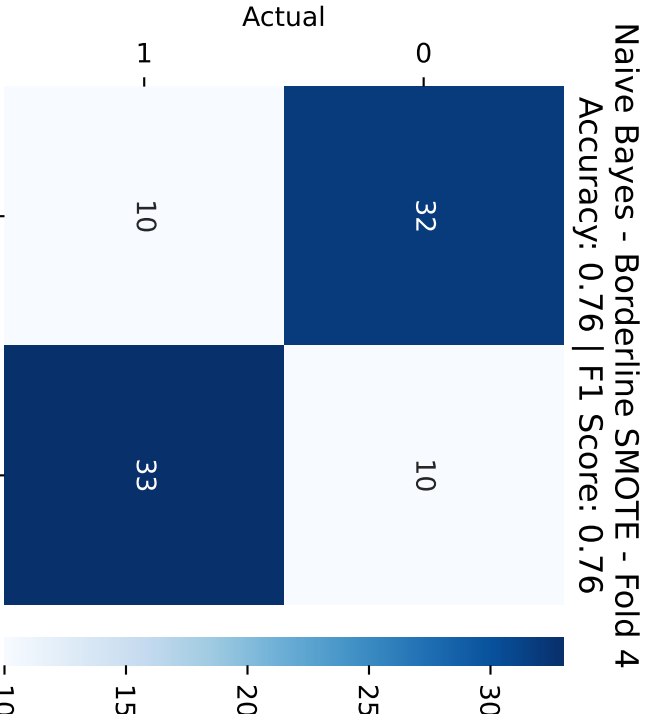
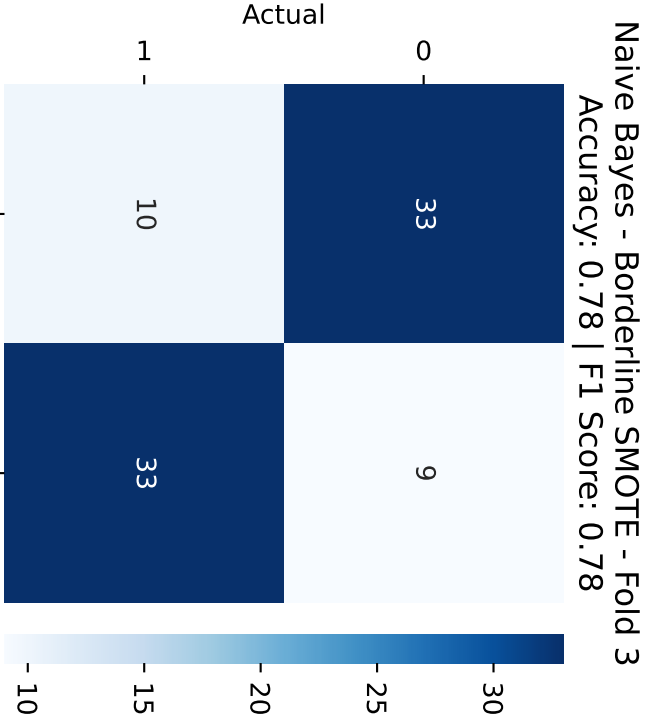
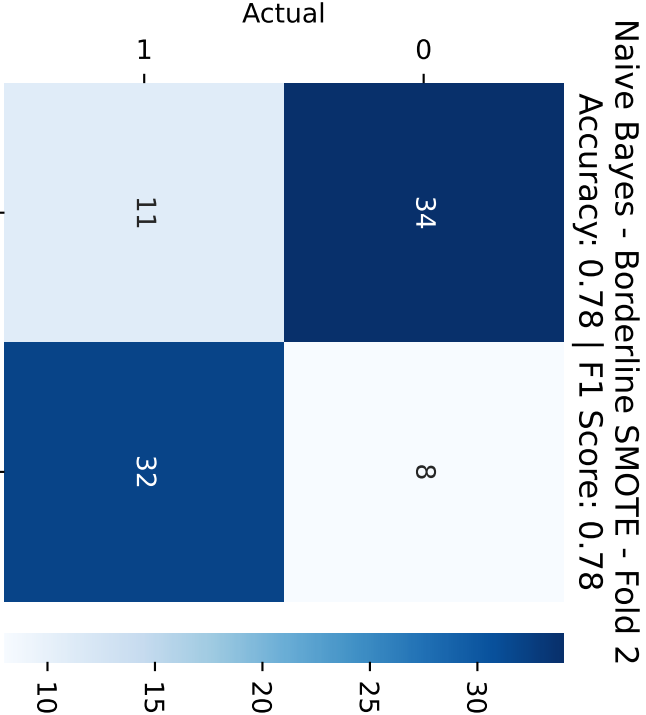
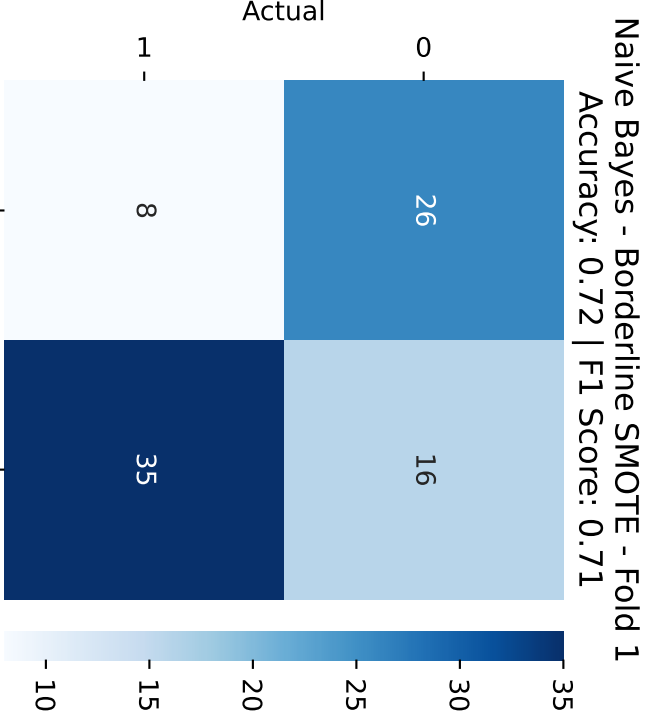


Decision Tree - Borderline SMOTE - Fold 9  
Accuracy: 0.86 | F1 Score: 0.86

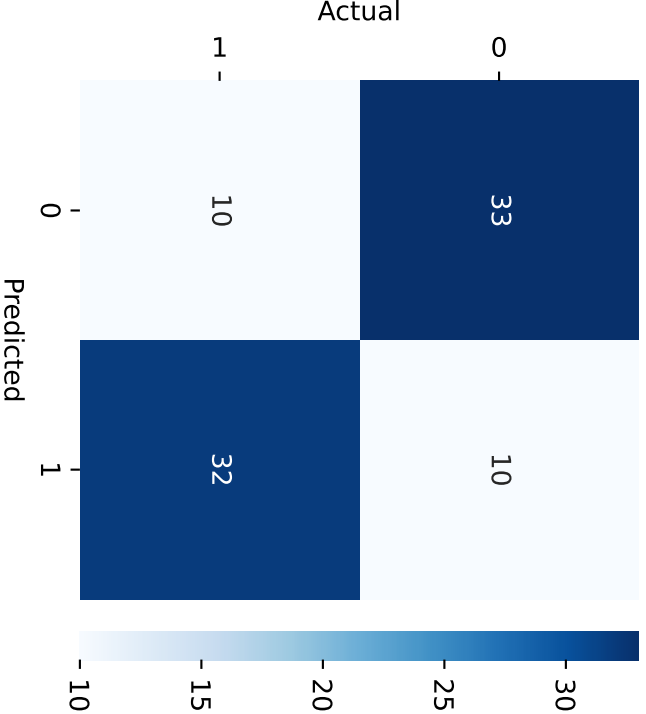


Decision Tree - Borderline SMOTE - Fold 10  
Accuracy: 0.87 | F1 Score: 0.87

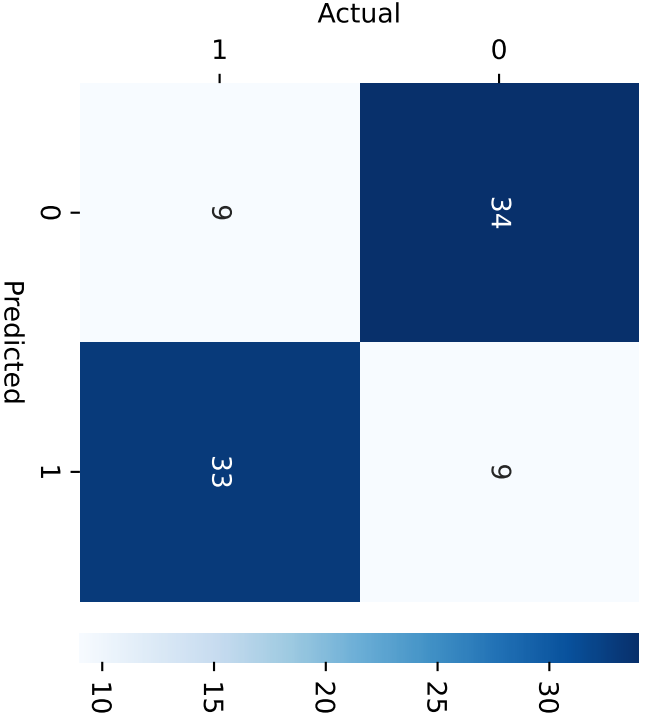




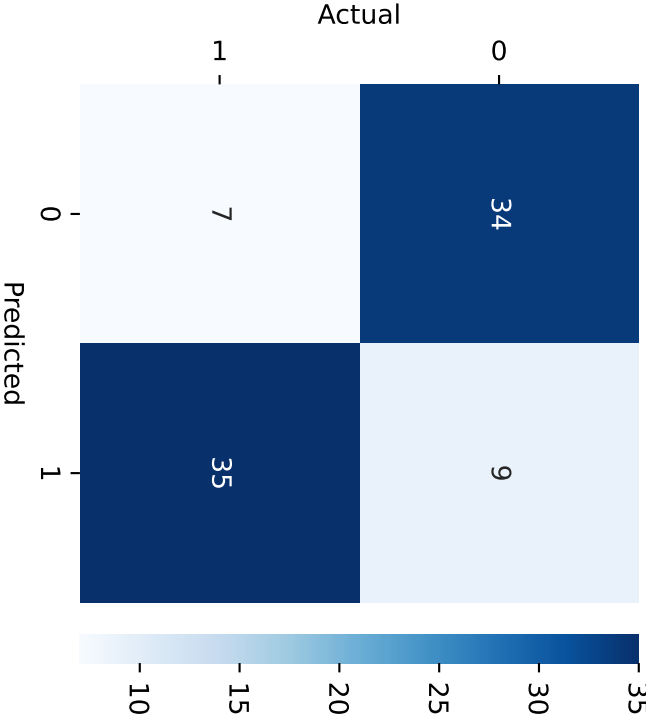
Naive Bayes - Borderline SMOTE - Fold 6  
Accuracy: 0.76 | F1 Score: 0.76



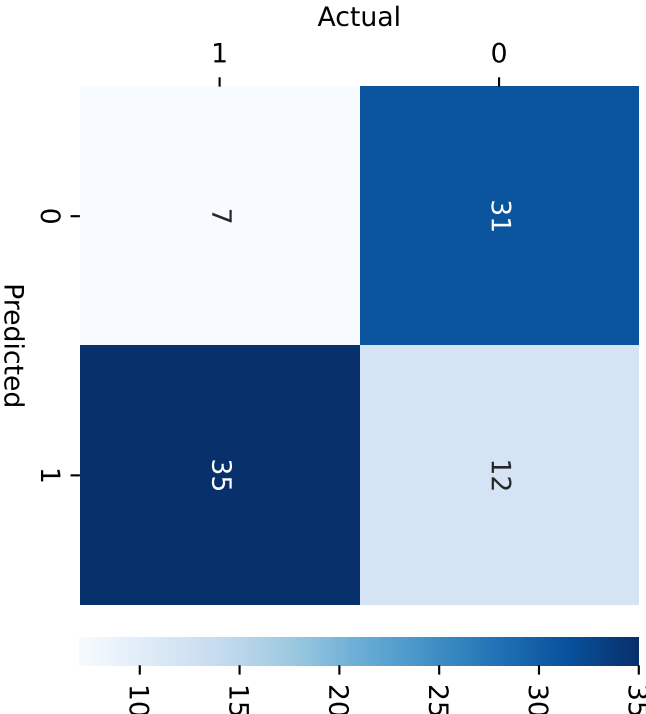
Naive Bayes - Borderline SMOTE - Fold 7  
Accuracy: 0.79 | F1 Score: 0.79



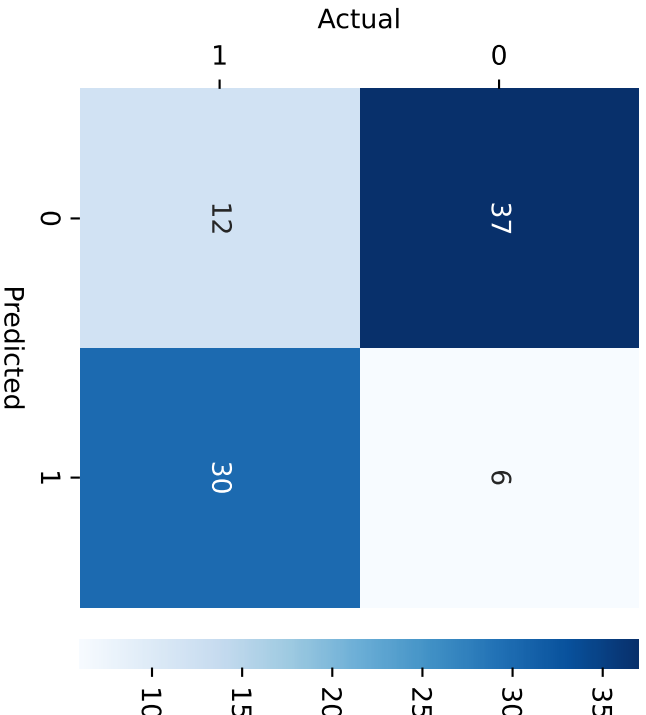
Naive Bayes - Borderline SMOTE - Fold 8  
Accuracy: 0.81 | F1 Score: 0.81

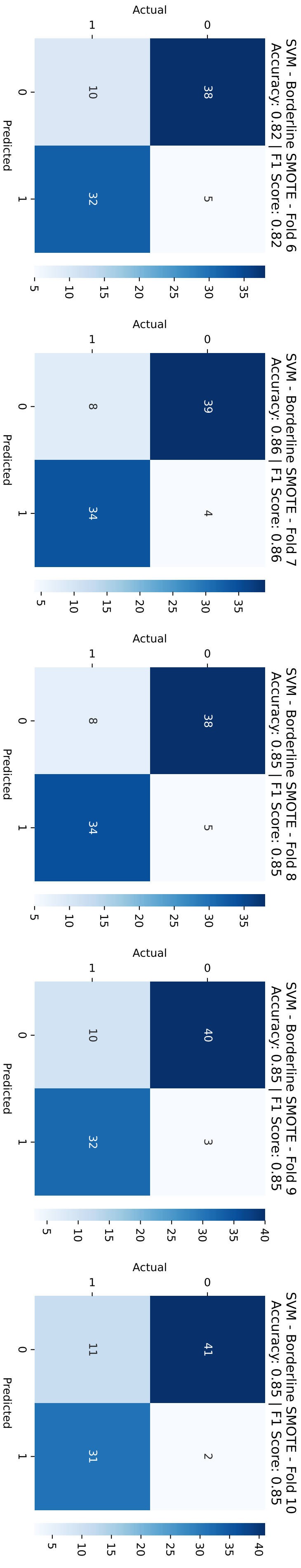
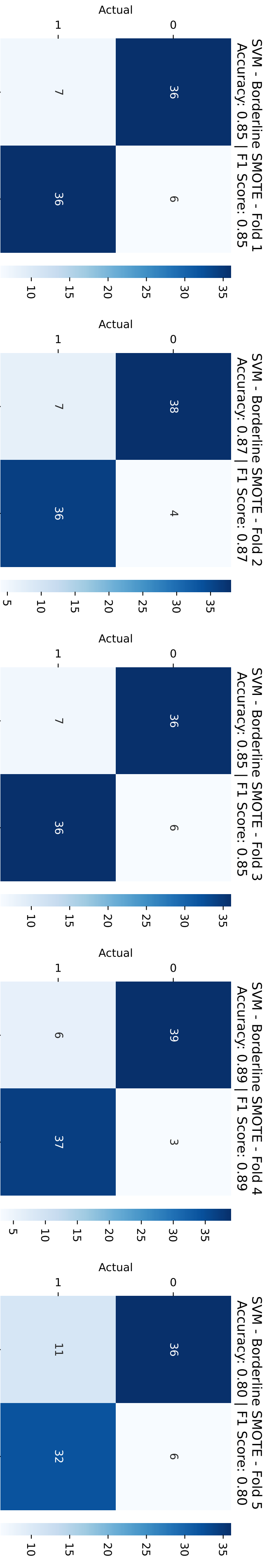


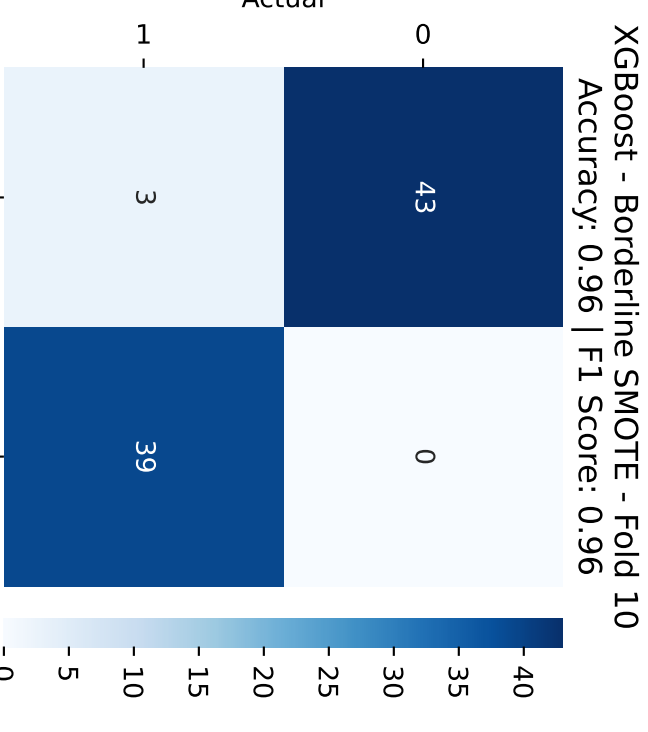
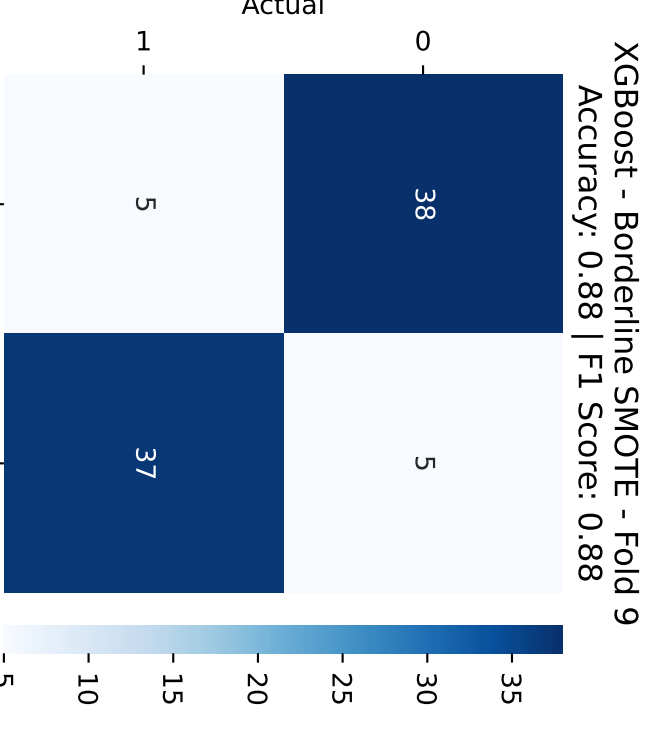
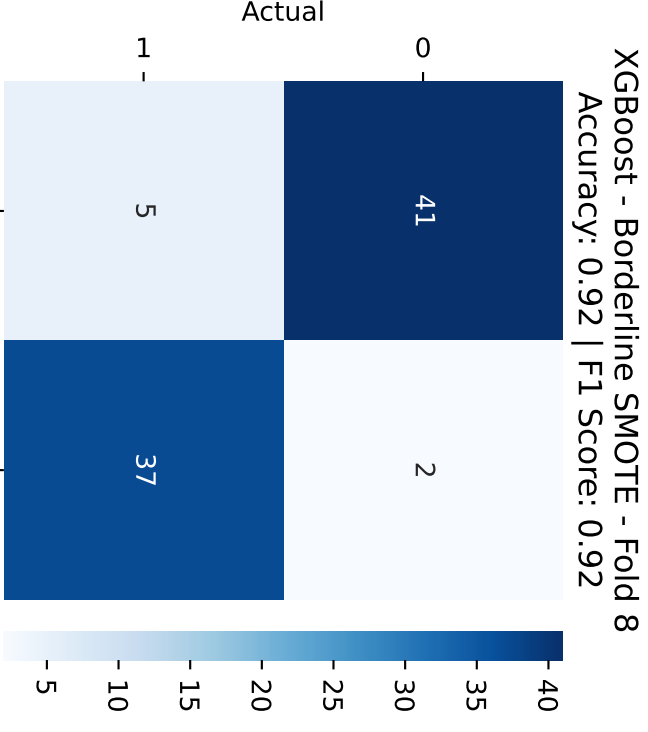
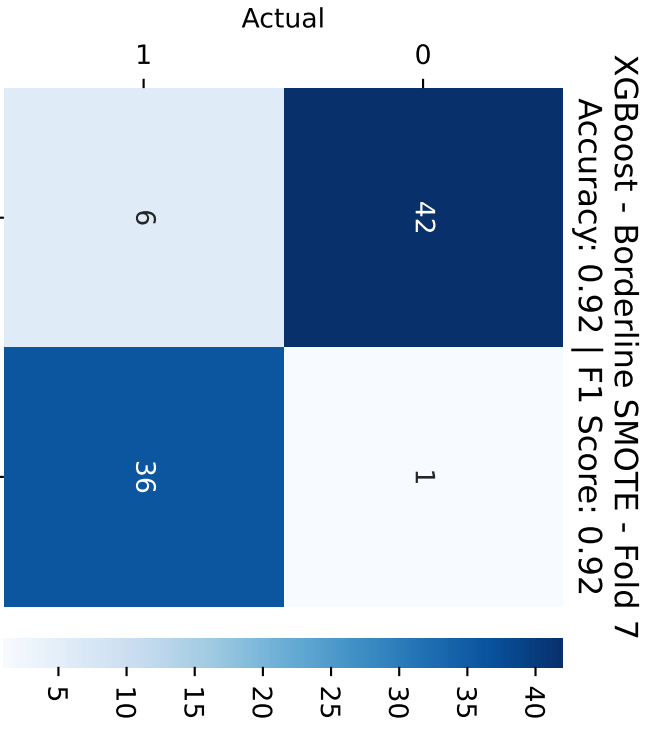
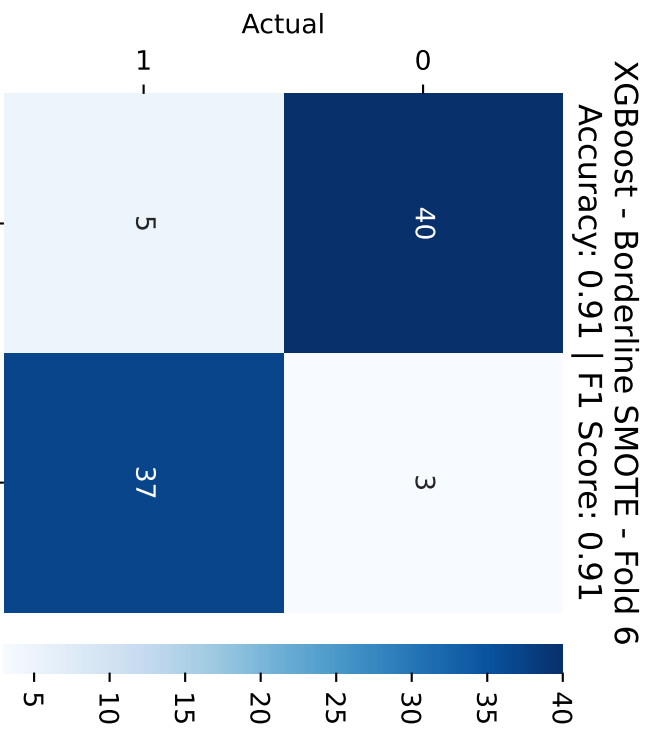
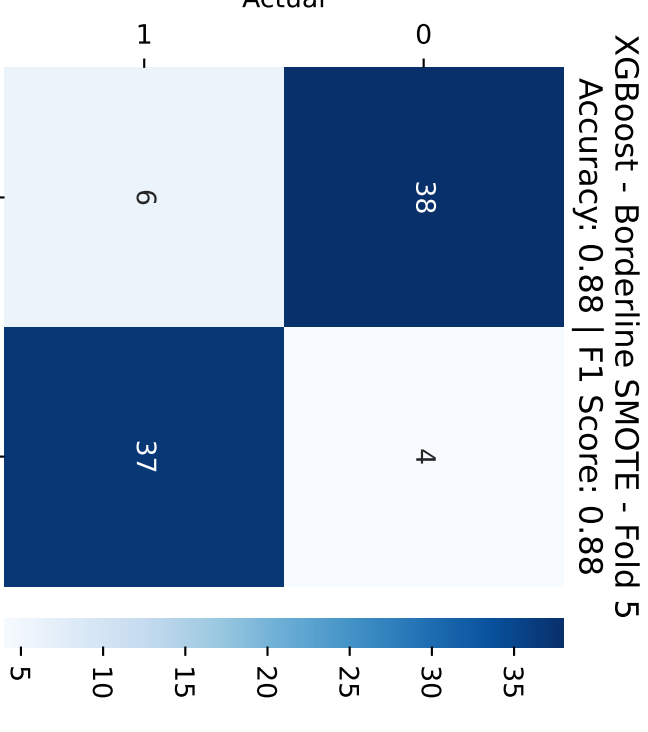
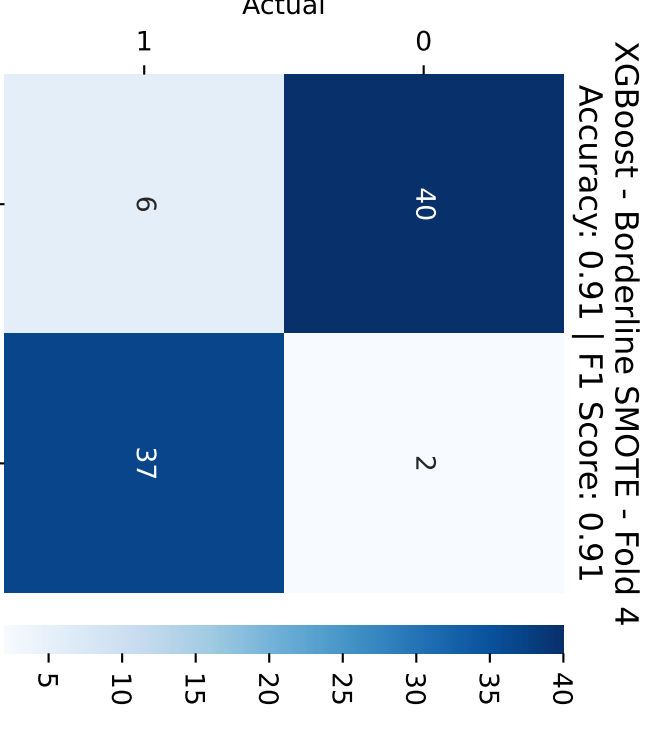
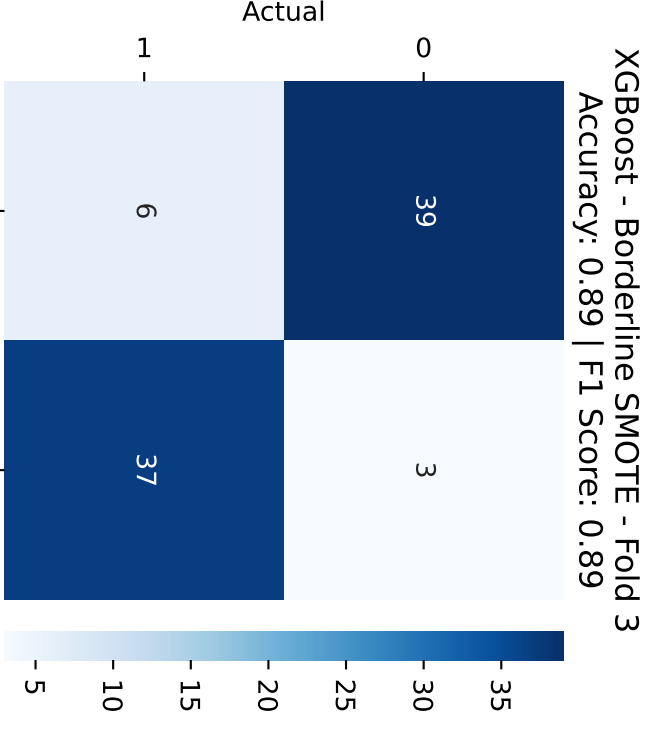
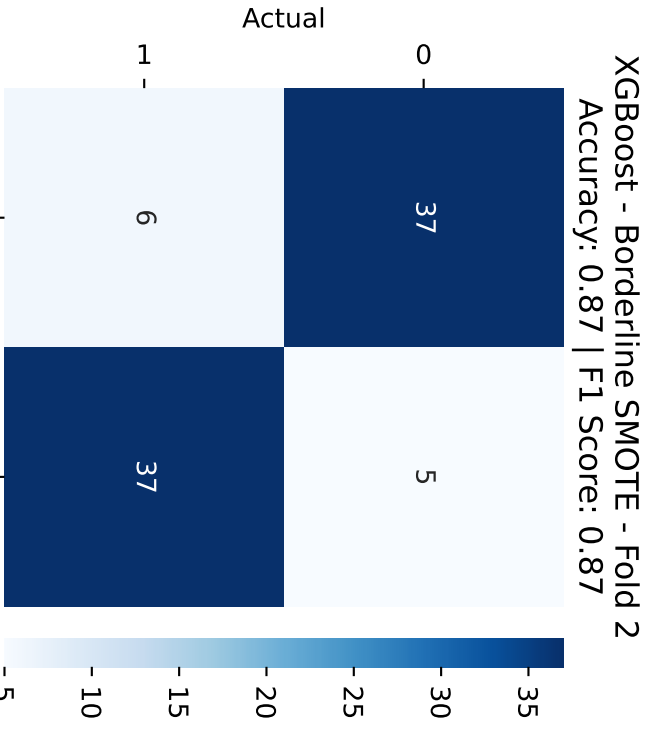
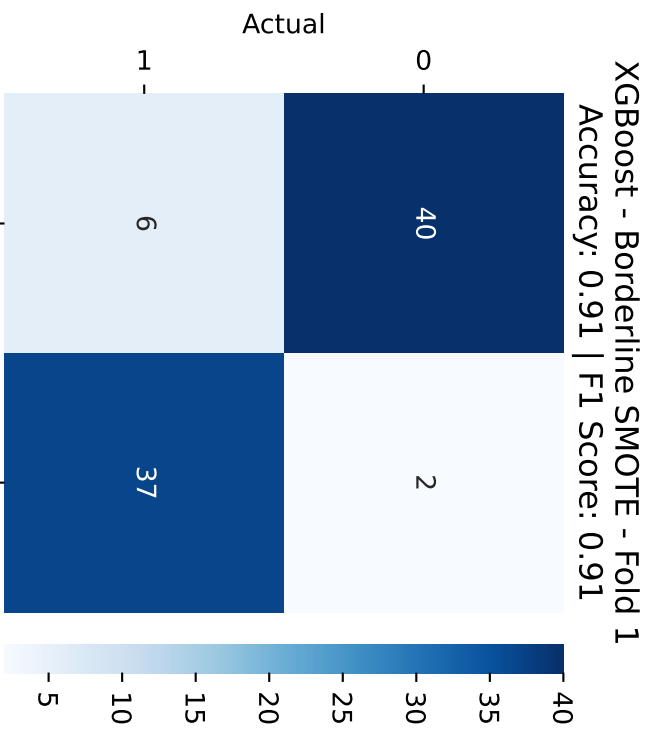
Naive Bayes - Borderline SMOTE - Fold 9  
Accuracy: 0.78 | F1 Score: 0.78

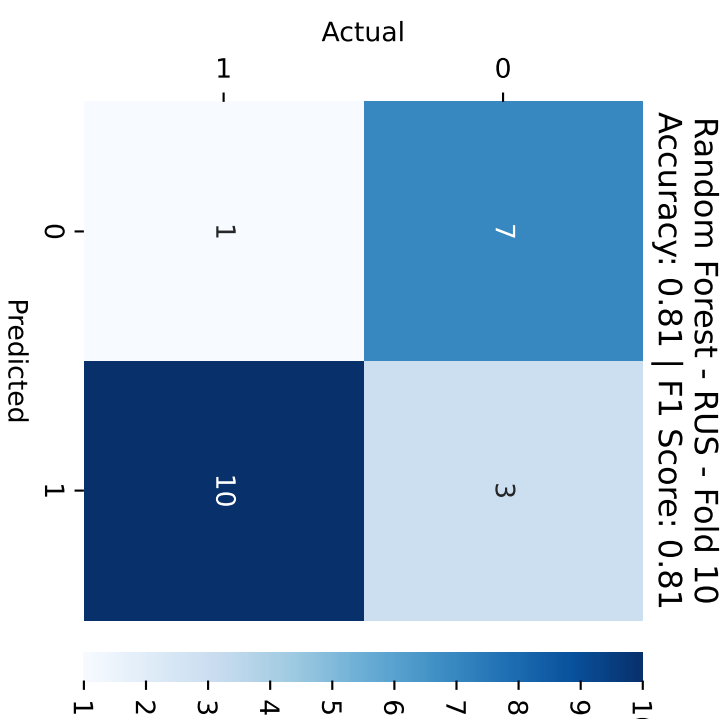
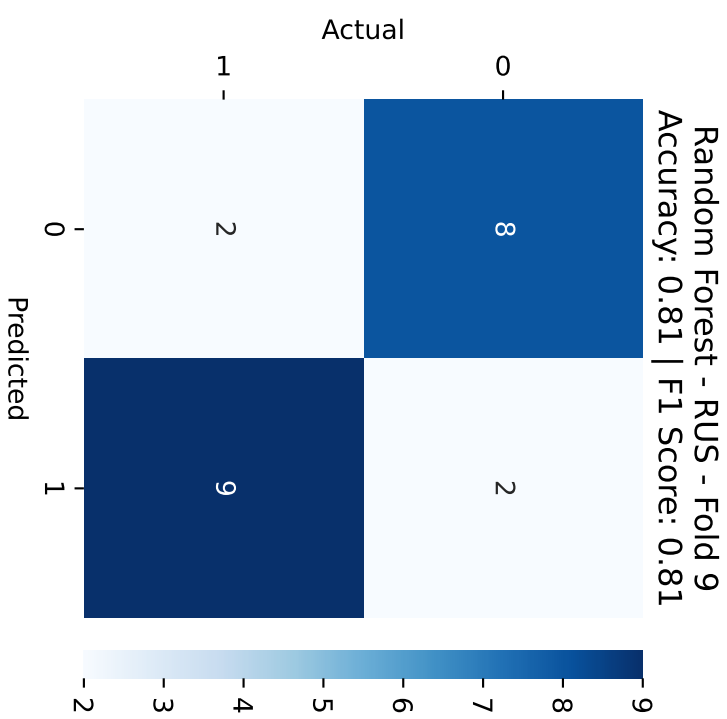
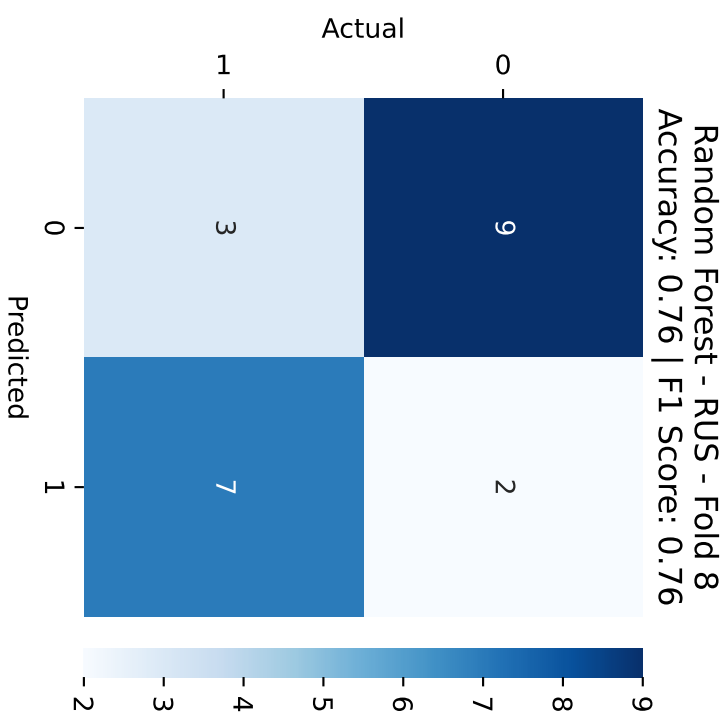
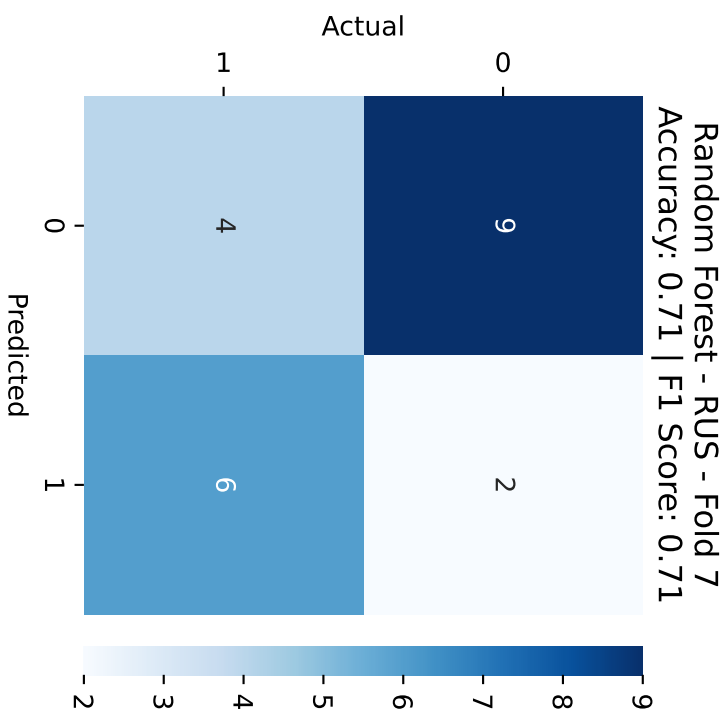
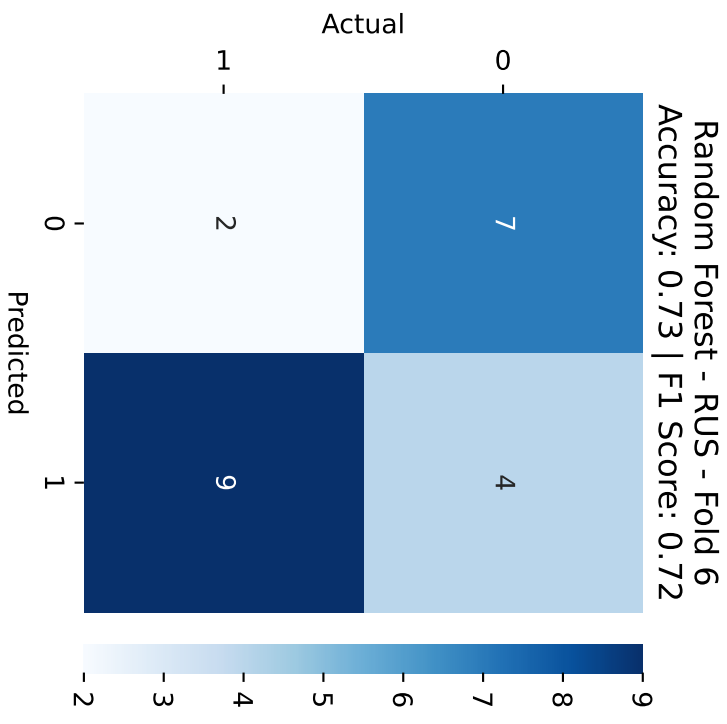
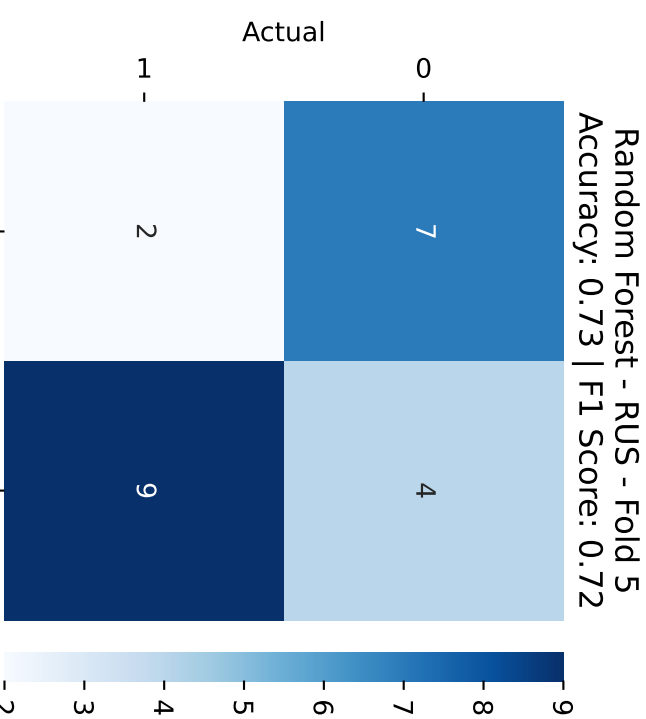
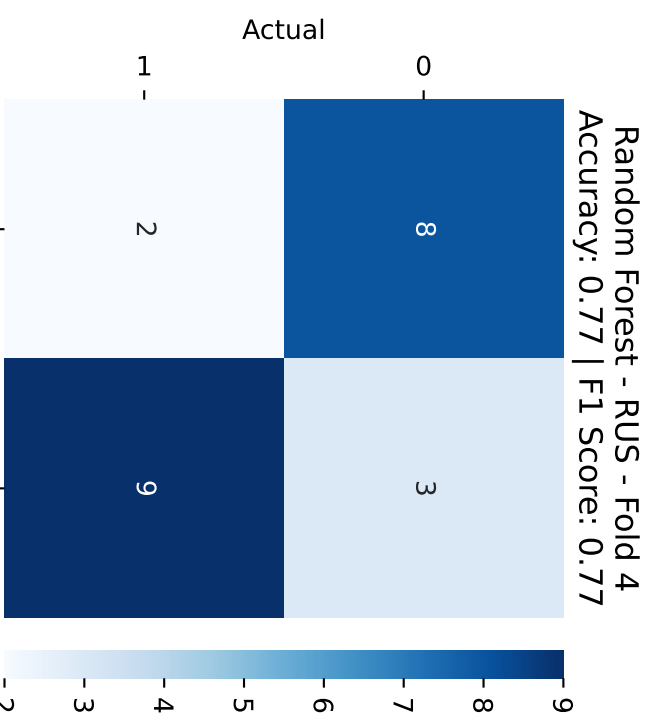
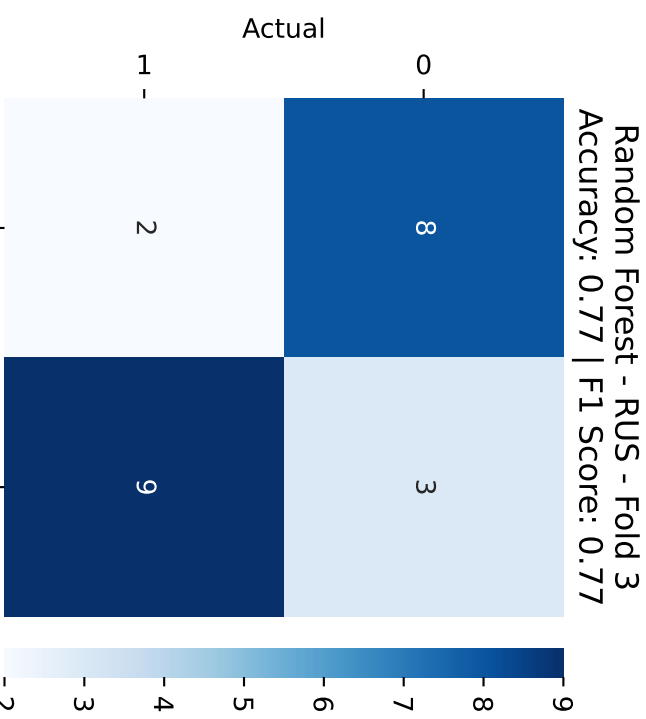
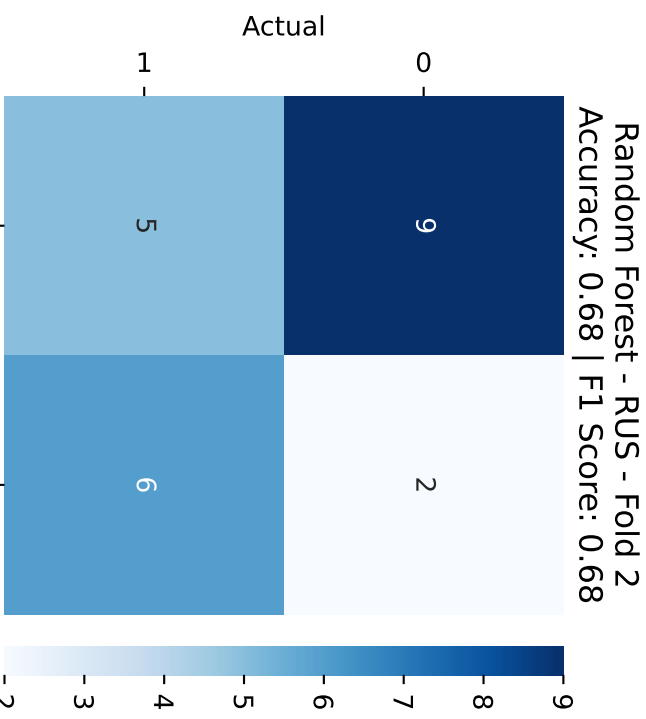
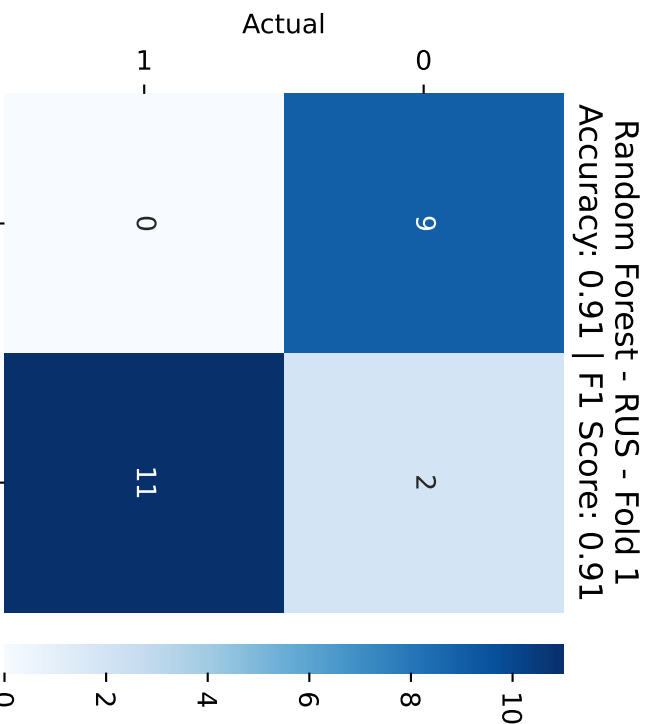


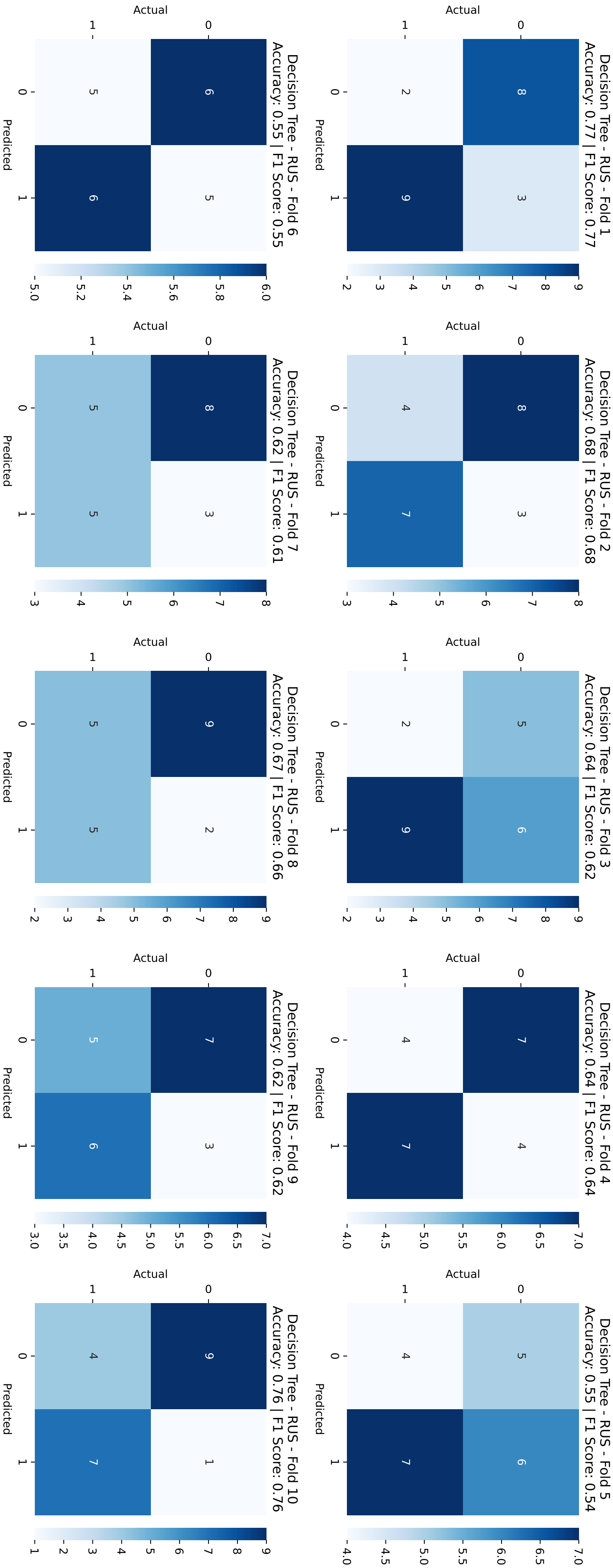
Naive Bayes - Borderline SMOTE - Fold 10  
Accuracy: 0.79 | F1 Score: 0.79

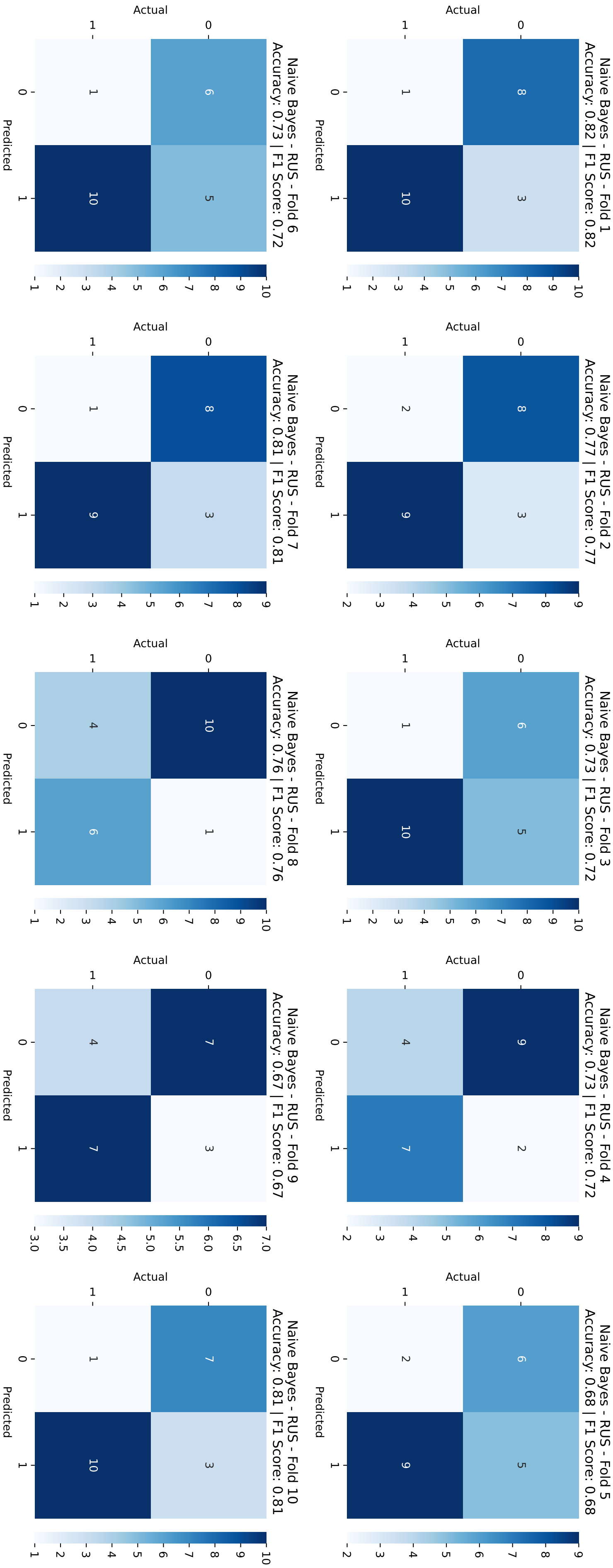


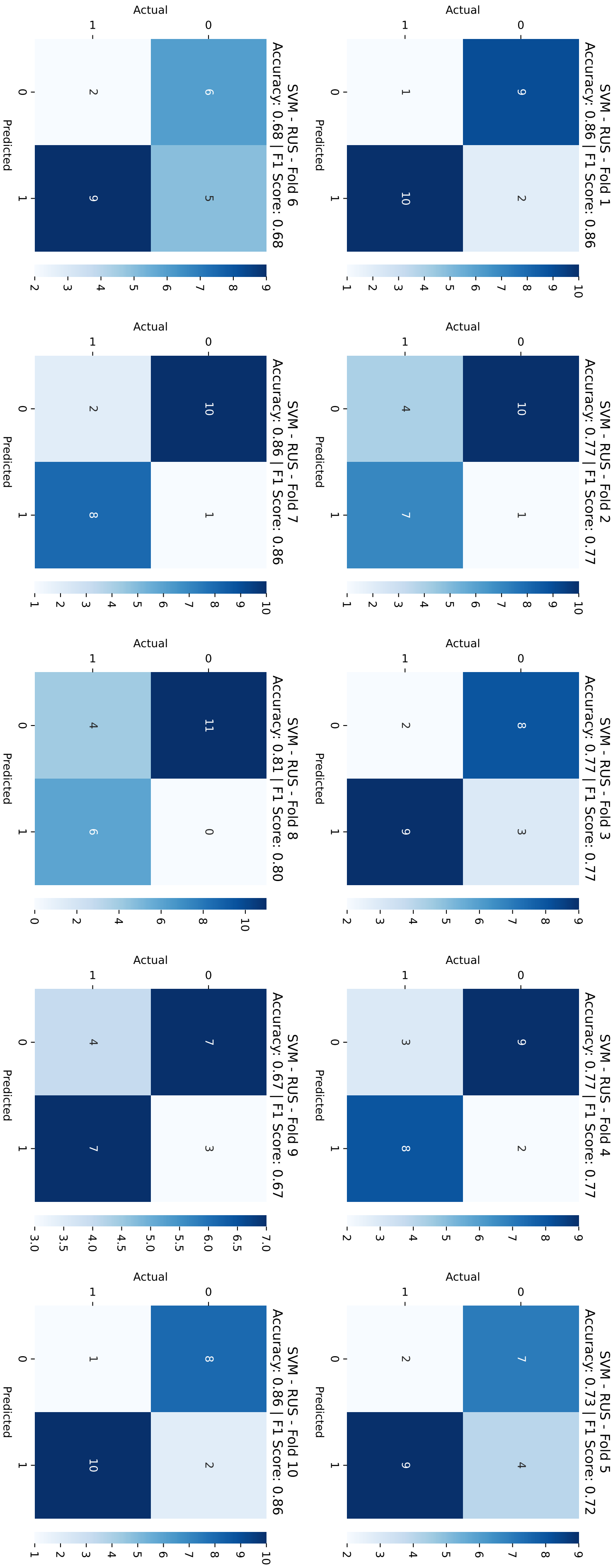


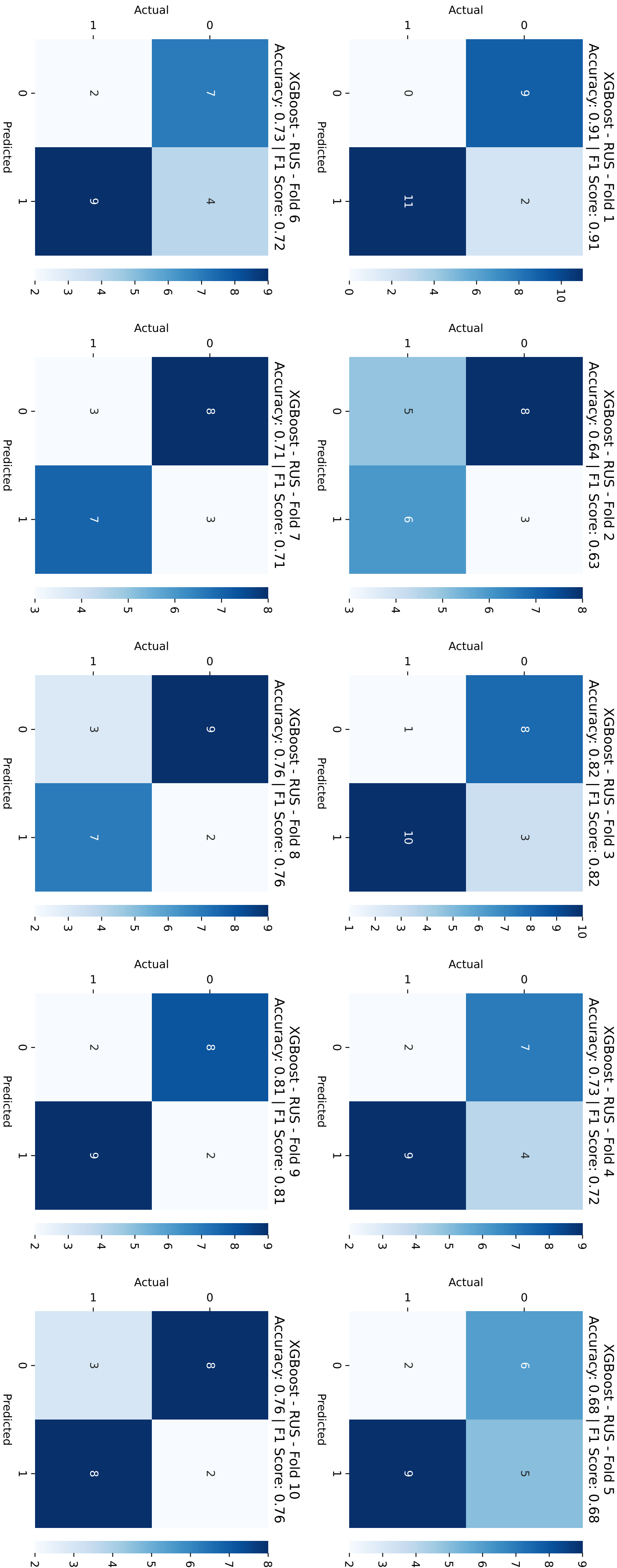


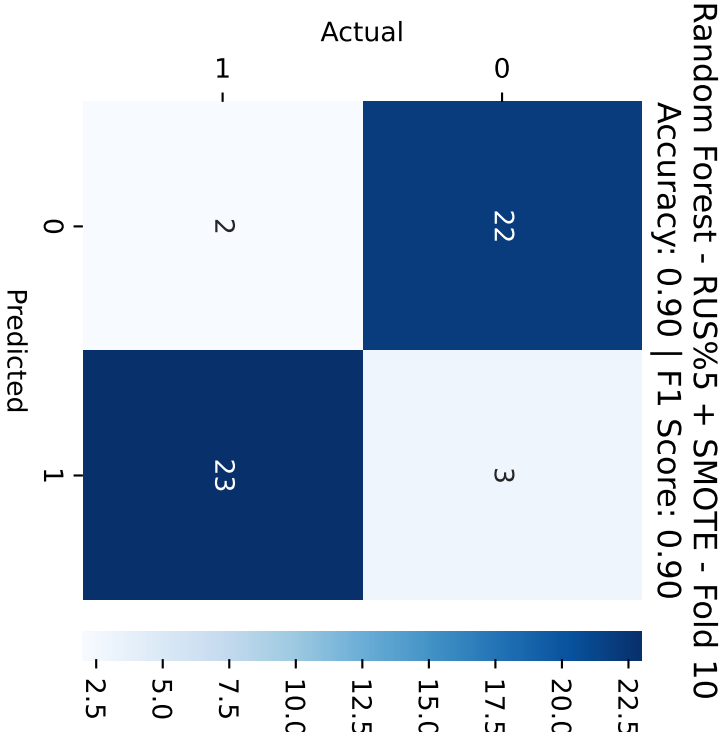
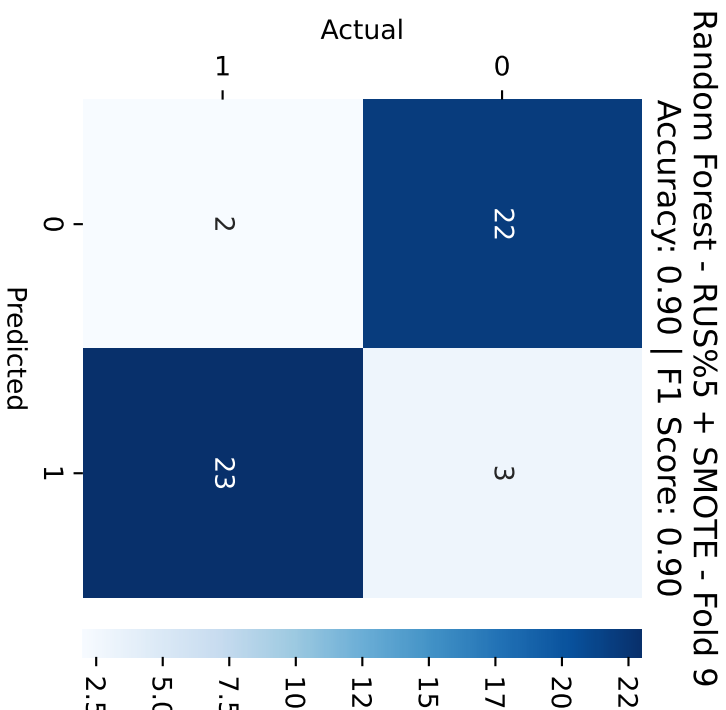
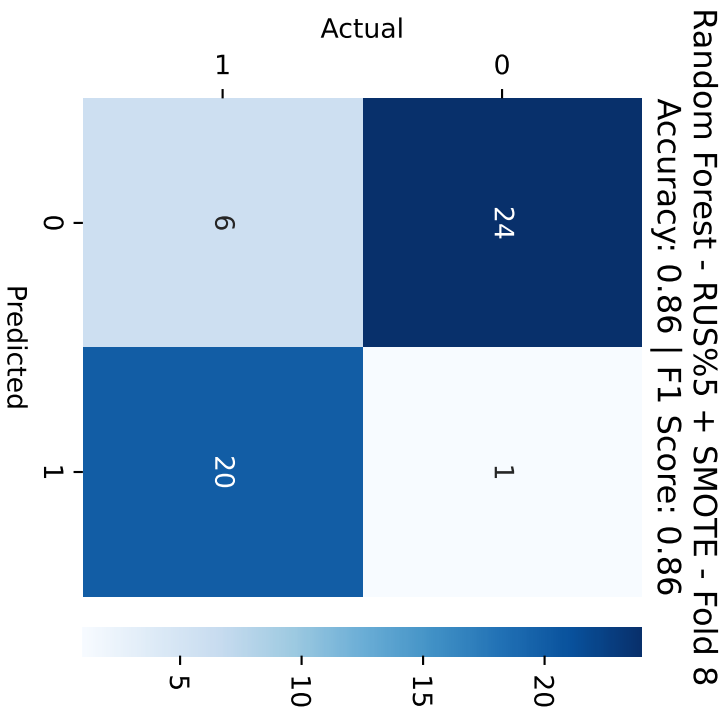
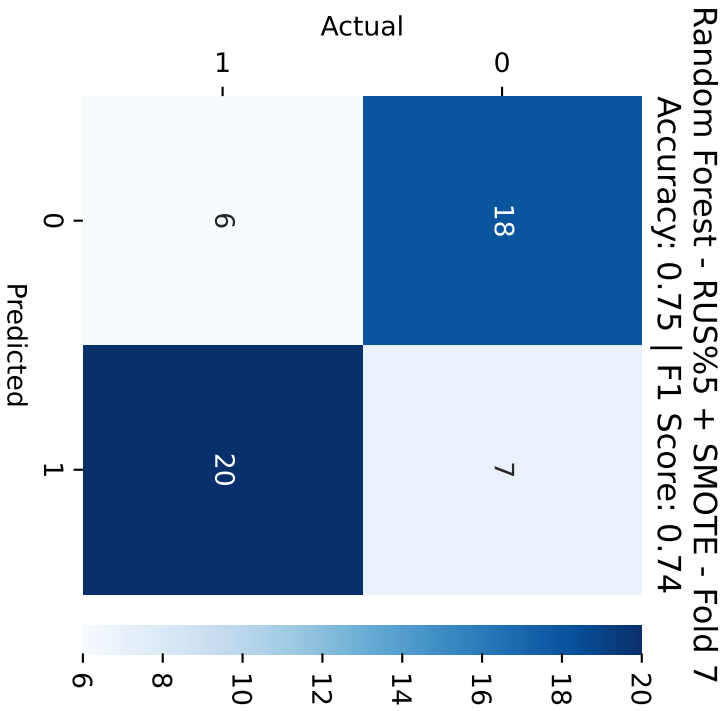
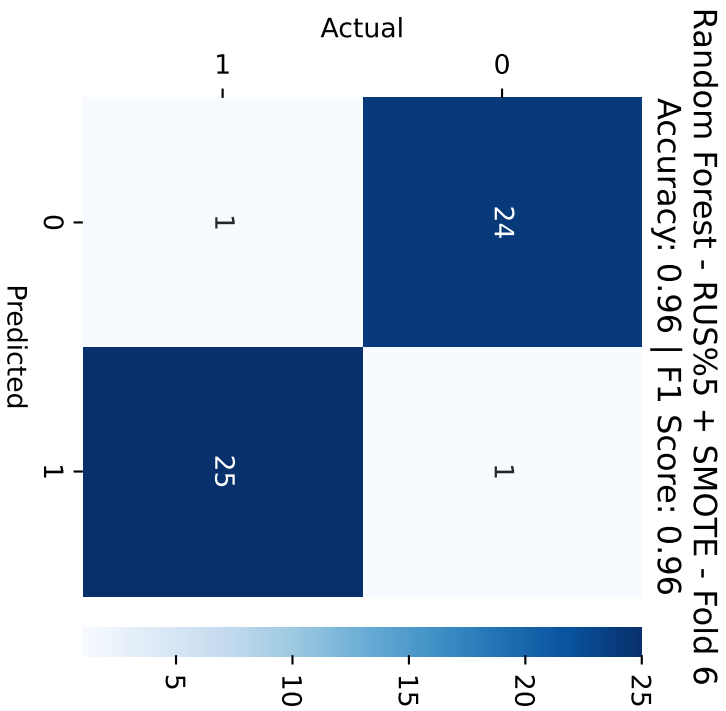
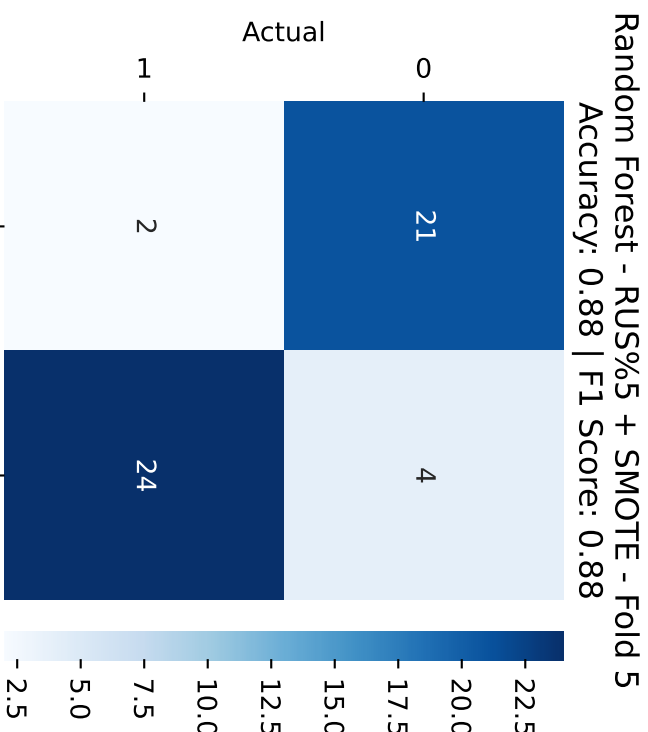
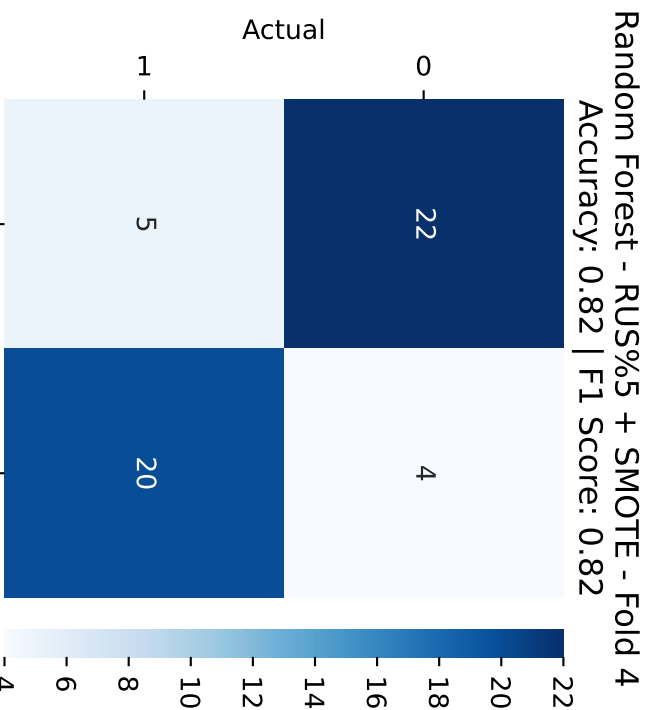
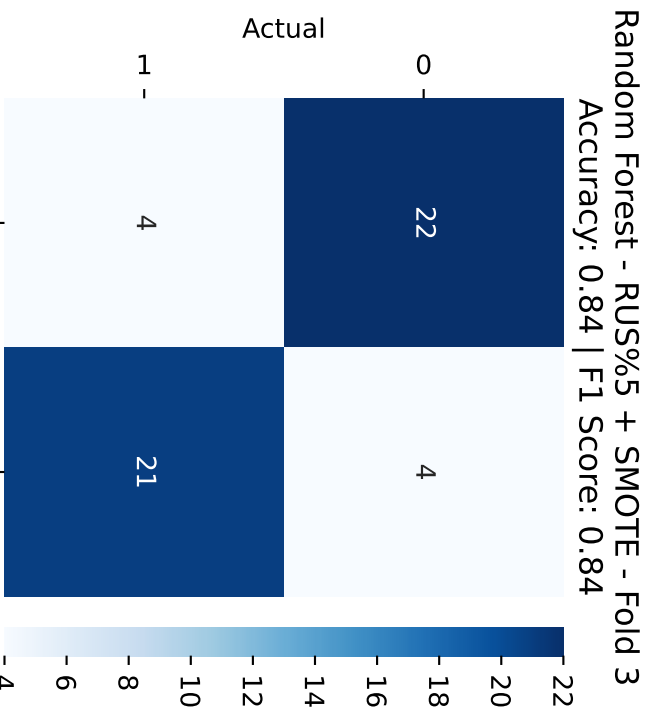
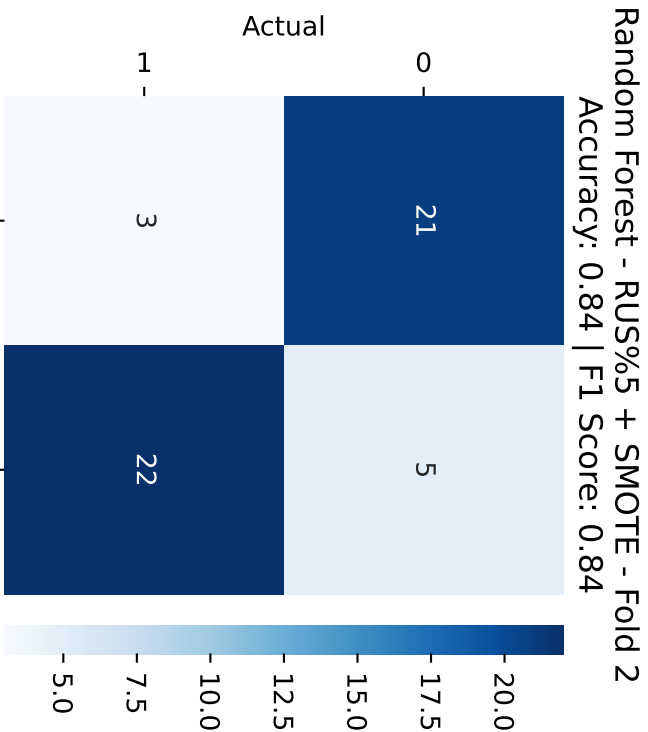
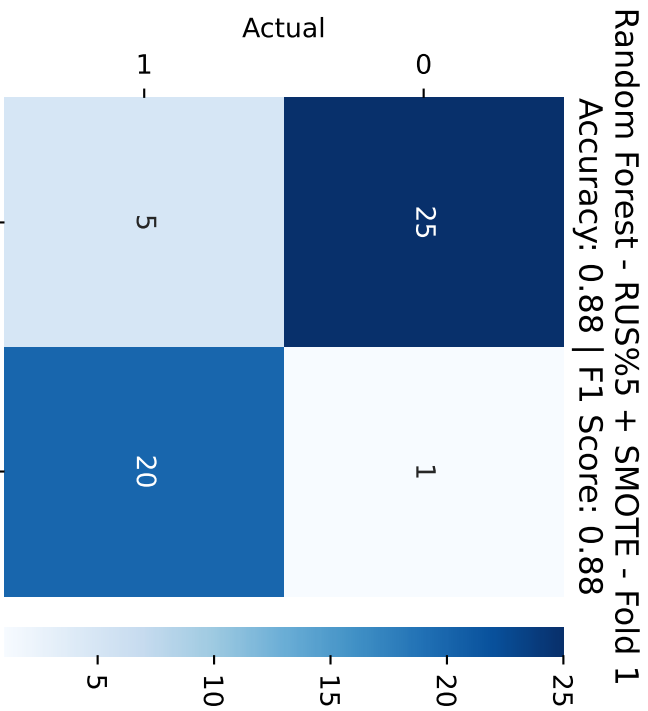


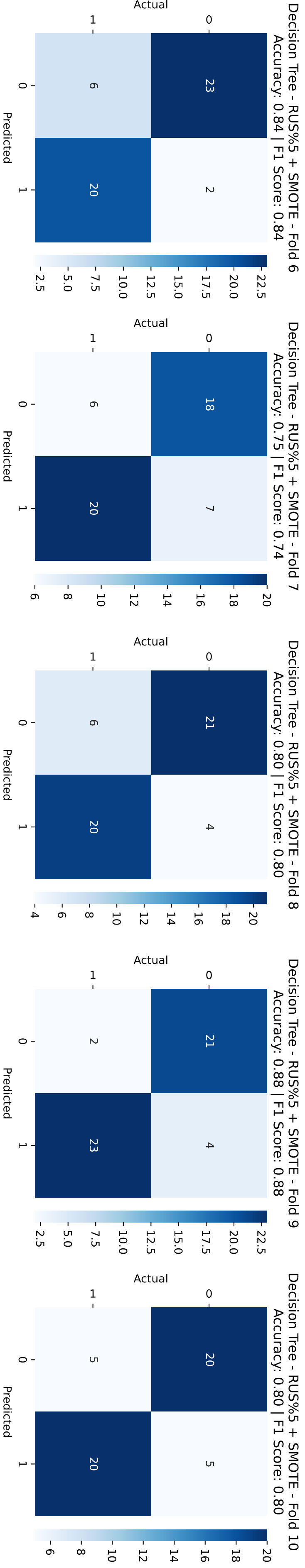
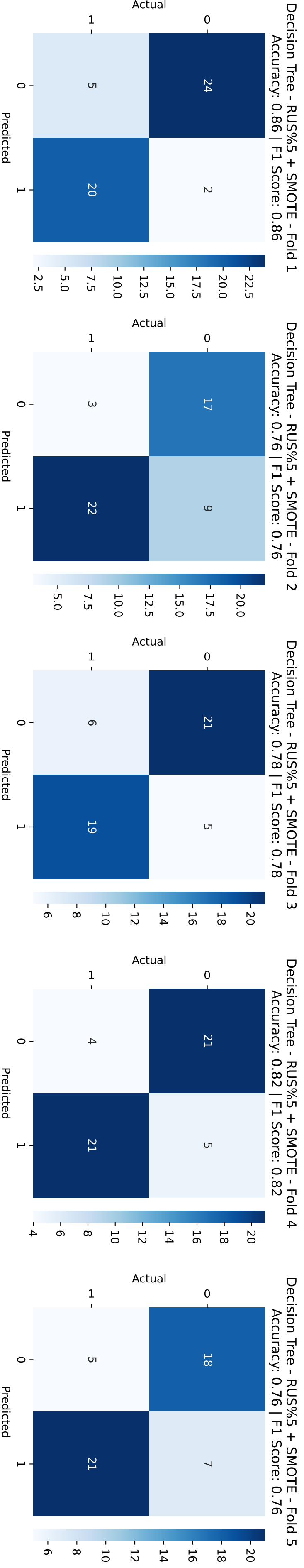


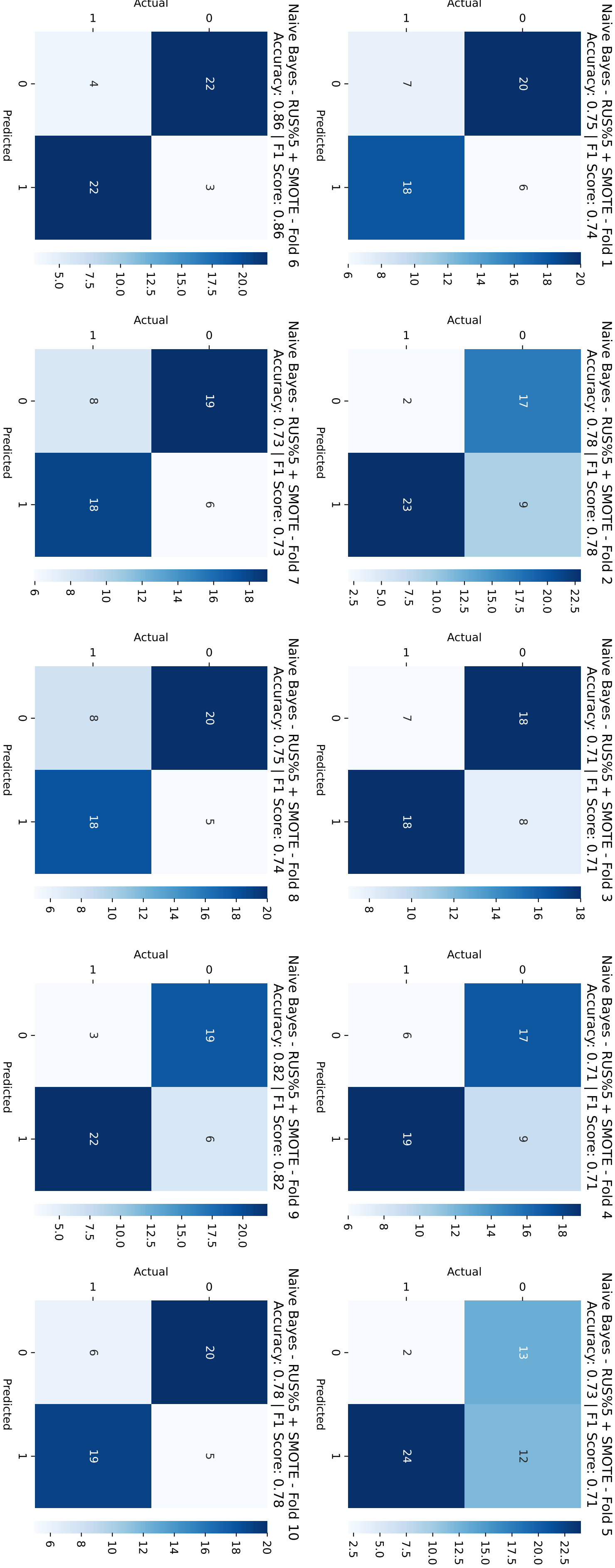


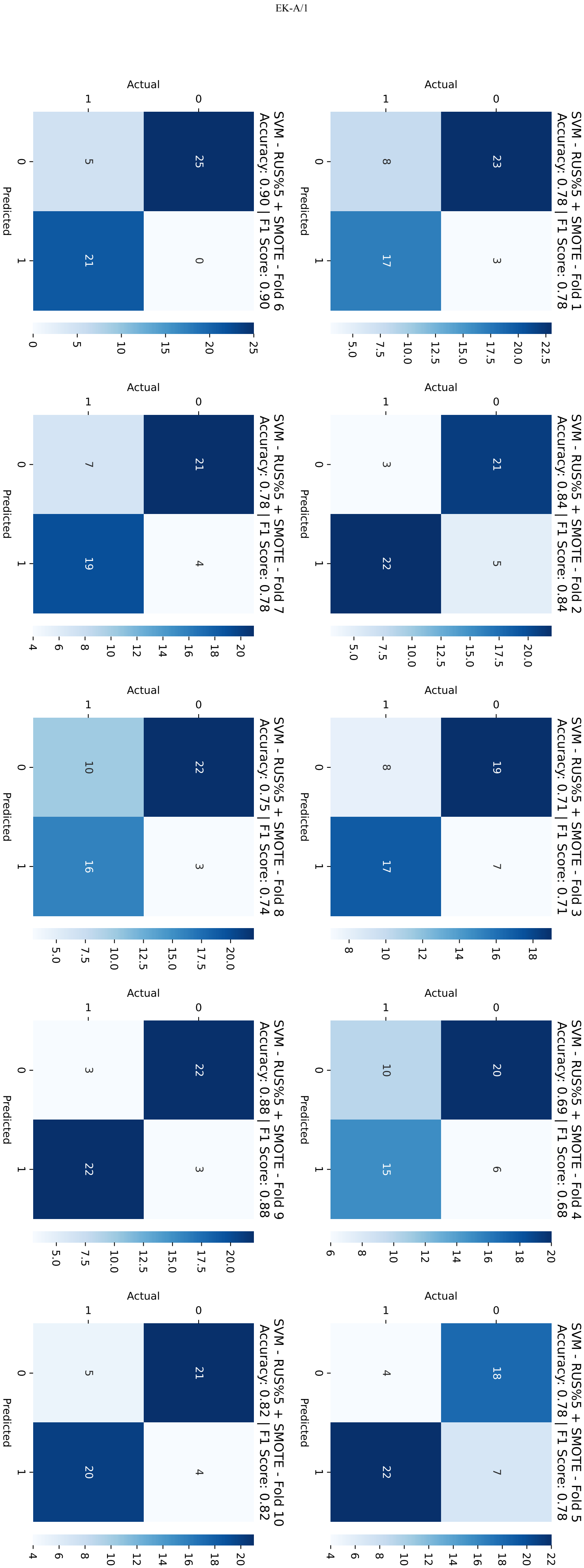


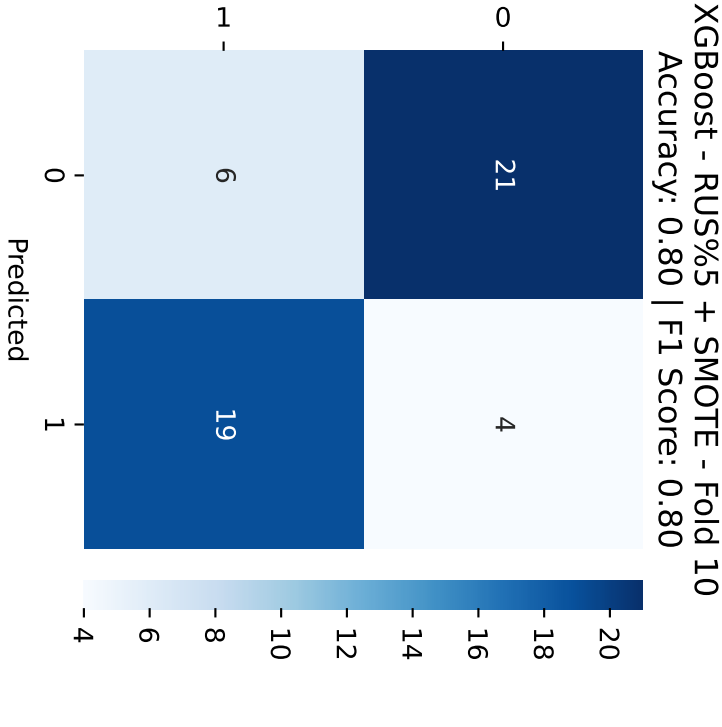
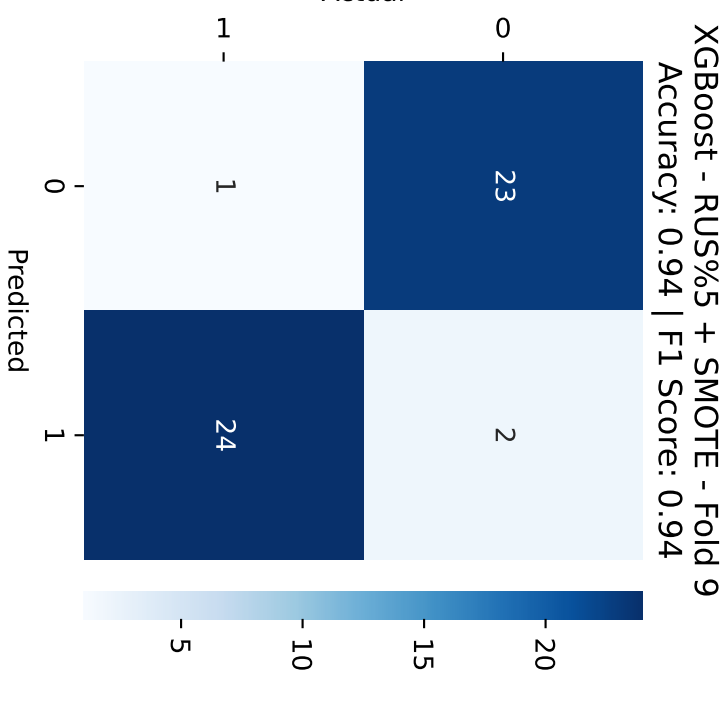
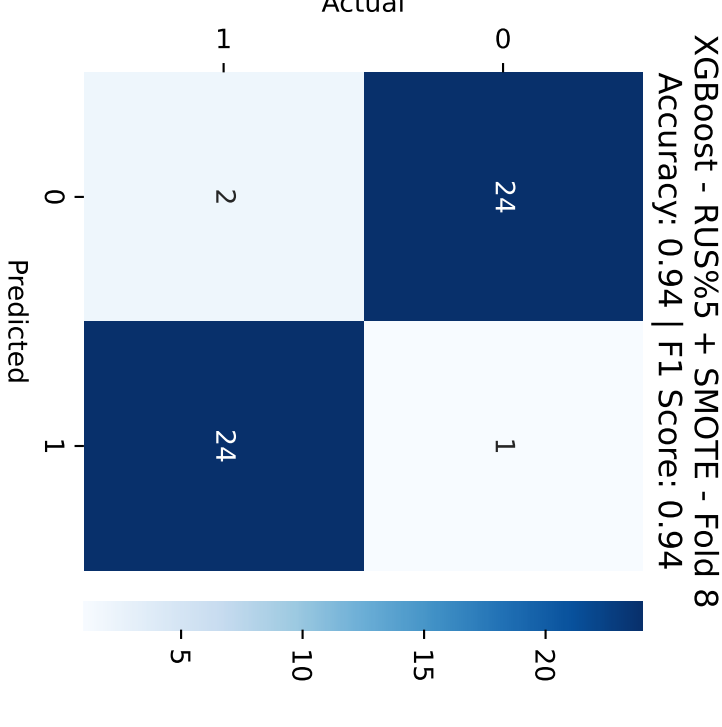
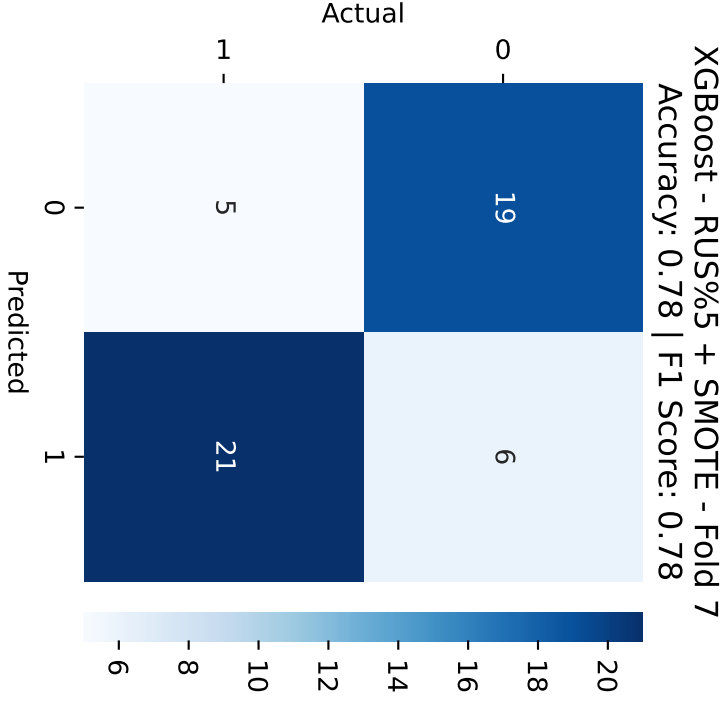
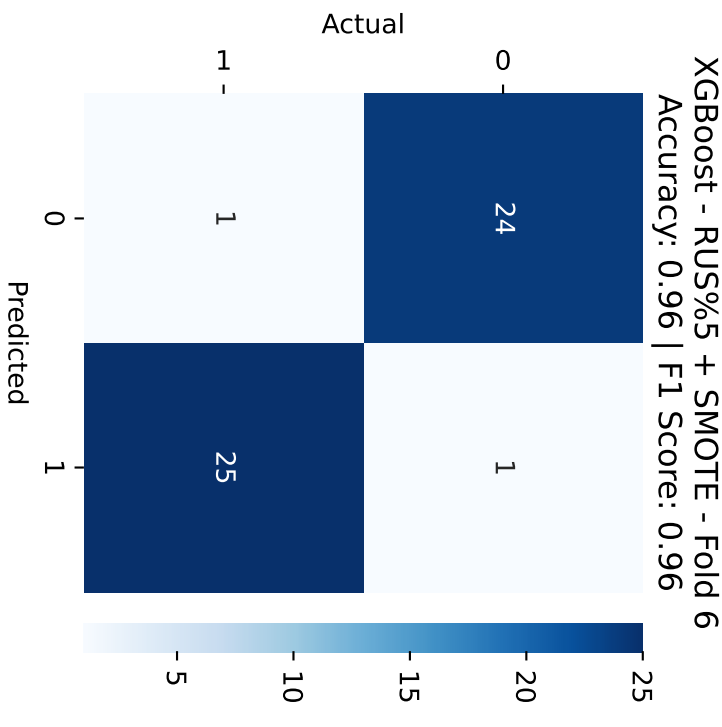
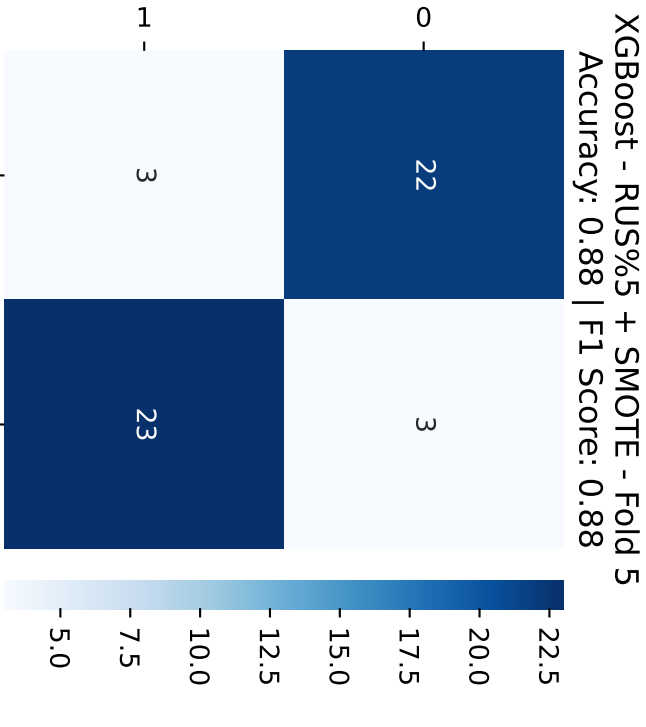
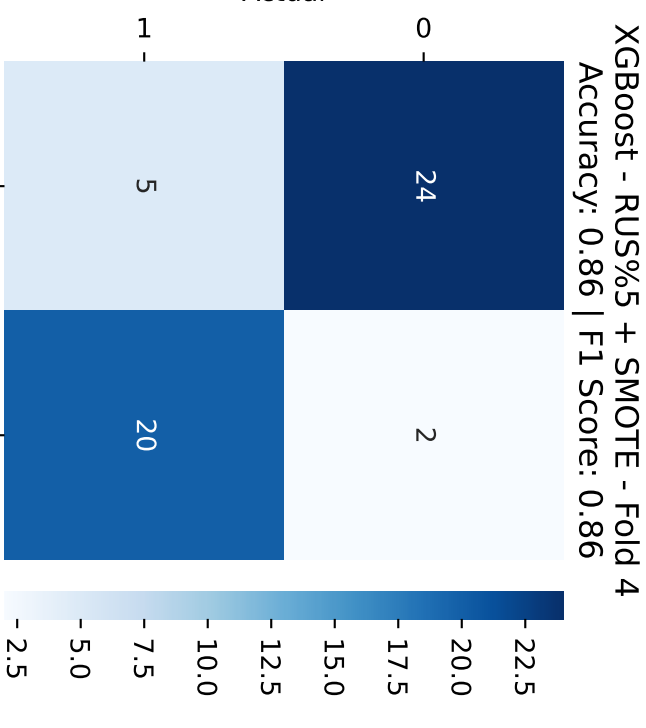
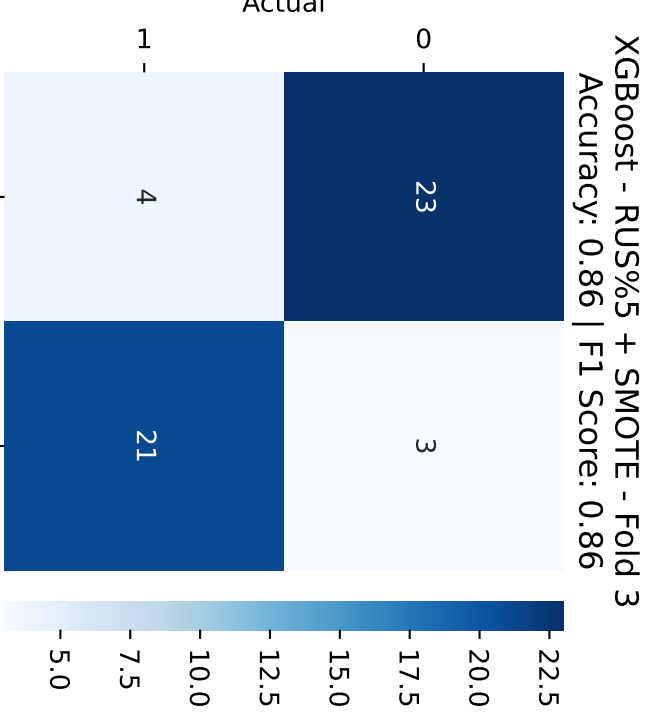
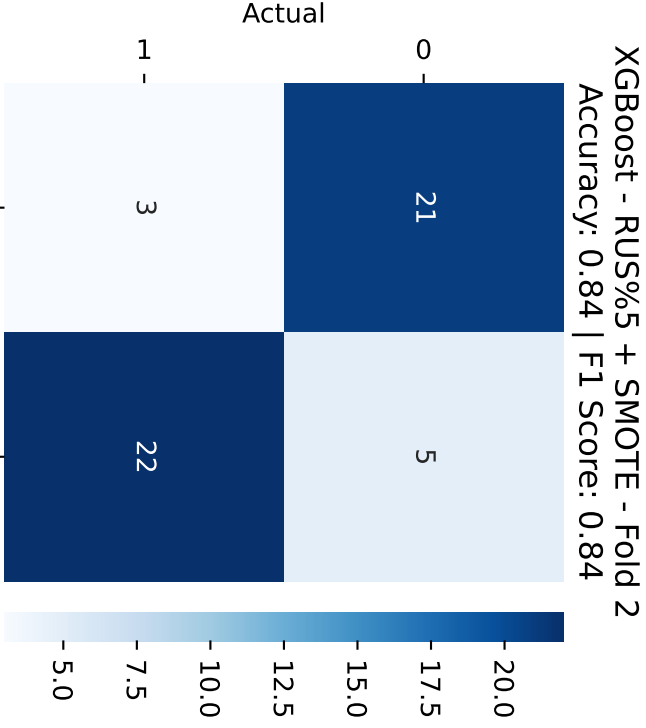
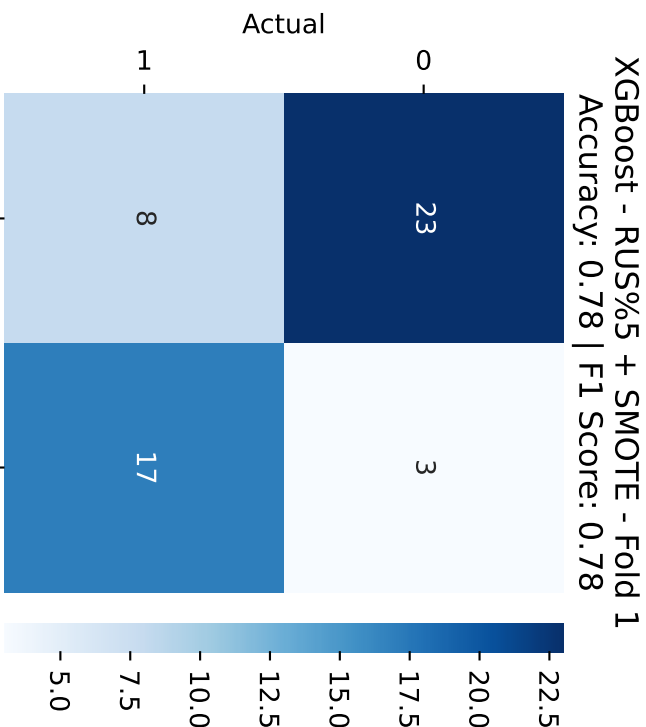


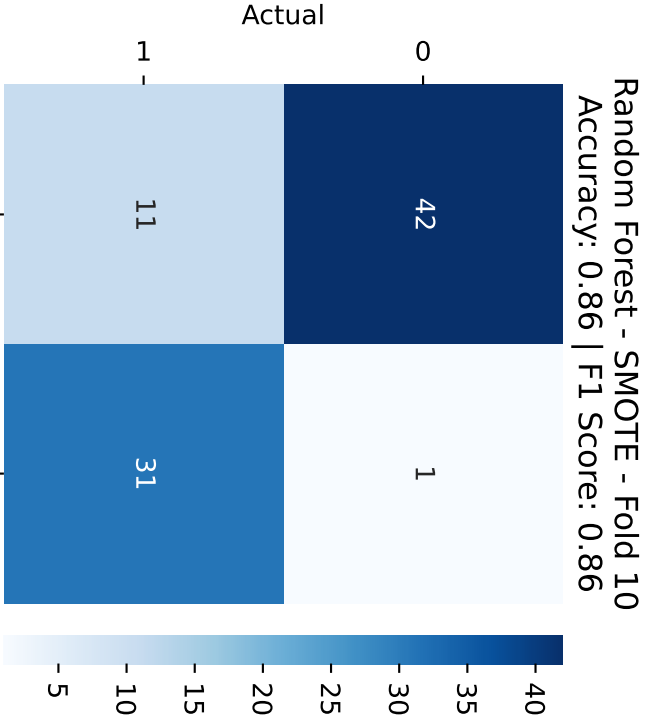
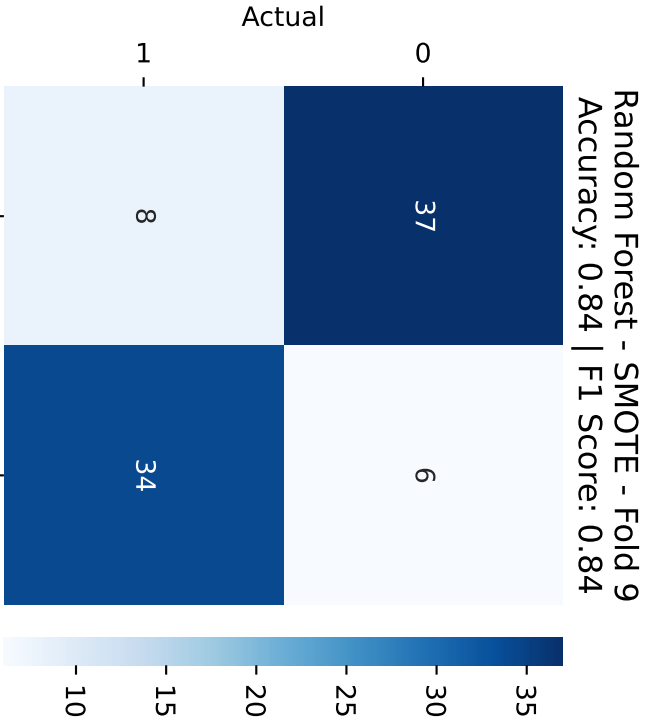
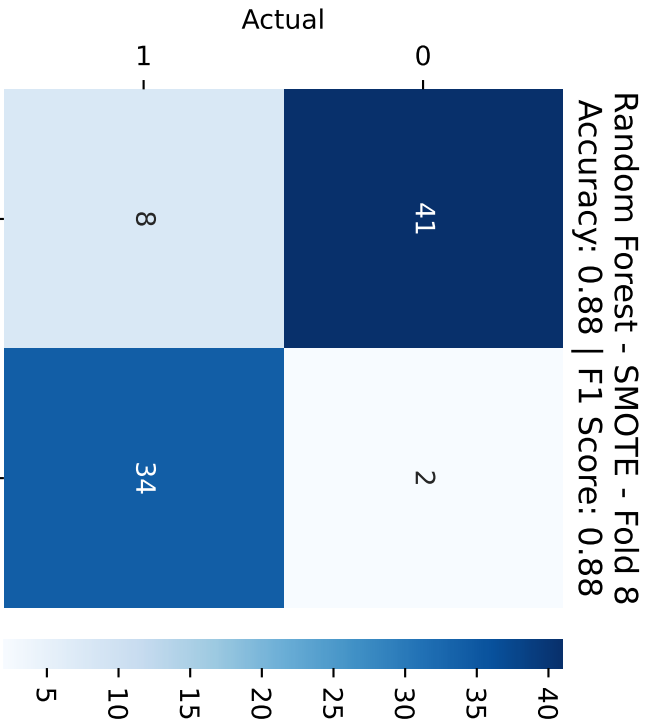
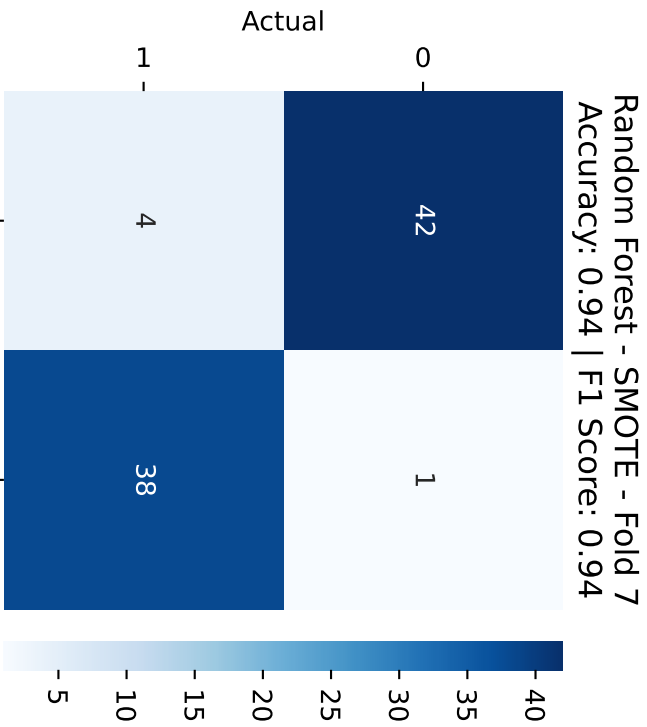
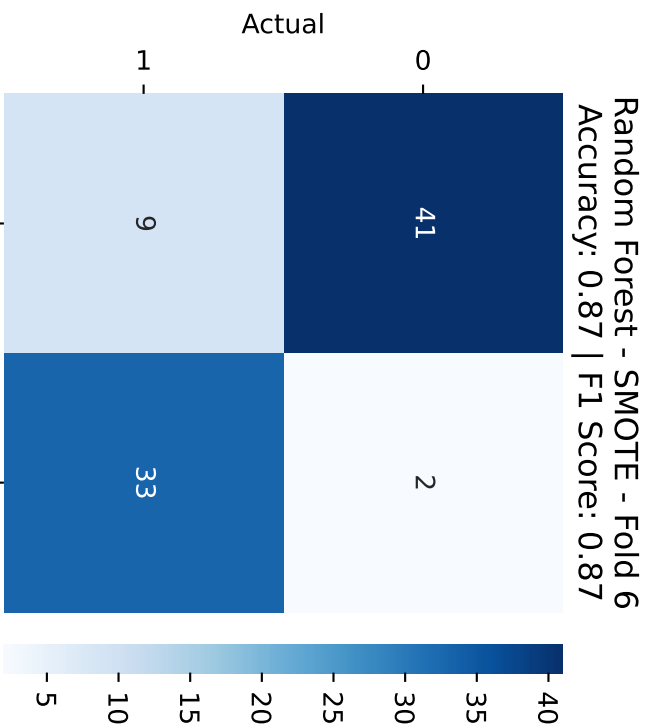
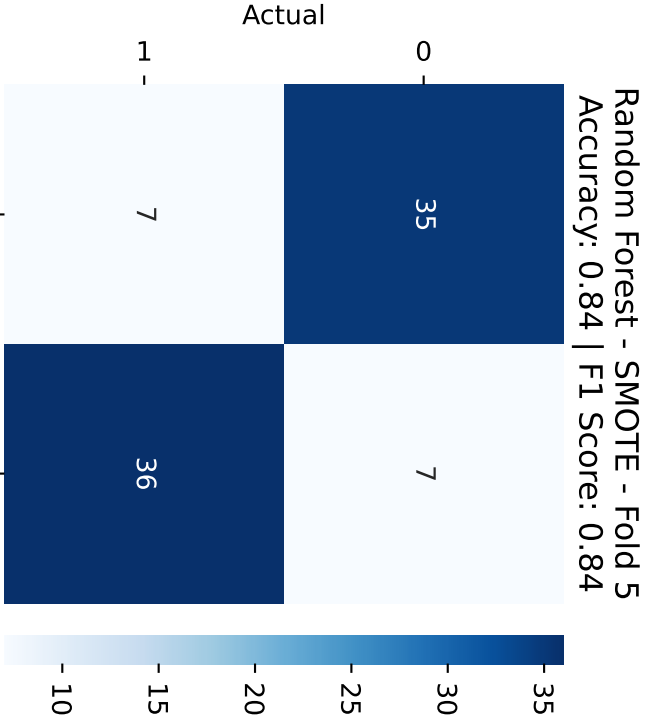
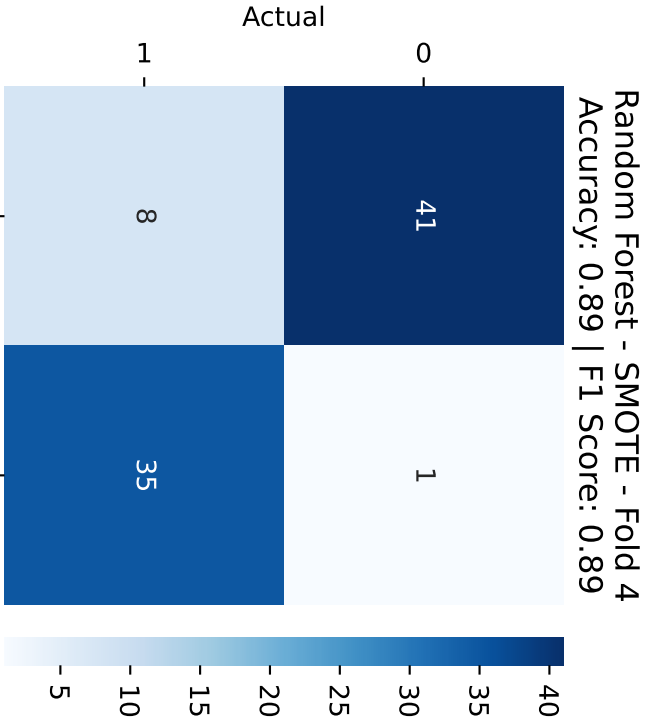
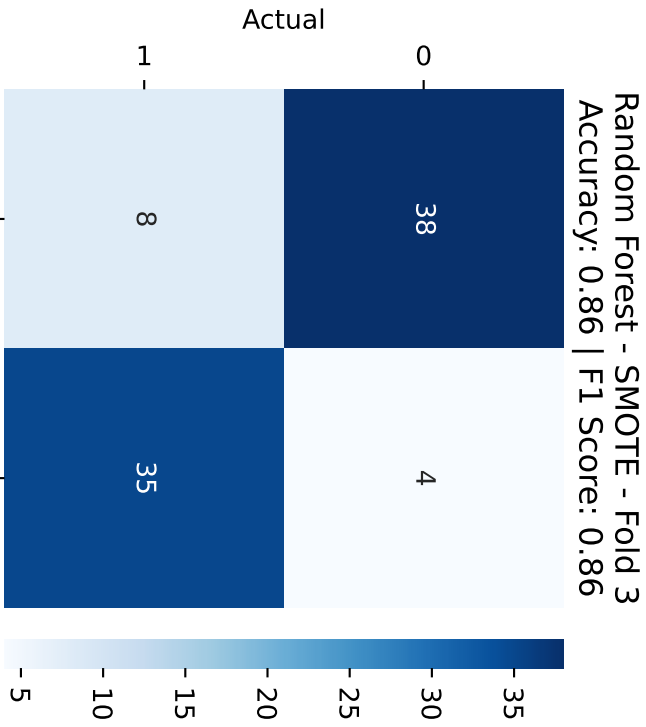
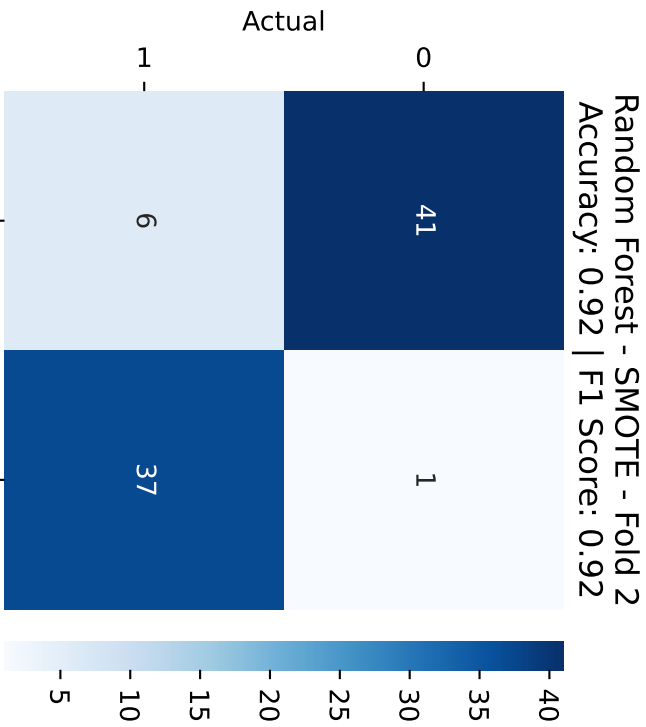
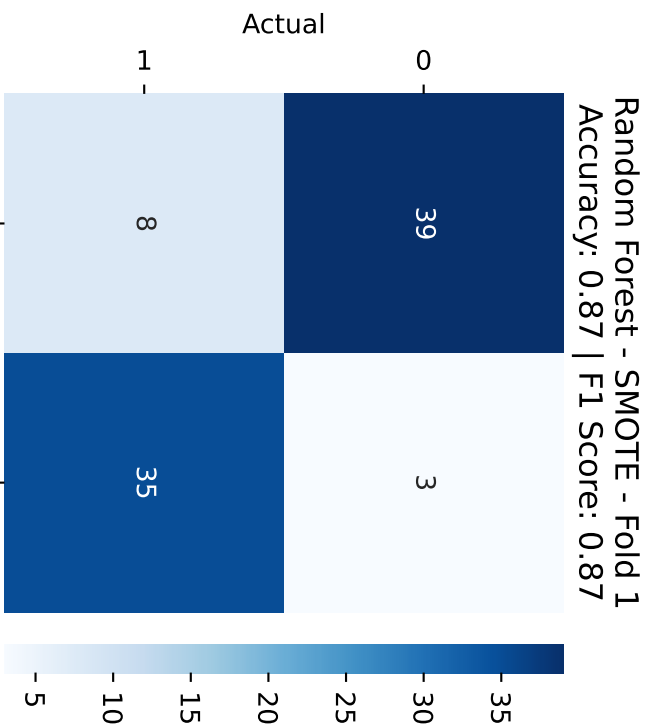


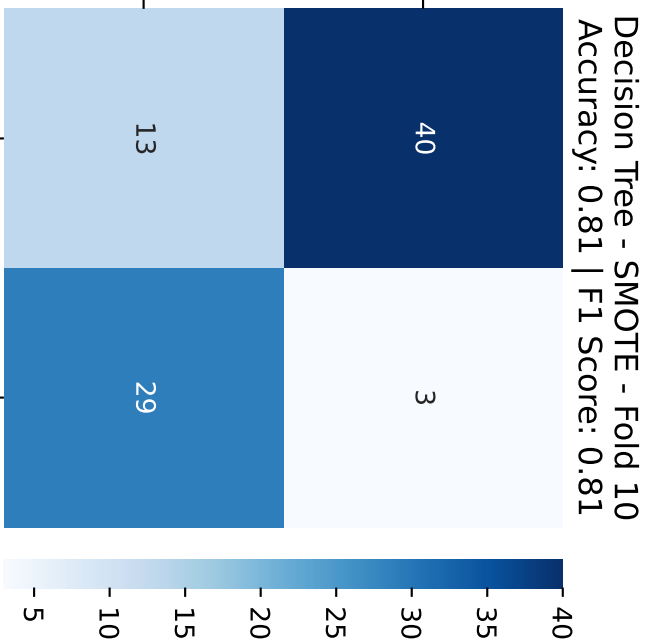
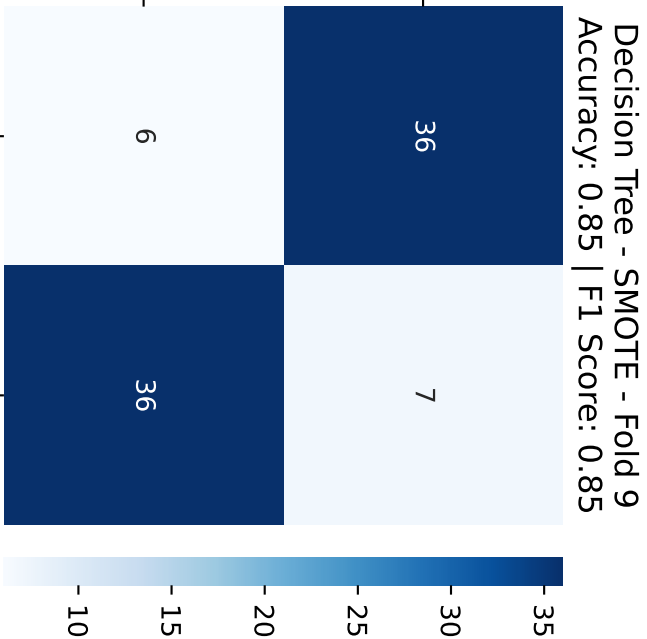
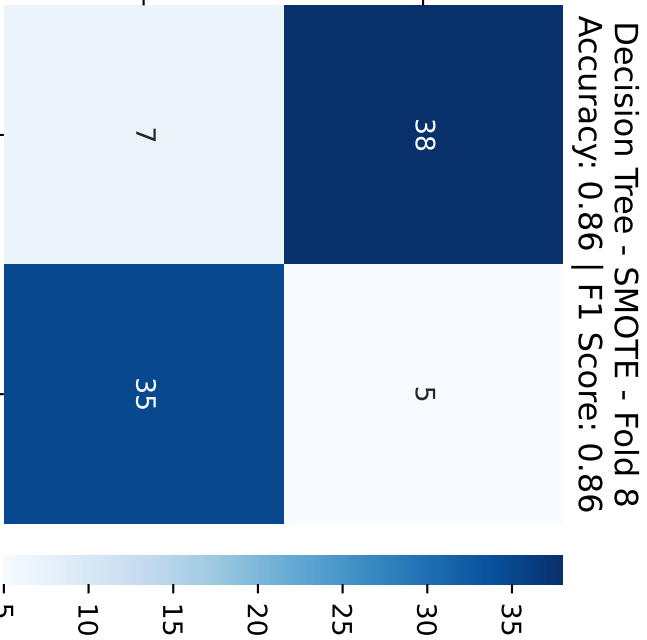
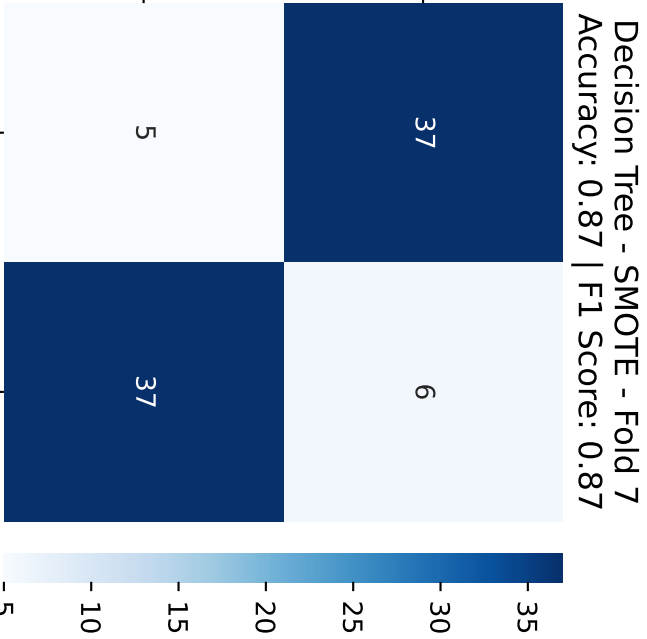
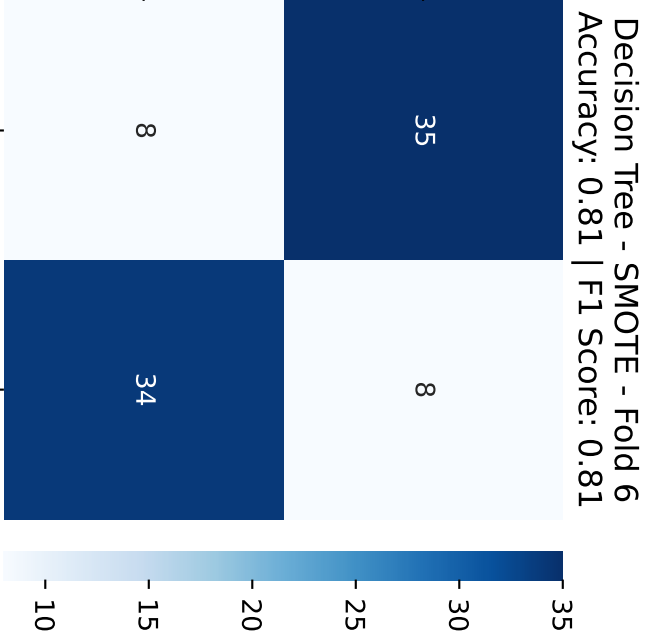
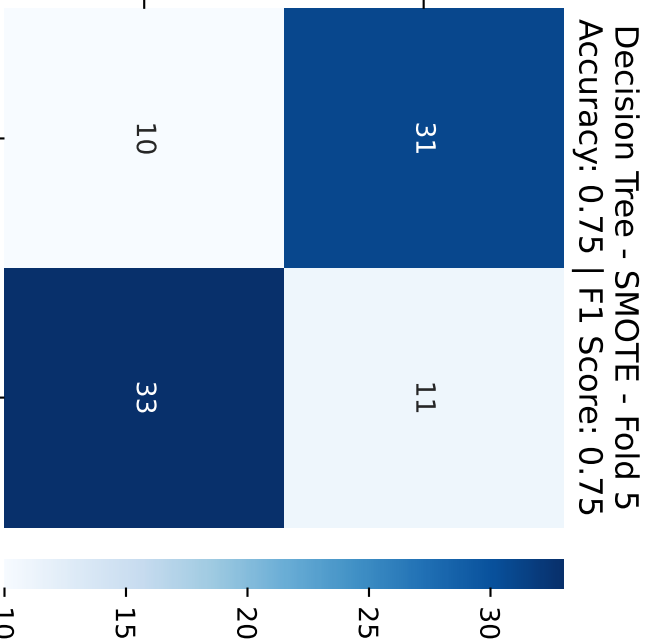
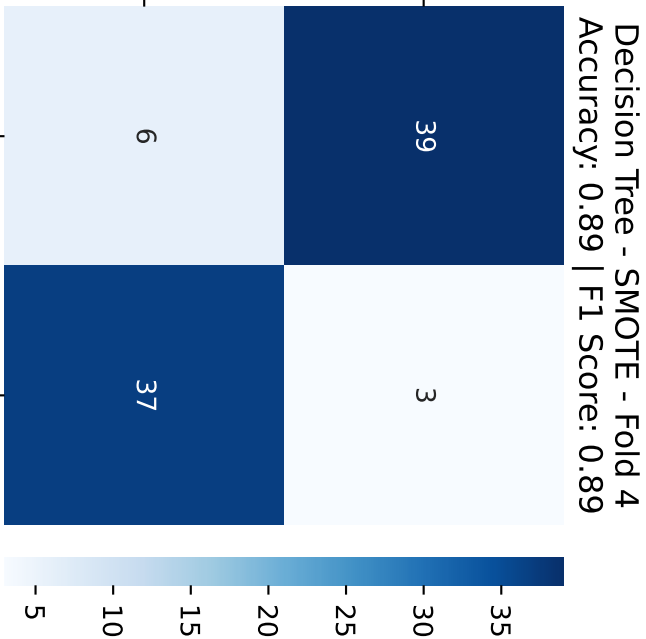
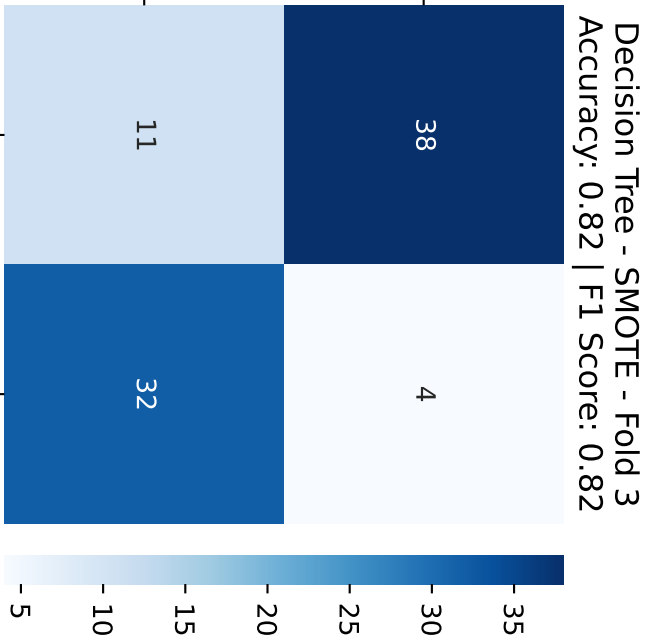
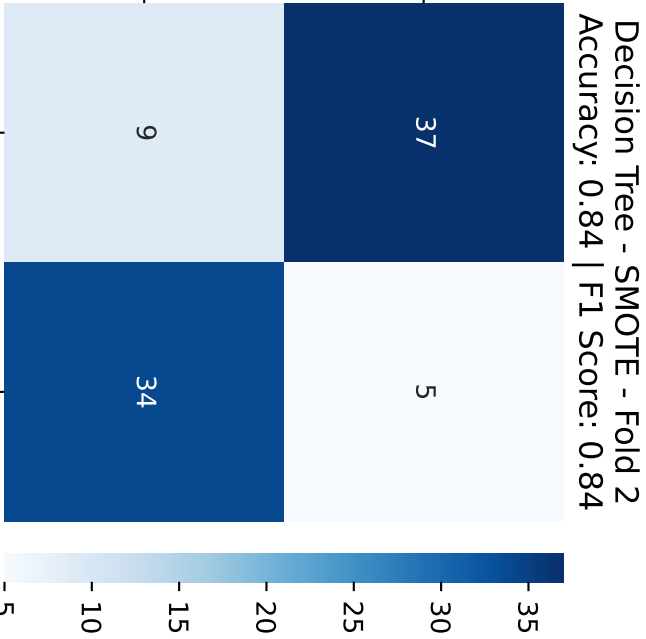
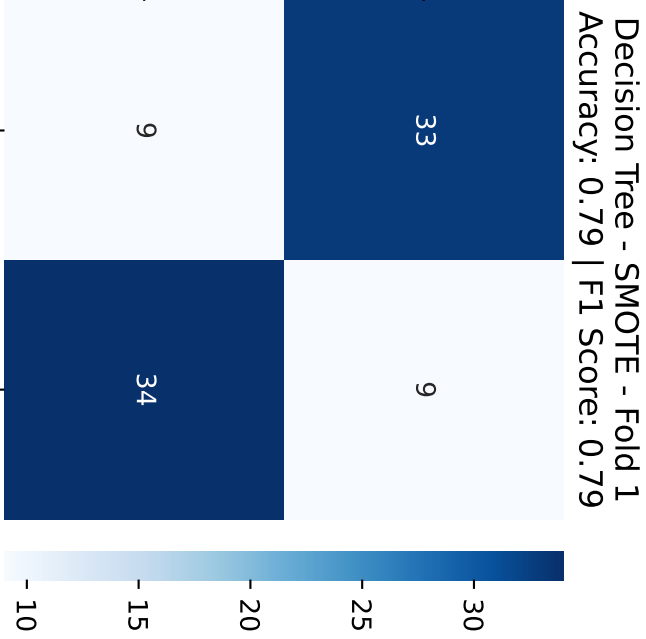


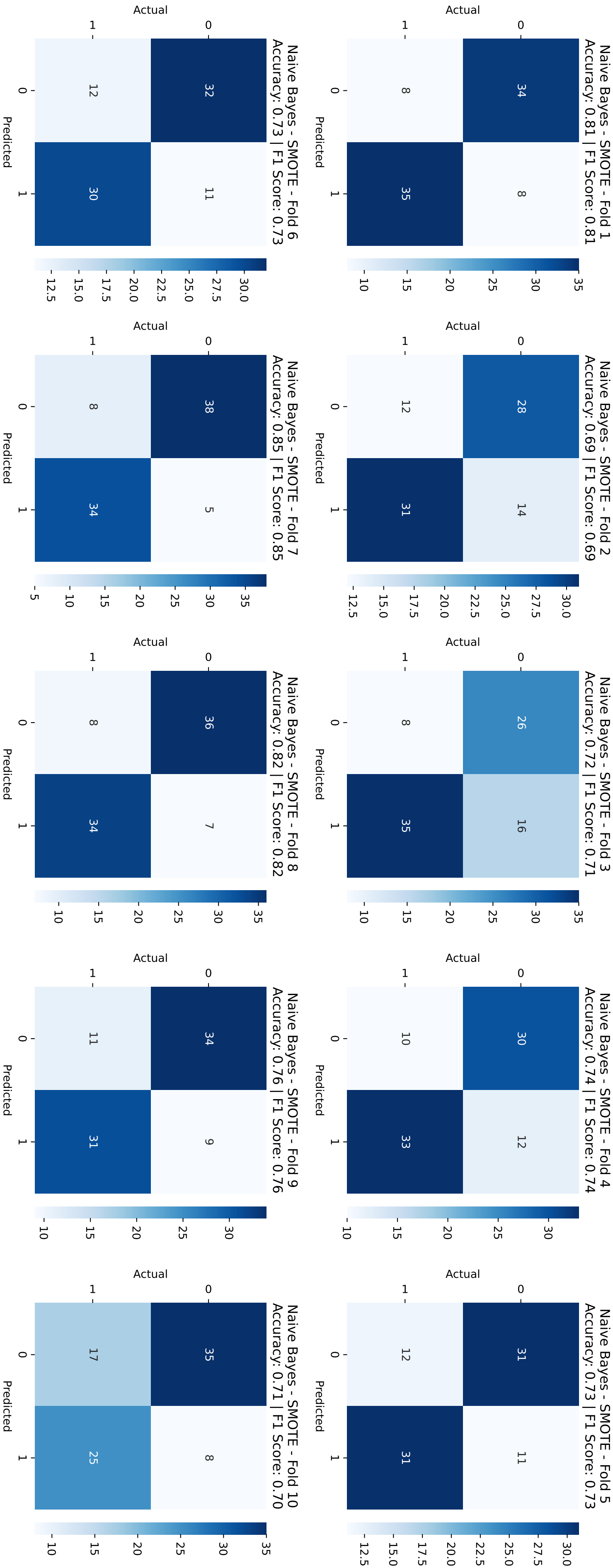


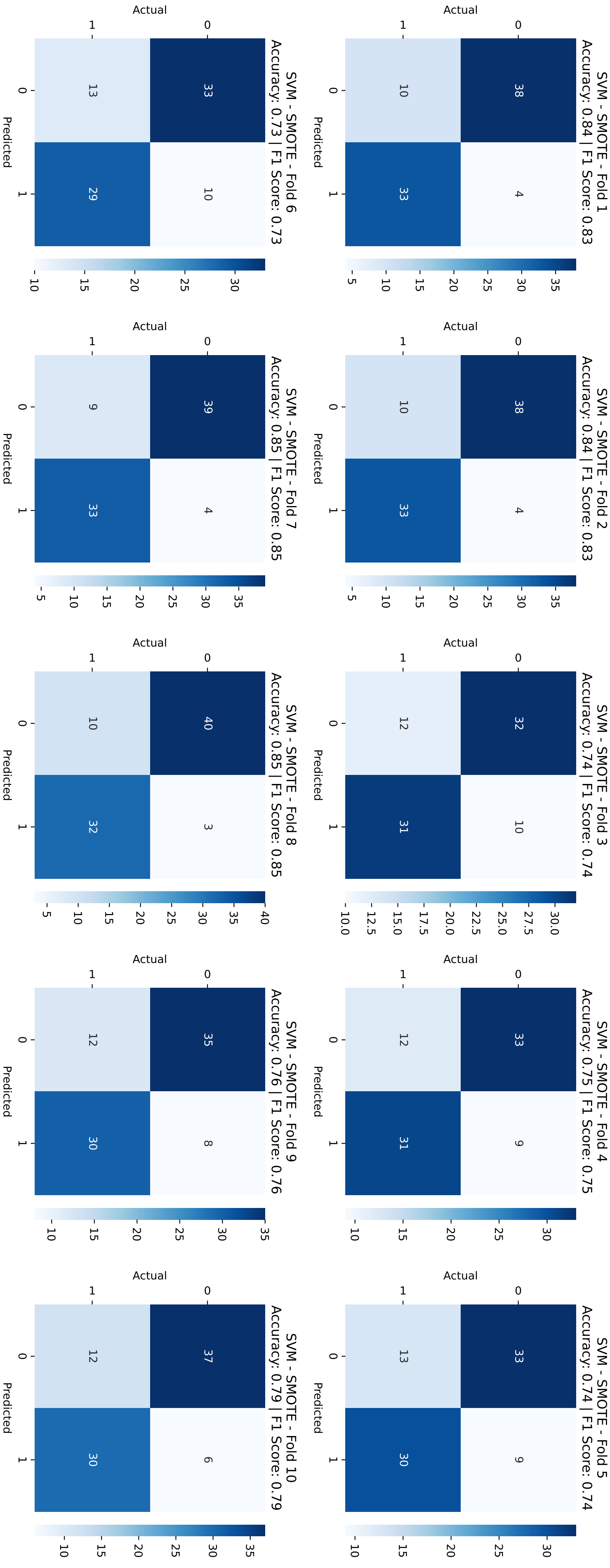


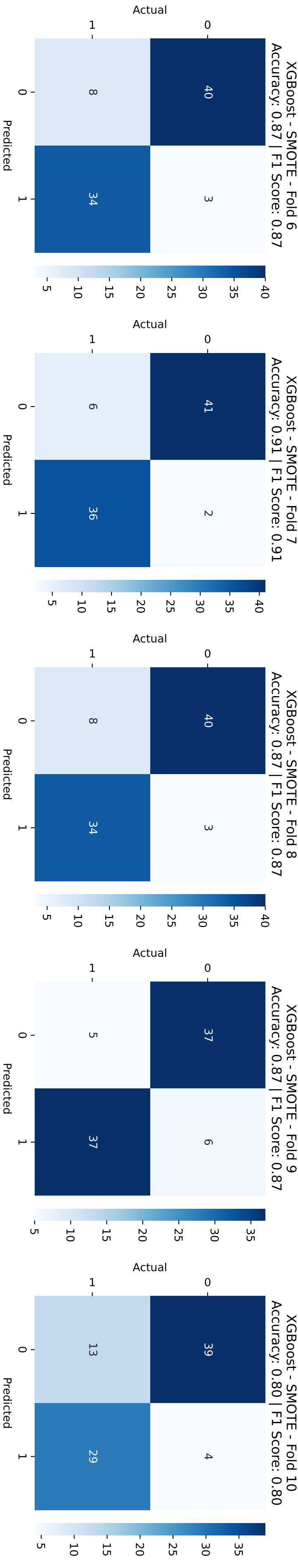
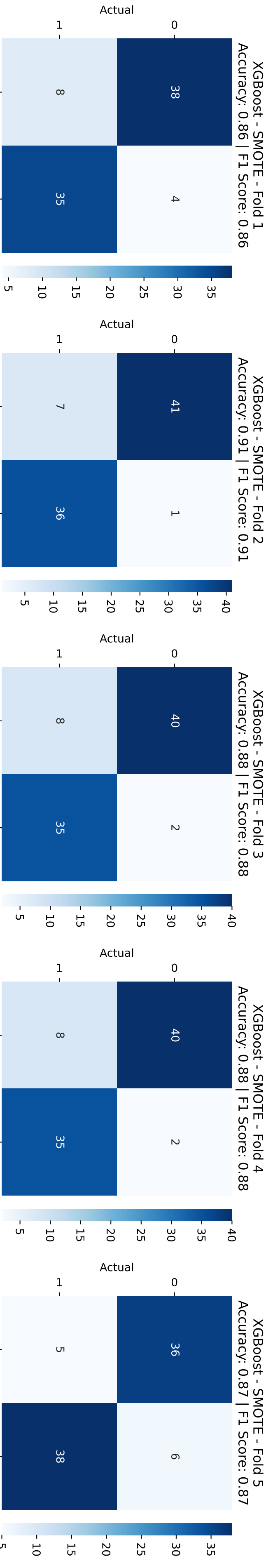


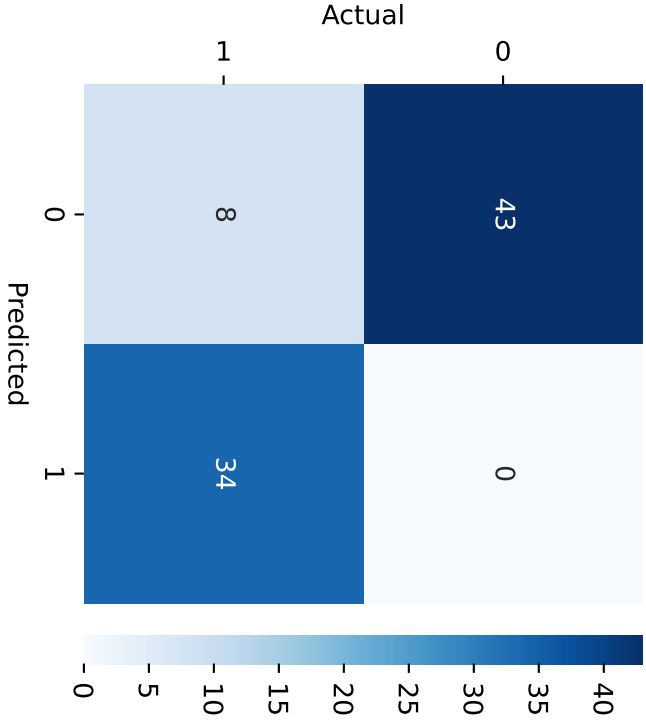
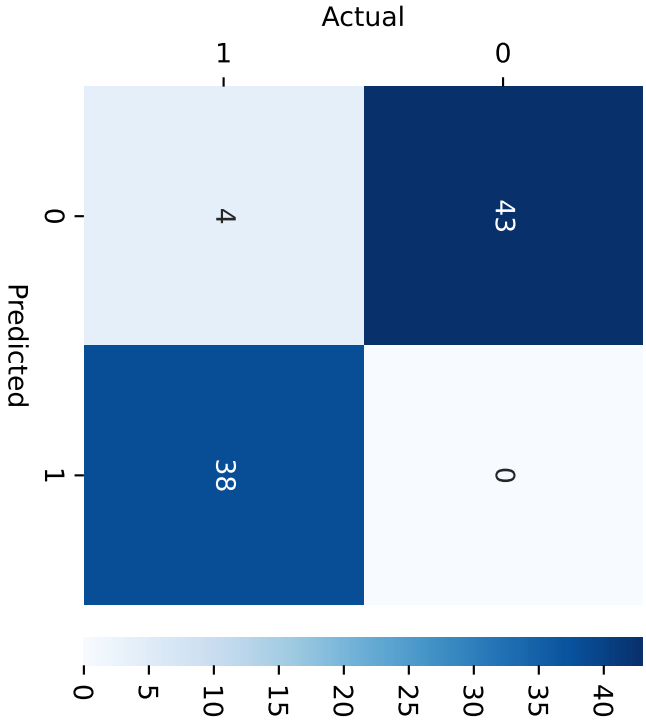
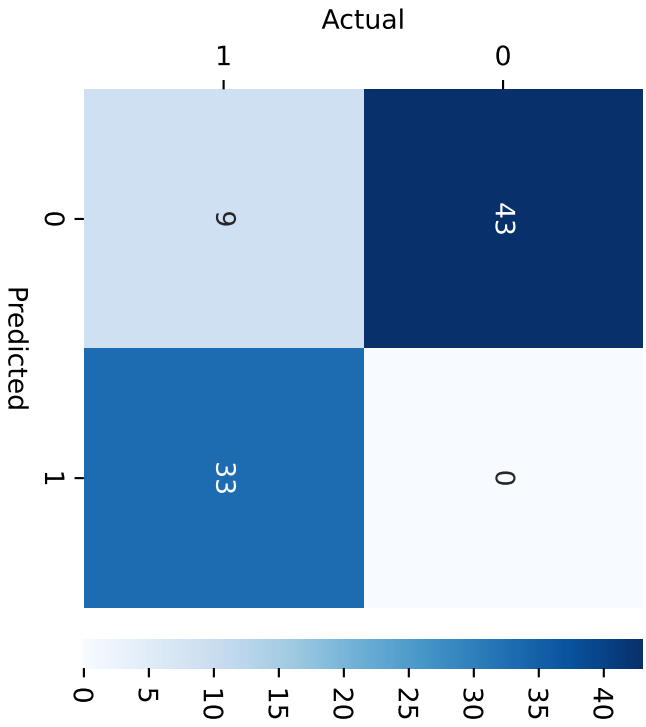
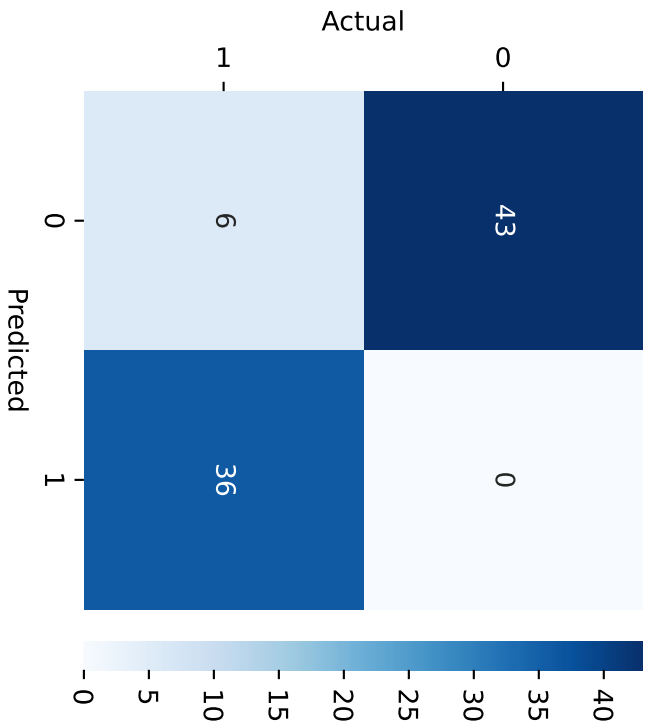
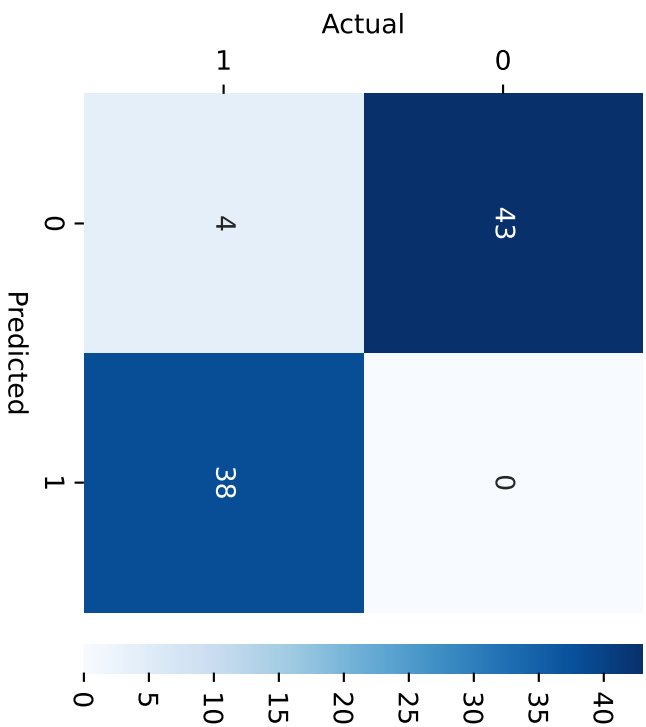
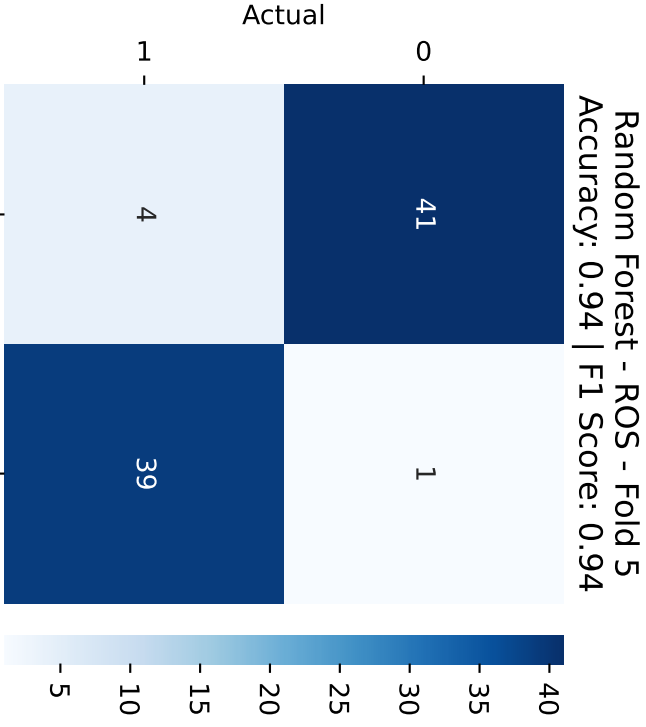
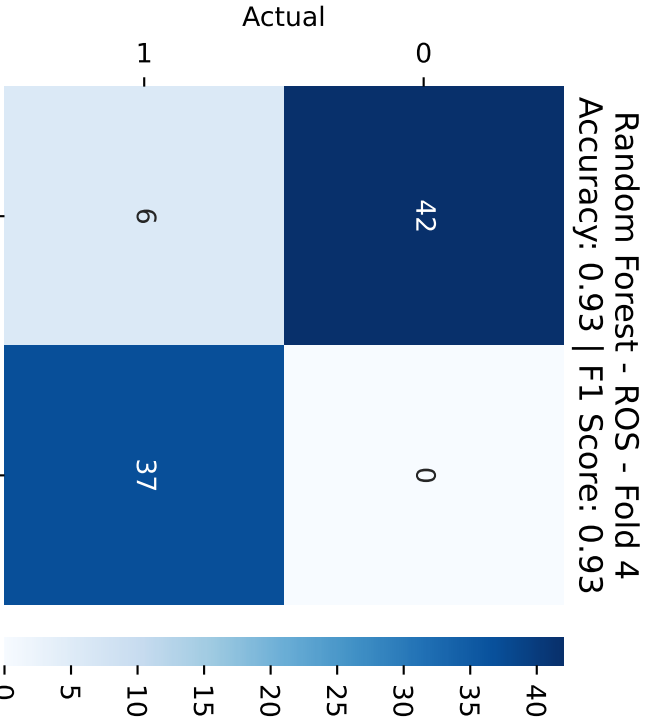
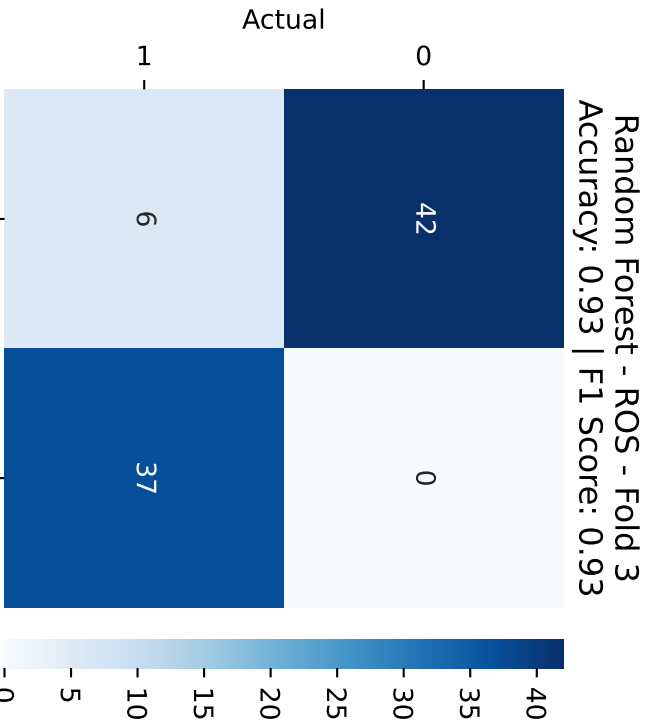
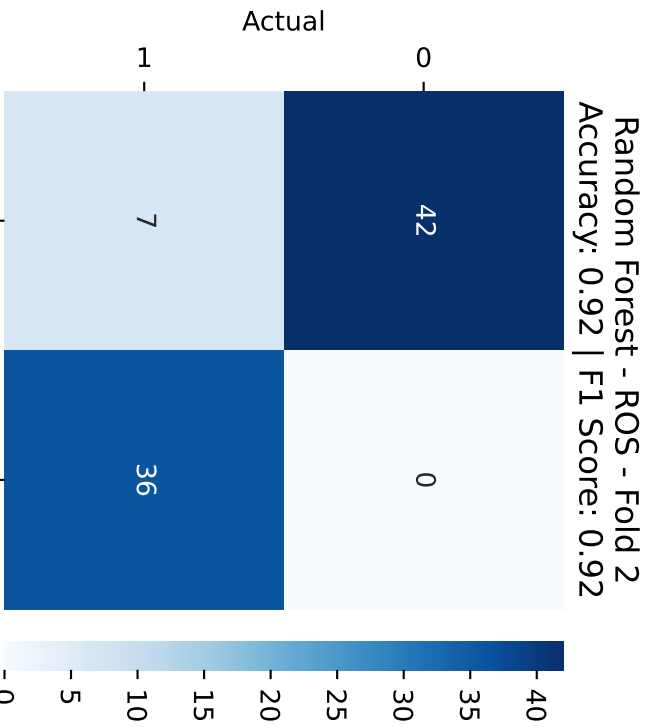
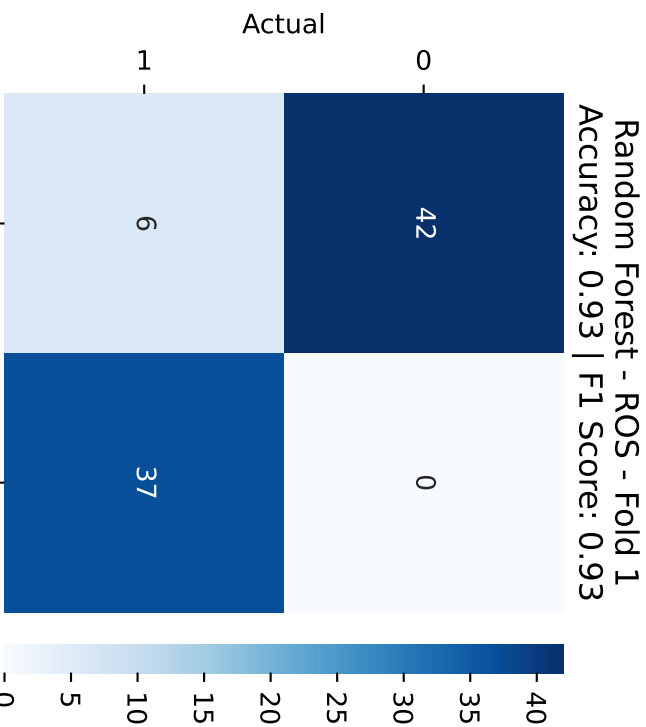


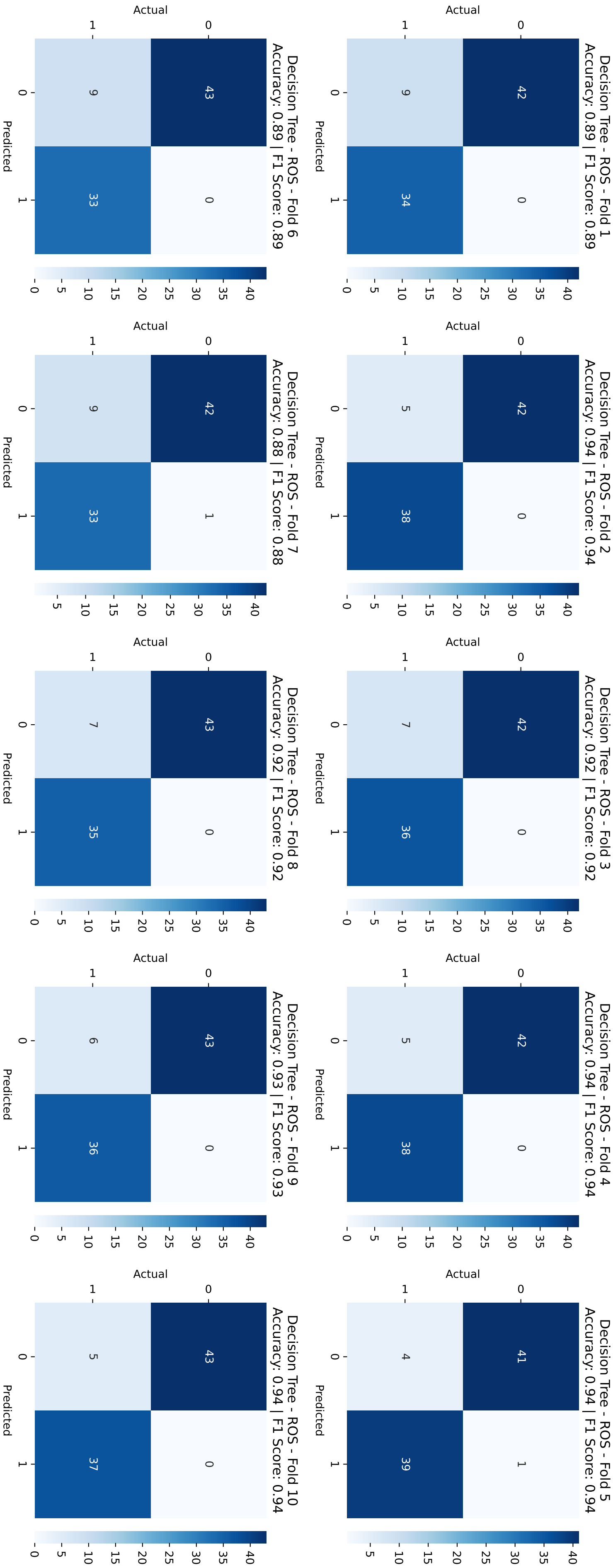


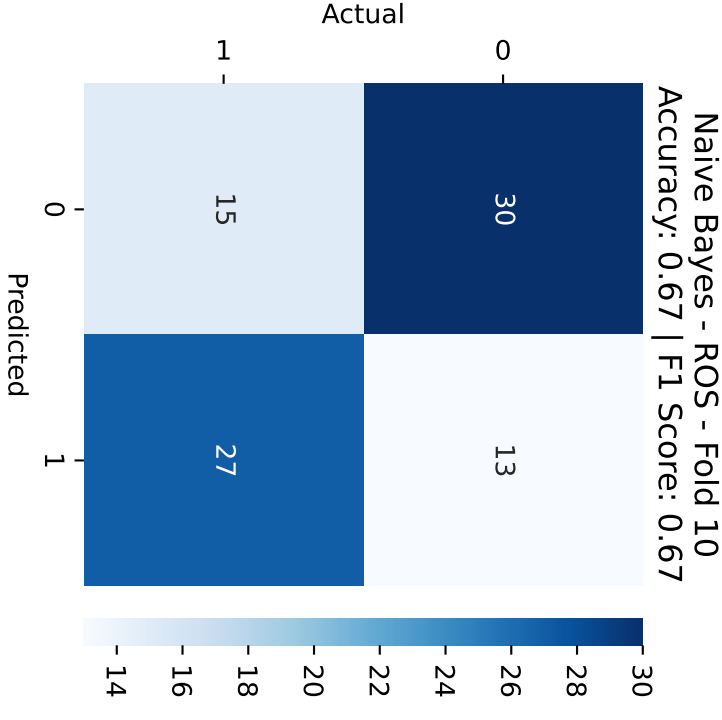
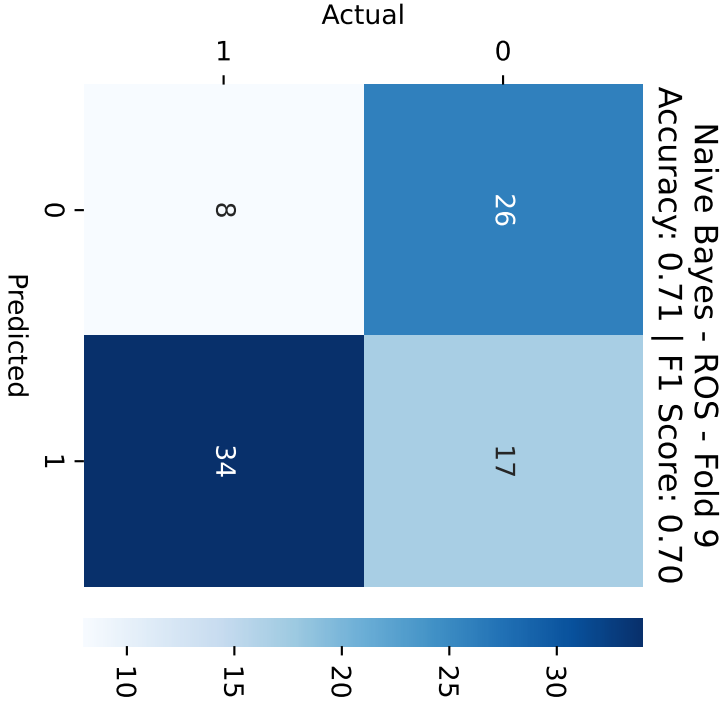
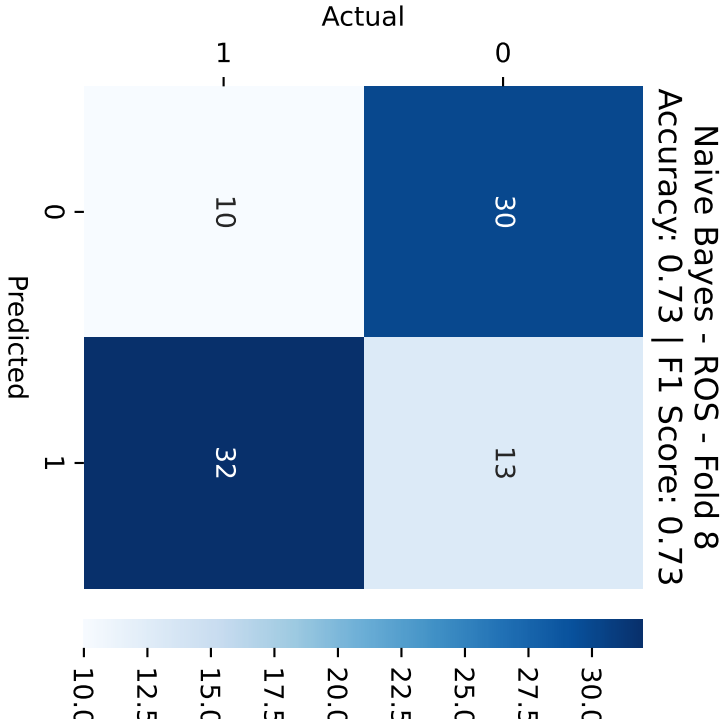
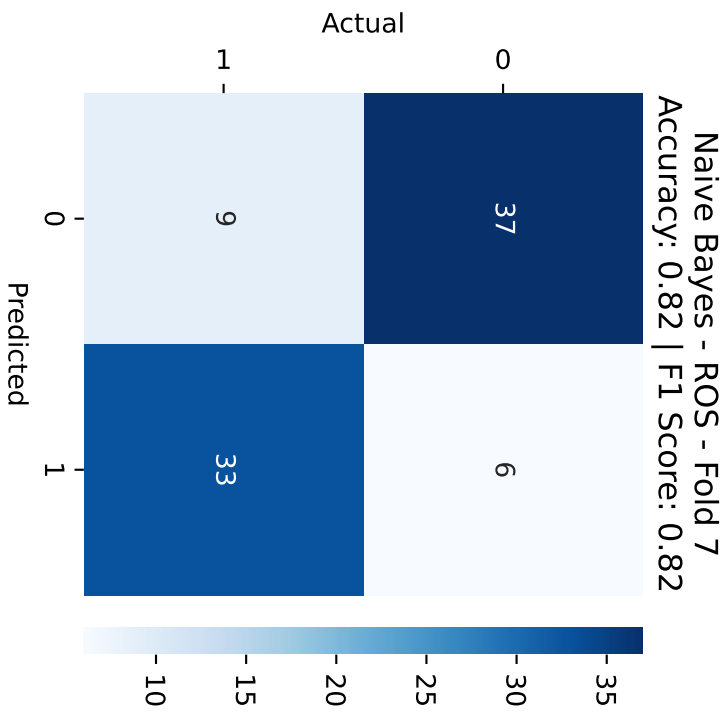
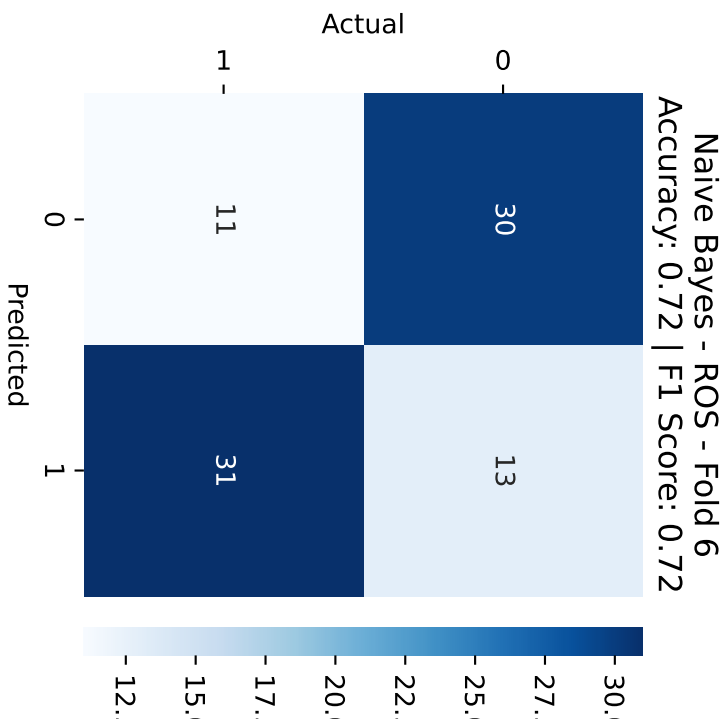
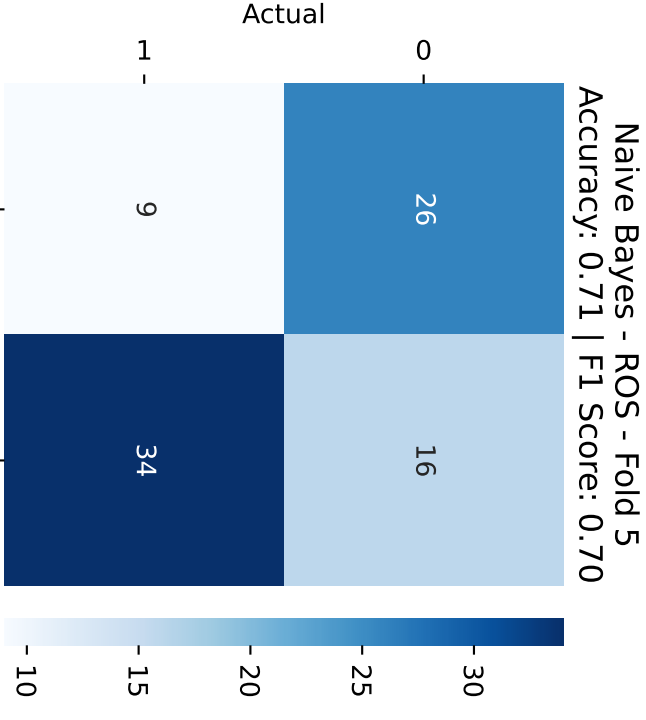
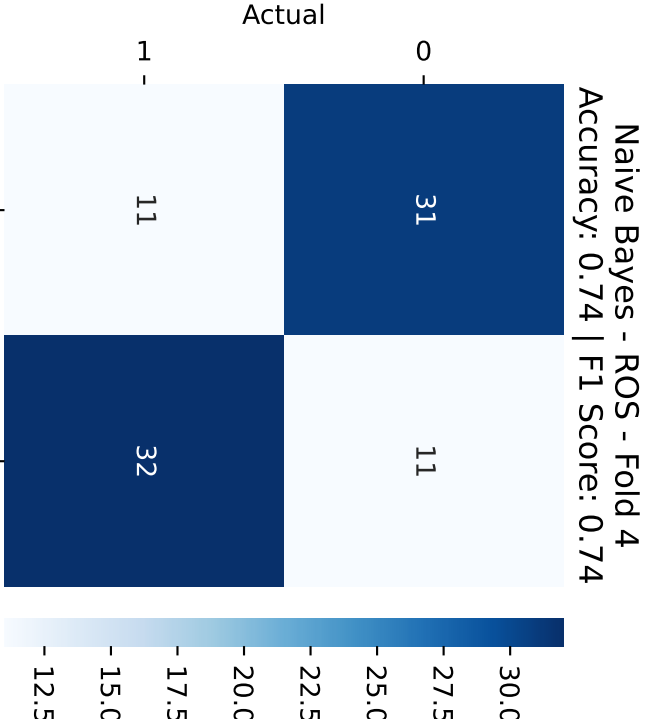
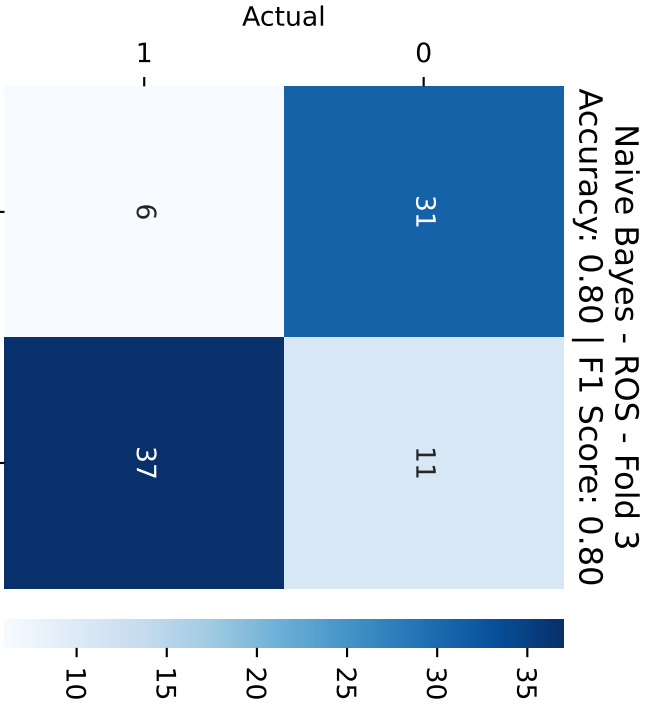
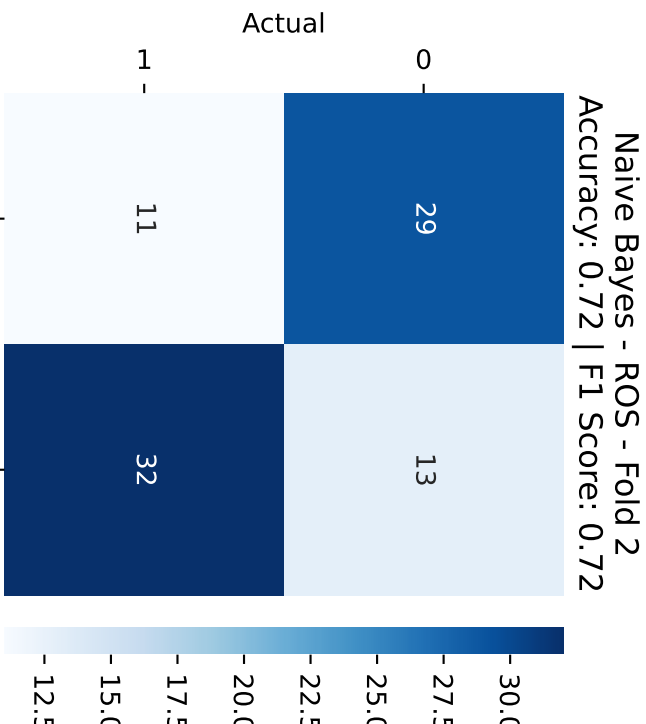
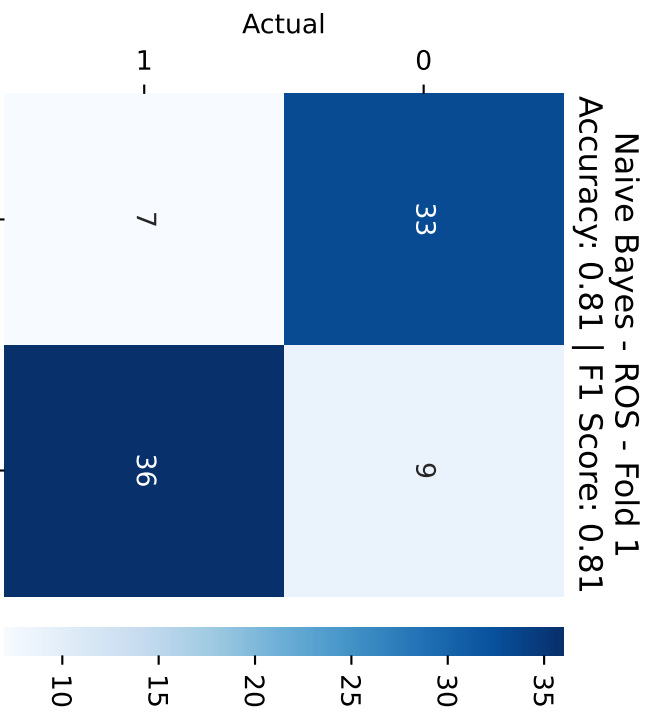


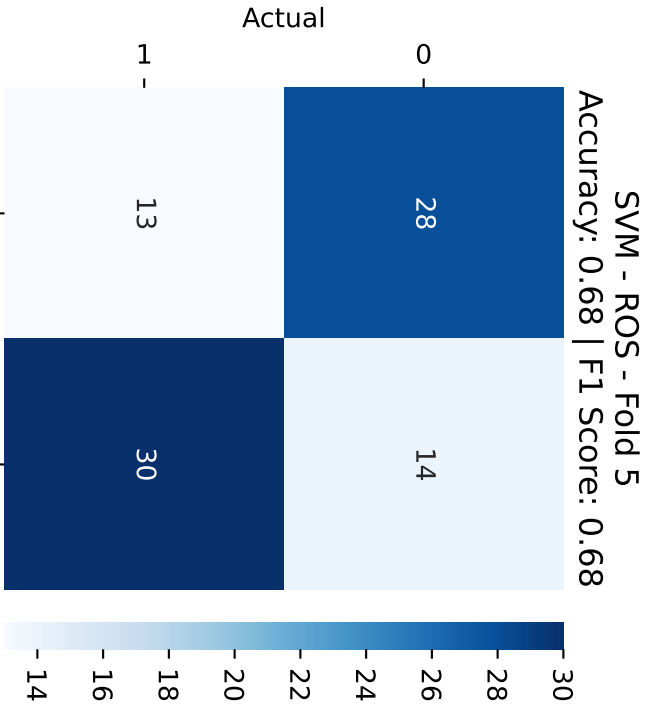
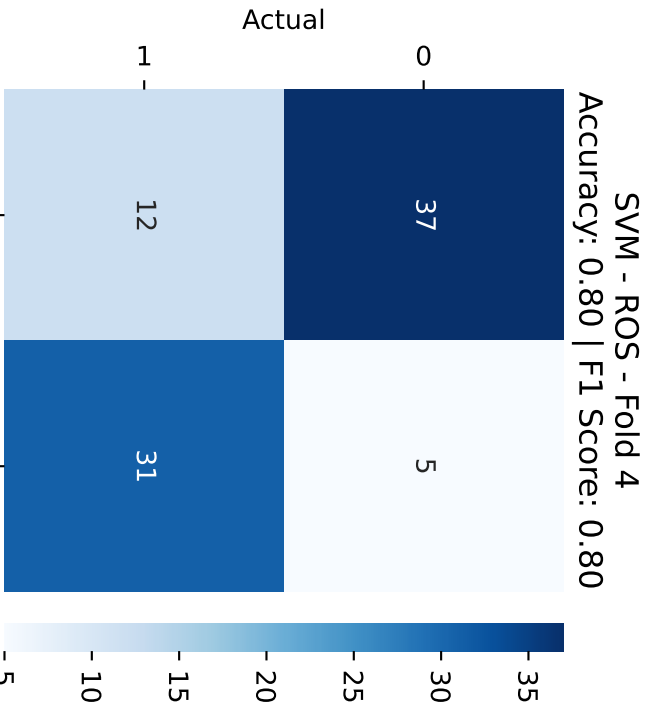
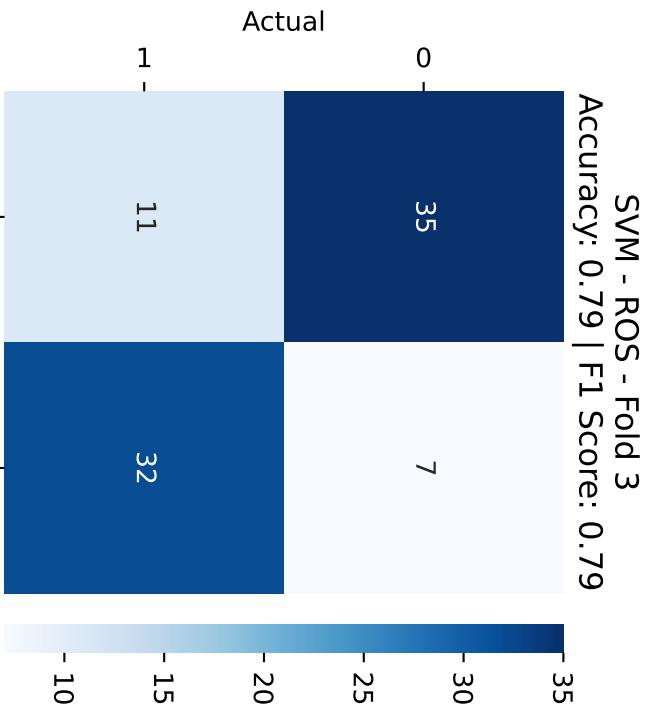
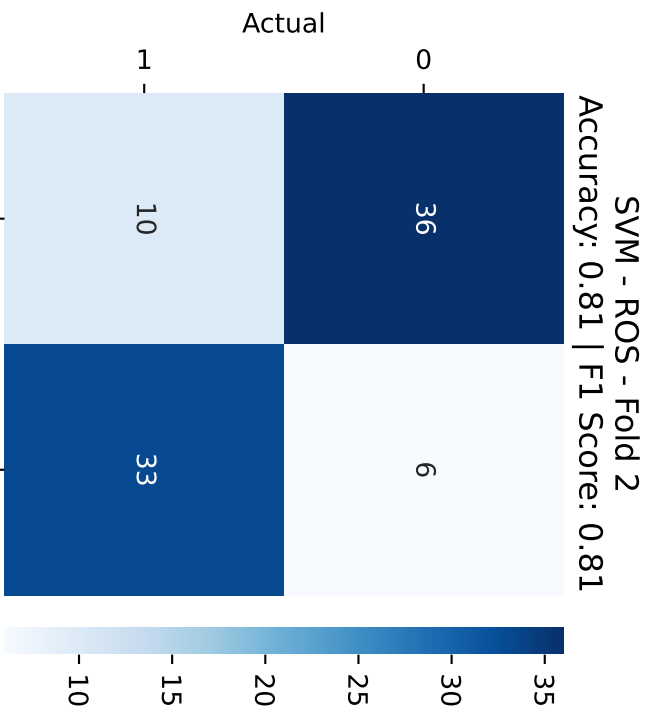
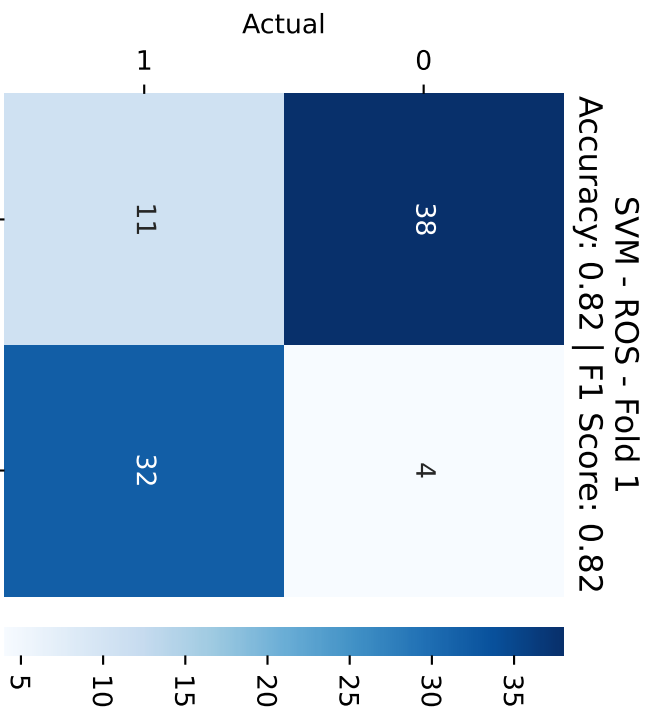




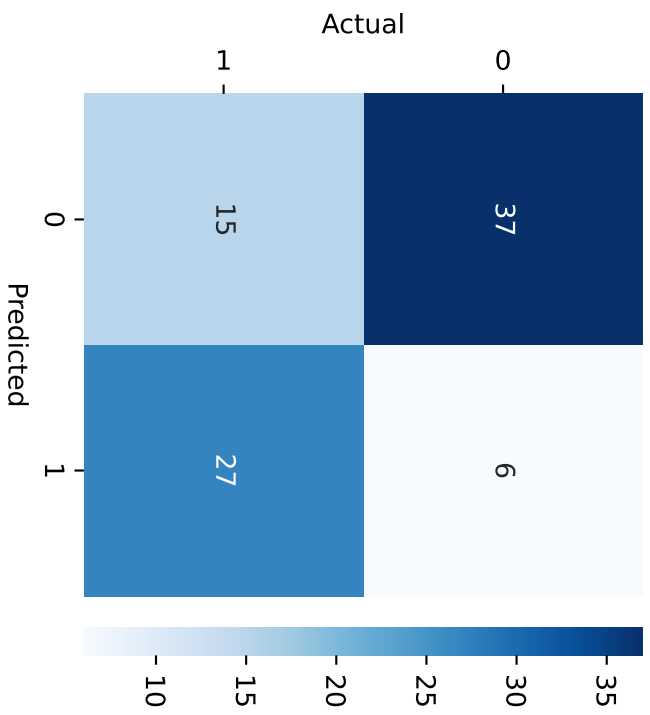




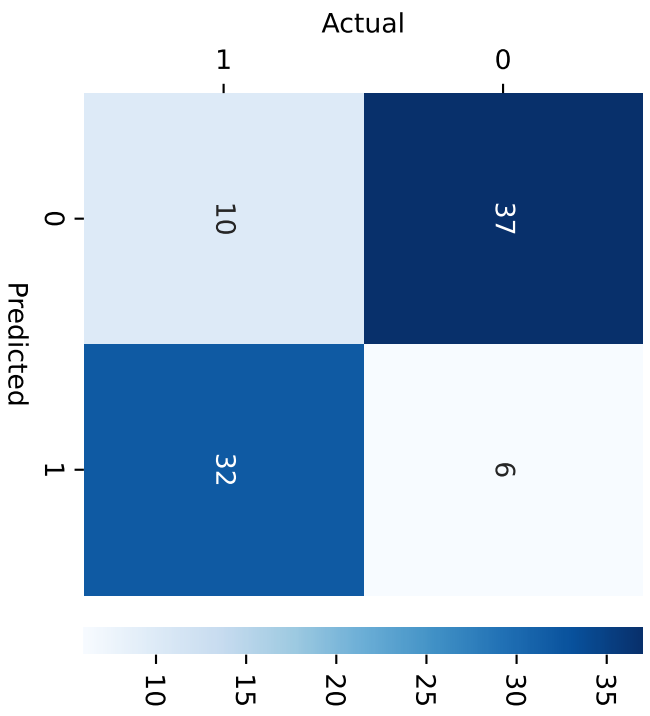




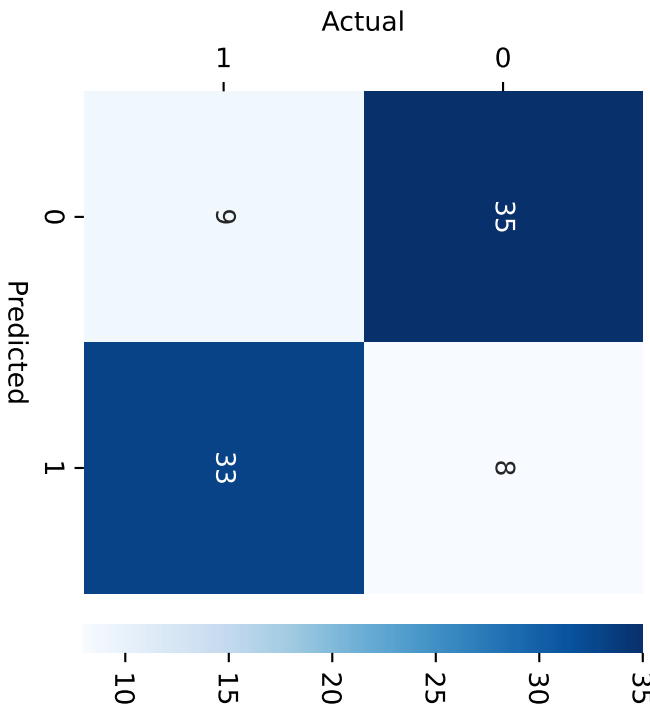
SVM - ROS - Fold 6  
Accuracy: 0.75 | F1 Score: 0.75



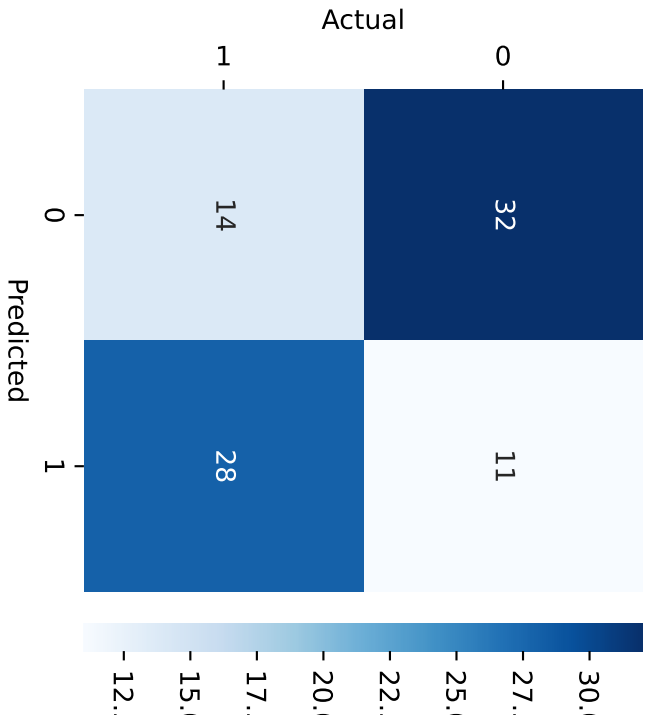
SVM - ROS - Fold 7  
Accuracy: 0.81 | F1 Score: 0.81



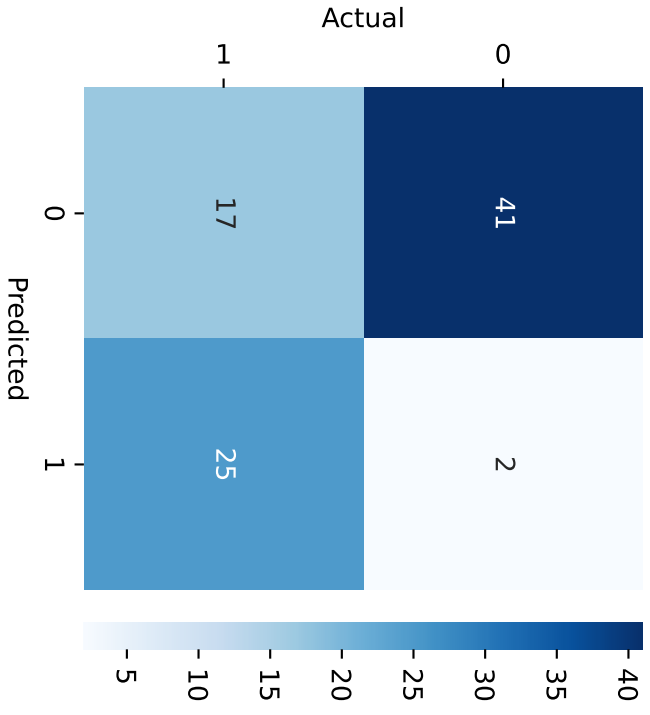
SVM - ROS - Fold 8  
Accuracy: 0.80 | F1 Score: 0.80

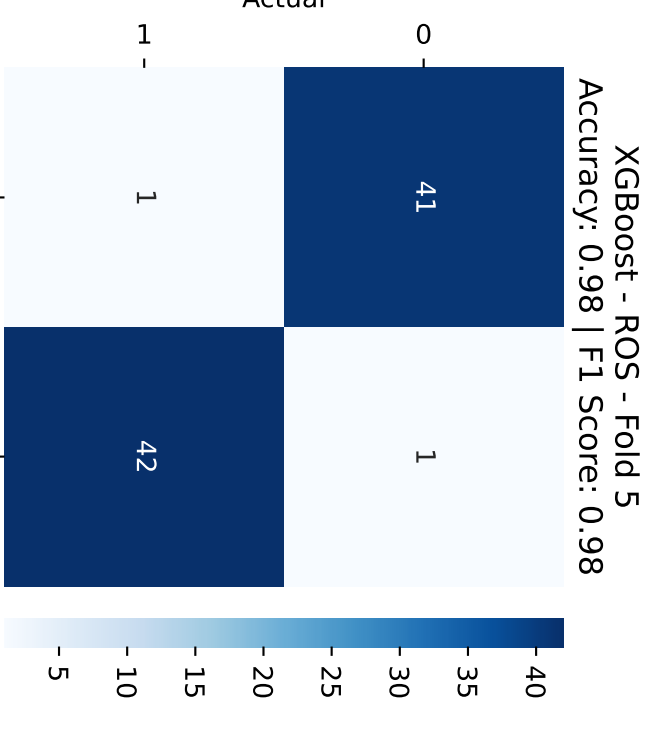
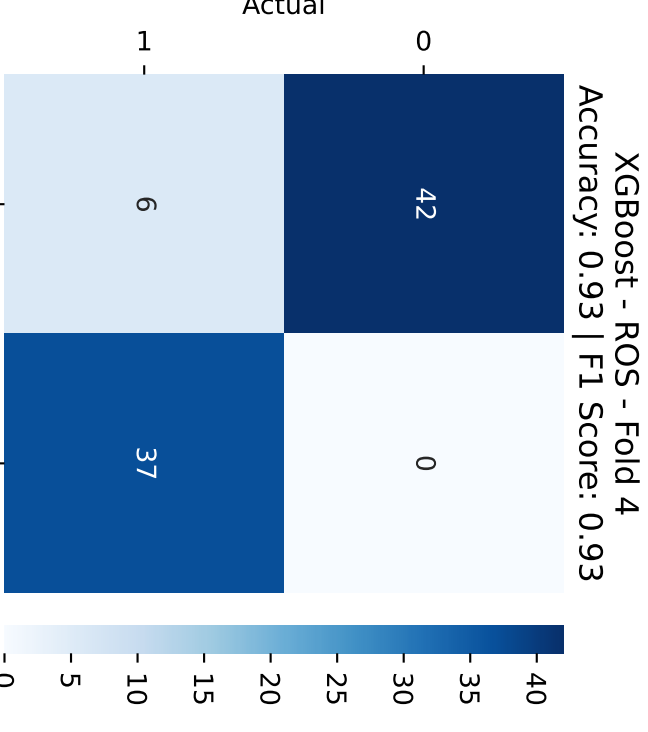
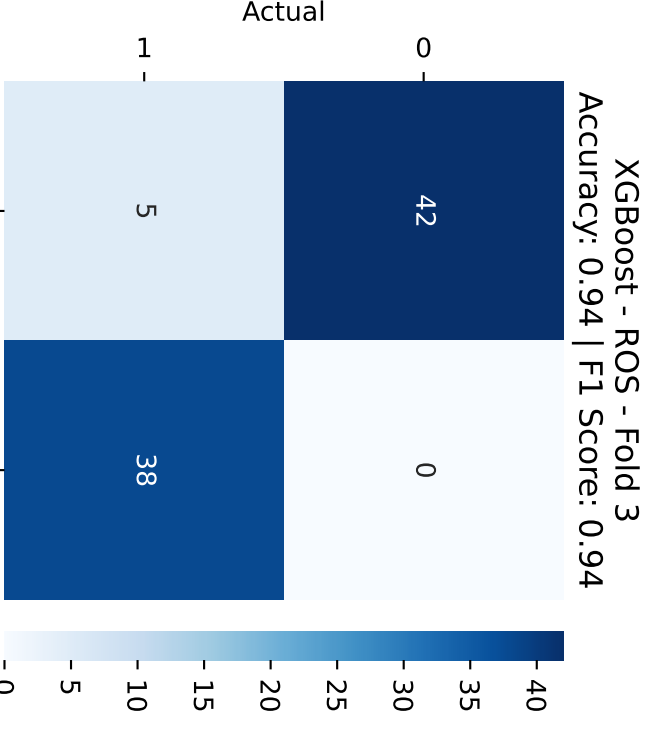
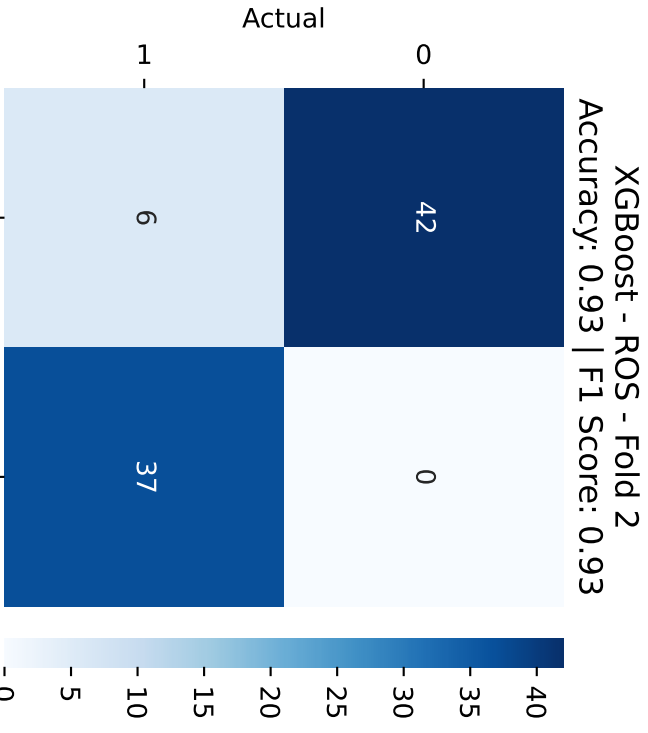
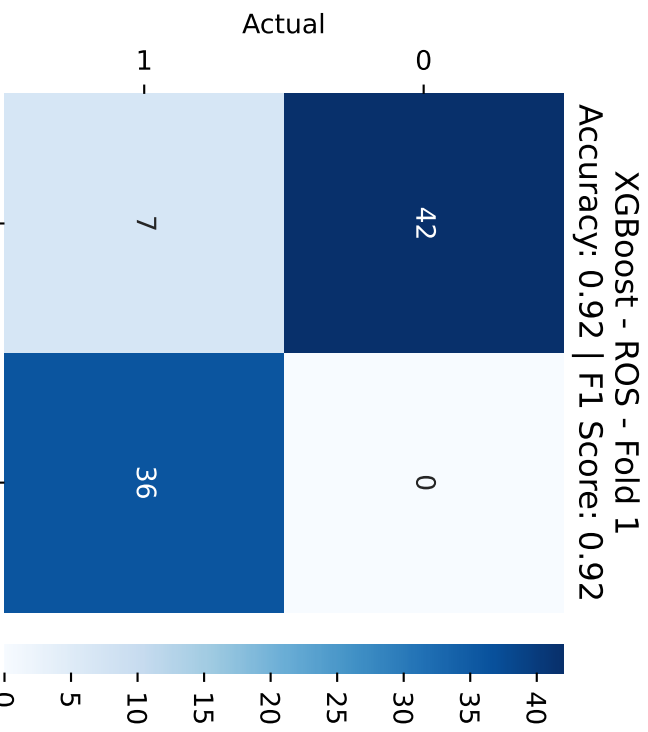


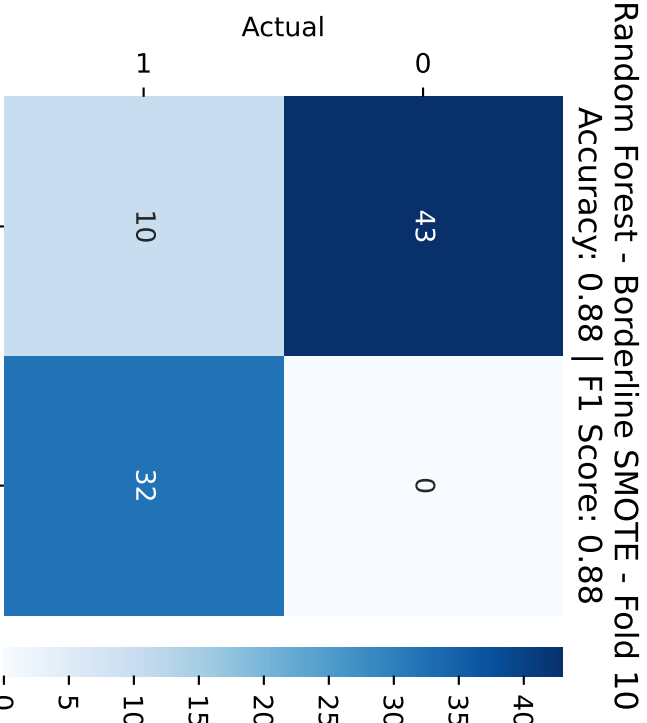
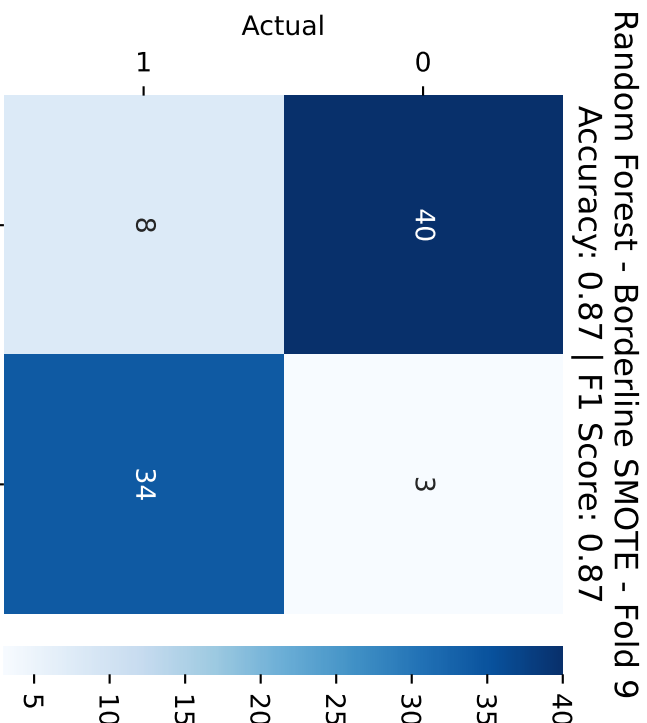
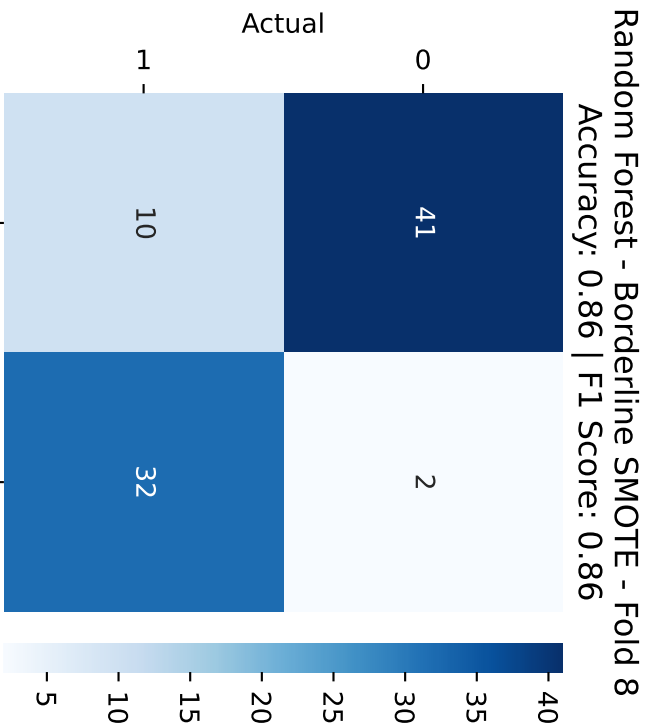
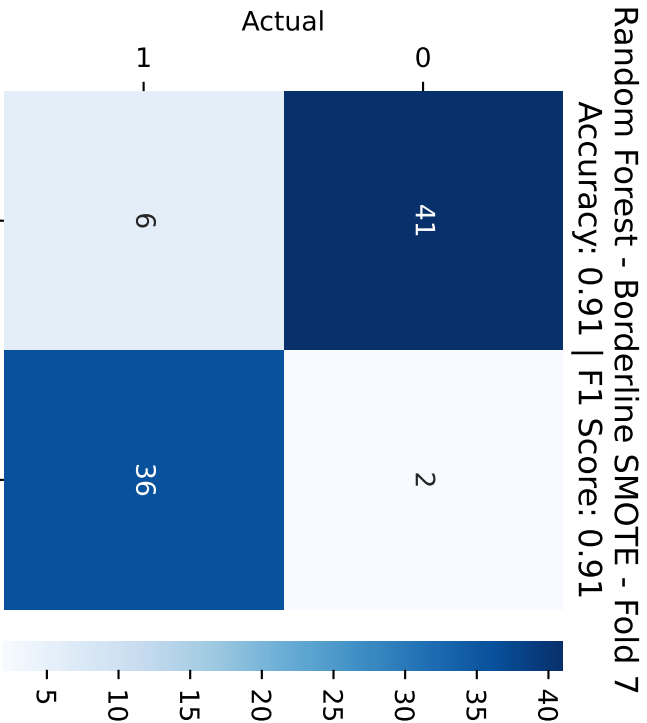
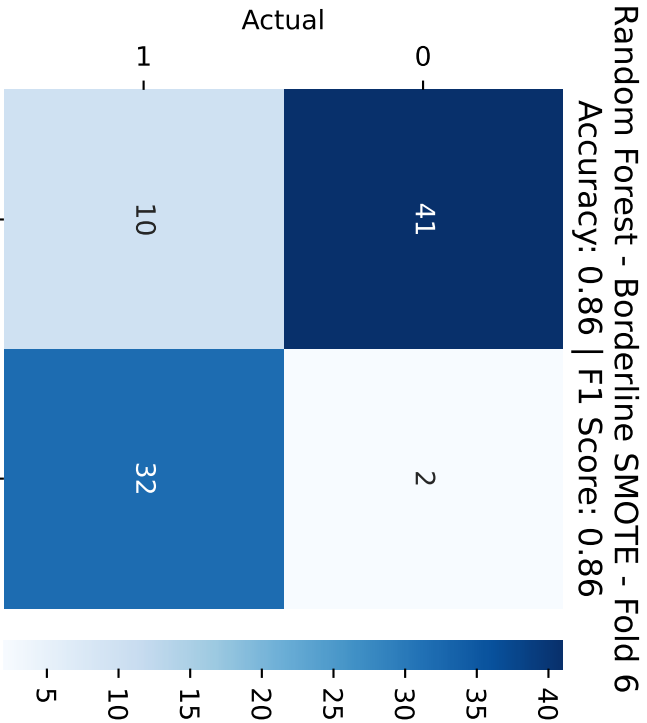
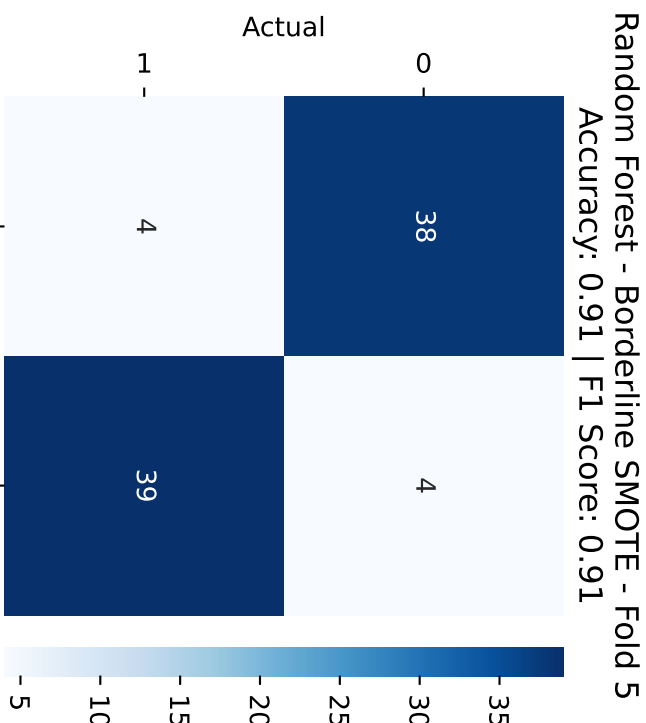
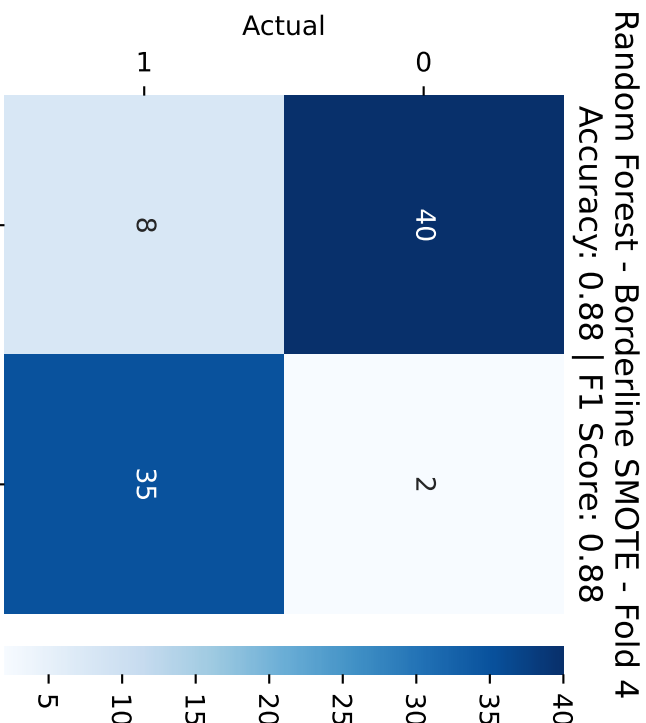
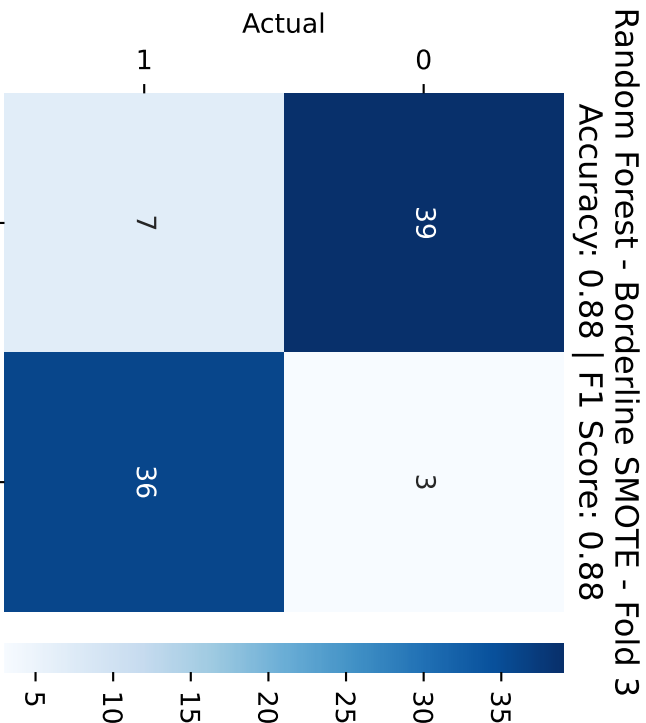
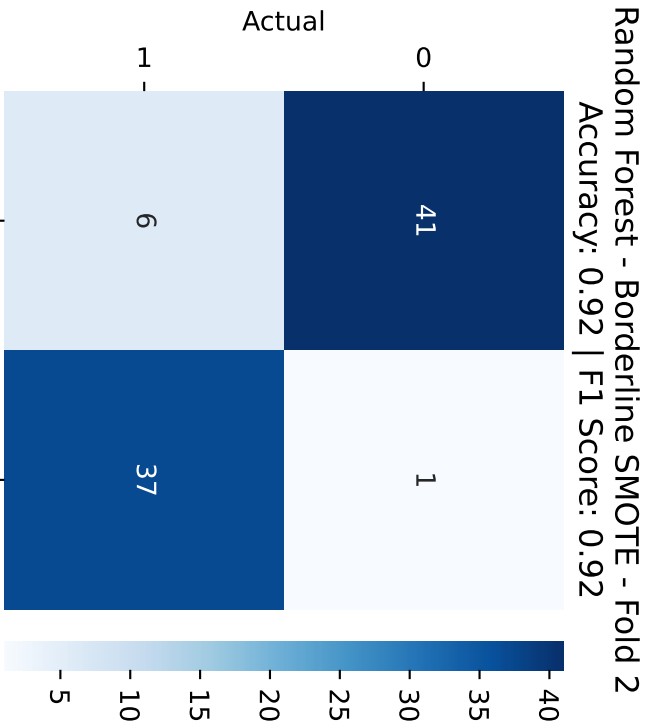
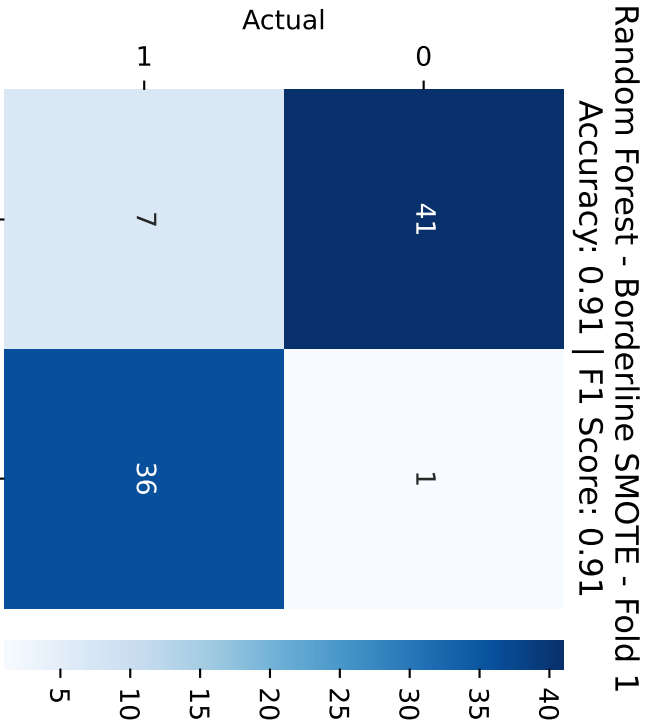
SVM - ROS - Fold 9  
Accuracy: 0.71 | F1 Score: 0.71

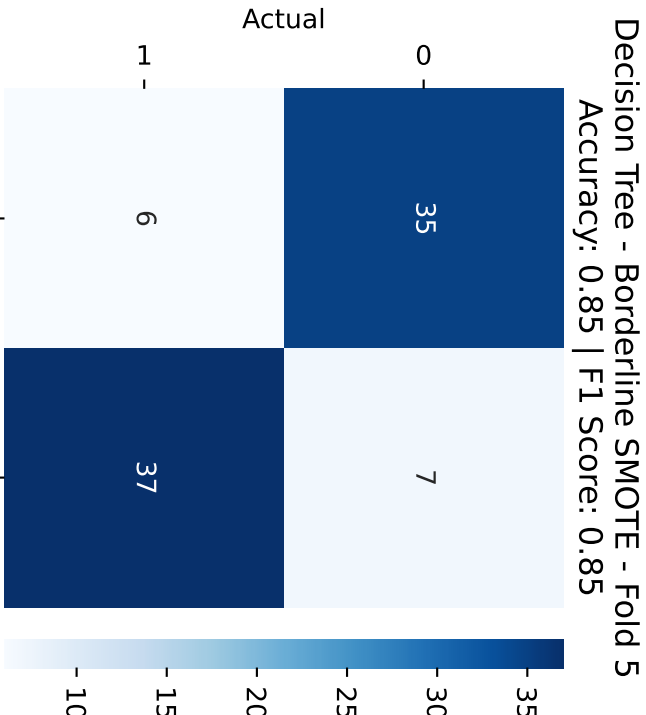
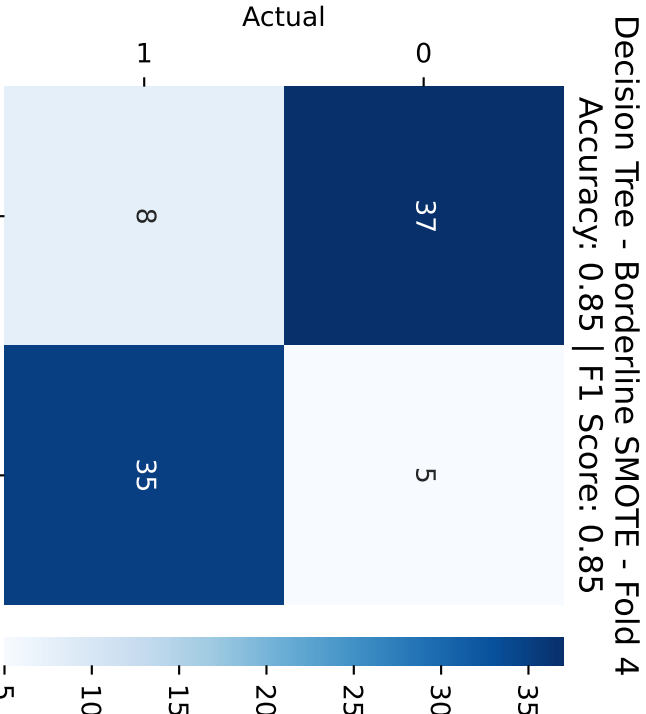
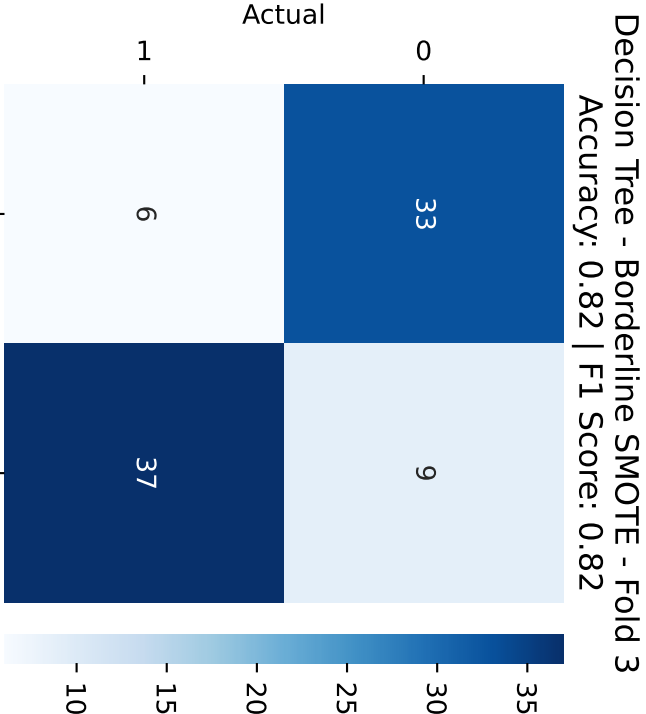
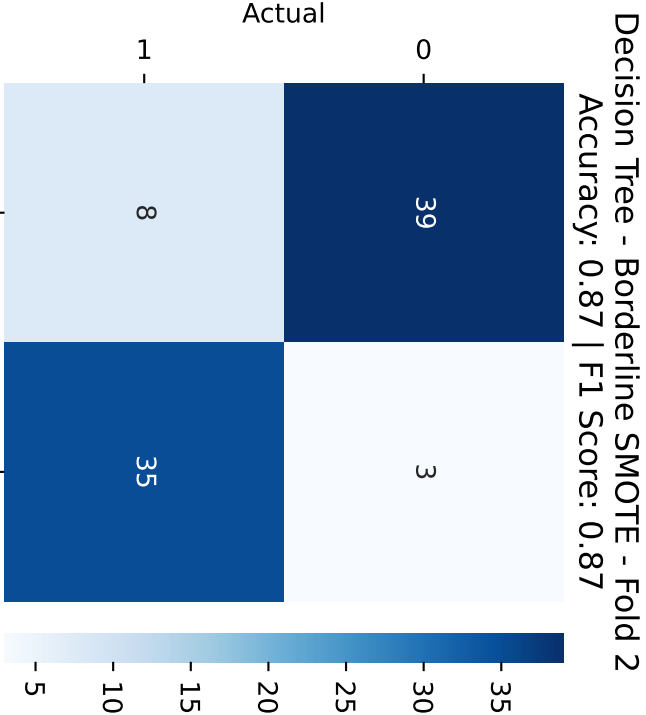
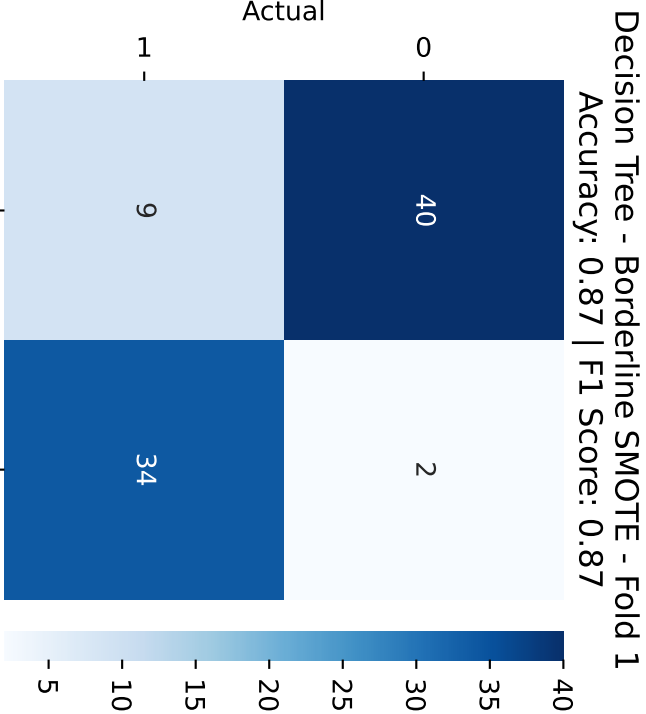


SVM - ROS - Fold 10  
Accuracy: 0.78 | F1 Score: 0.77

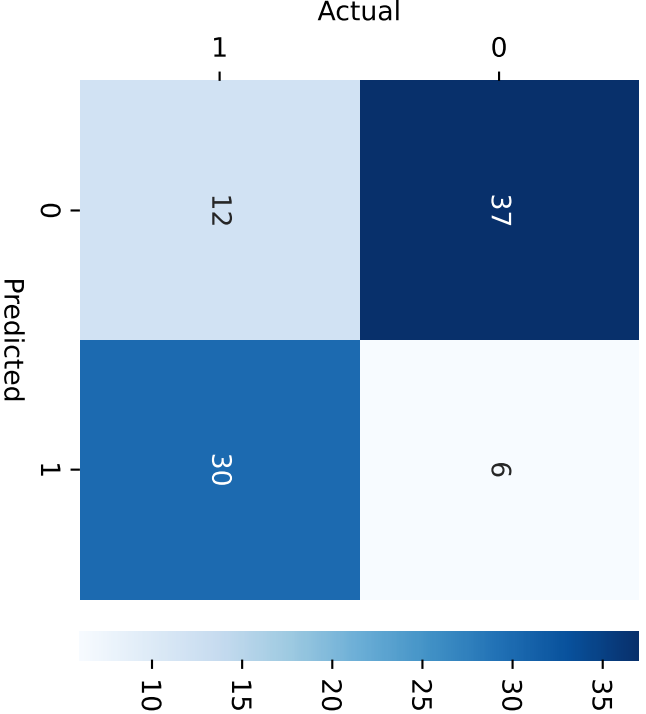




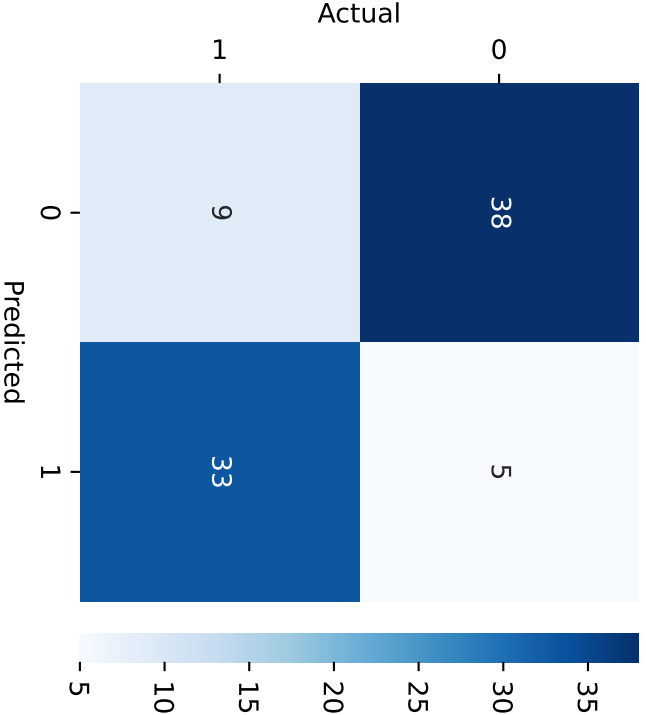




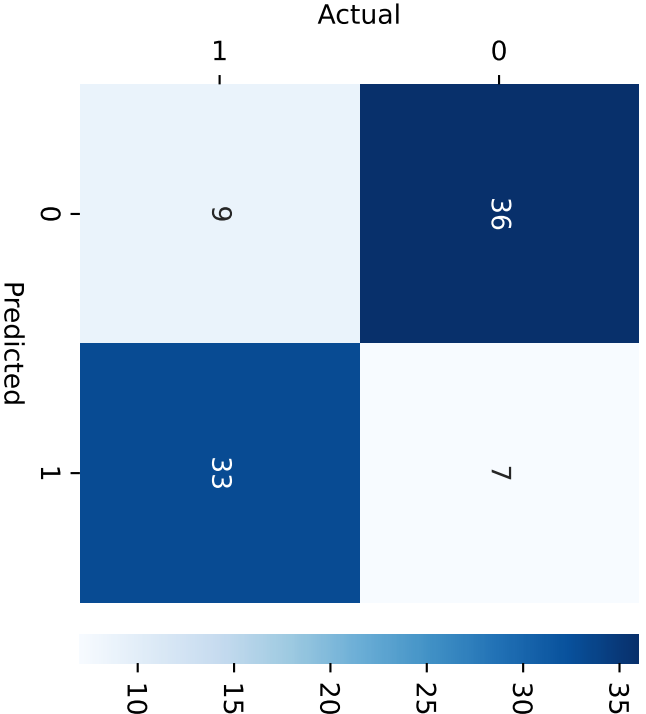
Decision Tree - Borderline SMOTE - Fold 6  
Accuracy: 0.79 | F1 Score: 0.79



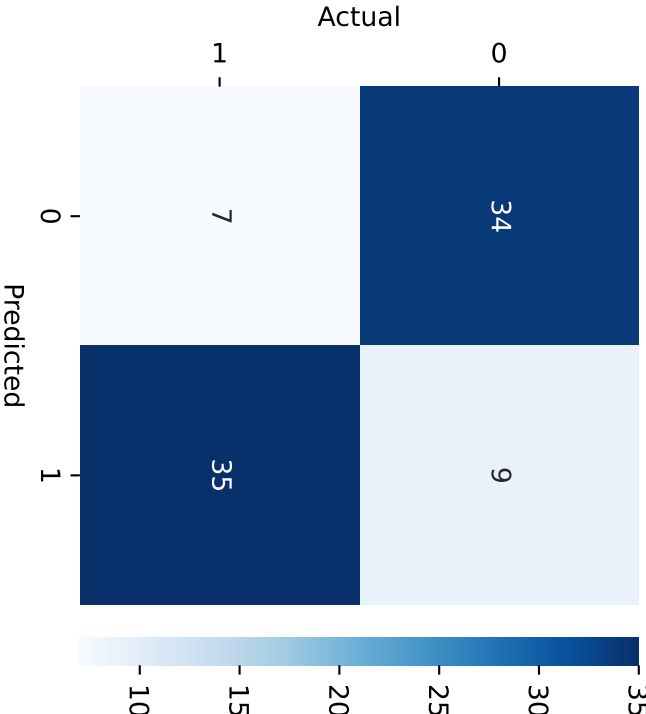
Decision Tree - Borderline SMOTE - Fold 7  
Accuracy: 0.84 | F1 Score: 0.83



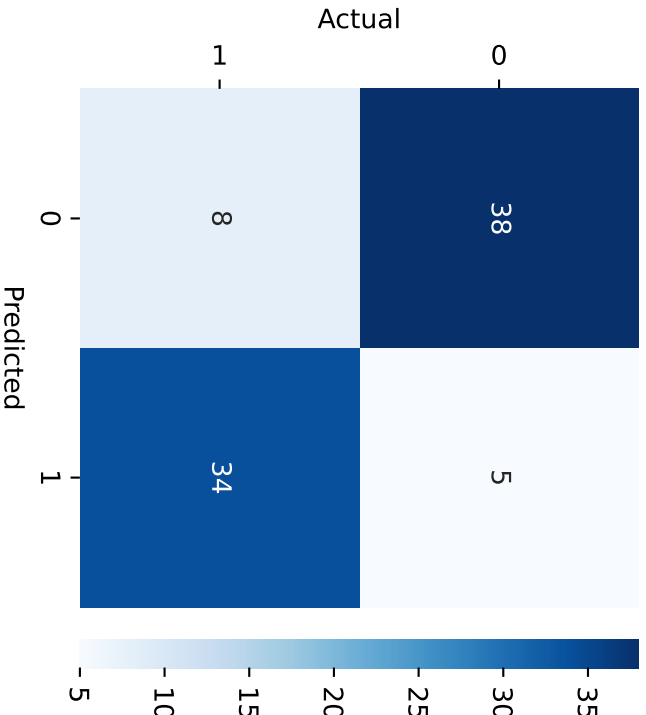
Decision Tree - Borderline SMOTE - Fold 8  
Accuracy: 0.81 | F1 Score: 0.81

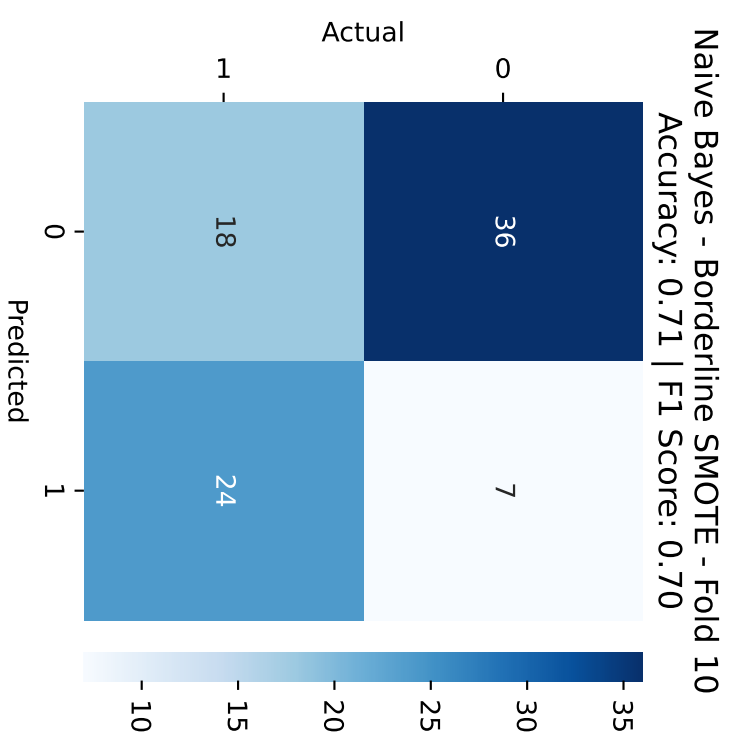
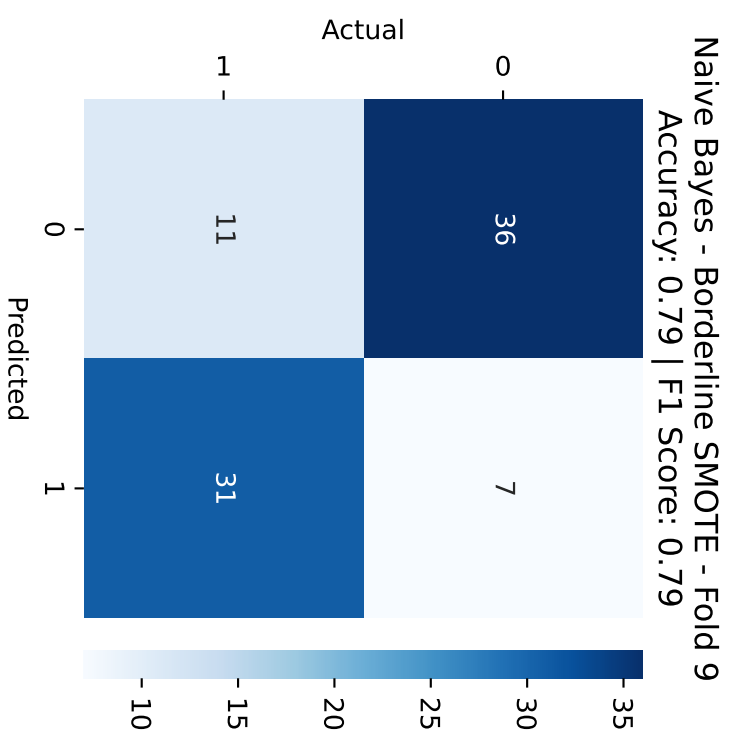
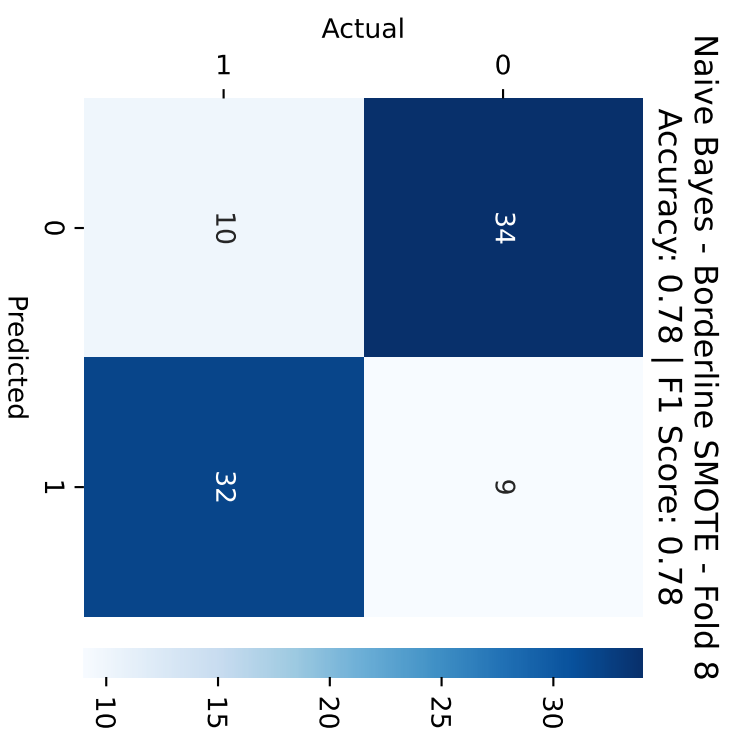
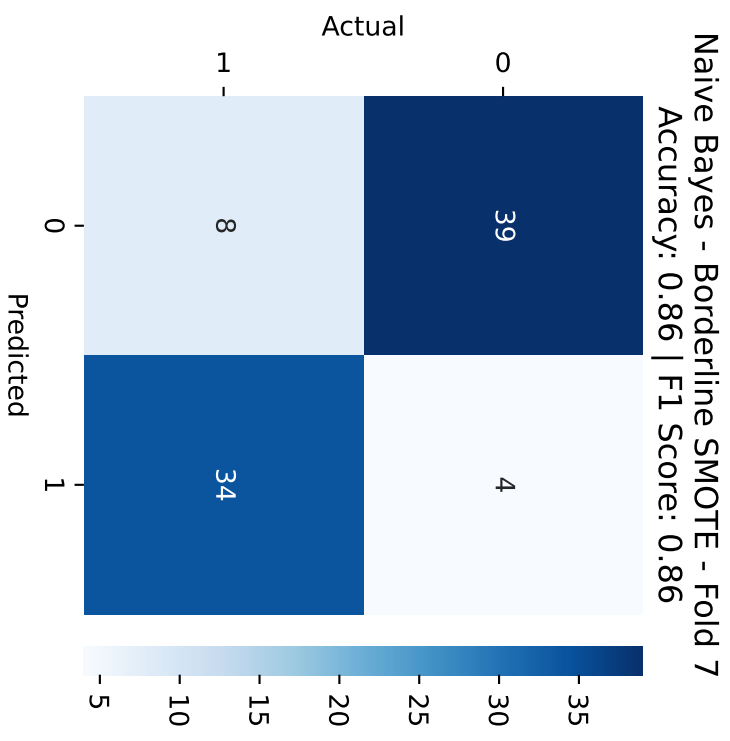
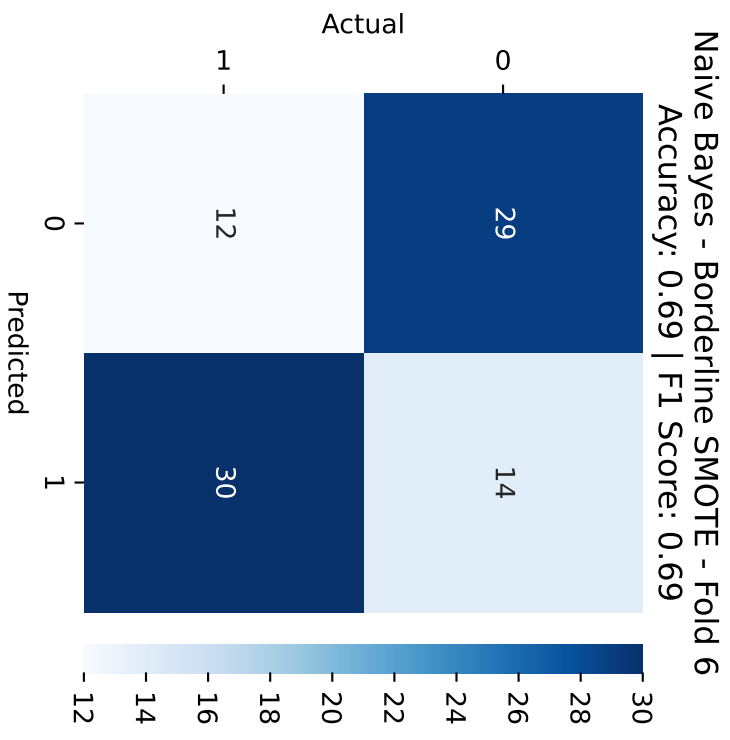
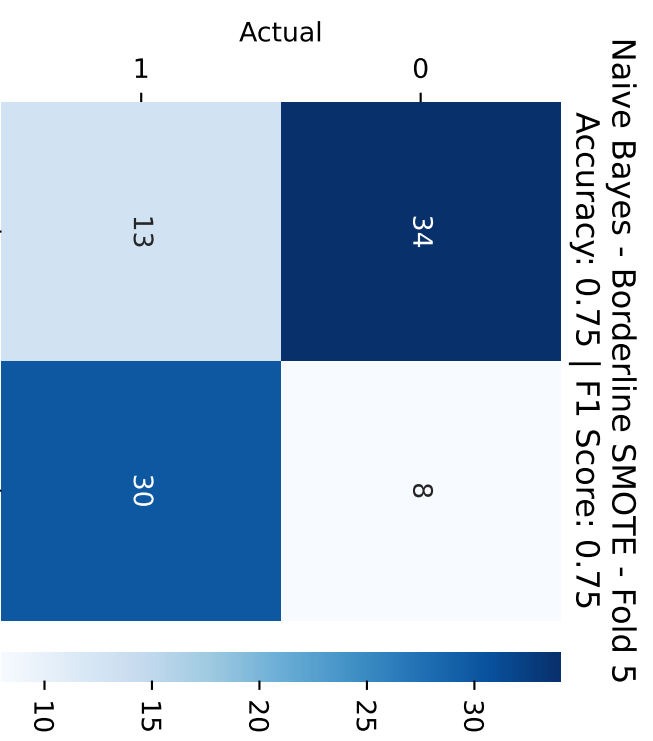
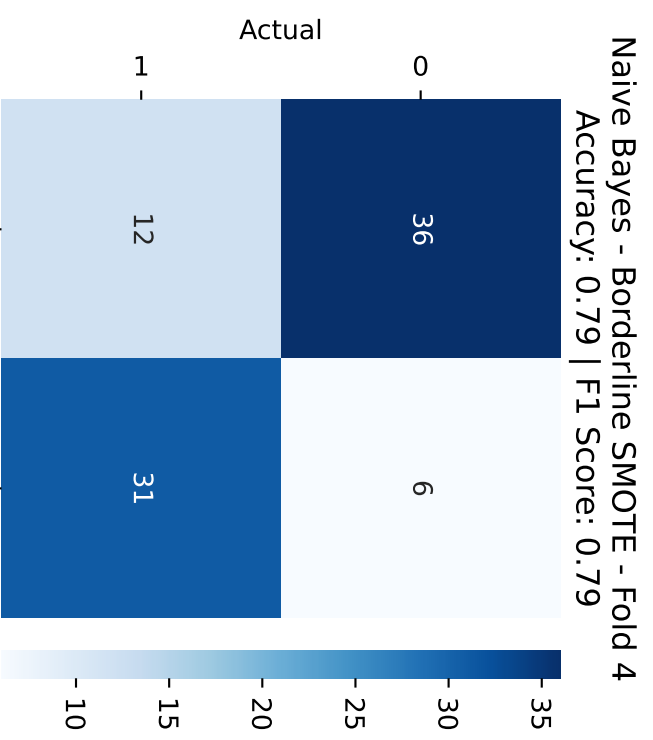
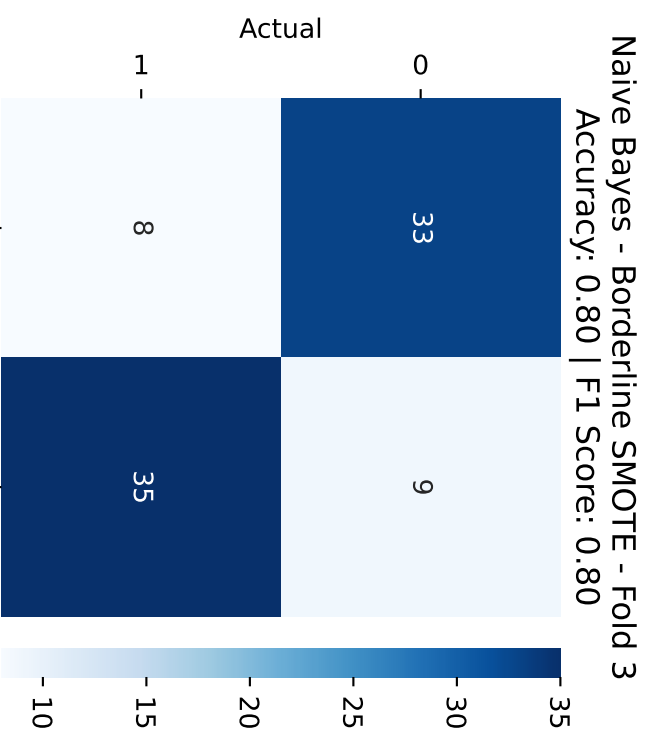
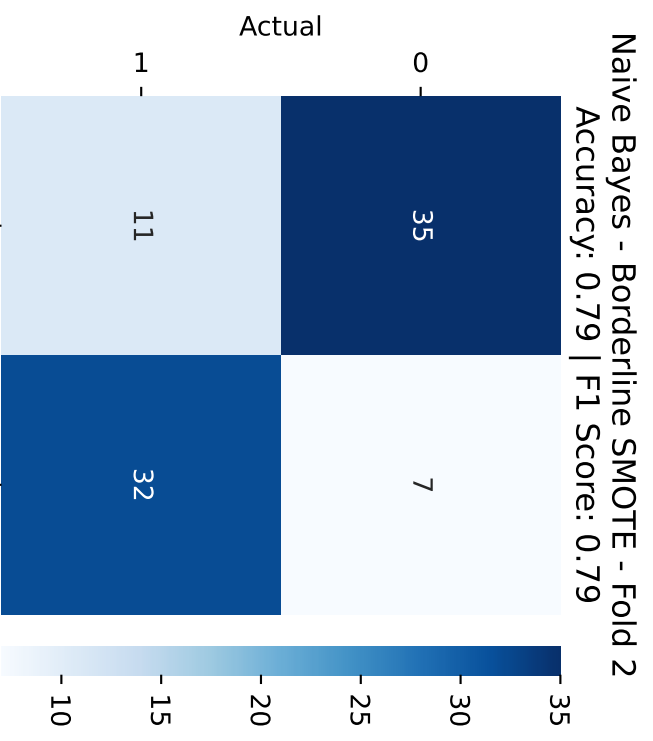
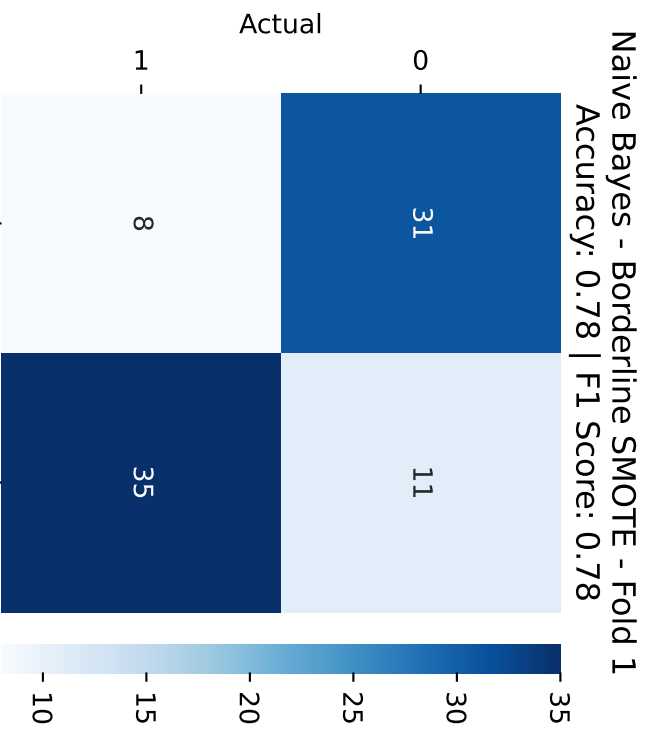


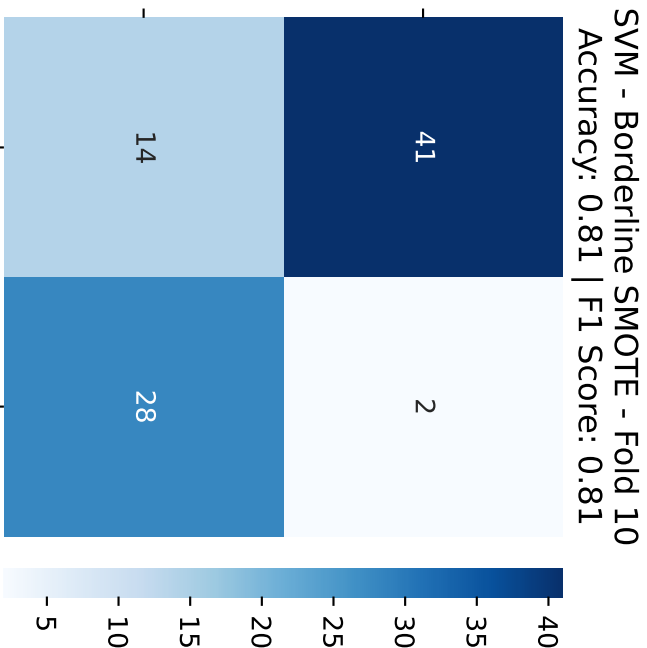
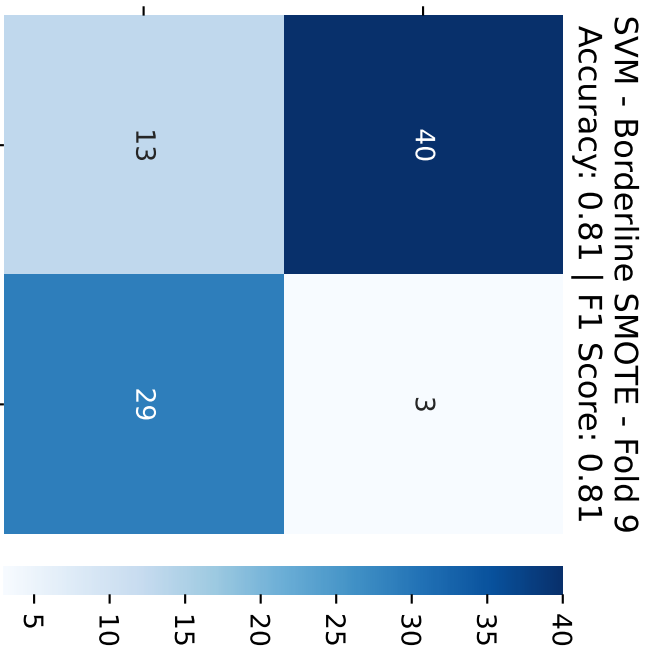
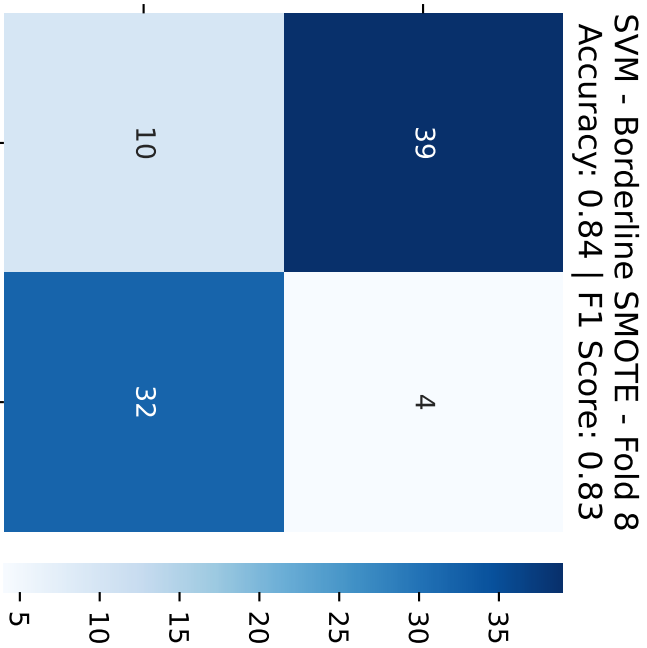
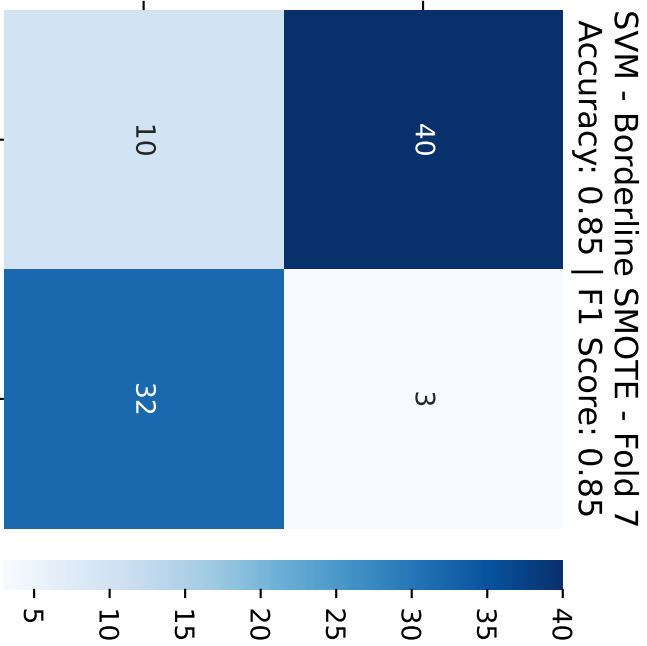
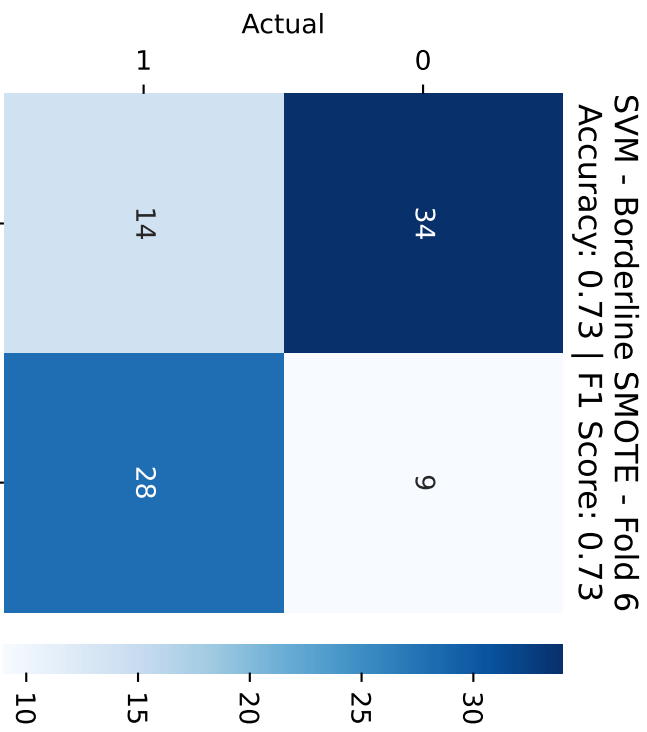
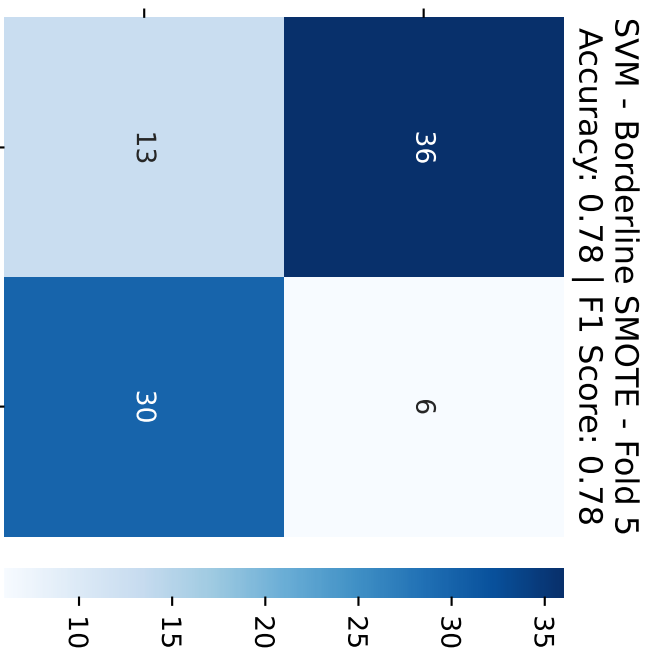
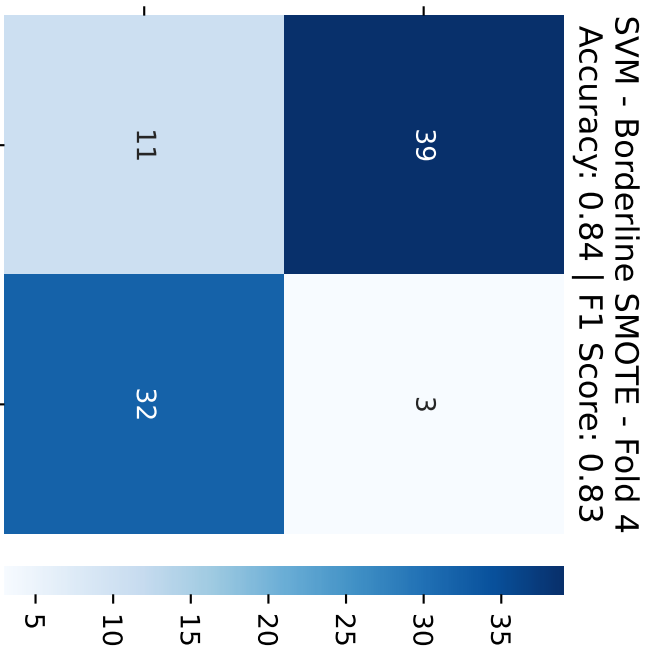
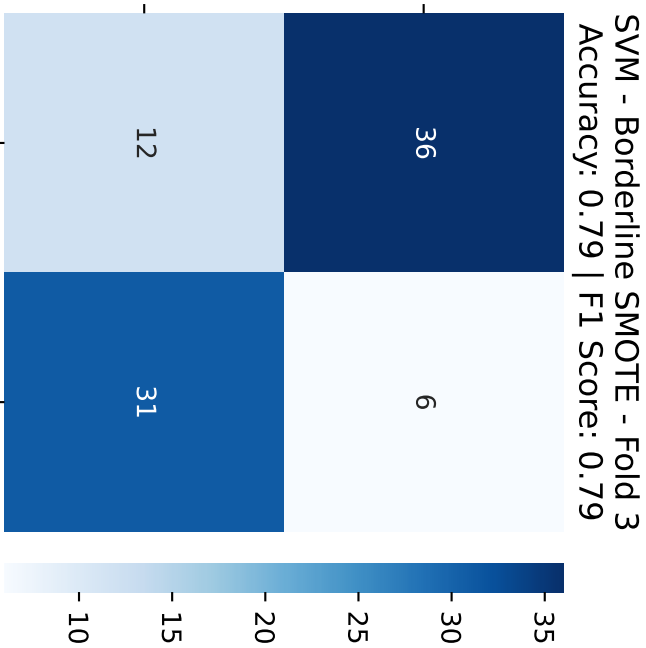
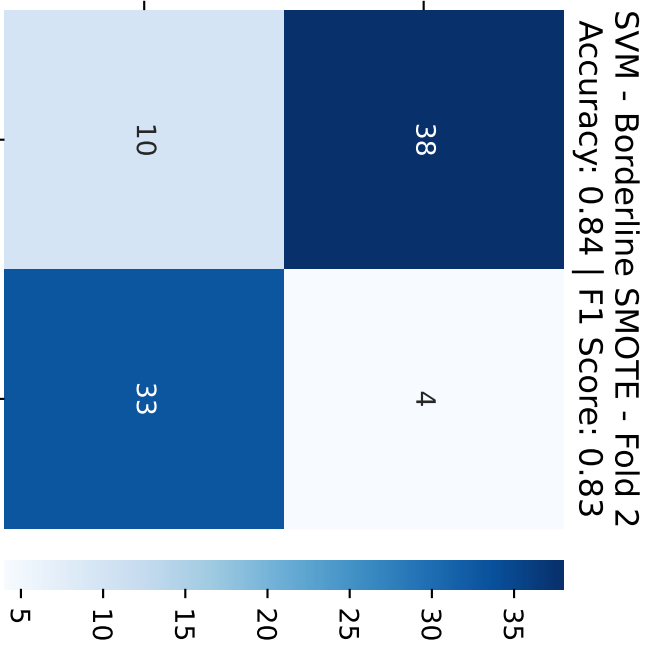
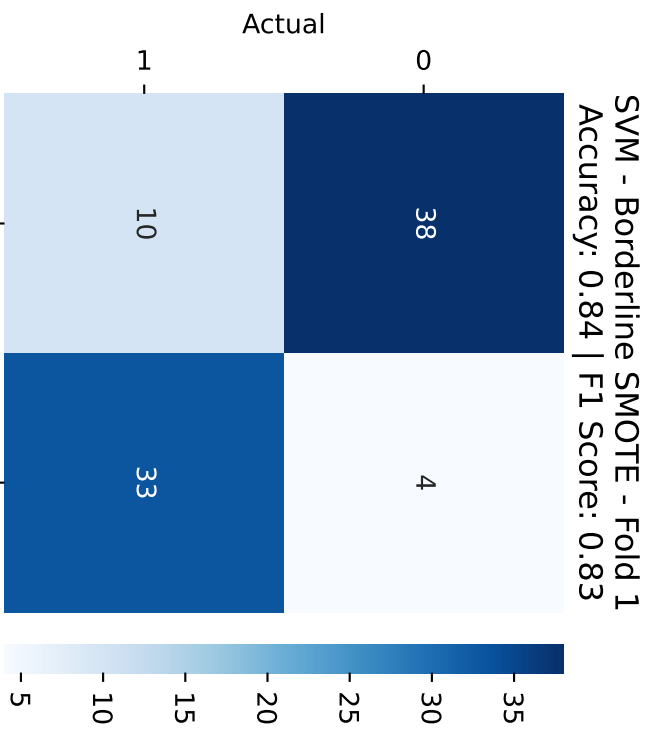
Decision Tree - Borderline SMOTE - Fold 9  
Accuracy: 0.81 | F1 Score: 0.81

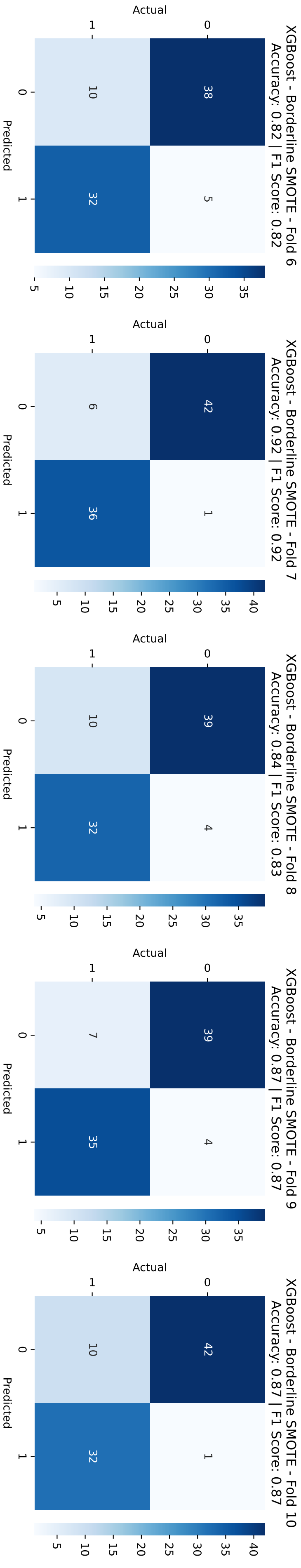
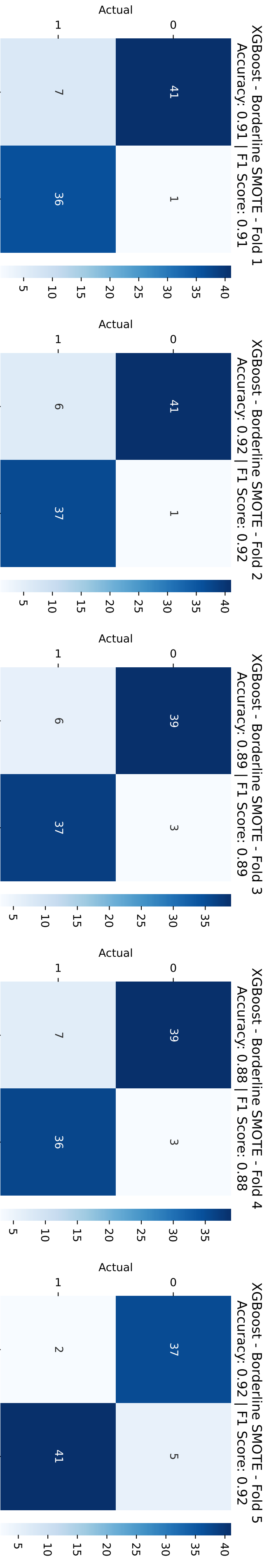


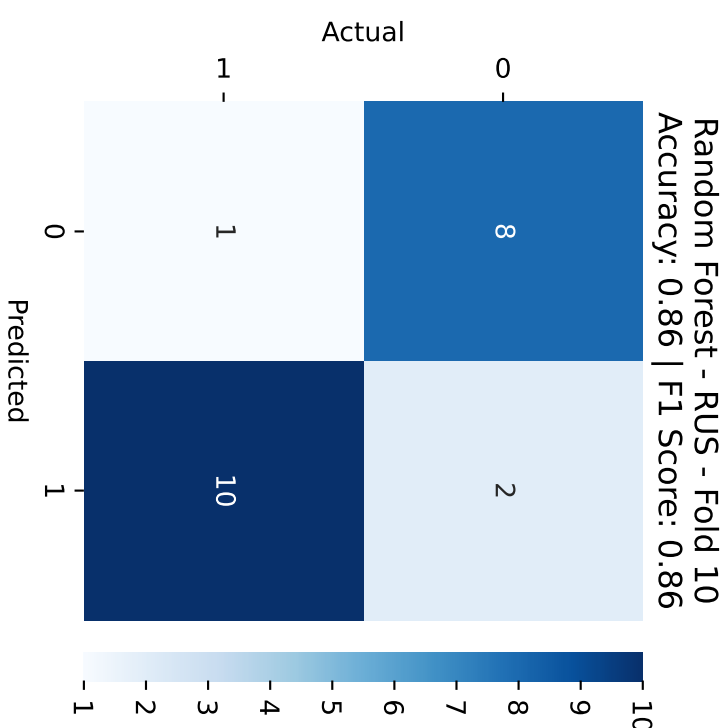
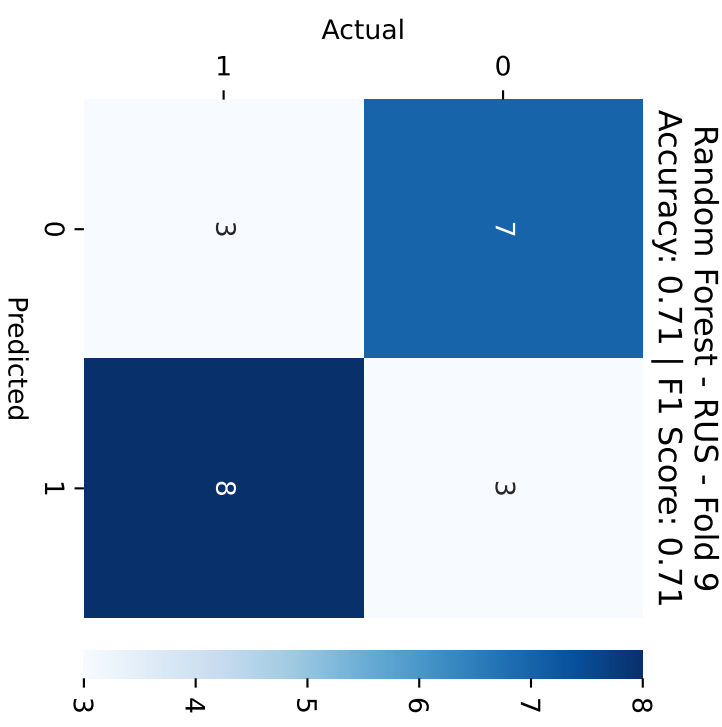
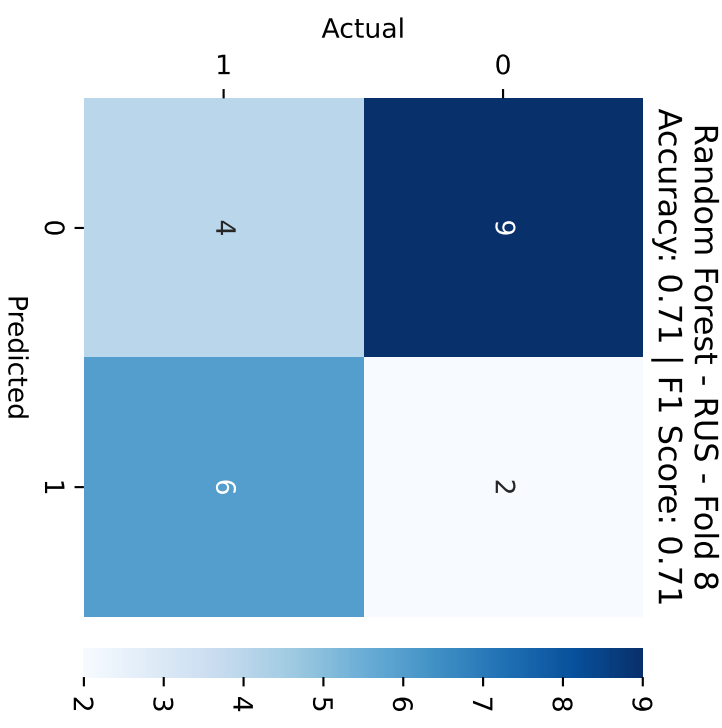
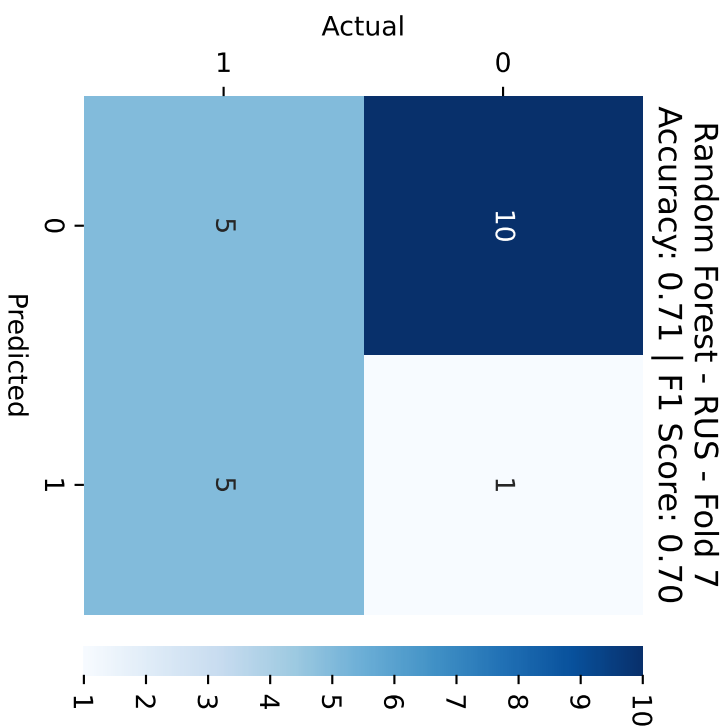
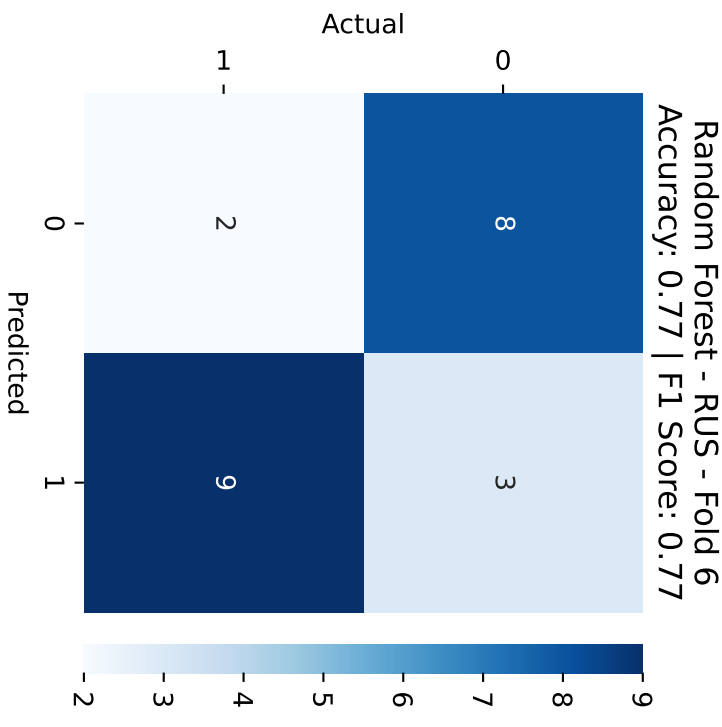
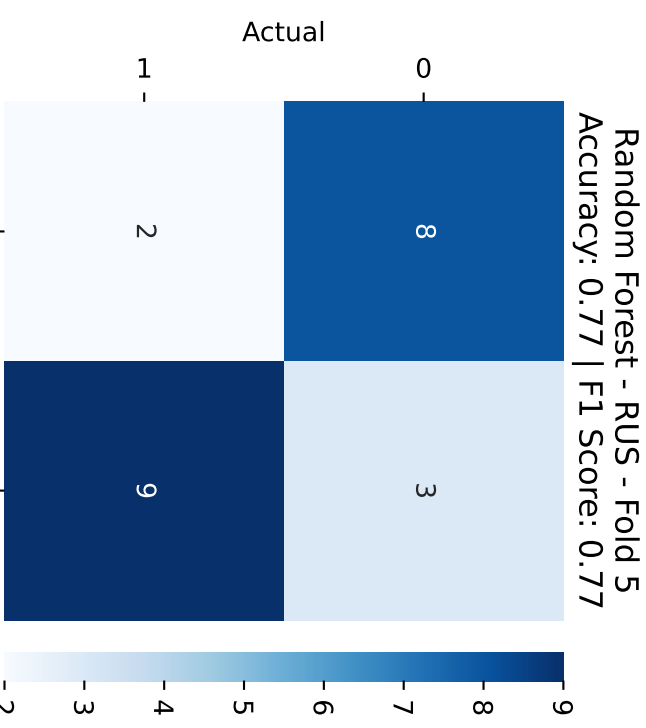
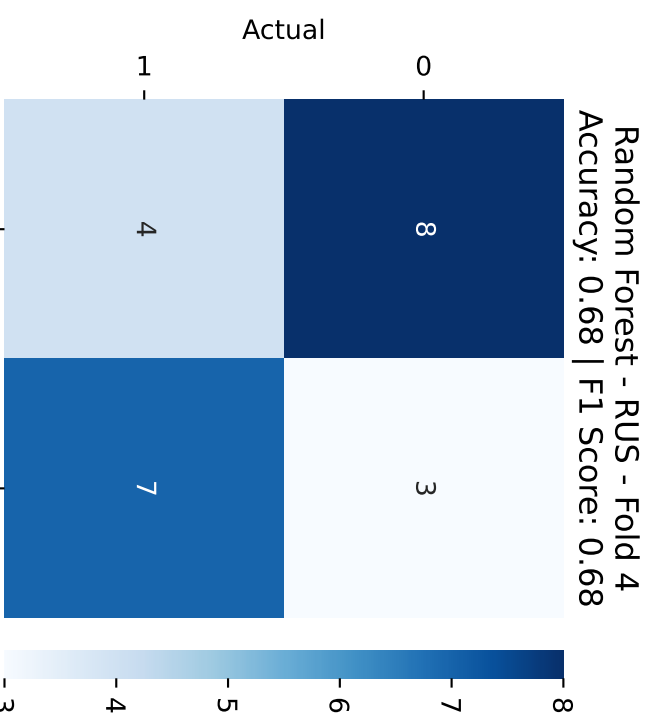
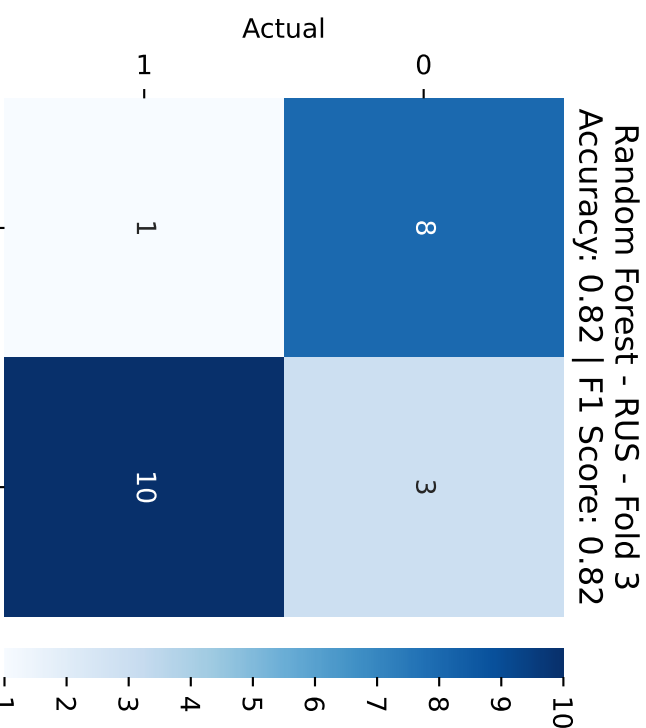
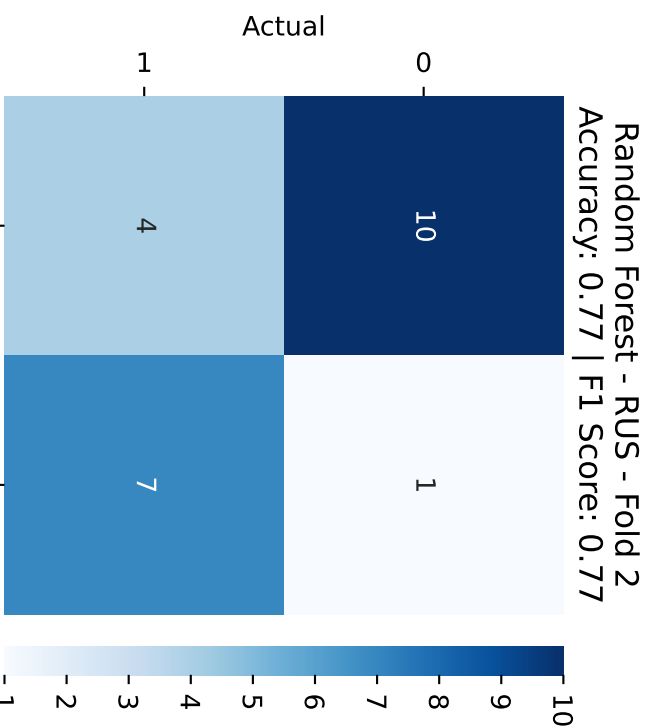
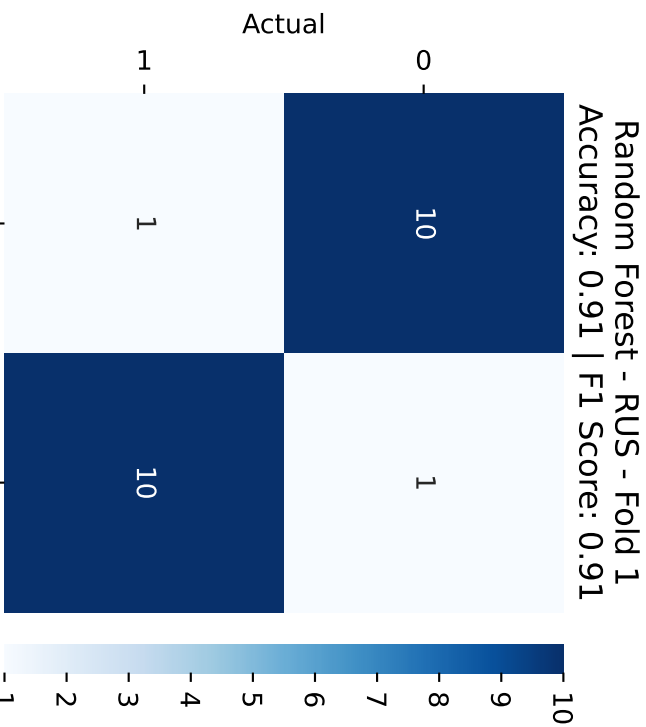
Decision Tree - Borderline SMOTE - Fold 10  
Accuracy: 0.85 | F1 Score: 0.85

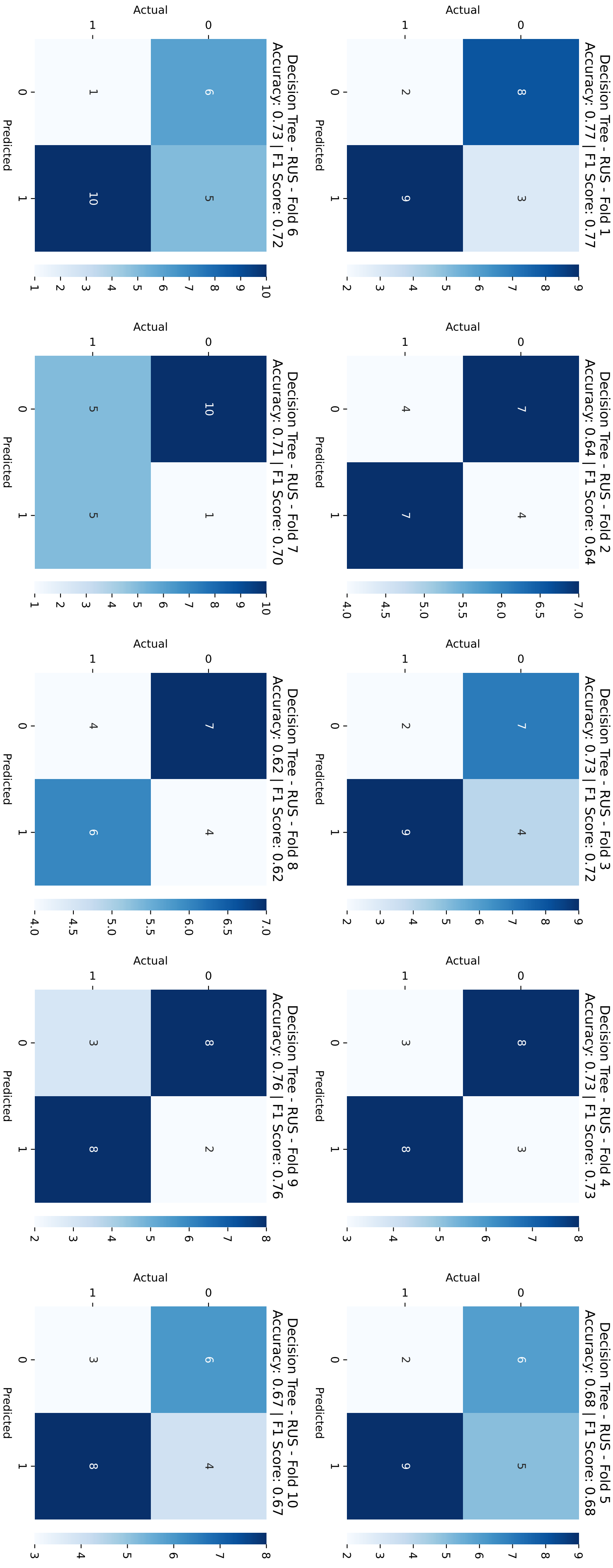


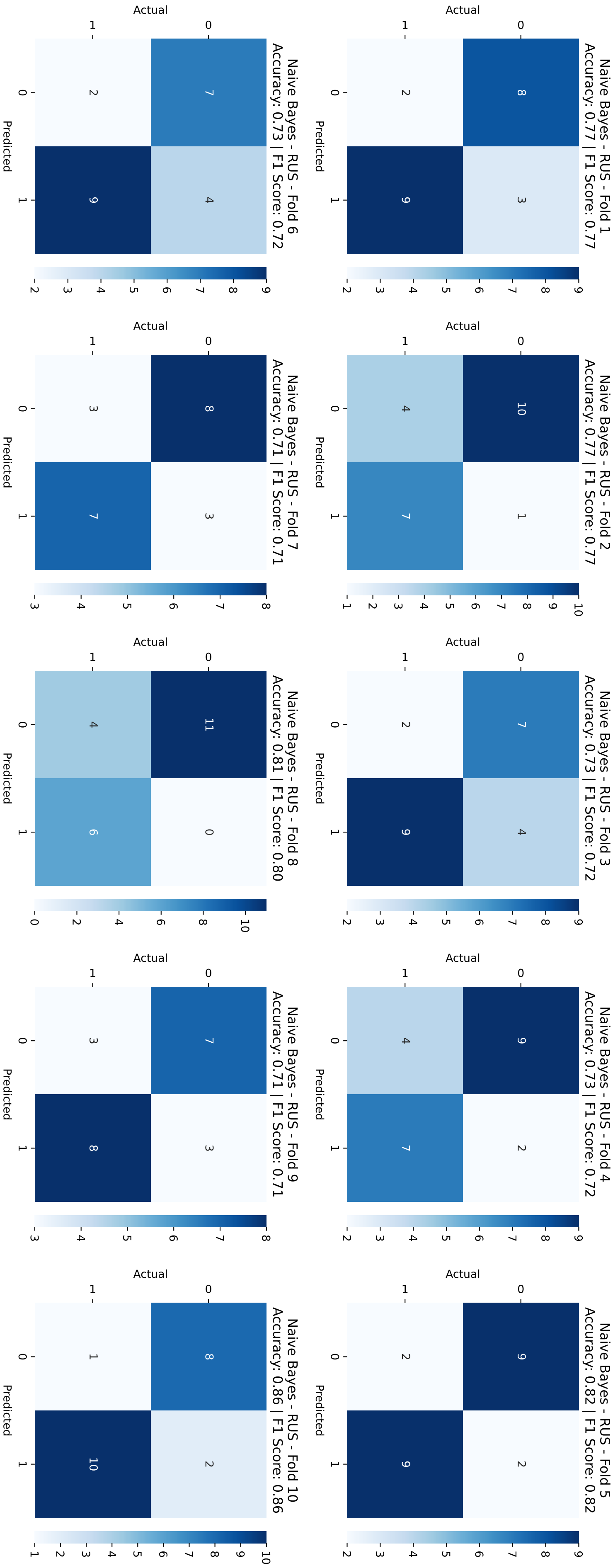


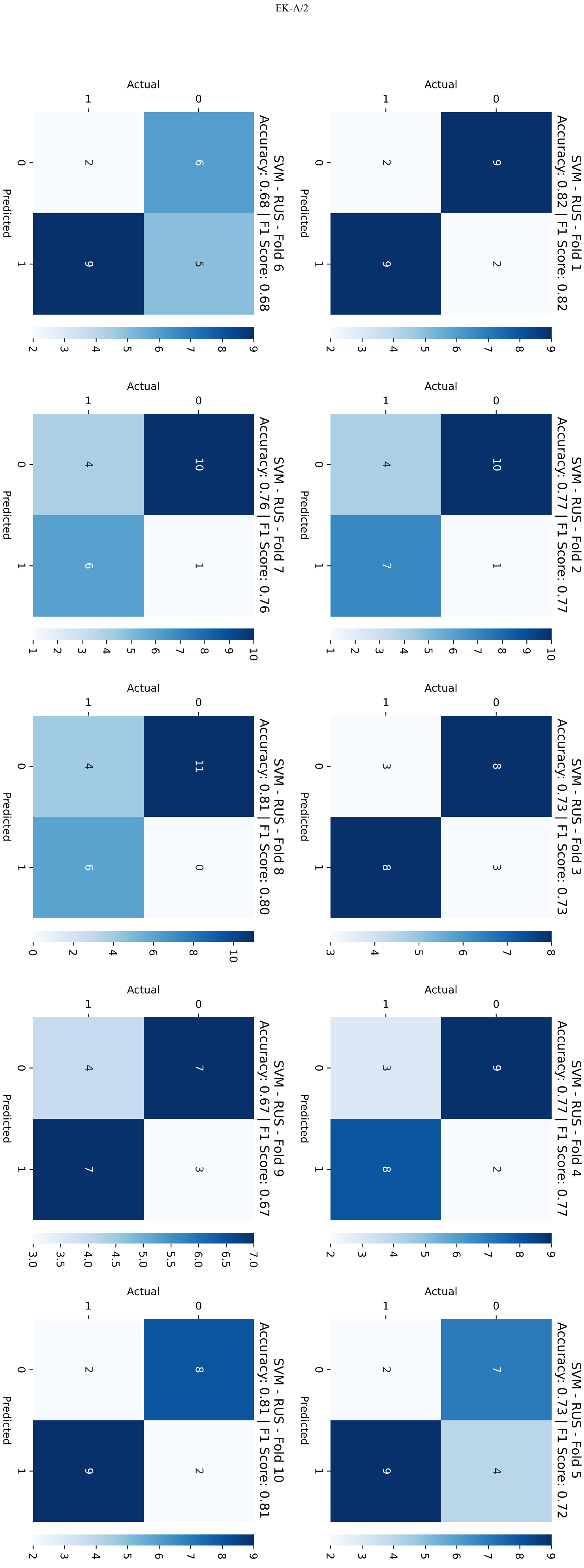


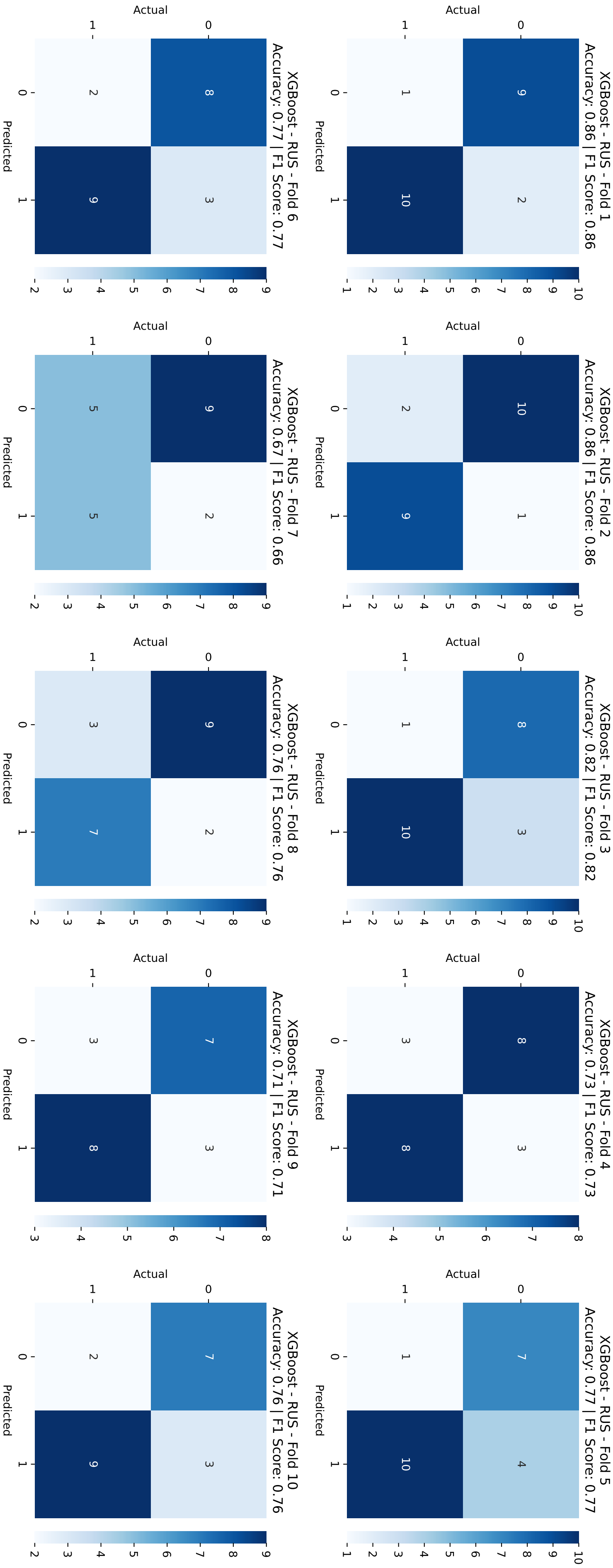


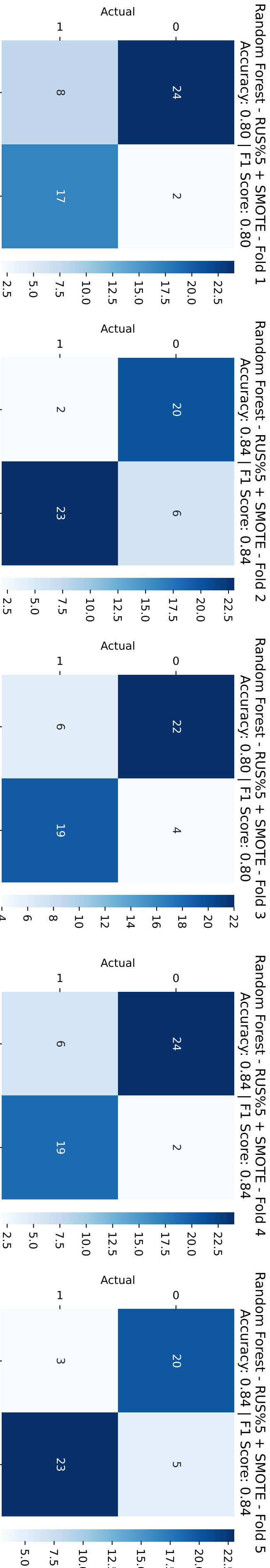


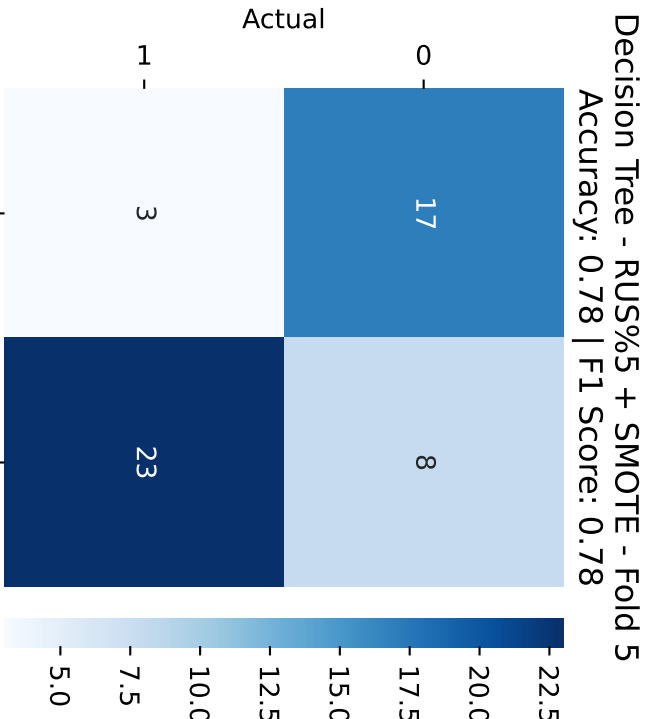
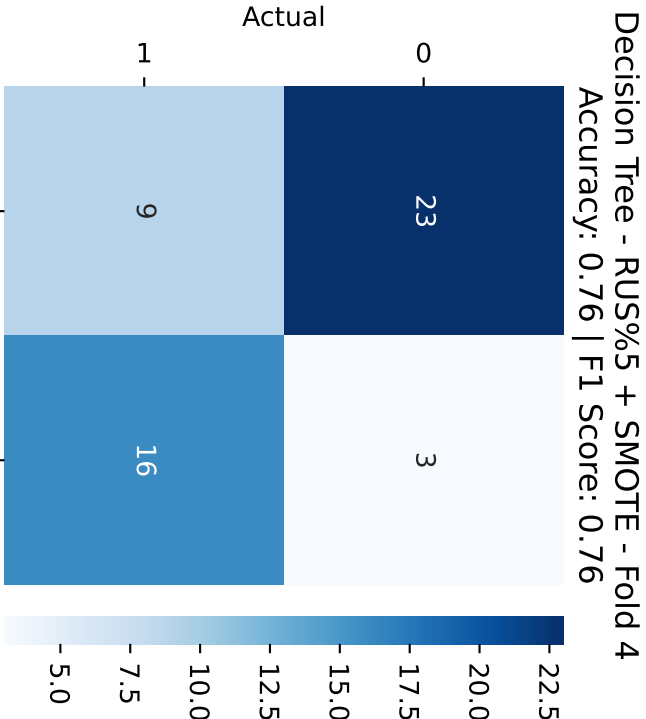
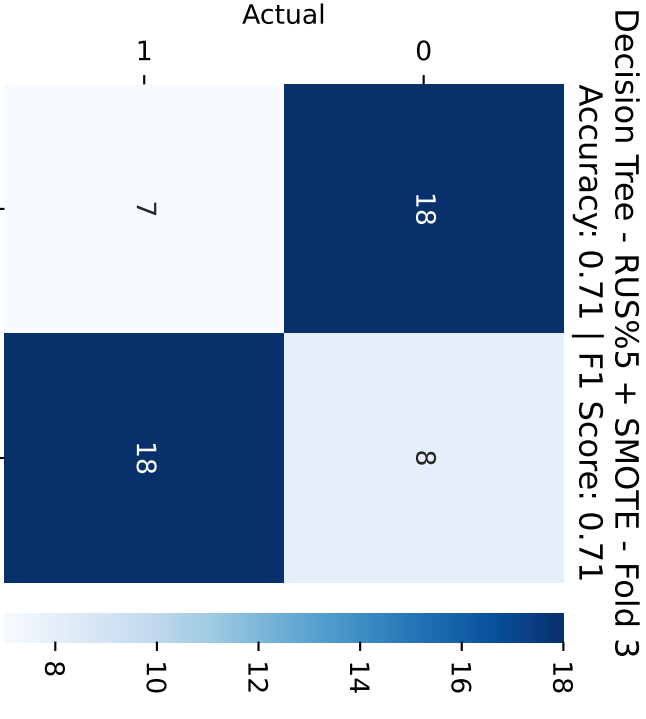
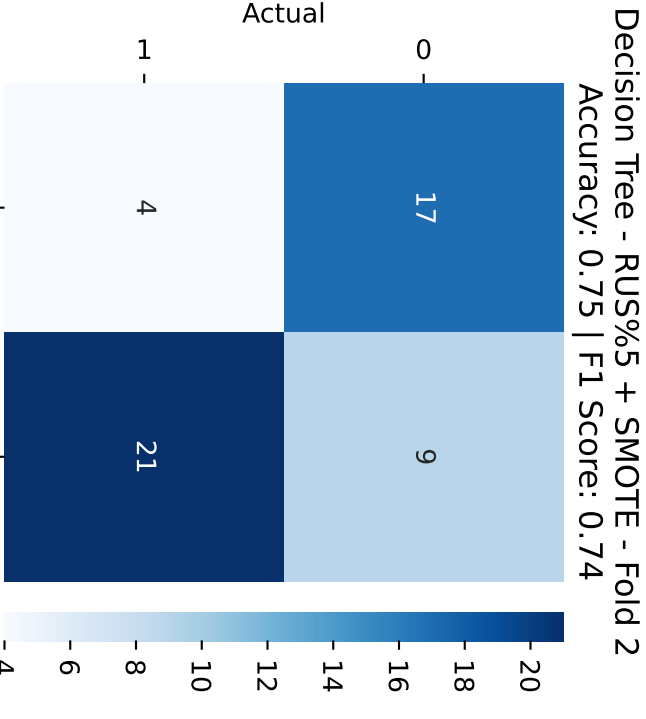
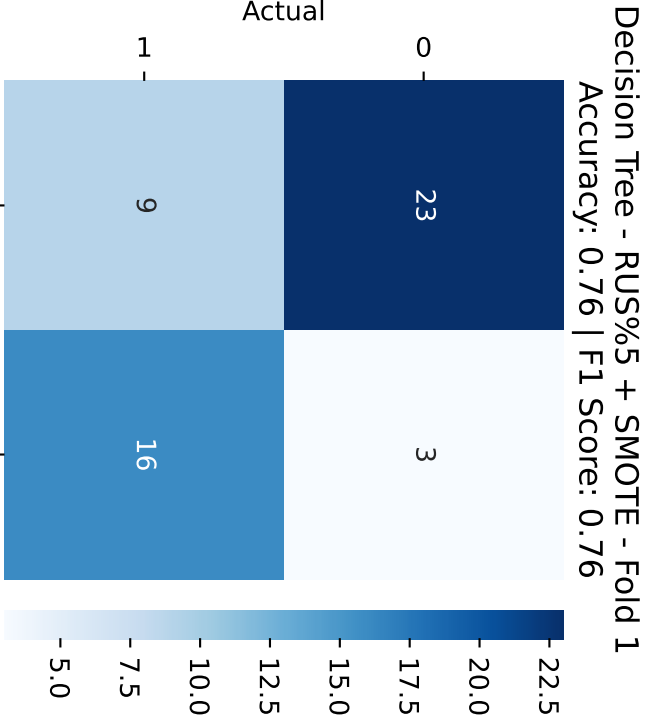




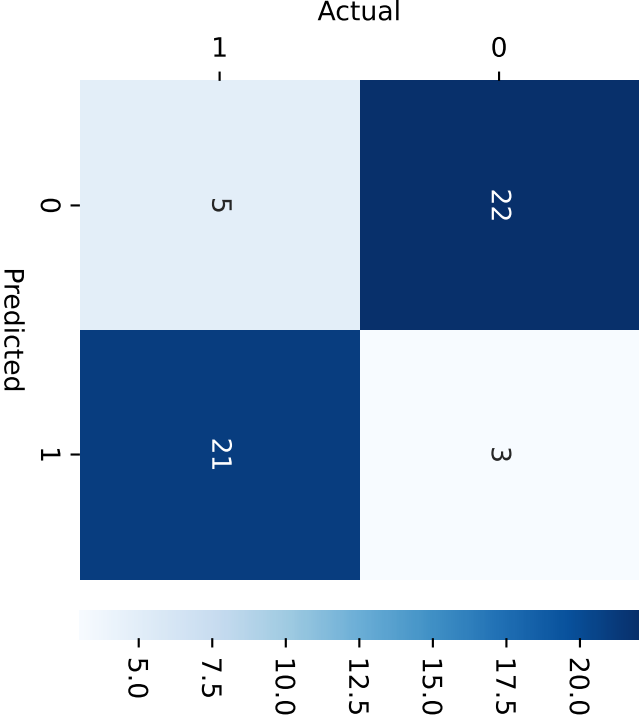




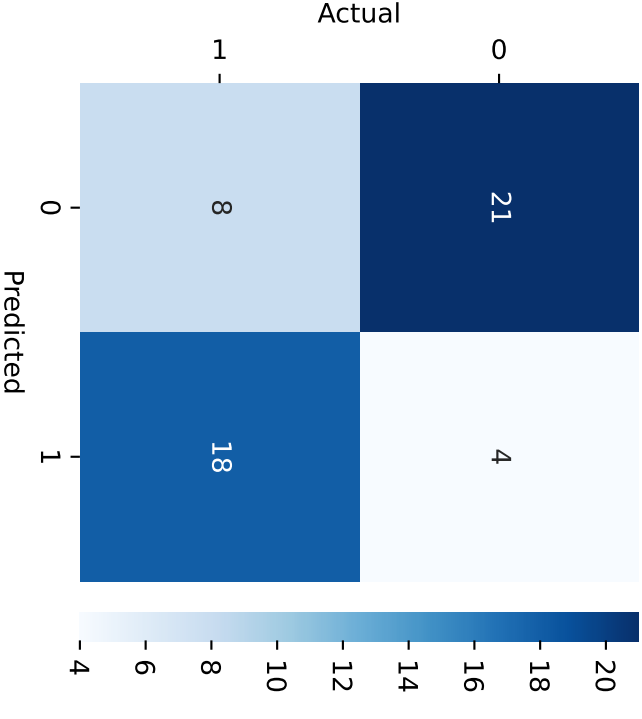




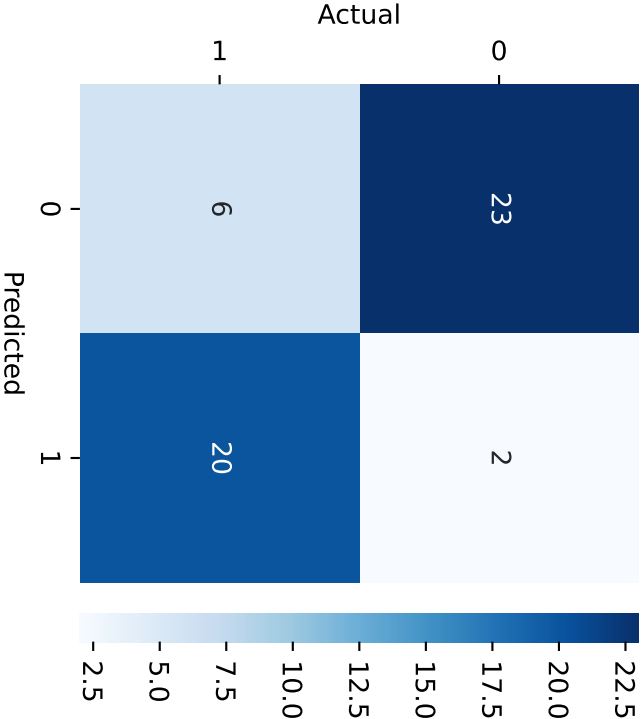
Decision Tree - RUS%5 + SMOTE - Fold 6  
Accuracy: 0.84 | F1 Score: 0.84



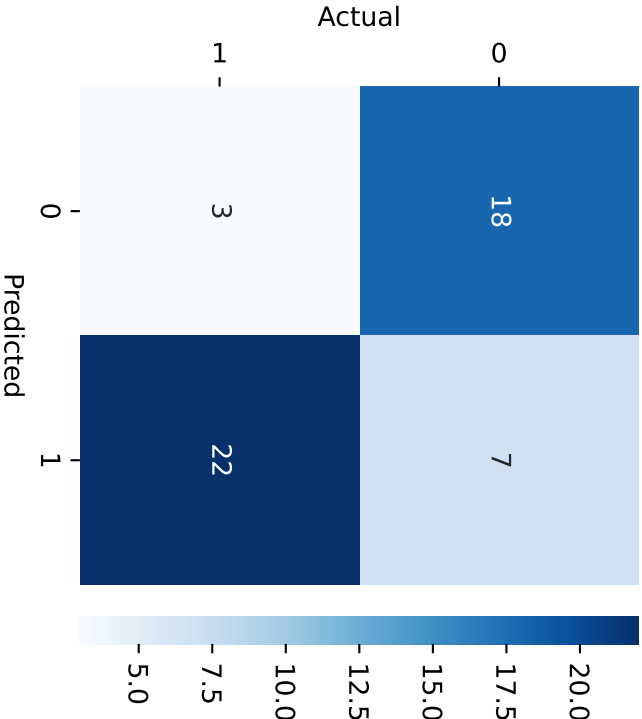
Decision Tree - RUS%5 + SMOTE - Fold 7  
Accuracy: 0.76 | F1 Score: 0.76



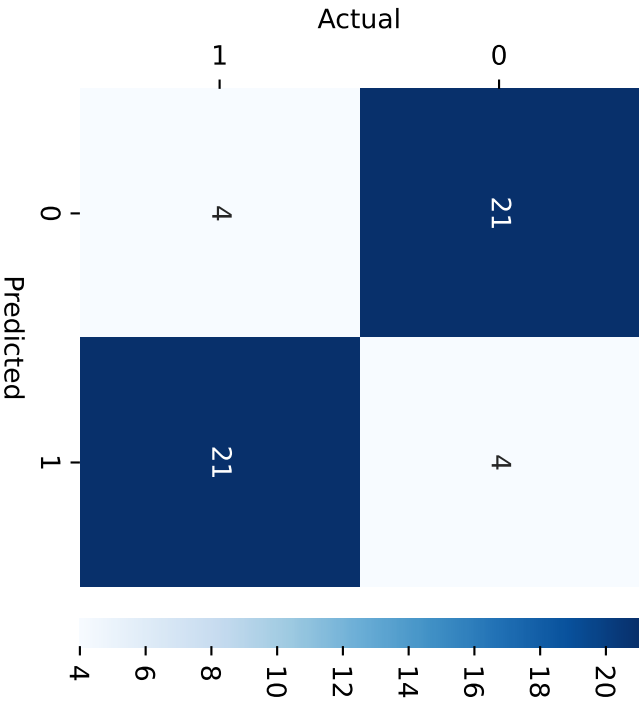
Decision Tree - RUS%5 + SMOTE - Fold 8  
Accuracy: 0.84 | F1 Score: 0.84

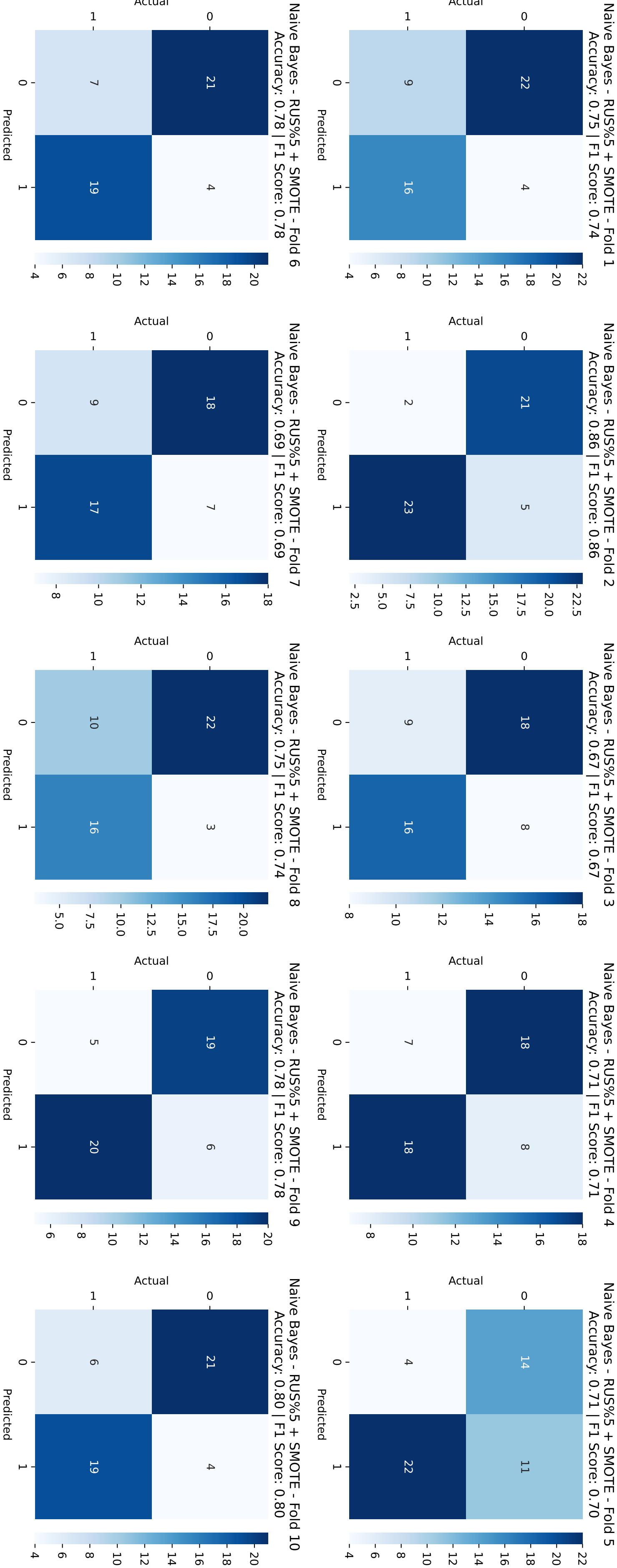


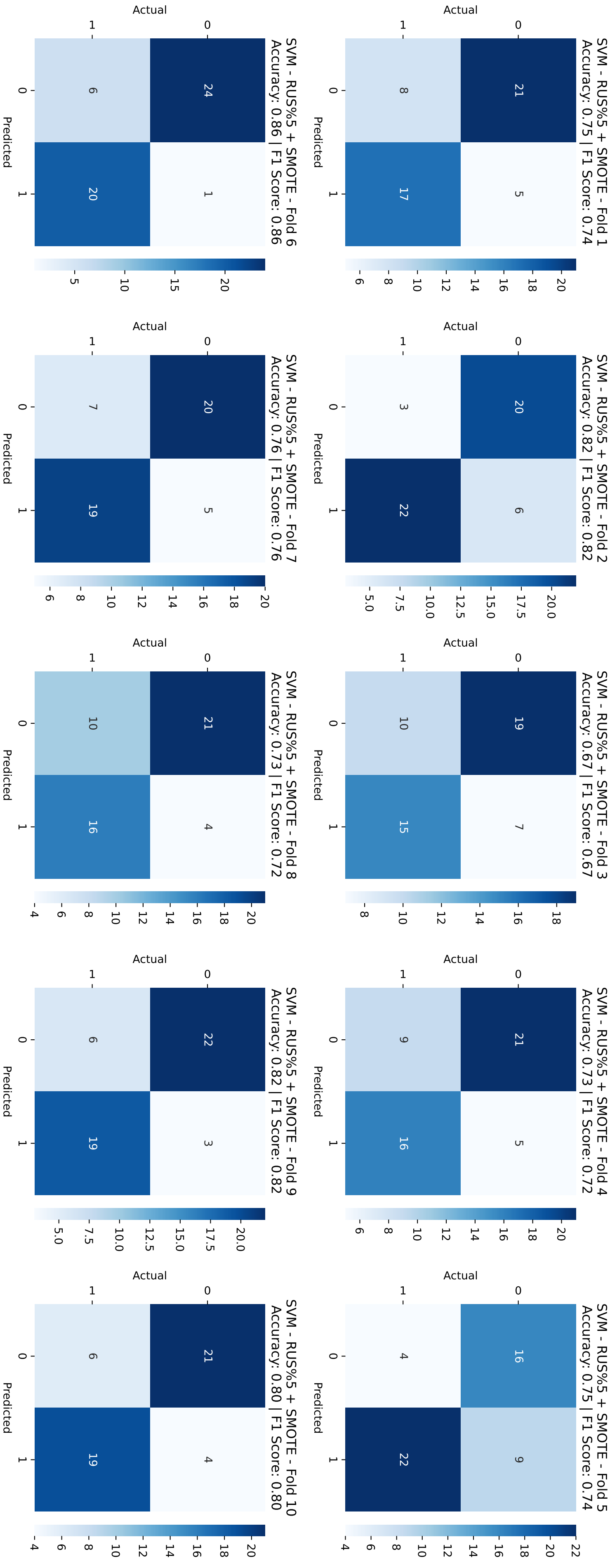
Decision Tree - RUS%5 + SMOTE - Fold 9  
Accuracy: 0.80 | F1 Score: 0.80

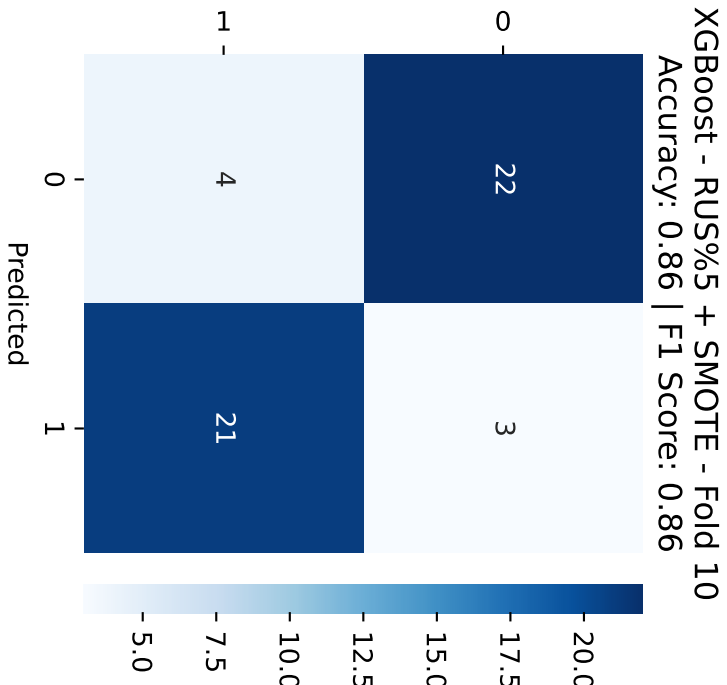
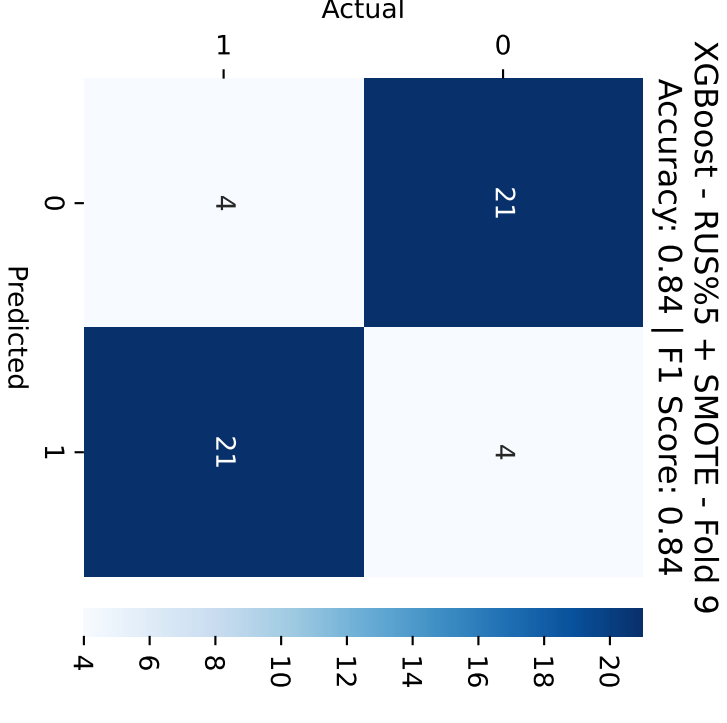
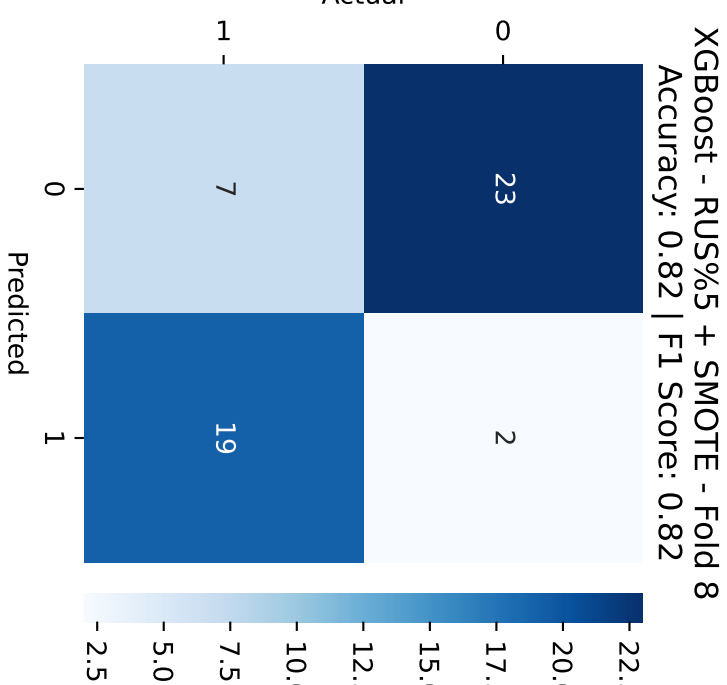
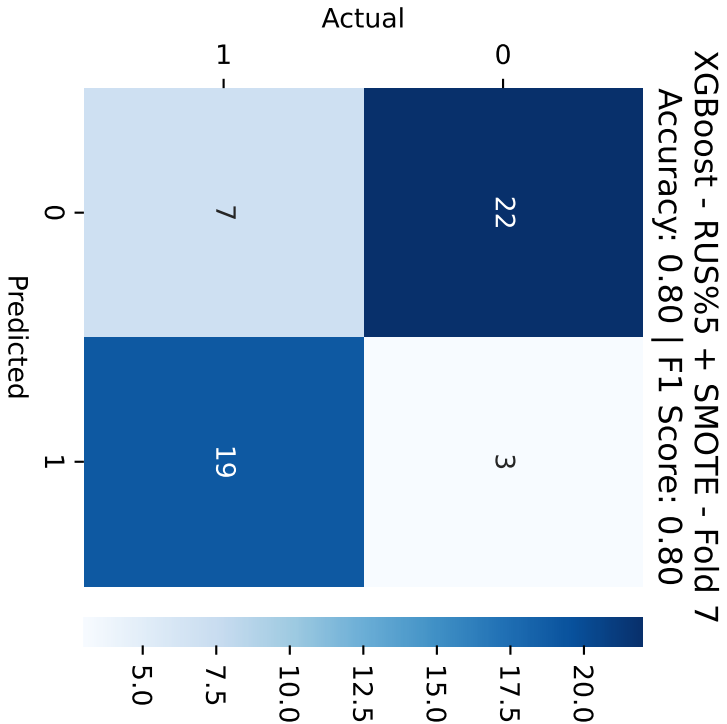
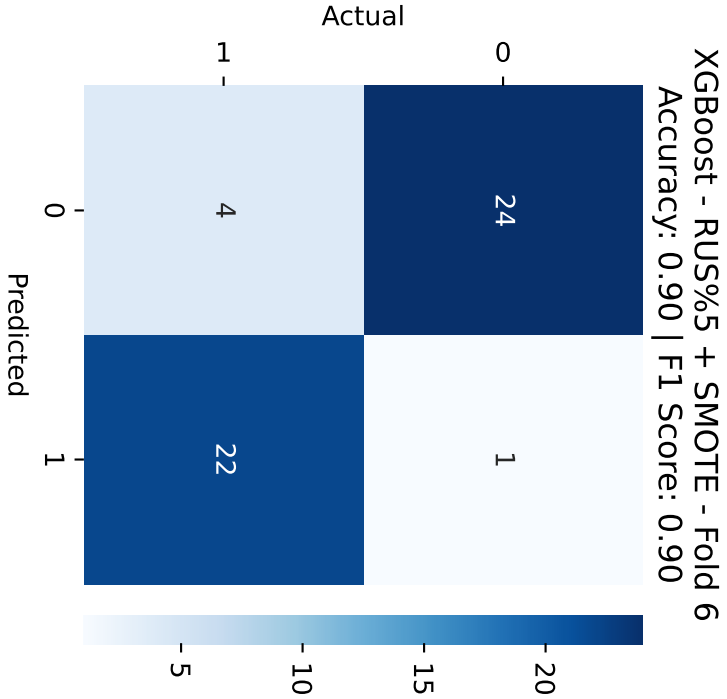
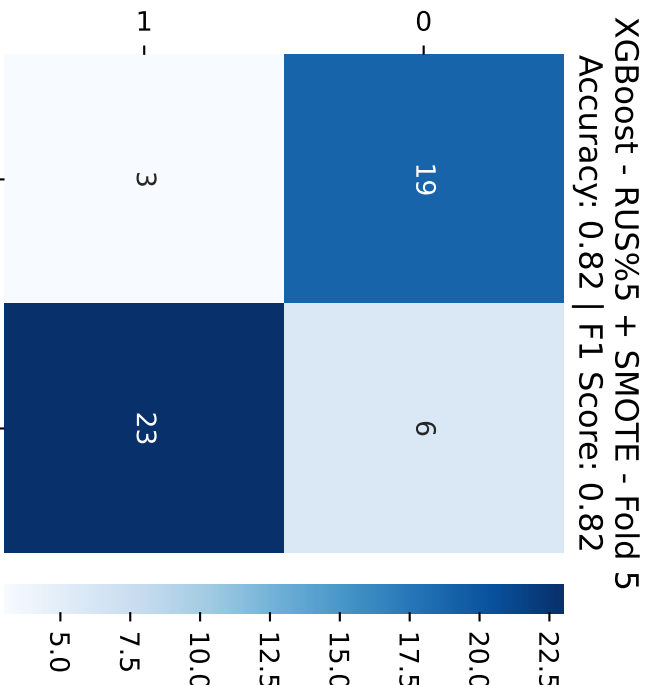
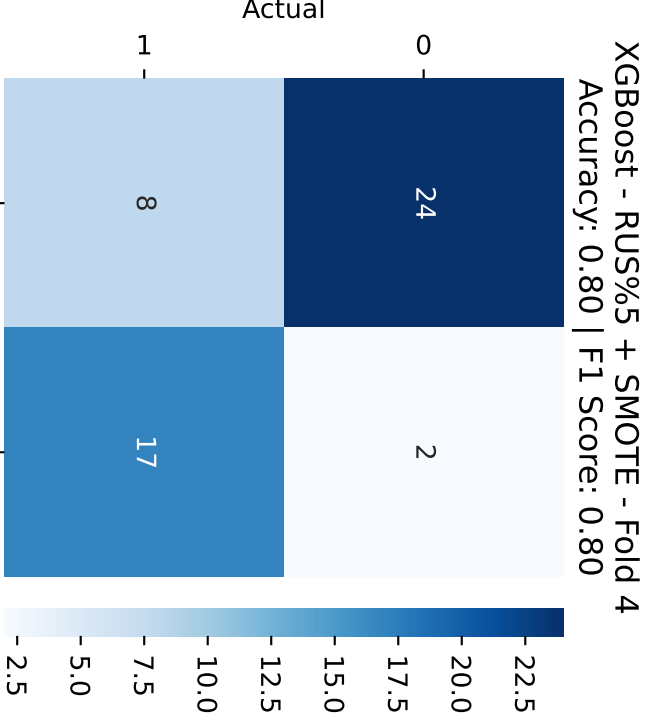
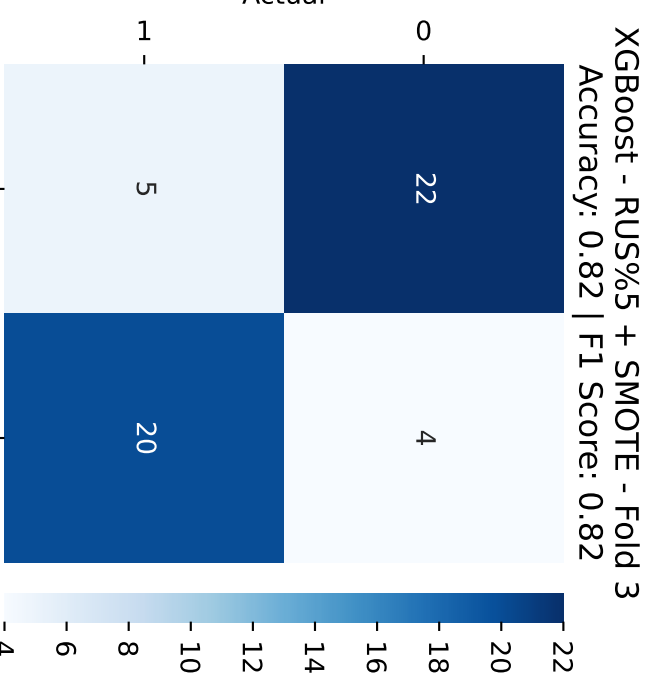
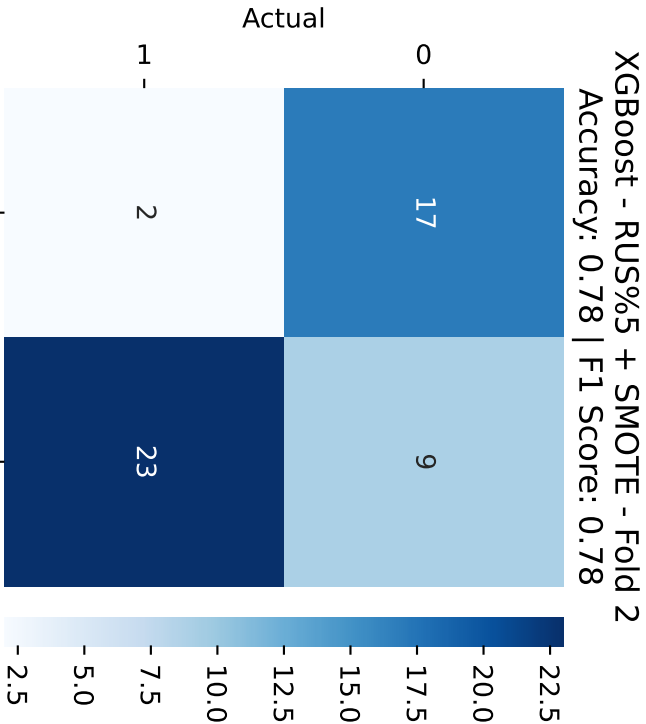
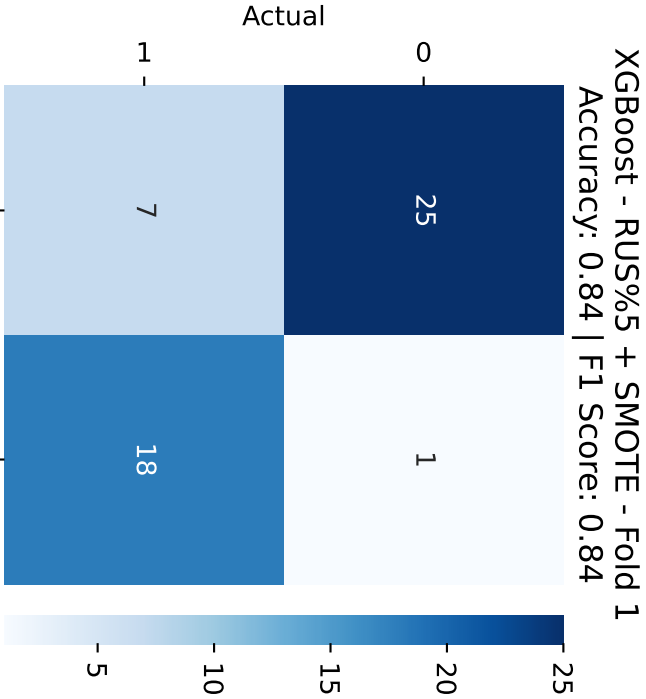


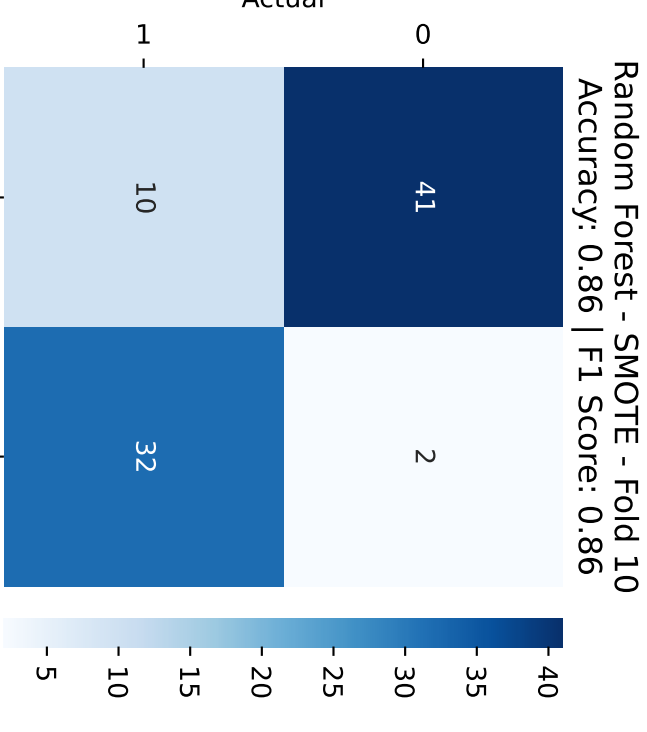
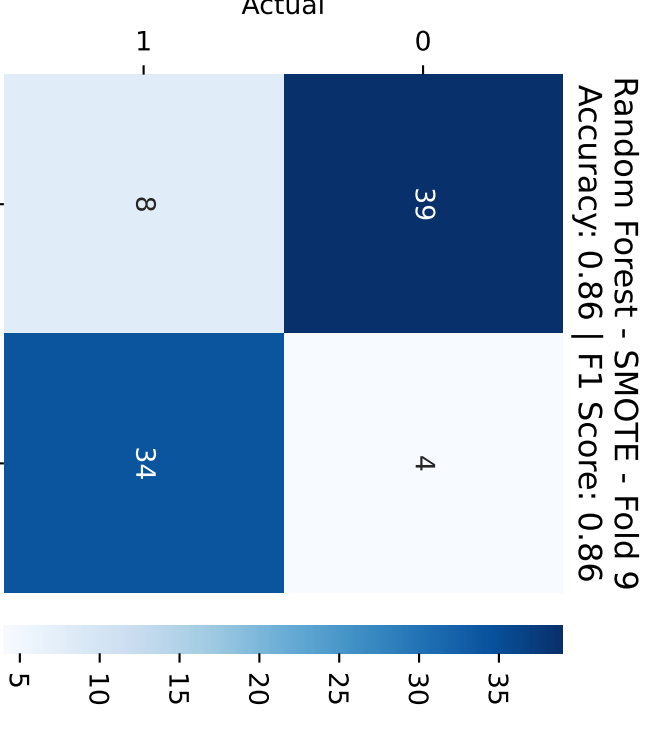
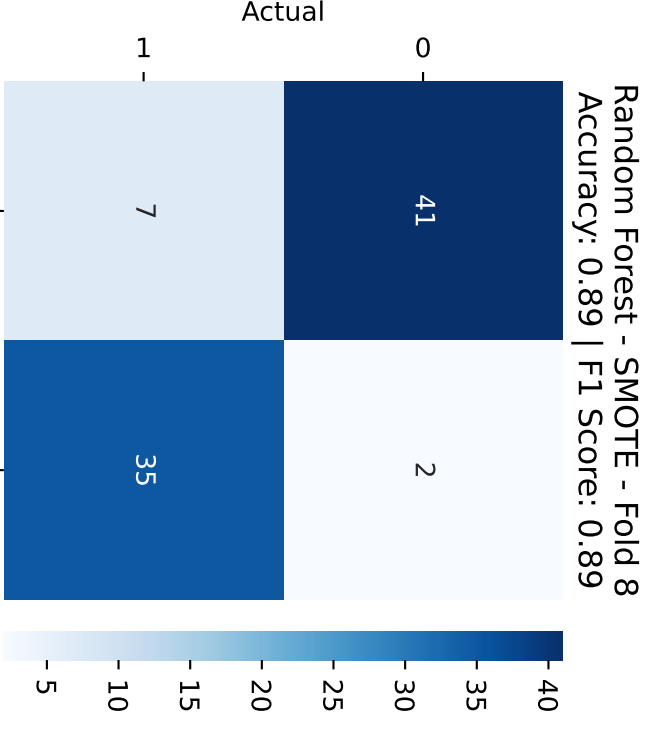
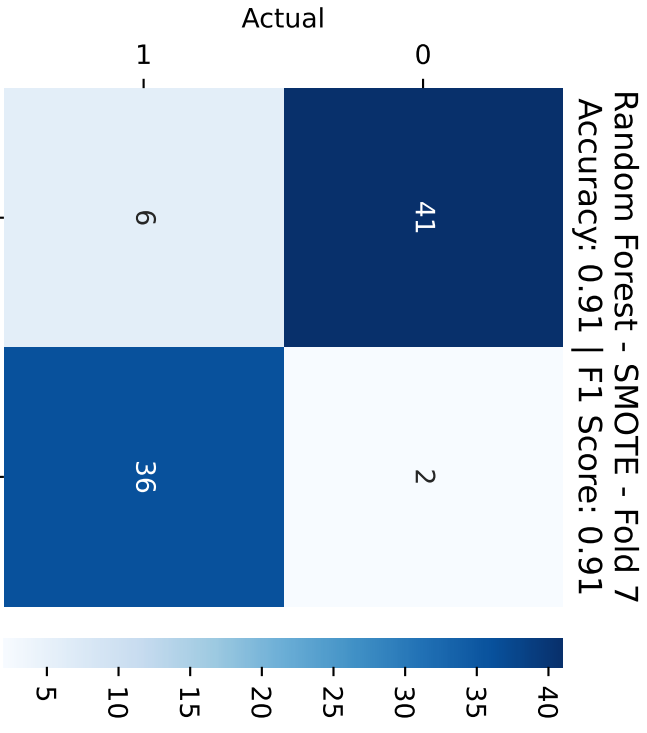
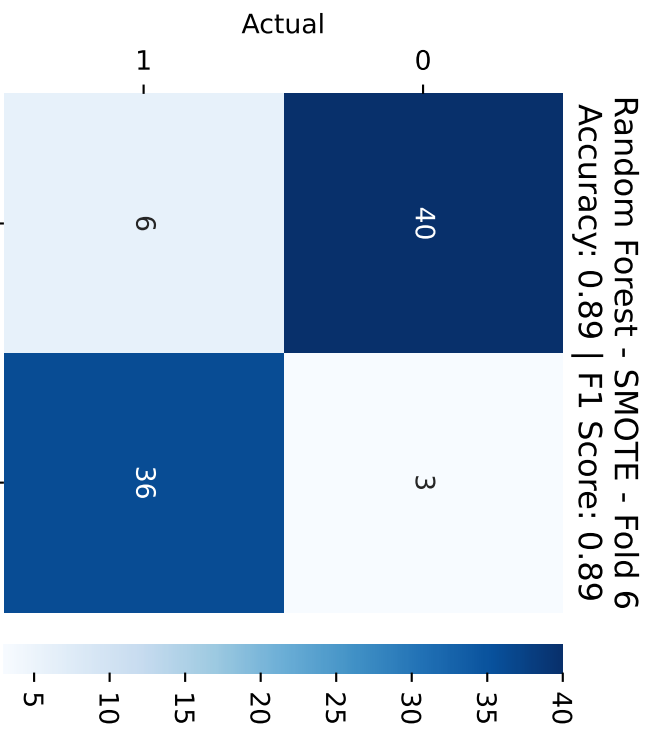
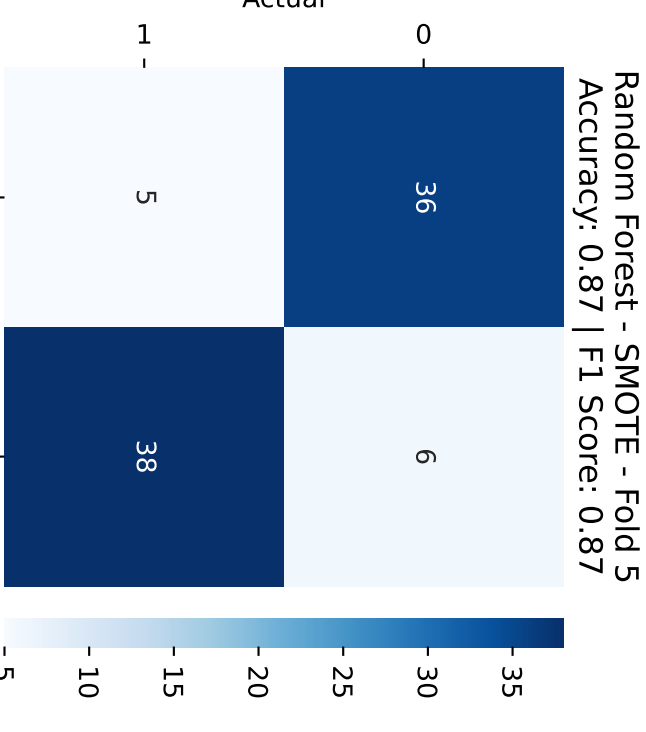
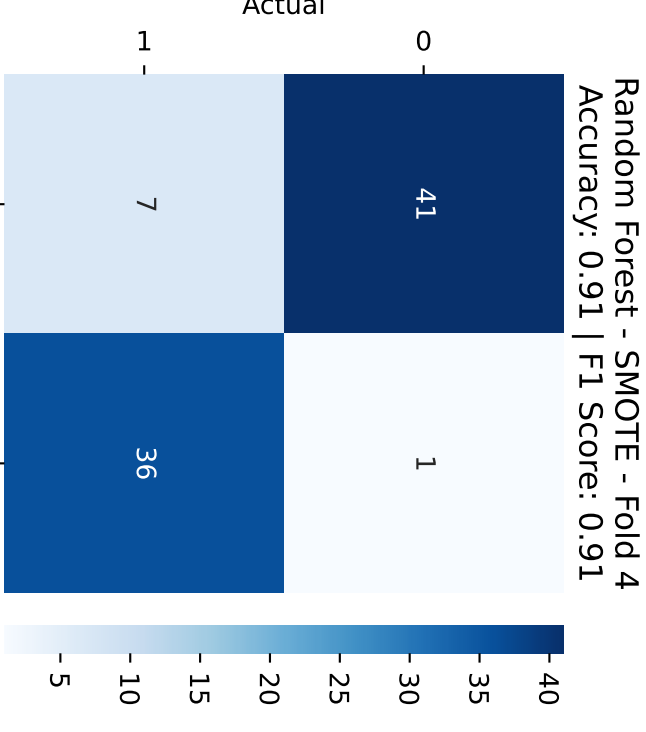
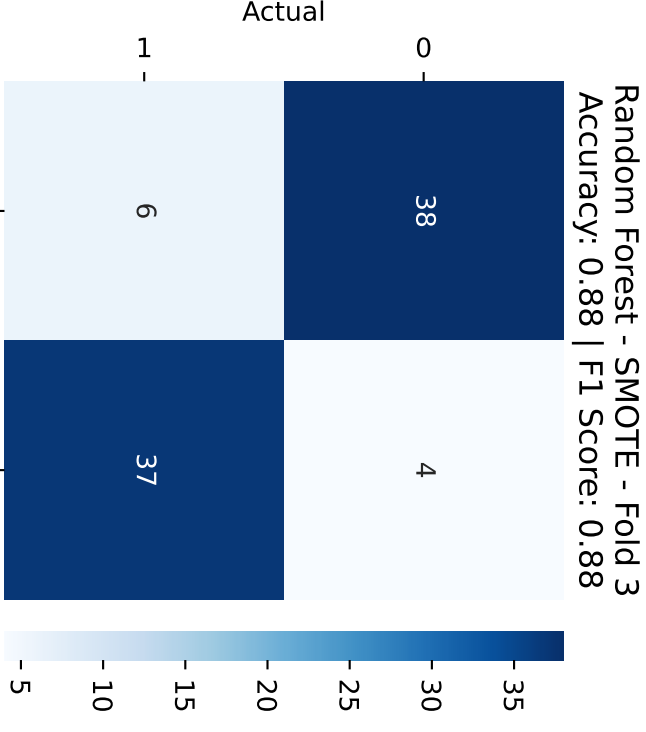
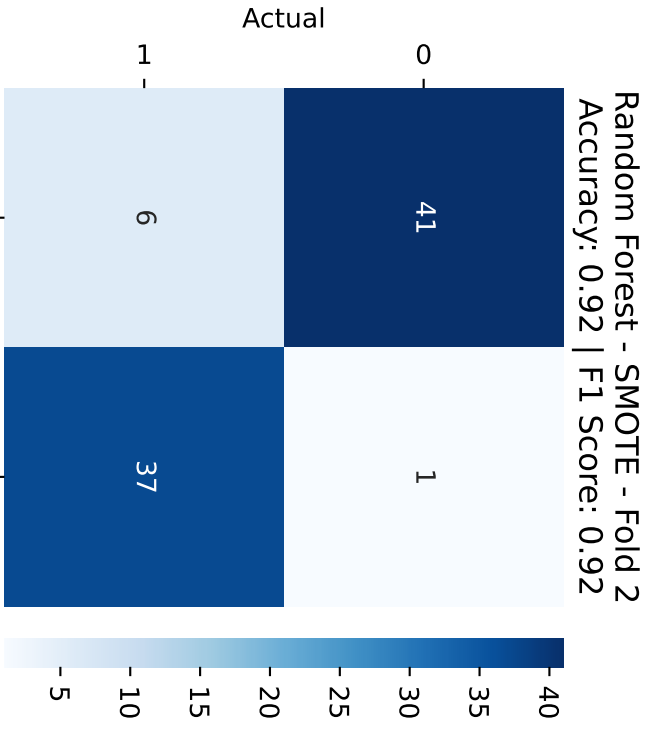
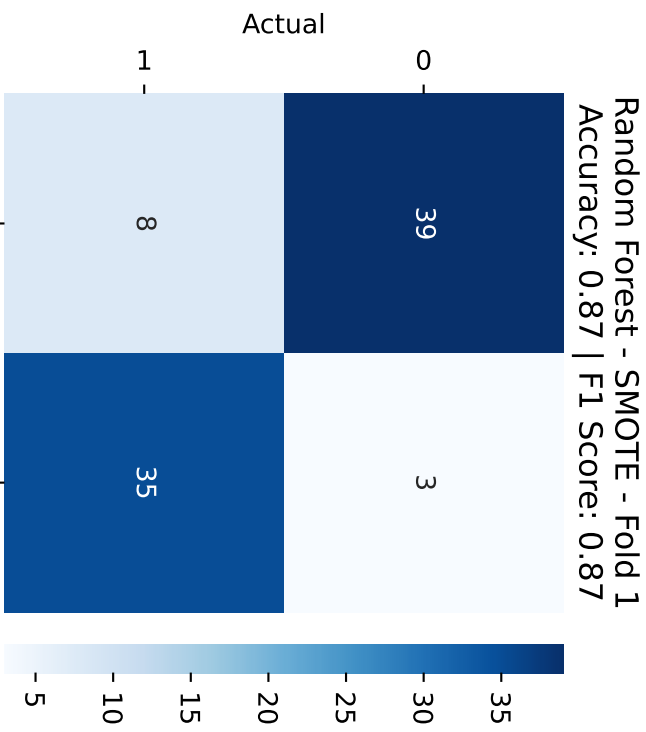
Decision Tree - RUS%5 + SMOTE - Fold 10  
Accuracy: 0.84 | F1 Score: 0.84

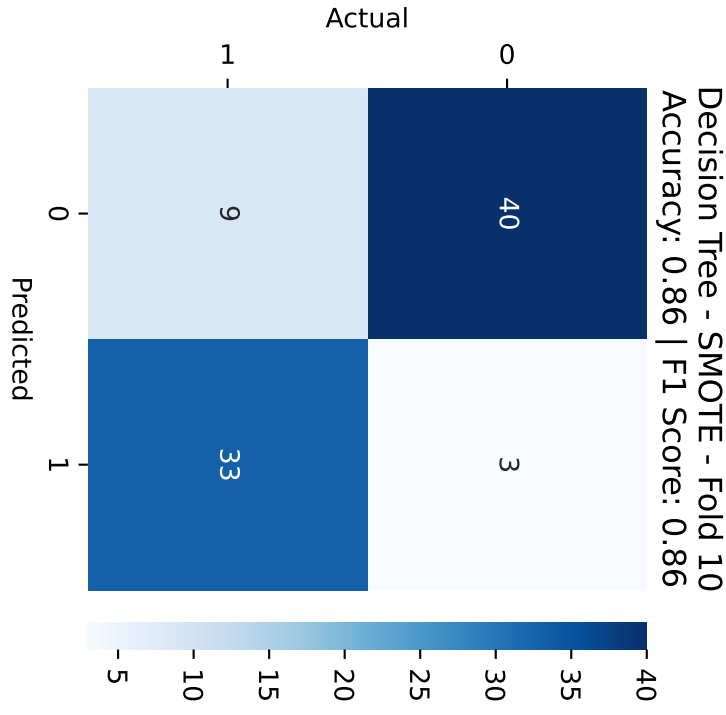
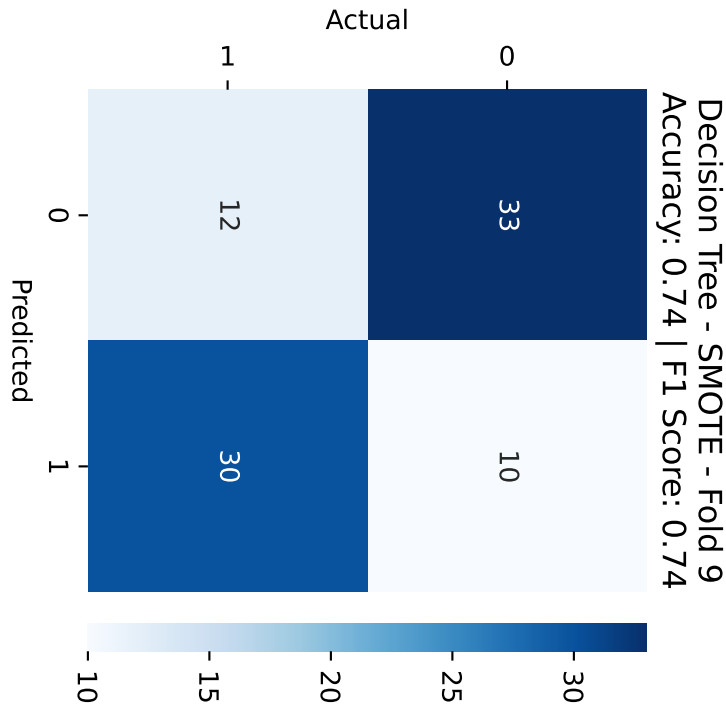
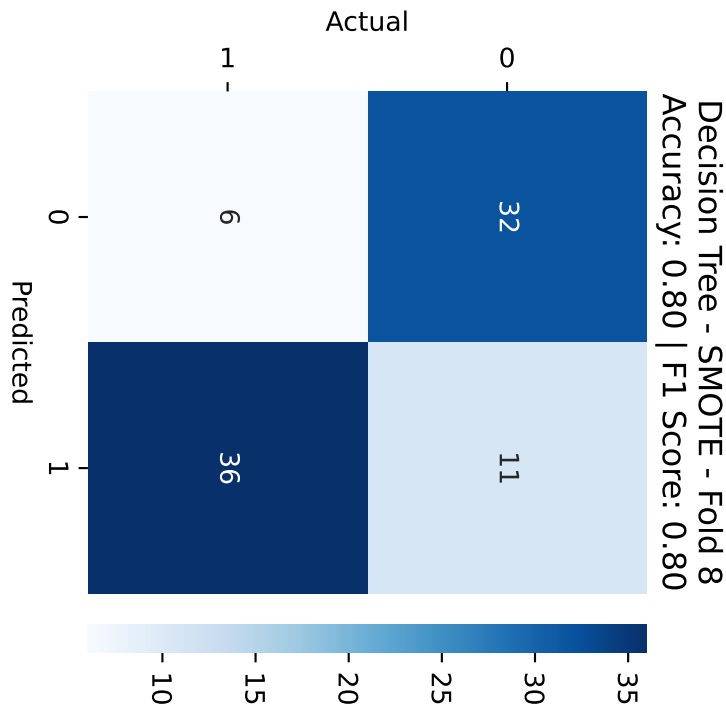
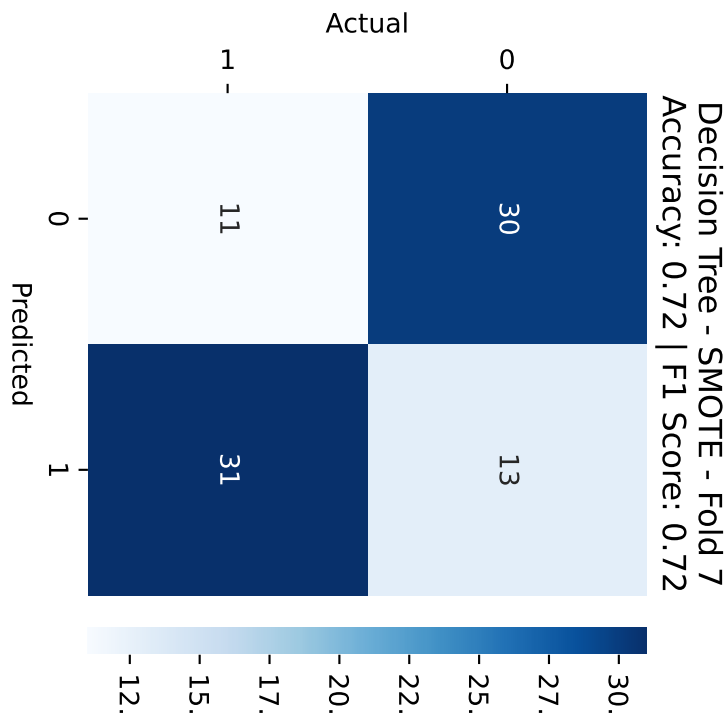
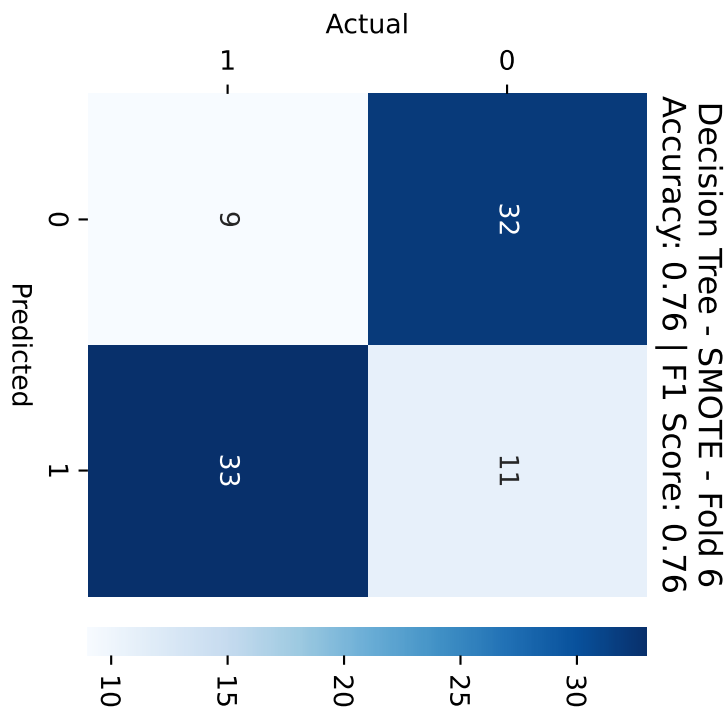
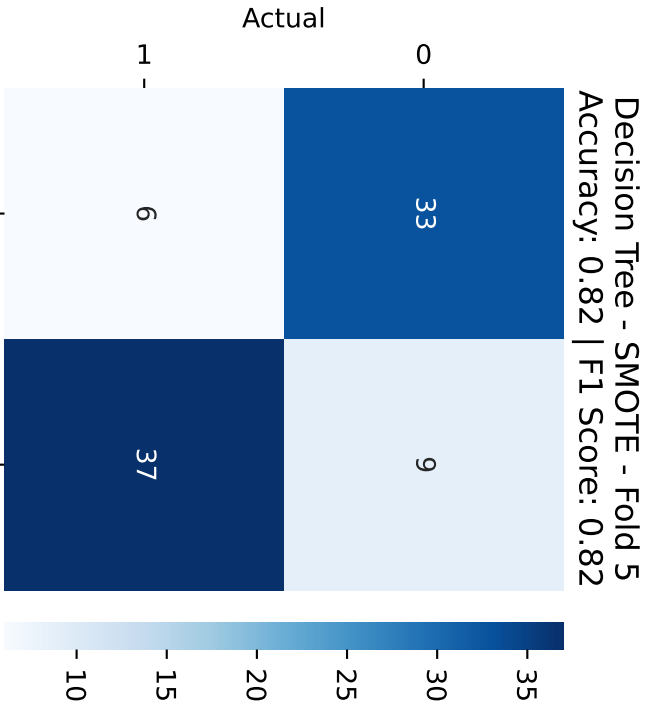
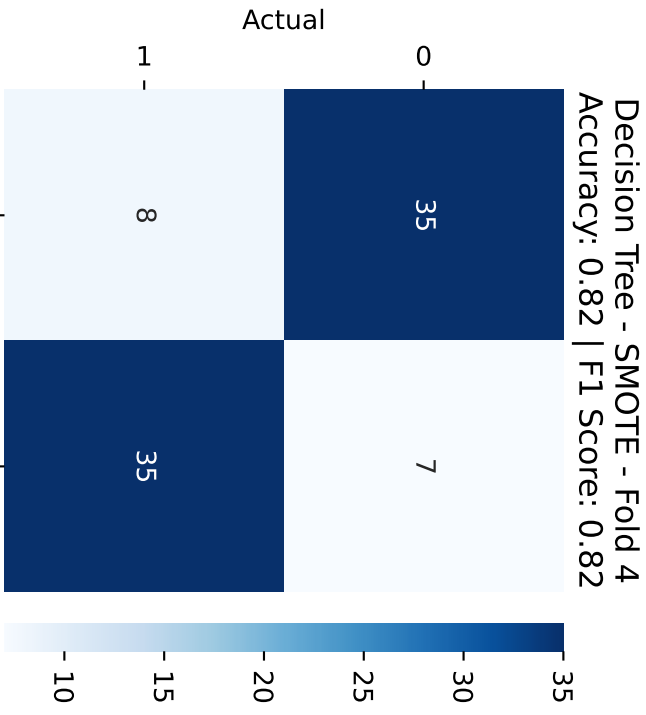
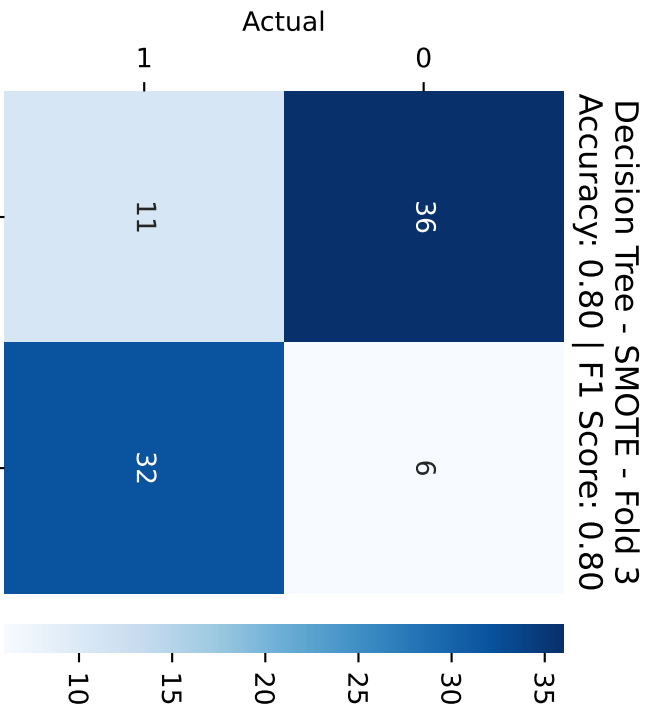
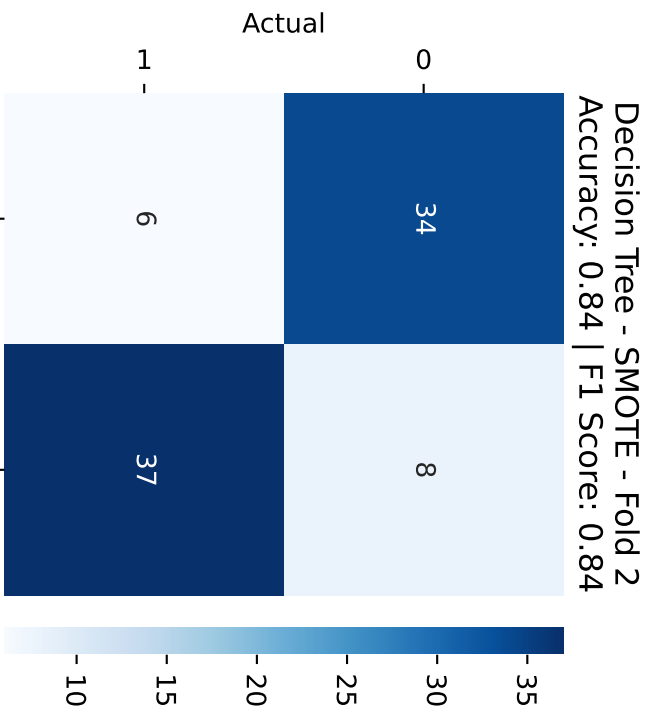
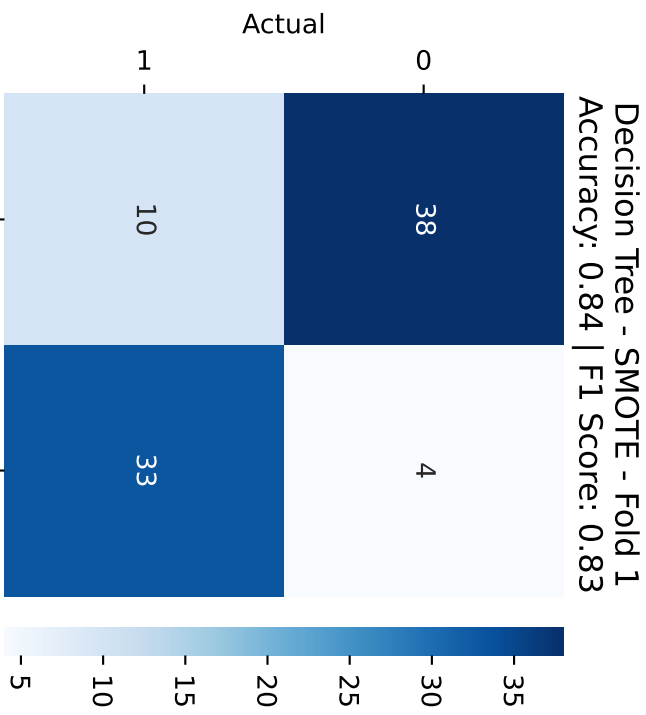


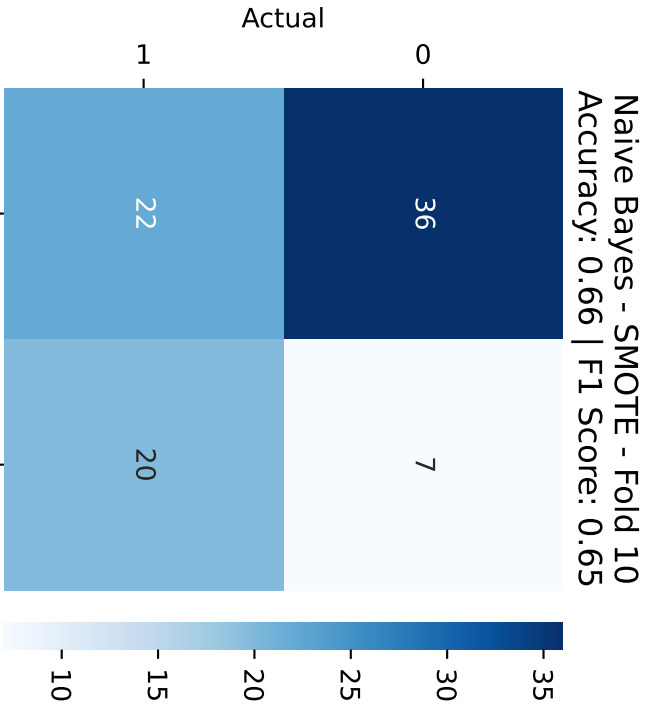
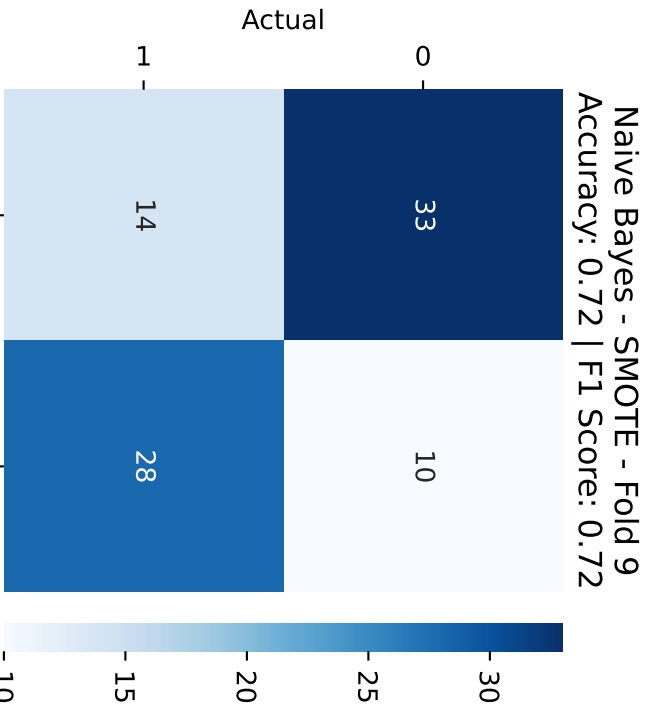
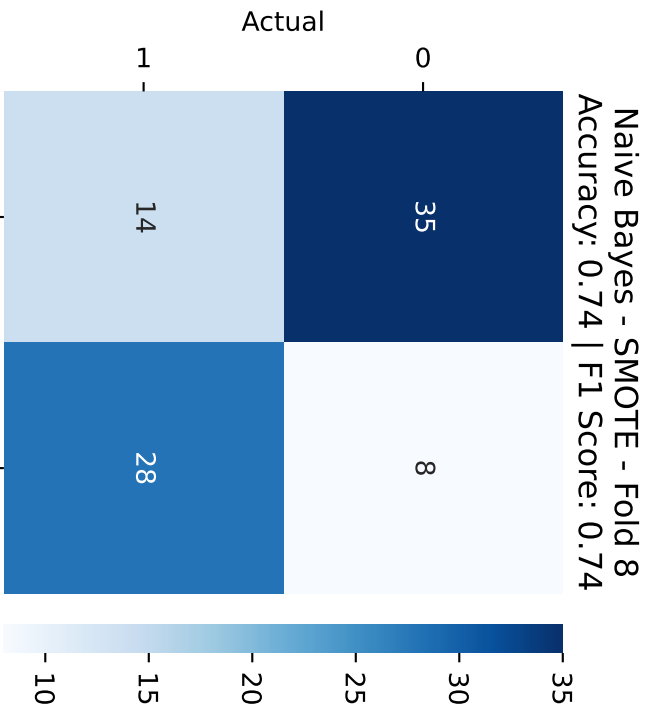
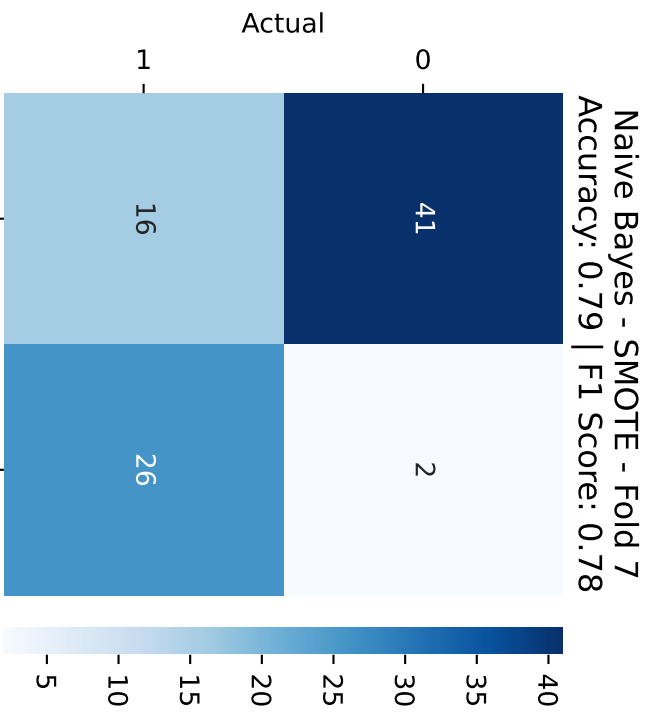
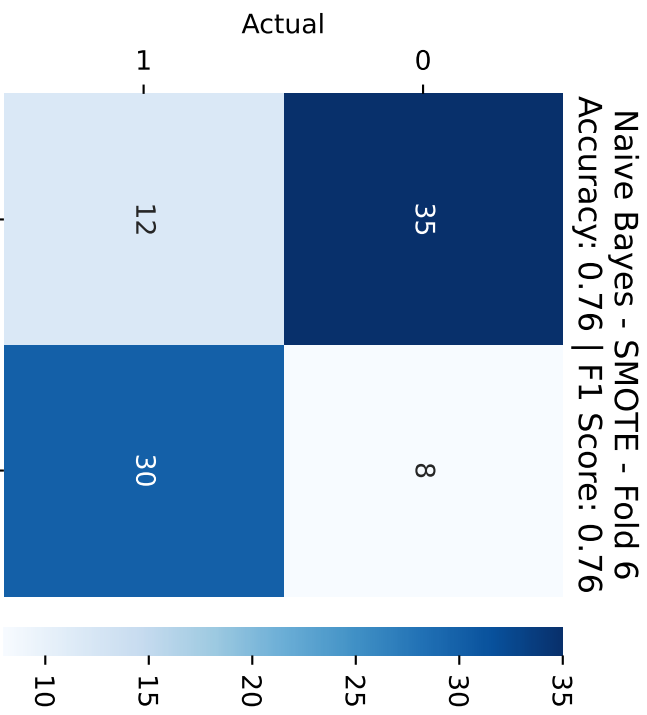
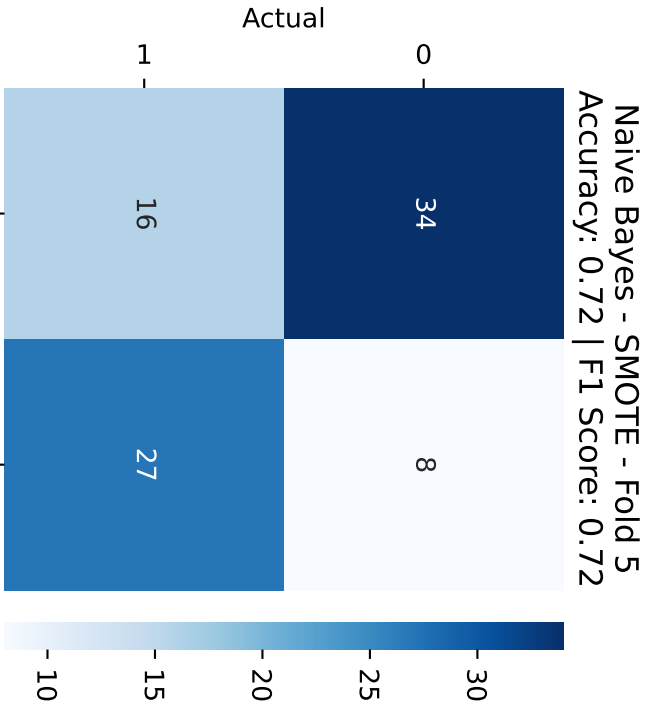
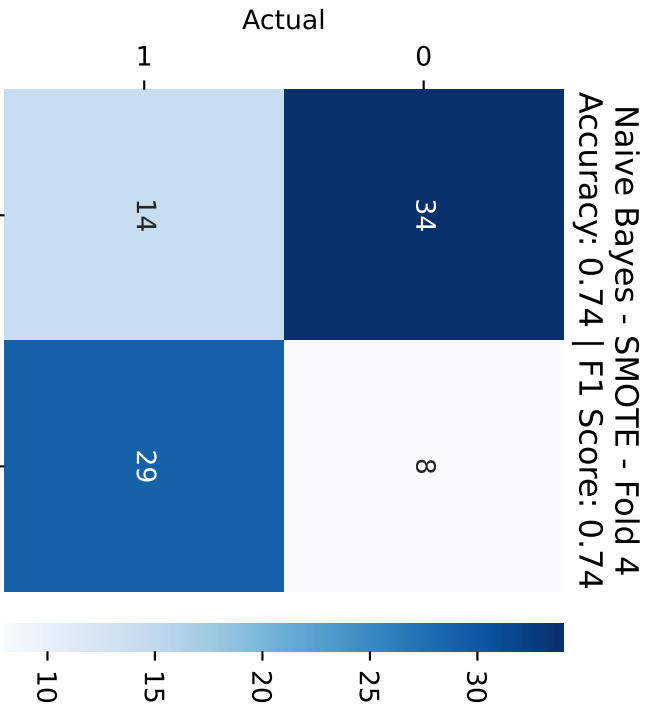
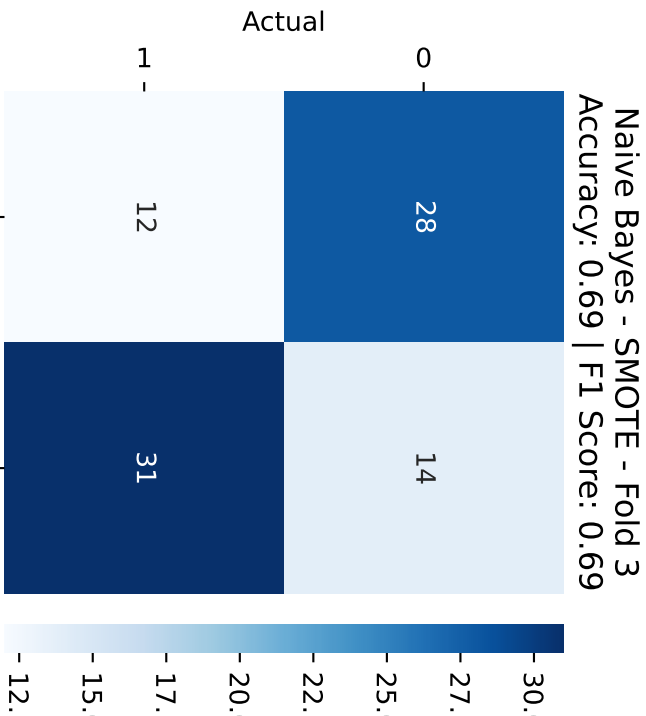
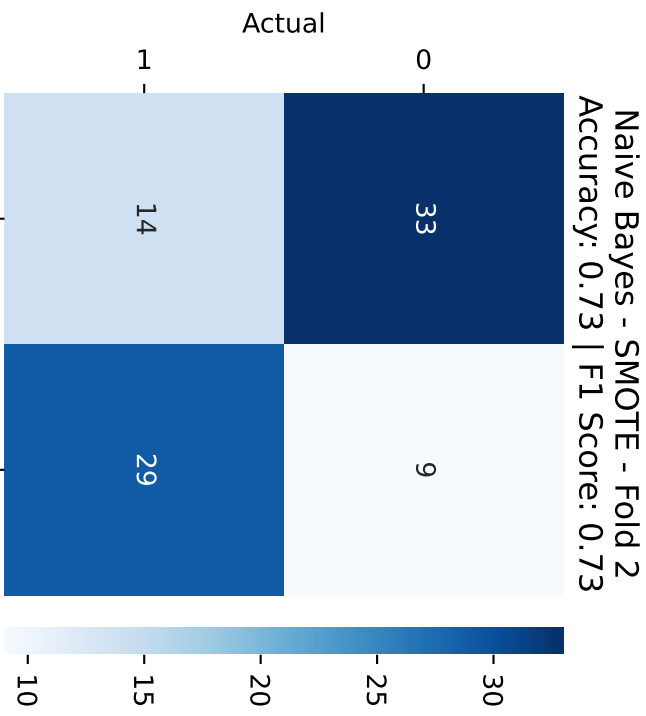
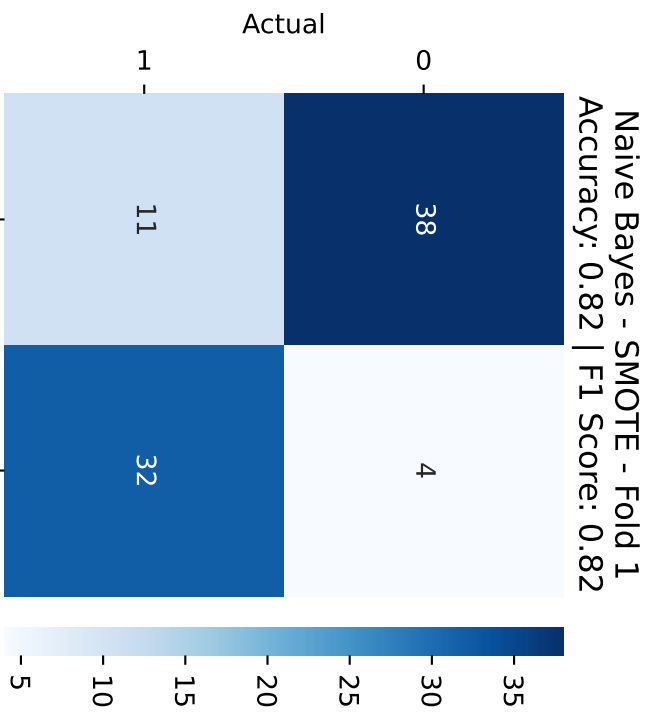


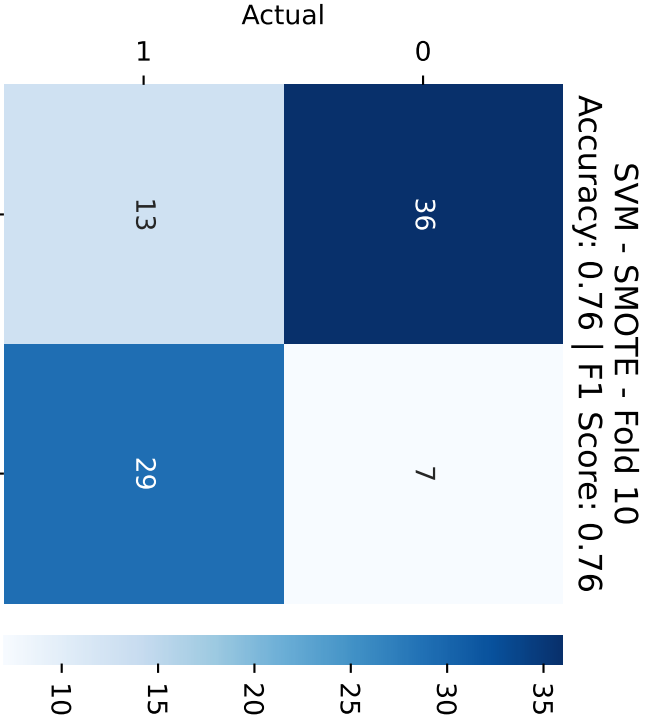
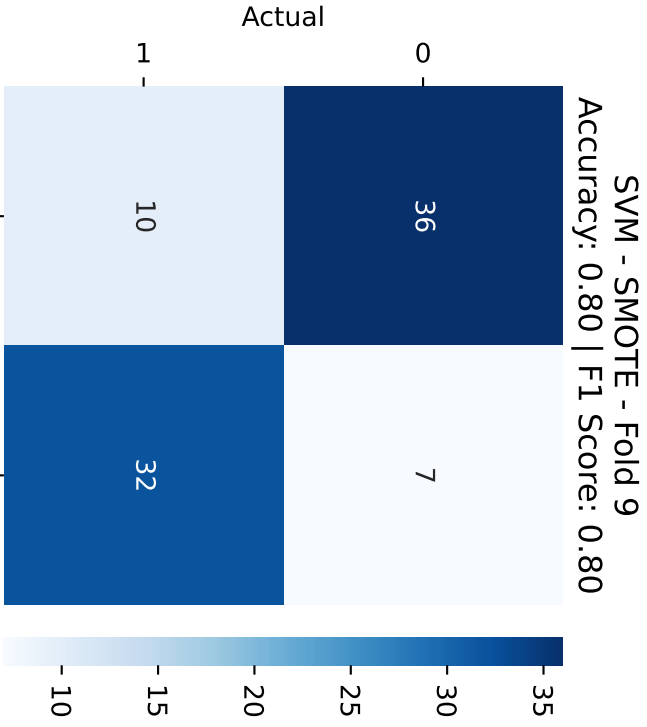
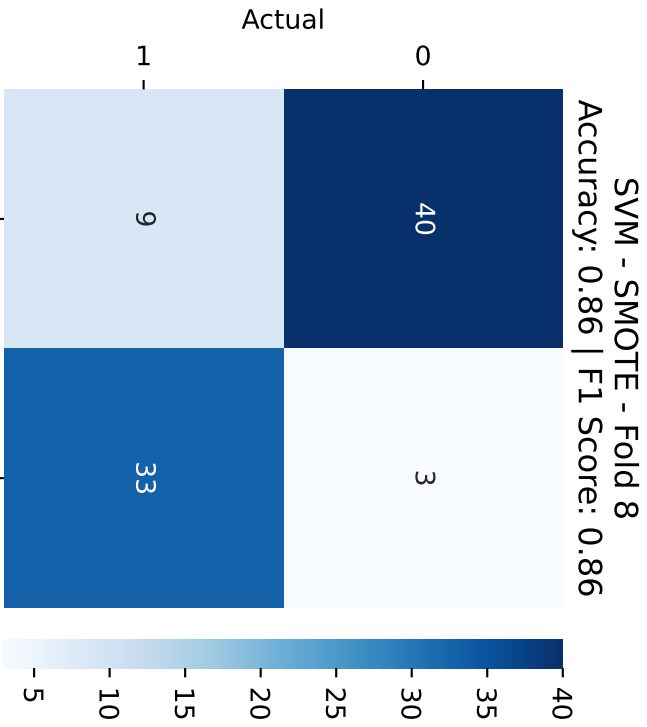
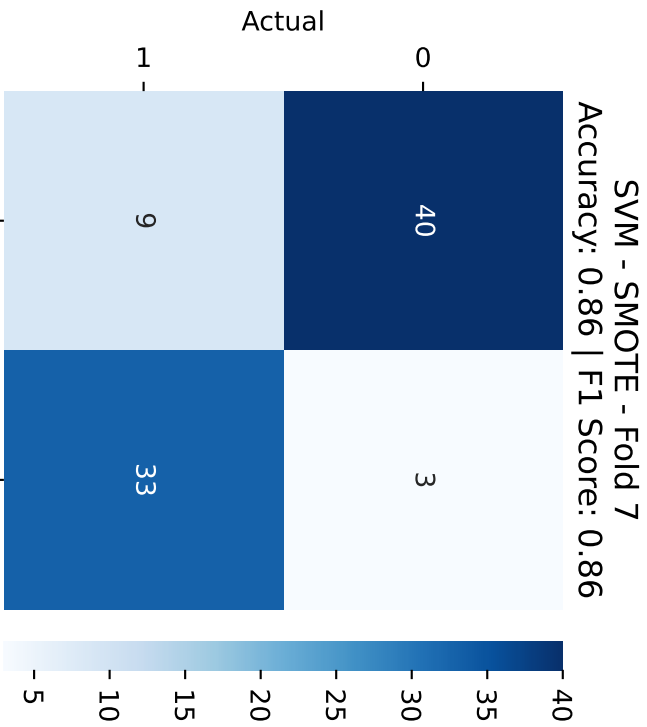
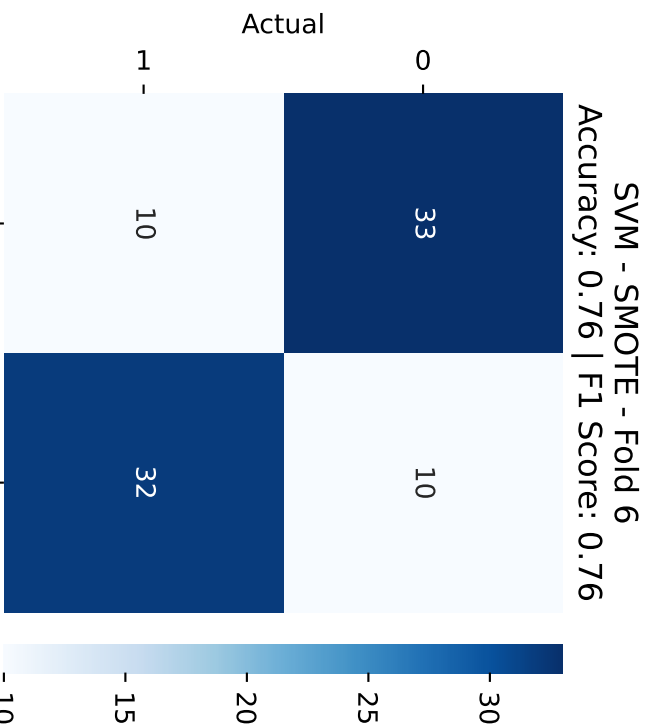
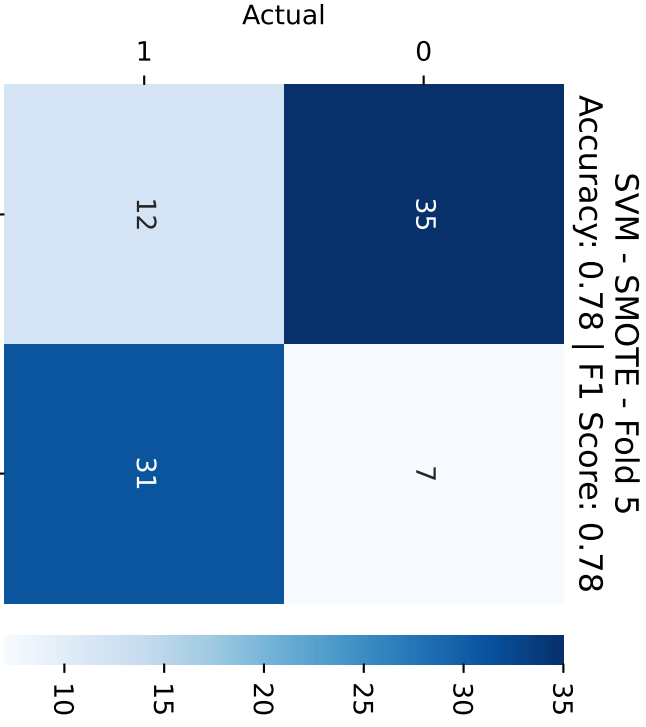
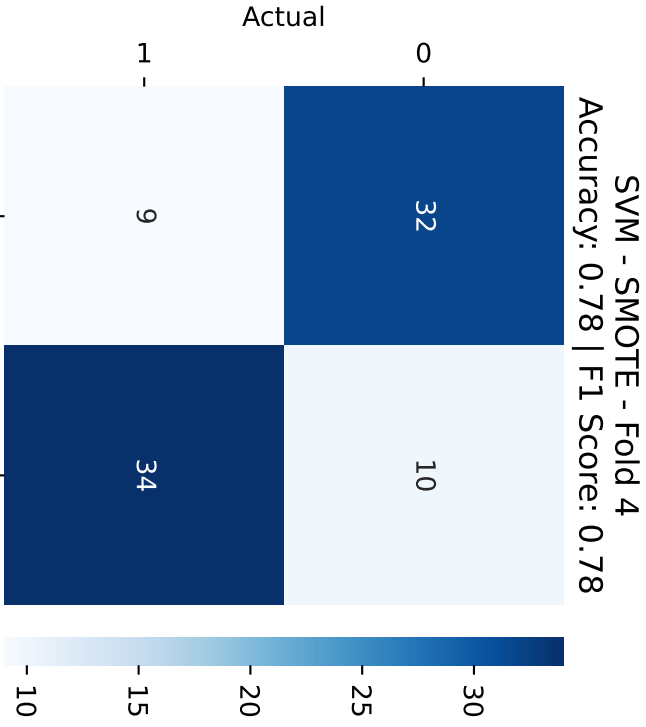
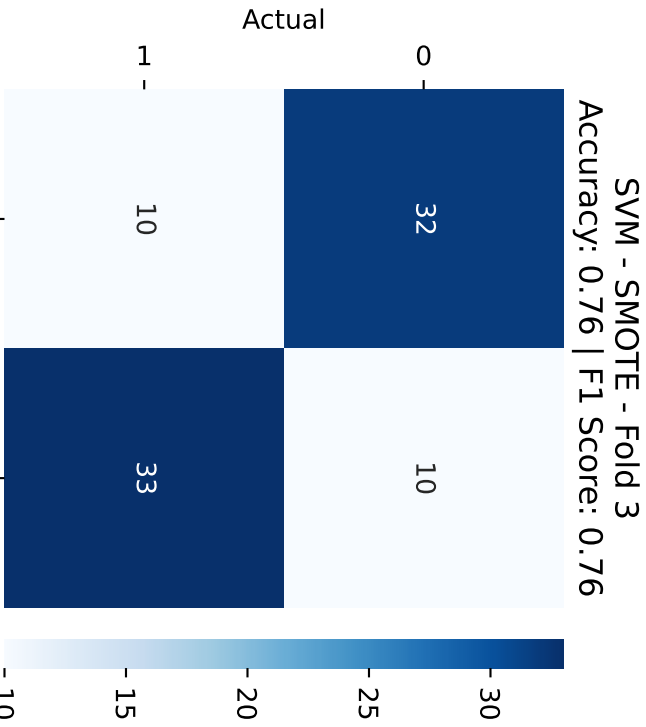
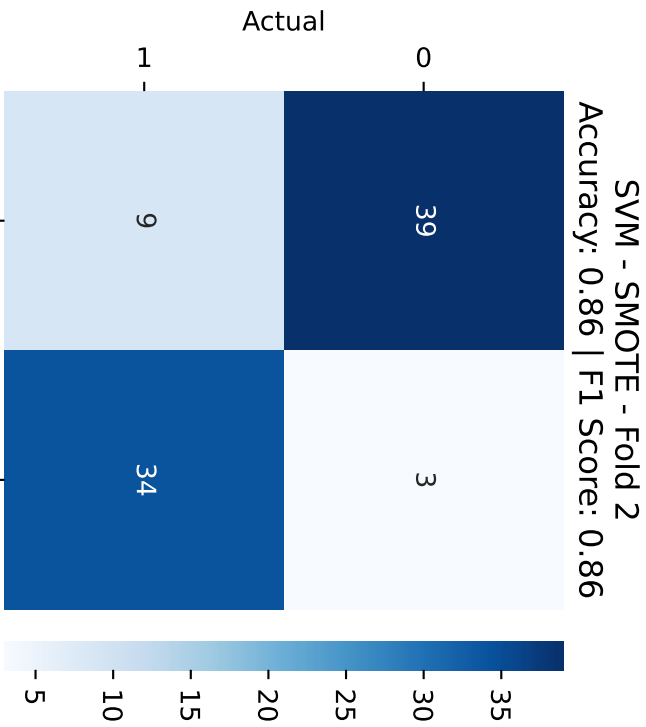
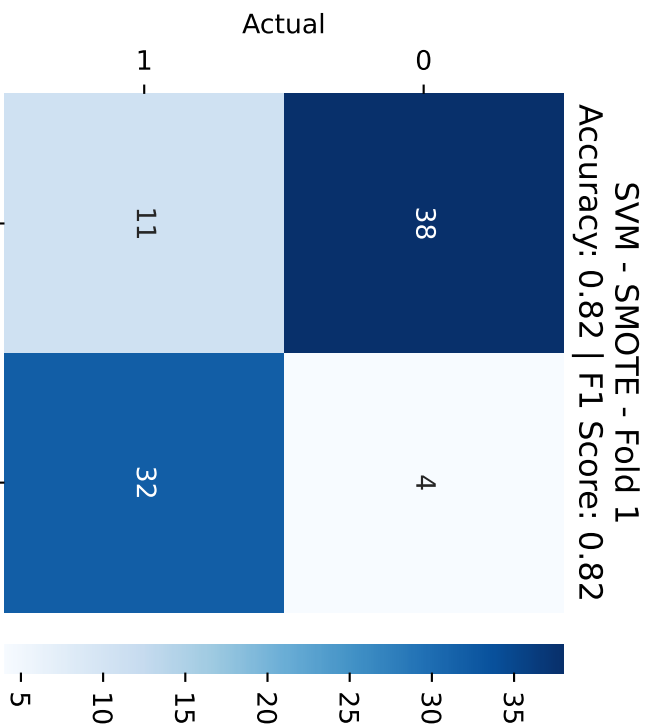


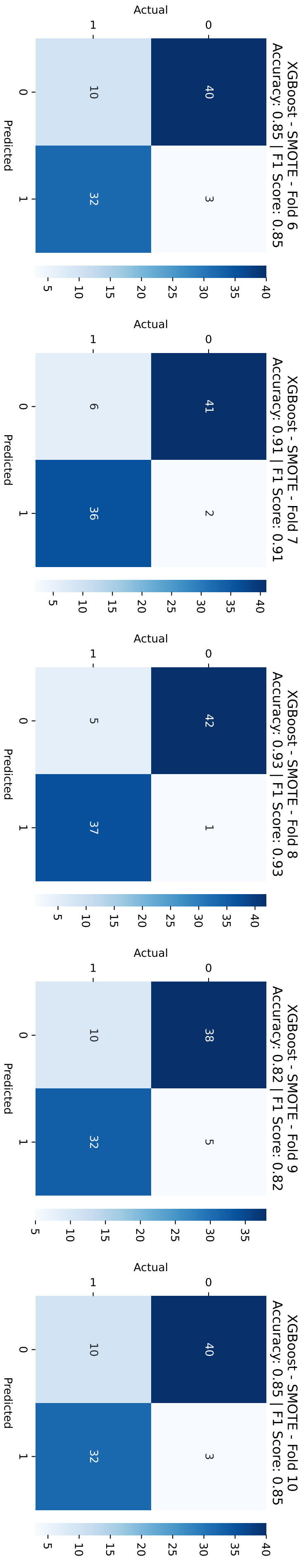
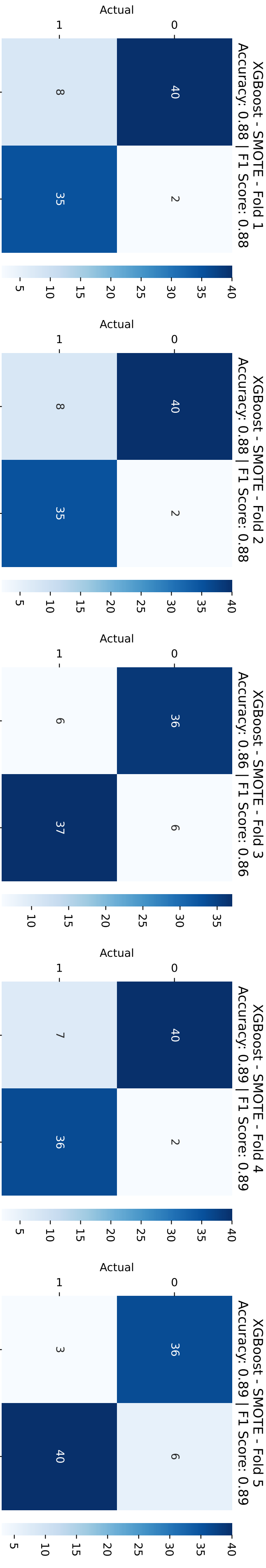


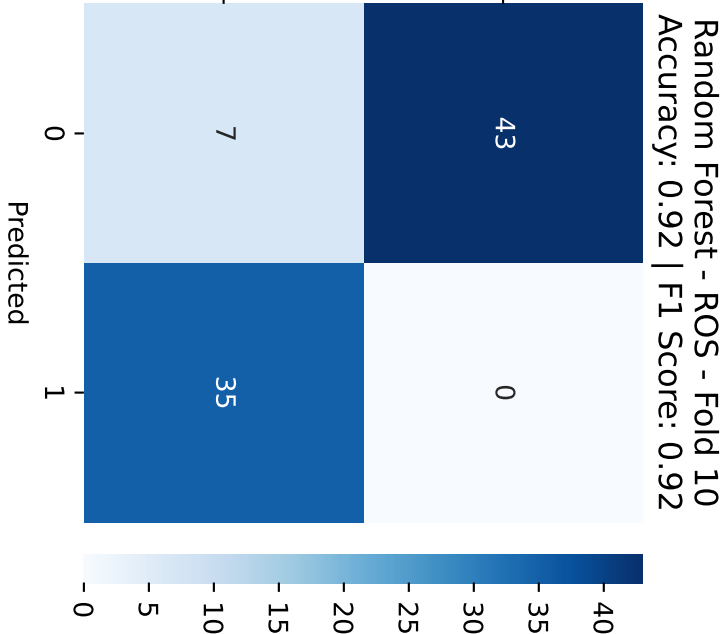
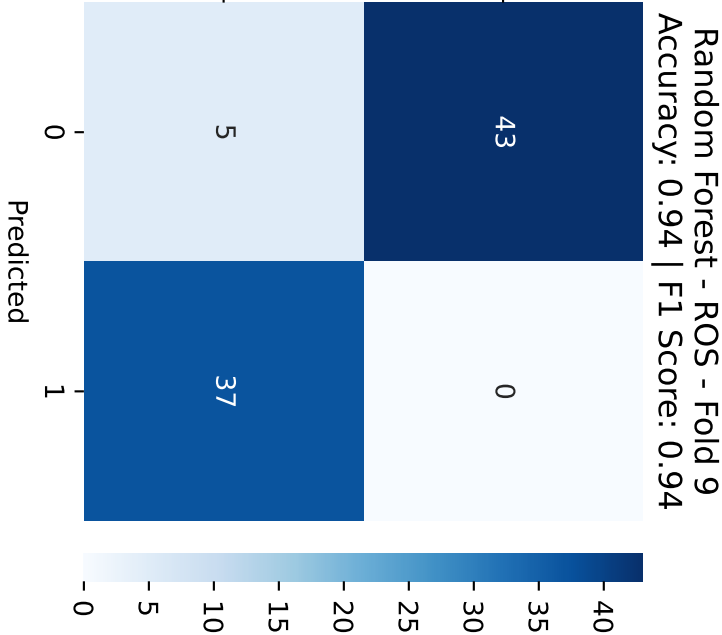
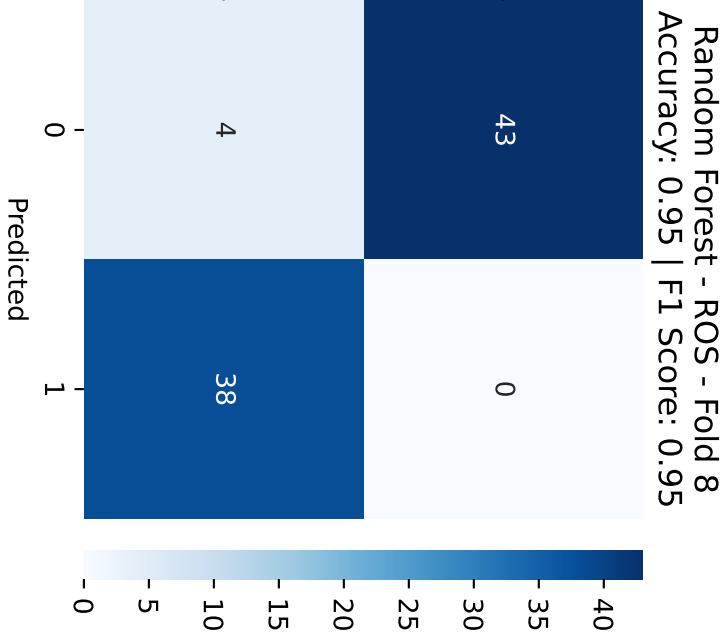
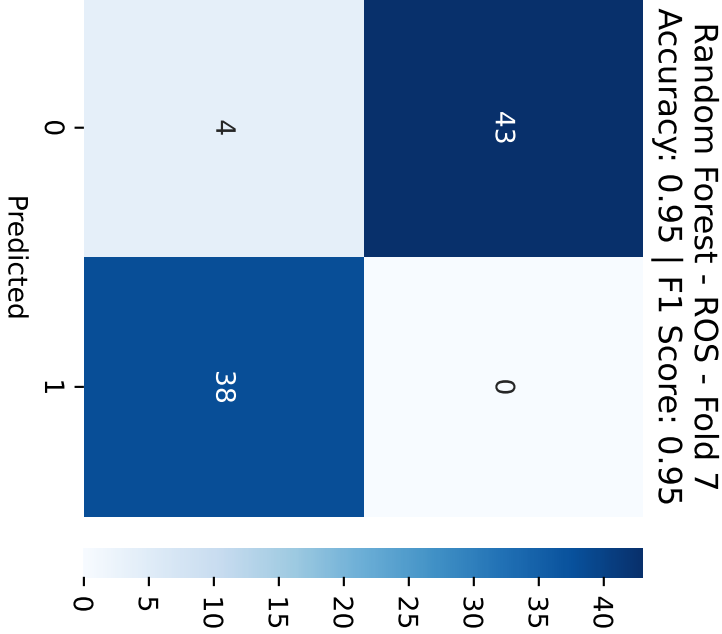
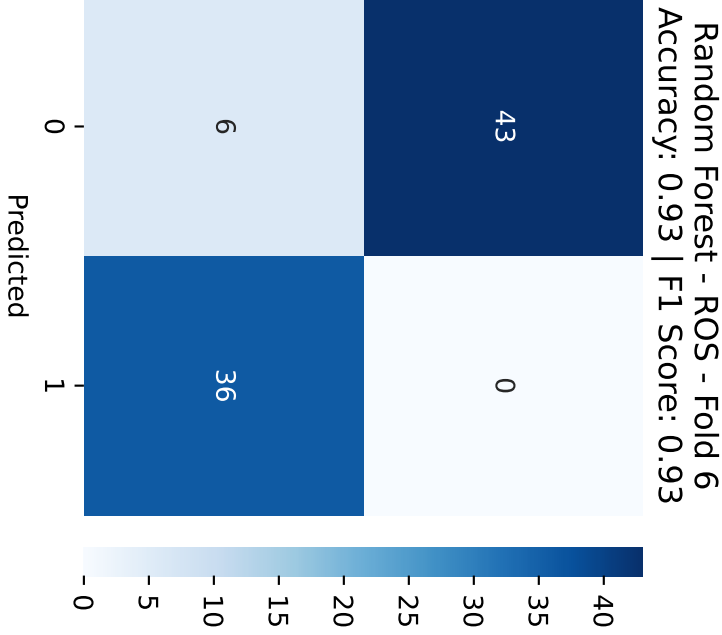
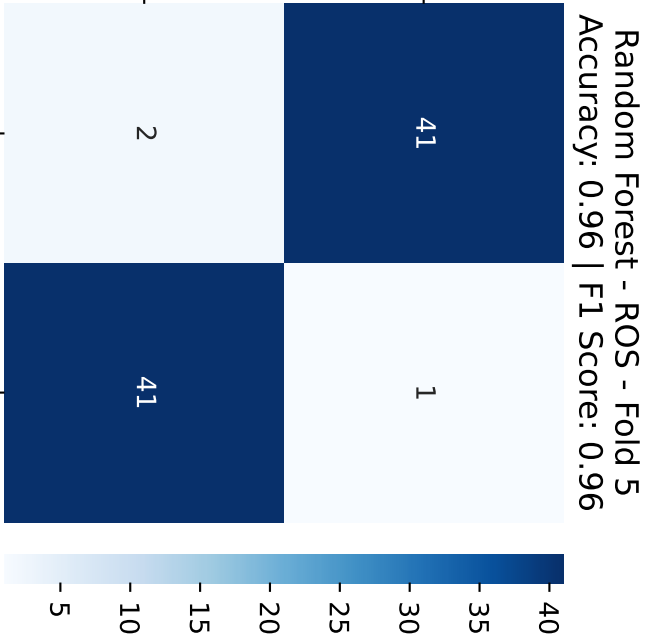
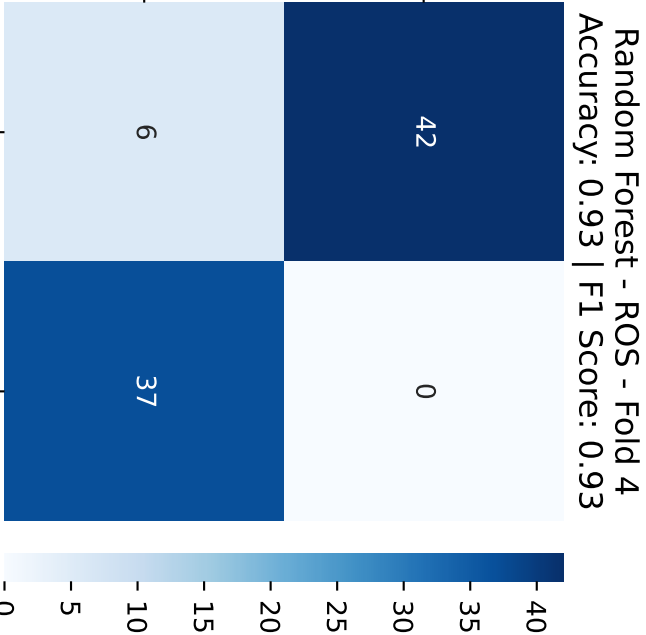
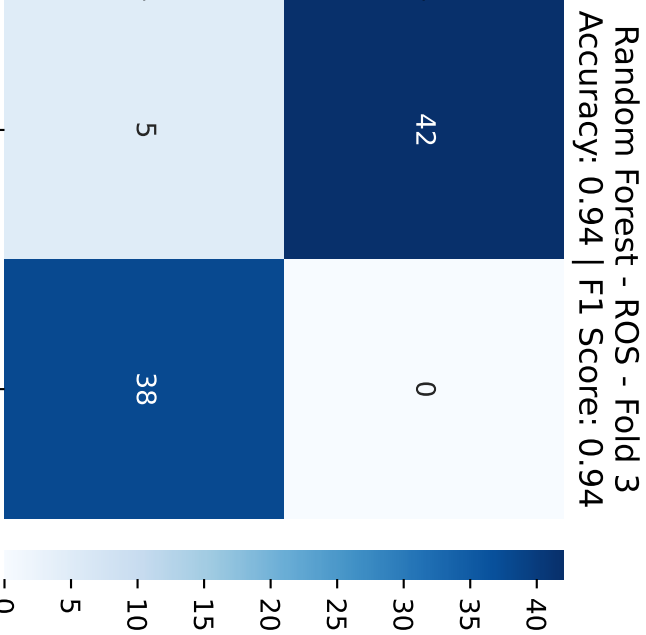
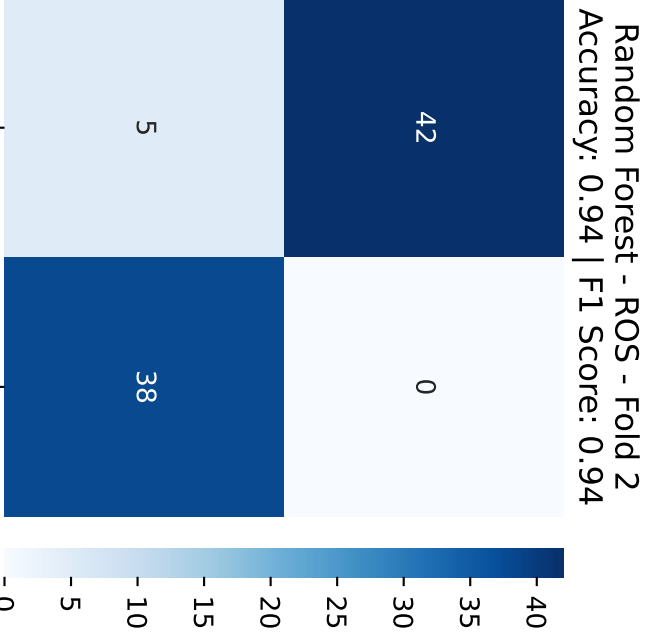
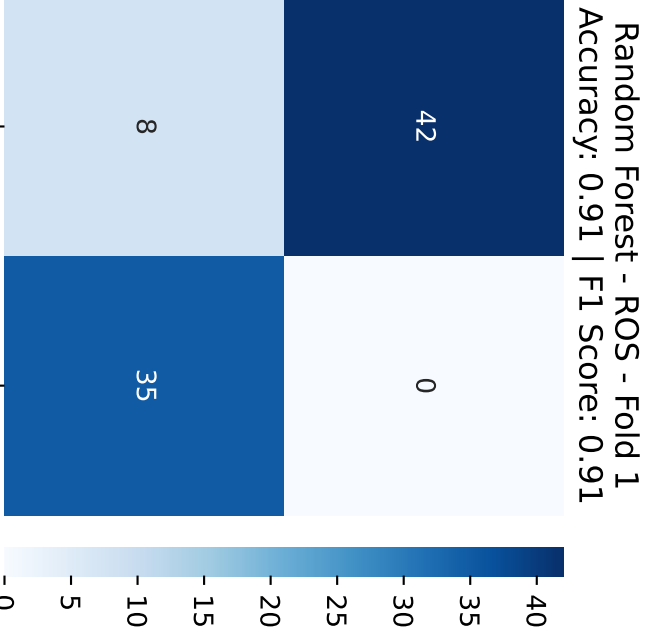


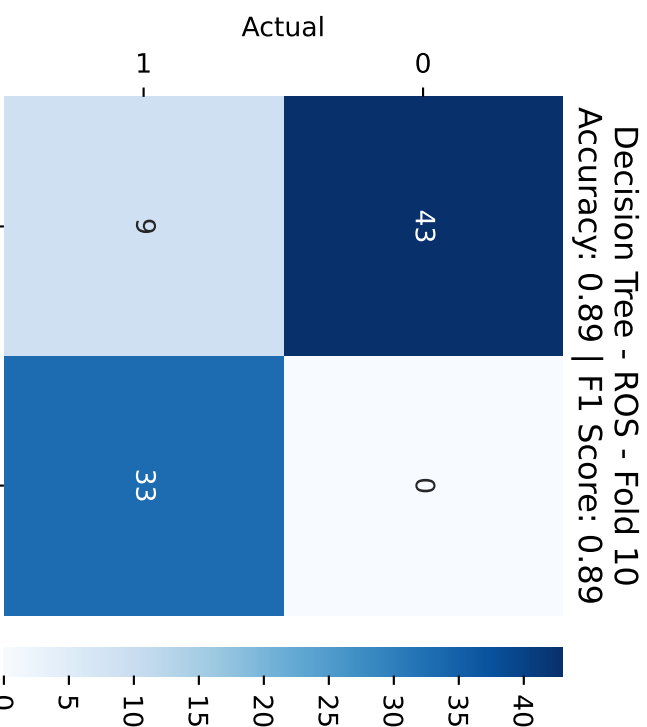
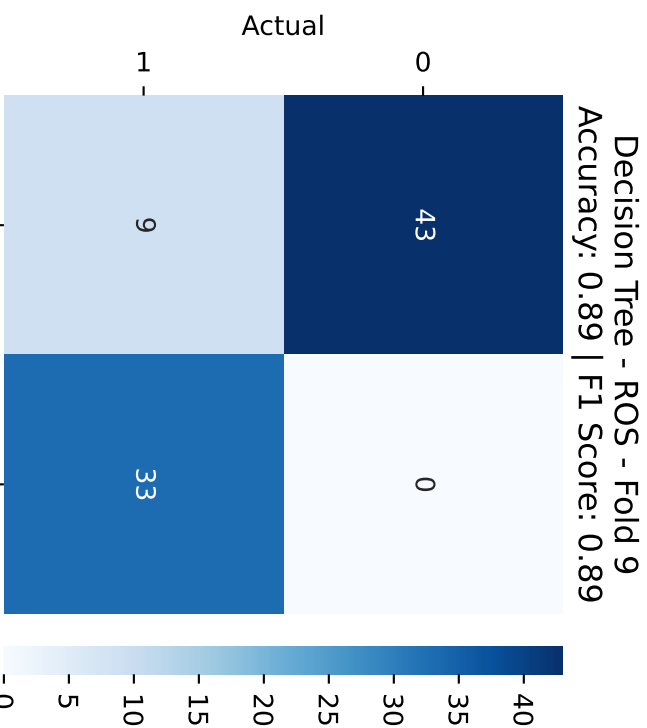
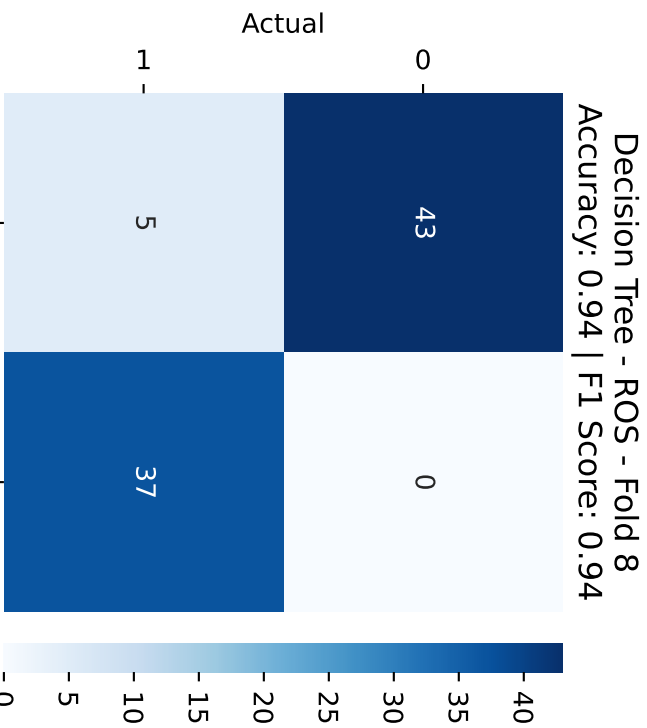
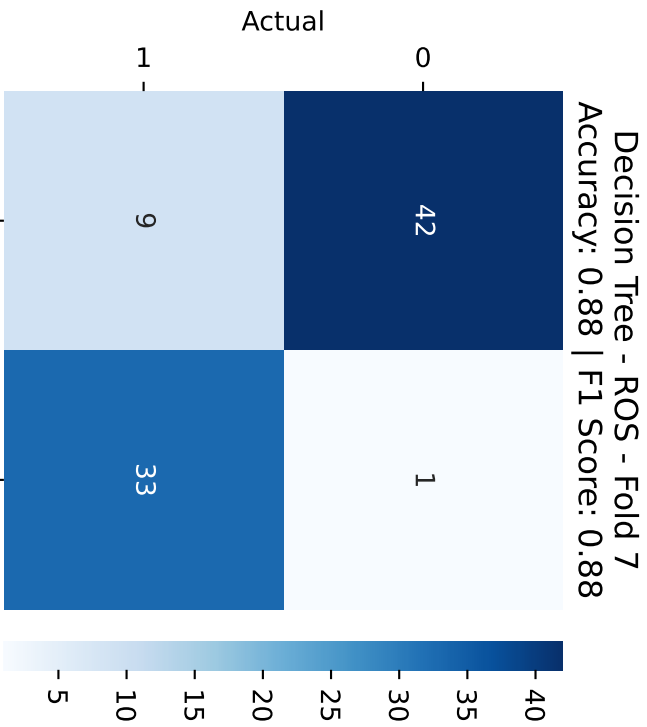
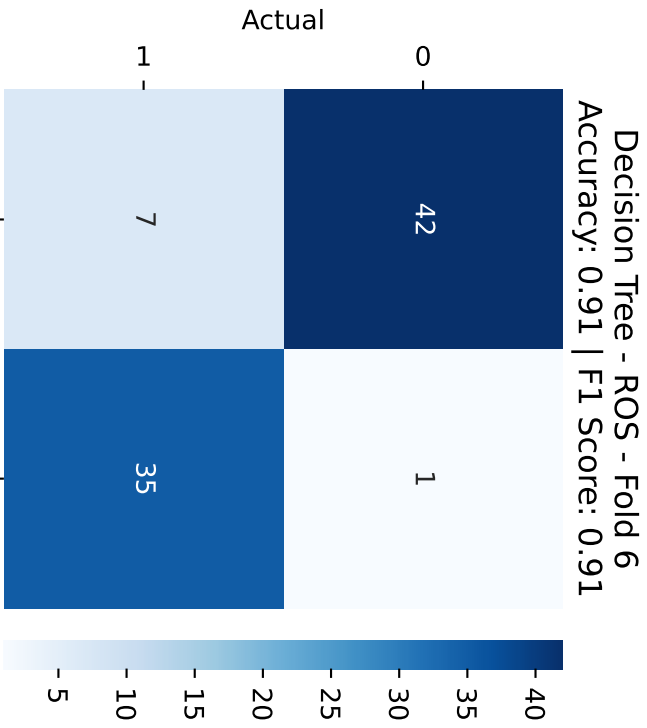
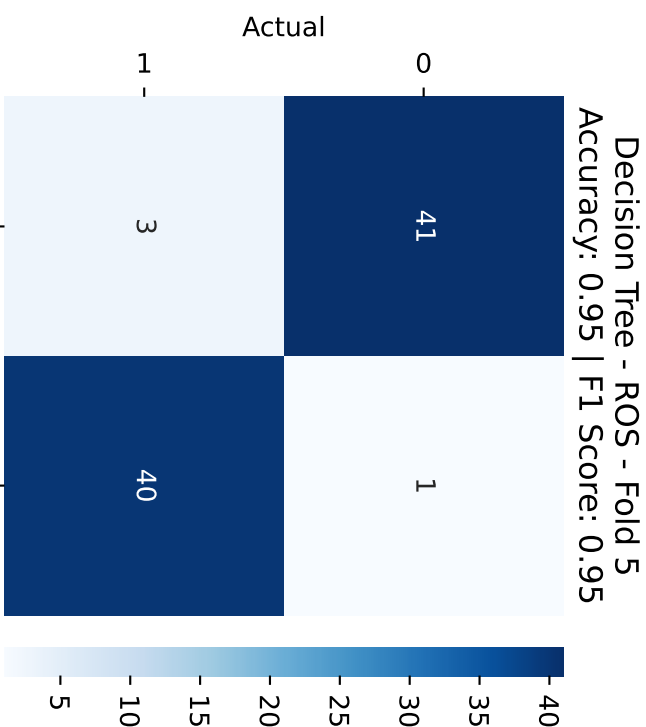
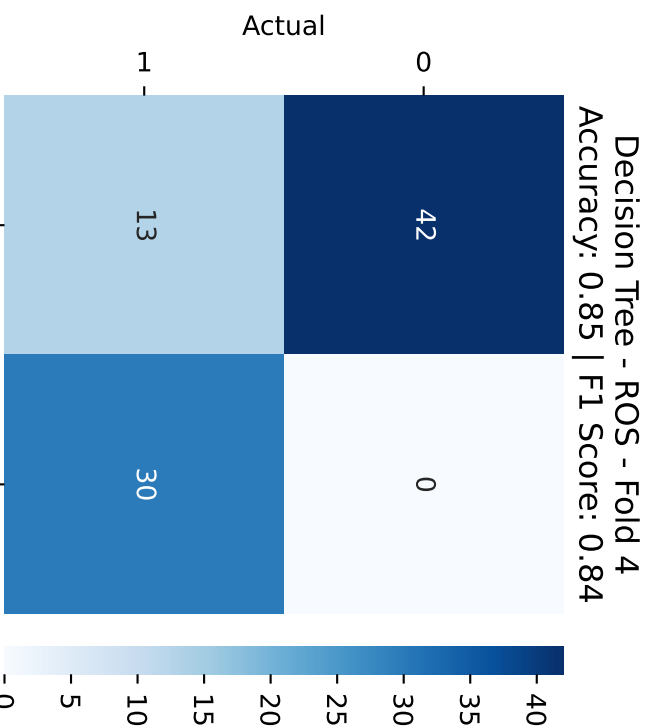
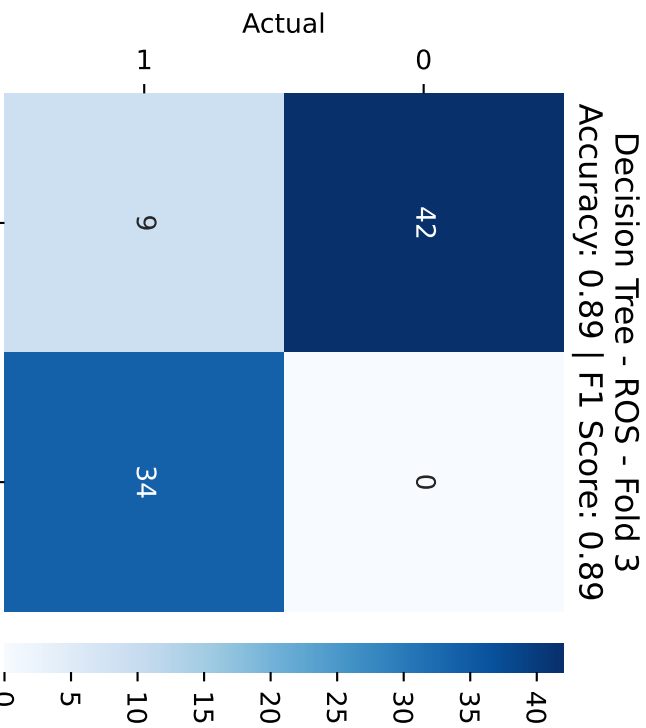
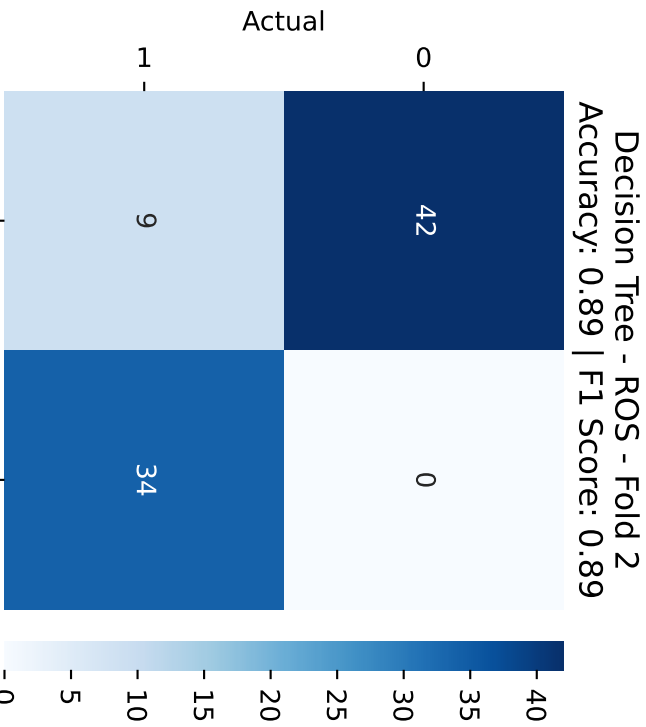
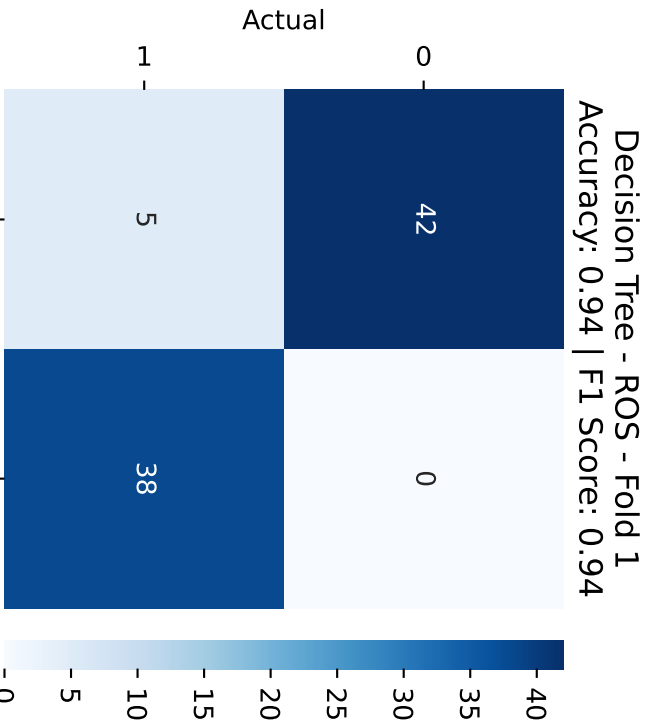


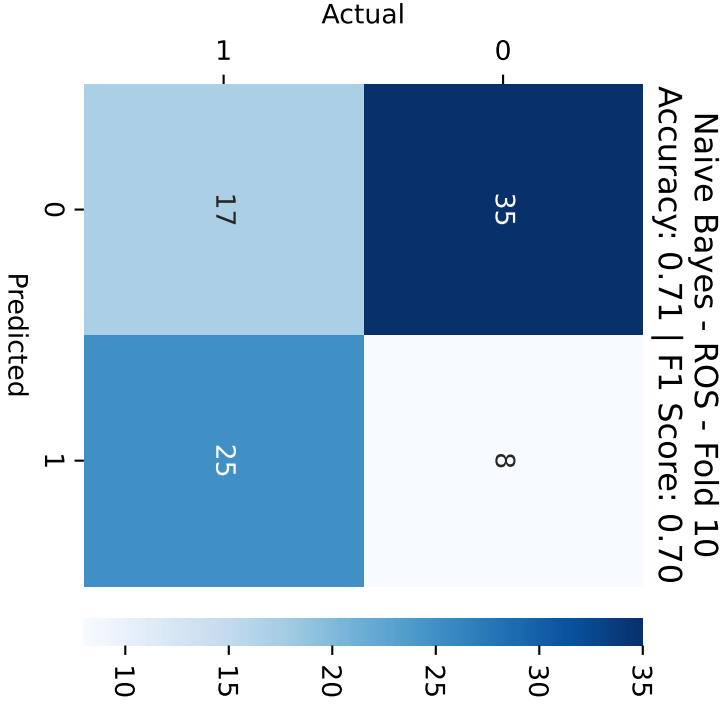
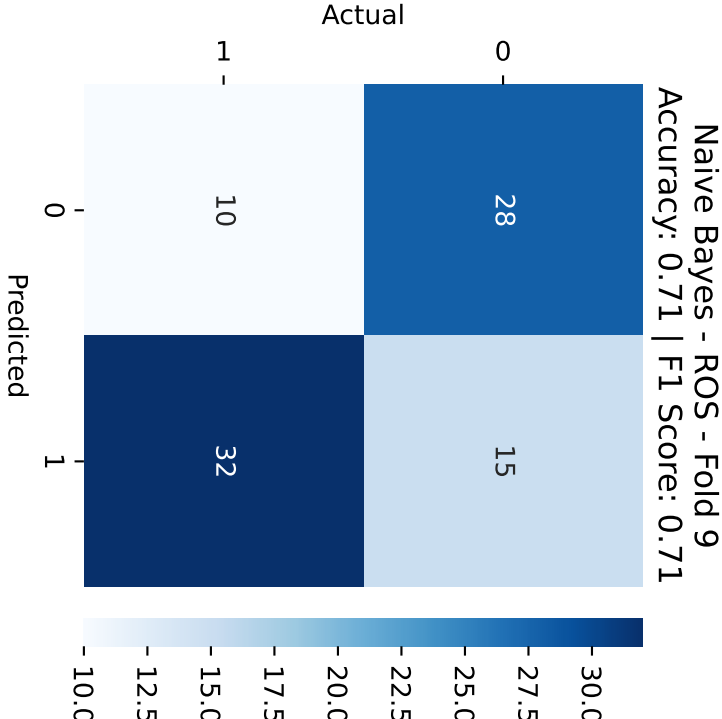
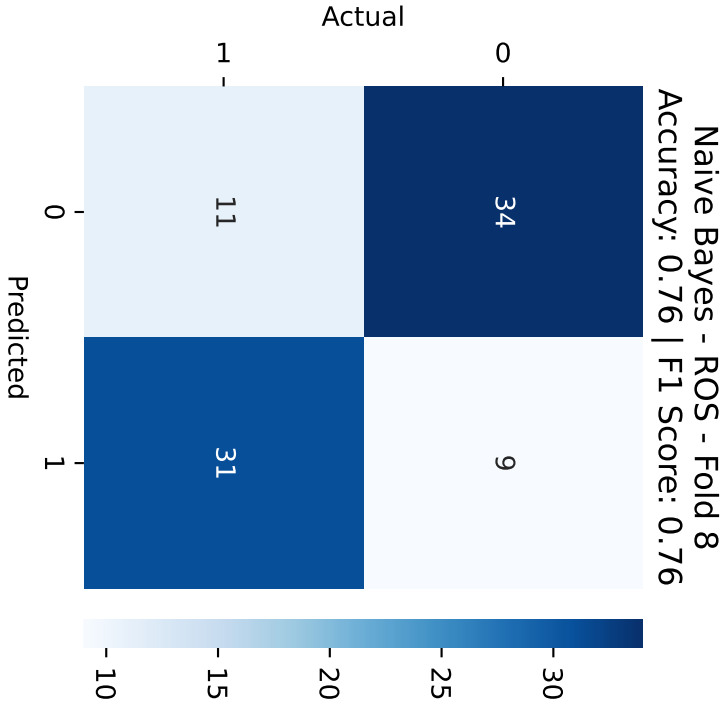
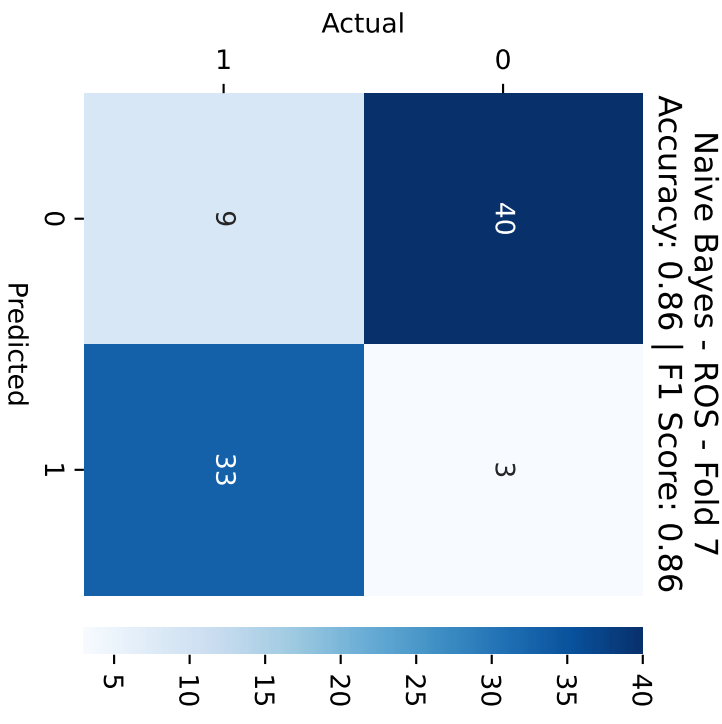
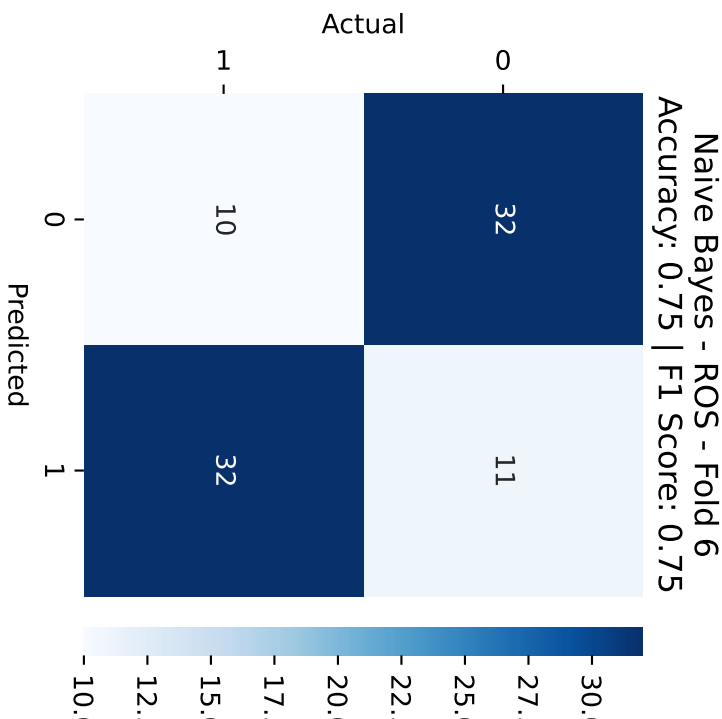
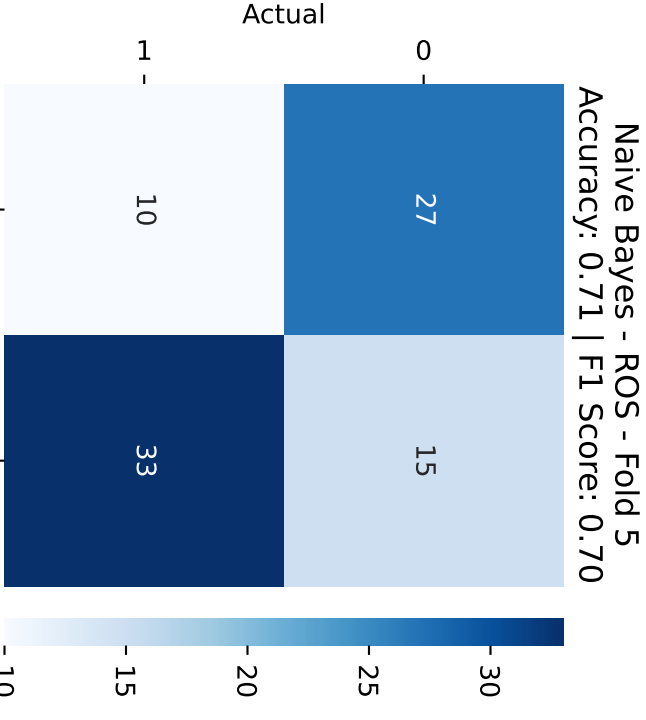
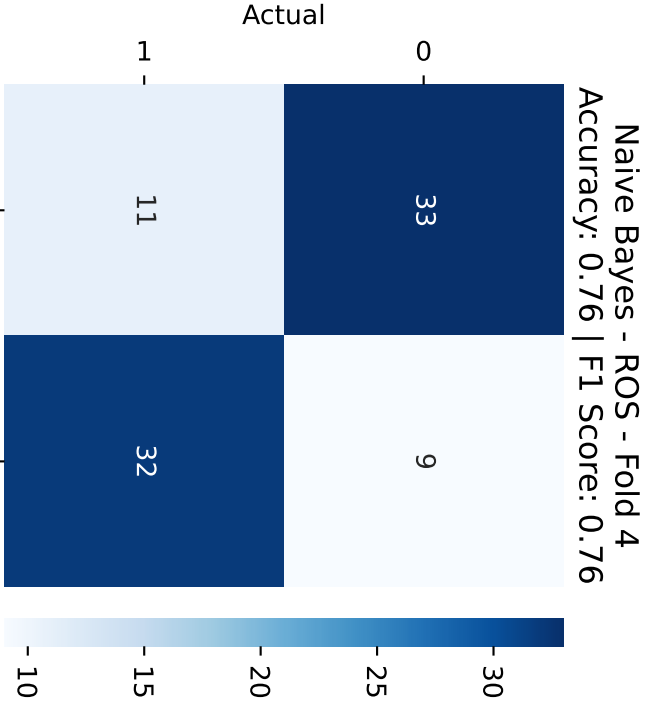
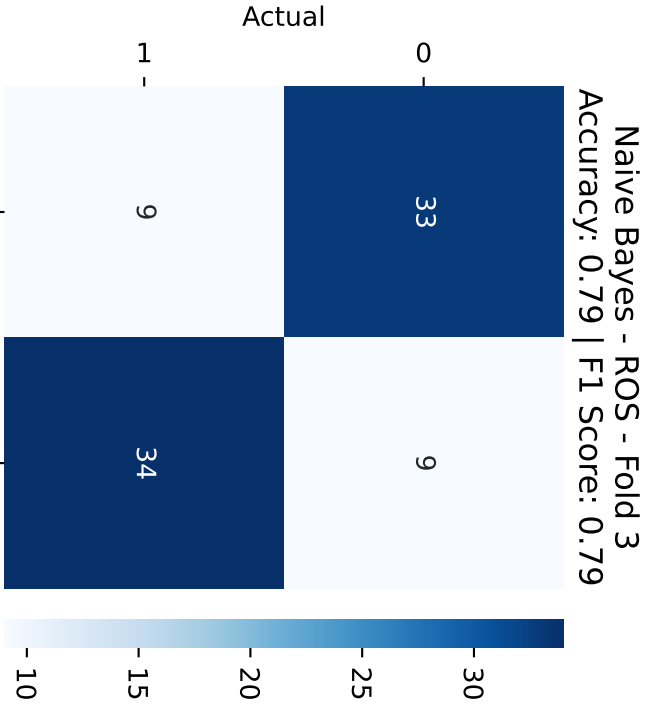
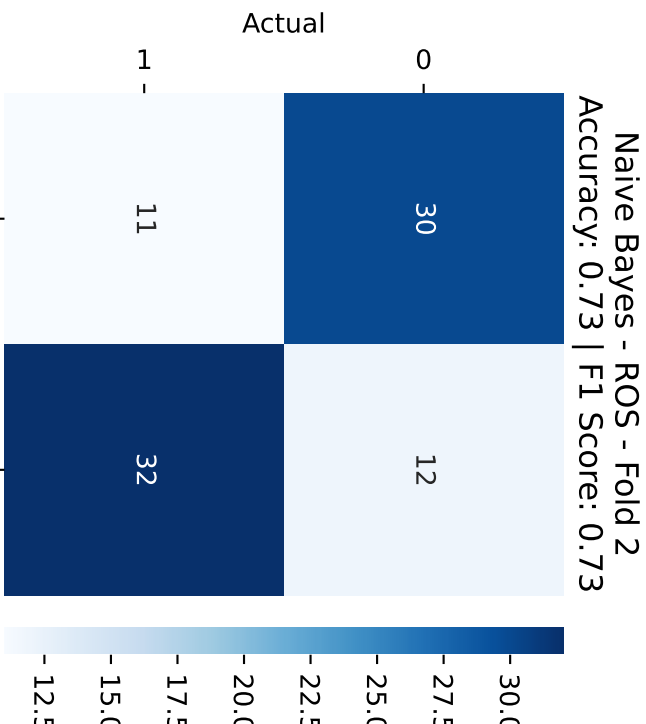
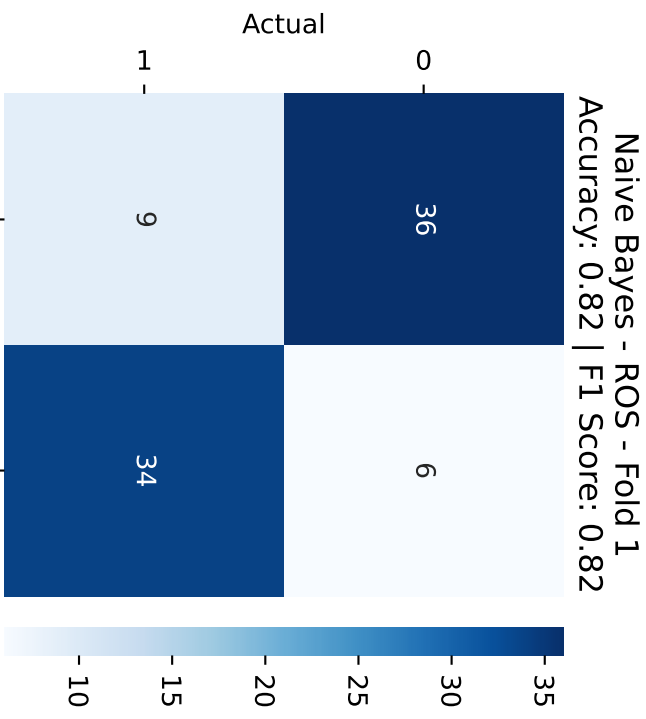


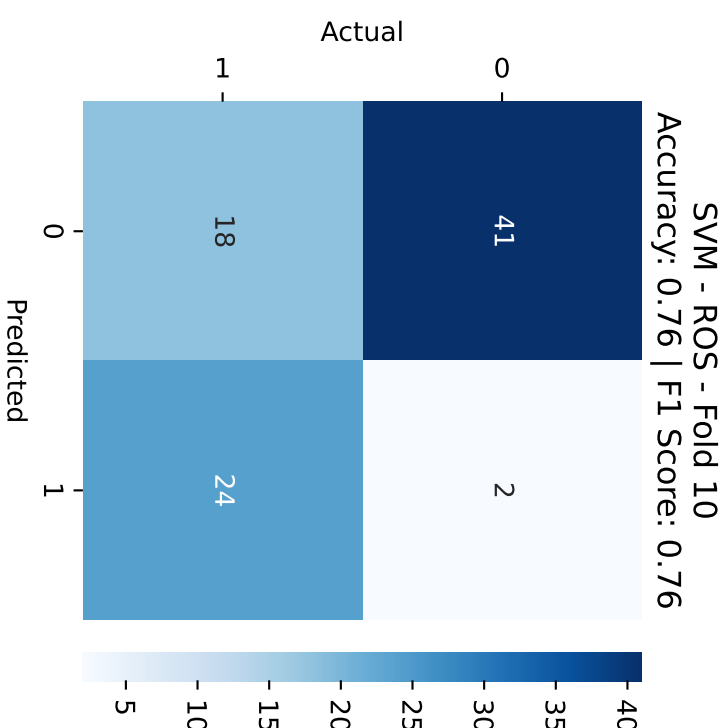
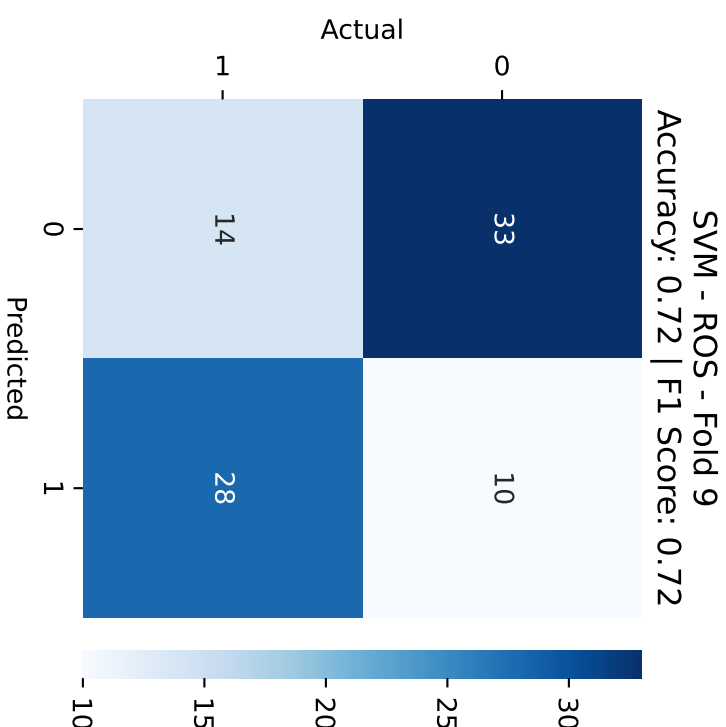
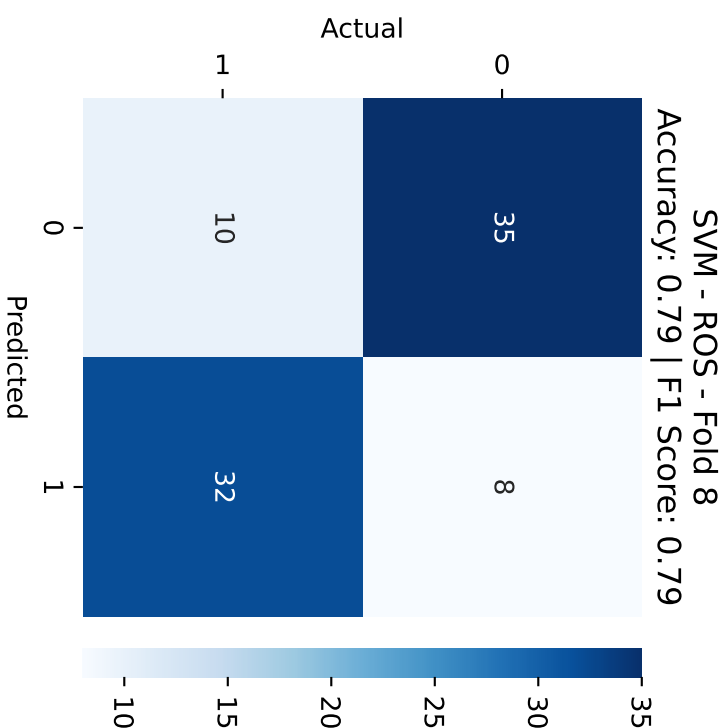
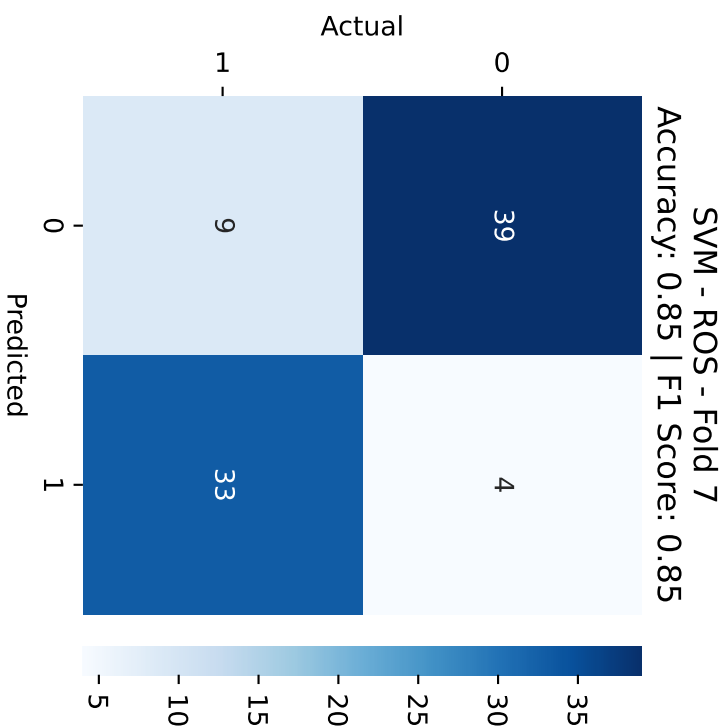
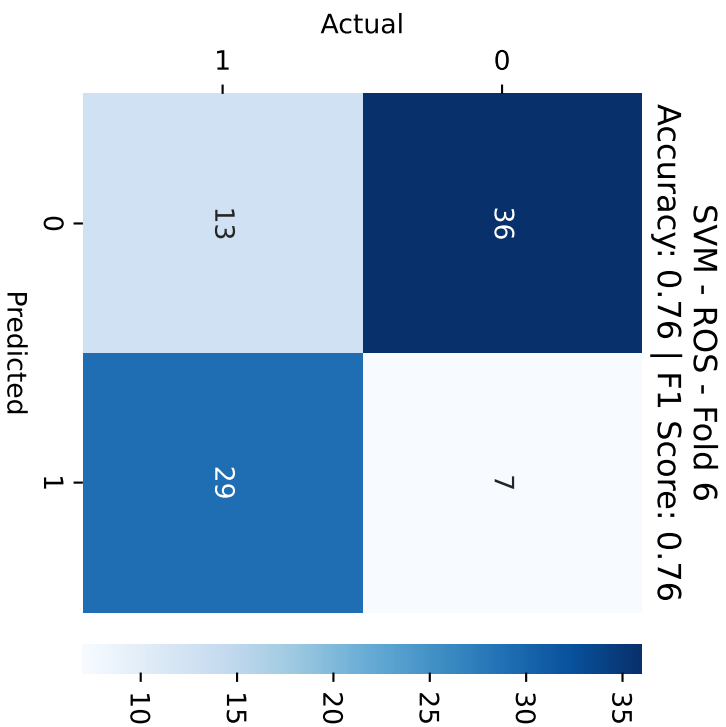
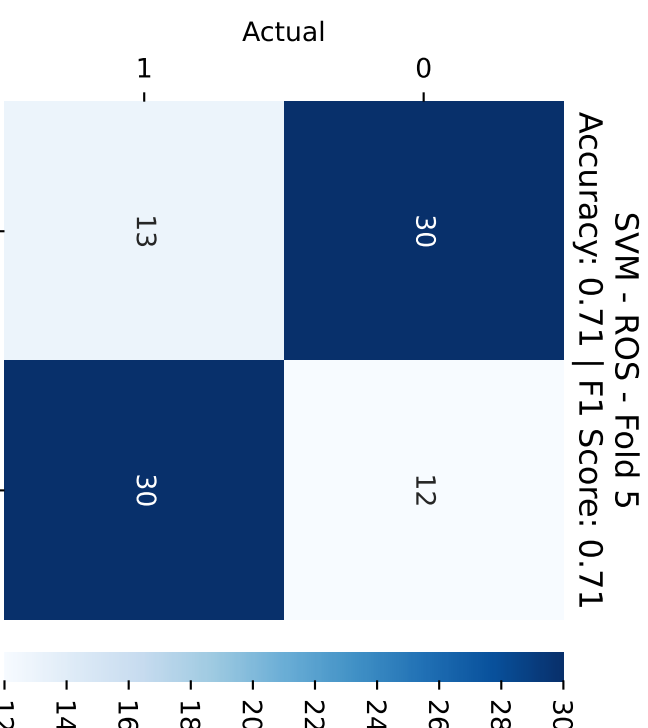
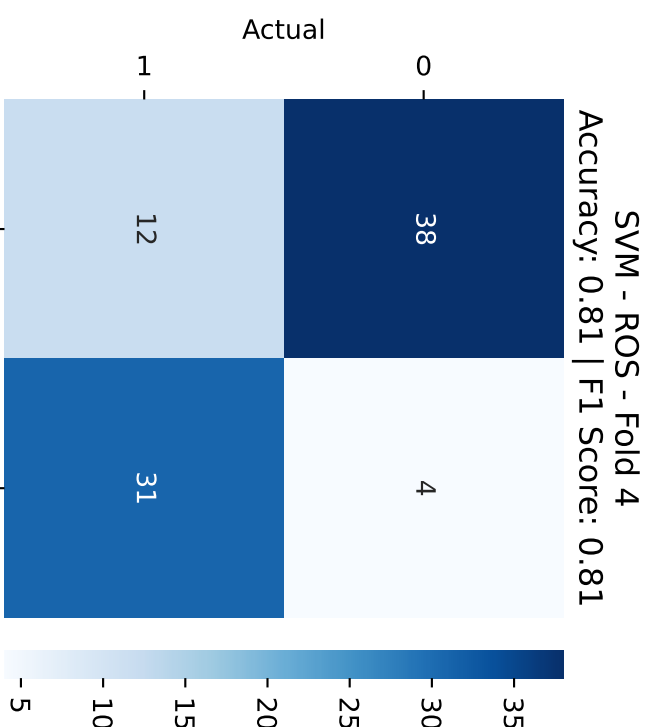
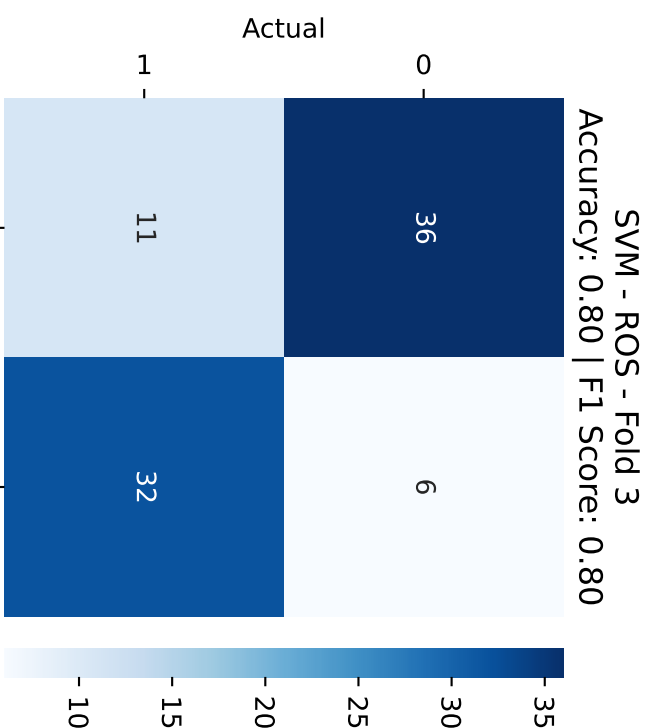
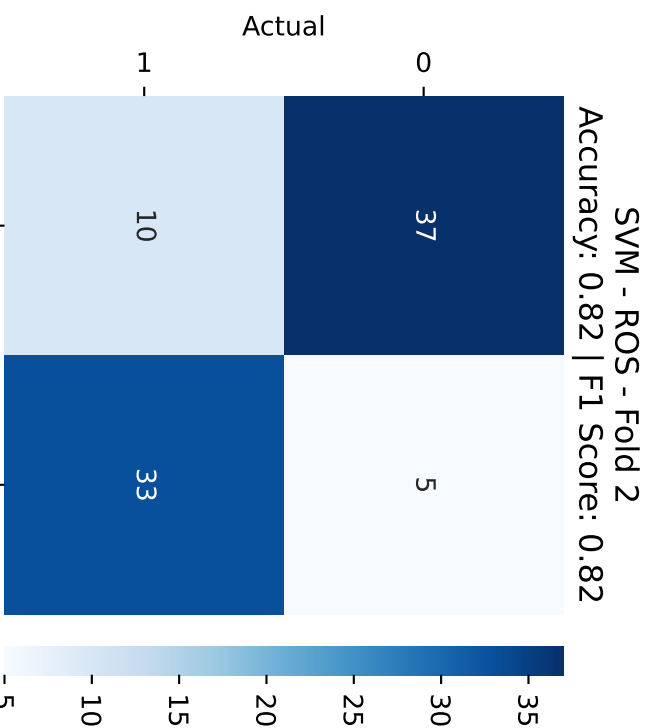
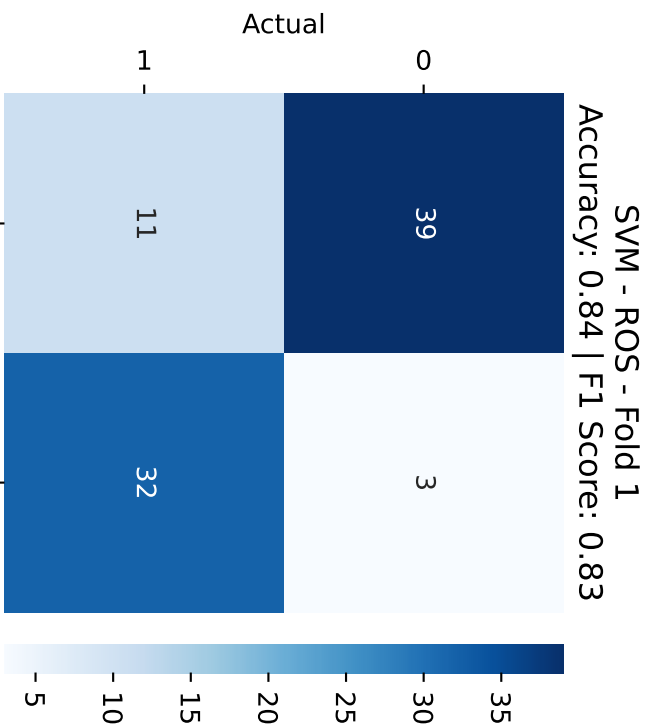


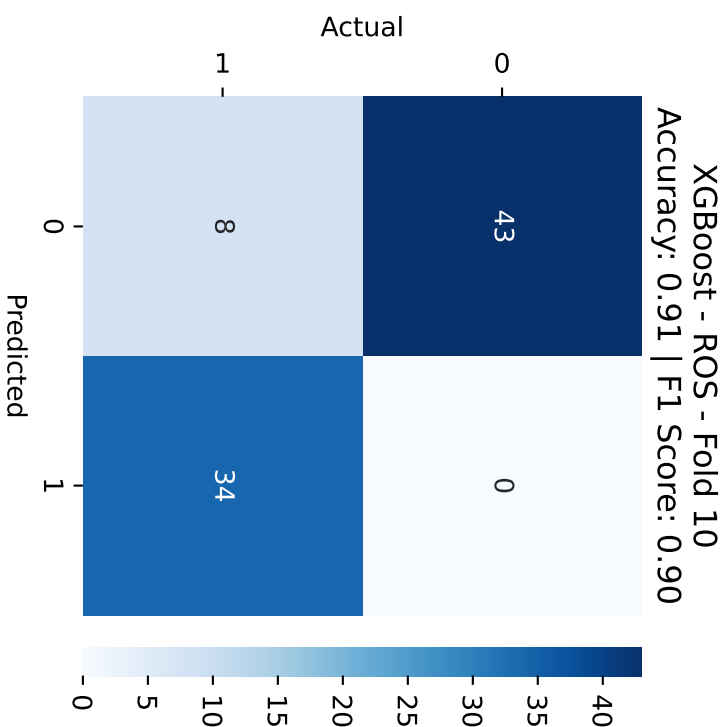
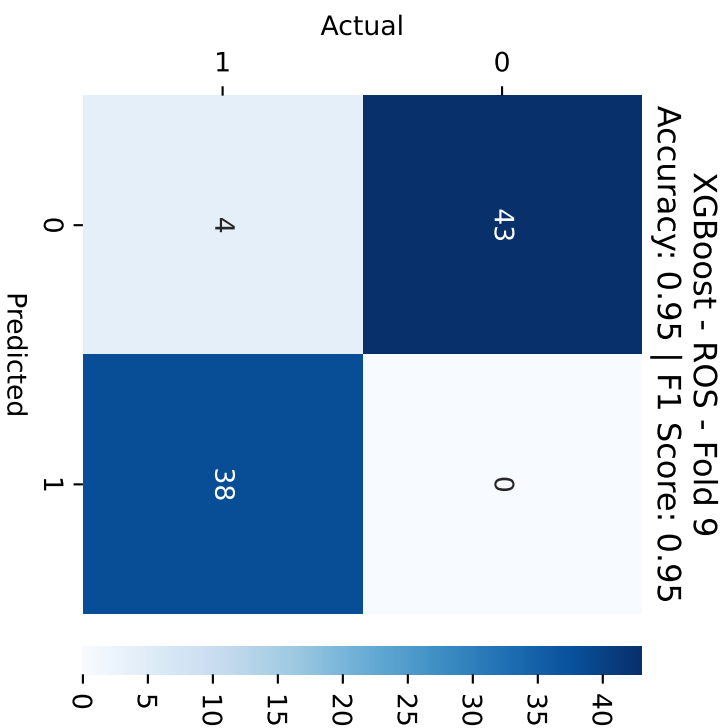
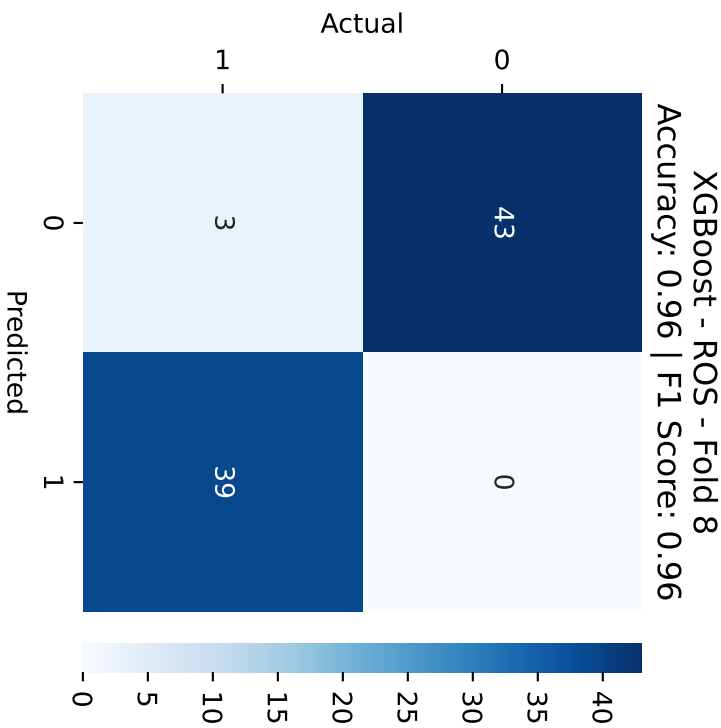
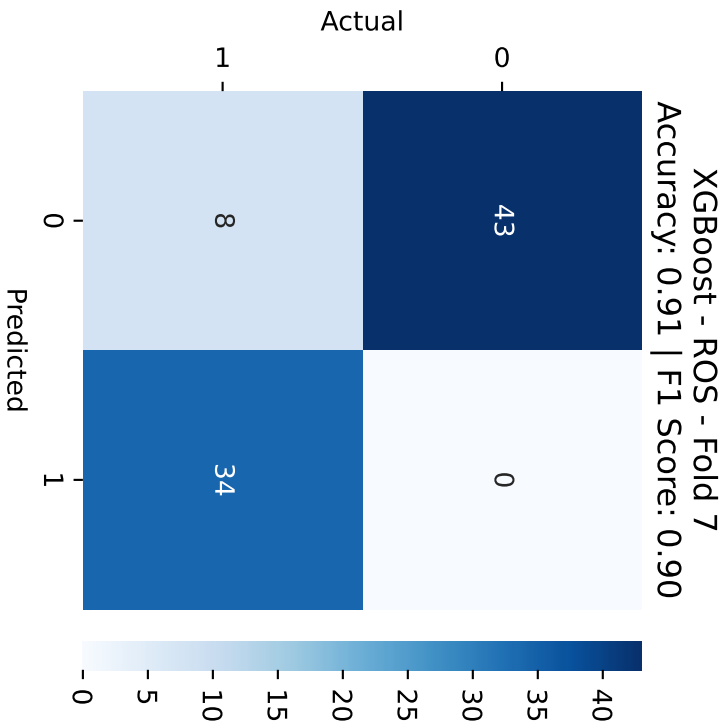
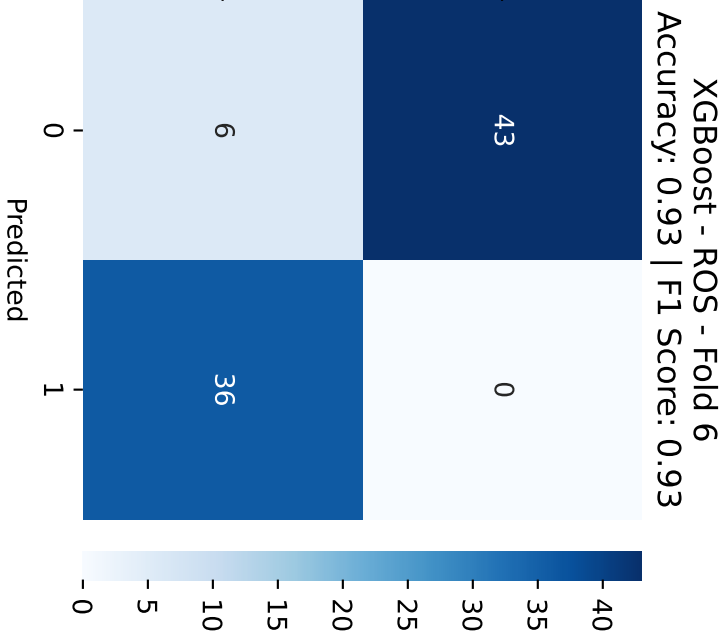
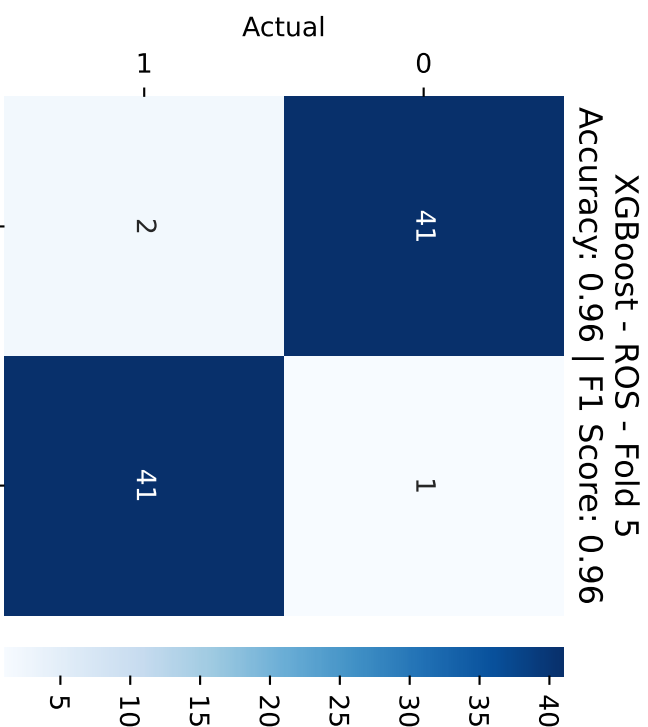
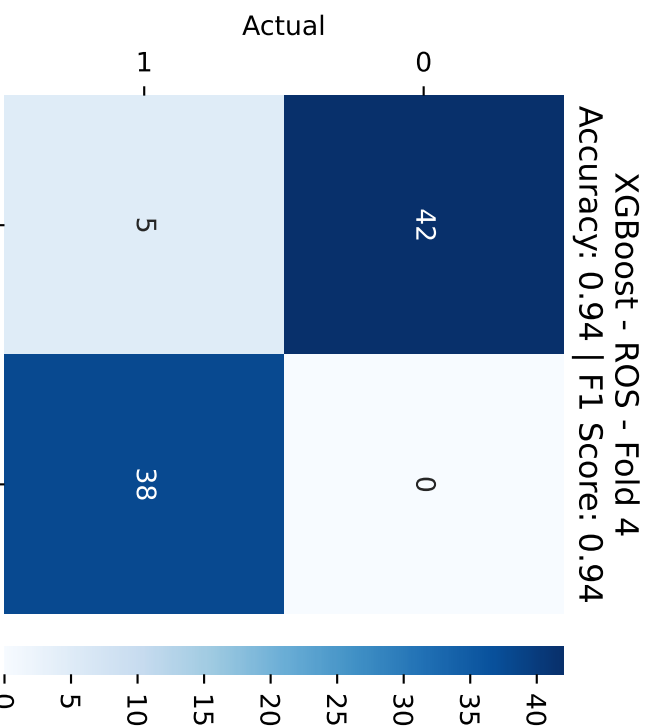
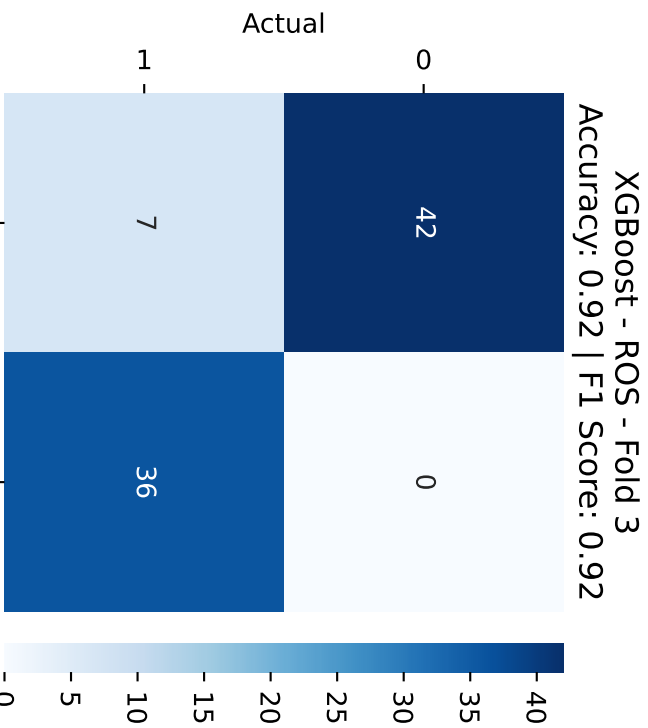
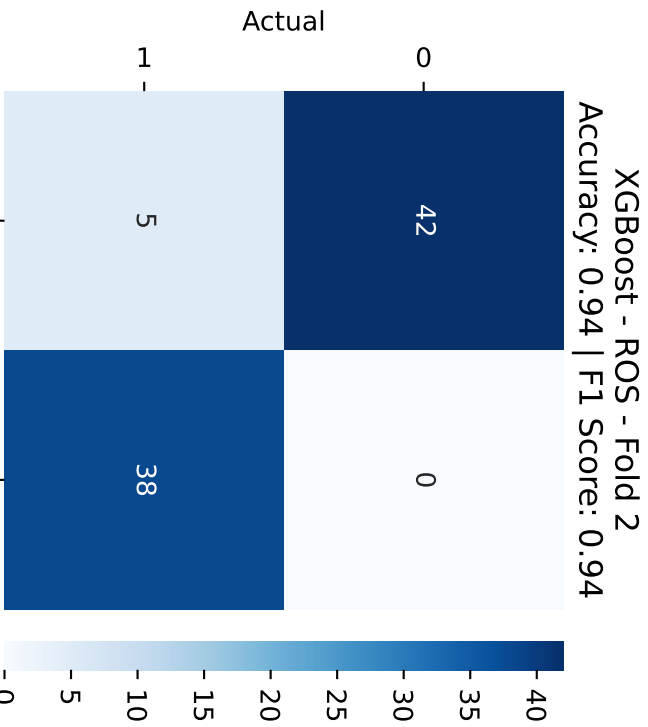
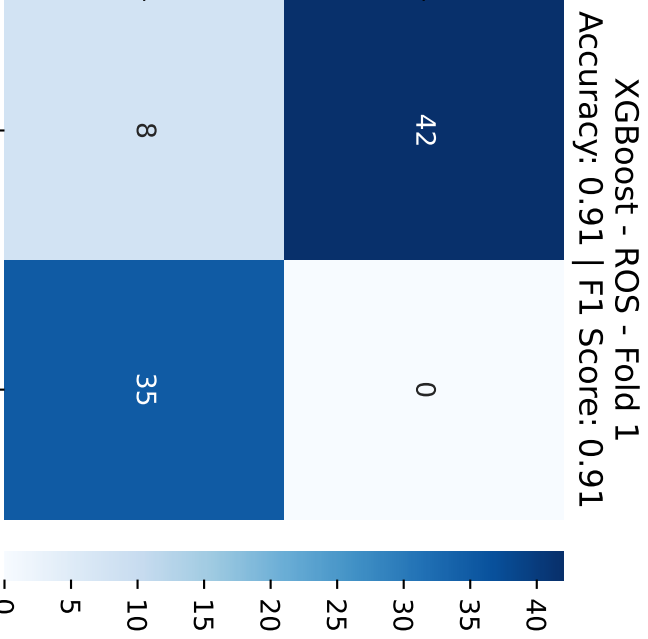


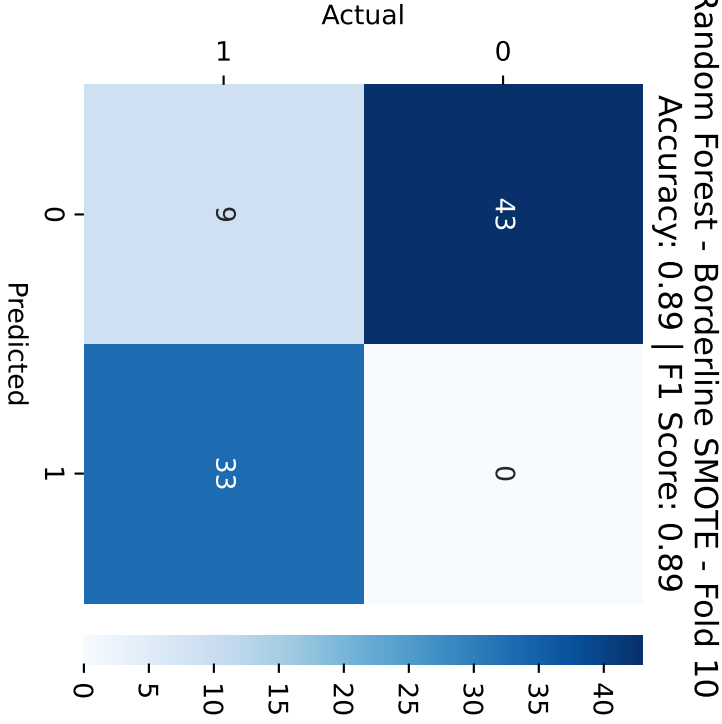
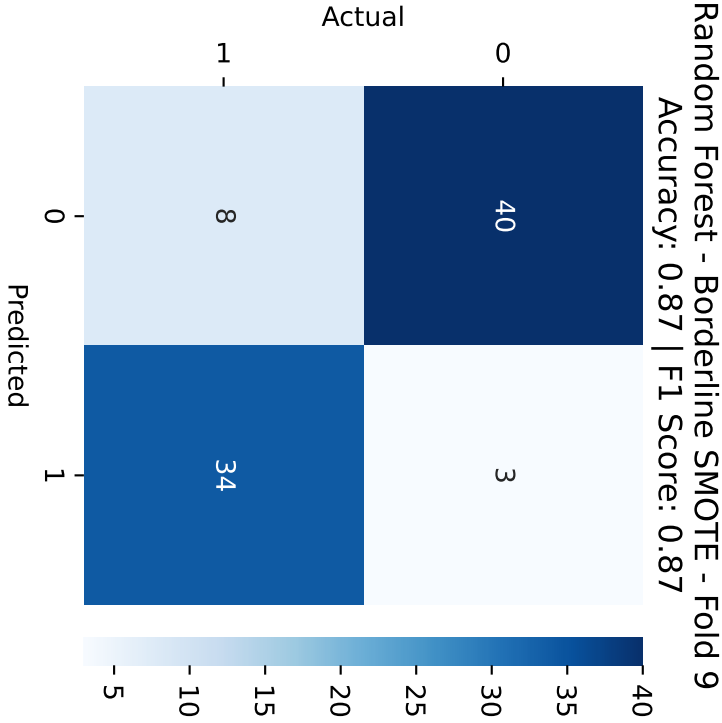
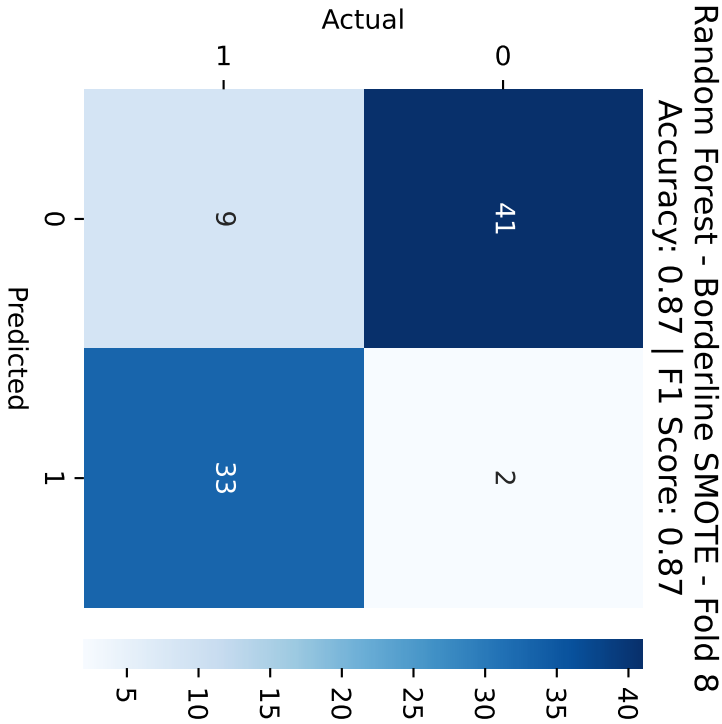
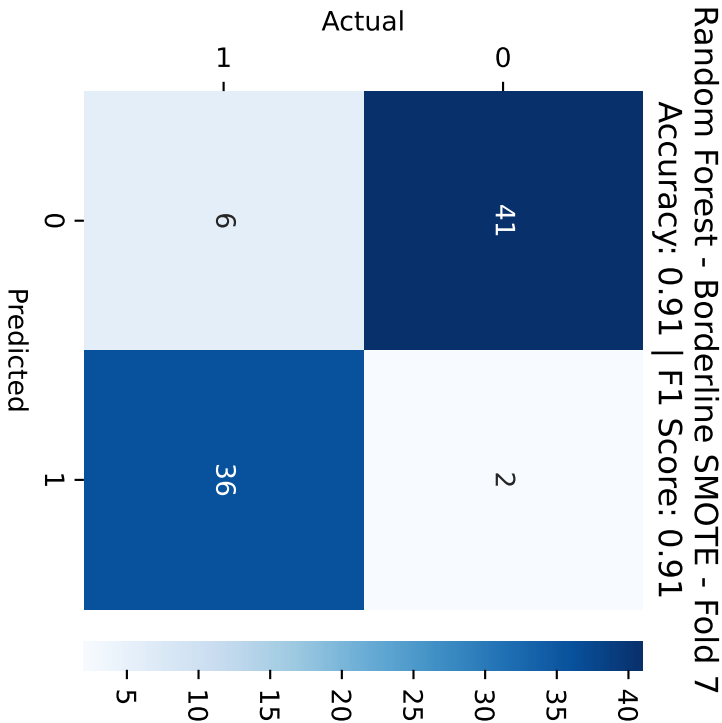
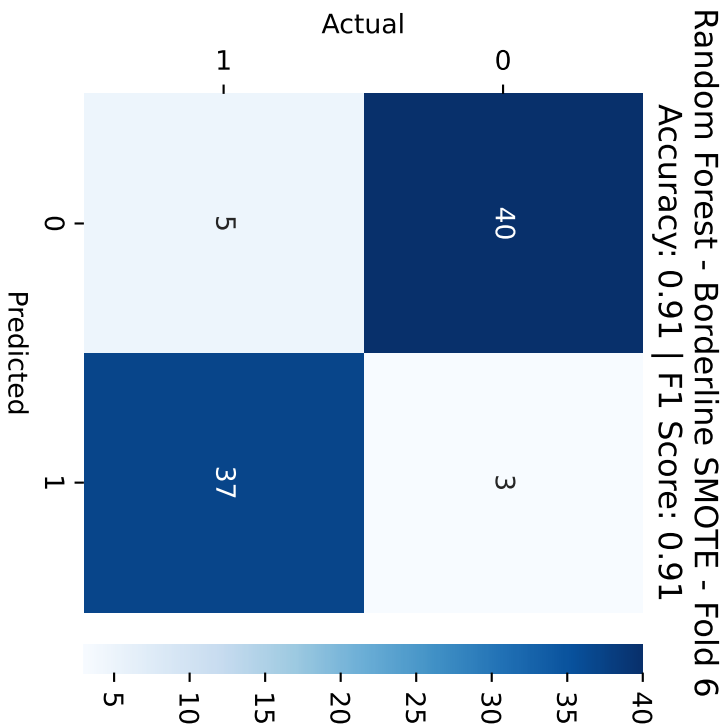
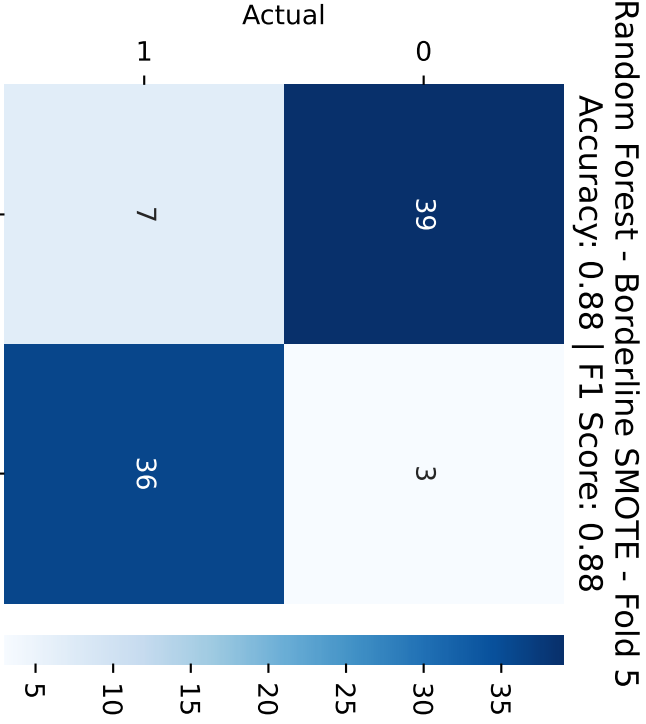
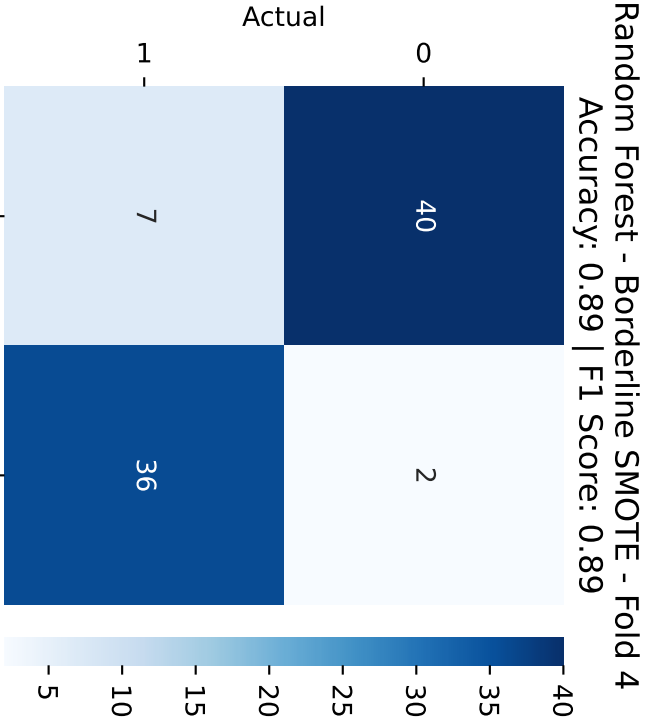
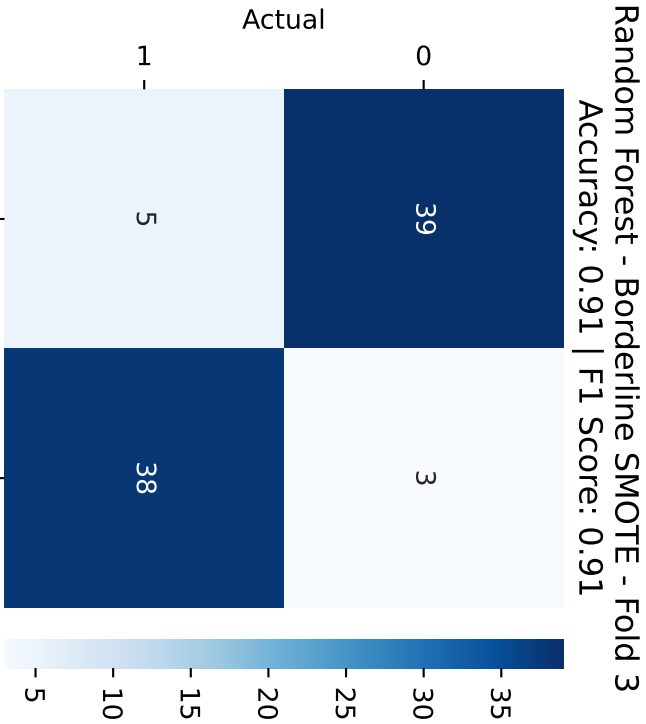
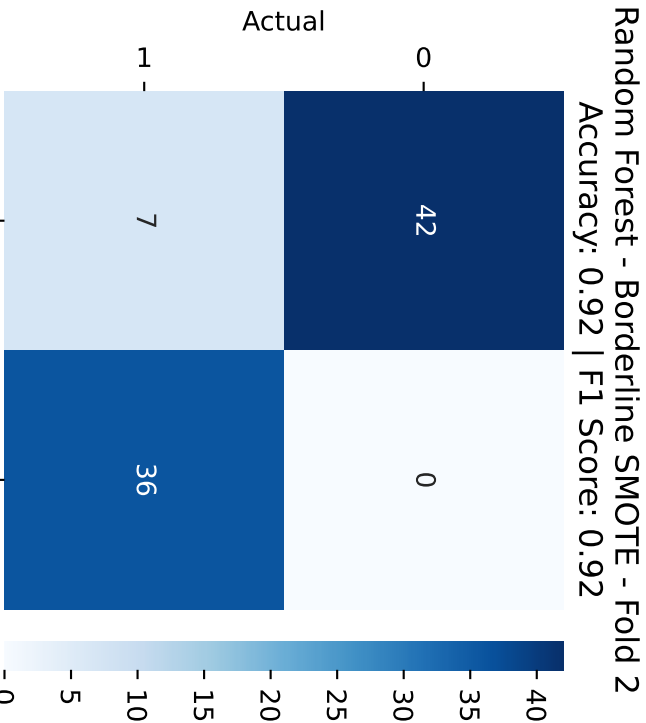
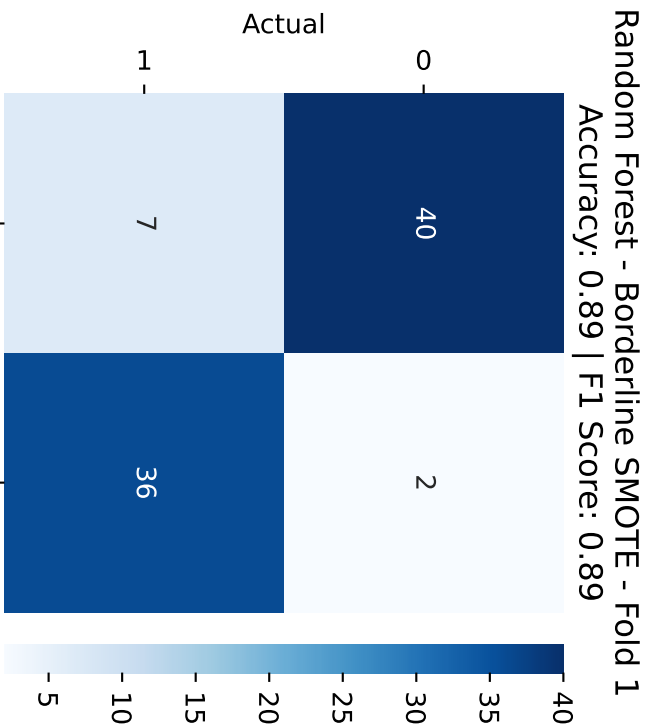


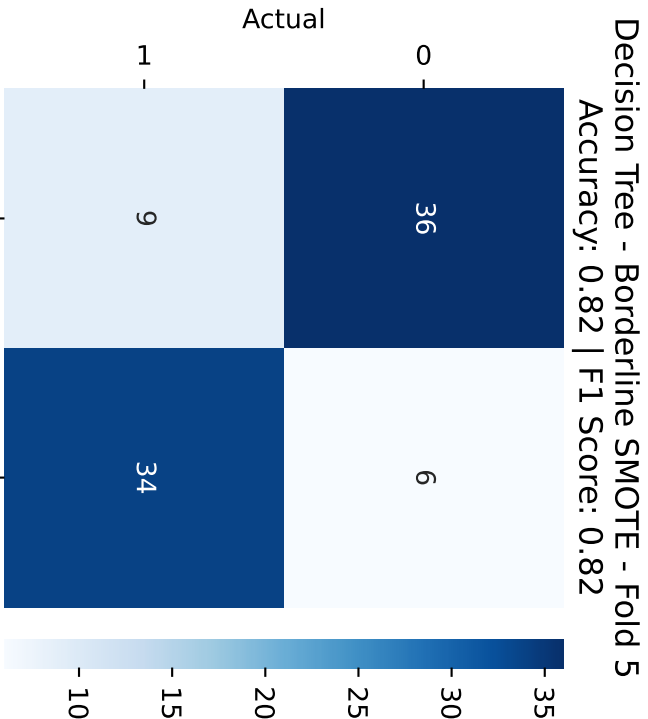
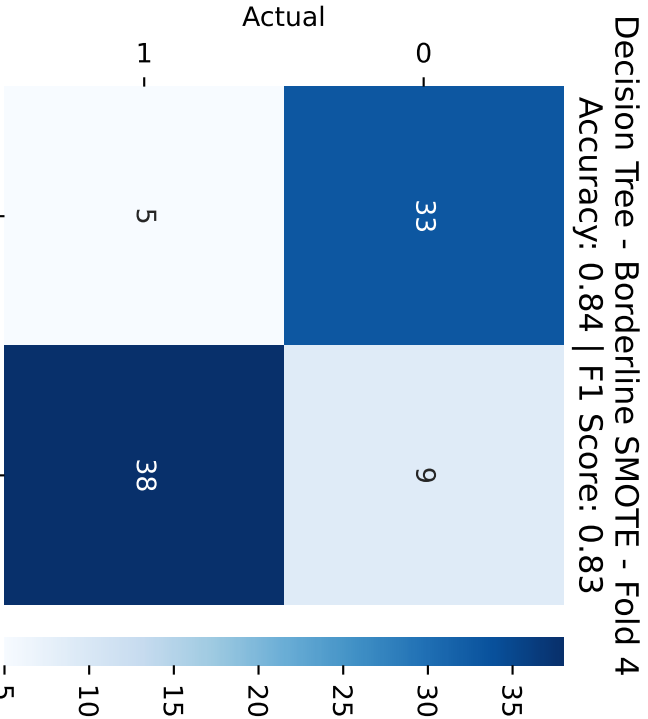
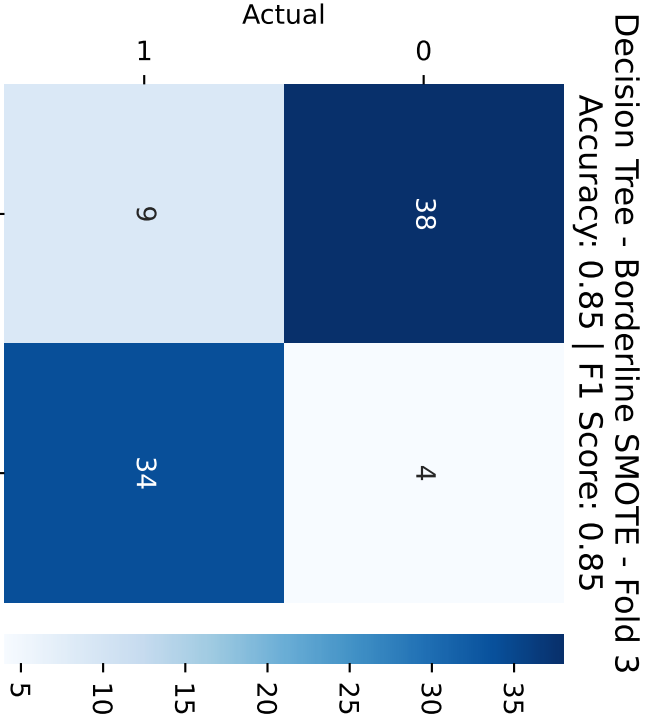
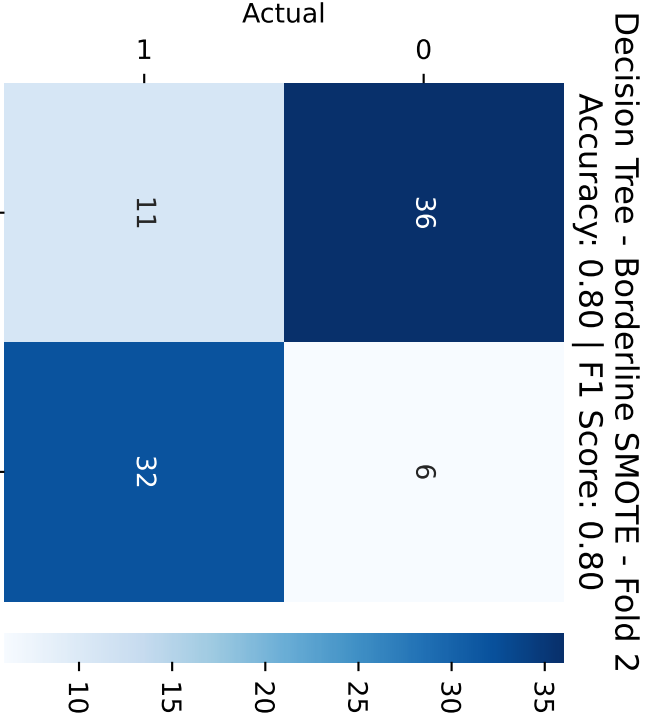
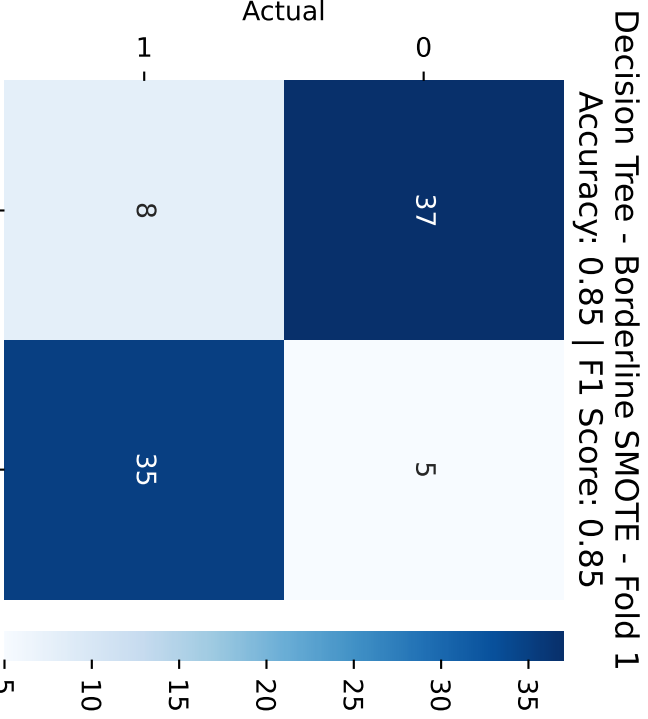




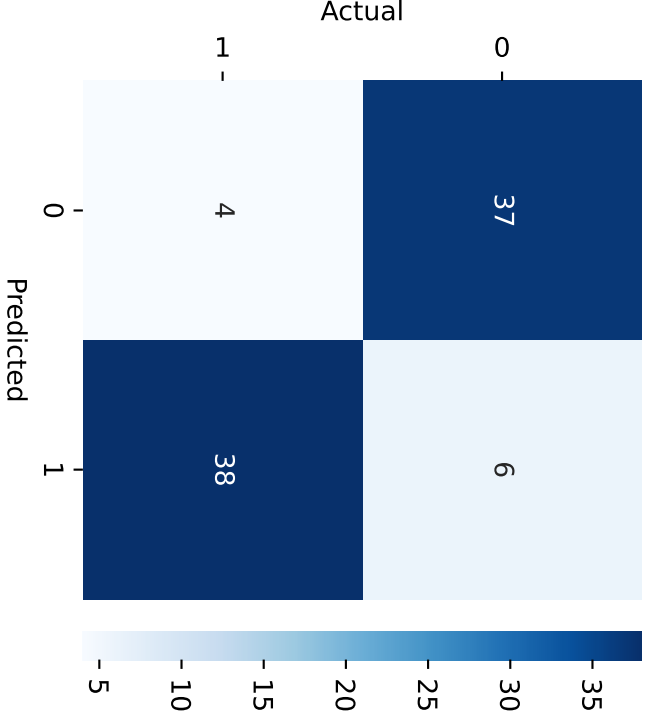




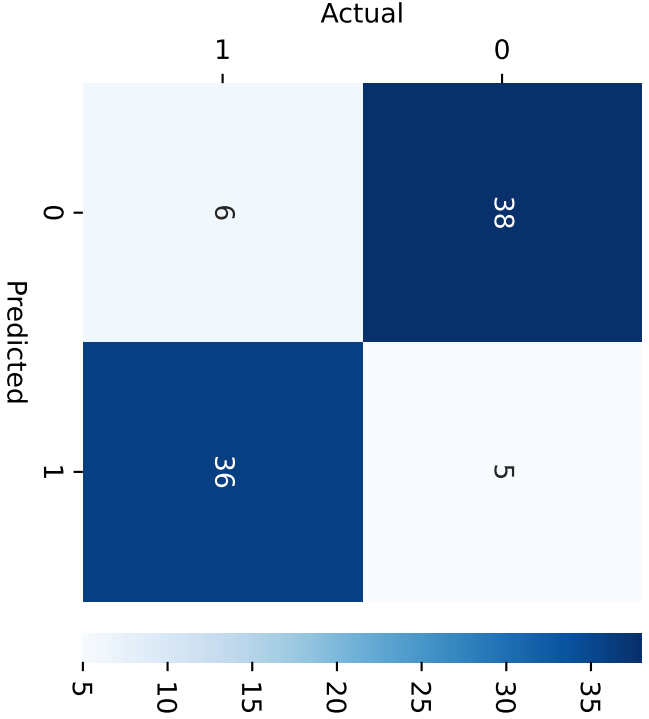




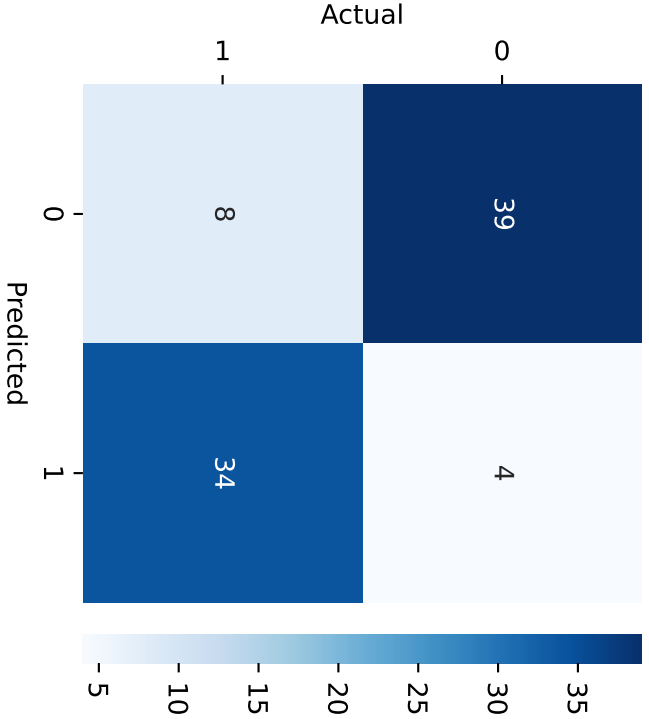
Decision Tree - Borderline SMOTE - Fold 6  
Accuracy: 0.88 | F1 Score: 0.88



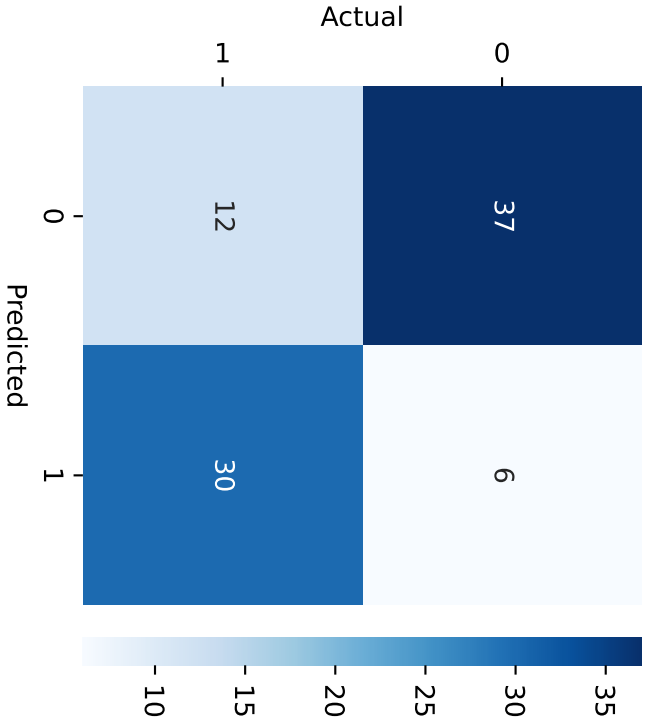
Decision Tree - Borderline SMOTE - Fold 7  
Accuracy: 0.87 | F1 Score: 0.87



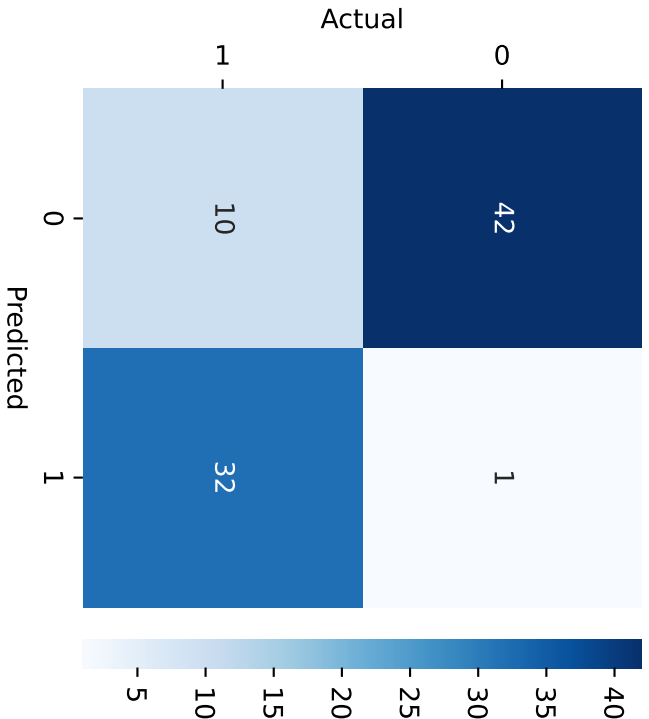
Decision Tree - Borderline SMOTE - Fold 8  
Accuracy: 0.86 | F1 Score: 0.86

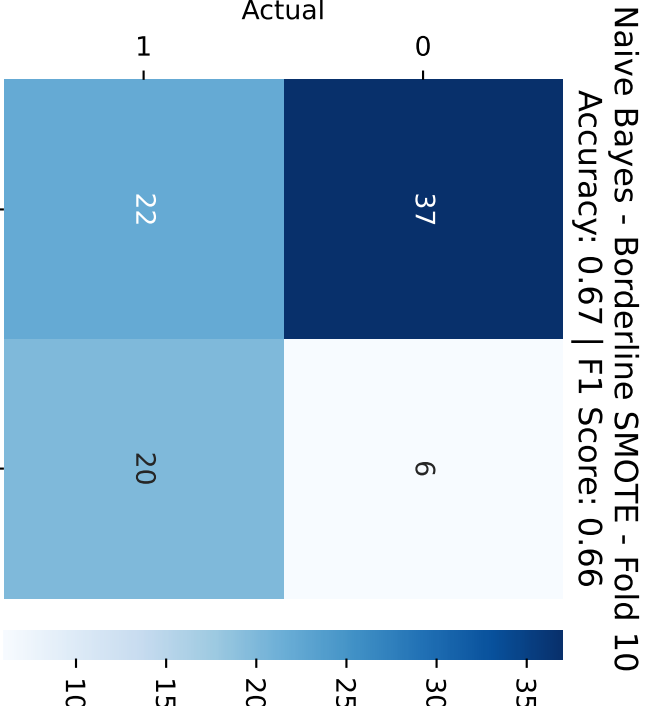
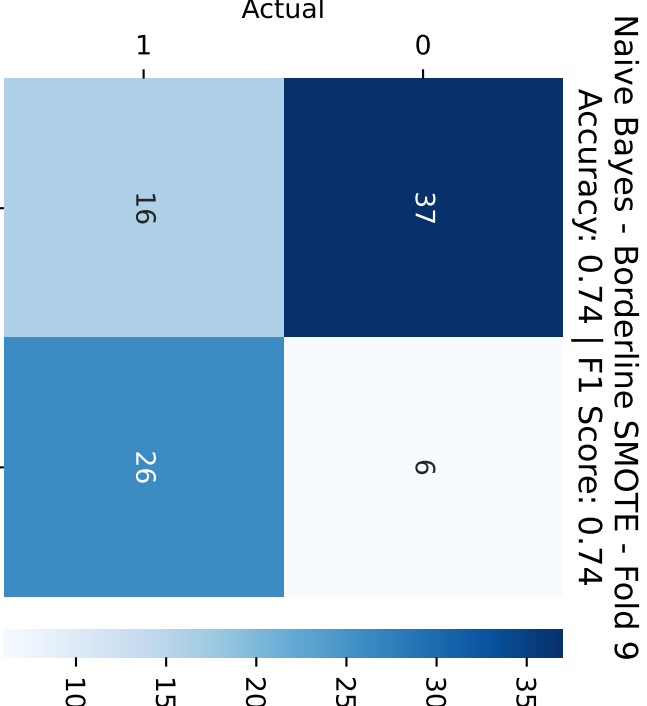
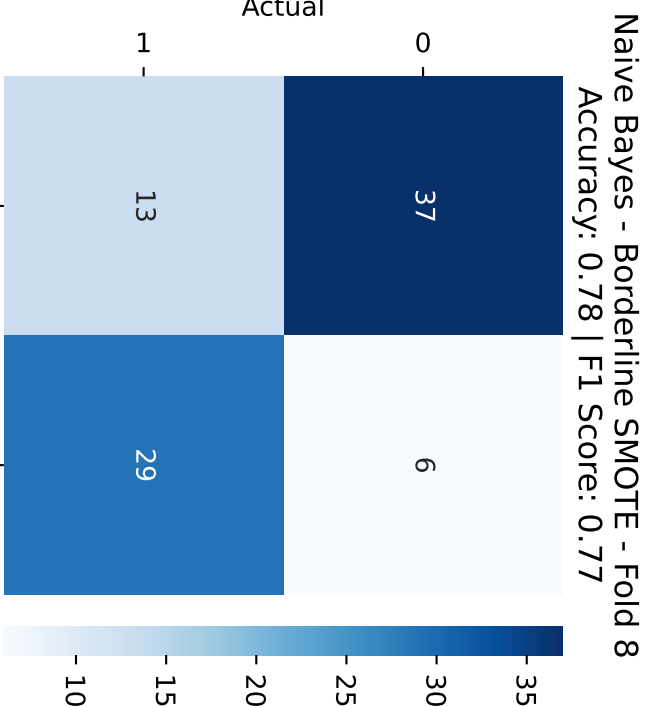
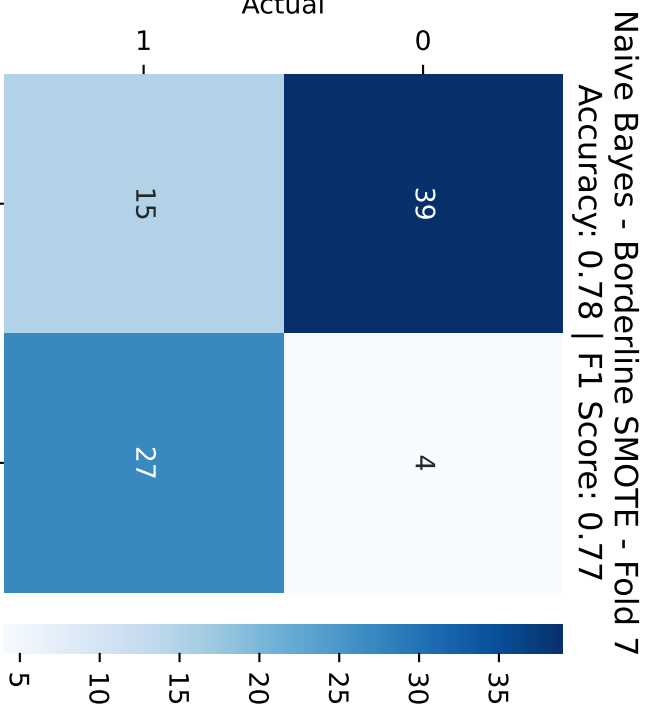
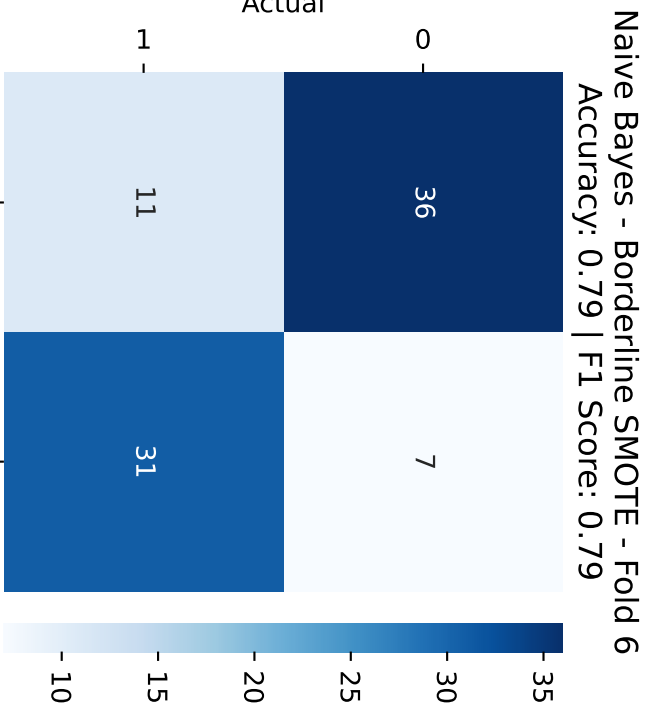
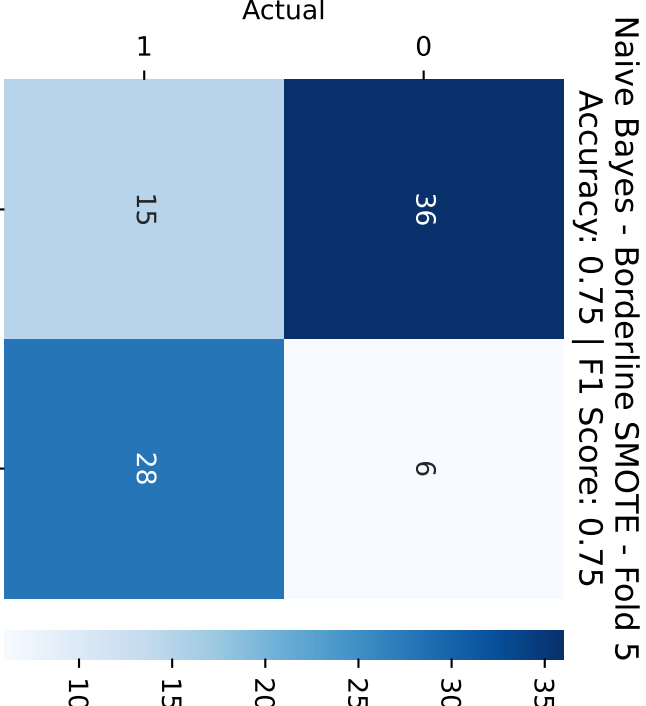
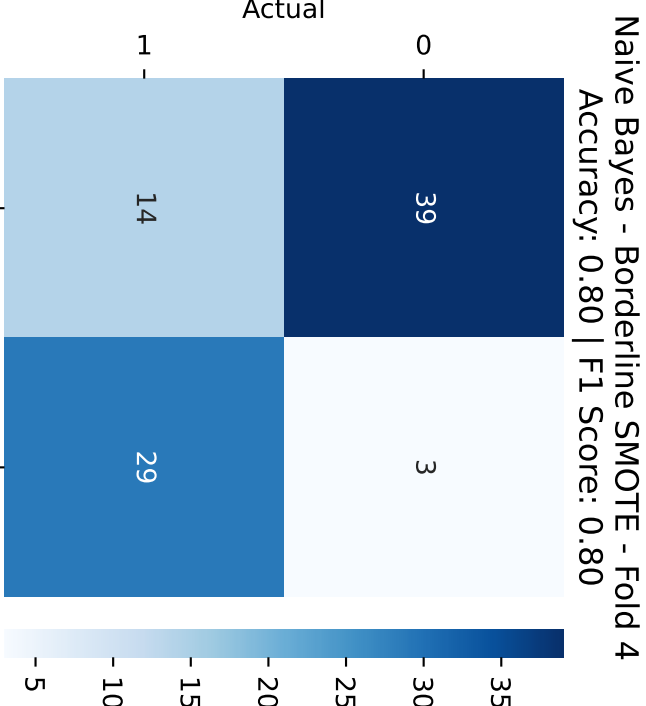
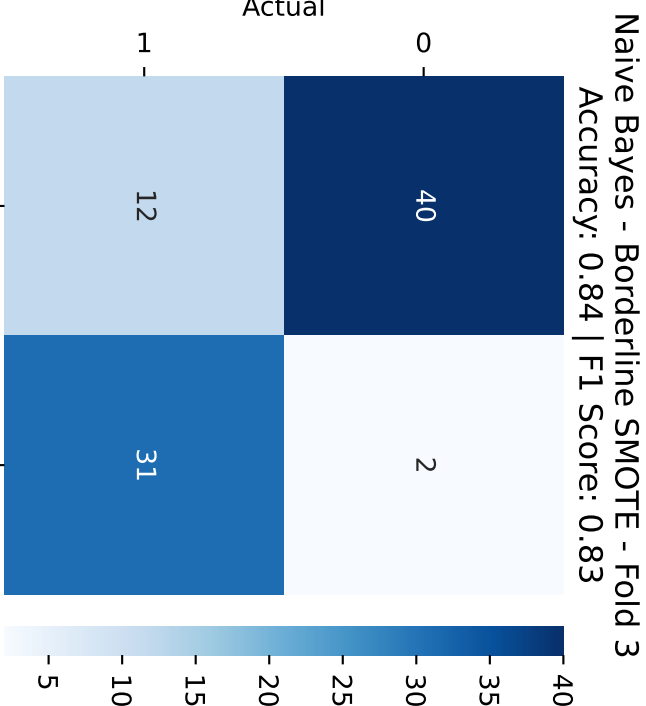
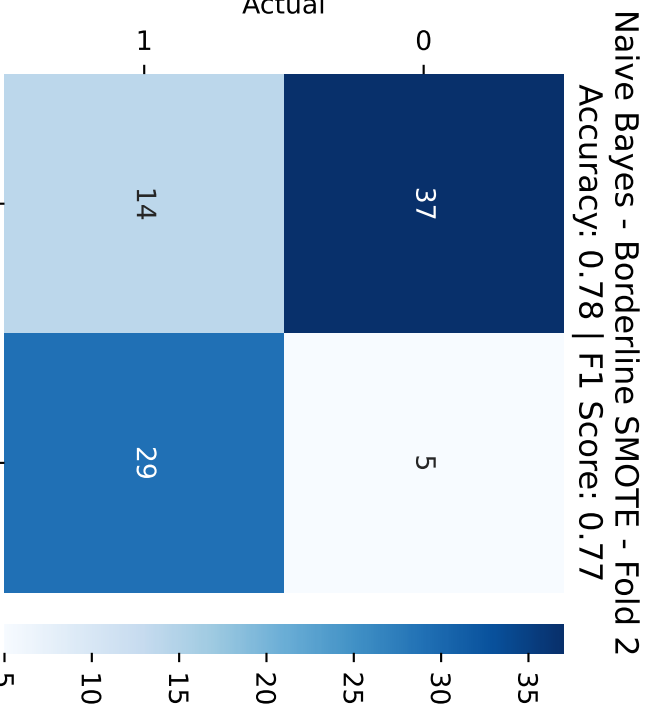
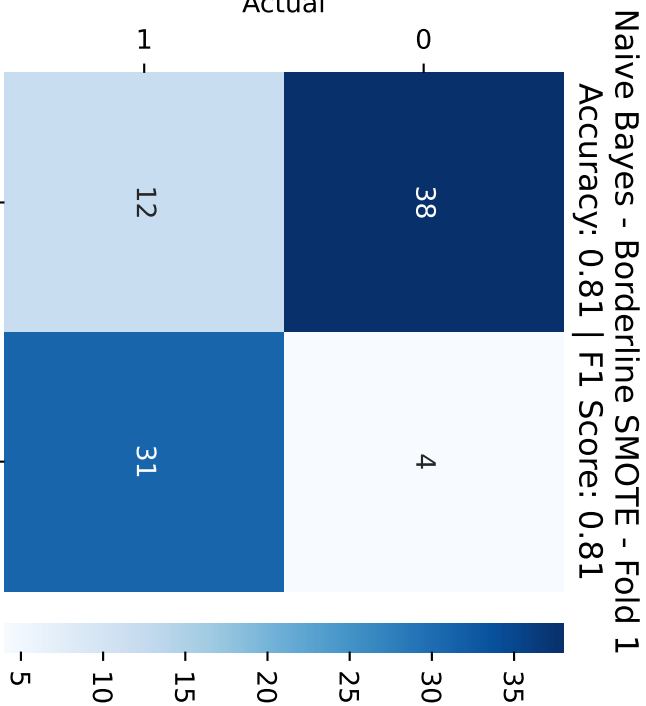


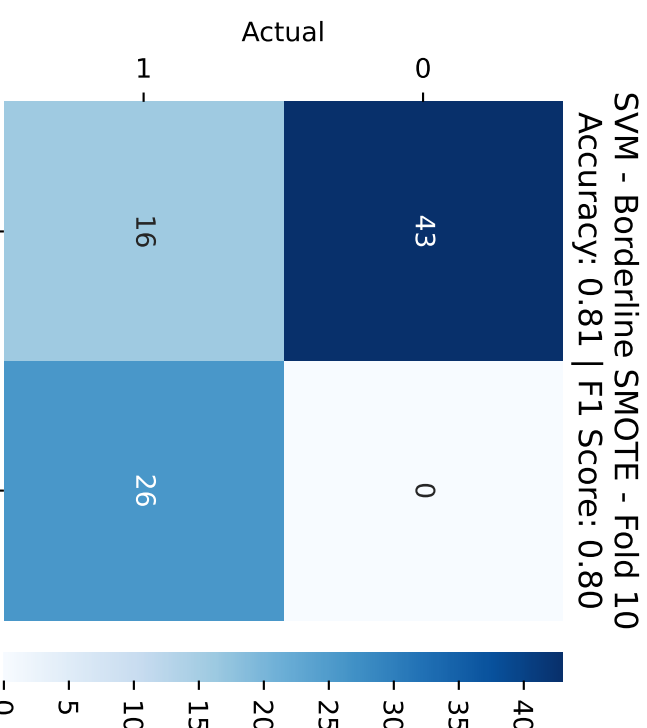
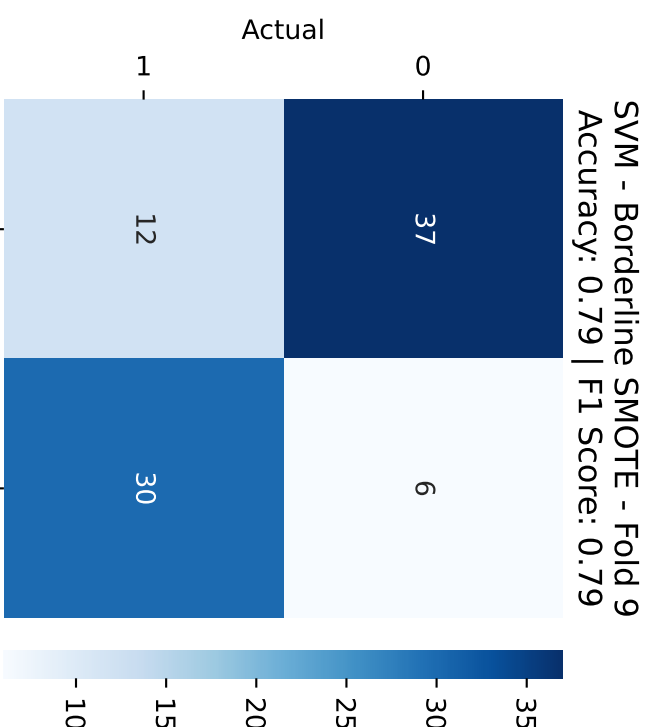
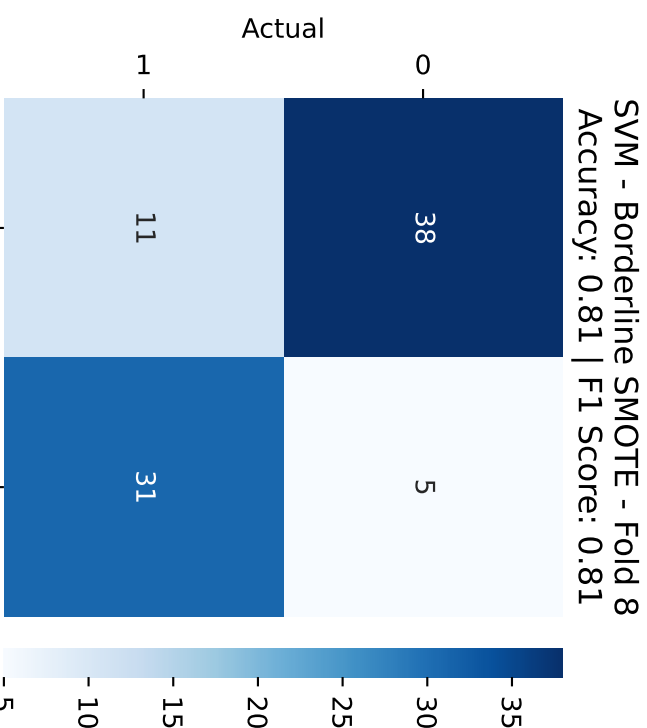
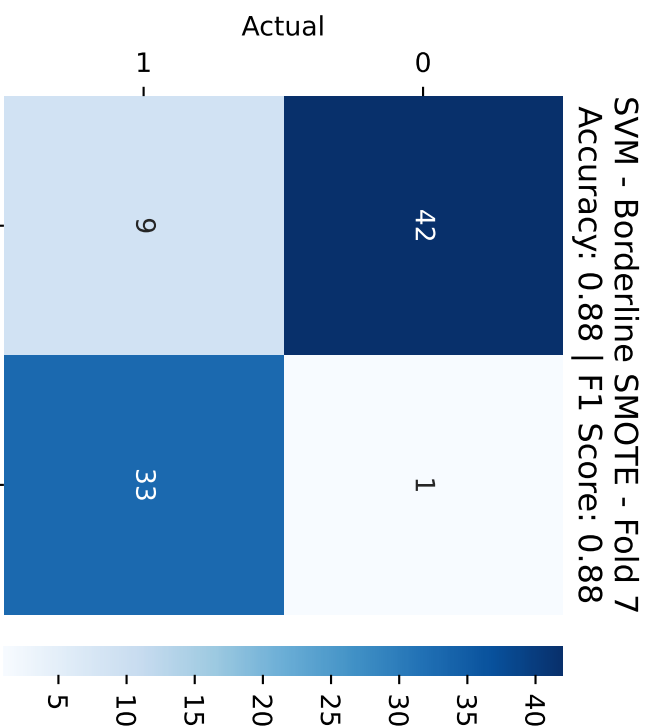
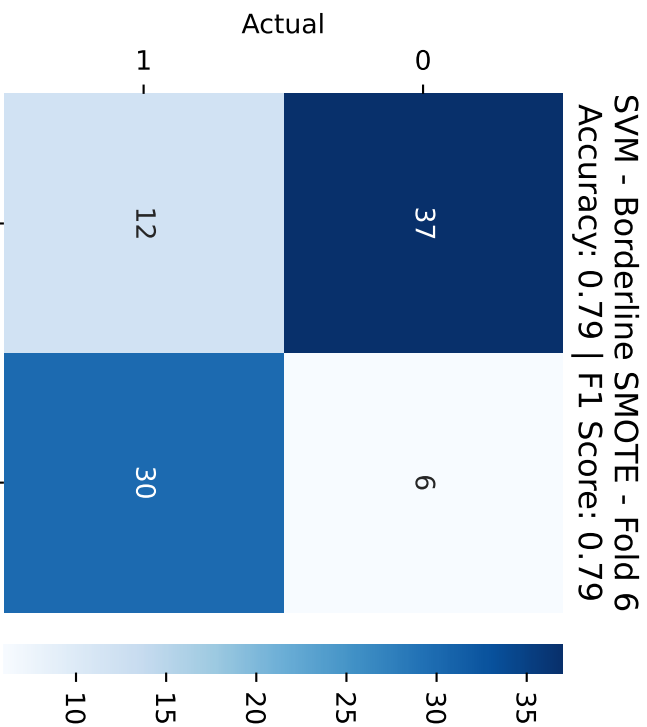
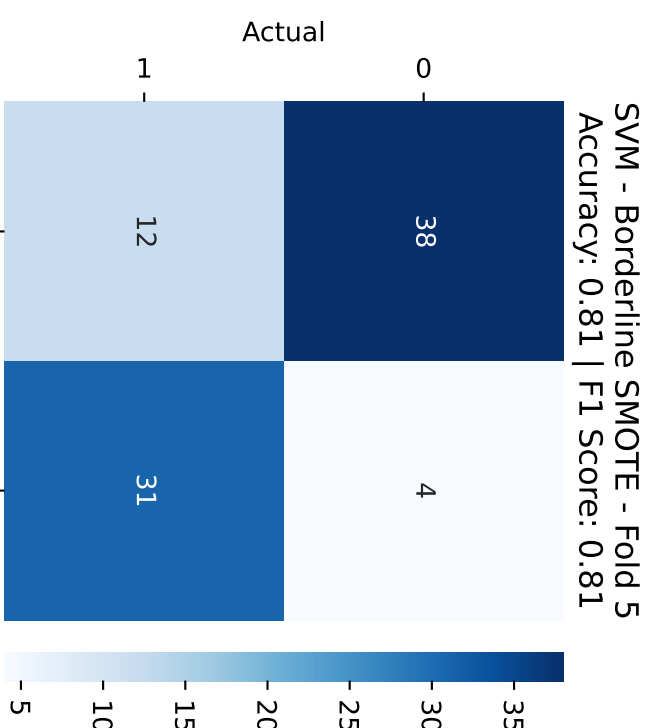
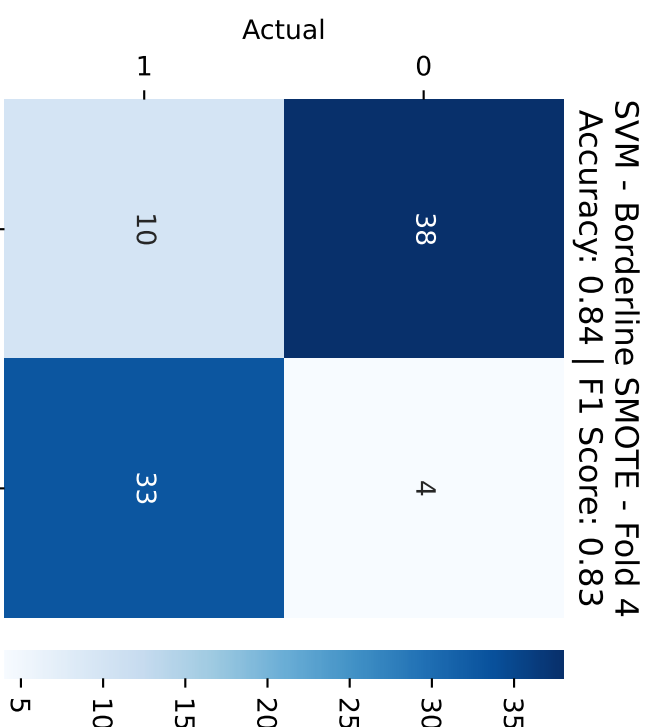
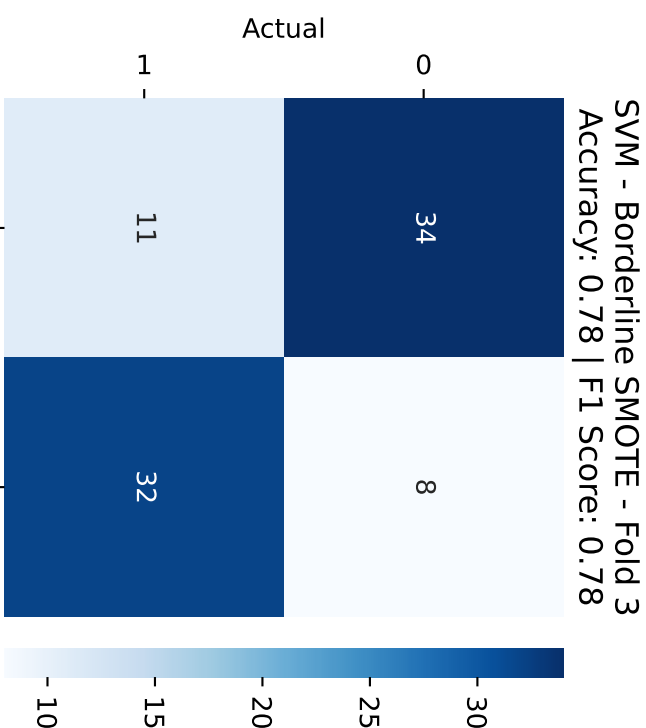
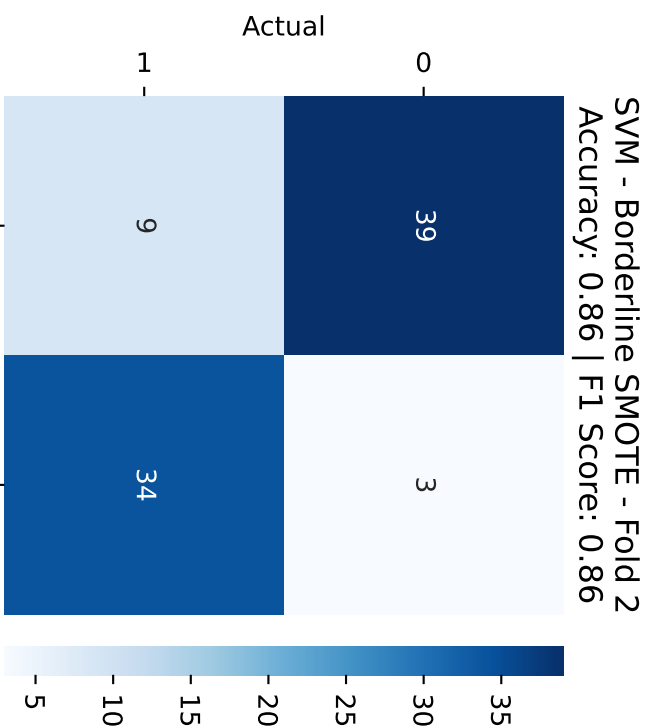
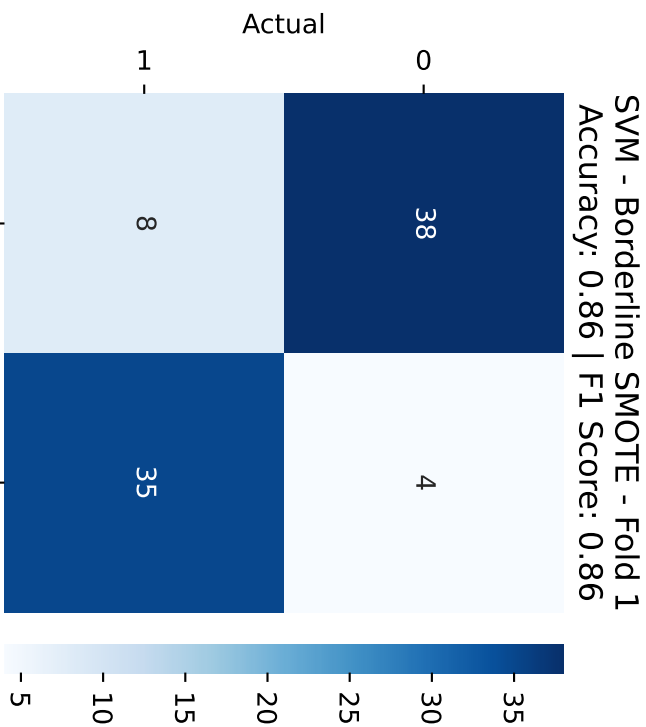
Decision Tree - Borderline SMOTE - Fold 9  
Accuracy: 0.79 | F1 Score: 0.79

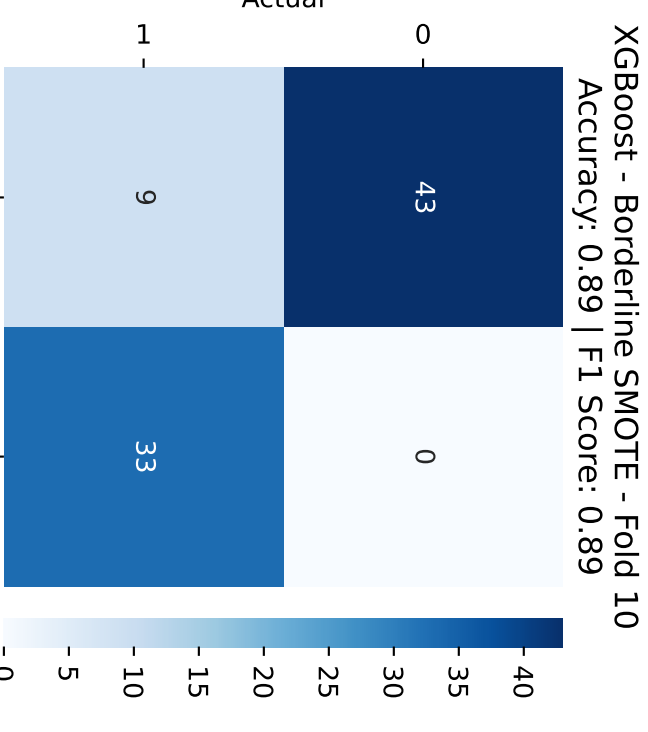
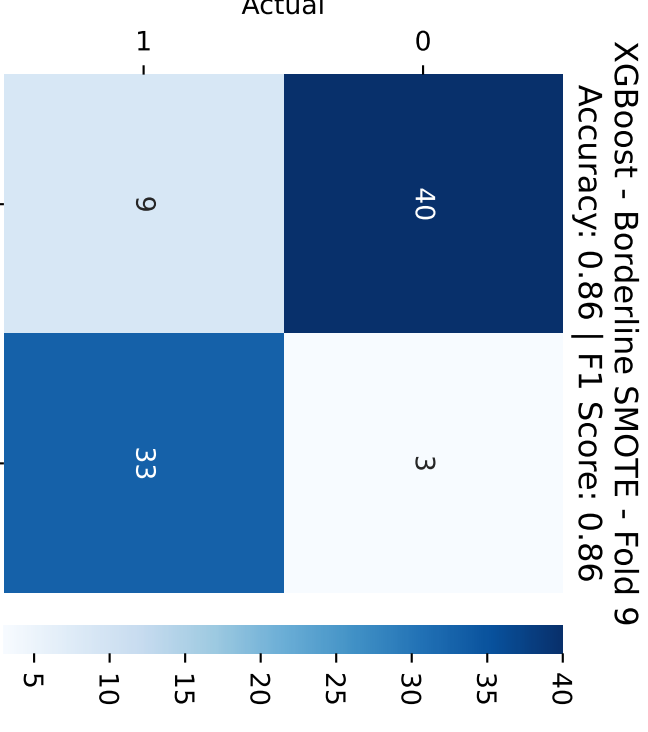
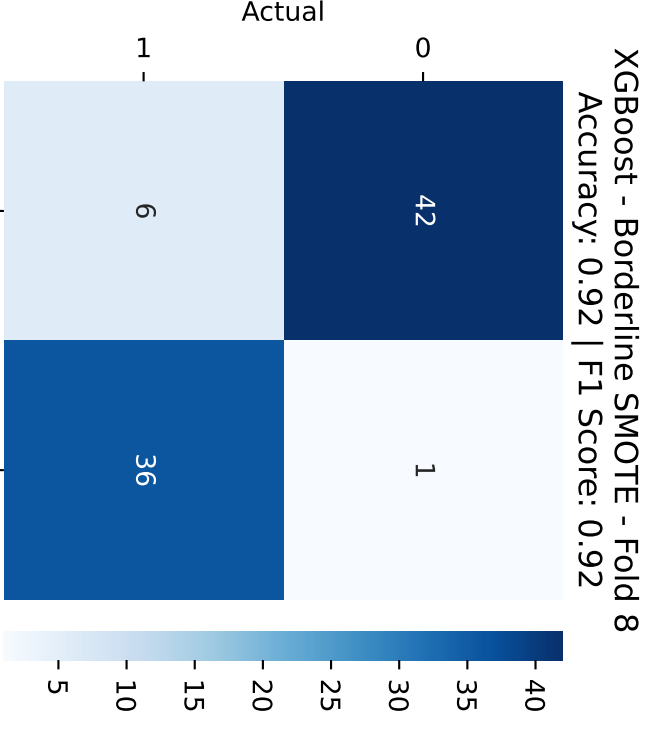
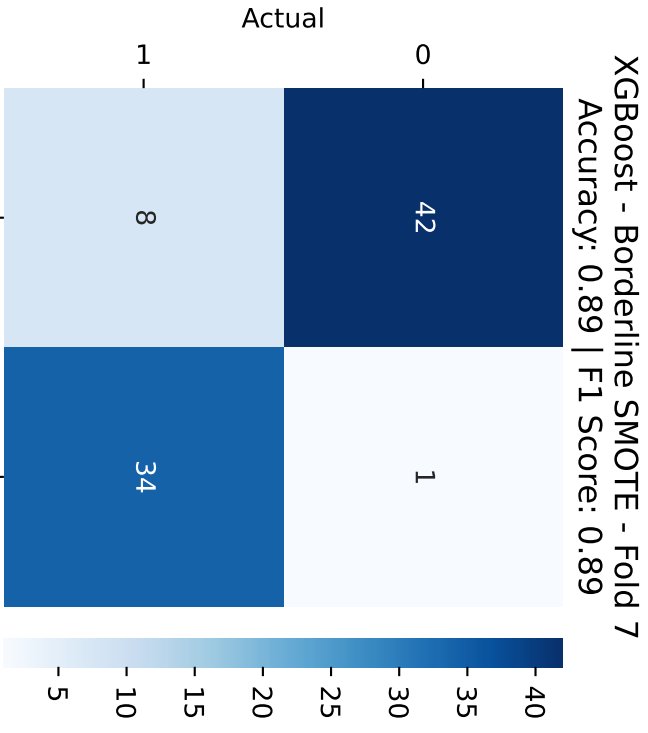
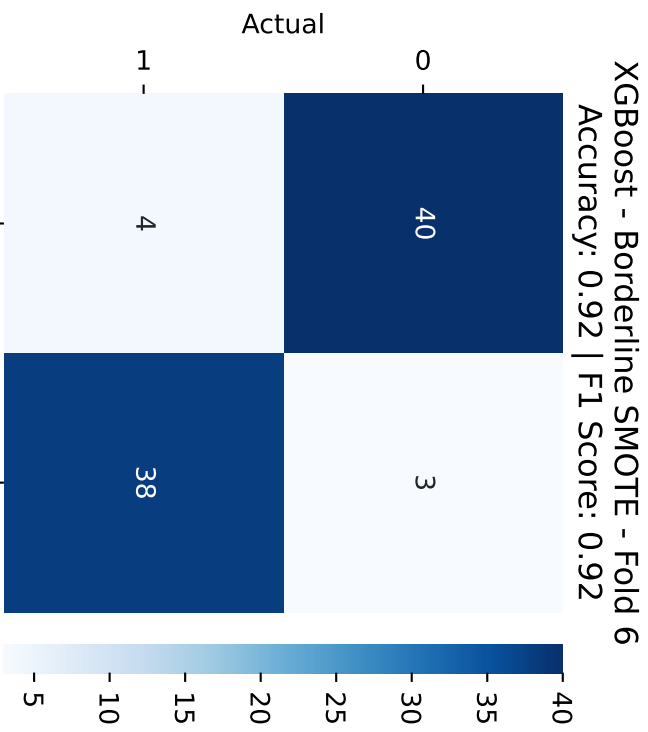
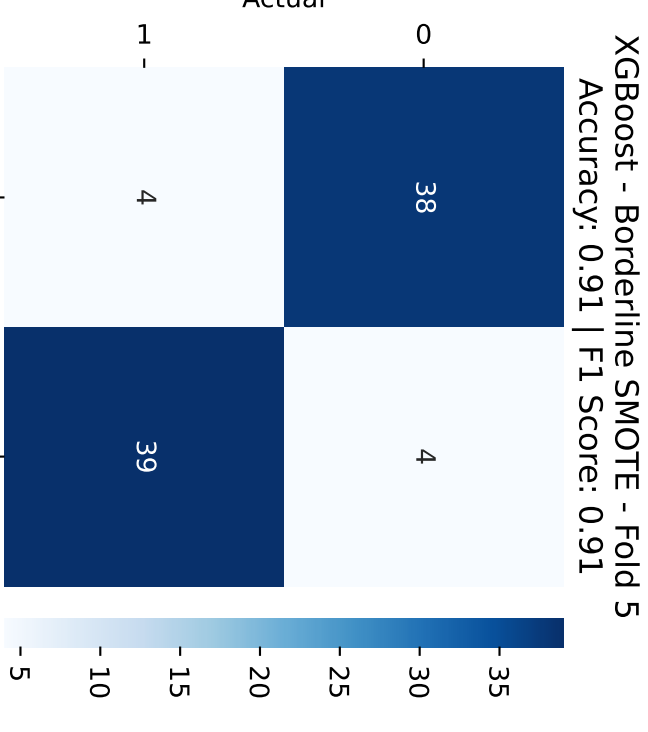
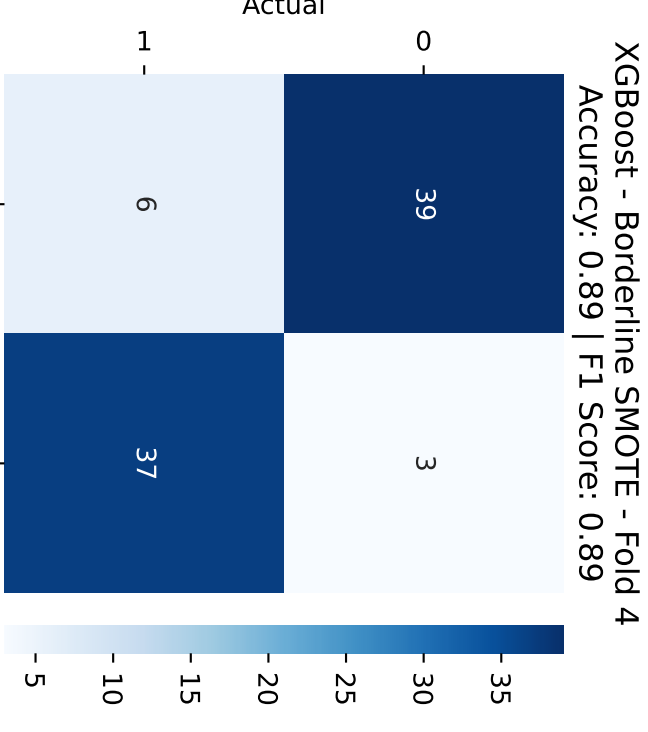
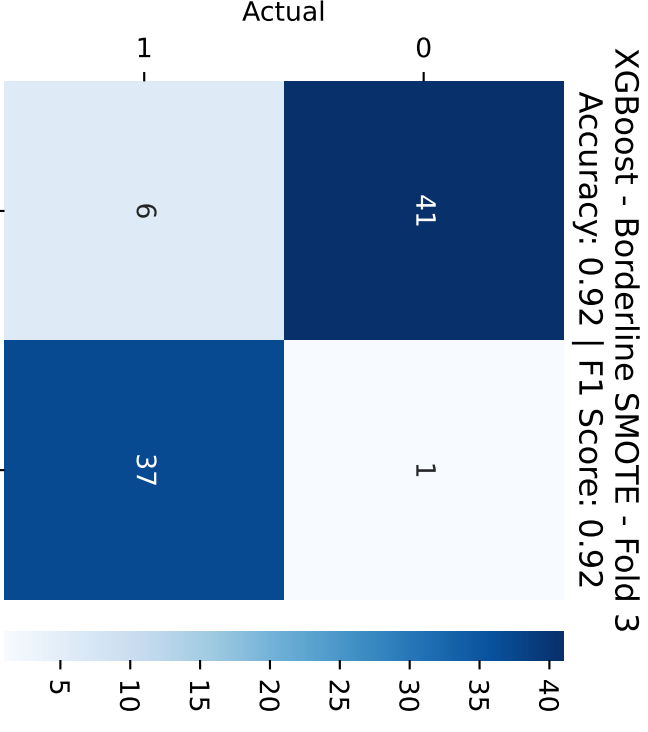
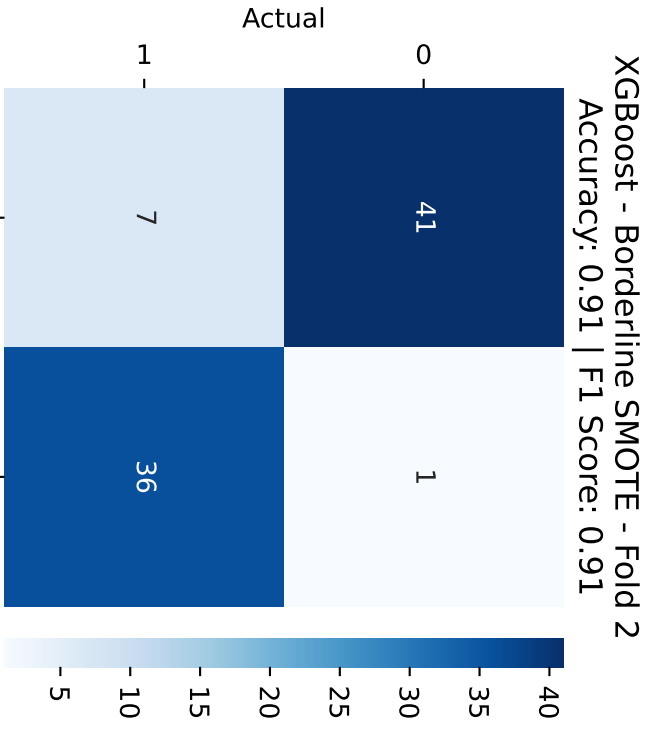
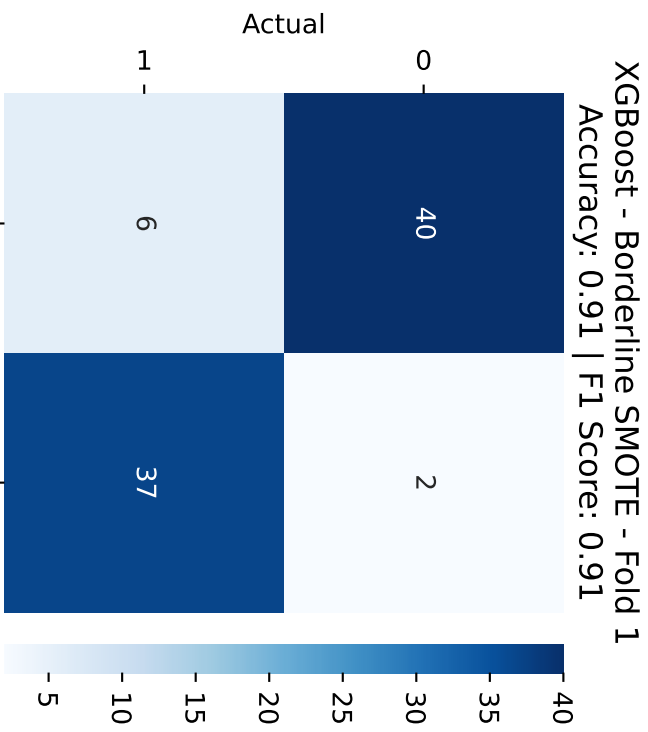


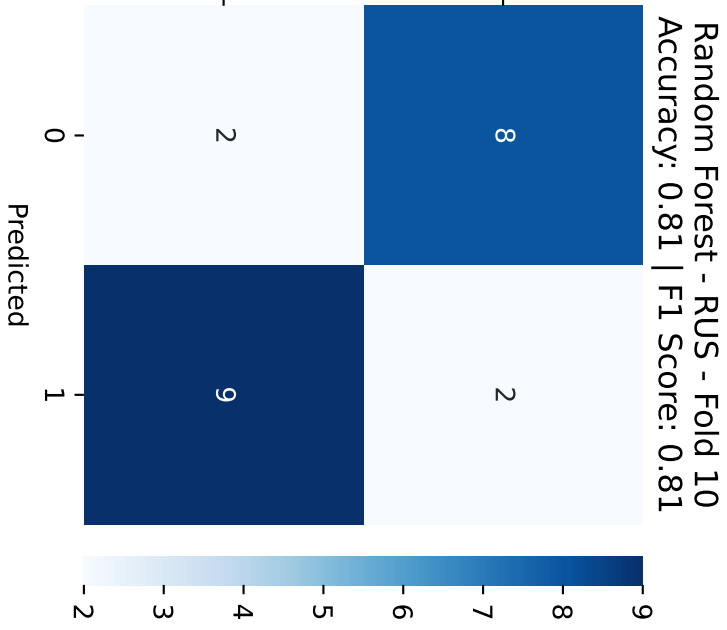
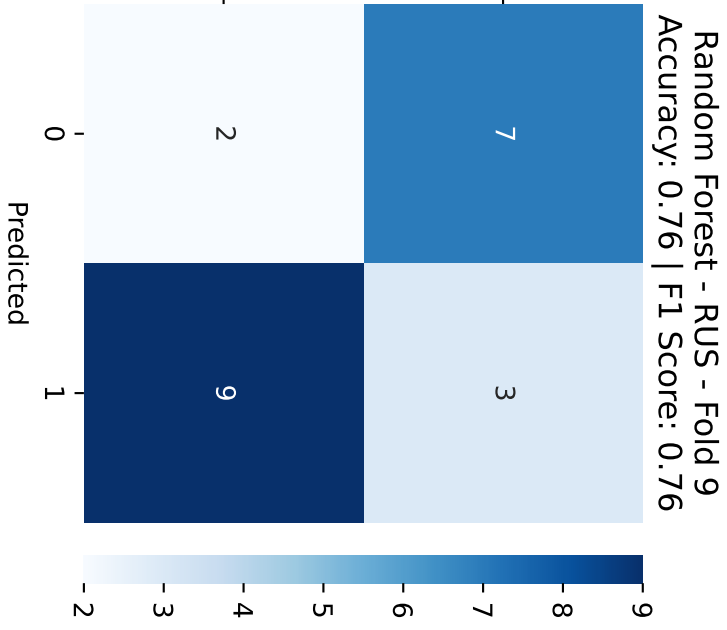
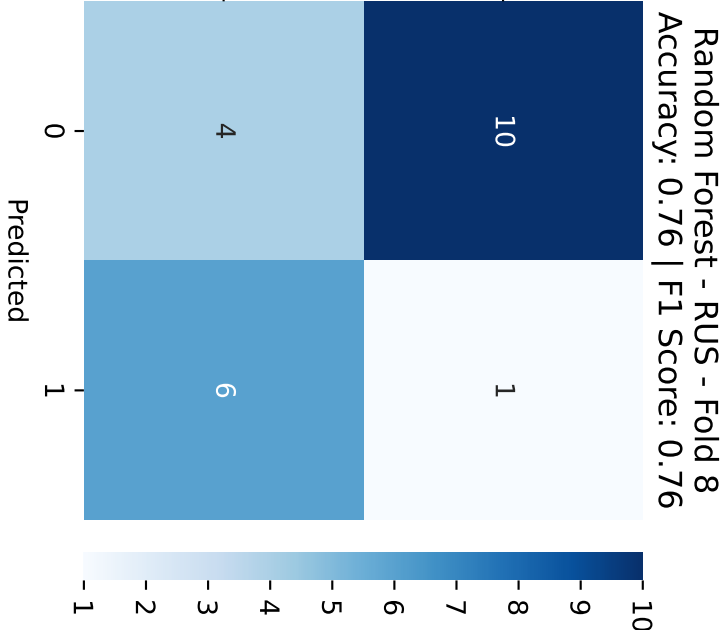
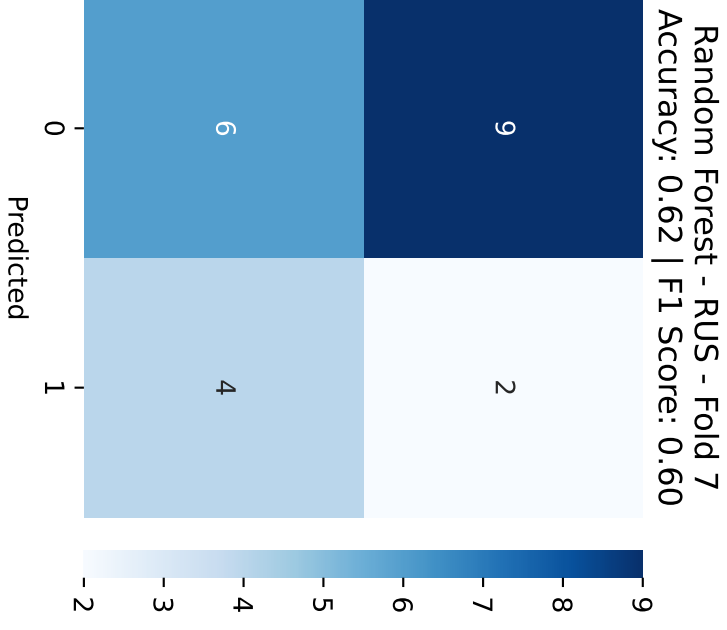
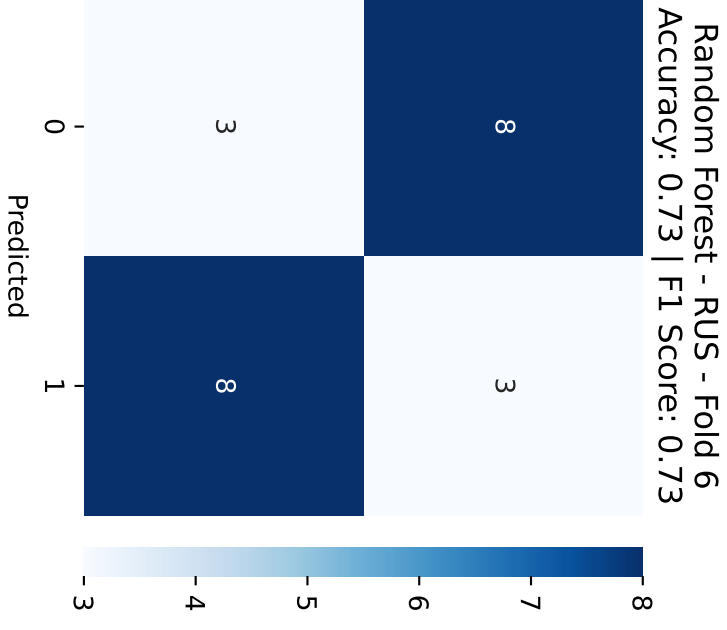
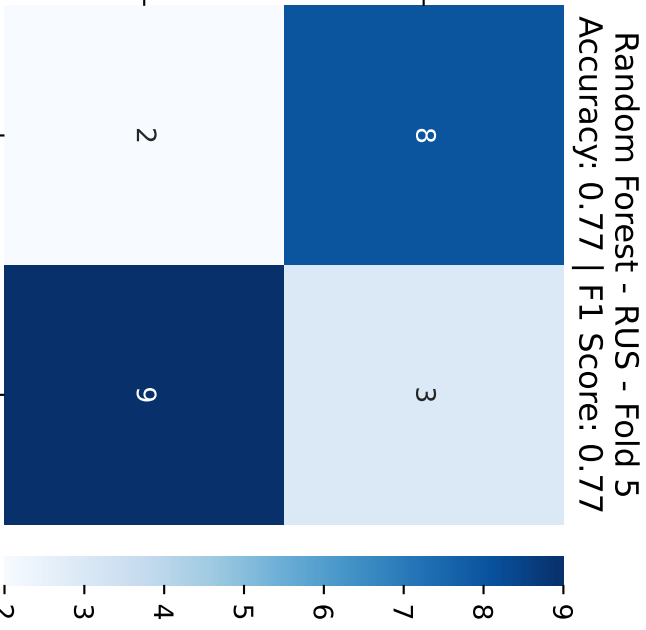
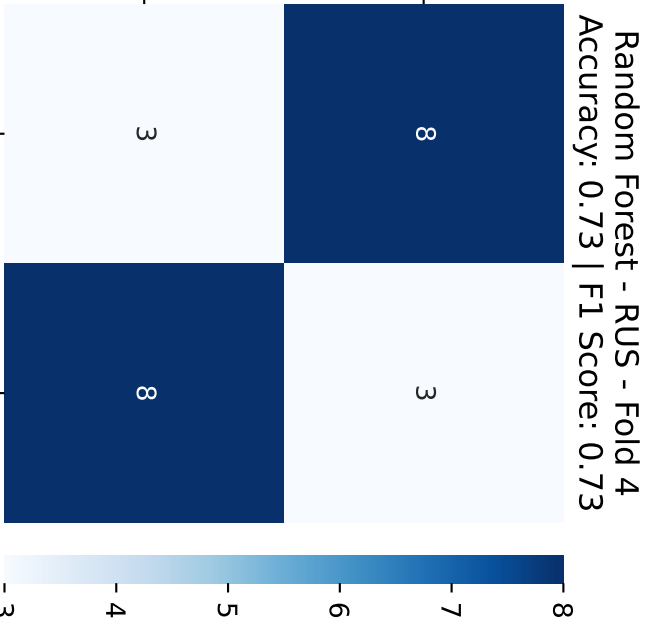
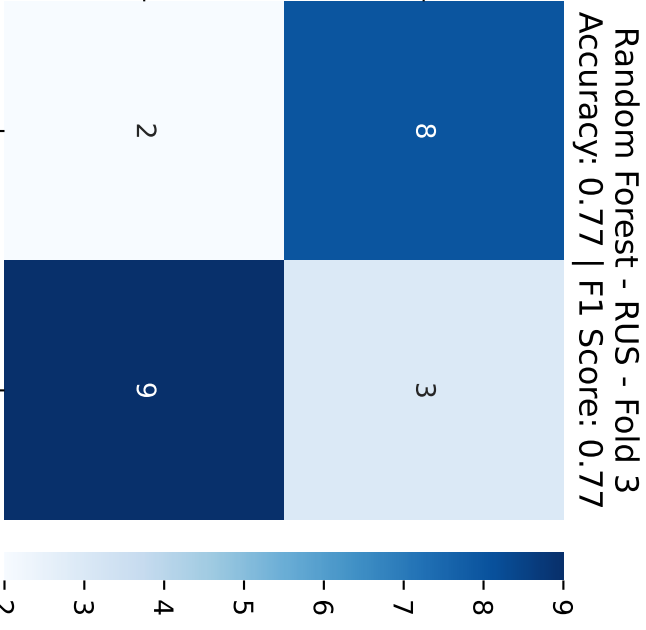
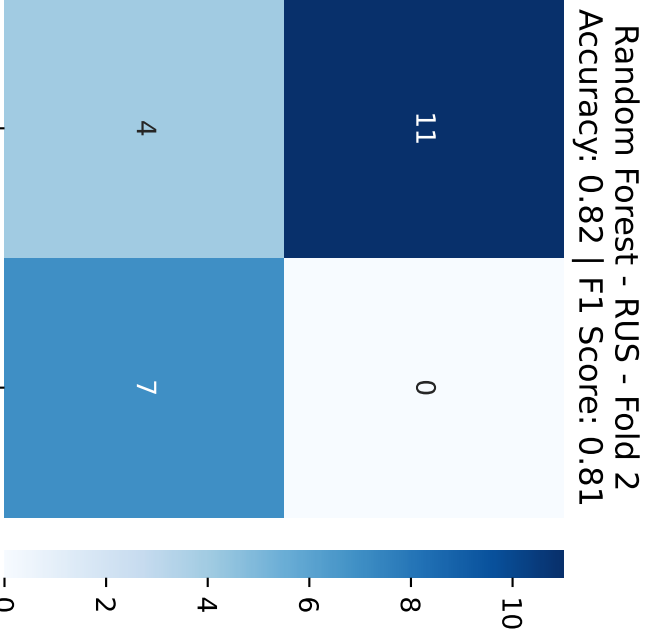
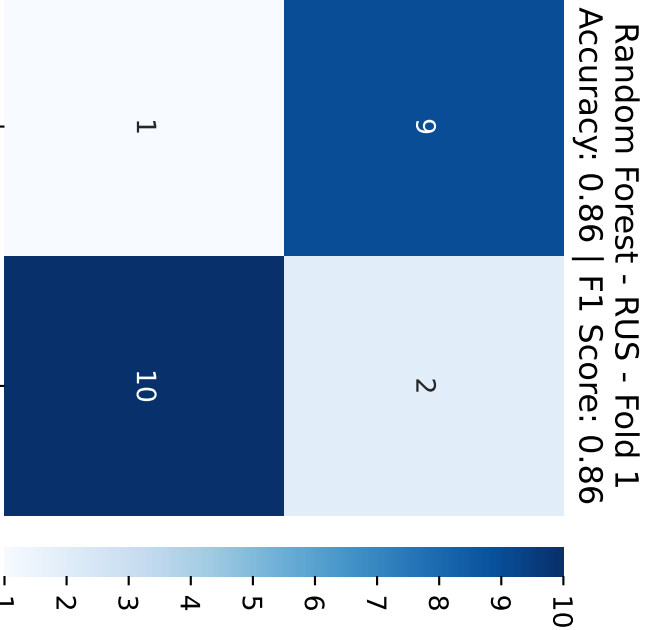
Decision Tree - Borderline SMOTE - Fold 10  
Accuracy: 0.87 | F1 Score: 0.87

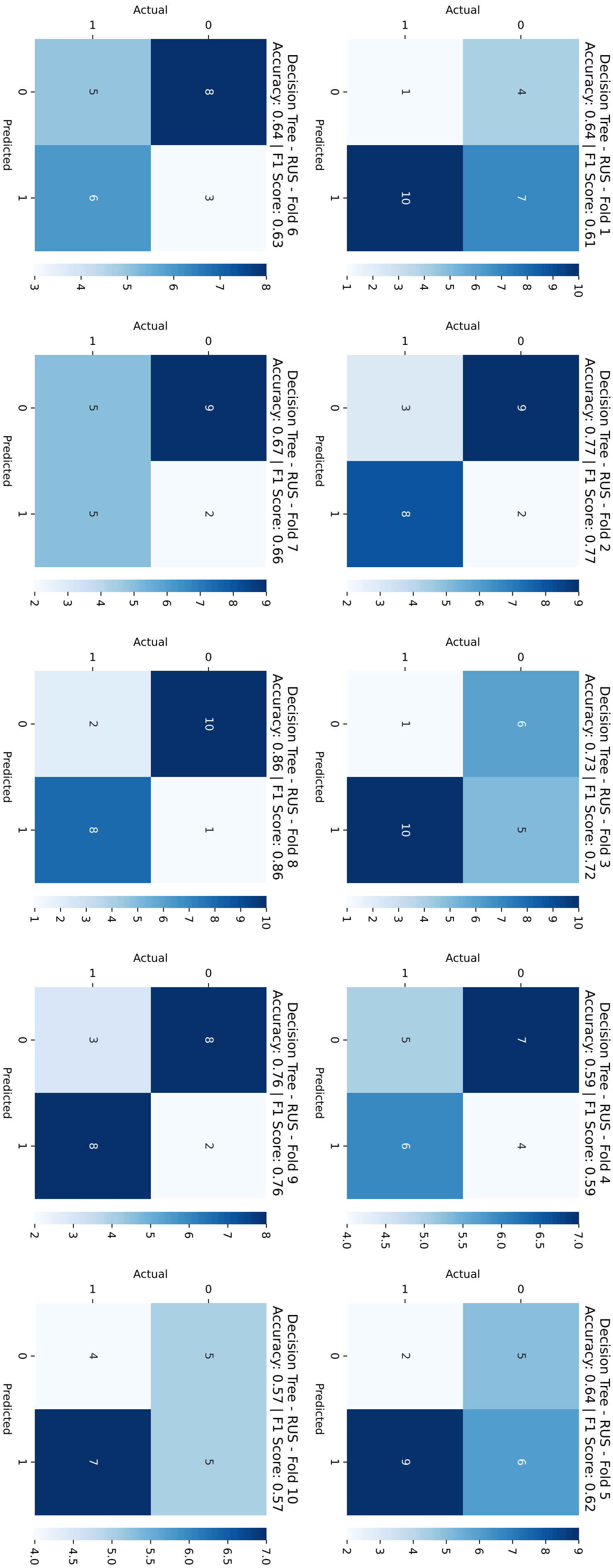


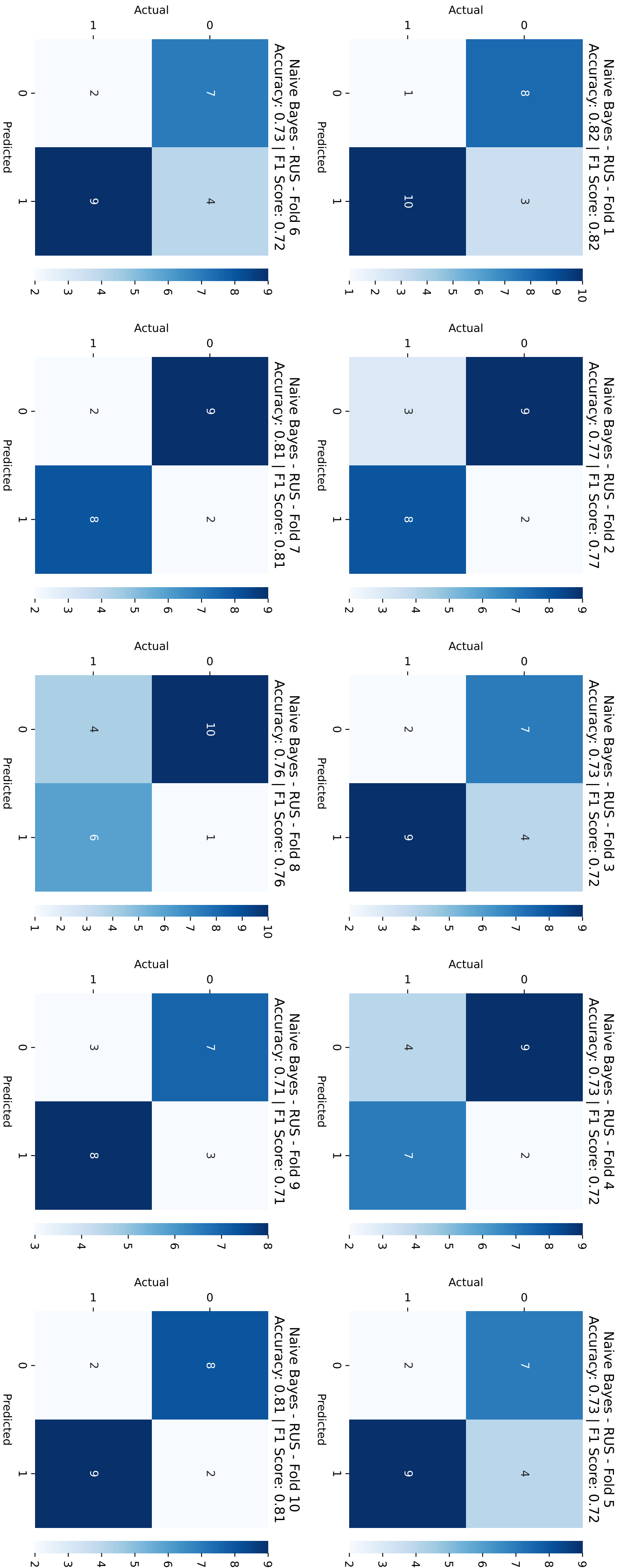


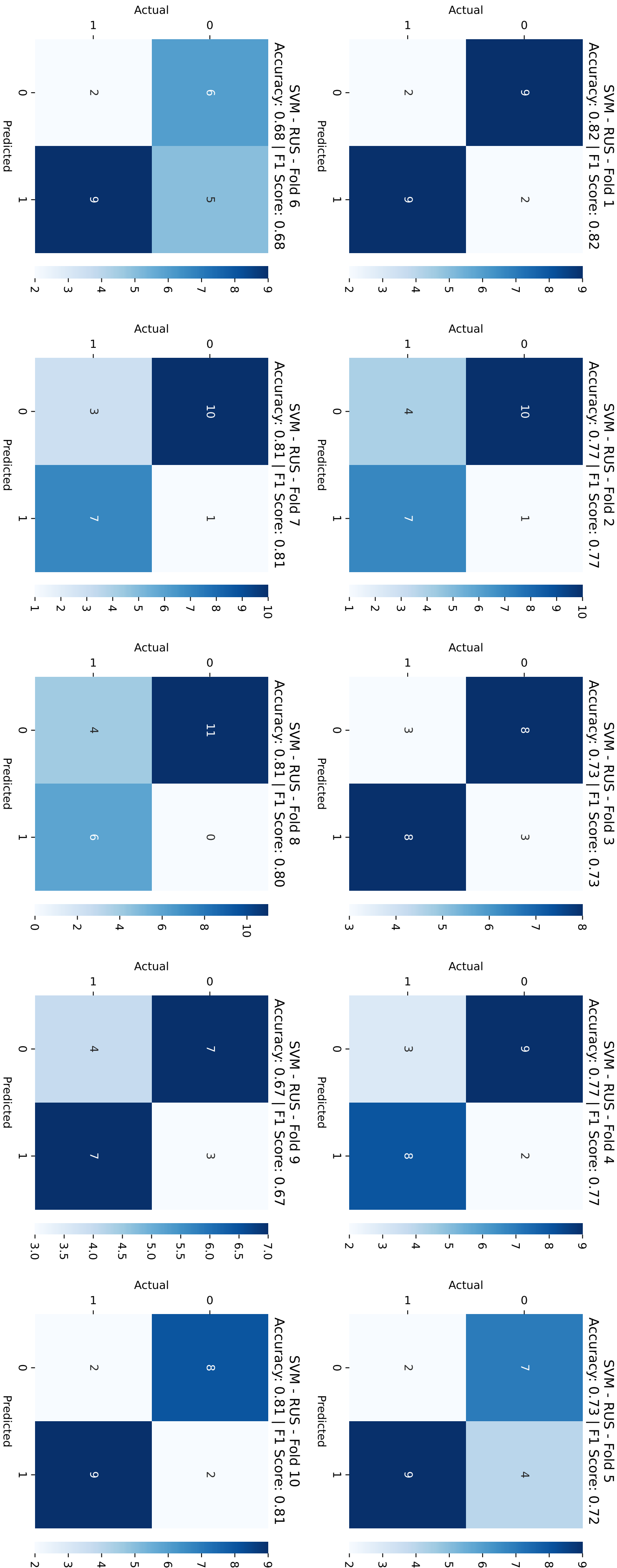


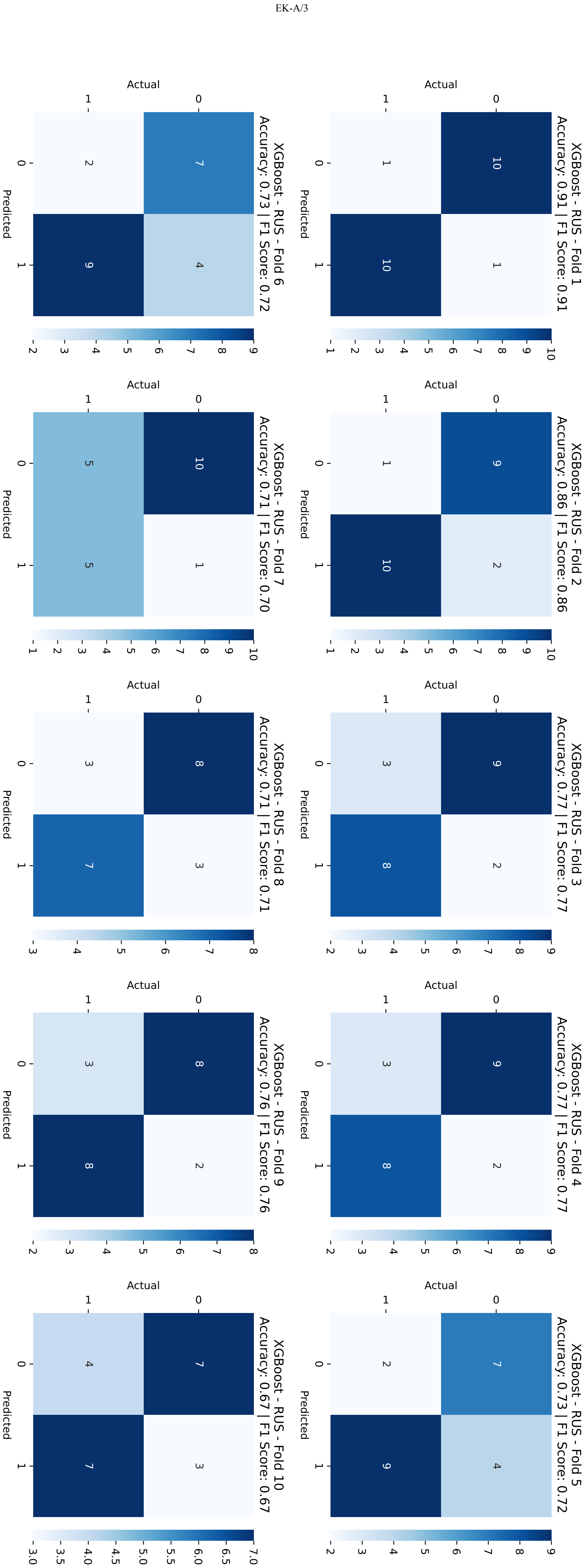


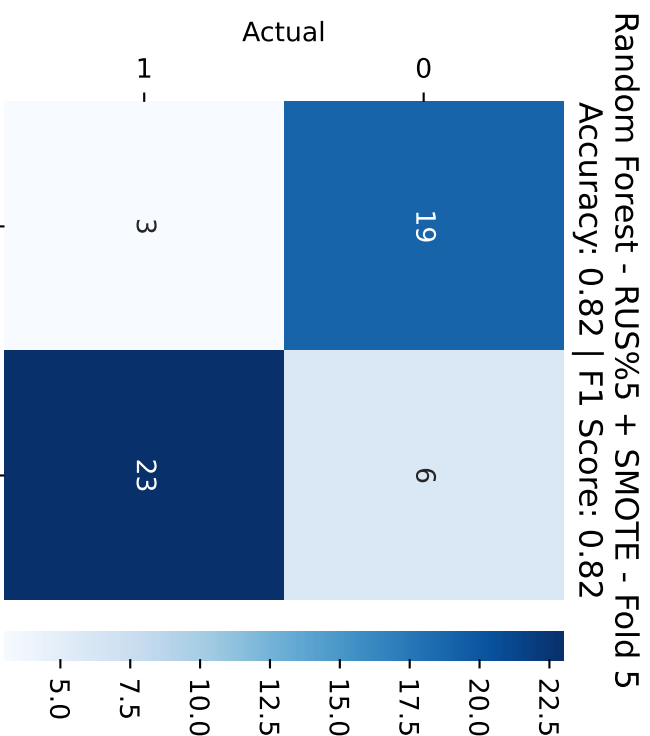
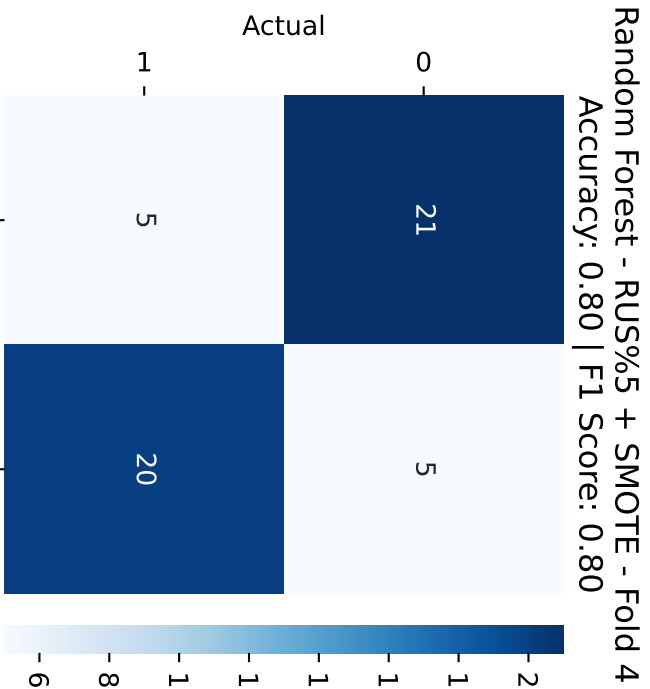
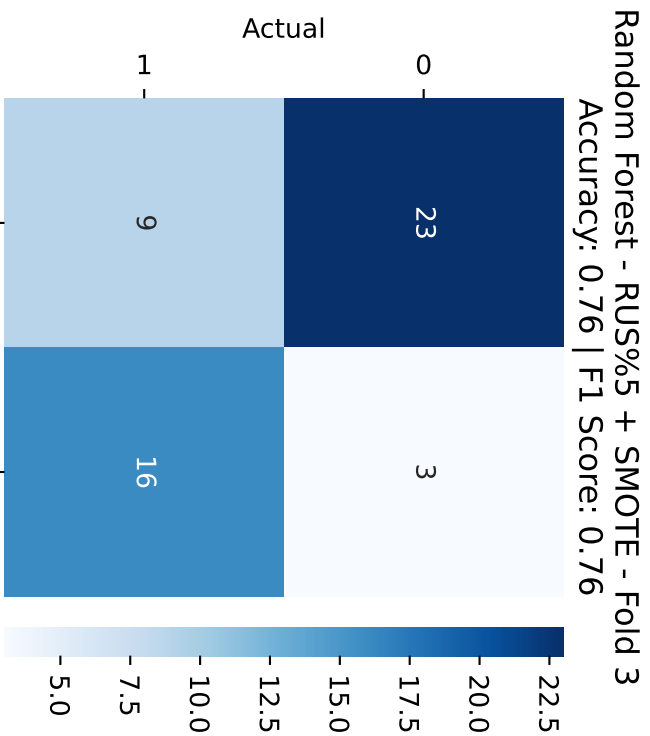
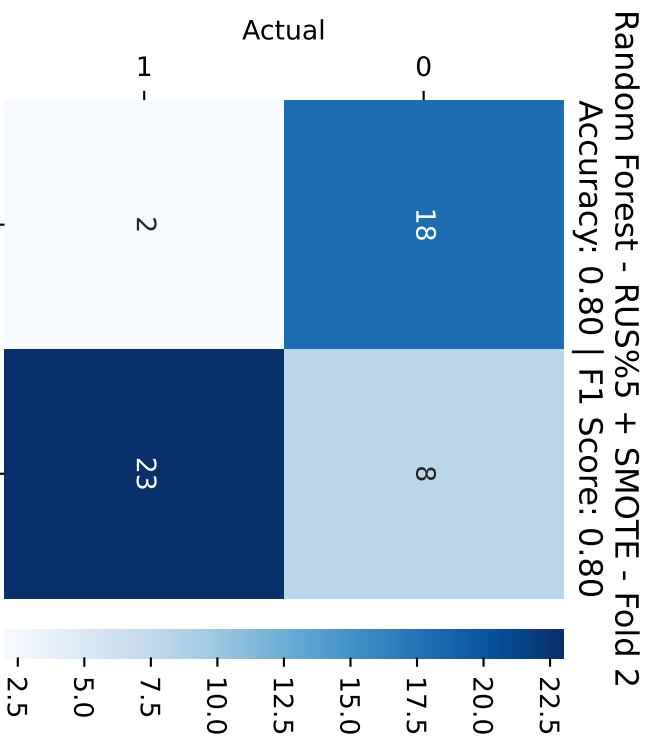
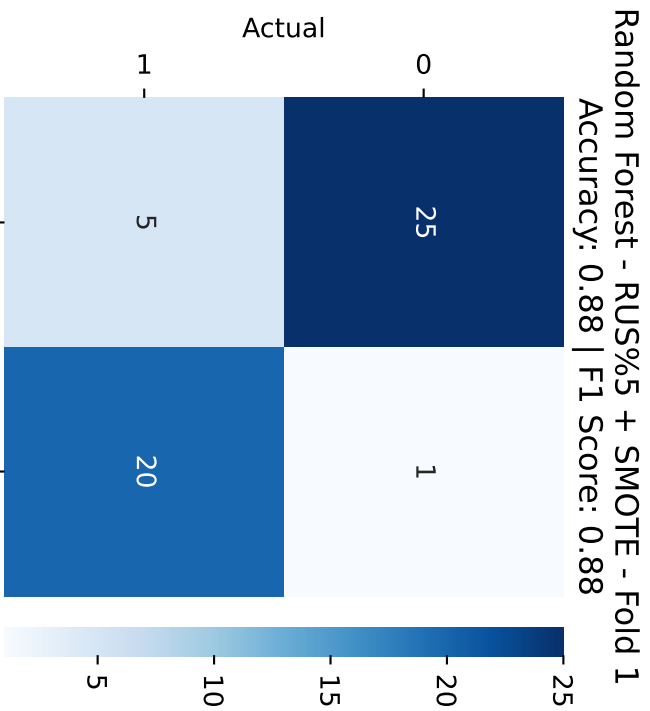




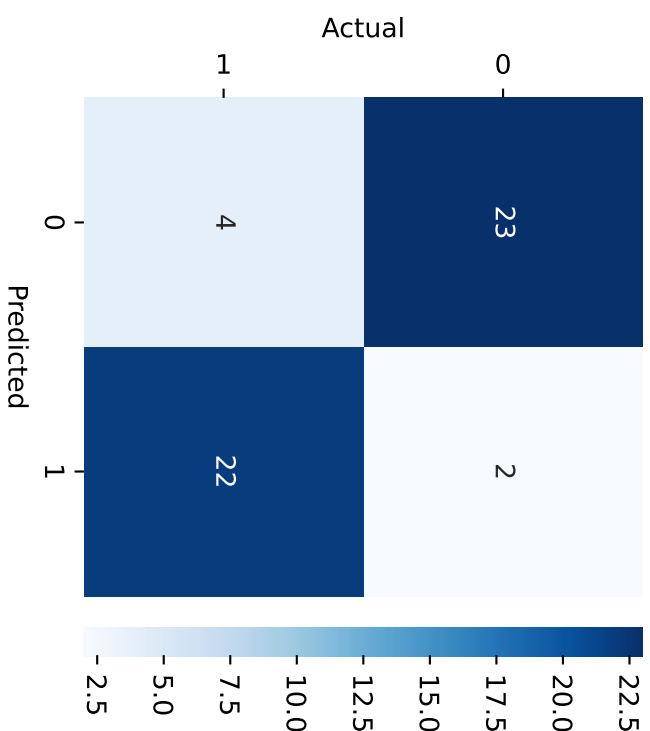




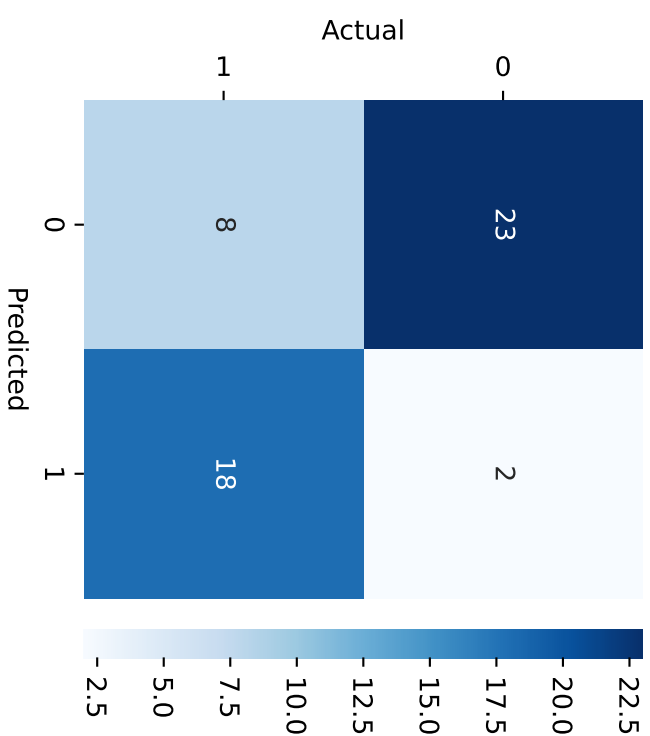




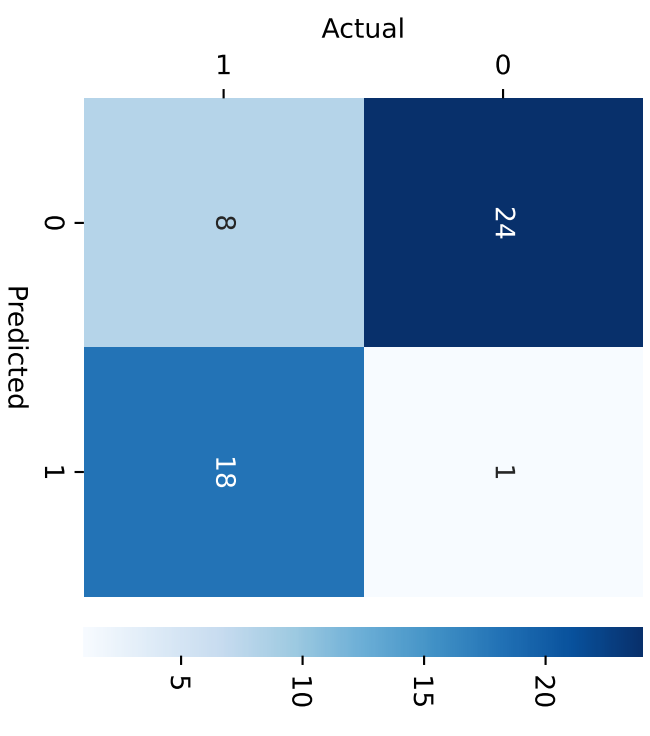
Random Forest - RUS%5 + SMOTE - Fold 6  
Accuracy: 0.88 | F1 Score: 0.88



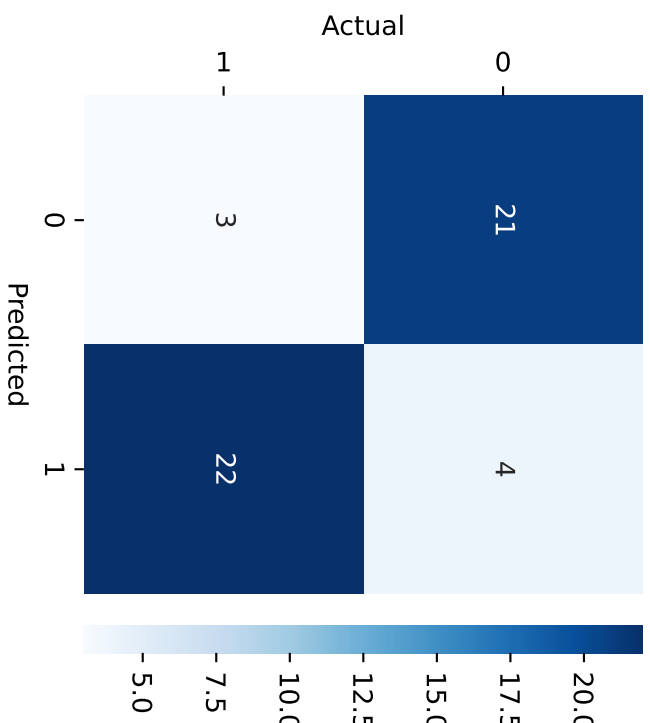
Random Forest - RUS%5 + SMOTE - Fold 7  
Accuracy: 0.80 | F1 Score: 0.80



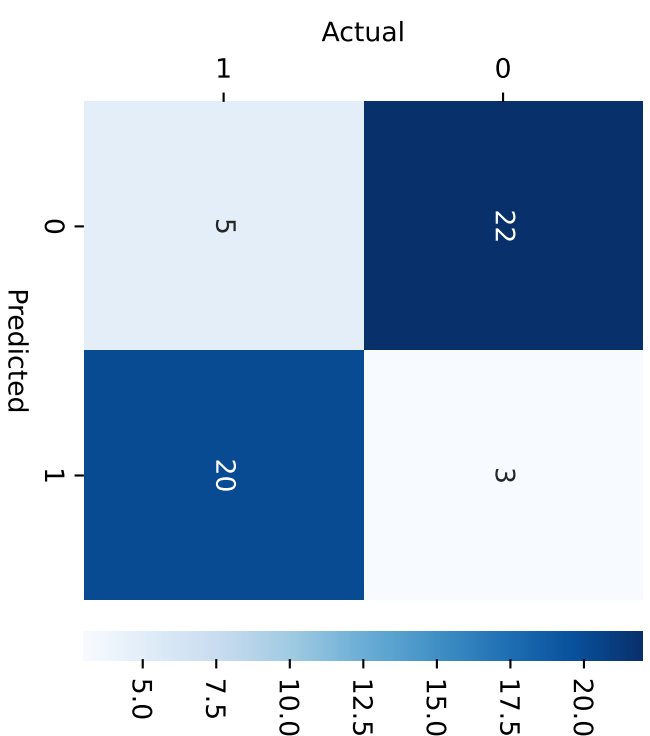
Random Forest - RUS%5 + SMOTE - Fold 8  
Accuracy: 0.82 | F1 Score: 0.82

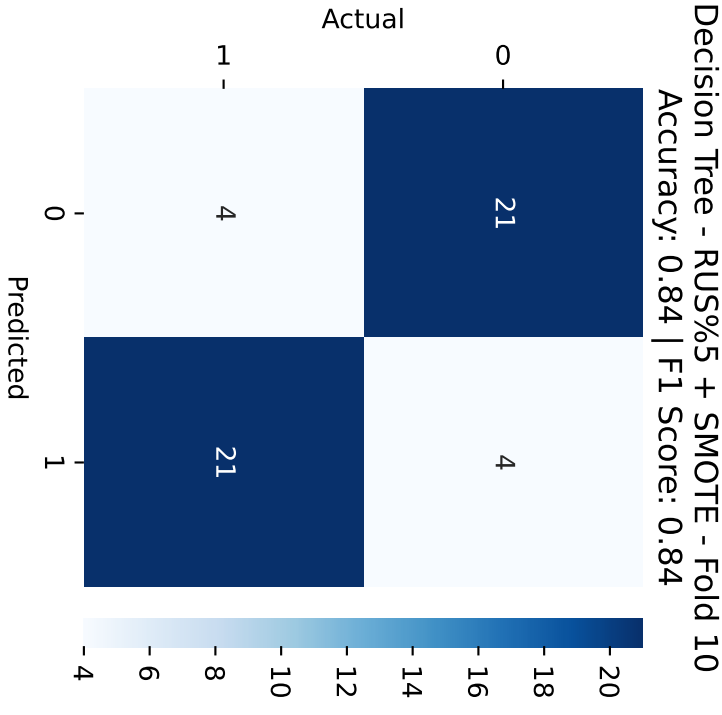
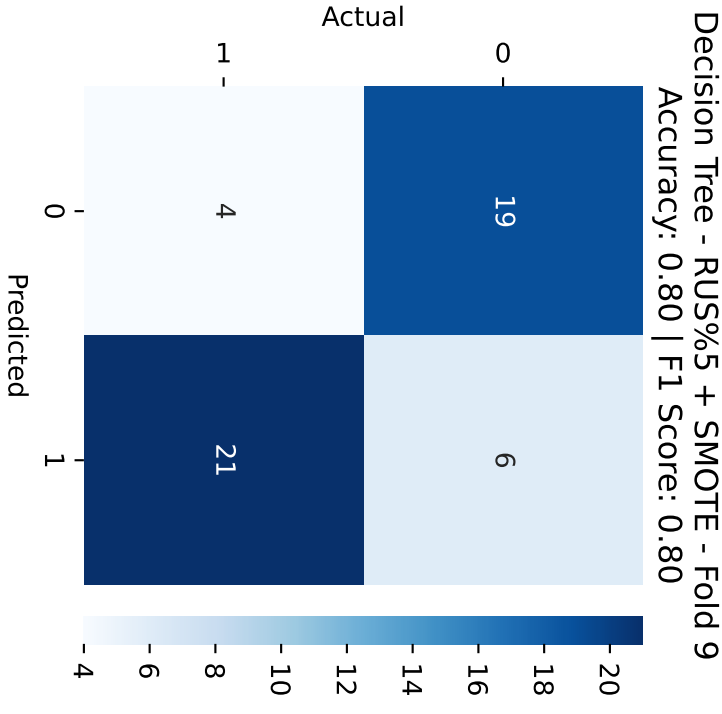
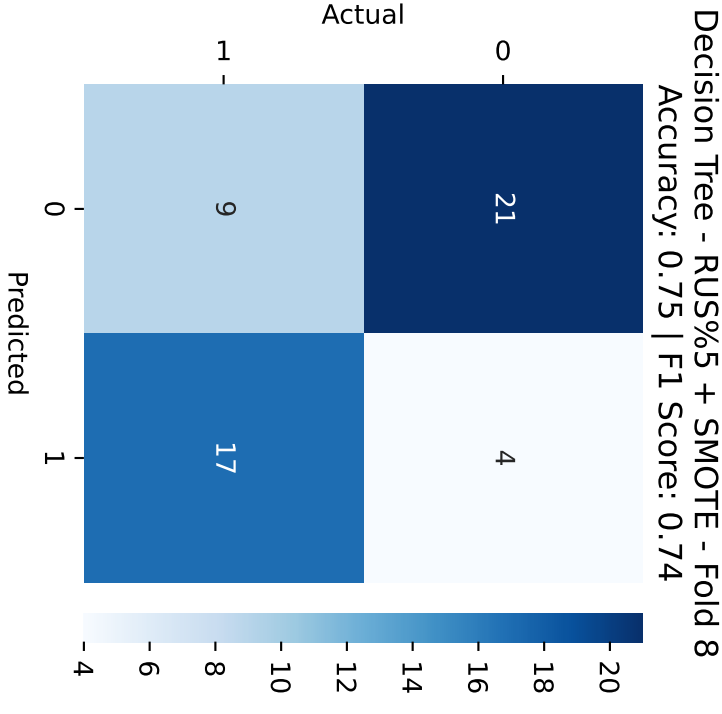
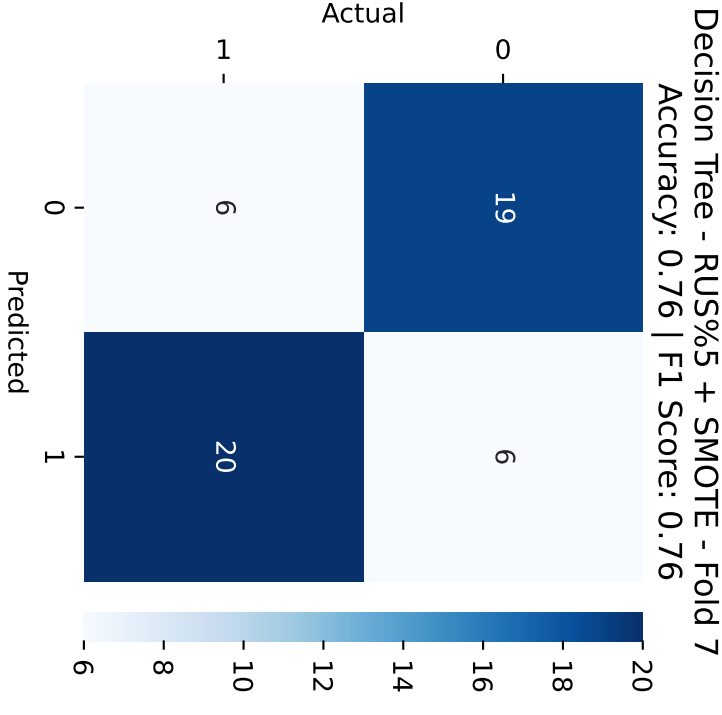
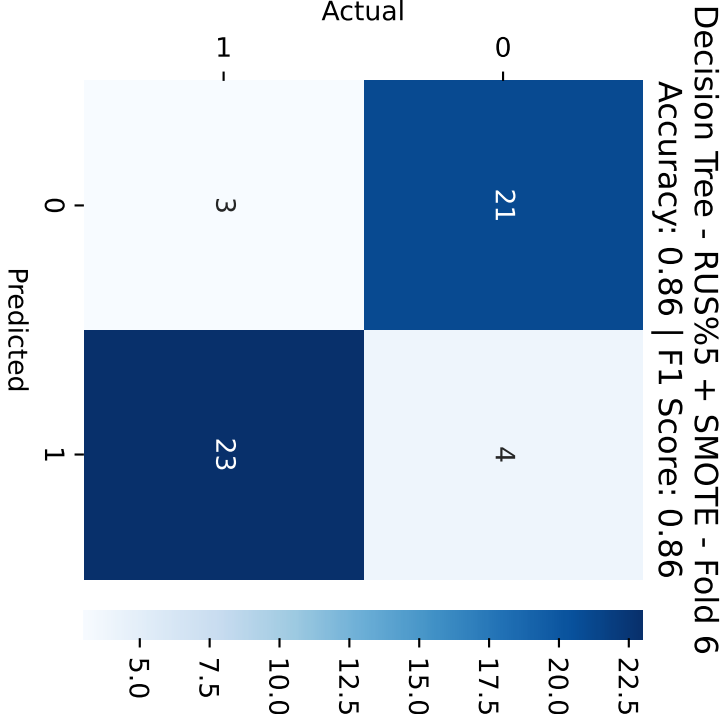
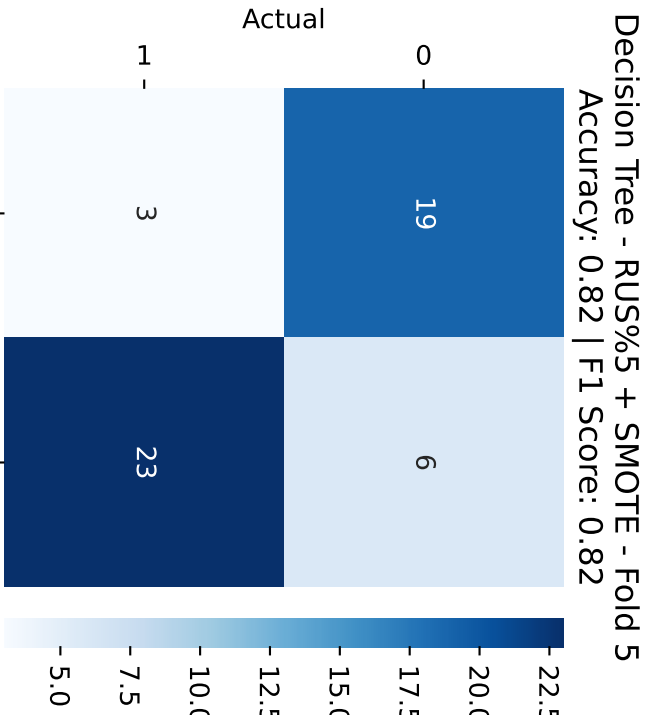
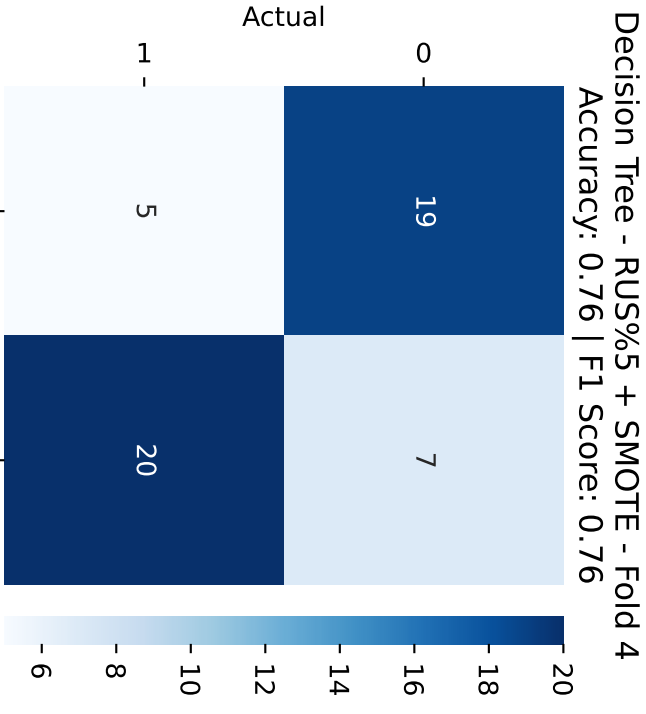
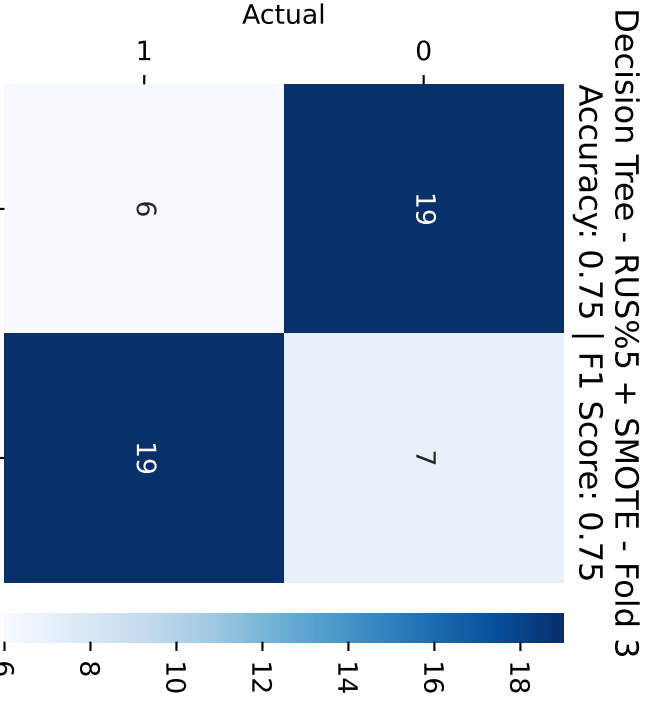
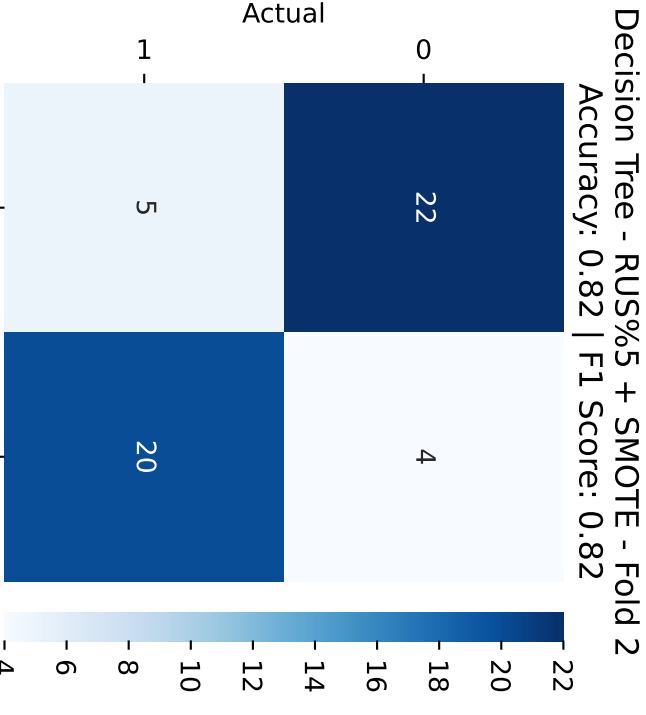
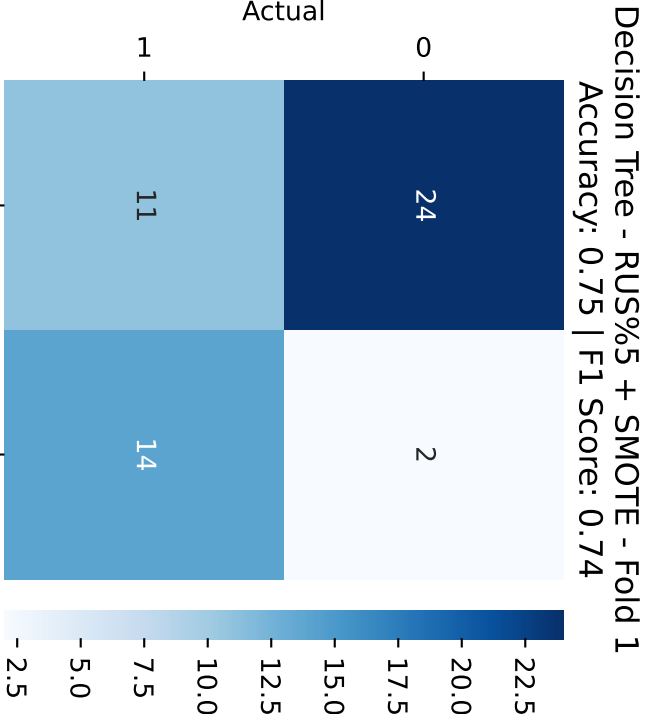


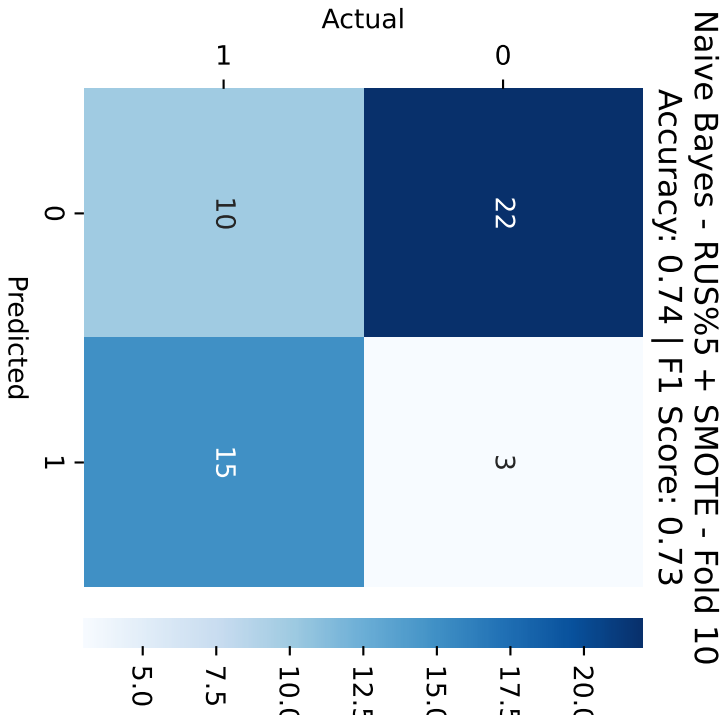
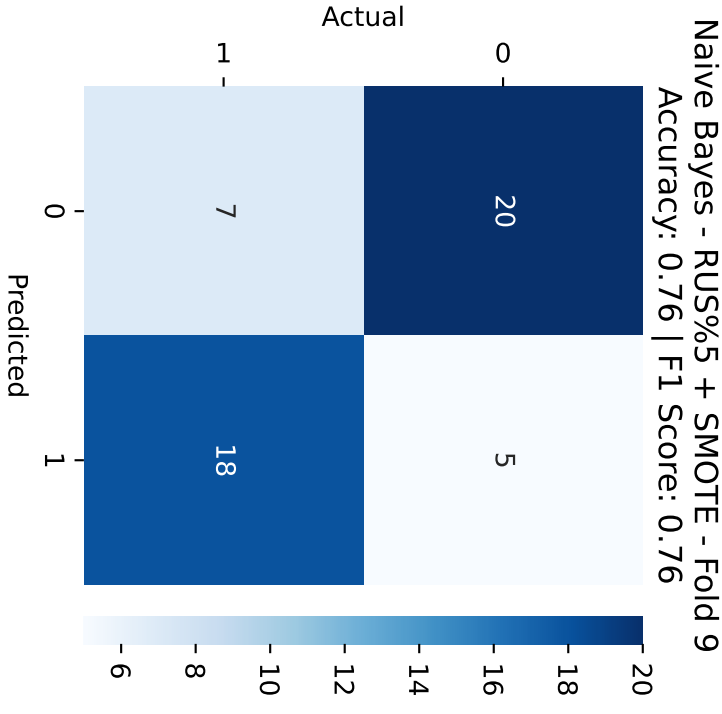
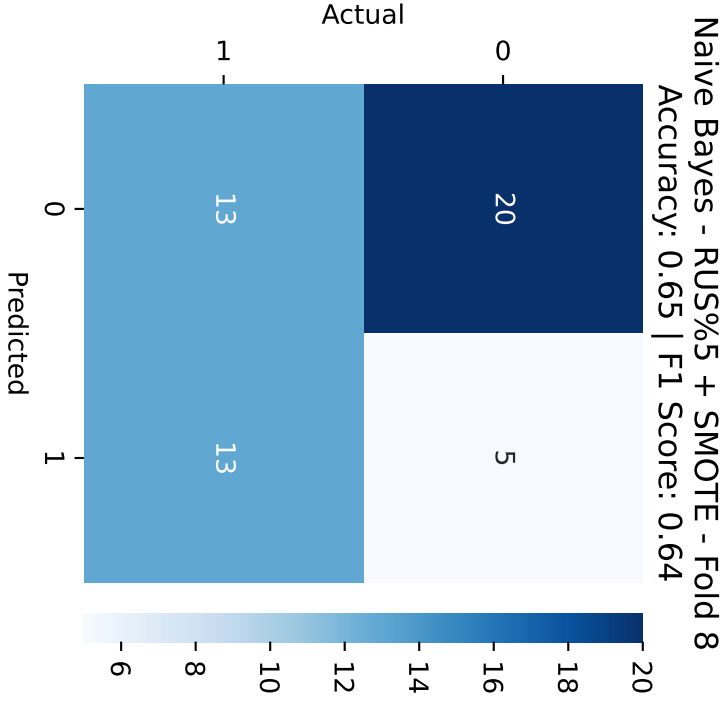
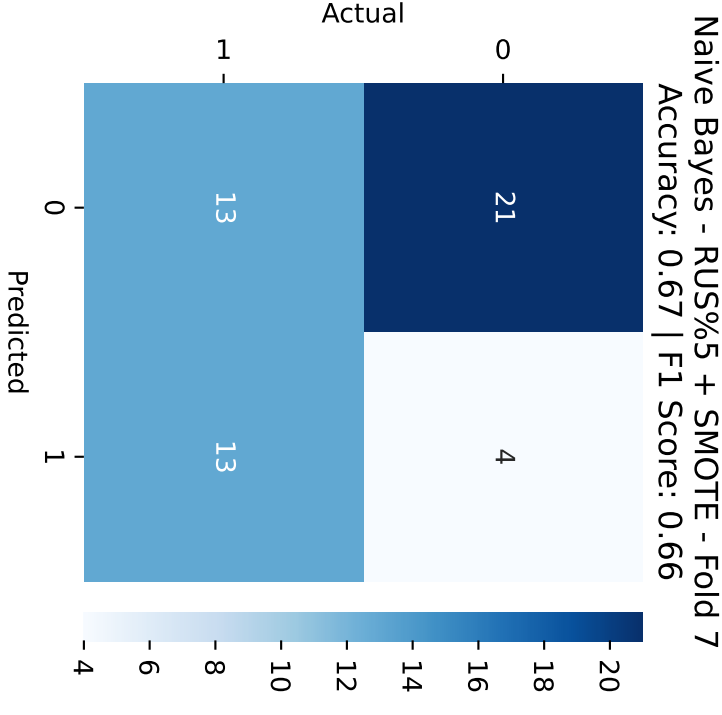
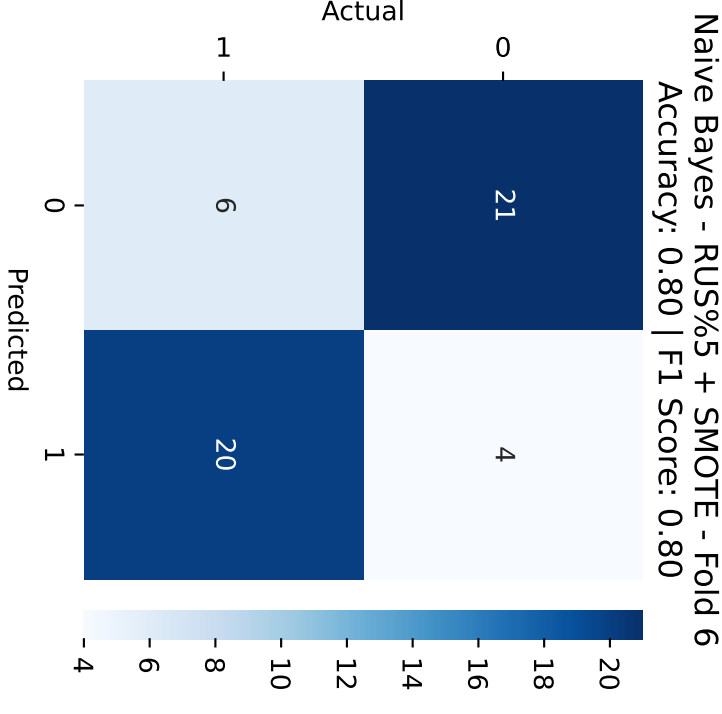
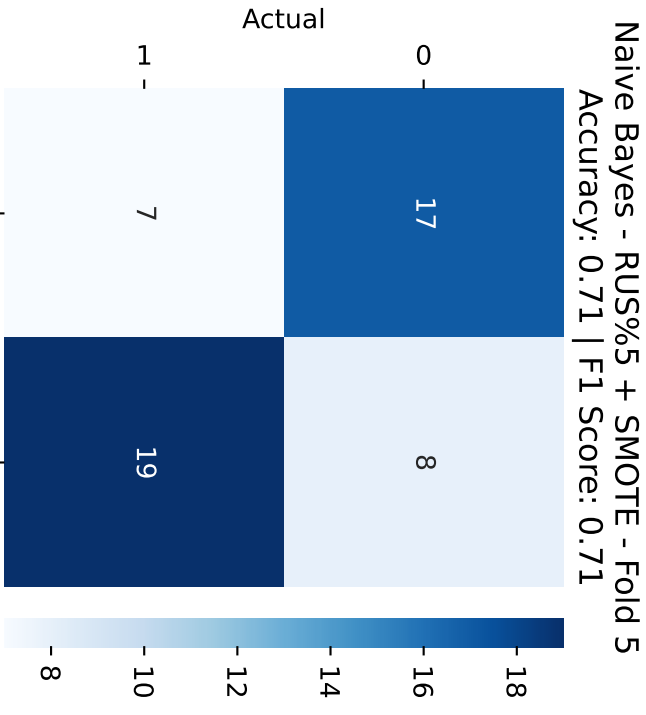
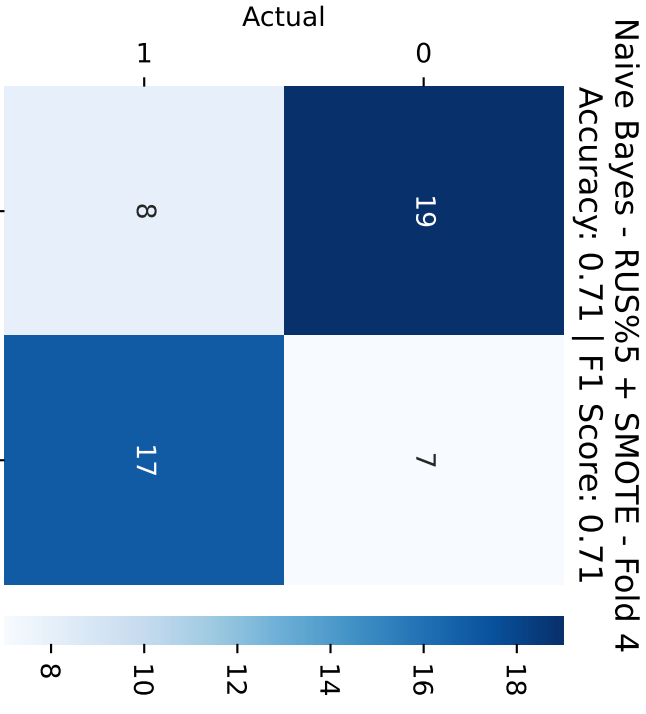
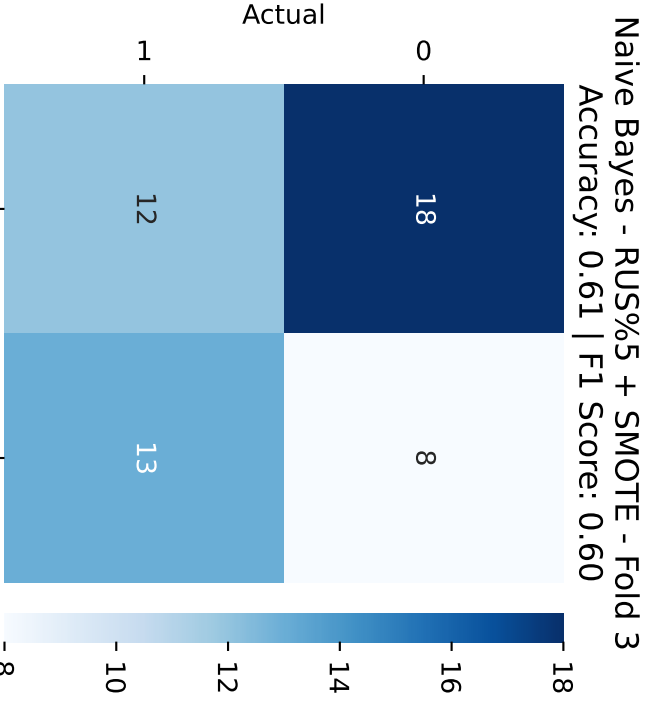
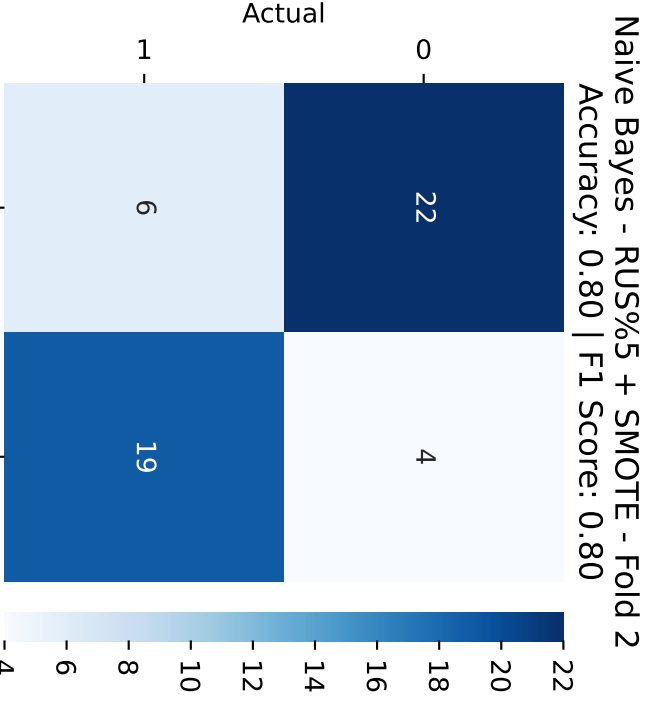
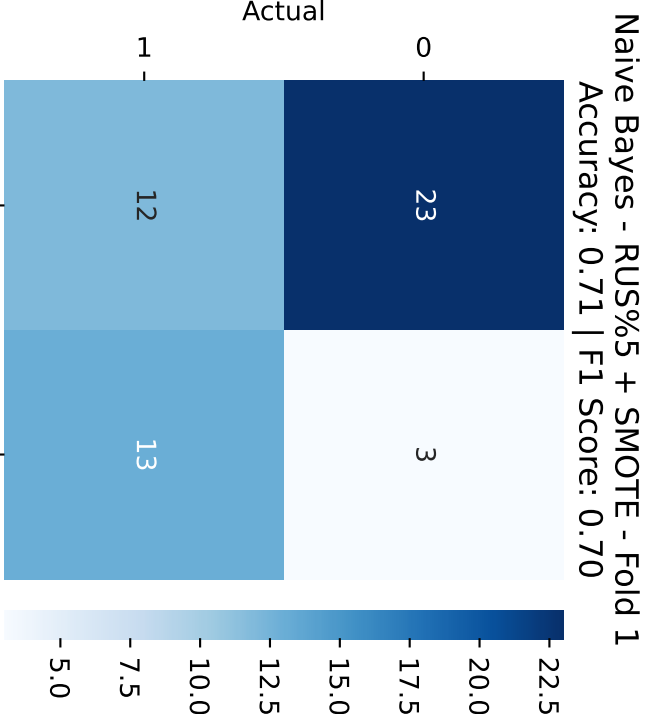
Random Forest - RUS%5 + SMOTE - Fold 9  
Accuracy: 0.86 | F1 Score: 0.86

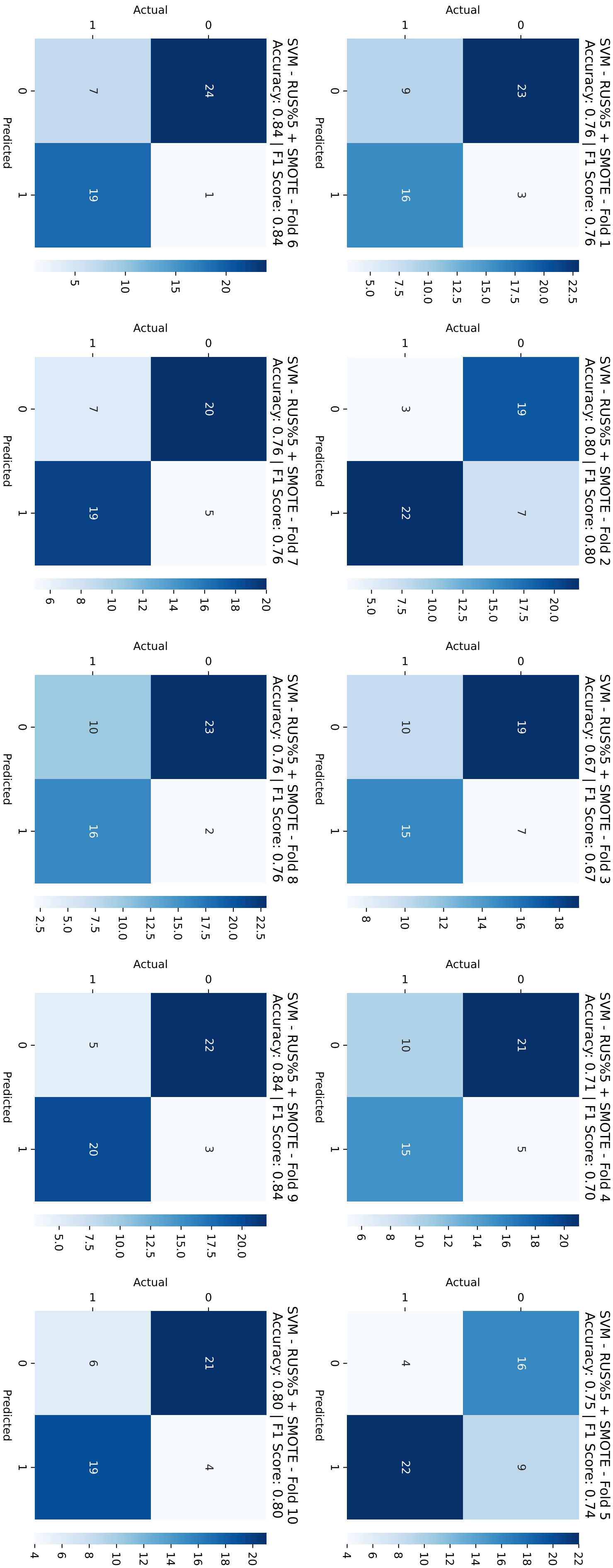


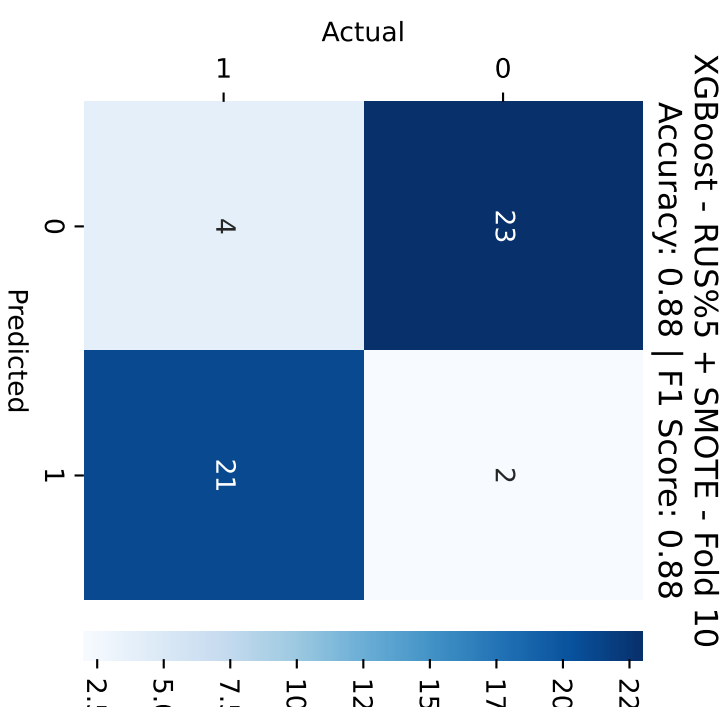
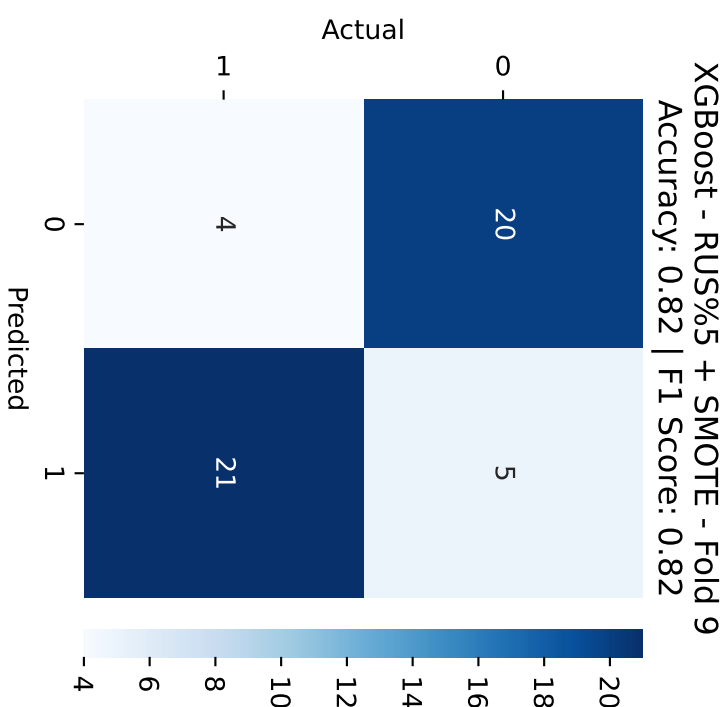
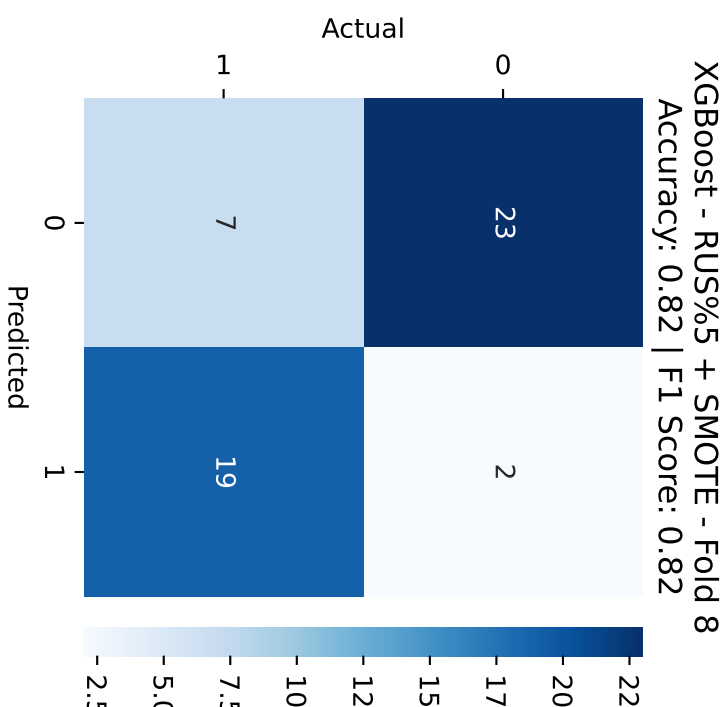
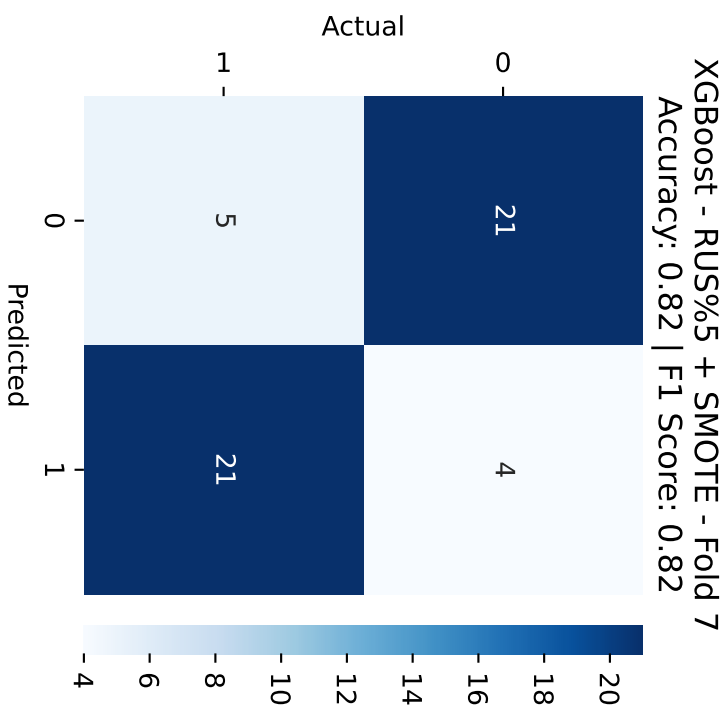
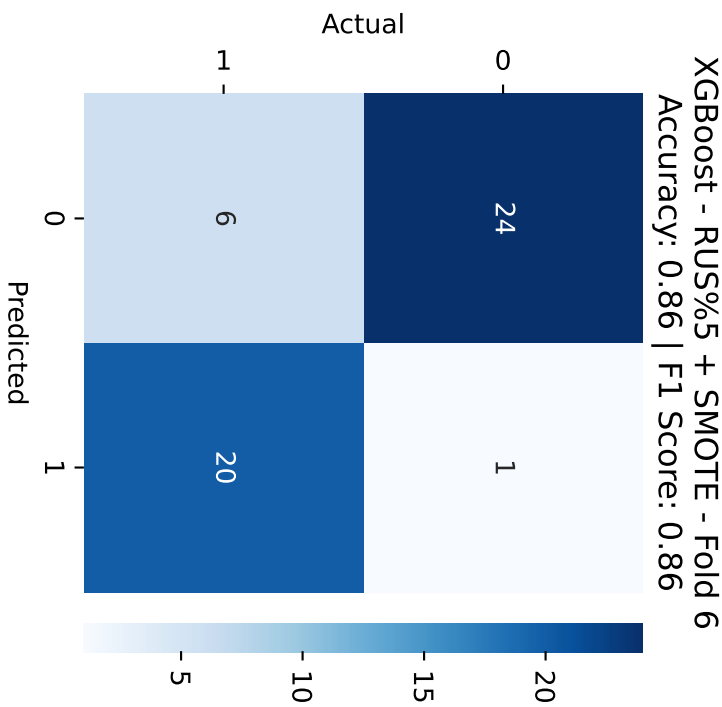
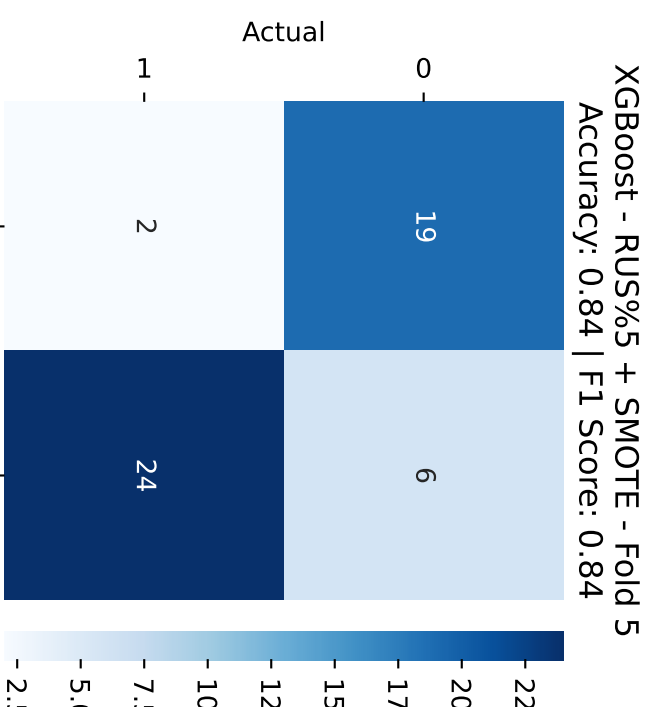
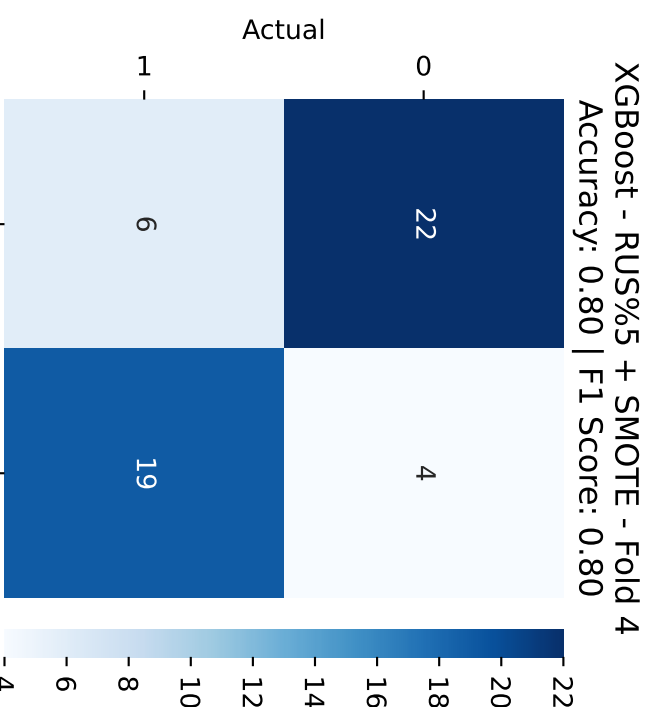
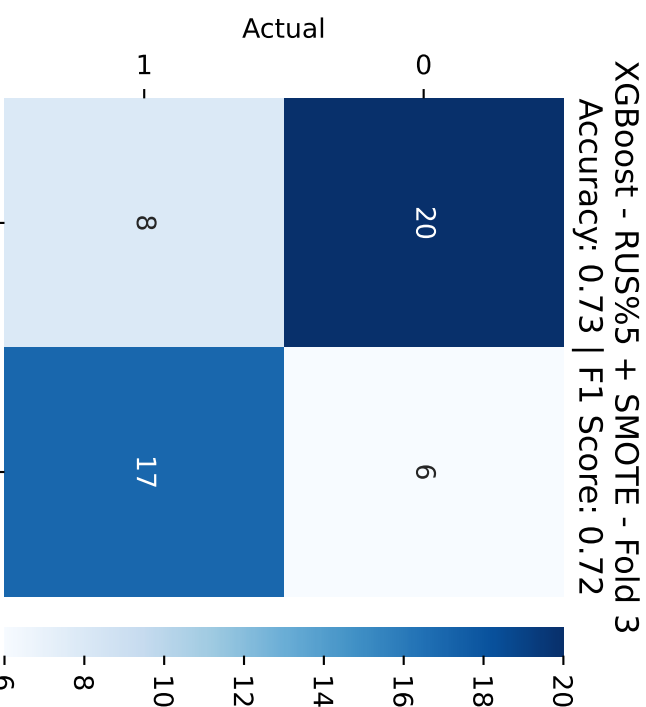
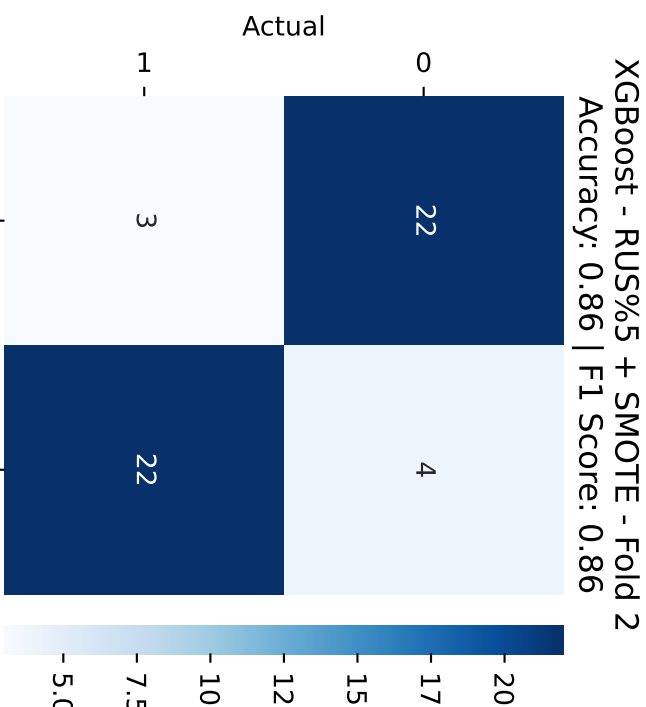
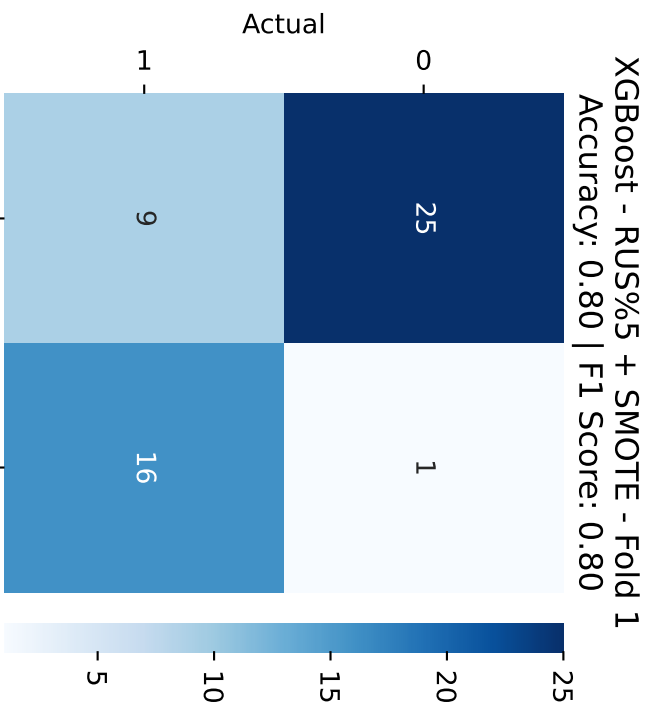
Random Forest - RUS%5 + SMOTE - Fold 10  
Accuracy: 0.84 | F1 Score: 0.84

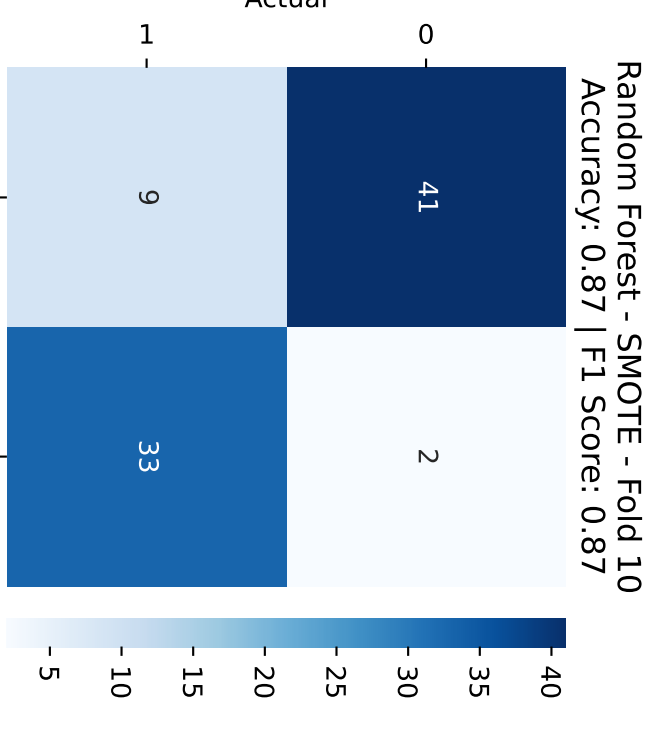
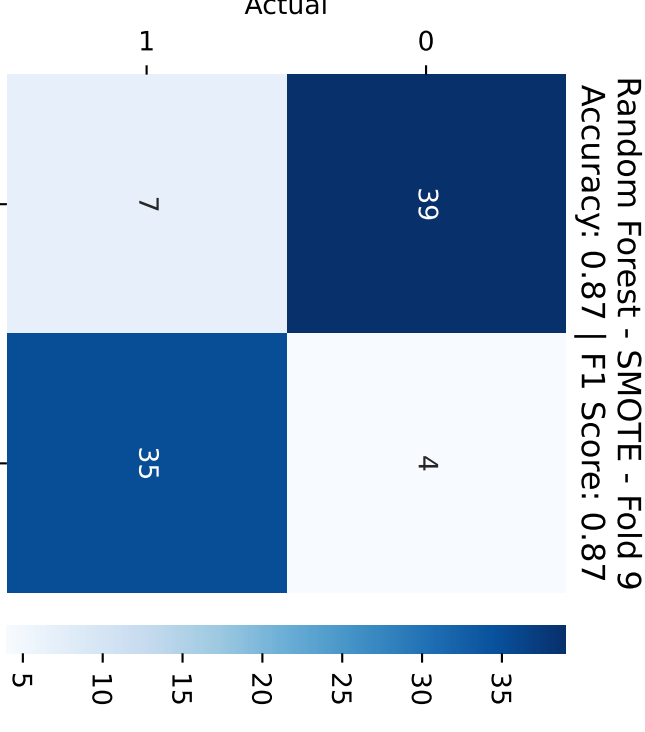
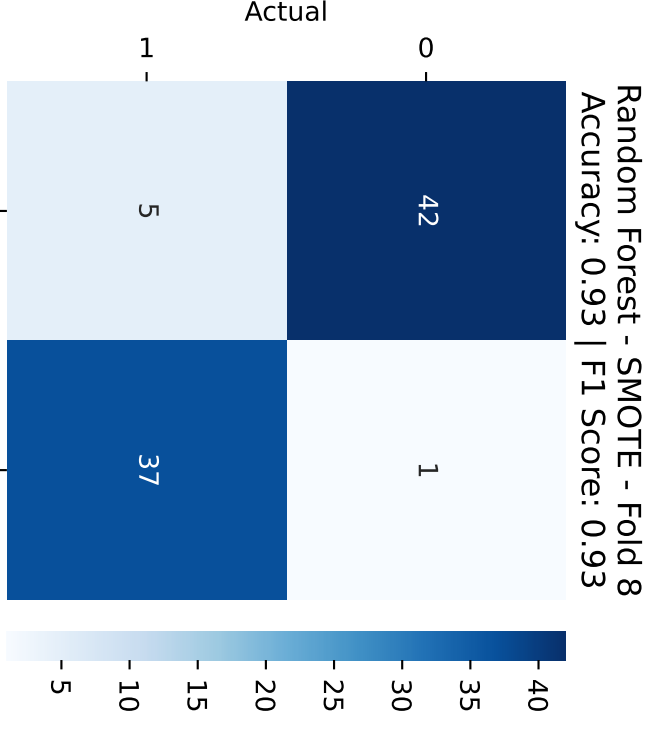
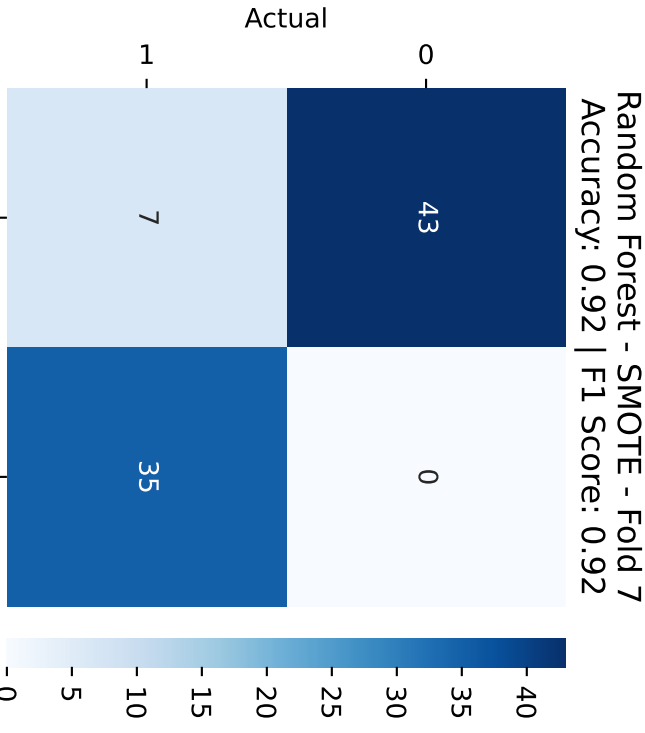
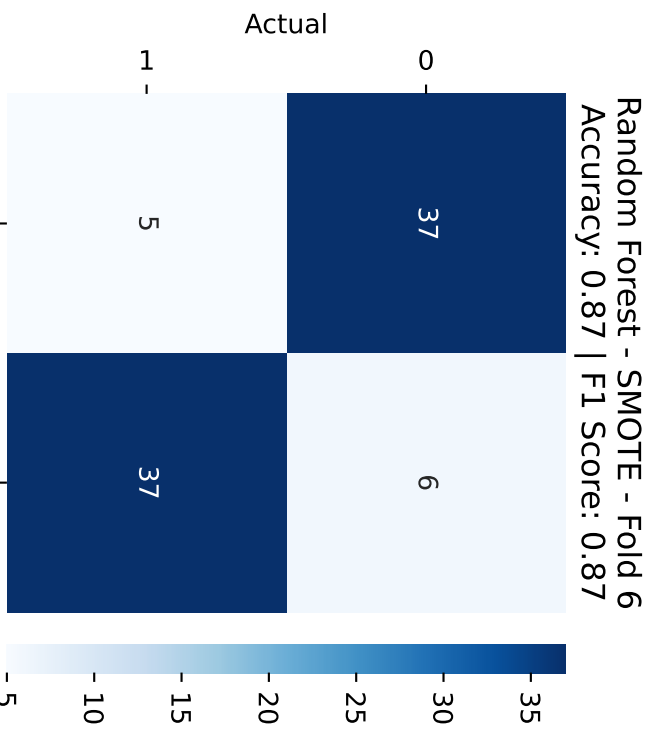
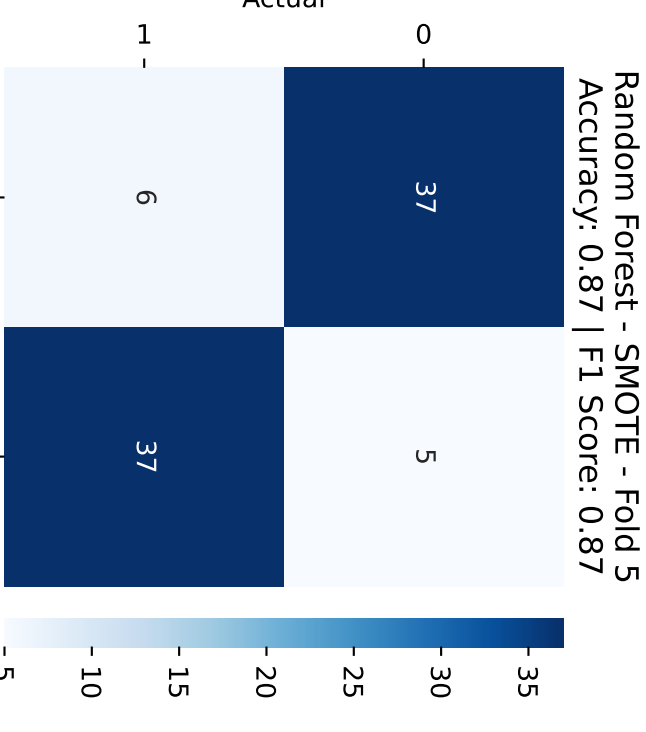
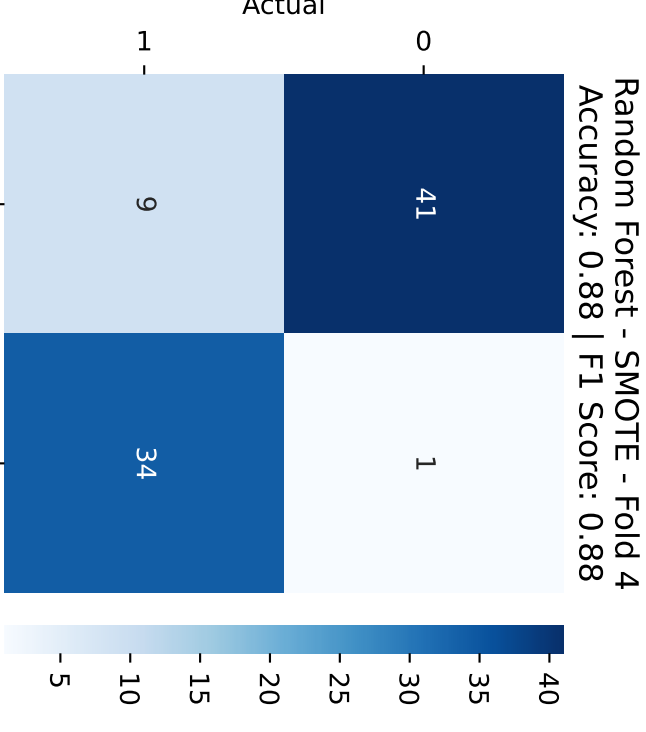
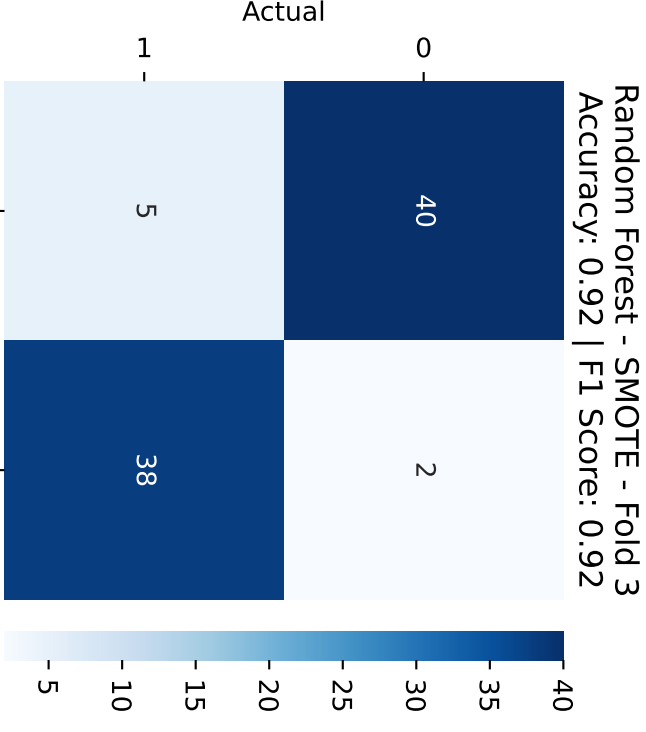
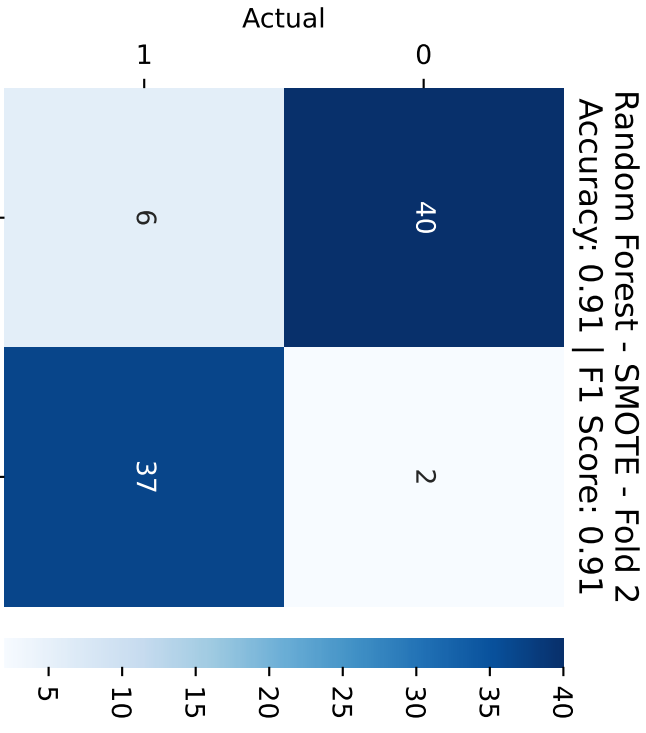
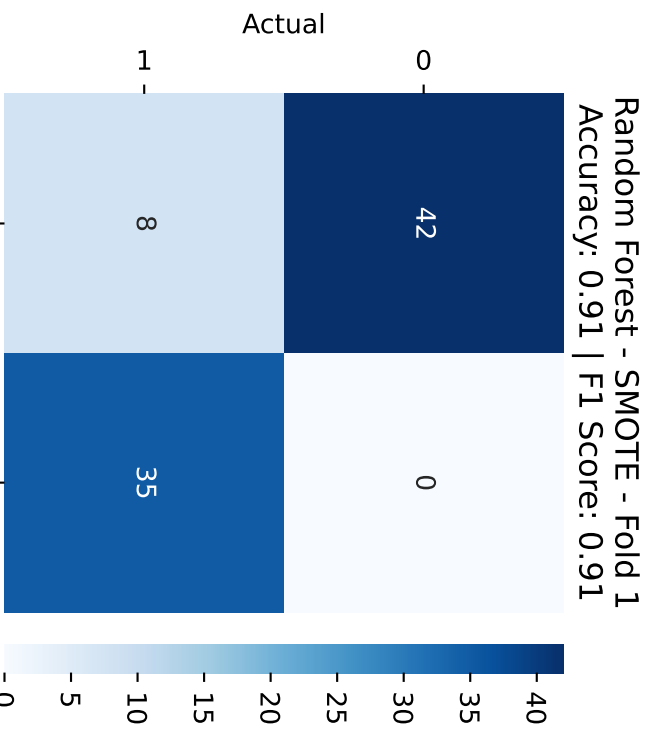


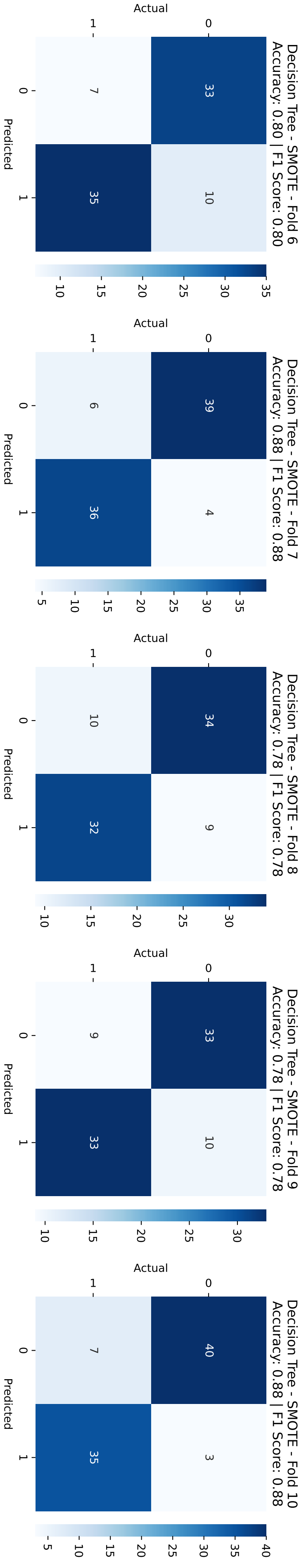
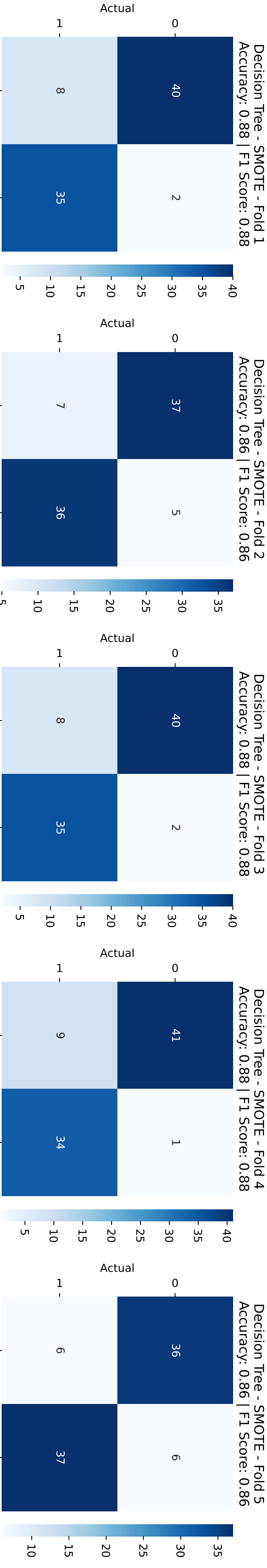


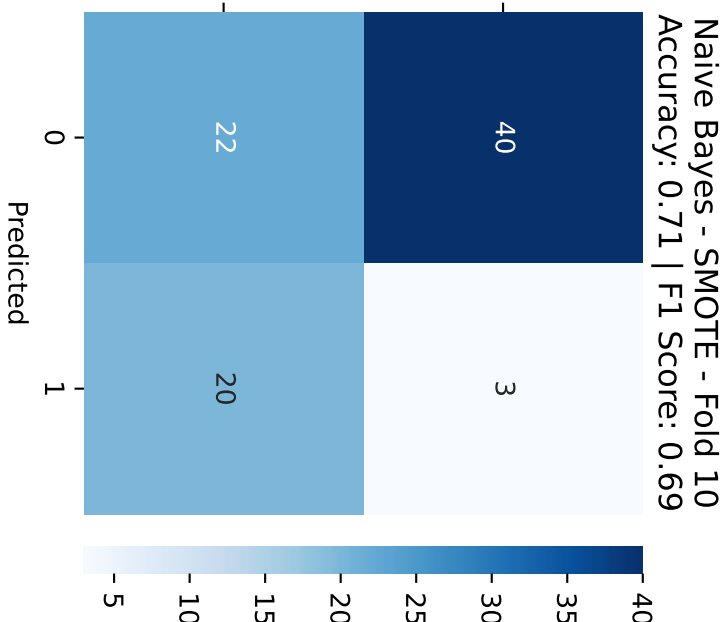
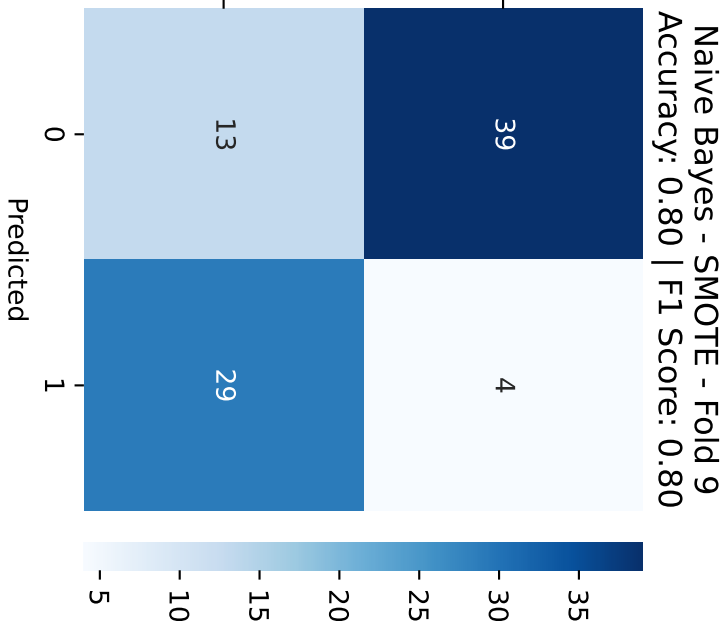
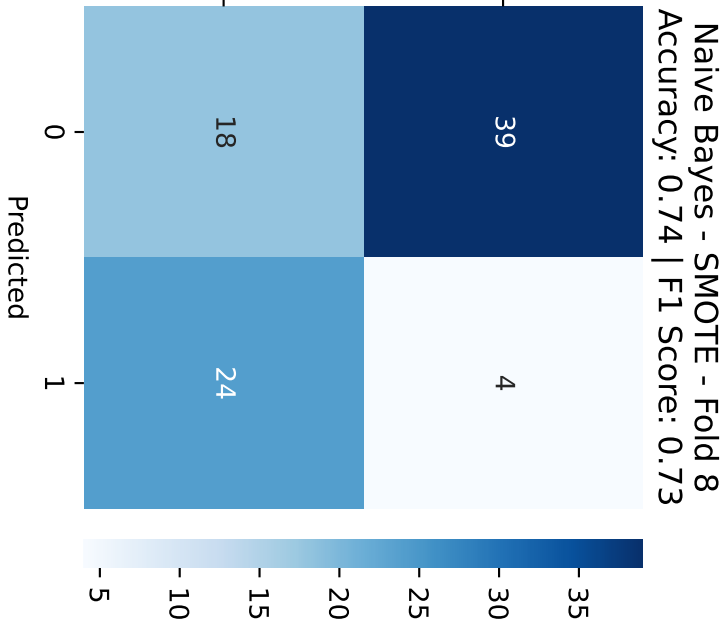
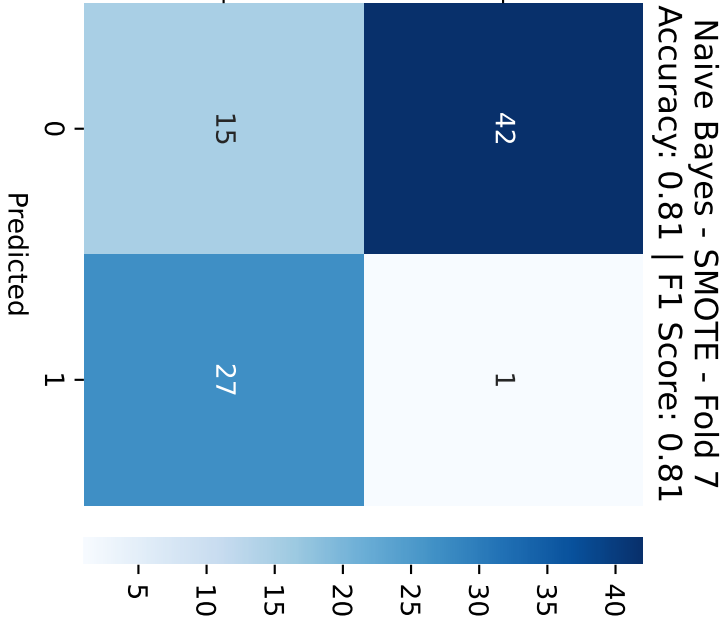
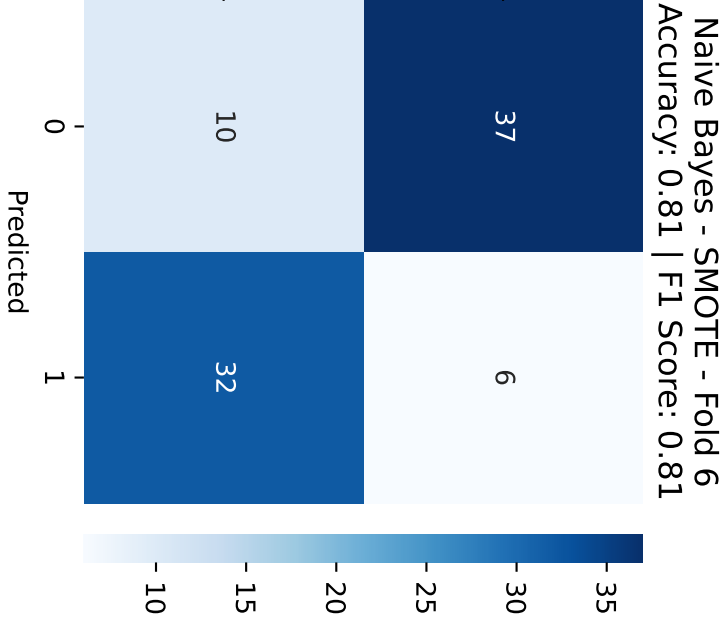
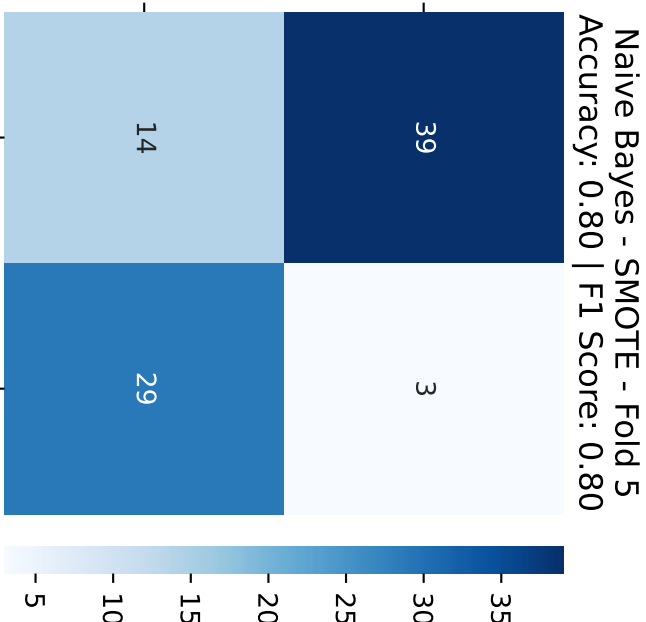
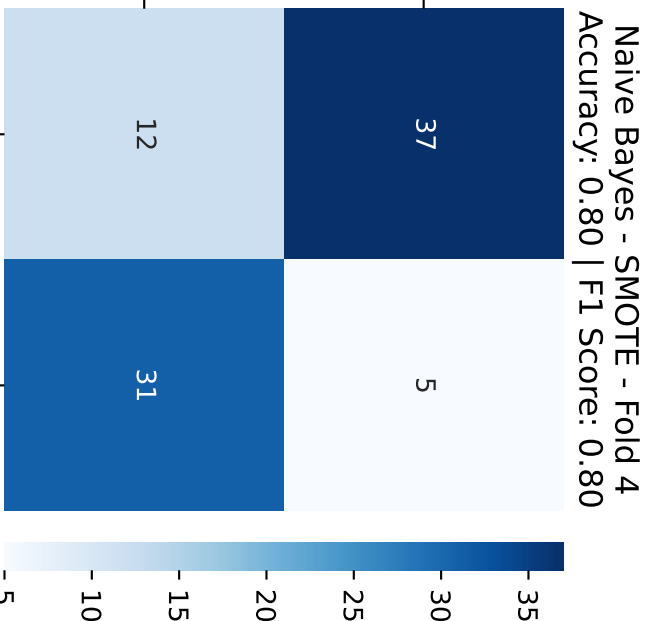
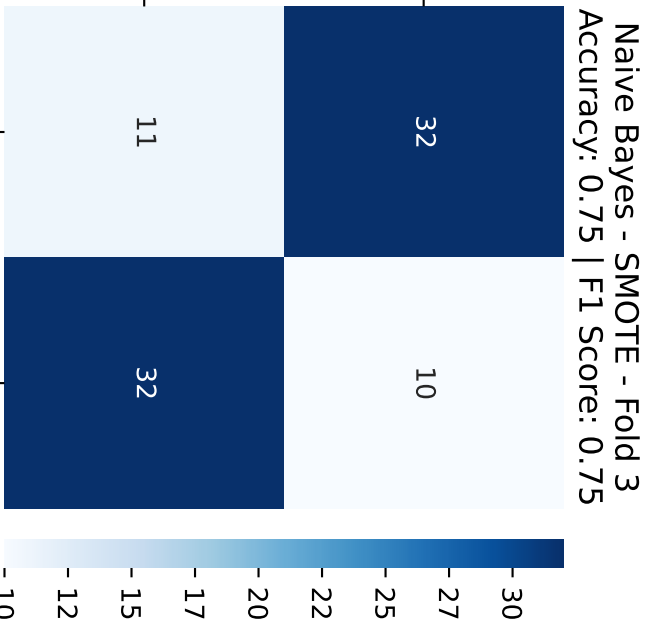
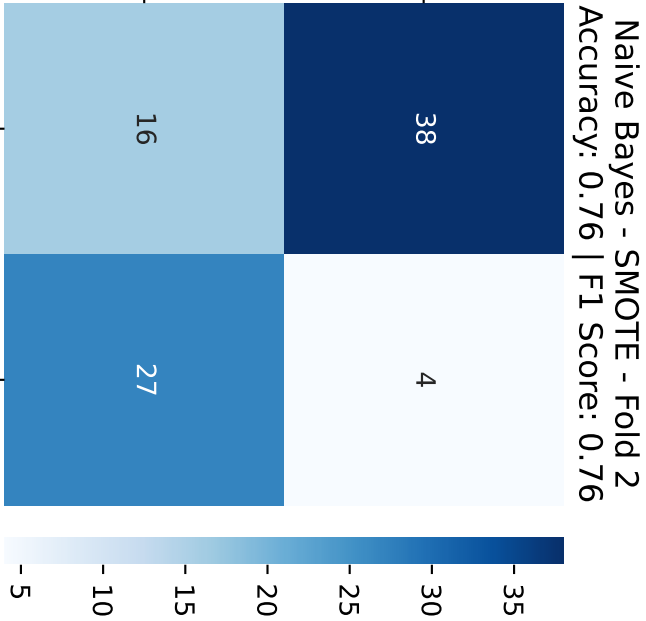
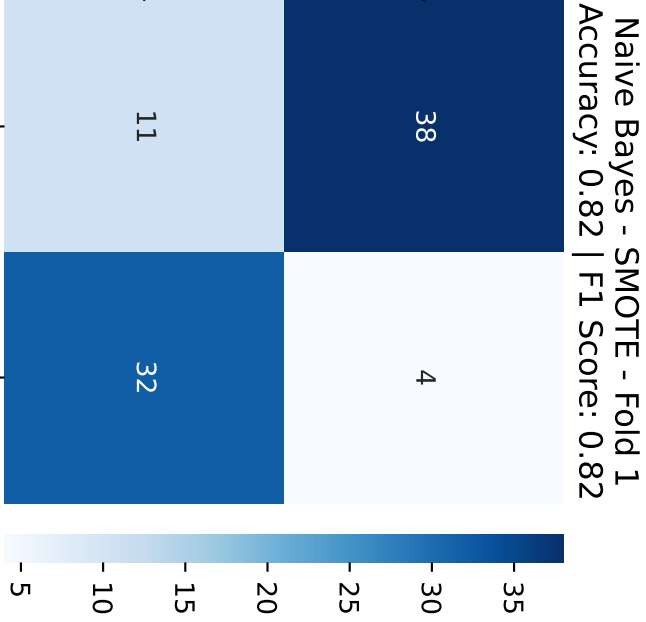


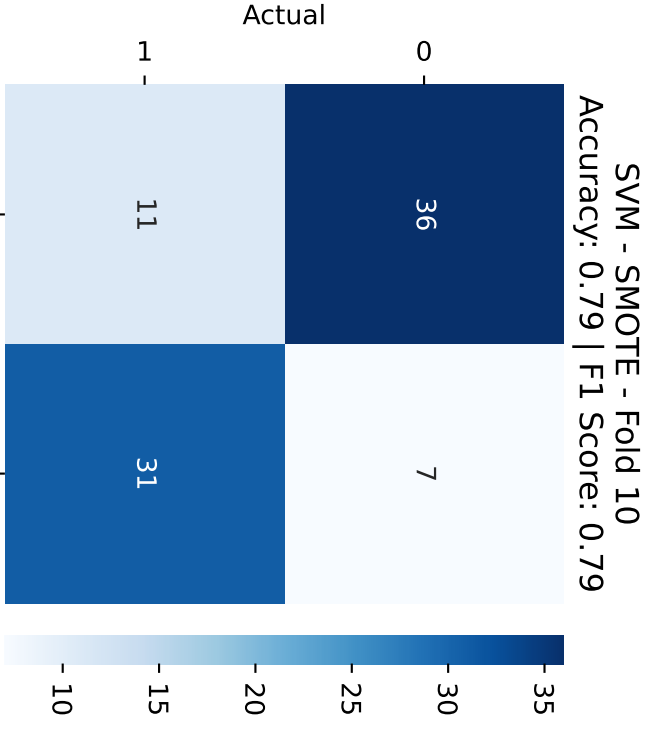
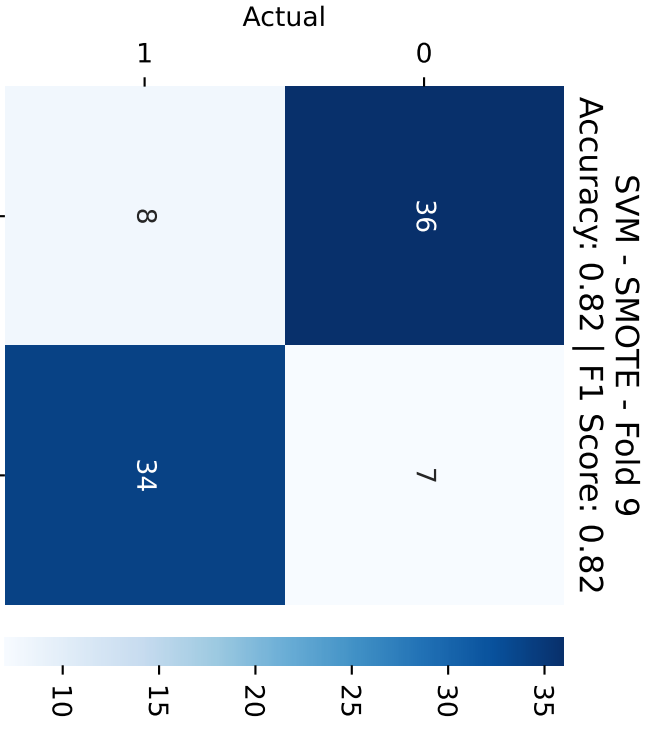
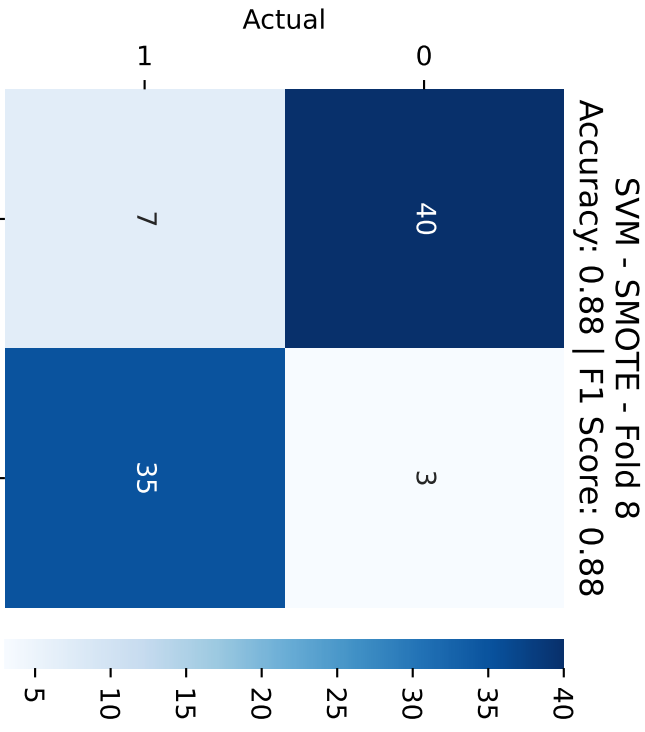
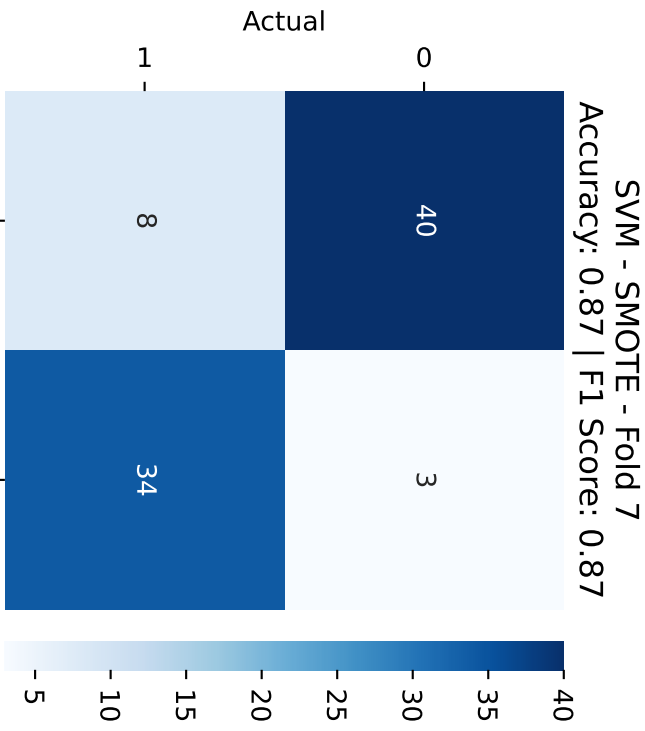
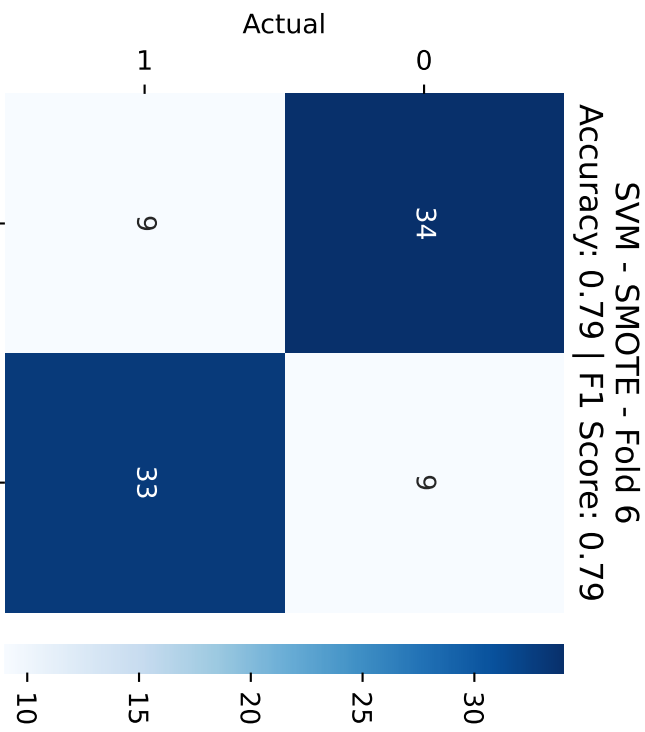
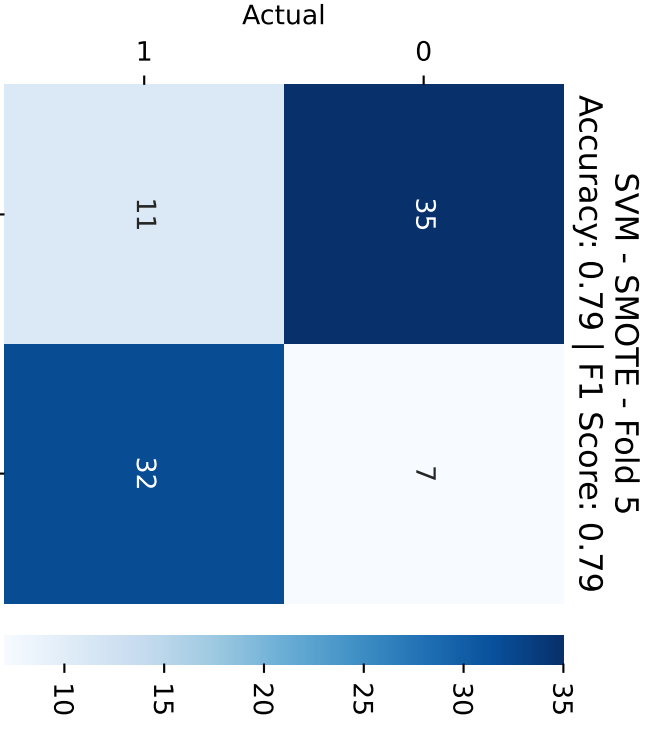
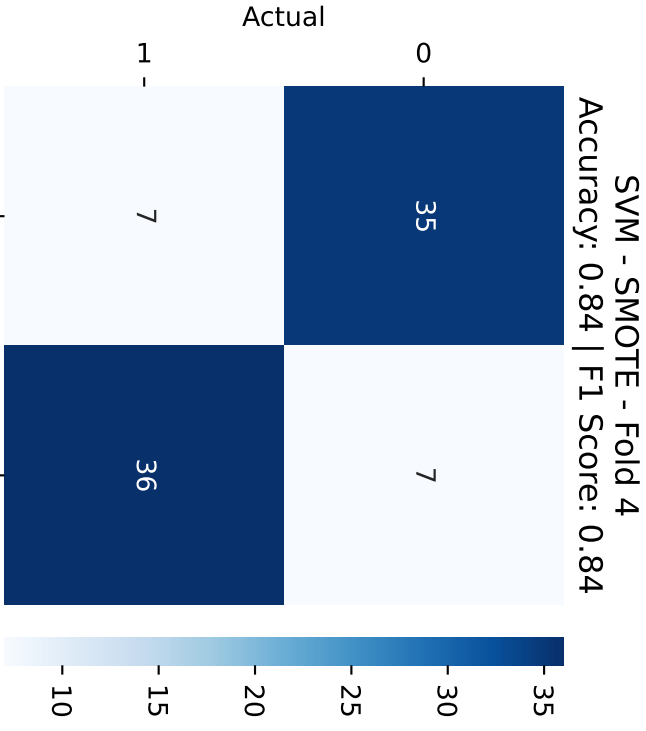
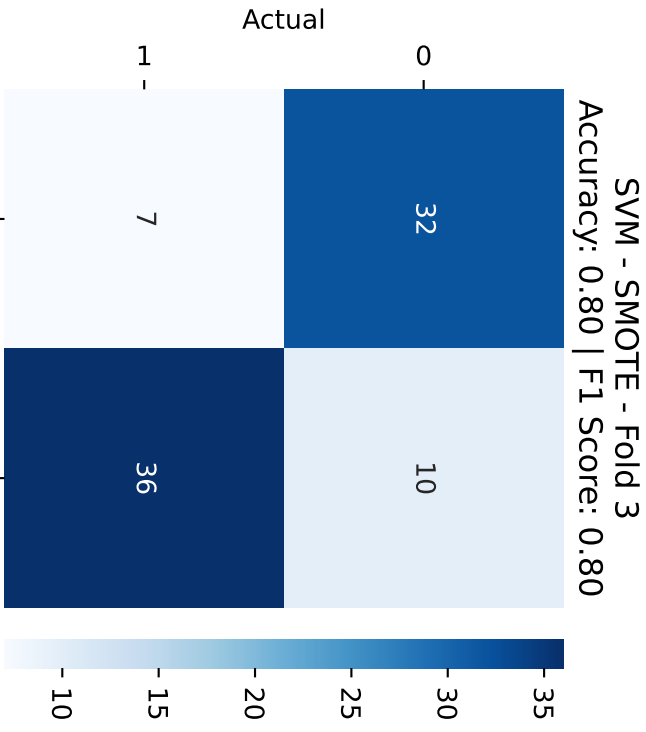
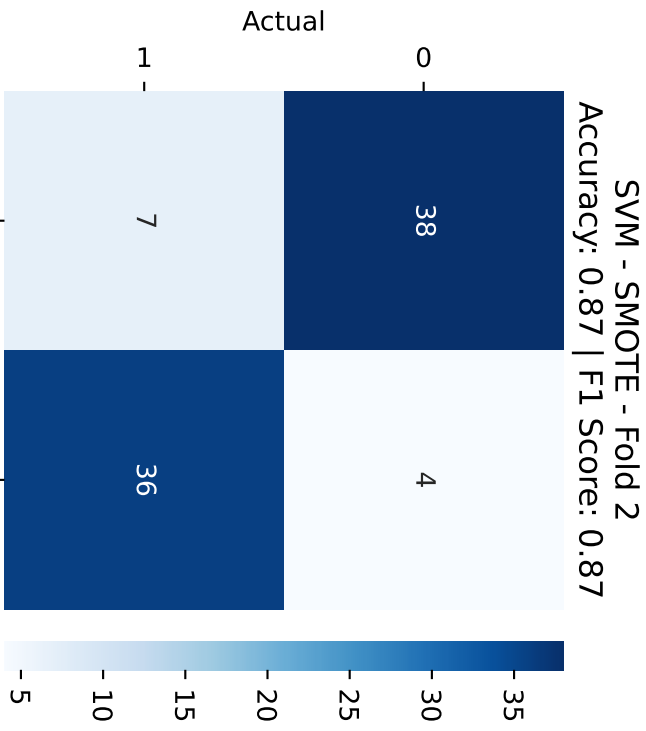
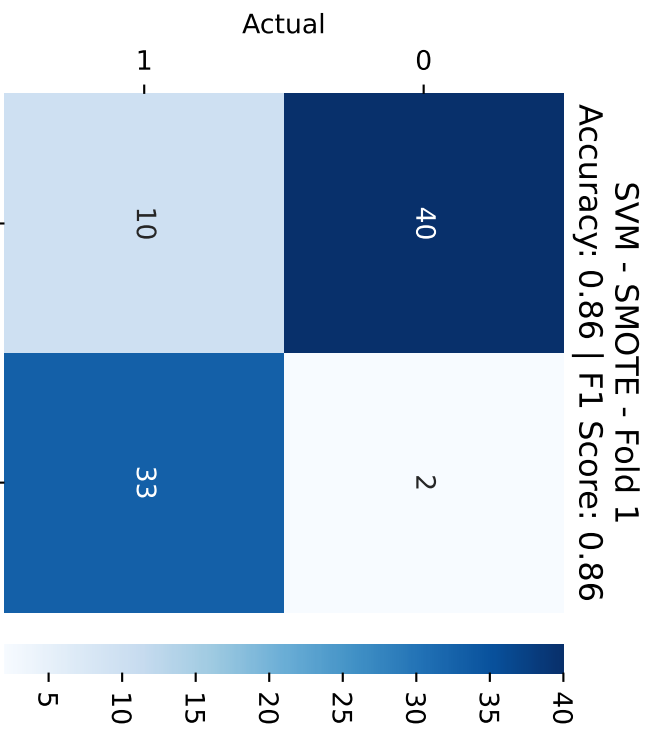


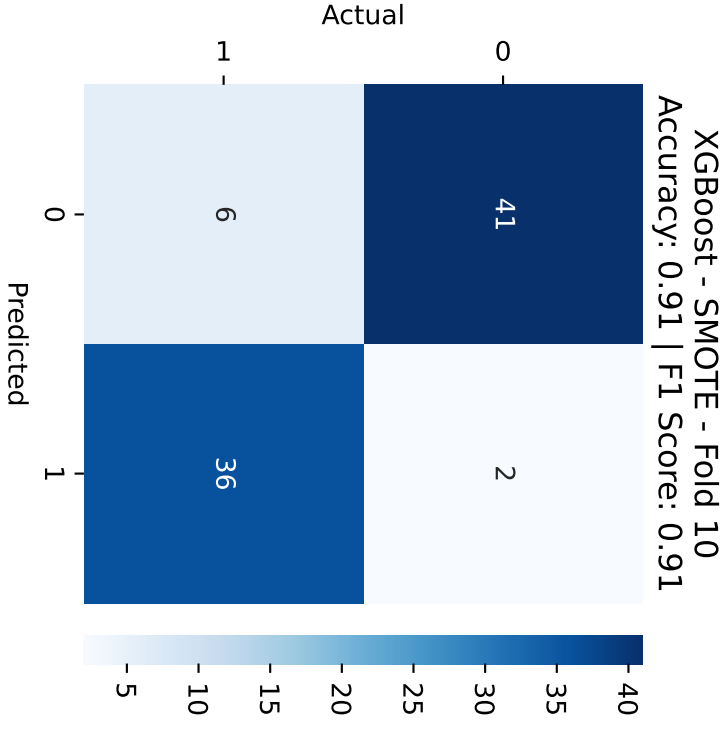
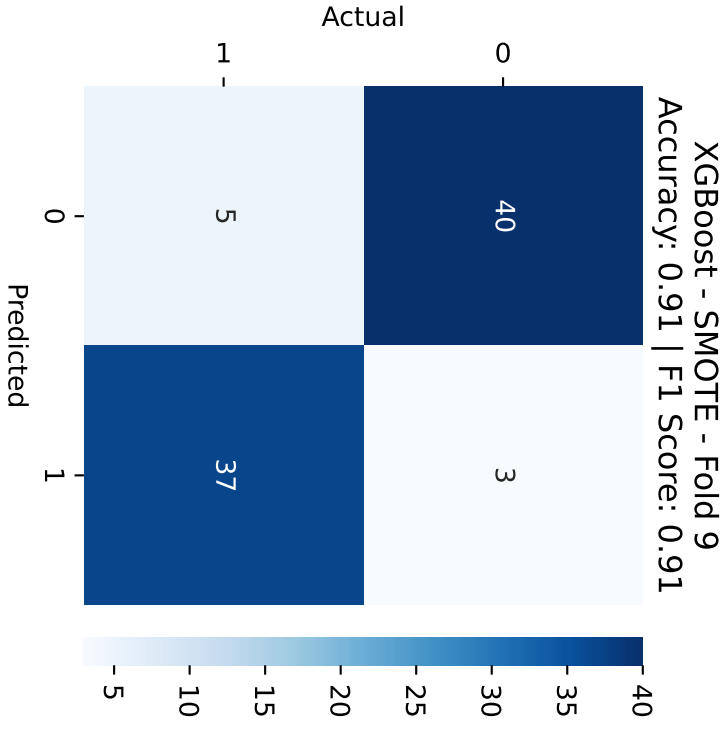
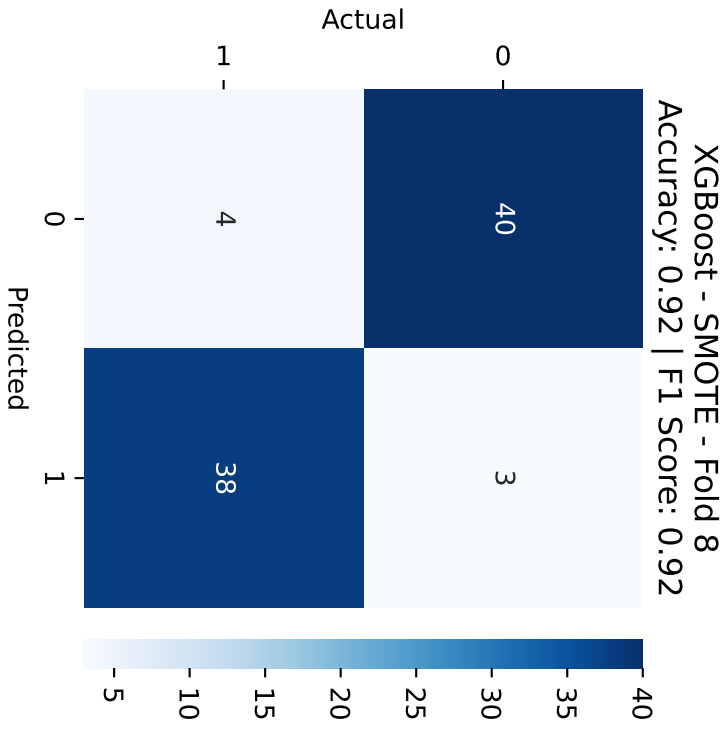
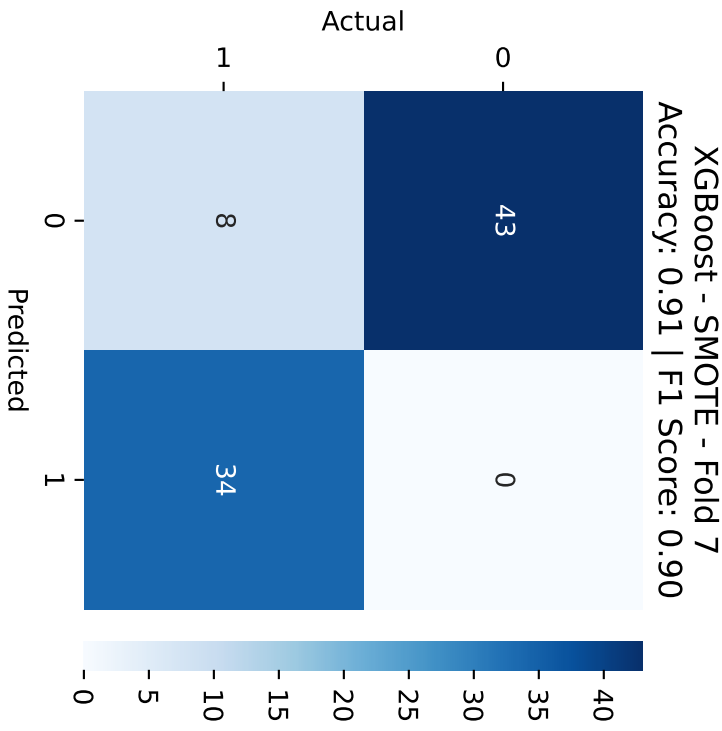
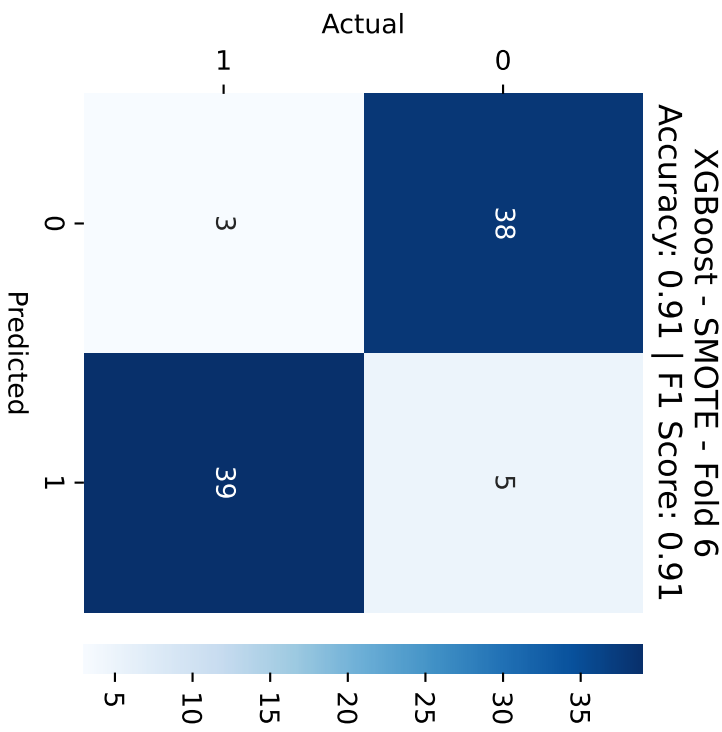
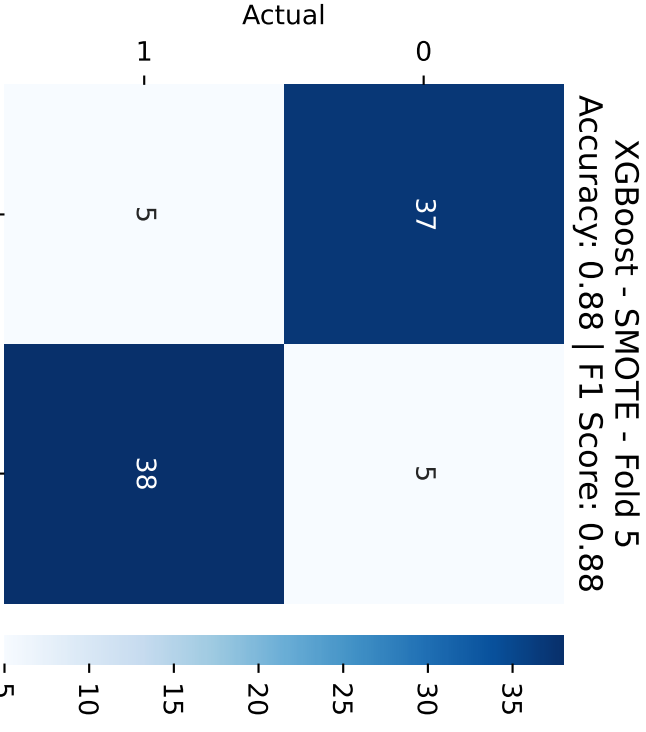
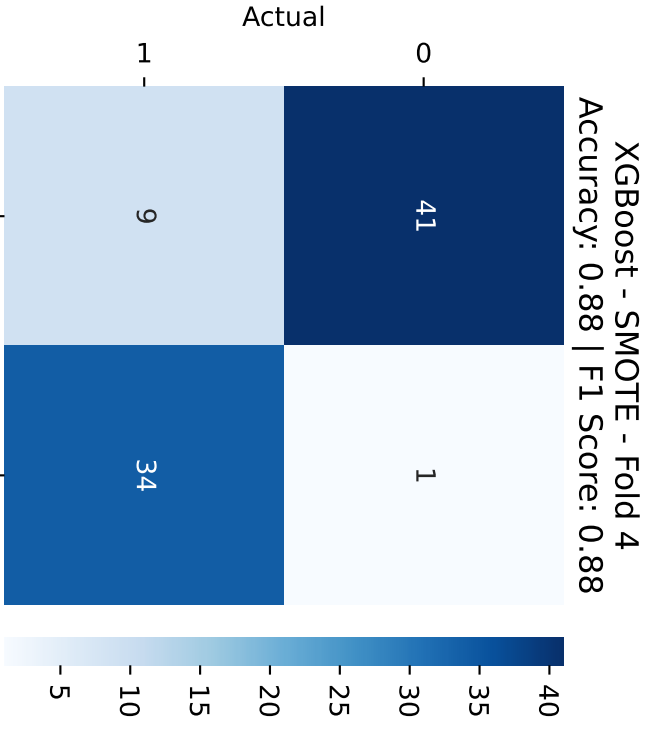
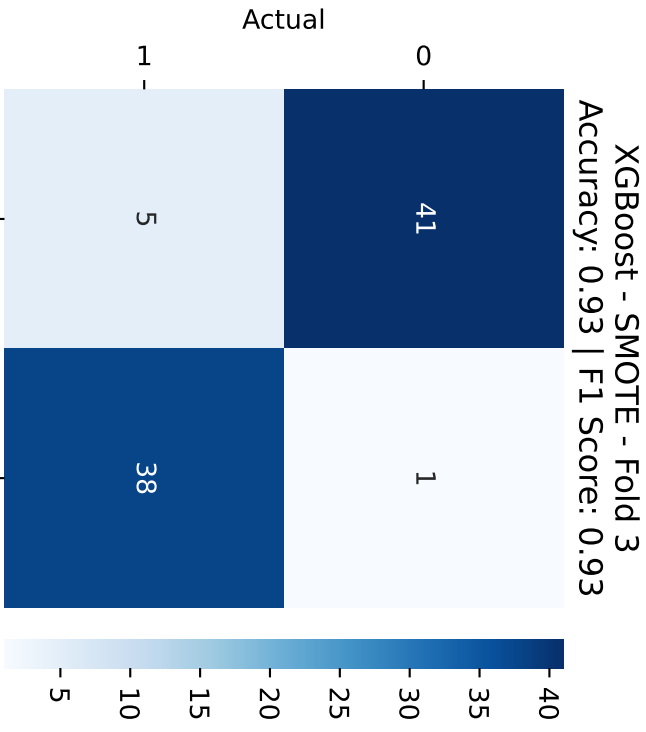
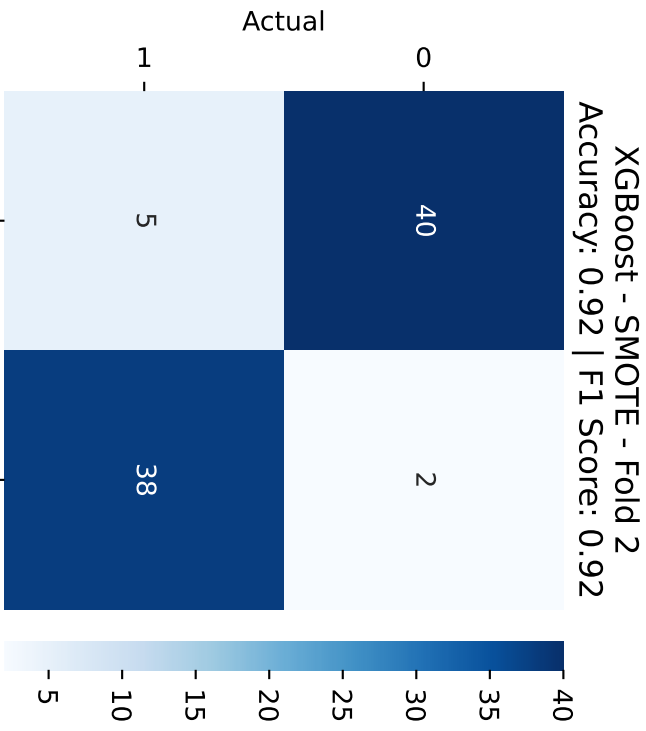
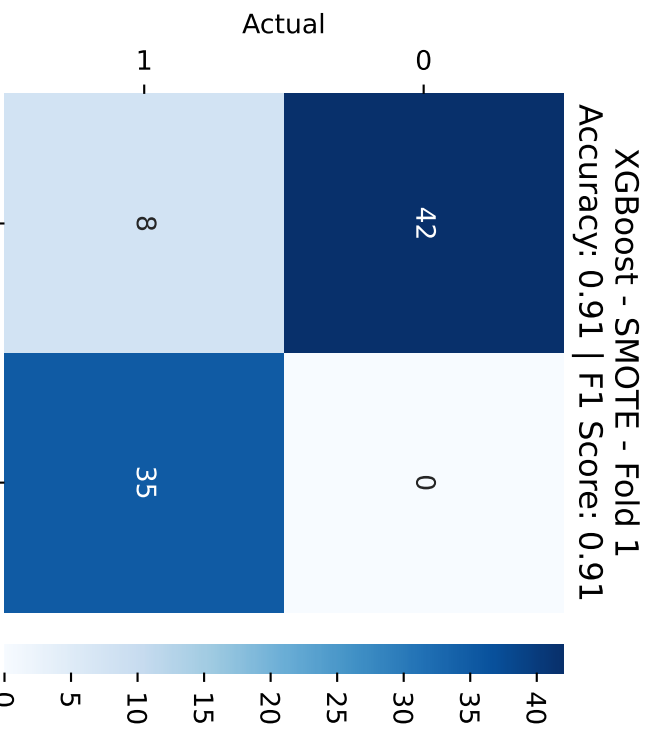


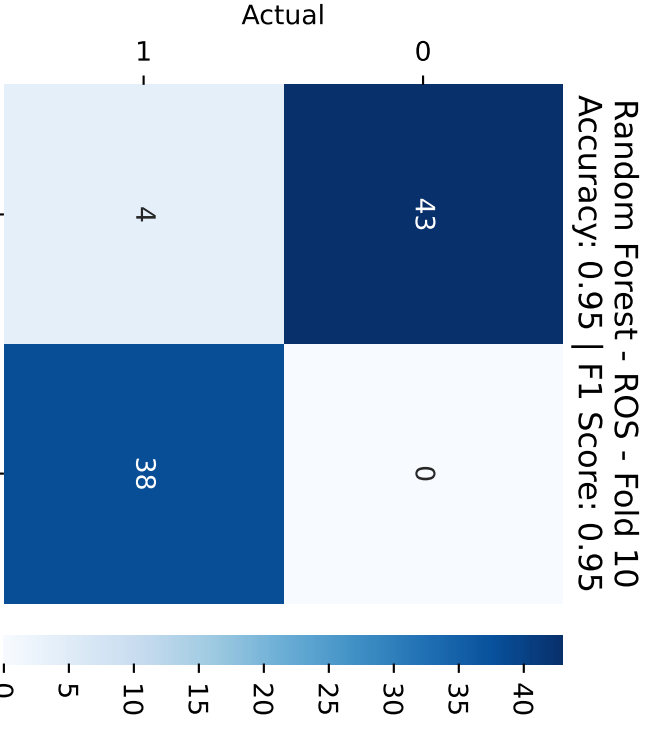
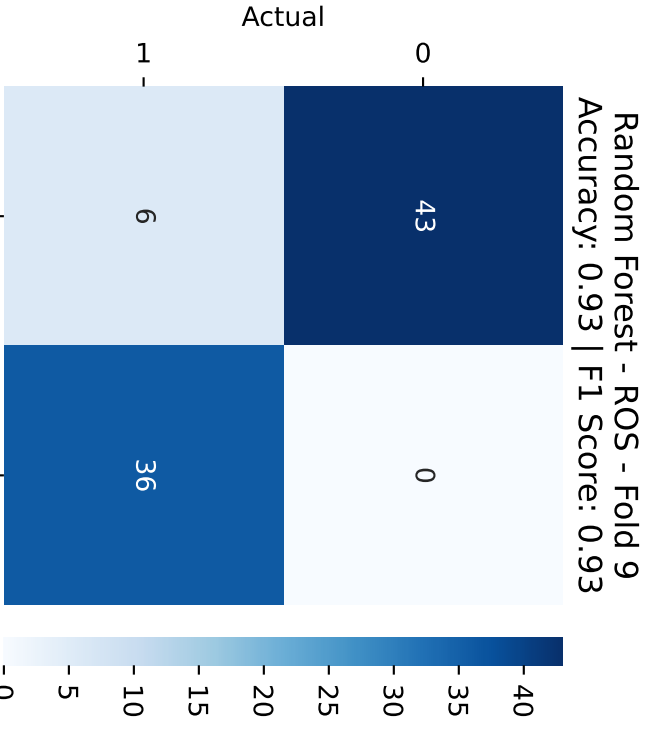
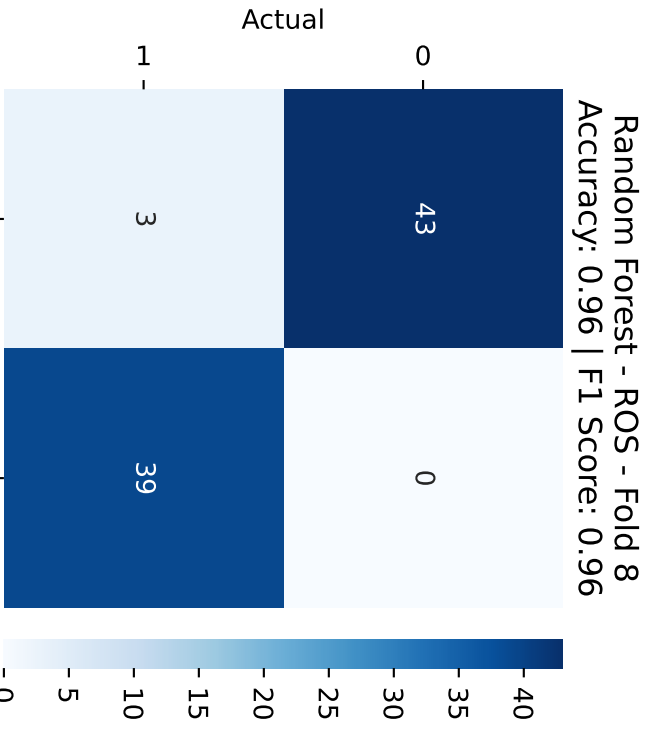
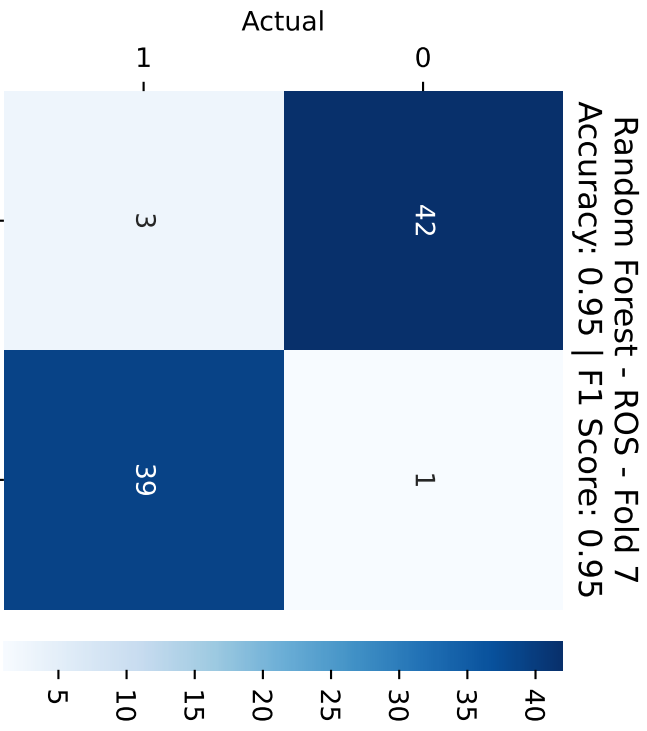
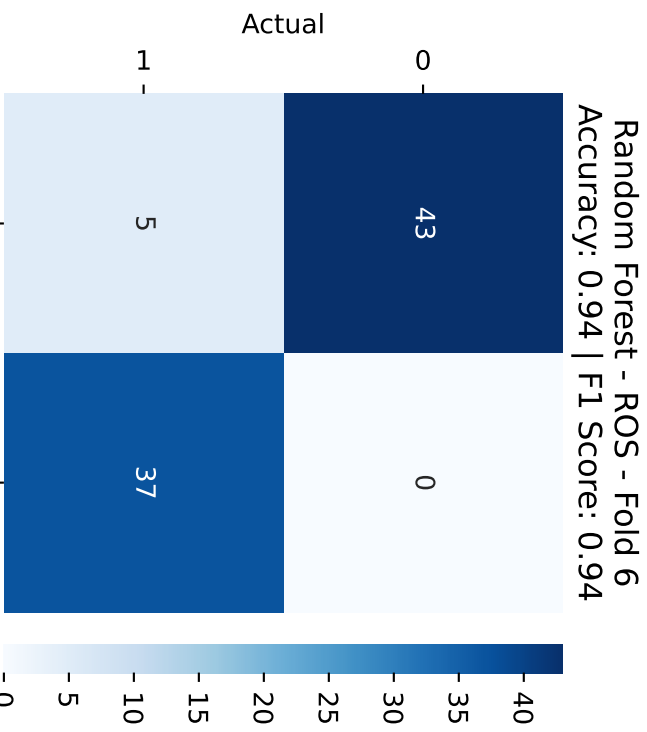
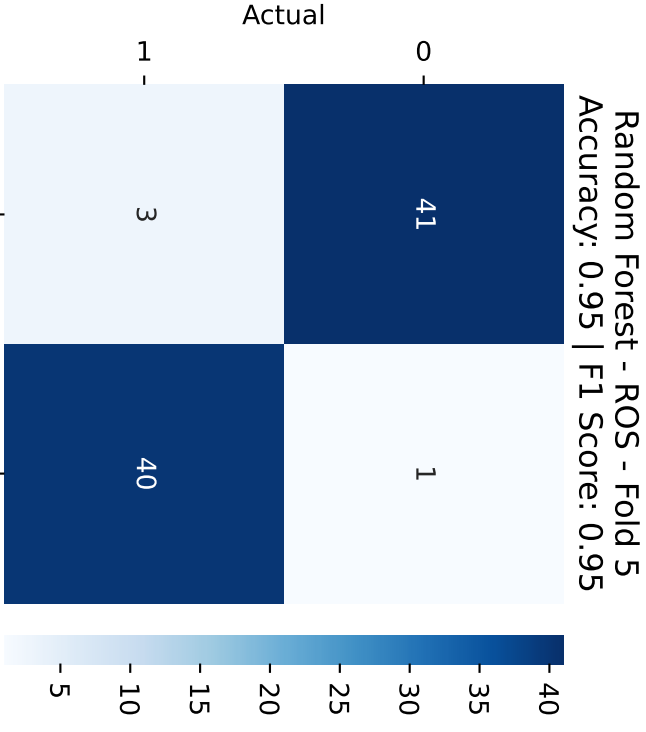
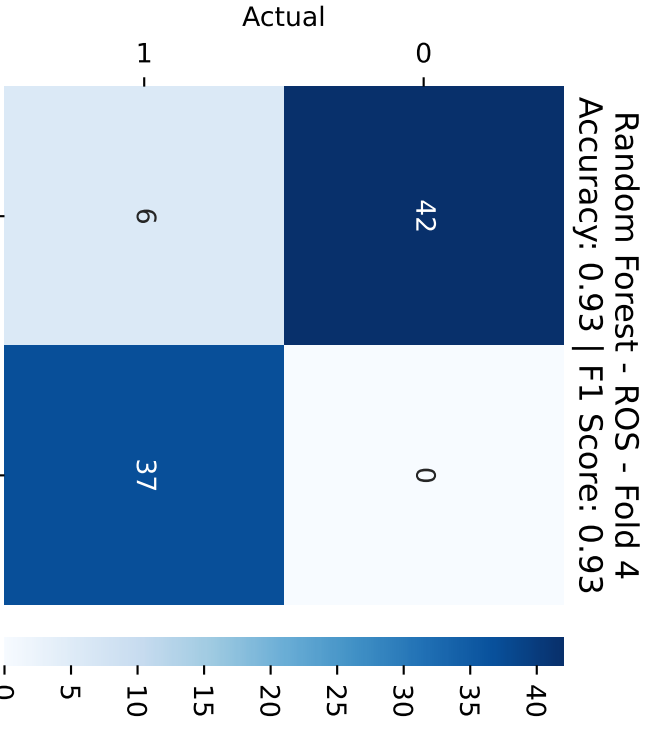
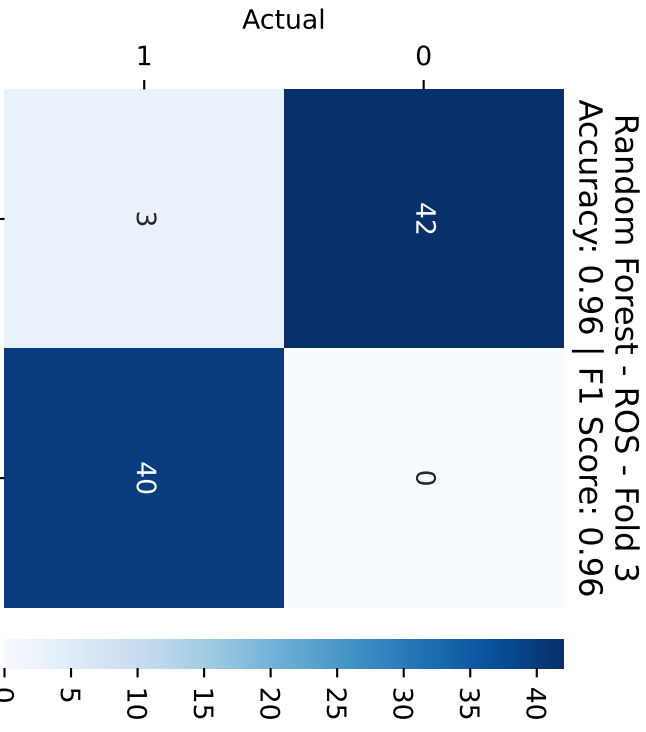
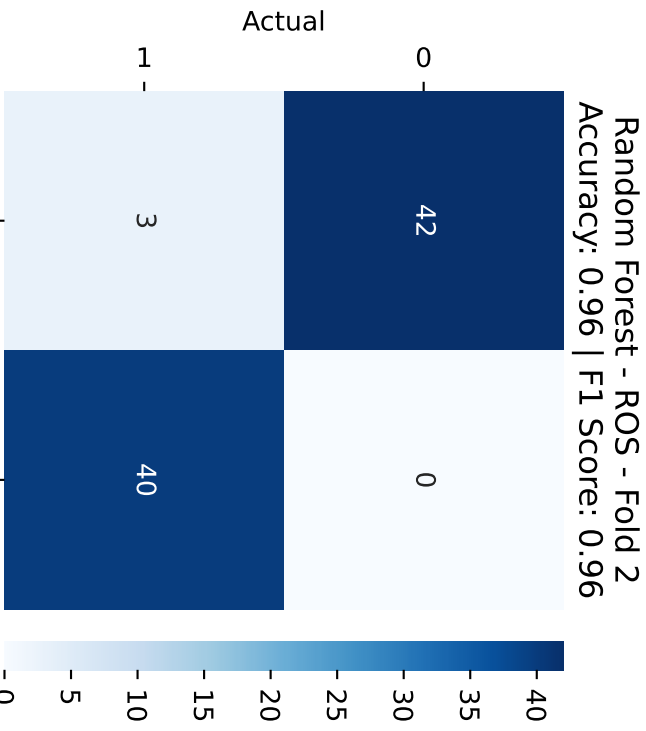
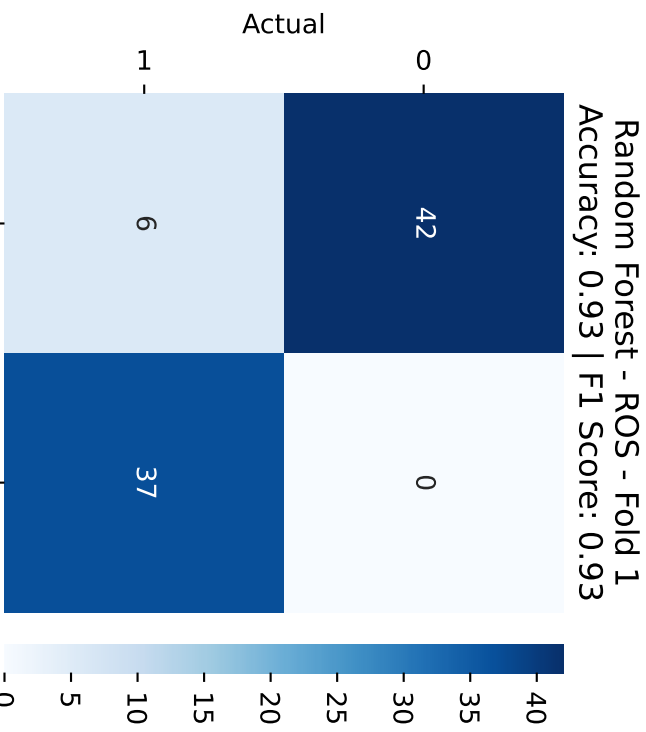


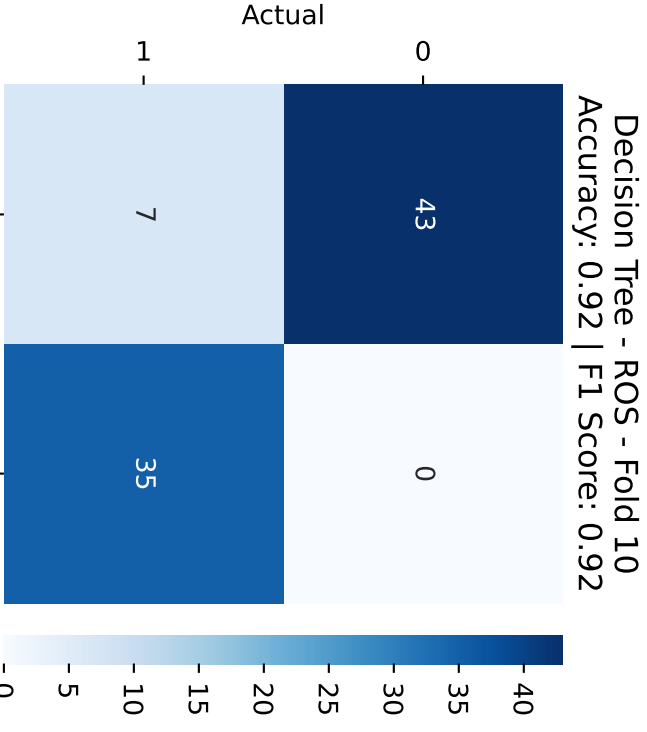
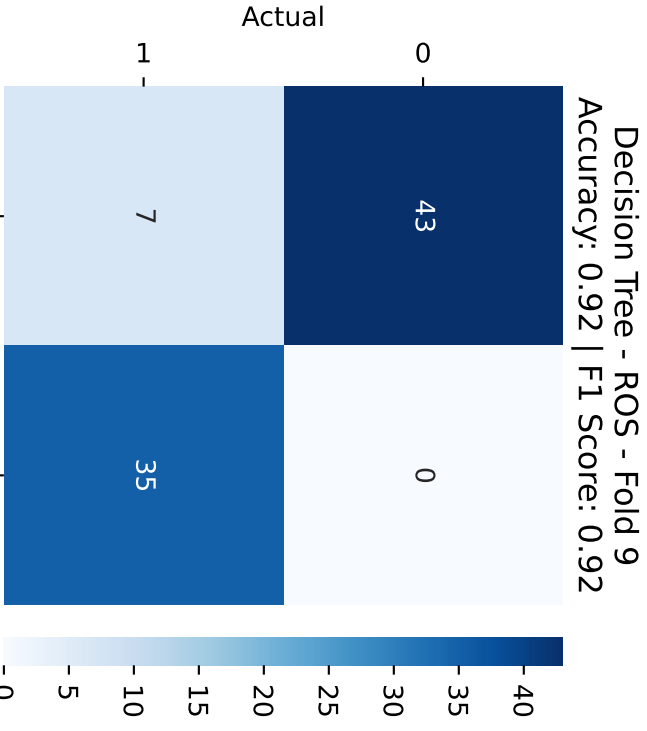
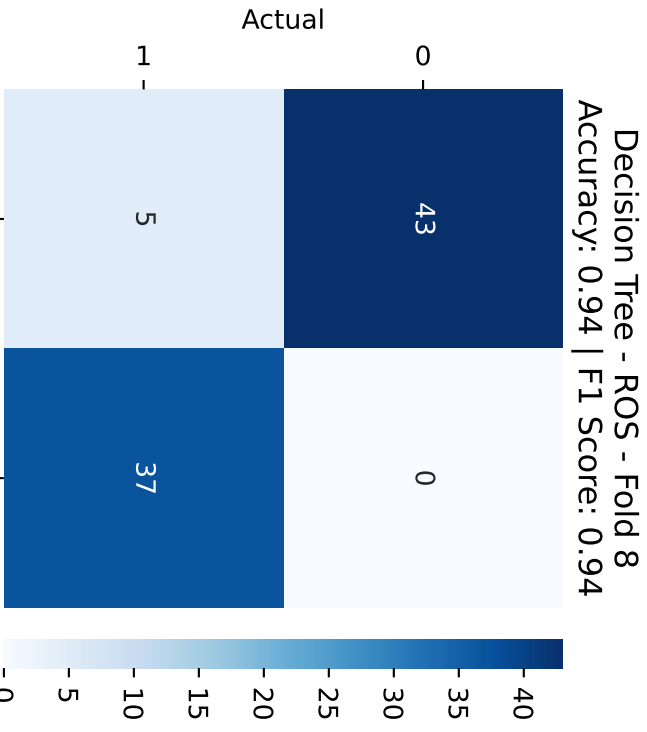
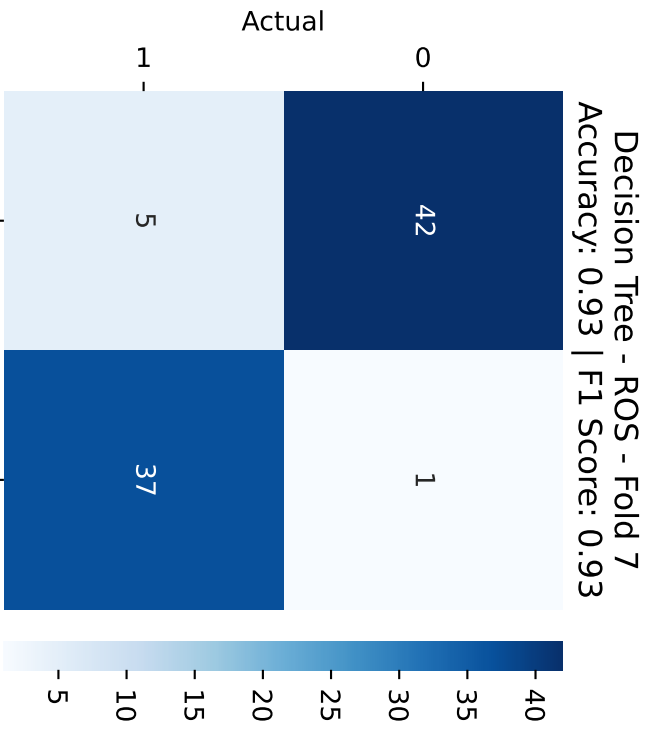
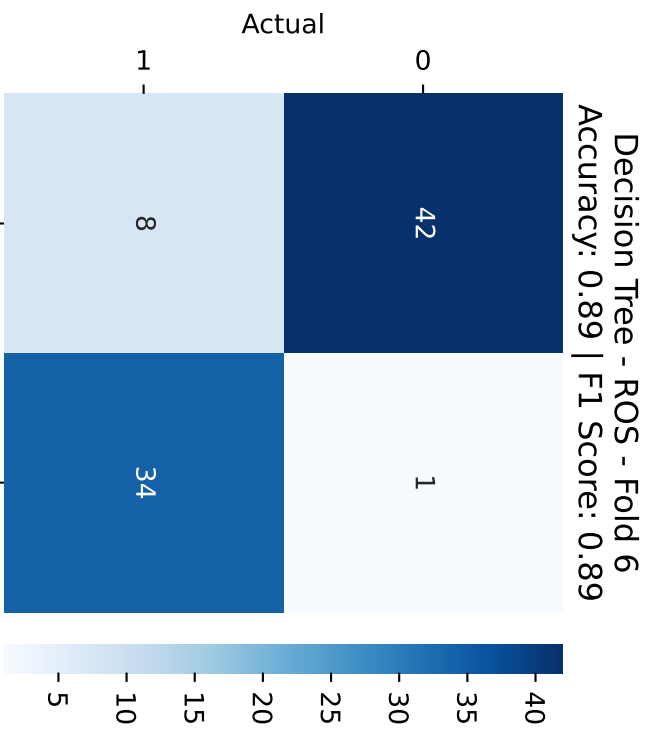
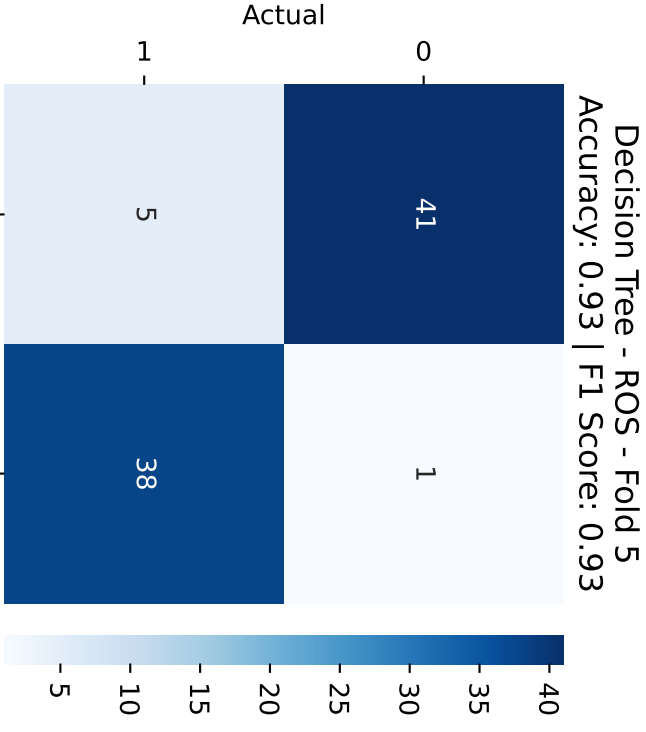
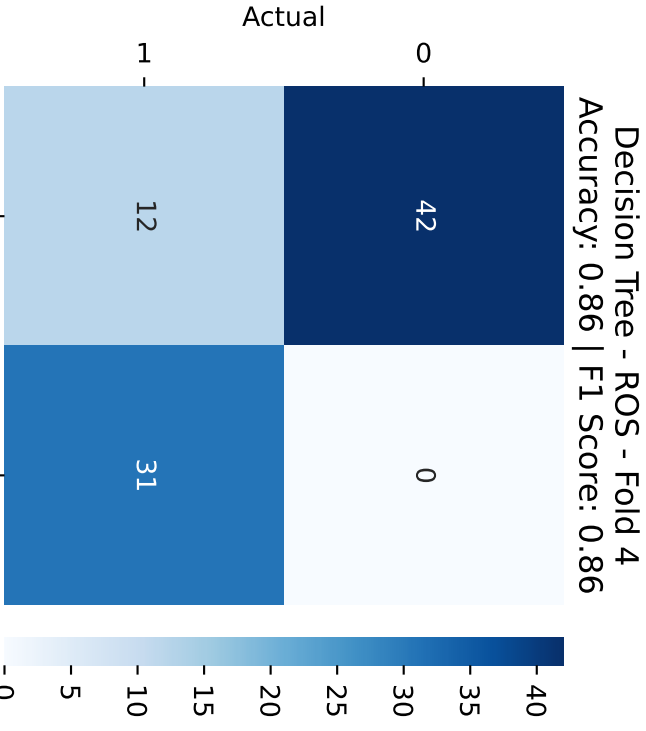
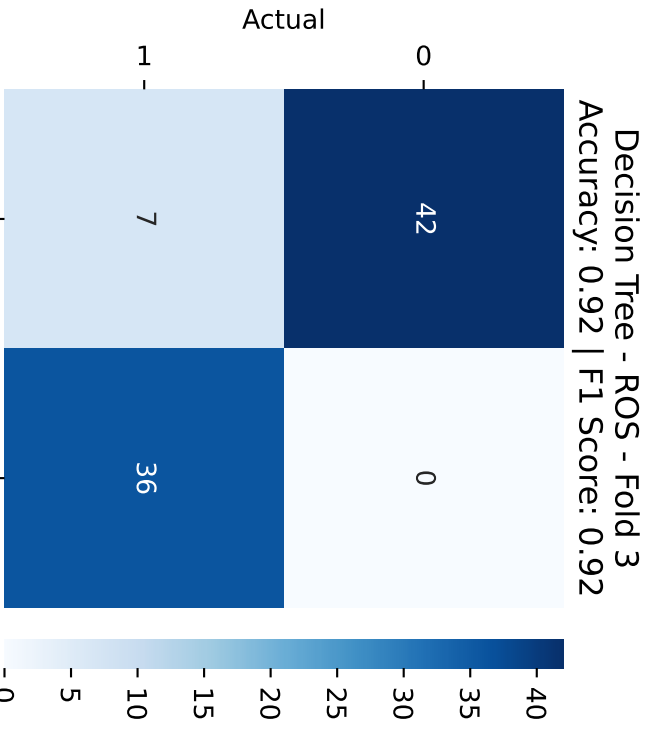
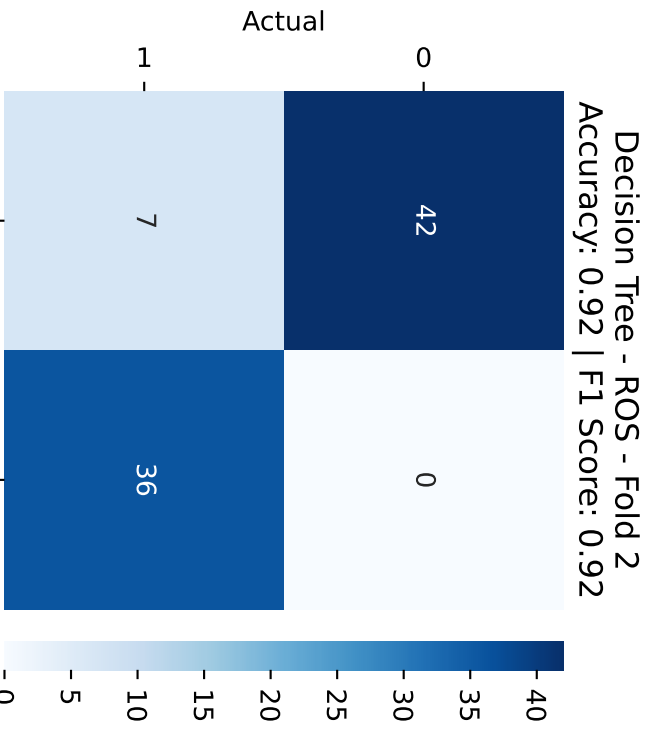
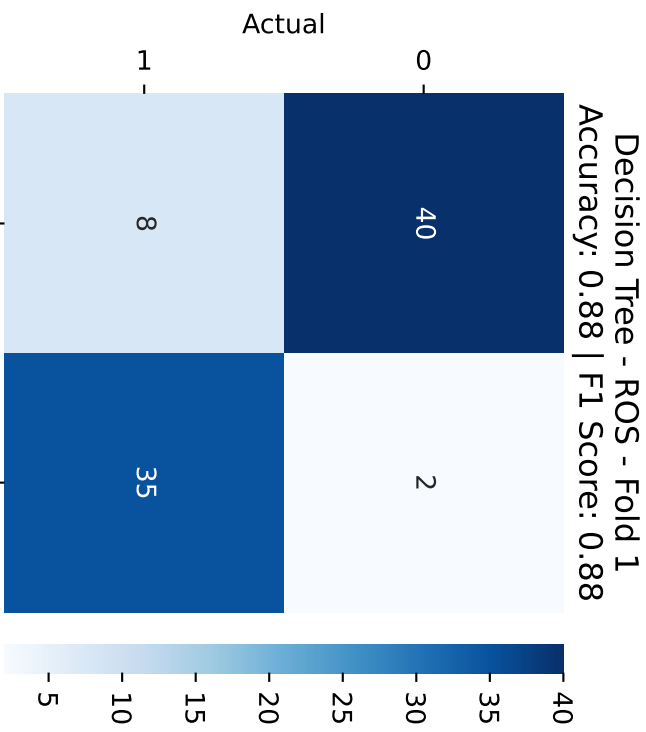


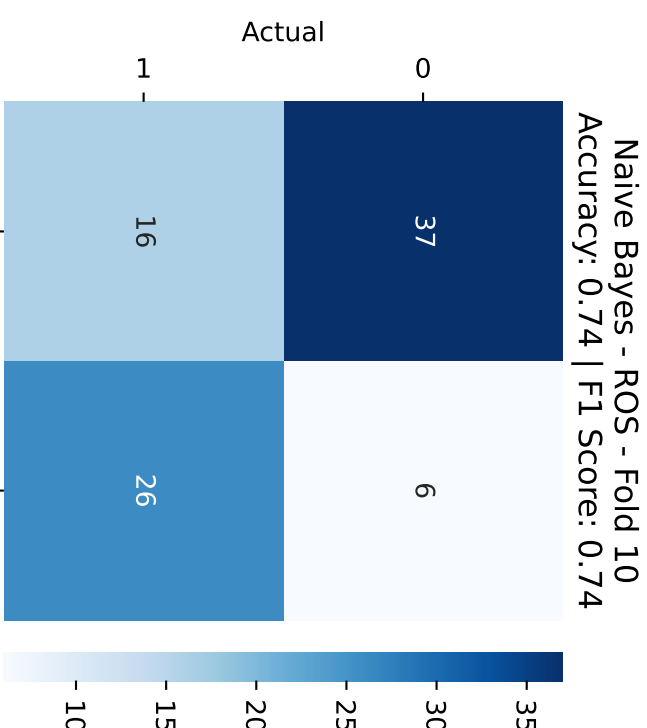
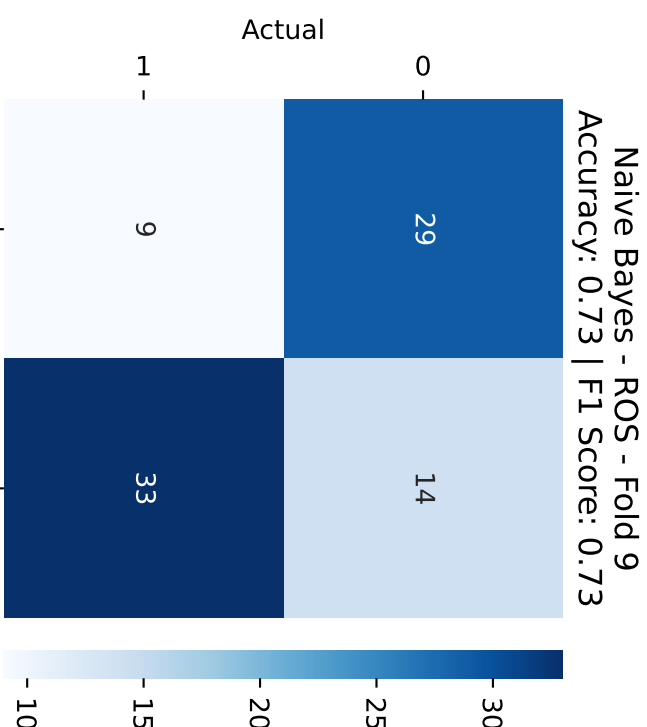
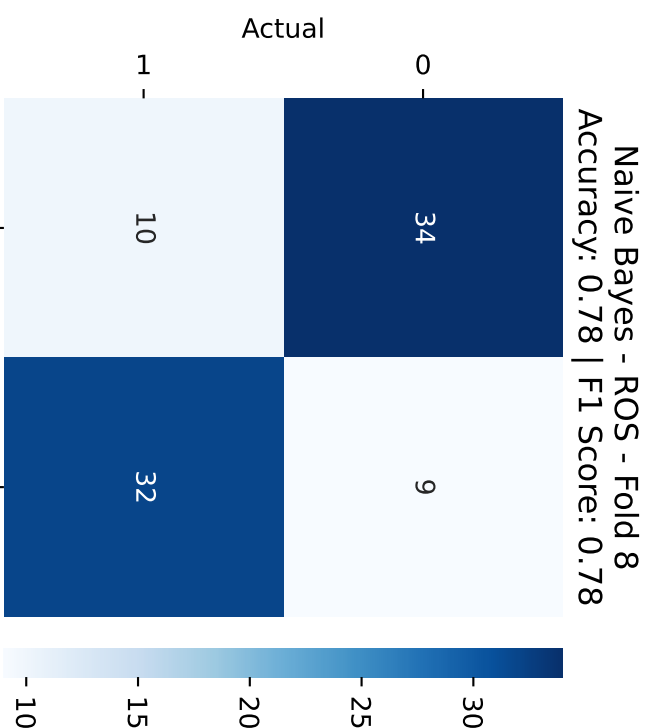
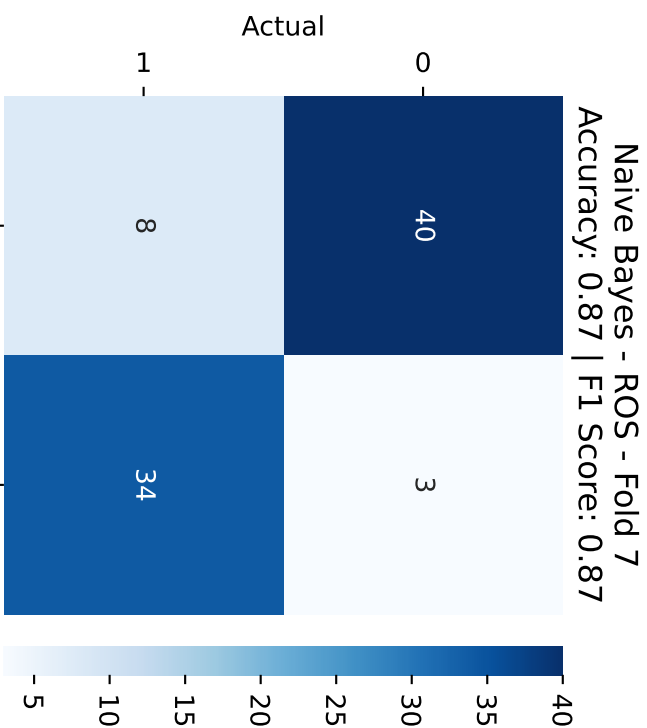
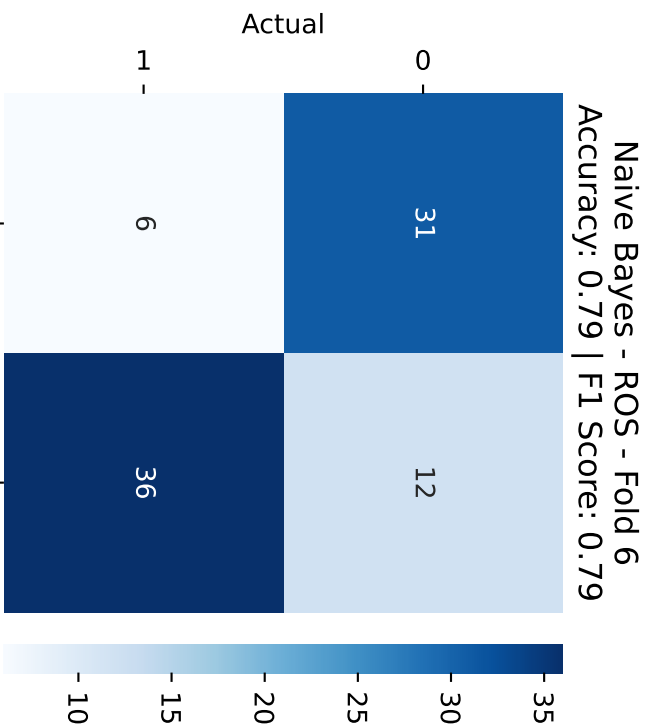
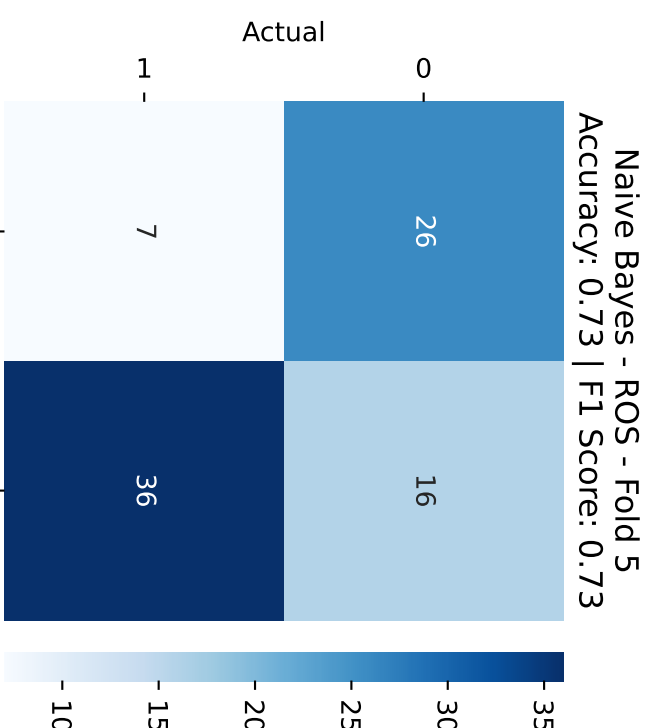
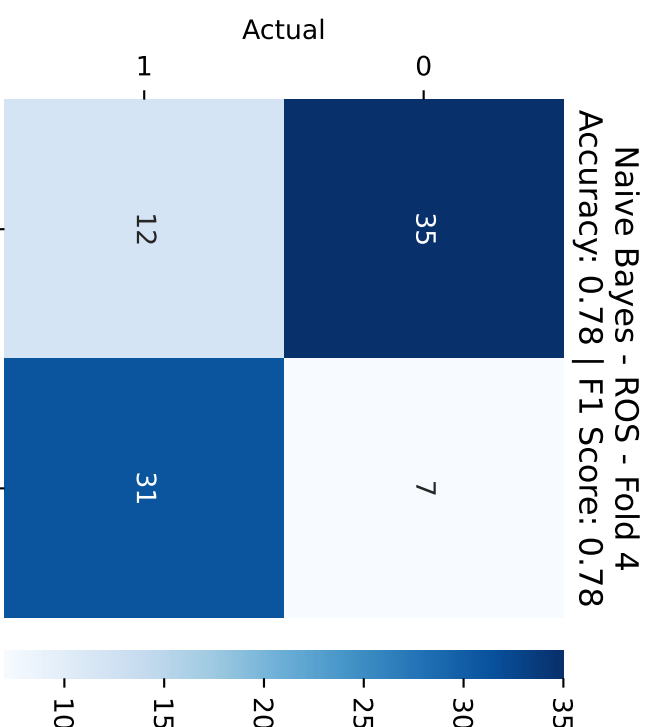
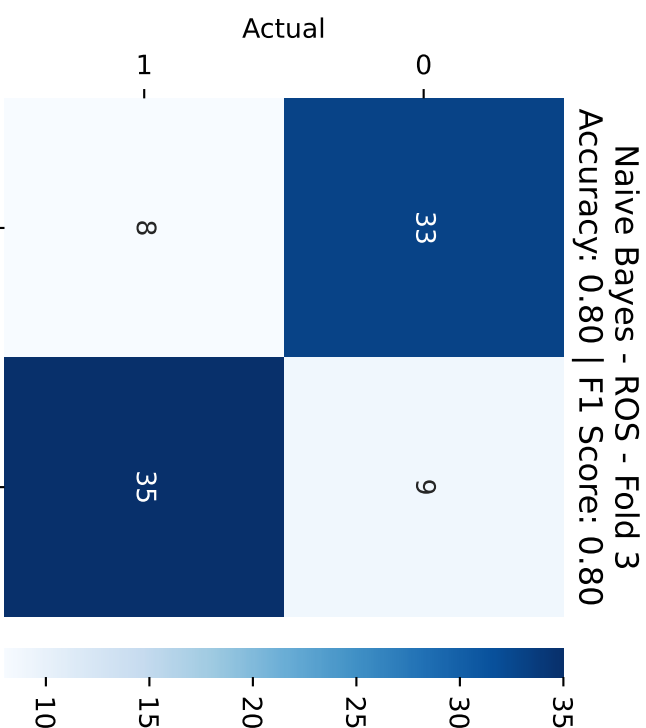
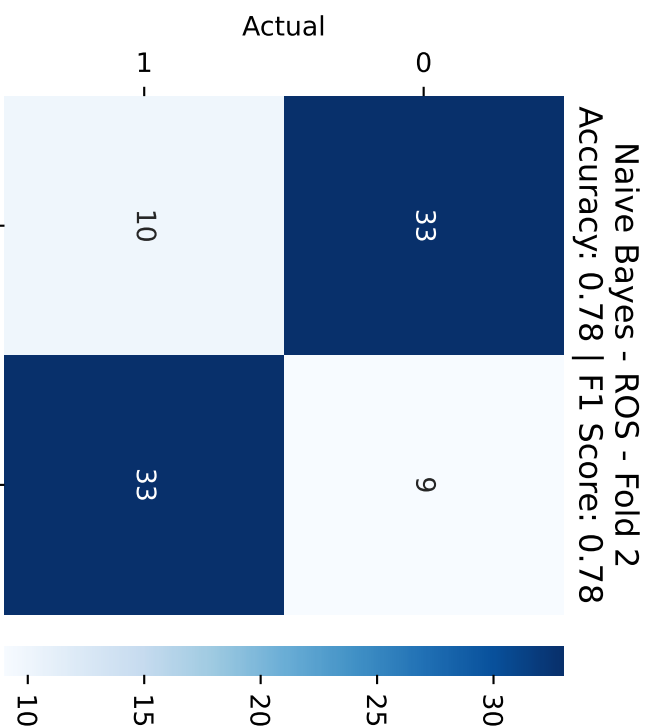
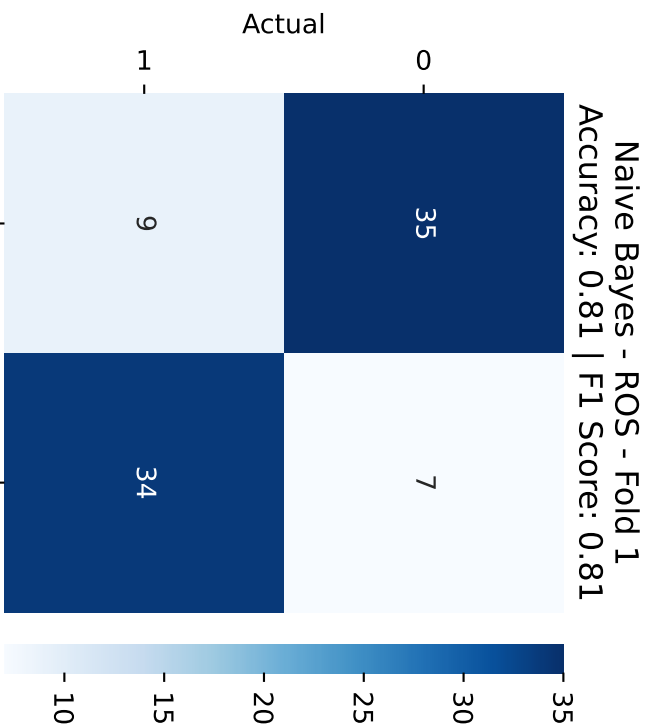


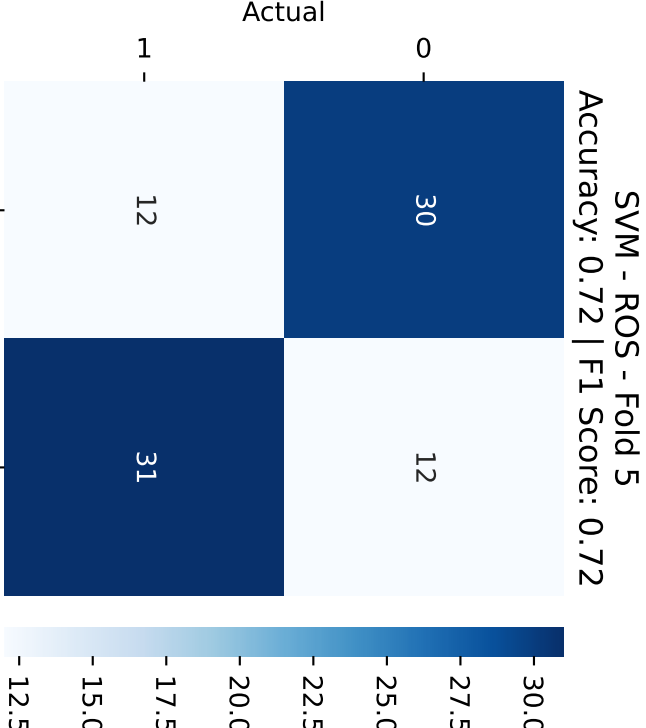
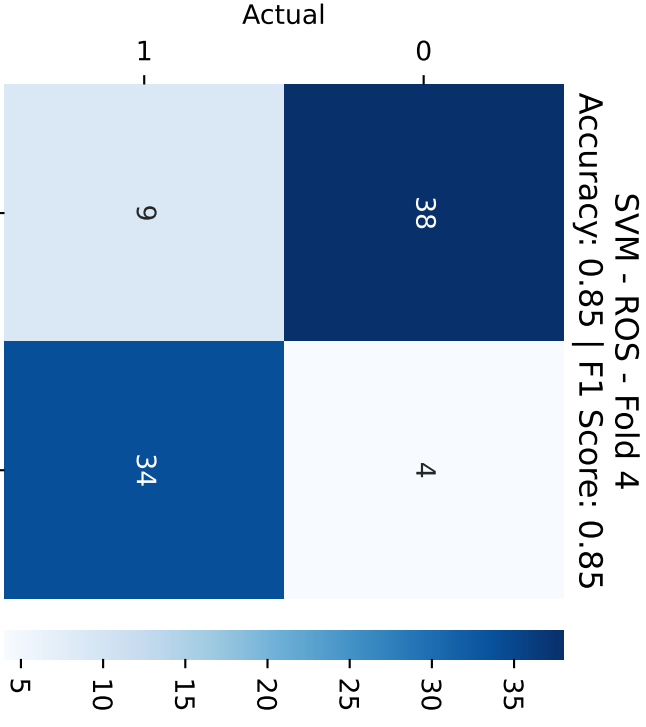
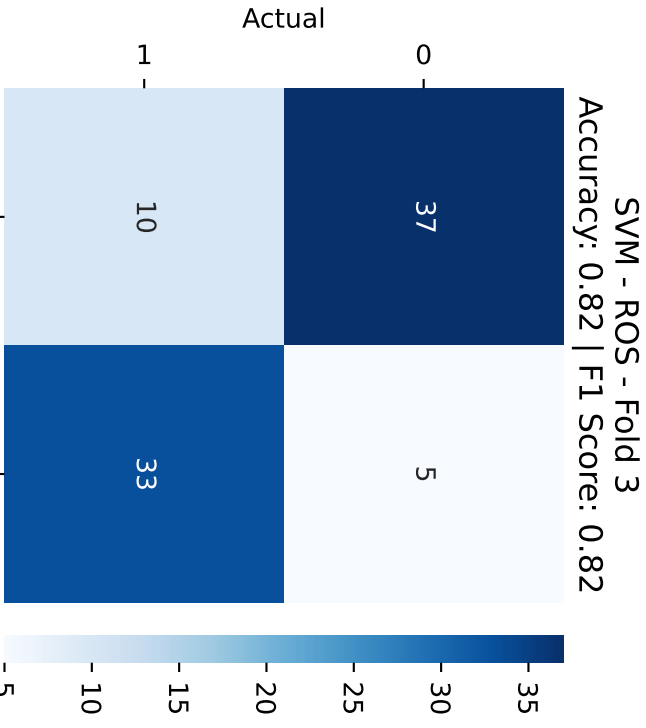
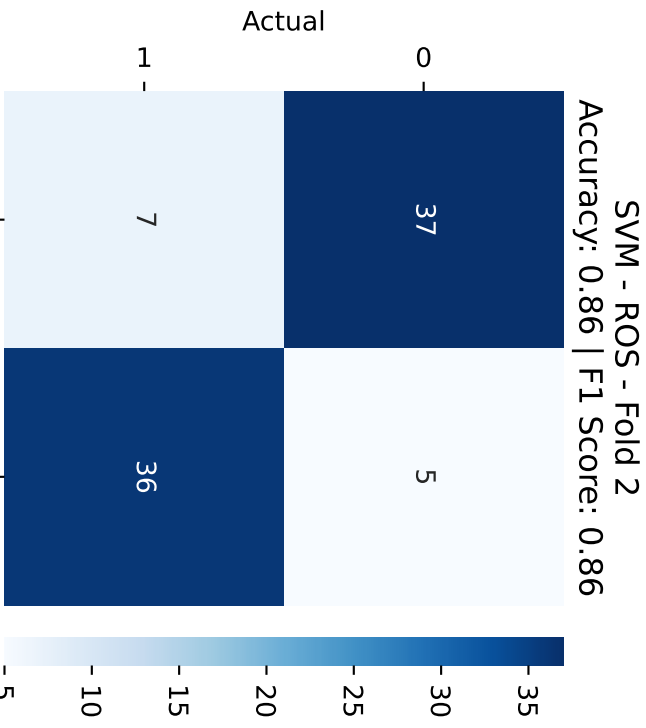
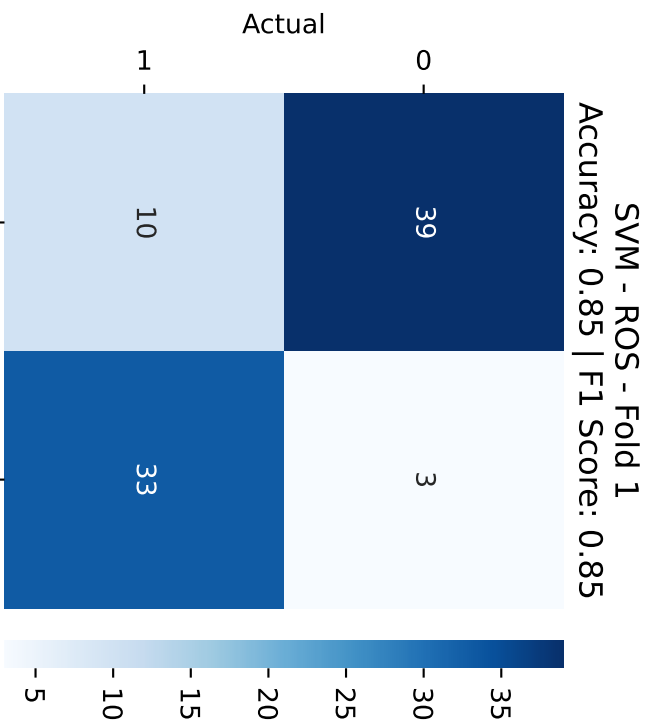




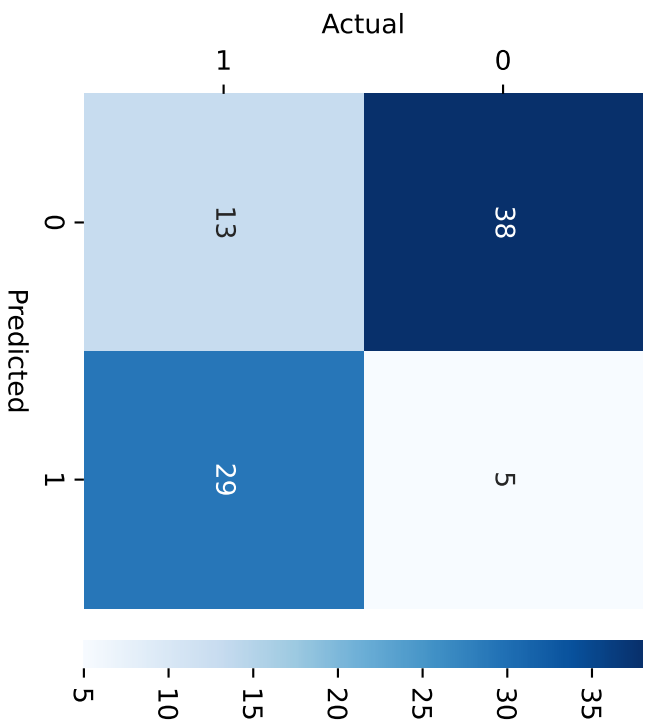




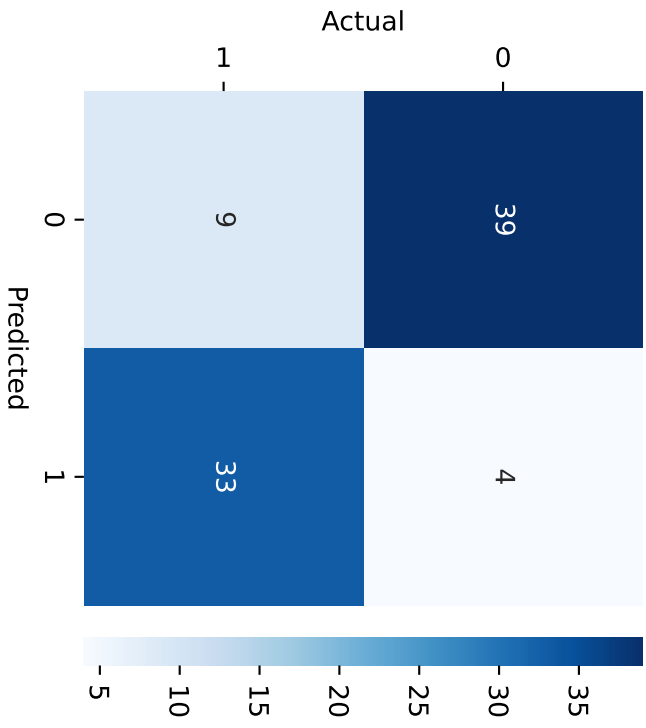




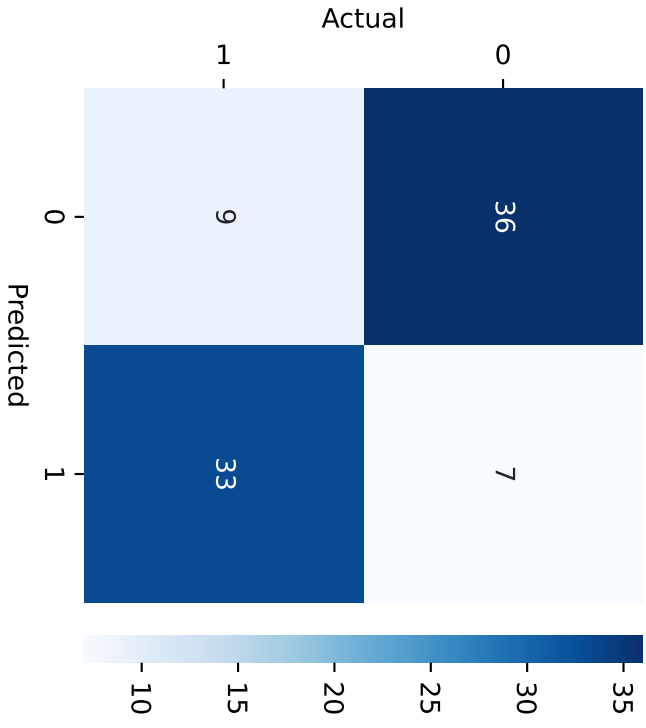
SVM - ROS - Fold 6  
Accuracy: 0.79 | F1 Score: 0.79



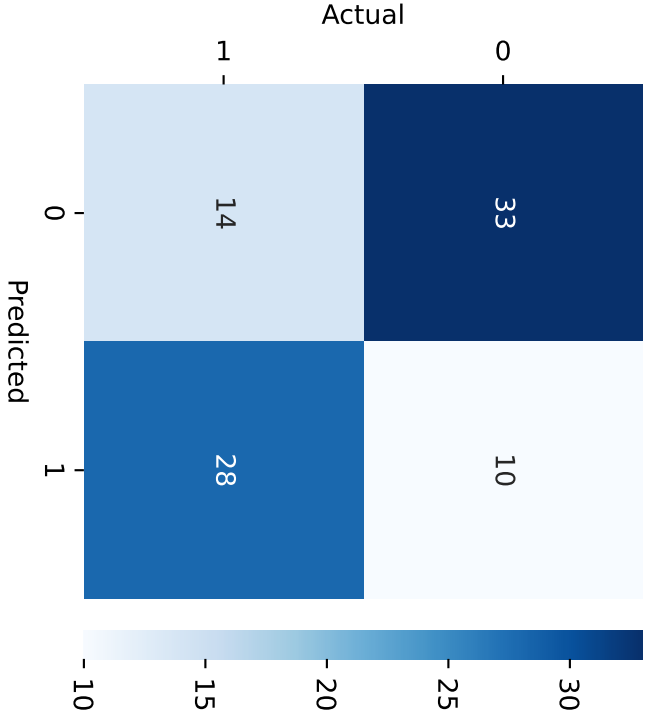
SVM - ROS - Fold 7  
Accuracy: 0.85 | F1 Score: 0.85



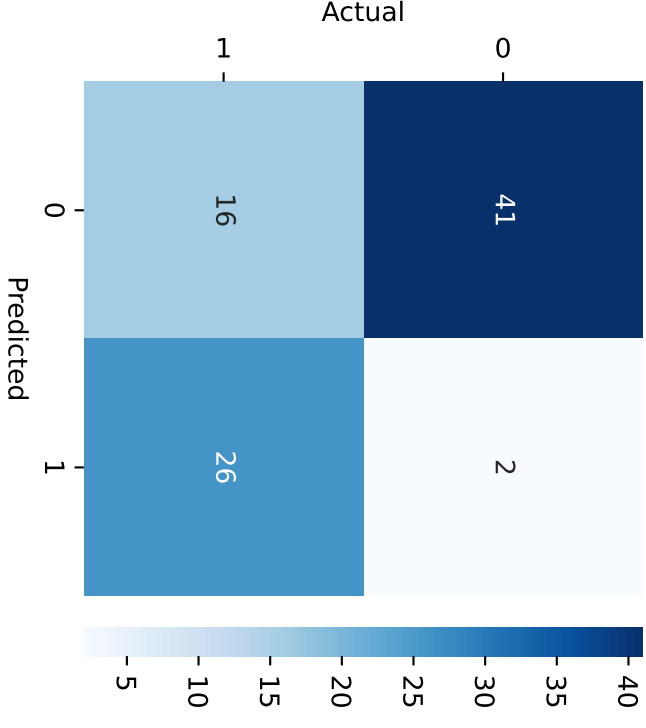
SVM - ROS - Fold 8  
Accuracy: 0.81 | F1 Score: 0.81

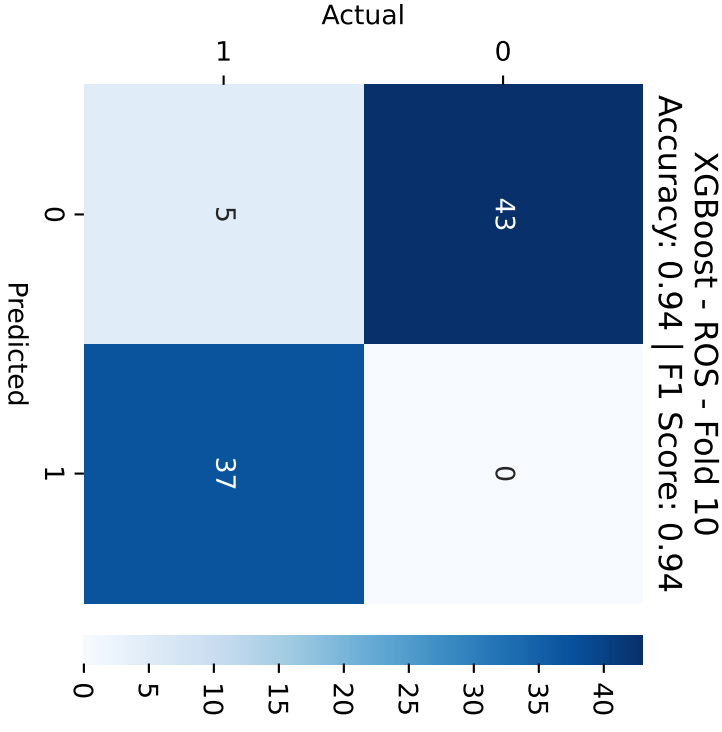
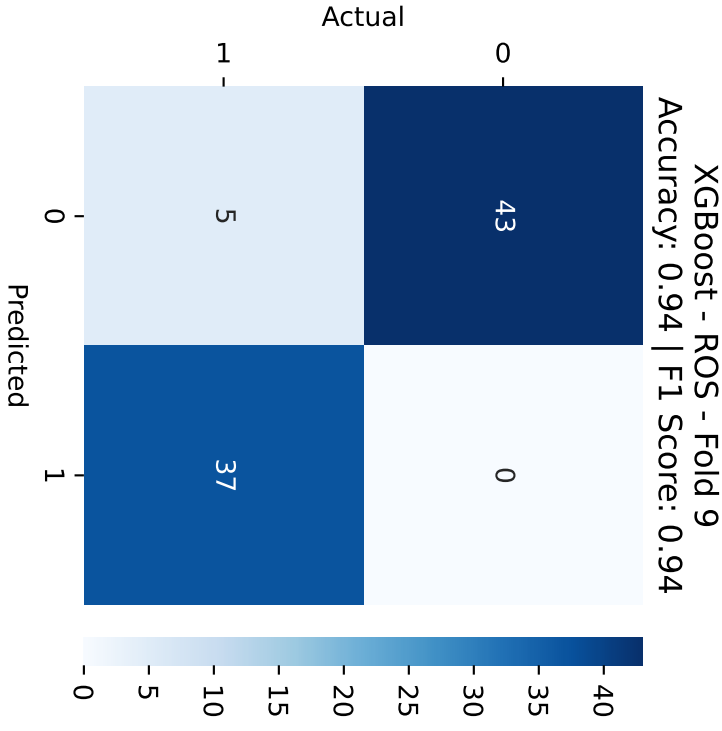
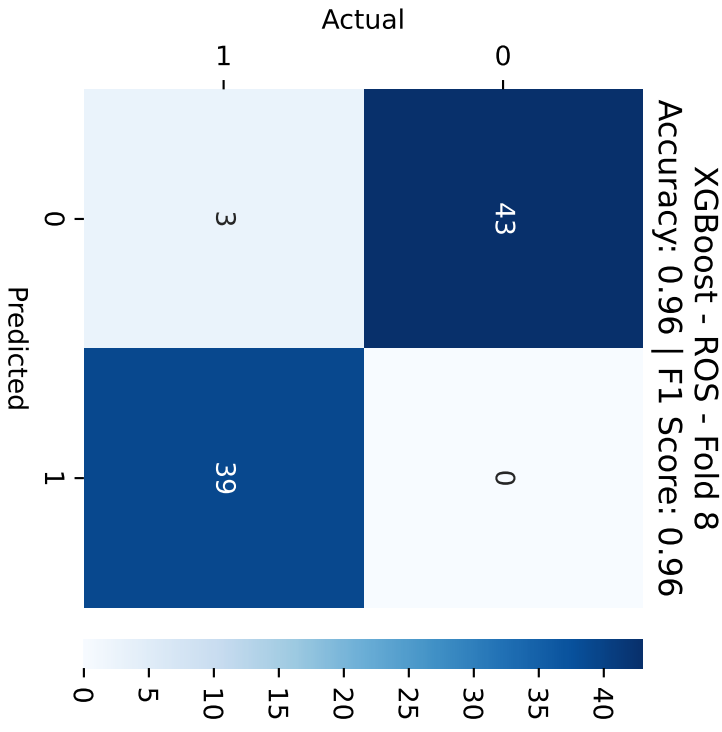
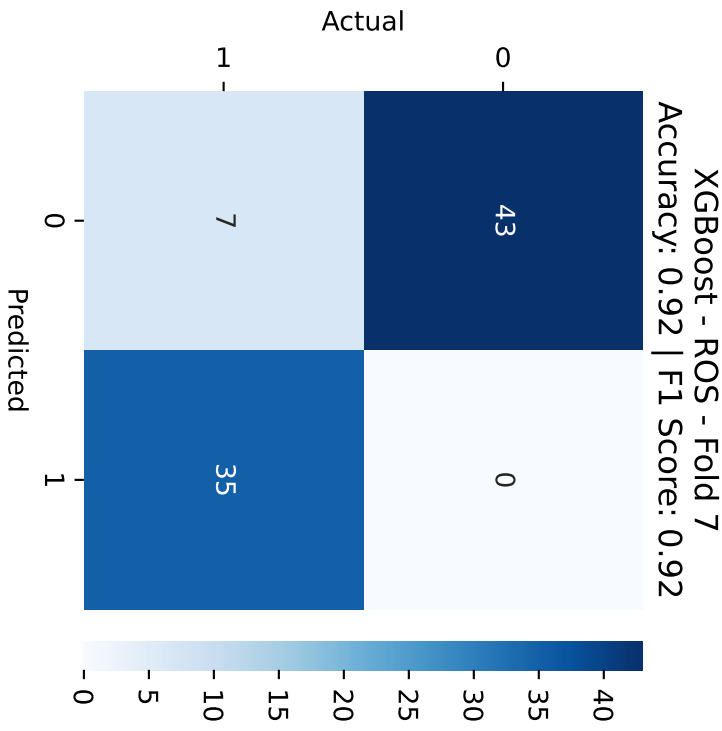
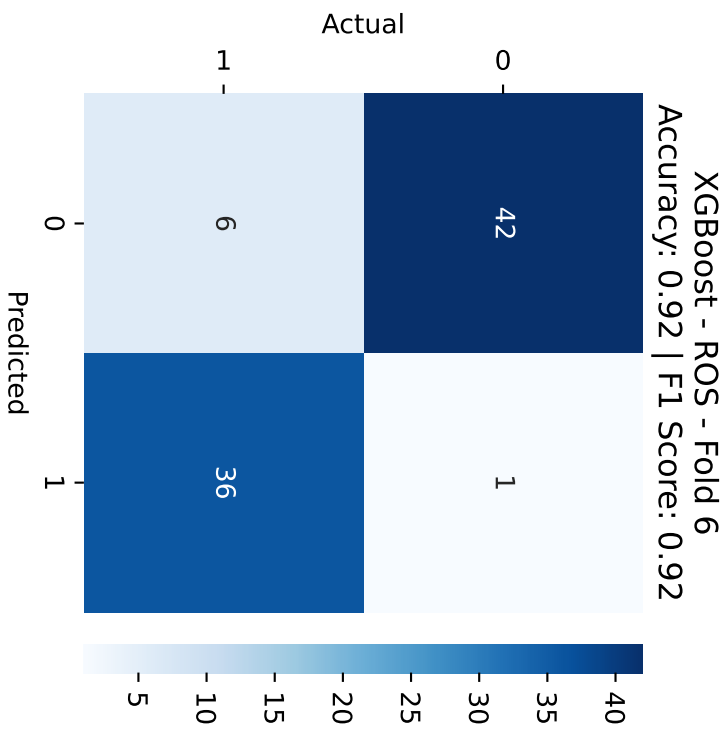
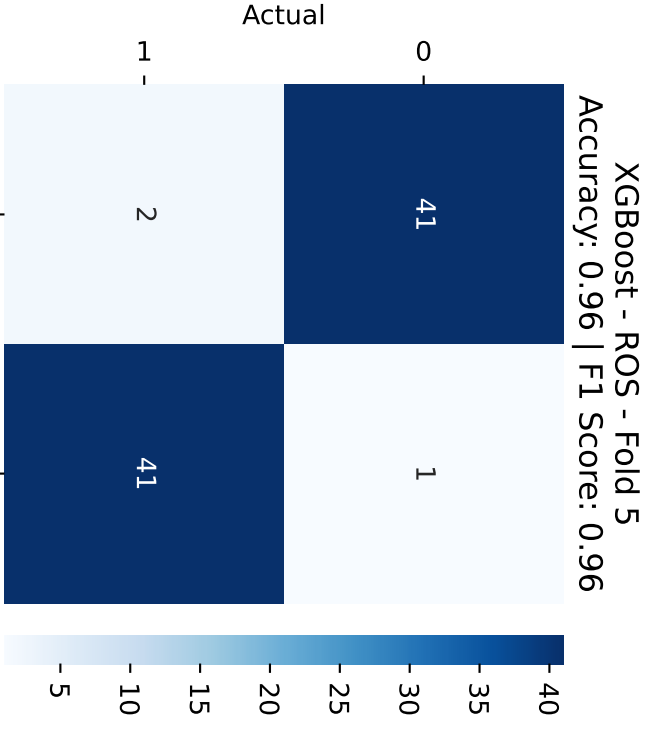
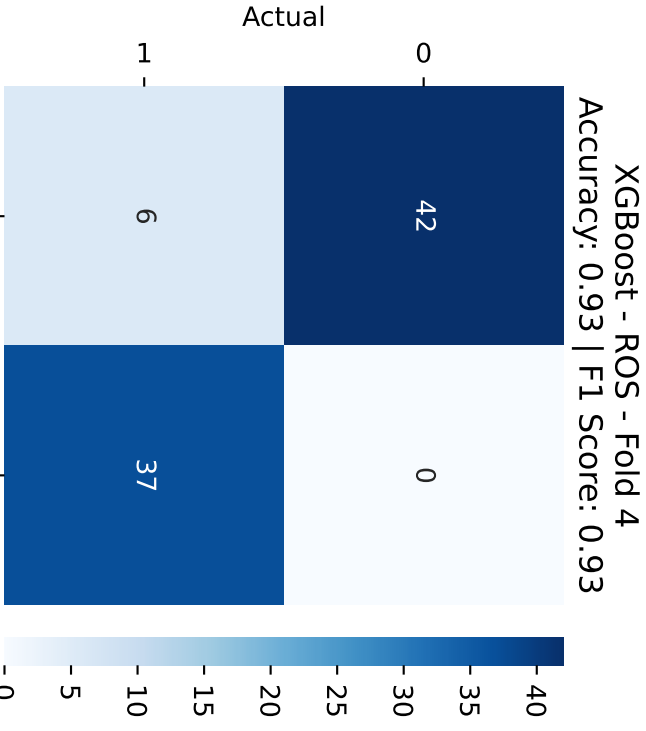
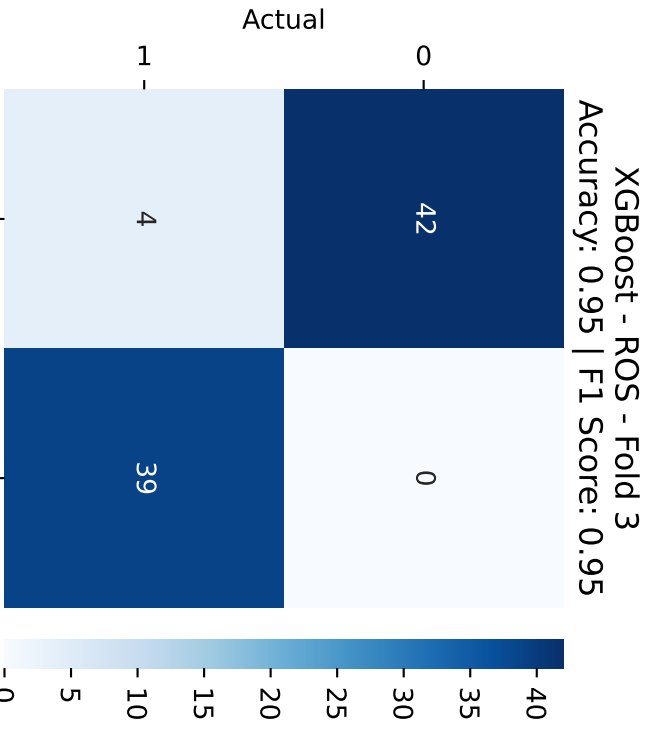
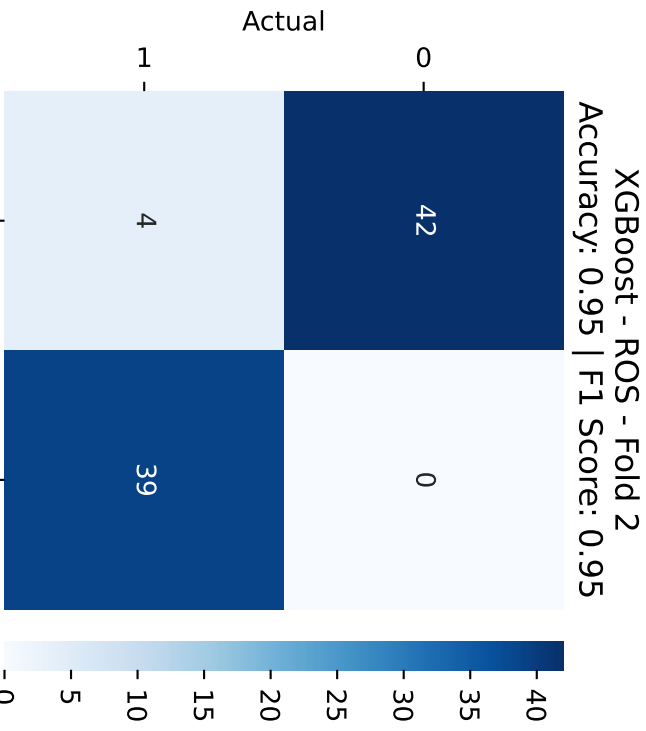
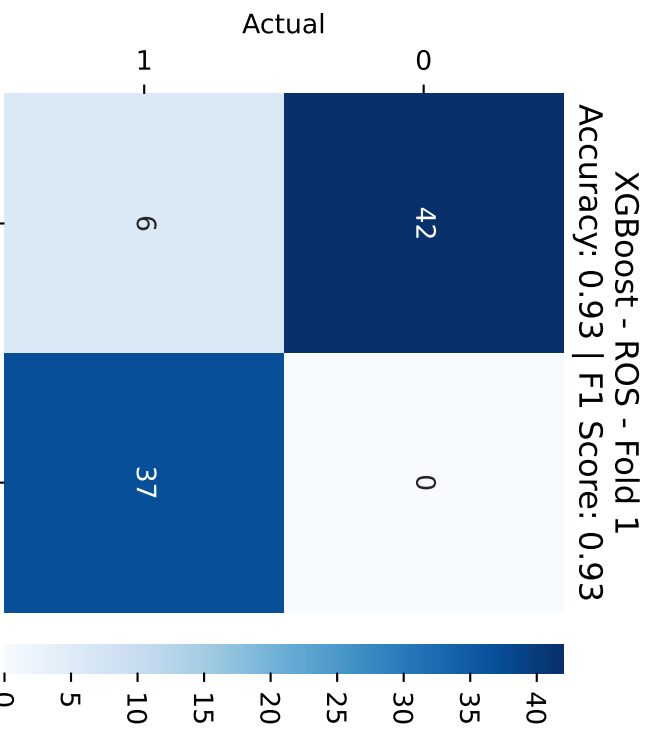


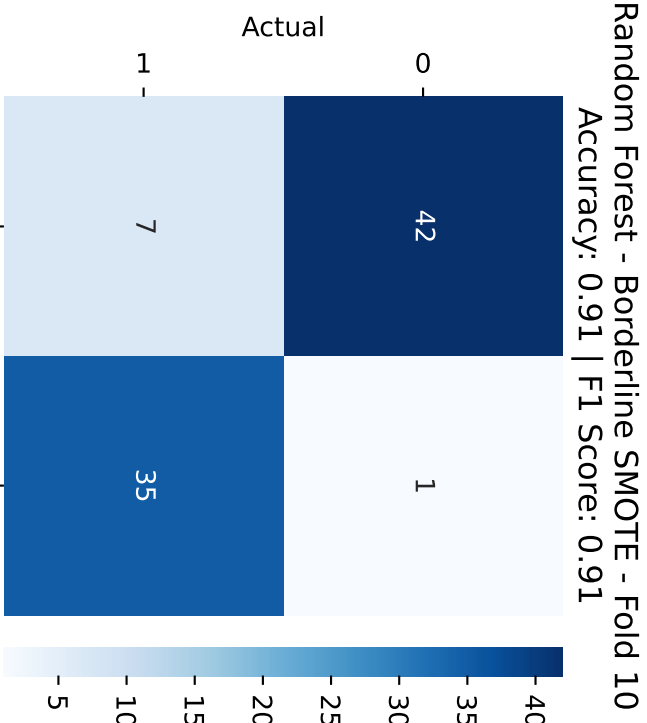
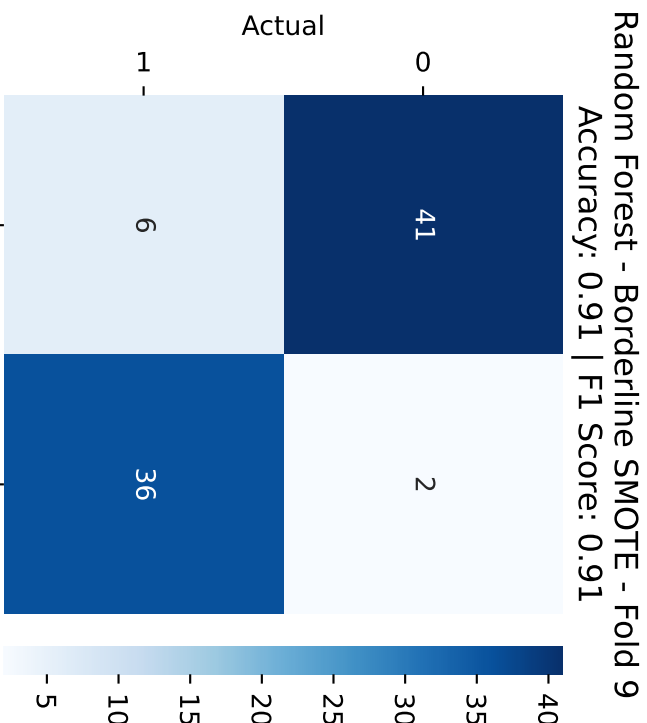
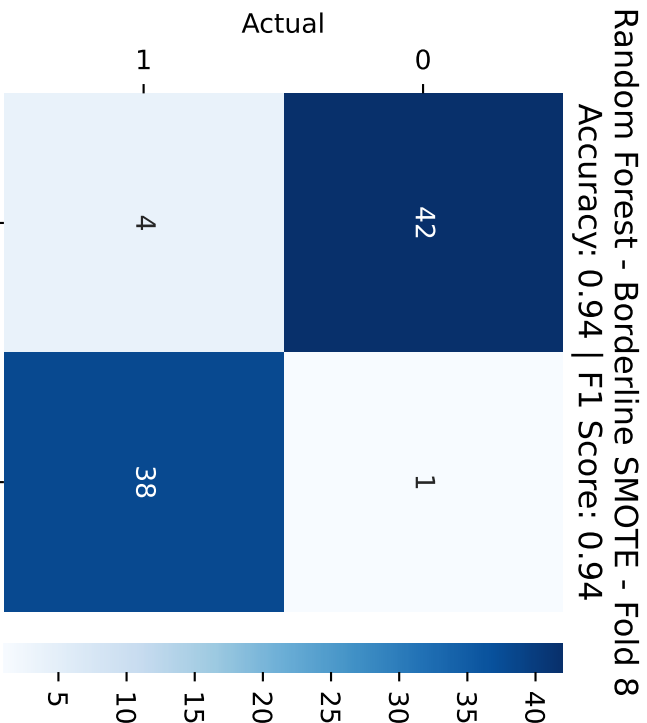
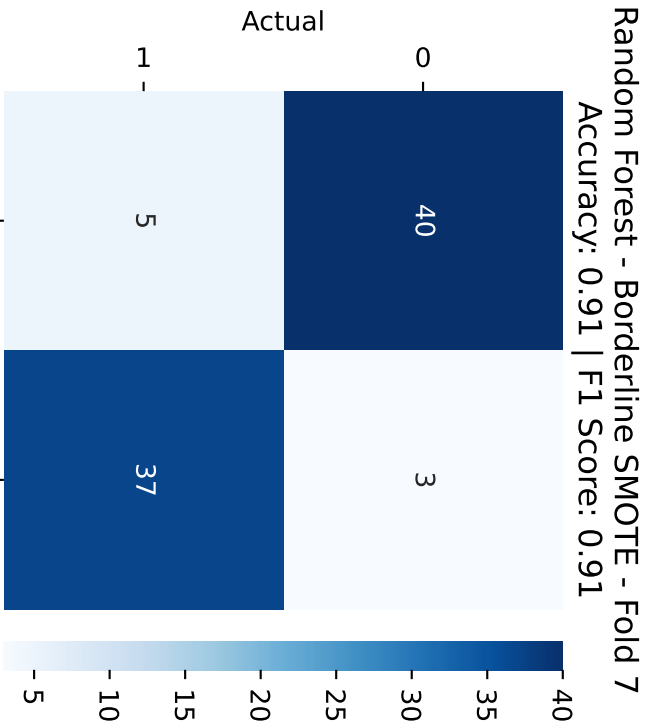
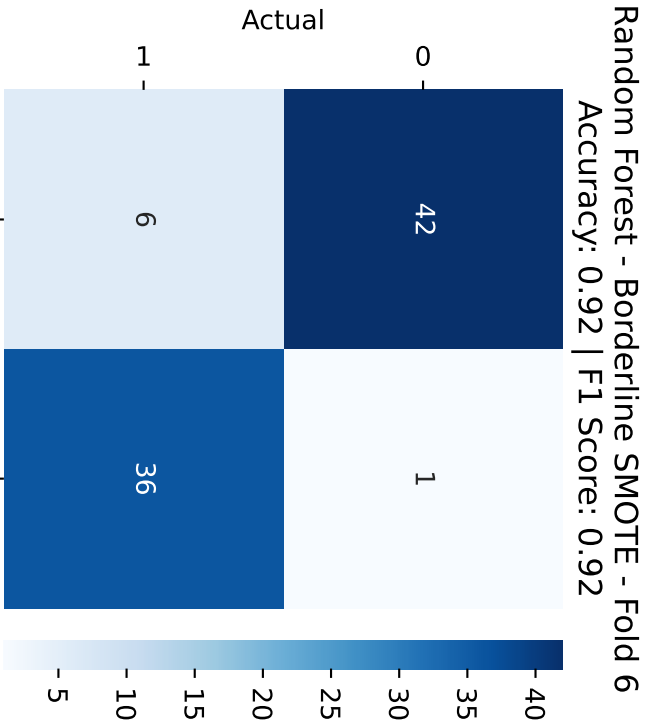
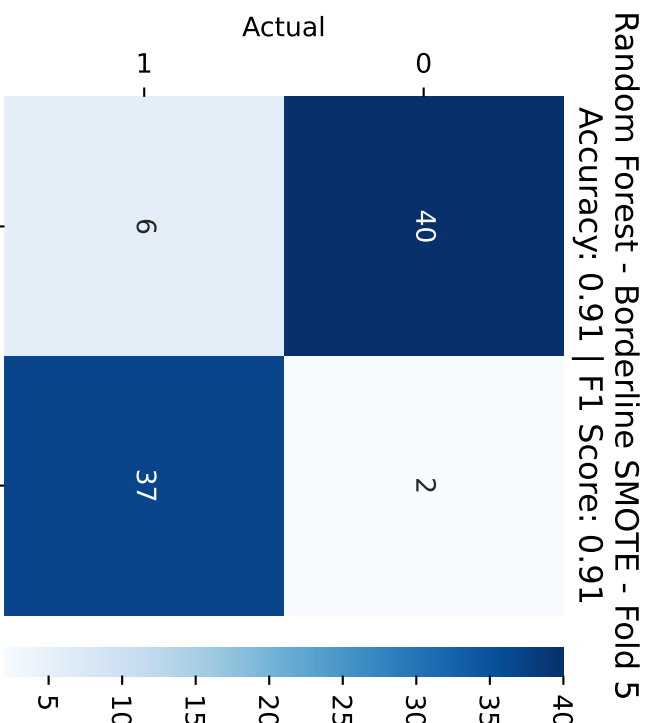
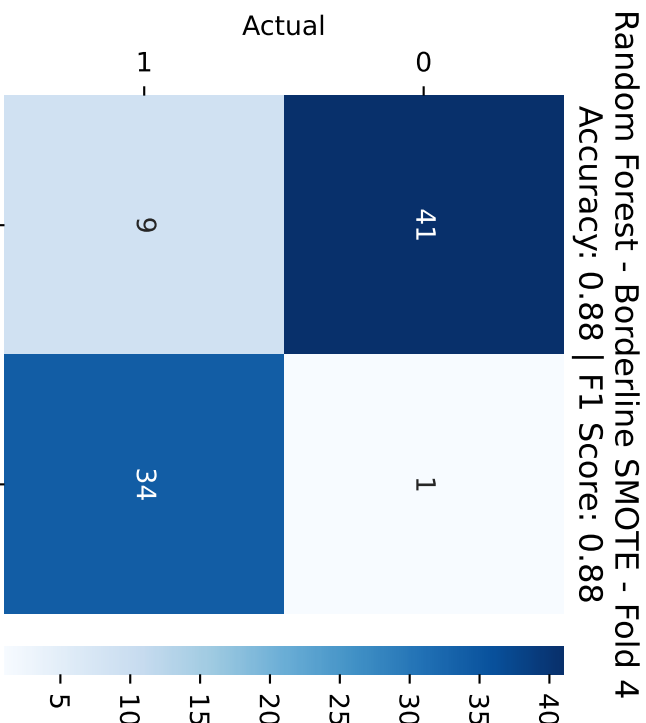
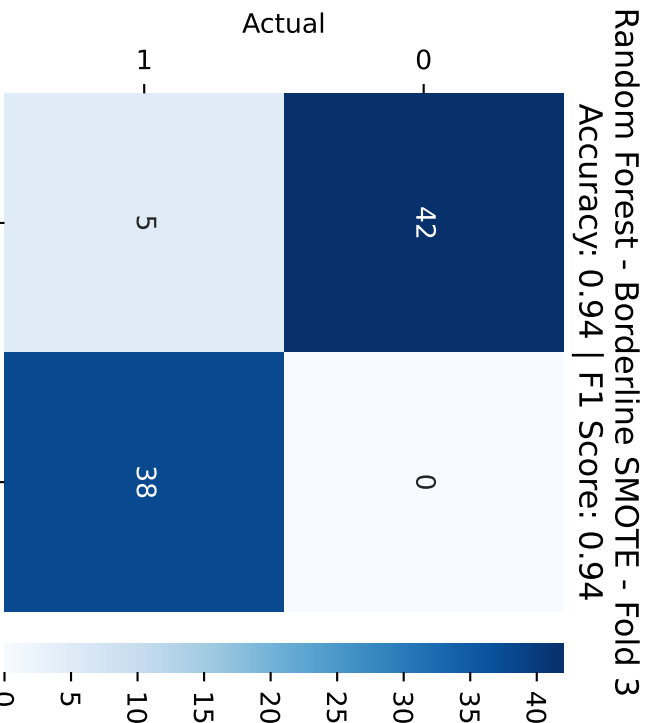
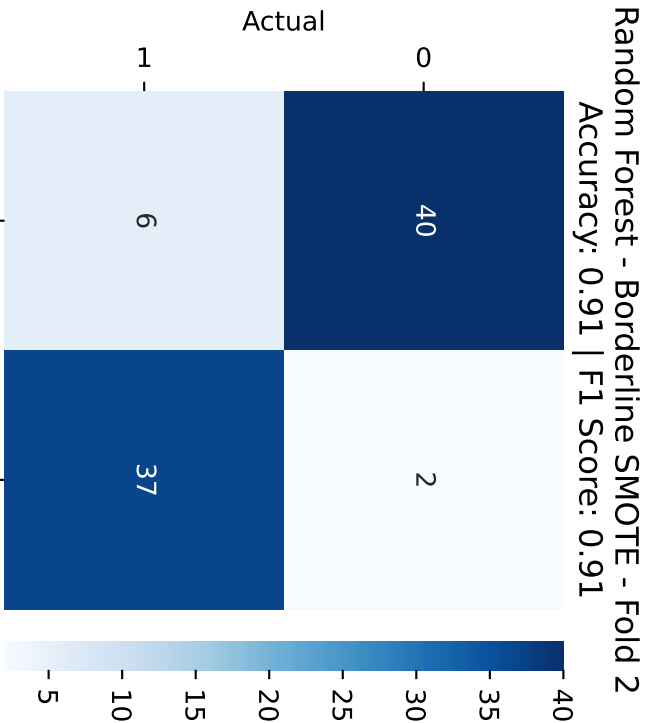
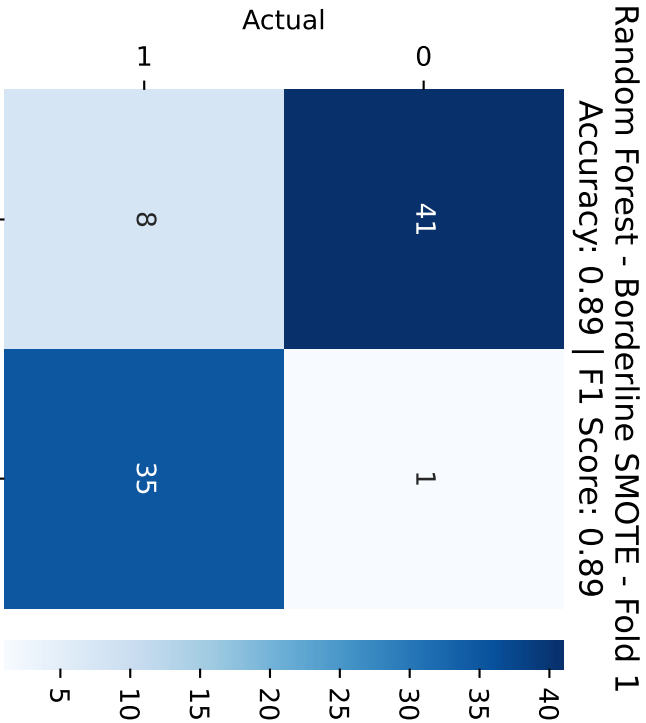
SVM - ROS - Fold 9  
Accuracy: 0.72 | F1 Score: 0.72

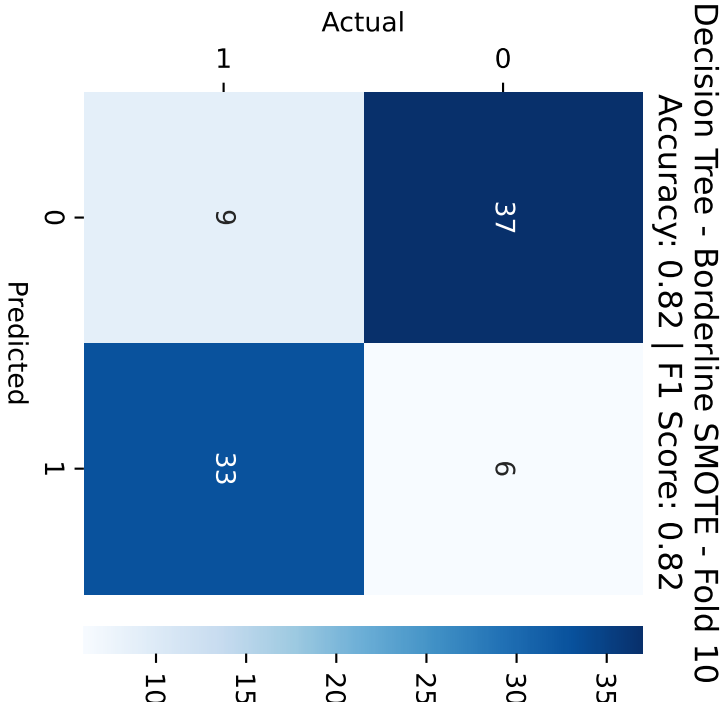
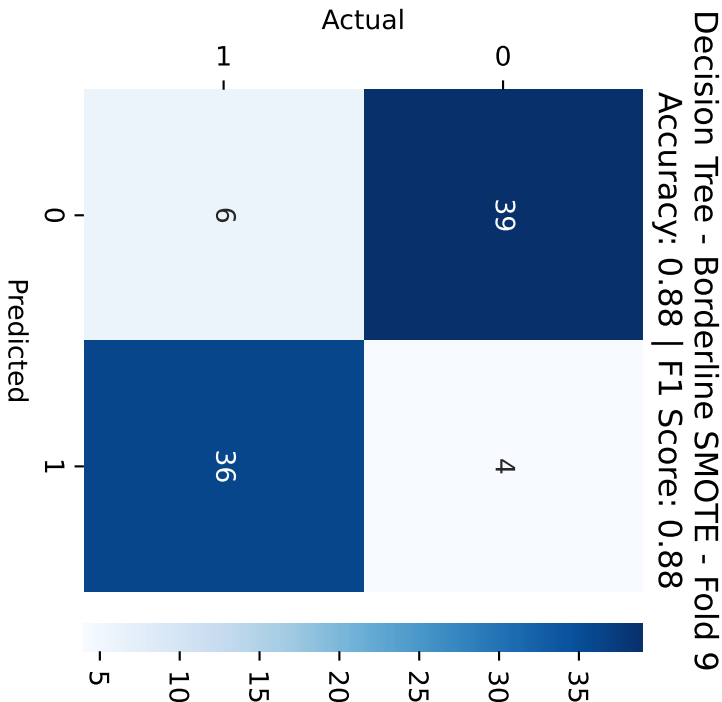
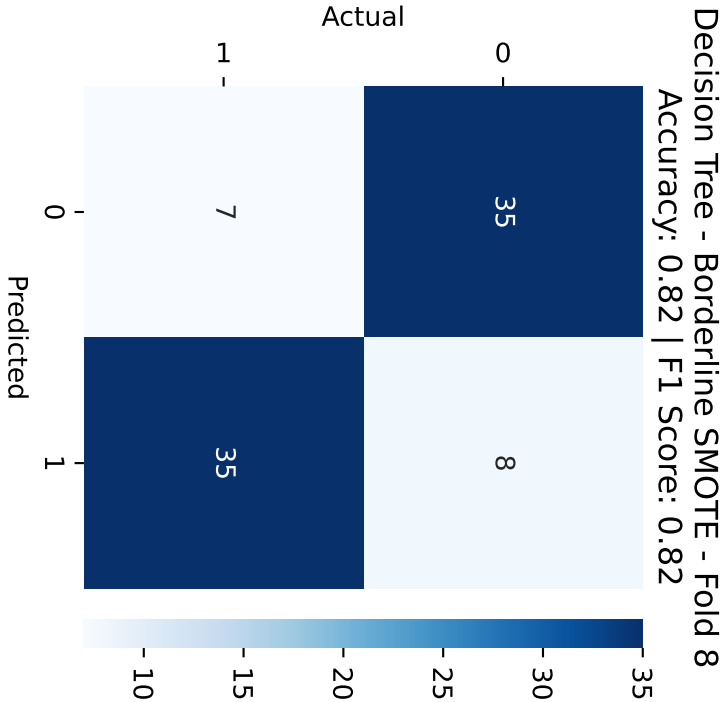
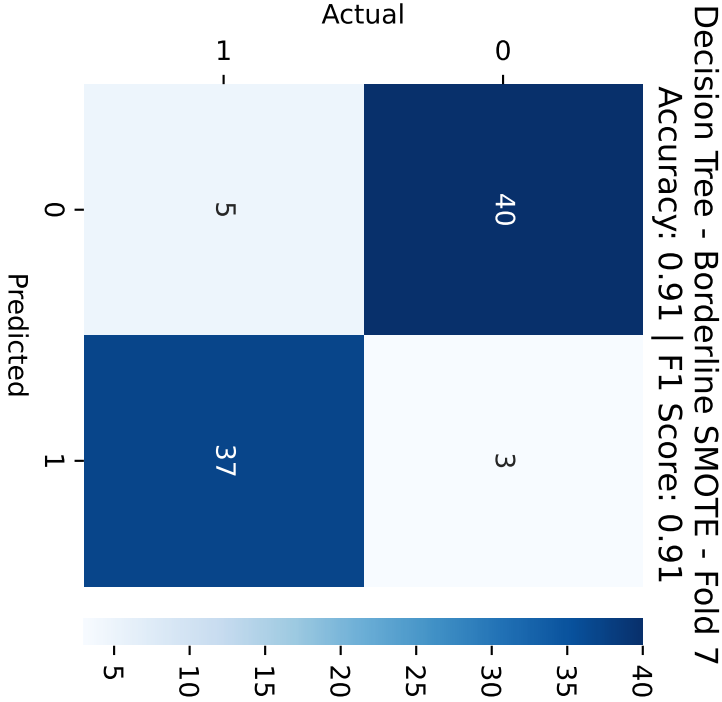
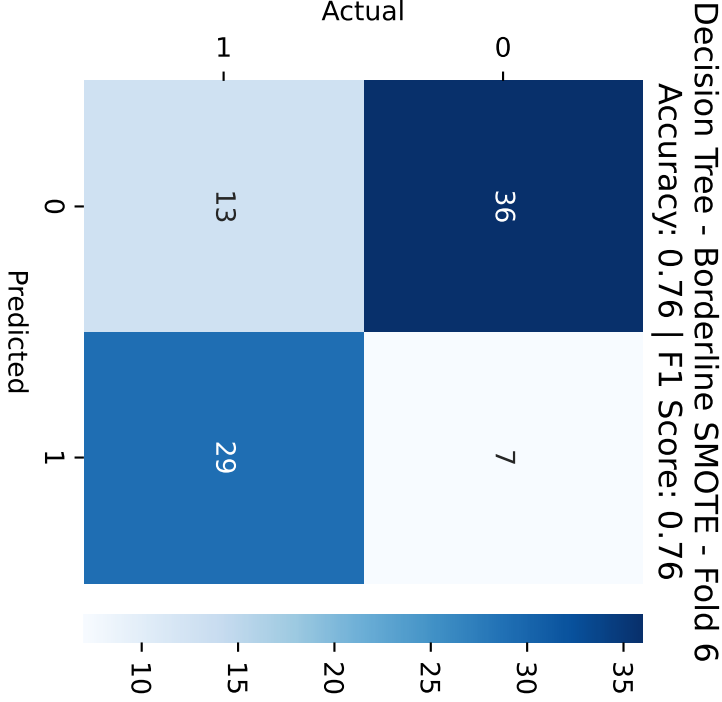
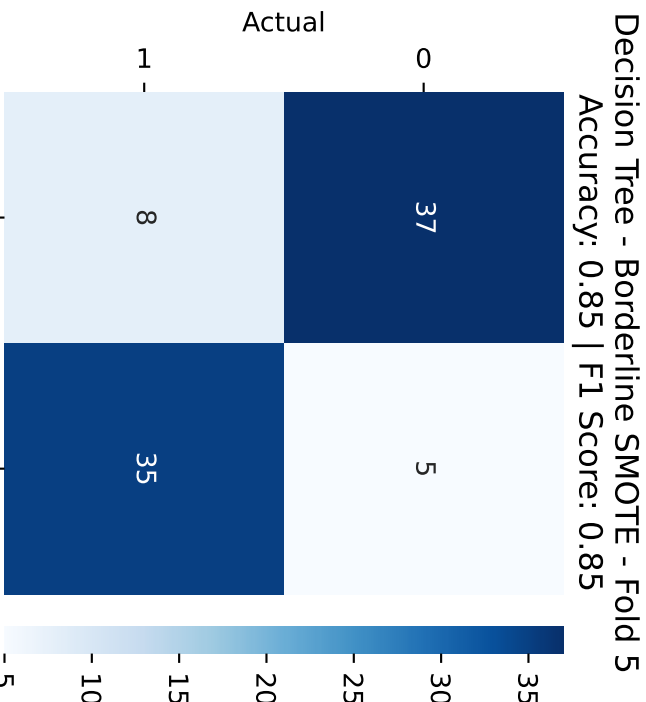
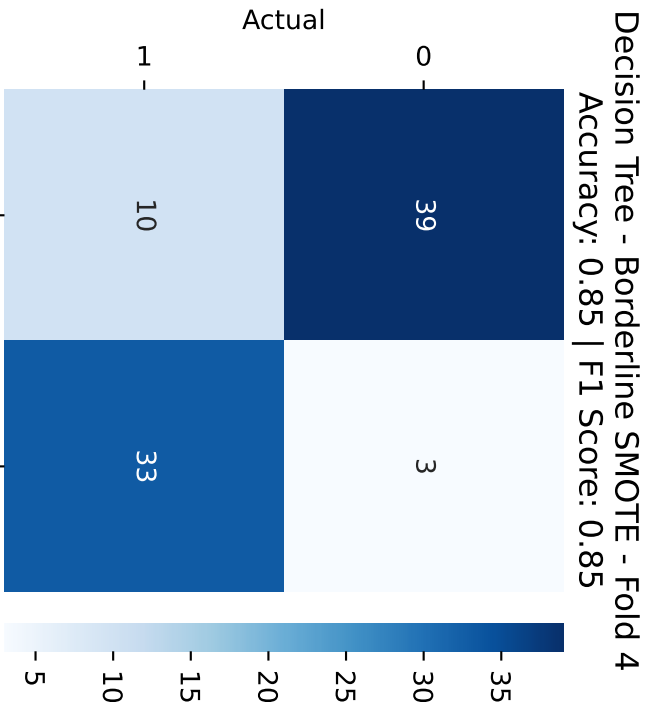
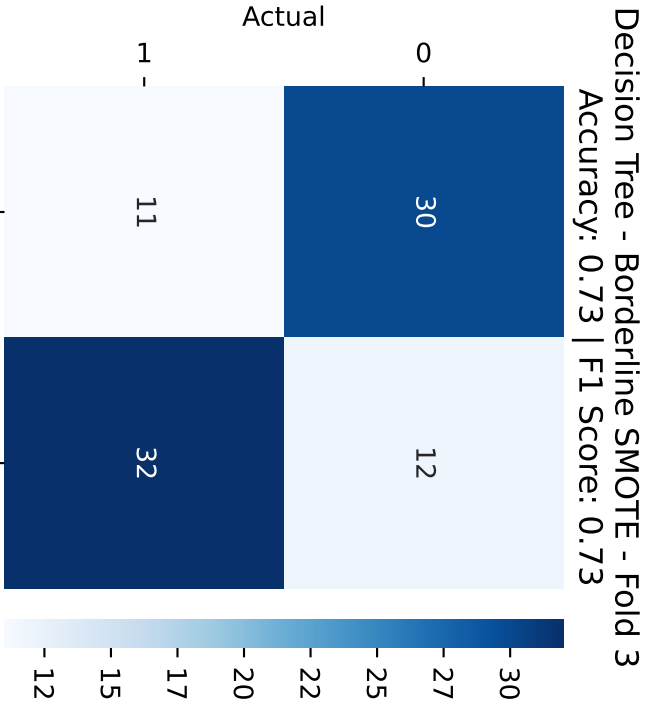
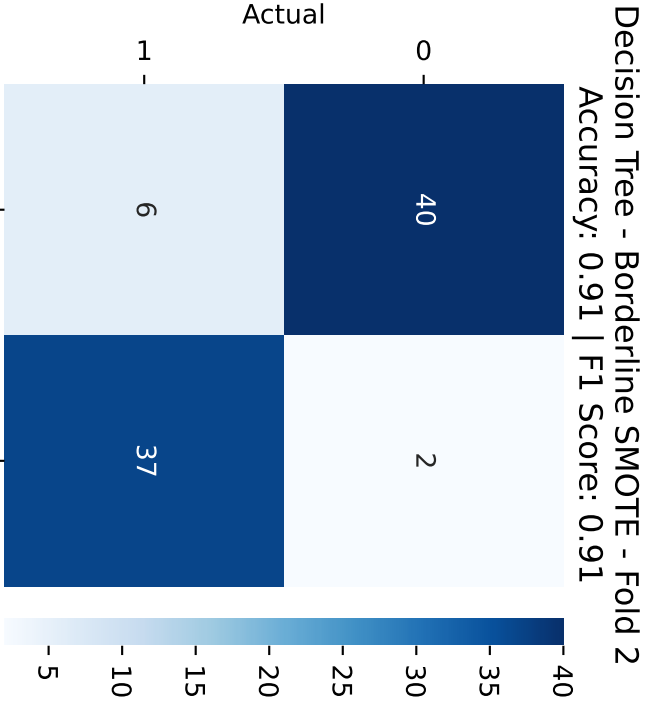
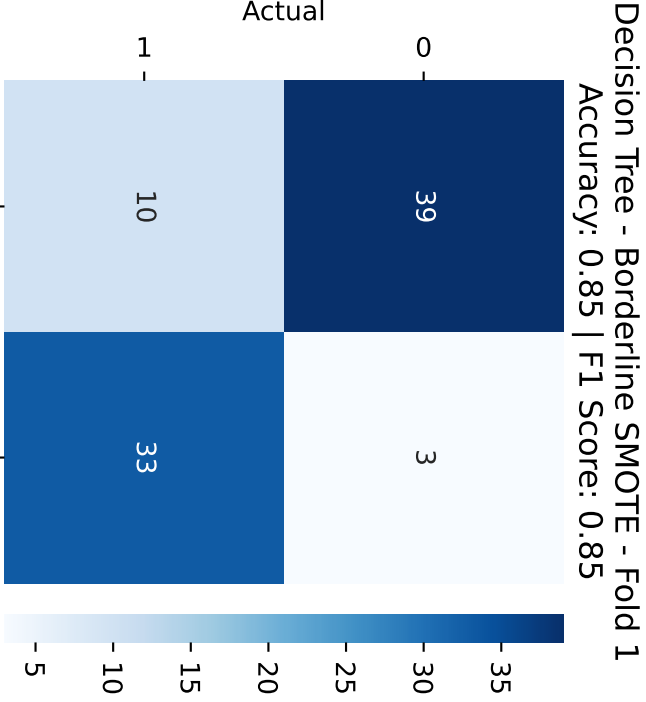


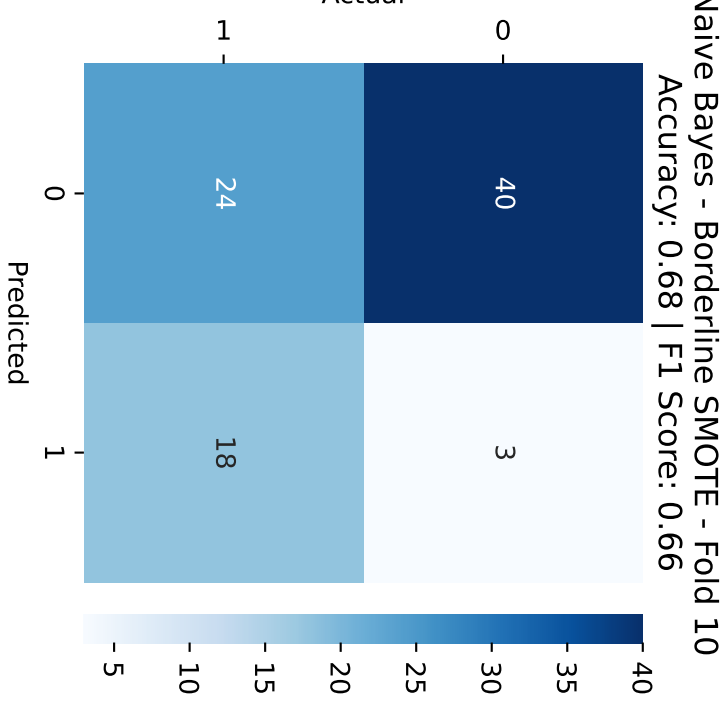
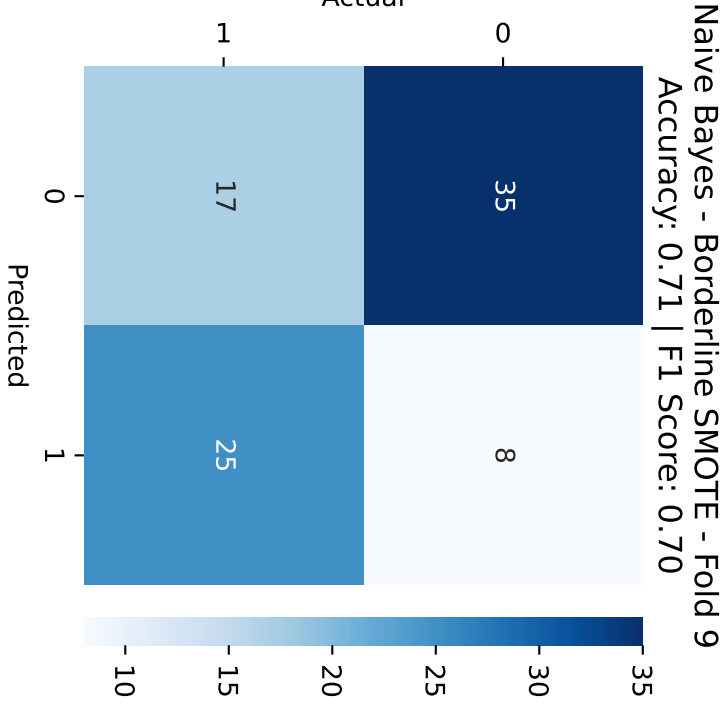
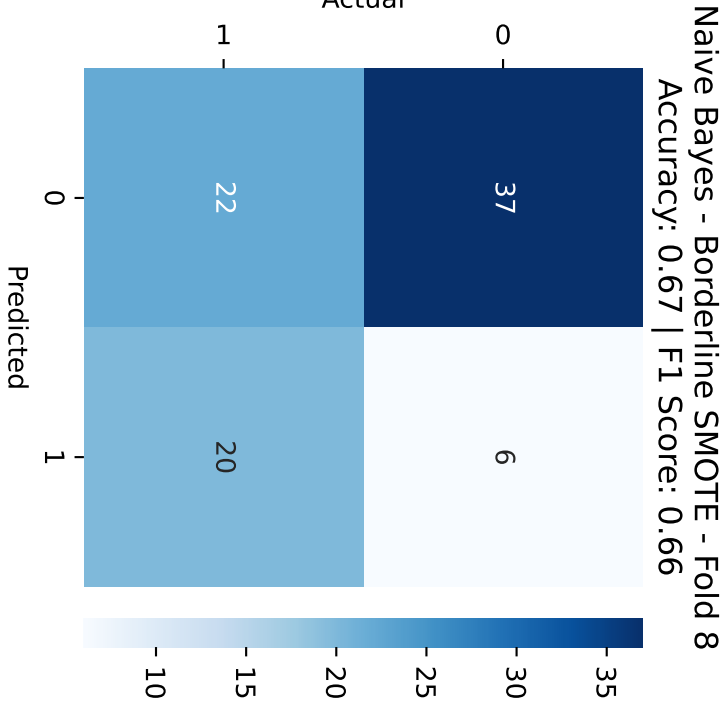
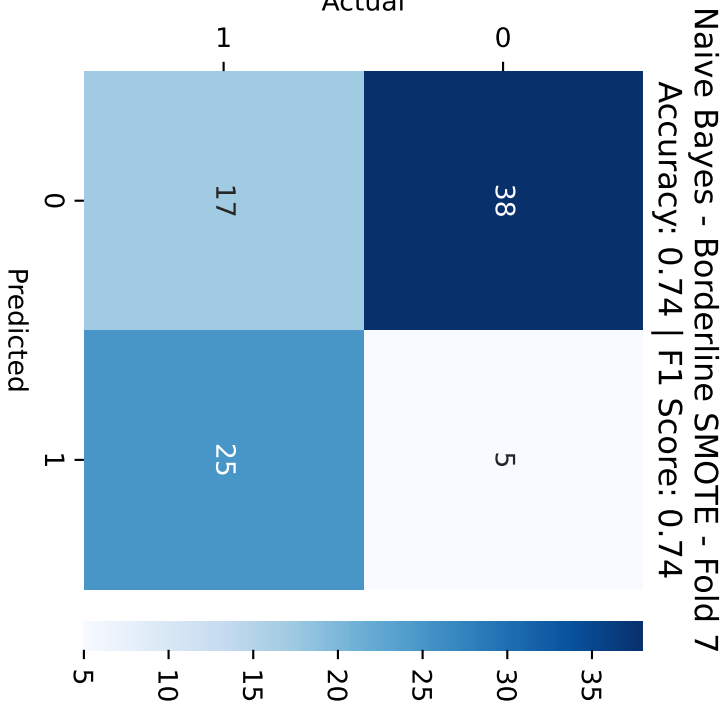
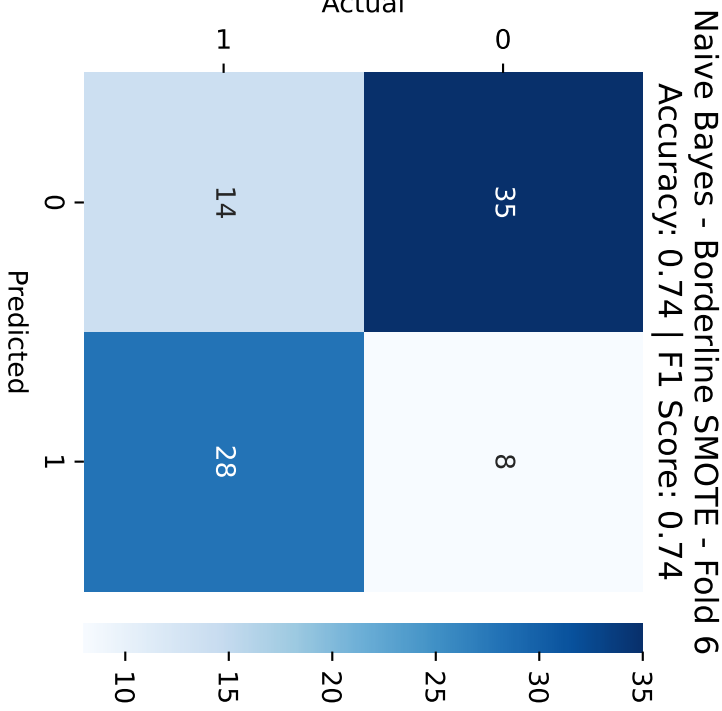
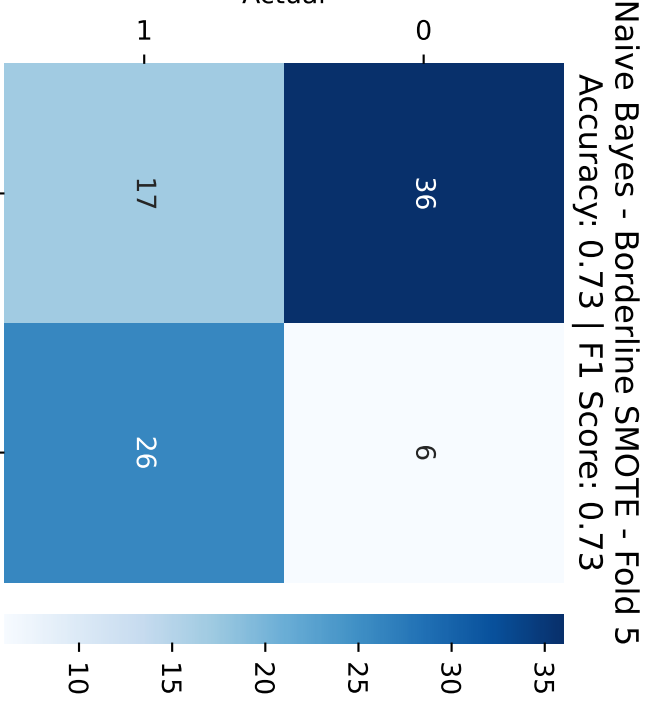
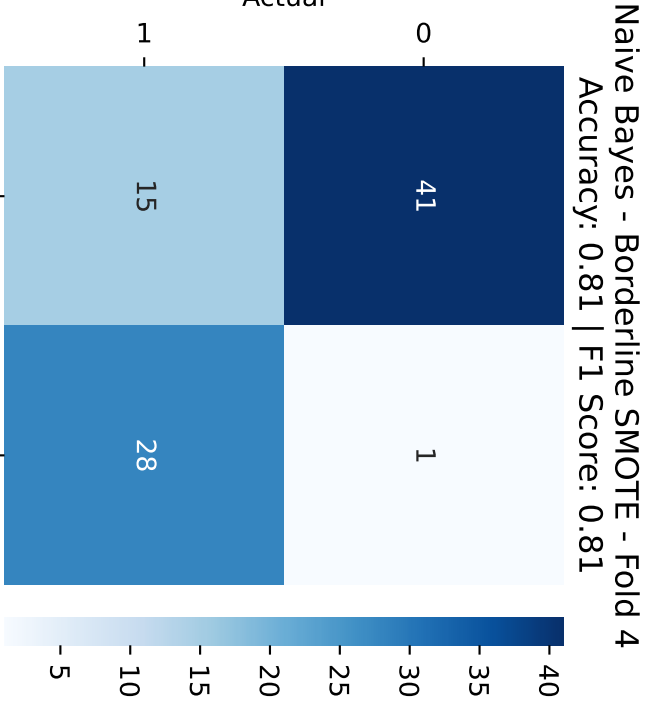
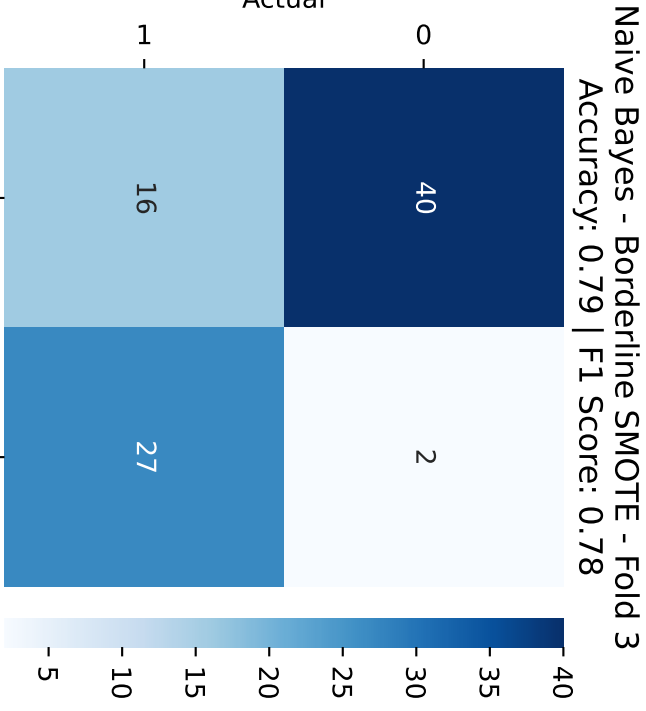
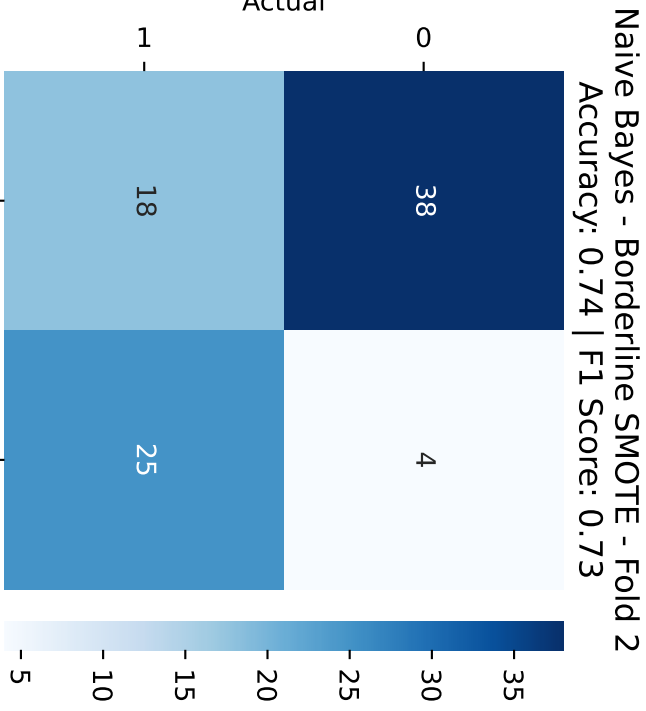
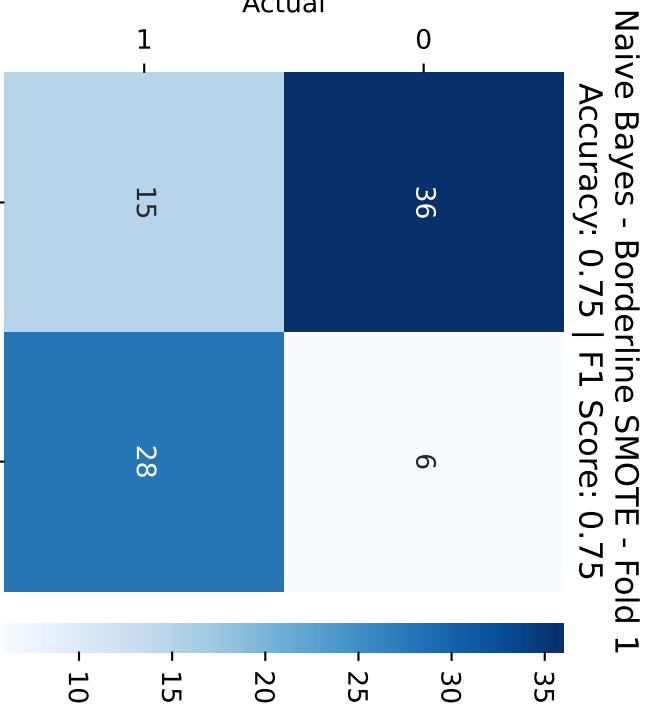
SVM - ROS - Fold 10  
Accuracy: 0.79 | F1 Score: 0.78

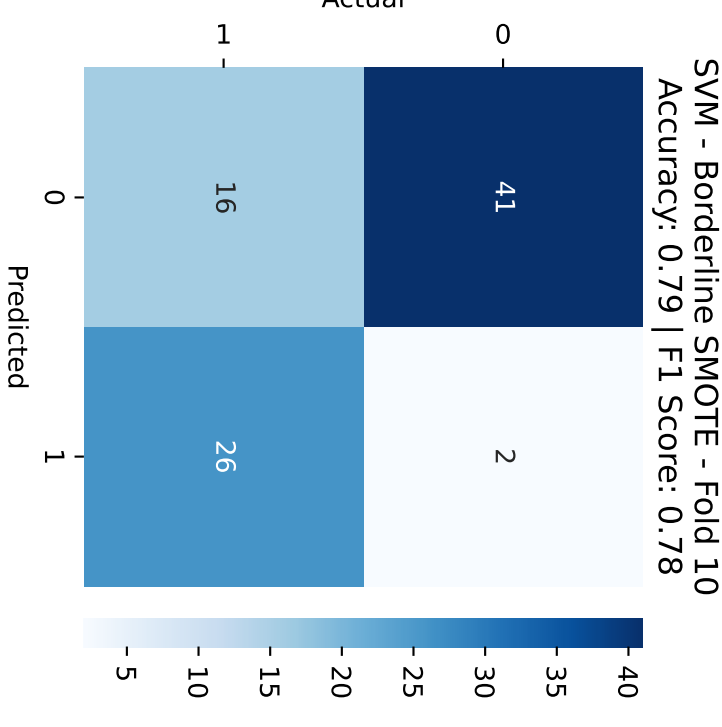
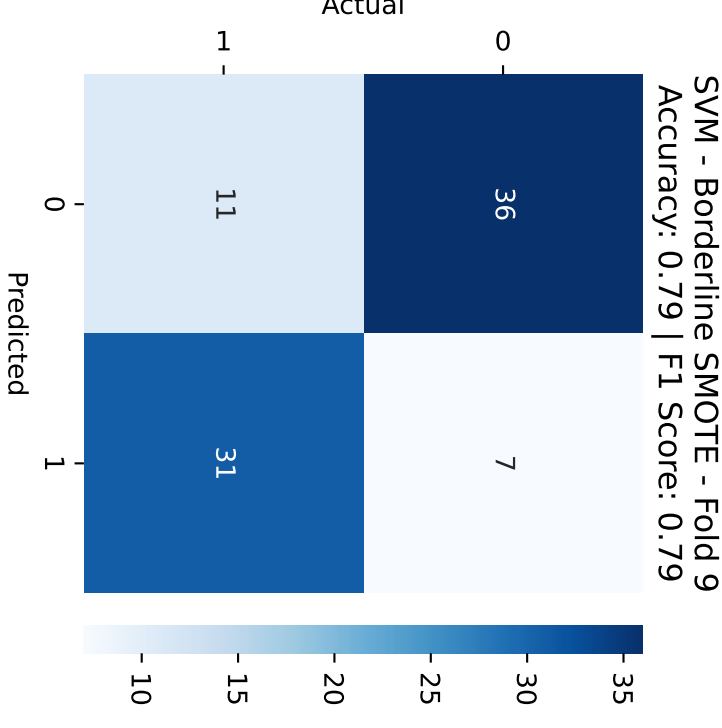
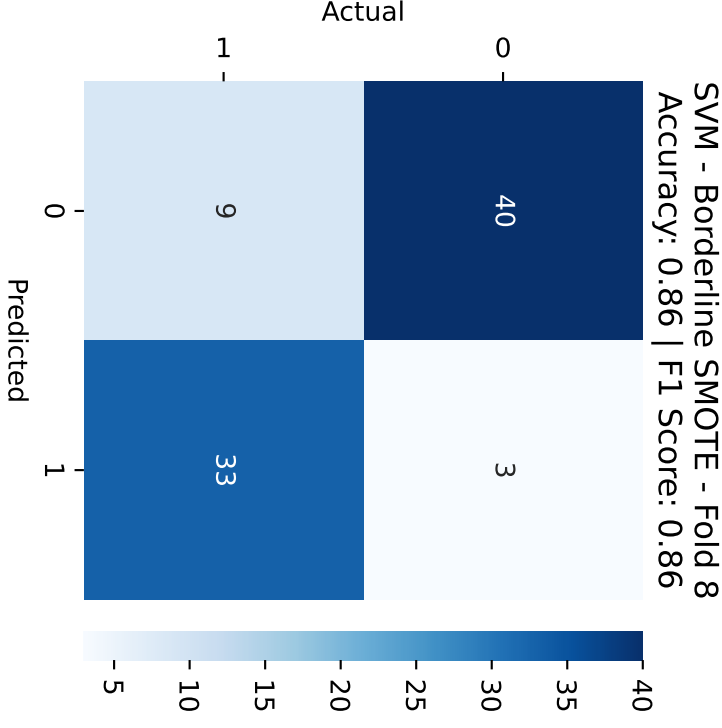
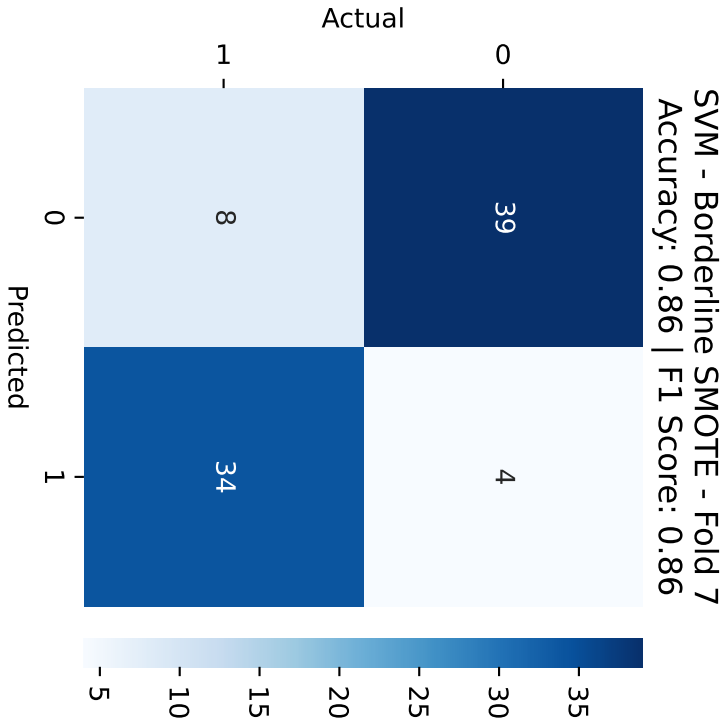
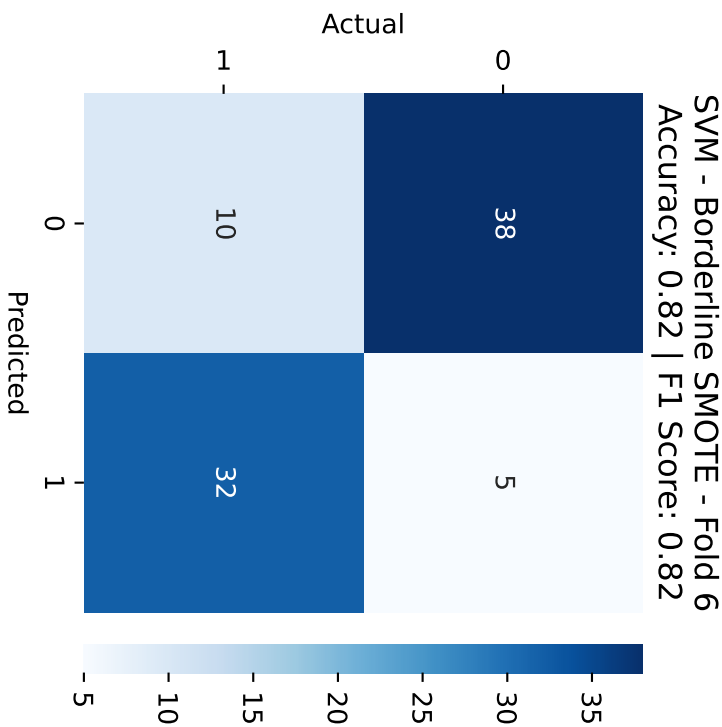
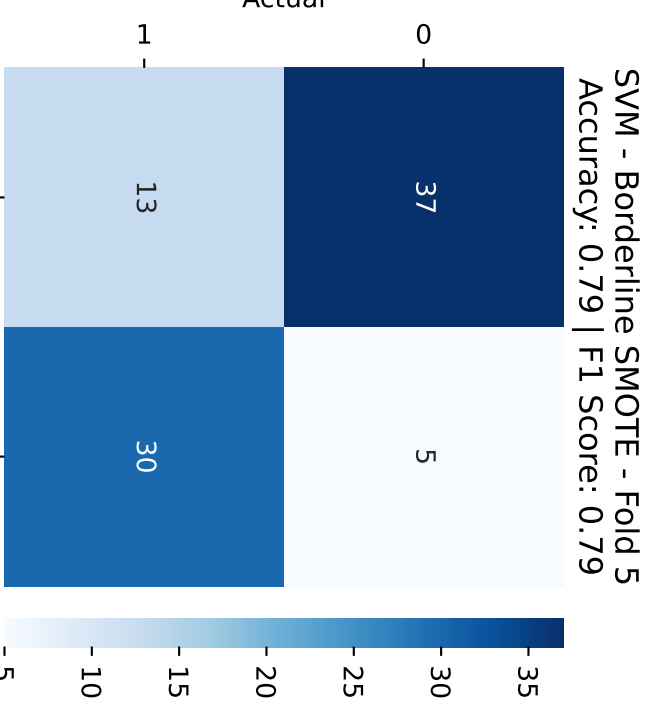
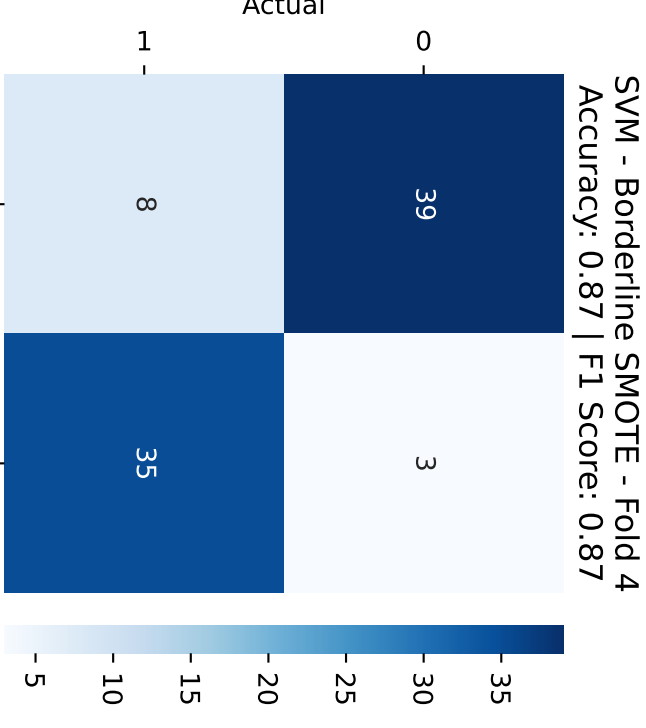
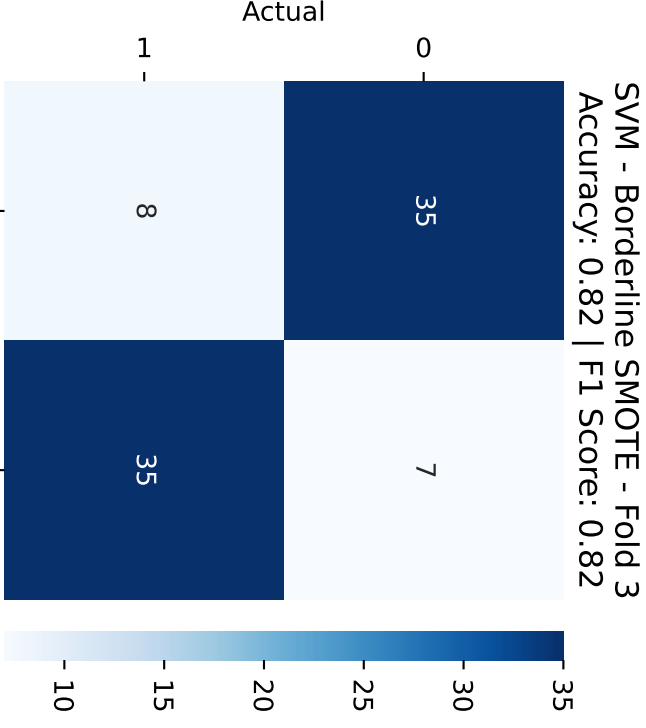
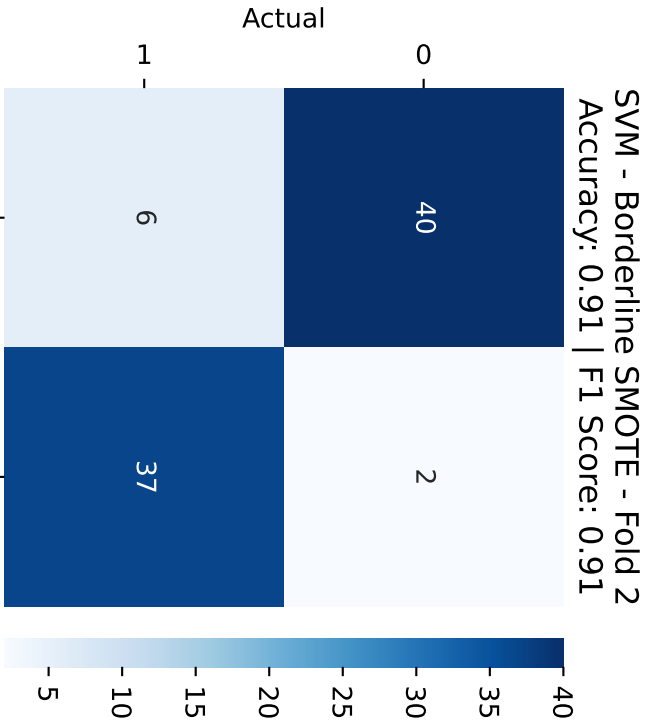
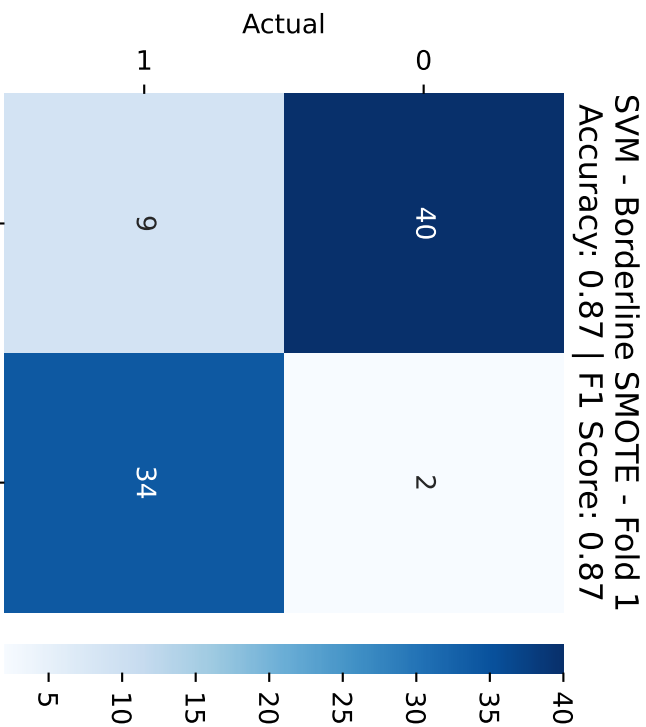


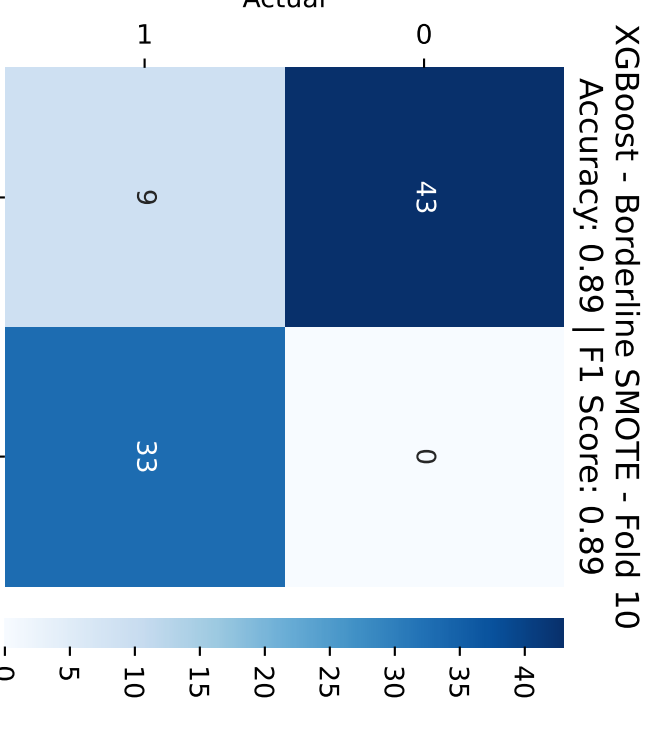
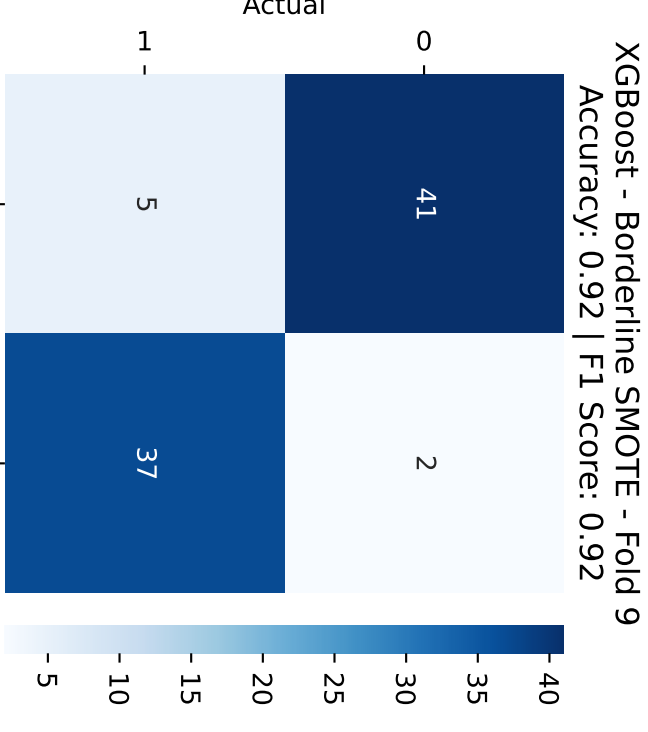
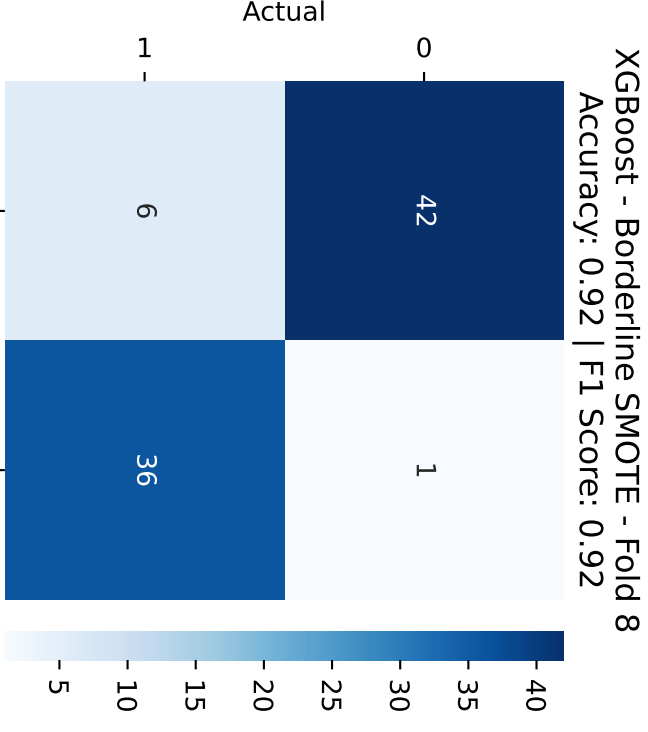
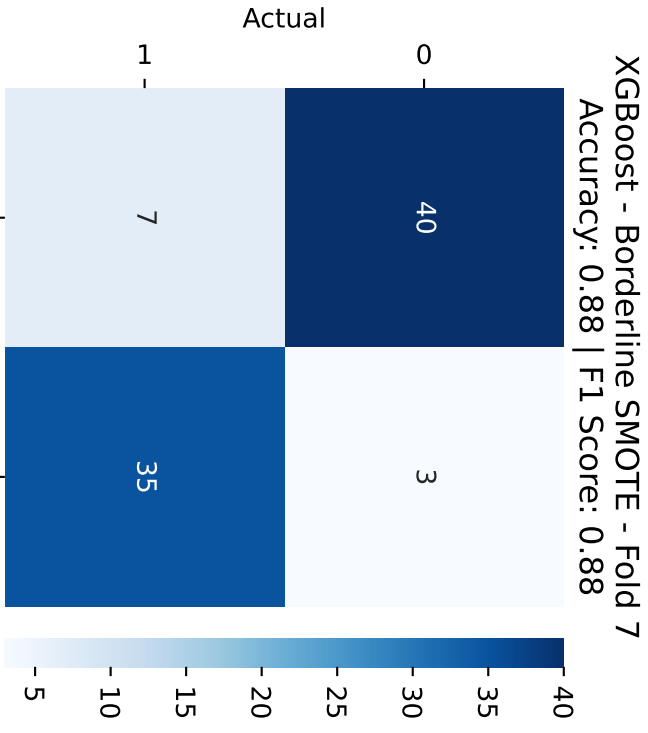
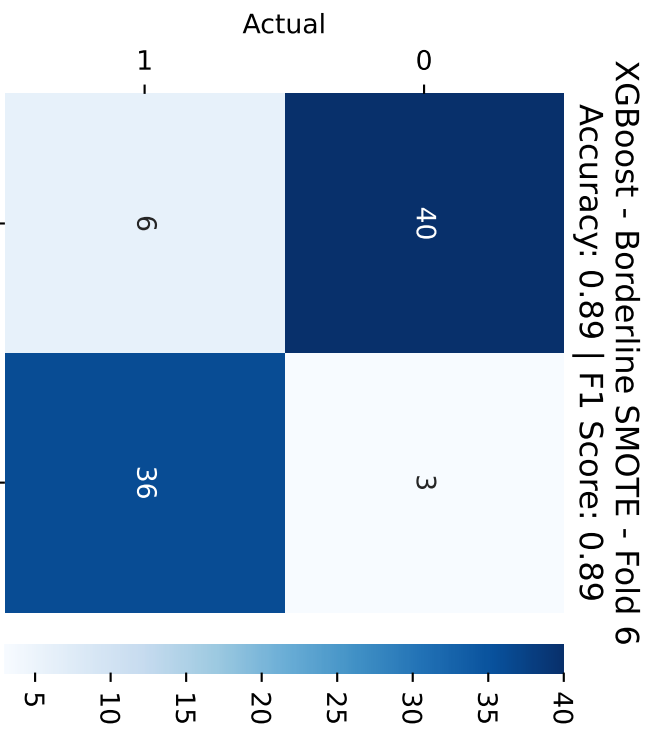
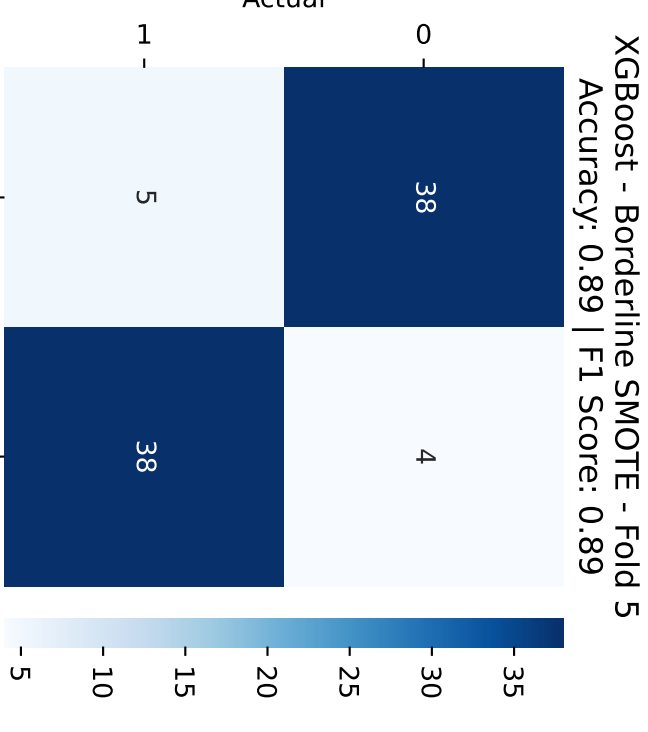
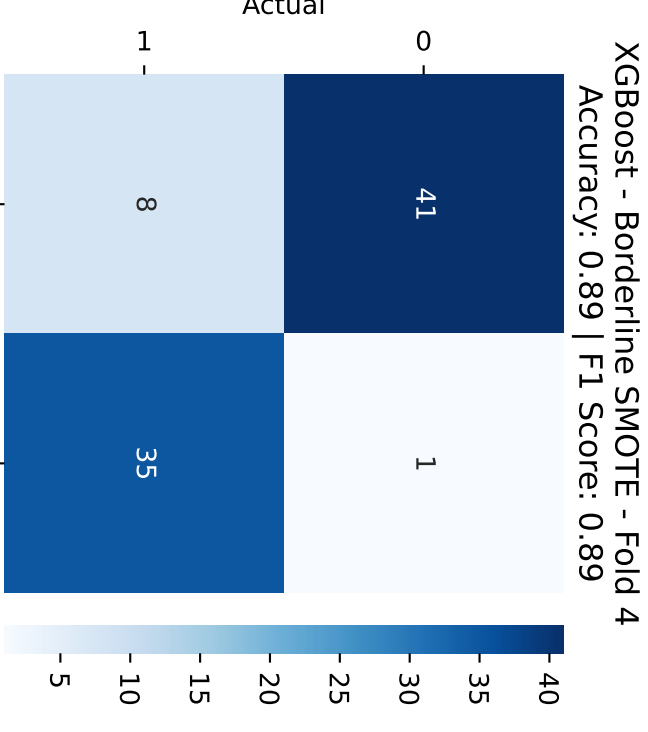
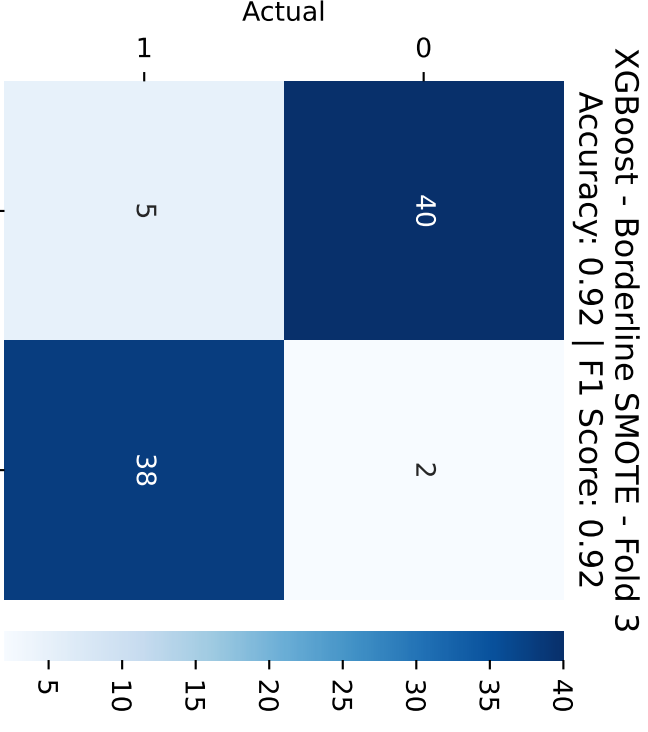
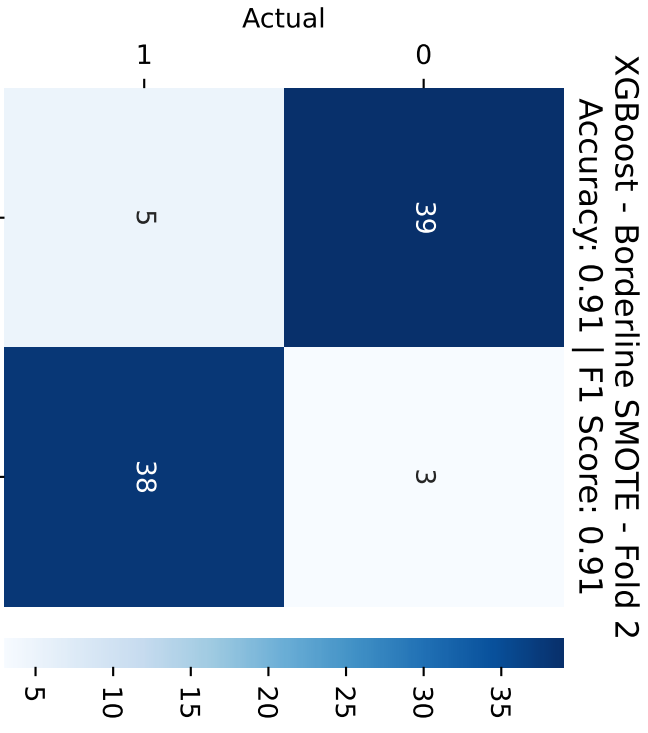
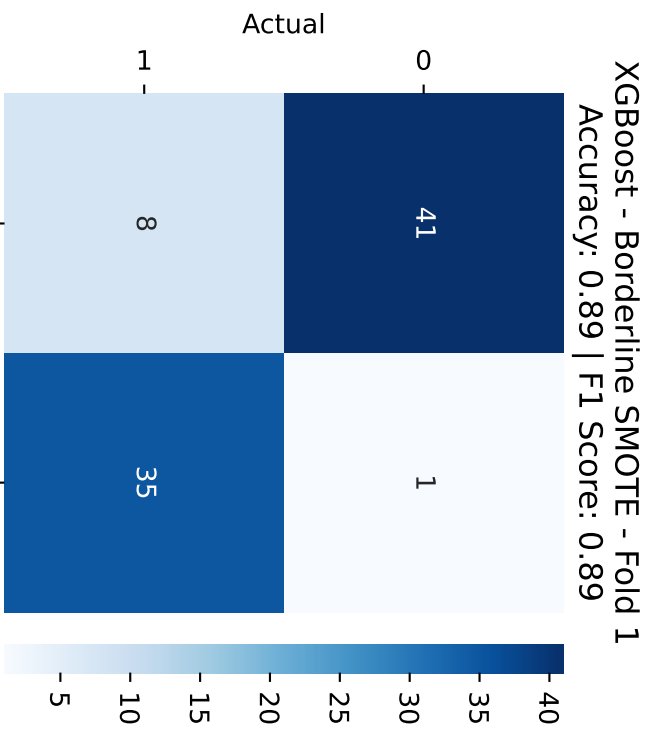


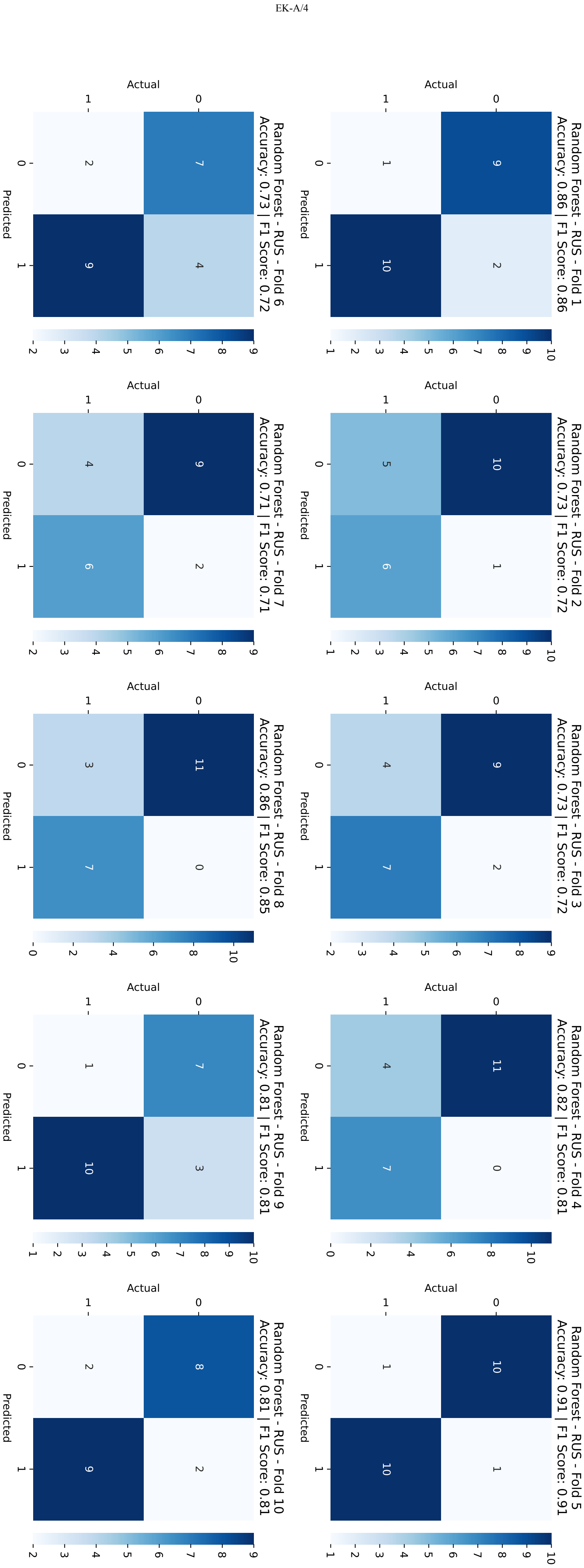


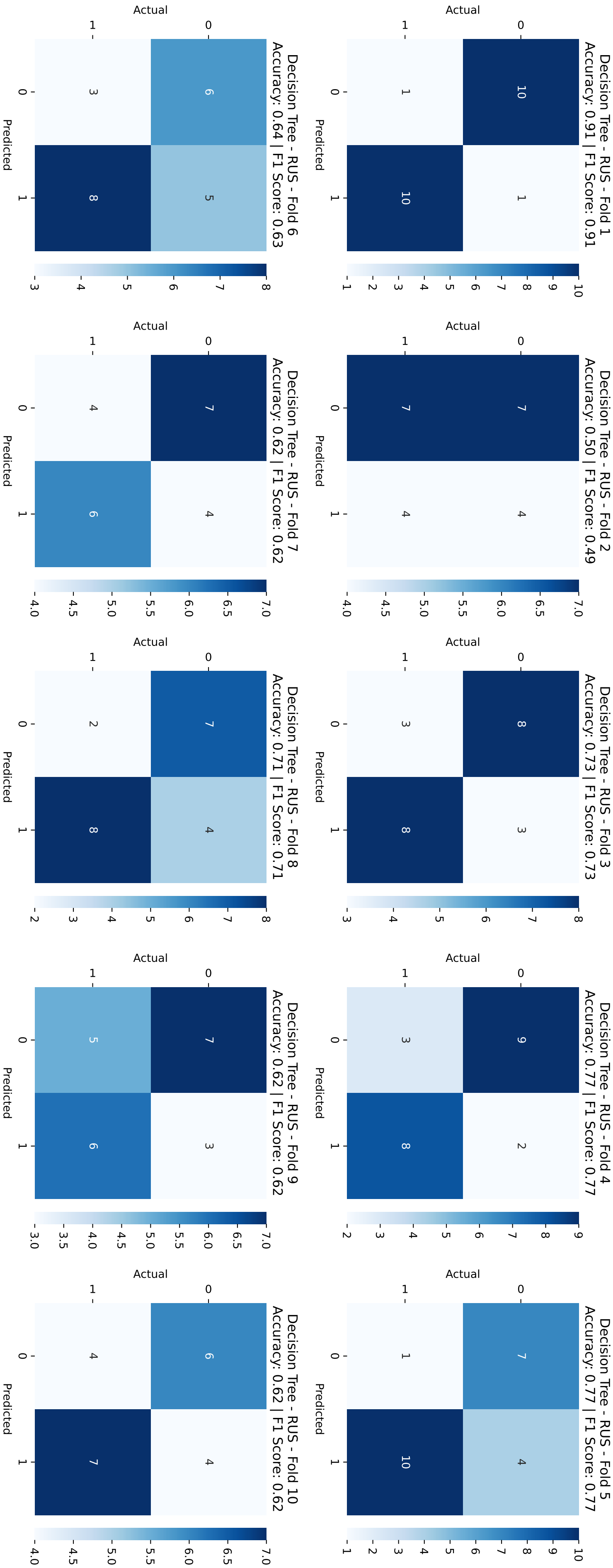


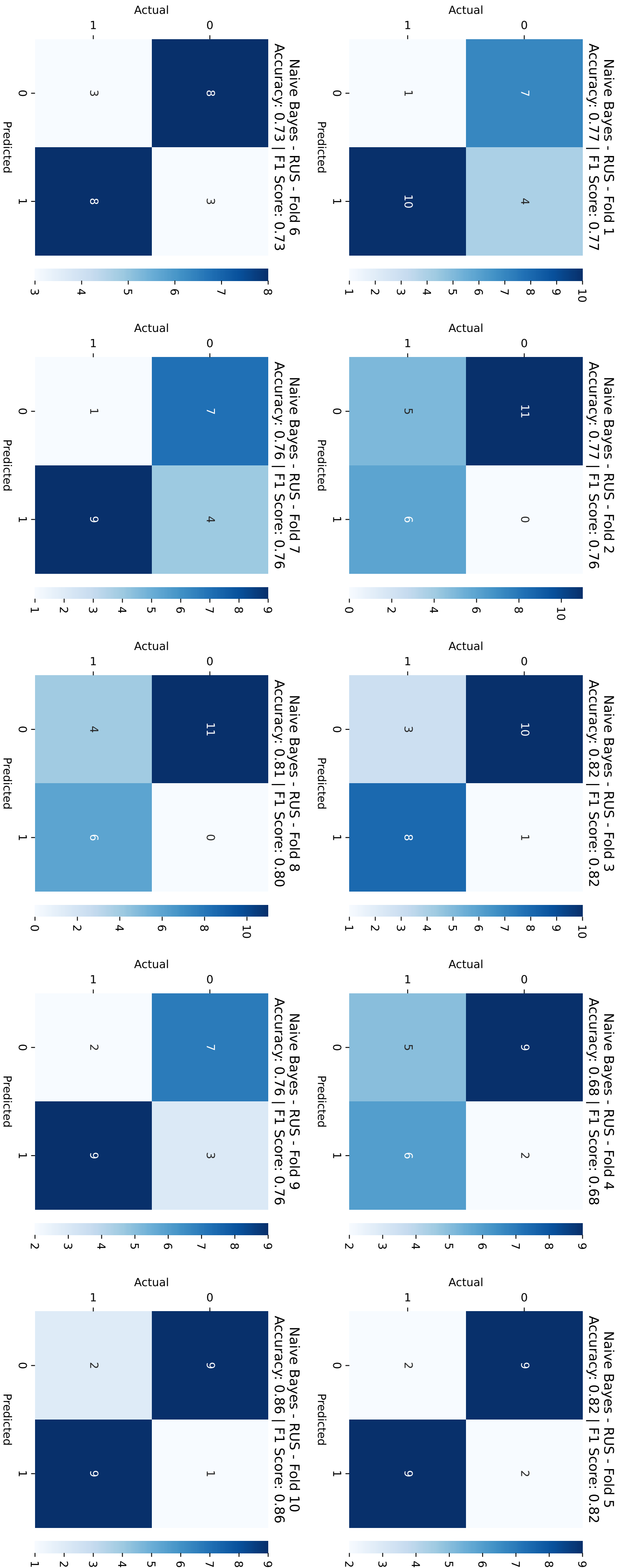


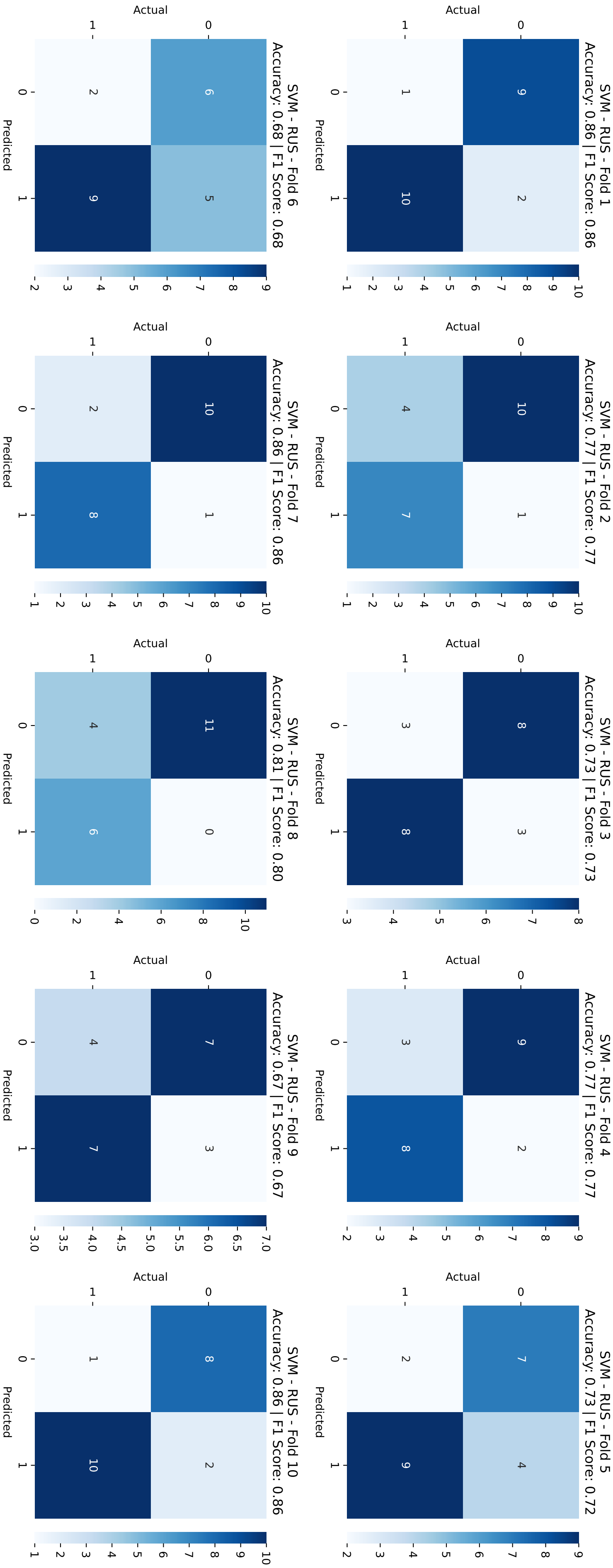


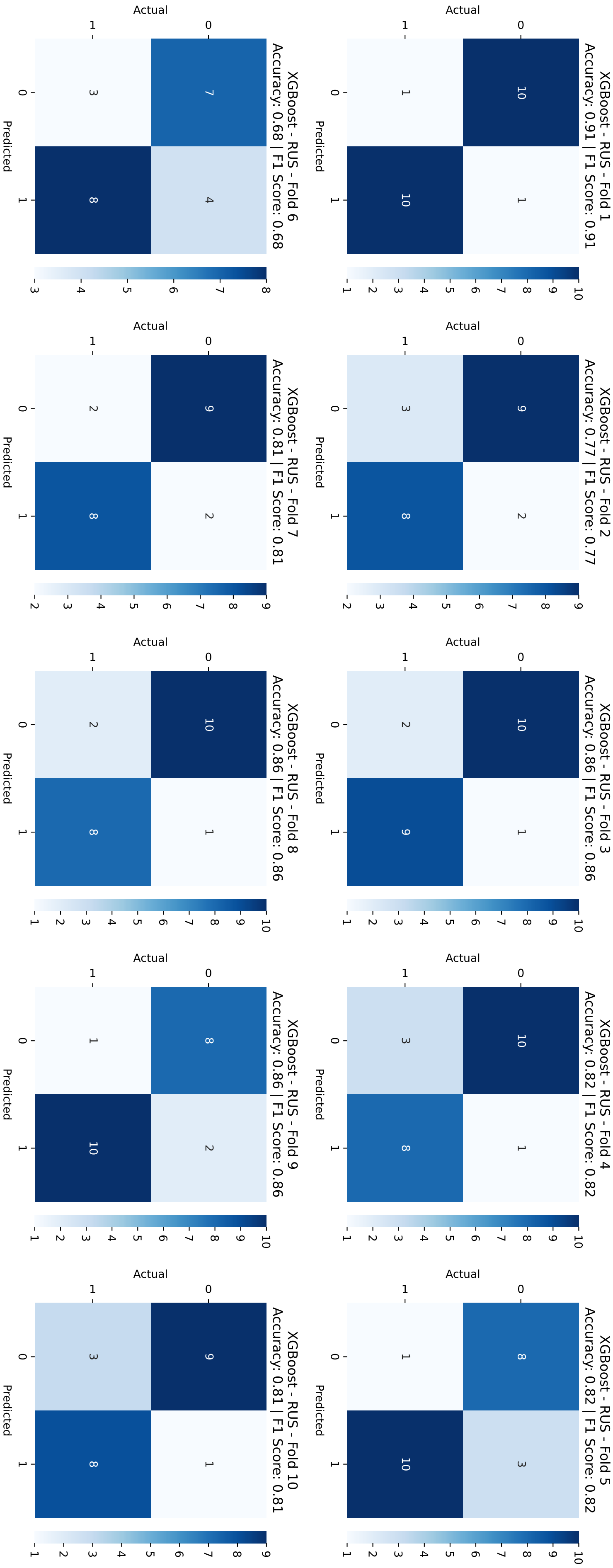


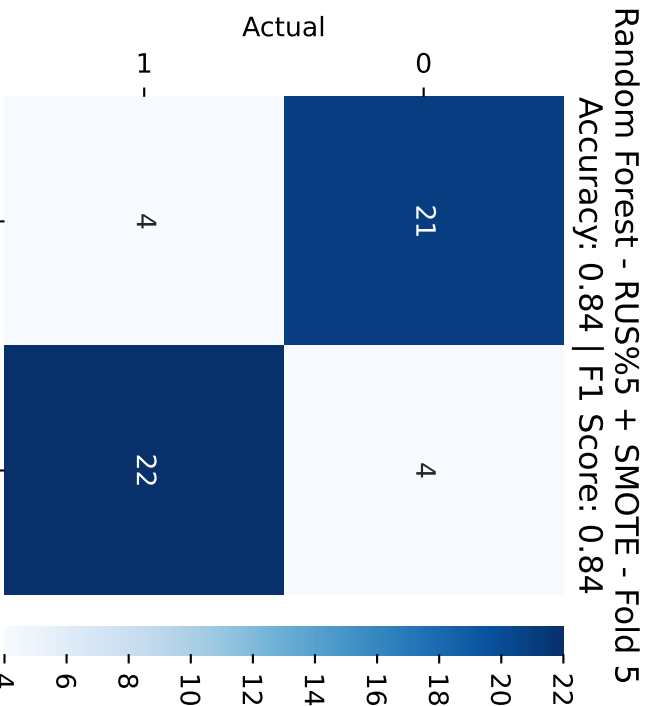
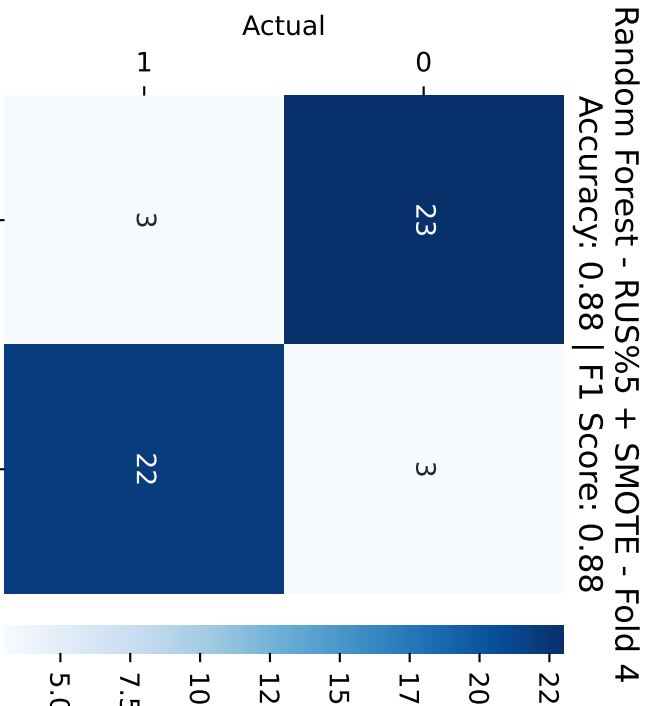
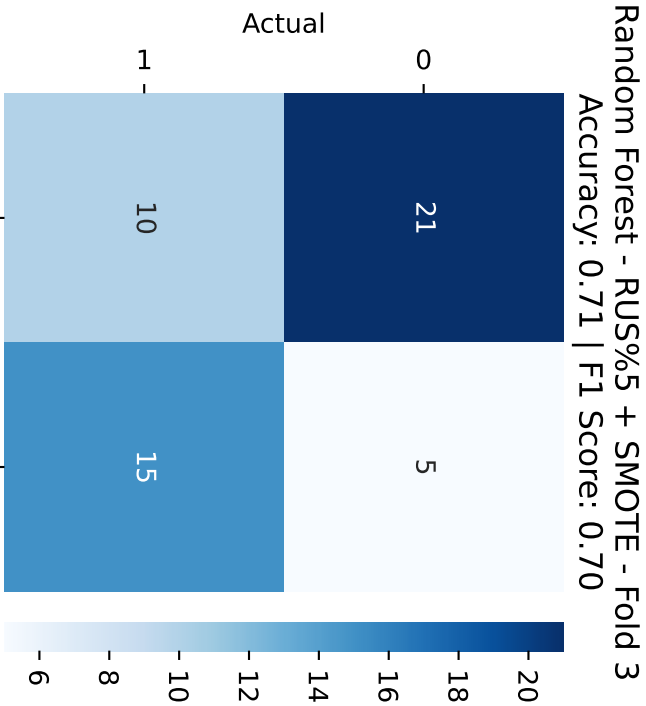
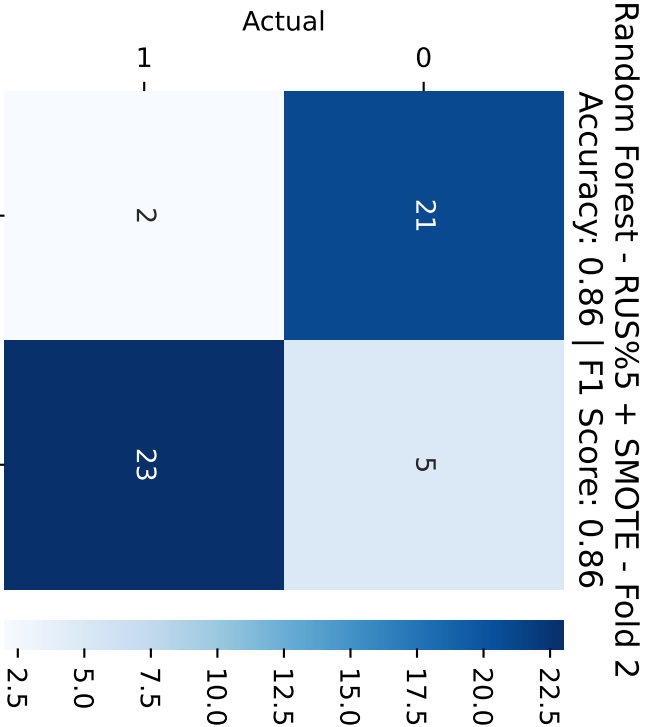
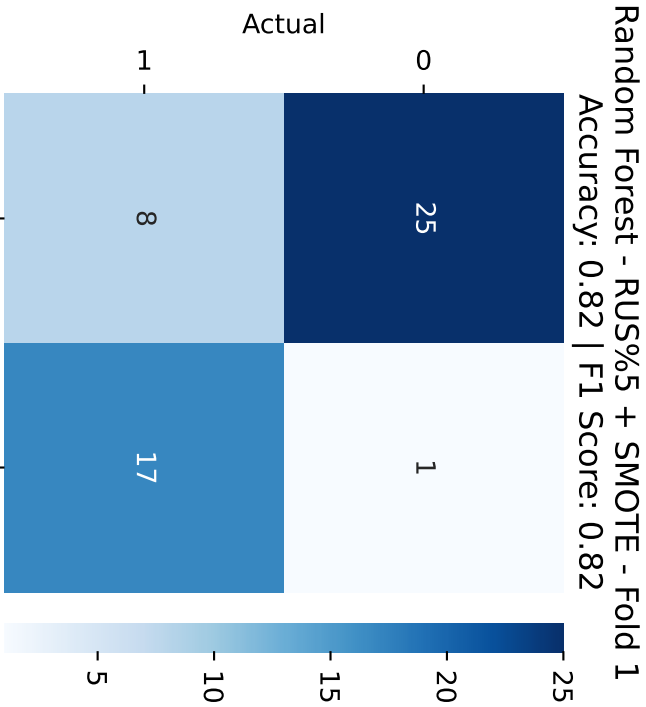


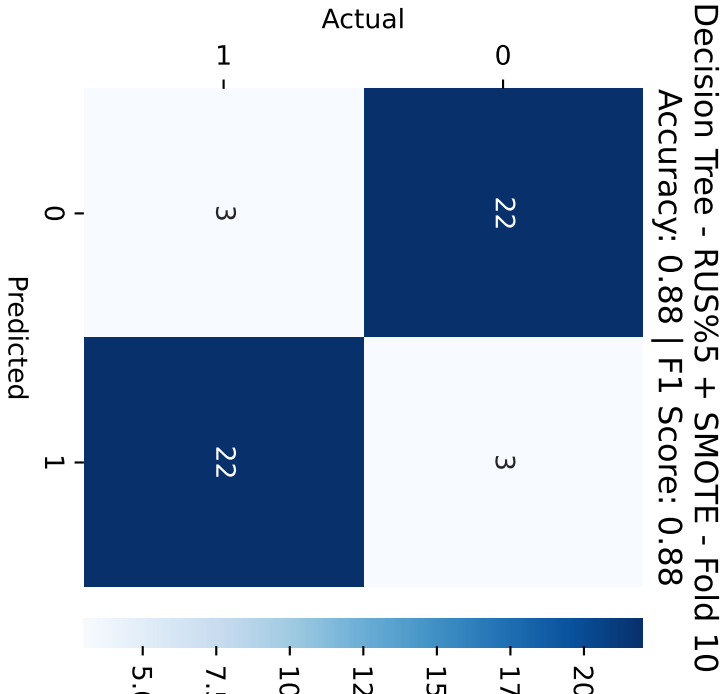
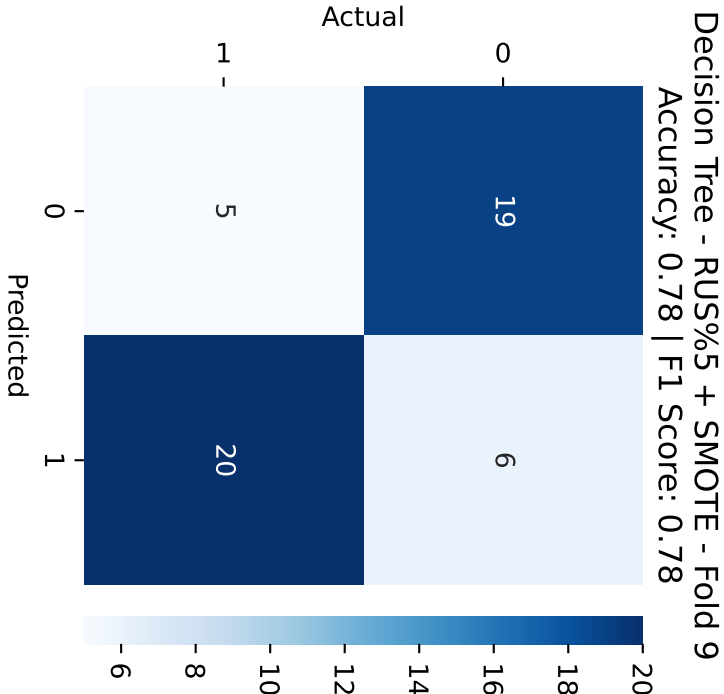
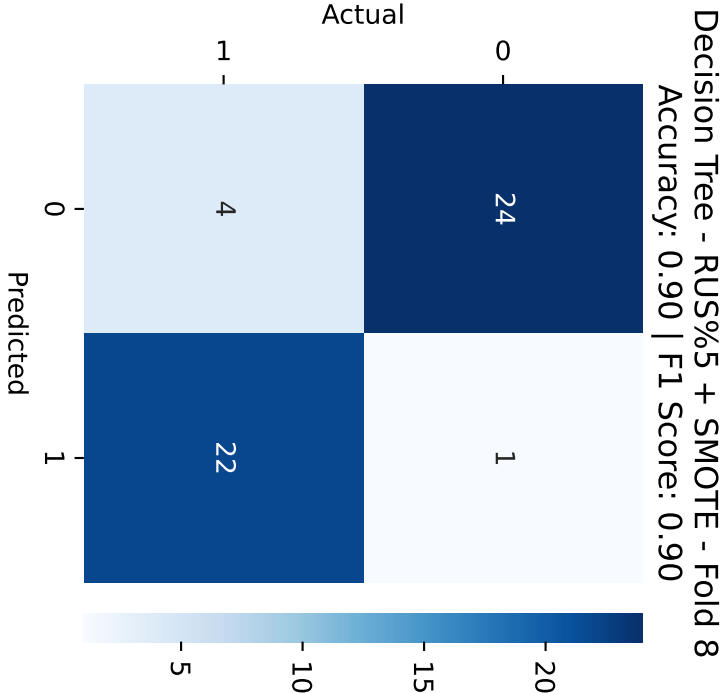
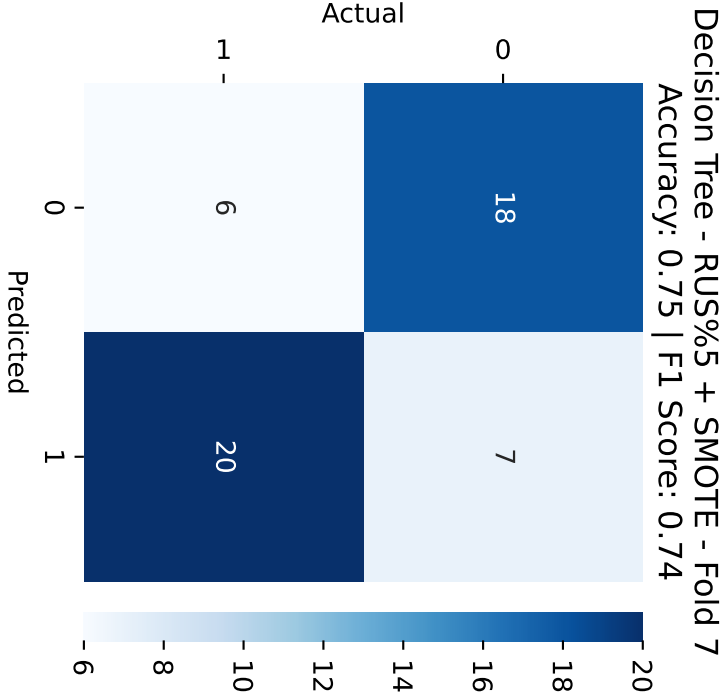
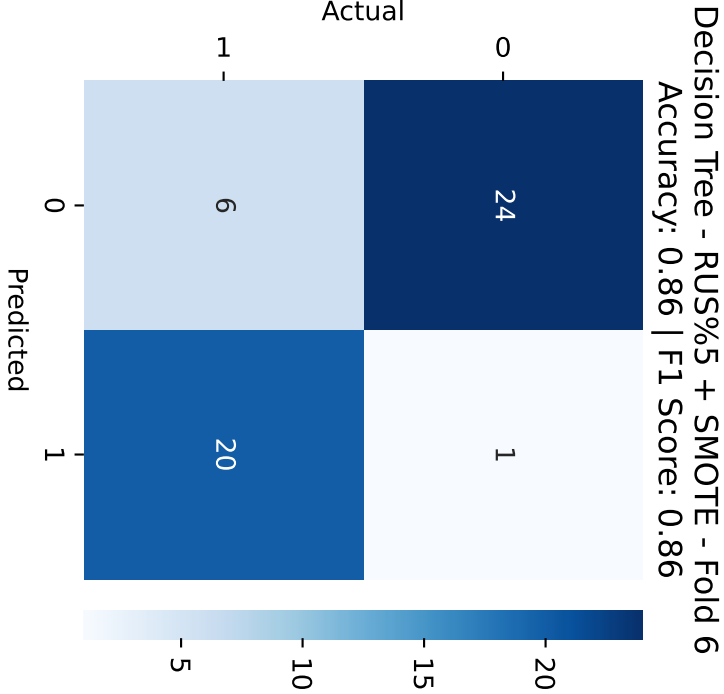
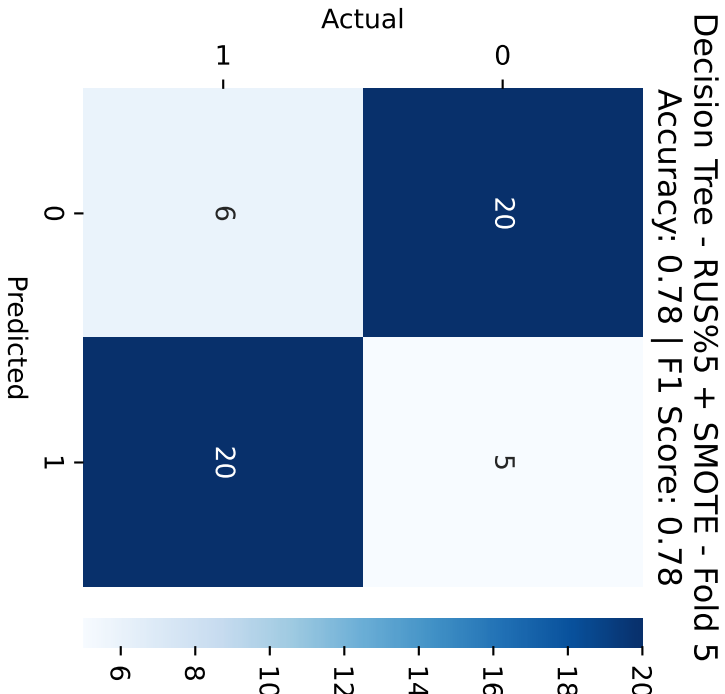
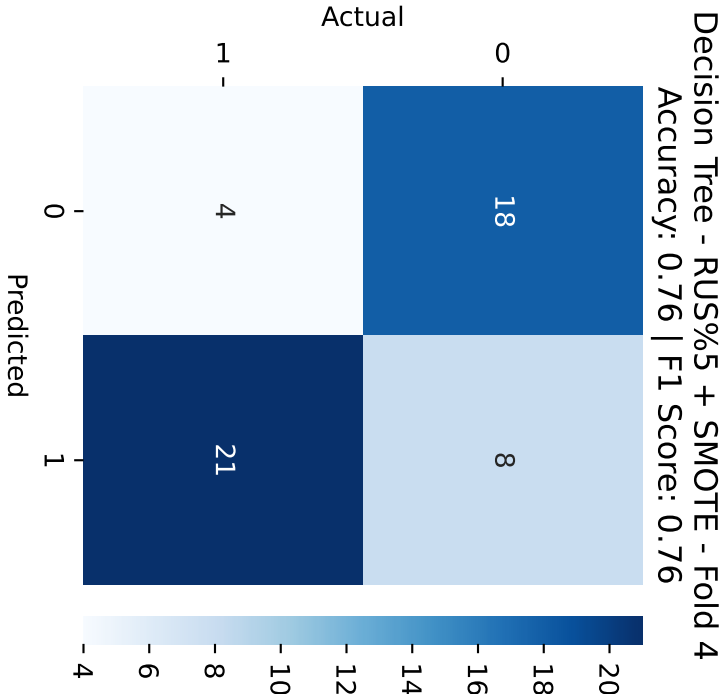
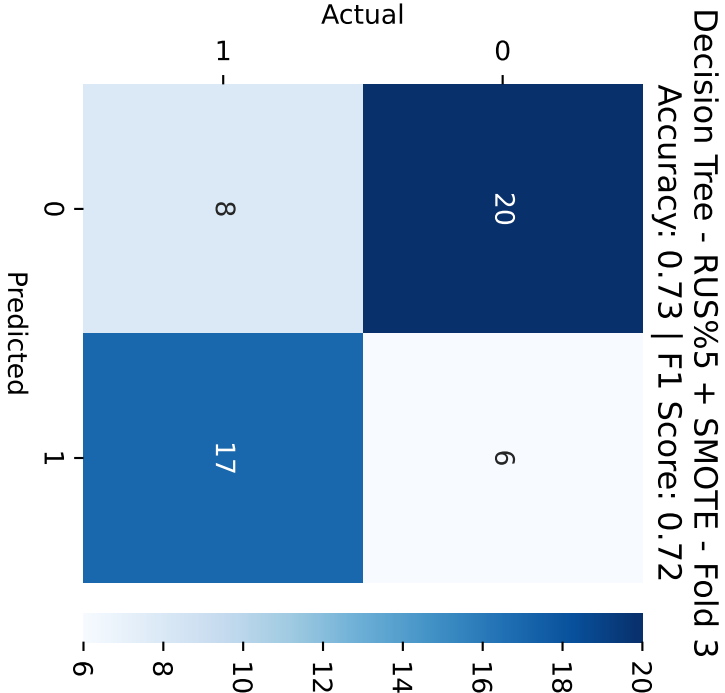
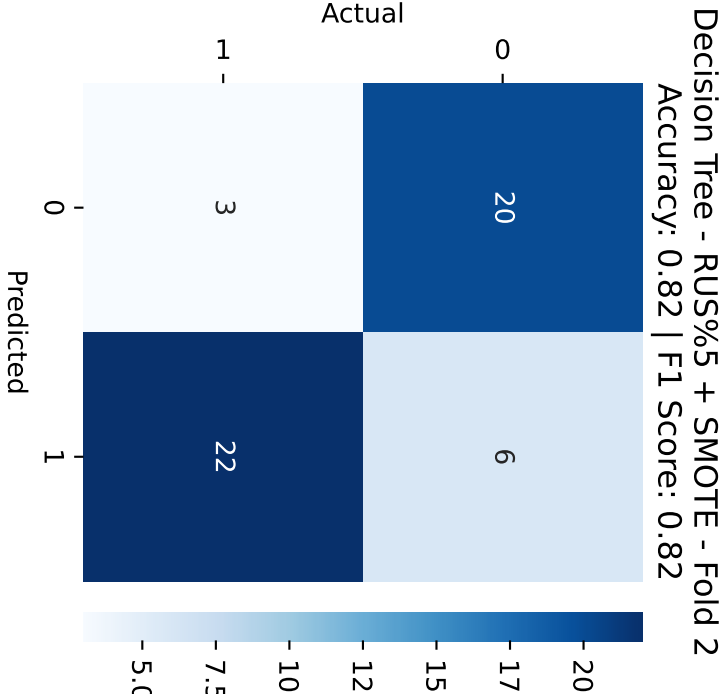
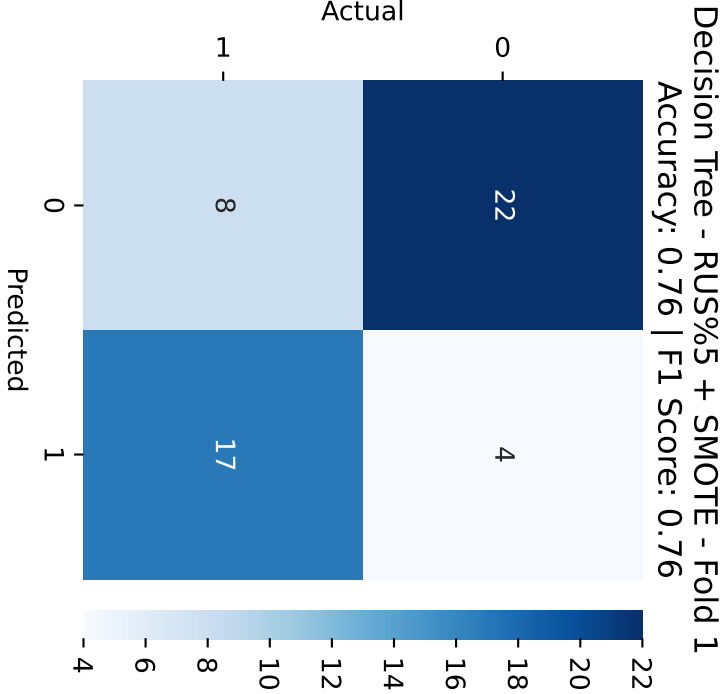


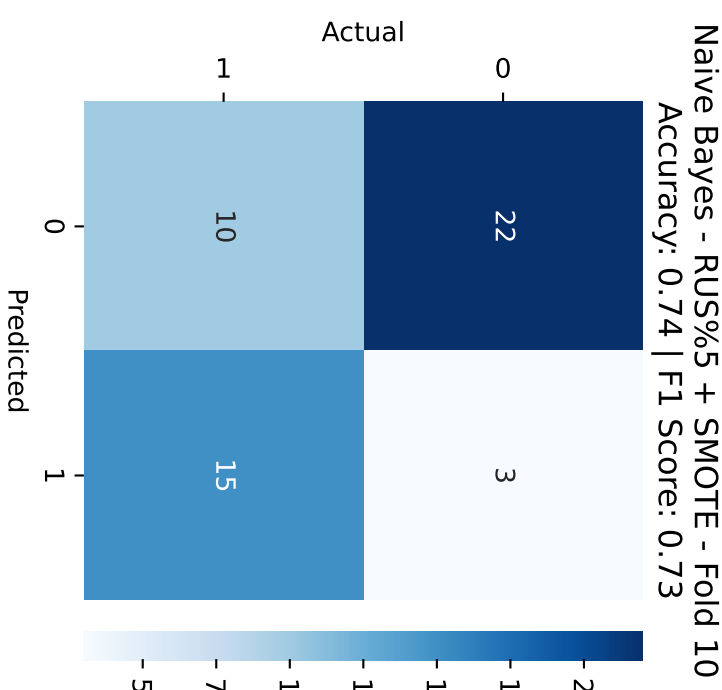
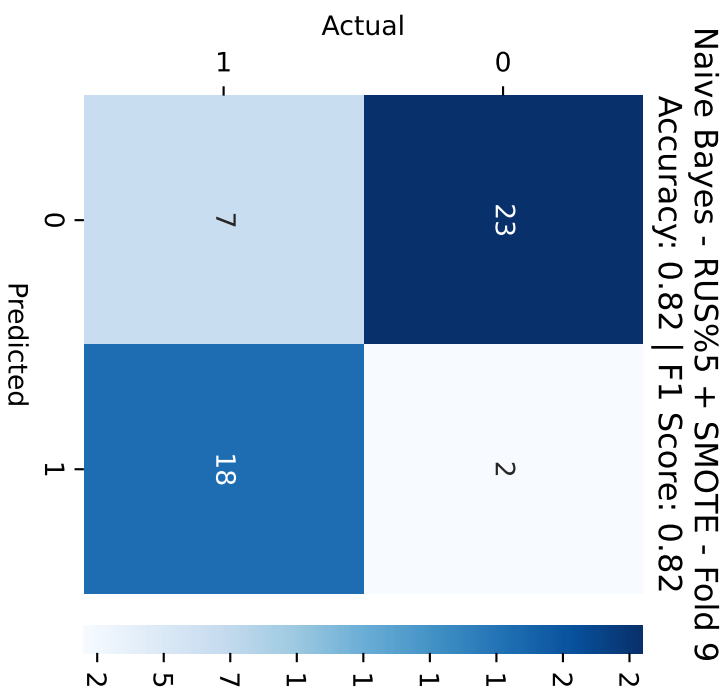
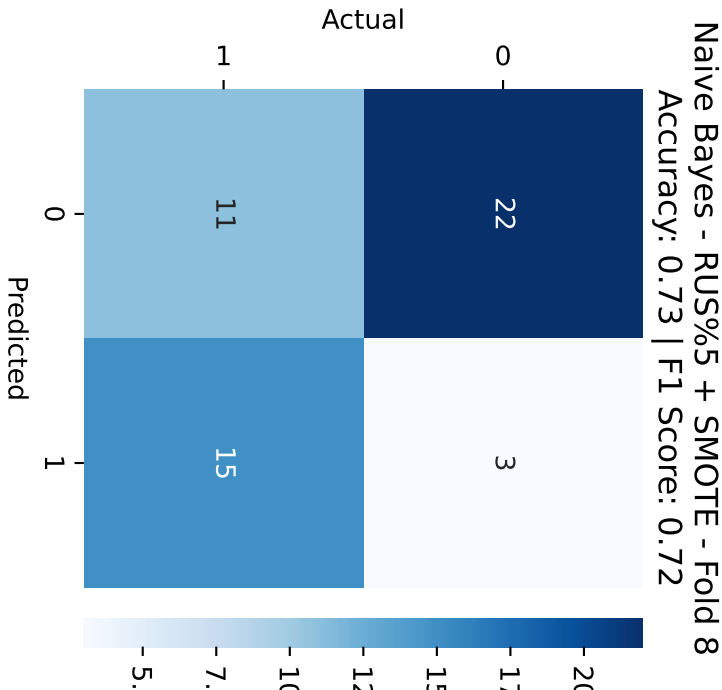
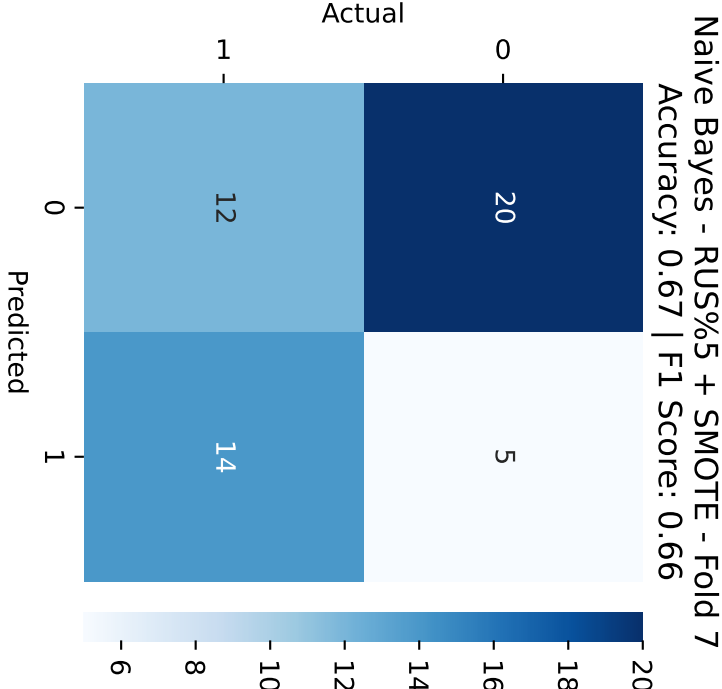
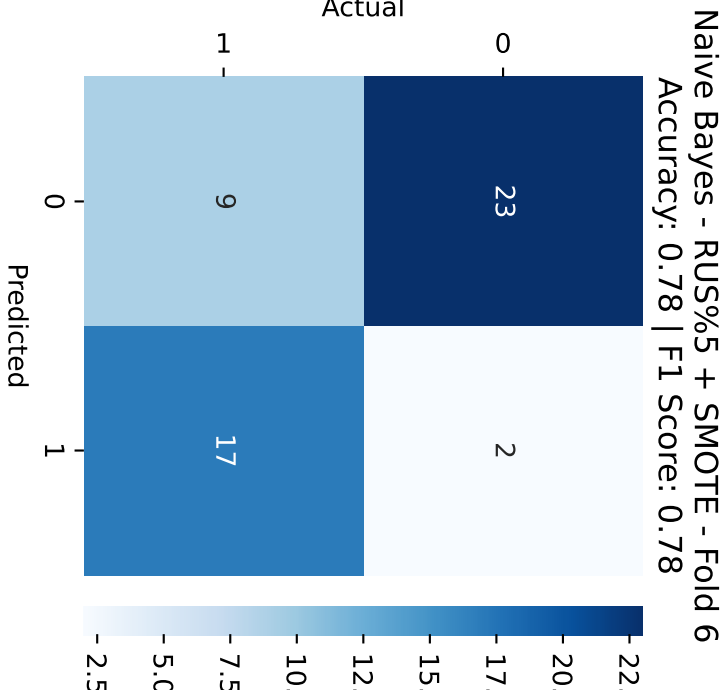
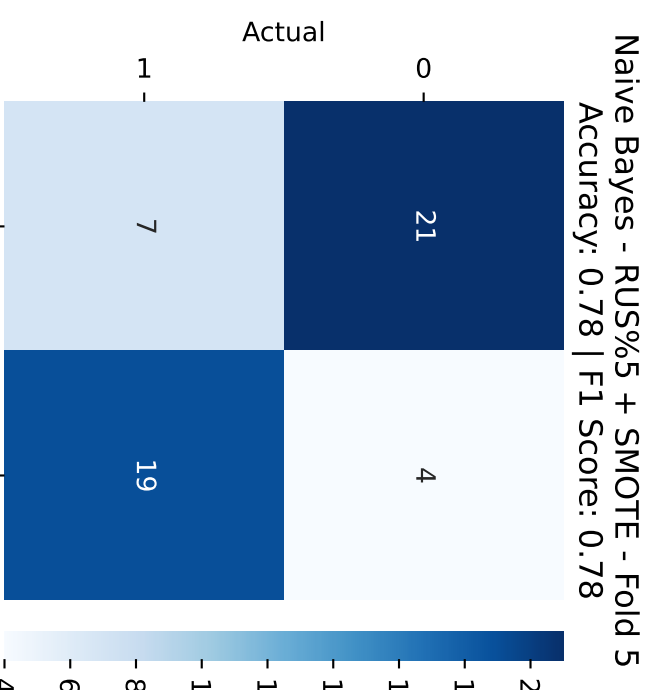
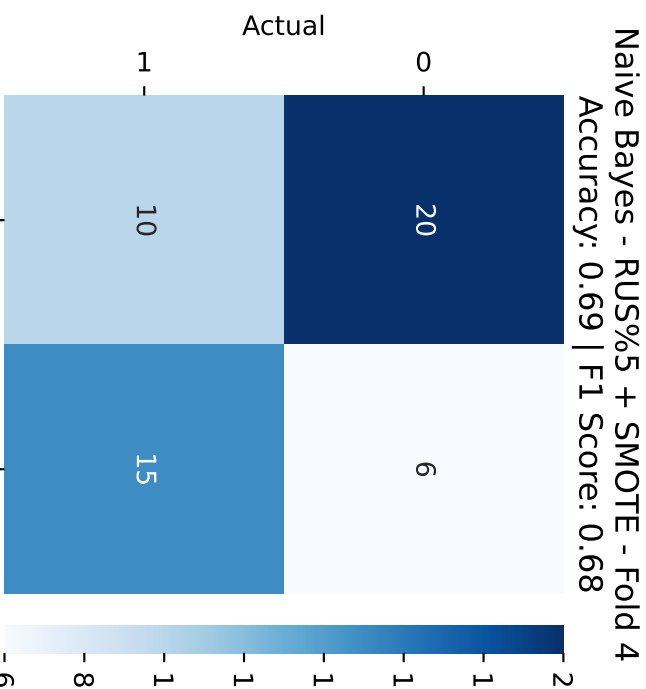
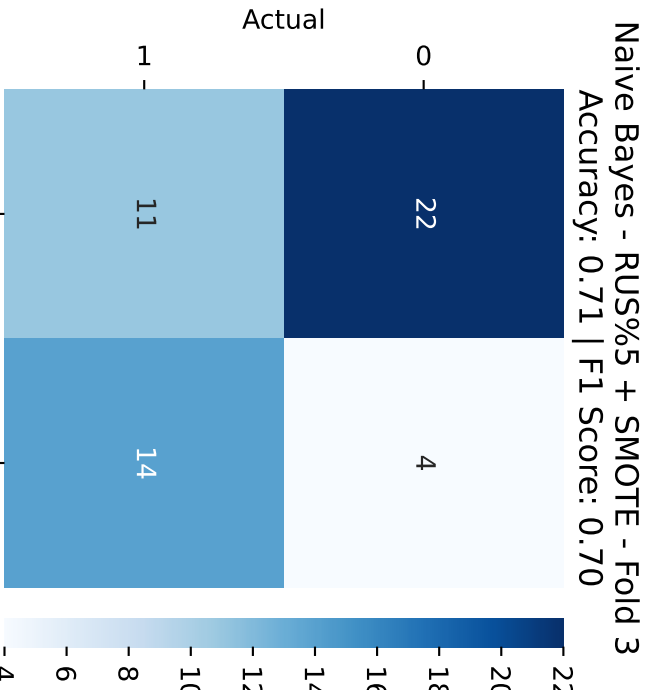
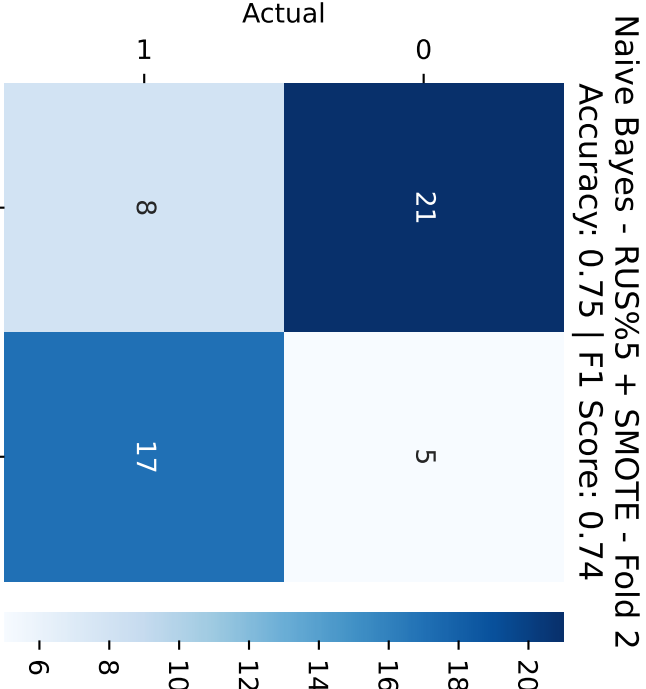
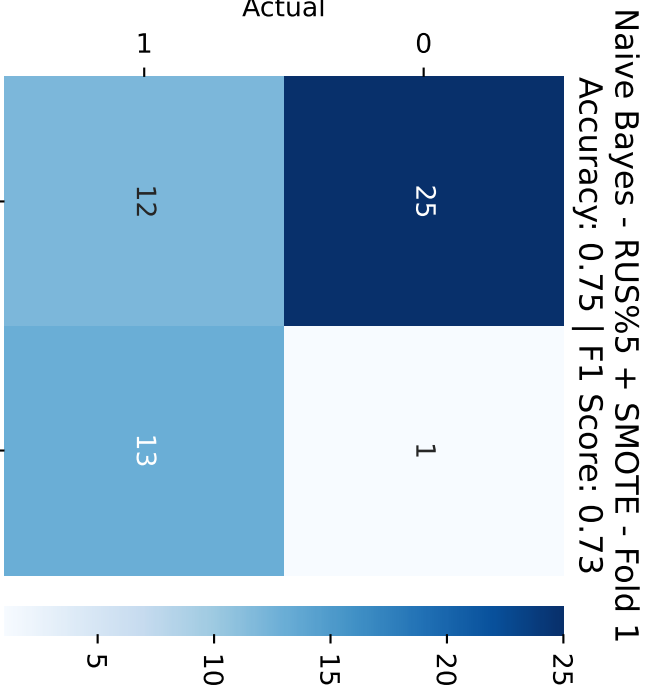


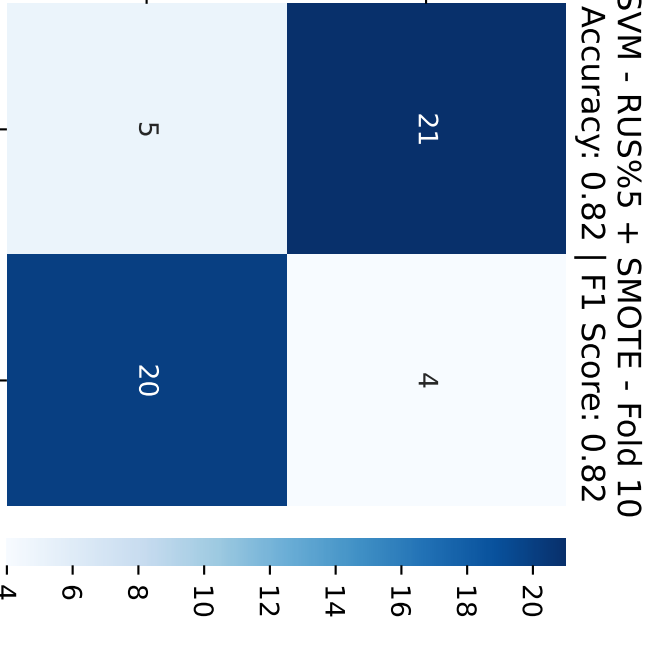
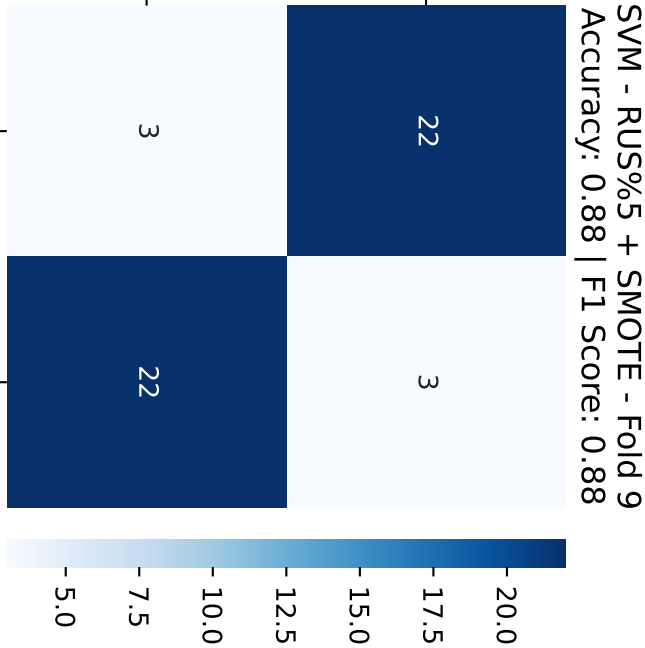
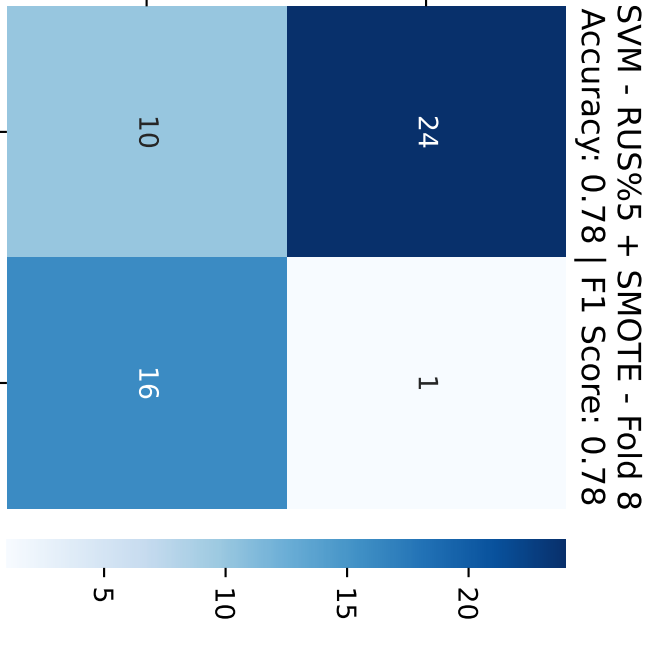
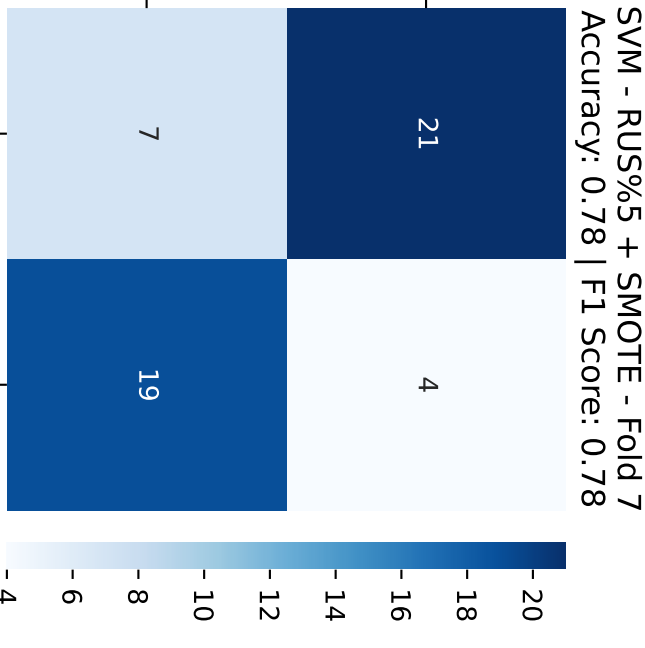
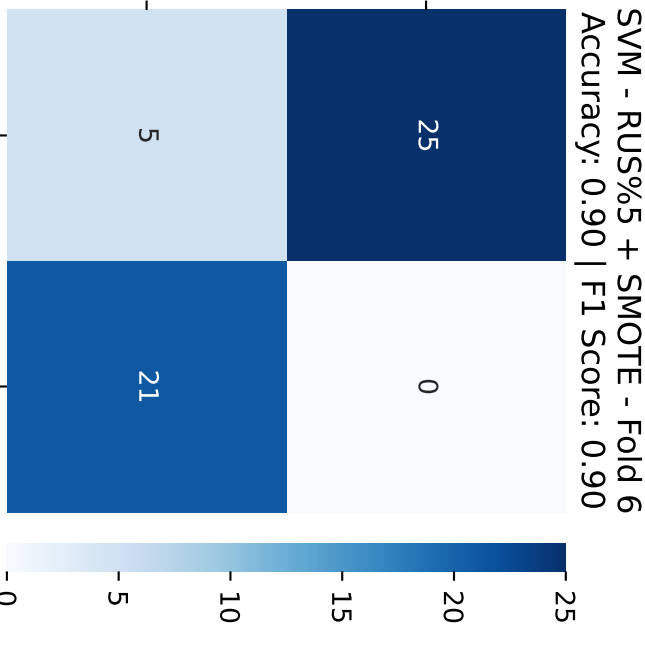
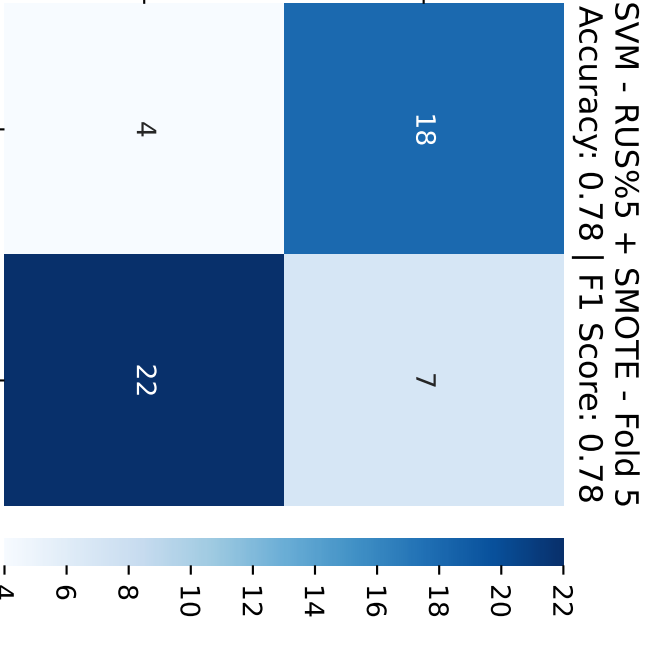
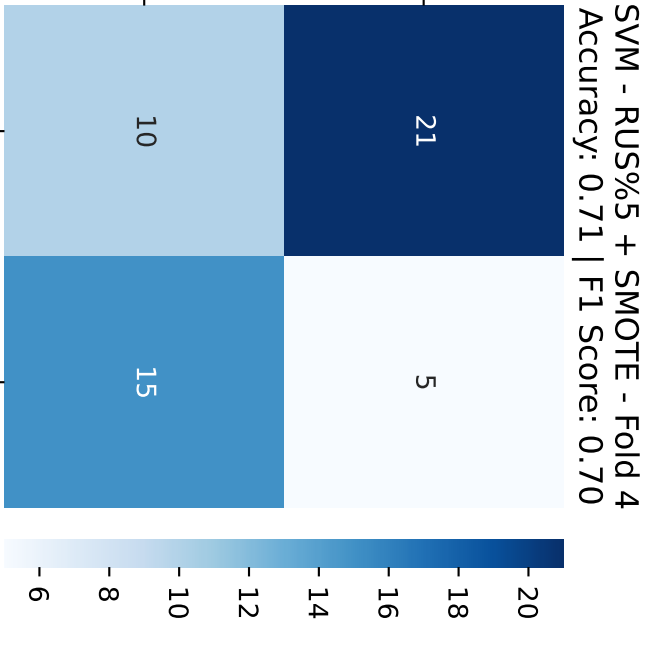
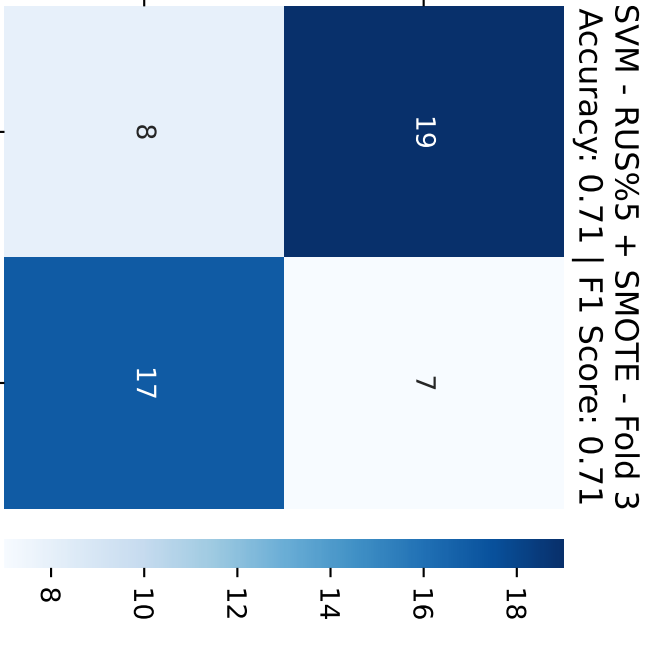
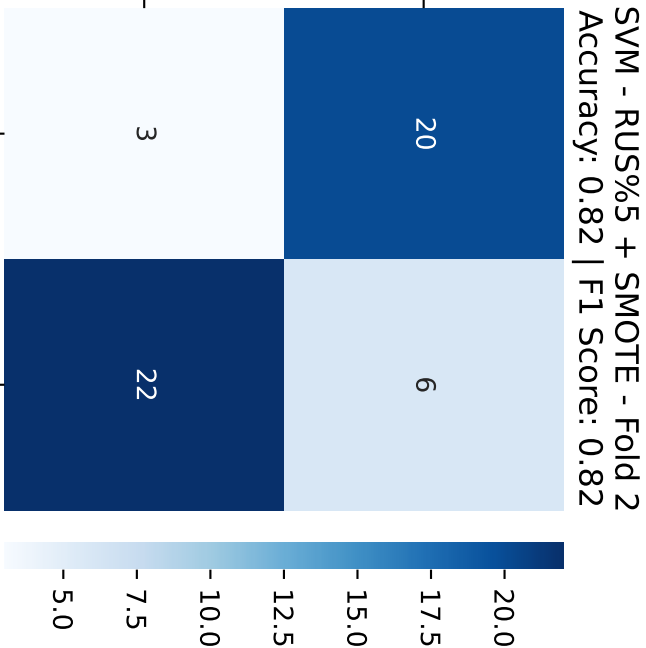
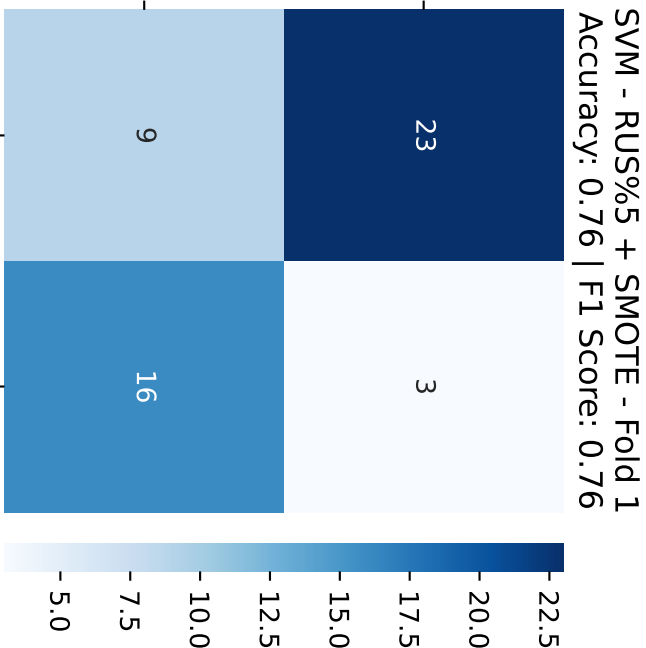


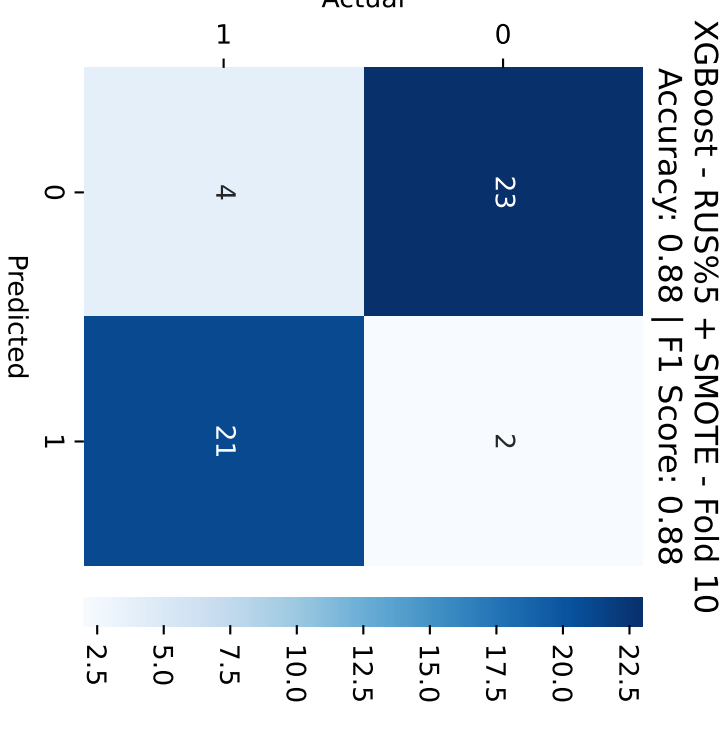
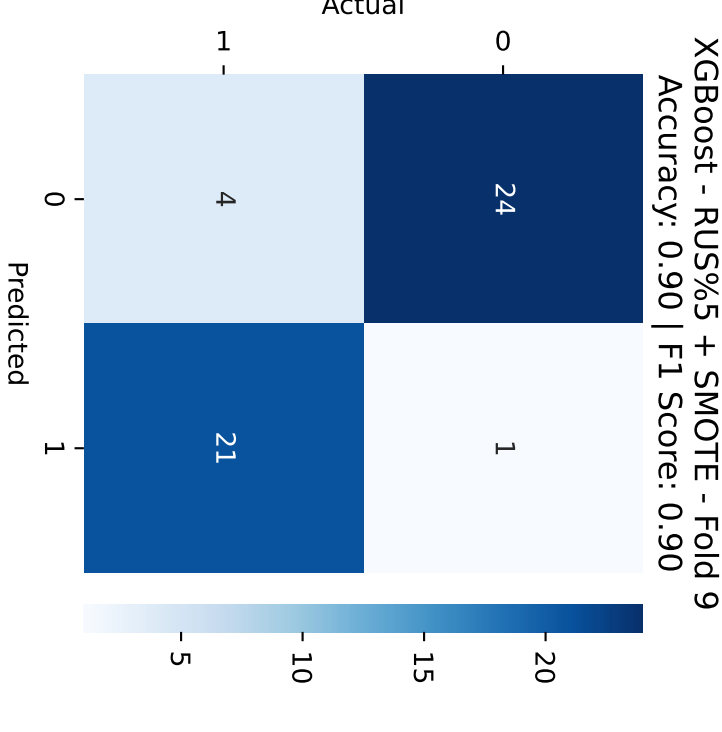
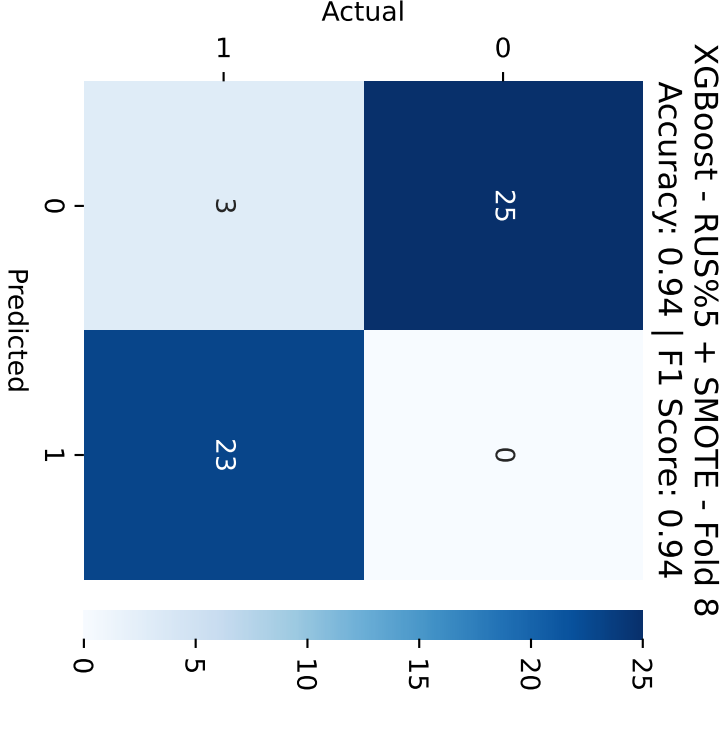
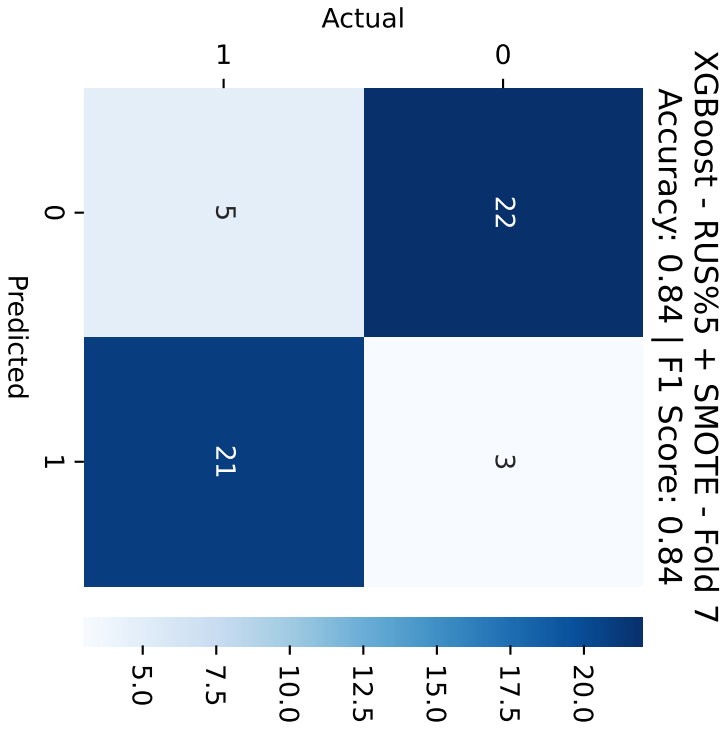
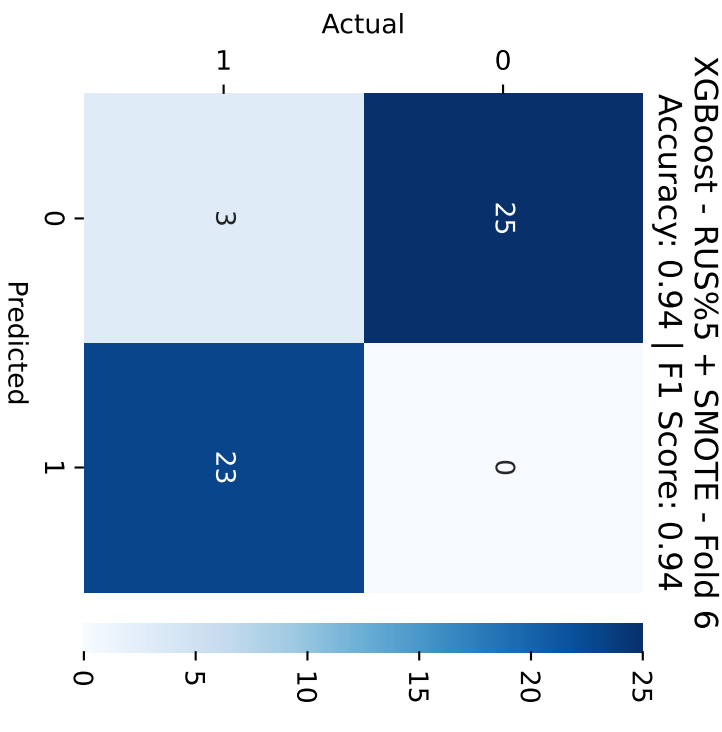
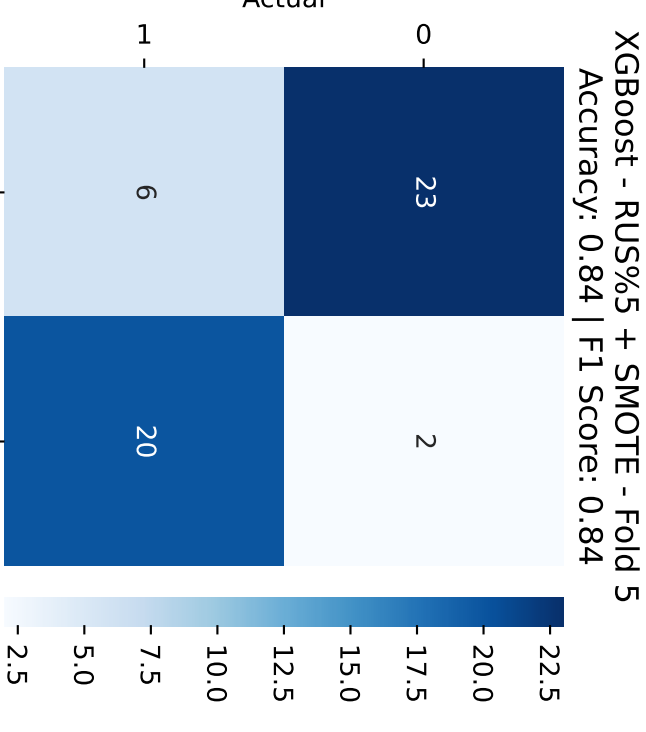
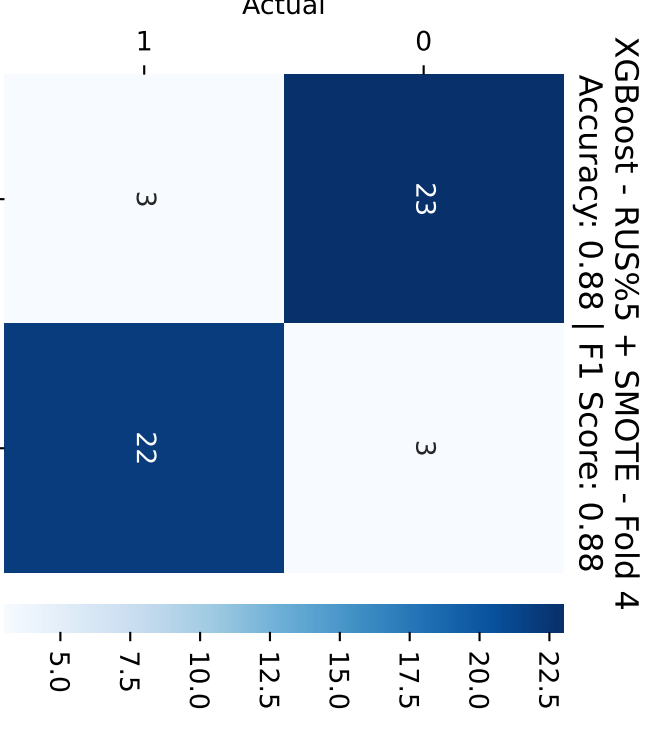
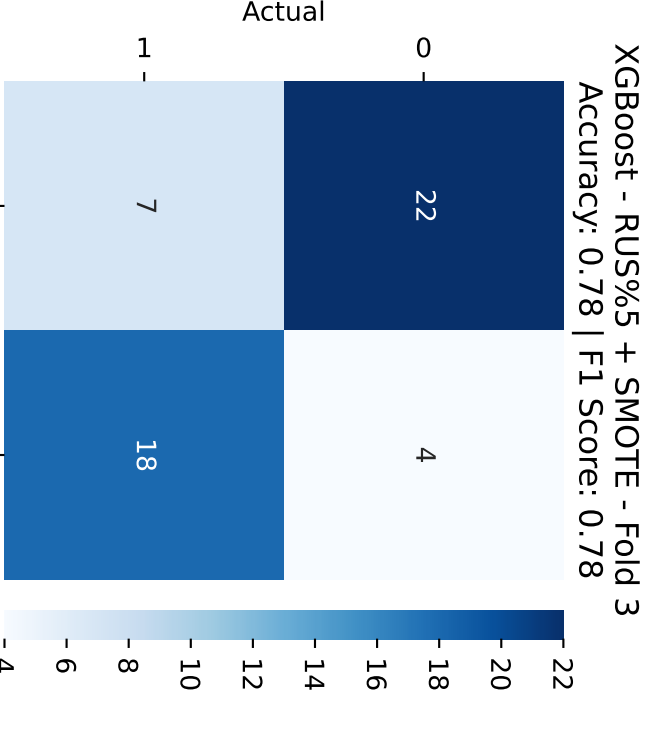
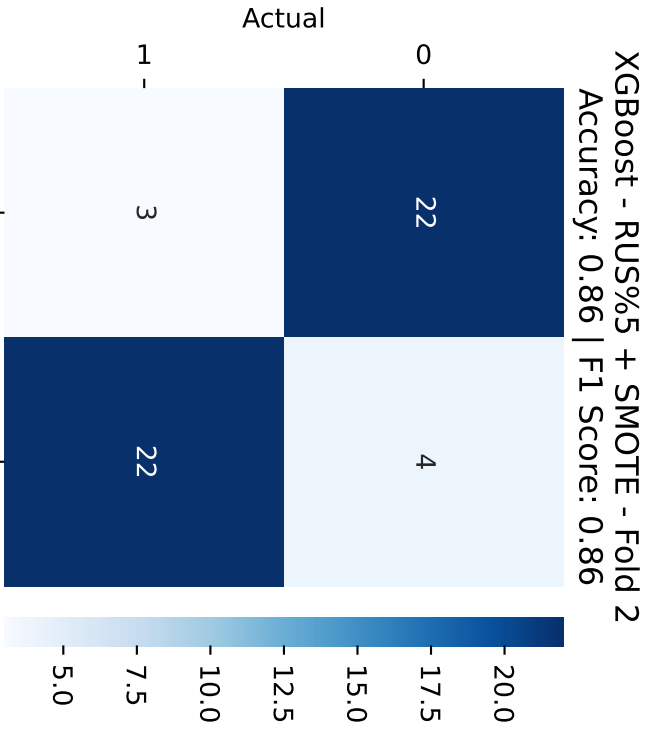
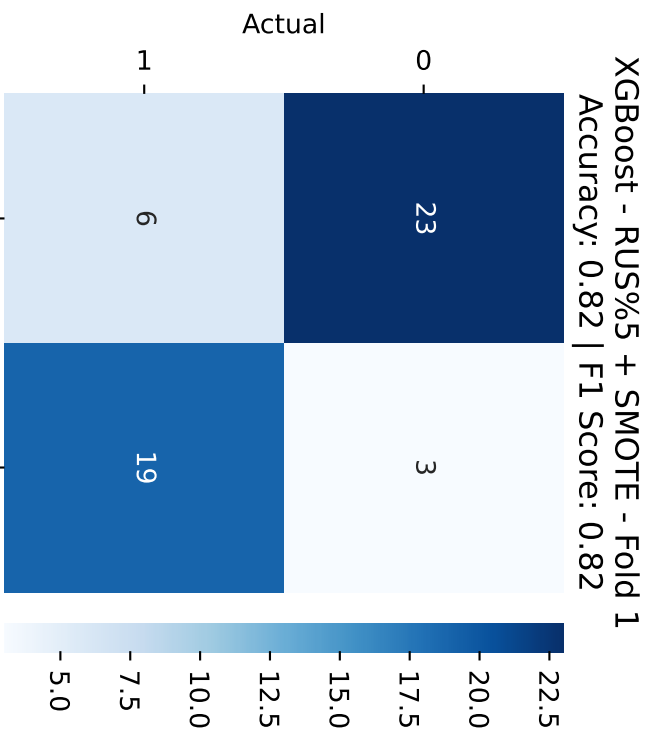












## **EK B**

Çalışma sonucunda elde edilen her bir PCA yaklaşımına ait örnekleme ve sınıflandırma algoritmalarına ait karmaşıklık matrisi grafiklerini içerir.



| PCA Approach | Sampler | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix  |
|--------------|---------|---------------|------|-------------|-------------|-------------------|
| No_PCA       | SMOTE   | Random Forest | 1    | 0,894117647 | 0,893823897 | [[40 2] [ 7 36]]  |
| No_PCA       | SMOTE   | Random Forest | 2    | 0,917647059 | 0,917165932 | [[42 0] [ 7 36]]  |
| No_PCA       | SMOTE   | Random Forest | 3    | 0,905882353 | 0,905882353 | [[38 4] [ 4 39]]  |
| No_PCA       | SMOTE   | Random Forest | 4    | 0,905882353 | 0,905725838 | [[40 2] [ 6 37]]  |
| No_PCA       | SMOTE   | Random Forest | 5    | 0,894117647 | 0,894088321 | [[37 5] [ 4 39]]  |
| No_PCA       | SMOTE   | Random Forest | 6    | 0,894117647 | 0,893941397 | [[40 3] [ 6 36]]  |
| No_PCA       | SMOTE   | Random Forest | 7    | 0,905882353 | 0,905331497 | [[42 1] [ 7 35]]  |
| No_PCA       | SMOTE   | Random Forest | 8    | 0,941176471 | 0,941078554 | [[42 1] [ 4 38]]  |
| No_PCA       | SMOTE   | Random Forest | 9    | 0,882352941 | 0,882352941 | [[38 5] [ 5 37]]  |
| No_PCA       | SMOTE   | Random Forest | 10   | 0,941176471 | 0,941160178 | [[41 2] [ 3 39]]  |
| No_PCA       | SMOTE   | Decision Tree | 1    | 0,858823529 | 0,858588757 | [[38 4] [ 8 35]]  |
| No_PCA       | SMOTE   | Decision Tree | 2    | 0,905882353 | 0,905804095 | [[37 5] [ 3 40]]  |
| No_PCA       | SMOTE   | Decision Tree | 3    | 0,788235294 | 0,787647059 | [[31 11] [ 7 36]] |
| No_PCA       | SMOTE   | Decision Tree | 4    | 0,823529412 | 0,823039828 | [[37 5] [10 33]]  |
| No_PCA       | SMOTE   | Decision Tree | 5    | 0,882352941 | 0,882255119 | [[36 6] [ 4 39]]  |
| No_PCA       | SMOTE   | Decision Tree | 6    | 0,847058824 | 0,846420594 | [[39 4] [ 9 33]]  |
| No_PCA       | SMOTE   | Decision Tree | 7    | 0,905882353 | 0,905804095 | [[40 3] [ 5 37]]  |
| No_PCA       | SMOTE   | Decision Tree | 8    | 0,870588235 | 0,870372818 | [[39 4] [ 7 35]]  |
| No_PCA       | SMOTE   | Decision Tree | 9    | 0,8         | 0,8         | [[34 9] [ 8 34]]  |
| No_PCA       | SMOTE   | Decision Tree | 10   | 0,905882353 | 0,905882353 | [[39 4] [ 4 38]]  |
| No_PCA       | SMOTE   | Naive Bayes   | 1    | 0,823529412 | 0,823480535 | [[34 8] [ 7 36]]  |
| No_PCA       | SMOTE   | Naive Bayes   | 2    | 0,729411765 | 0,728282589 | [[28 14] [ 9 34]] |
| No_PCA       | SMOTE   | Naive Bayes   | 3    | 0,729411765 | 0,725993808 | [[26 16] [ 7 36]] |
| No_PCA       | SMOTE   | Naive Bayes   | 4    | 0,705882353 | 0,703582313 | [[26 16] [ 9 34]] |
| No_PCA       | SMOTE   | Naive Bayes   | 5    | 0,705882353 | 0,705392769 | [[28 14] [11 32]] |
| No_PCA       | SMOTE   | Naive Bayes   | 6    | 0,764705882 | 0,764314595 | [[31 12] [ 8 34]] |
| No_PCA       | SMOTE   | Naive Bayes   | 7    | 0,835294118 | 0,834836601 | [[38 5] [ 9 33]]  |
| No_PCA       | SMOTE   | Naive Bayes   | 8    | 0,823529412 | 0,823382781 | [[34 9] [ 6 36]]  |
| No_PCA       | SMOTE   | Naive Bayes   | 9    | 0,752941176 | 0,752255759 | [[30 13] [ 8 34]] |
| No_PCA       | SMOTE   | Naive Bayes   | 10   | 0,788235294 | 0,788235294 | [[34 9] [ 9 33]]  |
| No_PCA       | SMOTE   | SVM           | 1    | 0,858823529 | 0,857721818 | [[40 2] [10 33]]  |
| No_PCA       | SMOTE   | SVM           | 2    | 0,870588235 | 0,870480706 | [[38 4] [ 7 36]]  |
| No_PCA       | SMOTE   | SVM           | 3    | 0,8         | 0,799667083 | [[32 10] [ 7 36]] |
| No_PCA       | SMOTE   | SVM           | 4    | 0,847058824 | 0,847058824 | [[36 6] [ 7 36]]  |
| No_PCA       | SMOTE   | SVM           | 5    | 0,776470588 | 0,776284855 | [[34 8] [11 32]]  |
| No_PCA       | SMOTE   | SVM           | 6    | 0,8         | 0,8         | [[34 9] [ 8 34]]  |
| No_PCA       | SMOTE   | SVM           | 7    | 0,894117647 | 0,893941397 | [[40 3] [ 6 36]]  |
| No_PCA       | SMOTE   | SVM           | 8    | 0,882352941 | 0,882026144 | [[40 3] [ 7 35]]  |
| No_PCA       | SMOTE   | SVM           | 9    | 0,823529412 | 0,823480535 | [[36 7] [ 8 34]]  |
| No_PCA       | SMOTE   | SVM           | 10   | 0,788235294 | 0,786995869 | [[37 6] [12 30]]  |
| No_PCA       | SMOTE   | XGBoost       | 1    | 0,894117647 | 0,894029668 | [[39 3] [ 6 37]]  |
| No_PCA       | SMOTE   | XGBoost       | 2    | 0,941176471 | 0,941127594 | [[41 1] [ 4 39]]  |
| No_PCA       | SMOTE   | XGBoost       | 3    | 0,941176471 | 0,941127594 | [[41 1] [ 4 39]]  |
| No_PCA       | SMOTE   | XGBoost       | 4    | 0,894117647 | 0,894029668 | [[39 3] [ 6 37]]  |
| No_PCA       | SMOTE   | XGBoost       | 5    | 0,894117647 | 0,894088321 | [[37 5] [ 4 39]]  |
| No_PCA       | SMOTE   | XGBoost       | 6    | 0,847058824 | 0,84680424  | [[38 5] [ 8 34]]  |
| No_PCA       | SMOTE   | XGBoost       | 7    | 0,882352941 | 0,881664371 | [[41 2] [ 8 34]]  |
| No_PCA       | SMOTE   | XGBoost       | 8    | 0,929411765 | 0,929353072 | [[41 2] [ 4 38]]  |
| No_PCA       | SMOTE   | XGBoost       | 9    | 0,917647059 | 0,91762425  | [[40 3] [ 4 38]]  |
| No_PCA       | SMOTE   | XGBoost       | 10   | 0,952941176 | 0,952902048 | [[42 1] [ 3 39]]  |

| PCA Approach | Sampler | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix  |
|--------------|---------|---------------|------|-------------|-------------|-------------------|
| No_PCA       | ROS     | Random Forest | 1    | 0,941176471 | 0,941013276 | [[42 0] [ 5 38]]  |
| No_PCA       | ROS     | Random Forest | 2    | 0,929411765 | 0,929117647 | [[42 0] [ 6 37]]  |
| No_PCA       | ROS     | Random Forest | 3    | 0,964705882 | 0,964676556 | [[42 0] [ 3 40]]  |
| No_PCA       | ROS     | Random Forest | 4    | 0,952941176 | 0,952862919 | [[42 0] [ 4 39]]  |
| No_PCA       | ROS     | Random Forest | 5    | 0,952941176 | 0,952928148 | [[41 1] [ 3 40]]  |
| No_PCA       | ROS     | Random Forest | 6    | 0,964705882 | 0,964647132 | [[43 0] [ 3 39]]  |
| No_PCA       | ROS     | Random Forest | 7    | 0,917647059 | 0,917003048 | [[43 0] [ 7 35]]  |
| No_PCA       | ROS     | Random Forest | 8    | 0,952941176 | 0,952810458 | [[43 0] [ 4 38]]  |
| No_PCA       | ROS     | Random Forest | 9    | 0,941176471 | 0,940930998 | [[43 0] [ 5 37]]  |
| No_PCA       | ROS     | Random Forest | 10   | 0,964705882 | 0,964647132 | [[43 0] [ 3 39]]  |
| No_PCA       | ROS     | Decision Tree | 1    | 0,929411765 | 0,929392222 | [[40 2] [ 4 39]]  |
| No_PCA       | ROS     | Decision Tree | 2    | 0,952941176 | 0,952862919 | [[42 0] [ 4 39]]  |
| No_PCA       | ROS     | Decision Tree | 3    | 0,905882353 | 0,905147879 | [[42 0] [ 8 35]]  |
| No_PCA       | ROS     | Decision Tree | 4    | 0,941176471 | 0,941013276 | [[42 0] [ 5 38]]  |
| No_PCA       | ROS     | Decision Tree | 5    | 0,941176471 | 0,941127594 | [[41 1] [ 4 39]]  |
| No_PCA       | ROS     | Decision Tree | 6    | 0,976470588 | 0,976451024 | [[43 0] [ 2 40]]  |
| No_PCA       | ROS     | Decision Tree | 7    | 0,882352941 | 0,880531267 | [[43 0] [10 32]]  |
| No_PCA       | ROS     | Decision Tree | 8    | 0,941176471 | 0,940930998 | [[43 0] [ 5 37]]  |
| No_PCA       | ROS     | Decision Tree | 9    | 0,941176471 | 0,940930998 | [[43 0] [ 5 37]]  |
| No_PCA       | ROS     | Decision Tree | 10   | 0,929411765 | 0,928998623 | [[43 0] [ 6 36]]  |
| No_PCA       | ROS     | Naive Bayes   | 1    | 0,823529412 | 0,823480535 | [[34 8] [ 7 36]]  |
| No_PCA       | ROS     | Naive Bayes   | 2    | 0,729411765 | 0,728282589 | [[28 14] [ 9 34]] |
| No_PCA       | ROS     | Naive Bayes   | 3    | 0,776470588 | 0,775537791 | [[30 12] [ 7 36]] |
| No_PCA       | ROS     | Naive Bayes   | 4    | 0,788235294 | 0,788235294 | [[33 9] [ 9 34]]  |
| No_PCA       | ROS     | Naive Bayes   | 5    | 0,670588235 | 0,667268296 | [[24 18] [10 33]] |
| No_PCA       | ROS     | Naive Bayes   | 6    | 0,741176471 | 0,739156666 | [[28 15] [ 7 35]] |
| No_PCA       | ROS     | Naive Bayes   | 7    | 0,776470588 | 0,776470588 | [[33 10] [ 9 33]] |
| No_PCA       | ROS     | Naive Bayes   | 8    | 0,752941176 | 0,750457133 | [[28 15] [ 6 36]] |
| No_PCA       | ROS     | Naive Bayes   | 9    | 0,694117647 | 0,690264146 | [[25 18] [ 8 34]] |
| No_PCA       | ROS     | Naive Bayes   | 10   | 0,8         | 0,799944606 | [[35 8] [ 9 33]]  |
| No_PCA       | ROS     | SVM           | 1    | 0,847058824 | 0,846165302 | [[39 3] [10 33]]  |
| No_PCA       | ROS     | SVM           | 2    | 0,835294118 | 0,835248518 | [[36 6] [ 8 35]]  |
| No_PCA       | ROS     | SVM           | 3    | 0,8         | 0,799445138 | [[36 6] [11 32]]  |
| No_PCA       | ROS     | SVM           | 4    | 0,858823529 | 0,858235294 | [[39 3] [ 9 34]]  |
| No_PCA       | ROS     | SVM           | 5    | 0,729411765 | 0,72933682  | [[30 12] [11 32]] |
| No_PCA       | ROS     | SVM           | 6    | 0,788235294 | 0,786101047 | [[38 5] [13 29]]  |
| No_PCA       | ROS     | SVM           | 7    | 0,847058824 | 0,846420594 | [[39 4] [ 9 33]]  |
| No_PCA       | ROS     | SVM           | 8    | 0,811764706 | 0,811608191 | [[36 7] [ 9 33]]  |
| No_PCA       | ROS     | SVM           | 9    | 0,741176471 | 0,740961263 | [[33 10] [12 30]] |
| No_PCA       | ROS     | SVM           | 10   | 0,776470588 | 0,768772632 | [[41 2] [17 25]]  |
| No_PCA       | ROS     | XGBoost       | 1    | 0,941176471 | 0,941013276 | [[42 0] [ 5 38]]  |
| No_PCA       | ROS     | XGBoost       | 2    | 0,941176471 | 0,941013276 | [[42 0] [ 5 38]]  |
| No_PCA       | ROS     | XGBoost       | 3    | 0,964705882 | 0,964676556 | [[42 0] [ 3 40]]  |
| No_PCA       | ROS     | XGBoost       | 4    | 0,941176471 | 0,941013276 | [[42 0] [ 5 38]]  |
| No_PCA       | ROS     | XGBoost       | 5    | 0,952941176 | 0,952928148 | [[41 1] [ 3 40]]  |
| No_PCA       | ROS     | XGBoost       | 6    | 0,917647059 | 0,917003048 | [[43 0] [ 7 35]]  |
| No_PCA       | ROS     | XGBoost       | 7    | 0,929411765 | 0,928998623 | [[43 0] [ 6 36]]  |
| No_PCA       | ROS     | XGBoost       | 8    | 0,952941176 | 0,952810458 | [[43 0] [ 4 38]]  |
| No_PCA       | ROS     | XGBoost       | 9    | 0,952941176 | 0,952810458 | [[43 0] [ 4 38]]  |
| No_PCA       | ROS     | XGBoost       | 10   | 0,952941176 | 0,952810458 | [[43 0] [ 4 38]]  |

| PCA Approach | Sampler          | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix  |
|--------------|------------------|---------------|------|-------------|-------------|-------------------|
| No_PCA       | Borderline SMOTE | Random Forest | 1    | 0,858823529 | 0,858588757 | [[38 4] [ 8 35]]  |
| No_PCA       | Borderline SMOTE | Random Forest | 2    | 0,870588235 | 0,870480706 | [[38 4] [ 7 36]]  |
| No_PCA       | Borderline SMOTE | Random Forest | 3    | 0,941176471 | 0,941127594 | [[41 1] [ 4 39]]  |
| No_PCA       | Borderline SMOTE | Random Forest | 4    | 0,894117647 | 0,894029668 | [[39 3] [ 6 37]]  |
| No_PCA       | Borderline SMOTE | Random Forest | 5    | 0,894117647 | 0,894117647 | [[38 4] [ 5 38]]  |
| No_PCA       | Borderline SMOTE | Random Forest | 6    | 0,847058824 | 0,84680424  | [[38 5] [ 8 34]]  |
| No_PCA       | Borderline SMOTE | Random Forest | 7    | 0,882352941 | 0,882026144 | [[40 3] [ 7 35]]  |
| No_PCA       | Borderline SMOTE | Random Forest | 8    | 0,917647059 | 0,91762425  | [[40 3] [ 4 38]]  |
| No_PCA       | Borderline SMOTE | Random Forest | 9    | 0,905882353 | 0,905882353 | [[39 4] [ 4 38]]  |
| No_PCA       | Borderline SMOTE | Random Forest | 10   | 0,941176471 | 0,940930998 | [[43 0] [ 5 37]]  |
| No_PCA       | Borderline SMOTE | Decision Tree | 1    | 0,8         | 0,799944606 | [[33 9] [ 8 35]]  |
| No_PCA       | Borderline SMOTE | Decision Tree | 2    | 0,811764706 | 0,811608191 | [[33 9] [ 7 36]]  |
| No_PCA       | Borderline SMOTE | Decision Tree | 3    | 0,823529412 | 0,823529412 | [[35 7] [ 8 35]]  |
| No_PCA       | Borderline SMOTE | Decision Tree | 4    | 0,8         | 0,8         | [[34 8] [ 9 34]]  |
| No_PCA       | Borderline SMOTE | Decision Tree | 5    | 0,894117647 | 0,894088321 | [[37 5] [ 4 39]]  |
| No_PCA       | Borderline SMOTE | Decision Tree | 6    | 0,776470588 | 0,774722558 | [[37 6] [13 29]]  |
| No_PCA       | Borderline SMOTE | Decision Tree | 7    | 0,882352941 | 0,882026144 | [[40 3] [ 7 35]]  |
| No_PCA       | Borderline SMOTE | Decision Tree | 8    | 0,870588235 | 0,870372818 | [[39 4] [ 7 35]]  |
| No_PCA       | Borderline SMOTE | Decision Tree | 9    | 0,858823529 | 0,858823529 | [[37 6] [ 6 36]]  |
| No_PCA       | Borderline SMOTE | Decision Tree | 10   | 0,870588235 | 0,870372818 | [[39 4] [ 7 35]]  |
| No_PCA       | Borderline SMOTE | Naive Bayes   | 1    | 0,717647059 | 0,714801396 | [[26 16] [ 8 35]] |
| No_PCA       | Borderline SMOTE | Naive Bayes   | 2    | 0,776470588 | 0,776284855 | [[34 8] [11 32]]  |
| No_PCA       | Borderline SMOTE | Naive Bayes   | 3    | 0,776470588 | 0,776470588 | [[33 9] [10 33]]  |
| No_PCA       | Borderline SMOTE | Naive Bayes   | 4    | 0,764705882 | 0,764705882 | [[32 10] [10 33]] |
| No_PCA       | Borderline SMOTE | Naive Bayes   | 5    | 0,764705882 | 0,76464074  | [[33 9] [11 32]]  |
| No_PCA       | Borderline SMOTE | Naive Bayes   | 6    | 0,764705882 | 0,764705882 | [[33 10] [10 32]] |
| No_PCA       | Borderline SMOTE | Naive Bayes   | 7    | 0,788235294 | 0,788235294 | [[34 9] [ 9 33]]  |
| No_PCA       | Borderline SMOTE | Naive Bayes   | 8    | 0,811764706 | 0,811712592 | [[34 9] [ 7 35]]  |
| No_PCA       | Borderline SMOTE | Naive Bayes   | 9    | 0,776470588 | 0,775850448 | [[31 12] [ 7 35]] |
| No_PCA       | Borderline SMOTE | Naive Bayes   | 10   | 0,788235294 | 0,786995869 | [[37 6] [12 30]]  |
| No_PCA       | Borderline SMOTE | SVM           | 1    | 0,847058824 | 0,847058824 | [[36 6] [ 7 36]]  |
| No_PCA       | Borderline SMOTE | SVM           | 2    | 0,870588235 | 0,870480706 | [[38 4] [ 7 36]]  |
| No_PCA       | Borderline SMOTE | SVM           | 3    | 0,847058824 | 0,847058824 | [[36 6] [ 7 36]]  |
| No_PCA       | Borderline SMOTE | SVM           | 4    | 0,894117647 | 0,894029668 | [[39 3] [ 6 37]]  |
| No_PCA       | Borderline SMOTE | SVM           | 5    | 0,8         | 0,799445138 | [[36 6] [11 32]]  |
| No_PCA       | Borderline SMOTE | SVM           | 6    | 0,823529412 | 0,822792993 | [[38 5] [10 32]]  |
| No_PCA       | Borderline SMOTE | SVM           | 7    | 0,858823529 | 0,858431373 | [[39 4] [ 8 34]]  |
| No_PCA       | Borderline SMOTE | SVM           | 8    | 0,847058824 | 0,84680424  | [[38 5] [ 8 34]]  |
| No_PCA       | Borderline SMOTE | SVM           | 9    | 0,847058824 | 0,845862803 | [[40 3] [10 32]]  |
| No_PCA       | Borderline SMOTE | SVM           | 10   | 0,847058824 | 0,845126935 | [[41 2] [11 31]]  |
| No_PCA       | Borderline SMOTE | XGBoost       | 1    | 0,905882353 | 0,905725838 | [[40 2] [ 6 37]]  |
| No_PCA       | Borderline SMOTE | XGBoost       | 2    | 0,870588235 | 0,870588235 | [[37 5] [ 6 37]]  |
| No_PCA       | Borderline SMOTE | XGBoost       | 3    | 0,894117647 | 0,894029668 | [[39 3] [ 6 37]]  |
| No_PCA       | Borderline SMOTE | XGBoost       | 4    | 0,905882353 | 0,905725838 | [[40 2] [ 6 37]]  |
| No_PCA       | Borderline SMOTE | XGBoost       | 5    | 0,882352941 | 0,88232037  | [[38 4] [ 6 37]]  |
| No_PCA       | Borderline SMOTE | XGBoost       | 6    | 0,905882353 | 0,905804095 | [[40 3] [ 5 37]]  |
| No_PCA       | Borderline SMOTE | XGBoost       | 7    | 0,917647059 | 0,917303397 | [[42 1] [ 6 36]]  |
| No_PCA       | Borderline SMOTE | XGBoost       | 8    | 0,917647059 | 0,917509975 | [[41 2] [ 5 37]]  |
| No_PCA       | Borderline SMOTE | XGBoost       | 9    | 0,882352941 | 0,882352941 | [[38 5] [ 5 37]]  |
| No_PCA       | Borderline SMOTE | XGBoost       | 10   | 0,964705882 | 0,964647132 | [[43 0] [ 3 39]]  |

| PCA Approach | Sampler | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix |
|--------------|---------|---------------|------|-------------|-------------|------------------|
| No_PCA       | RUS     | Random Forest | 1    | 0,909090909 | 0,908333333 | [[ 9 2] [ 0 11]] |
| No_PCA       | RUS     | Random Forest | 2    | 0,681818182 | 0,675789474 | [[9 2] [5 6]]    |
| No_PCA       | RUS     | Random Forest | 3    | 0,772727273 | 0,772256729 | [[8 3] [2 9]]    |
| No_PCA       | RUS     | Random Forest | 4    | 0,772727273 | 0,772256729 | [[8 3] [2 9]]    |
| No_PCA       | RUS     | Random Forest | 5    | 0,727272727 | 0,725       | [[7 4] [2 9]]    |
| No_PCA       | RUS     | Random Forest | 6    | 0,727272727 | 0,725       | [[7 4] [2 9]]    |
| No_PCA       | RUS     | Random Forest | 7    | 0,714285714 | 0,71031746  | [[9 2] [4 6]]    |
| No_PCA       | RUS     | Random Forest | 8    | 0,761904762 | 0,760815081 | [[9 2] [3 7]]    |
| No_PCA       | RUS     | Random Forest | 9    | 0,80952381  | 0,80952381  | [[8 2] [2 9]]    |
| No_PCA       | RUS     | Random Forest | 10   | 0,80952381  | 0,806878307 | [[ 7 3] [ 1 10]] |
| No_PCA       | RUS     | Decision Tree | 1    | 0,772727273 | 0,772256729 | [[8 3] [2 9]]    |
| No_PCA       | RUS     | Decision Tree | 2    | 0,681818182 | 0,68115942  | [[8 3] [4 7]]    |
| No_PCA       | RUS     | Decision Tree | 3    | 0,636363636 | 0,623931624 | [[5 6] [2 9]]    |
| No_PCA       | RUS     | Decision Tree | 4    | 0,636363636 | 0,636363636 | [[7 4] [4 7]]    |
| No_PCA       | RUS     | Decision Tree | 5    | 0,545454545 | 0,541666667 | [[5 6] [4 7]]    |
| No_PCA       | RUS     | Decision Tree | 6    | 0,545454545 | 0,545454545 | [[6 5] [5 6]]    |
| No_PCA       | RUS     | Decision Tree | 7    | 0,619047619 | 0,613756614 | [[8 3] [5 5]]    |
| No_PCA       | RUS     | Decision Tree | 8    | 0,666666667 | 0,657254902 | [[9 2] [5 5]]    |
| No_PCA       | RUS     | Decision Tree | 9    | 0,619047619 | 0,617316017 | [[7 3] [5 6]]    |
| No_PCA       | RUS     | Decision Tree | 10   | 0,761904762 | 0,75863572  | [[9 1] [4 7]]    |
| No_PCA       | RUS     | Naive Bayes   | 1    | 0,818181818 | 0,816666667 | [[ 8 3] [ 1 10]] |
| No_PCA       | RUS     | Naive Bayes   | 2    | 0,772727273 | 0,772256729 | [[8 3] [2 9]]    |
| No_PCA       | RUS     | Naive Bayes   | 3    | 0,727272727 | 0,717948718 | [[ 6 5] [ 1 10]] |
| No_PCA       | RUS     | Naive Bayes   | 4    | 0,727272727 | 0,725       | [[9 2] [4 7]]    |
| No_PCA       | RUS     | Naive Bayes   | 5    | 0,681818182 | 0,675789474 | [[6 5] [2 9]]    |
| No_PCA       | RUS     | Naive Bayes   | 6    | 0,727272727 | 0,717948718 | [[ 6 5] [ 1 10]] |
| No_PCA       | RUS     | Naive Bayes   | 7    | 0,80952381  | 0,808658009 | [[8 3] [1 9]]    |
| No_PCA       | RUS     | Naive Bayes   | 8    | 0,761904762 | 0,755182073 | [[10 1] [ 4 6]]  |
| No_PCA       | RUS     | Naive Bayes   | 9    | 0,666666667 | 0,666666667 | [[7 3] [4 7]]    |
| No_PCA       | RUS     | Naive Bayes   | 10   | 0,80952381  | 0,806878307 | [[ 7 3] [ 1 10]] |
| No_PCA       | RUS     | SVM           | 1    | 0,863636364 | 0,863354037 | [[ 9 2] [ 1 10]] |
| No_PCA       | RUS     | SVM           | 2    | 0,772727273 | 0,768421053 | [[10 1] [ 4 7]]  |
| No_PCA       | RUS     | SVM           | 3    | 0,772727273 | 0,772256729 | [[8 3] [2 9]]    |
| No_PCA       | RUS     | SVM           | 4    | 0,772727273 | 0,772256729 | [[9 2] [3 8]]    |
| No_PCA       | RUS     | SVM           | 5    | 0,727272727 | 0,725       | [[7 4] [2 9]]    |
| No_PCA       | RUS     | SVM           | 6    | 0,681818182 | 0,675789474 | [[6 5] [2 9]]    |
| No_PCA       | RUS     | SVM           | 7    | 0,857142857 | 0,856489049 | [[10 1] [ 2 8]]  |
| No_PCA       | RUS     | SVM           | 8    | 0,80952381  | 0,8003663   | [[11 0] [ 4 6]]  |
| No_PCA       | RUS     | SVM           | 9    | 0,666666667 | 0,666666667 | [[7 3] [4 7]]    |
| No_PCA       | RUS     | SVM           | 10   | 0,857142857 | 0,856489049 | [[ 8 2] [ 1 10]] |
| No_PCA       | RUS     | XGBoost       | 1    | 0,909090909 | 0,908333333 | [[ 9 2] [ 0 11]] |
| No_PCA       | RUS     | XGBoost       | 2    | 0,636363636 | 0,633333333 | [[8 3] [5 6]]    |
| No_PCA       | RUS     | XGBoost       | 3    | 0,818181818 | 0,816666667 | [[ 8 3] [ 1 10]] |
| No_PCA       | RUS     | XGBoost       | 4    | 0,727272727 | 0,725       | [[7 4] [2 9]]    |
| No_PCA       | RUS     | XGBoost       | 5    | 0,681818182 | 0,675789474 | [[6 5] [2 9]]    |
| No_PCA       | RUS     | XGBoost       | 6    | 0,727272727 | 0,725       | [[7 4] [2 9]]    |
| No_PCA       | RUS     | XGBoost       | 7    | 0,714285714 | 0,714285714 | [[8 3] [3 7]]    |
| No_PCA       | RUS     | XGBoost       | 8    | 0,761904762 | 0,760815081 | [[9 2] [3 7]]    |
| No_PCA       | RUS     | XGBoost       | 9    | 0,80952381  | 0,80952381  | [[8 2] [2 9]]    |
| No_PCA       | RUS     | XGBoost       | 10   | 0,761904762 | 0,761904762 | [[8 2] [3 8]]    |

| PCA Approach | Sampler       | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix  |
|--------------|---------------|---------------|------|-------------|-------------|-------------------|
| No_PCA       | RUS%5 + SMOTE | Random Forest | 1    | 0,882352941 | 0,881439532 | [[25 1] [ 5 20]]  |
| No_PCA       | RUS%5 + SMOTE | Random Forest | 2    | 0,843137255 | 0,843016591 | [[21 5] [ 3 22]]  |
| No_PCA       | RUS%5 + SMOTE | Random Forest | 3    | 0,843137255 | 0,843137255 | [[22 4] [ 4 21]]  |
| No_PCA       | RUS%5 + SMOTE | Random Forest | 4    | 0,823529412 | 0,823393508 | [[22 4] [ 5 20]]  |
| No_PCA       | RUS%5 + SMOTE | Random Forest | 5    | 0,882352941 | 0,88208061  | [[21 4] [ 2 24]]  |
| No_PCA       | RUS%5 + SMOTE | Random Forest | 6    | 0,960784314 | 0,960784314 | [[24 1] [ 1 25]]  |
| No_PCA       | RUS%5 + SMOTE | Random Forest | 7    | 0,745098039 | 0,744901734 | [[18 7] [ 6 20]]  |
| No_PCA       | RUS%5 + SMOTE | Random Forest | 8    | 0,862745098 | 0,861683165 | [[24 1] [ 6 20]]  |
| No_PCA       | RUS%5 + SMOTE | Random Forest | 9    | 0,9         | 0,899959984 | [[22 3] [ 2 23]]  |
| No_PCA       | RUS%5 + SMOTE | Random Forest | 10   | 0,9         | 0,899959984 | [[22 3] [ 2 23]]  |
| No_PCA       | RUS%5 + SMOTE | Decision Tree | 1    | 0,862745098 | 0,862107938 | [[24 2] [ 5 20]]  |
| No_PCA       | RUS%5 + SMOTE | Decision Tree | 2    | 0,764705882 | 0,761965656 | [[17 9] [ 3 22]]  |
| No_PCA       | RUS%5 + SMOTE | Decision Tree | 3    | 0,784313725 | 0,784147621 | [[21 5] [ 6 19]]  |
| No_PCA       | RUS%5 + SMOTE | Decision Tree | 4    | 0,823529412 | 0,823529412 | [[21 5] [ 4 21]]  |
| No_PCA       | RUS%5 + SMOTE | Decision Tree | 5    | 0,764705882 | 0,76416122  | [[18 7] [ 5 21]]  |
| No_PCA       | RUS%5 + SMOTE | Decision Tree | 6    | 0,843137255 | 0,842411038 | [[23 2] [ 6 20]]  |
| No_PCA       | RUS%5 + SMOTE | Decision Tree | 7    | 0,745098039 | 0,744901734 | [[18 7] [ 6 20]]  |
| No_PCA       | RUS%5 + SMOTE | Decision Tree | 8    | 0,803921569 | 0,803770739 | [[21 4] [ 6 20]]  |
| No_PCA       | RUS%5 + SMOTE | Decision Tree | 9    | 0,88        | 0,879807692 | [[21 4] [ 2 23]]  |
| No_PCA       | RUS%5 + SMOTE | Decision Tree | 10   | 0,8         | 0,8         | [[20 5] [ 5 20]]  |
| No_PCA       | RUS%5 + SMOTE | Naive Bayes   | 1    | 0,745098039 | 0,744901734 | [[20 6] [ 7 18]]  |
| No_PCA       | RUS%5 + SMOTE | Naive Bayes   | 2    | 0,784313725 | 0,78078202  | [[17 9] [ 2 23]]  |
| No_PCA       | RUS%5 + SMOTE | Naive Bayes   | 3    | 0,705882353 | 0,705882353 | [[18 8] [ 7 18]]  |
| No_PCA       | RUS%5 + SMOTE | Naive Bayes   | 4    | 0,705882353 | 0,705202836 | [[17 9] [ 6 19]]  |
| No_PCA       | RUS%5 + SMOTE | Naive Bayes   | 5    | 0,725490196 | 0,713314358 | [[13 12] [ 2 24]] |
| No_PCA       | RUS%5 + SMOTE | Naive Bayes   | 6    | 0,862745098 | 0,862745098 | [[22 3] [ 4 22]]  |
| No_PCA       | RUS%5 + SMOTE | Naive Bayes   | 7    | 0,725490196 | 0,725279035 | [[19 6] [ 8 18]]  |
| No_PCA       | RUS%5 + SMOTE | Naive Bayes   | 8    | 0,745098039 | 0,744509124 | [[20 5] [ 8 18]]  |
| No_PCA       | RUS%5 + SMOTE | Naive Bayes   | 9    | 0,82        | 0,819349659 | [[19 6] [ 3 22]]  |
| No_PCA       | RUS%5 + SMOTE | Naive Bayes   | 10   | 0,78        | 0,779911965 | [[20 5] [ 6 19]]  |
| No_PCA       | RUS%5 + SMOTE | SVM           | 1    | 0,784313725 | 0,781791079 | [[23 3] [ 8 17]]  |
| No_PCA       | RUS%5 + SMOTE | SVM           | 2    | 0,843137255 | 0,843016591 | [[21 5] [ 3 22]]  |
| No_PCA       | RUS%5 + SMOTE | SVM           | 3    | 0,705882353 | 0,705655847 | [[19 7] [ 8 17]]  |
| No_PCA       | RUS%5 + SMOTE | SVM           | 4    | 0,68627451  | 0,683838753 | [[20 6] [10 15]]  |
| No_PCA       | RUS%5 + SMOTE | SVM           | 5    | 0,784313725 | 0,783312474 | [[18 7] [ 4 22]]  |
| No_PCA       | RUS%5 + SMOTE | SVM           | 6    | 0,901960784 | 0,90120226  | [[25 0] [ 5 21]]  |
| No_PCA       | RUS%5 + SMOTE | SVM           | 7    | 0,784313725 | 0,783815413 | [[21 4] [ 7 19]]  |
| No_PCA       | RUS%5 + SMOTE | SVM           | 8    | 0,745098039 | 0,740924206 | [[22 3] [10 16]]  |
| No_PCA       | RUS%5 + SMOTE | SVM           | 9    | 0,88        | 0,88        | [[22 3] [ 3 22]]  |
| No_PCA       | RUS%5 + SMOTE | SVM           | 10   | 0,82        | 0,819927971 | [[21 4] [ 5 20]]  |
| No_PCA       | RUS%5 + SMOTE | XGBoost       | 1    | 0,784313725 | 0,781791079 | [[23 3] [ 8 17]]  |
| No_PCA       | RUS%5 + SMOTE | XGBoost       | 2    | 0,843137255 | 0,843016591 | [[21 5] [ 3 22]]  |
| No_PCA       | RUS%5 + SMOTE | XGBoost       | 3    | 0,862745098 | 0,862639395 | [[23 3] [ 4 21]]  |
| No_PCA       | RUS%5 + SMOTE | XGBoost       | 4    | 0,862745098 | 0,862107938 | [[24 2] [ 5 20]]  |
| No_PCA       | RUS%5 + SMOTE | XGBoost       | 5    | 0,882352941 | 0,882352941 | [[22 3] [ 3 23]]  |
| No_PCA       | RUS%5 + SMOTE | XGBoost       | 6    | 0,960784314 | 0,960784314 | [[24 1] [ 1 25]]  |
| No_PCA       | RUS%5 + SMOTE | XGBoost       | 7    | 0,784313725 | 0,784147621 | [[19 6] [ 5 21]]  |
| No_PCA       | RUS%5 + SMOTE | XGBoost       | 8    | 0,941176471 | 0,941176471 | [[24 1] [ 2 24]]  |
| No_PCA       | RUS%5 + SMOTE | XGBoost       | 9    | 0,94        | 0,93997599  | [[23 2] [ 1 24]]  |
| No_PCA       | RUS%5 + SMOTE | XGBoost       | 10   | 0,8         | 0,799679487 | [[21 4] [ 6 19]]  |

| PCA Approach | Sampler | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix  |
|--------------|---------|---------------|------|-------------|-------------|-------------------|
| PCA_90       | SMOTE   | Random Forest | 1    | 0,870588235 | 0,870229207 | [[39 3] [ 8 35]]  |
| PCA_90       | SMOTE   | Random Forest | 2    | 0,917647059 | 0,917418586 | [[41 1] [ 6 37]]  |
| PCA_90       | SMOTE   | Random Forest | 3    | 0,858823529 | 0,858588757 | [[38 4] [ 8 35]]  |
| PCA_90       | SMOTE   | Random Forest | 4    | 0,894117647 | 0,893499055 | [[41 1] [ 8 35]]  |
| PCA_90       | SMOTE   | Random Forest | 5    | 0,835294118 | 0,835294118 | [[35 7] [ 7 36]]  |
| PCA_90       | SMOTE   | Random Forest | 6    | 0,870588235 | 0,869576218 | [[41 2] [ 9 33]]  |
| PCA_90       | SMOTE   | Random Forest | 7    | 0,941176471 | 0,941078554 | [[42 1] [ 4 38]]  |
| PCA_90       | SMOTE   | Random Forest | 8    | 0,882352941 | 0,881664371 | [[41 2] [ 8 34]]  |
| PCA_90       | SMOTE   | Random Forest | 9    | 0,835294118 | 0,835157167 | [[37 6] [ 8 34]]  |
| PCA_90       | SMOTE   | Random Forest | 10   | 0,858823529 | 0,85663752  | [[42 1] [11 31]]  |
| PCA_90       | SMOTE   | Decision Tree | 1    | 0,788235294 | 0,788235294 | [[33 9] [ 9 34]]  |
| PCA_90       | SMOTE   | Decision Tree | 2    | 0,835294118 | 0,835020217 | [[37 5] [ 9 34]]  |
| PCA_90       | SMOTE   | Decision Tree | 3    | 0,823529412 | 0,822498425 | [[38 4] [11 32]]  |
| PCA_90       | SMOTE   | Decision Tree | 4    | 0,894117647 | 0,894029668 | [[39 3] [ 6 37]]  |
| PCA_90       | SMOTE   | Decision Tree | 5    | 0,752941176 | 0,752872749 | [[31 11] [10 33]] |
| PCA_90       | SMOTE   | Decision Tree | 6    | 0,811764706 | 0,811764706 | [[35 8] [ 8 34]]  |
| PCA_90       | SMOTE   | Decision Tree | 7    | 0,870588235 | 0,870588235 | [[37 6] [ 5 37]]  |
| PCA_90       | SMOTE   | Decision Tree | 8    | 0,858823529 | 0,858706143 | [[38 5] [ 7 35]]  |
| PCA_90       | SMOTE   | Decision Tree | 9    | 0,847058824 | 0,847058824 | [[36 7] [ 6 36]]  |
| PCA_90       | SMOTE   | Decision Tree | 10   | 0,811764706 | 0,808850026 | [[40 3] [13 29]]  |
| PCA_90       | SMOTE   | Naive Bayes   | 1    | 0,811764706 | 0,811764706 | [[34 8] [ 8 35]]  |
| PCA_90       | SMOTE   | Naive Bayes   | 2    | 0,694117647 | 0,69386331  | [[28 14] [12 31]] |
| PCA_90       | SMOTE   | Naive Bayes   | 3    | 0,717647059 | 0,714801396 | [[26 16] [ 8 35]] |
| PCA_90       | SMOTE   | Naive Bayes   | 4    | 0,741176471 | 0,740961263 | [[30 12] [10 33]] |
| PCA_90       | SMOTE   | Naive Bayes   | 5    | 0,729411765 | 0,729411765 | [[31 11] [12 31]] |
| PCA_90       | SMOTE   | Naive Bayes   | 6    | 0,729411765 | 0,72933682  | [[32 11] [12 30]] |
| PCA_90       | SMOTE   | Naive Bayes   | 7    | 0,847058824 | 0,84680424  | [[38 5] [ 8 34]]  |
| PCA_90       | SMOTE   | Naive Bayes   | 8    | 0,823529412 | 0,823480535 | [[36 7] [ 8 34]]  |
| PCA_90       | SMOTE   | Naive Bayes   | 9    | 0,764705882 | 0,764510239 | [[34 9] [11 31]]  |
| PCA_90       | SMOTE   | Naive Bayes   | 10   | 0,705882353 | 0,702167183 | [[35 8] [17 25]]  |
| PCA_90       | SMOTE   | SVM           | 1    | 0,835294118 | 0,834607843 | [[38 4] [10 33]]  |
| PCA_90       | SMOTE   | SVM           | 2    | 0,835294118 | 0,834607843 | [[38 4] [10 33]]  |
| PCA_90       | SMOTE   | SVM           | 3    | 0,741176471 | 0,741104814 | [[32 10] [12 31]] |
| PCA_90       | SMOTE   | SVM           | 4    | 0,752941176 | 0,752735893 | [[33 9] [12 31]]  |
| PCA_90       | SMOTE   | SVM           | 5    | 0,741176471 | 0,740746055 | [[33 9] [13 30]]  |
| PCA_90       | SMOTE   | SVM           | 6    | 0,729411765 | 0,728961347 | [[33 10] [13 29]] |
| PCA_90       | SMOTE   | SVM           | 7    | 0,847058824 | 0,846420594 | [[39 4] [ 9 33]]  |
| PCA_90       | SMOTE   | SVM           | 8    | 0,847058824 | 0,845862803 | [[40 3] [10 32]]  |
| PCA_90       | SMOTE   | SVM           | 9    | 0,764705882 | 0,764052288 | [[35 8] [12 30]]  |
| PCA_90       | SMOTE   | SVM           | 10   | 0,788235294 | 0,786995869 | [[37 6] [12 30]]  |
| PCA_90       | SMOTE   | XGBoost       | 1    | 0,858823529 | 0,858588757 | [[38 4] [ 8 35]]  |
| PCA_90       | SMOTE   | XGBoost       | 2    | 0,905882353 | 0,905490196 | [[41 1] [ 7 36]]  |
| PCA_90       | SMOTE   | XGBoost       | 3    | 0,882352941 | 0,881862745 | [[40 2] [ 8 35]]  |
| PCA_90       | SMOTE   | XGBoost       | 4    | 0,882352941 | 0,881862745 | [[40 2] [ 8 35]]  |
| PCA_90       | SMOTE   | XGBoost       | 5    | 0,870588235 | 0,870552392 | [[36 6] [ 5 38]]  |
| PCA_90       | SMOTE   | XGBoost       | 6    | 0,870588235 | 0,870048195 | [[40 3] [ 8 34]]  |
| PCA_90       | SMOTE   | XGBoost       | 7    | 0,905882353 | 0,905620915 | [[41 2] [ 6 36]]  |
| PCA_90       | SMOTE   | XGBoost       | 8    | 0,870588235 | 0,870048195 | [[40 3] [ 8 34]]  |
| PCA_90       | SMOTE   | XGBoost       | 9    | 0,870588235 | 0,870588235 | [[37 6] [ 5 37]]  |
| PCA_90       | SMOTE   | XGBoost       | 10   | 0,8         | 0,797473684 | [[39 4] [13 29]]  |

| PCA Approach | Sampler | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix  |
|--------------|---------|---------------|------|-------------|-------------|-------------------|
| PCA_90       | ROS     | Random Forest | 1    | 0,929411765 | 0,929117647 | [[42 0] [ 6 37]]  |
| PCA_90       | ROS     | Random Forest | 2    | 0,917647059 | 0,917165932 | [[42 0] [ 7 36]]  |
| PCA_90       | ROS     | Random Forest | 3    | 0,929411765 | 0,929117647 | [[42 0] [ 6 37]]  |
| PCA_90       | ROS     | Random Forest | 4    | 0,929411765 | 0,929117647 | [[42 0] [ 6 37]]  |
| PCA_90       | ROS     | Random Forest | 5    | 0,941176471 | 0,941127594 | [[41 1] [ 4 39]]  |
| PCA_90       | ROS     | Random Forest | 6    | 0,952941176 | 0,952810458 | [[43 0] [ 4 38]]  |
| PCA_90       | ROS     | Random Forest | 7    | 0,929411765 | 0,928998623 | [[43 0] [ 6 36]]  |
| PCA_90       | ROS     | Random Forest | 8    | 0,894117647 | 0,892780186 | [[43 0] [ 9 33]]  |
| PCA_90       | ROS     | Random Forest | 9    | 0,952941176 | 0,952810458 | [[43 0] [ 4 38]]  |
| PCA_90       | ROS     | Random Forest | 10   | 0,905882353 | 0,904933799 | [[43 0] [ 8 34]]  |
| PCA_90       | ROS     | Decision Tree | 1    | 0,894117647 | 0,893053057 | [[42 0] [ 9 34]]  |
| PCA_90       | ROS     | Decision Tree | 2    | 0,941176471 | 0,941013276 | [[42 0] [ 5 38]]  |
| PCA_90       | ROS     | Decision Tree | 3    | 0,917647059 | 0,917165932 | [[42 0] [ 7 36]]  |
| PCA_90       | ROS     | Decision Tree | 4    | 0,941176471 | 0,941013276 | [[42 0] [ 5 38]]  |
| PCA_90       | ROS     | Decision Tree | 5    | 0,941176471 | 0,941127594 | [[41 1] [ 4 39]]  |
| PCA_90       | ROS     | Decision Tree | 6    | 0,894117647 | 0,892780186 | [[43 0] [ 9 33]]  |
| PCA_90       | ROS     | Decision Tree | 7    | 0,882352941 | 0,881167249 | [[42 1] [ 9 33]]  |
| PCA_90       | ROS     | Decision Tree | 8    | 0,917647059 | 0,917003048 | [[43 0] [ 7 35]]  |
| PCA_90       | ROS     | Decision Tree | 9    | 0,929411765 | 0,928998623 | [[43 0] [ 6 36]]  |
| PCA_90       | ROS     | Decision Tree | 10   | 0,941176471 | 0,940930998 | [[43 0] [ 5 37]]  |
| PCA_90       | ROS     | Naive Bayes   | 1    | 0,811764706 | 0,811608191 | [[33 9] [ 7 36]]  |
| PCA_90       | ROS     | Naive Bayes   | 2    | 0,717647059 | 0,717412286 | [[29 13] [11 32]] |
| PCA_90       | ROS     | Naive Bayes   | 3    | 0,8         | 0,799165392 | [[31 11] [ 6 37]] |
| PCA_90       | ROS     | Naive Bayes   | 4    | 0,741176471 | 0,741176471 | [[31 11] [11 32]] |
| PCA_90       | ROS     | Naive Bayes   | 5    | 0,705882353 | 0,703582313 | [[26 16] [ 9 34]] |
| PCA_90       | ROS     | Naive Bayes   | 6    | 0,717647059 | 0,717568888 | [[30 13] [11 31]] |
| PCA_90       | ROS     | Naive Bayes   | 7    | 0,823529412 | 0,823235661 | [[37 6] [ 9 33]]  |
| PCA_90       | ROS     | Naive Bayes   | 8    | 0,729411765 | 0,72918693  | [[30 13] [10 32]] |
| PCA_90       | ROS     | Naive Bayes   | 9    | 0,705882353 | 0,702925158 | [[26 17] [ 8 34]] |
| PCA_90       | ROS     | Naive Bayes   | 10   | 0,670588235 | 0,670314334 | [[30 13] [15 27]] |
| PCA_90       | ROS     | SVM           | 1    | 0,823529412 | 0,822498425 | [[38 4] [11 32]]  |
| PCA_90       | ROS     | SVM           | 2    | 0,811764706 | 0,811451676 | [[36 6] [10 33]]  |
| PCA_90       | ROS     | SVM           | 3    | 0,788235294 | 0,787883136 | [[35 7] [11 32]]  |
| PCA_90       | ROS     | SVM           | 4    | 0,8         | 0,798831548 | [[37 5] [12 31]]  |
| PCA_90       | ROS     | SVM           | 5    | 0,682352941 | 0,682264962 | [[28 14] [13 30]] |
| PCA_90       | ROS     | SVM           | 6    | 0,752941176 | 0,749820433 | [[37 6] [15 27]]  |
| PCA_90       | ROS     | SVM           | 7    | 0,811764706 | 0,81124183  | [[37 6] [10 32]]  |
| PCA_90       | ROS     | SVM           | 8    | 0,8         | 0,799944606 | [[35 8] [ 9 33]]  |
| PCA_90       | ROS     | SVM           | 9    | 0,705882353 | 0,705392769 | [[32 11] [14 28]] |
| PCA_90       | ROS     | SVM           | 10   | 0,776470588 | 0,768772632 | [[41 2] [17 25]]  |
| PCA_90       | ROS     | XGBoost       | 1    | 0,917647059 | 0,917165932 | [[42 0] [ 7 36]]  |
| PCA_90       | ROS     | XGBoost       | 2    | 0,929411765 | 0,929117647 | [[42 0] [ 6 37]]  |
| PCA_90       | ROS     | XGBoost       | 3    | 0,941176471 | 0,941013276 | [[42 0] [ 5 38]]  |
| PCA_90       | ROS     | XGBoost       | 4    | 0,929411765 | 0,929117647 | [[42 0] [ 6 37]]  |
| PCA_90       | ROS     | XGBoost       | 5    | 0,976470588 | 0,976470588 | [[41 1] [ 1 42]]  |
| PCA_90       | ROS     | XGBoost       | 6    | 0,941176471 | 0,940930998 | [[43 0] [ 5 37]]  |
| PCA_90       | ROS     | XGBoost       | 7    | 0,905882353 | 0,905331497 | [[42 1] [ 7 35]]  |
| PCA_90       | ROS     | XGBoost       | 8    | 0,929411765 | 0,928998623 | [[43 0] [ 6 36]]  |
| PCA_90       | ROS     | XGBoost       | 9    | 0,941176471 | 0,940930998 | [[43 0] [ 5 37]]  |
| PCA_90       | ROS     | XGBoost       | 10   | 0,894117647 | 0,892780186 | [[43 0] [ 9 33]]  |

| PCA Approach | Sampler          | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix  |
|--------------|------------------|---------------|------|-------------|-------------|-------------------|
| PCA_90       | Borderline SMOTE | Random Forest | 1    | 0,905882353 | 0,905490196 | [[41 1] [ 7 36]]  |
| PCA_90       | Borderline SMOTE | Random Forest | 2    | 0,917647059 | 0,917418586 | [[41 1] [ 6 37]]  |
| PCA_90       | Borderline SMOTE | Random Forest | 3    | 0,882352941 | 0,882157298 | [[39 3] [ 7 36]]  |
| PCA_90       | Borderline SMOTE | Random Forest | 4    | 0,882352941 | 0,881862745 | [[40 2] [ 8 35]]  |
| PCA_90       | Borderline SMOTE | Random Forest | 5    | 0,905882353 | 0,905882353 | [[38 4] [ 4 39]]  |
| PCA_90       | Borderline SMOTE | Random Forest | 6    | 0,858823529 | 0,857400698 | [[41 2] [10 32]]  |
| PCA_90       | Borderline SMOTE | Random Forest | 7    | 0,905882353 | 0,905620915 | [[41 2] [ 6 36]]  |
| PCA_90       | Borderline SMOTE | Random Forest | 8    | 0,858823529 | 0,857400698 | [[41 2] [10 32]]  |
| PCA_90       | Borderline SMOTE | Random Forest | 9    | 0,870588235 | 0,870048195 | [[40 3] [ 8 34]]  |
| PCA_90       | Borderline SMOTE | Random Forest | 10   | 0,882352941 | 0,880531267 | [[43 0] [10 32]]  |
| PCA_90       | Borderline SMOTE | Decision Tree | 1    | 0,870588235 | 0,869832178 | [[40 2] [ 9 34]]  |
| PCA_90       | Borderline SMOTE | Decision Tree | 2    | 0,870588235 | 0,870229207 | [[39 3] [ 8 35]]  |
| PCA_90       | Borderline SMOTE | Decision Tree | 3    | 0,823529412 | 0,823235661 | [[33 9] [ 6 37]]  |
| PCA_90       | Borderline SMOTE | Decision Tree | 4    | 0,847058824 | 0,846931743 | [[37 5] [ 8 35]]  |
| PCA_90       | Borderline SMOTE | Decision Tree | 5    | 0,847058824 | 0,847016463 | [[35 7] [ 6 37]]  |
| PCA_90       | Borderline SMOTE | Decision Tree | 6    | 0,788235294 | 0,786995869 | [[37 6] [12 30]]  |
| PCA_90       | Borderline SMOTE | Decision Tree | 7    | 0,835294118 | 0,834836601 | [[38 5] [ 9 33]]  |
| PCA_90       | Borderline SMOTE | Decision Tree | 8    | 0,811764706 | 0,811608191 | [[36 7] [ 9 33]]  |
| PCA_90       | Borderline SMOTE | Decision Tree | 9    | 0,811764706 | 0,811712592 | [[34 9] [ 7 35]]  |
| PCA_90       | Borderline SMOTE | Decision Tree | 10   | 0,847058824 | 0,84680424  | [[38 5] [ 8 34]]  |
| PCA_90       | Borderline SMOTE | Naive Bayes   | 1    | 0,776470588 | 0,776098504 | [[31 11] [ 8 35]] |
| PCA_90       | Borderline SMOTE | Naive Bayes   | 2    | 0,788235294 | 0,787883136 | [[35 7] [11 32]]  |
| PCA_90       | Borderline SMOTE | Naive Bayes   | 3    | 0,8         | 0,799944606 | [[33 9] [ 8 35]]  |
| PCA_90       | Borderline SMOTE | Naive Bayes   | 4    | 0,788235294 | 0,787352941 | [[36 6] [12 31]]  |
| PCA_90       | Borderline SMOTE | Naive Bayes   | 5    | 0,752941176 | 0,752255759 | [[34 8] [13 30]]  |
| PCA_90       | Borderline SMOTE | Naive Bayes   | 6    | 0,694117647 | 0,694032962 | [[29 14] [12 30]] |
| PCA_90       | Borderline SMOTE | Naive Bayes   | 7    | 0,858823529 | 0,858431373 | [[39 4] [ 8 34]]  |
| PCA_90       | Borderline SMOTE | Naive Bayes   | 8    | 0,776470588 | 0,776408677 | [[34 9] [10 32]]  |
| PCA_90       | Borderline SMOTE | Naive Bayes   | 9    | 0,788235294 | 0,787647059 | [[36 7] [11 31]]  |
| PCA_90       | Borderline SMOTE | Naive Bayes   | 10   | 0,705882353 | 0,700399578 | [[36 7] [18 24]]  |
| PCA_90       | Borderline SMOTE | SVM           | 1    | 0,835294118 | 0,834607843 | [[38 4] [10 33]]  |
| PCA_90       | Borderline SMOTE | SVM           | 2    | 0,835294118 | 0,834607843 | [[38 4] [10 33]]  |
| PCA_90       | Borderline SMOTE | SVM           | 3    | 0,788235294 | 0,787352941 | [[36 6] [12 31]]  |
| PCA_90       | Borderline SMOTE | SVM           | 4    | 0,835294118 | 0,834008787 | [[39 3] [11 32]]  |
| PCA_90       | Borderline SMOTE | SVM           | 5    | 0,776470588 | 0,775164672 | [[36 6] [13 30]]  |
| PCA_90       | Borderline SMOTE | SVM           | 6    | 0,729411765 | 0,728282589 | [[34 9] [14 28]]  |
| PCA_90       | Borderline SMOTE | SVM           | 7    | 0,847058824 | 0,845862803 | [[40 3] [10 32]]  |
| PCA_90       | Borderline SMOTE | SVM           | 8    | 0,835294118 | 0,83433012  | [[39 4] [10 32]]  |
| PCA_90       | Borderline SMOTE | SVM           | 9    | 0,811764706 | 0,808850026 | [[40 3] [13 29]]  |
| PCA_90       | Borderline SMOTE | SVM           | 10   | 0,811764706 | 0,807603041 | [[41 2] [14 28]]  |
| PCA_90       | Borderline SMOTE | XGBoost       | 1    | 0,905882353 | 0,905490196 | [[41 1] [ 7 36]]  |
| PCA_90       | Borderline SMOTE | XGBoost       | 2    | 0,917647059 | 0,917418586 | [[41 1] [ 6 37]]  |
| PCA_90       | Borderline SMOTE | XGBoost       | 3    | 0,894117647 | 0,894029668 | [[39 3] [ 6 37]]  |
| PCA_90       | Borderline SMOTE | XGBoost       | 4    | 0,882352941 | 0,882157298 | [[39 3] [ 7 36]]  |
| PCA_90       | Borderline SMOTE | XGBoost       | 5    | 0,917647059 | 0,917509975 | [[37 5] [ 2 41]]  |
| PCA_90       | Borderline SMOTE | XGBoost       | 6    | 0,823529412 | 0,822792993 | [[38 5] [10 32]]  |
| PCA_90       | Borderline SMOTE | XGBoost       | 7    | 0,917647059 | 0,917303397 | [[42 1] [ 6 36]]  |
| PCA_90       | Borderline SMOTE | XGBoost       | 8    | 0,835294118 | 0,83433012  | [[39 4] [10 32]]  |
| PCA_90       | Borderline SMOTE | XGBoost       | 9    | 0,870588235 | 0,870372818 | [[39 4] [ 7 35]]  |
| PCA_90       | Borderline SMOTE | XGBoost       | 10   | 0,870588235 | 0,86895356  | [[42 1] [10 32]]  |

| PCA Approach | Sampler | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix |
|--------------|---------|---------------|------|-------------|-------------|------------------|
| PCA_90       | RUS     | Random Forest | 1    | 0,909090909 | 0,909090909 | [[10 1] [ 1 10]] |
| PCA_90       | RUS     | Random Forest | 2    | 0,772727273 | 0,768421053 | [[10 1] [ 4 7]]  |
| PCA_90       | RUS     | Random Forest | 3    | 0,818181818 | 0,816666667 | [[ 8 3] [ 1 10]] |
| PCA_90       | RUS     | Random Forest | 4    | 0,681818182 | 0,68115942  | [[8 3] [4 7]]    |
| PCA_90       | RUS     | Random Forest | 5    | 0,772727273 | 0,772256729 | [[8 3] [2 9]]    |
| PCA_90       | RUS     | Random Forest | 6    | 0,772727273 | 0,772256729 | [[8 3] [2 9]]    |
| PCA_90       | RUS     | Random Forest | 7    | 0,714285714 | 0,700549451 | [[10 1] [ 5 5]]  |
| PCA_90       | RUS     | Random Forest | 8    | 0,714285714 | 0,71031746  | [[9 2] [4 6]]    |
| PCA_90       | RUS     | Random Forest | 9    | 0,714285714 | 0,714285714 | [[7 3] [3 8]]    |
| PCA_90       | RUS     | Random Forest | 10   | 0,857142857 | 0,856489049 | [[ 8 2] [ 1 10]] |
| PCA_90       | RUS     | Decision Tree | 1    | 0,772727273 | 0,772256729 | [[8 3] [2 9]]    |
| PCA_90       | RUS     | Decision Tree | 2    | 0,636363636 | 0,636363636 | [[7 4] [4 7]]    |
| PCA_90       | RUS     | Decision Tree | 3    | 0,727272727 | 0,725       | [[7 4] [2 9]]    |
| PCA_90       | RUS     | Decision Tree | 4    | 0,727272727 | 0,727272727 | [[8 3] [3 8]]    |
| PCA_90       | RUS     | Decision Tree | 5    | 0,681818182 | 0,675789474 | [[6 5] [2 9]]    |
| PCA_90       | RUS     | Decision Tree | 6    | 0,727272727 | 0,717948718 | [[ 6 5] [ 1 10]] |
| PCA_90       | RUS     | Decision Tree | 7    | 0,714285714 | 0,700549451 | [[10 1] [ 5 5]]  |
| PCA_90       | RUS     | Decision Tree | 8    | 0,619047619 | 0,619047619 | [[7 4] [4 6]]    |
| PCA_90       | RUS     | Decision Tree | 9    | 0,761904762 | 0,761904762 | [[8 2] [3 8]]    |
| PCA_90       | RUS     | Decision Tree | 10   | 0,666666667 | 0,665141114 | [[6 4] [3 8]]    |
| PCA_90       | RUS     | Naive Bayes   | 1    | 0,772727273 | 0,772256729 | [[8 3] [2 9]]    |
| PCA_90       | RUS     | Naive Bayes   | 2    | 0,772727273 | 0,768421053 | [[10 1] [ 4 7]]  |
| PCA_90       | RUS     | Naive Bayes   | 3    | 0,727272727 | 0,725       | [[7 4] [2 9]]    |
| PCA_90       | RUS     | Naive Bayes   | 4    | 0,727272727 | 0,725       | [[9 2] [4 7]]    |
| PCA_90       | RUS     | Naive Bayes   | 5    | 0,818181818 | 0,818181818 | [[9 2] [2 9]]    |
| PCA_90       | RUS     | Naive Bayes   | 6    | 0,727272727 | 0,725       | [[7 4] [2 9]]    |
| PCA_90       | RUS     | Naive Bayes   | 7    | 0,714285714 | 0,714285714 | [[8 3] [3 7]]    |
| PCA_90       | RUS     | Naive Bayes   | 8    | 0,80952381  | 0,8003663   | [[11 0] [ 4 6]]  |
| PCA_90       | RUS     | Naive Bayes   | 9    | 0,714285714 | 0,714285714 | [[7 3] [3 8]]    |
| PCA_90       | RUS     | Naive Bayes   | 10   | 0,857142857 | 0,856489049 | [[ 8 2] [ 1 10]] |
| PCA_90       | RUS     | SVM           | 1    | 0,818181818 | 0,818181818 | [[9 2] [2 9]]    |
| PCA_90       | RUS     | SVM           | 2    | 0,772727273 | 0,768421053 | [[10 1] [ 4 7]]  |
| PCA_90       | RUS     | SVM           | 3    | 0,727272727 | 0,727272727 | [[8 3] [3 8]]    |
| PCA_90       | RUS     | SVM           | 4    | 0,772727273 | 0,772256729 | [[9 2] [3 8]]    |
| PCA_90       | RUS     | SVM           | 5    | 0,727272727 | 0,725       | [[7 4] [2 9]]    |
| PCA_90       | RUS     | SVM           | 6    | 0,681818182 | 0,675789474 | [[6 5] [2 9]]    |
| PCA_90       | RUS     | SVM           | 7    | 0,761904762 | 0,755182073 | [[10 1] [ 4 6]]  |
| PCA_90       | RUS     | SVM           | 8    | 0,80952381  | 0,8003663   | [[11 0] [ 4 6]]  |
| PCA_90       | RUS     | SVM           | 9    | 0,666666667 | 0,666666667 | [[7 3] [4 7]]    |
| PCA_90       | RUS     | SVM           | 10   | 0,80952381  | 0,80952381  | [[8 2] [2 9]]    |
| PCA_90       | RUS     | XGBoost       | 1    | 0,863636364 | 0,863354037 | [[ 9 2] [ 1 10]] |
| PCA_90       | RUS     | XGBoost       | 2    | 0,863636364 | 0,863354037 | [[10 1] [ 2 9]]  |
| PCA_90       | RUS     | XGBoost       | 3    | 0,818181818 | 0,816666667 | [[ 8 3] [ 1 10]] |
| PCA_90       | RUS     | XGBoost       | 4    | 0,727272727 | 0,727272727 | [[8 3] [3 8]]    |
| PCA_90       | RUS     | XGBoost       | 5    | 0,772727273 | 0,768421053 | [[ 7 4] [ 1 10]] |
| PCA_90       | RUS     | XGBoost       | 6    | 0,772727273 | 0,772256729 | [[8 3] [2 9]]    |
| PCA_90       | RUS     | XGBoost       | 7    | 0,666666667 | 0,657254902 | [[9 2] [5 5]]    |
| PCA_90       | RUS     | XGBoost       | 8    | 0,761904762 | 0,760815081 | [[9 2] [3 7]]    |
| PCA_90       | RUS     | XGBoost       | 9    | 0,714285714 | 0,714285714 | [[7 3] [3 8]]    |
| PCA_90       | RUS     | XGBoost       | 10   | 0,761904762 | 0,760815081 | [[7 3] [2 9]]    |

| PCA Approach | Sampler       | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix  |
|--------------|---------------|---------------|------|-------------|-------------|-------------------|
| PCA_90       | RUS%5 + SMOTE | Random Forest | 1    | 0,803921569 | 0,800694572 | [[24 2] [ 8 17]]  |
| PCA_90       | RUS%5 + SMOTE | Random Forest | 2    | 0,843137255 | 0,842411038 | [[20 6] [ 2 23]]  |
| PCA_90       | RUS%5 + SMOTE | Random Forest | 3    | 0,803921569 | 0,803467683 | [[22 4] [ 6 19]]  |
| PCA_90       | RUS%5 + SMOTE | Random Forest | 4    | 0,843137255 | 0,841919376 | [[24 2] [ 6 19]]  |
| PCA_90       | RUS%5 + SMOTE | Random Forest | 5    | 0,843137255 | 0,842774147 | [[20 5] [ 3 23]]  |
| PCA_90       | RUS%5 + SMOTE | Random Forest | 6    | 0,843137255 | 0,842411038 | [[23 2] [ 6 20]]  |
| PCA_90       | RUS%5 + SMOTE | Random Forest | 7    | 0,784313725 | 0,782644973 | [[22 3] [ 8 18]]  |
| PCA_90       | RUS%5 + SMOTE | Random Forest | 8    | 0,843137255 | 0,842411038 | [[23 2] [ 6 20]]  |
| PCA_90       | RUS%5 + SMOTE | Random Forest | 9    | 0,82        | 0,819927971 | [[21 4] [ 5 20]]  |
| PCA_90       | RUS%5 + SMOTE | Random Forest | 10   | 0,86        | 0,859943978 | [[22 3] [ 4 21]]  |
| PCA_90       | RUS%5 + SMOTE | Decision Tree | 1    | 0,764705882 | 0,760833487 | [[23 3] [ 9 16]]  |
| PCA_90       | RUS%5 + SMOTE | Decision Tree | 2    | 0,745098039 | 0,743125877 | [[17 9] [ 4 21]]  |
| PCA_90       | RUS%5 + SMOTE | Decision Tree | 3    | 0,705882353 | 0,705882353 | [[18 8] [ 7 18]]  |
| PCA_90       | RUS%5 + SMOTE | Decision Tree | 4    | 0,764705882 | 0,760833487 | [[23 3] [ 9 16]]  |
| PCA_90       | RUS%5 + SMOTE | Decision Tree | 5    | 0,784313725 | 0,781791079 | [[17 8] [ 3 23]]  |
| PCA_90       | RUS%5 + SMOTE | Decision Tree | 6    | 0,843137255 | 0,843016591 | [[22 3] [ 5 21]]  |
| PCA_90       | RUS%5 + SMOTE | Decision Tree | 7    | 0,764705882 | 0,763616558 | [[21 4] [ 8 18]]  |
| PCA_90       | RUS%5 + SMOTE | Decision Tree | 8    | 0,843137255 | 0,842411038 | [[23 2] [ 6 20]]  |
| PCA_90       | RUS%5 + SMOTE | Decision Tree | 9    | 0,8         | 0,798711755 | [[18 7] [ 3 22]]  |
| PCA_90       | RUS%5 + SMOTE | Decision Tree | 10   | 0,84        | 0,84        | [[21 4] [ 4 21]]  |
| PCA_90       | RUS%5 + SMOTE | Naive Bayes   | 1    | 0,745098039 | 0,74211673  | [[22 4] [ 9 16]]  |
| PCA_90       | RUS%5 + SMOTE | Naive Bayes   | 2    | 0,862745098 | 0,86242799  | [[21 5] [ 2 23]]  |
| PCA_90       | RUS%5 + SMOTE | Naive Bayes   | 3    | 0,666666667 | 0,66640996  | [[18 8] [ 9 16]]  |
| PCA_90       | RUS%5 + SMOTE | Naive Bayes   | 4    | 0,705882353 | 0,705882353 | [[18 8] [ 7 18]]  |
| PCA_90       | RUS%5 + SMOTE | Naive Bayes   | 5    | 0,705882353 | 0,699390201 | [[14 11] [ 4 22]] |
| PCA_90       | RUS%5 + SMOTE | Naive Bayes   | 6    | 0,784313725 | 0,783815413 | [[21 4] [ 7 19]]  |
| PCA_90       | RUS%5 + SMOTE | Naive Bayes   | 7    | 0,68627451  | 0,686033183 | [[18 7] [ 9 17]]  |
| PCA_90       | RUS%5 + SMOTE | Naive Bayes   | 8    | 0,745098039 | 0,740924206 | [[22 3] [10 16]]  |
| PCA_90       | RUS%5 + SMOTE | Naive Bayes   | 9    | 0,78        | 0,779911965 | [[19 6] [ 5 20]]  |
| PCA_90       | RUS%5 + SMOTE | Naive Bayes   | 10   | 0,8         | 0,799679487 | [[21 4] [ 6 19]]  |
| PCA_90       | RUS%5 + SMOTE | SVM           | 1    | 0,745098039 | 0,743914742 | [[21 5] [ 8 17]]  |
| PCA_90       | RUS%5 + SMOTE | SVM           | 2    | 0,823529412 | 0,823121702 | [[20 6] [ 3 22]]  |
| PCA_90       | RUS%5 + SMOTE | SVM           | 3    | 0,666666667 | 0,665119278 | [[19 7] [10 15]]  |
| PCA_90       | RUS%5 + SMOTE | SVM           | 4    | 0,725490196 | 0,723358909 | [[21 5] [ 9 16]]  |
| PCA_90       | RUS%5 + SMOTE | SVM           | 5    | 0,745098039 | 0,74211673  | [[16 9] [ 4 22]]  |
| PCA_90       | RUS%5 + SMOTE | SVM           | 6    | 0,862745098 | 0,861683165 | [[24 1] [ 6 20]]  |
| PCA_90       | RUS%5 + SMOTE | SVM           | 7    | 0,764705882 | 0,764524887 | [[20 5] [ 7 19]]  |
| PCA_90       | RUS%5 + SMOTE | SVM           | 8    | 0,725490196 | 0,722293265 | [[21 4] [10 16]]  |
| PCA_90       | RUS%5 + SMOTE | SVM           | 9    | 0,82        | 0,819349659 | [[22 3] [ 6 19]]  |
| PCA_90       | RUS%5 + SMOTE | SVM           | 10   | 0,8         | 0,799679487 | [[21 4] [ 6 19]]  |
| PCA_90       | RUS%5 + SMOTE | XGBoost       | 1    | 0,843137255 | 0,840555658 | [[25 1] [ 7 18]]  |
| PCA_90       | RUS%5 + SMOTE | XGBoost       | 2    | 0,784313725 | 0,78078202  | [[17 9] [ 2 23]]  |
| PCA_90       | RUS%5 + SMOTE | XGBoost       | 3    | 0,823529412 | 0,823393508 | [[22 4] [ 5 20]]  |
| PCA_90       | RUS%5 + SMOTE | XGBoost       | 4    | 0,803921569 | 0,800694572 | [[24 2] [ 8 17]]  |
| PCA_90       | RUS%5 + SMOTE | XGBoost       | 5    | 0,823529412 | 0,822710206 | [[19 6] [ 3 23]]  |
| PCA_90       | RUS%5 + SMOTE | XGBoost       | 6    | 0,901960784 | 0,901734279 | [[24 1] [ 4 22]]  |
| PCA_90       | RUS%5 + SMOTE | XGBoost       | 7    | 0,803921569 | 0,803013798 | [[22 3] [ 7 19]]  |
| PCA_90       | RUS%5 + SMOTE | XGBoost       | 8    | 0,823529412 | 0,822164069 | [[23 2] [ 7 19]]  |
| PCA_90       | RUS%5 + SMOTE | XGBoost       | 9    | 0,84        | 0,84        | [[21 4] [ 4 21]]  |
| PCA_90       | RUS%5 + SMOTE | XGBoost       | 10   | 0,86        | 0,859943978 | [[22 3] [ 4 21]]  |

| PCA Approach | Sampler | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix  |
|--------------|---------|---------------|------|-------------|-------------|-------------------|
| PCA_95       | SMOTE   | Random Forest | 1    | 0,870588235 | 0,870229207 | [[39 3] [ 8 35]]  |
| PCA_95       | SMOTE   | Random Forest | 2    | 0,917647059 | 0,917418586 | [[41 1] [ 6 37]]  |
| PCA_95       | SMOTE   | Random Forest | 3    | 0,882352941 | 0,88232037  | [[38 4] [ 6 37]]  |
| PCA_95       | SMOTE   | Random Forest | 4    | 0,905882353 | 0,905490196 | [[41 1] [ 7 36]]  |
| PCA_95       | SMOTE   | Random Forest | 5    | 0,870588235 | 0,870552392 | [[36 6] [ 5 38]]  |
| PCA_95       | SMOTE   | Random Forest | 6    | 0,894117647 | 0,893941397 | [[40 3] [ 6 36]]  |
| PCA_95       | SMOTE   | Random Forest | 7    | 0,905882353 | 0,905620915 | [[41 2] [ 6 36]]  |
| PCA_95       | SMOTE   | Random Forest | 8    | 0,894117647 | 0,893675796 | [[41 2] [ 7 35]]  |
| PCA_95       | SMOTE   | Random Forest | 9    | 0,858823529 | 0,858431373 | [[39 4] [ 8 34]]  |
| PCA_95       | SMOTE   | Random Forest | 10   | 0,858823529 | 0,857400698 | [[41 2] [10 32]]  |
| PCA_95       | SMOTE   | Decision Tree | 1    | 0,835294118 | 0,834607843 | [[38 4] [10 33]]  |
| PCA_95       | SMOTE   | Decision Tree | 2    | 0,835294118 | 0,835157167 | [[34 8] [ 6 37]]  |
| PCA_95       | SMOTE   | Decision Tree | 3    | 0,8         | 0,799445138 | [[36 6] [11 32]]  |
| PCA_95       | SMOTE   | Decision Tree | 4    | 0,823529412 | 0,823529412 | [[35 7] [ 8 35]]  |
| PCA_95       | SMOTE   | Decision Tree | 5    | 0,823529412 | 0,823235661 | [[33 9] [ 6 37]]  |
| PCA_95       | SMOTE   | Decision Tree | 6    | 0,764705882 | 0,76464074  | [[32 11] [ 9 33]] |
| PCA_95       | SMOTE   | Decision Tree | 7    | 0,717647059 | 0,717568888 | [[30 13] [11 31]] |
| PCA_95       | SMOTE   | Decision Tree | 8    | 0,8         | 0,799445138 | [[32 11] [ 6 36]] |
| PCA_95       | SMOTE   | Decision Tree | 9    | 0,741176471 | 0,740961263 | [[33 10] [12 30]] |
| PCA_95       | SMOTE   | Decision Tree | 10   | 0,858823529 | 0,857997246 | [[40 3] [ 9 33]]  |
| PCA_95       | SMOTE   | Naive Bayes   | 1    | 0,823529412 | 0,822498425 | [[38 4] [11 32]]  |
| PCA_95       | SMOTE   | Naive Bayes   | 2    | 0,729411765 | 0,728661069 | [[33 9] [14 29]]  |
| PCA_95       | SMOTE   | Naive Bayes   | 3    | 0,694117647 | 0,69386331  | [[28 14] [12 31]] |
| PCA_95       | SMOTE   | Naive Bayes   | 4    | 0,741176471 | 0,740098039 | [[34 8] [14 29]]  |
| PCA_95       | SMOTE   | Naive Bayes   | 5    | 0,717647059 | 0,715443636 | [[34 8] [16 27]]  |
| PCA_95       | SMOTE   | Naive Bayes   | 6    | 0,764705882 | 0,764052288 | [[35 8] [12 30]]  |
| PCA_95       | SMOTE   | Naive Bayes   | 7    | 0,788235294 | 0,781882353 | [[41 2] [16 26]]  |
| PCA_95       | SMOTE   | Naive Bayes   | 8    | 0,741176471 | 0,739661617 | [[35 8] [14 28]]  |
| PCA_95       | SMOTE   | Naive Bayes   | 9    | 0,717647059 | 0,716862745 | [[33 10] [14 28]] |
| PCA_95       | SMOTE   | Naive Bayes   | 10   | 0,658823529 | 0,647074017 | [[36 7] [22 20]]  |
| PCA_95       | SMOTE   | SVM           | 1    | 0,823529412 | 0,822498425 | [[38 4] [11 32]]  |
| PCA_95       | SMOTE   | SVM           | 2    | 0,858823529 | 0,858235294 | [[39 3] [ 9 34]]  |
| PCA_95       | SMOTE   | SVM           | 3    | 0,764705882 | 0,764705882 | [[32 10] [10 33]] |
| PCA_95       | SMOTE   | SVM           | 4    | 0,776470588 | 0,776408677 | [[32 10] [ 9 34]] |
| PCA_95       | SMOTE   | SVM           | 5    | 0,776470588 | 0,775850448 | [[35 7] [12 31]]  |
| PCA_95       | SMOTE   | SVM           | 6    | 0,764705882 | 0,764705882 | [[33 10] [10 32]] |
| PCA_95       | SMOTE   | SVM           | 7    | 0,858823529 | 0,857997246 | [[40 3] [ 9 33]]  |
| PCA_95       | SMOTE   | SVM           | 8    | 0,858823529 | 0,857997246 | [[40 3] [ 9 33]]  |
| PCA_95       | SMOTE   | SVM           | 9    | 0,8         | 0,799667083 | [[36 7] [10 32]]  |
| PCA_95       | SMOTE   | SVM           | 10   | 0,764705882 | 0,763328743 | [[36 7] [13 29]]  |
| PCA_95       | SMOTE   | XGBoost       | 1    | 0,882352941 | 0,881862745 | [[40 2] [ 8 35]]  |
| PCA_95       | SMOTE   | XGBoost       | 2    | 0,882352941 | 0,881862745 | [[40 2] [ 8 35]]  |
| PCA_95       | SMOTE   | XGBoost       | 3    | 0,858823529 | 0,858823529 | [[36 6] [ 6 37]]  |
| PCA_95       | SMOTE   | XGBoost       | 4    | 0,894117647 | 0,893823897 | [[40 2] [ 7 36]]  |
| PCA_95       | SMOTE   | XGBoost       | 5    | 0,894117647 | 0,893941397 | [[36 6] [ 3 40]]  |
| PCA_95       | SMOTE   | XGBoost       | 6    | 0,847058824 | 0,845862803 | [[40 3] [10 32]]  |
| PCA_95       | SMOTE   | XGBoost       | 7    | 0,905882353 | 0,905620915 | [[41 2] [ 6 36]]  |
| PCA_95       | SMOTE   | XGBoost       | 8    | 0,929411765 | 0,929215686 | [[42 1] [ 5 37]]  |
| PCA_95       | SMOTE   | XGBoost       | 9    | 0,823529412 | 0,822792993 | [[38 5] [10 32]]  |
| PCA_95       | SMOTE   | XGBoost       | 10   | 0,847058824 | 0,845862803 | [[40 3] [10 32]]  |

| PCA Approach | Sampler | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix  |
|--------------|---------|---------------|------|-------------|-------------|-------------------|
| PCA_95       | ROS     | Random Forest | 1    | 0,905882353 | 0,905147879 | [[42 0] [ 8 35]]  |
| PCA_95       | ROS     | Random Forest | 2    | 0,941176471 | 0,941013276 | [[42 0] [ 5 38]]  |
| PCA_95       | ROS     | Random Forest | 3    | 0,941176471 | 0,941013276 | [[42 0] [ 5 38]]  |
| PCA_95       | ROS     | Random Forest | 4    | 0,929411765 | 0,929117647 | [[42 0] [ 6 37]]  |
| PCA_95       | ROS     | Random Forest | 5    | 0,964705882 | 0,964705882 | [[41 1] [ 2 41]]  |
| PCA_95       | ROS     | Random Forest | 6    | 0,929411765 | 0,928998623 | [[43 0] [ 6 36]]  |
| PCA_95       | ROS     | Random Forest | 7    | 0,952941176 | 0,952810458 | [[43 0] [ 4 38]]  |
| PCA_95       | ROS     | Random Forest | 8    | 0,952941176 | 0,952810458 | [[43 0] [ 4 38]]  |
| PCA_95       | ROS     | Random Forest | 9    | 0,941176471 | 0,940930998 | [[43 0] [ 5 37]]  |
| PCA_95       | ROS     | Random Forest | 10   | 0,917647059 | 0,917003048 | [[43 0] [ 7 35]]  |
| PCA_95       | ROS     | Decision Tree | 1    | 0,941176471 | 0,941013276 | [[42 0] [ 5 38]]  |
| PCA_95       | ROS     | Decision Tree | 2    | 0,894117647 | 0,893053057 | [[42 0] [ 9 34]]  |
| PCA_95       | ROS     | Decision Tree | 3    | 0,894117647 | 0,893053057 | [[42 0] [ 9 34]]  |
| PCA_95       | ROS     | Decision Tree | 4    | 0,847058824 | 0,843689409 | [[42 0] [13 30]]  |
| PCA_95       | ROS     | Decision Tree | 5    | 0,952941176 | 0,952928148 | [[41 1] [ 3 40]]  |
| PCA_95       | ROS     | Decision Tree | 6    | 0,905882353 | 0,905331497 | [[42 1] [ 7 35]]  |
| PCA_95       | ROS     | Decision Tree | 7    | 0,882352941 | 0,881167249 | [[42 1] [ 9 33]]  |
| PCA_95       | ROS     | Decision Tree | 8    | 0,941176471 | 0,940930998 | [[43 0] [ 5 37]]  |
| PCA_95       | ROS     | Decision Tree | 9    | 0,894117647 | 0,892780186 | [[43 0] [ 9 33]]  |
| PCA_95       | ROS     | Decision Tree | 10   | 0,894117647 | 0,892780186 | [[43 0] [ 9 33]]  |
| PCA_95       | ROS     | Naive Bayes   | 1    | 0,823529412 | 0,823382781 | [[36 6] [ 9 34]]  |
| PCA_95       | ROS     | Naive Bayes   | 2    | 0,729411765 | 0,72933682  | [[30 12] [11 32]] |
| PCA_95       | ROS     | Naive Bayes   | 3    | 0,788235294 | 0,788235294 | [[33 9] [ 9 34]]  |
| PCA_95       | ROS     | Naive Bayes   | 4    | 0,764705882 | 0,76464074  | [[33 9] [11 32]]  |
| PCA_95       | ROS     | Naive Bayes   | 5    | 0,705882353 | 0,704654988 | [[27 15] [10 33]] |
| PCA_95       | ROS     | Naive Bayes   | 6    | 0,752941176 | 0,752941176 | [[32 11] [10 32]] |
| PCA_95       | ROS     | Naive Bayes   | 7    | 0,858823529 | 0,857997246 | [[40 3] [ 9 33]]  |
| PCA_95       | ROS     | Naive Bayes   | 8    | 0,764705882 | 0,764510239 | [[34 9] [11 31]]  |
| PCA_95       | ROS     | Naive Bayes   | 9    | 0,705882353 | 0,705066379 | [[28 15] [10 32]] |
| PCA_95       | ROS     | Naive Bayes   | 10   | 0,705882353 | 0,702167183 | [[35 8] [17 25]]  |
| PCA_95       | ROS     | SVM           | 1    | 0,835294118 | 0,834008787 | [[39 3] [11 32]]  |
| PCA_95       | ROS     | SVM           | 2    | 0,823529412 | 0,823039828 | [[37 5] [10 33]]  |
| PCA_95       | ROS     | SVM           | 3    | 0,8         | 0,799445138 | [[36 6] [11 32]]  |
| PCA_95       | ROS     | SVM           | 4    | 0,811764706 | 0,810295757 | [[38 4] [12 31]]  |
| PCA_95       | ROS     | SVM           | 5    | 0,705882353 | 0,705882353 | [[30 12] [13 30]] |
| PCA_95       | ROS     | SVM           | 6    | 0,764705882 | 0,763328743 | [[36 7] [13 29]]  |
| PCA_95       | ROS     | SVM           | 7    | 0,847058824 | 0,846420594 | [[39 4] [ 9 33]]  |
| PCA_95       | ROS     | SVM           | 8    | 0,788235294 | 0,788059215 | [[35 8] [10 32]]  |
| PCA_95       | ROS     | SVM           | 9    | 0,717647059 | 0,716862745 | [[33 10] [14 28]] |
| PCA_95       | ROS     | SVM           | 10   | 0,764705882 | 0,755478662 | [[41 2] [18 24]]  |
| PCA_95       | ROS     | XGBoost       | 1    | 0,905882353 | 0,905147879 | [[42 0] [ 8 35]]  |
| PCA_95       | ROS     | XGBoost       | 2    | 0,941176471 | 0,941013276 | [[42 0] [ 5 38]]  |
| PCA_95       | ROS     | XGBoost       | 3    | 0,917647059 | 0,917165932 | [[42 0] [ 7 36]]  |
| PCA_95       | ROS     | XGBoost       | 4    | 0,941176471 | 0,941013276 | [[42 0] [ 5 38]]  |
| PCA_95       | ROS     | XGBoost       | 5    | 0,964705882 | 0,964705882 | [[41 1] [ 2 41]]  |
| PCA_95       | ROS     | XGBoost       | 6    | 0,929411765 | 0,928998623 | [[43 0] [ 6 36]]  |
| PCA_95       | ROS     | XGBoost       | 7    | 0,905882353 | 0,904933799 | [[43 0] [ 8 34]]  |
| PCA_95       | ROS     | XGBoost       | 8    | 0,964705882 | 0,964647132 | [[43 0] [ 3 39]]  |
| PCA_95       | ROS     | XGBoost       | 9    | 0,952941176 | 0,952810458 | [[43 0] [ 4 38]]  |
| PCA_95       | ROS     | XGBoost       | 10   | 0,905882353 | 0,904933799 | [[43 0] [ 8 34]]  |

| PCA Approach | Sampler          | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix |
|--------------|------------------|---------------|------|-------------|-------------|------------------|
| PCA_95       | Borderline SMOTE | Random Forest | 1    | 0,894117647 | 0,893823897 | [[40 2] [ 7 36]] |
| PCA_95       | Borderline SMOTE | Random Forest | 2    | 0,917647059 | 0,917165932 | [[42 0] [ 7 36]] |
| PCA_95       | Borderline SMOTE | Random Forest | 3    | 0,905882353 | 0,905856296 | [[39 3] [ 5 38]] |
| PCA_95       | Borderline SMOTE | Random Forest | 4    | 0,894117647 | 0,893823897 | [[40 2] [ 7 36]] |
| PCA_95       | Borderline SMOTE | Random Forest | 5    | 0,882352941 | 0,882157298 | [[39 3] [ 7 36]] |
| PCA_95       | Borderline SMOTE | Random Forest | 6    | 0,905882353 | 0,905804095 | [[40 3] [ 5 37]] |
| PCA_95       | Borderline SMOTE | Random Forest | 7    | 0,905882353 | 0,905620915 | [[41 2] [ 6 36]] |
| PCA_95       | Borderline SMOTE | Random Forest | 8    | 0,870588235 | 0,869576218 | [[41 2] [ 9 33]] |
| PCA_95       | Borderline SMOTE | Random Forest | 9    | 0,870588235 | 0,870048195 | [[40 3] [ 8 34]] |
| PCA_95       | Borderline SMOTE | Random Forest | 10   | 0,894117647 | 0,892780186 | [[43 0] [ 9 33]] |
| PCA_95       | Borderline SMOTE | Decision Tree | 1    | 0,847058824 | 0,846931743 | [[37 5] [ 8 35]] |
| PCA_95       | Borderline SMOTE | Decision Tree | 2    | 0,8         | 0,799445138 | [[36 6] [11 32]] |
| PCA_95       | Borderline SMOTE | Decision Tree | 3    | 0,847058824 | 0,846634517 | [[38 4] [ 9 34]] |
| PCA_95       | Borderline SMOTE | Decision Tree | 4    | 0,835294118 | 0,834836601 | [[33 9] [ 5 38]] |
| PCA_95       | Borderline SMOTE | Decision Tree | 5    | 0,823529412 | 0,823382781 | [[36 6] [ 9 34]] |
| PCA_95       | Borderline SMOTE | Decision Tree | 6    | 0,882352941 | 0,88232037  | [[37 6] [ 4 38]] |
| PCA_95       | Borderline SMOTE | Decision Tree | 7    | 0,870588235 | 0,870552392 | [[38 5] [ 6 36]] |
| PCA_95       | Borderline SMOTE | Decision Tree | 8    | 0,858823529 | 0,858431373 | [[39 4] [ 8 34]] |
| PCA_95       | Borderline SMOTE | Decision Tree | 9    | 0,788235294 | 0,786995869 | [[37 6] [12 30]] |
| PCA_95       | Borderline SMOTE | Decision Tree | 10   | 0,870588235 | 0,86895356  | [[42 1] [10 32]] |
| PCA_95       | Borderline SMOTE | Naive Bayes   | 1    | 0,811764706 | 0,810295757 | [[38 4] [12 31]] |
| PCA_95       | Borderline SMOTE | Naive Bayes   | 2    | 0,776470588 | 0,77422312  | [[37 5] [14 29]] |
| PCA_95       | Borderline SMOTE | Naive Bayes   | 3    | 0,835294118 | 0,833219156 | [[40 2] [12 31]] |
| PCA_95       | Borderline SMOTE | Naive Bayes   | 4    | 0,8         | 0,796912281 | [[39 3] [14 29]] |
| PCA_95       | Borderline SMOTE | Naive Bayes   | 5    | 0,752941176 | 0,750457133 | [[36 6] [15 28]] |
| PCA_95       | Borderline SMOTE | Naive Bayes   | 6    | 0,788235294 | 0,787647059 | [[36 7] [11 31]] |
| PCA_95       | Borderline SMOTE | Naive Bayes   | 7    | 0,776470588 | 0,772303679 | [[39 4] [15 27]] |
| PCA_95       | Borderline SMOTE | Naive Bayes   | 8    | 0,776470588 | 0,774722558 | [[37 6] [13 29]] |
| PCA_95       | Borderline SMOTE | Naive Bayes   | 9    | 0,741176471 | 0,737168786 | [[37 6] [16 26]] |
| PCA_95       | Borderline SMOTE | Naive Bayes   | 10   | 0,670588235 | 0,657670127 | [[37 6] [22 20]] |
| PCA_95       | Borderline SMOTE | SVM           | 1    | 0,858823529 | 0,858588757 | [[38 4] [ 8 35]] |
| PCA_95       | Borderline SMOTE | SVM           | 2    | 0,858823529 | 0,858235294 | [[39 3] [ 9 34]] |
| PCA_95       | Borderline SMOTE | SVM           | 3    | 0,776470588 | 0,776284855 | [[34 8] [11 32]] |
| PCA_95       | Borderline SMOTE | SVM           | 4    | 0,835294118 | 0,834607843 | [[38 4] [10 33]] |
| PCA_95       | Borderline SMOTE | SVM           | 5    | 0,811764706 | 0,810295757 | [[38 4] [12 31]] |
| PCA_95       | Borderline SMOTE | SVM           | 6    | 0,788235294 | 0,786995869 | [[37 6] [12 30]] |
| PCA_95       | Borderline SMOTE | SVM           | 7    | 0,882352941 | 0,881167249 | [[42 1] [ 9 33]] |
| PCA_95       | Borderline SMOTE | SVM           | 8    | 0,811764706 | 0,810662994 | [[38 5] [11 31]] |
| PCA_95       | Borderline SMOTE | SVM           | 9    | 0,788235294 | 0,786995869 | [[37 6] [12 30]] |
| PCA_95       | Borderline SMOTE | SVM           | 10   | 0,811764706 | 0,80438293  | [[43 0] [16 26]] |
| PCA_95       | Borderline SMOTE | XGBoost       | 1    | 0,905882353 | 0,905725838 | [[40 2] [ 6 37]] |
| PCA_95       | Borderline SMOTE | XGBoost       | 2    | 0,905882353 | 0,905490196 | [[41 1] [ 7 36]] |
| PCA_95       | Borderline SMOTE | XGBoost       | 3    | 0,917647059 | 0,917418586 | [[41 1] [ 6 37]] |
| PCA_95       | Borderline SMOTE | XGBoost       | 4    | 0,894117647 | 0,894029668 | [[39 3] [ 6 37]] |
| PCA_95       | Borderline SMOTE | XGBoost       | 5    | 0,905882353 | 0,905882353 | [[38 4] [ 4 39]] |
| PCA_95       | Borderline SMOTE | XGBoost       | 6    | 0,917647059 | 0,91762425  | [[40 3] [ 4 38]] |
| PCA_95       | Borderline SMOTE | XGBoost       | 7    | 0,894117647 | 0,893289633 | [[42 1] [ 8 34]] |
| PCA_95       | Borderline SMOTE | XGBoost       | 8    | 0,917647059 | 0,917303397 | [[42 1] [ 6 36]] |
| PCA_95       | Borderline SMOTE | XGBoost       | 9    | 0,858823529 | 0,857997246 | [[40 3] [ 9 33]] |
| PCA_95       | Borderline SMOTE | XGBoost       | 10   | 0,894117647 | 0,892780186 | [[43 0] [ 9 33]] |

| PCA Approach | Sampler | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix |
|--------------|---------|---------------|------|-------------|-------------|------------------|
| PCA_95       | RUS     | Random Forest | 1    | 0,863636364 | 0,863354037 | [[ 9 2] [ 1 10]] |
| PCA_95       | RUS     | Random Forest | 2    | 0,818181818 | 0,811965812 | [[11 0] [ 4 7]]  |
| PCA_95       | RUS     | Random Forest | 3    | 0,772727273 | 0,772256729 | [[8 3] [2 9]]    |
| PCA_95       | RUS     | Random Forest | 4    | 0,727272727 | 0,727272727 | [[8 3] [3 8]]    |
| PCA_95       | RUS     | Random Forest | 5    | 0,772727273 | 0,772256729 | [[8 3] [2 9]]    |
| PCA_95       | RUS     | Random Forest | 6    | 0,727272727 | 0,727272727 | [[8 3] [3 8]]    |
| PCA_95       | RUS     | Random Forest | 7    | 0,619047619 | 0,600732601 | [[9 2] [6 4]]    |
| PCA_95       | RUS     | Random Forest | 8    | 0,761904762 | 0,755182073 | [[10 1] [ 4 6]]  |
| PCA_95       | RUS     | Random Forest | 9    | 0,761904762 | 0,760815081 | [[7 3] [2 9]]    |
| PCA_95       | RUS     | Random Forest | 10   | 0,80952381  | 0,80952381  | [[8 2] [2 9]]    |
| PCA_95       | RUS     | Decision Tree | 1    | 0,636363636 | 0,607142857 | [[ 4 7] [ 1 10]] |
| PCA_95       | RUS     | Decision Tree | 2    | 0,772727273 | 0,772256729 | [[9 2] [3 8]]    |
| PCA_95       | RUS     | Decision Tree | 3    | 0,727272727 | 0,717948718 | [[ 6 5] [ 1 10]] |
| PCA_95       | RUS     | Decision Tree | 4    | 0,590909091 | 0,590062112 | [[7 4] [5 6]]    |
| PCA_95       | RUS     | Decision Tree | 5    | 0,636363636 | 0,623931624 | [[5 6] [2 9]]    |
| PCA_95       | RUS     | Decision Tree | 6    | 0,636363636 | 0,633333333 | [[8 3] [5 6]]    |
| PCA_95       | RUS     | Decision Tree | 7    | 0,666666667 | 0,657254902 | [[9 2] [5 5]]    |
| PCA_95       | RUS     | Decision Tree | 8    | 0,857142857 | 0,856489049 | [[10 1] [ 2 8]]  |
| PCA_95       | RUS     | Decision Tree | 9    | 0,761904762 | 0,761904762 | [[8 2] [3 8]]    |
| PCA_95       | RUS     | Decision Tree | 10   | 0,571428571 | 0,569467146 | [[5 5] [4 7]]    |
| PCA_95       | RUS     | Naive Bayes   | 1    | 0,818181818 | 0,816666667 | [[ 8 3] [ 1 10]] |
| PCA_95       | RUS     | Naive Bayes   | 2    | 0,772727273 | 0,772256729 | [[9 2] [3 8]]    |
| PCA_95       | RUS     | Naive Bayes   | 3    | 0,727272727 | 0,725       | [[7 4] [2 9]]    |
| PCA_95       | RUS     | Naive Bayes   | 4    | 0,727272727 | 0,725       | [[9 2] [4 7]]    |
| PCA_95       | RUS     | Naive Bayes   | 5    | 0,727272727 | 0,725       | [[7 4] [2 9]]    |
| PCA_95       | RUS     | Naive Bayes   | 6    | 0,727272727 | 0,725       | [[7 4] [2 9]]    |
| PCA_95       | RUS     | Naive Bayes   | 7    | 0,80952381  | 0,80952381  | [[9 2] [2 8]]    |
| PCA_95       | RUS     | Naive Bayes   | 8    | 0,761904762 | 0,755182073 | [[10 1] [ 4 6]]  |
| PCA_95       | RUS     | Naive Bayes   | 9    | 0,714285714 | 0,714285714 | [[7 3] [3 8]]    |
| PCA_95       | RUS     | Naive Bayes   | 10   | 0,80952381  | 0,80952381  | [[8 2] [2 9]]    |
| PCA_95       | RUS     | SVM           | 1    | 0,818181818 | 0,818181818 | [[9 2] [2 9]]    |
| PCA_95       | RUS     | SVM           | 2    | 0,772727273 | 0,768421053 | [[10 1] [ 4 7]]  |
| PCA_95       | RUS     | SVM           | 3    | 0,727272727 | 0,727272727 | [[8 3] [3 8]]    |
| PCA_95       | RUS     | SVM           | 4    | 0,772727273 | 0,772256729 | [[9 2] [3 8]]    |
| PCA_95       | RUS     | SVM           | 5    | 0,727272727 | 0,725       | [[7 4] [2 9]]    |
| PCA_95       | RUS     | SVM           | 6    | 0,681818182 | 0,675789474 | [[6 5] [2 9]]    |
| PCA_95       | RUS     | SVM           | 7    | 0,80952381  | 0,806878307 | [[10 1] [ 3 7]]  |
| PCA_95       | RUS     | SVM           | 8    | 0,80952381  | 0,8003663   | [[11 0] [ 4 6]]  |
| PCA_95       | RUS     | SVM           | 9    | 0,666666667 | 0,666666667 | [[7 3] [4 7]]    |
| PCA_95       | RUS     | SVM           | 10   | 0,80952381  | 0,80952381  | [[8 2] [2 9]]    |
| PCA_95       | RUS     | XGBoost       | 1    | 0,909090909 | 0,909090909 | [[10 1] [ 1 10]] |
| PCA_95       | RUS     | XGBoost       | 2    | 0,863636364 | 0,863354037 | [[ 9 2] [ 1 10]] |
| PCA_95       | RUS     | XGBoost       | 3    | 0,772727273 | 0,772256729 | [[9 2] [3 8]]    |
| PCA_95       | RUS     | XGBoost       | 4    | 0,772727273 | 0,772256729 | [[9 2] [3 8]]    |
| PCA_95       | RUS     | XGBoost       | 5    | 0,727272727 | 0,725       | [[7 4] [2 9]]    |
| PCA_95       | RUS     | XGBoost       | 6    | 0,727272727 | 0,725       | [[7 4] [2 9]]    |
| PCA_95       | RUS     | XGBoost       | 7    | 0,714285714 | 0,700549451 | [[10 1] [ 5 5]]  |
| PCA_95       | RUS     | XGBoost       | 8    | 0,714285714 | 0,714285714 | [[8 3] [3 7]]    |
| PCA_95       | RUS     | XGBoost       | 9    | 0,761904762 | 0,761904762 | [[8 2] [3 8]]    |
| PCA_95       | RUS     | XGBoost       | 10   | 0,666666667 | 0,666666667 | [[7 3] [4 7]]    |

| PCA Approach | Sampler       | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix |
|--------------|---------------|---------------|------|-------------|-------------|------------------|
| PCA_95       | RUS%5 + SMOTE | Random Forest | 1    | 0,882352941 | 0,881439532 | [[25 1] [ 5 20]] |
| PCA_95       | RUS%5 + SMOTE | Random Forest | 2    | 0,803921569 | 0,801638047 | [[18 8] [ 2 23]] |
| PCA_95       | RUS%5 + SMOTE | Random Forest | 3    | 0,764705882 | 0,760833487 | [[23 3] [ 9 16]] |
| PCA_95       | RUS%5 + SMOTE | Random Forest | 4    | 0,803921569 | 0,803921569 | [[21 5] [ 5 20]] |
| PCA_95       | RUS%5 + SMOTE | Random Forest | 5    | 0,823529412 | 0,822710206 | [[19 6] [ 3 23]] |
| PCA_95       | RUS%5 + SMOTE | Random Forest | 6    | 0,882352941 | 0,882262443 | [[23 2] [ 4 22]] |
| PCA_95       | RUS%5 + SMOTE | Random Forest | 7    | 0,803921569 | 0,801638047 | [[23 2] [ 8 18]] |
| PCA_95       | RUS%5 + SMOTE | Random Forest | 8    | 0,823529412 | 0,820639835 | [[24 1] [ 8 18]] |
| PCA_95       | RUS%5 + SMOTE | Random Forest | 9    | 0,86        | 0,859943978 | [[21 4] [ 3 22]] |
| PCA_95       | RUS%5 + SMOTE | Random Forest | 10   | 0,84        | 0,83974359  | [[22 3] [ 5 20]] |
| PCA_95       | RUS%5 + SMOTE | Decision Tree | 1    | 0,745098039 | 0,735925238 | [[24 2] [11 14]] |
| PCA_95       | RUS%5 + SMOTE | Decision Tree | 2    | 0,823529412 | 0,823393508 | [[22 4] [ 5 20]] |
| PCA_95       | RUS%5 + SMOTE | Decision Tree | 3    | 0,745098039 | 0,745098039 | [[19 7] [ 6 19]] |
| PCA_95       | RUS%5 + SMOTE | Decision Tree | 4    | 0,764705882 | 0,764524887 | [[19 7] [ 5 20]] |
| PCA_95       | RUS%5 + SMOTE | Decision Tree | 5    | 0,823529412 | 0,822710206 | [[19 6] [ 3 23]] |
| PCA_95       | RUS%5 + SMOTE | Decision Tree | 6    | 0,862745098 | 0,862639395 | [[21 4] [ 3 23]] |
| PCA_95       | RUS%5 + SMOTE | Decision Tree | 7    | 0,764705882 | 0,764705882 | [[19 6] [ 6 20]] |
| PCA_95       | RUS%5 + SMOTE | Decision Tree | 8    | 0,745098039 | 0,743125877 | [[21 4] [ 9 17]] |
| PCA_95       | RUS%5 + SMOTE | Decision Tree | 9    | 0,8         | 0,799679487 | [[19 6] [ 4 21]] |
| PCA_95       | RUS%5 + SMOTE | Decision Tree | 10   | 0,84        | 0,84        | [[21 4] [ 4 21]] |
| PCA_95       | RUS%5 + SMOTE | Naive Bayes   | 1    | 0,705882353 | 0,695298351 | [[23 3] [12 13]] |
| PCA_95       | RUS%5 + SMOTE | Naive Bayes   | 2    | 0,803921569 | 0,803467683 | [[22 4] [ 6 19]] |
| PCA_95       | RUS%5 + SMOTE | Naive Bayes   | 3    | 0,607843137 | 0,604798441 | [[18 8] [12 13]] |
| PCA_95       | RUS%5 + SMOTE | Naive Bayes   | 4    | 0,705882353 | 0,705655847 | [[19 7] [ 8 17]] |
| PCA_95       | RUS%5 + SMOTE | Naive Bayes   | 5    | 0,705882353 | 0,705655847 | [[17 8] [ 7 19]] |
| PCA_95       | RUS%5 + SMOTE | Naive Bayes   | 6    | 0,803921569 | 0,803770739 | [[21 4] [ 6 20]] |
| PCA_95       | RUS%5 + SMOTE | Naive Bayes   | 7    | 0,666666667 | 0,657206675 | [[21 4] [13 13]] |
| PCA_95       | RUS%5 + SMOTE | Naive Bayes   | 8    | 0,647058824 | 0,639314033 | [[20 5] [13 13]] |
| PCA_95       | RUS%5 + SMOTE | Naive Bayes   | 9    | 0,76        | 0,759615385 | [[20 5] [ 7 18]] |
| PCA_95       | RUS%5 + SMOTE | Naive Bayes   | 10   | 0,74        | 0,734802122 | [[22 3] [10 15]] |
| PCA_95       | RUS%5 + SMOTE | SVM           | 1    | 0,764705882 | 0,760833487 | [[23 3] [ 9 16]] |
| PCA_95       | RUS%5 + SMOTE | SVM           | 2    | 0,803921569 | 0,803013798 | [[19 7] [ 3 22]] |
| PCA_95       | RUS%5 + SMOTE | SVM           | 3    | 0,666666667 | 0,665119278 | [[19 7] [10 15]] |
| PCA_95       | RUS%5 + SMOTE | SVM           | 4    | 0,705882353 | 0,70244238  | [[21 5] [10 15]] |
| PCA_95       | RUS%5 + SMOTE | SVM           | 5    | 0,745098039 | 0,74211673  | [[16 9] [ 4 22]] |
| PCA_95       | RUS%5 + SMOTE | SVM           | 6    | 0,843137255 | 0,841310437 | [[24 1] [ 7 19]] |
| PCA_95       | RUS%5 + SMOTE | SVM           | 7    | 0,764705882 | 0,764524887 | [[20 5] [ 7 19]] |
| PCA_95       | RUS%5 + SMOTE | SVM           | 8    | 0,764705882 | 0,759542689 | [[23 2] [10 16]] |
| PCA_95       | RUS%5 + SMOTE | SVM           | 9    | 0,84        | 0,83974359  | [[22 3] [ 5 20]] |
| PCA_95       | RUS%5 + SMOTE | SVM           | 10   | 0,8         | 0,799679487 | [[21 4] [ 6 19]] |
| PCA_95       | RUS%5 + SMOTE | XGBoost       | 1    | 0,803921569 | 0,798319328 | [[25 1] [ 9 16]] |
| PCA_95       | RUS%5 + SMOTE | XGBoost       | 2    | 0,862745098 | 0,862745098 | [[22 4] [ 3 22]] |
| PCA_95       | RUS%5 + SMOTE | XGBoost       | 3    | 0,725490196 | 0,724854757 | [[20 6] [ 8 17]] |
| PCA_95       | RUS%5 + SMOTE | XGBoost       | 4    | 0,803921569 | 0,803467683 | [[22 4] [ 6 19]] |
| PCA_95       | RUS%5 + SMOTE | XGBoost       | 5    | 0,843137255 | 0,841919376 | [[19 6] [ 2 24]] |
| PCA_95       | RUS%5 + SMOTE | XGBoost       | 6    | 0,862745098 | 0,861683165 | [[24 1] [ 6 20]] |
| PCA_95       | RUS%5 + SMOTE | XGBoost       | 7    | 0,823529412 | 0,823529412 | [[21 4] [ 5 21]] |
| PCA_95       | RUS%5 + SMOTE | XGBoost       | 8    | 0,823529412 | 0,822164069 | [[23 2] [ 7 19]] |
| PCA_95       | RUS%5 + SMOTE | XGBoost       | 9    | 0,82        | 0,819927971 | [[20 5] [ 4 21]] |
| PCA_95       | RUS%5 + SMOTE | XGBoost       | 10   | 0,88        | 0,879807692 | [[23 2] [ 4 21]] |

| PCA Approach | Sampler | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix  |
|--------------|---------|---------------|------|-------------|-------------|-------------------|
| PCA_99       | SMOTE   | Random Forest | 1    | 0,905882353 | 0,905147879 | [[42 0] [ 8 35]]  |
| PCA_99       | SMOTE   | Random Forest | 2    | 0,905882353 | 0,905725838 | [[40 2] [ 6 37]]  |
| PCA_99       | SMOTE   | Random Forest | 3    | 0,917647059 | 0,917578631 | [[40 2] [ 5 38]]  |
| PCA_99       | SMOTE   | Random Forest | 4    | 0,882352941 | 0,881434848 | [[41 1] [ 9 34]]  |
| PCA_99       | SMOTE   | Random Forest | 5    | 0,870588235 | 0,870588235 | [[37 5] [ 6 37]]  |
| PCA_99       | SMOTE   | Random Forest | 6    | 0,870588235 | 0,870588235 | [[37 6] [ 5 37]]  |
| PCA_99       | SMOTE   | Random Forest | 7    | 0,917647059 | 0,917003048 | [[43 0] [ 7 35]]  |
| PCA_99       | SMOTE   | Random Forest | 8    | 0,929411765 | 0,929215686 | [[42 1] [ 5 37]]  |
| PCA_99       | SMOTE   | Random Forest | 9    | 0,870588235 | 0,870372818 | [[39 4] [ 7 35]]  |
| PCA_99       | SMOTE   | Random Forest | 10   | 0,870588235 | 0,869576218 | [[41 2] [ 9 33]]  |
| PCA_99       | SMOTE   | Decision Tree | 1    | 0,882352941 | 0,881862745 | [[40 2] [ 8 35]]  |
| PCA_99       | SMOTE   | Decision Tree | 2    | 0,858823529 | 0,858784444 | [[37 5] [ 7 36]]  |
| PCA_99       | SMOTE   | Decision Tree | 3    | 0,882352941 | 0,881862745 | [[40 2] [ 8 35]]  |
| PCA_99       | SMOTE   | Decision Tree | 4    | 0,882352941 | 0,881434848 | [[41 1] [ 9 34]]  |
| PCA_99       | SMOTE   | Decision Tree | 5    | 0,858823529 | 0,858823529 | [[36 6] [ 6 37]]  |
| PCA_99       | SMOTE   | Decision Tree | 6    | 0,8         | 0,799833818 | [[33 10] [ 7 35]] |
| PCA_99       | SMOTE   | Decision Tree | 7    | 0,882352941 | 0,882255119 | [[39 4] [ 6 36]]  |
| PCA_99       | SMOTE   | Decision Tree | 8    | 0,776470588 | 0,776408677 | [[34 9] [10 32]]  |
| PCA_99       | SMOTE   | Decision Tree | 9    | 0,776470588 | 0,776470588 | [[33 10] [ 9 33]] |
| PCA_99       | SMOTE   | Decision Tree | 10   | 0,882352941 | 0,882026144 | [[40 3] [ 7 35]]  |
| PCA_99       | SMOTE   | Naive Bayes   | 1    | 0,823529412 | 0,822498425 | [[38 4] [11 32]]  |
| PCA_99       | SMOTE   | Naive Bayes   | 2    | 0,764705882 | 0,760333863 | [[38 4] [16 27]]  |
| PCA_99       | SMOTE   | Naive Bayes   | 3    | 0,752941176 | 0,752941176 | [[32 10] [11 32]] |
| PCA_99       | SMOTE   | Naive Bayes   | 4    | 0,8         | 0,798831548 | [[37 5] [12 31]]  |
| PCA_99       | SMOTE   | Naive Bayes   | 5    | 0,8         | 0,796912281 | [[39 3] [14 29]]  |
| PCA_99       | SMOTE   | Naive Bayes   | 6    | 0,811764706 | 0,81124183  | [[37 6] [10 32]]  |
| PCA_99       | SMOTE   | Naive Bayes   | 7    | 0,811764706 | 0,806117647 | [[42 1] [15 27]]  |
| PCA_99       | SMOTE   | Naive Bayes   | 8    | 0,741176471 | 0,733411765 | [[39 4] [18 24]]  |
| PCA_99       | SMOTE   | Naive Bayes   | 9    | 0,8         | 0,797473684 | [[39 4] [13 29]]  |
| PCA_99       | SMOTE   | Naive Bayes   | 10   | 0,705882353 | 0,689506572 | [[40 3] [22 20]]  |
| PCA_99       | SMOTE   | SVM           | 1    | 0,858823529 | 0,857721818 | [[40 2] [10 33]]  |
| PCA_99       | SMOTE   | SVM           | 2    | 0,870588235 | 0,870480706 | [[38 4] [ 7 36]]  |
| PCA_99       | SMOTE   | SVM           | 3    | 0,8         | 0,799667083 | [[32 10] [ 7 36]] |
| PCA_99       | SMOTE   | SVM           | 4    | 0,835294118 | 0,835294118 | [[35 7] [ 7 36]]  |
| PCA_99       | SMOTE   | SVM           | 5    | 0,788235294 | 0,787883136 | [[35 7] [11 32]]  |
| PCA_99       | SMOTE   | SVM           | 6    | 0,788235294 | 0,788235294 | [[34 9] [ 9 33]]  |
| PCA_99       | SMOTE   | SVM           | 7    | 0,870588235 | 0,870048195 | [[40 3] [ 8 34]]  |
| PCA_99       | SMOTE   | SVM           | 8    | 0,882352941 | 0,882026144 | [[40 3] [ 7 35]]  |
| PCA_99       | SMOTE   | SVM           | 9    | 0,823529412 | 0,823480535 | [[36 7] [ 8 34]]  |
| PCA_99       | SMOTE   | SVM           | 10   | 0,788235294 | 0,787647059 | [[36 7] [11 31]]  |
| PCA_99       | SMOTE   | XGBoost       | 1    | 0,905882353 | 0,905147879 | [[42 0] [ 8 35]]  |
| PCA_99       | SMOTE   | XGBoost       | 2    | 0,917647059 | 0,917578631 | [[40 2] [ 5 38]]  |
| PCA_99       | SMOTE   | XGBoost       | 3    | 0,929411765 | 0,929294379 | [[41 1] [ 5 38]]  |
| PCA_99       | SMOTE   | XGBoost       | 4    | 0,882352941 | 0,881434848 | [[41 1] [ 9 34]]  |
| PCA_99       | SMOTE   | XGBoost       | 5    | 0,882352941 | 0,882352941 | [[37 5] [ 5 38]]  |
| PCA_99       | SMOTE   | XGBoost       | 6    | 0,905882353 | 0,905856296 | [[38 5] [ 3 39]]  |
| PCA_99       | SMOTE   | XGBoost       | 7    | 0,905882353 | 0,904933799 | [[43 0] [ 8 34]]  |
| PCA_99       | SMOTE   | XGBoost       | 8    | 0,917647059 | 0,91762425  | [[40 3] [ 4 38]]  |
| PCA_99       | SMOTE   | XGBoost       | 9    | 0,905882353 | 0,905804095 | [[40 3] [ 5 37]]  |
| PCA_99       | SMOTE   | XGBoost       | 10   | 0,905882353 | 0,905620915 | [[41 2] [ 6 36]]  |

| PCA Approach | Sampler | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix  |
|--------------|---------|---------------|------|-------------|-------------|-------------------|
| PCA_99       | ROS     | Random Forest | 1    | 0,929411765 | 0,929117647 | [[42 0] [ 6 37]]  |
| PCA_99       | ROS     | Random Forest | 2    | 0,964705882 | 0,964676556 | [[42 0] [ 3 40]]  |
| PCA_99       | ROS     | Random Forest | 3    | 0,964705882 | 0,964676556 | [[42 0] [ 3 40]]  |
| PCA_99       | ROS     | Random Forest | 4    | 0,929411765 | 0,929117647 | [[42 0] [ 6 37]]  |
| PCA_99       | ROS     | Random Forest | 5    | 0,952941176 | 0,952928148 | [[41 1] [ 3 40]]  |
| PCA_99       | ROS     | Random Forest | 6    | 0,941176471 | 0,940930998 | [[43 0] [ 5 37]]  |
| PCA_99       | ROS     | Random Forest | 7    | 0,952941176 | 0,952902048 | [[42 1] [ 3 39]]  |
| PCA_99       | ROS     | Random Forest | 8    | 0,964705882 | 0,964647132 | [[43 0] [ 3 39]]  |
| PCA_99       | ROS     | Random Forest | 9    | 0,929411765 | 0,928998623 | [[43 0] [ 6 36]]  |
| PCA_99       | ROS     | Random Forest | 10   | 0,952941176 | 0,952810458 | [[43 0] [ 4 38]]  |
| PCA_99       | ROS     | Decision Tree | 1    | 0,882352941 | 0,881862745 | [[40 2] [ 8 35]]  |
| PCA_99       | ROS     | Decision Tree | 2    | 0,917647059 | 0,917165932 | [[42 0] [ 7 36]]  |
| PCA_99       | ROS     | Decision Tree | 3    | 0,917647059 | 0,917165932 | [[42 0] [ 7 36]]  |
| PCA_99       | ROS     | Decision Tree | 4    | 0,858823529 | 0,856200318 | [[42 0] [12 31]]  |
| PCA_99       | ROS     | Decision Tree | 5    | 0,929411765 | 0,929294379 | [[41 1] [ 5 38]]  |
| PCA_99       | ROS     | Decision Tree | 6    | 0,894117647 | 0,893289633 | [[42 1] [ 8 34]]  |
| PCA_99       | ROS     | Decision Tree | 7    | 0,929411765 | 0,929215686 | [[42 1] [ 5 37]]  |
| PCA_99       | ROS     | Decision Tree | 8    | 0,941176471 | 0,940930998 | [[43 0] [ 5 37]]  |
| PCA_99       | ROS     | Decision Tree | 9    | 0,917647059 | 0,917003048 | [[43 0] [ 7 35]]  |
| PCA_99       | ROS     | Decision Tree | 10   | 0,917647059 | 0,917003048 | [[43 0] [ 7 35]]  |
| PCA_99       | ROS     | Naive Bayes   | 1    | 0,811764706 | 0,811712592 | [[35 7] [ 9 34]]  |
| PCA_99       | ROS     | Naive Bayes   | 2    | 0,776470588 | 0,776470588 | [[33 9] [10 33]]  |
| PCA_99       | ROS     | Naive Bayes   | 3    | 0,8         | 0,799944606 | [[33 9] [ 8 35]]  |
| PCA_99       | ROS     | Naive Bayes   | 4    | 0,776470588 | 0,775850448 | [[35 7] [12 31]]  |
| PCA_99       | ROS     | Naive Bayes   | 5    | 0,729411765 | 0,725993808 | [[26 16] [ 7 36]] |
| PCA_99       | ROS     | Naive Bayes   | 6    | 0,788235294 | 0,787352941 | [[31 12] [ 6 36]] |
| PCA_99       | ROS     | Naive Bayes   | 7    | 0,870588235 | 0,870048195 | [[40 3] [ 8 34]]  |
| PCA_99       | ROS     | Naive Bayes   | 8    | 0,776470588 | 0,776408677 | [[34 9] [10 32]]  |
| PCA_99       | ROS     | Naive Bayes   | 9    | 0,729411765 | 0,728661069 | [[29 14] [ 9 33]] |
| PCA_99       | ROS     | Naive Bayes   | 10   | 0,741176471 | 0,737168786 | [[37 6] [16 26]]  |
| PCA_99       | ROS     | SVM           | 1    | 0,847058824 | 0,846165302 | [[39 3] [10 33]]  |
| PCA_99       | ROS     | SVM           | 2    | 0,858823529 | 0,858784444 | [[37 5] [ 7 36]]  |
| PCA_99       | ROS     | SVM           | 3    | 0,823529412 | 0,823039828 | [[37 5] [10 33]]  |
| PCA_99       | ROS     | SVM           | 4    | 0,847058824 | 0,846634517 | [[38 4] [ 9 34]]  |
| PCA_99       | ROS     | SVM           | 5    | 0,717647059 | 0,717647059 | [[30 12] [12 31]] |
| PCA_99       | ROS     | SVM           | 6    | 0,788235294 | 0,786101047 | [[38 5] [13 29]]  |
| PCA_99       | ROS     | SVM           | 7    | 0,847058824 | 0,846420594 | [[39 4] [ 9 33]]  |
| PCA_99       | ROS     | SVM           | 8    | 0,811764706 | 0,811608191 | [[36 7] [ 9 33]]  |
| PCA_99       | ROS     | SVM           | 9    | 0,717647059 | 0,716862745 | [[33 10] [14 28]] |
| PCA_99       | ROS     | SVM           | 10   | 0,788235294 | 0,781882353 | [[41 2] [16 26]]  |
| PCA_99       | ROS     | XGBoost       | 1    | 0,929411765 | 0,929117647 | [[42 0] [ 6 37]]  |
| PCA_99       | ROS     | XGBoost       | 2    | 0,952941176 | 0,952862919 | [[42 0] [ 4 39]]  |
| PCA_99       | ROS     | XGBoost       | 3    | 0,952941176 | 0,952862919 | [[42 0] [ 4 39]]  |
| PCA_99       | ROS     | XGBoost       | 4    | 0,929411765 | 0,929117647 | [[42 0] [ 6 37]]  |
| PCA_99       | ROS     | XGBoost       | 5    | 0,964705882 | 0,964705882 | [[41 1] [ 2 41]]  |
| PCA_99       | ROS     | XGBoost       | 6    | 0,917647059 | 0,917303397 | [[42 1] [ 6 36]]  |
| PCA_99       | ROS     | XGBoost       | 7    | 0,917647059 | 0,917003048 | [[43 0] [ 7 35]]  |
| PCA_99       | ROS     | XGBoost       | 8    | 0,964705882 | 0,964647132 | [[43 0] [ 3 39]]  |
| PCA_99       | ROS     | XGBoost       | 9    | 0,941176471 | 0,940930998 | [[43 0] [ 5 37]]  |
| PCA_99       | ROS     | XGBoost       | 10   | 0,941176471 | 0,940930998 | [[43 0] [ 5 37]]  |

| PCA Approach | Sampler          | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix  |
|--------------|------------------|---------------|------|-------------|-------------|-------------------|
| PCA_99       | Borderline SMOTE | Random Forest | 1    | 0,894117647 | 0,893499055 | [[41 1] [ 8 35]]  |
| PCA_99       | Borderline SMOTE | Random Forest | 2    | 0,905882353 | 0,905725838 | [[40 2] [ 6 37]]  |
| PCA_99       | Borderline SMOTE | Random Forest | 3    | 0,941176471 | 0,941013276 | [[42 0] [ 5 38]]  |
| PCA_99       | Borderline SMOTE | Random Forest | 4    | 0,882352941 | 0,881434848 | [[41 1] [ 9 34]]  |
| PCA_99       | Borderline SMOTE | Random Forest | 5    | 0,905882353 | 0,905725838 | [[40 2] [ 6 37]]  |
| PCA_99       | Borderline SMOTE | Random Forest | 6    | 0,917647059 | 0,917303397 | [[42 1] [ 6 36]]  |
| PCA_99       | Borderline SMOTE | Random Forest | 7    | 0,905882353 | 0,905804095 | [[40 3] [ 5 37]]  |
| PCA_99       | Borderline SMOTE | Random Forest | 8    | 0,941176471 | 0,941078554 | [[42 1] [ 4 38]]  |
| PCA_99       | Borderline SMOTE | Random Forest | 9    | 0,905882353 | 0,905620915 | [[41 2] [ 6 36]]  |
| PCA_99       | Borderline SMOTE | Random Forest | 10   | 0,905882353 | 0,905331497 | [[42 1] [ 7 35]]  |
| PCA_99       | Borderline SMOTE | Decision Tree | 1    | 0,847058824 | 0,846165302 | [[39 3] [10 33]]  |
| PCA_99       | Borderline SMOTE | Decision Tree | 2    | 0,905882353 | 0,905725838 | [[40 2] [ 6 37]]  |
| PCA_99       | Borderline SMOTE | Decision Tree | 3    | 0,729411765 | 0,72933682  | [[30 12] [11 32]] |
| PCA_99       | Borderline SMOTE | Decision Tree | 4    | 0,847058824 | 0,846165302 | [[39 3] [10 33]]  |
| PCA_99       | Borderline SMOTE | Decision Tree | 5    | 0,847058824 | 0,846931743 | [[37 5] [ 8 35]]  |
| PCA_99       | Borderline SMOTE | Decision Tree | 6    | 0,764705882 | 0,763328743 | [[36 7] [13 29]]  |
| PCA_99       | Borderline SMOTE | Decision Tree | 7    | 0,905882353 | 0,905804095 | [[40 3] [ 5 37]]  |
| PCA_99       | Borderline SMOTE | Decision Tree | 8    | 0,823529412 | 0,823529412 | [[35 8] [ 7 35]]  |
| PCA_99       | Borderline SMOTE | Decision Tree | 9    | 0,882352941 | 0,882255119 | [[39 4] [ 6 36]]  |
| PCA_99       | Borderline SMOTE | Decision Tree | 10   | 0,823529412 | 0,823235661 | [[37 6] [ 9 33]]  |
| PCA_99       | Borderline SMOTE | Naive Bayes   | 1    | 0,752941176 | 0,750457133 | [[36 6] [15 28]]  |
| PCA_99       | Borderline SMOTE | Naive Bayes   | 2    | 0,741176471 | 0,734500467 | [[38 4] [18 25]]  |
| PCA_99       | Borderline SMOTE | Naive Bayes   | 3    | 0,788235294 | 0,782773109 | [[40 2] [16 27]]  |
| PCA_99       | Borderline SMOTE | Naive Bayes   | 4    | 0,811764706 | 0,80690943  | [[41 1] [15 28]]  |
| PCA_99       | Borderline SMOTE | Naive Bayes   | 5    | 0,729411765 | 0,725234262 | [[36 6] [17 26]]  |
| PCA_99       | Borderline SMOTE | Naive Bayes   | 6    | 0,741176471 | 0,739661617 | [[35 8] [14 28]]  |
| PCA_99       | Borderline SMOTE | Naive Bayes   | 7    | 0,741176471 | 0,735454182 | [[38 5] [17 25]]  |
| PCA_99       | Borderline SMOTE | Naive Bayes   | 8    | 0,670588235 | 0,657670127 | [[37 6] [22 20]]  |
| PCA_99       | Borderline SMOTE | Naive Bayes   | 9    | 0,705882353 | 0,702167183 | [[35 8] [17 25]]  |
| PCA_99       | Borderline SMOTE | Naive Bayes   | 10   | 0,682352941 | 0,660582738 | [[40 3] [24 18]]  |
| PCA_99       | Borderline SMOTE | SVM           | 1    | 0,870588235 | 0,869832178 | [[40 2] [ 9 34]]  |
| PCA_99       | Borderline SMOTE | SVM           | 2    | 0,905882353 | 0,905725838 | [[40 2] [ 6 37]]  |
| PCA_99       | Borderline SMOTE | SVM           | 3    | 0,823529412 | 0,823529412 | [[35 7] [ 8 35]]  |
| PCA_99       | Borderline SMOTE | SVM           | 4    | 0,870588235 | 0,870229207 | [[39 3] [ 8 35]]  |
| PCA_99       | Borderline SMOTE | SVM           | 5    | 0,788235294 | 0,786582727 | [[37 5] [13 30]]  |
| PCA_99       | Borderline SMOTE | SVM           | 6    | 0,823529412 | 0,822792993 | [[38 5] [10 32]]  |
| PCA_99       | Borderline SMOTE | SVM           | 7    | 0,858823529 | 0,858431373 | [[39 4] [ 8 34]]  |
| PCA_99       | Borderline SMOTE | SVM           | 8    | 0,858823529 | 0,857997246 | [[40 3] [ 9 33]]  |
| PCA_99       | Borderline SMOTE | SVM           | 9    | 0,788235294 | 0,787647059 | [[36 7] [11 31]]  |
| PCA_99       | Borderline SMOTE | SVM           | 10   | 0,788235294 | 0,781882353 | [[41 2] [16 26]]  |
| PCA_99       | Borderline SMOTE | XGBoost       | 1    | 0,894117647 | 0,893499055 | [[41 1] [ 8 35]]  |
| PCA_99       | Borderline SMOTE | XGBoost       | 2    | 0,905882353 | 0,905856296 | [[39 3] [ 5 38]]  |
| PCA_99       | Borderline SMOTE | XGBoost       | 3    | 0,917647059 | 0,917578631 | [[40 2] [ 5 38]]  |
| PCA_99       | Borderline SMOTE | XGBoost       | 4    | 0,894117647 | 0,893499055 | [[41 1] [ 8 35]]  |
| PCA_99       | Borderline SMOTE | XGBoost       | 5    | 0,894117647 | 0,894117647 | [[38 4] [ 5 38]]  |
| PCA_99       | Borderline SMOTE | XGBoost       | 6    | 0,894117647 | 0,893941397 | [[40 3] [ 6 36]]  |
| PCA_99       | Borderline SMOTE | XGBoost       | 7    | 0,882352941 | 0,882026144 | [[40 3] [ 7 35]]  |
| PCA_99       | Borderline SMOTE | XGBoost       | 8    | 0,917647059 | 0,917303397 | [[42 1] [ 6 36]]  |
| PCA_99       | Borderline SMOTE | XGBoost       | 9    | 0,917647059 | 0,917509975 | [[41 2] [ 5 37]]  |
| PCA_99       | Borderline SMOTE | XGBoost       | 10   | 0,894117647 | 0,892780186 | [[43 0] [ 9 33]]  |

| PCA Approach | Sampler | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix |
|--------------|---------|---------------|------|-------------|-------------|------------------|
| PCA_99       | RUS     | Random Forest | 1    | 0,863636364 | 0,863354037 | [[ 9 2] [ 1 10]] |
| PCA_99       | RUS     | Random Forest | 2    | 0,727272727 | 0,717948718 | [[10 1] [ 5 6]]  |
| PCA_99       | RUS     | Random Forest | 3    | 0,727272727 | 0,725       | [[9 2] [4 7]]    |
| PCA_99       | RUS     | Random Forest | 4    | 0,818181818 | 0,811965812 | [[11 0] [ 4 7]]  |
| PCA_99       | RUS     | Random Forest | 5    | 0,909090909 | 0,909090909 | [[10 1] [ 1 10]] |
| PCA_99       | RUS     | Random Forest | 6    | 0,727272727 | 0,725       | [[7 4] [2 9]]    |
| PCA_99       | RUS     | Random Forest | 7    | 0,714285714 | 0,71031746  | [[9 2] [4 6]]    |
| PCA_99       | RUS     | Random Forest | 8    | 0,857142857 | 0,853109244 | [[11 0] [ 3 7]]  |
| PCA_99       | RUS     | Random Forest | 9    | 0,80952381  | 0,806878307 | [[ 7 3] [ 1 10]] |
| PCA_99       | RUS     | Random Forest | 10   | 0,80952381  | 0,80952381  | [[8 2] [2 9]]    |
| PCA_99       | RUS     | Decision Tree | 1    | 0,909090909 | 0,909090909 | [[10 1] [ 1 10]] |
| PCA_99       | RUS     | Decision Tree | 2    | 0,5         | 0,490526316 | [[7 4] [7 4]]    |
| PCA_99       | RUS     | Decision Tree | 3    | 0,727272727 | 0,727272727 | [[8 3] [3 8]]    |
| PCA_99       | RUS     | Decision Tree | 4    | 0,772727273 | 0,772256729 | [[9 2] [3 8]]    |
| PCA_99       | RUS     | Decision Tree | 5    | 0,772727273 | 0,768421053 | [[ 7 4] [ 1 10]] |
| PCA_99       | RUS     | Decision Tree | 6    | 0,636363636 | 0,633333333 | [[6 5] [3 8]]    |
| PCA_99       | RUS     | Decision Tree | 7    | 0,619047619 | 0,619047619 | [[7 4] [4 6]]    |
| PCA_99       | RUS     | Decision Tree | 8    | 0,714285714 | 0,712987013 | [[7 4] [2 8]]    |
| PCA_99       | RUS     | Decision Tree | 9    | 0,619047619 | 0,617316017 | [[7 3] [5 6]]    |
| PCA_99       | RUS     | Decision Tree | 10   | 0,619047619 | 0,619047619 | [[6 4] [4 7]]    |
| PCA_99       | RUS     | Naive Bayes   | 1    | 0,772727273 | 0,768421053 | [[ 7 4] [ 1 10]] |
| PCA_99       | RUS     | Naive Bayes   | 2    | 0,772727273 | 0,760348584 | [[11 0] [ 5 6]]  |
| PCA_99       | RUS     | Naive Bayes   | 3    | 0,818181818 | 0,816666667 | [[10 1] [ 3 8]]  |
| PCA_99       | RUS     | Naive Bayes   | 4    | 0,681818182 | 0,675789474 | [[9 2] [5 6]]    |
| PCA_99       | RUS     | Naive Bayes   | 5    | 0,818181818 | 0,818181818 | [[9 2] [2 9]]    |
| PCA_99       | RUS     | Naive Bayes   | 6    | 0,727272727 | 0,727272727 | [[8 3] [3 8]]    |
| PCA_99       | RUS     | Naive Bayes   | 7    | 0,761904762 | 0,75863572  | [[7 4] [1 9]]    |
| PCA_99       | RUS     | Naive Bayes   | 8    | 0,80952381  | 0,8003663   | [[11 0] [ 4 6]]  |
| PCA_99       | RUS     | Naive Bayes   | 9    | 0,761904762 | 0,760815081 | [[7 3] [2 9]]    |
| PCA_99       | RUS     | Naive Bayes   | 10   | 0,857142857 | 0,857142857 | [[9 1] [2 9]]    |
| PCA_99       | RUS     | SVM           | 1    | 0,863636364 | 0,863354037 | [[ 9 2] [ 1 10]] |
| PCA_99       | RUS     | SVM           | 2    | 0,772727273 | 0,768421053 | [[10 1] [ 4 7]]  |
| PCA_99       | RUS     | SVM           | 3    | 0,727272727 | 0,727272727 | [[8 3] [3 8]]    |
| PCA_99       | RUS     | SVM           | 4    | 0,772727273 | 0,772256729 | [[9 2] [3 8]]    |
| PCA_99       | RUS     | SVM           | 5    | 0,727272727 | 0,725       | [[7 4] [2 9]]    |
| PCA_99       | RUS     | SVM           | 6    | 0,681818182 | 0,675789474 | [[6 5] [2 9]]    |
| PCA_99       | RUS     | SVM           | 7    | 0,857142857 | 0,856489049 | [[10 1] [ 2 8]]  |
| PCA_99       | RUS     | SVM           | 8    | 0,80952381  | 0,8003663   | [[11 0] [ 4 6]]  |
| PCA_99       | RUS     | SVM           | 9    | 0,666666667 | 0,666666667 | [[7 3] [4 7]]    |
| PCA_99       | RUS     | SVM           | 10   | 0,857142857 | 0,856489049 | [[ 8 2] [ 1 10]] |
| PCA_99       | RUS     | XGBoost       | 1    | 0,909090909 | 0,909090909 | [[10 1] [ 1 10]] |
| PCA_99       | RUS     | XGBoost       | 2    | 0,772727273 | 0,772256729 | [[9 2] [3 8]]    |
| PCA_99       | RUS     | XGBoost       | 3    | 0,863636364 | 0,863354037 | [[10 1] [ 2 9]]  |
| PCA_99       | RUS     | XGBoost       | 4    | 0,818181818 | 0,816666667 | [[10 1] [ 3 8]]  |
| PCA_99       | RUS     | XGBoost       | 5    | 0,818181818 | 0,816666667 | [[ 8 3] [ 1 10]] |
| PCA_99       | RUS     | XGBoost       | 6    | 0,681818182 | 0,68115942  | [[7 4] [3 8]]    |
| PCA_99       | RUS     | XGBoost       | 7    | 0,80952381  | 0,80952381  | [[9 2] [2 8]]    |
| PCA_99       | RUS     | XGBoost       | 8    | 0,857142857 | 0,856489049 | [[10 1] [ 2 8]]  |
| PCA_99       | RUS     | XGBoost       | 9    | 0,857142857 | 0,856489049 | [[ 8 2] [ 1 10]] |
| PCA_99       | RUS     | XGBoost       | 10   | 0,80952381  | 0,808658009 | [[9 1] [3 8]]    |

| PCA Approach | Sampler       | Model         | Fold | Accuracy    | F1 Score    | Confusion Matrix |
|--------------|---------------|---------------|------|-------------|-------------|------------------|
| PCA_99       | RUS%5 + SMOTE | Random Forest | 1    | 0,823529412 | 0,819634121 | [[25 1] [ 8 17]] |
| PCA_99       | RUS%5 + SMOTE | Random Forest | 2    | 0,862745098 | 0,86242799  | [[21 5] [ 2 23]] |
| PCA_99       | RUS%5 + SMOTE | Random Forest | 3    | 0,705882353 | 0,70244238  | [[21 5] [10 15]] |
| PCA_99       | RUS%5 + SMOTE | Random Forest | 4    | 0,882352941 | 0,882352941 | [[23 3] [ 3 22]] |
| PCA_99       | RUS%5 + SMOTE | Random Forest | 5    | 0,843137255 | 0,843137255 | [[21 4] [ 4 22]] |
| PCA_99       | RUS%5 + SMOTE | Random Forest | 6    | 0,901960784 | 0,90120226  | [[25 0] [ 5 21]] |
| PCA_99       | RUS%5 + SMOTE | Random Forest | 7    | 0,803921569 | 0,803770739 | [[21 4] [ 6 20]] |
| PCA_99       | RUS%5 + SMOTE | Random Forest | 8    | 0,921568627 | 0,921508296 | [[24 1] [ 3 23]] |
| PCA_99       | RUS%5 + SMOTE | Random Forest | 9    | 0,84        | 0,84        | [[21 4] [ 4 21]] |
| PCA_99       | RUS%5 + SMOTE | Random Forest | 10   | 0,84        | 0,83974359  | [[22 3] [ 5 20]] |
| PCA_99       | RUS%5 + SMOTE | Decision Tree | 1    | 0,764705882 | 0,762879065 | [[22 4] [ 8 17]] |
| PCA_99       | RUS%5 + SMOTE | Decision Tree | 2    | 0,823529412 | 0,823121702 | [[20 6] [ 3 22]] |
| PCA_99       | RUS%5 + SMOTE | Decision Tree | 3    | 0,725490196 | 0,724854757 | [[20 6] [ 8 17]] |
| PCA_99       | RUS%5 + SMOTE | Decision Tree | 4    | 0,764705882 | 0,763616558 | [[18 8] [ 4 21]] |
| PCA_99       | RUS%5 + SMOTE | Decision Tree | 5    | 0,784313725 | 0,784313725 | [[20 5] [ 6 20]] |
| PCA_99       | RUS%5 + SMOTE | Decision Tree | 6    | 0,862745098 | 0,861683165 | [[24 1] [ 6 20]] |
| PCA_99       | RUS%5 + SMOTE | Decision Tree | 7    | 0,745098039 | 0,744901734 | [[18 7] [ 6 20]] |
| PCA_99       | RUS%5 + SMOTE | Decision Tree | 8    | 0,901960784 | 0,901734279 | [[24 1] [ 4 22]] |
| PCA_99       | RUS%5 + SMOTE | Decision Tree | 9    | 0,78        | 0,779911965 | [[19 6] [ 5 20]] |
| PCA_99       | RUS%5 + SMOTE | Decision Tree | 10   | 0,88        | 0,88        | [[22 3] [ 3 22]] |
| PCA_99       | RUS%5 + SMOTE | Naive Bayes   | 1    | 0,745098039 | 0,731403673 | [[25 1] [12 13]] |
| PCA_99       | RUS%5 + SMOTE | Naive Bayes   | 2    | 0,745098039 | 0,743914742 | [[21 5] [ 8 17]] |
| PCA_99       | RUS%5 + SMOTE | Naive Bayes   | 3    | 0,705882353 | 0,699390201 | [[22 4] [11 14]] |
| PCA_99       | RUS%5 + SMOTE | Naive Bayes   | 4    | 0,68627451  | 0,683838753 | [[20 6] [10 15]] |
| PCA_99       | RUS%5 + SMOTE | Naive Bayes   | 5    | 0,784313725 | 0,783815413 | [[21 4] [ 7 19]] |
| PCA_99       | RUS%5 + SMOTE | Naive Bayes   | 6    | 0,784313725 | 0,78078202  | [[23 2] [ 9 17]] |
| PCA_99       | RUS%5 + SMOTE | Naive Bayes   | 7    | 0,666666667 | 0,661208577 | [[20 5] [12 14]] |
| PCA_99       | RUS%5 + SMOTE | Naive Bayes   | 8    | 0,725490196 | 0,71946647  | [[22 3] [11 15]] |
| PCA_99       | RUS%5 + SMOTE | Naive Bayes   | 9    | 0,82        | 0,818181818 | [[23 2] [ 7 18]] |
| PCA_99       | RUS%5 + SMOTE | Naive Bayes   | 10   | 0,74        | 0,734802122 | [[22 3] [10 15]] |
| PCA_99       | RUS%5 + SMOTE | SVM           | 1    | 0,764705882 | 0,760833487 | [[23 3] [ 9 16]] |
| PCA_99       | RUS%5 + SMOTE | SVM           | 2    | 0,823529412 | 0,823121702 | [[20 6] [ 3 22]] |
| PCA_99       | RUS%5 + SMOTE | SVM           | 3    | 0,705882353 | 0,705655847 | [[19 7] [ 8 17]] |
| PCA_99       | RUS%5 + SMOTE | SVM           | 4    | 0,705882353 | 0,70244238  | [[21 5] [10 15]] |
| PCA_99       | RUS%5 + SMOTE | SVM           | 5    | 0,784313725 | 0,783312474 | [[18 7] [ 4 22]] |
| PCA_99       | RUS%5 + SMOTE | SVM           | 6    | 0,901960784 | 0,90120226  | [[25 0] [ 5 21]] |
| PCA_99       | RUS%5 + SMOTE | SVM           | 7    | 0,784313725 | 0,783815413 | [[21 4] [ 7 19]] |
| PCA_99       | RUS%5 + SMOTE | SVM           | 8    | 0,784313725 | 0,778192554 | [[24 1] [10 16]] |
| PCA_99       | RUS%5 + SMOTE | SVM           | 9    | 0,88        | 0,88        | [[22 3] [ 3 22]] |
| PCA_99       | RUS%5 + SMOTE | SVM           | 10   | 0,82        | 0,819927971 | [[21 4] [ 5 20]] |
| PCA_99       | RUS%5 + SMOTE | XGBoost       | 1    | 0,823529412 | 0,822710206 | [[23 3] [ 6 19]] |
| PCA_99       | RUS%5 + SMOTE | XGBoost       | 2    | 0,862745098 | 0,862745098 | [[22 4] [ 3 22]] |
| PCA_99       | RUS%5 + SMOTE | XGBoost       | 3    | 0,784313725 | 0,783312474 | [[22 4] [ 7 18]] |
| PCA_99       | RUS%5 + SMOTE | XGBoost       | 4    | 0,882352941 | 0,882352941 | [[23 3] [ 3 22]] |
| PCA_99       | RUS%5 + SMOTE | XGBoost       | 5    | 0,843137255 | 0,842411038 | [[23 2] [ 6 20]] |
| PCA_99       | RUS%5 + SMOTE | XGBoost       | 6    | 0,941176471 | 0,941040567 | [[25 0] [ 3 23]] |
| PCA_99       | RUS%5 + SMOTE | XGBoost       | 7    | 0,843137255 | 0,843016591 | [[22 3] [ 5 21]] |
| PCA_99       | RUS%5 + SMOTE | XGBoost       | 8    | 0,941176471 | 0,941040567 | [[25 0] [ 3 23]] |
| PCA_99       | RUS%5 + SMOTE | XGBoost       | 9    | 0,9         | 0,899638699 | [[24 1] [ 4 21]] |
| PCA_99       | RUS%5 + SMOTE | XGBoost       | 10   | 0,88        | 0,879807692 | [[23 2] [ 4 21]] |

## KİŞİSEL YAYINLAR VE ESERLER

C. Ozturk, O. Sagir, and U. Vural, “Machine Learning Approaches to Predict Thyroid Cancer Recurrence: A Comparative Study,” in *2024 9th International Conference on Computer Science and Engineering (UBMK)*, Antalya, Türkiye, 26–28 Ekim 2024, IEEE

C. Öztürk ve M. Kara, “Practical comparison of Nessus, Qualys and OpenVAS vulnerability scanner tools,” in *Proceedings of the 5th International Eurasian Conference on Science, Engineering and Technology (EurasianSciEnTech 2024)*, Kocaeli, Türkiye, 26–28 Haziran 2024, s. 710–716.

E. Dursun, C. Öztürk, N. B. Albayrak, N. Duru, and V. Yazıcı, “SEM ile Beyin Modellemesi Simülasyonlarının Kullanımı ve Etkileri Üzerine Bir Derleme,” in *Proceedings of the 1st Kocaeli International Engineering–Natural Sciences and Health Technologies Conference (1st Kocaeli National Engineering–Natural Sciences and Health Technologies Workshop)*, Kocaeli, Türkiye, 11–13 Ekim 2024, pp. 71–74.

A. Derin, C. Öztürk, N. B. Albayrak, and N. Duru, “Yapay Zekâ Yardımı ile Kemik Yaşı Tespiti,” in *Abstract Book of the 1st Kocaeli International Engineering–Natural Sciences and Health Technologies Conference (1st Kocaeli National Engineering–Natural Sciences and Health Technologies Workshop)*, Kocaeli, Türkiye, 11–13 Ekim 2024, p. 59.