CUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

MSc THESIS

Artificial Intelligence Based Sales Cancellation/Return Forecasting Software for E-commerce Industry

Zehra Sude SARI

Department of Computer Engineering

July 2025

CUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

MSc THESIS APPROVAL

Artificial Intelligence Based Sales Cancellation/Return Forecasting Software for E-commerce Industry

---

Zehra Sude SARI

Department of Computer Engineering

This master's thesis was evaluated by the following Jury Members on 02/07/2025 and was approved by unanimity / majority of votes.

Jury : Prof. Dr. Mehmet Fatih AKAY (Advisor) …………

: Prof. Dr. İbrahim Taner OKUMUŞ …………

: Asst. Prof. Dr. Barış ATA …………

This Thesis was written in the Department of Computer Engineering, Institute of Natural and Applied Sciences.

Thesis Number:

Prof. Dr. Sadık DİNÇER
Director
Institute of Natural and Applied Sciences

CONTENTS

CUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

MSc THESIS

Artificial Intelligence Based Sales Cancellation/Return Forecasting Software for E-commerce Industry

Zehra Sude SARI

Advisor: Prof. Dr. M. Fatih AKAY

Department of Computer Engineering

ABSTRACT

In today's rapidly evolving digital landscape, e-commerce has moved far beyond serving as a simple sales channel and has become a core component of strategic business planning. This study aims to develop machine learning based classification models for order cancellation and return prediction. These models were built using a real-world dataset of 11,000 transaction records collected from an e-commerce company between May 2020 and August 2024. 10,000 records were used for training the models, while the remaining 1,000 were reserved for independent testing. The training dataset was balanced across three classes: completed, canceled, and returned orders. Two distinct modeling approaches were adopted. The first approach utilized a multi-class classification framework to predict all three outcomes in a single model. The second approach employed separate binary classification models for cancellations and returns, allowing for more targeted predictions. To construct these models, five machine learning algorithms were applied: Logistic Regression (LR), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), and Deep Neural Network (DNN). To enhance model performance, four feature selection strategies were evaluated: minimum Redundancy Maximum Relevance (mRMR), Relief-F, F-Classification, and a Hybrid method combining the three. Among these, Relief-F and Hybrid approach yielded the most robust results, particularly when used in conjunction with RF and XGBoost.The results indicated that the binary classification approach produced superior predictive accuracy and sensitivity compared to the multi-class approach. SVM models demonstrated strong performance in binary classification scenarios, whereas LR and DNN models underperformed. Overall, the integration of advanced feature selection techniques with RF and XGBoost significantly improved forecasting capabilities, offering e-commerce businesses a valuable tool for optimizing operational planning and decision-making processes.

Keywords: E-commerce, Order Cancellation and Return, Prediction, Machine Learning, Feature Selection

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

E-ticaret Sektörü için Yapay Zeka Tabanlı Satış İptal/İade Tahminleme Yazılımı

Zehra Sude SARI

Danışman: Prof. Dr. M. Fatih Akay

Bilgisayar Mühendisliği Anabilim Dalı

ÖZ

Günümüzün hızla değişen dijital ortamında e-ticaret, basit bir satış kanalı olmanın ötesine geçmiş ve stratejik iş planlamasının temel bileşenlerinden biri hâline gelmiştir. Bu çalışma, sipariş iptali ve iade tahmini için makine öğrenmesi tabanlı sınıflandırma modelleri geliştirmeyi amaçlamaktadır. Modeller, Mayıs 2020 ile Ağustos 2024 tarihleri arasında bir e-ticaret şirketinden toplanan 11.000 işlem kaydından oluşan gerçek bir veri kümesi kullanılarak oluşturulmuştur. Bu kayıtların 10.000'i model eğitimi için, kalan 1.000'i ise bağımsız testler için ayrılmıştır. Eğitim veri kümesi, tamamlanan, iptal edilen ve iade edilen siparişler olmak üzere üç sınıf arasında dengelenmiştir. İki farklı modelleme yaklaşımı benimsenmiştir. İlk yaklaşım, tüm bu üç sonucu tek bir modelde tahmin etmek amacıyla çok sınıflı sınıflandırma çerçevesini kullanmıştır. İkinci yaklaşım ise, iptal ve iade durumları için ayrı ayrı ikili sınıflandırma modelleri geliştirerek daha hedefli tahminler yapılmasını sağlamıştır. Bu modelleri oluşturmak için beş makine öğrenmesi algoritması uygulanmıştır: Lojistik Regresyon (LR), Rastgele Orman (RF), Aşırı Gradyan Artırma (XGBoost), Destek Vektör Makineleri (SVM) ve Derin Sinir Ağı (DNN). Model performansını artırmak amacıyla dört öznitelik seçme stratejisi değerlendirilmiştir: minimum Artıklık Maksimum İlgililik (mRMR), Relief-F, F-Sınıflandırma ve bu üçünün birleşiminden oluşan Hibrit yöntem. Bunlar arasında, özellikle RF ve XGBoost ile birlikte kullanıldığında, Relief-F ve Hibrit yaklaşımlar en güçlü sonuçları vermiştir. Sonuçlar, ikili sınıflandırma yaklaşımının çok sınıflı yaklaşıma kıyasla daha yüksek doğruluk ve duyarlılık sağladığını ortaya koymuştur. SVM modelleri, ikili sınıflandırma senaryolarında güçlü bir performans sergilerken; LR ve DNN modelleri daha düşük başarı göstermiştir. Genel olarak, gelişmiş öznitelik seçme tekniklerinin RF ve XGBoost ile entegrasyonu, tahminleme yeteneklerini önemli ölçüde artırmış ve e-ticaret işletmeleri için operasyonel planlama ve karar alma süreçlerini optimize etmede değerli bir araç sunmuştur.

Anahtar Kelimeler: E-ticaret, Sipariş İptali ve İade, Tahmin, Makine Öğrenmesi, Öznitelik Seçme

GENİŞLETİLMİŞ ÖZET

Günümüzde dijitalleşmenin etkisiyle birlikte e-ticaret, yalnızca geleneksel satış süreçlerinin dijital ortama aktarılması değil, aynı zamanda tüketici davranışlarının analizine dayalı yeni iş modellerinin geliştirilmesini de içeren stratejik bir alan hâline gelmiştir. Kullanıcıların çevrimiçi alışverişe yönelmesi, işletmelere küresel pazarlara erişim, düşük operasyonel maliyet ve veri odaklı karar verme gibi avantajlar sunarken; bu dönüşüm, aynı zamanda yeni operasyonel zorlukları da beraberinde getirmiştir. Bu zorluklardan en öne çıkanları, sipariş iptalleri ve ürün iade süreçleridir.

Sipariş iptali ve ürün iadesi, e-ticaretin doğası gereği fiziksel deneyim eksikliğinden kaynaklanan başlıca sorunlardandır. Tüketiciler, ürünü teslim almadan önce deneyimleyemedikleri için beklentiyle gerçek arasında oluşan fark, kararlarını değiştirmelerine neden olabilmektedir. Özellikle giyim, ayakkabı, aksesuar gibi kişisel tercihlere bağlı sektörlerde bu durum daha da belirginleşmektedir. Bu davranış biçimi, tüketici haklarının korunmasına yönelik yasal düzenlemelerle birleştiğinde, e-ticaret işletmeleri için ciddi lojistik ve maliyet yükleri oluşturabilmektedir.

Türkiye'de çevrimiçi alışveriş yapan tüketicilere yasal olarak siparişlerini belirli bir süre içinde koşulsuz iade hakkı tanınmaktadır. Bu uygulama tüketici güvenini artırırken, küçük ve orta ölçekli satıcılar için iade süreçlerinin yönetimi hâlâ büyük bir operasyonel problem teşkil etmektedir. Geciken iadeler, stok planlamasında sapmalar, müşteri memnuniyetsizliği ve artan kargo maliyetleri, işletmelerin kârlılığını doğrudan etkilemektedir. Dünyanın diğer bölgelerinde de benzer eğilimler görülmektedir. Avrupa Birliği ülkelerinde iade süreleri ve haklar tüketici lehine düzenlenmiş olsa da, lojistik altyapısı zayıf olan ülkelerde uygulama açısından sorunlar yaşanabilmektedir. Amerika Birleşik Devletleri gibi bazı gelişmiş pazarlarda ise özellikle büyük e-ticaret platformları koşulsuz iade politikaları uygulayarak müşteri memnuniyetini en üst düzeye çıkarmayı hedeflemektedir. Ancak bu yaklaşım, küçük satıcılar için ciddi operasyonel yük ve maliyet anlamına gelmektedir. Küresel ölçekte değerlendirildiğinde, iptal ve iade süreçlerinin yönetimi, hem yasal düzenlemelere hem de teknolojik altyapı kapasitesine bağlı olarak değişkenlik göstermektedir.

Tüm nedenler göz önüne alındığında, sipariş iptali ve ürün iadesi olasılıklarını önceden öngörebilecek yazılımlara olan gereksinim hızla artmaktadır. İptal ya da iade riski yüksek olan ürünleri ve müşteri profillerini önceden tanımlayabilen tahminleme yazılımları; kampanya kurguları, stok optimizasyonu, müşteri ilişkileri yönetimi ve lojistik planlamada önemli ölçüde fayda sağlayabilir. Veri temelli yaklaşımların kullanılması yalnızca operasyonel süreçlerin verimliliğini artırmakla kalmaz, aynı zamanda müşteri memnuniyeti ile marka bağlılığı üzerinde de doğrudan olumlu etkiler yaratabilir. Söz konusu ihtiyaçlara yanıt üretmek amacıyla yürütülen tez çalışması, gerçek e-ticaret verisiyle desteklenmiş makine öğrenmesi tabanlı bir tahminleme altyapısı geliştirmeyi hedeflemektedir.

Çalışma kapsamında, Inveon tarafından sağlanan ve Mayıs 2020 ile Ağustos 2024 tarihleri arasında toplanan 11.000 satırlık sipariş verisi kullanılmıştır. Veri seti; sipariş tarihi, ürün kategorisi, ödeme yöntemi, uygulanan indirim, müşteri iletişim tercihleri, ürün rengi ve bedeni gibi çeşitli niteliklere sahiptir. Siparişlerin tamamlanan, iptal edilen ve iade edilen olmak üzere üç sınıfa ayrıldığı bu veri kümesinde, 10.000 satır model eğitimi, 1.000 satır ise genel performans değerlendirmeleri için test sürecinde kullanılmıştır. Eğitim verisinde sınıf dağılımı dengeli olacak şekilde yapılandırılmıştır.

Modelleme süreci üç ayrı deneysel çalışma çerçevesinde gerçekleştirilmiştir. Birinci deneysel çalışmada, model eğitimi başlamadan önce öznitelik seçimi yapılmış ve ardından 10 katlı çapraz doğrulama uygulanmıştır. İkinci deneysel çalışmada ise öznitelik seçimi, her çapraz doğrulama katı içerisinde ayrı ayrı yürütülmüştür. Bu iki yapı sayesinde, öznitelik seçme işleminin zamanlamasının model performansına etkisi değerlendirilmiştir. Üçüncü deneysel çalışmada ise ilk iki çalışmada elde edilen en başarılı modeller, daha önce hiç kullanılmamış 1.000 satırlık test seti üzerinde değerlendirilerek genellenebilirlik testine tabi tutulmuştur.

Tahminleme sürecinde iki temel yaklaşım uygulanmıştır. İlk yaklaşımda, tüm siparişler tek bir sınıflandırma modeli içerisinde değerlendirilmiş ve aynı anda üç farklı duruma ait tahminleme yapılmıştır. İkinci yaklaşımda ise iptal ve iade süreçleri birbirinden ayrılmış ve her biri için bağımsız tahminleme modelleri geliştirilmiştir. Her iki yaklaşım da eş zamanlı olarak değerlendirilmiş; öznitelik seçimi, başarı metrikleri ve test sonuçları bakımından karşılaştırılmıştır.

Model geliştirme aşamasında LR, RF, XGBoost, SVM ve DNN algoritmaları kullanılmıştır. Bu algoritmaların tahmin performanslarını artırmak amacıyla dört farklı öznitelik seçme yöntemi değerlendirilmiştir: mRMR, Relief-F, F-Sınıflandırma ve bu üç yöntemin birleşiminden oluşturulan Hibrit öznitelik seçme algoritması.

Sipariş iptali ve ürün iadesi gibi kritik süreçlerin tahmin edilmesi, e-ticaret operasyonlarında büyük önem taşımaktadır. Ancak saha uygulamalarında gözlemlenen temel zorluklardan biri, bu süreci yönetecek ekiplerin her zaman ileri düzey kodlama bilgisine sahip olmamasıdır. Teknik bilgi seviyesi sınırlı olan kullanıcıların da bu tahminleme sürecine aktif katılım sağlayabilmesi adına, kod yazma zorunluluğu olmaksızın kullanılabilecek, erişilebilir ve kullanıcı dostu sistemlere duyulan ihtiyaç artmaktadır. Bu ihtiyacı karşılamak amacıyla, hiperparametrik esnekliğe sahip bir arayüz geliştirilmiştir.

Söz konusu arayüz; veri yükleme, öznitelik seçimi, algoritma belirleme, model yapılandırması, tahmin çıktılarının alınması ve başarı metriklerinin görselleştirilmesi gibi işlemleri tek bir platform üzerinde bütünleşik biçimde sunmaktadır. Python diliyle geliştirilen yapı, farklı algoritmalar ve öznitelik seçme yöntemleri için çok sayıda senaryo oluşturulmasına imkân tanımakta; bu sayede deneysel esneklik ve kullanım kolaylığı aynı anda sağlanmaktadır. Arayüzün bu yapısı, modelleme sürecini sadece doğruluk açısından değil, erişilebilirlik ve sürdürülebilirlik yönünden de güçlü hâle getirmiştir.

Tahminleme sürecinin başarısını değerlendirmek amacıyla hata metrikleri hem sınıf bazında hem de genel olarak analiz edilmiştir. Tamamlanan, iptal edilen ve iade edilen siparişler için ayrı ayrı kesinlik, duyarlılık ve F1-skoru değerleri hesaplanmış; böylece modellerin her sınıf üzerindeki başarımı ayrıntılı biçimde incelenmiştir. Buna ek olarak, genel sistem performansını değerlendirebilmek için genel kesinlik, genel duyarlılık, genel F1-skoru ve doğruluk değerleri de göz önünde bulundurulmuştur. Farklı öznitelik seçme yöntemlerinin aynı algoritma üzerindeki etkisi bu metrikler üzerinden karşılaştırılmış ve analiz edilmiştir. Ayrıca, modellerin sınıflar arası ayrım başarısını daha iyi ortaya koymak amacıyla karmaşıklık matrisi sonuçları da değerlendirme sürecine dahil edilmiştir. Bu çok yönlü yaklaşım sayesinde, modellerin hangi koşullarda daha etkili sonuçlar verdiği net biçimde ortaya konmuştur.

İlk iki deneysel çalışmanın karşılaştırmalı analizine göre, iptal ve iade durumlarının ayrı ayrı ele alındığı ikili sınıflandırma yaklaşımı, çok sınıflı yapıya kıyasla daha yüksek doğruluk ve duyarlılık değerleri sunmuştur. İkili yapıda geliştirilen modeller, özellikle problem odaklı tahminleme senaryolarında daha etkili ve uygulamaya dönük sonuçlar vermiştir. Her iki yaklaşımda da RF ve XGBoost tabanlı modeller, tutarlı ve yüksek başarı metrikleriyle öne çıkmıştır. SVM tabanlı modeller ikili yapılarda tatmin edici sonuçlar verse de, çok sınıflı yapılarda görece geride kalmıştır. DNN ve LR tabanlı modeller ise genel olarak daha düşük başarı göstermiştir.

Modelleme sürecinde kullanılan dört farklı öznitelik seçme yönteminin etkileri de ayrı ayrı analiz edilmiştir. Relief-F ve Hibrit yöntem, özellikle RF ve XGBoost tabanlı modellerle kullanıldığında güçlü ve tutarlı sonuçlar üretmiştir. Relief-F yöntemi, her iki sınıflandırma yaklaşımında da iptal ve iade tahmininde F1 skoru açısından en yüksek düzeylere ulaşmıştır. F-Sınıflandırma yöntemi de benzer şekilde başarılı olmakla birlikte, DNN ve SVM tabanlı modellerle kullanıldığında daha düşük performans sergilemiştir. mRMR yöntemi ise tüm senaryolarda en düşük başarıyı göstermiş, bazı durumlarda öznitelik seçimi yapılmayan modellere kıyasla daha zayıf sonuçlar vermiştir.

Üçüncü deneysel çalışmada, önceki iki çalışmada en yüksek başarıyı gösteren modeller, bağımsız bir test veri seti üzerinde yeniden değerlendirilmiş ve benzer başarı düzeylerini korumuştur. Bu sonuçlar, önerilen sistemin genellenebilirliğini ve gerçek dünya uygulamaları için uygunluğunu ortaya koymuştur. Özellikle RF ve XGBoost tabanlı modellerin kararlı performansı, bu iki yöntemi iptal ve iade tahminleme süreçlerinde öne çıkan teknikler haline getirmiştir. Elde edilen bulgular, doğru model seçimi ve etkili öznitelik seçiminin, tahminleme sistemlerinin başarısında belirleyici olduğunu açıkça göstermektedir.

Ayrıca, öznitelik seçme işleminin ne zaman yapıldığının model başarısı üzerinde anlamlı bir etkisi olup olmadığı da değerlendirilmiştir. Birinci deneysel çalışmada öznitelik seçimi çapraz doğrulama öncesinde uygulanırken; ikinci çalışmada her bir kat içerisinde bağımsız olarak gerçekleştirilmiştir. Elde edilen sonuçlar, bu iki strateji arasında başarı metrikleri açısından belirgin bir fark olmadığını ortaya koymuştur. Bu bulgu, öznitelik seçme adımının konumundan ziyade kullanılan yöntemin ve modelin başarısı üzerinde daha belirleyici olduğunu göstermektedir.

Bu tez çalışması, e-ticaret işletmeleri açısından doğrudan uygulanabilir bir tahminleme altyapısı sunması bakımından büyük önem taşımaktadır. Geliştirilen yazılım sayesinde, işletmeler sipariş iptali veya iadesi gibi maliyetli süreçleri daha önceden öngörebilecek; bu da hem kaynak planlamasının etkinliğini artırmakta hem de müşteri deneyimini daha sürdürülebilir hâle getirecektir. Modüler yapıya sahip arayüz ve farklı algoritmalarla test edilmiş tahminleme modelleri, farklı ölçeklerdeki firmalar için esnek ve uyarlanabilir çözümler sunma potansiyeli taşımaktadır. Elde edilen bulgular, veri temelli karar destek sistemlerinin operasyonel verimlilik ve müşteri memnuniyeti açısından sunduğu katkıyı net biçimde ortaya koymaktadır. Bu bağlamda, tez çalışması yalnızca akademik değil, aynı zamanda sektörel anlamda da yüksek bir katma değer sunmaktadır.

Gelecek çalışmalarda, sınıf dengesizliğinin daha belirgin olduğu senaryoların değerlendirilmesi önemli bir araştırma alanı olarak öne çıkmaktadır. Bu tez çalışmasında kullanılan veri kümesinde tamamlanan, iptal edilen ve iade edilen siparişler dengeli şekilde dağıtılmıştır. Ancak gerçek dünya e-ticaret verilerinde iptal ve iade oranlarının genellikle oldukça düşük olduğu bilinmektedir. Bu nedenle, sınıf yapısı dengesiz olan gerçek veri setleriyle yeniden modelleme yapılması; sınıf dengesizliğinin model performansı üzerindeki etkilerini daha sağlıklı analiz edebilmek adına faydalı olacaktır. Ayrıca, topluluk öğrenmesi gibi ileri düzey modelleme yaklaşımlarının test edilmesi, tahmin başarımını ve geliştirilen sistemin sağlamlığını daha da artırma potansiyeline sahiptir.

ACKNOWLEDGEMENTS

I express my deepest gratitude to my supervisor, Prof. Dr. M. Fatih AKAY, for his invaluable guidance, insightful feedback, and continuous support throughout the preparation and completion of this thesis. His academic expertise and mentorship were essential in shaping the direction of this study.

I would also like to thank Prof. Dr. İbrahim Taner OKUMUŞ and Asst. Prof. Dr. Barış ATA for their valuable suggestions and constructive contributions, and for serving as members of my thesis committee.

Furthermore, I sincerely thank Inveon for providing access to real-world e-commerce data, which played a critical role in the empirical analysis and validation of the proposed models.

LIST OF TABLES

LIST OF FIGURES

SYMBOLS AND ABBREVIATIONS

AdaBoost : Adaptive Boosting
ANN : Artificial Neural Network
ANOVA : Analysis of Variance
AR : Augmented Reality
B2B : Business to Business
B2C : Business to Consumer
B2G : Business to Government
C2B : Consumer to Business
C2C : Consumer to Consumer
CSCL : Class-Supervised Contrastive Learning
CNN : Convolutional Neural Networks
DBSCAN : Density-Based Spatial Clustering of Applications with Noise
DeepOCP : Deep Order Cancellation Prediction Model
DNN : Deep Neural Network
DT : Decision Tree
EDA : Exploratory Data Analysis
EDI : Electronic Data Interchange
GBRT : Gradient Boosting Regression Trees
G2C : Government to Citizen
HC : Hierarchical Clustering
HyperGo : Hypergraph-based Local Graph Cut
ICC : Initial Collection Centers
K-NN : Nearest Neighbors
LASSO : Least Absolute Shrinkage and Selection Operator
LoGraph : Local Graph Algorithm
LR : Logistic Regression
MAD : Mean Absolute Deviation
MAE : Mean Absolute Error
MILP : Mixed Integer Linear Programming
MLP : Multi-Layer Perceptron
mRMR : Minimum Redundancy Maximum Relevance
PIA : Permutation Importance Analysis

R^2 : Correlation Coefficient

RAE : Relative Absolute Error

RF : Random Forest

RL : Reverse Logistics

RMSE : Root Mean Square Error

RRSE : Root Relative Squared Error

SFS : Sequential Forward Selection

SHAP : SHapley Additive exPlanations

SVM : Support Vector Machine

TCN : Temporal Convolutional Network

VMD-LSTM : Variational Mode Decomposition - Long Short-Term Memory

VR : Virtual Reality

XGBoost : Extreme Gradient Boosting

1. INTRODUCTION

Historical development of electronic commerce is directly related to the transformation in information and communication technologies. The foundations of electronic business models were laid in the 1960s with the use of Electronic Data Interchange (EDI) systems. Thanks to these systems, it became possible to share commercial documents digitally, and data transfer between businesses significantly improved in terms of both speed and accuracy. The infrastructures that emerged during this period formed the first forms of digital commerce (Gian & Ikate, 2021).

Following this early foundation, the 1980s and 1990s marked a significant leap in both technological capabilities and public adoption. With the widespread use of computers and the development of electronic payment systems in the 1980s, significant progress was made in the transfer of commerce to the digital environment. Basic commercial activities such as banking, invoice management and order transactions were transferred to the electronic environment. However, the massification of e-commerce became possible in the 1990s with the widespread use of the Internet and the opening of the World Wide Web to the public.

Pioneer online platforms such as Amazon and eBay, established during this period, demonstrated the commercial potential of the Internet and popularized online shopping; users acquired the habit of purchasing products and services without going to physical stores. At the same time, the development of digital payment solutions and secure connection protocols increased users' trust in e-commerce. Thus, commercial interaction in the digital environment became more attractive for both consumers and businesses.

Over time, this transformation process of e-commerce was not limited to large enterprises; it spread on a global scale with the integration of small and medium-sized enterprises into this system. Today, e-commerce is not only a sales channel, but also an integral part of many business functions such as global marketing, supply chain management and customer relations.

In parallel with these developments, the rapid spread of Internet technologies continued to reshape the structure of commerce. E-commerce emerged as a new form of commerce in the 1990s, coinciding with the rapid spread of Internet technologies. The increasing accessibility of the Internet for individual users paved the way for the sale of products and services through digital platforms. In its initial phase, e-commerce was regarded as an experimental domain; users were cautious about engaging with such systems due to concerns regarding payment security and reliability. Nevertheless, companies like Amazon and eBay, founded during this period, showcased the transformative potential of online shopping (Charles Ntumba et al., 2023).

As the e-commerce infrastructure matured, key enhancements in usability and consumer trust further accelerated its growth. Although early e-commerce activities were largely based on simple interfaces, limited payment infrastructure, and low-level user interaction, these structures developed rapidly over time; with advancements in technology, more secure, faster and user-friendly systems began to emerge. The use of secure payment protocols (e.g. SSL) and the spread of digital wallets such as PayPal have increased consumers' trust in the system; retailers have also started to invest more in online platforms.

One of the factors that provided the greatest momentum in the transformation in the retail sector has been the rapid development of mobile technology. The increase in the use of smartphones and tablets has made e-commerce accessible from anywhere, at any time, rather than limited to desktop computers. This has also changed user habits to a great extent; shopping has become a faster, more spontaneous and personalized process. Mobile-compatible interfaces, instant notifications, easy payment systems and location-based campaigns have digitized the retail experience, providing a more direct and interactive connection with the consumer (Gouveia & Mamede, 2022).

Along with mobile technologies, experiential technologies also began to play a more central role in shaping consumer expectations. Developing technology has not only facilitated access, but also made the shopping experience more impressive and multi-dimensional. Augmented Reality (AR) and Virtual Reality (VR) technologies have found application areas especially in the furniture, decoration and clothing sectors with the concept of “try before you buy”. In this way, users can experience how they will see the product in their homes or on themselves in a virtual environment without physically interacting with it. Such applications both enrich the user experience and contribute to the operational efficiency of businesses by reducing return rates.

The rise of social media platforms has also been an important turning point in the evolution of e-commerce. Consumers now discover products not only from e-commerce sites but also through social media; brands are present in these areas where users are present with their content and advertising strategies. This transformation can be read as part of a transition from traditional advertising to digital marketing. Consumer reviews, user-generated content and social proof are among the main factors that shape purchasing decisions today (Charles Ntumba et al., 2023).

In addition to these digital communication and interaction dynamics, e-commerce also began to reshape the very boundaries of markets. The growth of e-commerce has eliminated geographical boundaries. Thanks to cross-border trade, small and medium-sized businesses can reach not only local but also global markets. While this creates new market opportunities, it has also brought about new problem areas such as logistics, customs, taxation and legal regulations. For this reason, it has become important for businesses to approach digitalization processes with a holistic perspective — not only technically, but also operationally (Gouveia & Mamede, 2022).

Data analytics and artificial intelligence systems enable e-commerce to become a strategic management tool rather than just a sales channel. Insights obtained by analyzing user data enable more accurate critical decisions such as customer segmentation, stock optimization and dynamic pricing. At the same time, thanks to personalization applications, recommendation systems specific to each user can be developed, which both increases sales and supports user loyalty.

Ultimately, all these technological and behavioral shifts have redefined the e-commerce ecosystem. It has deeply affected the shopping behavior of not only retailers but also consumers. It is expected that this transformation will gain different dimensions with the new tools that technology will offer in the future.

In light of this evolution, it is evident that e-commerce has gradually been divided into different types and has taken on a more systematic and functional structure. The most common types of e-commerce are explained below in line with these structures.

1.1. Types of E-Commerce

As e-commerce evolved into a more structured and functional domain, it has been categorized based on the nature of the transaction and the parties involved. These classifications help to understand the dynamics and business models that govern digital commerce in various contexts (Gian & Ikate, 2021).

One of the most prevalent forms is Business to Consumer (B2C), in which companies sell directly to end users through platforms like Amazon, Trendyol, or Hepsiburada. Another major model is Business to Business (B2B), where commercial transactions occur between businesses—commonly manufacturers, wholesalers, and suppliers—usually under long-term agreements. The Consumer to Consumer (C2C) model, enabled by marketplaces such as Sahibinden.com or eBay, facilitates individual users to sell products or services to each other.

Additionally, the Consumer to Business (C2B) model allows individuals (e.g., freelancers, influencers) to offer value or services to companies, while Business to Government (B2G) and Government to Citizen (G2C) structures reflect the digitization of procurement and public services. With the emergence of omnichannel strategies, hybrid forms of these models have also surfaced, blurring the lines between traditional categories and adapting to the evolving needs of consumers and businesses alike.

1.2. Advantages and Limitations of E-Commerce

The widespread adoption of e-commerce has introduced significant benefits for both businesses and consumers. Its most remarkable advantage is the ability to operate beyond the limitations of time and space—allowing users to shop at any moment from any location, and businesses to reach customers far beyond their geographical boundaries. Operational costs are reduced through the elimination of physical retail space, and customer data can be analyzed to create personalized offers and targeted marketing strategies (Taher, 2021).

However, this convenience comes with substantial operational and strategic challenges. One of the most critical limitations lies in sales cancellation and product return processes. Unlike traditional retail, the inability to physically experience products before purchase often leads to a mismatch between expectation and reality, resulting in higher return rates. These returns, in turn, increase logistics costs, disrupt inventory planning, and directly affect business profitability. Moreover, the ease with which customers can cancel orders introduces volatility into demand forecasting and complicates supply chain coordination.

Security concerns also persist in online environments, especially related to the storage and transmission of personal and financial information. Despite improvements in encryption and fraud prevention technologies, some users remain reluctant to share sensitive data online. Thus, while e-commerce continues to offer flexibility and scalability, the sector must continually address the issues of trust, reverse logistics, and customer satisfaction to ensure sustainable growth (Taher, 2021).

1.3. Cancellation and Return Processes: A Comparison between Türkiye and International Markets

Among the structural limitations discussed above, the issue of order cancellations and product returns represents one of the most operationally challenging aspects of modern e-commerce. As purchasing decisions increasingly shift to online channels, where consumers cannot physically evaluate products prior to purchase, return rates have climbed considerably, placing growing pressure on logistics networks, inventory systems, and customer service operations.

These challenges manifest differently across national and international markets due to variations in legal regulations, consumer rights protections, logistics infrastructure, and platform policies. In Türkiye, return and cancellation policies are regulated under the Consumer Protection Law, which grants consumers the right to withdraw from a purchase within 14 days of delivery without stating a reason. This legal assurance supports consumer confidence, but in practical terms, smaller-scale sellers often struggle with the return management process, leading to issues such as delayed refunds, miscommunication, and lack of integration with logistics providers.

In contrast, return policies in international markets exhibit a broader diversity in both structure and implementation. For instance, within the European Union, the 14-day withdrawal right is uniformly guaranteed by law, but the execution and enforcement of this right can differ by country. Nations with strong consumer rights traditions—such as Germany or the Netherlands—typically provide easy return mechanisms, whereas others with weaker logistical capacity may create friction in the process.

In the United States, major e-commerce platforms such as Amazon and Walmart have pioneered liberal return policies, often offering unconditional returns and fully automated workflows. While this model enhances customer satisfaction, it also imposes significant financial strain on small-to-medium-sized businesses, which face challenges in absorbing return costs, restocking fees, and operational inefficiencies.

Additionally, cross-border returns present a separate set of complications. Tariffs, customs clearance delays, currency fluctuations, and language barriers can make return processing both expensive and time-consuming. For instance, a product returned from Europe to a Turkish seller may be delayed at customs or incur return shipping charges that outweigh the product's value.

While Türkiye benefits from a clear legal framework, practical challenges in logistics and customer communication often hinder seamless returns. International markets, on the other hand, may offer more flexible systems but must contend with cost-related pressures and cross-cultural operational differences. In both contexts, the efficient management of cancellations and returns has become a critical strategic factor in achieving customer satisfaction and maintaining business sustainability.

The primary objective of this thesis is to develop a machine learning-based software solution that predicts order cancellations and product returns for e-commerce companies and/or marketplace platforms. This software is designed to support businesses in executing operational planning based on data-driven insights, enhancing customer satisfaction, and managing inventory and logistics processes more efficiently.

1.4. Contributions of This Thesis

This thesis offers several innovative contributions to the e-commerce sector, which can be summarized as follows:

1. Dual Strategy with Multi-Class and Binary Classification Models: Two distinct classification strategies have been employed for predicting cancellations and returns. In the first approach, the cancellation-completion-return statuses are modeled within a single framework as a multi-class classification problem. In the second approach, two separate

binary classification models have been developed independently for cancellations and returns.

2. Item-Level Prediction Modeling: Rather than general order-level forecasting, the system performs forecasting at the individual product level within each order. This allows businesses to conduct product-level risk analysis and implement more flexible decision-support mechanisms.
3. Advanced Feature Selection: In addition to applying well-established feature selection algorithms—mRMR, Relief-F, and F-Classification—a new Hybrid feature selection algorithm has been developed based on the consensus of the outputs generated by these three methods.
4. Development of an Interface with Hyperparametric Flexibility: A graphical user interface has been developed to allow users to perform essential modeling processes—such as data uploading, feature selection, model configuration, hyperparameter tuning, and performance analysis—within a unified and user-friendly environment.
5. Modeling with an Extended Feature Set: Models have been developed using a rich dataset that includes customer behaviors, order history, product attributes, and campaign-related information. This multidimensional feature set has demonstrated improved predictive accuracy compared to conventional models.

This thesis is structured into six concise chapters, each addressing a specific dimension of the research. Chapter 1 provides an overview of the e-commerce sector; explains the problem identified around order cancellations and returns, and highlights the operational and strategic impacts of these situations on businesses. Chapter 2 provides a comprehensive literature review focusing on existing studies related to cancellation and return prediction, identifying methodological gaps in the current body of knowledge and positioning the present study accordingly. Chapter 3 describes the technical methodology, detailing the machine learning algorithms, feature selection techniques, and the structure of the dataset; it also introduces a custom-developed interface designed with hyperparametric flexibility to enable complete end-to-end modeling. Chapter 4 explains the model configuration process, including hyperparameter optimization and performance evaluation metrics such as accuracy, precision, recall, and F1-score, and presents a comparative analysis of model performance across different experimental settings. Chapter 5 synthesizes the results from the experimental studies, discusses their practical implications, and provides recommendations for future research and directions for system improvement. Chapter 6 compiles all academic references cited throughout the study, serving as a valuable resource for further exploration of the subject matter.

2. LITERATURE REVIEW

In this section, various academic studies on order cancellation and product return prediction in e-commerce will be examined in detail. This review aims to identify trends in literature and provide a framework for future research.

(Clement and Spinler, 2025) proposed a machine learning-based model to predict the probability of return of orders in the e-commerce fashion sector. The main goal of the study is to reduce both environmental impact and cost by associating return estimates with disposable and reusable packaging options. For this purpose, more than 1.9 million transaction data of a large fashion retailer in Europe were used. L1 regularization, Shapley values (SHAP), XGBoost gain, and Sequential Forward Selection (SFS) methods were compared for feature selection; XGBoost algorithm showed the best classification success. Model results revealed that integrating return estimates into the packaging strategy can increase the reusable packaging usage rate by up to 20% without increasing the cost per shipment.

(Zhang et al.,2025) developed an innovative approach using publisher behavior features extracted from three different modes (visual, auditory, and linguistic) to predict product return rates in live broadcast e-commerce. This model, called ContraFormer, is developed on a Transformer structure with an Encoder-Decoder architecture and includes the Class-Supervised Contrastive Learning (CSCL) technique. The model was tested with 2584 product live broadcasts and 864 different product data obtained from the TikTok China platform. ContraFormer achieved a 25% reduction in Mean Absolute Error (MAE) in product return prediction compared to existing time series models and multimodal fusion approaches. This approach was particularly effective in performing cross-modal analysis of publisher behaviors and modeling behavioral consistency.

(Karl, 2024) examined the current state of the art of algorithms for predicting product returns in e-commerce and aimed to expand the body of knowledge in this area by combining information from previous reviews. First, four different reviews were synthesized and then the literature on product returns was enriched by analyzing 28 additional publications on consumer returns. This review was the first to include an e-commerce perspective and highlight advanced algorithms and metrics for return predictions.

(Sharma, 2024) analyzed customer behaviors along with the increasing online product returns of developments in the e-commerce sector. In the study, a dataset consisting of 100,000 orders and 13 attributes was used. In order to classify and predict whether the customer will return the product, models such as LR, Adaptive Boosting (AdaBoost), Decision Tree (DT), Naive Bayes, XGBoost, K-Nearest Neighbors (K-NN) and RF were applied, and their performances were compared. Customer segmentation was performed using Logistic Regression and the classification threshold method, and customers were divided into high and low return risk categories. The results obtained show that developing strategies can contribute to reducing online product return rates.

(Reis, 2024) analyzed the product return rate of an e-commerce company and examined the financial impact of returns in line with the company's loss and profit data. Four different machine learning prediction models, namely the HP filter, the MED filter, the AR model, and the Theta model, were used to develop strategies to reduce return rates. The results revealed that the Theta model exhibited the highest performance. It was also determined that profitability would decrease if a strategy to reduce the amount of returned products was not implemented.

(Adebanjo et al., 2023) analyzed order cancellation behaviors and stock policies within the scope of the cash on delivery system. In the study, 315 survey data targeting orderers in the United Kingdom were used and evaluated with three different machine learning models. According to the analysis results, the Logistic Regression (LR) model showed the highest performance with 84% prediction accuracy and 69% precision rate. LR analysis revealed that customer order cancellations have a negative impact on the seller's optimal pricing and stock policy.

(Khouja and Hammami, 2023) developed a model for online retailers to optimize pricing, stock quantity and product return policies in the face of opportunistic consumer behavior (wardrobing). The study examined the effect of in-season price discounts to reduce inventory remaining at the end of the season. In addition, credit card and gift card (Store Credit or Gift Credit - SC/GC) cash return policies were compared, and it was determined that SC/GC return was more advantageous for e-retailers. The results show that costly returns were reduced, and sellers gain financial advantage with the application of the model.

(Abdulla et al., 2023) aimed to reduce cancellation rates without increasing return rates by analyzing the factors affecting order cancellation behavior in e-commerce. In the study, product and order process specific factors of order cancellations were examined using both econometric and quasi-experimental methods using a transactional data set. In addition, the effects of tightening order cancellation policies were evaluated. The results showed that changes in order cancellation policies provided a significant decrease in order cancellation rates without increasing return rates.

(Sun et al., 2023) analyzed the relationship between the cancellation probability of online taxi orders and user behavior. The study determined the factors that cause users to cancel taxi orders and used a deep learning-based model for the first time for this purpose. The Deep Residual Network-Based Ride-Hailing Order Cancellation Probability (DeepOCP) model was developed, and its effectiveness was tested with the Didi Chengdu Express public dataset. It was also stated that this study was the first research using a deep learning model to predict the probability of online ride-hailing orders.

(Thibbotuwawa et al., 2023) proposed a Reverse Logistics (RL) network model to optimize returns management in e-commerce. The study presented an approach that guides e-commerce companies to efficiently manage their RL networks and develop integrated decision-making processes for product returns. In the model, the hierarchical clustering method was applied to determine return trends and mixed integer linear programming was used. The decision variables included the optimal return volumes to be directed from Initial Collection Centers (ICC) to recycling centers and the relevant logistics parameters. The effectiveness of the model was tested with a case study on an e-commerce company, and it was found that it provided 39.9% cost savings compared to traditional RL network operations.

(Xu et al, 2023) designed a multimodal analytical framework to predict product returns in live broadcast e-commerce environments. In the study, experiments were conducted using a real-world dataset consisting of 5,664 products collected from Taobao live broadcast. Manual combinations of features in different formats (text, image, video) were made and the effects of these features on the classification performance were investigated. Experimental results showed that the use of multimodal signals from products and servers was effective in predicting product returns.

(Sweidan et al, 2023) addressed the problem of data-driven product return prediction using a large-scale real-world dataset. In the study, two different datasets were used: The first dataset contained 20,262 orders and 38 attributes of new customers, and the second dataset contained 15,738 orders and 56 attributes of customers with at least one order. In the study, 10 attributes that had the most impact on the classification performance were determined and analyzed. The results showed that well-calibrated probabilistic estimators achieved high precision and reasonable recall values.

(Dzyabura et al, 2023) aimed to predict product return rates before launch, in order to decide whether the product should be included in the retailer's range. The study conducted analyses on identifying and designing products with low probability of return using machine learning and image-based interpretable features. The results showed that retailers can increase their profitability by 8.3% by selecting products with low probability of return features.

(Şahinbaş, 2022) proposed a model to forecast the tendency to cancel orders and the parameters affecting order cancellation. In the study, a dataset consisting of 20 features, including 30,321 customers and orders, was used. The importance levels of the features were determined by correlation analysis and the 10 most important features were selected. Machine learning techniques such as Artificial Neural Network (ANN), SVM, and RF were used for order cancellation forecasting. As a result of the analyses, the RF model exhibited the best performance with 86% accuracy and an 88% F1-Score.

(Tüylü and Eroglu, 2022) aimed to forecast the product return rate using ensemble learning algorithms such as voting, bagging, stacking, and random subspace. The study used a dataset containing

9,106 orders and 21 attributes. The importance levels of the attributes were determined and analyzed by applying correlation analysis. As a result of the experiments, high-performance results were obtained with stacking and voting algorithms. When the results were compared, it was observed that a correlation coefficient of 83.89% was reached.

(Rajasekaran and Priyadarshini, 2021) aimed to develop a prototype model to predict the return rate of any product. Tests were conducted on different machine learning algorithms to optimize the performance of the proposed model. The results showed that the proposed model has the potential to support revenue growth and reduce return/loss rates in the e-commerce sector.

(Granov, 2021) aimed to predict the purchase, return and abandonment behavior of customers. A dataset consisting of 1,355,533 customers and 26 attributes were used in the study. A model was developed that classifies customers as losers, returners and loyal customers, and this classification was carried out in line with exploratory data analysis, the company's prior knowledge and expert opinions. Attribute selection was made using correlation analysis and distribution analysis methods. The proposed model classified losers, returners and loyal customers with accuracy rates of 0.68, 0.75 and 0.98, respectively.

(Chidroop and Moharir, 2020) aimed to predict the tendency of a customer to cancel an order. In the study, order data and historical data of customers were used, and machine learning algorithms were applied to these datasets. In the analysis results, it was determined that the DT model exhibited the best performance in accuracy, precision, recall and F1-Score metrics with rates of 93.1%, 91.6%, 92.6% and 92.1%, respectively.

(Cui et al., 2020) aimed to forecast the return volume at the period level using various machine learning methods. A dataset containing 331,390 products and 19 attributes was used in the study. Experimental results showed that the Least Absolute Shrinkage and Selection Operator (LASSO) model exhibited the best performance compared to other models.

(Imran and Amin, 2020) aimed to estimate product returns for e-commerce and identify the most important features affecting returns to address the return problem. Forecasts were developed in this direction using four different data mining techniques. A dataset containing 3,070 orders and 15 attributes was used in the study. The importance levels of the attributes were determined using the TabNet algorithm, and the 5 most important attributes were selected.

(Kedia et al., 2019) proposed a DNN -based approach for return forecasting. In the study, users' tastes and product features were modeled using Bayesian personalized ranking-based product embedding vectors. In addition, a separate set of embedding vectors representing users' body shape and size was created with a skip-gram-based model. The DNN model was integrated to forecast the probability of return using the obtained vectors. New features were derived by applying feature engineering on the dataset consisting of 24 features, and the most important features were determined to enhance model performance.

(Joshi et al., 2018) aimed to develop a product return prediction model based on product size, user and brand using historical data of one of the major e-commerce companies in India. Network science and machine learning techniques were employed in the study. In the analysis conducted across three main sectors of different sizes, the results obtained using the simple Random Walk (RW) method showed an improvement in the F-score ranging from 10% to 25% compared to the initial data.

(Li et al., 2018) proposed the HyperGo framework to forecast whether the customer will return the product after creating the shopping cart. HyperGo models past purchase and return records with a new hypergraphic representation method by utilizing the cart composition information. It was also stated that HyperGo was suitable for big data processing. The effectiveness and efficiency of HyperGo were verified in the experimental analysis performed using e-commerce datasets.

(Zhu et al., 2018) proposed a weighted hybrid graph model representing information from purchase and return history to forecast product returns. This graph model consists of customer nodes, product nodes, undirected edges reflecting customer return history, directed edges that differentiate between actions that result in a purchase and those that do not, and feature-based connections that assess the similarity between a customer and a product. Experiments were conducted on e-commerce datasets to evaluate the performance of the proposed graph, and the effectiveness of the model was tested.

Table 1.1**.** Literature review

Study	Methods	Dataset Size	Number of Features	Feature Selection	Performance Metric
Clement and Spinler, 2025	XGBoost, ANN, LR	Train: 1,942,028 Test: 215,781	104	L1 Regularization, XGBoost Gain, SHAP, SFS	ROC AUC: up to 83.5% Accuracy: up to 75.6%
Zhang et al., 2025	ContraFormer (Transformer + CSCL)	2,584 live-stream sessions / 864 products	Visual (6), Auditory (3), Textual (3)	Permutation Importance Analysis (PIA)	25% reduction in MAE, improved R^2
Sharma, 2024	LogR, DT, RF, NB, AdaBoost, XGBoost, KNN	100,000 order	13	The importance level of the features was determined with RF and then the 10 most important eatures were selected.	Accuracy values LR = 0.76 DT = 0.69 NB = 0.77 AdaBoost = 0.77 XGBoost = 0.77 RF = 0.73 KNN = 0.72

Table 1.1 (continued)

Adebanjo et al., 2023	LogR, RF, SVM	Random selected 300 customers	6	-	Accuracy, Preccision and Recall values respectively LR = 0.84, 0.69, 0.79 RF = 0.78,0.68, 0,.75 SVM = 0.79, 0.69, 0.57
Sun et al., 2023	ARIMA, SVR, RNN, TCN, DBSCAN, VMD-LSTM, DMVST-Net, DeepOCP	-	4	The importance level of the features was determined with RF.	MAE, MAPE and RMSE values respectively ARIMA = 20.1, 21, 26 SVR = 10.1, 9.8, 13.3 RNN = 8.3, 8.6, 13.1 TCN = 6.3, 7.9, 9.5 DBSCAN = 6.1, 7.8, 9.6 VMD-LSTM = 4.4, 4.9, 6.8 DMVST-Net = 3.5, 4.4, 5.5 DeepOCP = 3, 3.6, 5.0
Thibbotuwawa et al., 2023	Hierarchical Clustering (HC), Mixed Integer Linear Programming (MILP)	Sales data in 2,484 cities	-	-	Cost Savings = 39.9% savings achieved.
Xu et al, 2023	-	5,664 product	-	By manually combining features in different formats (text, image, video), the effects of features on classification performance were investigated.	-
Sweidan et al, 2023	LogR, GB, Platt Scaling GB (GBCall)	First dataset (new customers): 20,262 orders Second dataset (customers with at least 1 order): 15,738 orders	First dataset: 38 Second dataset: 56	The 10 features that most affect the classification performance have been determined.	Accuracy, AUC, Precision and Recall values respectively First Dataset LR = 0.74, 0.71, 0.71, 0.29 GB = 0.74, 0.72, 0.72, 0.28 GBCall = 0.74, 0.71, 0.71, 0.30 Second Dataset LR = 0.71, 0.73, 0.67, 0.43 GB = 0.73, 0.77, 0.69, 0.50

Table 1.1 (continued)

					GBCall = 0.72, 0.77, 0.69, 0.47
Dzyabura et al, 2023	GBRT, LASSO, CNN (ResNet-152), SHAP	1,231,055 order, 4,585 products	2.048 CNN based + traditional and human labeled features.	LASSO, SHAP	For CNN + GBRR Model Out-of-Sample R^2 = %46.88 Mean Absolute Deviation (MAD) = 8.12
Şahinbaş, 2022	SVM, XGBoost, RF, ANN	30,321 customers and orders	20	The importance of the attributes was determined using correlation analysis and then the 10 most important attributes were selected.	F1 Scores RF - Tuned RF = 0.88, 0.87 XGBoost - Tuned XGBoost = 0.88, 0.88 LR - Tuned LR = 0.85, 0.84 ANN = 0.87 SVM = 0.87
Tüylü and Eroglu, 2022	M5P, M5Rules, Decision Table, Linear Regression, SMOreg	9,106 order	20	The importance level of the features was determined by correlation analysis.	Stacking model R ≈ 0.86 MAE ≈ 0.01 RMSE ≈ 0.01
Granov, 2021	K-Prototypes (Clustering) + Bias-Reduced LR (Classification)	1,355,533 customers	25	Company Insights + Exploratory Data Analysis (EDA)	Accuracy values Churn: 0.86 Return: 0.75 Loyalty: 0.98
Chidroop and Moharir, 2020	Boosted DT, Decision Forests, NN, Locally Deep SVM	88,612 orders	~14 (order + customer history combined)	-	Accuracy = 0.93 Precision = 0.92 Recall = 0.93 F1 Score = 0.92 (BDT-Based Model)
Cui et al., 2020	LASSO, Elastic Net, SCAD, RF, GB	331,390 transactions	1.717	19 features were selected with LASSO.	R^2 = 0.93, MSE = 177.8
Imran and Amin, 2020	DT, XGBoost, LightGBM, CatBoost, TabNet	10,000 orders (resampled: 3,070)	15	The importance of the features was determined using TabNet and then the 5 most important features were selected.	TPR: 0.83 F2-Score: 0.76 AUC: 0.67 (TabNet-Based Model)
Kedia et al., 2019	DNN, GB, BPR Embeddings, Skip-gram Sizing Vectors	-	24	New features were derived by feature engineering, and then	AUC = 0.83 Precision = 0.74 Recall = 0.34 (Deep NN model)

Table 1.1 (continued)

				important features were determined to increase model performance.	
Joshi et al., 2018	Network-based clustering (Random Walk, FastGreedy, Infomap, WalkTrap, Node2Vec) + SVM	-	-		F1-score increase: 10–25% over baseline (Best Combo Model)
Li et al., 2018	Hypergraph-based Local Graph Cut (HyperGo) + Dual-Level Return Prediction	3,600,000 baskets, 557,000 products	-	Features were constructed based on basket-level similarity and clustering of historical return patterns.	Dataset A (North America): AUC: 0.70 Precision: 0.81 F0.5:0.74 Dataset B (Europe): AUC: 0.72 Precision: 0.68 F0.5:0.58
Zhu et al., 2018	Hybrid Graph (HyGraph) + Local Graph Algorithm (LoGraph)	Brand C: 161,000 products Brand E: 364,000 products Brand G: 33,000 products	-	Features were constructed from the hybrid graph using purchase-return edges and product similarity weights.	Precision values Brand C: 0.82 Brand E: 0.87 Brand G: 0.76

3. DATASET GENERATION

Within the scope of the thesis study, a dataset has been created using customer shopping data belonging to Inveon, an e-commerce platform provider company, between May 6, 2020 and August 6, 2024, and has been used for forecasting. The dataset contains various demographic and transaction-based attributes at the product, order and customer levels. The data set's attributes are listed in Table 3.1 along with descriptions of each.

Table 3.1. Attributes and Descriptions of Dataset

Attribute Name	Description
OrderId	Unique order ID
OrderNumber	Order number
OrderItemId	Unique order item ID
ProductId	Unique product ID
CategoryId	Unique product category ID
OrderPrice	Total amount of the order
PaymentMethodSystemName	Payment method used
OrderPlatform	Platform where the order was placed
OrderCounty	Order delivery county
OrderStateProvince	Order delivery province
OrderCreationDate	Order creation date (UTC)
CampaignUsed	Whether a campaign/discount was used
GenderLabel	Customer gender
CustomerAddress	Customer address
NumberOfOrdersCustomerCancelled	Number of orders canceled by the customer
NumberOfOrdersCustomerReturned	Number of orders returned by the customer
NumberOfOrdersCustomerCompleted	Number of orders completed by the customer
OrderItemStatus	Cancellation or return status of the order item
ProductPrice	Unit price of the product
IsProductDiscounted	Whether the product is on sale
CategoryName	Name of the product category
Color	Color of the product
Size	Size of the product
ProductSpecifications	Technical/other specifications of the product
AllowSmsCommunication	SMS communication permission status
AllowEmailCommunication	Email communication permission status
ShippingMethod	Selected shipping method
CustomerId	Customer's unique ID

The dataset consists of 11,000 rows in total, and each row represents an order item. 10,000 rows of this data have been used for the model development process. In this context, there are 4,000 completed, 3,000 returned and 3,000 canceled order data in a balanced way.

Following the model development process, a separate test set of 1,000 rows was created to measure the overall success of the models. A similarly balanced class distribution has been achieved in this test set: 400 completed, 300 returned and 300 canceled orders. The distribution of the dataset has been presented in Table 3.2.

Table 3.2. Data Set Distribution

Data Type	Row Count	Completed	Returned	Canceled
Train	10,000	4,000	3,000	3,000
Test	1,000	400	300	300

4. OVERVIEW OF METHODS

4.1. Forecasting Methods

4.1.1. Logistic Regression

LR is a data mining classification method used in classification problems (Peduzzi, 1996). Unlike linear regression, it is still widely used today due to its ability to include qualitative variables in the model. This method offers flexibility in terms of being applicable on ordered and unordered categorical data. In LR analysis, the suitability of independent variables to the model structure is very important. Therefore, only meaningful and effective variables should be included in the model, while meaningless or incompatible variables should be excluded. In addition, in order for the model to make accurate predictions, the data set should be normalized in order to reduce the effect of outliers. The presence of outliers can disrupt the structure of the model and reduce the reliability of the results.

4.1.2. Random Forest

RF algorithm allows the creation of a large number of models and classifications by training each decision tree with different observation examples. It is widely preferred due to its ease of use and flexible structure that can be adapted to both classification and regression problems. The overfitting problem, which is frequently encountered in DT, can be greatly reduced with RF. Because this model aims to increase generalization success by creating multiple decision trees at the same time. In addition, as the number of trees increases, the probability of the model producing correct results increases (Bonaccorso, 2017).

4.1.3. Extreme Gradient Boosting

XGBoost is a high-performance version of the Gradient Boosting (GB) algorithm, developed with various optimizations. It is widely preferred in classification and regression problems due to its high predictive power, ability to prevent over-learning, ability to cope with missing data and fast processing ability (Chen & Guestrin, 2016).

The model learns by sequentially optimizing decision trees and focuses on incorrect classifications at each stage. The main purpose of boosting algorithms is to increase overall accuracy by strengthening weak learners. XGBoost provides successful results in many applications such as text classification, customer behavior prediction and fraud detection.

4.1.4. Support Vector Machine

SVM is a method used to model the relationship between a continuous (real number type) dependent variable and one or more independent variables. Unlike traditional regression approaches, SVM does not rely on specific distribution assumptions; instead, it attempts to model the relationship between input and output variables by learning directly. In this respect, it is a machine learning technique that can provide more flexible and effective results, especially in nonlinear data structures.

4.1.5. Deep Neural Network

DNNs are ANN's with more than three layers. They are called "deep" because they contain more than one hidden layer in addition to the input and output layers. These structures are derived from Multi-Layer Perceptron (MLP), and the learning process is defined as deep learning (Kriegeskorte and Golan, 2019). DNNs can reuse the features obtained in a hidden layer in higher layers. In this way, they can efficiently learn functions that contain component structures and can approximately model many natural functions with fewer parameters. Thanks to these capabilities, deep neural networks are widely used in the fields of machine learning and artificial intelligence today.

4.2. Feature Selection Methods

4.2.1. Minimum Redundancy Maximum Relevance (mRMR)

mRMR is a widely used filter-based feature selection algorithm that aims to select a subset of features that are highly informative for the target variable while being minimally redundant with respect to each other. The core idea is to achieve a balance between maximum relevance to the target and minimum redundancy among the selected features. (Wu et al., 2023)

Let S be the subset of selected features, x_i a feature candidate, and c the target class. The mRMR criterion can be mathematically defined as follows:

$$mRMR(x_{\mathrm{i}}) = I(x_{\mathrm{i}}; c) - \frac{1}{|S|} \sum_{\{x_{\mathrm{j}} \in S\}} I(x_{\mathrm{i}}; x_{\mathrm{j}})$$

Where:

- $I(x_{\mathrm{i}}; c)$ denotes the mutual information between feature x_{i} and the target variable c, representing the relevance of the feature.
- $I(x_{\mathrm{i}}; x_{\mathrm{j}})$ measures the mutual information between the candidate feature x_{i} and each already selected feature x_{j}, representing redundancy.

In each iteration, the feature with the highest mRMR score is selected and added to the set S. The process continues until a predefined number of features is selected or a stopping criterion is met.

4.2.2. Relief-F

Relief-F is a feature selection algorithm based on instance-based learning that evaluates the quality of attributes according to how well their values distinguish between instances that are near each other. It is particularly effective for handling multiclass classification problems and can capture feature dependencies that are often missed by univariate selection methods.

Relief-F operates by randomly selecting an instance and then identifying its nearest neighbors from the same class (nearest hit) and from different classes (nearest miss). The algorithm updates feature weights by rewarding those that differentiate between classes and penalizing those that vary within the same class. This process is repeated for a user-defined number of iterations, producing a ranking of feature importance scores. Relief-F is especially useful in detecting interactions among features and has been shown to perform well in high-dimensional classification tasks (Urbanowicz et al., 2018)

4.2.3. F-Classification

F-Classification is a univariate method used for feature selection in classification problems. It independently evaluates the relationship between each feature and the target variable, and this relationship is measured based on the one-way ANOVA (Analysis of Variance) F-test.

The method computes an F-statistic and a corresponding p-value for each feature. A low p-value indicates that the feature creates statistically significant differences between classes and is therefore more informative for the model. Features with p-values below a certain threshold (commonly 0.05) are selected and ranked for inclusion in the modeling process.

4.2.4. Hybrid Feature Selection Algorithm

The Hybrid feature selection algorithm developed in this study aims to combine the strengths of three well-established feature selection techniques - mRMR, Relief-F, and F-Classification—to build a robust and balanced decision mechanism for selecting relevant features.

The algorithm follows a structured three-stage process:

1. Stage – Majority Voting: The ranked outputs of the three individual feature selection methods are compared. If at least two algorithms rank the same feature highly, that feature is selected based on majority voting. This stage ensures that commonly agreed-upon important features are prioritized.

2. Stage – Correlation Score Evaluation: If no agreement is reached in the voting stage, a correlation score is determined for every proposed feature. There are two elements in this score:
   - Inconsistency Score (IS): Measures ranking inconsistency of a feature across the three algorithms.
   - Distance to Rank 1 (D): Indicates how far a feature is from the top-ranked feature in the lists.

The total correlation score is the sum of IS and D. The feature with the lowest combined score is selected, as it is both consistently ranked and close to the top.

3. Stage – Priority-Based Selection: If multiple features share the same correlation score, a predefined priority order of the algorithms is applied. This priority is determined based on the average performance of each algorithm on the dataset used in the study.

This hybrid feature selection approach effectively balances relevance and redundancy, resulting in a feature set that enhances model accuracy and generalizability. The schematic representation of the hybrid feature selection algorithm is shown in Figure 4.1.

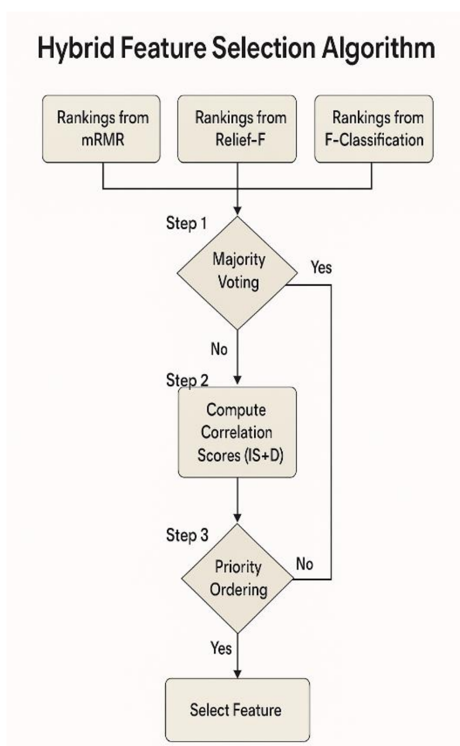


Figure 4.1. Flowchart of the Hybrid Feature Selection Algorithm

5. INTERFACE WITH HYPERPARAMETRIC FLEXIBILITY

To manage the training and testing processes of machine learning-based classification models in a user-friendly and flexible environment, an interface with high-level hyperparameter configurability was developed. The interface is designed to enable users to compare different model types, apply various feature selection algorithms, and analyze multiple hyperparameter combinations in detail. Developed using the Python programming language, the interface leverages open-source libraries such as Tkinter, Pandastable, and Matplotlib. The entire modeling pipeline - including data uploading, feature selection, model setup, hyperparameter configuration, prediction generation, and performance evaluation- can be executed within a single integrated platform. The main functionalities of the interface are summarized as follows:

- **Data Upload:** Users can upload training and test datasets in .csv or .xlsx formats. Uploaded data are displayed in tabular format within the interface.
- **Data Partitioning and Scaling:** Users can define the ratio between training and test sets. Three scaling options are available: none, min-max, and standardization.
- **Feature and Target Selection:** Users can select independent (input) and dependent (target) variables from the dataset via the interface.
- **Feature Selection:** The interface supports the application of mRMR, Relief-F, F-Regression, and the proposed hybrid feature selection algorithm, allowing users to identify the most relevant variables.
- **Model Selection and Hyperparameter Configuration:** The interface supports various predictive models, including DNN, RF, XGBoost, SVM, and LR. Each model has its own set of configurable hyperparameters. Additionally, grid search can be used for hyperparameter optimization.
- **Model Training:** Once a model is selected and configured, it can be trained on the uploaded data. The training loss is monitored and displayed graphically during the training process.
- **Prediction:** Users can generate predictions using the test dataset and access the results in tabular format.
- **Performance Evaluation:** The interface calculates and displays performance metrics such as accuracy, F1-score, precision, and recall.
- **Visualization:** Model's performance is visualized using a confusion matrix, enabling class-level performance analysis.

- **Model Saving and Loading:** Trained models can be saved and reloaded for future use without the need for retraining.

Through this interface, users can flexibly manage the preprocessing and modeling processes, perform experimental hyperparameter tuning, and strategically conduct feature selection to build highly accurate and interpretable predictive models. The output screens of the developed interface are presented between Figure 2 and Figure 6.

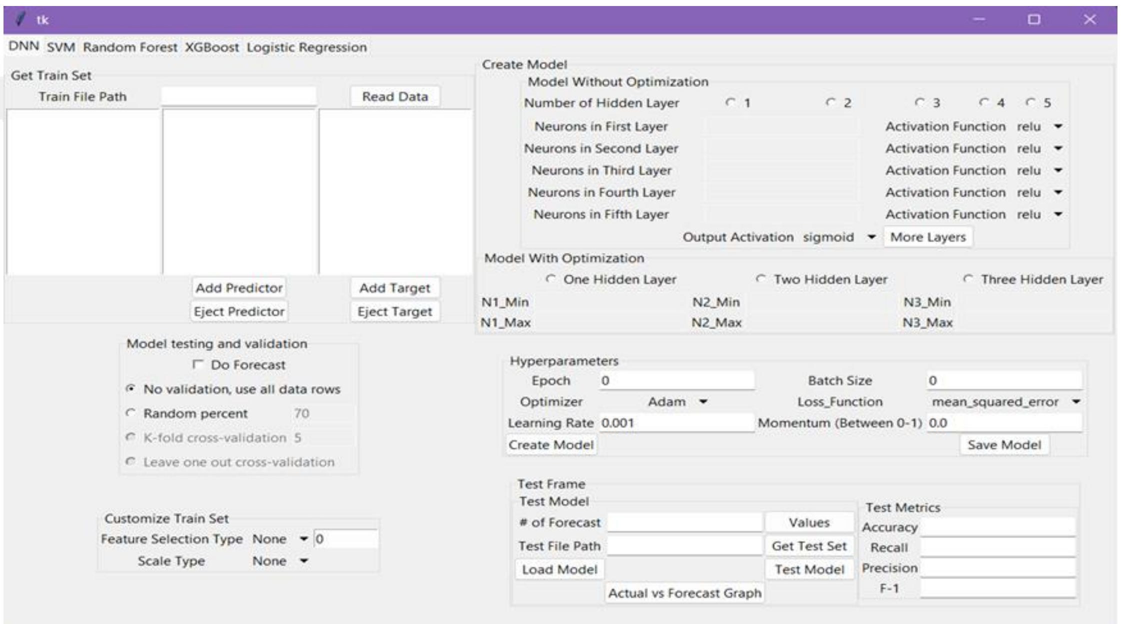


Figure 5.1. DNN

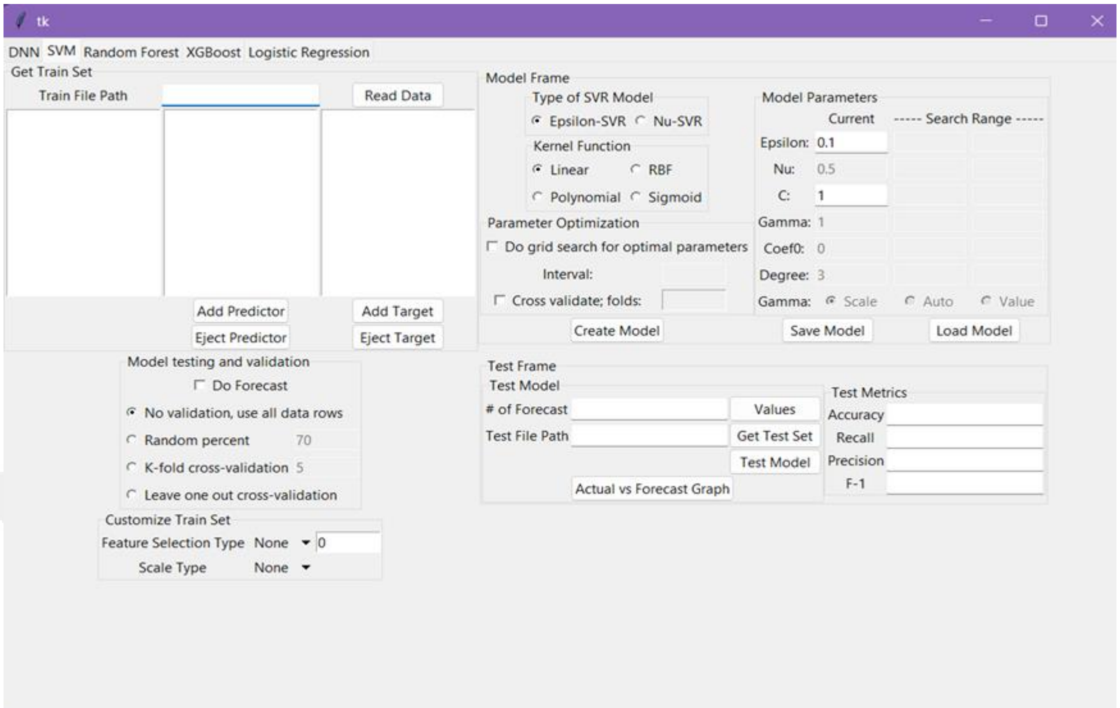


Figure 5.2. SVM

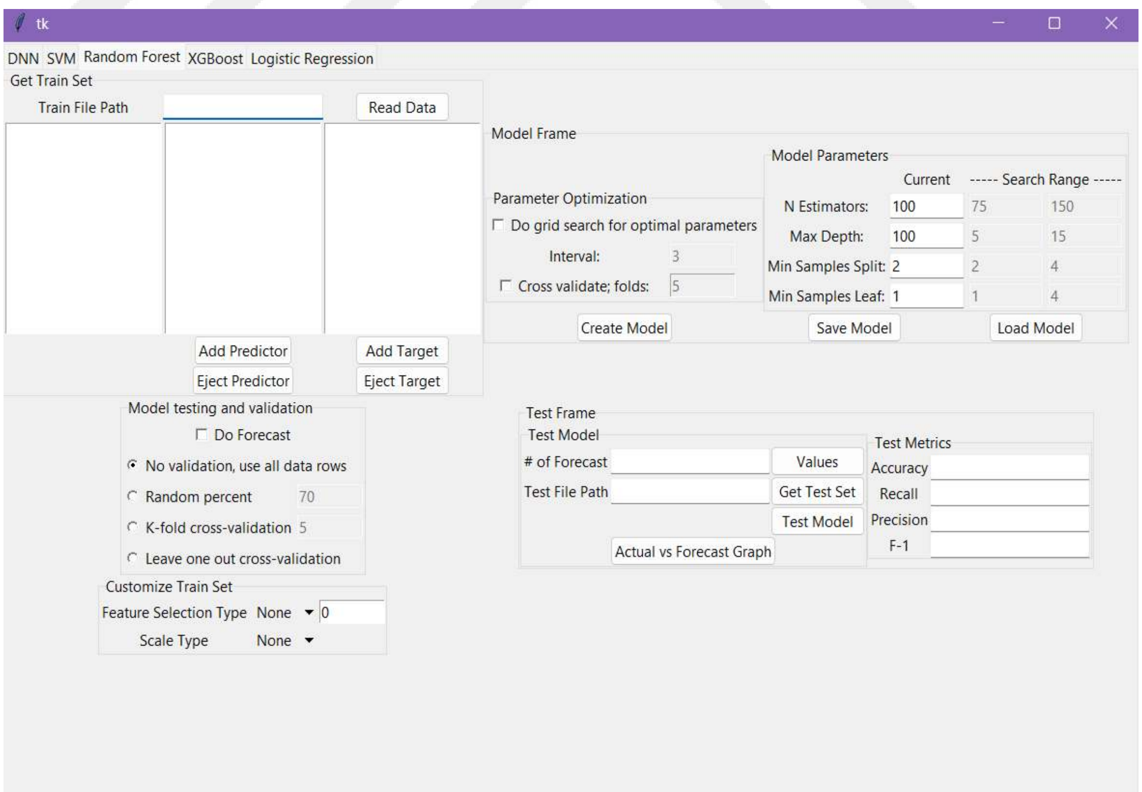


Figure 5.3. RF

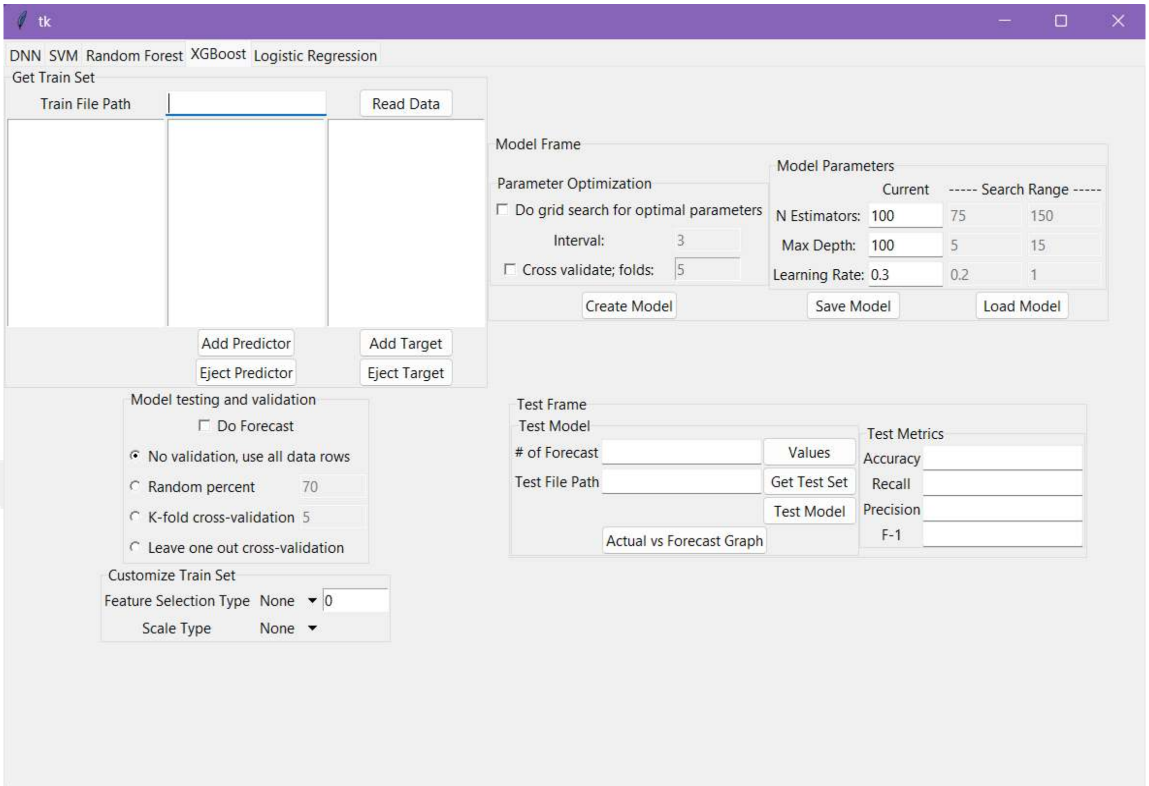


Figure 5.4. XGBoost

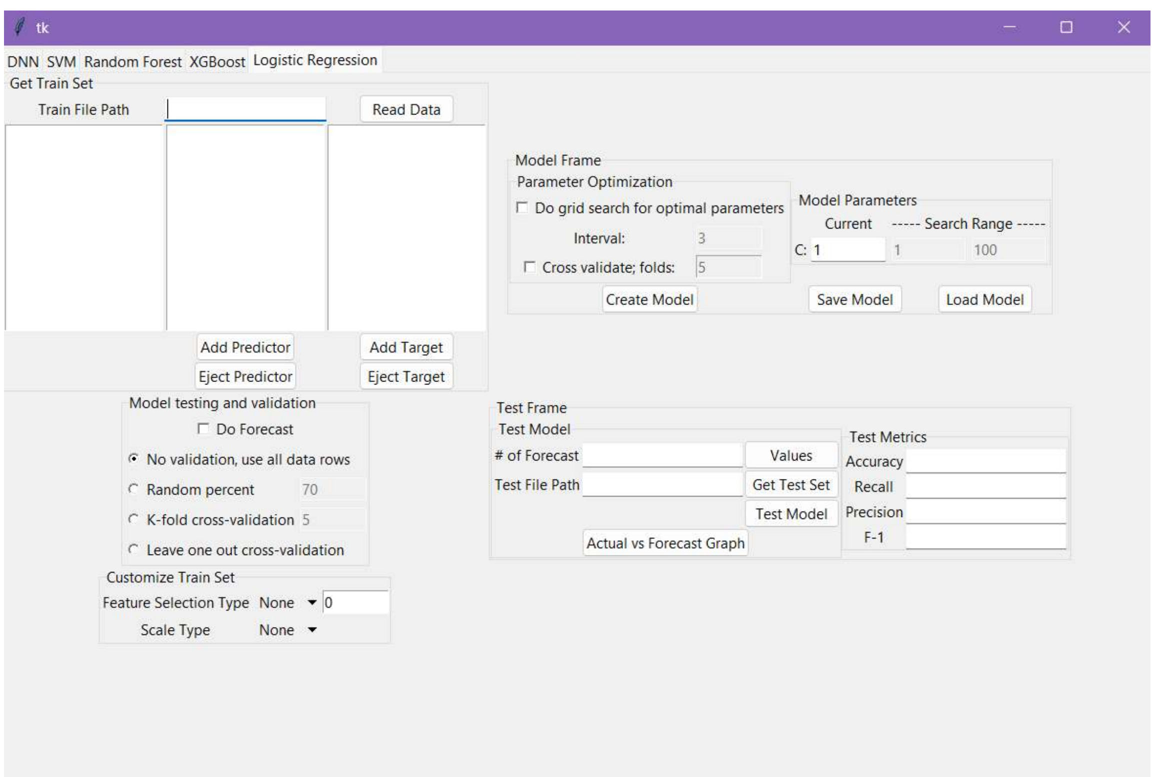


Figure 5.5. LR

6. RESULT AND DISCUSSION

6.1. Prediction Framework

The prediction process has been structured with three different experimental studies. The first two studies have used 10,000 rows of training data and have applied the 10-fold cross-validation method. However, the main difference between these two studies lies in the timing of the application of the feature selection process. In the first study, the feature selection process has been carried out before the cross-validation process; in the second study, feature selection has been performed separately in each fold. This design has enabled a comparative examination of the impact of feature selection on model performance under different scenarios.

In both studies, two different modeling approaches have been adopted for sales cancellation and return prediction. In the first approach, the problem has been considered in a multi-class classification structure, and completed, returned, and canceled orders have been classified within a single model framework. In the second approach, the problem has been treated as a binary classification task; two separate prediction models have been developed for cancellation and return situations.

LR, RF, XGBoost, SVM, and DNN algorithms have been used in the model development process. In the feature selection phase, mRMR, Relief-F, and F-Classification algorithms, as well as a hybrid feature selection method created by combining these three, have been applied. The effects of these methods on model performance have been analyzed comparatively. While the training process has been conducted using 10-fold cross-validation, a grid search has been applied with 3-fold cross-validation during the hyperparameter optimization phase. The grid search hyperparameter ranges for the chosen methods are shown in Table 6.1.

Table 6.1. Hyperparameter Ranges Used for Grid Search

LR	Inverse Regularization Strength (C)	[0.001, 0.01, 0.1, 1, 10, 100, 1000]
RF	Number of Trees (n_estimators)	[50, 100, 200]
	Maximum Tree Depth (max_depth)	None
	Minimum Samples Split	[2, 5, 10]
	Minimum Samples per Leaf	[1, 2, 4]
XGBoost	Maximum Tree Depth (max_depth)	[3, 5, 7, 9]
	Learning Rate (learning_rate)	[0.01, 0.1, 0.2, 0.3]
	Number of Trees (n_estimators)	[50, 100, 200]
SVM	Regularization Parameter (C)	[0.1, 1, 10, 100, 1000]
	Kernel Type	['rbf']
	Kernel Coefficient (gamma)	['scale', 'auto']
DNN	Number of Neurons	[32, 64]
	Learning Rate	[0.001, 0.01]
	Batch Size	[128, 256]

In the third study, a 1,000-line test dataset created independently of the training data has been used to evaluate the generalizability of the modeling process. On this test data, the performances of the most successful models obtained through both multi-class and binary classification approaches have been evaluated, and the accuracy of the results has been tested.

6.2. Forecasting Models

Key metrics like Precision, Recall, F1-score, and Accuracy were used to determine the performance of the generated models by the prediction process described in Section 4.5. The results were shown in tabular form after these metrics were computed independently for completed, canceled, and returned orders. Confusion matrices, which are included in the Appendix, were also created to show classification performance.

6.3. First Experimental Study: Pre-Cross-Validation Feature Selection

6.3.1. Multi-Class Classification Based Cancellation-Return Prediction Models

For the first study, 12 features were selected using mRMR, Relief-F, F-Classification, and Hybrid feature selection algorithms, and the selected features are presented in Table 6.2.

Table 6.2 Selected Features Using Feature Selection Algorithms

	mRMR	**Relief-F**	**F-Classification**	**Hybrid**
1	CancelledOrderCount	Price	PaymentMethodSystemName	PaymentMethodSystemName
2	OrderWeekend	County	OrderPlatform	OrderPlatform
3	OrderDay	StateProvince	StateProvince	StateProvince
4	County	OrderDay	County	County
5	ReturnedOrderCount	Renk	IsCampaignUsed	Price
6	IsDiscounted	CategoryName	Gender	OrderDay
7	AllowSmsCommunication	OrderMonth	CancelledOrderCount	CancelledOrderCount
8	OrderMonth	OrderWeekday	ReturnedOrderCount	Renk
9	Gender	ReturnedOrderCount	CompletedOrderCount	CategoryName
10	OrderWeekday	CompletedOrderCount	Price	OrderMonth
11	OrderPlatform	AllowEmailCommunication	IsDiscounted	OrderWeekday
12	Beden	AllowSmsCommunication	CategoryName	ReturnedOrderCount

In Table 6.3 through Table 6.7 the error metrics obtained by the models developed with the multi-class classification approach are presented for LR, RF, XGBoost, SVM and DNN algorithms, respectively.

Table 6.3. Error Metrics Obtained with LR

	LR	LR + mRMR	LR + Relief-F	LR + F-Class	LR + Hybrid
Completed orders – Precision	0.60	0.50	0.55	0.54	0.49
Completed orders – Recall	0.76	0.70	0.81	0.86	0.69
Completed orders – F1-score	0.67	0.58	0.65	0.67	0.57
Canceled orders – Precision	0.56	0.53	0.35	0.63	0.54
Canceled orders – Recall	0.52	0.51	0.30	0.28	0.51
Canceled orders – F1-score	0.54	0.52	0.32	0.38	0.52
Returned orders – Precision	0.71	0.65	0.64	0.73	0.66
Returned orders – Recall	0.52	0.33	0.33	0.57	0.33
Returned orders – F1-score	0.60	0.44	0.44	0.64	0.44
All orders – Precision	0.63	0.56	0.51	0.63	0.56
All orders – Recall	0.60	0.52	0.48	0.57	0.51
All orders – F1-score	0.60	0.52	0.47	0.56	0.51
Accuracy	0.62	0.53	0.51	0.60	0.53

Table 6.4. Error Metrics Obtained with RF

	RF	RF + mRMR	RF + Relief-F	RF + F-Class	RF + Hybrid
Completed orders – Precision	0.97	0.96	0.98	0.97	0.98
Completed orders – Recall	0.92	0.84	0.93	0.92	0.88
Completed orders – F1-score	0.95	0.89	0.95	0.94	0.93
Canceled orders – Precision	0.96	0.91	0.97	0.95	0.94
Canceled orders – Recall	0.99	0.98	0.99	0.99	1.00
Canceled orders – F1-score	0.97	0.94	0.98	0.97	0.97
Returned orders – Precision	0.93	0.87	0.93	0.92	0.90
Returned orders – Recall	0.96	0.95	0.96	0.95	0.97
Returned orders – F1-score	0.95	0.91	0.95	0.94	0.93
All orders – Precision	0.95	0.91	0.96	0.95	0.94
All orders – Recall	0.96	0.92	0.96	0.95	0.95
All orders – F1-score	0.96	0.91	0.96	0.95	0.94
Accuracy	0.96	0.91	0.96	0.95	0.94

Table 6.5. Error Metrics Obtained with XGBoost

	XGBoost	XGBoost + mRMR	XGBoost + Relief-F	XGBoost + F-Class	XGBoost + Hybrid
Completed orders – Precision	0.97	0.95	0.97	0.97	0.97
Completed orders – Recall	0.91	0.85	0.92	0.92	0.88
Completed orders – F1-score	0.94	0.90	0.95	0.95	0.93
Canceled orders – Precision	0.95	0.91	0.96	0.95	0.94
Canceled orders – Recall	0.99	0.97	0.99	0.99	1.00
Canceled orders – F1-score	0.97	0.94	0.98	0.97	0.97
Returned orders – Precision	0.93	0.88	0.93	0.93	0.90
Returned orders – Recall	0.96	0.94	0.96	0.95	0.96
Returned orders – F1-score	0.94	0.91	0.94	0.94	0.93
All orders – Precision	0.95	0.91	0.95	0.95	0.94
All orders – Recall	0.95	0.92	0.96	0.95	0.95
All orders – F1-score	0.95	0.91	0.96	0.95	0.94
Accuracy	0.95	0.91	0.96	0.95	0.94

Table 6.6. Error Metrics Obtained with SVM

	SVM	SVM + mRMR	SVM + Relief-F	SVM + F-Class	SVM + Hybrid
Completed orders – Precision	0.94	0.89	0.93	0.91	0.94
Completed orders – Recall	0.86	0.79	0.84	0.82	0.82
Completed orders – F1-score	0.90	0.84	0.88	0.86	0.87
Canceled orders – Precision	0.90	0.85	0.87	0.86	0.88
Canceled orders – Recall	0.98	0.94	0.96	0.92	0.97
Canceled orders – F1-score	0.94	0.89	0.91	0.89	0.92
Returned orders – Precision	0.90	0.85	0.90	0.86	0.87
Returned orders – Recall	0.92	0.89	0.93	0.90	0.93
Returned orders – F1-score	0.91	0.87	0.92	0.88	0.90
All orders – Precision	0.91	0.86	0.90	0.88	0.89
All orders – Recall	0.92	0.87	0.91	0.88	0.90
All orders – F1-score	0.92	0.87	0.90	0.88	0.90
Accuracy	0.91	0.87	0.90	0.88	0.90

Table 6.7. Error Metrics Obtained with DNN

	DNN	DNN + mRMR	DNN + Relief-F	DNN + F-Class	DNN + Hybrid
Completed orders – Precision	0.89	0.83	0.80	0.83	0.84
Completed orders – Recall	0.85	0.75	0.80	0.82	0.77
Completed orders – F1-score	0.87	0.79	0.80	0.82	0.80
Canceled orders – Precision	0.90	0.81	0.81	0.85	0.84
Canceled orders – Recall	0.92	0.86	0.81	0.81	0.88
Canceled orders – F1-score	0.91	0.84	0.81	0.83	0.86
Returned orders – Precision	0.86	0.79	0.84	0.81	0.81
Returned orders – Recall	0.88	0.84	0.85	0.86	0.86
Returned orders – F1-score	0.87	0.82	0.84	0.83	0.83
All orders – Precision	0.88	0.81	0.82	0.83	0.83
All orders – Recall	0.89	0.82	0.82	0.83	0.84
All orders – F1-score	0.89	0.81	0.82	0.83	0.83
Accuracy	0.88	0.81	0.82	0.83	0.83

6.3.2. Binary Classification Based Cancellation Prediction Models

12 features were selected using mRMR, Relief-F, F-Classification, and Hybrid feature selection algorithms, and the selected features are presented in Table 6.8.

Table 6.8. Selected Features Using Feature Selection Algorithms

	mRMR	Relief-F	F-Classification	Hybrid
1	CancelledOrderCount	County	PaymentMethodSystem Name	County
2	OrderWeekend	OrderDay	OrderPlatform	CancelledOrderCount
3	AllowSmsCommunica tion	Price	StateProvince	OrderDay
4	OrderDay	Beden	County	PaymentMethodSystem Name
5	IsDiscounted	Renk	IsCampaignUsed	OrderPlatform
6	County	OrderWeekday	Gender	StateProvince
7	OrderWeekday	StateProvince	CancelledOrderCount	Price
8	ReturnedOrderCount	CategoryName	ReturnedOrderCount	Beden
9	OrderMonth	CompletedOrder Count	CompletedOrderCount	Renk
10	CategoryName	OrderMonth	Price	OrderWeekday
11	OrderPlatform	ReturnedOrderCo unt	IsDiscounted	CategoryName
12	AllowEmailCommuni cation	CancelledOrderC ount	CategoryName	CompletedOrderCount

In Table 6.9 through Table 6.13, the error metrics obtained by the models developed with the multi-class classification approach are presented for LR, RF, XGBoost, SVM and DNN algorithms, respectively.

Table 6.9. Error Metrics Obtained with LR

	LR	LR + mRMR	LR + Relief-F	LR + F-Class	LR + Hybrid
Completed orders – Precision	0.74	0.69	0.73	0.66	0.72
Completed orders – Recall	0.81	0.79	0.84	0.91	0.86
Completed orders – F1-score	0.77	0.74	0.78	0.76	0.79
Canceled orders – Precision	0.71	0.65	0.73	0.75	0.75
Canceled orders – Recall	0.61	0.52	0.60	0.38	0.56
Canceled orders – F1-score	0.66	0.58	0.66	0.51	0.64
All orders – Precision	0.72	0.67	0.73	0.71	0.74
All orders – Recall	0.71	0.66	0.72	0.64	0.71
All orders – F1-score	0.72	0.66	0.72	0.64	0.71
Accuracy	0.73	0.67	0.73	0.68	0.73

Table 6.10. Error Metrics Obtained with RF

	RF	RF + mRMR	RF + Relief-F	RF + F-Class	RF + Hybrid
Completed orders – Precision	0.99	1.00	0.99	1.00	0.99
Completed orders – Recall	0.97	0.95	0.97	0.98	0.97
Completed orders – F1-score	0.98	0.98	0.98	0.99	0.98
Canceled orders – Precision	0.96	0.94	0.96	0.97	0.96
Canceled orders – Recall	0.99	1.00	0.99	1.00	0.99
Canceled orders – F1-score	0.98	0.97	0.98	0.98	0.97
All orders – Precision	0.98	0.97	0.98	0.98	0.98
All orders – Recall	0.98	0.98	0.98	0.99	0.98
All orders – F1-score	0.98	0.97	0.98	0.98	0.98
Accuracy	0.98	0.97	0.98	0.98	0.98

Table 6.11. Error Metrics Obtained with XGBoost

	XGBoost	XGBoost + mRMR	XGBoost + Relief-F	XGBoost + F-Class	XGBoost + Hybrid
Completed orders – Precision	1.00	1.00	1.00	1.00	1.00
Completed orders – Recall	0.96	0.94	0.97	0.96	0.96
Completed orders – F1-score	0.98	0.97	0.98	0.98	0.98
Canceled orders – Precision	0.95	0.93	0.96	0.95	0.95
Canceled orders – Recall	0.99	1.00	0.99	1.00	0.99
Canceled orders – F1-score	0.97	0.96	0.97	0.97	0.97
All orders – Precision	0.97	0.96	0.98	0.97	0.98
All orders – Recall	0.98	0.97	0.98	0.98	0.98
All orders – F1-score	0.98	0.97	0.98	0.98	0.98
Accuracy	0.98	0.97	0.98	0.98	0.98

Table 6.12. Error Metrics Obtained with SVM

	SVM	SVM + mRMR	SVM + Relief-F	SVM + F-Class	SVM + Hybrid
Completed orders – Precision	0.99	0.98	0.98	0.95	0.97
Completed orders – Recall	0.91	0.88	0.91	0.89	0.90
Completed orders – F1-score	0.95	0.93	0.94	0.92	0.93
Canceled orders – Precision	0.89	0.86	0.89	0.87	0.88
Canceled orders – Recall	0.98	0.98	0.97	0.94	0.97
Canceled orders – F1-score	0.94	0.91	0.93	0.90	0.92
All orders – Precision	0.94	0.92	0.93	0.91	0.92
All orders – Recall	0.95	0.93	0.94	0.92	0.93
All orders – F1-score	0.94	0.92	0.93	0.91	0.93
Accuracy	0.94	0.92	0.93	0.91	0.93

Table 6.13. Error Metrics Obtained with DNN

	DNN	DNN + mRMR	DNN + Relief-F	DNN + F-Class	DNN + Hybrid
Completed orders – Precision	0.98	0.96	0.97	0.95	0.97
Completed orders – Recall	0.93	0.89	0.93	0.90	0.92
Completed orders – F1-score	0.95	0.92	0.95	0.93	0.94
Canceled orders – Precision	0.92	0.86	0.91	0.88	0.90
Canceled orders – Recall	0.97	0.95	0.96	0.93	0.96
Canceled orders – F1-score	0.94	0.90	0.93	0.91	0.93
All orders – Precision	0.95	0.91	0.94	0.91	0.93
All orders – Recall	0.95	0.92	0.94	0.92	0.94
All orders – F1-score	0.95	0.91	0.94	0.92	0.93
Accuracy	0.95	0.91	0.94	0.92	0.94

6.3.3. Binary Classification Based Return Prediction Models

12 features were selected using mRMR, Relief-F, F-Classification, and Hybrid feature selection algorithms, and the selected features are presented in Table 6.14.

Table 6.14. Selected Features Using Feature Selection Algorithms

	mRMR	**Relief-F**	**F-Classification**	**Hybrid**
1	ReturnedOrderCount	County	PaymentMethodSystemName	ReturnedOrderCount
2	OrderWeekend	OrderDay	OrderPlatform	County
3	Gender	CompletedOrderCount	StateProvince	PaymentMethodSystemName
4	PaymentMethodSystemName	StateProvince	County	OrderPlatform
5	StateProvince	ReturnedOrderCount	IsCampaignUsed	StateProvince
6	Price	OrderMonth	Gender	OrderDay
7	CancelledOrderCount	OrderWeekday	CancelledOrderCount	CompletedOrderCount
8	OrderPlatform	CategoryName	ReturnedOrderCount	OrderMonth
9	Beden	CancelledOrderCount	CompletedOrderCount	OrderWeekday
10	OrderMonth	Gender	Price	CategoryName
11	AllowSmsCommunication	AllowEmailCommunication	IsDiscounted	CancelledOrderCount
12	County	AllowSmsCommunication	CategoryName	Gender

In Table 6.15 through Table 28, the error metrics obtained by the models developed with the multi-class classification approach are presented for LR, RF, XGBoost, SVM and DNN algorithms, respectively.

Table 6.15. Error Metrics Obtained with LR

	LR	LR + mRMR	LR + Relief-F	LR + F-Class	LR + Hybrid
Completed orders – Precision	0.79	0.65	0.79	0.79	0.79
Completed orders – Recall	0.91	0.91	0.91	0.91	0.91
Completed orders – F1-score	0.84	0.76	0.85	0.84	0.84
Returned orders – Precision	0.84	0.74	0.85	0.85	0.85
Returned orders – Recall	0.68	0.36	0.67	0.67	0.67
Returned orders – F1-score	0.75	0.48	0.75	0.75	0.75
All orders – Precision	0.82	0.70	0.82	0.82	0.82
All orders – Recall	0.79	0.63	0.79	0.79	0.79
All orders – F1-score	0.80	0.62	0.80	0.80	0.80
Accuracy	0.81	0.67	0.81	0.81	0.81

Table 6.16. Error Metrics Obtained with RF

	RF	RF + mRMR	RF + Relief-F	RF + F-Class	RF + Hybrid
Completed orders – Precision	0.98	0.98	0.97	0.97	0.98
Completed orders – Recall	0.94	0.89	0.94	0.94	0.94
Completed orders – F1-score	0.96	0.93	0.96	0.96	0.96
Returned orders – Precision	0.92	0.87	0.93	0.93	0.92
Returned orders – Recall	0.97	0.97	0.96	0.97	0.97
Returned orders – F1-score	0.95	0.92	0.94	0.95	0.94
All orders – Precision	0.95	0.92	0.95	0.95	0.95
All orders – Recall	0.95	0.93	0.95	0.95	0.95
All orders – F1-score	0.95	0.93	0.95	0.95	0.95
Accuracy	0.95	0.93	0.95	0.95	0.95

Table 6.17. Error Metrics Obtained with XGBoost

	XGBoost	XGBoost + mRMR	XGBoost + Relief-F	XGBoost + F-Class	XGBoost + Hybrid
Completed orders – Precision	0.97	0.97	0.97	0.97	0.97
Completed orders – Recall	0.94	0.90	0.94	0.94	0.94
Completed orders – F1-score	0.96	0.94	0.96	0.96	0.95
Returned orders – Precision	0.92	0.88	0.92	0.92	0.92
Returned orders – Recall	0.97	0.96	0.97	0.97	0.96
Returned orders – F1-score	0.95	0.92	0.94	0.94	0.94
All orders – Precision	0.95	0.93	0.95	0.95	0.95
All orders – Recall	0.95	0.93	0.95	0.95	0.95
All orders – F1-score	0.95	0.93	0.95	0.95	0.95
Accuracy	0.95	0.93	0.95	0.95	0.95

Table 6.18. Error Metrics Obtained with SVM

	SVM	SVM + mRMR	SVM + Relief-F	SVM + F-Class	SVM + Hybrid
Completed orders – Precision	0.95	0.93	0.95	0.95	0.96
Completed orders – Recall	0.90	0.87	0.90	0.90	0.90
Completed orders – F1-score	0.92	0.90	0.93	0.92	0.93
Returned orders – Precision	0.88	0.84	0.88	0.87	0.88
Returned orders – Recall	0.93	0.92	0.94	0.94	0.95
Returned orders – F1-score	0.90	0.88	0.91	0.91	0.91
All orders – Precision	0.91	0.89	0.92	0.91	0.92
All orders – Recall	0.92	0.89	0.92	0.92	0.93
All orders – F1-score	0.91	0.89	0.92	0.92	0.92
Accuracy	0.92	0.89	0.92	0.92	0.92

Table 6.19. Error Metrics Obtained with DNN

	DNN	DNN + mRMR	DNN + Relief-F	DNN + F-Class	DNN + Hybrid
Completed orders – Precision	0.94	0.93	0.93	0.93	0.93
Completed orders – Recall	0.92	0.83	0.91	0.90	0.91
Completed orders – F1-score	0.93	0.88	0.92	0.91	0.92
Returned orders – Precision	0.89	0.80	0.88	0.87	0.88
Returned orders – Recall	0.92	0.91	0.91	0.91	0.91
Returned orders – F1-score	0.91	0.85	0.89	0.89	0.90
All orders – Precision	0.92	0.86	0.90	0.90	0.91
All orders – Recall	0.92	0.87	0.91	0.90	0.91
All orders – F1-score	0.92	0.87	0.90	0.90	0.91
Accuracy	0.92	0.87	0.91	0.90	0.91

6.4. Second Experimental Study: In-Fold Feature Selection during Cross-Validation

6.4.1. Multi-Class Classification Based Cancellation-Return Prediction Models

In Table 6.20 through Table 6.24, the error metrics obtained by the models developed with the multi-class classification approach are presented for LR, RF, XGBoost, SVM and DNN algorithms, respectively.

Table 6.20. Error Metrics Obtained with LR

	LR	LR + mRMR	LR + Relief-F	LR + F-Class	LR + Hybrid
Completed orders – Precision	0.60	0.50	0.57	0.54	0.57
Completed orders – Recall	0.76	0.70	0.83	0.86	0.83
Completed orders – F1-score	0.67	0.58	0.67	0.67	0.67
Canceled orders – Precision	0.56	0.54	0.60	0.63	0.60
Canceled orders – Recall	0.52	0.51	0.35	0.28	0.35
Canceled orders – F1-score	0.54	0.52	0.44	0.38	0.44
Returned orders – Precision	0.71	0.65	0.73	0.73	0.73
Returned orders – Recall	0.52	0.33	0.58	0.57	0.58
Returned orders – F1-score	0.60	0.44	0.64	0.64	0.64
All orders – Precision	0.63	0.56	0.63	0.63	0.63
All orders – Recall	0.60	0.51	0.59	0.57	0.59
All orders – F1-score	0.60	0.52	0.59	0.56	0.59
Accuracy	0.62	0.53	0.61	0.60	0.61

Table 6.21. Error Metrics Obtained with RF

	RF	RF + mRMR	RF + Relief-F	RF + F-Class	RF + Hybrid
Completed orders – Precision	0.97	0.96	0.98	0.97	0.98
Completed orders – Recall	0.92	0.84	0.92	0.92	0.92
Completed orders – F1-score	0.95	0.90	0.95	0.94	0.95
Canceled orders – Precision	0.96	0.91	0.95	0.95	0.95
Canceled orders – Recall	0.99	0.98	0.99	0.99	0.99
Canceled orders – F1-score	0.97	0.94	0.97	0.97	0.97
Returned orders – Precision	0.93	0.87	0.93	0.92	0.93
Returned orders – Recall	0.96	0.95	0.96	0.95	0.96
Returned orders – F1-score	0.95	0.91	0.94	0.94	0.94
All orders – Precision	0.95	0.92	0.95	0.95	0.95
All orders – Recall	0.96	0.93	0.96	0.95	0.96
All orders – F1-score	0.96	0.92	0.95	0.95	0.96
Accuracy	0.96	0.92	0.95	0.95	0.95

Table 6.22. Error Metrics Obtained with XGBoost

	XGBoost	XGBoost + mRMR	XGBoost + Relief-F	XGBoost + F-Class	XGBoost + Hybrid
Completed orders – Precision	0.97	0.96	0.98	0.97	0.98
Completed orders – Recall	0.91	0.85	0.92	0.92	0.91
Completed orders – F1-score	0.94	0.90	0.95	0.95	0.94
Canceled orders – Precision	0.95	0.91	0.95	0.95	0.95
Canceled orders – Recall	0.99	0.98	0.99	0.99	0.99
Canceled orders – F1-score	0.97	0.94	0.97	0.97	0.97
Returned orders – Precision	0.93	0.88	0.93	0.93	0.92
Returned orders – Recall	0.96	0.94	0.96	0.95	0.96
Returned orders – F1-score	0.94	0.91	0.94	0.94	0.94
All orders – Precision	0.95	0.91	0.95	0.95	0.95
All orders – Recall	0.95	0.92	0.96	0.95	0.95
All orders – F1-score	0.95	0.92	0.95	0.95	0.95
Accuracy	0.95	0.92	0.95	0.95	0.95

Table 6.23. Error Metrics Obtained with SVM

	SVM	SVM + mRMR	SVM + Relief-F	SVM + F-Class	SVM + Hybrid
Completed orders – Precision	0.94	0.90	0.93	0.91	0.93
Completed orders – Recall	0.86	0.79	0.86	0.82	0.86
Completed orders – F1-score	0.90	0.84	0.89	0.86	0.89
Canceled orders – Precision	0.90	0.85	0.88	0.86	0.88
Canceled orders – Recall	0.98	0.94	0.95	0.92	0.95
Canceled orders – F1-score	0.94	0.89	0.91	0.89	0.91
Returned orders – Precision	0.90	0.85	0.90	0.86	0.90
Returned orders – Recall	0.92	0.89	0.93	0.90	0.93
Returned orders – F1-score	0.91	0.87	0.91	0.88	0.91
All orders – Precision	0.91	0.87	0.90	0.88	0.90
All orders – Recall	0.92	0.88	0.91	0.88	0.91
All orders – F1-score	0.92	0.87	0.91	0.88	0.91
Accuracy	0.91	0.87	0.91	0.88	0.91

Table 6.24. Error Metrics Obtained with DNN

	DNN	DNN + mRMR	DNN + Relief-F	DNN + F-Class	DNN + Hybrid
Completed orders – Precision	0.88	0.84	0.85	0.82	0.84
Completed orders – Recall	0.85	0.76	0.84	0.82	0.83
Completed orders – F1-score	0.87	0.79	0.85	0.82	0.84
Canceled orders – Precision	0.90	0.82	0.87	0.84	0.85
Canceled orders – Recall	0.92	0.87	0.86	0.83	0.87
Canceled orders – F1-score	0.91	0.84	0.87	0.84	0.86
Returned orders – Precision	0.86	0.80	0.85	0.83	0.85
Returned orders – Recall	0.89	0.85	0.87	0.83	0.85
Returned orders – F1-score	0.88	0.82	0.86	0.83	0.85
All orders – Precision	0.88	0.82	0.86	0.83	0.85
All orders – Recall	0.89	0.82	0.86	0.83	0.85
All orders – F1-score	0.89	0.82	0.86	0.83	0.85
Accuracy	0.88	0.82	0.86	0.83	0.85

6.4.2. Binary Classification Based Cancellation Prediction Models

In Table 6.25 through Table 6.29, the error metrics obtained by the models developed with the multi-class classification approach are presented for LR, RF, XGBoost, SVM and DNN algorithms, respectively.

Table 6.25. Error Metrics Obtained with LR

	LR	LR + mRMR	LR + Relief-F	LR + F-Class	LR + Hybrid
Completed orders – Precision	0.74	0.69	0.69	0.66	0.69
Completed orders – Recall	0.81	0.79	0.88	0.91	0.86
Completed orders – F1-score	0.77	0.74	0.77	0.76	0.76
Canceled orders – Precision	0.71	0.65	0.74	0.75	0.72
Canceled orders – Recall	0.61	0.52	0.47	0.38	0.47
Canceled orders – F1-score	0.66	0.58	0.58	0.51	0.57
All orders – Precision	0.72	0.67	0.72	0.71	0.70
All orders – Recall	0.71	0.66	0.67	0.64	0.67
All orders – F1-score	0.72	0.66	0.67	0.64	0.67
Accuracy	0.73	0.68	0.70	0.68	0.69

Table 6.26. Error Metrics Obtained with RF

	RF	RF + mRMR	RF + Relief-F	RF + F-Class	RF + Hybrid
Completed orders – Precision	0.99	1.00	1.00	1.00	1.00
Completed orders – Recall	0.97	0.95	0.97	0.98	0.97
Completed orders – F1-score	0.98	0.98	0.99	0.99	0.99
Canceled orders – Precision	0.96	0.94	0.97	0.97	0.97
Canceled orders – Recall	0.99	1.00	1.00	1.00	1.00
Canceled orders – F1-score	0.98	0.97	0.98	0.98	0.98
All orders – Precision	0.98	0.97	0.98	0.98	0.98
All orders – Recall	0.98	0.98	0.99	0.99	0.99
All orders – F1-score	0.98	0.97	0.98	0.98	0.98
Accuracy	0.98	0.97	0.98	0.98	0.98

Table 6.27. Error Metrics Obtained with XGBoost

	XGBoost	XGBoost + mRMR	XGBoost + Relief-F	XGBoost + F-Class	XGBoost + Hybrid
Completed orders – Precision	0.99	1.00	1.00	1.00	1.00
Completed orders – Recall	0.97	0.95	0.97	0.98	0.97
Completed orders – F1-score	0.98	0.98	0.99	0.99	0.99
Canceled orders – Precision	0.96	0.94	0.97	0.97	0.97
Canceled orders – Recall	0.99	1.00	1.00	1.00	1.00
Canceled orders – F1-score	0.98	0.97	0.98	0.98	0.98
All orders – Precision	0.98	0.97	0.98	0.98	0.98
All orders – Recall	0.98	0.98	0.99	0.99	0.99
All orders – F1-score	0.98	0.97	0.98	0.98	0.98
Accuracy	0.98	0.97	0.98	0.98	0.98

Table 6.28. Error Metrics Obtained with SVM

	SVM	SVM + mRMR	SVM + Relief-F	SVM + F-Class	SVM + Hybrid
Completed orders – Precision	0.99	0.98	0.97	0.95	0.96
Completed orders – Recall	0.91	0.87	0.91	0.89	0.91
Completed orders – F1-score	0.95	0.92	0.94	0.92	0.93
Canceled orders – Precision	0.89	0.85	0.89	0.87	0.88
Canceled orders – Recall	0.98	0.98	0.96	0.94	0.94
Canceled orders – F1-score	0.94	0.91	0.92	0.90	0.91
All orders – Precision	0.94	0.92	0.93	0.91	0.92
All orders – Recall	0.95	0.93	0.93	0.92	0.93
All orders – F1-score	0.94	0.92	0.93	0.91	0.92
Accuracy	0.94	0.92	0.93	0.91	0.92

Table 6.29. Error Metrics Obtained with DNN

	DNN	DNN + mRMR	DNN + Relief-F	DNN + F-Class	DNN + Hybrid
Completed orders – Precision	0.98	0.97	0.95	0.94	0.95
Completed orders – Recall	0.93	0.86	0.90	0.88	0.91
Completed orders – F1-score	0.95	0.91	0.92	0.91	0.93
Canceled orders – Precision	0.91	0.84	0.87	0.86	0.88
Canceled orders – Recall	0.97	0.96	0.93	0.93	0.93
Canceled orders – F1-score	0.94	0.90	0.90	0.89	0.91
All orders – Precision	0.94	0.90	0.91	0.90	0.92
All orders – Recall	0.95	0.91	0.92	0.91	0.92
All orders – F1-score	0.95	0.91	0.91	0.90	0.92
Accuracy	0.95	0.91	0.91	0.90	0.92

6.4.3. Binary Classification Based Return Prediction Models

In Table 6.30 through Table 6.34, the error metrics obtained by the models developed with the multi-class classification approach are presented for LR, RF, XGBoost, SVM and DNN algorithms, respectively.

Table 6.30. Error Metrics Obtained with LR

	LR	LR + mRMR	LR + Relief-F	LR + F-Class	LR + Hybrid
Completed orders – Precision	0.79	0.65	0.79	0.79	0.79
Completed orders – Recall	0.91	0.91	0.91	0.91	0.91
Completed orders – F1-score	0.84	0.76	0.85	0.84	0.85
Returned orders – Precision	0.84	0.74	0.85	0.85	0.85
Returned orders – Recall	0.68	0.34	0.69	0.67	0.68
Returned orders – F1-score	0.75	0.46	0.76	0.75	0.76
All orders – Precision	0.82	0.69	0.82	0.82	0.82
All orders – Recall	0.79	0.62	0.80	0.79	0.79
All orders – F1-score	0.80	0.61	0.80	0.80	0.80
Accuracy	0.81	0.67	0.81	0.81	0.81

Table 6.31. Error Metrics Obtained with RF

	RF	RF + mRMR	RF + Relief-F	RF + F-Class	RF + Hybrid
Completed orders – Precision	0.98	0.97	0.98	0.97	0.98
Completed orders – Recall	0.94	0.89	0.94	0.94	0.94
Completed orders – F1-score	0.96	0.93	0.96	0.96	0.96
Returned orders – Precision	0.92	0.87	0.93	0.93	0.93
Returned orders – Recall	0.97	0.97	0.97	0.97	0.97
Returned orders – F1-score	0.95	0.91	0.95	0.95	0.95
All orders – Precision	0.95	0.92	0.95	0.95	0.95
All orders – Recall	0.95	0.93	0.96	0.95	0.96
All orders – F1-score	0.95	0.92	0.96	0.95	0.95
Accuracy	0.95	0.92	0.96	0.95	0.95

Table 6.32. Error Metrics Obtained with XGBoost

	XGBoost	XGBoost + mRMR	XGBoost + Relief-F	XGBoost + F-Class	XGBoost + Hybrid
Completed orders – Precision	0.97	0.97	0.98	0.97	0.98
Completed orders – Recall	0.94	0.90	0.94	0.94	0.94
Completed orders – F1-score	0.96	0.93	0.96	0.96	0.96
Returned orders – Precision	0.92	0.88	0.92	0.92	0.92
Returned orders – Recall	0.97	0.96	0.97	0.97	0.97
Returned orders – F1-score	0.95	0.92	0.95	0.94	0.95
All orders – Precision	0.95	0.92	0.95	0.95	0.95
All orders – Recall	0.95	0.93	0.95	0.95	0.96
All orders – F1-score	0.95	0.92	0.95	0.95	0.95
Accuracy	0.95	0.92	0.95	0.95	0.95

Table 6.33. Error Metrics Obtained with SVM

	SVM	SVM + mRMR	SVM + Relief-F	SVM + F-Class	SVM + Hybrid
Completed orders – Precision	0.95	0.93	0.96	0.95	0.96
Completed orders – Recall	0.90	0.86	0.92	0.90	0.91
Completed orders – F1-score	0.92	0.89	0.94	0.92	0.93
Returned orders – Precision	0.88	0.83	0.90	0.87	0.89
Returned orders – Recall	0.93	0.92	0.95	0.94	0.95
Returned orders – F1-score	0.90	0.87	0.92	0.91	0.91
All orders – Precision	0.91	0.88	0.93	0.91	0.92
All orders – Recall	0.92	0.89	0.93	0.92	0.93
All orders – F1-score	0.91	0.88	0.93	0.92	0.92
Accuracy	0.92	0.88	0.93	0.92	0.92

Table 6.34. Error Metrics Obtained with DNN

	DNN	DNN + mRMR	DNN + Relief-F	DNN + F-Class	DNN + Hybrid
Completed orders – Precision	0.94	0.92	0.93	0.92	0.93
Completed orders – Recall	0.92	0.84	0.90	0.90	0.90
Completed orders – F1-score	0.93	0.88	0.92	0.91	0.92
Returned orders – Precision	0.90	0.81	0.88	0.87	0.87
Returned orders – Recall	0.92	0.90	0.90	0.90	0.91
Returned orders – F1-score	0.91	0.85	0.89	0.89	0.89
All orders – Precision	0.92	0.86	0.90	0.90	0.90
All orders – Recall	0.92	0.87	0.90	0.90	0.91
All orders – F1-score	0.92	0.86	0.90	0.90	0.90
Accuracy	0.92	0.87	0.90	0.90	0.90

6.5. Third Experimental Study: Independent Test Set Evaluation of Selected Models

The error metrics of the models developed with RF in the first approach and with RF+Hybrid (cancellation prediction) and XGBoost+F-Class (return prediction), respectively, in the second approach are presented in Tables 6.35, 6.36, and 6.37.

Table 6.35. Error Metrics Obtained with RF

	RF
Completed orders – Precision	0.96
Completed orders – Recall	1.00
Completed orders – F1-score	0.98
Canceled orders – Precision	0.98
Canceled orders – Recall	0.91
Canceled orders – F1-score	0.95
Returned orders – Precision	0.92
Returned orders – Recall	0.97
Returned orders – F1-score	0.95
All orders – Precision	0.95
All orders – Recall	0.96
All orders – F1-score	0.96
Accuracy	0.96

Table 6.36. Error Metrics Obtained with RF + Hybrid

	RF + Hybrid
Completed orders – Precision	1.00
Completed orders – Recall	0.95
Completed orders – F1-score	0.98
Canceled orders – Precision	0.94
Canceled orders – Recall	1.00
Canceled orders – F1-score	0.97
All orders – Precision	0.97
All orders – Recall	0.98
All orders – F1-score	0.97
Accuracy	0.97

Table 6.37. Error Metrics Obtained with XGBoost + F-Classification

	XGBoost + F-Class
Completed orders – Precision	0.99
Completed orders – Recall	0.90
Completed orders – F1-score	0.94
Returned orders – Precision	0.88
Returned orders – Recall	0.98
Returned orders – F1-score	0.93
All orders – Precision	0.93
All orders – Recall	0.94
All orders – F1-score	0.93
Accuracy	0.93

6.6. Model Performance Evaluation Based on Studies

6.6.1. First Experimental Study

Approach-Based Evaluation

- The multi-class classification approach is based on modeling cancellation and refund situations under more than one category. In this approach, RF and XGBoost-based models showed the highest performance and F1 scores reached 0.96. However, SVM-based models showed lower performance, F1 score was limited to 0.92. DNN and LR-based models showed sufficient but relatively lower performance.
- The binary classification approach created more focused prediction models by considering cancellation and refund situations separately. In this approach, RF and XGBoost-based models again showed the highest performance, F1 scores reached up to 0.98. SVM-based models showed performance in the range of 0.94-0.95. DNN and LR-based models fell behind RF and XGBoost-based models.

- In general, the binary classification approach provided better results than multi-class classification in terms of both precision and sensitivity and provided a more effective solution.

Method-Based Evaluation

- RF and XGBoost-based models showed the highest performance in all approaches. While the F1 scores in the first approach were in the range of 0.95-0.96, in the second approach they varied between 0.97-0.98 in cancellation prediction and 0.95-0.98 in return prediction. These models provided the most successful results compared to other models in both cancellation and return prediction.
- SVM-based models showed satisfactory performance especially in the second approach. While the F1 scores were at the level of 0.92 in the first approach, the results in the range of 0.94-0.95 in cancellation prediction and 0.93-0.95 in return prediction were obtained in the second approach. It showed a strong performance in binary classification but fell behind the RF and XGBoost-based models in multi-class classification.
- DNN-based models generally showed limited performance. While F1 scores remained between 0.88-0.89 in the first approach, they varied between 0.91-0.92 in cancellation prediction and 0.87-0.91 in return prediction in the second approach. Lower results were obtained compared to other models, especially in return prediction.
- LR-based models exhibited the lowest performance in all approaches. F1 scores remained between 0.56-0.60 in the first approach, they varied between 0.64-0.72 in cancellation prediction and 0.62-0.80 in return prediction in the second approach. These results show that LR-based models provide limited success in the specific problem.

Evaluation Based on Feature Selection Algorithms

- mRMR method reached F1 scores of 0.92 in the first approach with RF and XGBoost-based models. In the second approach, it performed well in cancellation prediction and reached an F1 score of 0.97, while it remained limited to 0.92 in return prediction. When used with SVM and DNN based models, it generally showed lower performance.
- Relief-F method provided strong results in both approaches with RF and XGBoost based models. In the first and second approaches, F1 scores were generally at the levels of 0.96-

0.98. However, when used with DNN and LR based models, lower performance was observed in these models.

- F-Classification method produced acceptable results in the first and second approaches with RF and XGBoost based models. F1 scores were at the levels of 0.95-0.96. However, it showed low performance with SVM and DNN based models and F1 scores remained between 0.87-0.93.
- Hybrid method has achieved successful results especially in the second approach and F1 scores have reached 0.98 levels in cancellation and return predictions in RF and XGBoost based models. It has also shown a balanced performance in the first approach but remained at similar levels with mRMR and Relief-F methods.

6.6.2. Second Experimental Study

Approach Based Evaluation

- The multi-class classification approach is based on modeling cancellation and refund situations under more than one category. In this approach, RF and XGBoost based models showed the highest performance and F1 scores reached 0.96. However, SVM based models showed a lower performance, F1 score was limited to 0.92. DNN and LR based models showed relatively lower performance.
- The binary classification approach created more focused prediction models by considering cancellation and refund situations separately. In this approach, RF and XGBoost based models again showed the highest performance, F1 scores reached up to 0.98. SVM based models showed performance in the range of 0.94-0.95. DNN and LR based models lagged behind RF and XGBoost based models.
- In general, the binary classification approach provided better results than multi-class classification in terms of both precision and sensitivity and provided a more effective solution.

Method-Based Evaluation

- RF and XGBoost based models showed the highest performance in all approaches. While F1 scores were in the range of 0.95-0.96 in the first approach, they varied between 0.98 in cancellation prediction and 0.95-0.96 in return prediction in the second approach. These models provided the most successful results compared to other models in both cancellation and return prediction.
- SVM based models showed satisfactory performance especially in the second approach. While F1 scores remained at the level of 0.92 in the first approach, results were obtained between 0.94-0.95 in cancellation prediction and 0.89-0.93 in return prediction in the second approach. It showed a strong performance in binary classification but fell behind RF and XGBoost based models in multi-class classification.
- DNN based models showed limited performance in general. In the first approach, F1 scores remained between 0.88-0.89, while in the second approach, they were between 0.94-0.95 for cancellation prediction and 0.92 for return prediction. Lower results were obtained compared to other models, especially in cancellation prediction.
- LR-based models exhibited the lowest performance in all approaches. In the first approach, F1 scores remained between 0.60-0.63, while in the second approach, they were between 0.64-0.72 for cancellation prediction and 0.80 for return prediction. These results show that LR-based models provide limited success in the specific problem.

Evaluation Based on Feature Selection Algorithms

- mRMR method exhibited a generally poor performance and produced lower results compared to models used without feature selection even in RF and XGBoost-based models. It was observed that mRMR could not provide effective feature selection in the first and second approaches.
- Relief-F method showed strong performance with XGBoost and RF-based models and increased F1 scores to 0.96-0.98 levels in the first and second approaches. This method provided consistent and effective results especially in cancellation and refund predictions. However, it was observed that the performance decreased when used with DNN and LR-based models.
- F-Classification method, similar to Relief-F, achieved very successful results with XGBoost and RF-based models. F1 scores were at 0.95-0.96 levels in the first and second approaches.

However, it was observed that this method provided lower performance in DNN and SVM-based models and F1 scores remained in the range of 0.83-0.92.

- Hybrid method showed a very similar performance to Relief-F and F-Classification methods, but generally achieved slightly lower results than these methods. Despite this, it has achieved satisfactory success by reaching F1 scores of 0.95-0.98, especially with XGBoost and RF-based models.

In general,

- The comparative analysis of the first and second experimental studies revealed that the binary classification approach outperforms the multi-class classification framework in terms of accuracy and recall when predicting order cancellations and product returns. Particularly in problem-specific forecasting scenarios, binary models delivered more effective and application-oriented results.
- Across all approaches, models based on RF and XGBoost consistently achieved the highest and most stable performance, positioning them as the most prominent techniques for cancellation and return prediction tasks. While SVM-based models yielded satisfactory results, especially in binary classification scenarios, models based on LR and DNN generally underperformed in comparison to other methods.
- With regard to feature selection algorithms, both Relief-F and the Hybrid method demonstrated stronger and more consistent outcomes, especially when paired with RF and XGBoost. In contrast, the mRMR algorithm consistently underperformed relative to its counterparts.
- The primary distinction between the first and second experimental setups, namely, applying feature selection prior to versus within each fold of cross-validation—did not lead to a significant difference in model performance.
- Finally, in the third study, which utilized an independent test dataset, the top-performing models maintained their high performance, confirming the robustness of the proposed system and indicating its strong potential for successful real-world deployment.

7. CONCLUSIONS

Within the scope of this thesis, machine learning-based classification models were developed for sales cancellation and product return prediction using order data belonging to a company operating in the e-commerce sector. The modeling process was carried out with a three-stage structure: in the first two experimental studies, cross-validation-based modeling was performed on the training data, while in the third study, the most successful models obtained were evaluated on an independent test dataset.

In the first experimental study, feature selection was made before cross-validation; in the second study, feature selection algorithms were applied separately for each layer. Two different modeling approaches were adopted in both studies. In the first approach, completed, canceled and returned orders were handled with a multi-class classification structure in a single model. In the second approach, cancellation and return situations were modeled separately within the framework of binary classification. Thanks to this applied structure, the effect of feature selection timing on model performance could be analyzed.

In the comparative analyzes made on a method basis, it was seen that RF and XGBoost-based models achieved the highest F1 scores in both experimental studies and gave the most successful results. While SVM-based models provide satisfactory results especially in binary classification structures; DNN and LR-based models generally lag behind other methods. When evaluated in terms of feature selection algorithms, it was determined that Relief-F and Hybrid approaches provide higher success and consistency especially when used with RF and XGBoost-based models. On the other hand, mRMR algorithm exhibited relatively low performance in both experimental studies. As a result of the comparative analysis of the first and second experimental studies, it was seen that the binary classification approach provides higher accuracy and precision than multi-class classification in terms of sales cancellation and product return prediction. That approach, which provides more effective results especially in cases requiring problem-oriented prediction, provides an advantage in terms of application success.

In future studies, it is important to evaluate scenarios where class imbalance is more pronounced. In the data set used in this thesis, completed, canceled and returned orders are equally balanced, and it is known that cancellation and return rates are generally quite low in real-world data. Therefore, in order to better analyze the effects of unbalanced class structures, remodeling can be done with real datasets with unbalanced class distribution. In addition, testing advanced approaches such as ensemble learning has the potential to further improve model performance.

8. REFERENCES

Al Imran, A., & Amin, M. N. (2020). Predicting the return of orders in the e-tail industry accompanying with model interpretation. Procedia Computer Science, 176, 1170–1179.

Abdulla, H., Letizia, P., & Souza, G. (2023). Order cancellation behavior in online retailing: An empirical investigation. SSRN. https://doi.org/10.2139/ssrn.4406252

Adebanjo, S., Banchani, E., & Mishiwo, S. H. (2023). Development of machine learning techniques for determining optimal pricing and inventory policy with order cancellation under the cash-on-delivery payment scheme. Advance Journal of Current Research, 8(12), 18–41.

Bonaccorso, G. (2017). Machine Learning Algorithms (pp. 167–170).

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). San Francisco.

Chidroop, I., & Moharir, M. (2020). Predicting the propensity of order cancellation in the e-commerce domain.

Clement, L., & Spinler, S. (2025). From returns to re-usage: A data-driven strategy for sustainable packaging–a case study in e-commerce. Journal of Cleaner Production, 145584.

Cui, H., Rajagopalan, S., & Ward, A. R. (2020). Predicting product return volume using machine learning methods. European Journal of Operational Research, 281(3), 612–627.

Dzyabura, D., El Kihal, S., Hauser, J. R., & Ibragimov, M. (2023). Leveraging the power of images in managing product return rates. Marketing Science, 42(6), 1125–1142.

Fan Zhang, & O'Donnell, L. J. (2020). Support vector regression. In Machine Learning (pp. 123–140). Academic Press.

Gian, M., & Ikate, S. (2021). Development of electronic business from the historical point of view of an e-commerce concept. Journal Dimensie Management and Public Sector, 2(2), 19–24.

Gouveia, F. D., & São Mamede, H. (2022). Digital transformation for SMEs in the retail industry. Procedia Computer Science, 204, 671–681.

Granov, A. (2021). Customer loyalty, return and churn prediction through machine learning methods: For a Swedish fashion and e-commerce company.

Joshi, T., Mukherjee, A., & Ippadi, G. (2018, August). One size does not fit all: Predicting product returns in e-commerce platforms. In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 926–927). IEEE.

Karl, D. (2024). Evaluating advanced product return forecasting algorithms: A (meta-) review integrating consumer returns research. International Journal of Business Forecasting and Marketing Intelligence, 9(2), 213–241.

Khouja, M., & Hammami, R. (2023). Optimizing price, order quantity, and return policy in the presence of consumer opportunistic behavior for online retailers. European Journal of Operational Research, 309(2), 683–703.

Kriegeskorte, N., & Golan, T. (2019). Neural network models and deep learning – A primer for biologists. arXiv, 1–14.

Li, J., He, J., & Zhu, Y. (2018, July). E-tail product return prediction via hypergraph-based local graph cut. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 519–527).

Ntumba, C., Aguayo, S., & Maina, K. (2023). Revolutionizing retail: A mini review of e-commerce evolution. Journal of Digital Marketing and Communication, 3(2), 100–110.

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. Journal of Clinical Epidemiology, 49(12), 1373–1379.

Rajasekaran, V., & Priyadarshini, R. (2021). An e-commerce prototype for predicting the product return phenomenon using optimization and regression techniques. In Advances in Computing and Data Sciences: 5th International Conference, ICACDS 2021 (pp. 230–240). Springer International Publishing.

Reis, M. F. P. (2024). Forecasting the product return rate in the e-commerce context (Doctoral dissertation).

Şahinbaş, K. (2022). Predicting order cancellations for e-commerce domain: A proposed model based on retailing experience. İnsan ve Toplum Bilimleri Araştırmaları Dergisi, 11(3), 1493–1514.

Sharma, N. (2024). Analyzing customer behavior patterns & predicting online product return intentions: A data mining approach (Doctoral dissertation, Hochschule Rhein-Waal).

Sun, H., Lv, Z., Li, J., Xu, Z., & Sheng, Z. (2023). Will the order be canceled? Order cancellation probability prediction based on deep residual model. Transportation Research Record, 2677(6), 142–160.

Sweidan, D., Johansson, U., Alenljung, B., & Gidenstam, A. (2023, July). Improved decision support for product returns using probabilistic prediction. In 2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE) (pp. 1567–1573). IEEE.

Taher, G. (2021). E-commerce: Advantages and limitations. International Journal of Academic Research in Accounting, Finance and Management Sciences, 11(1), 153–165.

Thibbotuwawa, A., Nanayakkara, P. R., Fernando, W. M., Jayalath, M. M., Perera, H. N., & Nielsen, P. (2023). A reverse logistics network model for handling e-commerce returns. IFAC-PapersOnLine, 56(2), 138–143.

Tüylü, A. N. A., & Eroglu, E. (2022). The prediction of product return rates with ensemble machine learning algorithms. Journal of Engineering Research.

Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J. H. (2018). Relief-based feature selection: Introduction and review. Journal of Biomedical Informatics, 85, 189–203.

Wu, H., Yang, T., Li, H., & Zhou, Z. (2023). Air quality prediction model based on mRMR–RF feature selection and ISSA–LSTM. Scientific Reports, 13(1), 12825.

Xu, W., Zhang, X., Chen, R., & Yang, Z. (2023). How do you say it matters? A multimodal analytics framework for product return prediction in live streaming e-commerce. Decision Support Systems, 172, 113984.

Zhang, W., Xie, R., Quan, P., & Ma, Z. (2025). Product return prediction in live streaming e-commerce with cross-modal contrastive transformer. Decision Support Systems, 114470.

Zhu, Y., Li, J., He, J., Quanz, B. L., & Deshpande, A. A. (2018, July). A local algorithm for product return prediction in e-commerce. In IJCAI (pp. 3718–3724).

9. CURRICULUM VITAE

Zehra Sude SARI. Turkey. She graduated from Piri Reis Anatolian High School and received her B.Sc. degree in Computer Engineering from Çukurova University in 2022. In 2023, she started her M.Sc. studies in the same department. Since 2022, she has been working as an R&D consultant, providing services to various companies. Her consultancy scope includes TÜBİTAK-funded projects, R&D center consultancy, technopark project applications, and European Union (EU) projects. Her main area of expertise focuses on the design, management, and end-to-end execution of machine learning-based projects. As part of this thesis study, two academic papers have been prepared. The first one has been published as a full-text paper in the SETSCI Conference Proceedings:

1. Sarı, Z. S., Taşkapı, B., Dağdaş, G., & Akay, M. F. (2024, November). Development of Machine Learning-Based Sales Cancellation/Return Forecasting Models for the E-Commerce Industry. In SETSCI-Conference Proceedings (Vol. 19, pp. 17–21). SETSCI-Conference Proceedings.
2. Sarı, Z. S., Yurdagül, H. H., Taşkapı, B., Dağdaş, G., & Akay, M. F. (2025, June). Multi-class machine learning models for predicting order cancellations and returns in e-commerce. Proceedings of the 9th International Symposium on Innovative Approaches in Smart Technologies, Gaziantep, Türkiye.

APPENDICES

10.1. First Experimental Study: Pre-Cross-Validation Feature Selection

10.1.1. Multi-Class Classification Based Cancellation-Return Prediction Models

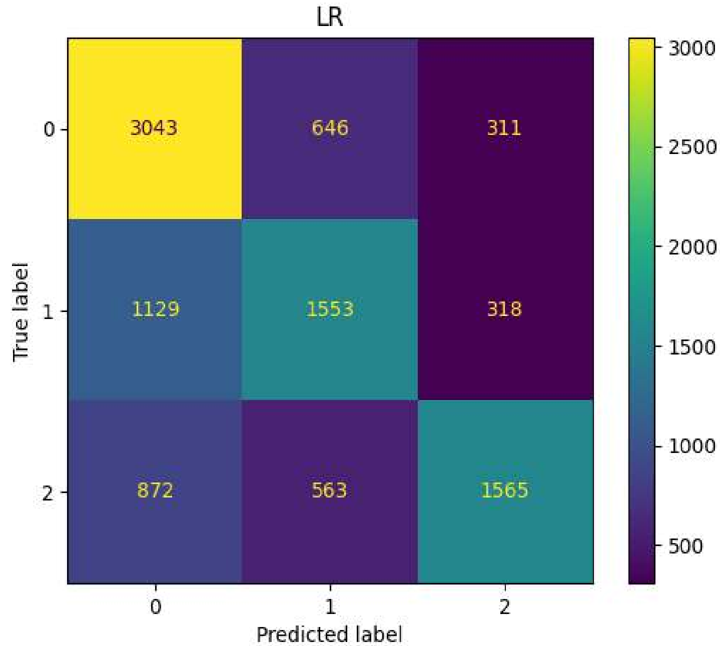


Figure 10.1. Confusion Matrix Obtained with LR

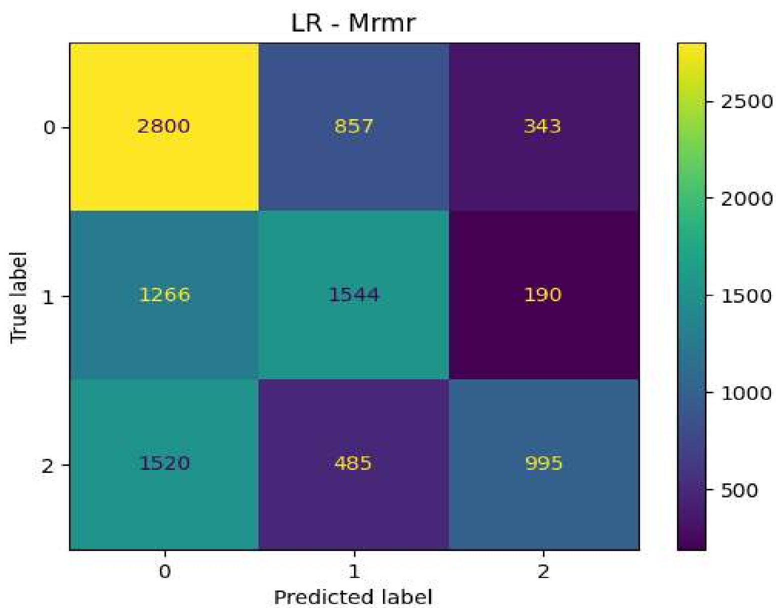


Figure 10.2. Confusion Matrix Obtained with LR+mRMR

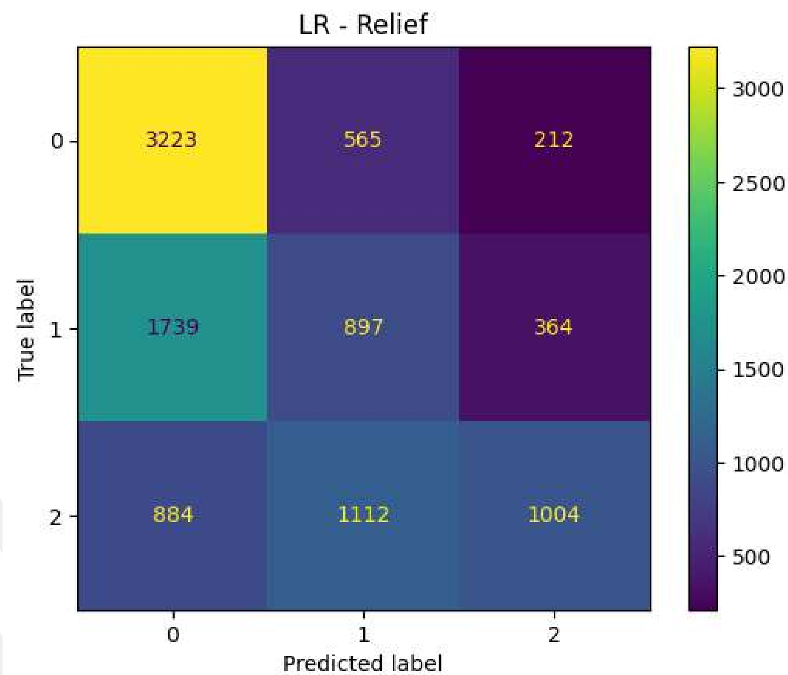


Figure 10.3. Confusion Matrix Obtained with LR+Relief-F

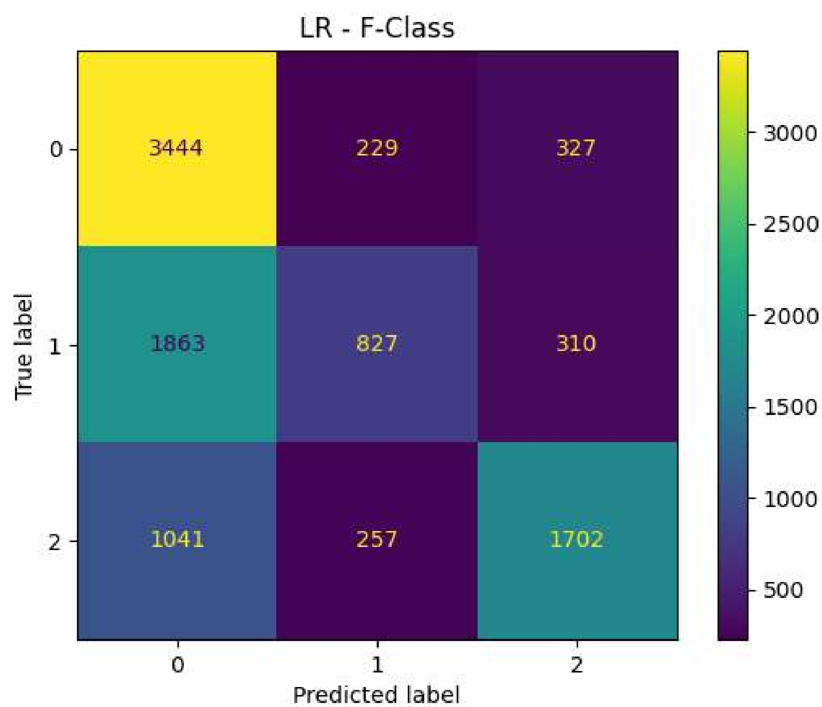


Figure 10.4. Confusion Matrix Obtained with LR+F-Classification

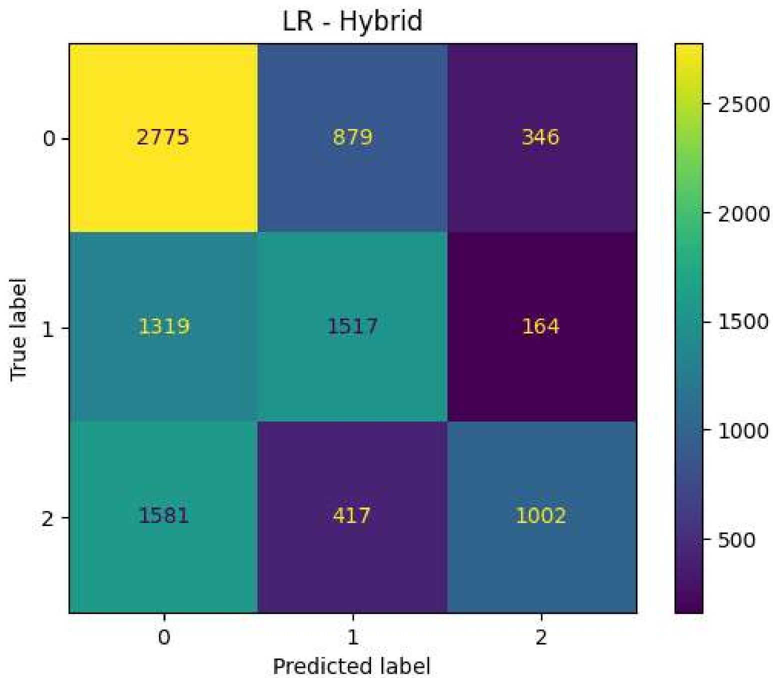


Figure 10.5. Confusion Matrix Obtained with LR+Hybrid

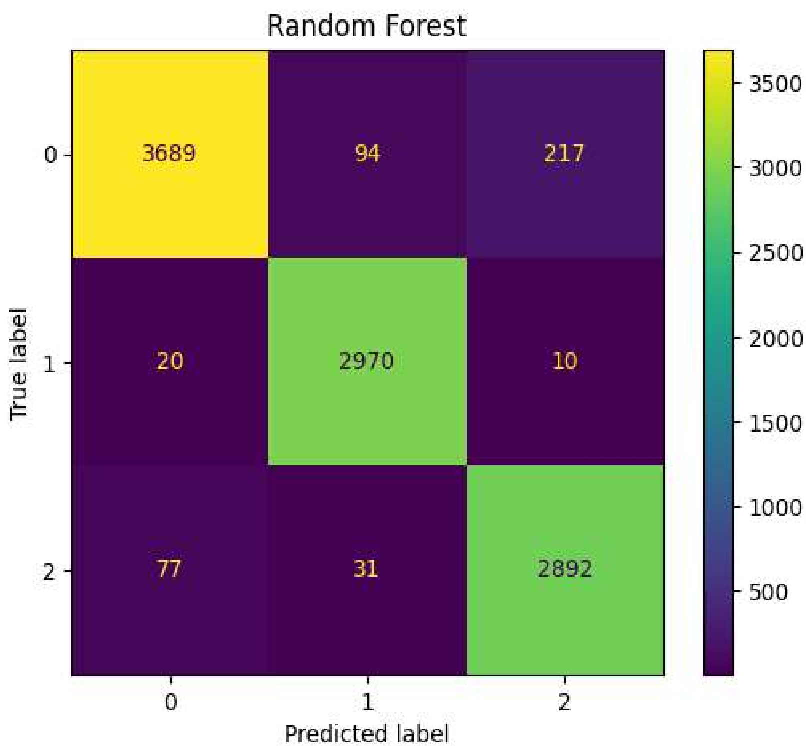


Figure 10.6. Confusion Matrix Obtained with RF

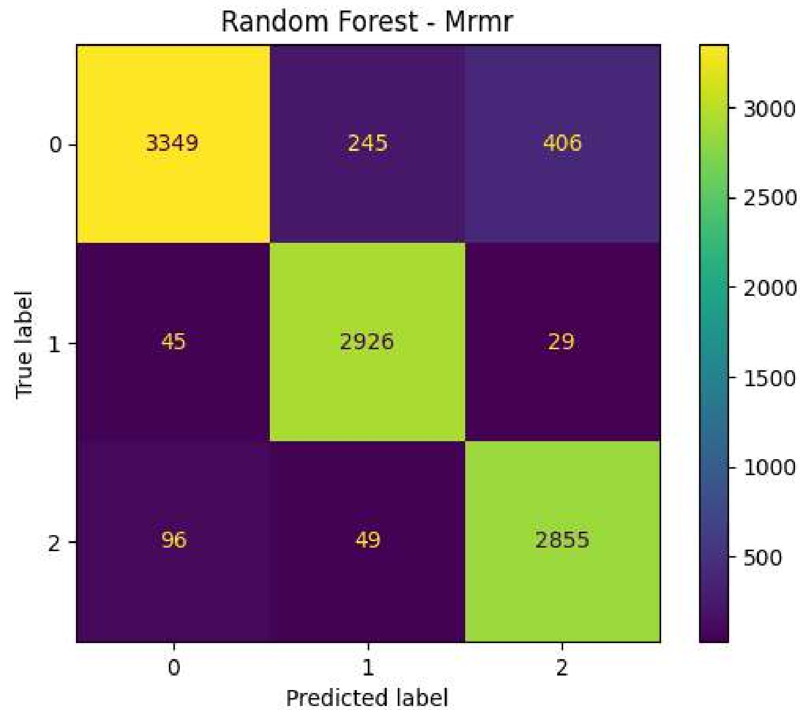


Figure 10.7. Confusion Matrix Obtained with RF+mRMR

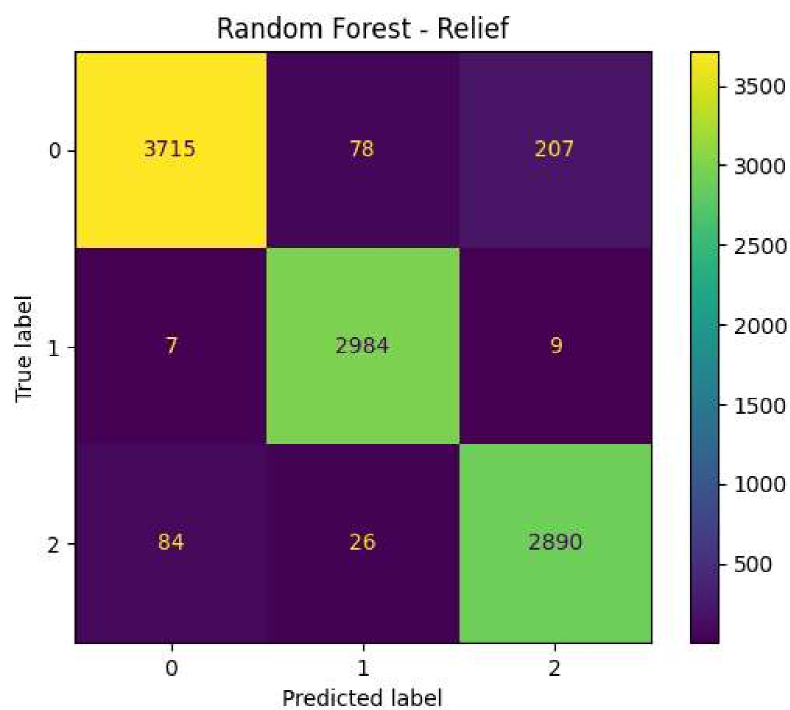


Figure 10.8. Confusion Matrix Obtained with RF+Relief-F

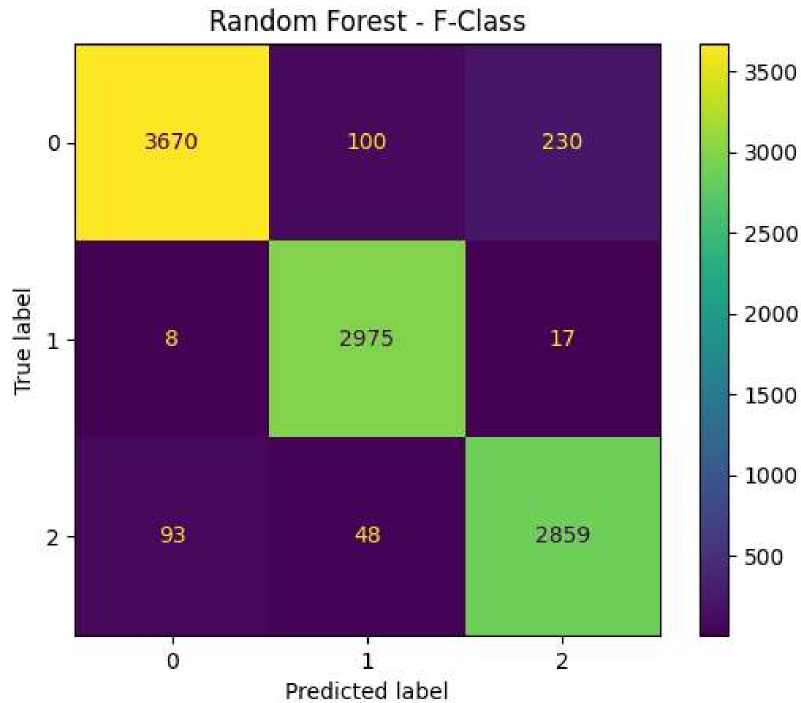


Figure 10.9. Confusion Matrix Obtained with RF+F-Classification

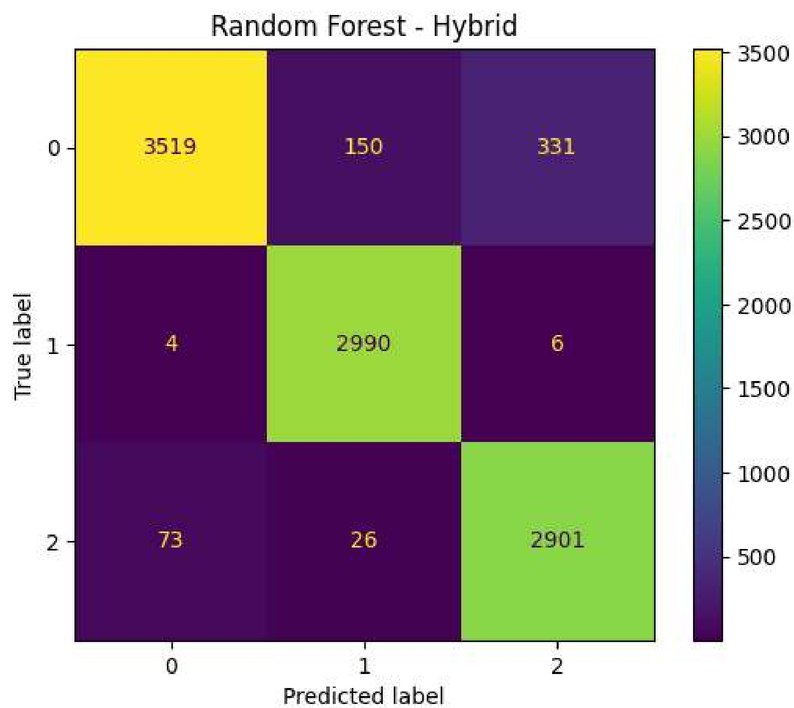
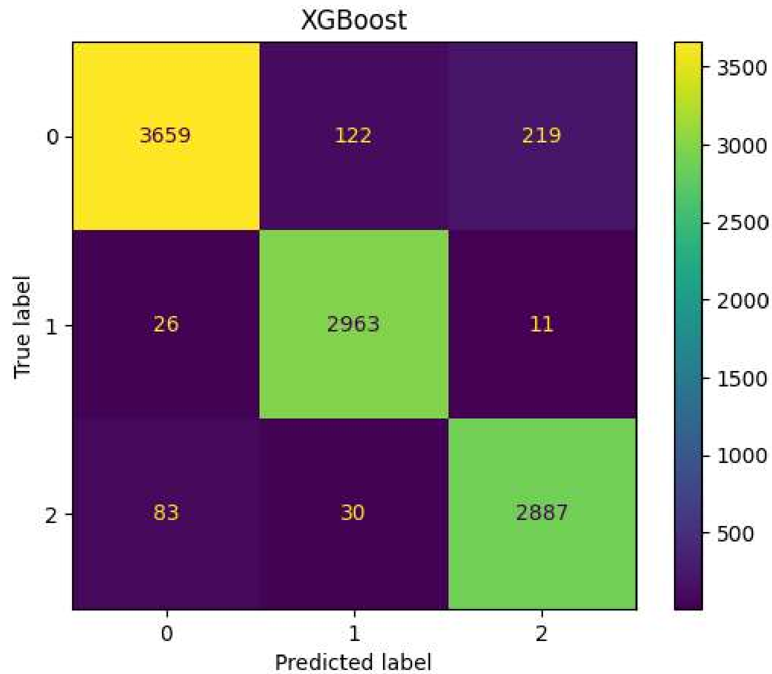


Figure 10.10. Confusion Matrix Obtained with RF+Hybrid

Figure 10.11. Confusion Matrix Obtained with XGBoost



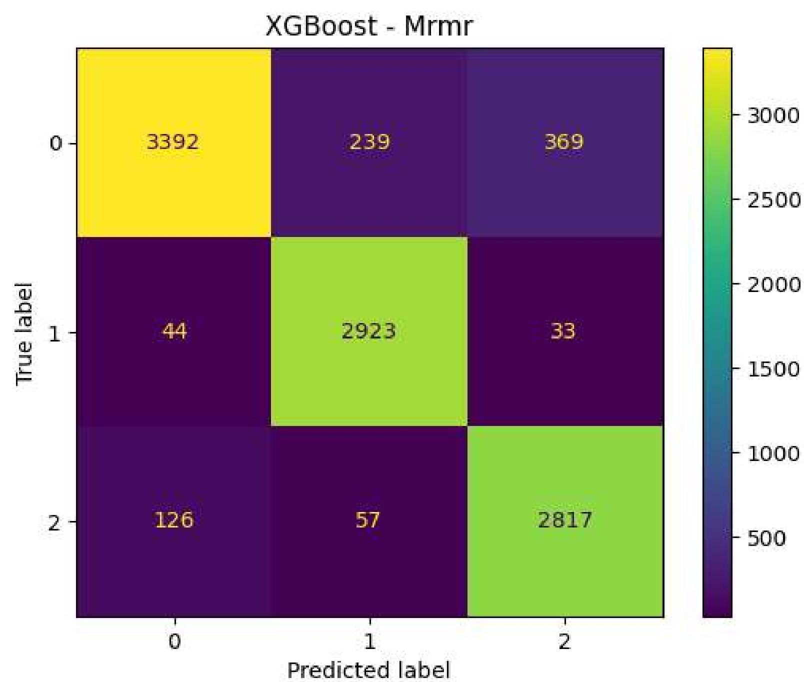


Figure 10.12. Confusion Matrix Obtained with XGBoost+mRMR

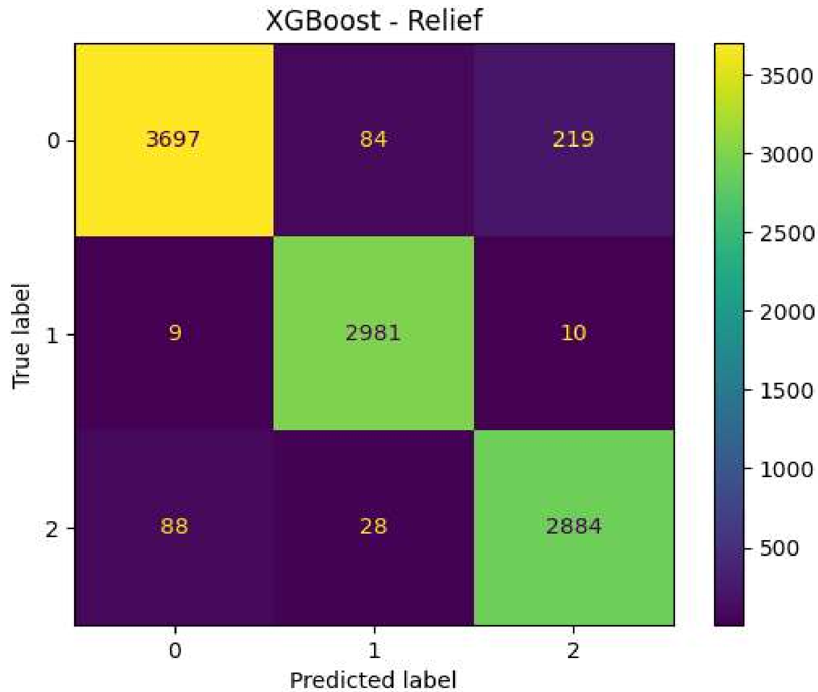


Figure 10.13. Confusion Matrix Obtained with XGBoost+Relief-F

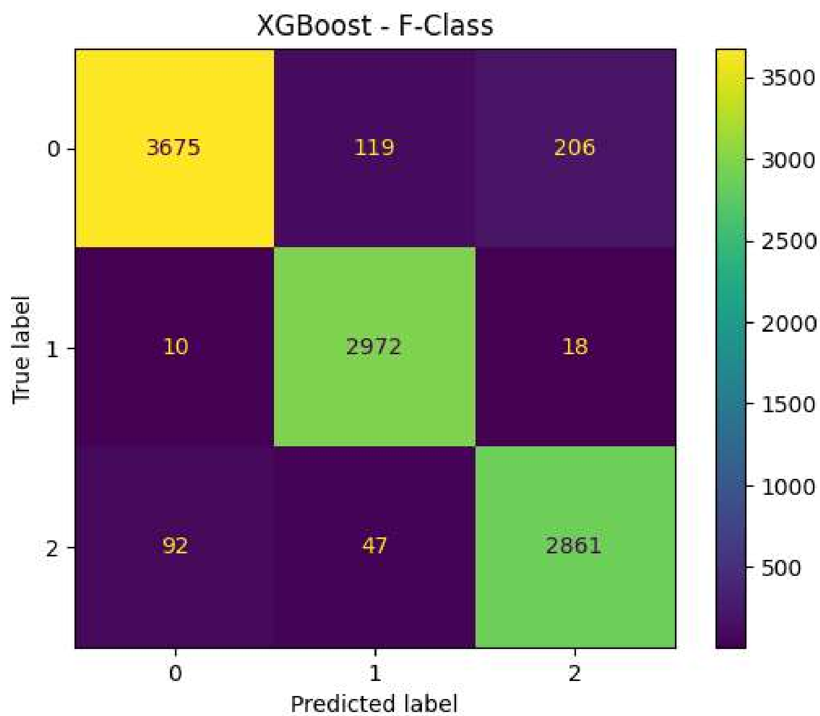


Figure 10.14. Confusion Matrix Obtained with XGBoost+F-Classification

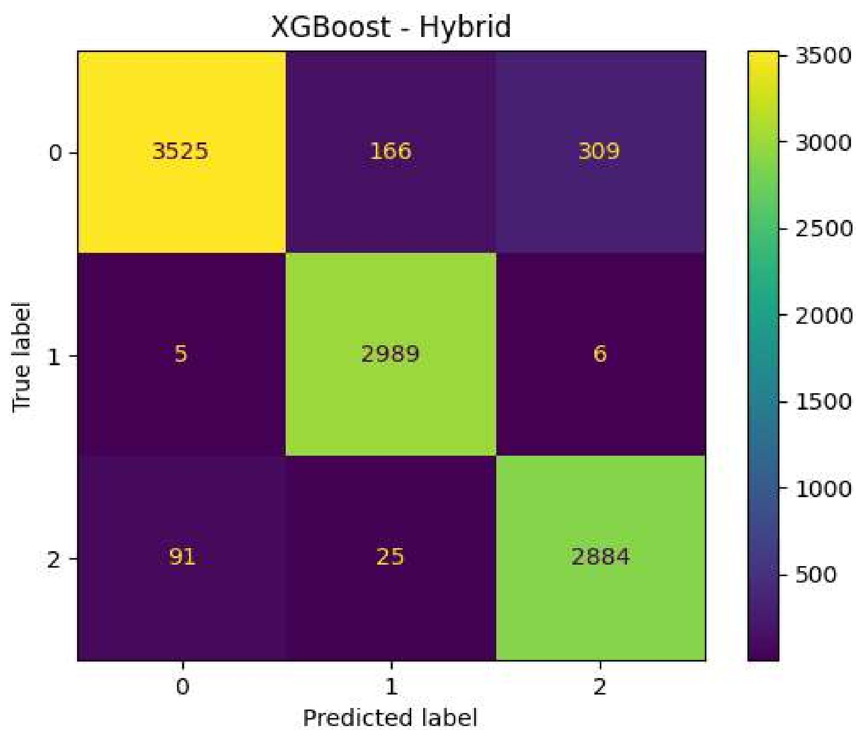


Figure 10.15. Confusion Matrix Obtained with XGBoost+Hybrid

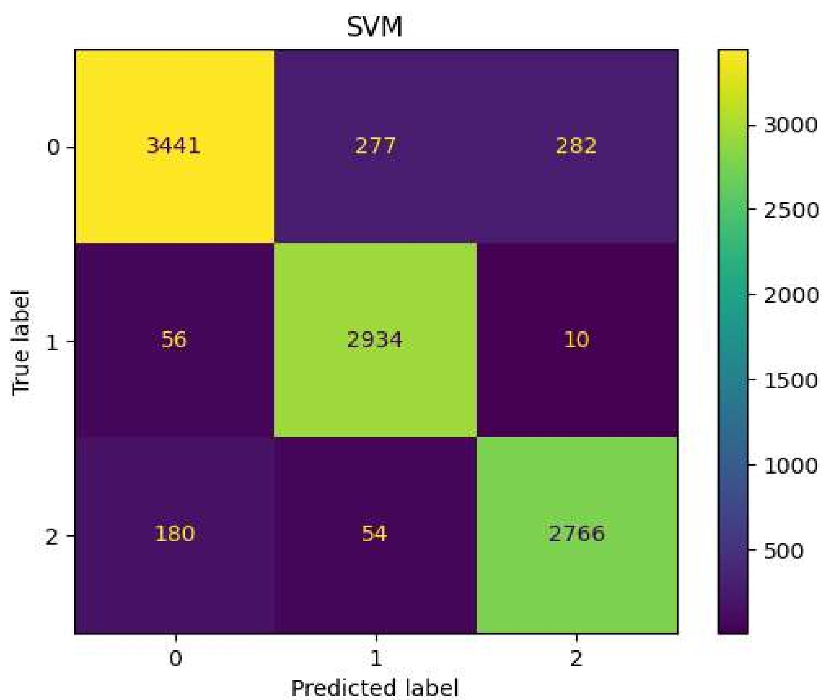


Figure 10.16. Confusion Matrix Obtained with SVM

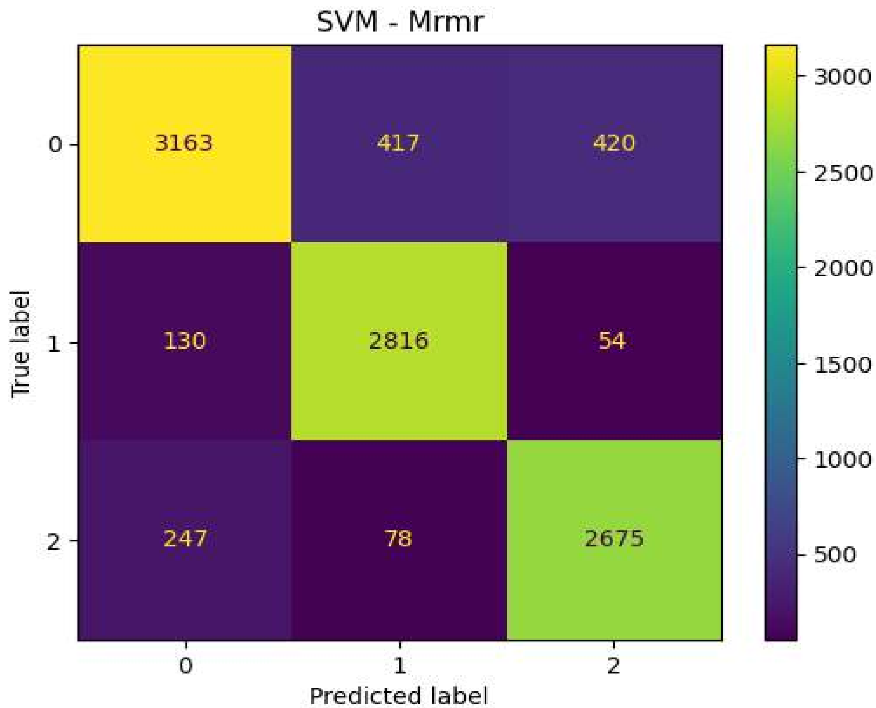


Figure 10.17. Confusion Matrix Obtained with SVM+mRMR

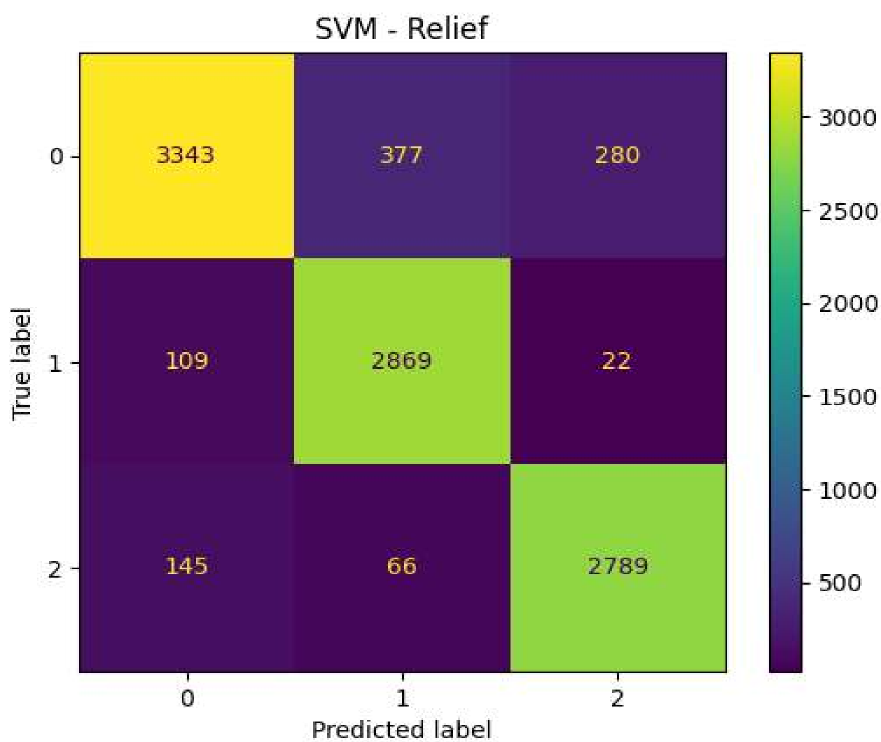


Figure 10.18. Confusion Matrix Obtained with SVM+Relief-F

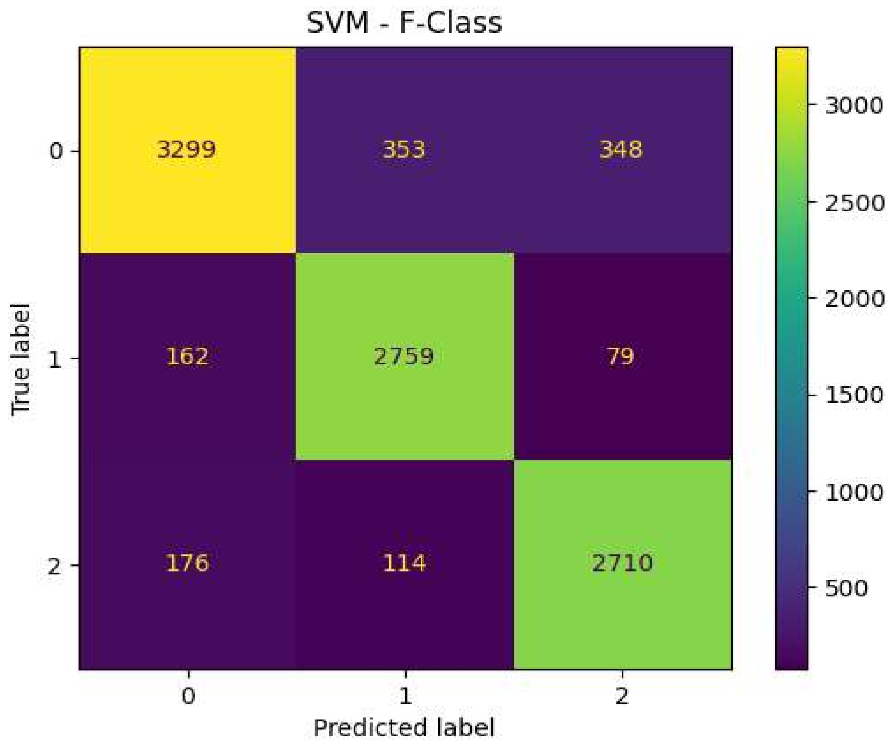


Figure 10.19. Confusion Matrix Obtained with SVM+F-Classification

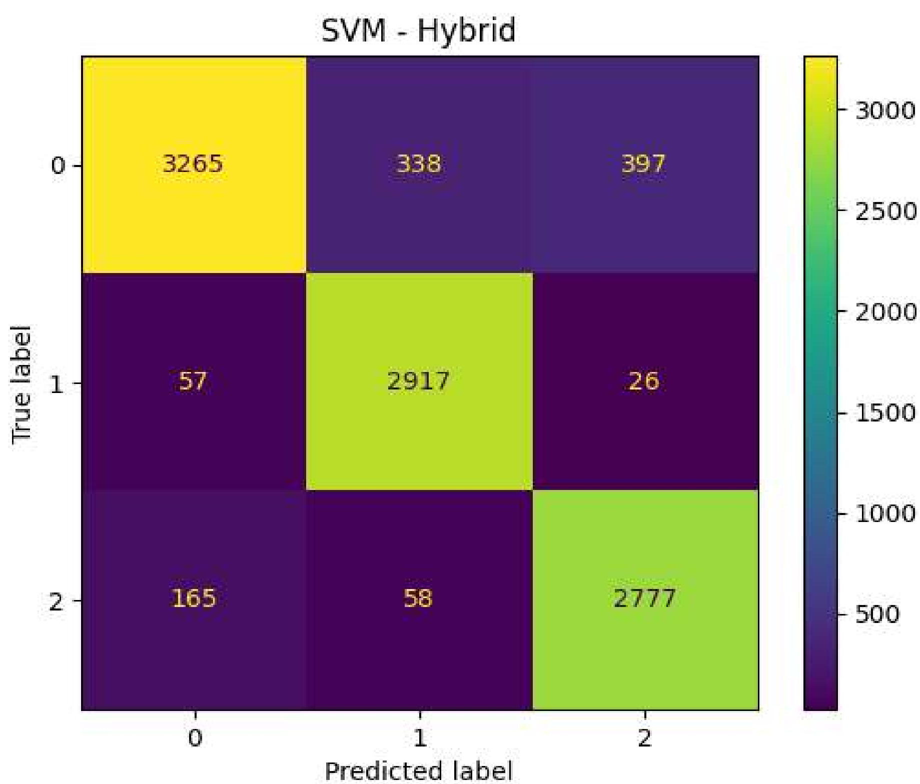


Figure 10.20. Confusion Matrix Obtained with SVM+Hybrid

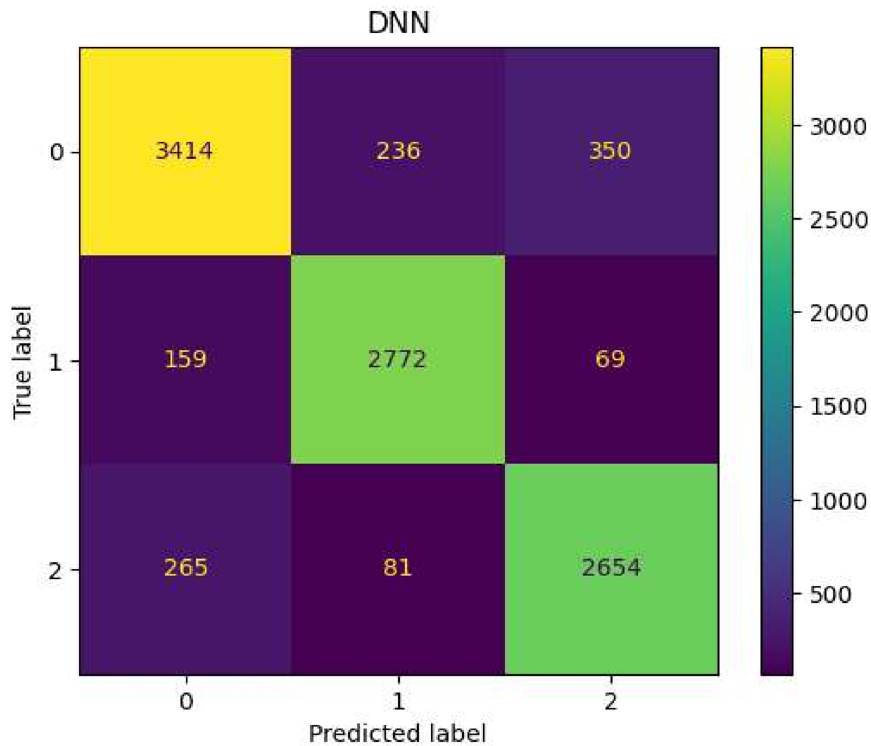


Figure 10.21. Confusion Matrix Obtained with DNN

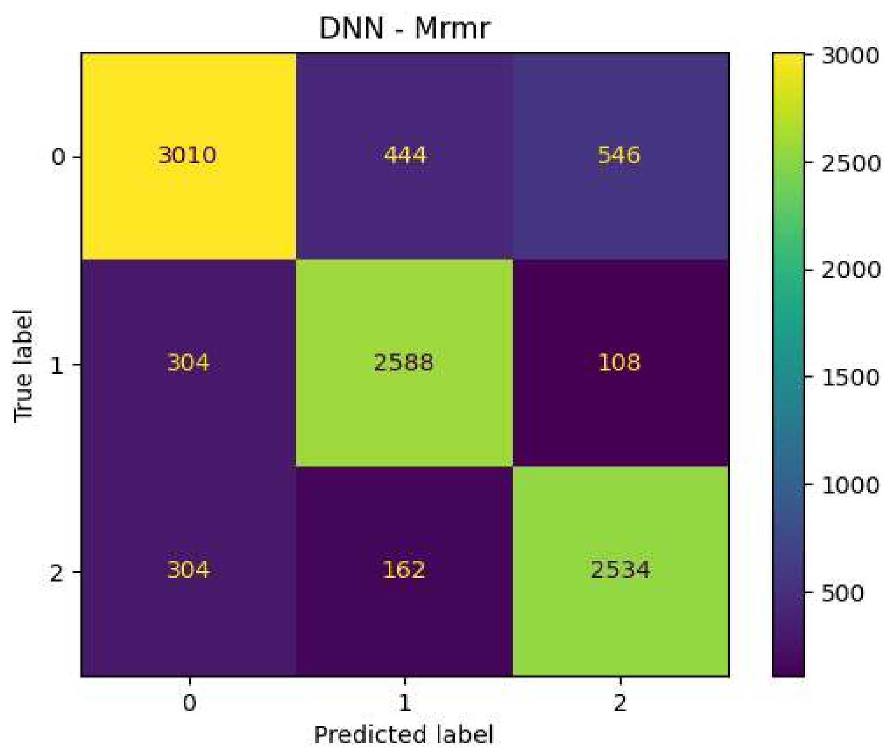


Figure 10.22. Confusion Matrix Obtained with DNN+mRMR

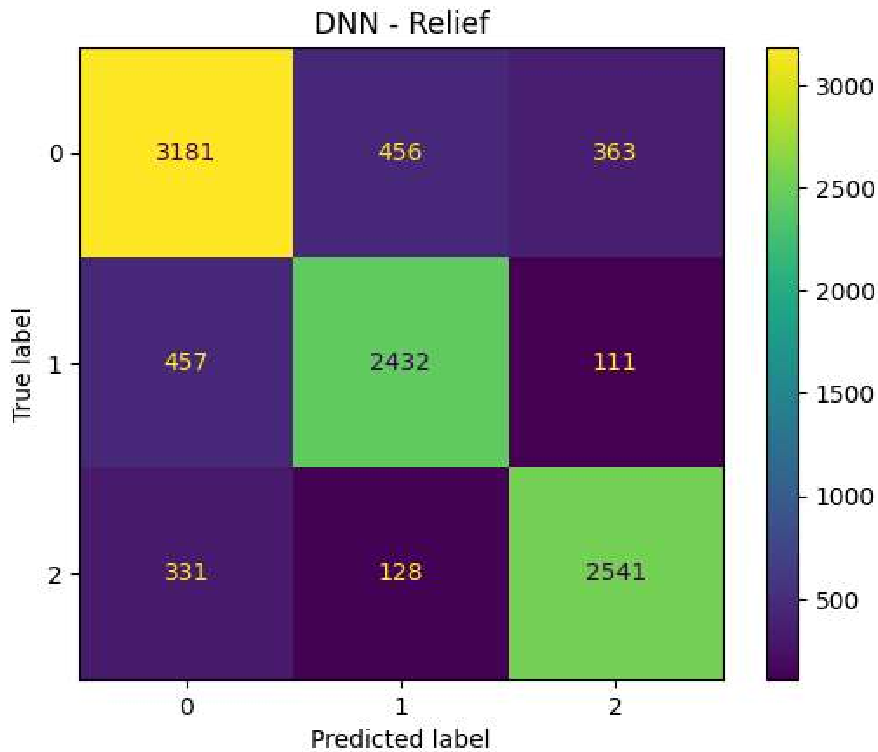


Figure 10.23. Confusion Matrix Obtained with DNN+Relief-F

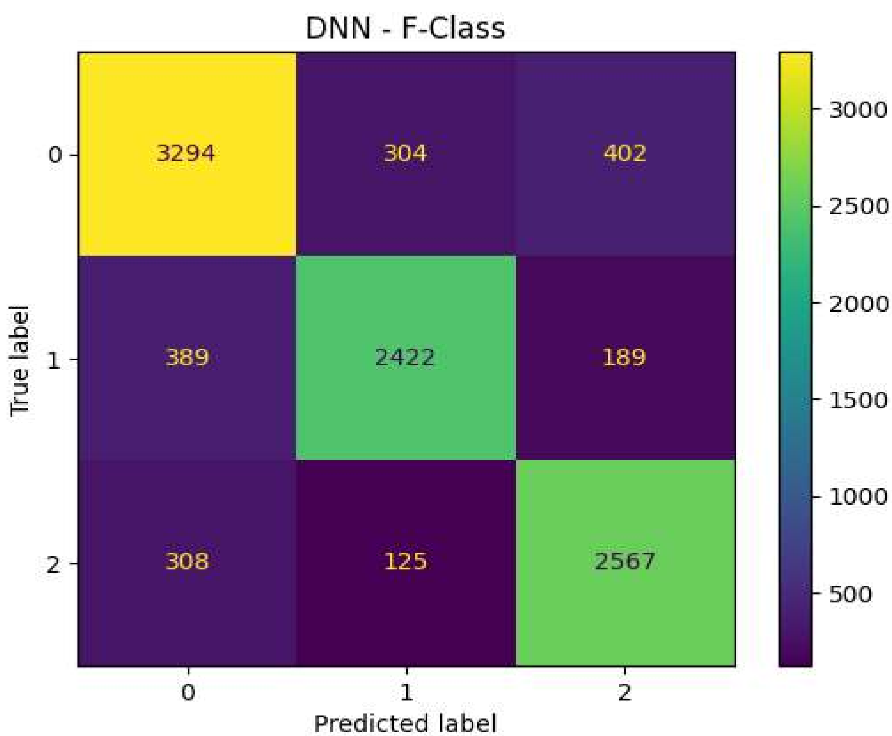


Figure 10.24. Confusion Matrix Obtained with DNN+F-Classification

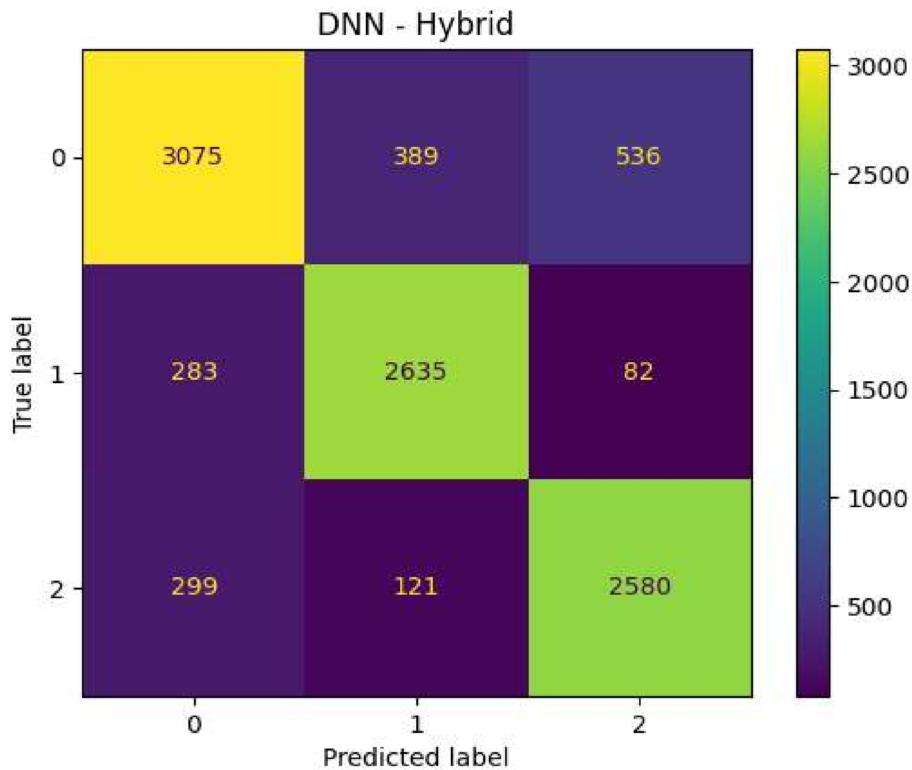


Figure 10.25. Confusion Matrix Obtained with DNN+Hybrid

10.1.2. Binary Classification Based Cancellation Prediction Models

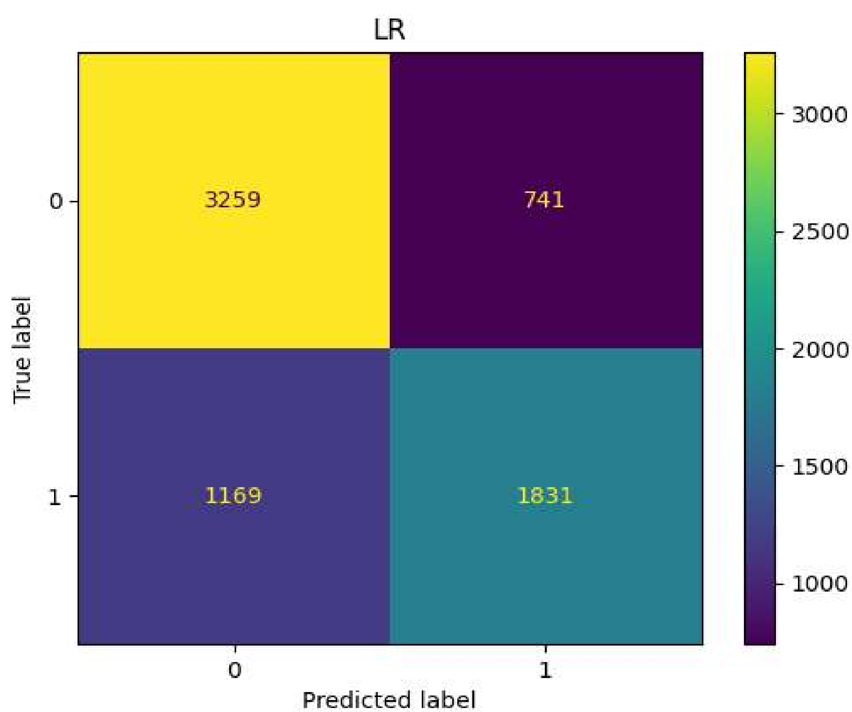


Figure 10.26. Confusion Matrix Obtained with LR

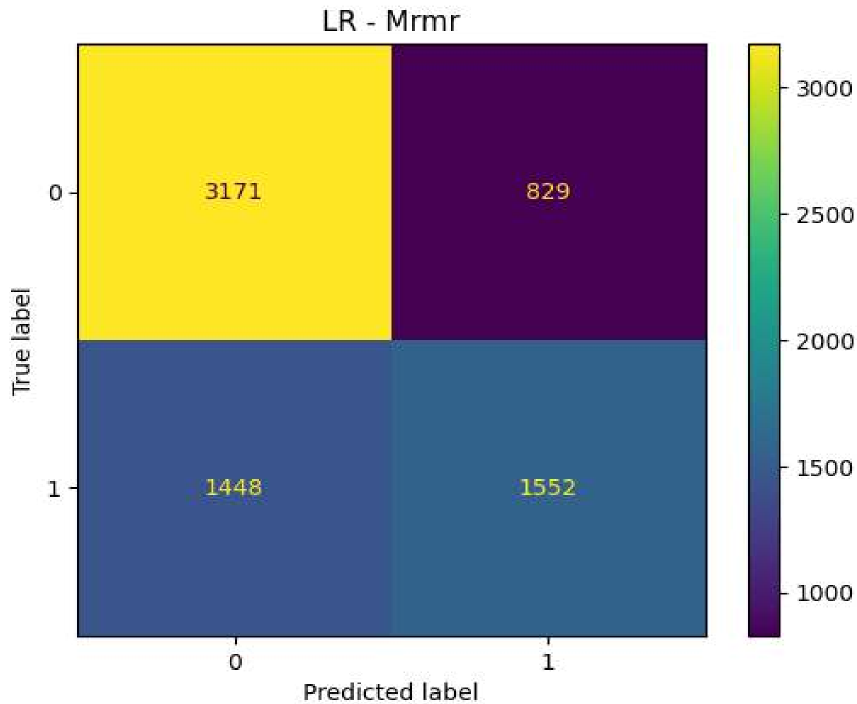


Figure 10.27. Confusion Matrix Obtained with LR+mRMR

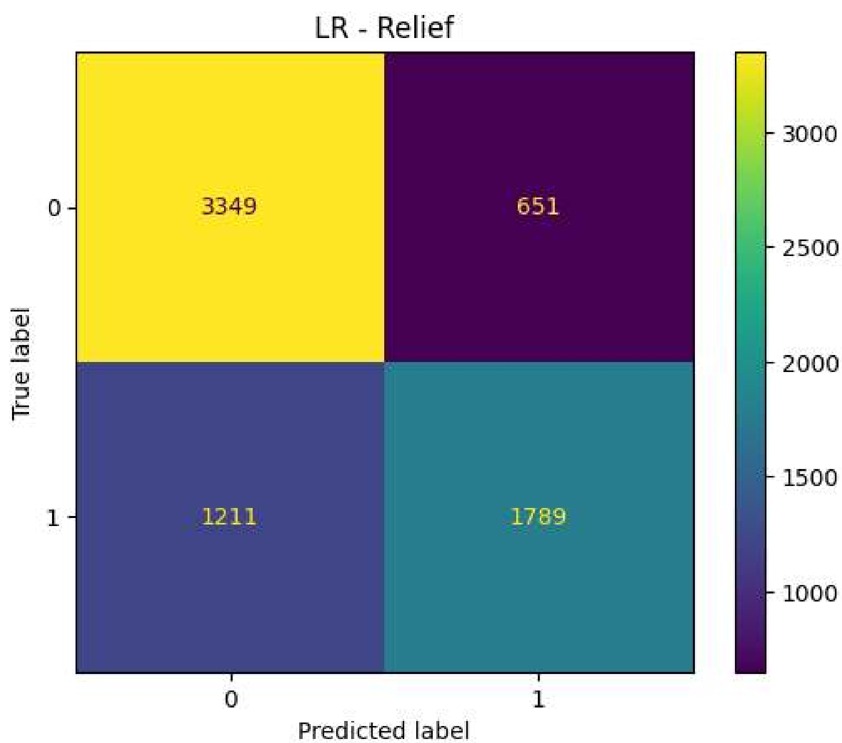


Figure 10.28.Confusion Matrix Obtained with LR+Relief-F

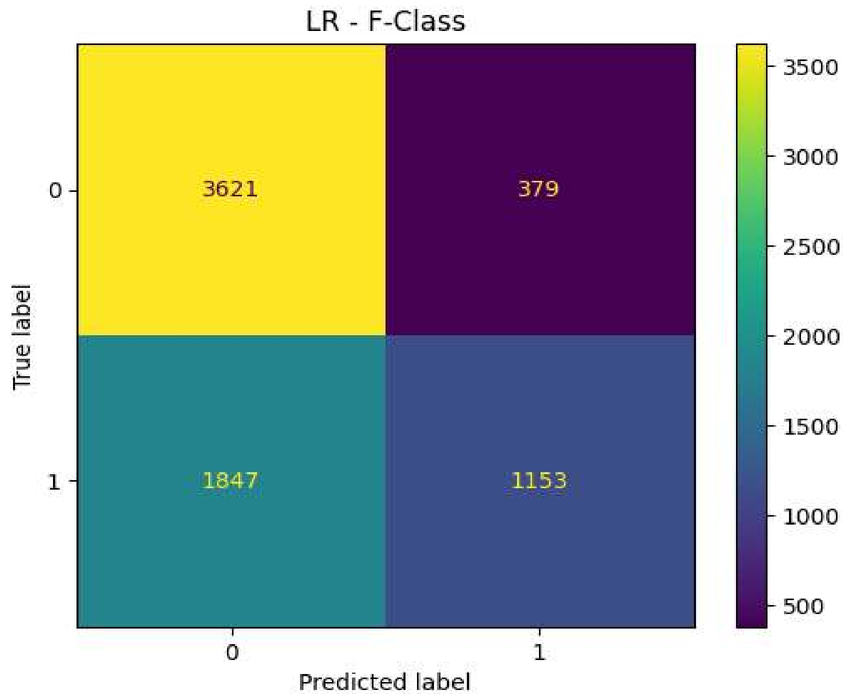


Figure 10.29. Confusion Matrix Obtained with LR+F-Classification

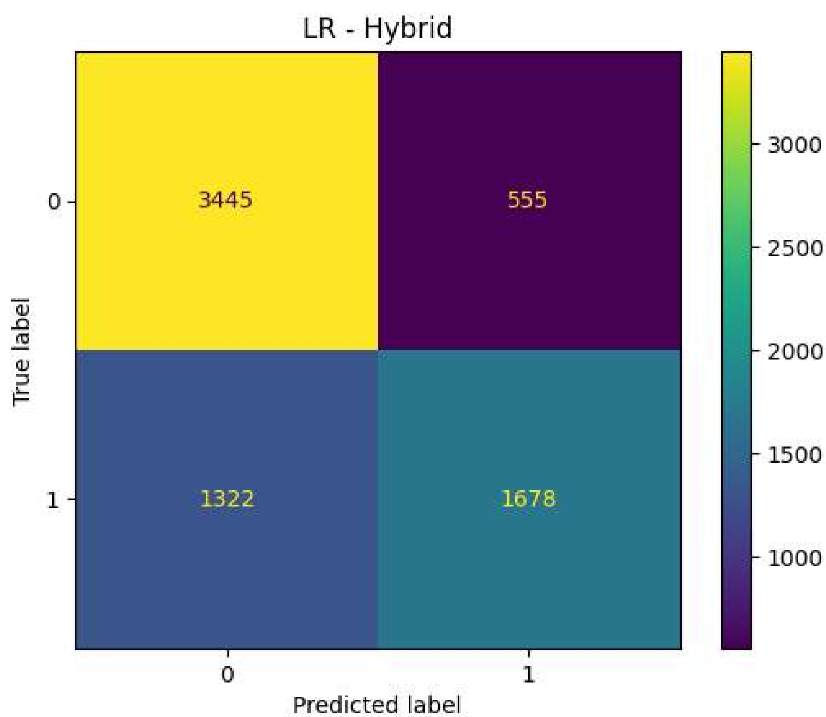


Figure 10.30. Confusion Matrix Obtained with LR+Hybrid

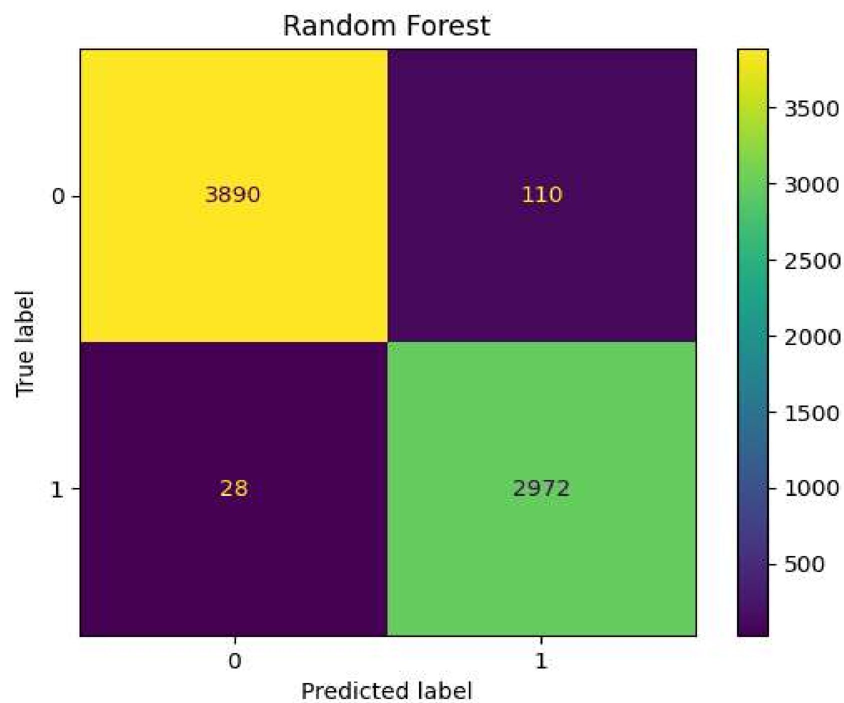


Figure 10.31. Confusion Matrix Obtained with RF

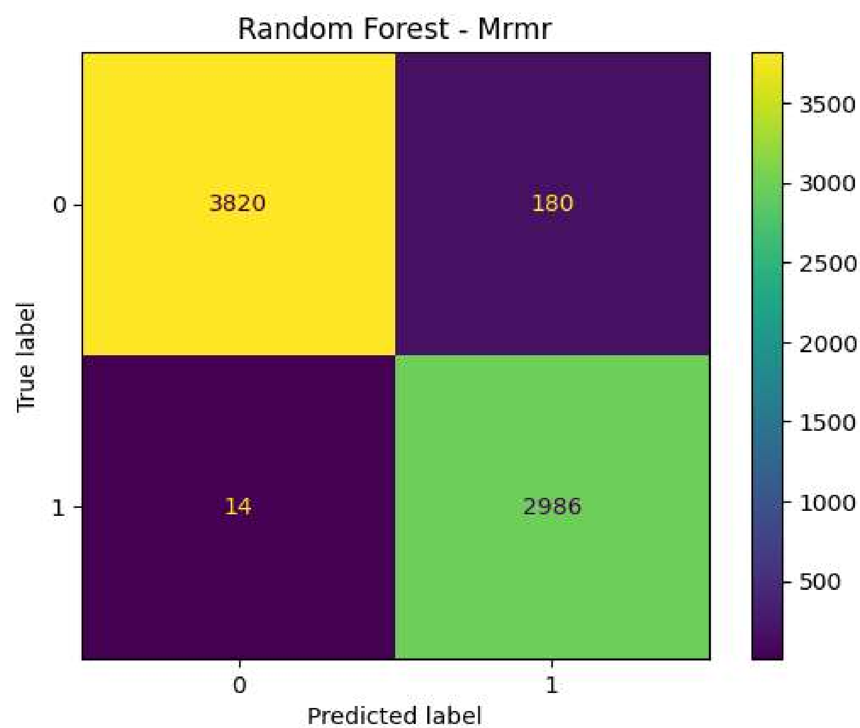


Figure 10.32. Confusion Matrix Obtained with RF+mRMR

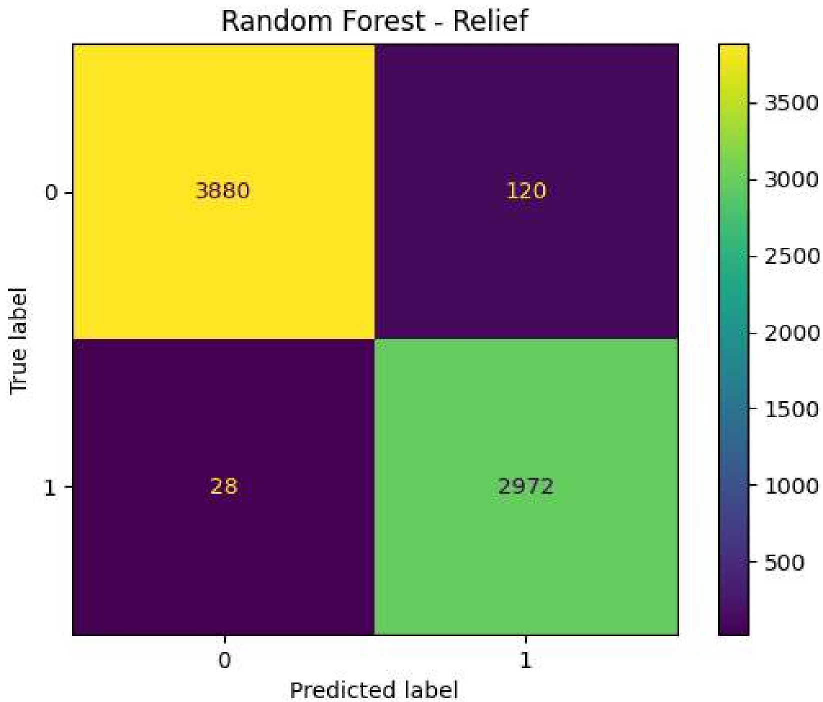


Figure 10.33. Confusion Matrix Obtained with RF+Relief-F

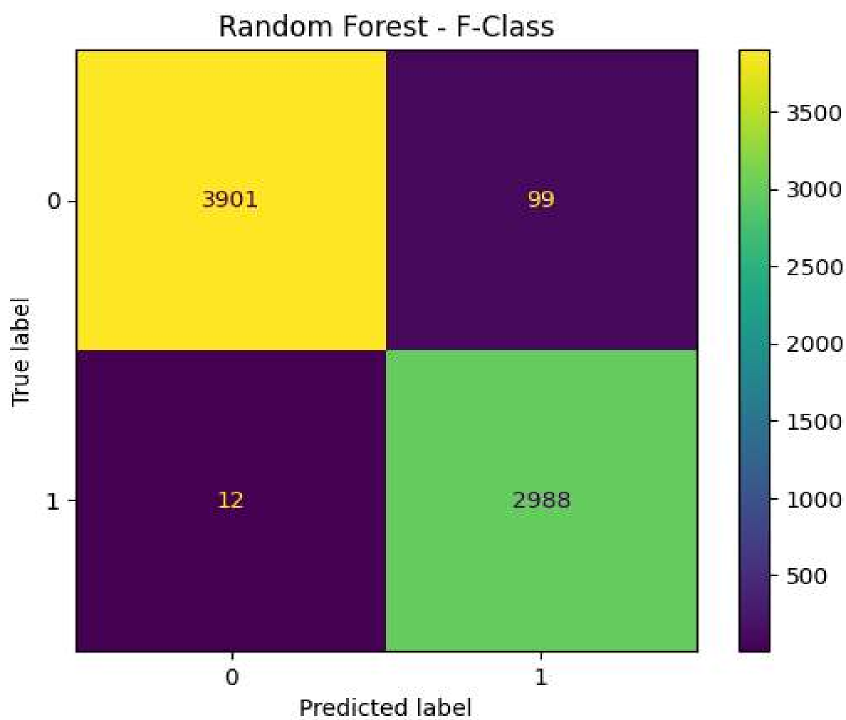


Figure 10.34. Confusion Matrix Obtained with RF+F-Classification

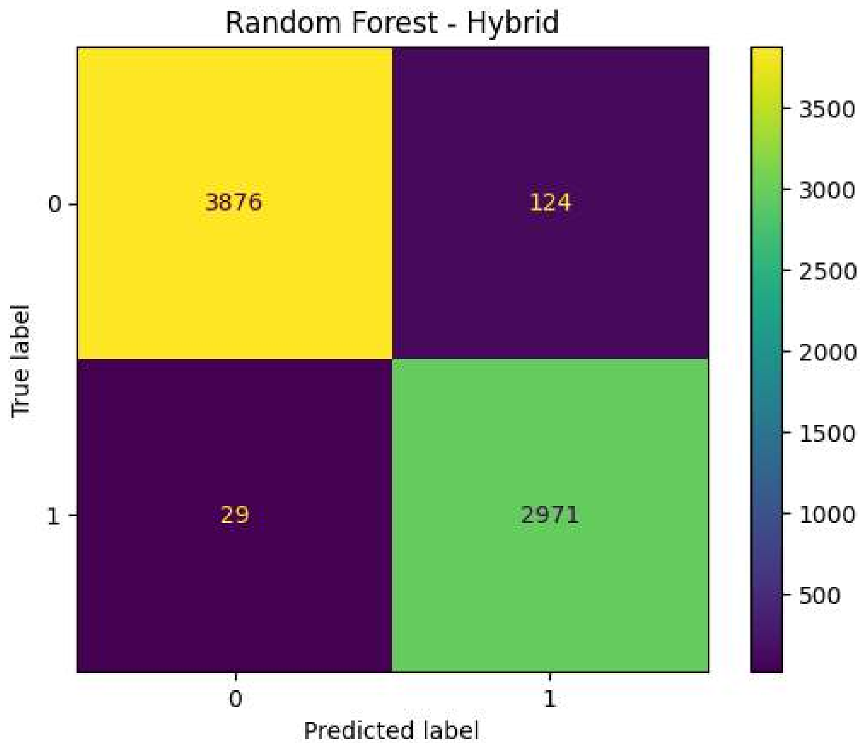


Figure 10.35. Confusion Matrix Obtained with RF+Hybrid

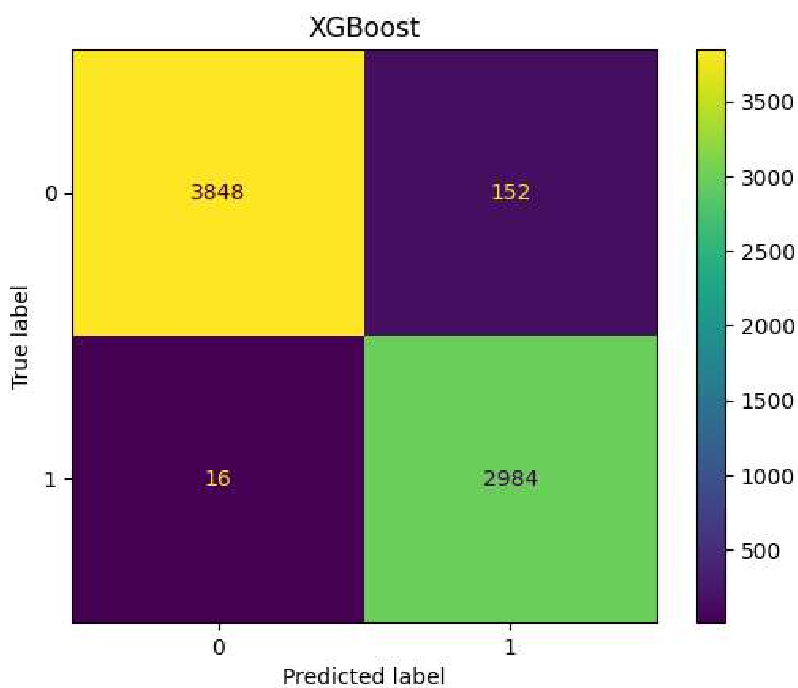


Figure 10.36. Confusion Matrix Obtained with XGBoost

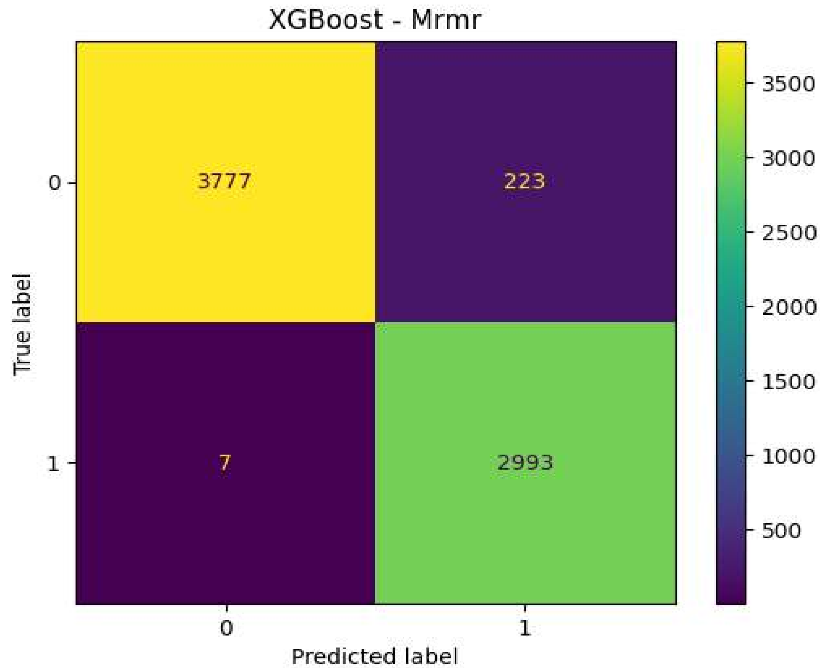


Figure 10.37. Confusion Matrix Obtained with XGBoost+mRMR

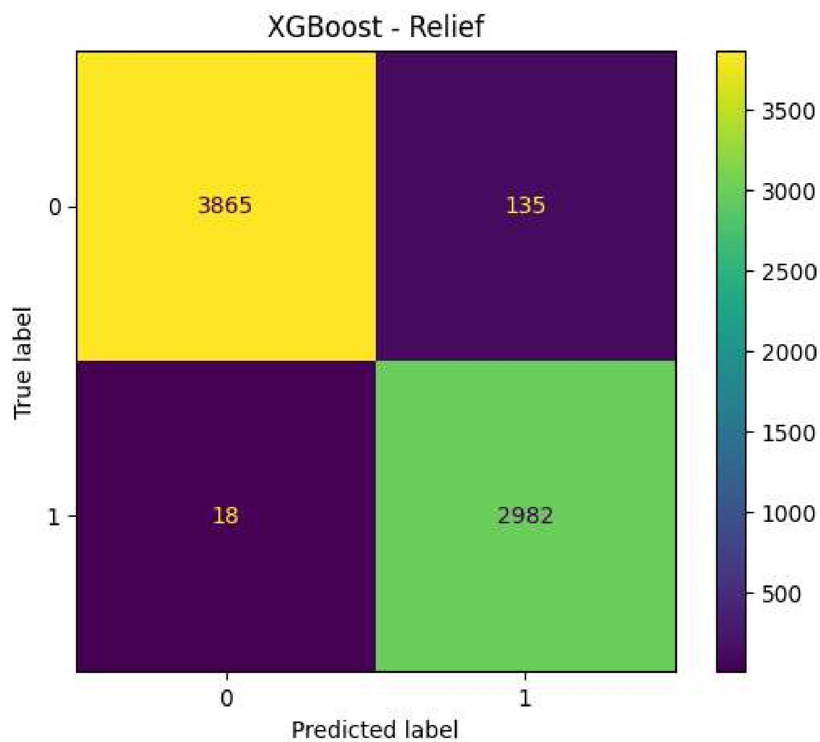


Figure 10.38. Confusion Matrix Obtained with XGBoost+Relief-F

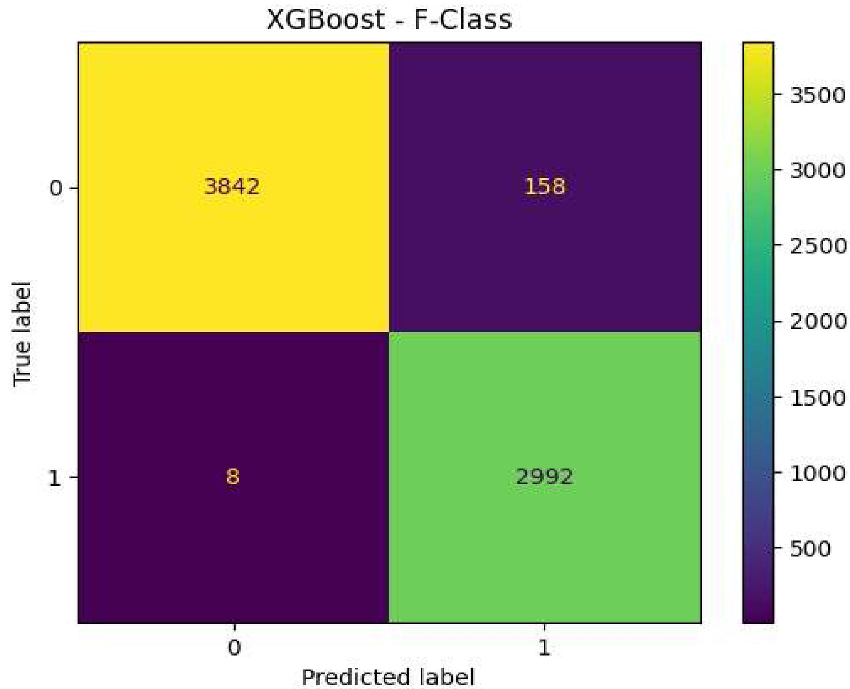


Figure 10.39. Confusion Matrix Obtained with XGBoost+F-Classification

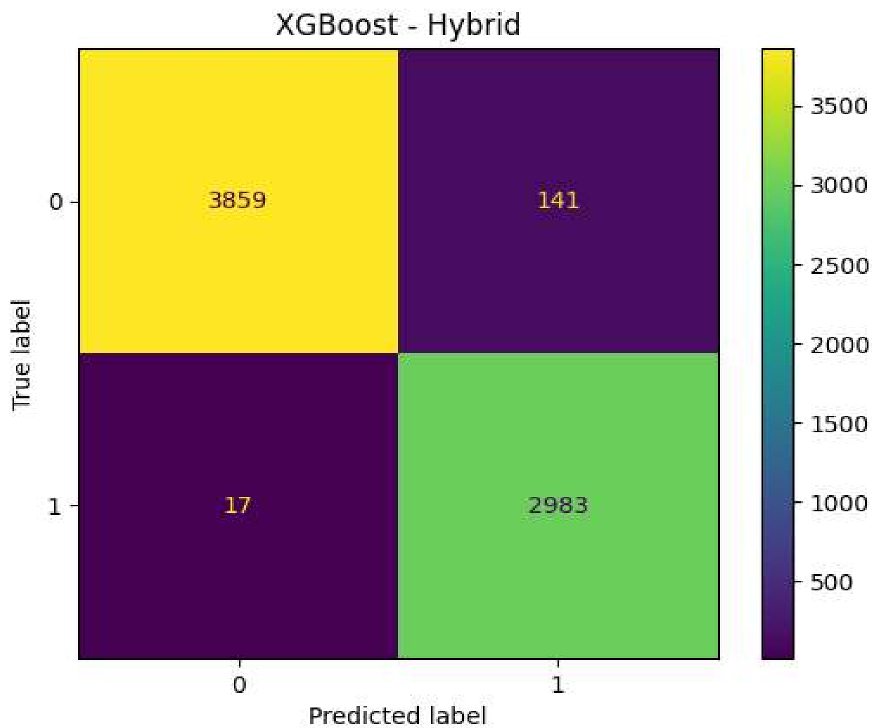


Figure 10.40. Confusion Matrix Obtained with XGBoost+Hybrid

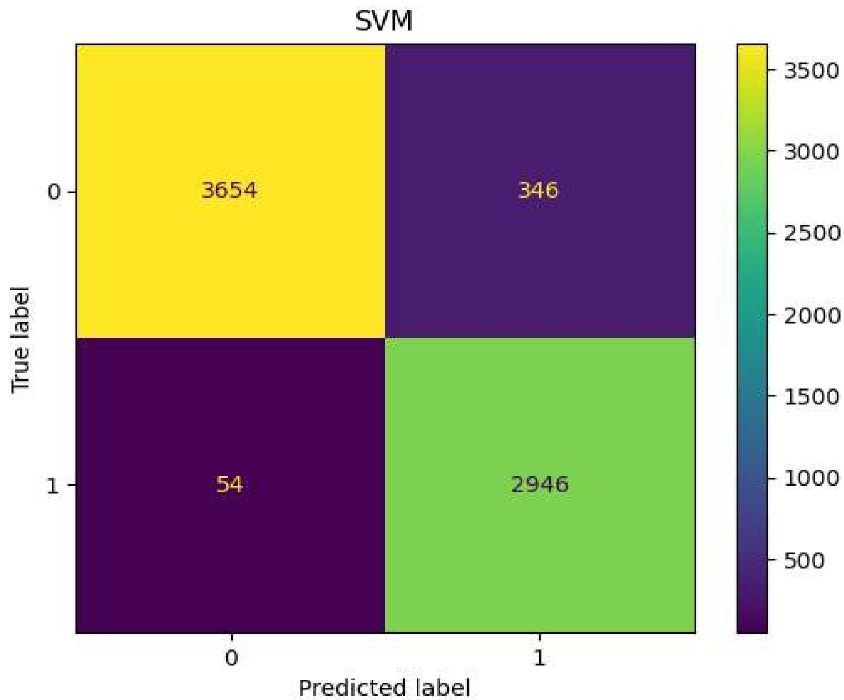


Figure 10.41. Confusion Matrix Obtained with SVM

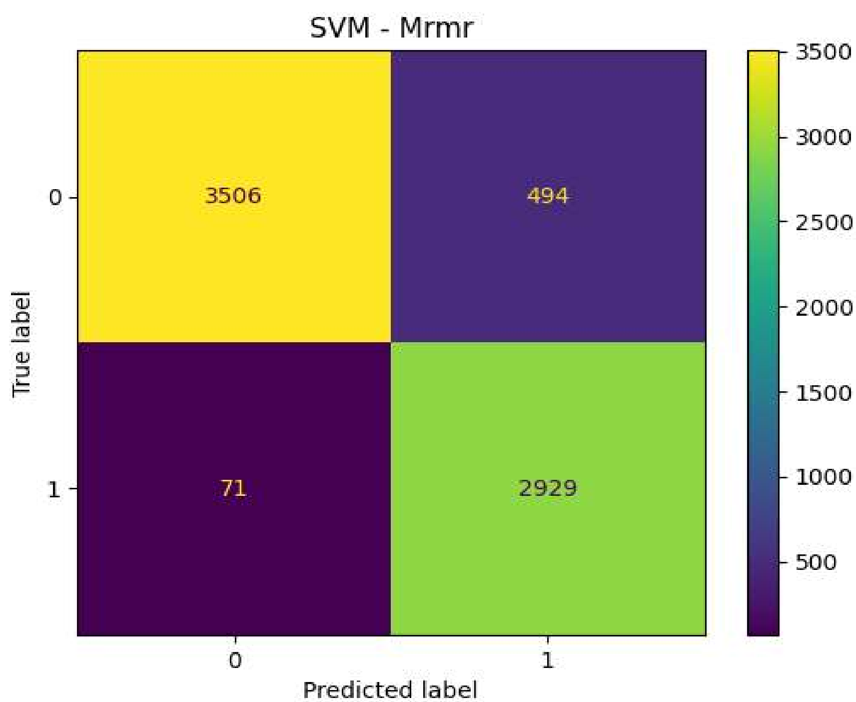


Figure 10.42. Confusion Matrix Obtained with SVM+mRMR

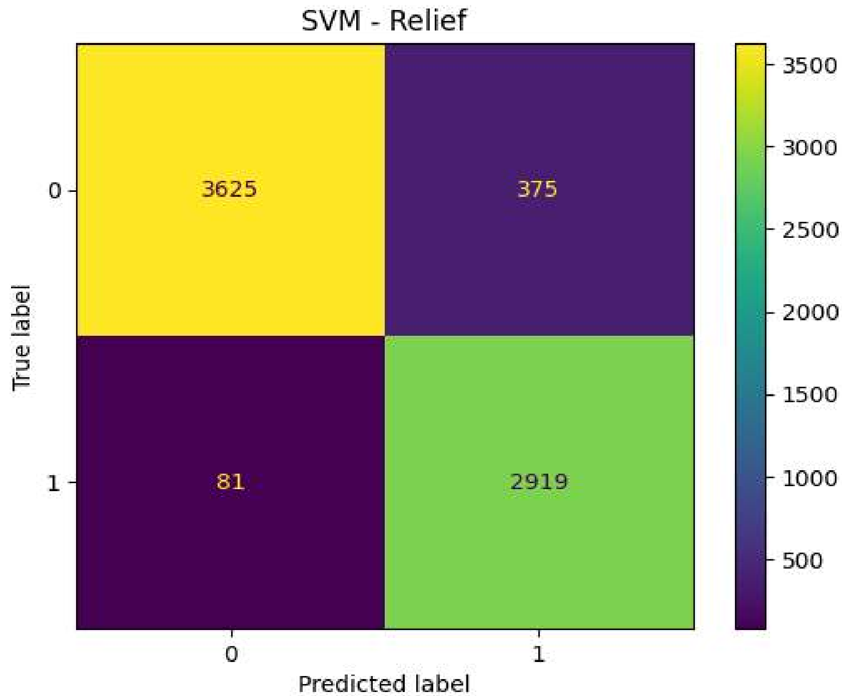


Figure 10.43. Confusion Matrix Obtained with SVM+Relief-F

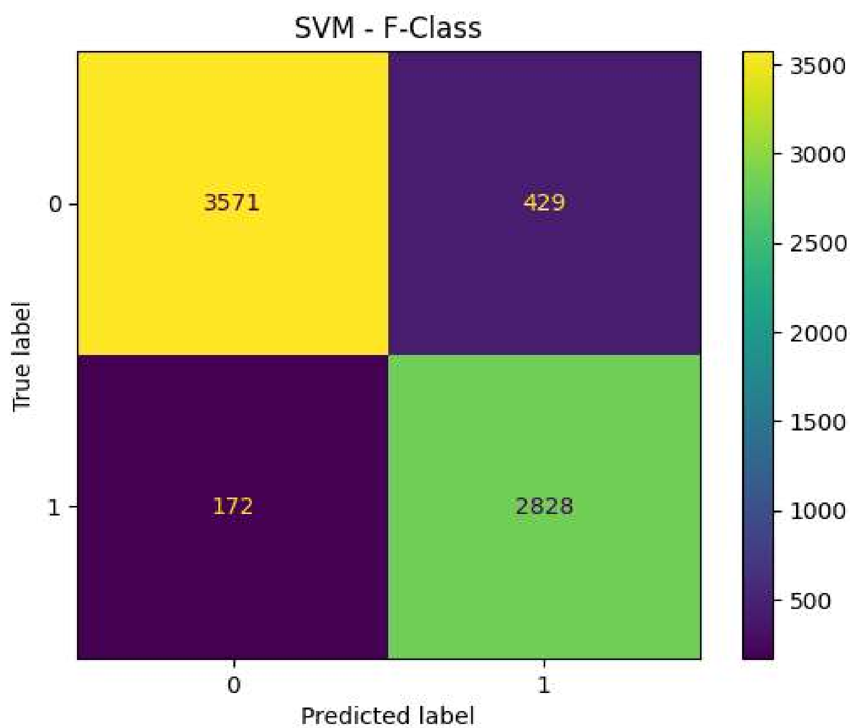


Figure 10.44. Confusion Matrix Obtained with SVM+F-Classification

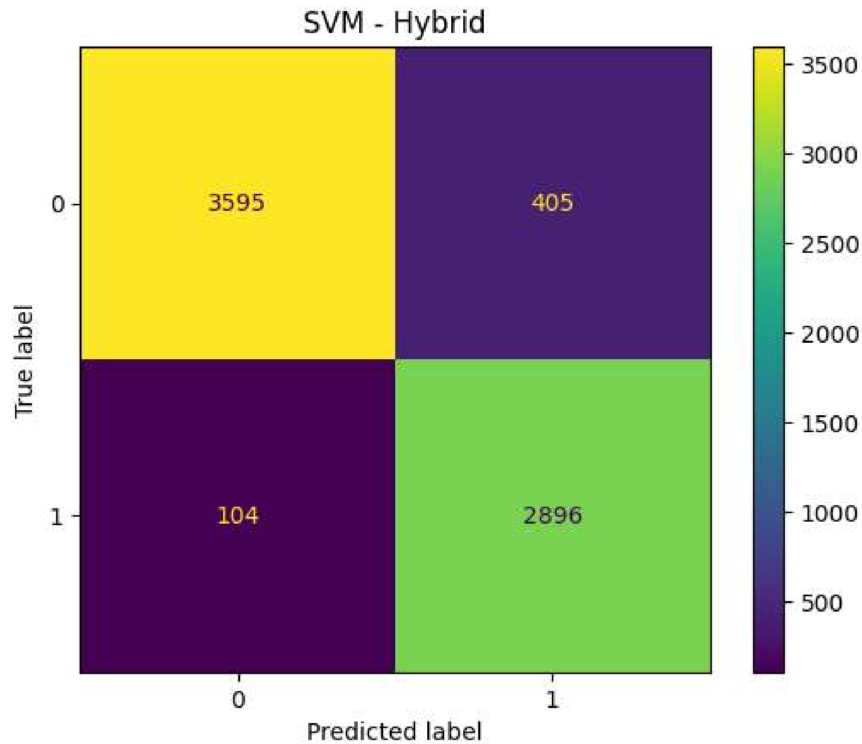


Figure 10.45. Confusion Matrix Obtained with SVM+Hybrid

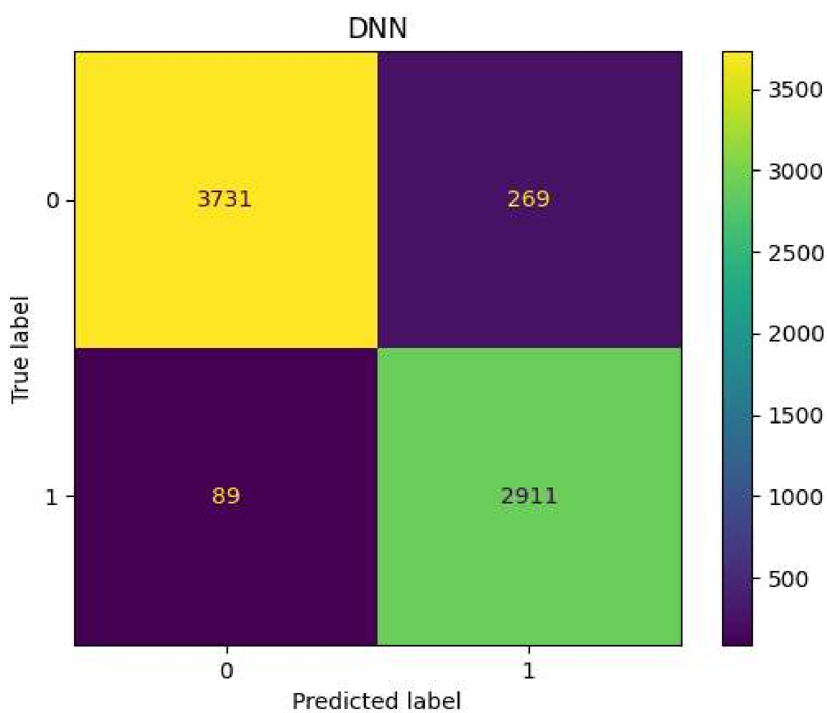


Figure 10.46. Confusion Matrix Obtained with DNN

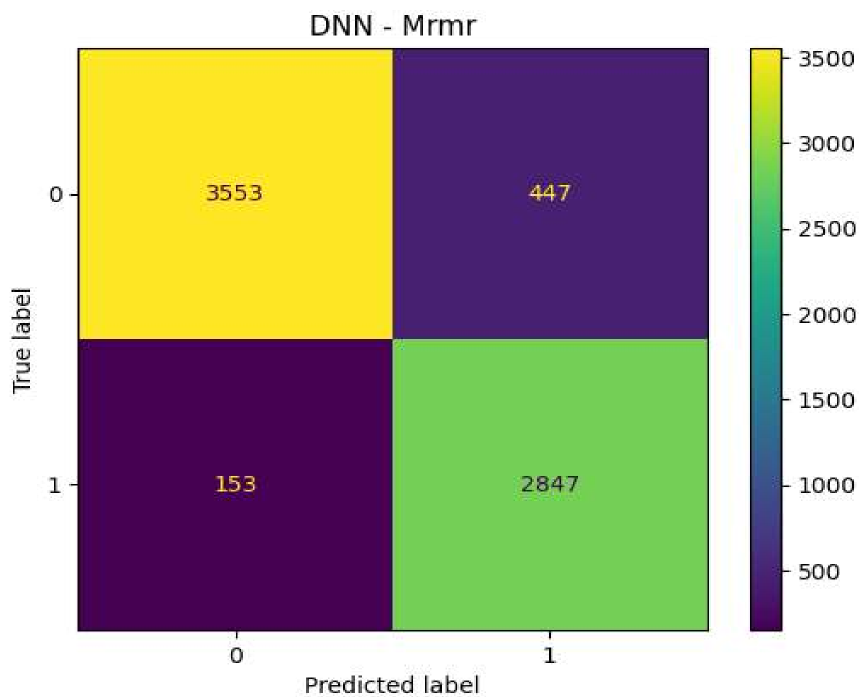


Figure 10.47. Confusion Matrix Obtained with DNN+mRMR

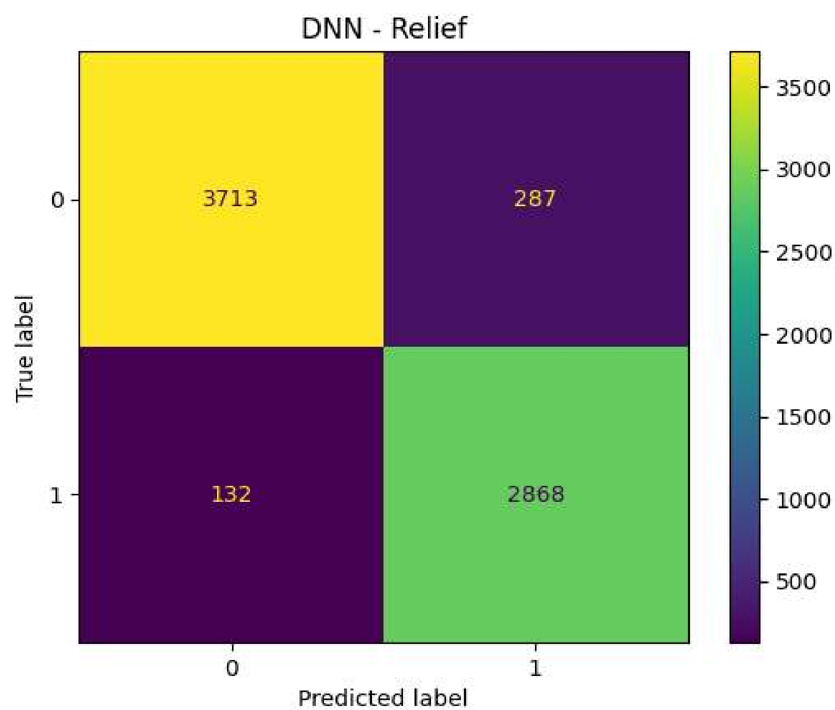


Figure 10.48. Confusion Matrix Obtained with DNN+Relief-F

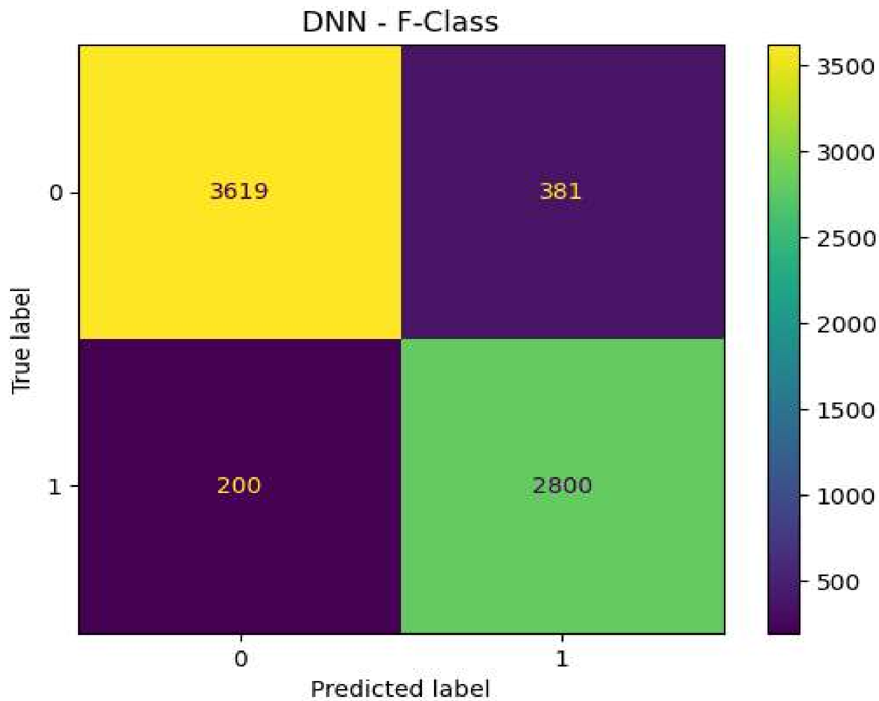


Figure 10.49. Confusion Matrix Obtained with DNN+F-Classification

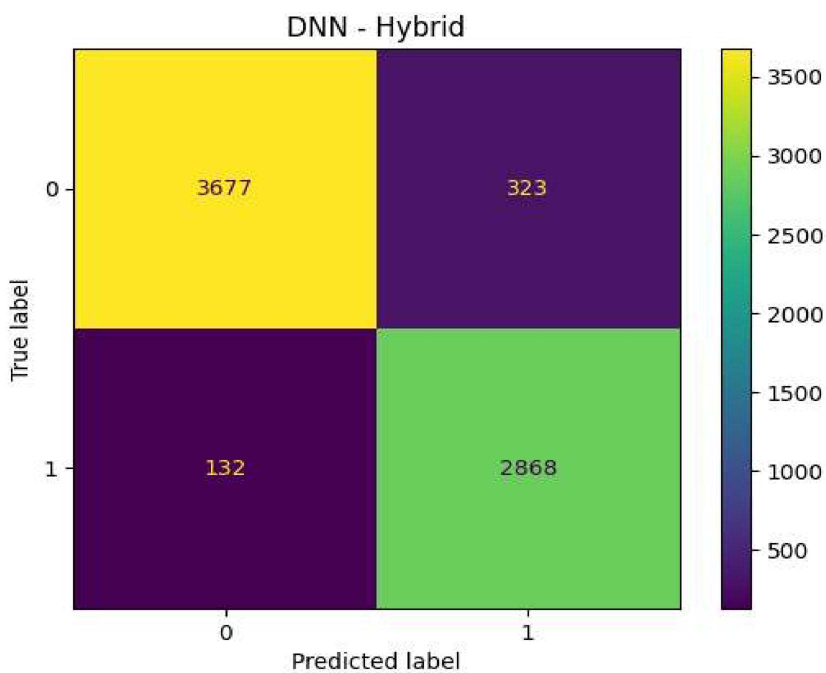


Figure 10.50. Confusion Matrix Obtained with DNN+Hybrid

10.1.3. Binary Classification Based Return Prediction Models

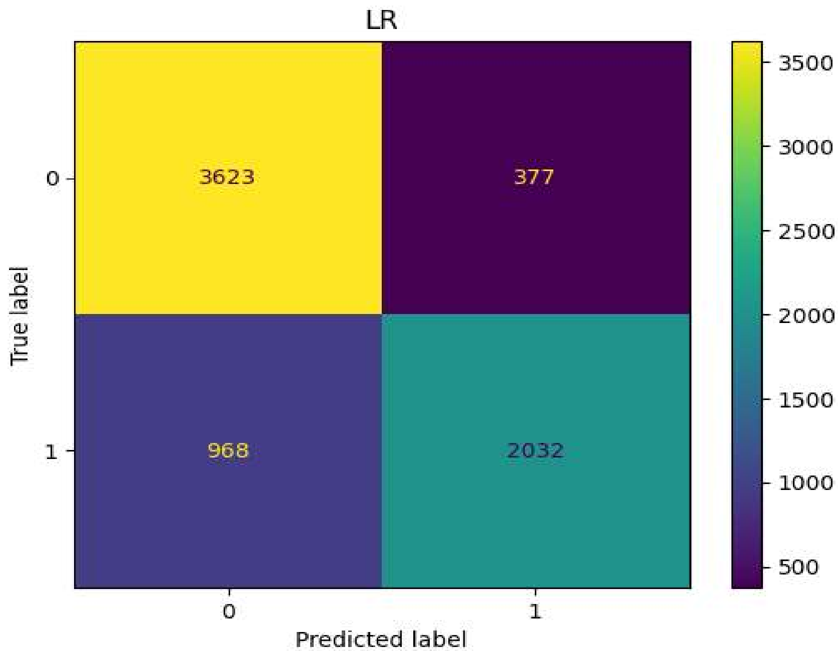


Figure 10.51. Confusion Matrix Obtained with LR

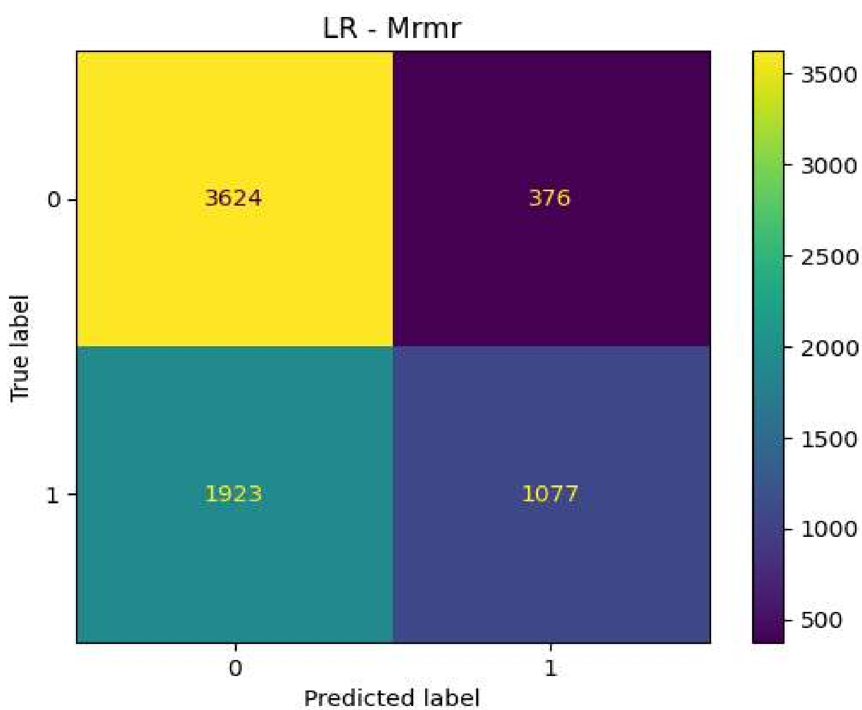


Figure 10.52. Confusion Matrix Obtained with LR+mRMR

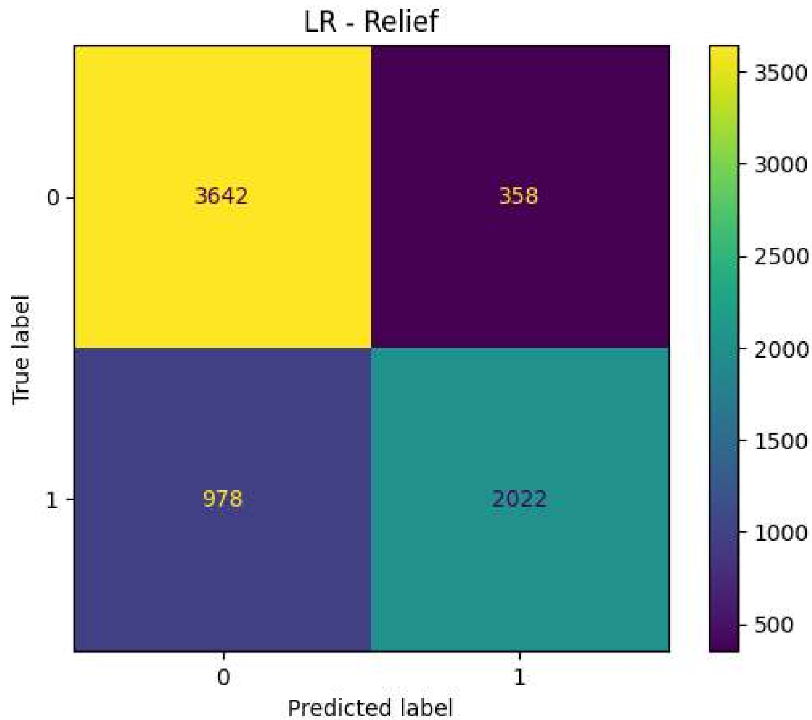


Figure 10.53. Confusion Matrix Obtained with LR+Relief-F

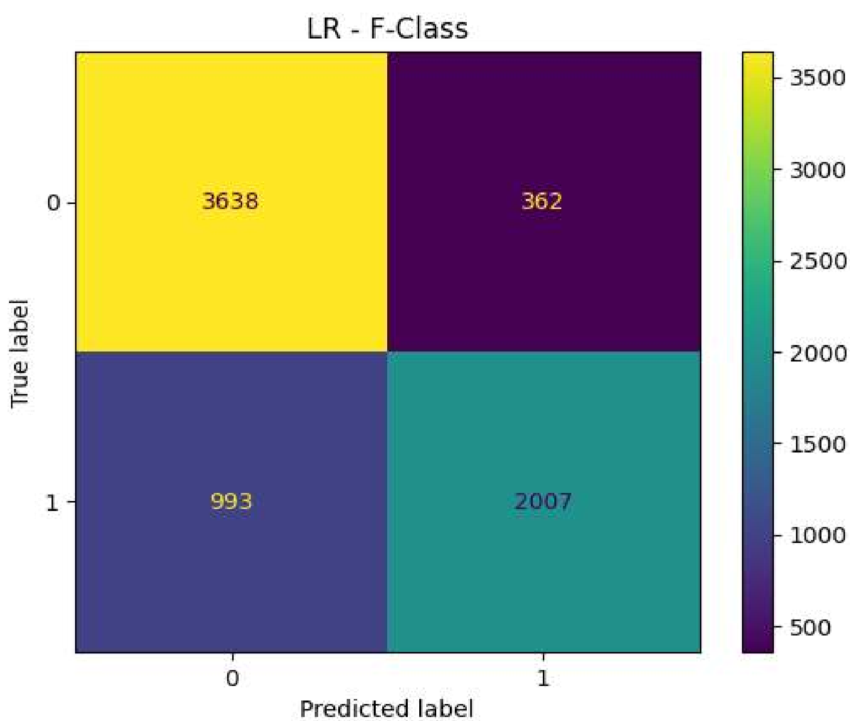


Figure 10.54. Confusion Matrix Obtained with LR+F-Classification

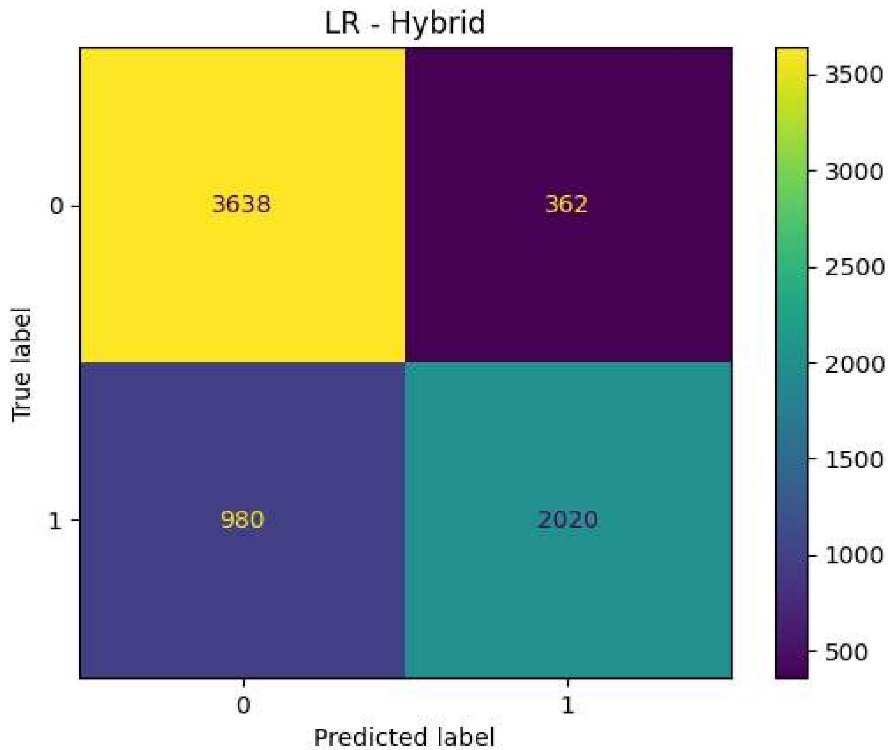


Figure 10.55. Confusion Matrix Obtained with LR+Hybrid

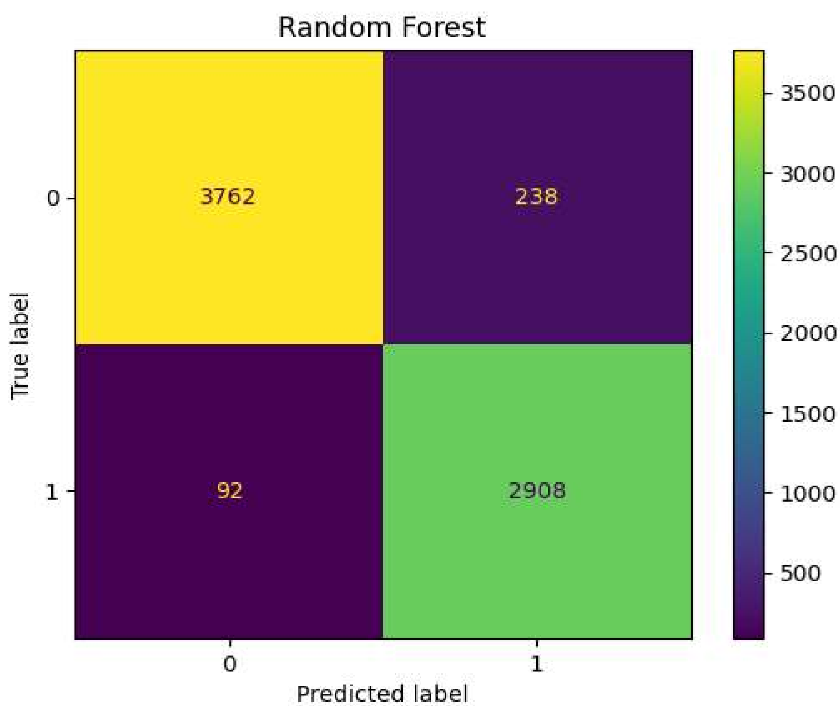


Figure 10.56. Confusion Matrix Obtained with RF

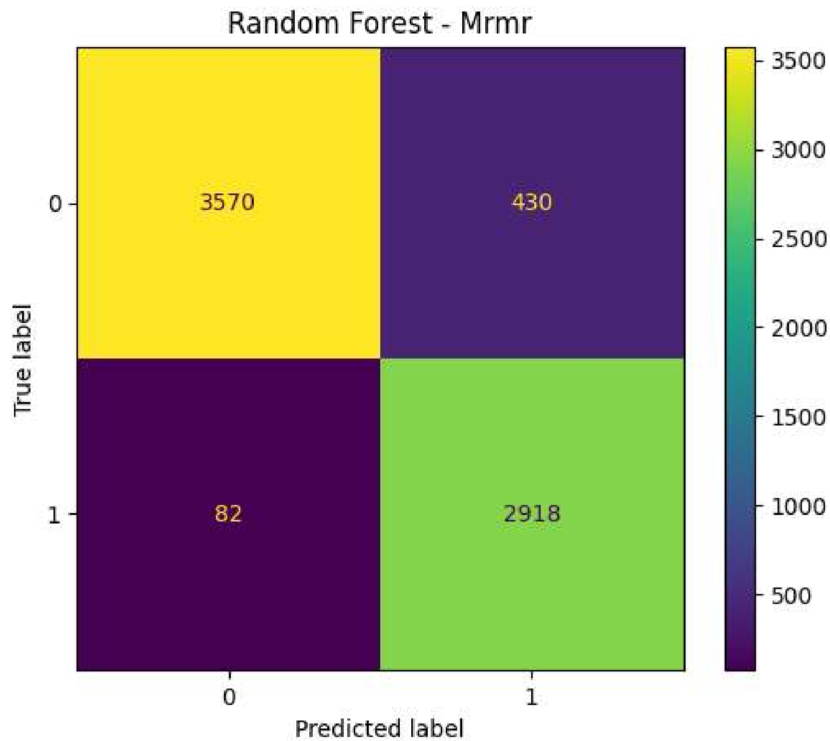


Figure 10.57. Confusion Matrix Obtained with RF+mRMR

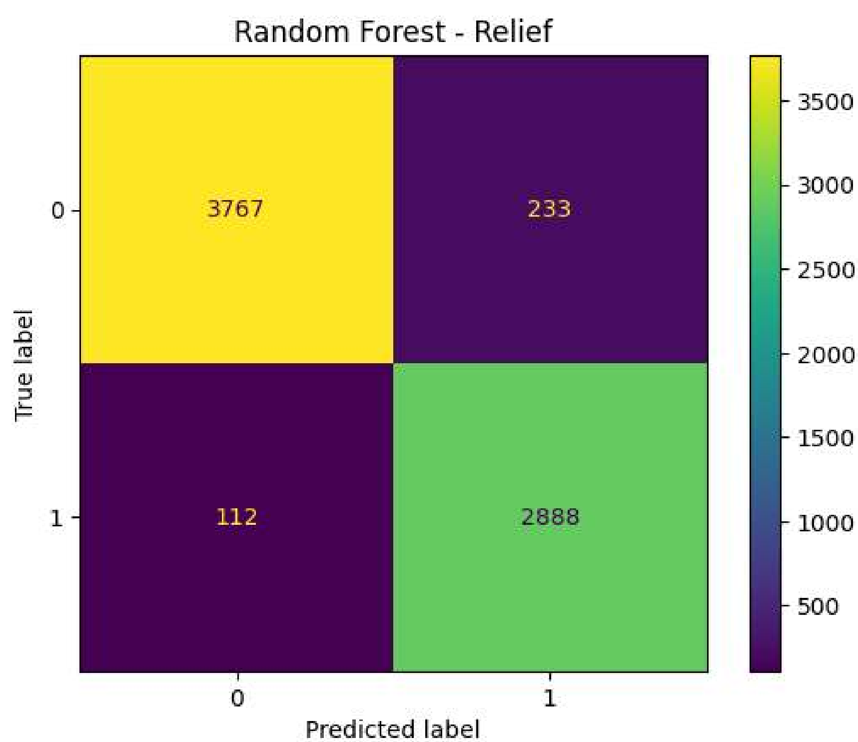


Figure 10.58. Confusion Matrix Obtained with RF+Relief-F

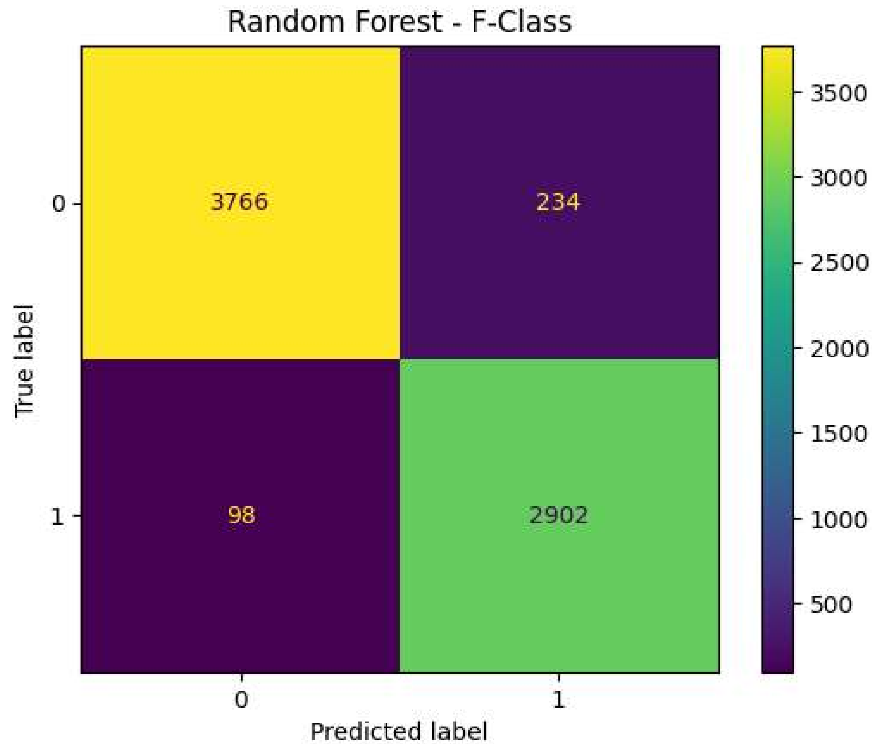


Figure 10.59. Confusion Matrix Obtained with RF+F-Classification

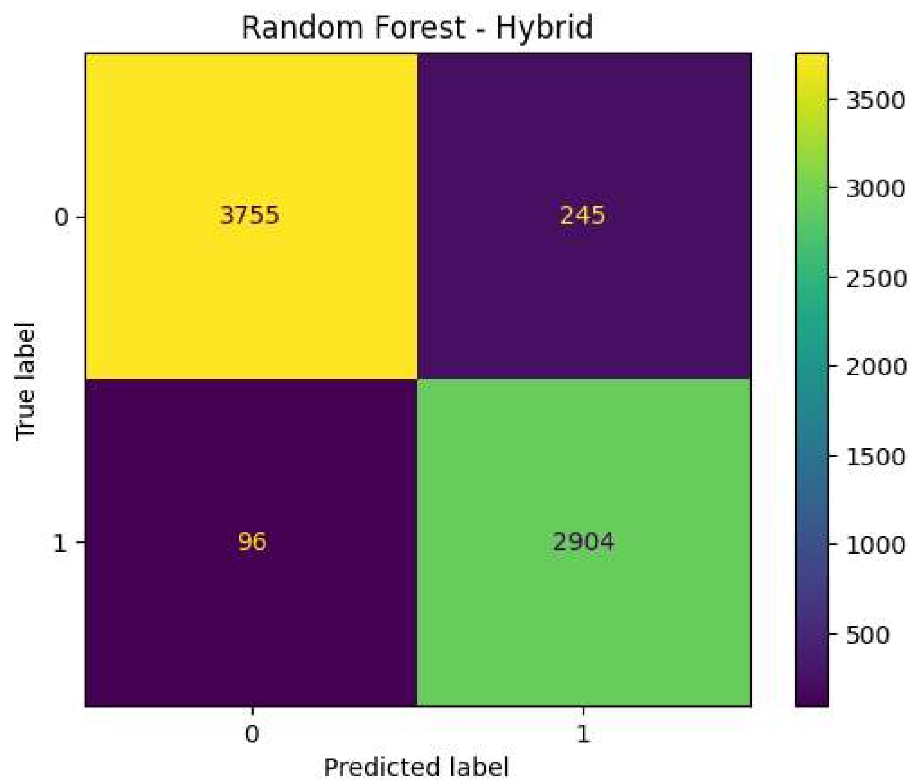


Figure 10.60. Confusion Matrix Obtained with RF+Hybrid

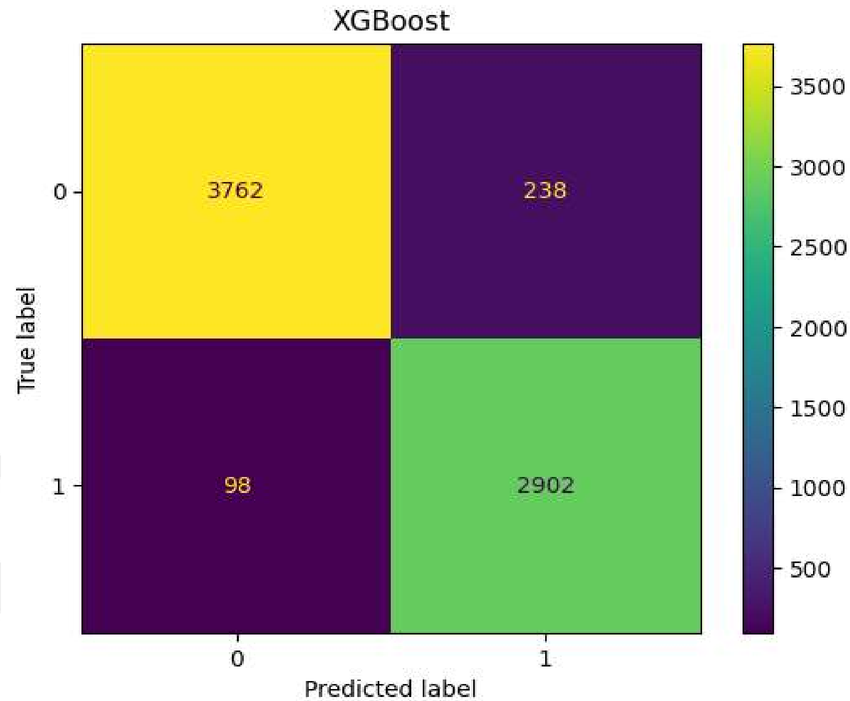


Figure 10.61. Confusion Matrix Obtained with XGBoost

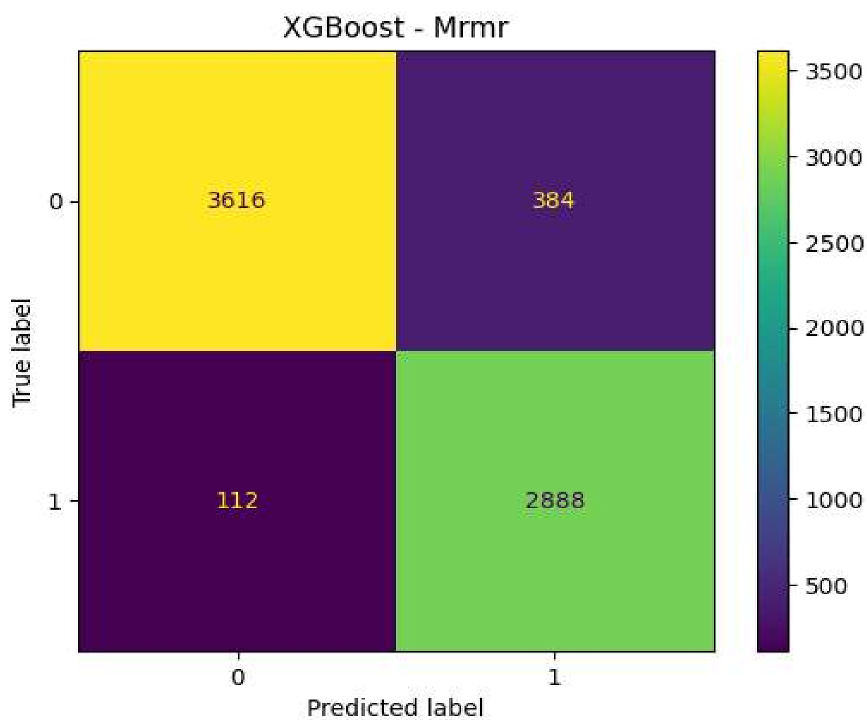


Figure 10.62. Confusion Matrix Obtained with XGBoost+mRMR

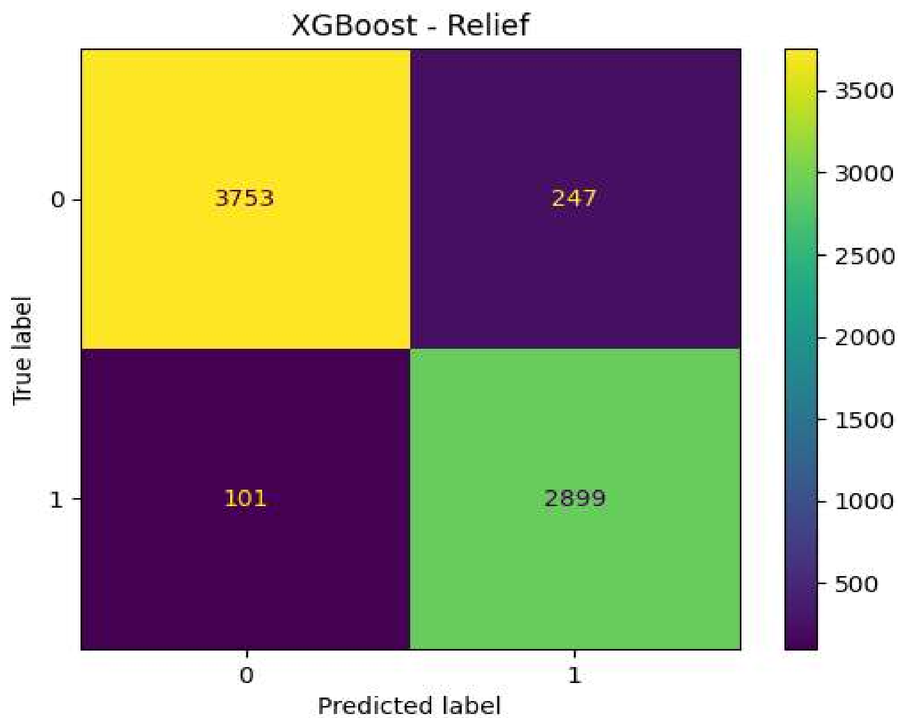


Figure 10.63. Confusion Matrix Obtained with XGBoost+Relief-F

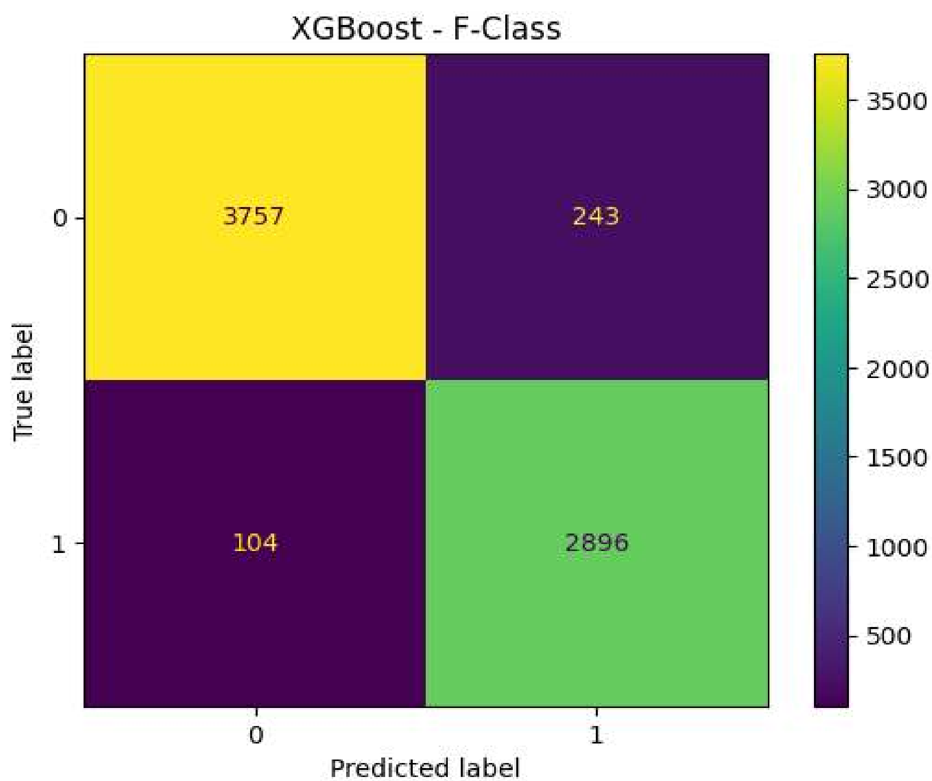


Figure 10.64. Confusion Matrix Obtained with XGBoost+F-Classification

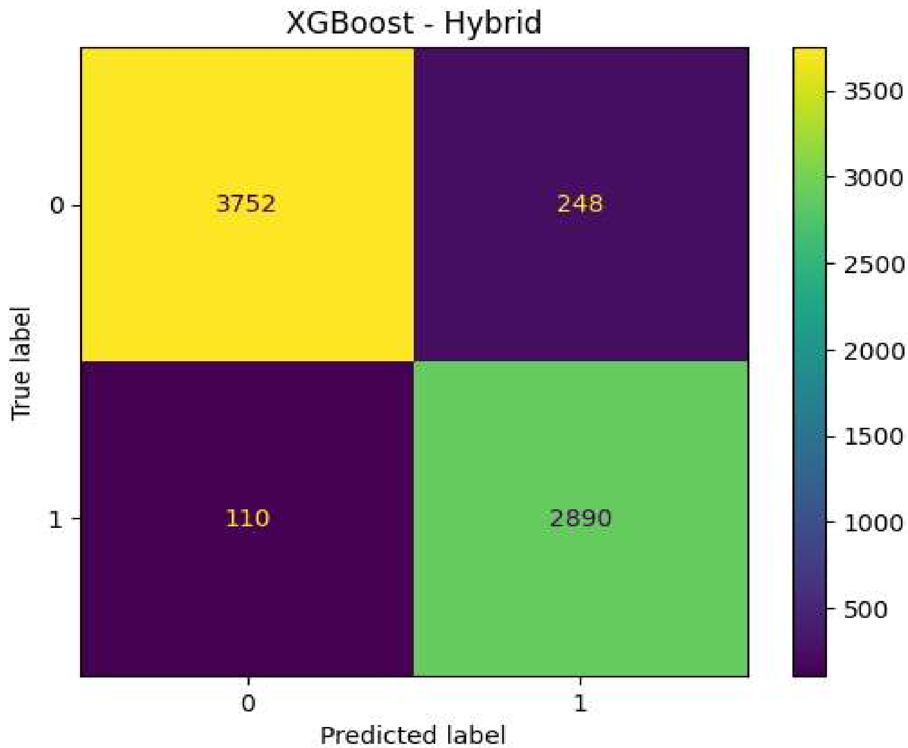


Figure 10.65. Confusion Matrix Obtained with XGBoost+Hybrid

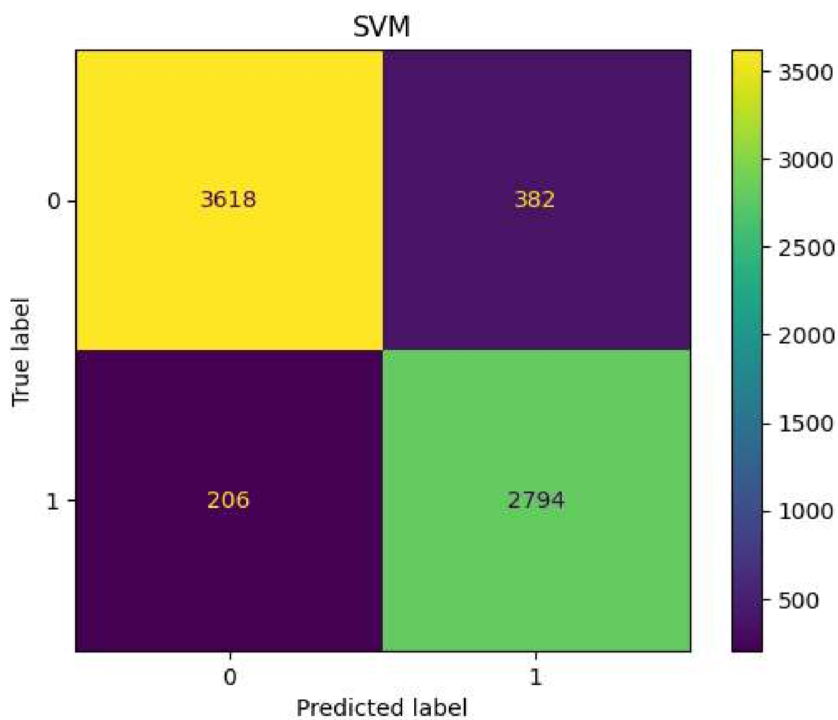


Figure 10.66. Confusion Matrix Obtained with SVM

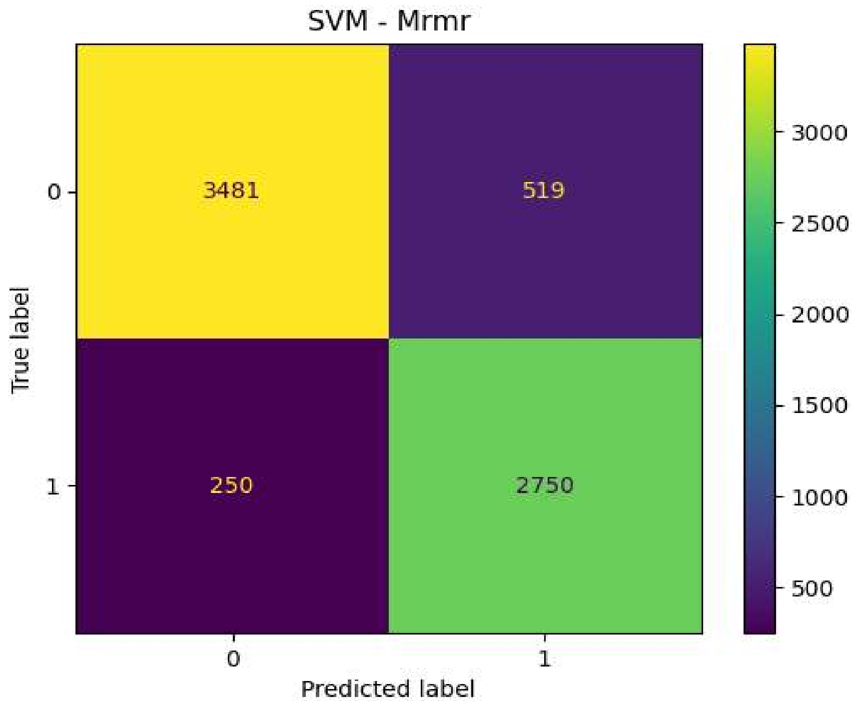


Figure 10.67 Confusion Matrix Obtained with SVM+mRMR

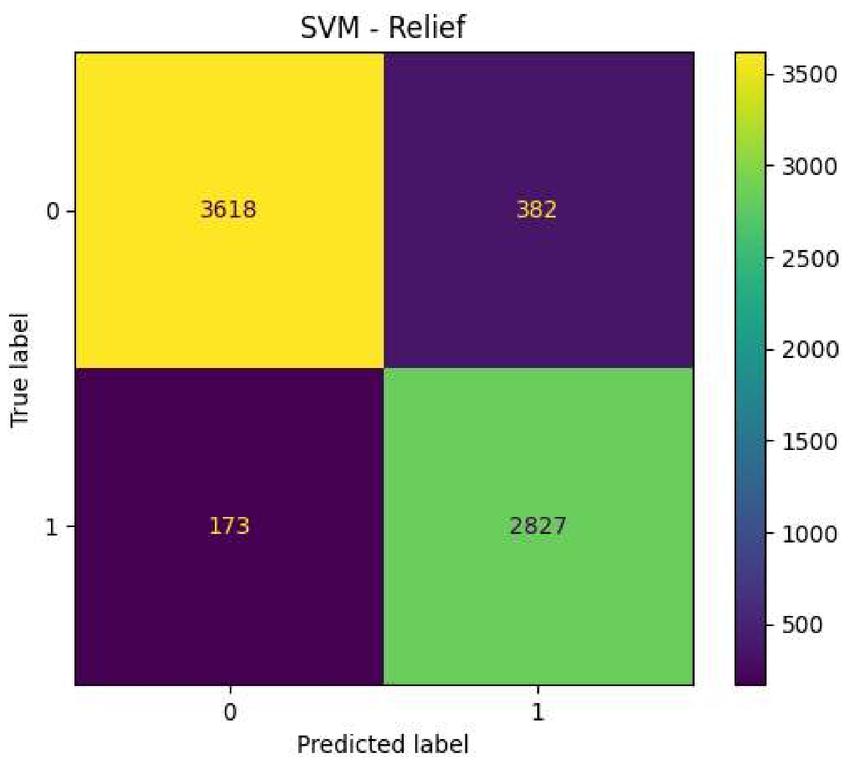


Figure 10.68. Confusion Matrix Obtained with SVM+Relief-F

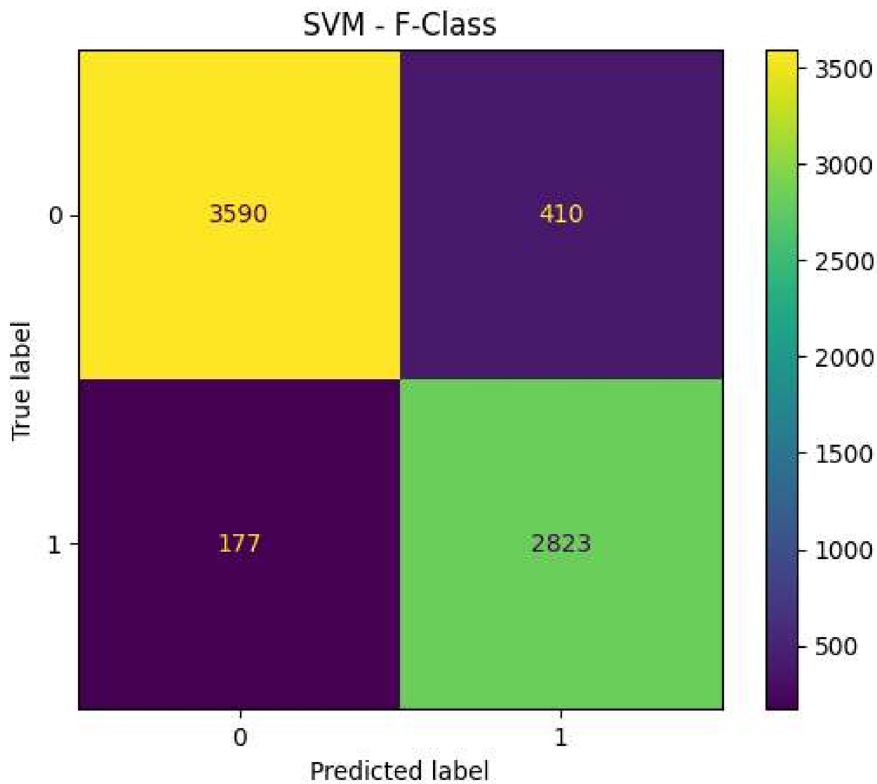


Figure 10.69. Confusion Matrix Obtained with SVM+Relief-F

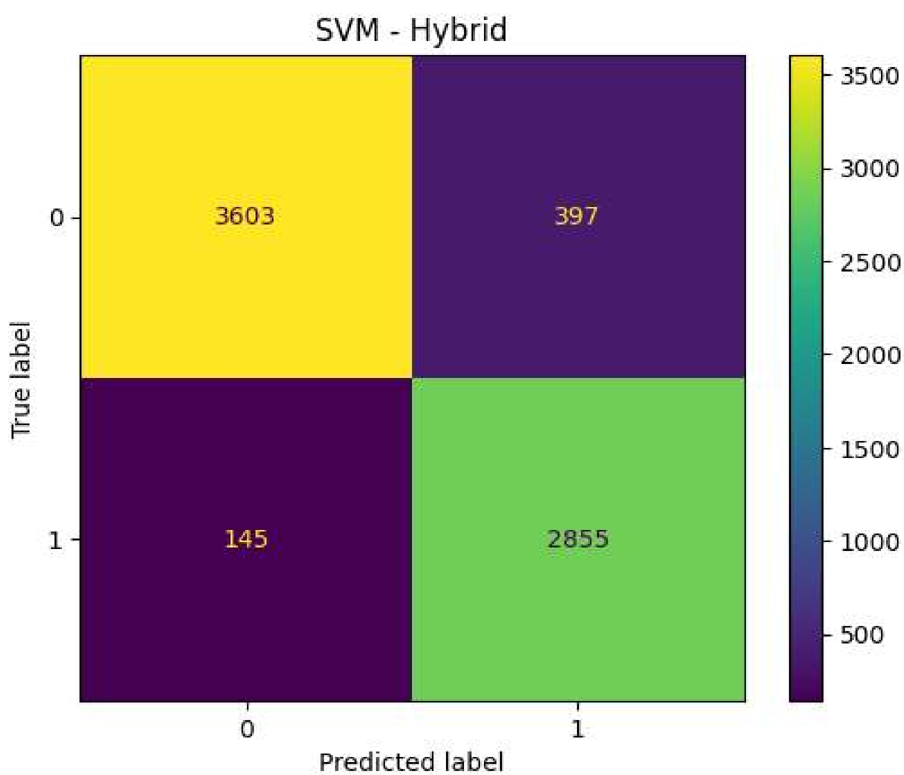


Figure 10.70. Confusion Matrix Obtained with SVM+Hybrid

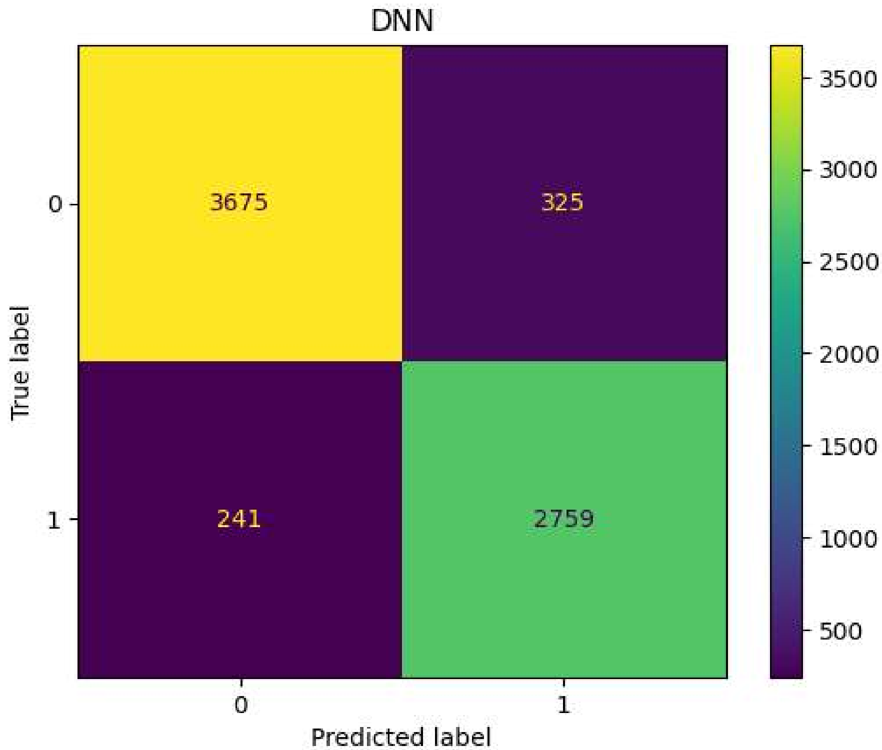


Figure 10.71. Confusion Matrix Obtained with DNN

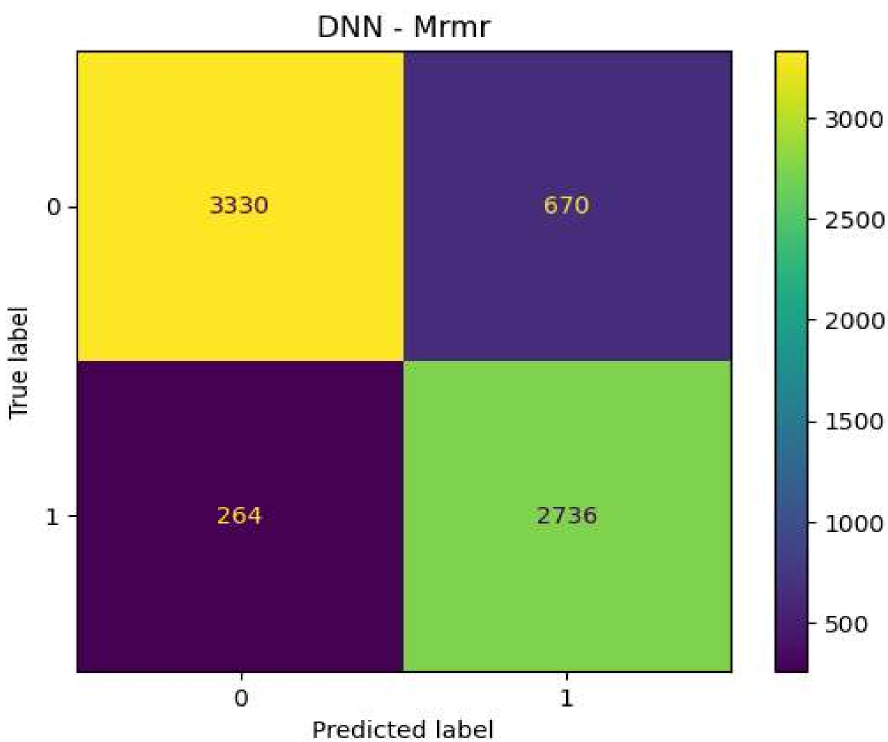


Figure 10.72. Confusion Matrix Obtained with DNN+mRMR

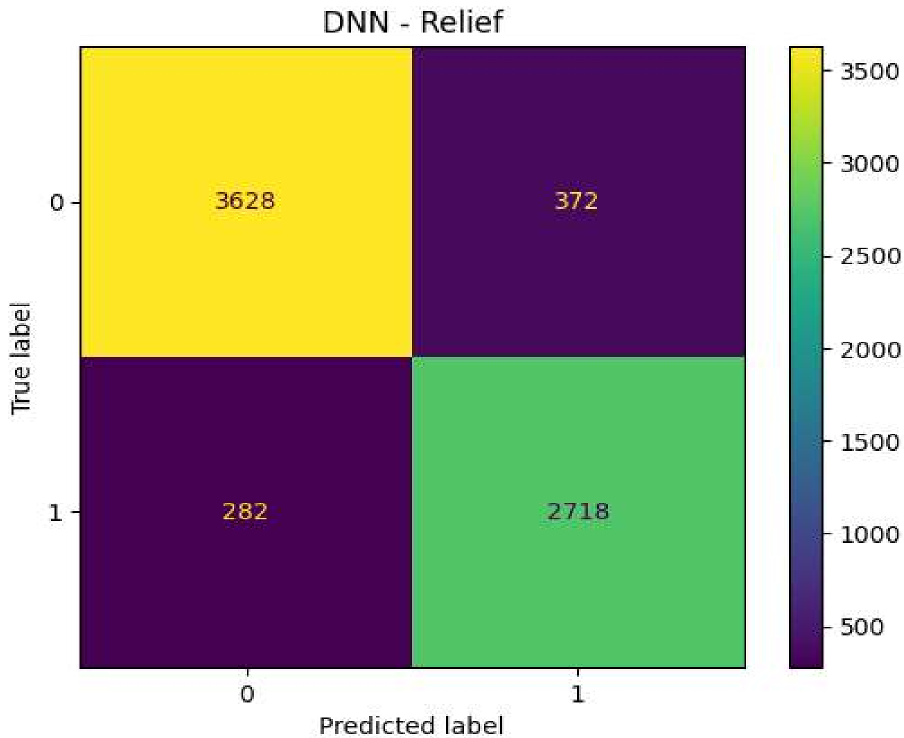


Figure 10.73. Confusion Matrix Obtained with DNN+Relief-F

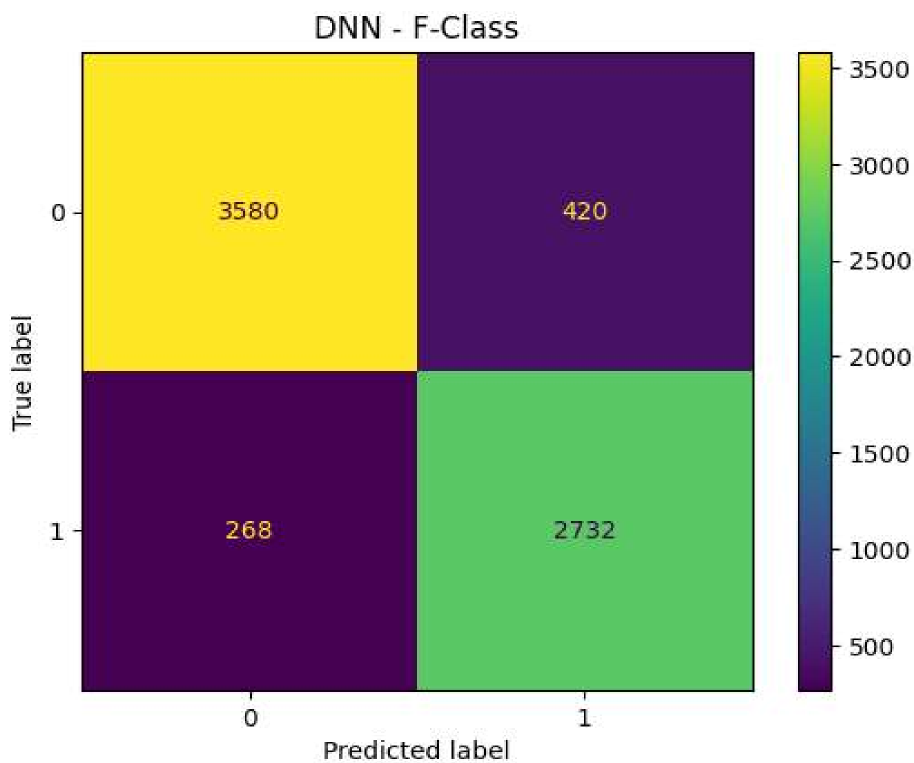


Figure 10.74. Confusion Matrix Obtained with DNN+F-Classification

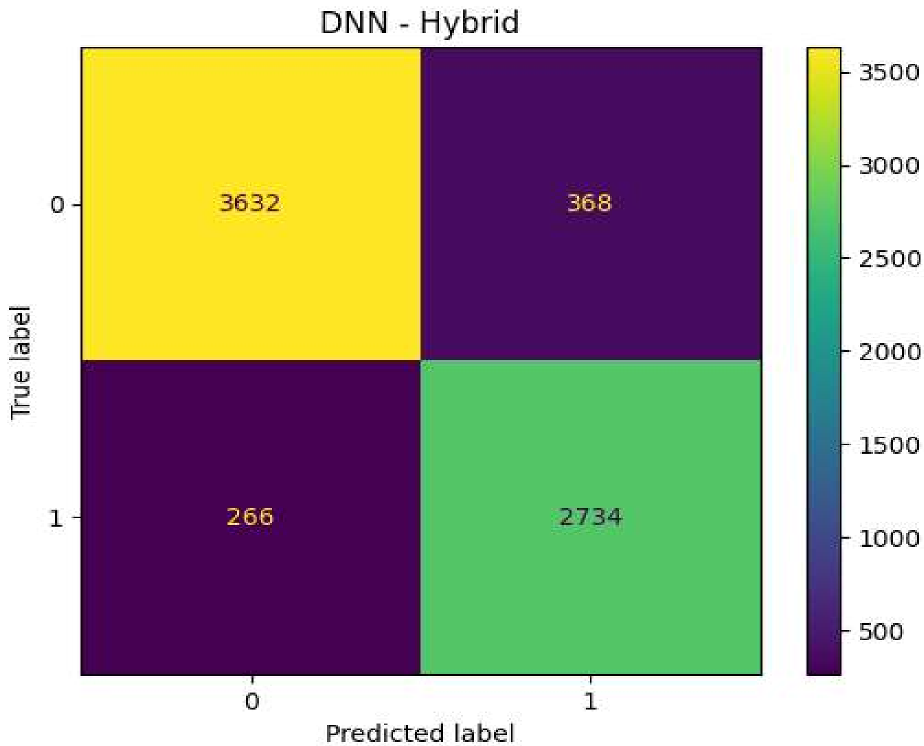


Figure 10.75. Confusion Matrix Obtained with DNN+Hybrid

10.2. Second Experimental Study: In-Fold Feature Selection during Cross-Validation

10.2.1. Multi-Class Classification Based Cancellation-Return Prediction Models

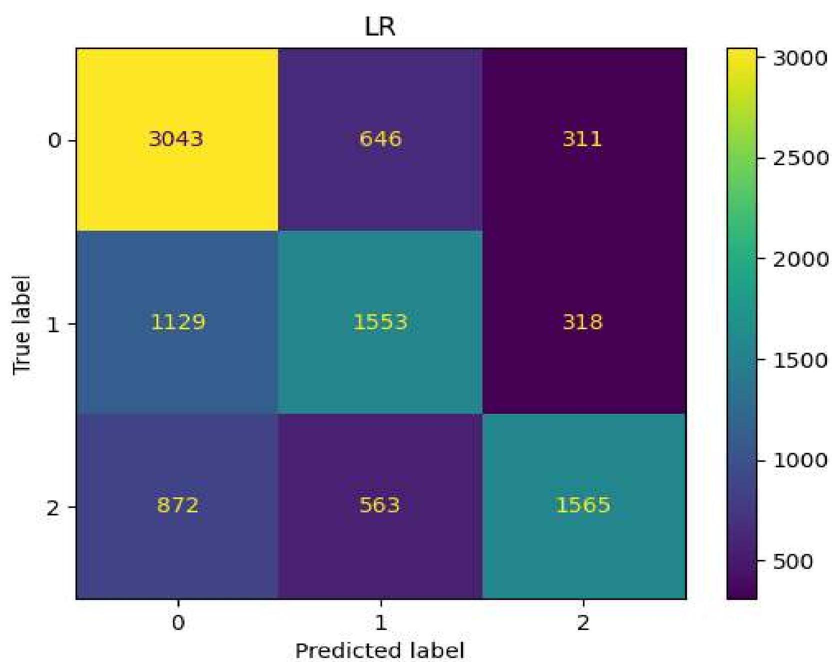


Figure 10.76. Confusion Matrix Obtained with LR

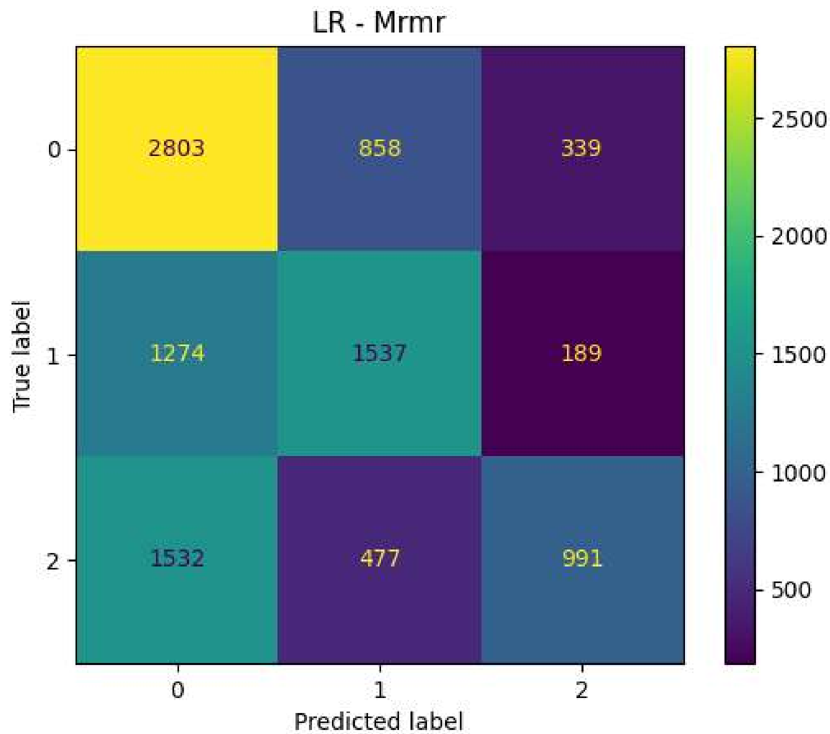


Figure 10.77. Confusion Matrix Obtained with LR+mRMR

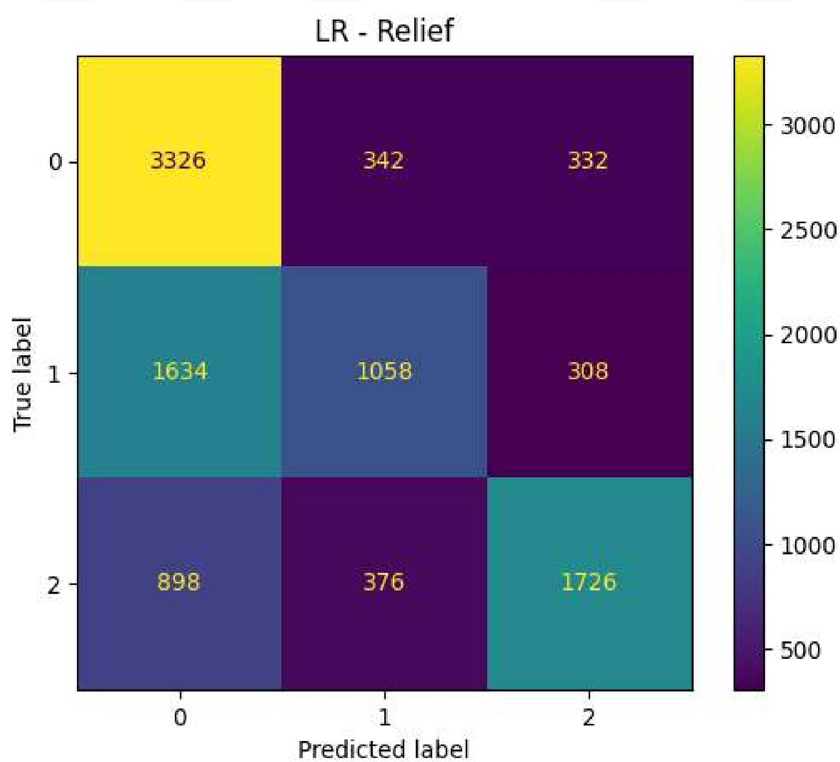


Figure 10.78. Confusion Matrix Obtained with LR+Relief-F

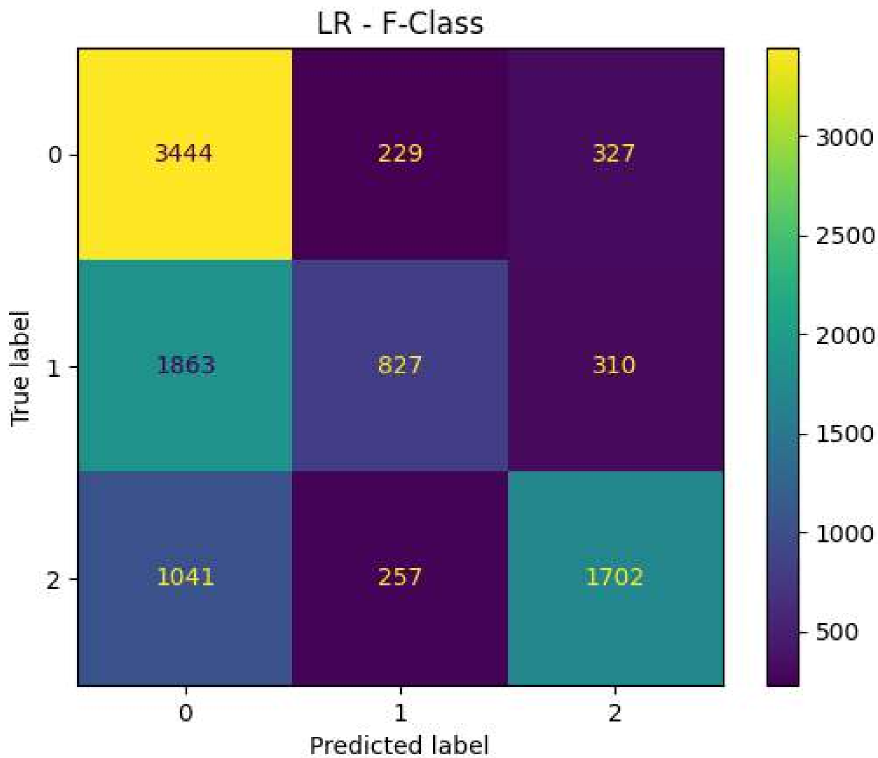


Figure 10.79. Confusion Matrix Obtained with LR+F-Classification

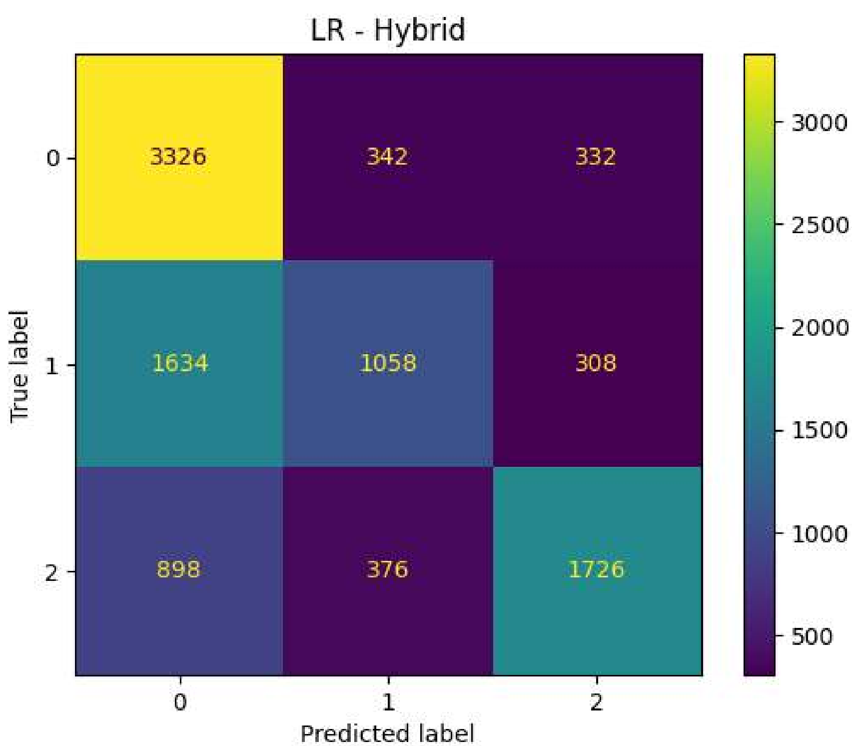


Figure 10.80. Confusion Matrix Obtained with LR+Hybrid

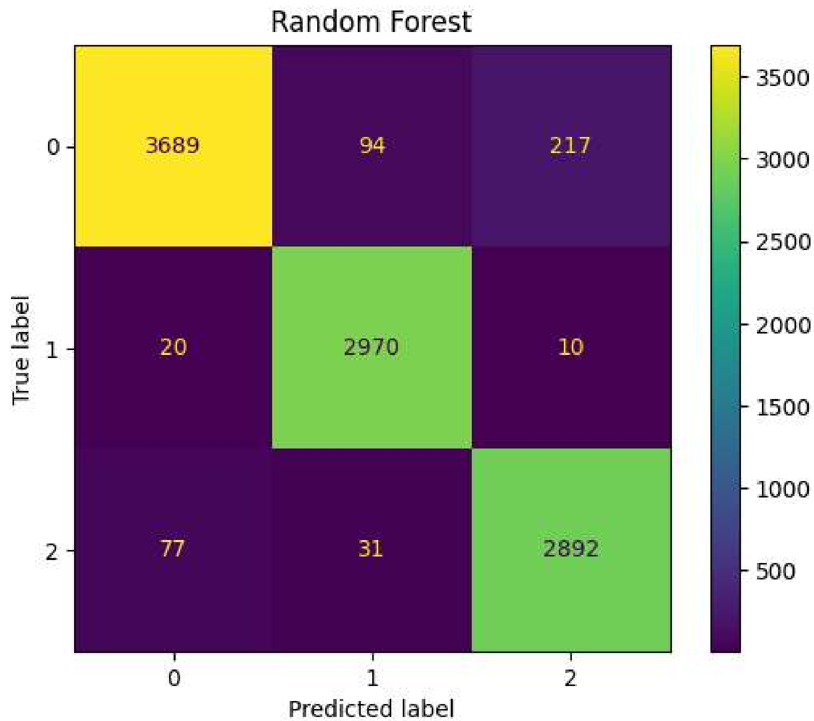


Figure 10.81. Confusion Matrix Obtained with RF

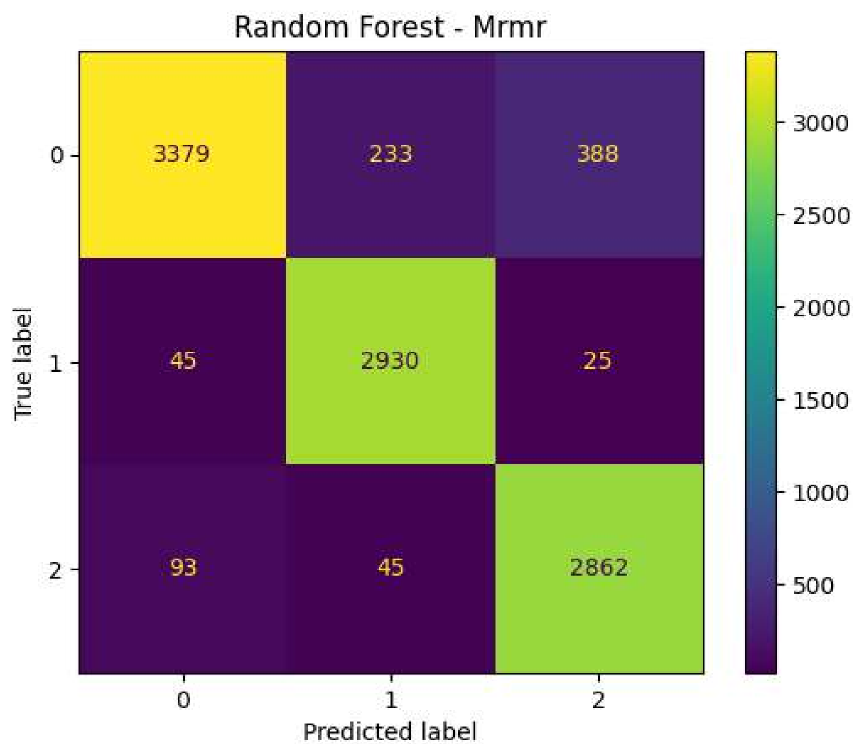


Figure 10.82. Confusion Matrix Obtained with RF+mRMR

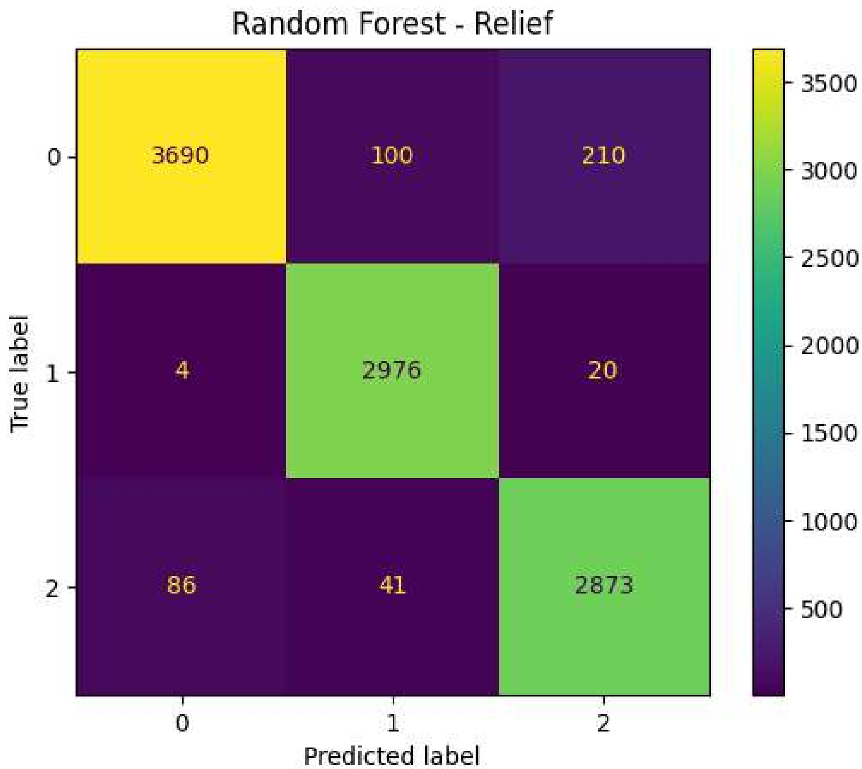


Figure 10.83. Confusion Matrix Obtained with RF+Relief-F

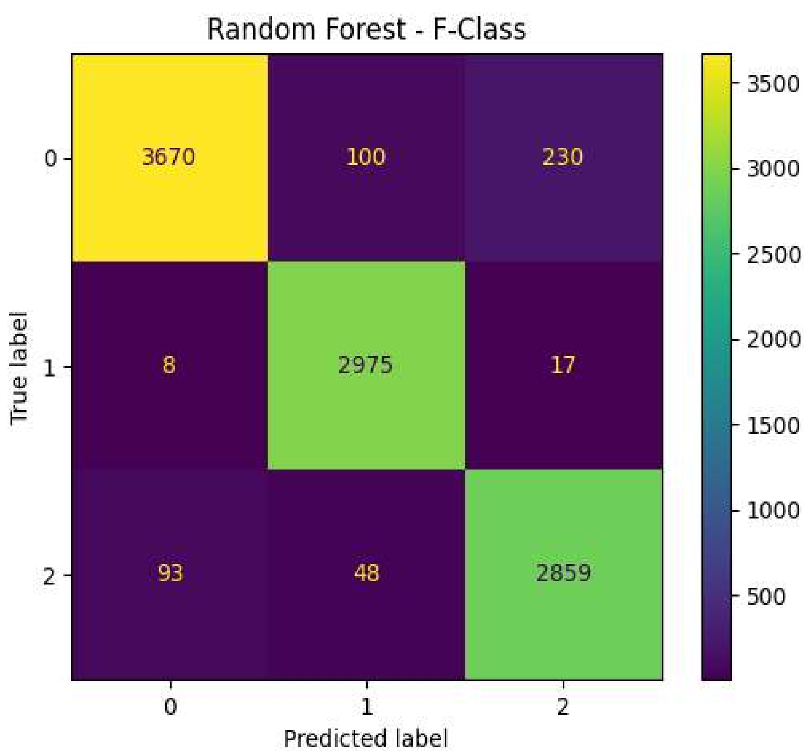


Figure 10.84. Confusion Matrix Obtained with RF+F-Classification

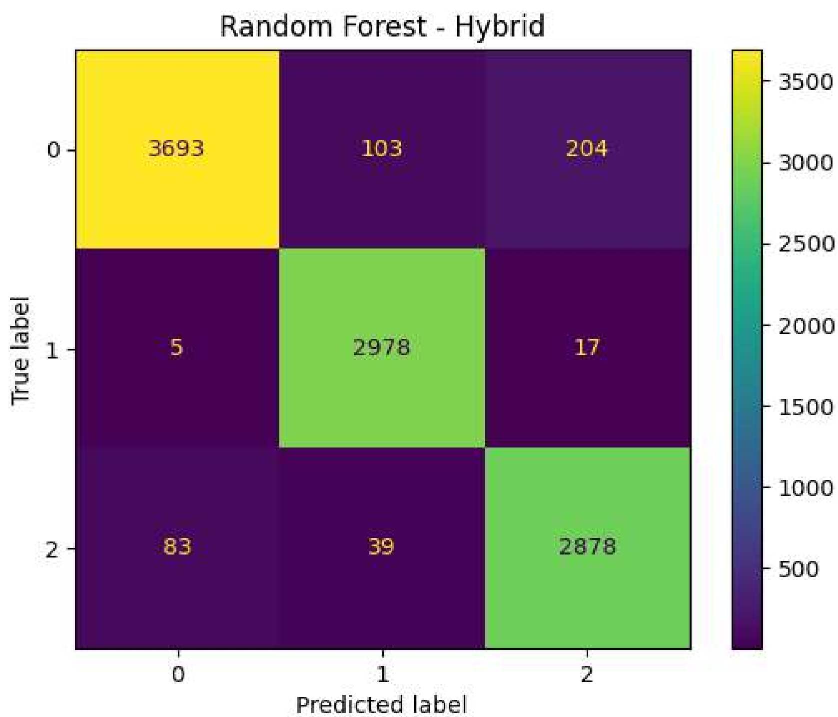


Figure 10.85. Confusion Matrix Obtained with RF+Hybrid

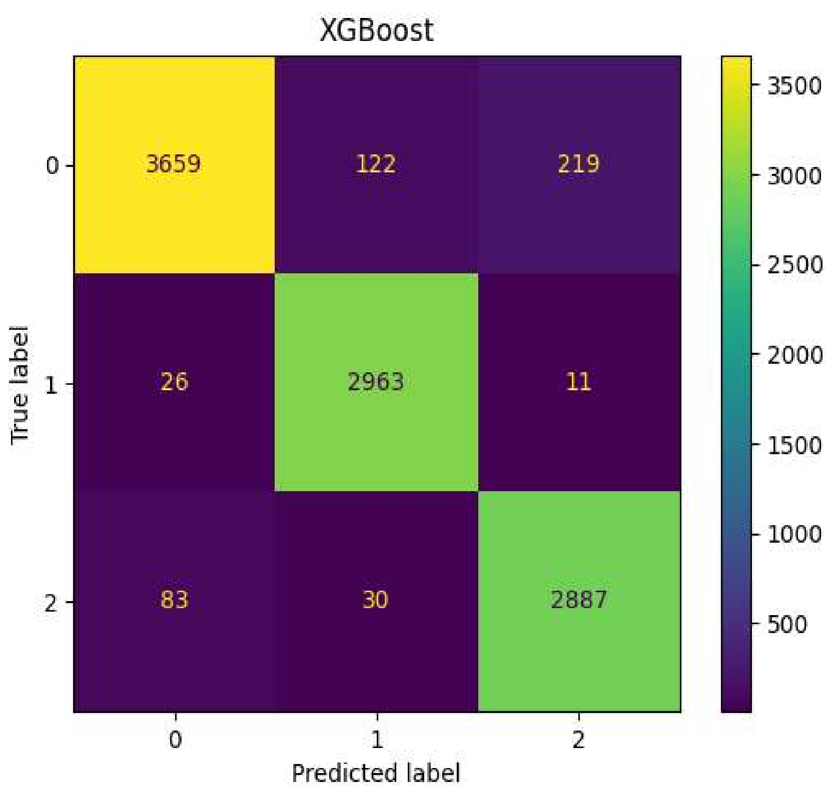


Figure 10.86. Confusion Matrix Obtained with XGBoost

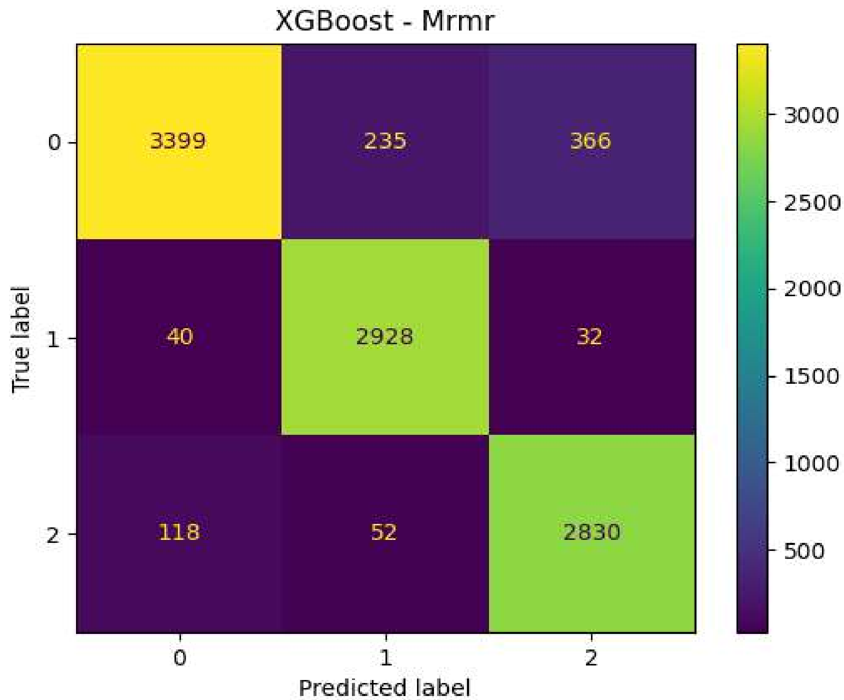


Figure 10.87. Confusion Matrix Obtained with XGBoost+mRMR

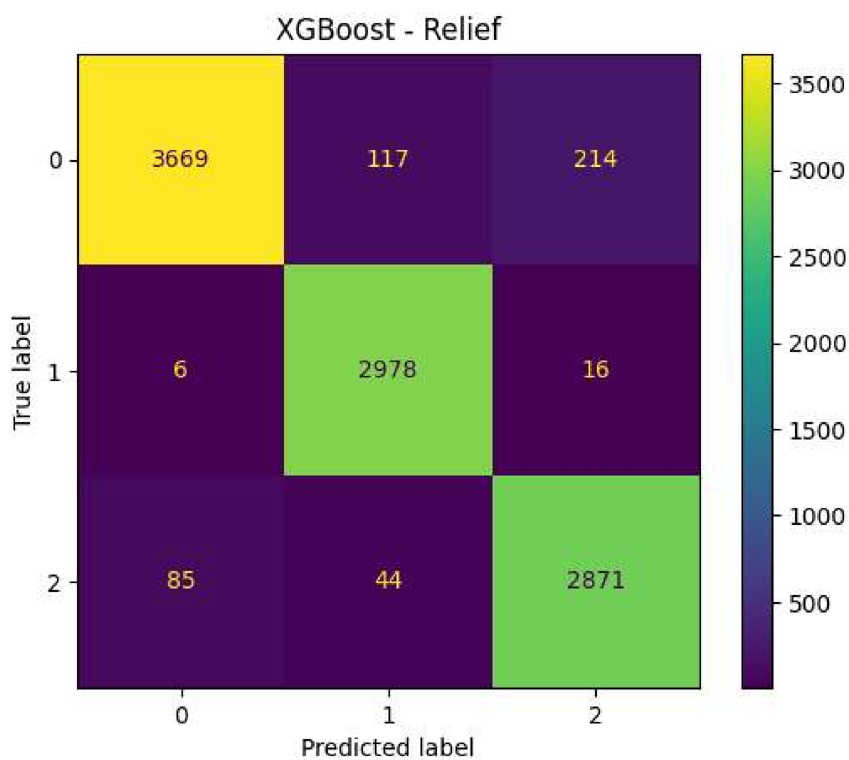


Figure 10.88. Confusion Matrix Obtained with XGBoost+Relief-F

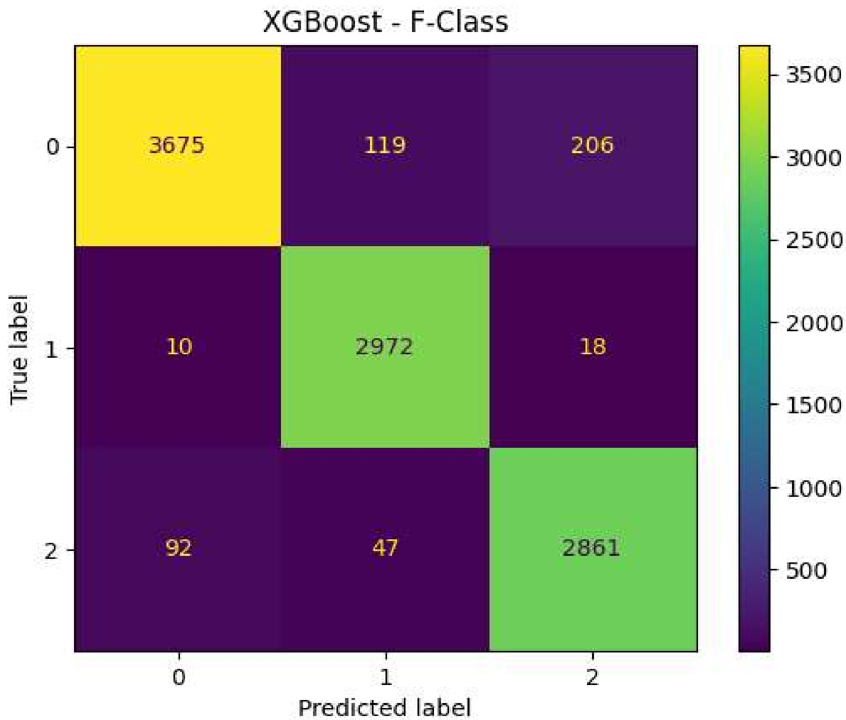


Figure 10.89. Confusion Matrix Obtained with XGBoost+F-Classification

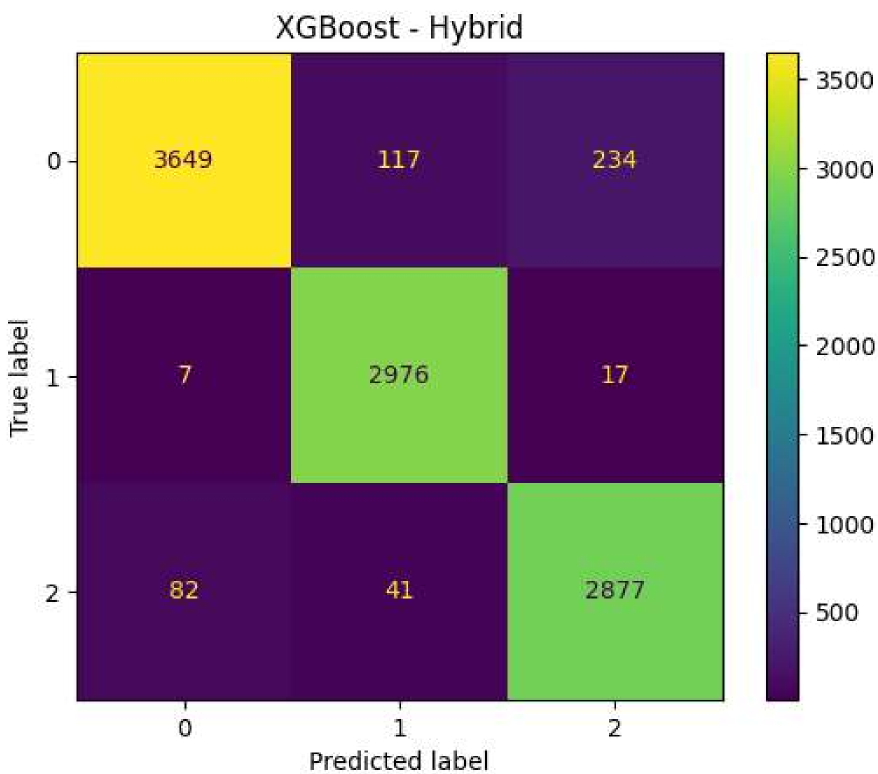


Figure 10.90. Confusion Matrix Obtained with XGBoost+Hybrid

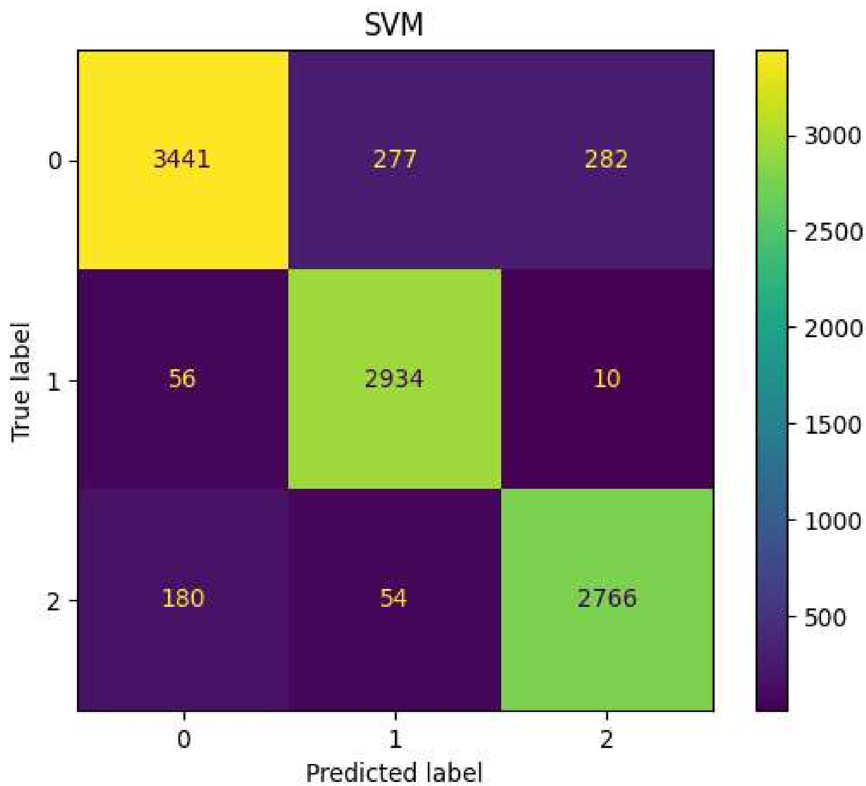


Figure 10.91. Confusion Matrix Obtained with SVM

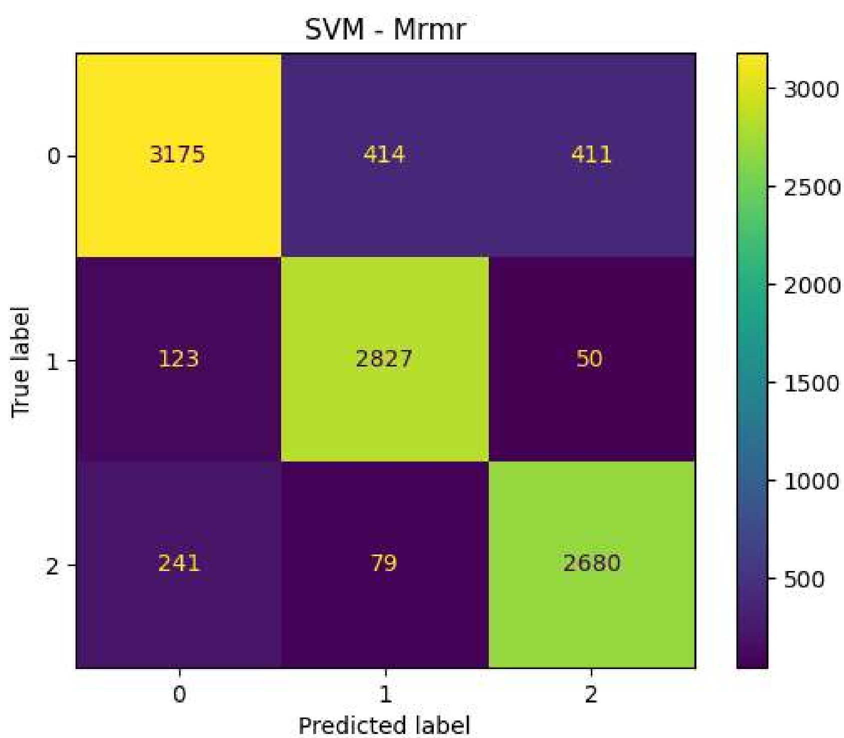


Figure 10.92. Confusion Matrix Obtained with SVM+mRMR

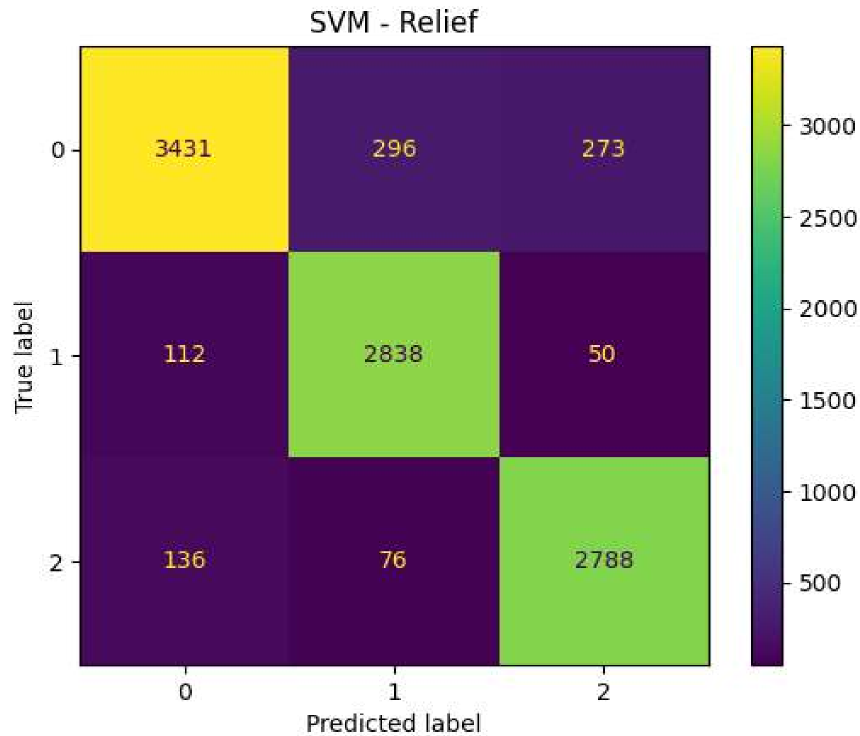


Figure 10.93. Confusion Matrix Obtained with SVM+Relief-F

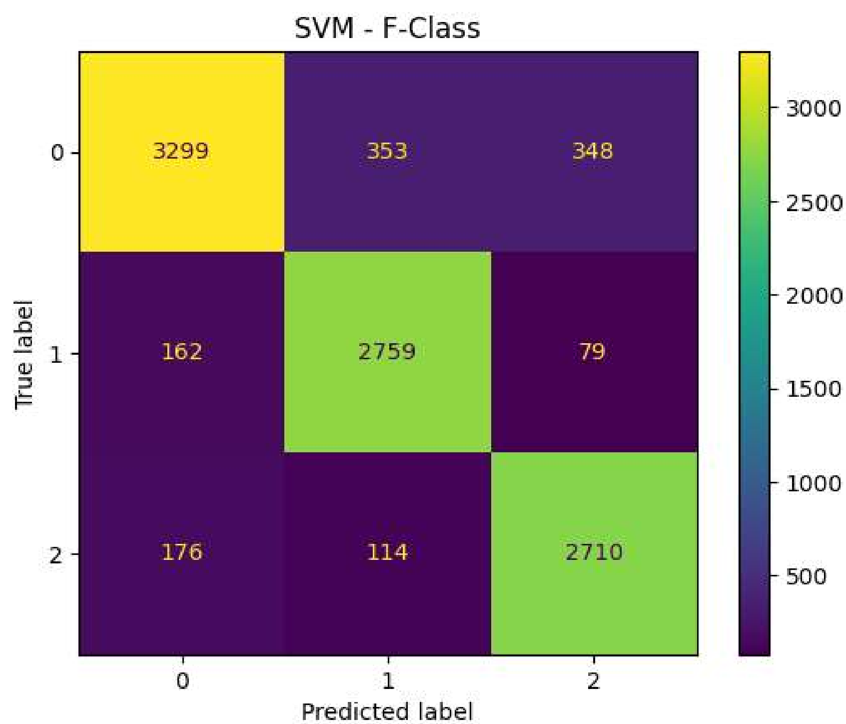


Figure 10.94. Confusion Matrix Obtained with SVM+F-Classification

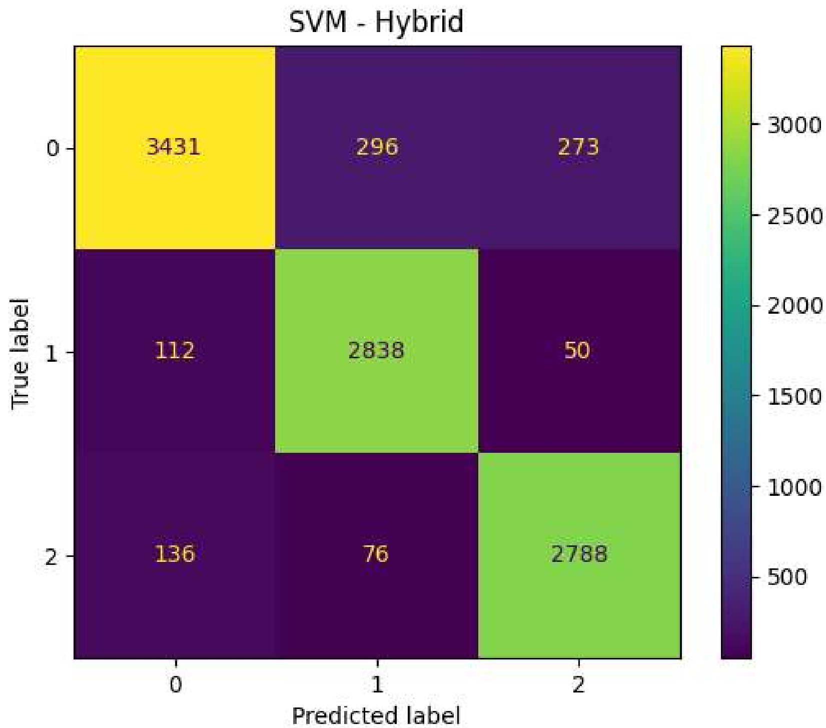


Figure 10.95. Confusion Matrix Obtained with SVM+Hybrid

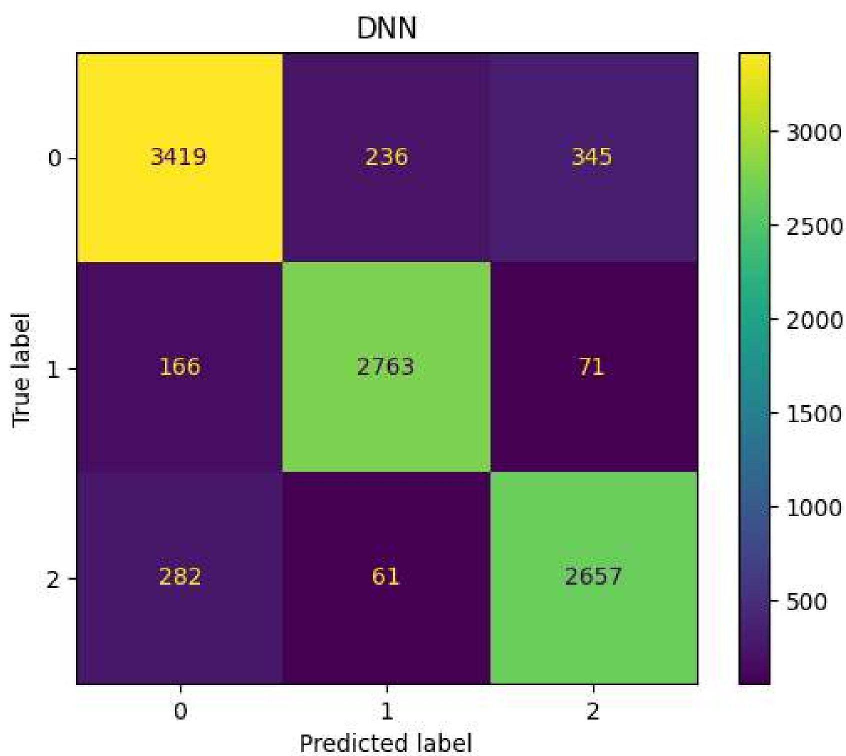


Figure 10.96. Confusion Matrix Obtained with DNN

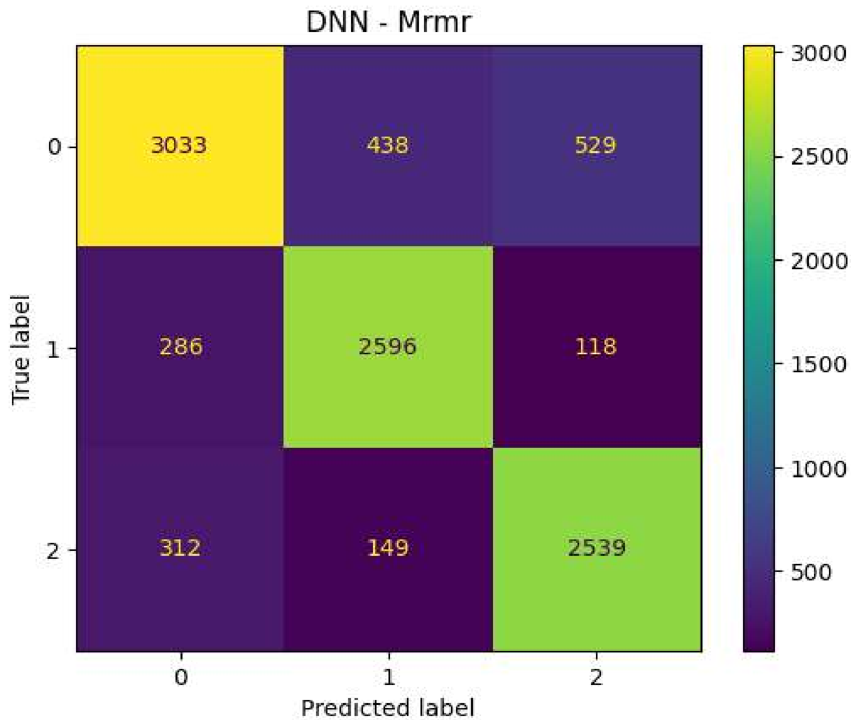


Figure 10.97. Confusion Matrix Obtained with DNN+mRMR

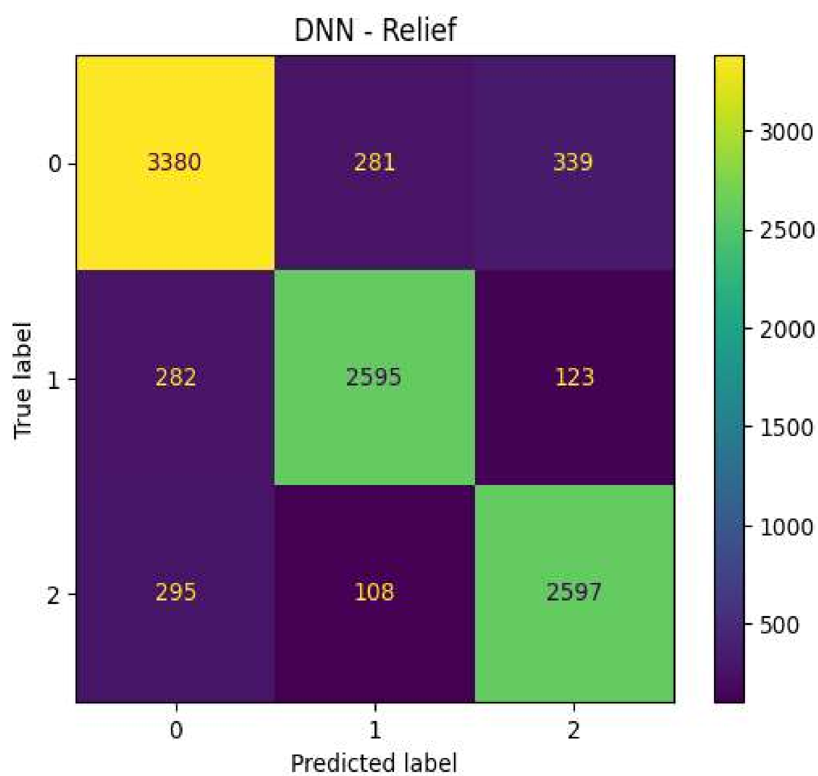


Figure 10.98. Confusion Matrix Obtained with DNN+Relief-F

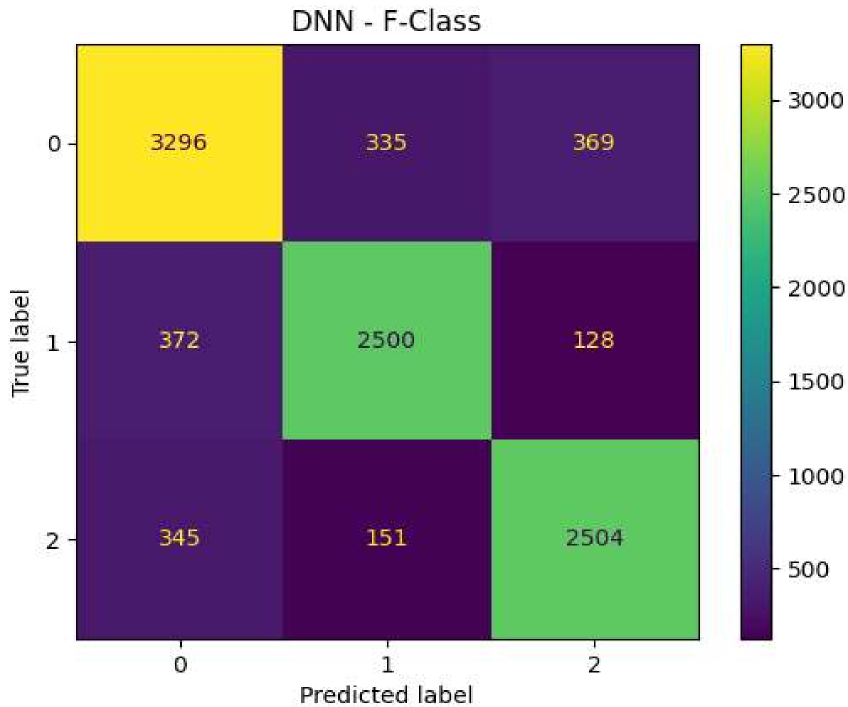


Figure 10.99. Confusion Matrix Obtained with DNN+F-Classification

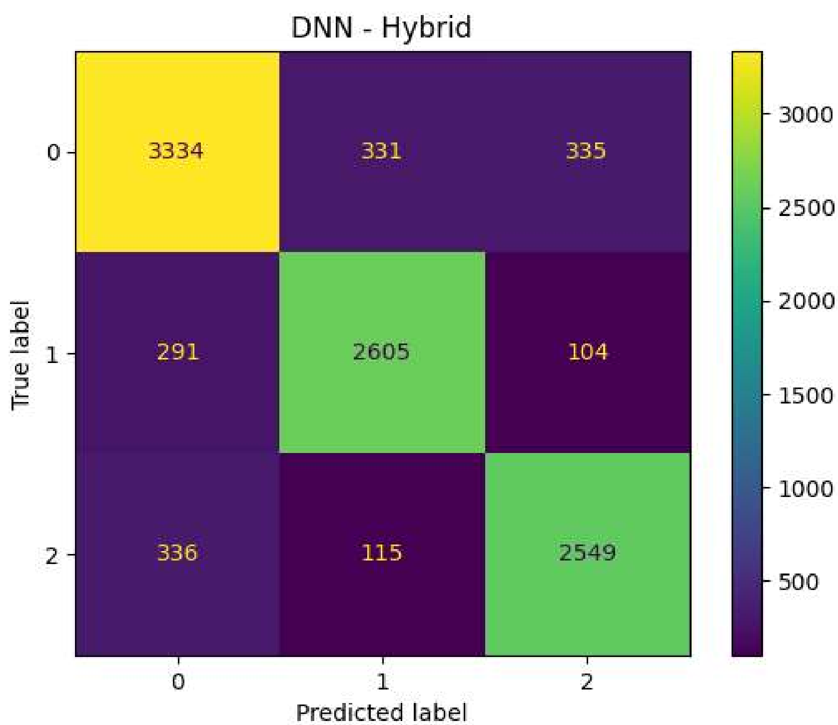


Figure 10.100. Confusion Matrix Obtained with DNN+Hybrid

10.2.2. Binary Classification Based Cancellation Prediction Models

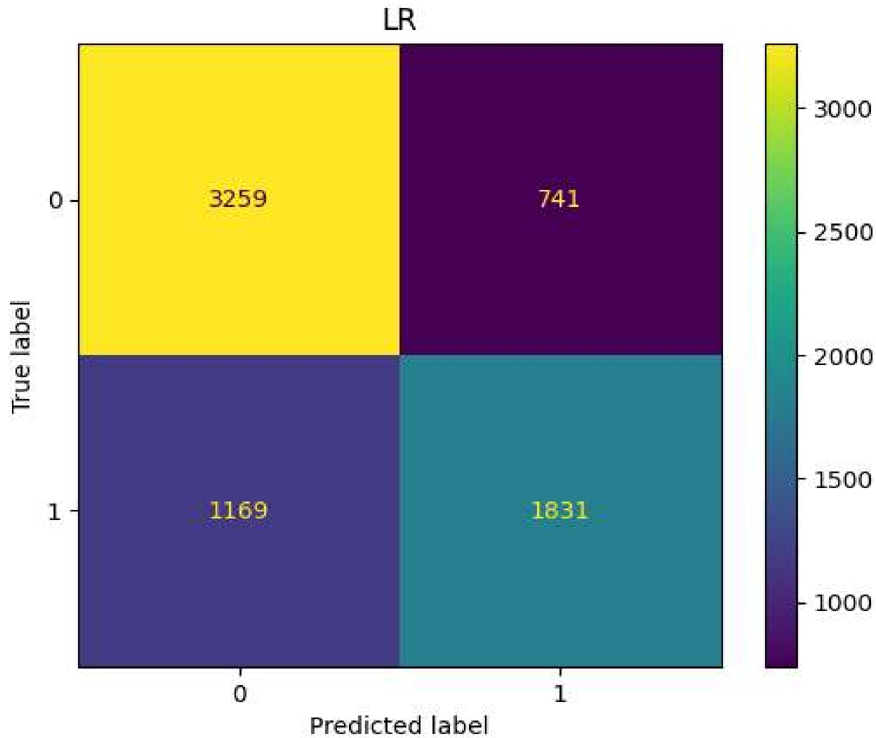


Figure 10. 101. Confusion Matrix Obtained with LR

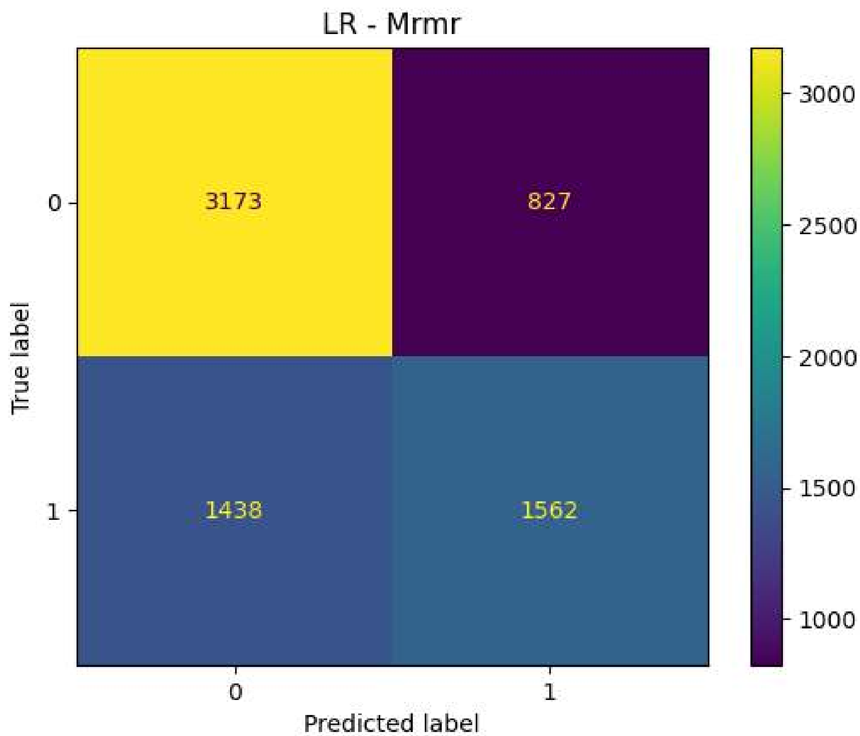


Figure 10.102. Confusion Matrix Obtained with LR+mRMR

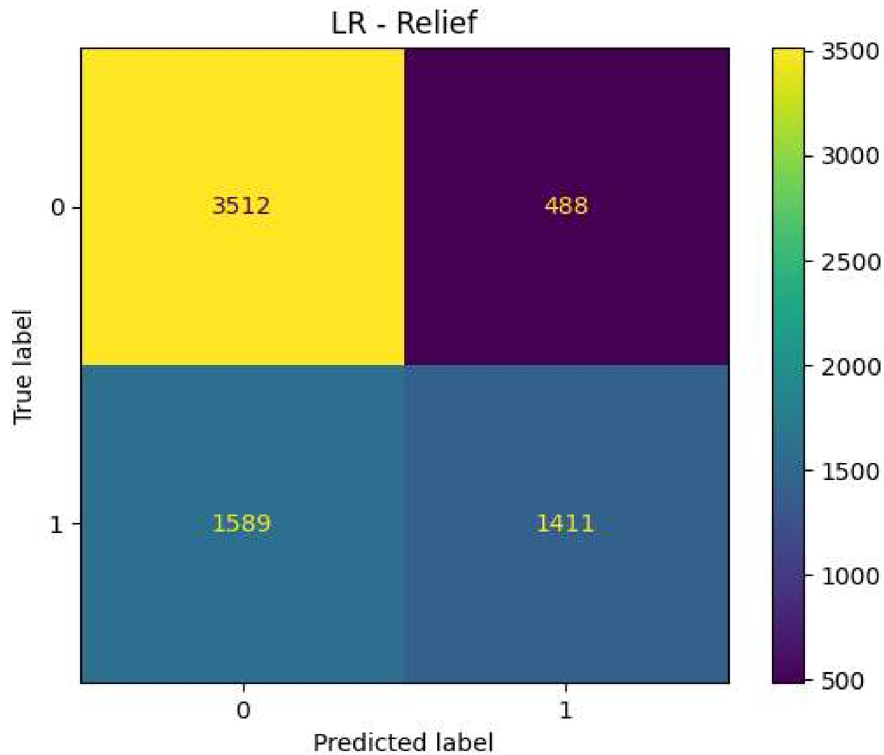


Figure 10.103. Confusion Matrix Obtained with LR+Relief-F

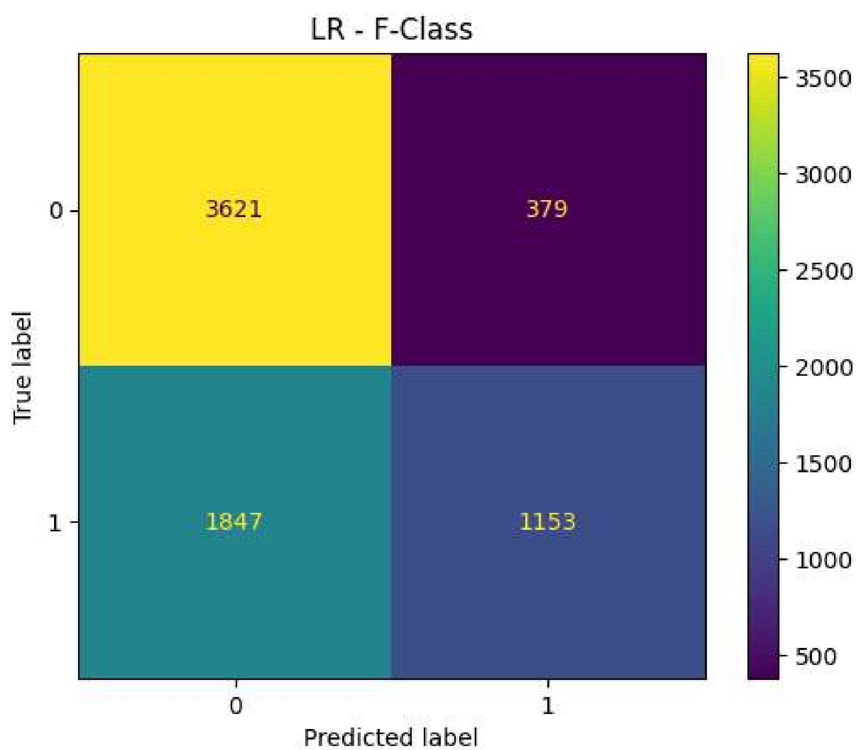


Figure 10.104. Confusion Matrix Obtained with LR+F-Classification

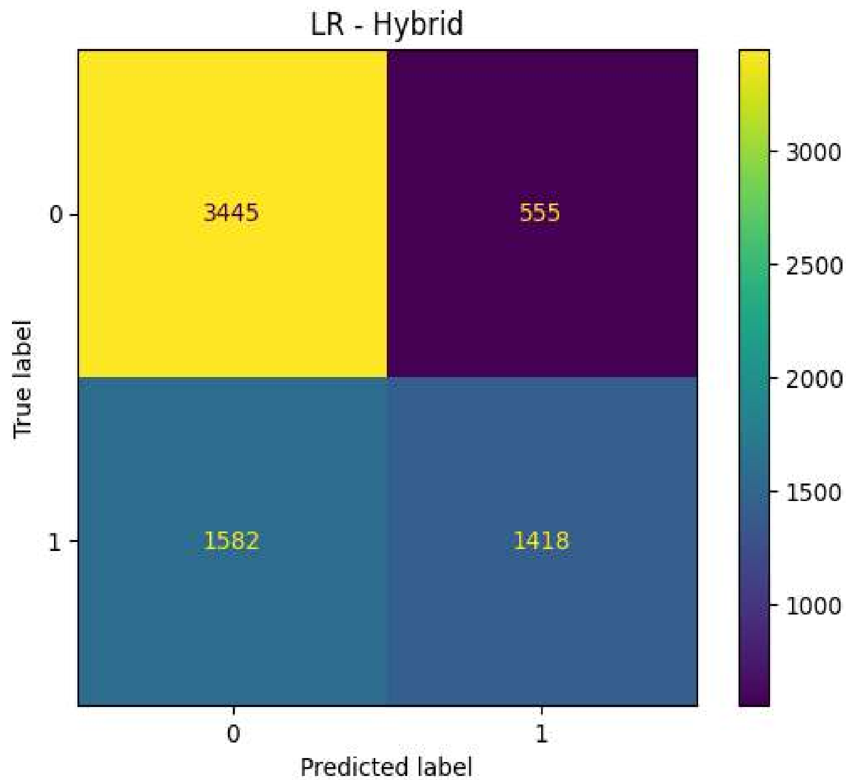


Figure 10.105. Confusion Matrix Obtained with LR+Hybrid

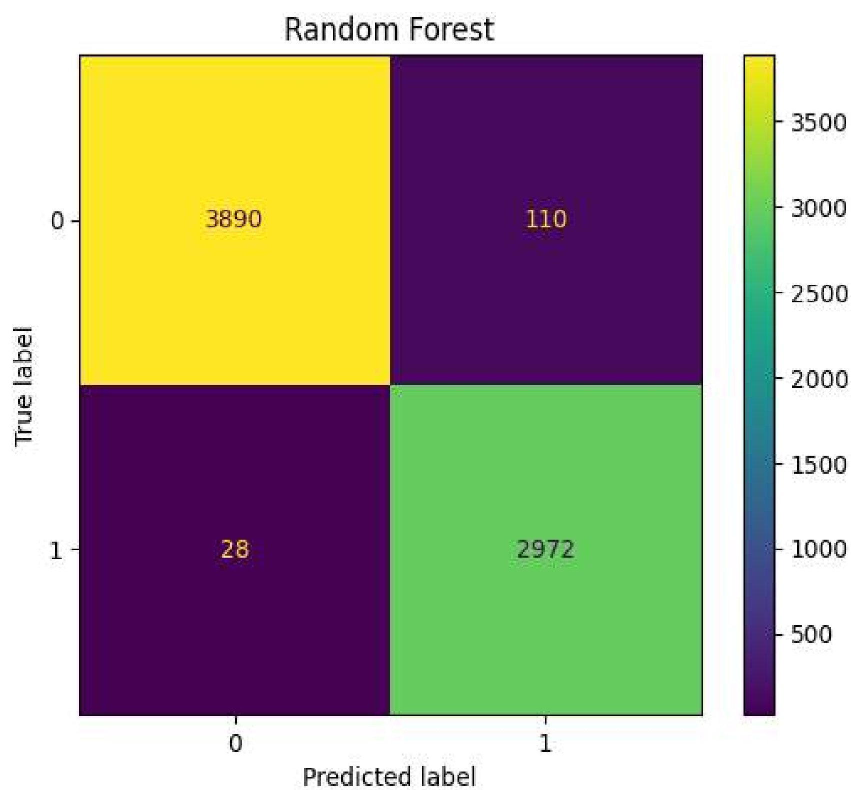


Figure 10.106. Confusion Matrix Obtained with RF

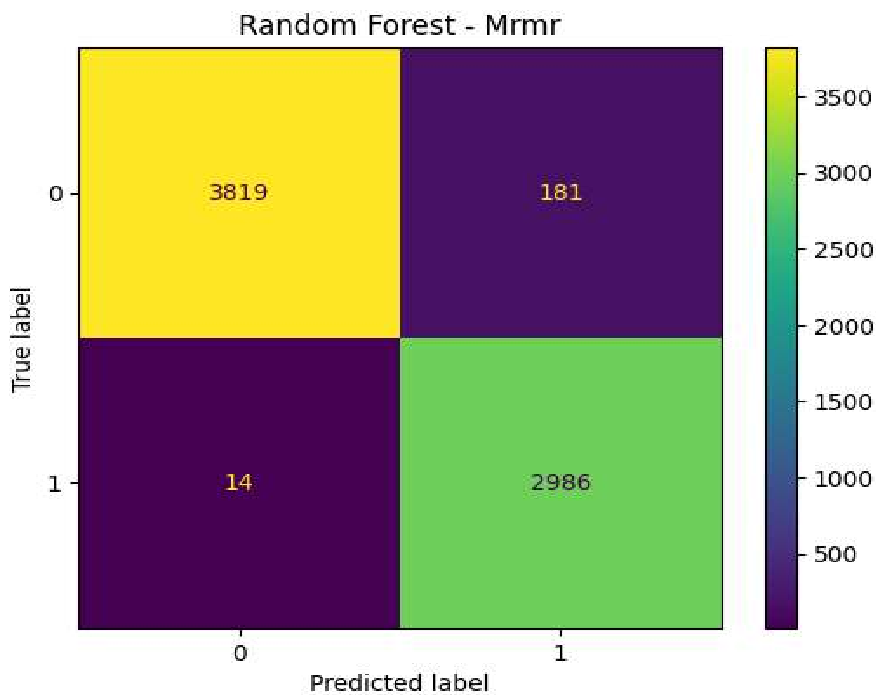


Figure 10.107. Confusion Matrix Obtained with RF+mRMR

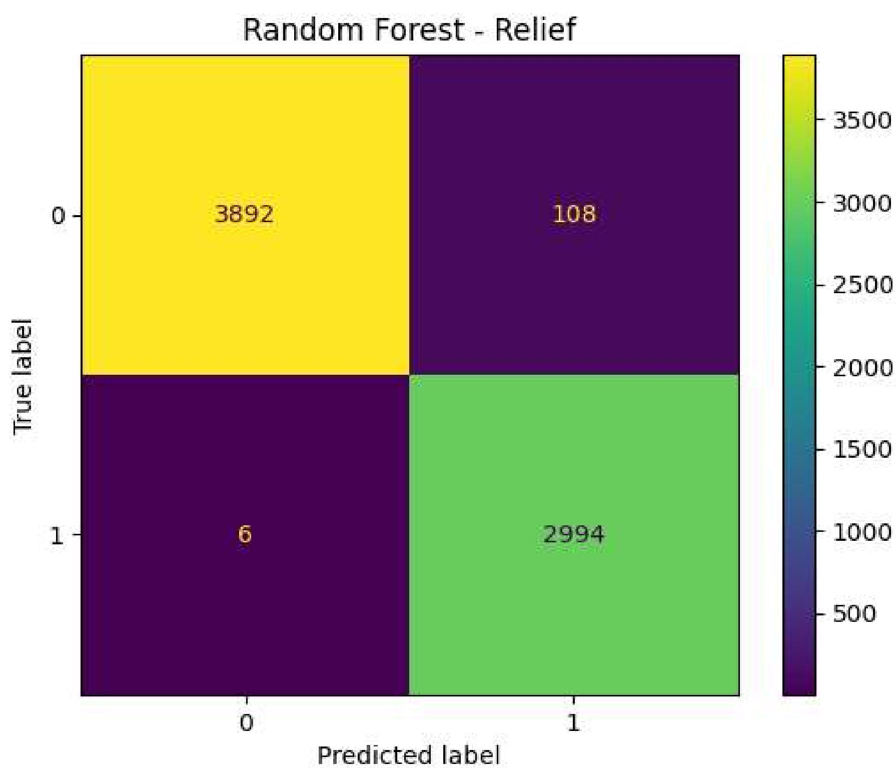


Figure 10.108. Confusion Matrix Obtained with RF+Relief-F

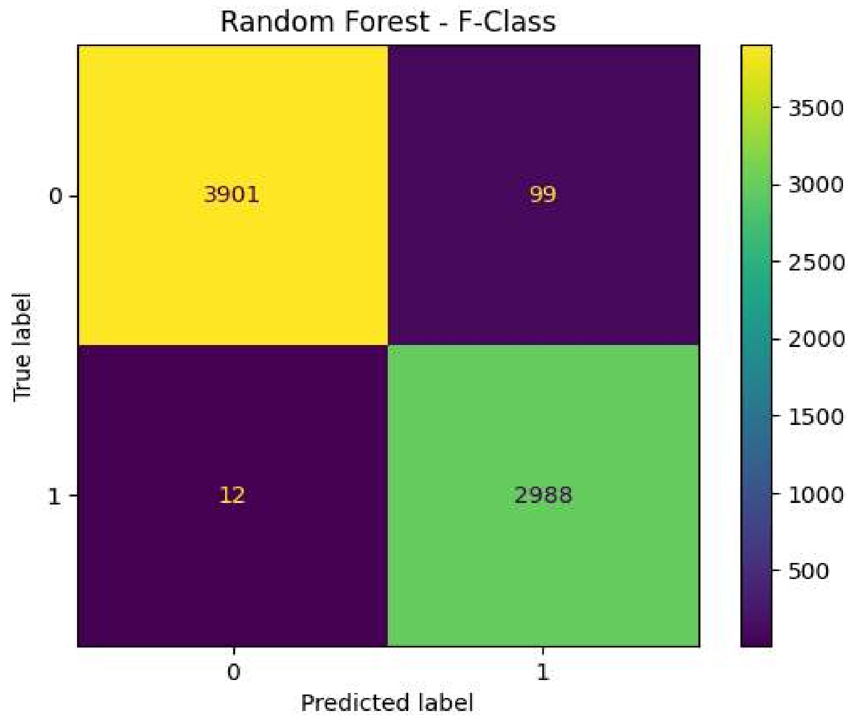


Figure 10.109. Confusion Matrix Obtained with RF+F-Classification

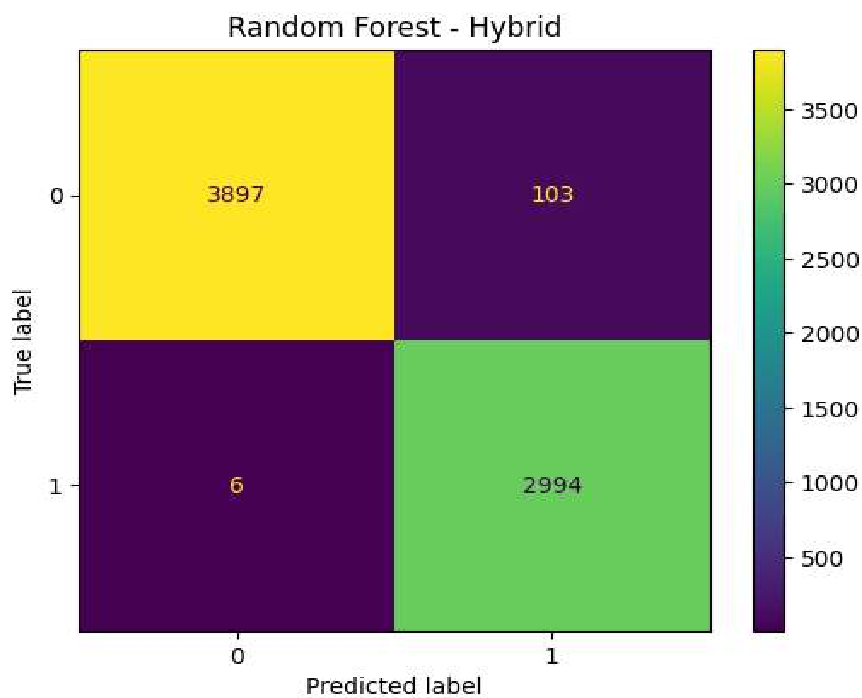


Figure 10.110. Confusion Matrix Obtained with RF+Hybrid

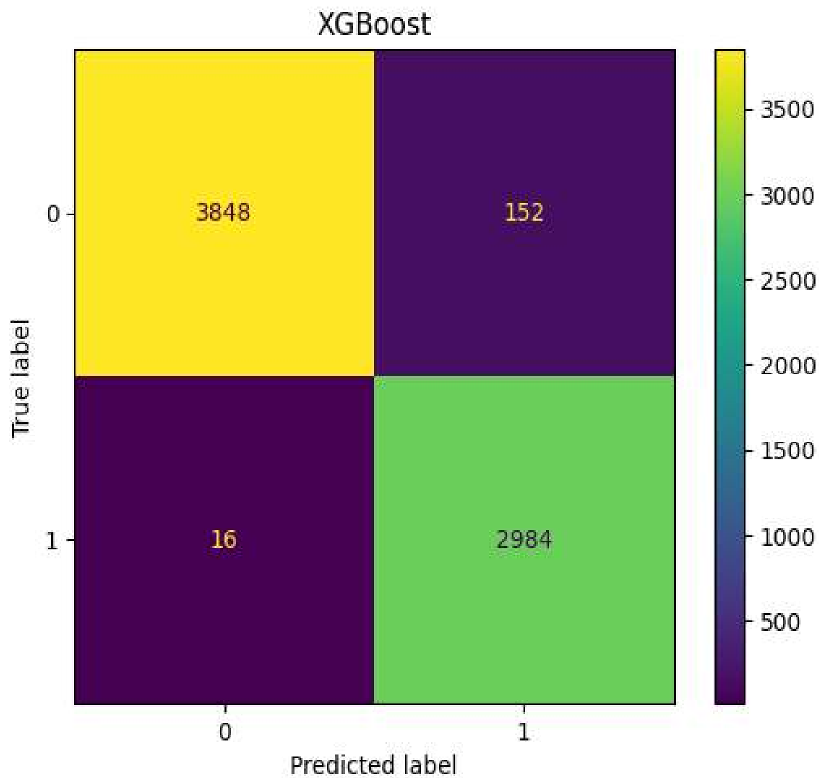


Figure 10.111. Confusion Matrix Obtained with XGBoost

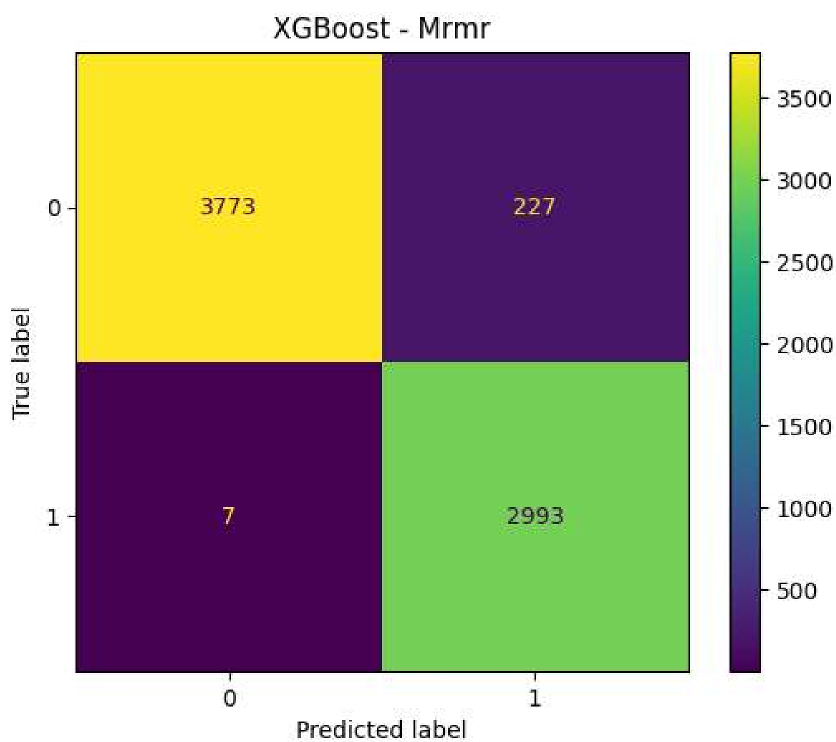


Figure 10.112. Confusion Matrix Obtained with XGBoost+mRMR

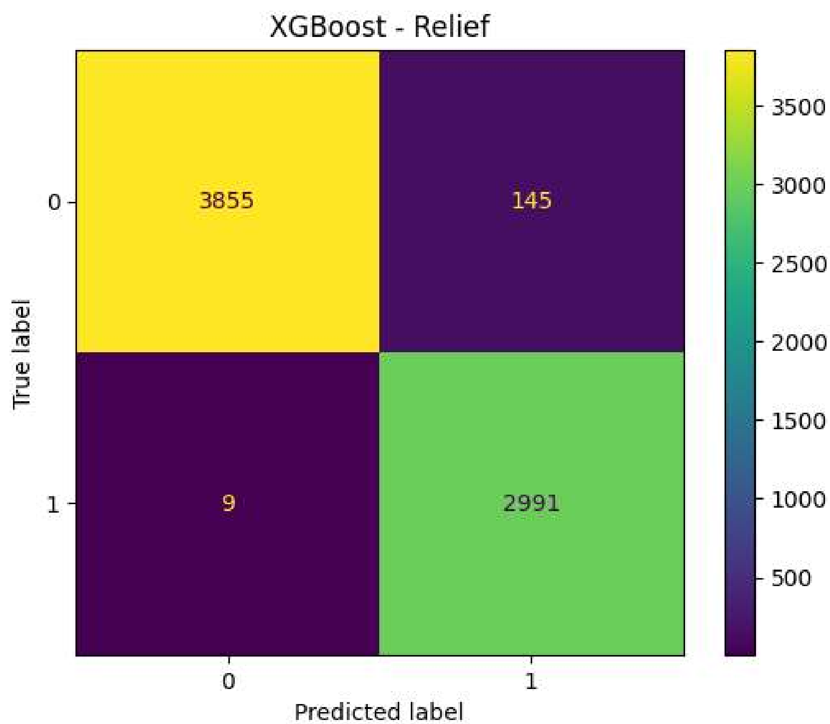


Figure 10.113. Confusion Matrix Obtained with XGBoost+Relief-F

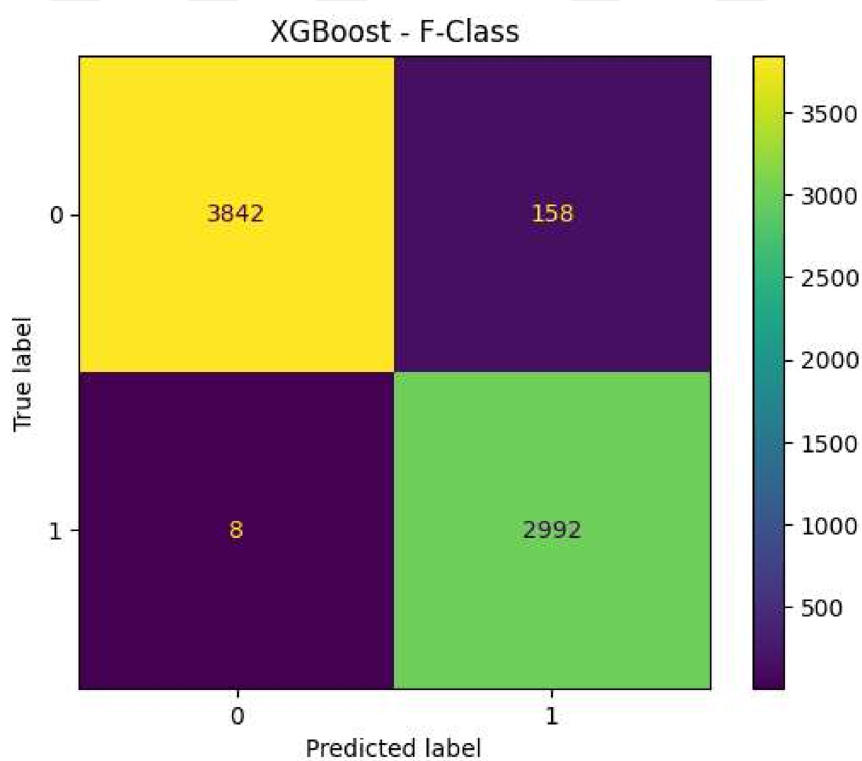


Figure 10.114. Confusion Matrix Obtained with XGBoost+F-Classification

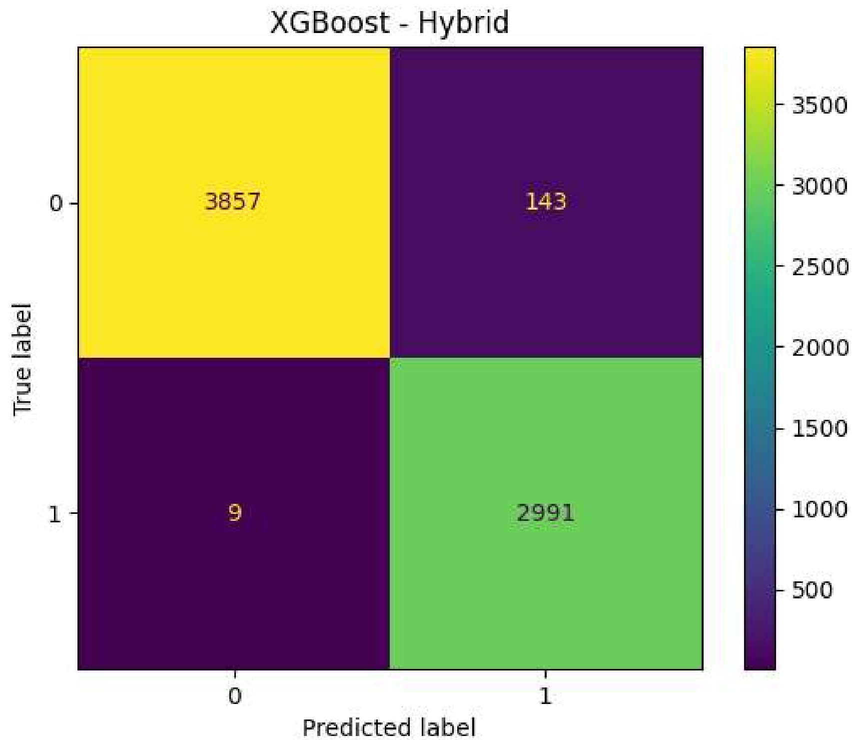


Figure 10.115. Confusion Matrix Obtained with XGBoost+Hybrid

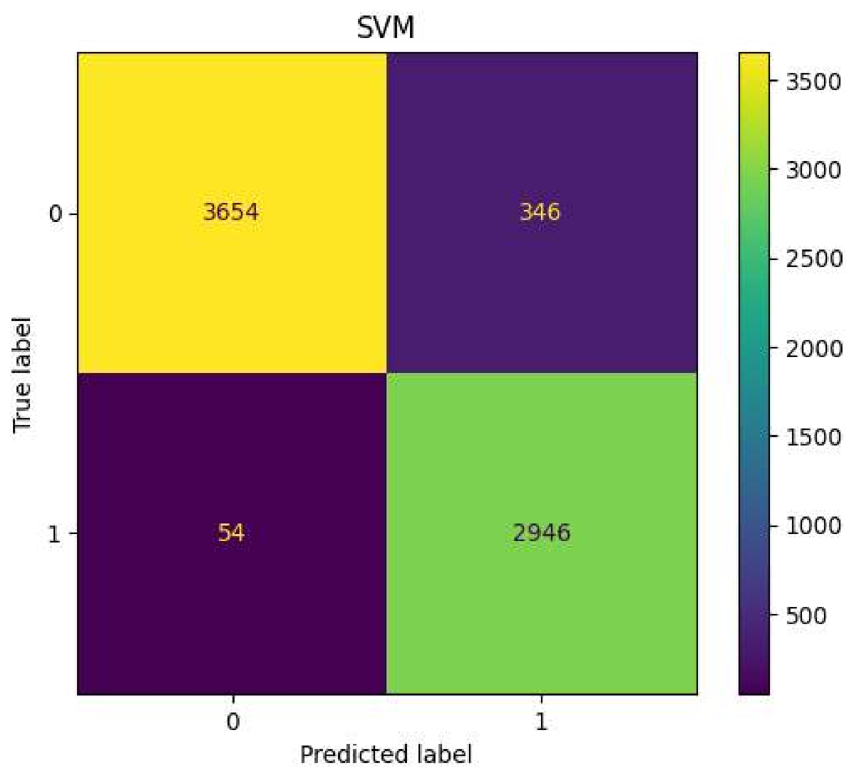


Figure 10.116. Confusion Matrix Obtained with SVM

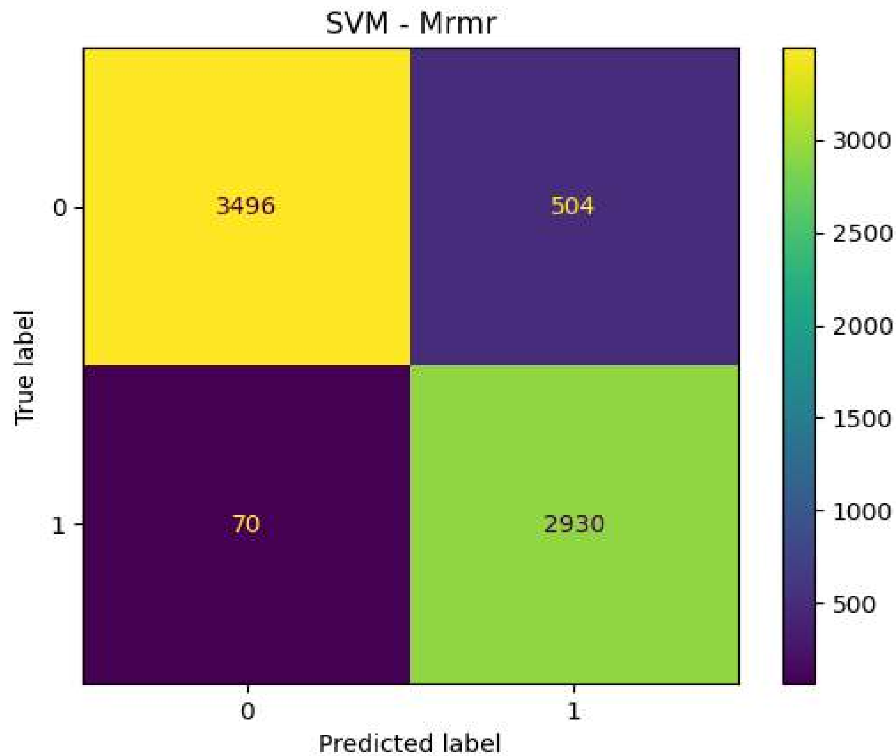


Figure 10.117. Confusion Matrix Obtained with SVM+mRMR

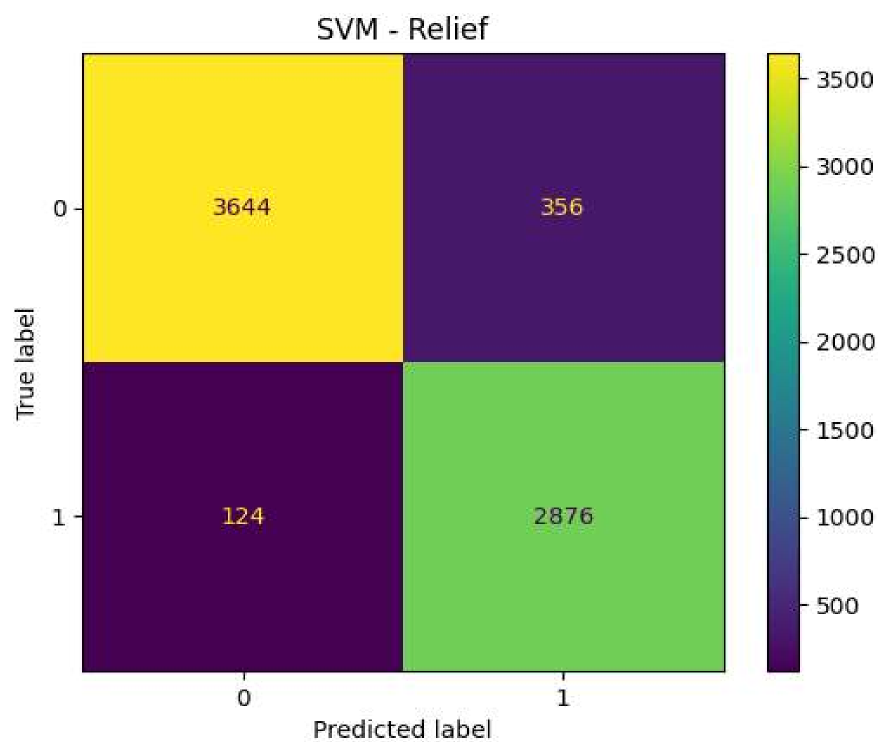


Figure 10.118. Confusion Matrix Obtained with SVM+Relief-F

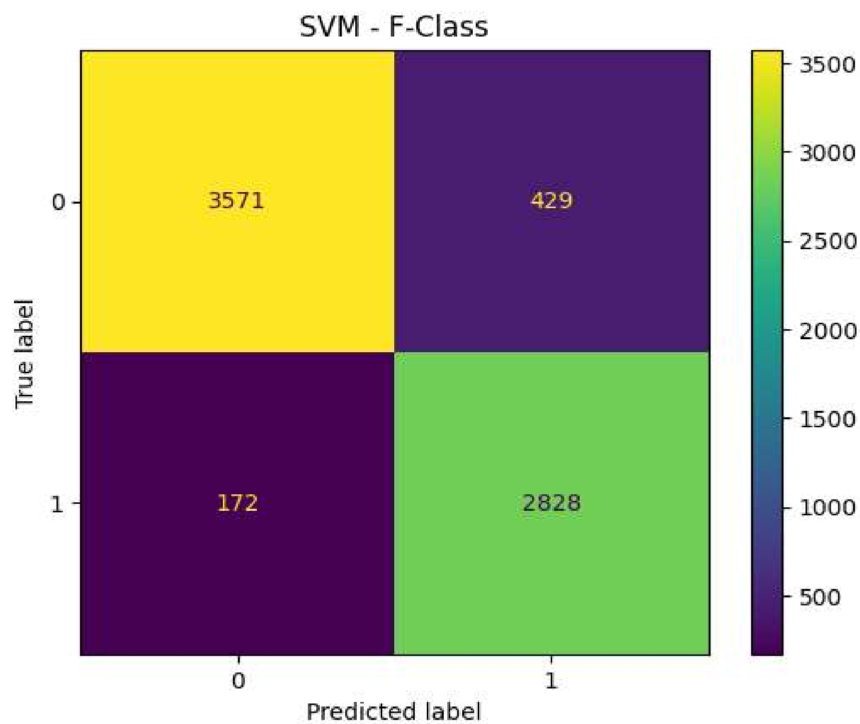


Figure 10.119. Confusion Matrix Obtained with SVM+F-Classification

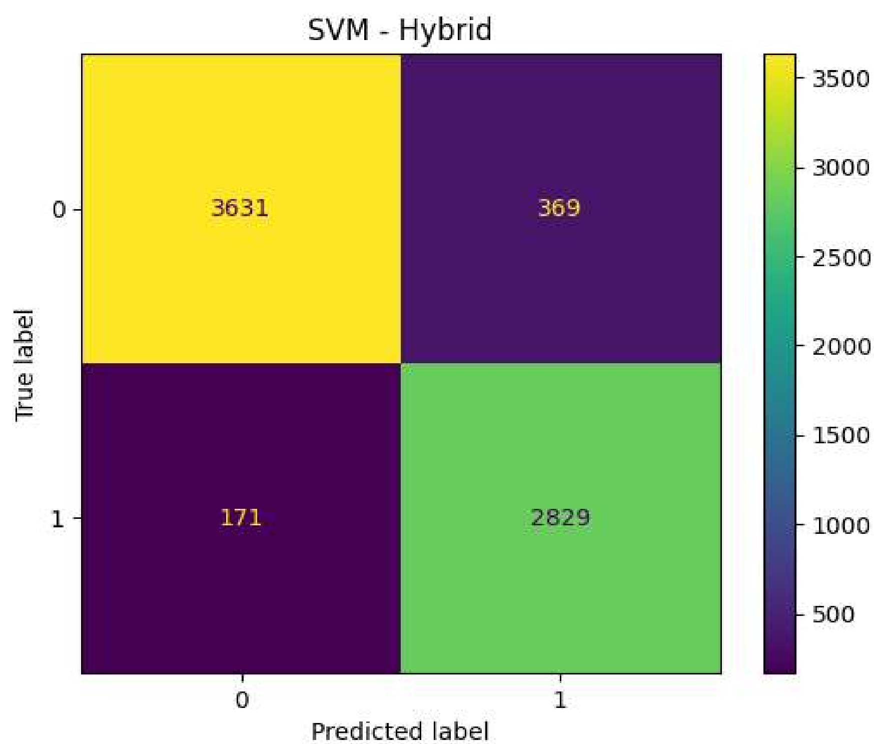


Figure 10.120. Confusion Matrix Obtained with SVM+Hybrid

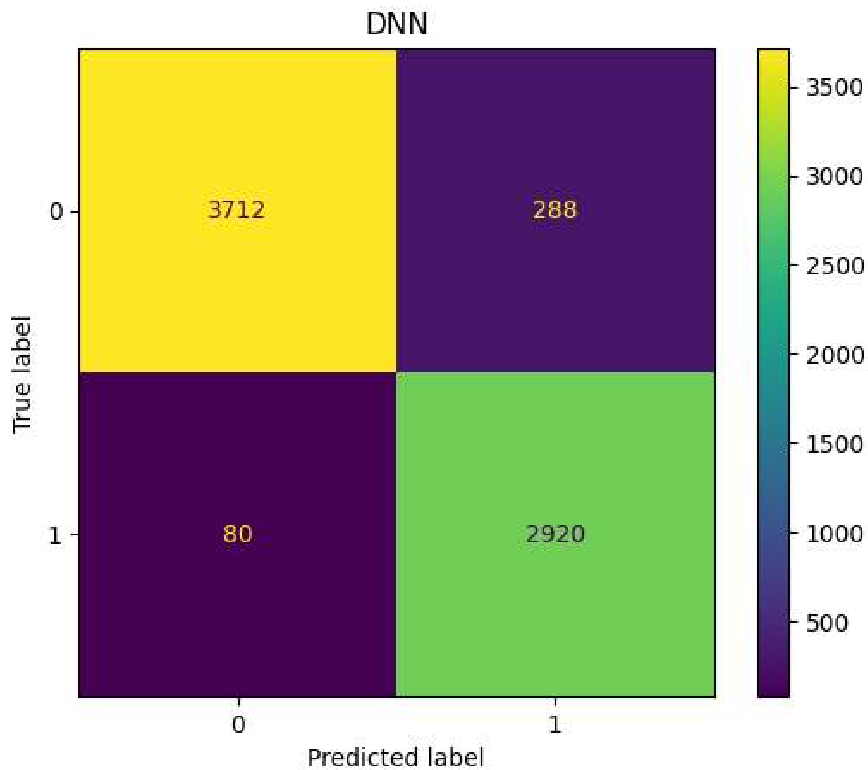


Figure 10.121. Confusion Matrix Obtained with DNN

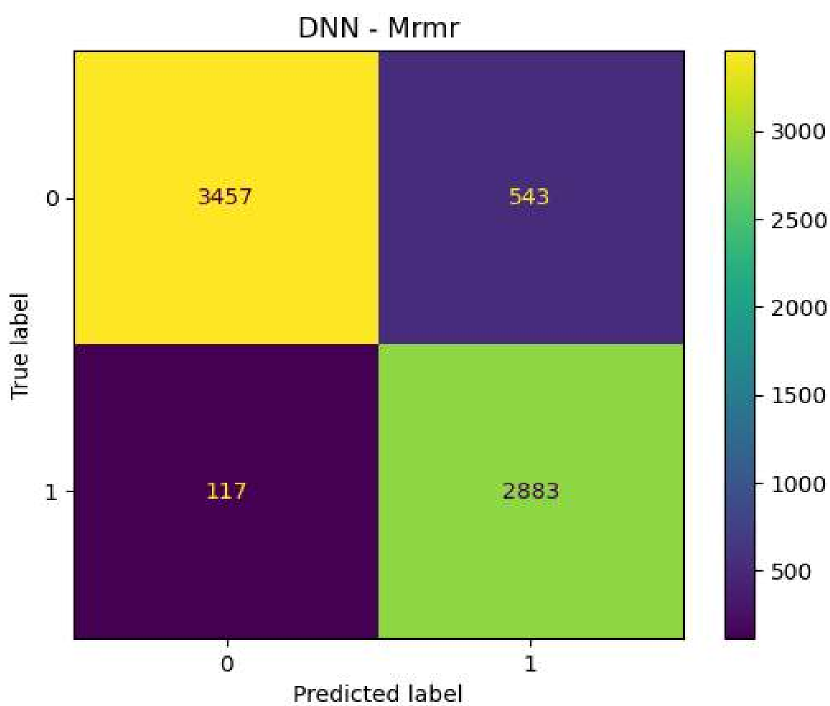


Figure 10.122. Confusion Matrix Obtained with DNN+mRMR

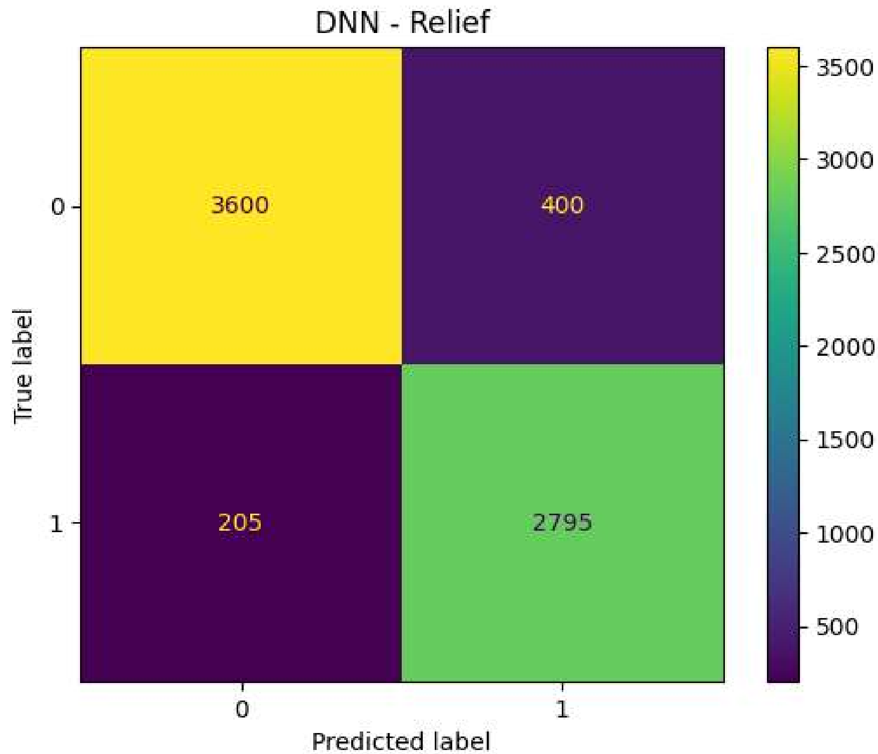


Figure 10.123 Confusion Matrix Obtained with DNN+Relief-F

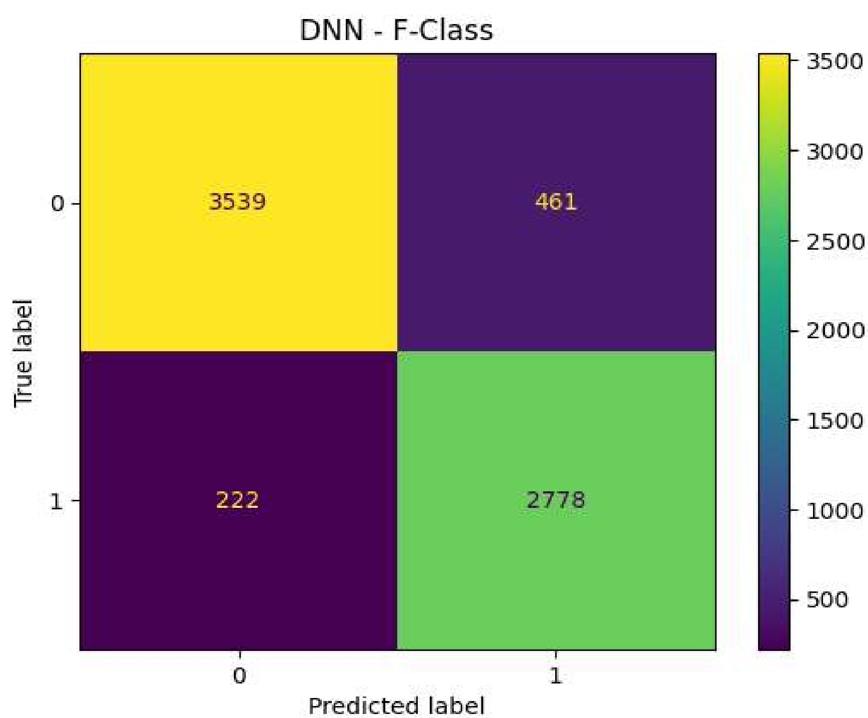


Figure 10.124. Confusion Matrix Obtained with DNN+F-Classification

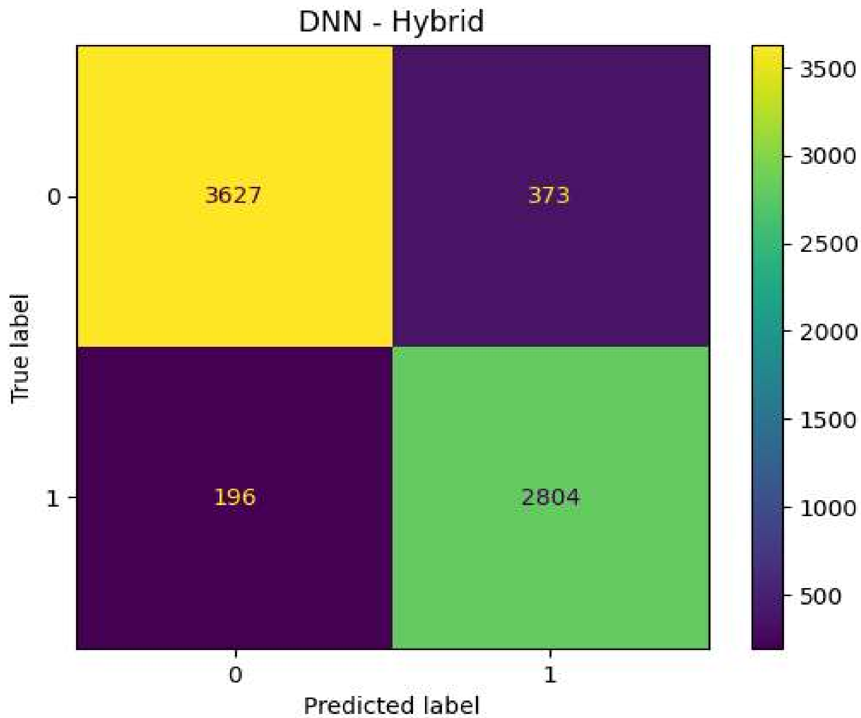


Figure 10.125. Confusion Matrix Obtained with DNN+Hybrid

10.2.3. Binary Classification Based Return Prediction Models

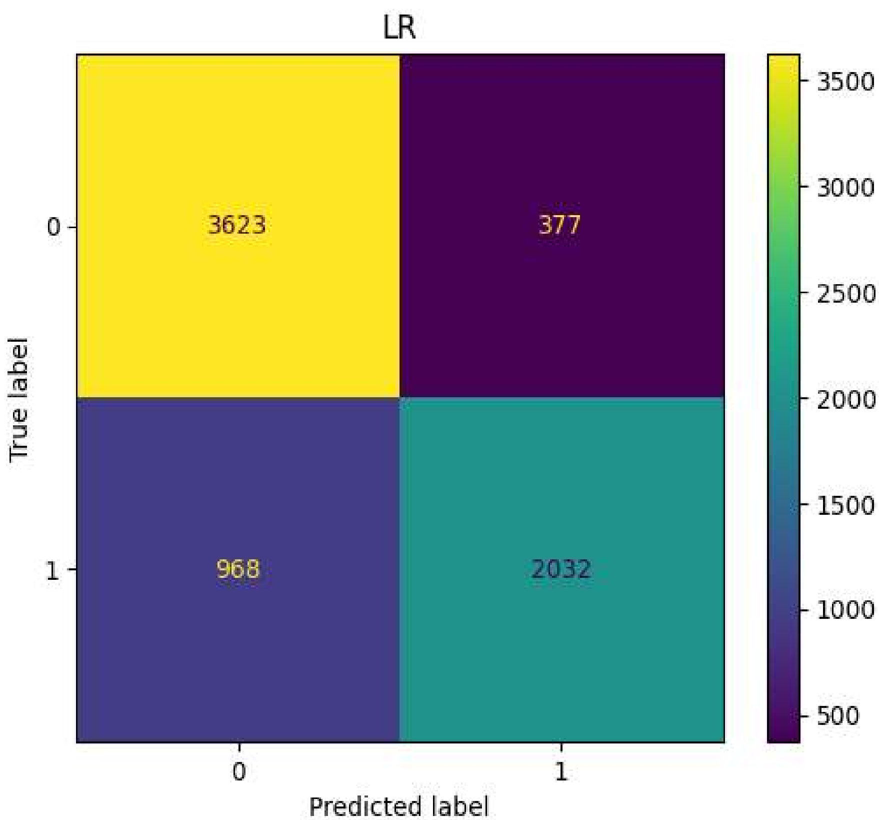


Figure 10.126. Confusion Matrix Obtained with LR

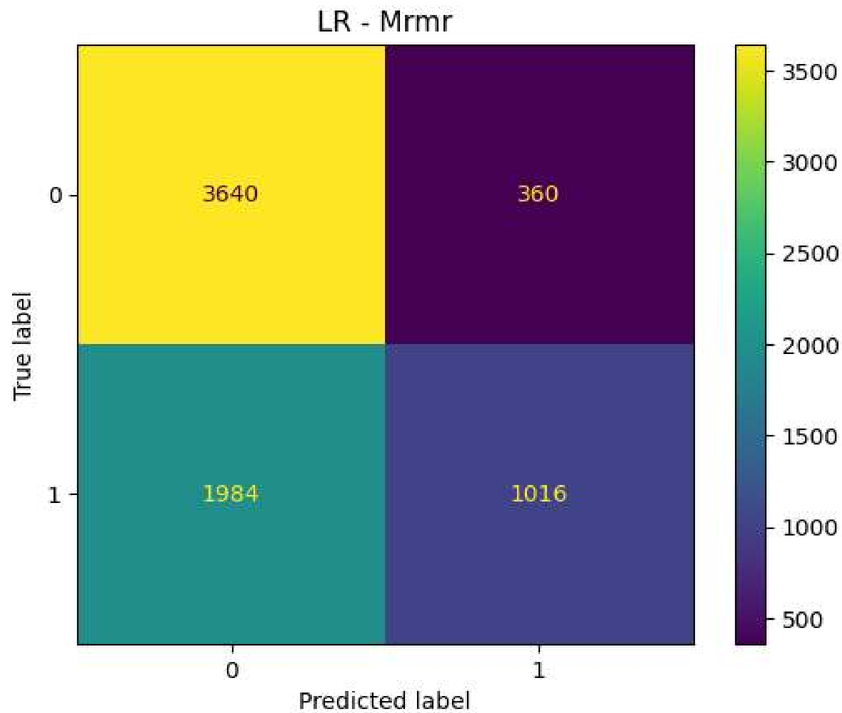


Figure 10.127. Confusion Matrix Obtained with LR+mRMR

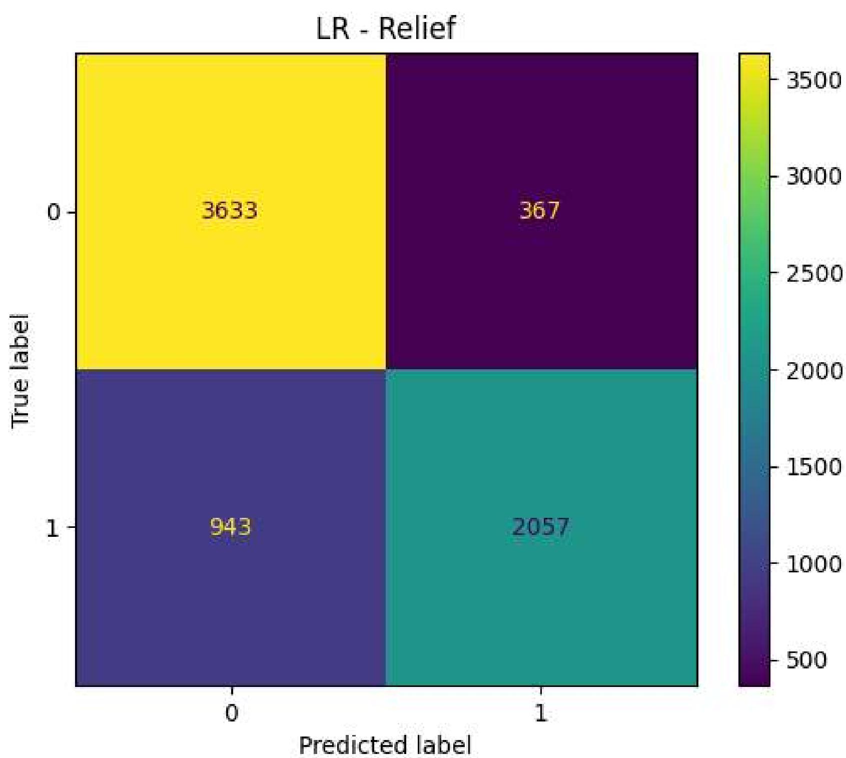


Figure 10.128. Confusion Matrix Obtained with LR+Relief-F

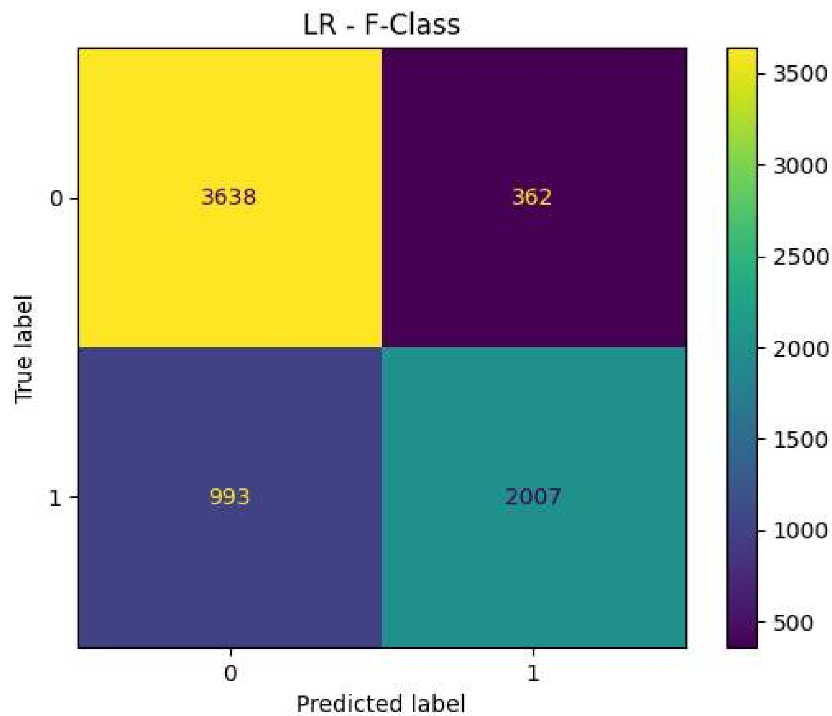


Figure 10.129. Confusion Matrix Obtained with LR+F-Classification

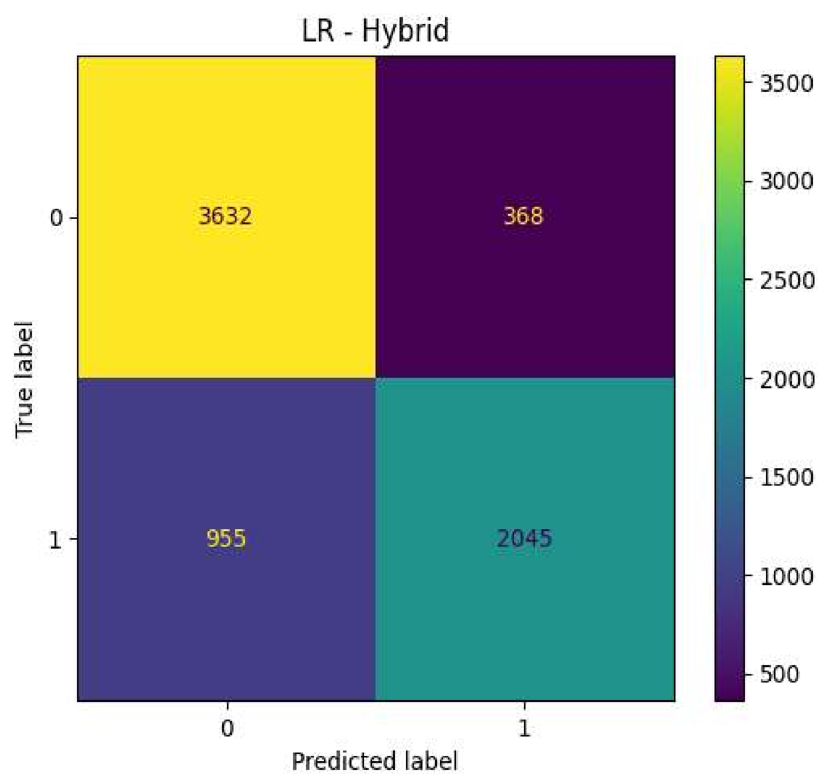


Figure 10.130. Confusion Matrix Obtained with LR+Hybrid

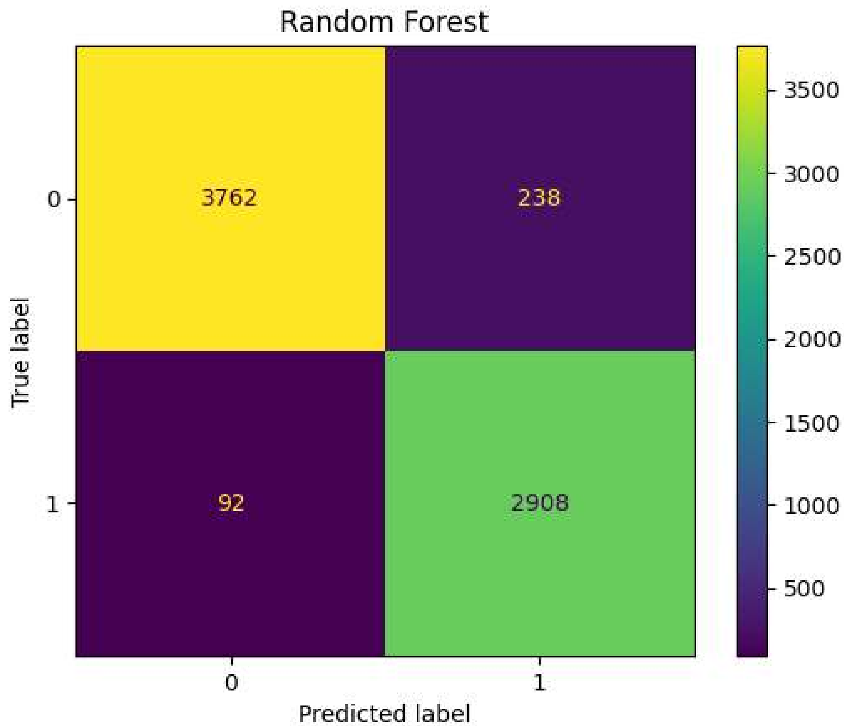


Figure 10.131. Confusion Matrix Obtained with RF

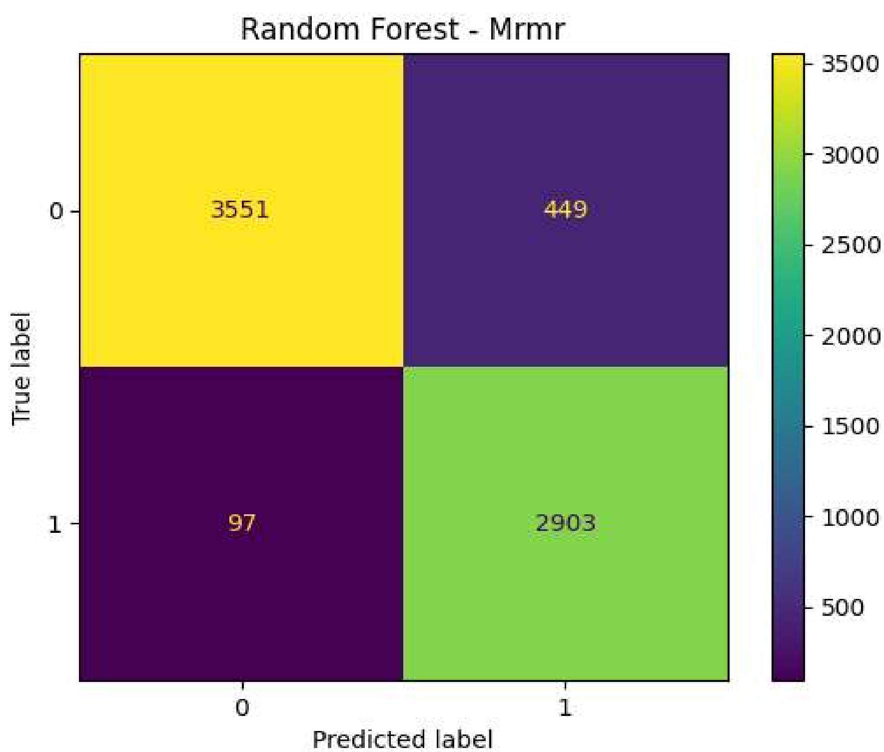


Figure 10.132. Confusion Matrix Obtained with RF+mRMR

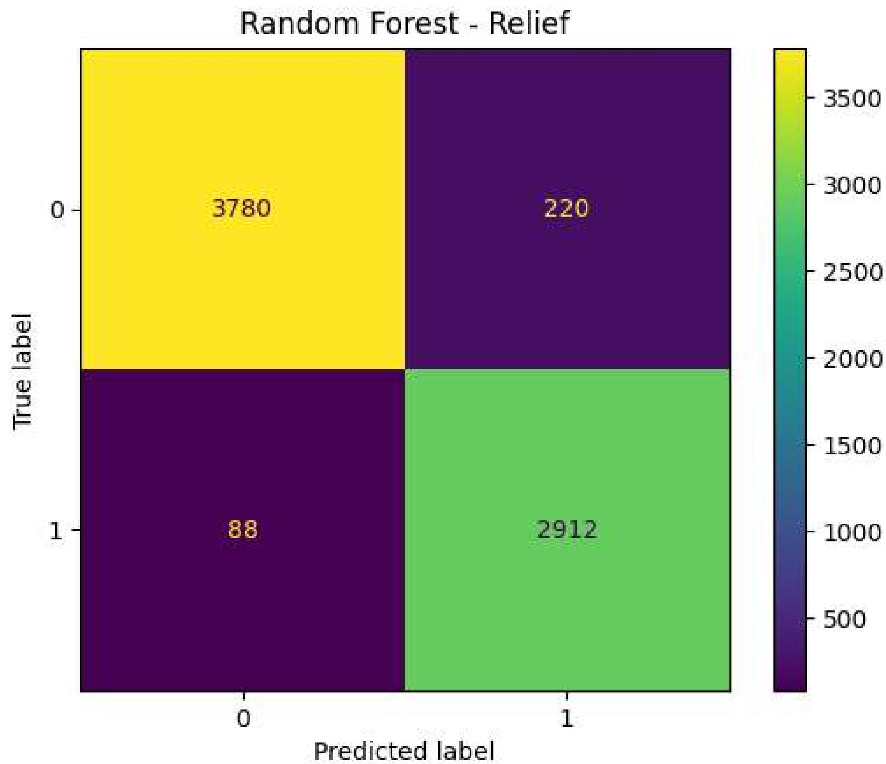


Figure 10.133 Confusion Matrix Obtained with RF+Relief-F

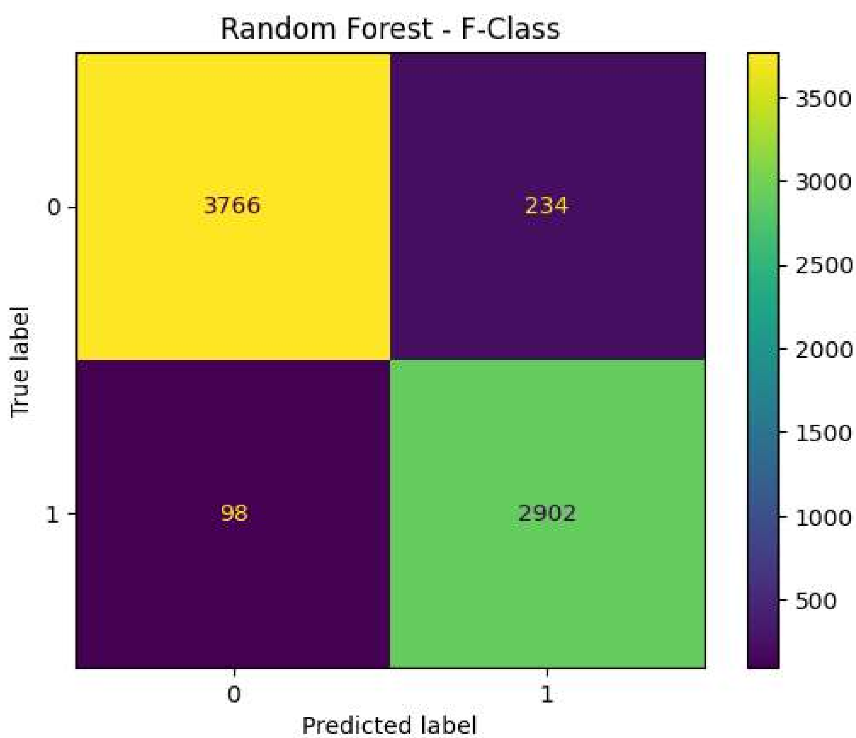


Figure 10.134. Confusion Matrix Obtained with RF+F-Classification

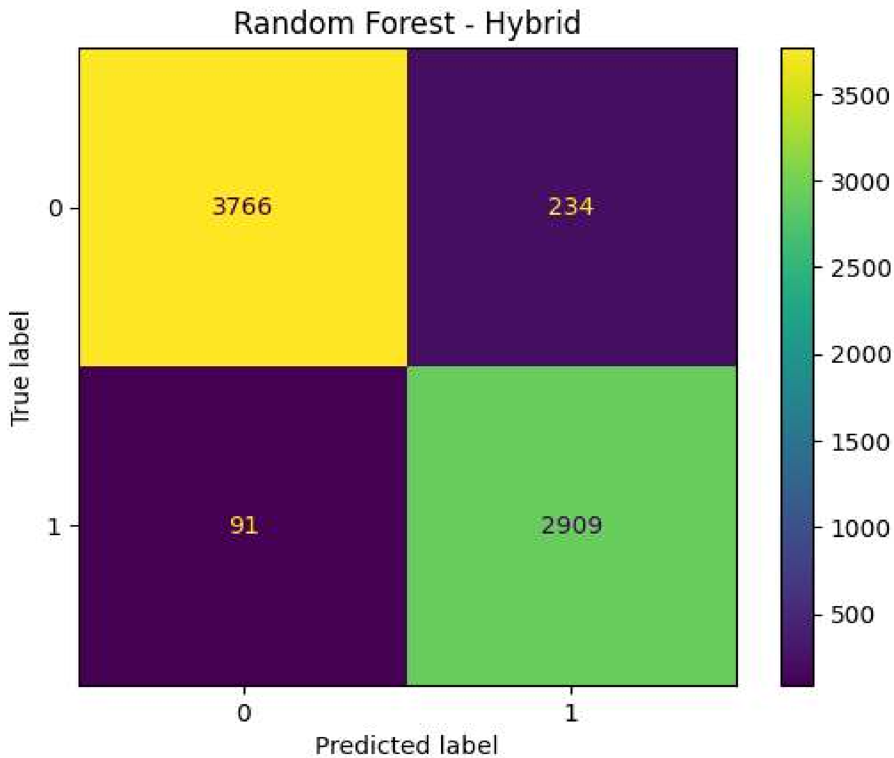


Figure 10.135. Confusion Matrix Obtained with RF+Hybrid

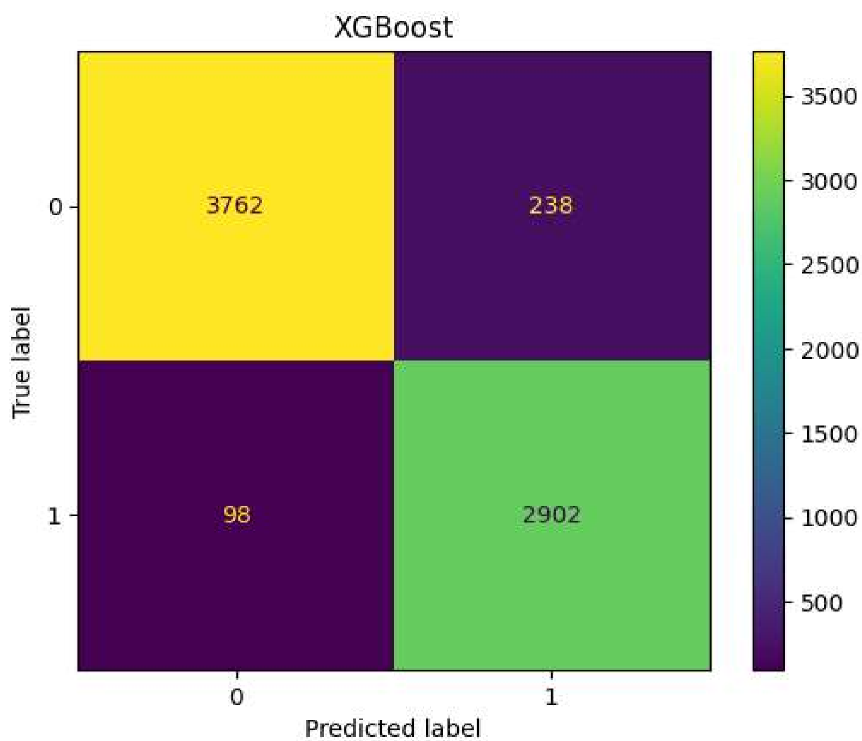


Figure 10.136. Confusion Matrix Obtained with XGBoost

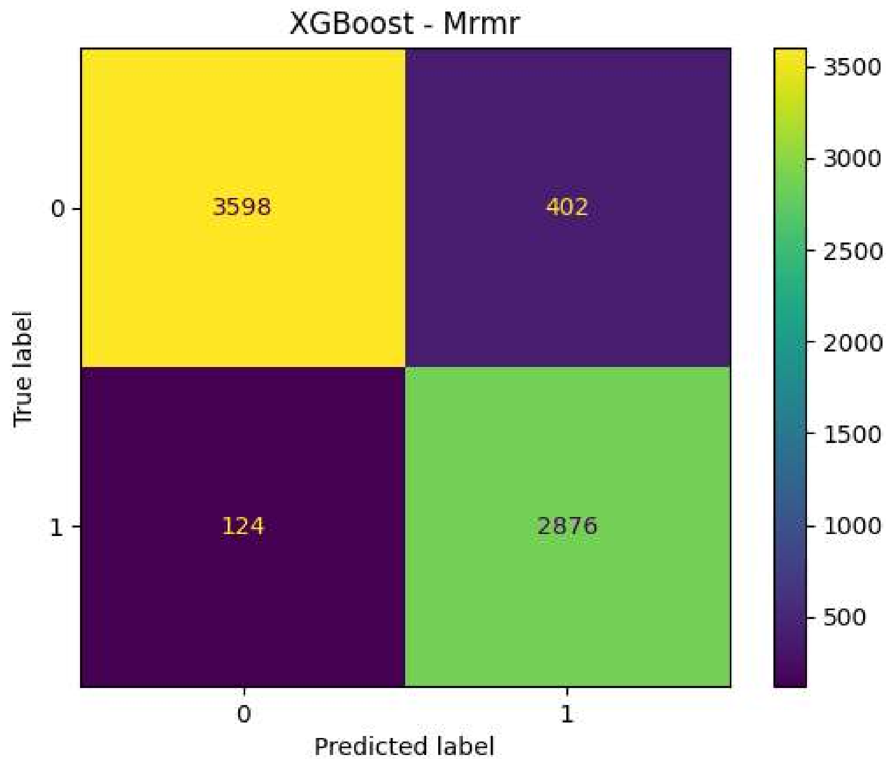


Figure 10.137. Confusion Matrix Obtained with XGBoost+mRMR

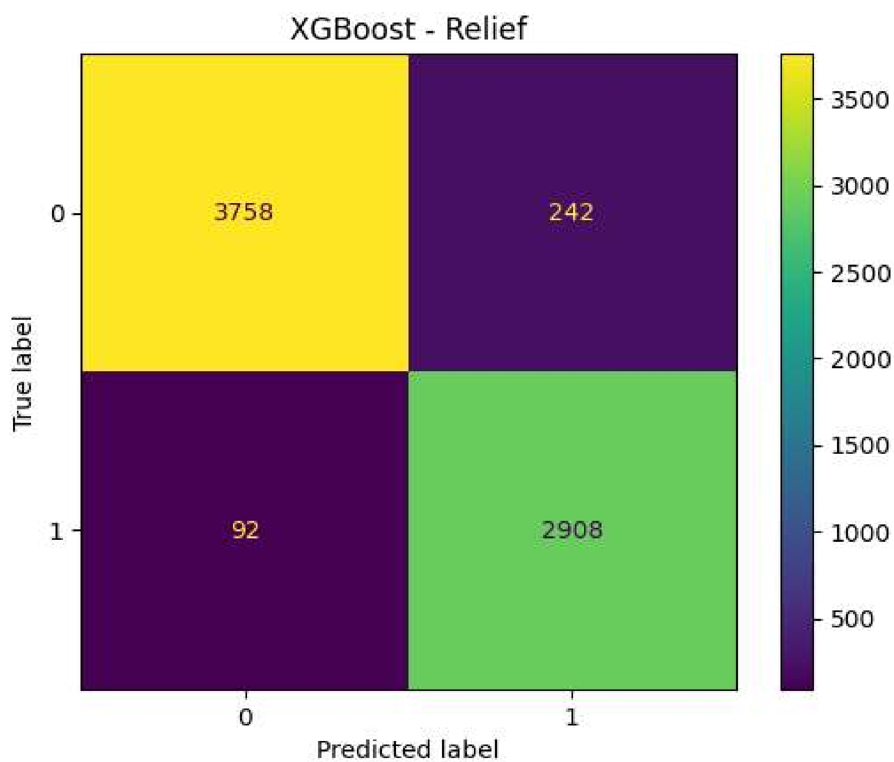


Figure 10.138. Confusion Matrix Obtained with XGBoost+Relief-F

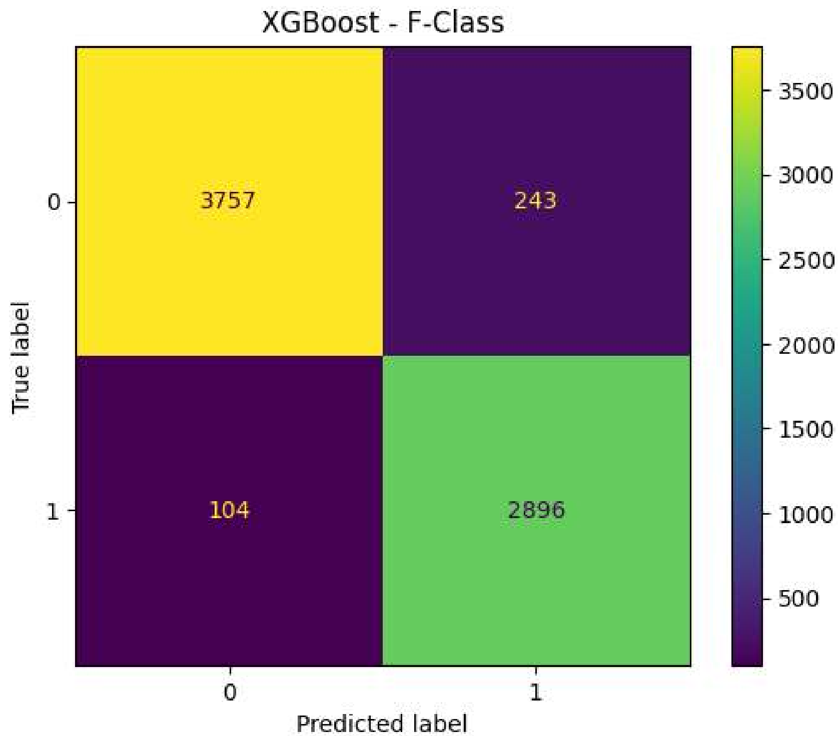


Figure 10.139. Confusion Matrix Obtained with XGBoost+F-Classification

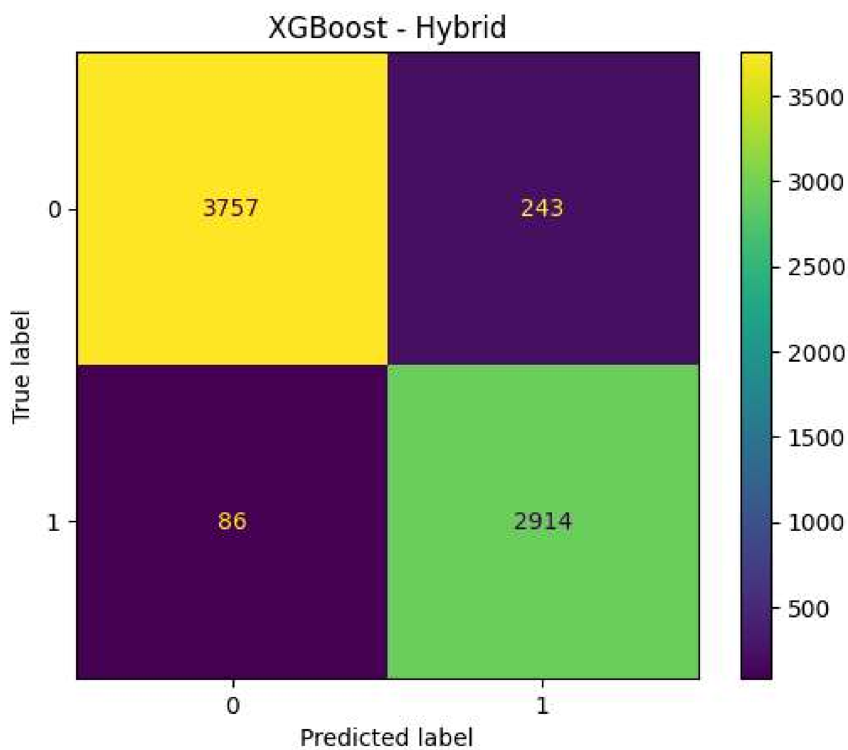


Figure 10.140. Confusion Matrix Obtained with XGBoost+Hybrid

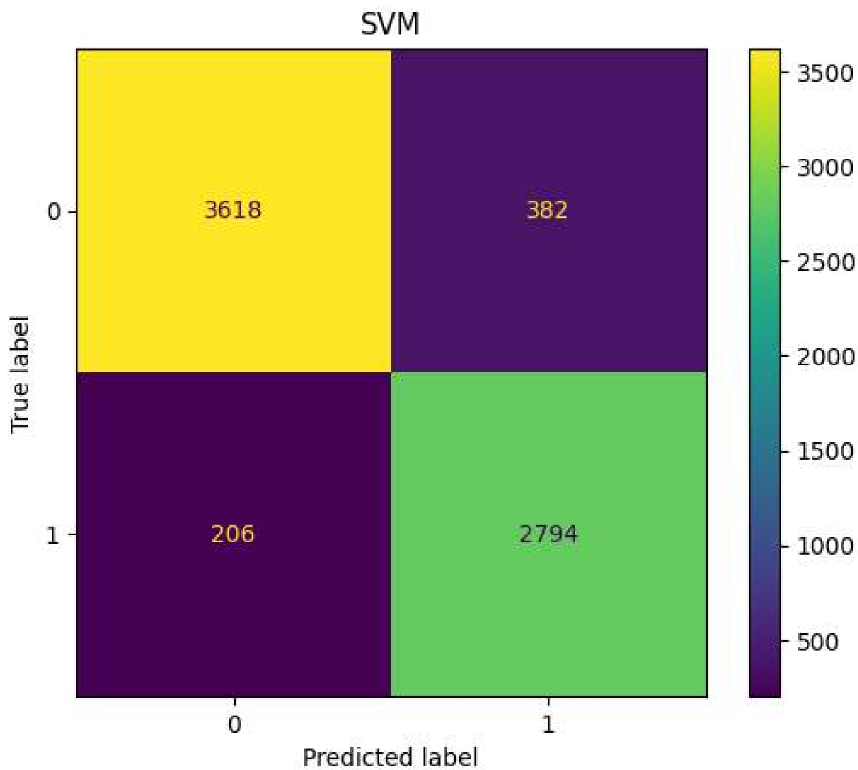


Figure 10.141. Confusion Matrix Obtained with SVM

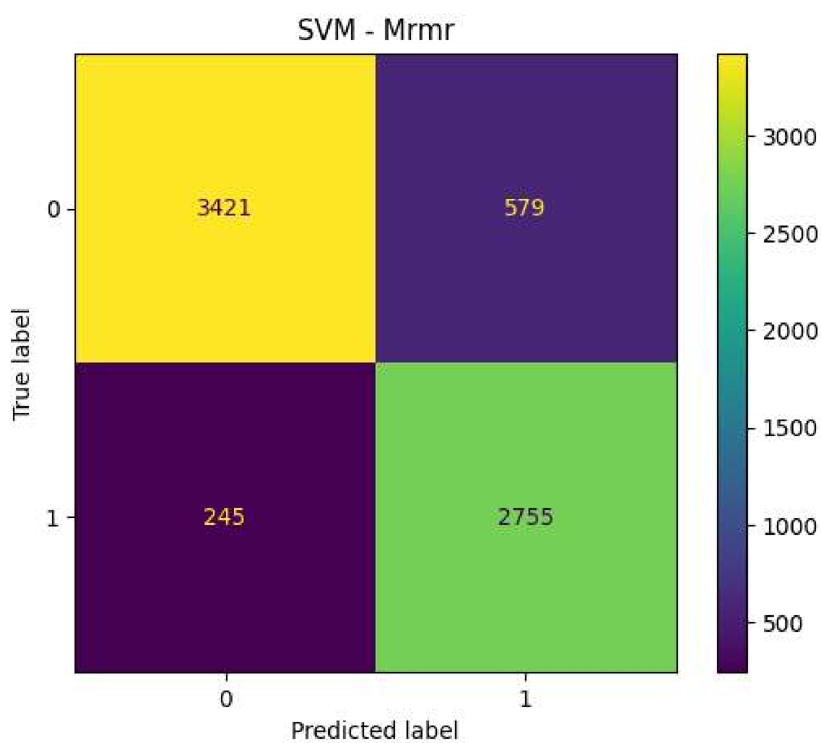


Figure 10.142. Confusion Matrix Obtained with SVM+mRMR

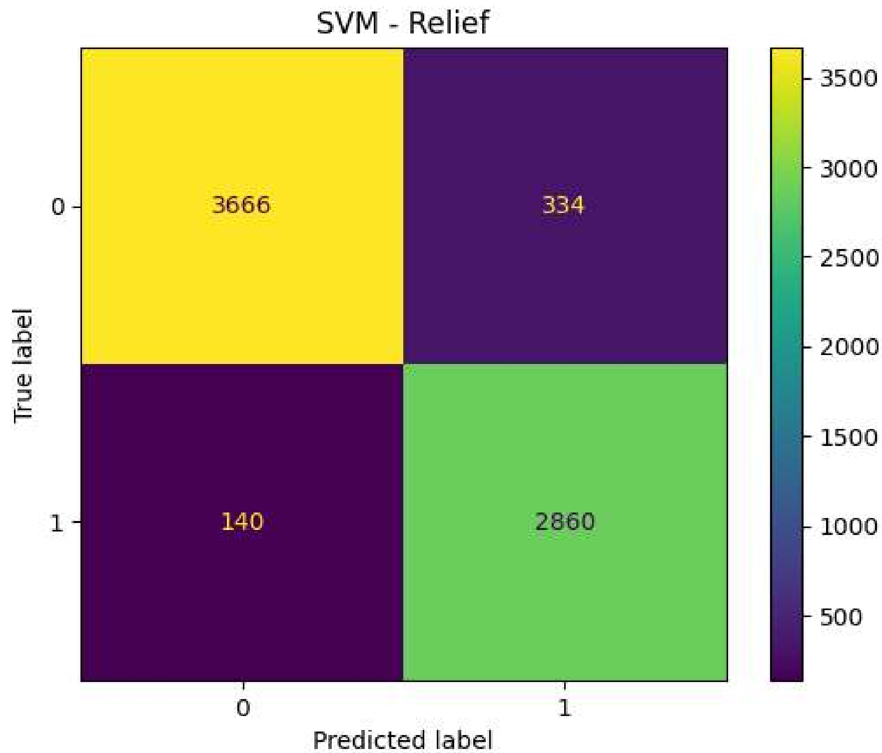


Figure 10.143. Confusion Matrix Obtained with SVM+Relief-F

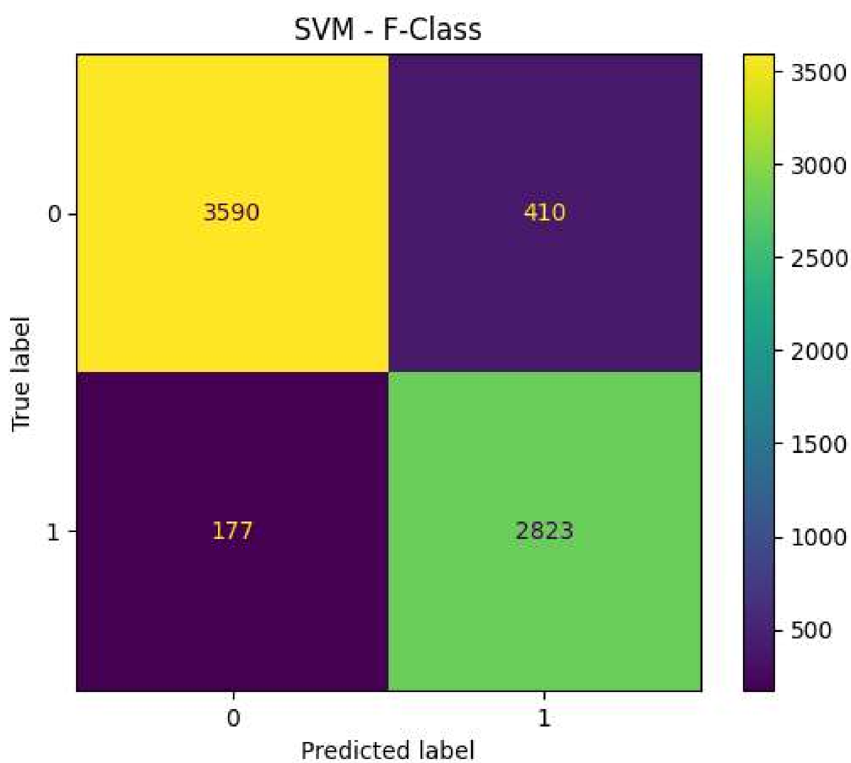


Figure 10.144. Confusion Matrix Obtained with SVM+F-Classification

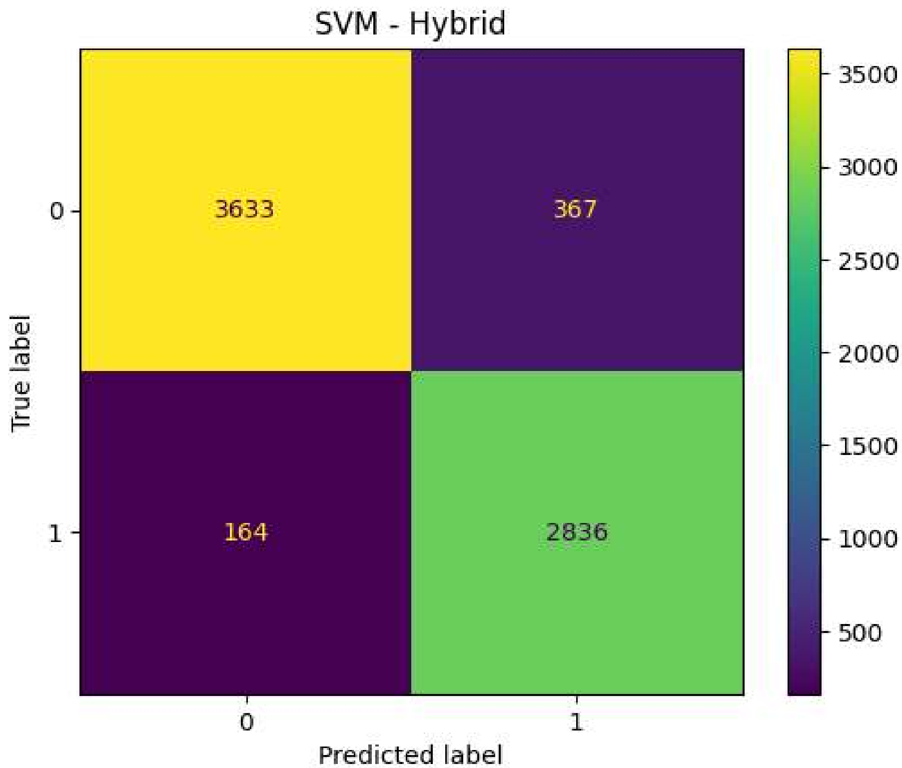


Figure 10.145. Confusion Matrix Obtained with SVM+Hybrid

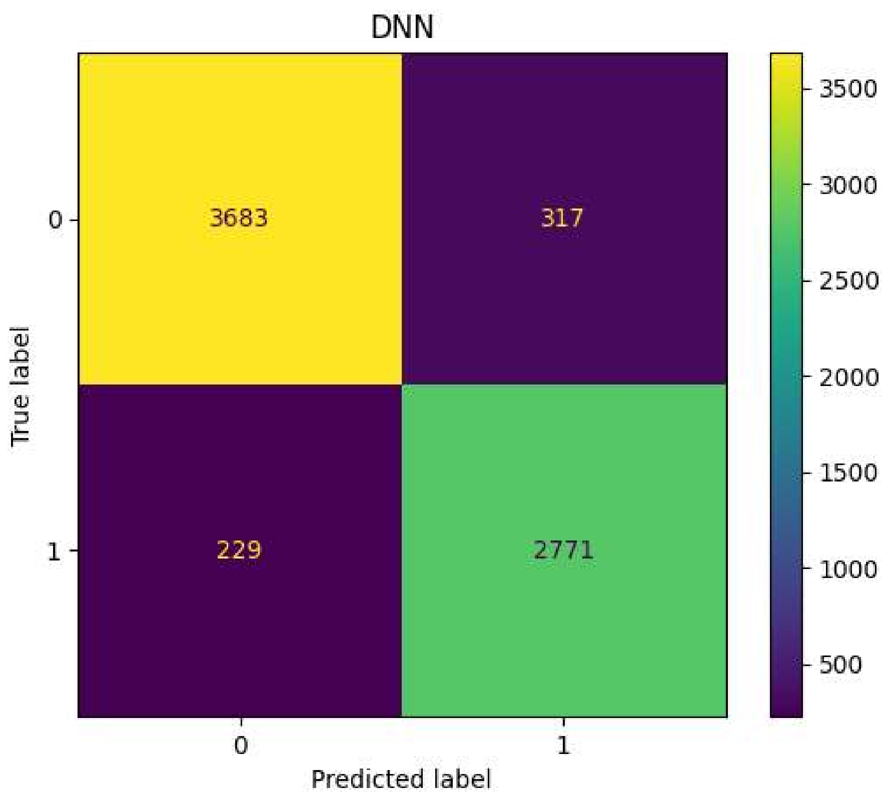


Figure 10.146. Confusion Matrix Obtained with DNN

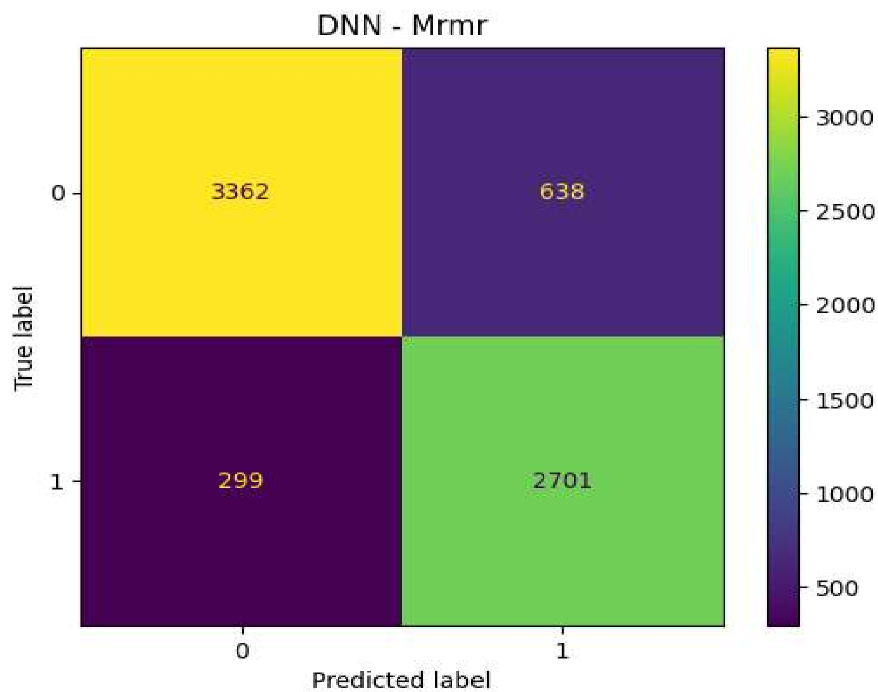


Figure 10.147. Confusion Matrix Obtained with DNN+mRMR

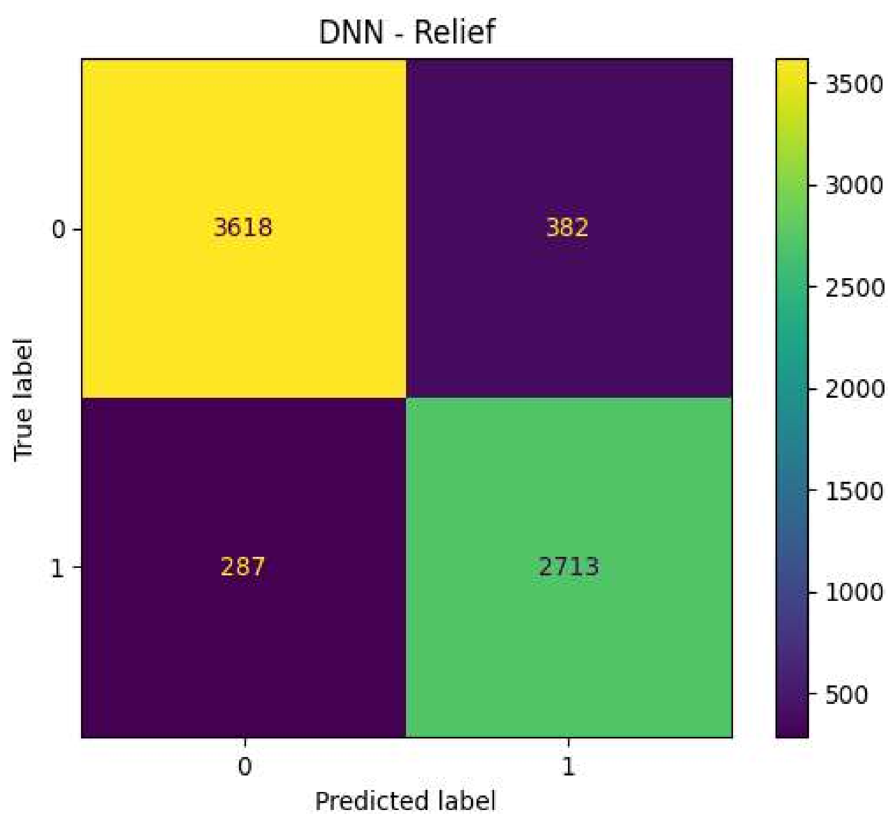


Figure 10.148. Confusion Matrix Obtained with DNN+Relief-F

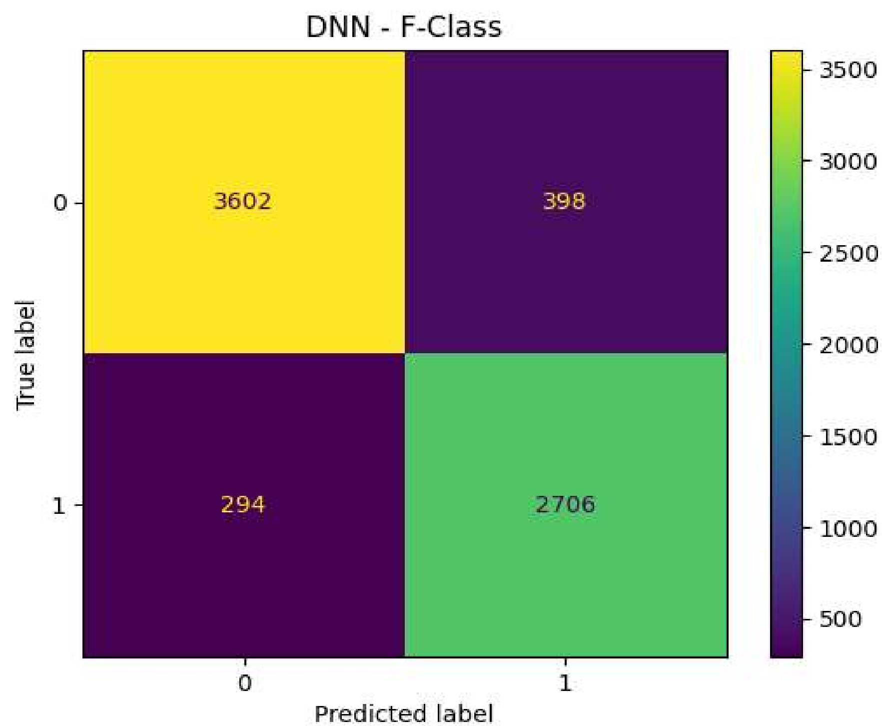


Figure 10.149. Confusion Matrix Obtained with DNN+F-Classification

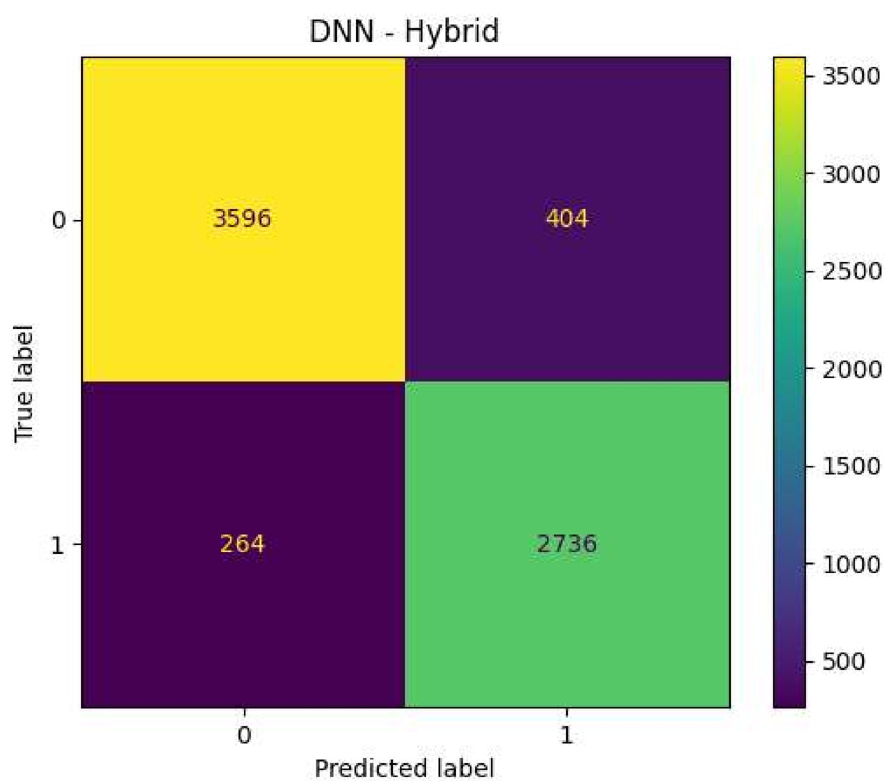


Figure 10.150. Confusion Matrix Obtained with DNN+Hybrid

10.3. Third Experimental Study: Independent Test Set Evaluation of Selected Models

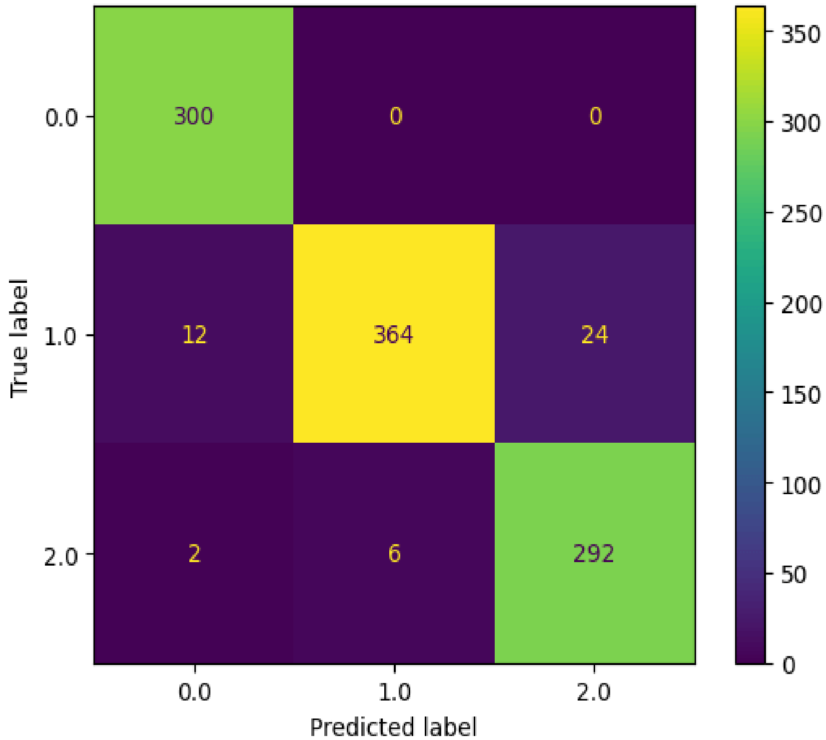


Figure 10.151. Confusion Matrix Obtained with RF

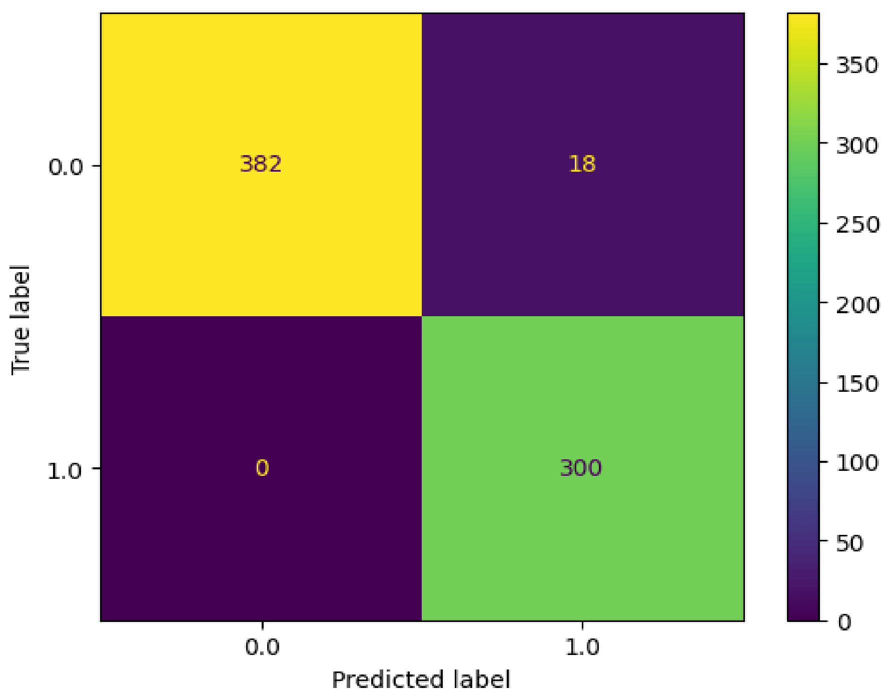


Figure 10.152. Confusion Matrix Obtained with RF+Hybrid

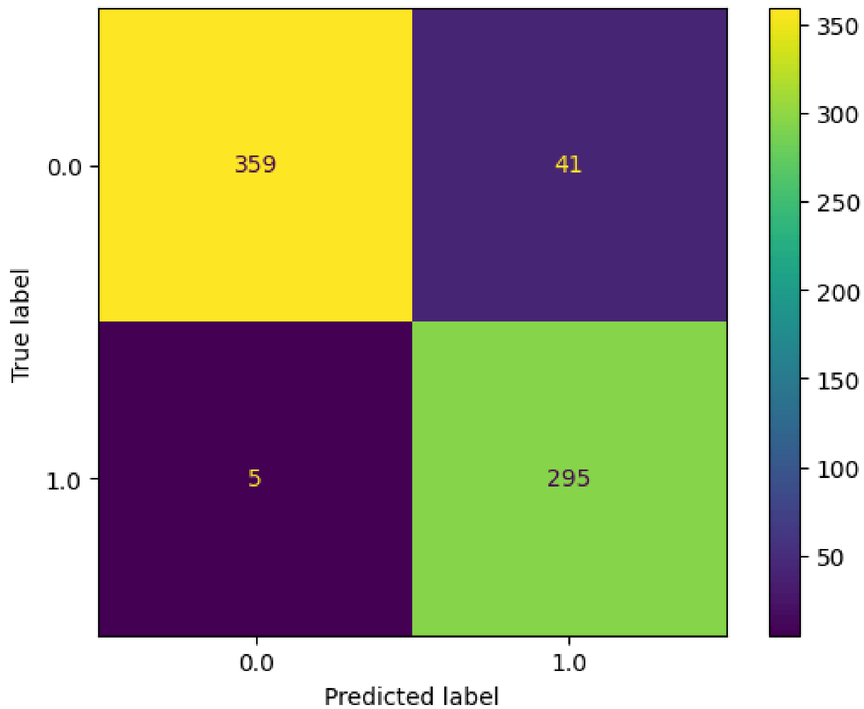


Figure 10.153. Confusion Matrix Obtained with XGBoost+F-Classification