



**MEME KANSERİ HASTALARININ SAĞ
KALIM TAHMİNİ İÇİN ÖZNİTELİK SEÇİMİ
VE SINIFLANDIRMA ALGORİTMALARININ
KARŞILAŞTIRILMASI
Yüksek Lisans Tezi**

Gizem Yağmur ÖZKAN

Eskişehir 2019

**MEME KANSERİ HASTALARININ SAĞ KALIM TAHMİNİ İÇİN ÖZNİTELİK
SEÇİMİ VE SINIFLANDIRMA ALGORİTMALARININ
KARŞILAŞTIRILMASI**

Gizem Yağmur ÖZKAN

YÜKSEK LİSANS TEZİ

**Bilgisayar Mühendisliği Anabilim Dalı
Danışman: Dr. Öğr. Üyesi Sevcan YILMAZ GÜNDÜZ**

**Eskişehir
Eskişehir Teknik Üniversitesi
Lisansüstü Eğitim Enstitüsü
Kasım 2019**

JÜRİ VE ENSTİTÜ ONAYI

Gizem Yağmur ÖZKAN'ın "MEME KANSERİ HASTALARININ SAĞ KALIM TAHMİNİ İÇİN ÖZNİTELİK SEÇİMİ VE SINIFLANDIRMA ALGORİTMALARININ KARŞILAŞTIRILMASI" başlıklı tezi 15/11/2019 tarihinde aşağıdaki jüri tarafından değerlendirilerek "Eskişehir Teknik Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca, Bilgisayar Mühendisliği Anabilim dalında Yüksek Lisans tezi olarak kabul edilmiştir.

<u>Jüri Üyeleri</u>	<u>Unvanı Adı Soyadı</u>	<u>İmza</u>
Üye (Tez Danışmanı)	: Dr. Öğr. Üyesi Sevcan YILMAZ GÜNDÜZ	
Üye	: Doç. Dr. Cihan KALELİ	
Üye	: Dr. Öğr. Üyesi Efnan ŞORA GÜNAL	

Prof. Dr. Murat TANIŞLI
Lisansüstü Eğitim Enstitüsü Müdürü

ÖZET

MEME KANSERİ HASTALARININ SAĞ KALIM TAHMİNİ İÇİN ÖZNETELİK SEÇİMİ VE SINIFLANDIRMA ALGORİTMALARININ KARŞILAŞTIRILMASI

Gizem Yağmur ÖZKAN

Bilgisayar Mühendisliği Anabilim Dalı
Eskişehir Teknik Üniversitesi, Lisansüstü Eğitim Enstitüsü, Kasım 2019

Danışman: Dr. Öğr. Üyesi Sevcan YILMAZ GÜNDÜZ

Meme kanseri günümüzde özellikle kadınlar arasında en sık görülen hastalıklardan biri haline gelmiş, bu durum da hastaların hayatta kalma tahmininin önemini artırmıştır. Bu çalışmada The Surveillance, Epidemiology, and End Results (SEER) meme kanseri veri seti kullanılarak makine öğrenme algoritmalarının sağ kalım tahmini üzerinde sağladığı başarılar karşılaştırılmış ve öznetelik seçiminin bu algoritmaların başarısı üzerinde sağladığı etki incelenmiştir. Çalışmanın ilk aşamasında veri setindeki sağ kalım tahmini için anlamlı olabilecek özneteliklerin tamamı kullanılarak 4 farklı makine öğrenme algoritması bu veri seti üzerinde Waikato Environment for Knowledge Analysis (Weka) ortamında çalıştırılmıştır. Bunlar: Naive Bayes, J48, Çok Amaçlı Evrimsel Bulanık Sınıflandırıcı (ÇAEBS), Destek Vektör Makineleridir (DVM). En başarılı algoritma olarak J48 algoritması belirlenmiştir. Çalışmanın ikinci aşamasında ise filtre ve sarmalama öznetelik seçim yöntemleri birlikte kullanılarak elde edilen veri setleri üzerinde ilk aşamada kullanılan sınıflandırma algoritmaları test edilmiş ve başarıları karşılaştırılmıştır. Daha az öznetelikle daha yüksek başarılar elde edilebildiği gözlemlenmiştir.

Anahtar Sözcükler: Meme kanseri, Öznetelik seçimi, Sağ kalım, Sınıflandırma algoritmaları

ABSTRACT

FEATURE SELECTION AND COMPARISON OF CLASSIFICATION ALGORITHMS FOR SURVIVAL OF BREAST CANCER PATIENTS

Gizem Yağmur ÖZKAN

Department of Computer Engineering
Eskişehir Technical University, Institute of Graduate Programs, November 2019

Supervisor: Asst. Prof. Dr. Sevcan YILMAZ GÜNDÜZ

Breast cancer has become one of the most common diseases, especially among women, increasing the importance of predicting survival. In this study, the success of machine learning algorithms on survival prediction was compared using the Surveillance, Epidemiology, and End Results (SEER) breast cancer data set and the effect of attribute selection on the success of these algorithms was examined. In the first stage of the study, 4 different machine learning algorithms were run on this data set in Waikato Environment for Knowledge Analysis (Weka) using all of the attributes that might be meaningful for survival estimation in the data set. These are: Naive Bayes, J48, Multiobjective Evolutionary Fuzzy Classifier (MEFC), Support Vector Machines (SVM). The most successful algorithm is J48 algorithm. In the second stage of the study, the classification algorithms used in the first stage were tested on the data sets obtained by using filter and wrapping attribute selection methods together and their successes were compared. It has been observed that higher achievements can be achieved with less attribute.

Keywords: Breast cancer, Feature selection, Survival, Classifications algorithms

ÖNSÖZ

Meme kanseri günümüzde en yaygın hastalıklardan biridir ve bu konudaki çalışmalar giderek artmakta ve önem kazanmaktadır. Hastalığın gidişatını öngörebilmek, hastalığın tedavi sürecinde oldukça önemlidir. Bu çalışma da hastaların sağ kalımını en başarılı şekilde tahmin edebilmek için makine öğrenmesi yöntemlerinin başarıları karşılaştırılmış, ayrıca öznitelik seçiminin bu yöntemlerin başarısına etkisi incelenmiştir.

Tez çalışmam sırasında bana kıymetli bilgi, birikim ve tecrübeleriyle yol gösteren değerli danışman hocam sayın Dr. Öğr. Üyesi Sevcan YILMAZ GÜNDÜZ'e teşekkür ve saygılarımı sunarım.

Öğrenim hayatım boyunca benden maddi, manevi desteklerini hiçbir zaman esirgemeyen sevgili aileme sonsuz teşekkür ve sevgilerimi sunarım. Ayrıca çalışmam boyunca bana destek olup, motivasyonumu yükselten Işık GÖKMEN'e çok teşekkür ederim.

Gizem Yağmur ÖZKAN

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmamın Eskişehir Teknik Üniversitesi tarafından kullanılan “bilimsel intihal tespit programı”yla tarandığını ve hiçbir şekilde “intihal içermediğini” beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

Gizem Yağmur ÖZKAN

İÇİNDEKİLER

Sayfa

BAŞLIK SAYFASI	i
JÜRİ VE ENSTİTÜ ONAYI.....	ii
ÖZET	iii
ABSTRACT.....	iv
ÖNSÖZ	v
ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ.....	vi
SİMGELER VE KISALTMALAR DİZİNİ.....	xiii
1. GİRİŞ	1
2. LİTERATÜR.....	3
3. ALGORİTMALAR.....	6
3.1. Öznitelik Seçim Algoritmaları.....	6
3.1.1. Bilgi kazanımı.....	6
3.1.2. Kazanım Oranı.....	7
3.1.3. Korelasyon Tabanlı Öznitelik Seçimi	8
3.1.4. Sıralı İleri Seçim	8
3.2. Sınıflandırma Algoritmaları.....	9
3.2.1. Naive Bayes.....	9
3.2.2. J48	10
3.2.3. Çok Amaçlı Evrimsel Bulanık Sınıflandırıcı.....	11
3.2.4. Destek Vektör Makineleri	13
4. VERİ SETİ VE YÖNTEM	17
4.1. Veri Seti Tanımı.....	17
4.2. Veri Seti Ön İşleme.....	17
4.2.1. Kayıp Verilerin Belirlenmesi	17
4.2.2. Sınıf Verisinin Oluşturulması.....	18
4.2.3. Özniteliklerin Belirlenmesi	18
4.2.4. Kayıp Verilerin Doldurulması.....	24
4.2.5. Normalizasyon.....	24
4.3. Değerlendirme Ölçütleri	25
4.3.1. Doğruluk.....	25
4.3.2. F-Skor	25

5. DENEYSEL ÇALIŞMALAR VE SONUÇLARI	27
5.1. Makine Öğrenmesi Algoritmaları ile Sağ Kalım Tahmini	27
5.2. Veri Seti Üzerinde Öznitelik Seçimi Çalışması	27
5.3. Bilgi Kazanımı ile Öznitelik Seçimi	28
5.4. Kazanım Oranı ile Öznitelik Seçimi	34
5.5. Korelasyon Tabanlı Öznitelik Seçimi	40
6. DEĞERLENDİRME	48
6.1. Veri Setinin Tamamı İle Yapılan Çalışma Değerlendirmeleri	48
6.2. Filtre Öznitelik Seçim Yöntemleriyle Yapılan Çalışma Değerlendirmeleri	49
6.3. Sarmalama Öznitelik Seçim Yöntemiyle Yapılan Çalışma Değerlendirmeleri	52
6.3.1. Naive Bayes Algoritması Sonuçlarının Değerlendirmesi	55
6.3.2. J48 Algoritması Sonuçlarının Değerlendirmesi	56
6.3.3. ÇAEBS Algoritması Sonuçlarının Değerlendirmesi	59
6.3.4. DVM Algoritması Sonuçlarının Değerlendirmesi	60
6.4. Genel Değerlendirme	63
7. SONUÇ	68
KAYNAKÇA	70

TABLÖLAR DİZİNİ

Sayfa

Tablo 4.1. Meme kanseri ile ilişkisi olmayan öznitelikler ve açıklamaları	18
Tablo 4.2. Uygulanamaz veri içeren öznitelikler	20
Tablo 4.3. Öznitelikler ve kayıp veri değerleri	20
Tablo 4.4. Veri setinde bulunan öznitelikler ve açıklamaları	21
Tablo 5.1. Sınıflandırma algoritmalarının karşılaştırılması	27
Tablo 5.2. Bilgi kazanımı öznitelik seçimi sonucu seçilen öznitelikler.....	28
Tablo 5.3. Bilgi Kazanımı ile seçilen özniteliklerle elde edilen doğruluk ve F-Skor değerleri.....	29
Tablo 5.4. Bilgi Kazanımı ile seçilmiş veri setinde Naive Bayes ile ileri aşamalı seçim sonucu.....	30
Tablo 5.5. Bilgi Kazanımı ile seçilmiş veri setinde J48 ile ileri aşamalı seçim sonucu	31
Tablo 5.6. Bilgi Kazanımı ile seçilmiş veri setinde ÇAEBS ile ileri aşamalı seçim sonucu	32
Tablo 5.7. Bilgi Kazanımı ile seçilmiş veri setinde DVM ile ileri aşamalı seçim sonucu	33
Tablo 5.8. Kazanım oranı öznitelik seçimi sonucu seçilen öznitelikler.....	34
Tablo 5.9. Kazanım Oranı ile seçilen özniteliklerle elde edilen doğruluk ve F-Skor değerleri.....	36
Tablo 5.10. Kazanım Oranı ile seçilmiş veri setinde Naive Bayes ile ileri aşamalı seçim	36
Tablo 5.11. Kazanım Oranı ile seçilmiş veri setinde J48 ile ileri aşamalı seçim sonucu	37
Tablo 5.12. Kazanım Oranı ile seçilmiş veri setinde ÇAEBS ile ileri aşamalı seçim sonucu.....	39
Tablo 5.13. Kazanım oranı ile seçilmiş veri setinde DVM ile ileri aşamalı seçim sonucu	39
Tablo 5.14. Korelasyon tabanlı öznitelik seçimi sonucu seçilen öznitelikler	41
Tablo 5.15. Kazanım Oranı ile seçilen özniteliklerle elde edilen doğruluk ve F-Skor değerleri.....	42

Tablo 5.16. Korelasyon ile seçilmiş veri setinde Naive Bayes ile ileri aşamalı seçim sonucu.....	42
Tablo 5.17. Korelasyon ile seçilmiş veri setinde J48 ile ileri aşamalı seçim sonucu	43
Tablo 5.18. Korelasyon ile seçilmiş veri setinde ÇAEBS ile ileri aşamalı seçim sonucu	45
Tablo 5.19. Korelasyon ile seçilmiş veri setinde DVM ile ileri aşamalı seçim sonucu	46
Tablo 6.1. Öznitelik kümelerine göre sınıflandırma algoritmalarının başarı yüzdeleri.....	51
Tablo 6.2. Geçmişte yapılan meme kanseri sağ kalım çalışmalarında kullanılan öznitelikler	53
Tablo 6.3. Naive Bayes sınıflandırma algortiması için seçilen meme kanseri öznitelikleri	55
Tablo 6.4. J48 sınıflandırma algortiması için seçilen meme kanseri öznitelikleri	56
Tablo 6.5. ÇAEBS sınıflandırma algortiması için seçilen meme kanseri öznitelikleri	59
Tablo 6.6. DVM sınıflandırma algortiması için seçilen meme kanseri öznitelikleri	60
Tablo 6.7. Sınıflandırma algoritmalarının farklı öznitelik seçim yöntemleriyle oluşturulmuş meme kanseri veri seti üzerindeki karşılaştırmalar	63

ŞEKİLLER DİZİNİ

Sayfa

Şekil 3.1. İki gruba en uzakta bulunan vektör	13
Şekil 3.2. Doğrusal olarak ayrılan verilerin sınıflandırılması	14
Şekil 3.3. Doğrusal olarak birbirinden ayırt edilemeyen verilerin sınıflandırılması ..	16
Şekil 4.1. Hata Matrisi	25
Şekil 5.1. Naive Bayes en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Bilgi Kazanımı).....	31
Şekil 5.2. J48 en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Bilgi Kazanımı).....	32
Şekil 5.3. ÇAEBS en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Bilgi Kazanımı)	33
Şekil 5.4. DVM en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Bilgi Kazanımı).....	34
Şekil 5.5. Naive Bayes en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Kazanım Oranı).....	37
Şekil 5.6. J48 en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Kazanım Oranı)	38
Şekil 5.7. ÇAEBS en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Kazanım Oranı)	39
Şekil 5.8. DVM en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Kazanım Oranı)	40
Şekil 5.9. Naive Bayes en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Korelasyon).....	43
Şekil 5.10. J48 en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Korelasyon)	45
Şekil 5.11. ÇAEBS en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Korelasyon)	46
Şekil 5.12. DVM en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Korelasyon)	47
Şekil 6.1. 73 öznitelikle elde edilmiş doğruluk ve F-Skor değerleri	48
Şekil 6.2. Bilgi Kazanımı öznitelik seçim yöntemi ile seçilen özniteliklerle sıralama algoritmaları doğruluk ve F-Skor karşılaştırması.....	49

Şekil 6.3. Kazanım Oranı öznitelik seçim yöntemi ile seçilen özniteliklerle sıralama algoritmaları doğruluk ve F-Skor karşılaştırması.....	50
Şekil 6.4. Korelasyon Tabanlı öznitelik seçim yöntemi ile seçilen özniteliklerle sıralama algoritmaları başarı karşılaştırması	51
Şekil 6.5. Filtre öznitelik seçim yöntemleri doğruluk karşılaştırması	52
Şekil 6.6. Naive Bayes algoritması doğruluk ve F-Skor karşılaştırması	56
Şekil 6.7. J48 algoritması doğruluk ve F-Skor karşılaştırması.....	58
Şekil 6.8. ÇAEBS algoritması doğruluk ve F-Skor karşılaştırması	60
Şekil 6.9. DVM algoritması doğruluk ve F-Skor karşılaştırması	62
Şekil 6.10. Farklı yöntemlerle seçilmiş öznitelikler ile sınıflandırma algoritmalarının başarı karşılaştırması.....	64
Şekil 6.11. Geçmişte yapılan çalışmalarla başarı karşılaştırması	66

SİMGELER VE KISALTMALAR DİZİNİ

α_i	: Lagrange çarpanları
μ	: Popülasyon Büyüklüğü
λ	: Oluşan Çocuk Sayısı
ξ	: N Tane Negatif Olmayan Değişken
ϕ	: Uzaylar Arası Dönüşüm Fonksiyonu
Γ	: Sınıflandırıcı
b	: Eşik değeri
AJCC	: American Joint Committee on Cancer (Amerika Kanser Komitesi)
AYA	: Adolescents And Young Adults (Ergenler ve Genç Yetişkinler)
BD	: Bölünme Değeri
BK	: Bilgi Kazanımı
CR	: Doğruluk Kriterinin Maksimum Olması
CS	: Collaborative Stage (Ortak Aşama)
ÇAEBS	: Çok Amaçlı Evrimsel Bulanık Sınıflandırıcı
ÇKA	: Çok Katmanlı Algılayıcılar
DN	: Doğru Negatif
DP	: Doğru Pozitif
DVM	: Destek Vektör Makineleri
E	: Entropi
ENORA	: Evolutionary NOn dominated sorting with RAdial slots (Radyal Yarıklerle Evrimsel Baskısız Sıralama)
EOD	: Extent of Disease (Hastalığın Kapsamı)
g_s	: Bulanık kümeler arasındaki minimum benzerlik oranı
ICCC	: International Classification of Childhood Cancer (Uluslararası Çocukluk Kanseri Sınıflaması)
IHS	: Indian Health Service (Hint Sağlık Servisi)
KO	: Kazanım Oranı
M_Γ	: Bulanık Kural
NL	: Maksimum farklı bulanık küme sayısının
NR	: Kural sayısının minimum olması
ÖA	: Öğrenme Algoritması

P	: Olasılık
r	: Öznitelik altkümeleri ile sınıf arasındaki korelasyon
R_j^Γ	: Bulanık Kural Kümesi
SEER	: The Surveillance, Epidemiology, and End Results
SİS	: Sıralı İleri Seçim
SS	: Summary Stage (Özet Aşama)
$S(\Gamma)$	: Bulanık kümeler arasındaki benzerlik oranı
T	: Tüm olası bulanık sınıflandırıcılar
TNM	: Tümör Nod Metastaz
w	: DVM normal vektörü
w_{ik}	: Çıktı Katmanındaki Ağırlıklar
W_{ji}	: Ağırlık
$\vec{w}_k$	: Kategorik Veriler
Weka	: Waikato Environment for Knowledge Analysis
$\vec{x}_k$	: Nümerik Veriler
y_k	: Sınıf Verisi
YN	: Yanlış Negatif
YP	: Yanlış Pozitif

1. GİRİŞ

Meme kanseri, meme dokusunu oluşturan hücrelerin kontrolsüz bir şekilde çoğalması durumudur. Dünya Sağlık Örgütüne göre meme kanseri kadınlar arasında en sık görülen kanser çeşididir. Her yıl 2,1 milyon kadını etkilemektedir ve kadınlar arasında en çok ölüme sebebiyet veren kanser türüdür (http-1). Ayrıca meme kanseri nadir de olsa erkeklerde de ortaya çıkabilmektedir. 2018 yılında 627.000 kadının meme kanserinden öldüğü tespit edilmiştir ki bu da kadın ölümlerinin yaklaşık %15'ini oluşturmaktadır (http-1).

Kanser sonucunda sağ kalım tahmini dünyada her zaman gündemde olan kritik konulardan biri olmuştur. Kanser teşhisi ve tedavisi, hastalığın seyrini belirleyen önemli kriterlerdir. Erken teşhis, meme kanserinin tedavi edilebilmesi ve dolayısıyla hayatta kalma oranının artırılabilmesi için büyük öneme sahiptir. Kanser yayılımı ve aşaması yine hastalığın seyrini belirlemede oldukça önemlidir. Kanser yayılması ve ilerlemesi halinde tedavi zorlaşmakta, dolayısıyla bu da hastanın sağ kalım süresini negatif yönde etkilemektedir. Meme kanseri tedavisinde uygulanan yöntemler ise şu şekildedir: cerrahi, ilaçla tedavi (kemoterapi), ışınlı tedavi (radyoterapi) ve hormon ile tedavi (hormonoterapi). Tedavilerin uygunluğu ve hayatta kalma oranının yüksek olması, sağlık uzmanları ve araştırmacılar için büyük ilgi çeken bir konu haline gelmiştir (Shukla vd., 2018, s. 199–208). Tedavinin seyrinin belirlenmesi açısından hastanın sağ kalım süresini doğru tahmin edebilme büyük önem taşımaktadır. Tıpta bilgisayar destekli uygulamalar son zamanlarda oldukça yaygın hale gelmiştir. Çünkü bilgisayar destekli uygulamalar hem büyük verilerin yorumlanması hem de aralarındaki örüntünün bulunması açısından oldukça yeteneklidir.

Geçmişten bu yana farklı yöntemler kullanılarak meme kanseri hastalarının sağ kalım durumunu belirlemek adına çalışmalar yapılmıştır. Bu çalışma, meme kanseri hastalarının sağ kalım durumunun anlaşılabilmesi için karar yardımı sağlayan algoritmaların doğruluk değerlerinin tespiti ve karşılaştırmasını ayrıca öznitelik seçiminin bu algoritmaların başarı oranları üzerindeki etkisinin incelenmesini içerir. Çalışmanın motivasyonu, daha az öznitelikle daha başarılı sağ kalım tahminleri yapmak olmuştur. Bu tür çalışmalarda kullanılan veri setinin büyüklüğü ve gerçek verilerden oluşması büyük önem taşımaktadır. Bu yüzden bu çalışmada Amerika Birleşik Devletleri'nde en kapsamlı veri kaynaklarından biri olan SEER (The Surveillance,

Epidemiology, and End Results) programında bulunan meme kanseri veri seti kullanılmıştır.



2. LİTERATÜR

Meme kanseri konusu geçmişten günümüze bilimsel çalışmalara konu olmuştur. Meme kanserinin tespiti, tekrarlama veya meme kanseri hastalarının sağ kalım süreleri hakkında önemli çalışmalar yapılmıştır. Meme kanseri tespitinde yapay sinir ağları, destek vektör makineleri, karar ağaçları vs. gibi birçok makine öğrenmesi yöntemi kullanılmıştır.

Meme kanserinde sağ kalım süresi için makine öğrenmesi alanında birçok çalışma yapılmıştır. Sağ kalım tahmini çalışmalarında birçok farklı yöntem test edilmiş ve yeni yöntemler ve öneriler ortaya konulmuştur. Tee ve Gazala (2011, s. 621-625), veri madenciliği ve bilgi keşfi teknolojilerine dayanarak geliştirilen web tabanlı prototip karar destek sistemini sunmaktadır. SEER meme kanseri veri tabanını kullanarak üç farklı veri seti elde etmişlerdir. İlk veri setinde hastaların teşhis tarihinden sonra 5 yıl hayatta kalma oranını tahmin etmektir. İkinci veri setinde hastaların teşhis tarihinden sonra hayatta kalma süresini tahmin etmektedir. Sonuncu veri setinde ise tümörün evresini öngörmek için çalışma yapmışlardır. Khan, Choi ve Shin (2008, s. 5148–5151) meme kanseri sağ kalım tahmini için melez bir sistem önermişlerdir. Bu melez sistem bulanık kurallar ve karar ağacı sistemleri birlikte kullanılarak elde edilmiştir. SEER veri tabanını kullanarak yapılan çalışma sonucu önerilen sistemin bireysel olarak kullanılan sistemlere göre daha başarılı olduğu sonucuna varmışlardır. Abdel-Zaher ve Eldeib (2016, s. 139-144), Derin İnanç Ağlarını kullanarak otomatik bir meme kanseri tanı sistemi sunmuşlardır. Fan vd. (2010), tedavi teknikleri ve meme kanseri sağ kalım süresindeki ilişkiyi SEER meme kanseri veri tabanını kullanarak veri madenciliği teknikleri ile incelemişlerdir. Daha sonra Apriori Algoritmasını uygulayarak, birliktelik kural çıkarımları elde etmişlerdir.

Genellikle sağ kalım konusunda eşik değeri 5 yıl olarak belirlenmiştir. Bununla birlikte, eşik değerini farklı kabul eden çalışmalar da bulunmaktadır. Örneğin, Boughorbel Al-Ali ve Elkum (2016, s. 1-15), 2, 5, 8 ve 11 yıl için farkı analizler yapmışlardır. Yine Burke vd. (1997, s. 857-862), 10 yıllık sağ kalımı tahmin eden bir çalışma yapmıştır. Lundin vd. (1999, s. 281-286), 5, 10 ve 15 yıllık meme kanseri sağ kalımını tahmin etmek için bir çalışmada bulunmuşlardır.

Meme kanserinde sağ kalımı tahmin ederken en iyi yöntemi belirlemek adına birçok yöntemin karşılaştırıldığı çalışmalar yapılmıştır. Delen, Walker ve Kadam (2005, s. 113-127) meme kanserinde sağ kalım tahmininde 3 farklı yöntem incelemişlerdir.

Bunlar Yapay Sinir Ağları (Çok Katmanlı Algılayıcılar), Karar Ağaçları (C5.0) ve Lojistik Regresyon yöntemleridir. Bunların arasında en başarılı yöntem Karar Ağacı Algoritması (C5.0) olarak belirlenmiştir. Ravdin ve Clark (1992, s. 285–293) hastanın sağ kalım şansını tahmin etmek için Yapay Sinir Ağı yöntemini kullanmışlar ve bunun sonucunu bir regresyon modeli ile karşılaştırmışlardır. Sonuç olarak benzer doğruluk tahminlerinde bulunduklarını görmüşlerdir. Ya-Qin, Cheng ve Zhang (2009, s. 1-4), meme kanseri sağ kalımı öngörmek için çeşitli algoritmalar uygulamışlardır. Başta dengesiz verilerde C5.0 algoritması kullanılmıştır. Daha sonra dengesiz veri engelini ortadan kaldırarak için çeşitli testler uygulamışlar ve Torbalama Algoritması ile de sağ kalımı tahmin etmişlerdir. Park vd. (2013, s. 2194-2205), meme kanseri sağ kalımını değerlendirmek için SEER veri tabanını kullanarak, Destek Vektör Makineleri, Yapay Sinir Ağları ve Yarı-denetimli Öğrenme gibi üç farklı yöntemi karşılaştırmışlardır. Veri seti, 16 temel özelliğe sahip 162.500 kayıttan oluşmaktadır. En başarılı yöntem olarak Yarı-denetimli Öğrenme belirlenmiştir. Daha sonra farklı parametre kombinasyonlarından oluşan 25 farklı deney yapmışlar ve bir model parametre değişiminden ne kadar az etkileniyorsa o kadar sağlamdır demişlerdir. Endo, Shibata ve Tanaka (2008, s. 11-16), SEER meme kanseri veri tabanını kullanarak yaşam süresinin tespiti için 7 farklı algoritmayı karşılaştırmışlardır. Bu algoritmalar, Lojistik Regresyon, Yapay Sinir Ağları, Naive Bayes, Bayes Ağları, Naive Bayes ile karar ağaçları, Karar Ağaçları (ID3) ve Karar Ağaçları (J48)'dir. Bunların arasından en başarılı yöntem Lojistik Regresyon olarak belirlenmiştir. Kate ve Nadig (2017, s. 304-311) meme kanseri sağ kalımını üç farklı makine öğrenmesi yöntemi kullanarak her bir aşamada ayrı ayrı tahmin etmek için modeller oluşturmuşlardır. Daha sonra bunları eski geleneksel modeller ile karşılaştırmışlar ve en iyi sonucun kendi aşamalarında eğitilen modeller ile alındığını gözlemlemişlerdir. Burke, Rosen ve Goodman (1995, s. 1063-1067), meme kanseri sağ kalım tahmininde çeşitli istatistiksel modellerin doğruluğunu Yapay sinir ağı modelleriyle karşılaştırmışlardır. Çalışmada sağ kalım 5 yıl olarak belirlenmiş ve çalışma bu şekilde şekillenmiştir. Kullandıkları istatistiksel modeller, Patolojik TNM Evreleme Modeli, Temel Bileşen Analizi, Sınıflandırma ve Regresyon Ağaçları ve Lojistik Regresyondur. Yapay sinir ağı modelleri ise Kaskad Korelasyonu Yapay Sinir Ağları, Eşlenik Gradyan Kökenli Yapay Sinir Ağları, Olasılıksal Yapay Sinir Ağları ve Geri Yayılım Yapay Sinir Ağlarıdır. Bu çalışmada en iyi sonuç veren modelin Geri Yayılım Yapay Sinir Ağı Modeli olduğu tespit edilmiştir. Lundin vd.

(1999, s. 281-286), 5, 10 ve 15 yıllık meme kanseri sağ kalımını tahmin etmek için Yapay Sinir Ağı ve Lojistik Regresyon Modellerini kullanmışlardır. Veri seti üzerinde yapılan çalışmada 5 yıllık sağ kalım için Yapay Sinir Ağı modeli Lojistik Regresyona göre daha yüksek başarı elde etmiştir.

Öznitelik seçimi algoritmaları ile ilgili detaylı bir çalışma yapılmamasına rağmen meme kanseri sağ kalımını belirleme çalışmalarında yararlanılmıştır. Shukla vd. (2017, s. 199-208), Veri setini kümeleme algoritmalarından yararlanarak ön işleme sürecinden geçirmişler ve Çok Katmanlı Algılayıcılar (ÇKA) algoritmasını hem ön işlemde geçirilmiş hem de geçirilmemiş veri seti üzerinde test etmişlerdir. Kümeleme algoritmaları ile ön işlemde geçirilen çalışmanın daha iyi sonuç verdiğini görmüşlerdir.

Veri setinde bulunan eksik verileri tamamlamak için yapılan çalışmalarda farklı yollar izlenmiştir. Eksik verileri verinin ortalama değeri ile dolduran çalışmalar bulunmaktadır (Rathore, Tomar ve Agarwal, s. 459–464). Delen, Walker ve Kadam (2005, s. 113-127), ise çalışmalarında bazı özniteliklerin eksik verilerini farklı özniteliklerin kombinasyonu ile doldurmuşlar, bazı özniteliklerin eksik verilerini ise ilgili kayıtları silerek ortadan kaldırmışlardır. Shukla vd. (2017, s. 199-208), ise eksik verilerin doldurulmasında farklı bir yöntem izlemişler ve özniteliklerin eksik verilerini ilgili niteliğin en sık değeri ile doldurmuşlardır.

3. ALGORİTMALAR

3.1. Öznitelik Seçim Algoritmaları

Öznitelik seçimi, yüksek sayıda öznitelikle yapılan çalışmaların getirdiği zorluğun üstesinden gelmek ve etkin bir sınıflandırma modeli oluşturabilmek için sıkça kullanılır. Öznitelik seçimi ile mevcut öznitelikler arasından etkin bir öznitelik alt kümesinin seçilmesi amaçlanır. Böylelikle, veri setinde yer alan gereksiz/ilgisiz özniteliklerin ortadan kaldırılması ile hem öğrenme algoritmasının çalışma süresi kısaltılmış hem de daha iyi bir sınıflandırma modeli oluşturulmuş olur (Dash ve Liu, 1997, s. 131-132)

Öznitelik seçim yöntemleri temel olarak 3 gruba ayrılırlar. Bunlardan ilki filtre yöntemleridir. Filtre yöntemleri, verinin genel özelliğine dayanarak öznitelik alt kümelerinin faydasını belirlerler. Belirli bir öğrenme algoritmasına dayalı değillerdir. İkinci yöntem ise sarmalama yöntemleridir. Sarmalama yöntemlerinde öznitelik seçimi belirli bir öğrenme algoritmasına dayalıdır. Algoritma sonucuna dayalı olarak sonuçlara ulaşıldığı için filtreleme yöntemlerine göre daha başarılıdır, ancak daha yavaş çalışırlar. Üçüncü yöntem ise gömülü yöntemlerdir. Gömülü yöntemlerde algoritma kendi öznitelik seçim yönteminden yararlanarak öznitelik seçimi ve sınıflandırmasını aynı anda gerçekleştirir.

Bu çalışmada, filtre yöntemlerinden Bilgi Kazanımı, Kazanım Oranı Ve Korelasyon Tabanlı Seçim; sarmalama yöntemlerinden ise Sıralı İleri Öznitelik Seçimi kullanılmıştır.

3.1.1. Bilgi kazanımı

Bilgi kazanımı kavramı, bir çalışmada kullanılan veri setindeki hangi özniteliklerin kullanımında daha yüksek verim elde edileceğini belirlerken kullanılan öznitelik seçim yöntemlerinden biridir. Bilgi kazanımı bir sınıfa ait en yüksek bilgiye sahip olan öznitelikleri belirleyebilmektedir. Bilgi kazanımının temeli entropi kavramına dayanır. Entropi, bir örneklem içerisindeki kayıtların bir öznitelik üzerindeki düzensizliğini ölçmektedir (Mitchell, 1997). Eğer bir özniteliğin tüm değerleri aynı sınıfa aitse, bu öznitelik düzensiz değildir. Entropi bilgi kazanımı ile ters orantılıdır. Entropi formülü (3.1)'de verilmiştir.

$$E(X) = \sum_{i=1}^m -p(x_i) \log_2 p(x_i) \quad (3.1)$$

Burada m değeri analiz edilen sınıfların sayısını temsil etmektedir. X sınıf değeridir ve $p(x)$ ise rastgele X değişkeni için marjinal olasılık yoğunluk işlevidir.

Gözlenen X değerleri bir A özniteliğinin değerlerine göre bölünmüşse ve A tarafından indüklenen bölümlere göre X 'in entropisi indüklenmeden önceki X entropisinden az ise o zaman X ile A arasında bir ilişki vardır denebilir. Bu durumdaki entropi değeri (3.2)'deki gibi ifade edilir.

$$E(X|A) = \sum_{\alpha \in A} p(X|A = \alpha) E(X|A = \alpha) \quad (3.2)$$

Burada $p(X|A=\alpha)$, A α değeri aldığındaki X 'in koşullu olasılığını, $E(X|A=\alpha)$ ise $A=\alpha$ değeri aldığındaki X 'in koşullu entropisini ifade eder. $E(X|A)$ ise A 'ya verilen koşul altında X 'nin koşullu entropisidir.

Yüksek entropi daha fazla düzensizlik ifade eder ve yüksek entropili öznitelik üzerinde yapılacak bir bölünme ile düzensizlik azalmış olacaktır (Var ve İnan, 2018, s.327). Buna dayanarak bilgi kazanımı denklem (3.3)'teki şekilde ifade edilebilir.

$$BK_A = E(X) - E_A(X) \quad (3.3)$$

3.1.2. Kazanım Oranı

Kazanım Oranı öznitelik seçim yöntemi çok fazla farklı değere sahip özniteliklere yönelik sapmayı azaltmak için önerilmiş bir yöntemdir. Özniteliğin her bir değerine düşen kayıt sayısı az olduğu için hesaplanan entropi değeri düşük, bilgi kazancı ise büyük çıkacaktır. Bu durumda olabilecek hataları indirmek için kazanım oranı öznitelik seçim yöntemi, bilgi kazanımı öznitelik seçim yöntemi modifiye edilerek oluşturulmuştur (Praveena Priyadarsini vd., 2011, s. 203). Kazanım oranı her bir öznitelik için bölünme bilgisi hesaplar. Bir özniteliğin bilgi kazanımı değeri, o özniteliğe ait bilgi kazanım değerinin bölünme değerine bölünmesiyle elde edilir. Bölünme değerin hesaplanması (3.4)'te verilmiştir.

$$BD_A(D) = - \sum_{j=1}^v (|D_j|/|D|) \log_2(|D_j|/|D|) \quad (3.4)$$

Verilen denklemde BD_A A özniteliğine ait bölünme değerini, D veri setini, v ise veri setinin kaç parçaya ayrıldığıdır.

Kazanım oranı (3.5) numaralı denklem ile ifade edilmektedir.

$$KO(A) = BK_A / BD_A(D) \quad (3.5)$$

Verilen denklemde $KO(A)$ A özniteliliğine ait kazanım oranını, BK_A A özniteliliğine ait bilgi kazanımını, BD_A ise A özniteliliğine ait bölünme değerini ifade eder.

3.1.3. Korelasyon Tabanlı Öznitelik Seçimi

Korelasyon Tabanlı Öznitelik Seçimi birbirinden bağımsız özniteliklerden oluşan ve sınıf ile ilişkili öznitelik altkümelerini değerlendirir. Eğer bir özniteliliğin sınıf ile alakası az ise düşük korelasyon gösterir ve bu da algoritmalarda göz ardı edilebilir. Korelasyon Tabanlı Öznitelik Seçimi, öznitelikler arasındaki etkileşimi bulmayı hedef alır ve her bir özniteliliğin bireysel tahmin yeteneğini gereksiz fazlalık derecesini göz önünde bulundurarak, bir öznitelik altkümelerinin değerini ölçer (Karegowda vd., 2010, s. 272-273). Korelasyon katsayıları öznitelik altkümeleri ve sınıf arasındaki korelasyonu tahmin etmek için kullanılır.

Sınıf değişkeni ile öznitelik altkümeleri arasındaki ilişkiyi ifade eden denklem (3.6)'da verilmiştir.

$$r_{zc} = \frac{k\bar{r}_{zi}}{\sqrt{k + k(k-1)\bar{r}_{ii}}} \quad (3.6)$$

Burada r_{zc} öznitelik altkümeleri ile sınıf değişkeni arasındaki ilişkiyi ifade etmek üzere, k öznitelik altkümelerinin sayısı, r_{zi} altküme öznitelikleri ve sınıf değişkeni arasındaki ortalama korelasyonu, r_{ii} ise altküme öznitelikleri arasındaki interkorelasyonu ifade eder.

3.1.4. Sıralı İleri Seçim

Sarmalama öznitelik seçim yöntemlerinden biri olan Sıralı İleri Seçim eğitim için en uygun öznitelik setini bulmayı hedefleyen bir öznitelik seçim yöntemidir. Sarmalama yöntemlerine öznitelik seçimi belirli bir öğrenme algoritmasına dayalı yapıldığı için, Sıralı İleri Seçim de bir öğrenme algoritması baz alınarak yapılır. Öznitelik seçme işlemi azdan çoğa doğru yapıldığı için ileri olarak adlandırılmıştır. (Kaynar vd., 2018, s.181) Öznitelik seçimine boş bir öznitelik kümesi ile başlanır. İlgili öğrenme algoritması her bir öznitelik üzerinde tekrarlanır ve durma kriteri sağlanıncaya kadar devam eder. Durma kriteri belirli bir öznitelik sayısına ulaşılması, belirli sayıda iterasyona ulaşılması, herhangi bir özelliğin eklenip çıkarılmasının daha iyi bir sonuç vermemesi, yeterince iyi bir öznitelik kümesine ulaşılması gibi kriterler olabilir. Bu

çalışmada herhangi bir özelliğin eklenip çıkarılmasının daha iyi bir sonuç vermemesi durma kriteri olarak belirlenmiştir.

Sıralı İleri Seçimin kaba kodu aşağıdaki gibidir:

Öznitelik Kümesi: $F = \{f_1, f_2, f_3, \dots, f_d\}$

Seçilen Öznitelik Alt Kümesi: S

Öğrenme Algoritması: $\ddot{O}A$

1. $S = \emptyset$ ve $\ddot{O}A_{acc}(S)=0$
2. $f_k = \max(\ddot{O}A_{acc}(S \cup f_i)) \quad i = 0 \text{’dan } d \text{’ye kadar}$
 $S = S \cup f_k$
 $F = F - f_k$
3. $\ddot{O}A_{acc}(S(k)) < \ddot{O}A_{acc}(S(k-1))$ olana kadar 2. madde tekrarlanır.

3.2. Sınıflandırma Algoritmaları

3.2.1. Naive Bayes

Naive Bayes istatistiksel bir sınıflandırma yöntemidir. Olasılık hesaplama yöntemlerini kullanarak, bir durumun hangi kategoriye ait olduğunu belirlemek amacı ile kullanılır. Naive Bayes yönteminde koşullu olasılık hesaplamaları temel alınmıştır. Bağımsız değişkenlerin, bağımlı değişkenler üzerindeki etkileri kullanılarak, yeni bir verinin sınıflandırılma işlemi gerçekleştirilir (Atak, 2014).

Bu yöntemde özniteliklerin birbirinden bağımsız olduğu kabul edilir. Fakat değişkenlerin bağımsızlığının garanti edilemediği durumlarda bile iyi sonuçlar verdiği gözlenmiştir (Russell ve Norvig, 2003).

Naive Bayes algoritmasının çalışma süresi, veri seti dinamik olarak değişmiyor ise oldukça kısadır. Fakat veri setinin dinamik olarak değiştiği durumlarda eğitim sürecini tekrarlamak gerekeceğinden daha uzun zaman alacaktır. Öznitelikler arasındaki bağımsızlık göz önünde bulundurularak düşünüldüğünde, Naive Bayes diğer modellere göre daha iyi performans gösterir ve daha az egzersiz verisine ihtiyaç duyar.

Test edilen veri setinde bulunan özniteliklerden biri, egzersiz veri setinde bulunmuyorsa, olasılık değeri model tarafından 0 atanır. Bu durumda tahmin yapmak imkânsız hale gelir. Bu durumu çözmek için bazı yöntemler kullanılmıştır. Laplace dönüşümü bu yöntemlerden biridir.

Naive Bayes formülü (3.7) numaralı denklemde verilmiştir.

$$P(C_i|f) = \frac{P(f|C_i) + P(C_i)}{P(f)} \quad (3.7)$$

C hedef sınıf değerini, f ise öznitelikleri temsil etmek üzere, $P(f|X)$ hedefin niteliğe göre koşullu olasılığını, $P(f|C)$ niteliğin hedefe bağlı koşullu olasılığını, $P(f)$ niteliğin önceki olasılık değerini, $P(C)$ ise hedefin önceki olasılık değerini gösterir.

3.2.2. J48

J48 algoritması, karar ağacı algoritmalarından biridir ve C4.5 algoritmasının açık kaynaklı versiyonudur. Bu yöntem, öğrenme aşamasında, normalize edilmiş, bilgi kazancı kavramını kullanır (Demirhan, 2015).

Karar ağaçlarında en tepede bulunan kök öznitelik, sınıf değerini belirlemede en etkili olan özniteliktir. Köke yerleştirilen öznitelik uygun dallara ayrılır. Ayrılan her dal, en baştaki kök niteliğin seçilmesine benzer bir yöntemle seçilen karar düğümlerine bağlanır. Bu durum bölünecek bir öznitelik kalmayıncaya kadar devam eder ve son olarak yapraklar oluşur.

J48 algoritması kök ve karar düğümlerinin belirlenmesinde bilgi kazanımı yöntemini kullanmaktadır. Bunun için ise öncelikle entropi hesaplanır. Entropi formülü (3.8) numaralı denklemde verilmiştir.

$$E(S) = - \sum_{x \in X} P(x) \log_2 P(x) \quad (3.8)$$

Burada $E(S)$ x özniteliğinin entropi değerini, $P(x)$ ise x özniteliğinde bulunan her bir değer olasılığını ifade etmektedir.

Bilgi kazanımı formülü ise (3.9) numaralı denklemde verilmiştir.

$$BK = E(S) - \sum_{t \in T} P(t)E(t) \quad (3.9)$$

Burada BK bilgi kazanımını, $P(t)$ bilgi kazanımı hesaplanacak niteliğe ait bir altkümenin olasılık değerini, $E(t)$ ise bilgi kazanımı hesaplanacak niteliğe ait bir altkümenin entropisini ifade etmektedir.

Kazanımı en yüksek öznitelik kök olarak seçilir. Bu yöntem karar düğümleri için de uygulanarak ağaç yapısı oluşturulur.

J48 algoritması, anlaşılması ve uygulanması kolay bir sınıflandırma yöntemidir. Algoritmanın çalışma zamanı karmaşıklığı, ağaç derinliği ile bağlantılıdır. Ağaç

derinliği ise örnek sayısına bağlıdır. Bu nedenle ağaç boyutu, örnek sayısı arttıkça doğrusal olarak artar (Bhargava vd., 2013, s. 1114-1119).

3.2.3. Çok Amaçlı Evrimsel Bulanık Sınıflandırıcı

Girdileri çıktılara eşleyen bulanık eğer-ise kuralları kümesi bulanık model olarak bilinir (Dwivedi ve Tripath, 2015, s. 1103). Bulanık model temel olarak 3 bileşenden oluşur: bulanık kurallar setinden oluşan bir kural tabanı, bulanık kuralların üyelik fonksiyonlarını belirtecek bir veri tabanı ya da sözlük, uygulama sürecini verilen belirli bir koşulda veya kural ile yerine getirmek için nedensel bir metodoloji (Yen ve Wang, 1999, s. 14). Bulanık modellerin geliştirilmesinde doğruluk ve yorumlanabilirlik iki önemli kriterdir. Fakat sistemlerin verdiği sonuç doğruluğu arttıkça yorumlanabilirliği azalmakta, yorumlanabilirlik arttıkça doğruluk azalmaktadır. Bulanık sistemlerin iki önemli kriteri yorumlanabilirlik ve doğruluk, iki nesnel optimizasyon problemi olarak tanımlanabilir ve bu tür bir problemler Çok Amaçlı Evrimsel Algoritmalar tarafından çözülmüştür.

Çok Amaçlı Evrimsel Bulanık Sınıflandırıcı, Radyal Yarıklarla Evrimsel Baskısız Sıralama (ENORA) çok amaçlı evrimsel algoritmayı kullanarak bulanık kural tabanlı bir sınıflandırıcı oluşturur (http-2). Çok Amaçlı Evrimsel Algoritma ile iki ana hedef olan doğruluk ve yorumlanabilirlik için optimizasyon yapılır. Amaç doğruluğu en üst seviyeye çıkarıp bulanık kural sayısını en aza indirmektir.

N giriş verisini içeren bir veri kümesini göz önüne alalım $\Delta = \{d_1, \dots, d_N\}$. Her bir girdi d_k , $\vec{x}_k$ vektörü nümerik verileri, $\vec{w}_k$ vektörü kategorik verileri, y_k ise sınıf verisini ifade etmek üzere $d_k = (\vec{x}_k, \vec{w}_k, y_k)$ formunda olsun. M_Γ , $R_1^\Gamma, \dots, R_{M_\Gamma}^\Gamma$ bulanık kurallar tarafından oluşturulmuş bir $\Gamma \in T$ sınıflandırıcısını ele alalım. Burada T tüm olası bulanık sınıflandırıcılardır. Ayrıca her bir çıktı z sınıfı için en az bir tane kural olduğunu varsayalım. Her bir bulanık kural R_j^Γ ($j = 1, \dots, M_\Gamma$) nümerik veriler için p tane $\tilde{A}_{ij}^\Gamma$ ($i = 1, \dots, p$) bulanık set, kategorik verileri için q tane sayısal değer $B_{ij}^\Gamma \in \{1, \dots, v_i\}$ çıktı verisi için bir tane $C_j^\Gamma \in \{1, \dots, z\}$ sayısal değer içerir. $R_1^\Gamma, \dots, R_{M_\Gamma}^\Gamma$ kural tabanı olarak adlandırılır. Her bir $\tilde{A}_{ij}^\Gamma$ ($i = 1, \dots, p$) ($j = 1, \dots, M_\Gamma$) bulanık set $\mu_{\tilde{A}_{ij}^\Gamma} : x_i \rightarrow [0,1]$ $x_i = [l_i, u_i] \subset \mathbb{R}$ şeklinde ifade edilen üyelik fonksiyonu ile tanımlanabilir. Üyelik fonksiyonları kümesi ise veri tabanı olarak adlandırılır.

Bulanık bir kural (3.10) numaralı denklemdeki gibi ifade edilir:

$$\begin{aligned}
R_j^\Gamma: & \text{Eğer } x_1 = \tilde{A}_{1j}^\Gamma \text{ ve ... ve } x_p = \tilde{A}_{pj}^\Gamma \text{ ve} \\
& w_i = B_{1j}^\Gamma \text{ ve ... ve } w_q = B_{qj}^\Gamma \\
& \text{ise } y = C_j^{\Gamma'} \text{ dir}
\end{aligned} \tag{3.10}$$

Bulanık bir sınıflandırıcı, veri tabanı ve kural tabanı birleşiminden oluşur ve bilgi tabanı olarak adlandırılır (Jiménez vd., 2013, s. 202). Ayrıca, kural tabanlı sınıflandırıcı, spesifik bir girdi için sınıf etiketi atamak amacıyla kuraldaki bilgileri kullanan nedensellik metodu ile karakterize edilir (Jiménez vd., 2013, s. 202). Optimizasyon işlemi (3.11) numaralı denklemde ifade edilmiştir.

$$\begin{aligned}
& CR(\Gamma, \Delta) \text{ maksimize edilir} \\
& NR(\Gamma) \text{ minimize edilir} \\
& M_\Gamma \geq M_{\min} \\
& M_\Gamma \leq M_{\max} \\
& S(\Gamma) \leq g_s \\
& NL(\Gamma) \leq L_{\max}
\end{aligned} \tag{3.11}$$

$\Delta = \{d_1, \dots, d_N\}$ N girişli bir veri kümesi ve her bir d_k , ($k = 1, \dots, N$) girdisi $d_k = (\vec{x}_k, \vec{w}_k, y_k)$ formunda olduğu kabul edilir. $\vec{x}_k$ gerçek input girdilerini, $\vec{w}_k$ kategorik girdileri y_k ise çıktı verisini ifade etmek üzere; $CR(\Gamma, \Delta)$ 'nin maksimize edilmesi doğruluk kriterinin maksimum olmasını ifade eder. $NR(\Gamma)$ 'nin minimize edilmesi kuralı ise kural sayısının minimize edilmesini ifade eder. $M_\Gamma \geq M_{\min}$ ve $M_\Gamma \leq M_{\max}$ kural sayısının belirlenmiş minimum ve maksimum kural sayısı arasında olacağını garanti eder. $S(\Gamma) \leq g_s$ kuralı bulanık kümeler arasındaki minimum benzerlik oranının g_s 'den ($0 < g_s \leq 1$) küçük eşit olduğu kısıtını ifade eder. Bulanık sınıflandırıcılar arasındaki benzerlik değeri $S(\Gamma)$ (3.12a)'deki şekilde hesaplanır.

$$\begin{aligned}
S(\Gamma) &= \max_j \text{Sim}(\tilde{A}_{ij}^\Gamma, \tilde{A}_{ik}^\Gamma) \\
& j, k = 1, \dots, M_\Gamma \\
& i = 1, \dots, p
\end{aligned} \tag{3.12a}$$

$$\begin{aligned}
& \tilde{A}_{ij}^\Gamma \neq \tilde{A}_{ik}^\Gamma \\
\text{Sim}(\tilde{A}, \tilde{B}) &= \max \left\{ \frac{|\tilde{A} \cap \tilde{B}|}{|\tilde{A}|}, \frac{|\tilde{A} \cap \tilde{B}|}{|\tilde{B}|} \right\}
\end{aligned} \tag{3.12b}$$

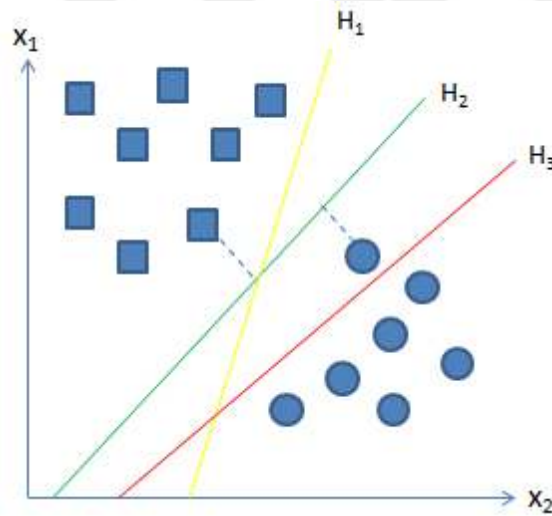
$NL(\Gamma) \leq L_{\max}$ ile bir Γ sınıflandırıcısı için hesaplanmış maksimum farklı bulanık küme sayısının belirtilen $L_{\max}$ değerinden küçük olmasını ifade eder.

Bu çalışmada kullanılan ENORA, hayatta kalma için $(\mu + \lambda)$ kullanan Pareto tabanlı elitist bir algoritmadır. Burada μ popülasyon büyüklüğü, λ ise yaratılan çocuk sayısını ifade eder. $(\mu + \lambda)$ tekniği hayatta en iyi çocuk ve ebeveynlerin kalmasını sağlar, bu yüzden elitist bir tekniktir (Jiménez vd., 2013, p. 208).

3.2.4. Destek Vektör Makineleri

Destek Vektör Makineleri (DVM) sınıflandırmada kullanılan etkili algoritmalarından biridir. DVM’de amaç, veri noktalarını mümkün olduğu kadar iyi sınıflandıran ve yine mümkün olduğu kadar iki sınıf noktaya ayıran optimum ayırıcı düzlemin bulunmasıdır (Güraksın, 2015). Sınıflandırma için iki grup arasında bir sınır çizilerek iki grup birbirinden ayrılır. Çizilen sınır her iki grubun üyelerine de en uzak yerdedir.

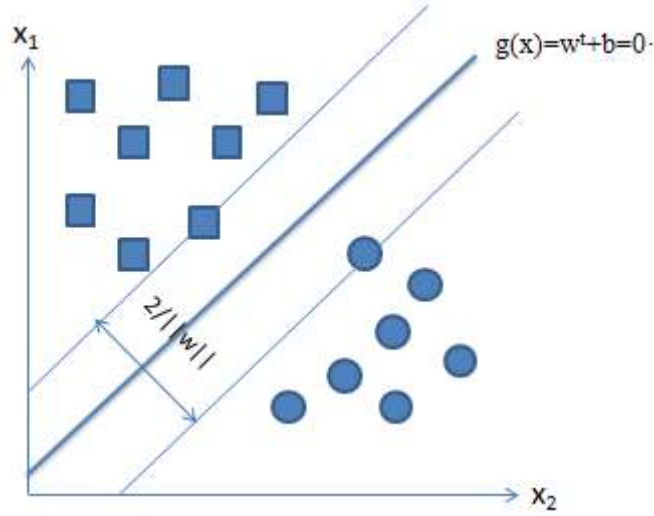
Şekil 3.1’te iki gruba da en uzakta bulunan vektörün seçilmesi ile ilgili bir görsel bulunmaktadır.



Şekil 3.1. İki gruba en uzakta bulunan vektör

DVM ile sınıflandırma, doğrusal veya doğrusal olmayan fonksiyonlar yardımıyla yapılır. Doğrusal destek vektör makineleri, doğrusal olarak birbirinden ayrılabilen verilerin sınıflandırılmasında kullanılır.

Şekil 3.2’de doğrusal olarak ayrılabilen verilerin sınıflandırılmasına ait bir görsel verilmiştir.



Şekil 3.2. Doğrusal olarak ayrılan verilerin sınıflandırılması

$(x_1, \dots, x_n)$ veri seti ve $y_i \in (-1, +1)$ sınıf etiketini ifade etmek üzere $w^T x + b = 0$ iki sınıfı ayıran hiper düzlemi ifade eder. Burada w normal vektörü, b ise eşik değeridir. En yakın nokta veya noktaların ayırıcı düzleme uzaklığı $1/\|w\|$ ile hesaplanır. Toplam uzunluk ise $2/\|w\|$ olarak ifade edilir. Sınıfların hiper düzleme göre belirlenmesi aşağıdaki şekilde ifade edilir.

Ayırıcı düzlem ile en yakın veri noktası arasındaki mesafe (sınır) maksimum ise ayırıcı düzlem optimum ayırıcı düzlemdir. Ayırıcı düzlem ile bir x verisi arasındaki mesafe $g(x)/\|w\|$ formülü ile bulunur (Çomak, 2008). DVM’de amaç bu mesafeyi maksimize etmektir. Çünkü mesafe ne kadar fazla ise sınıflandırma işlemi o kadar başarılı olacaktır. Bu mesafeyi maksimum değere çıkartan en uygun hiper düzlem (3.13)’teki eşitlik ile bulunur:

$$\text{Minimize } \frac{1}{2} \|w\|^2 \quad (3.13)$$

$$\text{Bağlı olarak } y_i(w^T x_i + b) \geq 1 \quad \forall i$$

Bu durum (3.14)’teki şekilde ikinci dereceden bir optimizasyon problemine dönüştürülebilir. α_i Lagrange çarpanlarını ifade etmek üzere:

$$\begin{aligned} \text{Maks: } L(\alpha) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1, j=2}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \\ \text{Bağlı olarak } \alpha_i &\geq 0, \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned} \quad (3.14)$$

Her sınıflandırma işleminde sınıf verileri doğrusal olarak birbirinden ayıramayabilir. Bu durumda doğrusal DVM’de verilen fonksiyonlara (3.15)’daki şekilde N tane negatif olmayan değişken (ξ) eklenerek genelleştirme yapılabilir. Bu şekilde az sayıda yanlış sınıflandırılacak veriler hesaba katılır.

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i \quad i = 1, 2, \dots, N \quad (3.15)$$

Bu durumda optimum ayırıcı düzlemi bulmadaki eşitlikler ise (3.16)’deki şekilde ifade edilir:

$$\begin{aligned} \text{Min: } & -\frac{1}{2} w \cdot w + C \sum \xi_i \\ \text{Bağlı olarak } & y_i(w \cdot x_i + b) \geq 1 - \xi_i \quad i = 1, 2, \dots, N \quad \xi \geq 0 \end{aligned} \quad (3.16)$$

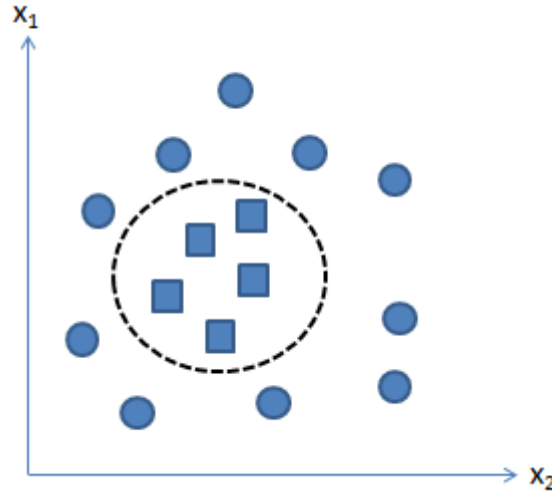
Burada C değeri kullanıcı tarafından seçilen ve değeri 0’dan büyük olan ($C > 0$) bir sabittir ve ceza parametresi olarak adlandırılır.

Doğrusal olarak ayırt edilemeyen sınıflar için doğrusal olmayan destek vektör makineleri devreye girer. Verilerin doğrusal olarak ayırt edilemediği durumlarda veriler daha yüksek boyutlu farklı bir uzaya taşınır. Bu durumda uzaylar arası dönüşüm yapabilen dönüşüm fonksiyonları kullanılır ($\phi(x)$).

$x \in R_n$ gözlem vektörünü daha yüksek dereceden bir uzayda z vektörlerine dönüştürerek, bu yeni uzayda doğrusal sınıflandırıcıları elde etmek mümkün olabilir (Güraksın, 2015). Bu z vektörünün yer aldığı özellik uzayını F ile gösterirsek, ϕ ifadesini, $R_n \rightarrow R_f$ eşlemesini yapmak üzere $z = \phi(x)$ biçiminde (3.17)’deki gibi ifade edebiliriz (Güraksın, 2015).

$$x \in R^n \rightarrow z(x) = [a_1, \phi_1(x), \dots, a_n, \phi_n(x)]^T \in R^F \quad (3.17)$$

Şekil 3.3’de doğrusal olarak birbirinden ayırt edilemeyen verilerin sınıflandırılmasına ait bir görsel verilmiştir.



Şekil 3.3. Doğrusal olarak birbirinden ayırt edilemeyen verilerin sınıflandırılması

Burada $\phi_1(x), \dots, \phi_n(x)$ fonksiyonları yerine tek bir fonksiyon kullanılabilir ve buna da çekirdek fonksiyonu denir. Çekirdek fonksiyonları sınıflandırmanın başarısını belirlemede önemlidir. (3.18)'te bazı çekirdek fonksiyonlarına yer verilmiştir.

$$\begin{aligned}
 \text{Doğrusal: } K(x_i, x_j) &= x_i^T x_j \\
 \text{Polinom: } K(x_i, x_j) &= (1 + x_i^T x_j)^p \\
 \text{Gauss: } K(x_i, x_j) &= e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}
 \end{aligned} \tag{3.18}$$

DVM iki sınıf içeren verilerin sınıflandırılması üzerine kurulmuştur. Çoklu sınıf içeren verilerin sınıflandırılmasında birkaç farklı yöntem benimsenmiştir. Bu yöntemlerden biri bire karşı kalanlar yaklaşımıdır ve bir sınıf verisinin diğer kalanlardan ayırt edilmesini amaçlar. Bu durumda k adet sınıf için k adet DVM modeli oluşturulması gerekir. Bir diğer yöntem bire karşı bir yaklaşımdır. Bu yaklaşımda ise k adet sınıf için $k*(k_1)/2$ sayıda sınıflandırıcı oluşturulur. Başka bir yöntem ise Hata Düzeltme Çıkış Kodları prensibidir. Bu durumda k sınıflı sınıflandırma problemi m sayıda ikili sınıflandırma problemine dönüştürülür. Sonra m sayıda çıkış değeri hesaplanarak daha sonra bunlar birleştirilir.

4. VERİ SETİ VE YÖNTEM

4.1. Veri Seti Tanımı

Bu çalışmada, The Surveillance, Epidemiology, and End Results (SEER) Programında bulunan meme kanseri veri seti kullanılmıştır.

Ulusal Kanser Enstitüsüne ait olan SEER Programı, Amerika Birleşik Devletleri'ndeki kanser insidansı ve hayatta kalma konusunda güvenilir bir bilgi kaynağıdır ([http-3](http://3)). SEER Programı, hastaya ait verileri rutin olarak toplar ve bunları veri tabanında tutar. SEER Programında farklı yıllarda yayımlanan farklı veri tabanları bulunmaktadır. Bu çalışmada, “Incidence – Seer 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2018 Sub 1975-2016 (varying)” veri tabanı kullanılmıştır. Veri tabanında meme, kolon ve rektum, kadın genital, erkek genital, lenfoma ve lösemi, solunum, üriner sistem, diğer sindirim sistemi kanserlerine ve diğer kanser tiplerine ait veriler bulunmaktadır. Kullanılan araştırma veri tabanında toplamda 10,440,123 kayıt bulunmaktadır, bu kayıtların 1,695,772'si meme kanserine aittir.

4.2. Veri Seti Ön İşleme

Veri seti, çalışmada ele alınan problemi inceleyebilmek adına bir ön işleme sürecinden geçirilmiştir. Hasta eğer 5 yıldan az yaşamış, hastalık tanısı ve ölüm sebebi meme kanseri veya hastalık tanısı meme kanseri ve hala hayatta olan hastalar veri setine dâhil edilmiştir. Bu durumda veri seti, hastalık tanısı sadece meme kanseri olan ve yaşam süresi bilinen hastaları içermektedir. Ayrıca veri setinde bulunan özniteliklerden bazılarının kodlama sistemi 2004 yılından sonra değiştiği için, çalışmaya sadece 2004 yılından sonra meme kanseri teşhisi konmuş hastalar dâhil edilmiştir.

Hatalı olduğu düşünülen kayıtlar da veri setinden silindikten sonra sonuç olarak, 202 tane öznitelik içeren 370.153 kayıt elde edilmiştir.

4.2.1. Kayıp Verilerin Belirlenmesi

SEER programında kayıp veriler her bir öznitelik için ayrı bir kod ile tanımlanmıştır. Bu kod tanımları veri tabanı açıklamalarında her bir öznitelik için ayrı ayrı belirtilmiştir. Weka tarafından algılanan kayıp veri değeri “NaN” olduğu için, açıklamalarda belirtilen değerlere istinaden her bir öznitelik için kayıp veriler Weka arayüzünden “NaN” olarak set edilmiştir.

4.2.2. Sınıf Verisinin Oluşturulması

Sınıf niteliğinin oluşturulmasında daha önceki yapılan çalışmalar baz alınarak sağ kalan ve ölen olarak sınıflandırılmıştır. Yine önceki çalışmalar baz alınarak sağ kalım için eşik değeri 5 yıl olarak belirlenmiştir. Sınıf niteliğinin belirlenmesinde “Sağ Kalım Ayı (Survival Month)”, “Ölüm Sebebi Tanısı (Cause Of Death To Site Recode) ve “Hayati Durum (Vital Status Recode)” öznitelikleri kullanılmıştır.

Sınıf verisi;

Eğer “Sağ Kalım Ayı” <60 ve “Ölüm Sebebi Tanısı” = “Meme Kanseri” ise

Ölen (Değer = 0)

Eğer “Sağ Kalım Ayı” >=60 ve “Hayati Durum” = “Hayatta” ise

Sağ Kalan (Değer = 1)

şeklinde belirlenmiştir.

4.2.3. Özniteliklerin Belirlenmesi

SEER meme kanseri veri setinde toplam 202 öznitelik bulunmaktadır. Bu özniteliklerden bazıları tek bir değer içermekte bazıları ise tamamına yakını kayıp veri içermektedir.

Öznitelikler belirlenirken öncelikle tek değer içeren öznitelikler algoritmaların eğitilmesinde etkisi olmayacağından çıkarılmıştır.

EK- 1’de tek değer içerdiği için çıkarılan öznitelikler verilmiştir.

Tek bir değer içerek öznitelikleri çıkardıktan sonra geriye 113 öznitelik kalmıştır. Kalan öznitelikler arasından meme kanseri ile alakasız ve sağ kalımı etkilemeyecek olan öznitelikler belirlenmiş ve bu öznitelikler de veri setinden çıkarılmıştır.

Tablo 4.1 ‘de bu özniteliklere yer verilmiştir.

Tablo 4.1. Meme kanseri ile ilişkisi olmayan öznitelikler ve açıklamaları

Öznitelik Adı	Açıklama
Year of diagnosis	Hastalığın teşhis yılı
SEER registry	Katılımcılar için SEER kayıt defterine atanan benzersiz bir kod
PRCDA 2016	Tanı bölgesinin PRCDA tarafından servis edilip edilmediğini tanımlayan alan
PRCDA Region	PRCDA ile çalışılırken kullanılan bölgeler

Tablo 4.1. (Devam) Meme kanseri ile ilişkisi olmayan öznitelikler ve açıklamaları

Öznitelik Adı	Açıklama
TNM Edition Number (2016+)	Vakayı evrelemek için kullanılan AJCC kılavuzunun baskısı (2016+)
CS version input current (2004+)	Herhangi bir alan güncellendikten veya kaydedildikten sonra CS versiyonunu gösteren alan (2004+)
CS version input original (2004+)	Başlangıçta CS alanlarını kodlamak için kullanılan versiyon numarasını gösteren alan (2004+)
CS version derived (2004-2015)	CS çıktı alanlarını türetmek için kullanılan en son CS versiyon numarasını gösteren alan (2004-2015)
Record Number Recode	Hastanın tümörlerinin kayıt numarası
IHS Link	Dosyaların Hint Sağlık Hizmeti dosyalarıyla bağlantılı olup olmadığını gösteren alan
Month of diagnosis	Teşhisin konulduğu ay
Month of diagnosis recode	Teşhisin konulduğu ay(yeniden kodlanmış)
Patient ID	Hasta için kullanılan kayıt numarası
Type of Reporting Source	Vakayı özetlemek için kullanılan kaynak belgeleri tanımlayan alan
Insurance Recode (2007+)	Hastanın sigorta durumunu gösteren alan (2007+)
Diagnostic Confirmation	Kanser varlığını doğrulamak için kullanılan en iyi yöntem
CS Tumor Size/Ext Eval (2004-2015)	CS Tumor Size ve CS Extension kodlarının kullanılan tanı yöntemlerine göre nasıl belirlendiğini kaydeden alan
CS Reg Node Eval (2004-2015)	CS Regional node alanının nasıl kodlandığını gösteren kodu kaydeden alan (2004-2015)
CS Mets Eval (2004-2015)	CS Mets at dx kodunun kullanılan teşhis yöntemlerine göre nasıl belirlendiğini kaydeden alan (2004-2015)
CS site-specific factor 6 (2004+ varying by schema)	Patolojik tümör boyutunun CS Tümör Boyutunda nasıl kodlandığını gösteren kodu kaydeden alan (2004+ şemaya göre değişken)

Çıkarılan alakasız öznitelikler sonucunda geriye 93 tane öznitelik kalmıştır. 93 tane öznitelik içinden kayıtlar için büyük oranda uygulanamaz veri içeren öznitelikler bulunmaktadır. Uygulanamaz veri içeren kayıtların silinmesi veri setini anlamsız hale getirip çok daraltacağı için bu özniteliklerin veri setinden çıkarılması uygun görülmüştür. Bu öznitelikler Tablo 4.2’te verilmiştir.

Tablo 4.2. *Uygulanamaz veri içeren öznitelikler*

Öznitelik Adı	Açıklama
Louisiana 2005 - 1st vs 2nd half of year	2005 yılının ilk yarısında teşhis edilen Louisiana vakalarını, yılın ikinci yarısında teşhis edilenlerden ayırmak için kullanılan alan
Lymphoma - Ann Arbor Stage (1983-2015)	Lenfomanın anatomik evreleri (1983-2015)
CS site-specific factor 4 (2004+ varying by schema)	Bölgesel lenf nodlarının immünohistokimyası (2004+ şemaya göre değişken)
CS site-specific factor 5 (2004+ varying by schema)	Bölgesel lenf nodlarının moleküler çalışmaları (2004+ şemaya göre değişken)
Primary by International Rules	Uluslararası kurallara göre tümörün primer olup olmadığını belirten alan

Geriye kalan öznitelikler arasında çok büyük oranda kayıp veri içeren öznitelikler bulunmaktadır. Bu öznitelikler de veri setinden çıkarılmıştır. Çıkarılan öznitelikler Tablo 4.3'te verilmiştir.

Tablo 4.3. *Öznitelikler ve kayıp veri değerleri*

Öznitelik Adı	Açıklama	Kayıp Veri (%)
Derived SEER Cmb Stg Grp (2016+)	Türetilmiş SEER birleştirilmiş aşama grubunun türetilmiş algoritmik hesaplamasının sonuçları (2016+)	%98
Derived SEER Combined T (2016+)	Türetilmiş SEER birleştirilmiş T sınıflandırmasının türetilmiş algoritmik hesaplamasının sonuçları (2016+)	%98
Derived SEER Combined N (2016+)	Türetilmiş SEER birleştirilmiş N sınıflandırmasının türetilmiş algoritmik hesaplamasının sonuçları (2016+)	%98
Derived SEER Combined M (2016+)	Türetilmiş SEER birleştirilmiş M sınıflandırmasının türetilmiş algoritmik hesaplamasının sonuçları (2016+)	%98
Derived SEER Combined T Src (2016+)	Türetilmiş SEER birleştirilmiş T'nin türetilmiş algoritmik hesaplaması için seçilen kaynak bilgisi (2016+)	%98
Derived SEER Combined N Src (2016+)	Türetilmiş SEER birleştirilmiş N'nin türetilmiş algoritmik hesaplaması için seçilen kaynak bilgisi (2016+)	%98
Derived SEER Combined M Src (2016+)	Türetilmiş SEER birleştirilmiş M'nin türetilmiş algoritmik hesaplaması için seçilen kaynak bilgisi (2016+)	%98
Tumor Size Summary (2016)	Genellikle cerrahi rezeksiyon numunesinde ölçülen katı primer tümörün en doğru ölçümü (2016+)	%99
Mets at DX-Distant LN (2016+)	Uzak lenf nodlarının/ düğümlerinin ilgili bir metastatik bölge olup olmadığını tanımlama (2016+)	%98
Mets at DX-Other (2016+)	Kemik, beyin, karaciğer veya akciğerde yakalanmayan herhangi bir uzak metastaz (2016+)	%98

Geriye kalan 78 öznitelikten sınıf verisi oluşturmak için kullanılan 3 öznitelik de (Survival month, COD to site recode, Vital status recode) veri setinden çıkarılmıştır. Ayrıca “COD to site recode” özniteliği ile aynı şeyi ifade eden “COD to site rec KM”

ve “Vital status recode” özniteliği ile aynı şeyi ifade eden ”SEER cause-specific death classification” öznitelikleri de veri setinden çıkarılmıştır.

Yukarıdaki veri seri ön işleme sonucunda 73 öznitelik ve 1 sınıf verisi içeren veri seti elde edilmiştir. Veri setinin öznitelikleri Tablo 4.4’te verilmiştir.

Tablo 4.4. Veri setinde bulunan öznitelikler ve açıklamaları

Öznitelik Adı	Açıklama
Age recode with <1 year olds	Tanı konulduğunda hastanın yaşı
Race recode (White, Black, Other)	İrk (yeniden kodlama) (Beyaz, Siyahi, Diğer)
Sex	Cinsiyet
County	Eyalet
State-county	İdari bölge-eyalet
Behavior recode for analysis	Davranış (Analiz için yeniden kodlama)
AYA site recode/WHO 2008	Özellikle ergen ve genç erişkinlere ait verileri analiz etmek için kullanılan bir bölge/histoloji kodu.
ICCC site recode ICD-O-3/WHO 2008	Genel olarak çocuklar hakkındaki verileri analiz etmek için kullanılan bir bölge/histoloji kodu.
Primary Site - labeled	Primer tümörün oluştuğu bölge (etiketlenmiş)
Primary Site	Primer tümörün oluştuğu bölge
Histologic Type ICD-O-3	Histolojik tip (ICD-O-3’e göre)
Behavior code ICD-O-3	Davranış kodu (ICD-O-3’e göre)
Grade	Sınıf
Laterality	Organ üzerinde tümörün bulunduğu taraf
ICD-O-3 Hist/behav	Histoloji/Davranış (ICD-O-3’e göre)
ICD-O-3 Hist/behav, malignant	Histoloji/Davranış (Kötü huylu tümörler) (ICD-O-3’e göre)
Histology recode - broad groupings	Histoloji (yeniden kodlama-geniş gruplandırma)
ICCC site rec extended ICD-O-3/WHO 2008	Ergen ve genç erişkinlerde kullanılan konum/histoloji (ICD-O-3/WHO 2008’e göre)
Site recode B ICD-O-3/WHO 2008	Birincil bölge ve histolojiye dayalı bir yeniden kodlama (ICD-O-3/WHO 2008’e göre)

Tablo 4.4. (Devam) Veri setinde bulunan öznitelikler ve açıklamaları

Öznitelik Adı	Açıklama
Summary stage 2000 (1998+)	Özet kanser evresi (1998+)
SEER Combined Summary Stage 2000 (2004+)	Özet kanser evresi (2000) ve Türetilmiş özet kanser evresi (2000) alanlarının birleştirilmesiyle elde edilmiş alan (2004+)
SEER historic stage A (1973-2015)	1973-2003 için EOD kodlama sistemi, 2004-2015 için ise CS kodlama sistemi kullanılarak elde edilmiş kanser evresini tanımlayan alan (1973-2015)
Derived SS1977 (2004-2015)	SEER özet kanser evresi 1997 alanından türetilerek oluşturulmuş kanser evresini tanımlayan alan (2004-2015)
Derived AJCC T, 7th ed (2010-2015)	TNM türetilmiş evreleme sisteminin T bileşeni (AJCC 7. baskı) (2010-2015)
Derived AJCC T, 7th ed (2010-2015)	TNM türetilmiş evreleme sisteminin T bileşeni (AJCC 7. baskı) (2010-2015)
Derived AJCC N, 7th ed (2010-2015)	TNM türetilmiş evreleme sisteminin N bileşeni (AJCC 7. baskı) (2010-2015)
Derived AJCC M, 7th ed (2010-2015)	TNM türetilmiş evreleme sisteminin M bileşeni (AJCC 7. baskı) (2010-2015)
Derived AJCC Stage Group, 6th ed (2004-2015)	Kanser evresi türetilmiş grupları (AJCC 6. baskı) (2004-2015)
Breast - Adjusted AJCC 6th Stage (1988-2015)	Kanser evresi (AJCC 6. baskı) (1988-2015)
Derived AJCC T, 6th ed (2004-2015)	TNM türetilmiş evreleme sisteminin T bileşeni (AJCC 6. baskı) (2004-2015)
Derived AJCC N, 6th ed (2004-2015)	TNM türetilmiş evreleme sisteminin N bileşeni (AJCC 6. baskı) (2004-2015)
Derived AJCC M, 6th ed (2004-2015)	TNM türetilmiş evreleme sisteminin M bileşeni (AJCC 6. baskı) (2004-2015)
Breast - Adjusted AJCC 6th T (1988-2015)	TNM evreleme sisteminin T bileşeni (AJCC 6. baskı) (1988-2015)
Breast - Adjusted AJCC 6th N (1988-2015)	TNM evreleme sisteminin N bileşeni (AJCC 6. baskı) (1988-2015)
Breast - Adjusted AJCC 6th M (1988-2015)	TNM evreleme sisteminin M bileşeni (AJCC 6. baskı) (1988-2015)
RX Summ--Surg Prim Site (1998+)	İlk çalışmanın veya ilk tedavi sürecinin bir parçası olarak gerçekleştirilen birincil bölgenin dokusunu çıkaran ve / veya yok eden bir cerrahi prosedür (1998+)
RX Summ--Surg Oth Reg/Dis (2003+)	Uzak lenf düğümlerinin veya diğer dokuların veya organların birincil alanın ötesinde cerrahi olarak çıkarılması prosedürü (2003+)
Reason no cancer-directed surgery	Ameliyat yapılmamasının nedeni
Regional nodes examined (1988+)	Çıkarılan ve incelenen toplam bölgesel lenf nodu sayısını kaydeden alan (1988+)
Regional nodes positive (1988+)	Metastaz içerdiği tespit edilen bölgesel lenf düğümlerinin tam sayısı (1988+)

Tablo 4.4. (Devam) Veri setinde bulunan öznitelikler ve açıklamaları

Öznitelik Adı	Açıklama
SEER Combined Mets at DX-bone (2010+)	Kemikte metastaz (2010+)
SEER Combined Mets at DX-brain (2010+)	Beyinde metastaz (2010+)
SEER Combined Mets at DX-liver (2010+)	Karaciğerde metastaz (2010+)
SEER Combined Mets at DX-lung (2010+)	Akciğerde metastaz (2010+)
Breast Subtype (2010+)	Meme alt tipi (2010+)
ER Status Recode Breast Cancer (1990+)	Yeniden kodlanmış östrojen durumu (1990+)
PR Status Recode Breast Cancer (1990+)	Yeniden kodlanmış progesteron durumu (1990+)
Derived HER2 Recode (2010+)	SSF9 (IHC), SSF11 (FISH), SSF13 (CISH), SSF14 (Other) verilerinden oluşturulmuş özet yeniden kodlama (2010+)
CS tumor size (2004-2015)	Tümör boyutu (2004-2015)
CS extension (2004-2015)	Menşе organı içindeki primer tümörün bitişik büyümesini veya komşu organlara doğrudan uzantısını tanımı (2004-2015)
CS lymph nodes (2004-2015)	Tanı anında kanserle ilgili bölgesel lenf düğümleri (2004-2015)
CS mets at dx (2004-2015)	Tanı anında metastatik tutulumun uzak bölgeleri (2004-2015)
CS site-specific factor 1	Östrojen reseptör
CS site-specific factor 2	Progesteron reseptör
CS site-specific factor 3	Pozitif Ipsilateral Seviye I-II Aksiller Lenf Nodu Sayısı
CS site-specific factor 7	Nottingham veya Bloom-Richardson (BR) Puanı / Notu
CS site-specific factor 15	Özet test sonuçları
Sequence number	Hastanın tüm ömrü boyunca meydana gelen tümörlerin sayı ve dizisi
First malignant primary indicator	İlk kötü huylu tümör indikatörü
Total number of in situ/malignant tumors for patient	Yayılmamış/ kötü huylu tümörlerin sayıca toplamı
Total number of benign/borderline tumors for patient	İyi huylu/sınırdaki tümörlerin sayıca toplamı
Behavior code ICD-O-2	Davranış kodu (ICD-O-2'ye göre)

Tablo 4.4. (Devam) Veri setinde bulunan öznitelikler ve açıklamaları

Öznitelik Adı	Açıklama
Histology ICD-O-2	Histoloji (ICD-O-2'ye göre)
Recode ICD-O-2 to 9	Birincil bölge ve morfoloji(ICD-O-2 kodlama sisteminden ICD-O-9 kodlama sistemine dönüştürülmüş)
Recode ICD-O-2 to 10	Birincil bölge ve morfoloji(ICD-O-2 kodlama sisteminden ICD-O-10 kodlama sistemine dönüştürülmüş)
Age recode with single ages and 85+	Tanı sırasındaki yaşı (yeniden kodlama)
Race recode (W, B, AI, API)	İrk (yeniden kodlama)
Origin recode NHIA (Hispanic, Non-Hisp)	Köken (yeniden kodlama)
Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic)	İrk ve köken (yeniden kodlama)
Age at diagnosis	Tanı sırasında yaşı
Race/ethnicity	İrk/etnik köken
Year of birth	Doğum yılı
Marital status at diagnosis	Tanı sırasında evlilik durumu
Survival	Hastanın hastalığı yenip yenemediğini gösteren sınıf değeri

4.2.4. Kayıp Verilerin Doldurulması

Bir öznitelikte yer alan kayıp veriler, kategorik öznitelikler için o öznitelikte en sık geçen değer ile doldurulmuştur. Nümerik öznitelikler için ise değerlerin ortalaması yöntemi kullanılarak doldurulmuştur.

4.2.5. Normalizasyon

Algoritmalar çalıştırılmadan önce, veri setindeki nümerik değerler normalizasyon işlemine tabi tutulmuştur. Weka ara yüzündeki ön işleme algoritmalarından “Normalize” çalıştırılarak, sayısal verilerin değer aralığı 0.0-1.0 aralığına indirgenmiştir.

4.3. Değerlendirme Ölçütleri

4.3.1. Doğruluk

Doğruluk ölçütü sınıflandırma model sonuçlarını değerlendirmede en sık kullanılan yöntemlerden biridir.

İkili sınıflandırma problemleri için hata matrisi Şekil 4.1’deki gibi tanımlanır.

		Tahmin Edilen	
		1	-1
Gerçek	1	DP	YN
	-1	YP	DN

Şekil 4.1. Hata Matrisi

Burada ilk satırı ve ilk sütun elemanı gerçek etiketleri 1 olan ve -1 olarak sınıflandırılan değerleri ifade eder ve Doğru Pozitif (DP) olarak adlandırılır. Hata matrisinin ilk satırı ve ikinci sütun elemanı, gerçek etiketlerin 1 olduğu ancak bunlar -1 olarak sınıflandırıldığı örnek sayısını temsil eder ve Yanlış Negatif (YN) olarak adlandırılır. Yanlış Pozitif (YP) olarak adlandırılan ikinci satır 1. Sütun elemanı gerçek etiketleri -1 olan ancak 1 olarak sınıflandırılan örnekler için kullanılır. Son olarak ikinci satır ikinci sütun elemanı ise Doğru Negatif (DN) adlandırılır ve etiketi gerçekte -1 olup -1 olarak sınıflandırılan örnekleri ifade eder.

Buradan yola çıkarak doğruluk değeri (4.1)’de ifade edilmiştir.

$$\text{doğruluk} = \frac{DP + DN}{DP + YN + YP + DN} \quad (4.1)$$

Denklemden anlaşılacağı üzere doğruluk değeri, doğru olarak sınıflandırılan kayıtların toplamının (DP + DN), kayıt sayısının tamamına (DP + YN + YP + DN) bölünmesi ile elde edilir.

4.3.2. F-Skor

F-Skor hassasiyet ve hatırlatma metriklerinin harmonik ortalamasıdır (Öztürk, 2018, s.33).

Hassasiyet formülasyonu (4.2)’de verilmiştir.

$$\text{Hassasiyet} = \frac{DP}{DP + YP} \quad (4.2)$$

Hatırlatma metriği ise (4.3)'teki gibi ifade edilir.

$$\text{Hassasiyet} = \frac{DP}{DP + YN} \quad (4.3)$$

Buradan yola çıkılarak F-Skor (4.4)'teki gibi tanımlanır.

$$F - \text{Skor} = \frac{2}{\frac{1}{\text{hassasiyet}} + \frac{1}{\text{hatırlatma}}} = \frac{2DP}{2DP + YP + YN} \quad (4.4)$$



5. DENEYSEL ÇALIŞMALAR VE SONUÇLARI

5.1. Makine Öğrenmesi Algoritmaları ile Sağ Kalım Tahmini

SEER veri tabanından alınmış olan meme kanseri veri seti üzerinde ön işleme sonucunda 73 öznitelik ve 1 sınıf değerinden oluşan ve toplam 370.153 kayıt içeren bir veri seti elde edilmiştir. Elde edilen veri setinde kayıp veriler en sık görülen verilerle doldurulmuş ve normalizasyon işleminden geçirilmiştir. Veri seti eğitim veri seri ve test veri seti olarak ikiye ayrılmıştır. Veri setinin %66,667'si eğitim için, geriye kalan %33.333'ü ise test için kullanılmıştır. Eğitim setinde bulunan kayıt sayısı 246.769, test setinde kullanılan kayıt sayısı ise 123.384'tür. Veri seti üzerinde Naive Bayes, J48, Çok Amaçlı Evrimsel Bulanık Sınıflandırıcı (ÇAEBS), Destek Vektör Makineleri (DVM), algoritmaları Weka aracılığıyla çalıştırılıp sonuçlar kaydedilmiştir.

Çalıştırılan 4 algoritmanın her biri farklı başarı oranları ile sonuç vermiştir. En yüksek oranda doğru sonuç veren algoritma %93,0234 ile J48 algoritması, en düşük oranda doğru sonuç veren algoritma ise %83,3033 ile Naive Bayes algoritması olmuştur. Algoritmaların başarı yüzdeleri Tablo 5.1'de verilmiştir.

Tablo 5.1. Sınıflandırma algoritmalarının karşılaştırılması

Algoritma	Başarı Yüzdesi (%)	F-Skor	Çalışma Süresi
Naive Bayes	83,3033	0,852	8,20
J48	93,0234	0,925	84,39
ÇAEBS	88,3559	0,856	4029,75
DVM	91,4349	0,902	17830,39

Yapılan uygulamada algoritmaların çalışma süreleri oldukça farklılık göstermektedir. En hızlı sonuç üreten algoritma 8,2 saniye ile Naive Bayes algoritması, en yavaş sonuç üreten algoritma ise 17830,39 saniye ile DVM algoritmasıdır.

5.2. Veri Seti Üzerinde Öznitelik Seçimi Çalışması

Veri seti üzerinde algoritmaların çalıştırılması sonucu elde edilen başarılı sonuçlara rağmen hastaya ait 73 öz niteliği toplamının zorluğu ve 73 öznitelik içinde bazı özniteliklerin başarıyı düşürüyor olabileceği durumu öznitelik çalışması yapılmasını gerekli kılmıştır. Çalışmanın bu aşamasında, 3 farklı filtre öznitelik seçim

yöntemi kullanılmıştır. Bu öznitelik seçimleri filtre öznitelik seçim sınıfına ait olup Bilgi Kazanımı Yöntemi, Kazanım Oranı Yöntemi ve Korelasyon Tabanlı Öznitelik Seçimidir. 73 öznitelik içeren veri seti üzerinde belirtilen öznitelik seçim yöntemleri uygulanmış ve ilk 24 öznitelik veri seti olarak belirlenmiştir.

5.3. Bilgi Kazanımı ile Öznitelik Seçimi

Veri seti üzerinde ilk olarak Bilgi Kazanımı öznitelik seçim yöntemi uygulanmıştır. Tablo 5.2’de öznitelikler ve sağladıkları fayda değerleri, fayda değerine göre büyükten küçüğe sıralanarak verilmiştir.

Tablo 5.2. *Bilgi kazanımı öznitelik seçimi sonucu seçilen öznitelikler*

Öznitelik Adı	Fayda Değeri
Breast - Adjusted AJCC 6th Stage (1988-2015)	0.159207
SEER Combined Summary Stage 2000 (2004+)	0.153205
Breast - Adjusted AJCC 6th T (1988-2015)	0.151887
RX Summ--Surg Prim Site (1998+)	0.107738
CS mets at dx (2004-2015)	0.099417
Derived AJCC M, 6th ed (2004-2015)	0.09857
Derived AJCC N, 6th ed (2004-2015)	0.085606
Reason no cancer-directed surgery	0.085344
CS extension (2004-2015)'	0.077583
CS lymph nodes (2004-2015)'	0.075608
CS tumor size (2004-2015)'	0.067464
ICD-O-3 Hist/behav	0.048557
Histologic Type ICD-O-3	0.030828
Regional nodes positive (1988+)	0.030451
Behavior recode for analysis	0.03017

Tablo 5.2 (Devam) Bilgi kazanımı öznitelik seçimi sonucu seçilen öznitelikler

Öznitelik Adı	Fayda Değeri
AYA site recode/WHO 2008	0.03017
First malignant primary indicator	0.03017
CS site-specific factor 3 (2004+ varying by schema)	0.028931
Age at diagnosis	0.027121
Regional nodes examined (1988+)	0.024553
Grade	0.02287
Year of birth	0.022728
PR Status Recode Breast Cancer (1990+)	0.016464
ER Status Recode Breast Cancer (1990+)	0.015392

Bilgi Kazanımı ile seçilen özniteliklerle sınıflandırma algoritmaları çalıştırılmıştır. Elde edilen doğruluk ve F-Skor değerleri Tablo 5.3’da verilmiştir.

Tablo 5.3. Bilgi Kazanımı ile seçilen özniteliklerle elde edilen doğruluk ve F-Skor değerleri

Algoritma	Doğruluk (%)	F-Skor
Naive Bayes	87,7123	0,883
J48	93,269	0,928
ÇAEBS	88,3551	0,846
DVM	91,7372	0,906

Bilgi Kazanımı öznitelik seçim yöntemiyle elde edilen veri setlerinin Naive Bayes, J48, ÇAEBS ve DVM algoritmaları ile başarıları test edildikten sonra, bu veri setleri üzerinde sarmalama öznitelik seçim yöntemlerinden Sıralı İleri Seçim yöntemi uygulanarak öznitelik sayısını azaltmak ve başarı oranını artırmak amaçlanmıştır.

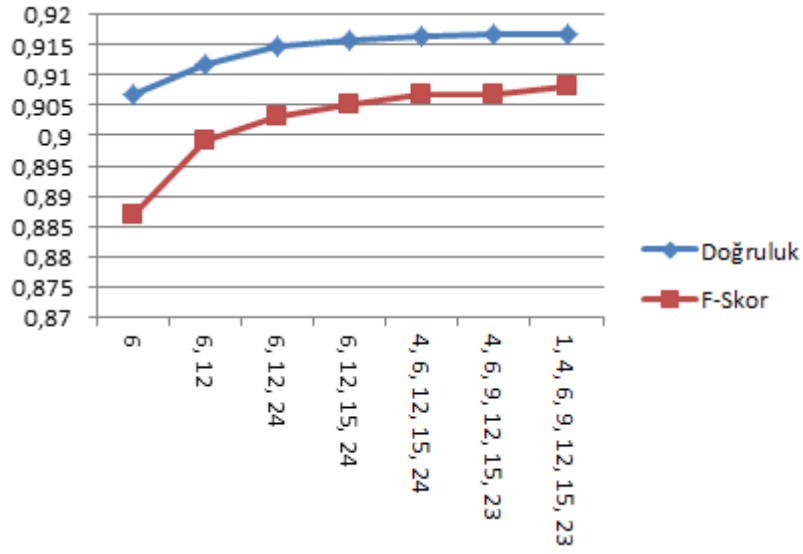
Bilgi Kazanımı seçilmiş öznitelikler üzerinde Sıralı İleri Seçim yöntemi uygulanarak elde edilen başarı oranları ayrıntılı şekilde EK- 2’de verilmiştir.

Sıralı ileri seçimin Naive Bayes algoritması ile uygulanması sonucu seçilen öznitelikler Tablo 5.4’de verilmiştir.

Tablo 5.4. *Bilgi Kazanımı ile seçilmiş veri setinde Naive Bayes ile ileri aşamalı seçim sonucu*

Öznitelik kod numarası	Bilgi Kazanımı ile seçilen öznitelikler
1	Behavior recode for analysis
4	Grade
6	SEER Combined Summary Stage 2000
9	Derived AJCC M, 6th ed
12	Reason no cancer-directed surgery
15	ER Status Recode Breast Cancer
23	Age at diagnosis

Seçilen 7 tane öznitelik ile Naive Bayes algoritmasının elde ettiği doğruluk değeri %91,6747, F-Skor değeri ise 0,9081’dir. Şekil 5.1’de seçilen öznitelik kümelerinin sırasıyla doğruluk ve F-Skor değerleri verilmiştir.



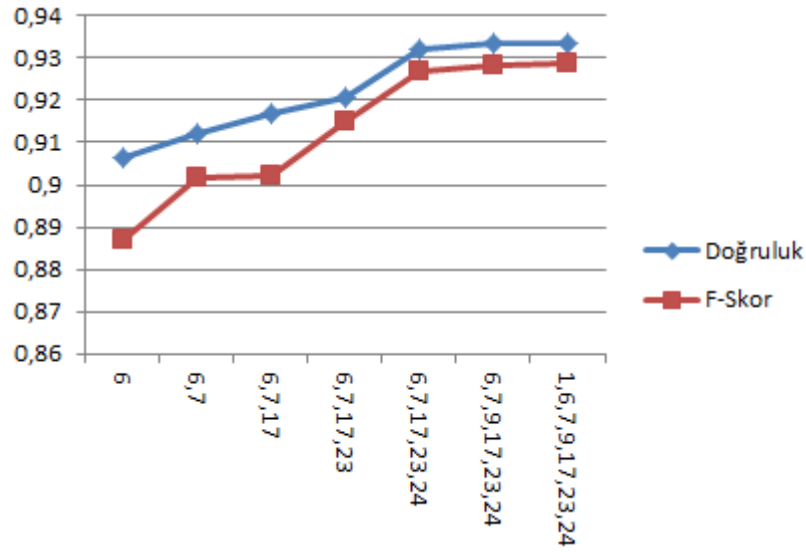
Şekil 5.1. Naive Bayes en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Bilgi Kazanımı)

Bilgi Kazanımı sonucu elde edilmiş özniteliklerle Sıralı İleri Seçimin J48 algoritması ile uygulanması sonucu seçilen öznitelikler Tablo 5.5’da verilmiştir.

Tablo 5.5. Bilgi Kazanımı ile seçilmiş veri setinde J48 ile ileri aşamalı seçim sonucu

Öznitelik kod numarası	Bilgi Kazanımı ile seçilen öznitelikler
1	Behavior recode for analysis
6	SEER Combined Summary Stage 2000
7	Breast - Adjusted AJCC 6th Stage
9	Derived AJCC M, 6th ed
17	CS tumor size
23	Age at diagnosis
24	Year of birth

Seçilen 7 öznitelik ile J48 algoritmasının elde ettiği doğruluk değeri %93,3589, F-Skor değeri ise 0,9286’dır. Şekil 5.2’de seçilen öznitelik kümelerinin sırasıyla doğruluk ve F-Skor değerleri verilmiştir.



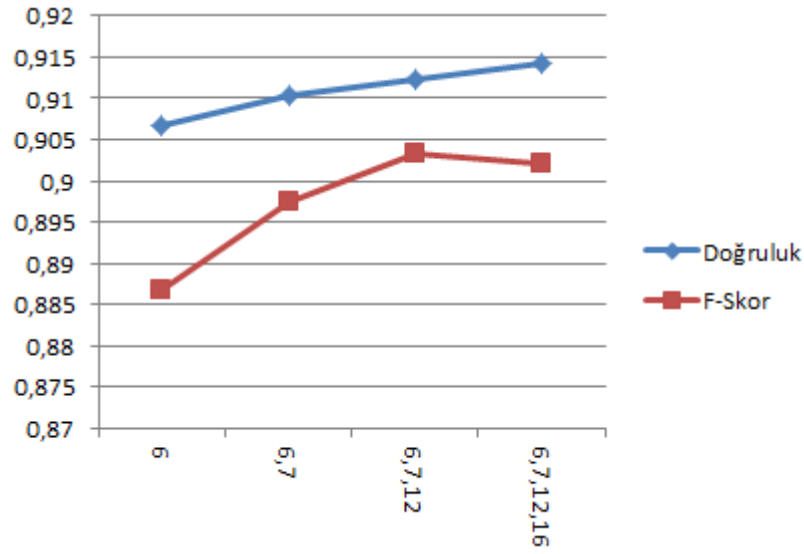
Şekil 5.2. J48 en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Bilgi Kazanımı)

Sıralı İleri Seçimin ÇAEBS algoritması ile uygulanması sonucu seçilen öznitelikler Tablo 5.6’da verilmiştir

Tablo 5.6. Bilgi Kazanımı ile seçilmiş veri setinde ÇAEBS ile ileri aşamalı seçim sonucu

Öznitelik kod numarası	Bilgi Kazanımı ile seçilen öznitelikler
6	SEER Combined Summary Stage 2000
7	Breast - Adjusted AJCC 6th Stage
12	Reason no cancer-directed surgery
16	PR Status Recode Breast Cancer

Seçilen 4 öznitelik ile ÇAEBS algoritmasının elde ettiği doğruluk değeri %91,4154, F-Skor değeri ise 0,9021’dir. Şekil 5.3’de seçilen öznitelik kümelerinin sırasıyla doğruluk ve F-Skor değerleri verilmiştir.



Şekil 5.3. ÇAEBS en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Bilgi Kazanımı)

Sıralı İleri Seçimin DVM algoritması ile uygulanması sonucu seçilen öznitelikler Tablo 5.7’te verilmiştir.

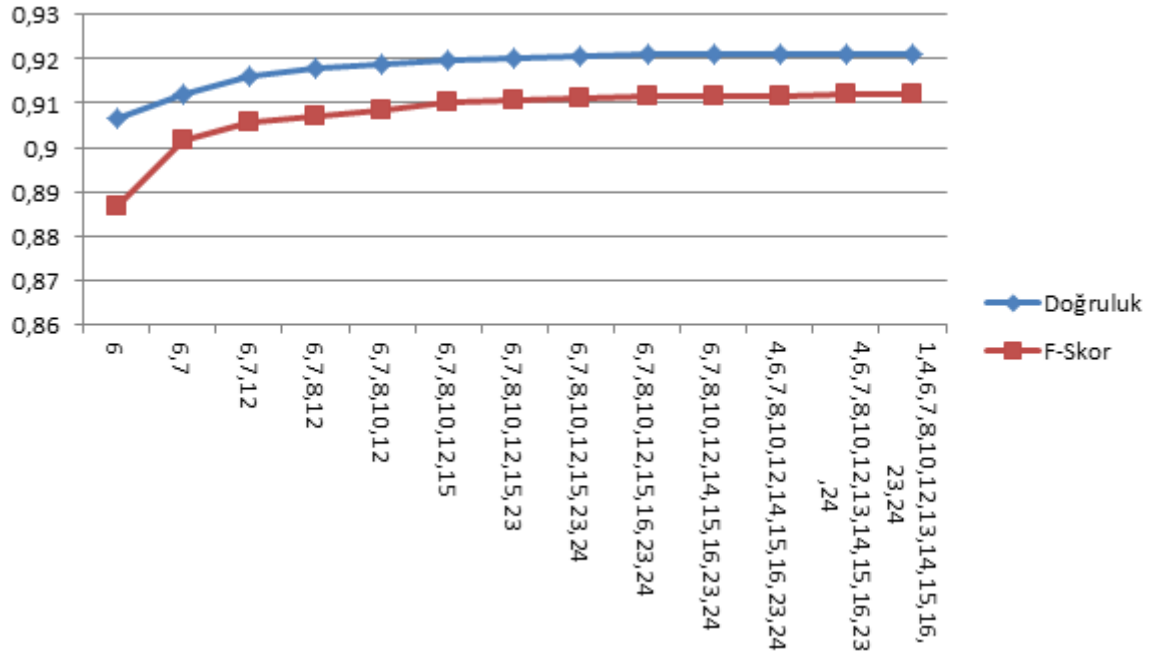
Tablo 5.7. Bilgi Kazanımı ile seçilmiş veri setinde DVM ile ileri aşamalı seçim sonucu

Öznitelik kod numarası	Bilgi Kazanımı ile seçilen öznitelikler
1	Behavior recode for analysis
4	Grade
6	SEER Combined Summary Stage 2000
7	Breast - Adjusted AJCC 6th Stage
8	Derived AJCC N, 6th ed
10	Breast - Adjusted AJCC 6th T
12	Reason no cancer-directed surgery
14	Regional nodes positive
15	ER Status Recode Breast Cancer
16	PR Status Recode Breast Cancer
23	Age at diagnosis

Tablo 5.7 (Devam) Bilgi Kazanımı ile seçilmiş veri setinde DVM ile ileri aşamalı seçim sonucu

Öznitelik kod numarası	Bilgi Kazanımı ile seçilen öznitelikler
24	Year of birth

Seçilen 12 öznitelik ile DVM algoritmasının elde ettiği doğruluk değeri % 92,1286, F-Skor değeri ise 0,912'dir. Şekil 5.4'de seçilen öznitelik kümelerinin sırasıyla doğruluk ve F-Skor değerleri verilmiştir.



Şekil 5.4. DVM en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Bilgi Kazanımı)

5.4. Kazanım Oranı ile Öznitelik Seçimi

Kazanım oranı öznitelik seçimi algoritmasına göre öznitelik seçimi yapılarak Tablo 5.8'te öznitelikler ve fayda değerleri, fayda değerine göre büyükten küçüğe sıralanarak verilmiştir.

Tablo 5.8. Kazanım oranı öznitelik seçimi sonucu seçilen öznitelikler

Öznitelik	Fayda Değeri
Breast - Adjusted AJCC 6th M (1988-2015)	0.350644
CS mets at dx (2004-2015)	0.25913
Reason no cancer-directed surgery	0.186277

Tablo 5.8 (Devam) *Kazanım oranı öznitelik seçimi sonucu seçilen öznitelikler*

Öznitelik	Fayda Değeri
Summary stage 2000 (1998+)	0.084852
Breast - Adjusted AJCC 6th Stage (1988-2015)	0.064588
CS extension (2004-2015)	0.056234
Breast - Adjusted AJCC 6th T (1988-2015)	0.056076
Breast - Adjusted AJCC 6th N (1988-2015)	0.055967
CS lymph nodes (2004-2015)	0.043844
AYA site recode/WHO 2008	0.041994
Behavior code ICD-O-3	0.041994
First malignant primary indicator	0.041994
SEER Combined Mets at DX-bone (2010+)	0.034647
RX Summ--Surg Prim Site (1998+)	0.030923
SEER Combined Mets at DX-lung (2010+)	0.030706
SEER Combined Mets at DX-liver (2010+)	0.030106
SEER Combined Mets at DX-brain (2010+)	0.024847
ICD-O-3 Hist/behav, malignant	0.023487
ER Status Recode Breast Cancer (1990+)	0.020442
Regional nodes positive (1988+)	0.020022
CS site-specific factor 3 (2004+ varying by schema)	0.018553
PR Status Recode Breast Cancer (1990+)	0.017406
CS tumor size (2004-2015)	0.017032
Histology recode - broad groupings	0.016266

Kazanım Oranı ile seçilen özniteliklerle sınıflandırma algoritmaları çalıştırılmıştır. Elde edilen doğruluk ve F-Skor değerleri Tablo 5.9’da verilmiştir.

Tablo 5.9. *Kazanım Oranı ile seçilen özniteliklerle elde edilen doğruluk ve F-Skor değerleri*

Algoritma	Doğruluk (%)	F-Skor
Naive Bayes	87,9142	0,88
J48	92,4796	0,92
ÇAEBS	89,0018	0,86
DVM	91,3619	0,90

Kazanım oranı öznitelik seçimi ile elde edilen veri setlerinin Naive Bayes, J48, ÇAEBS ve DVM algoritmaları ile başarıları test edildikten sonra, bu veri setleri üzerinde sarmalama öznitelik seçim yöntemlerinden Sıralı İleri Seçim yöntemi uygulanarak öznitelik sayısını azaltmak ve başarı oranını artırmak amaçlanmıştır.

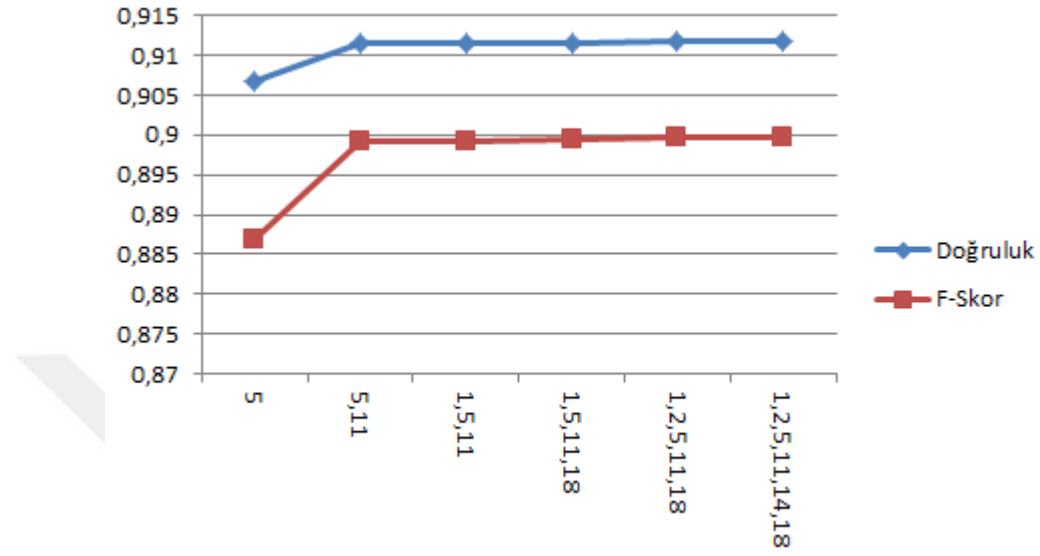
Kazanım oranı öznitelik seçim algoritmasıyla seçilmiş öznitelikler üzerinde sıralı ileri seçim yöntemi uygulanarak elde edilen başarı oranları ayrıntılı şekilde EK- 3’te verilmiştir.

Sıralı İleri Seçimin Naive Bayes algoritması ile uygulanması sonucu seçilen öznitelikler Tablo 5.10’de verilmiştir.

Tablo 5.10. *Kazanım Oranı ile seçilmiş veri setinde Naive Bayes ile ileri aşamalı seçim*

Öznitelik kod numarası	Kazanım Oranı ile seçilen öznitelikler
1	AYA site recode/WHO 2008
2	Behavior code ICD-O-3
5	Summary stage 2000
11	Reason no cancer-directed surgery
14	SEER Combined Mets at DX-brain
18	PR Status Recode Breast Cancer

Seçilen 6 öz nitelik ile Naive Bayes algoritmasının elde ettiği doğruluk değeri %91,1795, F-Skor değeri ise 0,8998'dir. Şekil 5.5'te seçilen öz nitelik kümelerinin sırasıyla doğruluk ve F-Skor değerleri verilmiştir.



Şekil 5.5. Naive Bayes en iyi öz nitelikler kümesi doğruluk ve F-Skor değer grafiği (Kazanım Oranı)

Sıralı ileri seçimin J48 algoritması ile uygulanması sonucu seçilen öz nitelikler Tablo 5.11'de verilmiştir.

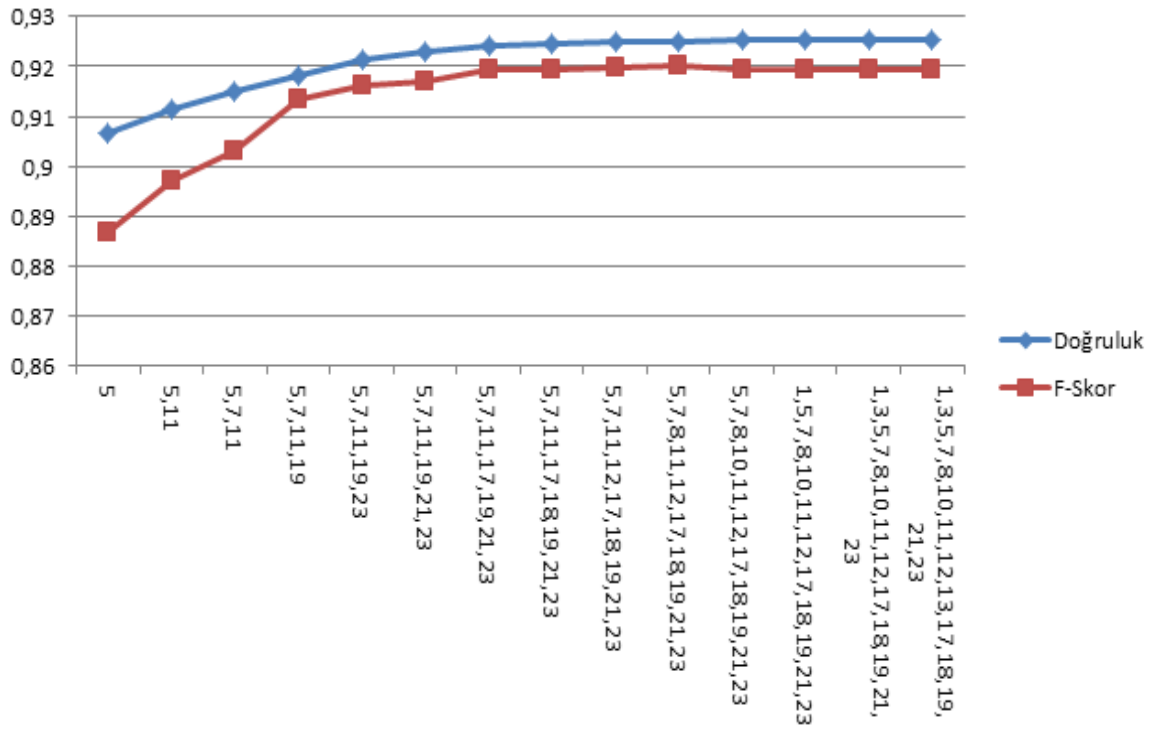
Tablo 5.11. Kazanım Oranı ile seçilmiş veri setinde J48 ile ileri aşamalı seçim sonucu

Öznitelik kod numarası	Kazanım Oranı ile seçilen öz nitelikler
1	AYA site recode/WHO 2008
3	ICD-O-3 Hist/behav, malignant
5	Summary stage 2000
7	Breast - Adjusted AJCC 6th T
8	Breast - Adjusted AJCC 6th N (1988-2015)
10	RX Summ--Surg Prim Site
11	Reason no cancer-directed surgery
12	Regional nodes positive

Tablo 5.11 (Devam) Kazanım Oranı ile seçilmiş veri setinde J48 ile ileri aşamalı seçim sonucu

Öznitelik kod numarası	Kazanım Oranı ile seçilen öznitelikler
13	SEER Combined Mets at DX-bone
17	ER Status Recode Breast Cancer
18	PR Status Recode Breast Cancer
19	CS tumor size
21	CS lymph nodes (2004-2015)
23	CS site-specific factor 3

Seçilen 14 öznitelik ile J48 algoritmasının elde ettiği doğruluk değeri %92,5273, F-Skor değeri ise 0,9195'dir. Şekil 5.6'de seçilen öznitelik kümelerinin sırasıyla doğruluk ve F-Skor değerleri verilmiştir.



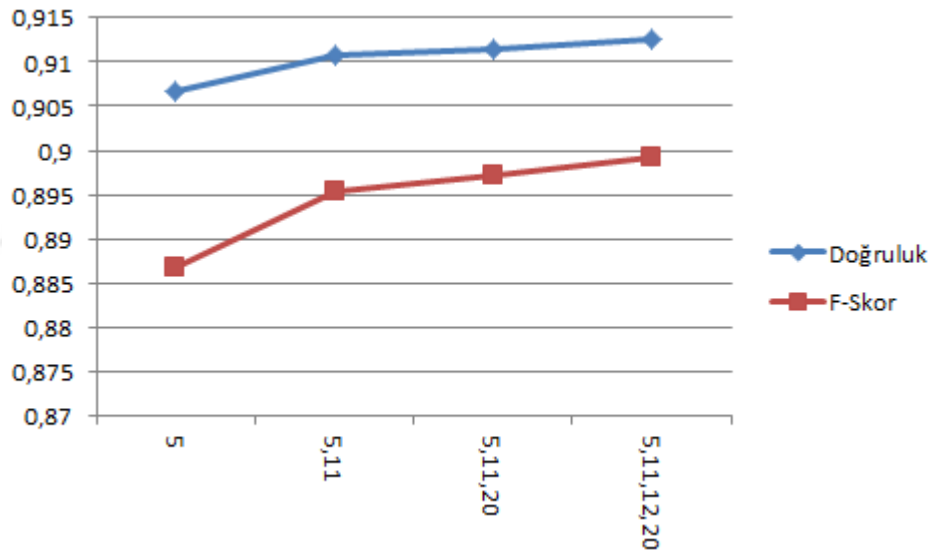
Şekil 5.6. J48 en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Kazanım Oranı)

Sıralı ileri seçimin ÇAEBS algoritması ile uygulanması sonucu seçilen öznitelikler Tablo 5.12'de verilmiştir.

Tablo 5.12. *Kazanım Oranı ile seçilmiş veri setinde ÇAEBS ile ileri aşamalı seçim sonucu*

Öznitelik kod numarası	Kazanım Oranı ile seçilen öznitelikler
5	Summary stage 2000
11	Reason no cancer-directed surgery
12	Regional nodes positive
20	CS extension

Seçilen 4 öznitelik ile ÇAEBS algoritmasının elde ettiği doğruluk değeri %91,2573, F-Skor değeri ise 0,8993'dir. Şekil 5.7'de seçilen öznitelik kümelerinin sırasıyla doğruluk ve F-Skor değerleri verilmiştir.



Şekil 5.7. *ÇAEBS en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Kazanım Oranı)*

Sıralı ileri seçimin DVM algoritması ile uygulanması sonucu seçilen öznitelikler Tablo 5.13'de verilmiştir.

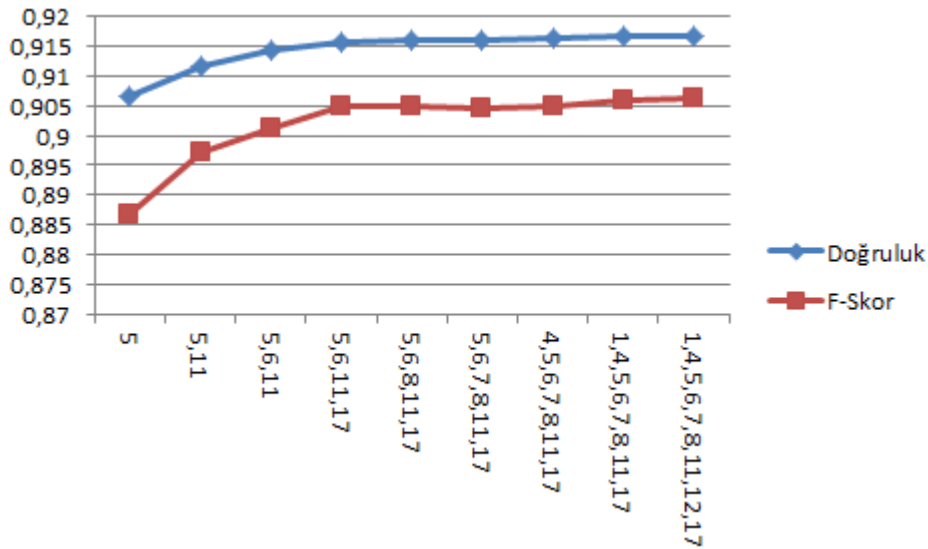
Tablo 5.13. *Kazanım oranı ile seçilmiş veri setinde DVM ile ileri aşamalı seçim sonucu*

Öznitelik kod numarası	Kazanım Oranı ile seçilen öznitelikler
1	AYA site recode/WHO 2008
4	Histology recode - broad groupings

Tablo 5.13. (Devam) Kazanım oranı ile seçilmiş veri setinde DVM ile ileri aşamalı seçim sonucu

Öznitelik kod numarası	Kazanım Oranı ile seçilen öznitelikler
5	Summary stage 2000
6	Breast - Adjusted AJCC 6th Stage
7	Breast - Adjusted AJCC 6th T
8	Breast - Adjusted AJCC 6th N (1988-2015)
11	Reason no cancer-directed surgery
17	ER Status Recode Breast Cancer

Seçilen 8 öznitelik ile DVM algoritmasının elde ettiği doğruluk değeri %91,6755, F-Skor değeri ise 0,9062'dir. Şekil 5.8'de seçilen öznitelik kümelerinin sırasıyla doğruluk ve F-Skor değerleri verilmiştir.



Şekil 5.8.DVM en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Kazanım Oranı)

5.5. Korelasyon Tabanlı Öznitelik Seçimi

Veri setine Korelasyon Tabanlı öznitelik seçimi uygulanması sonucu. Tablo 5.14'te öznitelikler ve fayda değerleri, fayda değerine göre büyükten küçüğe sıralanarak verilmiştir.

Tablo 5.14. *Korelasyon tabanlı öznitelik seçimi sonucu seçilen öznitelikler*

Öznitelik	Fayda Değeri
Derived AJCC M, 6th ed (2004-2015)	0.452419
CS mets at dx (2004-2015)	0.44208
Reason no cancer-directed surgery	0.404277
SEER Combined Mets at DX-bone (2010+)	0.276362
Regional nodes positive (1988+)	0.239432
CS site-specific factor 3 (2004+ varying by schema)	0.233882
Breast - Adjusted AJCC 6th N (1988-2015)	0.205187
SEER Combined Mets at DX-lung (2010+)	0.202908
SEER Combined Mets at DX-liver (2010+)	0.189824
CS lymph nodes (2004-2015)	0.189017
Summary stage 2000 (1998+)	0.186214
Breast Subtype (2010+)	0.183979
Behavior recode for analysis	0.175314
AYA site recode/WHO 2008	0.175314
First malignant primary indicator	0.175314
PR Status Recode Breast Cancer (1990+)	0.164956
ER Status Recode Breast Cancer (1990+)	0.16204
Derived AJCC Stage Group, 7th ed (2010-2015)	0.157594
CS site-specific factor 7 (2004+ varying by schema)	0.132028
Age at diagnosis	0.123916
SEER Combined Mets at DX-brain (2010+)	0.11224
RX Summ--Surg Prim Site (1998+)	0.104191

Tablo 5.14 (Devam) Korelasyon tabanlı öznitelik seçimi sonucu seçilen öznitelikler

Öznitelik	Fayda Değeri
CS site-specific factor 15 (2004+ varying by schema)	0.101145
Derived AJCC T, 7th ed (2010-2015)	0.100742

Elde edilen öznitelikler ile ilk çalışmada kullanılan eğitim ve test veri setlerinden yeni veri setleri elde edilmiştir. Yeni veri setleri ile Naive Bayes, J48, ÇAEBS ve DVM algoritmaları çalıştırılarak sonuçlar karşılaştırılmıştır. Yapılan çalışmanın sonucunda elde edilen doğruluk yüzdeleri Tablo 5.15'te verilmiştir.

Tablo 5.15. Kazanım Oranı ile seçilen özniteliklerle elde edilen doğruluk ve F-Skor değerleri

Algoritma	Doğruluk (%)	F-Skor
Naive Bayes	89,2028	0,892
J48	92,696	0,921
ÇAEBS	88,1289	0,863
DVM	91,5038	0,892

Korelasyon Tabanlı Öznitelik Seçim algoritmasıyla seçilmiş öznitelikler üzerinde sıralı ileri seçim yöntemi uygulanarak elde edilen başarı oranları ayrıntılı şekilde EK-4'te verilmiştir.

Sıralı ileri seçimin Naive Bayes algoritması ile uygulanması sonucu seçilen öznitelikler Tablo 5.16'te verilmiştir.

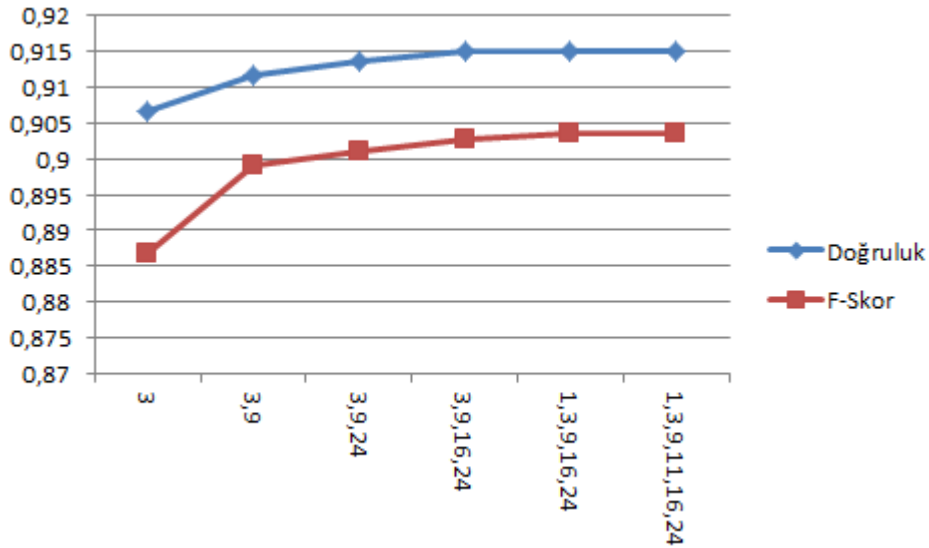
Tablo 5.16. Korelasyon ile seçilmiş veri setinde Naive Bayes ile ileri aşamalı seçim sonucu

Öznitelik kod numarası	Korelasyon ile seçilen öznitelikler
1	Behavior recode for analysis
3	Summary stage 2000

Tablo 5.16.(Devam) Korelasyon ile seçilmiş veri setinde Naive Bayes ile ileri aşamalı seçim sonucu

Öznitelik kod numarası	Korelasyon ile seçilen öznitelikler
9	Reason no cancer-directed surgery
11	SEER Combined Mets at DX-bone
16	ER Status Recode Breast Cancer
24	Age at diagnosis

Seçilen 6 öznitelik ile Naive Bayes algoritmasının elde ettiği doğruluk değeri %91,5118, F-Skor değeri ise 0,9035'tür. Şekil 5.9'da seçilen öznitelik kümelerinin sırasıyla doğruluk ve F-Skor değerleri verilmiştir.



Şekil 5.9.Naive Bayes en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Korelasyon)

Sıralı ileri seçimin J48 algoritması ile uygulanması sonucu seçilen öznitelikler Tablo 5.17'da verilmiştir.

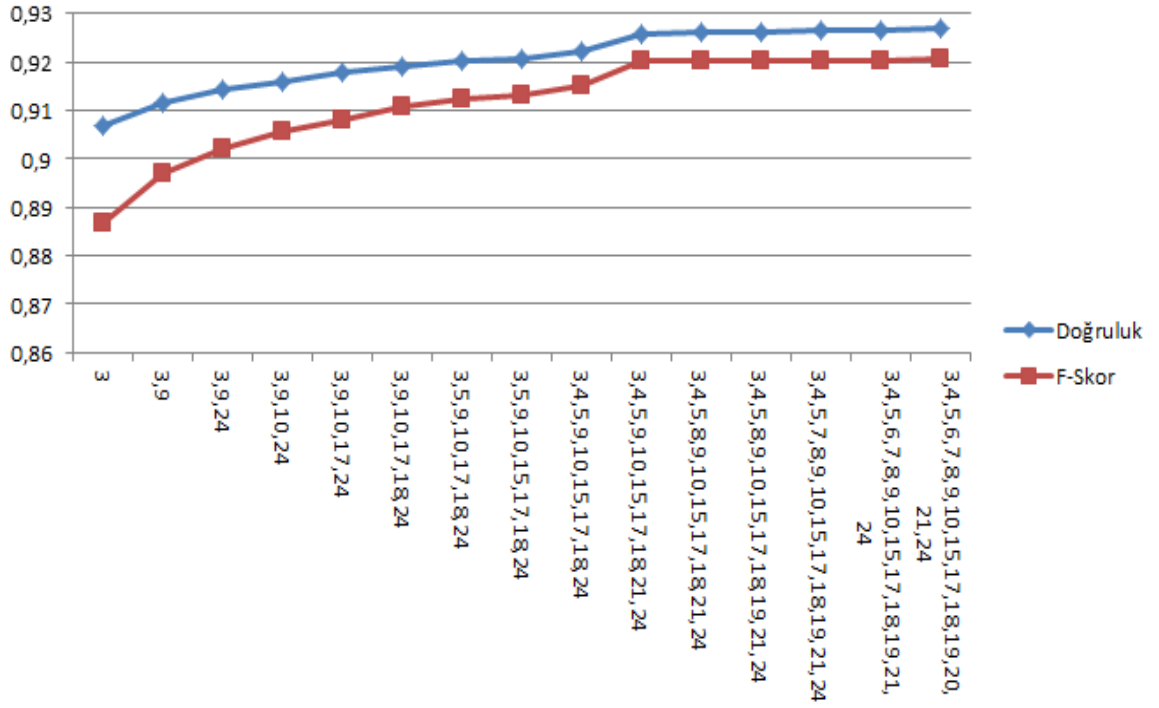
Tablo 5.17. Korelasyon ile seçilmiş veri setinde J48 ile ileri aşamalı seçim sonucu

Öznitelik kod numarası	Korelasyon ile seçilen öznitelikler
3	Summary stage 2000
4	Derived AJCC Stage Group, 7th ed

Tablo 5.17.(Devam) Korelasyon ile seçilmiş veri setinde J48 ile ileri aşamalı seçim sonucu

Öznitelik kod numarası	Korelasyon ile seçilen öznitelikler
5	Derived AJCC T, 7th ed
6	Derived AJCC M, 6th ed
7	Breast - Adjusted AJCC 6th N
8	RX Summ--Surg Prim Site
9	Reason no cancer-directed surgery
10	Regional nodes positive
15	Breast Subtype
17	PR Status Recode Breast Cancer
18	CS lymph nodes
19	CS mets at dx
20	CS site-specific factor 3
21	'CS site-specific factor 7
24	Age at diagnosis

Seçilen 15 öznitelik ile J48 algoritmasının elde ettiği doğruluk değeri %92,6886, F-Skor değeri ise 0,9207'dir. Şekil 5.10'da seçilen öznitelik kümelerinin sırasıyla doğruluk ve F-Skor değerleri verilmiştir.



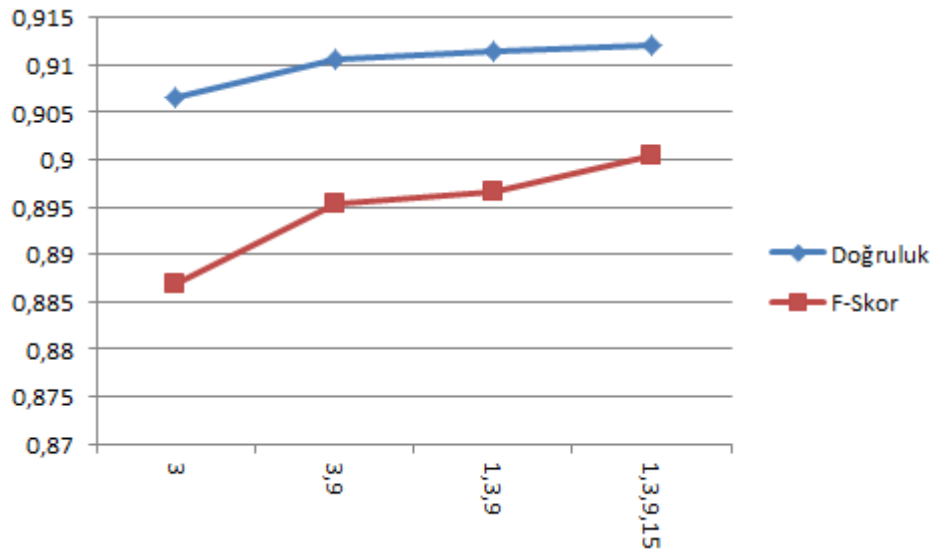
Şekil 5.10. J48 en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Korelasyon)

Sıralı ileri seçimin ÇAEBS algoritması ile uygulanması sonucu seçilen öznitelikler Tablo 5.18’de verilmiştir.

Tablo 5.18. Korelasyon ile seçilmiş veri setinde ÇAEBS ile ileri aşamalı seçim sonucu

Öznitelik kod numarası	Korelasyon ile seçilen öznitelikler
1	Behavior recode for analysis
3	Summary stage 2000
9	Reason no cancer-directed surgery
15	Breast Subtype

Seçilen 4 öznitelik ile ÇAEBS algoritmasının elde ettiği doğruluk değeri %91,2038, F-Skor değeri ise 0,9005’dir. Şekil 5.11’de seçilen öznitelik kümelerinin sırasıyla doğruluk ve F-Skor değerleri verilmiştir.



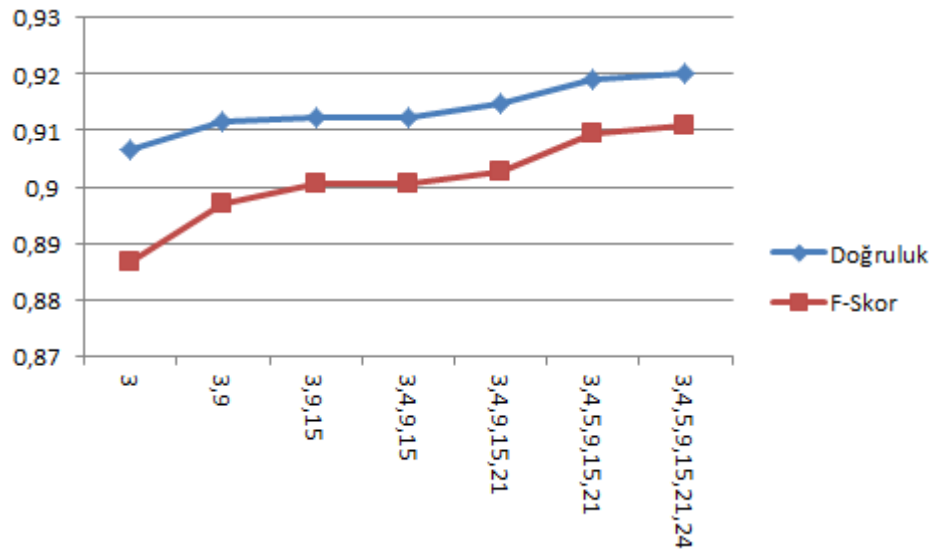
Şekil 5.11. ÇAEBS en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Korelasyon)

Sıralı ileri seçimin DVM algoritması ile uygulanması sonucu seçilen öznitelikler Tablo 5.19’da verilmiştir.

Tablo 5.19. Korelasyon ile seçilmiş veri setinde DVM ile ileri aşamalı seçim sonucu

Öznitelik kod numarası	Korelasyon ile seçilen öznitelikler
3	Summary stage 2000
4	Derived AJCC Stage Group, 7th ed
5	Derived AJCC T, 7th ed
9	Reason no cancer-directed surgery
15	Breast Subtype
21	'CS site-specific factor 7
24	Age at diagnosis

Seçilen 7 öznitelik ile DVM algoritmasının elde ettiği doğruluk değeri % 92,0119, F-Skor değeri ise 0,9108’dir. Şekil 5.12’da seçilen öznitelik kümelerinin sırasıyla doğruluk ve F-Skor değerleri verilmiştir.



Şekil 5.12. DVM en iyi öznitelikler kümesi doğruluk ve F-Skor değer grafiği (Korelasyon)

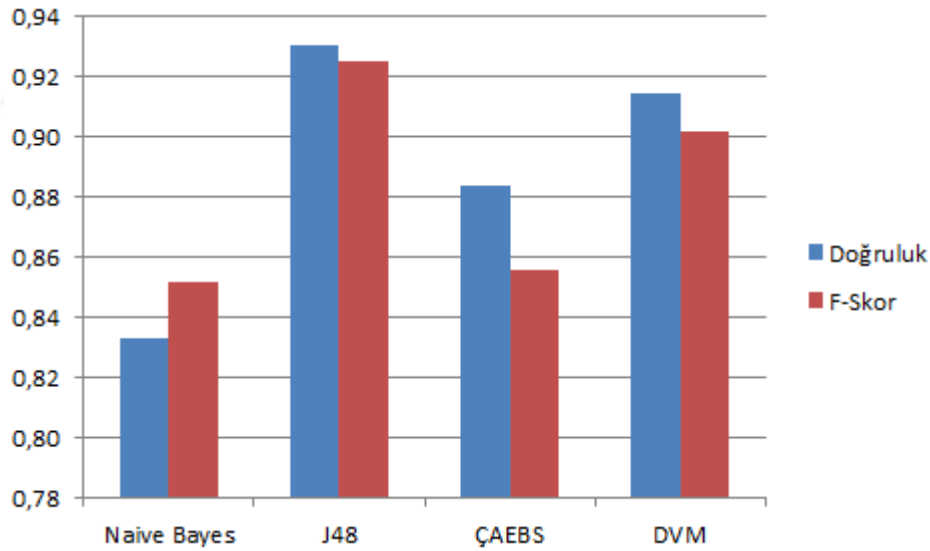
6. DEĞERLENDİRME

6.1. Veri Setinin Tamamı İle Yapılan Çalışma Değerlendirmeleri

Bu çalışmada meme kanseri hastalarının sağ kalımı SEER meme kanseri veri seti üzerinde çalışılarak sınıflandırma algoritmaları kullanılarak tespit edilmeye çalışılmıştır. Gerek sınıflandırma algoritmalarının başarısını artırabilmek, gerekse fazla sayıda olan özniteliklerin sayısını azaltmak adına öznitelik seçimi yapılarak elde edilen sonuçlar incelenmiştir.

Çalışmaya SEER meme kanseri veri seti incelenerek başlanmış ve yapılan ön çalışma neticesinde 73 öznitelik ve 1 sınıf değerinden oluşan ve toplam 370.153 kayıt içeren bir veri seti elde edilmiştir. Veri setindeki kayıtların 246.796'sı eğitim, 123.384'ü ise test için ayrılmıştır. İlk olarak elde edilen veri seti üzerinde Naive Bayes, J48, ÇAEBS ve DVM sınıflandırma algoritmaları uygulanarak sonuçlar incelenmiştir.

Doğruluk ve F-Skor değerlerinin grafiksel olarak karşılaştırması Şekil 6.1'deki gibidir.



Şekil 6.1. 73 öznitelikle elde edilmiş doğruluk ve F-Skor değerleri

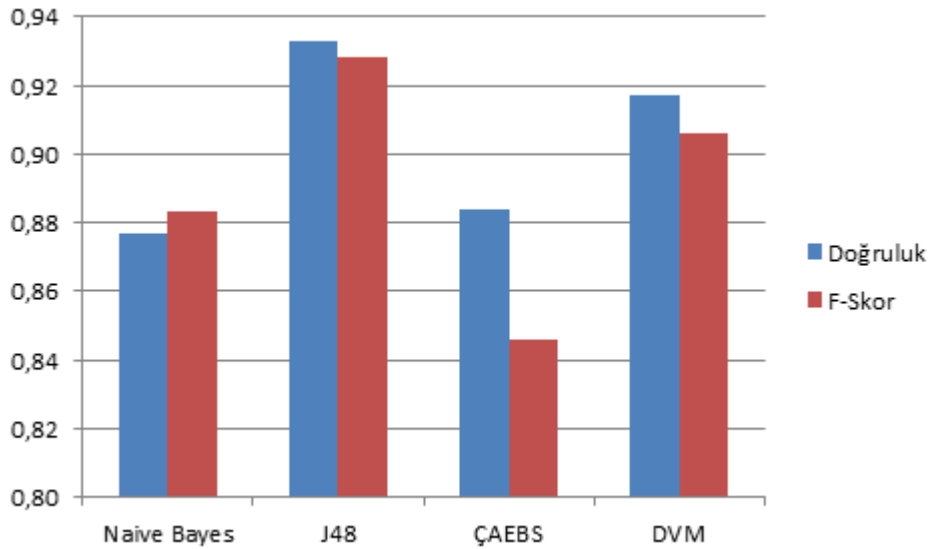
Çalışmanın bu aşaması için sonuç olarak, 123.384 kayıttan 114.776 kaydı doğru sınıflandıran J48 algoritması %93,0234 ile en yüksek başarı oranını sergilemiştir, dolayısıyla diğer denenen algoritmalara göre oldukça başarılı bir performans göstermektedir. Yine 123.384 kayıttan 112.816 doğru kayıt sınıflandırması ile yüksek başarı gösteren ikinci algoritma %91,4349 ile DVM algoritmasıdır. Fakat DVM algoritmasının çalışma süresi 17830,39 saniye ile en yavaş çalışan algoritmadır.

Çalışma süresi açısından 8,2 saniye ile hızlı bir algoritma olmasına rağmen Naive Bayes algoritması %83,3033 ile en düşük başarı gösteren sınıflandırma algoritmasıdır.

6.2. Filtre Öznitelik Seçim Yöntemleriyle Yapılan Çalışma Değerlendirmeleri

Çalışmanın öznitelik seçimi aşamasında, veri setine 3 ayrı filtre öznitelik seçim yöntemi uygulanmıştır. Bu filtre öznitelik seçim yöntemleri Bilgi Kazanımı, Kazanım Oranı ve Korelasyon Tabanlı öznitelik seçimi yöntemleridir. Öznitelik seçimi sonucunda 3 ayrı veri seti elde edilmiş ve bu veri setleri üzerinde ayrı ayrı Naive Bayes, J48, ÇAEBS ve DVM uygulanmıştır.

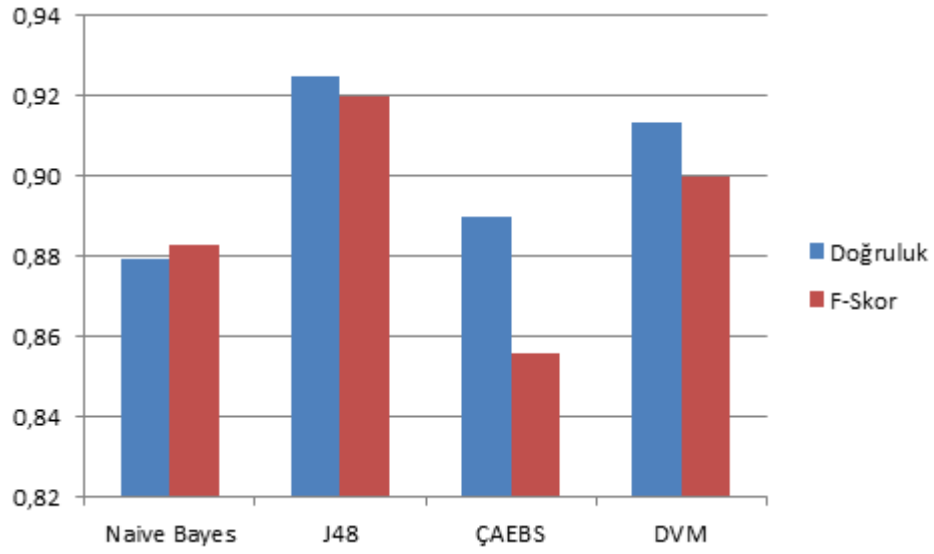
Bilgi kazanımı yöntemi ile elde edilmiş veri setine sınıflandırma algoritmalarının uygulanması sonucu elde edilen başarıların grafiksel gösterimi Şekil 6.2'deki gibidir.



Şekil 6.2. Bilgi Kazanımı öznitelik seçim yöntemi ile seçilen özniteliklerle sıralama algoritmaları doğruluk ve F-Skor karşılaştırması

Bilgi kazanımı yöntemi ile elde edilmiş öznitelikler ile yapılan çalışma sonucunda en yüksek başarı gösteren sınıflandırma algoritması %93,269 ile J48 olmuştur. En düşük başarıyı ise %87,7123 ile Naive Bayes göstermiştir.

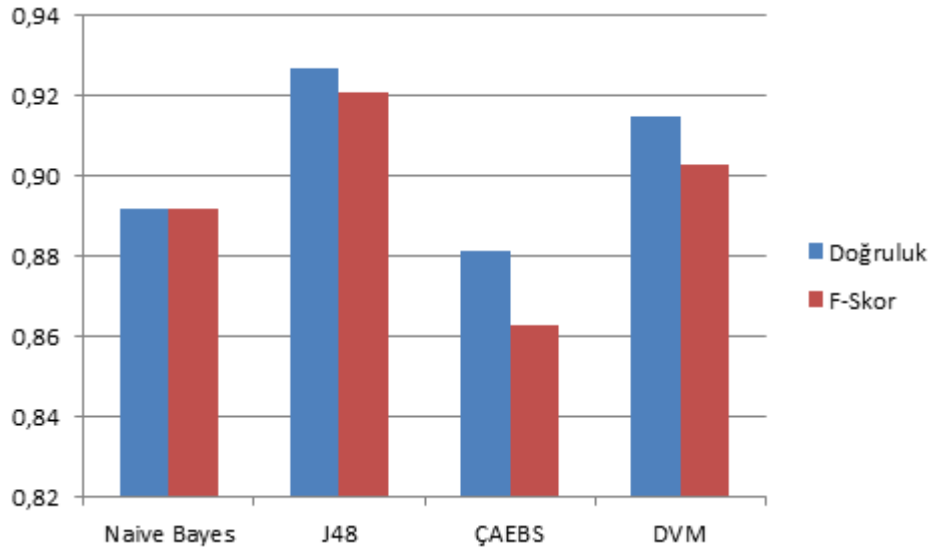
Kazanım Oranı yöntemi ile elde edilmiş veri setine sınıflandırma algoritmalarının uygulanması sonucu elde edilen başarıların grafiksel gösterimi Şekil 6.3'deki gibidir.



Şekil 6.3. *Kazanım Oranı öznitelik seçim yöntemi ile seçilen özniteliklerle sıralama algoritmaları doğruluk ve F-Skor karşılaştırması*

Kazanım Oranı yöntemi ile elde edilmiş öznitelikler ile yapılan çalışma sonucunda en yüksek başarı gösteren sınıflandırma algoritması %92,4796 ile J48 olmuştur. En düşük başarıyı ise %87,9142 ile Naive Bayes göstermiştir.

Korelasyon tabanlı yöntem ile elde edilmiş veri setine sınıflandırma algoritmalarının uygulanması sonucu elde edilen başarıların grafiksel gösterimi Şekil 6.4'deki gibidir.



Şekil 6.4. Korelasyon Tabanlı öznitelik seçim yöntemi ile seçilen özniteliklerle sıralama algoritmaları başarı karşılaştırması

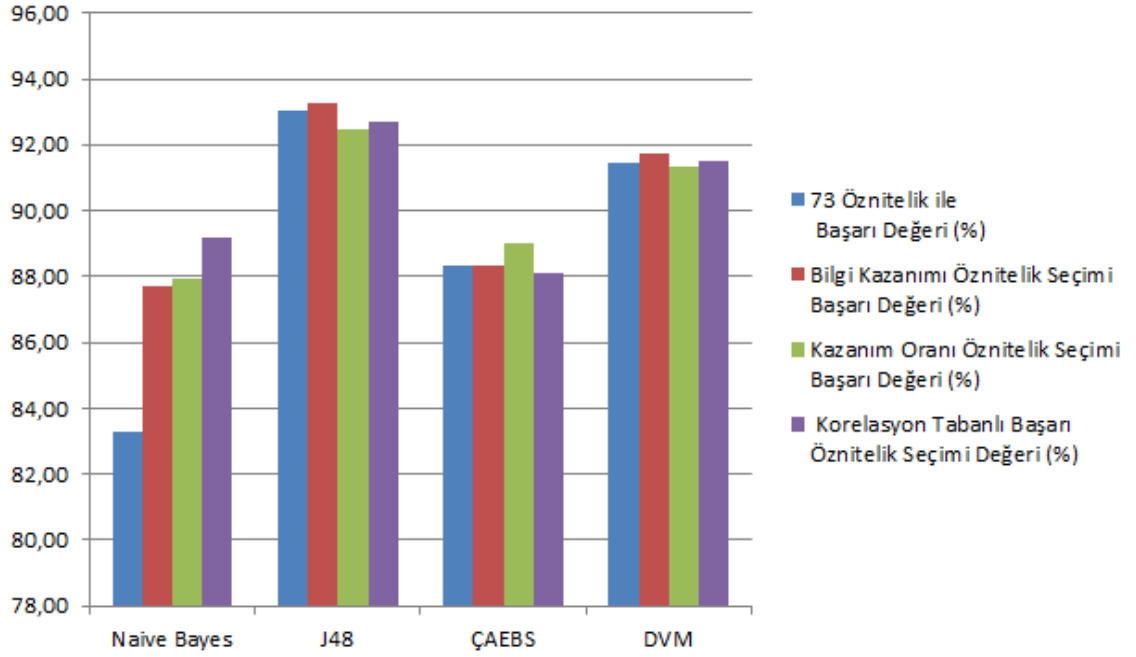
Korelasyon tabanlı yöntem ile elde edilmiş öznitelikler ile yapılan çalışma sonucunda en yüksek başarı gösteren sınıflandırma algoritması %92,696 ile J48 olmuştur. En düşük başarıyı ise %88,1289 ile ÇAEBS göstermiştir.

3 ayrı veri üzerinde çalıştırılmış sınıflandırma algortimalarının başarıları birbirleriyle ve 73 öznitelikle yapılan çalışma ile karşılaştırılmış, Tablo 6.1 ve Şekil 6.5'te gösterilen sonuçlar elde edilmiştir.

Tablo 6.1. Öznitelik kümelerine göre sınıflandırma algoritmalarının başarı yüzdeleri

Algoritma	73 öznitelik ile (%)	Bilgi Kazanımı (%)	Kazanım Oranı(%)	Korelasyon (%)
Naive Bayes	83,3033	87,7123	87,9142	89,2028
J48	93,0234	93,269	92,4796	92,696
ÇAEBS	88,3559	88,3551	89,0018	88,1289
DVM	91,4349	91,7372	91,3619	91,5038

Yapılan çalışma sonucunda J48 ve DVM algoritmaları Bilgi Kazanımı öznitelik seçim yöntemi algoritması ile seçilmiş özniteliklerle en yüksek başarıyı gösterirken, Naive Bayes algoritması en yüksek başarıyı Korelasyon Tabanlı yöntem ile seçilmiş öznitelikler ile göstermişlerdir. ÇAEBS ise Kazanım Oranı öznitelik seçim yöntemiyle seçilmiş özniteliklerle en yüksek başarısını elde etmiştir.



Şekil 6.5. Filtre öznitelik seçim yöntemleri doğruluk karşılaştırması

Naive Bayes sınıflandırma algoritması en yüksek başarıyı Korelasyon Tabanlı öznitelik seçim yöntemiyle seçilmiş veri seti ile göstermiştir. J48 ve DVM ise en yüksek başarıyı Bilgi Kazanımı ile seçilmiş veri seti ile göstermiştir. ÇAEBS ise Kazanım Oranı öznitelik seçim yöntemiyle seçilmiş veri seti ile göstermiştir. Naive Bayes, J48, ÇAEBS ve DVM'in en yüksek doğruluk değerleri sırasıyla %89,2028, %93,269, %89,0018 ve %91,7372'dir.

6.3. Sarmalama Öznitelik Seçim Yöntemiyle Yapılan Çalışma Değerlendirmeleri

Daha önceki çalışmalarda sağ kalım konusu ile SEER meme kanseri veri seti üzerinde birçok çalışma yapılmış ve çalışmalarda çoğunlukla farklı özniteliklerden oluşan veri setleri kullanılmıştır.

Daha önceki yapılan çalışmalarda en çok kullanılan öznitelikler Tablo 6.2’de verilmiştir.

Tablo 6.2. Geçmişte yapılan meme kanseri sağ kalım çalışmalarında kullanılan öznitelikler

Öznitelik	Delen vd.	Tee ve Gazala	Jahanbazi ve Nadimi	Ya-Qin vd.	Rathore vd.	Park vd.	Kate ve Nadig	Bellaachia ve Guven	Endo vd.
Irk	X	X	X	X	X	X	X	X	X
Davranış	X	X	X	-	-	X	X	X	X
Sınıf	X	X	X	X	-	X	X	X	X
Histolojik Tip	X	X	X	X	-	X	X	-	-
Medeni Durum	X	X	X	X	X	X	X	X	X
Birincil Bölge	X	X	X	X	X	X	X	X	-
Tümör Uzantısı	X	X	X	X	X	X	X	X	-
Bölgeye Özel Ameliyat Kodu	X	X	X	X	-	X	X	X	-
Tümörün Bulunduğu Taraf	-	-	-	-	X	-	-	-	-
Lenf Nodu İçerim Durumu	X	X	X	X	X	X	X	X	-
Radyasyon	X	X	X	X	X	X	X	X	X
Kanser Evresi	X	-	X	X	-	X	X	X	X

Tablo 6.2. (Devam) Geçmişte yapılan meme kanseri sağ kalım çalışmalarında kullanılan öznitelikler

Öznitelik	Delen vd.	Tee ve Gazala	Jahanbazi ve Nadimi	Ya-Qin vd.	Rathore vd.	Park vd.	Kate ve Nadig	Bellaachia ve Guven	Endo vd.
Yaş	X	X	X	X	X	-	X	X	X
Tümör Boyutu	X	X	X	X	X	X	X	X	-
Nod Sayısı	X	-	X	X	-	X	X	X	-
Pozitif Nod Sayısı	X	X	X	X	-	X	X	X	-
Primer Nod Sayısı	X	-	X	X	X	X	-	X	X
Progesteron	-	X	-	-	-	-	-	-	-
Östrojen	-	X	-	-	-	-	-	-	-
Ameliyat Edilmeme Nedeni	-	-	X	-	-	-	-	-	X
İlk Kötü Huylu Tümör İndikatörü	-	-	X	X	X	-	-	-	-
Cerrahi Radyasyon Sırası	-	-	-	-	X	-	-	-	X

Tablodan görüldüğü üzere Irk, Birincil Bölge, Medeni Durum, Lenf Nodu İçerim Durumu gibi bazı öznitelikler çoğu çalışmada veri setine dahil edilmiştir. Progesteron, Östrojen, Ameliyat Edilmeme Nedeni gibi birkaç önemli olabilecek öznitelik ise sadece bir veya birkaç çalışmada kullanılmıştır.

6.3.1. Naive Bayes Algoritması Sonuçlarının Değerlendirmesi

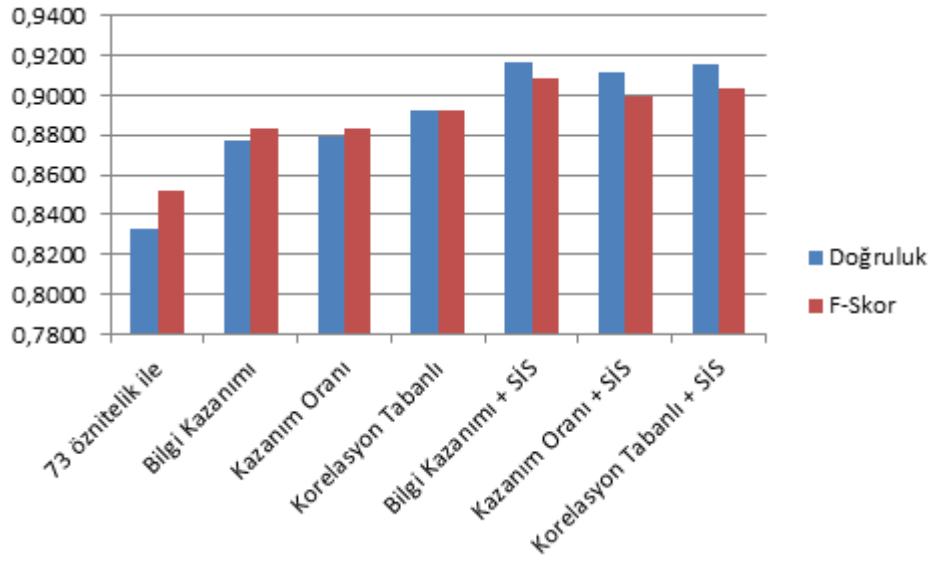
Naive Bayes ile çalıştırılmış Sıralı İleri Seçim algoritmasının seçtiği özniteliklerin karşılaştırılması Tablo 6.3’de verilmiştir. İlgili öznitelik literatürde yapılan çalışmalarda da kullanıldıysa “Literatür” sütununda “X” sembolü ile belirtilmiştir.

Tablo 6.3. Naive Bayes sınıflandırma algoritması için seçilen meme kanseri öznitelikleri

Öznitelik Açıklaması	Literatür	Bilgi Kazanımı + SİS	Kazanım Oranı + SİS	Korelasyon Tabanlı + SİS
Yaş	X	Age at diagnosis	-	Age at diagnosis
Davranış	X	Behavior recode for analysis	Behavior code ICD-O-3	Behavior recode for analysis
Kanser Evresi	X	SEER Combined Summary Stage 2000	Summary stage 2000	Summary stage 2000
Sınıf	X	Grade	-	-
M Evresi	-	Derived AJCC M, 6th ed	-	-
Östrojen	X	ER Status Recode Breast Cancer	-	ER Status Recode Breast Cancer
Progesteron	X	-	PR Status Recode Breast Cancer	-
Kanser Bölgesi	-	-	AYA site recode/WHO	-
Beyinde Metastaz	-	-	SEER Combined Mets at DX-brain	-
Kemikte Metastaz	-	-	-	SEER Combined Mets at DX-bone
Ameliyat edilmeme nedeni	X	Reason no cancer-directed surgery	Reason no cancer-directed surgery	Reason no cancer-directed surgery

Üç veri setine de Sıralı İleri Seçim Naive Bayes ile uygulandığında ortak olarak seçilen öznitelikler Tümör Davranışı, Kanser Evresi ve Ameliyat Edilmeme nedeni ile ilgili öznitelikler olmuştur. Bilgi Kazanımı ve Sıralı İleri Seçim Sonucu seçilen öznitelik sayısı 7, Kazanım Oranı ve Sıralı İleri Seçim Sonucu seçilen öznitelik sayısı 6, Kazanım Oranı ve Sıralı İleri Seçim Sonucu seçilen öznitelik sayısı 6’dır.

Bu öznitelikler ile Naive Bayes algoritması çalıştırıldığında elde edilen başarı oranları, 73 öznitelikle elde edilen başarı oranı ve sadece filtre öznitelik seçim yöntemleri uygulanarak seçilen öznitelikler ile başarı oranlarının grafiksel karşılaştırması Şekil 6.6’de verilmiştir.



Şekil 6.6. Naive Bayes algoritması doğruluk ve F-Skor karşılaştırması

Naive Bayes algoritması ile en yüksek başarıyı elde eden veri seti Bilgi Kazanımı ve Sıralı İleri Seçim algoritması kombinasyonu ile elde edilen 7 öznitelikli veri seti olmuştur. %91,6747 doğruluk oranı elde edilmiştir. Bilgi Kazanımı ve Sıralı ileri Seçim kombinasyonu olan yöntemde seçilen özniteliklerden M Evresi ve Sınıf diğer hibrit yöntemlerle seçilen öznitelikler arasında bulunmamaktadır. Bu durumda bahsedilen öznitelikler Naive Bayes algoritmasında daha yüksek başarı elde edilmesine katkı sağlamıştır.

6.3.2. J48 Algoritması Sonuçlarının Değerlendirmesi

Bilgi Kazanımı, Kazanım Oranı ve Korelasyon Tabanlı öznitelik seçim yöntemleri ile oluşturulmuş veri setleri üzerine Sıralı İleri Öznitelik Seçim algoritmasının J48 ile uygulanması sonucunda seçilen öznitelikler Tablo 6.4’da verilmiştir. İlgili öznitelik literatürde yapılan çalışmalarda da kullanıldıysa “Literatür” sütununda “X” sembolü ile belirtilmiştir.

Tablo 6.4. J48 sınıflandırma algoritması için seçilen meme kanseri öznitelikleri

Öznitelik Açıklaması	Literatür	Bilgi Kazanımı + SİS	Kazanım Oranı + SİS	Korelasyon Tabanlı + SİS
Yaş	X	Age at diagnosis	-	Age at diagnosis
Davranış	X	Behavior recode for analysis	-	-

Tablo 6.4 (Devam) J48 sınıflandırma algoritması için seçilen meme kanseri öznitelikleri

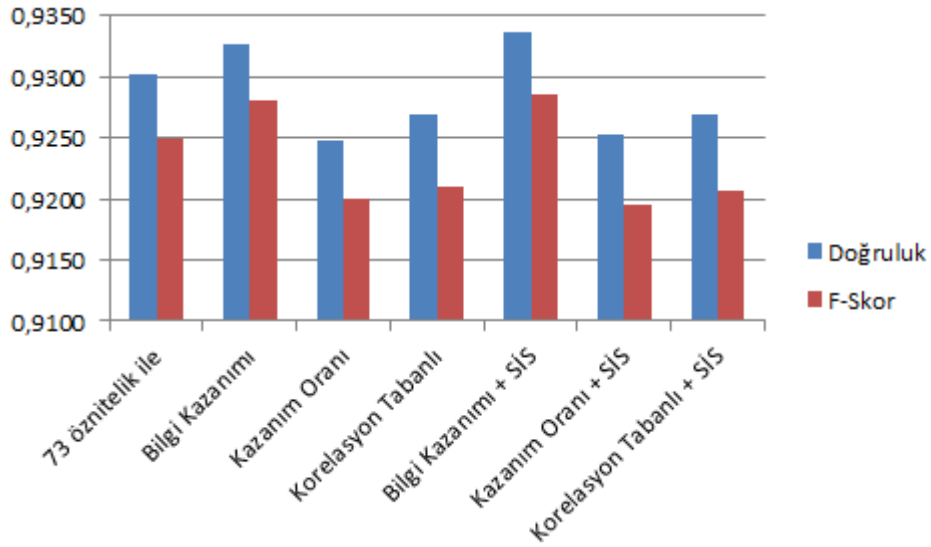
Öznitelik Açıklaması	Literatür	Bilgi Kazanımı + SİS	Kazanım Oranı + SİS	Korelasyon Tabanlı + SİS
Kanser Evresi	X	SEER Combined Summary Stage 2000	Summary stage 2000	Summary stage 2000
Histoloji/Davranış (Kötü Huylu)	X	-	ICD-O-3 Hist/behav, malignant	-
TNM evreleme	-	Breast - Adjusted AJCC 6th Stage	-	Derived AJCC Stage Group, 7th ed
T Evresi	-	-	Breast - Adjusted AJCC 6th T	Derived AJCC T, 7th ed
M Evresi	-	Derived AJCC M, 6th ed	-	Derived AJCC M, 6th ed
N Evresi	-	-	Breast - Adjusted AJCC 6th N (1988-2015)	Breast - Adjusted AJCC 6th N
Tümör Boyutu	X	CS tumor size	CS tumor size	-
Lenf Nodu Sayısı	X	-	CS lymph nodes (2004-2015)	CS lymph nodes
Doğum Yılı	-	Year of birth	-	-
Birincil Bölge Ameliyat Prosedürü	-	-	RX Summ--Surg Prim Site	RX Summ--Surg Prim Site
Östrojen	X	-	ER Status Recode Breast	-
Progesteron	X	-	PR Status Recode Breast Cancer	-
Kanser Bölgesi	-	-	AYA site recode/WHO 2008	-
Bölgesel Pozitif Nod Sayısı	X	-	Regional nodes positive	Regional nodes positive
Beyinde Metastaz	-	-	-	Breast Subtype
Kemikte Metastaz	-	-	SEER Combined Mets at DX-bone Cancer	PR Status Recode Breast Cancer
Ameliyat edilmeme nedeni	X	-	Reason no cancer-directed surgery	Reason no cancer-directed surgery
Pozitif Ipsilateral Seviye I-II Aksiller Lenf Nodu Sayısı	-	-	CS site-specific factor 3	CS site-specific factor 3
Metastatik tutulumun uzak bölgeleri	-	-	-	CS mets at dx

Tablo 6.4 (Devam) J48 sınıflandırma algoritması için seçilen meme kanseri öznitelikleri

Öznitelik Açıklaması	Literatür	Bilgi Kazanımı + SİS	Kazanım Oranı + SİS	Korelasyon Tabanlı + SİS
(BR) Puanı	-	-	-	'CS site-specific factor 7

Üç veri setine de sıralı ileri öznitelik seçimi J48 ile uygulandığında ortak olarak seçilen öznitelik Kanser Evresi ile ilgili öznitelik olmuştur. Bilgi Kazanımı ve Sıralı İleri Seçim Sonucu seçilen öznitelik sayısı 7, Kazanım Oranı ve Sıralı İleri Seçim Sonucu seçilen öznitelik sayısı 14, Kazanım Oranı ve Sıralı İleri Seçim Sonucu seçilen öznitelik sayısı 15'tir. Bilgi Kazanımı ve Sıralı İleri Seçim ile seçilmiş öznitelik sayısı hem diğer yöntemlere göre oldukça az, hem de başarısı yüksektir.

Bu öznitelikler ile J48 algoritması çalıştırıldığında elde edilen başarı oranları, 73 öznitelikle elde edilen başarı oranı ve sadece filtre öznitelik seçim yöntemleri uygulanarak seçilen öznitelikler ile başarı oranlarının grafiksel karşılaştırması Şekil 6.7'de verilmiştir.



Şekil 6.7. J48 algoritması doğruluk ve F-Skor karşılaştırması

J48 algoritması ile en yüksek başarıyı elde eden veri seti Bilgi Kazanımı ve Sıralı İleri Seçim algoritması kombinasyonu ile elde edilen veri seti olmuştur. % 93,3589 doğruluk oranı elde edilmiştir. Bilgi Kazanımı ve Sıralı ileri Seçim kombinasyonu olan yöntemde seçilen özniteliklerden Davranış ve Doğum Yılı diğer hibrit yöntemlerle seçilen öznitelikler arasında bulunmamaktadır. Bu durumda bahsedilen öznitelikler J48 algoritmasında daha yüksek başarı elde edilmesine katkı sağlamıştır.

6.3.3. ÇAEBS Algoritması Sonuçlarının Değerlendirmesi

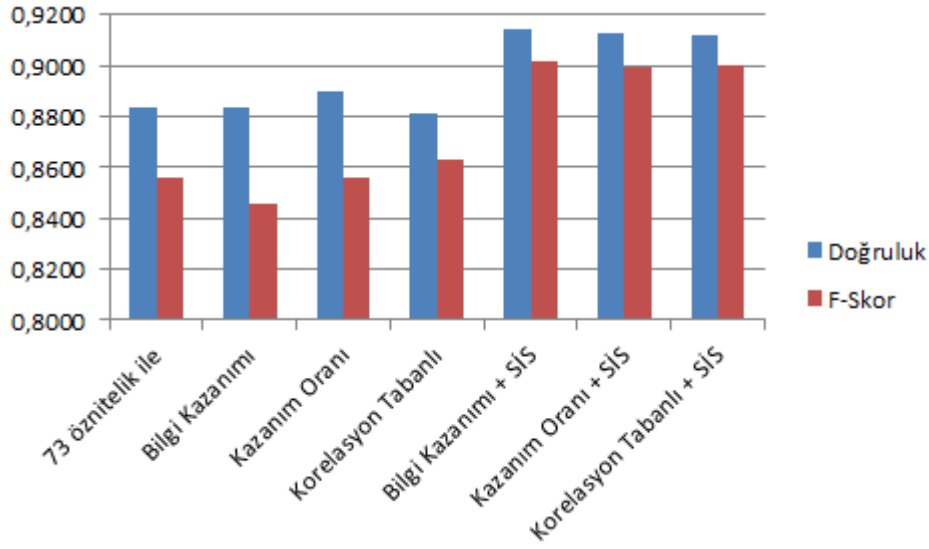
Bilgi kazanımı, kazanım oranı ve korelasyon tabanlı öznitelik seçim yöntemleri ile oluşturulmuş veri setleri üzerine sıralı ileri öznitelik seçim algoritmasının ÇAEBS ile uygulanması sonucunda seçilen öznitelikler Tablo 6.5’de verilmiştir. İlgili öznitelik literatürde yapılan çalışmalarda da kullanıldıysa “Literatür” sütununda “X” sembolü ile belirtilmiştir.

Tablo 6.5. ÇAEBS sınıflandırma algoritması için seçilen meme kanseri öznitelikleri

Öznitelik Açıklaması	Literatür	Bilgi Kazanımı + SIS	Kazanım Oranı + SIS	Korelasyon Tabanlı + SIS
Davranış	X	-	-	Behavior recode for analysis
Kanser Evresi	X	SEER Combined Summary Stage 2000	Summary stage 2000	Summary stage 2000
Bölgesel Pozitif Nod Sayısı	X	-	Regional nodes positive	-
TNM Evreleme	-	Breast - Adjusted AJCC 6th Stage	-	-
Meme Alt Tipi	-	-	-	Breast Subtype
Tümör Uzantısı	X	-	CS extension	-
Progesteron	X	PR Status Recode Breast Cancer	-	-
Ameliyat edilmeme nedeni	X	Reason no cancer-directed surgery	Reason no cancer-directed surgery	Reason no cancer-directed surgery

Üç veri setine de sıralı ileri öznitelik seçimi ÇAEBS ile uygulandığında ortak olarak seçilen öznitelikler Kanser Evresi ve Ameliyat Edilememe Nedeni ile ilgili öznitelikler olmuştur. Her bir veri seti için seçilen öznitelik sayısı 4 olmuştur.

Bu öznitelikler ile ÇAEBS algoritması çalıştırıldığında elde edilen başarı oranlarının grafiksel karşılaştırması Şekil 6.8’de verilmiştir.



Şekil 6.8. ÇAEBS algoritması doğruluk ve F-Skor karşılaştırması

ÇAEBS algoritması ile en yüksek başarıyı elde eden veri seti Bilgi Kazanımı ve Sıralı İleri Seçim algoritması kombinasyonu ile elde edilen veri seti olmuştur. % 91,4154 doğruluk oranı elde edilmiştir. Bilgi Kazanımı ve Sıralı ileri Seçim kombinasyonu olan yöntemde seçilen özneliklerden TNM Evreleme ve Progesteron diğer hibrit yöntemlerle seçilen öznelikler arasında bulunmamaktadır. Bu durumda bahsedilen öznelikler ÇAEBS algoritmasında daha yüksek başarı elde edilmesine katkı sağlamıştır.

6.3.4. DVM Algoritması Sonuçlarının Değerlendirmesi

Bilgi kazanımı, kazanım oranı ve korelasyon tabanlı öznelik seçim yöntemleri ile oluşturulmuş veri setleri üzerine sıralı ileri öznelik seçim algoritmasının DVM ile uygulanması sonucunda seçilen öznelikler Tablo 6.6'da verilmiştir. İlgili öznelik literatürde yapılan çalışmalarda da kullanıldıysa "Literatür" sütununda "X" sembolü ile belirtilmiştir.

Tablo 6.6. DVM sınıflandırma algoritması için seçilen meme kanseri öznelikleri

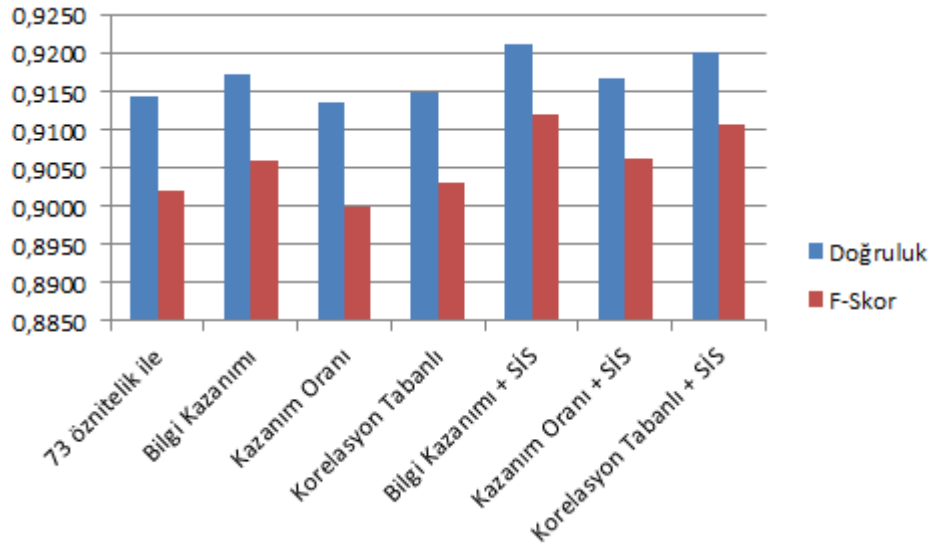
Öznitelik	Literatür	Bilgi Kazanımı + SİS	Kazanım Oranı + SİS	Korelasyon Tabanlı + SİS
Davranış	X	Behavior recode for analysis	-	-
Bölge		-	AYA site recode/WHO 2008	

Tablo 6.6 (Devam) DVM sınıflandırma algoritması için seçilen meme kanseri öznitelikleri

Öznitelik	Literatür	Bilgi Kazanımı + SİS	Kazanım Oranı + SİS	Korelasyon Tabanlı + SİS
Histoloji	X	-	Histology recode - broad groupings	-
Sınıf	X	Grade	-	-
Kanser Evresi	X	SEER Combined Summary Stage 2000	Summary stage 2000	Summary stage 2000
Bölgesel Pozitif Nod Sayısı	X	Regional nodes positive	-	-
TNM Evreleme		Breast - Adjusted AJCC 6th Stage	Breast - Adjusted AJCC 6th Stage	Derived AJCC Stage Group, 7th ed
Meme Alt Tipi		-	-	Breast Subtype
T Evresi		Breast - Adjusted AJCC 6th T	Breast - Adjusted AJCC 6th T	Derived AJCC T, 7th ed
N Evresi		Derived AJCC N, 6th ed	Breast - Adjusted AJCC 6th N (1988-2015)	-
Progesteron	X	PR Status Recode Breast Cancer	-	-
Ameliyat edilmeme nedeni	X	Reason no cancer-directed surgery	Reason no cancer-directed surgery	Reason no cancer-directed surgery
Östrojen	X	ER Status Recode Breast Cancer	ER Status Recode Breast Cancer	-
Yaş	X	Age at diagnosis	-	Age at diagnosis
Hastanın Doğum Yılı		Year of birth	-	-
(BR) Puanı		-	-	'CS site-specific factor 7

Üç veri setine de sıralı ileri öznitelik seçimi DVM ile uygulandığında ortak olarak seçilen öznitelikler Kanser evresi, TNM evreleme, T evresi ve Ameliyat edilememe nedeni ile ilgili öznitelikler olmuştur. Bilgi Kazanımı ve Sıralı İleri Seçim Sonucu seçilen öznitelik sayısı 12, Kazanım Oranı ve Sıralı İleri Seçim Sonucu seçilen öznitelik sayısı 8, Kazanım Oranı ve Sıralı İleri Seçim Sonucu seçilen öznitelik sayısı 7'tir.

Bu öznitelikler ile DVM algoritması çalıştırıldığında elde edilen başarı oranlarının grafiksel karşılaştırması Şekil 6.9'de verilmiştir.



Şekil 6.9. DVM algoritması doğruluk ve F-Skor karşılaştırması

DVM algoritması ile en yüksek başarıyı elde eden veri seti Bilgi Kazanımı ve Sıralı İleri Seçim algoritması kombinasyonu ile elde edilen veri seti olmuştur. % 91,7372 doğruluk oranı elde edilmiştir. 12 tane öznitelik seçilerek %91,7372 başarı göstermiştir. Bunu takip eden en yüksek başarı Korelasyon Tabanlı ve Sıralı İleri Seçim kombinasyonu ile seçilen öznitelikler tarafından sağlanmıştır. Korelasyon Tabanlı ve Sıralı İleri Seçim kombinasyonu tarafından 7 öznitelik seçilmiş ve %92,0119 başarı elde edilmiştir. Kazanım Oranı ve Sıralı İleri Seçim kombinasyonu ile 8 öznitelik seçilmiş ve %91,6755 başarı göstermiştir.

Bilgi Kazanımı ve Sıralı ileri Seçim kombinasyonu olan yöntemde seçilen özniteliklerden Tümör Davranışı, Tümör Sınıfı, Bölgesel Pozitif Nod Sayısı, Progesteron ve Hastanın Doğum Yılı diğer hibrit yöntemlerle seçilen öznitelikler arasında bulunmamaktadır. Bu durumda bahsedilen öznitelikler DVM algoritmasında daha yüksek başarı elde edilmesine katkı sağlamıştır.

6.4. Genel Değerlendirme

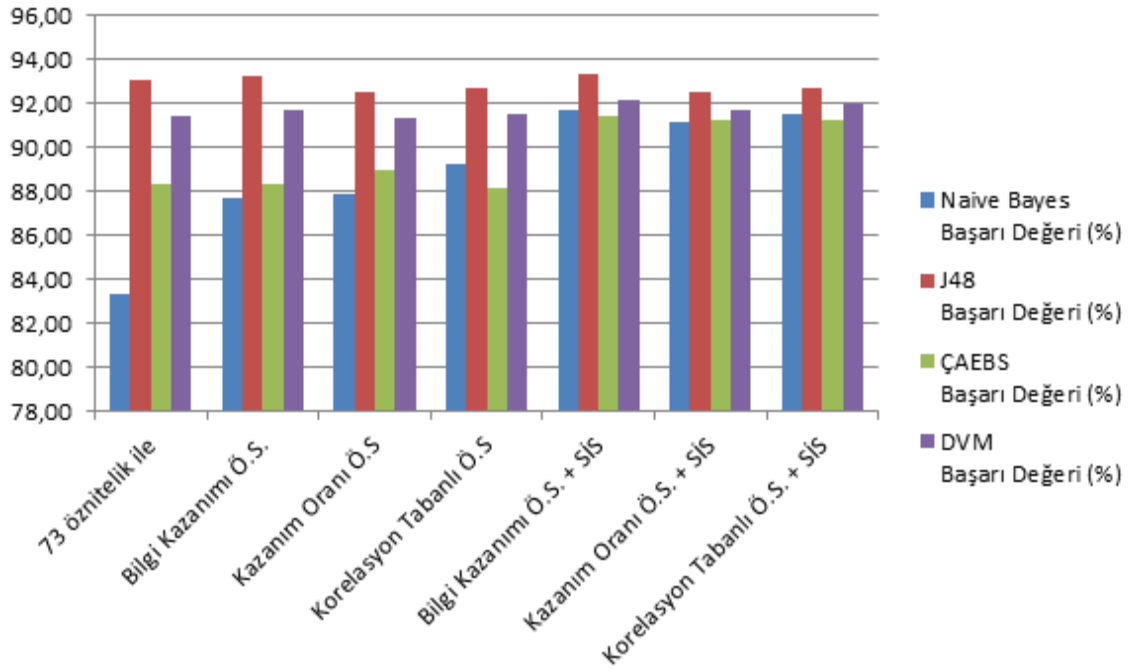
Tablo 6.7’da yapılan çalışmanın özet tablosu verilmiştir. Tabloda başarı oranları ve parantez içinde de kaç tane öznitelik kullanılarak bu başarının elde edildiği gösterilmektedir.

Tablo 6.7. Sınıflandırma algoritmalarının farklı öznitelik seçim yöntemleriyle oluşturulmuş meme kanseri veri seti üzerindeki karşılaştırmalar

Algoritmalar	Metrik	73 öznitelik ile	Bilgi Kazanımı	Kazanım Oranı	Korelasyon Tabanlı	Bilgi Kazanımı + SİS	Kazanım Oranı + SİS	Korelasyon Tabanlı + SİS
Naive Bayes	Doğruluk (%)	83,3033	87,7123	87,9142	89,2028	91,6747	91,1795	91,5118
	F-Skor	0,8520	0,8830	0,8830	0,8920	0,9081	0,8998	0,9035
	Öznitelik Sayısı	73	24	24	24	7	6	6
J48	Doğruluk (%)	93,0234	93,2690	92,4796	92,6960	93,3589	92,5273	92,6886
	F-Skor	0,9250	0,9280	0,9200	0,9210	0,9286	0,9195	0,9207
	Öznitelik Sayısı	73	24	24	24	7	14	15
ÇAEBS	Doğruluk (%)	88,3559	88,3551	89,0018	88,1289	91,4154	91,2573	91,2038
	F-Skor	0,8560	0,8460	0,8560	0,8630	0,9021	0,8993	0,9005
	Öznitelik Sayısı	73	24	24	24	4	4	4
DVM	Doğruluk (%)	91,4349	91,7372	91,3619	91,5038	92,1286	91,6755	92,0119
	F-Skor	0,9020	0,9060	0,9000	0,9030	0,9120	0,9062	0,9108
	Öznitelik Sayısı	73	24	24	24	12	8	7

Tablodan görüldüğü üzere e yüksek başarı Bilgi Kazanımı ve Sıralı İleri Seçim yöntemiyle seçilmiş öznitelikler ile çalıştırılan J48 algoritmasıyla elde edilmiştir. Diğer hibrit yöntemler J48 algoritması için 14 ve 15 öznitelik seçerek daha düşük performans sergilemişlerdir. Diğer taraftan en az öznitelikle en yüksek başarıyı elde eden algoritma başarı Bilgi Kazanımı ve Sıralı İleri Seçim yöntemiyle seçilmiş öznitelikler ile ÇAEBS'dır. Diğer hibrit yöntemlerle de ÇAEBS için 4 öznitelik seçilmiş fakat bu özniteliklerle elde edilen başarı daha düşük kalmıştır.

Şekil 6.10'da tüm yöntemlerin başarıları görsel olarak karşılaştırılmıştır.



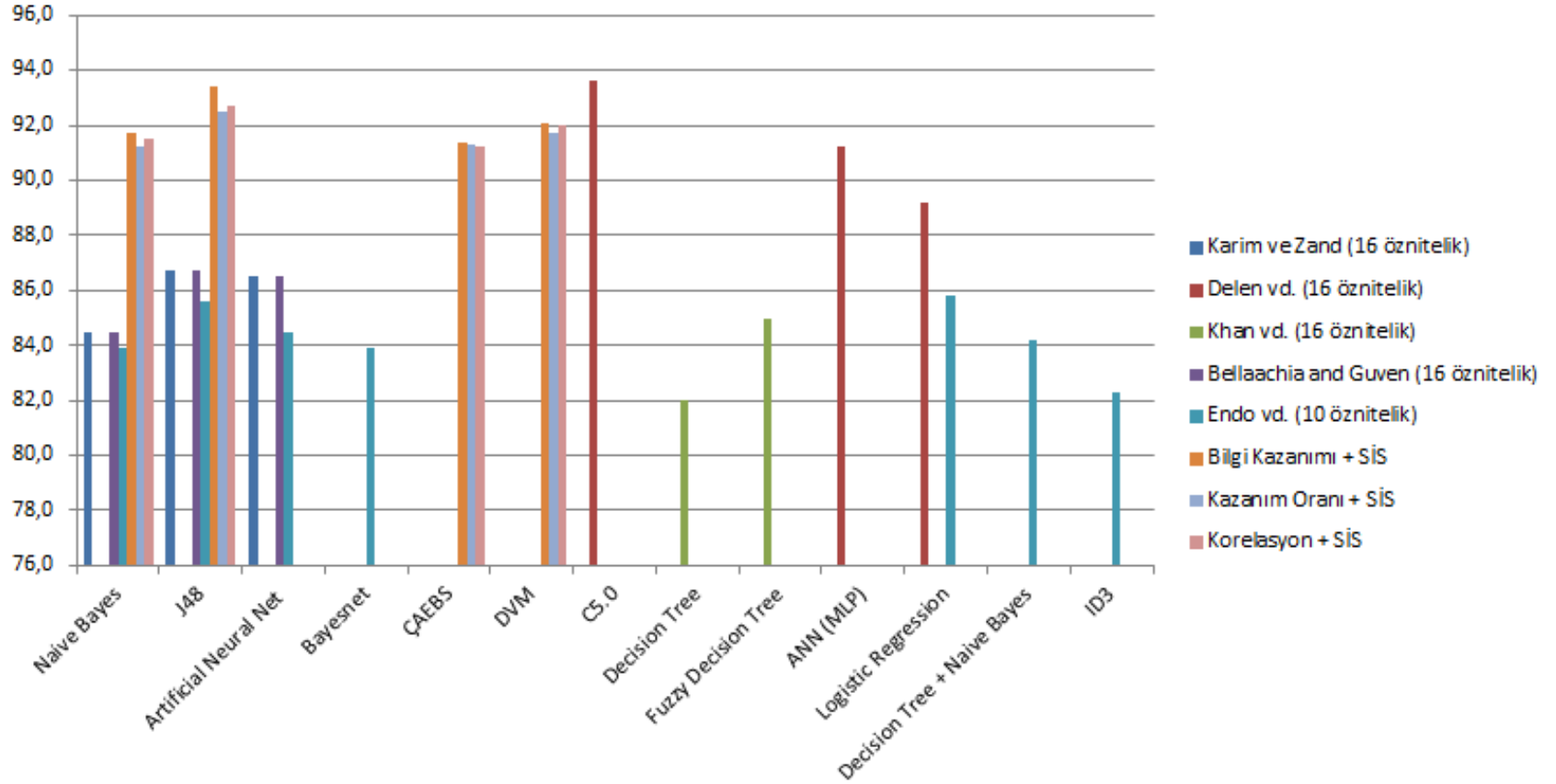
Şekil 6.10. Farklı yöntemlerle seçilmiş öznitelikler ile sınıflandırma algoritmalarının başarı karşılaştırması

Tüm sınıflandırma algoritmaları en yüksek başarılarını Bilgi Kazanımı ve Sıralı İleri Seçimin birlikte kullanılması ile seçilen öznitelikler ile elde etmişlerdir. Filtre ve sarmalama öznitelik seçim yöntemi olan Sıralı İleri Seçimin birlikte kullanılması, filtre yöntemleriyle elde edilen başarıları göre çok daha iyi sonuçlar vermiş aynı zamanda öznitelik kümesini de önemli ölçüde daraltmıştır. En yüksek başarı oranını Bilgi Kazanımı ve Sıralı İleri Seçim ile seçilmiş öznitelikler sağlamasına rağmen diğer yöntemler de aynı sayıda veya daha az öznitelik ile Bilgi Kazanımı ve Sıralı İleri Seçim ile elde edilen başarıları oldukça yaklaştırmışlardır. Örneğin ÇAEBS için 3 hibrit yöntemle de 4 öznitelik seçilmiştir. Bilgi Kazanımı ve Sıralı İleri Seçim ile elde edilen başarı ÇAEBS için %91,4154 iken diğer yöntemlerle elde edilen başarılar %91,2573 ve %91,2038'dir. Yine DVM algoritmasında Bilgi Kazanımı ve Sıralı İleri Seçim 12

öznitelikle %92,1286 başarı elde etmesine rağmen, Kazanım Oranı ve Sıralı İleri Seçim 8 öznitelikle %91,6755, Korelasyon ve Sıralı İleri Seçim 7 öznitelikle %92,0119 başarı sağlamıştır.



Şekil 6.11’de geçmişte yapılan çalışmaların elde ettiği başarılar ile çalışmada filtre ve sarmalama yöntemi kullanılarak elde edilmiş başarıların karşılaştırması verilmiştir.



Şekil 6.11. Geçmişte yapılan çalışmalarla başarı karşılaştırması

Geçmişten bu yana yapılan çalışmalarda en yüksek başarı Delen vd. tarafından bir karar ağacı algoritması olan C5.0 %93,6 olarak elde edilmiştir. Delen vd. çalışmalarında 16 tane öznelik kullanmışlardır. Yapılan bu çalışmada ise en yüksek başarı Bilgi Kazanımı ve Sıralı İleri Seçim kullanılarak seçilen özneliklerle elde edilen veri seti ile J48 algoritması kullanılarak elde edilmiştir. Elde edilen başarı oranı %93,4'tür ve 7 öznelik ile elde edilmiştir. Yapılan deneysel çalışmada yine farklı algoritmalar ve öznelik seçim yöntemleri kullanılarak elde edilen başarılar geçmişte yapılan çalışmalara göre daha yüksektir.



7. SONUÇ

Geçmişten günümüze meme kanserinde sağ kalım için makine öğrenmesi yöntemlerinin başarısı hakkında birçok çalışma yapılmasına rağmen, öznelik seçiminin sağ kalımı tahmin etme üzerindeki etkisi incelenmemiştir. Bu çalışmada makine öğrenmesi yöntemlerinin sağ kalımı tahmin etme üzerindeki başarısının incelenmesinin yanı sıra, öznelik seçiminin bu başarıya etkisi de çalışmanın konusu olmuştur.

Çalışmada SEER meme kanseri veri seti üzerinde 4 farklı makine öğrenmesi algoritması test edilmiştir. Bu algoritmalarından Naive Bayes, J48 ve DVM algoritmaları daha önce yapılan çalışmalarda da test edilmiş olup, ÇAEBS daha önceki çalışmalarda kullanılmayan bir algoritmadır. Veri setindeki tüm özneliklerle yapılan çalışmada J48 algoritmasının %93,0234 ile en yüksek başarıyı elde ettiği görülmüştür. Öznelik seçiminin tahmin üzerindeki etkisini incelemek için Bilgi Kazanımı, Kazanım Oranı ve Korelasyon Tabanlı filtre öznelik seçim yöntemleri sırayla, sarmalama öznelik seçim yöntemlerinden Sıralı İleri Seçim ile birleştirilerek öznelik seçimi yapılmıştır. Öznelik seçimi sonucunda elde edilen özneliklerle oluşturulan veri setleri makine öğrenmesi algoritmalarıyla test edilmiş ve sonuçlar hem sadece filtre öznelik seçimi kullanılarak oluşturulan veri setleriyle elde edilen başarılarla, hem de tüm veri setindeki öznelikler kullanılarak elde edilen başarılarla karşılaştırılmıştır. En yüksek başarıyı elde eden algoritma Bilgi Kazanımı ve Sıralı İleri Seçimin birlikte kullanılmasıyla seçilen özneliklerle oluşturulmuş veri setiyle test edilen J48 algoritmasıdır. Başarısı %93,3589 ve veri setindeki öznelik sayısı 7'dir. En az sayıda öznelikle iyi bir başarı sergileyen algoritma ise 4 öznelikle %91,4154 başarı gösteren ÇAEBS olmuştur. Bahsedilen 4 öznelik Bilgi Kazanımı ve Sıralı İleri Seçimin birlikte kullanılmasıyla elde edilmiştir. Önerilen yöntemlerle seçilen öznelikler arasında geçmişteki çalışmalarda da kullanılan Kanser Evresi, Yaş, Ameliyat Edilmeme Nedeni gibi özneliklerin yanı sıra, daha önceki çalışmalarda hiç kullanılmamış olan Meme Alt Tipi, Beyinde ve Kemikte Metastaz gibi öznelikler bulunmaktadır. Çalışmadaki kullanılan tüm yöntemlerde ortak olarak kullanılan öznelikler Kanser Evresi ve Ameliyat Edilmeme Nedenidir. Bu da meme kanserinde sağ kalımı tahmin etmede geçmişte kullanılanlardan başka özneliklerin de önemli olabileceğini göstermiştir.

Sonuç olarak meme kanseri sağ kalım tahmininde makine öğrenmesi algoritmalarının önerilen öznelik seçim yöntemleri ile seçilmiş özneliklerde daha iyi

sonular elde ettiėi gzlemlenmiř, hem de daha az sayıda znitelikle daha iyi sonular elde edilebileceėi kanıtlanmıřtır.



KAYNAKÇA

- Abdel-Zaher, A.M. ve Eldeib, A.M. (2016). Breast cancer classification using deep belief networks. *Expert Systems With Applications*, 46, 139-144.
- Atak, F. (2014). Gerçek ağ verisi üzerinde veri madenciliği uygulamalarının karşılaştırılması. Yüksek Lisans Tezi. Ankara: Gazi Üniversitesi, Bilişim Enstitüsü.
- Bhargava, N., Sharma, G., Bhargava, R. ve Mathuria, M. (2013). Decision Tree Analysis on J48 Algorithm for Data Mining. *IJARCSSE*, 3, 1114-1119.
- Boughorbel S., Al-Ali R. ve Elkum N. (2016). Model comparison for breast cancer prog- nosis based on clinical data. *PLoS ONE*, 11, 1–15.
- Burke H.B., Goodman P.H., Rosen D.B., Henson D.E., Weinstein J.N., Harrell Jr F.E., Marks, J.R., Winchester, D.P., Bostwick D.G. (1997). Artificial neural networks improve the accuracy of cancer survival prediction. *Cancer*, 79, 857—862.
- Çomak, E. (2008). Destek Vektör Makinelerinin Etkin Eğitimi İçin Yeni Yaklaşımlar. Doktora Tezi. Konya: Selçuk Üniversitesi, Fen Bilimleri Enstitüsü.
- Dash, M. ve Liu, H., (1997). Feature selection for classification. *Intelligent Data Analysis*, 1, 131-156
- Delen, D., Walker, G. Ve Kadam, A. (2005). Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial Intelligence in Medicine*, 34, 113-127.
- Demirhan, T. (2015). Makine Öğrenmesi Algoritmalarının Karmaşıklık Ve Doygunluk Analizinin Bir Veri Kümesi Üzerinde Gerçekleştirilmesi. Doktora Tezi. Edirne: Trakya Üniversitesi, Fen Bilimleri Enstitüsü.
- Endo, A., Shibata, T. ve Tanaka, H. (2008). Comparison of Seven Algorithms to Predict Breast Cancer Survival. *Biomedical Soft Computing and Human Sciences*, 13, 11-16.
- Fan, Q., Zhu, C., Xiao, J., Wang, B., Yin, L., Xu, X. ve Rong, F. (2010). An Application of Apriori Algorithm in SEER Breast Cancer Data. *International Conference on Artificial Intelligence and Computational Intelligence*.
- Güraksın, G.E. (2015). Yapay Zeka Tekniklerini Kullanarak Kemik Yaşı Tespiti. Doktora Tezi. Konya: Selçuk Üniversitesi, Fen Bilimleri Enstitüsü.

- Jiménez, F., Sánchez, G. Ve Juárez, J.M. (2013) Multi-objective evolutionary algorithms for fuzzy classification in survival prediction. *Artificial Intelligence and Medicine*, 60, p. 197-219.
- Karegowda, A.G., Manjunath, A.S. ve Jayaram, M.A. (2010). Comparative Study Of Attribute Selection Using Gain Ratio And Correlation Based Feature Selection. *International Journal of Information Technology and Knowledge Management*, 2 (2), 271-277.
- Kate, R.J. ve Nadig, R. (2017) Stage-specific predictive models for breast cancer survivability. *International Journal of Medical Informatics*, 97, 304-311.
- Khan, M.U. , Choi, J.P. , Shin, H. ve Kim, M. (2008). Predicting breast cancer survivability using fuzzy decision trees for personalized healthcare. *EMBS 30 Yıllık Uluslararası IEEE Tıp ve Biyoloji Derneği Mühendislik Mühendisliği Konferansında yayımlanan bildiri*, 5148–5151.
- Kaynar, O., Arslan, H., Görmez, Y. ve Işık, Y.E. (2018). Makine Öğrenmesi ve Öznitelik Seçim Yöntemleriyle Saldırı Tespiti. *Bilişim Teknolojileri Dergisi*, 11 (2), 175-185.
- Lundin M., Lundin J., Burke H.B., Toikkanen S., Pylkkanen L., Joensuu H. (1999). Artificial neural networks applied to survival prediction in breast cancer. *Oncology*, 57, 281—286.
- Mitchell, T.M. (1997). *Machine Learning*, McGraw-Hill Inc. New York NY-USA.
- Öztürk, S: (2018). Derin Evrimsel Sinir Ağlarında Özellik Seçim Yöntemlerinin Etkisi Yüksek Lisans Tezi. Adana: Çukurova Üniversitesi, Fen Bilimleri Enstitüsü.
- Park K., Ali A., Kim D., An Y., Kim M., Shin H. (2013). Robust predictive model for evaluating breast cancer survivability. *Engl Appl Artif Intell*, 26, 2194–2205.
- Praveena Priyadarsini, R., Valarmathi ve M.L., Sivakumari, S. (2011). Gain Ratio Based Feature Selection Method For Privacy Preservation. *Ictact Journal On Soft Computing*, 01 (4), 201-205.
- Rathore, N., Tomar, D. ve Agarwal, S. (2014) Predicting the survivability of breast cancer patients using ensemble approach. *IEEE Computer Society*, 459–464.
- Ravdin, P. M. ve Clark, G. M. (1992). A practical application of neural network analysis for predicting outcome of individual breast cancer patients. *Breast Cancer Research and Treatment*, 22, 285–293.

- Russell, S.J. ve Norvig, P. (2003). Artificial Intelligence: A Modern Approach, (2. Sürüm). Yayın yeri: Pearson Education.
- Shukla, N., Hagenbuchner, M., Win, K.T., Yang, J. (2018). Breast cancer data analysis for survivability studies and prediction. Computer Methods and Programs in Biomedicine, 155, 199–208.
- Tee, I.C. ve Gazala A.H. (2011). A Novel Breast Cancer Prediction System. IEEE, 978-1-61284-922-5/11, 621-625.
- Var, E. ve İnan, A. (2018). Sınıflandırma için diferansiyel mahremiyete dayalı öznitelik seçimi. Journal of the Faculty of Engineering and Architecture of Gazi University, 33:1, 323-336.
- Ya-Qin, L., Cheng, W. ve Lu,Z. (2009). Decision tree based predictive models for breast cancer survivability on imbalanced data. IEEE, 978-1-4244-2902-8/09, 1-4.
- http-1:<https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/> (Erişim tarihi:10.03.2019)
- http-2:
<http://weka.sourceforge.net/packageMetaData/MultiObjectiveEvolutionaryFuzzyClassifier/index.html> (Erişim tarihi:12.04.2019)
- http-3: <https://seer.cancer.gov/about/overview.html> (Erişim tarihi:07.01.2019)

ÖZGEÇMİŞ

Adı Soyadı : Gizem Yağmur ÖZKAN

Yabancı Dil : İngilizce

Doğum Yeri ve Yılı : 11.09.1988/Niğde

E-Posta : gy_ozkan@anadolu.edu.tr

Eğitim ve Mesleki Geçmişi:

- 2012, Eskişehir Osmangazi Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği
- 2012-2015, Yazılım Mühendisi, Eroğlu Yazılım, Yazılım Birimi
- 2015-Halen, Yazılım Mühendisi, Havelsan A.Ş., Eğitim ve Simülasyon Teknolojileri Birimi

Ödülleri:

- 2012, Birincilik, Bilgisayar Mühendisliği Bölüm Birinciliği, Eskişehir

Mesleki Birlik/Dernek/Kuruluş Üyelikleri:

- 2015, Bilgisayar Mühendisleri Odası, Ankara.

EK- 1. Tek Değer İçeren Öznitelikler

Öznitelik Adı	Açıklama
In research data	Verinin SEER analizlerinde kullanılıp kullanılmadığını gösteren alan
Site recode ICD-O-3/WHO 2008	Birincil bölge ve histoloji kodlarına dayalı yeniden kodlama
Lymphoma subtype recode/WHO 2008	Lenfomada ergen ve genç yetişkinlerde kullanılan birincil bölge/histoloji kodu
TNM 7/CS v0204+ Schema	Site/histoloji bilgisini belirli bir şema altında toplayan alan. AJCC 7. baskı evre ve T, N, M analizlerinde kullanılır
CS Schema - AJCC 6th Edition	Site/histoloji bilgisini belirli bir şema altında toplayan alan. AJCC 6. baskı evre ve T, N, M analizlerinde kullanılır
Histology recode - Brain groupings	Beyin ile ilgili histoloji kodlarıdır.
SEER summary stage 2000 (2001-2003)	2001-2003 yılları arasında kullanılan EOD kodlama sisteminden türetilmiş özet evreleme kodları (2001-2003)
Seer summary stage 1977 (1995-2000)	1995-2000 yılları arasında kullanılan EOD kodlama sisteminden türetilmiş özet evreleme kodları (1995-2000)
AJCC stage 3rd edition (1988-2003)	1988-2003 yılları arasında EOD kodlama sisteminden türetilmiş evreleme kodları (1988-2003).
SEER modified AJCC stage 3rd (1988-2003)	1988-2003 yılları arasında EOD kodlama sisteminden türetilmiş modifiye edilmiş evreleme kodları (1988-2003)
T value – based on AJCC 3rd (1988-2003)	1988-2003 yılları arasında EOD kodlama sisteminden türetilmiş T evreleme kodları (1988-2003)
N value – based on AJCC 3rd (1988-2003)	1988-2003 yılları arasında EOD kodlama sisteminden türetilmiş N evreleme kodları (1988-2003)
M value – based on AJCC 3rd (1988-2003)	1988-2003 yılları arasında EOD kodlama sisteminden türetilmiş M evreleme kodları (1988-2003)
RX Summ – Scope Reg LN Sur (2003+)	Bölgesel Lenf Düğümü Ameliyatının Kapsamı ve prosedürünü açıklayan alan Meme kanseri için kullanılmaz. (2003+)

Öznitelik Adı	Açıklama
Scope of reg lymph nd surg (1998-2002)	Birincil sitenin ameliyatı sırasında veya 1998-2002 tanılı vakalar için lenf düğümlerine yapılan cerrahi işlem prosedürünü tanımlayan alan (1998-2002)
RX Summ – Reg LN Examined (1998-2002)	Gerçekleştirilen ameliyatla birlikte incelenen bölgesel lenf düğümlerinin sayısını kaydeden alan (1998-2002)
Surg of oth reg/dis sited (1988-2002)	1988-2002 yılları arasında kullanılmış uzak lenf nodlarının cerrahi çıkarılma prosedürünü tanımlayan alan (1988-2002)
Site specific surgery (1973-1997varying detail by year and site)	Kansere yönelik olan ya da olmayan tedavinin ilk aşamasında uygulanan cerrahi prosedürü ifade eden alan (1973-1997 yıl ve bölgeye göre değişken)
Lymph-vascular Invasion (2004+ varying by schema)	Penis ve testis ile ilgili şemalarda kullanılan lenfi vasküler yayılımı ifade eden alan (2004+ şemaya göre değişken)
CS site-specific factor 8 (2004+ varying by schema)	Epidermal Büyüme Faktörü reseptörü 2 (HER2) için immünohistokimya laboratuvar değerini kaydeden alan (2004+ şemaya göre değişken)
CS site-specific factor 9 (2004+ varying by schema)	Epidermal Büyüme Faktörü reseptörü 2 (HER2) için immünohistokimya laboratuvar sonuç ve yorumunu kaydeden alan (2004+ şemaya göre değişken)
CS site-specific factor 10 (2004+ varying by schema)	Epidermal Büyüme Faktörü reseptörü 2 (HER2) için FISH test sonuçlarını kaydeden alan (2004+ şemaya göre değişken)
CS site-specific factor 11 (2004+ varying by schema)	Epidermal Büyüme Faktörü reseptörü 2 (HER2) için FISH test yorumunu kaydeden alan (2004+ şemaya göre değişken)
CS site-specific factor 12 (2004+ varying by schema)	Epidermal Büyüme Faktörü reseptörü 2 (HER2) için CISH test sonuçlarını kaydeden alan (2004+ şemaya göre değişken)
CS site-specific factor 13 (2004+ varying by schema)	Epidermal Büyüme Faktörü reseptörü 2 (HER2) için CISH test yorumunu kaydeden alan (2004+ şemaya göre değişken)
CS site-specific factor 16 (2004+ varying by schema)	Östrojen reseptörleri (ER), progesteron reseptörleri (PR) ve İnsan Epidermal Büyüme Faktörü Alıcısı 2 (HER2) bilgilerinin kombinasyonu olan alan (2004+ şemaya göre değişken)
CS site-specific factor 25 (2004+ varying by schema)	Konuma özgü faktörlerin (SSF) tanımlanmadığını gösteren alan (2004+ şemaya göre değişken)
SEER other cause of death classification	Kişinin kanser dışındaki nedenlerden öldüğünü belirten alan
EOD 10 – Prostate path ext (1995-2003)	Prostat kanseri için sadece radikal prostatektomiden elde edilen bilgileri yansıtmayı için ek bir alan (1995-2003)

Öznitelik Adı	Açıklama
EOD 10 – extent (1988-2003)	Tümörün en uzak metastazını belirten alandır. 1988-2003 yılları arasındaki vakalar için kullanılan alan (1988-2003)
EOD 10 – nodes (1988-2003)	Tümörün dahil olduğu en yüksek spesifik lenf düğümü zincirini kaydeden alandır. 1988-2003 yılları arasındaki vakalar için kullanılan alan (1988-2003)
EOD 10 – size (1988-2003)	Primer tümörün en büyük boyutunu milimetre cinsinden kaydeder. 1988-2003 yılları arasındaki vakalar için kullanılan alan (1988-2003)
Tumor marker 1 (1990-2003)	Meme, prostat ve testis vakaları için prognostik göstergeleri kaydeden alandır. 1990-2003 yılları arasındaki vakalar için kullanılmıştır.
Tumor marker 2 (1990-2003)	Meme ve testis vakaları için prognostik göstergeleri kaydeden alandır. 1990-2003 yılları arasındaki vakalar için kullanılan alan (1990-2003)
Tumor marker 3 (1990-2003)	Testis vakaları için prognostik göstergeleri kaydeden alandır. 1990-2003 yılları arasındaki vakalar için kullanılan alan (1990-2003)
Coding system-EOD (1973-2003)	Tümöre uygulanan SEER EOD kodunun türünü gösteren alan (1973-2003)
2-Digit NS EOD part 1 (1973-1982)	Evreleri tanımlamak için kullanılan 2 basamaklı bölgeye özgü olmayan hastalık derecesi kodlama sisteminin ilk basamağını ifade eden alan (1973-1982)
2-Digit NS EOD part 2 (1973-1982)	Evreleri tanımlamak için kullanılan 2 basamaklı bölgeye özgü olmayan hastalık derecesi kodlama sisteminin ikinci basamağını ifade eden alan (1973-1982)
2-Digit SS EOD part 1 (1973-1982)	Evreleri tanımlamak için kullanılan 2 basamaklı bölgeye özgü hastalık derecesi kodlama sisteminin ilk basamağını ifade eden alan (1973-1982)
2-Digit SS EOD part 2 (1973-1982)	Evreleri tanımlamak için kullanılan 2 basamaklı bölgeye özgü hastalık derecesi kodlama sisteminin ikinci basamağını ifade eden alan (1973-1982)
Expanded EOD(1) – CP53 (1973-1982)	Evreleri tanımlamak için kullanılan 13 basamaklı bölgeye özgü hastalık derecesi kodlama sisteminin 1. basamağını ifade eden alan (1973-1982)
Expanded EOD(2) – CP54 (1973-1982)	Evreleri tanımlamak için kullanılan 13 basamaklı bölgeye özgü hastalık derecesi kodlama sisteminin 2. basamağını ifade eden alan (1973-1982)
Expanded EOD(1,2) – CP53,CP54 (1973-1982)	Evreleri tanımlamak için kullanılan 13 basamaklı bölgeye özgü hastalık derecesi kodlama sisteminin 1 ve 2. basamağını ifade eden alan (1973-1982)
Expanded EOD(3) – CP55 (1973-1982)	Evreleri tanımlamak için kullanılan 13 basamaklı bölgeye özgü hastalık derecesi kodlama sisteminin 3 basamağını ifade eden alan (1973-1982)

Öznitelik Adı	Açıklama
Expanded EOD(4) – CP56 (1973-1982)	Evreleri tanımlamak için kullanılan 13 basamaklı bölgeye özgü hastalık derecesi kodlama sisteminin 4. basamağını ifade eden alan (1973-1982)
Expanded EOD(5) – CP57 (1973-1982)	Evreleri tanımlamak için kullanılan 13 basamaklı bölgeye özgü hastalık derecesi kodlama sisteminin 5. basamağını ifade eden alan (1973-1982)
Expanded EOD(6) – CP58 (1973-1982)	Evreleri tanımlamak için kullanılan 13 basamaklı bölgeye özgü hastalık derecesi kodlama sisteminin 6. basamağını ifade eden alan (1973-1982)
Expanded EOD(7) – CP59 (1973-1982)	Evreleri tanımlamak için kullanılan 13 basamaklı bölgeye özgü hastalık derecesi kodlama sisteminin 7. basamağını ifade eden alan (1973-1982)
Expanded EOD(8) – CP60 (1973-1982)	Evreleri tanımlamak için kullanılan 13 basamaklı bölgeye özgü hastalık derecesi kodlama sisteminin 8. basamağını ifade eden alan (1973-1982)
Expanded EOD(9) – CP61 (1973-1982)	Evreleri tanımlamak için kullanılan 13 basamaklı bölgeye özgü hastalık derecesi kodlama sisteminin 9. basamağını ifade eden alan (1973-1982)
Expanded EOD(10) – CP62 (1973-1982)	Evreleri tanımlamak için kullanılan 13 basamaklı bölgeye özgü hastalık derecesi kodlama sisteminin 10. basamağını ifade eden alan (1973-1982)
Expanded EOD(11) – CP63 (1973-1982)	Evreleri tanımlamak için kullanılan 13 basamaklı bölgeye özgü hastalık derecesi kodlama sisteminin 11. basamağını ifade eden alan (1973-1982)
Expanded EOD(12) – CP64 (1973-1982)	Evreleri tanımlamak için kullanılan 13 basamaklı bölgeye özgü hastalık derecesi kodlama sisteminin 12. basamağını ifade eden alan (1973-1982).
Expanded EOD(13) – CP65 (1973-1982)	Evreleri tanımlamak için kullanılan 13 basamaklı bölgeye özgü hastalık derecesi kodlama sisteminin 13. basamağını ifade eden alan (1973-1982)
EOD 4 – extent (1983-1987)	Evreleri tanımlamak için kullanılan 4 basamaklı bölgeye özgü hastalık derecesi kodlama sistemine göre tümörün en uzak metastazını belirten alan (1983-1987)
EOD 4 – nodes (1983-1987)	Evreleri tanımlamak için kullanılan 4 basamaklı bölgeye özgü hastalık derecesi kodlama sistemine göre tümörün dahil olduğu en yüksek spesifik lenf düğümü zincirini kaydeden alan (1983-1987)
EOD 4 – size (1983-1987)	Evreleri tanımlamak için kullanılan 4 basamaklı bölgeye özgü hastalık derecesi kodlama sistemine göre primer tümörün en büyük boyutunu milimetre cinsinden kaydeden alan (1983-1987)
Survival months flag	Hayatta kalma süresi için kullanılan alan
Type of follow up expected	Vakalar için beklenen takip türünü kodlayan alan

Öznitelik Adı	Açıklama
SS seq # - mal+ins (most detail)	SEER veri tabanındaki tüm tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası (çok detay)
SS seq # 1975+ - mal+ins (most detail)	1975 yılı ve sonrası tanı konulan tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası (çok detay)
SS seq # 1992+ - mal+ins (most detail)	1992 yılı ve sonrası tanı konulan tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası (çok detay)
SS seq # 2000+ - mal+ins (most detail)	2000 yılı ve sonrası tanı konulan tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası (çok detay)
Site – mal+ins (most detail)	Tümör bölgesi (çok detay)
SS seq # - mal (most detail)	SEER veri tabanındaki tüm kötü huylu tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası alanıdır (çok detay)
SS seq # 1975+ - mal (most detail)	1975 yılı ve sonrası tanı konulan kötü huylu tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası alanıdır (çok detay)
SS seq # 1975+ - mal (most detail)	1992 yılı ve sonrası tanı konulan kötü huylu tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası alanıdır (çok detay)
SS seq # 2000+ - mal (most detail)	2000 yılı ve sonrası tanı konulan kötü huylu tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası alanıdır. (çok detay)
Site - malignant (most detail)	Kötü huylu tümör bölgesi (çok detay)
SS seq # - mal+ins (mid detail)	SEER veri tabanındaki tüm tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası alanıdır. (orta detay)
SS seq # 1975+ - mal+ins (mid detail)	1975 yılı ve sonrası tanı konulan tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası alanıdır. (orta detay)
SS seq # 1992+ - mal+ins (mid detail)	1992 yılı ve sonrası tanı konulan tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası alanıdır. (orta detay)
SS seq # 2000+ - mal+ins (mid detail)	2000 yılı ve sonrası tanı konulan tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası alanıdır. (orta detay)
Site - mal+ins (mid detail)	Tümör bölgesi (orta detay)

Öznitelik Adı	Açıklama
SS seq # - mal (mid detail)	SEER veri tabanındaki tüm kötü huylu tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası alanıdır. (orta detay)
SS seq # 1975+ - mal (mid detail)	1975 yılı ve sonrası tanı konulan kötü huylu tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası alanıdır. (orta detay)
SS seq # 1992+ - mal (mid detail)	1992 yılı ve sonrası tanı konulan kötü huylu tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası alanıdır. (orta detay)
SS seq # 2000+ - mal (mid detail)	2000 yılı ve sonrası tanı konulan kötü huylu tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası alanıdır. (orta detay)
Site - malignant (mid detail)	Kötü huylu tümör bölgesi (orta detay)
SS seq # - mal+ins (least detail)	SEER veri tabanındaki tüm tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası (az detay)
SS seq # 1975+ - mal+ins (least detail)	1975 yılı ve sonrası tanı konulan tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası (az detay)
SS seq # 1992+ - mal+ins (least detail)	1992 yılı ve sonrası tanı konulan tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası (az detay)
SS seq # 2000+ - mal+ins (least detail)	2000 yılı ve sonrası tanı konulan tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası (az detay)
Site - mal+ins (least detail)	Tümör bölgesi (az detay)
SS seq # - mal (least detail)	SEER veri tabanındaki tüm kötü huylu tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası (az detay)
SS seq # 1975+ - mal (least detail)	1975 yılı ve sonrası tanı konulan kötü huylu tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası (az detay)
SS seq # 1992+ - mal (least detail)	1992 yılı ve sonrası tanı konulan kötü huylu tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası (az detay)
SS seq # 2000+ - mal (least detail)	2000 yılı ve sonrası tanı konulan kötü huylu tümörler göz önüne alınarak tümörün bölgeye özgü dizi numarası (az detay)
Site - malignant (least detail)	Kötü huylu tümör bölgesi (az detay)

EK- 2. Bilgi Kazanımı öznitelik seçim yöntemine göre seçilen özniteliklerle oluşturulmuş veri seti üzerinde Sıralı İleri Seçim

ÖZNİTELİK KOD NUMARALARI

Öznitelik kod numarası	Bilgi Kazanımı yöntemiyle seçilen öznitelikler
1	Behavior recode for analysis
2	AYA site recode/WHO 2008
3	Histologic Type ICD-O-3
4	Grade
5	ICD-O-3 Hist/behav
6	SEER Combined Summary Stage 2000
7	Breast - Adjusted AJCC 6th Stage
8	Derived AJCC N, 6th ed
9	Derived AJCC M, 6th ed
10	Breast - Adjusted AJCC 6th T
11	RX Summ--Surg Prim Site
12	Reason no cancer-directed surgery

Öznitelik kod numarası	Bilgi Kazanımı yöntemiyle seçilen öznitelikler
13	Regional nodes examined
14	Regional nodes positive
15	ER Status Recode Breast Cancer
16	PR Status Recode Breast Cancer
17	CS tumor size
18	CS extension
19	CS lymph nodes
20	CS mets at dx
21	CS site-specific factor 3
22	First malignant primary indicator
23	Age at diagnosis
24	Year of birth

NAIVE BAYES

Öznitelik	6	6, 12	6, 12, 24	6, 12, 15, 24	4, 6, 12, 15, 24	4, 6, 9, 12, 15, 23	1, 4, 6, 9, 12, 15, 23
1	87,5502	90,3455	91,1601	91,3813	91,5896	91,5929	91,6747
2	87,5502	90,3455	91,1601	91,3813	91,5896	91,5929	91,6747
3	87,8160	90,6033	91,0766	91,3894	91,4988	91,4802	91,5540
4	87,5502	90,7848	91,2030	91,5531	91,6350	-	-
5	87,8614	90,3593	91,0709	91,3635	91,4997	91,4794	91,5726
6	90,6649	-	-	-	-	-	-
7	90,4079	89,9395	90,7994	91,2217	91,2881	91,2614	91,4113
8	88,2488	89,7985	90,3261	90,7030	91,1220	90,8318	91,1001
9	90,5214	90,6649	91,1204	91,4794	91,5305	91,6618	-
10	90,5806	90,3699	90,9947	91,1666	91,3060	91,3805	91,4502
11	89,9030	90,8172	90,4728	90,8140	91,0109	90,9761	91,1106
12	89,9054	91,1714	-	-	-	-	-
13	86,9140	89,6777	90,5457	91,0263	91,1171	90,9664	91,1965

Öznitelik		6	6, 12	6, 12, 24	6, 12, 15, 24	4, 6, 12, 15, 24	4, 6, 9, 12, 15, 23	1, 4, 6, 9, 12, 15, 23
14	87,0769	89,9144	90,3658	90,6673	90,8618	90,8116	90,8691	90,8367
15	87,5502	90,8416	91,2865	91,5807	-	-	-	-
16	87,5502	90,8091	91,2646	91,5475	91,5450	91,3878	91,5596	91,4802
17	87,5502	90,3050	90,9664	91,2776	91,3975	91,3124	91,5945	91,4445
18	88,7854	89,9387	90,1389	90,2961	90,5303	90,7143	90,9453	90,7216
19	88,1791	89,9038	90,6195	90,9615	91,0085	90,8237	91,0936	90,9485
20	90,5214	90,6649	91,1195	91,4786	91,5296	91,6609	91,5653	91,6374
21	87,0834	89,8795	90,3763	90,6130	90,7338	90,7897	90,8278	90,8083
22	87,5502	90,3455	91,1601	91,3813	91,5896	91,5929	91,6747	91,6164
23	87,8055	91,0450	91,4648	-	-	-	-	-
24	87,7058	90,9696	91,3935	91,3416	91,4916	91,5142	91,6188	91,5702

J48

Öznitelik		6	6,7	6,7,17	6,7,17,23	6,7,17,23,24	6,7,9,17,23,24	1,6,7,9,17,23,24
1	87,5502	90,6649	91,1990	91,6780	92,0702	93,1887	93,3589	-
2	87,5502	90,6649	91,1990	91,6780	92,0702	93,1887	93,3589	93,3589
3	87,8160	90,7354	91,2395	91,6958	92,0354	93,1028	93,1563	93,1255
4	87,5502	90,7872	91,2654	91,7436	92,1091	93,1336	93,1587	93,1158
5	87,8614	90,7354	91,2395	91,6958	92,0354	93,1028	93,1563	93,1547
6	90,6649	-	-	-	-	-	-	-
7	90,4079	91,1990	-	-	-	-	-	-
8	88,2488	91,0815	91,4624	91,9365	92,3855	93,1522	93,1936	93,1936
9	90,5214	90,6649	91,1990	91,6780	92,0670	93,3435	-	-
10	90,5806	90,9704	91,1852	91,6885	92,0240	93,1449	93,2714	93,2811
11	89,9030	90,9137	91,4210	91,8555	92,1999	92,9366	92,9658	92,9788
12	89,9054	91,1560	91,5710	91,9171	92,2923	93,2584	93,3289	93,3281
13	87,5502	91,0944	91,5296	91,8360	92,2607	93,1068	93,2357	93,2495

Öznitelik		6	6,7	6,7,17	6,7,17,23	6,7,17,23,24	6,7,9,17,23,24	1,6,7,9,17,23,24
14	87,6466	91,1836	91,4899	91,8425	92,3377	93,1652	93,1668	93,0452
15	87,5502	90,8659	91,3278	91,7639	92,1124	93,1960	93,2641	93,2179
16	87,5502	90,8553	91,2760	91,7955	92,1067	93,1498	93,2276	93,1984
17	87,7034	91,1609	91,6780	-	-	-	-	-
18	88,7854	91,1228	91,3141	91,6982	92,0305	93,1409	93,1457	93,1603
19	88,1791	91,0061	91,4454	91,8247	92,1983	92,9772	93,1692	93,1741
20	90,5214	90,6649	91,1990	91,6780	92,0500	93,3411	93,3435	93,1927
21	87,6296	91,1293	91,4608	91,8433	92,3134	93,1879	93,2065	93,2098
22	87,5502	90,6649	91,1990	91,6780	92,0702	93,1887	93,3589	93,3589
23	87,8639	91,1268	91,6585	92,0702	-	-	-	-
24	87,7577	91,0507	91,5604	91,8846	93,1887	-	-	-

ÇAEBS

Öznitelik		6	6,7	6,7,12	6,7,12,16
1	87,5502	90,6649	91,1706	91,1187	90,9850
2	87,5502	90,6649	91,1212	91,0604	90,6454
3	87,8160	90,6957	91,0231	91,1001	91,0766
4	87,5502	90,7872	90,9785	91,0182	90,6632
5	87,7390	90,6973	91,1082	91,0563	91,2192
6	90,6649				
7	90,4079	91,0377			
8	88,2634	90,7808	90,9664	90,9599	90,6478
9	90,5214	90,6649	90,9445	90,9623	90,8180
10	90,5595	90,9988	91,0150	90,7727	90,9299
11	89,9030	90,6624	91,0247	91,1374	90,7281
12	89,9054	90,8845	91,2346		
13	87,5502	90,6649	90,7573	90,8270	91,1698

Öznitelik		6	6,7	6,7,12	6,7,12,16
14	87,6458	90,7986	91,0020	90,8237	91,2654
15	87,5502	90,6649	90,9088	91,0045	90,9420
16	87,5502	90,8537	90,7054	91,4154	
17	87,7091	90,7621	90,8740	90,7532	91,1430
18	88,1402	90,9656	91,1009	91,0150	91,0004
19	87,9068	90,6989	90,7548	91,2776	90,8861
20	90,5214	90,6632	90,7856	91,1593	91,0628
21	87,6296	90,7808	90,7954	91,0425	91,1212
22	87,5502	90,6649	90,9453	90,7759	90,9356
23	87,8639	91,0126	91,1657	90,6649	90,8594
24	87,7577	90,9364	90,8002	90,9785	90,9380

DVM						
lik	6	6,7	6,7,12	6,7,8,12	6,7,8,10,12	6,7,8,10,12,15
87,5502	90,6649	91,1730	91,5945	91,7979	91,8684	91,9641

[illegible]

Öznitelik		6	6,7	6,7,12	6,7,8,12	6,7,8,10, 12	6,7,8,10, 12,15	6,7,8,10, 12,15,23	6,7,8,10, 12,15,23 ,24	6,7,8,10, 12,15,16 ,23,24	6,7,8,10, 12,14,15 ,16,23, 24	4,6,7,8,1 0,12,14, 15,16,23 ,24	4,6,7,8,1 0,12,14, 15,16,23 ,24	1,4,6,7,8 ,10,12,1 4,15,16, 23,24
13	87,5502	90,6649	91,1730	91,5945	91,7955	91,8684	91,9665	92,0135	92,0646	92,0937	92,1043	92,1262	92,1262	92,1270
14	87,5680	90,6973	91,1730	91,5945	91,7955	91,8684	91,9657	92,0127	92,0613	92,1051	-	-	-	-
15	87,5502	90,8659	91,2987	91,6496	91,8409	91,9657	-	-	-	-	-	-	-	-
16	87,5502	90,8553	91,2946	91,6237	91,8530	91,9624	91,9835	92,0459	92,0994	-	-	-	-	-
17	87,5502	90,6632	91,1730	91,5945	91,7955	91,8684	91,9649	92,0143	92,0573	92,1027	92,0994	92,1140	91,9276	91,9365
18	88,7862	91,1260	91,2038	91,5037	91,7112	91,7404	91,7890	91,7963	91,8206	91,8838	91,8830	91,9162	91,9203	91,9300
19	88,1435	90,9566	91,3408	91,6998	91,6763	91,7517	91,7931	91,8109	91,8320	91,9008	91,9033	91,9154	92,1229	92,1197
20	90,5182	90,6649	91,1730	91,5637	91,7971	91,8652	91,9624	91,9892	92,0475	92,0791	92,0978	92,1189	92,1253	92,1278
21	87,5559	90,6795	91,1730	91,5945	91,7955	91,8684	91,9657	92,0111	92,0581	92,1043	92,1043	92,1197	92,1286	92,1213
22	87,5502	90,6649	91,1730	91,5945	91,7979	91,8684	91,9641	91,9892	92,0516	92,0913	92,0970	92,1164	-	-
23	87,5502	90,8440	91,1965	91,6050	91,8020	91,8903	92,0103	-	-	-	-	-	-	-
24	87,5502	90,8488	91,1933	91,6058	91,8003	91,8782	91,9787	92,0589	-	-	-	-	-	-

EK- 3. Kazanım Oranı öznitelik seçim yöntemine göre seçilen özniteliklerle oluşturulmuş veri seti üzerinde Sıralı İleri Seçim

ÖZNİTELİK KOD NUMARALARI

Öznitelik kod numarası	Kazanım Oranı yöntemiyle seçilen öznitelikler
1	AYA site recode/WHO 2008
2	Behavior code ICD-O-3
3	ICD-O-3 Hist/behav, malignant
4	Histology recode - broad groupings
5	Summary stage 2000
6	Breast - Adjusted AJCC 6th Stage
7	Breast - Adjusted AJCC 6th T
8	Breast - Adjusted AJCC 6th N (1988-2015)
9	Breast - Adjusted AJCC 6th M
10	RX Summ--Surg Prim Site
11	Reason no cancer-directed surgery
12	Regional nodes positive

Öznitelik kod numarası	Kazanım Oranı yöntemiyle seçilen öznitelikler
13	SEER Combined Mets at DX-bone
14	SEER Combined Mets at DX-brain
15	SEER Combined Mets at DX-liver
16	SEER Combined Mets at DX-lung
17	ER Status Recode Breast Cancer
18	PR Status Recode Breast Cancer
19	CS tumor size
20	CS extension
21	CS lymph nodes (2004-2015)
22	CS mets at dx (2004-2015)
23	CS site-specific factor 3
24	First malignant primary indicator

NAIVE BAYES

Öznitelik		5	5,11	1,5,11	1,5,11,18	1,2,5,11,18	1,2,5,11,14,18
1	87,5502	90,6649	91,1601	-	-	-	-
2	87,5502	90,6649	91,1601	91,1601	91,1787	-	-
3	87,8614	90,6430	91,0815	91,0823	91,0425	90,9866	90,9882
4	87,6110	90,6333	91,1106	91,1001	91,1147	91,0790	91,0766
5	90,6649	-	-	-	-	-	-
6	90,4079	90,3204	90,5563	90,5644	90,6170	90,0716	90,0724
7	90,5806	90,7224	91,1171	91,0904	90,9785	90,9737	90,9834
8	88,1686	90,3666	90,7589	90,7702	90,5028	90,4890	90,4898
9	90,4517	90,6649	91,0839	91,1601	91,1512	91,1479	91,1382
10	89,9030	91,0571	90,5700	91,0596	91,0482	90,5401	90,5401
11	89,9054	91,1576	-	-	-	-	-
12	87,0769	89,9760	90,3853	90,3828	90,3391	90,3269	90,3374
13	88,6249	90,6673	91,1584	91,1609	91,1520	91,1674	91,1674

Öznitelik		5	5,11	1,5,11	1,5,11,18	1,2,5,11,18	1,2,5,11,14,18
14	87,7196	90,6657	91,1584	91,1609	91,1674	91,1795	-
15	88,0721	90,6657	91,1584	91,1609	91,1487	91,1674	91,1674
16	88,0900	90,6665	91,1584	91,1609	91,1487	91,1674	91,1674
17	87,5502	90,6649	91,1276	91,1617	89,5813	89,5934	89,5764
18	87,5502	90,6649	91,1512	91,1666	-	-	-
19	87,5502	90,6389	91,1406	91,1422	91,1609	90,4258	90,8091
20	88,7854	89,9387	91,0336	90,1048	90,8869	90,9072	90,9080
21	88,1791	88,5673	90,5465	90,5174	90,2645	90,1089	90,1097
22	90,5214	90,6632	91,0839	91,1576	91,1528	91,1349	91,1520
23	87,0834	89,9922	90,4015	90,3990	90,3463	90,3334	90,3342
24	87,5502	90,6649	91,1601	91,1601	91,1787	91,1787	91,1795

	J48						
lik	5	5,11	5,7,11	5,7,11, 19	5,7,11, 19,23	5,7,11, 19,21,23	5,7,11, 17,19, 1,23

[illegible]

Öznitelik	5	5,11	5,7,11	5,7,11, 19	5,7,11, 19,23	5,7,11, 19,21,2 3	5,7,11, 17,19,2 1,23	5,7,11, 17,18,1 9,21,23	5,7,11, 12,17,1 8,19,21 ,23	5,7,8,1 1,12,17 ,18,19, 21,23	5,7,8,1 0,11,12 ,17,18, 19,21, 23	1,5,7,8, 10,11,1 2,17,18 ,19,21, 23	1,3,5,7, 8,10,11 ,12,17, 18,19, 21,23	1,3,5,7, 8,10,11 ,12,13, 17,18, 19,21, 23	
24	87,550	90,665	91,151	91,487	91,834	92,134	92,278	92,444	92,465	92,470	92,506	92,523	92,522	92,526	92,527

ÇAEBS

Öznitelik	5	5,11	5,11,20	5,11,12,20
1	87,5502	90,6649	91,1187	91,1220
2	87,5502	90,6649	91,1406	91,0045
3	87,8557	90,6657	91,0296	90,8602
4	87,61103	90,6713	91,0579	91,0993
5	90,6649	-	-	-
6	90,4079	90,9785	91,0733	91,1017
7	90,5595	90,9461	91,0507	91,0669
8	88,1686	90,6649	91,0515	90,9307

Öznitelik		5	5,11	5,11,20	5,11,12,20
9	90,4517	90,6649	90,9988	91,0896	90,9137
10	89,9030	90,6649	91,0661	90,7929	91,0579
11	89,9054	91,0693	-	-	-
12	87,6458	90,9485	91,0280	91,2573	-
13	88,6249	90,6673	91,0661	91,0507	90,8894
14	87,7196	90,6657	91,1001	91,0863	91,0126
15	88,0721	90,6657	91,0515	91,0215	90,9680
16	88,0900	90,6665	91,1050	91,1058	91,0523
17	87,5502	90,6649	91,1187	91,0879	90,8942
18	87,5502	90,6649	90,9104	91,0620	91,0677
19	87,7091	90,6649	91,0920	90,9420	90,8035
20	88,1402	90,9445	91,1463	-	-
21	87,9068	90,6592	90,9615	90,9696	90,6981

Öznitelik		5	5,11	5,11,20	5,11,12,20
22	90,5214	90,6649	91,0693	91,0507	91,0523
23	87,6296	90,6122	91,0644	91,0725	91,0393
24	87,5502	90,6649	91,1406	91,1025	91,0280

DVM

Öznitelik		5	5,11	5,6,11	5,6,11,17	5,6,8,11,17	5,6,7,8,11,17	4,5,6,7,8,11,17	1,4,5,6,7,8,11,17
1	87,5502	90,6649	91,1512	91,4437	91,5775	91,6066	91,6107	91,6755	-
2	87,5502	90,6649	91,1512	91,4437	91,5775	91,6066	91,6107	91,6755	91,6561
3	87,8639	90,6957	91,1017	91,3821	91,3959	91,3862	91,4300	91,4697	91,4672
4	87,6110	90,6713	91,1487	91,4405	91,5556	91,5588	91,6358	-	-
5	90,6649	-	-	-	-	-	-	-	-
6	90,4079	90,9348	91,4437	-	-	-	-	-	-
7	90,5806	91,0158	91,3805	91,5280	91,5483	91,6131	-	-	-
8	88,1686	90,6649	91,1487	91,4437	91,6083	-	-	-	-

Öznitelik		5	5,11	5,6,11	5,6,11,17	5,6,8,11,17	5,6,7,8,11,17	4,5,6,7,8,11,17	1,4,5,6,7,8,11,17
9	90,4517	90,6649	91,1512	91,4437	91,5791	91,6075	91,6091	91,6755	91,6569
10	89,9030	91,0709	91,1009	91,4429	91,4437	91,5402	91,6253	91,6075	91,6123
11	89,9054	91,1512	-	-	-	-	-	-	-
12	87,5680	90,6649	91,1512	91,4437	91,5791	91,6075	91,5888	91,6366	91,6755
13	88,6249	90,6665	91,1512	91,4413	91,5791	91,5937	91,6042	91,6699	91,6415
14	87,7196	90,6657	91,1512	91,4413	91,5791	91,5953	91,5856	91,6650	91,6415
15	88,0721	90,6649	91,1512	91,4413	91,5791	91,5953	91,6042	91,6650	91,6415
16	88,0900	90,6665	91,1512	91,4413	91,5791	91,5953	91,6042	91,6682	91,6415
17	87,5502	90,6649	91,1512	91,5815	-	-	-	-	-
18	87,5502	90,6649	91,1512	91,5596	91,5604	91,5613	91,6350	91,6561	91,6731
19	87,5502	90,6649	91,1512	91,4437	91,5799	91,6066	91,5912	91,6366	91,6601
20	88,7862	91,0604	91,4129	91,4146	91,5378	91,5572	91,6034	91,5345	91,5694
21	88,1435	90,6527	91,1374	91,4291	91,4300	91,5086	91,5791	91,5450	91,5442

Öznitelik		5	5,11	5,6,11	5,6,11,17	5,6,8,11,17	5,6,7,8,11,17	4,5,6,7,8,11,17	1,4,5,6,7,8,11,17
22	90,5182	90,6649	91,1487	91,4340	91,5637	91,5402	91,6253	91,6147	91,6010
23	87,5559	90,6649	91,1512	91,4437	91,5791	91,6075	91,5888	91,6350	91,6739
24	87,5502	90,6649	91,1512	91,4437	91,5775	91,6066	91,6107	91,6755	91,6561

EK- 4. Korelasyon Tabanlı öznitelik seçim yöntemine göre seçilen özniteliklerle oluşturulmuş veri seti üzerinde Sıralı İleri Seçim

ÖZNİTELİK KOD NUMARALARI

Öznitelik kod numarası	Korelasyon
1	Behavior recode for analysis
2	AYA site recode/WHO 2008
3	Summary stage 2000
4	Derived AJCC Stage Group, 7th ed
5	Derived AJCC T, 7th ed
6	Derived AJCC M, 6th ed
7	Breast - Adjusted AJCC 6th N
8	RX Summ--Surg Prim Site
9	Reason no cancer-directed surgery
10	Regional nodes positive
11	SEER Combined Mets at DX-bone
12	SEER Combined Mets at DX-brain

Öznitelik kod numarası	Korelasyon
13	SEER Combined Mets at DX-liver
14	SEER Combined Mets at DX-lung
15	Breast Subtype
16	ER Status Recode Breast Cancer
17	PR Status Recode Breast Cancer
18	CS lymph nodes
19	CS mets at dx
20	CS site-specific factor 3
21	'CS site-specific factor 7
22	CS site-specific factor 15
23	'First malignant primary indicator
24	Age at diagnosis

NAIVE BAYES

Öznitelik		3	3,9	3,9,24	3,9,16,24	1,3,9,16,24	1,3,9,11,16,24
1	87,5502	90,6649	91,1601	91,3343	91,5005	-	-
2	87,5502	90,6649	91,1601	91,3343	91,5005	91,3975	91,4737
3	90,6649	-	-	-	-	-	-
4	89,1087	90,7467	91,0620	91,3019	91,4551	91,3578	91,4267
5	88,0300	90,5871	90,9429	91,2784	91,4786	91,3911	91,4721
6	90,5214	90,6649	91,0839	91,3611	91,4186	91,4680	91,4421
7	88,1686	90,3666	90,7589	91,0166	91,0377	90,7978	90,9202
8	89,9030	91,0571	90,5700	90,8108	90,9129	91,2136	91,2354
9	89,9054	91,1576	-	-	-	-	-
10	87,0769	89,9760	90,3853	90,6203	90,6333	90,6138	90,6179
11	88,6249	90,6673	91,1584	91,3829	91,4680	91,5118	-
12	87,7196	90,6657	91,1584	91,3789	91,4859	91,4956	91,4899
13	88,0721	90,6657	91,1584	91,3700	91,4916	91,5013	91,4997

Öznitelik		3	3,9	3,9,24	3,9,16,24	1,3,9,16,24	1,3,9,11,16,24
14	88,0900	90,6665	91,1584	91,3692	91,4916	91,5029	91,4980
15	87,5502	90,6665	91,1998	91,4916	91,2606	91,1982	91,2476
16	87,5502	90,6649	91,1276	91,4956	-	-	-
17	87,5502	90,6649	91,1512	91,4899	91,4551	91,2743	91,4000
18	88,1791	88,5673	90,5465	90,9469	90,8334	90,6632	90,7500
19	90,5214	90,6632	91,0839	91,3611	91,4202	91,4680	91,4421
20	87,0834	89,9922	90,4015	90,6284	90,5992	90,6138	90,6000
21	87,5502	90,5384	91,1690	91,4802	91,4721	91,4056	91,4575
22	87,5502	90,6649	91,1617	91,3765	91,4462	91,3278	91,4146
23	87,5502	90,6649	91,1601	91,3343	91,5005	91,3975	91,4737
24	87,8055	90,9647	91,3643	-	-	-	-

J48

Öznitelik	3	3,9	3,9,24	3,9,10, 24	3,9,10, 17,24	3,9,10, 17,18, 24	3,5,9,1 0,17,1 8,24	3,5,9,1 0,15,1 7,18,2 4	3,4,5,9 10,15, 17,18, 24	3,4,5,9 10,15, 17,18, 21,24	3,4,5,8 9,10,1 5,17,1 8,21, 24	3,4,5,8 9,10,1 5,17,1 8,19, 21,24	3,4,5,7 8,9,10 15,17, 18,19, 21,24	3,4,5,6 7,8,9, 10,15, 17,18, 19,21, 24	3,4,5,6 7,8,9, 10,15, 17,18, 19,20, 21,24	
1	87,550	90,665	91,151	91,444	91,611	91,786	91,855	92,013	92,049	92,214	92,573	92,634	92,633	92,646	92,668	92,689
2	87,550	90,665	91,151	91,444	91,611	91,786	91,855	92,013	92,049	92,214	92,573	92,634	92,633	92,646	92,668	92,689
3	90,665	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	89,109	90,747	91,151	91,417	91,590	91,786	91,900	91,940	92,212	-	-	-	-	-	-	-
5	88,030	90,824	91,115	91,457	91,660	91,846	92,013	-	-	-	-	-	-	-	-	-
6	90,521	90,665	91,183	91,444	91,614	91,748	91,910	91,958	92,005	92,180	92,571	92,631	92,646	92,672	-	-
7	88,169	90,665	91,151	91,546	91,684	91,830	91,938	91,983	92,116	92,238	92,606	92,598	92,650	-	-	-
8	89,903	91,071	91,151	91,494	91,620	91,750	91,894	92,004	92,067	92,233	92,632	-	-	-	-	-
9	89,905	91,151	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	87,647	90,959	91,151	91,603	-	-	-	-	-	-	-	-	-	-	-	-
11	88,625	90,667	91,205	91,442	91,609	91,771	91,882	91,989	92,045	92,211	92,569	92,628	92,638	92,652	92,672	92,685

Öznitelik	3	3,9	3,9,24	3,9,10, 24	3,9,10, 17,24	3,9,10, 17,18, 24	3,5,9,1 0,17,1 8,24	3,5,9,1 0,15,1 7,18,2 4	3,4,5,9 10,15, 17,18, 24	3,4,5,9 10,15, 17,18, 21,24	3,4,5,8 9,10,1 5,17,1 8,21, 24	3,4,5,8 9,10,1 5,17,1 8,19, 21,24	3,4,5,7 8,9,10 15,17, 18,19, 21,24	3,4,5,6 7,8,9, 10,15, 17,18, 19,21, 24	3,4,5,6 7,8,9, 10,15, 17,18, 19,20, 21,24	
12	87,720	90,666	91,151	91,442	91,609	91,783	91,862	92,013	92,028	92,212	92,571	92,629	92,638	92,653	92,672	92,689
13	88,072	90,666	91,149	91,442	91,609	91,783	91,864	92,013	92,051	92,215	92,574	92,628	92,645	92,660	92,672	92,689
14	88,090	90,666	91,149	91,442	91,609	91,784	91,864	92,009	92,052	92,215	92,574	92,633	92,642	92,656 2	92,672	92,689
15	87,550	90,796	91,149	91,479	91,775	91,859	91,995	92,047	-	-	-	-	-	-	-	-
16	87,550	90,665	91,245	91,547	91,753	91,752	91,921	91,967	92,059	92,196	92,591	92,620	92,621	92,670	92,685	92,683
17	87,502	90,665	91,173	91,545	91,783	-	-	-	-	-	-	-	-	-	-	-
18	88,179	90,665	91,183	91,460	91,717	91,903	-	-	-	-	-	-	-	-	-	-
19	90,521	90,665	91,151	91,444	91,613	91,767	91,927	91,971	92,020	92,197	92,586	92,635	-	-	-	-
20	87,630	90,764	91,151	91,595	91,587	91,777	91,884	92,013	92,032	92,167	92,593 0	92,628 7	92,628 7	92,668 4	92,689	
21	87,550	90,665	91,162	91,502	91,637	91,819	91,949	92,011	92,125	92,571	-	-	-	-	-	-
22	87,550	90,665	91,201	91,440	91,599	91,766	91,885	92,019	92,063	92,214	92,572	92,625	92,638	92,656	92,672	92,674
23	87,550	90,665	91,172	91,444	91,611	91,786	91,855	92,013	92,049	92,214	92,573	92,634	92,633	92,646	92,668	92,689

Öznitelik	3	3,9	3,9,24	3,9,10, 24	3,9,10, 17,24	3,9,10, 17,18, 24	3,5,9,1 0,17,1 8,24	3,5,9,1 0,15,1 7,18,2 4	3,4,5,9 10,15, 17,18, 21,24	3,4,5,9 10,15, 17,18, 21,24	3,4,5,8 9,10,1 5,17,1 8,21, 24	3,4,5,8 9,10,1 5,17,1 8,19, 21,24	3,4,5,7 8,9,10 15,17, 18,19, 21,24	3,4,5,6 7,8,9, 10,15, 17,18, 19,21, 24	3,4,5,6 7,8,9, 10,15, 17,18, 19,20, 21,24
24	87,864	91,039	91,440	-	-	-	-	-	-	-	-	-	-	-	-

ÇAEBS

Öznitelik	3	3,9	1,3,9	1,3,9,15
1	87,5502	90,6649	91,1406	-
2	87,5502	90,6649	91,1187	91,0693
3	90,6649	-	-	-
4	89,1152	90,7281	91,0498	91,0474
5	88,1759	90,7978	90,8869	90,9907
6	90,5214	90,6649	91,0693	91,0158
7	88,1686	90,6649	91,0515	91,0920
8	89,9030	90,6649	91,0661	91,0693

Öznitelik		3	3,9	1,3,9	1,3,9,15
9	89,9054	91,0693	-	-	-
10	87,6458	90,9485	91,0280	90,8845	90,6341
11	88,6249	90,6673	91,0661	91,0563	90,9955
12	87,7196	90,6657	91,1001	91,1187	91,1463
13	88,0721	90,6657	91,0515	91,1447	91,0847
14	88,0900	90,6665	91,1050	91,1001	90,9429
15	87,5502	90,7962	91,0709	91,2038	-
16	87,5502	90,6649	91,1187	90,9899	91,0377
17	87,5502	90,6649	90,9104	91,1187	90,8026
18	87,9068	90,6592	90,9615	91,0344	91,0320
19	90,5214	90,6649	91,0693	91,1228	91,0450
20	87,6296	90,6122	91,0644	90,7135	91,1050
21	87,5502	90,5384	91,0061	91,0677	90,7054

Öznitelik		3	3,9	1,3,9	1,3,9,15
22	87,5502	90,6649	91,0725	91,1114	91,0142
23	87,5502	90,6649	91,1406	91,0733	91,0944
24	87,8639	90,9323	90,7840	91,0661	90,5879

DVM

Öznitelik		3	3,9	3,9,15	3,4,9,15	3,4,9,15,21	3,4,5,9,15,21	3,4,5,9,15,21,24
1	87,5502	90,6649	91,1512	91,2298	91,3343	91,4348	91,9243	92,0078
2	87,5502	90,6649	91,1512	91,2298	91,3343	91,4348	91,9243	92,0078
3	90,6649	-	-	-	-	-	-	-
4	89,1087	90,7467	91,0936	91,2298	-	-	-	-
5	88,0300	90,8245	91,1795	91,3473	91,4518	91,9235	-	-
6	90,5214	90,6649	91,1512	91,2063	91,3457	91,4340	91,9219	91,9981
7	88,1686	90,6649	91,1487	91,2298	91,3343	91,4834	91,9487	91,9965
8	89,9030	91,0709	91,1009	91,3473	90,9696	91,0069	91,8271	91,8441

Öznitelik		3	3,9	3,9,15	3,4,9,15	3,4,9,15,21	3,4,5,9,15,21	3,4,5,9,15,21, 24
9	89,9054	91,1512	-	-	-	-	-	-
10	87,5680	90,6649	91,1512	91,1495	91,3465	91,4340	91,9357	91,9940
11	88,6249	90,6665	91,1512	91,2298	91,3343	91,3846	91,9154	91,9397
12	87,7196	90,6657	91,1512	91,2152	91,3343	91,3854	91,9154	91,9381
13	88,0721	90,6649	91,1512	91,2152	91,3343	91,3854	91,9171	91,9470
14	88,0900	90,6665	91,1512	91,2152	91,3343	91,3854	91,9146	91,9414
15	87,5502	90,7962	91,2298	-	-	-	-	-
16	87,5502	90,6649	91,1512	91,2152	91,3003	91,3757	91,9551	92,0086
17	87,5502	90,6649	91,1512	91,2273	91,3068	91,3789	91,9543	92,0119
18	88,1435	90,6527	91,1374	91,2273	91,0709	91,0750	91,8498	91,8652
19	90,5182	90,6649	91,1487	91,1706	91,2662	91,3246	91,9008	91,9130
20	87,5559	90,6649	91,1512	91,1982	91,3465	91,4340	91,9316	91,9916
21	87,5502	90,5384	91,1487	91,2298	91,4745	-	-	-
22	87,5502	90,6649	91,1512	91,2014	91,3343	91,4194	91,9316	91,9746

Öznitelik		3	3,9	3,9,15	3,4,9,15	3,4,9,15,21	3,4,5,9,15,21	3,4,5,9,15,21, 24
23	87,5502	90,66499	91,1512	91,2273	91,3343	91,4348	91,9243	92,0078
24	87,5502	90,6649	91,1544	91,2298	91,3651	91,4462	92,0119	-