

KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ELEKTRİK- ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

**KALİTELERİNE GÖRE İÇECEKLERİN FARKLI ALGORİTMALAR İLE
SINIFLANDIRILMASI**

YÜKSEK LİSANS TEZİ

Elektrik- Elektronik Müh. Yeşim ER

OCAK 2017
TRABZON



**KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

ELEKTRİK- ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

**KALİTELERİNE GÖRE İÇECEKLERİN FARKLI ALGORİTMALAR İLE
SINIFLANDIRILMASI**

Elektrik- Elektronik Müh. Yeşim ER

**Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsünde
" ELEKTRİK YÜKSEK MÜHENDİSİ"
Unvanı Verilmesi İçin Kabul Edilen Tezdir.**

Tezin Enstitüye Verildiği Tarih : 07 / 12 / 2016

Tezin Savunma Tarihi : 05 / 01 / 2017

Tez Danışmanı : Doç. Dr. Ayten ATASOY

Trabzon 2017

**KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**Elektrik- Elektronik Mühendisliği Anabilim Dalında
Yeşim ER Tarafından Hazırlanan**

**KALİTELERİNE GÖRE İÇECEKLERİN FARKLI ALGORİTMALAR İLE
SINIFLANDIRILMASI**

**başlıklı bu çalışma, Enstitü Yönetim Kurulunun 13 / 12 / 2016 gün ve 1680 sayılı
kararıyla oluşturulan jüri tarafından yapılan sınavda
YÜKSEK LİSANS TEZİ
olarak kabul edilmiştir.**

Jüri Üyeleri

Başkan : Doç. Dr. Ayten ATASOY

.....

Üye : Yrd. Doç. Dr. Yusuf SEVİM

.....

Üye : Yrd. Doç. Dr. Rifat BENVENİSTE

.....

Prof. Dr. Sadettin KORKMAZ

Enstitü Müdürü

ÖNSÖZ

Bu tez, Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü Elektrik-Elektronik Mühendisliği Anabilim Dalı, Elektrik Mühendisliği Yüksek Lisans Programı'nda yapılan bir çalışmadır.

Öncelikle yüksek lisans tezi danışmanlığımı üstlenerek, gerek konu seçiminde ve gerekse çalışmaların yürütülmesinde yardımlarını esirgemeyen sayın hocam Doç. Dr. Ayten ATASOY'a en içten teşekkürlerimi sunuyorum.

Çalışmam süresince sabrıyla, ilgisiyle ve pozitif enerjisiyle bana destek veren ve her an yanımda olan Eda KARACA'ya teşekkürlerimi sunarım.

Ayrıca bu yaşa gelmemde büyük emekleri geçen sevgilerini ve güvenlerini hiç eksik etmeyen, her an yanımda olan aileme çok teşekkür ederim.

Yeşim ER
Trabzon 2017

TEZ ETİK BEYANNAMESİ

Yüksek Lisans Tezi olarak sunduğum “Kalitelerine Göre İçeceklerin Farklı Algoritmalar ile Sınıflandırılması” başlıklı bu çalışmayı baştan sona kadar danışmanım Doç. Dr. Ayten ATASOY’ un sorumluluğunda tamamladığımı, başka kaynaklardan aldığım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiğimi, çalışma sürecinde bilimsel araştırma ve etik kurallara uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi beyan ederim. 05/01/2017

Yeşim ER

İÇİNDEKİLER

	<u>Sayfa No</u>
ÖNSÖZ	III
TEZ ETİK BEYANNAMESİ.....	IV
İÇİNDEKİLER.....	V
ÖZET	VII
SUMMARY	VIII
ŞEKİLLER DİZİNİ.....	IX
TABLolar DİZİNİ.....	XII
SEMBOLLER DİZİNİ	XV
1. GENEL BİLGİLER	1
1.1. Giriş.....	1
1.2. Elektronik Dilin Tanımı ve Temelleri.....	2
1.3. Elektronik Dil ile Gıda Analizi İçin Örnek Ön İşlem Şartları	3
1.4. Elektronik Dil ile Tat ve Tat Bileşenlerinin Ayrımı	3
1.5. Elektronik Dilin Gıdadaki Uygulama Alanları	4
1.5.1. Zeytinyağı ve Bitkisel Yağlar	4
1.5.2. Süt ve Süt Ürünleri.....	6
1.5.3. Balık ve Et.....	8
1.5.4. Bal	11
1.5.5. Meyve ve Sebzeler	13
1.5.6. Farklı İçecekler	14
1.5.7. Diğer Gıda Bağlantılı Elektronik Dil Uygulamaları	15
1.6. Elektronik Dil ile Su Ortamı Takibi.....	16
1.7. Elektronik Dil ile Süreç İzleme.....	17
1.8. Sınıflandırma Algoritmaları	19
1.8.1. Destek Vektör Makinesi Algoritması	19
1.8.1.1. Lineer Ayrılabilme Durumu	20
1.8.1.2. Lineer Ayrılamama Durumu	23
1.8.1.3. Çok Sınıflı Destek Vektör Makineleri	25
1.8.2. k-En Yakın Komşuluk Algoritması	25

1.8.3.	Rastgele Orman Algoritması.....	27
1.9.	Özellik Seçimi.....	29
1.9.1.	Filtre Modelli Öznitelik Seçme Algoritmaları	29
1.9.1.1.	Bilgi Kazancı	30
1.9.1.2.	Kazanım Oranı	31
1.9.2.	Sarmal Modelli Öznitelik Seçme Algoritmaları	31
1.10.	Boyut İndirgeme	32
1.10.1.	Temel Bileşenler Analizi	32
1.11.	Örnekleme Yöntemleri.....	35
1.11.1.	Azınlıktan Yapay Örnek Üretme Tekniği.....	35
1.11.2.	Rastgele Alt Örnekleme	37
1.12.	Sınıflandırma Performans Ölçütleri	37
1.12.1.	Duyarlılık	38
1.12.2.	Anma.....	38
1.12.3.	F-Ölçeği	39
1.12.4.	Alıcı İşletim Karakteristiği	39
2.	YAPILAN ÇALIŞMALAR, BULGULAR VE İRDELEME.....	40
2.1.	Giriş.....	40
2.2.	Özniteliklerin Tanıtılması	40
2.2.1.	Özniteliklerin Radar Çizimleri.....	43
2.2.2.	Özniteliklerin Kutu Bıyık Grafikleri.....	48
2.2.3.	Özniteliklerin Saçılım Grafikleri	53
2.3.	Eğitim ve Test Veri Setlerinin Hazırlanması	57
2.4.	Sınıflandırma Algoritmalarına Ait Bulgular	58
2.5.	Boyut İndirgemeye Ait Bulgular.....	61
2.6.	Özellik Seçimine Ait Bulgular.....	63
2.6.1.	Filtre Modelli Özellik Seçim Algoritmalarına Ait Bulgular	63
2.6.2.	Sarmal Modelli Özellik Seçim Algoritmalarına Ait Bulgular	77
2.7.	Örnekleme Yöntemlerine Ait Bulgular.....	78
3.	SONUÇLAR	82
4.	KAYNAKLAR	83
ÖZGEÇMİŞ		

Yüksek Lisans Tezi

ÖZET

KALİTELERİNE GÖRE İÇECEKLERİN FARKLI ALGORİTMALAR İLE
SINIFLANDIRILMASI

Yeşim ER

Karadeniz Teknik Üniversitesi
Fen Bilimleri Enstitüsü
Elektrik- Elektronik Mühendisliği Anabilim Dalı
Danışman: Doç. Dr. Ayten ATASOY
2017, 89 Sayfa

Bu tez çalışması kapsamında, kalitelere göre içecekler farklı algoritmalar kullanılarak sınıflandırılmaya çalışılmıştır. Çalışmada kullanılan veri seti, UCI (University of California at Irvine) Makine Öğrenmesi veri tabanından alınmıştır. Veri seti, farklı kalitelerdeki kırmızı ve beyaz şarap örneklerini içeren ayrı iki veri kümesinden oluşmaktadır. İlk olarak, bu iki veri kümesi birleştirilmiş ve şarap örnekleri kırmızı ve beyaz şarap olarak sınıflandırılmıştır. Sonra her bir veri kümesindeki farklı kalitelere sahip kırmızı (6 farklı kalite) ve beyaz (7 farklı kalite) şarap örneklerine ayrı olarak kalite sınıflandırılması yapılmıştır. Daha sonra, 13 farklı kaliteden oluşan kırmızı ve beyaz şarap örneklerine kalite ve renk sınıflandırılması yapılmıştır. Destek Vektör Makinesi, k-En Yakın Komşuluk ve Rastgele Orman bu çalışmada kullanılan sınıflandırma algoritmalarıdır. Veri setine boyut indirgeme yöntemlerinden Temel Bileşenler Analizi, özellik seçme yöntemlerinden filtre tabanlı (Bilgi Kazancı, Kazanım Oranı) ve sarmal tabanlı yöntemler uygulanarak sınıflandırma işlemleri tekrarlanmıştır. Veri setindeki dengesizlikler SMOTE (Synthetic Minority Over-sampling Technique) rastgele örnek artırma ve örnek azaltma algoritmaları kullanılarak giderilmiş ve aynı sınıflandırma işlemleri tekrarlanmıştır. Çalışmada kullanılan performans ölçekleri duyarlılık, anma, F-ölçeği ve alıcı işletim karakteristiğidir.

Anahtar Kelimeler: Elektronik Dil, Destek Vektör Makinesi, k-En Yakın Komşuluk, Rastgele Orman, Sınıflandırma, Özellik Seçme

Master Thesis

SUMMARY

CLASSIFICATION OF BEVERAGES ACCORDING TO THEIR QUALITIES
USING DIFFERENT ALGORITHMS

Yeşim ER

Karadeniz Technical University
The Graduate School of Natural and Applied Sciences
Electrical and Electronics Engineering Graduate Program
Supervisor: Assoc. Prof. Ayten ATASOY
2017, 89 Pages

In the scope of this thesis, beverages are classified according to their qualities using different algorithms. The data set used is taken from the UCI (University of California, Irvine) Machine Learning Database. The data set is composed of two different data sets that include white and red wine samples. Firstly, the two data sets were merged and the wine samples are classified as red and white. Then, classification process were employed separately on red (6 different qualities) and white wine samples (7 different qualities) that were of different qualities. After that, quality and colour classification were applied on 13 different qualities of red and white wine samples. Support Vector Machine, k-Nearest Neighbours and Random Forest are the classification algorithms that are used in this study. Principle Component Analysis were applied on the data set for reduction purposes, as well as filter-based (Information Gain, Gain Ratio) and spiral-based methods for feature selection; and with using these, the classification process were repeated. The imbalance in the data set were eliminated by using SMOTE (Synthetic Minority Over-sampling Technique) random over-sampling and under-sampling algorithms and the same classification processes were re-applied. The performance measures that are used in the study are the precision, recall, F-measure and receiver operating characteristic.

Key Words: Electronic Tongue, Support Vector Machines, k-Nearest Neighbours, Random Forest, Classification, Feature Selection

ŞEKİLLER DİZİNİ

Sayfa No

Şekil 1.1.	“Jawda” pastörize sütünün 5 saklama günü için hibrit e-burun ve voltametrik e-dil ile ölçümlerinin TBA skor grafiği [35].	8
Şekil 1.2.	a) Kalın film teknolojisinde 16 elektrottan oluşan e-dilin tasarımı b) Metal elektrot tepkisinin TBA sonuçları [40].	9
Şekil 1.3.	Bal ayrımı için çoklu sensör sistemi [47].	12
Şekil 1.4.	Farklı üzüm çeşitlerinin TBA skoru grafiği [9].	13
Şekil 1.5.	E-dilin TBA ile farklı tipteki içecekleri ayırma yetisi [51].	15
Şekil 1.6.	Çoklu-sensör dizisi ölçüm kurulumunun şematik diyagramı [58].	17
Şekil 1.7.	E-dilin 3 ve 6 aydır yıllandırılan şarapla ilgili verilerinin TBA grafiği [60].	18
Şekil 1.8.	Optimum hiper düzlem ve destek vektörleri [62].	20
Şekil 1.9.	Lineer olarak birbirinden ayrılabilen veriler arasındaki muhtemel en büyük boşluk [63].	20
Şekil 1.10.	Kernel fonksiyonu ile verinin daha yüksek bir boyuta dönüştürülmesi [62].	24
Şekil 1.11.	k-EYK yöntemi için basit bir örnek [64].	26
Şekil 1.12.	RO algoritmasının genel mimarisi [66].	28
Şekil 1.13.	SMOTE algoritma adımları	36
Şekil 2.1.	Kırmızı şarap örneklerinin duysal değerlendirmelerinin histogram grafiği	43
Şekil 2.2.	Beyaz şarap örneklerinin duysal değerlendirmelerinin histogram grafiği	43
Şekil 2.3.	3. kalite kırmızı şarap için a) normal değerler b) ters sırada değerler ile özniteliklerin radar çizimi	44
Şekil 2.4.	4. kalite kırmızı şarap için a) normal değerler b) ters sırada değerler ile özniteliklerin radar çizimi	44
Şekil 2.5.	5. kalite kırmızı şarap için a) normal değerler b) ters sırada değerler ile özniteliklerin radar çizimi	44
Şekil 2.6.	6. kalite kırmızı şarap için a) normal değerler b) ters sırada değerler ile özniteliklerin radar çizimi	45
Şekil 2.7.	7. kalite kırmızı şarap için a) normal değerler b) ters sırada değerler ile özniteliklerin radar çizimi	45
Şekil 2.8.	8. kalite kırmızı şarap için a) normal değerler b) ters sırada değerler ile özniteliklerin radar çizimi	45

Şekil 2.9.	3. kalite beyaz şarap için a) normal değerler b) ters sırada değerler ile öz niteliklerin radar çizimi	46
Şekil 2.10.	4. kalite beyaz şarap için a) normal değerler b) ters sırada değerler ile öz niteliklerin radar çizimi	46
Şekil 2.11.	5. kalite beyaz şarap için a) normal değerler b) ters sırada değerler ile öz niteliklerin radar çizimi	46
Şekil 2.12.	6. kalite beyaz şarap için a) normal değerler b) ters sırada değerler ile öz niteliklerin radar çizimi	47
Şekil 2.13.	7. kalite beyaz şarap için a) normal değerler b) ters sırada değerler ile öz niteliklerin radar çizimi	47
Şekil 2.14.	8. kalite beyaz şarap için a) normal değerler b) ters sırada değerler ile öz niteliklerin radar çizimi	47
Şekil 2.15.	9. kalite beyaz şarap için a) normal değerler b) ters sırada değerler ile öz niteliklerin radar çizimi	48
Şekil 2.16.	3. kalite kırmızı şarap örneklerinin a) 4 öz niteliğinin b) 7 öz niteliğinin tüm veriler için dağılımı.....	49
Şekil 2.17.	4. kalite kırmızı şarap örneklerinin a) 4 öz niteliğinin b) 7 öz niteliğinin tüm veriler için dağılımı.....	49
Şekil 2.18.	5. kalite kırmızı şarap örneklerinin a) 4 öz niteliğinin b) 7 öz niteliğinin tüm veriler için dağılımı.....	49
Şekil 2.19.	6. kalite kırmızı şarap örneklerinin a) 4 öz niteliğinin b) 7 öz niteliğinin tüm veriler için dağılımı.....	50
Şekil 2.20.	7. kalite kırmızı şarap örneklerinin a) 4 öz niteliğinin b) 7 öz niteliğinin tüm veriler için dağılımı.....	50
Şekil 2.21.	8. kalite kırmızı şarap örneklerinin a) 4 öz niteliğinin b) 7 öz niteliğinin tüm veriler için dağılımı.....	50
Şekil 2.22.	3. kalite beyaz şarap örneklerinin a) 4 öz niteliğinin b) 7 öz niteliğinin tüm veriler için dağılımı.....	51
Şekil 2.23.	4. kalite beyaz şarap örneklerinin a) 4 öz niteliğinin b) 7 öz niteliğinin tüm veriler için dağılımı.....	51
Şekil 2.24.	5. kalite beyaz şarap örneklerinin a) 4 öz niteliğinin b) 7 öz niteliğinin tüm veriler için dağılımı.....	51
Şekil 2.25.	6. kalite beyaz şarap örneklerinin a) 4 öz niteliğinin b) 7 öz niteliğinin tüm veriler için dağılımı.....	52
Şekil 2.26.	7. kalite beyaz şarap örneklerinin a) 4 öz niteliğinin b) 7 öz niteliğinin tüm veriler için dağılımı.....	52
Şekil 2.27.	8. kalite beyaz şarap örneklerinin a) 4 öz niteliğinin b) 7 öz niteliğinin tüm veriler için dağılımı.....	52

Şekil 2.28.	9. kalite beyaz şarap örneklerinin a) 4 özniteliğinin b) 7 özniteliğinin tüm veriler için dağılımı.....	53
Şekil 2.29.	Kırmızı ve beyaz şarap örnekleri için saçılım grafiği matrisi.....	54
Şekil 2.30.	Farklı kalitelardaki kırmızı şarap örnekleri için saçılım grafiği matrisi	55
Şekil 2.31.	Farklı kalitelardaki beyaz şarap örnekleri için saçılım grafiği matrisi	56
Şekil 2.32.	Kırmızı şarap örneklerinin .arff formatlı dosyasının WordPad’de görünümü	58
Şekil 2.33.	Bilgi kazancı algoritmasıyla seçilen özniteliklerin sayısına göre a) 10 kat çapraz doğrulama b) %66 eğitim c) %80 eğitim verileri için kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının anma değerine göre karşılaştırılması.....	69
Şekil 2.34.	Kazanım oranı algoritmasıyla seçilen özniteliklerin sayısına göre a) 10 kat çapraz doğrulama b) %66 eğitim c) %80 eğitim verileri için kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının anma değerine göre karşılaştırılması.....	70
Şekil 2.35.	Bilgi kazancı algoritmasıyla seçilen özniteliklerin sayısına göre a) 10 kat çapraz doğrulama b) %66 eğitim c) %80 eğitim verileri için beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının anma değerine göre karşılaştırılması.....	75
Şekil 2.36.	Kazanım oranı algoritmasıyla seçilen özniteliklerin sayısına göre a) 10 kat çapraz doğrulama b) %66 eğitim c) %80 eğitim verileri için beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının anma değerine göre karşılaştırılması.....	76

TABLolar DİZİNİ

	<u>Sayfa No</u>
Tablo 1.1. Karışıklık matrisi	38
Tablo 2.1. Kırmızı şarap örneklerinin fiziko-kimyasal veri istatistikleri.....	41
Tablo 2.2. Beyaz şarap örneklerinin fiziko-kimyasal veri istatistikleri	41
Tablo 2.3. Karışık veri setinin kırmızı ve beyaz şaraplar olarak sınıflandırılmasının başarı sonuçları	59
Tablo 2.4. Kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları.....	60
Tablo 2.5. Beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları.....	60
Tablo 2.6. Kırmızı ve beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları	61
Tablo 2.7. Kırmızı şarap örneklerinin TBA uygulandıktan sonra kalitelerine göre sınıflandırılmasının başarı sonuçları	62
Tablo 2.8. Beyaz şarap örneklerinin TBA uygulandıktan sonra kalitelerine göre sınıflandırılmasının başarı sonuçları	62
Tablo 2.9. Kırmızı ve beyaz şarap örneklerinin TBA uygulandıktan sonra kalitelerine göre sınıflandırılmasının başarı sonuçları	63
Tablo 2.10. Kırmızı şarap örneklerine ait özniteliklerin bilgi kazancı ağırlıkları.....	64
Tablo 2.11. Kırmızı şarap örneklerine ait özniteliklerin kazanım oranı ağırlıkları.....	64
Tablo 2.12. Beyaz şarap örneklerine ait özniteliklerin bilgi kazancı ağırlıkları	65
Tablo 2.13. Beyaz şarap örneklerine ait özniteliklerin kazanım oranı ağırlıkları	65
Tablo 2.14. Bilgi kazancı/kazanım oranı algoritmalarıyla seçilen en iyi 3 öznitelikle kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları.....	66
Tablo 2.15. Bilgi kazancı algoritmasıyla seçilen en iyi 4 öznitelikle kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları	66
Tablo 2.16. Kazanım oranı algoritmasıyla seçilen en iyi 4 öznitelikle kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları	66
Tablo 2.17. Bilgi kazancı algoritmasıyla seçilen en iyi 5 öznitelikle kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları	67
Tablo 2.18. Kazanım oranı algoritmasıyla seçilen en iyi 5 öznitelikle kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları	67

Tablo 2.19.	Bilgi kazancı/kazanım oranı algoritmalarıyla seçilen en iyi 6 öznitelikle kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları.....	67
Tablo 2.20.	Bilgi kazancı/kazanım oranı algoritmalarıyla seçilen en iyi 7 öznitelikle kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları.....	68
Tablo 2.21.	Bilgi kazancı/kazanım oranı algoritmalarıyla seçilen en iyi 8 öznitelikle kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları.....	68
Tablo 2.22.	Bilgi kazancı/kazanım oranı algoritmalarıyla seçilen en iyi 3 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları.....	71
Tablo 2.23.	Bilgi kazancı algoritmasıyla seçilen en iyi 4 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları	71
Tablo 2.24.	Kazanım oranı algoritmasıyla seçilen en iyi 4 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları	72
Tablo 2.25.	Bilgi kazancı algoritmasıyla seçilen en iyi 5 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları	72
Tablo 2.26.	Kazanım oranı algoritmasıyla seçilen en iyi 5 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları	72
Tablo 2.27.	Bilgi kazancı algoritmasıyla seçilen en iyi 6 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları	73
Tablo 2.28.	Kazanım oranı algoritmasıyla seçilen en iyi 6 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları	73
Tablo 2.29.	Bilgi kazancı algoritmasıyla seçilen en iyi 7 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları	73
Tablo 2.30.	Kazanım oranı algoritmasıyla seçilen en iyi 7 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları	74
Tablo 2.31.	Bilgi kazancı algoritmasıyla seçilen en iyi 8 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları	74
Tablo 2.32.	Kazanım oranı algoritmasıyla seçilen en iyi 8 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları	74
Tablo 2.33.	Sarmalama yöntemiyle seçilen özniteliklerle kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları.....	77
Tablo 2.34.	Sarmalama yöntemiyle seçilen özniteliklerle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları.....	78
Tablo 2.35.	Kırmızı şarap örneklerinden elde edilen yeni eğitim kümelerinin örnek sayıları.....	78
Tablo 2.36.	Beyaz şarap örneklerinden elde edilen yeni eğitim kümelerinin örnek sayıları.....	79

Tablo 2.37.	SMOTE algoritmasıyla yeni eğitim kümeleri oluşturulan kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları	79
Tablo 2.38.	SMOTE ve rastgele alt örnekleme algoritmalarıyla yeni eğitim kümeleri oluşturulan beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları	80
Tablo 2.39.	Kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının tüm yöntemlerle anma değerlerinin karşılaştırılması	81
Tablo 2.40.	Beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının tüm yöntemlerle anma değerlerinin karşılaştırılması	81



SEMBOLLER DİZİNİ

Ag	Gümüş
AgCl	Gümüş Klorür
Ag ₂ CO ₃	Gümüş Karbonat
Ag ₂ O	Gümüş Oksit
Ag ₂ S	Gümüş Sülfür
Al	Alüminyum
Au	Altın
ATP	Adenozin Trifosfat
Ca	Kalsiyum
Cd	Kadmiyum
Cl	Klor
Co	Kobalt
CO ₂	Karbondioksit
Cr	Krom
Cu	Bakır
Cu ₂ O	Bakır I Oksit
CuS	Bakır II Sülfür
Fe	Demir
H	Hidrojen
HCl	Hidroklorik Asit
H ₂ O ₂	Hidrojen Peroksit
H ₂ SO ₄	Sülfürik Asit
Ir	İridyum
ITO	İndiyum Kalay Oksit
KBr	Potasyum Bromür
KCl	Potasyum Klorür
K ₂ SO ₄	Potasyum Sülfat
Mg	Magnezyum
Mn	Mangan
Na	Sodyum

NaCl	Sodyum Klorür
NaClO ₂	Sodyum Klorid
Ni	Nikel
Pb	Kurşun
Pb(NO ₃) ₂	Kurşun Nitrat
Pd	Paladyum
Pt	Platin
PVC	Polivinil Klorür
Rh	Rodyum
SO ₂	Kükürt Dioksit
SO ₄	Sülfat
Zn	Çinko
TVB-N	Total Volatile Basic Nitrogen (Toplam Uçucu Bazik Nitrojen)
HACCP	Hazard Analysis and Critical Control Points (Tehlike Analizleri ve Kritik Kontrol Noktaları)
IUPAC	International Union of Pure and Applied Chemistry (Uluslararası Temel ve Uygulamalı Kimya Birliği)
DP	Doğru Pozitif
DN	Doğru Negatif
YP	Yanlış Pozitif
YN	Yanlış Negatif
ÇLR	Çoklu Lineer Regresyon
ÇRA	Çoklu Regresyon Analizi
DFA	Diskriminant Fonksiyon Analizi
DVM	Destek Vektör Makinesi
GYSA	Geri Yayılımlı Sinir Ağı
HKA	Hiyerarşik Küme Analizi
KA	Küme Analizi
KEKK	Kısmi En Küçük Kareler
k-EYK	k-En Yakın Komşuluk
LOEKK	Lineer Olmayan En Küçük Kareler
RO	Rastgele Orman

SDA	Standart Diskriminant Analizi
SİA	Standart İlişki Analizi
TBA	Temel Bileşenler Analizi
TBR	Temel Bileşenler Regresyonu
YSA	Yapay Sinir Ağları



1. GENEL BİLGİLER

1.1. Giriş

Gıda dağıtımı ve güvenliği modern toplumda oldukça gerçek sorunlardır. Gıda sektöründeki güvenlik, gıdanın erişilebilirliği, eşit dağılımı, kalitesi ve güvenliğinin sağlanması için gereklidir. Güvenlik ise ürünün kimyasal bileşimi, fiziksel özellikleri, mikrobiyal ve toksik kontaminasyonu ve saklama koşullarına bağlıdır, bunlar ISO 22000 standartları ve Tehlike Analizleri ve Kritik Kontrol Noktaları (HACCP-Hazard Analysis and Critical Control Points) tarafından düzenlenir [1].

Bugünlerde tüketiciler yeme düzenleri açısından gittikçe daha seçici ve talepkar olmaya başlamışlardır. Gıdanın tadı ve kalitesi tüketicinin kendi tercihlerini ifade etmesi için başlıca araçlardır. Tüketicinin gıda kalitesi ve güvenliği konularındaki ihtiyaçlarının artışı gıda sertifikasyonunda yeni tekniklerin gelişmesine yol açmıştır. Fakat bu tekniklerin çoğu zaman alan tekniklerdir ve gelişmiş aparatlara ve kalifiye personele ihtiyaç duymaktadırlar [2]. Bu sınırlamalardan dolayı alternatif analitik metotlar sağlanmalıdır. Gıda değerlendirmesinde alternatif yaklaşımlara parlak bir örnek ise elektronik dilin (e-dilin) çoklu-sensör sisteminin insan tat alma sistemini taklit etmesidir. E-dil gıdaların kimyasal bileşimini belirleyebilir ve farklı gıda tatlarını tanıyabilir. Ayrıca, gıda ürünlerinin yapay duyuşal değerlendirmesi kolaylıkla insan algısıyla ilişkilendirilebilmektedir. Modern e-dil hem laboratuvar ortamında hem de endüstriyel gıda üretim sürecinde gıda kalitesinin hızlı ve yıkıcı olmayan bir değerlendirmesini yapmaya izin verir. E-dilin gıda maddesi analizi uygulamaları ile ilgili birkaç detaylı inceleme önceden yayınlanmıştır [3,4]. AlphaMOS'un (Fransa) α -Astree e-dili, Anritsu Corp'un (Japonya) elektronik tat algılama sistemi Insent[®], McScience'in (Güney Kore) çoklu dizi kimyasal sensörü ve Sensor Systems'in (St. Petersburg, Rusya) e-dili gibi birçok sistem piyasada hazır bulunmaktadır. Bu yapay tat sistemlerinin yanında farklı araştırma grupları tarafından geliştirilmiş ev yapımı aletler, çeşitli gıda değerlendirme görevleri için artmakta olan ihtiyacı karşılamak amacıyla geniş çaplı olarak kullanılmaktadır.

1.2. Elektronik Dilin Tanımı ve Temelleri

E-dil, ilk olarak 1990'lı yıllarda Toko ve arkadaşları [5-7] tarafından insan tat alma reseptörlerini taklit etme amacıyla ileri sürülmüştür. E-dilin amacı beş temel tat olan tuzlu (sodyum klorür (NaCl), potasyum klorür (KCl) ve potasyum bromür (KBr)), ekşi (hidroklorik asit (HCl), sitrik asit ve asetik asit), acı (kinin), tatlı (sakkaroz) ve umami (monosodyum glutamat) tatlarını incelemektir. Daha sonra, astrenjan ve keskin tatlar gibi diğer tat çeşitleri de incelenmiş ve e-dilin yiyecek değerlendirme, içecek ayırımı ve hatta zararlı madde ayırt etme gibi pek çok farklı uygulamaları araştırılmıştır.

Uluslararası Temel ve Uygulamalı Kimya Birliği (IUPAC-International Union of Pure and Applied Chemistry) tanımına göre, e-dil “Düşük seçiciliğe sahip pek çok sensör içeren ve sinyal işleme için Örüntü Tanıma veya Çok Değişkenli Veri Analizi'ne dayanan gelişmiş matematiksel prosedürler kullanan bir çoklu sensör sistemidir.” [8]. Modern e-dil sistemlerinde kullanılan sensör tipleri arasında voltametrik [9-11], potansiyometrik [12-14], impedimetrik veya kapasitif sensörler [15,16], optik sensörler [12], enzim bazlı biyosensörler [9] ve hibrit veya bölümlendirilmiş cihazlar [12,13] bulunmaktadır.

E-dil ile belirli olmayan kimyasal sensörlerin belirli iyon seçici elektrotların dizileri olarak uygulanması [17,18] ve yeni özüt alma yöntemleri [11,16] sunulmuştur. E-dilden elde edilen global bilgiler, ürünlerin tat karakteristiklerinin dijital parmak izlerini elde etmek için hala kullanılmaktadır. Fakat aynı zamanda özel bileşenlerin somut parametreleri ve içerikleri sensör ve kemometrik tekniklerin uygun olarak seçilmesiyle hesaplanabilir. Yapay algılama sistemleri tarafından yapılan bir ölçüm oldukça fazla miktarda bilgi üretir, bu bilgiler kemometrik metotlar için kullanılır. En sık kullanılan yöntemler Yapay Sinir Ağları (YSA), Temel Bileşenler Analizi (TBA), Küme ve Hiyerarşik Küme Analizleri (KA, HKA), Destek Vektör Makinesi (DVM), çeşitli regresyon metotları: Kısmi En Küçük Kareler (KEKK), Çoklu Lineer Regresyon (ÇLR) ve Temel Bileşen Regresyonudur (TBR).

E-dil teknolojisinde kullanılan çeşitli algılama materyalleri arasında metalik sensörler [11], karbon pasta [19,20], kalkojenit cam elektrotlar [13], polimerik filmler [12,17], moleküler baskılanmış polimerler [21] ve çoklu-transdüksiyonlu kaplamalar [12] vardır. İlerleyen kısımlarda, gıda endüstrisindeki bu materyallere bağlı bazı seçilmiş e-dil uygulamaları anlatılmıştır.

1.3. Elektronik Dil ile Gıda Analizi İçin Örnek Ön İşlem Şartları

Önceden e-dil cihazları çoğunlukla sıvıları analiz etmede kullanılırken, son zamanlarda katı, yağlı, lifli veya sıvı olmayan gıda maddelerinin değerlendirilmesiyle ilgili pek çok araştırma yapılmıştır. Bu bağlamda, e-dil analizinde örnek faz hali (sıvı vs.) ve belirli ön işlem prosedür şartları konusunda çeşitli uygulama sorunlarıyla karşılaşmaktadır. Bu yüzden katı gıda örnekleri doğru ölçüm ve hassas materyallerle fiziksel olarak yakın temasa geçebilmesi için uygun faz haline dönüştürülmelidir.

Örneğin, katı gıdalar ezilmeli veya kıyılmalı [18,22], soğuk örnekler sensör operasyon sıcaklığına kadar ısıtılmalı, sıcak olanlar soğutulmalı, homojen olmayanlar sonikasyonla homojenize edilmelidir [16]. Ayrıca modern e-dil sistemlerinde uygulanan sensörlerin çoğunun genelde sıvı fazında (ve hatta daha da sıklıkla sıvı ortamlarda) operasyon gösterdiği göz önüne alınarak, örnek ıslatma, seyreltme veya sensör dostu çözücülerle özüt çıkarma gerekli olabilir [11,16]. Bu süreçler yemek yerken insan ağzının gıdaya uyguladığı ön işlemlere benzerdir: Gıda parçalarının boyutlarını küçültmek için çiğneme, sindirim yolundaki geçişe hazırlamak amacıyla tükürük enzimleriyle öncül gıda dönüşümünü başlatmak için tükürükleme ve dilin beş temel tat için hassas olan belirli bölgelerinde gerçekleşen tat alma.

1.4. Elektronik Dil ile Tat ve Tat Bileşenlerinin Ayrımı

Tat değerlendirmesi ve tat belirleyen bileşen tespiti, 1990'ların başında ilk raporlandırıldığından beri e-dil sistemlerinin en yaygın uygulamalarından biridir [5]. Gıdanın genel tadını etkileyen beş temel tat tatlı, tuzlu, acı, ekşi ve umamidir. Yakın zamanda “kokumi” adında “karmaşıklık”, “damakta kalıcılık” içeren bir tadı anlatmak için yeni bir terim ortaya atılmıştır. Temel tatlardan sorumlu olan moleküller: tatlı için karbonil grupları içeren şekerler, tuzlu için genelde sodyum (Na) iyonları, ekşi için hidrojen (H) iyonları ve genelde gıda maddesine “lezzetli” umami tadını katmak için eklenen monosodyum glutamattır. Pek çok bileşik ve bunların arasında pek çok doğal toksin acı tada sahiptir; bu yüzden doğal koruma tepkisi olarak insan dili acılığa karşı hassastır. Referans alınan acı tattaki bileşiklerden biri kinindir. Bugüne kadar bilinen en güçlü kokumi madde glutamin, valin ve glisin amino asitlerinden oluşan bir peptit molekülüdür.

Geçmişte, temel tat bileşenleri içeriklerini gıda tadına bağlama girişimleri, tat hissi gıdanın kimyasal, fizikokimyasal ve biyolojik parametreleri tarafından belirlenen karmaşık bir süreç olduğundan genelde başarılı değildi [23]. Yine de temel tat bileşenleri sıklıkla yapay algı sistemlerinin nitel tek-tatlı karakterizasyon görevlerinde kalibrasyon için kullanılır.

Toyota (2011) ve arkadaşları, gıdada tatlılıktan sorumlu olan bileşenleri tespit etmek konusunda detaylı çalışmalar yapmışlar ve elektrolit olmayan ve aspartan gibi pozitif yüklü yapay tatlandırıcı bileşenler için tat alma sistemine uygulanmak üzere tatlılık sensörü geliştirmişlerdir [24]. İnterdijital kapasitör e-dilin şeker çözeltilerindeki şeker içeriğini belirleme uygulaması Angkawisittpan ve Manasri (2012) tarafından raporlanmıştır [15]. Diğer uygulamalar ise ilaçların acılık değeri değerlendirmeleri [25], süt ürünleri protein hidrolizatları [26], farklı tat artırıcılarından gelen umami tadının değerlendirilmesi ve karşılaştırılması [27], farklı ilaç formülasyonlarının ayrımı, tatlandırıcı ve farklı tatların eklenmesi ile ilaçlardaki aktif içeriğin (normalde acı) maskelenmesinin değerlendirilmesi ile ilgilidir [28].

1.5. Elektronik Dilin Gıdadaki Uygulama Alanları

Konvansiyonel yiyecek analizinde kimyasal, biyokimyasal, fiziko-kimyasal ve mikrobiyal prensiplere dayalı pek çok yöntem kullanılarak yemeğin karakteristiğine doğrudan etkiyen değişik bileşiklerin varlığı veya konsantrasyonunun belirlenmesi amaçlanmıştır. Bu konvansiyonel yaklaşımlar zaman alıcı olmuş ve bunları gerçekleştirmek için karmaşık ve pahalı donanımların olduğu laboratuvarlara ihtiyaç duyulmuştur. Ayrıca, analizler sahada gözlem yapmaya uygun olmayan uzman personel tarafından yapılmaktadır. Karşılaştırma yapmak gerekirse, e-dil, nitel ve nicel analizleri daha hızlı ve kolay prosedürler izleyerek yapılabilmesi sayesinde yiyecek analizinde umut vadetmektedir.

1.5.1. Zeytinyağı ve Bitkisel Yağlar

En çok tüketilen gıda maddelerinden biri olan bitkisel yağ ve özellikle zeytinyağı, son yirmi yılda çoklu-sensör sistemleri tarafından aktif olarak değerlendirilmiştir ve yapay

algılama sistemleri ile bitkisel yağ analizine olan ilgi hala oldukça fazladır. Geleneksel olarak yapay koku alma sistemleri, yani e-burunlar çoğunlukla bitkisel yağların aroma değerlendirilmesine uygulanmıştır [2]. Gerekli örnek ön işlemlerine bağlı deneysel sorunlar dolayısıyla e-dilin kullanımı çok daha azdı. E-dilde kullanılan elektrokimyasal sensörlerin doğru işlemlerini sağlamak için yağların özleri çıkarılmalı veya hidrofolik çözücülerle seyreltilmelidirler. Pazarda bulunan bitkisel yağların sınıflarının ve markalarının çok çeşitli olmasından dolayı e-dil sistemlerinin yağ analizindeki ana uygulamaları; yağın bitkisel kökenini tespit etmek [12], sınıfını tespit etmek veya coğrafik kökeninin istikrarını kontrol etmek [14], ürünün seyreltilmemiş olmasını ve kalitesinin temini [10,12], farklı yağ bileşenlerinin ve onların içerik niceliklerinin tespiti [12] ve tat değerlendirmesidir [14,20].

Men (2013) ve arkadaşları, sırasıyla mısır, ayçiçeği, soya fasulyesi, susam ve yer fıstığı yağlarını içeren farklı yenebilir yağların sınıflandırılması için yapılmış tek bir altın (Au) elektrota dayanan voltametrik bir e-dil sunmuşlardır. Yağ örneklerinin 30°C'de 5 dakika boyunca iyonsuzlaştırılmış su ile özütü çıkartılmış ve sonra sıvı fazda e-dil ile analiz edilmiştir. Sistem tüm yağları ayırt edebilmiştir, en iyi tespitler TBA ile elde edilmiştir. Son zamanlarda aynı e-dil cihazı ve TGS Figaro gaz sensörlerinden oluşan e-burun kızartılmış gıda üretiminde eski kızartma yağı ve yeni yağın karışım oranını tespit etmek için uygulanmıştır [10].

Tek zeytin çeşidinden elde edilen Portekiz ve İspanyol yağları, 40 adet ekran baskılı ve çapraz hassaslığa sahip dört lipidik katkı maddesinin önceden belirlenmiş farklı kütle oranlarında birleşimleriyle doplanmış polivinil klorür (PVC) zarlar ile modifiye edilmiş potansiyometrik sensörlerden oluşan potansiyometrik bir e-dil ile analiz edilmiştir [14]. Her yağdaki polar bileşenler hidroetanolik bir solüsyon kullanılarak ayrılmıştır. Yağlar zeytin çeşitlerine göre doğru sınıflandırılmıştır ve e-dilin hassaslığının %97'den fazla olduğu gözlenmiştir. Ayrıca, sistemin zeytinyağındaki polar bileşenleri ayırt etme yeteneği dolaylı olarak acılık, burukluk veya keskinlik gibi duyuşal özellikleri değerlendirme ihtimalini ortaya çıkarmıştır.

Kare dalga voltametrisine dayanan ve modifiye edilmiş karbon pasta sensörlerden oluşan bir e-dil Apetrei ve Apetrei (2014) tarafından saf zeytinyağının farklı oranlarda ayçiçek yağı, soya fasulyesi yağı ve mısır yağı ile karıştırılmasını tespit etmek için uygulanmıştır [19]. Karıştırılan yağların konsantrasyon seviyesi %5 ile %10 arasındayken saflığı bozulmuş yağların doğru sınıflandırılması ve polifenolün içeriğinin

değerlendirilmesi ihtimali gösterilmiştir. Daha önceden aynı yazarlar saf zeytinyağının fenolik içeriğinin polipirel ile modifiye edilmiş ekran baskılı elektrotlar kullanılarak tespit edimesini sunmuşlardır [29].

Aynı anda e-burun, e-dil ve e-göz kullanan bölümlendirilmiş yapay algılama sistemlerinin farklı derecelerde acılıklara sahip zeytinyağlarının karakterizasyonunda [20] ve Fas sızma zeytinyağlarının gelişmiş sınıflandırılmasında [30] uygulanması kısa süre önce sunulmuştur. Apetrei (2010) ve arkadaşları, metal oksit gaz sensörlerine dayalı e-burun, zeytinyağı için modifiye edilmiş karbon pasta elektrotlu voltametrik e-dilin ve 380 nm'den 780 nm'ye kadar değişen seri ışık yayan diyotlar kullanılarak kaydedilen geçirgenlik spektrumuna bağlı e-gözün kombinasyonunun; 25 saf zeytinyağının farklı zeytinler için ayrılmasında kullanılmasını sunmuşlardır. Tahmini kare kök ortalama hata karesi 0.099'dan azdır. Elektronik panel verisi ile yağ acılığı endeksi (uzmanlar tarafından puanlanmış) ve polifenolik içerik (kromatografik metotlarla ölçülen) arasındaki ilişki KEKK metodu ile belirlenmiştir ve hem kalibrasyonda hem de doğrulamada 0.9'dan büyük ilinti katsayıları bulunmuştur. 5 farklı Fas sızma zeytinyağını coğrafik kökenlerine göre ayırmak için, piyasadaki 5 adet metal oksit gaz sensöründen yapılan e-burun ve platinyum (Pt), Au, camısı karbon ve indiyum kalay oksit (ITO) elektrot bazlı voltametrik e-dilden gelen verileri belirlemek için düşük seviye soyut veri kaynaştırması yapılmıştır [30]. Elektrokimyasal ölçümler için yağlar yardımcı elektrolit olarak tetrabutül-amonyum tetrafloroborat içeren diklorometanda çözölmüş ve 30°C'de su banyosunda analiz edilmiştir.

1.5.2. Süt ve Süt Ürünleri

Süt ürünleri endüstrisi pek çok süt, yoğurt ve diğer fermente içecekler, yumuşak ve eski peynirler ve ekşi kremalardan oluşan geniş bir ürün yelpazesi sunmaktadır. Süt ürünleri endüstrisinde pek çok fermentasyon yöntemi, mikroorganizmalar ve gıda katkıları ve ayrıca farklı memeli hayvanlardan gelen çiğ süt kullanılmaktadır. Ürün homojenliğini koruma, tazelik garantisi ve saflığı koruma gerekliliği süt ürünlerinde e-dil uygulamalarının sayısının gittikçe artmasına neden olmaktadır [31,32].

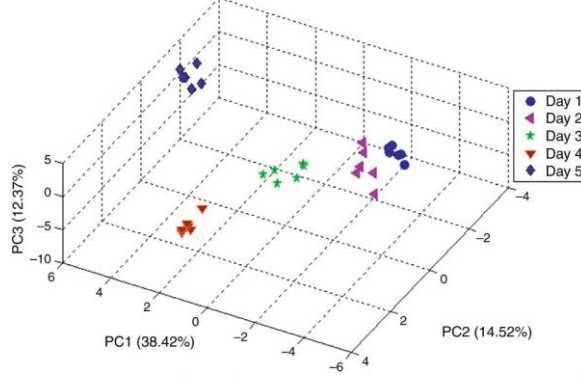
Wei ve Wang (2011) sığır sütündeki 6 farklı antibiyotik katkısının tespit edilmesi için 5 metalik elektrottan oluşan voltametrik e-dil geliştirmişlerdir [11]. Büyümeyi sağlamak, terapi veya propilaktik sebeplerle düzenli olarak ilaç verilen hayvanlarda

antibiyotikler süte geçebilir. Çiğ sütte antibiyotiklerin tespiti önemlidir, çünkü süütün fermentasyonu olumsuz yönde etkilenebilir veya tüketicilerde alerjiler ortaya çıkabilir. Süt örneklerinin içine antibiyotik örnekleri karıştırılmıştır (1 mg antibiyotik 100 µl dimetil sülfoksit karıştırılmış, sonra süt ile 100 ml'lik çözeltilere seyreltilmiştir.). Değişik antibiyotiklerle katılanmış sığır süütünün Diskriminant Fonksiyon Analizi (DFA) yöntemiyle açıkça ayrılmasının olası olduğu gösterilmiştir. Son zamanlarda aynı araştırma grubu dört metalik elektrotta dayanan (Au, gümüş (Ag), Pt, paladyum (Pd)) voltametrik bir e-dilin 26 pastörize sütün açıldıktan sonraki 72 saat içinde 7 farklı zamanda kalite ve saklama süresi gözleminde kullanıldığını [31] ve yoğurtların çeşitliliğini ve fiziksel özelliklerini değerlendirmede kullanıldığını [32] rapor etmişlerdir.

Değişik tatlara sahip farklı markaların probiyotik sütlerinin izlenmesi ve sınıflandırılması için Hruskar (2010) ve arkadaşları tarafından iki yöntemin karşılaştırılması, sensör analizi ve kimyasal alan etkili transistörlere dayalı α -Astree potansiyometrik e-dili kullanılarak yapılmıştır [33]. Eğitimli panelistler tarafından değerlendirilmiş görünüm, kıvam, renk, koku ve tatlar ile e-dilin, iki farklı sıcaklıkta (4°C ve 25°C) 20 günlük bir periyotta saklanmış sütlere tepkisi YSA ve KEKK metotlarıyla ilişkilendirilmiştir. E-dilin probiyotikle mayalanmış sütü tadına göre sınıflara ayırması, saklama boyunca oluşan bozunumu izlemesi ve duyuşal karakteristikleri tahmin edip bunları süütün kalitesiyle bağdaştırmasının olası olduğu gösterilmiştir. α -Astree e-dilin bir başka uygulaması ise 40 farklı probiyotik süt örneğinde aynı anda etanol, asetaldehit, diasetil, laktik asit, asetik asit ve sitrik asit içeriklerini belirlemektir [34]. E-dilden elde edilen veri, uygun enzimlerle yapılan enzimatik metotla belirlenen süt mikro bileşenleriyle YSA kullanılarak ilişkilendirilmiştir.

4 metal elektrottan (Pt, Au, Ag ve camısı karbon) oluşan voltametrik e-dil ve hibrit bir e-burun (ev yapımı ve ticari metal oksit gaz sensörlerinden, sıcaklık probundan ve nem sensöründen oluşan) Bougrini (2014) ve arkadaşları tarafından farklı pastörize süt markalarını ayırmak ve saklama sürelerini tanımak için kullanılmıştır [35]. Ölçümler yapılmadan önce süt örneklerinin ortam sıcaklığına ulaşmasına izin verilmiş ve süt örnekleri e-dil ve e-burun tarafından başka bir ön işlem yapılmadan analiz edilmiştir. Aynı değerlendirildiğinde, e-dil verisi ilk saklama gününde markaların açıkça ayrılmasına izin vermiş ve bu toplam varyansın %80.8'ini oluşturmuştur. E-dil ve e-burun verisinin birleşimi (orta seviyeli soyutlama füzyonu uygulanmıştır) tüm süt saklama günlerinin Şekil 1.1'de gösterilen TBA sınıflandırmasını ortaya çıkarmıştır. Birleştirilmiş veriye

uygulanan DVM, pastörize sütün saklama günlerinin tam olarak tespit edilmesini sağlamıştır.



Şekil 1.1. “Jawda” pastörize sütünün 5 saklama günü için hibrit e-burun ve voltametrik e-dil ile ölçümlerinin TBA skor grafiği [35].

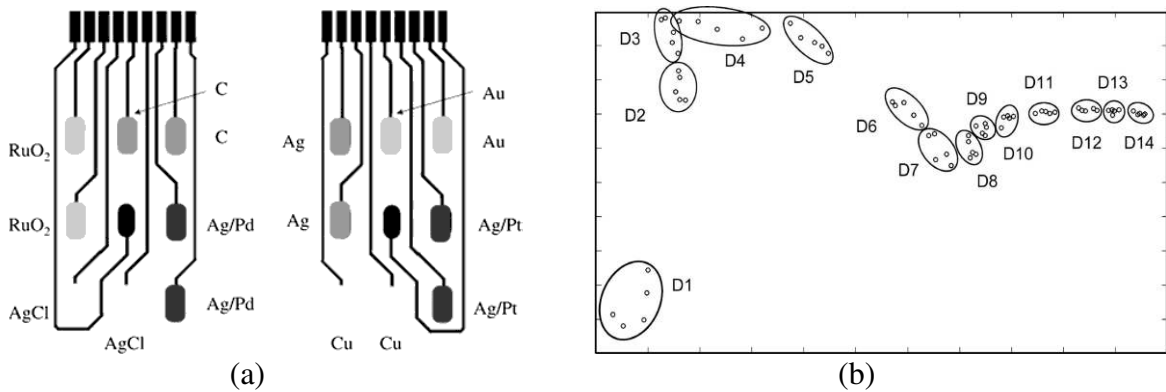
1.5.3. Balık ve Et

Et ve deniz ürünleri gibi besleyici ve proteince zengin gıda maddelerinin doğru kalite ve güvenlik kontrolü, bu ürünlerin dengeli beslenmede çok önemli olmasından dolayı çok ilgi görmektedir. Et ve balık pazarı endüstrisinde tazelik en önemli faktördür ve raf ömrü ile bozulma süresinin izlenmesi için çoklu-sensör analizinin etkili bir araç olduğu gözlenmiştir. E-dil sisteminin, balık tazeliği kontrolü [36], kalite [37] ve tat değerlendirmesi [38] uygulamalarında kullanımı birkaç çalışmada sunulmuştur.

Kaneki (2004) ve arkadaşları, domuz etinin tazeliğini değerlendirmek için potansiyometrik katı hal elektrotları kullanan bir e-dil öne sürmüşlerdir [39]. Putresin ve dimetil sülfür gibi etin çürümesinin ilk basamağında oluşan organik bileşikler tespit edebilmek için Pt, bakır sülfür (CuS) ve gümüş sülfür (Ag₂S) katı hal elektrotlarını seçmişlerdir. Domuz etinin nitel değerlendirmesi için her elektrottaki elektriksel potansiyelden elde edilen veriler üzerine TBA uygulanmıştır ve domuz etinin tazelik derecesinin nitel olarak elektrotlardaki potansiyelin analiziyle ölçülmesinin işe yarar olduğu gösterilmiştir. Öte yandan, domuz etinin tazeliğinin nicel olarak belirlenmesi ise elektrot potansiyelleri ve canlı bakteri sayısı üzerinde yapılan Çoklu Regresyon Analizi (ÇRA) yöntemi ile gerçekleştirilmiştir. Canlı bakteri sayısının deneysel ve tahmini sayıları

arasındaki ilişki ÇRA ile analiz edilmiştir. Sonuçlar determinasyon katsayısının 0,762 olduğunu göstermiştir, bu da lineer bir ilinti olduğuna işaret eder.

Gil (2008) ve arkadaşları tarafından yapılan bir başka çalışmada ise 16 potansiyometrik elektrottan oluşan e-dil, balık tazeliği analizinde kullanılmıştır [40]. Potansiyometrik elektrot çeşitleri metal, metal oksit, çözünmeyen metal tuzları ve grafitten oluşmuştur. Tüm elektrotlar, e-dil dizisini oluşturmak için değişik mürekkeplerle bir yüzey materyaline aktarılmışlardır. Şekil 1.2 (a), 16 elektrotla üretilmiş elektrot dizisini göstermektedir. Doku, pH, renk, mikrobiyal analiz, toplam uçucu bazik nitrojen (TVB-N- Total Volatile Basic Nitrogen) ve biyogenik aminler gibi balığın tazeliğini gösteren etmenler zamana göre belirlenmiştir. E-dilin balık filetolarının değişimini değerlendirmedeki yetkinliği ölçülmüştür. E-dil, TBA ve YSA ile örnekleri farklı zaman aralıklarına sınıflandırmada kullanılmıştır. Şekil 1.2 (b) farklı metalik elektrot tepkilerinin TBA sonuçlarını göstermektedir. Balık filetolarının farklı günlere ayrılması konusunda tatmin edici küme sonuçları elde edilmiştir. Ayrıca e-dil, nicel KEKK modelleri oluşturularak kimyasal ve biyokimyasal analizlerin sonuçlarını tahmin etmede de kullanılmıştır. Toplam biyogenik aminler, pH, TVB-N ve mikrobiyal analiz parametreleri ve 16 elektrottan oluşan e-dil arasında kayda değer ilişki bulunmuş ve ilinti katsayılarının 0,98'den büyük olduğu kaydedilmiştir.



Şekil 1.2. a) Kalın film teknolojisinde 16 elektrottan oluşan e-dilin tasarımı b) Metal elektrot tepkisinin TBA sonuçları [40].

Rodríguez-Mendez (2009) ve arkadaşları, sazangiller ailesi balıklarının bozulma işlemindeki farklı aminleri (amonyak, dimetilamin, trimetilamin, kadaverin ve histamin) ftalosiyanınle modifiye edilmiş ekran baskılı elektrotlar ve voltametrik karbon pasta

dizileri kullanarak tespit etmiştir [41]. Balık tazeliği ve balık bozulma süresi belirlemesinin TBA ve KEKK Diskriminant Analizi yapılmıştır. Ölçümden önce, 1 g balık kası kesiti alınmış, 25 ml 0.1 M (mol/L) KCl'de 5 dakika sonikasyon yapılarak özütü çıkarılmış; sıvı faz filtrasyon ile ayrılmış ve e-dil ölçümlerinde kullanılmıştır. Aynı araştırma grubu yakın zamanda Tirsi balığı (Ringa balığı) tazeliğinin izlenmesi için polipirel modifiyeli ekran baskılı elektrotların kullanıldığı voltametrik bir e-dil sunmuşlardır [36].

Biri soy metaller (iridyum (Ir), rodyum (Rh), Pt, Au) diğeri değerli metallerden (Ag, kobalt (Co), bakır (Cu), nikel (Ni)) oluşan ikili diziler halinde gruplanmış metal elektrotlar içeren voltametrik e-dilin taze morina balığının raf ömrü değerlendirmesi üzerine bir uygulama Ruiz-Rico (2013) ve arkadaşları tarafından sunulmuştur [37]. E-dil ölçümleri oda sıcaklığında doğrudan balık örnekleri üzerinde yapılmıştır. TVB-N, pH, nem, adenosin trifosfat (ATP) bağlantılı bileşenler, mezofilik bakteri ve enterobakterler sayımları standart yöntemlerle değerlendirilmiş ve e-dil verisiyle ilişkilendirilmiştir. Balık bozunumunda en önemli faktörlerden ikisi olan TVB-N ve mezofilik bakteri için başarılı bir KEKK elde edilmiş ve böylece voltametrik e-dilin morina balığı bozunumu değerlendirmedeki işe yararlılık potansiyeli doğrulanmıştır.

E-diller tadın değerlendirilmesinden çok sınıflandırmada kullanılmaktadır. Tat değerlendirmesi için Liu (2014) ve arkadaşları tarafından çığ kültür kirpi (balon) balığının tat peptitlerinin yapısal özelliklerini izole etmeyi ve incelemeyi hedefleyen ilginç bir araştırma sunulmuştur [38]. Balık kasının farklı kısımları izole edilmiş ve bu kısımlardaki matris destekli lazer desorpsiyon iyonizasyon uçuş spektrometresi metoduyla belirlenen tat peptitlerinin peptit yapılarına göre farklı tatlar (umami, acı, kokumi vs.) ortaya çıkarma ihtimalleri belirlenmiştir.

Gil (2011) ve arkadaşları, 10 günlük bir zaman diliminde soğutulmuş saklanan domuz eti filetosunun tazeliğini izlemek için Au, Ag, Cu, çinko (Zn), kurşun (Pb) ve grafit potansiyometrik elektrotlardan oluşan bir dizi uygulamışlardır [42]. Etten veri toplama elektrot dizisinin ve referans elektrotun doğrudan et örneğine saplanması ile gerçekleştirilmiştir. Her ölçümde yeni bir et parçası analiz edilmiştir. Veriler en az 5 dakika sonra alınmış, böylece elektrotların dengeye ulaşmaları sağlanmıştır. TBA ve YSA analizleri etin bozulma süresini belirlemenin mümkün olduğunu göstermiştir. Aynı araştırma grubu darbe voltametrisine dayanan, soy (Au, Pt, Rh, Ir, Ag) ve değerli (Ni, Co, Cu) elektrotlardan oluşan e-dilin salamuradaki ve kıyılmış etteki klorid, nitrit ve nitrat konsantrasyonlarının belirlenmesi uygulamasını sunmuşlardır [22].

Güney ve Atasoy (2013) istavrit balıklarının tazeliğini değerlendirmek için 8 metal oksit sensöründen oluşan e-burun cihazını geliştirmişlerdir [43]. İstavrit balıkları, 4°C’de buzlu balık pazarı depolama rezervuarında depolanmıştır. Balıklar yakalandıktan sonraki depolanma sürelerine göre, 1, 3, 5, 7, 9, 11 ve 13 günlük balıklar olarak sınıflandırılmış ve tazeliklerinin değişimi incelenmiştir. Çalışmada YSA, k-EYK ve önerilen karar ağacı yapısıyla sınıflandırma işlemi gerçekleştirilmiş, en yüksek doğruluk %97.22 ile önerilen karar ağacı ile elde edilmiştir. Önerilen algoritma DVM ve k-EYK algoritmalarının karar ağacı yapısında birleşiminden oluşmaktadır.

1.5.4. Bal

Balın besleyici ve iyileştirici olması onu en değerli doğal ürünlerden biri yapar. Tüketicilerin taleplerinin karşılanması için ticari balın karakterizasyonu gereklidir. Yüzlerce bileşenin karışımından oluşan ve ağdalı (viskozlu) bir madde olan balın karakterizasyonunda örneklerin ön işleme sokulması ve bileşenlerine ayrılmasına bağlı çeşitli sorunlarla karşılaşılır. Standart enstrümental metotlara bir alternatif olarak, balın analizi için e-dil kullanımının çeşitli uygulamaları önceden sunulmuştur [13,44]. Bu çalışmaların ana odağı ise balın botanik kökenini tespit etmek ve coğrafi sınıflamalar yapmaktır. E-dil analizinden önce bala sıklıkla uygulanan bir ön işlem prosedürü ise balı suda çözmek ve elde edilen örneklerde ölçümler yapmaktır.

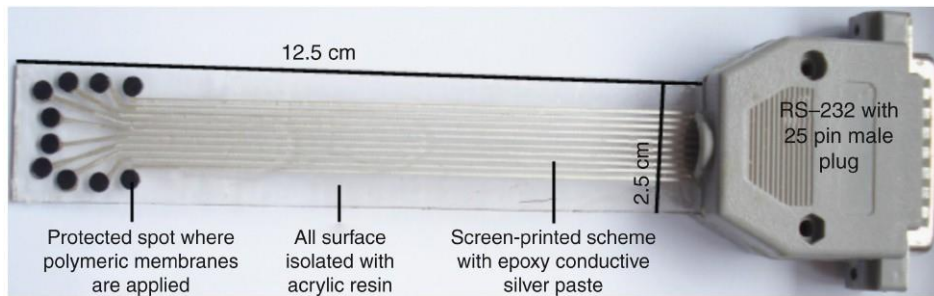
Wei ve Wang (2011) çeşitli monofloral balların metal voltametrik e-dil ve TBA ile floral kökenlerinin ayrılabilmesi olasılığını göstermişlerdir [44]. Daha yakın zamanda ise aynı yazarlar önceden bahsettikleri voltametrik e-dilin, monofloral balın floral ve coğrafi kökenini belirleme performansını α -Astree potansiyometrik e-dil ile karşılaştırmışlardır [45]. İki cihaz da KEKK Regresyonu, TBR ve En Küçük Kareler DVM metotlarıyla balların kategorilerini doğru tahmin edebilmişler ve farklı coğrafi kökenlere sahip dört farklı balı tahmin etmede kullanılan regresyon modellerinin oldukça kararlı olduğu görülmüştür.

Zakaria (2011) ve arkadaşları, farklı botanik kökenlerden gelen balların ayrımını gerçekleştirmiş ve Cyranose 320 e-burnu ve 7 farklı iyon seçici sensörden yapılmış kalkojenit bazlı potansiyometrik e-dili kullanarak şeker katışıklı balı ayrı ve birleştirilmiş ölçümlerle izlemiştir [13]. En yüksek sınıflandırma %94.44 ile sensör füzyonu uygulanarak Olasılık Tabanlı Sinir Ağında gözlenmiştir.

Ulloa (2013) ve arkadaşları, dört ticari Portekiz balının botanik kökenlerine göre, empedans e-dilinin ve morötesi görünür yakın kızılötesi spektroskopisinin birleşiminin TBA ve KA kemometrik teknikleriyle desteklenmesi sonucunda başarıyla ayrılmasını sunmuşlardır [16]. E-dil alüminyum (Al), Au, Pt ve ITO'dan oluşan dört elektrot levhadan oluşmaktadır. Kapasitans (C, nF), kondüktans (G, μ S) ve kondüktans/açısal frekans oranı verisi e-dilden alınmış, veri tepki eğrisi fitinin katsayılarını bulmak için ön işleminden geçirilmiş ve sonra bu katsayılar balın sınıflandırılması için spektroskopinin fit edilmiş verisi ile birlikte çok yönlü TBA yardımıyla analiz edilmiş ve %100 sınıflama yapılmıştır. Birleştirilmiş verinin, balın floral kökenindeki çeşitliliği belirlemede yalnızca e-dil kullanımından ve özellikle de sadece spektroskopik verinin kullanımından daha iyi olduğu gösterilmiştir.

Escriche (2012) ve arkadaşları çeşitli metal (Au, Ag, Cu) ve metalik bileşiklerden (gümüş oksit (Ag_2O), gümüş klorür (AgCl), gümüş karbonat (Ag_2CO_3), bakır oksit (Cu_2O)) yapılmış elektrotlara dayanan potansiyometrik e-dil ile bal örneklerini uygulanan termal işlemlere (çiğ, sıvılaştırılmış ve pastörize edilmiş) göre üçe ayırma girişiminde bulunmuşlardır [46]. Tatmin edici bir ayırım bulunamamış bunun yerine TBA ve YSA analizleri, e-dilin balın botanik kökenini sınıflamakta işe yarar olduğunu göstermiştir.

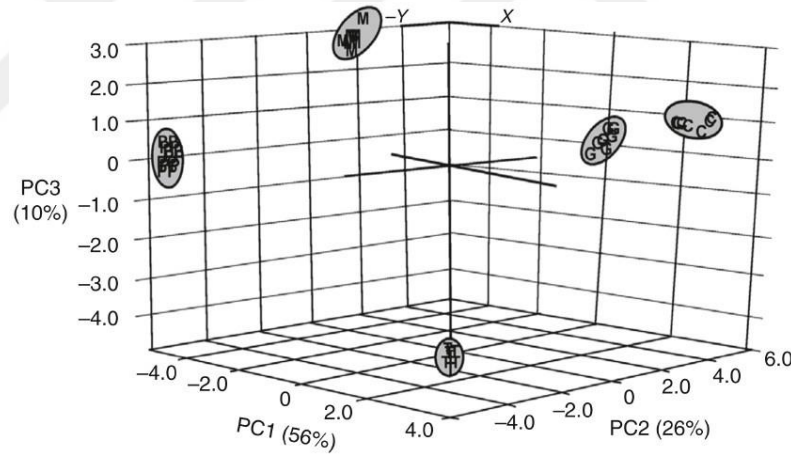
Bir başka e-dil ise 65 monofloral Portekiz balının sınıflandırılması için kullanılmıştır ve bu e-dil 20 adet çapraz seçiciliğe sahip lipit zardan oluşan potansiyometrik bir e-dildir [47]. Multi sensör sistemi bir ekran baskı yöntemi kullanılarak PVC'nin her iki tarafına da basılmıştır. Şekil 1.3'te bal ayırımı için çoklu sensör sistemi gösterilmektedir. E-dilin, balların renklerine göre (beyaz, açık amber, koyu amber) bir ön seçimi yapıldıktan sonra çok çeşitli floral kökenlerden gelen balları ayırabildiği gösterilmiştir.



Şekil 1.3. Bal ayırımı için çoklu sensör sistemi [47].

1.5.5. Meyve ve Sebzeler

Çoklu sensör sisteminin bir başka kullanım alanı da meyve ve sebzelerin tazeliğinin, tatlılığının ve içerdiği besinlerin, antioksidanların ve vitaminlerin nicel değerlendirilmesinin yapılmasıdır. Üzüm, şarabın ham maddesi olduğu için diğer meyvelere göre daha önemlidir. Üzüm analizleri genellikle fenolik antioksidanların değerlendirilmesine odaklıdır. Medina-Plaza (2014) ve arkadaşları, fenolik içeriklerine göre farklı tiplerdeki üzümleri ayırmak için, fenolik oksidazları (tirozinaz ve lakkaz) üzerine kurulu, nano yapılı, voltametrik biyosensörler içeren bir e-dil sistemi sunmuşlardır [9]. Üzüm örnekleri %50 oranında sulandırılmış şıralar biçiminde test edilmiştir. Şekil 1.4'te gösterilen TBA skoru grafiği, biyo-e-dilin, yapısındaki fenolik grupların sayısına göre fenolleri birbirinden ayırabildiğini ve 5 farklı üzüm çeşidini başarıyla ayırt edebildiğini göstermektedir.



Şekil 1.4. Farklı üzüm çeşitlerinin TBA skoru grafiği [9].

Campos (2013) ve arkadaşları, 7 çeşit İspanyol üzümünü ayırt etmek için metalik voltametrik e-dilin kullanımını sunmuştur [48]. Üzümler toplandıktan sonra, suyu posasından ayrılıp analiz edilmiştir. E-dilin toplam asidite ve şeker içeriği için iyi bir KEKK tahmin yeteneği %15'lik bir hata payıyla raporlanmıştır.

Kutyla-Olesiuk (2013) ve arkadaşları, mısır yapraklarında Pb birikimini değerlendirmek için minyatür iyon seçici elektrot dizisi üzerine kurulu analitik sürekli akış yöntemi kullanmışlardır [17]. 3-4 haftalık bitkilerden hasat edilen yapraklar, 5-10 mM kurşun nitrat ($Pb(NO_3)_2$) çözeltisinde 24 saat bekletilip, dondurarak kurutulmuştur.

Potansiyometrik ölçümler için, bitki materyali (0.2 g kuru ağırlık) 0.3 ml %96'lık sülfürik asit (H_2SO_4) ve 10 ml %30'luk hidrojen peroksit (H_2O_2) ile karıştırılmış ve sürekli akış modunda test edilmiştir. KEKK Diskriminant Analizi ile geliştirilen e-dilin, biyo endikasyon ya da fitoremediasyon süresince bitkilerin yetiştirme koşullarının tahmin edilmesinde potansiyel bir araç olduğu gösterilmiştir. Aynı araştırmacılar, farklı metotlarla kurutulan, çiğ ve kuru elmalardan alınan sulu özütlerin nicel ve nitel analizinde hibrit e-dil kullanmışlardır [18]. Sistem 5 adet potansiyometrik iyon seçici elektrot, Au ve glukoz oksidaz bazlı amperometrik sensörler, spektrofotometrik ve kondüktometrik ölçümler içermektedir. Çeşitli ölçüm tekniklerinden gelen verilerin kombinasyonunun, ayrı ayrı uygulanan metotlara göre kuru elma özütü örneklerinin ayrıştırılmasını daha iyi sağladığı gösterilmiştir.

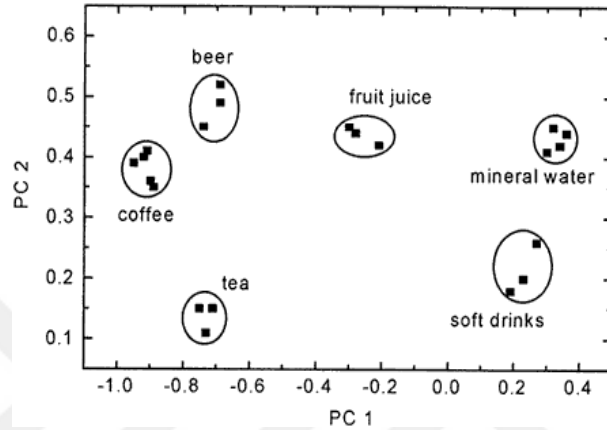
Beullens (2006) ve arkadaşları, dört domates kültürünün şeker ve asit profillerini belirlemek için 27 potansiyometrik sensörden oluşan bir e-dil kullanmışlardır [49]. Bu bilgilere dayanarak, domates kültürleri arasındaki şeker ve asit profili farklılıklarını tespit etmek için TBA ve Standart Diskriminant Analizi (SDA) gibi farklı çok değişkenli veri analizi teknikleri kullanılmıştır. E-dilin ve sönümlenmiş toplam yansıtma-Fourier dönüşümü kızılötesi spektroskopinin potansiyelleri, KEKK metodu kullanılarak örneklerin kimyasal içeriğini tahmin etmede kullanılmıştır. E-dilin, farklı domates kültürlerini ayırt edebildiği SDA kullanılarak kanıtlanmıştır.

1.5.6. Farklı İçecekler

E-dilin en yaygın uygulamalarından bir başkası da içecek tanıma ve ayırımıdır. Bira, şarap, çay, maden suyu, kahve gibi içeceklerin endüstrisinde nitel analize büyük bir talep vardır [50]. Çeşitli kromatograflar, spektrometreler gibi konvansiyonel analitik araçlarla karşılaştırıldığında e-dil içecek analizinde esneklik, basitlik ve maliyet açısından üstünlük göstermektedir.

Vlasov (2000) ve arkadaşları, farklı içecek türlerini doğru şekilde ayırmak için spesifik olmayan solüsyon sensörlerinin dizisine dayanan bir e-dil kullanmışlardır [51]. E-dil sensör dizisi iki tür potansiyometrik sensör içermektedir: (1) konvansiyonel klorür, sodyum, potasyum seçici sensörler ve pH sensörleri, (2) özel tasarlanmış, spesifik olmayan, kalkojenit camsı materyallere bağlı çapraz hassaslığa sahip gelişmiş sensörler. Bu elektrotlara bağlı olarak, e-dil içeceklerin nitel ve nicel analizinde kullanılmıştır. E-

dilin TBA ile ayırma işlemi Şekil 1.5'te gösterilmektedir ve görüldüğü üzere farklı içecekler ayrılabilir. Ayrıca e-dilin bazı iyonlar (Cu, demir (Fe), mangan (Mn), Zn, kalsiyum (Ca), magnezyum (Mg), Na, klor (Cl) ve sülfat (SO_4)) için nicel performansı da değerlendirilmiş ve kabul edilebilir nicel analiz hata paylarıyla tatmin edici sonuçlar elde edilmiştir.



Şekil 1.5. E-dilin TBA ile farklı tipteki içecekleri ayırma yetisi [51].

Polshin (2010) ve arkadaşları, biranın nicel analizi için 18 potansiyometrik kimyasal sensörden oluşan bir e-dil ortaya koymuşlardır [52]. 50 Belçika ve Hollanda birasının e-dil ve konvansiyonel analitik tekniklerle gerçek öz, gerçek fermentasyon derecesi, alkol içeriği, pH derecesi, acılık, renk, polifenol ve karbondioksit (CO_2) içeriği gibi değişik fiziko-kimyasal parametreleri ölçülmüştür. E-dil verisi ve fiziko-kimyasal veri arasındaki ilişkiyi incelemek için Standart İlişki Analizi (SİA) kullanılmıştır. Fiziko-kimyasal parametrelerin tahmini için e-dil verisine dayalı KEKK kalibrasyon modeli kurulmuştur. Sonuçlar e-dilin gerçek öz, acılık, alkol ve polifenol içeriği gibi biranın niteliğini değerlendirmekte kullanılabilecek parametreleri tahmin etme yetisi olduğunu göstermektedir.

1.5.7. Diğer Gıda Bağlantılı Elektronik Dil Uygulamaları

Yakın zamanlarda yapılan gıda analizine bağlı diğer e-dil uygulamaları belirli proteinlerin profillenmesi [21], bakteri ve gıda patojenlerinin belirlenmesi [46,53], alerjen taraması [54] ve evcil hayvan yemi kontrolü [55] ile uğraşmaktadır.

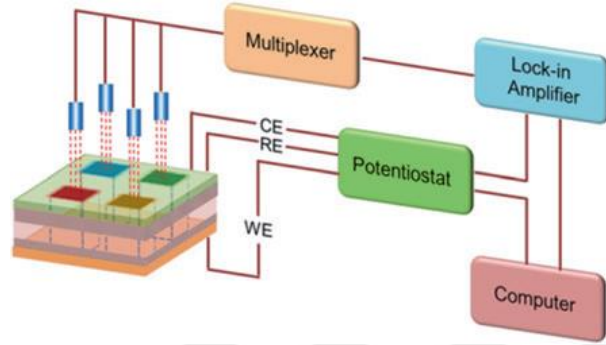
1.6. Elektronik Dil ile Su Ortamı Takibi

İnsanlar için yüksek toksisiteye (zehirliliğe) sahip olan su ortamı kirliliği dünyada büyük ilgi uyandırmıştır. Ancak çevresel takip geniş bir yelpazedeki farklı kimyasal bileşiklerin ve türlerinin sahada aynı anda ölçülmesini gerektirmektedir [56]. Sağlamlık, hassaslık ve geniş seçicilik sahada takibi engelleyen temel faktörlerdir. Küresel seçiciliğe, iyi bir kararlılığa ve yüksek hassaslığa sahip bir e-dil çevresel takip için umut vadeden bir yaklaşımdır. Çapraz seçicilikli e-dili kullanarak örneklerin gerçek zamanlı, çoklu bileşenli ölçümünü basit ve uygun bir araçla gerçekleştirmek olasıdır. Nicel analizde, yetkin bilgilerin elde edildiği sensörlerden gelen verilerin analizi için, e-dilde TBA ve YSA gibi pek çok veri işleme yöntemleri kullanılmaktadır. Aslında su ortamı takibi için farklı e-dillere dayanan pek çok çalışma yapılmıştır.

Di Natale (1997) ve arkadaşları, solüsyonlardaki birçok kimyasal türün konsantrasyonunu aynı anda tespit etmek için iyon-seçici elektrotlardan oluşan bir sensör dizisi kullanmışlardır [57]. E-dil sisteminde, kalkojenit camdan farklı şekillerde doplanmış (katkılanmış) elektrotlardan ve konvansiyonel elektrotlardan oluşan toplam 22 elektrot, sekiz katyon ve anyonun (Cu, kadmiyum (Cd), Fe, krom (Cr), Zn, Cl, SO₄ ve H) çapraz seçici ölçümlerinde kullanılmıştır. E-dilden en iyi performansı almak için ÇLR, KEKK, Lineer Olmayan En Küçük Kareler (LOEKK), Geri Yayılımlı Sinir Ağı (GYSA) da dahil olmak üzere farklı yaklaşımlar kullanılmıştır. Farklı analiz yaklaşımlarının dirençli kalibrasyon modellerini oluşturmak için yaklaşık 150 kimyasal solüsyon ölçülmüştür. Sonuçlar, modüler modellerin çeşitli yaklaşımların birlikte kullanımıyla farklı iyonların konsantrasyonlarının tahminindeki hataları oldukça azaltılabileceğini göstermiştir. Oldukça iyi olan tahmini konsantrasyon sonuçları (mutlak hata %6'dan az) elde edilmiş ve e-dilin çevre izleme için uygulanabilirliği doğrulanmıştır.

Konvansiyonel e-dillerdeki çapraz-seçiciliğe rağmen parazit reddi ve sensörlerin seçiciliğini arttırmaya yönelik bazı çalışmalar da yapılmıştır. Ha (2012) ve arkadaşları, ağır metal katyonlarının belirlenmesi için ışııkla etkileşen potansiyometrik sensör ve PVC membran modifikasyonları kullanılan çoklu-sensör dizisini kullanmışlardır [58]. Parazit, silikon substratın bir kısmının borla fazlaca doplanması ve yüzeyin kalın oksit tabakasıyla kaplanması ile azaltılmıştır. Çoklu-sensör dizisinin ölçüm kurulumunun şematik diyagramı Şekil 1.6'da verilmiştir. Elektro-kimyasal sınıflandırma için elektrot sistemi, potansiyel kontrolü için ise potansiyostat kullanılmıştır. Sonuçlar, farklı ışııkla etkileşen

potansiyometrik sensörlerin Zn, Cd ve Pb'ye yüksek hassasiyet gösterdiklerini kanıtlamıştır. Ayrıca uzun vadedeki kararlılık da 4 saat sürekli testte standart sapmanın 0.12mV'tan az olmasıyla kanıtlanmıştır.



Şekil 1.6. Çoklu-sensör dizisi ölçüm kurulumunun şematik diyagramı [58].

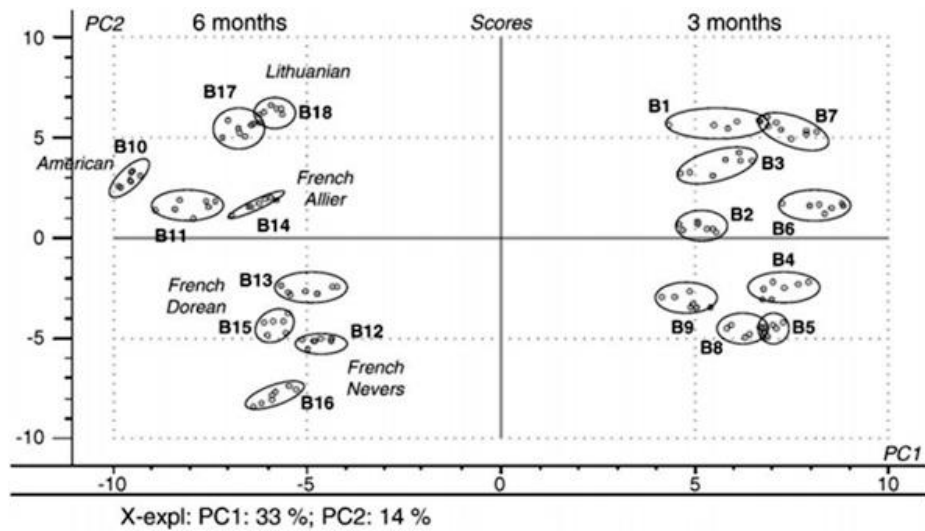
Spektroskopi ve analitik kimya gibi çevresel analiz için kullanılan konvansiyonel analitik yöntemlerin yanı sıra, su ortamı takibi için e-dil de çok önemli bir role sahiptir. Diğer yaklaşımlarla karşılaştırıldığında, e-dil, fark etme zamanı, ekipman masrafı, sahada ölçüm ve çoklu-elemanlı analiz için gözle görülür avantajlar sağlar.

1.7. Elektronik Dil ile Süreç İzleme

E-dil, son yıllarda çeşitli alanlardaki potansiyel uygulamaları sebebiyle hızla gelişmiştir. Yukarıda bahsi geçen uygulamaların yanında, e-dil endüstride süreç izleme amacıyla da kullanılmaktadır. Süt ve süt ürünleri endüstrisi, bira endüstrisi ve fermantasyon endüstrisi gibi endüstrilerde sürecin kalite parametrelerini izlemek çok önemlidir. Normalde bu endüstrilerde süreç kontrolü farklı olayların zamanları kontrol edilerek yapılır ve gerçek zamanlı bilgi süreç sırasında elde edilemez. Süreç izleme için analitik araçlar, süreçteki karmaşık ortamlara bağlı olabilmektedir ve kirlenme ihtimaline karşı ortama hiçbir ekstra madde girmemelidir. Dolayısıyla elektrokimya gibi bazı yaklaşımların gerçekleştirilmesi süreç referans elektrotla izlendiğinden oldukça zordur. Basit ve dayanıklı e-dil, endüstriyel uygulamalarda umut vadeden bir araç olarak görülmektedir.

Winqvist (2005) ve arkadaşları, süt ürünleri endüstrisi uygulaması için özel tasarlanmış bir voltametrik e-dil öne sürmüşlerdir [59]. Biri Au, Pt ve Rh işlenmiş dış malzemesinden üç elektrot ve diğeri de Au, Pt, Rh ve paslanmaz çelik işlenmiş, Polieter Eter Ketondan dört elektrot olmak üzere iki farklı tip elektrot kullanılmıştır. Referans elektrotta kirlilik durumunda, iki elektrot konfigürasyonu ile yüksek genlikli işaret voltametrisi ile akım ölçülmektedir. Tüm elektrotlar iletkenliği, yoğunluğu ve sıcaklığı gerçek zamanlı takip edebilmek için doğrudan süreç hattına batırılmışlardır. Çalışma, farklı kaynaklardan farklı zamanlarda alınan sütün belirlenebildiğini göstermiştir. Bu bilgi süt ve süt ürünleri endüstrisinde çok önemli olan tadı bozulmuş sütler için rehberlik etmektedir.

Parra (2006) ve arkadaşları, kırmızı şarabın yıllanmasını izlemek ve meşe varillerde yıllandırılan şarap örneklerini değişik özelliklerine göre (ahşabın kökeni ve pişirilme miktarı) ayırmak için voltametrik elektrotlara dayanan yenilikçi bir hibrit sensör dizisi kullanmışlardır [60]. Hibrit sensör dizisi, bir dizi katkılanmış polipirol sensörlerin dahil olduğu kimyasal yollarla düzenlenmiş elektrotlar, metalofitalosiyenin bileşiklerle düzenlenmiş karbon macunu elektrotlar ve perilen imit türevleriyle düzenlenmiş karbon macunu elektrotlar olmak üzere üç familyadan oluşmaktadır. Sensör materyallerindeki çeşitlilik, seriyi oluşturan sensörlerin tepkilerinde yüksek bir çapraz-seçicilik elde edilmesine olanak sağlamıştır. Şekil 1.7, 3 ay (B1–B9) ve 6 ay (B10–B18) dokuzar meşe varilinde yıllandırılan şarap için TBA skor grafiğini göstermektedir.



Şekil 1.7. E-dilin 3 ve 6 aydır yıllandırılan şarapla ilgili verilerinin TBA grafiği [60].

Şarap örnekleri, 3 ay yıllandırılan kırmızı şaraplar sağda, 6 ay yıllandırılanlar solda olmak üzere iki gruba ayrılmıştır. Sonuçlar, e-dilin meşe varillerdeki yıllandırma sürecini izleme yeterliliğine sahip olduğunu işaret etmektedir.

1.8. Sınıflandırma Algoritmaları

1.8.1. Destek Vektör Makinesi Algoritması

Temelleri ilk olarak Vapnik ve Chervonenkis tarafından hesaplanabilir öğrenme teorisinin önemli bir parçasını oluşturan ve öğrenmenin temel teorisi olarak bilinen Vapnik-Chervonenkis Teorisi kapsamında 1960'lı yıllarda atılan destek vektör makineleri (DVM), 1992 yılında Vapnik, Baser ve Guyon tarafından sunulmuştur [61].

DVM el yazısı tanıma, zaman serisi analizi, konuşma tanıma, meme kanseri teşhisi gibi birçok alanda başarı ile uygulanmıştır. Diğer sınıflandırma yöntemleri ile karşılaştırıldığında, eğitim süresinin oldukça uzun olmasına rağmen, yüksek güvenilirliği, ezbere öğrenmeye olan mukavemeti ve lineer olmayan sınıflandırmadaki başarı düzeyleri ile DVM tercih edilen bir yöntem olmuştur.

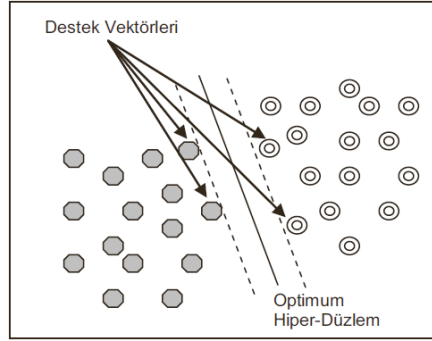
Bu yöntem, sınıflandırmayı bir doğrusal ya da doğrusal olmayan bir fonksiyon yardımıyla yerine getirir. DVM yöntemi, veriyi birbirinden ayırmak için en uygun fonksiyonun tahmin edilmesi esasına dayanır. Daha çok makine öğrenmesi yöntemleri arasında yer alan bu yöntem günümüzde veri madenciliği alanında da tercih edilmeye başlanmıştır.

DVM hem lineer olarak ayırt edilebilen hem de lineer olarak ayırt edilemeyen veri kümelerini sınıflandırabilir. Uygun bir dönüşüm ile her zaman veri bir hiper düzlem ile iki sınıfa ayrılabilir. Hiper düzleme en yakın öğrenme verileri destek vektörleri olarak adlandırılır.

DVM esasen çok boyutlu ve lineer olmayan sınıflandırma için geliştirilmiştir. Ancak konunun anlaşılabilmesi için, DVM kavramları ve süreci ilk olarak lineer ve iki boyutlu bir örnek üzerinde açıklanacaktır. Sadece lineer olarak sınıflandırılabilir ve iki boyutlu veri için, DVM en basit olarak aşağıdaki şekilde tanımlanabilir.

DVM karar doğrusuna bağımlı olarak belirlenen destek noktaları arasındaki genişliği maksimize etmeyi amaçlayan denetimli bir öğrenme algoritmasıdır.

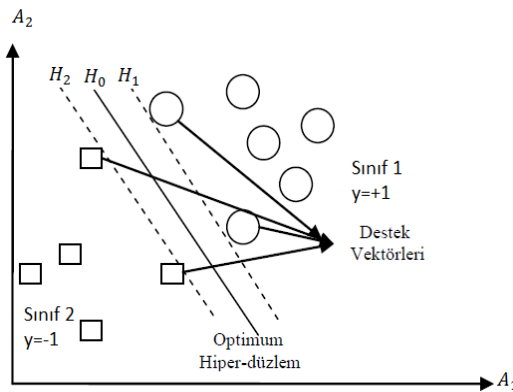
Şekil 1.8’de görüldüğü üzere sınırı maksimuma çıkararak en uygun ayrımı yapan hiper düzlem, optimum hiper düzlem ve sınır genişliğini sınırlandıran noktalar ise destek vektörleri olarak adlandırılır.



Şekil 1.8. Optimum hiper düzlem ve destek vektörleri [62].

1.8.1.1. Lineer Ayrılabilme Durumu

Lineer olarak ayrılabilme durumunda sınıflar ayırıcı aşırı düzlem ile ayrılabilir. DVM, bu aşırı düzlemin örnek gruplarına eşit uzaklıkta olmasını amaçlayan bir yöntemdir. Örnekle açıklarsak, Şekil 1.9 üzerinde yer alan H_1 ve H_2 bizim en uygun aşırı düzlemlerimizdir. Yani sınıflara olan mesafesini minimum tutan, aradaki mesafeyi de maksimum tutan düzlemlerdir. Bu iki aşırı düzlemin ortasından geçen H_0 ise optimal ayırma düzlemi adı verilen bir düzlemdir. Bu düzlem, en uygun aşırı düzlemlerimizin tam ortasından geçen düzlemdir.



Şekil 1.9. Lineer olarak birbirinden ayrılabilen veriler arasındaki muhtemel en büyük boşluk [63].

Bir hiper düzlem üzerindeki noktalar cinsinden H_0 düzlemi aşağıdaki gibi ifade edilebilir:

$$H_0 : W^T X + b = 0 \quad (1.1)$$

Bu ifadeyi şu şekilde de yazabiliriz:

$$\sum_{i=1}^n w_i x_i + b = 0 \quad (1.2)$$

Burada W ağırlık vektörünü $W = \{w_1, w_2, \dots, w_n\}$, n ise niteliklerin sayısını göstermektedir. İfade içinde yer alan b ise sabit bir sayıyı göstermektedir. Kolaylık olsun diye A_1 ve A_2 diye iki niteliğin var olduğunu kabul edelim. Eğitim kümesi ise, iki boyutlu uzay söz konusu olduğu için $X = (x_1, x_2)$ biçimindedir. Burada x_1 ve x_2 değerleri X için A_1 ve A_2 'nin değerleri olarak göz önüne alınır. H_1 hiper düzlemi şu şekilde ifade edilebilir:

$$H_1 : W^T X + b = +1 \quad (1.3)$$

Benzer biçimde H_2 hiper düzlemi,

$$H_2 : W^T X + b = -1 \quad (1.4)$$

biçiminde ifade edilir. (1.1) ifadesinde belirtilen hiper düzlemin üst tarafında kalan noktalar aşağıdaki eşitsizliğe uymaktadır:

$$W^T X + b > 0, \quad y_1 = +1 \quad (1.5)$$

Benzer biçimde hiper düzlemin alt kısmında kalan noktalar da aşağıdaki eşitsizliğe uymaktadır.

$$W^T X + b < -1, \quad y_2 = -1 \quad (1.6)$$

(1.5) ve (1.6) eşitsizlikleri birleştirilerek tek bir eşitsizlik biçimine dönüştürülebilir:

$$y_i(W^T X + b) - 1 \geq 0 \quad \forall_i \quad (1.7)$$

Burada H_1 ve H_2 hiper düzlemlerini göz önüne alalım. Bu düzlemler üzerindeki gözlemler destek vektör adını alır. Bir destek vektör ile (1.1) ile gösterilen hiper düzlemi arasındaki uzaklık, P hiper düzlemi üzerindeki bir nokta olmak üzere,

$$d = \frac{|WX_p \mp b|}{\|w\|} = \frac{|w_1 x_{1p} + w_2 x_{2p} + \dots + w_n x_{np} + b|}{\sqrt{w_1^2 + w_2^2 + \dots + w_n^2}} \quad (1.8)$$

bağıntısı ile bulunur. O halde X_1 destek vektörü ile H_2 hiper düzlemi arasındaki uzaklık ve (1.3) ve (1.4)'ten dolayı,

$$d = \frac{|W^T X_1 + b|}{\|w\|} = \frac{1}{\|w\|} \quad (1.9)$$

biçiminde hesaplanır. Bu durumda X_1 destek vektörü ile H_2 hiper düzlemi arasındaki uzaklık,

$$m = 2d = \frac{2}{\|w\|} \quad (1.10)$$

olarak belirlenir.

Optimum hiper düzlemin sınırının maksimuma çıkarılması için $\|w\|$ ifadesinin minimum hale getirilmesi gerekir. Bu durumda en uygun hiper düzlemin belirlenmesi aşağıdaki sınırlı optimizasyon probleminin çözümünü gerektirir.

$$\min \left(\frac{1}{2} \|w\|^2 \right) \quad (1.11)$$

Buna bağlı sınırlamalar ise;

$$y_i(wx_i + b) - 1 \geq 0 \text{ ve } y_i \in \{1, -1\} \quad (1.12)$$

şeklinde ifade edilir. Bu optimizasyon problemi Lagrange denklemleri kullanılarak çözülebilir. Bu işlem sonrasında;

$$L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^k a_i y_i (wx_i + b) + \sum_{i=1}^k a_i \quad (1.13)$$

eşitliği elde edilir. Sonuç olarak doğrusal olarak ayrılabilen iki sınıflı bir problem için karar fonksiyonu aşağıdaki şekilde yazılabilir.

$$f(x) = \text{sign} \left(\sum_{i=1}^k \lambda_i y_i (xx_i) + b \right) \quad (1.14)$$

1.8.1.2. Lineer Ayrılama Durumu

Birçok problemde verilerin doğrusal olarak ayrılması mümkün değildir. Bu durumda eğitim verilerinin bir kısmının optimum hiper düzlemin diğer tarafında kalmasından kaynaklanan problem pozitif bir yapay değişkenin (ξ_i) tanımlanması ile çözülür. Sınırın maksimum hale getirilmesi ve yanlış sınıflandırma hatalarının minimum hale getirilmesi arasındaki denge pozitif değerler alan ve C ile gösterilen bir düzenleme parametresi ($0 < C < \infty$) tanımlanmasıyla kontrol edilebilir. Düzenleme parametresi ve yapay değişken kullanılarak doğrusal olarak ayırım yapılamayan veriler için optimizasyon problemi:

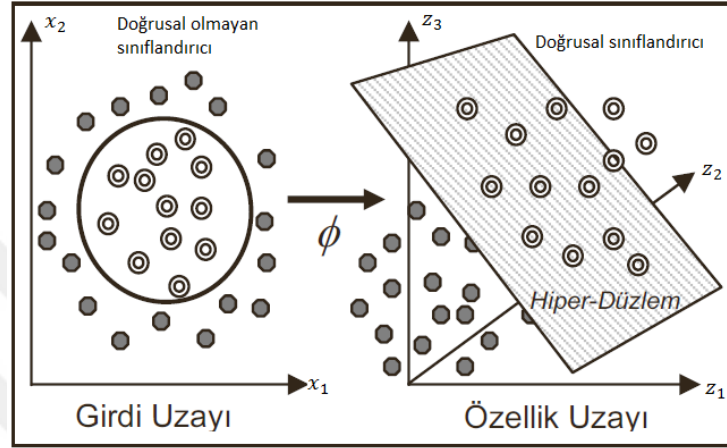
$$\min \left[\frac{\|w\|^2}{2} + C \cdot \sum_{i=1}^r \xi_i \right] \quad (1.15)$$

şeklini alır. Buna bağlı sınırlamalar ise;

$$y_i(w \cdot \varphi(x_i) + b) - 1 \geq 1 - \xi_i \quad i = (1, 2, \dots, n), \xi_i \geq 0 \quad (1.16)$$

şeklinde ifade edilir.

İfade edilen optimizasyon probleminin çözümü için Şekil 1.10'da görüleceği üzere girdi uzayında doğrusal olarak ayrılamayan veri, özellik uzayı olarak tanımlanan yüksek boyutlu bir uzayda görüntülenir. Böylece verilerin doğrusal olarak ayrımı yapılabilen ve sınıflar arasındaki hiper düzlem belirlenebilmektedir.



Şekil 1.10. Kernel fonksiyonu ile verinin daha yüksek bir boyuta dönüştürülmesi [62].

DVM matematiksel olarak $K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$ şeklinde ifade edilen bir kernel fonksiyonu yardımıyla doğrusal olmayan dönüşümler yapabilmekte ve bu şekilde verilerin yüksek boyutta doğrusal olarak ayrımına imkan sağlamaktadır. Sonuç olarak, kernel fonksiyonu kullanarak doğrusal olarak ayrılamayan iki sınıflı bir problemin çözümü ile ilgili karar kuralı aşağıdaki şekilde yazılabilir:

$$f(x) = \text{sign} \left(\sum_i a_i y_i \varphi(x) \cdot \varphi(x_i) + b \right) \quad (1.17)$$

DVM ile gerçekleştirilecek bir sınıflandırma işlemi için kullanılacak kernel fonksiyonu ve bu fonksiyona ait optimum parametrelerin belirlenmesi esastır.

1.8.1.3. Çok Sınıflı Destek Vektör Makineleri

Şimdiye kadar bahsedilen konular, sınıf sayısının 2 olması durumunda geçerli olan işlem adımlarıydı. Sınıf sayısının 2'den fazla olması durumunda farklı yaklaşımlar söz konusudur. DVM'nde çok sınıf probleminin çözümü için 2 yöntem kullanılmaktadır:

1-) Lagrange fonksiyonunun çok sınıfla işlem yapabilecek hale getirilmesidir. Yalnız bu yöntem, sınıf sayısının artması ile hata miktarının da artması nedeniyle tercih edilen bir yöntem değildir.

2-) DVM'nin ikili sınıflar halinde sınıflandırmalar yapacak biçimde çalıştırılmasını sağlamaktır. Bu yaklaşımda bire karşı bir, bire karşı hepsi ve döngüsel olmayan graf gibi farklı yöntemler kullanılmaktadır.

Bire karşı bir yaklaşımında her örnek veri seti, diğer örnek veri setleri ile ayrı ayrı eğitilir. Bire karşı hepsi yönteminde, her bir örnek veri seti bir sınıf, geriye kalan sınıfların tamamı bir sınıf olarak kabul edilir ve eğitim bu şekilde yapılır. Döngüsel olmayan graf yöntemi ise bire karşı bir yöntemiyle aynıdır. Eğitim bire karşı bir yöntemine göre yapılır, fakat test aşamasında eğitim örnekleri kontrol edilmek yerine, sınıflandırılacak elemanın ait olmadığı düşünülen sınıflar elenerek işlem yapılır.

1.8.2. k-En Yakın Komşuluk Algoritması

Sınıflandırma problemini benzerlik fonksiyonlarını kullanarak çözen kontrollü ve denetimli bir öğrenme algoritmasıdır. Bu yöntem, sınıfları belli olan bir örnek kümesindeki gözlem değerlerinden yararlanarak, örneğe katılacak yeni bir gözlemin hangi sınıfa ait olduğunu belirlemek amacıyla kullanılır [63].

Söz konusu bu yöntem örnek kümedeki gözlemlerin her birinin, sonradan belirlenen bir gözlem değerine olan uzaklıklarının hesaplanması ve en küçük uzaklığa sahip k sayıda gözlemin seçilmesi esasına dayanmaktadır. Uzaklıkların hesaplanmasında, i ve j noktaları için farklı hesaplama metotları bulunmaktadır. Bunlardan en çok kullanılan uzaklık bağıntısı (1.18) ile ifade edilen Öklid uzaklığıdır ve bu çalışmada bu bağıntı tercih edilmektedir.

$$d(i, j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (1.18)$$

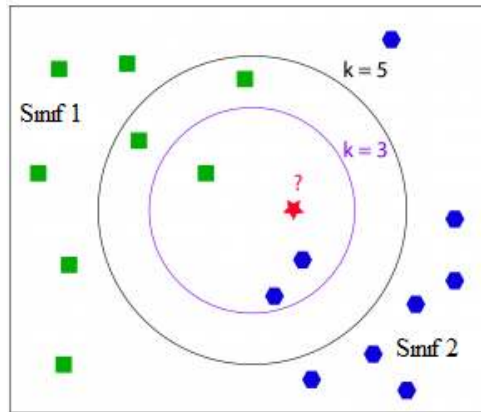
Diğer bir uzaklık ölçüsü Manhattan uzaklığıdır. Bu uzaklık, gözlemler arasındaki mutlak uzaklıkların toplamı alınarak hesaplanır. Söz konusu uzaklık şu şekilde ifade edilir:

$$d(i, j) = \sum_{k=1}^p \left(|x_{ik} - x_{jk}| \right) \quad i, j = 1, 2, \dots, n; k = 1, 2, \dots, p \quad (1.19)$$

p sayıda değişken göz önüne alınarak gözlem değerleri arasındaki uzaklığın hesaplanması söz konusu ise Minkowski uzaklık bağıntısı kullanılabilir. Söz konusu uzaklık şu şekilde hesaplanır:

$$d(i, j) = \left[\sum_{k=1}^p \left(|x_{ik} - x_{jk}|^m \right) \right]^{\frac{1}{m}} \quad i, j = 1, 2, \dots, n; k = 1, 2, \dots, p \quad (1.20)$$

Şekil 1.11’de k değerinin 3 ve 5 olduğu durumlar için k-EYK algoritmasının sınıflandırma örneği verilmiştir. Sistemde yeşil dikdörtgen (sınıf 1) ve mavi altıgen (sınıf 2) olmak üzere iki farklı sınıf vardır. Kırmızı yıldız, sınıfı bilinmeyen veridir. Bu yöntemle göre bilinmeyen verinin sınıfı, en çok komşuluğu bulunduğu sınıf olarak kabul edilir. k değerinin 3 alındığı durumda, kırmızı yıldız, mavi altıgen sınıfına karşılık gelir, çünkü en yakın 3 komşusu 2 adet mavi altıgen 1 adet yeşil dikdörtgen içerir. k değerinin 5 olduğu durumda ise kırmızı yıldız, yeşil dikdörtgen olarak sınıflandırılır, çünkü en yakın 5 komşusu mavi altıgenlerden daha fazla yeşil dikdörtgen içerir.



Şekil 1.11. k-EYK yöntemi için basit bir örnek [64].

1.8.3. Rastgele Orman Algoritması

Breiman tarafından ortaya konulan Rastgele Orman (RO) yöntemi, çok sayıda karar ağacından oluşan bir toplu öğrenme yöntemidir [65]. Ormanın eğitimi aşamasında karar ağaçları oluşturulmakta ve tekil ağaçlardan en çok oyu alan sınıf karar sınıfı olarak belirlenmektedir.

Ormanda yer alan her bir karar ağacı, bootstrap tekniği ile orijinal veri setinden örneklem seçilmesi ve her karar düğümünde tüm değişkenler içinden belirlenen sayıda rastgele değişkenin seçilmesi ile oluşturulmaktadır. Bu yöntemde CART algoritması ile ağaçlar oluşturulur ve ağaçlar budanmaz. CART algoritması veri setinin hangi değişkenden başlayarak dallara ayrılacağına “bilgi kazancını” kullanarak karar verir. Ayrıca dallara ayrılmak için seçilen değişkenin uygun test kriteri “gini katsayısı” ile belirlenir. Veri seti, test ve öğrenme olmak üzere iki bölüme ayrılmamış olsa da, veri setinin bütünü ele alınarak model kurulabilmekte ve modelin performansı modelin iç hatası olarak da bilinen Out-Of-Bag (OOB) hata oranı ile ölçülebilmektedir. Oluşturulan her ağaca önceden hesaplanan OOB hata oranına göre bir ağırlık verilir. En düşük hata oranına sahip ağaç, en yüksek ağırlığı, en yüksek hata oranına sahip ağaç en düşük ağırlığı alır. Her ağaç belirlenen ağırlığa göre yaptığı sınıf tahmini için bir oy verme işlemine tabi tutulur. Ağaçların oyu oluştuktan sonra RO algoritmasında bu ağırlıklı oylar toplanır.

RO algoritması aşağıdaki adımlardan oluşur:

- 1) Orijinal veri setinden n tane bootstrap örnekleme yapılır. Her örneklemin $2/3$ 'ü ağacı oluşturmak için öğrenme verisi olarak kullanılır (inBag).
- 2) Her bootstrap örnekleme için budanmamış sınıflama veya regresyon ağacı aşağıdaki adımlar kullanılarak oluşturulur.
 - inBag veri setinden her düğümde bütün tahmin değişkenleri içerisinde en iyi değişkeni seçmek yerine rastgele m tane tahmin değişkeni seçilir ve bunların içerisinde en iyi dallara ayıracak (en çok bilgi kazancı sağlayacak) olanı belirlenir.
 - Belirlenen tahmin değişkeni için en iyi dallanma kriteri gini indeksi ile hesaplanır ve hesaplanan değere göre veri seti her düğümde iki alt dala ayrılır.
 - Yukarıdaki verilen işlemler aşağıya doğru yaprak düğüm elde edilinceye kadar her düğümde tekrar edilir.

Breiman tarafından varsayılan m değeri regresyon ağaçları kurulurken $p/3$, sınıflama ağaçları kurulurken ise $p^{1/2}$ olarak önerilmiştir. Burada p değeri toplam tahmin edici değişken sayısını ifade etmektedir.

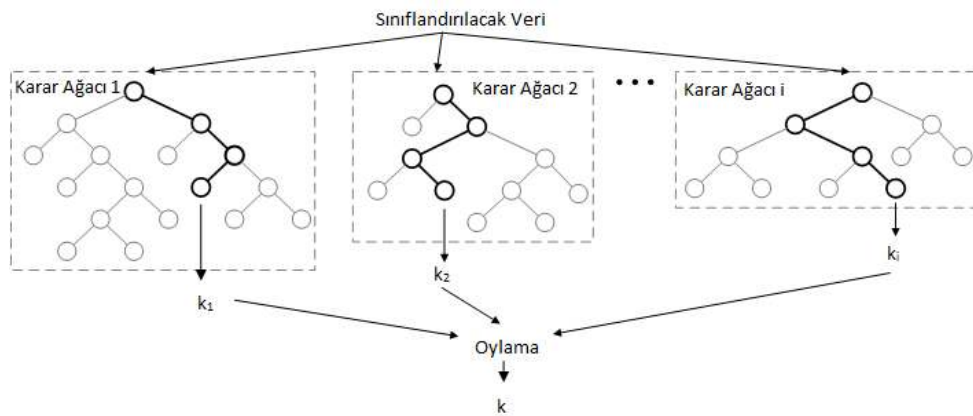
3) n tane ağacın ayrı ayrı yapmış olduğu tahminleri bir araya getirerek yeni bir tahminde bulunulur:

- Sınıflama ağaçları için en çok oyu alan sınıf final tahmini olarak seçilir.
- Regresyon ağaçları için ise yapılan oylamanın ortalaması alınarak nihai tahmin yapılır.

Veri setinden hata oranını hesaplamak için;

- Her karar ağacı oluşturulurken, bootstrap aşamasında, bootstarp örnekleme, ağaç oluşturulacak veri (inBag) ve ağaç oluşturmak için kullanılmayan veri (OOB verisi) olmak üzere ikiye ayrılır. OOB verisiyle ağaç test edilir ve hata oranı tahmini yapılır.
- Bireysel ağaçların yaptığı OOB tahminleri bir araya getirilir. Bu tahminlerden ormanın hata oranı kestirimi yapılır.

RO algoritması ile sınıflandırma işlemi şu şekilde özetlenebilir. Şekil 1.12'de RO algoritmasının genel mimarisi verilmiştir. Sınıflandırılacak veri rastgele alt kümelerle ayrılıp alt kümelerin her birinden karar ağacı oluşturulur. Veri her bir karar ağacına göre sınıflandırma işlemine tabi tutulur. Karar ağaçlarından çıkan sonuçlar oylama yöntemiyle toplanarak veri en çok oyu alan sınıfa dahil edilir.



Şekil 1.12. RO algoritmasının genel mimarisi [66].

RO algoritmasının avantajları çok yüksek sayıda veriyi hızlı bir şekilde sınıflandırabilmesi, eksik veriler için öngöründe bulunabilmesi, yüksek başarımla çalışmasıdır.

1.9. Özellik Seçimi

Özellik seçimi, kümeleme veya regresyon işlemleri için kullanılacak özelliklerin belirlenmesi aşamasında, tüm özellik kümesi sütunlarından bağımlı değişkenle olan ilişkiyi açıklamada, ilgisiz sütunların elenmesi ve açıklayıcı gücü yüksek sütun alt kümelerinin belirlenmesi işidir [67].

Özellik seçimi ile özelliklerin tümü değil, seçilen bir alt kümesi sınıflandırma için kullanılır. Bu da daha düşük boyutlu uzayda çalışmak anlamına gelir. İlk bakışta, veri kümesindeki tüm özelliklerin analize konu edilmesiyle, sınıflandırma veya bağımlı değişkeni açıklayan regresyon algoritmalarının başarılı sonuçlar vereceği akla gelmektedir. Oysa bu düşünce pek çok özellik içeren veri kümelerinde her zaman doğru olmayabilir. Veri kümesindeki her özellik bağımlı değişken hakkında açıklayıcı ya da tahminleyici bilgi taşımayabilir. Dolayısıyla özelliklerin tahminleyici bilgi taşıma durumuna göre ayırt edilip analize konu edilmesi gerekir. Başka bir ifadeyle, veri kümesi içindeki bazı özellikler işlem performansını olumsuz etkileyecek gürültüye sahip olduğundan bu özelliklerin veri kümesi içinden silinmesi, işlem sonucunun doğruluğunun artmasında etkili olabilmektedir. Diğer taraftan algoritmalarda kullanılacak veri boyutunun azaltılması da işlem gücü, hafıza ihtiyacı ve depolama alanı gibi işlem süreci üzerinde etkili konularda zaman tasarrufu sağlar. Özellik seçimi için iki genel yaklaşım bulunmaktadır.

1.9.1. Filtre Modeli Öznitelik Seçme Algoritmaları

Filtre modelli öznitelik seçme algoritmaları istatistiksel varsayımlar kullanarak niteliklerin sonuca ulaşmaya katkısını hesaplayarak nitelikleri puanlar ve eşikleme yaparak öznitelikleri seçer. Filtre modelli öznitelik seçme algoritmalarının en büyük avantajı hızlı ve sınıflandırma algoritmasından bağımsız olmasıdır. Bunun yanında sınıflandırma başarı oranının düşük olması dezavantajıdır.

1.9.1.1. Bilgi Kazancı

Bilgi kazancı, entropiye dayalı özellik seçim yöntemidir. Entropi, bir sistemdeki düzensizliğin ya da belirsizliğin ölçüsüdür, (1.21) nolu formüldeki gibi ifade edilmektedir. Entropi 0 ve 1 aralığında değerler alır ve 1 değerine yaklaştıkça belirsizlik artar. Yüksek entropiye sahip veri daha çok bilgi içerir.

$$E(D) = - \sum_{k=1}^m p_i \log_i(p_i) \quad (1.21)$$

p_i, D veri kümesindeki “ i ” sınıfının olasılığıdır ve “ i ” sınıfına düşen örnek sayısının tüm veri kümesindeki toplam örnek sayısına bölünmesiyle elde edilir.

Bilgi Kazancı yöntemi, en ayırt edici özelliği belirlemek için kullanılır ve veri kümesindeki her bir özellik için ölçülür. D veri kümesi, n tane alt bölüme X özelliğinden bölünecekse X ’e ait bilgi kazancı hesaplanması (1.22) numaralı formülle gerçekleştirilir.

$$\text{BilgiKazancı}(D, X) = E(D) - \sum_{k=1}^n p(D_i) E(D_i) \quad (1.22)$$

$E(D)$; veri kümesinin X üzerinden bölünmeden önceki entropisini, $E(D_i)$; i alt bölümünün X üzerinden bölünme olduktan sonraki entropisini ve $p(D_i)$ ise i alt bölümünün X üzerinden bölünme olduktan sonraki olasılığını göstermektedir [65]. Veri kümesinin bölünmeden önceki belirsizliğinin yüksek olması, verinin bilgi verici niteliğinin olduğunu göstermektedir. Bölünmeden sonraki belirsizliğinin düşük çıkmasıysa bu yöntemin veriyi dallara ayırma işlemini düzgün yaptığını göstermektedir. (1.22) numaralı formüle göre $E(D)$ ’nin yüksek çıkarken $p(D_i)E(D_i)$ çarpımları toplamının düşük çıkması bilgi kazancını artırmaktadır.

1.9.1.2. Kazanım Oranı

Bilgi kazancı yöntemi çok çeşitli değerlere sahip özellikleri seçme eğiliminde olduğundan sonuçları sapmalı bir yöntemdir. Bu sapmanın azaltılması amacıyla Kazanım Oranı yöntemi oluşturulmuştur. Sapmayı azaltmak için bölünme bilgisi kullanılmaktadır. Bölünme bilgisi (1.23) numaralı formülde gösterilmektedir.

$$BölünmeBilgisi(S) = -\sum_{i=1}^v \left(\frac{|S_i|}{|S|} \right) \log_2 \left(\frac{|S_i|}{|S|} \right) \quad (1.23)$$

Kazanım Oranı, bilgi kazancı değerlerini, bölünme bilgisine oranlayarak bir çeşit normalizasyona tabi tutar. Bu terim nitelik değerinin veriyi nasıl böldüğü konusunda hassastır.

$$KazanımOranı(A) = BilgiKazancı(A) / BölünmeBilgisi(S) \quad (1.24)$$

(1.23) ve (1.24) numaralı formüller kullanılarak en yüksek kazanım oranına sahip özellikler belirlenmiş olur.

1.9.2. Sarmal Modeli Öznitelik Seçme Algoritmaları

Sürekli bir arama işlemi gerçekleştiren ve bu arama işlemini gerçekleştirirken bir sınıflandırıcı algoritmasına bağlı olmasından dolayı bu yapılar sarmal yöntemler olarak ifade edilmişlerdir. Sarmalayıcı yöntemler olarak bilinen özellik seçim metotları sınıflandırıcılar üzerinde denemeler yaparak özelliklerin en iyi alt kümesini elde etmeye çalışır.

Alt dizilerin araştırılmasında adım adım ileriye doğru seçim, adım adım geriye doğru eliminasyon gibi deterministik veya benzetilmiş tavlama, genetik algoritmalar gibi sezgisel yöntemler kullanılır. Adım adım ileriye doğru seçimde boş bir öznitelikler dizisi ile başlanır ve her aşamada en iyi olduğu öngörülen öznitelik dizisi ilave edilir. Adım adım geriye doğru eliminasyon yaklaşımında ise bütün özniteliklerden meydana gelen dizi ile başlanarak, bu defa en kötü olduğu öngörülen öznitelik her aşamada diziden uzaklaştırılır.

Her alt dizi bir öğrenme algoritması ile eğitilerek başarı düzeyleri belirlenir ve en başarılı bulunan alt dizi seçilir.

Özellik seçimi yöntemlerinin kötü yanı ise seçim işleminin belirli bir sınıflandırıcı üzerinde denenerek veya belirli bir dizme yöntemi kullanılarak yapılmasıdır. Bu yüzden seçim yapılan sınıflandırıcı ile elde edilen özellik alt kümesi, başka sınıflandırıcılar için uygun olmayabilir ve performansı arttırmayabilir. Aynı şekilde uygulanan filtre sonucu elde edilen azaltılmış özellikler, belirli kategorizasyon işlemlerinin performansına kötü etki edebilir.

1.10. Boyut İndirgeme

Bu tez çalışmasında, veri madenciliğinde veri boyutunu azaltmak için kullanılan klasik yöntemlerden biri olan temel bileşenler analizi (TBA) yöntemi kullanılmıştır.

1.10.1. Temel Bileşenler Analizi

TBA, orijinal değişkenleri, temel bileşen adı verilen ve daha az sayıda olan ilişkili değişkene dönüştürür. Karhunen Loeve yöntemi olarak da bilinen TBA bir değişkenler kümesinin varyans-kovaryans yapısını, bu değişkenlerin doğrusal birleşimleri vasıtasıyla açıklayarak, boyut indirgenmesi ve yorumlanmasını sağlayan, çok değişkenli bir istatistik yöntemidir [68].

Bu yöntemde p adet değişken; doğrusal, ortogonal ve birbirinden bağımsız olma özelliklerini taşıyan k ($k \leq p$) tane yeni değişkene dönüştürülmektedir. Dönüştürülen k adet değişken, p adet değişkenin açıkladığının tümünü açıklayabilmekte başka bir deyişle p adet değişkeni temsil edebilmektedir. Bu indirgemedede önemli varyans kaybı olmamaktadır. Aslında yeni oluşturulan bu k adet değişken gerçek değişkenlerin doğrusal birleşimidir.

TBA veri kümesinin varyansını en iyi yansıtan ve bilginin önemli yönlerini gösteren özellikler olan alt düzey temel bileşenlerin karakteristiğini koruyarak özellik çıkarımı yapar. Bu da özellik kümesinin yeni bir hiper uzaya iz düşümünün özdeğerleri ve özvektörleri kullanılarak alınması ile gerçekleşir. Birinci temel bileşen en yüksek varyansa sahip özelliklerin doğrusal kombinasyonundan oluşur. Bu aynı zamanda en yüksek

özdeğere sahip olan özvektördür. İkinci temel bileşenin varyansı daha düşüktür ve birinci temel bileşen ile dik açı yapar. Esasen orijinal özellik sayısı kadar temel bileşen mevcuttur ve bunlar en yüksek özdeğerliden en düşük özdeğerliye doğru sıralanırlar.

TBA belirtildiği gibi, boyutların büyüklüğünü azaltarak ya da iki boyutu birleştirip tek bir boyut yaparak verinin bir bakıma sıkıştırılmasını da sağlar. Bunun için yapılacak işlemler aşağıdaki gibidir:

- Veriler gerekli durumlarda ortalamayla düzgünleştirilir.
- Kovaryans matrisi hesaplanır.
- Özdeğer ve özvektörler hesaplanır.
- İndirgeme için özellik vektörü seçilir ve indirgeme çarpımı yapılır.

Düzgünleştirme işlemi her zaman yapılması zorunlu bir işlem değildir. (1.25) numaralı ifade ile veri kümesinin ortalaması bulunur.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1.25)$$

Daha sonra ortalama her bir veriden çıkarılarak (1.26) ifadesindeki yeni düzgünleştirilmiş küme elde edilir.

$$x_i = x_i - \bar{x} \quad (1.26)$$

Temel bileşenler yönteminde işlemlere artık bu yeni düzgünleştirilmiş verilerle devam edilir. Standart sapma, verilerin nasıl yayıldığına (saçıldığına) dair ölçümsel olarak bilgi verir ve (1.27) numaralı formülle hesaplanır. Veri değerlerinin yayılımının özetlenmesi için kullanılan bir ölçüdür.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (1.27)$$

Varyans verinin yayılımı ile ilgili bir başka ölçüm bilgisi veren kavramdır. Genellikle değişimi ölçmek için kullanılır. Varyans, standart sapmanın karesidir ve (1.28) numaralı formülle hesaplanır.

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (1.28)$$

Kovaryans, iki rasgele değişkenin beraber değişimlerini inceleyen bir istatistiktir ve (1.29) numaralı formülle hesaplanabilir.

$$\text{cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1} \quad (1.29)$$

Kovaryans değeri, pozitif ise her iki değişkenin birlikte arttığı; negatif ise biri artarken diğerrinin azaldığı; sıfır ise bu iki değişkenin bağımsız olduğu yorumu yapılabilir. İki den fazla değişkene bakıldığında kovaryans matrisi kullanılır. Kovaryans matrisindeki diyagonal değerler değişkenlerin varyans değerlerine eşittir. Kovaryans matris $\text{cov}(a, b) = \text{cov}(b, a)$ özelliğinden dolayı simetrik bir yapı sergilemektedir ve (1.30) ile ifade edilebilir.

$$C = \begin{bmatrix} \text{cov}(x, x) & \text{cov}(x, y) & \text{cov}(x, z) \\ \text{cov}(y, x) & \text{cov}(y, y) & \text{cov}(y, z) \\ \text{cov}(z, x) & \text{cov}(z, y) & \text{cov}(z, z) \end{bmatrix} \quad (1.30)$$

TBA algoritması N-boyutlu m nesneli bir veri kümesi $\{x_1, \dots, x_m\}$ için, kovaryans matrisinin $(\sum)$, (1.31) nolu ifade ile hesaplanması ile başlar:

$$\sum = \frac{1}{m} \sum_{i=1}^m (x_i - \bar{x})(x_i - \bar{x})^T \quad (1.31)$$

Hesaplanan kovaryans matrisine ait özdeğerler $(\lambda_1, \lambda_2, \dots, \lambda_N)$, (1.32) numaralı ifade ile hesaplanır. Bu özdeğerlere karşılık gelen özvektörler $(u_1, u_2, \dots, u_N)$ aranan temel bileşenlerdir.

$$|C - \lambda I| = 0 \quad (1.32)$$

Bileşenlerin varyansları özdeğerlere eşittir ve bu özdeğerlerin büyükten küçüğe doğru sıralanmasıyla birlikte aslında bileşenlerin varyansları sıralanmış olmaktadır. En büyük değere sahip K tane özdeğer seçilir ve U dönüşüm matrisi K tane özvektör ile oluşturulur:

$$\begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_K \end{bmatrix} = \begin{bmatrix} u_1^T \\ u_2^T \\ \dots \\ u_K^T \end{bmatrix} (x - \bar{x}) = U^T (x - \bar{x}) \quad (1.33)$$

K sayısı PoV (Proportion of Variance) kriterine göre seçilir. PoV değerinin belirli bir eşik değeri (0.9, 0.95 gibi) olan T'den büyük olmasına göre belirlenir.

$$PoV = \frac{\sum_{i=1}^K \lambda_i}{\sum_{i=1}^N \lambda_i} > T \quad (1.34)$$

Bu kriter gereği veri bilgisinin %90 veya %95'i korunmuş olur.

1.11. Örneklem Yöntemleri

Örneklem yöntemleri, eğitim verilerindeki sınıf dağılımını ya azınlık örneklerini tekrarlayarak ya da yeni azınlık örnekleri üretmek (aşırı örneklem) ya da çoğunluk sınıfından örnekleri kaldırarak (alt örneklem) dengelemeyi amaçlamaktadır. Aşırı örneklem ve alt örneklem için çeşitli teknikler önerilmiştir.

1.11.1. Azınlıktan Yapay Örnek Üretme Tekniği

Chawla (2012) ve arkadaşları tarafından önerilen Azınlıktan Yapay Örnek Üretme Tekniği (Synthetic Minority Oversampling Technique, SMOTE), en fazla bilinen ve en sık

kullanılan yeniden örnekleme yöntemlerinden biridir [69]. SMOTE, var olan azınlık örneklerinden yola çıkarak ve interpolasyon yaparak yeni yapay azınlık örnekleri oluşturur. “Yapay” örnekler oluşturma yaklaşımı, el yazısı karakterlerinin tanınmasına yönelik geliştirilen bir yöntemden ilham alınarak geliştirilmiştir.

Fazla eğitim verisi, gerçek verinin rotasyonu veya çarpıtılmasıyla oluşturulur. Bu yöntem önce her bir azınlık örneğinin en yakın k tane komşusunu bulur; sonra en yakın komşularından rastgele bir tanesini seçer. Sonra azınlık sınıf örneği ile en yakın komşusunu birbirine bağlayan çizginin üzerinde yeni bir azınlık sınıf örneği oluşturur. Bu prosedür her iki sınıfın da eşit sayıda elemanı olana kadar tekrar eder. SMOTE algoritma adımları Şekil 1.13’te gösterilmektedir.

Algoritma SMOTE(T, N, k)

Giriş: Azınlık sınıfı örnekleri sayısı T ; SMOTE $N\%$ miktarı; En yakın komşu sayısı k

Çıkış: $(N/100) * T$ sentetik azınlık sınıf örnekleri

1. (* Eğer $N, 100\%$ 'den az ise, azınlık sınıf örneklerini onların yalnızca rastgele yüzdelerini SMOTE olacak şekilde rastgele hale getirin *)
2. **if** $N < 100$
3. **then** T azınlık sınıf örneklerini rastgele hale getirin
4. $T = (N/100) * T$
5. $N = 100$
6. **endif**
7. $N = (int)(N/100) (* SMOTE miktarı, 100\%$ 'ün (veya 100lük) tümleşik (integral) çarpımları olarak varsayılacaktır. *)
8. $k =$ En yakın komşuların sayısı
9. $numattr's =$ Özellik (nitelik) Sayısı
10. $Sample[][]$: Orijinal (özgün) azınlık sınıf örnekleri sıralaması
11. $newindex$: Üretilen sentetik örneklerin sayısı sayılarak kayıt, 0'dan başlatılır.
12. $Synthetic[][]$: Sentetik örneklerin sıralaması
13. (* Sadece her azınlık sınıf örnekleri için k en yakın komşularını hesaplayın. *)
13. **for** $i \leftarrow 1$ **to** T
14. i için k en yakın komşularını hesaplayın ve indisleri $nmarray$ içine kaydedin.
15. $Populate(N, i, nmarray)$
16. **endfor**

$Populate(N, i, nmarray)$ (* Sentetik örnekleri üreten fonksiyon. *)

17. **while** $N \neq 0$
 18. 1 ve k arasında rastgele bir sayı seçin, nm olarak işaretleyin. Bu adım, i 'nin k en yakın komşularından birini seçer.
 19. **for** $attr \leftarrow 1$ **to** $numattr's$
 20. Hesapla: $dif = Sample[nmarray[nm]][attr] - Sample[i][attr]$
 21. Hesapla: $gap = 0$ ve 1 arasında rastgele sayı.
 22. $Synthetic[newindex][attr] = Sample[i][attr] + gap * dif$
 23. **endfor**
 24. $newindex++$
 25. $N = N - 1$
 26. **endwhile**
 27. **return** (* Yerleştirmenin Sonu *)
- Pseudo-Code'un Sonu.

Şekil 1.13. SMOTE algoritma adımları

İki boyutlu uzayda (6, 4) orijinal örnek ve onun en yakın komşusu (4, 3) olsun. (6, 4) k-en yakın komşuları tanımlanabilen bir örnektir, (4, 3) onun k-en yakın komşularından birisidir.

$$\begin{aligned} f_{11} &= 6 & f_{21} &= 4 & f_{21} - f_{11} &= 4 - 6 = -2 \\ f_{12} &= 4 & f_{22} &= 3 & f_{22} - f_{12} &= 3 - 4 = -1 \end{aligned}$$

Yeni yapay azınlık örnek:

$$(f'_1, f'_2) = (6, 4) + \text{rand}(0-1) * (-2, -1) \text{ ile üretilir.}$$

1.11.2. Rastgele Alt Örnekleme

Rastgele alt örnekleme, sınıf dağılımlarını dengelemek için fazla sayıdaki sınıfın örneklerini yok etme amacıyla kullanılan, sezgisel olmayan bir yöntemdir [70]. Bu yöntemin en büyük dezavantajı, kullanışlı veya önemli örnekleri yok etme potansiyelidir.

1.12. Sınıflandırma Performans Ölçütleri

Performans ölçekleri değerlendirme için karışıklık matrisinde tanımlı değerleri kullanır. Bu değerler, verilen bir veri kümesinde gerçekte ait olduğu kategori ve tahmin sonucu öngörülen kategorisi uyuşan ve uyuşmayan örneklerin sayısını göstermektedir [71].

Bir veri kümesinde gerçekte pozitif sınıfta olup, sınıflandırmada pozitif sınıfta yer alan bir örneğe doğru pozitif (DP) adı verilir. Örnek, gerçekte negatif sınıfa ait olup sınıflandırma sonucunda negatif sınıfta yer alırsa doğru negatif (DN) adını alır. Gerçekte negatif sınıfta olmasına rağmen pozitif sınıfta tahmin edilen örneğe yanlış pozitif (YP) denir. Pozitif sınıfta olmasına rağmen negatif sınıfta tahmin edilen örnekler ise yanlış negatif (YN) olarak isimlendirilmektedir. Tablo 1.1'de örneğin doğru ya da yanlış sınıfta olması ile ilgili tüm durumlar verilmiştir. Karışıklık matrisindeki değerlerden yararlanarak geliştirilen performans ölçekleri alt bölümlerde açıklanacaktır.

Tablo 1.1. Karışıklık matrisi

		Öngörülen Sınıf	
		Pozitif	Negatif
Gerçek Sınıf	Pozitif	Doğru Pozitif	Yanlış Negatif
	Negatif	Yanlış Pozitif	Doğru Negatif

1.12.1. Duyarlılık

Kategorizasyon işlemleri için duyarlılık (precision) ile bir örneğin doğru kategoride olması durumu ölçülür. Duyarlılık, doğru sınıflandırılmış pozitif örnek sayısının pozitif sınıflandırılmışların sayısına oranı olarak (1.35) bağıntısıyla elde edilir.

$$Duyarlılık = \frac{DP}{DP + YP} \quad (1.35)$$

1.12.2. Anma

Kategorizasyonun hassasiyetini ölçmek için kullanılan ölçeğe anma adı verilir. Doğru pozitiflerin oranı olarak da bilinir ve (1.36) bağıntısıyla elde edilir.

$$Anma = \frac{DP}{DP + YN} \quad (1.36)$$

Tek başına anma ölçeği kullanmak performansı ölçmek için yeterli değildir. Örneğin veri kümesindeki tüm örnekler bir sınıftaymış gibi bir sonuç elde edilirse, anma değeri %100 olur. Bu sebeple tek başına duyarlılık ya da anma yerine ikisinin bileşkesi yöntemleri kullanmak daha sağlıklı ölçüm sonuçları ortaya çıkarmaktadır.

1.12.3. F-Ölçeği

Duyarlılık ve anma ölçeklerinin ağırlıklı harmonik ortalamasına F-Ölçeği adı verilir. Duyarlılık ve anma eşit ağırlıkta olduğu için bu ölçek F_1 ölçeği olarak da bilinir. F_1 ölçeği (1.37)'de verilmiştir.

$$F_1 = 2 \times \frac{\text{duyarlılık} \times \text{anma}}{\text{duyarlılık} + \text{anma}} = \frac{2 \times DP}{2 \times DP + YP + YN} \quad (1.37)$$

Genel olarak kategorizasyon performansının ölçülmesinde (1.37)'de verilmiş olan F_1 ölçeği kullanılır. Kategoriler arası dağılımı homojen olmayan ya da belirli bir kategoride çok az sayıda örnek barındıran veri kümelerindeki kategorizasyon işlemlerinde başarıyı ölçmek için de bu ölçeğin kullanımı daha uygundur.

1.12.4. Alıcı İşletim Karakteristiği

DP örnek sayısının YP örnek sayısına olan oranı hassasiyet olarak bilinir. Eğer bu oran grafik eğri olarak çizilirse, bu eğri alıcı işletim karakteristiği (receiver operating characteristic, ROC) adını alır. ROC eğrisi aynı zamanda DP oranının YP oranına bölümünden elde edilen değerler ile de çizdirilebilir. Burada DP oranı, DP'lerin tüm gerçek pozitiflere (DP+YN) oranı, YP oranı ise YP'lerin tüm gerçek negatiflere (DN+YP) oranı ile hesaplanır.

ROC eğrisi altında kalan alanın büyük olması modelin kategorizasyonda başarılı olduğunu gösterdiğinden, genellikle modellerin performanslarının karşılaştırılmasında kullanılır

2. YAPILAN ÇALIŞMALAR, BULGULAR VE İRDELEME

2.1. Giriş

Çalışmada kullanılan veri seti, UCI Makine Öğrenmesi veri tabanından alınmıştır [72]. Bu veri seti, ilgili adresten ulaşılabilen, araştırma amacıyla herkese açık bir şarap kalitesi veri setidir.

Kırmızı ve beyaz şaraplar kullanılarak iki veri seti hazırlanmıştır. İki veri seti de Portekiz'in Minho bölgesinden gelen "Vinho Verde" şarabının kırmızı ve beyaz çeşitlerine bağlıdır. Veriler Mayıs 2004'ten Şubat 2007'ye kadar, sadece korunmuş ve resmi sertifikasyon kurumu tarafından test edilmiş kök örneklerden toplanmıştır [73]. Veri setleri, kırmızı şarap için 11 özellik ve 1599 örnek, beyaz şarap için aynı 11 özellik ve 4898 örnek içermektedir. Girdiler objektif testler (pH değerleri gibi) içermekte ve çıktılar duyuşal verilere (uzmanlar tarafından yapılan en az 3 değerlendirmenin medyanı alınmıştır) dayanmaktadır. Her uzman şarap kalitesini 0 (çok kötü) ve 10 (son derece mükemmel) arasında değerlendirmiştir.

Veriler, şarap örneği test sürecini üretici talebinden laboratuvara ve oradan da sensör analizine otomatik yöneten bilgisayarlı bir sistemle (iLab) kaydedilmiştir. Her giriş belli bir teste (analitik ve sensör) karşılık gelmektedir ve son veri tabanı tek bir sayfaya (.csv) aktarılmıştır.

Ön işlem aşamasında, veri tabanı, her satırda belli bir şarap örneği (tüm testlerle) olacak şekilde çevrilmiştir. Bu veri tabanında örneklerdeki etkin özelliklerin kaybını engellemek amacıyla, sadece en yaygın fiziko-kimyasal testler seçilmiştir. Ayrıca, kırmızı ve beyaz şaraplar oldukça değişik tatlara sahip oldukları için analizleri ayrıca gerçekleştirilecektir.

2.2. Özniteliklerin Tanıtılması

Özellikler; fiziko-kimyasal testlere dayalı sabit asitlik, uçucu asitlik, sitrik asit, artık şeker, kloridler, serbest kükürt dioksit (SO₂), toplam SO₂, yoğunluk, pH, sülfidler

(potasyum sülfat (K_2SO_4)), alkol değerlerini içeren 11 farklı girdi değişkenlerinden ve duyusal verilere dayalı kalite çıktı değişkeninden oluşmaktadır.

Tablo 2.1 ve Tablo 2.2, sırasıyla kırmızı ve beyaz şarap veri setlerinin, 11 farklı fiziko-kimyasal özelliklerini ve veri istatistiklerini (her bir özellik için tüm örneklerin minimum, maksimum, ortalama ve standart sapma değerlerini) sunmaktadır.

Tablo 2.1. Kırmızı şarap örneklerinin fiziko-kimyasal veri istatistikleri

Özellikler (Birim)	Min	Max	Ort.	Std. S.
Sabit Asitlik (g/dm^3)	4.600	15.90	8.320	1.741
Uçucu Asitlik (g/dm^3)	0.120	1.580	0.528	0.179
Sitrik Asit (g/dm^3)	0.000	1.000	0.271	0.195
Artık Şeker (g/dm^3)	0.900	15.50	2.539	1.410
Kloridler (g/dm^3)	0.012	0.611	0.087	0.047
Serbest SO_2 (mg/dm^3)	1.000	72.00	15.87	10.46
Toplam SO_2 (mg/dm^3)	6.000	289.0	46.47	32.89
Yoğunluk (g/cm^3)	0.990	1.004	0.997	0.002
pH	2.740	4.010	3.311	0.154
Sülfidler (g/dm^3)	0.330	2.000	0.658	0.170
Alkol (%)	8.400	14.90	10.42	1.066

Tablo 2.2. Beyaz şarap örneklerinin fiziko-kimyasal veri istatistikleri

Özellikler (Birim)	Min	Max	Ort.	Std. S.
Sabit Asitlik (g/dm^3)	3.800	14.20	6.855	0.844
Uçucu Asitlik (g/dm^3)	0.080	1.100	0.278	0.101
Sitrik Asit (g/dm^3)	0.000	1.660	0.334	0.121
Artık Şeker (g/dm^3)	0.600	65.80	6.391	5.072
Kloridler (g/dm^3)	0.009	0.346	0.046	0.022
Serbest SO_2 (mg/dm^3)	2.000	289.0	35.31	17.01
Toplam SO_2 (mg/dm^3)	9.000	440.0	138.4	42.50
Yoğunluk (g/cm^3)	0.987	1.039	0.994	0.003
pH	2.720	3.820	3.188	0.151
Sülfidler (g/dm^3)	0.220	1.080	0.490	0.114
Alkol (%)	8.000	14.20	10.51	1.231

Sabit Asitlik: Asitlik şarabın “taze, mayhoş ve ekşi” özelliklerine atıfta bulunmaktadır. Şarapla ilgili çoğu asit sabittir veya uçucu değildir.

Uçucu Asitlik: Sabit asitten daha uçucu veya daha çok buharlaşan organik asitlerdir. Yüksek seviyede olması şarapta hoş olmayan, sirkemsi bir tada sebep olan asetik asittir.

Sitrik Asit: Şarapta bulunan ve “tazelik” ekleyen uçucu olmayan asitlerden biridir. Az miktarda bulunur.

Artık Şeker: Fermantasyon durduktan sonra geride kalan şeker miktarıdır. 1 gram/litre’den az şaraplar bulmak çok nadir olduğu gibi 45 gram/litre’den fazla olan şaraplar tatlı kabul edilir.

Kloridler: Şaraptaki tuz miktarıdır.

Serbest SO₂: Moleküler SO₂ (çözünmüş gaz halinde) ve bisülfid iyonu arasındaki denge durumunda bulunan serbest durumdaki SO₂. Mikrobiyal büyümeyi ve şarabın oksidasyonunu önler.

Toplam SO₂: SO₂’nin serbest ve bağlı formlarının miktarıdır. Az miktarlardaki SO₂ çoğu zaman şarapta saptanamaz ancak serbest SO₂ konsantrasyonunun 50 ppm’den fazla olduğu durumlarda SO₂ şarabın kokusunda ve tadında oldukça belirgindir.

Yoğunluk: Şarabın yoğunluğu, alkol ve şeker içeriği yüzdesine bağlı olarak suyunkine yakındır.

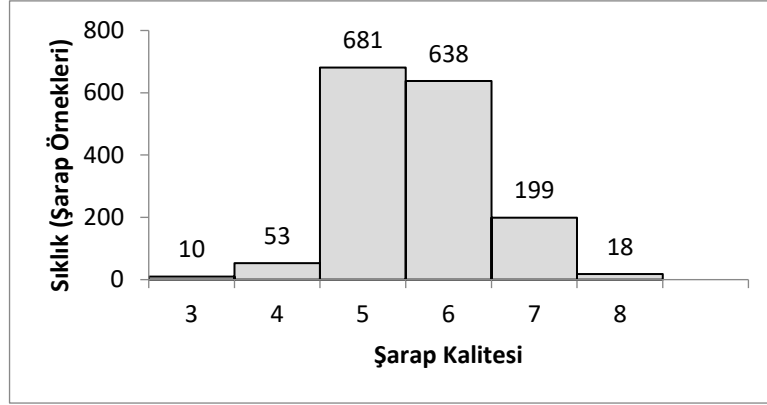
pH: Şarabın 0’dan (çok asidik) 14’e (çok bazik) kadar olan bir aralıkta ne kadar asidik veya bazik olduğunu anlatır. Çoğu şarap pH aralığında 3-4 arasındadır.

Sülfidler: Şarabın içinde bulunan, antimikrobiyal ve antioksidan olarak davranan SO₂ gazı seviyelerine katkıda bulunabilen bir maddedir.

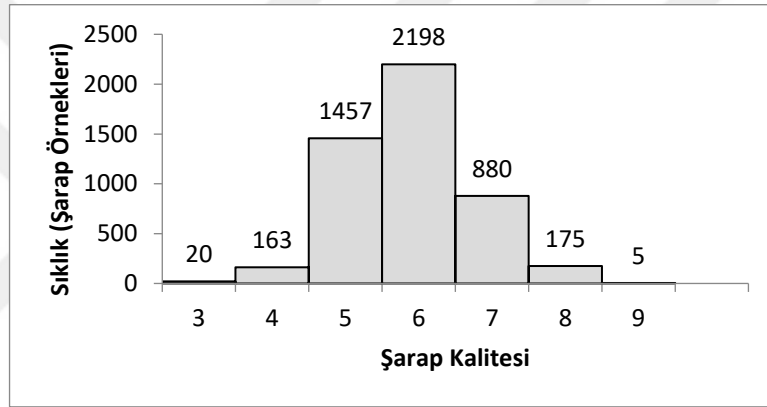
Alkol: Şarabın alkol içeriği yüzdesidir.

Kalite: 0 (çok kötü) ile 10 (son derece mükemmel) arasında ölçeklendirilen en az üç duyusal değerlendiriciyle (kör duyular ile) değerlendirilmiştir. Son duyusal puan bu üç değerlendirmenin medyanı ile belirlenmiştir.

Şekil 2.1 ve Şekil 2.2, sırasıyla kırmızı ve beyaz şarap örneklerinin kalitelerine göre veri setindeki örnek sayısını göstermektedir.



Şekil 2.1. Kırmızı şarap örneklerinin duysal değerlendirmelerinin histogram grafiği

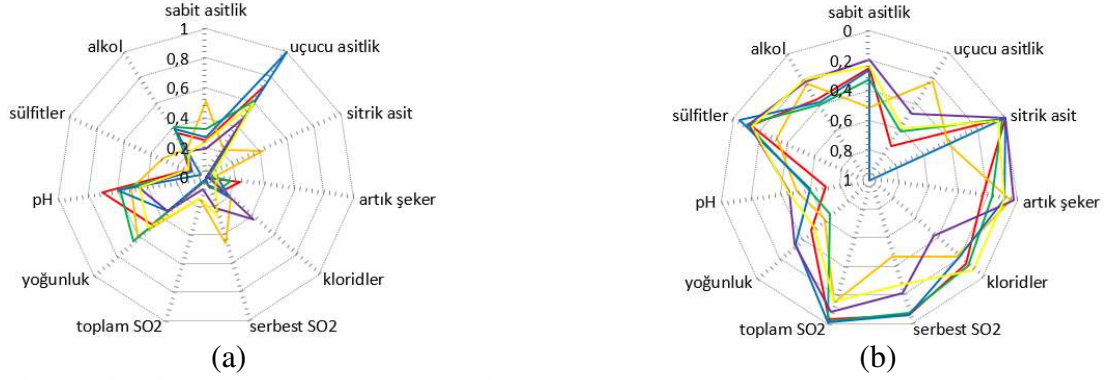


Şekil 2.2. Beyaz şarap örneklerinin duysal değerlendirmelerinin histogram grafiği

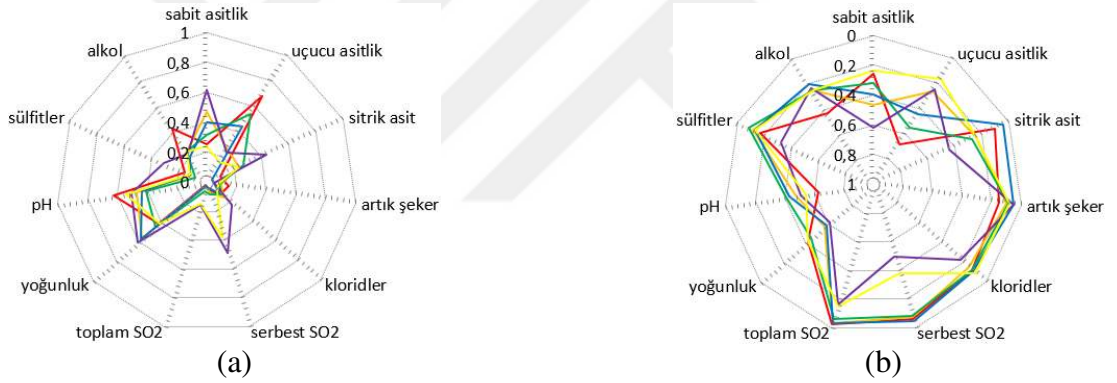
2.2.1. Özniteliklerin Radar Çizimleri

Farklı kalitelardaki şarap örneklerinin örüntüsü için çizilen radar grafikleri Şekil 2.3-2.15 arasında yer almaktadır. Bu radar çizimlerinde şarap ölçümlerinden elde edilen özniteliklerin her biri radar grafikte bir eksene karşılık gelmektedir. Grafikteki farklı renkler ise aynı kalitedeki şarap örneklerinden rastgele seçilmiş 6 tane örneği ifade etmektedir. Bu grafikleri çizebilmek için veri setine min-max normalizasyonu uygulanmıştır. Bu yöntemde, bir grup verinin içerisindeki en büyük ve en küçük değerler ele alınır. Diğer bütün veriler, bu değerlere göre normalleştirilir. Buradaki amaç en küçük değeri 0 ve en büyük değeri 1 olacak şekilde normalleştirmek ve diğer bütün verileri 0-1 aralığına yaymaktır. Örneğin, Şekil 2.3'te farklı renklerdeki radar çizimleri 3. kalite kırmızı şarap için seçilen farklı 6 örneği, radardaki 11 eksen ise 11 öz niteliği

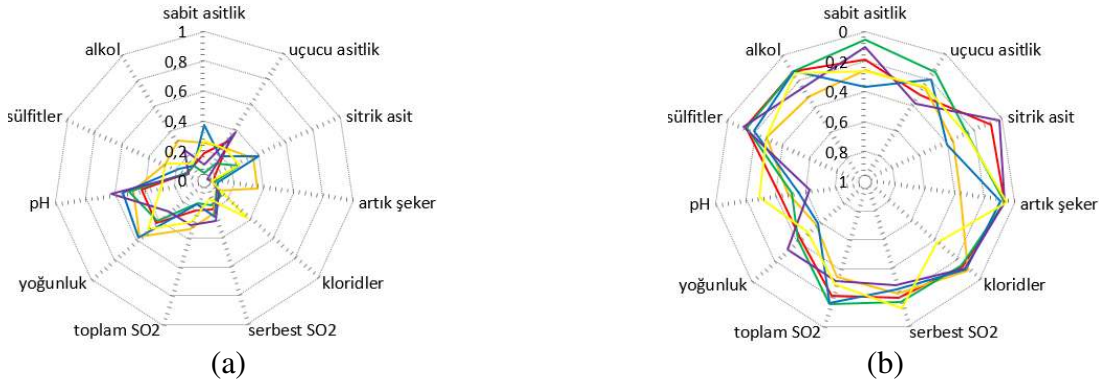
göstermektedir. Böylece bu öz nitelikler ile ortaya çıkan 3. kalite kırmızı şarap örüntüsü elde edilmektedir.



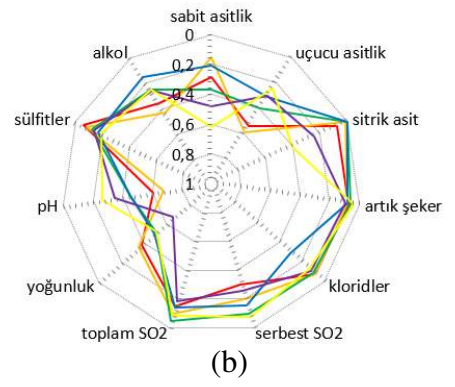
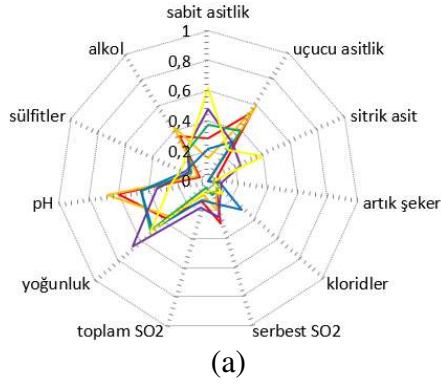
Şekil 2.3. 3. kalite kırmızı şarap için a) normal değerler b) ters sırada değerler ile öz niteliklerin radar çizimi



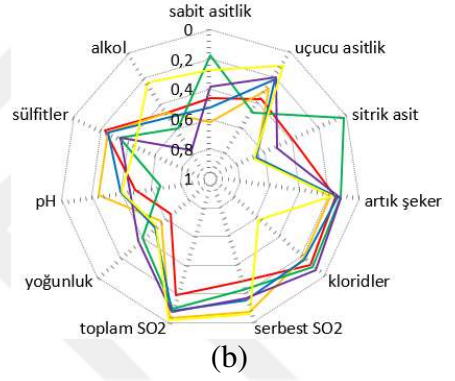
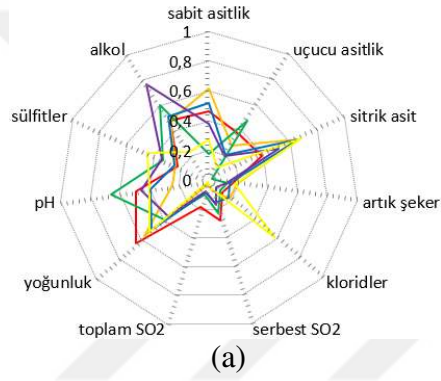
Şekil 2.4. 4. kalite kırmızı şarap için a) normal değerler b) ters sırada değerler ile öz niteliklerin radar çizimi



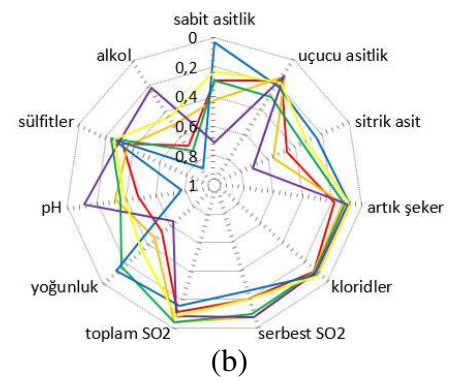
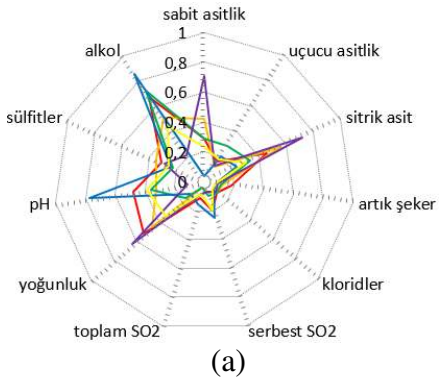
Şekil 2.5. 5. kalite kırmızı şarap için a) normal değerler b) ters sırada değerler ile öz niteliklerin radar çizimi



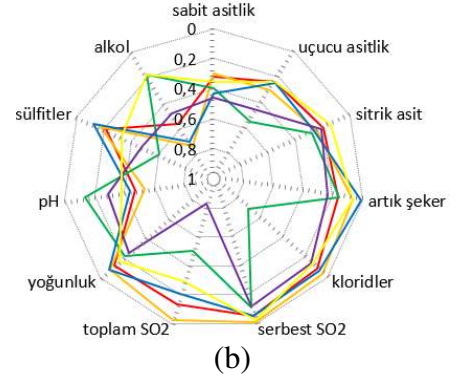
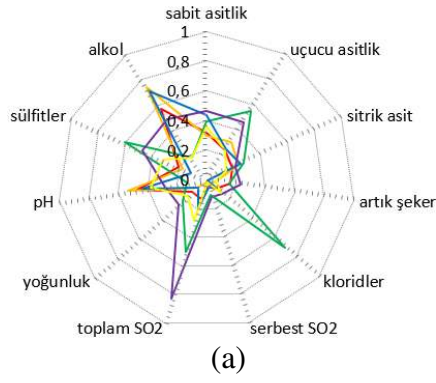
Şekil 2.6. 6. kalite kırmızı şarap için a) normal değerler b) ters sırada değerler ile özniteliklerin radar çizimi



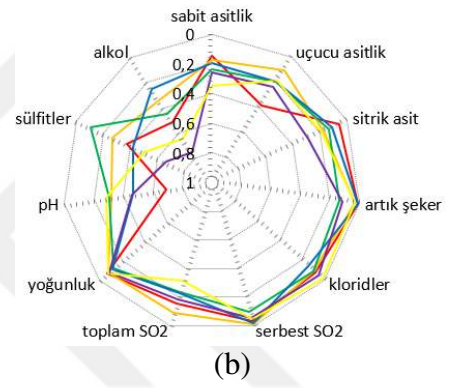
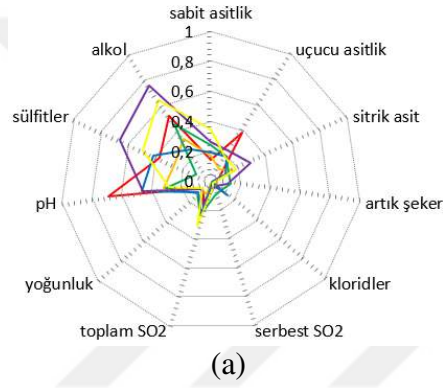
Şekil 2.7. 7. kalite kırmızı şarap için a) normal değerler b) ters sırada değerler ile özniteliklerin radar çizimi



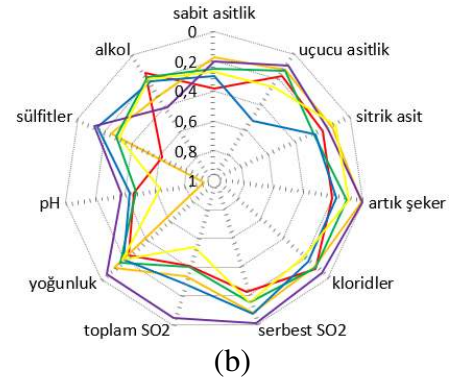
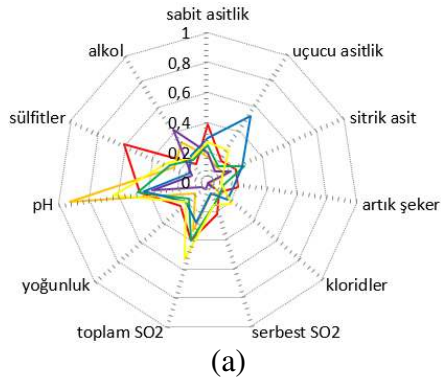
Şekil 2.8. 8. kalite kırmızı şarap için a) normal değerler b) ters sırada değerler ile özniteliklerin radar çizimi



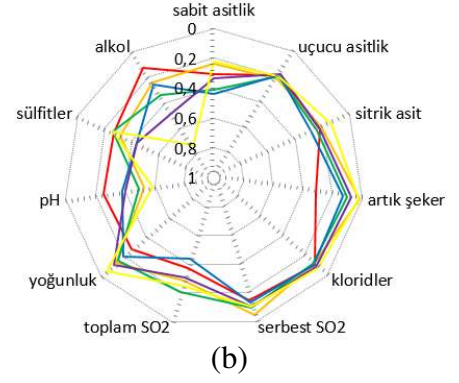
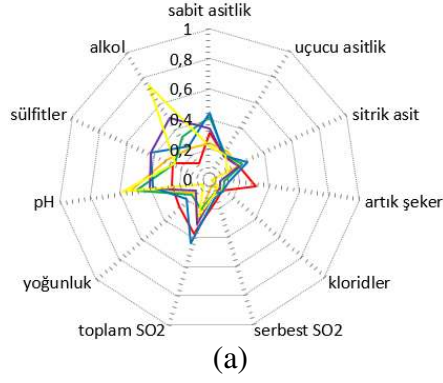
Şekil 2.9. 3. kalite beyaz şarap için a) normal değerler b) ters sırada değerler ile özniteliklerin radar çizimi



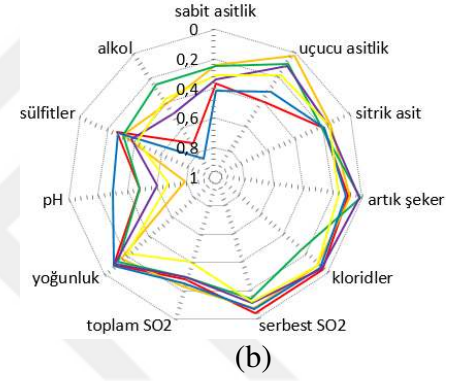
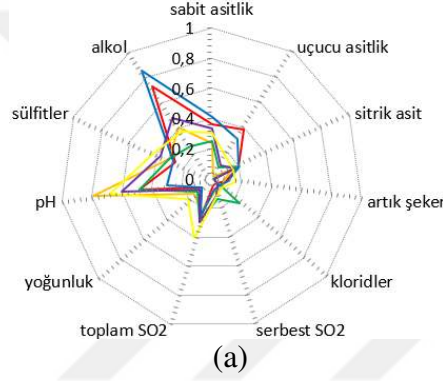
Şekil 2.10. 4. kalite beyaz şarap için a) normal değerler b) ters sırada değerler ile özniteliklerin radar çizimi



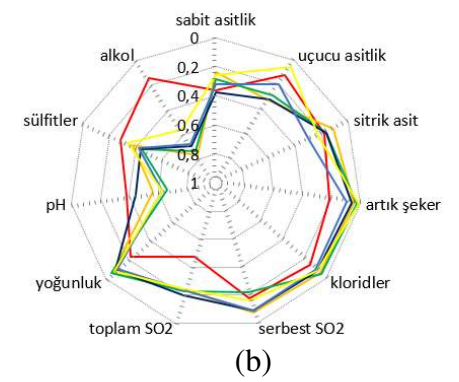
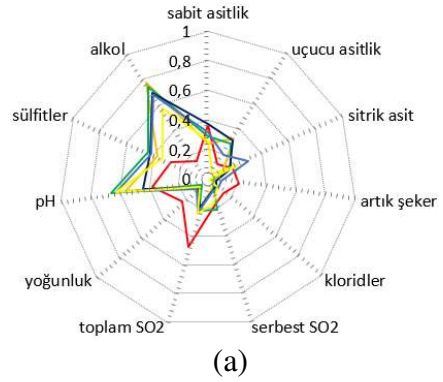
Şekil 2.11. 5. kalite beyaz şarap için a) normal değerler b) ters sırada değerler ile özniteliklerin radar çizimi



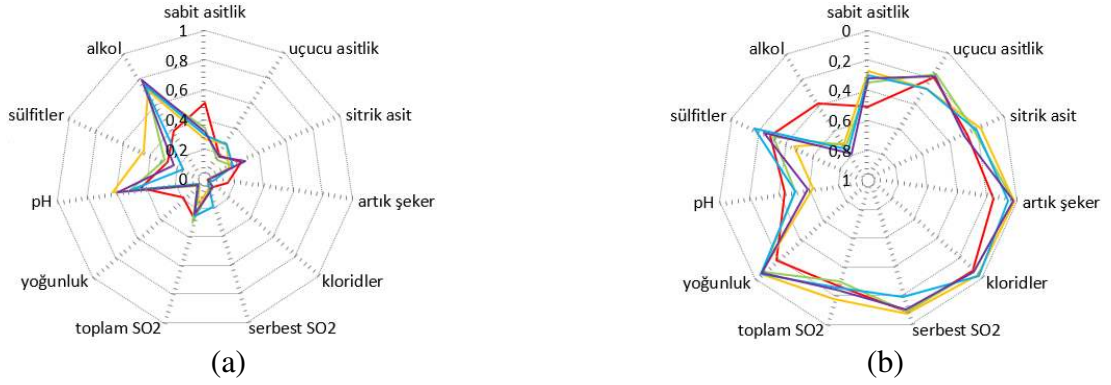
Şekil 2.12. 6. kalite beyaz şarap için a) normal değerler b) ters sırada değerler ile özniteliklerin radar çizimi



Şekil 2.13. 7. kalite beyaz şarap için a) normal değerler b) ters sırada değerler ile özniteliklerin radar çizimi



Şekil 2.14. 8. kalite beyaz şarap için a) normal değerler b) ters sırada değerler ile özniteliklerin radar çizimi

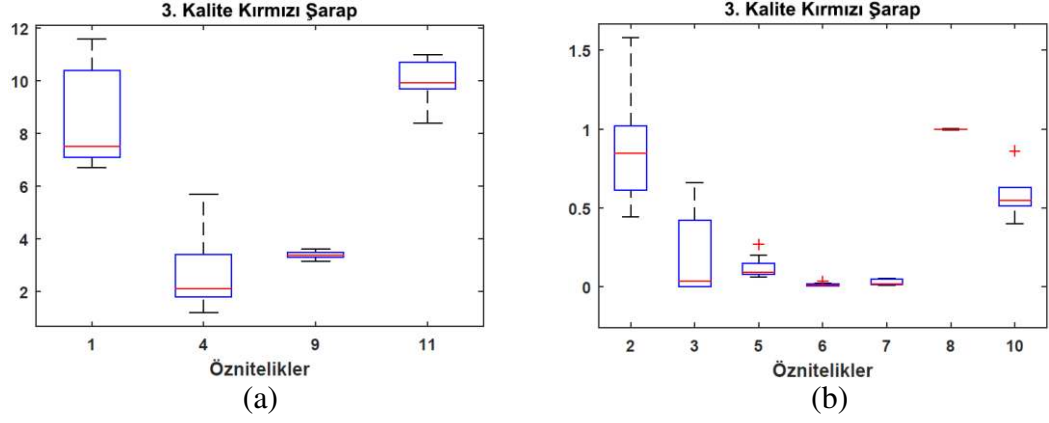


Şekil 2.15. 9. kalite beyaz şarap için a) normal değerler b) ters sırada değerler ile öz niteliklerin radar çizimi

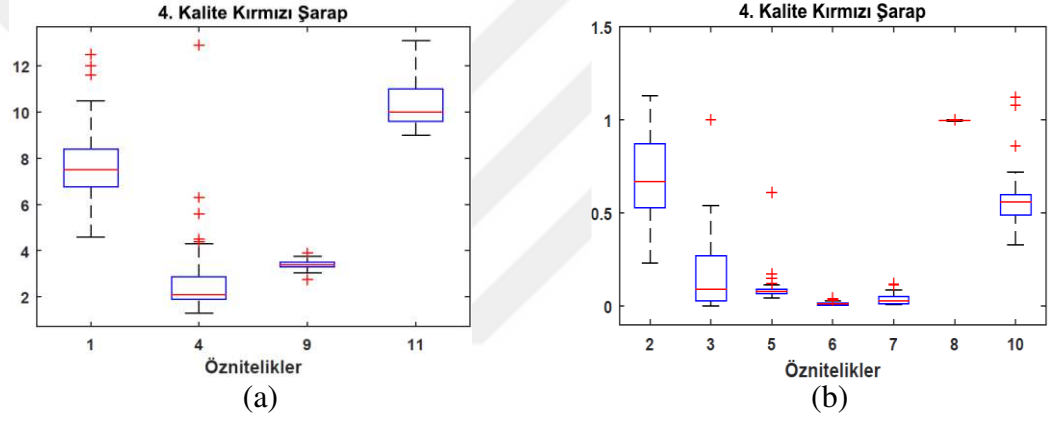
2.2.2. Öz niteliklerin Kutu Bıyık Grafikleri

Şarap kalitesi veri setindeki öz niteliklerin tüm veriler için dağılımını gözlemlemek için kutu-bıyık grafikleri oluşturulmuştur. Kutu-bıyık grafiklerinin yatay ekseninde öz nitelikler varken, dikey ekseninde o öz niteliklerin dağılımı gösterilmektedir. Grafikler oluşturulurken, ortalama, üst ve alt çeyreklikler, en küçük ve en büyük veri değerleri istatistiki ölçümlerinden faydalanılır. Farklı kalitelere göre şarap örneklerinin öz niteliklerinin dağılımını gözlemlemek için çizilen kutu-bıyık grafikleri Şekil 2.16-2.28 arasında yer almaktadır.

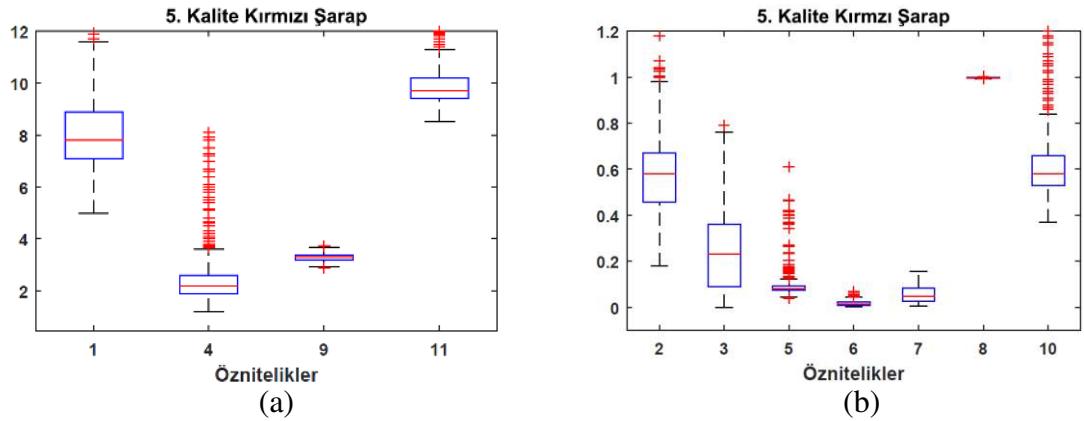
Şekil 2.16-2.28’de çizilen kutu-bıyık grafiklerinin yatay ekseninde yer alan sayılar öz nitelikleri ifade etmektedir. Sabit asitlik, uçucu asitlik, sitrik asit, artık şeker, kloridler, serbest SO₂, toplam SO₂, yoğunluk, pH, sülfidler ve alkol öz nitelikleri sırasıyla 1’den 11’e kadar numaralandırılmıştır. Örneğin, Şekil 2.28 (a)’da yatay eksenindeki 1, 4, 9 ve 11 sayıları sırasıyla sabit asitlik, artık şeker, pH ve alkol öz niteliklerini ifade ederken, Şekil 2.28 (b)’deki 2, 3, 5, 6, 7, 8 ve 10 sayıları ise sırasıyla uçucu asitlik, sitrik asit, kloridler, serbest SO₂, toplam SO₂, yoğunluk ve sülfidler öz niteliklerini temsil etmektedir. Dikey ekseninde ise bu öz niteliklerin değerleri bulunmaktadır. Bu grafikler sayesinde tüm öz niteliklerin dağılımı gözlenebilmektedir.



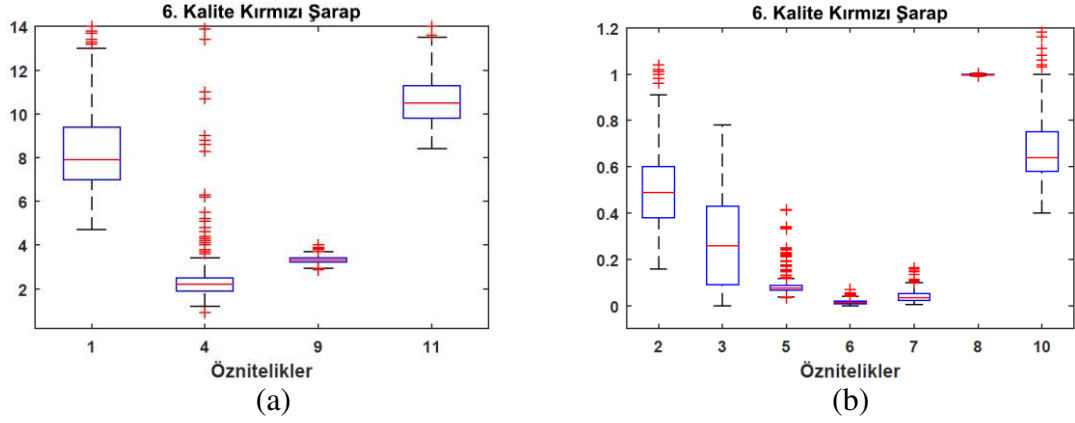
Şekil 2.16. 3. kalite kırmızı şarap örneklerinin a) 4 özniteliğinin b) 7 özniteliğinin tüm veriler için dağılımı



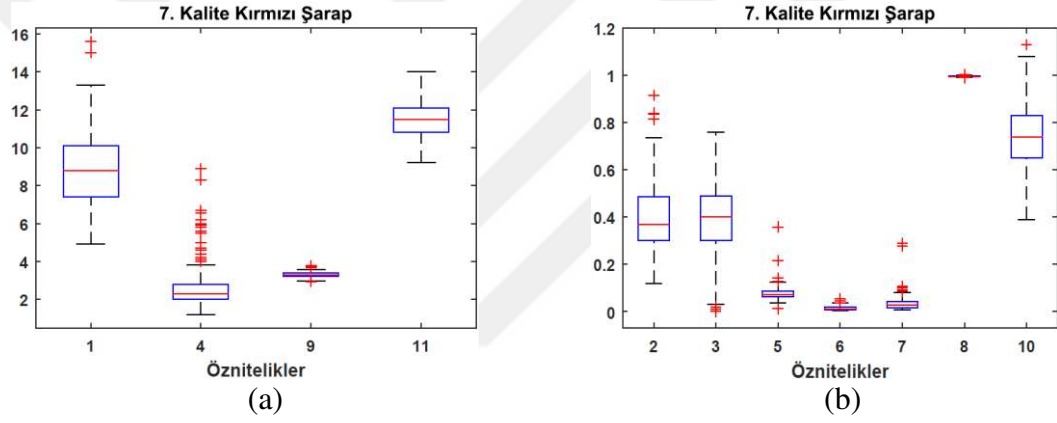
Şekil 2.17. 4. kalite kırmızı şarap örneklerinin a) 4 özniteliğinin b) 7 özniteliğinin tüm veriler için dağılımı



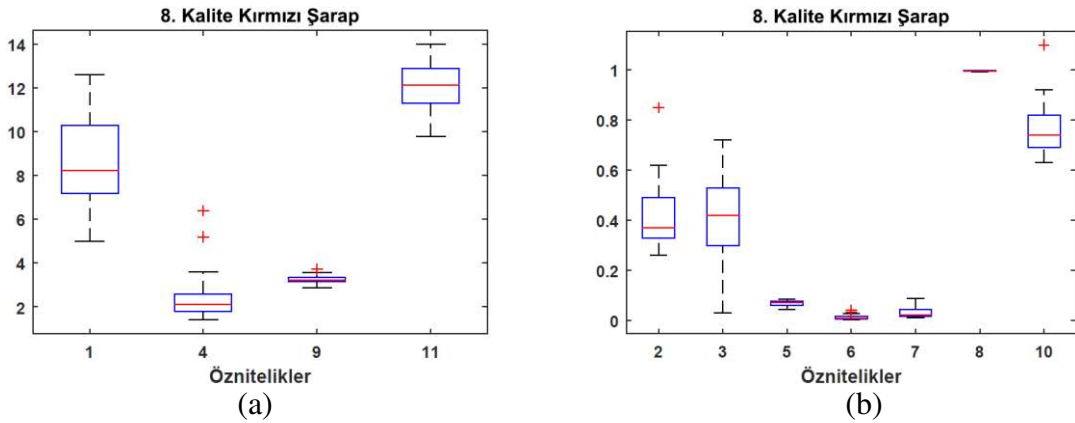
Şekil 2.18. 5. kalite kırmızı şarap örneklerinin a) 4 özniteliğinin b) 7 özniteliğinin tüm veriler için dağılımı



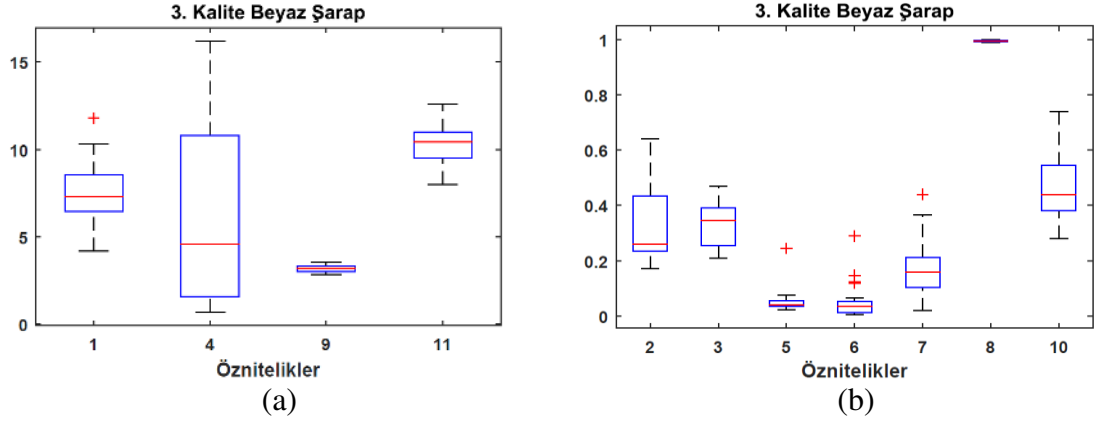
Şekil 2.19. 6. kalite kırmızı şarap örneklerinin a) 4 özniteliğinin b) 7 özniteliğinin tüm veriler için dağılımı



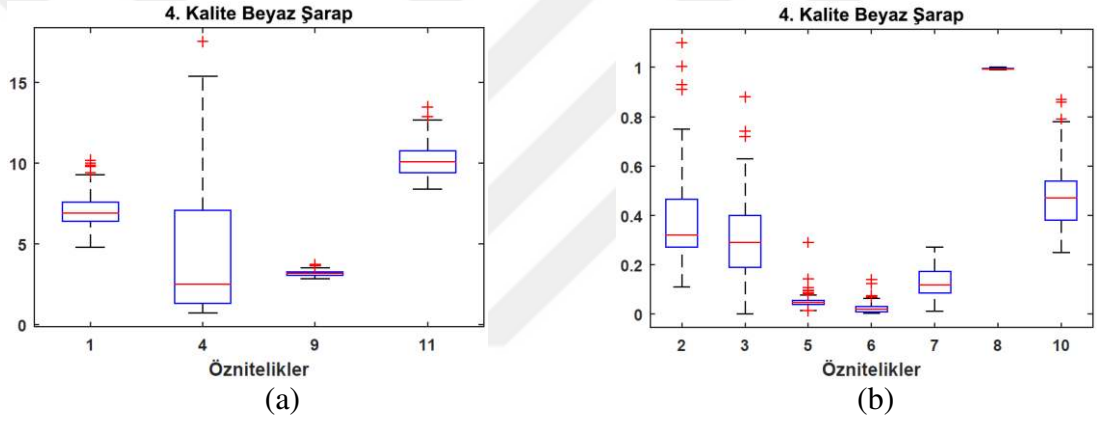
Şekil 2.20. 7. kalite kırmızı şarap örneklerinin a) 4 özniteliğinin b) 7 özniteliğinin tüm veriler için dağılımı



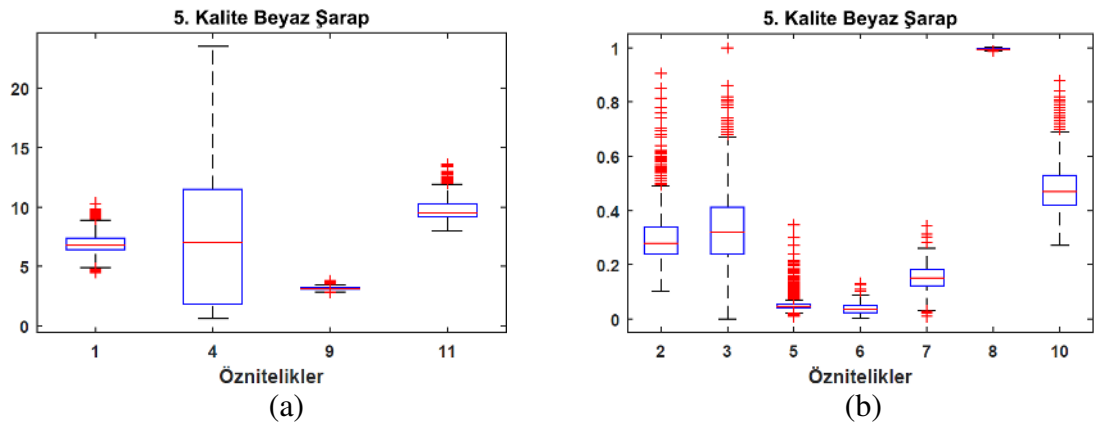
Şekil 2.21. 8. kalite kırmızı şarap örneklerinin a) 4 özniteliğinin b) 7 özniteliğinin tüm veriler için dağılımı



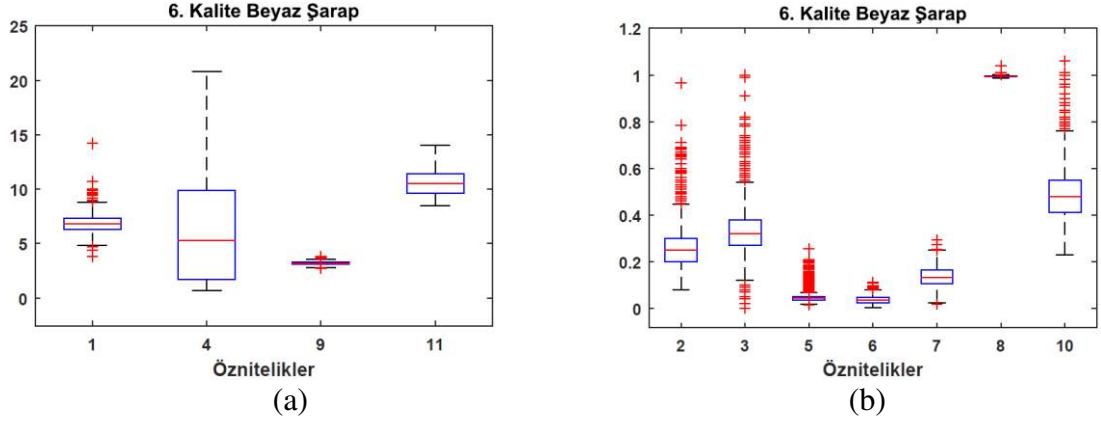
Şekil 2.22. 3. kalite beyaz şarap örneklerinin a) 4 özniteliğinin b) 7 özniteliğinin tüm veriler için dağılımı



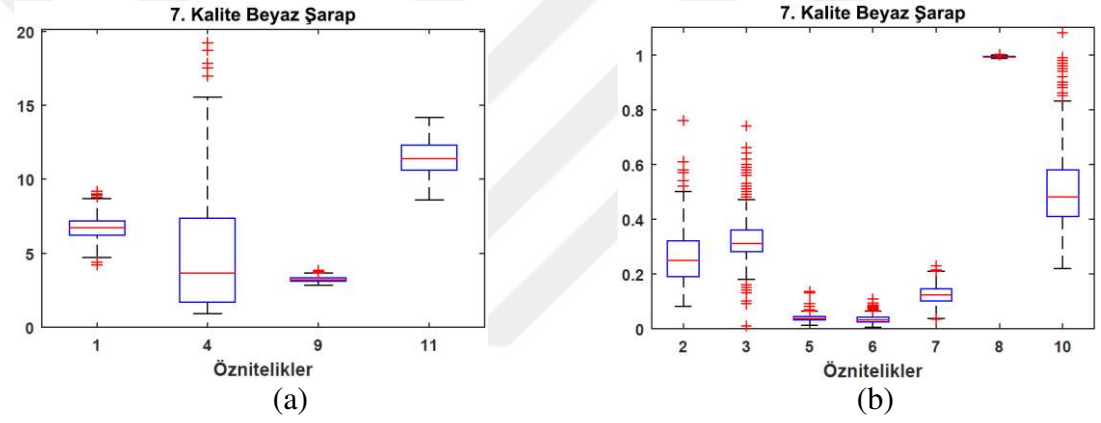
Şekil 2.23. 4. kalite beyaz şarap örneklerinin a) 4 özniteliğinin b) 7 özniteliğinin tüm veriler için dağılımı



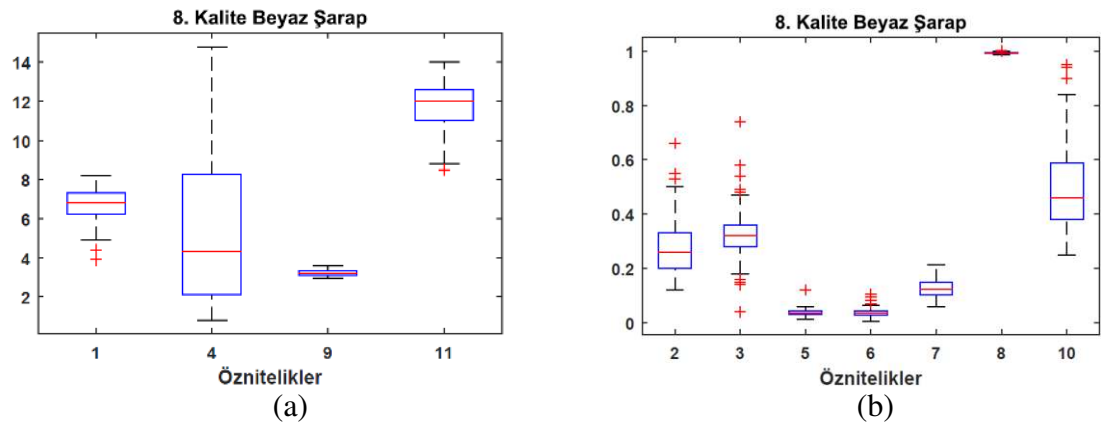
Şekil 2.24. 5. kalite beyaz şarap örneklerinin a) 4 özniteliğinin b) 7 özniteliğinin tüm veriler için dağılımı



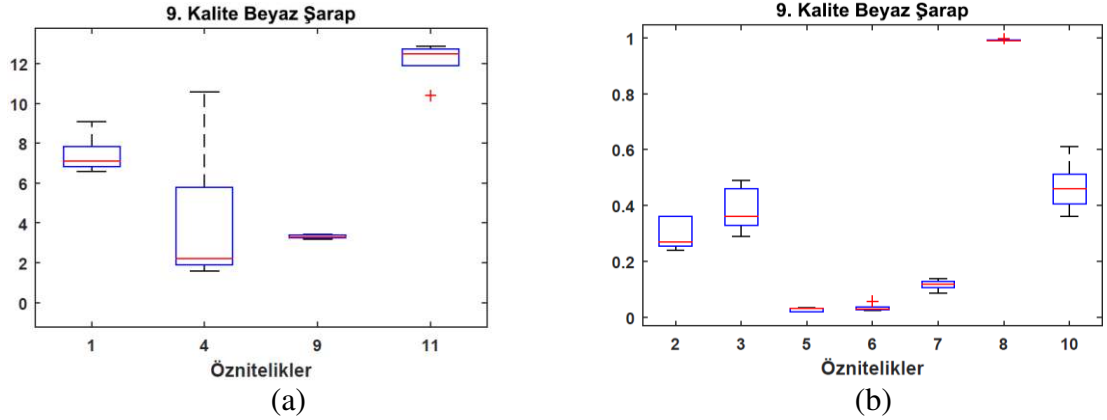
Şekil 2.25. 6. kalite beyaz şarap örneklerinin a) 4 öz niteliğinin b) 7 öz niteliğinin tüm veriler için dağılımı



Şekil 2.26. 7. kalite beyaz şarap örneklerinin a) 4 öz niteliğinin b) 7 öz niteliğinin tüm veriler için dağılımı



Şekil 2.27. 8. kalite beyaz şarap örneklerinin a) 4 öz niteliğinin b) 7 öz niteliğinin tüm veriler için dağılımı



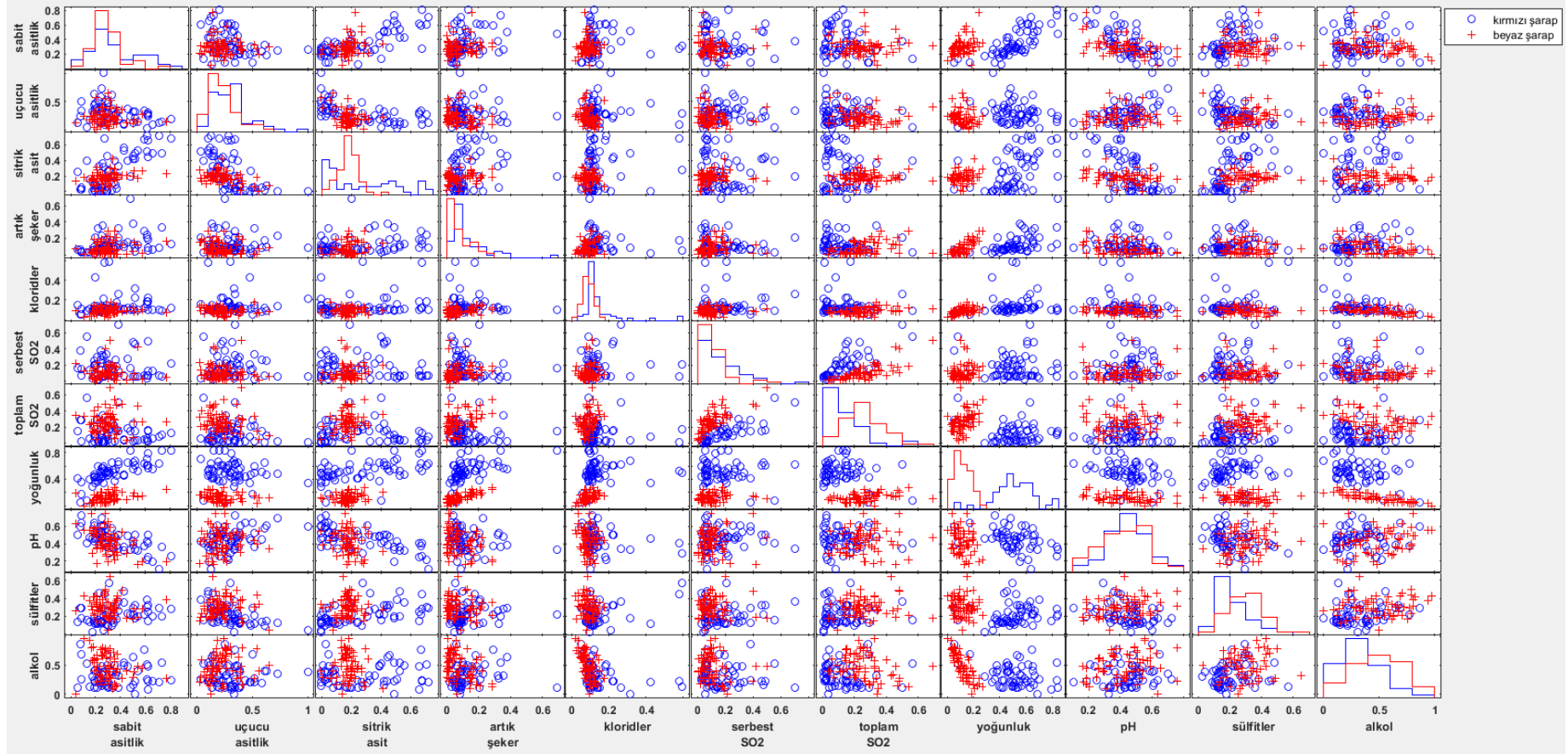
Şekil 2.28. 9. kalite beyaz şarap örneklerinin a) 4 özniteliğinin b) 7 özniteliğinin tüm veriler için dağılımı

2.2.3. Özniteliklerin Saçılım Grafikleri

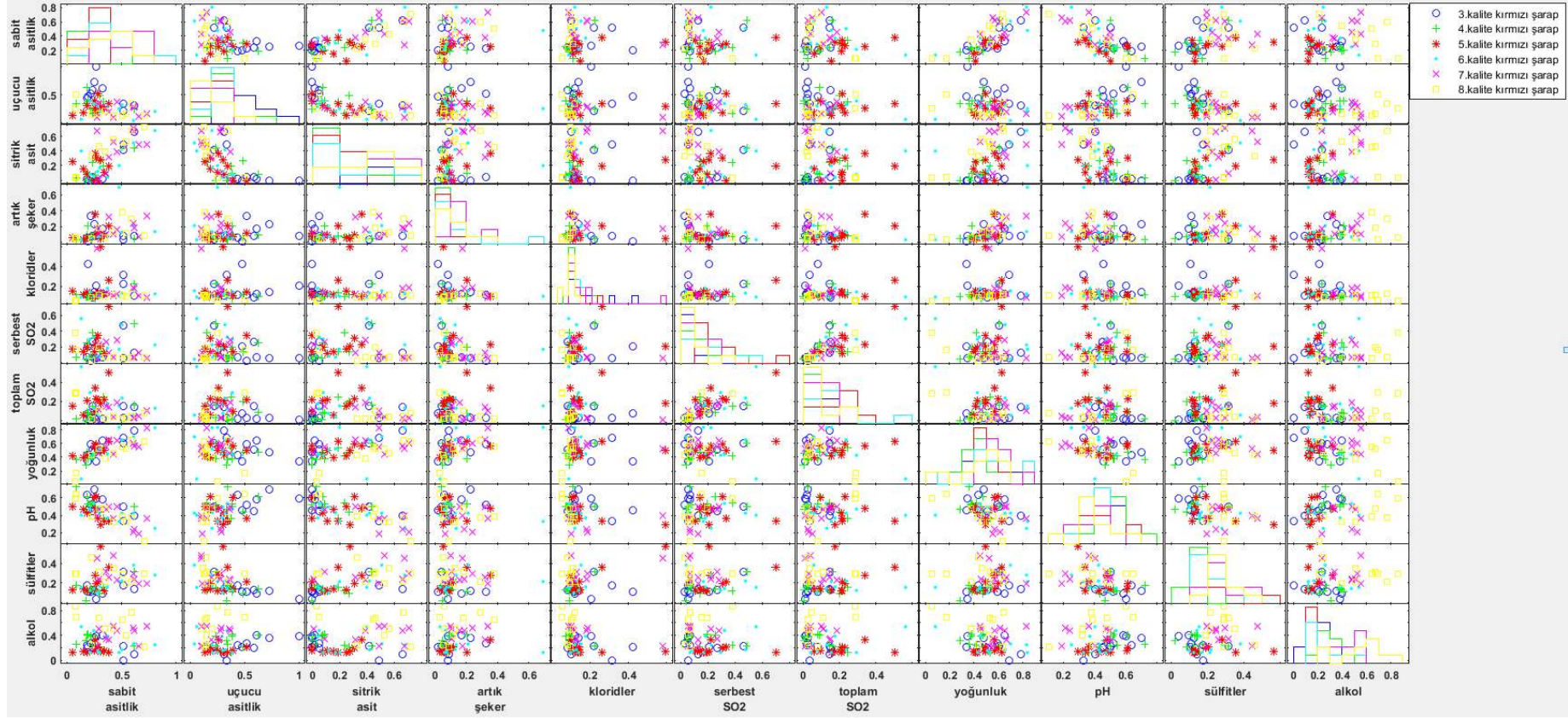
Saçılım grafiği, iki değişken arasındaki ilişkiyi veya birlikteliği ortaya koyar. Saçılım grafiği, x değerlerine karşı tekabül eden y değerlerinin bir grafiğidir. Dikey eksen (y değişkeni) genellikle tepki değişkenidir. Yatay eksen (x değişkeni) genellikle tepkiye bağlı olarak varsaydığımız değişkenlerdir.

Bir dizi $X_1, X_2, \dots, X_k$ değişkenleri kümesinde, saçılım grafiği matrisi, tek sayfa üzerinde değişkenlerin tüm ikili saçılım grafiklerini matris formatında içerir. Yani, k adet değişken varsa, saçılım grafiği matrisi k adet satır ve k adet sütundan oluşacaktır. Bu matrisin i . satırı ve j . sütunu X_j 'ye karşı X_i grafiği olacaktır. Her satır ve her sütun, tekil bir saçılım grafiğini tanımlar. i . satırı ve j . sütunu için olan tekil grafik, düşey eksen X_i ve yatay eksen X_j olacak şekilde tanımlanır.

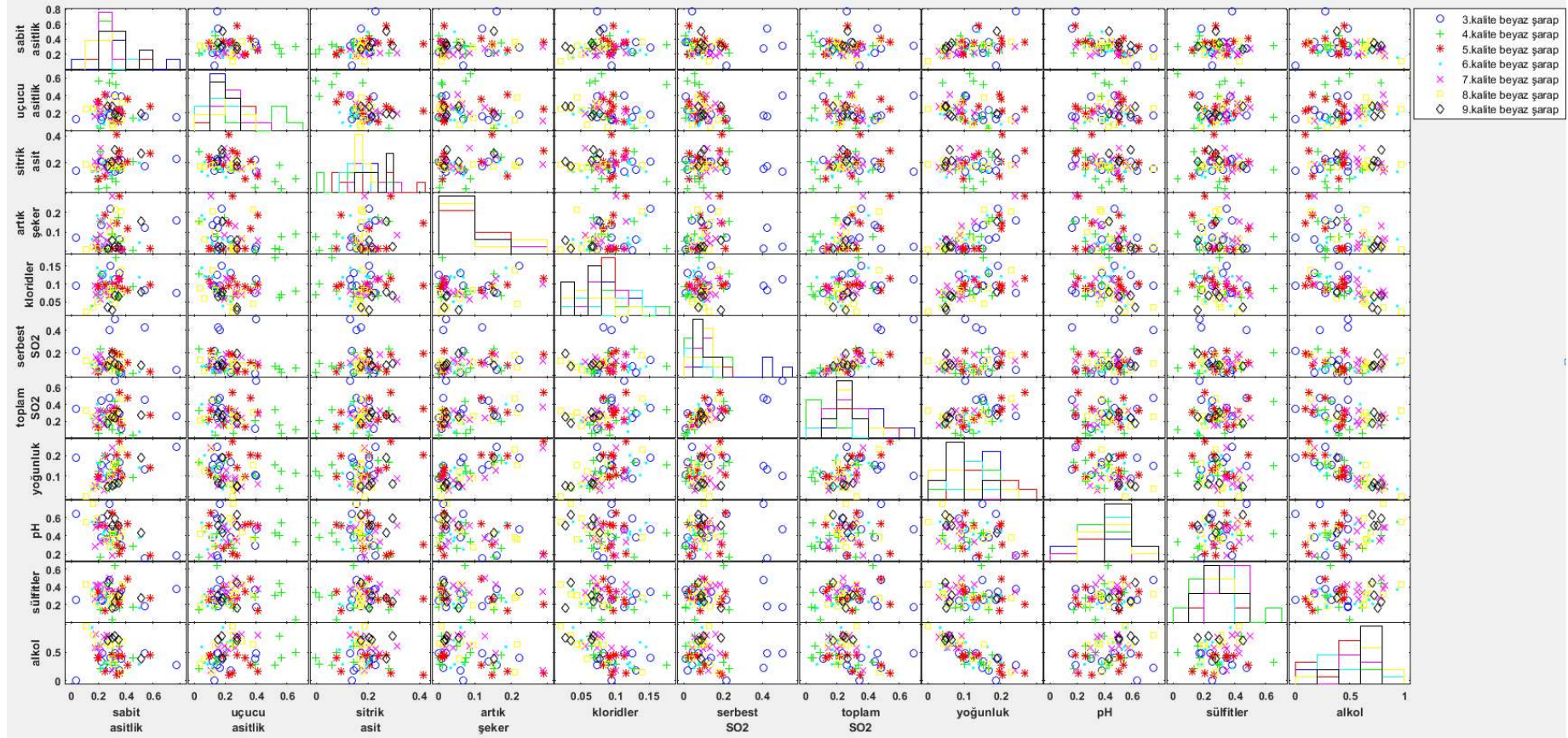
Veri setine ait saçılım grafikleri matrisleri Şekil 2.29-2.31 arasında yer almaktadır. Şekil 2.29'da kırmızı ve beyaz şarap örnekleri, Şekil 2.30'da farklı kalitelerdeki kırmızı şarap örnekleri, Şekil 2.31'de farklı kalitelerdeki beyaz şarap örneklerinin tüm özniteliklerinin saçılım grafikleri gösterilmiştir. Bu saçılım grafiklerini oluşturabilmek için Şekil 2.29 için veri setinden 60 adet kırmızı ve 60 adet beyaz şarap örneği rastgele seçilmiştir. Aynı şekilde farklı kalitelerdeki kırmızı ve beyaz şarap örnekleri için her bir kaliteden rastgele 10'ar örnek seçilmiştir.



Şekil 2.29. Kırmızı ve beyaz şarap örnekleri için saçılım grafiği matrisi



Şekil 2.30. Farklı kalitelardaki kırmızı şarap örnekleri için saçılım grafiği matrisi



Şekil 2.31. Farklı kalitelardaki beyaz şarap örnekleri için saçılım grafiği matrisi

2.3. Eğitim ve Test Veri Setlerinin Hazırlanması

K kat çapraz doğrulama yönteminde başlangıç veri seti, rassal olarak, her biri yaklaşık olarak eşit büyüklükte olan K tane ayrık alt kümeye; $A_1, A_2, \dots, A_K$ 'ya ayrılır. Eğitim ve test süreçleri K defa tekrarlanır. Yani ilk iterasyonda $A_2, A_3, \dots, A_K$ alt kümeleri bir model elde etmek için eğitim seti olarak kullanılır ve model A_1 üzerinde test edilir, ikinci iterasyonda $A_1, A_3, \dots, A_K$ alt kümeleri eğitim seti olarak kullanılır ve model A_2 üzerinde test edilir. Yapılan K adet uygulama sonucunda elde edilen K adet sonuçların ortalaması yöntemin sonucu olmaktadır.

Bu tez çalışmasında, test edilecek verilerin başarı değerlendirmesi 3 farklı şekilde ele alınmıştır. Bu değerlendirmeler, 10 kat çapraz doğrulama; tüm veri setinin %66'sının eğitim seti geri kalanının test seti; tüm veri setinin %80'inin eğitim seti geri kalanının test seti olduğu durumlarıdır.

[72]'den alınan veri seti .csv formatından .arff formatına dönüştürülmüştür. Şekil 2.32'de kırmızı şarap örneklerinin .arff formatlı veri dosyasının bir kısmı gösterilmektedir. İngilizce, Attribute Relationship File Format kelimelerinin baş harflerinden oluşmuştur. Arff dosya yapısı, Weka'ya özel olarak geliştirilmiştir ve dosya, metin yapısında tutulmaktadır. Dosyanın ilk satırında, dosyadaki ilişki tipi (relation) tutulmakta olup ikinci satırdan itibaren veri kümesindeki özellikler (attributes) yazılmaktadır. Özelliklerin hemen ardından veri kümesi yer alır ve veri kümesindeki her satır bir örneğe işaret etmektedir. Ayrıca veri kümesindeki her örneğin her özelliği arasında da virgül ayırıcı kullanılmaktadır.

Aşağıdaki örnek dosyada, sınıflandırma için kullanılan öznitelik değerleri bir dosya içerisinde 10 örnek içerecek şekilde gösterilmiştir. Bu değerler tip olarak sayısal değerler olduğundan 'numeric' olarak ifade edilmiştir. Ancak bu değerler 'nominal', 'real', 'string' ve 'date' tiplerinde de olabilir. 'Nominal' tahmin değeridir ve bir tanım kümesini ele alır. 'Real' reel sayılar kümesinden bir değer verileceğinde kullanılır. 'String' veri kümesinin bu özelliğinin serbest yazı şeklinde olabileceğini ifade eder. 'Date' veri kümesinin bu özelliğinin tarih olduğunu ifade eder.

```

@relation whatever

@attribute 'fixed acidity' numeric
@attribute 'volatile acidity' numeric
@attribute 'citric acid' numeric
@attribute 'residual sugar' numeric
@attribute chlorides numeric
@attribute 'free sulfur dioxide' numeric
@attribute 'total sulfur dioxide' numeric
@attribute density numeric
@attribute pH numeric
@attribute sulphates numeric
@attribute alcohol numeric
@attribute quality {5,6,7,4,8,3}

@data
7.8,0.88,0,2.6,0.098,25,67,0.9968,3.2,0.68,9.8,5
7.8,0.76,0.04,2.3,0.092,15,54,0.997,3.26,0.65,9.8,5
11.2,0.28,0.56,1.9,0.075,17,60,0.998,3.16,0.58,9.8,6
7.4,0.7,0,1.9,0.076,11,34,0.9978,3.51,0.56,9.4,5
7.4,0.66,0,1.8,0.075,13,40,0.9978,3.51,0.56,9.4,5
7.9,0.6,0.06,1.6,0.069,15,59,0.9964,3.3,0.46,9.4,5
7.3,0.65,0,1.2,0.065,15,21,0.9946,3.39,0.47,10,7
7.8,0.58,0.02,2,0.073,9,18,0.9968,3.36,0.57,9.5,7
7.5,0.5,0.36,6.1,0.071,17,102,0.9978,3.35,0.8,10.5,5
6.7,0.58,0.08,1.8,0.097,15,65,0.9959,3.28,0.54,9.2,5

```

Şekil 2.32. Kırmızı şarap örneklerinin .arff formatlı dosyasının WordPad’de görünümü

2.4. Sınıflandırma Algoritmalarına Ait Bulgular

Weka, veri madenciliği işlemleri için kullanılan makine öğrenmesi algoritmalarının bulunduğu bir kütüphanedir. Tez kapsamında, Weka yazılımının 3.9.0 versiyonu kullanılmıştır. Bu tez çalışmasında destek vektör makineleri, k-en yakın komşuluk ve rastgele orman algoritmaları sınıflandırma için kullanılmıştır. Bu algoritmalara ait Weka yazılımında oluşturulan modeller ve bu modellere ait sonuçlar detaylı bir şekilde incelenmiştir. Çalışmada kullanılan performans ölçekleri duyarlılık, anma, F-ölçeği ve ROC alanıdır.

k-EYK algoritması için uzaklık ölçütü olarak Öklid uzaklık ölçüsü seçilmiştir. Karar ormanını oluşturacak karar ağacı sayısı Breiman tarafından önerilen ve varsayılan değer olan 100 olarak seçilmiştir. Her düğümde rastgele seçilecek değişken sayısı, toplam değişken sayısının karekökünün tam sayıya yuvarlandığı yöntem kullanılarak belirlenmiştir [65]. Bu yöntemle göre, bu çalışmada kullanılan veri setinde 11 değişken olduğundan bireysel ağaçlar oluşturulurken her düğümde $\sqrt{11} \cong 4$ değişken kullanılması uygundur.

[72]'den alınan veri seti, farklı kalitelerdeki kırmızı ve beyaz şarap örneklerini içeren iki veri setinden oluşmaktadır. İlk aşamada, bu iki veri setini kırmızı ve beyaz şaraplar olarak sınıflandırabilmek için, iki veri setine de ayrı olarak kırmızı ve beyaz etiketli yeni değişkenler eklenmiştir ve tek bir veri setine dönüştürülmüştür. Tablo 2.3, kırmızı ve beyaz şarap örneklerinin sınıflandırılmasının performans sonuçlarını sunmaktadır.

Tablo 2.3. Karışık veri setinin kırmızı ve beyaz şaraplar olarak sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	99.1	99.1	99.1	98.6
	k-EYK	99.2	99.2	99.2	99.0
	RO	99.5	99.5	99.5	99.8
%66 eğitim %34 test	DVM	99.2	99.2	99.2	98.7
	k-EYK	99.2	99.2	99.2	98.9
	RO	99.4	99.4	99.4	99.9
%80 eğitim %20 test	DVM	99.2	99.2	99.2	98.6
	k-EYK	99.2	99.2	99.2	98.7
	RO	99.5	99.5	99.5	99.9

Şarap örneklerinin kırmızı ve beyaz olarak sınıflandırılmasının en başarılı sonucu, her üç test değerlendirmesi için de RO algoritmasıyla elde edilmiştir. RO algoritması ile sırasıyla bu üç test değerlendirmesi için % 99.5075, % 99.3662 ve %99.5381 doğruluk oranı elde edilmiştir. Tablo 2.3, RO algoritmasının diğer algoritmalarından daha iyi performans gösterdiğini açıkça göstermektedir.

İkinci aşamada ise veri setleri şarap örneklerinin kalite sınıflandırmasını yapmak için kullanılmıştır. Kırmızı şarap veri seti 3 ile 8 arasında değişen toplam 6 kalite sınıfından, beyaz şarap veri seti ise 3 ile 9 arasında değişen toplam 7 kalite sınıfından oluşmaktadır. Tablo 2.4 ve Tablo 2.5, sırasıyla kırmızı ve beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının performans sonuçlarını sunmaktadır.

Tablo 2.4. Kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	48.1	58.3	52.7	70.6
	k-EYK	64.3	64.8	64.5	72.7
	RO	67.6	69.9	68.2	86.3
%66 eğitim %34 test	DVM	49.2	58.5	53.4	70.3
	k-EYK	61.6	61.6	61.4	70.6
	RO	65.6	69.1	67.3	85.7
%80 eğitim %20 test	DVM	49.7	59.1	53.9	70.8
	k-EYK	65.6	65.6	65.5	72.8
	RO	69.6	71.9	70.5	87.2

Tablo 2.5. Beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	39.6	52.1	44.1	66.7
	k-EYK	65.1	65.4	65.2	75.0
	RO	71.0	70.4	69.5	87.3
%66 eğitim %34 test	DVM	39.8	51.4	43.8	65.3
	k-EYK	60.6	60.7	60.6	71.3
	RO	67.1	66.6	65.5	83.9
%80 eğitim %20 test	DVM	39.4	51.2	43.7	65.8
	k-EYK	63.0	63.3	63.0	73.1
	RO	69.8	68.7	67.4	85.7

RO algoritmasının, DVM ve k-EYK algoritmalarına kıyasla kırmızı ve beyaz şarap örneklerini kalitelerine göre sınıflandırırken daha üstün performans gösterdiği Tablo 2.4 ve Tablo 2.5'te görülmektedir. Kırmızı şarap örneklerinin kalitelerine göre RO algoritmasıyla sınıflandırılmasının sonucunda, her üç test değerlendirmesi için sırasıyla %69.906, %69.1176 ve %71.875 doğruluk oranı elde edilmiştir. Beyaz şarap örneklerinin kalitelerine göre RO algoritmasıyla sınıflandırılmasının sonucunda, her üç test değerlendirmesi için sırasıyla %70.3757, %66.6006 ve %68.6735 doğruluk oranı elde edilmiştir.

Üçüncü aşamada ise veri seti, kırmızı ve beyaz şarap örneklerinin birlikte kalite sınıflandırmasını yapmak için birleştirilmiştir. Bu durumda veri seti, 3 ile 8 arasında değişen toplam 6 adet kırmızı şarap kalitesine ait örnekler ve 3 ile 9 arasında değişen toplam 7 adet beyaz şarap kalitesine ait örneklerden oluşmaktadır. Veri seti toplamda 13 sınıflı hale gelmiştir. Tablo 2.6, kırmızı ve beyaz şarap örneklerinin birlikte kalite sınıflandırılmasının performans sonuçlarını göstermektedir.

Tablo 2.6. Kırmızı ve beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	43.1	53.0	45.7	81.5
	k-EYK	64.6	65.0	64.7	78.3
	RO	70.3	70.6	69.4	93.1
%66 eğitim %34 test	DVM	40.5	52.5	45.0	81.2
	k-EYK	59.6	60.0	59.7	74.7
	RO	66.4	67.0	65.8	90.9
%80 eğitim %20 test	DVM	41.7	54.1	46.4	81.3
	k-EYK	63.3	63.5	63.3	76.5
	RO	69.5	69.0	67.7	92.0

Her bir aşamada yapılan analizler sonucunda, şarap örneklerinin sınıflandırılmasında RO algoritmasının DVM ve k-EYK algoritmalarından daha üstün, k-EYK algoritmasının da DVM algoritmasından daha üstün olduğu görülmüştür.

2.5. Boyut İndirgemeye Ait Bulgular

Bu tez çalışmasında, sınıflandırma başarımını arttırmak amacıyla özneliklere boyut indirgeme yöntemlerinden temel bileşenler analizi Weka yazılımında uygulanmıştır. Boyut indirgenmesi gerçekleştirildikten sonra şarap örneklerinin kalitelerine göre sınıflandırılması süreci aynı algoritmalar kullanılarak değerlendirilmiştir.

Tablo 2.7-2.9 özneliklere TBA uygulandıktan sonra şarap veri setinin kalitelerine göre sınıflandırılmasının sonuçlarını göstermektedir.

Tablo 2.7. Kırmızı şarap örneklerinin TBA uygulandıktan sonra kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	47.8	58.0	52.4	69.7
	k-EYK	64.3	64.8	64.5	72.7
	RO	68.4	71.2	69.4	86.4
%66 eğitim %34 test	DVM	49.2	58.3	53.3	69.8
	k-EYK	62.4	62.3	62.1	70.1
	RO	66.9	70.0	68.3	84.9
%80 eğitim %20 test	DVM	47.8	56.9	51.9	69.3
	k-EYK	68.0	67.8	67.8	74.4
	RO	71.4	73.4	72.1	87.8

Tablo 2.8. Beyaz şarap örneklerinin TBA uygulandıktan sonra kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	39.8	52.2	44.0	66.3
	k-EYK	64.5	64.7	64.6	74.4
	RO	70.7	69.9	68.8	86.9
%66 eğitim %34 test	DVM	39.8	51.4	43.8	65.2
	k-EYK	61.0	60.9	60.9	71.4
	RO	66.4	66.2	64.9	83.5
%80 eğitim %20 test	DVM	39.4	51.2	43.6	65.9
	k-EYK	63.5	63.6	63.5	73.7
	RO	68.1	67.4	66.3	85.4

Tablo 2.4 ile Tablo 2.7 karşılaştırıldığında kırmızı şarap veri setine TBA uygulamak DVM algoritmasında başarıyı düşürürken, k-EYK ve RO algoritmalarında başarıyı arttırmıştır. Bunun aksine, beyaz şarap örnekleri için TBA uygulamak DVM algoritması için bir değişikliğe sebep olmazken diğer algoritmalarda başarıyı düşürdüğü Tablo 2.5 ve Tablo 2.8’den gözlenmiştir.

Tablo 2.9. Kırmızı ve beyaz şarap örneklerinin TBA uygulandıktan sonra kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	45.7	52.8	45.3	80.9
	k-EYK	64.0	64.3	64.1	77.6
	RO	70.3	70.1	68.9	92.6
%66 eğitim %34 test	DVM	40.2	51.3	43.5	80.1
	k-EYK	58.8	59.1	58.8	74.2
	RO	65.8	65.6	64.5	90.7
%80 eğitim %20 test	DVM	41.5	53.6	45.7	80.4
	k-EYK	62.5	62.7	62.5	76.6
	RO	67.4	68.0	66.6	91.6

Kırmızı ve beyaz şarap örneklerinin kalite sınıflandırmasının birlikte yapıldığı durumda veri setine TBA uygulamanın tüm algoritmalarda başarıyı düşürdüğü Tablo 2.6 ve Tablo 2.9'dan görülmektedir.

2.6. Özellik Seçimine Ait Bulgular

2.6.1. Filtre Modelli Özellik Seçim Algoritmalarına Ait Bulgular

Filtre modeli özellik seçme algoritmaları oluşturulurken ilk işlem adımı olarak veri setinin tüm özniteliklerine algoritma uygulanarak her niteliğin istatistiksel olarak bulunduğu sınıf etiketi ile ne kadar fazla ilişkili ve diğer nitelikler ile ne kadar az ilişkili olduğunu belirten bir katsayı hesaplanır.

Çalışmada kullanılan veri setindeki mevcut 11 öznitelikten, sınıf bilgisine erişmede hangilerinin daha fazla katkı verdiğini bulabilmek için filtre modeli özellik seçme algoritmalarından bilgi kazancı ve kazanım oranı algoritmaları kullanılmıştır.

Tablo 2.10 ve Tablo 2.11, sırasıyla kırmızı şarap örneklerine ait özniteliklerin bilgi kazancı ve kazanım oranı algoritmaları yardımıyla bulunan ağırlıklarını göstermektedir. Tablo 2.12 ve Tablo 2.13 de ise beyaz şarap örneklerine ait özniteliklerin bilgi kazancı ve kazanım oranı ağırlıkları bulunmaktadır.

Tablo 2.10. Kırmızı şarap örneklerine ait özniteliklerin bilgi kazancı ağırlıkları

Öznitelikler	Bilgi Kazancı Ağırlıkları
Alkol	0.2324
Uçucu Asitlik	0.1282
Sülfidler	0.1194
Sitrik Asit	0.0774
Toplam SO ₂	0.0703
Yoğunluk	0.0570
Kloridler	0.0364
Sabit Asitlik	0.0216
Artık Şeker	0.0000
pH	0.0000
Serbest SO ₂	0.0000

Tablo 2.11. Kırmızı şarap örneklerine ait özniteliklerin kazanım oranı ağırlıkları

Öznitelikler	Kazanım Oranı Ağırlıkları
Alkol	0.1213
Uçucu Asitlik	0.0897
Sülfidler	0.0791
Yoğunluk	0.0784
Toplam SO ₂	0.0622
Sitrik Asit	0.0501
Kloridler	0.0477
Sabit Asitlik	0.0343
Artık Şeker	0.0000
pH	0.0000
Serbest SO ₂	0.0000

Tablo 2.10 ve Tablo 2.11'e bakıldığında kırmızı şarap örnekleri için her iki algoritmada da en iyi 3 özniteliğin alkol, uçucu asitlik ve sülfidler olduğu söylenebilir. Artık şeker, pH ve serbest SO₂ ise en kötü 3 özniteliktir. Özniteliklere bir sıralama yapıldığında her iki algoritma için de en iyi 3, 6, 7 ve 8 öznitelikleri, aynı öznitelikleri ifade etmektedir.

Tablo 2.12. Beyaz şarap örneklerine ait özniteliklerin bilgi kazancı ağırlıkları

Öznitelikler	Bilgi Kazancı Ağırlıkları
Alkol	0.2078
Yoğunluk	0.1177
Kloridler	0.0774
Sitrik Asit	0.0672
Toplam SO ₂	0.0629
Artık Şeker	0.0618
Serbest SO ₂	0.0597
Uçucu Asitlik	0.0571
pH	0.0162
Sülfidler	0.0148
Sabit Asitlik	0.0125

Tablo 2.13. Beyaz şarap örneklerine ait özniteliklerin kazanım oranı ağırlıkları

Öznitelikler	Kazanım Oranı Ağırlıkları
Alkol	0.0754
Yoğunluk	0.0674
Kloridler	0.0590
Sabit Asitlik	0.0393
Toplam SO ₂	0.0366
Serbest SO ₂	0.0357
Uçucu Asitlik	0.0344
Sitrik Asit	0.0334
Artık Şeker	0.0305
pH	0.0179
Sülfidler	0.0102

Beyaz şarap örnekleri için en iyi 3 öznitelik ise her iki algoritma içinde alkol, yoğunluk ve kloridlerdir.

Tablo 2.14-2.21, kırmızı şarap verilerine bilgi kazancı, kazanım oranı algoritmaları ile öznitelik seçme uygulandıktan sonra seçilen özniteliklerle gerçekleştirilen sınıflandırma performans sonuçlarını sunmaktadır.

Tablo 2.14. Bilgi kazancı/kazanım oranı algoritmalarıyla seçilen en iyi 3 öznitelikle kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	45.6	55.5	50.0	68.1
	k-EYK	61.8	62.0	61.9	70.5
	RO	64.6	66.4	65.2	81.8
%66 eğitim %34 test	DVM	47.2	56.3	51.3	68.4
	k-EYK	59.2	59.4	59.2	68.7
	RO	62.5	63.4	62.4	80.7
%80 eğitim %20 test	DVM	47.1	56.3	51.3	68.8
	k-EYK	61.0	60.6	60.7	69.7
	RO	65.9	66.3	65.3	82.4

Tablo 2.15. Bilgi kazancı algoritmasıyla seçilen en iyi 4 öznitelikle kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	45.7	55.7	50.1	68.1
	k-EYK	64.3	64.2	64.2	72.2
	RO	66.0	67.5	66.4	83.2
%66 eğitim %34 test	DVM	46.5	55.7	50.7	67.9
	k-EYK	62.4	62.3	62.3	72.0
	RO	63.1	64.3	63.4	81.2
%80 eğitim %20 test	DVM	47.3	56.6	51.5	68.9
	k-EYK	64.1	64.1	64.0	72.4
	RO	63.9	63.8	63.4	83.4

Tablo 2.16. Kazanım oranı algoritmasıyla seçilen en iyi 4 öznitelikle kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	45.5	55.5	49.9	67.9
	k-EYK	62.6	63.1	62.9	71.4
	RO	66.0	67.7	66.3	83.2
%66 eğitim %34 test	DVM	47.2	56.3	51.3	68.1
	k-EYK	59.9	60.1	60.0	70.0
	RO	66.1	66.2	65.0	81.3
%80 eğitim %20 test	DVM	47.1	56.3	51.3	68.7
	k-EYK	61.9	61.9	61.9	70.7
	RO	69.2	69.4	68.5	82.4

Tablo 2.17. Bilgi kazancı algoritmasıyla seçilen en iyi 5 öznitelikle kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	47.6	57.7	52.2	69.9
	k-EYK	64.7	65.0	64.9	73.1
	RO	67.9	69.4	68.1	84.9
%66 eğitim %34 test	DVM	49.9	59.0	53.9	70.7
	k-EYK	64.7	64.0	64.1	72.9
	RO	63.4	65.1	63.9	83.0
%80 eğitim %20 test	DVM	49.6	58.8	53.7	70.4
	k-EYK	66.0	65.3	65.4	73.0
	RO	66.7	68.1	67.1	85.5

Tablo 2.18. Kazanım oranı algoritmasıyla seçilen en iyi 5 öznitelikle kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	47.5	57.6	52.1	69.9
	k-EYK	65.0	65.5	65.2	73.9
	RO	68.0	69.7	68.4	85.3
%66 eğitim %34 test	DVM	49.1	58.3	53.2	69.5
	k-EYK	61.3	62.1	61.6	71.1
	RO	64.2	66.2	64.9	83.4
%80 eğitim %20 test	DVM	50.1	59.4	54.3	70.8
	k-EYK	61.9	61.9	61.8	70.2
	RO	67.4	68.8	68.0	84.4

Tablo 2.19. Bilgi kazancı/kazanım oranı algoritmalarıyla seçilen en iyi 6 öznitelikle kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	47.7	57.8	52.2	69.9
	k-EYK	64.9	65.2	65.0	73.1
	RO	67.2	69.2	67.7	85.6
%66 eğitim %34 test	DVM	49.9	59.2	54.1	70.2
	k-EYK	61.8	61.9	61.8	70.5
	RO	66.0	67.8	66.6	84.1
%80 eğitim %20 test	DVM	50.0	59.4	54.2	71.0
	k-EYK	62.3	61.9	62.0	72.4
	RO	67.5	68.1	67.1	85.4

Tablo 2.20. Bilgi kazancı/kazanım oranı algoritmalarıyla seçilen en iyi 7 öznitelikle kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

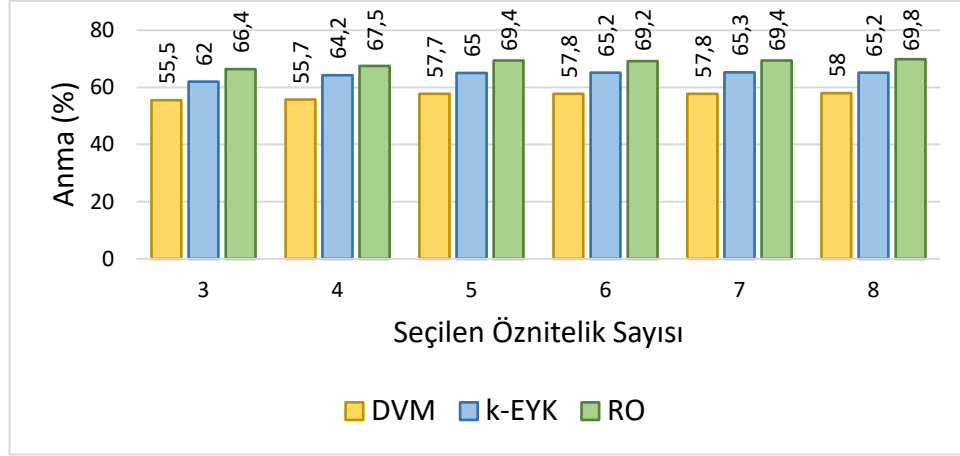
Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	47.7	57.8	52.3	70.0
	k-EYK	64.9	65.3	65.1	73.3
	RO	67.5	69.4	67.9	86.4
%66 eğitim %34 test	DVM	49.9	59.2	54.1	70.4
	k-EYK	62.0	62.1	62.0	71.0
	RO	62.8	65.8	64.3	85.1
%80 eğitim %20 test	DVM	49.1	58.4	53.3	70.0
	k-EYK	62.3	61.6	61.7	70.3
	RO	67.3	69.7	68.5	86.6

Tablo 2.21. Bilgi kazancı/kazanım oranı algoritmalarıyla seçilen en iyi 8 öznitelikle kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

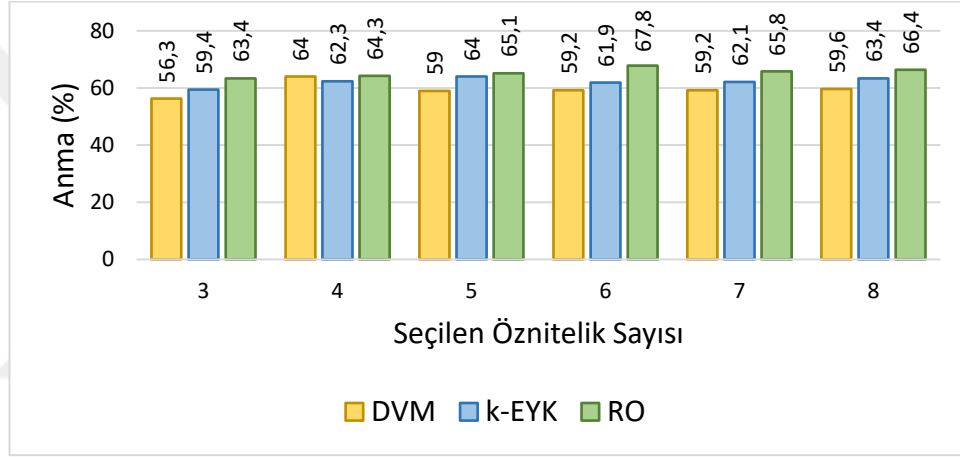
Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	47.8	58.0	52.4	70.1
	k-EYK	64.8	65.2	64.9	73.0
	RO	67.9	69.8	68.2	86.3
%66 eğitim %34 test	DVM	50.3	59.6	54.4	71.1
	k-EYK	62.9	63.4	63.0	71.8
	RO	63.3	66.4	64.8	85.0
%80 eğitim %20 test	DVM	49.1	58.1	53.2	70.4
	k-EYK	65.4	65.6	65.3	72.9
	RO	67.6	69.7	68.5	86.8

Tablo 2.14-2.21'e bakıldığında her iki algoritma ile seçilen özniteliklerle, kırmızı şarap örneklerinin kalitelerine göre sınıflandırma sonuçlarının Tablo 2.4'e yakın sonuçlar ürettiği görülmüştür. Bilgi kazancı ve kazanım oranı algoritmalarının hemen hemen birbirleriyle benzer sonuçlar verdikleri gözlemlenmiştir.

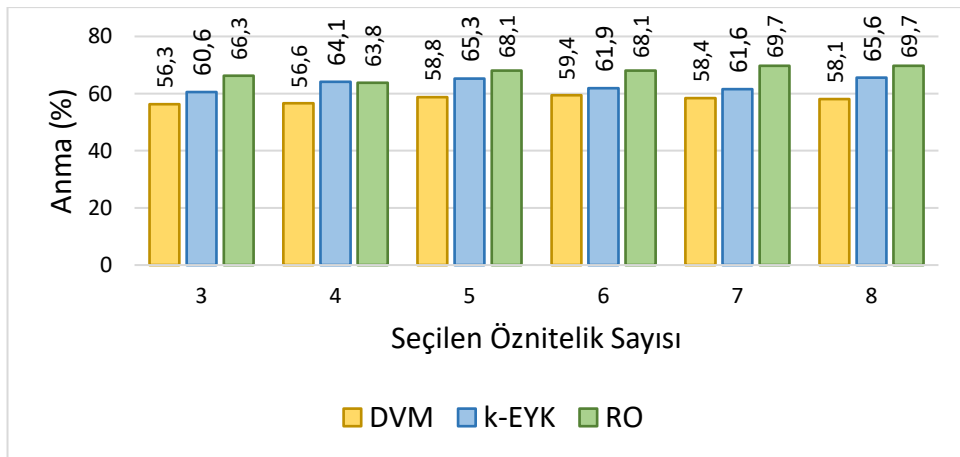
Şekil 2.33 ve Şekil 2.34 kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının, sırasıyla bilgi kazancı ve kazanım oranı algoritmalarıyla seçilen öznitelik sayısına göre anma değeri başarı sonuçlarını göstermektedir.



(a)

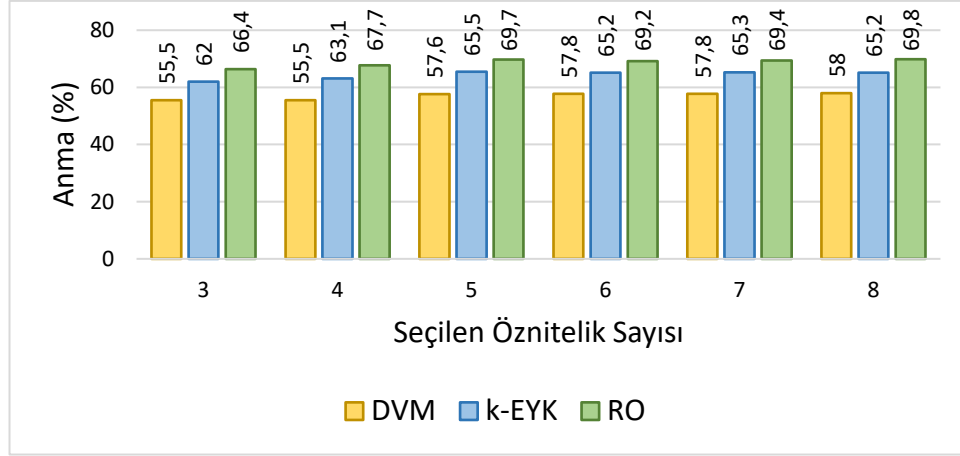


(b)

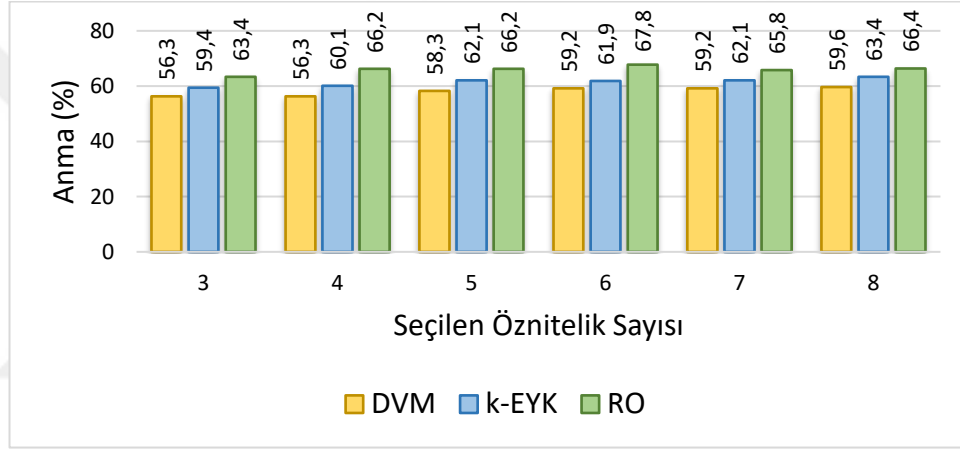


(c)

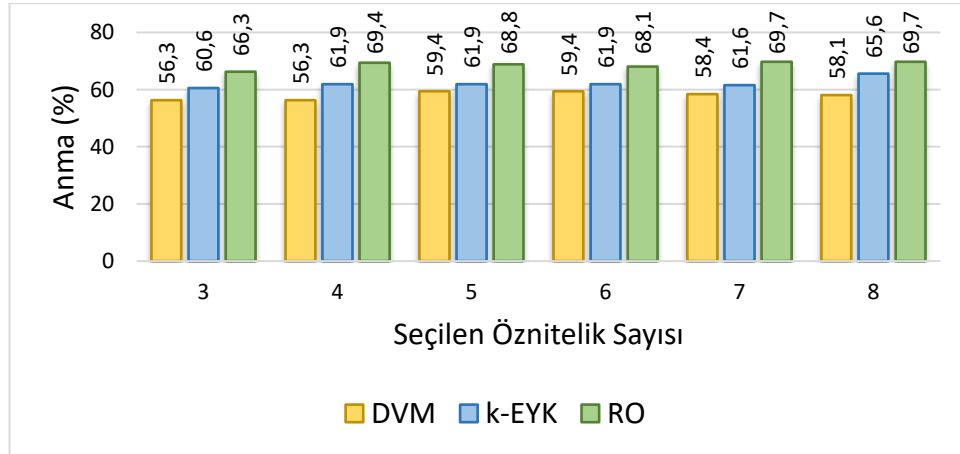
Şekil 2.33. Bilgi kazancı algoritmasıyla seçilen özniteliklerin sayısına göre a) 10 kat çapraz doğrulama b) %66 eğitim c) %80 eğitim verileri için kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının anma değerine göre karşılaştırılması



(a)



(b)



(c)

Şekil 2.34. Kazanım oranı algoritmasıyla seçilen özniteliklerin sayısına göre a) 10 kat çapraz doğrulama b) %66 eğitim c) %80 eğitim verileri için kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının anma değerine göre karşılaştırılması

Tablo 2.22-2.32, beyaz şarap verilerine bilgi kazancı, kazanım oranı algoritmaları ile öznitelik seçme uygulandıktan sonra seçilen özniteliklerle gerçekleştirilen sınıflandırma performans sonuçlarını sunmaktadır.

Şekil 2.35 ve Şekil 2.36 beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının, sırasıyla bilgi kazancı ve kazanım oranı algoritmalarıyla seçilen öznitelik sayısına göre anma değeri başarı sonuçlarını göstermektedir.

Tablo 2.22. Bilgi kazancı/kazanım oranı algoritmalarıyla seçilen en iyi 3 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	36.7	49.1	41.9	63.3
	k-EYK	60.5	60.6	60.5	72.0
	RO	62.9	63.5	62.7	80.5
%66 eğitim %34 test	DVM	36.8	48.3	41.6	62.3
	k-EYK	56.1	56.2	56.0	67.9
	RO	58.2	58.9	58.0	76.6
%80 eğitim %20 test	DVM	36.0	47.8	40.9	62.3
	k-EYK	58.1	58.9	58.4	70.5
	RO	60.8	61.6	60.7	78.4

Tablo 2.23. Bilgi kazancı algoritmasıyla seçilen en iyi 4 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	36.6	49.0	41.8	63.2
	k-EYK	61.6	61.6	61.7	72.4
	RO	65.0	65.1	64.3	82.3
%66 eğitim %34 test	DVM	36.7	48.3	41.6	62.3
	k-EYK	56.6	56.4	56.4	68.6
	RO	59.3	60.1	59.1	78.5
%80 eğitim %20 test	DVM	36.0	47.8	40.9	62.4
	k-EYK	59.5	60.1	59.7	70.8
	RO	61.2	61.5	60.4	80.3

Tablo 2.24. Kazanım oranı algoritmasıyla seçilen en iyi 4 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	36.6	49.1	41.9	63.6
	k-EYK	62.2	62.2	62.2	72.7
	RO	64.7	65.0	64.2	82.6
%66 eğitim %34 test	DVM	36.9	48.6	41.8	62.6
	k-EYK	55.4	55.8	55.6	67.6
	RO	59.0	59.7	58.8	78.4
%80 eğitim %20 test	DVM	36.2	48.0	41.1	62.6
	k-EYK	59.2	59.8	59.4	70.5
	RO	63.0	63.0	61.9	80.9

Tablo 2.25. Bilgi kazancı algoritmasıyla seçilen en iyi 5 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	36.8	49.2	42.0	63.4
	k-EYK	61.8	61.8	61.8	72.5
	RO	66.7	66.6	65.7	83.7
%66 eğitim %34 test	DVM	37.0	48.7	41.9	62.6
	k-EYK	57.4	57.4	57.3	69.1
	RO	62.5	62.3	61.4	79.4
%80 eğitim %20 test	DVM	36.5	48.4	41.5	62.9
	k-EYK	58.6	58.9	58.5	70.0
	RO	62.9	62.4	61.2	81.4

Tablo 2.26. Kazanım oranı algoritmasıyla seçilen en iyi 5 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	36.8	49.3	42.1	63.7
	k-EYK	61.3	61.5	61.4	72.2
	RO	65.7	65.6	64.7	83.8
%66 eğitim %34 test	DVM	36.9	48.6	41.8	62.7
	k-EYK	57.0	56.9	56.9	68.2
	RO	60.9	60.7	59.6	79.4
%80 eğitim %20 test	DVM	36.1	47.9	41.0	62.6
	k-EYK	61.6	61.5	61.5	71.9
	RO	63.9	63.4	62.2	82.1

Tablo 2.27. Bilgi kazancı algoritmasıyla seçilen en iyi 6 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	36.9	49.4	42.2	63.6
	k-EYK	62.7	62.9	62.8	73.0
	RO	67.7	67.3	66.4	84.7
%66 eğitim %34 test	DVM	37.1	48.8	42.0	62.4
	k-EYK	58.4	58.4	58.4	69.9
	RO	62.6	62.6	61.5	80.0
%80 eğitim %20 test	DVM	36.1	47.9	41.0	62.2
	k-EYK	60.7	60.9	60.7	71.6
	RO	64.7	63.9	62.5	81.9

Tablo 2.28. Kazanım oranı algoritmasıyla seçilen en iyi 6 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	37.0	49.6	42.4	64.1
	k-EYK	62.1	62.3	62.2	72.6
	RO	66.9	66.9	66.1	85.1
%66 eğitim %34 test	DVM	37.4	49.2	42.4	63.4
	k-EYK	58.6	58.8	58.7	69.9
	RO	63.6	63.7	62.7	82.3
%80 eğitim %20 test	DVM	36.3	48.2	41.3	62.7
	k-EYK	63.5	63.6	63.4	73.8
	RO	63.7	64.0	63.0	83.8

Tablo 2.29. Bilgi kazancı algoritmasıyla seçilen en iyi 7 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	37.0	49.5	42.3	64.0
	k-EYK	62.9	63.1	63.0	73.4
	RO	68.8	68.4	67.6	85.9
%66 eğitim %34 test	DVM	37.8	49.7	42.8	63.3
	k-EYK	58.2	58.0	58.0	69.5
	RO	65.1	64.8	63.7	82.1
%80 eğitim %20 test	DVM	36.5	48.4	41.5	62.9
	k-EYK	61.9	61.8	61.7	72.2
	RO	65.7	65.1	63.8	84.0

Tablo 2.30. Kazanım oranı algoritmasıyla seçilen en iyi 7 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

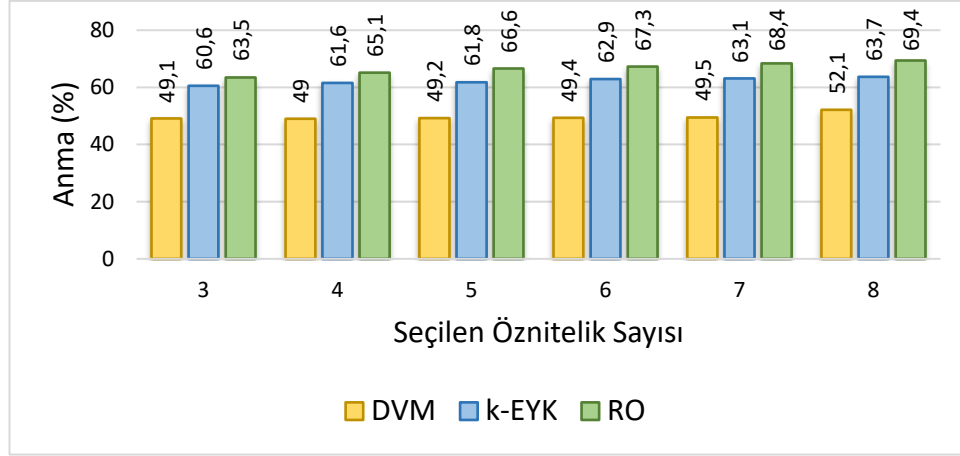
Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	39.5	52.1	44.1	66.4
	k-EYK	63.4	63.6	63.5	73.3
	RO	69.2	69.0	68.1	86.3
%66 eğitim %34 test	DVM	40.2	51.7	44.0	65.4
	k-EYK	60.2	60.1	60.0	70.8
	RO	65.4	65.3	64.2	83.1
%80 eğitim %20 test	DVM	40.0	51.5	43.9	65.6
	k-EYK	62.5	62.2	62.3	73.3
	RO	66.8	66.3	65.1	84.3

Tablo 2.31. Bilgi kazancı algoritmasıyla seçilen en iyi 8 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

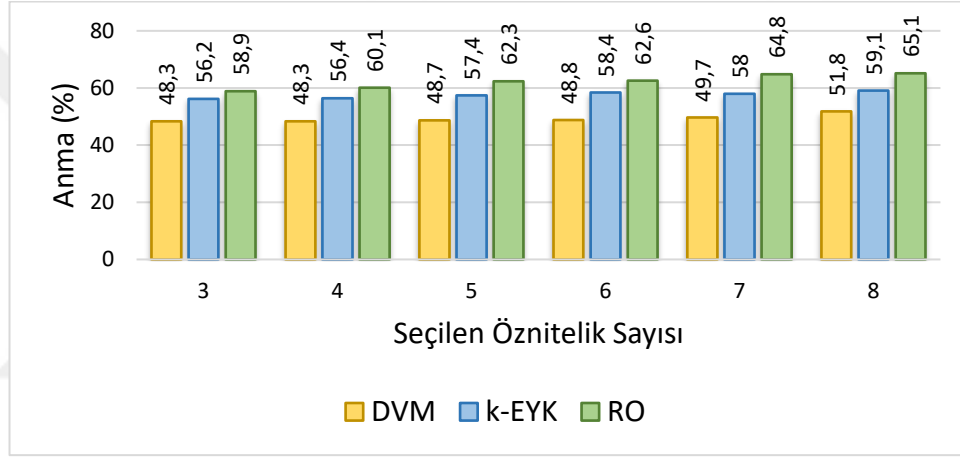
Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	39.9	52.1	44.0	66.3
	k-EYK	63.6	63.7	63.6	74.0
	RO	69.7	69.4	68.5	86.5
%66 eğitim %34 test	DVM	40.4	51.8	44.2	65.3
	k-EYK	59.4	59.1	59.1	70.2
	RO	65.0	65.1	64.0	82.5
%80 eğitim %20 test	DVM	39.4	50.6	43.0	65.0
	k-EYK	61.1	61.1	61.0	71.8
	RO	67.2	66.8	65.5	84.5

Tablo 2.32. Kazanım oranı algoritmasıyla seçilen en iyi 8 öznitelikle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

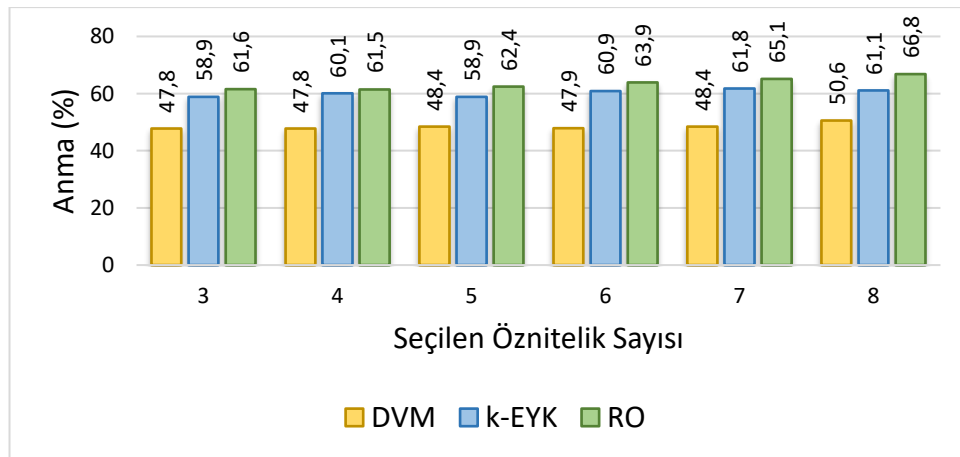
Test Verilerinin Seçimi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
10 kat çapraz doğrulama	DVM	39.5	52.1	44.0	66.4
	k-EYK	63.6	63.8	63.7	73.7
	RO	69.7	69.2	68.3	86.4
%66 eğitim %34 test	DVM	40.3	51.7	44.0	65.4
	k-EYK	60.4	60.0	60.1	71.2
	RO	65.4	65.3	64.2	82.9
%80 eğitim %20 test	DVM	39.9	51.6	44.0	65.7
	k-EYK	62.4	61.6	61.9	72.8
	RO	66.5	65.7	64.5	84.8



(a)

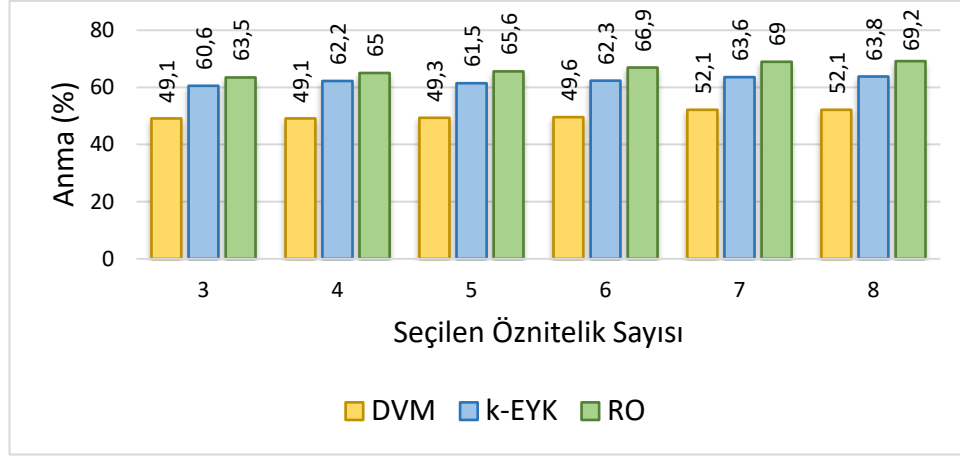


(b)

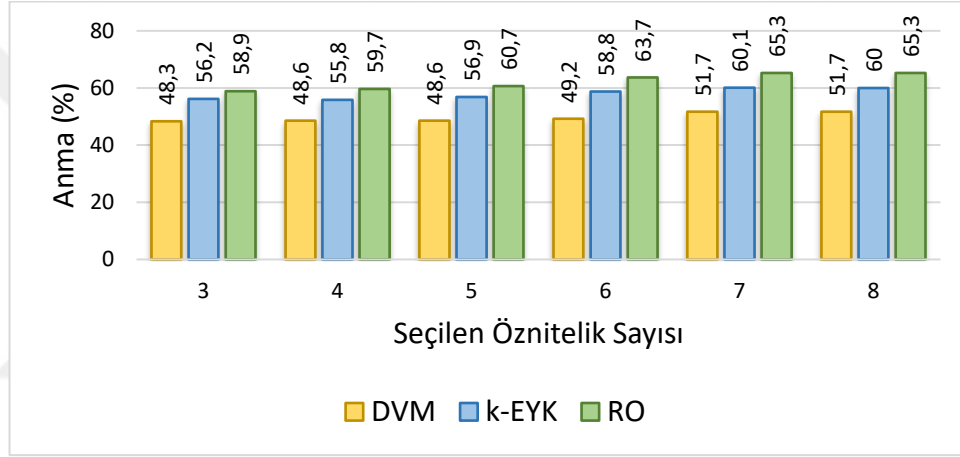


(c)

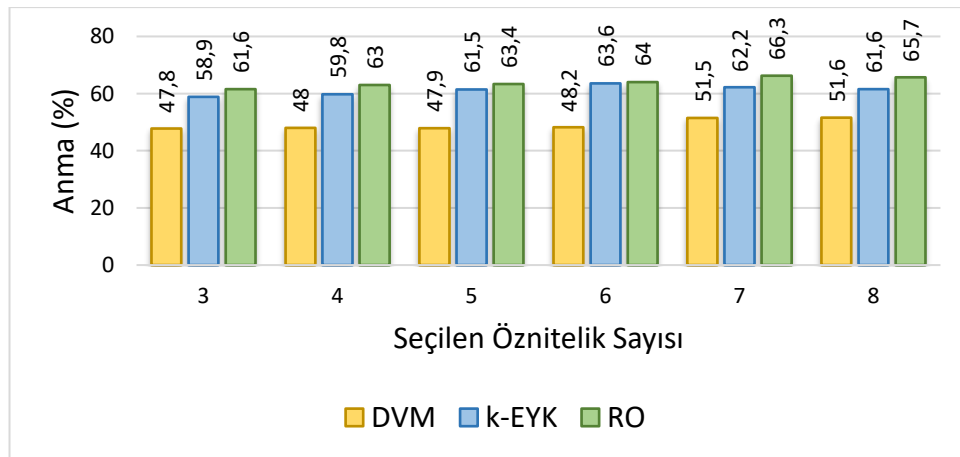
Şekil 2.35. Bilgi kazancı algoritmasıyla seçilen özniteliklerin sayısına göre a) 10 kat çapraz doğrulama b) %66 eğitim c) %80 eğitim verileri için beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının anma değerine göre karşılaştırılması



(a)



(b)



(c)

Şekil 2.36. Kazanım oranı algoritmasıyla seçilen özniteliklerin sayısına göre a) 10 kat çapraz doğrulama b) %66 eğitim c) %80 eğitim verileri için beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının anma değerine göre karşılaştırılması

Filtre modellenli öznitelik seçme algoritmalarından bilgi kazancı ve kazanım oranı her iki veri setine de uygulandığında belirlenen en iyi özniteliklerin birbirine yakın sınıflandırma sonuçları ürettiği gözlenmiştir.

2.6.2. Sarmal Modelli Özellik Seçim Algoritmalarına Ait Bulgular

Sarmalama yöntemi, belirli bir sınıflandırma algoritması kullanarak özellik uzayını bazı özellikleri ihmal ederek araştırır ve ihmal edilen özelliklerin tahmin sonuçlarına nasıl etki ettiğini analiz eder. Öğrenme sürecinde bir özelliğin belirgin bir fark ortaya çıkarması o özelliğin önemli olduğuna ve öğrenme açısından kaliteli bir özellik olduğu anlamına gelir.

DVM, k-EYK ve RO algoritmaları için sarmalama yöntemi ile özelliklerin en iyi alt kümesi elde edilmeye çalışılmıştır. Veri setindeki özellikler 1'den 11'e kadar sırasıyla sabit asitlik, uçucu asitlik, sitrik asit, artık şeker, kloridler, serbest SO₂, toplam SO₂, yoğunluk, pH, sülfidler ve alkol olmak üzere numaralandırılmıştır.

Tablo 2.33 ve Tablo 2.34 sırasıyla kırmızı ve beyaz şarap verilerine sarmalama yöntemi ile özellik seçimi yapıldıktan sonra seçilen özelliklerle gerçekleştirilen sınıflandırma sonuçlarını vermektedir.

Tablo 2.33. Sarmalama yöntemiyle seçilen özniteliklerle kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Seçilen Özellikler	Sınıflandırıcılar	Test Verilerinin Seçimi	Duyarlılık (%)	Anma (%)	F-Ölçütü (%)	ROC Alanı (%)
1, 2, 3, 5, 6, 7, 9, 10, 11	DVM	10 kat ÇD	48.2	58.5	52.8	70.3
		%34 test	50.0	59.4	54.2	71.0
		%20 test	49.6	59.1	53.9	70.6
1, 6, 7, 8, 11	k-EYK	10 kat ÇD	63.7	64.7	64.2	71.8
		%34 test	64.5	64.5	64.4	73.6
		%20 test	65.6	65.0	65.2	74.1
1, 2, 5, 7, 8, 10, 11	RO	10 kat ÇD	67.5	69.9	68.2	86.1
		%34 test	65.1	68.4	66.7	86.1
		%20 test	66.7	68.4	67.3	86.6

Tablo 2.34. Sarmalama yöntemiyle seçilen özneliliklerle beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Seçilen Özellikler	Sınıflandırıcılar	Test Verilerinin Seçimi	Duyarlılık (%)	Anma (%)	F-Ölçütü (%)	ROC Alanı (%)
2, 4, 7, 8, 9, 11	DVM	10 kat ÇD	39.9	52.4	44.4	66.4
		%34 test	39.7	51.2	43.7	65.1
		%20 test	39.4	51.0	43.5	65.3
1, 2, 3, 4, 5, 6, 8, 9, 10, 11	k-EYK	10 kat ÇD	64.9	65.1	65.0	74.8
		%34 test	61.0	61.1	61.0	71.8
		%20 test	62.9	62.8	62.7	72.7
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11	RO	10 kat ÇD	71.0	70.4	69.5	87.3
		%34 test	67.1	66.6	65.5	83.9
		%20 test	69.8	68.7	67.4	85.7

2.7. Örneklemeye Yöntemlerine Ait Bulgular

Örneklemeye yöntemleri, eğitim verilerindeki sınıf dağılımını dengelemek için kullanılmıştır.

Tablo 2.35 veri setindeki (I) kırmızı şarap örneklerinden, SMOTE ve rastgele alt örneklemeye ile oluşturulan yeni eğitim kümelerinin (II, III ve IV) her bir kalitedeki örnek sayılarını göstermektedir. Tablo 2.36 ise veri setindeki (I) beyaz şarap örneklerinden SMOTE ve rastgele alt örneklemeye ile oluşturulan yeni eğitim kümelerinin (II, III, IV ve V) her bir kalitedeki örnek sayılarını gösterir.

Tablo 2.35. Kırmızı şarap örneklerinden elde edilen yeni eğitim kümelerinin örnek sayıları

Kalite	Örnek Sayısı			
	I	II	III	IV
k3	10	52	198	680
k4	53	53	198	681
k5	681	681	681	681
k6	638	638	638	681
k7	199	199	199	680
k8	18	52	198	680

Tablo 2.36. Beyaz şarap örneklerinden elde edilen yeni eğitim kümelerinin örnek sayıları

Kalite	Örnek Sayısı				
	I	II	III	IV	V
b3	20	163	880	879	2198
b4	163	163	879	879	2197
b5	1457	1457	1457	879	2197
b6	2198	2198	2198	879	2198
b7	880	880	880	879	2197
b8	175	175	879	879	2198
b9	5	163	880	879	2198

Örneğin, Tablo 2.35’te gösterilen II numaralı yeni eğitim kümesini oluşturabilmek için k3 ve k8 kalitesindeki kırmızı şarap örneklerinden SMOTE algoritmasıyla yeni yapay azınlık örnekler üretilmiştir ve her iki kalitedeki şarap örnekleri sayısı 52’ye çıkarılmıştır.

Sınıflandırma algoritmalarının performansı eğitim kümelerinin sırasıyla II, III ve IV olduğu durumlar için I numaralı gerçek veri setiyle test edilmiştir. Tablo 2.37’de kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları gösterilmektedir.

Tablo 2.37. SMOTE algoritmasıyla yeni eğitim kümeleri oluşturulan kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Veri Kümesi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
II eğitim I test	DVM	46.8	55.6	49.0	71.2
	k-EYK	63.3	63.7	63.5	72.1
	RO	67.6	69.6	68.1	87.6
III eğitim I test	DVM	52.7	52.6	47.2	76.1
	k-EYK	68.2	68.3	68.2	79.1
	RO	73.0	73.1	72.5	91.9
IV eğitim I test	DVM	54.4	55.6	54.5	84.2
	k-EYK	78.1	78.4	78.2	87.0
	RO	83.3	83.8	83.4	97.1

Beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları Tablo 2.38’de yer almaktadır.

Tablo 2.38. SMOTE ve rastgele alt örnekleme algoritmalarıyla yeni eğitim kümeleri oluşturulan beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı sonuçları

Veri Kümesi	Sınıflandırıcılar	Duyarlılık (%)	Anma (%)	F-Ölçeği (%)	ROC Alanı (%)
II eğitim I test	DVM	35.4	49.1	40.2	67.2
	k-EYK	65.5	65.7	65.6	75.1
	RO	72.1	70.7	69.8	88.8
III eğitim I test	DVM	44.3	46.0	41.5	75.8
	k-EYK	72.6	72.7	72.6	83.5
	RO	78.3	77.7	77.5	95.0
IV eğitim I test	DVM	45.2	46.8	45.4	79.3
	k-EYK	71.9	72.4	72.1	84.1
	RO	77.0	77.8	77.2	95.7
V eğitim I test	DVM	46.8	48.4	47.1	80.1
	k-EYK	79.1	79.5	79.2	88.4
	RO	84.6	85.0	84.7	97.9

Oluşturulan yeni eğitim kümeleriyle, şarap örnekleri kalitelerine göre sınıflandırıldığında en yüksek başarı oranına RO algoritmasıyla ulaşıldığı Tablo 2.37 ve Tablo 2.38’de görülmektedir.

Tablo 2.37’deki sonuçlara bakıldığında en yüksek başarımın kırmızı şarap örnekleri için IV numaralı eğitim kümesinin kullanılmasıyla ortaya çıktığı gözlenmektedir. Bu IV numaralı eğitim kümesi farklı kalitelerdeki şarap örneklerinin sayısının eşit olduğu eğitim kümesidir. Kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının başarı oranının %69.9’dan %83.8’e çıktığı Tablo 2.4 ve Tablo 2.37 ile gözlenmiştir.

Aynı şekilde beyaz şarap örnekleri için en yüksek başarımın V numaralı eğitim kümesiyle elde edildiği Tablo 2.38’de gösterilmiştir. Bu eğitim kümesinde de farklı kalitelerdeki şarap örneklerinin sayısı yaklaşık olarak birbirine eşittir. Beyaz şarap örnekleri için başarı oranının %70.4’ten %85.0’a çıktığı Tablo 2.5 ile Tablo 2.38’de mevcuttur.

Kırmızı ve beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının yapılan tüm yöntemlerle anma değerine göre karşılaştırılması sırasıyla Tablo 2.39 ve Tablo 2.40’ta sunulmuştur. Bilgi kazancı ve kazanım oranı algoritmalarında seçilen en iyi 8 öznitelikle anma başarımı değerlendirilmiştir. SMOTE algoritmasında ise kırmızı şarap örnekleri için IV, beyaz şarap örnekleri için V numaralı eğitim kümesi kullanıldığında oluşan anma başarımı değerlendirmeye alınmıştır.

Tablo 2.39. Kırmızı şarap örneklerinin kalitelerine göre sınıflandırılmasının tüm yöntemlerle anma değerlerinin karşılaştırılması

Test Verilerinin Seçimi	Sınıflandırıcılar	Ham	TBA	Bilgi Kazancı	Kazanım Oranı	Sarmal	SMOTE
10 kat çapraz doğrulama	DVM	58.3	58.0	58.0	58.0	58.5	55.6
	k-EYK	64.8	64.8	65.2	65.2	64.7	78.4
	RO	69.9	71.2	69.8	69.8	69.9	83.8
%66 eğitim %34 test	DVM	58.5	58.3	59.6	59.6	59.4	-
	k-EYK	61.6	62.3	63.4	63.4	64.5	-
	RO	69.1	70.0	66.4	66.4	68.4	-
%80 eğitim %20 test	DVM	59.1	56.9	58.1	58.1	59.1	-
	k-EYK	65.6	67.8	65.6	65.6	65.0	-
	RO	71.9	73.4	69.7	69.7	68.4	-

Tablo 2.40. Beyaz şarap örneklerinin kalitelerine göre sınıflandırılmasının tüm yöntemlerle anma değerlerinin karşılaştırılması

Test Verilerinin Seçimi	Sınıflandırıcılar	Ham	TBA	Bilgi Kazancı	Kazanım Oranı	Sarmal	SMOTE
10 kat çapraz doğrulama	DVM	52.1	52.2	52.1	52.1	52.4	48.4
	k-EYK	65.4	64.7	63.7	63.8	65.1	79.5
	RO	70.4	69.9	69.4	69.2	70.4	85.0
%66 eğitim %34 test	DVM	51.4	51.4	51.8	51.7	51.2	-
	k-EYK	60.7	60.9	59.1	60.0	61.1	-
	RO	66.6	66.2	65.1	65.3	66.6	-
%80 eğitim %20 test	DVM	51.2	51.2	50.6	51.6	51.0	-
	k-EYK	63.3	63.6	61.1	61.6	62.8	-
	RO	68.7	67.4	66.8	65.7	68.7	-

Cortez (2009) ve arkadaşları, kırmızı ve beyaz şarap örneklerini kalitelerine göre sınıflandırırken genel doğruluk oranını sırasıyla %62.4 ve %64.6 bulmuşlardır [73]. Bu tez çalışmasında ise en yüksek doğruluk oranı kırmızı ve beyaz şarap örneklerini kalitelerine göre sınıflandırırken sırasıyla %83.8 ve %85.0 olarak elde edilmiştir.

3. SONUÇLAR

Bu tez çalışması kapsamında, kalitelere göre içecekler farklı algoritmalar kullanılarak sınıflandırılmaya çalışılmıştır. Çalışmada kullanılan veri seti, farklı kalitelerdeki kırmızı ve beyaz şarap örneklerini içeren ayrı iki veri kümesinden oluşmaktadır. Destek Vektör Makinesi, k-En Yakın Komşuluk ve Rastgele Orman bu çalışmada kullanılan sınıflandırma algoritmalarıdır.

Çalışmada kullanılan performans ölçekleri duyarlılık, anma, F-ölçeği ve alıcı işletim karakteristiğidir. Sınıflandırma algoritmaları Weka yazılımında modellenerek, şarap örneklerinin renklerine ve kalitelere göre sınıflandırılması gerçekleştirilmiştir. Hem renklerine göre sınıflandırma yapılırken hem de kalitelere göre sınıflandırma işlemi gerçekleştirilirken en yüksek başarımlarına RO algoritmasıyla ulaşılmıştır.

Veri setine boyut indirgeme yöntemlerinden Temel Bileşenler Analizi uygulanıp sınıflandırma süreci tekrarlanmıştır. Boyut indirgemenin kırmızı şarap örnekleri için RO algoritmasında başarımlarını arttırırken, beyaz şarap örneklerinde ise başarımlarını düşürdüğü gözlenmiştir.

En iyi öznelikleri belirlemek için filtre modeli Bilgi Kazancı ve Kazanım Oranı ile sarmal modeli yöntemler uygulanarak sınıflandırma işlemleri tekrarlanmıştır.

Veri setindeki şarap örneklerinin sayısındaki dengesizlik Azınlıktan Yapay Örnek Üretme ve Rastgele Alt Örnekleme yöntemleri ile giderilerek sınıflandırma süreci tekrarlanmıştır. Azınlıktan Yapay Örnek Üretme (SMOTE) algoritmasının kullanımıyla kırmızı şarap örneklerinin kalitelere göre sınıflandırılmasının en yüksek başarımlar oranı %83.8 olarak, beyaz şarap örneklerinde ise %85.0 olarak elde edilmiştir.

4. KAYNAKLAR

1. Soman, R. and Raman, M., HACCP System-Hazard Analysis and Assessment, Based on ISO22000:2005 Methodology, Food Control, 69 (2016) 191-195.
2. Pico, Y., Chemical Analysis of Food: Techniques and Applications, First Edition, Academic Press, MA, 2012.
3. Śliwińska, M., Wisniewska, P., Dymerski, T., Namiesnik, J. and Wardencki, W., Food Analysis Using Artificial Senses, Journal of Agricultural and Food Chemistry, 62, 7 (2014) 1423-1448.
4. Escuder-Gilabert, L. and Peris, M., Review: Highlights in Recent Applications of Electronic Tongues in Food Analysis, Analytica Chimica Acta, 665, 1 (2010) 15-25.
5. Hayashi, K., Yamanaka, M., Toko, K. and Yamafuji, K., Multichannel Taste Sensor Using Lipid Membranes, Sensors and Actuators B: Chemical, 2, 3 (1990) 205-213.
6. Toko, K., Matsuno, T., Yamafuji, K., Hayashi, K., Ikezaki, H., Sato, K., Toukubo, R. and Kawarai, S., Multichannel Taste Sensor Using Electric Potential Changes in Lipid Membranes, Biosensors and Bioelectronics, 9, 4-5 (1994) 359-364.
7. Toko, K., RETRACTED: Electronic Tongue, Biosensors and Bioelectronics, 13, 6 (1998) 701-709.
8. Vlasov, Y., Legin, A., Rudnitskaya, A., Di Natale, C. and D'Amico, A., Nonspecific Sensor Arrays ("Electronic Tongue") for Chemical Analysis of Liquids, IUPAC Technical Report, Pure and Applied Chemistry, 77, 11 (2005) 1965-1983.
9. Medina-Plaza, C., de Saja, J.A. and Rodriguez-Mendez, M.L., Bioelectronic Tongue Based on Lipidic Nanostructured Layers Containing Phenol Oxidases and Lutetium Bisphthalocyanine for the Analysis of Grapes, Biosensors and Bioelectronics, 57 (2014) 276-283.
10. Men, H., Chen, D., Zhang, X., Liu, J. and Ning, K., Data Fusion of Electronic Nose and Electronic Tongue for Detection of Mixed Edible-Oil, Journal of Sensors, 2014 (2014) 1-7.
11. Wei, Z. and Wang, J., Detection of Antibiotic Residues in Bovine Milk by a Voltammetric Electronic Tongue System, Analytica Chimica Acta, 694, 1-2 (2011) 46-56.
12. Tortora, L., Stefanelli, M., Mastroianni, M., Lvova, L., Di Natale, C., D'Amico, A., Filippini, D., Lundstrom, I. and Paolesse, R., The Hyphenated CSPT-Potentiometric Analytical System: An Application for Vegetable Oil Quality Control, Sensors and Actuators B: Chemical, 142, 2 (2009) 457-463.

13. Zakaria, A., Md Shakaff, A.Y., Masnan, M.J., Ahmad, M.N., Adom, A.H., Jaafar, M.N., Ghani, S.A., Abdullah, A.H., Abdul Aziz, A.H., Kamarudin, L.M., Subari, N. and Fikri, N.A., A Biomimetic Sensor for the Classification of Honeys of Different Floral Origin and the Detection of Adulteration, Sensors, 11 (2011) 7799-7822.
14. Dias, L.G., Fernandes, A., Veloso, C.A., Machado, A.A.S.C., Pereira, J.A. and Peres, A.M., Single-Cultivar Extra Virgin Olive Oil Classification Using a Potentiometric Electronic Tongue, Food Chemistry, 160 (2014) 321-329.
15. Angkawisittpan, N. and Manasri, T., Determination of Sugar Content in Sugar Solutions Using Interdigital Capacitor Sensor, Measurement Science Review, 12, 1 (2012) 8-13.
16. Ulloa, P.A., Guerra, R., Cavaco, A.M., Rosa da Costa, A.M., Figueira, A.C. and Brigas, A.F., Determination of the Botanical Origin of Honey by Sensor Fusion of Impedance E-Tongue and Optical Spectroscopy, Computers and Electronics in Agriculture, 94 (2013) 1-11.
17. Kutyla-Olesiuk, A., Ciosek, P., Romanowska, E. and Wroblewski, W., Effect of Lead Accumulation in Maize Leaves on Their Chemical Images Created by a Flow-Through Electronic Tongue, Talanta, 103 (2013) 179-185.
18. Kutyla-Olesiuk, A., Nowacka, M., Wesoly, M. and Ciosek, P., Evaluation of Organoleptic and Texture Properties of Dried Apples by Hybrid Electronic Tongue, Sensors and Actuators B: Chemical, 187 (2013) 234-240.
19. Apetrei, I.M. and Apetrei, C., Detection of Virgin Olive Oil Adulteration Using a Voltammetric E-Tongue, Computers and Electronics in Agriculture, 108 (2014) 148-154.
20. Apetrei, C., Apetrei, I.M., Villanueva, S., de Saja, J.A., Gutierrez-Rosales, F. and Rodriguez-Mendez, M.L., Combination of an E-Nose, an E-Tongue and an E-Eye for the Characterisation of Olive Oils With Different Degree of Bitterness, Analytica Chimica Acta, 663, 1 (2010) 91-97.
21. Bueno, L., El-Sharif, H.F., Salles, M.O., Boehm, R.D., Narayan, R.J., Paixao, T.R.L.C. and Reddy, S.M., MIP-Based Electrochemical Protein Profiling, Sensors and Actuators B: Chemical, 204 (2014) 88-95.
22. Campos, I., Masot, R., Alcaniz, M., Gil, L., Soto, J., Vivancos, J.L., Garcia-Breijo, E., Labrador, R.H., Barat, J.M. and Martinez-Manez, R., Accurate Concentration Determination of Anions Nitrate, Nitrite and Chloride in Minced Meat Using a Voltammetric Electronic Tongue, Sensors and Actuators B: Chemical, 149, 1 (2010) 71-78.
23. Rosenthal, A.J., Food Texture: Measurement and Perception, First Edition, Springer, New York, 1999.

24. Yasuura, M., Tahara, Y., Ikezaki, H. and Toko, K., Development of a Sweetness Sensor for Aspartame, a Positively Charged High-Potency Sweetener, Sensors, 14, 4 (2014) 7359-7353.
25. Liu, R., Zhang, X., Zhang, L., Gao, X., Li, H., Shi, J. and Li, X., Bitterness Intensity Prediction of Berberine Hydrochloride Using an Electronic Tongue and a GA BP Neural Network, Experimental and Therapeutic Medicine, 7, 6 (2014) 1696-1702.
26. Newman, J., Egan, T., Harbourne, N., O'Riordan, D., Jacquier, J.C. and O'Sullivan, M., Correlation of Sensory Bitterness in Dairy Protein Hydrolysates: Comparison of Prediction Models Built Using Sensory, Chromatographic and Electronic Tongue Data, Talanta, 126 (2014) 46-53.
27. Yang, Y., Chen, Q., Shen, C., Zhang, S., Gan, Z., Hu, R., Zhao, J. and Ni, Y., Evaluation of Monosodium Glutamate, Disodium Inosinate and Guanylate Umami Taste by an Electronic Tongue, Journal of Food Engineering, 116, 3 (2013) 627-632.
28. Ciosek, P., Wesoly, M., Zabadaj, M., Lisiecka, J., Sollohub, K., Cal, K. and Wroblewski, W., Towards Flow-Through/Flow Injection Electronic Tongue for the Analysis of Pharmaceuticals, Sensors and Actuators B: Chemical, 207, (2015) 1087-1094.
29. Apetrei, I.M. and Apetrei, C., Voltammetric E-Tongue for the Quantification of Total Polyphenol Content in Olive Oils, Food Research International, 54, 2 (2013) 2075-2082.
30. Haddi, Z., Alami, H., El Bari, N., Tounsi, M., Barhoumi, H., Maaref, A., Jaffrezic-Renault, N. and Bouchikhi, B., Electronic Nose and Tongue Combination for Improved Classification of Moroccan Virgin Olive Oil Profiles, Food Research International, 54, 2 (2013) 1488-1498.
31. Wei, Z., Wang, J. and Zhang, X., Monitoring of Quality and Storage Time of Unsealed Pasteurized Milk by Voltammetric Electronic Tongue, Electrochimica Acta, 88 (2013) 231-239.
32. Wei, Z., Wang, J. and Jin, W., Evaluation of Varieties of Set Yogurts and Their Physical Properties Using a Voltammetric Electronic Tongue Based on Various Potential Waveforms, Sensors and Actuators B: Chemical, 177 (2013) 684-694.
33. Hruskar, M., Major, N. and Krpan, M., Application of a Potentiometric Sensor Array as a Technique in Sensory Analysis, Talanta, 81, 1-2 (2010) 398-403.
34. Hruskar, M., Major, N., Krpan, M. and Vahcic, N., Simultaneous Determination of Fermented Milk Aroma Compounds by a Potentiometric Sensor Array, Talanta, 82, 4 (2010) 1292-1297.

35. Bougrini, M., Tahri, K., Haddi, Z., El Bari, N., Llobet, E., Jaffrezic-Renault, N. and Bouchikhi, B., Aging Time and Brand Determination of Pasteurized Milk Using a Multisensor E-Nose Combined with a Voltammetric E-Tongue, Materials Science and Engineering: C, 45 (2014) 348-358.
36. Apetrei, I.M., Rodriguez-Mendez, M.L., Apetrei, C. and de Saja, J.A., Fish Freshness Monitoring Using an E-Tongue Based on Polypyrrole Modified Screen-Printed Electrodes, IEEE Sensors Journal, 13, 7 (2013) 2548-2554.
37. Ruiz-Rico, M., Fuentes, A., Masot, R., Alcaniz, M., Fernandez-Segovia, I. and Manuel Barat, J., Use of the Voltammetric Tongue in Fresh Cod (*Gadus Morhua*) Quality Assessment, Innovative Food Science & Emerging Technologies, 18 (2013) 256-263.
38. Liu, Y., Qui, C.Y., Wang, X.C., Miao, X.D., Zhang, J.J. and Xiong, Z.H., Isolation and Identification of Flavor Peptides in Cultured Puffer Fish (*Takifugu Obscurus*), Modern Food Science and Technology, 30, 8 (2014) 38-42.
39. Kaneki, N., Miura, T., Shimada, K., Tanaka, H., Ito, S., Hotori, K., Akasaka, C., Ohkubo, S. and Asano, Y., Measurement of Pork Freshness Using Potentiometric Sensor, Talanta, 62, 1 (2004) 215-219.
40. Gil, L., Barat, J.M., Escriche I., Garcia-Breijo, E., Martinez-Manez, R. and Soto, J., An Electronic Tongue for Fish Freshness Analysis Using a Thick-Film Array of Electrodes, Microchimica Acta, 163, 1-2 (2008) 121-129.
41. Rodriguez-Mendez, M.L., Gay, M., Apetrei, C. and de Saja, J.A., Biogenic Amines and Fish Freshness Assessment Using a Multisensor System Based on Voltammetric Electrodes. Comparison between CPE and Screen-Printed Electrodes, Electrochimica Acta, 54, 27 (2009) 7033-7041.
42. Gil, L., Barat, J., Baigts, D., Martinez-Manez, R., Soto, J., Garcia-Breijo, E., Aristoy, M.C, Toldra, F. and Llobet, E., Monitoring of Physical-Chemical and Microbiological Changes in Fresh Pork Meat Under Cold Storage by means of a Potentiometric Electronic Tongue, Food Chemistry, 126, 3 (2011) 1261-1268.
43. Güney, S. ve Atasoy, A., Fish Freshness Assessment by Using Electronic Nose, 36th International Conference on Telecommunications and Signal Processing (TSP), Haziran 2013, Roma, İtalya, Bildiriler Kitabı, 742-746.
44. Wei, Z. and Wang, J., Classification of Monofloral Honeys by Voltammetric Electronic Tongue with Chemometrics Method, Electrochimica Acta, 56, 13 (2011) 4907-4915.
45. Wei, Z. and Wang, J., Tracing Floral and Geographical Origins of Honeys by Potentiometric and Voltammetric Electronic Tongue, Computers and Electronics in Agriculture, 108 (2014) 112-122.

46. Escriche, I., Kadar, M., Domenech, E. and Gil-Sanchez, L., A Potentiometric Electronic Tongue for the Discrimination of Honey According to the Botanical Origin. Comparison with Traditional Methodologies: Physicochemical Parameters and Volatile Profile, Journal of Food Engineering, 109, 3 (2012) 449-456.
47. Sousa, M.E.B.C., Dias, L.G., Veloso, A.C.A., Estevinho, L., Peres, A.M. and Machado, A.A.S.C., Practical Procedure for Discriminating Monofloral Honey with a Broad Pollen Profile Variability Using an Electronic Tongue, Talanta, 128 (2014) 284-292.
48. Campos, I., Bataller, R., Armero, R., Manuel Gandia, J., Soto, J., Martinez-Manez, R. and Gil-Sanchez, L., Monitoring Grape Ripeness Using a Voltammetric Electronic Tongue, Food Research International, 54, 2 (2013) 1369-1375.
49. Beullens, K., Kirsanov, D., Irudayaraj, J., Rudnitskaya, A., Legin, A., Nicolai, B.M. and Lammertyn, J., The Electronic Tongue and ATR-FTIR for Rapid Detection of Sugars And Acids in Tomatoes, Sensors and Actuators B: Chemical, 116, 1-2 (2006) 107-115.
50. Legin, A., Rudnitskaya, A., Vlasov, Y., Di Natale, C., Davide, F. and D'Amico, A., Tasting of Beverages Using an Electronic Tongue, Sensors and Actuators B: Chemical, 44, 1-3 (1997) 291-296.
51. Vlasov, Y.G., Legin, A., Rudnitskaya, A., D'Amico, A. and Di Natale, C., «Electronic Tongue»-New Analytical Tool for Liquid Analysis on the Basis of Non-Specific Sensors and Methods of Pattern Recognition, Sensors and Actuators B: Chemical, 65, 1-3 (2000) 235-236.
52. Polshin, E., Rudnitskaya, A., Kirsanov, D., Legin, A., Saison, D., Delvaux, F., Delvaux, F.R., Nicolai, B.M. and Lammertyn, J., Electronic Tongue as a Screening Tool for Rapid Analysis of the Quality Parameters of Beer, Talanta, 81, 1-2 (2010) 88-94.
53. Poshtiban, S., Singh, A., Fitzpatrick, G. and Evoy, S., Bacteriophage Tail-Spike Protein Derivatized Microresonator Arrays for Specific Detection of Pathogenic Bacteria, Sensors and Actuators B: Chemical, 181 (2013) 410-416.
54. Peres, A.M., Dias, L.G., Veloso, A.C.A., Meirinho, S.G., Sa Morais, J. and Machado, A.A.S.C., An Electronic Tongue for Gliadins Semiquantitative Detection in Foodstuffs, Talanta, 83, 3 (2011) 857-864.
55. Eves, V., Hullar, I. and Romvari, R., Electronic Nose and Tongue for Pet Food Classification, Agriculturae Conspectus Scientificus, 78, 3 (2013) 225-228.
56. Krantz-Rülcker, C., Stenberg, M., Winqvist, F. and Lundström, I., Electronic Tongues for Environmental Monitoring Based on Sensor Arrays and Pattern Recognition: A Review, Analytica Chimica Acta, 426, 2 (2001) 217-226.

57. Di Natale, C., Macagnano, A., Davide, F., D'Amico, A., Legin, A., Vlasov, Y., Rudnitskaya, A. and Selezenev, B., Multicomponent Analysis on Polluted Waters by means of an Electronic Tongue, Sensors and Actuators B: Chemical, 44, 1-3 (1997) 423-428.
58. Ha, D., Hu, N., Wu, C.X., Kirsanov, D., Legin, A., Khaydukova, M. and Wang, P., Novel Structured Light-Addressable Potentiometric Sensor Array Based on PVC Membrane for Determination of Heavy Metals, Sensors and Actuators B: Chemical, 174 (2012) 59-64.
59. Winqvist, F., Bjorklund, R., Krantz-Rülcker, C., Lundström, I., Östergren, K. and Skoglund, T., An Electronic Tongue in the Dairy Industry, Sensors and Actuators B: Chemical, 111-112 (2005) 299-304.
60. Parra, V., Arrieta, A.A., Fernandez-Escudero, J.A., Iniguez, M., de Saja, J.A. and Rodriguez-Mendez, M.L., Monitoring of the Ageing of Red Wines in Oak Barrels by means of an Hybrid Electronic Tongue, Analytica Chimica Acta, 563, 1-2 (2006) 229-237.
61. Boser, B., Guyon, I.M. and Vapnik, V.N., A Training Algorithm for Optimal Margin Classifiers, 5th Annual Workshop on Computational Learning Theory, July 27-29, 1992, New York, NY: ACM Press, 144-152.
62. Kavzoğlu, T. ve Çölkesen, İ., Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi, Harita Dergisi, 144, 7 (2010) 73-82.
63. Özkan, Y., Veri Madenciliği Yöntemleri, 2. Basım, Papatya Yayıncılık Eğitim, İstanbul, 2013.
64. <http://www.coxdocs.org/lib/exe/fetch.php?w=300&tok=bc385a&media=perseus:user:activities:matrixprocessing:learning:knn.png>, 20 Kasım 2016.
65. Breiman, L., Random Forests, Machine Learning, 45, 1 (2001) 5-32.
66. <http://file.scirp.org/Html/6-9101686/f799e10c-50bd-48ec-9344-49d767083be5.jpg>, 20 Kasım 2016.
67. Han, J., Kamber, M. and Pei, J., Data Mining: Concepts and Techniques, Third Edition, Morgan Kaufmann Publishers, MA, 2012.
68. Akman, M., Veri Madenciliğine Genel Bakış ve Random Forests Yönteminin İncelenmesi: Sağlık Alanında Bir Uygulama, Yüksek Lisans Tezi, Ankara Üniversitesi, Sağlık Bilimleri Enstitüsü, 2010.

69. Chawla, N.V., Bowyer, K.W., Kegelmeyer, W.P. and Hall, L.O., SMOTE: Synthetic Minority Over-sampling Technique, Journal of Artificial Intelligence Research, 16, 1 (2002) 321-357.
70. Batista, G.E.A.P.A., Prati, R.C. and Monard, M.C., A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data, ACM SIGKDD Explorations Newsletter, 6, 1 (2004) 20-29.
71. Powers, D.M.W., Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness&Correlation, Journal of Machine Learning Technologies, 2, 1 (2011) 37-63.
72. UCI Machine Learning Repository, Wine quality data set, [Online], Available: <https://archive.ics.uci.edu/ml/datasets/Wine+Quality>.
73. Cortez, P., Cerdeira, A., Almeida, F., Matos, T. and Reis, J., Modelling Wine Preferences by Data Mining from Physicochemical Properties, Decision Support Systems, 47, 4 (2009) 547-553.

ÖZGEÇMİŞ

Yeşim ER, 06.09.1991 tarihinde Trabzon'da doğdu. İlköğretimini Aydınlıkevler İlköğretim Okulu'nda, ortaöğretimini Trabzon Anadolu Lisesi'nde 2009 yılında tamamlamıştır. 2013 yılında Karadeniz Teknik Üniversitesi Elektrik-Elektronik Mühendisliği Bölümünden mezun oldu. 2014 yılında Gümüşhane Üniversitesi Elektrik-Elektronik Mühendisliği Devreler ve Sistemler Anabilim Dalı'na ÖYP araştırma görevlisi olarak atandı. Aynı sene Orta Doğu Teknik Üniversitesi Yabancı Diller Yüksek Okulu tarafından gerçekleştirilen İngilizce Dil Eğitim Programı'na katılmış ve başarılı olmuştur. 2015 yılı Eylül ayından itibaren Karadeniz Teknik Üniversitesi Elektrik-Elektronik Mühendisliği Devreler ve Sistemler Anabilim Dalı'nda araştırma görevlisi olarak lisansüstü çalışmalarını sürdürmektedir. Yabancı dil olarak İngilizce bilmektedir.