

DOKUZ EYLÜL ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

**ÇAMAŞIR MAKİNELERİNDE FARKLI LEKE
GRUPLARI İÇİN YIKAMA PERFORMANSINA
ETKİ EDEN PARAMETRELERİN DENETİMLİ
ÖĞRENME ALGORİTMALARI İLE
TAHMİNLENMESİ**

Merih Şükrü AKGÜN

Ekim, 2023

İZMİR

**ÇAMAŞIR MAKİNELERİNDE FARKLI LEKE
GRUPLARI İÇİN YIKAMA PERFORMANSINA
ETKİ EDEN PARAMETRELERİN DENETİMLİ
ÖĞRENME ALGORİTMALARI İLE
TAHMİNLENMESİ**

**Dokuz Eylül Üniversitesi Fen Bilimleri Enstitüsü
Yüksek Lisans Tezi
İstatistik Anabilim Dalı, Veri Bilimi Programı**

Merih Şükrü AKGÜN

Ekim, 2023

İZMİR

YÜKSEK LİSANS TEZİ SINAV SONUÇ FORMU

MERİH ŞÜKRÜ AKGÜN tarafından **PROF.DR. NESLİHAN DEMİREL** yönetiminde hazırlanan “**ÇAMAŞIR MAKİNELERİNDE FARKLI LEKE GRUPLARI İÇİN YIKAMA PERFORMANSINA ETKİ EDEN PARAMETRELERİN DENETİMLİ ÖĞRENME ALGORİTMALARI İLE TAHMİNLENMESİ**” başlıklı tez tarafımızdan okunmuş, kapsamı ve niteliği açısından bir Yüksek Lisans tezi olarak kabul edilmiştir.

Prof.Dr. Neslihan DEMİREL

Danışman

Prof.Dr. Burcu HÜDAVERDİ

Jüri Üyesi

Dr.Öğr.Üyesi Özge ELMASTAŞ GÜLTEKİN

Jüri Üyesi

Prof. Dr. Okan FISTIKOĞLU

Müdür

Fen Bilimleri Enstitüsü

TEŞEKKÜR

Tez çalışmamın her aşamasında bilgi ve tecrübeleriyle beni yönlendiren desteklerini hiç esirgemeyen değerli hocam Prof. Dr. Neslihan Demirel'e, bu süreçteki tüm katkıları için hocam Prof. Dr. Burcu Hüdaverdi'ye, paylaştığı tecrübeleri ve makaleleriyle tez çalışmamda değerli katkılar sunan Prof. Dr. Rainer Stammeringer'a ve tez çalışması boyunca her zaman destek olan aileme ve arkadaşlarıma teşekkürlerimi sunarım.

Bu tez çalışmasına 2210-D Sanayiye Yönelik Yurt İçi Yüksek Lisans Burs Programı kapsamında maddi destek veren Türkiye Bilimsel ve Teknik Araştırma Kurumu (TÜBİTAK)'na, tezin uygulama bölümünde kullanılan verilerin temininde ve yorumlanmasında her zaman destek olan sanayi danışmanım Vestel Beyaz Eşya Ar-Ge Performans Geliştirme Grubu Başmühendisi Tolga Şen'e ve Vestel Beyaz Eşya Ar-Ge Teknik Müdürü Barkın Elmas'a çok teşekkür ederim.

Merih Şükrü AKGÜN

ÇAMAŞIR MAKİNELERİNDE FARKLI LEKE GRUPLARI İÇİN YIKAMA PERFORMANSINA ETKİ EDEN PARAMETRELERİN DENETİMLİ ÖĞRENME ALGORİTMALARI İLE TAHMİNLENMESİ

ÖZ

Ev tipi çamaşır makinelerinin yıkama performansının, Avrupa'daki eko tasarım gereklilikleri gibi belirli sınır değerlere uyması gerekmektedir. Çamaşır makinelerinde lekelerin yıkama performansları için belirlenen sınır değerleri EN 60456 standardına göre elde edilmelidir. Yıkama performansını etkileyen yıkama süresi, su seviyesi, deterjan miktarı, motor yoğunluğu, hacim başına yıkanan yük miktarı ve sıcaklık olarak belirlenen faktörlerin çeşitli düzeyleri için yıkamalar gerçekleştirilmektedir. Bu tez çalışmasında Vestel Beyaz Eşya firmasından sebum, karbon, kan, kakao ve kırmızı şarap lekesi olan kumaşlar üzerinde çeyrek yükle, yarım yükle ve tam yükle elde edilen yıkama performans değerleri temin edilmiştir. 5 leke için yıkama performans değerleri ile tüm lekelerin yıkama performans değeri toplamı ve 3 ayrı yük için toplam 18 veri kümesi elde edilmiştir. Elde edilen veri kümeleri çoklu doğrusal regresyon, regresyon ağacı, torbalama (bagging), rassal ormanlar regresyonu, XGBoost regresyonu, destek vektör regresyonu (doğrusal çekirdek ve polinom çekirdek) ve k-en yakın komşu regresyonu algoritmaları ile modellenmiştir. Tüm modellerin performanslarının karşılaştırılması için hata kareler ortalaması, kök hata kareler ortalaması ve ortalama mutlak hata metrikleri kullanılmıştır. 18 veri kümesinin 8'inde rassal ormanlar regresyonu, 7'sinde bagging ve 3'ünde XGBoost modeli en iyi model olarak tespit edilmiştir. Sonuç olarak ağaç tabanlı algoritmalar kullanan modeller olarak ön plana çıkmıştır.

Anahtar kelimeler: Çamaşır makinesi, yıkama performansı, denetimli makine öğrenmesi algoritmaları

ESTIMATION OF PARAMETERS AFFECTING WASHING PERFORMANCE OF DIFFERENT STAIN GROUPS IN WASHING MACHINES WITH SUPERVISED LEARNING ALGORITHMS

ABSTRACT

The washing performance of household washing machines must comply with certain limit values, such as eco-design requirements in Europe. The limit values determined for the washing performance of stains in washing machines must be obtained according to the EN 60456 standard. Washing is carried out at various levels of factors affecting washing performance, such as washing time, water level, detergent amount, motor intensity, amount of washed load per volume and temperature. In this thesis study, washing performance values obtained at quarter load, half load and full load on fabrics with sebum, carbon, blood, cocoa and red wine stains were obtained from Vestel company. The washing performance value for 5 stains and the sum of the washing performance value of all stains were obtained, and a total of 18 data sets were obtained for 3 separate loads. The resulting data sets were modeled with multiple linear regression, regression tree, bagging, random forest regression, XGBoost regression, support vector regression (linear kernel and polynomial kernel) and k-nearest neighbor regression algorithms. To compare the performances of all models, mean square error, root mean square error and mean absolute error metrics were used. Out of 18 datasets, random forests regression was found to be the best model in 8, bagging in 7 and XGBoost in 3. As a result, tree-based algorithms have come to the fore as the best models.

Keywords: Washing machine, washing performance, supervised machine learning algorithms

İÇİNDEKİLER

	Sayfa
YÜKSEK LİSANS TEZİ SINAV SONUÇ FORMU	ii
TEŞEKKÜR	iii
ÖZ	iv
ABSTRACT	v
İÇİNDEKİLER	vi
ŞEKİLLER LİSTESİ	ix
TABLolar LİSTESİ	xii
KISALTMALAR	xiv
BÖLÜM BİR GİRİŞ	1
BÖLÜM İKİ ÇAMAŞIR YIKAMA İŞLEMİ VE TEMEL UNSURLARI	3
2.1 Yıkama Süreci	3
2.2 Yıkama Sürecine Etki Eden Faktörler	4
2.2.1 Kimya	4
2.2.2 Süre	5
2.2.3 Sıcaklık	5
2.2.4 Mekanik	5
2.2.5 Su	6
BÖLÜM ÜÇ MAKİNE ÖĞRENİMİ	8
3.1 Denetimli Öğrenme	8
3.1.1 Çoklu Doğrusal Regresyon	9
3.1.2 Regresyon Ağacı	10
3.1.3 Torbalama	12
3.1.4 Rassal Ormanlar Regresyonu	13
3.1.5 Aşırı Gradyan Artırma	13
3.1.6 Destek Vektör Regresyonu	15
3.1.7 K-En Yakın Komşu	16
3.2 Model Performans Metrikleri	17

3.2.1 Hata Kareler Ortalaması	17
3.2.2 Kök Hata Kareler Ortalaması.....	17
3.2.3 Mutlak Hata Ortalaması.....	18
BÖLÜM DÖRT ANALİZ VE UYGULAMA	19
4.1 Veri Kümesinin İncelenmesi	19
4.2 Modellerin Eğitilmesi ve Performans Metrikleri.....	20
4.2.1 Sebum Lekesi Değişkeni.....	21
4.2.1.1 Çeyrek Yük	21
4.2.1.2 Yarım Yük.....	23
4.2.1.3 Tam Yük.....	25
4.2.2 Karbon Lekesi Değişkeni.....	27
4.2.2.1 Çeyrek Yük	27
4.2.2.2 Yarım Yük.....	29
4.2.2.3 Tam Yük.....	31
4.2.3 Kan Lekesi Değişkeni	33
4.2.3.1 Çeyrek Yük	33
4.2.3.2 Yarım Yük.....	35
4.2.3.3 Tam Yük.....	37
4.2.4 Kakao Lekesi Değişkeni	39
4.2.4.1 Çeyrek Yük	39
4.2.4.2 Yarım Yük.....	41
4.2.4.3 Tam Yük.....	43
4.2.5 Kırmızı Şarap Lekesi Değişkeni	45
4.2.5.1 Çeyrek Yük	45
4.2.5.2 Yarım Yük.....	47
4.2.5.3 Tam Yük.....	49
4.2.6 Toplam Değişkeni.....	51
4.2.6.1 Çeyrek Yük	51
4.2.6.2 Yarım Yük.....	53
4.2.6.3 Tam Yük.....	55

BÖLÜM BEŞ DEĞERLENDİRME VE SONUÇ.....	57
KAYNAKLAR.....	62



ŞEKİLLER LİSTESİ

Sayfa

Şekil 2.1	Sinner çevrimi (Lasic, 2014).	4
Şekil 4.1	a) Sebum değişkeni için çeyrek yük ile oluşturulan bagging modelinin hata kareler ortalamasının artışı grafiği, b) Sebum değişkeni için çeyrek yük ile oluşturulan bagging modelinin düğüm saflığının artışı grafiği	22
Şekil 4.2	a) Sebum değişkeni için yarım yük ile oluşturulan rassal ormanlar regresyonu modelinin hata kareler ortalamasının artışı grafiği, b) Sebum değişkeni için yarım yük ile oluşturulan rassal ormanlar regresyonu modelinin düğüm saflığının artışı grafiği	24
Şekil 4.3	a) Sebum değişkeni için tam yük ile oluşturulan bagging modelinin hata kareler ortalamasının artışı grafiği, b) Sebum değişkeni için tam yük ile oluşturulan bagging modelinin düğüm saflığının artışı grafiği	26
Şekil 4.4	a) Karbon değişkeni için çeyrek yük ile oluşturulan rassal ormanlar regresyonu modelinin hata kareler ortalamasının artışı grafiği, b) Karbon değişkeni için çeyrek yük ile oluşturulan rassal ormanlar regresyonu modelinin düğüm saflığının artışı grafiği	28
Şekil 4.5	a) Karbon değişkeni için yarım yük ile oluşturulan bagging modelinin hata kareler ortalamasının artışı grafiği, b) Karbon değişkeni için yarım yük ile oluşturulan bagging modelinin düğüm saflığının artışı grafiği	30
Şekil 4.6	a) Karbon değişkeni için tam yük ile oluşturulan XGBoost modelinin kazanç grafiği, b) Karbon değişkeni için tam yük ile oluşturulan XGBoost modelinin kapsama grafiği, c) Karbon değişkeni için tam yük ile oluşturulan XGBoost modelinin sıklık grafiği	31

Şekil 4.7	a) Kan değişkeni için çeyrek yük ile oluşturulan rassal ormanlar regresyonu modelinin hata kareler ortalamasının artışı grafiği, b) Kan değişkeni için çeyrek yük ile oluşturulan rassal ormanlar regresyonu modelinin düğüm saflığının artışı grafiği	33
Şekil 4.8	a) Kan değişkeni için yarım yük ile oluşturulan XGBoost modelinin kazanç grafiği, b) Kan değişkeni için yarım yük ile oluşturulan XGBoost modelinin kapsama grafiği, c) Kan değişkeni için yarım yük ile oluşturulan XGBoost modelinin sıklık grafiği	35
Şekil 4.9	a) Kan değişkeni için tam yük ile oluşturulan rassal ormanlar regresyonu modelinin hata kareler ortalamasının artışı grafiği, b) Kan değişkeni için tam yük ile oluşturulan rassal ormanlar modelinin düğüm saflığının artışı grafiği.	37
Şekil 4.10	a) Kakao değişkeni için çeyrek yük ile oluşturulan rassal ormanlar regresyonu modelinin hata kareler ortalamasının artışı grafiği, b) Kakao değişkeni için çeyrek yük ile oluşturulan rassal ormanlar regresyonu modelinin düğüm saflığının artışı grafiği	39
Şekil 4.11	a) Kakao değişkeni için yarım yük ile oluşturulan rassal ormanlar modelinin hata kareler ortalamasının artışı grafiği, b) Kakao değişkeni için yarım yük ile oluşturulan rassal ormanlar regresyonu modelinin düğüm saflığının artışı grafiği	41
Şekil 4.12	a) Kakao değişkeni için tam yük ile oluşturulan bagging modelinin hata kareler ortalamasının artışı grafiği, b) Kakao değişkeni için tam yük ile oluşturulan bagging modelinin düğüm saflığının artışı grafiği	43
Şekil 4.13	a) Kırmızı şarap değişkeni için çeyrek yük ile oluşturulan XGBoost modelinin kazanç grafiği, b) Kırmızı şarap değişkeni için çeyrek yük ile oluşturulan XGBoost modelinin kapsama grafiği, c) Kırmızı şarap değişkeni için çeyrek yük ile oluşturulan XGBoost modelinin sıklık grafiği	45

Şekil 4.14	a) Kırmızı şarap değişkeni için yarım yük ile oluşturulan rassal ormanlar modelinin hata kareler ortalamasının artışı grafiği, b) Kırmızı şarap değişkeni için yarım yük ile oluşturulan rassal ormanlar regresyonu modelinin düğüm saflığının artışı grafiği	47
Şekil 4.15	a) Kırmızı şarap değişkeni için tam yük ile oluşturulan rassal ormanlar regresyonu modelinin hata kareler ortalamasının artışı grafiği, b) Kırmızı şarap değişkeni için tam yük ile oluşturulan rassal ormanlar regresyonu modelinin düğüm saflığının artışı grafiği	49
Şekil 4.16	a) Toplam değişkeni için çeyrek yük ile oluşturulan bagging modelinin hata kareler ortalamasının artışı grafiği, b) Toplam değişkeni için çeyrek yük ile oluşturulan bagging modelinin düğüm saflığının artışı grafiği	51
Şekil 4.17	a) Toplam değişkeni için yarım yük ile oluşturulan bagging modelinin hata kareler ortalamasının artışı grafiği, b) Toplam değişkeni için yarım yük ile oluşturulan bagging modelinin düğüm saflığının artışı grafiği.	53
Şekil 4.18	a) Toplam değişkeni için tam yük ile oluşturulan bagging modelinin hata kareler ortalamasının artışı grafiği, b) Toplam değişkeni için tam yük ile oluşturulan bagging modelinin düğüm saflığının artışı grafiği	55

TABLÖLAR LİSTESİ

Sayfa

Tablo 4.1	Sebum değişkeni için çeyrek yük ile oluşturulan modellere ait performans metrikleri	21
Tablo 4.2	Sebum değişkeni için yarım yük ile oluşturulan modellere ait performans metrikleri	23
Tablo 4.3	Sebum değişkeni için tam yük ile oluşturulan modellere ait performans metrikleri	25
Tablo 4.4	Karbon değişkeni için çeyrek yük ile oluşturulan modellere ait performans metrikleri	27
Tablo 4.5	Karbon değişkeni için yarım yük ile oluşturulan modellere ait performans metrikleri	29
Tablo 4.6	Karbon değişkeni için tam yük ile oluşturulan modellere ait performans metrikleri.	31
Tablo 4.7	Kan değişkeni için çeyrek yük ile oluşturulan modellere ait performans metrikleri	32
Tablo 4.8	Kan değişkeni için yarım yük ile oluşturulan modellere ait performans metrikleri	34
Tablo 4.9	Kan değişkeni için tam yük ile oluşturulan modellere ait performans metrikleri	36
Tablo 4.10	Kakao değişkeni için çeyrek yük ile oluşturulan modellere ait performans metrikleri	38
Tablo 4.11	Kakao değişkeni için yarım yük ile oluşturulan modellere ait performans metrikleri	40
Tablo 4.12	Kakao değişkeni için tam yük ile oluşturulan modellere ait performans metrikleri	42

Tablo 4.13 Kırmızı şarap değişkeni için çeyrek yük ile oluşturulan modellere ait performans metrikleri	44
Tablo 4.14 Kırmızı şarap değişkeni için yarım yük ile oluşturulan modellere ait performans metrikleri	46
Tablo 4.15 Kırmızı şarap değişkeni için tam yük ile oluşturulan modellere ait performans metrikleri	48
Tablo 4.16 Toplam değişkeni için çeyrek yük ile oluşturulan modellere ait performans metrikleri	50
Tablo 4.17 Toplam değişkeni için yarım yük ile oluşturulan modellere ait performans metrikleri	52
Tablo 4.18 Toplam değişkeni için tam yük ile oluşturulan modellere ait performans metrikleri	54
Tablo 4.19 Tüm modellerin RMSE değerleri.....	60

KISALTMALAR

MLR : Çoklu Doğrusal Regresyon (Multiple Linear Regression)

SSE : Hata Kareler Toplamı (Sum of Squares Error)

XGBoost : Aşırı Gradyan Artırma (Extreme Gradient Boosting)

SVM : Destek Vektör Makinesi (Support Vector Machine)

SVR : Destek Vektör Regresyonu (Support Vector Regression)

K-NN : K-En Yakın Komşu (K-Nearest Neighbor)

MSE : Hata Kareler Ortalaması (Mean Squared Error)

RMSE : Kök Hata Kareler Ortalaması (Root Mean Squared Error)

MAE : Mutlak Hata Ortalaması (Mean Absolute Error)

BÖLÜM BİR

GİRİŞ

Temizliğin, sağlıklı bir yaşam sürdürebilmek için hayati bir önemi vardır. İnsan sağlığının korunmasında kişisel hijyen ve çevre temizliğinin rolü büyüktür. Bununla birlikte eşyaların ve kıyafetlerin temizliğini sağlamak da sağlıklı bir yaşam sürdürmek için gereklidir (Özyay, 2016). Kumaş yüzeyinde virüslerin veya Gram negatif bakterilerin birkaç saat, Staphylococcus Aureus, Clostridium Difficile veya mantar sporları gibi türlerin, günler veya haftalar boyunca kalabildiğine dair göstergeler vardır. Özellikle Staphylococcus Aureus enfeksiyonlarının yayılmasında, giysiler ve ev çamaşırları yoluyla bulaşmasının önemli bir rolü bulunmaktadır (Bloomfield vd., 2011). Bu nedenle çamaşır temizliği sağlıklı yaşamın ana unsurlarından biridir.

Tarih boyunca insanlar çamaşır temizliğini sağlamak için çeşitli yöntemler geliştirmiştir. Antik Mısır'da, giysiler sodyum karbonat tuzlarıyla ovularak yıkanmıştır. Antik Romalılar ise üre içinde kaynatmayı ve dövmeyi tercih etmişlerdir. Günümüze kadar devam eden yıkama işlemi, çamaşır makinesinin icadı ile daha da gelişerek günlük hayatımızın ayrılmaz bir parçası haline gelmiştir (Özyay, 2016).

İlk mekanik çamaşır makineleri 1850'lerde kullanılmaya başlanmıştır. 20. yüzyıl boyunca hijyen, insanlar için giderek daha önemli hale gelmiştir. İlk elektrikli çamaşır makinesi 1908'de Amerika Birleşik Devletleri'nde üretilmiş ve 1950'lerde çamaşır makinesi teknolojisi, ısıtıcıların ve yüksek hızlı sıkma özelliklerinin eklenmesiyle daha da geliştirilmiştir. 1960'lı yıllarda dikey eksenli bir makinede tek tamburda yıkama, durulama ve sıkma işlemi yapılabilmekteydi. Daha sonra yatay eksenli makineler geliştirilmiştir. 20. yüzyılın sonunda daha az su ve enerji kullanan, yıkama işlemini daha düşük sıcaklıklarda geliştirebilen elektromekanik kontrollü çamaşır makineleri ortaya çıkmıştır (Efe, 2015).

Çamaşır makinesi, yıkama işlemini gerçekleştiren tambur ve kazan, tamburu döndüren motor, su çekmek için kullanılan vana ve hortum bağlantıları gibi birçok parçadan oluşan elektrikli bir cihazdır (Özyay, 2016). Çamaşır makinesinin temel çalışma prensibi, deterjanın kirli çamaşırların etrafında türbülanslı bir akış oluşturmaktır (Ali vd., 2019).

Çamaşır makinesinde yıkama işlemi; su alımı, ana yıkama, durulama ve sıkma adımlarının sırasıyla uygulanması ile gerçekleştirilmektedir (Özyay, 2016). Yıkama işlemine etki eden faktörleri ise ilk olarak yıkama döngüsünü dört parametrenin bir fonksiyonu olacak şekilde Sinner (1960) tarafından tanımlamıştır. Bu parametreler; sıcaklık, süre, mekanik ve kimyadır. Parametrelerden birinin yoğunluğu azalırsa, diğer parametrelerden biri veya birkaçı artırılarak kayıp telafi edilmelidir. Bununla birlikte, su parametresi de ıslak temizlemede yüksek bir öneme sahiptir. Stamminger (2010), tüm faktörlerin birleştirici unsurunu temsil eden suyu beşinci parametre olarak eklemiştir. Literatürde bu parametreleri içeren çalışmalar mevcut olup (Efe, 2015; Lambert, 2016; Özyay, 2016), ancak çamaşır makinesi üreticisinin test verilerini makine öğrenmesi algoritmaları ile değerlendirildiği bir çalışmaya rastlanmamıştır. Bu tez çalışmasında, Vestel Beyaz Eşya firmasının ürettikleri çamaşır makinelerinin farklı leke çeşitlerine (sebum, karbon, kan, kakao ve kırmızı şarap) göre yıkama performanslarının tahminlenmesi ve yıkama performanslarının artırılması amaçlanmıştır. Ayrıca tüm leke çeşitlerinin yıkama performans değerlerinin toplanması ile elde edilen toplam yıkama performansının tahminlenmesi ve yıkama performansının artırılması da amaçlanmıştır. Farklı çamaşır makineleri ile kurulacak makine öğrenmesi algoritmaları sayesinde, test edilmesi düşünülen bir makinenin yıkama performansı tahmin edilebilecektir. Bu durum üreticiyi testlerden önce parametrelerin ayarlanmasında daha hassas kararlar almasını sağlayacaktır. Alınan bu kararlar neticesinde daha kesin sonuçlarla testler yapılarak, gerçekleştirilen test sayısı düşürülecektir. Vestel Beyaz Eşya firmasının test verimliliğini arttırmak, sanayinin bu testler için harcadığı iş gücü ve zamandan, ayrıca su ve elektrik tüketiminden tasarruf edilmesini sağlayacaktır.

Çalışmanın ikinci bölümünde çamaşır yıkama işlemi ve temel unsurları hakkında bilgiler verilmiştir. Üçüncü bölümde kullanılan makine öğrenimi, denetimli öğrenme algoritmaları ve model performans metriklerinin detayları anlatılmıştır. Dördüncü bölümde veri kümesi tanıtılmış, modeller oluşturulmuş ve model performans metrikleri incelenmiştir. Beşinci bölümde çalışmanın sonuçlarından bahsedilmiştir.

BÖLÜM İKİ

ÇAMAŞIR YIKAMA İŞLEMİ VE TEMEL UNSURLARI

Eski zamanlarda insanlar, kıyafetlerindeki lekeleri dere kenarında kayalara vurarak temizlemişlerdir. Ancak bu zahmetli iş, teknolojiadaki gelişmeler sayesinde çamaşır makinesindeki bir düğmeye basarak yapılabilecek kadar basit bir hale gelmiştir. Çamaşır makinesi lüksten çok bir ihtiyaçtır (Zuheros, 2011).

Çamaşır makinesinin çalışma yöntemi şu şekilde açıklanabilir: Çamaşırlar makinenin tamburuna yerleştirilir ve daha sonra belirli bir program seçilerek yıkama işlemi başlatılır. Bu aşamadan sonra makine su giriş vanası yardımıyla su almaya başlar. Deterjanın dağılması için alınan su, önce deterjan kutusu üzerinde bulunan duşlama sistemi yardımıyla deterjan çekmecesinin içerisindeki deterjanı süpürür, ardından kazan bağlantı hortumundan tambura deterjanlı su verilir. Tambur, çamaşırları ıslatmak ve yıkama işlemini gerçekleştirmek için belirli bir hızda döner. Deterjanlı su belli bir süre ve sıcaklıkta yıkanmak üzere kazana alındıktan sonra, yıkama suyu tahliye pompası tarafından boşaltılır ve sıkma işlemine geçilir. Bu adım sırasında tambur, çamaşırdan olabildiğince fazla su çekmek için merkezkaç kuvveti ilkesini kullanarak yüksek hızda döner. Tambur içindeki çamaşırlar sürekli olarak tambur çeperine itildiği için devamlı olarak su atılır. Sıkma işlemi bittiğinde tambur kontrollü bir şekilde durur ve çamaşırlar kolayca makineden çıkartılabilir (Özyay, 2016).

2.1 Yıkama Süreci

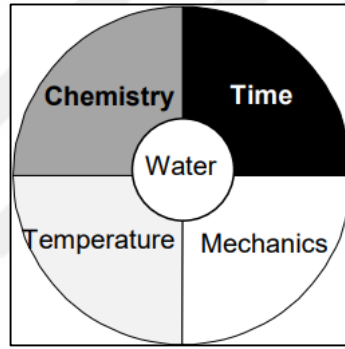
Çamaşır makinesi temel olarak yıkama işlemini gerçekleştiren bir cihazdır. Bunun için öncelikle yıkama işlemi ve bunu etkileyen faktörler göz önünde bulundurulmalıdır. Çamaşır yıkama, sulu çözeltide çeşitli fiziksel ve kimyasal etkileşimlerin meydana geldiği karmaşık bir süreçtir. Geniş anlamda yıkama, tekstil ürünlerinin yüzeyindeki suda çözünen maddelerin ve kirlerin, su veya sulu bir deterjan solüsyonu vasıtasıyla uzaklaştırılması olarak tanımlanabilir (Lasic, 2014).

Yıkama işlemi dört ana döngüye ayrılabilir. İlk döngü, temizliğin çamaşırların su ve temizlik maddeleri ile döndürülerek yapıldığı yıkama döngüsüdür. İkinci aşama, kullanılan tüm deterjan ve temizlik maddelerini su ile temizlemek için yapılan

durulama döngüsüdür. Tüm döngülerden önemli ölçüde daha kısa süren üçüncü döngü, dağıtım döngüsüdür. Bu programda, son döngüde oluşabilecek dengesiz yükü azaltmak için çamaşırlar susuz orta hızda tamburlanır. Döndürerek kurutma ise son döngüdür. Santrifüj kuvvetlerinden yararlanır, yıkama ve durulama devrelerinde çamaşırların emdiği suyu çıkarır (Karakaş, 2011).

2.2 Yıkama Sürecine Etki Eden Faktörler

Sinner (1960) tarafından geliştirilen Sinner çevrimi adlı çamaşır temizleme modeli, çamaşır yıkama performansını etkileyen dört faktörü ortaya koymaktadır. Bu faktörler sıcaklık, süre, mekanik ve kimyadır. Stamminger (2010), tüm faktörlerin birleştirici unsurunu temsil eden suyu beşinci parametre olarak eklemiştir. Şekil 2.1’de Sinner çevrimi verilmiştir.



Şekil 2.1 Sinner çevrimi (Lasic, 2014)

Sinner çevriminin dört faktöründen her biri, diğer üç faktörle belirli bir dereceye kadar ikame edilebilir ve elde edilen yıkama performansı aynı kalır. Örneğin, sıcaklıktaki düşüş, diğer faktörlerin, yani kimya, süre veya mekanik gibi bir veya daha fazla faktörün artırılmasıyla kısmen telafi edilebilir (Lasic, 2014).

2.2.1 Kimya

Kimyasal etki, yıkama işleminin önemli bir parçasıdır. Su ve deterjanın birleşmesi ile oluşan solüsyondaki deterjan oranı arttıkça veya azaldıkça kimyasal etki artmakta veya azalmaktadır. Bir deterjanın en önemli işlevi, kiri ve lekeleri uzaklaştırmak, kiri yıkama solüsyonunda askıda tutmak ve tekrar tekstil yüzeyine yapışmasını engellemektir (Özyay, 2016).

Deterjan konsantrasyonu arttıkça yıkama performansı da artmaktadır. Ancak deterjan konsantrasyonu çok yüksek olduğu durumlarda, yüksek köpük üretimi mekanik etkinin azalmasına neden olduğundan yıkama performansını olumsuz etkileyecektir (Lasic, 2014).

2.2.2 Süre

Tekstilden kiri çıkarmak için, yıkama sürecinin kimya ve mekanik faktörleri, belirli bir yıkama sıcaklığında belirli bir süre boyunca yük ile etkileşime girmelidir. Yıkama işleminin süresi, deterjanın etkili olacağı süreyi belirlemektedir. Daha uzun temizleme süreleri, kirin uzaklaştırılmasını artıracak ve böylece temizleme performansını iyileştirecektir (Lasic, 2014).

Yıkama işleminin başında soğuk giriş suyunun ısıtılması gerekmektedir. Isınma süresi, önceden ayarlanmış sıcaklığa ulaşmak için gereken süredir. Isınma süresi çamaşır miktarına ve dolayısıyla su miktarına, suyun giriş sıcaklığına ve çamaşır makinesinin ısıtma kapasitesine bağlıdır (Smulders vd., 2002; Lasic, 2014).

2.2.3 Sıcaklık

Çamaşır yıkama işleminin bir diğer önemli faktörü sıcaklıktır. Sıcaklık arttıkça leke çıkarma etkinliği de artmaktadır. Bunun nedeni, kimyasal reaksiyonları hızlandırması, yağ lekelerini kolayca çözmesi ve kirin kumaşa bağlanma kuvvetini zayıflatmasıdır (Smulders vd., 2002).

Maksimum yıkama sıcaklığı, çamaşırın özelliklerine göre değişmektedir. Yüksek sıcaklıklar kirin çözünmesini arttırırken, yıkama solüsyonunun kir bağlama kapasitesi azalarak kirin kumaşa yeniden yapışmasına neden olabilmektedir. Ayrıca, sentetik liflerin örgüsü sıcaklık yükseldikçe bozulabilmektedir (Lasic, 2014).

2.2.4 Mekanik

Mekanik etki, elle yıkama durumunda tekstillerin sürtülmesi, dövülmesi, gerilmesi ve sıkılması ile yapılmaktadır. (Smulders vd., 2002). Çamaşır makinelerinde mekanik etki ise tambur çapı, su seviyesi, tambur devir sayısı, tamburda bulunan kanatların konumu ve sayısı, yük miktarı ve yıkama süresi gibi faktörlere bağlıdır (Lasic, 2014).

Çamaşıra verilen mekanik etkinin derecesi, kuru çamaşır ağırlığının yıkama çözeltisi hacmine olan oranı değiştirilerek ayarlanabilmektedir. Düşük seviye yıkama çözeltisi ve nispeten hızlı tambur dönüşü, çamaşır üzerinde yüksek mekanik etkiye sebep olmaktadır. Bunun aksine, yüksek seviye yıkama çözeltisi ve daha yavaş tambur dönüşü ise düşük mekanik etkiye sebep olmaktadır (Lasic, 2014).

Yüksek mekanik hareket kir çıkarma performansında artışa sebep olmakla birlikte güçlü tekstil hasarlarının oluşumuna da sebep olması muhtemeldir (Özyay, 2016).

2.2.5 Su

Su, Stammering (2010) tarafından ıslak temizlikte suyun temel unsur olarak önemini göstermek için beşinci faktör olarak sunulmuştur. Su, deterjanın çözülmesinde, deterjan bileşenlerinin ve ısıнын çamaşıra taşınmasında, hidrodinamik akış direnci yoluyla kir parçacıkları üzerinde mekanik kuvvet sağlanmasında, durulama işlemi sırasında salınan kirlerin liflerden uzaklaştırılmasında ve deterjan çözeltisini seyreltmek için etken madde görevi görmektedir (Özyay, 2016).

Çamaşır makinelerinin ana işlevlerinden biri olan yıkama, su kullanımını gerektirir. Su miktarı, çamaşır makinesinin performansını belirleyen ana faktörlerden biridir. Suyun yıkama işlemine en önemli katkısı çözücü özelliğidir. Bu özellik deterjanlar, çözünebilir tuzlar ve kirler için bir taşıma ortamı görevi görmektedir. Ayrıca yıkama işleminden alınan su miktarı, enerji ve yıkama performansını belirlemektedir. Her yıkama programı için tekstilin cinsine ve yük miktarına bağlı olarak istenen performans sağlayacak belirli bir su miktarı vardır. Yıkama adımındaki su miktarı belirlenen miktardan az olduğu durumda çamaşırlar yeterince ıslanmamakta ve suda çözünmüş temizlik kimyasalları (örneğin toz veya sıvı deterjanlar) çamaşırların kirli veya lekeli bölgelerine nüfuz edememektedir. Yıkama döngüsündeki su miktarı belirtilen miktardan fazla olduğu durumda, temizlik kimyasallarının konsantrasyonu azalmakta ve belirtilen yıkama sıcaklığına ulaşamamaktadır. Bu da yıkama performansının düşmesine neden olmaktadır (Özyay, 2016).

Su temel olarak yüzey ve yeraltı akiferleri olmak üzere iki kaynaktan sağlanmaktadır. Yağmur suyu, nehirlere, göllere ve yeraltı su kaynaklarına doğru akarken mineral maddeleri çözer. Çözünmüş madde kalsiyum, magnezyum ve demir

gibi çok çeşitli minerallerden oluşur. Kalsiyum ve magnezyum iyonları (sülfat, klorür, karbonat ve bikarbonat olarak bulunur) suyun sert olmasına neden olur. Çamaşır suyundeki sertlik sorun oluşturur. Çünkü buna neden olan mineraller sabun ve deterjanların temizleme etkisini engeller (Abeliotis vd., 2015).

Tipik olarak, çok sert suyla gerçekleştirilen yıkamalarda düşük yıkama performansı elde edilir. Bununla birlikte, suyun sertliği çamaşır makinesi bileşenleri olan ısıtıcı, metal parçalar vb. zarar verebilir (Özyay, 2016).



BÖLÜM ÜÇ

MAKİNE ÖĞRENİMİ

Makine öğrenimi, bilgi teknolojisi, istatistik, olasılık, yapay zeka, psikoloji, nörobiyoloji ve birçok disiplinde sunulan geniş bir alanı temsil etmektedir (Nasteski, 2017). Bilgisayarların önceden var olan verileri otomatik olarak okumasını ve yorumlanmasını kolaylaştıran bir teknolojidir. İstatistik ve bilgisayar bilimlerinin bir bileşimi olan makine öğrenimi terimi, 1959'da Arthur Samuel tarafından ortaya atılmıştır (Bansal vd., 2022).

Makine öğrenimi, bilgisayarlara insan beynini taklit etmeyi öğretmenin ötesine geçerek, istatistik alanı ile öğrenme süreçlerinin temel istatistiksel hesaplama teorilerini üreten geniş bir disiplin haline gelmiştir (Nasteski, 2017).

Makine öğrenimi, matematiksel modeller oluşturmak için algoritmalarından yararlanmaktadır. Bu algoritmalar geçmiş verileri kullanarak yeni veriler için tahminler yapar. Metin tespiti, nefret söylemi tespiti, öneri sistemi ve yüz tespiti gibi farklı birçok alanda kullanılmaktadır (Bansal vd., 2022).

Algoritmanın istenen sonucuna bağlı olarak, makine öğrenimi algoritmaları denetimli öğrenme ve denetimsiz öğrenme olarak iki başlık altında incelenebilir. Denetimli öğrenmede, algoritmalar girdileri istenen çıktılarla eşleyen bir fonksiyon üretmektedir. Denetimsiz öğrenme ise, bir dizi girdiyi modellemektedir. Burada etiketli örnekler mevcut değildir. Denetimli öğrenme ve denetimsiz öğrenmenin yanında, yarı denetimli öğrenme, takviyeli öğrenme ve öğrenmeyi öğrenme gibi öğrenme türleri de dahil edilebilir (Nasteski, 2017).

3.1 Denetimli Öğrenme

Denetimli öğrenmede amaç görülmemiş gözlemler için yanıt değerleri tahmin etmek veya bağımlı ve bağımsız değişkenler arasındaki ilişki hakkında çıkarımlarda bulunmak için model oluşturmaktır (James vd., 2013).

Değişkenler nicel veya nitel olarak tanımlanabilir. Nicel değişkenler sayısal değerler alırken, nitel değişkenler sınıf veya kategoriye temsil eden değerler almaktadır. Bağımlı değişkenin nicel olduğu problemlere regresyon problemleri, nitel

olduğu problemlere ise sınıflandırma problemleri denir (James vd., 2013; Hastie vd., 2009).

Bu çalışmada makine öğrenimi algoritmalarından çoklu doğrusal regresyon, regresyon ağacı, torbalama, rassal ormanlar regresyonu, aşırı gradyan artırma, destek vektör regresyonu ve k-en yakın komşu algoritmaları ele alınmıştır.

3.1.1 Çoklu Doğrusal Regresyon

Çoklu doğrusal regresyon (Multiple Linear Regression - MLR), tek bir bağımlı değişken ile birden fazla bağımsız değişken arasındaki ilişkiyi modellemek için kullanılan istatistiksel bir yöntemdir.

Bir MLR modeli, bağımlı değişken Y ve p adet bağımsız değişken $X_1, X_2, \dots, X_p$ ve n adet gözlemden oluşmaktadır. Y ve $X_1, X_2, \dots, X_p$ arasındaki ilişki doğrusal bir model olarak şu şekilde formüle edilir:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (3.1)$$

burada $\beta_0, \beta_1, \dots, \beta_p$ regresyon katsayıları olarak adlandırılan parametrelerdir ve ε ise rassal hata terimidir (Chatterjee ve Hadi, 2013).

X_j , j 'inci bağımsız değişkeni temsil eder ve diğer tüm X değerleri sabit tutulduğunda, β_j , X_j 'deki bir birimlik artışın Y üzerindeki ortalama etkisi olarak yorumlanmaktadır (James vd., 2013).

$\beta_0, \beta_1, \dots, \beta_p$ parametrelerinin tahmin edilmesi gereklidir. En küçük kareler yöntemi kullanılarak, yani hata kareler toplamını (sum of squares error - SSE) en aza indirilerek yapılmaktadır. Hatalar

$$\varepsilon_i = y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_p x_{ip}, \quad i = 1, 2, \dots, n \quad (3.2)$$

şeklinde temsil edilir. SSE ise

$$S(\beta_0, \beta_1, \dots, \beta_p) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_p x_{ip})^2 \quad (3.3)$$

şeklinde temsil edilir. $S(\beta_0, \beta_1, \dots, \beta_p)$ değerini en aza indirmek üzere $\beta_0, \beta_1, \dots, \beta_p$ 'ye göre türevleri alınarak sıfıra eşitlenmesinden elde edilen normal denklemlerin çözümüyle parametre tahminlerine ulaşılr. $\hat{\beta}_0$ sabit terim, $\hat{\beta}_j$, X_j kestiricisinin regresyon katsayısının tahmini olarak adlandırılır (Chatterjee ve Hadi, 2013).

Tahmin edilen regresyon katsayılarını kullanarak en küçük kareler regresyon denklemi şu şekilde yazılır:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_p X_p \quad (3.4)$$

$\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$, değerleri en küçük kareler regresyon katsayısı tahminleridir. Her gözlem için hesaplanır:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip}, \quad i = 1, 2, \dots, n \quad (3.5)$$

$\hat{y}_i$ tahmin değerlerini gösterir. Gerçek değer ile tahmin değeri arasındaki farka artık denir ve şu şekilde gösterilir:

$$e_i = y_i - \hat{y}_i, \quad i = 1, 2, \dots, n \quad (3.6)$$

Aşağıdaki formülle:

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2 \quad (3.7)$$

SSE'yi minimum yapan $\beta_0, \beta_1, \dots, \beta_p$ değerleri elde edilmiş olur (Chatterjee ve Hadi, 2013).

3.1.2 Regresyon Ağacı

Regresyon ağacı yöntemi 1984 yılında, L. Breiman, J. Friedman, R. Olshen ve C. Stone tarafından, ikili karar ağaçlarının oluşturulmasını anlatan Classification and Regression Trees (CART) adlı kitapta yayınlanmıştır (Breiman, 2017).

Regresyon ağacı parametrik olmayan bir yöntemdir. Parametrik olmayan yöntemler, parametrelerin verilen örnekten tahmin edildiği belirli bir modelin varsayıldığı parametrik yaklaşımlardan farklıdır. Genellikle uygun bir parametrik

model bilinmeyebilir, parametrelerin tahmini çok maliyeti olabilir ya da parametrelerin tahmini optimal olmayabilir. Bu gibi durumlarda, parametrik olmayan yaklaşımlar tercih edilmektedir (Alpaydin, 1997).

Regresyon ağacı, eğitim verilerindeki N gözlemin her biri için p girdi ve bir yanıtta oluşur. Yani $i = 1, 2, \dots, N$ için (x_i, y_i) , $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$. Algoritmanın bölme değişkenlerine ve bölme noktalarına ayrıca ağacın nasıl bir topolojiye (şekle) sahip olması gerektiğine otomatik olarak karar vermesi gerekir. Öncelikle M adet $R_1, R_2, \dots, R_M$ bölgeleri elde edilir ve yanıt her bir bölgede sabit c_m olarak modellenir (Hastie vd., 2009).

$$f(x) = \sum_{m=1}^M c_m I(x \in R_m) \quad (3.8)$$

Kriter olarak $\sum (y_i - f(x_i))^2$ toplamının en küçük olması kabul edildiğinde, en iyi $\hat{c}_m$, R_m bölgesindeki y_i değerlerinin ortalamasıdır. Yani, her bir bölgede bulunan y_i değerlerinin ortalaması:

$$\hat{c}_m = \text{ortalama}(y_i | x_i \in R_m) \quad (3.9)$$

en iyi tahmin olan $\hat{c}_m$ değerini verir (Hastie vd., 2009).

Minimum kareler toplamı cinsinden en iyi ikili bölümü bulmak genellikle hesaplama açısından mümkün değildir. Tüm verilerle başlayarak, bir bölme değişkeni j ve bölme noktası s belirlenir ve yarı düzlemler çifti tanımlanır:

$$R_1(j, s) = \{X | X_j \leq s\} \text{ ve } R_2(j, s) = \{X | X_j > s\} \quad (3.10)$$

Sonrasında aşağıdaki denklemi çözecek olan bölme değişkeni j ve bölme noktası s aranır:

$$\min_{j,s} [\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2] \quad (3.11)$$

Herhangi bir j ve s seçimi için, iç minimizasyon şu şekilde çözülür:

$$\hat{c}_1 = \text{ortalama}(y_i | x_i \in R_1(j, s)) \text{ ve } \hat{c}_2 = \text{ortalama}(y_i | x_i \in R_2(j, s)) \quad (3.12)$$

Her bir bölme değişkeni için, bölme noktası s 'nin belirlenmesi çok hızlı bir şekilde yapılabilir ve bu nedenle tüm girdileri tarayarak en iyi çiftin (j, s) belirlenmesi mümkündür (Hastie vd., 2009).

3.1.3 Torbalama

Torbalama (bagging), Leo Breiman, tarafından 1996 yılında önerilmiştir. Regresyon ve sınıflandırma problemleri için kullanılan parametrik olmayan bir yöntemdir (Breiman, 1996).

Karar ağaçları yüksek varyanstan muzdariptir. Eğitim verileri rastgele iki kısma ayrıldığında ve her iki kısma regresyon ağacı uygulandığında, birbirinden oldukça farklı sonuçlar elde edilebilmektedir. Diğer bir yandan düşük varyanslı bir prosedürde ise farklı veri kümelerine tekrar tekrar uygulandığında da sonuçlar benzer olmaktadır. Örneklem büyüklüğünün açıklayıcı değişkenlerin sayısına oranı büyük olduğunda, doğrusal regresyon düşük varyansa sahip olma eğiliminde olacaktır. Torbalama istatistiksel bir öğrenme yönteminin varyansını azaltmaya yönelik bir yöntemdir (James vd., 2013).

Bu yaklaşımda orijinal örneklemde aynı genişlikte yerine koyarak tekrarlı B rassal örneklem elde edilir. Bu örneklemde bootstrap örneklemeleri denir. Daha sonra $\hat{f}^{*b}(x)$ elde etmek için yöntem b 'inci bootstrap eğitim seti üzerinde eğitilir ve son olarak tüm tahminlerin ortalaması alınır:

$$\hat{f}_{bag}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^{*b}(x) \quad (3.13)$$

Buna torbalama denir (James vd., 2013).

Regresyon ağaçlarına torbalama uygulamak için, B bootstrap eğitim setini kullanarak B regresyon ağacı oluşturulur. Daha sonra elde edilen tahminlerin ortalaması alınır. B ağacın ortalamasının alınması varyansı azaltır (James vd., 2013).

3.1.4 Rassal Ormanlar Regresyonu

Rassal ormanlar regresyon ve sınıflama problemleri için 2001 yılında Leo Breiman tarafından önerilmiş parametrik olmayan bir yöntemdir (Breiman, 2001).

Rassal ormanlar regresyonu yönteminde, torbalama yönteminde olduğu gibi, bootstrap eğitim örnekleri üzerinde bir dizi karar ağacı oluşturulur. Ancak bu karar ağaçları oluşturulurken, regresyon problemlerinde p değişkenden rastgele seçilen m değişkenden ağaçlar kurulur ve $m \approx p/3$ olarak belirlenir (James vd., 2013).

Veri kümesinde çok güçlü bir tahmin edici ile orta derecede güçlü tahmin edicilerin olduğu bir durumda, torbalanmış ağaçların toplanmasında ağaçların çoğu veya tamamı için bu güçlü tahmin ediciyi en üstteki bölmede kullanacaktır. Böylece torbalanmış ağaçların tümü birbirine oldukça benzer elde edilecektir. Dolayısıyla torbalanmış ağaçlardan gelen tahminler yüksek düzeyde korelasyonlu olacaktır. Yüksek düzeyde korelasyonlu birçok niceliğin ortalamasının alınması, birçok korelasyonsuz niceliğin ortalamasının alınması kadar varyansta büyük bir azalmaya yol açmamaktadır (James vd., 2013).

Rassal ormanlar, her bir bölünmeyi değişkenlerin yalnızca bir alt kümesini dikkate alarak bu sorunun üstesinden gelmektedir. Bu nedenle, bölünmelerin ortalama $(p - m)/p$ 'si güçlü tahmin ediciyi dikkate almayacak ve böylece diğer tahmin edicilerin daha fazla şansı olacaktır. Bu sayede ortaya çıkan ağaçların ortalaması daha az değişkenle daha güvenilir hale gelir (James vd., 2013).

3.1.5 Aşırı Gradyan Artırma

Aşırı gradyan artırma (extreme gradient boosting – XGBoost), gradyan artırma yöntemi temelli bir algoritmadır (Friedman, 2001; Friedman, 2002).

XGBoost'un amaç fonksiyonu, modelin sapmasını ve aşırı uyumu önlemek için normal terimi temsil eden iki farklı bölümden oluşur (Chen ve Guestrin, 2016). $D = \{(x_i, y_i)\}$, n örnek ve m özellik içeren bir veri kümesini temsil eder; burada tahmin değişkeni, k temel modelden oluşan bir eklemeli modeldir. Örnek tahmin sonuçları aşağıdaki gibidir:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in \varphi \quad (3.14)$$

$$\varphi = \{f(x) = w_s(x)\} (s: R^m \rightarrow T, w_s \in R^T) \quad (3.15)$$

Burada;

$\hat{y}_i$, i . değere ait tahmin değerini,

x_i , i . gözlem değerini,

$f_k(x_i)$, i . gözleme ait tahmin skorunu,

φ , s 'nin ağaç yapısı parametreleri olan regresyon ağacı kümesini,

$f(x)$ yaprakların ağırlığını,

w yaprak sayısını

temsil eder (Qiu vd., 2021).

XGBoost'un amaç fonksiyonu, geleneksel kayıp fonksiyonunu ve model karmaşıklığını içerir. Algoritmanın operasyonel verimliliğini değerlendirmek için kullanılabilir. Denklem 3.16'da ilk terim geleneksel kayıp fonksiyonunu, ikinci terim ise modelin karmaşıklığını temsil etmektedir:

$$Obj = \sum_{i=1}^m l(y_i, \hat{y}_i^{(t-1)} + f_i(x_i)) + \Omega(f_k) \quad (3.16)$$

$$\Omega(f_k) = \gamma T + \frac{1}{2\lambda w^2} \quad (3.17)$$

Bu formüllerin her ikisinde de i veri kümesindeki örnek sayısını, m ise k . ağaca aktarılan toplam veri miktarını temsil etmektedir. γ ve λ ise ağacın karmaşıklığını ayarlamak için kullanılır. Düzenleme terimi, son öğrenme ağırlığını yumuşatabilir ve aşırı uyumu önleyebilir (Qiu vd., 2021).

3.1.6 Destek Vektör Regresyonu

Destek vektör makinesi (Support Vector Machine - SVM) büyük ölçüde AT&T Bell Laboratuvarlarında Vapnik ve çalışma arkadaşları tarafından geliştirilmiştir (Boser vd., 1992; Cortes ve Vapnik, 1995; Guyon vd., 1992; Schilkop vd., 1995; Schölkopf vd., 1996; Vapnik vd., 1996). SVM, regresyon ve sınıflandırma problemleri için kullanılan parametrik olmayan bir yöntemdir.

SVM uyarlanabilir marj tabanlı kayıp fonksiyonlarını kullanır ve eğitim verilerini daha yüksek boyutlu özellik uzayına yansıtır. İyi genelleme kabiliyetine sahip en iyi karar kuralını bulmaktadır. Nihai destek vektör regresyonu (Support Vector Regression - SVR) modeli şu biçimi alır:

$$Y = f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) k(x_i, x) + b = \sum_{i=1}^N w_i k(x_i, x) + b \quad (3.18)$$

burada $k(x_i, x)$ çekirdek fonksiyonu, α_i ve α_i^* Lagrangian çarpanları ve b ayarlanabilir bir sabittir (Cheng vd., 2017).

Çekirdek, iki gözlemin benzerliğini ölçen bir fonksiyondur. Bu çalışmada doğrusal ve polinom çekirdek fonksiyonları kullanılmıştır. Doğrusal çekirdek fonksiyonu:

$$k(x_i, x) = x_i x \quad (3.19)$$

şeklinde ifade edilir. Doğrusal çekirdek esas olarak Pearson korelasyonunu kullanarak bir çift gözlemin benzerliğini ölçer. Polinom çekirdek fonksiyonu ise:

$$k(x_i, x) = (x_i x + 1)^d \quad (3.20)$$

şeklinde ifade edilir. d dereceli bir polinom çekirdeği olarak bilinir; burada d , pozitif bir tam sayıdır. SVM algoritmasında standart doğrusal çekirdek yerine $d > 1$ olan bir çekirdeğin kullanılması, çok daha esnek bir karar sınırına yol açar. Bu orijinal özellik uzayı yerine d dereceli polinomları içeren daha yüksek boyutlu bir uzayda bir SVM elde edilmesini sağlar (Cheng vd., 2017).

3.1.7 K-En Yakın Komşu

Evelyn Fix ve Joseph Lawson Hodges Jr. tarafından geliştirilen k-en yakın komşu (K-Nearst Neighbor - K-NN) yöntemi, 1951 yılında parametrik olmayan bir sınıflandırma yöntemi olan diskriminant analizi olarak tanıtılmıştır. Ancak makale resmi olarak yayınlanmamıştır (Taunk vd., 2019).

K-NN algoritması bir denetimli öğrenme tekniğidir. Sadeliği ile bilinmektedir. Hem regresyon hem de sınıflandırma problemleri için kullanılmaktadır (Bansal vd., 2022; Taunk vd., 2019). Bu sebeple veri kümesi üstünde herhangi bir varsayımda bulunmamaktadır (Taunk vd., 2019).

K-NN algoritması, tembel öğrenme yöntemi olarak da bilinir. Çünkü eğitim veri seti ile öğrenme süreci, yeni veri setinin tahmini için talep olana kadar gerçekleşmemektedir (Aha, 1997).

K-NN algoritmasındaki “K”, yeni veri noktasının komşu sayısını ifade etmektedir. Algoritmanın en önemli süreci, en uygun k değerine karar vermektir (Bansal vd., 2022). k için hangi değer en etkili sonucu verdiğini görmek için algoritmayı farklı değerlerle çalıştırmak gerekmektedir. Regresyonda sürekli değer tahmini, k adet en yakın komşularının ortalamasıdır (Taunk vd., 2019).

Hedef gözlemlerin tahmini, en yakın k gözleminin ağırlıklı ortalamaları şu şekilde hesaplanmaktadır:

$$\hat{y}_j = \frac{\sum_{i=1}^k w_{ij} y_i}{\sum_{i=1}^k w_{ij}} \quad (3.21)$$

Burada, k en yakın komşu sayısıdır. y_i , i . komşunun gözlenen yanıt değeri, $\hat{y}_j$, j . hedef gözlemin tahmin değeridir. w_{ij} , j . hedef değeri için i . gözlem ağırlığıdır. Ağırlık aşağıdaki şekilde hesaplanır:

$$w_{ij} = \frac{\left(\frac{1}{1+d_{ij}}\right)^{pm}}{\sum_{i=1}^k \left(\frac{1}{1+d_{ij}}\right)^{pm}} \quad (3.22)$$

burada d_{ij} i ve j arasındaki benzerlik mesafesi ve pm ağırlıklandırma parametresidir ($i \neq j$). d_{ij} şu şekilde tanımlanmıştır:

$$d_{ij} = \sum_{l=1}^L c_l |x_{il} - x_{jl}| \quad (3.23)$$

burada L bağımsız değişkenlerin sayısı ve c ilgili ağırlıklarıdır (Haara ve Kangas, 2012).

3.2 Model Performans Metrikleri

Belirli bir veri kümesi üzerinde istatistiksel bir öğrenme yönteminin performansını değerlendirmek için tahminlerinin gerçekte gözlemlenen verilerle ne kadar iyi eşleştiğini ölçmek için kullanılan bazı metrikler vardır. Bir başka ifade ile, belirli bir gözlem için tahmin edilen yanıt değerinin, o gözlem için gerçek yanıt değerine ne ölçüde yakın olduğunu nicelleştirmek gerekir (James vd., 2013). Bu çalışmada belirlenen metrikler hata kareler ortalaması, kök hata kareler ortalaması ve mutlak hata ortalamasıdır.

3.2.1 Hata Kareler Ortalaması

Regresyon ayarında, en sık kullanılan ölçülerden biri hata kareler ortalaması (mean squared error - MSE)'dir: MSE:

$$MSE = \frac{1}{n} \sum_{i=1}^n e_i^2 \quad (3.24)$$

şeklinde ifade edilir (Willmott ve Matsuura, 2005).

Tahmin edilen değerler, gerçek değerlere ne kadar yakınsa MSE değerleri de o kadar küçük olacaktır. Sıfıra yakın MSE değeri, tahminleyicilerin daha iyi performans gösterdiği anlamına gelir.

3.2.2 Kök Hata Kareler Ortalaması

Kök hata kareler ortalaması (root mean squared error - RMSE), MSE'nin kareköküne eşittir. Denklem

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \quad (3.25)$$

şeklinde gösterilir (Willmott ve Matsuura, 2005).

RMSE değeri 0 ile ∞ aralığında değerler alabilir. 0'a yakın değer alması daha iyi performans sergilediğini gösterir.

3.2.3 Mutlak Hata Ortalaması

Mutlak hata ortalaması (mean absolute error - MAE), her bir gerçek değer ile o değere karşılık gelen tahmin değeri arasındaki farkların mutlak değerlerinin ortalamasıdır ve

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (3.26)$$

şeklinde ifade edilir (Willmott ve Matsuura, 2005).

MAE sıfırdan büyük değerler alır. 0'a yakın değer alması ilgili algoritmanın daha iyi performans sergilediğini gösterir.

BÖLÜM DÖRT

ANALİZ VE UYGULAMA

4.1 Veri Kümesinin İncelenmesi

Vestel veri kümesi, Vestel Beyaz Eşya firmasının ürettikleri çamaşır makinelerinin yıkama performans değerlerinin temin edilmesiyle oluşturulmuştur. Yıkama performans değerleri, firmaya ait 39 farklı çamaşır makinesiyle elde edilmiştir. Yıkama standartlarının sağlanması açısından sebum, karbon, kakao, kan ve kırmızı şarap için ayrı ayrı 5 cm x 5 cm boyutunda lekeli kumaşlar hazır olarak yurt dışından temin edilmekte ve bu 5 lekenin bulunduğu kumaş ile 1 adet lekesiz kumaş 70 cm x 100 cm boyutundaki pamuklu kumaşa dikilmektedir. Çamaşır makinelerinin yıkama performansını değerlendirmek için, bu kumaşların yıkama sonrası spektrofometre ile ölçülen reflektans değerleri kullanılmaktadır.

Vestel veri kümesindeki bağımsız değişkenler:

- Hacim başına yıkanan yük miktarı; çamaşır makinelerinin tambur hacminin yıkanan yük miktarına oranını belirtir.
- Deterjan dozajı; yıkama için kullanılan deterjan miktarını belirtir. Testlerde kullanılan deterjan miktarı, EN 60456 standardında belirlenmiştir.
- Kg başına ana yıkama suyu tüketimi; yıkama boyunca harcanan suyun yük miktarına oranını belirtir.
- Yükteki sıcaklık; yıkama sırasında çamaşırın ulaştığı sıcaklığın celsius ölçeğindeki değeridir.
- Motor yoğunluğu; tambur dönüşünü sağlayan motorun, yıkama boyunca geçen sürenin yüzde kaçında çalıştığını belirtir.
- Ana yıkama süresi; yıkama boyunca geçen süreyi belirtir.

Bağımlı değişkenler:

- Sebum lekesi, karbon lekesi, kan lekesi, kakao lekesi ve kırmızı şarap lekesi; lekeli kumaşların yıkama sonrasında her leke için spektrofotometre ile ölçülen reflektans değerleridir.
- Toplam; tüm lekeler için elde edilen reflektans değerlerinin toplamı.

Vestel veri kümesi 400 gözlem ve 12 değişkenden oluşmakta ve veri kümesinde eksik değer bulunmamaktadır.

4.2 Modellerin Eğitilmesi ve Performans Metrikleri

Yürütülen bu çalışmada, yıkama performans değerleri çeyrek yük, yarım yük ve tam yük ayrı ayrı ele alınmıştır. Testler sonucunda ulaşılan yıkama performans değerlerinin %30'u çeyrek yük, %40'ı yarım yük ve geriye kalan %30'u ise tam yük içindir. Böylece 5 leke için yıkama performans değeri ile tüm lekelerin yıkama performans değeri toplamı ve 3 ayrı yük için toplam 18 veri kümesi elde edilmiştir. Tüm veri kümeleri MLR, regresyon ağacı, bagging, rassal ormanlar regresyonu, XGBoost regresyonu, destek vektör regresyonu (doğrusal çekirdek ve polinom çekirdek) ve K-NN regresyonu algoritmaları ile modellenmiştir. Tüm modellerin performanslarının karşılaştırılması için RMSE, MSE ve MAE metrikleri kullanılmıştır.

Modeller oluşturulmadan önce verilerin %80'i modellerin eğitimi için, kalan %20'lik bölümü ise performanslarının değerlendirilmesi için ayrılmıştır. Modeller R istatistiksel programlama diliyle oluşturulmuştur.

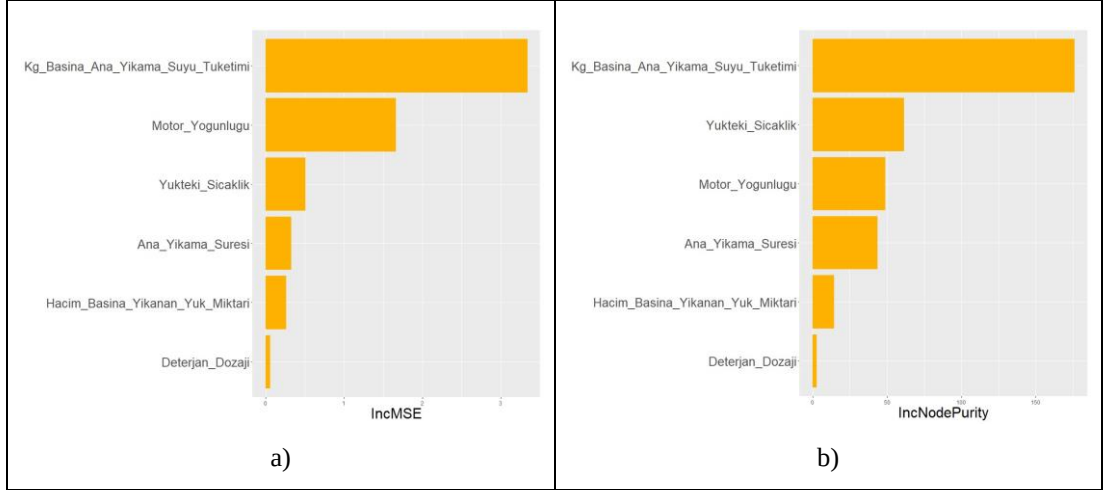
4.2.1 Sebum Lekesi Değişkeni

4.2.1.1 Çeyrek Yük

Tablo 4.1 Sebum lekesi değişkeni için çeyrek yük ile oluşturulan modellere ait performans metrikleri

Modeller	RMSE	MSE	MAE
Bagging	1.453570	2.112866	1.203551
Rassal Ormanlar Regresyonu	1.480113	2.190734	1.216147
XGBoost	1.514631	2.294107	1.314601
K-NN	1.687806	2.848688	1.225833
SVR Polinom	1.721777	2.964515	1.230827
SVR Doğrusal	1.880455	3.536110	1.427054
Regresyon Ağacı	1.940420	3.765229	1.611568
MLR	1.965805	3.864388	1.575193

Sebum lekesi değişkeni için çeyrek yük ile oluşturulan Tablo.4.1'deki modeller incelendiğinde, ağaç tabanlı modellerin iyi performans verdiği görülmektedir. RMSE ve MSE değerlerinde ilk üç sırayı bagging, rassal ormanlar regresyonu ve XGBoost modelleri, MAE değerinde ilk iki sırayı bagging ve rassal ormanlar regresyonu modeli almıştır. Değişken öneminin incelenmesi için seçilen model, tüm performans metriklerinde en iyi değerlere sahip olan bagging modelidir.



Şekil 4.1 a) Sebum lekesi değişkeni için çeyrek yük ile oluşturulan bagging modelinin hata kareler ortalamasının artışı grafiği, b) Sebum lekesi değişkeni için çeyrek yük ile oluşturulan bagging modelinin düğüm saflığının artışı grafiği

Şekil 4.1.a incelendiğinde, özellikle kg başına ana yıkama suyu tüketimi değişkeninin modelde önemli olduğu görülmektedir. Ardından motor yoğunluğu değişkeni ön plana çıkmaktadır. Bu değişkenleri sırasıyla yükteki sıcaklık, ana yıkama süresi, hacim başına yıkanan yük miktarı ve deterjan dozajı değişkenleri takip etmektedir.

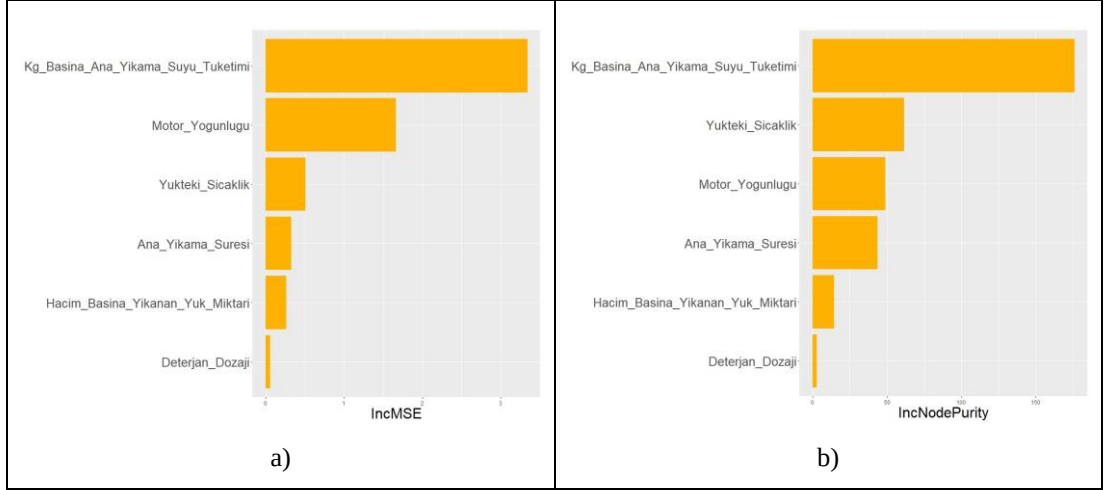
Şekil 4.1.b incelendiğinde, kg başına ana yıkama suyu tüketimi değişkeninin modelde önemli olduğu görülmektedir. Sırasıyla yükteki sıcaklık, motor yoğunluğu, ana yıkama süresi, hacim başına yıkanan yük miktarı ve deterjan dozajı şeklinde sıralanmaktadır.

4.2.1.2 Yarım Yük

Tablo 4.2 Sebum lekesi değişkeni için yarım yük ile oluşturulan modellere ait performans metrikleri

Modeller	RMSE	MSE	MAE
Rassal Ormanlar Regresyonu	0.9243660	0.8544525	0.7613979
Bagging	0.9449956	0.8930167	0.7630957
XGBoost	1.0241559	1.0488953	0.8785742
Regresyon Ağacı	1.1331901	1.2841198	1.0150641
K-NN	1.1470638	1.3157553	1.0010846
SVR Doğrusal	1.1866672	1.4081790	0.9937804
MLR	1.2616531	1.5917684	1.0759808
SVR Polinom	1.2762552	1.6288274	1.1381542

Sebum lekesi değişkeni için yarım yük ile elde edilen ve Tablo 4.2’de gösterilen modellerde, ağaç tabanlı modellerin performansı ön plana çıkmaktadır. RMSE ve MSE değerleri incelendiğinde ilk dört sırada rassal ormanlar regresyonu, bagging, XGBoost ve regresyon ağacı, MAE değeri için ise ilk üç sırada rassal ormanlar regresyonu, bagging ve XGBoost modelleri yer almaktadır. Tüm performans metriklerinde birinci olan rassal ormanlar regresyonu modeli, değişken öneminin incelenmesi için seçilmiştir.



Şekil 4.2 a) Sebum lekesi değişkeni için yarım yük ile oluşturulan rassal ormanlar regresyonu modelinin hata kareler ortalamasının artışı grafiği, b) Sebum lekesi değişkeni için yarım yük ile oluşturulan rassal ormanlar regresyonu modelinin düğüm saflığının artışı grafiği

Şekil 4.2.a incelendiğinde, kg başına ana yıkama suyu tüketimi değişkeninin modelde önemli olduğu görülmektedir. Ayrıca motor yoğunluğu değişkeni de ön plana çıkmaktadır. Bu değişkenleri takip eden sırayla yükteki sıcaklık, ana yıkama süresi, hacim başına yıkanan yük miktarı ve deterjan dozajı değişkenleri takip etmektedir.

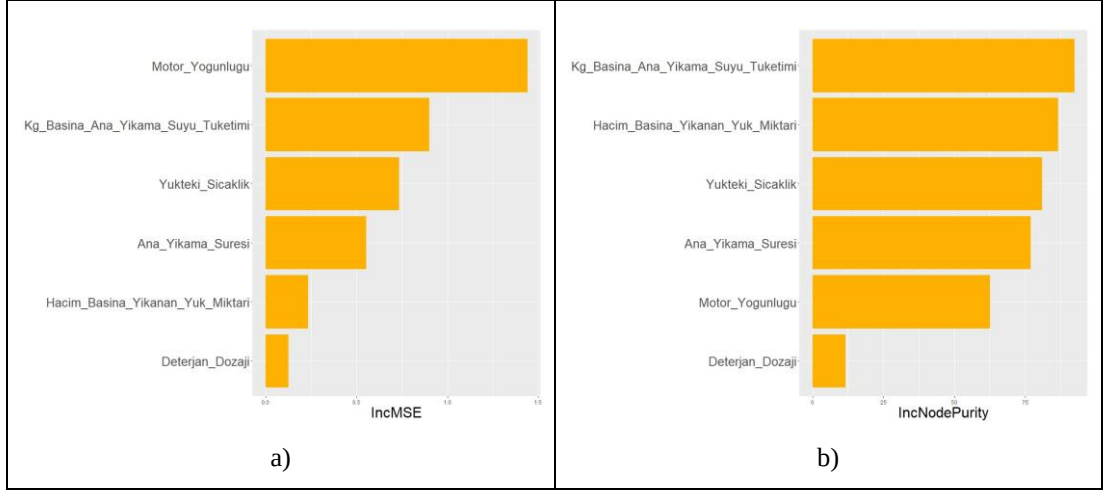
Şekil 4.2.b incelendiğinde, tekrar kg başına ana yıkama suyu tüketimi değişkeninin model için önemli olduğu görülmektedir. Kg başına ana yıkama suyu tüketimi değişkenini yükteki sıcaklık, motor yoğunluğu ve ana yıkama süresi değişkenleri takip etmiştir. Hacim başına yıkanan yük miktarı ve deterjan dozajı değişkenlerinin ise diğer değişkenlere kıyasla modelde öneminin daha az olduğu görülmektedir.

4.2.1.3 Tam Yk

Tablo 4.3 Sebum lekesi deęiřkeni iin tam yk ile oluřturulan modellere ait performans metrikleri

Modeller	RMSE	MSE	MAE
Bagging	1.624298	2.638343	1.369724
Rassal Ormanlar Regresyonu	1.646797	2.711942	1.389940
XGBoost	1.734781	3.009464	1.430697
Regresyon Aęacı	2.441221	5.959561	1.808097
MLR	2.533736	6.419816	2.022104
SVR Polinom	2.610989	6.817264	1.984840
K-NN	2.663582	7.094669	2.002299
SVR Doęrusal	2.988453	8.930854	2.150534

Sebum lekesi deęiřkeni iin tam yk ile oluřturulan modellerde, aęa tabanlı modellerin performanslarının dięer modellerin nnde olduęu Tablo 4.3’de grlmektedir. Tm performans metriklerinde en iyi deęerlere sahip olan model bagging modelidir. Daha sonra rassal ormanlar regresyonu modeli, ardından XGBoost modeli ve son olarak regresyon aęacı modeli gelmektedir. Deęiřken neminin incelenmesi iin tm performans metriklerinde birinci olan bagging modeli seilmiřtir.



Şekil 4.3 a) Sebum lekesi değişkeni için tam yük ile oluşturulan bagging modelinin hata kareler ortalamasının artışı grafiği, b) Sebum lekesi değişkeni için tam yük ile oluşturulan bagging modelinin düğüm saflığının artışı grafiği

Şekil 4.3.a incelendiğinde, motor yoğunluğu değişkeninin modelde önem sırasında ilk sırayı almıştır. Kalan değişkenler kg başına ana yıkama suyu tüketimi, yükteki sıcaklık, ana yıkama süresi, hacim başına yıkanan yük miktarı ve deterjan dozajı şeklinde sıralanmaktadır.

Şekil 4.3.b incelendiğinde, özellikle kg başına ana yıkama suyu tüketimi, hacim başına yıkanan yük miktarı, yükteki sıcaklık ve ana yıkama süresi değişkenlerinin modelde önemli olduğu görülmektedir. Ayrıca motor yoğunluğu değişkeni de ön plana çıkmaktadır. Deterjan dozajı değişkeni ise önem sırasında son sırada yer almıştır.

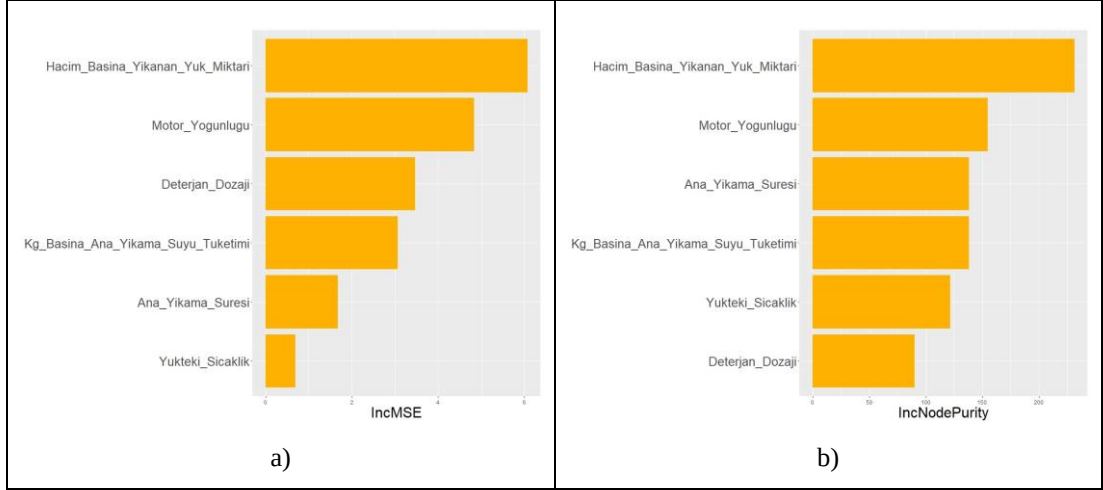
4.2.2 Karbon Lekesi Değişkeni

4.2.2.1 Çeyrek Yük

Tablo 4.4 Karbon lekesi değişkeni için çeyrek yük ile oluşturulan modellere ait performans metrikleri

Modeller	RMSE	MSE	MAE
Rassal Ormanlar Regresyonu	1.992571	3.970340	1.230724
Bagging	2.057325	4.232586	1.228137
K-NN	2.172169	4.718318	1.672500
SVR Doğrusal	2.200959	4.844220	1.709957
SVR Polinom	2.282116	5.208052	1.396377
XGBoost	2.427185	5.891228	1.485895
Regresyon Ağacı	2.632427	6.929670	1.764753
MLR	2.777585	7.714977	2.314309

Karbon lekesi değişkeni için çeyrek yük ile oluşturulan Tablo 4.4'deki sekiz model incelendiğinde, en iyi performansı gösteren iki model ağaç tabanlı modellerdir. RMSE ve MSE değerlerinde en iyi model rassal ormanlar regresyonu modeli olmuştur. Arkasından bagging modeli gelmektedir. MAE değerinde ise en iyi model bagging modeli olmuştur. Arkasından rassal ormanlar regresyonu modeli gelmektedir. Rassal ormanlar regresyonu modeli, değişken öneminin incelenmesi için seçilen modeldir.



Şekil 4.4 a) Karbon lekesi değişkeni için çeyrek yük ile oluşturulan rassal ormanlar regresyonu modelinin hata kareler ortalamasının artışı grafiği, b) Karbon lekesi değişkeni için çeyrek yük ile oluşturulan rassal ormanlar regresyonu modelinin düğüm saflığının artışı grafiği

Şekil 4.4.a incelendiğinde, hacim başına yıkanan yük miktarı değişkeni ön plana çıkmaktadır. Modelde önem sırası incelendiğinde diğer değişkenler motor yoğunluğu, deterjan dozajı, kg başına ana yıkama suyu tüketimi, ana yıkama süresi ve yükteki sıcaklık şeklinde sıralanmaktadır.

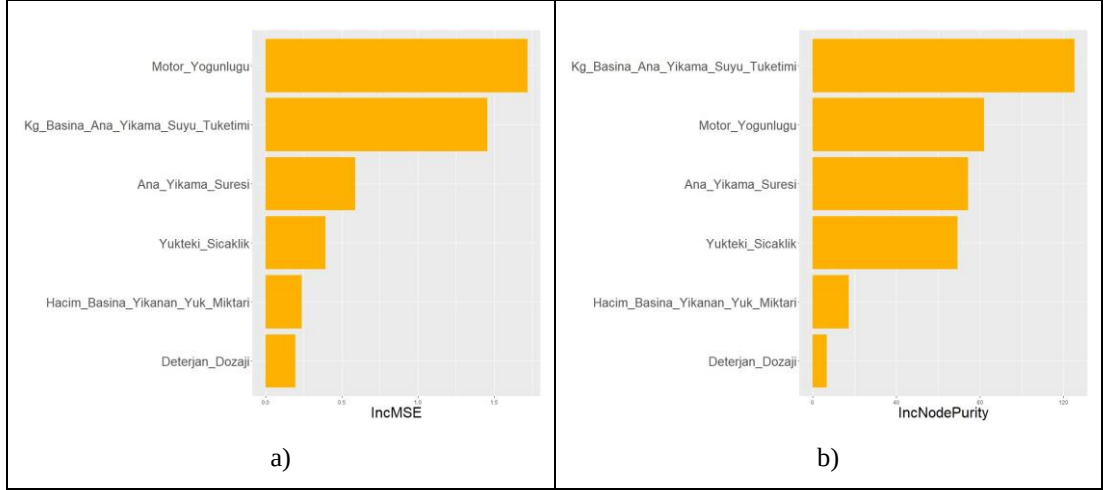
Şekil 4.4.b içinde önem sırasında ilk iki sırayı alan değişkenlerde benzer bir durum geçerlidir. Hacim başına yıkanan yük miktarı ve motor yoğunluğu değişkenleri modelde ön plana çıkmıştır. Diğer değişkenlerin sırası ana yıkama süresi, kg başına ana yıkama suyu tüketimi, yükteki sıcaklık ve deterjan dozajı şeklindedir.

4.2.2.2 Yarım Yük

Tablo 4.5 Karbon lekesi değişkeni için yarım yük ile oluşturulan modellere ait performans metrikleri

Modeller	RMSE	MSE	MAE
Bagging	0.9378708	0.8796016	0.7487787
XGBoost	0.9478575	0.8984339	0.7761901
Rassal Ormanlar Regresyonu	0.9557084	0.9133785	0.7692179
Regresyon Ağacı	1.1339386	1.2858167	0.8278336
K-NN	1.3555060	1.8373966	0.9841562
SVR Polinom	1.3708078	1.8791141	1.0610638
SVR Doğrusal	1.4106619	1.9899671	1.1242075
MLR	1.5437958	2.3833054	1.2172913

Tablo 4.5’de karbon lekesi değişkeni için yarım yük ile oluşturulan modellerin performans metrikleri incelendiğinde ağaç tabanlı modeller diğer modellerin önünde yer almıştır. Tüm performans metriklerinde bagging, XGBoost, rassal ormanlar regresyonu ve regresyon ağacı modelleri ilk dört sırayı almıştır. Değişken öneminin incelenmesi için seçilen model, tüm performans metriklerinde en iyi değerlere sahip olan bagging modelidir.



Şekil 4.5 a) Karbon lekesi değişkeni için yarım yük ile oluşturulan bagging modelinin hata kareler ortalamasının artışı grafiği, b) Karbon lekesi değişkeni için yarım yük ile oluşturulan bagging modelinin düğüm saflığının artışı grafiği

Şekil 4.5.a incelendiğinde özellikle motor yoğunluğu ve kg başına ana yıkama suyu tüketimi değişkenlerinin modelde önemli oldukları görülmektedir. Ana yıkama süresi ve yükteki sıcaklık değişkenleri takip etmiş, hacim başına yıkanan yük miktarı ve deterjan dozajı değişkenleri ise son iki sırayı almıştır.

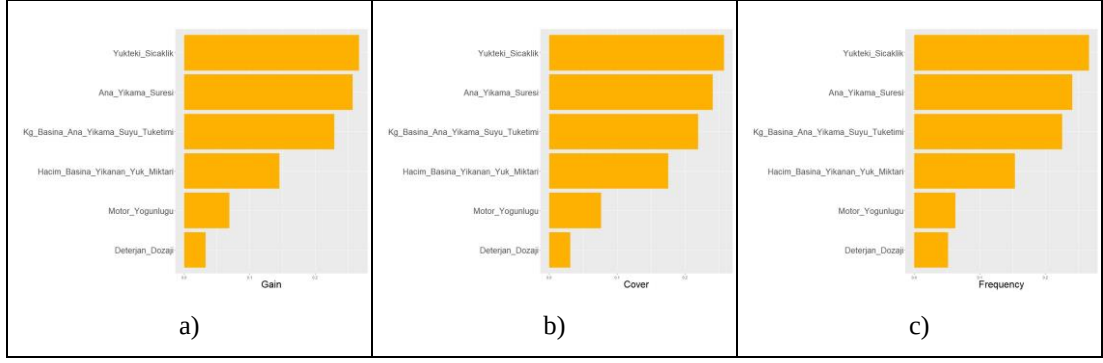
Şekil 4.5.b incelendiğinde ise modeldeki önem sırasında kg başına ana yıkama suyu tüketimi değişkeni ilk sırayı almıştır. Daha sonra motor yoğunluğu, ana yıkama süresi ve yükteki sıcaklık değişkenleri model için önem sırasında ön plana çıkmıştır. Hacim başına yıkanan yük miktarı ve deterjan dozajı değişkenleri önceki grafikte olduğu gibi tekrar son iki sırayı almıştır.

4.2.2.3 Tam Yük

Tablo 4.6 Karbon lekeli değişkeni için tam yük ile oluşturulan modellere ait performans metrikleri

Modeller	RMSE	MSE	MAE
XGBoost	1.517015	2.301335	1.309499
Bagging	1.569974	2.464817	1.251226
Rassal Ormanlar Regresyonu	1.627998	2.650376	1.298118
Regresyon Ağacı	2.464229	6.072427	2.001346
SVR Doğrusal	2.507685	6.288485	1.867300
SVR Polinom	2.631023	6.922283	1.942114
K-NN	2.692794	7.251142	1.962880
MLR	2.715035	7.371415	2.016073

Karbon lekeli değişkeni için tam yük ile oluşturulan Tablo 4.6'deki modeller incelendiğinde, ağaç tabanlı modeller dikkat çekmektedir. RMSE ve MSE değerleri en iyi, MAE değeri üçüncü en iyi model olan XGBoost modeli ön plana çıkmıştır. Bagging, rassal ormanlar regresyonu ve regresyon ağacı modelleri en iyi modeller arasında üst sıralarda yer almıştır. XGBoost modeli, değişken öneminin incelenmesi için seçilmiştir.



Şekil 4.6 a) Karbon lekesi değişkeni için tam yük ile oluşturulan XGBoost modelinin kazanç grafiği, b) Karbon lekesi değişkeni için tam yük ile oluşturulan XGBoost modelinin kapsama grafiği, c) Karbon lekesi değişkeni için tam yük ile oluşturulan XGBoost modelinin sıklık grafiği

Şekil 4.6.a, Şekil 4.6.b ve Şekil 4.6.c incelendiğinde yükteki sıcaklık, ana yıkama süresi ve kg başına ana yıkama suyu tüketimi değişkenleri en önemli değişkenler olarak görülmektedir. Bu değişkenleri hacim başına yıkanan yük miktarı, motor yoğunluğu ve deterjan dozajı değişkenleri takip etmektedir.

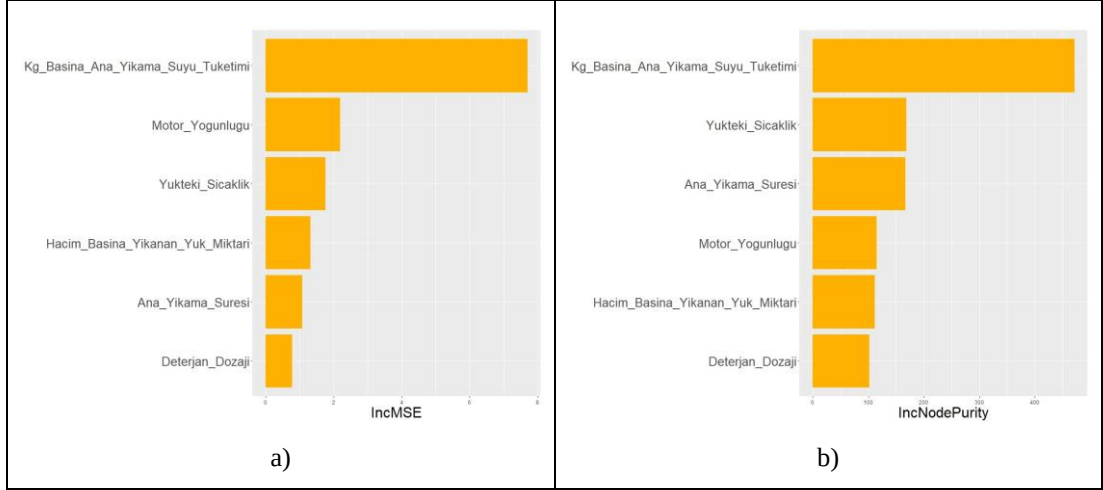
4.2.3 Kan Lekesi Değişkeni

4.2.3.1 Çeyrek Yük

Tablo 4.7 Kan lekesi değişkeni için çeyrek yük ile oluşturulan modellere ait performans metrikleri

Modeller	RMSE	MSE	MAE
Rassal Ormanlar Regresyonu	2.188376	4.788990	1.756077
Bagging	2.206045	4.866634	1.751524
XGBoost	2.312994	5.349941	1.819962
SVR Polinom	2.462067	6.061773	1.868672
SVR Doğrusal	2.544880	6.476412	1.940502
Regresyon Ağacı	2.879889	8.293760	2.314763
MLR	2.996442	8.978667	2.499658
K-NN	3.067271	9.408153	2.089375

Kan lekesi değişkeni ile çeyrek yük için oluşturulan Tablo 4.7'deki modeller incelendiğinde, tüm performans metriklerinde ilk üç sırayı ağaç tabanlı model almıştır. RMSE ve MSE değerlerinde en iyi performansı, MAE değerinde ikinci en iyi performansı gösteren rassal ormanlar regresyonu modeli ön plana çıkmaktadır. Bagging modeli, MAE değerinde en iyi performansı, RMSE ve MSE değerlerinde ikinci en iyi performansı gösteren model olmuştur. Bu modelleri XGBoost modeli takip etmektedir. Değişken öneminin incelenmesi için seçilen model, rassal ormanlar regresyonu modelidir.



Şekil 4.7 a) Kan lekesi değişkeni için çeyrek yük ile oluşturulan rassal ormanlar regresyonu modelinin hata kareler ortalamasının artışı grafiği, b) Kan lekesi değişkeni için çeyrek yük ile oluşturulan rassal ormanlar regresyonu modelinin düğüm saflığının artışı grafiği

Şekil 4.7.a incelendiğinde kg başına ana yıkama suyu tüketimi değişkeni dikkat çekmektedir. Değişken birinci sırayı alarak modelde en önemli değişken olmuştur. Ardından sıralama motor yoğunluğu, yükteki sıcaklık, hacim başına yıkanan yük miktarı, ana yıkama süresi ve deterjan dozajı şeklinde devam etmektedir.

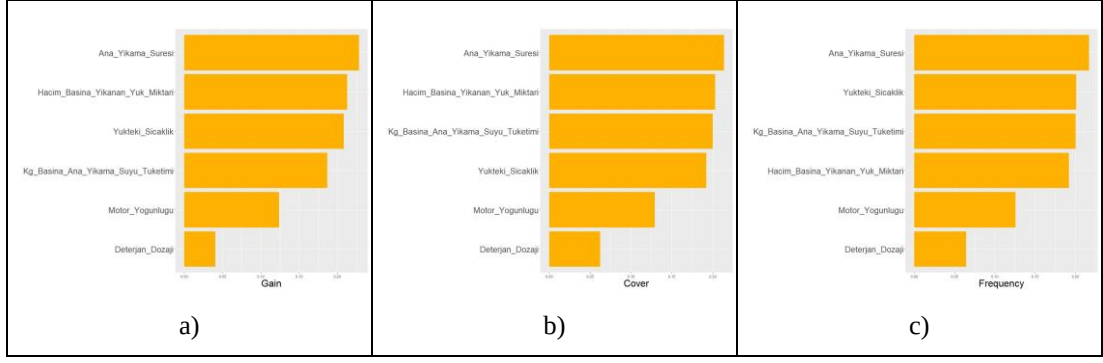
Şekil 4.7.b incelendiğinde de benzer bir durum mevcuttur. Burada da kg başına ana yıkama suyu tüketimi değişkeni en önemli değişken olmuştur. Yükteki sıcaklık, ana yıkama süresi, motor yoğunluğu, hacim başına yıkanan yük miktarı ve deterjan dozajı değişkenleri takip eden sıraları almıştır.

4.2.3.2 Yarım Yük

Tablo 4.8 Kan lekesi değişkeni için yarım yük ile oluşturulan modellere ait performans metrikleri

Modeller	RMSE	MSE	MAE
XGBoost	1.046515	1.095193	0.8225110
Rassal Ormanlar Regresyonu	1.072756	1.150805	0.8416368
SVR Polinom	1.074458	1.154460	0.7854437
Regresyon Ağacı	1.096527	1.202372	0.8936399
Bagging	1.110334	1.232841	0.8706868
SVR Doğrusal	1.116859	1.247374	0.8186004
K-NN	1.137019	1.292812	0.8964489
MLR	1.280799	1.640447	0.9316050

Kan lekesi değişkeni için yarım yük ile oluşturulan Tablo 4.8'deki modeller incelendiğinde RMSE ve MSE değerlerinde ilk üç sırayı XGBoost, rassal ormanlar regresyonu ve SVR polinom çekirdek modeli almıştır. MAE değerlerinde ise ilk üç sırayı SVR polinom çekirdek, SVR doğrusal çekirdek ve XGBoost modeli almıştır. Değişken öneminin incelenmesi için XGBoost modeli seçilmiştir.



Şekil 4.8 a) Kan lekesi değişkeni için yarım yük ile oluşturulan XGBoost modelinin kazanç grafiği, b) Kan lekesi değişkeni için yarım yük ile oluşturulan XGBoost modelinin kapsama grafiği, c) Kan lekesi değişkeni için yarım yük ile oluşturulan XGBoost modelinin sıklık grafiği

Şekil 4.8.a incelendiğinde, ana yıkama süresi değişkeni ile hacim başına yıkanan yük miktarı, yükteki sıcaklık ve kg başına yıkama suyu tüketimi değişkenlerinin modelde önemli olduğu görülmektedir. Bu değişkenleri, motor yoğunluğu değişkeni ve deterjan dozajı değişkeni takip etmiştir.

Şekil 4.8.b incelendiğinde, Şekil 4.8.a ile arasındaki fark, kg başına ana yıkama suyu tüketimi değişkeninin yükteki sıcaklık değişkeninin önüne geçmiş olmasıdır.

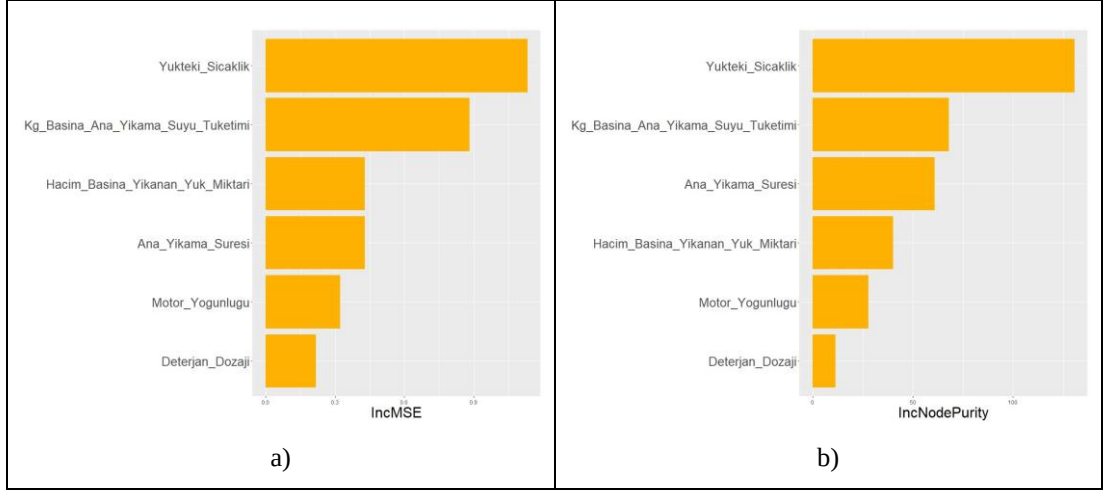
Şekil 4.8.c ile Şekil 4.8.b arasındaki fark ise, yükteki sıcaklık değişkeni ile hacim başına yıkanan yük miktarı değişkeni yer değiştirerek, yükteki sıcaklık değişkeni ikinci sırayı, hacim başına yıkanan yük miktarı değişkeni ise dördüncü sırayı almış olmasıdır.

4.2.3.3 Tam Yk

Tablo 4.9 Kan lekesi deęiřkeni iin tam yk ile oluřturulan modellere ait performans metrikleri

Modeller	RMSE	MSE	MAE
Rassal Ormanlar Regresyonu	1.134089	1.286157	0.8672452
Bagging	1.144493	1.309864	0.8605325
Regresyon Aęacı	1.167093	1.362107	0.9107162
SVR Polinom	1.262932	1.594998	1.0206943
XGBoost	1.306671	1.707390	0.9882027
SVR Doğrusal	1.329329	1.767115	1.1064714
K-NN	1.360453	1.850831	1.1004054
MLR	1.391912	1.937420	1.1200052

Tablo 4.9'daki kan lekesi deęiřkeni iin tam yk ile oluřturulan modeller incelendięinde, aęa tabanlı modeller n plana ıkmaktadır. RMSE ve MSE deęerlerinde birinci, MAE deęerinde ikinci sırada yer alan rassal ormanlar regresyonu modeli dikkat ekmektedir. RMSE ve MSE deęerlerinde ikinci ve nc sırayı bagging ve regresyon aęacı modelleri, MAE deęerinde ise birinci sırayı bagging, nc sırayı regresyon aęacı ve drdnc sırayı XGBoost modeli almıřtır. Rassal ormanlar regresyonu modeli, deęiřken neminin incelenmesi iin seilen model olmuřtur.



Şekil 4.9 a) Kan lekesi değişkeni için tam yük ile oluşturulan rassal ormanlar regresyonu modelinin hata kareler ortalamasının artışı grafiği, b) Kan lekesi değişkeni için tam yük ile oluşturulan rassal ormanlar modelinin düğüm saflığının artışı grafiği

Şekil 4.9.a incelendiğinde, yükteki sıcaklık ve ardından kg başına ana yıkama suyu tüketim değişkenlerinin modelde öneminin yüksek olduğu görülmektedir. Sırasıyla hacim başına yıkanan yük miktarı, ana yıkama süresi, motor yoğunluğu ve deterjan dozajı değişkenleri önem sırasında yerlerini almıştır.

Şekil 4.9.b incelendiğinde yükteki sıcaklık değişkeninin tekrar modelde en önemli değişken olduğu görülmektedir. Ardından takip eden sırayla kg başına ana yıkama suyu tüketimi, ana yıkama süresi, hacim başına yıkanan yük miktarı, motor yoğunluğu ve deterjan dozajı değişkenleri gelmektedir.

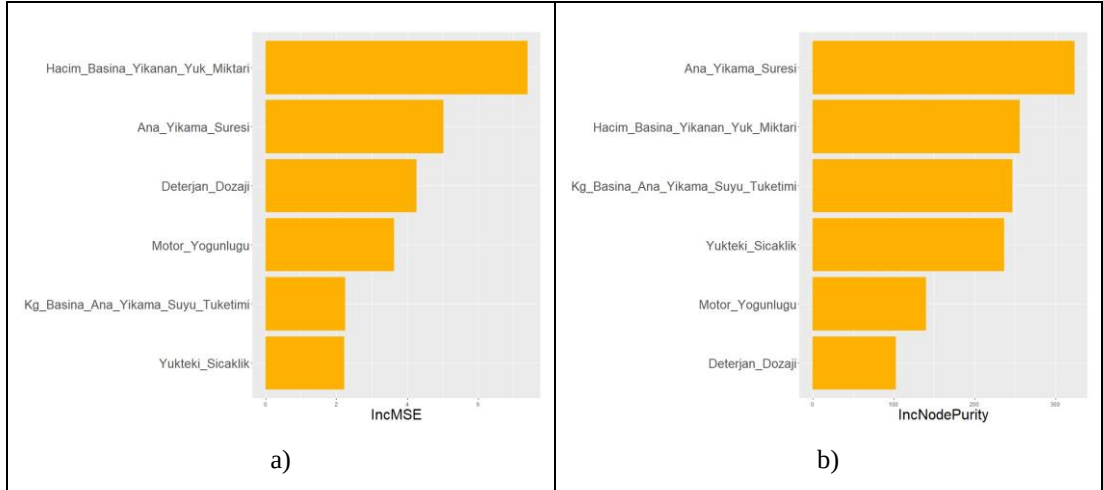
4.2.4 Kakao Lekesi Değişkeni

4.2.4.1 Çeyrek Yük

Tablo 4.10 Kakao lekesi değişkeni için çeyrek yük ile oluşturulan modellere ait performans metrikleri

Modeller	RMSE	MSE	MAE
Rassal Ormanlar Regresyonu	4.263581	18.17813	2.268466
Bagging	4.308521	18.56335	2.259268
SVR Doğrusal	4.692279	22.01748	2.625997
XGBoost	4.737810	22.44685	2.562727
K-NN	4.918829	24.19488	3.160583
Regresyon Ağacı	5.016871	25.16900	2.893564
MLR	5.204144	27.08311	3.911188
SVR Polinom	5.405378	29.21811	2.627255

Tablo 4.10’da kakao lekesi değişkeni için çeyrek yük ile elde edilen modellerde, ağaç tabanlı modellerin dikkat çekici performansları devam etmektedir. Rassal ormanlar regresyonu modeli RMSE ve MSE değerinde birinci, MAE değerinde ikinci olarak ön plana çıkmıştır. RMSE ve MSE değerinde bagging, SVR doğrusal çekirdek ve XGBoost modelleri, rassal ormanlar regresyonu modelini takip etmiştir. MAE değerinde ise bagging modeli birinci, XGBoost modeli üçüncü olmuştur. Değişken öneminin incelenmesi için seçilen model rassal ormanlar regresyonu modeli olmuştur.



Şekil 4.10 a) Kakao lekesi değişkeni için çeyrek yük ile oluşturulan rassal ormanlar regresyonu modelinin hata kareler ortalamasının artışı grafiği, b) Kakao lekesi değişkeni için çeyrek yük ile oluşturulan rassal ormanlar regresyonu modelinin düğüm saflığının artışı grafiği

Şekil 4.10.a incelendiğinde dikkat çeken değişken hacim başına yıkanan yük miktarı değişkeni olmuştur. Bu değişkeni ana yıkama süresi, deterjan dozajı ve motor yoğunluğu değişkenleri takip etmiştir. Kg başına ana yıkama suyu tüketim ve yükteki sıcaklık değişkenleri ise önem sırasında son iki sırayı almıştır.

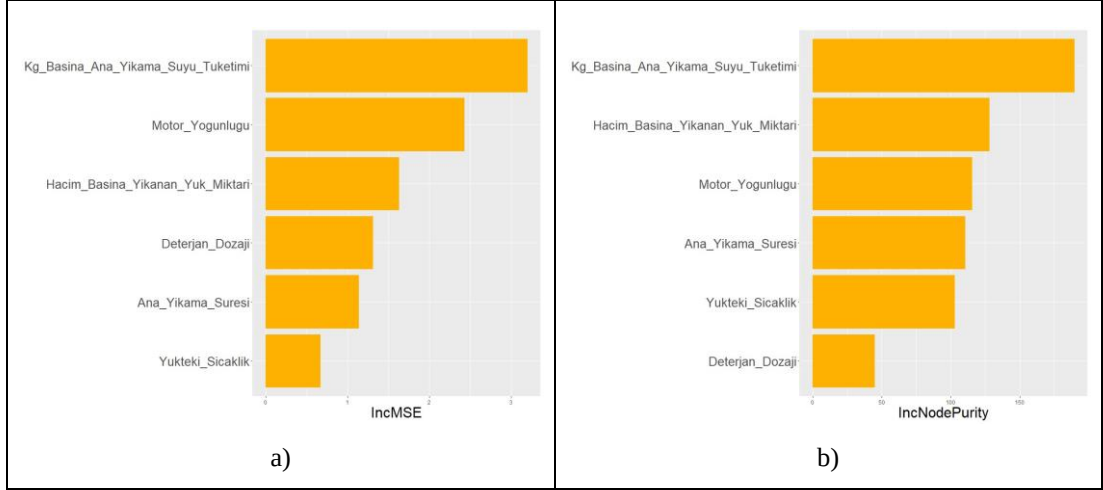
Şekil 4.10.b incelendiğinde ise ana yıkama süresi değişkeni öne çıkmaktadır. Kalan değişkenler ise hacim başına yıkanan yük miktarı, kg başına ana yıkama suyu tüketimi, yükteki sıcaklık, motor yoğunluğu ve deterjan dozajı şeklinde sıralanmaktadır.

4.2.4.2 Yarım Yük

Tablo 4.11 Kakao lekesi değişkeni için yarım yük ile oluşturulan modellere ait performans metrikleri

Modeller	RMSE	MSE	MAE
Rassal Ormanlar Regresyonu	1.126484	1.268967	0.9713120
Bagging	1.141710	1.303503	0.9760537
XGBoost	1.243200	1.545545	1.0378541
Regresyon Ağacı	1.414668	2.001287	1.1147383
SVR Polinom	1.516725	2.300456	1.2861064
K-NN	1.559743	2.432799	1.3377734
SVR Doğrusal	1.599923	2.559753	1.3284978
MLR	1.858138	3.452677	1.5046932

Kakao lekesi değişkeni için yarım yük ile kurulan Tablo 4.11'deki modeller incelendiğinde, ağaç tabanlı modellerin fark yarattığı görülmektedir. Tüm performans metriklerinde sıralama rassal ormanlar regresyonu, bagging, XGBoost ve regresyon ağacı modeli şeklinde olmuştur. Üç performans metriğinde de ilk sırada yer alan rassal ormanlar regresyonu modeli, değişken öneminin incelenmesi için seçilmiştir.



Şekil 4.11 a) Kakao lekesi değişkeni için yarım yük ile oluşturulan rassal ormanlar modelinin hata kareler ortalamasının artışı grafiği, b) Kakao lekesi değişkeni için yarım yük ile oluşturulan rassal ormanlar regresyonu modelinin düğüm saflığının artışı grafiği

Şekil 4.11.a incelendiğinde, özellikle kg başına ana yıkama suyu tüketimi değişkeninin modelde önemli olduğu görülmektedir. Diğer değişkenler motor yoğunluğu, hacim başına yıkanan yük miktarı, deterjan dozajı, ana yıkama süresi ve yükteki sıcaklık şeklinde sıralanmaktadır.

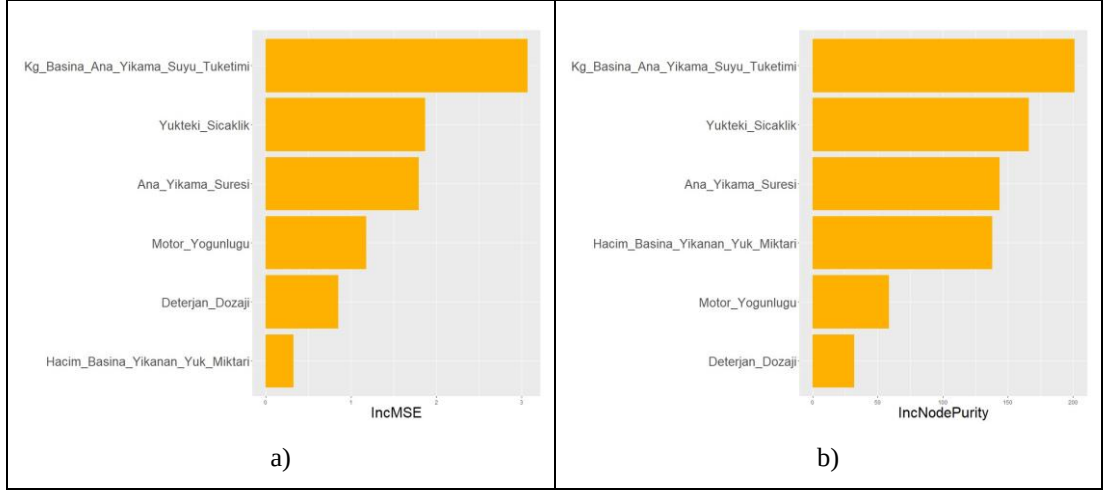
Şekil 4.11.b, Şekil 4.11.a gibi benzer bir şekilde kg başına ana yıkama suyu tüketimi değişkeni dikkat çekmektedir. Hacim başına yıkanan yük miktarı, motor yoğunluğu, ana yıkama süresi ve yükteki sıcaklık değişkenleri, kg başına ana yıkama suyu tüketimi değişkenini takip etmiştir. Deterjan dozajı değişkeni ise modelde önem sırasında son sırayı almıştır.

4.2.4.3 Tam Yk

Tablo 4.12 Kakao lekesi deęiřkeni iin tam yk ile oluřturulan modellere ait performans metrikleri

Modeller	RMSE	MSE	MAE
Bagging	2.030178	4.121624	1.550691
Rassal Ormanlar Regresyonu	2.033713	4.135989	1.563802
XGBoost	2.231591	4.979998	1.703264
SVR Doğrusal	2.728994	7.447407	2.210850
Regresyon Ağacı	2.769637	7.670892	2.052055
SVR Polinom	2.925223	8.556931	2.300046
K-NN	2.977686	8.866614	2.420288
MLR	2.990488	8.943017	2.304603

Tablo 4.12’de kakao lekesi deęiřkeni iin tam yk ile elde edilen sekiz model iinde tm performans metriklerinde ilk  sırayı ağaç tabanlı modeller almıřtır. Sıralama ise bagging, rassal ormanlar regresyonu ve XGBoost modelleri řeklinedir. Tm performans metriklerinde en iyi deęerlere sahip olan bagging modeli, deęiřken neminin incelenmesi iin seilmiřtir.



Şekil 4.12 a) Kakao lekesi değişkeni için tam yük ile oluşturulan bagging modelinin hata kareler ortalamasının artışı grafiği, b) Kakao lekesi değişkeni için tam yük ile oluşturulan bagging modelinin düğüm saflığının artışı grafiği

Şekil 4.12.a incelendiğinde kg başına ana yıkama suyu tüketimi değişkeni ön plandadır. Daha sonra önem sırasına göre yükteki sıcaklık, ana yıkama süresi, motor yoğunluğu ve deterjan dozajı değişkenleri gelmektedir. Hacim başına yıkanan yük miktarı değişkeni ise son sırada yer almaktadır.

Şekil 4.12.b incelendiğinde de yine kg başına ana yıkama suyu tüketimi değişkeni önem sırasında ilk sıradadır. Şekil 4.12.a ile kıyaslandığında hacim başına yıkanan yük miktarı değişkeni iki sıra yükselerek, son iki sırayı motor yoğunluğu ve deterjan dozajı değişkenlerinin almasını sağlamıştır.

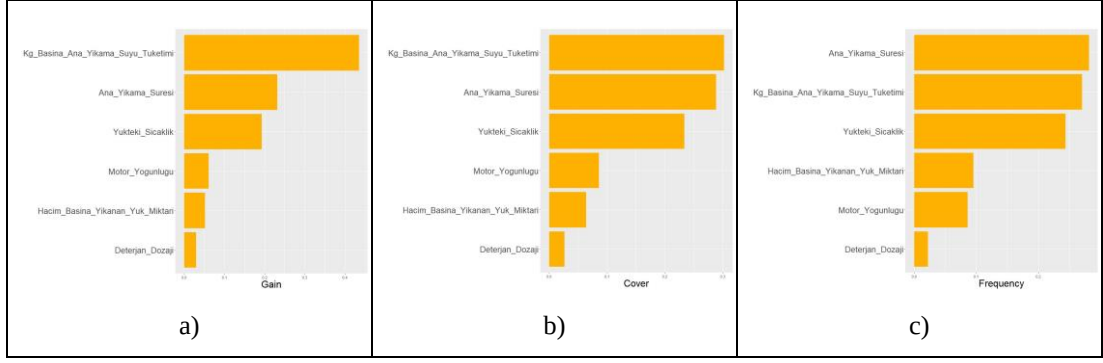
4.2.5 Kırmızı Şarap Lekesi Değişkeni

4.2.5.1 Çeyrek Yük

Tablo 4.13 Kırmızı şarap lekesi değişkeni için çeyrek yük ile oluşturulan modellere ait performans metrikleri

Modeller	RMSE	MSE	MAE
XGBoost	1.992571	3.970340	1.230724
Rassal Ormanlar Regresyonu	2.057325	4.232586	1.228137
SVR Polinom	2.172169	4.718318	1.672500
Bagging	2.200959	4.844220	1.709957
SVR Doğrusal	2.282116	5.208052	1.396377
MLR	2.427185	5.891228	1.485895
Regresyon Ağacı	2.632427	6.929670	1.764753
K-NN	2.777585	7.714977	2.314309

Kırmızı şarap lekesi değişkeni için çeyrek yük ile oluşturulan Tablo 4.13'deki modeller incelendiğinde, en iyi performansı gösteren iki modelin ağaç tabanlı model olduğu görülmektedir. RMSE ve MSE değerlerinde birinci sırada XGBoost modeli, ikinci sırada rassal ormanlar regresyonu modeli ve üçüncü SVR polinom çekirdek modeli yer almaktadır. MAE değerinde ise birinci sırada rassal ormanlar regresyonu modeli, ikinci sırada XGBoost modeli ve üçüncü sırada SVR doğrusal çekirdek modeli yer almaktadır. Değişken öneminin incelenmesi için XGBoost modeli seçilmiştir.



Şekil 4.13 a) Kırmızı şarap lekesi değişkeni için çeyrek yük ile oluşturulan XGBoost modelinin kazanç grafiği, b) Kırmızı şarap lekesi değişkeni için çeyrek yük ile oluşturulan XGBoost modelinin kapsama grafiği, c) Kırmızı şarap lekesi değişkeni için çeyrek yük ile oluşturulan XGBoost modelinin sıklık grafiği

Şekil 4.13.a ve Şekil 4.13.b incelendiğinde değişkenlerin önem sırası aynı olmuştur. Kg başına yıkama suyu tüketimi değişkeni ilk sırayı alırken, arkasından ana yıkama süresi, yükteki sıcaklık, motor yoğunluğu, hacim başına yıkanan yük miktarı ve deterjan dozajı değişkenleri gelmektedir.

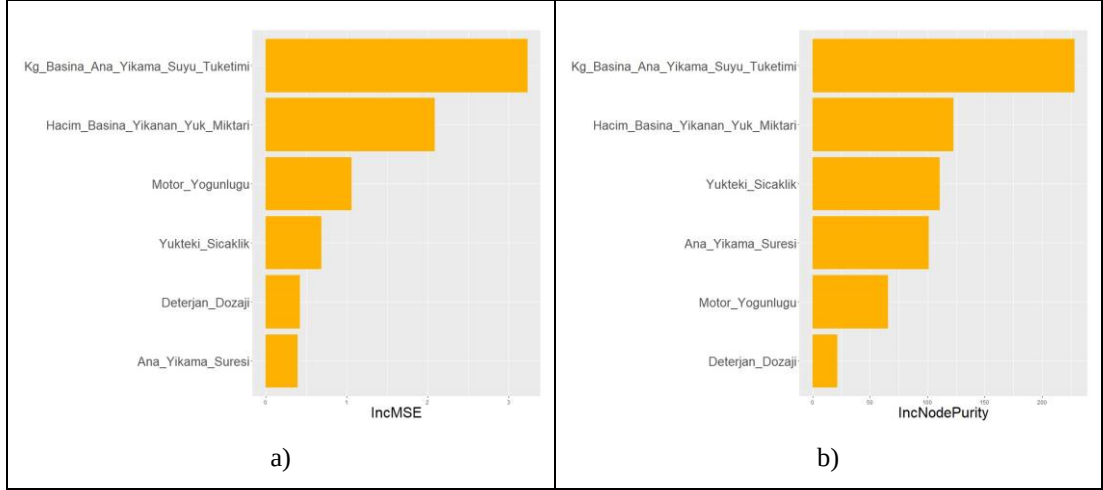
Şekil 4.13.c incelendiğinde ise sıralama ana yıkama süresi, kg başına yıkama suyu tüketimi, yükteki sıcaklık, hacim başına yıkanan yük miktarı, motor yoğunluğu ve deterjan dozajı değişkenleri şeklindedir.

4.2.5.2 Yarım Yük

Tablo 4.14 Kırmızı şarap lekesi değişkeni için yarım yük ile oluşturulan modellere ait performans metrikleri

Modeller	RMSE	MSE	MAE
Rassal Ormanlar Regresyonu	1.501142	2.253426	1.147570
XGBoost	1.577063	2.487126	1.215025
Bagging	1.580323	2.497422	1.186175
SVR Polinom	1.736900	3.016821	1.396814
K-NN	1.766745	3.121388	1.379219
Regresyon Ağacı	1.924124	3.702253	1.599096
SVR Doğrusal	1.968143	3.873587	1.710964
MLR	2.238182	5.009460	2.010402

Tablo 4.14’de kırmızı şarap lekesi değişkeni için yarım yük ile oluşturulmuş modellere incelendiğinde, performans metriklerinde ilk üç sırayı ağaç tabanlı modellerin aldığı görülmektedir. Tüm performans metriklerinde rassal ormanlar regresyonu modeli birinci olmuştur. RMSE ve MSE değerlerinde XGBoost modeli ikinci, bagging modeli ise üçüncü, MAE değerinde ise bagging modeli ikinci, XGBoost modeli üçüncü olmuştur. Değişken öneminin incelenmesi için seçilen model, tüm performans metriklerinde en iyi performansı gösteren rassal ormanlar regresyonu modelidir.



Şekil 4.14 a) Kırmızı şarap lekesi değişkeni için yarım yük ile oluşturulan rassal ormanlar modelinin hata kareler ortalamasının artışı grafiği, b) Kırmızı şarap lekesi değişkeni için yarım yük ile oluşturulan rassal ormanlar regresyonu modelinin düğüm saflığının artışı grafiği

Şekil 4.14.a incelendiğinde, kg başına ana yıkama suyu tüketimi değişkeni, ardından hacim başına yıkanan yük miktarı değişkeni önem sırasında ilk iki sırayı almıştır. Ardından motor yoğunluğu, yükteki sıcaklık, deterjan dozajı ve ana yıkama süresi değişkenleri takip etmiştir.

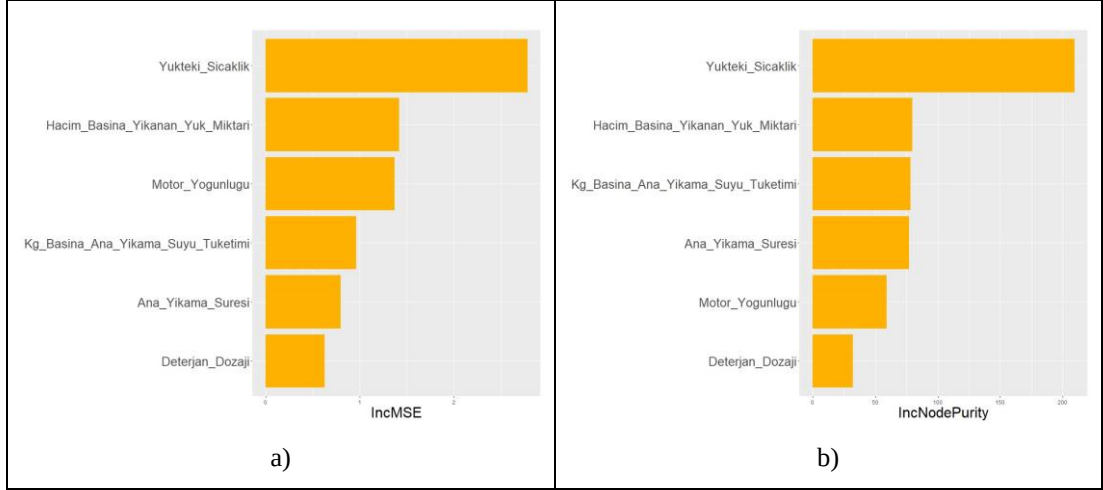
Şekil 4.14.b, Şekil 4.14.a ile aynı şekilde ilk iki sırayı, kg başına ana yıkama suyu tüketimi değişkeni ile hacim başına yıkanan yük miktarı değişkeni almıştır. Bu değişkenleri ise yükteki sıcaklık, ana yıkama süresi, motor yoğunluğu ve deterjan dozajı değişkenleri takip etmiştir.

4.2.5.3 Tam Yk

Tablo 4.15 Kırmızı şarap lekesi deęiřkeni iin tam yk ile oluřturulan modellere ait performans metrikleri

Modeller	RMSE	MSE	MAE
Rassal Ormanlar Regresyonu	1.904373	3.626635	1.477604
Bagging	2.021486	4.086406	1.534247
XGBoost	2.128223	4.529332	1.592474
SVR Polinom	2.327895	5.419094	1.733888
K-NN	2.372576	5.629118	1.934201
Regresyon Aęacı	2.576388	6.637773	1.765694
SVR Doğrusal	2.687856	7.224572	2.072772
MLR	2.691670	7.245087	2.067223

Kırmızı şarap lekesi deęiřkeni iin tam yk ile elde edilen sekiz model iinde en iyi performansları gsteren Tablo 4.15’deki modeller, aęaç tabanlı modeller olmuřtur. Tm performans metriklerinde birinci model rassal ormanlar regresyonu modeli, ikinci model bagging modeli ve nc model XGBoost modeli olmuřtur. Deęiřken neminin incelenmesi iin tm performans metriklerinde en iyi deęerlere sahip olan rassal ormanlar regresyonu modeli seilmiřtir.



Şekil 4.15 a) Kırmızı şarap lekesi değişkeni için tam yük ile oluşturulan rassal ormanlar regresyonu modelinin hata kareler ortalamasının artışı grafiği, b) Kırmızı şarap lekesi değişkeni için tam yük ile oluşturulan rassal ormanlar regresyonu modelinin düğüm saflığının artışı grafiği

Şekil 4.15.a incelendiğinde önem sırasında yükteki sıcaklık değişkeni ön plana çıkmaktadır. Daha sonrasında sırasıyla hacim başına yıkanan yük miktarı, motor yoğunluğu, kg başına ana yıkama suyu tüketimi, ana yıkama süresi ve deterjan dozajı değişkenleri gelmiştir.

Şekil 4.15.b incelendiğinde de yükteki sıcaklık değişkeni önem sırasında belirgin bir şekilde fark yaratmıştır. Ardından hacim başına yıkanan yük miktarı, kg başına ana yıkama suyu tüketimi, ana yıkama süresi ve motor yoğunluğu değişkenleri gelmektedir. Deterjan dozajı değişkeni ise önem sırasında son sırada yer almaktadır.

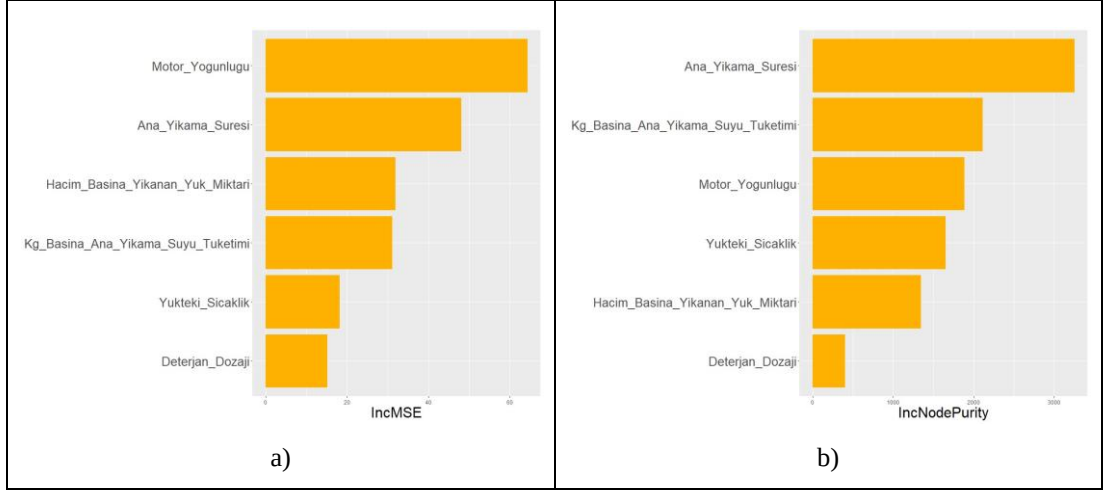
4.2.6 Toplam Değişkeni

4.2.6.1 Çeyrek Yük

Tablo 4.16 Toplam değişkeni için çeyrek yük ile oluşturulan modellere ait performans metrikleri

Modeller	RMSE	MSE	MAE
Bagging	9.036074	81.65063	5.467522
Rassal Ormanlar Regresyonu	9.056070	82.01240	5.561803
XGBoost	10.033020	100.66149	6.218096
K-NN	10.084994	101.70710	6.844167
Regresyon Ağacı	10.456108	109.33019	7.347278
SVR Doğrusal	10.815489	116.97480	7.044730
SVR Polinom	12.351135	152.55054	6.622358
MLR	12.501609	156.29023	9.682154

Tüm lekelerin yıkama performans değerleri toplamıyla elde edilen toplam değişkeni için çeyrek yük ile Tablo 4.16’da oluşturulan modellerin tüm performans metriklerinde ilk üç sırayı ağaç tabanlı modeller almıştır. En iyi performans gösteren model bagging modeli olurken, rassal ormanlar regresyonu modeli ikinci ve XGBoost modeli ise üçüncü en iyi model olmuştur. Değişken öneminin incelenmesi için seçilen model, tüm performans metriklerinde birinci olan bagging modelidir.



Şekil 4.16 a) Toplam değişkeni için çeyrek yük ile oluşturulan bagging modelinin hata kareler ortalamasının artışı grafiği, b) Toplam değişkeni için çeyrek yük ile oluşturulan bagging modelinin düğüm saflığının artışı grafiği

Şekil 4.16.a incelendiğinde, özellikle motor yoğunluğu, ardından ana yıkama süresi değişkenleri, modeldeki önem sırasında diğerlerinden ayrılmaktadır. Daha sonrasında hacim başına yıkanan yük miktarı ve kg başına ana yıkama suyu tüketimi değişkenleri gelirken, son iki sırada yükteki sıcaklık ve deterjan dozajı değişkenleri yer almaktadır.

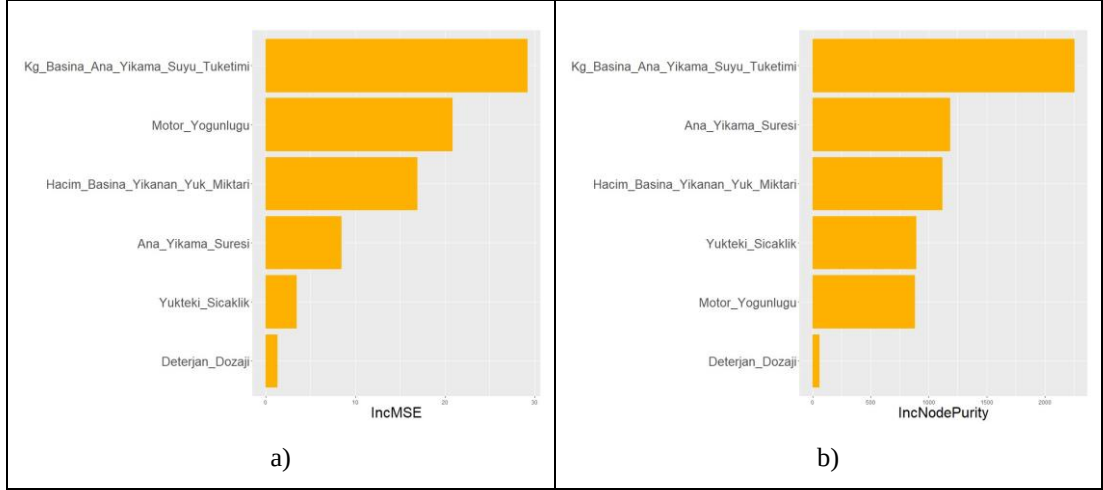
Şekil 4.16.b incelendiğinde ise ana yıkama süresi değişkeni dikkat çekmektedir. Ardından kg başına ana yıkama suyu tüketimi, motor yoğunluğu, yükteki sıcaklık ve hacim başına yıkanan yük miktarı değişkenleri gelirken, deterjan dozajı değişkeni önem sırasında son sırada yer almaktadır.

4.2.6.2 Yarım Yük

Tablo 4.17 Toplam değişkeni için yarım yük ile oluşturulan modellere ait performans metrikleri

Modeller	RMSE	MSE	MAE
Bagging	3.724738	13.87367	3.031125
Rassal Ormanlar Regresyonu	3.727160	13.89172	3.069601
XGBoost	3.804314	14.47280	3.049511
Regresyon Ağacı	4.384425	19.22318	3.646699
K-NN	4.419684	19.53360	3.593125
SVR Polinom	4.681223	21.91385	3.886895
SVR Doğrusal	5.553194	30.83797	4.457863
MLR	6.314237	39.86958	5.634158

Tablo 4.17’de toplam değişkeni için yarım yük ile oluşturulan modeller arasında ağaç tabanlı modeller fark yaratmaktadır. RMSE ve MSE değerlerinde ilk dört sırayı bagging, rassal ormanlar regresyonu, XGBoost ve regresyon ağacı modelleri alırken, MAE değerinde ilk üç sırayı bagging, XGBoost ve rassal ormanlar regresyonu modelleri almıştır. Üç performans metriğinde de ilk sırada yer alan bagging modeli, değişken öneminin incelenmesi için seçilmiştir.



Şekil 4.17 a) Toplam değişkeni için yarım yük ile oluşturulan bagging modelinin hata kareler ortalamasının artışı grafiği, b) Toplam değişkeni için yarım yük ile oluşturulan bagging modelinin düğüm saflığının artışı grafiği

Şekil 4.17.a incelendiğinde ilk sırada kg başına ana yıkama suyu tüketimi değişkeni yer almış, ardından motor yoğunluğu, hacim başına yıkanan yük miktarı, ana yıkama süresi, yükteki sıcaklık ve deterjan dozajı gelmektedir.

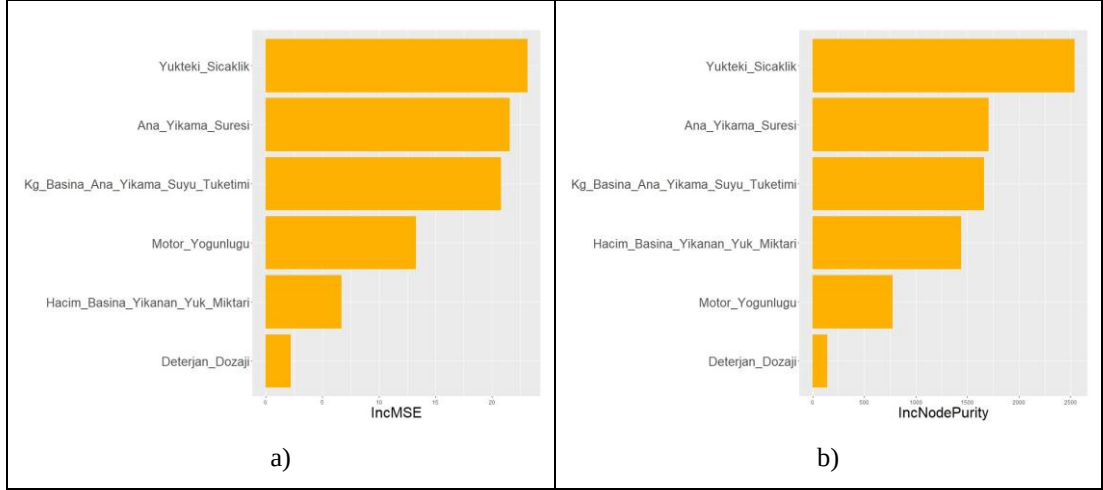
Şekil 4.17.b incelendiğinde ise Şekil 4.17.a ile aynı şekilde kg başına ana yıkama suyu tüketimi değişkeni, modeldeki önem sırasında ilk sırada yer almıştır. Daha sonrasında sırasıyla ana yıkama süresi, hacim başına yıkanan yük miktarı, yükteki sıcaklık ve motor yoğunluğu değişkenleri gelmektedir. Deterjan dozajı önem sırasında son sırada yer almaktadır.

4.2.6.3 Tam Yk

Tablo 4.18 Toplam deęiřkeni iin tam yk ile oluřturulan modellere ait performans metrikleri

Modeller	RMSE	MSE	MAE
Bagging	6.426984	41.30612	5.036965
Rassal Ormanlar Regresyonu	6.480887	42.00189	5.095090
XGBoost	8.607913	74.09616	6.642152
Regresyon Aęacı	9.692493	93.94442	6.295390
K-NN	10.513387	110.53130	7.657016
SVR Polinom	10.595598	112.26670	7.718438
MLR	10.886022	118.50548	7.446465
SVR Doğrusal	11.357551	128.99397	7.807631

Toplam deęiřkeni iin tam yk ile elde edilen Tablo 4.18'deki modeller incelendięinde, aęa tabanlı modeller fark yaratmaktadır. Tm performans metriklerinde ilk iki sırada bagging ve rassal ormanlar regresyonu modelleri yer almıřtır. RMSE ve MSE deęerlerinde nc ve drdnc sırada XGBoost modeli ile regresyon aęacı modeli, MAE deęerinde ise nc ve drdnc sırada regresyon aęacı modeli ile XGBoost modeli bulunmaktadır. Tm performans metriklerinde en iyi deęerlere sahip olan bagging modeli, deęiřken neminin incelenmesi iin seilmiřtir.



Şekil 4.18 a) Toplam değişkeni için tam yük ile oluşturulan bagging modelinin hata kareler ortalamasının artışı grafiği, b) Toplam değişkeni için tam yük ile oluşturulan bagging modelinin düğüm saflığının artışı grafiği

Şekil 4.18.a incelendiğinde yükteki sıcaklık, ana yıkama süresi ve kg başına ana yıkama suyu tüketimi değişkenleri dikkat çekmektedir. Daha sonra önem sırasında motor yoğunluğu ve hacim başına yıkanan yük miktarı değişkenleri gelmektedir. Deterjan dozajı değişkeni ise son sırada yer almaktadır.

Şekil 4.18.b incelendiğinde ise önceki grafikte benzer bir şekilde, ilk üç sırada yükteki sıcaklık, ana yıkama süresi ve kg başına ana yıkama suyu tüketimi değişkenleri yer almaktadır. Ardından sırasıyla hacim başına yıkanan yük miktarı ve motor yoğunluğu değişkenleri gelmekte ve son sıra deterjan dozajı değişkeni bulunmaktadır.

BÖLÜM BEŞ

DEĞERLENDİRME VE SONUÇ

Vestel Beyaz Eşya firmasıyla yürütülen bu çalışmada, firmadan temin edilen sebum, karbon, kan, kakao ve kırmızı şarap lekelerinin yıkama performans değerleri ve tüm lekelerin yıkama performans değeri toplamı ile çeyrek yük, yarım yük ve tam yük için toplam 18 veri kümesi oluşturulmuştur. Bu veri kümeleri MLR, regresyon ağacı, bagging, rassal ormanlar regresyonu, XGBoost regresyonu, SVR (doğrusal çekirdek ve polinom çekirdek) ve K-NN modelleri ile modellenmiştir. Oluşturulan tüm modellerin değerlendirilmesinde hata kareler ortalaması, kök hata kareler ortalaması ve ortalama mutlak hata performans metrikleri kullanılmıştır.

Çalışma sonucunda, tüm lekeler ayrı ayrı incelendiğinde sebum lekesi için oluşturulan modellerde, kg başına ana yıkama suyu tüketimi değişkeni bir istisna dışında tüm yüklerde, modelde önemi ilk sırada yer almıştır. İstisnayı oluşturan durum ise tam yük ile oluşturulan bagging modelinin hata kareler ortalamasının artışı grafiğidir. Ancak burada da ikinci sırayı alarak, modelde önem sırasında üst sıralarda olmayı başarmıştır. Sonuç olarak kg başına ana yıkama suyu tüketimi değişkeni, sebum lekesi için oluşturulan modellerde oldukça önemli bir değişkendir. Bununla birlikte deterjan dozajı değişkeni ise tüm yüklerde son sırada yer almıştır. Sebum lekesi için oluşturulan modellerde önemi en düşük değişken deterjan dozajı değişkenidir.

Karbon lekesi için oluşturulan modellerde ise çeyrek yükte, hacim başına yıkanan yük miktarı değişkeni, ardından motor yoğunluğu değişkeni dikkat çekmektedir. Yarım yükte ise motor yoğunluğu ve kg başına ana yıkama suyu tüketimi değişkenleri ön plana çıkmıştır. Tam yükte ise yükteki sıcaklık, ana yıkama süresi ve kg başına ana yıkama suyu tüketimi değişkenleri ilk üç sırayı almıştır. Deterjan dozajı değişkeni, çeyrek yük ile oluşturulan rassal ormanlar regresyonu modelinin hata kareler ortalaması artışı grafiği hariç tüm durumlarda en son sırada yer alarak modellerde önemi en düşük değişken olmuştur.

Kan lekesi için oluşturulan modeller incelendiğinde, çeyrek yükte kg başına ana yıkama suyu tüketimi değişkeni ön plana çıkmıştır. Yarım yükte ise kg başına ana yıkama suyu

değişkeninin ana yıkama süresi, hacim başına yıkanan yük miktarı ve yükteki sıcaklık değişkenleriyle birlikte modelde önemi yüksektir. Kg başına ana yıkama suyu değişkeni tam yükte, yükteki sıcaklık değişkeninin arkasından gelerek modelde etkisi en yüksek değişkenlerden olmuştur. Kan lekesi için oluşturulan tüm modellerde kg başına ana yıkama suyu değişkeninin öneminin yüksek olduğu tespit edilmiştir. Bununla birlikte yarım yük ve tam yükte, yükteki sıcaklık değişkeninin de önemi yüksektir. Deterjan dozajı değişkeni ise modellerde önem sırasında son sırada yer alarak önemi en düşük değişken olmuştur.

Kakao lekesi için oluşturulan modellerde, çeyrek yükte hacim başına yıkanan yük miktarı ve ana yıkama süresi değişkeni değişkenlerinin önemi yüksektir. Yarım yükte ve tam yükte kg başına ana yıkama suyu tüketimi değişkeni ön plana çıkmıştır. Deterjan dozajı değişkeni ise çeyrek yük ile oluşturulan rassal ormanlar regresyonu modelinin hata kareler ortalamasının artışı grafiği hariç diğer tüm durumlarda alt sıralarda yer almıştır.

Kırmızı şarap için çeyrek yükte oluşturulan modellerde özellikle kg başına yıkama suyu tüketimi değişkeni ön plana çıkmıştır. Bununla birlikte ana yıkama süresi ve yükteki sıcaklık değişkenlerinin de önemli olduğu söylenebilir. Yarım yükte de kg başına yıkama suyu tüketimi değişkeninin modeli önemi yüksektir. Tam yükte ise yükteki sıcaklık değişkeninin önemi yüksektir. Kırmızı şarap lekesi için de deterjan dozajı değişkeni önem sıralarında en alt sıralarda yer bulabilmiştir.

Toplam değişkeni için çeyrek yük ile oluşturulan modellerde ana yıkama süresi ve motor yoğunluğu değişkenleri ön plana çıkmıştır. Yarım yükte kg başına ana yıkama suyu tüketimi değişkeni önem sırasında ilk sırada yer almıştır. Tam yükte ise yükteki sıcaklık, ana yıkama süresi ve kg başına ana yıkama suyu tüketimi değişkenlerinin model için önemli olduğu görülmüştür. Deterjan dozajı değişkeni ise tüm modellerde önem sırasında son sırada yer bulmuştur.

Yapılan testlerde değişkenlerin aldığı değerler göz önünde bulundurulduğunda, kg başına ana yıkama suyu tüketimi değişkeni yıkama performansı üzerinde oldukça etkili olmuştur. Genellikle tam yükte yapılan testlerde, yükteki sıcaklık değişkeni de ön plana çıkmıştır. Bu durumun sebebi, sıcaklık değeri çeyrek ve yarım yüke kıyasla

tam yük için yapılan testlerde daha geniş aralıkta deęerler almaktadır. Deterjan dozajı deęişkeninin modellerde öneminin düşük olmasının sebebi ise deterjan dozajı EN 60456 standardına göre belirlenmiş olması ve aldığı deęerlerin etkisinin anlaşılmasını sağlayacak kadar deęişkenlik göstermemesidir. Bu sayede, çamaşır makinesi üreticisi, testlerde çamaşır makinesinin performansını daha güvenilir bir şekilde gözlemleyebilmektedir.



Tablo 4.19 Tüm modellerin RMSE değerleri

	MLR	Regresyon Ağacı	Bagging	Rassal Ormanlar Regresyonu	XGBoost	SVR Doğrusal	SVR Polinom	K-NN
Sebum	Çeyrek Yük	1.965805	1.453570	1.480113	1.514631	1.880455	1.721777	1.687806
	Yarım Yük	1.2616531	0.9449956	0.9243660	1.0241559	1.1866672	1.2762552	1.1470638
	Tam Yük	2.533736	1.624298	1.646797	1.734781	2.988453	2.610989	2.663582
Karbon	Çeyrek Yük	2.777585	2.057325	1.992571	2.427185	2.200959	2.282116	2.172169
	Yarım Yük	1.5437958	0.9378708	0.9557084	0.9478575	1.4106619	1.3708078	1.3555060
	Tam Yük	2.715035	1.569974	1.627998	1.517015	2.507685	2.631023	2.692794
Kan	Çeyrek Yük	2.996442	2.206045	2.188376	2.312994	2.544880	2.462067	3.067271
	Yarım Yük	1.280799	1.110334	1.072756	1.046515	1.116859	1.074458	1.137019
	Tam Yük	1.391912	1.144493	1.134089	1.306671	1.329329	1.262932	1.360453
Kakao	Çeyrek Yük	5.204144	4.308521	4.263581	4.737810	4.692279	5.405378	4.918829
	Yarım Yük	1.858138	1.141710	1.126484	1.243200	1.599923	1.516725	1.559743
	Tam Yük	2.990488	2.030178	2.033713	2.231591	2.728994	2.925223	2.977686
Kırmızı Şarap	Çeyrek Yük	2.427185	2.200959	2.057325	1.992571	2.282116	2.172169	2.777585
	Yarım Yük	2.238182	1.580323	1.501142	1.577063	1.968143	1.736900	1.766745
	Tam Yük	2.691670	2.021486	1.904373	2.128223	2.687856	2.327895	2.372576
Toplam	Çeyrek Yük	12.501609	9.036074	9.056070	10.033020	10.815489	12.351135	10.084994
	Yarım Yük	6.314237	3.724738	3.727160	3.804314	5.553194	4.681223	4.419684
	Tam Yük	10.886022	6.426984	6.480887	8.607913	11.357551	10.595598	10.513387

Çalışmada oluşturulan 18 modelin Tablo 4.19’da verilen RMSE değerleri incelendiğinde ağaç tabanlı modeller dikkat çekicidir. Kırmızı renk ile gösterilen RMSE değeri en iyi olan modellerin 8’i rassal ormanlar regresyonu, 7’si bagging ve 3’ü XGBoost modeli olmuştur. Mavi renk ile gösterilen RMSE değerinde ikinci en iyi modellerin 8’i rassal ormanlar regresyonu, 8’i bagging ve 2’si XGBoost modelidir. Yeşil renk ile gösterilen RMSE değerinde üçüncü en iyi modellerin 9’u XGBoost, 2’si rassal ormanlar regresyonu, 2’si SVR polinom çekirdek, 1’i regresyon ağacı, 1’i bagging, 1’i SVR doğrusal çekirdek ve 1’i K-NN modeli olmuştur. Sonuç olarak tüm lekelerde ağaç tabanlı modeller oldukça iyi performans göstermiştir.

Yapılan çalışmada, çamaşır makinesi üreticileri ürettikleri veya üretecekleri çamaşır makineleri için testler gerçekleştirmeden önce, oluşturulan modelleri kullanarak leke çıkarma performansını tahmin edebilirler. Bu sayede iyi performans vermeyecek değerlerle testler gerçekleştirilmesinin önüne geçilerek iş gücü, zaman, su ve elektrik tüketiminden tasarruf edilmesi sağlanacak ve test verimliliği artacaktır.

KAYNAKLAR

- Abeliotis, K., Candan, C., Amberg, C., Ferri, A., Osset, M., Owens, J., & Stamminger, R. (2015). Impact of water hardness on consumers' perception of laundry washing result in five European countries. *International Journal of Consumer Studies*, 39(1), 60-66.
- Aha, D. W. (1997). Lazy learning. In *Lazy learning* (pp. 7-10). Dordrecht: Springer Netherlands.
- Ali, A., Yadav, K. S., & Yadav, A. (2019). Mechanical washing machine. *International Journal of Research in Engineering, Science and Management*, 2(6), 76-80.
- Alpaydin, E. (1997). Voting over multiple condensed nearest neighbors. *Lazy learning*, 115-132.
- Bansal, M., Goyal, A., & Choudhary, A. (2022). A comparative analysis of K-Nearest Neighbour, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. *Decision Analytics Journal*, 100071.
- Bloomfield, S. F., Exner, M., Signorelli, C., Nath, K. J., & Scott, E. A. (2011). The infection risks associated with clothing and household linens in home and everyday life settings, and the role of laundry. In *International Scientific Forum on Home Hygiene*.
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152).
- Breiman, L. (1996). Bagging predictors. *Machine learning*, 24, 123-140.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- Breiman, L. (2017). *Classification and regression trees*. Routledge.
- Chatterjee, S., & Hadi, A. S. (2013). Regression analysis by example. John Wiley & Sons.

- Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- Cheng, K., Lu, Z., Wei, Y., Shi, Y., & Zhou, Y. (2017). Mixed kernel function support vector regression for global sensitivity analysis. *Mechanical Systems and Signal Processing*, 96, 201-214.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20, 273-297.
- Efe, S. (2015). *Deterjan Bileşenlerinin Temizleme Aktivitesi Üzerine Mekanik Hareket Ve Sıcaklık Etkisinin İncelenmesi* (Doctoral dissertation, Fen Bilimleri Enstitüsü).
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4), 367-378.
- Guyon, I., Boser, B., & Vapnik, V. (1992). Automatic capacity tuning of very large VC-dimension classifiers. *Advances in neural information processing systems*, 5.
- Haara, A., & Kangas, A. S. (2012). Comparing K nearest neighbours methods and Linear regression-is there reason to select one Over the other?. *Math. Comput. For. Nat. Resour. Sci.*, 4(1), 50-65.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.
- Karakaş, S. (2011). *Vector control of PMSM in washing machine application* (Doctoral dissertation).

- Lambert, E., Maitra, W., Scheid, F., Niestrath, M., Gorny, S., & Stamminger, R. (2016). Differentiated evaluation of washing performance in washing machines of test stain strips as a function of temperature, washing duration and load size. *Tenside Surfactants Detergents*, 53(5), 424-437.
- Lasic, E. (2014). *Sustainable use of washing machine: modeling the consumer behavior related resources consumption in use of washing machines* (Doctoral dissertation, Universitäts-und Landesbibliothek Bonn).
- Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons. b*, 4, 51-62.
- Özyay, B. (2016). *Önden Yüklemeli Çamaşır Makinası İçin Farklı Leke Gruplarının Temizlenmesini Sağlayacak Bir Yıkama Algoritmasının Oluşturulması* (Doctoral dissertation, Fen Bilimleri Enstitüsü).
- Qiu, Y., Zhou, J., Khandelwal, M., Yang, H., Yang, P., & Li, C. (2021). Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration. *Engineering with Computers*, 1-18.
- Schölkopf, P. B., Burges, C., & Vapnik, V. (1995). Extracting support data for a given task. In *Proceedings, First International Conference on Knowledge Discovery & Data Mining. AAAI Press, Menlo Park, CA* (pp. 252-257).
- Schölkopf, B., Burges, C., & Vapnik, V. (1996). Incorporating invariances in support vector learning machines. In *Artificial Neural Networks—ICANN 96: 1996 International Conference Bochum, Germany, July 16–19, 1996 Proceedings 6* (pp. 47-52). Springer Berlin Heidelberg.
- Sinner, H. (1960). *Über das Waschen mit Haushaltwaschmaschinen: in welchem Umfange erleichtern Haushaltwaschmaschinen und-geräte das Wäschehaben im Haushalt?*. Haus+ Heim-Verlag.
- Smulders, E., Rähse, W., von Rybinski, W., Steber, J., Sung, E., & Wiebel, F. (2002). *Laundry detergents* (p. 52). Weinheim: Wiley-Vch.
- Stamminger, R. (2010). Reinigen. *Bonn: aid infodienst*, 28.

- Taunk, K., De, S., Verma, S., & Swetapadma, A. (2019). A brief review of nearest neighbor algorithm for learning and classification. In *2019 international conference on intelligent computing and control systems (ICCS)* (pp. 1255-1260). IEEE.
- Vapnik, V., Golowich, S., & Smola, A. (1996). Support vector method for function approximation, regression estimation and signal processing. *Advances in neural information processing systems*, 9.
- Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1), 79-82.
- Zuheros, R. G. (2011). *Structural Analysis of a Washing Ma-chine Through Its Loading Case: Redesign of The Tripod AndThe Front Cover* (Doctoral dissertation, Sweden: University of Skovde).