



T.C.
SELÇUK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ELMAS FİYAT TAHMİNİ İÇİN MAKİNE
ÖĞRENMESİ YÖNTEMLERİNİN
İNCELENMESİ

Erkan ÜNSAL

YÜKSEK LİSANS TEZİ

Bilgisayar Mühendisliği Anabilim Dalı

Ocak-2025
KONYA
Her Hakkı Saklıdır

TEZ KABUL VE ONAYI

Erkan Ünsal tarafından hazırlanan “Elmas Fiyat Tahmini İçin Makine Öğrenmesi Yöntemlerinin İncelenmesi” adlı tez çalışması 02/01/2025 tarihinde aşağıdaki jüri tarafından oy birliği / ~~oy çokluğu~~ ile Selçuk Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı’nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri

İmza

Başkan

Prof. Dr. Nurettin DOĞAN

.....

Danışman

Prof. Dr. Humar KAHRAMANLI ÖRNEK

.....

Üye

Dr. Öğr. Üyesi İnayet Hakkı ÇİZMECİ

.....

Üye

Unvanı Adı SOYADI

.....

Üye

Unvanı Adı SOYADI

.....

Yukarıdaki sonucu onaylarım.

Prof. Dr. Ömer Faruk YÜKSEL
FBE Müdürü

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Erkan ÜNSAL

02.01.2025

ÖZET

YÜKSEK LİSANS TEZİ

ELMAS FİYAT TAHMİNİ İÇİN MAKİNE ÖĞRENMESİ YÖNTEMLERİNİN İNCELENMESİ

Erkan ÜNSAL

Selçuk Üniversitesi Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Humar KAHRAMANLI ÖRNEK

2025, 42 Sayfa

Jüri

Prof. Dr. Humar KAHRAMANLI ÖRNEK

Prof. Dr. Nurettin DOĞAN

Dr. Öğr. Üyesi İnayet Hakkı ÇİZMECİ

Başta mücevherat olmak üzere birçok alanda çeşitli amaçlar için kullanılan elmasın, karmaşık bir fiyatlandırma sistemine sahip olması elmas ticaretinde sorunlara yol açmaktadır. Bu kapsamda, farklı kodlama ve normalizasyon tekniklerinin uygulandığı çeşitli veri ön işleme adımlarından geçirilerek elde edilen 4 değişik veri setiyle özgün makine öğrenmesi yöntemleri kullanılarak elmas fiyat tahmini yapılmaktadır. Elmas fiyat tahmininde kullanılan doğrusal (linear), k-en yakın komşu (k-nearest neighbors), destek vektör (support vector), karar ağacı (decision tree), rastgele orman (random forest), aşırı gradyan güçlendirme (extreme gradient boosting) ve çok katmanlı algılayıcı (multi-layer perceptron) yöntemlerinin performansları, yaygın olarak kullanılan MAE (Ortalama Mutlak Hata (Mean Absolute Error)), WMAE (Ağırlıklı Ortalama Mutlak Hata (Weighted Mean Absolute Error)), MSE (Ortalama Kare Hata (Mean Squared Error)), RMSE (Karekök Ortalama Kare Hata (Root Mean Squared Error)), R^2 (R-Kare (R-Squared)), Belirlilik Katsayısı (Coefficient of Determination)), MAPE (Ortalama Mutlak Yüzde Hata (Mean Absolute Percentage Error)) ve çalışmada geliştirilen yenilikçi metrikler QMAE (Kantil Ortalama Mutlak Hata (Quantile Mean Absolute Error)), AMAE (Asimetrik Ortalama Mutlak Hata (Asymmetric Mean Absolute Error)), QMAPE (Kantil Ortalama Mutlak Yüzde Hata (Quantile Mean Absolute Percentage Error)), AMAPE (Asimetrik Ortalama Mutlak Yüzde Hata (Asymmetric Mean Absolute Percentage Error)), MPAE (Ortalama Orantılı Mutlak Hata (Mean Proportion Absolute Error)), QMPAE (Kantil Ortalama Orantılı Mutlak Hata (Quantile Mean Proportion Absolute Error)), AMPAE (Asimetrik Ortalama Orantılı Mutlak Hata (Asymmetric Mean Proportion Absolute Error)) ile karşılaştırmalı olarak incelenmektedir. Bu incelemeler sonucunda, en uygun yöntemin belirlenmesi ve geliştirilen yenilikçi metriklerin geniş kullanım alanlarına sunulması hedeflenmektedir.

Anahtar Kelimeler: Elmas, Fiyat, Makine, Metrik, Öğrenme, Tahmin

ABSTRACT

MS THESIS

INVESTIGATION OF MACHINE LEARNING METHODS FOR DIAMOND PRICE PREDICTION

Erkan ÜNSAL

**SELÇUK UNIVERSITY INSTITUTE OF SCIENCES
DEPARTMENT OF COMPUTER ENGINEERING**

Advisor: Prof. Dr. Humar KAHRAMANLI ÖRNEK

2025, 42 Pages

Jury

Prof. Dr. Humar KAHRAMANLI ÖRNEK

Prof. Dr. Nurettin DOĞAN

Asst. Prof. Dr. İnayet Hakkı ÇİZMECİ

Diamond, widely used for various purposes, particularly in jewelry, has a complex pricing system that poses challenges in diamond trading. In this context, diamond price prediction is performed using original machine learning methods on four distinct datasets obtained through various data preprocessing steps, where different encoding and normalization techniques are applied. The performance of methods employed for diamond price prediction, namely linear, k-nearest neighbors, support vector, decision tree, random forest, extreme gradient boosting, and multi-layer perceptron, is investigated comparatively using widely adopted metrics such as MAE (Mean Absolute Error), WMAE (Weighted Mean Absolute Error), MSE (Mean Squared Error), RMSE (Root Mean Squared Error), R^2 (R-Squared, Coefficient of Determination), MAPE (Mean Absolute Percentage Error), alongside innovative metrics developed in this study such as QMAE (Quantile Mean Absolute Error), AMAE (Asymmetric Mean Absolute Error), QMAPE (Quantile Mean Absolute Percentage Error), AMAPE (Asymmetric Mean Absolute Percentage Error), MP AE (Mean Proportion Absolute Error), QMPAE (Quantile Mean Proportion Absolute Error), and AMPAE (Asymmetric Mean Proportion Absolute Error). These investigations aim to identify the most suitable method and to introduce the innovative metrics developed in this study for broader application areas.

Keywords: Diamond, Learning, Machine, Metric, Prediction, Price

ÖNSÖZ

Bu tez, elmas fiyat tahmini üzerine yapılan bir çalışmayı içermektedir. Çalışmanın gerçekleştirilmesini sağlayan özgün makine öğrenmesi yöntemleri yaygın olarak kullanılan ve çalışmada geliştirilen yenilikçi metrikler ile karşılaştırmalı şekilde incelenmektedir. Bu bağlamda, çalışmanın hem elmas ve fiyatlandırması ile ilgili bir rehber niteliği taşıması hem de yenilikçi metriklerin geniş kullanım alanlarına sunulması hedeflenmektedir.

Çalışma süresince, bana her aşamada rehberlik eden ve değerli bilgi birikimlerini paylaşan danışman hocam Sayın Prof. Dr. Humar KAHRAMANLI ÖRNEK'e ve bana olan inancı sayesinde her zaman motivasyon kaynağım olan Sayın Prof. Dr. Şakir TAŞDEMİR'e saygılarımı ve minnetlerimi arz ederim. Ayrıca, beni daima destekleyen nişanlım Feyza Nur ÖZTÜRK ve tüm aileme en derin duygularıyla sevgilerimi sunarım.

Başta kıymetli jüri üyeleri olmak üzere doğrudan veya dolaylı olarak katkıda bulunan herkese emekleri için teşekkür ederim. En büyük temennim, tez çalışmamın literatüre katkı sağlaması ve ilgili alanlarda yeni çalışmalar için yol gösterici olmasıdır. Saygılarımla,

Erkan ÜNSAL
KONYA-2025

İÇİNDEKİLER

ÖZET	iv
ABSTRACT.....	v
ÖNSÖZ	vi
İÇİNDEKİLER	vii
KISALTMALAR	ix
1. GİRİŞ.....	1
2. KAYNAK ARAŞTIRMASI	4
3. ELMAS	7
3.1. Elmas Fiyatlandırma	9
3.1.1. Kesim	11
3.1.2. Renk	12
3.1.3. Berraklık	14
3.1.4. Karat.....	15
4. MATERYAL VE YÖNTEM.....	16
4.1. Çalışma Düzeni.....	16
4.2. Veri Seti	16
4.3. Veri Ön İşleme.....	17
4.3.1. Veri kodlama.....	18
4.3.1.1. Etiket kodlama (Label encoding).....	18
4.3.1.2. İkili kodlama (Binary encoding).....	18
4.3.2. Veri normalizasyonu	19
4.3.2.1. Min-Max normalizasyonu.....	19
4.3.2.2. Z-Score normalizasyonu	19
4.4. Veri Setlerinin Keşfi	20
4.5. Model Seçimleri ve Kurulumları	20
4.5.1. Doğrusal (Linear - LinearRegression)	21
4.5.2. K-en yakın komşu (K-nearest neighbors - KNeighborsRegressor)	21
4.5.3. Destek vektör (Support vector - SVR).....	21
4.5.4. Karar ağacı (Decision tree - DecisionTreeRegressor)	21
4.5.5. Rastgele orman (Random forest - RandomForestRegressor)	22
4.5.6. Aşırı gradyan güçlendirme (eXtreme gradient boosting - XGBRegressor) ..	22
4.5.7. Çok katmanlı algılayıcı (Multi-layer perceptron - MLPRegressor)	22
5. METRİKLER, ARAŞTIRMA SONUÇLARI VE İNCELEMELER	23
5.1. Ortalama Mutlak Hata (Mean Absolute Error).....	23
5.2. Ağırlıklı Ortalama Mutlak Hata (Weighted Mean Absolute Error)	24
5.3. Kantil Ortalama Mutlak Hata (Quantile Mean Absolute Error)	25
5.4. Asimetrik Ortalama Mutlak Hata (Asymmetric Mean Absolute Error)	26

5.5. Ortalama Kare Hata (Mean Squared Error).....	27
5.6. Karekök Ortalama Kare Hata (Root Mean Squared Error)	28
5.7. R-Kare (R-Squared), Belirlilik Katsayısı (Coefficient of Determination)	29
5.8. Ortalama Mutlak Yüzde Hata (Mean Absolute Percentage Error).....	30
5.9. Kantil Ortalama Mutlak Yüzde Hata (Quantile Mean Absolute Percentage Error)	31
5.10. Asimetrik Ortalama Mutlak Yüzde Hata (Asymmetric Mean Absolute Percentage Error)	32
5.11. Ortalama Orantılı Mutlak Hata (Mean Proportion Absolute Error)	33
5.12. Kantil Ortalama Orantılı Mutlak Hata (Quantile Mean Proportion Absolute Error).....	34
5.13. Asimetrik Ortalama Orantılı Mutlak Hata (Asymmetric Mean Proportion Absolute Error)	35
6. SONUÇLAR.....	37
KAYNAKLAR	38

KISALTMALAR

4C	: Cut (Kesim), Color (Renk), Clarity (Berraklık), Carat (Karat)
AdaBoost	: Adaptive Boosting (Adaptif Güçlendirme)
AMAE	: Asymmetric Mean Absolute Error (Asimetrik Ortalama Mutlak Hata)
AMAPE	: Asymmetric Mean Absolute Percentage Error (Asimetrik Ortalama Mutlak Yüzde Hata)
AMPAE	: Asymmetric Mean Proportion Absolute Error (Asimetrik Ortalama Orantılı Mutlak Hata)
GIA	: Gemological Institute of America (Amerika Gemoloji Enstitüsü)
ISO	: International Organization for Standardization (Uluslararası Standart Teşkilatı)
LASSO	: Least Absolute Shrinkage and Selection Operator (En Küçük Mutlak Daralma ve Seçim Operatörü)
MAE	: Mean Absolute Error (Ortalama Mutlak Hata)
MAPE	: Mean Absolute Percentage Error (Ortalama Mutlak Yüzde Hata)
MPAE	: Mean Proportion Absolute Error (Ortalama Orantılı Mutlak Hata)
MSE	: Mean Squared Error (Ortalama Kare Hata)
QMAE	: Quantile Mean Absolute Error (Kantil Ortalama Mutlak Hata)
QMAPE	: Quantile Mean Absolute Percentage Error (Kantil Ortalama Mutlak Yüzde Hata)
QMPAE	: Quantile Mean Proportion Absolute Error (Kantil Ortalama Orantılı Mutlak Hata)
R²	: R-Squared (R-Kare), Coefficient of Determination (Belirlilik Katsayısı)
RMSE	: Root Mean Squared Error (Karekök Ortalama Kare Hata)
SVM	: Support Vector Machine (Destek Vektör Makinesi)
USD	: United States dollar (Amerikan doları)
WMAE	: Weighted Mean Absolute Error (Ağırlıklı Ortalama Mutlak Hata)
XGBoost	: eXtreme Gradient Boosting (Aşırı Gradyan Güçlendirme)

1. GİRİŞ

Karbon elementinin modifikasyonu olan elmas, doğal oluşum süreci kadar karmaşık bir fiyatlandırma sistemine sahiptir (Hershey, 1940; Robertson, 1994; Chu, 2001; Shirey ve Shigley, 2013; Bracking ve Sharife, 2014). Buna rağmen, diğer maddelerden öne çıkmasını sağlayan özellikleri, elmasın hem çeşitli sektörlerde ve uygulama alanlarında kullanılabilmesini hem de bazı endüstriyel işlemleri gerçekleştirebilmesini sağlayarak farklı avantajların elde edilmesine olanak tanır (Hershey, 1940; Shirey ve Shigley, 2013; Wang ve ark., 2020; Smith ve ark., 2021). Bu durum, taleplerin yoğunlaşmasına yol açmaktadır ancak keşfinden itibaren her zaman değer gören elmasın popülaritesinin ve pahalılığının asıl sebebi, mücevher amaçlı tercih edilmesidir (Epstein, 1982; Bauer, 2012). Elmas ticaretini destekleyen işletmelerin ve ülkelerin girişimleri, güçlü pazarlama ağı ve e-ticaretin yaygınlaşması, her türden elmas talebine yanıt verilebilmesini ve önemli ölçüde tüketici kitlesinin oluşmasını desteklemektedir (Epstein, 1982; Yarnell, 2004; Bauer, 2012). Tüketici kitlesinin, üreticilerinin ve satıcılarının en büyük sorunlarından biri elmas fiyatlandırmasıdır çünkü fiyatlandırmada ciddi düzeyde tutarsızlıklar ortaya çıkabilmektedir (Bracking ve Sharife, 2014; Wolff, 2016; Mamonov ve Triantoro, 2018; Wu ve Cosguner, 2020).

Doğal olarak meydana gelen, nispeten nadir bulunan ve yüksek ekonomik değere sahip metalik kimyasal elementler olan değerli metaller, önemli özellikleri sayesinde hem para, mücevherat, sanat ve endüstri gibi çeşitli alanlarda yaygın olarak kullanılmaktadırlar hem de yatırım ve değer saklama açısından uygun görülmektedirler (Preston, 1960; Poole Jr, 2004; Menase, 2009; Soos, 2011; Duckenfield, 2016; Umemura ve ark., 2022). Fiyatlandırmaları da nispeten basit faktörlere dayanmaktadır (Menase, 2009; Duckenfield, 2016). ISO (Uluslararası Standart Teşkilatı (International Organization for Standardization)) 4217 para birimi koduna sahip altın, gümüş, platin ve paladyum gibi en bilinen değerli metaller, kütleleri ve saflık oranlarına göre fiyatlandırılırken bazı durumlarda işçilik, işlem maliyeti, menşee ve sertifika gibi etki düzeyi düşük dış değişkenler de göz önüne alınabilmektedir (Menase, 2009; Duckenfield, 2016; Aharon ve Qadan, 2018). Değerli metaller gibi doğal olarak meydana gelen, nispeten nadir bulunan ve önemli özellikleri sayesinde birçok farklı alanda kullanılan değerli taşlarda ise fiyatlandırma çok daha farklı ve karmaşıktır (Hurlbut ve Kammerling, 1991; Chu, 2001; Matlins ve Bonanno, 2009; Bauer, 2012). Kesilip cilalanarak takı, mücevher veya süs eşyası yapmak için kullanılan ve mineral kristal parçaları olan değerli

taşların elmas, yakut, zümrüt ve safir dışındakileri genellikle yarı değerli taşlar olarak kabul edilmektedirler. Yarı değerli taşların fiyatlandırması, değerli metal fiyatlandırmasındaki stratejilere benzer şekilde basit faktörlere dayanarak yapılmaktadır ve fiyatları da elmas, yakut, zümrüt ve safirden daha düşüktür. Bu durum nadiren de olsa değişiklik gösterebilir, örneğin bazen tsavorite adı verilen yeşil garnet, orta kalitedeki bir zümrütten çok daha pahalı olabilmektedir. Dolayısıyla, değerli ve yarı değerli terimlerinin sadece ticari bağlamda kullanılması bu gibi durumlarda yanıltıcı olabilmektedir (Hurlbut ve Kammerling, 1991; Alden, 2009; Bauer, 2012). Elmas, zümrüt, yakut ve safir gibi değerli taşların fiyatlandırmasında da diğerlerinde olduğu gibi kütle önemli bir faktör olsa da kesim, renk, berraklık gibi estetik ve kalite unsurları daha önemli olabilmektedirler (Chu, 2001; Matlins ve Bonanno, 2009). Elmasın yüksek ekonomik değeri bu faktörlerin standartlaşmasını desteklemektedir ancak elmastan genellikle daha ucuz olan zümrüt, yakut ve safirde nadirlik seviyesi ve pazar talebi fiyatlandırmada çok daha büyük bir rol oynayabilmektedir.

Elmas fiyatlandırmasında dikkate alınan faktörler elmas ticaretini kompleks bir duruma dönüştürmektedir (Chu, 2001; Matlins ve Bonanno, 2009). Bu faktörlerin elmas fiyatına farklı oranlardaki etkileri, fiyatlandırmada ortaya çıkan tutarsızlıklar ve fiyatlandırma için mutlaka uzmana ihtiyaç duyulması gibi unsurlar, problemin üzerinde durulmasını gerektirmektedir (Chu, 2001; Matlins ve Bonanno, 2009; Bracking ve Sharife, 2014; Wolff, 2016; Mamonov ve Triantoro, 2018; Wu ve Cosguner, 2020). Çalışmada, “diamonds” veri seti ve özgün makine öğrenmesi yöntemleri ile elmas fiyat tahmini yapılarak problemin çözümüne odaklanılmaktadır (Agrawal, 2016). Aynı zamanda, yöntemlerin performansları yaygın olarak kullanılan ve çalışmada geliştirilen yenilikçi metrikler ile karşılaştırmalı şekilde incelenerek en uygun yönteme ulaşmaya çalışılmaktadır.

Bölüm 2’de, fiyat tahminine yoğunlaşılacak alanlardaki çalışmalara, değerli metaller ve değerli taşların fiyat tahmini çalışmalarına, farklı veya çalışmada kullanılan veri seti ile gerçekleştirilen özgün veya benzer türde elmas fiyat tahmini çalışmalarına ait değerlendirmeler verilerek çalışmamızın gerekliliği anlatılmaktadır. Bölüm 3’te, odaklanılan değerli taş elmas, birçok yönü ile ayrıntılı olarak ele alınmaktadır. Bu bölümde, elmas fiyatlandırma da kapsamlı şekilde açıklanmaktadır. Bölüm 4’te, çalışma düzeni, veri seti ve veri ön işleme detaylandırılmaktadır. Ayrıca, veri ön işleme sonunda elde edilen 4 farklı veri setine ait korelasyon ısı haritalarını gösteren veri setlerinin keşfi ve elmas fiyat tahmini için seçilen tüm modellerin ve kurulum aşamalarının tanıtıldığı

model seçimleri ve kurulumları da bu bölümde yer almaktadır. Bölüm 5’te tüm metrikler ayrı başlık altında sunulmakta ve her metriğin sonucu kendine ait başlıkta verilerek karşılaştırmalı incelemeler yapılmaktadır. Son bölümde sonuçlar paylaşılmaktadır.



2. KAYNAK ARAŞTIRMASI

Ekonomi ve ticarete temel rol oynayan fiyatlar, genellikle arz talep çerçevesinde belirlense de değişken koşullar fiyatlandırmaya farklı parametreler ve stratejileri dahil etmektedir (Sammot-Bonnici ve Channon, 2015). Bu durum fiyatlandırma işlemini karmaşıktırırken fiyat tahmini yapmayı da son derece zorlaştırmaktadır. Fiyat tahminindeki zorluk, çalışmaları da etkilemektedir. Fiyat tahmini çalışmaları, döviz, hisse senedi, borsa ve kripto para gibi alanlarda yoğunlaşmaktadır. Döviz alanında, Ince ve Trafalis (2006) tarafından yapılan çalışma gibi döviz kuru tahminlerine yönelik çalışmalar ön plana çıkmaktadır. Hisse senedi alanında, Leung ve ark. (2014) yaptığı çalışma gibi hisse senetlerinin fiyat tahminlerine yönelik çalışmalar değer taşımaktadır. Borsa alanında, Patel ve ark. (2015) yaptığı çalışma gibi ulusal ya da uluslararası borsa endekslerinin tahminlerine yönelik çalışmalar yer almaktadır. Kripto para alanında ise Chen ve ark. (2021) yaptığı çalışma gibi kripto paraların fiyat tahminlerine yönelik çalışmalar giderek artmaktadır.

Değerli metaller ve değerli taşların fiyat tahmini çalışmalarının büyük çoğunluğunda, altın ve gümüşün ayrı ayrı veya birlikte ele alındığı gözlemlenmektedir (Manjula ve Karthikeyan, 2019; Sadowsky, 2021). Bu çalışmalardaki tahminler, altın ve gümüşün geçmiş zamanlardaki fiyatları baz alınarak yapılmaktadır (Manjula ve Karthikeyan, 2019; Goel ve ark., 2022). Pandey ve ark. (2019) yaptığı çalışmada da bu bağlamda hareket edilmektedir ancak söz konusu yaklaşım genellikle yeterli olsa da değerli taşlar ve özellikle elmas için eksik kalmaktadır (Bracking ve Sharife, 2014; Wolff, 2016; Mamonov ve Triantoro, 2018; Wu ve Cosguner, 2020). Bu durum, elmas başta olmak üzere değerli taşların fiyat tahmini çalışmalarına yoğunlaşılmasını gerektirmektedir. Mamonov ve Triantoro (2018) ile Chaijunla ve Taninpong (2021) tarafından yapılan çalışmalar gibi farklı veri setleri ile özgün sonuçlara ulaşılan birkaç elmas fiyat tahmini çalışması mevcuttur. Ayrıca, çalışmada kullanılan “diamonds” veri setiyle yapılan değişik türde çalışmalar da görülebilmektedir (Swain ve ark., 2023; Basha ve ark., 2024; Zabala-Blanco ve ark., 2024). Bunlar hem elmas fiyat tahmini çalışmalarının hem de kullanılan veri setinin önemini vurgulamaktadır.

Çalışmada kullanılan “diamonds” veri seti ile gerçekleştirilen, paralel veya karşıt sonuçlar elde edilmiş benzer türde elmas fiyat tahmini çalışmaları bulunmaktadır. Mihir ve ark. (2021) yaptığı çalışmada doğrusal (linear), karar ağacı (decision tree), rastgele orman (random forest), CatBoost (kedi güçlendirme), SVM (Destek Vektör Makinesi

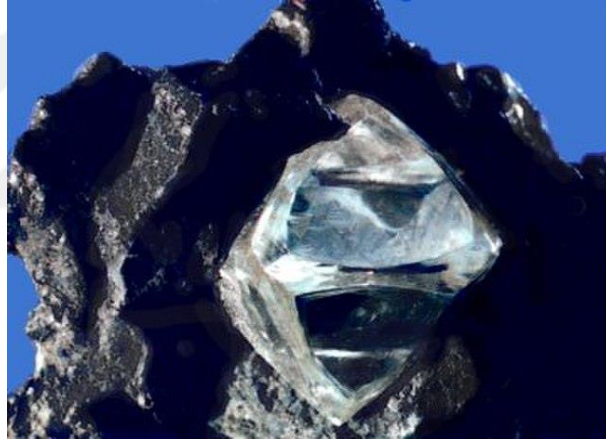
(Support Vector Machine)), k-en yakın komşu (k-nearest neighbors), huber kaybı (huber loss), ekstra ağaçlar (ekstra trees), pasif agresif (passive aggressive), bayesçi (bayesian) ve XGBoost (Aşırı Gradyan Güçlendirme (eXtreme Gradient Boosting)) ile elmas fiyat tahmini yapılarak en uygun yöntemin 271.10 MAE (Ortalama Mutlak Hata (Mean Absolute Error)) skoruna sahip kedi güçlendirme olduğu belirtilmektedir. Basysyar ve Dwilestari (2022) tarafından yapılan çalışmada sırt (ridge), LASSO (En Küçük Mutlak Daralma ve Seçim Operatörü (Least Absolute Shrinkage and Selection Operator)), Elastik Net (ElasticNet) ve rastgele orman ile elmas fiyat tahmini yapılarak daha zayıf bir MAE skoru (538.72) ile en uygun yöntemin rastgele orman olduğu gösterilmektedir. Fitriani ve ark. (2022) yaptığı çalışmada ise sadece k-en yakın komşu ve LASSO ile elmas fiyat tahmini yapılarak en uygun yöntemin $0.906 R^2$ (R-Kare (R-Squared), Belirlilik Katsayısı (Coefficient of Determination)) skoruna sahip k-en yakın komşu olduğu tespit edilmektedir. Basysyar ve Dwilestari (2022) tarafından yapılan çalışma ile paralellik gösteren Sharma ve ark. (2021) yaptığı çalışmada doğrusal, LASSO, sırt, karar ağacı, rastgele orman, Elastik Net, AdaBoost (Adaptif Güçlendirme (Adaptive Boosting)) ve gradyan güçlendirme (gradient boosting) ile elmas fiyat tahmini yapılarak en uygun algoritmanın $0.979 R^2$ skoruna sahip rastgele orman olduğu ortaya konulmaktadır. Fajar ve ark. (2024) yaptığı çalışmada da rastgele orman, k-en yakın komşu ve AdaBoost ile elmas fiyat tahmini yapılarak aynı sonuca ulaştıkları ve $0.981 R^2$ skoruna sahip rastgele orman yönteminin en uygun yöntem olduğu sonucuna ulaşılmaktadır. Bu çalışmalarda, derin öğrenme dahil edilmeden problemin çözülmek istenmesi, performans karşılaştırmaları için ihtiyaç duyulan eşit koşullara ve çalışma bazlı gerekliliklere yeterince dikkat edilmemesi, bazı performans metriklerinin değerlendirilmemesi ve performans karşılaştırmalarının ilk veya tekil sonuçlar üzerinden gerçekleştirilmesi gibi belirgin noktalar bulunmaktadır. Çalışmada, bu noktalara büyük önem verilerek hassas ve detaylı şekilde yaklaşılmaktadır. Böylece hem makine öğrenmesinden hem de derin öğrenmeden yararlanarak daha kapsamlı bir perspektif benimsenmektedir. Performans karşılaştırmalarında tutarlı ve doğru sonuçların elde edilebilmesini sağlayan eşit koşulların titizlikle oluşturulmasına ve çalışma bazlı gerekliliklerin de göz önünde bulundurulmasına, özen gösterilmektedir (Ünsal ve ark., 2022). Yaygın olarak kullanılan performans metriklerine ek olarak çalışmada geliştirilen yenilikçi metriklerin de dahil edilmesi, incelemeleri literatürde ilk olacak şekilde güçlendirmektedir. Çoklu sonuçlar ile gerçekleştirilen değerlendirmeler ise güvenilir ve etkili performans karşılaştırmalarına

imkân sunmakta ve en doğru sonuca yaklaştırmaktadır. Tüm bunlar çalışmanın özgünlüğünü ve gerekliliğini yansıtmaktadır.



3. ELMAS

Çok az türde yabancı maddenin onu kirletebilmesine imkân veren kristal yapıda düzenlenmiş son derece yoğun bir atom düzenine sahip elmas, karbon elementinin katı formudur (Hershey, 1940; Robertson, 1994; Shirey ve Shigley, 2013; Cartigny ve ark., 2014). Elmas kübik olarak adlandırılan atomları sayesinde, Şekil 3.1.'de gösterildiği gibi genellikle diğer değerli taşlara göre kaliteli ve büyük boyutlarda bulunmaktadır. En az 3000 yıl önce ilk olarak Hindistan'da çıkarıldıkları ve tanındıkları düşünülmektedir (Hershey, 1940). Bu bölgedeki Dravidian halkı da nehir yataklarında topladıkları kaba görünümlü siyah elmasları, ticarete değişim aracı olarak kullanmışlardır. Hatta, elmas kütlesini carob ağacı (keçiboynuzu) tohumlarına göre ölçtükleri için, şu anda kuyumculukta yaygın kullanılan bir birim olan karatın, “carob” sözcüğünden türediği ileri sürülmektedir (Bakırcı, 2023).



Şekil 3.1. Doğal elmas

Dünya'nın manto tabakasının farklı derinliklerinde çeşitli mineraller, yüksek basınç ve sıcaklık altında karbon içeren sıvılar tarafından çözünerek yerlerini doğal elmaslara bırakırlar (Hershey, 1940; Robertson, 1994; King, 2012; Shirey ve Shigley, 2013; Cartigny ve ark., 2014). 1-3,5 milyar yaşında oldukları tahmin edilen elmaslar, milyonlarca yıl boyunca saklı kaldıktan sonra yaklaşık 100 milyon yıl önce başlayan volkanik patlamalarla yüzeye doğru taşınarak kimberlit ve lamproit olarak bilinen magmatik kayaların içinde birikmişlerdir (Hershey, 1940; Robertson, 1994; King, 2012; Shirey ve Shigley, 2013). Yüksek oranda sıkıştırılmış kömürden oluştuklarına dair büyük bir yanlışlık bulunmaktadır. Kömür, gömülü tarih öncesi bitkilerden meydana gelirken,

tarihlenen elmasların çoğu ilk kara bitkilerinden çok daha eskidir (King, 2012; Shirey ve Shigley, 2013; Cartigny ve ark., 2014). 1772 yılında Fransız bilim adamı Antoine Lavoisier, güneş ışınlarını mercek aracılığıyla oksijen atmosferindeki bir elmas üzerinde yoğunlaştırmış ve yanma sonucu ortaya çıkan tek ürünün karbondioksit olduğunu göstererek elmasın karbondan oluştuğunu kanıtlamıştır (Lavoisier, 1772). 1797'de bu deneyi tekrarlayıp genişleten İngiliz kimyager Smithson Tennant, yanan elmas ve grafitin aynı miktarda gaz açığa çıkardığını göstererek bu maddelerin kimyasal eşdeğerliğini ortaya koymuştur (Tennant, 1797). Bilinen diğer katı karbon formu olan grafit, oda sıcaklığında ve basınçta kimyasal olarak stabil durumdadır ancak elmas metastabil durumda olduğu için bazı koşullar altında düşük oranda grafitte dönüşebilmektedir. Aynı şekilde grafitten elmasa dönüşüm de mümkündür (Cartigny ve ark., 2014).

Kimyasal yapısı ve mükemmel biyolojik uyumluluğu sayesinde elmasın, olağanüstü sertlik, üstün dayanıklılık ve çok yüksek termal iletkenlik gibi doğal maddeler arasında öne çıkmasını sağlayan eşsiz özellikleri bulunmaktadır (Hershey, 1940; Robertson, 1994; King, 2012; Chen ve Zhang, 2013; Shirey ve Shigley, 2013; Cartigny ve ark., 2014; Wang ve ark., 2020; Smith ve ark., 2021). Bu özellikler elmasın, kesme ve cilalama araçları gibi çoğu endüstriyel işlemleri gerçekleştirebilmesini ve farklı disiplinlerden ilgi görmesini sağlamaktadır (Chen ve Zhang, 2013; Wang ve ark., 2020; Smith ve ark., 2021). Nadirliği ve pahalılığı dolayısıyla farklı metodolojilerle yapayları üretilse de optik teknikler ve termal iletkenlik ölçümleri ile doğalları ayırt edilebilmektedir (Werner ve Locher, 1998; Prinsloo ve ark., 2011). Dünya'da elmasın nadirliği nispeten geçerli bir yargı olsa da bu durumun uzayda farklılık gösterdiği düşünülmektedir. Karbon açısından zengin yıldızlarda bolca mevcut olduğu hatta bazı güneş dışı gezegenlerin neredeyse tamamının elmastan meydana geldiği öne sürülmektedir (Planck, 2011).

Elmaslar, antik çağlarda süs ve dekoratif eşyalarda ve eski Hindistan'da dini ikonlarda kullanıldıkları günden beri değerli taşlar olarak değerlendirilmektedirler (Hershey, 1940; Alden, 2009). Sadece diğer elmaslar tarafından çizilebildiği için cilasını iyi koruyan elmasların sertliği, değerli taş olarak kullanımlarına katkıda bulunmaktadır. Diğer pek çok değerli taşın aksine, çizilmeye karşı dayanıklılığı nedeniyle günlük kullanıma çok uygundur bu yüzden her gün takılan nişan yüzüğü ve alyanslarda tercih edilerek popülaritelerini giderek artırmaktadırlar (Bauer, 2012). Görece nadirliği ve maddi değeri dolayısıyla mücevher niteliğinde kullanılan elmasın, insan gözüne hoş görünmesini sağlayan optik bir özelliği de bulunmaktadır. "Adamantine" adıyla bilinen

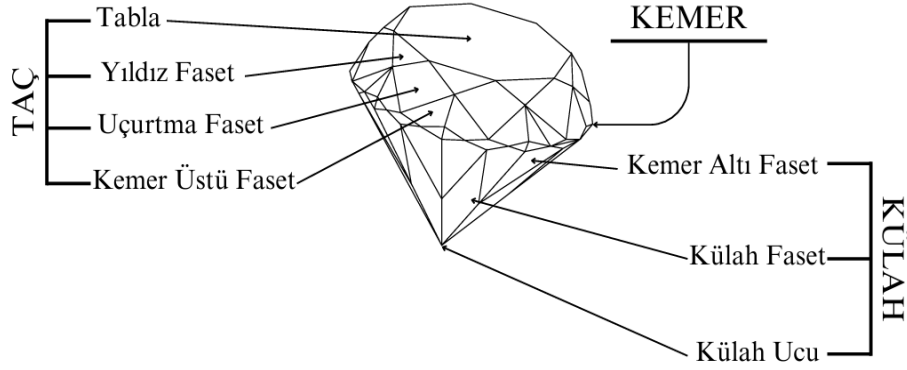
özellik, metalik olmayan en yüksek parlaklığı ifade eder (Sharma ve ark., 2021). Elmasın yüzeyine çarpan ışık, bu özellik sayesinde büyük oranda yansıtılarak ışıldamasını mümkün kılar. Elmasın ışıltısı da dayanıklılığı, modası, estetik güzelliği ve güçlü pazarlama stratejileri gibi günden güne değerini arttıran faktörlerdendir. Elmasın popülaritesine en büyük katkı, Güney Afrika'da ilk büyük ölçekli elmas madenini kuran De Beers Consolidated Mines tarafından yapılmaktadır. 1930'larda başlayan ve 20. yüzyılın ortalarına kadar devam eden bir reklam kampanyası aracılığıyla De Beers, elmasları nişan sürecinin önemli parçası ve imrenilen statü sembolü haline getirdi (Epstein, 1982). Yaygın olarak, elmasın değerinin nadirliğinden kaynaklandığına inanılsa da alexandrite gibi değerli taşlarla karşılaştırıldığında oldukça yaygındır ve her yıl tonlarca çıkarıldığı bilinmektedir (Epstein, 1982; Yarnell, 2004). Bu yüzden, elmas fiyatlandırması üzerinde dikkatle durulması gereken bir konudur.

3.1. Elmas Fiyatlandırma

Fiyatlar, bir ürün veya hizmet karşılığında verilen para ya da kaynağın belirlenmesinde mekanizma görevi görmektedirler. Fiyatlandırma yaparken kullanılan farklı parametreler ve stratejiler fiyatlandırma işlemi karmaşılaştırırken fiyat tahmini yapmayı son derece zorlaştırmaktadır (Sammuto-Bonnici ve Channon, 2015). Elmas fiyatlandırmasında da sıkça karşılaşılan bu durum, birçok soruna yol açmaktadır (Bracking ve Sharife, 2014; Wolff, 2016; Mamonov ve Triantoro, 2018; Wu ve Cosguner, 2020). Yirminci yüzyılın ortalarında gemologlar, elmaslar başta olmak üzere değerli taşları, kıymetleri açısından en önemli özelliklere göre değerlendirmek için yöntemler geliştirmişlerdir. Bu yöntemlerden 1953 yılında GIA (Amerika Gemoloji Enstitüsü (Gemological Institute of America)) tarafından geliştirilen ve “4C (kesim (cut), renk (color), berraklık (clarity), karat (carat))” olarak bilinen standart, uluslararası alanda kabul görmektedir. 4C, başta elmas olmak üzere tüm değerli taşların temel tanımlayıcısı olarak kullanılmaktadır (Chu, 2001; GIA, 2002; Matlins ve Bonanno, 2009). Endüstride kullanılan elmaslarda, 4C’de belirtilen özellikler önem arz etmese de mücevheratta kullanılan elmaslara, 4C’nin her niteliği ayrı ayrı ele alınarak değer biçilmektedir (Chu, 2001; Matlins ve Bonanno, 2009; Bracking ve Sharife, 2014; Wang ve ark., 2020; Smith ve ark., 2021). Bir elmasın kesim, renk, berraklık ve karat özelliklerini bilmek fiyat aralığını makul şekilde oluşturmak için yeterlidir (Chu, 2001; Matlins ve Bonanno, 2009; Bracking ve Sharife, 2014). Bireysel elmas satın alacak tüketicilere 4C’yi dikkate

almaları tavsiye edilmektedir. Bazı durumlarda, elmasın, kaynağı, değerlendirildiği gemoloji enstitüsü, markası ve pazar talebi de değerini etkileyebilmektedir (Matlins ve Bonanno, 2009). Karat ve kesim matematiksel olarak tanımlanmaktadır ancak berraklık ve renk gemoloji eğitimi almış kişiler tarafından değerlendirilmektedir ve bu nedenle yorumda hafif farklılıklara açıktır.

Elmasın kaba bir taş olarak başlangıç formundan mücevher haline gelene kadar şekillendirilme ve cilalanması, elmas kesim işlemi olarak adlandırılmaktadır. Elmas kesim işlemi yapılarak elmaslara birçok farklı şekil verilebilmektedir. Elmas kesim işlemleri yüzlerce yıl boyunca geliştirildi ve en büyük başarı 1919'da Marcel Tolkowsky tarafından elde edildi. Bir elmasa yukarıdan bakıldığında, ışığın geri dönmesi ve saçılması için ideal şekli hesaplayarak Şekil 3.2.'deki gibi yuvarlak (round) şeklini geliştirdi (Tolkowsky, 1919). Yuvarlak şekil, en popüler elmas şeklidir ve piyasadaki elmasların çoğunluğu bu şekildedir. Diğer elmas şekilleri genellikle, moda ve marka farklılaştırma girişimleri yüzünden ortaya çıkmaktadır. Maksimum ışık miktarını yansıtmak için yuvarlak şekil elmasın kesilmesi gereken açılar ve uzunluk oranlarına ilişkin matematiksel kurallar vardır. Diğer şekil taşlar, bu düzene yeteri kadar uyum sağlayamazlar. Bu yüzden, fiyatlandırmada 4C'nin tam olarak uygulanabileceği elmas, yuvarlak şeklinde kesilmiş olandır (Chu, 2001; Matlins ve Bonanno, 2009).



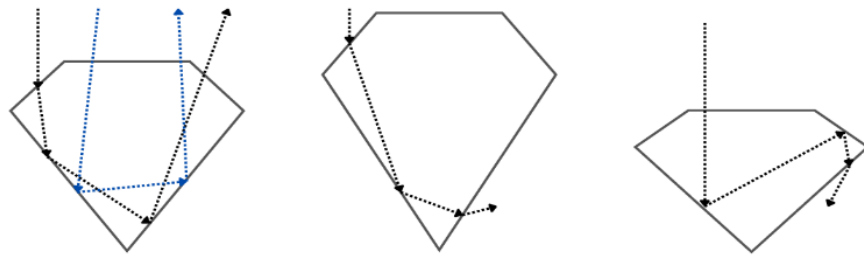
Şekil 3.2. Yuvarlak elmas ve bölümleri

Bölümleri Şekil 3.2.'de gösterilen yuvarlak şekil elmas, üç kısım ve 57 fasetten oluşmaktadır (Tolkowsky, 1919). Elması sağlam tutan ve cilalı ya da fasetli olan orta kısma kemer denir. Kalın kemerli bir elmas daha mat görünür, ince kemerli elmas ise baskıya karşı daha az dayanıklıdır. Kemerin üst tarafında taç kısmı yer almaktadır. Taç bölümünde 1 tabla, 8 yıldız faseti, 8 uçurtma (bezel) faseti ve 16 kemer üstü faseti olmak

üzere toplam 33 faset bulunmaktadır. Taç daha fazla faset sayısına sahip olduğu için ışığı çeşitli renklere dönüştürerek yansıma ve parlaklık vermektedir. Kemerin alt tarafında ise 16 kemer altı faseti ve 8 külâh faseti olmak üzere toplam 24 faset bulunan külâh kısmı yer almaktadır. Külâh ne kadar doğru açıyla kesilmiş ise, elmasa giren ışık o oranda yansyarak taç bölümünden çıkmaktadır. Böylece, elmasta optimum parlaklık sağlanmaktadır. Kimi zaman külâh ucuna da faset atılabilmektedir. Bu faset ile elmasın en hassas kısmı olan külâh ucunda oluşabilecek hasarlar önlenmeye çalışılmaktadır. Elmas, Dünya'nın en sert doğal maddesi olsa bile, belirli yönlerden gelen darbelere karşı dayanıksız ve kırılabilir olabilmektedir.

3.1.1. Kesim

Elmasın kesim kalitesi, 4C'de en önemli kriter olarak kabul edilmektedir (Chu, 2001; GIA, 2002; Matlins ve Bonanno, 2009). İdeal kesilmiş elmasın karatının daha yüksek, berraklığının ve renginin ise gerçekte olduğundan daha kaliteli görünebileceği genel olarak kabul edilmektedir. Elmasın kesim kalitesi, ışığı yansıtma ve kırma yeteneğini belirlemektedir. Sadece birkaç derece hizada kesilmiş fasetlere sahip elmas, düşük performans gösteren bir taş olarak kalabilmektedir. İdeal kesim kalitesine sahip elmas, ışığı yukarı doğru yansıtmalı ve üstten bakıldığında beyaz görünmelidir (Tolkowsky, 1919). Kesimler arasındaki fark Şekil 3.3.'te gösterilmektedir.



Şekil 3.3. Elmas kesimleri

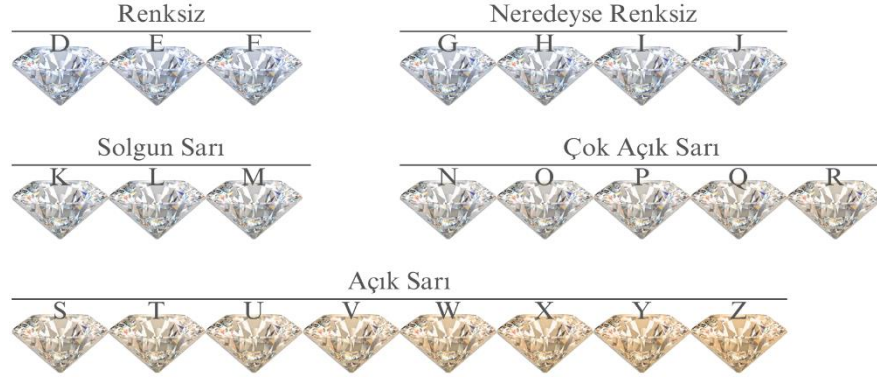
Elmasın ideal oranlarına ilişkin birçok farklı teori olsa da genelde yuvarlak elmas şeklini geliştiren Marcel Tolkowsky tarafından hesaplanan boyutlar dikkate alınmaktadır. Elmasın kemerinin uçtan uca kadar olan ölçüsünü çap, tabladan külâh ucuna kadar olan yüksekliğini derinlik olarak kabul ettiğimiz durumda, ideal oranlar şu şekildedir: tabla yüzdesi (tabla çapının çapa bölümü) %53, derinlik yüzdesi (derinliğin çapa bölümü)

%59.3 (k lah ucu y ksekl đi ve kemer kalınlıđına iliřkin ayarlamalar hari ), k lah a ısı (kemer ile k lah fasetleri arasındaki a ı) 40.45 , ta  a ısı (kemer ile tacın u urtma fasetleri arasındaki a ı) 34.30 , k lah y zdesi (k lah derinliđinin  apa b l m ) %43.1 ve ta  y zdesi (ta  derinliđinin  apa b l m ) %16.2 (Tolkowsky, 1919). Elmasın  zellikleri, Tolkowsky tarafından hesaplanan boyutlardan ne kadar uzaksa ıřık o kadar az yansımaktadır.

Elmas kesim kalitesi, elmas sahipleri arasında genellikle kesim oranları belirtilerek anlatılmamaktadır. Bunun yerine ideal kesime en uzaktan en yakına olmak  zere sırasıyla “Fair”, “Good”, “Very Good”, “Premium” ve “Ideal” olarak sınıflandırılarak belirtilmektedir. Sınıflandırma sayısı ve isimleri deđiřiklik g sterse de mantık s rekli aynıdır. T m  zellikleri aynı, sadece kesimleri farklı iki elmas arasında bile devasa fiyat deđiřimleri g r lebilmektedir. Kesimin, fiyata olan etki oranı her elmasta deđiřebilmektedir (Chu, 2001; Matlins ve Bonanno, 2009; Bracking ve Sharife, 2014; Wolff, 2016; Mamonov ve Triantoro, 2018; Wu ve Cosguner, 2020).

3.1.2. Renk

Kimyasal olarak saf ve yapısal olarak m kemm l olan bir elmasın, hi bir renk tonu veya renk barındırmayarak tamamen řeffaf halde olması beklenir. Genelde, m cevher boyutundaki dođal elmasların  ođu kusurludur. Elmasın rengi, kimyasal kirliliđinden ve yapısal kusurlarından etkilenebilmektedir (Hershey, 1940). Tonuna ve yođunluđuna bađlı olarak, elmasın rengi deđerini azaltabilir veya artırabilir.  rneđin, beyaz elmasların  ođunda daha sarı bir ton algılanabildiđi zaman fiyatta indirim  gidilirken, yođun pembe veya mavi elmaslar  nemli  l  de daha deđerli olabilmektedir. En yaygın yabancı madde olan nitrojen, elmasın yapısındaki az miktardaki karbon atomunun yerini alır ve sarı ila kahverengi arasında bir renk tonuna sebebiyet verir (Walker, 1979). Bu etki, en nadir elmaslar hari  hemen hemen t m beyaz elmaslarda mevcuttur.



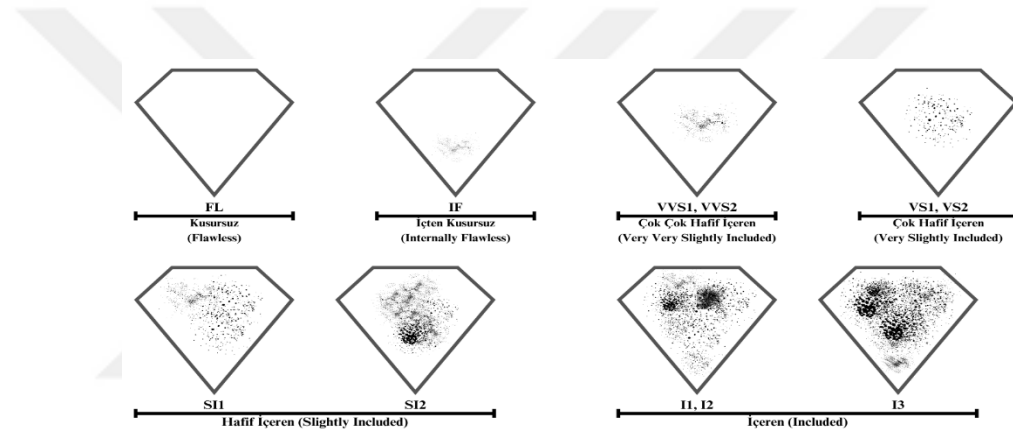
Şekil 3.4. Elmas renkleri

GIA tarafından oluşturulan derecelendirme sistemi, elmasların rengini D’den Z’ye kadar sınıflandırmaktadır. Şekil 3.4.’te gösterilen sistem, standartlaştırılmış ve dikkatle kontrol edilen aydınlatma koşullarının yanı sıra, bilinen renk derecelerine sahip doğal elmaslardan oluşan bir referans seti kullanılmaktadır (GIA, 2002). “D-F” dereceliler “renksiz (ekstra beyaz)”, “G-J” dereceliler “neredeyse renksiz (beyaz)”, “K-M” dereceliler “solgun sarı (beyaza yakın)”, “N-R” dereceliler “çok açık sarı (çok açık renkli)” ve “S-Z” dereceliler “açık sarı (açık renkli)” olarak kabul edilmektedir. Yüksek renk dereceli elmaslar, daha nadirdir ve yoğun talep görmektedirler. Bu nedenle, düşük renk dereceli elmaslara göre daha pahalıdır. “Z” dereceli elmaslar da nadirdir ve parlak sarı rengi oldukça değerlidir. Sarı veya kahverengi tonlarının aksine diğer renkteki elmaslar, daha nadir ve değerlidir. Soluk pembe veya mavi renk tonu bile elmasın değerini artırabilirken, yoğun bir renk genellikle daha arzu edilir olarak düşünülür ve yüksek fiyatlar biçilmektedir. Çeşitli yabancı maddeler ve yapısal kusurlar pırlantalarda sarı, pembe, mavi, kırmızı, yeşil, kahverengi ve diğer tonlar dahil olmak üzere farklı renklere neden olabilmektedir. Ayrıca, karbonado adı verilen siyah veya gri renkli çok kristalli elmas içeren bir tür taş da mevcuttur. Elmasın en sert şeklidir ve aşındırıcı olarak kullanılmaktadır ancak bazen değerli taş olarak da tercih edilmektedir.

Elmas rengi, 4C’de kesimden sonra en önemli kriter olarak düşünülmektedir. Tüm özellikleri aynı, sadece renkleri farklı iki elmas arasında bile devasa fiyat değişimleri görülebilmektedir. Rengin, fiyata olan etki oranı her elmasta değişebilmektedir (Chu, 2001; Matlins ve Bonanno, 2009; Bracking ve Sharife, 2014; Wolff, 2016; Mamonov ve Triantoro, 2018; Wu ve Cosguner, 2020).

3.1.3. Berraklık

Berraklık, elmasın yapısal iç kalıntılarının (kusurlarının) ölçüsüdür. Kalıntılar, beyazımsı veya bulanık görünebilen küçük çatlaklar, yabancı madde kristalleri ve başka elmaslara ait kristaller gibi yapısal kusurlar olabilir. Kalıntıların sayısı, boyutu, rengi, göreceli konumu, yönelimi ve görünürlüğü, elmasın yoruma dayalı berraklığını etkileyebilmektedir. GIA'nın, elmasa 10 kat büyütme altında bakıldığında eğitimli bir profesyonel tarafından görülebilen kalıntıları temel alan berraklığı derecelendirmek için geliştirdiği Şekil 3.5.'te gösterilen sistemde netlik ölçeği, “kusursuz (FL)”, “içten kusursuz (IF)”, “çok çok hafif içeren (VVS)”, “çok hafif içeren (VS)”, “hafif içeren (SI)” ve “içeren (I)” olarak sınıflandırılmaktadır (GIA, 2002).



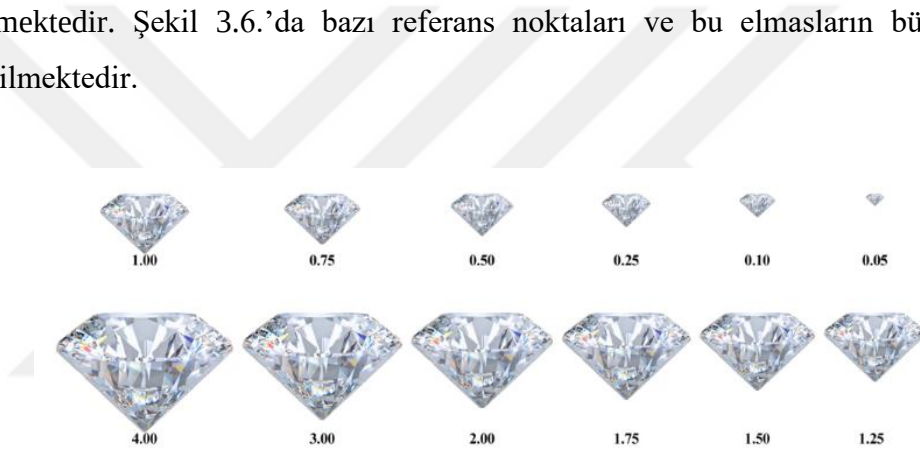
Şekil 3.5. Elmas berraklıkları

Çıkarılan tüm elmasların yalnızca yaklaşık %20'si değerli taş olarak kullanılmaya uygun görülecek berraklık derecesine sahiptirler, diğer elmaslar endüstriyel kullanıma ayrılmaktadırlar (Chen ve Zhang, 2013). Değerli taş olarak kullanılan elmasların önemli kısmı, bir veya daha fazla kalıntı içerirler. Bu kalıntıların çoğunun, çeşitli kuyumculuk teknikleri ile gizlenmesi mümkündür ancak büyük kalıntılar elmasın ışığı iletme ve dağıtma yeteneğini etkileyebilmektedir (Alden, 2009). Yüzeye yakın büyük çatlaklar, kırılma olasılığını da artırabilmektedir.

Elmas berraklığı, 4C'de kesim ve renkten sonra en önemli kriter olarak düşünülmektedir. Tüm özellikleri aynı, sadece berraklıkları farklı iki elmas arasında bile devasa fiyat değişimleri görülebilmektedir. Berraklığın, fiyata olan etki oranı her elmasta değişebilmektedir (Chu, 2001; Matlins ve Bonanno, 2009; Bracking ve Sharife, 2014; Wolff, 2016; Mamonov ve Triantoro, 2018; Wu ve Cosguner, 2020).

3.1.4. Karat

Karat, elmasın kütlesini belirtmek için kullanılmaktadır. Bir karat 200 miligrama denk gelmektedir. Elmaslarda karat dışındaki diğer özelliklerin tümü eşdeğer olduğunda, karat arttıkça fiyat da artmaktadır. Karat başına fiyat, artan boyutla birlikte doğrusal olarak artmaz çünkü büyük elmaslar daha nadir ve mücevheratta daha çok tercih edilmektedir. Bunun yerine, referans noktası olarak belirlenen karat ağırlıklarında keskin sıçramalar bulunmaktadır. Referans karatlarından daha ağır elmaslara olan talep, daha hafif olanlara göre çok daha yüksektir. Örnek olarak, 0.99 karatlık elmasın karat başına fiyatı, talep farklılıkları nedeniyle 1.01 karatlık elmasa göre çok daha düşük olabilmektedir. Şekil 3.6.'da bazı referans noktaları ve bu elmasların büyüklükleri gösterilmektedir.



Şekil 3.6. Elmas referans karatları

Tüm özellikleri aynı, sadece karatları farklı iki elmas arasında bile devasa fiyat değişimleri görülebilmektedir. Karatın, fiyata olan etki oranı her elmasta değişebilmektedir (Chu, 2001; Matlins ve Bonanno, 2009; Bracking ve Sharife, 2014; Wolff, 2016; Mamonov ve Triantoro, 2018; Wu ve Cosguner, 2020).

Haftalık elmas fiyat listesi olan Rapaport Elmas Raporu, New York Rapaport Group kurucusu Martin Rapaport tarafından farklı elmas grupları için yayınlanmaktadır (Rapaport, 2024). Bu listeler perakende fiyat tabanı olarak kabul edilebilmektedir.

4. MATERYAL VE YÖNTEM

4.1. Çalışma Düzeni

Çalışma süreci boyunca, veri bilimi ve makine öğrenmesi alanlarında yaygın olarak kullanılan bir programlama dili olan Python kullanılmaktadır. Python, bilimsel hesaplama ve makine öğrenmesi için kapsamlı kütüphane desteği sağlamasıyla ön plana çıkmaktadır. Çalışmada da Pandas, Scikit-learn, Matplotlib, Seaborn, Numpy, XGBoost ve Plotly kütüphanelerinden yararlanılmaktadır (Lee, 2019; Raschka ve Mirjalili, 2019). Ana geliştirme ortamı olarak kod çalıştırma, görselleştirme desteği ve etkileşimli arayüz gibi özellikleri bulunan Jupyter Notebook tercih edilmektedir (Lee, 2019). Ayrıca çalışma, HP Victus 16-D1048NT 68P33EA model laptop üzerinde gerçekleştirilmektedir (HP, 2022).

4.2. Veri Seti

Çalışmada kullanılan “diamonds” veri seti, kullanıcıların veri kümeleri yayınlamasını, bulmasını, modeller keşfedip oluşturmasını, diğer veri bilimcileri ve makine öğrenimi mühendisleriyle çalışmasını sağlayan veri bilimi ortamı Kaggle internet sitesi aracılığı ile elde edilmiştir (Agrawal, 2016; Lee, 2019). İlk ve son beş satırı Şekil 4.1.’de gösterilen “diamonds” veri seti 53940 satır ve dizin sayacı hariç 10 sütun (değişken) içeren bir veri çerçevesidir. Şekil 4.1.’de veri setine ait bilgiler ve değişkenlerin veri tipleri de gösterilmektedir.

	Unnamed: 0	carat	cut	color	clarity	depth	table	price	x	y	z	
	0	1	0.23	Ideal	E	SI2	61.5	55.0	326	3.95	3.98	2.43
	1	2	0.21	Premium	E	SI1	59.8	61.0	326	3.89	3.84	2.31
	2	3	0.23	Good	E	VS1	56.9	65.0	327	4.05	4.07	2.31
	3	4	0.29	Premium	I	VS2	62.4	58.0	334	4.20	4.23	2.63
	4	5	0.31	Good	J	SI2	63.3	58.0	335	4.34	4.35	2.75
	...	...	...	...	...	...	...	...	...	...	...	...
53935	53936	0.72	Ideal	D	SI1	60.8	57.0	2757	5.75	5.76	3.50	
53936	53937	0.72	Good	D	SI1	63.1	55.0	2757	5.69	5.75	3.61	
53937	53938	0.70	Very Good	D	SI1	62.8	60.0	2757	5.66	5.68	3.56	
53938	53939	0.86	Premium	H	SI2	61.0	58.0	2757	6.15	6.12	3.74	
53939	53940	0.75	Ideal	D	SI2	62.2	55.0	2757	5.83	5.87	3.64	

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 53940 entries, 0 to 53939
Data columns (total 11 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Unnamed: 0  53940 non-null  int64
1   carat       53940 non-null  float64
2   cut         53940 non-null  object
3   color       53940 non-null  object
4   clarity     53940 non-null  object
5   depth       53940 non-null  float64
6   table       53940 non-null  float64
7   price       53940 non-null  int64
8   x           53940 non-null  float64
9   y           53940 non-null  float64
10  z           53940 non-null  float64
dtypes: float64(6), int64(2), object(3)
memory usage: 4.5+ MB
```

53940 rows x 11 columns

Şekil 4.1. Veri seti

Veri setinde bulunan ilk sütun tamsayılarla dolu dizin sayacıdır. İkinci sütun ilk değişken “carat”, elmasların kütlesi olan karatı ondalıklı sayılarla ifade etmektedir. Üçüncü sütun ikinci değişken “cut”, elmasların kesimini düşük kalitelden yüksek kaliteliye doğru sırasıyla “Fair”, “Good”, “Very Good”, “Premium” ve “Ideal” sınıflarıyla tanımlamaktadır. Dördüncü sütun üçüncü değişken “color”, elmasların rengini düşük kalitelden yüksek kaliteliye doğru sırasıyla “J”, “I”, “H”, “G”, “F”, “E” ve “D” sınıflarıyla göstermektedir. Beşinci sütun dördüncü değişken “clarity”, elmasların berraklığını düşük kalitelden yüksek kaliteliye doğru sırasıyla “I1”, “SI2”, “SI1”, “VS2”, “VS1”, “VVS2”, “VVS1”, “IF” sınıflarıyla nitelendirmektedir. Altıncı sütun beşinci değişken “depth”, elmasların derinlik yüzdesini ondalıklı sayılarla belirtmektedir. Bu değişken, elmasın derinlik ölçüsünün (“z”), elmasın uzunluk (“x”) ve genişlik (“y”) ölçüsünün ortalamasına bölünmesi ve 100 ile çarpılmasıyla elde edilmektedir. Yedinci sütun altıncı değişken “table”, elmasın tabla genişliğinin ortalama kemer genişliğine bölünmesi ve 100 ile çarpılmasının sonucu olarak verilmektedir. Sekizinci sütun yedinci değişken “price”, elmasların fiyatını USD (Amerikan doları (United States dollar)) cinsinden tamsayı olarak temsil etmektedir. Dokuzuncu sütun sekizinci değişken “x”, elmasların uzunluğunu; onuncu sütun dokuzuncu değişken “y”, elmasların genişliğini; on birinci sütun onuncu değişken “z”, elmasların derinliğini; ondalıklı olarak milimetre cinsinden yansıtmaktadır (Agrawal, 2016).

4.3. Veri Ön İşleme

Veri kodlama ve veri normalizasyon adımlarına geçilmeden önce veri seti üzerinde çeşitli temel ön işleme adımları gerçekleştirilmektedir. Bunlar sırasıyla şunlardır:

- (1) İlk sütun dizin sayacının kaldırılması.
- (2) Uzunluk (“x”) sütununda 0 veya 30’dan büyük değerlerin bulunduğu satırların kaldırılması.
- (3) Genişlik (“y”) sütununda 0 veya 30’dan büyük değerlerin bulunduğu satırların kaldırılması.
- (4) Derinlik (“z”) sütununda 0 veya 30’dan büyük değerlerin bulunduğu satırların kaldırılması.
- (5) Derinlik yüzdesi (“depth”) sütununda 45’ten küçük veya 75’ten büyük değerlerin bulunduğu satırların kaldırılması.

- (6) Tabla yüzdesi (“table”) sütununda 40’tan küçük veya 80’den büyük değerlerin bulunduğu satırların kaldırılması.
- (7) Sütunların mantıksal düzende (“carat”, “cut”, “color”, “clarity”, “depth”, “table”, “x”, “y”, “z”, “price”) sıralanması.
- (8) Veri setinin 10 defa rastgele biçimde karıştırılması.

Bu temel adımlar, mantıksız veri noktalarının çıkarılması, ham verilerin modelleme sürecine uygun hale getirilmesi ve doğruluğun artırılması için önemlidir (Lee, 2019; Raschka ve Mirjalili, 2019). Bu 8 temel veri ön işleme adımı ile 1 sütun ve 30 satır kaldırılarak veri setimizde gerçek bir elması temsil eden 53910 satır kalmaktadır.

4.3.1. Veri kodlama

Özellikle kategorik verilerin sayısallaştırılması için kullanılan veri kodlama, makine öğrenmesi modellerinin kategorik verilerle daha etkili çalışmasına ve doğru sonuçlar üretmesine aracılık eder (Lee, 2019; Raschka ve Mirjalili, 2019). 8 Temel ön işleme adımından sonra 2 farklı veri kodlama tekniği ile sayısallaştırma yapılmaktadır.

4.3.1.1. Etiket kodlama (Label encoding)

Veri kodlama tekniği olan etiket kodlamada her kategoriye benzersiz bir tamsayı atanır (Lee, 2019; Raschka ve Mirjalili, 2019). Çalışmada kalite ilişkilerine bağlı olarak, “cut” sütunundaki [“Fair”, “Good”, “Very Good”, “Premium”, “Ideal”] değerlerine sırasıyla [1, 2, 3, 4, 5]; “color” sütunundaki [“J”, “I”, “H”, “G”, “F”, “E”, “D”] değerlerine sırasıyla [1, 2, 3, 4, 5, 6, 7]; “clarity” sütunundaki [“I1”, “SI2”, “SI1”, “VS2”, “VS1”, “VVS2”, “VVS1”, “IF”] değerlerine sırasıyla [1, 2, 3, 4, 5, 6, 7, 8] atanmaktadır. Çalışmada bu işlemten sonra elde edilen veri seti “LE” olarak adlandırılmaktadır.

4.3.1.2. İkili kodlama (Binary encoding)

Veri kodlama tekniği olan ikili kodlamada önce kategorilere etiket kodlamada olduğu gibi benzersiz bir tamsayı atanır ve ardından atanan numaralar ikili sayı sistemine dönüştürülür (Lee, 2019; Raschka ve Mirjalili, 2019). Çalışmada kalite ilişkilerine bağlı olarak, “cut” sütunundaki [“Fair”, “Good”, “Very Good”, “Premium”, “Ideal”] değerlerine sırasıyla [001, 010, 011, 100, 101]; “color” sütunundaki [“J”, “I”, “H”, “G”,

“F”, “E”, “D”] değerlerine sırasıyla [001, 010, 011, 100, 101, 110, 111]; “clarity” sütunundaki [“I1”, “SI2”, “SI1”, “VS2”, “VS1”, “VVS2”, “VVS1”, “IF”] değerlerine sırasıyla [0001, 0010, 0011, 0100, 0101, 0110, 0111, 1000] atanmaktadır. Daha sonra her değişken için bit düzeyi kadar sütun oluşturulur. Yani “cut” için 3, “color” için 3 ve “clarity” için 4 yeni sütun oluşturularak, her kategorinin sahip olduğu bitler bu sütunlara eklenir. Bu işlemlerden sonra elde edilen veri seti “BE” olarak adlandırılmaktadır.

4.3.2. Veri normalizasyonu

Verilerin ölçeğini düzenlemek ve farklı aralıklarda bulunan özellikleri ortak bir ölçeğe getirmek için kullanılan veri normalizasyonu, tüm değişkenlerin eşit ağırlıkta katkıda bulunmasını sağlamak ve algoritmaların daha hızlı ve doğru çalışmasına yardımcı olmaktadır (Lee, 2019; Raschka ve Mirjalili, 2019). Veri kodlama işleminden sonra elde edilen “LE” ve “BE” olarak adlandırılan 2 veri setinde 2 farklı teknik ile normalizasyon yapılmaktadır.

4.3.2.1. Min-Max normalizasyonu

Verileri belirli bir aralığa (genellikle 0 ile 1 arasında) dönüştürmek için kullanılan veri normalizasyon tekniği Min-Max, bu işlemi her bir veri noktası için belirtilen formüle göre yeniden ölçeklendirir:

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4.1)$$

Burada, X : orijinal değer, X_{min} : değişkenin minimum değeri ve X_{max} : değişkenin maksimum değeridir (Lee, 2019).

4.3.2.2. Z-Score normalizasyonu

Verilerin ortalama değeri etrafında standart sapma birimlerinde ölçeklendirme yapan Z-Score tekniği, verileri ortalaması 0 ve standart sapması 1 olan bir dağılıma dönüştürür. Her bir veri noktası için belirtilen formülü kullanır:

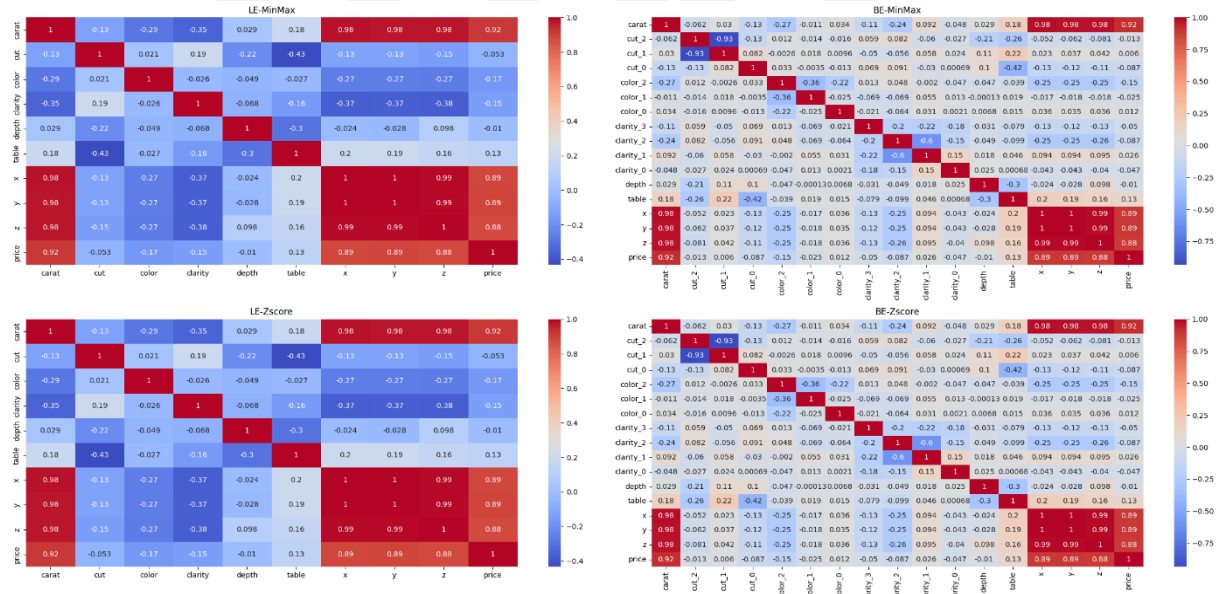
$$Z = \frac{X - \mu}{\sigma} \quad (4.2)$$

Burada, X : orijinal değer, μ : değişkenin ortalaması ve σ : değişkenin standart sapmasıdır (Lee, 2019).

Veri kodlama işleminden sonra elde edilen “LE” ve “BE” veri setlerinde Min-Max ve Z-Score ile normalizasyon yapılarak “LE-MinMax”, “BE-MinMax”, “LE-Zscore” ve “BE-Zscore” olarak adlandırılan dört adet veri seti elde edilmektedir.

4.4. Veri Setlerinin Keşfi

Veri ön işleme adımlarından sonra elde edilen “LE-MinMax”, “BE-MinMax”, “LE-Zscore” ve “BE-Zscore” veri setlerinin korelasyon matrisleri için oluşturulan ısı haritaları Şekil 4.2.’de gösterilmektedir. Bu haritalar sayesinde sütunlar arasındaki ilişkiler rahat bir şekilde gözlemlenebilmektedir.



Şekil 4.2. Korelasyon ısı haritaları

4.5. Model Seçimleri ve Kurulumları

Hedeflenen kapsamda, farklı makine öğrenmesi yöntemleri kullanılarak elmas fiyat tahmin modelleri oluşturulmaktadır. Bu çeşitlilik, modellerin değerlendirilmesinde çok yönlü bir bakış açısı sunmaktadır. Çalışmada oluşturulan dört farklı veri setinde,

“price” hedef değişkeni, “price” değişkeni hariç tüm değişkenler giriş değişkeni olarak seçilmektedir. Modellerin karşılaştırılmasının adil bir şekilde yapılabilmesi için her modelin varsayılan hiper parametreleri kullanılmaktadır. Ayrıca modellere, aşırı öğrenme problemine karşı önlem olarak K-Katlı Çapraz Doğrulama (K-Fold Cross-Validation) kullanılarak 10 katlı çapraz doğrulama uygulanmaktadır (Lee, 2019; Raschka ve Mirjalili, 2019).

4.5.1. Doğrusal (Linear - LinearRegression)

Bağımlı ve bağımsız değişkenler arasındaki doğrusal ilişkiyi modelleyen, basit ve yorumlanabilir bir yöntemdir. Makine öğrenmesi modellerinin temelini oluşturan bu yöntem, hedef değişkenin, giriş değişkenlerinin ağırlıklı toplamı olarak ifade edildiği bir doğrusal fonksiyon oluşturur (Lee, 2019; Basha ve ark., 2023). Daha karmaşık yöntemlerle karşılaştırma yapmak için başlangıç noktası olarak tercih edilmektedir.

4.5.2. K-en yakın komşu (K-nearest neighbors - KNeighborsRegressor)

Bu yöntem, tahmini yapılacak veri noktası için en yakın komşunun değerlerini kullanarak hedef değişkeni belirler. Modelin doğrudan eğitim verisi üzerinden öğrenmesini sağlar (Lee, 2019; Raschka ve Mirjalili, 2019). Farklı çalışma prensibi ve doğrusal olmayan ilişkileri yakalama potansiyeli nedeniyle kullanılmaktadır.

4.5.3. Destek vektör (Support vector - SVR)

Verileri daha yüksek bir boyuta taşıyarak doğrusal olmayan sınırlar oluşturan ve bunları tahmin için kullanan bir yöntemdir. Kernel fonksiyonları sayesinde karmaşık ilişkileri etkili bir şekilde modelleyebilir (Lee, 2019; Mihir ve ark., 2021). Bu kabiliyeti seçilmesinde önemli bir etkidir.

4.5.4. Karar ağacı (Decision tree - DecisionTreeRegressor)

Bu yöntem, veriyi özelliklerine göre ayırarak tahmin yapan bir ağaç yapısı kullanır. Her düğümde bir karar alınarak, sonuca ulaşılır ve model, verinin her bir parçası

için tahminler yapar (Raschka ve Mirjalili, 2019; Mihir ve ark., 2021). Karar verme sürecindeki şeffaflık ve ağaç yapısının gücü sebebiyle öne çıkmaktadır.

4.5.5. Rastgele orman (Random forest - RandomForestRegressor)

Birden fazla karar ağacının birleştirilmesiyle oluşan bir topluluk yöntemidir. Her bir ağaç bağımsız olarak tahminler yapar ve nihai sonuç, bu tahminlerin ortalaması alınarak genelleştirilir. Modelin daha güvenilir ve stabil sonuçlar üretmesine yardımcı olur. Aşırı öğrenmeye karşı dayanıklı olması ve yüksek genelleme yeteneği ile ön plana çıkmaktadır (Raschka ve Mirjalili, 2019; Basha ve ark., 2023). Birden fazla karar ağacının birleşimi, modelin hata payını azaltarak daha doğru tahminler yapılmasını sağlar, bu da özellikle karmaşık ve gürültülü veri setlerinde faydalıdır. Önemli özellikleri sayesinde çalışmada yer bulmaktadır.

4.5.6. Aşırı gradyan güçlendirme (eXtreme gradient boosting - XGBRegressor)

Gradyan artırma yöntemlerinin bir türevi ve güçlü bir topluluk yöntemidir. Gradyan azalma tekniğini kullanarak, modelin hatalarını her iterasyonda düzeltir ve daha doğru sonuçlar elde edilmesini sağlar. Büyük veri setlerinde yüksek doğruluk sağlama kapasitesi ve modelin hızlı öğrenme yeteneği nedeniyle önem arz etmektedir (Mihir ve ark., 2021; Basha ve ark., 2023). Performansı ve veri üzerinde çok daha hassas tahminler yapabilme yeteneği, onu güçlü bir seçenek haline getirmektedir.

4.5.7. Çok katmanlı algılayıcı (Multi-layer perceptron - MLPRegressor)

Yapay sinir ağı yöntemidir ve çok katmanlı ağlar kullanarak doğrusal olmayan ilişkileri modelleyebilir. Bu yöntem, daha karmaşık ve yüksek boyutlu verilerde başarılı olabilir, çünkü sinir ağları büyük veri kümelerinde örüntüleri tanıyıp öğrenebilmektedir (Taud ve Mas, 2018). Derin öğrenme algoritmalarının temelini ve gücünü temsil eden model olduğu için önemli bir seçimdir.

5. METRİKLER, ARAŞTIRMA SONUÇLARI VE İNCELEMELER

Model tahminlerinin doğruluklarını ölçmek için farklı yaklaşımlar benimseyen ve farklı formüllere sahip birçok metrik kullanılmaktadır. Bu metrikler hem normalize hem de denormalize veriler üzerinden hesaplanarak sunulmaktadır. Normalize veriler üzerinden yapılan hesaplamalar, model performansını ölçek bağımsız olarak değerlendirirken, denormalize veriler üzerinden yapılan hesaplamalar, hata değerlerini gerçek ölçekte ifade ederek sonuçların anlamlılığını artırmayı hedeflemektedir (Chicco ve ark., 2021).

Metriklerin formüllerinde a_i : gerçek değeri, p_i : tahmin değerini, n : örneklem sayısını belirtmektedir.

5.1. Ortalama Mutlak Hata (Mean Absolute Error)

Çalışmalarda yaygın olarak kullanılan bir metrik olan MAE (Ortalama Mutlak Hata (Mean Absolute Error)), modellerin tahminlerinin gerçek değerlerden ne kadar saptığını ölçen bir hata metriğidir. Tahmin hatalarının büyüklüğünü yorumlamak için kullanılır. MAE'nin düşük olması, modelin daha doğru tahminler yaptığını gösterir. Belirtilen formüle göre hesaplanır (Chicco ve ark., 2021):

$$MAE = \frac{1}{n} \sum_{i=1}^n |a_i - p_i| \quad (5.1)$$

Şekil 5.1.'de gösterilen MAE sonuçlarına göre, normalize veriler üzerinden hesaplanan MAE için en iyi kombinasyon LE-MinMax veri seti ve RandomForestRegressor modeli olarak ortaya çıkmaktadır. Denormalize veriler üzerinden hesaplanan MAE için ise en iyi kombinasyon LE-Zscore veriseti ve RandomForestRegressor modeli olarak gözlemlenebilmektedir.

Normalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	0.04327	0.01833	0.04012	0.0189	0.01426	0.01482	0.02023
BE-MinMax	0.04244	0.01934	0.0467	0.01919	0.01466	0.01524	0.02134
LE-Zscore	0.20073	0.09185	0.08177	0.08775	0.06615	0.06854	0.07778
BE-Zscore	0.1969	0.09833	0.08475	0.0894	0.06803	0.07067	0.07843

Denormalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	800.3851	339.1111	742.1982	349.5976	263.8595	274.2602	374.3517
BE-MinMax	785.0959	357.8176	863.89	354.9886	271.2507	281.9897	394.8846
LE-Zscore	800.3925	366.2459	326.0754	349.883	263.762	273.3031	310.1447
BE-Zscore	785.1031	392.0658	337.9484	356.4703	271.2529	281.8166	312.7274

Şekil 5.1. MAE sonuçları

5.2. Ağırlıklı Ortalama Mutlak Hata (Weighted Mean Absolute Error)

WMAE (Ağırlıklı Ortalama Mutlak Hata (Weighted Mean Absolute Error)), MAE'nin ağırlıklandırılan bir varyasyonudur ve hata hesaplamalarında farklı veri noktalarına farklı önem seviyeleri atamak için kullanılır. Bu metrik, her bir hata terimini, ilgili veri noktasına atanan bir ağırlık ile çarparak ortalama hata değerini hesaplar. WMAE, özellikle bazı veri noktalarının diğerlerinden daha kritik olduğu veya daha fazla öneme sahip olduğu durumlarda kullanışlıdır. Düşük bir WMAE değeri, ağırlıklı veri noktalarındaki tahminlerin daha doğru olduğunu gösterir. Belirtilen formüle göre hesaplanır (Cleger-Tamayo ve ark., 2012):

$$WMAE = \frac{\sum_{i=1}^n w_i \cdot |a_i - p_i|}{\sum_{i=1}^n w_i} \quad (5.2)$$

Burada, w_i :ağırlık katsayısını temsil eder ve çalışmada her değer toplam ağırlığa oranı biçiminde belirlenerek değeri daha yüksek varlıkların (elmasların) önem derecesinin artırılması hedeflenmektedir. Şu şekilde hesaplanır:

$$w_i = \frac{a_i}{\sum_{i=1}^n a_i} \quad (5.3)$$

Şekil 5.2.'de gösterilen WMAE sonuçlarına göre, normalize veriler üzerinden hesaplanan WMAE için en iyi kombinasyon LE-MinMax veri seti ve

RandomForestRegressor modeli olarak ortaya çıkmaktadır. Denormalize veriler üzerinden hesaplanan WMAE için de en iyi kombinasyon, normalize veriler üzerinden hesaplanan WMAE için ortaya çıkan en iyi kombinasyon ile aynıdır. Bu metrik, Z-Score ile normalize edilmiş veri setlerinde hesaplanırken negatif değere dönüşebilir çünkü WMAE hesaplanırken baz alınan ağırlıklar negatif olabilmektedir.

Normalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	0.07145	0.0384	0.04344	0.04051	0.03084	0.03123	0.0378
BE-MinMax	0.07015	0.04153	0.04738	0.04113	0.03167	0.03211	0.03791
LE-Zscore	-17.10124	-12.59734	-9.8389	-13.15487	-9.89999	-10.03927	-10.00568
BE-Zscore	-16.82565	-13.53317	-9.73897	-12.9231	-10.13978	-10.23955	-9.78544

Denormalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	1321.6337	710.2954	803.5303	749.3098	570.4426	577.6694	699.2572
BE-MinMax	1297.5443	768.1958	876.3037	760.7618	585.828	593.863	701.1627
LE-Zscore	-68186.7348	-50228.6073	-39230.0398	-52451.6004	-39473.6313	-40028.9683	-39895.0197
BE-Zscore	-67087.8698	-53959.9828	-38831.5841	-51527.4762	-40429.7077	-40827.5142	-39016.8775

Şekil 5.2. WMAE sonuçları

5.3. Kantil Ortalama Mutlak Hata (Quantile Mean Absolute Error)

Çalışmada oluşturulan QMAE (Kantil Ortalama Mutlak Hata (Quantile Mean Absolute Error)), hata hesaplamalarında pozitif ve negatif sapmalara farklı ağırlıklar atayarak, tahmin performansını daha detaylı bir şekilde değerlendiren bir metriktir. Bu metrik, negatif hataları bir katsayı ile çarparken, pozitif hatalara farklı bir katsayı uygular. Bu sayede, belirli türdeki hatalara daha fazla önem verilmesi sağlanır. QMAE, özellikle tahmin sapmalarının yönünün (pozitif veya negatif) farklı etkiler oluşturduğu senaryolarda kullanışlıdır. Düşük bir QMAE değeri, modelin bu farklı önem seviyelerini dikkate alarak daha dengeli tahminler yaptığını gösterir. Belirtilen formüle göre hesaplanır:

$$QMAE = \frac{1}{n} \sum_{i=1}^n q_i \cdot |a_i - p_i| \quad (5.4)$$

Burada, q_i :kantili temsil eder ve çalışmada satıcı perspektifi benimsenerek gerçek değerden düşük tahminlere daha yüksek bir katsayı atanmaktadır. Şu şekilde hesaplanır:

$$q_i = \begin{cases} 0.9, & a_i > p_i \\ 0.1, & a_i \leq p_i \end{cases} \quad (5.5)$$

Şekil 5.3.'te gösterilen QMAE sonuçlarına göre, normalize veriler üzerinden hesaplanan QMAE için en iyi kombinasyon LE-MinMax veri seti ve RandomForestRegressor modeli olarak ortaya çıkmaktadır. Denormalize veriler üzerinden hesaplanan QMAE için ise en iyi kombinasyon LE-Zscore veriseti ve RandomForestRegressor modeli olarak gözlemlenebilmektedir.

Normalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	0.02163	0.00923	0.01555	0.00936	0.00707	0.00742	0.00893
BE-MinMax	0.02122	0.01022	0.03492	0.0095	0.00723	0.00761	0.00949
LE-Zscore	0.10035	0.04629	0.03659	0.04336	0.03279	0.03444	0.0399
BE-Zscore	0.09844	0.05069	0.03731	0.04438	0.03358	0.03534	0.03769

Denormalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	400.1254	170.7224	287.6768	173.044	130.7452	137.2165	165.1779
BE-MinMax	392.4831	189.0837	645.8418	175.6834	133.8122	140.7789	175.6268
LE-Zscore	400.1291	184.565	145.8807	172.8894	130.7401	137.3242	159.0966
BE-Zscore	392.4867	202.1191	148.7631	176.9467	133.8951	140.9234	150.2728

Şekil 5.3. QMAE sonuçları

5.4. Asimetrik Ortalama Mutlak Hata (Asymmetric Mean Absolute Error)

Çalışmada oluşturulan AMAE (Asimetrik Ortalama Mutlak Hata (Asymmetric Mean Absolute Error)), tahmin hatalarını asimetrik olarak değerlendiren bir metriktir. Tahmin değeri ve gerçek değer arasındaki duruma göre cezalandırma uygulanır. Bu yaklaşım, belirli türdeki tahmin hatalarının daha az önemli olduğu durumlarda kullanışlıdır. Modellerin tahmin hatalarını asimetrik bir perspektifle değerlendirmesini sağlayarak, belirli hataların daha toleranslı bir şekilde ele alınmasını mümkün kılar. Düşük bir AMAE değeri, modelin bu asimetrik değerlendirme çerçevesinde daha doğru tahminler yaptığını gösterir. Belirtilen formüle göre hesaplanır:

$$AMAE = \frac{1}{n} \sum_{i=1}^n e_i \quad (5.6)$$

Burada, e_i :asimetrik ceza fonksiyonunu temsil eder ve çalışmada satıcı perspektifi benimsenerek gerçek değerden düşük tahminlerin ceza fonksiyonu daha baskın olacak şekilde tasarlanmaktadır. Şu şekilde hesaplanır:

$$e_i = \begin{cases} |a_i - p_i|, & a_i > p_i \\ \sqrt{|a_i - p_i|}, & a_i \leq p_i \end{cases} \quad (5.7)$$

Şekil 5.4.'te gösterilen AMAE sonuçlarına göre, normalize veriler üzerinden hesaplanan AMAE için en iyi kombinasyon LE-MinMax veri seti ve RandomForestRegressor modeli olarak ortaya çıkmaktadır. Denormalize veriler üzerinden hesaplanan AMAE için de en iyi kombinasyon, normalize veriler üzerinden hesaplanan AMAE için ortaya çıkan en iyi kombinasyon ile aynıdır.

Normalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	0.12146	0.06421	0.12998	0.06157	0.05397	0.05655	0.07898
BE-MinMax	0.12072	0.06516	0.0791	0.06213	0.0552	0.05778	0.07998
LE-Zscore	0.31537	0.17066	0.17428	0.15582	0.13378	0.13965	0.1533
BE-Zscore	0.31275	0.17774	0.18364	0.15785	0.13681	0.14339	0.16007

Denormalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	413.6857	178.4893	282.5374	179.7103	136.8299	143.9207	169.2455
BE-MinMax	405.9998	199.0798	704.9324	182.392	139.886	147.5486	179.7998
LE-Zscore	413.6895	192.7734	150.3539	179.4855	136.8365	144.133	167.2473
BE-Zscore	406.0036	211.6387	153.0304	183.7949	139.9872	147.7497	156.5018

Şekil 5.4. AMAE sonuçları

5.5. Ortalama Kare Hata (Mean Squared Error)

MSE (Ortalama Kare Hata (Mean Squared Error)), MAE gibi tahmin modellerinin doğruluğunu değerlendirmek için yaygın olarak kullanılan bir hata metriğidir. Büyük sapmaların daha fazla cezalandırılmasını sağlayarak, büyük hatalara karşı daha duyarlı hale gelir. Düşük bir MSE değeri, modelin tahminlerinin gerçek değerlere daha yakın olduğunu ve büyük sapmaların az olduğunu gösterir. Belirtilen formüle göre hesaplanır (Chicco ve ark., 2021):

$$MSE = \frac{1}{n} \sum_{i=1}^n (a_i - p_i)^2 \quad (5.8)$$

Şekil 5.5.'te gösterilen MSE sonuçlarına göre, normalize veriler üzerinden hesaplanan MSE için en iyi kombinasyon LE-MinMax veri seti ve XGBRegressor modeli olarak ortaya çıkmaktadır. Denormalize veriler üzerinden hesaplanan MSE için de en iyi kombinasyon, normalize veriler üzerinden hesaplanan MSE için ortaya çıkan en iyi kombinasyon ile aynıdır.

Normalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	0.00428	0.00127	0.00246	0.00154	0.00084	0.00082	0.00116
BE-MinMax	0.00417	0.00162	0.00319	0.00158	0.00087	0.00088	0.00121
LE-Zscore	0.09218	0.03067	0.02118	0.03319	0.01793	0.01773	0.02005
BE-Zscore	0.08973	0.03616	0.02216	0.03406	0.01874	0.01877	0.0194

Denormalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	1465406.7189	433131.192	842299.5329	525541.8185	286120.9939	281711.3247	397998.7865
BE-MinMax	1426509.8521	552632.3367	1090682.5284	540797.8231	298668.3251	300592.8644	414634.6421
LE-Zscore	1465433.9019	487669.514	336786.4321	527580.7754	285025.26	281837.1168	318702.2411
BE-Zscore	1426536.3136	574825.3201	352356.0009	541466.182	297980.6412	298384.1668	308449.972

Şekil 5.5. MSE sonuçları

5.6. Karekök Ortalama Kare Hata (Root Mean Squared Error)

RMSE (Karekök Ortalama Kare Hata (Root Mean Squared Error)), MSE'nin karekökü alınarak hesaplanan bir hata metriğidir ve tahmin modellerinin doğruluğunu değerlendirmek için yaygın olarak kullanılır. Bu metrik, tahmin hatalarının büyüklüğünü ölçerken, hataların karesini alarak büyük sapmalara daha fazla ağırlık verir. Karekök işlemi ise, hata birimini orijinal veriyle aynı hale getirir, bu da yorumlamayı kolaylaştırır. Düşük bir RMSE değeri, modelin tahminlerinin gerçek değerlere daha yakın olduğunu ve büyük sapmaların az olduğunu gösterir. Belirtilen formüle göre hesaplanır (Chicco ve ark., 2021):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (a_i - p_i)^2} \quad (5.9)$$

Şekil 5.6.'da gösterilen RMSE sonuçlarına göre, normalize veriler üzerinden hesaplanan RMSE için en iyi kombinasyon LE-MinMax veri seti ve XGBRegressor modeli olarak ortaya çıkmaktadır. Denormalize veriler üzerinden hesaplanan RMSE için de en iyi kombinasyon, normalize veriler üzerinden hesaplanan RMSE için ortaya çıkan en iyi kombinasyon ile aynıdır.

Normalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	0.06544	0.03556	0.04961	0.03916	0.02889	0.02868	0.0341
BE-MinMax	0.06457	0.04016	0.05645	0.03971	0.02951	0.02961	0.03479
LE-Zscore	0.30358	0.17504	0.14543	0.18203	0.13374	0.13306	0.1415
BE-Zscore	0.29953	0.19002	0.14876	0.18436	0.13674	0.1369	0.13921

Denormalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	1210.4542	657.8143	917.7138	724.3817	534.3423	530.4531	630.6806
BE-MinMax	1194.2862	742.7839	1044.245	734.493	545.8125	547.7863	643.5604
LE-Zscore	1210.4654	697.9122	579.8709	725.8131	533.2508	530.5301	564.2109
BE-Zscore	1194.2973	757.65	593.1448	735.097	545.2252	545.8409	555.0801

Şekil 5.6. RMSE sonuçları

5.7. R-Kare (R-Squared), Belirlilik Katsayısı (Coefficient of Determination)

R^2 (R-Kare (R-Squared), Belirlilik Katsayısı (Coefficient of Determination)), bir modelin bağımlı değişkendeki (hedef değişken) varyasyonu ne kadar iyi açıkladığını ölçen ve yaygın olarak kullanılan istatistiksel bir metriktir. 0 ile 1 arasında bir değer alır. 1'e yaklaştıkça modelin hedef değişken üzerindeki açıklayıcılığı artar. Tahmin edilen değerlerin gerçek değerlere olan yakınlığını göstererek, modelin genel performansını özetler. Belirtilen formüle göre hesaplanır (Chicco ve ark., 2021):

$$R^2 = 1 - \frac{\sum_{i=1}^n (a_i - p_i)^2}{\sum_{i=1}^n (a_i - \bar{a})^2} \quad (5.10)$$

Burada, $\bar{a}$: gerçek değerlerin ortalamasını temsil eder ve şu şekilde hesaplanır:

$$\bar{a} = \frac{1}{n} \sum_{i=1}^n a_i \quad (5.11)$$

Şekil 5.7.'de gösterilen R^2 sonuçlarına göre, normalize veriler üzerinden hesaplanan R^2 için en iyi kombinasyon LE-MinMax veri seti ve XGBRegressor modeli olarak ortaya çıkmaktadır. Denormalize veriler üzerinden hesaplanan R^2 için de en iyi kombinasyon, normalize veriler üzerinden hesaplanan R^2 için ortaya çıkan en iyi kombinasyon ile aynıdır.

Normalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	0.9077	0.97275	0.94696	0.96696	0.98201	0.98228	0.97493
BE-MinMax	0.91015	0.96524	0.9313	0.96601	0.98123	0.98109	0.9739
LE-Zscore	0.9077	0.96932	0.97881	0.96683	0.98209	0.98228	0.97994
BE-Zscore	0.91015	0.96385	0.97784	0.96597	0.98127	0.98124	0.98059

Denormalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	0.9077	0.9728	0.947	0.967	0.982	0.9823	0.9749
BE-MinMax	0.9101	0.9652	0.9313	0.966	0.9812	0.9811	0.9739
LE-Zscore	0.9077	0.9693	0.9788	0.9668	0.9821	0.9823	0.9799
BE-Zscore	0.9101	0.9639	0.9778	0.966	0.9813	0.9812	0.9806

Şekil 5.7. R^2 sonuçları

5.8. Ortalama Mutlak Yüzde Hata (Mean Absolute Percentage Error)

MAPE (Ortalama Mutlak Yüzde Hata (Mean Absolute Percentage Error)), tahmin hatalarını yüzde cinsinden ifade eden bir hata metriğidir ve gerçek değerlere göre tahmin hatalarının büyüklüğünü ölçer. Hata boyutunu anlaşılır bir şekilde ifade ettiği için farklı ölçeklerdeki veri kümeleriyle çalışırken oldukça kullanışlıdır. Düşük bir MAPE değeri, modelin tahminlerinin gerçek değerlere daha yakın olduğunu ve yüzdesel hata oranının düşük olduğunu gösterir. Ancak, MAPE sıfıra yakın gerçek değerler için hesaplama zorlukları yaratabilir. Belirtilen formüle göre hesaplanır (Chicco ve ark., 2021):

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{a_i - p_i}{a_i} \right| \quad (5.12)$$

MAPE'nin anlamlı sonuç verebilmesi için denormalize veriler üzerinden hesaplanması gerekir. Şekil 5.8.'de gösterilen MAPE sonuçlarına göre, en iyi

kombinasyon LE-Zscore veri seti ve RandomForestRegressor modeli olarak ortaya çıkmaktadır.

Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	43.4121	9.2708	55.9642	8.2845	6.2909	7.1572	13.8725
BE-MinMax	42.3238	9.8661	64.9941	8.4089	6.4734	7.4001	16.0121
LE-Zscore	43.4141	10.1122	11.832	8.2705	6.2903	7.0929	9.725
BE-Zscore	42.3257	10.9314	13.1417	8.4682	6.4715	7.3992	10.5181

Şekil 5.8. MAPE sonuçları

5.9. Kantil Ortalama Mutlak Yüzde Hata (Quantile Mean Absolute Percentage Error)

Çalışmada oluşturulan QMAPE (Kantil Ortalama Mutlak Yüzde Hata (Quantile Mean Absolute Percentage Error)), QMAE’de olduğu gibi farklı sapmalara farklı katsayılar atayan ve tahmin hatalarını yüzde cinsinden ifade eden bir metriktir. Bu sayede, belirli yönlerdeki hataların performans üzerindeki etkisini ayrı ayrı değerlendirme olanağı sunar. Düşük bir QMAPE değeri, modelin hedef değişkenin varyasyonunu daha iyi açıkladığını ve tahmin performansının yüksek olduğunu gösterir. Belirtilen formüle göre hesaplanır:

$$QMAPE = \frac{100}{n} \sum_{i=1}^n q_i \cdot \left| \frac{a_i - p_i}{a_i} \right| \quad (5.13)$$

Burada, q_i :kantili temsil eder ve QMAE’de olduğu gibi satıcı perspektifi benimsenerek gerçek değerden düşük tahminlere daha yüksek bir katsayı atanmaktadır. Şu şekilde hesaplanır:

$$q_i = \begin{cases} 0.9, & a_i > p_i \\ 0.1, & a_i \leq p_i \end{cases} \quad (5.14)$$

MAPE’de olduğu gibi anlamlı sonuç verebilmesi için denormalize veriler üzerinden hesaplanması gerekir. Şekil 5.9.’da gösterilen QMAPE sonuçlarına göre, en iyi kombinasyon LE-Zscore veri seti ve RandomForestRegressor modeli olarak ortaya çıkmaktadır.

Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	24.8017	3.9573	12.6295	3.7839	2.7636	3.2119	5.114
BE-MinMax	24.2163	4.1633	55.8362	3.8285	2.8114	3.3168	6.8156
LE-Zscore	24.8031	4.2385	3.9686	3.7799	2.7604	3.1929	4.3428
BE-Zscore	24.2177	4.5305	3.8048	3.8501	2.8123	3.3162	4.5632

Şekil 5.9. QMAPE sonuçları

5.10. Asimetrik Ortalama Mutlak Yüzde Hata (Asymmetric Mean Absolute Percentage Error)

Çalışmada oluşturulan AMAPE (Asimetrik Ortalama Mutlak Yüzde Hata (Asymmetric Mean Absolute Percentage Error)), AMAE gibi asimetrik bir hata ölçüm yöntemi olup, pozitif ve negatif tahmin hatalarını farklı şekillerde ele alır. Bu yöntem, belirli yönlerdeki hataların diğerlerinden daha az kritik olduğu durumlarda kullanışlıdır. Asimetrik bir hassasiyetle performansın değerlendirilmesine yardımcı olan AMAPE'nin düşük bir değere sahip olması, modelin hedef değişkeni doğru bir şekilde tahmin ettiğini gösterir. Belirtilen formüle göre hesaplanır:

$$AMAPE = \frac{100}{n} \sum_{i=1}^n e_i \quad (5.15)$$

Burada, e_i :asimetrik ceza fonksiyonunu temsil eder ve AMAE'de olduğu gibi satıcı perspektifi benimsenerek gerçek değerden düşük tahminlerin ceza fonksiyonu daha baskın olacak şekilde tasarlanmaktadır. Şu şekilde hesaplanır:

$$e_i = \begin{cases} \left| \frac{a_i - p_i}{a_i} \right|, & a_i > p_i \\ \sqrt{\left| \frac{a_i - p_i}{a_i} \right|}, & a_i \leq p_i \end{cases} \quad (5.16)$$

MAPE'de olduğu gibi anlamlı sonuç verebilmesi için denormalize veriler üzerinden hesaplanması gerekir. Şekil 5.10.'da gösterilen AMAPE sonuçlarına göre, en iyi kombinasyon LE-MinMax veri seti ve RandomForestRegressor modeli olarak ortaya çıkmaktadır.

Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	53.4428	18.6798	53.4447	16.2564	14.5292	16.0591	25.1479
BE-MinMax	52.9212	19.0479	69.6422	16.3836	14.8404	16.4211	26.6615
LE-Zscore	53.4446	19.8421	23.0446	16.2457	14.5341	15.9498	18.9935
BE-Zscore	52.9229	20.6632	24.9201	16.4469	14.8411	16.4374	20.2513

Şekil 5.10. AMAPE sonuçları

5.11. Ortalama Orantılı Mutlak Hata (Mean Proportion Absolute Error)

Çalışmada oluşturulan en önemli metrik MP AE (Ortalama Orantılı Mutlak Hata (Mean Proportion Absolute Error)), her tahmin hatasını WMAE'ye benzer şekilde orantılayarak hesaplar ve aynı zamanda düşük değere sahip veri noktalarının önem derecelerinin aşırı zayıflamasını engeller. Her veri noktasının gerçek değerine göre göreceli önemini belirten bir orantı kullanır ve WMAE'deki negatif değerlerin oluşturduğu problemi de çözer. Ayrıca, toplam hatayı WMAE'deki gibi ağırlıklar toplamına değil örneklem sayısına böler. Özellikle gerçek değerlerin geniş bir aralıkta değiştiği durumlarda hataları göreceli büyüklüklerine göre değerlendirmek için kullanışlıdır. Düşük bir MP AE değeri, modelin hedef değişken üzerindeki tahmin doğruluğunun yüksek olduğunu gösterir. Belirtilen formüle göre hesaplanır:

$$MPAE = \frac{1}{n} \sum_{i=1}^n |w_i| \cdot |a_i - p_i| \quad (5.17)$$

Burada, w_i :orantı katsayısını temsil eder ve çalışmada her değer en yüksek değere orantısı biçiminde belirlenerek değeri daha yüksek varlıkların (elmasların) önem derecesinin artırılması hedeflenmektedir. Şu şekilde hesaplanır:

$$w_i = \frac{a_i}{a_{max}} \quad (5.18)$$

Şekil 5.11.'de gösterilen MP AE sonuçlarına göre, normalize veriler üzerinden hesaplanan MP AE için en iyi kombinasyon LE-MinMax veri seti ve RandomForestRegressor modeli olarak ortaya çıkmaktadır. Denormalize veriler

üzerinden hesaplanan MPAE için ise en iyi kombinasyon LE-Zscore veriseti ve RandomForestRegressor modeli olarak gözlemlenebilmektedir.

Normalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	0.01394	0.0075	0.00848	0.00791	0.00602	0.0061	0.00738
BE-MinMax	0.01369	0.00811	0.00925	0.00803	0.00618	0.00627	0.0074
LE-Zscore	0.06237	0.03009	0.02558	0.02889	0.02217	0.02264	0.02548
BE-Zscore	0.06117	0.03224	0.02665	0.02953	0.02278	0.02343	0.02492

Denormalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	267.3313	142.1576	167.0055	149.8299	114.026	115.5762	140.582
BE-MinMax	262.4451	153.6024	183.0572	152.1416	117.1143	118.8215	141.3529
LE-Zscore	267.3327	152.1426	124.203	149.7121	113.9163	115.3081	125.6599
BE-Zscore	262.4465	162.4678	127.3825	152.6495	117.0878	118.9337	122.2428

Şekil 5.11. MPAE sonuçları

5.12. Kantil Ortalama Orantılı Mutlak Hata (Quantile Mean Proportion Absolute Error)

Çalışmada oluşturulan QMPAE (Kantil Ortalama Orantılı Mutlak Hata (Quantile Mean Proportion Absolute Error)), farklı sapmaları farklı şekilde değerlendirir ve belirli yönlerdeki hataların performans üzerindeki etkisini ayrı ayrı değerlendirme olanağı sunar. Veri noktalarını farklı orantılandırma yöntemleriyle, gerçek değerın büyüklüğüne göre hataların önem derecesini ayırt eder. Düşük bir QMPAE değeri, modelin hedef değişken üzerindeki tahminlerinin daha doğru ve tutarlı olduğunu gösterir. Belirtilen formüle göre hesaplanır:

$$QMPAE = \frac{1}{n} \sum_{i=1}^n |q_i| \cdot |a_i - p_i| \quad (5.19)$$

Burada, q_i :kantili temsil eder ve çalışmada baskın satıcı perspektifi benimsenerek sadece gerçek değerden düşük tahminler değerlendirmeye dahil edilmekte ve katsayı atanmaktadır. Şu şekilde hesaplanır:

$$q_i = \begin{cases} \frac{a_i}{a_{max}}, & a_i > p_i \\ 0, & a_i \leq p_i \end{cases} \quad (5.20)$$

Şekil 5.12’de gösterilen QMPAE sonuçlarına göre, normalize veriler üzerinden hesaplanan QMPAE için en iyi kombinasyon LE-MinMax veri seti ve RandomForestRegressor modeli olarak ortaya çıkmaktadır. Denormalize veriler üzerinden hesaplanan QMPAE için ise en iyi kombinasyon LE-Zscore veriseti ve RandomForestRegressor modeli olarak gözlemlenebilmektedir.

Normalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	0.02069	0.00923	0.01301	0.00976	0.00687	0.00706	0.01005
BE-MinMax	0.02051	0.01074	0.00791	0.00995	0.00713	0.00736	0.00908
LE-Zscore	0.10122	0.03994	0.03137	0.03654	0.02646	0.02736	0.03082
BE-Zscore	0.10035	0.04457	0.03562	0.03789	0.02755	0.02867	0.03102

Denormalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	391.7722	173.81	247.8369	184.1765	129.3257	133.236	189.3025
BE-MinMax	388.4039	202.0143	159.8363	187.6946	134.1839	138.8923	171.6157
LE-Zscore	391.7748	191.2359	153.2665	182.4034	129.3052	133.7398	147.192
BE-Zscore	388.4065	213.193	173.1126	187.9585	134.2505	139.5243	146.0023

Şekil 5.12. QMPAE sonuçları

5.13. Asimetrik Ortalama Orantılı Mutlak Hata (Asymmetric Mean Proportion Absolute Error)

Çalışmada oluşturulan AMPAE (Asimetrik Ortalama Orantılı Mutlak Hata (Asymmetric Mean Proportion Absolute Error)), tahmin hatalarını orantılayarak hesaplayan ve asimetrik olarak değerlendiren bir metriktir. Tahmin değeri ve gerçek değer arasındaki duruma göre cezalandırma uygulanır. Modellerin tahmin hatalarını hem orantılayarak önem derecelerini belirginleştirir hem de asimetrik bir perspektifle değerlendirerek belirli hataların daha toleranslı bir şekilde ele alınmasını mümkün kılar. Düşük bir AMPAE değeri, modelin bu asimetrik değerlendirme çerçevesinde daha doğru tahminler yaptığını gösterir. Belirtilen formüle göre hesaplanır:

$$AMPAE = \frac{1}{n} \sum_{i=1}^n |w_i| \cdot e_i \quad (5.21)$$

Burada, w_i :orantı katsayısını, e_i :asimetrik ceza fonksiyonunu temsil eder ve satıcı perspektifi benimsenerek gerçek değerden düşük tahminlerin ceza fonksiyonu daha baskın olacak şekilde tasarlanmaktadır. Şu şekilde hesaplanırlar:

$$w_i = \frac{a_i}{a_{max}} \quad (5.22)$$

$$e_i = \begin{cases} |a_i - p_i|, & a_i > p_i \\ \sqrt{|a_i - p_i|}, & a_i \leq p_i \end{cases} \quad (5.23)$$

Şekil 5.13'te gösterilen AMPAE sonuçlarına göre, normalize veriler üzerinden hesaplanan AMPAE için en iyi kombinasyon LE-MinMax veri seti ve RandomForestRegressor modeli olarak ortaya çıkmaktadır. Denormalize veriler üzerinden hesaplanan AMPAE için ise en iyi kombinasyon LE-Zscore veriseti ve RandomForestRegressor modeli olarak gözlemlenebilmektedir.

Normalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	0.02813	0.01802	0.01944	0.01954	0.01601	0.01629	0.01935
BE-MinMax	0.02772	0.01841	0.0199	0.01972	0.01633	0.01661	0.01988
LE-Zscore	0.07842	0.04456	0.04382	0.04231	0.03511	0.03655	0.04007
BE-Zscore	0.07762	0.04665	0.04621	0.04299	0.03583	0.03765	0.04021

Denormalize							
Set\Model	LinearR	KNeighborsR	SVR	DecisionTreeR	RandomForestR	XGBR	MLPR
LE-MinMax	175.2891	86.6497	102.7464	83.3834	66.2889	66.7575	77.6368
BE-MinMax	171.7897	100.667	122.9701	85.2771	68.3296	69.0218	75.0851
LE-Zscore	175.2903	94.0406	67.1753	83.1386	66.2837	67.0563	75.6331
BE-Zscore	171.7908	104.7723	71.0133	86.2291	68.3245	69.1073	69.7903

Şekil 5.13. AMPAE sonuçları

6. SONUÇLAR

Elmas ve fiyatlandırması ile ilgili bir rehber niteliği taşıyan bu çalışmada oluşturulan 4 farklı veri seti, hem farklı kodlama ve normalizasyon tekniklerinin hem de modellerin karşılaştırılması için kapsamlı bir zemin yaratmaktadır. Bu veri setleriyle, farklı makine öğrenmesi yöntemleri kullanılarak elmas fiyat tahmin modelleri kurulmakta ve elde edilen sonuçlar üzerinden yöntemlerin performansları, yaygın olarak kullanılan ve çalışmada geliştirilen yenilikçi metrikler ile karşılaştırmalı şekilde incelenerek çok boyutlu bir perspektiften değerlendirme imkânı sunulmaktadır.

Elde edilen inceleme sonuçları, farklı kodlama, normalizasyon ve makine öğrenmesi yöntemlerinin fiyat tahmininde bazı güçlü ve zayıf yönleri sahip olduğunu ortaya koymaktadır. Normalize veriler üzerinden yapılan hesaplamalarda MAE, WMAE, QMAE, AMAE, MPA, QMPAE ve AMPAE için en iyi kombinasyon LE-MinMax veri seti ve RandomForestRegressor modeli olurken; MSE, RMSE ve R^2 için en iyi kombinasyon LE-MinMax veri seti ve XGBRegressor modeli olarak görülebilmektedir. Denormalize veriler üzerinden yapılan hesaplamalarda WMAE, AMAE, MSE, RMSE ve R^2 için en iyi kombinasyon değişmeyerek normalize veriler üzerinden yapılan hesaplamalarda ortaya çıkan en iyi kombinasyonlarla aynıdır. Ancak, MAE, QMAE, MAPE, QMAPE, MPAAE, QMPAAE, AMPAAE için en iyi kombinasyon LE-Zscore veri seti ve RandomForestRegressor modeli olmaktadır. Tüm değerlendirmeler göz önüne alındığında etiket kodlama, Min-Max normalizasyon ve rastgele orman yöntemlerinin ön plana çıktığına dair bir genelleneme yapılabilmektedir.

Gelecekte, bu çalışmanın başta yenilikçi metrikler olmak üzere birçok yönü, farklı sektörlerdeki tahmin çalışmaları veya model değerlendirmeleri için kullanılabilir. Çalışma, farklı yönleri ve yenilikçi yaklaşımları ile elde edilebilecek bulguların zenginleştirilmesi ve farklı perspektiflerin aynı anda değerlendirilebilmesi için potansiyel alanlar sunmaktadır.

KAYNAKLAR

- Agrawal, S., 2016, *Diamonds* [online], Kaggle, Web adresi: <https://www.kaggle.com/datasets/shivam2503/diamonds> [Ziyaret Tarihi: 01 Ocak 2023].
- Aharon, D. Y. ve Qadan, M., 2018, What drives the demand for information in the commodity market?, *Resources Policy*, 59, 532-543.
- Alden, N., 2009, *Simply Gemstones: Designs for Creating Beaded Gemstone Jewelry*, Crown Publishing Group.
- Bakırcı, Ç. M., 2023, *The Grand Diamond Lie: How Do They Trick the Lovers?* [online], Youtube, Web adresi: <https://www.youtube.com/watch?v=Ln3AfeqktHI> [Ziyaret Tarihi: 20 Kasım 2023].
- Basha, M. S. A., Oveis, P. M., Prabavathi, C., Lakshmi, M. B. ve Sucharitha, M. M., 2023, An efficient machine learning approach: Analysis of supervised machine learning methods to forecast the diamond price, *International Conference for Advancement in Technology (ICONAT)*, 1-6.
- Basha, M. S. A., Oveis, P. M., Raju, R. P. K. ve Sucharitha, M. M., 2024, Gems of Prediction: From Clarity to Carats-Unveiling Diamond Prices with Machine Learning in Waikato Environment for Knowledge Analysis, *2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT)*, 1-7.
- Basysyar, F. M. ve Dwilestari, G., 2022, Comparison Of Machine Learning Algorithms For Predicting Diamond Prices Based On Exploratory Data Analysis, *International Journal of Engineering Applied Sciences and Technology*, 7 (5), 71-79.
- Bauer, M., 2012, *Precious stones*, 1, Courier Corporation.
- Bracking, S. ve Sharife, K., 2014, Rough and polished: A case study of the diamond pricing and valuation system, *Leverhulme Centre for the Study of Value working paper series* 4.
- Cartigny, P., Palot, M., Thomassot, E. ve Harris, J. W., 2014, Diamond formation: a stable isotope perspective, *Annual Review of Earth and Planetary Sciences*, 42 (1), 699-732.
- Chaijunla, T. ve Taninpong, P., 2021, Comparative Study of Predicting Diamond Ring Prices in Online Retail Shop, *Data Science and Engineering (DSE) Record*, 2 (1), 1-13.
- Chen, W., Xu, H., Jia, L. ve Gao, Y., 2021, Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants, *International Journal of Forecasting*, 37 (1), 28-43.
- Chen, Y. ve Zhang, L., 2013, *Polishing of diamond materials: mechanisms, modeling and implementation*, Springer.
- Chicco, D., Warrens, M. J. ve Jurman, G., 2021, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, *Peerj computer science*, 7, e623.
- Chu, S., 2001, Pricing the C's of diamond stones, *Journal of Statistics Education*, 9 (2).
- Cleger-Tamayo, S., Fernández-Luna, J. M. ve Huete, J. F., 2012, On the Use of Weighted Mean Absolute Error in Recommender Systems, *RUE@ RecSys*, 24-26.
- Duckenfield, M., 2016, *The monetary history of gold: a documentary history, 1660-1999*, Routledge.

- Epstein, E. J., 1982, Have you ever tried to sell a diamond?, *Atlantic Monthly*, 363, 23-34.
- Fajar, F. I., Maylawati, D. S. a., Zulfikar, W. B., Wahana, A., Lukman, N. ve Ramdhani, M. A., 2024, Random Forest, K-Nearest Neighbors, and Adaboost Regressor in Diamond Price Forecasting, *2024 12th International Conference on Cyber and IT Service Management (CITSM)*, 1-7.
- Fitriani, S. A., Astuti, Y. ve Wulandari, I. R., 2022, Least absolute shrinkage and selection operator (lasso) and k-nearest neighbors (k-nn) algorithm analysis based on feature selection for diamond price prediction, *International Seminar on Machine Learning, Optimization, and Data Science (ISMODE)*, 135-139.
- GIA, 2002, *Diamond Quality Factors* [online], Gemological Institute of America, Web adresi: <https://www.gia.edu/diamond-quality-factor> [Ziyaret Tarihi: 17 Aralık 2023].
- Goel, S., Saxena, M., Sarangi, P. K. ve Rani, L., 2022, Gold and silver price prediction using hybrid machine learning models, *Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC)*, 390-395.
- Hershey, J. W., 1940, The book of diamonds: their curious lore, properties, tests and synthetic manufacture, *Hearthside Press*.
- HP, 2022, *HP Victus Laptop 16-D1048NT* [online], HP Development Company, Web adresi: <https://www.hpstore.com.tr/hp-victus-laptop-16-d1048nt-intel-core-i5-12500h-8gb-ram-512gb-ssd-nvidia-geforce-rtx3050-161-inc-fhd-144-hz-freedos-siyah> [Ziyaret Tarihi: 16 Ekim 2024].
- Hurlbut, C. S. ve Kammerling, R. C., 1991, *Gemology*, Wiley.
- Ince, H. ve Trafalis, T. B., 2006, A hybrid model for exchange rate prediction, *Decision Support Systems*, 42 (2), 1054-1062.
- King, H., 2012, *How Do Diamonds Form?* [online], Geology, Web adresi: <https://geology.com/articles/diamonds-from-coal> [Ziyaret Tarihi: 10 Eylül 2023].
- Lavoisier, A. L., 1772, Premier mémoire sur la destruction du diamant par la feu, *Mémoires de l'Académie royale des sciences*, 564-616.
- Lee, W. M., 2019, Python machine learning, *John Wiley & Sons*.
- Leung, C. K.-S., MacKinnon, R. K. ve Wang, Y., 2014, A machine learning approach for stock price prediction, *Proceedings of the international database engineering & applications symposium*, 18, 274-277.
- Mamonov, S. ve Triantoro, T., 2018, Subjectivity of diamond prices in online retail: insights from a data mining study, *Journal of theoretical and applied electronic commerce research*, 13 (2), 15-28.
- Manjula, K. ve Karthikeyan, P., 2019, Gold price prediction using ensemble based machine learning techniques, *International Conference on Trends in Electronics and Informatics (ICOEI)*, 3, 1360-1364.
- Matlins, A. L. ve Bonanno, A. C., 2009, *Jewelry & Gems-The Buying Guide*, 7, GemStone Press.
- Menase, M., 2009, Altın piyasası ve Türkiye'de altın fiyatlarını etkileyen faktörlerin analizi, Yüksek Lisans, *Marmara Üniversitesi*.
- Mihir, H., Patel, M. I., Jani, S. ve Gajjar, R., 2021, Diamond price prediction using machine learning, *International Conference on Communication, Computing and Industry 4.0 (C2I4)*, 2, 1-5.
- Pandey, A. C., Misra, S. ve Saxena, M., 2019, Gold and diamond price prediction using enhanced ensemble learning, *International Conference on Contemporary Computing (IC3)*, 12, 1-4.

- Patel, J., Shah, S., Thakkar, P. ve Kotecha, K., 2015, Predicting stock market index using fusion of machine learning techniques, *Expert systems with applications*, 42 (4), 2162-2172.
- Planck, M., 2011, *A Planet Made of Diamond* [online], Institute for Radio Astronomy, Web adresi: <https://www.astronomy.com/science/a-planet-made-of-diamond/> [Ziyaret Tarihi: 24 Eylül 2023].
- Poole Jr, C. P., 2004, Encyclopedic dictionary of condensed matter physics, *Academic Press*.
- Preston, E., 1960, Platinum in the glass industry, *Platinum Metals Review*, 4 (1), 2-9.
- Prinsloo, G., Spektorov, Y., Linde, O., Epstein, A. ve Fischler, S., 2011, The global diamond industry: Lifting the veil of mystery, *Bain & Company*.
- Rapaport, M., 2024, *The Rapaport Diamond Price List* [online], Rapaport, Web adresi: <https://rapaport.com/diamond-price-list/> [Ziyaret Tarihi: 01 Aralık 2024].
- Raschka, S. ve Mirjalili, V., 2019, Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2, *Packt publishing ltd*.
- Robertson, J., 1994, Diamond-like carbon, *Pure and applied chemistry*, 66 (9), 1789-1796.
- Sadorsky, P., 2021, Predicting gold and silver price direction using tree-based classifiers, *Journal of risk and financial management*, 14 (5), 198.
- Sammot-Bonnici, T. ve Channon, D. F., 2015, Pricing strategy, *Wiley encyclopedia of management*, 1-3.
- Sharma, G., Tripathi, V., Mahajan, M. ve Srivastava, A. K., 2021, Comparative analysis of supervised models for diamond price prediction, *International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 11, 1019-1022.
- Shirey, S. B. ve Shigley, J. E., 2013, Recent Advances In Understanding The Geology Of Diamonds, *Gems & Gemology*, 49 (4), 188-222.
- Smith, E. M., Ni, P., Shirey, S. B., Richardson, S. H., Wang, W. ve Shahar, A., 2021, Heavy iron in large gem diamonds traces deep subduction of serpentinized ocean floor, *Science Advances*, 7 (14), eabe9773.
- Soos, A., 2011, *Gold Mining Boom Increasing Mercury Pollution Risk* [online], OilPrice, Web adresi: <https://oilprice.com/Metals/Gold/Gold-Mining-Boom-Increasing-Mercury-Pollution-Risk.html> [Ziyaret Tarihi: 09 Kasım 2023].
- Swain, D., Chaturvedi, A. ve Rathi, V., 2023, GemInsight: Unleashing Random Forest for Diamond Quality Forecasting, *2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBE)*, 1-4.
- Taud, H. ve Mas, J.-F., 2018, Multilayer perceptron (MLP), *Geomatic approaches for modeling land change scenarios*, 451-455.
- Tennant, S., 1797, Iv. on the nature of the diamond, *Philosophical transactions of the Royal Society of London*, 87, 123-127.
- Tolkowsky, M., 1919, Diamond Design: A Study of the Reflection and Refraction of Light in a Diamond, *E. F. N. Spon, Ltd*.
- Umemura, T., Sato, K., Kusaka, Y. ve Satoh, H., 2022, Palladium, *Handbook on the Toxicology of Metals*, Elsevier, 649-662.
- Ünsal, E., Çınar, A. C. ve Taşdemir, Ş., 2022, A Performance Comparison of Python, C++ and Java Programming Languages, *New Trends In Engineering, Tablet Kitabevi*, 59-71.
- Walker, J., 1979, Optical absorption and luminescence in diamond, *Reports on progress in physics*, 42 (10), 1605.

- Wang, Z., Ma, H., Fang, S., Yang, Z., Miao, X., Chen, L. ve Jia, X., 2020, Synthesis and characterization of gem diamond single crystals in Fe-C system under high temperature and high pressure, *Journal of Crystal Growth*, 531, 125371.
- Werner, M. ve Locher, R., 1998, Growth and application of undoped and doped diamond films, *Reports on progress in physics*, 61 (12), 1665.
- Wolff, F.-C., 2016, Bargaining powers of buyers and sellers on the online diamond market: a double perspective non-parametric analysis, *Annals of Operations Research*, 244 (2), 697-718.
- Wu, C. ve Cosguner, K., 2020, Profiting from the decoy effect: A case study of an online diamond retailer, *Marketing Science*, 39 (5), 974-995.
- Yarnell, A., 2004, The many facets of man-made diamonds, *Chemical & engineering news*, 82 (5), 26-31.
- Zabala-Blanco, D., Ramírez, J., Ahumada-García, R., Flores-Calero, M., Palacios-Játiva, P. ve Saire-Yujra, R. V., 2024, Diamond Prices Estimation with ELM-Based Regressors, *2024 43rd International Conference of the Chilean Computer Science Society (SCCC)*, 1-8.

