



DOĞAL GAZ BORU HATTI İNŞAATI MALİYETLERİNİN ÇOKLU DOĞRUSAL REGRESYON VE K-EN YAKIN KOMŞULUK YÖNTEMLERİ İLE TAHMİNİ

Coşkun ÇAKMAK^{1*}, Mürsel ERDAL²

¹Baskent Elektrik Dağıtım A.Ş., Baskent Headquarters, 06530, Ankara, Türkiye

²Gazi University, Faculty of Technology, Department of Civil Engineering, 06560, Ankara, Türkiye

Özet: Bu çalışmada, Türkiye sınırları içerisinde yapılacak olan doğal gaz boru hattı (DGBH) maliyetlerinin ön tahmini için makine öğrenmesi algoritmaları kullanılarak modeller geliştirilmiştir. Bunun için, 1997-2022 yılları arasında Türkiye’de tamamlanmış DGBH projelerinden elde edilen veriler kullanılmıştır. Projelerin boru çapı, hat uzunluğu, hat vanası sayısı, take-off vana sayısı ve pig istasyonu sayısı gibi değişkenleri, maliyet tahmininde bağımsız değişkenler olarak belirlenmiştir. Veri setinin nicel anlamda yetersiz ve veri kalitesinin ortalama bir seviyede olmasından dolayı, klasik makine öğrenmesi tahmin süreçleri yürütülemez. Bu nedenle, mevcut veri seti eğitim ve test bölümlerine ayrılmadan, bütün veri kullanılarak çalışılmış ve Çoklu Doğrusal Regresyon (ÇDR) ile K-En Yakın Komşu (KNN) algoritmalarına konumlandırıldığında modelin uygun bir şekilde performans gösterip göstermediği incelenmiştir. Bu çalışma, ileride veri kalitesinin ve sayısının artması durumunda, klasik makine öğrenmesi tahmin süreçlerinin yürütülemezliği konusunda ön fikir vermesi amacıyla gerçekleştirilmiştir. Her iki farklı yöntem denemesinde de benzer ve ortalama düzeyde belirleme katsayıları (R^2) elde edilmiştir. Sonuç olarak, bu çalışmada, DGBH projelerinde ön maliyet tahminlerinin hassasiyetini iyileştirmek için ÇDR ve KNN yöntemlerinin etkinliği karşılaştırılmış ve sektöre önemli bir katkı sağlayacağı değerlendirilmiştir. Gelecekte yapılacak çalışmaların daha geniş veri setleri ve farklı model teknikleri kullanarak maliyet tahminlerinin doğruluğunu artırabileceği ve sektör paydaşlarına yol gösterici olabileceği öngörülmektedir.

Anahtar kelimeler: Doğal gaz boru hatları, Makine öğrenmesi, Maliyet tahmini


Estimation of Natural Gas Pipeline Construction Costs Using Multiple Linear Regression and K-Nearest Neighbor Methods


Abstract: In this study, models were developed using machine learning algorithms for the preliminary estimation of natural gas pipeline (NGP) costs within the borders of Türkiye. For this purpose, data obtained from NGP projects completed in Türkiye between 1997 and 2022 were used. Variables such as pipe diameter, line length, number of line valves, number of take-off valves and number of pigging stations of the projects were determined as independent variables in the cost estimation. Since the data set was quantitatively insufficient and the data quality was at an average level, classical machine learning estimation processes could not be carried out. For this reason, the existing data set was studied using the entire data without dividing it into training and test sections, and it was examined whether the model performed appropriately when positioned in Multiple Linear Regression (MLR) and K-Nearest Neighbor (KNN) algorithms. This study was carried out to provide a preliminary idea about whether classical machine learning estimation processes can be carried out in the future if the data quality and number increase. Similar and average coefficients of determination (R^2) were obtained in both different method trials. As a result, in this study, the effectiveness of the MLR and KNN methods was compared to improve the accuracy of preliminary cost estimates in NGP projects and it was evaluated that it will make a significant contribution to the sector. It is anticipated that future studies can increase the accuracy of cost estimates by using larger data sets and different model techniques and can guide the sector stakeholders.

Keywords: Natural gas pipelines, Machine-learning, Cost estimate

*Sorumlu yazar (Corresponding author): Baskent Elektrik Dağıtım A.Ş., Baskent Headquarters, 06530, Ankara, Türkiye

E mail: coskuncakmak90@gmail.com (C. ÇAKMAK)

Coşkun ÇAKMAK  <https://orcid.org/0000-0002-8138-272X>

Mürsel ERDAL  <https://orcid.org/0000-0002-9338-6162>

Gönderi: 31 Temmuz 2024

Kabul: 05 Kasım 2024

Yayınlanma: 15 Kasım 2024

Received: July 31, 2024

Accepted: November 05, 2024

Published: November 15, 2024

Cite as: Çakmak C, Erdal M. 2024. Estimation of natural gas pipeline construction costs using multiple linear regression and k-nearest neighbor methods. BSJ Eng Sci, 7(6): 1327-1337.

1. Giriş

Kabul edilebilir bir doğruluk seviyesinde yapılmış maliyet tahminleri proje başarısını doğrudan etkilemektedir (Adeli ve Wu, 1998; Kim ve ark., 2004; Arage ve Dharwadkar, 2017; Ibrahim ve Elshwafy, 2021; Sueri ve Erdal, 2022). İnşaat sektörü projelerinde de ön maliyet tahmininin doğruluğu, proje maliyeti, işin

tamamlanma süresi ve ürün/işçilik kalitelerinin belirlenmesinde kritik bir rol oynamaktadır. İnşaat sektörü, nicelik ve nitelik yönünden diğer sektörlerle göre daha fazla risk faktörü ve belirsizlik parametresi barındırmaktadır (Birgönül ve Dikmen, 1996). Bu nedenle doğru ve güvenilir maliyet tahminleri yapmanın önemi daha da önem kazanmaktadır.



Doğal gaz boru hattı (DGBH) projeleri, büyük ölçekli yatırımlardır ve belirsizlik, hat uzunluğu ile doğru orantılı bir şekilde artmaktadır. Bu projelerde maliyet tahminlerinin doğruluğu, projelerin finansal sürdürülebilirliği ve operasyonel verimliliği açısından da hayati bir öneme sahiptir (Ugur ve ark., 2019). Hat uzunluğu, boru çapı, kullanılan malzemeler ve işçilik maliyetleri gibi değişkenler, proje maliyetini doğrudan etkileyen faktörler arasındadır (Parker, 2004; Rui ve ark., 2011a; Thaduri, 2012; Govan ve Reinschmidt, 2013; Erdal, 2021; Kaiser, 2021; Çakmak ve Erdal, 2022). Bu faktörlerin doğru bir şekilde tahmin edilmesi, projenin bütçe aşımına uğramadan tamamlanmasını ve kaynakların etkin bir şekilde kullanımını sağlamak için elzemdir.

Maliyet tahminlerinin doğruluğu, sadece proje yöneticileri için değil, aynı zamanda yatırımcılar, müteahhitler ve tedarikçiler için de kritik bir öneme sahiptir (Çakmak ve Erdal, 2022). Doğru maliyet tahminleri, yatırımcıların nakit akış eğrilerini daha hassas yönetmelerini sağlarken, uygulayıcı paydaşların da iş planlarını daha doğru bir şekilde yapmalarına olanak tanır.

Bu çalışmanın amacı, Türkiye'de gerçekleştirilen DGBH projelerinin maliyet tahminlerine odaklanmaktadır. Veri seti, 1997-2022 yılları arasında tamamlanmış projeleri içermektedir. Ancak, elde edilen verilerin sınırlı olması ve bazı projelere ilişkin eksik bilgiler, çalışmanın genel geçerliliğini kısıtlayabilir. Gelecekte yapılacak çalışmalar için daha geniş ve kapsamlı veri setlerinin kullanılması, maliyet tahminlerinin doğruluğunu artıracaktır. Kısıtlı bilgilerle maliyet tahmini yapabilmek için Çoklu Doğrusal Regresyon (ÇDR) ve K-En Yakın Komşu (KNN) modelleri oluşturulmuş ve sonuçları karşılaştırılmıştır.

2. Literatür Araştırması

İnşaat sektörü özelinde maliyet tahminini konu edinen ve regresyon ile hafıza tabanlı yöntemlerin kullanıldığı çalışma sayısı oldukça fazladır. Ancak özel olarak DGBH projelerine odaklanılan çalışma sayısı, diğer üst yapı ve alt yapı projelerine oranla literatürde kısıtlı bir alanı kaplamaktadır. Bu çalışmaların çoğunluğu Kuzey Amerika kıtasında yer alan projelerin verileri ile ortaya konan tahmin çalışmalarıdır. Yerel literatürde, DGBH projeleri ile ilgili bu projelerin jeopolitik önemi ve enerji dünyasındaki yerini tartışan çalışmalar bulunmakla birlikte maliyet anlamında irdelendiği herhangi bir çalışmaya rastlanılmamıştır. Bu çalışma, Türkiye'de gerçekleştirilen projelerin maliyet tahminlerine odaklanarak, literatürdeki bu boşluğu doldurmayı amaçlamaktadır. Bu bölümde kronolojik sıra göz önüne alınarak DGBH inşaatlarının maliyet tahminine odaklanılan çalışmalar sunulmuştur.

Zhao (2000) çalışmasında, DGBH inşaat maliyetlerinde kara üzerinde yapılanlar için %80, kara üzerinde olmayanlar için ise %84 kadarlık payı malzeme ve işçilik maliyetinin kapladığını belirtmiştir. Parker (2004) çalışmasında, doğal gaz, akaryakıt ve petrol boru

hatlarının maliyetlerinden yola çıkarak hidrojen iletim hatlarında ne tür maliyet farklılıkları olabileceğini araştırmıştır. Parker, yaptığı araştırma ile birlikte doğal gaz boru hattı inşaat maliyetlerinin %26 malzeme maliyeti, %45 işçilik, %22 irtifak hakkı ve %7 çeşitli (araştırma, inceleme, sigortalar, izinler vb.) giderlerden oluştuğu sonucuna varmıştır. Yılmaz (2005) çalışmasında, petrol ve DGBH ile ilgili genel bir değerlendirme yapmıştır. İnşası devam eden ve inşaatı planlanan boru hatlarının detaylarının verildiği çalışmada araştırmacı, boru hatlarının Türkiye ekonomisine katkıda bulunacağını dile getirmiştir. Rui ve ark. (2011) çalışmalarında, 1992-2008 yılları arasında tamamlanmış 412 adet boru hattı projesinin maliyetlerini malzeme, işçilik, geçiş hakkı ve genel/çeşitli giderler üzerinden değerlendirmişlerdir. Çalışmada, işin tekrar tekrar yapılarak öğrenilmesinin işçilik maliyetlerini zamanla aşağıya doğru çektiğini belirtmişlerdir. Araştırmacılar yaptıkları bir diğer çalışmada ise aynı boru hatlarının bölgesel düzeyde maliyet farklılıklarını ortaya koymayı amaçlamışlardır. Çalışma kapsamında Amerika Birleşik Devletleri (ABD); Batı, Merkez, Güney Batı, Güney Doğu, Orta Batı ve Kuzey Doğu olarak farklı bölgelere bölünmüş ve en uygun maliyetlere Merkez bölgede ulaşıldığı belirlenmiştir (Rui ve ark., 2011a). Thaduri (2012) çalışmasında, 2000-2008 yılları arasında ABD ve Kanada sınırları içerisinde tamamlanmış 180 adet doğal gaz boru hattına ve 136 adet basınç istasyonuna ait verileri toplamış ve tanımlayıcı istatistik verileri sunmuştur. Yapılan değerlendirmeler sonucunda araştırmacı, doğal gaz boru hattı maliyetlerinde en yüksek paya sahip maliyet kalemlerinin %40 ile işçilik ve %31 ile malzeme maliyeti olduğunu belirtmiştir. Rui ve ark. (2012), çalışmalarında boru hattı basınç istasyonu inşaatlarındaki maliyet hatalarını istatistiksel olarak sunmuşlardır. Ulvestad ve Overland (2012) doğal gaz ve CO₂ fiyatlarındaki değişimlerin LNG ve boru hattı taşımacılığının maliyet verimliliği üzerindeki etkilerini araştırdıkları bir çalışma gerçekleştirmişlerdir. Kaiser (2021) çalışmasında, ABD'deki doğal gaz boru hattı inşaatı ve devre dışı bırakma maliyetlerini detaylandırmış ve bu maliyetlerin bileşenleri ile teknik yeterliliklerin etkilerini incelemiştir. Ayrıca çalışmada, 2014-2019 yılları arasında gerçekleştirilen 99 adet projeye ait verilerle yapılan regresyon analizinde, boru hatlarının maliyetine hat uzunluğu ve boru çapı gibi faktörlerin önemli etkilerinin olduğu belirtilmiştir. Erdal (2021) çalışmasında, boru hattı projelerinin maliyet tahmini için 50 projelik bir veri seti kullanarak 6 farklı makine öğrenmesi algoritmasının başarı düzeyini incelemiştir. 10 katlı çapraz doğrulama tekniği uygulanan çalışmada, tüm yöntemler 0,79 ve üzerinde R² değeri elde etmiş ve Model Trees Regression yöntemi en başarılı yöntem olarak belirlenmiştir.

3. Materyal ve Yöntem

DGBH projeleri için bir erken maliyet tahmin modeli geliştirmek amacıyla Türkiye sınırları içerisinde yapılmış

olan 107 adet projenin bilgilerine ulaşılmıştır. İnternet siteleri, ihale duyuru ve bilgilendirme kaynakları, yüklenici firmaların kamuya açık sunduğu bilgiler, geçmiş meclis tutanakları ve Kamu İhale Kurumu sayfaları incelenerek, projelerin maliyet tahmininde kullanılacak parametreler elde edilmiştir. Detaylı incelemeler sonucunda 1997-2022 yılları arasında yapımı tamamlanmış DGBH projelerinin boru çapı, boru hattı uzunluğu, hat vanası sayısı, take-off vana sayısı ve

pig istasyonu sayısı bilgileri, maliyet tahmininde kullanılacak bağımsız değişkenler olarak belirlenmiştir. Bu parametrelerin maliyet tahmini konusunda fikir verebileceği değerlendirilmiştir. Bu çalışma kapsamında metin içerisinde ve grafiklerde kullanılan değişkenlerin bilgi ve kısaltmaları Tablo 1’de, elde edilen verilerin tanımlayıcı istatistikleri ise Tablo 2’de verilmiştir.

Tablo 1. Değişkenlerin araştırma içerisinde gösterimi

Değişken	Model içerisinde kullanım	Kısaltma
Çap	diameter	D
Hat uzunluğu	length	L
Hat vanası	line_valve	LV
Take-off vanası	take_off_valve	TOV
Pig istasyonu	pigging_station	PS
Km başına birim maliyet	cost_km	-
Düzeltilmiş maliyet	adj_cost	C

Tablo 2. Veri setinin tanımlayıcı istatistikleri

Değişkenler	Adet	Ortalama	Standart Sapma	En Küçük	25. Yüzdelik	Medyan	75. Yüzdelik	En Büyük
diameter (inç)	107	19,68	13,37	6,00	10,00	14,00	36,00	48,00
length (km)	107	80,20	77,36	1,57	32,32	59,00	102,65	501,00
line_valve	107	2,40	2,73	0,00	1,00	2,00	3,00	17,00
take_off_valve	107	1,93	3,44	0,00	1,00	1,00	2,00	34,00
pigging_station	107	1,55	0,97	0,00	1,00	1,00	2,00	6,00
cost_km (bin \$)	107	304,96	366,40	29,51	75,75	157,62	374,33	1.627,34
adj_cost (milyon \$)	107	34,72	79,25	0,91	3,15	6,20	27,67	563,35

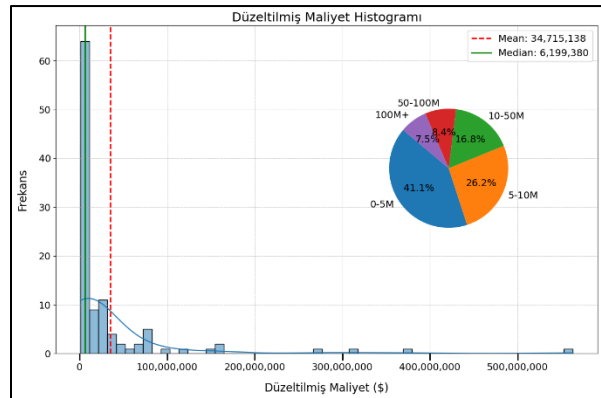
Tablo 2’de yer alan adj_cost parametresi, projelerin tamamlandığı tarihlerde ortaya çıkan ve Amerikan Doları cinsinden belirtilen maliyet değerlerinin, ABD’de kullanılan Consumer Price Index (CPI) değerlerine göre Mayıs 2024 tarihine dönüştürülmüş hâlini ifade etmektedir. Örnek bir hesaplama aşağıda sunulmuştur.

Mart 2004 tarihinde tamamlanmış bir projenin maliyeti yaklaşık 14.000.000,00 \$ civarındadır. İlgili tarihte ABD tüketici fiyat endeksi (CPI) değeri ise 187,1 şeklindedir. Mayıs 2024 CPI değeri ise 313,225’tir. Bu durumda projenin düzeltilmiş maliyeti aşağıda şekilde ortaya çıkar. (eşitlik 1);

$$\begin{aligned} & \text{Düzeltilmiş maliyet} \\ &= \frac{\text{Proje Tamamlanan Dönem Maliyeti}}{\text{Proje Tamamlanan Dönem CPI}} \times \text{Dönüştürülecek ayın CPI değeri} \quad (1) \\ & \text{Düzeltilmiş maliyet} = \frac{14.000.000,00}{187,1} \times 313,225 \\ &= 23.437.467,60 \$ \end{aligned}$$

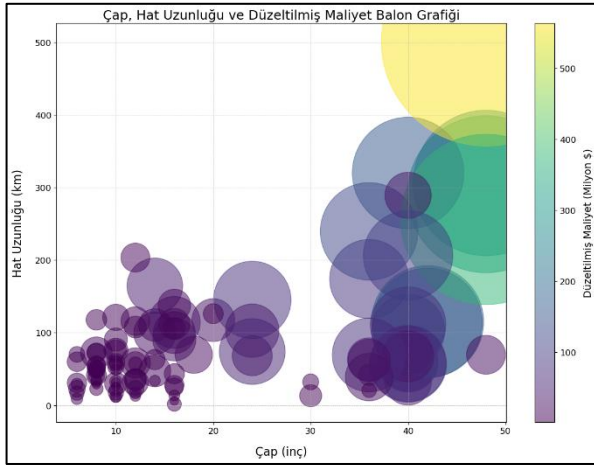
Tablo 2’de yer alan, cost_km parametresi ise hesaplanan adj_cost parametresinin DGBH uzunluğuna bölünmesi ile elde edilen bir değişkendir. Bu parametre ile DGBH projelerinde km başına oluşan birim maliyetler belirlenmiştir. cost_km değişkeni ayrıca bu çalışmanın model bölümünde bahsedilen uç değer atma işlemi için

de kullanılacak bir parametredir. Bu değişken ile birlikte adj_cost ve length değişkenleri birlikte değerlendirilmiş ve uç değer tespit adımında kullanılmıştır. Veri setinde küçük, orta, büyük ve mega projeler yer aldığından, büyük ve mega proje sayısının diğer proje sayısından daha az olmasından kaynaklı sağa çarpık bir veri dağılımı söz konusudur. Veri setinin düzeltilmiş maliyet histogram grafiği Şekil 1’de verilmiştir.



Şekil 1. Maliyet histogram grafiği.

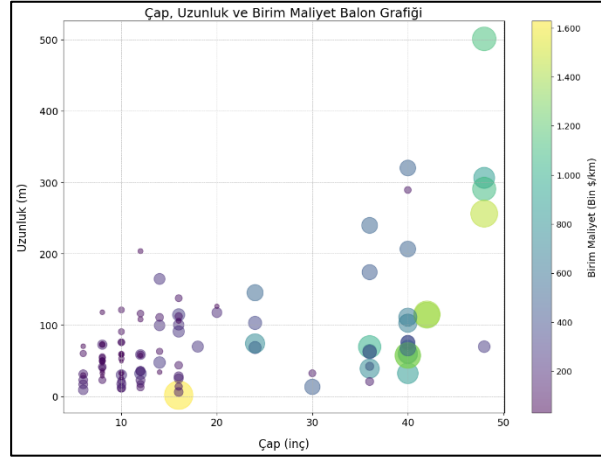
Veri setinin daha iyi anlaşılabilmesi için boru çapı ve hat uzunluğuna bağlı olarak maliyet değişkeninin yayılımını gösteren grafik Şekil 2’de gösterilmiştir.



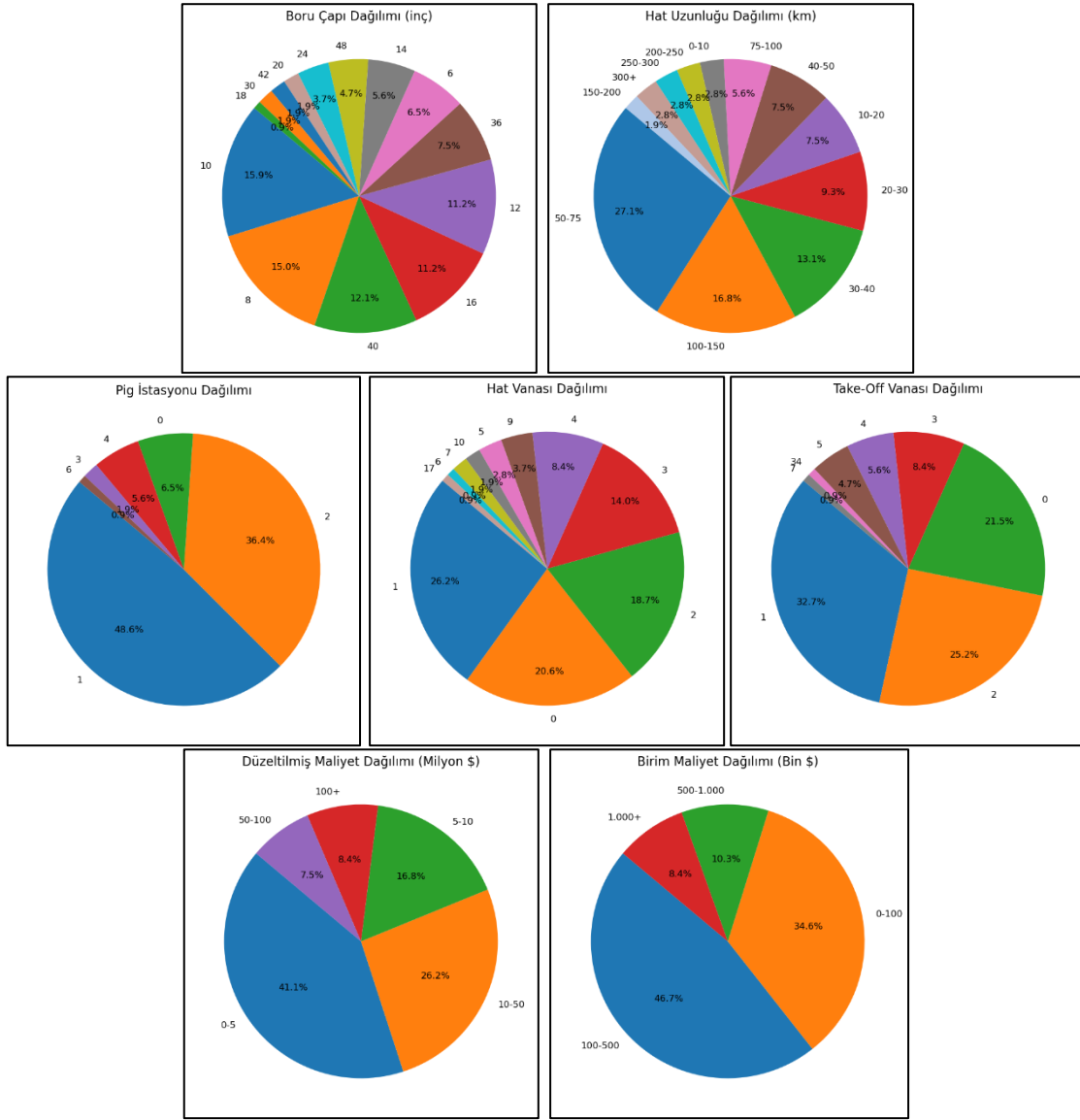
Şekil 2. Çap ve hat uzunluğu değişkenlerine göre toplam proje maliyeti.

Şekil 2 detaylı bir şekilde incelendiğinde, genel eğilim boru çapı ve hat uzunluğunun artmasıyla proje maliyetinin de arttığı yönündedir. Bu durumda, bu değişkenlerle birlikte bağımlı değişken olan düzeltilmiş maliyet verisinin doğru orantılı olduğu yorumunu yapmak mümkündür. Ancak bazen düşük çap ve hat uzunluklarında dâhi yüksek maliyetlerin ortaya çıkabildiği de grafikten görülebilir. Her iki parametre ile birlikte, km başına ortaya çıkan birim maliyet verisinde de orantılı bir artış gerçekleşmektedir. Ancak buna

rağmen düşük hat uzunluğuna ve boru çapına sahip projelerde de yüksek km başına birim maliyet değerleri gözlemlenebilmektedir. Hatta en yüksek km başına birim maliyet verisi, çap ve hat uzunluğu değişkeninin oldukça düşük olduğu bir bölgede yer almaktadır. Çap ve hat uzunluğu parametrelerine karşın km başına düşen birim maliyeti gösteren grafik Şekil 3'te sunulmuştur. Ayrıca veri setinde yer alan tüm bağımlı-bağımsız değişkenlerin pasta grafikleri Şekil 4'te gösterilmiştir.



Şekil 3. Çap ve hat uzunluğu değişkenlerine göre km başına maliyet.



Şekil 4. DGBH proje maliyet tahmin parametrelerinin pasta grafikleri.

3.1. Modeller

Mevcut veri seti içerisinde yer alan, veriyi sağa çarpık kılan ve uç değerler olarak adlandırılan değişkenler analiz sonuçlarını olumsuz yönde etkilemektedir. Bu durum ortaya konulan modellerin başarı metriklerini de aşağıya doğru çekmektedir. Bundan dolayı, veri setinin bir miktar daha normal dağılıma yaklaşmasını sağlayacak olan uç değerli veri atma işlemi gerçekleştirilmiştir. IQR (Interquartile Range) ile uç değer tespit işlemi yaygın olarak kullanılmaktadır (Hubert ve ark., 2007). Bu araştırma özelinde veri setindeki uç değerleri tespit etmek için de IQR yöntemi kullanılmıştır. IQR değerinin hesaplanması için aşağıdaki matematiksel işlemin (Eşitlik 2) yapılması gerekmektedir.

$$IQR = Q3 - Q1 \quad (2)$$

Burada;

- Q1 = Toplam veri sayısının en düşük %25'lik kısmını ayıran değerdir.
- Q3 = Toplam veri sayısının en düşük %75'lik kısmını (en yüksek %25) ayıran değerdir.

IQR hesaplandıktan sonra uç değer tespiti için alt sınır ve üst sınırların hesaplanması gerekmektedir. Sınır değerler aşağıdaki şekilde (Eşitlik 3 ve Eşitlik 4) hesaplanmıştır.

$$\text{Alt sınır} = Q1 - 1.5 \times IQR \quad (3)$$

$$\text{Üst sınır} = Q3 + 1.5 \times IQR \quad (4)$$

Gözlem değerlerinde alt sınırın altında kalan ve/veya üst sınırın üstünde olan tüm gözlemler uç değer olarak nitelendirilmiş ve veri setinden çıkarılmıştır. Jupyter notebook ortamında Python aracılığıyla yazılan kod Şekil 5'te verilmiştir. Sonuç çıktısı ise Tablo 3'de sunulmuştur.

```
# IQR yöntemi ile outlier tespiti ve temizleme fonksiyonu
def remove_outliers_iqr(df, columns):
    Q1 = df[columns].quantile(0.25)
    Q3 = df[columns].quantile(0.75)
    IQR = Q3 - Q1
    is_not_outlier = ~((df[columns] < (Q1 - 1.5 * IQR)) | (df[columns] > (Q3 + 1.5 * IQR))).any(axis=1)
    return df[is_not_outlier]
```

Şekil 5. Uç değer tespit fonksiyonu kodu

Tablo 3. Modele esas veri seti tanımlayıcı istatistikleri

Değişkenler	Adet	Ortalama	Standart Sapma	En Küçük	25. Yüzdalık	Medyan	75. Yüzdalık	En Büyük
diameter (inç)	88	15,48	10,20	6,00	8,00	12,00	16,00	48,00
length (km)	88	59,01	37,53	6,00	31,64	54,53	74,29	203,47
line_valve	88	1,69	1,49	0,00	1,00	1,00	3,00	7,00
take_off_valve	88	1,99	3,72	0,00	1,00	1,00	2,00	34,00
pigging_station	88	1,52	0,96	0,00	1,00	1,00	2,00	6,00
cost_km (bin \$)	88	173,40	164,98	29,51	64,93	114,54	206,09	776,82
adj_cost (milyon \$)	88	10,33	11,93	0,91	2,54	5,19	12,91	55,60

Tablo 3’de yer alan tanımlayıcı istatistikler, uç değerlerden temizlenmiş veri setinin istatistikleridir. Bu veri seti ile modeller oluşturulmuş ve model başarı parametreleri R^2 değerleri üzerinden gözlemlenmiştir. Karşılaştırma yapmak için kök ortalama kare hatası (RMSE) değeri hesaplanmıştır. R^2 , modelde bağımsız değişkenlerin bağımlı değişkeni ne kadar iyi açıkladığını gösteren istatistiksel bir ölçüttür. 0 ile 1 arasında bir değer alır ve değer 1’e yakın olması, modelin veriyi daha iyi açıkladığı anlamına gelir. RMSE ise modelin tahmin ettiği değerler ile gerçek değerler arasındaki ortalama farkı ölçen bir hata metriğidir. Değerin düşük olması, modelin tahminlerinin gerçeğe daha yakın olduğunu gösterir. R^2 ve RMSE değerlerinin hesaplanma yöntemi Eşitlik 5 ve Eşitlik 6’da gösterilmiştir.

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

Burada;

- y_i , i-inci gözlemin gerçek değerini,
- \bar{y} , bağımlı değişkenin ortalama değerini,
- \hat{y}_i , modelin tahmin ettiği i-inci gözlemin değerini,
- n , toplam gözlem sayısını temsil eder.

3.1.1. Çoklu doğrusal regresyon (ÇDR)

ÇDR, bir bağımlı değişkenin birden fazla bağımsız değişken ile olan ilişkisinin modellenmesine odaklanılan basit bir yöntemdir. ÇDR, bağımsız değişkenlerin bağımlı değişken üzerindeki etkilerini anlamak, tahmin etmek ve analiz etmek için kullanılır. ÇDR modeli, aşağıdaki genel denkleme (Eşitlik 7) dayanır:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (7)$$

Burada;

- β_0 modelin sabit terimidir.
- $\beta_1, \beta_2, \dots, \beta_n$ her bir bağımsız değişkenin katsayılarıdır. Bu katsayılar, bağımsız değişkenlerin

bağımlı değişken üzerindeki etkisini gösterir.

- ε ise hata terimidir ve modelin açıklayamadığı rastgele hataları içerir.

ÇDR analizi statsmodels kütüphanesi kullanılarak yapılmıştır. Statsmodels kütüphanesi bu tarz regresyon problemleri için kullanıcıya sağladığı ve model hakkında bilgi veren yararlı ara yüzüyle istatistiksel modellemeler amacıyla yaygın olarak kullanılan kütüphanelerden birisidir (Seabold ve Perktold, 2010). Model süreci için adımlar ve model çıktıları aşağıda verilmiştir.

- a) Model süreci aşamaları için gerekli kütüphanelerin import edilmesi: Jupyter notebook ortamında kütüphanelerin yüklenebilmesi için genellikle import ifadesi kullanılır. Şekil 6’daki kod parçasığı bu işlemin yapılması için yazılması gereken metni göstermektedir.

```
import statsmodels.api as sm
```

Şekil 6. Statsmodels kütüphanesi import kodu.

- b) Verinin okunması: Bu adımda Pandas kütüphanesi yardımı ile excel dosyasındaki veriler okunmuştur. Excel dokümanında istenen sütun değerlerinin okutulabildiği kod parçasığı Şekil 7’de verilmiştir.

```
ngpc = pd.read_excel(
    "C:\\Users\\cosku\\ngpc_veriler.xlsx",
    usecols=["diameter",
            "length",
            "line_valve",
            "take_off_valve",
            "pigging_station",
            "adj_cost",
            "cost_km"] )
```

Şekil 7. Excel verilerinin okutulması.

- c) Bağımlı-bağımsız değişkenlerin belirlenmesi ve sabit eklenmesi: Bu adımda excel dokümanından çekilen değişkenlerin hangilerinin bağımsız, hangilerinin bağımlı değişken olduğunun modele

tanıtılma işlemi yapılmaktadır. Şekil 8'deki kod parçası "ngpc" isimli veri setinin değişkenlerini gruplamaktadır. Statsmodels içerisindeki "add_constant" fonksiyonu yardımıyla da bağımsız değişken X'e sabit eklenmektedir. Böylece Eşitlik 7'de yer alan β_0 sabiti, sonuç denklemi elde edildiğinde denklem içerisinde yer alabilmektedir.

```
X = ngpc.drop(["adj_cost", "cost_km"], axis=1)
y = ngpc["adj_cost"]
X=sm.add_constant(X)
```

Şekil 8. Değişken gruplaması ve sabit eklenmesi.

d) Modelin fit edilmesi ve sonuçların gösterimi: Makine öğrenmesi algoritmalarında verilerin modele uyup

uymadığının kontrolü ve modelin eğitilmesi amacıyla "fit()" fonksiyonu kullanılmaktadır. Bu adımda değişkenler statsmodels kütüphanesinin ÇDR algoritmasına gönderilmekte ve bağımsız değişkenlerin bağımlı değişken üzerindeki açıklayabilme oranları tespit edilmektedir. Bu işlemin gerçekleşmesini sağlayan kod parçasığı Şekil 9'da gösterilmiştir.

```
tüm_veri_çdr = sm.OLS(y, X).fit()
tüm_veri_çdr.summary()
```

Şekil 9. Değişkenlerin modele fit edilmesi.

Bu adımda yer alan summary fonksiyonu ile Tablo 4 ve Tablo 5'de yer alan sonuçlar elde edilmiştir.

Tablo 4. Regresyon modeli başarı metrikleri

Dep. Variable:	adj_cost	R-squared:	0,581
Model:	OLS	Adj. R-squared:	0,556
Method:	Least Squares	F-statistic:	22,75
No. Observations:	88	Prob (F-statistic):	3,01e-14
Df Residuals:	82	Log-Likelihood:	-1520,0
Df Model:	5	AIC:	3052
Covariance Type:	nonrobust	BIC:	3067

Tablo 5. Değişken katsayıları ve anlamlılık düzeyleri

Parametreler	coef	std err	t	P> t	[0.025	0.975]
const	-4,43e+06	2,69e+06	-1,647	0,103	-9,79e+06	9,22e+05
diameter	6,35+05	9,21e+04	6,893	0,000	4,52e+05	8,18e+05
length	1,10e+05	4,04e+04	2,717	0,008	2,94e+04	1,9e+05
line_valve	5,32e+06	1,04e+06	0,511	0,611	-1,54e+06	2,6e+06
take_off_valve	2,37e+05	2,39e+05	0,992	0,324	-2,39e+05	7,13e+05
pigging_station	-1,91e+06	9,75e+05	-1,963	0,053	-3,85e+06	2,56e+04

Tablo 4 incelendiğinde model, bağımlı değişkendeki değişkenliğin %58 oranında açıklanabildiğini göstermektedir. Tablo 5, boru çapı ve hat uzunluğu parametrelerinin, proje maliyeti üzerinde en etkili değişkenler olduğunu ortaya koymaktadır. Diğer değişkenlerin ise 0,05'ten daha büyük p değerlerine sahip olmaları sebebi ile ÇDR ile model kurulurken bağımlı değişken üzerindeki varyans değerinin açıklanmasında yetersiz oldukları sonucu ortaya çıkmaktadır. Ancak pigging_station değişkeni sınır değere oldukça yakın olduğundan dolayı tekrarlanacak olan modelde kullanılmasına karar verilmiştir. Bu kapsamda ÇDR modellemesi için boru çapı, hat uzunluğu ve pig istasyonu parametreleri kullanılıp, diğer parametreler

dışarıda tutularak tekrarlanmıştır. Ayrıca model içerisinde bir sabit olmasının anlamlı olmadığı 0,103 (P>0,05) değerine sahip olan p değerinden anlaşılmaktadır. Dolayısıyla bu durum modelin başarısını etkileyebileceğinden, düzenlenecek olan modelde veri setine sabit ekleme işlemi gerçekleştirilmemiştir. Böylece genelleme açısından daha doğru bir modelin ortaya çıkacağı değerlendirilmiştir. Şekil 10'da yalnızca model için anlamlı olan değişkenlerin tutulmasına, diğer değişkenlerin dışarıda kalmasına ve denklemde sabit bir katsayı olmadan regresyon kurma işlemi gerçekleştirmeye olanak tanıyacak kod parçasığına yer verilmiştir. Tablo 6 ve Tablo 7'de ise bu kod parçasığının sonuç çıktısı yer almaktadır.

```
X = ngpc_clean.drop(["adj_cost", "cost_km", "line_valve", "take_off_valve"], axis=1)
y = ngpc_clean["adj_cost"]

tüm_veri_çdr = sm.OLS(y, X).fit()
tüm_veri_çdr.summary()
```

Şekil 10. İstatistiksel olarak anlamlı (p<0,05) bağımsız değişkenlerle model fit etme işlemi

Tablo 6. Üç değişkenli regresyon modeli başarı metrikleri

Dep. Variable:	adj_cost	R-squared:	0,748
Model:	OLS	Adj. R-squared:	0,739
Method:	Least Squares	F-statistic:	53,65
No. Observations:	88	Prob (F-statistic):	8,54e-16
Df Residuals:	85	Log-Likelihood:	-1552,4
Df Model:	3	AIC:	3051
Covariance Type:	nonrobust	BIC:	3058

Tablo 7. Üç değişkenli regresyon modeli değişken katsayıları ve anlamlılık düzeyleri

Parametreler	coef	std err	t	P> t	[0.025	0.975]
diameter	5,39e+05	6,89e+04	7,825	0,000	4,02e+05	6,76e+05
length	1,171e+05	2,21e+04	5,306	0,000	7,32e+04	1,61e+05
pigging_station	-2,887e+06	7,38e+05	-3,912	0,000	-4,35e+06	-1,42e+06

Eşitlik 6 aracılığıyla hesaplanan RMSE değeri ise 7.679.373,12 \$ düzeyinde hesaplanmıştır. Bu değer uç veriler atıldıktan sonra ortaya çıkan tanımlayıcı istatistiklere bakıldığında düzeltilmiş maliyet verisinin ortalamasına yakın bir değerdir ve görece yüksek olduğu değerlendirilebilir.

Son durumda DGBH projelerinin maliyet tahmininde kullanılabilecek model denklemi Eşitlik 8'de verilmiştir.

$$C(\$)=539.049,5 \times D + 117.093,69 \times L - 2.887.226,57 \times PS \quad (8)$$

3.1.2. K-en yakın komşu (KNN)

KNN, gözetimli öğrenme algoritmaları arasında yer alan ve hem sınıflandırma hem de regresyon problemlerinde kullanılan basit ve etkili bir yöntemdir. Bu yöntem, yeni bir verinin sınıfını belirlemek veya değerini tahmin etmek için en yakın K komşusunun etiketlerine veya değerlerine bakar. Bu algoritma, belirli bir noktanın yakınındaki diğer noktalarla olan benzerliğine dayanarak çalışır (Bentley, 1975).

KNN algoritması şu temel adımları içerir:

- I. **Veri Kümesinin Hazırlanması:** Algoritma, veriler üzerinde çalışır. Her bir veri noktası, özellik vektörlerine ve etiketlerine sahiptir.
- II. **Mesafe Hesaplama:** Yeni bir veri noktası geldiğinde, bu noktanın tüm eğitim veri noktalarına olan mesafesi hesaplanır. Genellikle Öklidyen mesafesi kullanılır, ancak Manhattan veya Minkowski mesafesi gibi diğer mesafe metrikleri de kullanılabilir. Öklidyen mesafesi, iki veri noktası arasındaki en kısa doğrusal mesafeyi hesaplar ve Eşitlik 9 ile ifade edilir:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (9)$$

Manhattan mesafesi, koordinat eksenlerine paralel yolların toplam uzunluğunu hesaplar (Eşitlik 10).

$$d(p, q) = \sum_{i=1}^n |p_i - q_i| \quad (10)$$

Minkowski mesafesi, genelleştirilmiş bir mesafe hesaplama yöntemidir ve Eşitlik 11'de verilmiştir.

$$d(p, q) = \left(\sum_{i=1}^n |p_i - q_i|^p \right)^{1/p} \quad (11)$$

Burada:

- p ve q veri noktalarını,
- p_i ve q_i veri noktalarının i'inci özellik değerlerini temsil eder.

III. **K Komşusunun Seçimi:** Hesaplanan mesafelerden en küçük olan K tanesi seçilir. Bu komşular, yeni veri noktasının en yakın K komşusu olarak belirlenir.

IV. **Tahmin Yapma:** Regresyon problemlerinde, K komşusunun ortalama değeri alınarak tahmin yapılır. KNN regresyon için kullanılan karar kuralı Eşitlik 12 ile hesaplanır:

$$\hat{y} = \frac{1}{K} \sum_{i=1}^K y_{N_k(x_i)} \quad (12)$$

Burada:

- \hat{y} yeni veri noktasının tahmin edilen değerini,
- $y_{N_k(x_i)}$ K komşusunun gerçek değerlerini temsil eder.

Bu bölümde scikit-learn kütüphanesi kullanılarak KNN analizi yapılmıştır. Scikit-learn, makine öğrenmesi algoritmaları ve veri işleme adımları için geniş kapsamlı bir kütüphanedir (Buitinck ve ark., 2011). KNN modeli için adımlar ve çıktıları aşağıda açıklanmıştır.

KNN, yukarıda bahsedildiği şekilde mesafelere dayalı bir tahmin modeli kurduğu ve mevcut veri setinde birbirinden farklı ölçeklerde değişkenler yer aldığı için değişkenlerin ölçeklendirilmesi gereklidir. Bu bölümde ÇDR düzeyinde basit olmayan, modelin en iyi parametrelerinin araştırıldığı (hiperparametre ayarlama) bir model oluşturulmuştur. Model için adımlar ve çıktıları aşağıda açıklanmıştır.

- a) Model aşamaları için gerekli kütüphanelerin import edilmesi: KNN modeli için gerekli import işlemini gerçekleştiren kod Şekil 11'de gösterilmiştir.

```
from sklearn.neighbors import KNeighborsRegressor
```

Şekil 11. KNN kütüphanesi import kodu.

- b) Verinin okunması: Veri seti değerlerinin okutulabildiği kod parçacığı Şekil 12'de verilmiştir.


```
ngpc = pd.read_excel(
    "C:\\Users\\cosku\\ngpc_veriler.xlsx",
    usecols=["diameter",
            "length",
            "line_valve",
            "take_off_valve",
            "pigging_station",
            "adj_cost",
            "cost_km"] )
```

Şekil 12. Excel verilerinin okutulması.

- c) Uç değerler atılarak bağımlı-bağımsız değişkenlerin belirlenmesi; IQR yönteminin fonksiyonu ve değişken sınıflamasını sağlayan kod Şekil 13'te verilmiştir.

```
outlier_degiskenler = ["cost_km", "adj_cost", "length"]
ngpc_clean= remove_outliers_iqr(ngpc, outlier_degiskenler)
X=ngpc_clean.drop(["cost_km", "adj_cost"], axis=1)
y=ngpc_clean["adj_cost"]
```

Şekil 13. Uç değerlerin atılması ve değişken gruplaması.

- d) Standardizasyon: Standartlaştırma, KNN algoritmasının performansını ve doğruluğunu artıran kritik bir adımdır. Bu süreç, farklı birimlerdeki değişkenlerin aynı ölçek üzerinde karşılaştırılabilir olmasını sağlar. Özellikle veri setindeki değişkenlerin farklı ölçeklerde olduğu durumlarda, standartlaştırma adımı KNN algoritmasının daha etkili çalışmasına katkıda bulunur. Scikit-learn bünyesinde yer alan StandardScaler'ın import edilmesi ve değişkenlerin standart sapmasının 1, ortalamalarının ise 0 olmasını sağlayacak kod parçacığı Şekil 14'te verilmiştir.

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

Şekil 14. Değişkenlerin standartlaştırılması.

- e) Modelin oluşturulması ve fit edilmesi: Şekil 15'te KNN model nesnesinin hiperparametreleri görülmektedir.

```
# KNN modeli (Tune Öncesi)
knn = KNeighborsRegressor()
knn.fit(X_scaled, y)
knn
```

```
KNeighborsRegressor(algorithm='auto', leaf_size=30, metric='minkowski',
                    metric_params=None, n_jobs=None, n_neighbors=5, p=2,
                    weights='uniform')
```

Şekil 15. Değişkenlerin modele fit edilmesi ve varsayılan parametre değerleri.

Şekil 15'ten görüldüğü üzere KNN modeli varsayılan olarak belirli parametrelerle modeli oluşturmaktadır. Bunlardan en önemlileri ve sonraki adımlarda ayarlamaya çalışılacak olanlar ise `n_neighbors` ve `leaf_size` parametreleridir. `n_neighbors` parametresi yeni bir veri noktası oluşturulurken modelin en yakınındaki

kaç gözleme uğrayacağını belirtir. `leaf_size` ise KNN içerisinde ağaç yapıları şeklinde saklanan veri kümelerinde en fazla ne kadar veri tutulması gerektiğini gösteren bir parametredir. Modelde görülen değerler bir başlangıç noktası olarak düşünülmelidir. Bu başlangıç noktasının en iyi değerler olabileceği ihtimâli de göz ardı edilmemelidir.

- f) Tahmin ve sonuçlar: Varsayılan parametreler ile Şekil 16'da kod parçacığı gösterilen model çalıştırıldığında $R^2=0,690$ ve $RMSE=6.625.707,23$ \$ değerleri elde edilmiştir. Bu öncül değerlendirmenin ardından modelin `n_estimator` ve `leaf_size` parametreleri değiştirilerek, en iyi belirleme katsayısına ulaşılmaya çalışılmıştır.

```
y_pred = knn.predict(X_scaled)
R2_score_train_BT = r2_score(y, y_pred)
RMSE_train_BT = np.sqrt(mse(y, y_pred))
```

Şekil 16. Model başarı parametreleri olan R^2 ve RMSE hesaplaması.

- g) Hiperparametre ayarlama işlemi ve sonrasında tahmin ve sonuçlar: Hiperparametre ayarlama işlemi, yani KNN model nesnesinin en iyi hangi parametrelerle çalıştığının araştırılması sürecinde GridSearchCV yöntemi kullanılmıştır. GridSearchCV, bir modelin hiperparametrelerini sistematik olarak denemek ve en iyi performans gösteren kombinasyonu belirlemek için kullanılan bir çapraz doğrulama yöntemidir. Bu yöntem, modelin doğruluğunu artırmak ve aşırı uyum sorunlarını azaltmak amacıyla farklı parametre setlerini değerlendirir (Şekil 17).

```
knn_params = {'n_neighbors': [3, 4, 5, 6, 7, 8, 9, 10],
              'leaf_size': [20, 30, 40] }

# KNN modeli için en iyi hiperparametreleri bulma
knn_grid = GridSearchCV(KNeighborsRegressor(),
                        knn_params,
                        cv=5,
                        scoring='r2')
knn_grid.fit(X_scaled, y)

# En iyi parametreler ile KNN modelini eğitme
best_knn = knn_grid.best_estimator_

# Modeli eğit
best_knn.fit(X_scaled, y)

# Tune sonrası tahmin yap
y_pred = best_knn.predict(X_scaled)

R2_score_train_AT = r2_score(y, y_pred)
RMSE_train_AT = np.sqrt(mse(y, y_pred))
```

Şekil 17. Hiperparametre ayarlama işlemi ve model başarı metriklerinin hesaplanması.

Hiperparametre ayarlama işlemi sonrası elde edilmiş parametreler ile model çalıştırıldığında $R^2=0,660$ ve $RMSE=6.934.296,43$ \$ değerleri ortaya çıkmıştır. Bu değer ayarlanmış hiperparametreler ile elde edildiği için güvenilirliği daha fazladır.

4. Tartışma ve Sonuç

Bu çalışmada, Türkiye'de gerçekleştirilen DGBH proje maliyetlerinin, çeşitli proje parametreleri ile iki farklı model aracılığıyla tahmin başarısı belirlenmeye çalışılmıştır. Verilerde yer alan nicel kısıtlılık sebebiyle klasik model üretme ve tahmin yapma sürecinden farklı olarak mevcut veriler ÇDR ve KNN modellerine yerleştirilmiş ve maliyet üzerindeki açıklama kapasitesi düzeyi tespit edilmeye çalışılmıştır. Yöntemlerde bağımsız değişkenler olarak boru çapı, hat uzunluğu, hat vanası sayısı, take-off vana sayısı ve pig istasyonu sayısı parametreleri kullanılmıştır.

ÇDR modelinde, boru çapı ve hat uzunluğu parametreleri maliyet üzerinde en etkili değişkenler olarak belirlenmiş ve bu değişkenlerin bağımlı değişkendir varyans değerinin %75'ini açıklayabildiği ($R^2=0,748$) belirlenmiştir. KNN modelinde varsayılan parametrelerle elde edilen belirleme katsayısı $R^2 = 0,690$ olup, ayarlanmış hiperparametreler ile bu değer 0,660 olarak bulunmuştur.

Sonuçlar, mevcut veri setinin kısıtlılığına rağmen her iki yöntemin de maliyet tahmininde makul ancak yeterli olmayan bir performans sergilediğini göstermektedir. Elde edilen R^2 değerleri yüksek olmasa da, veri kalitesinin ve sayısının artması durumunda tahmin doğruluğunun önemli ölçüde artabileceğini göstermektedir. Daha geniş ve kaliteli veri setleri elde edilerek maliyet tahmin modellerinin doğruluğu artırılabilir. Mevcut veri setinin kısıtlılığı nedeniyle sadece belirli parametreler kullanılmıştır. Veri setine daha fazla ve farklı parametrelerin (Örn; zemin cinsleri, işçilik ücretleri, kamulaştırma bedelleri vb.) eklenmesinin mümkün olduğu koşullarda, hem klasik makine öğrenmesi süreçlerinin işletilebileceği hem de model performans başarı metriklerinde gözle görülür iyileşmeler sağlanacağı düşünülmektedir. SMOGN (Synthetic Minority Over-sampling Technique for Regression) gibi sentetik veri artırma yöntemleri kullanılarak bu tarz tahmin çalışmalarının tekrarlanması ve bu sayede modellerin genelleme özelliğinin artırılması mümkün olabilecektir. Farklı makine öğrenmesi algoritmaları kullanılarak (Örn; ağaç yöntemleri, rastgele ormanlar, destek vektör makineleri vb.) maliyet tahminlerinin doğruluğu artırılabilir. Ayrıca veri ön işleme süreçlerinin iyileştirilmesi ve uç değerlerin daha etkin yönetilmesinin de model performansını artırabileceği değerlendirilmiştir. Bu çalışma, doğalgaz boru hattı projelerinde maliyet tahminleri konusunda önemli bir başlangıç noktası oluşturmaktadır ve gelecekte yapılacak araştırmalara ışık tutacaktır.

Katkı Oranı Beyanı

Yazarların katkı yüzdeleri aşağıda verilmiştir. Yazarlar makaleyi incelemiş ve onaylamıştır.

	C.Ç.	M.E.
K	70	30
T	70	30
Y	10	90
VTI	40	60
VAY	80	20
KT	90	10
YZ	80	20
KI	20	80
GR	80	20
PY	10	90

K= kavram, T= tasarım, Y= yönetim, VTI= veri toplama ve/veya işleme, VAY= veri analizi ve/veya yorumlama, KT= kaynak tarama, YZ= Yazım, KI= kritik inceleme, GR= gönderim ve revizyon, PY= proje yönetimi

Çatışma Beyanı

Yazarlar bu çalışmada hiçbir çıkar ilişkisi olmadığını beyan etmektedirler.

Etik Onay Beyanı

Bu araştırmada hayvanlar ve insanlar üzerinde herhangi bir çalışma yapılmadığı için etik kurul onayı alınmamıştır.

Kaynaklar

- Adeli H, Wu M. 1998. Regularization Neural Network for construction cost estimation. J Constr Eng Manag, 124(1): 18-24.
- Arage SS, Dharwadkar NV. 2017. Cost estimation of civil construction projects using machine learning paradigm. International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), February 13-15, Tamil Nadu, India, pp: 594-599.
- Bentley JL. 1975. Multidimensional binary search trees used for associative searching. Commun ACM, 18(9): 509-517.
- Birgönül TM, Dikmen İ. 1996. İnşaat projelerinin risk yönetimi. İMO Tek Derg, 97: 1305-1326.
- Buitinck L, Louppe G, Blondel M, Fabien P, Mueller A, Olivier G, Niculae V, Prettenhofer P, Gramfort A, Grobler J, Layton R, VanderPlas J. 2011. Scikit-learn: Machine learning in Python. J Mach Learn Res, 12: 2825-2830.
- Çakmak C, Erdal M. 2022. Preliminary estimation of natural gas pipeline construction costs with regression analysis. 7th International Project and Construction Management Conference, October 10-12, İstanbul, Türkiye, pp: 328-337.
- Erdal H. 2021. Prediction of pipeline projects construction costs utilizing machine learning techniques. International Marmara Sciences Congress (Spring 2021), May 12-14, Kocaeli, Türkiye, pp: 218-223.
- Govan P, Reinschmidt K. 2013. Benchmarking natural gas pipeline projects. Pipelines 2013 Conference, June 5-7, Texas, US, pp: 1532-1542.
- Hubert M, Veeken SVD. 2007. Outlier detection for skewed data. J Chemom, 22(3-4): 235-246.
- Ibrahim AH, Elshwadfy LM. 2021. Factors affecting the accuracy

- of construction project cost estimation in Egypt. *Jordan J Civ Eng*, 15(3): 329-344.
- Kaiser MJ. 2021. A review of onshore and offshore pipeline construction and decommissioning cost in the USA - Part 1: Specifications, cost estimation and onshore construction. *Int J Oil Gas Coal Technol*, 27(3): 247-285.
- Kim GH, An SH, Kang KI. 2004. Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Build Environ*, 39(10): 1235-1242.
- Parker NC. 2004. Using natural gas transmission pipeline costs to estimate hydrogen pipeline costs. URL=<https://www.researchgate.net/publication/254396811> (accessed date: July, 15, 2024).
- Rui Z, Metz P, Wang X, Chen G, Zhou X, Reynolds D. 2012. Inaccuracy in pipeline compressor station construction cost estimation. *SPE Annual Technical Conference and Exhibition*, October 20-22, Texas, US, pp: 4219-4234.
- Rui Z, Metz PA, Reynolds DB, Chen G, Zhou X. 2011a. Regression models estimate pipeline construction costs. *Oil Gas J*, 109(14): 120-127.
- Rui Z, Metz PA, Reynolds DB, Chen G, Zhou X. 2011b. Historical pipeline construction cost analysis. *Int J Oil Gas Coal Technol*, 4(3): 244-263.
- Seabold S, Perktold J. 2010. *Statsmodels: econometric and statistical modeling with Python*. Proceedings of the 9th Python in Science Conference, June-July, Texas, US, pp: 92-96.
- Sueri M, Erdal M. 2022. Early Estimation of sewerage line costs with regression analysis. *Gazi Univ J Sci*, 35(3): 822-832.
- Thaduri RK. 2012. Oil and gas pipeline construction cost analysis and developing regression models for cost estimation. MSc Thesis, Texas A&M University, Institute of Technology & Science, Texas, US, pp: 34-35.
- Ugur LO, Kanit R, Erdal H, Namli E, Erdal HI, Baykan UN, Erdal M. 2019. Enhanced predictive models for construction costs: A Case study of Turkish mass housing sector. *Comput Econ*, 53(4): 1403-1419.
- Ulvestad M, Overland I. 2012. Natural gas and CO2 price variation: Impact on the relative cost-efficiency of LNG and pipelines. *Int J Environ Stud*, 69(3): 407-426.
- Yılmaz NF. 2005. Petrol ve doğal gaz boru hatları üzerine genel bir değerlendirme. *Tesisat Müh Derg*, 87: 4-14.
- Zhao J. 2000. Interim report IR-00-054 diffusion, costs and learning in the development of international gas transmission lines. URL=<https://pure.iiasa.ac.at/id/eprint/6192/> (accessed date: July, 15, 2024).