

KOBİ'ler için Makine Öğrenimi ile Karbon Ayak İzi Skoru Tahmini

Carbon Footprint Score Estimation with Machine Learning for SMEs

Halil Efe Çakmak

Muğla Sıtkı Koçman Üniversitesi Bilgisayar Mühendisliği Bölümü, h.efcakmak88@gmail.com

ÖZET

Bu makalede KOBİ'lerin iklim krizine olan negatif etkilerini azaltmaya yarayacak yeşil inovasyonlardan biri olmayı hedefleyen karbon ayak izi skoru tahmini yapan bir makine öğrenimi modeli üzerine durulmuştur. Yazarın çalışmaya katkı sağlayacak veri kümesine erişememesi sebebiyle modelde kullanılmak üzere sentetik veri seti üretilmiş, ön işleme adımlarından geçirilmiş ve görselleştirme işlemleri yapıp analizlerde bulunulmuştur. Ardından KNN, SVM, Naive Bayes, Decision Tree Classifier, Random Forest Classifier, XGBoost, LightGBM ve CatBoost gibi sınıflandırma algoritmaları veri üzerinde test edilip Accuracy, Precision, Recall ve F-Score ölçütleri ile de değerlendirilmiştir. Sonuç olarak karbon ayak izi skorlamada en iyi sonuç 0.9750 Accuracy ve 0.9749 F-Score değerleriyle SVM algoritmasında elde edilmiştir. Bu çalışma sayesinde KOBİ'ler, üretmekte oldukları karbon ayak izi skorlarını öğrenebilir ve çıkan sonuçlara göre yeşil inovasyonları benimsemeye yeni stratejiler geliştirebilirler.

Anahtar Kavramlar: KOBİ'ler, Yeşil İnovasyon, Karbon Ayak İzi, Makine Öğrenimi, Sınıflandırma Algoritmaları.

ABSTRACT

This article focuses on a machine learning model designed to estimate carbon footprint scores, which aims to be one of the green innovations that will help reduce the negative impacts of the climate crisis caused by SMEs. A synthetic dataset of SMEs was generated, preprocessed, visualized, and analyzed in this work because the author could not access the dataset that would contribute the study. Next, the dataset was tested with different classification algorithms like KNN, SVM, Naive Bayes, Decision Tree Classifier, Random Forest Classifier, XGBoost, LightGBM, CatBoost and evaluated using Accuracy, Precision, Recall, and F-Score. The SVM algorithm had the best results in carbon footprint scoring, with an accuracy of 0.9750 and an F-Score of 0.9749. Through this study, SMEs can learn about the carbon footprint scores they are producing and utilize the results to adopt new strategies for green innovations.

Keywords: SMEs, Green Innovation, Carbon Footprint, Machine Learning, Classification Algorithms.

1 GİRİŞ

Son dönemlerde yayınlanan çalışmalara göre, iklim değişikliği dünyada giderek daha da olumsuz etkilere sebep olmakta ve ciddi bir endişe kaynağı haline gelmektedir [1]. Ülkelerin ekonomik gelişmişlik seviyeleri arttıkça çevreye yönelik negatif etkileri de doğrusal bir şekilde artmaktadır [2]. Bu nedenle, daha yaşanılabilir bir gelecek için toplumların ve işletmelerin sürdürülebilir iş modellerini benimsemesi gerekmektedir [3]. Sürdürülebilirliğin bir ölçütü olan karbon ayak izinin tanımlanmasında yıllar içinde değişiklikler gözlenmiştir. Günümüzdeki tanımı ise, belirli sınırlar içerisinde bireylerin, organizasyonların, etkinliklerin, ürünlerin ve süreçlerinin tarafından üretilen ve atmosfere salınan farklı sera gazlarının CO₂-e (Karbon Dioksit Eşdeğeri) miktarı cinsinden ifade edilmesidir [4].

Küçük ve Orta Ölçekli İşletmeler (KOBİ'ler), küresel ekonominin büyük bir bölümünü oluşturdukları ve çevre üzerinde önemli bir etkiye sahip olduklarından dolayı yeşil inovasyonların teşvik edilmesi gereken büyük bir sektördür [5]. KOBİ tanımı ülkeden ülkeye göre değişiklik göstermektedir. Örneğin Avrupa Birliği'nde, 250'den az çalışanın çalıştığı ve cirosu 50 milyon Euro'yu geçmeyen firmalara KOBİ denirken, Amerika Birleşik Devletleri'nde 100 ile 1500 çalışanın çalıştığı ve 2.1 ila 21.5 milyon ABD Doları arasında cirosu olan şirketlere KOBİ denir [6]. KOBİ'ler, OECD (the Organization for Economic Co-operation and Development) ülkelerindeki işletmelerin %99'unu oluşturmakta ve işe alımların üçte ikisini gerçekleştirmektedir [7].

Doğal çevre üzerindeki olumsuz etkileri bakımından KOBİ'ler tekil olarak daha küçük etkiye sahip olsada toplu olarak değerlendirildiğinde seragazı üretiminin önemli bir payını oluşturmaktadır. Bu yüzden KOBİ'lerin yeşil inovasyonları benimsemesi, küresel iklim hedeflerine ulaşılmasında kritik bir rol oynamaktadır. [8]. Hatta KOBİ'ler boyutları, daha az hiyerarşik ve daha merkezi yapıları nedeniyle hızlı bir şekilde aksiyon alıp yeşil inovasyonları benimseyebilirler [9]. Şirketler tarafından atılacak sürdürülebilirlik adımları, yeşil markalaşmaya sebep olabilir hatta dahası, şirketlerin itibarını yükseltebilir ve müşterilerin güvenini artırabilir [10]. Son yapılan çalışmalar, müşterilerin sürdürülebilir ürünleri satın alma isteğindeki artışı göstermektedir. Bu artmakta olan talep de, yeşil inovasyonları benimsemiş KOBİ'lerin ekonomik olarak büyümesine ve kendi sektöründeki diğer şirketlerle rekabet etmesine olanak sağlar [11].

Avrupa Birliği, 2050 yılında iklim-nötr kıta olma hedefini gerçekleştirmek için 2019 yılında Avrupa Yeşil Mutabakatını açıklamıştır. Bu mutabakat, iklim krizini hafifletecek bir dizi düzenlemeler içermekte ve birçok sektörde köklü değişiklikler öngörmektedir. Örneğin Yeşil Mutabakat ile sıfır emisyon sürecinin hızlandığı Türkiye Cumhuriyeti'nde, Ticaret Bakanlığı ile Hazine ve Maliye Bakanlığının 2053 net sıfır emisyon hedefine ulaşılmasına katkı sağlamak adına şirketlerin yeşil dönüşüm süreci için finansman sağlaması, Yeşil Mutabakatın KOBİ'ler için de önemli bir fırsat olduğunu göstermektedir [12]. Ayrıca şirketler, hükümetlerin belirlediği yasal düzenlemeler sayesinde daha kolay bir şekilde yeşil inovasyonları benimseyebilir ve çevreye zarar veren şirketlere yönelik uygulanan yasal yaptırımlardan da kaçınabilirler [13].

Güneş enerjisi kullanımı, atıkların geri dönüştürülmesi, sürdürülebilir ürün ve hizmet sunumu, sertifikasyonlar aracılığıyla sürdürülebilirlik hedeflerinin belirlenmesi ve döngüsel ekonomik yaklaşımların benimsenmesi gibi adımlar, KOBİ'ler arasında en sık başvurulan sürdürülebilirlik stratejileridir [10]. Gupta ve Barua tarafından yapılan çalışmada, yeşil inovasyonların uygulanmasındaki engeller ele alınmış ve beş kategoriye ayrılmıştır: yönetimsel ve organizasyonel eksiklikler, girişimciler ve çalışanların yeşil inovasyon hakkındaki bilgi eksikliği, finansal ve ekonomik engeller, hükümet desteğinin yetersizliği ve politikaların zayıflığı, teknolojik yetersizlikler ve belirsizlikler [14]. Ancak, El-Gayar ve Fritz'in yaptığı araştırma bizlere bilimsel ve teknolojik gelişmelerin çevresel yönetim ve yeşil inovasyonlarını teşvik ettiğini ve olumlu bir rol oynadığını gösteren örnekler sunmaktadır [15]. Ayrıca, Wang, Li, Lin ve Wei'nin yayınladığı çalışmada, dijital teknolojilerinin gelişimi ile enerji verimliliğini artırma ve endüstriyel yapıyı modernize etme yollarıyla karbon emisyonlarını azaltarak, firma sürdürülebilirliğini artırdığı gösterilmiştir [16].

Bu makalede, KOBİ'lerin sürdürülebilirlik hedeflerini gerçekleştirmede karşılaştıkları teknolojik ve bilişimsel yetersizlikleri azaltacak makine öğrenimi çözümleri üzerine durulacaktır. Her bir KOBİ'nin karbon ayak izi skorunu, önceden sağlanmış veriler doğrultusunda tahmin edebilecek bir yapay zeka modeli geliştirilmeye çalışılacaktır. Bu modelin kullanımıyla KOBİ'ler; buldukları ofis veya fabrika gibi iş yerlerini sürdürülebilir hale getirip karbon ayak izi üretimlerini azaltabilirler, enerji tasarrufu sağlayabilirler ve kaynak yönetimlerini daha kolay yapabilirler, çeşitli kurumlar tarafından sağlanan yeşil sertifikasyonları kazanabilirler, sürdürülebilir ürün ve hizmet sunumunu sağlayabilirler, en önemlisi çevre dostu adımlar atıp iklim krizini hafifletebilirler.

2 LİTERATÜR ARAŞTIRMASI

Jordanis ve Evgenia'nın yaptıkları çalışmada, KOBİ'lerin ihtiyaçlarına uyarlanabilecek bir karbon ayak izi hesaplama uygulaması geliştirilmiştir. Şirketler bu uygulamada; üretim süreçlerini tasarlayabilir, üretilecek ürünün hammaddelerini seçebilir, üretim araçlarını belirleyebilir, tedarikçilerini ve depolama seçeneklerini yönetebilirler. Bunlar gibi birkaç veri uygulamaya sağlandığında, tek bir ürünün üretebileceği karbon emisyon miktarı bulunabilir. Böylelikle KOBİ'ler üretim süreçleri boyunca ne kadar karbon ayak izi ürettiklerini öğrenebilir ve sonuçlara göre üretim süreçlerini daha sürdürülebilir olacak şekilde güncelleyip geliştirebilirler [17].

Jerome ve araştırma ekibinin yapmış oldukları çalışmada, Fransa'da bulunan araştırma laboratuvarlarındaki üretilen karbon ayak izlerinin analizinin ve tahmininin yapılabilirdiği açık kaynaklı yazılım olan GES 1point5 aracı öne sürülmüştür. Bu yazılım aracı sayesinde Fransa'daki 470'den fazla araştırma laboratuvarındaki karbon ayak izi verileri bir veri tabanında toplanabilmiştir. GES 1point5 yazılımının çalışması için gereken birkaç veri girdisi şunlardır: laboratuvar üye sayısı, laboratuvarın elektrik ve gaz tüketimi, işletilen araç sayısı, kullanılan dijital cihaz sayısı, laboratuvar üyelerinin işe gidip gelme verileri ve mesleki seyahatler hakkındaki verilerdir. Toplanan bu veriler, Fransız emisyon faktörleri ile birleştirilir ve sonuç olarak bir araştırma laboratuvarının karbon ayak izi bulunulabilir [18].

Mateusz, Elzbieta ve Magdalena tarafından yapılan çalışmada, binaların karbon ayak izlerini hesaplayan bir makine öğrenimi modeli oluşturulmuştur. Bu model hesaplamalarını yapabilmesi için binanın genişliği, yüksekliği, kat sayısı, binanın yalıtımı, duvar ve çatı alanı, binanın her yönündeki pencere alanı ve çevresindeki kentsel düzenin gölge üzerindeki etkisi gibi verileri girdi olarak alıp binanın yaşam süresi boyunca ne kadar karbon ayak izi üretebileceğini tahmin etmektedir. Böylelikle inşaat mühendisleri ve mimarlar gibi inşaat sektöründe çalışan kişilerin, daha sürdürülebilir binalar yapabilmesi kolaylaştırılmıştır [19].

Milczarski'i tarafından yapılan çalışmada, iklim değişikliğinin gıda güvenliğine olan olumsuz etkisi üzerine durulmuş ve bu etkileri azaltmaya yönelik makine öğrenimi yöntemlerine odaklanılmıştır. Donmuş gıda üretimi yapan firmaların, üretim süreçlerindeki karbon ayak izlerinin optimizasyonu üzerine çalışılmıştır. Bu çalışmada KNN, Multilayer Perceptron, Decision Tree Classifier, Random Forest Classifier ve SVM sınıflandırma algoritmaları kullanılmıştır [20].

Jankovic ve arkadaşlarının yaptığı çalışmada, Nüfus ve farklı kaynaklara dayalı enerji tüketimini girdi verisi olarak alan ve KNNReg (K-Nearest Neighbors Regression), RFR (Random Forest Regression), ANN ReLU (Artificial Neural Network with Rectified Linear Unit) ve ANN SPOCU (Artificial Neural Network with Scaled Polynomial Constant Unit) gibi dört farklı makine öğrenimi modelini kullanan bir sürdürülebilirlik uygulaması üzerine çalışılmıştır. Modeller farklı makine öğrenimi ölçütleri kullanılarak karşılaştırılmıştır ve KNNReg modeli ile diğer modellerden daha iyi bir sonuca varılmıştır. Yapmış oldukları çalışma ile, ülkelerin ve birçok sektördeki şirketin ekolojik ayak izinin tahminini yaparak çevresel sürdürülebilirliğe destek olmayı hedeflemişlerdir [21].

LİU, Yang ve Zhang tarafından yapılan çalışmada, Endüstri 4.0'daki imalat sektörünün üretim süreçlerindeki karbon ayak izini azaltmaya yönelik makine öğrenimi modellerine odaklanılmıştır. Gerçek veri analizleri ile endüstriyel fırınların

karbon emisyonlarını hesaplayan makine öğrenimi teknikleri geliştirilmiştir. Şirketler bu teknikler sayesinde karbon ayak izlerini azaltabilir ve daha sürdürülebilir bir üretim sürecine sahip olabilirler [22].

Henderson ve arkadaşlarının yapmış oldukları çalışmada, makine öğrenimi sistemlerinin enerji tüketimlerini ve karbon ayak izlerinin hesaplanmasını kolaylaştıracak ve daha doğru raporlanmasını sağlayacak Experiment-Impact-Tracker adlı uygulama üzerinde çalışılmıştır. Bu sürdürülebilirlik uygulaması hesaplamaları yapabilmek için; kullanılan yazılımın türünü, CPU ve GPU donanım bilgilerini, enerji şebekesi bölgesindeki ortalama karbon yoğunluğunu ve bellek kullanımı gibi verileri girdi olarak almaktadır. Uygulamanın kullanımı ile şirketler enerji verimliliklerini artırabilir ve karbon ayak izi miktarlarını azaltabilirler [23].

Literatür taraması yapılırken KOBİ'lerin tüketim verilerine dayanan bir karbon ayak izi skorlama çalışması bulunamamıştır. Bu makalenin amacı KOBİ'ler tarafından kullanılacak bir karbon ayak izi skoru tahmini yapabilen makine öğrenimi modeli geliştirerek literatüre katkı sağlamaktır.

3 VERİ KÜMESİ VE ANALİZİ

Makalenin bu bölümünde, çalışmada kullanılacak veri kümesi elde edilmeye çalışılmıştır ve veri ön işleme adımlarının uygulanması açıklanmıştır.

3.1 Sentetik Verinin Üretimi

KOBİ'ler için yapılan araştırmaların yetersizliğinden dolayı işletmelerin doğrudan ürettikleri emisyonlarına yönelik veri kıtlığı yaşanmaktadır. Bu soruna çözüm üretmek için Statistics Denmark tarafından pilot çalışma yapılmıştır. Çalışmada Danimarka'daki KOBİ'lerin elektrik ve gaz tüketimi araştırılıp verileri toplanmıştır. Gerçek elektrik ve gaz tüketim verileri, önceden üretilmiş ulusal tahminlerle karşılaştırılmış ve sonuç olarak tahminlerin gerçek verilere çok yakın olduğu görülmüştür. Statistics Denmark tarafından yapılan çalışmanın, sadece sınırlı emisyon verilerini kapsadığı ve Danimarka gibi enerji tüketiminin %50'sinin enerji yoğun KOBİ'lerin %1'inden azına ait bir ülkenin verilerini içerdiği için, bu makalede pilot çalışma tarafından sağlanan verilerin yetersiz ve kullanışsız olduğuna karar verilmiştir [8].

Yazarın, KOBİ'lerin aylık tüketmekte oldukları elektrik, su ve doğalgaz miktarını içeren veri kümesine ayrıca üretmekte oldukları atık miktarını içeren veri kümesine erişememesi neticesinde, makalede sentetik veri üretimi ve kullanımı gerçekleştirilmiştir. Sentetik verinin üretimini sağlamak için OpenAI tarafından geliştirilen ChatGPT-4o yapay zeka modeline başvurulmuştur [24]. Verinin doğru ve eksiksiz elde edilmesi için girilen prompt şu şekildedir:

“Bir araştırmacı gibi davranmanı istiyorum. Senden şirketlerin; aylık elektrik kullanımını MWh (megawatt hour) cinsinden, aylık su kullanımını m³ (cubic meter) cinsinden, aylık doğal gaz kullanımını scf (standard cubic foot) cinsinden ve aylık atık miktarını ise ton cinsinden bulduran, 1000 satırlık bir sentetik veri tablosunun senin tarafından oluşturulmasını istiyorum.”

Tablo 1: ChatGPT'den Elde Edilen Veri Seti.

Şirket_No	Elektrik Tük. MWh	Su Tük. m ³	Doğalgaz Tük. scf	Atık Miktarı. ton
şirket_1	478,2703976	393,2109624	36165,91851	6,832962925
şirket_2	106,5729528	969,5565844	26819,67751	40,03513023
şirket_3	272,4394958	1375,291472	41969,62017	7,48019621
şirket_4	308,1040805	1273,019645	31187,10223	3,648856013

3.2 Verinin İşlenmesi

Bu bölümde kullanılmak üzere emisyon faktörleri, EPA (U.S. Environmental Protection Agency) ve The Environment Agency of Japan araştırmaları tarafından sağlanmıştır [25-26]. Tablo 2'de her bir değişkenin araştırmalar tarafından belirlenmiş emisyon faktörleri gösterilmiştir.

Tablo 2: Emisyon Faktörleri.

Elektrik Tüketimi Emisyon Faktörü	373,45 kg CO ₂ / MWh
Su Tüketimi Emisyon Faktörü	0,59 kg CO ₂ / m ³
Doğalgaz Tüketimi Emisyon Faktörü	0,05444 kg CO ₂ / scf
Atık Üretimi Emisyon Faktörü	376,00 kg CO ₂ / ton

Ardından yazar tarafından sentetik veri setindeki her bir değişken, araştırmalar sonucunda bulunan Tablo 2'deki emisyon faktörleri ile çarpılmış ve sonuç olarak değişkenlerin tekil olarak ürettikleri karbon ayak izi miktarları bulunmuştur. Sonrasında bütün değişkenler toplanarak her bir şirketin aylık ürettikleri karbon ayak izi miktarı hesaplanmış ve Tablo 3'te gösterilmiştir.

Tablo 3: Şirketlerin Toplam Karbon Ayak İzi.

Şirket_No	Elektrik Karbon Su Karbon Ayak Doğalgaz Karbon Atık Karbon Toplam Karbon Ayak İzi Eşdeğeri	Karbon İzi Eşdeğeri	Karbon Ayak izi Eşdeğeri	Karbon Ayak İzi Eşdeğeri	Karbon Ayak İzi Eşdeğeri
şirket_1	178610,08	231,9944678	1968,872604	2569,19406	183380,1411
şirket_2	39799,66924	572,0383848	1460,063244	15053,20897	56884,97983
şirket_3	101742,5297	811,4219684	2284,826122	2812,553892	107651,3317
şirket_4	115061,4689	751,0815908	1697,825845	1371,969861	118882,3462

3.3 Verinin Skorlanması

Veri analizinin bu bölümünde ise şirketler ve aylık ürettikleri toplam karbon ayak izi miktarları, bir K-means kümeleme algoritması kullanılarak altı farklı kümeye bölünmüştür. Ardından bu altı ayrı kümenin ortalama karbon ayak izi miktarı hesaplanmıştır ve sonucunda her bir şirket, A'dan C-'ye olarak belirlenmiş skor aralığında, toplam ürettikleri karbon ayak izi miktarının arttığı biçimde A, A-, B, B-, C ve C- olarak skorlanmıştır.

Tablo 4: Şirketlerin Elde Ettikleri Skorlar.

Şirket_No	Toplam Karbon Ayak İzi Miktarı	Ait olunan Kümenin Rakamı	Elde edilen Skor
şirket_1	183380,1411	3	C-
şirket_2	56884,97983	1	A-
şirket_3	107651,3317	4	B
şirket_4	118882,3462	0	B-

3.4 Sütun Dönüşümü (Encoding)

Veri setindeki ‘‘Karbon Ayak İzi Skoru’’ sütunu, kategorik bir veri olduğu ve sıralı bir yapıya sahip olduğu için label encoding yöntemi uygulanmıştır ve tablo 5’te gösterilmiştir. Böylelikle makine öğrenimi modellerinin, karbon ayak izi skorları arasındaki toplam karbon ayak izi miktarının artan ilişkisini anlaması kolaylaştırılmıştır.

Tablo 5: Label Encoding İşlemi.

Karbon Ayak İzi Skoru	Encoded Karbon Ayak İzi Skoru
A	0
A-	1
B	2
B-	3
C	4
C-	5

3.5 Kullanılacak Verinin Hazırlanması

Bu bölümde ise makalenin ilerleyen bölümlerindeki makine öğrenimi modellerini eğitmek ve test etmek için kullanılacak veri seti hazırlanmıştır. Sentetik veri setinden elde edilen şirketlerin aylık elektrik, su ve doğalgaz tüketimleri, aylık üretilen atık miktarı ve önceki veri adımlarından elde edilen şirketlerin karbon ayak izi rakam skorları birleştirilerek bir veri seti haline getirilmiştir ve tablo 6’da ilk dört satırı gösterilmiştir.

Tablo 6: Modellerde Kullanılacak Veri Seti.

Şirket_No	Elektrik Tük. MWh	Su Tük. m ³	Doğalgaz Tük. scf	Atık Miktarı. ton	Karbon Ayak İzi Rakam Skoru
şirket_1	478,2703976	393,2109624	36165,91851	6,832962925	5
şirket_2	106,5729528	969,5565844	26819,67751	40,03513023	1
şirket_3	272,4394958	1375,291472	41969,62017	7,48019621	2
şirket_4	308,1040805	1273,019645	31187,10223	3,648856013	3

3.6 Verinin Açıklanması ve Özellikleri

Veri ön işleme adımlarının sonucunda elde edilip tablo 6’da gösterilen veri setinin her bir değişkeninin açıklaması ve veri tipleri tablo 7’de gösterilmiştir.

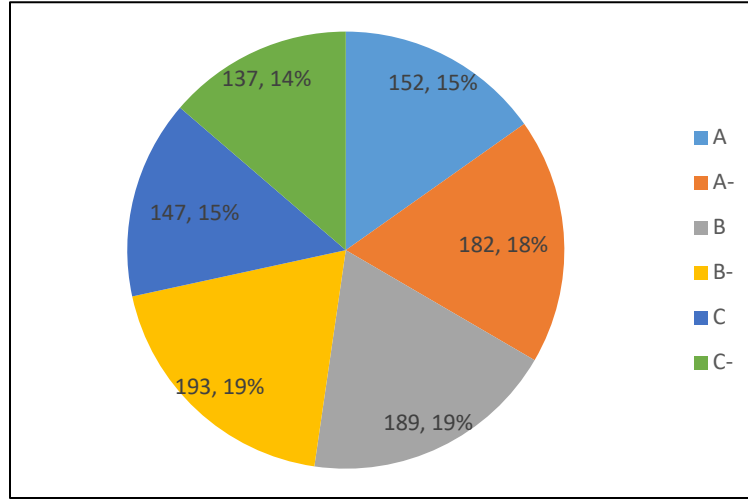
Tablo 7: Veri Setindeki Değişkenlerin Öznitelikleri.

Değişken	Açıklama	Veri Tipi
Şirket_No	Şirketin kimliği	object
Elektrik Tük. MWh	Aylık tüketilen elektrik miktarı	Float64
Su Tük. m ³	Aylık tüketilen su miktarı	Float64
Doğalgaz Tük. scf	Aylık tüketilen doğalgaz miktarı	Float64
Atık Miktarı. ton	Aylık üretilen atık miktarı	Float64
Karbon Ayak İzi Rakam Skoru	Şirketin karbon ayak izinin rakam skoru	İnt64

3.7 Veri Analizi ve Görselleştirme

Makalenin bu adımında verilerin grafiklerle görselleştirilmesi, aralarındaki ilişkilerin incelenmesi ve anlamlı sonuçlara varılması amaçlanmıştır. Böylelikle makalenin ilerleyen adımlarındaki makine öğrenimi modelleri için daha stratejik kararlar alınması sağlanmıştır.

Grafik 1'de her bir skor başına düşen şirket sayısı dilim grafiği şeklinde ve yüzdelik değerleriyle birlikte verilmiştir. Karbon ayak izi skoru, kategorik bir veri türü olduğu ve okurlar tarafından daha iyi anlaşılabilmesi için dilim grafik türü seçilmiştir. Skor grafiğindeki B ve B- skorları, %38'lik oranlarıyla en büyük paya sahip olurken; C ve C- skorları %29'luk oranlarıyla en küçük paya sahip olmuştur. Grafikteki bu dağılımın nispeten dengeli olduğu sonucuna varılmıştır.

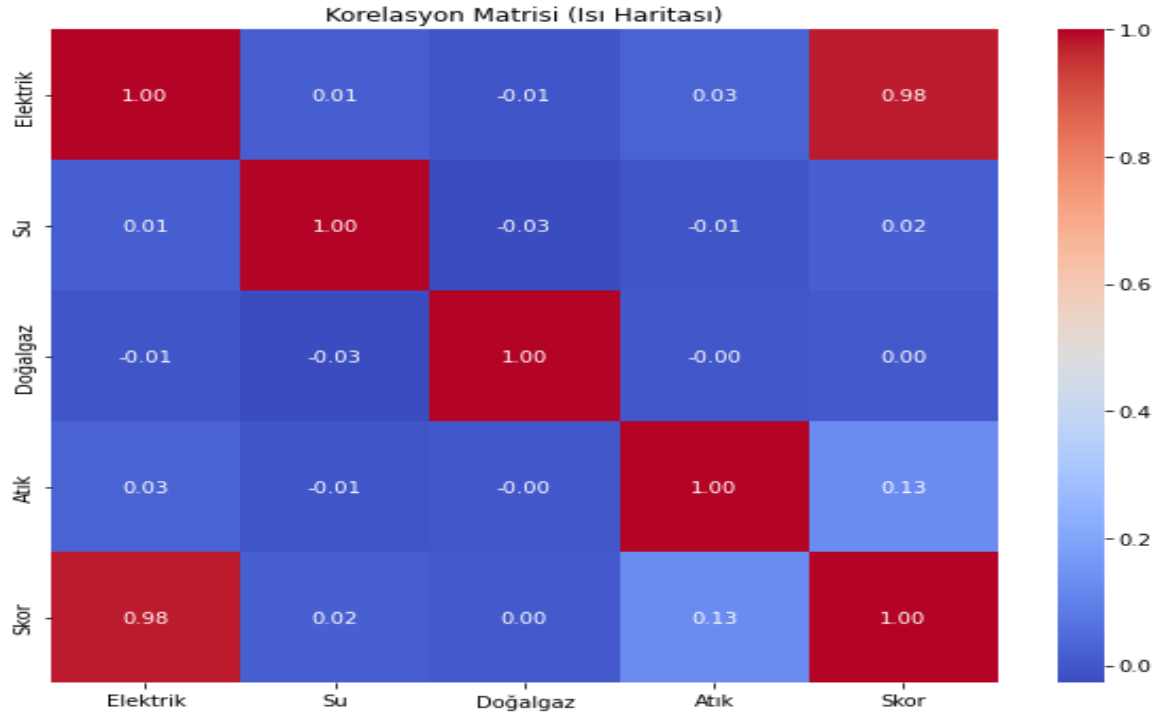


Grafik 1: Her Bir Skor Başına Düşen Şirket Sayısı Dilim Grafiği

Tablo 8'de her bir skordaki elektrik, su, doğalgaz tüketimlerinin ve atık üretimlerinin ortalama miktarları gösterilmiştir. Tabloda kırmızı tonları daha düşük tüketimi ve üretimi gösterirken, mavi tonları ise daha yüksek tüketim ve üretim ortalamalarını göstermektedir. Elektrik tüketiminde A skorundan C- skoruna doğru belirgin bir artış gözlenmiştir. Su ve doğalgaz tüketiminde ise nispeten dengeli bir dağılım gözlenmiştir. En çok su tüketimi C skorunda gözlenirken, en çok doğal gaz tüketimi ise B skorunda gözlenmiştir. Atık üretiminde de A skorundan C- skoruna doğru belirgin bir artış gözlenmiştir.

Tablo 8: Her Bir Skordaki Değişkenlerin Ortalama Değerleri.

	A	A-	B	B-	C	C-
Elektrik Tük. Ort. MWh	87,687	156,188	230,783	315,268	398,26	466,432
Su Tük. Ort. m ³	1091,283	1042,46	1109,571	1094,319	1117,517	1081,217
Doğalgaz Tük. Ort. Scf	27412,15	27745,39	28716,37	28178,83	27577,57	27845,79
Atık Miktarı Ort. ton	21,54682	25,0114	25,91254	26,43968	26,54835	28,19058



Grafik 2: Veri Setindeki Değişkenler Arasındaki İlişkiyi Temsil Eden Korelasyon Isı Haritası

Grafik 2'deki korelasyon matrisi, veri setindeki değişkenler arasındaki ilişkileri anlamlandırarak incelemek için seçilmiş ve Spyder IDE yazılımı üzerinden hazırlanmıştır [27]. Hücreler içerisinde bulunan katsayılar, her iki değişken arasındaki ilişkiyi negatif, nötr ya da pozitif olarak göstermektedir. Makaledeki veri seti kullanılarak elde edilen korelasyon matrisinden anlaşılacağı üzere elektrik tüketimi ile elde edilen skor rakamı arasında pozitif güçlü bir ilişki, elektrik tüketimi ile su tüketimi ve atık üretimi arasında zayıf pozitif ilişki ve doğalgaz tüketimi arasında da zayıf negatif bir ilişki, su tüketimi ile skor rakamı arasında zayıf pozitif ilişki, su tüketimi ile doğalgaz tüketimi ve atık üretimi arasında zayıf negatif ilişki, doğalgaz ile atık üretimi ve skor rakamı arasında nötr bir ilişki ve atık üretimi ile skor rakamı arasında zayıf pozitif bir ilişki olduğu gözlenmiştir.

4 MODELLEME

Makalenin bu adımında, şirketlerin karbon ayak izi skorunu bulmaya yönelik tahmin modelleri ve kullanılan modellerin değerlendirilmeleri üzerine çalışılmıştır. Veri kümesi hazırlama ve analizi adımlarından sonra elde edilen veri kümesi, makalenin bu aşamasında kullanılacak makine öğrenimi modellerinin eğitim ve tahmin işlemleri için kullanılmıştır. Makine öğrenimi sınıflandırma algoritmalarından KNN (K-Nearest Neighbors), SVM (Support Vector Machine), Naive Bayes, Decision Tree Classifier, Random Forest Classifier, XGBoost (Extreme Gradient Boosting), LightGBM (Light Gradient Boosting Machine) ve CatBoost (Categorical Boosting) algoritmaları kullanılmıştır. Modellerin değerlendirilmesi için ise Accuracy, Precision, Recall ve F-Score ölçütleri kullanılmıştır.

4.1 Algoritmaların Açıklanması

Bu bölümde karbon ayak izi skorlama çalışmasında kullanılacak makine öğrenimi algoritmaları açıklanmıştır.

4.1.1 KNN

Makine öğrenimi alanında sıkça kullanılan ve diğer algoritmalarından görece daha basit olan KNN sınıflandırma algoritması; etiketsiz verileri, en yakın etiketli verilerle karşılaştırarak sınıflandırır. Bu algoritmada veriler arasındaki mesafe ölçüm yöntemi ve k değeri sonuçların doğruluğu açısından önemli bir rol oynamaktadır. K değeri bir veriyi sınıflandırırken bakılacak en yakın komşu sayısını belirtmektedir hatta aşırı öğrenme ya da zayıf öğrenme arasında denge kurmayı sağlamaktadır. Makalede KNN modelinin kullanımına karar verilmiş ve mesafe ölçüm yöntemi olarak öklidyen mesafesi kullanılıp, k değeri 3 olarak alınmıştır [28].

4.1.2 SVM

Denetimli öğrenme yöntemlerinden biri olan SVM, özellikle sınıflandırma problemlerinde öne çıkan güçlü bir makine öğrenimi algoritmasıdır. SVM performansını önemli ölçüde etkileyen çekirdek fonksiyonları, kullanılan veri kümesi ile ilişkilidir ve genel olarak lineer, polinom ve sigmoid fonksiyonları tercih edilir. Bu çalışmadaki veri setinde sınıflandırma üzerine çalışıldığından bu algoritmanın kullanılması ve lineer çekirdek fonksiyonunu içermesi uygun görülmüştür [29].

4.1.3 Naive Bayes

Thomas Bayes'in geliştirdiği Bayes teoremine dayanan Naive Bayes algoritmasının temel amacı, verinin en iyi şekilde sınıflandırılmasıdır. Olasılıksal sınıflandırma algoritmalarının arasında yer alan Naive Bayes, algoritmasının basitliği ve doğruluğu nedeniyle sınıflandırmada önemli bir yeri vardır. Algoritmanın sınıflandırma problemlerindeki başarısı sebebiyle makalede kullanılmasına karar verilmiştir [30].

4.1.4 Decision Tree Classifier

Karar ağaçları, her bir verinin farklı birçok karar fonksiyonlarıyla bir sınıfa atandığı sınıflandırma algoritmasıdır. Bu algoritma ağaç diyagramına benzemektedir hatta genel olarak bir kök düğüm, birçok iç düğüm ve sınıf sayısı kadar son düğüme sahiptir. Karar ağaçlarının sınıflandırmadaki başarısı ve geniş uygulama alanlarının olmasından dolayı makalede kullanılması uygun görülmüştür [31].

4.1.5 Random Forest Classifier

Birden fazla karar ağacının birleşimiyle oluşan rastgele orman algoritması, hem sınıflandırma hem de regresyon problemlerinde kullanılabilir. Çoklu ağaç yapısına sahip olması, hata oranını ve aşırı öğrenmeyi azaltmada önemli bir rol oynamaktadır. Rastgele ormanlar güçlü bir sınıflandırma tekniği olduğu için makalede kullanılmasına karar verilmiştir [32].

4.1.6 XGBoost

Gradient Boosting, zayıf sınıflandırma modellerinin hatalarını analiz ederek ve adım adım düzelterek modelleri iyileştiren güçlü bir makine öğrenimi algoritmasıdır. XGBoost ise Gradient Boosting algoritmasının gelişmiş bir versiyonudur. Hem çoğu zaman sonuç elde etmede daha hızlı performans gösterdiği hem de genel olarak daha isabetli sonuçlar verdiği için makalenin modelleme aşamasında kullanılmasına uygun görülmüştür [33].

4.1.7 LightGBM

LightGBM, Gradient Boosting sınıflandırma algoritmasının daha gelişmiş bir türevidir. Gelişimine katkıda bulunan iki yenilikçi teknik bulunmaktadır. Bunlar GOSS (Gradient-based One-Side Sampling) ve EFB (Exclusive Feature Bundling)

teknikleridir. GOSS tekniği ile gereksiz verilerle zaman kaybedilmeyen, EFB tekniği ile ise hesaplama yükü azaltılarak verimi artırılan LightGBM sınıflandırma algoritmasının makalede kullanılmasına karar verilmiştir [34].

4.1.8 CatBoost

Kategorik özelliklere sahip veri kümelerinde önemli avantajlar sunan güçlü bir makine öğrenimi tekniğidir. Gradyan güçlendirme algoritması ile etkin bir şekilde kategorik verileri işleyebilir. Diğer gradyan güçlendirme yöntemlerinin aksine hem kategorik verileri sayılara dönüştürmeden işleyebilir hem de daha hızlı bir şekilde eğitim ve tahminde bulunabilir. Makalede bulunan veri setindeki karbon ayak izi skoru gibi bir kategorik veri bulunduğu için CatBoost tekniğinin çalışmada kullanılması uygun görülmüştür [35].

4.2 Algoritmaların değerlendirilmesi

Makine öğrenimi projelerinin önemli bir aşaması olan modellerin değerlendirilmesi, ölçütlerinin belirlenmesi ve açıklanması, makalenin bu adımında yapılmıştır. Her bir sınıflandırma algoritmasının değerlendirilmesi için Accuracy, Precision, Recall ve F-Score ölçütleri kullanılmıştır. Bu ölçütlerin kolay açıklanabilmesi için binary confusion matrix kullanılmış ve tablo 9'da gösterilmiştir [36].

Tablo 9: Binary Confusion Matrix

		Actual Data	
		True (1)	False (0)
Predicted Result	Positive (1)	TP	FP
	Negative (0)	FN	TN

4.2.1 Accuracy

Accuracy doğru tahminlerin toplamının, tüm tahminlerin toplamına bölünmesi ile hesaplanmaktadır.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (1)$$

4.2.2 Precision

Precision ise doğru pozitif tahmin sayısının, toplam pozitif tahmin sayısına bölünmesiyle hesaplanmaktadır.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

4.2.3 Recall

Recall hesaplanış şekli ise doğru pozitif tahmin sayısının, doğru pozitif ve yanlış negatif tahminlerin toplamına bölümüdür.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

4.2.4 F-Score

F-Score testin doğruluğunu ölçen bir ölçüttür ve Precision ile Recall ölçütlerini içermektedir.

$$F - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

4.2.5 Algoritmaların Değerlendirilme Sonuçları

Makalenin bu bölümünde şirketlerin karbon ayak izi skorunun hesaplanabilmesi için kullanılan sınıflandırma algoritmalarının her birinin, Spyder IDE yazılımı üzerinden Accuracy, Precision, Recall ve F-score değerlendirilme ölçütleri hesaplanmıştır ve tablo 10'da gösterilmiştir [27].

Tablo 10: Sınıflandırma Algoritmalarının Değerlendirilme Sonuçları

	Accuracy	Precision	Recall	F-Score
KNN	0.7600	0.7700	0.7600	0.7600
SVM	0.9750	0.9770	0.9750	0.9749
Naive Bayes	0.8800	0.8800	0.8800	0.8700
Decision Tree Classifier	0.9200	0.9200	0.9200	0.9100
Random Forest Classifier	0.9400	0.9440	0.9400	0.9395
XGBoost	0.9400	0.9414	0.9400	0.9396
LightGBM	0.9400	0.9500	0.9400	0.9400
CatBoost	0.9400	0.9500	0.9400	0.9400

5 SONUÇ

Bu çalışma, küresel ekonominin büyük bir kısmını oluşturan ve çevreye etkisi azımsanamayacak kadar büyük olan KOBİ'lerin iklim krizine olan etkilerini hafifletecek teknolojik çözümler aramak için yazılmıştır. Bu amaç doğrultusunda şirketlerin karbon ayak izi skorunu tahmin edebilen bir makine öğrenimi modeli oluşturulmuştur.

Makalenin giriş kısmında, KOBİ'nin genel tanımlaması yapılmış ve yeşil inovasyonların benimsenmesinde karşılaşılan engellerden bahsedilmiştir. Sonraki literatür araştırması kısmında, çeşitli alanlarda karbon ayak izi araştırması yapan makalelere yer verilmiştir. Veri kümesi bölümünde ise yazarın çalışmaya katkı sağlayacak veri kümesine erişememesi neticesinde modellerde kullanılmak üzere sentetik veri üretimi yapılmıştır. Üretilen sentetik veri kümesi üzerinde çalışılmış ve analizlerde bulunulmuştur. Ardından KNN, SVM, Naive Bayes, Decision Tree Classifier, Random Forest Classifier, XGBoost, LightGBM ve CatBoost gibi sınıflandırma algoritmaları veri üzerinde test edilmiştir. Sonuç olarak karbon ayak izi skorlamada en iyi sonuç 0.9750 Accuracy değeriyle SVM algoritması ile elde edilirken en kötü sonuç 0.7600 Accuracy değeriyle KNN algoritması ile elde edilmiştir.

Böylelikle makalede üzerine çalışılan makine öğrenimi modeli, KOBİ'lerin sürdürülebilirlik hedeflerine ulaşmasında yardımcı olabilecek bir araç haline gelebilmiştir. Bu modelin kullanımı sayesinde KOBİ'ler; çevre dostu stratejiler geliştirebilir, çalışma ortamlarını sürdürülebilir hale getirebilir, kaynak yönetimlerini düzenleyebilir, ürettikleri karbon emisyonlarını azaltabilir, yeşil sertifikasyonları kazanmada önemli adımlar atabilir, şirketlerinin yeşil markalaşmasına sebep olup itibarını arttırabilir ve en önemlisi yeşil inovasyonlarla iklim krizini hafifletebilirler.

Tüm bu çalışmaların neticesinde geliştirilen modelin doğruluğunu arttırmak ve kullanım alanını genişletmek için gelecekteki çalışmalarda; KOBİ'lere ait gerçek dünya verilerinin toplanmasına yönelik araştırmalar arttırılmalı, veri kümesindeki değişkenlerin emisyon faktörleri KOBİ'lerin buldukları bölgeye göre düzeltilmeli, her bir KOBİ'ye yönelik modelin girdi değişkenleri optimize edilmeli ve OptiWisdom firması tarafından geliştirilen Kolay.ai gibi KOBİ'lere yönelik teknolojik çözümler sunan yazılımlarla entegre edilmelidir [37].

6 REFERANSLAR

- [1] Thomas Dietz, Rachael L. Shwom, Cameron T. Whitley. 2020. Climate Change and Society. From <https://doi.org/10.1146/annurev-soc-121919-054614>.
- [2] Xiao Luo, Weiye Zhang. 2020. Green Innovation Efficiency: a Threshold Effect Of Research and Development. From <http://doi.org/10.1007/s10098-020-01977-x>.
- [3] Jacob Guinot. 2020. Changing the Economic Paradigm: Towards a Sustainable Business Model. From <https://doi.org/10.18280/ijssdp.150502>. Spain.
- [4] Divya Pandey, Madhoolika Agrawal, Jai Shanker Pandey. 2010. Carbon footprint: current methods of estimation. From <https://doi.org/10.1007/s10661-010-1678-y>.
- [5] Margarida Rodrigues, Mario Franco. 2023. Green Innovation in Small and Medium-Sized Enterprises (SMEs): a Qualitative Approach. From <http://doi.org/10.3390/su15054510>. Portugal.
- [6] Claudia P. Ribau, Antonio Carrizo Moreira, Mario Raposo. 2016. SME Internationalization Research: Mapping the State of the Art. From <https://doi.org/10.1002/cjas.1419>.
- [7] OECD (2018), Environmental Policy Toolkit for SME Greening in EU Eastern Partnership Countries, OECD Green Growth Studies, OECD Publishing, Paris. <http://dx.doi.org/10.1787/9789264293199-en>.
- [8] OECD (2023), Assessing greenhouse gas emissions and energy consumption in SMEs. From <https://doi.org/10.1787/ac8e6450-en>.
- [9] Avi Fiegenbaum, Aneel Karnani. 1991. Output Flexibility-A Competitive Advantage for Small Firms. From <https://www.jstor.org/stable/2486341>. Michigan, U.S.A.
- [10] Lenka Wildnerova, Carlo Menon, Robert Dehghan, Jan Kinne, David Lenz. 2024. Which SMEs are greening? Cross-country evidence from one million websites. From <https://doi.org/10.1787/dd00999-en>.
- [11] Anil Kumar, Gyan Prakash, Gaurav Kumar. 2021. Does environmentally responsible purchase intention matter for consumers? A predictive sustainable model developed through an empirical study. From <https://doi.org/10.1016/j.jretconser.2020.102270>.
- [12] Türkiye Cumhuriyeti Ticaret Bakanlığı. Yeşil Mutabakat. From <https://ticaret.gov.tr/dis-iliskiler/yesil-mutabakat>.
- [13] Klaus Rennings, Andreas Ziegler, Kathrin Ankele, Esther Hoffmann. 2005. The influence of different characteristics of the EU environmental management and auditing scheme on technical environmental innovations and economic performance. From <https://doi.org/10.1016/j.ecolecon.2005.03.013>. Berlin, Germany.
- [14] Himanshu Gupta, Mukesh Kumar Barua. 2018. A framework to overcome barriers to green innovation in SMEs using BWM and Fuzzy TOPSIS. From <https://doi.org/10.1016/j.scitotenv.2018.03.173>.
- [15] Omar F. El-Gayar, Brian D. Fritz. 2006. Environmental management information systems (EMIS) for sustainable development: A conceptual overview. From <https://aisel.aisnet.org/cais/vol17/iss1/34>.
- [16] Hui Wang, Yueyue Li, Weifen Lin, Wendong Wei. 2023. How does digital technology promote carbon emission reduction? Empirical evidence based on e-commerce pilot city policy in China. From <https://doi.org/10.1016/j.jenman.2022.116524>. China.
- [17] Iordanis Eleftheriadis, Evgenia Anagnostopoulou. 2024. Developing a Tool for Calculating the Carbon Footprint in SMEs. From <https://doi.org/10.3390/su16051905>.
- [18] Jerome Mariette, Odile Blanchard, Olivier Berne, Olivier Aumont, Julian Carrey, Anne-Laure Ligozat, Emmanuel Lellouch, Philippe-Emmanuel Roche, Gael Guennebaud, Joel Thanwerdas, Philippe Bardou, Gerald Salin, Elise Maigne, Sophie Servan, Tamara Ben-Ari. 2022. An open-source tool to assess the carbon footprint of research. From <https://iopscience.iop.org/article/10.1088/2634-4505/ac84a4/meta>.
- [19] Mateusz Ploszaj-Mazurek, Elzbieta Rynska, Magdalena Grochulska-Salak. 2020. Methods to Optimize Carbon Footprint of Buildings in Regenerative Architectural Design with the Use of Machine Learning, Convolutional Neural Network, and Parametric Design. From <https://doi.org/10.3390/en13205289>.
- [20] Piotr Milczarski. 2021. Application of Machine Learning Methods in Carbon Footprint Optimization. From <https://doi.org/10.1109/IDAACS53288.2021.9660994>. Poland
- [21] Radmila Jankovic, Ivan Mihajlovic, Nada Strbac, Alessia Amelio. 2021. Machine learning models for ecological footprint prediction based on energy parameters. From <https://link.springer.com/article/10.1007/s00521-020-05476-4>.
- [22] Zhao Liu, Gangying Yang, Yi Zhang. 2023. Carbon footprint assessment in manufacturing Industry 4.0 using machine learning with intelligent Internet of things. From <https://doi.org/10.1007/s00170-023-12183-6>.
- [23] Peter Henderson, Jieru Hu, Joshua Romoff, Emma Brunskill, Dan Jurafsky, Joelle Pineau. 2020. Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. From <https://www.jmlr.org/papers/v21/20-312.html>.
- [24] OpenAI. From <https://openai.com/>.
- [25] United States Environmental Protection Agency, 2024 GHG Emission Factors Hub. From <https://www.epa.gov/climateleadership/ghg-emission-factors-hub>.
- [26] Yasutoshi Shimizu, Satoshi Dejima, Kanako Toyosada. 2012. The CO2 Emission Factor of Water in Japan. From <https://doi.org/10.3390/w4040759>.
- [27] Spyder IDE. The Scientific Python Development Environment. From <https://www.spyder-ide.org/>.
- [28] Zhongheng Zhang. 2016. Introduction to machine learning: k-nearest neighbors. From <https://doi.org/10.21037%2Fatm.2016.03.37>.
- [29] DURGESH K. SRIVASTAVA, LEKHA BHAMBHU. 2010. DATA CLASSIFICATION USING SUPPORT VECTOR MACHINE. From <https://www.researchgate.net/publication/285663733>.

- [30] Indika Wickramasinghe, Harsha Kalutarage. 2020. Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. From <https://doi.org/10.1007/s00500-020-05297-6>.
- [31] Philip H. Swain, Hans Hauska. 1977. The decision tree classifier: Design and potential. From <https://doi.org/10.1109/TGE.1977.6498972>.
- [32] LEO BREIMAN. 2001. Random Forests. From <https://doi.org/10.1023/A:1010933404324>.
- [33] Ramraj Santhanam, Sunil Raman, Nishant Uzir and Shatadeep Banerjee. 2017. Experimenting XGBoost Algorithm for Prediction and Classification of Different Datasets. From <https://www.researchgate.net/publication/318132203>.
- [34] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye and Tie-Yan Liu. 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. From <https://dl.acm.org/doi/10.5555/3294996.3295074>.
- [35] Anna Veronika Dorogush, Vasily Ershov, Andrey Gulin. 2018. CatBoost: gradient boosting with categorical features support. From https://ui.adsabs.harvard.edu/link_gateway/2018arXiv181011363V/doi:10.48550/arXiv.1810.11363.
- [36] Zeljko Vujovic. 2021. Classification Model Evaluation Metrics. From <http://dx.doi.org/10.14569/IJACSA.2021.0120670>.
- [37] Şadi Evren Şeker. 2023. KOBİ'lere Özel Basit Yapay Zeka Çözümü : Kolay.AI. From <https://ybsansiklopedi.com/cilt-11-sayi-1-ocak-2023/>.