

Veri Madenciliği Algoritmaları Kullanarak Türkiye'deki Elektrik Tüketicileri İçin En Uygun Tarife Seçim Önerisi

Best Tariff Proposal for Electricity Consumers in Turkey Using Data Mining Algorithms

¹Seda Balta and ^{*2}Cüneyt Bayılmış

¹Teknoloji Fakültesi, Bilişim Sistemleri Mühendisliği Kocaeli Üniversitesi, Türkiye

^{*2}Bilgisayar ve Bilişim Bilimleri Fakültesi, Bilgisayar Mühendisliği Sakarya Üniversitesi, Türkiye

Özet

Teknolojinin gelişmesiyle birlikte elektriğe bağımlılık gün geçtikçe artmaktadır. Elektrik tüketimi arttıkça maliyette artmaktadır. Bu nedenle bireyler, optimum fayda-maliyet ilişkisini arayarak tasarruf etmeyi amaçlamaktadırlar. Bu çalışmada ilk olarak 1500 satırlık, gündüz, puant ve gece değerlerine sahip olan veri seti oluşturuldu ve bu veri seti kullanılarak en uygun algoritmayı seçmek için Roc analizi yapıldı. Sonra meskenlerdeki toplam elektrik tüketimi, ROC analizi sonucu seçilen lojistik regresyon algoritması ile RapidMiner yazılımında analiz edildi. Abonelere en uygun tarife seçimi yapıldı. Sonuç olarak,% 100 doğru tahmin elde edildi.

Anahtar kelimeler: Elektrik tüketimi, veri madenciliği, lojistik regresyon, rapidMiner, sınıflandırma algoritmaları

Abstract

With the development of technology, dependence on electricity is increasing day by day. Cost increases as electricity consumption increases. For this reason, individuals aim to save by searching for the optimum benefit-cost relationship. In this study, firstly 1500 rows of data created with day, peak, night values and Roc analysis was used to select the most appropriate algorithm for the data set. Later total electricity consumption in households was analyzed by selecting logistic regression which is data mining algorithm in RapidMiner. The most appropriate tariff selection was provided to the subscribers. As a result, 100% accurate estimation was obtained.

Key words: Electric consumption, data mining, logistic regression, rapidMiner, classification algorithms

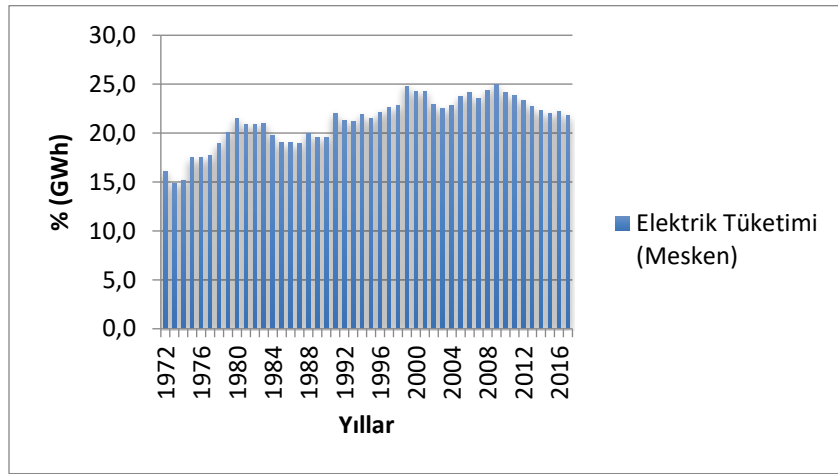
1. Giriş

Elektrik enerjisi dünya genelinde talebi ve tüketimi hızlıca artan kaliteli, verimli, modern ve en esnek yapıda olan enerji türüdür. Bu nedenle gündelik hayatımızda birçok yerde geniş kullanım alanına sahiptir [1].Nüfus, şehirleşme, sanayileşmenin artması ve insan hayatını kolaylaştırmada en önemli unsur olan teknolojinin gün geçtikçe ilerlemesi elektrik enerjisine olan bağımlılığı giderek arttırmaktadır [2].Buna bağlı olarak Şekil 1'de Türkiye Elektrik Dağıtım Anonim

*Corresponding author: Address: Faculty of Technology, Information Systems Engineering Kocaeli University, 41380, Kocaeli TURKEY. E-mail address: seda.balta@kocaeli.edu.tr, Phone: +902623032260

Şirketi'nin 1972–2017 yılları arası toplam elektrik tüketimi görülmektedir. Şekil 1'e göre yıllar geçtikçe elektrik tüketimi dalgalanmalarla birlikte artış göstermektedir.

İnsanlar gerek temel ihtiyaçlarını karşılamak gerekse hayatlarını kolaylaştırmak için her geçen gün sayısı artan elektrik ev aletlerine sahip olmaktadır. Bu kapsamda gelişmekte olan teknoloji insanların elektrik talebine büyük oranda etki etmektedir [3]. Elektrik tüketimlerinin artmasıyla birlikte insanlar bazı sorunlarla karşı karşıya kalmaktadır. Tüketimi icra ederken maliyeti minimuma indirmeyi, faydaları ise maksimuma çıkarmayı amaç haline getirmektedirler. Bu mantıkla optimum maliyet–fayda ilişkisini bulmak amaçlı davranışlarını gözden geçirmek zorundadırlar [4].



Şekil 1. 1972-2017 yılları arasında Türkiye’de meskenlerdeki elektrik tüketiminin toplam elektrik tüketimine yüzdesel oranı

Bu çalışmada, mesken kullanıcılarının elektrik tüketim davranışlarının veri madenciliği teknikleri ile analizi ve kullanıcı için en uygun tarife seçimi sunulmaktadır. Bu amaçla Türkiye’deki elektrik faturalandırma sistemi incelenmiş ve verilere RapidMiner ortamında veri madenciliği teknikleri uygulanmıştır.

Makalenin geri kalan kısmı şu şekilde organize edilmiştir. Bölüm 2’de veri madenciliği alanındaki algoritmaların kullanım alanları örneklendirilmiştir. Bölüm 3.1’de veri setinin hazırlanması, Bölüm 3.2’de Lojistik regresyon seçilmesinin sebebi anlatılmaktadır. Lojistik regresyon algoritmasının RapidMiner ortamında elektrik tüketim verileri üzerinde uygulaması Bölüm 4’te verilmektedir. Bölüm 4.1’de benzetim sonuçları, Bölüm 5’te çalışmanın sonuçları ele alınmaktadır.

2. İlgili Çalışmalar

Tablo 1’de ilgili çalışmalarda gösterildiği gibi, çeşitli veri setleri ve algoritmalar kullanılarak sınıflandırma işlemleri yapılmıştır. Bu sınıflandırma işlemlerine göre algoritmaların başarımları

oranları gösterilmiştir. Bu çalışmada çeşitli sınıflandırma algoritmaları kullanılarak önce en uygun algoritma seçilecek, daha sonra seçilen algoritma ile sınıflandırma işlemi yapılarak başarı oranı gösterilecektir.

Tablo 1. Önerilen sistem için ilgili çalışmaların karşılaştırılması

İlgili Çalışmalar	Uygulanan Algoritmalar	Bağımsız Değişkenler	Motivasyon
[5]	Regresyon Analizi	Isıtma Soğutma Havalandırma sistemleri	Enerji tasarrufu sağlamak
[6]	K-Means	Hava şartları Isıtma Soğutma Havalandırma sistemleri	Enerji tasarrufu sağlamak
[7]	Sinir ağları Fuzzy teorisi Gap İstatistik Algoritması	Gürültülü veri Ayrıklı veri Eksik veri	Enerji tüketim veri niteliklerini incelemek
[8]	K-Means	Bulaşık makinesi Çamaşır makinesi Kurutma makinesi Kettle Mikrodalga fırın vb.	Enerji tasarrufu sağlamak
[9]	Çoklu lineer regresyon	Ev boyutu Çocuk sayısı Oda sayısı Toplam Gelir	Enerji tasarrufu sağlamak
[10]	Yapay sinir ağları	Nüfus Toplam Üretim İthalat İhracat	Enerji tüketimini tahmin etmek
Önerilen Sistem	Lojistik Regresyon	Gündüz Puant Gece	Enerji tasarrufu sağlamak

3. Önerilen Sistem

3.1. Veri setinin hazırlanması

Türkiye'deki elektrik faturalandırma sistemine göre faturalar genel olarak 3 ana bileşenden oluşmaktadır.

- Enerji tüketim tutarı,
- Zaman dilimlerine göre enerji tüketimleri,
- Vergiler,

Faturalandırma yapılırken dikkat edilen bazı hususlar vardır. Bu hususların en önemlisi zaman dilimlerine göre ücretlenme yapılmasıdır. Kabaca bir gün (24 saat), Tablo 2’de gösterildiği gibi 3 farklı zaman dilimine bölünmektedir. Bu dilimler gündüz, puant ve gece olarak adlandırılmıştır.

Tablo 2. Bir günün saat dilimlerine ayrılması

Zaman Dilimleri	Adlandırma	Saat Dilimleri
Gündüz	T1	06:00-17:00
Puant	T2	17:00-22:00
Gece	T3	22:00-06:00

Faturalarda yer alan tarife tipinin anlamı çoğunlukla kullanıcılar tarafından bilinmemektedir. Faturalar tek zamanlı ve çok zamanlı tarife tipi olarak iki farklı ücretlendirme türü esas alınarak hesaplanabilmektedir. Bütün aboneler varsayılan olarak tek zamanlı tarife türüne üyedir.

Tek zamanlı tarifede, (1)’de gösterildiği gibi zaman dilimlerine bakılmaksızın kullanılan toplam enerji tek tür ücretlendirmeye tabii tutulur.

$$\text{Aktif_tüketim (kWh)} * \text{Aktif_birimfiyat (TL/KwH)} = \text{Tüketim tutarı} \quad (1)$$

Çok zamanlı tarifede ise (2), (3) ve (4)’te gösterildiği gibi hangi zaman diliminde ne kadar elektrik tüketildiği ile ilişkili ücretlendirmeler esas alınır.

$$\text{Gündüz : T1_tüketim (kWh)} * \text{T1_birimfiyat} = \text{T1_tüketimtutarı} \quad (2)$$

$$\text{Puant : T2_tüketim (kWh)} * \text{T2_birimfiyat} = \text{T2_tüketimtutarı} \quad (3)$$

$$\text{Gece : T3_tüketim (kWh)} * \text{T3_birimfiyat} = \text{T3_tüketimtutarı} \quad (4)$$

Faturalandırma ücretleri Tablo 3’te gösterildiği gibi her bir zaman diliminde farklılık göstermektedir. Zaman dilimlerine göre ücretlendirmeler göz önüne alındığında en düşük ücretlendirme gece saatlerinde olduğundan elektrikli aletlerin gece çalıştırılması en uygun çözüm olacaktır.

Tablo 3. Tarife türüne göre saat bazlı ücretlendirme

Tarife Tipi	Zaman Dilimi	Birim Fiyat (TL)
Tek Zamanlı Tarife	Gündüz-Puant-Gece	0,2791
Çok Zamanlı Tarife	Gündüz	0,2846
Çok Zamanlı Tarife	Puant	0,4848
Çok Zamanlı Tarife	Gece	0,1245

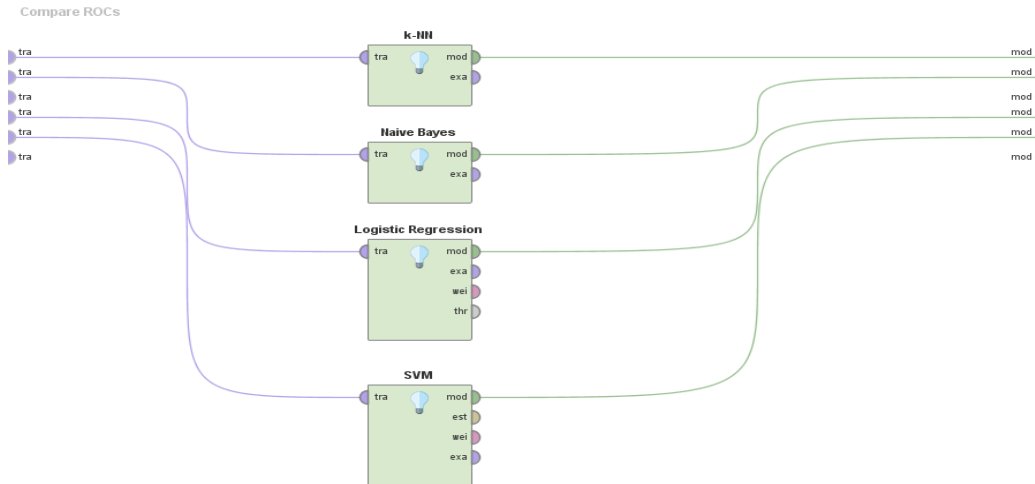
Bu çalışmada 1500 adet kullanıcının gündüz, puant ve gece değerleri Şekil 2'deki gibi rastgele olarak belirlenmiştir. Öncelikle yukarıda anlatılan faturalandırma sistemine göre kullanıcının elektrik tüketiminin hangi tarifeye uygun olduğu belirlenerek sınıflandırma yapılmıştır. Daha sonra ilgili yazılım aracılığıyla sınıflandırma algoritmaları kullanarak tarife tipini en uygun seçen algoritma analiz edilmiştir.

Gündüz	Puant	Gece	Tarife
164	314	307	Tek
315	194	139	Tek
143	178	359	Çok
187	354	210	Tek
268	148	359	Çok

Şekil 2. Veri seti örneği

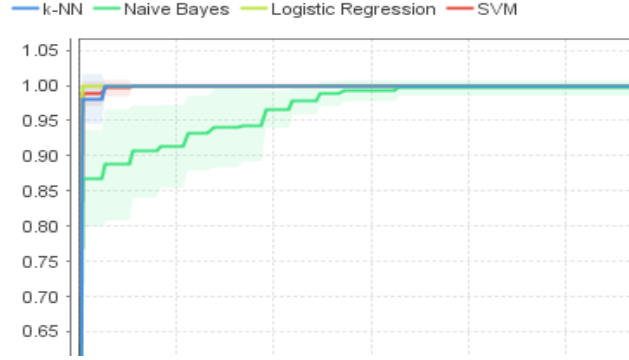
3.2. Algoritma seçimi

Bu çalışmada 1500 adet kullanıcıdan alındığı varsayılan veriyi analiz edebilmek için kullanılacak olan en uygun algoritmanın seçiminde RapidMiner yazılımı kullanılmaktadır. RapidMiner aracılığıyla belirlenen popüler algoritmalar arasında Şekil 3'te gösterildiği gibi ROC analizi yapılarak veri setine en uygun algoritma seçilmektedir. Şekilden görüldüğü üzere k-NN, Naive Bayes, Logistik Regresyon ve SVM algoritmaları ROC analizi gerçekleştirilmektedir.



Şekil 3. ROC Analizi

Şekil 4'te gösterildiği gibi ROC analizinde 1.00 tam sayısına ilk kez en yakın sonucu veren algoritma en yüksek doğruluk oranını verecektir. Yapılan ROC analizine göre rastgele olarak belirlenen veri setinde tarife seçimi yaparken kullanılacak, en yüksek başarıım yüzdesine sahip olan algoritmanın Lojistik Regresyon olduğu görülmektedir.



Şekil 4. ROC analiz sonucu

Lojistik regresyon bir bağımlı değişkenin bir veya birden fazla bağımsız değişken ile arasındaki ilişkiyi bulmak için kullanılan istatistiksel bir yöntemdir [11]. Her bağımsız değişken ağırlıklar ile çarpılır ve toplanır. 0 ile 1 arasında sonuç bulmak için bu çıkan değer sigmoid fonksiyonuna eklenir. Sonuç 0 ile 0,5 arasındaysa 0 kabul edilir ve ilişki yoktur denilir. Fakat sonuç 0,5 ile 1 arasındaysa 1 kabul edilir. Bağımlı ve bağımsız değişkenler arasında bir ilişkinin varlığından söz edilebilir. En iyi ağırlıkları ve regresyon katsayılarını bulmak önemlidir.

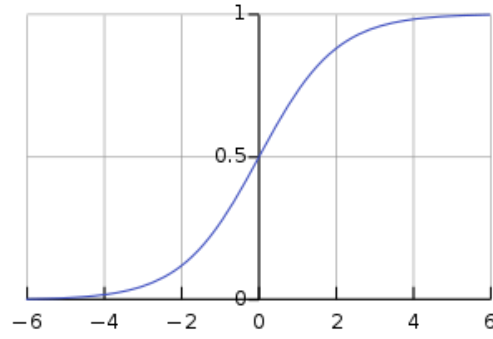
(5), (6) ve (7)'de görüldüğü gibi bir model, bir olayın gerçekleşme olasılığının gerçekleşmeme olasılığına bölünmesinin ardından doğal logaritmasının alınması ile oluşmaktadır [12].

$$\text{Logit}(p) = \log [p/(1-p)] = \ln [p/(1-p)] \quad (5)$$

$$\text{Logit}[p(x)] = \log [p(x)/1-p(x)] = a + b_1x_1 + b_2x_2 + \dots + b_ix_i \quad (6)$$

$$P = \exp(a + b_1x_1 + b_2x_2 + \dots + b_ix_i) / 1 + \exp(a + b_1x_1 + b_2x_2 + \dots + b_ix_i) \quad (7)$$

Logaritmik dağılımın kullanılmasının nedeni dağılımı normalleştirerek hata payını en aza indirmektir. Eşitlikler neticesinde kategorik değişken Şekil 5'te görüldüğü gibi 1 ve 0 arasındayken $-\infty$ ile $+\infty$ arasında değerler almaktadır.



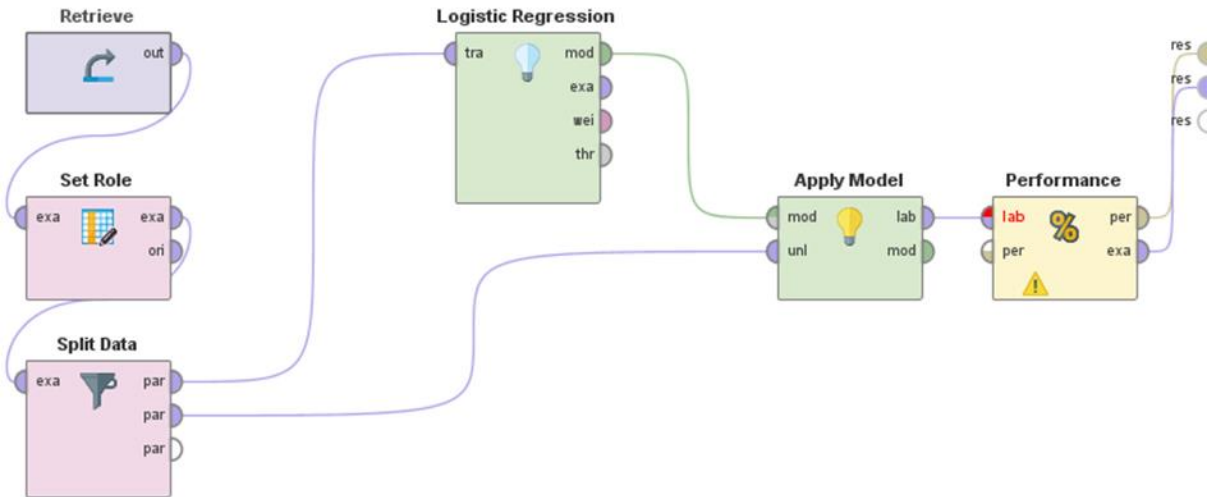
Şekil 5. Lojistik regresyon sigmoid fonksiyonu grafiği[13]

4. Uygulamanın Gerçekleştirilmesi

Bu bölümde Lojistik Regresyon algoritmasının veri seti üzerinde kullanımını gerçekleştirilmektedir. Tasarım oluşturulurken Şekil 5'teki gibi bir iş akışı göz önünde bulundurulmuştur.

- Veri setini al.
- Sınıflandırma yapılacak sütunun seçimi yap.
- Tüm veriyi eğitim ve test verisi olarak iki farklı veri setine ayır.
- Eğitim verisine Lojistik Regresyon algoritmasını uygulayarak öğrenmeyi sağla.
- Test verisine Lojistik Regresyon uygula.
- Test verisine uygulanan algoritmanın performansını ölç.

Yukarıda anlatılan iş akışına göre Şekil 5'te RapidMiner programında oluşturulan tasarım görülmektedir.



Şekil 6. RapidMiner tasarım ekranı

4.1. Benzetim sonuçları

1500 adet verilik veri setimiz %70 eğitim verisi, %30 test verisi olarak ayrılmaktadır. Şekil 7’de gösterildiği gibi veri setinde bulunan gündüz, puant ve gece değerlerine lojistik regresyon uygulanmıştır. Bu verilere göre örneğin 481.satır tahmini yapılırken uygulanan algoritma sonucunda, ‘Çok’ tarifesi olma olasılığı 0,126 ‘Tek’ tarifesi olma olasılığı ise 0,874 olarak belirlenmiştir. Algoritma Seçimi bölümünde anlatılanlar doğrultusunda ‘Tek’ tarifesine dahil edilmiştir.

Row No.	Tarife	prediction(Tarife)	confidence(Cok)	confidence(Tek)	Gunduz	Puant	Gece
479	Tek	Tek	0	1	233	690	685
480	Tek	Tek	0	1	1928	1699	237
481	Tek	Tek	0.126	0.874	819	120	182

Şekil 7. Sınıflandırma tahmin tablosu

Tablo 4’te elde edilen sonuçlara göre veri setimizde 549 adet ‘Çok’, ‘951’ adet ‘Tek’ zamanlı tarifeye ait kullanıcı bulunmaktadır. Uygulanan algoritma sonucunda bu verilerin tamamı sınıflandırılırken %100 doğruluk oranı elde edilmiştir.

Tablo 4. RapidMiner performans vektörü

	Doğru ‘Çok’	Doğru ‘Tek’	Sınıf Kesinliği
Tahmin ‘Çok’	549	0	% 100
Tahmin ‘Tek’	0	951	% 100
Sınıfın Hatırlanması	% 100	% 100	

5. Sonuç

Bu çalışma sonuçlarına göre Türkiye’de varolan her bir meskenin gündüz, puant ve gece toplam elektrik tüketimleri aylık bazda incelenip, Lojistik regresyon uygulandığında elektrik faturalandırma sisteminde tasarrufa gidilebilir. Ayrıca çalışmada sadece meskenlerdeki elektrik tüketimi miktarları incelenmiştir. Diğer abone grupları olan ticarethane, sanayi, tarımsal sulama ve aydınlatma abone grupları içinde aynı analizler yapılarak çalışma genişletilebilir ve bu alanlarda da tasarrufa gidilebilir.

References

- [1] Haliloğlu EY, Tutu BE. Short-term electricity power demand forecasting for Turkey. Journal of Yasar University 2018;13(51):243-255.
- [2] Nişancı M. Türkiye’de elektrik enerjisi talebi ve elektrik tüketimi ile ekonomik büyüme arasındaki ilişki. Sosyal Ekonomik Araştırmalar Dergisi 2005;5(9):107-121.
- [3] Akan Y, Tak S. Türkiye elektrik enerjisi ekonometrik talep analizi. Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi 2003;17:1-2.
- [4] Güloğlu B, Akın E. Türkiye’de hane halkları elektrik talebinin belirleyicileri: sıralı logit yaklaşımı. Siyaset, Ekonomi ve Yönetim Araştırmaları Dergisi 2014;2(3):1-20.
- [5] Moreno M, Ubeda B, Skarmeta A, Zamora M. How can we tackle energy efficiency in iot based smart buildings?. Sensors;14(6):9582-9614.
- [6] Terroso-Saenz F, Gonzalez-Vidal A, Ramallo-Gonzalez AP, Skarmeta AF. An open IoT platform for the management and analysis of energy data. Future Generation Computer Systems 2019;92:1066-1079
- [7] Chen W, Zhou K, Yang S, Wu C. Data quality of electricity consumption data in a smart grid environment. Renewable and Sustainable Energy Reviews 2017;75:98-105.
- [8] Gajowniczek K, Zabkowski T. Data mining techniques for detecting household characteristics based on smart meter data. Energies 2015;8(7):7407-7427.
- [9] Beckel C, Sadamori L, Staake T, Santini S. Revealing household characteristics from smart meter data. Energy 2014;78:397-410.
- [10] Kavaklioglu K, Ceylan H, Ozturk HK, Canyurt OE. Modeling and prediction of Turkey’s electricity consumption using artificial neural networks. Energy Conversion and Management 2009;50(11):2719-2727.
- [11] Ürük E. İstatistiksel uygulamalarda lojistik regresyon analizi. Marmara Üniversitesi Sosyal Bilimler Enstitüsü. İstanbul: Yüksek Lisans Tezi; 2007.
- [12] Pandey R. Introduction to Logistic Regression, <https://www.quora.com/What-exactly-is-a-logistic-regression-algorithm-in-machine-learning-What-are-its-applications> (26.09.2019)
- [13] Bozarık E. Sınır ağları ve derin öğrenme : Lojistik Regresyon, <https://medium.com/deep-learning-turkiye/sinir-aglari-ve-derin-ogrenme-iii-lojistik-regresyon-cc9686981c6b> (20.08.2019)