

Veri Madenciliği Yöntemleri İle Meteorolojik Verilerden Kayıp Güneş Işınım Değerlerinin Tahmini

Duygu Sağaltıcı, Fatma Didem Alay, Cihan Efil, Nagehan İlhan

Harran Üniversitesi, Elektrik Elektronik ve Bilgisayar Mühendisliği Bölümü, Şanlıurfa
e-posta: duyugasaltici@harran.edu.tr

Geliş Tarihi: 04.07.2018

Kabul Tarihi: 31.10.2018

Özet

Günümüzde doğaya zarar vermemekle birlikte, sınırsız ve dünyanın her yerinde mevcut olan yenilenebilir kaynakların kullanımı önceki yıllara oranla ciddi miktarda artış göstermektedir. Yenilenebilir enerji kaynakları arasında kullanımı giderek yaygınlaşan sistemler, güneş enerjisini elektrik enerjisine dönüştüren PV panellerdir. Güneş enerjisi sistemlerinde enerji elde etmede en önemli etkenlerden biri güneş radyasyon değeridir. Sistem tasarlanırken panel eğim açısının nasıl belirleneceğine güneş ışınlarının panele gelme açıları değerlendirilerek karar verilir. Güneş radyasyonu sıcaklık, nem, basınç, rüzgar, güneşli gün sayısı ve coğrafi konuma bağlı olarak değişim göstermektedir. Çalışmada Harran Üniversitesi bünyesinde yer alan GAPYENEV merkezindeki meteorolojik gözlem istasyonundan 2013 yılından itibaren alınan meteorolojik veriler kullanılmaktadır. Bu istasyondan toplanan verilerdeki kayıp güneş ışınım değerlerinin çeşitli veri madenciliği yöntemleriyle tahmin edilmesi çalışmaları yapılarak Şanlıurfa ili için güneş ışınım değerlerinin doğruluğu artırılmakta, böylelikle güneş panelleri ile yenilenebilir enerji üretiminde enerji veriminin yükseltmesi amaçlanmaktadır.

Anahtar kelimeler: veri madenciliği; kayıp veri tahmini; yapay sinir ağları; güneş radyasyon değeri

Estimation Of Missing Solar Radiation Values Of Meteorological Data Using Data Mining Techniques

Abstract

Today, many scientists have sought to find renewable resources that are unlimited and available all over the world to prevent any possible harm and damage in the nature. Photovoltaic (PV) panels are getting one of the most popular and widespread among renewable energy sources. It basically converts solar energy into electricity. One of the most important factors in obtaining energy is the solar radiation intensity. The optimum angle of the panel for the sake of the maximum efficiency is determined by studying the angle of sun rays coming towards the pan. Solar radiation varies with respect to temperature, humidity, pressure, wind, number of sunny days and geographical location. The meteorological data, used in this study, are collected from the meteorological observation station at the GAPYENEV center in Harran University since 2013. The accuracy of solar radiation intensity for Şanlıurfa province is increased by estimating the losses in solar radiation collected from GAPYENEV with various data mining methods. By doing this, we plan to improve the efficiency in renewable energy with solar panels.

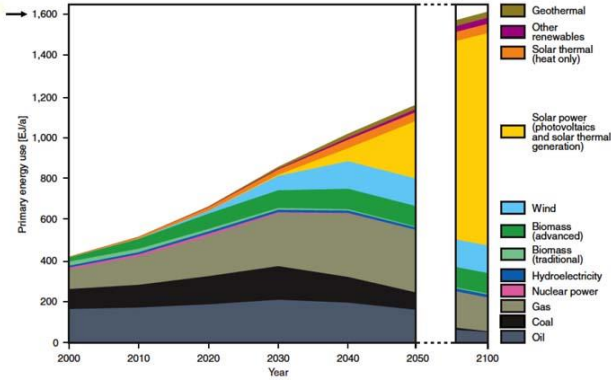
Keywords: data mining; missing data values; artificial neural networks; solar global radiance

1. Giriş

Günümüzde enerji tüketimi artan nüfus ve sanayi gelişmeleri ile daha büyük boyutlara ulaşmış ve kısıtlı kaynaklarla karşılanamayacak hale gelmiştir. Bununla birlikte enerji üretiminde kullanılan kaynaklar doğaya ve insan sağlığına zarar veren etmenlere sahiptir. Bu nedenle pek çok bilim insanı doğaya zarar vermemekle birlikte, sınırsız ve dünyanın her yerinde mevcut olan yenilenebilir kaynaklar arayışına çıkmıştır. Yenilenemeyen enerji kaynakları kısa vadede daha ucuz olarak nitelendirilse de uzun vadede yenilenebilir enerji

kaynakları sınırsızdır ve hammadde maliyeti yoktur [1] [2].

Yenilenebilir enerji kaynakları arasında kullanımı giderek yaygınlaşan sistemler, güneş enerjisini elektrik enerjisine dönüştüren PV panellerdir. Ancak bu sistemlerin kurulum maliyeti yüksektir ve enerji verimliliğini düşüren pek çok etmen söz konusudur. Küresel Değişim Alman Danışma Konseyi (German Advisory Council on Global Change, WBGU, 2003) tahminlerine göre geleceğin en yaygın enerji kaynağı Şekil 1 de görüldüğü üzere güneş enerjisi ve sistemleri olacaktır [3].



Şekil 1. Dünya birincil enerji kaynaklarının evrimi(WBGU, 2003).

Güneş radyasyon değeri ve fotovoltaik, güneş enerjisi sistem tasarımı için temeli oluşturan verilerdendir. Sistem tasarlanırken panel eğim açısının nasıl belirleneceğine güneş ışınlarının panele gelme açıları değerlendirilerek karar verilir. Güneş radyasyonu sıcaklık, nem, basınç, rüzgar, güneşli gün sayısı ve coğrafi konuma bağlı olarak değişim göstermektedir. Güneş radyasyon değerlerinin ölçümünde kullanılacak teknikler ve çeşitli ölçüm cihazları mevcuttur. Ancak meteorolojik verilerin elde edilmesi diğer ölçümlere göre daha kolaydır [4]. Bu verilerin elde edilmesinde ışınım değerlerinin kayıplar içermesi büyük bir sorun oluşturmaktadır. Literatürde birçok çalışmada kayıp veri çözümünde veri madenciliği yöntemlerinin yüksek doğruluk verdiği gözlemlenmiştir [1] [5].

Literatürde küresel radyasyon değeri tahminleri üzerinde durulmuş ve çeşitli yöntemler geliştirilmiştir. Kadir Bakırcı, Erzurum ili için yatay yüzeye gelen anlık global radyasyonu basit bir hesaplama yöntemi kullanarak tahmin etmiştir [6]. 2009 yılında yapılan bu çalışmadan günümüze gelindiğinde global radyasyon değerini ölçen materyaller kullanılmaktadır. Malezya'da yapılan başka bir çalışmada ise sınırlı sayıdaki radyasyon ağı ve materyallerin maddi anlamda zorlayıcı olması nedenleriyle farklı bir yol izlenmiştir. Engel-Cox ve ark. Malezya'da güneş enerjisi teknolojisinin performansı için güneş ve ilgili meteorolojik verilerin değerlendirilmesini yapmışlardır. Güneş radyasyon verilerini tahmin etmek için sıcaklık, nem, rüzgar hızı, yağış, bulutluluk oranı, enlem, boylam ve parçacık miktarlarını girdi verisi olarak kullanarak yapay sinir ağları (YSA) metodunu kullanmışlardır. Daha kolay ulaşılan meteorolojik verilerden yararlanılarak yapay sinir ağları ile küresel radyasyon değerine ulaşmak hedeflenmiştir [2]. Güneydoğu

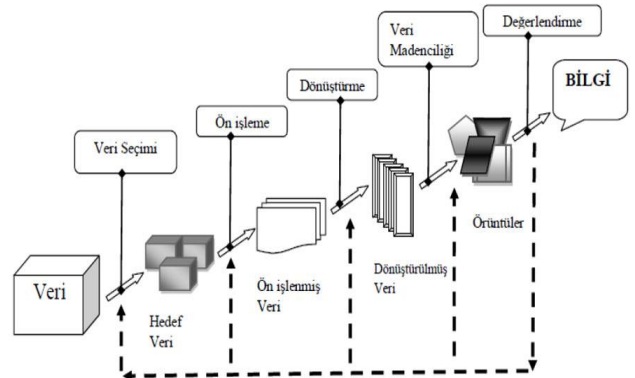
Anadolu bölgemizde yer alan Gaziantep ilimizde yapılan diğer bir çalışmada Nermin Şarlak ve Aytaç Güven güneş radyasyon verilerini genetik programlama tekniğini kullanarak tahmin etmeye çalışmışlardır. Yapılan model karşılaştırmaları sonucunda mevcut veriler ile en iyi tahmin sonuçlarını lineer genetik programlama tekniğinin verdiği tespit edilmiştir.

Meteorolojik verilerden elde edilen parametrelerdeki kayıp ışınım değerlerinin tahmini için yapılan çalışmalarda Lineer Genetik Programlama tekniği kullanılarak iyi sonuçlar elde edildiği görülmüştür. Bu çalışmada ise Yapay Sinir Ağları ile diğer veri madenciliği tekniklerinin kullanılmasıyla meteorolojik verilerden elde edilen ışınım değerlerindeki kayıp verilerin tahmini yapılarak sonuçlar karşılaştırılmakta ve en uygun yöntemin değerlendirilmesi yapılmaktadır.

2. Veri Madenciliği

Çağımızda teknolojinin gelişmesiyle birlikte veri akışı da büyük bir hız kazanmıştır. Öyle ki veriler işlem yapmayı zorlaştıracak kadar büyümüş ve anlamsız bilgi yığınlarına dönüşmeye başlamıştır.

Veri madenciliği büyük ve bir kenara atılmış verilerden anlamlı sonuçlar, bilinmeyen ya da öngörülemeyen bağıntılar ve tahmin yöntemleri elde eden bir analiz tekniğidir. Veri madenciliği yöntemlerinin kullanılabilmesi ve doğru bilgiye ulaşmak için verilerin bir dizi ön işlemden geçerek analiz edilebilecek hale gelmesi gerekmektedir [7].



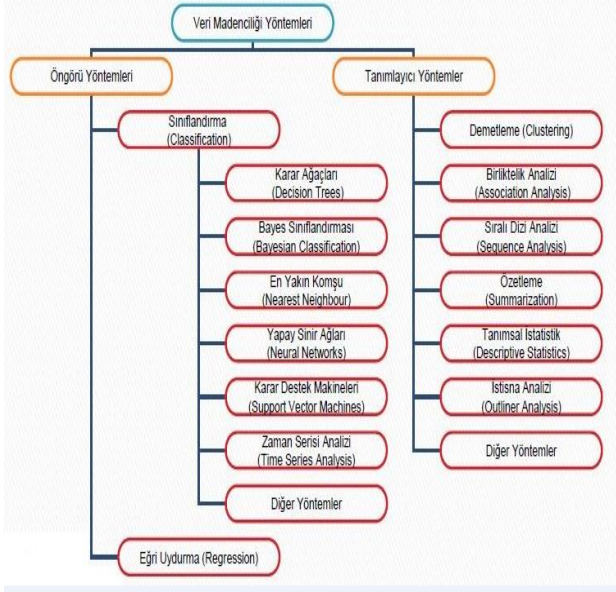
Şekil 2. Veri Madenciliği Bilgi Keşif Süreci.

Şekil-2'de bilgiye ulaşmak amacıyla izlenmesi gereken bilgi keşif süreci görülmektedir [8]. Veri madenciliğinde en uzun süreç verinin hazırlanması

kısımdır ve yaklaşık %80'lik bir zaman dilimini kapsamaktadır. Verinin ön işleme basamakları veri temizleme, veri bütünleştirme, veri indirgeme, veri dönüştürme ve veri ayrıklaştırma olarak beş maddeye ayrılabilir [9].

2.1. Veri Madenciliği Yöntemleri

Veri madenciliği konusunda geliştirilen yöntemler arasında sıklıkla kullanılanlar sınıflandırma, kümeleme ve birliktelik analizidir. Sınıflandırma yöntemleri Şekil-3'de görüldüğü üzere öngörü yöntemlerinden biri olarak kabul edilir [10].



Şekil 3. Veri madenciliği yöntemleri.

Sınıflandırmada amaç verilerdeki gizli örüntüyü açığa çıkarmaktır. Verilerin sınıflandırılması için belirli kurallar oluşturmak amacıyla veri tabanının bir kısmı eğitim geriye kalan kısmı da test olarak ayrılır. Ortaya çıkan kurallar ile veriler sınıflara ayrılır ve ortaya çıkan sonuç karar vermeyi sağlar [11].

2.2. Eksik Değerler

Bazı uygulama ya da cihazlardan alınan veriler eksik bilgiler içerebilir. Bu da veri madenciliği yöntemlerinin kullanımında istenmeyen bir durumdur. Veri temizleme yöntemleri kullanılarak eksik değerler veri kümesinden atılabilir. Veriler önem arz ediyor ise kayıp değer regresyon ya da karar ağaçları gibi sınıflandırma yöntemleri kullanılarak tahmin edilebilir. Bu çalışmada veri setindeki kayıp değerler karar ağaçları, regresyon, yapay sinir ağları ve en yakın komşu gibi yöntemler kullanılarak tahmin edilmiştir [11].

3. Veriler ve Yöntem

Çalışmada kullanılan veriler GAPYENEV meteorolojik gözlem istasyonundan alınmıştır. 2013-2018 yıllarını kapsayan veriler küresel güneş radyasyonu, güneş süresi, sıcaklık, nem, rüzgâr hızı, rüzgâr yönü, basınç ve toplam yağış parametrelerinden oluşmaktadır.

TMSTAMP	GlobGunRad	Gunessuresi	Sicaklik	NNem	rzgortyon	rzgorthiz	rzgmaxyon	rzgmaxhiz	BASINC	YAGIS_Tot
05.03.2013 11:50	809	5,083	13,98	30,66	322	5,803	325,5	7,79	950,7495	0
05.03.2013 12:00	?	10	14,38	32,65	328,8	5,463	318,1	7,25	950,5808	0
05.03.2013 12:10	804	10	13,7	30,12	339,8	5,173	343	7,79	950,417	0
05.03.2013 12:20	798,3	10	14,17	26,28	328,6	5,237	342,3	8,27	950,7418	0,4
05.03.2013 12:30	?	10	13,76	26,74	338	5,026	24,25	7,61	950,4167	0
05.03.2013 12:40	783,1	10	13,96	25,67	331	5,111	357,7	8,51	950,0916	0
05.03.2013 12:50	776,2	10	14,37	29,1	314	4,663	306,5	7,73	950,4167	0
05.03.2013 13:00	?	10	13,97	26,61	302,5	6,688	305,5	10,09	950,4167	0,3
05.03.2013 13:10	748,3	10	14	25,07	312,7	6,744	300,5	8,89	950,0916	0
05.03.2013 13:20	701,7	9,83	14,24	24,59	306,5	5,897	321,4	9,93	950,1729	0
05.03.2013 13:30	716,1	10	14,17	25	310,3	6,283	301,5	8,47	950,0916	0

Şekil 4. GAPYENEV meteorolojik gözlem istasyonundan alınan 2013-2018 yılları arası veriler

GAPYENEV' den alınan verilerin içerdiği kayıp değerlerin tahmininde kullanılan yöntemler:

- Çok katmanlı algılayıcılar (MLP)
- En yakın komşu Algoritması (IBK)
- Randomtree
- Reptree

Gece alınan ölçümlerde küresel radyasyon değerlerinin 0 olması gereksiz veri olarak nitelendirilmiş ve veriler arasından çıkarılmıştır. Böylece 263088 adet veri 139025 veriye indirgenmiştir. Ayrıca veriler yıllık değerlerine göre 5 sınıfa ayrılmıştır. 2013, 2014, 2015, 2016, 2017-2018 olmak üzere ayrılan veriler üzerinde her bir veri seti için yöntemler denenmiştir. Veri setlerindeki kayıp değerlerin artması ile gösterdiği performans ölçülmek istenmiş bu nedenle %1, %5, %10 ve %15 kayıp değer içeren veriler analize alınmıştır. Toplamda 80 analiz ve sonuçlarına göre hata hesaplamaları yapılmıştır.

4. Analiz Sonuçları

Eksik değerlerin tahmini için veriler WEKA programı formatına uygun hale getirildikten sonra belirtilen yöntemler ile analiz edilmiştir. Tablo-1, Tablo-2, Tablo-3, Tablo-4 ve Tablo-5 ' te 2013-2018 yılları

arası uygulanan yöntemler ve kayıp değerlerin yüzde sayılarına göre elde edilen sonuçlar görülmektedir.

Tablo 1. 2013 yılına ait güneş radyasyonu tahmini.

Sınıflandırıcılar	%1	%5	%10	%15
MLP (Multilayer Perceptron)	$R^2=0.5879$ RMSE= 205.2321	$R^2=0.4960$ RMSE= 224.4405	$R^2=0.4894$ RMSE=225.0956	$R^2=0.6099$ RMSE= 196.9192
IBK (En yakın komşu)	$R^2=0.8353$ RMSE=129.7448	$R^2=0.7882$ RMSE=145.4932	$R^2=0.7778$ RMSE=148.4931	$R^2=0.7660$ RMSE=152.5148
Random Tree	$R^2=0.6552$ RMSE=187.7277	$R^2=0.5625$ RMSE= 209.1112	$R^2=0.5937$ RMSE=200.7970	$R^2=0.5460$ RMSE=212.4341
Reptree	$R^2=0.6409$ RMSE=191.5877	$R^2=0.6959$ RMSE= 174.3465	$R^2=0.6410$ RMSE=188.7395	$R^2=0.6511$ RMSE= 186.2389

Tablo 2. 2014 yılına ait güneş radyasyonu tahmini.

Sınıflandırıcılar	%1	%5	%10	%15
MLP (Multilayer Perceptron)	$R^2=0.5811$ RMSE= 195.2218	$R^2=0.5917$ RMSE=197.2958	$R^2=0.5481$ RMSE=205.4621	$R^2=0.5569$ RMSE= 207.5298
IBK (En yakın komşu)	$R^2=0.7842$ RMSE=140.1361	$R^2=0.7554$ RMSE= 152.7156	$R^2=0.7706$ RMSE=146.3831	$R^2=0.7936$ RMSE=141.6476
Random Tree	$R^2=0.6543$ RMSE=177.3600	$R^2=0.5889$ RMSE=197.9834	$R^2=0.5656$ RMSE=201.4428	$R^2=0.5811$ RMSE=201.7815
Reptree	$R^2=0.7152$ RMSE=160.9823	$R^2=0.6578$ RMSE= 180.6303	$R^2=0.64333$ RMSE=182.5258	$R^2=0.6781$ RMSE=0.6781

Tablo 3. 2015 yılına ait güneş radyasyonu tahmini.

Sınıflandırıcılar	%1	%5	%10	%15
MLP (Multilayer Perceptron)	$R^2=0.6587$ RMSE= 174.6027	$R^2=0.6120$ RMSE=184.2345	$R^2=0.5885$ RMSE= 192.0560	$R^2=0.5178$ RMSE=208.8588
IBK (En yakın komşu)	$R^2=0.7702$ RMSE=143.2801	$R^2=0.7560$ RMSE=146.0966	$R^2=0.7534$ RMSE=148.6913	$R^2=0.7454$ RMSE=151.7572
Random Tree	$R^2=0.5583$ RMSE=198.6381	$R^2=0.5504$ RMSE=198.3371	$R^2=0.5433$ RMSE= 202.3357	$R^2=0.5167$ RMSE=209.0892
Reptree	$R^2=0.6689$ RMSE=171.9771	$R^2=0.6889$ RMSE=164.9802	$R^2=0.6673$ RMSE=172.6927	$R^2=0.6635$ RMSE=174.4725

Tablo 4. 2016 yılına ait güneş radyasyonu tahmini.

Sınıflandırıcılar	%1	%5	%10	%15
MLP (Multilayer Perceptron)	$R^2=0.6277$ RMSE= 189.5406	$R^2=0.4034$ RMSE=232.8366	$R^2=0.6373$ RMSE=184.5801	$R^2=0.3789$ RMSE=242.6277
IBK (En yakın komşu)	$R^2=0.7302$ RMSE=161.3348	$R^2=0.7845$ RMSE=139.9512	$R^2=0.7608$ RMSE=149.8916	$R^2=0.7686$ RMSE=148.0904
Random Tree	$R^2=0.5806$ RMSE=201.1530	$R^2=0.5568$ RMSE=200.6844	$R^2=0.5814$ RMSE=198.2837	$R^2=0.5611$ RMSE=203.9597
Reptree	$R^2=0.6741$ RMSE=177.3149	$R^2=0.6865$ RMSE=168.7802	$R^2=0.6647$ RMSE=177.4589	$R^2=0.6685$ RMSE=177.2470

Tablo 5. 2017-2018 yılına ait güneş radyasyonu tahmini.

Sınıflandırıcılar	%1	%5	%10	%15
MLP (Multilayer Perceptron)	$R^2=0.4508$ RMSE= 216.7416	$R^2=0.5379$ RMSE=205.8243	$R^2=0.3958$ RMSE=234.2911	$R^2=0.4336$ RMSE=226.7472
IBK (En yakın komşu)	$R^2=0.7816$ RMSE=136.6746	$R^2=0.7727$ RMSE=144.3720	$R^2=0.7686$ RMSE=144.9977	$R^2=0.7073$ RMSE=162.9979
Random Tree	$R^2=0.4661$ RMSE=213.7100	$R^2=0.6118$ RMSE= 188.6552	$R^2=0.5795$ RMSE=195.4433	$R^2=0.5985$ RMSE=190.9012
Reptree	$R^2=0.6894$ RMSE=162.9923	$R^2=0.7069$ RMSE=163.9333	$R^2=0.6877$ RMSE=168.4479	$R^2=0.6769$ RMSE=171.2586

5. Sonuç ve Öneriler

Yapılan analizler ve değerlendirmeler sonucunda eksik değerlerin tahmin edilmesinde en yüksek başarıyı 'En Yakın Komşu Yöntemi (IBK)' nin verdiği görülmüştür.

Genel tahmin olarak ise karar ağaçları yöntemlerinden hibrit bir yöntem olan 'RandomTree' algoritması yüksek başarı göstermiştir (Şekil 5).

Gün geçtikçe geliştirilen yeni yöntemler ile veri madenciliğinin birleştirilmesinin daha iyi sonuç vereceği düşünülmektedir.

Teşekkür

Bu çalışmada kullanılan veriler Harran Üniversitesi GAP Yenilenebilir Enerji ve Enerji Verimliliği Merkezi (GAPYENEV) meteoroloji istasyonundan alınmıştır.

Kaynaklar

- [1] A. G. Nermin ŞARLAK, «Global Güneş Radyasyon Tahmini: Gaziantep,» *İMO Teknik Dergi*, cilt 460, pp. 7561-7568, 2016.
- [2] N. N. J. F. J.A. Engel-Cox, «Evaluation of Solar and Meteorological Data Relevant to Solar Energy Technology Performance in Malaysia,» *Journal of Sustainable Energy & Environment*, cilt 3, pp. 115-124, 2012.
- [3] H. A. Frindt Vidal, *Analysis of solar tracking systems for photovoltaic power plants considering experimental and computer modeled results for Munich, Germany, Santiago de Chile*: Harald Frindt, 2009.
- [4] O. B. Hüsamettin Bulut, «Simple model for the generation of daily global solar-radiation data in Turkey,» *Applied Energy*, cilt 84, p. 477-491, 2007.
- [5] A. T. H. Mori, «A data mining method for selecting input variables for forecasting model of global solar radiation,» %1 içinde *Transmission and Distribution*

- Conference and Exposition (T&D), Orlando, FL, USA, 2012.
- [6] K. BAKIRCI, «Yatay Yüzeğe Gelen Anlık Global Güneş Işınımının Tahmini İçin Basit Bir Hesaplama Metodu,» *Isı Bilimi ve Tekniği Dergisi*, cilt 29, no. 2, pp. 53-58, 2009.
- [7] İ. B. AYDİLEK, *Veri Kümelerindeki Eksik Değerlerin Yeni Yaklaşımlar Kullanılarak Hesaplanması*, Konya: Selçuk Üniversitesi, 2013.
- [8] N. T. M. Y. Serkan SAVAŞ, «Veri Madenciliği Ve Türkiye'deki Uygulama Örnekleri,» *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, cilt 11, no. 21, pp. 1-23, 2012.
- [9] S. Piramuthu, «Evaluating feature selection methods for learning in data mining applications,» *European Journal of Operational Research*, cilt 156, no. 2, pp. 483-494, 2004.
- [10] S. Akyokuş, «TBD veri Madenciliği Günü,» 18 Ocak 2018. [Çevrimiçi]. Available: <http://slideplayer.biz.tr>. [Erişildi: 20 Nisan 2018].
- [11] Y. ÖZKAN, *Veri Madenciliği Yöntemleri*, İstanbul: Papatya Yayıncılık Eğitim, 2016.