



Düzce Üniversitesi Bilim ve Teknoloji Dergisi

Araştırma Makalesi

Veri Madenciliği ile Üniversite Bilişim Teknik Servis Hizmetleri Analizi

 Serdar KIRIŞOĞLU ^{a*},  Abdurrahman YAKUPOĞLU ^b

^a *Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Düzce Üniversitesi, Düzce, TÜRKİYE*

^b *Elektrik-Elektronik ve Bilgisayar Mühendisliği EABD, Fen Bilimleri Enstitüsü, Düzce Üniversitesi, Düzce, TÜRKİYE*

* Sorumlu yazarın e-posta adresi: serdarkirisoglu@duzce.edu.tr

DOI: 10.29130/dubited.593830

ÖZET

Bu çalışmada Düzce Üniversitesi Teknik Servis Otomasyonu verileri üzerinde, birim personelinin performansını değerlendirmek ve yöneticiye karar destek sürecinde yardımcı olması için Veri Madenciliği (VM) bilimi kullanılmıştır. VM büyük ölçekli verilerden, anlamlı bilgi çıkarma veya geleceğe yönelik tahminlerde bulunma işi olarak adlandırılabilir. Sınıflandırma ise mevcut veriden geleceğe yönelik tahminlerde bulunma tekniğidir. Kurumsal işletmelerde ve kamu kurumlarında personel performans analizi yapabilmek ve ileriye yönelik karar destek süreçlerinde kullanmak için çok sayıda veri bulunmaktadır. Ancak yöneticilerin bu verileri ham hali ile, performans süreçlerinde kullanmaları çok zordur. Bu çalışmada birim yöneticilerine, VM'nin sınıflandırma tekniği ile personelin performans değerlendirilmesi ve karar destek sürecinde yardımcı olunması hedeflenmiştir. VM'nin sınıflandırma yöntemleri bu veriler üzerinde uygulanmış ve Derin Öğrenme yönteminin başarısı ortaya konulmuştur.

Anahtar Kelimeler: *Veri Madenciliği, Derin Öğrenme, Personel Performans Analizi*

Analysis of University Informatics Technical Services Datas With Data Mining

ABSTRACT

In this study, Data Mining (DM) science was used to evaluate the performance of unit personnel and to assist the manager in decision support process on Düzce University Technical Service Automation data. The DM can be called the task of extracting meaningful information from large-scale data or making predictions for the future. Classification is the technique of making predictions from the available data. There is a lot of data available to perform personnel performance analysis in corporate enterprises and public institutions and to use them in advanced decision support processes. However, it is very difficult for managers to use this data in its raw form in performance processes. In this study, it is aimed to assist the unit managers in the performance evaluation and decision support process of the personnel by using DM's classification technique. DM's classification methods were applied on these data and the success of Deep Learning was demonstrated.

Keywords: *Data Mining, Deep Learning, Personnel Performance Analysis*

I. GİRİŞ

Teknolojinin gelişmesi, veri tabanlarının boyutlarının büyümesi ve internet hızının artması ile konumdan bağımsız bir şekilde büyüyen veriye erişerek bu veriden anlamlı bilgi çıkartmak giderek daha fazla önem kazanmıştır. VM'nin bu denli gelişen, büyüyen ve internet hızı artan dijital dünyada kullanımını neredeyse zorunlu kılan sebeplerden biri ise hiç şüphesiz müşterilerin tatmini ve memnuniyetleri kavramlarının önem kazanmasıdır [1],[2]. Kamu kurumlarında ve kurumsal işletmelerde teknolojik cihazların altyapıdaki kullanımı, donanımsal açıdan bakım ve onarımı, bu kurumların paydaşlarına daha iyi hizmet verebilmeleri için hayati önem arz etmektedir. Konum olarak birbirine uzak, çok fazla personeli ve teknolojik ekipmanı olan bir kurumun bu tip talepleri takip edip cevap verebileceği kurumsal bir altyapıya sahip olması gerekmektedir. Düzce Üniversitesi Bilgi İşlem Daire Başkanlığı Üniversite personelinin bu tip ihtiyaçlarını karşılayabilmek ve takip edebilmek için web tabanlı bir talep bildirim sistemi kullanmaktadır [3]. Bu sistemde, bilişim personeli birimlerden gelen talepleri görüp talebi yerine getirmekte, talebin durumu ile ilgili sisteme bilgi girebilmektedir. Ayrıca talepte bulunan kurum personeli talebin hangi aşamada olduğu hakkında bilgi sahibi olabilmektedir. Bu sayede birim amirlerinin Bilgi İşlem personeli ile ilgili performans değerlendirmesi ve geleceğe yönelik tahminlerde bulunmasını sağlayacak birçok veriye sahip olmasının önü açılmıştır.

Personelin performansının artması, performans yönetimi stratejisini ve personel verimliliğini artırır, yönetimin karar vermesini destekler ve işletmelere sürdürülebilir kalkınma sağlamada yardımcı olur. Bu sebeple Xiaofan ve arkadaşları temel VM teorisini ve insan kaynakları performans yönetiminin durumunu, performans analizindeki karar ağacının uygulanması yoluyla, çalışanların performansını etkileyen gerçek nedenleri bularak, performans düzenlemelerini keşfederek, işletmeyi yönetmede etkili bir yol sunmuşlardır [4].

Dan Hou ve arkadaşları insan gücü kullanımının arttırılmasını hedefleyen çalışmada, işletme personelinin performans değerlendirmesini analiz etmek, işletme personelinin yapısını kavramak ve daha sonra yeni personelin performansını tahmin etmek için bulanık veri madenciliği algoritmasını kullanmışlardır. Bu algoritma ayrıca deneysel olarak test edilmiş ve çok iyi sonuçlar verdiği ortaya koyulmuştur [5].

VM ile sadece personel performansı değerlendirmesi değil öğrencilerin akademik olarak performansları gibi daha birçok performans analiz edilebilir. Kurniawan ve arkadaşının çalışmasında okullardaki öğrenci performansını akademik olarak tahmin etmek için veri ambarı ve veri madenciliği tekniklerinde uygulanabilecek bir model önermişlerdir. VM teknikleri, temel bilgileri veri ambarından çıkarmak ve veri ambarında depolanan değişkenler arasındaki ilişkileri araştırmak için kullanılmaktadır. Adı geçen çalışmada, başarısı düşük öğrencilere nasıl yardım edilebileceği, ders veya modül uygunluğunun nasıl değerlendirilebileceği ve okullardaki öğrencilerin akademik performansını artırmak için yapılması gereken müdahaleler VM yöntemleri kullanılarak açıklanmaya çalışılmıştır [6].

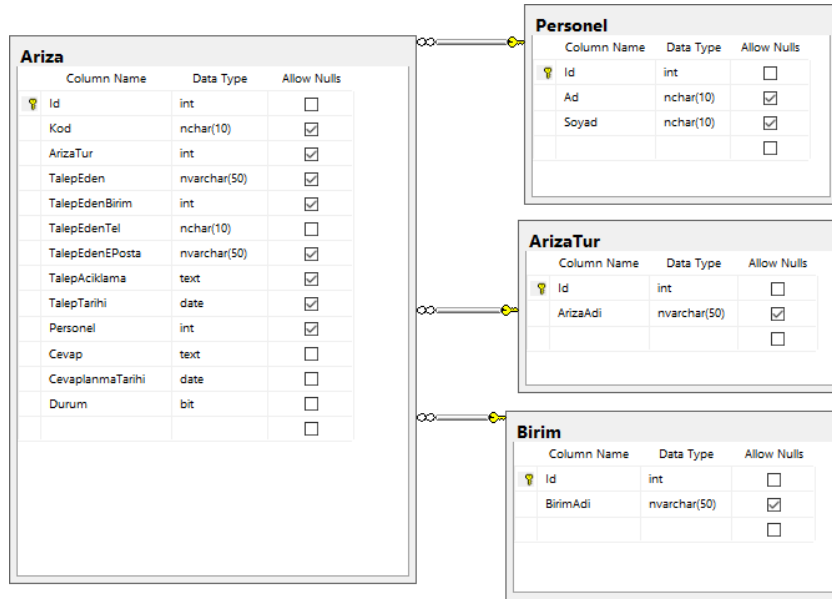
Benzer bir uygulama ise Ağaoğlu'nun C5.0 sınıflandırma tekniği kullanarak akademik camiada öğretim elemanlarının performansını değerlendirmeye yönelik çalışmasıdır. Ağaoğlu çalışmasında ayrıca karar ağacı, destek vektör makineleri, yapay sinir ağları ve diskriminant analizi olacak şekilde 4 farklı sınıflandırma tekniğini de kullanarak karşılaştırma yapmıştır. Ağaoğlu'nun analizleri, öğretim elemanlarının öğrencilerin algılarına dayalı başarılarının, temel olarak öğrencilerin derse ilgilerine bağlı olduğunu göstermiştir [7].

Literatürde personel performansının hangi etmenlere bağlı olduğundan tutun, motivasyonunun artırılmasına dair neler yapılabileceğini ortaya koyan ve VM ile yapılmış çok sayıda çalışma bulunmaktadır [8]-[11].

II. TEKNİK SERVİS VERİLERİ ÜZERİNDEN VM İLE PERFORMANS DEĞERLENDİRMESİ

A. ÇALIŞMADA KULLANILAN VM SÜREÇLERİ

Talep bildirim sistemi ASP.Net ortamında C# kodlama dili ile yazılmış bir web tabanlı uygulamadır. Bu uygulamanın kullandığı veritabanı ise MS-Sql [12] programı kullanılarak hazırlanmıştır. Ham verilerin VM yöntemleri uygulanmadan önce VM süreçlerinden geçirilmesi gerekmektedir. Bu çalışmada veriler, Veri Temizleme- Veri Bütünleştirme ve Veri İndirgeme olacak şekilde 3 ayrı VM sürecine tabi tutulmuştur. Bu süreçlerde farklı tablolardaki veriler sayısal değerlerden metinsel değerlere dönüştürülmüş, uygulamanın algoritması gereği bir talep için oluşturulmuş birden çok kayıt birleştirilmiş ve VM'yi etkileyecek ancak sonuçları etkilemeyecek eksik veriler temizlenmiş ya da sayısına göre uygun değerlerle doldurulmuştur [13].



Şekil 1. Veritabanı şeması

B. VERİLER VE ÖZELLİKLERİ

Veritabanından alınan ham veriler VM süreçlerinden geçirildikten sonra elde edilen veri setinde toplamda 5458 adet veri bulunmaktadır. Bu verilerden tahmin işlemine tabi tutulacak özellik ise arızanın giderilme süresidir. Bu süre program aracılığı ile ve Bilgi İşlem Daire Başkanlığı Hizmet standartları Envanterinde yer alan arızanın türü için farklı farklı belirlenmiş olan arıza giderilme gün sayısına bağlı olarak Uzun-Kısa-Normal olmak üzere 3 farklı değer alabilir. Bu çalışma birim yöneticilerine karar destek sürecinde kullanılacağından Tablo 1'de gösterilmiş olan sayılara sahip, arızanın hangi birimden geldiği arızanın türü ve talebe kimin cevap vereceği özellikleri ise tahmin yapılırken kullanılmıştır.

Tablo 1. Veri setindeki benzersiz kayıtlar

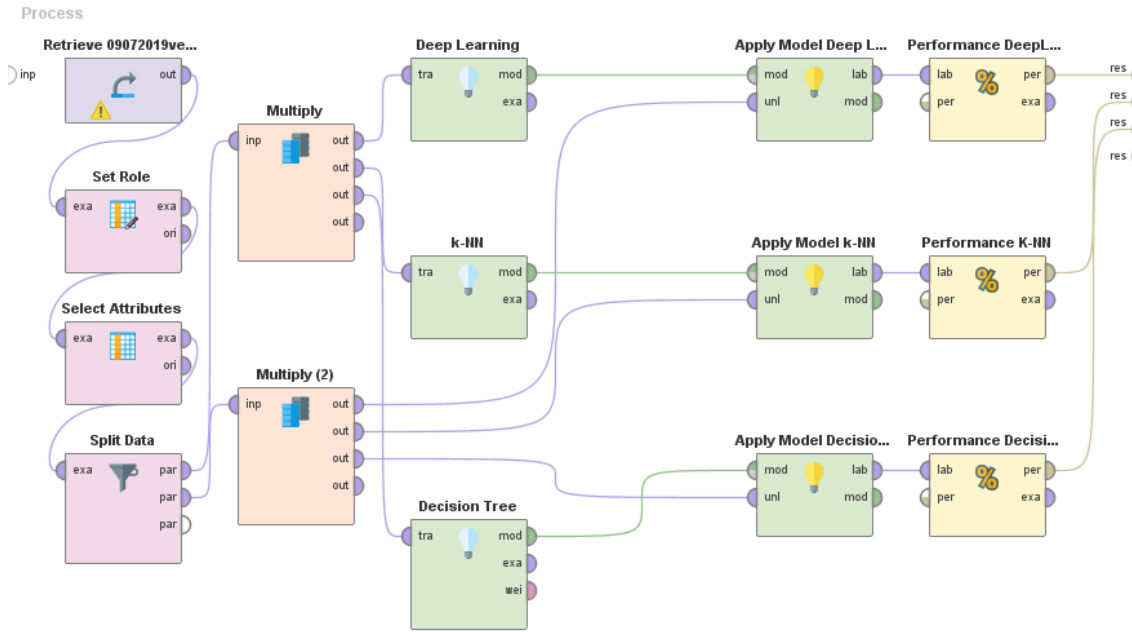
Talebe Cevap Veren Personel Sayısı	Arıza Talep Türü Sayısı	Talep Oluşturan Birim Sayısı
12	20	47

C. YÖNTEM

SPSS, WEKA, RapidMiner, Orange, SAS, KNIME, PSS Clementine, Angoss, KXEN, Sciptella ETL gibi programlar VM yöntemleri ile veri analiz etmek için geliştirilmiş birçok açık ve kapalı kaynak kodlu yazılımlardan bazılarıdır. Bu çalışmada RapidMiner [14] adlı yazılım kullanılmıştır. Bu çalışmada, literatürde yer alan personel performans değerlendirmesinde en çok kullanılan algoritmalar, Derin Öğrenme, k-NN ve Karar Ağacı algoritmaları kullanılmıştır.

RapidMiner programında VM’de kullanılacak olan Şekil 2’de gösterilen model için veri setinin yüklenmesi, tahmin edilecek olan özelliğin seçilmesi hangi özelliklere göre tahmin yapılacağı bilgilerinin girilmesi ve tahmin yapmak üzere veri setinin bir kısmının eğitim kalan diğer kısmının da test verisi olarak bölünmesi gerekmektedir.

Bu çalışmada veri setinin %60’ı eğitim verisi %40’ı test verisi olarak kullanılmıştır. Bu aşamalardan sonra performans ölçümü için uygulanan modellere eğitim ve tahmin verilerinin sonuçları Performans katmanına konulmuştur.



Şekil 2. Rapidminer ile modelleme

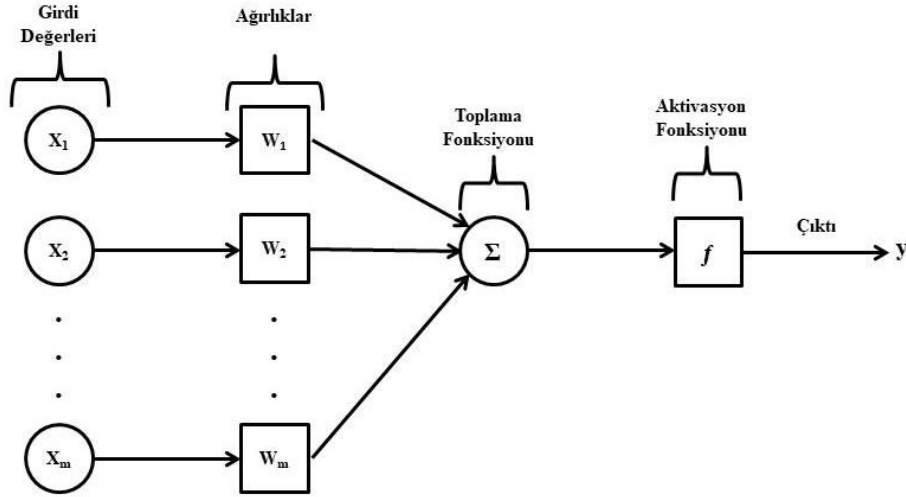
Performans bloğundan çıkan sonuç, uygulanan modelin doğruluk oranını göstermektedir. Bu oran, tahmin edilecek verinin alabileceği değerleri, sistemin doğru ya da yanlış tahmin etmesi ile şekillenir. Veri setindeki tahmin edilecek verinin %40 olması, sisteme giren verinin, bu oranda ve rastgele olarak veri setinden çıkarılarak, tahmin edilecek olan veri olarak belirlenmesi demektir. Bu yüzde %40’lık verinin aslında tahmin edilecek özelliği bilindiği için gerçek değeri ve tahmin değeri başarı veya doğruluk oranının hesaplanmasında kullanılır. Bu çalışmada veri setindeki tahmin edilecek verinin özelliği kısa, normal ve uzun olarak belirlenmiştir. Örneğin bir verinin gerçek değeri kısa tahmin değeri kısa, gerçek değeri normal tahmin değeri normal ve gerçek değeri uzun tahmin değeri de uzun ise doğruluk oranı artar, aksi durumlarda doğruluk oranı düşer.

C.1. Başlıca Sınıflandırma Yöntemleri

Model oluşturma sürecinde sınıflandırmada kullanılan bazı yöntemler vardır. Bunlardan başlıcaları k-NN yani k- en yakın komşu, Naive Bayes, Decision Tree, Random Forest gibi yöntemlerdir [15] [16] [4] [17]. Yapay sinir ağları ise geçmişte çok popüler olmuş fakat daha sonra kullanım oranı düşmüştür. Onun yerini çalışma mantığı aynı temele dayanan Derin Öğrenme almıştır.

C.1.1. Derin Öğrenme

Derin öğrenme temelde YSA'nın çalışma mantığı ile aynıdır. 1990'larda YSA'nın veriler üzerindeki doğruluk oranı çok düşüktü. Bunun sebepleri arasında, eğitim için kullanılan veri setlerinin küçüklüğü, bilgisayar işlem gücünün azlığı, kullanılan aktivasyon fonksiyonları gösterilebilir. Ancak günümüzde bu kısıtların aşılmasıyla, Şekil 3'te temel yapısı gösterilen YSA daha fazla katman, nöron ve aktivasyon fonksiyonlarına sahip olarak Derin Öğrenme adıyla popüler hale gelmiştir. Bir sonraki bölümde Derin Öğrenmenin, büyük veri setlerinde diğer VM yöntemlerine olan üstünlüğü açıkça ortaya konulmuştur.



Şekil 3. Yapay Sinir Ağı hücresi

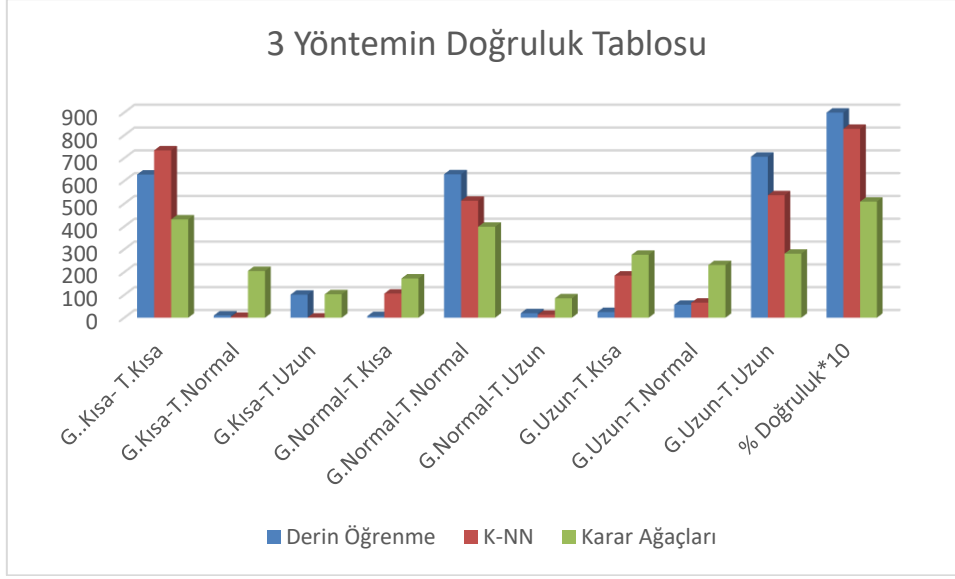
D. BULGULAR

Sistem Şekil 2'deki gibi tasarlandıktan sonra çalıştırılmış ve her bir model için Tablo 2'deki sonuçlar elde edilmiştir. Diğer modellerin sonuçları birleştirilmiş bir şekilde Şekil 4'te sunulmuştur.

Tablo 2. Derin Öğrenme yöntemi sonucu

	true Kısa	true Normal	true Uzun	class precision
pred. Kısa	628	7	25	95.15%
pred. Normal	10	629	57	90.37%
pred. Uzun	101	20	706	85.37%
class recall	84.98%	95.88%	89.59%	

Bu tabloya göre 5458 adet verinin %40'ı olan 2183 adet veri, tahmin ve performans ölçümünde kullanılmıştır. Derin öğrenme yönteminin doğruluk oranı 89,92% bulunmuştur. Buna ek olarak k- En yakın komşu yöntemi için k=5 olarak seçilmiş ve 82,82%'lik bir doğruluk oranı bulunmuştur. Karar ağaçları ise 50,89%'lik doğruluk oranına sahip olmuştur.

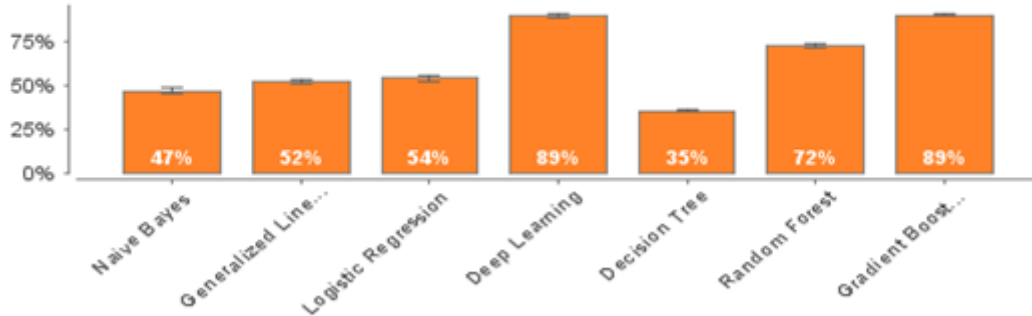


Şekil 4. Karşılaştırma tablosu

Şekil 4’te modellerin verilerin gerçek ve tahmin değerlerine oranları ve grafiğin en solunda ise bu 3 yöntemin ortalama doğruluk oranları verilmiştir. Talebin gerçekte kısa olup tahminde kısa olarak tahmin edilmesinde en başarılı yöntem K-NN yani k-en yakın komşu yöntemi olsa da ortalamada en yüksek başarıyı Derin Öğrenme ortaya koymuştur.

Diğer VM modellerinin ne kadar başarıya ve ne kadar sürede sonuca varabileceği ise RapidMiner programının Auto Model dediğimiz işlem sürecine tabi tutulması sonucu Şekil 5 ve Şekil 6’da doğruluk oranları ve işlem süreleri ile sonuçlanmıştır.

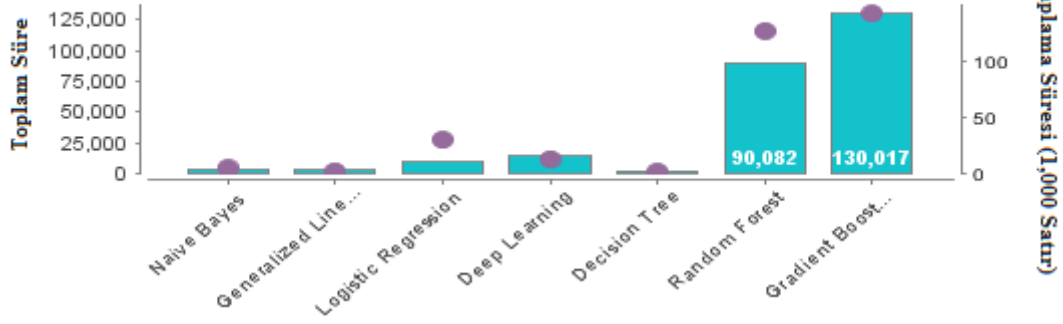
Doğruluk Oranları



Şekil 5. Modellerin doğruluk oranları

Auto model’de en başarılı iki model Gradient Boosted Trees ve Derin Öğrenme olarak öne çıkmıştır.

Çalışma Süreleri (ms)



Şekil 6. Modellerin çalışma süreleri

Auto model’de işlem sürelerine bakıldığında en hızlı sonuç veren yöntem Karar Ağaçları olurken en yavaş yöntem ise Gradien Boosted Trees olmuştur. Karar Ağaçları veri seti üzerinde en düşük doğruluk oranına sahip olduğundan bu çalışma için uygun bir yöntem değildir.

III. SONUC

Performans değerlendirme, personelin işletme amaçlarının gerçekleşmesine yapmış olduğu katkının ölçülmesidir. Diğer bir deyişle personelin görev ve sorumluluklarını ne denli etkin olarak yerine getirip getirmediğinin ölçülmesidir. Yapılan araştırmalar, çalışanların işteki başarıları bakımından büyük farklılıklar gösterdiklerini, aynı kişinin başarısında da zaman içinde önemli değişimler olabildiğini göstermiştir. İyi planlanmış bir performans değerlendirmesi yöneticiye, çalışana ve genel anlamda organizasyona fayda sağlar. Düzenli yapılan performans analizleri çalışana daha verimli çalışmaya yöneltilir.

Bu çalışma kapsamında VM süreci ve kullanılan yöntemler hakkında bilgi verilmiştir. Derin Öğrenme, k-En Yakın Komşu ve Karar Ağacı sınıflandırma yöntemleri veri seti üzerinde uygulanmış ve tahmin oranları karşılaştırılmıştır. En yüksek tahmin Derin Öğrenme yöntemi ile elde edilmiştir. K-En Yakın Komşu ve Derin Öğrenme yöntemleri birbirine yakın tahminde bulunurken Karar Ağacı yönteminin tahmin oranı çok düşük kalmıştır.

Sonuç olarak çalışanların performansları birim ve talep türü özelliklerine göre ölçülmüştür. Bu çalışma mevcut veriler ışığında, birim yöneticisinin karar alma sürecinde, bu çalışmadaki analiz sonuçlarını dikkate alarak, taleplerin daha hızlı cevaplanmasını veya bazı önlemler almasına yardımcı olacak sonuçlar ortaya koymaktadır.

IV. KAYNAKLAR

- [1] M. J. A. Berry ve G. S. Linoff, "Data mining techniques: For marketing, sales, and customer relationship management," 2. baskı, Indianapolis, USA: Wiley, ss. 1-19, 2004.
- [2] S. N. Sumathi, S., Sivanandam, *Introduction to data mining and its application*, Berlin, Germany: Springer, 2006. ss. 325-327.
- [3] Talep Bildirim Sistemi. (2013). <http://www.bidbtalep.duzce.edu.tr/>.

- [4] X. Chen ve F. Wang, "Application of data mining on enterprise human resource performance management," *3rd International Conference on Information Management, Innovation Management and Industrial Engineering*, Kunming, China, 2010, ss. 151–153.
- [5] X. D. Hou, Y. F. Dong, H. P. Liu, ve J. H. Gu, "Application of fuzzy data mining in staff performance assessment," *Proc. Sixth Int. Conf. Mach. Learn. Cybern. ICMLC 2007*, Hong Kong, China, 2007, ss. 835–838.
- [6] Y. Kurniawan ve E. Halim, "Use data warehouse and data mining to predict student academic performance in schools: A case study (perspective application and benefits)," *IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*, Kuta, Indonesia, 2013, ss. 98–103.
- [7] M. Ağaoglu, "Predicting instructor performance using data mining techniques in higher education," *IEEE Access*, c. 4, ss. 2379–2387, 2016.
- [8] K. Yaralıoğlu, "Performans değerlendirmede analitik hiyerarşi proses," *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, c. 1, s. 16, ss. 129–142, 2001.
- [9] P. T. Costa, R. R. McCrae, ve G. G. Kay, "Persons, places, and personality: career assessment using the revised NEO personality inventory," *Journal of Career Assessment*, c. 3, s. 2, ss. 123–139, 1995.
- [10] A. Yelboğa, "İnsan kaynakları yönetiminde performans değerlendirilmesi için geliştirilen bir ölçeğin psikometrik özelliklerinin incelenmesi," Yüksek lisans tezi, Eğitim Bilimleri, Eğitim Bilimleri Enstitüsü, Ankara Üniversitesi, Ankara, Türkiye, 2003.
- [11] A. Yelboğa, "Kişilik özellikleri ve iş performansı arasındaki ilişkinin incelenmesi," *"İş, Güç" Endüstri İlişkileri ve İnsan Kaynakları Dergisi*, c. 8, s. 2, ss. 196–217, 2006.
- [12] "Microsoft-Structured Query Language (2014). Microsoft.
- [13] Y. Özkan, *Veri madenciliği yöntemleri*, 2. Baskı. İstanbul, Türkiye: Papatya Yayıncılık Eğitim, 2013, ss. 37.
- [14] RapidMiner Studio 9.3. (2019). GmbH.
- [15] H. Küçük, C. Tepe, ve İ. Emiroğlu, "K-en yakın komşu algoritması ve destek vektör makinesi yöntemleri ile EMG işaretlerinin sınıflandırılması", ss. 1–4, 2013.
- [16] P. Bermejo, J. A. Gámez, ve J. M. Puerta, "Improving the performance of Naive Bayes multinomial in e-mail foldering by introducing distribution-based balance of datasets," *Expert Syst. Appl.*, c. 38, s. 3, ss. 2072–2080, 2011.
- [17] L. Breiman, *Random Forests*, Netherlands: Kluwer Academic, 2001, ss. 5-32.