

VERİ MADENCİLİĞİ YÖNTEMLERİ İLE İŞ YAŞAM DENGESİNDE YIPRANMA DURUMU TAHMİNİ¹

Dr. Öğr. Üyesi Ufuk ÇELİK * 

ÖZET

İş yaşam dengesinde çalışanların yıpranma durumunun tahmini, şirketlerde insan kaynakları departmanı için çalışan performansını anlamaya yönelik önemli bir göstergedir. Bu tür göstergeler için klasik istatistik analizleri yerine veri madenciliği ile tahminleme yapmak daha efektif sonuçlar verecektir. Bu çalışmada iş hayatında yıpranma ve dolayısıyla işi bırakma durumunun tahmini veri madenciliği yöntemlerinden karar ağacı ve destek vektör makinesi yöntemleri ile gerçekleştirilerek performans sonuçları karşılaştırılmıştır. Analiz için örnek veri seti IBM şirketi Watson Analytics programı kapsamında sunulan bir veritabanından alınmıştır. Veri seti 1470 adet çalışanın 35 farklı öz niteliği içermektedir. Çalışmada yıpranmayı etkileyen faktörler belirlenmiş ve karar ağacı ile destek vektör makinesi yöntemlerinin tahmin performansında sırasıyla %84.09 ve %91.36 doğruluk oranları elde edilmiştir.

Anahtar Kelimeler: Veri Madenciliği, İş Analitiği, İşten Ayrılma, Veri Örnekleme

JEL Kodları: E37, D70, J63, C80

ESTIMATION OF EMPLOYEE ATTRITION IN BUSINESS LIFE BALANCE WITH DATA MINING METHODS

ABSTRACT

The estimation of employee attrition in business life balance is an important indicator to understand the employee performance for human resources department in companies. For such indicators, rather than classical statistical analysis, estimating with data mining will provide more effective results. In this study, the prediction of attrition and hence the turnover situations were estimated with decision tree and support vector machine methods from data mining and performance result were compared. The sample dataset for analysis was taken from a database provided by the IBM company in scope of Watson Analytics programme. The data set contains 35 different attributes of 1470 employees. Factors affecting attrition were determined in the study and the estimation performance of

¹ Bu makale, Bandırma Onyedil Eylül Üniversitesi tarafından düzenlenen IV. Uluslararası Uygulamalı Ekonomi ve Finans Konferansı (ICOAEF'18) kapsamında sözlü olarak sunulan "Alt-Örnekleme ve Üst-Örnekleme Yöntemleri ile Çalışanların İşten Ayrılma Tahmini" isimli bildirinin genişletilmiş bir çalışmasıdır.

* Bandırma Onyedil Eylül Üniversitesi, Ö. S. Uygulamalı Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü Bandırma-Balıkesir/Türkiye e-mail: ucelik@bandirma.edu.tr

the decision tree and support vector machine methods were obtained for accuracy rates of 84.09% and 91.36% respectively.

Keywords: *Data Mining, Business Analytics, Employee Turnover, Data Sampling*

JEL Codes: *E37, D70, J63, C80.*

1. GİRİŞ

İş yaşam dengesinde çalışanların memnuniyetsizliği birçok açıdan olumsuz etkiler yaratmaktadır (Korkmaz ve Erdoğan, 2014). Bu bağlamda çalışan memnuniyetinin ölçülmesi şirketler için gerekli bir analizdir. Çalışan memnuniyetinin zayıf olması şirketi terk etmesi için başlıca faktördür. Personelin işyerine bağlılık durumlarının temelinde örgütsel adalet algısı yer almaktadır. Aşırı ve az iş yükü, yetersiz ücret, çalışma saatleri düzensizliği, fiziksel koşullar veya liyakat gibi durumlar bu algıyı etkileyen faktörler olarak sayılabilir (Black ve Porter, 2000). Çalışan memnuniyeti ile müşteri memnuniyeti arasında da pozitif bir ilişki olduğu aşikârdır (Chi ve Gursoy, 2009). Memnuniyetsizliği gidermek adına öncelikle uygun analizlerin yapıp sebeplerin belirlenmesi gerekir (Altıntaş, 2006; Mobley, 1977, 1982). Ayrıca personelin yıpranmaya bağlı olarak işi bırakma ihtimalleri de doğru şekilde analiz edilerek muhtemel bir personel kaybından önce gerekli düzenlemeler yapılmalıdır.

İnsan kaynakları departmanı tarafından çalışan memnuniyeti ve dolayısıyla işten ayrılma ihtimalinin belirlenmesi adına öncelikle doğru verilerin alınması elzemdir (Ranjan, Goyal ve Ahson, 2008). Bu verilerin, veri madenciliği yöntemleri ile analizinde farklı kurumlara göre sonuçlar değişebilecektir. İşten ayrılma tahmininde kullanılacak yöntemler konusunda farklı analizlerin incelenmesi ile konuya etkili bir yaklaşım sağlanacaktır (Nagadevara, Srinivasan ve Valk, 2008).

2. VERİ

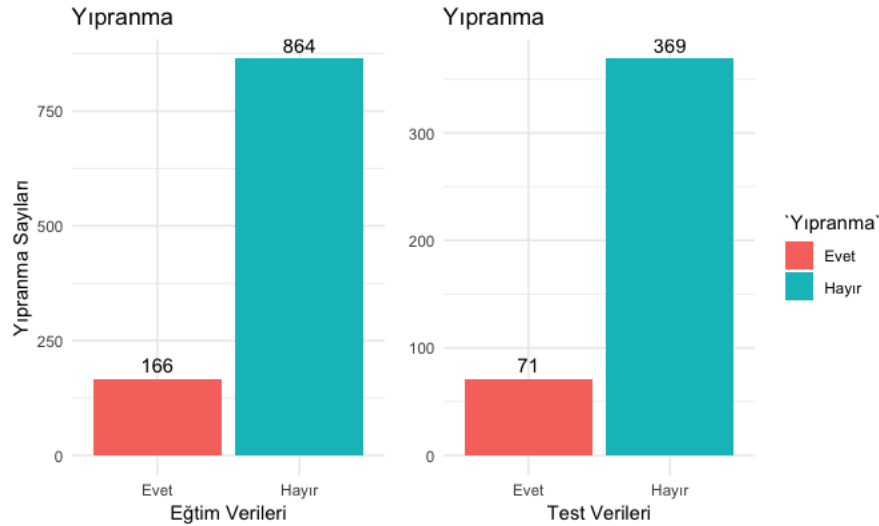
Çalışmada kullanılan veriler IBM Şirketinin Watson Analytics sisteminde tanıtılan çalışanların yıpranma durumu için hazırlanan örnek veritabanından alınmıştır (Stacker, 2015). Veriler içinde aynı değerlere sahip “Standart Saatler”, “18 Yaş Üstü” ve “Çalışan Sayısı” ile gereksiz olan “Çalışan Numarası” değişkenleri temizlenmiştir. İstatistik detaylı özeti Ek 1’ile gösterilen veriler 1470 adet personelin 30 farklı değişkeni ve 1 adet “Yıpranma” durumu değişkenine göre “Evet” veya “Hayır” şeklinde sınıflandırılmıştır. Verilerde 22 adet sayısal ve 9 adet kategorik değişken bulunmaktadır. Verilerin R programlama ortamına alınması ve düzenlenmesi kodları Şekil 1 ile gösterilmiştir.

Şekil 1. Verilerin Düzenlemesi R Kodları

```
1 library(data.table)
2 library(caret)
3 library(rpart.plot)
4 library(e1071)
5 library(ROCR)
6
7 veriseti <- fread('myData.csv', sep = ';', header = TRUE, stringsAsFactors = TRUE)
8 veriseti$ÇalışanNumarası <- NULL
9 veriseti$`18YaşÜstü` <- NULL
10 veriseti$StandartSaatler <- NULL
11 veriseti$ÇalışanSayısı <- NULL
12
13 set.seed(2025)
14 VeriOrnekleme <- createDataPartition(y=veriseti$Yıpranma, p=0.7, list=FALSE)
15 EğitimVerileri <- veriseti[VeriOrnekleme, ]
16 TestVerileri <- veriseti[-VeriOrnekleme, ]
```

Şekil 2 ile gösterilen %70 eğitim ve %30 test verileri şeklinde ayrılan veri setinde test verilerine bakıldığında 440 veriden 369 tanesinde Yıpranma için hayır ve 71 tanesinde evet olduğu anlaşılmaktadır. Grafik için R paketlerinden “ggplot” özelliği kullanılmıştır (Ginestet, 2011; Wickham, 2016).

Şekil 2. Eğitim ve Test Verileri Yıpranma Sayıları



3. YÖNTEM

Çalışma için analizler destek vektör makinesi ve karar ağacı yöntemleri kullanılarak R programlama dili ile gerçekleştirilmiştir (Jed Wing vd., 2018; Meyer, Dimitriadou, Hornik, Weingessel ve Leisch, 2018; Sing, Sander, Beerenwinkel ve Lengauer, 2005; Therneau ve Atkinson, 2018). Açık kaynak kodlu betik bir dil olan R kayıtları vektör, liste, veri çerçevesi gibi özel veri yapıları sayesinde ön tanımlı fonksiyonlar ile veriye bellek üzerinden doğrudan erişmek yerine sembol ve referanslarla kontrol eder (Ihaka ve Gentleman, 1996; Özdemir vd., 2018; “The R Project for Statistical Computing”, 01.12.2018). Fonksiyonlar R-CRAN web sitesi üzerindeki paketlerden çağırılmaktadır (“The Comprehensive R Archive Network”, 01.12.2018).

3.2. Karar Ağacı Algoritması

Denetimli bir öğrenme algoritması olan karar ağacı, verileri özniteliklerin sınıf niteliği ayrımında veriler üzerindeki yaygın değerine göre oluşturduğu kurallarla kategorize eder. (Cover ve Thomas, 2006; Quinlan, 1999). Özniteliklerin veriler üzerindeki belirleyici yaygınlığını hesaplamak için farklı fonksiyonlar mevcuttur. Bu çalışmada bilgi kazancı formülüne göre özniteliklerin veriyi ne kadar temsil ettikleri hesaplanarak kurallar oluşturulmuştur (Therneau ve Atkinson, 2018).

Bir veri setine ait bilgi kazancını hesaplamasında S , C_i sınıfından $i = (1, \dots, n)$ olmak üzere s_i veri içeriyorsa herhangi bir veriyi sınıflandırmak için gereken bilgi Denklem 2 ile hesaplanır.

$$Bilgi(s_1, s_2, \dots, s_n) = -\sum_{i=1}^n \frac{s_i}{S} \left(\log_2 \left(\frac{s_i}{S} \right) \right) \quad (2)$$

Bir A değişkeninin (a_1, a_2, \dots, a_v) değerleri ile entropisini bulmak için ise Denklem 3 kullanılır.

$$Entropi(A) = \sum_{j=1}^v \left(\frac{s_{1j} + \dots + s_{nj}}{S} \right) \cdot Bilgi(s_1, s_2, \dots, s_n) \quad (3)$$

A değişkenini kullanarak ağacın dallanmasıyla elde edilen bilgi kazancı Denklem 4 ile gösterilmiştir.

$$Kazanç(A) = Bilgi(s_1, s_2, \dots, s_n) - Entropi(A) \quad (4)$$

Her bir veri özniteliği için hesaplanan kazanç değerlerinden en yüksek olanı alınarak ağaç oluşturulmaya başlanır. Her bir dal ayrımı için kalan veriler için tekrar kazanç hesabı yapılır ve en yüksek olana göre ayrıma devam edilir. İşlem tüm verilerin sınıflandırılması ile sona erer.

Bilgi kazancı formülüne göre hesaplanan karar ağacı algoritması için R dilinde yapılan kodlama Şekil 3 ile gösterilmiştir.

Şekil 3. Karar Ağacı R Dili Kodlaması

```

1 # Karar Ağacı Algoritması 10 katlı Çapraz Doğrulama ile kontrolü
2 ka.kontrol <- trainControl(method = "repeatedcv", number = 10, repeats = 3)
3 # en uygun modelin seçimi
4 ka.ayar <- train(Yıpranma ~., data = EgitimVerileri, method = "rpart",
5               parms = list(split = "information"),
6               trControl=ka.kontrol,
7               tuneLength = 10)
8 # model gösterimi
9 ka.ayar
10 # en uygun modele göre karar ağacı şeması oluşturulması
11 prp(ka.ayar$finalModel, box.palette = "auto", varlen = 0)
12
13 # test verilerinin tahminlemesi karşılaştırma matrisi gösterimi
14 ka.tahmin.uygunluk <- predict(ka.ayar, newdata = TestVerileri)
15 confusionMatrix(table(ka.tahmin.uygunluk, TestVerileri$Yıpranma))

```

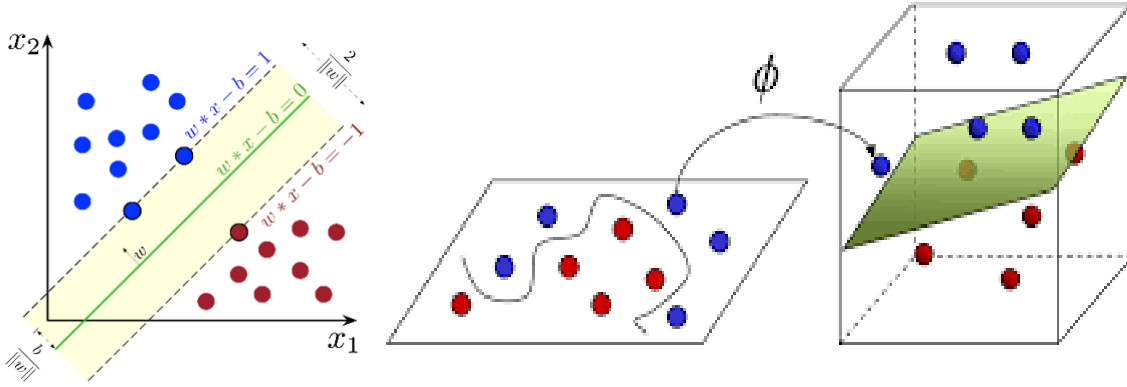
3.2. Destek Vektör Makinesi Algoritması

Veriler içinde herhangi bir değerden en uzak iki sınıf arasındaki sınırı çizen vektörden dolayı bu ismi almıştır (Ülgen, 2017). Şekil 4 ile gösterilen ve Denklem 5 ile hesaplanan bu vektör iki sınıfa ait veriyi en uygun şekilde ayırmak yani sınıflandırmak için kullanılır (Cortes ve Vapnik, 1995).

$$(\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n) \xrightarrow{\text{eğitim seti için}} \vec{w} \cdot \vec{x} - b = 0 \quad (5)$$

Burada y_i 1 veya -1 şeklinde \vec{x}_i noktasının ait olduğu sınıfı belirtirken, \vec{x} noktaların oluşturduğu veri setini temsil eder. $\vec{w} \cdot \vec{x} - b = 1$ ise bir sınıf değerinin üstünde olduğunu, $\vec{w} \cdot \vec{x} - b = -1$ ise diğer sınıf değerinin altında olduğunu belirtir. Sıfıra eşit olması durumu ayırım vektörünü belirtir. Sınıf verileri arasındaki uzaklığı maksimum yapmak için $\|w\|$ değeri minimize edilmelidir.

Şekil 4. Destek Vektör Makinesi Sınıflandırması



Kaynak: (Jordan, 2017)

Vektörün veriyi ayırma durumuna göre doğrusal veya doğrusal olmayan şeklinde iki tür algoritma vardır (Akpınar, 2014). Doğrusal destek vektör makinesi, ayırım yapan doğrunun eğim hesabında bir karar fonksiyonu kullanır (Osuna, Freund ve Girosi, 1997). Doğrusal olmayan ayırmalarda ise polinom (çok terimli) ve gausyen radyal tabanlı çekirdek bir fonksiyonla hesaplama yapılır (Jed Wing vd., 2018).

Destek vektör makinesi için svm() fonksiyonu kullanılan R dili kodlaması Şekil 5 ile gösterilmiştir.

Şekil 5. Destek Vektör Makinesi R Dili Kodlaması

```
1 # Destek Vektör Makinesi
2 ind <- sample(2, nrow(veriseti), replace = TRUE, prob=c(0.7, 0.3))
3
4 # Destek Vektör Makinesi gama ve cost değerleri performans ayarlaması
5 dvm.ayar <- tune(svm, Yıpranma~., data = veriseti[ind == 1,],
6               ranges = list(gamma = 2^(-8:1), cost = 2^(0:4)),
7               tunecontrol = tune.control(sampling = "fix"))
8 # dvm ayarlama sonuçları gösterimi
9 dvm.ayar
10
11 # Eğer ayarlama sonucu parametre marjini ile örtüşüyorsa (ör, gamma = 2^-8),
12 # parametreler genişletilebilir
13 # dvm.ayar sonucundan gelen gama ve cost parameterleri kullanılır.
14 dvm.model <- svm(Yıpranma~., data=veriseti[ind==1,],
15               cost=8, gamma=0.0078125, probability=TRUE)
16 dvm.olasilik <- predict(dvm.model, type="prob",
17                       newdata=veriseti[ind==2,], probability=TRUE)
18
19 dvm.model.tahmin <- predict(dvm.model, newdata=TestVerileri)
20 confusionMatrix(table(dvm.model.tahmin, TestVerileri$Yıpranma))
```

3.3. ROC analizi

Bir sınıflandırma şeklinde yapılan bir tahminlemenin performans değerlendirmesi alıcı işlem karakteristiği (Receiver Operating Characteristics) ROC analizi yapılarak sonuçların hassasiyet ve kesinlik ölçütleri hesaplanır (Gribskov ve Robinson, 1996).

Bir sınıflama sonucuna göre ROC analizi için aşağıdaki hesaplamalar incelenir.

- DP (doğru pozitif), gerçekte evet olan ve algoritmanın da evet bulduğu yıpranma verileri
- DN (doğru negatif), gerçekte hayır olan ve algoritmanın hayır bulduğu yıpranma verileri
- YP (yanlış pozitif), gerçekte evet olan ama algoritmanın hayır bulduğu yıpranma verileri
- YN (yanlış negatif), gerçekte hayır olan ama algoritmanın evet bulduğu yıpranma verileri
- Hassasiyet DP/(DP+YN) oranı
- Kesinlik DN/(DN+YP) oranı

ROC analizinde, hassasiyeti tespit etme ve kesinliği eleme kabiliyeti arasındaki dengeye bakılarak tahminleme için kullanılan yöntemin performansı anlaşılır. Kesinlik ve hassasiyet arasındaki dengeyi değerlendirmek için ROC eğrisi ve altında kalan alan kullanılır. ROC puanı 1'e yaklaştıkça pozitifler daha iyi bir şekilde negatiflerden ayrılır (Sing vd., 2005).

Şekil 6. ROC Analizi R Dili Kodlaması

```

1 # Karar Ağacı ROC analizi
2 ka.tahmin.olasilik = predict(ka.ayar, newdata = TestVerileri, type = "prob")[,2]
3 # sınıflandırma olasılıklarına göre tahminleme
4 ka.tahmin = prediction(ka.tahmin.olasilik, TestVerileri$Yıpranma)
5 # Karar Ağacı ROC eğrisi çizimi
6 plot(performance(ka.tahmin, "tpr", "fpr"), col="red",
7 xlab="Yanlış Pozitif Oranı", ylab="Doğru Pozitif Oranı")
8 abline(0, 1, lty = 2)
9
10 # Destek Vektör Makinesi ROC analizi
11 dvm.olasilik.rocr <- prediction(attr(dvm.olasilik, "probabilities")[,2],
12 veriseti[ind==2, 'Yıpranma'])
13 dvm.olasilik.performans <- performance(dvm.olasilik.rocr, "tpr", "fpr")
14 plot(dvm.olasilik.performans, col="blue", add=TRUE)
15
16 # Etiketlerin gösterimi
17 legend(0.5, 0.3, c('Karar Ağacı', 'Destek Vektör Makinesi'),
18 col=c("red", "blue"), lty=4:2)

```

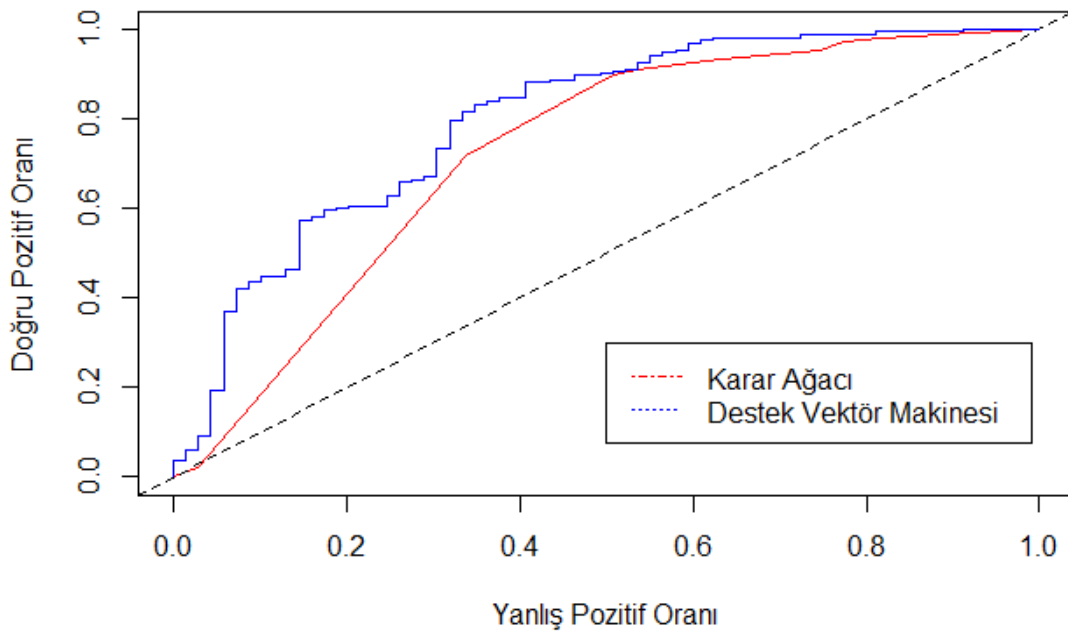
4. BULGULAR

Karar ağacı ve destek vektör makinesi yöntemleri sonucu elde edilen sınıflandırma başarısı, yapılan ROC analizi sonuçlarına göre elde edilen performans değerleri Tablo 1 ile ve çizilen ROC eğrisi Şekil 7 ile gösterilmiştir.

Tablo 1. ROC Analizi Sonuçları

Algoritmalar	DP	DN	YP	YN	Hassasiyet	Kesinlik	ROC puanı
Karar Ağacı	18	352	17	53	0.25352	0.95393	0.8409
Destek Vektör Makinesi	38	364	5	33	0.53521	0.98645	0.9136

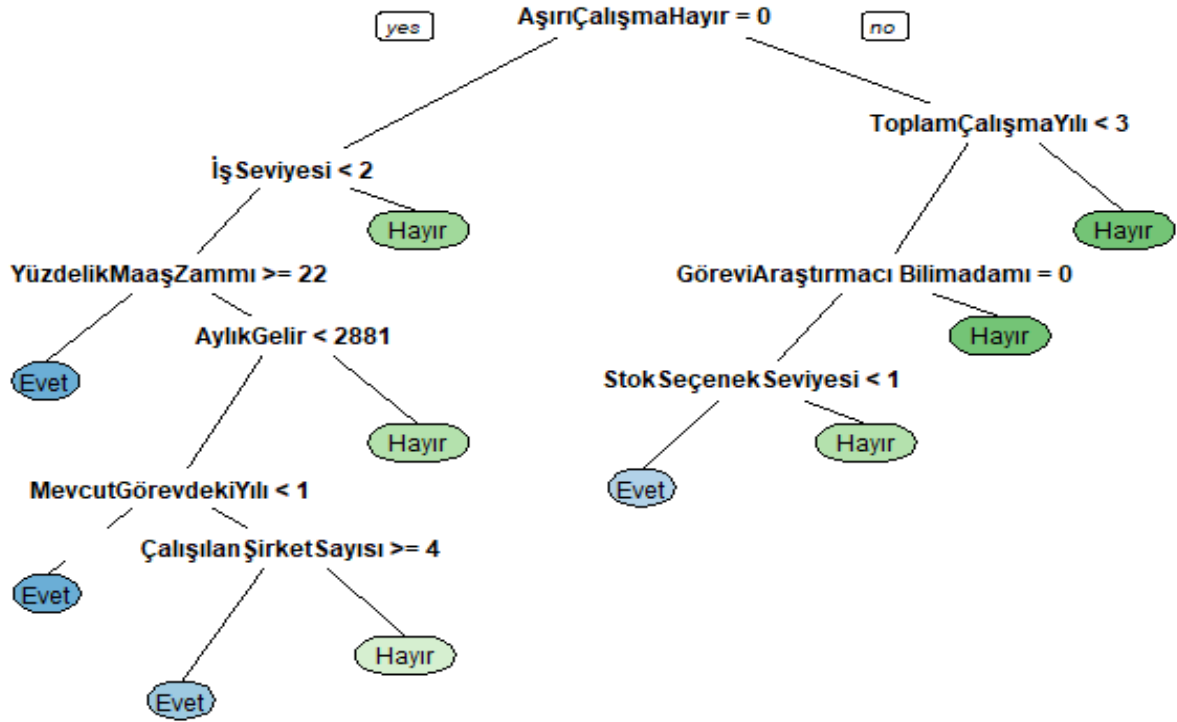
ROC analizi sonuçlarına göre destek vektör makinesi algoritması daha başarılı bir tahminleme sonucu sergilemektedir (Robin vd., 2011; Sing vd., 2005).

Şekil 7. ROC Eğrisi

4.1. Karar Ağacı Analiz Sonuçları

Karar ağacı analizi için öncelikle rastgele sayı üretme aracı çekirdek (seed) değeri ayarlanır. Bu değer algoritmanın her çalışmasında farklı sonuçlar üretmesini engeller. Eğitim için tüm veri setinin %70'i (1029 adet) ve test için %30'u (441 adet) alınır. %84.09 doğruluk oranı çıkaran karar ağacı Şekil 8 ile gösterilmiştir (Jacobucci, 2018).

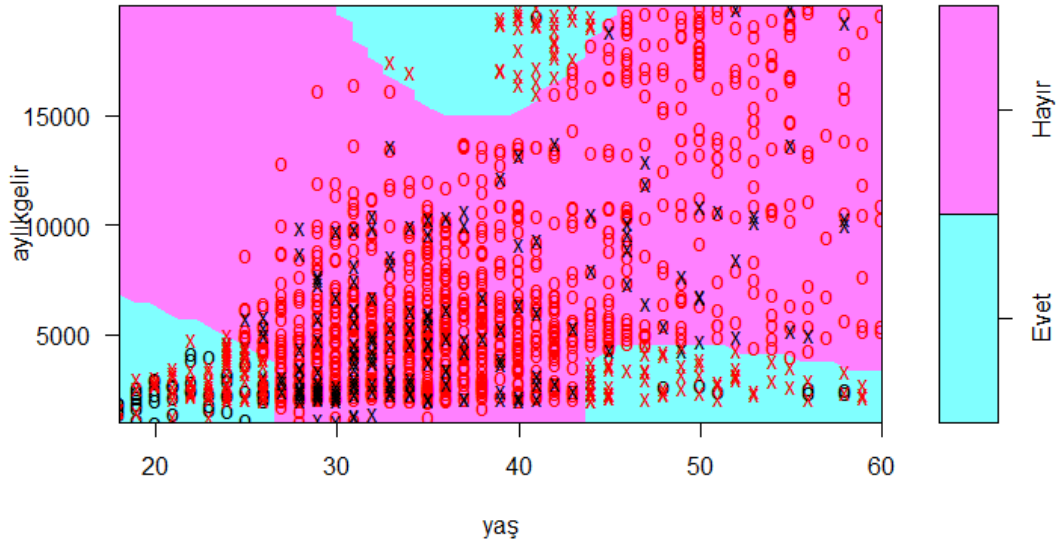
Şekil 8. Karar Ağacı Şeması



4.2. Destek Vektör Makineleri Analiz Sonuçları

Karar destek vektörü sonucu test verilerinde %91.36 doğruluk oranı elde edilmiştir. Örnek bir gösterim için yaşa göre aylık gelir durumunun yıpranmaya etkisi incelendiğinde Şekil 9 ile destek vektör makinesi sınıflaması görülmektedir (Torgo, 2011).

Şekil 9. Yaş ve Aylık Gelir Durumuna Göre Yıpranma Sınıflandırması



5. SONUÇ VE TARTIŞMA

Bu çalışmada elde edilen sonuçlara göre karar ağacı ile yapılan sınıflamada %84.09 doğruluk oranı bulunurken, destek vektör makinesi ile yapılan sınıflamada %91.36 doğruluk oranı görülmüştür.

Karar ağacı şemasına göre yıpranmayı belirleyici faktör olarak “Aşırı Çalışma” dikkat çekmektedir. Aşırı çalışma durumunu takip eden diğer faktörlere bakıldığında ise sırasıyla “İş Seviyesi” 2’den küçük olanlar ve “Toplam Çalışma Yılı” 3’den küçük olanlar izlemektedir. “Yüzdelik Maaş Zammı”, “Aylık Gelir” ile “Görev Türü” diğer önemli faktörler olarak karşımıza çıkmaktadır.

ROC analizi sonuçları incelendiğinde her iki yöntem içinde hassasiyet değerleri düşük çıkarken, kesinlik değeri yüksek olarak gözlemlenmiştir. Destek vektör makinesi 38 adet doğru pozitif değeri bulurken karar ağacı 18 tane doğru pozitif değeri ile destek vektör makinesine göre daha düşük bir performans sergilemiştir. Yanlış pozitif ve doğru negatif değerlerinde iki yöntemde birbirinden farklı sonuçlar çıkardığı anlaşılmaktadır. Doğru negatif değerlerinde ise her iki yöntemde birbirine yakındır.

İş yaşam dengesi durumuna göre yıpranma faktörleri ele alındığında bu çalışmada yapılan analizlerin şirketlere personel memnuniyeti açısından farklı bir bakış açısı kazandıracakı aşikârdır. İnsan kaynakları departmanı hem çalışanları hem de müşterileri için bu tür analizlerle memnuniyet durumunu daha iyi bir şekilde ortaya koyabilecektir. İleriye dönük çalışmalarda bu çalışmadaki analizlere ek olarak veri madenciliğinin diğer yöntemlerinin incelenmesi de daha efektif sonuçlar verecektir.

REFERENCES

Akpınar, P. D. H. (2014) "DATA Veri Madenciliği Veri Analizi", İstanbul, Papatya Bilim Yayınevi

Altıntaş, F. Ç. (2006) "Bireysel Değerlerin Örgütsel Adalet ve Sonuçları İlişkisinde Yönlendirici

Etkisi: Akademik Personel Üzerinde Bir Analiz", Dokuz Eylül Üniversitesi İşletme Fakültesi

Dergisi, 7(2): 19–40.

Black, J. S. ve Porter, L. W. (2000) "Management: Meeting New Challenges", USA, Prentice Hall

Chi, C. G. ve Gursoy, D. (2009) "Employee satisfaction, customer satisfaction, and financial performance: An empirical examination", *International Journal of Hospitality Management*, 28(2): 245–253.

Cortes, C., ve Vapnik, V. (1995) "Support-vector networks", *Machine Learning*, 20(3): 273–297.

Cover, T. M., ve Thomas, J. A. (2006) "Elements of Information Theory 2nd Edition", USA, Willey.

Ginestet, C. (2011) "ggplot2: Elegant Graphics for Data Analysis", *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(1): 245–246.

Gribskov, M., ve Robinson, N. L. (1996) "Use of receiver operating characteristic (ROC) analysis to evaluate sequence matching", *Computers & Chemistry*, 20(1): 25–33.

Ihaka, R., ve Gentleman, R. (1996) "R: a language for data analysis and graphics", *Journal of Computational and Graphical Statistics*, 5(3): 299–314.

Jacobucci, R. (2018) "dtree: Decision Trees R package version 0.4.2.", <https://cran.r-project.org/package=dtree>, (01.11.2018)

Jed Wing, M. K. C., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., vd. (2018) "caret: Classification and Regression Training", <https://cran.r-project.org/web/packages/caret>, (01.12.2018)

Jordan, J. (2017) "Support vector machines. Jeremy Jordan Thoughts, ideas, and new things I've learned.", <https://www.jeremyjordan.me/support-vector-machines/>, (01.11.2018)

Korkmaz, O. ve Erdoğan, E. (2014) "İş Yaşam Dengesinin Örgütsel Bağlılık ve Çalışan Memnuniyetine Etkisi", *Ege Akademik Bakış Dergisi*, 14(4): 541–557.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. ve Leisch, F. (2018) "e1071: Misc Functions of the Department of Statistics", *Probability Theory Group (Formerly: E1071)*, TU Wien.

Mobley, W. H. (1977) "Intermediate linkages in the relationship between job satisfaction and employee turnover", *Journal of Applied Psychology*, 62(2): 237–240.

Mobley, W. H. (1982) "Employee Turnover, Causes, Consequences, and Control.", USA, Addison-Wesley series on managing human resources. Addison-Wesley.

Nagadevara, V., Srinivasan, V. ve Valk, R. (2008) "Establishing a Link Between Employee Turnover and Withdrawal Behaviours: Application of Data Mining Techniques.", *Research & Practice in Human Resource Management*, 16(2): 81–99.

Osuna, E., Freund, R. ve Girosi, F. (1997) "Support Vector Machines: Training and Applications", *Yönetim ve Ekonomi Araştırmaları Dergisi / Journal of Management and Economics Research*

Cambridge, MA, USA: Massachusetts Institute of Technology.

Özdemir, A. F., Yıldıztepe, E. ve Binar, M. (2018) "İstatistiksel Yazılım Geliştirme Ortamı: R",
Akademik Bilişim'10 – XII. Akademik Bilişim Konferansı, Muğla

Quinlan, J. R. (1999) "Simplifying decision trees", *International Journal of Human-Computer Studies*,
51(2): 497–510.

Ranjan, J., Goyal, D. P. ve Ahson, S. I. (2008). "Data mining techniques for better decisions in human
resource management systems", *International Journal of Business Information Systems*, 3(5):
464–481.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C. ve Müller, M. (2011)
"pROC: an open-source package for R and S+ to analyze and compare ROC curves", *BMC
Bioinformatics*, 12(1): 77.

Sing, T., Sander, O., Beerenwinkel, N. ve Lengauer, T. (2005) "ROCR: visualizing classifier
performance in R", *Bioinformatics*, 21(20): 7881.

Stacker, M. (2015) "SAMPLE DATA: HR Employee Attrition and Performance",
<https://www.ibm.com/communities/analytics/watson-analytics-blog/hr-employee-attrition/>,
01.11.2018

"The Comprehensive R Archive Network. R-Project.", <https://cran.r-project.org/> 01.11.2018

"The R Project for Statistical Computing. R-Project.", <https://www.r-project.org/>, 01.11.2018

Therneau, T. ve Atkinson, B. (2018) "rpart: Recursive Partitioning and Regression Trees".

Torgo, L. (2011) "Data mining with R: learning with case studies", USA, Chapman and Hall/CRC.

Ülgen, E. K. (2017) "Makine Öğrenimi Bölüm-4 (Destek Vektör Makineleri)",
[https://medium.com/@k.ulgen90/makine-ogrenimi-bolum-4-destek-vektor-makineleri-
2f8010824054](https://medium.com/@k.ulgen90/makine-ogrenimi-bolum-4-destek-vektor-makineleri-2f8010824054), 01.11.2018

Wickham, H. (2016) "ggplot2: Elegant Graphics for Data Analysis", New York, Springer-Verlag.

Ek 1: Veri Seti Özeti

No	Değişken	İstatistikler / Değerler	Frekanslar (% Geçerlilik)	Grafik
1	Yaş [tamsayı]	ort (sd) : 36.92 (9.14) min < med < mak 18 < 36 < 60 IQR (CV) : 13 (0.25)	43 farklı değer	
2	Seyahat [faktör]	1. Nadiren Seyahat 2. Seyahat Yok 3. Sık Seyahat	1043 (71.0%) 150 (10.2%) 277 (18.8%)	
3	Günlük Oran [tamsayı]	ort (sd) : 802.49 (403.51) min < med < mak 102 < 802 < 1499 IQR (CV) : 692 (0.5)	886 farklı değer	
4	Departman [faktör]	1. ARGE 2. Satış 3. İnsan Kaynakları	961 (65.4%) 446 (30.3%) 63 (4.3%)	
5	Evden Uzaklık [tamsayı]	ort (sd) : 9.19 (8.11) min < med < mak 1 < 7 < 29 IQR (CV) : 12 (0.88)	29 farklı değer	
6	Eğitim [tamsayı]	ort (sd) : 2.91 (1.02) min < med < mak 1 < 3 < 5 IQR (CV) : 2 (0.35)	1: 170 (11.6%) 2: 282 (19.2%) 3: 572 (38.9%) 4: 398 (27.1%) 5: 48 (3.3%)	
7	Eğitim Alanı [faktör]	1. Diğer 2. Pazarlama 3. Teknik Derece 4. Tıp 5. Yaşam Bilimleri 6. İnsan Kaynakları	82 (5.6%) 159 (10.8%) 132 (9.0%) 464 (31.6%) 606 (41.2%) 27 (1.8%)	
8	Ortam Memnuniyeti [tamsayı]	ort (sd) : 2.72 (1.09) min < med < mak 1 < 3 < 4 IQR (CV) : 2 (0.4)	1: 284 (19.3%) 2: 287 (19.5%) 3: 453 (30.8%) 4: 446 (30.3%)	
9	Cinsiyet [faktör]	1. Bay 2. Bayan	882 (60.0%) 588 (40.0%)	
10	Saatlik Oran [tamsayı]	ort (sd) : 65.89 (20.33) min < med < mak 30 < 66 < 100 IQR (CV) : 35.75 (0.31)	71 farklı değer	
11	İş İlgisi [tamsayı]	ort (sd) : 2.73 (0.71) min < med < mak 1 < 3 < 4 IQR (CV) : 1 (0.26)	1: 83 (5.6%) 2: 375 (25.5%) 3: 868 (59.1%) 4: 144 (9.8%)	
12	İş Seviyesi [tamsayı]	ort (sd) : 2.06 (1.11) min < med < mak 1 < 2 < 5 IQR (CV) : 2 (0.54)	1: 543 (36.9%) 2: 534 (36.3%) 3: 218 (14.8%) 4: 106 (7.2%) 5: 69 (4.7%)	

No	Değişken	İstatistikler / Değerler	Frekanslar (% Geçerlilik)	Grafik
13	Görevi [faktör]	1. Araştırma Yöneticisi 2. Araştırmacı Bilim adamı 3. Laboratuvar Teknisyeni 4. Müdür 5. Satış Temsilcisi 6. Satış Yöneticisi 7. Sağlık Temsilcisi 8. Üretim Yöneticisi 9. İnsan Kaynakları	80 (5.4%) 292 (19.9%) 259 (17.6%) 102 (6.9%) 83 (5.6%) 326 (22.2%) 131 (8.9%) 145 (9.9%) 52 (3.5%)	
14	İş Başarısı [tamsayı]	ort (sd) : 2.73 (1.1) min < med < mak 1 < 3 < 4 IQR (CV) : 2 (0.4)	1: 289 (19.7%) 2: 280 (19.1%) 3: 442 (30.1%) 4: 459 (31.2%)	
15	Medeni Durum [faktör]	1. Bekar 2. Boşanmış 3. Evli	470 (32.0%) 327 (22.2%) 673 (45.8%)	
16	Aylık Gelir [tamsayı]	ort (sd) : 6502.93 (4707.96) min < med < mak 1009 < 4919 < 19999 IQR (CV) : 5468 (0.72)	1349 farklı değer	
17	Aylık Oran [tamsayı]	ort (sd) : 14313.1 (7117.79) min < med < mak 2094 < 14235.5 < 26999 IQR (CV) : 12414.5 (0.5)	1427 farklı değer	
18	Çalışılan Şirket Sayısı [tamsayı]	ort (sd) : 2.69 (2.5) min < med < mak 0 < 2 < 9 IQR (CV) : 3 (0.93)	0: 197 (13.4%) 1: 521 (35.4%) 2: 146 (9.9%) 3: 159 (10.8%) 4: 139 (9.5%) 5: 63 (4.3%) 6: 70 (4.8%) 7: 74 (5.0%) 8: 49 (3.3%) 9: 52 (3.5%)	
19	Aşırı Çalışma [faktör]	1. Evet 2. Hayır	416 (28.3%) 1054 (71.7%)	
20	Yüzdeler Maaş Zammı [tamsayı]	ort (sd) : 15.21 (3.66) min < med < mak 11 < 14 < 25 IQR (CV) : 6 (0.24)	15 farklı değer	
21	Performans Oranı [tamsayı]	ort (sd) : 3.15 (0.36) min < med < mak 3 < 3 < 4 IQR (CV) : 0 (0.11)	3: 1244 (84.6%) 4: 226 (15.4%)	
22	İlişki Memnuniyeti [tamsayı]	ort (sd) : 2.71 (1.08) min < med < mak 1 < 3 < 4 IQR (CV) : 2 (0.4)	1: 276 (18.8%) 2: 303 (20.6%) 3: 459 (31.2%) 4: 432 (29.4%)	

No	Değişken	İstatistikler / Değerler	Frekanslar (% Geçerlilik)	Grafik
23	Stok Seçenek Seviyesi [tamsayı]	ort (sd) : 0.79 (0.85) min < med < mak 0 < 1 < 3 IQR (CV) : 1 (1.07)	0: 631 (42.9%) 1: 596 (40.5%) 2: 158 (10.8%) 3: 85 (5.8%)	
24	Toplam Çalışma Yılı [tamsayı]	ort (sd) : 11.28 (7.78) min < med < mak 0 < 10 < 40 IQR (CV) : 9 (0.69)	40 farklı değer	
25	Geçen Yılın Eğitim Süreleri [tamsayı]	ort (sd) : 2.8 (1.29) min < med < mak 0 < 3 < 6 IQR (CV) : 1 (0.46)	0: 54 (3.7%) 1: 71 (4.8%) 2: 547 (37.2%) 3: 491 (33.4%) 4: 123 (8.4%) 5: 119 (8.1%) 6: 65 (4.4%)	
26	İş Yaşam Dengesi [tamsayı]	ort (sd) : 2.76 (0.71) min < med < mak 1 < 3 < 4 IQR (CV) : 1 (0.26)	1: 80 (5.4%) 2: 344 (23.4%) 3: 893 (60.8%) 4: 153 (10.4%)	
27	Şirketteki Yılı [tamsayı]	ort (sd) : 7.01 (6.13) min < med < mak 0 < 5 < 40 IQR (CV) : 6 (0.87)	37 farklı değer	
28	Mevcut Görevdeki Yılı [tamsayı]	ort (sd) : 4.23 (3.62) min < med < mak 0 < 3 < 18 IQR (CV) 5 (0.86)	19 farklı değer	
29	Son Promosyon Sonrası Yılı [tamsayı]	ort (sd) : 2.19 (3.22) min < med < mak 0 < 1 < 15 IQR (CV) : 3 (1.47)	16 farklı değer	
30	Mevcut Amiri Çalıştığı Yıl [tamsayı]	ort (sd) : 4.12 (3.57) min < med < mak 0 < 3 < 17 IQR (CV) : 5 (0.87)	18 farklı değer	
31	Yıpranma [faktör]	1. Evet 2. Hayır	237 (16.1%) 1233 (83.9%)	

Kaynak: R programlama dili summarytools 0.8.8 (R version 3.4.3) tarafından oluşturulmuştur