





Received: June 27, 2019  
Accepted: October 27, 2019  
Published Online: December 31, 2019

AJ ID: 2018.07.02.MIS.03  
DOI: 10.17093/alphanumeric.583502  
**Research Article**


## Forecasting of University Students' Turkish Language Course Exam Results According to Their Exam Preparation Levels

Emrah Aydemir, Ph.D. \* 

Assist. Prof., Department of Faculty of Engineering and Architecture Ahi Evran University, Kırşehir, Turkey, aydemir.emrah23@gmail.com

Feyzi Kaysi, Ph.D. 

Res. Assist., Vocational School of Technical Sciences Istanbul University Cerrahpasa, Istanbul, Turkey, feyzikaysi@gmail.com

Sevinç Gülseçen, Ph.D. 

Prof., Department of Informatics, Istanbul University, Istanbul, Turkey gulsecen@istanbul.edu.tr

\* Kırşehir Ahi Evran Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bağbaşı Yerleşkesi, Kırşehir, Türkiye

### ABSTRACT

There are many algorithms and software in the field of forecasting. These methods are also used in education. There are studies on forecasting of student achievement in education. Forecasting of academic achievement of university students is important in terms of seeing possible situations. In this study, the achievement of the students towards Turkish Language course was forecasting with data mining methods. 160 student data, in a state university, were included in the study. For the data obtained, prediction models developed by DecisionStump, RandomTree, RandomForest, REPTree and M5P methods were created and compared with each other. 10-fold cross-validation method was used in the separation of data for training and test. In models, it will affect the student's passing grade; program, type of OSS entrance, OSS entrance score, OSS entrance rankings, the previous semester grade point average, midterm exam grade, studying status, the current study and how many points are expected from the exam, how the exam passed and final exam score was taken into consideration. Among the models, it was seen that the model established with best results with 10.16 mean absolute error and 0.72 correlation coefficient. As a result of the study, it is thought that students can take precautions by predicting the passing grade.

### Keywords:

Forecasting, University Students, Turkish Language, Data Mining

## Üniversite Öğrencilerinin Türk Dili Dersi Sınav Sonuçlarının Sınava Hazırlık Düzeylerine Göre Tahminlenmesi

### ÖZ

Günümüzde tahminleme alanında çok sayıda algoritma ve yazılım bulunmaktadır. Buna imkân tanıyan yöntemler eğitim alanında da kullanılmaktadır. Eğitimde öğrenci başarısı üzerine çalışmalar yer almaktadır. Üniversite öğrencilerinin akademik başarılarının tahminlenmesi olası durumların görülebilmesi açısından önem arz etmektedir. Bu çalışmada öğrencilerin Türk Dili dersine yönelik başarıları veri madenciliği yöntemleriyle tahmin edilmiştir. Çalışmaya Türkiye'deki bir devlet üniversitesinde Türk Dili dersini alan 160 öğrenci verisi dahil edilmiştir. Elde edilen veriler için DecisionStump, RandomTree, RandomForest, REPTree ve M5P yöntemleri ile geliştirilen tahmin modelleri oluşturulmuş ve birbirleriyle karşılaştırılmıştır. Verilerin eğitim ve test olarak ayrıştırılmasında 10-katlı çapraz doğrulama yöntemi kullanılmıştır. Modellerde öğrencinin ders geçme notunu etkileyecek program, ÖSS giriş türü, ÖSS giriş puanı, ÖSS giriş sıralaması, bir önceki dönemin not ortalaması, vize notu, çalışma durumu, mevcut çalışma ile sınavdan kaç puan beklendiği, sınav nasıl geçti ve final sınav puanı dikkate alınmıştır. Sonuç olarak en iyi ortalama mutlak hata 10.16 ve korelasyon katsayısı 0.72 ile tahminler üretildiği görülmüştür. Çalışma sonucunda öğrencilerin ders geçme notunu önceden tahmin ederek önlemler alınabileceği düşünülmektedir.

### Anahtar Kelimeler:

Tahminleme, Üniversite Öğrencileri, Türk Dili, Veri Madenciliği

## 1. Giriş

Günümüzde muazzam düzeyde ifade edilebilecek teknolojik gelişmeler sürekli olarak dijital bilgi miktarını arttırmaktadır. Bununla birlikte bu verilerin işlenmesi zorlaşmaktadır (Fayyad, Piatetsky-Shapiro ve Smyth, 1996). Özellikle elde edilen verilerin işlenmesinde faydalı bilgiye ulaşmak büyük önem taşımaktadır (Bramer, 2013). Bu aşamada ise veri madenciliği kavramı bir çözüm yolu olarak görülebilmektedir. Thuarisingham (2003) veri madenciliğini, büyük miktardaki verinin bilgisayarlar aracılığıyla işlenerek geleceğe yönelik tahminlemelerin yapılması şeklinde ifade etmiştir. Sieber (2008) ise veri madenciliğini, verinin farklı bir bakış açısıyla analiz edilmesi ve bu verilerle faydalı yeni bilgilerin ortaya çıkarılması süreci olarak belirtmiştir. Elde edilen verilerin işlenmesi algoritma ve yazılımlar sayesinde daha kolay hale gelmiştir. Algoritmalar genel olarak tahminleme veya sınıflama teknikleri üzerine kuruludur (Weiss ve Kulikowski, 1991).

Veri madenciliği yöntemlerinin denetimsiz, yarı-denetimli ve denetimli öğrenme yaklaşımı olmak üzere üç türü bulunmaktadır. Denetimli öğrenme yaklaşımında algoritma, etiketleri bilinen bir dizi örnekle çalışır. Bu etiketler, sınıflandırma için nominal değerler veya regresyon durumunda sayısal değerler olabilir. Denetimsiz öğrenme yaklaşımında ise örneklerin etiketleri bilinmemekte ve algoritma, tipik olarak örnekleri, kümeleme görevini karakterize eden öznel değerlerinin benzerliğine göre gruplandırmayı amaçlamaktadır. Yarı-denetimli öğrenme yaklaşımında ise etiketli örneklerin küçük bir alt kümesi mevcut olmasına rağmen bunları çok sayıda etiketlenmemiş örnekle birlikte kullanmayı esas almaktadır (Neelamegam ve Ramaraj, 2013). Karar ağacı bu öğrenme yaklaşımları arasından denetimli öğrenme yaklaşımı içerisinde yer alır. Aynı zamanda karar ağaçları, verilerin belirli özellik değerlerine göre sınıflandırılmasına yarar (Kuyucu, 2012).

Alan yazında öğrenci başarısını tahminlemeye yönelik çalışmalar bulunmaktadır. Bu çalışmalardan bazıları; Aksu (2018)'nin öğrencilerin PISA başarısını tahminlemesi, Aydemir (2019)'in ders geçme notlarının tahmini, Gök (2017)'ün akademik başarının tahmin edilmesi, del Campo-Ávila, Muñoz, Ruiz ve Bueno (2015)'nin öğrenme başarısının tahminlenmesi, Tsiakmaki ve diğ. (2018)'nin öğrenci notlarını tahminlemesi ve Martinez Abad ile López (2017)'in akademik başarıyı etkileyen faktörleri tespit etmesi şeklindedir. Bu çalışma örnekleri çoğaltılabilir.

Üniversite öğrencilerinin mezuniyet kriterleri arasında öğrenim gördükleri program için belirlenen düzeydeki akademik başarının elde edilmesi yer almaktadır. Bu başarının belirlenmesindeki en önemli etken, derslere yönelik yapılan sınavlardan elde edilen sonuçlardır. Bu sonuçların okunması esnasında değerlendirici tarafından yapılabilecek eksik veya hatalı okuma veya sınav sonuçlarının otomasyon sistemine hatalı girilmesi gibi olumsuzluklar olabilmektedir. Bu tür olumsuzluklar öğrencilerin başarı ortalamalarını olumsuz etkileyebilmektedir. Bu olumsuz durumu kontrol etmek veya olası hataların önüne geçebilmek amacıyla hazırlanacak bir tahminleme ile ek bir değerlendirme sağlanabilir. Aynı zamanda öğrencilerin belirli değişkenler göz önünde bulundurulması ilgili derse yönelik notunu tahmin etmesi de sağlanabilir. Bu çalışmanın amacı, üniversite öğrencilerinin bazı kriterlere göre sınav sonuçlarının tahmin edilmesini sağlamaktır. Bu sayede sınav sonuçları ile tahminleme sonuçlarının karşılaştırılması sağlanabilir. Böylece olası olumsuz durumların önüne geçilebilir.

## 2. Yöntem

Bu çalışmada 2018-2019 Eğitim-Öğretim yılının bahar dönemine ait Türk Dili II dersini alan ön lisans öğrencilerinin verileri kullanılmıştır. Veriler aşağıdaki başlıklardan oluşmakta olup, Program, ÖSS Giriş Türü, ÖSS Giriş Puanı, ÖSS Giriş Sıralaması, Bir Önceki Dönemin Not Ortalaması, Vize Notu verileri öğrenci bilgi sisteminden elde edilmiştir. Çalışma Durumu ve Mevcut Çalışma ile Sınavdan Kaç Puan Beklendiği sınav öncesi ve Sınav Nasıl Geçti ise sınav sonrası öğrencilere sorulmuştur.

- Program
- ÖSS Giriş Türü (TYT, YGS4, YGS5, Sınavsız Geçiş, YGS6)
- ÖSS Giriş Puanı
- ÖSS Giriş Sıralaması
- Bir Önceki Dönemin Not Ortalaması
- Vize Notu
- Çalışma Durumu (Hiç Çalışmadım, Çok Çalıştım, Çalıştım, Çalışmadım, Belirtilmedi)
- Mevcut Çalışma ile Sınavdan Kaç Puan Beklendiği
- Sınav Nasıl Geçti (İyi Geçti, Kötü Geçti, Çok İyi Geçti, Çok Kötü Geçti, Belirtilmedi)
- Final Sınav Puanı

Ham veriler öncelikle bir veritabanına yüklenmiş ve ön işleme yapılmıştır. Öğrenci bilgi sistemi ile kâğıttan elde edilen veriler birleştirilmiş ve sınava girmeyen öğrencilerin verileri çıkarılmıştır. Tüm işlemler sırasında öğrencileri tanımlayabilecek herhangi bir veri elde edilmemiştir. Tüm işlemler sonrası 179 öğrenci verisinden 160 öğrenci verisi kalmıştır. Öğrencilerin programlara göre sayıları Tablo 1’de verilmiştir.

Program	Öğrenci Sayısı
Bankacılık ve Sigortacılık	42
Basım ve Yayın Teknolojileri	12
Fotoğrafçılık ve Kameramanlık	29
Grafik Tasarımı	44
Kooperatifçilik	8
Lojistik	25
<b>Toplam</b>	<b>160</b>

**Tablo 1.** Programlara Göre Öğrenci Sayıları

## 3. Karar Ağaç Algoritmaları

Karar ağaçları hem kolay anlaşılmayı sağlayan grafiksel bir arayüz sunar hem de yüksek başarı oranı sunar. Bu özellikleri sebebiyle en çok tercih edilen tahmin yöntemleri arasında oldukları söylenebilir. Karar ağaçları genel olarak veriyi alt gruplara bölerek dallanma yapar ve her dal yeni bir kuraldır. Günümüzde WEKA (Waikato Environment for Knowledge Analysis) yazılımı birçok alanda elde edilen büyük verilerin işlenmesinde başarılı bir şekilde kullanılmaktadır (Kuyucu, 2012). Bu yazılım Yeni Zelanda’daki Waikato Üniversitesi tarafından geliştirilmiştir (Aydemir,

2018). Aşağıdaki karar ağaç algoritmaları Weka programında varsayılan olarak sunulan ve yaygın olarak kullanılan algoritmalar (Witten, Frank ve Hall, 2016).

- DecisionStump, kategorik ya da sayısal bir sınıf içeren veri kümeleri için tek seviyeli bir ikili karar ağacı oluşturur.
- RandomTree, her düğümde verilen belirli sayıda rasgele özelliğe dayanan bir test seçer ve budama yapmaz.
- RandomForest, rastgele ağaçları birleştirerek rastgele ormanlar inşa eder.
- REPTree, bilgi kazancı / değişimini azaltma yöntemini kullanarak bir karar veya regresyon ağacı oluşturur ve bu hatayı azaltılmış hata budaması kullanarak eritir. Hız için optimize edilmiş olup sadece bir kez sayısal nitelikler için değerleri sıralar
- M5P, M5 modelini kullanarak öğrenmeyi gerçekleştirir ve kuralları çıkarır.

Buradaki çalışmada öğrencilerin sınav notları sayısal olarak tahmin edilmiştir. Yukarıda bahsedilen yöntemler de sayısal tahminler elde etmek için kullanılır ve tahmin başarısını ölçmek için korelasyon katsayısı (r), ortalama mutlak hata (MAE) ve karekök ortalama hata (RMSE) değerleri kullanılır. Bu değerler aşağıdaki formüller yardımıyla hesaplanır.

$$r = \frac{\sum(xy) - \frac{(\sum x)(\sum y)}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

$$MAE = \frac{\sum_{i=1}^n |\hat{\theta}_i - \theta_i|}{n}$$

$n$  = örneklem sayısı

$\theta_i$  =  $i$  sıra numaralı gerçek talep

$\hat{\theta}_i$  =  $i$  sıra numaralının tahmin edilen talebi

$i$  = örneklem sırası

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{\theta}_i - \theta_i)^2}{n}}$$

$n$  = örneklem sayısı

$\theta_i$  =  $i$  sıra numaralı gerçek talep

$\hat{\theta}_i$  =  $i$  sıra numaralının tahmin edilen talebi

$i$  = örneklem sırası

#### 4. Bulgular

Beş farklı ağaç algoritması kullanarak Weka programı ile elde edilen verilere yönelik tahminler yapılmıştır. Verilerin test ve eğitim olarak ayrıştırılmasında 10-katlı çapraz doğrulama yöntemi kullanılmıştır. Weka programı varsayılan olarak çapraz doğrulama katsayısını 10 olarak belirlemiştir. Verilerin test ve eğitim olarak ayrıştırılmasında en

yüksek başarı oranı 10 katlı çapraz doğrulama yönteminde elde edilmektedir (Witten, Frank ve Hall, 2016).

Algoritma	Korelasyon Katsayısı (r)	Ortalama Mutlak Hata (MAE)	Karekök Ortalama Hata (RMSE)
DecisionStump	0.4901	11.9092	14.727
RandomTree	0.5414	10.9437	16.0874
RandomForest	0.7270	10.1653	12.8113
REPTree	0.6774	10.3285	13.5407
M5P	0.6680	10.6533	13.7203

**Tablo 2.** Kullanılan Algoritmalar ve Başarı Değerleri

Tablo 2'deki veriler incelendiğinde; kullanılan algoritmaların başarı oranlarına göre en yüksek başarı oranı RandomForest algoritması ile elde edilmiştir. Bu algoritmadaki gerçek veriler ile tahmin verileri arasındaki korelasyon katsayısı 0.7270 çıkararak veriler arasındaki uyumun iyi olduğunu göstermektedir. Ortalama mutlak hata (MAE) 10.1653 bulunması ile gerçek notlar ile tahmin edilen notların yaklaşık 10 puan aşağıya da yukarı sapmalar ile tahminler elde edildiği söylenebilir. Notların 0-100 aralığında olduğu düşünüldüğünde yaklaşık %10.16 sapmalar ile tahmin yapılmaktadır.

## 5. Tartışma Sonuç ve Öneriler

2018-2019 akademik yılı bahar dönemine ait Türk Dili II dersini alan öğrencilerin öğrenim görmekte olduğu program, ÖSS giriş türü, ÖSS giriş puanı, ÖSS giriş sıralaması, bir önceki dönem not ortalaması, ara sınav notu, sınava çalışma durumu, mevcut çalışma ile sınavdan kaç puan beklendiği, sınavın nasıl geçtiği verileri dikkate alınarak final sınav puanı tahmin edilmeye çalışılmıştır. Altı farklı programdaki toplam 160 öğrenciye ait verilerin Weka programında beş farklı ağaç algoritması ile analizi yapılmış ve en iyi başarı oranı RandomForest algoritmasında elde edilmiştir. Gerçek veriler ile tahmin verileri arasındaki korelasyon katsayısı 0.72 bulunmuş ve 0-100 aralığındaki notlar 10.16 ortalama mutlak hata ile tahmin edilmiştir. Benzer bir çalışma Aydemir (2019) tarafından yapılmış ve Bagging yöntemi ile 0.80 korelasyon katsayısı elde edilmiştir. Bu çalışma öğrencilerin ders çalışma durumlarını dikkate alması nedeniyle farklılık göstermektedir. Aksu (2018) tarafından gerçekleştirilen öğrencilerin PISA başarısının tahmin edildiği çalışmada da buradaki gibi RandomForest algoritması en iyi başarı oranını sağlamıştır. Fakat bu çalışma sınıflandırma yöntemlerini kullanması nedeniyle başarı ölçütleri buradaki çalışmadan farklılık göstermektedir. Gök (2017)'ün çalışmasında ise öğrencilere yapılan anket ile elde edilen verilere makine öğrenmesi algoritmaları uygulanmış ve yapılan not tahmininde RandomForest algoritması en iyi başarı oranını vermiştir. Öğrenme zorluklarının önceden bilinmesi nihai sonucu etkileyecek önlemler alınmasına yardımcı olur (Tsiakmaki ve diğ., 2018). Bu çalışmada öğrencinin ders çalışma durumundan alınacak notu tahmin etmesi, çalışma düzeyinin yeterliliği ile ilgili öğrenciye geri dönüt sağlayacaktır. Martinez Abad ve Chaparro Caso López (2017) de akademik performansın göstergeleri arasında en çok kişisel faktörlerin ve sonrasında ise okul ile ilişkiler ve sosyal faktörlerin önemli olduğunu belirtmiştir. Bu çalışmada daha çok kişisel faktörlere odaklanılmıştır. Yeni çalışmalarda ders geçme notuna etki eden farklı unsurları da dikkate alan tahminler yapılabilir. Bu sayede başarı oranının artırılması sağlanabilir.

## Kaynakça

- Aksu, G. (2018). PISA Başarısını Tahmin Etmede Kullanılan Veri Madenciliği Yöntemlerinin İncelenmesi. Hacettepe Üniversitesi, Eğitim Bilimleri Ana Bilim Dalı, Doktora Tezi.
- Aydemir, E. (2018). Weka ile Yapay Zeka. Seçkin Yayınevi, Ankara.
- Aydemir, E. (2019). Ders Geçme Notlarının Veri Madenciliği Yöntemleriyle Tahmin Edilmesi. Avrupa Bilim ve Teknoloji Dergisi, (15), 70-76.
- Bramer, M. (2013). Principles of Data Mining (2nd ed.). London: Springer-Verlag.
- del Campo-Ávila, J., Muñoz, R. C., Ruiz, F. A. T., & Bueno, R. M. (2015). Mining web-based educational systems to predict student learning achievements. International Journal of Artificial Intelligence and Interactive Multimedia, 3(2), 49-54.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996). The KDD process of extracting useful knowledge from volumes of data. Commun. ACM, 39 (11), 27-34.
- Gök, M. (2017). Makine Öğrenmesi Yöntemleri İle Akademik Başarının Tahmin Edilmesi. Gazi Üniversitesi Fen Bilimleri Dergisi, 5(3): 139-148.
- Kuyucu, Y. E. (2012). Lojistik regresyon analizi (LRA), yapay sinir ağları (YSA) ve sınıflandırma ve regresyon ağaçları (C&RT) yöntemlerinin karşılaştırılması ve tıp alanında bir uygulama. Yüksek Lisans Tezi. Gaziosmanpaşa Üniversitesi Sağlık Bilimleri Enstitüsü, Tokat.
- Martinez Abad, F., & Chaparro Caso López, A. A. (2017). Data-mining techniques in detecting factors linked to academic achievement. School Effectiveness and School Improvement, 28(1), 39-55.
- Neelamegam, S., Ramaraj, E. (2013) Classification algorithm in Data mining: An Overview, International Journal of P2P Network Trends and Technology (IJPTT), 4 (8), 369-374.
- Sieber, J. E. (2008). Data mining: knowledge discovery for human research ethics, J Empir Res Hum Res Ethics, 3 (3), 1-2.
- Thuarisingham, B. M., (2003). Web Data Mining and Applications in Business Intelligence and Counter Terrorism, USA: CRC Press LLC, Boca Raton, FL.
- Tsiakmaki, M., Kostopoulos, G., Koutsonikos, G., Pierrakeas, C., Kotsiantis, S., & Ragos, O. (2018, July). Predicting University Students' Grades Based on Previous Academic Achievements. In 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA) (pp. 1-6). IEEE.
- Weiss, S. M., Kulikowski, C. A. (1991). Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems. San Mateo, CA: Morgan Kaufmann.
- Witten, I. H., Frank, E., Hall, M. (2016) Data minig: Practical machine learning tools and techniques, United States of America: Morgan Kaufmann publications.