

Veri Madenciliği Yöntemleriyle Ortaöğretim Öğrencilerinin Akademik Başarılarının Tahmin Edilmesi

Bilgisayar Mühendisliği, Teknoloji Fakültesi, Gazi Üniversitesi, Ankara, Türkiye

Yusuf Doğu 23181616038

yusuf.dogu@gazi.edu.tr

MOHAMMED ABDULHAMID ALBAWENDI 22181616071

mabdulhamid.albawendi@gazi.edu.tr

ÖZET

Bu çalışmada, Portekiz'deki eğitim sorunu üzerine iki farklı liseden toplanan öğrenci verileri kullanılarak, ortaöğretim öğrencilerinin Portekizce ve Matematik derslerindeki başarılarının tahmin edilmesi amaçlanmıştır. Veri seti, öğrencilerin demografik bilgileri, sosyal ve ailevi faktörler, dönem içi notları gibi çeşitli özellikleri içermektedir.

Veri analizi sürecinde, veri ön işleme adımları gerçekleştirilmiş; eksik veri kontrolü, değişkenlerin tür analizi ve istatistiksel özetleri çıkarılmıştır. Daha sonra, kategorik ve sayısal değişkenler ayrılarak görselleştirme çalışmaları yapılmıştır.

Veri seti detaylı şekilde analiz edilip anlamlandırıldıktan sonra G1, G2 ve G3 hedef sütunları bazında korelasyon analizi yapabilmek adına, öncelikle kategorik sütunlar OneHotEncoder ile sayısal hale getirilmiştir. Sonrasında hem Matematik hem Portekizce verileri için diğer sütunların hedef sütunlardaki korelasyonları incelenmiştir. Bununla birlikte, hedef sütunlarla fazlasıyla benzerlik gösteren G1, G2, G3 sütunları ve korelasyonu çok düşük olan sütunlar, daha kaliteli bir veri seti elde edebilme amacıyla veri setlerinden çıkarılmıştır. Ayrıca, normalde G1, G2 ve G3 olan hedef sütunların birbirleriyle yüksek korelasyon göstermeleri nedeniyle, ortalamaları alınıp tek bir sütuna indirgenmiştir. Test aşaması için ise tüm sütunlar MinMaxScaler ile 0-1 arasında ölçeklendirilip; Random Forest, Decision Trees, Support Vector Machines ve Yapay Sinir Ağları gibi makine öğrenme modelleriyle doğruluk değerleri test edilip geliştirilmeye çalışılmıştır. Sonuç olarak, öğrencilerin Portekizce ve Matematik derslerindeki başarılarını tahmin etmede, geçmiş başarısızlıklar, devamsızlık sayısı ve yükseköğretim isteği önemli faktörler olarak belirlenmiştir. Bu çalışmanın bulguları, Portekiz'deki eğitim problemine yeni çözümler üretebilmenin yanı sıra, genel olarak öğrenci başarısındaki faktörlerin geniş bir analizini sunmaktadır.

Anahtar Kelimeler: Veri Bilimi, Makine Öğrenme Modelleri, Eğitim Analitiği

Giriş

Eğitim, ailede başlar ve bireylerin ömür boyu bilgi, beceri ve yeteneklerini geliştirmeleri için hayati bir öneme sahip olmakla birlikte, toplumların uzun vadeli gelişimi için temel bir faktördür. Ancak, dünya genelinde olduğu gibi Portekiz'de de öğrenci başarısızlığı ve okul terk oranları, eğitim sistemlerinin karşılaştığı ciddi problemler arasında yer almaktadır. Bu problemin temel kaynağı olarak Matematik ve Portekizce derslerindeki başarı eksikliği öne

çıkılmaktadır; bir anlamda, öğrencilerin bu temel iki dersteki eksikliği diğer derslere de yansımaktadır.

Bu çalışmada, Portekiz'deki iki farklı liseden elde edilen gerçek öğrenci verileri kullanılarak, öğrencilerin yıl sonu notlarının tahmin edilmesi amaçlanmıştır. Çalışmada, öğrencilerin demografik bilgileri, sosyal ve ailevi durumları ile dönem içi notları gibi çeşitli özellikler analiz edilmiştir. Veri madenciliği ve makine öğrenmesi teknikleri kullanılarak, öğrenci performansını etkileyen faktörler belirlenmeye ve başarı tahmin modelleri geliştirilmeye çalışılmıştır.

Bu araştırmanın temel amacı, öğrenci başarısına etki eden önemli değişkenleri tespit ederek öğrencinin okul başarı sürecini daha derin şekilde anlayabilmek; bununla birlikte, öğrencinin eğitim sürecini bitirme kararı almadan önce geri bildirimlerle yol gösterebilmektir. Ayrıca, bu çalışma kapsamında aşağıdaki araştırma sorularına yanıt aranmıştır:

Öğrenci performansını tahmin etmek için hangi değişkenler daha etkilidir?

- Hangi makine öğrenme modelleri, en yüksek doğrulukla öğrenci başarısını tahmin edebilmektedir?
- Demografik ve sosyal faktörler, öğrencilerin akademik başarılarında nasıl bir rol oynamaktadır?

Bu bağlamda yapılan analizler ve elde edilen sonuçlar, eğitim politikalarının iyileştirilmesine ve öğrenci başarı oranlarının artırılmasına katkı sağlamayı hedeflemektedir.

2. Materyal ve Yöntem

2.1. Veri Seti

Bu çalışmada, Portekiz'deki iki farklı devlet okulundan toplanan gerçek öğrenci verileri kullanılarak, ortaöğretim öğrencilerinin akademik performanslarının tahmin edilmesi hedeflenmiştir. Toplanan veri seti yalnızca öğrencilerin notlarını değil, aynı zamanda onların sosyal, demografik ve ailevi özelliklerini de kapsamaktadır. Bu kapsamlı veri seti sayesinde, öğrenci başarısını etkileyen çok boyutlu faktörlerin analizi mümkün hâle gelmiştir.

Veri seti, iki temel derse (Matematik ve Portekizce) odaklanmaktadır ve her ders için ayrı veri kümeleri oluşturulmuştur. Öğrencilerin birinci ve ikinci dönem notları (G1 ve G2) ile final notları (G3) birlikte kaydedilmiştir. Ayrıca, öğrencilerin yaşları, cinsiyetleri, yaşadıkları yerleşim tipi (kentsel veya kırsal), aile büyüklükleri, ebeveynlerin eğitim seviyeleri ve meslekleri gibi demografik bilgiler; internet erişimi, okul destek programlarına katılım, özel ders alımı gibi okul dışı faktörler; boş zaman süreleri ve alkol tüketim alışkanlıkları gibi sosyal etkileşim verileri de toplanmıştır.

Toplamda 397 + 649 öğrenci kaydının bulunduğu veri setinde, 33 farklı değişken yer almaktadır. Bu değişkenlerin bazıları doğrudan akademik başarıyı yansıtan objektif ölçütlerken, bazıları öğrencinin sosyal çevresi, yaşam tarzı ve aile desteği gibi dolaylı faktörleri temsil etmektedir.

Çalışmada kullanılan veri setinde eksik veri bulunmamakta, tüm değişkenler eksiksiz olarak kaydedilmiştir. Böylece, modelleme ve analiz aşamalarında eksik veri ile uğraşmadan doğrudan veri madenciliği teknikleri uygulanabilmektedir. Bununla birlikte, gerçek veriler ele alındığından sayısal değişkenler için aykırı değer bastırma yöntemi tercih edilmemiştir.

Bu çalışma kapsamında, öğrencilerin akademik performanslarını etkileyen faktörlerin derinlemesine anlaşılması ve başarı tahminlerinin yapılabilmesi için kapsamlı bir veri analiz süreci yürütülmüş; böylece eğitimcilerin gelecekte öğrencilerin başarısızlık risklerini erken aşamalarda tespit edebilmeleri için bilgi temelli bir yaklaşım oluşturulmuştur.

Değişken Adı	Açıklama
sex	Öğrencinin cinsiyeti (ikili: kız veya erkek)
age	Öğrencinin yaşı (sayısal: 15 ile 22 arasında)
school	Öğrencinin okulu (ikili: Gabriel Pereira veya Mousinho da Silveira)
address	Öğrencinin ev adresi türü (ikili: kentsel veya kırsal)
Pstatus	Ailenin birlikte yaşama durumu (ikili: birlikte yaşıyor veya ayrı)
Medu	Annenin eğitimi (sayısal: 0'dan 4'e kadar)
Mjob	Annenin mesleği (nominal: öğretmen, sağlıkla ilgili, kamu hizmeti, evde veya diğer)
Fedu	Babanın eğitimi (sayısal: 0'dan 4'e kadar)
Fjob	Babanın mesleği (nominal: öğretmen, sağlıkla ilgili, kamu hizmeti, evde veya diğer)
guardian	Öğrencinin velisi (nominal: anne, baba veya diğer)
famsize	Aile büyüklüğü (ikili: ≤ 3 veya > 3)
famrel	Aile içi ilişkilerin kalitesi (sayısal: 1 - çok kötü, 5 - çok iyi)

Değişken Adı	Açıklama
reason	Bu okulun seçilme nedeni (nominal: eve yakın, okulun itibarı, ders tercihi veya diğer)
traveltime	Evden okula ulaşım süresi (sayısal: 1 - < 15 dk, 2 - 15-30 dk, 3 - 30 dk-1 saat, 4 - >1 saat)
studytime	Haftalık çalışma süresi (sayısal: 1 - < 2 saat, 2 - 2-5 saat, 3 - 5-10 saat, 4 - > 10 saat)
failures	Geçmişteki sınıf başarısızlıkları (sayısal: $1 \leq n < 3$, aksi takdirde 4)
schoolsup	Ekstra eğitim desteği (ikili: evet veya hayır)
famsup	Aileden eğitim desteği (ikili: evet veya hayır)
activities	Ekstra-kürsü aktiviteleri (ikili: evet veya hayır)
paidclass	Ücretli ek dersler (ikili: evet veya hayır)
internet	Evde internet erişimi (ikili: evet veya hayır)
nursery	Anaokuluna gitme durumu (ikili: evet veya hayır)
higher	Yükseköğrenim almak isteme durumu (ikili: evet veya hayır)
romantic	Romantik ilişki durumu (ikili: evet veya hayır)
freetime	Okul sonrası boş zaman (sayısal: 1 - çok düşük, 5 - çok yüksek)
goout	Arkadaşlarla dışarı çıkma sıklığı (sayısal: 1 - çok düşük, 5 - çok yüksek)
Walc	Hafta sonu alkol tüketimi (sayısal: 1 - çok düşük, 5 - çok yüksek)
Dalc	Çalışma günü alkol tüketimi (sayısal: 1 - çok düşük, 5 - çok yüksek)
health	Mevcut sağlık durumu (sayısal: 1 - çok kötü, 5 - çok iyi)
absences	Okul devamsızlık sayısı (sayısal: 0 ile 93 arasında)
G1	Birinci dönem notu (sayısal: 0 ile 20 arasında)
G2	İkinci dönem notu (sayısal: 0 ile 20 arasında)
G3	Final notu (sayısal: 0 ile 20 arasında)

Tablo 1: Her iki veri setinde bulunan değişkenler ve açıklamaları

2.2. Kullanılan Makine Öğrenmesi Algoritmaları

Makine öğrenmesi, yapay zekâ alanının önemli bir dalı olup bilgisayarların verilerden öğrenerek karar verme ve tahmin yapabilmesini sağlayan bir disiplindir. Günümüzde, eğitim alanı da dâhil olmak üzere birçok sektörde veri madenciliği ve makine öğrenmesi teknikleri yoğun şekilde kullanılmaktadır. Özellikle öğrenci başarısının tahmin edilmesi gibi karmaşık problemler için, geleneksel yöntemlerin ötesinde daha etkili ve doğru sonuçlar elde etmek amacıyla makine öğrenmesi algoritmalarına başvurulmaktadır.

Bu algoritmaların uygulanmasından önce, veri seti üzerinde ön işleme (preprocessing) adımları gerçekleştirilmiştir. Eksik veri kontrolü, kategorik değişkenlerin sayısallaştırılması (Label Encoding, One-Hot Encoding gibi teknikler) ve veri temizleme işlemleri yapılmıştır. Özellikle değişkenlerin farklı veri tiplerine göre ayrılması ve istatistiksel özetlerinin çıkarılması süreci, modelleme aşamasının doğruluğunu artırmıştır. Gerekli sayısallaştırmalar yapıldıktan sonra MinMaxScaler kullanılarak tüm veri seti 0–1 aralığına indirgenmiştir.

Veri madenciliği süreci kapsamında ayrıca veri görselleştirme teknikleri kullanılarak değişkenlerin dağılımları incelenmiş ve bazı önemli özellikler belirlenmiştir. Modelleme aşamasında ise bu çalışma özelinde yalnızca sınıflandırma (binary classification) üzerinde durulmuştur.

Bu doğrultuda kullanılan makine öğrenmesi algoritmaları şunlardır:

Karar Ağaçları (Decision Trees):

Karar ağaçları, veriyi belirli kurallara göre dallara ayırarak sınıflandıran denetimli öğrenme algoritmalarıdır. Her düğüm bir özelliği, her dal bir karar kuralını ve her yaprak ise bir sonucu temsil eder. Bu çalışmada, karar ağaçları kullanılarak öğrencilerin başarı düzeyleri farklı parametre kombinasyonlarına göre analiz edilmiştir.

Rastgele Ormanlar (Random Forest):

Rastgele Orman algoritması, birçok karar ağacının birleşiminden oluşan güçlü bir topluluk (ensemble) yöntemidir. Torbalama (bagging) tekniği ile farklı alt veri kümeleri üzerinde bağımsız ağaçlar eğitilir ve bu ağaçların oylamasıyla daha sağlam ve genelleştirilebilir tahminler elde edilir. Bu yöntem, özellikle veri setindeki değişkenlikleri ve aşırı öğrenmeyi (overfitting) azaltmada etkili olmuştur.

Destek Vektör Makineleri (Support Vector Machines – SVM):

SVM, veriyi en iyi ayıracak hiper düzlemi (hyperplane) bulmayı amaçlayan güçlü bir sınıflandırma algoritmasıdır. Özellikle yüksek boyutlu veri setlerinde etkili performans göstermektedir. Bu çalışmada, öğrenci verilerinin farklı sınıflara ayrılması sürecinde SVM algoritması ile başarı tahminleri yapılmıştır.

Yapay Sinir Ağları (Artificial Neural Networks – ANN):

İnsan beyninin çalışma şekliyle esinlenilerek geliştirilen yapay sinir ağları, çok katmanlı yapıları sayesinde karmaşık ilişkileri modelleyebilme yeteneğine sahiptir. Bu çalışmada, öğrencilerin not tahminlerinde yapay sinir ağları da kullanılarak veriler arasındaki doğrusal olmayan ilişkiler tespit edilmeye çalışılmıştır.

Elde edilen bulgular, hangi makine öğrenmesi algoritmasının öğrenci başarısını tahmin etmede daha yüksek doğruluk sağladığını ve hangi değişkenlerin başarı üzerinde daha etkili olduğunu ortaya koymuştur.

2.3. Literatür Taraması

Öğrenci performansının tahmini ve öğrenci davranışlarının analiz edilmesi amacıyla yapılan çalışmalar, veri madenciliği ve makine öğrenmesi tekniklerinin eğitim alanında güçlü araçlar olduğunu göstermektedir. Literatürde bu konuda yapılan araştırmalar, çeşitli yöntemlerin farklı veri setleri ve senaryolar üzerinde başarıyla uygulanabileceğini ortaya koymuştur.

Firdausiah Mansur ve Yusof (2018) tarafından yapılan çalışmada, öğrenci davranışlarının gözlemlenmesi ve sınıflandırılması için K-means kümeleme algoritması kullanılmıştır. Araştırmacılar, öğrenci başarılarını etkileyen faktörleri belirlemek amacıyla öğrenci performans verileri ve öğretmen niteliklerini içeren veriler toplamış, ardından belirli öznelitliklerin seçimiyle K-means algoritmasını uygulamışlardır. Sonuç olarak, öğrenci başarısı ile öğretmen özellikleri arasında güçlü bir ilişki tespit edilmiştir. Bu çalışma, e-öğrenme ortamlarında öğrenci-öğretmen etkileşiminin optimize edilmesi ve kişiselleştirilmiş eğitim stratejilerinin geliştirilmesi açısından önemli bir katkı sunmuştur.

Karagiannopoulos ve arkadaşları (2007) tarafından gerçekleştirilen başka bir çalışmada, veri setlerinde sıklıkla karşılaşılan dengesiz dağılım (imbalanced data) problemi ele alınmıştır. Araştırmacılar, azınlık sınıfların tahmin doğruluğunu artırmak amacıyla “A Wrapper for Reweighting Training Instances” adını verdikleri bir yöntem önermişlerdir. Bu yöntem, öğrenme algoritmalarını değiştirmeksizin eğitim örneklerinin ağırlıklarını yeniden düzenlemektedir. Özellikle Naive Bayes, C4.5 ve 5NN gibi sınıflandırıcılar üzerinde yapılan deneylerde, önerilen yaklaşımın azınlık sınıfların tahmin doğruluğunu önemli ölçüde

artırdığı ve genel performansı iyileştirdiği gözlemlenmiştir. Bu bulgular, öğrenci performansı tahmini gibi dengesiz veri setlerine sahip uygulamalarda dikkate alınması gereken önemli bir stratejiyi ortaya koymaktadır.

Cortez ve Silva (2008), “Using Data Mining to Predict Secondary School Student Performance” başlıklı çalışmalarında, veri madenciliği yöntemlerinin lise düzeyindeki öğrenci performansını öngörmedeki etkinliğini incelemişlerdir. Bu çalışmada, öğrencilerin akademik başarılarını etkileyebilecek faktörler (okula devamsızlık, ebeveyn eğitim düzeyi, çalışma saatleri gibi) analiz edilmiş ve karar ağaçları ile Naive Bayes sınıflandırıcıları kullanılarak modeller oluşturulmuştur. Elde edilen sonuçlar, %72’ye varan doğruluk oranlarına ulaşılabilceğini göstermiştir. Araştırmacılar, öğrenci başarısızlıklarının erken dönemde tespit edilerek önleyici müdahalelerde bulunulabileceğini ve böylece eğitim süreçlerinin iyileştirilebileceğini belirtmişlerdir.

Son olarak, Helwig (2017) tarafından gerçekleştirilen “Adding Bias to Reduce Variance in Psychological Results: A Tutorial on Penalized Regression” başlıklı çalışmada, cezalı regresyon yöntemlerinin özellikle küçük ve orta ölçekli veri setlerinde daha güvenilir ve tekrarlanabilir sonuçlar üretebileceği tartışılmıştır. Araştırmada, Ridge Regression, Lasso ve Elastic Net gibi yöntemler detaylı bir şekilde açıklanmış ve bu yöntemlerin klasik OLS regresyona kıyasla daha düşük varyans ve daha yüksek tahmin doğruluğu sunduğu gösterilmiştir. Çalışmada ayrıca cezalı regresyonun yüksek değişkenlik gösteren veri setlerinde overfitting riskini azalttığı ve psikolojik araştırmalar gibi küçük örneklem büyüklüğüne sahip çalışmalarda model performansını artırdığı vurgulanmıştır. Çalışmanın bir bölümünde, öğrenci matematik performansını tahmin etmek amacıyla 30 farklı değişken kullanılarak yapılan uygulama örneği sunulmuş ve cezalı regresyon yöntemlerinin daha doğru ve yorumlanabilir sonuçlar ürettiği gözlemlenmiştir.

Bu literatür taraması sonucunda, farklı yöntemlerin öğrenci performansı tahmininde önemli avantajlar sunduğu ve özellikle veri dengesizliği ile değişken çokluğu gibi zorlukların üstesinden gelmede cezalı regresyon tekniklerinin etkili olduğu anlaşılmaktadır. Bu bağlamda, mevcut çalışmanın temel amacı, öğrenci davranış kalıplarını anlamak, başarıyı etkileyen faktörleri belirlemek ve veri madenciliği teknikleri kullanarak tahmin modelleri geliştirerek eğitim süreçlerinin iyileştirilmesine katkıda bulunmaktır.

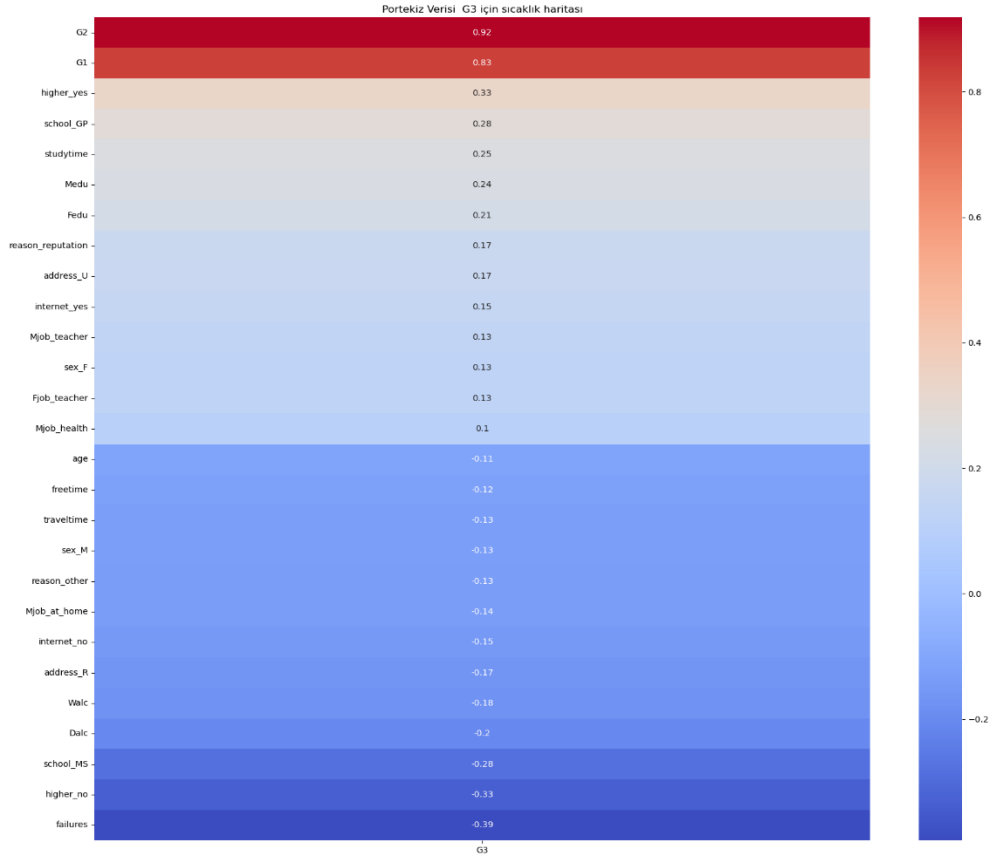
3. Bulgular ve Tartışma

Bu çalışmada, Portekiz'deki iki farklı liseden elde edilen 649 + 397 öğrencilik veri seti kullanılarak öğrencilerin akademik başarılarını etkileyen faktörler analiz edilmiş ve çeşitli makine öğrenmesi algoritmaları uygulanarak sınav notlarının ortalaması tahmin edilmeye çalışılmıştır. Kullanılan modellerin performansı, bu proje özelinde yalnızca sınıflandırma bağlamında değerlendirilmiştir.

Veri seti analizinde öne çıkan ilk bulgu, öğrencilerin birinci (G1) ve ikinci dönem (G2) notlarının final notu (G3) ile yüksek korelasyon içinde olmasıdır. Özellikle bu iki değişkenin modele dâhil edilmesi durumunda, tahmin doğruluğu ciddi şekilde artmıştır. Bu sebepten ötürü, G1, G2 ve G3 notlarının ortalaması alınarak modeller eğitilmiştir. Bunun yanı sıra, geçmiş başarısızlıklar (failures), devamsızlık sayısı (absences), aile eğitimi düzeyi (Medu, Fedu) ve öğrencinin haftalık çalışma süresi (studytime) gibi değişkenler de akademik performansı etkileyen güçlü faktörler olarak belirlenmiştir.



Matematik verisindeki değişkenlerin G3 üzerindeki korelasyonları . (sıcaklık haritasında korelasyonu en az %10 olan değişkenler ele alınmıştır.)



Portekizce verisindeki değişkenlerin G3 üzerindeki korelasyonları .(sıcaklık haritasında korelasyonu en az %10 olan değişkenler ele alınmıştır.)

Matematik ve Portekizce derslerine ait veriler, MinMaxScaler ile ölçeklendirildikten sonra iki şekilde ele alınmıştır: indirgenmiş sütunlardan oluşan veri setleri ve ham halleri. Uygulanan makine öğrenmesi algoritmaları arasında Karar Ağaçları (Decision Tree), Rastgele Orman (Random Forest), Destek Vektör Makineleri (SVM) ve Yapay Sinir Ağları (ANN) yer almıştır. Her bir model için hiperparametre optimizasyonu yapılabilmesi adına GridSearch yöntemi kullanılmış ve K-Fold yöntemi ile en iyi train-test ayrımı belirlenerek en doğru model eğitilmiştir.

	Matematik Verisi (18 sütun)	Matematik Verisi (33 sütun)	Portekizce Verisi (20 sütun)	Portekizce Verisi (33 sütun)
Model				
Random Forest	73.42%	74.68%	84.50%	84.50%
Decision Tree	69.62%	68.35%	78.46%	78.46%
SVM	78.48%	70.89%	81.40%	81.40%
YSA (ANN)	72.15%	75.95%	81.40%	81.54%

Sonuçlara baktığımızda, yaptığımız sütun indirgemelerinin maliyet-kazanç ilişkisi açısından çok faydalı olduğu görülmektedir. Örneğin, 18 değişken ile 33 değişkenlik veriyle aynı işlevi gören modeller oluşturulabilmiştir.

Sonuçlar, literatürdeki çalışmalarla uyumlu şekilde, veri madenciliği tekniklerinin öğrenci başarısını öngörmede güçlü araçlar olduğunu doğrulamaktadır. Özellikle Random Forest ve SVM modellerinin, çok değişkenli bu veri setinde daha etkili olduğu gözlemlenmiştir.

Bu sonuçlar neticesinde, bir öğrencinin notuna etki eden en önemli değişkenler hakkında net yorumlar elde edilmiştir. Ayrıca verinin doğru analiz edilip, sütunların uygun şekilde seçilmesinin model öğrenme ve doğruluk düzeyine olan etkisi açıkça ortaya konmuştur.

4. Sonuç ve Öneriler

Sonuçlara gelecek olursak, şu temel sorulara geri dönelim: **Eğitim nedir? Eğitimin amacı nedir?**

Eğitimin amacı; öğrencinin öğrenmek, öğretmenin ise öğretmek için sınıfa girdiği bir öğrenme sürecidir. Bir ülkede eğitim anlamında bir problem varsa, bu durumun üstesinden gelebilmek için sadece dönem sonu notlarını değerlendirmek yeterli ve doğru bir yaklaşım olmayabilir.

Buradaki temel problem, öğretmenin öğrenciye yeterli düzeyde yol gösterememesi olabilir. Aynı zamanda, bir ailenin çocuğunu nasıl eğittiği de büyük önem taşır. Bir öğretmen, öğretme ve yol gösterme görevini yerine getirmiyorsa, burada yalnızca bir eğitim sorunu değil, aynı zamanda bir **ahlak sorunu** söz konusudur. Ahlakın temeli ise ailede başlar; aileler ise ülkenin kültürel yapısından beslenerek çocuklarına yön verirler.

Belki de Portekiz'deki eğitim kültürüne de detaylı bir bakış atmak oldukça değerlidir. Yani, öğretmenlerin temel amacı gerçekten derse gelip sadece zamanın geçmesini beklemek midir, yoksa öğrencilere yön vermek midir? Bu sorunun yanıtı, eğitim sisteminin verimliliğiyle doğrudan ilişkilidir.

Öneriler:

1. **İleride yapılacak çalışmalarda**, sadece notlar ve demografik veriler değil; psikolojik, motivasyonel ve duygusal zekâ gibi nitel veriler de modele dâhil edilerek daha kapsayıcı bir başarı tahmin sistemi kurulabilir.
2. **Veri setine zaman serisi analizi** gibi yaklaşımlar uygulanarak öğrencilerin başarı trendleri dinamik olarak izlenebilir. Bu veriler, öğretmen-veli iş birliğini güçlendirecek geri bildirim sistemlerinde kullanılabilir.
3. **Öğretmen niteliğine dair değişkenler** de veri setine eklenmelidir. Sınavların zorluk düzeyi, öğretim yöntemi, öğretmenin iletişim tarzı gibi faktörler, öğrenci başarısı üzerinde belirleyici olabilir.

KAYNAKÇA

1. Bishop, C. M., *Pattern Recognition and Machine Learning*, Springer, 2006.
2. Breiman, L., "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
3. Cortez, P., Cerdeira, A., Almeida, F., Matos, T., and Reis, J., "Modeling student performance: A case study," in *Proceedings of the 5th International Conference on Educational Data Mining*, 2008.
4. Cortez, P., and Cerdeira, A., "Wine quality data set," University of Minho, 2009. [UCI Repository].
5. Cortes, C., and Vapnik, V., "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
6. Goodfellow, I., Bengio, Y., and Courville, A., *Deep Learning*, MIT Press, 2016.
7. Hijazi, S. T., and Naqvi, S. M. M. R., "Factors affecting students' performance: A case of private colleges," *Bangladesh e-Journal of Sociology*, vol. 3, no. 1, pp. 1–10, 2006.
8. Kotsiantis, S., Pintelas, P., and Kanellopoulos, D., "Preventing student dropout in distance learning using machine learning techniques," in *Proceedings of the 7th International Conference on Knowledge-Based Intelligent Information & Engineering Systems*, 2004.
9. Larose, D. T., *Discovering Knowledge in Data: An Introduction to Data Mining*, Wiley-Interscience, 2005.
10. McCulloch, W. S., and Pitts, W., "A Logical Calculus of the Ideas Immanent in Nervous Activity," *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, 1943.
11. Mitchell, T. M., *Machine Learning*, McGraw-Hill, 1997.
12. Quinlan, J. R., "Induction of Decision Trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.