Makale Türü: Araştırma Makalesi
Gönderim Tarihi: 31 Aralık 2019 **Kabul Tarihi:** 31 Mayıs 2020

Kısa Sınavların, Ödevlerin ve Projelerin Dönem Sonu Sınavına Olan Etkilerinin Farklı Makine Öğrenmesi Teknikleri ile Araştırılması

Analysis of the Effects of Quizzes, Homeworks and Projects on Final Exam with Different Machine Learning Techniques

Karani Kardaş[1], *Altay Güvenir*[2]

[1]*Havelsan A.Ş, Komuta Kontrol ve Savunma Teknolojileri Direktörlüğü, Ankara, Türkiye*

kkardas@havelsan.com.tr

[2]*Bilkent Üniversitesi, Bilgisayar Mühendisliği Bölümü, Ankara, Türkiye*
guvenir@cs.bilkent.edu.tr

Öz

Bu çalışma Bilkent Üniversitesi Bilgisayar Mühendisliği Bölümü'nde verilen Programlama Dilleri dersi kapsamında yapılan kısa sınavların, ödevlerin ve projelerin dönem sonu sınavına olan etkisini sunar. Bu bağlamda çalışmanın ilk kısmında, öğrenci verilerinden kısa sınav bilgileri ile Lineer Regresyon yöntemi kullanılarak lineer bir model üretilmeye çalışılmış ve dönem sonu sınav notu tahmin edilmeye çalışılmıştır. Çalışmanın ikinci kısmında ise En Yakın Komşu (KNN), Lineer Regresyon, Lojistik Regresyon, Destek Vektör Makineleri (SVM), Lineer Destek Vektör Makineleri ve Çok Katmanlı Algılayıcılar (MLP) makine öğrenmesi yöntemleri kullanılarak kısa sınavların, ödevlerin ve projelerin ayrı ayrı ve birlikte dönem sonu sınav notuna etkisi gösterilmeye çalışılmıştır. Çalışma, kullanılan makine öğrenmesi yöntemlerinin karşılaştırmalı performansını göstermektedir. Deneysel sonuçlar lineer modelimizin iyileştirilmesi gerektiğini belirtirken, sınıflandırma işlemlerinin başarılı bir şekilde elde edildiğini göstermektedir.

Anahtar kelimeler: Öğrenci performans tahmini, öğrenci performansı, kısa sınav, dönem sonu sınavı, makine öğrenmesi, lineer regresyon

Abstract

This study presents the effect of quizzes, assignments and projects on the final exam of the Programming Languages course given in the Department of Computer Engineering at Bilkent University. In this context, in the first part of the study, a linear model is tried to be produced by using the linear regression method from the student data and the final exam grade is estimated. In the second part of the study, K-Nearest Neighbors(KNN), Linear Regression, Logistic Regression, Support Vector Machines(SVM), Linear Support Vector Machinesand Multi-layer Perceptron (MLP) machine learning methods are used to show the effect of quizzes, assignments and projects separately and together on the final exam grade. The study shows the comparative performance of machine learning methods used in the experiment. Experimental results show that our linear model needs to be improved, while classification operations are achieved successfully.

Keywords: Student performance prediction, student performance, quiz, final exam, machine learning, linear regression

1. Giriş

Eğitim hiç şüphesiz birçok insanı ve ülkelerin geleceklerini etkileyen çok önemli bir kavramdır. Eğitim kurumlarının ana hedeflerinden biri eğitim kalitesini ve performansı arttırmak için bireysel performansları izlemektir. Bireysel zayıf yönlerin tespiti ve öğrencilerin gelecek için doğru yöne yönlendirilebilmesi daha iyi bir gelecek için önemlidir [1]. Eğitim kurumlarından biri olan üniversitelerde de öğrencilerin derste gösterdiği performans dikkat edilmesi gereken bir husustur ve fakülte, enstitü gibi alt kurumlara kaliteyi ve performansı arttırma adına önemli bilgiler sağlar.

Öğrencilerin akademik performans ölçümü zor bir süreçtir. Bu durumun ana nedeni ise öğrencilerin ders performansını etkileyen birçok faktörün olmasıdır [1 ve 2]. Ders kalitesini arttırmaya yönelik ders kapsamında yapılan kısa sınavlar (İng. quiz), ödevler, ara sınavlar ve projeler bu faktörler arasında düşünülebilir. Bu çalışmalar dönem boyunca öğrenme sürecini olumlu yönde etkiler, öğrencilere geri bildirimde bulunulmasına yardımcı olur ve dönem sonu sınavı öncesi öğrencilerin iyi ve yararlı yönde uyarılmasını sağlar. Ayrıca bu şekilde bir geri bildirimin yararlarından biri de derste başarısız olma veya dersi bırakma olasılığı yüksek olarak değerlendirilen öğrencilere yararlı bir uyarı sağlamasıdır [3]. Dönem sonu geçme notunu da etkileyen bu çalışmalara öğrenciler tarafından yeterli düzeyde önem verilmesi gerekmektedir. Bahsedilen çalışmalar sadece öğrencilere değil, dersi veren öğretmenlere de büyük faydalar sağlar. Öğretmenler de dersin dönem boyunca öğrenciler tarafından anlaşılmada zorluk çekilen konular hakkında bilgi sahibi olup, öğrencilerin gelişim takibini yapabilir ve dersin kalitesinin arttırılmasına yönelik olarak bu bilgileri kullanabilirler. Bu şekilde öğrenci performansının takibi özel ilgi gerektiren öğrenciler için de faydalar sağlar [4].

Çevrimiçi eğitim veren sistemlerin artması ile birlikte bu sistemler üzerinden milyonlarca insan, gerek öğrenci veya gerek

de iş yaşamında aktif rol alan çalışanlar olarak çeşitli eğitimler almaktadırlar [5]. Bu sistemler üzerinden esnek ve kolay bir şekilde ihtiyaç duyulan eğitimler alınabilir [6]. Bu sistemler için de öğrenci takibi, sistemin kalitesinin ve kullanılabilirliğinin arttırılmasında önemli yer tutar. Doğru yönlendirilmeyen öğrenciler eğitimlerden yeterli faydayı sağlayamazlar ve bu sistemlerin kullanımı düşer.

Erken not tahmini öğrencinin performansını arttırmada oldukça etkilidir ve öğrenci gelişimini izleme konusunda bir çözüm olarak düşünülebilir [4]. Günümüzde öğrencilerin eğitim bilgileri ile ilgili çok fazla veri kayıt edilebilmektedir. Bu verilerin çeşitliliği ve boyutu sürekli ve hızlı bir şekilde artmaktadır [7]. Bu veriler ile farklı yöntemler kullanılarak çıkarımlar ve geleceğe dönük tahminler yapılabilir. Makine öğrenmesi yöntemlerinin son zamanlarda yaygınlaşması ve birçok alanda başarılı bir şekilde uygulanması ile birlikte bu yöntemlerin öğrenci performansı analizi alanında da kullanımı oldukça yaygındır. Makine öğrenmesi teknikleri yardımı ile geçmiş öğrenci performansının analizi, öğrencilerin gelecek performansları hakkında bir perspektif verebilmektedir [4 ve 6]. Bu teknikler ile öğrenci not tahminleri yapılabilir, öğrenci performansı arttırılabilir ayrıca öğretmene dersin gidişatı hakkında faydalar sağlayabilir [4].

Ders boyunca yapılan kısa sınavlar, ödevler ve projeler öğrencilerin performansını izlemek ve hem öğrencilere hem de öğretmenlere geri bildirim vermek adına yararlıdır. Verilen derse aktif katılımın, yapılan kısa sınavlar ile ölçülebileceği değerlendirilebilir. Çünkü kısa sınava girebilmesi için öğrencilerin öncelikle derse katılması gerekmektedir. Ayrıca öğrencilerin kısa sınavlardan iyi bir not alabilmesi için derse gelmeden önce hazırlık yapmaları, dersi aktif olarak dinlemeleri ve derse etkili bir şekilde dikkatlerini vermeleri gerekir. Dönem boyunca verilen ödevler ve projeler öğrencilerin derse ilgisini arttırır. Öğrencilerin dersi takip etmelerini ve derse çalışıp araştırma yapmalarını kolaylaştırır.

Bu çalışmada, yukarıda önemi bahsedilen eğitim performansını arttırma amacıyla Bilkent Üniversitesi Bilgisayar Mühendisliği Bölümü'nde verilen Programlama Dilleri (CS315) dersi kapsamında yapılan kısa sınavların, ödevlerin ve projelerin dönem sonu sınavına olan etkisi incelenmiştir. Çalışmada 2017 ve 2018 yıllarında ve farklı öğretmenlerden bu dersi alan yedi şubenin 330 öğrenciye ait not bilgileri kullanılmıştır. İlk olarak, öğrenci verilerinden kısa sınav bilgileri ile dönem sonu sınav notu tahmin edilmeye çalışılmış, sonrasında birçok faklı makine öğrenmesi yöntemleri kullanılarak kısa sınavların, ödevlerin ve projelerin ayrı ayrı ve birlikte dönem sonu sınav notuna etkisi gösterilmeye çalışılmıştır.

Çalışma organizasyonu şu şekildedir: ikinci bölümde alanyazı taraması sonuçları sunulmuş ve ilgili çalışmalar hakkında bilgiler verilmiştir. Üçüncü bölümde kullanılan makine öğrenmesi yöntemleri hakkında kısa bilgiler verilmiştir. Kullanılan veri kümesi hakkında detaylı bilgiler ve veri kümesinde yapılan ön hazırlık işlemleri dördüncü bölümde anlatılmıştır. Beşinci bölümde yapılan deneyler ve sonuçları detaylı bir şekilde sunulmuştur. Çalışma, sonuçlar ve gelecek çalışmaların değerlendirilmesiyle sonlandırılmıştır.

2. İlgili Çalışmalar

Alanyazında eğitim performansını ölçme ve arttırma adına farklı yöntemler kullanan birçok çalışma vardır. Rao vd. öğrencilerin öğrenme performansını sınıflandırma yöntemi ile tahminlemeye çalışır ve öğrencileri dört sınıfa ayırır [1]. Çalışmada J48, Naïve Bayesian sınıflandırma ve Random Forest algoritmaları kullanılıp karşılaştırılmıştır. Random Forest algoritması en iyi sonucu vermiştir. Büyüklüğü 200 civarı olan veri kümesinde nümerik değerlerden çok gruplanmış değerler vardır. Kabul tipi, cinsiyet, eğitim materyali, aile geliri gibi sosyal ve demografik 20 öznitelik (İng. feature) dikkate alınmıştır.

Karar Ağaçları (Decision tree) yöntemi de eğitim performansı alanında kullanılmıştır. Bu yöntem ile kurallar üretilerek tahminler yapılabilir. Örneğin Al-Radaidehvd yapılan çalışmada üniversite öğrencilerinin C++ programlama dersinin 2005 yılı verileri kullanılarak, ders notları harf grupları olarak ön görülemeye çalışılmıştır [2]. Öğrenme işleminde 12 öznitelik kullanılmıştır. Çalışmada elde edilen sonuçların değerleri yeterince yüksek bulunmamış olup bunun sebebi yeterli verinin olmayışına bağlanmıştır. Khanyine Karar Ağaçları yöntemini kullanmış ve ortaokul öğrencilerinin önceki verileri kullanılarak ders geçme notu yine harf cinsinden tahminlenmeye çalışmıştır [8]. Çalışmada yaklaşık olarak %85 oranında doğru sınıflandırma başarısı elde edilmiştir.

Yapay Sinir Ağları bu alanda kullanılan bir başka makine öğrenmesi yöntemidir. Örneğin Agrawal ve Mavani öğrencilerin performans düzeylerini zayıf, orta, iyi ve çok iyi şeklindeki gruplar olarak öngörülemeye çalışmıştır [9]. Okubovd tarafından yapılan çalışmada ise öğrencilerin kısa sınav skorları tahmin edilmeye çalışılmıştır [3]. Veri kümesi olarak 2693 öğrencinin 10 farklı öğretmen tarafından verilen 15 dersteki derse katılım, aksiyonlar gibi kayıt verileri dikkate alınmıştır.

Kim ve arkadaşları tarafından yapılan bir çalışmada çevrimiçi kurs sitelerinden biri olan Udacity'den öğrenci verileri için Derin Öğrenme yöntemi kullanılarak öğrenci performans tahmini yapılmıştır [10].Pandey ve arkadaşlarının çalışmasında Bayes sınıflandırma yöntemi kullanılarak öğrencilerin önceki yıl verilerinden öğrenci performansı tahminlemeye çalışılmış, öğrenciler iki sınıf olarak gruplandırılmıştır [11]. Yadav ve Pal tarafından yapılan çalışmada ise mühendislik öğrencilerinin önceki yıl sınavlarından dönem sonu sınavı performansı tahminlemeye çalışılmıştır [7]. Öğrenciler 3 sınıf olarak gruplanmıştır. ID3, C4.5, ve CART karar ağacı ile sınıflandırma algoritmaları kullanılmış ve karşılaştırılmıştır. Çalışmaya göre C4.5 algoritması en iyi sonucu vermiştir.

Gadhavi ve Chirag tarafından gerçekleştirilen çalışmada tek öznitelik dikkate alınarak lineer regresyon modeli kullanılmıştır [6]. Öğrenci geçme notu, yapılmış testlerin ortalama verisi ile tahminlenmeye çalışılmıştır. Collaborative Filtering, Matrix Factorization ve Restricted Boltzmann Machines (RBM) yöntemleri ile üniversite öğrencilerinin akademik performansları tahmin etmeye çalışan [4], en iyi sonucu RBM yöntemi ile elde etmiştir. Veri kümesi 225 öğrencinin üç yıllık verilerinden oluşmaktadır. Verinin içeriği öğrencilerin ortaokul bilgilerinden, lise bilgilerinden, mülakat bilgilerinden, girilmiş test sonuçlarından ve diğer derslerden alınmış notlardan oluşmaktadır. Çalışmada birçok farklı derse göre değerlendirme yapılmıştır.

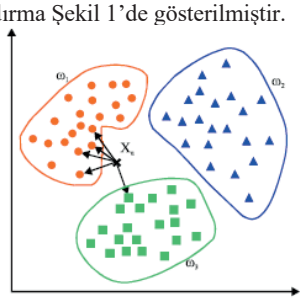
3. Kullanılan Makine Öğrenmesi Yöntemleri

Bu bölümde Makine Öğrenmesi konusunda bilgi verildikten sonra çalışmada kullanılan KNN (K-Nearest Neighbors) (En Yakın Komşu), Lineer Regresyon, Lojistik Regresyon, SVM (Support Vector Machines) (Destek Vektör Makineleri), Lineer SVM ve MLP (Multi-layer Perceptron) (Çok Katmanlı Algılayıcılar) yöntemleri hakkında kısa bilgi verilmiştir.

Makine öğrenmesi, insan zekâsının benzerinin yapımını amaçlayan Yapay Zekânın bir alt alanıdır ve bilgisayar yazılımlarının öğrenmesini sağlayan teknoloji olarak düşünülebilir [12 ve 13]. Makine öğrenmesinde yazılımların verilere göre davranışı açıkça programlanmadan eldeki

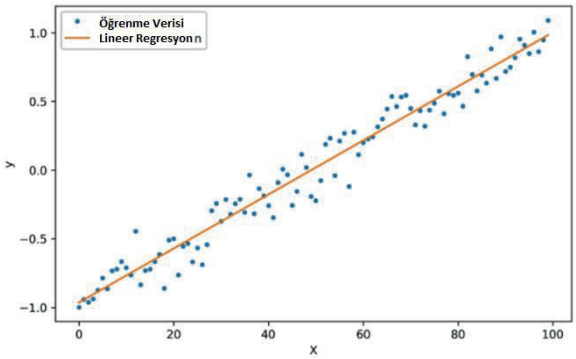
verilerden öğrenilen modeller ile gelecek verilerin tahminlenmesi sağlanır. Makine Öğrenmesi, bir problemin çözümünün ilgili ortamdan edinilen verilere göre modellenmesidir [12]. Son yıllarda oldukça yoğun çalışılan bir konu olan makine öğrenmesi bankacılık, alışveriş, reklamcılık, filtreleme, güvenlik, sosyal medya gibi birçok alanda çok çeşitli uygulamalarda kullanılmaktadır [14]. Makine öğrenmesi amacıyla kullanılan birçok yöntem ve algoritma vardır. Bazı yaklaşımlar tahmin yaparken, bazı yaklaşımlar sınıflandırma yapar. Bu yöntemler probleme veya eldeki verilere göre değişen farklı başarılara sahip olabilirler. Makine öğrenmesi algoritmaları girdi verilerini dikkate alarak, çıktı değerlerini kabul edilebilir bir aralıkta tahmin etmek için kullanır. Farklı algoritmalar, farklı özellikler ve farklı veriler ile çeşitli denemeler yapılarak en iyi model oluşturulmaya çalışılır.

KNN yöntemi öznitelik uzayındaki en yakın eğitim örneklerine dayanarak nesneleri sınıflandıran örüntü tanıma yöntemlerinden birisidir [15]. Bu yöntemde sınıflandırılmak istenen yeni örnek için, daha önceki sınıflandırılmış örneklerden *k* tanesinin her birinin yakınlığına bakılarak sınıflandırma yapılır [16]. Yeni örnek kendisine en yakın sınıfa dâhil edilir. Bu işlem sırasında *k* değişkeni örneğe en yakın olan *k* adet sınıf nitelikleri belli olan örneklerin sayısını belirtir. Yakınlık hesabında genelde öklit mesafesi kullanılır. KNN yöntemi kullanılarak yapılmış örnek bir sınıflandırma Şekil 1'de gösterilmiştir.



Şekil. 1. Örnek Bir KNN Sınıflandırması [17]

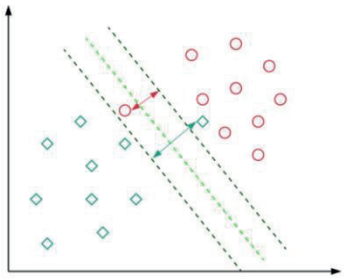
Lineer Regresyon iki değişken arasındaki ilişkiyi ölçmek için kullanılan makine öğrenmesi yöntemlerinden biridir [18]. Bu yöntemde iki değişken arasındaki bağlantıya en yakın doğrusal sonuç elde edilir. Örnek verilerinin mümkün olan en çoğunu kapsayacak şekilde geçecek bir doğru denklemi elde edilir. Bu denklem ile gelecekteki veriler öngörülür. Öznitelik sayısı birden fazla olursa Multilineer Regresyon olarak isimlendirilir. Lojistik Regresyon ise, çok değişkenli verileri sınıflandırmak için kullanılır [19]. Bağımlı değişkenin kategorik olduğu durumlar için (Örneğin Not durumu iyi veya kötü gibi) çok uygun bir yöntemdir. Örnek bir regresyon modeli Şekil 2'de gösterilmiştir.



Şekil. 2. Örnek Bir Lineer Regresyon Modeli [20]

SVM bir düzlemde bulunan iki sınıfı bir sınır ile ayırarak sınıflandırır [21]. Sınıflandırma için, bir düzlemde bulunan örnekler arasına bir sınır çizerek örnekleri iki gruba ayırır. Bu sınırın çizileceği yer iki grubun da üyelerine en uzak olan yerdir. Sınırın çizilmesi için iki gruba da yakın ve birbirine paralel iki çizgi çizilir ve bu çizgiler birbirine yaklaştırılarak sınır çizgisi üretilir. Özetle veriyi birbirinden ayırmak için en uygun fonksiyon tahmin edilmeye çalışılır [15]. Sınıflandırma için lineer bir sınır bulunamazsa, veriler başka birçok boyutlu uzaya taşınarak sınır aranır.

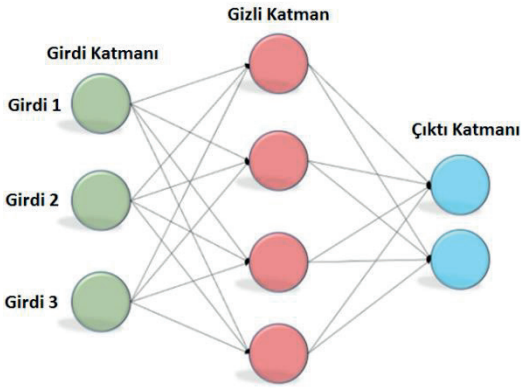
Lineer SVM, SVM'de kullanılan çekirdeğin lineer olarak seçilmiş halidir. SVM yöntemin basit bir yapısının olması ve uygulamalarda yüksek performans sağlaması nedenleri ile kullanımı yaygındır. Örnek bir SVM modeli Şekil 3'te gösterilmiştir.



Şekil. 3. Örnek Bir SVM Modeli [22]

MLP sınıflandırma yönteminde algılayıcılar Yapay Sinir Ağı modelidir [23]. Yapay sinir ağları; insan beynini özelliklerini taklit etmeye dayanır. Öğrenme yolu ile yeni bilgiler türetebilme, yeni bilgiler oluşturabilme ve keşfedebilme gibi yetenekleri otomatik olarak gerçekleştirmeye çalışır [15].

Eşik değerleri kullanılarak sınıflandırma yapılabilir. Çok katmanlı algılayıcılarda, bilgi girişinin yapıldığı girdi katmanı, bir veya daha fazla gizli katman, bir çıktı katmanı ve katmanlar arası ileri-geri yayılım olarak adlandırılan geçişler bulunur. Örnek bir MLP modeli Şekil 4'te gösterilmiştir.



Şekil. 4. Örnek Bir MLP Modeli [24]

4. Veri Ön Hazırlık İşlemleri

Makine öğrenmesi yöntemleri kullanılmadan önce veri kümesi hazırlanmıştır. Veriler Bilkent Üniversitesi Not Giriş Sistemi (STARS-AIRS) üzerinden indirilmiştir. 2017 ve 2018 yıllarındaki toplam 7 şubenin not bilgilerinden oluşan veri kümesinde öğrenci numara, isim, soy isim bilgileri, dönem boyunca gerçekleştirilmiş en az beş kısa sınav notu, iki proje notu, üç ödev notu, bir ara sınav notu, dönem sonu sınav notu,

tüm notların dönem sonu notuna olan ağırlık bilgileri ve dönem sonu notu yer almaktadır.

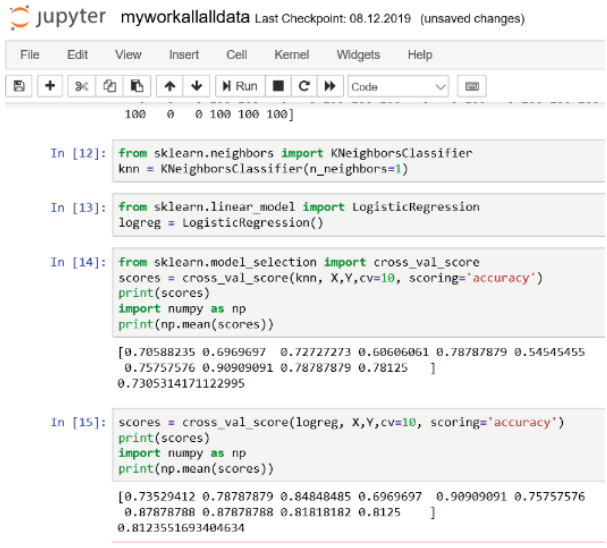
Öncelikle tüm veriler tek dosyada birleştirilmiştir. Daha sonra kullanılmayacak ve gereksiz olan veriler ve öğrencileri tanımlayan tüm bilgiler (numara, isim, soy isim) silinmiştir. Tüm notların dönem sonu notuna olan ağırlık bilgileri de kullanılmayacağından veri kümesinden silinmiştir. Ara sınav notu ile ilgili bir deney şimdilik olmadığından ve bu tür bir deney gelecek çalışma olarak bırakıldığından bu veri sütunu da veri kümesinden çıkarılmıştır. Dönem sonu notu bilgisi belli ağırlıklarla kısa sınav, ödev, proje, ara sınav ve dönem sonu sınavına bağlı bir değişken olduğundan deneylerde dikkate alınmamıştır.

Gereksiz veriler temizlendikten sonra veri satırlarındaki eksiklikler giderilmiştir. Olmayan veri değerleri için sıfır değeri girilmiştir. Örneğin bir kısa sınav notuna katılmayan veya teslim edilmemiş ödevler için öğrencinin ilgili not hücresinde veri yoksa sıfır değeri girilmiştir.

Tüm hazırlık işlemleri sonucunda veri kümesi yapımız en az beş kısa sınav notu, üç ödev notu, iki proje notu ve dönem sonu sınav notu olmak üzere 330 satır ve on bir sütundan oluşmaktadır.

5. Deneyler

Deneyler Python programlama dili [25] ve o dilde hazırlanmış olan makine öğrenmesi kütüphaneleri kullanılarak yapılmıştır. Geliştirme ortamı olarak Anaconda - Jupyter yazılım ortamı [26] kullanılmıştır. Örnek bir çalışma ortamı Şekil 5'te gösterilmiştir. Deneyler iki kısımdan oluşmaktadır. İlk olarak Lineer Regresyon yöntemi ile kısa sınav notlarından dönem sonu sınav notu tahmini yapılmış, daha sonra farklı birçok makine öğrenmesi yöntemleri ile ve farklı öznitelikler kullanılarak özniteliklerin dönem sonu sınav notuna etkisi anlaşılmaya çalışılmıştır.



Şekil. 5. Örnek Çalışma Ortamı

5.1 Lineer Regresyon Yöntemi ile Kısa Sınav Notlarından Dönem Sonu Sınav Notu Tahmini

Deney için "qort" ve "kator" isimli iki öznitelik kullanılmıştır. Veri kümesindeki şubeden şubeye değişen ve en az beş tane yapılmış kısa sınav notlarından ortalama ve yüzde olarak katılım oranı hesaplanmıştır. "qort", öğrencilerin kısa sınav notlarının ortalamasını, "kator" ise öğrencilerin kısa sınavlara katılım oranını temsil etmektedir. Belirtilen iki öznitelik ayrı ayrı ve birlikte kullanılarak toplam üç deneyde tahmin edilmeye çalışılacak bilgi ise dönem sonu sınav notudur. Veri kümesinden ilgili sütunlar oluşturulduğunda oluşan örnek veri Tablo I'de gösterilmiştir.

TABLO I. İlk Deney İçin Hazırlanmış Veri Örneği
(2 öğrenci için)

Veri No	qort	kator	Final
0	86	88	39
1	97	75	39

Bu verilerin birbirleri ile olan korelasyonu ise Tablo II'de gösterilmiştir.

TABLO II. Özniteliklerin Korelasyon Matrisi

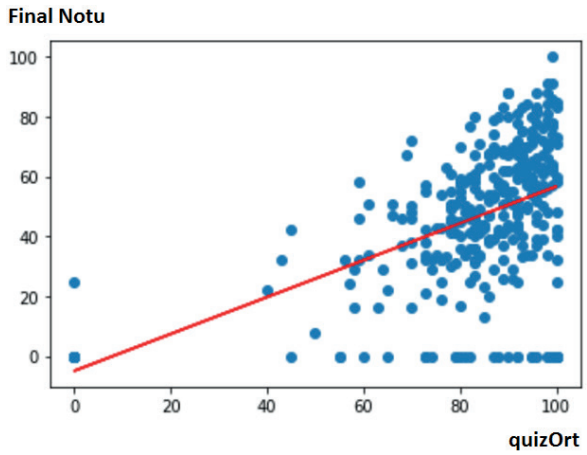
	qort	kator	Final
qort	1	0.406	0.470
kator	0.406	1	0.571
Final	0.470	0.571	1

5.1.1 İki Öznitelik Birlikte Kullanılarak Multilineer Regresyon Yöntemi İle Final Notu Tahmini

İlk iki öznitelik ve dönem sonu sınavından oluşan veri kümesi %80 i öğrenme verisi %20 si test verisi olarak bölünmüştür. Öğrenme verisi kullanılarak Multilineer Regresyon yöntemiyle model öğrenildikten sonra model test verisi ile denendiğinde ortalama skor sonucu 0.43 olarak bulunmuştur.

5.1.2 Kısa Sınav Notu Ortalaması Özelliği Kullanılarak Lineer Regresyon Yöntemi İle Final Notu Tahmini

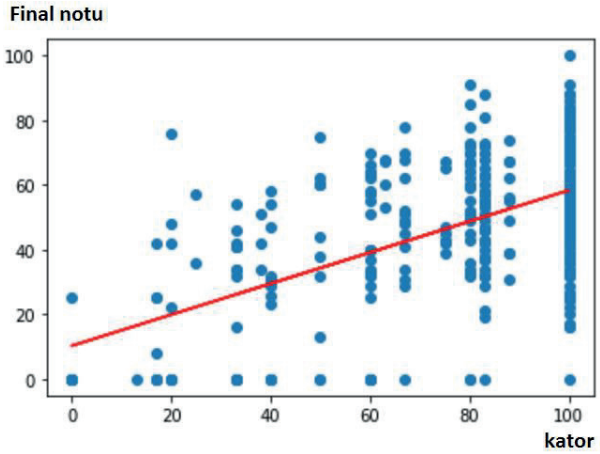
Kısa sınav notu ortalaması özniteliği ile model öğrenildikten sonra test edildiğinde başarı olasılığı0.22 olarak bulunmuştur. Elde edilen modelin katsayıları yaklaşık olarak -4.96 ve 0.62 olarak bulunmuştur. Model için bulunmuş regresyon doğrusu *Final notu = -4.96 * quizOrt + 0.6* şeklindedir ve Şekil 6'da gösterilmiştir.



Şekil. 6. Kısa Sınav Notu Ortalaması Özelliği İle Oluşturulmuş Denklem

5.1.3 Kısa Sınavlara Katılım Oranı Özelliği Kullanılarak Lineer Regresyon Yöntemi İle Final Notu Tahmini

Kısa sınavlara katılım oranı özelliği ile model öğrenildikten sonra test edildiğinde başarı olasılığı0.33 olarak bulunmuştur. Elde edilen modelin katsayıları yaklaşık olarak 10.29 ve 0.48 olarak bulunmuştur. Model için bulunmuş regresyon doğrusu *Final notu = 10.29 * kator + 0.6* şeklindedir ve Şekil 7'de gösterilmiştir.

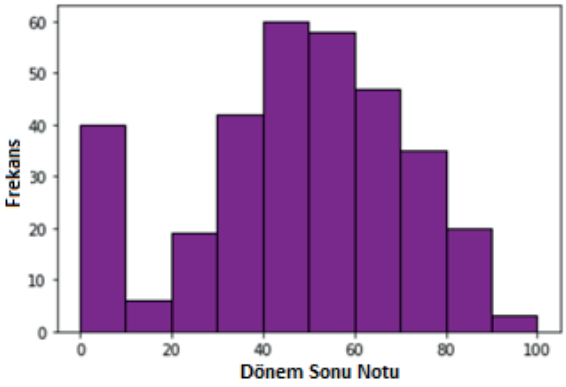


Şekil. 7. Kısa Sınavlara Katılım Oranı Özelliği İle OluşturulmuşDenklem

5.2 Makine Öğrenmesi Yöntemleri ile Farklı Özniteliklerin Dönem Sonu Sınav Notuna Etkisi

Bu kısımda anlatılan deneylerde makine öğrenmesi yöntemleri ile ve farklı öznitelikler kullanılarak özniteliklerin dönem sonu sınav notuna etkisi araştırılmıştır. Veri kümesindeki üç farklı yarıyıla ait toplam yedi şubenin yapılan kısa sınav sayıları değişkenlik göstermektedir. Tüm şubelerde en az beş kısa sınav yapıldığından beşten fazla kısa sınav yapılan şubeler için ilk beş kısa sınav notu dikkate alınmıştır.

Beş kısa sınav notu, üç ödev notu, iki proje notu ve dönem sonu sınav notu sütunlarından oluşan veri kümesi 330 öğrenciye ait veriden oluşmaktadır. Veri sayısının az oluşu nedeni ile yapılacak dönem sonu sınav notu iki sınıfa ayrılmıştır. Dönem sonu sınav notu ortalamasından büyük olan değerler için "yüksek", kalan değerler için "düşük" olmak üzere iki sınıf belirlenmiştir. Öğrencilerin dönem sonu sınav not dağılımını gösteren histogram Şekil 8'de gösterilmiştir.



Şekil. 8. Dönem Sonu Sınav Notu Dağılımı

Kısa sınav notu, ödev notu ve proje notu öznitelikleri ile ayrı ayrı ve birlikte kullanılarak farklı makine öğrenmesi yöntemleri sonucunda dönem sonu sınav notu sınıflandırılmıştır. Deneylerde kullanılan öznitelikler Tablo III'te gösterilmiştir.

TABLO III. Deneylerde Kullanılan Öznitelikler

Deney No	Öznitelik Sayısı	Öznitelikler
1	2	Kısa Sınav Notu Ortalaması, Kısa Sınav Katılım Ortalaması
2	5	5 Kısa Sınav Notu
3	3	3 Ödev Notu
4	2	2 Proje Notu
5	10	5 Kısa Sınav, 3 Ödev, 2 Proje Notu

Sınıflandırma sonuçları 10 katlı çapraz değerleme test yöntemi ile elde edilmiştir. Elde edilen sınıflandırma sonuçları kullanılan öznitelikler ve makine öğrenmesi yöntemlerine göre yüz üzerinden Tablo IV'te gösterilmiştir.

TABLO IV. Sınıflandırma Sonuçları

Deney No	**1**	**2**	**3**	**4**	**5**	**Yöntem Ortalaması**
KNN	69	69	72	72	73	71
Lojistik Regresyon	75	76	80	76	81	78
Lineer Regresyon	60	70	78	69	71	70
SVM	75	70	77	72	68	72
MLP	69	77	66	68	52	66
Deney Ortalaması	70	72	75	71	69	

Tablo IV'te gösterildiği gibi makine öğrenmesi yöntemlerinin deney sonuç ortalamaları ve deneylerin ortalaması belirtilmiştir. Buna göre tüm deneylere göre en iyi performans ortalaması %78 değerle Lojistik Regresyon yöntemidir. Deneylerin yöntemlere göre ortalamasına göre %75 ile ödev notlarının öznitelik olarak kullanıldığı üç numaralı deney en iyi performans ortalamasını göstermiştir. Tüm deneyler ve makine

öğrenmesi yöntemlerine göre en iyi performans, tüm özniteliklerin kullanıldığı Lojistik Regresyon makine öğrenmesi yöntemi ile %81 olarak elde edilmiştir.

6. Sonuçlar ve Değerlendirme

Bu çalışmada kısa sınavların, ödevlerin ve projelerin dönem sonu sınavına olan etkisi araştırılmıştır. Bilkent Üniversitesi Bilgisayar Mühendisliği Bölümü'nde verilen Programlama Dilleri dersini farklı öğretmenlerden almış toplam 330 öğrenci verisi ele alınmıştır. Bu veri kümesinde ön hazırlık işlemleri yapılmış, ilgili öznitelikler elde edilmeye çalışılmıştır. Öncelikle öğrencilerin kısa sınav bilgilerinden elde edilen öznitelikler ile Lineer Regresyon yöntemi ile dönem sonu sınavı denklemi oluşturulmaya çalışılmıştır. Sonrasında öğrencilerin kısa sınav, ödev ve proje bilgileri ile bu bilgilerden bağımsız olan dönem sonu sınav notu KNN, Lineer Regresyon, Lojistik Regresyon, Lineer SVM, SVM ve MLP makine öğrenmesi yöntemleri ile sınıflandırılmaya çalışılmıştır. Bu kısımda her öznitelik kümesine ve uygulanan her makine öğrenmesi yöntemine göre elde edilen sonuçlar karşılaştırılmış ve analiz edilmiştir. Ayrıca kullanılan her bir makine öğrenmesi yöntemi hakkında bilgiler verilmiştir.

Veri boyunun büyüklüğü dikkate alındığında elde edilen sonuçlar tatmin edicidir. Sonuç olarak dönem sonu sınav sonucu için oluşturulan lineer denklemin başarısı düşük olsa da seçilen özniteliklerin dönem sonu sınav notu ile ilgili olduğu ve bu özniteliklere göre dönem sonu sınav notunun başarılı olarak sınıflandırıldığı gözlemlenmiştir.

Gelecek çalışma olarak, başka derslerden de benzer bilgilerin toplanarak veri kümesinin arttırılması ve bu sayede dönem sonu sınav sonucu için oluşturulan denklemin iyileştirilmesi planlanmaktadır. Ayrıca ara sınav notu üzerinde de analizler yapmak ve kısa sınav, ödev ve proje bilgilerini ara sınav öncesi ve sonrası ayırarak daha detaylı analizler yapmak hedeflenmiştir. Başka ders verilerine göre elde edilmiş sonuçları da birbirleri ile karşılaştırmak yine geliştirilmesi planlanan başka bir hedeftir. Son olarak araştırılmak istenen bir diğer çalışma da kısa sınav, ödev ve proje bilgilerinin her birinin dönem sonu sınav notuna pozitif mi yoksa negatif mi etkisi olduğunu ölçmeye çalışmaktır.

Kaynaklar

[1] Rao, K. Prasada, M. C. Rao ve B. Ramesh. "Predicting learning behavior of students using classification techniques", International Journal of Computer Applications, 139.7: 15-19, 2016.

[2] Al-Radaideh, Q. A., Al-Shawakfa, E. M. ve Al-Najjar, M. I, "Mining student data using decision trees." International Arab Conference on Information Technology, Yarmouk University, Jordan. 2006.

[3] Okubo, F., Yamashita, T., Shimada, A., Taniguchi, Y. ve Konomi, S., "On the prediction of students' quiz score by recurrent neural network", CEUR Workshop Proceedings, 2163, 2018.

[4] Iqbal, Z., Qadir, J., Mian, A. ve Kamiran, F., "McGraw Hill, 199: A Case Study", 2017.

[5] Udemy, https://www.udemy.com/, [Erişim: 25.05.2020].

[6] Gadhavi, M. ve Chirag, P., "Student final grade prediction based on linear regression" Indian J. Comput. Sci. Eng. 8.3: 274-279, 2017.

[7] Yadav, S. K. ve Pal, S., "Data mining: A prediction for performance improvement of engineering students using classification", World of computer science and information technology journal, 2(2), pp. 51-56, 2012.

[8] Khan, B., Khiyal, M. S. H. ve Khattak, M.D., "Final grade prediction of secondary school student using decision tree.", International Journal of Computer Applications 115.21, 2015.

[9] Agrawal, H. ve Mavani, H., "Student performance prediction using machine learning." International Journal of Engineering Research and Technology, 4.03: 111-113. 2015.

[10] Kim, B., Vizitei, E.ve Ganapathi, V., "GritNet: Student performance prediction with deep learning." arXiv preprint arXiv:1804.07405, 2018.

[11] Pandey, U. K. ve Saurabh, P., "Data Mining: A prediction of performer or underperformer using classification", International Journal of Computer Science and information technology, 2(2), pp. 686-690, 2011.

[12] Mitchell, T., "Machine Learning", McGraw Hill, 1997.

[13] Nabıyev, Vasif V., "Yapay zeka: problemler-yöntemler-algoritmalar", Seçkin Yayıncılık, 2005.

[14] Öztemel, E., "Yapay sinir ağları", Papatya Yayıncılık, İstanbul, 2003.

[15] Karakoyun, M. ve Hacıbeyoğlu, M.,"Biyomedikal Veri Kümeleri ile Makine Öğrenmesi SınıflandırmaAlgoritmalarının İstatistiksel Olarak Karşılaştırılması", Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 2014.

[16] Liao, Y.ve Vemuri, V. R., "Use of k-nearest neighbor classifier for intrusion detection", Computers ve security, 21(5), 439-448, 2002.

[17] KNN Classifier, https://www.mathworks.com/matlabcentral/fileexchange/63621-knn-classifier?focused=7902851&tab=function, [Erişim: 25.05.2020].

[18] Seber, G. A.ve Lee, A. J., "Linear regression analysis", Vol. 329, John Wiley & Sons, 2012.

[19] Wright, R. E., "Logistic regression", 1995.

[20] How to do Linear Regression and Logistic Regression in Machine Learning?, https://mlfromscratch.com/machine-learning-introduction-8-linear-regression-and-logistic-regression/#/, [Erişim: 25.05.2020].

[21] Shawe-Taylor, J. ve Cristianini, N., "Support vector machines. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods", 93-112, 2000.

[22] Support Vector Machines, https://towardsdatascience.com/support-vector-machines-soft-margin-formulation-and-kernel-trick-4c9729dc8efe, [Erişim: 25.05.2020].

[23] J. A. Freeman ve D. M. Skapura, "Neural Networks Algorithms", Applications and Programming Techniques. New York, USA: Addison-Wesley Publishing Company, 1991.

[24] Multi Layer Perceptron (MLP) Models on Real World Banking Data, https://becominghuman.ai/multi-layer-perceptron-mlp-models-on-real-world-banking-data-f6dd3d7e998f, [Erişim: 25.05.2020].

[25] Python Programlama Dili, https://www.python.org, [Erişim: 25.05.2020].
[26] Anaconda Yazılım Ortamı, https://www.anaconda.com, [Erişim: 25.05.2020].

Karani KARDAŞ



Karani Kardaş, 1981 yılı Diyarbakır doğumludur. İlköğrenimini Diyarbakır'da orta ve lise öğrenimini ise İzmir'de tamamladı. 2003 yılında Ege Üniversitesi Bilgisayar Mühendisliği Bölümü'nden lisans, sırasıyla 2007 ve 2018 yıllarında ODTÜ Bilgisayar Mühendisliği Bölümü'nden yüksek lisans ve doktora derecelerini aldı. 2018 yılında Bahçeşehir Üniversitesi İşletme Bölümü'nden yüksek lisans derecesini aldı. 2003 yılından beri Havelsan'da Coğrafi Bilgi Sistemleri konusu ağırlıklı olmak üzere farklı projelerde yer aldı. Coğrafi Bilgi Sistemleri ve Makine Öğrenmesi; üzerinde çalıştığı konulardandır. Ayrıca Bilkent Üniversitesi'nde Yarı Zamanlı Öğretim Görevlisi olarak Programlama Dilleri ve Veritabanı Yönetim Sistemleri derslerini vermektedir.

Prof. Dr. H. Altay GÜVENİR



H. Altay Güvenir, 1957 yılında Bursa'da doğdu. İlk, orta ve lise öğrenimini Bursa'da tamamladı. Lisans ve yüksek lisans eğitimini İstanbul Teknik Üniversitesi, Elektrik Fakültesi, Elektronik ve Haberleşme Bölümünde, sırasıyla 1979 ve 1981 yıllarında tamamladı. 1979 -1982 yılları arasında İTÜ Elektrik Fakültesi, Telekomünikasyon kürsüsünde asistan olarak görev yaptı. Kazandığı Milli Eğitim Bakanlığı doktora bursuyla, 1982-1987 yılları arasında, ABD'deki Case Western Reserve Üniversitesinde Bilgisayar Mühendisliği ve Bilimleri alanında doktora çalışmalarını tamamladı. 1987-1988 yılları arasında Hacettepe Üniversitesi Bilgisayar Bilimleri ve Mühendisliği Bölümünde yardımcı doçent olarak görev yaptı. 1988 yılında yardımcı doçent olarak çalışmaya başladığı İ. D. Bilkent Üniversitesinde 1994 yılında doçent 2001 yılında profesör oldu. 2001 – 2019 yılları arasında Bilgisayar Mühendisliği Bölümünün başkanlığını yaptı. 1999 yılında TÜBİTAK Hüsamettin Tuğaç Vakfı Birincilik Ödülünü, 2020 yılında İ. D. Bilkent Üniversitesi Eğitimde Üstün Başarı Ödülünü aldı. 2015 yılından buyana Bilkent-TÜSİAD Bilgi Toplumu Forumu'nun direktörlüğünü yürütmektedir. Araştırma alanları yapay zekâ ve makine öğrenmesidir.