

Sınıf Öğretmenliği Öğrencilerinin Mezuniyet Notlarının Regresyon Analizi ve Yapay Sinir Ağları Yöntemleriyle Tahmini¹

Murat Altun
Akdeniz Üniversitesi-Türkiye
emurataltun@gmail.com

Doç. Dr. Kemal Kayıkçı
Akdeniz Üniversitesi-Türkiye
kemalkayikci@akdeniz.edu.tr

Doç. Dr. Sezgin Irmak
Akdeniz Üniversitesi-Türkiye
sezginakdeniz@edu.tr

Özet:

Bu çalışmada Akdeniz Üniversitesi Eğitim Fakültesi Sınıf Öğretmenliği Bölümü'nden 2012-2017 yılları arasında mezun 578 öğrencinin cinsiyet, medeni durum, kayıt yaşı ve 1. sınıf 1. dönem ara sınav puanları verisi kullanılarak mezuniyet notlarını tahmin etmek için modeller oluşturulmuştur. Anonimleştirilmiş öğrenci verisi üzerinde çoklu doğrusal regresyon analizi ve yapay sinir ağları ile oluşturulan modellerin çapraz değerlendirme sonuçlarında öğrenci başarısını kestirmede birbirine yakın sonuçlar verdiği görülmüştür. Ortalama mutlak hata yüzdesi değerleri baz alınarak yapılan değerlendirmede regresyon analizi modeli %94.30 başarı sağlarken yapay sinir ağları modeli %94.43 başarı sağlamıştır. Modeller öğrenci başarısına etki eden faktörlerin ağırlıklarını belirlemek için bulgular sağlamıştır. Araştırma kapsamında geliştirilen modellerin etkili olduğu eğitimde toplam kalite yönetimi bağlamında sıfır hata hedefinde proaktif pozisyon alarak öğrenci başarısını artırma gibi çalışmalarında kullanılabileceği söylenebilir.

Anahtar Kelimeler: Sınıf öğretmenliği, yapay sinir ağları, çoklu doğrusal regresyon, eğitsel veri madenciliği



**E-Uluslararası Eğitim
Araştırmaları Dergisi,**
Cilt: 10, Sayı: 3, 2019, ss.29-43

DOI: 10.19160/ijer.624839

Gönderim : 06.09.2019
Kabul : 28.10.2019

Önerilen Atıf

Altun, M., Kayıkçı, K., ve Irmak S. (2019). Sınıf Öğretmenliği Öğrencilerinin Mezuniyet Notlarının Regresyon Analizi ve Yapay Sinir Ağları Yöntemleriyle Tahmini, *E-Uluslararası Eğitim Araştırmaları Dergisi*, Cilt: 10, Sayı: 3, 2019, ss. 29-43, DOI: 10.19160/ijer.624839

¹ Akdeniz Üniversitesi Bilimsel Araştırma Projeleri Birimi tarafından desteklenmiştir (Proje Numarası: SDK-2018-4077). Kemal Kayıkçı danışmanlığında Murat Altun doktora tezi kapsamında hazırlanmıştır

GİRİŞ

Tüm örgütlerde olduğu gibi eğitim örgütleri de amaçları için vardır. Bundan dolayıdır ki örgütler, amaçlarını gerçekleştirebildikleri ölçüde varlıklarını sürdürebilirler. Örgütün amaçlarını gerçekleştirmesi ise öncelikle yönetimin sorumluluğundadır. Farklı bir anlatımla örgütler, "etkili" ve "yeterli" oldukları sürece varlıklarını sürdürebilirler (Aydın, 2014). Yönetim, örgütün amaçlarını gerçekleştirmek için insan ve madde kaynaklarını etkili ve verimli bir şekilde kullanmaktır (Bursalıoğlu, 1991). Ancak yönetimin bunu gerçekleştirmesi, yönetim biliminin ortaya koyduğu ilke, yöntem, teknik ve kuramlara uygun bir yönetsel anlayışı izlemesine bağlıdır. Toplumsal kültürü geliştirme ve yeni nesillere aktarma işlevi olan eğitim örgütlerinin (Uras, 2016) bu işlevini yerine getirebilmesi için sürekli kendini yenilemesi ve geliştirmesi beklenmektedir. Eğitim örgütlerinin bu beklentilerini gerçekleştirmesi, örgüt ve yönetim alanındaki bilimsel ve teknolojik gelişmelere ayak uydurmasına bağlıdır. Son zamanlarda küreselleşmenin de etkisine giren eğitim örgütleri kendilerini rekabet ortamı içerisinde bulmuşlardır. Küreselleşme ve yeni dünya düzenine uyum sağlamaya çalışan örgütlerin yönetsel anlayışları da buna uygun olarak değişmiş ve yenilenmiştir. Küresel pazarda başarı için sürekli iyileştirme esastır. Küresel pazar rekabeti olimpiyatlardaki yarışmalara benzer. Sürekli kendini geliştirmeyen atlet uzun dönemde yarışı önde yarışı bitiremez, bazen yarış çizgisine dahi ulaşamaz (Öztürk, 2009). Küreselleşme ve bilimsel gelişmeler doğrultusunda örgüt ve yönetim alanında öğrenen örgütler, toplam kalite yönetimi ve insan kaynakları yönetimi anlayışları (Toprakçı, 2017) ön plana çıkmıştır. Rekabete dayalı örgüt anlayışında az maliyet ve yüksek performans hedeflenmektedir. Toplam kalite yönetimi anlayışı "sıfır hata", süreç yönetimi "bilimsel yöntem ve istatistiksel düşünme" ve sürekli gelişim anlamına gelen "kaizen" ilkelerine dayanmaktadır (Kıngır, 2013; Öztürk, 2009). Çağdaş yönetim anlayışında yöneticilerin örgütlerini rekabet ortamında etkili ve verimli bir şekilde amacına ulaştırabilmeleri için toplam kaliteye dayalı örgütsel işleyişte tercih ettikleri yaklaşımlardan biri de "sıfır hata" için sürekli süreç iyileştirmeye dayalı süreç yönetimidir. Sürekli süreç iyileştirmek için süreci izleme, süreci kontrol etme (denetleme) ve süreci iyileştirme evreleri sırasıyla uygulanmalıdır. Sürekli süreç iyileştirmek için Planla, Uygula Kontrol Et ve Önlem al (PUKÖ) şeklinde ifade edilen Deming döngüsü kullanılabilir (Öztürk, 2009). PUKO döngüsüne dayalı sürekli denetim anlayışı sayesinde örgütsel işleyişte oluşan hedeften sapmalar kısa sürede tespit edilerek düzeltilmeleri için gerekli önlemler alınır, eksiklikler giderilir. Böylece sapma ve hataların devamından dolayı oluşacak zaman, emek, para ve diğer harcamaların boşa gitmesi veya ileride telafisi mümkün olmayan zararların olması önlenerek, örgütsel işleyişin amaca yönelik olması, daha etkili ve verimli bir işleyişin gerçekleşmesi sağlanabilir. Sapmaları önceden tespit ederek önlem almak ve süreci geliştirmek çağdaş eğitim denetiminin temel amacıdır (Aydın, 2007). Bu anlayış aynı zamanda insan kaynakları yönetiminin ilkelerinden birisi olan reaktif-tepkisel (sorunlar olduktan sonra çözümlenmeye çalışmak) yerine proaktif-önleyici (sorunlar oluşmadan tahmin ederek önlem alarak olmasını önlemek) örgüt anlayışını ifade eder (Tortop, Aykaç, Yayman ve Özer, 2007; Bingöl, 2010).

Örgütler proaktif pozisyonlar almak için özellikle bilgi ve iletişim teknolojilerinden, veri üretiminden ve veriye yönelik yeni tekniklerden yararlanmaktadır. İnternet bant genişliklerinin artması, kullanıcı sayısının artması, sosyal medya, akıllı cihazlar ve otomasyon sistemleri üretilen veri miktarını artırırken maliyeti azalan, erişimi kolaylaşan veri depolama aygıtları depolanan veri miktarını artırmaktadır. Bu değişim sadece veri miktarı artışında değil aynı zamanda değişik formlarda ve değişik biçimlerdeki verinin (görüntü, ses, video, metin, sensor değerleri) üretilmesi ve depolanmasında da kendini göstermektedir. İşlemci hızının ve bellek kapasitelerinin artması çok büyük miktarda veri üzerinde önceden uygulanması zor olan bazı istatistiksel yöntemlerin, matematiksel modellerin ve algoritmaların kullanılmasına ve bazı yeni teknikler geliştirilmesine olanak sağlamıştır. Büyük veri (big data) üzerinden değişik yaklaşımlar geliştirerek rekabet avantajına çevirmek için bilgi elde etmek temel motivasyonlardan biridir (Walter, 2000; Şeker, 2014; IDC, 2016). Veriye ve veriden bilgi çıkarılması sürecine farklı bakış açısıyla ortaya çıkan

veriden bilgiye giden tüm teknikleri çatısı altında birleřtiren Veri Madencilięi (VM) için alanyazında veri tabanlarında bilgi madencilięi (knowledge mining from databases), bilgi çıkarımı (knowledge extraction), veri ve örüntü analizi (data/pattern analysis), veri arkeolojisi gibi birçok kavram kullanılmaktadır. Bu kavramlardan veri tabanlarından bilgi keřfi KDD (Knowledge Discovery from Databases) en yaygın kullanılanıdır (Akgöbek ve Kaya, 2011). Bilgi keřfi süreci olarak adlandırılan veri madencilięi sürecinde hedef genelde "büyük miktardaki" veride gizli olan ve kolayca görülemeyen bilgilerdir (Fayyad, Piatetsky-Shapiro ve Smyth, 1996). Veri madencilięi veride/veri tabanlarında potansiyel örüntüleri keřfetme sürecidir. Üzerinde çalıřılan veri miktarı düşünöldüğünde manuel bir süreç deęildir. Veri madencilięi sonucunda elde edilen örüntüler mantıklı olmalı ve bir yarar saęlamalıdır (Witten ve Frank, 2005). Veri madencilięi anlamlı korelasyonları, örüntüleri ve eğilimleri keřfetme sürecidir ve örüntü tanıma teknolojilerinin yanı sıra, istatistiksel (regresyon, navie bayes), matematiksel teknikler kullanır (GG, 2019). Veri madencilięi tek başına bir çözüm deęil veri analiz teknikleri bütünüdür. Mevcut problemleri çözmek, kritik kararları almak veya geleceęe yönelik tahminleri yapmak için gerekli olan bilgileri elde etmeye yarayan bir araçtır (Argüden ve Erřahin, 2008).

Günümüzde birçok alanda yaygın bir řekilde kullanılan veri madencilięi karar verme, strateji oluřturma gibi kritik konularda yöneticilere bilgi kaynaęı olarak destek saęlayabilir. Veri madencilięinin kullanım alanlarından biri de eğitimdir. Türkiye'de YÖK Tez Kurulu web sayfasından yapılan taramada veri madencilięi ile ilgili eğitim alanında yapılan çalıřmaların eğitsel veri madencilięi, eğitimsel veri madencilięi ve eğitimde veri madencilięi adları ile kullanıldığı görölmektedir (YÖK, 2019). Geleneksel eğitimde ve uzaktan eğitimde veri madencilięi çalıřmaları, eğitimde veri madencilięi alanında yapılan çalıřmaları inceleme ve tanıma, öğrencilerin akademik başarı ve başarısızlıkları ile bunları etkileyen faktörlerin belirlenmesi, öğrenci özelliklerini belirleme ve bunlara göre öğrencilerin üzerine yoğunlařmış durumdadır (Özbay, 2015). Eğitsel veri madencilięi, eğitim ortamlarından özgün, büyük ölçekli veriyi keřfetmek için yöntemler geliřtiren ve bu yöntemleri öğrencileri ve öğrenme ortamlarını daha iyi anlamak için kullanan, yeni bir disiplindir (IEDMS, 2019). řekilde gösterildięi gibi eğitsel veri madencilięi eğitim sistemlerden elde edilen ham verinin eğitim yazılımlarının, geliřtiricilerin, öğretmenlerin ve arařtırmacıların kullanabileceęi bilgiye çevirme süreci olarak tanımlanmaktadır (García, Romero, Ventura ve de Castro, 2011). Eğitsel veri madencilięi eğitim örgütlerinin etkililięini ve verimlilięini artırmada yöneticilere ihtiyaç duydukları veriye dayalı bilgiyi saęlamak için kullanılabilirler. Eğitsel veri madencilięi ile öğrenci verinin analiz edilmesi, öğrenci başarı ve başarısızlık nedenlerinin tespit edilmesi, öğrenci başarılarının arttırılması, eğitim-öğretim ortamlarındaki aksaklıkların tespit edilmesi, daha etkili eğitim-öğretim ortamlarının oluřturulmasında yararlanılmaktadır (Özbay, 2015). Eğitsel veri madencilięi ile öğrenci akademik performansına yönelik tahmin modelleri oluřturulabilir ve bu tahmin modelleri öğrenciye rehberlik yapmak için kullanılabilir (Bienkowski, Feng ve Means, 2012). Eğitsel veri madencilięi ve öğrenme analitięi gibi yeni yaklařımlarla analiz edilmesi sayesinde öğrenci profillerinin (öğrenme stillerinin, davranıřlarının ve akademik başarı durumlarının) modellenmesi, benzer öğrenci profillerinin kümeleneşine olanak vererek öğrencilere kişiselleřtirilebilir eğitim ortamları, öğretim programları ve öğrenme stilleri sunulabilir (Bienkowski, Feng ve Means, 2012). Eğitsel veri madencilięi ile geliřtirilen modeller sayesinde geleneksel yöntemlerle çok fazla zaman ve emek gerektiren eğitsel iřlerin kısa zamanda ve otomatik olarak yapılması saęlanabilir (Lopez, Luna, Romero ve Ventura, 2012).

Arařtırma konusuyla ilgili alan yazında yapılan çalıřmalar incelendięinde eğitsel veri madencilięi ile öğrenci profilini belirleme, öğrenci başarısını tahmin etme veya okulu bırakma riskini belirlemeye yönelik çalıřmaların olduęu görölmektedir. Ayesha, Mustafa, Sattar ve Khan (2010) tarafından yapılan çalıřmada öğrenci bilgilerinden onların sınav notları tahmin etmek için bir model oluřturulmuřtur. Bu modelin final sınavından önce başarı durumu zayıf olabilecek öğrencilerin belirlenmesi ve performanslarının yükseltilmesi için gerekli önlemlerin alınabilmesi için kullanılması amaçlanmıřtır. Baradwaj ve Pal (2011) tarafından yapılan bir çalıřmada

Hindistan'da yer alan Purvanchal üniversitesinin Bilgisayar Uygulamaları Bölümü'nde gerçekleştirilmiş, bilgisayar uygulamaları dersindeki akademik başarıyı tahmin etmişlerdir. Başarı üzerinde etkili olan faktörleri saptamışlardır. [Tsai C-F., Tsai C-T., Hung ve Hwang \(2011\)](#) tarafından Tayvan'da bir üniversitede yapılan bir çalışmada öğrencilerin bilgisayar yeterlilik sınavı sonuçları tahmin edilmiştir. [Cunningham \(2017\)](#) tarafından yapılan çalışmada Kitlelesel Çevrimiçi Açık Kurs (Massive Open Online Courses) çevrim içi kurslara kayıtlı her bir öğrencinin çevrimiçi kursun ilk çalışma gününe dayanarak, öğrencilerin kursu ne zaman tamamlayacağını yaklaşık % 80'ini doğru tahmin edebilecek modeller geliştirilmiştir. [Whitlock \(2018\)](#) tarafından yapılan çalışmada Üniversiteden ayrılacak öğrencileri tanımlamak için öngörücü modeller geliştirmiştir. Geliştirilen tüm modellerde rastgele şansın ötesinde tahmin gücü olduğu görülmüştür. [Benton \(2018\)](#) tarafından yapılan çalışmada, üniversite öncesi demografik, akademik ve sosyal özellikleri kullanan modellerin Beyaz ve Afrika kökenli Amerikalı öğrencilerin birinci sınıf akademik başarıları ilişkisini belirlemeye yönelik modeller geliştirilmiştir. [Alsuwaiket \(2018\)](#) tarafından yapılan çalışmada öğrenciler, öğretmenler ve okul müdürleri için TIMSS 2015 4. Sınıf geçmiş geçmiş anketlerinden gelen maddeleri kullanarak matematikte öğrencinin başarısını tahmin etmek modeller oluşturulmuştur. [Akçapınar \(2014\)](#) doktora tezi çalışmasında 76 öğrencinin çevrimiçi öğrenme ortamlarına ait veri kullanılarak öğrencilerin derse ilişkin akademik performanslarını "geçti/kaldı" olarak *sınıflandıran*, dersten başarısız olacakları tahmin eden, benzer öğrenci profiline sahip öğrencileri *sınıflandıran* bir model geliştirmiştir. [Aydın \(2015\)](#) öğrencinin başarı durumu ile ilgili olarak dersin adı, e-hizmet faydalanma süreleri, öğrencinin dersi kaçınıcı kez aldığı, değerlendirilen sınavlarının ortalaması, öğrenci yaşı, deneme sınavlarında doğru cevapladığı soru sayısı, deneme sınavlarındaki başarı oranı gibi değişkenler kullanılarak öğrencilerin derslerdeki başarı durumları değişik *sınıflama* algoritmalarıyla tahmin edilmiştir. [Kılınç \(2015\)](#) tarafından yapılan çalışmada mevzuattaki başarısızlık nedeniyle öğrencilikten çıkarılma politikasının, öğrencilerin parasal durumlarının ve demografik özelliklerinin, etkilerinin eğitsel veri madenciliği teknikleriyle incelenmiştir. Eskişehir Osmangazi Üniversitesi Bilgisayar Mühendisliği Bölümü'nde 2008'den 2011'e kadar olan dört yıl içinde birinci sınıf derslerine başlayan öğrencilerin verisi kullanılmıştır. Uygulanan yöntemler sonucunda, atılma politikasındaki değişiklik ile öğrencilerin not durumları arasındaki ilişki bulunmuş, öğrencilerin eğitim sürelerinin, burs veya kredi alınmasıyla değişiklik gösterdiği ve öğrencilerin parasal durumlarıyla annelerinin meslekleri arasında bağlantı olduğu görülmüştür.

Çağdaş örgüt ve yönetim kuramları doğrultusunda etkili ve verimli bir kalite yönetimi için gerekli ilkeler arasında "bilimsel yöntem ve istatistiksel düşünme", "proaktif yönetim anlayışı", "sıfır hata", "sürekli süreç geliştirme- Kaizen" yer almaktadır. Bu doğrultuda Akdeniz Üniversitesi'nin bilimsel ve istatistiksel bir yöntemle sürekli olarak süreç iyileştirme ve proaktif bir yaklaşımla öğrencilerin gelecekteki başarı-başarısızlık durumlarını tahmin ederek bu konudaki eğitim ve yönetim sürecinde iyileştirmeye gitmesi büyük önem taşımaktadır. Bu çalışma, sınıf öğretmenliği bölümü öğrencilerinin mezuniyet notlarının bilimsel ve istatistiksel bir yöntem olan çoklu doğrusal regresyon ve yapay sinir ağları yöntemleriyle tahminini sağlayarak öğrencilerin gelecekteki olası başarısızlıklarını proaktif bir yaklaşımla önlemeye yardımcı olması ve üniversitede bu anlayışı yaygınlaştırmada ilk örnek olması açısından önemlidir. Yapılan çalışmalar öğrenci başarısına yönelik veri bazlı modellemelerin başarılı sonuçlar verdiği ve eğitsel veri madenciliği teknikleri altında oluşturulan modellerin işe yaradığını göstermektedir. Öğrencilere ilişkin verinin çoklu doğrusal regresyon ve yapay sinir ağları teknikleriyle modellenmesi ve sonuçların veri üzerinde değerlendirilerek modelin performansının ölçülmesi akademik başarının artırılmasında önemli bir rol oynayabilir.

Bu araştırmanın amacı Akdeniz Üniversitesi Eğitim Fakültesi Sınıf Öğretmenliği Bölümü'ndeki öğrencilerin akademik ve demografik verisini kullanarak mezuniyet notlarını tahmin edecek modeller geliştirmektir.

YÖNTEM

Arařtırma modeli:

Arařtırmada veri madencilięi tekniklerinden çoklu doęrusal regresyon analizi ve yapay sinir aęları kullanılmıřtır. Regresyon teknięi bir veri öęesini, gerçek deęerli tahmin deęiřkenine eřlemek için kullanılır. Bu fonksiyonun öęrenilmesini kapsar. Temel varsayım hedef verinin bilinen bir fonksiyon tipine (lineer, lojistik) uymasdır. Regresyon teknięinde daha sonra verilen veriyi en iyi modelleyen fonksiyon bulunmaya çalıřılır. Gerçek sonuçlar ile tahminler arasındaki farkı tespit eden hata analizi hangi fonksiyonun en iyi olduęuna karar vermede kullanılır (Dunham, 2013). Yapay sinir aęları insan beynindeki nöronlara çok benzeyen birbirine baęlı düęümleri olan sistemlerdir. Algoritmalar kullanarak ham verideki gizli kalıpları ve korelasyonları tanıyabilir, kümelere ayırabilir, sınıflandırabilir ve bu işlemleri sürekli öęrenip geliřtirebilirler. Yapay sinir aęları bilgisayarla görme, konuřma tanıma, makine çevirisi, sosyal aę filtreleme, video oyunları ve tıbbi tanı gibi çeřitli alanlarda geliřme göstermiřtir. Yapay sinir aęları birbirine baęlı bir yapay nöron grubundan oluşur ve bu, hesaplama için baęlantıya dayalı bir yaklařım kullanarak bilgiyi işler. Merkezi baęlantıya dayalı temel prensip, zihinsel olayların birbirine baęlı basit ve çoęu zaman tek biçimli birim aęlarıyla tanımlanabilmesidir. Çoęu durumda bir yapay sinir aęları, öęrenme ařamasında aę üzerinden akan dıř veya iç bilgilere dayanarak yapısını deęiřtiren uyarlamalı bir sistemdir. Modern sinir aęları genellikle girdiler ve çıktıları arasındaki karmařık iliřkileri modellemek ya da verideki modelleri bulmak için kullanılır. İleri beslemeli bir sinir aęı, birimler arasındaki baęlantıların yönlendirilmiř bir döngü oluřturmadıęı yapay bir sinir aęıdır. Bu aęda, bilgi giriř düęümlerinden gizli düęümlere (eęer varsa) çıkıř düęümlerine kadar sadece bir yönde ileri doęru hareket eder. Aęda hiç döngü yoktur. Geri yayılım algoritması iki ařamaya ayrılabilen denetimli bir öęrenme yöntemidir: yayılma ve aęrılık güncelleme. İki faz, aęın performansı yeterince iyi olana kadar tekrar edilir. Geri yayılım algoritmalarında, çıkıř deęerleri önceden tanımlanmıř bazı hata fonksiyonlarının deęerini hesaplamak için doęru cevap ile karřılařtırılır. Çeřitli tekniklerle, hata aędan geri beslenir. Bu bilgiyi kullanarak, algoritma, hata fonksiyonunun deęerini küçük bir miktar azaltmak için her baęlantının aęrılıęını ayarlar. Bu işlemleri yeterince fazla sayıda eğitim döngüsü için tekrarladıktan sonra, aę genellikle hesaplama hatasının küçük olduęu bir duruma dönuőecektir. Bu durumda, aęın belirli bir hedef işlevi öęrenecektir (řeker, 2008; SAS, 2019, Rapid Miner, 2019).

Evren ve Örnekleme

Bu arařtırmanın evrenini Akdeniz Üniversitesi Sınıf Öęretmenlięi Bölümü'nden 2012-2017 yılları arasından mezun olan toplam 696 öęrenci oluřturulmuřtur. Arařtırmada örnekleme alınma yoluna gidilmeyerek tüm öęrencilerin verisi kullanılmaya çalıřılmıřtır. Veri incelendięinde arařtırmada kullanılacak olan özelliklere iliřkin kayıp verinin olduęu kayıtlar arařtırmadan çıkarılmıřtır. Elde edilen veri setinde 578 öęrencinin verisi bulunmaktadır ve evrenin %83.05'ini temsil etmektedir.

Verinin Toplanması ve Analizi:

Arařtırma kapsamında kullanılan veri Akdeniz Üniversitesi öęrenci bilgi sistemi veri tabanından elde edilmiřtir. Arařtırma kapsamında kullanılan veri için veri kullanımında uyulacak etik ilkeler belirlenerek Akdeniz Üniversitesi Bilimsel Arařtırma ve Yayın Etięi Kurulundan uygulama için izin alınmıřtır. Gereki izin alındıktan sonra veriye eriřim için Bilgisayar Bilimleri Arařtırma ve Uygulama Merkezi'ne bařvuru yapılarak Sınıf Öęretmenlięi Bölümü öęrencilerine ait öęrenci kimlik bilgilerinin filtrelendięi, anonimleřtirilmiř veri alınmıřtır. Veri analizi için bir veri bilimi platformu Rapid Miner Studio (v. 9.2.001 Education Edition) kullanılmıřtır. Bu yazılım veri

analizi, veri ön işleme, veri madenciliği teknikleri ve algoritmalarının kullanıldığı, modellerin oluşturulduğu ve değerlendirildiği tümleşik veri bilimi platformudur. Rapid Miner platformundaki araçlar kullanılarak veri ön işleme işlemleri yapılan ve veri madenciliği teknikleri için hazır hale getirilen veri setindeki niteliklere ilişkin veri özellikler Tablo 1’de gösterilmiştir.

Tablo 1

Veri Setine Ait Özellikler

Nitelik Adı	Türü	Değer aralığı
Öğrenci Kod	Sayısal	1-1000
Cinsiyet	İkili grup (binominal)	Kadın (0) Erkek (1)
Medeni Durum	İkili grup (binominal)	Bekar (0) Evli (1)
Kayıt Yaşı	Sayısal	16-33
Eğitim Bilimine Giriş ara sınav puanı	Sayısal	0-100
Türkçe I Yazılı Anlatım ara sınav puanı	Sayısal	0-100
Atatürk İlkeleri ve İnkılap Tarihi I ara sınav puanı	Sayısal	0-100
İngilizce I ara sınav puanı	Sayısal	0-100
Bilgisayar I ara sınav puanı	Sayısal	0-100
Genel Biyoloji ara sınav puanı	Sayısal	0-100
Uygurluk Tarihi ara sınav puanı	Sayısal	0-100
Mezuniyet Notu	Sayısal	0-4

Modelin Değerlendirilmesi:

Modelin eğitimi ve testi için Rapid Miner’da yer alan Cross Validation (Çapraz Doğrulama) işleci kullanılmıştır. Bu işleç iç içe geçmiş bir işleçtir. İki alt işlemi vardır: Eğitim alt işlemi ve test alt işlemi. Eğitim alt işlemi bir modelin eğitimi için kullanılır. Eğitilen model daha sonra test alt işleminde uygulanır. Modelin performansı test aşamasında ölçülür. Veri seti girişi, eşit boyuttaki k altkümelerine ayrılır. K altkümelerinin test veri seti olarak tek bir altküme tutulur (yani test alt işleminin girişi). Kalan k-1 altkümeleri, eğitim veri seti olarak kullanılır (yani, Eğitim alt işleminin girişi). Çapraz doğrulama işlemi daha sonra k kez tekrarlanır, k alt setlerinin her biri test verisi olarak tam bir kez kullanılır. Tek bir kestirim üretmek için k yinelemelerden elde edilen k sonuçlarının ortalaması alınır veya başka şekilde birleştirilir (Rapid Miner, 2019).

Modellerin eğitiminde ve testinde k=10 olarak alınmıştır. Veri seti 10 eşit parçaya bölünerek her seferinde 9 parçası eğitim 1 parçası test için kullanılarak veri setinin tümü kullanılmıştır. Çapraz Doğrulama yönteminde veri setinin kullanımında karışık örnekleme “Shuffle Sampling” yöntemi kullanılmıştır. Modellerin değerlendirilmesinde kullanılan performans göstergeleri ve hesaplamalarına ilişkin bilgiler aşağıda yer almaktadır.

RMSE (Root Mean Squared Error)	: Ortalama Hata Kareleri Karekökü
MSE (Mean Squared Error)	: Ortalama Hata Kareleri
R ² (SC) Squared Correlation	: Determinasyon katsayısı
MAPE (Mean Absolute Percentage Error)	: Ortalama Mutlak Hata Yüzdesi

BULGULAR

Sınıf Öğretmenliği Bölümü Mezunlarına Ait Betimleyici İstatistikler

Bölümden mezun olan öğrencilerin verisi incelendiğinde 578 mezun öğrenciden; 366’sı (%63.32) kadın, 212’i (%36.68) erkek, 88’u evli (%15.22), 490’si (%84.78) bekarıdır. Öğrencilerin yaşlarına bakıldığında en küçük yaş 16.13, ortalama 18.95 ve en büyük yaş ise 32.47’dir. Mezun öğrencilerin 1. sınıf 1. dönem aldığı derslerin ara sınav puanları ve mezuniyet notlarına ilişkin betimleyici istatistikler Tablo 2’de yer almaktadır.

Tablo 2

Deęiřkenlere Ait Özellikler

Ders Adı/Ara Sınav Puanı	Min	Max	\bar{x}	SS
Eđitim Bilimine Giriř	20	100	72,73	18.96
Türkçe I Yazılı Anlatım	40	100	77.74	13.28
Atatürk İlkeleri ve İnkılap Tarihi I	23	100	71.12	16.88
İngilizce I	20	100	84.70	12.54
Bilgisayar I	5	100	76.82	16.04
Genel Biyoloji	12	97	50,83	17.12
Uygarlık Tarihi	30	100	78.86	16.07

Min: en küçük deęer, Max: En büyük deęer, \bar{x} : Ortalama, SS: Standart sapma Mezuniyet Notu: 4'lük sistem,

Modelleme

Sınıf Öğretmenlięi bölümü mezunu öğrencilerin mezuniyet puanları kestirmek için 2 adet model test edilmiştir. Öğrencilere ait cinsiyet, medeni durum gibi kategorik veri ikili grup (binominal deęer 0-1) ifade edilerek kullanılmıştır. 1. model çoklu doğrusal regresyon analizidir. Çoklu doğrusal regresyon analizinde amaç girdileri kullanarak çıktılar tahmin edecek bir model oluşturmaktır. Rapid Miner'da çoklu doğrusal regresyon işlecinde kullanılan parametreler ve açıklamaları aşağıda verilmiştir.

Nitelik seçiminde M5-prime algoritması kullanılmıştır. Eş bütünleşme niteliklerini çıkar "eliminate colinear features" özellięi seçilerek eş bütünleşme özellięi gösteren dersler analiz dışı bırakılmıştır. Minimum tolerans düzeyi "mintolerance" anlamlılık düzeyi 0.05 olarak seçilmiştir. Eğitim Bilimine Giriş dersi eş bütünleşme özellięi gösterdięi için çoklu doğrusal regresyon analizine girmemiştir. Oluşturulan çoklu doğrusal regresyon modeline ilişkin sonuçlar Tablo 3'te gösterilmiştir.

Tablo 3

Çoklu Doğrusal Regresyon Modeli ile Mezuniyet Notu Tahmini

Deęiřkenler	B	SH	SB	T	t	p
Atatürk İlkeleri ve İnkılap Tarihi I	.002	.001	.132	.973	3.600	.000*
Bilgisayar I	.002	.001	.110	.989	3.304	.001*
Genel Biyoloji	.004	.001	.209	.799	5.433	.000*
Türkçe I yazılı Anlatım	.001	.001	.037	.926	1.058	.290
Uygarlık Tarihi	.003	.001	.167	.956	4.980	.000*
İngilizce I	.004	.001	.155	.854	4.394	.000*
Yaş	-.004	.008	-.018	.946	-.528	.598
Medeni Durum	-.153	.028	-.191	.950	-5.488	.000*
Cinsiyet	-.220	.021	-.367	.922	-10.460	.000*
Sabit	2.266	.187	-	-	12.150	.000*

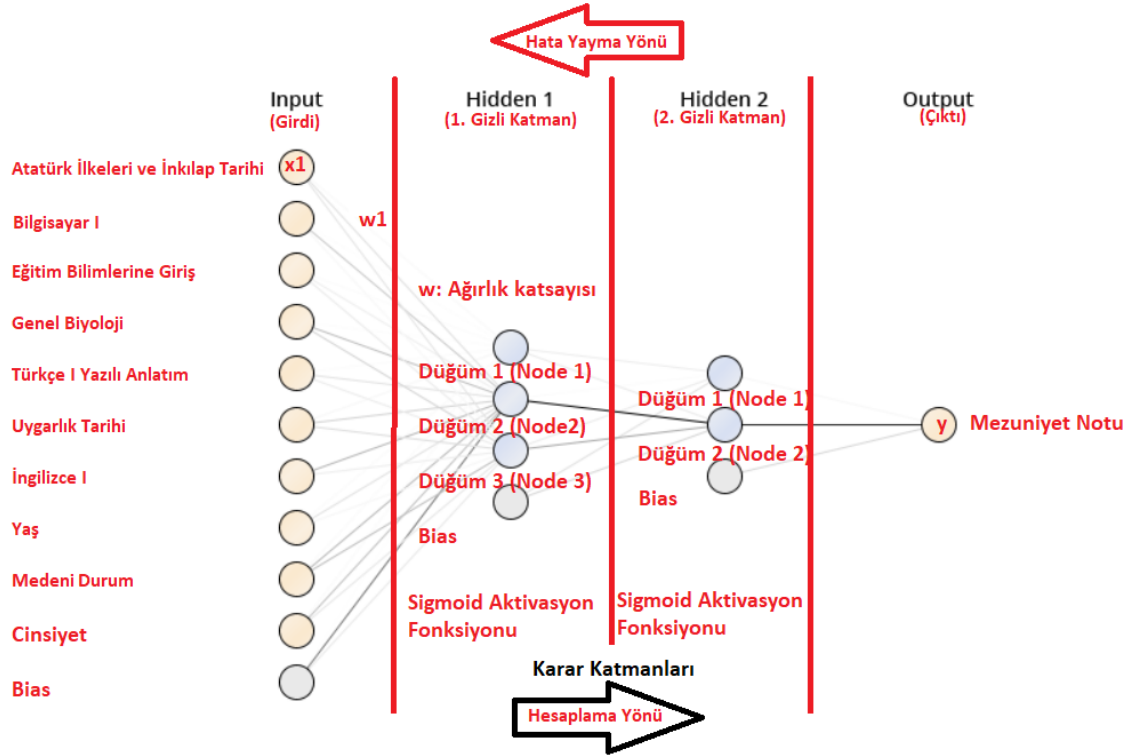
$R^2=.410$, * $p<.01$

Tablo 3 incelendięinde Genel Biyoloji ve İngilizce I derslerinin ara sınav puanları mezuniyet notunu hesaplamada en yüksek katsayıya sahiptir. İkinci sırada Uygarlık Tarihi dersi ara sınav puanı gelmektedir. Mezuniyet notunu hesaplamada olumsuz yönden etki eden deęişkenler ise evli olma durumu ve cinsiyet olarak erkek olma durumları görülmektedir. Analiz incelendięinde mezuniyet notuna "evli" olmanın ve "erkek" olmanın olumsuz etkisi olduęu görülmektedir. Çoklu doğrusal regresyon formülü aşağıdaki gibi ifade edilebilir.

Mezuniyet Notu= 2.266+ (.004*Genel Biyoloji) + (.004*İngilizce I) + (.003*Uygarlık Tarihi) + (.002*Atatürk İlkeleri ve İnkılap Tarihi I) +(.002*Bilgisayar I) + (.001*Türkçe I Yazılı Anlatım)- (0,151*Evli Olma) – (0,222*Cinsiyet)

Determinasyon katsayısına bakılarak ($R^2=.410$) 5 dersin ara sınav puanları, medeni durum ve cinsiyet değişkenleriyle oluşturulan çoklu doğrusal regresyon modelinin mezuniyet notundaki varyansın %41'ini açıkladığı söylenebilir.

2. Model yapay sinir ağları ile mezuniyet notu kestirim modelidir. Bu model Şekil 1'de görüldüğü üzere giriş katmanı, 1. gizli katman, 2. gizli katman ve çıkış katmanından oluşmaktadır.



Şekil 1: Yapay Sinir Ağı Modeli

Yapay sinir ağları modelinde (Şekil 1) girdi katmanında Sınıf Öğretmenliği Bölümü 1. sınıf 1. dönem derslerine ait ara sınav puanları, yaş, medeni durum ve cinsiyet değişkenleri yer almaktadır. Yapay sinir ağları modeli incelendiğinde karar katmanları 2 adet gizli katmandan oluşmaktadır: 1.gizli katman ve 2. gizli katman. Bu katmanlarda aktivasyon fonksiyonu olarak Sigmoid fonksiyonu kullanılmıştır. Çıktı katmanında ise tahmin edilecek değer olan mezuniyet notu yer almaktadır. Hesaplama yönü girdi katmanından çıktı katmanına doğru işlemektedir. Girdilerin çıktılar üzerinde etkisi ağırlıklar (w) ile belirlenmektedir. Ağırlıklar girdilerle çarpılarak ileri doğru iletilir. Aktivasyon fonksiyonu her nöron için net girdiyi işleyerek aktivasyon çıktısını üretir. Modelin eğitimi sırasında tahmin edilen değer ile gerçek değer arasındaki fark azaltılmaya çalışılır. Bunun için aktivasyon fonksiyonu üzerinden geri yayılım ile ağırlıklar öğrenme oranı ve momentum değeri eklenerek güncellenir. Tahmin edilen değerler ile gerçek değerler arasındaki fark minimum olana kadar veya iterasyon sayısı kadar model öğrenmeye devam edecektir. Şekil 1'de gösterilen yapay sinir ağları için değişik sayıda gizli "layer" katman ve katmanlar içinde "node" düğümler test edilmiştir. Bu modellerden hata oranı en düşük modele ait parametreler ve açıklamaları aşağıda sunulmuştur (Rapid Miner, 2019).

Girdi (Input): Sinir ağının girdilerini göstermektedir. Modele girecek olacak değişkenleri göstermektedir.

Gizli Katman (Hidden Layer): Bu parametre tüm gizli katmanların adını ve boyutunu açıklar. Sinir ağının yapısını bu parametre ile tanımlayabilir. Her liste girişinde yeni bir gizli katman açıklanır. Her giriş gizli katmanın adını ve boyutunu gerektirir. Katman adı isteğe bağlı olarak seçilebilir. Sadece modeli görüntülemek için kullanılır. Modelde 2 adet gizli katman

kullanılmıřtır (Hidden1, Hidden2). 1. Gizli katmanda 3+1 bias düğüm (node, nöron) 2. Gizli katmanda 2+1 bias düğüm yer almaktadır.

Öğrenme Döngüsü (Training Cycles): Bu parametre, sinir ağı eğitimi için kullanılan egzersiz çevrimlerinin sayısını belirtir. Geri yayılımda, çıkıř deęerleri önceden tanımlanmıř bazı hata fonksiyonlarının deęerini hesaplamak için doęru cevap ile karřılařtırılır. Hata daha sonra ağı üzerinden geri beslenir. Bu bilgiyi kullanarak, algoritma, hata fonksiyonunun deęerini küçük bir miktar azaltmak için her baęlantının ağırlıęını ayarlar. Bu parametre iřlemin kaę defa tekrarlanacaęını belirtmek için kullanılır. Varsayılan 1000 deęeri korunmuřtur.

Learning Rate (Öğrenme Oranı): parametresi her adımda ağırlıkları ne kadar deęiřtirdiđimizi belirler ve 0 deęerini alamaz. 0.02 deęerinde en düşük hata oranı elde edilmiřtir. Momentum (Moment): Momentum, önceki ağırlık güncellemesinin bir kısmını olana ekler. Bu, yerel maksimumlamayı önler ve optimizasyon yönlerini düzeltir. 0.02 deęerinde en düşük hata elde edilmiřtir. Yapay sinir aęları modelinin çalıřma yapısı řu řekilde açıklanabilir: Gizli katman 1' in 1. düğümü için: Gizli katman 1 için 1. adım girdilerin toplamı: $X_1 * W_{1,1} + X_2 * W_{2,1} + \dots + X_{10} * W_{10,1} + b$. Girdiler ve sabit deęer toplandıktan sonra 2. adımda toplam sonucu elde edilen deęer sigmoid fonksiyonu ile çıktı olarak gizli katman 2 düğümüne baęlanır. Aynı hesaplama iřlemleri tüm katmanlardaki tüm düğümler için hesaplanır. Yapay sinir aęları tahminlerdeki hatayı azaltmak için hatanın kaynaęı olan (geri yayılım algoritması) düğümün ağırlıęını (W) güncelleyerek ideal ağırlıkları elde eder. Tablo 4'te 1. gizli katmana ait ağırlık katsayıları gösterilmiřtir.

Tablo 4

Yapay Sinir Aęları Modeli 1. Gizli Katman Ağırlık ve Düğümler

Girdi /Ağırlık (W)	Düğüm 1	Düğüm 2	Düğüm 3
Atatürk İlkeleri ve İnkılap Tarihi I	-.055	-.332	.288
Bilgisayar I	.018	-.599	.080
Eđitim Bilimlerine Giriř	-.034	-.181	-.228
Genel Biyoloji	-.025	-.721	.472
Türkçe I Yazılı Anlatım	.030	-.317	-.214
Uygarlık Tarihi	-.083	-.486	.375
İngilizce I	.005	-.823	.188
Yař	.039	.327	-.093
Medeni Durum	-.047	.779	.717
Cinsiyet	.048	.700	.273
Bias	-.007	1.511	-.251

1.gizli katmanın çıktıları 2. gizli katmanın girdileri olarak iřlenmektedir. 1. gizli katmandaki 3 düğümünün çıktıları 2. gizli katmanda 2 düğümün girdileridir. 2. gizli katmana ait ağırlık katsayıları (W)Tablo 5'te gösterilmiřtir.

Tablo 5

YSA Modeli 2. Gizli Katman Ağırlık ve Düğümler

Girdi/Ağırlık (W)	Node1	Node2
Gizli katman 2 düğüm 1	-.196	-.220
Gizli katman 2 düğüm 2	-.035	-2.859
Gizli katman 2 düğüm 3	-.227	.970
Bias	-.335	-.462

Yapay sinir aęları ile oluřturulan modelin çıktıları 2. gizli katmanın 2 düğümünün çıktılarından elde edilen regresyon fonksiyonu sonucunda oluřan katsayılar: Gizli katman 2

düğüm 1: -.084, gizli katman 2 düğüm 2: 2.365 ve Threshold: -.452. Sınıf Öğretmenliği öğrencilerinin mezuniyet notunu tahmin etmek için çoklu doğrusal regresyon ve yapay sinir ağları ile oluşturulan modellerin performans göstergeleri karşılaştırılmalı olarak Tablo 6'da sunulmuştur.

Tablo 6

Modellerin Karşılaştırılması

	N	\bar{x}	\bar{x}_1	RMSE	MSE	R ²	MAPE (%)
Regresyon Modeli	578	3.175	3.176	.223	.050	.410	5.70
YSA Modeli			3.175	.219	.048	.429	5.57

\bar{x} : Gerçek Mezuniyet Notlarının Ortalaması

\bar{x}_1 : Tahmini Mezuniyet Notlarının Ortalaması

Tablo 6 incelendiğinde: Regresyon Modeli Ortalama Hata Kareleri Karekökü (RMSE)=.223, Ortalama Hata Kareleri (MSE)=.050, determinasyon katsayısı (R²)= .410 ve Ortalama Mutlak Hata Yüzdesi (MAPE)=5.70'dir. Yapay sinir ağları modeli incelendiğinde Hata Kareleri Karekökü (RMSE)=.219, Ortalama Hata Kareleri MSE= .048, determinasyon katsayısı (R²)= .429 ve Ortalama Mutlak Hata Yüzdesi (MAPE)=5.578'tir. Sınıf Öğretmenliği öğrencilerinin mezuniyet notunu tahmin etmede yapay sinir ağlarının çoklu doğrusal regresyon modeline göre daha başarılı olduğu söylenebilir. Yapay sinir ağları modelinde determinasyon katsayısı incelendiğinde modelin toplam varyansın %42.9'unu açıkladığı görülmektedir. Başka bir gösterge olan MAPE değerine bakıldığında ise yapay sinir ağları modelinin %94.43 doğrulukla (1-MAPE) öğrencilerin mezuniyet notunu tahmin etmeyi sağladığı görülmektedir. MAPE özellikle kestirim (forecast) modellerinde sıkça tercih edilen bir hata göstergesidir. MAPE hesaplanırken hatların mutlak değerinin alınması (pozitif negatif değerlerin hata değerini azaltmaması) güvenilirliğini artırmaktadır (MAPE, 2006).

SONUÇ TARTIŞMA ve ÖNERİLER

Araştırmada Akdeniz Üniversitesi Eğitim Fakültesi Sınıf Öğretmenliği Bölümü'nden mezun olan öğrencilerin 1. sınıf 1.dönem ara sınav puanları, cinsiyet, yaş ve medeni durum gibi değişkenler kullanılarak mezuniyet notlarını tahmin etmeye yönelik 2 model geliştirilmiştir. Geliştirilen modellerden çoklu doğrusal regresyon modelinin %94.30, yapay sinir ağları modelinin ise %94.43 doğrulukla (1-MAPE) öğrencilerin mezuniyet notunu tahmin kestirebildiği görülmektedir. Çoklu regresyon modeli sonuçları değerlendirildiğinde mezuniyet notuna etki eden demografik değişkenlerin cinsiyet ve medeni durum olduğu görülmektedir. Mezuniyet notuna etki eden ara sınav puanları ağırlıklarına göre sıralandığında Genel Biyoloji, Atatürk İlkeleri ve İnkılap Tarihi, İngilizce I, Bilgisayar I ve Uygurlık Tarihi dersleridir. Araştırmada oluşturulan modeller değerlendirildiğinde öğrencilerin mezuniyet notlarını erken akademik dönemlerde belirlenmesinde etkili olduğu söylenebilir. Yapılan benzer çalışmalar veri madenciliği teknikleri altında regresyon, yapay sinir ağları, karar ağaçları, sınıflandırma gibi tekniklerle öğrenci başarısının tahmin edilmesinde başarılı sonuçlar elde edildiği görülmektedir. Araştırmalar incelendiğinde çoğunlukla öğrenci profillerinin belirlenerek öğrencilerin başarı durumlarına göre sınıflandırılmalarına yönelik modeller oluşturulmuştur. Öğrencilerin akademik performansını yükseltmek, başarısızlık nedenlerini belirlemek ve başarısızlığa düşmeden gerekli önlemleri almak ve için bu tür çalışmalardan yararlandığı görülmektedir (Ayesha, Mustafa, Sattar ve Khan, 2010; Baradwaj ve Pal, 2011; Tsai C-F., Tsai C-T., Hung ve Hwang, 2011; Cunningham, 2017; Benton, 2018; Alsuwaiket, 2018; Whitlock, 2018; Akçapınar, 2014; Aydın, 2015; Kılınç, 2015). Türkiye'de e-devlet uygulamaları ve bu kapsamda Millî Eğitim Bakanlığı (MEB) tarafından kullanılan MEBBİS (MEB Bilişim Sistemleri), e-okul (öğrenci veli bilgi sistemleri), Eğitim Bilişim Ağı (EBA) üniversite otomasyon sistemleri, uzaktan eğitim sistemleri gibi uygulamalarda öğretmen ve öğrencilere ait çok büyük ölçeklerde ve çok çeşitli veri tutulmaktadır, ancak bu veriden "bilgi"

elde etmede yeterince yararlanılmadıđı söylenebilir (MEB, 2019). Özellikle eğitim alanında öğrencilerle, öğretmenlerle, öğretim ortamlarıyla, ölçme ve değerlendirme sonuçlarıyla ilgili tutulan devasa ölçekte veri yığınları düşünöldüğünde bu veri içerisinde saklı bilgilerin, örüntülerin keşfedilmesinin ve bunların eğitimin verimliliđi ve kalitesinin artırılması için kullanılmasının önemi daha iyi anlaşılacaktır. Bu çalışmanın toplam kalite yönetimi anlayışını uygulayan Akdeniz Üniversitesi'nin sürekli gelişme (Aydın, 2007; Öztürk, 2009), sorunları önceden kestirerek proaktif bir anlayışla önlemeye çalışma ve hedeflerine (Tortop, Aykaç, Yayman ve Özer, 2007; Bingöl, 2010) daha etkili ve verimli bir şekilde ulaşmasına katkı sağlaması beklenmektedir.

KAYNAKLAR

- Akçapınar, G. (2014). Çevrimiçi Öğrenme Ortamındaki Eğitim Verilerine Göre Öğrencilerin Akademik Performanslarını Veri Madenciliđi Yaklaşımı ile Modellenmesi. Yayınlanmamış doktora tezi Ankara: Hacettepe Üniversitesi Bilgisayar ve Öğretim Teknolojileri ABD.
- Akgöbek, Ö. ve Kaya, S. (2011). Veri madenciliđi teknikleri ile veri kümelerinden bilgi keşfi:medikal veri madenciliđi uygulaması. *e-Journal of New World Sciences Academy*, 6 (1), 237-245.
- Alsuwaiket, M. (2018). Measuring academic performance of students in higher education using data mining techniques (Order No. 13872716). Available from Pro Quest Dissertations & Theses Global. (2204780401).
- Argüden, Y. ve Erşahin, B. (2008). Veri madenciliđi veriden bilgiye masraftan değere. ARGE danışmanlık.
- Aydın, M. (2007). *Çağdaş Eğitim Denetimi*, 5. Baskı. Ankara: Hatibođlu Yayınları..
- Aydın, M. (2014). *Eğitim Yönetimi. Kavramlar, Kuramlar, Süreçler, İlişkiler*, 10. Baskı. Ankara: Gazi Kitabevi.
- Aydın, S. (2015). Veri madenciliđi ve Anadolu Üniversitesi açık öğretim sisteminde bir uygulama. Eğitim ve Öğretim Arařtırmaları Dergisi *Journal of Research in Education and Teaching*, 4 (3), 36-44.
- Ayesha, S., Mustafa, T., Sattar, A. R., ve Khan, M. I. (2010). Data mining model for higher education system. *Europen Journal of Scientific Research*, 43 (1), 24-29.
- Baradwaj, B. K. ve Pal, S. (2011). Mining educational data to analyze students' performance. *International Journal of Advanced Computer Science and Applications*, 2 (6), 63-69.
- Benton, J. (2018). Predicting first-year academic success of african american and White students at predominately White institutions (Order No. 10791386). Available from Pro Quest Dissertations & Theses Global.
- Bienkowski, M., Feng, M., ve Means, B. (2012). *Enhancing Teaching And Learning Through Educational Data Mining And Learning Analytics: An Issue Brief*. Washington
- Bingöl, D. (2010). *İnsan Kaynakları Yönetimi*. 7. Baskı. İstanbul: Beta Yayınları.
- Bursaliođlu, Z. (1991). *Okul Yönetiminde Yeni Yapı ve Davranış*.8. Baskı. Ankara: Pegem Yayınları.
- Cunningham, J. A. (2017). Predicting student success in a self-paced mathematics MOOC (Order No. 10272808). Available from Pro Quest Dissertations & Theses Global. (1900990574).
- Dunham, M. H. (2003). Data mining introductory and advanced topics. New Jersey: Prentice Hall
- Fayyad, U., Piatetsky-Shapiro, G., ve Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37-37.
- García, E., Romero, C., Ventura, S., ve De Castro, C. (2011). A collaborative educational association rule mining tool. *The Internet and Higher Education*, 14(2), 77-88.
- GG. (2019). Gartner Group Web Sayfası. 07 20, 2019 tarihinde <https://www.gartner.com/it-glossary/data-mining> adresinden alındı.
- IDC. (2016). IDC Global Research. 12 25, 2016 tarihinde <http://www.idc.com> adresinden alındı.

- IEDMS. (2019). International Education Data Mining Society. 07 20, 2019 tarihinde <http://www.educationaldatamining.org/> adresinden alındı.
- Kılınç, Ç. (2015). Üniversite öğrenci başarısı üzerine etki eden faktörlerin veri madenciliği yöntemleri ile incelenmesi (Yayımlanmamış yüksek lisans tezi Tez No. 415460). YÖK Başkanlığı Tez Merkezi https://tez.yok.gov.tr/UlusalTezMerkezi/TezGoster?key=WY5CM7tPNE2z_YM6pBu0t51AOffDtEC7P8BHiMbdbdZh0odTb6HLCYy8VAo6r2oS adresinden alındı.
- Kingir, S. (2013) *Toplam Kalite Yönetimi* 3. Baskı. Ankara: Nobel yayınları.
- Lopez, M. I., Luna, J. M., Romero, C., ve Ventura, S. (2012). Classification via clustering for predicting final marks based on student participation in forums. 5th *International Conference on Educational Data Mining*.
- MEB. (2019). 2023 Vizyonu. 12/09/2019 tarihinde <http://2023vizyonu.meb.gov.tr/> adresinden alındı.
- Mean Absolute Percentage Error (MAPE). (2006). in Encyclopedia of Production and Manufacturing Management (pp. 462–462). Springer US. https://doi.org/10.1007/1-4020-0612-8_580
- Özbay, Ö. (2015). Veri madenciliği kavramı ve eğitimde veri madenciliği uygulamaları. *Uluslararası Eğitim Bilimleri Dergisi* (5), 262-272.
- Öztürk, A. (2009). *Toplam Kalite Yönetimi ve Planlaması*. Bursa: Ekin Basım Yayın Dağıtım.
- Rapid Miner, (2019). Rapid Miner Documentation. 07 20, 2019 tarihinde <https://docs.rapidminer.com/> adresinden alındı.
- SAS. (2019). SAS. 07 20, 2019 tarihinde <http://documentation.sas.com/> adresinden sorgulanarak alındı.
- Şeker, Ş. E. (2008). Bilgisayar kavramları. 06 23, 2019 tarihinde <http://bilgisayarkavramlari.sadievrenseker.com/2008/10/02/yapay-sinir-aglari-artificial-neural-networks/> adresinden alındı
- Şeker, Ş. E. (2014). MIS Sözlük. 06 23, 2019 tarihinde <http://mis.sadievrenseker.com/> adresinden alındı.
- Toprakçı, E. (2017) Sınıf Yönetimi. Ankara: PegemA Yayınları
- Tortop, N., Aykaç, B., Yayman, H. Ve Özer, M. A. (2007). *İnsan Kaynakları Yönetimi*. 2. Baskı, Ankara: Nobel Yayınları.
- Tsai, C. F., Tsai, C. T., Hung, C. S. ve Hwang, P. S. (2011). Data mining techniques for identifying students at risk of failing a computer proficiency test required for graduation. *Australasian Journal of Educational Technology*, 27 (3), 481-498.
- Uras, M. (2016). *Eğitimin Sosyolojisi*. E. Toprakçı (Eds.) Eğitim, Pedagoji. S.217-270). Ankara: Ütopya yayınları.
- Walter, A. A. (2000). Data Mining Industry, Emerging Trends and New Opportunities, Yayımlanmamış YL Tezi. Massachusetts Institute of Technology.
- Whitlock, J. L. (2018). Using data science and predictive analytics to understand 4-year university student churn (Order No. 10904502)
- Witten, I. H., Frank, E. (2005). *Data Mining: Practical Machine Learning Tools And Techniques*. 2.Baskı, Morgan Kaufmann yayınları, Burlington.
- YÖK. (2019). Yükseköğretim Kurulu Başkanlığı Tez Merkezi. <https://tez.yok.gov.tr> adresinden sorgulanmıştır.

Estimation of Graduation Grades of Primary Education Students by Using Regression Analysis and Artificial Neural Networks²

Murat Altun
Akdeniz University-Turkey
emurataaltun@gmail.com

Assoc. Prof. Dr. Kemal Kayıkçı
Akdeniz University-Turkey
kemalkayikci@akdeniz.edu.tr

Assoc. Prof. Dr. Sezgin Irmak
Akdeniz University-Turkey
sezginakdeniz@akdeniz.edu.tr

Abstract

In this research, models were developed to estimate the graduation grades of 578 students who graduated from the Department of Primary Education, Faculty of Education, Akdeniz University between 2012-2017 by using the data related to gender, marital status, age of enrollment and midterm exam points of 1st semester of 1st grade. It has been seen that the models formed by regression and artificial neural networks on anonymized student data give close results in predicting student success as a result of cross-evaluation. In the evaluation based on Mean absolute percentage error values, the regression model achieved 94.30% success and the artificial neural networks model achieved 94.43% success. The models provided findings to determine the weights of the factors affecting student achievement. It can be said that the models developed with in the scope of the research can be used in studies such as zero error in the context of total quality management, taking proactive positions, increasing student achievement and guiding.

Keywords: Primary Education, artificial neural networks, regression, educational data mining



**E-International Journal
of Educational Research,
Vol: 10, No: 3, 2019, pp. 29-43**

DOI: 10.19160/ijer.624839

Received: 06.09.2019
Accepted: 28.10.2019

Suggested Citation:

Altun, M., Kayıkçı, K., ve Irmak S. (2019). Estimation of Graduation Grades of Primary Education Students by Using Regression Analysis and Artificial Neural Networks, *E-International Journal of Educational Research*, Vol: 10, No: 3, 2019, pp. 29-43, DOI: 10.19160/ijer.624839

² Supported by Akdeniz University Scientific Research Projects Unit (Project Number: SDK-2018-4077).

EXTENDED ABSTRACT

Problem: In the conception of competitive organization, low cost and high performance is targeted. In order to achieve this, total quality management approach is based on Kaizen principles which means "zero error", "scientific method of process management" and "statistical thinking and continuous improvement" (Kıngur, 2013; Öztürk, 2009). Together with process management based on continuous development and supervision, proactive management understanding based on prevention is essential for the effective functioning of educational organizations. In this study, it is important to predict student achievement in this respect.

Developing technology has enabled the use of some statistical methods, mathematical models and algorithms that are difficult to apply on a large amount of data and development of some new techniques. It is one of the main motivations to acquire information to develop competitive approaches through big data (Walter, 2000; Şeker, 2014; IDC, 2016). Many concepts such as information mining in databases, information extraction, knowledge and pattern analysis, data archeology. are used in the literature for Data Mining (DM), which combines all the techniques from data to information with this different point of view. Knowledge Discovery from Databases (KDD) is the most widely used (Akgöbek ve Kaya, 2011). In the DM process, called the discovery process, the target is information that is often hidden in large amounts of data and not easily visible (Fayyad, Piatetsky-Shapiro and Smyth, 1996). One of the usage area of data mining is education. In the scanning process of the web page of YÖK (higher education board) Thesis Committee in Turkey, it is seen that the studies carried out in the field of education related to data mining are conducted under the names educational data mining, educational data mining and data mining in education (YÖK, 2019). It is seen that the studies conducted in the field of educational data mining deal with such factors as determining and classifying the students who are at risk of academic failure, the factors affecting student success, student profiles (Ayesha, Mustafa, Sattar ve Khan, 2010; Baradwaj ve Pal, 2011; Tsai C-F., Tsai C-T., Hung ve Hwang, 2011; Cunningham, 2017; Benton, 2018; Alsuwaiket, 2018; Whitlock, 2018; Akçapınar, 2014; Aydın, 2015; Kılınç, 2015). The aim of this study is to develop a model that predicts graduation scores by using multi linear regression analysis and artificial neural network on the academic and demographic data of the students of the Department of Primary Education, Faculty of Education, Akdeniz University. With the models to be developed, it is considered important to predict the academic success of the students in the future and to carry out the necessary studies to increase the academic performance of the students.

Method: Multi linear regression analysis and artificial neural network were used in the research. Regression is used to map a data item to a real-valued forecasting variable. This includes learning the function. The basic assumption is that the target data conforms to a known function type (linear, logistic). In the regression technique, the function that best models the given data is tried to be found. The error analysis, which detects the difference between actual results and predictions, is used to decide which function is best. The error analysis, which detects the difference between actual results and predictions, is used to decide which function is best (Dunham, 2013). The artificial neural network consists of an interconnected group of artificial neurons that process information using a link-based approach to computation. The main principle based on central connection is that mental events can be defined by simple and often uniform unit networks connected to each other. In most cases, an artificial neural network is an adaptive system that changes its structure based on external or internal information flowing through the network during the learning phase (Şeker, 2008; SAS, 2019, RapidMiner, 2019). In this study, data of 578 students who completed 1st year 1st semester visa scores, graduation grade averages and demographic age, gender, marital status data of 696 students who graduated from Department of Primary Education, Faculty of Education, Akdeniz University between 2012-2017 as data set used. For the training and testing of the models, the k results obtained from iterations were averaged by

using the Cross-Validation processor in RapidMiner (RapidMiner, 2019). In order to measure the performance indicators of the models, root mean square error (RMSE), mean squared error (MSE), determination coefficient (R^2), mean absolute error percentage (MAPE) were used.

Findings: In the study, two models were developed to predict graduation grades by using the data related to gender, marital status, age of enrollment and midterm exam points of 1st semester of 1st grade of the students who graduated from the Department of Primary Education, Faculty of Education, Akdeniz University. Regression Model root mean square error was (RMSE) = 0.223, Mean Error Squares was (MSE) = 0.050, the determination coefficient was (R^2) = 0.410 and Mean Absolute Error Percentage was (MAPE) = 0.057. When the artificial neural network model was examined, the root mean square error was (RMSE) = 0.219, the Mean Error Squares was (MSE) = 0.048, the determination coefficient (R^2) = 0.429 and the Mean Absolute Error Percentage was (MAPE) = 0.057. When the determination coefficient of artificial neural network model was examined, it was seen that the model explains 42.9% of the total variance. When the MAPE, another indicator, was examined, it was seen that the ANN model enabled students to estimate the graduation grade with an accuracy of 94.43% (1-MAPE). When the results of multi linear regression were evaluated, it was seen that the demographic variables affecting the graduation grade were gender and marital status. The midterm exam points which affected the graduation grade were General Biology, Atatürk's Principles and History of Turkish Revolution, English I, Computer I, Turkish I and History of Civilization. When the models created in this research were evaluated, it could be said that it is effective in determining the graduation grades of the students in the early academic terms.

Discussion and Conclusion: It is seen that multi linear regression model with accuracy 94.30% and artificial neural network model with accuracy 94.43% can predict students' graduation grade (1-MAPE). When the results of multi linear regression were evaluated, it was seen that the demographic variables affecting the graduation grade were gender and marital status. It can be stated that students may benefit from such studies in order to improve their academic performance, to determine the reasons of their failure and to take the necessary precautions without failing (Ayesha, Mustafa, Sattar & Khan, 2010; Baradwaj & Pal, 2011; Tsai C-F., Tsai C-T., Hung & Hwang, 2011; Cunningham, 2017; Benton, 2018; Alsuwaiket, 2018; Whitlock, 2018; Akçapınar, 2014; Aydın, 2015; Kılınç, 2015). In applications such as electronic government applications and information systems used by the Ministry of Education in this context, e-school (students and parents in formation systems), Education Information Network, universities automation systems, distance education systems in Turkey, large scale and wide variety of data related to teachers and students are kept, but it can be said that these data are not used sufficiently to obtaining formation (MoNE, 2019). With this study, it was aimed to contribute to the continuous process development related to the education process of Akdeniz University which applies total quality management approach (Aydın, 2007, Öztürk, 2009), and to preventing its problems with a proactive approach by predicting them in advance more effectively and efficiently (Tortop, Aykaç, Yayman ve Özer, 2007; Bingöl, Turkey, 2010).