T.C.
İNÖNÜ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR ve ÖĞRETİM TEKNOLOJİLERİ EĞİTİMİ ANA BİLİM DALI
BİLGİSAYAR ve ÖĞRETİM TEKNOLOJİLERİ EĞİTİMİ BİLİM DALI

EĞİTSEL VERİ MADENCİLİĞİNİN ÖĞRENCİ BAŞARISININ KESTİRİMİNE YÖNELİK KULLANIMI

YÜKSEK LİSANS TEZİ

Harun EKİNCİ

Malatya-2022

T.C.
İNÖNÜ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR ve ÖĞRETİM TEKNOLOJİLERİ ANA BİLİM DALI
BİLGİSAYAR ve ÖĞRETİM TEKNOLOJİLERİ EĞİTİMİ BİLİM DALI

EĞİTSEL VERİ MADENCİLİĞİNİN ÖĞRENCİ BAŞARISININ KESTİRİMİNE YÖNELİK KULLANIMI

YÜKSEK LİSANS TEZİ

Harun EKİNCİ

Danışman: Prof. Dr. Olgun Adem KAYA

Malatya-2022

ONUR SÖZÜ

Prof. Dr. Olgun Adem KAYA danışmanlığında yüksek lisans tezi olarak hazırladığım **Eğitsel Veri Madenciliğinin Öğrenci Başarısının Kestirimine Yönelik Kullanımı** başlıklı bu çalışmanın bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın tarafımdan yazıldığını ve yararlandığım bütün yapıtların hem metin içinde hem de kaynakçada yöntemine uygun biçimde gösterilenlerden oluştuğunu belirtir, bunu onurumla doğrularım.

Harun EKİNCİ

ÖNSÖZ

Çalışma süresinde her türlü yol gösterici olan, olumlu tavrı ile beni motive edip cesaretlendiren, bilgi birikimiyle çalışmama katkılar sağlayan, beraber çalışmaktan gurur duyduğum değerli danışman hocam Prof. Dr. Olgun Adem KAYA' ya sonsuz teşekkür ederim. Ayrıca bu süreçte çalışmamda yol gösterici olan değerli hocam Dr. İlyas AKKUŞ' a ayrıca teşekkür ederim.

Mart, 2022

Harun EKİNCİ

ÖZET

EĞİTSEL VERİ MADENCİLİĞİNİN ÖĞRENCİ BAŞARISININ KESTİRİMİNE YÖNELİK KULLANIMI

EKİNCİ, Harun
Yüksek Lisans, İnönü Üniversitesi Eğitim Bilimleri Enstitüsü
Bilgisayar ve Öğretim Teknolojileri Eğitimi Ana Bilim Dalı
Bilgisayar ve Öğretim Teknolojileri Eğitimi Bilim Dalı

Tez Danışmanı: Prof. Dr. Olgun Adem KAYA
Mart-2022, XI+ 84 sayfa

Bu çalışmanın amacı, eğitimde büyük veri, öğrenme analitikleri ve veri madenciliği kavramlarının önemini belirtmek ve öğrenci akademik performansını arttırmaya yönelik EVM çalışması yapmaktır. Araştırmada üniversiteye yerleşme puanı, yerleşme puan türü, lise puanı, cinsiyet, il, yaş ve medeni durum değişkenlerinin üniversiteden 4 yılda mezun olabilme durumuna etkisinin tahminine yönelik bir lojistik regresyon modeli oluşturulmuştur. Ayrıca yukarıda belirtilen değişkenlere ek olarak, Bilgisayar I dersi geçme durumu ve notu değişkenlerinin Bilgisayar II dersi ile ilişkisinin olup olmadığı incelenip lojistik ve doğrusal regresyon modeli de oluşturulmuştur. Bu amaca yönelik olarak İnönü Üniversitesinin otomasyon sisteminde kayıtlı öğrenci bilgileri içerisinden 223.279 öğrenciye ait bir veri seti oluşturulmuştur. Bu veri seti EVM süreç tasarımlarından CRISP-DM iş süreç adımlarına uygun olarak işlenmiştir. Veri setinin düzenlenmesi ve modellerin oluşturulması RapidMiner Studio programı ile gerçekleştirilmiştir. Yapılan analiz sonucunda öğrenci mezuniyet durumu lojistik regresyon modelinin %76,80 ile yüksek düzeyde performans gösterdiği gözlenmiştir. Öğrencilerin üniversiteye kayıt sonrası kişisel ve akademik bilgileri ile mezuniyet süresinin kestirilebileceği ve bu değişkenler kullanılarak öğrencilere yönelik gerekli önlemlerin alınabileceği sonucuna ulaşılmıştır. Bilgisayar II dersi lojistik regresyon modelinin %79,34 ile yüksek düzeyde performans gösterdiği görülmüş, öğrencilerin kişisel ve akademik bilgileri ile bu dersten geçme durumlarının ikinci dönemin başında tahmin edile bileceği sonucuna ulaşılmıştır. Geliştirilen Bilgisayar II dersi geçme notu doğrusal regresyon modelinin düşük hatalı tahminlerde bulunabilmesi sayesinde, öğrencilerin kişisel bilgileri (Cinsiyet, Yaş, İl) ve akademik bilgileri (Üniversiteye

Yerleşme Puanı, Yerleşme Puan Türü, Lise Bitirme Puanı, Bilgisayar I Dersi Geçme Notu) ile bu dersten kaç puan ile geçeceği kestirilebilmiştir.

Yapılan araştırma sonucunda eğitim kurumlarında depolanan bilgilerin çok önemli olması nedeniyle, titizlikle depolanması ve anonim olarak erişime açılmasının ne kadar önemli olduğu vurgulanmıştır. Düzgün ve çok sayıda veri ile yapılan EVM tahmin modelleri ile öğrencilere erken dönemlerde gerekli uyarı ve desteklerin verilerek akademik başarıların arttırılabileceği sonucuna ulaşılmıştır.

Anahtar Kelimeler: Büyük Veri, Eğitsel Veri Madenciliği, Öğrenme Analitikleri, Regresyon Analizi

ABSTRACT

THE USE OF EDUCATIONAL DATA MINING TO PREDICT STUDENT SUCCESS

EKİNCİ, Harun
Master, İnönü University Institute of Educational Sciences
Computer and Instructional Technologies Education
Computer Education and Instructional Technologies

Dissertation Advisor: Prof. Dr. Olgun Adem KAYA
March-2022, XI+ 84 pages

The aim of this study is to emphasize the importance of big data, learning analytics and data mining concepts in education and to conduct educational data mining studies to increase the academic performance of students. In this study, a logistic regression model was created to estimate the effects of university placement score, placement score type, high school score, gender, province, age and marital status variables on being able to graduate from university in 4 years. In addition to the variables mentioned above, it was investigated whether the variables of passing the Introduction to Computers I course and grade were related to the success in the Introduction to Computers II course and a logistic and linear regression model were also created. For this purpose, a data set of 223,279 students was created from the information registered in the student automation system of İnönü University. This data set was processed in accordance with the CRISP-DM business process steps, which is one of the educational data mining process designs. The editing of the data set and the creation of the models were carried out with the RapidMiner Studio program. As a result of the analysis, it was observed that the student graduation status logistic regression model performed at a high level with 76.80. It has been concluded that the graduation period can be predicted by using the personal and academic information of the students after their registration to the university, and the necessary measures can be taken for the success of the students by using these variables. It was seen that the logistic regression model of the Introduction to Computers II course showed a high level of performance with 79.34%, and it was concluded that whether the students will pass this course or not can be predicted at the beginning of the second semester by using their personal and academic information. The fact that the developed Introduction to Computers II course passing grade linear regression model can make low erroneous

estimations made it possible to predict the grade that students will get from this course by using their personal (Gender, Age, Province) and academic (University Placement Score, Placement Score Type, High School Graduation Score, Introduction to Computers I Course Passing Grade) information.

As a result of the research, it was emphasized that since the information stored in educational institutions is valuable, it is very important that it is meticulously stored and made accessible anonymously. It has been concluded that academic achievement can be increased by giving the necessary warnings and supports to the students in the early stages with the educational data mining prediction models made with smooth and large numbers of data.

Keywords: Big Data, Educational Data Mining, Learning Analytics, Regression Analysis

İÇİNDEKİLER

Sayfa No

ŞEKİLLER LİSTESİ

Sayfa No

TABLOLAR LİSTESİ

Sayfa No

KISALTMALAR LİSTESİ

RMSE : Kök-Ortalama-Kare Hatası
RE : Mutlak Hata
ÖSYM : Öğrenci Seçme ve Yerleştirme Merkezi
OBS : Otomasyon Bilgi Sistemi
EVM : Eğitsel Veri Madenciliği
MEB : Millî Eğitim Bakanlığı
BIGCHEM : Kimyada Büyük Veri (Big Data in Chemistry)
HADOOP : Dağıtık Dosya Sistemi
SQL : Yapılandırılmış Sorgu Dili (Structured Query Language)
H-BASE : H Veri Tabanı (Apache Hive-Base)
K-NN : K-En Yakın Komşu Algoritması
COVİD-19 : Korona Virüs-19
KDD : Knowledge Discovery in Databases
SEMMA : **S**ample - Explore – Modify - Model - Assess
CRISP-DM : Endüstriden Bağımsız Standart Veri Madenciliği Süreci

BÖLÜM I

1. GİRİŞ

Bu bölümde, problem durumuna, problem cümlesi ve alt problemlere, çalışmanın amacına ve önemine, sayıltılara, sınırlılıklara ve tanımlara yer verilmiştir.

1.1. Problem Durumu

Bilgisayar ve internet teknolojileri kullanım oranlarının artmasına bağlı çeşitli alanlardaki teknolojik araçların gelişimi de hızlanmaktadır. Gelişen teknolojik araçların ve uygulamaların neredeyse tamamının internet veya yerel ağlara bağlantılı olması kişilerin ya da kurumların bilgilerinin sunucularda depolanmasını sağlamaktadır. Bu tür teknolojilerin yaygınlaşmaya başladığı günden itibaren sunucularda depolanan veriler ivmelenerek artarak petabaytlara varan boyutlarda veri yığınlarının oluşmasını sağlamıştır. Böylesine büyüklükte olan bilgi yığınlarına Büyük Veri (Big Data) denilmektedir (Devenport & Dyche, 2013; Picciotto, 2020; Williams vd., 2020).

Büyük Veri; kişi veya kurumların çeşitli bilgilerinin elektronik ortamda depolanması ile oluşan veri ambarlarıdır. Bu veri ambarlarına örnek olarak sosyal medya platformları, elektronik devlet sistemleri, iş bulma firmaları gibi veri depolama alanlarının bulunduğu kurum ve şirketler gösterilebilir. Bu ambarlarda depolanan veriler; kimlik bilgileri, iletişim bilgileri, hobiler, eğitim bilgileri, sağlık ve finans bilgileri gibi her alandan oluşan verilerdir (Williams vd., 2020; Pratsri & Nilsook, 2020). Bu denli farklı tür ve içerikteki veri yığınlarının oluşması kurumların ve araştırmacıların çeşitli araştırmalar ve analizler yapmalarına önemli ölçüde katkı sağlamaktadır.

Çeşitli kurumların ve kişilerin oluşturmuş olduğu büyük verilerin analiz edilerek bilgiye dönüştürülmesi ihtiyacı oluşmuş bu sayede veri madenciliği ortaya çıkmıştır (Can vd., 2012). Disiplinler arası bir yöntem olan veri madenciliği, büyük verilerin işlenip analiz edilerek eldeki sistemlerin geliştirilmesi ve sistemlerdeki eksikliklerin giderilmesi için anlamlı veriler oluşturma işlemidir (Özbay, 2015a; Molluzzo & Lawler, 2015; Jiang

vd., 2015; Clayton ve Halliday, 2017; Yap ve Drye, 2018; Bulut ve Yavuz, 2019; Czyzewska ve Mroczek, 2020). Veri madenciliği yöntemleri ile yapılan çalışmaların finans, sağlık, muhasebe, mimari ve endüstri alanında yoğunlaştığı gözlenmiştir (Uzun, Y. Uzun, F. ve Çakar, 2019; Pan, 2018).

Veri madenciliğinin çeşitli sektörlerde olduğu gibi eğitim alanında da yaygınlaşmaya başladığı görülmektedir. Eğitim alanında öğrenci, öğretmen, eğitim içeriği, eğitim ortamı, ölçme ve değerlendirme sonuçlarının oluşturmuş olduğu veri yığınları bulunmaktadır. Özellikle eğitimde teknoloji kullanımının artması eğitsel alanda ki bu verilerin dijital ortamlarda tutulmasını sağlamaktadır (Bezerra & Silva, 2020). Eğitim ortamında oluşan bu denli veri yığınları büyük veri tanımına uygun olarak büyük veri kapsamında ele alınıp düzenlenip etkili veri tiplerine çevrilerek veri madenciliği yöntemlerine uygun hale getirile bilinir (Özbay, 2015a). Eğitim ve öğretimin daha etkili bir şekilde gerçekleştirilmesi için dijital ortamlarda tutulan bu verilerin anlamlı bilgilere dönüştürülmesi adına eğitsel veri madenciliği kavramı ortaya çıkmıştır. Diğer bir değişle eğitsel veri madenciliği; eğitim, istatistik ve bilgisayar bilimi ile ilişkili olan disiplinler arası bir alandır (Jalota & Agrawal, 2019; Tekin ve Öztekin, 2018; Romero & Ventura, 2013).

EVM çalışmaları ile eğitimin kalitesinin arttırılması hedeflenmektedir. Eğitim kalitesinin artırımına yönelik olarak EVM çalışmalarını iki grupta özetlemek mümkün olabilir (Pena & Ayala, 2013). Çalışmaların bir kısmı sınıflandırmaya yönelik eğitim ve öğretimdeki mevcut durumu ortaya çıkartmak için yapılan çalışmalar, diğer bir kısmı ise tahmine yönelik olarak eğitimdeki sorunların giderilmesini sağlamak için yapılan çalışmalardır (Jalota & Agrawal, 2019; Tekin ve Öztekin, 2018; Molluzzo & Lawler, 2015). Eğitim ortamında oluşan veriler özellikle ölçme ve değerlendirme verileri ile öğrenci başarısı tahmin edilebilir ve tahmin sonuçlarına göre öğrencilere ek öğrenim hizmeti ya da eksik öğrenmeler yeniden kazandırılabilir. Bu amaçla, öğrenci başarısı öğrenme sürecinin başında tahmin edilerek gelecekteki akademik performansının arttırılabilmesi büyük önem taşımaktadır. EVM çalışmalarının genelinde veri seti olarak bir öğrenci grubu ya da belirli bir örneklem olduğu ve çalışmalar için gereken verilerin ise anket, görüşme ve sistem verileri gibi yöntemler ile elde edildiği görülmektedir (Akgün ve Bulut, 2020; Hussain vd., 2018). Farklı yöntemler ile elde edilen veri yığınları düzenlenip işlenerek öğretim teknolojilerinin hedeflerinden biri olan öğrenenlerin öğrenme süreçleri desteklenebilir (Bezerra & Silva, 2020; Şahin, 2018).

EVM ve bu alandaki çalışmaların eğitimin kalitesini arttırmaya yönelik olması öğretim teknolojilerinin ana hedefleri ile örtüştüğünü göstermektedir. Öğretim teknolojileri amaçları arasında öğrenenlerin öğrenme sürecini desteklemek ve performanslarını arttırmak gelmektedir (Şahin, 2018). Öğrenci başarısını arttırmak için yapılan çalışmalar içerisinde tahmine dayalı veri madenciliği yöntemleri kullanılmaktadır. Bu yöntemler, mevcut veriler ile geleceğe yönelik kestirimde bulunmayı sağlayabilecek istatistiksel modellerin oluşturulması üzerine kurulmuştur. Tahmine dayalı veri madenciliği yöntemlerinde yüksek doğruluğa sahip tahmin modellerinin oluşturulmasında regresyon analizi yönteminin öne çıktığı gözlenmiştir (Altun, 2019; Bahadır, 2013).

Eğitim ortamında oluşan verilerin veri madenciliği yöntemleri ile işlenmesi şeklinde tanımlanan EVM ile öğrenci başarısı tahmin edilebilir ve tahmin sonuçlarına göre öğrencilere ek öğrenim hizmeti ya da eksik öğrenmeler yeniden kazandırılabilir (Romero & Ventura, 2020; Salal vd., 2019; Aldowah vd., 2019; Özbay, 2015b; Tekin, 2014). Bu çalışmada eğitim alanında büyük veri kavramı tanımlanmış ve İnönü Üniversitesi öğrencilerine ait ölçme değerlendirme verileri ile öğrenci başarısının artırımına yönelik EVM uygulaması yapılmıştır. Araştırma sonucunda eğitim alanında büyük veriyi EVM ile işleyip öğrenci akademik performansının artmasına yönelik tahminlerde bulunulmasına yardım eden bir model önerisi sunulmuştur.

1.2. Araştırmanın Problemi

Bu araştırmada, Büyük Verinin eğitim alanında önemine dikkat çekmek ve öğrenme analitikleri kullanılarak öğrenci akademik başarısını arttırmak için tahmine dayalı eğitsel veri madenciliği modeli geliştirilmesi temel amaç olarak belirlenmiştir. Bu kapsamda İnönü Üniversitesi Otomasyon Bilgi Sistemindeki öğrenci verileri kullanılarak öğrencilerin mezuniyet süresi tahmini ve Bilgisayar II dersi mezuniyet durumu tahminine yönelik eğitsel veri madenciliği çalışması yapılmak istenmiştir. Böylece büyük verinin ve veri madenciliğinin eğitim alanındaki önemi vurgulanabilecektir. Bu amaca yönelik olarak araştırmanın problem cümlesi "Öğrencilerin mezun olma süreleri ve Bilgisayar II dersi başarı durumları EVM'nin lojistik ve doğrusal regresyon modelleri kullanılarak kestirilebilir mi?" şeklinde belirlenmiştir. Bu genel amaç çerçevesinde aşağıdaki sorulara yanıt aranmıştır:

1. Öğrencilerin mezun olma süreleri öğrencilerin kişisel bilgileri (Cinsiyet, Medeni Durum, Yaş, İl) ve akademik bilgileri (Üniversiteye Yerleşme Puanı, Yerleşme Puan Türü, Lise Bitirme Puanı, Mezuniyet Notu) kullanılarak lojistik regresyon ile kestirilebilir mi?
2. Bilgisayar II dersi geçme durumu öğrencilerin kişisel bilgileri (Cinsiyet, Medeni Durum, Yaş, İl) ve akademik bilgileri (Üniversiteye Yerleşme Puanı, Yerleşme Puan Türü, Lise Bitirme Puanı, Bilgisayar I Dersi Geçme Durumu ve Notu) kullanılarak lojistik regresyon ile kestirilebilir mi?
3. Bilgisayar II dersi geçme notu öğrencilerin kişisel bilgileri (Cinsiyet, Medeni Durum, Yaş, İl) ve akademik bilgileri (Üniversiteye Yerleşme Puanı, Yerleşme Puan Türü, Lise Bitirme Puanı, Bilgisayar I Dersi Geçme Durumu ve Notu) kullanılarak doğrusal regresyon ile kestirilebilir mi?

1.3. Araştırmanın Önemi

Bloom (1984) yatığı çalışmada, öğrenci başarısının arttırılmasında öğretici desteğinin iki standart sapma değerinde etkili olduğunu göstermiştir. Bu bağlam ile öğreticiler teknolojinin sunmuş olduğu imkanlar ile öğrencilere ek destekler sağlayabilir. Eğitim kurumlarında biriken veri yığınları ile anlık oluşan veriler kullanılarak EVM çalışmalarının gerçekleştirilmesi öğrenimi destekler nitelikte olacaktır. Bu sayede öğrenciler hakkında daha detaylı ve kullanılabilir bilgilere ulaşılarak, öğrencinin geleceğe yönelik akademik başarısı tahmin edilebilir ve gerekli önlemlerin alınması sağlanabilir.

Araştırma kapsamında eğitsel veri madenciliğine dayalı mezuniyet durumu modeli ile üniversiteye yerleşen öğrencilerin kayıt sonrası mezuniyet süreleri tahmin edilmek istenmiştir. Böylelikle öğrenim süresini uzatma durumu gösterebilecek öğrencilerin belirlenerek gerekli desteklerin verilmesi ve öğrencinin zamanında mezun olması sağlanabilir. Bilgisayar II dersi geçme durumu ve notu kestirimi ile de üniversiteye yerleşen öğrencilerin bu dersten kalabilecek ya da düşük kazanımlar ile dersi geçebilecek düzeyde olanlar tespit edilerek gerekli akademik önlemlerin alınabilmesi hedeflenmiştir. Böylelikle öğrencilerin çağın gerekliliği olan bilgisayar okur yazarlığı konusunda tam donanımlı bir şekilde yetişmeleri sağlanabilir. Kestirim modelleri ile İnönü Üniversitesi akademik ve idari birimlerine, öğrencilerin akademik performanslarını izleme, öngörme ve müdahale etme şansı sağlanabilir.

Yapılan araştırma, belirli bir zaman ve öğrenci grubundan ziyade üniversitede kayıtlı tüm öğrencileri kapsadığından eğitim alanında oluşan büyük veriden veri madenciliği yöntemi ile veri işleme çalışması olarak değerlendirilmiştir. Böylece eğitim alanında büyük veri ile ilgili literatüre katkı sağlaması ve EVM uygulamaları ile de bilişim teknolojileri eğitiminin öneminin tekrar hatırlatılması açısından önemlidir. Araştırma eğitim alanında oluşan büyük veriyi EVM ile işleyerek öğrencilerin akademik başarısını izleme, öğrencilere rehberlik etme ve öğrenme destek sistemlerinin geliştirilmesine katkı sağlaması açısından önemli olduğu düşünülmektedir.

1.4. Araştırmanın Sınırlıkları

Bu araştırma, İnönü Üniversitesi Otomasyon Bilgi Sistemi üzerinde kayıtlı aktif öğrenimine devam eden ve mezun olan öğrencilerin, kişisel ve akademik verileri kullanılarak RapidMiner Studio programı ile geliştirilmiş eğitsel veri madenciliği kestirim modelleri (Mezuniyet Süresi ve Bilişim Dersi tahmin modelleri) ile sınırlıdır. EVM kestirim modellerinde kullanılan veri seti cinsiyet, medeni durum, yaş, il, üniversiteye yerleşme puanı, yerleşme puan türü, üniversiteden mezun olma süresi, lise mezuniyet notu, bilgisayar dersi iki dönem notu ve dersi geçme durumu verileri ile sınırlıdır.

1.5. Varsayımlar

Araştırma kapsamında İnönü Üniversitesinde lisans düzeyinde eğitim almış ve almakta olan öğrenci verileri ile veri seti oluşturulmuştur. Öğrenci verilerinden fazla sayıda eksik verisi bulunan kişilerin verileri veri setinden çıkartılmıştır. Düzenlenen veri setine veri madenciliği teknikleri uygulanarak oluşturulan modeller, rastlantısal bağlantılara dayanmamaktadır.

1.6. Tanımlar

Büyük Veri (Big Data): Artan hacimlerde ve her zamankinden daha yüksek hıza ulaşan, daha fazla çeşitlilik gösteren verilerdir.

Öğrenme Analitikleri (Learning Analytics): Öğrenmeyi ve öğrenmenin ortamlarını en uygun hale getirmek ve bu ortamları daha iyi anlamak için öğrenenlerin bağlamlarına dair verilerin ölçülmesi, toplanması, analiz edilmesi ve raporlanmasıdır.

Veri Madenciliği (Data Mining): Büyük ölçekli veriler arasından faydalı bilgiye ulaşma ve bilginin keşfedilmesi işlemidir.

Eğitsel Veri Madenciliği (Educational Data Mining): Eğitim ortamlarında oluşan ve depolanan verilerin eğitimin kalitesini arttırmak için veri madenciliği yöntemleri ile işlenip kullanılması sürecidir.

Otomasyon Bilgi Sistemi (OBS): İnönü Üniversitesi öğrencilerinin verilerinin tutulduğu çeşitli türlerde ve boyutlarda veri içeren veri tabanları sistemidir (https://obs.inonu.edu.tr/). Araştırma kapsamında araştırmacıya veri tabanına salt okuyucu grubunda erişim yetkisi verilmiştir.

RapidMiner Studio: Veri bilimi ve istatistik çalışmaları için gerekli tüm özellikleri barındıran ve kolay kullanımlı ara yüzü ile rahat çalışma ortamı sağlayan bir makine öğrenimi programıdır.

Yerleşme Puanı: Öğrencilerin üniversiteye yerleşirken kullanmış oldukları puan.

Yerleşme Puan Türü: Öğrencinin üniversiteye yerleşirken kullandığı puanın türü (Sayısal, Sözel, Eşit Ağırlık, Dil, Özel Yetenek).

Mezuniyet Durumu: Öğrencinin lisans öğrenim süresi dört yıl (8 yarı dönem) içerisinde bitirebilmesi ya da bu süreyi aşması.

Bilgisayar I ve II Dersleri: Üniversitenin çeşitli fakültelerinde bilişim dersi olarak verilen derslerin genel adı olarak tanımlanmıştır. Ders içerikleri aynı olmasına karşı bölümlerde isimlerinin farklı olmasından dolayı bu isimle nitelendirilmişlerdir.

BÖLÜM II

2. KURAMSAL BİLGİLER VE İLGİLİ ARAŞTIRMALAR

Bu bölümde araştırma konusu ile ilgili alan yazın taraması kapsamında geçen veri, büyük veri, veri madenciliği kavramları tanımlanmıştır. Araştırma ile ilgili temel kavramların tanımlanmasının ardından eğitim alanında yapılan büyük veri çalışmaları ve eğitsel veri madenciliği ile ilgili çalışmalar sunulmuştur.

2.1. Veri Kavramı

İnsanlığın varlığından bu yana bir nesneyi, olayı veya durumu anlatmak ya da isimlendirmek için çeşitli şekiller, semboller ve harfler geliştirilmiştir. Veri, herhangi bir işlemden geçirilmemiş olan gözlem veya ölçüm yöntemleri ile elde edilen her türlü değerdir (Şeker, 2013). TDK (2021) bilişim alanında veriyi, "Olgu, kavram veya komutların, iletişim, yorum ve işlem için elverişli biçimli gösterimi." şeklinde tanımlamıştır. Bununla beraber verinin farklı şekillerdeki tanımları da mevcuttur. Prytherch (2005)'e göre veri; veri tabanında bulunan bilgilendirme amaçlı olarak kullanılan terimdir. Yılmaz (2009) göre ise veri; "tek başına bir anlam ifade etmeyen veya tek kullanılamayan, bilgilendirme ve bilgiye temel oluşturan ilişkilendirme, yorumlama, anlama ve analiz etmeye gerek duyulan ham bilgi şeklinde" tanımlamıştır.

Veri sayı, resim, ölçüm, metin, sembol, olay biçiminde temsil edilebilir. Deney, ölçüm, araştırma gibi yöntemler ile elde edilen değerler birer veri olarak nitelendirilir (Özen, 2014). Verinin tanımlanması kadar nasıl nitelendirilip gruplandırıldığı da önemli olmaktadır. Jeffery (2016) veriyi aşağıdaki gibi gruplara ayırarak nitelendirmiştir.

- Yapılandırılmış, yapılandırılmamış, yarı yapılandırılmış
- Statik, dinamik, akan
- Güvenli / açık, özel / halka açık
- Ücretli / ücretsiz

- Açık hükümet verisi
- Açık veri
- Büyük veri

Verinin tanımı ve nasıl nitelendirildiğinden sonra enformasyon kavramını da açıklamak gerekmektedir. Enformasyon, TDK (2021)' ya göre danışma, tanıtma, haber alma, haber verme ve haberleşme olarak tanımlanmıştır. Bir konu ile ilgili bilinmeyeni veya belirsizliği gidermeye yardımcı olan ifadelere enformasyon denilmektedir (Altun, 2019). Veri, gözlem ve ölçüm ile belirlenen bir değerdir ve işlenerek enformasyon haline getirilir. Enformasyon haline dönüşmüş veri deneyim, yorum ve fikirler ile bilgiye dönüştürülür (Altun, 2019). Veri organize edilip işlenerek daha değerli bilgilere ulaşmasında bilişim teknolojileri kullanılmaktadır. Verileri organize etmek için kullanılan bilişim teknolojilerinden biri veri tabanıdır. Veri tabanı bir bilgisayar sisteminde depolanmış yapılandırılmış veri kümesi şeklinde tanımlanabilir. Veri tabanları birbirleri ile ilişkili verilerin fiziksel ve mantıksal özelliklerinin tutulduğu ambarlardır (Altun, 2019). Veri tabanı genel olarak bir yönetim paneli tarafından kontrol edilir. Veri tabanı yönetim panelinde kayıt ekeme, silme, güncelleme ve sorgulama yazılımları kullanılarak veri hızlı bir şekilde işlenir (Oracle, 2019; Burma, 2009).

Artan veri depolama işlemleri veri tabanlarının çoğalmasını sağlamakta ve farklı alanlarda yüksek hacme sahip veri tabanları oluşturmaktadır. Bu denli yüksek hacimde veri tabanları birleştirilip işlenmesi büyük verinin temelini oluşturmaktadır. Araştırmalar, gözlemler, internet, sosyal medya gibi birçok kaynaktan elde edilen veri yığınları büyük veriyi oluşturur (Oracle, 2019; Jaiswal, 2018; Doğan ve Arslantekin, 2016).

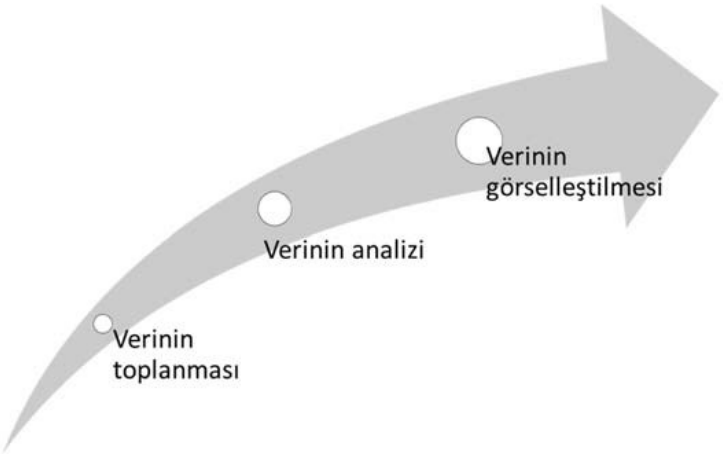
2.2. Büyük Veri Kavramı

Büyük veri ilk kez 2000 yılında Francis X. Diebold tarafından, "Makroekonomik Ölçümler ve Kestirim için Büyük Veri Dinamik Faktör Modelleri (Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting)" isimli bildiride 8. Dünya Ekonometri Kongresi'nde kullanılmıştır (Diebold, 2003; Jacobs, 2009; Gürsakal, 2013). Büyük veri genel olarak birbirinden farklı veri kaynaklarından toplanan verilerin analizi, işlenmesi ve depolanmasını ele alan bir kavramdır. Büyük verinin kesin bir tanımı

bulunmamakla beraber farklı açılardan tanımlama yapılmaktadır (Schönberger ve Cukier, 2013). Monino ve Sedkaoui (2016) büyük veriyi terim olarak, "organizasyon için kullanılan verinin hacmi kritik seviyeye ulaştığında gerekli olan yeni depolama teknolojileri ve yeni kullanım yöntem ve yaklaşımlarının kullanılması" şeklinde tanımlamıştır. Sigman ve diğerleri (2014) ise büyük veriyi çeşitli türde bilgilerin depolanması ve eş zamanlı olarak büyük miktarda verinin işlemesi şeklinde tanımlamışlardır. Büyük veri; Hadoop, yapılandırılmış SQL, yapılandırılmamış SQL, HIVE ve H-BASE sistemleri ile kullanılmaktadır (Sigman, 2014; Prinsloo vd., 2015).

Günümüzde internet kullanımının artması ve bilgiye erişimin kolay olması ile günlük üretilen veri miktarı katlanarak artmaktadır (Aktan, 2018). Yüksek miktarda veri transferi ve depolanması büyük veri kavramını oluşturmuş ve bu alanda çalışmaların yapılmasını sağlamıştır. Eğitim alanında büyük veri çalışmaları öğrenme analitiği olarak isimlendirilmiştir. Özellikle çevrim içi öğrenme ortamlarında oluşan büyük verilerin depolanması ve işlenmesi büyük veri kavramını öğrenme analitiği olarak isimlendirilmesine neden olmuştur. Bu alandaki çalışmalar ise genel olarak yükseköğretim düzeyinde ilgi görmüştür (Booth, 2012; Johnson, Adams, ve Cummins, 2012; Sin ve Muthu, 2015). Büyük verinin saklanması ve bu veriden değer üretmek oldukça zor olduğundan genellikle verilerden çeşitli örneklemler alınarak analizler yapılmaktadır. Fakat yapılan analizlerde örneklemden üretilen değerlerin evreni tam olarak yansıtmadığı gözlenebilir.

Büyük veri ile ölçülemeyen, saklanamayan, analiz edilemeyen ve paylaşılamayan bilgilerin büyük bir çoğunluğu düzenli veriye çevrilebilir (Schönberger ve Cukier, 2013). Hadoop gibi büyük veri teknolojileri ile tüm veri üzerinden çalışmalar yapılarak daha doğru, etkili ve kapsamlı sonuçlar elde edilmektedir (Sarıoğlu ve Koç, 2017). Büyük verinin hızlı gelişim göstermesi ile finans, sağlık, muhasebe, pazarlama ve eğitim alanlarında depolanan verilerin kullanım ihtiyacı oluşturmuştur. Bu alanlarda büyük veri işlenerek kullanılmaktadır. Büyük verinin işlenmesi ile gerçek değerinin ortaya çıkartılması üç aşamada gerçekleştirilir. Bunlar Şekil 2.1'de görüldüğü gibi, verinin toplanması, analiz edilmesi ve görselleştirilmesi şeklindedir (Bozkurt, 2016; Daniel, 2015).



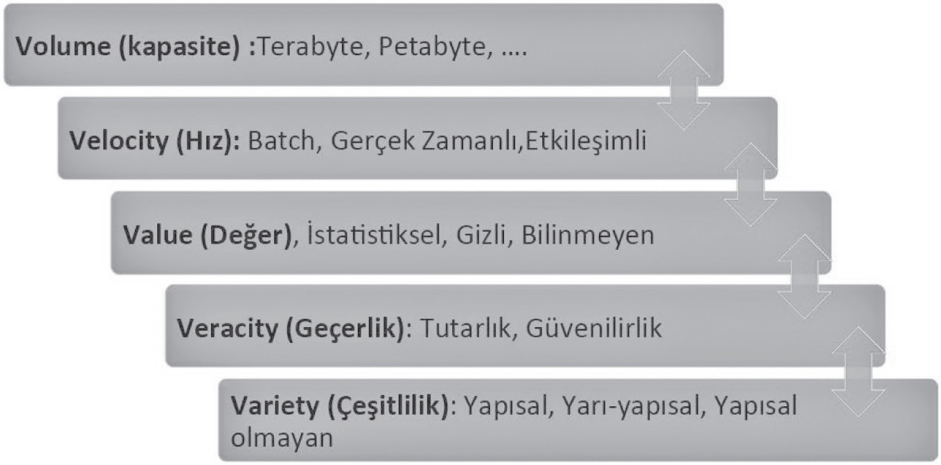
Şekil 2.1. Büyük Verinin Üç Temel Aşaması (Daniel, 2015).

Büyük veri yığınlarının oluşması ile bu yığınlardaki bilgilerin keşfedilip işlenmesi gerekliliğini oluşturmuş ve buda analitik yaklaşımlar kullanılması gereksinimi oluşturmuştur. Büyük verinin analiz edilebilmesi için soyut ve somut teknolojiler ile işlenebilir olması gerekmektedir. Soyut teknoloji; veriyi toplayan, analiz eden ve raporlayan yazılımlar olarak gösterilirken, somut teknoloji; soyut teknolojinin işlevsel hale getirileceği bilgisayar ve bilişim sistemleri olarak gösterilmektedir (Jacobs, 2009; Prinsloo vd., 2015; Bozkurt, 2016). Eğitim alanında oluşan büyük verinin soyut ve somut teknolojiler ile kullanılmasını sağlayan yaklaşıma ise öğrenme analitiği denilmektedir.

Büyük veri oluşturulurken ya da mevcut veri yığınlarının büyük veri olarak tanımlanması için aşağıda tanımlanan büyük veri bileşenlerine uygun olması gerekmektedir.

2.2.1. Büyük Veri Bileşenleri

Büyük veri bileşenleri genel olarak 3V (Hacim, Hız ve Çeşitlilik) olarak değerlendirilse de detaylı incelendiğinde temelde 5V olarak ele alınmaktadır. Bunlar, kapasite (volume), hız (velocity), değer (value), geçerlik (veracity) ve çeşitlilik (variety) olmak üzere beş bileşen ile ifade edilmektedir (Şekil 2.2).



Şekil 2.2. Büyük Veri Bileşenleri.

Hacim (Volume): Veri boyutunu terabayt ve petabayt gibi ölçü birimleri ile ifade edilen depolama alanlarıdır. Veri depolama alanları ve işleme maliyetlerinin ucuzlaması daha fazla verinin sunucularda depolanmasını sağlamaktadır (Atan, 2010). Büyük verinin bilinen en önemli özelliği ve problemi verinin hacmidir. Verinin hacminin büyük olması büyük veri olarak nitelendirilmesinde yeterli değildir. Veri miktarının istenilen zamanda analizinin yapılamadığı durumlarda veri hacmi problem olmasına büyük veri denilmektedir (Jaiswal, 2018; Sarıoğlu ve Koç, 2017).

Önceden tanımlı bir modele ait olmayan veriler yapılandırılmamış veri olarak nitelendirilir. Bu veriler; e-postalar, video kayıtları, resimler, sesler gibi birçok kaynaktan oluşmaktadır. Bu verileri analiz etmek ve işlemek güçleşmektedir. Bu nedenle bu bilgiler büyük veri kapsamına girip ve önemli bilgilerin %80'ini oluşturmaktadırlar (Grimes, 2005).

Hız (Velocity): Verinin üretilme hızıdır. Günlük hayatta kullanılan telefon veya internete bağlı tüm cihazlar yüksek hızlarda veri üretmektedir. Her gün insanların sosyal medyada birçok bilgi paylaşması bu hızın ne kadar büyük olduğunu göstermektedir. Sosyal medya platformlarından biri olan Youtube'a dakikada 48 saatlik video yüklenmektedir. Bu kadar hızlı bir şekilde artan büyük veri, verinin işlev sayısının ve çeşitliliğinin de hızlı bir şekilde artmasını sağlamaktadır. Verilerin bu kadar hızlı artması büyük verinin yüksek hızda bağlantısını ve geniş bant büyüklüğünü gerektirmektedir. Hızlı bir şekilde üretilen verinin gerçek zamanlı analiz edilmesi ve yönetilebilmesi ise büyük verinin diğer bir problemi olarak kaşımıza çıkmaktadır (Jaiswal, 2018; Schaeffer ve Olson, 2014).

Değer (Value): Verilerin analiz edilip önemli hale getirilmesi ile oluşan değerdir. Değer elde edilemeyen veriler anlamsızdır. Verilerden üretilen değer verinin içeriği, üretilme amacı ve uygulama alanına göre değişiklik göstermektedir. Verilerden bu özelliklere göre değer üretmek geleneksel yöntemler ile çok zor bir durumdur. Bu problemde verinin büyük veri bakış açısı ve büyük veri teknolojileri ile analiz edilmesini gerektirmektedir (Sarıoğlu ve Koç, 2017).

Doğruluk (Veracity): Büyük veri içerisinde düzensiz veri yığınları verilerin doğruluğunu olumsuz yönde etkilemektedir. Doğruluğundan emin olunamayan veriler ile yapılan analizler gerçek değerleri yansıtmazlar. Büyük veriye özgü teknolojiler ile verinin doğruluğundan ve analize uygunluğundan emin olunmalıdır (Jaiswal, 2018).

Çeşitlilik (Variety): Verilerin farklı formatlarda ve kaynaklarda olması büyük verinin bir diğer özelliğidir. Büyük veri farklı kaynaklardan oluştuğu için heterojen bir yapıdadır. Sosyal ağlarda metin ve görseller, veri tabanındaki isim kayıtları, ses kayıtları ham verilerdir. Bu verilerin kullanıma hazır olması nadir bir durumdur (Dumbill, 2013). Üretilen veriler yapısal, yarı yapısal ve yapısal olmayan formatlarda karşımıza çıkmaktadır. Yapısal olan veriler veri tabanlarında tutulan ilişkisel veriler, yarı yapısal veriler ise belirli başlıklar altında saklanan ve düzenlenebilen XML formatındaki verilerdir. JSON formatında yapısal olmayan veriler ise ses, video ve metin dosyalarından oluşan ve büyük verinin %80'ini oluşturan verilerdir. Telefon, tablet ve entegre devrelerden üretilen farklı yapılardaki verilerin bir arada tutulması verileri çıkar-dönüştür-yükle işlemlerinde yeni büyük veri teknolojilerinin kullanılmasını zorunluluk haline getirmiştir. (Jaiswal, 2018; Demirtaş ve Argan, 2015; Sarıoğlu ve Koç, 2017).

Veri yığınlarının büyük veri olarak değerlendirilmesi için yukarıda tanımları verilen 5V kavramlarının her birini barındırması gerekmektedir. Büyük veri tanım ve bileşenlerine uygun olduğu belirlenen veri yığınının büyük veri olarak nitelendirilmesi, veri kavramının önem kazanmasını sağlamış ve işlenmesi gerekliliğini ortaya çıkartmıştır. Büyük veriden anlamlı veriler elde edilmesi veri madenciliği kavramının ortaya çıkmasını sağlamıştır (Can vd., 2012).

2.3. Eğitimde Büyük Veri

Eğitim; bireye istenilen yönde yaşantılar yolu ile kazandırılan davranış değişikliğidir (Ertürk,1973). Çeşitli ortamlarda ve yaş grupların da gerçekleştirilen eğitim geçmişten günümüze farklı biçimlerde uygulanmış, günümüzde ise çağdaş eğitimin gerekliliği olan yapılandırmacı yaklaşım modeli ile verilmeye başlanmıştır. Bu modelde eğitim öğrenci merkezli bir yapı olmakta ve bilgi bireyler tarafından yapılandırılmaktadır. Bu yaklaşım ile bireyin ve toplumun ihtiyaçları ön planda tutulmakta ve hedefler bu ihtiyaçlara göre belirlenmektedir (Ayas & Haluk, 2014). İhtiyaçların belirlenmesi uzun süre ve maliyet gerektirmekte olduğu için bu yaklaşım tam anlamıyla uygulanamamakta ve eğitimin kalitesi beklenildiği şekilde ilerlememektedir. Bu ihtiyaçların belirlenmesi geçmiş yaşantı, öğrenme ve çeşitli verilerin işlenmesi ile daha etkili olacaktır. Bu nedenle büyük veri eğitimde ihtiyaçların ve hedeflerin belirlenmesinde, öğrencilerin eğitim başarılarının istenilen seviyede olmasında önemli bir etken olacaktır (Jacobs, 2009; Li, 2009).

Eğitim, gelişen teknoloji ile her geçen gün artan bilginin daha iyi ve etkili öğretimini sağlamak için sürekli olarak kendini güncel tutmalıdır. Eğitimin teknoloji ve bilimdeki gelişmeleri etkili ve doğru bir açıdan analiz edip kendini yenileye bilmesi için büyük veri ile çeşitli uygulamalar yapılmalı, yeni öğrenme yöntemleri oluşturulmalıdır (Pan, 2018; Jacobi, 2014). Ancak eğitim alanındaki büyük veri çalışmaları her geçen gün biraz daha gelişme göstermekte olmasına karşın diğer alanlar kadar ilgi görmemekte ve büyüyememektedir. Bunun en önemli nedenlerinden biri büyük veriye erişim sorunu olarak gösterilebilir. Birçok devlet kuruluşu büyük veriye erişimi sadece belirli bölümlerdeki verilerin erişime sunulması ile sınırlandırmaktadır. Bunlar genellikle; nüfus sayımı, enerji kullanımı ve bütçe raporları gibi sınırlı konulardır (Ohlhorst, 2013). Eğitim alanında çalışmaların artması ve daha sağlıklı sonuçların elde edilmesi için eğitim kuruluşlarının verilerini açık veri haline getirmeleri gerekmektedir (Sharar, 2017; Sarıoğlu ve Koç, 2017; Aktan, 2018).

Alan yazın incelendiğinde eğitimde büyük veri ile yapılan çalışmalarda genel olarak büyük veri tanımlanmış ve büyük verinin kullanılmasının eğitime çok büyük katkılar sağlanacağı belirtilmiştir (Shi-rui, Hong- feng, 2018). Yapılan çalışmalar yükseköğretim düzeyinde yoğunlaştığı görülmüş ve yükseköğretimde öğrenci ihtiyaçları, öğretim ihtiyaçları ve kurumların olanaklarını arttırma amaçlı çeşitli çalışmaların

yapıldığı gözlenmiştir (Picciano, 2012). Büyük verinin eğitim alanında kullanımına yönelik çalışmalardan bazıları aşağıda belirtilerek bu konuda ne gibi çalışmaların yapıldığı belirtilmek istenmiştir.

Li (2019) büyük veri uygulamaları ile yükseköğretimdeki öğrencilerin artan bilgi ve gelişmelere kütüphanelerde daha hızlı ve güvenli ulaşabileceklerini belirtmiştir. Zhang (2017), hızlı yaygınlaşan küreselleşmenin Çin'in kültürüne olan etkilerini incelemiş ve bu etkilerin olumsuz yönde geliştiğini belirtmiştir. Öğrencilerin kültürel değerlere göre yetiştirilmesi ve kültürel değerler konusunda araştırmalar yapılması için yenilikçi bir öğrenme gerektiğini belirtmiş. Bu yenilikçi öğrenmenin geliştirilmesi için ise büyük verinin anlaşılması ve işlenmesinin gerekli olduğunu belirtmiştir.

Tetko ve diğerleri (2016), kimya alanında artan veri hacmi ve bu verilerin işlenmesi için geliştirilen BIGCHEM projesinin kimya alanına etkilerini açıkladıkları makalede kimya alanında oluşan büyük verilerin işlenmesi için makine öğrenmesi ve veri madenciliği yöntemleri ile uygulamalar geliştirilebileceğini belirtmişlerdir. Kimya alanında daha etkili ve yeni çalışmalar yapılabilmesi için bu alanda çalışanların büyük veri eğitimi almaları gerektiği belirtilmiştir.

Olayinka ve diğerleri (2017), Washington DC merkezli Küresel Sağlık Üniversitesi üyelerine sağlık alanında çalışanlara yönelik büyük veri üzerine bir eğitimin gerekliliği hakkında bir çalışma yapmışlardır. Sigman ve diğerleri (2014), büyük verinin tanımı, depolanması ve kullanılan sistemleri incelemiş ve bir anket düzenlemişlerdir. Yapılan anket sonucuna göre kurumların yüzde 64'ünün büyük veri alanında yatırım yaptıklarını ve bunun bir istihdam alanı oluşturacağı belirlenmiştir. Bu araştırmada büyük verinin kullanılabilmesi ve işlenebilmesi için profesyonel anlamda büyük veri eğitiminin verilmesi gerektiği de belirtilmiştir.

Fiofanova (2021), eğitimde veriye dayalı yönetim ile eğitimdeki yöneticilerin kendilerini geliştirmesinde verilerden yararlanılması gerektiğini ifade etmiştir. Bu ortamın sağlanması için eğitsel verilerin elektronik ortamda kayıt altına alınmasının gerekliliğini belirtmiştir.

İncelenen çalışmalarda genel olarak büyük veri hakkında bilgilendirme yapıldığı ve eğitimde büyük veri kullanılarak birçok alanda verimliliğin arttırılacağı belirlenmiştir. Büyük veri kullanılarak eğitim alanında önemli gelişmelerin sağlanabileceği ve eğitim başarısının arttırılabileceği düşünülmektedir.

Günümüzde teknolojinin her alanda olduğu gibi eğitim alanında da yaygın kullanılması ve verilerin elektronik ortamda saklanması birçok açıdan fayda sağlamaktadır. Salgın nedeni ile dünya genelinde yaşanan sorunların bir kısmına büyük veri ile çözümler bulunmuştur. Bunlardan en önemlisi salgın sürecini değiştiren aşıların üretimidir. Ülkelerin sağlık sistemlerindeki veriler dünya Sağlık Örgütü tarafından toplanarak aşı geliştirmek isteyen firmalara bir büyük veri alt yapısı oluşturulmuştur. Eğitim alanında yaşanan problemlere de büyük veri ile çözümler üretilebilir mi? sorusunu akla getirmiştir. Birçok farklı değişkenin etkilediği eğitim sistemini böylesi bir dönemde en kısa zamanda ve en iyi bir şekilde yönetmek için sağlık sisteminde olduğu gibi büyük veri oluşturulup çeşitli değişkenler hesaplanarak uygun bir eğitim modeli geliştirilmesi sürecin daha verimli olmasını sağlayabilir. Ayrıca uzaktan eğitim sunucularında biriken veri hacminin daha önce olmadığı kadar büyük olması pandeminin olumlu sonuçlarından biri olarak görülebilir.

Araştırmada ele alınan konuların eğitim ve öğretim alanında büyük verinin katkısının incelenmesine olanak sağlayacağı düşünülmüştür. Büyük veri ile eğitim sistemindeki sorunların daha çabuk ve kolay bir şekilde belirlenebileceği ve ileri yönlü programların yapılmasında kolaylıklar sağlayacağı düşünülmektedir. Büyük verinin eğitim alanında kullanılmasına yönelik çalışmaların bu alana gereken önemin verilmesine katkı sağlayacağı ve olası sorun ya da modellerde büyük veri sistemlerine başvurulabileceği kanısını ortaya çıkmıştır. Büyük veri sistemlerinden modeller oluşturulması için veri madenciliği yöntemi kullanılır. Eğitim alanında oluşan büyük verinin işlenmesi ve anlamlı verilerin oluşturulması için ise eğitsel veri madenciliği yöntemleri kullanılmalıdır (Özbay, 2015b).

2.4. Veri Madenciliği Kavramı

Artan teknoloji kullanımı ile çeşitli bilgilerin depolanması sonucu veri tabanları oluşmaktadır. Veri tabanlarının kullanımının artması ile çözümlenemeyecek derecede karmaşık bilgi yığınları meydana gelmektedir. Veriler arasındaki ilişkilerin çözülmesi ve anlamlı verilerin ortaya çıkarılması için kullanılan yöntemler yetersiz kalmıştır. Böylesine büyük miktardaki verinin analiz edilmesi için bilgisayar teknolojileri, istatistik, veri tabanı teknolojileri ve diğer disiplinleri bir araya toplayan veri madenciliği ortaya çıkmıştır (Can ve arkadaşları, 2012).

Veri madenciliği eldeki sistemlerin geliştirilmesi ve sistemlerdeki eksikliklerin giderilmesi amacı ile var olan verilerden anlamlı bilgiler elde etme işlemidir (Özbay, 2015a). Veri madenciliği ile depolanmakta olan her veri kullanılarak analiz yapılabildiği için birçok alanda kullanılmaktadır. Bilimin bütün alanlarına yansımış ve bununla kalmayıp hayatın her alanında kullanılmaya başlanmıştır (Özdemir, 2016). Bunlara sağlık, endüstri, mühendislik, pazarlama ve eğitim alanları örnek gösterilebilir. Veri madenciliğinin bu denli fazla alanda kullanılması aşağıda sıralanan birçok tanımı beraberinde getirmiştir.

- Veri madenciliği, çok sayıda veri analiz yöntem ve aracı kullanılarak büyük veri tabanlarındaki gizli bilgi ve yapıyı açıklamaktır (Oğuzlar, 2004)
- Veri madenciliği; büyük veriden gizli kalmış, değerli ve kullanılabilir bilgilerin çıkarılma işlemidir (Koyuncugil, 2007)
- Veri madenciliği; matematik ve istatistiksel yöntemler kullanılarak büyük veriden anlamlı ilişkilerin, örüntülerin ve trendlerin keşfedilme sürecidir (Gartner Group, 2013)
- Veri madenciliği; analitik metot ve araçlar ile büyük yığınlar halindeki veriyi işlemektir (Gupta, 2014)
- Veri madenciliği; veri tabanlarındaki geçerli, yeni, faydalı ve anlaşılır örüntülerin açığa çıkartılması için gerçekleştirilen süreç (Akpınar, 2014).

Veri madenciliği tanımları incelendiğinde, büyük veri yığınlarından önemli veri elde etme işlemi veri madenciliği olarak görülmektedir. Değerli veriler ile karar vermeye yardımcı modeller geliştirilerek daha doğru sonuçların elde edilebileceği söylenebilir.

Çalışmalarda kullanılan yöntemler incelendiğinde, veri madenciliği uygulamalarının tanımlama ve tahmin yapmak amacı ile kullanıldığı görülmüştür. Tanımlayıcı modeller; yorumlanabilecek veriyi barındıran örüntüleri bulabilme, tahmin modellerinin ise; bağımsız değişkenler kullanılarak bağımlı değişkenin bilinmeyen yönünü bulmak ya da geleceğe ilişkin kestirimde bulanmak için yapılmaktadır (Özdemir, 2016).

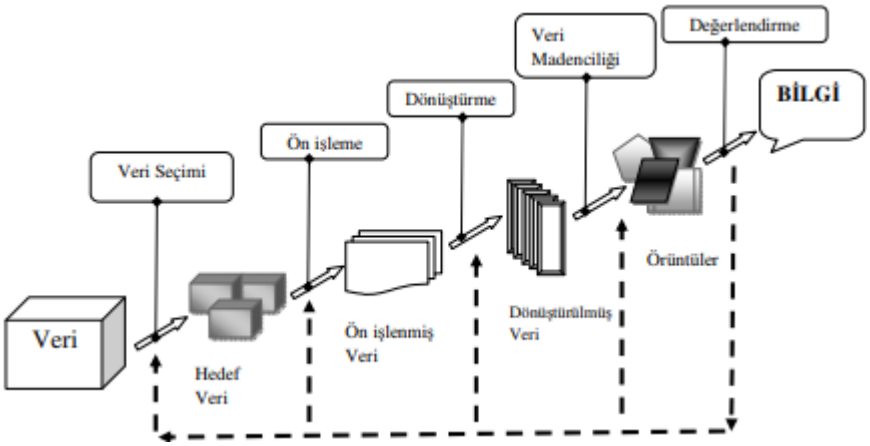
Veri madenciliği büyük veriden anlamlı veriler oluşturma işlemi olduğundan yapılması gereken özel bir süreç bulunmaktadır. Yapılan çalışmanın amacına ulaşabilmesi için veri madenciliği süreçlerinin takip edilmesi gerekmektedir.

2.4.1. Veri Madenciliği Süreci

Veri madenciliği uygulamalarının farklı disiplinleri içermesi ve büyük veri yığınlarından veri elde etmesi ortaya bazı güçlükler çıkartmaktadır. Bu güçlüklerin önlenmesi ve çözülmesi için standart bir veri madenciliği süreci oluşturulmuştur. Geliştirilmek istenen veri madenciliği uygulamalarında izlenmesi gereken sürecin temel adımları bulunmaktadır (Shearer,2000; Tufferry,2011; Özdemir, 2016). Aşağıda sıralanan temel adımların izlenmesi doğrultusunda kullanılacak süreç tasarımının belirlenmesi uygulamanın en iyi sonuç vermesini sağlayacaktır.

- Problemin/Hedeflerin tanımlanması
- Verilerin Hazırlanması
- Modelin kurulması ve değerlendirilmesi
- Modelin kullanılması
- Modelin izlenmesi

Veri madenciliği çalışmaları genel anlamda yukarıdaki adımlara uygun olarak gerçekleştirilir. Çalışmanın problemi belirlendikten sonra büyük veri yığınlarından problem durumuna uygun anlamlı veriler çekilerek bu veriler hazırlanır. Oluşturulmak istenen model için verilerin bir kısmı öğrenmeye ayrılır, kalan veriler ise test için kullanılarak model oluşturulur ve test edilir.



Şekil 2.3. Bilgi Keşfi Sürecinde Veri Madenciliği (Savaş vd. 2012).

Şekil 2.3'te görüldüğü gibi veri madenciliği süreci 5 aşamadan oluşmaktadır. Bu aşamalar aşağıda incelenmiştir.

Problemin/Hedefin Tanımlanması: Veri madenciliği sürecinin zor ve en önemli aşaması ilk aşamadır. Yapılacak çalışmanın amaçlarının belirlendiği bu aşama veri madenciliği hedeflerinin belirlendiği ve çalışmanın planının geliştirilme kısmını kapsamaktadır. Doğru bir şekilde belirlenmemiş amaç süreç içerisinde yanlışlıklara neden olacak ve gerçekleştirilmek istenen amaca hizmet etmeyecektir (Diller, 2016). Bu nedenle problemin sürecin başında doğru bir şekilde ortaya konulması veri madenciliği yapılacak çalışmaya bir yol çizmesini sağlayacaktır (Balaban ve Kartal, 2015).

Verinin Hazırlığı: Hedefler doğrultusunda verilerin toplanması ve toplanan veri setinin hedeflere ulaşacak düzeyde işlenmesi aşamasıdır. Bu bölümde veri setinden oluşturulmak istenen model ya da modeller için gerekli olan değişkenlerin seçimi, verilerin seçilip temizlenmesi ve bu verilerin analiz edilebilecek düzeye getirilme işlemleri yer almaktadır (Altun, 2019). Ayrıca bu adımda veri setinden alınan verilerin nominal (metin) türde olanların numerik (sayısal) türe dönüştürülmesi ve veri setinde eksik verisi bulunan verilerin atılması işlemleri uygulanmaktadır (Erşahin, 2008).

Modelin kurulması ve değerlendirilmesi: Bu aşama çalışmanın amacına ve problemine uygun olarak geliştirilmek istenen modelin seçimi, uygulanması ve optimal değerlere göre ayarlanmasını içermektedir. Çalışmada birden fazla model kullanılabilir ancak bazı modeller farklı problem durumu için yapılacaksa veri hazırlama sürecine tekrar dönülebilir (Özdemir, 2016). Değerlendirme, oluşturulan modelin sürecin ilk aşamasındaki hedeflere uygunluğunun test edilmesidir. Bu aşamada geliştirilen örüntüler yorumlanarak bilgiye dönüştürülür (Şeker, 2018).

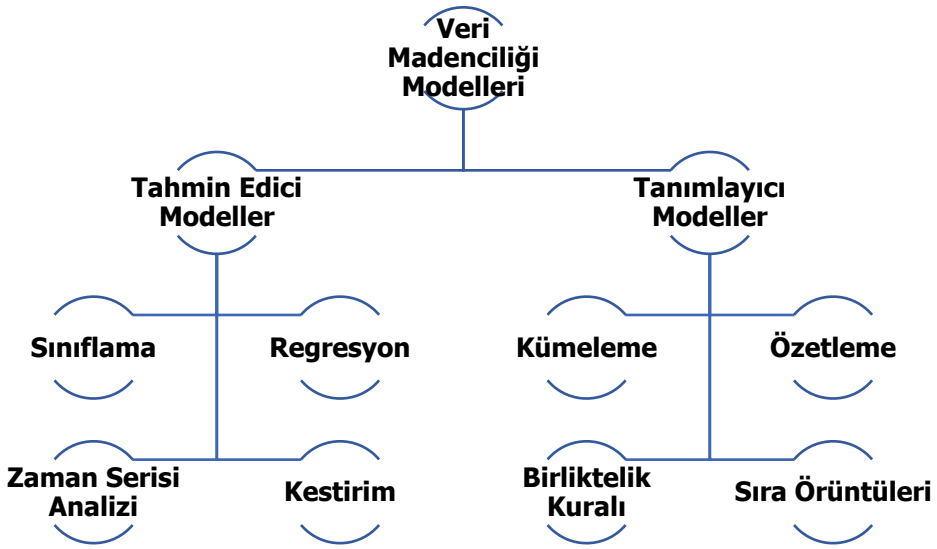
Modelin kullanılması: Oluşturulan veri madenciliği modelleri ile başarı sağlanan uygulamanın kullanılması ve başka uygulamalar için aktarıldığı bölümdür. Modelleme ile elde edilen bulguların düzenlenmesi ve hedefler ile birleştirilmesi gerekmektedir. İhtiyaç duyulması halinde uygulama adımı raporlama ya da başka sistemlere entegre edilebilir (Erşahin,2008).

Modelin izlenmesi: Bu aşama geliştirilen uygulamanın yayılma planının oluşturulup, takip ve bakımının planlandığı, raporların hazırlandığı ve projenin değerlendirildiği kısımdır. Kurulan model sürekli izlenip yeniden düzenlenmesi gerekecektir (Altun, 2019; Savaş, Topaloğlu ve Yılmaz, 2012).

Veri madenciliği standart sürecini temel alarak bazı danışmanlık şirketleri, standart sürece benzer adımların oluşturduğu kendine özgü yöntemler geliştirmişlerdir. Bu süreçlerden üç yaklaşım ön plana çıkmaktadır. Bunlar; veri tabanlarından bilgi keşfetme süreci olan KDD, SAS tarafından geliştirilen SEMMA ve araştırmada kullandığımız CRISP-DM (Endüstriden Bağımsız Standart Veri Madenciliği) süreçleridir (Diler, 2016; SVE, 2019).

2.4.2. Veri Madenciliği Modelleri

Veri madenciliğinde kullanılan modeller temelde iki kategoriye ayrılmaktadır. Bu modeller Şekil 2.4'te görüldüğü gibi tahmin etmeye ve tanımlamaya yönelik modellerdir (Aydın ve Özkul, 2015). Tahmin edici modeller, mevcut veriler kullanılarak bir model geliştirilir ve geliştirilen model ile sonuçların önceden bilinmeyen veri yığınları için sonuç tahmininde bulunulmasını sağlar. Tanımlayıcı modeller ise karar vermede rehberlik edecek veri örüntülerinin tanımlanmasını sağlamaktadır (Özbay, 2015a; Tekin ve Eymir, 2016).



Şekil 2.4. Veri Madenciliği Modelleri.

Veri madenciliği modellerinden Tahmin edici modeller; sınıflama, regresyon, zaman serisi analizi ve kestirim modelleridir. Tanımlayıcı modeller; kümeleme, özetleme, birliktelik kuralı ve sıra örüntüleri modelleridir.

2.4.2.1. Tanımlayıcı Modeller

Tanımlayıcı modeller veriler arasında gizli kalmış ilişki, bağlantı ve örüntülerin tanımlanmasını sağlamaktadır. Karar vermeye rehberlik etmede kullanılan tanımlayıcı modeller, var olan veriyi yorumlayarak alt veri setlerinin özelliklerinin tanımlanmasını sağlamaktadır. Veri setlerinin tanımlanması, yeni bir verinin mevcut yapıya dahil edilmesinde karar almaya yardımcı olur (Erşahin, 2008). Tanımlayıcı modeller aşağıda detaylı açıklanmıştır.

Kümeleme: Diğer bir adı bölümleme olan kümeleme benzer özelliklere sahip nesnelerin gruplara ayrılması işlemidir. Çok boyutlu ortamlarda kendine has özellikler sergileyen veri gruplarının oluşmasında kullanılır (Topuz, 2021). Benzer özellik gösteren veriler kümeleme yönteminde aralarındaki uzaklıklara göre gruplandırılır. Kümeleme işlemleri hiyerarşik (en yakın komşu ve en uzak komşu algoritmaları) ve hiyerarşik olmayan (k-ortalamalar) yöntemler ile yapılmaktadır (Diler, 2016).

Özetleme: Genelleme ya da nitelendirme olarak da isimlendirilen özetleme, veri tabanındaki özet bilgilerin ortaya çıkartılma işlemidir. Basit açıklamalar ile veriyi alt kümelere eşleyebilmek için verinin çeşitli parçalarına ulaşılması ile gerçekleştirilir. Veri tabanı içeriğindeki bilginin kısa olarak nitelendirilmesi işlemidir (Altun, 2019).

Birliktelik Kuralı: Veri tabanında yer alan veriler arasındaki ilişkilerin incelenerek, hangi olay ya da bulguların eş zamanlı olarak gerçekleşeceğinin tespit edilmesidir (Özkan, 2008). Literatürde market sepet analizi olarak da geçmektedir. Genelde pazarlama alanında kullanılan bu analiz yöntemi, tüketim alışkanlıklarının ve birlikte sunulacak ürünlerin tespit edilmesinde kullanılmaktadır (Balaban, 2016).

Birliktelik kuralı yönteminde çok büyük veri setlerinin kullanılması, örüntünün çıkartılmasında bilgisayar kaynakları ve hesaplamaların uzun sürmesine neden olduğundan yöntemi maliyetli hale getirmektedir. Pazarlama alanında bu modelin yaygın kullanılması bu alanda getirisinin yüksek düzeyde önemli olmasından kaynaklanmaktadır (Özcan, 2014; Balaban, 2016).

Sıra Örüntüleri: Birliktelik kuralına benzer bir yapıda olan bu yöntemde veriler arasındaki ilişki zamana bağlıdır. Belirli bir zaman aralığında olayların ilişkisini ele almaktadır. Bilgisayar ağı, Telekomünikasyon ağları ve bilimsel deneylerden toplanan veriler yapısı gereği aralarında bir ilişki bulundurur. Bu türden ilişkilerin tespitinde tanımlayıcı modellerden sıra örüntüleri yöntemi kullanılır (Altun, 2019). Bir mobilya mağazasının veri tabanında kanepe takımı alındıktan sonra takip eden diğer alışveriş de masa takımı alınması, meteoroloji veri tabanında sağanak yağış göstergesinden sonra sel afeti yaşanması sıra örüntülerine örnek olarak verilebilir (Argüden ve Erşahin, 2008).

2.4.2.2. Tahmin Edici Modeller

Tahmin edici modeller sonuçları belli olan veriler kullanılarak geleceğe yönelik kestirimde bulunmak için kullanılmaktadır. Tahmin modelleri, hedeflenen amaçlara yönelik veri setlerinde sonuçları bulunmayan ya da yeni girilen verilerin sonuçlarının tahmin edilmesinde kullanılmaktadır (Altun, 2019). Tahmin edici model ile mevcut sistemlerin iyileştirilmesine ve geleceğe yönelik planlamalar yapılabilmektedir (Zaimoğlu, 2018). Çalışmamızda yaptığımız üzere önceki eğitim dönemlerine ait veriler kullanılarak gelecek dönemler hakkında başarı durumu kestirimi yapılabilir. Böylece

gerekli önlemler alınıp başarı ve eğitimin kalitesi arttırılabilir.

Sınıflama: Belirli özellik gösteren verilerin gruplandırılıp kategorize edilmesi ile tahmin edilmek istenen değişkene yönelik sonuçların kestirimidir. Sınıflandırma modeli kategorize edilmiş bağımsız değişken ya da değişkenler ile bağımlı değişkenin tahmin edilme sürecidir (Zaki ve Wagner, 2014). Sınıflandırma algoritmalarında kategorize edilmiş verinin karakteristik özelliklerine bakılarak sınıflar tanımlanmaktadır. Veri setinde bulunan verinin örnek kısmı ile sınıf etiketlerini içeren tanımları kullanılarak oluşturulan öğrenme modeli sınıflandırma tahmini için kullanılmaktadır (Altun,2019). Sınıflandırma yöntemi için, genç kadınlar küçük araba satın alırken, yaşlı erkekler büyük araba satın alır, örneği verilebilir (Argüden ve Erşahin, 2008).

Zaman Serisi Analizi: Veri serinde bir özelliğin zamana bağlı olarak aldığı değerlerin incelenmesidir. Değerlerin ölçümü saatli, günlük, haftalık ve yıllık olmak üzere eşit zaman dilimlerinde yapılır. Zaman serisi analizi farklı işlevlerde kullanılabilir. Bunlar; farklı zaman serileri arasındaki benzerliğin bulunması, zaman serisinde çizelgenin davranışına karar vermek ve tarihsel zaman çizelgesi oluşturularak gelecek zaman değerlerinin tahmininde kullanılabilir. Tahmine dayalı zaman serisi modellerinde geçmiş verilerin kullanılması nedeniyle bu modele denetimli öğrenme modeli de denilmektedir (Aydın, 2007).

Kestirim: Veri setinde kayıtlı verilere bakılarak geleceğe yönelik olayların belli bir olasılık ile tahmin edilmesi sürecidir. Kestirim yöntemi sınıflandırmadan farklı olarak, verinin şu anki durumu yerine gelecekteki durumunu tahmin etmekte kullanılır. Kestirim yöntemi ile konuşma tanıma, makine öğrenmesi, desen tanıma teknikleri gibi uygulamalar yapılmaktadır. Geleceğe yönelik değer tahminlerinde zaman serisi ve regresyon analizinin yanı sıra kestirim yöntemi de kullanılmaktadır (Altun, 2019).

Regresyon: İki değişken arasında ilişkiyi belirleme ve bir değişken kullanılarak diğer değişken hakkında tahmin yapılmasını sağlayan istatistiksel bir hesaplama yöntemidir. Regresyon modelinde değişkenler arasındaki ilişkinin fonksiyonel şeklini, bağımlı ve bağımsız değişken olarak bir doğru denklemi üzerinde gösterir. Değişkenlerden birinin bilinmesi ile diğer değişken hakkında kestirimde bulunulmasını sağlar (Gamgam ve Altunkaynak, 2015).

Regresyon modeli oluşturulurken veri setinin bir bölümü ile doğrusal (lineer) ve lojistik fonksiyonlar oluşturulur. Verinin kalan bölümü ile de hedeflenen özellikleri en iyi modelleyen fonksiyon belirlenmeye çalışılır. Belirleme aşamasında gerçek veri ile tahminler arasındaki farkın bulunması, en az hata payına sahip fonksiyonun seçilmesinde karar vermeyi sağlamaktadır.

Çoklu doğrusal regresyon: Hedef (bağımlı) değişkenin, birden fazla bağımsız değişken arasındaki doğrusal ilişkiyi belirleyen ve bağımsız değişkenlerin bağımlı değişken üzerindeki etkisini ölçen bir analiz yöntemidir. Bağımsız değişkenler regresyon modelinde birlikte ele alınarak bağımlı değişken üzerindeki etkisine bakılır ve formülü şu şekildedir;

$$Y=\alpha + \beta_1 X_1 + \beta_1 X_1 + \beta_2 X_2 + \beta i Xi + \varepsilon$$

- **Y**; bağımlı değişken
- **X**; bağımsız değişken
- **α**; sabit olup X=0 olduğunda Y'nin aldığı değerdir
- **β**; regresyon katsayısı
- **ε**; hata terimi

Regresyon modelinde elde edilen değer ile gerçek değer arasındaki fark hata terimidir. Bu değerin küçük olması modelin tahmin değerinin yüksek olmasını sağlar.

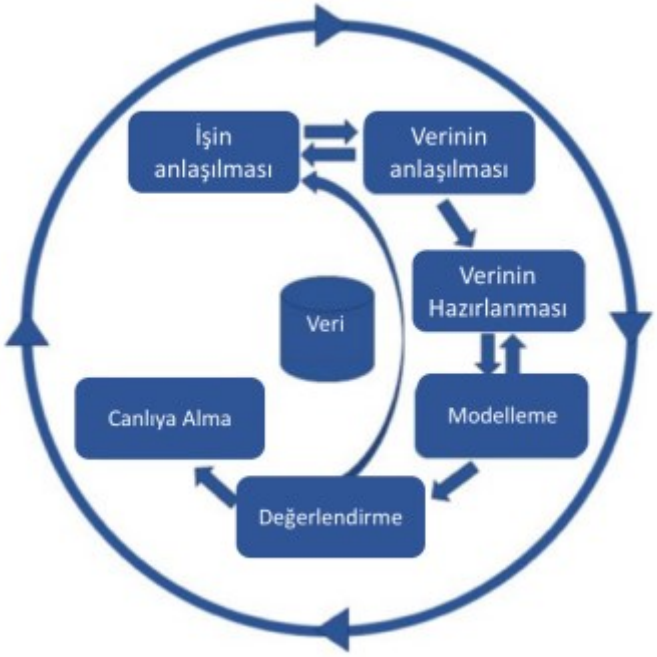
Lojistik regresyon: İki ya da daha fazla bağımsız değişken ile hedef değişken arasında kurulan ilişkiyi sınıflandırma yaparak belirleyen bir modeldir. Doğrusal regresyondan farklı olarak lojistik regresyonda hedef (bağımlı) değişken kesikli değer almaz kategoriktir (Gamgam ve Altunkaynak, 2015).

Bu çalışmada, öğrencilerin mezuniyet süresi ve başarı durumları kestirilmeye çalışıldığından tahmine dayalı modeller kullanılmıştır. Mezuniyet zamanı kestirimi için lojistik regresyon modeli, bilgisayar dersi başarı durumları kestirimi için çoklu doğrusal regresyon modeli kullanılmıştır. Eğitsel veri madenciliği uygulaması CRISP-DM iş süreci adımlarına uygun olarak gerçekleştirilmiştir. CRISP-DM iş süreci aşağıda tanımlanmış olup yöntem ve bulgular kısmında çalışma adımlarına uygun olarak detaylı açıklanmıştır.

2.5. CRISP-DM İş Akış Süreci

Türkçe karşılığı Çapraz Endüstri Veri Madenciliği Standart Süreci olan CRISP-DM bilgi oluşumu için veri madenciliğinin temel adımlarını belirten bir süreçtir (Yurdakul, 2015). CRISP-DM süreci Şekil 2.5'te görüldüğü gibi 6 aşamadan oluşmaktadır. Bunlar iş (hedef) sürecini anlama, veriyi anlama, veriyi hazırlama, modelleme, değerlendirme ve yayılım şeklindedir (SVE, 2019).

CRISP-DM süreci Şekil 2.5'te de görüldüğü gibi döngüler içermektedir. Bu döngüler ile problem durumuna göre verilerin elde edilmesi ve veri seçiminde yaşanan zorluklara veya fırsatlara göre iş analizinin yeniden gözden geçirilmesidir. Model ile verinin hazırlanması aşamasında geri dönüş yapılabilir. Bu bazı modellerin eksik veri ile çalışmamasından kaynaklanmaktadır. Bu sayede verinin hazırlanması aşamasına geri dönülüp gerekli veri ön hazırlığı yapılabilir. Sürecin değerlendirme aşamasında bulunan döngü en önemli olan kısımdır. Bunun sebebi oluşturulan model sonuçlarına göre tekrar başa dönülerek bütün adımların kontrol edilebilmesidir (Şeker, 2018).



Şekil 2.5. CRISP-DM Süreci (Şeker, 2018)**.**

Çalışmada kullanılan CRISP-DM süreci aşağıda belirtildiği gibi 6 adımdan oluşmaktadır. Çalışmaya uygun olarak adımların uygulanması yöntem ve bulgular kısmındaki başlıklarda detaylı açıklanmaktadır.

1. **İş/Problem Anlama:** Yapılacak çalışmanın amaçlarının belirlendiği bu adım veri madenciliği hedeflerinin belirlendiği ve çalışmanın planının geliştirilme kısmını kapsamaktadır. Problemin sürecin başında ortaya konulması veri madenciliği yapılacak çalışmaya bir yol çizmesini sağlamaktadır (Balaban ve Kartal, 2015).

2. **Veriyi Anlamak:** Bu adımda hedefler doğrultusunda işlenecek verinin bir nitelik kazanmasını sağlamak olarak belirtilebilir. İşlenecek verinin toplanması sürecini kapsamaktadır.

3. **Verinin Hazırlığı:** Toplanan veri setini hedeflere ulaşacak düzeyde işlenmesi aşamasıdır. Bu bölümde veri setinden oluşturulmak istenen model ya da modeller için gerekli olan veriler seçilip bu verilerin analiz edilebilecek düzeye getirilme işlemi yer almaktadır.

4. **Modelleme:** Bu aşama çalışmanın amacına ve problemine uygun olarak geliştirilmek istenen modelin seçimi, uygulanması ve optimal değerlere göre ayarlanmasını içermektedir. Çalışmada birden fazla model kullanılabilir ancak bazı modeller farklı problem durumu için yapılacaksa veri hazırlama sürecine tekrar dönülebilir (Özdemir, 2016).

5. **Değerlendirme:** Oluşturulan modelin sürecin ilk aşamasındaki hedeflere uygunluğunun test edildiği aşamadır. Bu aşama sonuçları değerlendirme, süreci değerlendirme ve sonraki adımları planlama aşamalarını barındırmaktadır.

6. **Uygulama:** Oluşturulan veri madenciliği modelleri ile başarı sağlanan uygulamanın başka uygulamalar için aktarıldığı bölümdür. Bu aşama geliştirilen uygulamanın yayılma planının oluşturulup, takip ve bakımının planlandığı, raporların hazırlandığı ve projenin değerlendirildiği kısımdır.

Bu çalışmada kullanılan CRISP- DM iş süreci yöntem bölümünde detaylı bir şekilde anlatılmış probleme özgü bilgiler verilmiştir.

2.6. Öğrenme Analitiği ve Eğitsel Veri Madenciliği

Teknolojinin hızlı gelişimi eğitim ve öğretim alanlarının değişmesine neden olmuştur. Eğitim alanında teknolojik gelişmelerden biride öğretim teknolojileridir. Öğretim teknolojileri "öğrenmeyi kolaylaştırmak ve performans artışı sağlamak amacıyla uygun teknolojik süreç ve kaynakların oluşturulması, kullanılması ve değerlendirilmesinin etik uygulamasıdır." şeklinde Januszewski ve Molenda (2013) tarafından tanımlanmıştır. Bu bağlamda bakıldığında öğrenme analitiklerinin ve eğitsel veri madenciliği ile eğitim ve öğretim kolaylaştırılarak performansı arttırılıp ve öğretim teknolojilerinin amaçları sağlanabilir.

Dijital gelişim ile birlikte eğitim ortamları e-öğrenme, öğrenme yönetim sistemleri ve çevrimi içi öğrenme alanlarına talep artmıştır. Bu denli teknoloji destekli öğrenmenin hayatımıza girmesi "öğrenme analitikleri" adlı bir alanın oluşmasına neden olmuştur (Tuzcu, 2018). Öğrenme analitiği kavramı ilk kez Siemens (2010) tarafından öğrenmeye yönelik kestirim yapmak ve öneriler sunabilmek için öğrencinin ürettiği veri, bilgi ve sosyal bağlantıların kullanılması olarak tanımlanmıştır. 2011 yılında 1. Uluslararası Öğrenme Analitikleri ve Bilgi Konferansında ise öğrenme ortamlarını optimize etmek amacıyla öğrenenler hakkında verilerin toplanması, ölçümü, analizi ve raporlanması olarak tanımlanmıştır (Bahçeci, 2015). Öğrenme analitiği öğrenenin ilerleme ve performansını tahmin etmek için veriler kullanılarak geliştirilen modellerin kullanımı ve bilgiye dayanarak hareket etme yöntemidir (Winne, 2017). Greller ve Drachsler (2012) öğrenme analitiğini Şekil 2.6'da görüldüğü gibi 6 adet bileşen ile açıklamışlardır.

Şekil 2.6. Öğrenme Analitiği Bileşenleri.

Öğrenme analitiği bileşenleri; amaçlar (tahmin, yansıtma), veri (korumalı, açık), paydaşlar (öğretmen, öğrenci, veli ve kurum), yeterlilikler (eleştirel düşünme, yorumlama) ve kısıtlar (gizlilik, etik) şeklindedir (Greller ve Drachsler, 2012). Öğrenme analitiklerinde teknoloji, algoritmalar ve eğitsel teoriler gibi çeşitli araçlar yer almaktadır. Bu türden araçların eğitim alanında kullanılması veri tabanlarında farklı türde, boyutta birçok verinin depolanması sağlanmakta ve bunlar büyük veriyi oluşturmaktadır. Eğitim alanında oluşan büyük veriyi anlamlı verilere dönüştürmek için araştırmacılar veri madenciliği yöntemini kullanmaktadırlar (Özbay, 2015b).

Eğitim ortamında oluşan büyük veri yığınları kullanılarak eğitim ve öğretimi etkili şekilde gerçekleştirmek için bu yığınlardaki verilerden anlamlı veriler oluşturulması işlemi ile eğitsel veri madenciliği ortaya çıkmıştır. Eğitsel veri madenciliği eğitim, istatistik ve bilgisayar bilimi ile ilişkili olan disiplinler arası bir alan olarak tanımlanmıştır (Romero ve Ventura, 2013). Eğitsel veri madenciliği yöntemi ile mevcut veriler ile şu anki durumun analizi ve geleceğe yönelik çıkarımlarda bulunmak mümkün olabilmektedir (Tekin ve Şengür, 2013). Eğitim alanına yönelik olarak öğrenci akademik performansını kestirmek için modeller oluşturulabilir ve bu modeller ile öğrenciye rehberlik edilebilir (Bienkowski, Feng ve Means, 2012).

EVM ve öğrenme analitiği çalışmaları yeni yaklaşımlar çerçevesinde öğrenci verilerini analiz ederek öğrenciye yönelik öğrenme stili, davranış ve akademik başarının modellenebilmesini sağlamaktadır. Buna ek olarak öğrencinin öğrenme profilleri belirlenip sınıflandırılarak öğrencilerin öğrenme stillerine uygun öğretim programları (grup ile öğrenim, bireysel öğrenim) ve öğrenme ortamları (sınıf içi, uzaktan eğitim, iş başında eğitim) düzenlenebilir (Bienkowski, Feng ve Means, 2012).

Geleneksel eğitim anlayışından çağdaş eğitime geçildiği süreç zarfında eğitim alanında çeşitli öğrenme ortamları test edilmiş ancak sadece sınıf ortamı kullanılmak istenmiştir. Çağdaş eğitim öğrenmenin yeri ve zamanının olmadığı bir eğitim sistemi getirmek istese de bu sistem tam olarak kullanılamamıştır. EVM çalışmaları geleneksel sınıf ortamı ve uzaktan eğitim olmak üzere iki eğitim sistemi üzerine yoğunlaşmıştır. Öğrenci çalışmalarının izlenmesinin zor olduğu geleneksel eğitim de EVM çalışmalarının kısıtlı olduğu, öğrenci izleniminin kolay yapılabildiği çeşitli imkanların ve verilerin depolanabildiği uzaktan eğitim sistemlerinde ise daha geniş bir alanda çalışmalar yapıldığı görülmüştür (Tuzcu, 2018).

Geleneksel eğitimde ve çağdaş eğitim sistemlerinde yapılan EVM çalışmaları, eğitim alanında veri madenciliğini inceleme, tanıtma, öğrenci akademik başarısını ölçme, akademik başarıya etki eden faktörleri belirleme ve öğrenci özelliklerini belirleme olarak görülmüştür (Zaiane, 2001; Özbay, 2015a). Öğrenci verilerinin analiz edilmesi ile başarı, başarısızlık nedenleri, başarının artırımı için neler yapılması gerektiği, üniversiteye yerleşme puanları ile ders başarısı arasındaki ilişkinin sorgulanması ve buna yönelik cevapların bulunması eğitimin kalitesini arttırabileceği düşünülmektedir (Akgöbek ve Çakır, 2009).

Eğitimin kalitesinin arttırılmasına yönelik öğrenci performansının yanı sıra öğreticinin nitelikleri ve yönetimin karar verme sürecindeki bilgisinin de geliştirilmesi gerekmektedir. Bunun sağlana bilmesi için bilgiye ulaşılması ve gerekli bilgilerin yönetim sistemlerinden elde edilmesi ile karar verme süreci geliştirilip eğitimin kalitesi arttırılabilir (Altun, 2019). Eğitimde büyük veri, veri madenciliği ve öğrenme analitiği yöntemleri ile eğitim yönetiminde veriye dayalı karar verme süreçlerinin kullanılabileceği MEB'in 2023 Eğitim Vizyon Belgesinde belirtilmiştir.

MEB (2018) yayınladığı bildiri de veriye dayalı yönetim anlayışı kazandırmayı hedefleyerek öğretmen, okul yöneticisi ve eğitim yöneticilerinin üzerindeki bürokratik iş yükünü azaltılacağını belirtmiştir. Bunun yanında veriye dayalı yönetim ile öğrenim ve öğretimin daha anlaşılır olacağını ve performansa dayalı bir eğitim öğretim sürecinin hayata geçireceğini hedeflemiştir. Bildiride öğrenci verilerinin analiz edilmesi ile her alanda desteğe ihtiyaç duyan öğrenciler belirlenerek gelişim planlarında gerekli eylemlere yer verilebileceği belirtilmiştir.

Verinin eğitim alanında veri madenciliği yöntemleri ile kullanılması karar vermeye yönelik strateji oluşturma gibi kritik konularda yöneticilere destek sağlaya bileceği belirlenmiştir. EVM ile bu konulara ek olarak akademik performansın arttırılması ve eğitimin sürekliliğinin arttırılması gibi konularda eğitim alanına önemli katkılar sağlanabilir. Eğitim ortamlarında oluşan veri yığınları içerisinde saklı bilgi ve örüntülerin veri madenciliği ile açığa çıkartılması, eğitimin verimliliğini ve kalitesini artıracağı düşünülmektedir.

2.7. İlgili Araştırmalar

Bu bölümde veri madenciliğinin eğitim alanında kullanılmasına yönelik olarak Türkiye'de ve Dünya'da yapılan araştırmalar incelenmiş olup sınıflandırmaya dayalı eğitsel veri madenciliği ve tahmine dayalı eğitsel veri madenciliği çalışmaları olarak iki ayrı başlık altında sırasıyla sunulmuştur.

2.7.1. Sınıflandırmaya Yönelik Eğitsel Veri Madenciliği Çalışmaları

Veri madenciliğinin eğitim alanında kullanımına yönelik olarak yurtiçi ve yurtdışında yapılan çalışmalar incelenmiş ve eğitsel veri madenciliği sınıflandırma çalışmalarında çoğunlukla öğrenci performansını arttırmaya yönelik çalışmaların yapıldığı tespit edilmiştir. Sınıflandırma modelleri ile, öğrencilerin başarılarına göre kümelenerek mevcut durumun analiz edildiği ve öğrenci başarısını etkileyen faktörlerin belirlenmesinin amaçlandığı görülmüştür.

EVM ile ilgili 2016, 2018 ve 2020 tarihli alanyazın tarama çalışmaları ve dergiler incelendiğinde 2003-2021 yılları arasında toplam 803 EVM çalışması yapıldığı gözlenmiştir. Yapılan çalışmalarda en çok akademik başarıya yönelik ve öğrenci mezuniyetine yönelik çalışmaların olduğu rapor edilmiştir (Akgün, 2020; Tekin ve Öztekin, 2018; Türel ve Baz, 2016). Ancak tez konusu ile alakalı olup önemli görülerek incelenen çalışmalar aşağıda sırasıyla verilmiştir.

Polat (2021) yaptığı çalışmada, MEB'e bağlı uzaktan eğitim veren açık öğretim liselerinde eğitim gören 2.317.130 öğrencinin verisini kullanarak öğrencilerin mezun olma, okul terk gibi durumlarını sınıflandırma yöntemleri ile tahmin etmiştir. Çalışmada J48, Decision Tree, k-NN, Naive Bayes ve Random Forest algoritmaları kullanılmış. Çalışma sonucunda J48 algoritmasıyla geliştirilen modelin %80,47 oranıyla en başarılı model olduğu görülmüştür. Bu modele göre öğrencilerin durumlarını tahmin etmede en önemli özelliğin toplam kredi sayısı olduğu tespit edilmiştir.

Uğuz, Şahin ve Yılmaz (2021) yaptıkları çalışmada, 2018 yılında yapılan PISA sınavına katılan 6890 öğrencinin Fen Bilimleri puanını tahmin etmek istemişlerdir. Çalışmada k-NN, Naive Bayes ve Random Forest yöntemleri kullanılarak aile eğitim durumu, ders çalışma süresi, bilgi ve iletişim teknolojileri kullanımı ve öğrenci algıları değişkenlerinin Fen Bilimleri puanını nasıl etkilediğini tespit etmişlerdir. Araştırma

sonucunda bağımlı değişkeni etkileyen bağımsız değişkenlerden aile eğitimi değişkeninin modele bir etkisinin olmadığı, k-NN modelinin performans değerinin %77 olduğu, Naive Bayes performans değerinin %55,06 olduğu ve Random Forest modelinin performans değerinin %62,22 olduğu gözlenmiştir.

Olgun (2021) yaptığı çalışmada, ters yüz eğitiminde öğrenci başarısını veri madenciliği sınıflandırma yöntemleri ile tahmin etmiştir. Çalışmada farklı bölümlerde öğrenim gören ve temel bilişim dersini alan 404 öğrencinin video izleme verilerini kullanarak Random Forest, Naive Bayes ve Destek Vektör makine algoritmalarını geliştirmiş. Çalışma sonucunda en yüksek düzeyde performans gösteren algoritmanın %83,11 ile Random Forest olduğu belirlenmiştir.

Topuz (2021) çalışmasında veri madenciliği analiz yazılımı WEKA ve Orange programlarının başarı değerlerini MEB tarafından yapılan ABIDE sınav sonuçlarını kullanarak karşılaştırmış. Çalışma sonucunda sınıflandırma modelleri k-NN ve Yapay Sinir Ağları algoritmalarında Orange'ın, Destek Vektör ve Naive Bayes algoritmalarında ise WEKA yazılımının daha yüksek doğru sınıflandırma yaptığı belirlenmiştir.

Kayhan (2019) yaptığı çalışma ile, EVM kullanılarak ön lisans öğrenimi gören öğrencilerin mezun olma zamanlarını tahmin edilmesini ve geç mezun olacak kişilerin tespit edilip gerekli önlemlerin alınmasını sağlamak istemiştir. Çalışmada farklı fakülte ve bölümlerde öğrenim gören ön lisans öğrenci verileri ile Karar ağacı, Naive Bayes, Random Forest ve Yapay Sinir Ağı modelleri oluşturmuş. Çalışma sonucunda oluşturulan modellerden en yüksek düzeyde performans gösteren Karar Ağaçları ve Naive Bayes yöntemi olduğu görülmüştür.

Chung ve Lee (2019) Kore Ulusal Eğitim Sisteminde verisi bulunan 160,715 lise öğrenci verileri ile yaptıkları çalışmada öğrencilerin okulu terk etme durumları tahmin edilmiştir. Tahmin modeli için öğrencilerin ilk aydaki derse geç kalma sayıları, mazeretli ve mazeretsiz devamsızlık günleri, kulüp aktiviteleri, gönüllü çalışma süreleri gibi devamsızlık ile ilgili veriler kullanılmıştır. Random Forest modeli ile öğrencilerin okulu terk etme tahmin sonucunu %95 başarı oranı ile doğru tahmin etmişlerdir. Tahmin modelinde en etkili değişkenin mazeretsiz devamsızlık olduğu tespit edilmiştir.

Aksu (2018) yaptığı çalışmada, 2015 yılında yapılan PISA sınavında öğrencilerin Fen okur-yazarlığı bakımından başarılı ve başarısız olarak sınıflandırılmasını sınıflandırma modelleri ile tahmin etmiş. Çalışma sonucunda en yüksek düzeyde performans gösteren ve hata düzeyleri düşük olan modelin Random Forest olduğunu belirtmiştir.

Tuzcu (2018) yaptığı çalışmada, ders yönetim sistemindeki öğrenci verilerini kullanarak Naive Bayes, Lineer ve Lojistik Regresyon, Derin Öğrenme, Karar Ağacı ve Random Forest modelleri ile derslerin başarı tahmini yapmış. Çalışma sonucunda her bir modelin farklı performans gösterdiği gözlenmiştir.

Iam-On ve Boongen (2017) yaptıkları çalışmada üniversiteyi bırakma eğiliminde olan öğrencileri kümeleme analizi ile belirlemeye çalışmışlar. Öğrencilerin demografik verileri, üniversite öncesi akademik verileri ve üniversitenin ilk yılındaki akademik verilerini kullanılarak k-Means algoritması ile iki küme elde etmişleridir. Kümeler üniversiteye kayıt öncesi iyi nota sahip öğrenciler ile orta ve düşük nota sahip öğrenciler şekline ayrılmış. Çalışma sonucunda iyi akademik geçmişe sahip öğrencilerin öğrenimlerine devam ettikleri, orta ve düşük akademik geçmişe sahip öğrencilerin ise eğitimine devam etmedikleri görülmüştür.

Can (2017) yaptığı çalışmada, üniversite öğrencilerine uygulanan dönem sonu ders değerlendirme anketi verileri kullanılarak, öğrencilerin akademik başarıları ile sorulara verdikleri cevaplar arasındaki ilişki incelemiş. Lojistik regresyon ve Karar Ağacı modellerinin kullanıldığı çalışmanın sonucunda öğrencilerin akademik başarılarında "Sınav, Proje ve Quizler öğrenmeye yardımcı oldu." Başlıklı sorunun belirleyici olduğu gözlenmiştir.

Sara ve diğerleri (2015) tarafından yapılan çalışmada Danimarka'da Lise öğrencilerinin okulu terk etme durumlarını tahmin edilmiş. Ders yönetim sisteminde kayıtlı 72.598 öğrencinin 17 farklı verisi ile Destek Vektör Makineleri, Gaussian Ernels, Random Forest, Karar Ağaçları ve Naive Bayes modelleri geliştirilmiş ve en yüksek performans gösteren model belirlenmek istenmiş. Çalışma sonucunda Random Forest modelinin %93,47 oranı ile en yüksek performansı gösterdiği belirlenmiştir.

Pehlivanoğlu ve Duru (2015) yaptıkları çalışmada, ortaokul öğrencilerinin cinsiyet, uyku durumu ve başarı durumu gibi özellikleri ile sosyal ağlar üzerindeki günlük etkinlikleri arasındaki ilişki incelenmiştir. Çalışma sonucunda kız ve erkek öğrencilerin en fazla Facebook ağını kullandıklarını ve bu ağı genelde tabletler üzerinden oyun amaçlı olarak kullandıkları belirlenmiştir.

Kılınç (2015) çalışmasında, öğrencilerin demografik durumlarının ve maddi olanaklarının öğrencilikten çıkartılma durumuna etkisini incelemiş ve bu kapsamda Eskişehir Osmangazi Üniversitesinde 2008-2011 yılları arasında birinci sınıfta okuyan Bilgisayar Mühendisliği Bölümündeki öğrencilerin verilerini kullanılarak sınıflandırma ve birliktelik modelleri geliştirmiştir. Geliştirilen modellerin sonuçları incelendiğinde, öğrencilikten çıkarılma ile öğrencilerin not durumları arasındaki ilişki bulunmuş, öğrencilerin eğitim sürelerinin, burs veya kredi alınmasıyla değişiklik gösterdiği ve öğrencilerin parasal durumlarıyla annelerinin meslekleri arasında bir bağlantı olduğu belirlenmiştir.

Yükseltürk, Özekeş ve Türel (2014) yaptıkları çalışmada çevrimiçi eğitime katılan 189 öğrencinin eğitimi bırakma durumları tahmin edilmiş. Anket ve ölçekler ile elde edilen cinsiyet, yaş, öğrenim durumu, önceki çevrimiçi deneyim, meslek, öz yeterlilik, hazırbulunuşluğu ve bırakma durumu verileri k-NN, Decision Tree, Naive Bayes ve Yapay Sinir Ağları yöntemlerinde kullanmışlar. Oluşturulan modellerin %79'un üzerinde performans gösterdiği, en yüksek performans değerini ise %87 doğruluk oranı ile k-NN modelinin gösterdiği belirlenmiştir. Çalışma sonucunda çevrimiçi eğitimlerde kişinin öz yeterliliği, hazırbulunuşluğu ve önceki çevrimiçi deneyimlerinin öğrenimi bırakmada etkili olduğu belirtilmiştir.

Özetlemek gerekirse EVM sınıflandırma uygulamalarında genel olarak öğrencinin akademik başarı durumu, mezun olma ve okulu bırakma gibi durumlarının farklı sınıflandırma algoritmaları ile tespiti ve tahminine yönelik çalışmaların yapıldığı gözlenmiştir. Sınıflandırma çalışmalarında başarı oranının yaklaşık olarak %70-80 aralığında sonuçlar göstermesinin başarılı tahmin modeli sonucu kabul edildiği tespit edilmiştir. Yapılan çalışmalarda paket programların kullanıldığı ve çalışmada kullanılan verilerin ise genel olarak öğrenci bilgi sistemlerinden elde edildiği belirlenmiştir.

2.7.2. Tahmine Yönelik Eğitsel Veri Madenciliği Çalışmaları

Veri madenciliğinin eğitim alanında kullanımına yönelik olarak yurtiçi ve yurtdışında yapılan çalışmalar incelenmiş, eğitsel veri madenciliği kestirim (tahmin) çalışmalarında öğrenci performansının tahmin edilerek akademik başarıyı arttırmaya yönelik çalışmaların yapıldığı tespit edilmiştir. Kestirim modelleri ile, öğrencilerin geçmiş akademik verileri kullanılarak geleceğe yönelik olarak başarının tahmin edilmesi ve öğrencilerin öğrenimlerini devam ettirmesine yönelik olarak gerekli önlemlerin alınmasının hedeflendiği görülmüştür.

EVM ile ilgili Aruğaslan ve Çivril'in (2021) alanyazın tarama çalışmasında EVM ile ilgili çalışmaların genelde yüksek lisans ve doktora öğrencilerinin yaptığı gözlenmiştir. Yapılan çalışmalarda %49,4 ile en çok öğrencinin akademik başarısının tahminine yönelik çalışmaların olduğu rapor edilmiştir. Ancak tez konusu ile alakalı olup önemli görülerek incelenen çalışmalar aşağıda sırasıyla verilmiştir.

Koç ve Akın (2022) yaptıkları çalışmada, 2019 yılında Türkiye'de lise giriş sınavına giren öğrenci verilerini kullanarak Regresyon ve Random Forest modelleri ile bir makine öğrenmesi geliştirmişlerdir. Çalışmada lise giriş sınavına; boşanma oranı, gayri safi yurtiçi hasıla, okur-yazar oranı yüksek öğrenim nüfusu gibi değişkenler açısından etkisi incelenmiş. Çalışma sonucunda oluşturulan iki modelin ayrı olarak değerlendirilmesinin tahmin etmede yetersiz olduğu bu yüzden bu iki modelin beraber kullanılması gerektiği belirlenmiştir.

Keser (2021) yaptığı çalışmada, orta okul öğrencilerinin akademik performanslarını tahmin etmek için bir yapay sinir ağı modeli geliştirmiş. Çalışma sonucu incelendiğinde öğrencilerin Matematik dersinin tahmin sonuçlarının %97 doğruluk ile, Portekizce dersinin tahmin sonuçlarının ise %97,6 doğruluk oranı ile başarı gösterdiği belirlenmiştir.

Karataşcı (2021) çalışmasında, ortaokulda öğrenim gören 5, 6, 7 ve 8. sınıf öğrencilerinin Matematik ve Fen Bilimleri derslerinin hedef kazanımlarını kazanıp kazanılmadığını ve iki ders arasındaki ilişkiyi incelemiş. Çalışma sonucunda 5,6,7 ve 8. Sınıf öğrencilerinin ders kazanımları incelendiğinde; Matematik dersi için sırasıyla %58,09 %51,92 %58,16 ve %50,57, Fen Bilimleri dersi için ise sırasıyla %54,84 %53,17 %46,1 %57,71 oranında hedef kazanıma ulaşıldığı belirlenmiştir. Derslerin hedef

kazanımlarının bu denli düşük çıkmasının Covid-19 salgını nedeni ile eğitime uzaktan devam edilmesi olduğu belirtilmiştir.

Yu ve diğerleri (2020) yaptıkları çalışmada öğrenci başarısını kısa ve uzun süreli olarak tahmin etmede öğrenim yönetim sistemi ve anket verilerinin (Öz yeterlilik, zaman yönetimi vb.) kullanılmasının etkilerini incelemişlerdir. 2000 öğrenci verisi kullanılarak yapılan çalışmada kısa süreli tahmin için ders başarı notları, uzun süreli tahmin için ise yıllık başarı not ortalaması kullanmışlardır. Çalışmanın sonucunda öğrenci yönetim sistemi verilerinin yüksek tahmin etme gücüne sahip olduğu, anket verilerinin ise çok düşük tahmin gücüne sahip olduğu sonucuna ulaşmışlardır. Öğrenci başarısını tahmin etmeye yönelik olarak öğrenci yönetim sistemlerinin kullanılmasının yüksek doğruluk sağlayacağı belirtilmiştir.

Altun, Kayıkçı ve Irmak (2019) yaptıkları çalışmada Akdeniz Üniversitesi Sınıf Öğretmenliği Bölümünden 2012-2017 yılları arasında mezun olmuş 578 öğrencinin çeşitli verilerini kullanarak mezuniyet notlarını tahmin etmişleridir. Çalışmada Yapay Sinir Ağı ve Regresyon analizi modelleri oluşturmuşlardır. Çalışma sonunda Regresyon analizi performans değerini %94,30, Yapay Sinir Ağı modelinin performans değerini ise %94,43 olduğu gözlenmiş ve bu modellerin kullanılması ile eğitimde toplam kalitenin ve öğrenci başarısının arttırılabileceği sonucuna ulaşmışlardır.

Berens ve diğerleri (2019) yaptıkları çalışma ile Almanya'da üniversitede öğrenim gören öğrencilerin idari bilgilerini kullanarak okulu bırakma durumunda olanları tahmin etmek istenmiş ve bırakma risk bulunan öğrencilerin erken tespiti ile gerekli önlemlerin alınmasını hedeflemişlerdir. Tahmin modelleri için öğrencilerin kişisel verileri (yaş, cinsiyet, doğduğu ülke ve bölge, göçmenlik bilgisi, sağlık sigortası), önceki eğitim bilgileri (üniversiteye giriş derecesi, giriş puanı, önceden okuduğu üniversite var ise dönem sayısı) ve akademik bilgileri (başarılı-başarısız sınav sayısı, derslerin ortalama geçme puanı, mezun olma ve öğrenimi bırakma durumu) kullanılmıştır. Çalışmada Regresyon analizi, Yapay Sinir Ağları, Karar Ağaçları ve AdaBoost yöntemleri kullanılmıştır. Çalışma sonucunda birinci dönem sonu verileri ile %79, dördüncü dönem sonu verileri ile %90 oranında başarılı tahminler gerçekleştirildiği görülmüştür.

Altun (2019) yaptığı çalışma ile, Akdeniz Üniversitesi Eğitim Fakültesi öğrenci verilerini kullanarak geliştirmiş olduğu veri madenciliği algoritmaları ile öğrenci mezuniyet notunun kestirimi ve enstitü yöneticilerine karar vermede rehber olacak bir model oluşturmuştur. Çalışma sonucunda 1. Dönem verileri kullanılarak yapılan modelin performans değeri %94-97 aralığında olduğu, alt modellerin lojistik regresyon ve karar ağacı modellerinin performans değerlerinin ise %72-87 aralığında performans gösterdiği gözlenmiştir.

Aksu (2018) çalışmasında, Türkiye'de öğrenim gören 5895 öğrencinin PISA fen okuryazarlığı başarısını tahmin etmek için Decision Sump, Hoeffding Tree, J.48, Lojistik regresyon, RepTree, Random Forest ve Rastgele Ağaç yöntemlerini kullanmıştır. Modellerde 12 farklı değişken kullanılmış ve en yüksek düzeyde performans gösteren model belirlenmeye çalışılmıştır. Uygulanan modellerden en yüksek düzeyde performans gösteren %72,35 ile lojistik regresyon yöntemi olduğu belirlenmiş ve başarının kestirimine yönelik olarak lojistik regresyon yönteminin kullanılabileceği belirtilmiştir.

Alsuwaiket (2018) yaptığı çalışma ile geçmişte yapılan anketler değerlendirilerek öğrencilerin matematik başarısını kestirecek model geliştirmiştir. Modelde beş farklı ülkedeki 230.000 öğrenci verilerinden; dil, okuma, beslenme, eğitimci deneyimleri ve öğrencinin matematiksel yeteneği gibi önemli değişkenler Yapay Sinir Ağı, Random Forest ve k-NN algoritmaları kullanılarak öğrenci başarısı tahmin edilmiştir. Çalışma sonucunda, Random Forest modelinin en yüksek doğrulukta olduğu ve en düşük RMSE değerine sahip olduğu gözlenmiştir.

Asif ve diğerlerinin (2017) yaptığı çalışmada, veri madenciliği yöntemleri ile lisans düzeyindeki öğrencilerin performansları incelenerek öğrenim süresi sonundaki akademik başarıyı tahmin etmişlerdir. Çalışma ile başarısızlık gösterecek öğrencilere zamanında uyarı ve destek hizmetleri verilmesi, başarı gösterecek öğrencilere ise tavsiye ve fırsatların sunulabileceği belirlemiştir.

Cunningham (2017) yaptığı çalışmada, çevrimiçi (online) kurslara kayıtlı öğrenci verilerini kullanarak kursun ilk gününde öğrencinin kursu ne zaman tamamlayacağını %80 oranında doğruluk ile tahmin edebilecek modeller geliştirmiştir.

Dalkılıç ve Aydın (2017), apriori algoritması ile öğrencilerin başarı durumlarını etkileyen faktörleri incelemiş. Çalışmanın sonucunda cinsiyet, bölüm türü, öğrenim türü, kayıtlı olunan yılın genel başarı durumu ve devamsızlık eğitimi üzerinde etkisinin olduğu sonucuna ulaşmışlardır.

Devasia, Vinushree ve Hedge'nin (2016) yaptığı çalışmada, Amrita Vishwa Vidyapeetham Üniversitesinin veri tabanında kayıtlı 700 öğrenci verisi kullanılarak gelecek dönemlere yönelik öğrenci başarısı kestirilmek istenmiştir. Veri setleri ile Naive Bayes, Regresyon, Decision Tree ve Yapay Sinir Ağı modelleri oluşturulmuştur ve bu modellerin performans değerleri kıyaslanarak en iyi model seçilmek istenmiştir. Uygulanan modellerin performans değerlerin incelendiğinde, en yüksek performans değeri gösteren modelin Naive Bayes olduğu gözlenmiştir.

Hamsa ve diğerleri (2016), bulanık mantık ve karar ağacı yöntemlerini kullanarak, bilgisayar, elektronik ve komünikasyon bölümlerinde öğrenim gören lisans ve yüksek lisans öğrencilerinin akademik performanslarının kestirimine yönelik bir uygulama gerçekleştirmişlerdir. Uygulamalarının performans değerleri ölçüldüğünde bulanık mantık modelinin performans değerinin yüksek seviyede olduğu gözlenmiştir.

Hassana ve Al-Razgan (2016) yaptıkları çalışma ile, üniversiteye yerleşen öğrencilerin üniversite dönem ortalamalarına etkisini belirlemeye yönelik regresyon modelleri geliştirmişlerdir. Geliştirilen regresyon modelleri incelendiğinde; lise not ortalamasının üniversite not ortalamasını, üniversite öncesi sınavlara göre daha fazla etkilediği görülmüş ve kayıt olunan yılın üniversite not ortalaması üzerinde beklenmedik bir etkisinin olduğunu belirlemişlerdir.

Natek ve Zwilling (2014), yüksek öğretimde veri madenciliği araçları ile öğrenci başarı oranı tahminine yönelik bir çalışma gerçekleştirmişlerdir. Yaptıkları çalışmada J-48 algoritması ile öğrenci başarısını %98 doğruluk ile tahmin etmişlerdir. Çalışma sonucunda yükseköğrenim sistemlerinde eğitsel veri madenciliği uygulamalarının öğrenci bilgi sistemlerine entegre bir biçimde çalıştırılması gerektiğini belirtmişlerdir.

Şengür (2013) çalışmasında, Fırat Üniversitesi Eğitim Fakültesi BÖTE bölümünden mezun olmuş 127 öğrencinin verilerini kullanarak mezuniyet notu kestirimine yönelik yapay sinir ağı ve karar ağacı modelleri karşılaştırılmıştır. Mezuniyet notu kestirimine yönelik iki farklı aşama gerçekleştirilmiş. İlk aşamada, öğrencinin

birinci ve ikinci sınıflardaki yıl sonu notu, ikinci aşamada ise, öğrencilerin bir, iki ve üçüncü sınıflara ait yıl sonu notları kullanılarak mezuniyet notu kestirimi yapılmıştır. Yapılan çalışmalar incelendiğinde ise her iki aşamada karar ağaçları algoritmasının yüksek performans gösterdiği sonucuna ulaşılmıştır.

Ekim (2011) yaptığı çalışma ile, Selçuk Üniversitesinde kayıtlı öğrenci verileri ile öğrenciler hakkında geleceğe yönelik kestirimlerde bulunmak için kestirim modelleri oluşturmuştur. Apriori algoritması ve karar ağacı modeli ile öğrencinin başarısına etki eden faktörler belirlenmeye çalışılmış. Modellerin sonucu incelendiğinde ailenin eğitim seviyesi ve gelir düzeyinin öğrenci başarısında en etkili faktörler olduğu belirlenmiştir.

Özetlemek gerekirse EVM tahmine yönelik uygulamalarında genel olarak öğrencinin akademik başarı notları ve mezuniyet notu kestirimine yönelik farklı algoritmalar ile tahminine yönelik çalışmaların yapıldığı gözlenmiştir. Tahmine yönelik modellemeyi oluşturmak için sınıflandırma, regresyon ve ayırma gibi tekniklerin kullanıldığı belirlenmiştir. Alanyazın taramasında öğrencilerin performansını tahmin etmede en popüler olanının sınıflandırma yöntemi olduğu gözlenmiştir (Çetintav, 2022; Topuz, 2021). Sınıflandırma tekniği ile öğrenci performansını tahmin etmek için kullanılan teknikler arasında karar ağacı, yapay sinir ağları, naiv bayes, K-en yakın komşu ve destek vektör makinesi yöntemleri kullanılmaktadır (Shahiri, Husian ve Rashid, 2015). İncelenen kestirim modellerinde başarı oranının yaklaşık olarak %70-90 aralığında sonuçlar göstermesinin başarılı tahmin modeli sonucu kabul edildiği tespit edilmiştir. Yapılan çalışmalarda paket programların kullanıldığı ve çalışmada kullanılan akademik verilerin ise genel olarak öğrenci bilgi sistemlerinden elde edildiği belirlenmiştir. EVM çalışmalarında genel olarak karşılaşılan sorunlar veriye erişim ve etik durumların çalışmaları etkilediği ve istenilen araştırmaların yapılmasında bu engellerin önemli sorunlar oluşturduğu sonucuna ulaşılmıştır (Çetintav ve diğerleri, 2022; Demiral ve diğerleri, 2017; Roberts ve diğerleri, 2016). Yaptığımız çalışma literatürün gözden geçirilmesi ile EVM alanında regresyon modellerinin daha yaygın kullanılması ve bu modeller ile hem sınıflandırma hem kestirim yöntemleri geliştirilerek, Türkiye'de EVM alanında gelecekte yapılacak çalışmalara katkı sağlanması açışından önemli olacaktır.

BÖLÜM III

3. YÖNTEM

Bu bölümde, araştırma modeli, evreni ve örneklem bilgileri, veri toplama süreci, çalışma sürecinde kullanılan veri toplama araçları ve elde edilen verilerin analizine ilişkin bilgiler yer almaktadır.

3.1. Araştırmanın Modeli

Bu araştırmada, eğitsel veri madenciliği süreç tasarımlarından CRISP-DM iş döngüsü tasarımı adımları kullanılarak tahmin çalışmaları yapılmıştır. CRISP-DM iş döngüsü veri madenciliği uygulama adımlarını iş analitiği ile birleştiren, iş süreçlerini anlama, veriyi anlama, verinin hazırlanması, modelleme, değerlendirme ve yayılım alt süreçlerinden oluşur (Schröer vd., 2021; Marinez-Plumed, 2019). Çalışmada, İnönü Üniversitesi Otomasyon Bilgi Sistemindeki öğrenci verileri kullanılarak öğrencilerin mezuniyet süresi tahmini ve Bilgisayar II dersi başarı durumu tahminine yönelik eğitsel veri madenciliği regresyon modelleri oluşturulmuştur. Araştırma büyük veriden öğrenme analitikleri ve EVM ile anlamlı veriler elde etme işlemi olduğundan Bilgisayar dersi seçilmiştir. Bilgisayar dersi, üniversitede bulunan fakülte ve bölümlerin hedefleri doğrultusunda kazandırılmak istenen temel seviyede bilişim teknolojilerinin kullanımına yönelik tüm bölümlerde verilen bir dersidir.

Çalışma da tahmine dayalı bir model oluşturulmak istendiğinden veri madenciliği regresyon tahmin modelleri kullanılmıştır. Eğitim alanında yapılan veri madenciliği çalışmaları eğitsel veri madenciliği olarak nitelendirilmektedir (Romero ve Ventura, 2013). Çalışmada EVM süreçleri takip edilerek OBS sisteminden elde edilen veriler ile geleceğe yönelik kestirimde bulunulmak istemişlerdir. Mezun olma süresine ilişkin Lojistik Regresyon, Bilgisayar II dersi başarı tahmini için doğrusal ve lojistik regresyon kullanılmıştır.

Çalışma eğitsel veri madenciliği adımlarına uygun olarak yürütülmüştür. EVM adımlarının tanımlandığı ve farklı uygulamalarda da kullanılabilen standart veri madenciliği sürecinin işlediği CRISP-DM (Veri madenciliği için sektörler arası standart süreç) süreç modeli izlenmiştir. Bu model EVM modelinin bağımsız çalışmalarda kullanılmasına imkân sağladığından ve EVM uygulama adımlarını içerdiğinden seçilmiştir. Çalışmada kullanılan CRISP-DM süreç modeli adımları aşağıda açıklanmış ve adımlar çalışmaya uygun olarak evren örneklem, veri toplama teknikleri ve verilerin analizi bölümlerinde ayrıntılı olarak açıklanmıştır.

3.2. Çalışma Grubu

Çalışmada araştırma problemleri doğrultusunda İnönü Üniversitesi Otomasyon Bilgi Sistemine kayıtlı mezun veya öğrenimine devam eden (N=223.279) öğrenci verileri kullanılmıştır. Çalışma grubu uygun örneklem yöntemine göre seçilmiştir. Uygun örneklem yöntemi bir evrenin tamamının ölçülemediği durumlarda, evreni en iyi şekilde temsil eden rassal olarak seçilmiş yeterli büyüklükteki veri kümeleridir (Baştürk ve Taştepe, 2013; Başkale, 2016). Uygun örneklem yönteminin seçilme nedeni ise, araştırma kapsamında gereken verilere erişimin verilmesi ve bu grubun mezun olma durumu ve öğrencilik özelliklerinin biliniyor olmasıdır. Çalışmada üniversiteyi bitirme notunun tahmini için mezun olan öğrenciler içerisinden eksik verisi bulunmayan kişilerin verileri ele alınmıştır. Bilgisayar dersi tahmini için ise bu dersi alan lisans düzeyinde eğitim almış öğrencilerden eksik verisi bulunamayan öğrencilerin verileri seçilmiştir. Çalışma bir veri madenciliği uygulaması olduğu için ele alınan veri setinin büyük olması sonuçların daha doğru ve geçerli olmasını sağlamaktadır.

3.3. Verilerin Toplanması

Çalışma kapsamında İnönü Üniversitesi OBS sistemindeki farklı veri tabanları kullanılarak bir veri yığını oluşturulmuş ve elde edilen veri ambarı ile araştırma problemlerine uygun analizler yapılmıştır. Verilerin kullanılmasında uygulanacak etik ilkeler belirlenerek İnönü Üniversitesi Araştırma ve Yayın Etiği Kurulundan gerekli izinler alınmıştır (Ek-1). Etik kurul izinleri alındıktan sonra üniversite Öğrenci İşleri Daire Başkanlığı tarafından araştırma için gerekli verilerin alımı için izin başvurusu yapılıp gerekli izin alınmıştır (Ek-2). Öğrenci İşleri Sistem yöneticileri tarafından

verilerin alınması için OBS sistemine şahsım adına Sistem Yönetimi Salt Okuyucu grubunda erişim verilmiştir. Verilerin toplanması detaylı olarak CRISP-DM iş süreci Veri Anlama bölümünde anlatılmıştır.

3.4. Verilerin Analizi

Araştırmanın bu bölümünde veriler analiz edilirken izlenen CRISP-DM süreç modelinin basamaklarına uygun olarak yapılan eğitsel veri madenciliği çalışması açıklanmıştır. Veriler veri madenciliği yazılımı olan RapidMiner Studio programı kullanılarak analiz edilmiştir. RapidMiner Studio görsel iş akışına ve tam otomasyona sahip, makine öğrenmesinden model geliştirilmesine kadar tüm veri bilim döngüsünü gerçekleştiren bir platformdur (RapidMiner, 2020; Wahyuni, S. 2018; Hofmann ve Klinkenberg, 2016).

3.4.1. Araştırma İş/Problem Anlama

Araştırmada problem durumunun belirlendiği CRISP-DM sürecinin ilk aşamasıdır. Bu aşamada, çalışmanın amacının belirlenip, mevcut durumun değerlendirilip, hedefler belirlenerek araştırma planı oluşturulmuştur.

Araştırma Amaçlarının Belirlenmesi. Araştırma büyük verinin eğitim alanında önemine dikkat çekmek ve bu alandaki etkilerinin anlaşılması için eğitsel veri madenciliği yöntemi ile İnönü Üniversitesi OBS sistemi üzerinde bulunan verilerin işlenerek öğrenci performansının kestirimi amaçlanmıştır. Geliştirilen modeller ile öğrenci akademik performansının arttırılması ve gerekli önlemlerin zamanında alınmasında kullanılabilir.

Araştırma Mevcut Durumunun Değerlendirilmesi. İnönü üniversitesi otomasyon sistemi 2016 yılına dek üniversite bünyesinde yer alan sunucular üzerinden hizmet vermiştir. 2016 yılı itibari ile üniversite Otomasyon Bilgi Sistemine (https://obs.inonu.edu.tr/oibs/login.aspx) geçiş yapmış ve öğrenci verileri bu sistemde depolanmaya başlamıştır. Eski sistemde yer alan veriler mevcut sisteme aktarılmış ancak aktarım esnasında verilerin bir bölümü eksik veya yanlış bir şekilde işlendiği tespit edilmiştir.

Otomasyon Bilgi Sistemi İnönü Üniversitesi Öğrenci işleri tarafından üniversite sunucuları üzerinden yönetilmektedir. OBS sistemi web tabanlı bir ara yüz tarafından erişim kolaylığı sağlanmaktadır. Çalışma için gerekli verilerin alınması için şahsıma yönetici gurubu salt okuyucu düzeyinde erişim izni verilmiştir.

Araştırma Hedeflerinin Belirlenmesi. Araştırmada öğrencilerin akademik performansını kestirebilmek hedeflenmiştir. Bu hedef doğrultusunda eğitsel veri madenciliği regresyon tahmin modelleri geliştirilmiştir. Çalışmada kullanılacak kestirim hedefleri aşağıda belirtilmiştir;

- Öğrencilerin mezun olma süreleri öğrencilerin kişisel bilgileri (Cinsiyet, Medeni Durum, Yaş, İl) ve akademik bilgileri (Üniversiteye Yerleşme Puanı, Yerleşme Puan Türü, Lise Bitirme Puanı, Mezuniyet Notu) ile kestirimi,
- Bilgisayar II dersi geçme durumu öğrencilerin kişisel bilgileri (Cinsiyet, Medeni Durum, Yaş, İl) ve akademik bilgileri (Üniversiteye Yerleşme Puanı, Yerleşme Puan Türü, Lise Bitirme Puanı, Bilgisayar I Dersi Geçme Durumu ve Notu) ile kestirimi.
- Bilgisayar II dersi geçme notu öğrencilerin kişisel bilgileri (Cinsiyet, Medeni Durum, Yaş, İl) ve akademik bilgileri (Üniversiteye Yerleşme Puanı, Yerleşme Puan Türü, Lise Bitirme Puanı, Bilgisayar I Dersi Geçme Durumu ve Notu) kullanılarak doğrusal regresyon ile kestirilebilir mi?

Araştırma Planı. Sürecin son aşaması olan bu kısımda araştırma süreci planı oluşturulur. Araştırma planı Tablo 3.1'de gösterildiği şekilde yürütülmüştür.

Tablo 3.1.

Araştırma Planı.

	İş Adımı	İlgili Birim
1	Araştırma yapılacak konunun seçilmesi ve ilgili birimlere bildirilerek gerekli izinlerin alınması.	Eğitim Bilimleri Enstitüsü Etik Kurul Yönetimi Rektörlük
2	Çalışmada kullanılacak verilere erişimin sağlanması.	Öğrenci İşleri Daire Başkanlığı Sistem Yönetimi / Otomasyon Bilgi Sistemi (OBS)
3	Verilerin alınması ilişkisel veri tabanlarının çözümlenmesi, veri tabanlarındaki bilgilerin keşfedilmesi ve tablolar arasındaki ilişkileri belirlenmesi.	Araştırmacı, (OBS) Sistem yetkililerinden bilgi alınması
4	Veri tabanlarından gerekli verilerin çekilmesi özet bilgi çıkartılması.	Araştırmacı
5	Veri ön işleme ve veri setinin hazırlanması	Araştırmacı
6	Modellerin Geliştirilmesi. Araştırma kapsamında uygulanacak eğitsel veri madenciliği modellerinin oluşturulması.	Araştırmacı
7	Değerlendirme. Oluşturulan modellerin amaca uygunluğunun ve başarı durumunun değerlendirilmesi.	Araştırmacı
8	Elde edilen modellerin uygulanması	Araştırmacı

3.4.2. Veriyi Anlama

CRISP-DM iş döngüsü sürecinde veriyi anlama aşaması ilk verinin toplanması, veriyi keşfetme ve verinin kalitesinin belirlenmesi alt aşamalarından oluşmaktadır. İlk veri, İnönü Üniversitesi Otomasyon Bilgi Sistemindeki veri tabanlarından elde edilmiştir. Öğrenci işleri veri tabanları kullanıcı kolaylığı açısından web tabanlı bir ara yüzden erişime sunulmaktadır. Sistemden veriler bu ara yüz üzerinden salt okuyucu grubu ile erişim sağlanarak elde edilmiştir. Çalışmada kullanılacak verilere erişim "https://obs.inonu.edu.tr/oibs/start.aspx?" web adresinden sağlanmış olup ara yüz içerisinde yer alan veri tabanlarından sorgu penceresinde çalışmada kullanılacak verilere

uygun sorgular ile elde edilmiştir. Veri tabanlarında yer alan verilerin kısaltılmış olması bu verilerin anlaşılmasında güçlük yaşanmasına neden olmuş, kısaltmalar ile ilgili gerekli bilgiler öğrenci işleri sistem yöneticilerinden alınmıştır. Çalışmada kullanılacak veriler Microsoft Excel dosya biçiminde sistemden çekilmiştir. OBS sisteminin web tabanlı çalışması kişisel bilgisayarın RAM belleğini önemli ölçüde kullanmaktadır. Buda verilerin çekilmesinde zorluklar çıkarttığından veriler küçük parçalar halinde çekilmiştir.

İlk verinin toplanmasından sonra veriyi tanımlamak amacı ile yerel bilgisayara kaydedilen Excel dosya sisteminde yer alan veriler öncelikle Excel programında incelenmiştir. Bu aşamada verilerin büyüklüğü, tipi gibi özelliklere bakılmış ve sayısal değerlerin ondalık, yüzdelik ya da bindelik değerlerinde yer alan hatalar düzeltilmiştir. Microsoft Excel dosya biçiminde küçük parçalar halinde yer alan veriler düzenlendikten sonra veri madenciliği programı RapidMiner Studio yazılımı ile birleştirilerek detaylı bir şekilde incelenmiştir. Elde edilen toplam veri yığını gerekli hataların (öğrenci ÖSYM notunun 456,69 yerine 45669 şeklinde girilmesi gibi) düzeltilmesi ile çalışmada kullanılacak veri seti hazırlanmıştır. Oluşturulan veri setinde yer alan bilgilerin OBS sisteminde hangi veri tabanında ve tablodan alındığı ve verinin türü gibi bilgiler Tablo 3.2'de gösterilmiştir.

Tablo 3.2.

Öğrenci Bilgi Sistemi Kullanılan Veri Tabanları, Tablolar ve Veriler.

	Veri Tabanı Adı	Tablo Adı	Veri	Türü
1	Öğrenci	Özlük Bilgileri	Öğrenci No	Sayı
		Ön Kayıt Bilgileri	Lise Puanı	Sayı
			Kayıt Tarihi	Tarih
		Kimlik Bilgileri	Medeni Durum	Metin
			Cinsiyet	Metin
			Yaş	Sayı
			İl	Metin
			Doğum Tarihi	Tarih
		Fakülte Bölüm Bilgisi	Fakülte adı	Metin
			Bölüm adı	Metin
		ÖSYM Bilgileri	Yerleşme Puanı	Sayı
			Yerleşme Puan Türü	Metin
2	Dönem Ders Bilgisi	Ders Hareket Bilgileri	Ders Adı	Metin
			Bölümü	Metin
			Dönemi	Metin
3	Öğrenci DK Sınavlar	Özlük Bilgileri	Öğrenci No	Sayı
		Ders Kayıt Bilgileri	Geçme Durumu	Metin
			Devam Bilgisi	Metin
		Sınav Bilgileri	Ortalama Notu	Sayı
			Harf Kodu	Metin
4	Müfredat ve Blogna	Fakülte Bölüm Bilgileri	Bölümü	Metin
		Müfredat Bilgileri	Ders Adı	Metin
			Ders Diğer Adı	Metin
			Müfredat Yılı	Tarih
5	Program Bilgileri	Fakülte Bölüm Bilgisi	Bölümü	Metin
			Programı	Metin
		ÖSYM Taban Puan Bilgileri	Yıl	Tarih
			ÖSYM Taban Puan	Sayı
6	Mezunlar Portalı	Özlük Bilgileri	Öğrenci No	Sayı
			Okuduğu Yıl	Sayı
		Diploma ve Mezuniyet Bilgileri	Mezuniyet Yılı	Tarih
			Mezuniyet Notu	Sayı

Otomasyon Bilgi Sisteminde öğrenci verileri, ders bilgileri ve personel bilgilerinin yer aldığı 10 adet veri tabanı bulunmaktadır. Bu veri tabanlarından yapılacak çalışmada kullanılacak verileri barındıran Tablo 3.2'de belirtildiği gibi altısı kullanılmış olup kullanılan veri tabanlarında yer alan tablolardan ise yine çalışma için gerekli olanlar kullanılmıştır.

Veriyi tanımlama aşamasından sonra verinin keşfi aşamasına geçilmiş ve bu aşamada tablolarda elde edilen verilerin içerikleri ile ilgili detaylı inceleme gerçekleştirilmiştir. Veri setinin birleştirilmesi ile veri setinde yer alan öğrenci bilgilerinden üniversiteye yerleşme puanı değerlerinden birçoğunun eksik olduğu görülmüştür. Benzer sorunun lise yerleşim puanında da olduğu görülmüş ve bu değişken değerlerindeki eksik verilerin veri tabanına kaydedilmemesi olduğu sonucuna ulaşılmıştır. Veri setindeki bu eksiklikler öğrencilerin kayıt yıllarına göre incelendiğinde genel olarak 2016 yılı ve öncesinde yoğun olduğu gözlenmiştir. Üniversite sistem yöneticilerine bu konu hakkında bilgi verildiğinde kayıp verinin kaynağının 2016 yılında yapılan sistem değişikliği olduğu ve yeni sisteme geçmiş verilerin tam olarak yüklenemediği cevabı alınmıştır.

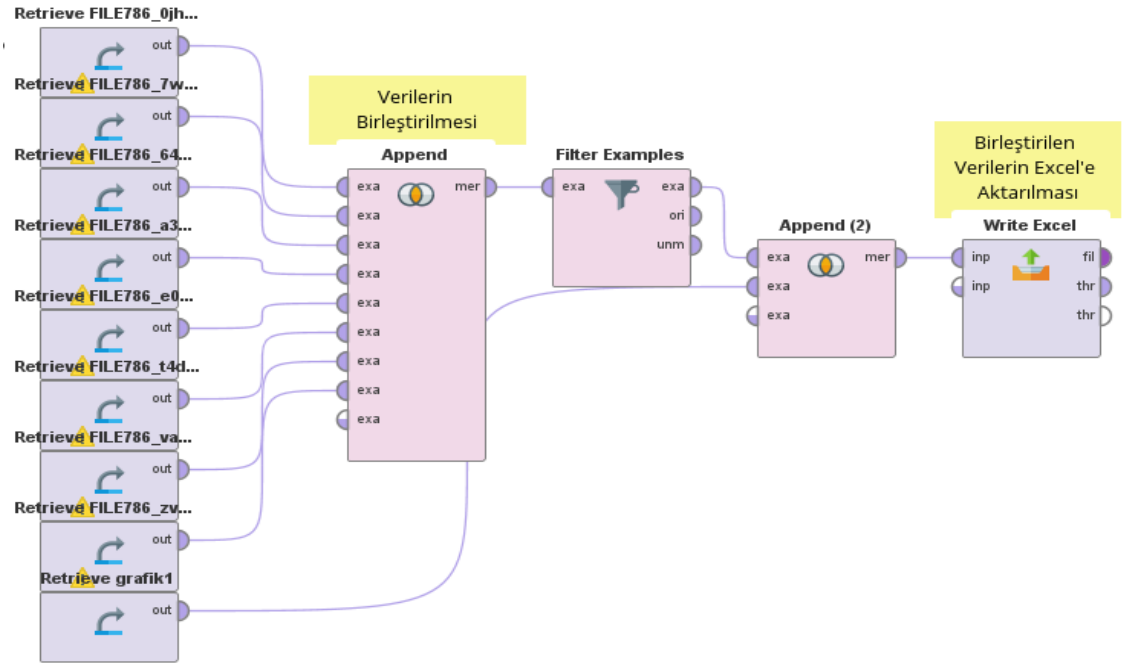
Bilişim dersi ile ilgili verilerin elde edilmesi sürecinde bu dersin fakülte ve bölümlerde farklı isimler ile verildiği tespit edilmiştir. Farklı isimler ile verilen dersin sistemden tespit edilmesi için Dönem Ders Bilgisi veri tabanı ile Müfredat ve Blogna veri tabanları detaylı bir şekilde incelenmiş ve dersin güncel isimleri tespit edilmiştir. Ancak veri elde etme esnasında az sayıda verinin olduğu gözlenmiş ve bunun nedeninin fakültelerde verilen bu dersinde yıllara göre ve müfredattaki değişikliklere göre aynı isimle görüldüğü ancak ders kodunun yenilendiği sonucuna varılmıştır. Bu nedenle veriler derslerin isimleri ile tek tek girilerek sistemden elde edilmiştir. Ders isimlerinin farklı olması nedeni ile bu derslerin içeriği bölümlerin sayfalarında incelenmiş ve aynı içeriğe sahip olduğu ve temel bilişim dersi olduğu sonucuna ulaşılmıştır. Fakülte ve bölüm programlarına göre isimlerinin değiştiği Bilişim dersine yönelik bilgiler Tablo 3.3'te gösterilmiştir. Farklı isimler ile isimlendirilen bu dersler birleştirildikten sonra genel isim olarak birinci dönem dersine "Bilgisayar I", ikinci dönem dersine ise "Bilgisayar II" ismi verilmiştir.

Tablo 3.3.

Fakülte ve Bölümlerde Bilgisayar Dersi.

Fakülte	Bölüm	Dersin Adı
Beden Eğitimi ve Spor Yüksekokulu	Beden Eğitimi ve Spor Öğretmenliği	Bilişim Teknolojileri Bilgisayar
	Beden Eğitimi ve Spor Eğitimi	
	Engelliler Egzersiz ve Spor Eğitimi	
	Engellilerde Beden Eğitimi ve Spor Eğitimi	
Eğitim Fakültesi	Eğitim Bilimleri	Temeli Bilgi Teknolojisi Bilişim Teknolojileri
	Türkçe ve Sosyal Bilimler Eğitimi	
	Yabancı Diller Eğitimi	
	Özel Eğitim Bölümü	
	Güzel Sanatlar Eğitimi	
	Matematik ve Fen Bilimleri	
Eczacılık Fakültesi	Eczacılık	Bilgisayar
Fen Edebiyat Fakültesi	Kimya	Temel Bilgi Teknolojileri
	Biyoloji	Temel Bilgi Teknolojisi
Güzel Sanatlar ve Tasarım Fakültesi	Müzik Bilimleri	Temel Bilgi Teknolojileri
	Resim	
	Seramik	Temel Bilgi Teknolojisi
	Grafik Tasarım	
Hemşirelik Fakültesi	Hemşirelik	Bilgisayar Bilgi Teknolojisi
İktisadi ve İdari Bilimler Fakültesi	Ekonometri	Temel Bilgi Teknolojileri
	Maliye	
	İşletme	Temel Bilgi Teknolojisi
Ziraat Fakültesi	Bahçe Bitkileri	Bilgisayar
	Bitki Koruma	
İletişim Fakültesi	Gazetecilik	Temel Bilgi Teknolojileri

Verinin keşfedilmesi aşamasından sonra veri seti EVM modellerinin oluşturulması için, farklı veri tabanlarında küçük parçalar halinde çekilmiş veriler Rapid Miner Studio programında Şekil 3.1'de görüldüğü gibi birleştirilmiştir. Veriler birleştirilirken öğrenci numarası ID (kimlik) olarak kullanılmış ve Append (ekle, birleştir) operatörü ile satır bazında birleştirilmiştir. Filter (filtre) operatörü ile bilişim dersi ile ilgili veriler dönemsel olarak ayrılmış ve dönemlerine göre birleştirilmiştir. Hazır olan veri seti programdan dışa Write Excel operatörü ile Excel dosya biçiminde dışa aktarılmış ve yedeklenmiştir.



Şekil 3.1. Verilerin RapidMiner Studio ile birleştirilmesi.

CRISP-DM iş sürecinin verinin anlaşılması aşamasının alt aşamaları olan ilk verinin toplanması, veriyi tanımlama ve veriyi keşfetme aşamaları ile veriler birleştirilip modele uygun veri seti oluşturulmuş ve veri setinin hazırlanması için sürecin diğer aşaması olan veri hazırlama bölümüne geçilmiştir.

3.4.3. Veri Hazırlama

CRISP-DM sürecinin bu aşamasında elde edilen veriler modelleme öncesinde incelenerek verinin kalitesini arttırmak ve modele uygun veri setinin hazırlanması için gerekli düzenlemelerin yapıldığı en önemli süreçtir. Bu sürecin çalışmanın amacına uygun olacak şekilde yürütülmesi, araştırmada zaman kaybını önleyecek ve çalışmanın daha verimli sonuçlar üretmesine sebep olacaktır.

Bu süreçte birleştirilmiş olan veri seti oluşturulacak modellere uygun olarak düzeltilip analize hazır hale getirilmesi gerekmektedir. Veri seti RapidMiner Studio programı ile modellerde kullanılacak değişkenler isimlendirilmiş ve mevcut durumuna bakılmıştır. Veri setinde 223.279 öğrenci verisi içerisinden 40.471 öğrenci aktif olarak üniversitede öğrenim görmekte ve 171.163 öğrencide mezun durumunda görünmektedir. Öğrenim durumu aktif olan öğrenciler ile mezun olan öğrenciler toplandığında veri setinde yer alan toplam öğrenci sayısına ulaşılmadığı görülmüş. Bu sınıflandırmada yer almayan öğrencilerin üniversiteden ayrıldığı ya da atıldığı sonucuna ulaşılmıştır.

Veri setinde mezun olan öğrenci verilerinden 78.697 öğrenci verisinin işlenebilir düzeyde veri içerdiği belirlenmiştir. Mezuniyet kestirimi için veri setinde bu kişilerin verileri kullanılmıştır. Bilgisayar dersi kestirim modeli için ise veri setinde kayıp verisi bulunmayan ve bu dersi almış 9.160 öğrenci verisi kullanılmıştır. Veri hazırlama sürecinde verilerin seçilmesi, biçimlendirilmesi ve düzenlenmesi RapidMiner Studio programı ile gerçekleştirilmiştir. Bu aşamaların anlaşıla bilmesi için programa ait bazı işleçler aşağıda açıklanmıştır.

- Retrieve (Alma): Programa analiz edilecek veriyi yükler. Programa alınan veriyi temsil eder ve çalışmalarda kullanılmak üzere gerekli yerlere sürükle bırak yöntemi ile eklenebilir.
- Select Attiributes (Nitelik Seçimi): Veri setinden kullanılacak olan veriyi seçmeye yarar. Bu operatör tablo şeklinde alınan veri setinden işlenmesi istenen sütunların seçiminde kullanılır.
- Filter (Filtreleme): Bu operatör ile veri setinde seçmek istediğimiz özelliklere göre veri seçmemizi sağlar. Örneğin; yaşı 30'dan büyük kişilerin seçimi için "Yaş>30" şeklinde parametreler kullanılarak yapılır.
- Multiply (Çoğaltıcı): Veri setini işlem sırasında klonlayarak verinin

özelliğini kaybetmeden farklı bir iş içinde aynı anda kullanılmasını sağlar.

- Set Role (Rol ayarlayıcı): Bu operatör, bir veya daha fazla niteliğin rolünü değiştirmek için kullanılır. Veri setinde model için kullanılacak değişkenlerin niteliklerini belirtmede kullanılmaktadır.
- Cross Validation (Çapraz Doğrulayıcı): Bu Operatör, bir öğrenme modelinin istatistiksel performansını tahmin etmek için çapraz doğrulama gerçekleştirir. Operatörün iç içe geçmiş iki alt süreci vardır. Eğitim alt süreci ve test alt süreci. Eğitim alt süreci, bir modeli eğitmek için kullanılır. Eğitilen model daha sonra test alt sürecinde uygulanır ve modelin performansı test aşamasında ölçülür.
- Join (Birleştirme): Bu operatör, girdi kümelerinin bir veya daha fazla niteliğini anahtar olarak kullanarak iki kümeyi birleştirir. Kimlik rolüne sahip bir öznitelik anahtar olarak seçilir, ancak anahtar olarak bir veya daha fazla öznitelikten oluşacak ise bir dizi seçilebilir. SQL dilinde kullanılan Join etiketi ile aynı özellikleri göstermektedir.
- Write (Yazdır): Bu komut veri setinde düzenlenen verilerin dışarı aktarılmasında kullanılır. Virgülle ayrılmış dosyalar için CSV formatında, Excel dosya türünde aktarım için XLS formatında dışa aktarım sağlamaktadır.

Veri ön işleme sürecinde 2016 yılı öncesine ait Üniversiteye Yerleşme Puanlarını ve Lise Puanı değerlerinin eksik oluğu ve bazı değerlerin yüzdelik değer ayıracı olarak nokta ya da virgül ile belirtilmeden sisteme kaydedildiği görülmüştür (Örnek: 450,00 sayısı 45000 şeklinde girilmesi). Bu veriler veri setinden tespit edilerek gerekli biçime dönüştürülmüştür. Eksik değere sahip değişkenler detaylı incelenmiş ve sadece bir özellikte eksik veri gösteren değerler (Örneğin: Sadece yerleşme puanı eksik olan öğrenci değeri) k-NN (k en yakın komşu) algoritmasına göre değere benzer özellikte bulunan en yakın 5 komşusunun değer ortalaması alınarak düzeltilmiştir. Birden fazla eksi veriye sahip veriler ise veri setinden çıkartılmıştır.

Veri setinde öğrenci Yaş değişkeni sistem üzerinde öğrencinin şimdiki yaşını gösterdiğinden veri setinden Doğum Tarihi ve Kayıt Yılı ve Mezuniyet Yılı değişkenleri kullanılarak modelde ele alınacak yaş değeri hesaplanmıştır. Çalışmada oluşturulacak doğrusal regresyon modelleri için kategorik yapıya sahip ya da metin özelliği gösteren değişkenler Tablo 3.3'de gösterildiği gibi sayısal değerlere dönüştürülmüştür.

Tablo 3.4.

Değişken Değerlerinin Sayısal Verilere Dönüştürülmesi.

Değişken	Normal Değer	Sayısal Değer
Cinsiyet	Erkek	1
	Kadın	0
Medeni Durum	Evli	1
	Bekar	0
Yerleşme Puan Türü	Sayısal	1
	Sözel	2
	Eşit Ağırlık	3
	TYT	4
	Dil	5
	Özel Yetenek	6

Doğrusal regresyon analizi için veri setinde kategorik yapıda bulunan cinsiyet, medeni durum ve yerleşme puan türü değişkenleri Tablo 3.3'de görüldüğü gibi sayısal verilere dönüştürülmüştür. Veri seti üzerinde ön işlemlerin yapılmasının ardından Modelleme sürecine geçilmiş ve çalışmada uygulanacak modeller oluşturulmuştur.

3.4.4. Model Oluşturma

Model oluşturma sürecinde araştırma problemi doğrultusunda eğitsel veri madenciliği tahmin modellerinden regresyon analizi kullanılmıştır. Eğitsel veri madenciliği yöntemi ile mevcut veriler ile şu anki durumun analizi ve geleceğe yönelik çıkarımlarda bulunmak mümkün olabilmektedir (Tekin ve Şengür, 2013). Bu nedenle araştırmada eğitsel veri madenciliği tahmin modellerinden regresyon analizi kullanılmıştır. Regresyon iki ya da daha fazla değişken arasında ilişki belirleme ve bu ilişki ile yeni verilere yönelik kestirimde bulunmayı sağlayan istatistiki analiz yöntemidir (Gamgam ve Altunkaynak, 2015). Çalışma amaçları doğrultusunda Mezuniyet Süresi kestirimi ve Bilgisayar dersi geçme durumu kestirimi için Lojistik regresyon yöntemi, Bilgisayar dersi not kestirimi için ise doğrusal regresyon yöntemi uygulanması karşılaştırılmıştır.

Regresyon modelinde elde edilen değer ile gerçek değer arasındaki fark hata terimidir. Bu değerin küçük olması modelin tahmin değerinin yüksek olduğunu gösterir. Regresyon modeli EVM sürecinde bir makine öğrenmesidir. Veri setinde bulunan verilerden bir miktarı öğrenmeye ayrılır böylelikle modelin oluşması için bir makine öğrenmesi gerçekleştirilir. Veri setinden ayrılmış diğer veri kümesi ise öğrenme sonucu

oluşmuş modeli test etmek için kullanılır. Bu sayede oluşturulan modelin tahmin değerleri ile test grubunda ele alınan gerçek değerler karşılaştırılır ve modelin başarı durumu, hata değeri tespit edilmiş olur.

3.4.4.1. Öğrenci Mezuniyet Süresi Kestirim Modeli

Öğrenci mezuniyet süresi kestirim modeli ile üniversiteye kayıt yaptıran öğrencilerin, kayıt sonrası üniversiteden öğrenim görecekleri fakülte ve bölümlerin normal öğrenim süresi içerisinde tamamlayabileceklerini tespit etmek amacı ile yapılmıştır. Bu model ile öğrencilerin normal süre dışında üniversiteyi bitirme ya da bırakma ihtimallerine karşın gerekli akademik ve idari önlemler alınarak bu süre zarfında mezun olabilmeleri hedeflenmektedir. Normal öğrenim süresi içerisinde üniversiteyi tamamlayan öğrenciler çalışma hayatına erken katılması sağlanıp ülkenin istihdam talebini karşılayabileceği düşünülmektedir.

Belirtilen hedefler çerçevesinde öğrenci kişisel bilgilerinden cinsiyet, medeni durum, yaş ve il değişkenleri ile akademik bilgilerinden üniversiteye yerleşme puanı, yerleşme puan türü, lise mezuniyet notu ve üniversite mezuniyet notu bağımsız değişkenleri kullanılarak üniversiteden mezun olma süresi tahmini amaçlanmıştır. Model oluşturulurken seçilen bağımsız değişkenlerin seçilme nedenleri aşağıda belirtilmiştir.

Cinsiyet; Öğrenim hayatında öğrencilerin meslek seçiminde ve akademik başarılarında toplumsal rollerden etkilendiği görülmektedir (Vatandaş, 2007). Bu etkinin başarıyı olumlu ya da olumsuz yönde etkiledi görülmüştür. Cinsiyet değişkeninin modelde kullanılması, toplumsal rollerden kaynaklı bu değişkenin öğrenci performansına ne derecede etki ettiğinin tespit edilmesini ve modelin tahmin etmede ki başarısının artırımıdır.

Medeni Durum; Bir kimsenin evli ya da bekar olması durumunu belirten bu değişkenin başarıya etkileri çeşitli çalışmalarda ele alınmıştır. Yapılan çalışmalar incelendiğinde medeni durumu evli olan kişilerin akademik çalışmalarında destekleyici bir etki gördükleri için başarılarının arttığı yönünde sonuçlar elde edilmiştir (Şenel ve Kutlu, 2015). Çalışmada bu değişkenin kullanılması ile öğrencilerin medeni durumlarının başarıdaki etkileri ölçmek istenmiş ve oluşturulan modelin performansını arttırmak hedeflenmiştir.

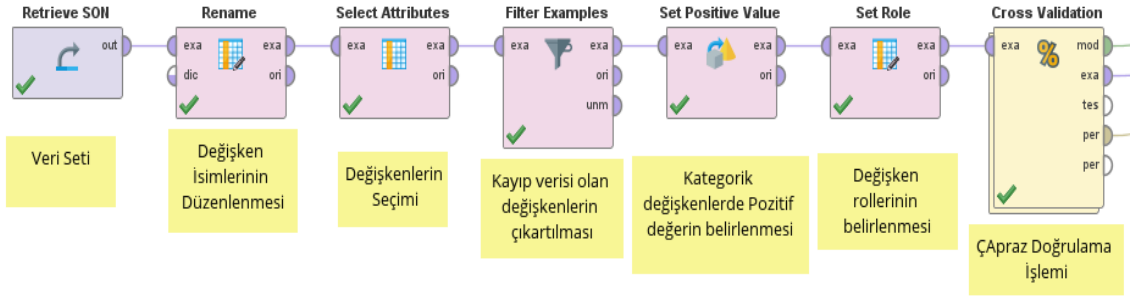
Yaş; Bireylerin yaşamlarında aldıkları eğitimler yaş ile ilişkilendirilmiştir. Bunun nedeni kişinin bilişsel algılama düzeyi yaşa bağlı olarak gelişme gösterdiğinden eğitim sistemleri ve eğitim ortamları bu kritere göre düzenlenmektedir. Yaşın ilerlemesi yaşamda deneyimin artmasını sağladığından başarıyı da arttırdığı yapılan çalışmalarda gözlenmiştir. Çalışmada bu değişkenin kullanılması ile öğrenme sürecinde başarının tespitinde kullanılacak modellerin kalitesini arttırmak hedeflenmiştir.

İl (Yerleşim yeri); Yerleşim yeri ile öğrencinin öğrenim gördüğü okul arasındaki mesafe öğrencinin başarısını etkilemektedir (Güvendir, 2014). Yerleşim yeri ile okulun farklı illerde olması öğrenci açısından maddi ve manevi bir zorluk oluşturduğundan başarı bu durumdan etkilenmektedir. Çalışmada bu değişkenin kullanılması, tahmin modellerinde daha verimli sonuçlara ulaşılacağı hedeflenmiştir.

Yerleşme puanı ve türü; Üniversitelerin öğrenci kabulünde kullanmış oldukları puan ve türüdür. Üniversite fakültelerindeki bölümlerin tercih edilişine göre belirlenen yerleşme puanı, öğrenim görmek isteyen öğrencilerin ÖSYM sınavına girerek aldığı puandır. Ülke genelinde başarı durumunun belirlendiği sınav, üniversitelerdeki başarıyı da arttırdığı bilinmektedir. Çalışmada bu değişkenlerin kullanılması oluşturulmak istenen tahmin modellerinin geçmiş başarıların etkisi ile ileriye yönelik kestirimde performans arttırıcı olarak görülmektedir.

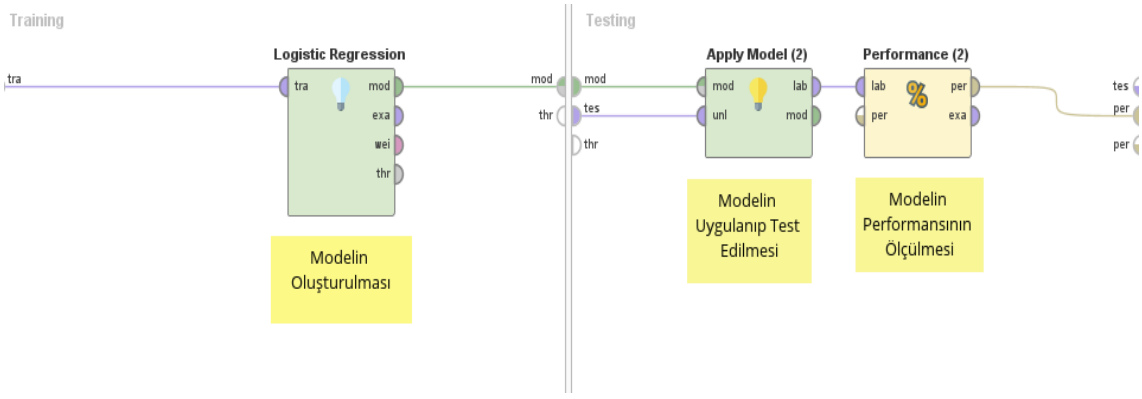
Lise mezuniyet notu; Öğrencilerin lise döneminde göstermiş oldukları akademik başarının yüksek öğretim sürecinde etkisini ölçmek ve geçmiş başarı durumu ile geleceğe yönelik başarı kestiriminde modelin performansını artırabileceği düşünülmektedir.

Değişkenler seçildikten sonra model için gerekli veriler birleştirilmiş ve işlenmek için RapidMiner Studio programındaki operatörler Şekil 3.2'de görüldüğü gibi eklenmiştir.



Şekil 3.2. Mezuniyet Süresi Lojistik Regresyon Modeli.

Programın işlem bölümüne veri seti (Retrieve) yüklenmiş ve değişken isimleri anlaşılır bir şekilde değiştirilmiştir. Değişken seçimi (Select Attributes) operatörü ile veri setinde model için gerekli olan değişkenler seçilmiştir. Filtre (Filter Examples) operatörü, kayıp veriye sahip değişkenlerin veri setinden çıkartılmasında kullanılarak modelin doğru sonuç vermesi sağlanmıştır. Pozitif değer seçme (Set Pozitive Value) operatörü kullanılarak değişkenlerde modelin alacağı pozitif değer (Cinsiyeti=Erkek olan ve Medeni Durumu=Bekar) belirtilmiştir. Rol belirleme (Set Role) operatörü il de modelde kullanılacak değişkenlerin ne amaçla kullanılacağı seçilmiştir. Bu kapsamda mezuniyet süresi değişkeni bağımlı değişken olarak Cinsiyet, Medeni Durum, İl, Yaş, Yerleşme Puanı ve Lise Puanı ise bağımsız değişken olarak belirtilmiştir. Fakülte değişkeni veri setinde grup belirleyen bir özellik olarak, Öğrenci No değişkeni ise her bir değerin bağımsız bir veri belirtmesi için kimlik (ID) olarak modelde belirtilmiştir.



Şekil 3.3. Mezuniyet Süresi Çapraz Doğrulama Süreci.

Veri setinden veriler seçilip rolleri belirlendikten sonra işlem yapmak için çapraz doğrulama (Cross Validation) operatörü ile 10 kat çapraz doğrulama yöntemi seçilmiştir.

Bu operatör ile veri seti on eşit parçaya ayrılarak dokuzu öğrenmeye biri test ayrılmıştır. Şekil 3.3'de görüldüğü gibi çapraz doğrulama operatörünün öğrenme (Training) bölümünde lojistik regresyon operatörü eklenip model oluşturulmuş, test (Testing) bölümünde ise modeli uygula (Apply Model) ve performans (Performance) operatörü eklenerek oluşturulan regresyon modeli test verisine uygulanıp modelin performansı ölçülmek istenmiştir.

3.4.4.2. Bilişim Dersi Lojistik Regresyon Modeli

Bilişim dersi lojistik regresyon kestirim modeli ile üniversitede öğrenim gören öğrencilerin almış oldukları bilişim derslerinin kalitesini arttırmak hedeflenmiştir. Bu hedef doğrultusunda üniversitede yer alan fakülte ve bölümlerde farklı isimlerde adlandırılan ancak ders içeriği aynı olan bilişim dersi için öğrenci verileri kullanılarak bu dersin ikinci döneminin başında dersi geçemeyecek öğrencileri belirleyip bunlara ek önlemler alınması ve ders başarısının arttırılması hedeflenmiştir.

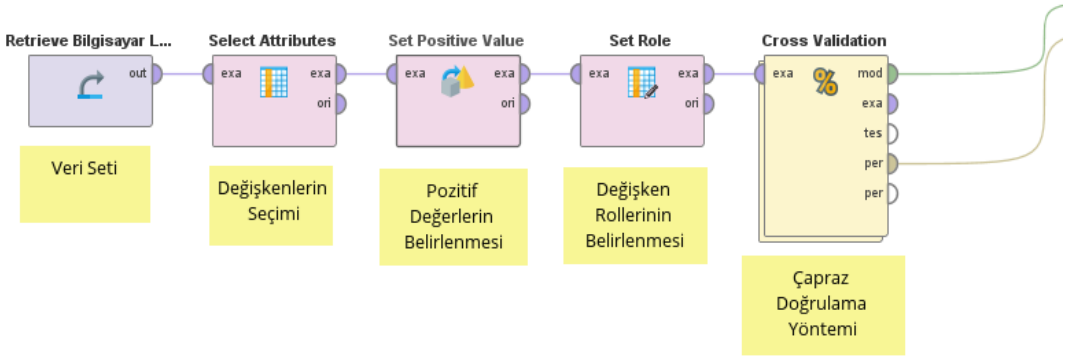
Belirtilen hedefler çerçevesinde öğrenci kişisel bilgilerinden cinsiyet, medeni durum, yaş ve il değişkenleri ile akademik bilgilerinden üniversiteye yerleşme puanı, yerleşme puan türü, lise mezuniyet notu, Bilgisayar I dersi geçme notu ve geçme durumu bağımsız değişkenleri kullanılarak Bilgisayar II dersi geçme durumu kestirilmek istenmiştir. Bilgisayar dersi alan bölümlerin ders isimleri Tablo 3.4'te gösterilmiştir.

Tablo 3.5.

Fakültelerde bilgisayar ders isimleri.

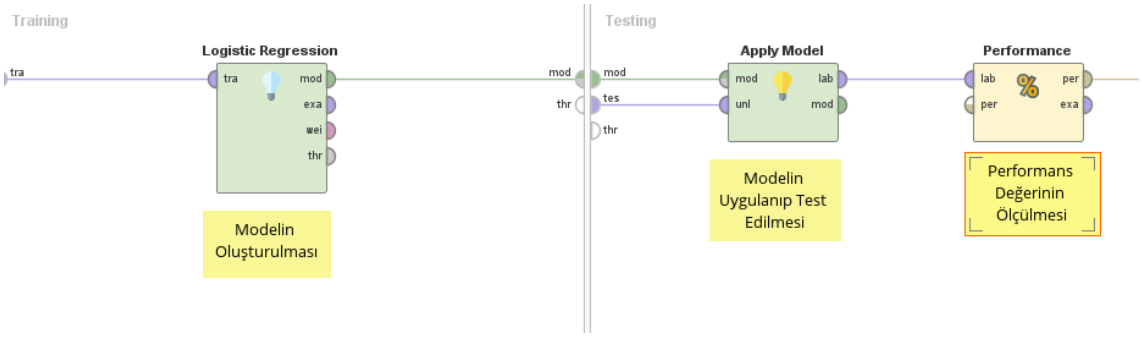
Fakülte	Bölüm	Ders Adı
Beden Eğitimi ve Spor Yüksekokulu	Beden Eğitimi ve Spor Öğretmenliği	Bilgisayar
	Engelliler Egzersiz ve Spor Eğitimi	
Eğitim Fakültesi	Temel Eğitim	Bilişim Teknolojileri
	Matematik ve Fen eğitimi	
	Yabancı Diller Eğitimi	
	Güzel Sanatlar Eğitimi	
	Türkçe ve Sosyal Bilimler Eğitimi	
	Eğitim Bilimleri	
	Özel Eğitim	
Güzel Sanatlar ve Tasarım Fakültesi	Peyzaj Mimarlığı	Temel Bilgi Teknolojileri
	Grafik Tasarım	
	Müzik Bilimleri	
	Resim	
	Seramik	
Spor Bilimleri Fakültesi	Beden Eğitimi ve Spor Eğitimi	Bilgisayar
	Engellilerde Beden ve Spor Eğitimi	Bilişim Teknolojileri
İktisadi ve İdari Bilimler Fakültesi	İktisat	Temel Bilgi Teknolojisi
	Ekonometri	
	Maliye	

Bilgisayar dersinin iki dönemde veren fakülteler Tablo 3.4'te gösterildiği gibidir. Bu fakültelerde yer alan bölümlerin bilişim derslerini hangi isim ile verdiği OBS veri tabanlarından Müfredat Blogna veri tabanı ile Ders Kayıt veri tabanları kullanılarak belirlenmiştir. Bu isimler Bilgisayar I ve Bilgisayar II isimleri ile ortak isimlendirilmiştir. Derslerin belirlenip verilerin çekilmesi ve hazırlanması sonrası Bilgisayar dersi lojistik regresyon tahmin modeli Şekil 3.4'te ki gibi oluşturulmuştur.



Şekil 3.4. Bilgisayar II Dersi Lojistik Regresyon Modeli.

Bilgisayar II dersi kestirim modeli için öncelikle RapidMiner Studio programının işlem bölümüne veri seti eklenmiş ve modelde kullanılacak değişkenler seçilmiştir. Cinsiyet ve medeni durum değişkenlerinin pozitif değerleri belirtilip analiz için değişkenlerin rolleri belirleme aşamasına geçilmiştir. Değişken rolleri Bilgisayar II dersi geçme durumu tahmin değişkeni (bağımlı değişken), Bilgisayar I geçme durumu ve puanı, Cinsiyet, Medeni durum, Yerleşme Puanı, Lise Puanı, İl, Yaş ve Yerleşme Puan Türü değişkenleri bağımsız değişken olarak belirtilmiştir. Özellikle yaş değişkeninin bu tahmin modellerinde önemli bir etkisinin olduğu düşünülmektedir. Sürekli değişen ilk ve ortaöğretim müfredatlarının etkileri yaş değişkeni üzerinden gözlenebileceği düşünülmektedir. Üniversite öncesi eğitimlerin müfredatlarında bilişim derslerine verilen ağırlığın giderek azalmasının etkilerinin bu dersteki başarı üzerinde etkili olabileceği düşünülmüştür. Bununla birlikte teknolojiye yakınlık ve bilgisayar becerilerindeki farklılaşmanın da yaşa bağlı olduğu düşünülmektedir. Bölüm değişkeni modelin öğrenme aşaması için grup belirtici olarak, Öğrenci No değişkeni ise kimlik (ID) olarak belirtilmiştir. Modeli oluşturup test etmek için son adıma çapraz doğrulama operatörüne geçilmiştir.



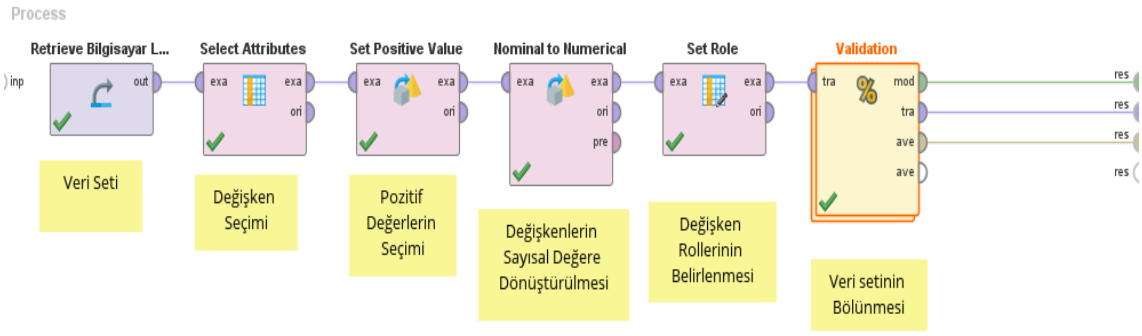
Şekil 3.5. Lojistik Regresyon Çapraz Doğrulama Süreci.

Veri setinin modelde kullanılması için Şekil 3.5'de görüldüğü üzere çapraz doğrulama yöntemi ile veri seti on parçaya ayırılarak dokuzu öğrenmede kullanılmış, kalan bir veri parçası ise test aşamasına gönderilmiştir. Çapraz doğrulama yönteminin öğrenme (Training) aşamasında lojistik regresyon modeli oluşturulmuş, test (Testing) aşamasında ise oluşturulan model test verisine uygulanıp performans değeri ölçülmüştür.

3.4.4.3. Bilişim Dersi Doğrusal Regresyon Modeli

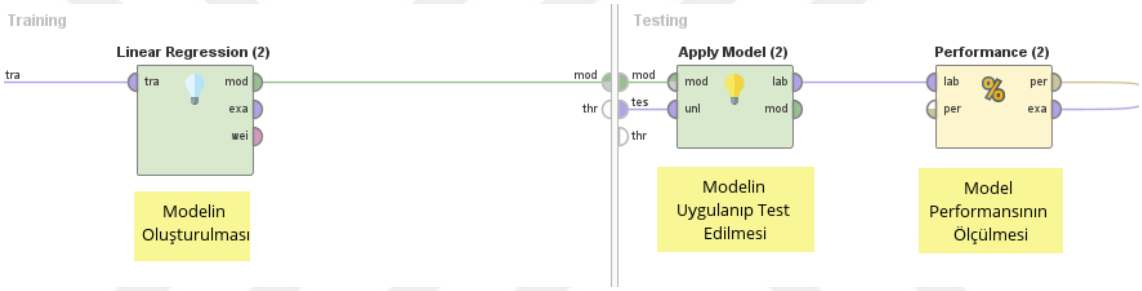
Bilişim dersi doğrusal regresyon kestirim modeli ile üniversitede öğrenim gören öğrencilerin almış oldukları bilişim derslerinin kalitesini ve öğrenci başarısını hedeflenmiştir. Bu model geliştirilerek bilişim dersi hedefleri doğrultusunda dersin kazanımlarının öğrenciye en iyi şekilde verilebileceği düşünülmüştür. Böylece öğrencinin meslek hayatında ve günlük yaşantısında bilişim teknolojilerini en etkili bir biçimde kullanabileceği amaçlanmıştır.

Bu amaç doğrultusunda bilişim dersinin lojistik regresyon modelinde kullanılan veri seti doğrusal regresyon için sayısal verilere dönüştürülmüş ve ders geçme durumu değişkenleri yerine ders ortalama notu değişkenleri veri setine eklenmiştir. Veri setinde öğrenci kişisel bilgilerinden cinsiyet, medeni durum, yaş ve il değişkenleri ile akademik bilgilerinden üniversiteye yerleşme puanı, yerleşme puan türü, lise mezuniyet notu, Bilgisayar I dersi geçme notu bağımsız değişkenleri kullanılarak Bilgisayar II dersi geçme notu kestirilmek istenmiştir.



Şekil 3.6. Bilgisayar II Dersi Doğrusal Regresyon Modeli.

Bilgisayar II dersi doğrusal regresyon modeli Şekil 3.6'da görüldüğü gibi öncelikle RapidMiner Studio programının işlem bölümüne veri seti eklenmiş ve modelde kullanılacak değişkenler seçilmiştir. Cinsiyet, medeni durum ve il değişkenlerinin pozitif değerleri belirtilip analiz için değişken değerleri sayısal değere dönüştürülmüştür. Değişken rolleri belirleme aşamasında; Bilgisayar II geçme notu tahmin değişkeni (bağımlı değişken) olarak, Bilgisayar I geçme notu, Cinsiyet, Medeni durum, Yerleşme Puanı, Lise Puanı, İl, Yaş ve Yerleşme Puan Türü değişkenleri bağımsız değişken olarak belirtilmiştir. Bölüm değişkeni modelin öğrenme aşaması için grup belirtici olarak, Öğrenci No değişkeni ise kimlik (ID) olarak belirtilmiştir. Modelin oluşturulup test edilmesi için Şekil 3.7'de görüldüğü gibi veriyi ayırma (Split Validation) operatörüne geçilmiştir.



Şekil 3.7. Doğrusal Regresyon Modelin Oluşturulup Test Edilme Süreci.

Split Validation (verinin ayrılması) operatörü ile veri seti değişkenlerden aynı oranda ve rastgele olarak iki parça oluşturur. Operatör veri setindeki verilerin %90'ını modelin öğrenme aşamasında, kalan %10'luk parçasını ise modelin test aşamasında kullanılmıştır. Doğrusal (Lineer) regresyon modeli Split Validation operatörünün öğrenme (Training) aşamasında oluşturulmuştur. Operatörün test (Testing) aşamasında ise oluşturulan model test verisine uygulanıp performans değeri ölçülmüştür.

CRISP-DM sürecinin değerlendirme ve uygulama aşaması tezin dördüncü bölümüne uygun görüldüğünde bu bölümlerde ele alınmıştır. Çalışmada elde edilen bilgiler Bulgular bölümünde anlatılarak Sonuç, Tartışma ve Öneriler bölümünde literatürde yapılan diğer çalışmalarla tartışılarak değerlendirilmiştir. Sonuçlar neticesinde gelecek çalışmalar için önerilerde bulunulmuştur.

BÖLÜM IV

4. BULGULAR VE YORUM

Bu bölümde, araştırma süresince elde edilen bulgulara ve yorumlara yer verilmiş CRISP-DM iş süreci ile yapılan eğitsel veri madenciliği tahmin işlemlerine yönelik bilgiler sunulmuştur.

4.1. Birinci Alt Probleme İlişkin Bulgular

Araştırmanın birinci alt problemi "*Öğrencilerin mezun olma süreleri öğrencilerin kişisel bilgileri (Cinsiyet, Medeni Durum, Yaş, İl) ve akademik bilgileri (Üniversiteye Yerleşme Puanı, Yerleşme Puan Türü, Lise Bitirme Puanı, Mezuniyet Notu) ile kestirilebilir mi?"* olarak belirlenmiştir. Mezuniyet süresi tahminine yönelik yapılan lojistik regresyon modelinin performans değeri Şekil 4.1'de gösterilmiştir.

Tablo 4.1.

Mezuniyet Süresi Lojistik Regresyon Performansı.

	Gerçekte Mezun	Gerçekte Uzatmış	Sınıflandırma Hassasiyeti %
Tahmin Mezun	60.169	18.028	76,5
Tahmin Uzatmış	222	278	55,60
Sınıf Hatırlama %	99,63	1,52	

Mezuniyet süresi lojistik regresyon modeli için veri setinde İnönü Üniversitesinin lisans eğitiminden mezun olmuş 78.697 öğrencinin verisi bulunmaktadır. Mezun olan öğrencilerin lisans öğrenim süresi olan 4 yılda (8 dönem) bitiren öğrenciler Mezun olarak belirtilmiş olup, bu sürenin üzerinde bir sürede bitiren öğrenci grubu ise uzatmış olarak belirtilmiştir.

Tablo 4.1 incelendiğinde gerçekte mezun olarak (normal öğrenim süresi içerisinde) bitiren 60.391 öğrencinin 60.169'unu (Başarı oranı: %99,63) modelin başarılı olarak tahmin ettiği görülmüştür. Burada üniversitenin lisans öğrenim süresi (4yıl/8 dönem) içerisinde mezun olanları %99,63 doğruluk ile tahmin edebildiği gözlenmiştir. Gerçekte uzatmış (normal öğrenim süresini aşmış) olarak bitiren 18.028 öğrenciden model 278'ini (Başarı oranı: %1,52) başarılı olarak tahmin etmiştir. Modelin zamanında mezun olanları yüksek doğrulukta tahmin etmesi gözlenmiş ancak mezuniyetini uzatmış öğrencileri düşük bir performansta tahmin ettiği gözlenmiştir. Modelin uzatmış öğrencileri düşük tahmin etmesi olarak örneklem sayısının olduğu tahmin edilmektedir. Lojistik regresyon modelinin genel performans değeri incelendiğinde "(doğru pozitif + doğru negatif) / öğrenci sayısı" formülünden %76,80 oranı ile yüksek bir doğrulukta tahmin yaptığı görülmüştür.

Tablo 4.2.

Lojistik Regresyon Değişken Değerleri.

Bağımsız Değişkenler	Coefficient	P-Value	Std. Coefficent	Std. Error	Z- Value
İli	0,121	0	0,06	0,02	6,38
Cinsiyeti	0,442	0	0,22	0,02	24,07
Medeni Durum	0,061	0,07	0,03	0,02	2,71
Lise Puanı	-0,002	0	-0,10	0,01	-3,71
Yerleşme Puanı	-0,007	0	-0,46	0	-16,4
Yaşı	-0,095	0	-0,53	0	-33,14
Mezuniyet Notu	-1,14	0	-0,61	0,02	-64,14
Intercept	6,435	0	-1,39	0,21	30,14

RapidMiner ile yapılan veri madenciliği çalışmalarında çıktı tablosunda değişken değerleri ve P anlamlılık değerleri önemlidir. Lojistik regresyon modelinde kullanılan bağımsız değişkenlerin bağımlı (tahmin) değişkenine etki değeri (Coefficient) Tablo 4.2'de görülmektedir. Bağımsız değişkenlerin P anlamlılık değerleri incelendiğinde standart ($P<0.05$) değerden düşük olduğu gözlenmiştir. Buda tabloda yer alan her bir değişkenin model için anlamlı bir değer olduğunu göstermektedir. Ancak tabloda yerleşme puan türü değişkeninin olmadığı görülmektedir. Bu değişkenin modelden çıkartılmasının nedeni P anlamlılık değerinin yüksek olmasıdır. Tabloda yer alan her bir

değişkenin model için anlamlı bir değer olduğunu göstermektedir.

Mezuniyet süresi tahmin modelinin hata değerleri ölçüldüğünde Logistic_Loss (Sınıflandırma hatası) değerinin 0.417 olduğu gözlenmiştir. Bu değerin sıfıra yakın olması modelin çok yüksek düzeyde başarılı bir sınıflandırma yaptığını göstermektedir. Modelin Kappa değerinin ise 0,017 gibi çok düşük bir seviyede çıkması modeldeki değişkenlerin bağımlı değişken ile uyumunun çok yüksek seviyede olduğunu göstermektedir.

4.2. İkinci Alt Probleme İlişkin Bulgular

Araştırmanın ikinci alt problemi "*Bilgisayar II dersi geçme durumu öğrencilerin kişisel bilgileri (Cinsiyet, Medeni Durum, Yaş, İl) ve akademik bilgileri (Üniversiteye Yerleşme Puanı, Yerleşme Puan Türü, Lise Bitirme Puanı, Bilgisayar I Dersi Geçme Durumu ve Notu) bilgileri kullanılarak lojistik regresyon ile kestirilebilir mi?"* olarak belirlenmiştir. Bu problem doğrultusunda oluşturulan regresyon modelinin performans sonucu Tablo 4.2'de gösterilmiştir.

Tablo 4.3.

Bilgisayar II Dersi Geçme Durumu Lojistik Regresyon Performansı.

	Gerçekte Geçti	Gerçekte Kaldı	Sınıflandırma Hassasiyeti %
Tahmin Geçti	2.822	635	81.63
Tahmin Kaldı	1.257	4.446	77.96
Sınıf Hatırlama %	69.18	87.50	

Bilgisayar II dersi geçme durumu lojistik regresyon modeli için İnönü Üniversitesinde lisans düzeyinde öğrenim görmüş ya da görmekte olan öğrencilerden Bilişim dersini iki dönem boyunca alan ve kayıp verisi bulunmayan 9.160 öğrenci verisi kullanılmıştır. Bu dersi alan öğrencilerin dersten geçme durumu geçti ve kaldı olarak ele alınmıştır. Tablo 4.3 incelendiğinde gerçekte dersten geçen 4.079 öğrencinin 2.822'sini (Başarı oranı: %69,18) modelin başarılı olarak tahmin ettiği görülmüştür. Bilgisayar II dersinden geçen öğrencilerin tahmin performansının %69,18 olması modelin orta

düzeyde doğrulukta tahmin yaptığını göstermektedir.

Tablo 4.3'te gerçekte dersten kalan 5.081 öğrencinin 4.446'sını (Başarı oranı: %87,50) ise modelin başarılı olarak tahmin ettiği görülmüştür. Dersten kalan öğrenci tahminini modelin %87,50 ile yüksek bir performans oranı ile doğru tahmin etmesi bu dersten başarısız olacak öğrencilerin belirlenmesinde oldukça önemlidir. Bilgisayar II dersi lojistik regresyon modelinin toplam doğru tahmin etme oranı ise "(doğru pozitif + doğru negatif) / öğrenci sayısı" formülünden %79,34 performans oranı ile yükske düzeyde bir tahmin gerçekleştirdiği belirlenmiştir.

Tablo 4.4.

Bilgisayar II Dersi Lojistik Regresyon Değişken Değerleri.

Bağımsız Değişkenler	**Coefficient**	**P Value**	**Std. Coefficent**	**Std. Error**	**Z Value**
ÖSYM Puan Türü TYT	-1,671	0	-1,67	0,41	-4,05
ÖSYM Puan Türü Özel Yetenek	0,211	0,49	0,21	0,31	0,68
ÖSYM Puan Türü Sözel	1,272	0	1,27	0,36	3,51
ÖSYM Puan Türü Dil	2,617	0	2,62	0,36	7,11
ÖSYM Puan Türü Eşit Ağırlık	2,465	0	2,46	0,3	8,24
Cinsiyet	-0,108	0,07	-0,05	0,06	-1,87
Medeni Durum	-0,082	0,25	-0,03	0,07	-1,14
İlk Dönem Dersi Geçme Durumu	0,063	0	0,73	0,09	17,54
İli	-0,063	0,28	0,03	0,06	1,07
Lise Diploma Notu	-0,012	0	-0,46	0	-11,88
Yerleşme Puanı	-0,008	0	-0,46	0	-9,37
İlk Dönem Notu	-0,011	0	-0,27	0	-6,82
Yaşı	0,069	0	0,16	0,01	5,50
Intercept	3,542	0	-1,47	0,54	6,61

Bu modelde kullanılan bağımsız değişkenlerin bağımlı (tahmin) değişkene etki değeri (Coefficient) Tablo 4.4'te görülmektedir. Değişkenlerin P anlamlılık değerleri incelendiğinde çoğu değişkenin standart ($P<0.05$) değerinden düşük olduğu görülmüş ancak ÖSYM puan türü Özel yetenek olan, Medeni durumu Bekar olan ve Lise Diploma

Notu değişkenlerinin anlamlılık değerlerinin yüksek olduğu görülmüştür.

P anlamlılık değerleri standart değere eşit ya da bu değerden düşük olan değişkenler bağımlı değişken üzerinde etki gösterdiğini ve bu değişkenlerin model için anlamlı olduğunu göstermektedir. Ancak P anlamlılık değeri yüksek çıkan ÖSYM puan türü Özel yetenek olan, Medeni durumu Bekar olan ve Lise Diploma Notu değişkenlerinin model için anlamsız olduğu ve modelden çıkartılması gerektiği sonucuna varılmıştır.

Modelde kullanılacak değişkenlerden birkaçının P anlamlılık değerlerinin standart değerden yüksek olması bağımlı değişken üzerinde anlamlı bir etkiye sahip olmadıklarını ifade etmektedir. Bu nedenle bu değişkenleri RapidMiner Studio yazılımı modelden çıkartarak hesaplamakta ve böylece model daha güvenilir olmaktadır.

Bilgisayar II dersi geçme durumu tahmin modelinin hata değerleri ölçüldüğünde Logistic_Loss (Sınıflandırma hatası) değerinin 0.404 olduğu gözlenmiştir. Buda modelin çok yüksek düzeyde başarılı bir sınıflandırma yaptığını göstermektedir. Modelin Kappa değerinin ise 0,57 gibi çok düşük bir seviyede çıkması modeldeki değişkenlerin bağımlı değişken ile uyumunun çok yüksek seviyede olduğunu göstermektedir.

4.3. Üçüncü Alt Probleme İlişkin Bulgular

Araştırmanın üçüncü alt problemi "*Bilgisayar II dersi geçme notu öğrencilerin kişisel bilgileri (Cinsiyet, Medeni Durum, Yaş, İl) ve akademik bilgileri (Üniversiteye Yerleşme Puanı, Yerleşme Puan Türü, Lise Bitirme Puanı, Bilgisayar I Dersi Geçme Durumu ve Notu) kullanılarak doğrusal regresyon ile kestirilebilir mi?"* olarak belirlenmiştir. Bu problem doğrultusunda oluşturulan doğrusal regresyon modeline ilişkin sonuçlar Şekil 4.3'te gösterilmiştir.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
Cinsiyeti	3.519	0.551	0.049	0.988	6.383	0.000	****
ÖSYM Puan Türü	-8.682	0.172	-0.398	0.922	-50.501	0	****
İli	3.073	0.540	0.044	0.966	5.692	0.000	****
Lise Diploma Notu	0.157	0.009	0.176	0.823	18.012	0	****
Yerleşme Puanı	0.078	0.006	0.130	0.802	13.255	0	****
İlk Dönem Notu	0.378	0.012	0.278	0.764	32.144	0	****
Yaşı	-0.863	0.115	-0.059	0.962	-7.525	0.000	****
(Intercept)	-7.959	3.837	?	?	-2.075	0.038	**

Şekil 4.1. Bilgisayar II Doğrusal Regresyon Modeli.

Doğrusal regresyon modelleri sonuçları incelenirken RapidMiner Studio yazılımı çalışmacılara kolaylık sağlamaktadır. Bu kolaylık çıktı bölümünde en solda yer alan "Code" sütunudur. Bu sütün oluşturulan model için bağımlı değişkenin hangi değişkenler tarafından ne düzeyde etkilendiklerini göstermektedir. "*" sembolü ile gösterilen bu değer eğer dört tane ise çok yüksek düzeyde, üç tane ise yüksek düzeyde, iki tane ise orta düzeyde, bir tane ise düşük düzeyde bağımlı değişkeni etkilediğini göstermektedir. Eğer "Code" bölümünde hiçbir değer yok ise bu bağımsız değişkenin bağımlı değişkeni hiç etkilemediğini göstermektedir.

Doğrusal regresyon modelinde bağımlı (tahmin) değişkeni etkileyen bağımsız değişkenler Şekil 4.3'te gösterilmiştir. Bağımsız değişkenlerin P anlamlılık değerleri sıfır ve sıfıra çok yakın olduğu görülmüştür. Ancak medeni durum değişkeninin modelden atıldığı gözlenmiştir. Bu da bu değişkenin modelde anlamlı bir etkisinin olmadığını göstermektedir. Değişkenlerin "Code" sembolleri incelendiğinde bütün değişkenlerin yüksek düzeyde bağımlı değişkene etki ettiği görülmüştür. Bağımsız değişken değerleri (Coefficient) kullanılarak doğrusal regresyon modeli aşağıda oluşturulmuştur.

Y= (3,519*Cinsiyet) + (-8,682*ÖSYM Puan Türü) + (3,073*İli) + (0,157*Lise Diploma Notu) + (0,078*Yerleşme Puanı) + (0,378*İlk Dönem Notu) + (-0,863*Yaşı) + (-7.959)

Geliştirilen Bilgisayar II dersi geçme notu modelinin tahmin performansının R^2=0,527 olduğu gözlenmiştir. R^2 değerinin 0,30'dan yüksek olması regresyon modelinin geçerli olduğunu göstermektedir. Modelin tahmin değerleri ile gerçek değerler arasındaki farkın ise RMSE=23,18 olduğu gözlenmiştir.

BÖLÜM V

5. SONUÇ, TARTIŞMA VE ÖNERİLER

Yapılan çalışmada geliştirilen modellerden elde edilen sonuçlar literatürde yer alan çalışma sonuçları ile karşılaştırılarak açıklanmış ve ileri çalışmalar için öneriler yapılmıştır. Bu kısım aynı zamanda veri madenciliği süreç tasarımlarından CRISP-DM iş sürecinin değerlendirme ve yayılım süreci olarak ele alınmıştır.

5.1. Sonuçlar ve Tartışma

Bu çalışmada yükseköğretimde öğrenim gören öğrencilerin kişisel ve akademik verileri kullanılarak üniversiteden mezun olma durumu ve Bilgisayar II dersi geçme durumu ve notu tahmin modelleri geliştirilmiştir. Aşağıda, elde edilen sonuçlar doğrultusunda modellerin performansı yorumlanmış, literatürdeki benzer ve farklı durumlar tartışılmıştır.

Çalışmanın birinci alt problemi doğrultusunda öğrencilerin mezun olma sürelerini etkileyen kişisel ve akademik değişkenlerin lojistik regresyon modelindeki performans değerleri belirlenmiştir. Otomasyon bilgi sisteminden lisans eğitim düzeyinde mezun olmuş 78.697 öğrenci verisi kullanılarak öğrencilerin mezun olma durumları tahmin edilmiştir. Üniversiteden lisans eğitim süresinde (4 yıl) ve öncesinde mezun olan kişiler "mezun" olarak sınıflandırılmış, lisans eğitim süresini aşan ya da ayrılan öğrenciler "uzatmış" şeklinde sınıflandırılmıştır. Oluşturulan lojistik regresyon modeli incelendiğinde cinsiyet (0,442), il (0,121), medeni durum (0,061), lise puanı (-0,002), yerleşme puanı (-0,007), yaşı (-0,095) ve mezuniyet notu (-1,140) katsayılarının bağımlı değişken olan mezun olma süresi üzerinde etkilerinin olduğu gözlenmiştir. Ayrıca P anlamlılık değerlerinin 0,05'ten küçük olması ile tüm değişkenlerin modelde anlamlı ölçüde etki gösterdikleri tespit edilmiştir. Modelin sınıflandırma hatası (logistic_loss) değerinin 0.417 olması modelin sınıflandırmada yaptığı hata payının çok düşük olduğunu göstermektedir.

Kappa değeri kontrol edildiğinde 0.017 gibi çok düşük bir değer çıkması modeldeki değişkenlerin mezun olma süresine çok yüksek seviyede uyum gösterdiğini belirtmektedir. Lojistik regresyon modelinin tahmin performans değeri incelendiğinde %76,80 olduğu görülmüş ve bu değer ile modelin mezun olma süresi tahmininde yüksek performans gösterdiği sonucuna ulaşılmıştır. Regresyon analizi tahmine yönelik bir model olsa da lojistik regresyon sınıflandırıcı özellik gösteren bir tahmin modelidir. Bu nedenle sınıflandırma modelleri ile birlikte incelenmesi daha sağlıklı olacaktır. Önceki çalışmalar incelendiğinde EVM alanında öğrencilerin mezun olma durumlarının tahmin etmede en çok sınıflandırma algoritmalarının tercih edildiği görülmüş ve bu algoritmalardan daha yüksek performans alındığı gözlenmiştir. Kayhan (2019) ve Şengür'ün (2013) çalışmaları incelendiğinde mezuniyet süresi tahmininde sınıflandırma algoritmalarının daha yüksek performans gösterdiği belirtilmiştir. Mezuniyet tahminine yönelik ilk yıl öğrenim verileri ile yapılan çalışmalarda performans değerlerinin %70-80 aralığında olduğu görülmüştür. Örneğin Polat (2021) geliştirdiği model ile öğrencilerin mezun olma durumlarını sınıflandırma algoritmalarından J48 algoritması ile %80,47 olarak tahmin etmiştir. Berens ve diğerleri (2019) geliştirdikleri regresyon, yapay sinir ağı ve AdaBoost modelleri ile öğrencilerin okulu bırakma ve mezun olma durumu tahminini, ilk dönem verileri ile %79, dördüncü dönem sonu verileri ile %90 oranında başarı ile tahmin etmişlerdir. Çalışmanın geliştirilen modelin performansı da yukarıda bahsedilen çalışmalardaki başarı oranlarına yakın gerçekleşmiştir.

Çalışmanın ikinci alt problemi kapsamında öğrencilerin Bilgisayar II dersinden geçme durumlarının tahminine yönelik öğrencilerin kişisel ve akademik bilgileri kullanılarak lojistik regresyon modeli geliştirilmiştir. Otomasyon bilgi sisteminde lisans düzeyinde bilişim dersi alan farklı fakülte ve bölümlerde öğrenim gören 9.160 öğrenci verisi kullanılmıştır. Farklı fakülte ve bölümlerde farklı isimlerde verilen ancak içeriğinin temel bilişim dersi olduğu dersler analiz için dönemlik olarak Bilgisayar I ve Bilgisayar II olarak ele alınmıştır. Bilgisayar II dersini alan öğrencilerden dersi geçen kişiler "geçti" olarak sınıflandırılmış olup, dersten başarısız sayılan ya da devamsızlıktan kalan öğrenciler "kaldı" şeklinde gruplandırılmıştır. Oluşturulan lojistik regresyon modeli incelendiğinde cinsiyet (-0.108), Bilgisayar I geçme durumu (1,558), lise puanı (-0,012), yerleşme puanı (-0,08), yaşı (0,069) ve ÖSYM puan türü TYT (-1,571) Sözel (1,272) Dil (2,617) Eşit Ağırlık (2.465) değişkenlerinin P anlamlılık değerleri standart değer olan

0,05'ten küçük olduğu için bu değişkenler modelde kullanılmış ve bu değişkenlerin Bilgisayar II dersi geçme durumunu etkiledikleri görülmüştür. Bağımsız değişkenlerden medeni durumu, ili ve ÖSYM puan türü Özel Yetenek olan değişkenlerin P anlamlılık değeri standardın üzerinde olması nedeni ile bu değişkenler modelden çıkartılmıştır. Modelin sınıflandırma hatası (logistic_loss) değerinin 0,404 olması, modelin sınıflandırma hata miktarının çok düşük olduğunu göstermektedir. Kappa değerinin 0,57 olması ise modeldeki değişkenlerin Bilgisayar II dersi ile yüksek düzeyde uyum sağladığını göstermektedir. Geliştirilen lojistik regresyon modelinin performans değeri %79,34 olarak ölçülmüş ve böylece modelin Bilgisayar II dersi geçme durumunu tahmin etmede yüksek başarı gösterdiği tespit edilmiştir. Geliştirilen modelin başarı oranının yüksek olduğu sonucuna EVM ile bir derse yönelik geçme durumu tahmin modelleri karşılaştırıldığında varılmıştır. Bu çalışmalara örnek olarak; Uğuz, Şahin ve Yılmaz (2021) Fen Bilimleri dersinin puanını tahmin etmede yaptıkları çalışmada sınıflandırma algoritmalarından k-NN ile %77 Naive Bayes ile %55,06 Random Forest yöntemi ile %62,22 olarak tahmin etmesi, Karataşcı'nın (2021) Matematik ve Fen Bilimleri derslerinin kazanımlarını tahmin için geliştirdiği modelinde %46-58 arasında tahmin etmesi ve Aksu'nun (2018) yaptığı çalışmada fen okuryazarlığı tahmininde geliştirdiği lojistik regresyon modelinin %72,35 ile performans göstermesi örnek olarak gösterilebilir. Yapılan sınıflandırma çalışmalarında belirtilen en iyi sonucu sınıflandırma algoritmalarının gösterdiği şeklindeki genel kanı yaptığımız çalışma sonucunda doğru olmadığı belirlenmiştir. Tahmin çalışmalarında gösterilen performans değerinin yüksek olduğu belirlenmiş ve EVM çalışmalarında modellerin kıyaslanmasından ziyade hedefe yönelik en iyi model seçimi yapılması gerektiği belirlenmiştir (Yu ve diğerleri 2020; Berens ve diğerleri, 2019; Altun, 2019).

Çalışmanın üçüncü alt problemi kapsamında öğrencilerin Bilgisayar II dersinden geçme notlarının tahminine yönelik öğrencilerin kişisel ve akademik bilgileri kullanılarak doğrusal (lineer) regresyon modeli geliştirilmiştir. Otomasyon bilgi sisteminde lisans düzeyinde bilişim dersi alan farklı fakülte ve bölümlerde öğrenim gören 9.160 öğrenci verisi kullanılmıştır. Farklı fakülte ve bölümlerde farklı isimlerde verilen ancak içeriğinin temel bilişim dersi olduğu dersler analiz için dönemlik olarak Bilgisayar I ve Bilgisayar II olarak ele alınmıştır. Bilgisayar II dersinin öğrenci notu hesaplanmasında dersten geçme durumları üniversitenin ders yönetmeliği gereği bağıl değerlendirme ile

hesaplanmaktadır. Ancak bir dersin kazanımlarının ne kadarının öğrenci tarafından kazanıldığı dersin geçme notu ile belirlenebilir. Bu yüzden öğretim üyeleri ya da öğretmenler derslerindeki gerekli kazanımların kazanılıp kazanılmadığını yaptıkları sınavlar ile ölçerler. Bu nedenle geliştirilen doğrusal regresyon modeli Bilgisayar dersini veren eğitimcilere dönemin başında öğrencilerin hangi aralıkta bir performans göstereceği hakkında bilgi sahibi olmasına imkân tanıyacağı düşünülmektedir. Oluşturulan doğrusal regresyon modeli incelendiğinde cinsiyet (3,519), Bilgisayar I geçme notu (0,378), lise puanı (0,157), yerleşme puanı (0,078), yaşı (0.863) ve ÖSYM puan türü (-8,682) il (3,073) değişkenlerinin P anlamlılık değerleri standart değer olan 0,05'ten küçük olduğu için bu değişkenler modelde kullanılmış ve bu değişkenlerin Bilgisayar II dersi geçme notunu etkiledikleri görülmüştür. Bağımsız değişkenlerden medeni durum değişkenin P anlamlılık değeri standardın üzerinde olması nedeni ile bu değişken modelden çıkartılmıştır. Doğrusal regresyon modelinin performans değeri olarak RMSE değeri yani modelin tahmin performans hatası değerine bakılır. RMSE değeri sıfır ila sonsuz arasında bir değer alabilir. Burada bu değerin sıfıra yakın olması modelin çok yüksek performans gösterdiği anlamına gelmektedir. Modelin RMSE değeri ölçüldüğünde 23,18 gibi bir değerin çıkması, modelin tahmin hata miktarının çok düşük olduğunu göstermektedir. Böylece modelin Bilgisayar II dersi geçme notunu tahmin etmede yüksek başarı gösterdiği tespit edilmiştir. Geliştirilen modelin tahmin başarısının yüksek olduğu sonucuna Hassana ve Al-Razgan'ın (2016) çalışmasında ulaştığı sonuç ile Devasia, Vinushree ve Hedge'nin (2016) yaptıkları çalışmadaki regresyon sonucunun, yapılan çalışma ile benzer çıkması ile belirlenmiştir. EVM ile bir derse yönelik geçme notu tahmin modelleri geliştirilerek geleceğe yönelik gerekli önlemler alınabilir ve ders başarı ve kazanımlarının arttırılması sağlanabilir (Çetintav ve diğerleri, 2022; Demiral ve diğerleri, 2017; Roberts ve diğerleri, 2016).

Alan yazındaki çalışmalar ile yapılan karşılaştırmalar sonucunda EVM modellerinde genel olarak başarı performans aralığı %70-80 olarak tespit edilmiştir (Olgun, 2021; Benens vd., 2019; Altun, 2019; Aksu, 2018; Yükseltürk, Özekeş ve Türel, 2014). Ancak bazı çalışmalarda %90 ile çok yüksek düzeyde başarı sağlandığı görülmüştür (Keser, 2021; Altun, Kayıkçı ve Irmak, 2019; Benens vd., 2019). Bunun nedeninin ise veri setindeki verilerin tam olmasından kaynaklandığı düşünülmektedir. Araştırmada elde ettiğimiz performans değerlerinin alan yazındaki çoğu çalışma ile benzer olduğu ve yeterli performans oranına sahip olduğunu göstermektedir. Buda

oluşturulan modeller ile başarılı tahminler yapılabileceğini göstermektedir (Topuz, 2021; Demiral ve diğerleri, 2017). EVM çalışmalarında Aksu (2018), Kayhan (2019), Berens ve diğerlerinin (2019) ve Topuz (2021) yaptıkları araştırmalarda olduğu gibi farklı modeller kullanarak sınıflandırma ve tahmin yöntemlerinde daha iyi modelin belirlenmesi çalışmalarının gerçek veriler üzerinde yapılan yeni çalışmalara rehber olacak nitelikte olduğu düşünülmektedir. Böylece gerçek veriler kullanılarak yapılan çalışmaların hedeflenen eğitimin kalitesinin arttırılmasına yönelik olması daha faydalı olacaktır (Altun, Kayıkçı ve Irmak, 2019; Demiral ve diğerleri, 2017; Roberts ve diğerleri, 2016).

Bu araştırma kapsamında EVM süreç tasarımlarından CRIPS-DM iş sürecinin izlenmesi ile İnönü Üniversitesinde öğrenim gömüş ve görmekte olan öğrenci verileri kullanılarak mezuniyet durumunun tespiti, Bilgisayar II dersinin geçme durumunun tespiti ve geçme notu tahmini yapılmıştır. Ortaya çıkan sonuçlar öğrenci akademik performansının izlenmesi ve gerekli önlemlerin alınabileceği konusunda umut verici olmakta ve bu alanda daha fazla çalışmanın yapılması gerekliliğini ortaya çıkarmaktadır.

5.2. Öneriler

Araştırma sonucunda oluşturulan modeller İnönü Üniversitesi öğrenci akademik performansını kestirmede, başarısızlığı düşecek öğrencileri ve geç mezun olacak öğrencilerin tespitinde kullanılabilir. Böylelikle üniversite yönetiminin ve öğretim üye ve elemanlarının gerekli önlemleri alması sağlanabilir ve akademik başarı arttırılabilir.

Eğitimin kalitesinin arttırılmasına yönelik yapılacak çalışmalarda eğitsel veri madenciliği alanın seçilmesi bu hedefe ulaşılmada daha hızlı ve daha doğru sonuçlara ulaşılmasını sağlayabilir. Çünkü bu çalışmalarda mevcut bilgi kullanılarak geleceğe yönelik durumların tespiti ve tahmini gerçekleştirilebilir. Bunu şöyle örneklendirmek daha iyi olacaktır. 2020 yılında dünya genelinde baş gösteren Covid-19 salgınının her alanda olduğu gibi eğitimi de aksattığı görülmüş ve bu alanda gerekli önlemlerin zamanında alınamamasına neden olmuştur. Bunun birçok nedeni olabilir ancak büyük veri ve EVM alanında yapılacak çalışmalar ile öğrencilerin geçmiş bilgileri kullanılarak gerçekçi, ekonomik ve doğru bir eğitim modeli seçilebilir. Bu nedenle eğitim alanında büyük veri ve veri madenciliği çalışmalarının arttırılması gerekmektedir.

Eğitim alanında veri madenciliği çalışmalarının yapılabilmesi için kurumların veriye erişmede zorluklar çıkartmaması ve mevcut verilerini daha düzenli ve sistematik bir biçimde veri tabanlarında muhafaza etmesi gerekmektedir. Yapılan çalışmada en çok karşılaştığımız sorun olan eksi veri ile geliştirilmek istenen modellerin veri setleri küçülmekte ve buda modellerin doğru tahmin oranlarını olumsuz yönde etkilemektedir. Bu nedenle literatürde incelenen çalışmalarda da bahsedildiği üzere eğitim kurumlarında özellikle yükseköğretimde veri bilimi, büyük veri ve veri tabanı sistemleri dersleri tüm öğrenci ve personellere verilmelidir. Böylece eğitimin teknoloji ile olan bağı daha da güçlendirilerek eğitimin kalitesinin arttırılabileceği düşünülmektedir.

KAYNAKÇA

Akgöbek, Ö. ve Çakır, F. (2009). Veri madenciliğinde uzman bir sistem tasarımı. *Akademik bilişim konferansları*. Şanlıurfa.

Akgün, K., ve Özek, M.B. (2020). Eğitsel Veri Madenciliği Yöntemi ile İlgili Yapılmış Çalışmaların İncelenmesi: İçerik Analizi. *Uluslararası Eğitim Bilim ve Teknoloji Dergisi*, 6 (3), 197-213.

Akpınar, H. (2014). *Data: Veri Madenciliği Veri Analizi*. (2. Basım) Papatya Yayıncılık Eğitim.

Akçapınar, G. (2014). Çevrimiçi öğrenme ortamındaki etkileşim verilerine göre öğrencilerin akademik performanslarının veri madenciliği yaklaşımı ile modellenmesi. Doktora Tezi, *Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Dergisi*, Ankara.

Aksu, G. (2018). PISA başarısını tahmin etmede kullanılan veri madenciliği yöntemlerinin incelenmesi. Doktora tezi, *Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Dergisi*, Ankara.

Aktan, E. (2018). Büyük veri: Uygulama alanları, analitiği ve güvenlik boyutu. *Bilgi Yönetimi Dergisi*, 1(1), 1-22.

Aldowah, H., Al-Samarraie, H. & Fauzy, W. M. (2019). Telematics and Informatics. https://doi.org/10.1016/j.tele.2019.01.007

Altun, M. (2019). Öğrenci akademik performansının kestirimine ilişkin bir model önerisi: Veri madenciliğine dayalı bir çalışma. Doktora Tezi, *Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü*, Antalya.

Altun, M., Kayıkçı, K., ve Irmak, S. (2019). Sınıf Öğretmenliği Öğrencilerinin Mezuniyet Notlarının Regresyon Analizi ve Yapay Sinir Ağları Yöntemleriyle Tahmini. *Uluslararası Eğitim Araştırmaları Dergisi*, 10(3), 29-43.

Alsuwaiket, M. (2018). Measuring academic performance of students in higher education using data mining techniques. Doctoral dissertation, *Available from ProQuest Dissertations & Theses Global*.

Argüden, Y. ve Erşahin, B. (2008). *Veri madenciliği veriden bilgiye masraftan değere.* ARGE Danışmanlık.

Asif, R., Merceron, A., Ali, S.A., Haider, N.G., (2017). Analyzing undergraduate students' performance using educational data mining. *Computers and Education* 113, 177–194. https://doi.org/10.1016/j.compedu.2017.05.007

Atan, S. (2016). Veri, Büyük Veri ve İşletmecilik. *Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 19(35).

Ayas, A., Haluk, Ö., Çalık, M., Çimer, A., Ekiz, D., Yiğit, N., ve İskurt, E. (2014). *Eğitim bilimine giriş*. Pegem Akademi.

Aydın, S. (2007). Veri madenciliği ve Anadolu Üniversitesi uzaktan eğitim sisteminde bir uygulama. Doktora Tezi, *Anadolu Üniversitesi Sosyal Bilimler Enstitüsü*, Eskişehir.

Aydın, S., ve Özkul, A. E. (2015). Veri madenciliği ve Anadolu Üniversitesi Açık Öğretim Sisteminde Bir Uygulama. *Eğitim ve Öğretim Araştırmaları Dergisi*, 4(3), 36-44.

Bahadir, E. (2013). Öğretmen Adaylarının Akademik Başarılarının Sınıflandırılmasında Lojistik Regresyon Analizi Yaklaşım. *Journal of Turkish Studies*. 8. 203-203. 10.7827/TurkishStudies.5397.

Bahçeci, F. (2015). Öğrenme Yönetim Sistemlerinde Kullanılan Öğrenme Analitikleri Araçlarının İncelenmesi*, Türkiye Eğitim Araştırmaları Dergisi*, 2, 1, 43-44.

Baltacı, A. (2018). Nitel araştırmalarda örnekleme yöntemleri ve örnek hacmi sorunsalı üzerine kavramsal bir inceleme. *Bitlis Eren Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 7(1), 231-274.

Başkale, H. (2016). Nitel araştırmalarda geçerlik, güvenirlik ve örneklem büyüklüğünün belirlenmesi. *DSpace Pamukkale Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 9(1), 23-28.

Baştürk, S. ve Taştepe, M. (2013). *Bilimsel Araştırma Yöntemleri: Evren ve Örneklem.* (2. Baskı). Vize Yayıncılık.

Berens, J., Schneider, K., Görtz, S., Oster, S., & Burghoff, J. (2018). Early detection of students at risk–predicting student dropouts using administrative student data and machine learning methods. *CESifo Working*. (7259).

Bezerra, L. N., & Silva, M. T. (2020). Educational Data Mining Applied to a Massive Course. *International Journal of Distance Education Technologies (IJDET),* 18(4), 17-30. http://doi.org/10.4018/IJDET.2020100102

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4-16.

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *ResearchGate,* https://www.researchgate.net/publication/308067314_Enhancing_teaching_and_learning_through_educational_data_mining_and_learning_analytics_An_issue_brief.

Booth, M. (2012). Learning analytics: The new black. *EDUCAUSE Review*, 47(4), 52-53.

Bozkurt, A. (2016). Öğrenme analitiği: e-öğrenme, büyük veri ve bireyselleştirilmiş öğrenme. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, 2(4), 55-81.

Bulut, O. ve Yavuz, H. C. (2019). Educational data mining: A tutorial for the rattle package in R. *International Journal of Assessment Tools in Education*,20-36. https://dx.doi.org/10.21449/ijate.627361.

Burma, A, Z. (2009). *Veri Tabanı Yönetim Sistemleri ve SQL/PL-SQL/T-SQL.* (1. Baskı) Seçkin Yayınevi.

Büyüköztürk, Ş., Kılıç-Çakmak, E., Akgün, Ö., Karadeniz, Ş., ve Demirel, F. (2013). *Bilimsel Araştırma Yöntemleri.* Ankara, Pegem Akademi.

Can, M.B., Eren, Ç., Koru, M., Özkan, Ö. ve Rzayeva, Z. (2012). *Veri Kümelerinden Bilgi Keşfi: Veri Madenciliği*, Başkent Üniversitesi Tıp Fakültesi XIV. Öğrenci Sempozyumunda sunuldu, Ankara.

Can, Ş. (2017). Veri Madenciliği ve Eğitim Sektöründe Bir Uygulama, Yüksek Lisans Tezi, Celal Bayar Üniversitesi, Sosyal Bilimler Enstitüsü, Manisa.

Çalışkan, S. K., ve Soğukpınar, İ. (2008). *KxKNN: K-Means ve K En Yakin Komşu Yöntemleri ile Ağlarda Nüfuz Tespiti.* EMO Yayınları, 120-24.

Çetintav, G., Çil, B. D., & Yılmaz, R. (2022). Eğitsel Veri Madenciliği ve Öğrenme Analitikleri Araştırmalarında Veri Gizliliği ve Etik Meseleler: Araştırmalar Üzerine Bir İnceleme. *Eğitim Teknolojisi Kuram ve Uygulama Dergisi*, 12(1), 113-146.

Chung, J. Y., & Lee, S. (2019). Dropout early warning systems for high school students using machine learning. *Children and Youth Services Review*, 96, 346-353.

Clayton, M., & Halliday, D. (2017). Big data and the liberal conception of education. *Theory and Research in Education*, 15(3), 290–305. https://doi.org/10.1177/1477878517734450

Cunningham, J. A. (2017). Predicting student success in a self-paced mathematics MOOC. Doctoral dissertation. USA: Arizona State University. *Available from ProQuest Dissertations & Theses Global*.

Czyzewska, M. & Mroczek, T. (2020). Data Mining in Entrepreneurial Competencies Diagnosis. *Education Sciences*, 10(8), 196.

Daniel, B. (2015). Big Data and analytics in higher education: Opportunities and challenges. *British journal of educational technology*, 46(5), 904-920.

Dalkılıç, F. ve Aydın, Ö. (2017). Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Öğrencilerinin Devamsızlık Davranışlarını Etkileyen Faktörler. *Yükseköğretim ve Bilim Dergisi,* (3), 546-553.

Demiral, G., Soba, M., ve Armutlu, Ş. (2017). Kütüphane veri tabanında veri madenciliği: Uşak Üniversitesi örneği. *Bartın Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 8(16),241-264.

Demirtaş, B. ve Argan, M. (2015). Büyük Veri ve Pazarlamadaki Dönüşüm: Kuramsal Bir Yaklaşım. *Pazarlama ve Pazarlama Araştırmaları Dergisi*, 8(15), 1-22.

Devasia, T., Vinushree, T. & Hedge, V. (2016). Prediction of students performance using educational data mining. International Conference on Data Mining and Advanced Computing. *SAPIENCE*, s. 91-95.

Davenport, T. H., & Dyche, J. (2013). Big data in big companies. *International Institute for Analytics*, 3(1-31).

Diebold, F. X. (2003). Big data dynamic factor models for macroeconomic measurement and forecasting. *In Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress of the Econometric Society*, 115-122.

Diler, S. (2016). *Veri madenciliği süreçleri ve karar ağaçları algoritmaları ile bir uygulama.* Yayınlanmamış Yüksek Lisans Tezi, Van Yüzüncü Yıl Üniversitesi, Van.

Doğan, K., ve Arslantekin, S. (2016). Büyük Veri: Önemi, Yapısı ve Günümüzdeki Durum. *Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi*, 56(1).

Ekim, U. (2011). Veri madenciliği algoritmalarını kullanarak öğrenci verilerinden birliktelik kurallarının çıkarılması, Yüksek lisans tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü.

Erşahin, B., (2008). *Veri madenciliği; veriden bilgiye, masraftan değere.* ARGE Danışmanlık, İstanbul. http://www.arge.com/wpcontent/uploads/2013/02/VeriMadenciligi.pdf,

Fiofanova, O.A. (2021), Eğitimde büyük veri yönetimi. Kamu Yönetimi. *Gosudarstvennaya Sluzhba*. 10.22394 /2070-8378

Gamgam, H. ve Altunkaynak, B. (2015). *SPSS uygulamalı regresyon analizi*. Seçkin Yayınevi, Ankara.

Gartner Group, (2013), IT Glossary. Gartner: http://www.gartner.com/it-glossary/data-mining

Güvendir, M. A. (2014). Öğrenci başarılarının belirlenmesi sınavında öğrenci ve okul özelliklerinin Türkçe başarısı ile ilişkisi. *Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi*, 39(172).

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Orallo, J. H., Kull, M., Lachiche, N. & Flach, P. A. (2019). CRISP-DM twenty years later: From data mining processes to data science trajectories. *IEEE Transactions on Knowledge and Data Engineering*. 10.1109/TKDE.2019.2962680.

Mills, G. E. & Gay, L. R. (2019). *Educational research: Competencies for analysis and applications*. Pearson. One Lake Street, Upper Saddle River, New Jersey 07458.

Gupta, G. K. (2014), *Introduction to Data Mining with Case Studies*. Delhi: PHI Learning Pvt. Ltd.

Grimes, S. (2005). Structure, Models and Meaning. InformationWeek. http://informationweek.com/software/business-intelligence/structure-models-and-meaning/59301538

Gürsakal, N. (2014). *Büyük Veri*. Dora Yayınları, Bursa.

Hamsa, H., Indiradevi, S. & Kizhakkethottam, J. J. (2016). Student academic performance prediction model using decision tree and fuzzy genetic algorithm, Procedia Technology. Pr*ocedia Teknology*.

Hassana, S. M. & Al-Razgan, M. S. (2016). Pre-University Exams Effect on Students GPA: A case Study in IT Department, Procedia Computer Science, 82, 127-131.

Hofmann, M., & Klinkenberg, R. (2016). *RapidMiner: Data mining use cases and business analytics applications.* CRC Press.

Hussain, S., Dahan, N. A., Ba-Alwib, F. M., & Ribata, N. (2018). Educational data mining and analysis of students' academic performance using WEKA. *Indonesian Journal of Electrical Engineering and Computer Science*, 9(2), 447-459. 10.11591/ijeecs.v9.i2.pp447-459.

Iam-On, N. & Boongoen, T. (2017). Generating descriptive model for student dropout: a review of clustering approach. *Human-centric Computing and Information Sciences*, 7(1), 1-24.

Jacobi, F., Jahn, S., Krawatzeck, R., Dinter, B., & Lorenz, A. (2014). Towards a design model for interdisciplinary information systems curriculum development, as exemplified by Big data analytics education. *European Conference on Information Systems*, Israel.

Jacobs, A. (2009). The pathologies of big data. *Communications of the ACM*, 52(8), 36-44.

Jaiswal, M. (2018). Big Data concept and imposts in business. *Manishaben Jaiswal'Big Data Concept and Imposts in Business' International Journal of Advanced and Innovative Research (IJAIR)* ISSN, 2278-7844.

Jalota, C. & Agrawal, R. (2019). Analysis of Educational Data Mining using Classification, *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon),* 243-247, 10.1109/COMITCon.2019.8862214.

Januszewski, A., & Molenda, M. (2013*). Educational technology: A definition with commentary.* (Two edition) Association for Educational Communications and Technology (AECT).

Jeffery, K. (2014). Data is the New Oil. *Best Practices for Data Management & Sharing*. Joint Research Centre (JRC). Ispra, Italy.

Jiang, Y. H., Javaad, S. S. & Golab, L. (2015). Data Mining of Undergraduate Course Evaluations. *Informatics in Education*, (1) 85-102. 10.15388/infedu.2016.05.

Johnson, L., Adams, S., & Cummins, M. (2012). The NMC horizon report: 2012 higher Education edition. Austin, *TX: New Media Consortium*.

Karasar, N. (2008). *Bilimsel Araştırma Yöntemi*, Nobel Yayın Dağıtım Ltd.

Karataşcı, M. (2021). Sayısal derslerde kazanımlara erişim düzeylerinin veri madenciliği ile analizi ve ortaokul öğrencileri üzerine bir uygulama. Yüksek Lisans Tezi, Sivas Cumhuriyet Üniversitesi, Sosyal Bilimler Enstitüsü, Sivas.

Kaya, G., ve Usluel, Y. K. (2011). Öğrenme-öğretme süreçlerinde BİT entegrasyonunu etkileyen faktörlere yönelik içerik analizi. *Dokuz Eylül Üniversitesi Buca Eğitim Fakültesi Dergisi*, (31), 48-67.

Kayhan, O. (2019). Uzaktan eğitim öğrencilerin mezuniyet durumlarının veri madenciliği yöntemleri ile tahmini: Amasya Üniversitesi Örneği. Yüksek Lisans Tezi, Amasya Üniversitesi / Fen Bilimleri Enstitüsü, Amasya.

Keser, S. B. (2021). Önerilen Yapay Sinir Ağı Algoritması ile Ortaokul Öğrencilerin Akademik Performansının Tahmini. *Veri Bilimi Dergisi*, 4(2), 19-32.

Kılınç, Ç. (2015). Üniversite öğrenci başarısı üzerine etki eden faktörlerin veri madenciliği yöntemleri ile incelenmesi, Yüksek lisans tezi, Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü.

Koç, T. ve Akın, P. (2022). Estimation of High School Entrance Examination Success Rates Using Machine Learning and Beta Regression Models. *Journal of Intelligent Systems: Theory and Applications*, 5(1), 9-15.

Koyuncugil, A. S. (2007). "Veri Madenciliği ve Sermaye Piyasalarına Uygulaması", Sermaye Piyasası Kurulu Araştırma Raporu, Araştırma Dairesi.

Li, X. (2019). A brief Analysis of University Libraries Information Literacy Education Innovation in the Big Data Era. *Shandong Technology and Business University*, China.

MEB, (2018). 2023Eğitimvizyonu. http://2023vizyonu.meb.gov.tr/doc/2023_EGITIM_VIZYONU.pdf.

Mohammed, A. F., Humbe, V. T., & Chowhan, S. S. (2016). *A review of big data environment and its related technologies.* International Conference on Information Communication and Embedded Systems (ICICES).

Molluzzo, J. C. & Lawler, J. P. (2015). A Proposed Concentration Curriculum Design for Big Data Analytics for Information Systems. *Information Systems Education Journal.* 13 (1).

Monino, J. L. & Sedkaoui, S. (2016). *Big data, open data and data development* (Vol. 3). John Wiley & Sons.

Natek, S., Zwilling, M. (2014). Student data mining solution–knowledge management system related to higher education institutions, Expert Systems with Applications.

Olayinka, O., Kekeh, M., Sheth-Chandra, M., & Akpinar-Elci, M. (2017). Big data knowledge in global health education. *Annals of Global Health*, 83(3-4), 676-681.

Oğuzlar, A. (2004) *Veri Madenciliğine Giriş*, Ekin Kitabevi, Bursa.

Olgun K. B. (2021). Ters yüz sınıflardaki video izleme davranışları incelenerek veri madenciliği ile başarının tahmin edilmesi. Doktora Tezi, İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.

Oracle. (2019). *Data mining consepts.* https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/process.htm#CHDFGCIJ

Özbay, Ö. (2015a). Öğretim yönetim sistemi üzerinde üniversite (lisans) düzeyindeki öğrenci hareketliliğinin veri madenciliği yöntemleriyle analizi. Yüksek lisans Tezi, Eğitim Bilimleri Enstitüsü.

Özbay, Ö. (2015b). Veri madenciliği kavramı ve eğitimde veri Madenciliği uygulamaları. *Uluslararası Eğitim Bilimleri Dergisi*, (5), 262-272.

Özcan, C. (2014). Veri madenciliğinin güvenlik uygulama alanları ve veri madenciliği ile sahtekarlık analizi. Yüksek lisans tezi, İstanbul Bilgi Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.

Özen, Ü. (2014). *Bilgi Sistemlerine Giriş: Temel Kavramlar.* Atatürk Üniversitesi AOF Yayınevi.

Öztürk, A. (2018). Açık ve uzaktan öğrenme ortamlarında eğitsel veri madenciliği. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, 4(2), 10-13.

Pan, J., Liu, L., Ke, G., & Davis, H. (2018). *Research on the Reform of Education under the Background of Big Data.* Advances in Social Science Education and Humanities Research, 79-85.

Pehlivanoğlu, M. K. ve Duru, N. (2015). Veri madenciliği teknikleri kullanılarak ortaokul öğrencilerinin sosyal ağ kullanım analizi: Kocaeli ili örneği. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 3(2), 508-517.

Peña-Ayala, A. (2014). *Educational data mining. Studies in Computational Intelligence*, *Springer.*

Picciotto, R. (2020). Evaluation and the Big Data Challenge. *American Journal of Evaluation*, 41(2), 166–181. https://doi.org/10.1177/1098214019850334 .

Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of asynchronous learning networks*, 16(3), 9-20.

Pratsri, S. & Nilsook, P. (2020). Design on Big data Platform-based in Higher Education Institute. *Canadian Center of Science and Education*, (4). 10.5539/hes.v10n4p36.

Prasad, D. V., BalaBhargavi, S., Jahnavi, M., & Vandana, C. (2019). Comparative Study of Big Data Analytics: Applications & Technologies. *The International journal of analytical and experimental modal analysis.*

Prinsloo, P., Archer, E., Barnes, G., Yuraisha, C. & Dion, Z (2015). Big(ger) Data as Better Data in Open Distance Learning. *International Review of Research in Open and Distributed Learning.*

Prytherch, R. (2005). *Harrod's Librarians' Glossary and Reference Book: A Dictionary of Over*, Hampshire: Ashgate Publishing Limited.

RapidMiner, (2020). RapidMiner Studio. https://rapidminer.com/products/studio/

Roberts, L., Chang, V., & Gibson, D. (2016b). Ethical considerations in adopting a universityand system-wide approach to data and learning analytics. *In B. Kei Daniel (Ed.), Big data and learning analytics in higher education* (pp. 89–108).

Romero, C., & Ventura, S. (2013). Data mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3(1), 12-27.

Romero, C. & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3), 1355. https://doi.org/10.1002/widm.1355

Sağıroğlu, Ş., ve Koç, O. (2017). Büyük veri ve açık veri analitiği: Yöntemler ve uygulamalar, Grafiker Yayınları, Ankara.

Salal, Y. K., Abdullaev, S. M., & Kumar, M. (2019). Educational data mining: Student performance prediction in academic. *International Journal of Engineering and Advanced Technology,* 8(4C), 54-59.

Sara, N. B., Halland, R., Igel, C., & Alstrup, S. (2015). High-school dropout prediction using machine learning: a Danish large-scale study. *In M. Verleysen (Ed.), Computational Intelligence and Machine Learning* (pp. 319-324).

Savaş, S., Topaloğlu, N., ve Yılmaz, M. (2012). Veri madenciliği ve Türkiye'deki uygulama örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 11(21), 1-23.

Schaeffer, D. M., & Olson, P. C. (2014). Big Data Options For Small And Medium Enterprises. *Review of Business Information Systems (RBIS)*, 18(1), 41–46. https://doi.org/10.19030/rbis.v18i1.8542

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181, 526-534. https://doi.org/10.1016/j.procs.2021.01.199.

Shahar, T. H. (2017). Educational justice and big data. *Theory and Research in Education*, 15(3), 306–320. https://doi.org/10.1177/1477878517737155

Siemens, G. & Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education, Educause review, 46(5), 30.

Sin, K., & Muthu, L. (2015). Application of big data in education data mining and learning analytics, A lterature review. *ICTACT Journal on Soft Computing*, 5(4), 1-035.

Şahin, M. (2018). *E-Öğrenme Ortamlarına Yönelik Öğrenme Analitiklerine Dayalı Müdahale Motoru Tasarımı ve Geliştirilmesi*. Doktora tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Mayer-Schönberger, V. & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of data warehousing,* 5(4), 13-22.

Sigman, B. P., Garr, W., Pongsajapan, R., Selvanadin, M., Bolling, K., & Marsh, G. (2014). Teaching big data: Experiences, lessons learned, and future directions. decision line, 45(1), 10-15.

Siemens, G., & Baker, R. S. D. (2012). *Learning analytics and educational data mining: towards communication and collaboration.* In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 252-254).

Silahtaroğlu, G. (2008). *Veri madenciliği.* (2. Basım), Papatya Yayınları, İstanbul.

Song, L., Li, Z., He, B., Xu, S., & Meng, Y. (2019). The Reform in Education of Big Data on the Measurement and Control of Architectural Major. *Xi'an University of Architecture and Technology*, 10.12677/ae.2019.91012.

SVE. (2019). *What is the CRISP-DM methodology*, Smart vision http://www.sv-europe.com/crisp-dm-methodology/

Şeker, Ş. (2013). *İş Zekâsı ve Veri Madenciliği*. Cinius Yayınları, İstanbul.

Şeker, Ş. E. (2018). *CRISP-DM: Endüstriler Arası Standart İşleme–Veri Madenciliği için (Cross Industry Standard Processing–Data Mining)*. YBS Ansiklopedi.

Şenel, S. ve Kutlu, Ö. (2015). Ankara üniversitesi uzaktan eğitim programına katılan öğrencilerin akademik başarılarını yordayan faktörler. *Journal of Measurement and Evaluation in Education and Psychology*, 6(2).

Şengür, D. (2013). Öğrencilerin akademik başarılarının veri madenciliği metotları ile tahmini. Yüksek lisans tezi, Fırat Üniversitesi Eğitim Bilimleri Enstitüsü.

Tan, Ş. (2010). *Öğretim ilke ve yöntemleri*. Pegem Akademi.

Türk Dil Kurumu. (1969). Türkçe sözlük (genişletilmiş baskı). Ankara: TDK.

Tekin, A., ve Öztekin, Z. (2018). Eğitsel Veri Madenciliği il ilgili 2006-2016 yılları arasında yapılan çalışmaların incelenmesi. Eğitim Teknolojisi Kuram ve Uygulama, *Dergi Park Akademik*, 8(2), 108-124.

Tekin, A. (2014). Early prediction of students' grade point averages at graduation: A data mining approach. *Eurasian Journal of Educational Research*, 54, 207-226.

Tetko, I. V., Engkvist, O., Koch, U., Reymond, J. L., & Chen, H. (2016). BIGCHEM: challenges and opportunities for big data analysis in chemistry. *Molecular informatics*, *35*(11-12), 615-621.

Topuz, S. (2021). Eğitsel Verilerde Weka ve Orange Veri Madenciliği Yazılımlarından Elde Edilen Analiz Sonuçlarının Karşılaştırılması. Yüksek Lisans tezi, Hacettepe Üniversitesi Eğitim Bilimleri Ana Bilim Dalı, Ankara.

Tufféry, S. (2011). *Data mining and statistics for decision making*. John Wiley & Sons.

Tuzcu, S. (2018). *Ders yönetim sistemi tabanlı veri madenciliği ve öğrenme analitiği*, Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, Eskişehir Osmangazi Üniversitesi.

Uğuz E., Şahin, S. ve Yılmaz, R. (2021). PİSA 2018 Fen Bilimleri Puanlarının Değerlendirilmesinde Eğitsel Veri Madenciliğinin Kullanımı. *Bilgi ve İletişim Teknolojileri Dergisi*, 3(2), 212-227.

Uzun, Y., Uzun, F. ve Çakar, E. (2021). *Veri Madenciliği Kullanım Alanları.* Uluslararası Mühendislik, Doğa ve Sosyal Bilimler Sempozyumunda sunuldu, Batman Üniversitesi.

Vatandaş, C. (2007). Toplumsal Cinsiyet ve Cinsiyet Rollerinin Algılanışı. *Istanbul Journal of Sociological Studies*, (35), 29-56.

Wahyuni, S. (2018). Implementasi Rapidminer Dalam Menganalisa Data Mahasiswa Drop Out. *Universitas Pembangunan Pancabudi Medan*, 10(2), 1899-1902. ISSN:1979-5408

Williams, T., Cheng, X., Majumder, M., Hastings, M., Suh, H., Dash, K. & Yeo, J. J. (2020). Collaborative Big Data Review for Educational Impact. *School Community Journal,* 30.

Winne, P. H. (2017), "*Learning analytics for self-regulated learning*", The Handbook of learning analytics, 241-249.

Xu, C., Wang, C., & Yang, N. (2019, October). *The supply-side precision reform and innovation of ideological and political education in colleges based on big data technology*. In 4th International Conference on Modern Management, Education Technology and Social Science (MMETSS 2019) (pp. 637-643). Atlantis Press.

Yap, A. Y. & Drye, S. (2018). The Challenges of Teaching Business Analytics: Finding Real Big Data for Business Students. *Information Systems Education Journal,* 16 (1).

Yılmaz, M. (2017). Enformasyon ve bilgi kavramları bağlamında enformasyon yönetimi ve bilgi yönetimi. *Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi*, 49(1).

Yurdakul, S. (2015). Veri madenciliği ile lise öğrenci performanslarının değerlendirilmesi, Yüksek Lisans Tezi, Kırıkkale Üniversitesi Fen Bilimleri Enstitüsü, XVII.

Yu, R., Li, Q., Fischer, C., Doroudi, S., ve Xu, D. (2020). Towards accurate and fair prediction of college success: evaluating different sources of student data. *In Proceedings of the 13th* International Conference on Educational Data Mining.

Yükseltürk, E., Özekeş, S. ve Türel, Y. K. (2014). Predicting dropout student: an application of data mining methods in an online education program. *European Journal of Open, Distance and e-learning*, 17(1).

Zaki, M. J., Meira Jr, W., & Meira, W. (2014). *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press.

Zaiane, O. R., (2001), Web Usage Mining for a Better Web-based Learning Environment, *Conference on Advanced Technology for Education*, 60-64, https://webdocs.cs.ualberta.ca/~zaiane/postscript/CATE2001.pdf.

Zhang, Y. N. (2017). Research on the innovation of college students' ideological and political education in big data era. *Journal of Jiamusi Vocational Institute*, 21(1), 143.

EKLER

EK 1: Etik Kurul Kararı

EK 2: Araştırma İzin Belgesi