

GAZİ

JOURNAL OF ENGINEERING SCIENCES

A Review of Studies Evaluating Students' Academic Performance with Machine Learning Techniques

Sema Kayalı^a, Serkan Savaş^b

Submitted: 21.04.2024 Revised: 13.07.2024 Accepted: 09.10.2024 doi:10.30855/gmbd.0705AR09

ABSTRACT

Keywords: Artificial intelligence in education, academic achievement, student performance, machine learning, attribute selection

^a Çankırı Karatekin University,
Graduate School of Natural and Applied
Sciences,
Dept. of Electronics and Computer
Engineering
18200 - Çankırı, Türkiye
Orcid: 0009-0005-3949-0021
e mail: semakanis@gmail.com

^b Kırıkkale University,
Engineering and Natural Sciences Faculty,
Dept. of Computer Engineering
71450 - Kırıkkale, Türkiye
Orcid: 0000-0003-3440-6271

^{*}Corresponding author:
serkansavas@kku.edu.tr

The application of machine learning in educational technologies is increasingly prevalent, supporting personalized learning and enhancing student performance. This study examines prominent methods and models in the literature for predicting or classifying students' academic achievements. Research conducted over the past 15 years was analyzed, focusing on algorithms, feature selection techniques, and evaluation metrics, which were statistically assessed. The analysis of the 10 most successful models is presented along with feature selection techniques and classification algorithms. Algorithms such as Artificial Neural Networks, Logistic Regression, Random Forest, and K-Nearest Neighbors demonstrated high accuracy rates. Information Gain emerged as the most frequently used feature selection technique, while Naive Bayes was identified as the most common classification algorithm. Key factors influencing student performance include parental education level, the quality of prior education, and family income. This study provides a comprehensive overview of the literature and serves as a reference point for future research in the field.

Öğrencilerin Akademik Performanslarını Makine Öğrenmesi Teknikleriyle Değerlendiren Çalışmaların İncelenmesi

ÖZ

Makine öğrenmesi, eğitim teknolojilerinde bireyselleştirilmiş öğrenmeyi desteklemek ve öğrenci başarısını artırmak için giderek daha fazla kullanılmaktadır. Bu çalışmada, öğrencilerin akademik başarılarının tahmin edilmesi veya sınıflandırılması üzerine literatürdeki önde gelen yöntemler ve modeller incelenmiştir. Son 15 yılda yapılan araştırmalar analiz edilerek kullanılan algoritmalar, öznitelik seçim teknikleri ve ölçüm metrikleri istatistiksel olarak değerlendirilmiştir. Çalışmada En başarılı 10 modelin analizi, öznitelik seçim teknikleri ve sınıflandırma algoritmalarıyla birlikte sunulmuştur. Yapay Sinir Ağları, Lojistik Regresyon, Rastgele Orman ve K-En Yakın Komşu gibi algoritmalar yüksek başarı oranlarıyla öne çıkmıştır. En sık tercih edilen öznitelik seçim tekniği Bilgi Kazancı, sınıflandırma algoritması ise Naive Bayes olmuştur. Öğrenci başarısını etkileyen temel faktörler arasında ebeveynlerin eğitim düzeyi, önceki eğitim kalitesi ve aile gelir düzeyi yer almaktadır. Çalışma, literatüre kapsamlı bir bakış sunarken, ileride yapılacak araştırmalar için bir rehber niteliğindedir.

Anahtar Kelimeler: Eğitimde yapay zeka, akademik başarı, öğrenci performansı, makine öğrenmesi, öznitelik seçimi

1. Giriş (Introduction)

Eğitim, bir ülkenin kalkınması, insanların daha iyi bir hayat sürdürebilmeleri ve bireylerin üretken olması için zorunludur. Eğitimde akademik başarının artırılması okul idaresi, öğretmen, öğrenci ve veli açısından her zaman istenilen bir durumdur. Eğitim öğretim sürecinin içerisindeki paydaşlar (veli, öğrenci, okul yönetimi ve öğretmen), geçmişten günümüze birçok sorunla karşılaşmıştır. Bu sorunlar öğrencilerin akademik başarılarını olumlu veya olumsuz yönde etkilemektedir. Aynı sınıf içinde eğitim gören öğrencilerin ailelerinin kültürel yapıları ve öğrencilerin farklı yaşam şekilleri bunlara örnek gösterilebilir. Öğrencilerin öğrenme ve bilgi edinme yolları birbirinden farklıdır. Aynı sınıftaki öğrencilerin akademik başarılarının birbirinden farklı olması bu durumu doğrulamaktadır [1].

Veri madenciliği ile büyük verilerin işlenerek anlamlandırılmasında büyük başarılar elde edilmektedir. Bu sebeple son zamanlarda veri madenciliği büyük önem kazanmıştır. Sağlık, ekonomi, iletişim, bankacılık hizmetleri, ulaşım ve eğitim-öğretim gibi pek çok sektörde karşılaşılan problemlere çözüm üretmek için gerekli tahminleri yapmaktadır [2-4]. Veri madenciliği uygulamalarının eğitim alanında yapılan çalışmalarına eğitsel veri madenciliği denmektedir. Bu çalışmalar ile eğitim sisteminde kalite ve verimliliği arttırmak, eğitim ile ilgili konuları analiz etmek ve eğitim süreçlerini iyileştirmek istenmektedir. Eğitsel veri madenciliği, eğitimdeki farklı paydaşlara (öğrenciler, öğretmenler, yöneticiler ve politika yapımcılar) önemli bilgiler sunarak, eğitim sisteminin genel performansını ve verimliliğini artırabilir. Bu uygulamalar sayesinde, eğitimin her aşamasında daha bilinçli kararlar alınabilir, kaynaklar daha verimli kullanılabilir ve öğrenci başarısı daha sürdürülebilir hale gelebilir. Teknolojinin eğitime entegrasyonu ile, bireysel öğrenme hızlarına ve kapasitelerine göre uyarlanmış eğitim programları geliştirilebilir ve böylece öğrencilerin başarısızlık oranları azaltılabilir [5-7]. Ülkemizin eğitim sisteminde rekabetçi bir ortam bulunmaktadır. Gelişen teknoloji ile eğitim kurumları, akademik gelişim için eğitsel veri madenciliği araçlarının ve tekniklerinin kullanımına giderek daha fazla önem vermektedirler. Bu sayede uygulanan eğitim programlarının başarıyı nasıl etkilediği tespit edilebilir. Programların etkilerinin önceden bilinmesi, gelecekte uygulanacak yeni eğitim modellerinin daha yararlı bir şekilde uygulanmasına imkân tanır. Eğitim sürecinin doğru bir şekilde planlanması ve yürütülmesi hem öğrenci hem de eğitmen için akademik başarıyı artıracaktır [8-10]. Makine öğrenmesi yöntemleri ile öğrenci başarısını artırmaya yönelik akademik başarı sınıflandırması ve tahmini oldukça popüler hale gelmiştir. Öğrencilerin bireysel ihtiyaçlarını ve öğrenme stillerini anlamaya yardımcı olan bu yöntemler, öğrencilere kişiselleştirilmiş eğitim deneyimleri sunarak onların potansiyellerini maksimize etmeyi hedefler. Ayrıca, öğrenci verilerinin analizi ile zorluk çekilen konular ve alanlar belirlenebilir, bu da öğretim stratejilerinin daha hedefli ve etkili bir şekilde oluşturulmasına imkan tanır [11].

Makine öğrenmesi istatistik ve matematik bilimlerinden yararlanılarak geliştirilen bütünleştirme ve sorun çözme becerileri açısından çok yararlı ve seçenekli bir çalışma yöntemidir [12]. Bilgisayar yazılımlarıyla çeşitli ortamlarda makineler eğitilerek otomatik bir şekilde problemlerin algılanması ve sonucunda çözüm üretilmesi sağlanır. Makine öğrenmesi ile geliştirilen sistemler, gelecek için öngörülerde bulunarak karar alma süreçlerine olumlu katkılar sağlayarak insan yaşamını kolaylaştırmaktadır. Bu sebeple bu çalışmanın amacı, öğrencilerin akademik başarısının farklı makine öğrenmesi yöntemleri ile sınıflandırılmasına yönelik çalışmaların kapsamlı olarak araştırılması ve bir araya getirilmesidir. Çalışmalarda kullanılan teknikler, başarımlar, bulgular ve katkılar da sunulmuştur. Bu araştırmanın, sonraki çalışmalar için bir temel oluşturması amaçlanmaktadır.

Literatürde öğrencilerin akademik başarılarının sınıflandırılmasına yönelik ulusal ve uluslararası çok sayıda çalışma vardır. Fakat bu çalışmaların yöntem ve teknikleri, sonuçları ve stratejileri açısından incelenmesi zorluğu tüm araştırmacılar için baş edilmesi gereken bir problemdir. Bu araştırma, bu problemin giderilmesi için önemli sonuçlar sunmaktadır. Akademik performansla ilgili literatürü incelemek için daha önce yapılmış bazı çalışmalar vardır ancak bunların çoğu genel literatür incelemeleridir ve genel öğrenci performans tahminine yöneliktir [13-15]. Bu çalışmada makine öğrenmesi alanındaki en iyi uygulamalar toplanarak incelenmesi amaçlanmıştır. Ayrıca sadece makine öğrenmesi teknikleriyle performans tahminine yönelik araştırmalar değil bu araştırmalar içerisinde öznitelik seçim teknikleri kullanılarak daha etkili sonuçlar üretmeyi amaçlayan literatür incelenmiştir.

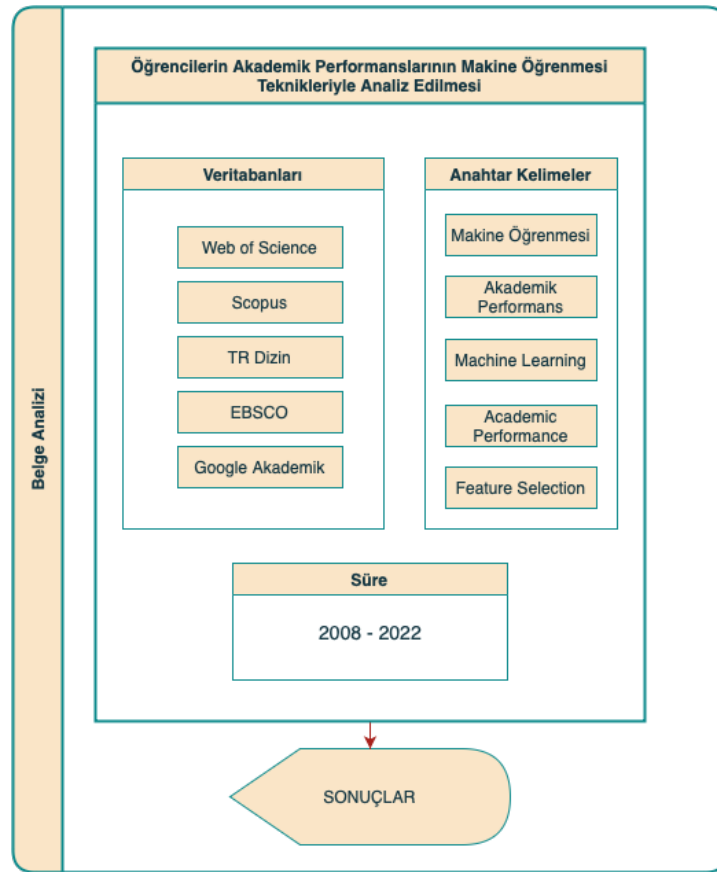
Bu çalışmanın ikinci bölümünde araştırma için uygulanan strateji belirtilmiştir. Üçüncü bölüm olan bulgular bölümünde akademik başarı tahmini veya sınıflandırmasında kullanılan farklı makine öğrenmesi algoritmaları ve öznitelik seçme yöntemlerinin literatürde bulunan benzer çalışmalarına yer verilmiştir. Dördüncü bölümünde, çalışmada özetlenen araştırmalar bir tablo halinde sunulmuş ve araştırmaların

kapsamlarına yönelik istatistik bilgileri tartışılmıştır. Sonuç bölümünde ise çalışmaya ait sonuçlar ve daha sonraki çalışmalara yönelik öneriler belirtilmiştir.

2. Uygulanan Strateji (Implemented Strategy)

Araştırma stratejisi, sistematik süreçler ve araçlar kullanarak veri toplamak ve analiz etmek olarak belirlenmiştir. Çalışmada elde edilen veriler farklı akademik veri tabanlarında yayınlanan makine öğrenmesi ve akademik performans değerlendirme çalışmalarını içermektedir. Mevcut çalışma, bu verileri analiz ederek, bu yayınların kapsam ve sonuçlarını araştırarak dünya genelinde öğrencilerin akademik başarılarının makine öğrenmesiyle değerlendirilmesi alanında atılan adımları incelemeyi ve tespit etmeyi amaçlamaktadır. Bu sayede eğitim öğretim uygulamalarında yapay zekâ teknolojilerinin etkin kullanımının sağlanması için atılması gereken adımlar, tespit edilen eksiklikler ve gelecekte yapılabilecek çalışmalar konusunda bir durum tespiti çalışması yapılmıştır.

Bu çalışmada nitel araştırma yöntemlerinden biri olan betimsel araştırma modeli kullanılmıştır. Araştırma felsefesi açısından temel bir araştırma modeli, uygulama yöntemi açısından ise kütüphane, arşiv, müze ve internet gibi mevcut kaynaklardan derlenen verilere (metin, resim vb.) dayalı bir belge araştırma modeli oluşturulmuştur [16]. Çalışmada kullanılan araştırma modeli Şekil 1'de gösterilmiştir.



Şekil 1. Araştırma Modeli
(Research Model)

Çalışmada kullanılan araştırma modeli Web of Science (WoS), Scopus, EBSCO, Google Scholar ve TR Dizin'den elde edilen çalışmalardan oluşmaktadır. Araştırmada kullanılan çalışmalar akademik veri tabanlarına kaydedilmiş ve araştırma konusuna göre sınıflandırılmıştır. Belirlenen anahtar kelimeler kullanılarak veri tabanlarından 2008-2022 yılları arasında, 15 yıllık süreçte gerçekleştirilen makine öğrenmesi ve akademik performans değerlendirme ile ilgili çalışmalar taranmıştır. Elde edilen çalışmalar araştırma konusuna göre ve daha sonra tarihsel sıraya göre düzenlenmiştir. Bu çalışmalar şöyle sınıflandırılmıştır:

- Yazar
- Çalışma Türü

- Veri Seti
- Öznitelik Seçim Teknikleri
- Yöntemler
- Kullanılan Araçlar
- Değerlendirme Ölçütleri

Araştırma kapsamında belirlenen veri tabanlarındaki makale, konferans bildirileri ve tez çalışmaları incelenmiş kitap veya kitap bölümü taraması gerçekleştirilmemiştir. Araştırmalar hem Türkçe hem de İngilizce anahtar kelimelerle yapılmış ve çalışma dilleri olarak da Türkçe ve İngilizce çalışmalar incelenmiştir. Ayrıca farklı veri tabanlarından elde edilen aynı yayınlar olmuştur ancak çiftleme yayınlara rastlanılmamıştır. Araştırma gerçekleştirilirken anahtar kelimeler birlikte kullanılmıştır. Örneğin makine öğrenmesi anahtar kelimesi tek başına kullanıldığında çok farklı sonuçlar getirebilecekken, bu anahtar kelime akademik performans anahtar kelimesi ile kullanıldığında sonuçlar daraltılarak hedefe yönelik çalışmalara ulaşılmasını sağlamaktadır. Bu nedenle anahtar kelimeler birlikte kullanılarak tarama gerçekleştirilmiştir.

3. Araştırma Bulguları (Research Findings)

Literatürde öğrenci performansını sınıflandırma ve tahminleme üzerine makine öğrenmesi alanında farklı çalışmalar yapılmıştır. Literatür araştırmasında incelenen yayınlara ait detaylar Tablo 1'de gösterilmiştir.

Tablo 1. İncelenen çalışmalara ait anahtar bilgiler (Key information about the studies reviewed)

Yazar	Çalışma Türü	Veri Seti	Öznitelik Seçim Teknikleri	Yöntemler	Kullanılan Araçlar	Değerlendirme Ölçütleri
[17]	Makale	578 kayıt 12 öznitelik	Yok	Çoklu Doğrusal Regresyon Analizi YSA (%94.43)	Rapid Miner	Doğruluk oranı Mutlak hata yüzdesi
[18]	Makale	421 kayıt 29 öznitelik	IG	YSA (%98.6) KNN, LR, SVM, DT, RF, NB	Orange	Doğruluk oranı Karışıklık matrisi ROC eğrisi
[19]	Makale	497 kayıt 9 öznitelik	Yok	DT, NB Kural Tabanlı (%71.3)	WEKA	Doğruluk oranı
[20]	Makale	300 kayıt 24 öznitelik	CFS, GR, IG, SU	J48, PART RF (%99) BayesNet	WEKA	Doğruluk oranı
[21]	Bildiri	56.000 kayıt 34 öznitelik	IG	DT, RF (%75.52), GBT, DL, NB, LR, GLM	Rapid Miner	Doğruluk oranı
[22]	Makale	17 öznitelik	Yok	GBM (%99)	Belirtilmemiş	ROC eğrisi
[23]	Bildiri	429.757 kayıt 11 öznitelik	Özellik seçimi kullanılmış fakat yöntem adı belirtilmemiş	C5.0 (%82.1), CHAID, LR, YSA, CART, QUEST	SPSS Clementine	Doğruluk oranı
[24]	Makale	44 öznitelik	IG Chi-Square MI MICH	KNN Hibrit C5.0 RF (99.98) Improved Deep Belief Networks	R	Doğruluk oranı
[25]	Makale	2.039 kayıt	Özellik seçimi kullanılmış fakat yöntem adı belirtilmemiş	LR NB YSA (%79) KA		Doğruluk oranı
[26]	Makale	2.850 Kayıt 33 öznitelik	IG BFS Genetik algoritma	KNN, KA, NB, Derin Öğrenme (%97.7)	Rapid Miner	Doğruluk oranı
[27]	Makale	Belirtilmemiş	Yok	NB, KNN, KA, RF, DVM, LR, YSA (%89.84)	Belirtilmemiş	Doğruluk oranı

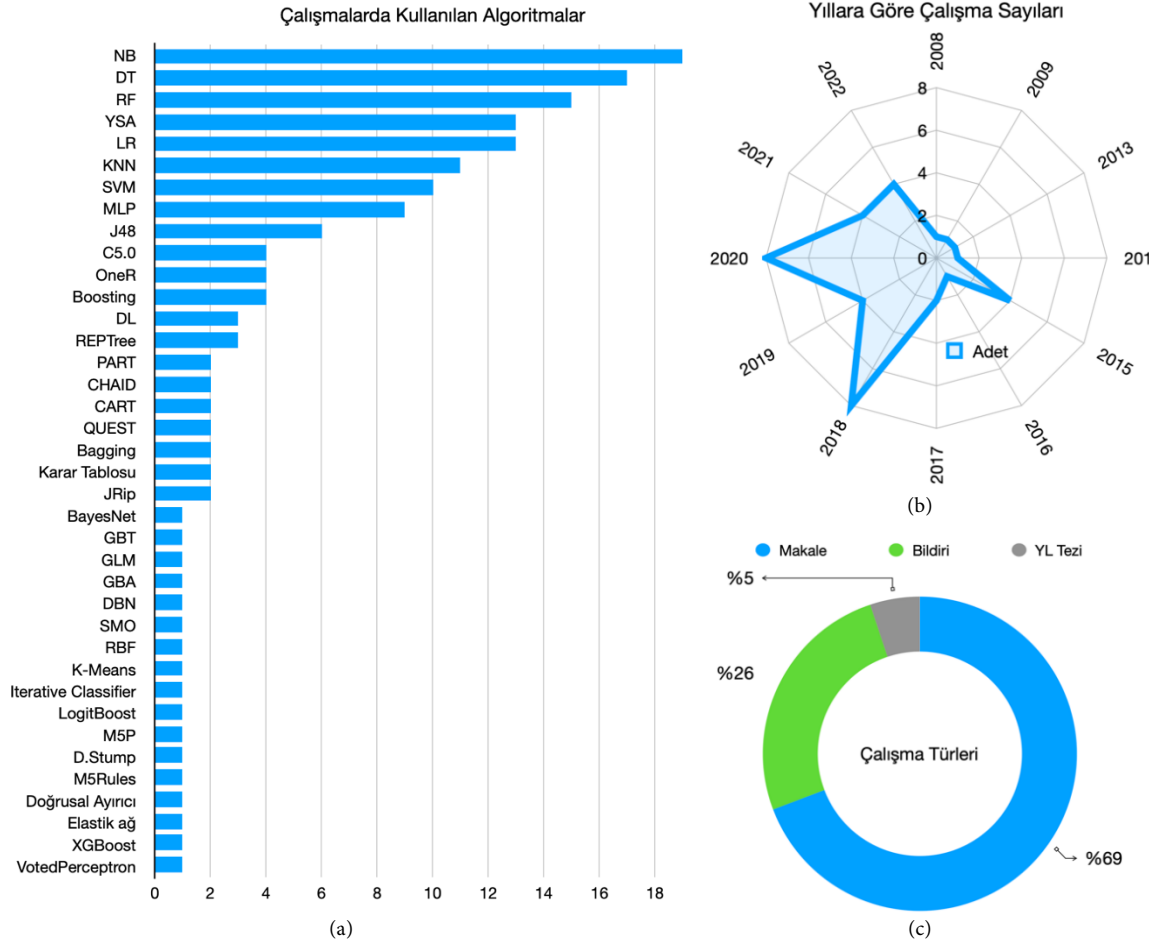
Tablo 1. Devamı (Contuniuue)

[28]	Makale	152 kayıt	Chi-Square IG SU ReliefF	MLP (%75), NB, SMO, J48, REPTree	WEKA	Doğruluk oranı
[29]	Makale	195.584 kayıt	Yok	YSA+MLP (%85) RBF	Belirtilmemi ş	Korelasyon katsayıları Doğruluk oranı
[30]	Makale	5.358 kayıt 8 öznitelik	Yok	K-Means C5.0 CART (%85) CHAID QUEST	SPSS Clementine	Kümeleme analizi Doğruluk oranı
[31]	Makale	3.775 kayıt	ANOVA	Kümeleme an. Gruplama yöntemi Regresyon Korelasyon teknikleri	Orange D-melt SPSS	Kümeleme analizi Gruplama yöntemi Regresyon
[32]	Makale	395 Kayıt 33 öznitelik	Yok	Iterative Classifier OneR (%92.15) LogitBoost YSA	WEKA	Duyarlılık Özgünlük Doğruluk oranı ROC eğrisi
[33]	Makale	5820 Kayıt 34 öznitelik	Yok	LR (%84.1)	Belirtilmemi ş	Doğruluk oranı Karışıklık matrisi
[34]	Bildiri	1.000 Kayıt 8 Öznitelik	Rastgele n- sınıfı ANOVA Select k-best	SVM (%72)	Jupiter Notebook (Python)	Doğruluk Oranı
[35]	Makale	1.395 kayıt 27 öznitelik	CFS	LR (%62-65), NB, SVM, KNN, RF, Bagging, MLP	WEKA	Doğruluk oranı
[36]	Makale	480 Kayıt 17 öznitelik	Hiper parametre seçimi; Izgara araması (%82.3)	KNN SVM AutoML (DRF) (%77.5 - %82.30)	H2O AutoML (Python)	Doğruluk oranı F-skoru RMSE MSE
[37]	Makale	3.794 Kayıt 13 öznitelik	Yok	YSA, M5P, D.Stump, M5Rules, DT, Bagging	WEKA	MSE Korelasyon katsayısı
[38]	Bildiri	1.044 Kayıt 33 öznitelik	CAE IGAE GRAE	NB, RF, J48, MLP, Karar Tablosu (%76.11), JRip, LR	WEKA	Doğruluk oranı
[39]	Bildiri	500 Kayıt 16 öznitelik	Belirtilmemiş	LR SVM (%78)	WEKA Jupyter Notebook	Doğruluk Kesinlik Duyarlılık F-skoru Karışıklık matrisi
[40]	Makale	395 Kayıt 33 öznitelik 650 kayıt 52 öznitelik	Özellik seçimi var	Hibrid (1. veri seti %90 ve 2. veri seti %85) (Kümeleme + sınıflandırma)		Doğruluk Kesinlik Duyarlılık F-skoru Karışıklık matrisi
[41]	Makale	-	Sarmalayıcı öznitelik seçimi	KNN (%92 - 2. veri seti), DT, NB (%87 - 1. veri seti), SVM, Doğrusal Ayırıcı	MATLAB	Doğruluk oranı
[42]	Makale	38 öznitelik	Özellik seçimi var	NB (%86) MLP C4.5 (J48)	WEKA	Doğruluk oranı
[43]	Bildiri	500 kayıt	Select K-Best	LR, NB, DT, Boosting (%75) RF	Python	Doğruluk oranı

Tablo 1. Devamı (Contunue)

[44]	Makale	3.225 Kayıt 57 öznitelik	CFS FSA IG Relief	LR (%90) Elastik ağ reg. (%90) RF XGBoost	R	Doğruluk oranı
[45]	Makale	1.044 Kayıt 33 öznitelik	Filtreleme Özyinelemeli özellik çıkarma metotları	YSA (%87.9-%97.6)	Python	Doğruluk Kesinlik Duyarlılık F-skoru Karışıklık matrisi
[46]	Bildiri	395 kayıt 33 öznitelik	Yok	C5.0, Boosted- C5, Regresyon Ağaçları, SVM, LR, RF Derin Ö. (%87)	R	Doğruluk oranı
[47]	Bildiri	6.882 Kayıt 15 öznitelik	mRMR Genetik algoritma SVM, IG	KNN (%91.12) DT NB YSA	RapidMiner	Doğruluk oranı
[48]	Yüksek Lisans Tezi	458 Kayıt 79 öznitelik	C4.5 CFS	C4.5 MLP KNN SVM (%85.5)	WEKA	Doğruluk Kesinlik Duyarlılık F-skoru Kappa değeri
[49]	Makale	1.044 Kayıt 33 öznitelik	CFS GR IG	NB, RF, DT, MLP, Karar Tablosu (%83.33)	MS SQL	Doğruluk F-skoru ROC eğrisi
[50]	Makale	1.969 kayıt 32 öznitelik	CFS Chi-Square GR IG Relief SU	VotedPerceptron NB OneR (%89) PART	WEKA	Doğruluk F-skoru ROC eğrisi
[51]	Bildiri	480 kayıt 16 öznitelik		MLP, DT, RF, OneR, NB, YSA (%79.1)	WEKA	Doğruluk Kesinlik Duyarlılık F-skoru Karışıklık matrisi
[52]	Yüksek Lisans Tezi	127 kayıt 49 öznitelik	Yok	YSA DT	MATLAB	Korelasyon katsayısı OKH, OMH
[53]	Makale	500 Kayıt 16 öznitelik 300 kayıt 24 öznitelik	Chi-Square CFS GR Ana Bileşenler Relief	MLP (%77), NB, DT, J48, RF, MLP, SL, OneR, Jrip, REPTree	WEKA	Doğruluk oranı
[54]	Makale	413 kayıt 34 öznitelik	Öznitelik seçimi yapılmış	NB DT MLP KNN	Belirtilmemi ş	Doğruluk oranı
[55]	Bildiri	1.044 kayıt 32 öznitelik	-	DT SVM RF YSA	R	Doğruluk oranı Karışıklık matrisi

Tablo 1’de sunulan öznitelik seçme ve çeşitli makine öğrenmesi yöntemlerini birleştirerek tahmin yapan çalışmalar incelenmiştir. İncelenen çalışmalara ait istatistiksel bilgiler özetlenmiştir. Yıllara göre gerçekleştirilen çalışma sayıları, bu çalışmaların türleri ve bu çalışmalarda kullanılan algoritmalar, Şekil 2’de sunulmuştur.

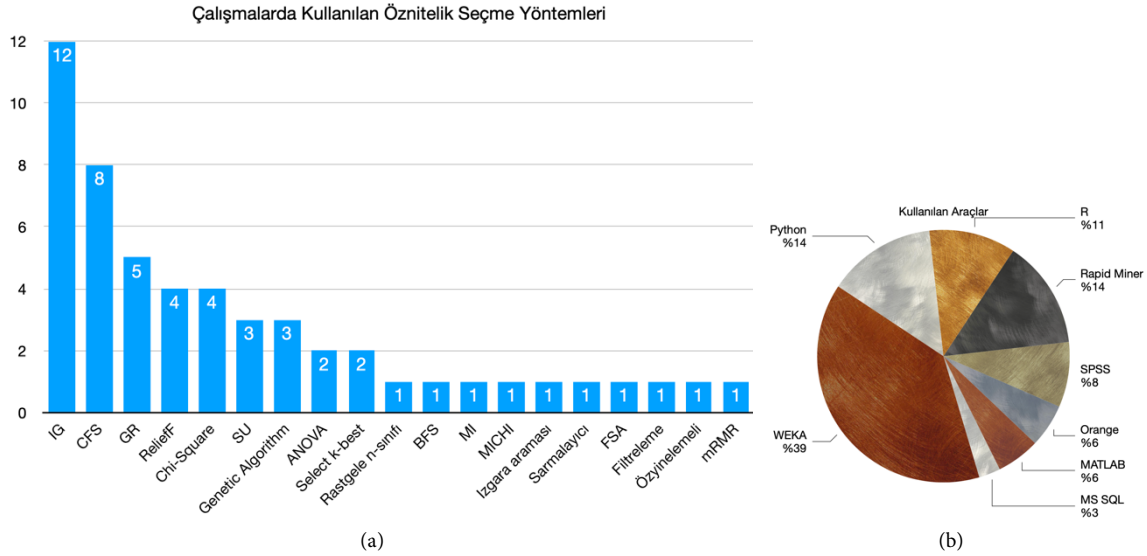


Şekil 2. Çalışma istatistikleri (a) çalışmalarda kullanılan algoritmalar (b) yıl bazında erişilen çalışma sayıları, (b) erişilen çalışma türü oranları

(Study statistics (a) algorithms used in the studies (b) number of studies accessed by year, (b) proportion of types of studies accessed)

Şekil 2.a'da görüldüğü gibi çalışmalarda en çok Naive Bayes (NB) algoritması test edilmiştir. Daha sonra Karar Ağaçları (Decision Tree-DT) gelmektedir. Üçüncü en sık kullanılan algoritma ise Rastgele Orman (Random Forest-RF) algoritmasıdır. Şekil 2.b'de görüldüğü gibi özellikle son 5 yılda bu alanda gerçekleştirilen çalışmaların sayısı artmaya başlamıştır. Ayrıca Şekil 2.c'de görüldüğü gibi çalışmaların büyük bir yüzdesi makale türündedir (%69). Elde edilen çalışmalardan %26 bildiri geri kalan %5'i ise tezli yüksek lisans çalışmasıdır.

Bu çalışmalarda birçok farklı yöntem uygulanmış, yeni yöntemler test edilmiş ve farklı öneriler ortaya konulmuştur. Bu çalışmaların hemen hemen hepsinde birden fazla teknik uygulanmış ve farklı algoritmalar test edilmiştir. Araştırmalarda kullanılan öznelik seçme yöntemleri ve kullanılan analiz araçları Şekil 3'te gösterilmiştir.



Şekil 3. Çalışmalarda kullanılan (a) öznitelik seçim yöntemleri ve (b) analiz araçları
(a) feature selection methods and (b) analysis tools used in the studies

Şekil 3.a'da en çok kullanılan öznitelik seçme tekniğinin Bilgi Kazancı (Information Gain-IG) teknik olduğu görülmektedir. İkinci sık kullanılan teknik Korelasyon Tabanlı Özellik Seçimi (Correlation-based Feature Selection-CFS) tekniğidir. Üçüncü sırada en yaygın kullanılan teknik ise Kazanç Oranı (Gain Ratio-GR) tekniğidir. Şekil 3.b'de görüldüğü gibi araştırmalarda en çok tercih edilen analiz aracı WEKA olmuş, bu aracı sırasıyla Python programlama dili, R programlama dili ve Rapid Miner aracı takip etmiştir. Öznitelik seçme yöntemleri kullanılarak yapılan ve öznitelik seçme yöntemleri kullanılmadan yapılan çalışmalar bu bölümde detaylı olarak açıklanmaktadır.

Altun ve arkadaşları (2019) çalışmalarında üniversite öğrencilerinin mezuniyet notlarını tahmin etmek üzere veri madenciliği modelleri geliştirmişlerdir. Bu amaçla; Akdeniz Üniversitesi Eğitim Fakültesi Sınıf Öğretmenliği Bölümü'nden 2012-2017 yılları arasında mezun olan 578 öğrencinin cinsiyet, medeni durum, kayıt yaşı ve birinci sınıf birinci dönem ara sınav puanlarına ilişkin veriler kullanılmıştır. Öğrenci bilgi sisteminden elde edilen verilere Rapid Miner platformundaki araçlar kullanılarak veri ön işleme uygulanmıştır. Elde edilen veri setinde öğrenci kod, cinsiyet, medeni durum, kayıt yaşı ve 8 derse ait ara sınav puanları olmak üzere 12 öznitelik yer almaktadır. Rapid Miner yazılımında, Çoklu Doğrusal Regresyon Analizi ve Yapay Sinir Ağları (YSA) ile tahmin modelleri oluşturulmuştur. Ortalama mutlak hata yüzdesi değerleri temel alınarak yapılan değerlendirmede; Regresyon analizi modeli ile %94.30 ve YSA modeli ile %94.43 başarı sağlandığı belirtilmiştir [17].

Yıldız ve Börekçi (2020), farklı okullardaki dokuzuncu sınıf öğrencilerinden toplanan eğitim ile ilgili veriler üzerinde bir fikir geliştirilmeye çalışmışlardır. Veriler; öğrenciler ve aileleri, çalışma rutinleri, öğrenme etkinliklerine katılma davranışları ve bilim hakkındaki düşüncelerini içermektedir. Araştırma toplanan verilerden sınavla öğrenci alan okulda okuyan ve sınavsız öğrenci alan okulda okuyan öğrencileri sınıflandırma problemi olarak ele alınmıştır. Oluşturulan modeller iki sınıf verisi üzerinden tahmin edilmiştir. Çalışmada, denetimli sınıflandırma algoritmalarının tahmin doğrulukları karşılaştırılmış ve sınıfların oluşumunda hangi değişkenlerin etkili olduğu tanımlanmıştır. Arayüz olarak Orange uygulaması kullanılmıştır. Tahmin doğruluğunu etkileyen faktörleri belirlemek için değişkenlerin bilgi kazanç katsayısını incelemişlerdir. Ailenin demografik değişkenlerinin, öğrencinin bilimsel inançlarının, çalışma rutinlerinin ve bazı derslere yönelik tutumlarının sınıflandırmayı etkilediği belirtmişlerdir. Kullanılan makine öğrenme algoritmaları: YSA, K-En Yakın Komşu (K-Nearest Neighbor - KNN), Lojistik Regresyon (Logistic Regression - LR), Destek Vektör Makinesi (Support Vector Machine - SVM), Karar Ağacı (Decision Tree - DT), Rastgele Orman (Random Forest - RF) ve Naive Bayes (NB) algoritmalarıdır. Makine öğrenmesi algoritmalarının

tahmin sonuçları karşılaştırıldığında YSA algoritmasının yaklaşık %98.6 başarı oranı ile en iyi model olduğu görülmüştür [18].

Ahmad ve arkadaşları (2015) çalışmalarında üniversite birinci sınıf öğrencilerinin Bilgisayar Bilimleri dersindeki akademik performansını tahmin etmek için bir veri madenciliği uygulaması gerçekleştirmişlerdir. Veriler, öğrencilerin demografik özelliklerini, önceki akademik kayıtlarını ve aile geçmiş bilgilerini içeren 2006-2007 ile 2013-2014 yılları arasındaki bilgileriyle oluşturulmuştur. Bu çalışmada, öğrencilerin not ortalamaları bağımlı değişken olarak seçilmiştir. Kullanılan diğer öznitelikler ırk, cinsiyet, aile geliri, üniversiteye giriş şekli ve üç konuda Malezya Eğitim Sertifikası notlarıdır. Çalışmada WEKA veri madenciliği yazılımı kullanılmıştır. Karar Ağaçları, Naive Bayes ve Kural Tabanlı sınıflandırma algoritmaları veri setine uygulanmış ve sonuçları karşılaştırılmıştır. Kural Tabanlı Algoritma, %71.3'lük doğruluk oranı ile karşılaştırılan üç algoritma arasında en yüksek oranı elde eden algoritma olmuştur. Ayrıca literatürdeki çalışmalar içerisinde Kural Tabanlı Algoritma kullanılan tek çalışma budur. Çalışmada tahmin modelinden çıkarılan bilgiler ile öğrencilerin ilk dönemdeki başarı düzeyinin belirlenebileceği böylece başarısızlık ihtimali bulunan öğrenciler için gerekli önlemlerin alınabileceği ve gerekli iyileştirmelerin yapılabileceği belirtilmiştir. Çalışmanın sınırlı tarafının, toplanan verilerdeki eksik değerler nedeniyle küçük veri boyutu olduğu belirtilmiştir [19].

Hussain ve arkadaşları (2018) çalışmalarında Assam'da yer alan üç kolejde okuyan öğrencilerin dönem sonu başarılarını tahmin etmek amacıyla veri madenciliği yöntemlerini kullanmışlardır. Veri seti üç kolejde okuyan 300 öğrencinin 24 nitelik altında sosyo-ekonomik, demografik ve akademik bilgileri toplanarak oluşturulmuştur. Veri temizleme ile 22 niteliğe düşen öznitelik sayısı WEKA programı ile yapılan öznitelik seçimi sonrası 12'ye indirgenmiştir. Öznitelik seçimi için korelasyon tabanlı öznitelik seçimi (Correlation based - CFS), kazanç oranı öznitelik seçimi (Gain Ratio - GR), bilgi kazancı öznitelik seçimi (Information Gain - IG), simetrik belirsizlik öznitelik seçimi (Symmetrical Uncertainty - SU) kullanmışlardır. Bu öznitelik seçim yöntemleri kullanılarak toplam 11 özniteliğin oldukça etkili olduğu bulunmuştur. WEKA programında J48, PART, RF ve BayesNet olmak üzere toplam dört sınıflandırma algoritması denenmiş ve karşılaştırılmıştır. Ayrıca Apriori algoritması ek olarak veri setindeki gizli kuralları ortaya çıkarmak için kullanılmıştır. RF algoritması öznitelik seçimli ve öznitelik seçimsiz olarak veri seti üzerinde uygulanmıştır. Öznitelik seçimli olarak RF algoritmasının daha iyi performans gösterdiği belirtilmiştir. Sonuçta RF algoritması %99 doğruluk oranı ile en başarılı algoritma olmuştur [20].

Saa ve diğerleri (2019) yaptıkları çalışmada öğrenci bilgi sisteminden çıkarılan yeni bir veri kümesine dayalı olarak öğrencilerin akademik performansını tahmin etmişlerdir. Veri kümesi Birleşik Arap Emirlikleri'ndeki özel bir üniversiteden alınmıştır. Veri kümesi, öğrencilerin bilgileriyle ilgili 34 öznitelik ve 56.000 kayıt içerir. RapidMiner yazılımını kullanmışlardır. DT, RF, Gradyan Artırılmış Ağaçlar (Gradient Boosted Trees - GBT), Derin öğrenme, NB, LR, Genelleştirilmiş Doğrusal Model (Generalized Linear Model - GLM) algoritmaları ile çalışmışlardır. IG öznitelik seçim tekniklerini kullanmışlardır. Öğrencilerin akademik performansı üzerinde doğrudan etkisi olan en önemli özelliklerin, öğrencilerin demografik özellikleri, öğrencilerin önceki performans bilgileri, ders ve öğretmen bilgileri ve öğrenci genel bilgileri olmak üzere dört ana kategoride olduğu belirtilmiştir. Öznitelik seçimi ile 17 öznitelik belirlenmiş ve bu öznitelikleri ile modelleri kurmuşlardır. Genel olarak, öğrencilerin akademik performansını tahmin etmek için en uygun veri madenciliği algoritmasının %75.52 doğruluk oranı ile RF olduğu belirtilmiştir [21].

Fernandes ve çalışma arkadaşları (2019) çalışmalarında 2015-2016 eğitim öğretim yılında Brezilya Federal Bölgesi'ndeki devlet okullarında öğrenim gören öğrencilerin akademik performanslarının tahminini gerçekleştirmeyi ve öğrenci başarısızlıklarını düşürmeyi amaçlamışlardır. Öğrencilerin performansının erken tahminiyle öğrencilerin mezuniyetlerini garanti altına alacak ve yüksek öğrenim için notlarını en azından geçer bir seviyeye yükseltmelerine yardımcı olacak şekilde gerçekleşmesinin sağlanabileceği belirtilmiştir. Çalışma lise üçüncü sınıftaki öğrencileri kapsamaktadır. Çalışılan veriler, Brezilya Federal Bölgesi Eğitim

Bakanlığı'nın bir havuzundan elde edilmiştir. Başlangıçta bu verilerin nicel bilgilerini elde etmek için tanımlayıcı bir istatistiksel analiz yapılmıştır. Bu aşamada, 17 özneliği kapsayan bir analiz gerçekleştirilmiştir. Bunlar okulun bulunduğu şehir, okul adı, sınıf dönemi, bir sınıftaki özel gereksinimli öğrencilerin varlığı, sınıf türü, öğrencinin adı, cinsiyet, yaş, öğrencinin ikamet ettiği yer, varsa öğrencinin özel ihtiyaçları, ilk iki ayın notları, devamsızlıklar ve öğrencinin okul yılını geçip geçmediği ve benzeri özellikleridir. Bu analizden sonra iki veri seti elde edilmiştir. İlk veri seti, okul yılının başlamasından önce elde edilen özneliliklerdir. İkincisi veri seti ise ilk veri setine ek olarak sömestrin başlamasından iki ay sonra toplanan devamsızlık, notlar ve okul dersleri gibi akademik değişkenleri içermektedir. Her veri seti için öğrenci performansını tahmin etmek üzere artırma tekniği ile geliştirilen karar ağacı modeli olan Gradyan Artırılmış Makine (Gradient Boosting Machine - GBM) ile sınıflandırma modelleri oluşturulmuştur. Oluşturulan sınıflandırma modelleri, Alıcı çalışma karakteristiği (Receiver Operating Characteristic - ROC) eğrisi kullanılarak değerlendirilmiştir. Sonuçta, notlar ve devamsızlık niteliklerinin yılsonu öğrenci performansını tahmin etmek için en önemli değişkenler olduğu belirtilmiştir. Artırma tekniği kullanılarak %99 doğruluk oranı elde etmişlerdir [22].

Aydın ve Özkul (2015) yaptıkları çalışmada Anadolu Üniversitesi açık öğretim sisteminde eğitim gören öğrencilerin performanslarını tahmin etmeye yönelik veri madenciliği uygulaması geliştirmişlerdir. Kullanılan veri seti 2004-2005 ve 2005-2006 öğrenim yıllarına ait öğrenci bilgi sistemi ve e-öğrenme sisteminden elde edilmiştir. Veri seti 180.554 öğrencinin 129 adet dersine ilişkin 429.757 kayıt içermektedir. SPSS Clementine yazılımında öznelilik seçme teknikleri uygulamışlardır. Öğrencinin akademik başarısını etkileyen 11 öznelilik belirlenmiştir. Bu öznelilikler: öğrencinin yaşı, ders adı, öğrencinin dersi kaçınıcı kez aldığı, e-hizmet faydalanma süreleri, sınavlarının ortalaması, deneme sınavlarında yaptığı doğru cevap sayısı ve deneme sınavlarındaki başarı durumudur. Öğrenci performansını değerlendirmeye yönelik tahmin modellerini oluşturmak için SPSS Clementine yazılımında LR, YSA, DT algoritmalarından olan C5.0, CHAID, C&RT ve QUEST modelleri oluşturulmuştur. Sonuçta %82.1 doğruluk oranı ile C5.0 ile elde edilen DT modelinin en iyi tahmin modeli olarak seçildiği belirtilmiştir [23].

Sokkhey ve Okazaki (2020) çalışmalarında eğitsel veri madenciliği kullanarak düşük performans gösteren öğrencileri tahmin etmeyi amaçlamışlardır. Bunun için Karma rastgele orman (Hibrit Random Forest) eğitsel veri madenciliği modelini önermişlerdir. Önerdikleri bu modeli web tabanlı bir sisteme dâhil etmişlerdir. Çalışmada kullanılan veriler; Kamboçya'daki farklı liselerden elde edilmiştir. Çalışmada öncelikle tahmin modellerinin karşılaştırmalı bir çalışması yapılmıştır. Karşılaştırmalar hem orijinal veri seti ile hem de farklı öznelilik seçme yöntemleri ile oluşturulmuş veri setleri ile gerçekleştirilmiştir. Bu amaçla yapılan dört sınıflandırıcının (KNN, Hibrit C5.0, Karma rastgele orman ve Geliştirilmiş derin inanç ağları) karşılaştırmalı deneyi, dört öznelilik seçme yönteminden (IG, Ki kare, Karşılıklı bilgi ve önerilen MICHİ yöntemi) elde edilen özellik kümeleri kullanılarak gerçekleştirilmiştir. Öğrenme çıktılarını iyileştirmek, uygun öznelilikler seçmek ve tahmin modellerinin performansını artırmak için karşılıklı bilgi (Mutual Information-MI) ve sıralı özellik puanlarına dayalı ki-kare algoritmalarının birleşimi olan MICHİ adlı yeni bir özellik seçme yöntemi tanımlanmıştır. Deneysel sonuçlara göre karma rastgele orman algoritmasının sınıflandırma sonucu diğer üç sınıflandırıcıdan daha iyi performans gösterdiği belirtilmiştir. Karma rastgele orman algoritması MICHİ öznelilik seçme yöntemiyle birlikte %99.98 doğruluk ile en başarılı sınıflandırma olmuştur. MICHİ'nin öğrenci performansının başarısı için baskın faktörleri belirlemede etkili bir özellik seçme yöntemi olduğu belirtilmiştir. Veri ön işleme için R programlama dili, R Studio geliştirme ortamında kullanılmıştır [24].

Mengash (2020) yaptığı çalışmada yükseköğretim kurumlarında akademik olarak iyi performans göstermesi muhtemel adayları seçmek için; başvuru sahiplerinin üniversitedeki akademik performansını tahmin etmek üzere veri madenciliği tekniklerini kullanmıştır. Veri seti olarak 2016-2019 arasında bir Suudi devlet üniversitesinin Bilgisayar Bilimi ve Bilgi Koleji'ne kaydolun 2.039 öğrenciye ait veriler kullanılmıştır. Çalışmada öznelilik seçim teknikleri kullanmışlardır ve sonucunda lise not ortalaması, scholastic achievement admission test puanı ve genel yetenek testi puan bilgileri sonuca en çok etki eden öznelilikler olarak

belirlemişlerdir. Çalışmada YSA, DT, SVM ve NB olmak üzere dört iyi bilinen veri madenciliği tekniği kullanılarak dört tahmin modeli geliştirilmiştir. Bu modeller arasında YSA algoritmaları ile %79'un üzerinde bir doğruluk oranına sahip olarak en başarılı modeli elde etmişlerdir [25].

Çiftçi ve arkadaşları (2018) çalışmalarında üniversitede çalışan akademisyenlerin performanslarını tahmin etmişlerdir. Veri seti olarak "Türkiye Öğrenci Değerlendirme Veri Seti (Türkiye Student Evaluation Data Set)" isimli bir veri seti kullanılmıştır. Veri seti 2.850 öğrencinin 3 akademisyeni değerlendirdiği anket sonuçlarını içermektedir. Öğrencilerin 13 farklı dersini bu 3 akademisyen vermektedir. Anket çalışması 2013 yılında 33 nitelik üzerinden gerçekleştirilmiştir. Veri setinde yer alan 33 tane nitelik genetik algoritma ile öznitelik seçimi yapılarak çalışmada kullanılmak üzere 19 öznitelige indirgenmiştir. Bilgi kazancı, genetik algoritma ve geri yönlü eliminasyon olmak üzere üç farklı öznitelik seçme yöntemi kullanmışlardır. En başarılı yöntemin genetik algoritma olduğunu belirtmişlerdir. Çalışmada, Rapidminer veri madenciliği yazılımı kullanılmıştır. Makine öğrenmesi yöntemleri olarak KNN, DT, NB ve derin öğrenme algoritmalarını uygulamışlardır. İndirgenen öznitelikler üzerinden derin öğrenme kullanımı ile %97.70'lik tahmin doğruluk performansına ulaşılmıştır [26].

Kumari ve arkadaşları (2020) çalışmalarında YSA algoritmasının veri madenciliğinde öğrenci performansını tahmin etmek için uygun bir sınıflandırıcı olup olmadığını araştırmışlardır. Kaggle sitesinden alınan sınavlarda öğrenci performansı (students performance in exams) adlı hazır veri setini kullanmışlardır. Bu veri seti; Amerika'da lise öğrencilerinden elde edilen verilerden oluşmaktadır. Veri seti, üç sınavdan alınan puanları ve sınav puanları üzerinde etkisi olduğu düşünülen öğrencilere ait çeşitli kişisel, sosyal ve ekonomik faktörleri içermektedir. Veri setinde 8 adet öznitelik yer almaktadır. Bu öznitelikler; cinsiyet, etnik köken, ebeveyn eğitim düzeyi, öğle yemeği, sınava hazırlık kursu, matematik, okuma ve yazma puanıdır. YSA algoritmasının uygulanabilirliğini araştırmak için NB, KNN, RF, DT, SVM ve LR olmak üzere altı sınıflandırıcıyla performans açısından karşılaştırılmıştır. %89.84 doğruluk oranı ile YSA algoritmasının ilgili veri setinde diğer sınıflandırma algoritmalarına göre performans olarak daha iyi olduğu ve öğrenci performansını tahmin etmek için etkili bir algoritma olduğunu belirtmişlerdir [27].

Kaur ve diğerleri (2015) sınıflandırma tabanlı algoritmalar kullanan tahmine dayalı bir veri madenciliği modeli ile lise öğrencileri arasında yavaş öğrenenleri tespit etmeyi amaçlamışlardır. Çalışmada 152 öğrenciye uygulanan anketten elde edilen veriler kullanılmıştır. Tüm uygulamalar WEKA aracı kullanılarak yapılmıştır. İlk önce veriler üzerinde öznitelik seçim teknikleri uygulanmıştır. Bu teknikler Ki kare, IG, SU ve ReliefF öznitelik seçim teknikleridir. Bu veriler üzerinde belirlenen öznitelikler kullanarak Çok katmanlı algılayıcı (Multilayer Perception - MLP), NB, Sıralı minimal optimizasyon (Sequential minimal optimization - SMO), J48 ve REPTree gibi çeşitli sınıflandırma algoritmaları uygulanmış ve sonuçlar karşılaştırılmıştır. En iyi sonucu %75 doğru tahmin yüzdesiyle MLP algoritmasının verdiği belirtilmiştir [28].

Yıldız Aybek ve Okur (2018) çalışmalarında, Anadolu Üniversitesi Açık öğretim Sisteminde 2014-2015 ve 2015-2016 akademik yıllarında Bilgi Teknolojileri dersini alan öğrencilerin yarıyıl sonu sınav puanlarını ve geçme oranlarını YSA ile tahmin etmeyi amaçlamışlardır. Bu amaçla 626.478 öğrencinin demografik özellikleri, eğitim durumu, ilgili dersin ara sınav, final ve başarı puanları ile ilgili veriler toplanmıştır. Veri ön işleme sonrasında eksik verilere sahip örneklerin çıkarılması sonucu 195.584 öğrencinin verileri MLP ve Radyal Taban Fonksiyonu (Radial Base Function - RBF) modelleri ile analiz edilmiştir. Veri setinin %70'i alınarak eğitim seti, %30'u alınarak test seti oluşturulmuştur. 6 parametrenin (gizli katman, minimum birim, maksimum birim, eğitim, etkinleştirme işlevi ve sürekli değişkenlerin yeniden ölçeklendirilmesi) farklı kombinasyonlarından oluşan 12 MLP ve 3 parametrenin (gizli katman, etkinleştirme fonksiyonu ve sürekli yeniden ölçeklendirme değişkenleri) çeşitli kombinasyonlarından oluşan 4 RBF oluşturulmuştur. Oluşturulan 16 ağ, öğrencilerin final sınav puanlarını tahmin etmek ve öğrencilerin geçme/kalma durumunu tahmin etmek için iki kez çalıştırılmıştır. Toplam 32 analiz yapılmıştır. İlk analiz sonucunda; 12 MLP ağı ile yapılan analizlerde tahmin edilen ve gerçekleşen puanlar arasında orta düzeyde korelasyon katsayıları görülmüştür.

RBF ağlarının ise final sınav puanlarını açıklamada tutarlı sonuçlar vermediği belirtilmiştir. İkinci analizde; RBF ağlarının geçme/kalma durumlarını veya öğrencileri sınıflandırmada başarılı olmadığı tespit edilmiştir. MLP ağlarının ise genel olarak daha başarılı olduğu belirtilmiştir. Bu çalışmada YSA modellerinin diğer matematiksel modellerle karşılaştırılması amaçlanmasa da farklı parametrelerle elde edilen %85-%87 arasında değişen tahmin doğruluk oranlarının elde edildiğini belirtmişlerdir [29].

Özbay ve Ersoy (2017) çalışmalarında matematik bölümü mezunu öğretmen adaylarının Öğrenme Yönetim Sistemi üzerindeki hareketliliği ile akademik başarıları arasındaki ilişkiyi veri madenciliği yöntemleri ile incelemişlerdir. Çalışmada 2012–2013 öğretim yılı bahar döneminde Başkent Üniversitesi’nde pedagojik formasyon eğitimi alan 40 öğrencinin öğrenme yönetim sistemine ait günlük kayıtları ve yılsonu akademik başarı notları kullanılmıştır. Akademik başarı notları ilgili öğretim elemanından temin edilmiştir. Veri seti üzerinde veri ön işleme işlemleri yapılmıştır. Veri setinde toplam 5,358 öğrenciye ait ve 8 öznitelik vardır. Çalışmada, öncelikle öğrenci akademik başarı gruplarının oluşturulması amacıyla tanımlayıcı modellerden kümeleme yöntemi kullanılmıştır. K-Ortalama algoritması kullanılarak öğrenci yılsonu akademik başarı notlarına göre kümeleme işlemi gerçekleştirilmiştir. Öğrenciler düşük, orta ve yüksek akademik başarı grubu olmak üzere 3 gruba ayrılmıştır. Daha sonra bu başarı grupları da girdi değişkeni olarak kullanılarak sınıflandırma yöntemlerinden karar ağacı yapısı ile öğrenci başarı durumları tahmin edilmiştir. Karar ağacı C5.0, CART, CHAID, QUEST algoritmaları kullanılarak modeller oluşturulmuştur. Oluşturulan modeller arasında en yüksek doğruluk oranını %85 olarak elde eden CART algoritması olduğu belirtilmiştir. Analizlerde SPSS Clementine programı kullanılmıştır. Sonuç olarak öğrenme yönetim sisteminde yüksek düzeyde hareketli olan öğrencilerin yüksek akademik başarı grubunda, orta düzeyde hareketli olan öğrencilerin orta akademik başarı grubunda ve düşük düzeyde hareketli olan öğrencilerin düşük akademik başarı grubunda yer aldıklarının tespit edildiği belirtilmiştir [30].

Baltacı (2018) çalışmasında, İlahiyat fakültesi öğrencilerine yönelik veri madenciliği alanında bir uygulama gerçekleştirmiştir. Öğrencilerin yaşı, cinsiyeti, anne ve babanın eğitim durumu, liseyi bitirdikten sonra kaçınıcılığında üniversiteye yerleştiği, yerleştikleri fakülte, tercih sırası, yerleşme puanı, üniversitede aldığı derslerin not ortalaması, üniversitede başarısız ders sayısı ve gibi verilerinden faydalanılarak, akademik başarıya etki eden faktörler bulunmak istenmiştir. Kullanılan veriler 2017 yılında 11 farklı devlet üniversitesinin İlahiyat fakültelerinde öğrenim gören 2. Sınıf öğrencilerinden temin edilmiştir. Veri seti üzerinde ön işleme olarak veri temizleme, veri dönüştürme ve veri ayrıştırma işlemleri yapılmıştır. Veri ön işleme sonucunda toplam 3,775 öğrenciye ait veri, makine öğrenmesi algoritmalarında kullanılmak üzere hazır hale getirilmiştir. Öğrencinin akademik başarısı, öğrencinin not ortalaması ve başarısız olunan ders sayısına göre incelenmiştir. Öğrencinin başarısız olduğu ders sayısı ve yılsonu not ortalaması bağımlı değişken olarak belirlenmiştir. Çalışmada veri madenciliği tekniklerinden kümeleme analizi, gruplama yöntemi, regresyon ve korelasyon ilişkilendirme teknikleri kullanılmıştır. İstatistik anlamlılık düzeyi $p < 0.05$ olarak kabul edilmiştir. Veri madenciliği analiz sonucuna göre yılsonu not ortalaması ve başarısız olunan ders sayısı arasında yüksek düzeyde negatif yönlü bir ilişki olduğu gözlemlenmiştir. Not ortalamasındaki artış, öğrencinin başarısız olduğu ders sayısının azalmasına neden olmaktadır. Ayrıca yılsonu not ortalaması ile yaş arasında da yüksek düzeyde negatif ilişki belirlenmiştir. Yaşı daha büyük öğrencilerin yılsonu not ortalamasının daha düşük olduğu belirtilmiştir. Böylelikle öğrencinin yaşı, not ortalamasını olumlu yönde etkilemediği görülmüştür. Ayrıca öğrencinin yaşı, başarısız olduğu ders sayısını olumlu yönde etkilediği de belirtilmiştir. Çalışma için “Orange Datamining”, “D-Melt” ve “SPSS Modeller” veri madenciliği programları kullanılmıştır [31].

Hazel Başer ve arkadaşları (2020) çalışmalarında, ortaöğretimde öğrenci başarısını olumsuz etkileyen faktörlerin belirlenmesi ve öğrencilerin karşılaştıkları sorunların tespit edilmesini amaçlamışlardır. Çalışmada kullanılan veri seti “UC Irvine Machine Learning Repository” veri tabanından elde edilmiştir. Veri seti 395 öğrencinin 33 özneliğinden oluşmaktadır. İlgili veriler okul raporu ve anketlerden elde edilmiştir. Veri içeriğindeki öznitelikler öğrencinin; yılsonu ders notları, kişisel bilgileri, okul içi ve dışı sosyal etkinliklerini kapsamaktadır. Veri seti üzerinde veri madenciliği yöntemlerinden Classifier, OneR, LogitBoost

ve YSA ile modeller kurulmuştur. Değerlendirmelerin objektif olarak yapılabilmesi için her sınıflandırma modeli 3 ayrı çapraz doğrulama (5, 7 ve 10 kat) bölünme yöntemi kullanılmıştır. Tüm modeller WEKA yazılımı kullanılarak gerçekleştirilmiştir. Modellerin değerlendirilmesi için doğruluk, özgünlük ve duyarlılık performans metrikleri kullanılmıştır. Sonuç olarak çalışmada uygulanan OneR algoritması en başarılı veri madenciliği yöntemi olmuştur. Doğruluk değerinin ortalama %92.15, özgünlük değerinin ortalama %86.6, duyarlılık değerinin ise ortalama %96 olduğu belirtilmiştir [32].

Can ve çalışma arkadaşları (2019) çalışmalarında veri madenciliği tekniklerinden LR analiziyle öğrencilerin ders başarılarını etkileyen faktörleri araştırmışlardır. Çalışmada, bir devlet üniversitesine ait dönem sonu ders değerlendirme anket verilerini içeren hazır bir veri seti kullanılmıştır. Veri seti, 5.820 öğrenci tarafından yanıtlanan anket verilerinden elde edilmiştir. Veri seti, öğrencilerin derslerine giren 3 farklı öğretim üyesi hakkında tanımlanmış 28 soru ve 5 tane de demografik bilgileri içeren kategorik değişken olmak üzere toplam 33 öznitelikten oluşmaktadır. Çalışmalarında kullandıkları veri setine LR algoritması ile sınıflandırma modeli kurmuşlardır. Elde edilen sonuçlara göre, öğrenci ders başarısını ders tekrar sayısının etkilediği belirtilmiştir. Ayrıca öğrencilerin mesleki gelişimlerini arttıran, öğrencilere yeni bakış açıları kazandıran uygulamaların ve derse katılımın ders başarısını olumlu olarak etkilediğinin görüldüğü belirtilmiştir. LR analizi sonucunda yapılan tahminlerde doğru sınıflandırma oranını %84.1 gibi oldukça yüksek bir değer elde etmişlerdir. Yani 5,820 adet verinin 4,895 tanesi doğru sınıflandırılırken yalnızca 925 tanesi yanlış sınıflandırılmıştır [33].

Gladshiya ve Sharmila (2021) çalışmalarında eğitim alanında öğrencilerin performansı tahmin etmek için öğrencilerin riskini belirlemeyi esas almışlardır. Öğrenci risklerini belirleyerek performansları makine öğrenimi algoritmaları kullanılarak tahmin etmeyi amaçlamışlardır. Öğrencinin performansını tahmin etmek için veri ön işleme ve özellik seçiminin çalışma yöntemlerini karşılaştırmışlardır. Performance of student isimli veri setini, www.kaggle.com adresinden temin etmişlerdir. Veri seti 1,000 satır ve 8 öznitelik içeren virgülle ayrılmış değer (csv) dosya biçimindedir. Nitelikler; cinsiyet, kentsel durum, ebeveyn eğitimi, öğle yemeği, ölçek hazırlık, matematik puanı, fen ve okuma puanlarıdır. Veri setinde öznitelik seçimi yöntemlerinden sınıflandırma ve filtreleme metotları kullanılmıştır. Bu yöntemlerden rastgele n-sınıf sınıflandırıcı, Varyans analizi filtresi, Select K-Best yöntemleri ile öznitelikler belirlenmiştir. Select K-Best öznitelik seçim metodunu önermektedirler. Belirlenen öznitelikler ile SVM algoritması kullanılarak veri seti eğitilmiş ve tahmin değerleri elde edilmiştir. Bu algoritma, Jupyter Notebook kullanılarak Python programlama dili kullanılarak uygulanmışlardır. Öznitelik seçimleri yapılarak doğruluk oranlarında önemli ölçüde farklılık gösterdiği belirtilmiştir. SVM algoritmasıyla rastgele n-sınıf sınıflandırıcı ile 3 sınıf için tahmin edilen doğruluk oranı %52 iken 4 sınıf için tahmin edilen doğruluk oranı %72' ye çıktığını belirtmişlerdir [34].

Abbasoğlu (2020) çalışmasında, eğitsel veri madenciliği yöntemleri kullanarak, ortaokul öğrencilerinin sosyoekonomik durumlarının ve demografik özelliklerinin yılsonu genel başarı ortalamalarına olan etkilerini analiz etmeyi amaçlamıştır. Örneklem olarak 2019-2020 eğitim-öğretim yılı ikinci dönemi içinde Yalova iline ait 4 farklı ortaokuldaki 5, 6, 7 ve 8. Sınıf düzeylerinde okuyan 1,395 ortaokul öğrencisi seçilmiştir. Çalışmada, E-okul Yönetim Bilgi sisteminden elde edilen öğrencilerin kişisel özelliklerine ilişkin veriler kullanılmıştır. Sistemden elde edilen veriler 27 bağımsız değişken ile temsil edilmiştir. Tahmin sonuçlarını artıracığı düşünülen korelasyona dayalı öznitelik seçim teknikleri kullanılarak 27 bağımsız değişken 8'e düşürülmüştür. Elde edilen veri seti üzerinde makine öğrenmesi algoritmaları kullanılarak yılsonu genel başarı ortalamaları tahmin edilmiştir. Tüm algoritmalar WEKA programında çalıştırılmıştır. Bu algoritmalar; LR, NB, Linear SVM, KNN, RF, Non-Linear SVM, Torbalama (Bagging), MLP algoritmalarıdır. Veri seti öznitelik seçimi yapılmadan, temel bileşen analizi yapılarak ve öznitelik seçimi uygulanarak 3 farklı veri seti üzerinden analiz edilmiş ve sonuçlar karşılaştırılmıştır. Modellerin performanslarını değerlendirmek için doğruluk oranı kullanılmıştır. Çalışmalarında, en iyi sonuca ulaşabilmek için 10 kat çapraz doğrulama ile bölünme yöntemi kullanarak modeller oluşturulmuştur. Deneysel sonuçlara göre sınıflandırma yöntemlerinden en başarılı

sonucu %62 ile %64.5 arasında doğruluk oranları ile LR algoritmalarıyla kurulan modellerden elde etmişlerdir [35].

Aghalarova ve Keser (2022) çalışmalarında öğrencilerin akademik performanslarını tahmin etmeyi amaçlamışlardır. Çalışmada kullanılan veri seti için otomatik makine öğrenmesi yöntemi ile en iyi modeli araştırmışlardır. Veri önileme, model seçimi ve hiper-parametre optimizasyonu gibi görevlerle uğraşmadan en iyi modelin otomatik makine öğrenmesi ile bulunabileceğini savunmuşlardır. Çalışmalarında H2O AutoML aracı kullanmışlardır. Tüm deneyler, Python'un 3.7.9 sürümü kullanılarak gerçekleştirilmişlerdir. Çalışmada kullanılan veri seti Kalboard 360 E-öğrenme sisteminden toplanmıştır. Veri seti 480 öğrenciye ait 17 öznelikten oluşmaktadır. Öznelikler; demografik, akademik ve davranışsal öznelikler olmak üzere üç esas kategoridedir. AutoML sayesinde veri önileme ve en uygun modelin seçimi otomatik gerçekleştirilmektedir. Dağıtılmış RF (Distributed Random Forest - DRF) algoritmasında varsayılan hiper-parametreler kullanıldığında doğruluk değerini %77.50 ve f-skor değeri %80.01 olarak elde etmişlerdir. Bu sebeple dağıtılmış RF algoritmasının en iyi algoritma olduğunu belirtmişlerdir. En uygun hiper-parametreleri bulabilmek için ızgara araması kullanmışlardır. Izgara araması ile bulunan en uygun hiper-parametreler ile tekrar dağıtılmış RF algoritması uygulandığında doğruluk değeri %82.30 ve f-skor değeri %82.50 olarak hesaplamışlardır. Önerilen yöntem ile geleneksel makine öğrenme yöntemlerinden KNN ve SVM algoritmaları karşılaştırılmıştır. KNN algoritmasında doğruluk değeri %61.1 ve f-skor değeri %61.01, SVM algoritmasında ise sırasıyla doğruluk değeri %64.5 ve f-skor değeri %64.4 olarak elde etmişlerdir. Elde edilen sonuçlar önerdikleri yöntemin etkinliğini ve başarısını kanıtlamaktadır [36].

Aydemir (2019) çalışmasında, öğrencinin başarı durumunu önceden öngörerek gerekli önlemlerin alınabileceğinden yola çıkarak öğrencilerinin Yabancı Dil dersinden geçme notlarını makine öğrenmesi yöntemleriyle tahmin etmeyi amaçlamıştır. Bu amaçla 2017-2018 eğitim-öğretim yılının tüm dönemlerinde Yabancı Dil dersini alan 3.794 ön lisans ve lisans öğrencisine ait veriler kullanılmıştır. Veri setinde 12 tanesi bağımsız ve bir tanesi bağımlı olmak üzere toplam 13 adet öznelik vardır. Bu öznelikler: ÖSS giriş türü, ÖSS giriş puanı, ÖSS giriş sıralaması, öğrenim tipi, fakülte, bölüm, program, program tipi, öğretmen, öğretmen unvanı, bir önceki dönemin not ortalaması, aktif dönem not ortalaması ve ders geçme notudur. Verilerin eğitim ve test olarak bölünmesinde 10-katlı çapraz doğrulama yöntemi kullanılmıştır. Çalışmada YSA, M5P, Karar Kütüğü, M5Rules, Karar Tabloları, Torbalama yöntemleri kullanılarak tahmin modelleri geliştirilmiştir. Tüm modeller WEKA programı aracılığıyla ile kurulmuştur. Veriler 19 farklı makine öğrenmesi yöntemleri kullanarak modeller kurulmuş, analiz edilmiş ve tahmin sonuçları elde edilmiştir. Tüm modellerin başarı oranları karşılaştırılmıştır. Torbalama yöntemi ile kurulan modelin 0.80 korelasyon katsayısı ve 1.22 ortalama mutlak hata ile tahmin yaparak en iyi sonucu verdiği belirtilmiştir [37].

Assistant ve çalışma arkadaşları (2021) çalışmalarında öğrencinin akademik performansını tahmin etmek için çapraz doğrulama yöntemi uygulayarak bazı öznelik seçim algoritmalarında kullanılan sınıflandırma algoritmaları ile temel sınıflandırma algoritmalarının tahmin sonuçlarını karşılaştırmışlardır. Öznelik seçim algoritmalarından, Korelasyon Öznelik Değerlendiricisi (Correlation Attribute Evaluator - CAE), Bilgi Kazancı Öznelik Değerlendiricisi (Information Gain Attribute Evaluator - IGAE) ve Kazanç Oranı Öznelik Değerlendiricisi (Gain Ratio Attribute Evaluator - GRAE) yöntemlerini kullanmışlardır. Bu sayede, sonucu tahmin etmek için herhangi bir madencilik sürecine katkıda bulunmayan özellikleri kaldırmayı amaçlamışlardır. Bu çalışmada halka açık iki veri kümesi kullanmışlardır. Veriler iki farklı Portekiz lisesinden gelen ortaokul öğrencileri tarafından derlenmiştir. Öğrencilerin notlarının yanı sıra sosyal, demografik ve okulla ilgili bilgilerin tümü veri toplamaya dâhil edilmiş. Her iki veri kümesi de toplam 33 öznelik içerir. WEKA üzerinde uygulanan sınıflandırma algoritmaları NB, RF, J48, MLP, JRip, LR ve Karar tablosu algoritmalarıdır. 10 kat çapraz doğrulama yöntemi ile farklı sınıflandırma algoritmalarının uygulama sonuçlarını karşılaştırmışlardır. Tüm özellikler kullanıldığında karar tablosu algoritması, %76.11 ile en yüksek doğruluk oranını elde etmiştir. Öznelik değerlendirici algoritmaları uygulandıktan sonra tahmin sonucu

üzerinde en fazla etkiye sahip olan on farklı öznelik seçilmiştir. Araştırmacıları bu işlemin doğruluk yüzdelerinde yüzde 10 ila 15 fark oluşturduğunu belirtmişlerdir [38].

Butto ve arkadaşları (2020) çalışmalarında öğrencilerin akademik performansını makine öğrenme algoritmaları ile tahmin etmeyi amaçlamışlardır. Çalışmada kullanılan veri seti 500 öğrencinin 16 farklı öznelik bilgisini içerir. Hem WEKA aracı hemde Jupyter not defterinde çalıştırılan Python 3 kullanılarak LR ve SVM algoritmalarını kullanarak sınıflandırma modellerini kurmuşlardır. Önerilen model üzerinde uygulanan algoritmaların sonuçları ve kalitesi, hatırlama, kesinlik ve f1-skoru ve doğruluk gibi dört değerlendirme ölçüsüne dayalı olarak karşılaştırmışlardır. Jupyter aracılığıyla elde edilen sonuçlara bakıldığında LR algoritması %71 doğruluk oranına, SVM algoritması %78 doğruluk oranına ulaştığı görülmüştür [39].

Karthikeyan ve arkadaşları (2017) çalışmalarında farklı veri madenciliği tekniklerine dayalı olarak öğrencilerin performansını artırmaya yönelik bir çalışma yapmışlardır. Bu çalışma için performansı göstermek için iki farklı boyutta veri seti kullanılmıştır. Birinci veri setinin adı Student Performance Dataset olup veri seti toplam 395 kayıt ve 33 özneliğe sahiptir. Bu veriler, iki Portekiz okulunun orta öğretimindeki Matematik ve Portekiz dili derslerindeki öğrenci başarısını ele almaktadır. Veri seti raporlar ve anketler kullanılarak toplanmıştır. İkinci veri seti College Real Dataset adında, CMS College of Science and Commerce'den temin edilmiştir. Veri setinde toplam 650 kayıt, 52 tane öznelik vardır. Öğrencilerin okula girmeden önce seçilmesinde kullanılan notlarıyla ilgili değişkenleri içerir. Geliştirilmiş performans dayalı bir Öğrenci Performans Tahmin Sistemi (SPPS) tasarlayıp uygulayarak optimizasyon tekniklerine dayalı bir öznelik seçim algoritması önermişlerdir. Akademik verileri kullanarak özellik seçimi ve topluluk sınıflandırma algoritmaları kullanmışlardır. Sınıflandırma algoritmaları olarak DT, NB, YSA, Kümeleme, Birliklik kuralı madenciliği, Tahmin, Zaman serisi analizi, Büyük verilerden bilgi keşfi için Sıralı model, Genetik algoritma ve KNN teknikleri kullanmışlardır. Çalışmalarında, iki veri madenciliği olan Kümeleme (Geliştirilmiş K-Ortalamları) ve Sınıflandırma (SVM) tekniklerini birleştirmişlerdir. Bunun adına CESVM-SPPS demişlerdir. Modelin performansını artırmak için birleştirilen sınıflandırıcılar ve KNN olmak üzere üç sınıflandırıcının birleştirmesini de önermektedirler. Deneylerde Level 1 (Geri Yayılımlı sinir ağı, KNN, CESVM) ve Level 2 (CESVM) seviyelerini kullanmışlardır. Deneysel sonuçlara göre CESVM-SPPS modelinde birinci veri seti için doğruluk oranı %84, ikinci veri seti için doğruluk oranı %82 olarak elde etmişlerdir. 2 Adımlı SPPS ile birlikte Level 1 sınıflandırıcı olarak SVM ve seviye 2 Sınıflandırıcı olarak CESVM uygulandığı zaman modelde başarı oranı yaklaşık olarak birinci veri seti için %90, ikinci veri seti için %85 olduğunu söylemişlerdir. Önerilen SPPS, CESVM-SPPS olarak adlandırarak tahmin hızını artırdığını ve tahmin doğruluğunu artırdığını savunmuşlardır [40].

Shreem ve arkadaşları (2022) çalışmalarında öğrencilerin performansını tahmin sistemleri oluştururken eğitim verilerinden anlamlı bilgi çıkarmanın karmaşık bir iş olduğunu, bu nedenle veri boyutunu azaltmak gerektiğini savunmuşlardır. Çalışmalarında sarmalayıcı öznelik seçim algoritması olarak geliştirilmiş bir ikili genetik algoritma (EBGA) önermişlerdir. Bir k-ortalama algoritmasına ve elektromanyetik benzeri mekanizma yöntemine dayalı yeni hibrit seçim mekanizması önerilmiştir. Önerilen EBGA ile beş farklı sınıflandırıcı (KNN, DT, NB, SVM ve Doğrusal Ayrıcı) arasında hibrit bir makine öğrenimi yaklaşımı kullanmışlardır. Bu yazıda UCI Machine Learning Repository'den elde edilen iki gerçek vaka çalışması kullanılmıştır. Elde edilen sonuçlar, önerilen yaklaşımın ikili genetik algoritmanın performansını artırma yeteneğini göstermiştir. Ayrıca, tüm sınıflandırıcıların performansları %1 ve %11 arasında iyileştirilmiştir. En iyi sonucu ilk veri seti için NB sınıflandırıcısıyla %87 doğruluk oranı, ikinci veri seti için KNN sınıflandırıcısıyla %92 doğruluk oranı elde edilmiştir. Önerilen algoritma Matlab R2019b'de kodlanmıştır [41].

Mueen arkadaşlarıyla birlikte (2016) çalışmalarında öğrenci performansını etkileyen özellikleri tespit etmeyi, öğrencilerin akademik performanslarını tahmin etmeyi ve farklı veri madenciliği algoritmalarıyla

sınıflandırma yapıp doğruluğunu karşılaştırmayı amaçlamışlardır. Çalışmada, aynı öğretim üyesi tarafından verilen 2014-2015 eğitim yılına ait Programlama Temelleri ve Gelişmiş İşletim Sistemleri lisans derslerine kayıt yaptıran öğrencilerin 38 tane farklı özelliklerinin olduğu verileri toplamışlardır. Veri madenciliği sınıflandırması için NB, MLP ve DT (J48) olmak üzere 3 farklı algoritma kullanmışlardır ve bu işlemler için WEKA aracını kullanmışlardır. Sınıflandırma tekniklerini uygulamadan önce veri kümesinin ön işleme yapılması gerektiğini ve bu işlemlerin sonucu doğrudan etkilediğini vurgulamışlardır. Ön işleme olarak özellik seçimi, dengesiz veri dağılımı ve veri dönüşümü yapılmasını önermektedirler. Oluşturdukları modelleri sınıflandırma algoritmalarında kullanmak üzere 10 kat çapraz doğrulama kullanmışlardır. Bunlardan dokuzu modeli eğitmek için ve biri modelin testi için kullanılmıştır. Üç sınıflandırıcının tahmini performansını karşılaştırmışlardır. Tüm veri seti ile yapılan ve özellik seçimi yapıldıktan sonra oluşturulan modellerin sonuçlarına göre her iki durumda da NB sınıflandırıcısının %86 ve %85.7 tahmin doğruluğu elde ederek diğer iki sınıflandırıcıdan daha iyi performans gösterdiğini belirtmişlerdir [42].

Gajwani ve Chakraborty (2021) çalışmalarında bir eğitim veri setinin belirli özelliklerine dayalı olarak bir öğrencinin akademik performansını tahmin etmek ve öğrencilerin notlarını doğrudan etkileme olasılığı en yüksek olan özelliklerin bir alt kümesini belirlemek ve böylece tahmini daha doğru hale getirmeyi amaçlamışlardır. Kullandıkları veri setindeki nitelikler öğrencilere ait demografik, davranışsal ve akademik bilgilerdir. Öznitelik seçiminden sonra çeşitli makine öğrenimi algoritmaları kullanılarak öğrenci performanslarını değerlendirmişlerdir. DT, LR, NB sınıflandırıcısı gibi denetimli öğrenme algoritmaları ve Torbalama, RF sınıflandırıcısı gibi topluluk makine öğrenimi algoritmaları ile deneyler yapmışlardır. Topluluk makine öğrenimi algoritmaları, bir veya daha fazla temel algoritmayı birleştirerek tahminin doğruluğunu olumlu yönde geliştirdiğini belirtmişlerdir. Yükseltme, Oylama ve RF sınıflandırıcıları veri kümesinde en iyi sonuçları vermiştir. Yükseltme %75'lik bir tahmin doğruluğu elde ederken, Oylama ve RF sınıflandırıcılarının her ikisi de %74.31'lik bir doğruluk elde etmiştir. Seçilen özelliklerden herhangi birinin çıkarılmasının sınıflandırıcıların doğruluğunu azalttığını belirtmişlerdir [43].

Bertolini ve arkadaşları (2021) çalışmalarında öğrencilerin fen dersi için başarılarını tahmin etmeyi, öğrenme ortamındaki eğitim kaygılarını ele almayı ve risk altındaki öğrencileri tespit etmeyi amaçlamışlardır. Kullanılan veri seti 3.225 öğrenciye ait bilgiler içermektedir. Veri setinde toplam 57 özellik vardır. Veri seti üzerinde dört veri madenciliği algoritması ve öznitelik seçim tekniklerinden 4 filtre özelliği kullanmışlardır. Bunlar CFS, Fisher puanlama algoritması (FSA), IG ve Relief özellik değerlendirmesidir. Tüm teknikler R ara yüzünde uygulanmıştır. Kullanılan sınıflandırma algoritmaları ise LR, Elastik ağ regresyonu, RF ve Aşırı gradyan artırma (XGBoost) algoritmalarıdır. Çalışmada, verilerin çapraz doğrulama adımıyla özellik seçimini bütünleştiren bir yöntem kullanmışlardır. Ayrıca filtre özelliği seçim teknikleri yine kullanarak verilerin alt kümeleri üzerindeki etkili özellik kümelerini karşılaştırılmıştır. Öznitelik seçim teknikleri CFS ve FSA'nın üniversite biyolojisi için veri setinde veri madenciliği yöntemlerinden LR ve Elastik ağ regresyonu için ortalama ROC eğrisi altında kalan alanı (Area Under Curve-AUC) önemli ölçüde iyileştirdiği gözlemlenmiştir. Ortalama AUC değerleri 0.90'ı aştığı belirtilmiştir. Sonuç olarak filtre yöntemlerinin veri madenciliği metodlarının performansını artırma potansiyeline sahip belirtmişlerdir [44].

Aghalarova ve Keser (2021) çalışmalarında ortaokul öğrencilerinin akademik performansını tahmin etmek amaçlamışlardır. Bunun için YSA algoritması önermektedirler. Çalışmalarında ortaokul öğrencilerinin akademik performanslarını Portekizce ve Matematik dersleri üzerinden sınıflandırma yöntemleri ile tahmin etmişlerdir. Tüm yöntemler Python programlama dilini kullanılarak gerçekleştirmişlerdir. İlk etapta veri seti üzerinde ön işleme işlemi yapmışlardır. Veri setinde kategorik değerler içeren öznitelikler sayısal değerlere dönüştürülmüştür. Veri setindeki dengesiz sınıf dağılımı için SMOTE yöntemi kullanmışlardır. Çalışmalarında kullanılan rastgele arama algoritması ile YSA modelinde kullanılacak en uygun hiperparametre setini belirlemişlerdir. Modelde sigmoid, ReLU ve Softmax aktivasyon fonksiyonları kullanmışlardır. Kayıp fonksiyonu olarak binary_crossentropy ve sparse_categorical_crossentropy kullanmışlardır. Optimizasyon fonksiyonu olarak RMSProp kullanmışlardır. Modeli değerlendirmek için

performans ölçütleri olan doğruluk, kesinlik, f-ölçüsü ve duyarlılık değerlerini kullanmışlardır. Algoritmanın başarısı hem öznitelik seçim işlemi olmadan tüm veri seti ile hem de öznitelik seçim işlemi ile indirgenen veri setleri ile sınıflandırma yapılarak tüm sonuçları değerlendirmişlerdir. Öznitelik seçme işlemi için öz yinelenmeli özellik çıkarma ve filtreleme yöntemleri kullanmışlardır. Öğrencilerin Portekizce ve Matematik derslerindeki başarıları 2-seviyeli ve 5-seviyeli sınıflandırma tekniği ile önerilen algoritma kullanılarak tahmin edilmektedir. Deney sonuçlarında, önerilen algoritma ile Matematik dersi için doğruluk değerleri 2-seviyeli için %97 ve 5-seviyeli sınıflandırma için doğruluk değerleri %92.3 elde edilmektedir. Portekizce dersi için ise bu değerler 2-seviyeli için %97.6 ve 5-seviyeli sınıflandırma için %87.9 olarak elde etmişlerdir. Sonuç olarak öznitelik sayısı arttıkça doğruluğun çok az fark ile değişiklik gösterdiğini belirtmişlerdir. Tüm modellerde 10 kat çapraz doğrulama uygulanmıştır. Rastgele arama algoritması ile YSA modelinin hiper-parametreleri optimize edildiğinde hem Portekizce hem de Matematik dersi için oluşturulan veri seti üzerinde yapılan 2-seviyeli ve 5-seviyeli sınıflandırma sonuçlarının iyileştiğini gözlemlemişlerdir [45].

Özkan ve arkadaşları (2018) öğrencilerin akademik başarısının değerlendirilmesinde bir döneme ilişkin akademik performansları; demografik, sosyal ve çevreyle ilgili diğer bazı özelliklerine ilişkin veriye dayanarak, yapay zekâ algoritmaları yardımıyla modellenmesini ve öngörülerin geliştirilmesini amaçlamışlardır. Çalışmalarında UCI Machine Learning Repository’de erişime açık olan “student performance” veri setini kullanmışlardır. Veri seti, Portekiz’de bir ortaöğretim okulunda 395 öğrencinin bilgilerini içermektedir. Veri kümesinde 33 öznitelik yer almaktadır. Model performans geçermesi için veri kümesinin tesadüfi olarak %90’ı modellerin eğitimi, kalan %10 ise test amacıyla ayırmışlardır. Eğitim verisiyle birlikte C5, Boosted C5, Regresyon Ağaçları, SVM, LR, RF, Derin öğrenme algoritmaları kullanılmış ve modellerin kodlaması için R programlama dilinden yararlanmışlardır. Çalışmada özgün bir derin öğrenme modeli tasarlanmış ve bu modelde girdi ile çıktı katmanları haricinde üç adet gizli katman kullanılmıştır. Bu katmanlardaki sinir hücresi sayısı 256 olarak belirlenmiş aktivasyon fonksiyonu olarak da “ReLU” kullanılmıştır. Çıktı katmanında ise 20 sinir hücresine yer verilmiştir. Momentum oranı 0.9, öğrenme oranı ise 0.005 olarak belirlenmiştir. Araştırmacılar, ağırlıkların başlangıç değerlerinin belirlenmesi için Xavier parametresini seçmişlerdir. Performans değerlendirmek için doğruluk ölçütünü dikkate almışlardır. Farklı algoritmalar ile oluşturulan sınıflandırma modellerinin test verisine uygulanması sonucunda C5.0 ile 0.86, Boosted C5 ile 0.82, Regresyon Ağaçlarıyla 0.84, SVM ile 0.78, LR ile 0.82, RF ile 0.84 ve derin öğrenme ile 0.87 doğruluk değerlerini elde etmişlerdir. Bu doğrultuda en uygun performansın derin öğrenme algoritmasıyla elde edildiğini belirtmişlerdir. Derin öğrenme modelini oluşturan özniteliklerden önem derecesi en yüksek öznitelikler sırasıyla “ailesi ile olan ilişki seviyesi, arkadaşlarıyla dışarı çıkışları ve annesinin eğitim durumunun en etkili değişkenler” olduğunu belirtmişlerdir. Bu sonuçla aile içi ilişkilerdeki seviyenin öğrenci performansı üzerindeki etkisini vurgulamaktadırlar [46].

Punlumjeak ve Rachbure (2015) çalışmalarında, bir sınıflandırma modeli uygulanmadan önce, en önemli özellikleri bulmak için veri ön işleme sürecinde özellik seçim yöntemi kullanılmasını önermişlerdir. Çalışmada kullanılan veriler, Rajamangala Teknoloji Üniversitesi Thanyaburi, Pathumthani, Tayland Mühendislik Fakültesi’nden 2004-2010 yılları arasında toplanmıştır. 463,956 kayıt, öğrenci kimliğine göre 6,882 kayıt olarak gruplandırılmıştır. Veri seti sadece birinci sınıf öğrencisi ve öğrencinin kabul verileri vardır. Veri seti 15 öznitelikten oluşmaktadır. Bu çalışmada, veri temizleme ve veri dönüştürme işlemi yapıldıktan sonra dört öznitelik seçme yöntemi kullanılmıştır. Bunlar; Genetik algoritmalar, SVM, IG ve her yöntem için minimum özellik kümesini bulmak için mRMR’dir. Daha sonrasında NB, DT, KNN ve YSA ile sınıflandırmalar gerçekleştirilmiştir. Çalışmalarında, sınıflandırıcıların doğruluğunu değerlendirerek her bir öznitelik seçim yönteminin performansını her bir sınıflandırma tekniği ile karşılaştırmışlardır. Doğru bir tahmin modeli için 10 kat çapraz doğrulama (Cross Validation) kullanmışlardır. Deneysel sonuçlar, 10 öznitelik seçilerek mRMR öznitelik seçim yönteminin KNN sınıflandırıcı ile %91.12 doğrulukla en iyi sonucu verdiğini göstermektedir. Mevcut çalışmanın sonucu, minimum ve anlamlı özellik bulmak için gelecekteki seçim avantajının öğrenci performansını sınıflandırmada daha etkili olduğunu göstermektedir. Çalışmada kullanılan yazılım RapidMiner Studio sürüm 6.40’tır [47].

Yakupoğlu (2018) tez çalışmasında ilgili okulda öğretmenin verdiği kararlar yürütülen, öğrencilerin 5 ana ders üzerindeki yetkinlik sınıflarına atanması işlemi, sınıflandırma algoritmalarıyla otomatikleştirilmeye çalışılmıştır. T.C. Millî Eğitim Bakanlığı'na bağlı özel bir okulda, orta öğretim 5., 6. ve 7. sınıf öğrencilerine ait veri setleri üzerinde eğitimsel veri madenciliğinin bir uygulaması yapılmıştır. Çalışma, 2015-2018 yılları arasındaki üç akademik yıl boyunca ortaokul seviyesinde öğrenim görmüş öğrencileri kapsamaktadır. Öğrenciye ait demografik bilgiler, aile bilgileri, sağlık bilgileri, öğrencinin tutum ve davranışına yönelik bilgiler ve akademik başarısını ve gayretini gösteren bilgiler yer almaktadır. Veri setinde 79 adet öznitelik vardır. Tüm veri seti kullanılarak öğrencilerin Matematik, Türkçe, Sosyal Bilgiler ve İngilizce dersleri olan üzere 5 ana ders üzerindeki yetkinlik sınıfları 4 farklı makine öğrenmesi algoritmasıyla tahmin edilmeye çalışılmıştır. Çalışmada kullanılan makine öğrenmesi algoritmaları C4.5 Karar ağacı, MLP, KNN ve SVM algoritmasıdır. Yapılan ilk deneyde, veri seti üzerinde herhangi bir ön işleme çalışması yapılmadan algoritmaların gösterdiği performanslar incelenmiştir. Veri ön işleme öncesi yapılan bu analizde, C4.5 ve SVM algoritmaları, MLP ve KNN algoritmalarına göre daha yüksek performans göstermiştir. Bir sonraki deneyde veri setindeki gürültülü ve eksik veriler temizlenerek veri ön işleme yapılmıştır. 2 farklı öznitelik seçme yöntemi uygulanmıştır. İlk yöntemde C4.5 algoritmasına göre daha az açıklayıcı değişkenler modelden çıkartılarak değişken sayısı azaltılmıştır. İndirgenen veri seti ile kurulan modellerden SVM algoritması dışında diğer algoritmalarda kayda değer bir başarı sağlanamamıştır. İkinci yöntemde WEKA özellik seçimi ile değişkenler belirlenmiştir. Tahmin edilen sınıflar üzerinde 0.1 korelasyon değerinin altında kalan değişkenlerin modelden çıkartılması yoluyla değişkenler azaltılmıştır. Tüm sınıflandırma algoritmalarının performans değerleri, kayda değer oranlarda yükselmiştir [48].

Kumar ve arkadaşları (2022) yaptıkları çalışmada eğitim kurumlarındaki öğrencilerin başarısının öğrenmesiyle ilgili herhangi bir örüntü keşfedilir ve doğru bir şekilde analiz edilirse gelecekte öğrencilerin akademik performansını artırmasına yardımcı olacağını savunmaktadırlar. Veri setinin birçok özellik içereceğini, ancak tüm bu özelliklerin veri kümesindeki modeli tahmin etmek için önemli olmadığını belirtmişlerdir. Öznitelik seçim tekniğini uygulayarak önemli olmayan özellikleri kaldırmak gerektiğini belirtmişlerdir. Öznitelik seçim yöntemlerinden CFS, IG ve GR özneliği değerlendiricisini kullanarak, öğrenci verilerindeki en önemli özellikleri tespit etmeyi amaçlamışlardır. Bu çalışmada, öğrenci tahminini tahmin etmek için halka açık iki veri seti kullanılmıştır. Öğrencilerin notlarının yanı sıra kişisel, ders dışı etkinliklere katılım ve demografik özellikleri ile ilgili verileri içerir. Tüm veriler okul raporları ve anketler kullanılarak toplanmıştır. İlk veri seti (395 kayıt), bir matematik dersindeki öğrenci performansına ilişkin verileri içerirken, ikinci veri seti (649 kayıt) bir Portekizce dil dersinden öğrenci performansına ilişkin verileri içerir. Her ikisinde de toplam 33 öznitelik bulunmaktadır. Çalışmalarında, sınıflandırma algoritmaları ile farklı öznitelik seçim yöntemleri arasındaki etkiyle bir karşılaştırma yapmışlardır. Sınıflandırma algoritmaları olarak NB, RF, DT, MLP ve Karar tablosu teknikleri kullanılmıştır. Öznitelik seçim teknikleri kullanmadan yapılan modellerde en yüksek başarıyı %80 doğruluk oranıyla Karar tablosu teknikleri verirken, öznitelik seçim teknikleri uygulandıktan sonra tüm sınıflandırma algoritmalarında doğruluk oranında %2 ile %6 arasında artış gözlemlenmiştir. %83.33 doğruluk oranıyla en yüksek başarıyı öznitelik seçim tekniklerinden bilgi kazanım özneliği değerlendiricisi uygulandıktan sonra kurulan Karar tablosu tekniklerinin verdiği belirtilmiştir [49].

Ramaswami ve Bhaskaran (2009) ortaokul öğrencilerine ait kişisel bilgi, sosyal çevre ve akademik bilgiler içeren verileri kullanarak öğrencilerin akademik başarı durumlarının tahmin edilmesini hedeflemektedirler. Veri setinde 1,969 kayıt ve 32 bağımsız öznitelik vardır. Çalışmalarında farklı öznitelik seçim yöntemleri kullanarak 6 adet indirgenmiş veri seti oluşturmuşlardır. Her bir veri seti ile modeller oluşturularak sonuçlar karşılaştırılmıştır. Öznitelik seçim yöntemi olarak; CFS, Ki-kare, GR, IG, SU ve Rahatlama öznitelik değerlendirme yöntemlerini kullanmışlardır. Modelleme için 4 farklı sınıflandırma algoritması kullanılmıştır. Bunlar NB, VotedPerceptron, OneR ve PART algoritmalarıdır. Dört sınıflandırıcının hepsi de IG niteliği tarafından oluşturulan özellik alt kümesi için iyi performans göstermiştir. OneR yöntemi %89'dan fazla olan

yüksek tahmin doğruluğu göstermiştir. Özellikle OneR tüm özellik alt kümeleri için sürekli olarak aynı düzeyde tahmin performansı sağladığı görülmüştür. Modellerin performansını değerlendirmek için Doğruluk, F-skoru ve ROC eğrisi kullanılmışlardır. Tüm modeller WEKA programında kurulmuştur [50].

Uzel ve arkadaşları (2018) öğrencilerin akademik başarısını önceden tahmini etmeyi ve başarıya etkisi fazla olan özellikleri belirlemeyi hedeflemişlerdir. Çalışmalarında birliktelik kuralları ve sınıflandırma yöntemlerini aynı anda uygulamışlardır. Kullanılan veri seti ilköğretim ve ortaöğretim seviyesindeki öğrencilere aittir. Veri seti 480 öğrenciye ait 16 öznitelikten oluşmaktadır. Sınıflandırma için MLP, DT, RF, Oylama sınıflandırıcıları (OneR), NB ve YSA yöntemlerini kullanmışlardır. 10-kat çapraz katlama yönteminden yararlanmışlardır. Tüm uygulamalar WEKA yazılımında gerçekleştirilmiştir. Modellerin performanslarının değerlendirilmesi için karmaşıklık matrisi, doğruluk, duyarlılık, kesinlik ve F-skoru dikkate alınmıştır. En yüksek doğruluk oranı YSA ile kurulan modelden elde etmişlerdir. Bu modelden elde edilen doğruluk oranı %79.1 olduğu belirtilmiştir. Ayrıca hangi özelliklerin öğrencilerin akademik performanslarının artmasıyla daha fazla ilişkili olduğu gözlemlenmiştir. Veri setindeki özelliklerden öğrencinin okuldaki devamsızlığı, cinsiyeti, ek ders takviyesi alıp almadığı, ailenin okuldan memnuniyeti gibi özellikler akademik başarıyı en çok etkileyen etmenler olduğu belirtilmiştir [51].

Şengür (2013) tezinde Fırat Üniversitesinde öğrenim gören Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü öğrencilerinin tüm ders not bilgileri ile mezuniyet notlarını tahmin etmeyi amaçlamışlardır. Böylece mezun olamayacak öğrenciler önceden bilenecek ve öğrencilere haber verilecektir. Aynı zamanda ortalaması belirli bir notun altında kalan öğrencileri uyarabileceğini ve önlem alınabileceğini belirtmiştir. 2011 yılında mezun olmuş 127 bölüm öğrencisinin 4 yıl içerisinde almış olduğu toplam 49 farklı dersin yılsonu notlarını kullanmıştır. Mezuniyet notu çıktı değişkendir ve 0-4 arası değerlerden oluşmaktadır. YSA ve DT algoritmaları kullanılarak diploma notu tahmin edilmiştir. Mezuniyet notunun tahmini için iki senaryo denemiştir. Senaryoları MATLAB ortamında gerçekleştirmiştir. İlk senaryoda, öğrencilerin birinci ve ikinci sınıfa ait derslerinin yılsonu notları kullanılarak mezuniyet notunu tahmin etmiştir. İkinci senaryo da ise mezuniyet notlarının tahminini ilk üç sınıf notları ile gerçekleştirmiştir. Başarı değerlendirmesi için korelasyon katsayısı ve ortalama kare hata fonksiyonları (OKH), modelin performansını daha etkili değerlendirmek için ortalama mutlak hata (OMH) yöntemini kullanılmıştır. Gerçekleştirilen çalışmalarında YSA'nın, DT algoritmasına göre daha iyi tahmin sonucu elde ettiğini ve ikinci senaryonun, birinci senaryoya göre daha iyi tahmin yaptığını belirtmiştir. İkinci senaryo için ortalama 7.2228 OMH değeri ve 0.2026 OKH değeri elde edilmiştir [52].

Zaffar ve çalışma arkadaşları (2018) çalışmalarında filtre öznitelik seçim algoritmalarının ve sınıflandırma algoritmalarının iki farklı öğrenci veri kümesi üzerindeki performansının analizini yapmışlardır. Birinci veri seti 16 öznitelik içeren 500 öğrenci kaydından oluşmaktadır. Veri setinde demografik, akademik ve davranışsal olmak üzere üç özellik kategorisi vardır. İkinci veri seti ise 24 öznitelik içeren 300 öğrenci kaydından oluşmaktadır. Veriler Hindistan'daki 3 farklı kolejden elde edilmiştir. Tüm modeller WEKA programının 3.9 sürümü kullanılarak oluşturulmuştur. Çalışmada altı farklı öznitelik seçim algoritması kullanılmıştır. Bunlar CFS, Ki-kare, filtre, GR, Ana bileşenler ve ReliefF öznitelik seçim yöntemleridir. Öznitelik seçim yöntemleri ile veri seti indirgenerek yeni veri setleri oluşturulmuştur. Oluşturulan veri setleriyle sınıflandırma modelleri kurulmuştur. DT, J48, RF, Rastgele ağaç, Bayesian Network, NB, NaiveBayesUpdateable, MLP, Basit Lojistik, SMO, Karar tablosu, OneR, Jrip, REPTree olmak üzere on beş sınıflandırma algoritması kullanılmıştır. 2 farklı veri setinde kullanılan öznitelik seçim yöntemlerinin önemli performans farkı olduğunu belirtmişlerdir. Doğruluk yüzdelerinde %10 ile %20 fark olduğu gözlemlenmiştir. Filtre öznitelik seçim yöntemlerinin performansı, öznitelik sayısı arttıkça düşmektedir. Bu sebeple çok sayıda öznitelik sayısına sahip olan öğrencinin akademik performansını tahmin etmek için sarmalayıcı özellik seçme teknikleri de değerlendirilebilir olduğunu belirtmişlerdir. Birinci veri seti üzerinde uygulanan Ki-kare seçim teknikleri ile MLP algoritmasının %77 ile en yüksek doğruluk sağladığı görülmüştür. İkinci veri seti için Ki-

kare özellik seçim yöntemleri ile Bayes Net sınıflandırma algoritması birlikte kullanıldığında %62 elde edilerek en yüksek doğruluk oranı sağladığı görülmüştür [53].

Acharya ve Sinha (2014) çalışmalarında yükseköğretimde öğrenim gören öğrencilerin kişisel bilgileri, sosyal çevre bilgileri, sosyo-ekonomik bilgileri, daha önceki okuluna ait mezuniyet bilgileri gibi verileri kullanılarak 2. dönem notlarını tahmin etmeyi amaçlamışlardır. Veri setinde 413 öğrenciye ait 34 adet bağımsız nitelik bulunmaktadır. Bağımlı nitelik 6 seviyeli olup kategorik bir yapıya sahiptir. Veri seti üzerinde öznelik seçim yöntemleri kullanarak başarıya en çok etki eden 8 bağımsız nitelik belirlenmiştir. İndirgenen veri seti ile farklı makine öğrenmesi algoritmaları kullanarak tahmin modelleri oluşturmuşlardır. Bu algoritmalar NB, DT, MLP ve KNN algoritmalarıdır. En başarılı model DT algoritmaları ile elde edilmiştir. Ailenin ekonomik durumu, 1. dönem not ortalaması, ortaöğretim not ortalaması, okula devamsızlığı, hangi dine mensup olduğu, sosyal çevre ve ortaöğretim okul türü niteliklerinin başarıyı önemli derecede etkilediğini belirtmişlerdir [54].

Cortez ve Silva (2008) ortaokul öğrencilerinin akademik başarılarının tahminine yönelik regresyon ve sınıflandırma modelleri oluşturmuşlardır. Çalışmada 2 farklı veri seti kullanmışlardır. Matematik veri setinde 395 kayıt, Portekizce veri setinde ise 649 kayıt mevcuttur. Her 2 veri setinde de 32 bağımsız nitelik mevcuttur. Bu nitelikler öğrencilerin sosyo-ekonomik, demografik, bir önceki döneme ait dönem not bilgilerini içerir. Bağımlı nitelik diye isimlendirilen hedef nitelikleri ise Portekizce ve Matematik derslerinin yılsonu not ortalamalarıdır. Not ortalamaları 2 seviyeli ve 5 seviyeli olmak üzere 2 çeşittir. Makine öğrenmesi algoritmalarından DT, SVM, RF ve YSA algoritmaları kullanılarak modeller kurulmuştur. Modeller R programlama dili kullanılarak uygulanmıştır. Çapraz doğrulama yöntemi kullanılarak veriyi bölme işlemi gerçekleştirilmiştir. Model performans değerlendirmesi için karmaşıklık matrisi ve doğruluk ölçütü kullanmışlardır. Oluşturulan tahmin modellerin performans sonuçlarına göre Portekizce dersi için en yüksek performansı DT algoritması, Matematik dersi için en yüksek performansı NB algoritması ile elde etmişlerdir. Başarıyı en fazla etkileyen özellikler anne ve babanın eğitim durumu, alkol tüketimleri, meslekleri ve öğrencinin geçmiş dönem not bilgileri olduğu belirlenmiştir [55].

4. Tartışma (Discussion)

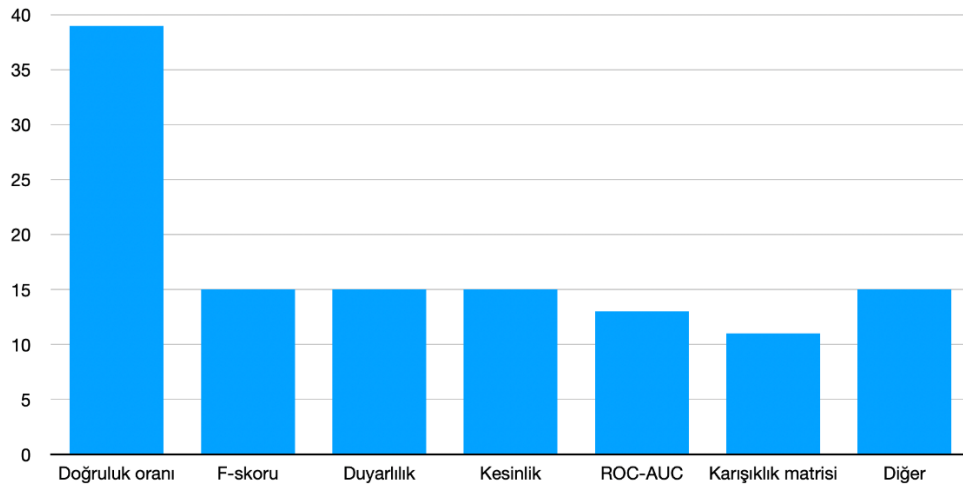
Bu bölümde, incelenen çalışmalarla ilgili özet bilgiler sunularak en başarılı modeller, en çok kullanılan yöntemler ve algoritmalar ile araştırmanın araçları ile ilgili bilgiler verilmiştir. İncelenen çalışmalarda öznelik seçim teknikleri kullanılan ve sınıflandırmada en başarılı 10 modellerin başarı oranları Tablo 2'de gösterilmiştir.

Tablo 2. En başarılı 10 modelin analizi (Analysis of the 10 most successful models)

Çalışma	Öznelik Seçim Tekniği	Model	Başarı Oranı (Doğruluk %)
[18]	IG	YSA	%98.6
[21]	IG	RF	%75.52
[38]	IG, CFS, GR	DT	%85 - %91
[43]	Select K-Best	Boosting	%75
[44]	FSA	LR	%90
[47]	mRMR	KNN	% 91.12
[48]	CFS	SVM	%85.5
[49]	IG	Karar Tablosu	%83.33
[50]	IG	OneR	%89
[53]	Chi-Square	MLP	%77

Bu çalışmada öğrencilerin akademik başarılarının önceden tahmin edilmesi veya sınıflandırılması, veri setindeki gereksiz veya sonucu aynı etkileyen özelliklerin tespiti ile ilgili ulusal ve uluslararası alanlarda yayınlanmış olan çalışmalar araştırılmıştır. İncelenen çalışmalarda bazılarında öznelik seçim teknikleri kullanılmış bazılarında ise kullanılmamıştır.

Literatürde kullanılan öznelik seçim teknikleri ve sınıflandırma algoritmalarının istatistikleri de analiz edilmiştir. Yapılan 39 çalışmada en fazla kullanılan algoritmanın NB algoritması olduğu, en fazla kullanılan öznelik seçim tekniklerinin de filtreleme yöntemlerinden IG öznelik seçim tekniği olduğu görülmektedir. IG öznelik seçim tekniği kullanılan Tablo 2'deki çalışmalar incelendiğinde başarı oranlarının %75.52 ile %98.6 arasında değiştiği görülmektedir. İncelenen çalışmalardan 12 tanesinde IG öznelik seçim tekniği, 8 çalışmada CFS, 5 çalışmada GR, 4 çalışmada Ki-kare (Chi-Square), 3 çalışmada Genetik algoritma öznelik seçim tekniği kullanılmıştır. Select k-best, rastgele n sınıflandırıcısı, mRmR, Simetrik belirsizlik, ANOVA, Karşılıklı Bilgi, Hibrit modeller, Geri yönlü emilasyon, Relief, Ana bileşenler öznelik seçim teknikleri de kullanılan tekniklerdir. Sınıflandırma algoritmaları sırasıyla 19 çalışmada NB, 17 çalışmada DT, 15 çalışmada RF, 13 çalışmada YSA, 13 çalışmada LR, 11 çalışmada KNN, 10 çalışmada SVM, 9 çalışmada MLP, 4 çalışmada Boosting algoritmaları kullanılmıştır. Bu algoritmaların haricinde J48, Bayes net, Cart, CS5, Chaid, Derin öğrenme, REPTree, OneR, Karar Tabloları ve Çoklu Regresyon algoritmaları da kullanılmış Kural Tabanlı Algoritma ise bir çalışmada kullanılmıştır. Çalışma ortamları incelendiğinde WEKA programının yoğun olarak kullanıldığı görülmüştür. Kullandıkları ortamları belirten çalışmalar içerisinde en çok kullanılan geliştirme ortamı sırası ile WEKA (14), Rapid Miner (5), SPSS (3), Python-Scikit Learn (5), R-Studio (4), MATLAB (2), Orange (2) şeklindedir. Modellerin performansını değerlendirmek için en çok kullanılan ölçüt doğruluk oranıdır. Literatür incelemesinde doğruluk oranından faydalanan çalışma sayısı 39'dur. Diğer değerlendirme ölçütlerinin kullanım sayıları; F-skoru (15), duyarlılık (15), kesinlik (15), ROC-AUC (13), karışıklık matrisi (11)'dir. Kappa skoru, OKH, OMH, RMSE, MSE gibi metrikler, çalışmalarda kullanılan diğer değerlendirme ölçütleridir. Kullanılan ölçütlere ait grafik Şekil 4'te verilmiştir.



Şekil 4. Çalışmalarda kullanılan değerlendirme ölçütleri
(Evaluation criteria used in the studies)

Son olarak literatür incelemesi sırasında öznelik seçimi algoritmalarını uygulayan çalışmalarda öznelik sonuçlarına bakılmıştır. Araştırmalarda kullanılan veri setlerinde, öznelik seçim işlemlerinden sonra belirlenerek kullanılan temel öznelikler incelenmiştir. Farklı çalışmalarda farklı veri setleri kullanılmasına rağmen, bu araştırmalardaki performansı etkileyen temel öznelikler şöyle özetlenebilir: anne ve babanın eğitim düzeyleri, öğrencilerin daha önce aldığı eğitim kalitesi, öğrencilerin okul dışı aldıkları kurs, ailenin gelir düzeyi, anne ve babanın eğitim durumu ve öğrencinin sağlık durumu. Bu öznelikler, farklı araştırmalarda ortak olarak kullanılan ve akademik performansı en çok etkileyen temel kriterler olarak bulgular arasında yer almaktadır.

5. Sonuç (Conclusion)

Eğitim, toplumların ilerlemesinde hayati önem taşımaktadır. Eğitim sistemlerinde yer alan her öğrenci kendine özgü ihtiyaçlara ve öğrenme tarzlarına sahiptir. Eğitim sistemleri ise bu çeşitliliği etkili bir şekilde

yönetebilme konusunda zorluklarla karşı karşıyadır. Son yıllarda makine öğrenimi teknolojilerinin eğitimde kullanımı, öğrenci performansının değerlendirilmesi ve akademik başarının artırılması için önemli bir araç olarak ortaya çıkmıştır.

Makine öğrenmesi teknikleri, öğrencilerin bireysel ihtiyaçlarını ve öğrenme stillerini daha iyi anlamak ve onlara özelleştirilmiş bir eğitim deneyimi sunmak için kullanılabilir. Böylece makine öğrenmesi teknikleriyle öğrencilerin potansiyelleri daha iyi ortaya çıkarılabilir. Ek olarak eğitim politikalarının geliştirilmesi ve eğitim programlarının etkinliğinin değerlendirilmesi için de kullanılarak eğitim sistemlerinin daha verimli ve etkili hale getirilmesine yardımcı olabilir. Makine öğrenmesi teknikleri özellikle öğrencilerin akademik zorluklarla karşılaşabileceği alanları belirlemek ve erken müdahale sağlamak için kullanılabilir. Öğrencilerin bireysel öğrenme hızlarına ve kapasitelerine göre zorlanabilecekleri alanlar erken müdahale kapsamında daha farklı öğretim yöntem ve teknikleriyle anlatılabilir. Konu tekrarları, ek sınavlar ve geribildirim mekanizmaları planlanarak öğrenmeler pekiştirilebilir. Böylece öğrencilerin başarısız olmaları engellenebilir ve akademik başarıları artırılabilir. Makine öğrenmesi teknolojilerinin eğitimde kullanımı bazı zorluklar ve endişeleri de beraberinde getirebilir. Öğrenci verilerinin toplanması ve kullanılması, gizlilik ve etik endişelere yol açabilir. Bu verilerin nasıl kullanılacağı, kimin erişebileceği ve nasıl korunacağı gibi konuları baş edilmesi gereken unsurlardan biridir. Teknolojilere eşit erişim sağlamak da önemlidir çünkü erişimdeki eşitsizlikler, teknolojilerin etkisini azaltabilir ve eşitsizlikleri derinleştirebilir.

Bu potansiyellerden yola çıkarak bu çalışmada öğrencilerin akademik başarılarını makine öğrenmesi algoritmaları ve özellik seçim teknikleriyle değerlendiren çalışmalar incelenmiştir. Bu çalışmalarda kullanılan teknikler ve algoritmalar bir araya getirilmiştir. Çalışmalar sonuçları bakımında da irdelenmiş, bulguları sunulmuştur. Ayrıca yıl bazında çalışma sayıları ve kullanılan araçlarla ilgili içgörüler oluşturmak amaçlı bilgiler de verilmiştir. Bu yönleriyle çalışmanın gelecekteki araştırmalara kapsamlı bir analiz noktası oluşturacağı düşünülmektedir. Bu çalışmanın literatüre katkıları gelecek çalışmaların başlangıç noktası olması açısından da önem az etmemektedir.

Gelecekteki çalışmalar makine öğrenmesinin eğitimde kullanımı açısından pek çok potansiyeli barındırmaktadır. Bu çalışmalarda makine öğrenmesi teknikleri kullanılarak, öğrencilerin bireysel ihtiyaçlarını daha iyi karşılayacak kişiselleştirilmiş öğrenme modelleri geliştirilebilir. Gelecekteki araştırmalar, bu modellerin etkinliğini ve uygulama yöntemlerini de inceleyebilir. Buna ek olarak farklı araştırmalarda daha sıkı gizlilik ve etik protokoller geliştirmeye odaklanılabilir. Verilerin nasıl korunacağı, kimin erişebileceği ve bu verilerin etik kullanımı konularında rehberlik edecek politikalar oluşturabilir. Ayrıca öğrencilerin akademik zorluklarla karşılaşabilecekleri alanları belirlemek ve erken müdahale sağlamak için gelecekteki çalışmalar, bu erken müdahale mekanizmalarının nasıl daha etkili hale getirilebileceğini ve hangi öğretim yöntemlerinin en iyi sonuçları verdiğini araştırabilir. Öğrenci performansını değerlendiren ve akademik başarıyı artıran makine öğrenimi algoritmalarının geliştirilmesine yönelik çalışmalar da devam etmelidir. Bu çalışmalar, mevcut algoritmaların iyileştirilmesi, yeni algoritmaların geliştirilmesi ve bu algoritmaların eğitim ortamında nasıl uygulanabileceği konularında derinlemesine analizler yapabilir. Uzun vadeli etki çalışmaları da bir diğer önemli araştırma alanı olabilir. Bu araştırmalar, bu tekniklerin öğrencilerin kariyerleri üzerindeki etkilerini ve uzun vadeli eğitim çıktılarındaki değişiklikleri değerlendirebilir.

Çıkar Çatışması Beyanı (Conflict of Interest Statement)

Yazarlar tarafından herhangi bir çıkar çatışması bildirilmemiştir.

Kaynaklar (References)

- [1] B. Aslan and M. C. Babadoğan, "Relationships between learning style preferences and gender, age and success level at 7th and 8th Grade," *Eurasian Journal of Educational Research*, vol. 21, pp. 35-48, 2005.
- [2] R. R. F. Mendes, F. B. de Voznika, A. A. Freitas, and J. C. Nievola, "Discovering Fuzzy Classification Rules with Genetic Programming

and Co-evolution,” in *Principles of Data Mining and Knowledge Discovery*, L. De Raedt ve A. Siebes, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 314-325, 2001.

[3] M. Ünver, E. Erdal, and A. Ergüzen, “Big Data Example In Web Based Learning Management Systems,” *International Journal of Advanced Computational Engineering and Networking*, vol. 6, no. 2, pp. 39-42, 2018.

[4] S. Buyrukoğlu, “A novel color labeled student modeling approach using e-learning activities for data mining,” *Univers Access Inf Soc*, vol. 22, no. 2, pp. 569-579, June 2023. doi:10.1007/S10209-022-00894-8/TABLES/7

[5] A. Peña-Ayala, “Educational data mining: A survey and a data mining-based analysis of recent works,” *Expert Syst Appl*, vol. 41, no. 4, pp. 1432-1462, March 2014. doi:10.1016/J.ESWA.2013.08.042

[6] B. G. Emiroglu and S. Sahin, “Analysis of Students’ Performances during Lab Sessions of Computer Networks Course,” *J Educ Techno Soc*, vol. 16, no. 3, pp. 329-346, 2013.

[7] R. Bütüner and M. H. Calp, “Estimation of the Academic Performance of Students in Distance Education Using Data Mining Methods,” *International Journal of Assessment Tools in Education*, vol. 9, no. 2, pp. 410-429, June 2022. doi:10.21449/IJATE.904456

[8] S. Kayalı and S. Buyrukoğlu, “Makine Öğrenmesi Yöntemleriyle Öğrencilerin Akademik Performanslarının Sınıflandırılması,” in *2nd International Conference On Educational Technology And Online Learning-ICETOL*, pp. 330, 2022

[9] S. Buyrukoglu, F. Batmaz, and R. Lock, “Semi-Automatic Assessment Approach to Programming Code for Novice Students,” in *Proceedings of the 8th International Conference on Computer Supported Education (CSEDU 2016)*, pp. 289-297, 2016

[10] İ. Yücedağ and O. Güler, “Fuzzy Logic Based Approach to Site Selection Problem of Vocational Secondary School Students,” *Hacettepe University Journal of Education*, vol. 32, no. 1, pp. 111-122, 2017.

[11] S. Agarwal, G. N. Pandey, and M. D. Tiwari, “Data Mining in Education: Data Classification and Decision Tree Approach,” *International Journal of e-Education, e-Business, e-Management and e-Learning*, vol. 2, no. 2, pp. 140-144, 2012. doi:10.7763/IJEEEE.2012.V2.97

[12] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. in *Springer Series in Statistics*. New York, NY: Springer New York, 2009. doi:10.1007/978-0-387-84858-7

[13] E. Alyahyan and D. Düştögör, “Predicting academic success in higher education: literature review and best practices,” *International Journal of Educational Technology in Higher Education*, vol. 17, no. 1, pp. 1-21, December 2020. doi:10.1186/S41239-020-0177-7/TABLES/15

[14] A. Hellas et al., “Predicting academic performance: A systematic literature review”, in *Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE, Association for Computing Machinery*, pp. 175-199, July 2018. doi:10.1145/3293881.3295783

[15] B. Albreiki, N. Zaki, and H. Alashwal, “A Systematic Literature Review of Student’ Performance Prediction Using Machine Learning Techniques”, *Educ Sci (Basel)*, vol. 11, no. 9, 2021. doi:10.3390/educsci11090552

[16] S. Savaş and S. Karataş, “Cyber governance studies in ensuring cybersecurity: an overview of cybersecurity governance,” *International Cybersecurity Law Review*, vol. 3, pp. 7-34, 2022. doi:10.1365/s43439-021-00045-4

[17] M. Altun, K. Kayıkçı ve S. Irmak, “Sınıf Öğretmenliği Öğrencilerinin Mezuniyet Notlarının Regresyon Analizi ve Yapay Sinir Ağları Yöntemleriyle Tahmini,” *e-Uluslararası Eğitim Araştırmaları Dergisi*, vol. 10, no. 3, pp. 29-43, 2019. doi:10.19160/ijer.624839

[18] M. Yıldız and C. Börekci, “Predicting Academic Achievement with Machine Learning Algorithms,” *Journal of Educational Technology and Online Learning*, vol. 3, no. 3, pp. 372-392, 2020. doi:10.31681/jetol.773206

[19] F. Ahmad, N. H. Ismail, and A. A. Aziz, “The prediction of students’ academic performance using classification data mining techniques,” *Applied Mathematical Sciences*, vol. 9, no. 129, pp. 6415-6426, 2015. doi:10.12988/AMS.2015.53289

[20] S. Hussain, N. A. Dahan, F. M. Ba-Alwi, and N. Ribata, “Educational Data Mining and Analysis of Students’ Academic Performance Using WEKA,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 9, no. 2, pp. 447-459, February 2018. doi:10.11591/IJEECS.V9.I2.PP447-459

[21] A. A. Saa, M. Al-Emran, and K. Shaalan, “Mining Student Information System Records to Predict Students’ Academic Performance,” in *The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019)*, A. E. Hassanien, A. T. Azar, T. Gaber, R. Bhatnagar, and M. F. Tolba, Ed., Cham: Springer International Publishing, pp. 229-239, 2020.

[22] E. Fernandes, M. Holanda, M. Victorino, V. Borges, R. Carvalho, and G. Van Erven, “Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil,” *J Bus Res*, vol. 94, pp. 335-343, January 2019. doi:10.1016/J.JBUSRES.2018.02.012

[23] S. Aydın and A. E. Özkul, “Data Mining and An Application in Anadolu University Open Education System,” *Journal of Research in Education and Teaching*, vol. 4, no. 3, pp. 36-44, 2015.

[24] P. Sökkhey and T. Okazaki, “Developing Web-based Support Systems for Predicting Poor-performing Students using Educational

- Data Mining Techniques,” *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 7, pp. 23-32, 2020. doi:10.14569/IJACSA.2020.0110704
- [25] H. A. Mengash, “Using data mining techniques to predict student performance to support decision making in university admission systems,” *IEEE Access*, vol. 8, pp. 55462-55470, 2020. doi:10.1109/ACCESS.2020.2981905
- [26] F. Çifçi, C. Kaleli, and S. Günel, “Predicting Instructor Performance by Feature Selection and Machine Learning Methods,” *Anadolu Journal of Educational Sciences International*, vol. 8, no. 2, pp. 419-440, August 2018. doi:10.18039/AJESI.454587
- [27] V. S. R. Kumari, S. Veesa, and S. R. Chevala, “An Efficient Comparison Neural Network Methods to Evaluate Student Performance,” *GRD Journal for Engineering*, vol. 6, no. 1, pp. 4-7, 2020.
- [28] P. Kaur, M. Singh, and G. S. Josan, “Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector,” *Procedia Comput Sci*, vol. 57, pp. 500-508, January 2015. doi:10.1016/j.PROCS.2015.07.372
- [29] H. S. Yıldız Aybek and M. R. Okur, “Predicting Achievement with Artificial Neural Networks: The Case of Anadolu University Open Education System,” *International Journal of Assessment Tools in Education*, vol. 5, no. 3, pp. 474-490, September 2018. doi:10.21449/IJATE.435507
- [30] Ö. Özbay and H. Ersoy, “Analysis of Student Dynamism into Learning Management System through Data Mining Methods,” *GUJGEF*, vol. 37, no. 2, pp. 523-558, August 2017.
- [31] A. Baltacı, “The Data Mining: Measurement of Academic Achievement in Faculty of Divinity Students by Data Mining,” *Religion and Science-the Journal of the Faculty of Islamic Sciences of The University of Mus Alparslan University*, vol. 1, no. 1, pp. 1-23, June 2018.
- [32] S. Hazel Başer, O. Hökelekli, and K. Adem, “Predicting the Performance of Students Studying in Secondary Education Using Data Mining Methods,” *Journal of Computer Science and Technologies*, vol. 1, no. 1, pp. 22-27, June 2020.
- [33] Ş. Can, T. Özdil, and C. Yılmaz, “Estimation of The Factors Affecting the Success of The University Students By Logistic Regression Analysis,” *International Review of Economics and Management*, vol. 6, no. 1, pp. 28-49, April 2018. doi:10.18825/IREMJOURNAL.349984
- [34] V. B. Gladshiya and K. Sharmila, “An Efficient Approach of Feature Selection and Metrics for Analyzing the Risk of the Students Using Machine Learning,” *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation, ICAECA 2021*, 2021. doi:10.1109/ICAECA52838.2021.9675507
- [35] B. Abbasoğlu, “Prediction of Academic Achievements of Secondary School Students with Educational Data Mining Methods,” *Veri Bilimi Dergisi*, vol. 3, no. 1, pp. 1-10, 2020.
- [36] S. Aghalarova and S. Bozkurt Keser, “Application of AutoML Technique for Predicting Academic Performance of Students,” *El-Cezeri*, vol. 9, no. 2, pp. 394-412, May 2022. doi:10.31202/ECJSE.946505
- [37] E. Aydemir, “Forecasting of The Course Learning Notes by Data Mining Methods,” *European Journal of Science and Technology*, vol. 15, pp. 70-76, March 2019. doi:10.31590/EJOSAT.518899
- [38] M. K. Assistant, N. Nidhi, S. Majithia, and N. Sharma, “Predictive Model for Students’ Academic Performance Using Classification and Feature Selection Techniques,” *Proceedings - 2021 2nd International Conference on Computational Methods in Science and Technology, ICCMST 2021*, pp. 106-111, 2021. doi:10.1109/ICCMST54943.2021.00032
- [39] S. Bhutto, I. F. Siddiqui, Q. A. Arain, and M. Anwar, “Predicting Students’ Academic Performance Through Supervised Machine Learning,” *ICISCT 2020 - 2nd International Conference on Information Science and Communication Technology*, February 2020. doi:10.1109/ICISCT49550.2020.9080033
- [40] K. Karthikeyan and P. Kavipriya, “On Improving Student Performance Prediction in Education Systems using Enhanced Data Mining Techniques,” *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 7, no. 5, pp. 935-941, May 2017. doi:10.23956/IJARCSE/SV7I5/0348
- [41] S. S. Shreem, H. Turabieh, S. Al Azwari, and F. Baothman, “Enhanced binary genetic algorithm as a feature selection to predict student performance,” *Soft comput*, vol. 26, no. 4, pp. 1811-1823, February 2022. doi:10.1007/S00500-021-06424-7/FIGURES/16
- [42] A. Mueen, B. Zafar, and U. Manzoor, “Modeling and Predicting Students’ Academic Performance Using Data Mining Techniques,” *International Journal of Modern Education and Computer Science*, vol. 8, no. 11, pp. 36, November 2016. doi:10.5815/IJMECS.2016.11.05
- [43] J. Gajwani and P. Chakraborty, “Students’ Performance Prediction Using Feature Selection and Supervised Machine Learning Algorithms,” in *International Conference on Innovative Computing and Communications*, A. and B. S. and H. A. E. and A. S. and J. A. Gupta Deepak and Khanna, Ed., Singapore: Springer Singapore, pp. 347-354, 2021.
- [44] R. Bertolini, S. J. Finch, and R. H. Nehm, “Enhancing data pipelines for forecasting student performance: integrating feature selection with cross-validation,” *International Journal of Educational Technology in Higher Education*, vol. 18, no. 1, pp. 1-23, December 2021. doi:10.1186/S41239-021-00279-6/FIGURES/8
- [45] S. Aghalarova and S. Bozkurt Keser, “Prediction of Secondary School Students’ Academic Performance with Proposed Artificial Neural Networks Algorithm,” *Veri Bilimi*, vol. 4, no. 2, pp. 19-32, August 2021.

- [46] Y. Özkan, F. Önay Koçoğlu, and Ç. Selçukcan Erol, "Prediction of Student Performance By Deep Learning Algorithm," in *7th International Conference on "Innovations in Learning for the Future": Digital Transformation in Education*, S. Gülseçen, Ç. Selçukcan Erol, Z. Ayvaz Reis, and Gezer Murat, Ed., İstanbul: İstanbul University Press, pp. 136-144, 2018.
- [47] W. Punlumjeak and N. Rachburee, "A comparative study of feature selection techniques for classify student performance," in *7th International Conference on Information Technology and Electrical Engineering: Envisioning the Trend of Computer, Information and Engineering, ICITEE 2015, 29-30 Sept. 2015, Mai, Thailand*, 2015. pp. 425-429. doi:10.1109/ICITEED.2015.7408984
- [48] Y. Yakupoğlu, "Eğitimsel veri madenciliği ve bir uygulaması," Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, İstanbul, 2018.
- [49] M. Kumar, Nidhi, B. Sharma, and D. Handa, "Building Predictive Model by Using Data Mining and Feature Selection Techniques on Academic Dataset," *International Journal of Modern Education and Computer Science*, vol. 14, no. 4, pp. 29, August 2022. doi:10.5815/IJMECS.2022.04.02
- [50] M. Ramaswami and R. Bhaskaran, "A Study on Feature Selection Techniques in Educational Data Mining," *J Comput*, vol. 1, no. 1, pp. 711, December 2009.
- [51] V. N. Uzel, S. S. Turgut, and S. A. Özel, "Prediction of Students' Academic Success Using Data Mining Methods," in *Innovations in Intelligent Systems and Applications Conference, ASYU 2018, Adana, Türkiye, 4-6 November 2018*, Buse Melis Ozyildirim, Tülay Yildirim, Eds. IEEE, 2018. doi:10.1109/ASYU.2018.8554006
- [52] D. Şengür, "Öğrencilerin akademik başarılarının veri madenciliği metotları ile tahmini," Fırat Üniversitesi, Elazığ, 2013.
- [53] M. Zaffar, M. A. Hashmani, K. S. Savita, and S. S. H. Rizvi, "A Study of Feature Selection Algorithms for Predicting Students Academic Performance," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 5, pp. 541-549, 2018. doi:10.14569/IJACSA.2018.090569
- [54] A. Acharya and D. Sinha, "Early Prediction of Students Performance using Machine Learning Techniques," *Int J Comput Appl*, vol. 107, no. 1, pp. 37-43, December 2014. doi:10.5120/18717-9939
- [55] P. Cortez and A. Silva, "Using Data Mining to Predict Secondary School Student Performance," in *Proceedings of 5th Annual Future Business Technology Conference, Porto, Portugal, 9-11 Ap. 2008*, pp. 5-12.

This is an open access article under the CC-BY license

