

Karar Ağacı Algoritmalarıyla Öğrencilerin Akademik Performans Verilerinin Sınıflandırılması

Büşra DUYGU, Doç. Dr. Nursal ARICI

Bilgisayar Mühendisliği, Fen Bilimleri Enstitüsü, Gazi Üniversitesi, Ankara, Türkiye

busra.duygu13@gmail.com, nursal@gazi.edu.tr

Özet

Eğitim sektörünün giderek büyüyen yapısında, yükseköğretim kurumları, öğrenci performansının akademik olarak iyileştirilmesi ve öğrencilerin okulu bırakmalarını önlemek için veri madenciliği araçlarını ve tekniklerini kullanmaktadır. Veriler sosyo-ekonomik, demografik ve çevresel yirmi dört özellik taşıyan üç yüz öğrencinin akademik bilgisinden oluşmaktadır. J48, Decision Stump, RepTree, Random Forest ve Random Tree karar ağacı sınıflandırma yöntemleri kullanılmıştır. Veri madenciliği aracı olarak 3.8.3 sürümlü WEKA seçilmiştir. Rasgele orman algoritmasının doğruluk ve sınıflandırıcı hatalarına dayanarak diğer sınıflandırıcılardan daha iyi performans gösterdiği gözlemlenmiştir.

Anahtar Kelimeler: karar ağaçları, veri madenciliği, öğrenci performans analizi, eğitimde veri madenciliği

Abstract

In the ever-growing structure of the education sector, higher education institutions use data mining tools and techniques to improve student performance academically and prevent students from dropping out of school. The data consist of academic knowledge of three hundred students with twenty-four characteristics of socio-economic, demographic and environmental characteristics. J48, Decision Stump, RepTree, Random Forest and Random Tree decision tree classification methods were used. WEKA with 3.8.3 version was chosen as data mining tool. It was observed that the random forest algorithm performed better than other classifiers based on accuracy and classifier errors.

Key Words: decision trees, data mining, students performance analysis, data mining in education

1. Giriş

Eğitimsel veri madenciliği, veri madenciliğinde ortaya çıkan bir alandır. Bu sürekli gelişen eğitim ortamında öğrenci performansını araştırmak ve analiz etmek için veri madenciliği araçlarını kullanmaktadır. Sonuçlarının tahmin edilmesiyle akademik açıdan hem iyi hem de kötü performans gösteren öğrencilere odaklanır. Bu sayede fakülteler ve eğitimciler için geribildirim oluşturulmakta ve verilerin görselleştirilmesiyle de öğrenme sürecinin daha iyi bir değerlendirilmesi sağlanmaktadır. Eğitim veri madenciliği eğitim kalitesinin iyileştirilmesi için kullanılmaya başlanılan bir araçtır. Modern eğitim kurumları stratejileri ve gelecek planları için veri madenciliğine ihtiyaç duymaktadır. Öğrencinin performansı kişisel, sosyal, ekonomik ve çevresel faktörler gibi çeşitli etmenlere bağlıdır [1, 2]. Üst düzey

eğitim kurumlarının yetkilileri, öğrencilerin performansındaki yeni pedagojik stratejilerin tasarlanmasına yol açabilecek eğilimleri ve davranışları anlamak için deneysel sonuçların sonucunu kullanabilirler [3].

Bu çalışmada ele alınan veri setinin boyutu ve miktarı göz önünde bulundurulduğunda verilerin incelenmesinde veri madenciliği yöntemlerinden faydalanılmaktadır. Öğrencilerin akademik performansını araştırmak için sınıflandırma algoritmalarından karar ağacı algoritmaları kullanılmıştır. Görsellik yönünden zengin olması, maliyetinin düşük olması ve kolay anlaşılması sebepleriyle karar ağaçları(KA) algoritmaları tercih edilmiştir.

Çalışma kapsamında J48, DecisionStump, RepTree, Random Tree ve Random Forest sınıflandırma algoritmaları ile çalışılmıştır. Sınıflandırmayı en iyi gerçekleştiren algoritmayı seçmek için her algoritmanın doğruluğuna göre karşılaştırma yapılmıştır.

Çalışmada çeşitli özellikler dikkate alınmıştır ve etkisi yüksek olan özellikleri bulmak için ilk önce özellik seçimi yapılmıştır. Faydalı ve anlamlı bilgiler elde etmek için gereksiz özellikleri veri kümesinden kaldırılarak temizleme işlemi yapılmıştır. Bu sayede veri madenciliği süreci daha hızlı, değerli ve anlamlı hale getirilmiştir. Araştırmada öğrencilerin dönem sonu yüzdesi bağımlı parametre olarak seçilmiştir. Yüzdeler "En İyi", "Çok İyi", "İyi", "Geçiş", "Başarısız" olarak sınıflandırılmıştır. Araştırmada WEKA (Bilgi Analizi için Waikato Çevre) veri madenciliği aracı kullanılmıştır. WEKA, Java'da yazılmış ve veri madencileri tarafından yaygın olarak kullanılan açık kaynaklı bir araçtır [4]. WEKA, makine öğrenme algoritmalarının çoğunun veri setine uygulanmasını ve sonuçların görselleştirilmesinde kullanılır.

Çalışma beş bölümden oluşmaktadır. İkinci bölümde literatür taraması yer almaktadır. Üçüncü bölümde çalışmada kullanılan veri, özellikleri ve analiz teknikleri hakkında temel bilgiler sunulmaktadır. Dördüncü bölümde veri seti analiz sonuçları, beşinci bölümde ise sonuç ve tartışma yer almaktadır.

2.Literatür Taraması

Sumitha ve diğ. ark. [5], kıdemli öğrenciler veri setini kullanarak öğrencinin gelecekteki öğrenme çıktılarını tahmin etmek için bir veri modeli geliştirmiş ve veri madenciliği sınıflandırma algoritmalarından karar ağaçları algoritmalarını kıyasladıklarında J48 algoritmasının verilerini temel alarak bu iş için en uygun olduğunu gözlemlemişlerdir.

Baradwaj ve Pal [6], "Önceki Dönem Notları", "Sınıf Test Notları" dahil olmak üzere çoklu performans göstergelerine sahip, 4 yıllık bir süre boyunca (2007-2010) belirli bir kurs programına kayıtlı olan 50 kişilik bir grup üzerinde bir araştırma yapmış , "Seminer Performansı", "Ödevler", "Genel Yeterlilik", "Devam", "Laboratuar Çalışması" ve "Dönem Sonu Notları" gibi öznitelikleri ele almıştır. Karar ağacı oluşturmak için ID3 karar ağacı algoritmasını seçmişlerdir. Bu çalışma başarısızlık oranını azaltmak ve bir sonraki dönem sınavı için uygun önlem almak için özel dikkat gerektiren öğrencileri belirlemeyi amaçlamıştır.

Abeer ve Elaraby [7], temel olarak sınıflandırma kurallarını oluşturmaya ve öğrencilerin seçtikleri bir ders programında öğrencilerin performansını tahmin etmeye odaklanan benzer bir araştırma yürütmüştür. Daha önce kaydedilmiş öğrencilerin verilerini, üniversite veri tabanından toplanan birden fazla özellik ile 6 yıl boyunca (2005-10) belirli bir kurs programında toplamışlar ve analiz etmişlerdir. Sonuç olarak, bu çalışma, seçilen ders programındaki öğrencilerin final notlarını bir dereceye kadar tahmin etmenin yanı sıra, özel dikkat gerektiren öğrencileri tanımlamaya da yardımcı olur.

Ahmad ve arkadaşları [8], bilgisayar bilimleri dersinin birinci sınıf lisans öğrencilerinin akademik performanslarını tahmin etmek için bir çerçeve tasarlamıştır. Veri kümesi, 8 yıllık verileri içermektedir. Toplanan veriler, önceki akademik kayıtlar, aile geçmişi ve demografik bilgiler dahil olmak üzere öğrencilerin kayıtlarının çeşitli yönlerini içermektedir. Öğrencilerin akademik performansını bulmak için Karar Ağacı, Naif Bayes ve Kural Tabanlı sınıflandırıcılar uygulanmış ve Kural Tabanlı sınıflandırıcının diğer sınıflandırıcılar arasında en iyisi olduğunu ve doğruluğunun% 71.3 olduğunu göstermiştir.

Khasanah ve diğ. ark. [9], öğrenci performansını tahmin etmek için yüksek etkili özelliklerin dikkatlice seçilebileceğini bulmak için bir çalışma yürütmüştür. Öğrenci verileri, Endüstri Mühendisliği Universitas Islam Indonesia bölümünden alınmıştır. Öğrenci performansının sınıflandırılması ve tahmini için Bayesian Ağ ve Karar Ağacı algoritmaları kullanılmıştır. Özellik Seçimi yöntemleri, öğrencinin ilk yarıyıldaki katılımının ve Not Ortalaması'nın özellikler listesine girdiğini göstermiştir. Doğruluk oranı göz önüne alındığında, Bayesian Ağ, kendi durumlarında Karar Ağacı sınıflandırmasını geride bırakmıştır.

Ankita A Nichat ve diğ. ark. [10] karar ağacı ve yapay sinir ağı tekniklerini kullanarak sınıflandırma modelleri oluşturdular. Öğrencilerin performansını artırmak için öğrencilerin güçlü ve zayıf yönlerine erişmek için çeşitli özellikler kullanmışlardır.

Hilal Almarabeh [11], üniversite öğrencilerinin performansını değerlendirmek için WEKA aracını kullanarak sınıflandırma algoritmalarının doğruluğunun, verilerin boyutuna ve doğasına bağlı olduğunu buldu. Yazar, Naive Bayes, Bayesian Network, Sinir Ağı, ID3 ve J48 sınıflandırma tekniklerini kullanmış, Bayesian Ağının doğruluk açısından diğerlerinden daha iyi performans gösterdiği sonucunu elde etmiştir.

3.Kullanılan Veri ve Yöntemler

3.1.Karar Ağaçları

Son yıllarda en yaygın kullanılan sınıflandırma ve tanımlama yöntemlerinden biri karar ağaçlarıdır. Çok fazla tercih edilmesindeki en önemli sebep ise ağaç yapılarının oluşturulmasında izlenen kuralların anlaşılabilir ve sade olmasıdır[12]. Güvenilirliklerinin yüksek olması, veri tabanı sistemlerine entegrasyonunun kolay olması da tercih edilmesinde etkilidir [13].

Karar ağaçları kök düğüm, dallar ve yapraklardan oluşmaktadır. Karar ağaçlarındaki yapraklar sınıflandırmanın olduğu yerler olup dallar ise sonucu ifade etmektedir. Kök düğümden yaprak düğümlere doğru tüme varım yöntemiyle ağaç oluşturulur[14]. Karar ağacı iki temel işlem adımı içerir. Bu işlemler bölme ve budama işlemleridir. KA oluştururken en önemli adım hangi öznitelik değerlerini baz alıp dallanmanın oluşturulacağına karar vermektir. Literatürde bu sorunu çözmek için bir çok yaklaşım geliştirilmiştir. Bilgi kazancı[15], Gini indeksi[16], Towing kuralı[16] ve Ki-Kare olasılık tablo istatistiği[17] yaklaşımları bunlardan en önemlileridir. Entropi kurallarını içeren bilgi kazancı ve bilgi kazancı oranı yaklaşımında eldeki tüm öznitelikler özyineli olarak test edilir ve bilgi kazancı en yüksek olan öznitelik dallanma için seçilir[18].

Formüller ile açıklayacak olursak :

Bir veri setinin A_1, A_2, \dots, A_n şeklinde birkaç tane sınıftan oluştuğu ve T 'nin sınıf değerini gösterdiği düşünülürse, bir sınıfa ait olasılık $P_i=(A_i/T)$ olur ve sınıflara ait Entropi ;

$$Entropi(T) = - \sum_{i=0}^n P_i \log_2(P_i) \quad (1)$$

Şeklinde hesaplanır. Veri setindeki B özniteliliğine göre T sınıf değerleri T_1, T_2, \dots, T_n şeklinde alt kümelere ayrıldığı göz önünde bulundursun. B öznitelik değeri kullanılarak T sınıf değerlerinin bölünmesi sonucunda elde edilecek kazanç,

$$Kazanç(B, T) = Entropi(T) - \sum_{i=0}^n \frac{|T_i|}{|T|} Entropi(T_i) \quad (2)$$

Şeklinde hesaplanır. T kümesi için B özniteliliğinin değerini belirlemede bölümlenme bilgisi kullanılır(eşitlik 3).

$$Bilgi(B) = - \sum_{i=0}^k \frac{|T_i|}{|T|} \log_2\left(\frac{|T_i|}{|T|}\right) \quad (3)$$

Kazanç oranı,

$$Kazanç\ oranı = \frac{Kazanç(B, T)}{Bilgi(B)} \quad (4)$$

şeklinde hesaplanır[19].

Diğer önemli adım ise budamadır. Budama, bir ağacın istenmeyen alt düğümlerini veya alt ağaçlarını ayıklama işlemidir[20].

Karar ağacı algoritmasında ortaya çıkan sorulardan biri, işlemler sonucunda elde edilen final ağacının en uygun büyüklüğüdür. Çok büyük bir ağaç, eğitim verilerinin üzerinde çok fazla risk oluşturur ve yeni örneklerle genellemenin zayıf olmasına neden olur. Tam tersi küçük bir ağaç örnek alanı hakkında önemli yapısal bilgileri yakalayamayabilir. Bununla birlikte, bir ağaç algoritmasının ne zaman durması gerektiğini söylemek zordur çünkü tek bir ekstra düğüm eklenmesinin hatayı önemli ölçüde azaltıp azalmayacağını söylemek mümkün değildir. Bu sorun ufuk etkisi(horizon effect) olarak bilinir. Ortak bir çözüm ise, her düğüm az sayıda örnek içerene kadar ağacı büyütme ve ardından ek bilgi sağlamayan düğümleri kaldırmak için budama kullanmaktır [21]. Budama işlemi sonucunda daha az karmaşık ve anlaşılabilir bir ağaç elde edilmiş olur. Böylece sadeleşen ağaç yapısının işlem karmaşıklığı da azalmış olur. Budama yapılırken iki yöntem kullanılır: ilki ağacın oluşturulma aşamasında küçültülmesine karar verilen ön budama işlemi, diğeri ise ağaç yapısı oluşturulduktan sonrasında gerçekleştirilen son budama işlemidir[22].

3.2.Kullanılan Veri

Çalışmada UCI Machine Learning Repository adlı veri seti paylaşım sitesinden “Öğrencilerin Akademik Verileri” kullanılmıştır. Veri kümesi 24 özniteliği olan 300 örnek içermektedir. Önerilen çerçeveye aşağıdaki Şekil 1'de gösterilmektedir.

Bu araştırmanın verileri, üç farklı kolejden toplanmıştır, bunlar Duliajan Koleji, Doomdooma Koleji ve Digboi Assam, Hindistan Koleji'dir. İlk olarak, yirmi dört özniteliğin verileri toplanılmıştır. Öğrencinin öznitelik olarak adı bir anlam ifade etmediğinden, öznitelik listesinden çıkarılmıştır. "Pratik kağıttaki işaretler" niteliği, eksik değerler içerdiğinden dolayı ön işleme aşamasında kaldırılmıştır. Son olarak, veri temizliğinden sonra yirmi iki özellik seçildi. Tablo-1 seçilen özellikleri ,muhtemel değerleri ile göstermektedir.

Özellik(Attributes)	Değerler(Value)
Cinsiyet(GE)	Kadın,Erkek(Female, Male)
Öğrencinin kasti(CST)	Genel(G),Program Kast(SC),Program Kabile(ST),Diğer Geri Sınıflar(OBC) ve Azınlıklar(MOBC)
X sınıfı Değerlendirme Yüzdesi (TNP)	En iyi(Best),Oldukça iyi(VG), İyi(Good),Geçer(Paas),Başarısız(Fail)
XII Sınıfı Değerlendirme yüzdesi(TWP)	En iyi(Best),Oldukça iyi(VG), İyi(Good),Geçer(Paas),Başarısız(Fail)
İç Değerlendirme Yüzdesi(IP)	En iyi(Best),Oldukça iyi(VG), İyi(Good),Geçer(Paas),Başarısız(Fail)
Dönem Sonu Yüzdesi(ESP)	En iyi(Best),Oldukça iyi(VG), İyi(Good),Geçer(Paas),Başarısız(Fail)
Önceki Döneme Ait Borç(ARR)	Evet(Yes),Hayır(No)
Medeni Hali(MS)	Evli(Married),Bekar(Unmarried)
Yaşadığı yer(LS)	Köy(Village), Kasaba(Town)
Kabul Kategorisi(AS)	Ücretsiz(Free),Ücretli(Paid)
Ailenin Aylık Geliri(FMI)	Oldukça yüksek(Very High),Yüksek(High),Orta üstü(Above Medium),Orta(Medium),Düşük(Low)
Aile Boyutu (FS)	Geniş(Large),Orta(Average), Küçük(Small)
Arkadaş Sayısı(NF)	Geniş(Large),Orta(Average), Küçük(Small)
Çalışma Saatleri(SH)	İyi(Good), Orta(Average), Kötü(Poor)
Derse Devam Yüzdesi(ATD)	İyi(Good), Orta(Average), Kötü(Poor)
Okula ulaşım Süreleri(TT)	Geniş(Large),Orta(Average), Küçük(Small)

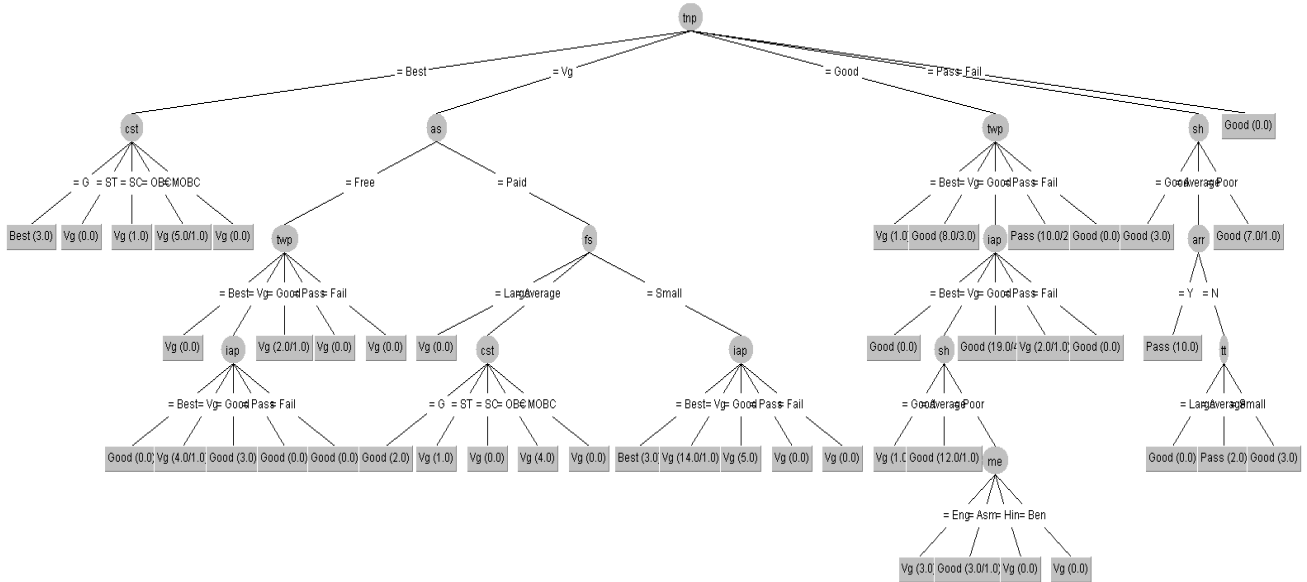
Kullanılan Dil(ME)	İngilizce(Eng),Hintçe(Hin),Assam(Asm),Bengalce(Ben)			
Baba Yeterliliği(FQ)	Okuma yazma bilmeyen(II),X Sınıfı seviyesi altında(Um),XII Seviyesi, Degree-sanat bilim sınavını geçen,PG			
Anne Yeterliliği(MQ)	Okuma yazma bilmeyen(II),X Sınıfı seviyesi altında(Um),XII Seviyesi(10), 12,sanat bilim sınavını geçen(Degree),PG			
Baba (FO)	Servis(Service), Diğerleri(Other)	Ticaret(Business),	Emekli(Retired),	Çiftçi(Farmer),
Anne (MO)	Servis(Service), Diğerleri(Other)	Ticaret(Business),	Emekli(Retired),	Çiftçi(Farmer),

Tablo 1: Veri kümesi Açıklaması

4. Veri Analizi

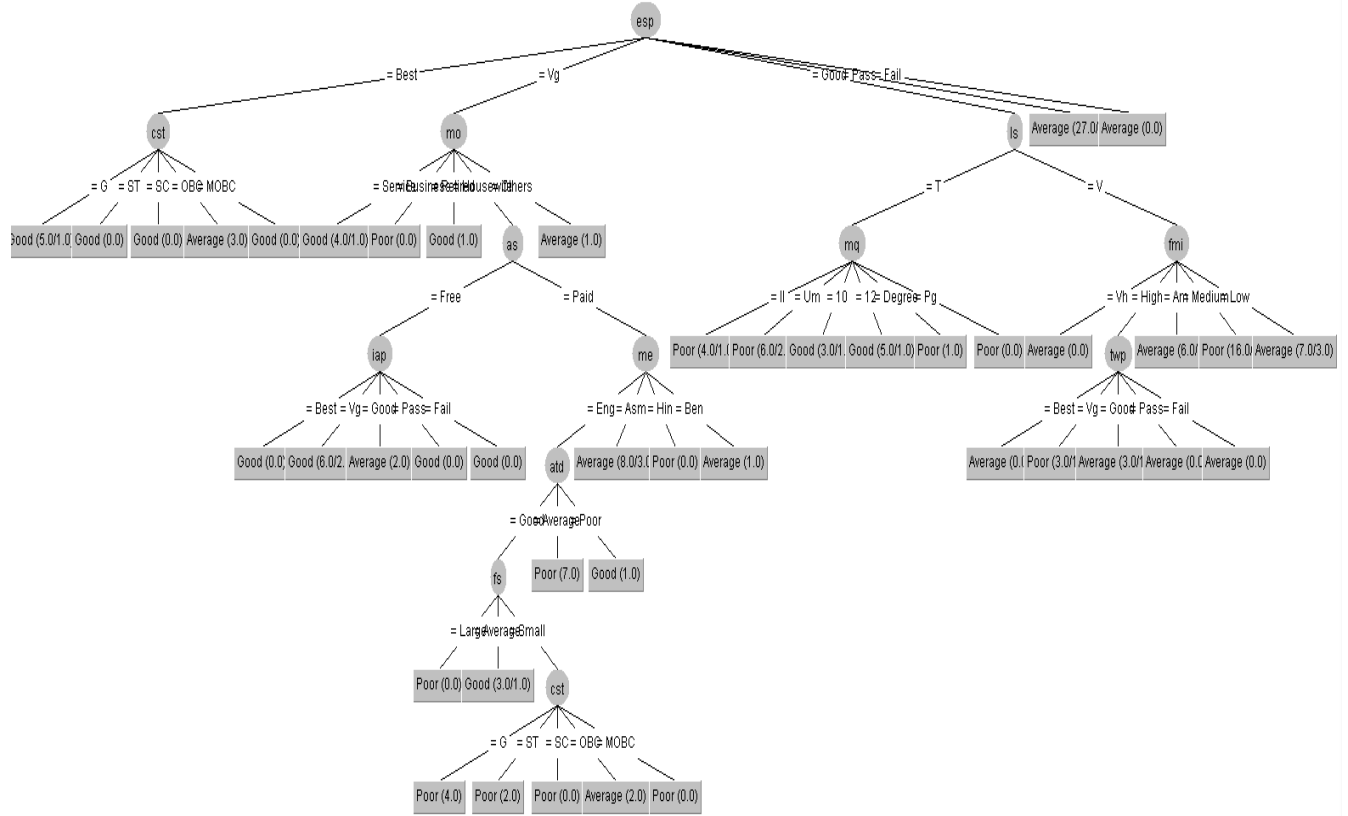
Waikato Üniversitesinde geliştirilmiş olan WEKA Programının (Waikato Environment for Knowledge Analysis) 3.8.3 sürümü kullanılmıştır. WEKA Programı, birçok sınıflandırma, kümeleme ve birliktelik kurallarına ait algoritmayı desteklemektedir ve tamamen açık kaynak kodlu bir yazılımdır. WEKA, arff dosya tipinin yanı sıra data, csv, c45, libsvm, dat, bsi, xrff, xrff.gz dosya tiplerini de desteklemektedir. Mevcut veriseti ile yapılan uygulama sonucunda izleyen tablodaki sonuçlar algoritmaların performans derecelerine göre sıralanarak sunulmuştur:

İlk olarak çalışmada karar ağacı algoritmalarının doğruluğunu ve sonucunu kötü etkilememesi sebebiyle veri üzerinde temizleme işlemleri yapılmıştır. Bunlardan “medeni hali” bilgisinde tek tip değer olduğu(bekar) için veri üzerinde bir etkisi olmadığı gerekçesiyle veri setinden kaldırılmıştır. Bu çalışmada öğrencilerin kabul kategorisi(ücretli-ücretsiz), çalışma kategorisi(iyi-orta-kötü), dönem sonu yüzdesi(başarısına) karar ağacı algoritmalarıyla detaylı incelenmiştir.



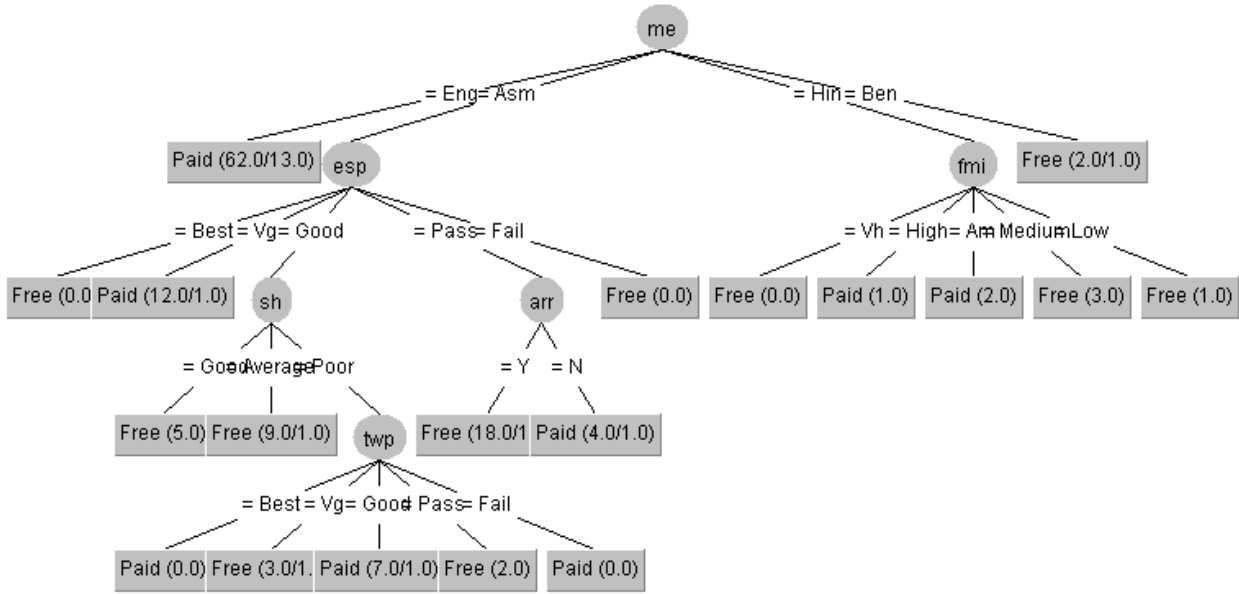
Şekil 1

Şekil 1 kullanılan algoritmalarından j48 ‘in karar ağacı dallanma görüntüsüdür. Kök düğüm olarak seçilen “Tnp: X sınıfı değerlendirme yüzdesi özneteliğinin bilgi kazancının en yüksek olduğu görülmektedir. Şekil içerisinde alt dallanmalar ve nihai sonuçları gösteren yapraklarda mevcuttur.



Şekil 2

Şekil 2 kullanılan algoritmalarından j48 ‘in karar ağacı dallanma görüntüsünde kök düğüm olarak seçilen “esp dönem sonu yüzdesi yani başarısını ifade eden özneteliğinin bilgi kazancının en yüksek olduğu görülmektedir. Öğrencinin çalışma saatleri özneteliği baz alınarak incelendiğinde aile gelir ve refah durumunun da etkisi olduğu görülmektedir.



Şekil 3

Şekil 3 de ise öğrencinin eğitimini ücretli ya da ücretsiz almasında etkili olan kök düğümlerde aile gelir durumu ve öğrencilerin akademik durumunun etkili öznelilikler olduğu gözlemlenmiştir. J48 karar ağacı ile yapılan bu sınıflandırmada %84.73 başarı oranı görülmektedir.

Çalışmanın Weka uygulaması üzerindeki sonuçların detaylı gösterimi paylaşılmıştır. Çalışmada karar ağaçlarına ait RandomForest, J48, RandomTree, DecisionStump, REPTree gibi bir çok algoritma denenmiş, başarımlar dereceleri tabloda karşılaştırılmıştır.

5.Sonuç ve Öneriler

Algoritma	Ele Alınan Öz Nitelik	Doğru sınıflandı rılan Örnek	Doğruluk Oranı %	Ağaç Boyutu	Çalışma süresi(sn)
Decision Stump	Çalışma saati(Sh)	60	45.80	-	0
	Başarı Yüzdesi(Esp)	75	57.21	-	0
	Kabul Kategorisi(As)	91	69.46	-	0
J48	Çalışma saati(Sh)	101	77.09	56	0.02

	Başarı Yüzdesi(Esp)	114	87.02	61	0
	Kabul Kategorisi(As)	111	84.73	25	0
REPTree	Çalışma saati(Sh)	75	57.25	11	0.01
	Başarı Yüzdesi(Esp)	91	69.46	14	0
	Kabul Kategorisi(As)	101	77.09	10	0
RandomForest	Çalışma saati(Sh)	131	100	-	0.05
	Başarı Yüzdesi(Esp)	131	100	-	0.03
	Kabul Kategorisi(As)	131	100	-	0.01
RandomTree	Çalışma saati(Sh)	131	100	229	0
	Başarı Yüzdesi(Esp)	131	100	221	0
	Kabul Kategorisi(As)	131	100	166	0

Tablo 2: Veri Seti Üzerine Uygulanan Karar Ağacı Algoritmalarının Karşılaştırılması

Karar ağaçları algoritmaları ile veri seti üzerindeki etkili faktörlerin birbirleriyle olan ilişkisi incelenmiş dönem sonu yüzdesi (Esp), çalışma saati(Sh) ve kabul kategorisi(As) özneliğine göre tahmin modelleri oluşturulmuştur. Karar ağacı teknikleri ve algoritmaları performansları ve doğrulukları test edilmiş ve doğrulanmıştır.

Bu çalışmada, toplanan bir eğitim veri setinden öğrencilerin notlarını etkili bir şekilde tahmin edebilecek niteliksel tahmin modelleri oluşturmak için birçok veri madenciliği yöntemi kullanılmıştır. İlk olarak, üniversite öğrencilerini hedef alan ve bunlarla ilgili birden fazla kişisel, sosyal ve akademik veri toplayan bir anket yapılmıştır. İkincisi, toplanan veri seti önceden işlenmiş ve veri madenciliği görevlerine uygun olması için düzenlenmiştir. Üçüncüsü, veri madenciliği görevlerinin uygulanması, sınıflandırma modelleri oluşturmak ve bunları test etmek için eldeki veri setinde sunulmuştur. Beş karar ağacı algoritması uygulanmıştır. Çalışmamızda veri seti az olmasına rağmen yüksek başarı elde edildiği görülmüştür. Özellikle Random Forest ve Random Tree algoritmalarında %100 başarı gözlemlenmiştir. En düşük başarı oranı ise Decision Stump algoritmasında elde edilmiştir. Bu çalışmada, öğrencinin akademik performansının tamamen çabalarına bağlı olmadığı, ancak, daha büyük etkilere sahip birçok çevresel faktörün olduğu da tespit edilmiştir. Literatürdeki sonuçlarla elde etmiş olduğumuz sonuçlar karşılaştırılmış ve sonuçların benzer özellikler gösterdiği ve bu benzerlikten yola çıkılarak öğrencilerin akademik performansının sınıflandırmasında karar ağaçlarından yararlanılabileceği sonucuna varılmıştır.

Sonuç olarak, bu çalışma, üniversitelere ve öğrencilere birçok yönden yardımcı olabilecek sonuçlar bulmak için üniversiteleri, öğrencilerinin verileri üzerinde düzenli olarak veri madenciliği görevleri gerçekleştirmeleri için motive edebilir ve yardımcı olabilir.

6.Kaynakça

1. Hijazi, S.T. and S.M.M.R. Naqvi, Factors Affecting Students' Performance, A Case Of Private Colleges. Bangladesh e-Journal of Sociology, 2006. 3(1): p. 10. Bodur,Ş., Soysal,Ş., Otizmin Erken Tanısı ve Önemi, 2004. cilt 13 ,sayı 10, 394-398
2. Bhardwaj, B.K. and S. Pal, Data Mining: A prediction for performance improvement using classification. International Journal of Computer Science and Information Security, 2011. 9(4): p. 5.
3. Strecht, P., et al., A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance. Proceedings of the 8th International Conference on Educational Data Mining, 2015: p. 3.
4. Kabakchieva, D., Predicting Student Performance by Using Data Mining Methods for Classification. Cybernetics and Information Technologies, 2013. 13(1): p. 12.
5. Sumitha, R. and E.S. Vinothkumar, Prediction of Students Outcome Using Data Mining Techniques. International Journal of Scientific Engineering and Applied Science (IJSEAS), 2016. 2(6): p. 8.
6. Baradwaj, B.K. and Pal, S., 2011. Mining Educational Data to Analyze Students' Performance. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011.
7. Ahmed, A.B.E.D. and Elaraby, I.S., 2014. Data Mining: A prediction for Student's Performance Using Classification Method. World Journal of Computer Application and Technology, 2(2), pp.43-47.
8. Ahmad, F., N.H. Ismail, and A. Abdulaziz, The Prediction of Students' Academic Performance Using Classification Data Mining Techniques. Applied Mathematical Sciences, 2015. 9(129): p. 12.
9. Khasanah, A.U. and Harwati, A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques. IOP Conf. Series: Materials Science and Engineering, 2017. 215(012036): p. 7.
10. Nichat, A.A. and D.A.B. Raut, Analysis of Student Performance Using Data Mining Technique. International Journal of Innovative Research in Computer and Communication Engineering, 2017. 2007(An ISO 3297): p. 5.
11. Almarabeh, H., Analysis of Students' Performance by Using Different Data Mining Classifiers. IJ. Modern Education and Computer Science, 2017. 9(8): p. 9-15.
12. Safavian S.R., Landgrebe D., 1991, "A survey of decision tree classifier methodology", IEEE Transactions on Systems Man and Cybernetics, 21, 660-674.
13. Göral,MA., "Kredi Kartı Başvuru Aşamasında Sahtecilik Tespiti İçin Bir Veri Madenciliği Modeli – Yük. Lisans Tezi İ.T.Ü. Fen Bilimleri Ens.. Ocak 2007- AKPINAR, 2000, s. 17.

14. Geetha, A. ve Nasira, G.M. (2014) “Data Mining for Meteorological Applications: Decision Trees for Modeling Rainfall Prediction” IEEE International Conference on Computational Intelligence and Computing Research.
15. Quinlan J.R., 1993, “C4.5: Programs for Machine Learning”, Morgan Kaufmann, San Mateo, CA,302 s
16. Breiman L., Friedman J.H., Olshen R.A. and Stone C.J., 1984, “Classification and Regression Trees”Monterey, CA: Wadsworth, 358 s
17. Mingers J., 1989, “An empirical comparison of pruning methods for decision tree induction”,Machine Learning, 4, 227–243
18. Yücebaş,S.,2018.Karmaşık Hastalıkların Teşhisinde Veri Madenciliği Yöntemlerinin Başarım Karşılaştırması, Çanakkale Onsekiz Mart University, Journal of Graduate School of Natural and Applied Sciences, 2018:4,1, 14-27///// Quinlan J.R., 1986. Induction of Decision Trees. Machine Learning. 1(1):81-106
19. Kavzoğlu, T., Çölkesen, İ., ”Karar Ağaçları İle Uydu Görüntülerinin Sınıflandırılması: Kocaeli Örneği” Harita Teknolojileri Elektronik Dergisi 2010, 2(1) 36-45
20. Bounsaythip, C. ve Esa, R. R. (2001). “Overview of Data Mining For Customer Behavior Modeling”, VTT Information Technology Research Report, Version:1, s.1-53
21. Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The Elements of Statistical Learning. Springer: 2001, pp. 269-272
22. Breiman L., Friedman J.H., Olshen R.A. and Stone C.J., 1984, “Classification and Regression Trees” Monterey, CA: Wadsworth, 358 s