

## Karar Ağaçlarıyla Öğrenci Verilerinin Sınıflandırılması

Mehmet Ali ALAN<sup>(\*)</sup>

**Özet:** Bu çalışmada, karar ağaçları yöntemiyle Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Fakültesi öğrencilerine ait veriler kullanılarak veri madenciliği yapılmıştır. Öğrencilere ait verilerden yararlanarak, hem bu verileri en başarılı sınıflandıran karar ağaçlarına ait algoritma, hem de bu algoritmanın üreteceği sınıflar tespit edilmeye çalışılmıştır. Çalışmanın sonucunda LADTree algoritmasının öğrenci verilerini sınıflandırmada en başarılı algoritma olduğu ve ondokuz değişik sınıf ürettiği anlaşılmıştır.

**Anahtar kelimeler:** Veri Madenciliği, Sınıflandırma, Karar Ağaçları  
**Classification of Students' Data by Decision Trees**

**Abstract:** In this study, a data mining application was conducted using the data from students of Faculty of Economics and Administrative Sciences in Cumhuriyet University. Both decision tree algorithm which classifies the data best and classes produced by this algorithm were tried to determine by benefiting from the students' data. As a result of the study, it is explored that LADTree algorithm was the best algorithm which classifies the students' data and nineteen classes were produced by this algorithm.

**Key Words:** Data Mining, Classification Analysis, Decision Trees

### I. Giriş

Veritabanları, rasyonel karar almayı sağlayacak gizli bilgiler bakımından zengindir. Sınıflandırma ve tahmin, gelecek veri trendlerinin tahmini veya önemli veri sınıflarının açıklanmasında kullanılan iki önemli veri analiz tekniğidir. Bu analizler büyük miktarlardaki verilerin daha iyi anlaşılmasında kullanışlı olabilmektedir (Han and Kamber, 2006:285).

Veritabanlarındaki veriler üzerinde farklı disiplinler, farklı amaçlarla istatistiksel ya da matematiksel analizler yapmaktadırlar. Söz konusu analiz tekniklerinden biri de veri madenciliği yöntemidir.

Günümüzde kurumlar büyük miktarlarda veri üretmekte, ancak bu veriler içinde anlamlı ve yararlı bilgiyi ortaya çıkarmakta zorluklar yaşamaktadırlar. Geleneksel istatistik yöntemlerle büyük boyuttaki veriyi çözümlenmek kolay değildir. Bu nedenle verileri işlemek ve çözümlenmek için özel yöntemlere gereksinim duyulmuştur. Veri madenciliği yöntemleri bu gereksinimi karşılamak üzere ortaya çıkmıştır (Özkan, 2008:4).

Bu çalışmanın amacı, veri madenciliği tekniğini kullanarak, Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Fakültesi öğrencilerine ait veriler yardımıyla sınıflandırma analizi yapmaktır. Bu amaçla adı geçen fakülte'deki mevcut kayıtlı olan öğrencilere ait veriler ele alınarak, veri madenciliğinin en yaygın kullanılan tekniklerinden "Karar Ağaçları" yöntemi kullanılmıştır.

<sup>(\*)</sup> Yrd. Doç. Dr. Cumhuriyet Üniversitesi, İİBF Yönetim Bilişim Sistemleri Bölümü.

Çalışma üç bölümden oluşmaktadır. Birinci bölümde sınıflandırma, Veri madenciliği ve Karar Ağaçları ile ilgili tanımlayıcı açıklamalar yer almaktadır. İkinci bölümde konuyla ilgili yapılmış çalışmalar gözden geçirilmektedir. Üçüncü bölümde ise Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Fakültesi öğrencilerine ait veriler üzerinde veri madenciliği yapılmıştır.

## II. Veri Madenciliği, Sınıflandırma ve Karar Ağaçları

Veri madenciliği, hem yararlı hem de anlaşılabilir verilerle, alışılmamış yollarla, verileri özetleyen ve gizli ilişkileri ortaya koyan bir analiz yöntemidir (Larose, 2006). Bu yöntem, öncelikle bilinmeyen desenlerin ortaya konması amacıyla bilimsel ve teknik veri araştıran, veritabanındaki bilgi keşfi süreçlerinden biridir (Rokach and Maimon, 2005:2).

Disiplinler arası nitelik taşıyan veri madenciliğini en yaygın kullanan bilim dalları; veritabanı sistemleri, istatistik, matematik, makine öğrenmesi, görselleme ve bilişim bilimleridir (Han and Kamber, 2006:29). Veri madenciliği, verinin bütünü kullanması bakımından diğer istatistiksel verilerden ayrılmaktadır. Bu yöntemle, geleneksel yollarla elde edilmiş küçük verilerle çalışma yerine daha kolay değerlendirme yapabilecek, yeni bağımsız veriler tercih edilebilmektedir (Weiss and Zhang, 2003:426).

Gartner Group'a göre veri madenciliği, istatistiksel ve matematiksel yöntemler kadar desen tanıma teknolojilerinin kullanılmasıyla, muazzam miktarlardaki depolanmış verilerin elenmesi ile yeni anlamlı birliktelikler, desenler ve trendler keşfetme sürecidir (Larose, 2005:2).

Veri madenciliği, reklamcılık, biyoinformatik, veritabanı pazarlama, dolandırıcılık tespiti, e-ticaret, sağlık, güvenlik gibi alanların içinde olduğu, farklı alanlarda uygulanabilen, değişik bakış açısı ve çalışması ile veri analizinden bilgi keşfetme süreci olarak bilinir (Jain, 2011).

Birliktelik kuralları, kümeleme, karar ağaçları, diskriminant analizi, yapay sinir ağları, genetik algoritmalar vb. birçok veri madenciliği algoritması vardır. Bu algoritmalar sıradan bilgiyi bulup çıkarmak ve bir yöneticinin kararlarını yönlendirebilen özel bilgiye ulaşmak için çeşitli alanlardan elde edilen verileri işlemek amacıyla kullanılır (Wu and Li, 2003).

En yaygın veri madenciliği algoritmaları ve modelleri içinde karar ağaçları, sınıflandırma ağaçları olarak ta adlandırılır (Bramer, 2007:6); birliktelik kuralları, kümeleme, sınıflandırma, çoklu lineer regresyon, sıralı örüntüler ve zaman serileri tahmini, örüntü tanıma ve özelliklerinin belirlenmesi sayılabilir. Sınıflandırma, regresyon ve zaman serisi analizleri gizli örüntülerin ortaya çıkarılmasında ve şekillendirilmesinde uygun iken, birliktelik kuralları, kümeleme ve sırasal keşif yaklaşımları, hava tahmini ve şiddeti araştırmak ve tanımlamak için yararlı araçlar olabilir (Tadesse 2009).

Sınıflandırma, günlük yaşamda çok sıklıkla başvurulan bir işlemdir. Sınıflandırma ile nesnelere bölünerek ayrıştırılır, yani karşılıklı olarak özel ya da genel kategorilerden her biri bir sınıf olarak atanabilir. Pek çok pratik karar

verme işlemi, bir sınıflandırma problemi olarak formüle edilebilir. Örneğin kişiler ya da nesnelere birçok kategoriden biri olabilir (Bramer, 2007:23).

Sınıflandırma, farklı sınıflardaki, değişik öğeleri ayırma sürecidir. Bu sınıflar, iş kuralları, sınıf sınırları veya bazı matematiksel fonksiyonlar olabilir. Sınıflandırma işlemi, sınıflandırılmış olan öğenin, bilinen bir sınıf değeri ile özellikleri arasındaki bir ilişki üzerine bina edilebilir. Bu sınıflandırma tipi, “denetimli öğrenme” olarak isimlendirilir. Eğer bir sınıfın bilinen örnekleri yoksa bu sınıflandırma denetimsizdir. En yaygın denetimsiz sınıflandırma yaklaşımı “kümeleme”dir. Kümeleme teknolojisinin en yaygın uygulamaları, perakende ürünlerde birliktelik analizi (market sepet analizi) ve dolandırıcılık tespiti (Nisbet, et al., 2009: 235).

Veri madenciliğinde denetimli öğrenme kavramı, bir sınıflandırma ile bilinen veriler temelinde bir sınıflandırma fonksiyonu öğretmek ya da bir sınıflandırma modeli inşa etmektir. Bu fonksiyon ya da model, veri tabanındaki verileri hedef niteliklere dönüştürür, dolayısıyla yeni veriler sınıf tahmininde kullanılabilir (Dong-Peng, et al., 2008:36).

Karar ağacı algoritması, veri madenciliği sınıflandırma algoritmalarından biridir ve bilgi teorisi ilkelerine dayanmaktadır. Bir karar ağacı algoritması kullanıcıların, gürültülü ve bilinmeyen verilerin üstesinden gelmesini ve kolayca yorumlanmasını sağlayan otomatik olarak dağıtabilen modeller oluşturur. Bir karar ağacı, karar ağaçlarının yapısını temsil eden düğümleri dallar ve yaprak düğümlerinden oluşur. Ağacın üst düğümü kök düğümü olarak adlandırılır; içindeki düğümler değişkenlerin değerleri üzerine ifa edilmiş olan testleri temsil ederler; dallar testler üzerindeki farklı sonuçları ve yaprak düğümleri örneklerin sınıfını temsil eder (Dong-Peng vd., 2008).

Karar ağaçları, sınıflandırma ve tahmin açısından güçlü ve popüler araçlardır. Bu yöntemin çekici tarafı, yapay sinir ağlarının aksine, karar ağaçlarının kuralları temsil etmesidir. Başka bir deyişle bunları yorumlamak daha kolaydır (Nisbet et al. 2009:465).

1970’lerin sonu ve 1980’lerin başlarında bir makine öğrenme araştırmacısı olan J. Ross Quinlan, ID3 (Iterative Dichotomiser) olarak bilinen bir karar ağacı algoritması geliştirdi. Bu çalışma ile daha önce E. B. Hunt, J. Marin ve P. T. Stone tarafından tanımlanmış olan *kavram öğrenme sistemleri* genişletilmiş oldu. Quinlan daha sonra ID3’ün bir devamı olan C4.5’i oluşturdu. Bu, daha yeni öğrenme algoritmaları için genellikle bir benchmark görevi görmektedir. 1984’de bir grup istatistikçi (L. Breiman, J. Friedman, R. Olshen ve C. Stone), ikili karar ağaçlarının oluşumunu açıklayan “Sınıflandırma ve Regresyon Ağaçları” (CART) adlı kitabı yayınladılar. ID3 ve CART aynı zamanda birbirinden bağımsız olarak ortaya çıktılar. Ancak eğitim değişkenlerinden (demetlerinden) öğrenme karar ağaçları için benzer bir yaklaşım ortaya koydular. Bu iki köşe taşı algoritma karar ağacı induksiyonu ile ilgili çok sayıda çalışmaya kaynaklık etti (Han and Kamber, 2006:292).

Veri madenciliğinde bir karar ağacı, hem sınıflandırıcıları, hem de regresyon modellerini temsil edecek şekilde kullanılabilir. Diğer yandan operasyonel araştırmalarda karar ağaçları, hiyerarşik kararlar modeli ve onların sonuçları ile ilişkili bilgiler ortaya koyar. Karar vericiler olması en fazla muhtemel hedefine ulaşacak şekilde strateji geliştirmek için karar ağaçlarını kullanır. Bir karar ağacı sınıflandırma amacıyla kullanıldığı zaman, genellikle “sınıflandırma ağacı”, regresyon amacıyla kullanıldığında ise “regresyon ağacı” olarak adlandırılır (Rokach and Maimon:2008:5).

Karar ağaçları hâlihazırdaki kullanımıyla en popüler tüme varım metodudur. Karar ağaçları genellikle iki aşamada oluşturulurlar. Büyüme olduğu zaman bu algoritma her bir düğümde sınıflar arasındaki en iyi özellik ayırt ediciyi (veri alt seti) ortaya çıkarır ve daha sonra o özelliğe dayalı olarak bu verileri iki yeni düğüm halinde bölümlere ayırır. Bu, her bir tabaka için bir sınıf tahsis edilinceye kadar ortaya çıkan veri alt setine tekrar tekrar uygulanır. İndirgemenin ikinci aşaması en iyi dengeye erişebilmek için ağacın en az yararlı dallarını kesmek suretiyle işletilir. Daha basit bir model genellikle daha sağlamdır. Yani yeni veriler hakkında daha doğru sonuçlar ortaya çıkarır. Nihai ağaç etiketlenmiş birkaç bölgedeki bu özel alanı bölümlere ayırır (Nisbet et al. 2009:300).

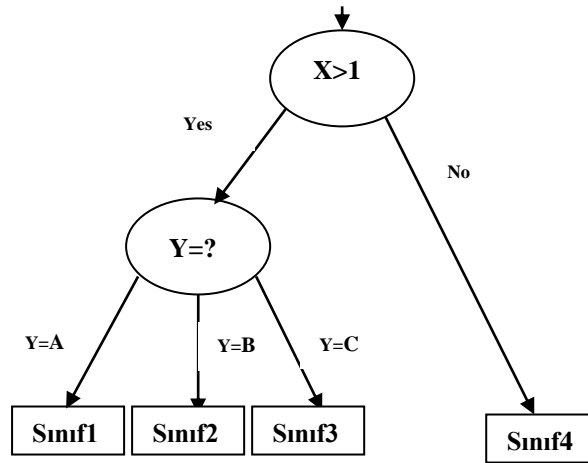
Karar ağacı yöntemlerinde parametrik istatistiksel varsayımlar yapılmaz. Öngörüler terminal düğümlerde birkaç mantıksal if-then şartıyla sunulabilirler. Normal bir veri dağılımında veya değişkenler ve tepki değişkeni arasındaki lineer ilişkilerde örtülü varsayımlar yoktur. Karar ağacı yöntemleri, değişkenlerin öngörü sağlayabildikleri zamanın ötesini analistlerin bilmedikleri yerlerde veri madenciliği için oldukça uygundur. Bu nedenle karar ağacı yöntemleri ilişkileri açığa çıkarabilir ve onları daha fazla bilişimsel yoğun yöntemlerin gözden kaçırdığı birkaç karar kuralı halinde ifadelendirebilirler (Nisbet et al. 2009:278-279).

Bir karar ağacı, öz nitelikler değeri üzerine bölünme olarak bilinen bir işlem tarafından oluşturulur yani outlook gibi bir öznitelik testi ve ardından olası değerlerin herbiri için bir dal oluşturmadır. Sürekli öznitelikli test durumunda normal olarak değer “daha az ya da eşit” ya da “daha büyük” bölünmüş değer olarak bilinen bir değer verir. Herbir dal, sadece bir sınıflandırma ile etiketli oluncaya kadar bölünme işlemi devam eder (Bramer, 2007:43-44).

Çekici bir sınıflandırma yöntemi olan karar ağacı, kök düğümden aşağı doğru yaprak düğümlerinde sonlanıncaya kadar uzayan, dallar tarafından bağlantıları sağlanmış, karar düğümlerinin bir koleksiyonunu içerir. Karar ağacı diyagramının en üstüne yerleştirilmiş olan kök düğümden başlayarak, dalın her bir olası sonucu ile karar düğümleri test edilir. Herbir dal daha sonra diğer bir karar düğmesi ya da sonlandırma yaprak düğmesine yol açar (Larose, 2005:107).

Verilerinden sınıflandırıcılar üretmek için özellikle etkili bir yöntem, bir karar ağacı üretmektir. Karar ağacı betimlemesi en yaygın olarak kullanılan mantık yöntemidir. Öncelikli olarak makine öğrenmesi ve uygulamalı istatistik literatüründe açıklanan çok sayıda karar ağacı ortaya çıkaran algoritmalar bulunmaktadır. Onlar girdi-çıkı örnekleri kümesinden oluşan karar ağaçları gibi öğrenme yöntemleri denetlenmektedir. Tipik bir karar ağacı öğrenme sistemi yukarıdan aşağıya bir stratejiyi benimser. O basit ama en basit olmayı gerektirmeyen bir ağacın bulunacağını garanti eder. Bir karar ağacı düğümlerden oluşur. Bir düğümün giden dalları, düğümde testin tüm olası sonuçlarına karşılık gelmektedir. X ve Y gibi iki giriş özellikli/öznelikli sınıflandırma örnekleri için basit bir karar ağacı aşağıda verilmiştir. (Kantardzic, 2003:140)

İki giriş özelliklerini X ve Y ile numunelerin sınıflandırılması için basit bir karar ağacı aşağıdaki şekilde verilebilir:



Şekil 1: Basit Bir Karar Ağacı

### III. Literatür Özeti

Konuyla ilgili literatürde farklı veri setleri üzerinden yapılmış çok sayıda çalışma bulunmaktadır. Bunlardan DIMIĆ et al. (2011), Moodle elektronik öğrenme materyallerini kullanarak öğrenci verilerini toplamış ve bu veriler ile hem sınıflandırma, hem kümeleme, hem de birliktelik kuralları teknikleri ile analiz yapmışlardır. Dener ve Diğerleri (2005), lisansüstü öğrencilerine ait verileri kullanarak çeşitli algoritmalarla sınıflandırmalar yapmış ve açık kaynak kodlu yazılımların başarımlarını derecelerini karşılaştırmışlardır. Albayrak ve Yılmaz (2009), İMKB verilerinden yararlanarak karar ağaçları ile sınıflandırma analizi metoduyla veri madenciliği yapmışlardır. Dolgun vd. (2009), bir telekomünikasyon kurumuna ait verileri

kullanarak karar ağaçları algoritmalarından yararlanarak terk eden müşteriye ait profil modeli oluşturmuşlardır. Ayık vd. (2007), Atatürk Üniversitesi öğrencilerine ait veritabanındaki tüm verileri kullanarak sınıflandırma analizi yapmışlardır. Kaya vd. (2012), Epileptik EEG işaretlerini karar ağaçları ve karar kurallarını kullanarak sınıflandırmış ve tanı performanslarının oldukça yüksek olduğunu tespit etmişlerdir.

#### IV. Veri Seti ve Yöntem

Bu çalışmada Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Fakültesi öğrencilerine ait veriler kullanılmıştır. Fakülte de 8 bölüm bulunmakta ve hepsinde de ikili öğretim yapılmaktadır. Veriler hem birinci öğretim öğrencilerine, hem de ikinci öğretim öğrencilerine aittir. Bu bağlamda 4106 öğrenciye ait 53378 veri girişi yapılmıştır. 2012 sonu ve 2013 yılı başı itibarıyla kayıtlı olan öğrencilere ait veriler Cumhuriyet Üniversitesi Bilgi İşlem Daire Başkanlığından Excel formatında alınmıştır. Excel makroları kullanılarak veri ambarı hazırlanmıştır. Veri madenciliğine ait Karar Ağaçlarından yararlanılarak sınıflandırma analizi yapılmıştır. Gerekli dönüşümler yapıldıktan sonra veriler “veriset.arff” adlı metin dosyasına yazdırılmıştır.

Verilerde bölümler; CEEI, İKTİSAT, İSLETME, KAMU, MALİYE, YBS, EKONOMETRİ ve UTL olarak tanımlanmıştır. Sınıf sütunu, 1’den 4’e kadar sınıfları, cinsiyet, “Erkek ve Bayan” değerlerini almaktadır. Gelir sütunu “600 TL’den Az”, “601-1200”, “1201-1800”, “1801-2400”, “2400 Üzeri” olmak üzere 5 grupta tanımlanmıştır. Ancak analizlerin kolaylığı açısından bu değerler için sınıflandırma öncesi her bir grup için sırasıyla 1’den 5’e kadar değerler atanmıştır. Baba Sağ mı, Baba Çalışıyor mu, Anne çalışıyormu, Katkı kredisi alıyor mu, Öğrenim kredisi alıyor mu sütunları “Evet/Hayır” değerlerini içermektedir. Kardeş sayısı ve okuyan kardeş sayısı, kayıt yılı nümerik değer olarak tanımlanmıştır. Kayıt yılı ise 2007 ve öncesi 2007, diğerleri ise 2008, 2009, 2010, 2011, 2012 ve 2013 şeklinde değerler almaktadır. Sınıflar ise “Burs\_Alıyor”, “Burs\_Almıyor” şeklinde tanımlanmıştır. Verilerin tüm alanlarında herhangi bir değer tanımlanmayan ya da “Bilinmiyor” biçiminde yer alan değerler için “?” değeri kullanılmıştır.

#### V. Uygulama

Yapılan çalışmada Waikato Üniversitesinde geliştirilmiş olan WEKA Programının (Waikato Environment for Knowledge Analysis) 3.6.9 sürümü kullanılmıştır. WEKA Programı, açık kaynak kodlu bir yazılımdır. Bu program pek çok sınıflandırma, kümeleme ve birliktelik kurallarına ait algoritmayı desteklemektedir. WEKA, metin tabanlı arff, arff.gz, names, data, csv, c45, libsvm, dat, bsi, xrf, xrf.gz dosya tiplerinin yanı sıra, veritabanlarını ve verilerin olduğu URL adreslerini de desteklemektedir.

Mevcut veriseti ile yapılan uygulama sonucunda izleyen tablodaki sonuçlar algoritmaların performans derecelerine göre sıralanarak sunulmuştur:

**Tablo 1.** Karar Ağaçları Analizine Ait Bazı Algoritmaların Başarım Dereceleri

Algoritmalar	Doğru sınıflandırılan Örnek	Kappa İstatistiği	Ortalama Mutlak Hata	Ortalama Hata Karekök	Görelü Mutlak Hata %	Görelü Hata Karekök %	TP Oranı	FP Oranı	F-Ölçütü
LADTree	4049	0.8485	0.0299	0.1131	30.0898	50.7753	0.986	0.203	0.986
ADTree	4043	0.8349	0.0874	0.1437	87.8594	64.5106	0.985	0.203	0.984
LMT	4039	0.8219	0.0279	0.1216	28.0478	54.5937	0.984	0.225	0.983
SimpleCART	4038	0.8205	0.0292	0.1227	29.3201	55.0979	0.983	0.221	0.983
FT	4037	0.8208	0.0217	0.1259	21.8482	56.5276	0.983	0.207	0.983
NBTree	4036	0.8098	0.022	0.122	22.157	54.7567	0.983	0.251	0.982
BFTree	4034	0.8063	0.0272	0.1261	27.3738	56.6262	0.982	0.247	0.982
J48	4034	0.8005	0.031	0.1295	31.1889	58.1485	0.982	0.273	0.981
REPTree	4030	0.7915	0.0327	0.1329	32.8823	59.659	0.981	0.273	0.98
J48Graft	4027	0.7773	0.0326	0.1358	32.7751	60.9767	0.981	0.304	0.979
RandomForest	4027	0.789	0.0316	0.1326	31.8013	59.5396	0.981	0.256	0.98
RandomTree	3996	0.7176	0.0292	0.1587	29.4083	71.2549	0.973	0.287	0.973
DecisionStump	3995	0.6397	0.0526	0.1622	52.8582	72.7931	0.973	0.489	0.968

Yapılan uygulama çalışmasında karar ağaçlarına ait pek çok algoritma denenmiş, başarım dereceleri tabloda verilmiştir. Bu çalışmada karar ağaçlarına ait; LADTree, ADTree, LMT, SimpleCart, FT, NBTree, BFTree, J48, REPTree, J48Graft, RandomForest, RandomTree ve DecisionStump algoritmaları kullanılarak modeller oluşturulmuş ve oluşturulan modellerin başarım dereceleri karşılaştırılmıştır.

LADTree algoritması, 4049 doğru sınıflandırılmış örnek derecesiyle en başarılı algoritma olarak gözükmektedir. Bu algoritmanın, sınıflararası uyumu veren kappa istatistiği 0.8485, birinci sınıftaki doğru olarak sınıflandırılmış kayıtların sayısını veren TP (True Positive) oranı 0.986, birinci sınıfta sınıflandırılmış, ikinci sınıftaki kayıtların sayısını veren FP (False Positive) oranı 0.203 olarak elde edilmiştir. Kesinlik ve duyarlılığın harmonik ortalaması olan F-ölçütü (Coşkun ve Baykal) ise 0.986 olarak bulunmuştur. LADTree algoritmasını, ADTree algoritması 4043 doğru örnek sınıflandırmasıyla takip etmektedir. Bu algoritmaların kappa istatistiği 0.8349, TP oranı 0.985, FP oranı, 0.203 ve F-Ölçütü 0.984 olarak bulunmuştur. Daha sonra ise diğer algoritmalar gelmektedir. F-ölçütü aşağıdaki formülle elde edilmektedir:

$$F - \text{Ölçütü} = \frac{2 \times \text{Duyarlılık} \times \text{Kesinlik}}{\text{Duyarlılık} + \text{Kesinlik}}$$

Formülde kullanılan kesinlik ve duyarlık ise aşağıdaki şekilde hesaplanmaktadır:

$$\text{Kesinlik} = \frac{TP}{TP + FP}$$

$$\text{Duyarlılık} = \frac{TP}{TP + FN}$$

Sınıflandırma başarısı en iyi olarak çıkan LADTree algoritmasının ürettiği sınıflar ise aşağıda verilen şekildeki gibidir:

```

: 0,0
| (1)Ogretim_kredi = Evet: 1.568,-1.568
| (1)Ogretim_kredi = Hayir: -1.186,1.186
| | (2)Kayit_yil = 2007: 1.766,-1.766
| | (2)Kayit_yil != 2007: -0.491,0.491
| | | (3)Kardes_say = 0: 0.861,-0.861
| | | | (4)Baba_Sagmi = Evet: -0.658,0.658
| | | | (4)Baba_Sagmi = Hayir: 1.216,-1.216
| | | | | (5)Gelir = 1: -1.611,1.611
| | | | | (5)Gelir != 1: -1.611,1.611
| | | | (3)Kardes_say != 0: -0.393,0.393
| (6)Sinif = 1: -0.395,0.395
| | (9)Baba_Sagmi = Evet: -0.504,0.504
| | (9)Baba_Sagmi = Hayir: 0.619,-0.619
| (6)Sinif != 1: 0.083,-0.083
| | (7)Kayit_yil = 2012: 1.945,-1.945
| | (7)Kayit_yil != 2012: -0.037,0.037
| | | (8)Bolum = ISLETME: 0.361,-0.361
| | | (8)Bolum != ISLETME: -0.12,0.12
Legend: Burs Aliyor, Burs Almiyor

```

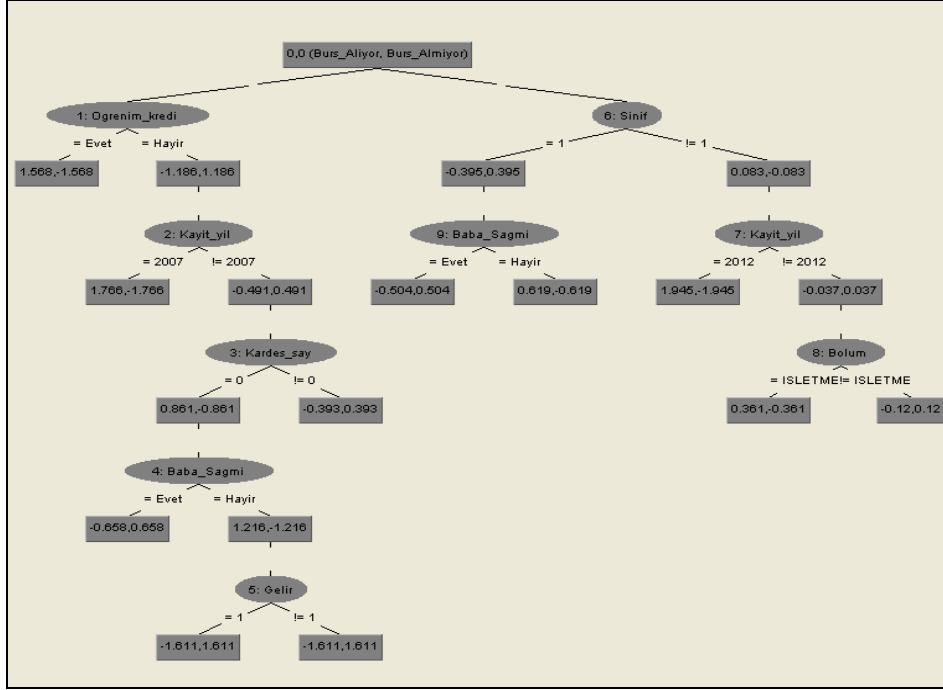
Şekil 2: LADTree Algoritmasının Ürettiği Sınıflar

Tam eğitimli set kullanılarak LADTree algoritması ile üretilen karar ağacı 28 dal ve 19 yapraktan (sınıf) oluşmuştur:

1. Öğrenim Kredisi alanlar bir sınıfta toplanmıştır.
2. Öğrenim Kredisi almayanlar bir sınıfta toplanmıştır.
3. Öğrenim kredisi almayan, kayıt yılı 2007 ve öncesine ait olanların sınıfı.
4. Öğrenim kredisi almayan ve kayıt yılı 2007'den sonra olanların sınıfı.
5. Öğrenim kredisi almayan, kayıt yılı 2007'den sonra ve kardeş sayısı sıfırdan farklı olanların sınıfı.



6. Öğrenim kredisi almayan, kayıt yılı 2007'den sonra ve kardeş sayısı sıfır olanların sınıfı.
7. Öğrenim kredisi almayan, kayıt yılı 2007'den sonra, kardeş sayısı sıfır olan ve babası sağ olanların sınıfı.
8. Öğrenim kredisi almayan, kayıt yılı 2007'den sonra, kardeş sayısı sıfır olan ve babası sağ olmayanların sınıfı.
9. Öğrenim kredisi almayan, kayıt yılı 2007'den sonra, kardeş sayısı sıfır olan, babası sağ olmayan ve gelir durumu 650 TL'den az olanların sınıfı.
10. Öğrenim kredisi almayan, kayıt yılı 2007'den sonra, kardeş sayısı sıfır olan, babası sağ olmayan ve gelir durumu 650 TL'den farklı olanların sınıfı.
11. Sınıfı 1 (bir) olanlar bir sınıfta toplanmıştır.
12. Sınıfı 1 (bir) olanlar ve babsı sağ olanların sınıfı.
13. Sınıfı 1 (bir) olanlar ve babsı sağ olmayanların sınıfı.
14. Sınıfı 1 (bir)den farklı olanlar bir sınıfta toplanmıştır.
15. Sınıfı 1 (bir)den farklı olan ve kayıt yılı 2012 olanların sınıfı.
16. Sınıfı 1 (bir)den farklı olanve kayıt yılı 2012 den farklı olanların sınıfı (0.037).
17. Sınıfı 1 (bir)den farklı olan ve kayıt yılı 2012 den farklı olan ve bölümü İşletme olanların sınıfı.
18. Sınıfı 1 (bir)den farklı olan, kayıt yılı 2012 den farklı olan, bölümü İşletmeden farklı olanların sınıfı.
19. Burs alanlar ve burs almayanlar şeklindeki ana sınıf.  
Satırların sonundaki değerler ise o sınıfın hangi aralıktaki değerleri aldığını göstermektedir.  
LADTree algoritmasının görsel ağacı ise aşağıdaki gibi oluşmuştur.



Şekil 3. LADTree Algoritması ile üretilen Görsel Ağaç

LADTree algoritmasının sunduğu görsel ağaç yapısından da anlaşılacağı gibi herbir dikdörtgen bir sınıfı temsil etmektedir. Şekilde ana sınıf tanımında yer alan (Burs\_Aliyor, Burs\_Almıyor) ile birlikte 19 sınıf oluşmuştur. Elips biçimli şekiller ise koşulları ifade etmektedir.

## VI. Sonuç ve Değerlendirme

Bu çalışmada, sınıflandırma yöntemlerinden karar ağaçları kullanılarak öğrencilere ait verilerle analiz yapılmıştır. Veri madenciliği, gizli, önemli, önceden bilinmeyen, yararlı bilgileri ortaya koyan bir veri analiz tekniğidir. Bu yöntemle, alışıl gelmiş analiz tekniklerinden farklı olarak, yalnızca sayısal verilerle değil, sayısal olmayan veriler ile de analizler yapılabilmekte ve gizli örüntüler ortaya çıkarılabilmektedir. Veri madenciliği, veri kümesinden rastgele örnek seçmek yerine, mevcut verilerin tümünü kullanması bakımından da pek çok analiz tekniğinden ayrılır. Karar ağaçları, veri madenciliğinin yalnızca sınıflandırma türünde kullanılan analiz tekniklerinden biridir. Karar ağaçları ile üretilen sonuçları yorumlamak, görsel özelliğinden dolayı oldukça kolaydır. Bu çalışmada, mevcut verilerle karar ağaçlarına ait pek çok algoritma denenmiş ve 4049 doğru sınıflanan örnek sayısı ile en başarılı algoritma olarak LADTree algoritması bulunmuştur. En çok doğru sınıflandırma yapan bu algoritma,

toplam 19 sınıf üretmiştir. Bu analizler sonucunda, öğrencilere ait verilerden yararlanılarak hangi sınıfa düşecekleri tahmin edilebilmektedir.

### Kaynaklar

- Albayrak, A.Sait ve Yılmaz, Şebnem (2009), **Veri Madenciliği: Karar Ağacı Algoritmaları Ve İMKB Verileri Üzerine Bir Uygulama** Süleyman Demirel Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt.14, Sayı.1, ss.31-52
- Ayık, Y. Ziya, Özdemir, Abdülkadir ve Yavuz Uğur (2007), **Lise Türü Ve Lise Mezuniyet Başarısının, Kazanılan Fakülte İle İlişkinin Veri Madenciliği Tekniği İle Analizi**, Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, Cilt 10, Sayı 2, ss.441-454
- Baykal, Abdullah ve Coşkun Cengiz (2011), **Veri Madenciliğinde Sınıflandırma Algoritmalarının Bir Örnek Üzerinde Karşılaştırılması**, <http://ab.org.tr/ab11/bildiri/67.pdf>, 18.01.2013
- Bose, I., Chun, L. A., Yue, L. V. W., Ines, L. H. W. and Helen, W. O. L., (2009), **Business Data Warehouse: The Case of Wal-Mart**, Data Mining Applications for Empowering Knowledge Societies, Ed. Hakikur Rahman, Information Science Reference, pp.189-198
- Bramer, Max (2007), **Principles of Data Mining**, Springer, London
- Dener, Murat, Dörerler Murat ve Orman Abdullah, (2009), **Açık Kaynak Kodlu Veri Madenciliği Programları: WEKA'da Örnek Uygulama**, [ab.org.tr/ab09/bildiri/42.pdf](http://ab.org.tr/ab09/bildiri/42.pdf), 18.01.2013
- Dolgun ve Diğerleri (2009), **Veri Madenciliğinde Yapısal Olmayan Verilerin Analizi: Metin ve Web Madenciliği**, İstatistikçiler Dergisi 2, ss.48-58
- Dong-Peng Yang, Li Jin-Lin, Lun Ran and Chao Zhou, (2008), Applications of Data Mining Methods in the Evaluation of Client Credibility, *Applications of Data Mining in E-Business and Finance* C. Soares et al. (Eds.), IOS Press, Amsterdam, pp.35-43
- Giudici, Paolo and Figini, Silvia, (2009), **Applied Data Mining For Business and Industry**, Second Edition, Wiley Publication, West Sussex
- Han, Jiawei and Kamber, Micheline, (2006), **Data Mining: Concepts and Techniques**, *Second Edition*, Morgan Kaufmann Publications, San Francisco
- Ivancsy, Renata and Vajk, Istvan, (2005), "Fast Discovery Of Frequent Itemsets: A Cubic Structure-Based Approach", *Informatica* 29, pp.71-78
- Jain, Y. K., Yadav, V. K. and Panday, G. S., (2011), "An Efficient Association Rule Hiding Algorithm for Privacy Preserving Data Mining", *International Journal On Computer Science And Engineering*, Vol. 3 No. 7, pp. 2792-2798.

- Kantardzic, Mehmed , (2003). **Data Mining: Concepts, Models, Methods, and Algorithms**, John Wiley & Sons J. B. Speed Scientific School, University of Louisville IEEE Computer Society
- Kaya Yılmaz, Ertuğrul Ö. Faruk ve Tekin Ramazan (2012), Batman University International participated Science and Culture Symposium, Batman University Journal of Life Sciences, Volume 1, Number 2, ss.403-413
- Larose, Daniel T., (2005), **Discovering Knowledge In Data**, Wiley Publication, New Jersey.
- Larose, Daniel T., (2006), **Data Mining Methods and Models**, A John Wiley & Sons, Inc., Publication, New Jersey
- Nisbet, R., Elder, J., and Miner, G., (2009), **Handbook of Statistical Analysis and Data Mining Applications**, Elsevier Inc, Burlington.
- Özkan, Yalçın (2008), **Veri Madenciliği Yöntemleri**, Papatya Yayınları, İstanbul
- Rokach, Lior and Maimon, Oded (2008), **Data Mining with Decision Trees**, World Scientific, New Jersey
- Tadesse, T., Wardlow, B. And Hayes, M.J. (2009), The Application of Data Mining for Drought Monitoring and Prediction, **Data Mining Applications for Empowering Knowledge Societies**, Edited by Hakikur Rahman, Information Science Reference, New York, pp.280-291
- Silahtaroglu, Gökhan (2008), **Veri Madenciliği**, Papatya Yayınları, İstanbul
- Wang, Chien-Hua and Lee, Wei-Hsuan & Pang, Chin-Tzong, (2010). “Applying Fuzzy FP-Growth To Mine Fuzzy Association Rules”, **World Academy of Science, Engineering And Technology**, 65, pp. 956-962
- Wu, Tong and Li Xiangyang (2003), Data Storage and Management, **The Handbook of Data Mining**, Edited by. Nong Ye, Lawrence Erlbaum Associates Publishers. London, pp.393-407
- Weiss, Sholom M. And Zhang, Tong (2003), Performance Analysis and Evaluation, **The Handbook of Data Mining**, Edited by. Nong Ye, Lawrence Erlbaum Associates Publishers. London, pp.436-439