



**VERİ MADENCİLİĞİ UYGULAMALARI İÇİN
YENİ BİR ARAYÜZ TASARIMI**

Yusuf ÇELİK

Yüksek Lisans Tezi

Yazılım Mühendisliği Anabilim Dalı

Danışman: Yrd. Doç. Dr. Murat KARABATAK

AĞUSTOS-2016

**T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

VERİ MADENCİLİĞİ UYGULAMALARI İÇİN YENİ BİR ARAYÜZ TASARIMI

YÜKSEK LİSANS TEZİ

Yusuf ÇELİK

(121137104)

Yazılım Mühendisliği Anabilim Dalı

Danışman: Yrd. Doç. Dr. Murat KARABATAK

Tezin Enstitüye Veriliş Tarihi : 26 Temmuz 2016

AĞUSTOS-2016

**T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**VERİ MADENCİLİĞİ UYGULAMALARI İÇİN YENİ BİR
ARAYÜZ TASARIMI**

YÜKSEK LİSANS TEZİ

Yusuf ÇELİK

(121137104)

Tezin Enstitüye Veriliş Tarihi : 26 Temmuz 2016

Tezin Savunulduğu Tarih : 09 Ağustos 2016

Tez Danışmanı : Yrd. Doç. Dr. Murat KARABATAK (F.Ü.)

Diğer Jüri Üyeleri : Prof. Dr. Asaf VAROL (F.Ü.)

Doç. Dr. M. Fatih TALU (İ.Ü.)

Ağustos -2016

ÖNSÖZ

Bilgisayar bilimleri konularından veri madenciliği, günümüzün en popüler çalışma alanlarının başında gelmektedir. Hayatımızın her alanında bilişim teknolojilerinin kullanımı artıkça depolanan veri miktarı da artmaktadır. Bu veriler içerisindeki önemli bilgileri keşfetmek için veri madenciliği yaygın olarak kullanılmaktadır. Bu nedenle popülerliği her gün daha da artan ve ders müfredatlarında daha çok yer bulan veri madenciliği dersinin önemi artmaktadır. Bu tez çalışmasında, veri madenciliği dersi için bir eğitsel araç tasarlanmıştır. Bu eğitsel araç, hem öğrencilerin dersi daha iyi kavraması hem de teoride gördükleri algoritmaları veriler üzerinde uygulama imkanı bulması açısından etkilidir. Aynı zamanda profesyonel olarak veri madenciliği alanı ile ilgilenen araştırmacıların da kullanabilecekleri bir yazılım özelliği taşımaktadır.

Tez çalışmam boyunca desteğini hiçbir zaman eksik etmeyen ve aynı mesleği paylaşmaktan gurur duyduğum çok değerli hocam Sayın Yrd. Doç. Dr. Murat KARABATAK'a teşekkürü bir borç bilirim. Tez çalışması süresince desteklerini esirgemeyen çok kıymetli meslektaşlarım ve dostlarım Sayın Öğr. Gör. Vedat TÜMEN, Sayın Okt. Mehmet AŞKIN, Sayın Dr. Özal YILDIRIM, Öğr. Gör. Nuray YILDIRIM ve Sayın Öğr. Gör. Yücel BÜRHAN'a her birine ayrı ayrı katkılarından dolayı ayrıca teşekkür ederim.

Yusuf ÇELİK
ELAZIĞ–2016

İÇİNDEKİLER

Sayfa No

ÖNSÖZ.....	I
İÇİNDEKİLER.....	II
ÖZET	IV
SUMMARY.....	V
ŞEKİLLER LİSTESİ.....	VI
TABLolar LİSTESİ.....	VII
KISALTMALAR LİSTESİ	VIII
1. GİRİŞ.....	1
1.1. Genel Bilgiler.....	1
1.2. Tezin Amacı.....	3
2. İLGİLİ ARAŞTIRMALAR	5
2.1. Bilgisayar Destekli Eğitim.....	5
2.1.1. Bilgisayar Destekli Öğretimin Amaçları	5
2.1.3. BDE'nin Sınırlılıkları	6
2.2. Yazılım	7
2.2.1. Yazılım Türleri.....	8
2.2.2. Yazılım Kalitesi	9
2.2.3. Yazılım Değerlendirme	11
2.2.4. Eğitsel Yazılım.....	12
3. VERİ MADENCİLİĞİ.....	16
3.1. Veri Madenciliği Tanımı	16
3.1.1. Veri Madenciliğinin Uygulama Alanları.....	17
3.1.2. Veri Madenciliğinin Farklı Tanımları	19
3.1.3. Veri Madenciliği Süreci.....	20
3.1.4. Veri Madenciliğinde Kullanılan Yaklaşımlar	21
3.1.5. Kümeleme Metotları.....	24
3.1.6. Bu Çalışmada Kullanılan Sınıflandırma Algoritmaları.....	26
4. EĞİTSEL VERİ MADENCİLİĞİ ARACI	33
4.1. Veri Madenciliği Aracı İçerisindeki Algoritmalar	35
4.1.1. İstatistiksel Sınıflandırma	35
4.1.2. Birliktelik Kuralları	36
4.1.3. Kümeleme Algoritmaları.....	36
4.1.4. Regresyon Ağaçları.....	38
4.1.5. Karar Ağaçları.....	40
4.1.6. Bellek Tabanlı algoritmalar	40
5. GELİŞTİRİLEN YAZILIMININ DEĞERLENDİRİLMESİ	42

5.1.	Evren ve Örneklem.....	42
5.2.	Akademik Başarıların Değerlendirmesi	43
5.3.	Yazılımının Öğrenci Görüşleri ile Değerlendirilmesi	45
6.	SONUÇLAR VE ÖNERİLER	49
	KAYNAKLAR	50



ÖZET

Veri Madenciliği, büyük miktardaki veriler arasından anlamlı verilerin elde edilmesini sağlayan bilgisayar bilimleri alanıdır. Tıp, Bankacılık, Görüntü İşleme, Otomotiv, Sinyal İşleme vb. birçok alanda yaygın olarak kullanılmaktadır. Günümüzde birçok alanda yaygın olarak kullanılan Veri Madenciliği, üniversitelerin bilişim ile ilgili bölümlerinde seçmeli veya zorunlu ders olarak da verilmektedir. Bu derslerin uygulama saatlerinde, mevcut veri madenciliği yazılımları kullanılmakta ancak bu yazılımlar eğitsel açıdan öğrenciye yeterince fayda sağlayamamaktadır. Bu tez çalışmasında, Veri Madenciliği dersinde kullanılmak üzere eğitsel bir veri madenciliği yazılımı geliştirilmeye çalışılmıştır. Böylece öğrencilerin veri madenciliği algoritmalarını daha iyi anlayabilmeleri, ara adımlarını görebilmeleri ve sonuçlarını daha iyi yorumlayabilmeleri hedeflenmiştir. Geliştirilen bu yazılım, Fırat Üniversitesi Teknoloji Fakültesi Yazılım Mühendisliği bölümünde veri madenciliği dersini alan öğrenciler üzerinde test edilmiştir. Bu amaçla öğrenciler kontrol ve deney grubu olarak iki gruba ayrılmıştır. Bu iki öğrenci grubu arasındaki akademik başarı durumlarının gözlemlenmesinin yanı sıra yazılımı kullanan öğrencilerden hazırlanan yazılım hakkında ölçek yolu ile öğrenci görüşleri toplanmıştır. Uygulama sonucunda elde edilen tüm bulgular değerlendirilerek hazırlanan yazılımın eğitimsel olarak katkıları detaylı bir şekilde incelenmiştir. Elde edilen sonuçlardan, geliştirilen veri madenciliği eğitsel yazılımını kullanan öğrenciler ile kullanmayanlar arasında anlamlı farklılıklar olduğu gözlemlenmiştir.

Anahtar Kelimeler: Veri Madenciliği, yazılım, eğitsel yazılım

SUMMARY

A New Interface Design for Data Mining Applications

Data Mining is the field of computational science which helps to obtain meaningful data from a large amount of data. It is widely used in many fields such as Medicine, Banking, Image Processing, Automotive and Signal Processing. Today, Data Mining is given as an elective or compulsory course at the relevant departments of computing at universities. In the practice hours of the course, the available Data Mining softwares is used, but the softwares is not so much beneficial to students from educational perspective. In this study, an educational data mining software has been developed to be used in Data Mining course. Thanks to the software, it is aimed that students will better understand data mining algorithms, and will be able to see the intermediate steps and will better interpret the results. The software was tested on students who took data mining course in the department of Software Engineering at the Technology Faculty of Firat University. For this purpose, students were divided into control and experimental groups. In addition to monitoring the academic achievement of two student groups, the opinion of students who used the software program was collected through scale tests. The educational contributions of the software were examined in detail after the assessment of the all the findings from the application. From the results, it has been observed that there are meaningful differences between the students who used data mining educational software and those who did not.

Keywords: Data Mining, software, educational software

ŞEKİLLER LİSTESİ

	<u>Sayfa No</u>
Şekil 2.1. Yazılım kalitesini etkileyen etmenler	10
Şekil 3.1. Veri Madenciliği Süreci	20
Şekil 3.2. Veri kümeleme	23
Şekil 3.3. KNN algoritması.....	26
Şekil 4.1. EVMA yazılımına ait ana sayfa görüntüsü.	33
Şekil 4.2. EVMA yazılımı içerisinde Bayes algoritmasının kullanımı.	341
Şekil 4.3. EVMA içerisindeki Apriori algoritmasının kullanımına ait bir örnek.	3541
Şekil 4.4. En yakın komşu algoritmasına ait örnek bir gösterim.	41
Şekil 4.5. K-ortalamlar algoritmasına ait örnek bir gösterim.....	41
Şekil 4.6. Twoing algoritmasına ait örnek bir gösterim.	41
Şekil 4.7. Gini Algoritmasına ait örnek bir gösterim.	41
Şekil 4.8. Veri madenciliği aracı içerisindeki Id3 algoritmasına ait bir görüntü.....	41
Şekil 4.9. Bellek tabanlı k-en yakın komşu algoritması.	41
Şekil 5. 1. Sınav sonuçları yazılım dağıtılan ve dağıtılmayan gruplar.....	43

TABLÖLAR LİSTESİ

	<u>Sayfa No</u>
Tablo 3.1. K-NN Algoritma yapısı.....	28
Tablo 5.1. Gruplara göre öğrenci sayıları ile bu öğrencilere ait akademik başarı ortalaması	428
Tablo 5.2. Öğrencileri ara sınav ile genel sınav puanlarının eğitici yazılım programını kullanmalarına göre t-testi sonuçları	43
Tablo 5.3. Hazırlanan anket çalışmasının değerlendirilmesi	45



KISALTMALAR LİSTESİ

BDE	: Bilgisayar Destekli Eğitim
BDÖ	: Bilgisayar Destekli Öğretim
CMM	: Capability Maturity Model
KNIME	: Konstanz Information Miner
K-NN	: K-En Yakın Komşu
SEI	: Software Engineering Institute
VM	: Veri Madenciliği
WEKA	: Waikato Environment for Knowledge Analysis
EVMA	: Eğitsel Veri Madenciliği Aracı

1. GİRİŞ

1.1. Genel Bilgiler

Yakın geçmişe kadar bir ülkenin gelişmişlik düzeyi dendiğinde akla ilk gelen sanayi ve hizmet sektörlerindeki gelişmeler olmuştur. Bu alanlardaki ilerlemelere göre ülkelerin gelişmişlik düzeyleri kıyaslanmaktaydı. Günümüzde ise bilgi daha değerli bir kavram olarak görülmekte ve gelişmişlik kriteri olarak bilgiye erişim hızı ön plana çıkmaktadır. Bilgisayar ve internetin hayatımıza girmesi ile birlikte artık bilgiye erişmek çok daha kolaylaşmış ve bu nedenle bu iki araç hayatımızda önemli bir yer edinmiştir.

Önceki dönemlerde öğretmenler ve ders kitapları klasik öğretim materyalleri olarak eğitimde kullanılan araçlar iken, günümüzde bilgisayar destekli araçlar da eğitimde sıklıkla kullanılan materyaller olmuştur. Bilgisayarın eğitim alanında kullanılması ile birlikte eğitimde verimliliğin artırılmasına yönelik çeşitli bilgisayar yazılımları geliştirilmeye başlanmış ve mevcut yöntemlerin eksikliklerinin giderilmesi amaçlanmıştır.

Bilgisayar günümüzde bir seçenek olmaktan öte tamamlayıcı bir unsur olarak “Bilgisayar Destekli Eğitim (BDE)” adıyla literatürde yerini almıştır. Köksal (1981) BDE’yi “Öğrencinin bilgisayar ile etkileşimi göz önünde tutarak; oluşturulmuş bir ders yazılımı ile karşılıklı etkileşimde bulunarak kendi öğrenme hızına göre kullanabildiği öğretim türü, bu durumla ilgili uygulama ve araştırma alanıdır” şeklinde tanımlar. Yalın’a göre BDE: “Öğrencilere ders içeriği ya da kavramları öğretmek için ilgili konular hakkında bilgisayar programlarıyla tasarlanan uygulamalarının kullanılmasıdır.” (Yalın, 2001 den aktaran Türker vd, 2015).

BDE’nin yaygınlaşması bireysel ve grup çalışmaları için yeni eğitim araçlarına ihtiyaç duyulmasına sebep olmuştur. Bilgisayar yardımı ile daha kaliteli eğitim amacı birleşince eğitim alanında kullanılacak yazılımların geliştirilmesi de kaçınılmaz olmuştur. Eğitim için özel olarak hazırlanan yazılım türüne “Eğitsel Yazılım” adı verilmektedir. Eğitsel yazılımlar doğru kullanıldıkları sürece öğrencilerin derslere olan ilgilerini arttırmakta, çoklu öğrenme ortamı yaratmakta, öğrenmeyi kolaylaştırmakta ve BDE’i öğrenci merkezli kılmaktadır (Çankaya ve Karamete, 2008). Bu nedenle; öğrencinin yeteneklerine, öğrenme ihtiyaçlarına, dersin amacına ve işleyişine uygun yazılımlar tercih edilmeli ve bu yazılımlar öğrencinin derse ilgisini artıracak nitelikte olmalıdır.

En son dizaynları tüm yönleriyle ele almak ve platforma bağlı kalmadan basit bir şekilde dizaynları oluşturmak için şimdiye kadar çok fazla bilgisayar destekli eğitim yazılımı oluşturulmuştur. Bu araçlar; mimarlık, makine mühendisliği, elektrik-elektronik mühendisliği gibi alanlarda hem eğitim hem de profesyonel ürün geliştirme aşamalarında verimli bir şekilde kullanılmaktadır.

Aydoğan (2008) yaptığı çalışmada “Bilişim Teknolojilerinin Temelleri” dersi için bir eğitim programı hazırlamış ve bu programı deney grubuna uygulamıştır. Kontrol grubuna ise dersi klasik yöntemle anlatmıştır. Hazırlanan yazılımın başarımını değerlendirmek için her iki grup üzerinde bir anket çalışması gerçekleştirmiştir. Hazırlanan yazılımın; tasarım, öğretim ve motivasyon açısından başarılı olduğu anket sonuçlarıncı tespit edilmiştir.

Küpçüoğlu (2008), ortaöğretim 10. Sınıf öğrencilerinin “Bilişim Teknolojileri Temelleri” dersi “Programlama Temelleri” modülünü deney grubuna BDE materyali kullanarak anlatmış, kontrol grubuna ise klasik öğrenme yöntemleri ile anlatım yapmıştır. Alınan geri dönüt, deney grubunun son test başarı puanlarının kontrol grubuna göre daha iyi olduğu gözlenmiştir.

Şen (2006) yaptığı çalışmada, meslek liselerinde elektronik devre uygulamaları için eğitsel yazılım kullanmış ve deney grubuna uygulamıştır. Yapılan çalışma, başarı oranı ve kalıcılık açısından değerlendirilmiştir. Elde edilen sonuçlara göre, başarı açısından deney grubu son test sonuçlarında kontrol grubundan belirgin bir şekilde üstün çıkmıştır. Kalıcılık testinde ise iki grup arasında anlamlı bir fark oluşmamıştır.

Eğitim amacı ile geliştirilen yazılım araçları hem dersler için bir öğretim materyali hem de bilimsel çalışmalarda kullanılabilen etkili araçlardır. Özellikle disiplinler arası kullanımı yaygın olan yapay zeka ve veri madenciliği gibi alanlar için geliştirilen eğitsel yazılımlar lisans ve lisans üstü çalışmalarda sıklıkla kullanılmaktadır. Veri madenciliği alanında kullanılmak üzere literatürde geliştirilen ve günümüzde yaygın olarak kullanılan çeşitli yazılımlar aşağıda verilmiştir.

Waikato Environment for Knowledge Analysis veya kısaca WEKA(URL-1, 2016) programı Waikato Üniversitesindeki yapay zeka grubu tarafından geliştirilmiş bir veri madenciliği uygulama yazılımıdır. Java programlama dili kullanılarak geliştirilmiş açık kaynak kodlu bir yazılımdır. WEKA içerisinde birçok makine öğrenmesi algoritmalarını içerir. Bu yazılım ile veriler üzerinde önışleme, birliktelik kuralı ile veri madenciliği, sınıflandırma, kümeleme gibi işlemler kolayca yapılabilir (Witten ve Frank, 2005).

Konstanz Üniversitesi tarafından geliştirilen bir veri madenciliği yazılımı olan Konstanz Information Miner (KNIME) (URL-2, 2016), Eclipse Rich Client Platformu ile geliştirilmiştir. Kullanıcılara kendi uygulama modüllerini geliştirebilmeleri adına yazılım geliştirme kiti sunmaktadır (Bilgin, 2009). Herhangi bir kurulum işlemine ihtiyaç duyulmadan çalıştırılabilir. Programın kabul ettiği dosya uzantıları; “.txt”, “.arff” veya “.table” şeklindedir.

Granada Üniversitesinin İspanya Ulusal Bilim Projeleri Kurumu desteği ile geliştirilmiş bir veri madenciliği yazılımı olan Keel (URL-3, 2016) Java tabanlı bir yazılımdır. Klasik veri madenciliği algoritmalarından ziyade; bulanık sınıflandırıcılar, yapay zeka tabanlı sınıflandırıcılar ve kural tabanlı kümeleme algoritmalarının birçoğunu içerir (Bilgin, 2009).

Orange (URL-4, 2016), Slovenya Ljubljana Üniversitesince geliştirilmiş bir diğer veri madenciliği uygulamasıdır. C++ programlama dili ile geliştirilmiş olan Orange yazılımının ara yüz ve grafikleri Qt3 kütüphanesi ve Python ile geliştirilmiştir (Bilgin, 2009).

1.2. Tezin Amacı

Bu tez çalışmasının genel amacı, Yazılım Mühendisliği, Bilgisayar Mühendisliği gibi bölümlerde verilmekte olan Veri Madenciliği dersinde kullanılmak üzere bir eğitsel yazılım geliştirmektir. Geliştirilecek olan bu yazılım ile veri madenciliği dersini alan öğrencilerin ders içeriğini daha iyi kavraması ve uygulamalı olarak ders içeriğini verimli bir şekilde öğrenebilmeleri hedeflenmektedir. Bu amaç için hem görsel olarak hem de barındırdığı veri madenciliği yöntemleri açısından zengin bir veri madenciliği yazılım aracı geliştirilmesi amaçlanmaktadır. Veri madenciliği dersini alan öğrenciler üzerinde uygulanacak olan yazılıma yönelik çeşitli değerlendirmelerin yapılabilmesi için çalışma kapsamında bir anket uygulaması gerçekleştirilmesi hedeflenmektedir. Bu anket çalışması ile öğrencilerin geliştirilen yazılım için memnuniyetlerinin ölçülmesi ve elde edilen dönütler ışığında ilerleyen zamanlarda yapılacak olan çalışmalar için bir fikir edinilmesi beklenmektedir. Anket çalışmasının yanı sıra, uygulamayı kullanan öğrencilere ait ders notları üzerinde ayrı bir istatistiksel çalışma yapılarak geliştirilen yazılımın etkilerinin tespit edilmesi hedeflenmektedir.

Tez çalışmasının bir diğer amacı ise, genel bir veri madenciliği uygulaması gerçekleştirerek bu alanda çalışma yapan bilim insanlarının kolaylıkla kullanabileceği bir veri madenciliği aracı hazırlamaktır. Geliştirilecek olan yazılımın kolay kullanılabilir olması ve görsel olarak kullanıcıya yol gösterebilecek nitelikte olması hedeflenmektedir. Böylece farklı disiplinlerde çalışma yapan bilim insanlarının da verilerini kolaylıkla değerlendirebilecekleri bir platform oluşturulması amaçlanmaktadır.

1.3. Tezin Organizasyonu ve Katkıları

Tezin giriş bölümünde, teze genel bir bakış açısı kazandırmaya yönelik temel bilgiler verilmiştir. Tezde yer alan diğer bölümlerin organizasyonu aşağıda verilmiştir.

Bölüm 2’de, Bilgisayar Destekli Eğitim konusunda temel bilgiler verilmiştir. Bu bölüm altında, BDE’nin olumlu ve olumsuz yönleri ele alınmıştır. Ayrıca eğitsel yazılım tanımı ve türleri açıklanarak eğitsel yazılımın yararları ilgili önemli bilgiler sunulmuştur.

Bölüm 3’de, Veri Madenciliği konusu detaylı bir şekilde verilmiştir. Veri Madenciliği uygulama alanları ve Veri Madenciliği süreçleri incelenecek veri madenciliği yöntemleri ile ilgili açıklayıcı bilgilere yer verilmiştir.

Bölüm 4’te, tez kapsamında hazırlanan veri madenciliği aracı hakkında detaylı bilgilere yer verilmiştir.

Bölüm 5’te Geliştirilen veri madenciliği aracının öğrenciler üzerindeki uygulama süreçleri anlatılarak elde edilen geri dönüşlere yer verilmiştir. Yazılıma ait memnuniyet düzeyi için elde edilen anket sonuçları ve bu sonuçlara ait değerlendirmeler bu bölüm altında sunulmuştur. Ayrıca yazılımın uygulandığı öğrenci gruplarına ait sınav notları ile ilgili bir değerlendirmeye bu bölüm altında yer verilmiştir.

2. İLGİLİ ARAŞTIRMALAR

2.1. Bilgisayar Destekli Eğitim

Teknolojik gelişmeler günümüzde bilgisayarın her alanda olduğu gibi eğitimde de kullanılmasını zorunlu kılmıştır. Hatta artık bilgisayar bir seçenek olmaktan çıkmış tamamlayıcı bir unsur olarak bilgisayar destekli eğitim adıyla literatürde yerini almıştır. Baki (Baki, 2002 den aktaran Kutluca ve Birgin, 2007) BDE'yi "Etkileşimli bir şekilde öğrencinin hatalarını eksikliklerini görmesini ve performansını tanımlamasını, buna yönelik dönütler alması; grafik, ses gibi multimedya araçları yardımıyla konu ya da derse ilgili olmasını sağlamak amacıyla bilgisayardan yararlanma biçimidir." şeklinde tanımlar. Bir diğer tanımda BDE: "Öğretme amacıyla kullanılan bir bilgisayar yardımıyla, öğretim sürecini ve öğrenci motivasyonunu güçlendiren, öğrencinin kendi durumuna istifade edebileceği, bilgisayar yardımıyla kendi kendine öğrenme sürecidir." (Şahin ve Yıldırım, 1999 dan aktaran Karamustafaoğlu, 2006).

2.1.1. Bilgisayar Destekli Öğretimin Amaçları

Bilgisayar Destekli öğretim (BDÖ) yönteminde bilgisayar kullanımının temel amacı, "materyalleri ya da bilgiyi en iyi şekilde kullanmada öğrenciye öğretim sürecine yardım edip, kalıcı öğrenmeyi sağlamaktır" (Uşun, 2000 den aktaran Yiğit ve Akdeniz, 2003).

BDÖ' nün öğrenciler için hedeflenen genel amaçlarını şu şekilde sıralayabiliriz;

- Öğrencinin derslere ilgisini arttırmak,
- Öğrencinin bilimsel düşünme yeteneğini geliştirmek,
- Öğrencinin hem bireysel hem grup çalışmalarını desteklemek,
- Öğretme yöntemlerini arttırmak,
- Problemlere farklı açılardan bakmayı sağlamak,
- Geleneksel öğretim yöntemlerini daha etkili hale getirmek,
- Öğrenme sürecini hızlandırmak,
- Öğrenmeye yardımcı materyal sayısını arttırmak,
- Öğretimi daha ucuz ve etkili hale getirmek,
- Gereksinmeye dayalı öğretimi gerçekleştirmek,
- Telafi edici öğretimi sağlamak,

- Öğretimde sürekli olarak niteliğin artmasını sağlamaktır.

Uşun (Uşun, 2000 den aktaran Yiğit ve Akdeniz, 2003), bu amaçlar göz önünde bulundurulduğunda BDE merkezinde öğrenci olduğunu belirtir.

2.1.2. BDE'nin Yararları

Öğrenme Hızı: BDE sayesinde öğrenciler, öğrenme hızlarına paralel olarak dersleri işlerler. Öğrencinin derse aktif bir şekilde katılımını sağlar. Eğitsel yazılım öğrenciye sorular sorar ve öğrencinin verdiği cevaplar doğrultusunda işlenmesine gerek duyulmayan konuları atlar, anlaşılmayan konuları ise tekrar verir. Bu sayede öğrenci hakkında geri dönüt sağlanır.

Zaman ve Ortamdan Bağımsızlık: BDE ile öğrenim gören bir öğrenci ders saati dışında da öğretim etkinliklerini uygulayabilir, konuyu tekrar edebilir, evde veya herhangi bir ortamda bilgisayar veya tablet yardımı ile uygulama şansı bulabilir.

Öğrenci Performansının İzlenebilmesi: Klasik öğrenme ortamlarında öğrenci performanslarının belirlenmesi oldukça güçtür. Dolayısı ile öğrenciyi yönlendirmek ve etkin bir öğretim gerçekleştirmek de zordur. BDE sayesinde bir öğrencinin sorulara verdiği cevaplar ve konuyu öğrenmek için harcadığı zaman bilgisayar tarafından kayıt edilebilir. Bu sayede öğrenci performansı belirlenebilir ve öğretmen dersi buna göre şekillendirebilir (Varol, 1997).

2.1.3. BDE'nin Sınırlılıkları

Öğrencilerin Sosyo-Psikolojik Gelişimlerinin Engellenmesi: Eğitsel yazılımların dikkat çekici özellikte hazırlanması ve özellikle öğretimsel oyunlar sebebiyle öğrencilerin bilgisayar başında uzun zaman geçirmesi nedeniyle öğrenciler diğer bireylerle iletişim kurmakta zorlanmaktadır. Bu durum sebebiyle de öğrencinin sosyo-psikolojik gelişimi olumsuz etkilenmektedir. Bu durumun önüne geçilebilmesi için bilgisayar destekli eğitimin yanı sıra öğrencilerin birbiri ile ve öğretmenler ile doğru iletişim sağlayacak öğretimsel faaliyetlerin planlanması ve uygulanması gerekmektedir.

Özel Donanım ve Beceri Gerektirmesi: BDE için kullanılacak eğitsel yazılımı kullanmak için bilgisayar ile birlikte özel donanımlara da ihtiyaç duyulabilir. İhtiyaç duyulan donanımların maliyetleri yüksek erişimleri zor olabilir. Bu durumda eğitimler için maliyet-fayda analizleri gerçekleştirilmeli ve eldeki kaynaklar doğru şekilde kullanılmalıdır.

Eğitim Programını Desteklememesi: Öğretim amacı ile kullanılan her materyalin öğretim amacına uygun olarak hazırlanması ve eğitim programını desteklemesi gerekmektedir. Eğitsel yazılım hazırlayıcılar tarafından gerçekleştirilen yazılımların birçoğunun bu amaca uygun olarak hazırlanmamıştır. Bu da BDE için en büyük sorunlardan biridir. Bu durumda eğitsel yazılım seçilir veya tasarlanırken kalite standartlarına uygun olmasına dikkat etmek gerekmektedir.

Öğretimsel Niteliğin Zayıf Olması: Programların eğitim amacına uygunluğunun yanında öğrenciye ilgili konuları aktif bir şekilde aktarabilmeli, öğrencinin derse aktif bir şekilde katılımını sağlamalıdır. Ama eğitsel yazılım olarak hazırlanan yazılımların çoğu bu nitelikten yoksundur hatta yazılı materyallerin elektronik ortama aktarılmış şekilleridir. Bu da BDE için önemli bir kısıttır (Odabaşı, 1997).

2.2. Yazılım

Yazılım hem ürün hem de ürünü oluşturmak için kullanılan materyaldir. Bu durumda yazılımın iki farklı rolü olduğu söylenebilir. Ürün olarak yazılım, bilgisayar donanımını kullanmaya yarayan materyaldir. Araç olarak baktığımızda ise tüm teknolojik cihazları kontrol etmemize yarayan programların oluşturulduğu bir ortam olarak tanımlanabilir. Yazılımın donanımdan farkları şu şekilde ifade edilebilir (Pressman, 2001):

- Yazılımlar geliştirilmiş veya inşa edilmişlerdir
- Yazılım yıpranmaz tükenmez
- Donanım birimleri montaj ile birleştirilen birimler olarak üretilse de yazılımlar özel olarak üretilmektedir.

2.2.1. Yazılım Türleri

Yazılımlar için genelde teknik açıdan sınıflandırmalar yapılmaktadır. Konu tabanlı sınıflandırma yapılması çok güç olduğu için eğitsel yazılımlar bu sınıflandırmanın dışında kalmaktadır. Yazılımları gördükleri işlevlere göre aşağıdaki şekilde sınıflara ayırabiliriz;

Sistem Yazılımları

Sistem yazılımları, donanım ile kullanıcılar arasındaki bağlantıyı sağlamaya yarayan yazılımlardır. İşletim sistemleri gibi çok temel, karmaşık ve büyük olabildikleri gibi editör ve derleyiciler gibi de daha küçük ve kısıtlı hizmetler için kullanılanları da mevcuttur.

Eş Zamanlı Yazılımlar

Gerçek dünya problemleri için hazırlanmış yazılımlardır. Olayları görüntüleme, analiz etme ve kontrol etme gibi işlemleri gerçekleştirirler. Gerçek zamanlı çalıştıkları için çok kısa sürede işlemleri gerçekleştirmeleri gerekmektedir.

İş Yazılımları

Büyük veri depolama gereksinimine sahip yazılım türleridir. Çalışan bilgileri, envanter hesapları gibi büyük boyutta bilgileri depolamanın yanı sıra büyük işlemleri organize etmeleri de uygulama alanlarını oldukça genişletmiştir.

Mühendislik ve Bilim Yazılımları

Bu tür yazılımlar cebirsel olarak yoğun ve karmaşık işlemlere dayalıdır. Jeoloji, biyoloji hatta astronomiye kadar geniş bir kullanım alanına sahiptir. Teknolojik gelişmeler ile birlikte bu yazılımlar geliştirmek için sayısal algoritmalar yerine kullanım alanına bağlı olarak daha özel ve interaktif yöntemler kullanılmaktadır.

Gömülü Yazılımlar

Belirli donanım birimlerinden oluşan sistemler için özel olarak tasarlanmış, sistemin kontrolünü sağlayan kullanımını kolaylaştıran; sistemlerin salt okunur belleğinde bulunan yazılımlardır. Cep telefonu, çamaşır makinesi, buzdolabı yazılımları örnek olarak verilebilir.

Kişisel Bilgisayar Yazılımları

Hesaplama araçları, kelime işleme, eğlence, veri tabanı yönetimi, grafik oluşturma gibi yazılımlar kişisel bilgisayar yazılımları olarak nitelendirilmektedir.

Web Tabanlı Yazılımlar

Sunucu ile istemci arasında, istemciye herhangi bir dosya veya uygulama yüklemesi yapmadan tarayıcı üzerinden çalışan yazılımlar uygulama yazılımları olarak adlandırılır.

Yapay Zeka Yazılımları

Basit matematiksel denklemler yardımı ile çözülemeyen, büyük ve karmaşık problemleri çözmek için oluşturulmuş yazılımlardır. Ürün tasarımı, medikal, tedavi, örüntü tanıma gibi alanlarda kullanılan bu yazılımlar uzman sistemler olarak da adlandırılır.

2.2.2. Yazılım Kalitesi

Yazılım geliştiricilerinin gittikçe arttığı ve belirli bir amaç için bile birden çok yazılımın üretildiği günümüzde, yazılım kalitesi önemli bir etken haline gelmiştir. Bir yazılımın kalitesi; kaynak, zaman ve maliyet olmak üzere üç kritere bağlı olarak belirlenir (Erdemir vd, 2011). Crosby kaliteyi ihtiyaç karşılama oranı olarak tanımlar (Crosby, 1979 tarafından aktarılan Mardanivd, 2013). Geliştirici açısından kalite, kolay test edilebilen, bakımı rahat yapılan, maliyet ve zaman hesabı optimuma yakın yazılımlardır. Yazılım kalitesini etkileyen faktörler Şekil 2.1’de gösterildiği gibi süreç, çevre, işgücü ve teknoloji olarak sıralanmaktadır (McConnell, 1996).



Şekil 2. 1. Yazılım kalitesini etkileyen etmenler

Yazılım kalitesinden bahsedilebilmesi için yazılımın iki farklı gözle incelemesi gerekmektedir. Bunlardan biri yazılımı geliştiren diğeri ise kullanan gözüdür. Kullanan için kaliteli yazılım istenenleri yerine getirebilen, kolay kullanılan ve hata vermeyen yazılımdır. Geliştiren ise maliyeti, yazılımın uyumluluğu ve kullanım alanının genişletilebilmesi konularını göz önünde bulundurmaktadır.

ISO/IEC 9126-1 modeli yazılım ürünlerinin kalitesini tanımlayan ve sınıflandıran bir modeldir. Bu modele göre yazılım kalitesinin sınıfları aşağıdaki şekildedir (Erdemir vd, 2011):

İşlevsellik: Yazılımın gereksinim duyulan ihtiyaçları karşılayabiliyor olabilmesidir.

Güvenilirlik: Hataya karşı duyarlı geri yükleme süreçleri veri kurtarma prosedürlerinin olmasıdır.

Kullanılabilirlik: Kullanıcı etkileşiminin yüksek olması, kolay öğrenilmesi gibi kriterleri taşımasıdır.

Verimlilik: İhtiyaç duyulan araçlardan daha fazla araç istemeyecek şekilde performanslı bir şekilde çalışmasıdır.

Bakılabilirlik: Yazılımın kalitesi geri dönük hata eksik ya da güncellemelerin yapıldığında maliyeti yükseltmeyecek şekilde olmasıdır. Bakılabilirliği kötü bir yazılımın maliyeti de yüksek olacaktır.

Taşınabilirlik: Yazılımın platforma bağımlı olmadan bir çok platformda çalışıyor olabilmesidir.

2.2.3. Yazılım Değerlendirme

Yazılım ihtiyacının artması ve buna bağlı olarak piyasadaki yazılımların artması ile birlikte kaliteli yazılıma duyulan ihtiyaç da artmıştır. Bu nedenle, yazılımın doğru değerlendirilmesi gerekliliği ortaya çıkmıştır (Kurtel ve Eren, 2008). Yazılım değerlendirmesi yazılımcılar tarafından gerçekleştirilmelidir. Değerlendirme işlemi, kaynak kodlar ile tasarım özelliklerinin uyumu şeklinde tanımlanabilir.

Kurtel ve Eren yazılım değerlendirme sayesinde yazılım kalitesinin artacağını düşünür ve yazılım değerlendirmenin gerekliliğini şöyle anlatır (Kurtel ve Eren, 2008):

- Yazılım değerlendirme yazılım kalitesinin artmasına yardımcı olur,
- Değerlendirme sonucunda yazılımda yapılacak değiştirmelerde kalite kriterlerinin baz alınmasını sağlar,
- Karmaşıklığı azaltır,
- Hataların minimize edilmesini sağlar,
- Yazılım projeleri ve bu projelerden elde edilen ürünlerin anlaşılmasına yardımcı olur,
- Yazılım projesi sürecinde; sürecin izlenmesi, analiz edilmesi, değerlendirilmesi ve kontrolünü sağlar.

ISO/IEC 15939 Sistem ve Yazılım Mühendisliği Ölçüm Süreci standardı ve ISO 9000 kalite yönetim sistemi, yazılım değerlendirme için temel endüstriyel standartlardır. ISO 9000 genel anlamda müşterilerin ihtiyaçlarına ve beklentilerine yönelik bir hizmetin gerçekleştirilmesi sırasında tüm süreci kapsayan ve sürekli gelişmeyi amaçlayan ve bu aşamaları değerlendiren bir sistemdir.

18 ISO 15504 (SPICE) standardı, yazılım geliştirme süreçlerinin analizi gerçekleştirmek için birçok bağımsız metodolojilerin alt yapısını oluşturduğu bir standart şeklinde karşımıza çıkmaktadır (Kurtel ve Eren, 2008). Software Engineering Institute (SEI)'nin geliştirmiş olduğu Yetenek Olgunluk Modeli (Capability Maturity Model-CMM) de yazılım süreçlerin olgunluğunun değerlendirilmesinde kullanılmaktadır.

İnsan etkinliğinin olduğu her alanda giderek daha da büyük öneme sahip olan yazılım değerlendirme, CMM ya da ISO 9000 standartları altında ele alınsa da bu önlemlerin yetersiz kalabileceğini belirten McDaniel (2002) yalnızca yazılım değerlendirmeye yönelik kalite yönetim uygulamaları çatısı altında oluşturulan standartlara ihtiyaç vardır.

2.2.4. Eğitsel Yazılım

Teknolojik gelişmelerle birlikte bilgisayarın her alanda olduğu gibi eğitim alanında da yer edinmesi ve BDE yönteminin yaygınlaşması ile birlikte ülkemizde de bu alanda çalışmalar yapılmaya başlanmıştır. Bilgisayar, derslerde daha etkin kullanılır olmuş, derslere yardımcı olması için yazılımlar geliştirilmeye başlanmıştır. Eğitim amacı için özel olarak üretilen yazılımlara “eğitsel yazılım” adı verilmektedir. Eğitsel yazılım kullanımı her geçen gün artmaktadır.

2.2.4.1. Eğitsel Yazılımların Yararları

Öğrencinin ihtiyaçlarına karşılık verebilen yazılımlar öğretmenin dersi işleyiş biçimini ve dersteki rolünü belirlemesine yardımcı olmaktadır. Kaliteli yazılımlar, öğretmenin derste işini oldukça kolaylaştırır. Bazı yazılımlar öğrenmeye yardımcı olacak tüm materyalleri içerdiğinden başka bir öğrenim aracına ihtiyaç duyulmayabilir (Akkoyunlu, 1996). Yazılım kullanımında amaç öğrencinin öğrenme sürecini kısaltmak ve öğrenme işlemi kolaylaştırmaktır. Hatalı, kontrolsüz veya aşırı kullanım tam tersine zararlı olabilmektedir (Bayram vd, 2007 den aktarılan Kelleci ve Şahinkayas, 2012).

Bayram (Bayram vd, 2007 den aktaran Kelleci ve Şahinkayas, 2012), öğretim yazılımlarının, uzmanlar tarafından da kabul edilmiş katkılarını aşağıdaki gibi sıralamaktadır:

- Bireysel veya grup çalışmaları için mevcut akışın eksiklerini kapatır ve gereksinimleri karşılar.
- Öğrenci merkezli eğitim olanağı sağlar.
- Öğretimi somutlaştırır.
- Tekrar imkânı sağlar.
- Öğrenme hızını artırır, öğrenme süresini kısaltır.
- Derse katılımı artırır, öğrencinin derste daha etkin olmasını sağlar.

- Çoklu öğrenme ortamı sağlar.
- Uygulama imkânı ile öğrenmeyi kalıcılaştırır.
- Bireysel becerileri göz önünde bulundurarak öğrenmeyi basit hale getirir.

2.2.4.2. Eğitsel Yazılım Türleri

Eğitsel yazılımlar değerlendirme testlerinden başarılı bir şekilde geçecek ve derslerde aktif olarak kullanılacak şekilde hazırlansa da bir yazılımın tüm gereksinimleri karşılaması neredeyse imkansızdır. Her eğitsel yazılım belirli bir amacı gerçekleştirmek için üretilir. Yazılım; ihtiyaçları karşılamalı, öğrencilerin becerilerini açığa çıkarmalı, öğrenme ortamına kolay adapte edilmelidir. Eğitsel yazılımların hazırlanmasında kullanılan teknikler (Alessi ve Trollip, 2001):

- Öğretici Yazılımlar (Tutorials)
- Alıştırma-Uygulama Yazılımları (Drill&Practice)
- Benzeşim Programları (Simulations)
- Eğitsel Oyunlar (Educational Games)
- Hipermedya (Hypermedia)

Öğretici Yazılımlar

Öğretici yazılımlar veya bire-bir eğitim yazılımları, öğretmenin rolünü yerine getirmeye çalışan yazılımlardır (Demir, 2004). Bu yazılımlar öğretmenin yaptığı; yeni bilgiler vermek, geri dönüt vermek, adaptasyonu sağlamak ve gerekli alıştırmaları yaptırmak gibi işlevleri sağlayabilmelidir. Kelleci'ye göre öğretici yazılımlar öğrenci başarılarını artırmada çok etkilidir (Kelleci, 2010). Öğretici yazılımlar iki farklı biçimde hazırlanır. Bunlar, “Doğrusal Öğretici Yazılımlar” ve “Dallara Ayrılmış Öğretici Yazılımlar” olarak sınıflandırmaktadır.

a. Doğrusal Öğretici Yazılımlar

Doğrusal öğretici yazılımlarda yer alan yazılım, bir başlık ya da kavram ile başlayıp bilgiyi sunar ve sorduğu sorular ile doğrusal bir yol izleyerek devam eder. Bu süreç tüm öğrenciler için aynıdır. Sorulan sorulara dahi aynı sıra ile cevap verilir. İlgi çekici

olmaması, öğrenci seviyesini ve kapasitesini gözetmeyişi, bilgisayarların aktif şekilde kullanılmaması sebebiyle bu yazılımlar pek kullanışlı bulunulmamaktadır (Kelleci, 2010).

b. Dallara Ayrılmış Öğretici Yazılımlar

Dallara ayrılmış öğretici yazılımlar, öğrenci merkezli olarak öğrencilerin seviyelerine ve öğrenme hızlarına göre hazırlandıklarından doğrusal öğretici yazılımlara göre daha kullanışlıdır. Geri dönüt ve ihtiyaç halinde ek bilgiler verebilen ayrıca sorulan sorulara verilen doğru cevaplar doğrultusunda gerek duyulmayan konuların atlanabildiği öğretici yazılım türüdür (Kelleci, 2010).

Alıştırma Uygulama Yazılımları

Uygulama, öğretim aşamasını kapsayan Alıştırma Uygulama Yazılımları yeni bilgiler de aktarabilen eğitsel yazılım türüdür. Alıştırma Uygulama yazılımları, öğretici yazılımların tamamlayıcısı niteliktedir (Kelleci, 2010). Alıştırma uygulama yazılımları basılı materyal olan yardımcı kaynaklar, çalışma sayfaları veya ders kitapları kullanılarak kolayca hazırlanabilir ve bu kaynaklardan daha etkili olabilirler. Bu yazılımlar eğitmek için tasarlansa da belirli bilgi birikimlerine sahip öğrencilerin seviyelerini geliştirme konusunda daha etkilidir (Öztürk ve İnan, 1998 den aktarılan Gül ve Yeşilyurt, 2011).

Benzetim Programları

Benzetim programlarının amacı, kullanıcılara gerçek hayatta gerçekleşmesi zor veya imkansız olan olayları daha güvenli bir şekilde sunmak ve deneyimlemelerini sağlamaktır (Demir, 2004). Uygulamalar kullanıcı deneyimleri ile ne kadar desteklenebilir ve gerçeğe ne kadar uygun gerçekleştirilebilirse öğrenme o kadar kalıcı olur (Akpınar, 1999).

Eğitsel Oyunlar

Eğitsel oyunlar, kullanıcıların geçmişte edindikleri bilgileri daha rahat bir ortamda tekrar etmelerine olanak sağlayan ve bunları pekiştirmelerini kolaylaştıran eğitsel yazılım türüdür (Güneş, 2009). Eğitsel oyunlar kullanıcıların; olgular, kavramlar, yöntemsel bilgiler, problem çözme becerileri, iletişim becerileri ve karar verme süreçleri gibi bazı psikomotor becerilerini de geliştirmelerine yardımcı olabilirler (Akpınar, 1999).

Benzetim programlarına çok benzer yönü bulunan eğitsel oyunlar genellikle simülasyon uygulamaları ile karıştırılmaktadır. Simülasyonlar öğrenmeyi, gerçek bir

olguyu taklit ederek ve bazen de eğlendirerek gerçekleştirmeye çabalar. Bir eğitsel oyun gerçeği modelleyebilir ancak her zaman modelleyeceği söylenemez (Öztürk ve İnan, 1998 den aktarılan Gül ve Yeşilyurt, 2011).

Hipermedya

Hipermedya eğitsel yazılımları, internet ortamındaki bilgilerin yönetilmesi için kullanılan ortak bir yöntem olarak adlandırılabilir. Hipermedya, öğrenmeyi yönlendirebilmek adına çeşitli erişim yöntemlerinin ortak bir veri tabanı yardımıyla kullanılması ile oluşur. Hipermedya yazılımının temel amacı öğrenmeyi kalıcı hale getirmektir. Hipermedya uygulamaları ile desteklenen öğrenme stratejileri; ileri düzeyde anlama, öğrencinin yönlendirilmesi, arama, erişim ve uygulama gibi kavramlardır (Alessi ve Trollip, 2001).

3. VERİ MADENCİLİĞİ

3.1. Veri Madenciliği Tanımı

Veritabanı ve bilişim teknolojileri 1960’lardan bu yana veri dosyaları sistemlerinden büyük veritabanı sistemlerine kadar sistematik bir şekilde gelişmektedir. Bu gelişme 1970’lerde ilişkisel veritabanı sistemlerini ortaya çıkardı. 1980’lerin ikinci yarısından itibaren ise de nesne tabanlı, geliştirilmiş-ilşkisel veri tabanları gibi gelişmiş veritabanı sistemlerinin ortaya çıkmasına neden oldu. Veri tabanı sistemlerindeki gelişmeler doğrultusunda 1980’lerin sonlarına doğru veri ambarları ve veri madenciliği gibi kavramların oluşmasını sağladı (Han vd, 2000).

Veri madenciliği, büyük miktarlardaki veriler içerisinde geleceğe yönelik tahminlerin gerçekleştirilmesine yardımcı olacak anlamlı ve kullanışlı bağlantı ve kuralların bilgisayar programları aracılığıyla aranması ve analizidir (Akpınar, 2000). Literatürde veri madenciliğiyle aynı anlama gelen ya da yakın bir anlam içeren farklı tanımlarla da karşılaşılabilir. Bunlar veri tabanlarında bilgi madenciliği, bilgi çıkarımı, veri/örüntü analizi, veri arkeolojisi ve veri taranmasıdır (Mitra ve Acharya, 2005). Ayrıca Veri Madenciliği, çok büyük miktardaki verilerin içindeki ilksileri inceleyerek aralarındaki bağlantıyı bulmaya yardımcı olan veri analizi tekniğidir (Akpınar, 2000).

Veri madenciliği; konunun uzmanlarından, gözlem ve deneyler sonucunda elde edilen verilerden üstü kapalı, önceden bilinmeyen ancak potansiyel olarak kullanışlı bilginin genelleme yapacak şekilde çıkarılmasıdır (Akgöbel ve Çakır, 2009).

Veri madenciliği yardımıyla, büyük veri kümelerinden oluşan veritabanı sistemleri içerisinde gizli kalmış bilgilerin elde edilmesi sağlanır. Bu işlem, istatistik, matematik disiplinleri, modelleme teknikleri, veritabanı teknolojisi ve çeşitli bilgisayar programları kullanılarak gerçekleştirilir (Akpınar, 2000).

Veri madenciliği, kullanıcıya iş süreçlerinde meydana gelen veriler arasındaki şablonları ve ilişkileri bulması konusunda yardım etmektedir. Veri madenciliği hayatımızın hemen hemen tüm alanlarında kullanılmaktadır. Bankacılık, pazarlama, sigortacılık, perakendecilik, basın-yayın telekomünikasyon, endüstri, bilim ve mühendislik alanlarında

yoğun olarak kullanılmaktadır. Veri madenciliği tüm bu sektörlerde bilgiye ve sonuca ulaşımı daha hızlı hale getirmektedir.

Veri madenciliği Elektrik – Elektronik Mühendisliğin çalışma alanlarında sinyalizasyon işlemede, güç kalitesi analizinde, harmonik hareketlerin tespiti ve incelenmesinde kullanılmaktadır (Yıldırım, 2010).

3.1.1. Veri Madenciliğinin Uygulama Alanları

Veri madenciliği teknolojinin bulunduğu hemen hemen tüm alanlarda uygulanmaktadır. Bilgisayar Bilimleri, Elektrik Elektronik mühendisliği, Endüstri, Telekomünikasyon alanları ver analizinde, Sağlık alanları olan; Biyotıp, Gen Mühendisliği ve DNA sıralama desenlerinin veri analizleri gibi birçok alanda, hastalık tansında, finans analizi, astronomi ve birçok alanda uygulanmaktadır.

Bununla birlikte günümüzde veri madenciliği teknikleri özellikle işletme gibi çeşitli alanlarda başarılı bir şekilde kullanılmaktadır. Bu uygulamalardan temel olanları ilgi alanlarına göre aşağıda özetlenmiştir (Dempstervd, 1977).

- Pazarlama

- Müşteri ürün arasında ilişki kurma,
- Müşteri ve demografik özellikleri arasındaki ilişkilerin kurulması,
- Kampanyalara cevap verme oranının artırılması,
- Yeni müşteri çekimi, mevcut müşterileri koruma
- Satılacak ürünleri tespit etme,
- Müşteri değerlendirme,
- Satış öngörüsü.

- Bankacılık ve Sigorta

- Tespit edilemeyen finansal denklemlerin bulunması,
- Dolandırıcılığın önüne geçme,
- Müşteri alışveriş bağlantıları,
- Kredi ihtiyaçları isteklerinin değerlendirilmesi,
- Yeni poliçe almak isteyen müşterilerin belirlenmesi,
- Sigortaya risk oluşturabilecek müşteri profili tespiti.

- ***Biyoloji, Tıp ve Genetik***

- Bitki türleri ıslahı,
- Gen haritasının incelenmesi ve hastalık tespiti,
- Farklı test sonuçlarının tahmini, tıbbi teşhis,
- Tedavi sürecinin belirlenmesi,
- Kanseri hücrelerin tespiti,
- Yeni virüs türlerinin keşfi ve sınıflandırılması,

- ***Kimya***

- Yeni kimyasal moleküllerin keşfi ve sınıflandırılması,
- İlaç geliştirilmesi ve keşfinde kullanılmaktadır.

- ***Yüzey Analizi ve Coğrafi Bilgi Sistemleri***

- Coğrafi özelliklere göre bölgeleri gruplandırma,
- Kentler için yerleşim yerleri belirleme,
- Kentlerdeki suç, fakir-zengin, ırk belirleme,
- Kentlere yerleştirilecek otomasyonların yer tespitinde kullanılmaktadır.

- ***Görüntü Tanıma ve Robot Görüş Sistemleri***

- Çeşitli algılayıcılar yolu ile elde edilmiş görüntülerden yararlanarak tanımlama yapabilecek tekniklerde kullanılmaktadır.

- ***Uzay Bilimleri ve Teknolojisi***

- Evrenimizdeki cisimlerin şekillerin tespiti
- Yeni galaksilerin keşfedilmesi,
- Yıldızların bulundukları yerlere göre sınıflandırılmasında kullanılmaktadır.

Veri madenciliği gruplama, veri özetleri oluşturma, değişikliklerin incelenmesi, sapmaların bulunması gibi bazı teknik yaklaşımları içerir. Başka bir deyişle, veri madenciliği, verilerin içerisindeki desenlerin, ilişkilerin, değişimlerin, düzensizliklerin, kuralların ve istatistiksel olarak önemli olan yapıların yarı otomatik olarak hızlı bir şekilde bulunmasıdır.

Veri Madenciliği uygulamalarını geliřtirmek için programlara gerek vardır. Bu kapsamda, SPSS Clementine, Excel, SAS, Angoss, KXEN, SQL Server, MATLAB ticari ve Rapid Miner (YALE), WEKA, R, C4.5, Orange, KNIME açık kaynak olmak üzere birçok program geliřtirilmiřtir (Yeřilova vd, 2004).

3.1.2. Veri Madenciliğinin Farklı Tanımları

Veri Madenciliği ile ilgili literatür taramasında çok sayıda bilgi yayın ve tezler bulunmaktadır. Genelde kiřilerin kullanım alanlarına göre veri madenciliğini tanımlamıřlardır. Ařağıda bunlardan birkaçı görönmektedir.

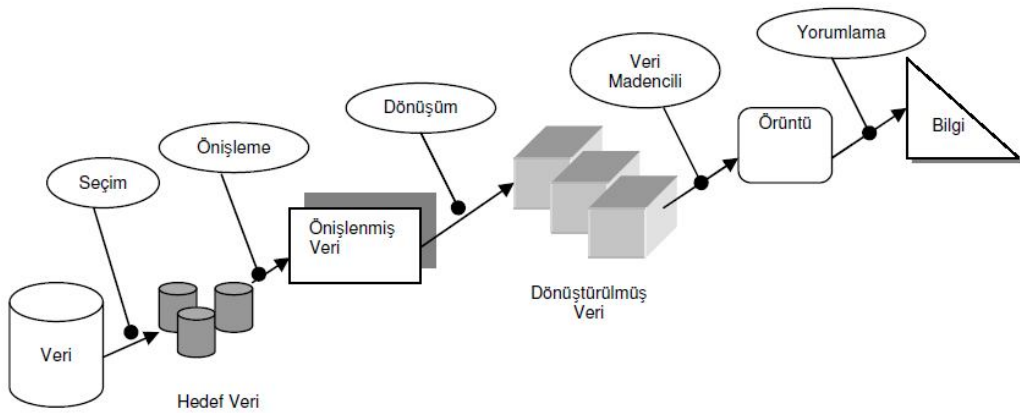
- Bazı veri madenciliğinin teknikleri tümevarım mantığı (inductivelogic) programlamasını, makineyi öğrenmeyi v.b. kapsar. VM sınıflandırmayı, birliktelik ve verilerin arasındaki sıralamayı yapmak için kurallar bulmayı içerir. Temel olarak, birisi gözlenen örneklerden ve desenlerden çıkarılan bilgi olan hipotezlere varır. Bu desenler verilen sorgulardan elde edilir ve her sorgu önce verilen sorguya bağıdır (Thuraisingham, 2003).
- Gartner Grup tarafından yapılan tanımda ise VM, istatistik ve matematik tekniklerle birlikte ilişki tanıma teknolojilerini kullanarak, depolama ortamlarında saklanmış bulunan veri yığınlarının elenmesi ile anlamlı yeni ilişki ve eğilimlerin keřfedilmesi sürecidir (Thuraisingham, 2003).
- Veri madenciliği, temel olarak bilgisayar destekli bir bilgi çözümleme işlemidir Veri madenciliği, ayrı sorgular vererek büyük miktarda olan veriden yararlı bilgi, desenler ve eğilimler (genelde önceden belli olmayan) çıkarabilmektir (Kalıkov, 2006).
- Veri madenciliği, verinin sahibine anlamlı ve yararlı olacak şekilde veri kümesinin içinde şüphe uyandırmayan ilişkileri bulmak ve veriyi yeni bir şekilde özetlemek için veri kümelerinin incelenmesidir (Kalıkov, 2006).
- Veri madenciliğini, önceden bilinmeyen, tam ve yararlı bilginin veriden çekilmesi olarak tanımlamış; veri madenciliğinin insanların kolayca anlayabileceği bir formda veriyi keřfeden ve sunan makine öğrenme, istatistik ve görselleřtirme teknikleri gibi teknikleri kullandığını belirtmiřtir (Savran, 2009).
- Veri madenciliği çok büyük veri yığınlarından kritik bilgilerin elde edilmesini ve bu bilgilerin ışığında stratejik kararlar alınmasını sağlar (Iřık, 2006).

- Veri madenciliği bir veya daha fazla makine öğrenme tekniğinin uygulanarak otomatik olarak bir veritabanı içinde bulunan verilerden bilgi çıkartılması, verilerin analiz edilmesi işlemidir (Han ve Kamber, 2000).
- Veri madenciliği, geniş veritabanlarından prediktif bilginin otomatik olarak çekilme sürecidir. Gelecek trendlerini tahmin eder ve uzmanların gözden kaçırdığı davranışları bulur (Han ve Kamber, 2000).
- Veri madenciliğinde amaç, büyük veritabanları içinde gizlenmiş anlamlı örüntülerin ve birlikteliklerin keşfedilerek mantıksal kurallara ya da görsel sunumlara dönüştürülmesidir (Işık, 2006).
- Veri madenciliğinin büyük hacimli verilerdeki örüntüleri araştıran matematiksel algoritmaları kullandığını söylemiştir. Davis' e göre veri madenciliği hipotezleri mesfeder, sonuçları birleştirmek için insan yeteneklerini de kullanır. Davis'e göre veri madenciliği sadece bir bilim değil, aynı zamanda bir sanattır (Han ve Kamber, 2000).

Sonuç olarak veri madenciliği, önceden bilinmeyen ilişki bulunması için bugünün endüstrisinde yaratılan büyük miktarlardaki veriyi analiz eden bir yoldur. Yüksek güçlü bilgisayarlara ve gereken yazılımlara kolay ve düşük fiyatlarla ulaşılabilmesi bu teknolojinin islemesini olanaklı kılmıştır (Han ve Kamber, 2000).

3.1.3. Veri Madenciliği Süreci

Veri madenciliği; seçim, ön işleme, dönüşüm, veri madenciliği algoritmaları, örüntü ve yorumlama gibi süreçlerden oluşmaktadır. Şekil 3.1'de veri madenciliği süreçlerine ait bir gösterim yapılmıştır.



Şekil 3. 1. Veri Madenciliği Süreci

Alternatif olarak Veri Madenciliği aslında bilgi keşfi sürecinin bir parçası şeklinde kabul görmektedir. Bilgi keşfi sürecinin aşamaları aşağıda verilmiştir (Dener vd, 2009).

- **Veri Temizleme:** Seçilen örneklem kümesinde yer alan hatalı tutanakların çıkarıldığı ve eksik nitelik değerlerinin nasıl ele alınacağı belirlenir.
- **Veri Bütünleştirme:** Seçilen örneklemden ilgisiz niteliklerin atıldığı ve tekrarlı kayıtların ayıklandığı adımdır.
- **Veri Seçme:** Bu adım, veri kümelerinden, keşfin gerçekleştirileceği uygulama alanına uygun örneklem kümesinin seçilmesidir.
- **Veri Dönüşümü:** Verinin Veri Madenciliği tekniğinden kullanılabilecek hale dönüşümünü gerçekleştirir.
- **Veri Madenciliği:** Kullanılacak veri madenciliği algoritmasının (sınıflandırma, kümeleme vb.) belirlenerek uygulanmasıdır.
- **Örüntü Değerlendirme:** Veri madenciliği sonucunda keşfedilen örüntülerin geçerlilik, yenilik, yararlılık ve basitlik kıstaslarına göre değerlendirilir.
- **Bilgi Sunumu:** Madenciliği yapılmış olan elde edilmiş bilginin kullanıcıya sunumunu gerçekleştirilir (Delen vd, 2005).

3.1.4. Veri Madenciliğinde Kullanılan Yaklaşımlar

Veri madenciliğinde üç tür yaklaşım kullanılmaktadır bunlar sınıflandırma, kümeleme, birliktelik analizi ve ardışık analizdir.

3.1.4.1.Sınıflandırma Yaklaşımı

Sınıflandırma, önceden belirlenen veri sınıflarına (hedef sınıf) göre model geliştirilip bu modelin sınıflandırma için kullanıldığı iki adımlı bir süreçten oluşmaktadır. Veri elemanının, daha önceden tanımlanmış farklı sınıflardan hangisine ait olduğunu keşfeden, tahmin edici fonksiyonun bulunmasını sağlar. Elde edilen modeller, karar ağaçları, matematiksel formüller ve sinir ağları gibi çeşitli yöntemlerle gösterilebilir. Sınıflama, kategorik değerlerin tahmin edilmesinde kullanılmaktadır.

Sınıflandırma ve tahmin yöntemlerinde aşağıdaki karşılaştırma kriterleri önerilmiştir;

- Tahmin doğruluğu (Predictive Accuracy): Modelin yeni ya da daha önceden bilinmeyen bir verinin sınıf etiketinin doğru olarak tahmin edilme yeteneğidir.

- Hız (speed): Modelin genelleştirilmesindeki hesaplama maliyetine karşılık gelir.
- Sağlamlık (Robustness): Modelin, gürültülü ya da kayıp veri verildiğinde doğru tahmin yapma yeteneğidir.
- Ölçeklendirilebilirlik (Scalability): Büyük miktarlarda veri verildiğinde, modeli etkin bir şekilde kurma yeteneğidir.
- Yorumlanabilirlik (Interpretability): Model tarafından sağlanan anlaşılma seviyesidir.

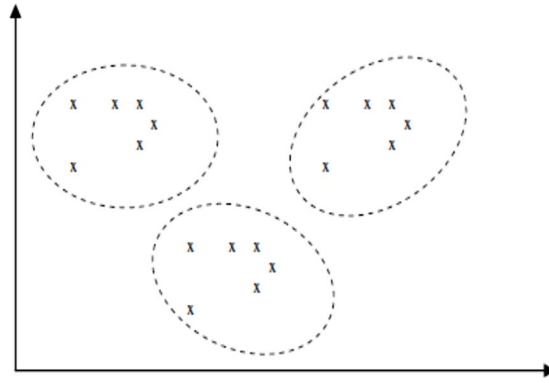
Veri Madenciliğinde yaygın olarak kullanılan sınıflandırma yöntemleri aşağıda verilmiştir.

1. Karar ağacı ile sınıflandırma yöntemi
2. Bayes ile sınıflandırma yöntemi
3. NaiveBayes ile sınıflandırma yöntemi
4. k- en yakın komşuluk ile sınıflandırma yöntemi
5. Vaka tabanlı nedenleşme ile sınıflandırma yöntemi
6. Genetik algoritma ile sınıflandırma yöntemi
7. Kaba küme teorisi ile sınıflandırma yöntemi
8. Genetik programlama ile sınıflandırma yöntemi

3.1.4.2. Kümeleme Yaklaşımı

Kümeleme, var olan veri grubunu birbirinden farklı özelliklere ve niteliklere sahip parçalara veya gruplara ayırma işlemini gerçekleştirmektedir. Kümeleme algoritması aynı özelliklere sahip veriler aynı kümede, farklı özellik ve benzerlikler taşıyan veriler de farklı kümelerde bulunmaktadır. Sınıflamanın tersine kümeleme önceden tanımlanacak hedef bir değişkene gereksinim duymaz.

Kümeleme incelenmesi, kişilerin ya da nesnelerin sınıflandırılmasını ayrıntılı bir şekilde tanımlamak için geliştirilmiştir. Bu nedene yönelik, ele alınan örnekte yer alan varlıklar aralarındaki benzerliklere göre kümelere ayrılır, daha sonra bu gruplara dahil edilen bireylerin profili ortaya konur. Şekil 3.2’de bir kümeleme görülmektedir (Işık, 2006).



Şekil 3. 2. Veri kümeleme

Kümeleme analizi ile yapay sinir ağları, şekil tanıma ve istatistik gibi birçok alanda çalışılmaktadır (Doğan, 2006).

3.1.4.2.1. Kümeleme Yaklaşımın Avantajları

Kümeleme yaklaşımının avantajları, genel olarak şu şekilde maddelendirilebilir.

- İlişkilerin görüntülenmesi: Kümelemenin en önemli özelliklerinden biri grafikler sayesinde sonucun görüntülenebilmesidir. Görsel sonuç benzerliklerin kolay tespit edilmesini sağlar.
- Anormalliklerin tespiti: Grafikler sayesinde aykırı durumlar kolayca tespit edilir, böylece sıra dışı veriler belirlenir.
- Diğer veri madenciliği teknikleri için örneklemelerin yaratılması: Karar ağaçları gibi bazı teknikler çok büyük veriler üzerinde çalışamazlar. Bu metotların uygulanabilmesi için kümeleme analizi uygulanarak öncelikle verinin bir bölümünün seçilmesi ve en uygun başlangıç noktalarının belirlenmesi sağlanır (Seidman, 2006).

3.1.4.2.2. Kümeleme Yaklaşımın Zayıflıkları

Sonuçların anlaşılması zordur: Takip edilmesi gereken belirli kurallar olmadığı için tahminlerin tamamen gerçek olması mümkün değildir.

- Farklı veri tiplerinde özellikler içeren nesnelerin karşılaştırılması zordur.
- Küme, birbirlerine yakın bireylerin çok boyutlu uzayda oluşturdukları birlik olarak ifade edilebilir. Bu durumda küme kavramı, "benzerlik" ve "uzaklık" kavramlarını çağrıştırmaktadır.

- Noktaların geometrik olarak gösterimlerinde ikiden fazla boyut olduğunda noktalar arasındaki uzaklıkları çok boyutlu olarak hesaplamak gerekir (Işık, 2006).

3.1.4.2.3. Hesaplama Kullarılan Öölütler

Çeşitli sayıda değışkene göre nesneler arasındaki benzerlik ve uzaklıkları hesaplamak için bazı ölçüler kullanılmaktadır.

a. Minkowski uzaklığı

- Manhattan (City-Blok) uzaklığı ($n=1$),
- Öklid (Euclidean) uzaklığı ($n=2$),
- “Supremum” ($L_{\max}$ norm, L_{∞} norm) uzaklığı ($n=\infty$)

b. Tchebyshev uzaklığı

c. Mahalanobis uzaklığı

d. Canberra uzaklığı

e. Bray Curtis (Sorensen) uzaklığı

f. Kosinüs benzerliği

g. Genişletilmiş Jaccard benzerliği

h. Pearson İlişkisi

i. Spearman benzerliği

3.1.5. Kümeleme Metotları

Kullanılan kümeleme metotları genel olarak şunlardır:

3.1.5.1. Bölünmeli Yöntem

Veritabanındaki nesnelerin mantıksal olarak gruplandırılıp analiz edildiğı yöntemdir. Merkez tabanlı kümeleri tespit etmede başarılıdır. Gruplandırma yapılırken değışik kriterler değerlendirilebilir. Yapılan gruplandırma analizin kalitesine etki eder. Başlıca bölünmeli kümeleme algoritmalarından bazıları k-means, fuzzy c-means, k-medoids, Clara ve Clarans’dır.

3.1.5.2. Hiyerarşik Yöntem

Bu yöntemde, analiz işleminden önce nesneler, hiyerarşik bir yapıya göre düzenlenir. Veriyi hiyerarşik yapıya dönüştürmek için farklı yaklaşımlar kullanılır. Bu yaklaşımlar aşağıdaki gibi iki tipte sınıflandırılabilir:

- **Birleştirme yaklaşımı (Agglomerative):** Her bir nesne ayrı bir grup oluşturur. Birbirine yakın nesneler veya gruplar birleştirilir.
 - **Bölme yaklaşımı (Divisive):** Bütün nesneler başlangıçta aynı kümededir. Her ötelemede küme daha küçük kümelere ayrıştırılır.
- Başlıca hiyerarşik kümeleme algoritmalarından bazıları Birch, Cure ve Chamelon' dur.

3.1.5.3. Yoğunluk Bazlı Yöntem

Birçok kümeleme yönteminde, nesnelerin birbirlerine göre farklılıkları değerlendirilerek kümeleme işlemi yapılır. Bu yöntemler farklı şekillerdeki kümelerin tespitinde yetersiz almaktadır. Yoğunluk bazlı yöntemler ise nesnelerin yoğunluğuna göre gruplama yapar. Komşuluk içindeki yoğunluk belli bir seviyeyi aşana kadar kümeler büyümeye devam eder. Bu yöntemin gürültülü verilerden etkilenme oranı düşüktür. Yoğunluk bazlı yöntemlere örnek olarak Dbscan ve Optics algoritmalarıdır.

3.1.5.4. Grid Bazlı Yöntem

Grid yapısı oluşturulacak şekilde nesne uzayı hücrelere bölünür. Bütün kümeleme işlemleri bu yapı üzerinde yapılır. Bu yöntemin temel avantajı, hızlı ve nesnelerin sayısından bağımsız olmasıdır. Bu tipteki metotlara örnek olarak Sting, Clique ve Wave Cluster verilebilir.

3.1.5.5. Model Bazlı Yöntem

Her küme için bir model belirlenir ve bu modele uyan veriler uygun kümeye yerleştirilir. Bu algoritmalar, veri noktalarının uzaydaki dağılımını yansıtan yoğunluk fonksiyonu ile kümeleri belirler. Standart istatistiklere dayanarak küme sayısı otomatik olarak belirlenir ve gürültülü ve sıra dışı verilerden az etkilenmesi nedeniyle güçlü bir analiz sağlar (Işık, 2006).

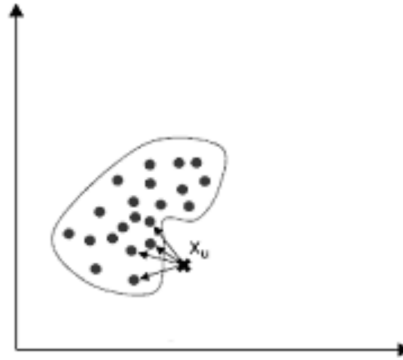
3.1.6. Bu Çalışmada Kullanılan Sınıflandırma Algoritmaları

Bu tez çalışmasında, sınıflandırma algoritmalarından olan Karar Ağaçları algoritmalarından Twoing ve Gini, K-nn ve Bayes algoritmaları kullanılmıştır. Bu algoritmalara ait genel bilgiler aşağıda verilmiştir.

3.1.6.1. K - NN Algoritması

K en yakın komşu algoritması, sınıflandırma problemini çözen denetimli öğrenme (sınıflandırma için öğrenme kümesi kullanır) algoritmalarından biridir (Çalışkan, 2008). Bu yöntem, örüntü tanımanın en klasik yöntemlerinden birisi olup, tespit edilmek istenen örüntünün en yakın komşusunun sınıfına dahil etmesiyle gerçekleştirir. Metot, örnek vektörün istatistiksel dağılımından bağımsız olup, yalnızca en yakın komşunun sınıfına göre bir sınıflandırma yaparak tanıma işlemini gerçekleştirir (Doğan, 2006)

KNN yönteminde sınıflandırma yapılacak verilerin öğrenme kümesindeki normal davranış verilerine benzerlikleri hesaplanarak en yakın olduğu düşünülen k verinin ortalamasıyla, belirlenen eşik değere göre sınıflara atamaları Şekil 3.3'te görüldüğü gibi yapılır (Çalışkan, 2008).



Şekil 3. 3. KNN algoritması

Bu yöntemde, örnek örüntünün vektörü alınarak, veri tabanındaki her bir vektöre olan uzaklığı hesaplanır. En yaygın kullanılan mesafe hesaplama yöntemi Öklid uzaklığı olsa da, başka herhangi bir mesafe hesaplama yöntemi de kullanılabilir. Tanımlanacak olan örnek örüntünün vektörü, veri tabanındaki kendisine en çok benzeyen, örüntünün sınıfından ayrılır. Böylece örnek örüntünün vektörü en çok hangi sınıfa dahil oluyorsa, örnek örüntü bu sınıf türünden bir örüntü olarak tanımlanır. Bu yöntem matematiksel olarak aşağıdaki gibi ifade edilebilir (Sirivd, 1998).

Kümeleme işleminde kümelerin merkez noktaları ve küme içindeki nesneler arası uzaklık için kullanılan hata parametresini hesaplamak gerekmektedir. Hata parametresi hesaplamak için, üç farklı hesaplama yöntemi kullanılmaktadır (Doğan, 2006).

Sınıflandırma sırasında çıkarılan özelliklerden sınıflandırılmak istenen yeni verinin daha önceki bireylerden k tane yakınlığa bakılır. Önce $k=3$ tane yeni bir eleman sınıflandırılmak istensin bu durumda önceden sınıflandırılmış elemanlardan en yakın üç tanesi alınır. Bu elemanlar hangi sınıfa aitse yeni elemanda o sınıfa eklenir. Uzaklık hesaplaması için Öklit yöntemi kullanılır (Doğan, 2006).

Verilen ekleme yani bir defa kaç bölümden sınıftan oluşuyorsa o kadar komşu (k) bulunur. k adet komşu arar ve hangi komşuya özellik bakımından yakınsa o sınıfa giren mesafeler hesaplanarak bulunur. KNN algoritması denetlemeli bir öğrenme algoritmasıdır.

Yakın komşu algoritması denetimli bir öğrenme algoritmasıdır. Bu algoritmada, sınıfı belirsiz bir örnek geldiğinde bu örnek var olan öğrenme bilgileri kullanılarak sınıflandırılmaktadır. Algoritma, sınıfı belirsiz olan bu yeni örneğin en yakın komşularına göre örneğin sınıfına karar verir (Doğan, 2006).

3.1.6.1.1. KNN Algoritmasının Adımları

- Test kümesindeki her verinin $X = \{x_1, x_2, x_3, x_4, \dots, x_N\}$, öğrenme kümesindeki verilere $D = \{d_1, d_2, d_3, d_4, \dots, d_M\}$ yakınlığı hesaplanır.

$$sim(x_i, d_l) = \frac{x_i \cdot d_l}{\|x_i\| \|d_l\|} \quad (3.1)$$

$$(i = \{1, 2, 3, \dots, n\}, l = \{1, 2, 3, \dots, m\})$$

- Her verinin öğrenme kümesindeki verilere olan yakınlıkları sıralanıp ilk “ k ” tanesi alınarak ortalamaları hesaplanır.

$$sim_{avg}(x_i) = \frac{\max(\sum_{l=1}^k sim(x_i, d_l))}{k} \quad (3.2)$$

Ortalama değerleri, belirlenen eşik değerinden büyük olanlar normal, küçük olanlar ise anormal olarak sınıflandırılır. Tablo 3.1’de k-nn algoritmasının yapısı görülmektedir.

Tablo 3. 1. K-NN Algoritma yapısı

o	-Öğrenme kümesi D belirlenir $D = \{d_1, d_2, d_3, \dots, d_M\}$;
o	- fortest kümesi(X) içindeki her veri xi için do
o	-- if xi bilinmeyen bir sistem çağrısı ise then
o	-----xi = anormal;
o	-- else
o	----- for öğrenme kümesi(D) içindeki her veri dl için do
o	-----benzerlik hesapla sim(xi, dl);
o	----- if sim(xi, dl) = 1 then
o	-----xi =normal; exit;
o	-----En büyük k tane sim(xi, dl) benzerliği bul;
o	-----En yakın k komsu için benzerlik ortalaması(sim_avg) hesapla;
o	----- if sim_avg(xi) > esik değeri then
o	-----xi =normal;
o	----- else
o	-----xi =anormal;

3.1.6.2. NaiveBayes Algoritması

Bayes sınıflandırıcılar istatistiksel sınıflandırıcılardır. Bayes verilerin üyelik olasılıklarını yani belirli bir sınıfa ait olma olasılıklarını tahmin eder. Bayes sınıflandırıcı aşağıda tarif edilen Bayes teoremine dayanmaktadır. Sınıflandırma algoritmalarının karşılaştırılmaları çalışmalarında, Bayes algoritması karar ağaçları ve yapay sinir ağları algoritmalarıyla performans olarak karşılaştırılmaktadır. Bayes sınıflandırıcı büyük veri tabanları üzerinde uygulandığında yüksek doğruluk ve hız sergilemiştir (Han ve Kamber, 2000.9).

Bir veri setindeki, herhangi bir örnek $X = \{x_1, x_2, x_m\}$ nitelik değerlerinden oluşsun, Toplam m adet sınıf olduğu varsayılırsa, sınıfı belirlenecek örneğe ilişkin olarak Denklem 2.3 ile olasılık hesaplamaları yapılır (Ren vd, 2009).

$$P(C_i|X) = \frac{P(X|C_i) P(C_i)}{P(X)} \quad (3.3)$$

$p(X|C_i)$:Sınıf i'den bir örneğin x olma olasılığı

$P(C_i)$:Sınıf i'nin ilk olasılığı

$p(X)$: Herhangi bir örneğin x olma olasılığı
 $P(C_i|X)$: X olan bir örneğin sınıf i 'den olma olasılığı (son olasılık)

Hesaplamalardaki işlem yükünü azaltmak için $P(X|C_i)$ olasılığı basitleştirilebilir. Bunun için, örneğe ait x_i değerlerinin birbirinden bağımsız olduğu kabul edilerek, Denklem 2.4 kullanılır (Delen vd, 2005).

$$P(X|C_i) = \prod_{k=1}^n P(x_k|C_i) = P(x_1|C_i) \cdot P(x_2|C_i) \dots P(x_n|C_i) \quad (3.4)$$

3.1.6.3. Bayes Teoremi

Üyeliği bilinmeyen X örneği için. H ise bu veri örneği X in C sınıfına ait olduğunu öngören bir hipotez olsun.

Bayes Teoremi $P(H|X) = \frac{P(X|H) P(H)}{P(X)}$ dir.

Sade Bayes sınıflandırıcı veya Bayes sınıflandırıcı aşağıdaki şekilde çalışır:

1. Her veri örneği n -boyutlu özellik vektörleri ile gösterilir. $X = (x_1, x_2, x_n)$ her veri örneği n özelliklerden alınan örnek üzerindeki n ölçümler ile tarif edilir, sırasıyla, $A_1, A_2, \dots, A_n$.

$$P(C_i|X) = \frac{P(X|C_i) P(C_i)}{P(X)}$$

2. $P(C_i) = \frac{S_i}{S}$ olabilmektedir. Burada S_i , C_i sınıfına ait eğitilen örnek sayısı ve S ise toplam eğitilen örnek sayısıdır.

3. Birçok nitelik barındıran veri setleri verilirse $P(X|C_i)$ aşırı derece hesap yükü gerektirir. $P(X|C_i)$ işleminde hesap yükünü azaltmak için, sınıf koşul bağımsızlığına ait saf varsayım uygulanır. Bu varsayım, özelliklere ait değerler bir diğerinden şartlı olarak bağımsızdır, örneğe ait verilen sınıf üyeliği, özellikler arasında bağımlılık ilişkisi yoktur.

Bu $P(X|C_i) = \prod_{k=1}^n P(x_k|C_i)$ olasılıklar $(x_1|C_i).P(x_2|C_i) \dots P(x_n|C_i)$ eğitim örneklerinde tahmin edilebilir, burada $P(C_i) = \frac{S_i}{S}$ burada S_{ik} A_k için x_k değerine sahip olan C_i sınıfına ait eğitim seti sayısı ve S_i C_i 'ye ait eğitim seti sayısıdır. $P(x_k|C_i) = \frac{S_{ik}}{S_i}$ ifadesinin sıfır olmasını engellemek için $P(x_k|C_i) = \frac{S_{ik}+1}{S_i+a_i}$ ifadesi kullanılır, a_i özellik sayısı örneğin “Gelir” için bu değer üçtür.

4. $P(X)$ ifadesi sabit olduğu için hesaplanmaz, çünkü önemli olan birbirlerine göre oransal büyüklükleridir. Bilinmeyen örnek X 'i sınıflandırmak için, her C_i sınıfı için $P(X|C_i) P(C_i)$ ifadesi hesaplanır. Burada hangisi büyükse o değer X için geçerli sınıf olur (Doğan, 2006).

3.1.6.4. Twoing Algoritması

Veri madenciliğinin sınıflandırma algoritmalarının içinde yer alır. Metot 1984'te *Breiman* tarafından ortaya atılmıştır. Karar düğümünün her seferinde ikiye ayrılma esasına dayanmaktadır. Yani karar ağacında sağ sol şeklinde ikili dallanma vardır. Bir düğümde ayrılabilir, seçilebilir iki düğüm vardır. CART algoritmasında bölünme işlemi, bir düğüme belirli bir kriter uygulanarak gerçekleştirilir. Bunun için öncelikle tüm niteliklerin var olduğu değerler dikkate alınır ve tüm eşleşmelerden sonra iki bölünme elde edilir (Özkan, 2013).

Adım 1:

- Nitelik değerlerine göre eğitim kümesi ikiye ayrılır. Bunlara *aday bölünme* adı veriliyor. Bir t düğümünde “sağ” ve “sol” olmak üzere iki ayrı dal bulunur. Burada, kümeler t_{sol} ve $t_{sağ}$ olarak adlandırılır.
- Her aday bölünme için P_{sol} ve $P(j|t_{sol})$ olasılıkları hesaplanır. Burada $P(j|t_{sol})$, bir j değerinin sol tarafta bulunma olasılığını ifade eder. Söz konusu j değerleri, sınıf değerlerinin yer aldığı nitelikler olarak isimlendirilir. Olasılık hesapları şu şekilde yapılır:

$$P_{Sol} = \frac{t_{sol} \text{ daki herbir nitelik deęerinin ilgili nitelik sütünundaki tekrar sayısı}}{\text{Eđitim kümesindeki kayıtların sayısı}} \quad (3.5)$$

$$P(j|t_{sol}) = \frac{t_{sol} \text{ daki kayıtların } j \text{ sayısı}}{t_{sol} \text{ daki herbir nitelik deęerinin ilgili nitelik sutunundaki tekrar sayısı}} \quad (3.6)$$

- Her aday bölünme için $P_{Saę}$ ve $P(j|t_{saę})$ olasılıkları da hesaplanmalıdır. Burada $P(j|t_{saę}), j$ deęerinin saę tarafta bulunma olasılıđını ifade eder.

$$P_{Saę} = \frac{t_{saę} \text{ daki herbir nitelik deęerinin ilgili nitelik sütünundaki tekrar sayısı}}{\text{Eđitim kümesindeki kayıtların sayısı}} \quad (3.7)$$

$$P(j|t_{saę}) = \frac{t_{saę} \text{ daki kayıtların } j \text{ sayısı}}{t_{saę} \text{ daki herbir nitelik deęerinin ilgili nitelik sutunundaki tekrar sayısı}} \quad (3.8)$$

- $\Phi(s|t)$, t düęümündeki s aday bölünmelerinin uygunluk ölçüsü olarak adlandırılır ve řu řekilde hesaplanır:

$$\Phi(s|t) = 2P_{sol}P_{saę} \sum_{j=1}^n |P(j|t_{sol}) - P(j|t_{saę})| \quad (3.9)$$

- $\Phi(s|t)$ deęerleri hesaplandıktan aralarında en büyük deęerin bulunduęu aday bölünme satırı dallanmanın yapılacađı satırı gösterir. Dallanma yapıldıktan sonra bu adıma bađlı olarak karar ađacı çizilir (Özkan, 2013).

Adım 2:

Algoritmanın birinci adımına dönölerek ađacın alt kümesine aynı işlemler uygulanır.

3.1.6.5.Gini Algoritması

Gini Algoritması da ikili bölünmelere dayalı bir sınıflandırma algoritmasıdır. Twoing algoritmasına benzer řekilde, nitelik deęerlerinin ikili ađaç řeklinde dallanma esasına dayanmaktadır. (Özkan, 2013).

Algoritma:

- Nitelik değerlerinin her biri, ikili olacak biçimde gruplanır. Elde edilen sınıf değerleri böylece sol ve sağ bölünmelere karşılık gelecek şekilde gruplandırılır.
- Niteliklerin her biri için sol ve sağ taraftaki bölünmeler için $Gini_{sol}$ ve $Gini_{sağ}$ değerleri hesaplanır.

$$Gini_{sol}=1-\sum_{i=1}^k \left(\frac{L_i}{|T_{sol}|} \right)^2 \quad (3.10)$$

$$Gini_{sağ}=1-\sum_{i=1}^k \left(\frac{R_i}{|T_{sağ}|} \right)^2 \quad (3.11)$$

Denklemden yer alan ifadeler şu şekildedir:

k: Sınıfların sayısı

T: Bir düğümdeki örnekler

$|T_{sol}|$: Sol taraftaki örneklerin sayısı

$|T_{sağ}|$: Sağ taraftaki örneklerin sayısı

L_i : Sol tarafta i kategorisindeki örneklerin sayısı

R_i : Sağ tarafta i kategorisindeki örneklerin sayısı

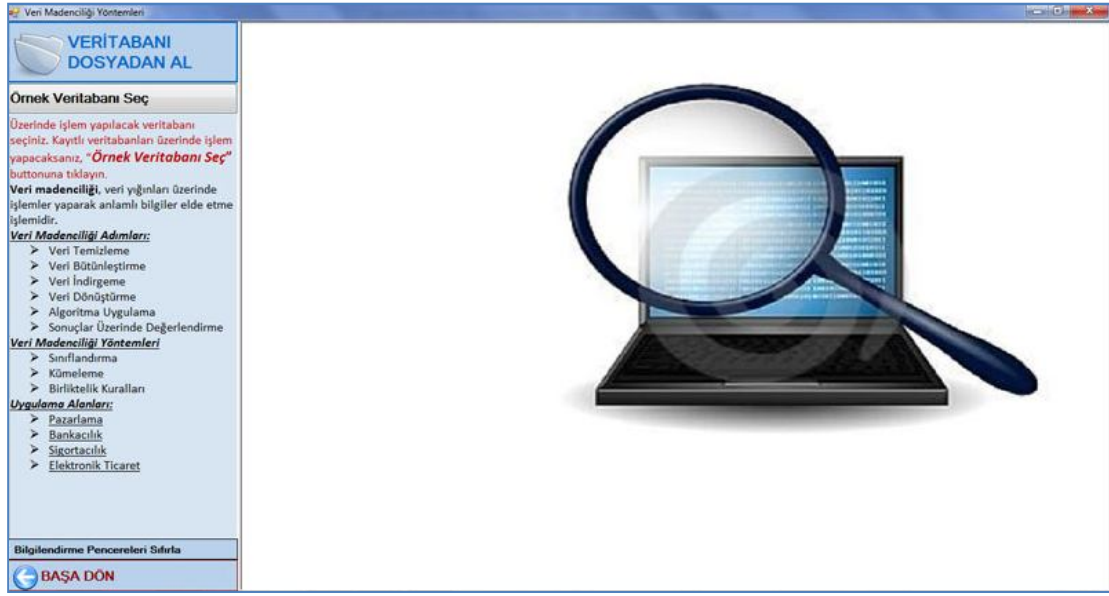
- Her j niteliği için, n eğitim kümesindeki satır sayısı olmak üzere aşağıdaki bağıntının değeri hesaplanır:

$$Gini_j = \frac{1}{n} (|T_{sol}| Gini_{sol} + |T_{sağ}| Gini_{sağ}) \quad (3.12)$$

- Her j niteliği için hesaplanan $Gini_j$ değerleri arasından en küçük olanı seçilir ve bölünme bu nitelik üzerinden gerçekleştirilir.
- En baştaki adıma dönülerek işlemlere devam edilir (Özkan, 2013).

4. EĞİTSEL VERİ MADENCİLİĞİ ARACI

Bu tez çalışmasında, veri madenciliği uygulamaları için EVMA(Eğitsel Veri Madenciliği Aracı) isimli bir yazılım geliştirilmiştir. Bu yazılım, hem eğitimde hem de veri madenciliği çalışması yapacak kullanıcılar için kullanışlı bir şekilde hazırlanmıştır. Geliştirilen EVMA yazılımının ana sayfa ekran görüntüsü Şekil 4.1’de verilmiştir.



Şekil 4. 1. EVMA yazılımına ait ana sayfa görüntüsü.

EVMA; farklı öğrenci grupları üzerinde bilgisayar destekli öğretimin katkılarını görmek için geliştirildiğinden, program’ a giriş kullanıcı adı ve şifre üzerinden yapılabilmektedir. Bunu esnek kılmak adına bir metin dosyası içerisine girilen bir şifre ile iptal edilebilmektedir. Sisteme giriş yapıldıktan sonra programda yapılabilecek işlemler sol tarafta yönlendirici bir bilgi panelinde verilmiştir. Bu panelde, kullanılabilecek algoritmalar ve veri madenciliği hakkında bilgiler yer almaktadır. Kullanıcılar giriş yaptıktan sonra, veri madenciliği algoritması uygulayacağı veri setini yazılımın sol üst kısmında bulunan “*VERİTABANINI DOSYADAN AL*” butonuna tıklayarak seçebilir. Kullanıcının, algoritmalar için kullanılabileceği dosya türleri Microsoft Excel dosyaları, metin dosyaları ve Microsoft Access dosyalarıdır. Kullanıcı “Örnek Veritabanı Seç” butonuna tıklayarak uygulamada kayıtlı olan veri setlerinden herhangi birisini seçilebilmektedir.

Her veri madenciliği algoritması için ayrı ayrı ipucu bilgi formları EVMA içerisinde oluşturulmuştur. Kullanıcı dışarıdan bir veri seti seçtiğinde ya da kayıtlı bir veri seti açtığında veri seti üzerinde düzenlemeler yapabileceği bir form açılmaktadır.

Bu form kapsamında yapılabilecek işlemler şunlardır;

- Normalleştirme Uygula
- Veri Temizleme
- Verileri Kodla
- Nitelik Kategorileştirme
- İşlemleri Geri Al
- KAYDET

Yukarıda verilen bu işlemler veri madenciliğinin temel aşamalarını içermekle birlikte her bir işleme ait işlevler kısaca aşağıdaki şekilde özetlenebilir;

- *Normalleştirme Uygula* seçeneğinde, “Min-Max” ve “Z-Score” normalleştirme seçenekleri bulunmaktadır. Seçimi yapılan Sütun/Nitelik üzerinde ilgili normalleştirme işlemi uygulanır. Seçilen niteliğinin verilerinin nümerik kontrolü yapıldıktan sonra verinin normalleştirilmesi gerçekleştirilir.
- *Veri Temizleme* seçeneğinde, veri setinin niteliklere göre benzersiz kayıtlarını gösterip kayıp/yanlış veri üzerinde silme ya da değiştirme işlemleri yapılarak veri temizleme süreci gerçekleştirilir.
- *Verileri Kodla* seçeneğinde, seçilen nitelik veya nitelikler sıfırdan başlayarak her benzersiz kayıt için birer birer artarak yeniden yazılır.
- *Nitelik Kategorileştirme* seçeneğinde, sayısal değerleri kategorik değerlere dönüştürme işlemi yapılmaktadır.
- *İşlemleri Geri Al* seçeneğinde, veri seti üzerinde yapılan bütün değişiklikler iptal edilir veri seti ilk yüklendiği hale getirilir.
- *KAYDET* seçeneğinde, form üzerinde yapılan değişikliklerle kalıcı olarak dosyaya kaydedilir.

Veri setine ait düzenlemeler yapıldıktan sonra, veri seti üzerinde uygulanacak veri madenciliği algoritmasının seçimi gerçekleştirilerek işlemler devam ettirilir.

4.1. Eğitsel Veri Madenciliği Aracı İçerisindeki Algoritmalar

EVMA içerisinde, veri madenciliği alanında kullanılan temel algoritmalar yer almaktadır. Bu yöntemler genel olarak; istatistiksel sınıflandırma, birliktelik kuralları, kümeleme algoritmaları, regresyon ağaçları, karar ağaçları ve bellek tabanlı algoritmalar şeklindedir. Bu yöntemler aynı zamanda bilgisayar bilimlerinde verilmekte olan veri madenciliği dersinin müfredatına da uygun olarak düzenlenmiştir.

4.1.1. İstatistiksel Sınıflandırma

4.1.1.1. Bayes Algoritması

Bayes algoritması veri madenciliği kapsamında sıklıkla kullanılan bir istatistiksel sınıflandırma algoritmasıdır. EVMA içerisinde hazırlanan Bayes algoritmasının kullanımı için, kullanıcı ilk önce sınıf sütun/nitelik seçimini gerçekleştirir. Daha sonra sınıfı bulunacak yeni kayıt için bilgilerin girileceği bir form açılır. Kullanıcı ilgili bilgileri girdikten sonra açılan formda kullanıcının girdiği bilgilerin tekrar sayıları ile bu bilgilerin karşılık geldiği sınıf sütunun benzersiz değerlerinin sayılarını yazar. Şekil 4.2’de görüldüğü gibi sayısal değerler için yine bütün sınıf değerleri üzerinden hesaplama gerçekleşmiş şekliyle yazar. Sonuç hesapla butonuna tıklandığında girilen kayıt için sınıf hesaplaması gerçekleşir ve ilgili kayıta bir sınıf belirlenmiş olur.

Veri Madenciliği Yöntemleri

Bayes

Bayes Sınıflandırıcılar

Bayes sınıflandırıcıları istatistiksel sınıflandırma teknikleri arasında yer alır.

Sınıf üyeliği bilinmeyen bir X kümesinin; bütün sınıflar için olasılıklarının hesaplanması ve hesaplanan olasılık değerlerinden büyük olanın seçilmesi esasına dayanır.

$P(X|C_i) = \prod_{k=1}^n P(x_k|C_i)$

X_i değerleri bağımsız olasılıkları;

En büyük sonrasal sınıflandırma yöntemi;

$C_{map} = \arg \max_c \prod_{k=1}^n P(x_k|C_i)$

Sfır Değer Sorunu;

Bir X kümesinin sınıf olasılıkları

BAŞA DÖN

örnekt.xlsx

EĞİTİM **YAŞ** **CİNSİYET** **KABUL**

BayesForm

Aday Bölünme	Nitelik Adı	Özellik	Aday Tekrar Sayısı	Sınıf/EVET	Sınıf/HAYIR
0	YAŞ	60	Sonrasal	0.02068801765	0.00013425469
1	EĞİTİM	ORTA	3	3	0
2	EĞİTİM	İLK	3	1	2
3	EĞİTİM	YÜKSEK	2	1	1
4	CİNSİYET	ERKEK	4	3	1
5	CİNSİYET	KADIN	4	2	2

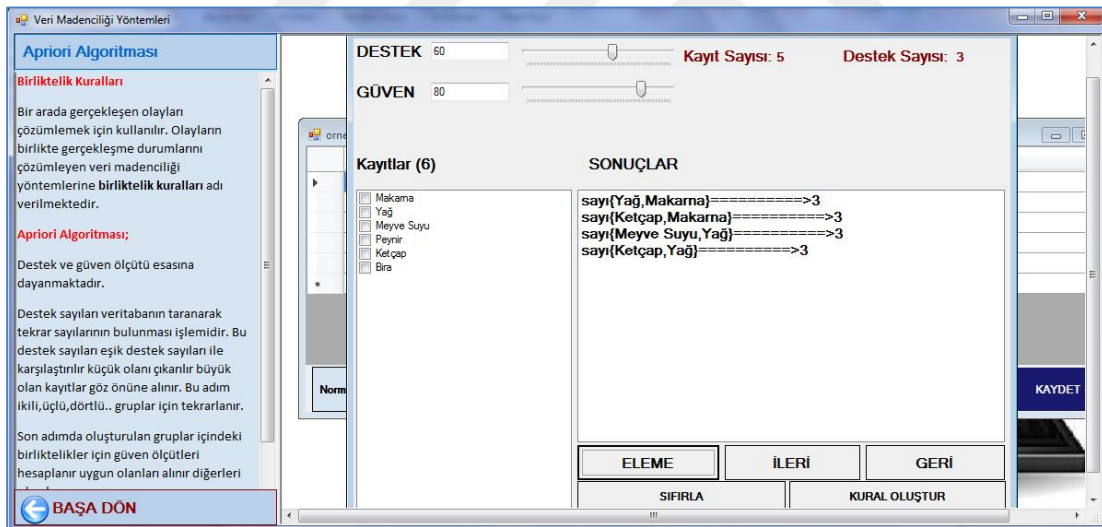
Sonuç Hesapla **Test Verisi: ORTA,60,ERKEK** **KAYDET**

Şekil 4.2 EVMA yazılımı içerisinde Bayes algoritmasının kullanımı.

4.1.2. Birliktelik Kuralları

4.1.2.1. Apriori Algoritması

Apriori algoritması, sıklıkla tekrar eden öge setleri üzerinde çıkarım gerçekleştirmektedir. Şekil 4.3’de, EVMA içerisinde Apriori algoritmasına ait bir uygulama örneği verilmiştir. Veri seti üzerindeki her kayıt için tekrar sayısı formda yer almaktadır. “İleri” butonu tıklandığında bu kayıtlar kendi aralarında gruplandırılarak yine veri setindeki sayıları listeye eklenir. Bu adımlar en uzun kayıt grubuna erişine dek devam eder. “Geri” butonuna tıklayarak belirtilen adımlar bir önceki duruma getirilir. Destek değeri girilip “Eleme” butonuna tıklandığında listede destek değerlerine uymayanlar silinir. Güven değeri girildikten sonra “Kural Oluştur” butonuna tıklandığında destek ve güven kriterlerine uygun kayıtlar için kurallar oluşturur. “Sıfırla” butonuna tıklandığında ise kayıtların ilk durumuna geri dönlür.



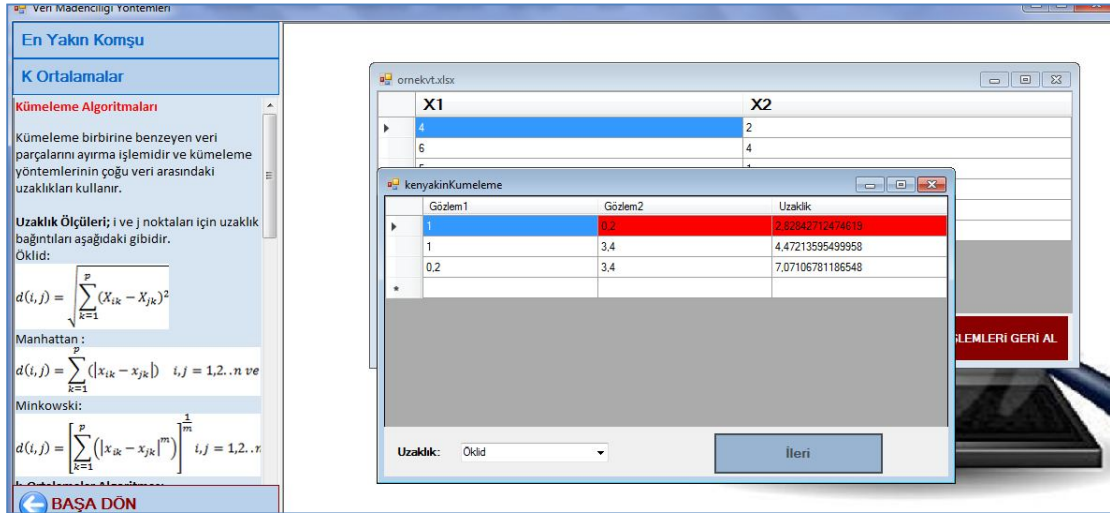
Şekil 4. 3. EVMA içerisindeki Apriori algoritmasının kullanımına ait bir örnek.

4.1.3. Kümeleme Algoritmaları

4.1.3.1. En Yakın Komşu

Kümeleme algoritmaları içerisinde en popüler olan en yakın komşu algoritması yazılım aracı içerisinde etkili bir şekilde gerçekleştirilmiştir. En yakın komşu

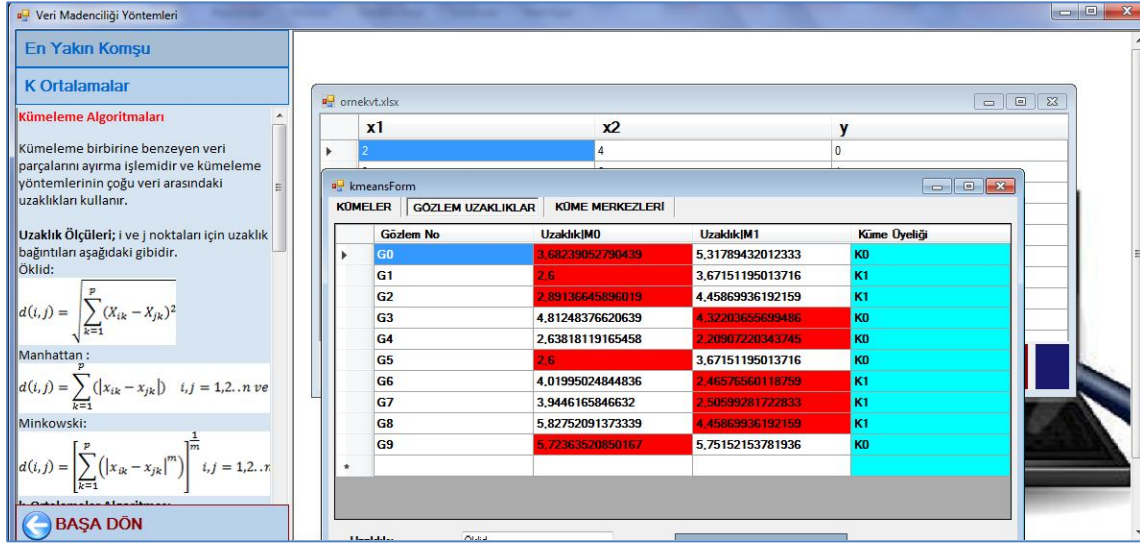
algoritmasının uygulamasında, veri seti seçildikten sonra açılan formda komşu uzaklıkları için uzaklık hesaplamaları için Öklid, Minkowski veya Manhattan seçeneklerinden biri seçilir ve ilgili uzaklıklar hesaplanır. Şekil 4.4'te görüldüğü gibi “İleri” butonuna tıklandığında en yakın komşuya göre kümeler oluşturulur. Veriye ait en yakın komşu kırmızı arka plan rengi ile gösterilmektedir. Her bir adımda bir nokta grubu birleştirilmektedir. Bu işlemler iki grup kalana kadar devam ettirilir.



Şekil 4. 4. En yakın komşu algoritmasına ait örnek bir gösterim.

4.1.3.2. K-Ortalamalar

EVMA içerisinde yer alan K-ortalamlar algoritması formunda üç sekme yer almaktadır. Bu sekmeler; kümeler, gözlem uzaklıkları, kayıtlar şeklindedir. Şekil 4.5’de görüldüğü gibi kümeler sekmesinde seçilen küme sayısı ve bu kümeye ait kayıtlar ile e^2 değerleri bulunmaktadır. Gözlem uzaklıkları sekmesinde ise gözlemlerin (kayıtların) küme merkezlerine olan uzaklıkları hesaplanmaktadır. İlgili gözlemlerin küme üyelikleri burada da bulunmaktadır. Küme merkezleri sekmesinde ise Kümeler ve veri setine ait niteliklere dair bilgiler bulunmaktadır.

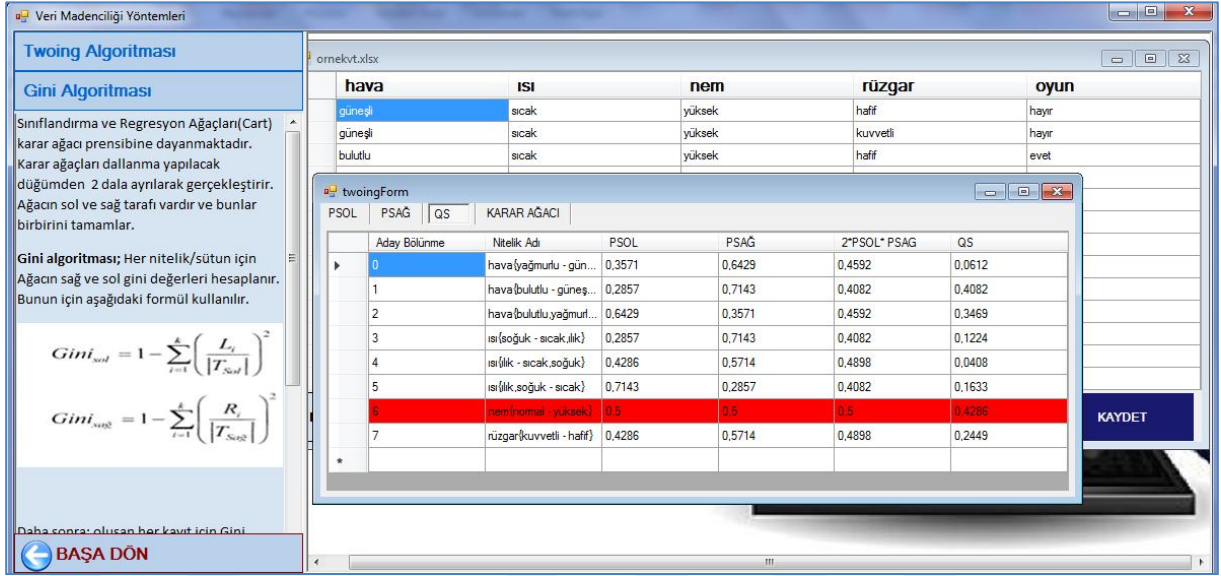


Şekil 4. 5. K-ortalamlar algoritmasına ait örnek bir gösterim.

4.1.4. Regresyon Ağaçları

4.1.4.1. Twoing algoritması

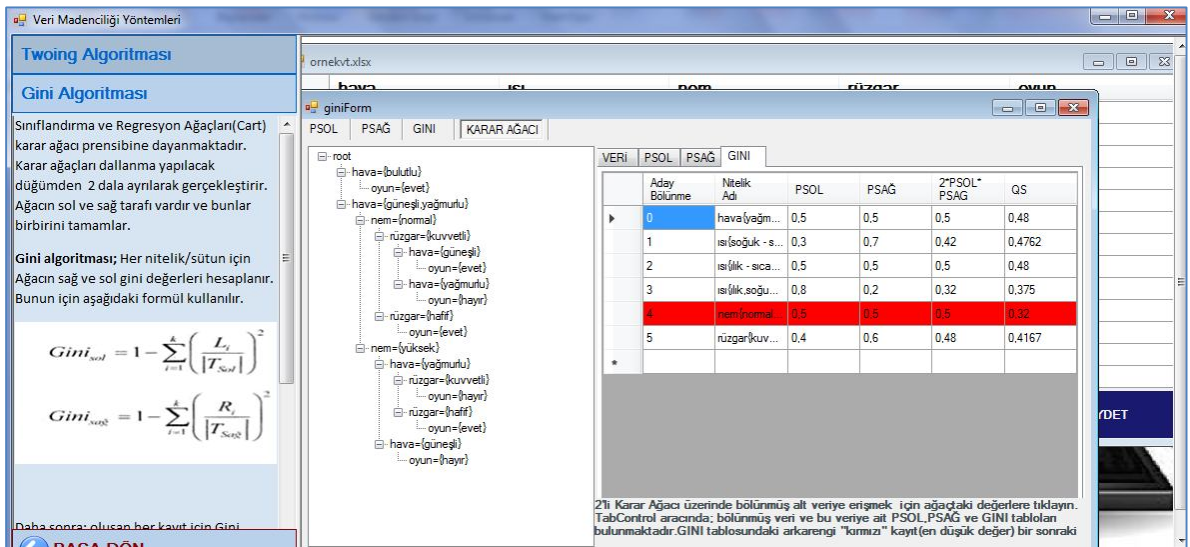
Twoing algoritması, ikili karar ağacı esasına dayalıdır. Sınıf sütunu seçildikten sonra form açılır. Açılan formda ilk adıma ait Twoing “Psol” ve “Psağ” değerleri tablo halinde ayrı sekmelerde tutulmaktadır. Yine bu veri setine ait TwoingQs değerleri “Qs” sekmesinde tablo halinde tutulmaktadır. Burada en yüksek Qs değeri (dallanmanın yapılacağı alt küme) kırmızı arka plan ile gösterilmiştir. Karar ağacı sekmesinde diğer adımlara ait veriler bulunmaktadır. Bu sekmede “.NET treeview” aracı kullanılarak ikili karar ağacı oluşturulmaktadır. Ağacın herhangi bir dalına tıklandığında ise Şekil 4.6’da görüldüğü gibi ilgili dala ait bölünmüş veriler Psol, Psağ ve twoing tablosu ile ilgili twoing Qs değerleri gibi ayrı ayrı sütunlarda verilmektedir. Burada dallanmanın yapılacağı en yüksek Qs değeri kırmızı olarak işaretli gelmektedir.



Şekil 4. 6. Twoing algoritmasına ait örnek bir gösterim.

4.1.4.2. Gini Algoritması

Gini algoritması, twoing algoritması gibi ikili ağaç karar ağacı esasına dayanmaktadır. Algoritmadaki hesaplamalar değişmekle beraber sonuç ikili ağaç şeklindedir. Bu nedenle form üzerinde kullanılan bütün araçlar twoing algoritmasındaki ile aynıdır. Gini algoritmasında; twoing algoritmasının tersine “giniindex” değeri en küçük olan değer üzerinden dallanma gerçekleştirdiğinden, ilgili sekmelerde Giniindex değerlerinden en küçük olanı Şekil4.7’de görüldüğü gibi kırmızı arka plan rengi ile gösterilmektedir.

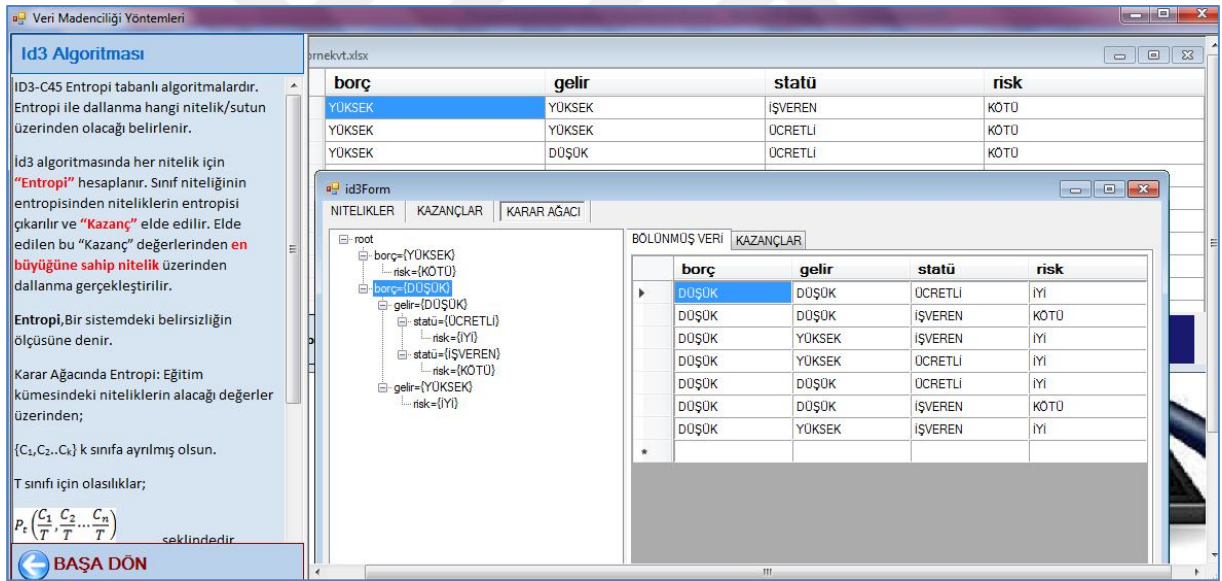


Şekil 4. 7. Gini Algoritmasına ait örnek bir gösterim.

4.1.5. Karar Ağaçları

4.1.5.1. Id3 Algoritması

Id3 algoritması, niteliklerin kazanç değerlerine göre dallanması ile oluşmaktadır. Hazırlanan araç üzerinde Id3 algoritması seçildiğinde form açılmadan önce oluşturulacak karar ağacına ait sınıf sütununun seçilmesi gerekmektedir. Formda ilk adım için kazanç değerleri ve nitelik alt değerlerine ait tekrar sayıları tablolarda tutulmaktadır. Bu değerler, tablolarda “Nitelik” ve “Kazançlar” sekmelerinde bulunmaktadır. Sonraki adımlar için Karar ağacı sekmesi bulunmaktadır. Bu algoritmada da karar ağacı aracı olarak “*NET treeview*” aracı kullanılmıştır. Şekil 4.8’de görüldüğü gibi ağacın dallarına tıklanarak alt veri setine ve ilgili veri setine ait Id3 kazanç değerleri bulunmaktadır. Kazanç değeri en yüksek değer bir sonraki dallanmayı gerçekleştirecek niteliğe aittir. Bu satırın arka plan rengi kırmızıdır.



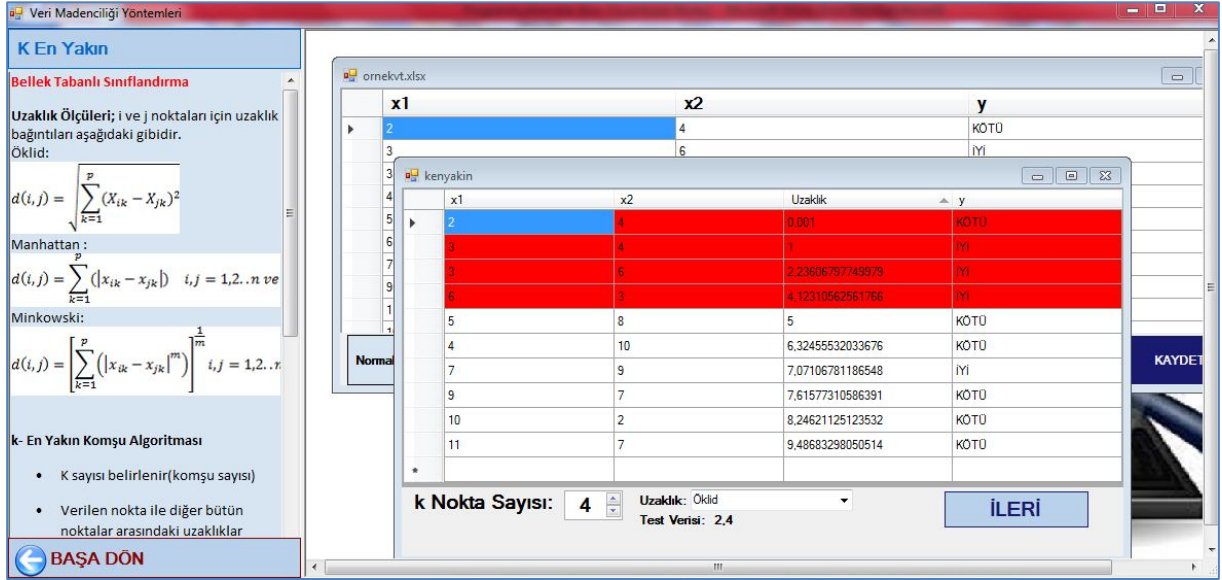
Şekil 4. 8. Veri madenciliği aracı içerisindeki Id3 algoritmasına ait bir görüntü.

4.1.6. Bellek Tabanlı algoritmalar

4.1.6.1. K En Yakın Komşu Algoritması

K-en yakın komşu algoritmasında girilen noktanın sınıfını bulmak için en yakın komşulara bakılmaktadır. Algoritmaya ait buton tıklandığında; sınıf sütunu seçilmesi istenecektir. Daha sonra sınıflandırılması istenen noktaya ait değerlerin girilmesi

istenmektedir. Form açıldıktan sonra açılan formda uzaklık seçimi ve k nokta sayısını değiştirmek için seçenekler bulunmaktadır. Şekil 4.9’da görüldüğü gibi en yakın k komşu sayısı girildikten ve uzaklık birimi seçildikten sonra “ileri” butonuna tıklanır ve en yakın k komşu seçilir. Daha sonra “sonuç hesapla” butonuna tıklandığında; en fazla tekrar eden sınıf ve ağırlıklı oylama için sonuçlar hesaplanır ve ana ekrana görünür.



Şekil 4. 9. Bellek tabanlı k-en yakın komşu algoritması.

5. GELİŞTİRİLEN YAZILIMININ DEĞERLENDİRİLMESİ

Bu bölümde, eğitsel yazılımın değerlendirilmesine ilişkin sonuçlara yer verilmiştir. Geliştirilen yazılımın değerlendirilmesine yönelik iki farklı yol izlenmiştir. Bunlardan birincisi, geliştirilen yazılımın öğrencilerin akademik başarıları üzerindeki katkılarına yönelik olup ikincisi ise anket yoluyla öğrenciler tarafından yazılımın değerlendirmesi şeklinde gerçekleştirilmiştir.

Bu iki farklı değerlendirme kriterine yönelik olarak öğrenciler iki gruba ayrılmış ve bu gruplardan birinin hazırlanan yazılımı kullanması diğerrinin ise yazılımı kullanmaması sağlanarak aralarında akademik başarı anlamında farklılıklar oluşup oluşmadığı gözlemlenmiştir. Eğitsel yazılım değerlendirme kriterlerinden biri de kullanıcıların yazılım hakkındaki görüşlerinin alınarak değerlendirilmesidir. Buna istinaden hazırlanan bir anket yardımı ile yazılımı kullanan öğrencilerden geri bildirimler alınmış ve bu geri bildirimler tez çalışmasında değerlendirilmiştir.

5.1. Evren ve Örneklem

Hazırlanan veri madenciliği yazılımının öğrencilere katkısının değerlendirilmesi amacıyla Fırat Üniversitesi Teknoloji Fakültesi Yazılım Mühendisliği'nde öğrenim görmekte olan öğrenciler bu çalışmada evren olarak seçilmiştir. Bu amaçla 2015-2016 eğitim-öğretim Bahar yarıyılında YMT 324 Veri Madenciliği dersini almakta olan öğrencilerden özellikle devam zorunluluğu olan öğrenciler örneklem olarak alınmıştır. Bu dersi alan öğrencilerden 120 kişi, 60 kişiden oluşan iki gruba ayrılmıştır. Birinci grubun öğrencileri, uygulama saatlerinde geliştirilen yazılımı kullanmaları, ikinci grup öğrencilerinin ise geliştirilen yazılımı kullanmadan geleneksel şekilde dersin uygulamalarına devam etmeleri sağlanmıştır. Bu iki grubun oluşturulması sırasında öğrencilerin geçmiş dönem notları dikkate alınmış ve her iki grubun Ağırlıklı Genel Not Ortalamalarının eşit olmasına özen gösterilmiştir. Bu şekilde denk iki grup oluşturulması ve akademik başarının bu iki grup arasında karşılaştırılabilmesi hedeflenmiştir. Tablo 5.1'de öğrenci gruplarına ait bilgiler yer almaktadır.

Tablo 5. 1. Gruplara Göre Öğrenci Sayıları ile Bu Öğrencilere Ait Akademik Başarı Ortalaması

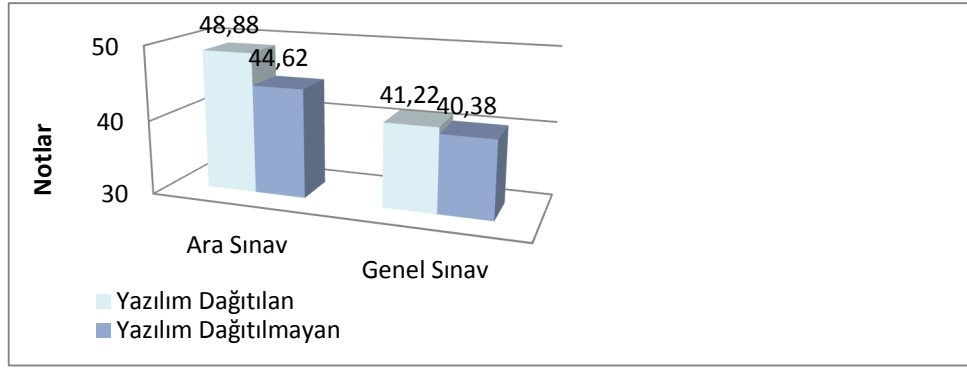
Grup	Sayı	Akademik Başarı Ortalaması
Yazılım <i>Dağıtılan</i> Öğrenci Grubu (1. Grup)	60	2.49
Yazılım <i>Dağıtılmayan</i> Öğrenci Grubu (2. Grup)	60	2.48

5.2. Akademik Başarıların Değerlendirmesi

Yazılımı kullanan ve kullanmayan öğrenci gruplarının başarısının tespiti için bu bölümünde eğitici veri madenciliği aracı kullanan öğrenciler ile kullanmayan öğrencilerin ara sınav ve genel sınav puan ortalamalarının karşılaştırılması amaçlanmıştır. Bu amaç doğrultusunda;

- Birbirine denk olarak oluşturulan bu öğrenci gruplarının ara sınav puanları ile genel sınav puanları arasında anlamlı düzeyde farklılık olup olmadığı araştırılmıştır.

Öğrencilerin ara sınav ve genel sınav puan ortalamalarında eğitici yazılım kullanılmasına bağlı anlamlı bir fark olup olmadığının belirlenmesi amacıyla gruplara bağımsız gruplar t-testi yapılmıştır. Bu teste ait analiz sonuçları Tablo 5.2.'de yer almaktadır. EVMA dağıtılan ve dağıtılmayan gruplar için sonuçlar Şekil 5.1'de gösterilmiştir.



Şekil 5.1. Sınav sonuçları EVMA dağıtılan ve dağıtilmayan gruplar.

Tablo 5. 2. Öğrencileri Ara sınav ile Genel Sınav Puanlarının Eğitici Yazılım Programını Kullanmalarına Göre T-Testi Sonuçları

Sınav	Grup	N	Ort	SS	Sd	T
Ara Sınav	Yazılım <i>Dağıtılan</i> Öğrenci Grubu (1. Grup)	60	48.88	19.26	118	1.16
	Yazılım <i>Dağıtilmayan</i> Öğrenci Grubu (2. Grup)	60	44.62	20.80		
Genel Sınav	Yazılım <i>Dağıtılan</i> Öğrenci Grubu (1. Grup)	60	41.22	17.61	118	0.22
	Yazılım <i>Dağıtilmayan</i> Öğrenci Grubu (2. Grup)	60	40.38	19.86		

Araştırma kapsamındaki öğrencilerin ara sınav ortalamaları arasında eğitici yazılım kullanıp kullanmamalarına göre farklılık olup olmadığına bakılmıştır. Yazılım dağıtılan öğrencilerin ara sınav not ortalamaları 48.88 iken yazılım dağıtilmayan öğrenci grubunun ara sınav not ortalaması 44.62 olduğu görülmektedir. Yazılım dağıtılan grubun ara sınav değerlendirmesinde daha başarılı olduğu gözlenmiştir. Öğrencilerin genel sınav puan ortalamalarında da yazılım kullanıp kullanmamalarına göre farklılık olup olmadığına bakılmıştır. Yazılım dağıtılan öğrencilerin genel sınav ortalamaları 41.22 iken yazılım dağıtilmayan öğrenci grubunun ortalaması 40.38 olarak bulunmuştur. Bu sonuca birbirine denk olan bu iki öğrenci grubunda, yazılım dağıtılan öğrenci grubunun hem ara sınav hem de genel sınav değerlendirmesinde daha başarılı olduğu görülmüştür.

Elde edilen bu sonuçlardan yola çıkarak; eğitici yazılım kullanan öğrenciler ile eğitici yazılım kullanmadan derse devam eden öğrencilerin ara sınav ve genel sınav ortalamaları arasında bir farklılık gözlenmiştir. Başka bir deyişle eğitici yazılım

öğrencilerin ara ve genel sınav ortalamaları, kullanmayan öğrencilere göre daha yüksek bulunmuştur.

5.3. Yazılımın Öğrenci Görüşleri ile Değerlendirilmesi

Bilgisayar destekli öğretim; yazılım, donanım, öğretmen eğitimi, laboratuvar ve yardımcı personel eğitimi gibi birçok unsuru kendi içerisinde barındırmaktadır. Bu öğeler içinde en fazla dikkat çeken ise ders yazılımı olarak kabul edilmekte ve hatta bilgisayar destekli öğretimin başarısının ders yazılımının kalitesi ile doğrudan orantılı olduğu ileri sürülmektedir. Bilgisayar destekli öğretimin etkinliği büyük ölçüde yazılımın niteliğine bağlıdır. İyi bir yazılım öğrenci başarısını olumlu yönde etkilerken, kötü hazırlanmış bir yazılım zaman kaybına ya da istenmedik davranışların kazanılmasına neden olabilmektedir.

Bu tez çalışması kapsamında, geliştirilen veri madenciliği yazılım aracının etkinliğini belirlemek için öğrenciler üzerinde bir anket çalışması yapılmıştır. Bu amaçla eğitsel yazılım değerlendirmeye yönelik (Kelleci, 2010), (Ateş, 2010) ve (Arslan, 2006) tarafından hazırlanmış anket çalışmaları incelenmiş ve geliştirilen EVMA yazılımına uygun olarak 22 maddelik bir anket hazırlanmıştır. Görüş belirleme amacı taşıdığı için anket üzerinde geçerlik ve güvenirlik çalışmasına gereksinim duyulmamıştır. Anket toplam 22 sorudan oluşmakta ve öğrencilerin maddelere katılma düzeyi 5’li Likert tipi ölçek ile değerlendirilmiştir. Her sorunun değerlendirilmesi için 1 ile 5 puan arasında ağırlıklandırılmış puanlama ile katılımcıların maddelere katılıp katılmadıklarının ölçülmesi hedeflenmiştir. Anketin örneklem grubunu, Fırat Üniversitesi Teknoloji Fakültesi Yazılım Mühendisliğinde Veri Madenciliği dersini alan ve eğitsel veri madenciliği yazılımı kullanan 60 öğrenci oluşturmaktadır. Ancak, bu öğrencilerden sadece 50 öğrenciye ulaşılmış ve 3 anket formunda da eksik veriler olmasından dolayı örneklem sayısı 47 olarak gerçekleşmiştir. Tablo 5.5’te katılımcıların 22 maddelik ankete vermiş olduğu cevaplar neticesinde her madde için kaç öğrencinin hangi cevapları verdiği görülmektedir.

Tablo 5. 3. Hazırlanan anket çalışmasının değerlendirilmesi

Maddeler	Katılmıyorum (1)	Kısmen Katılmıyorum (2)	Çoğunlukla Katılmıyorum (3)	Katılmıyorum (4)	Kesinlikle Katılmıyorum (5)	Ortalama (X)	Standart Sapma (S)
1. Yazılımdaki yönlendirmeler yeterince açık ve anlaşılırdır.	2	7	15	16	7	3,40	1,06
2. Yönlendirmeler kullanımı kolaylaştırıyor.	0	2	17	15	13	3,83	0,89
3. Yazılımın Ara yüzünün kullanımı kolaydır.	3	5	16	13	10	3,47	1,14
4. Yazılımın ekran tasarımı amaca uygundur.	3	6	16	17	5	3,32	1,05
5. Metinler, ekranda kolaylıkla okunabilecek biçimde yer almıştır.	1	5	12	14	15	3,79	1,08
6. Ekranlar ve bölümler arasında rahatça ilerlenebilir.	8	13	14	7	5	2,75	1,22
7. Her ekrandan ana menüye kolaylıkla gidilebilir.	12	9	6	9	11	2,96	1,55
8. Yazılımda kullanılan navigasyon(gezinti mekanizması) kullanışlıdır	6	8	22	5	6	2,94	1,15
9. Ekrandaki vurgulamalar, öğrenme gereksinimleri doğrultusunda düzenlenmiştir.	1	7	14	14	11	3,58	1,08
10. Yazılımda sağlanan yardım menüsü, kullanıcıların ihtiyaçlarına cevap verebilir.	2	8	21	12	4	3,17	0,97
11. Yazılımın kullanımı esnasında çıkan “hata uyarıları” yeterince anlaşılırdır.	9	11	10	11	5	2,83	1,29
12. Yazılım kesintiye uğramaksızın rahatça yönetilebilir.	3	13	16	10	5	3,02	1,09
13. Yazılım, Veri Madenciliği dersinin içeriğine uygundur.	0	1	5	15	26	4,40	0,77
14. Yazılım, dersin uygulamasını daha kolay yapabilmeme imkân tanımıştır.	2	1	10	18	16	3,95	1,02
15. Yazılım, derste işlediğim konuları daha iyi anlayabilmemi sağlamıştır.	1	5	12	18	11	3,70	1,02
16. Yazılım sayesinde algoritmaların ara adımlarını takip etme şansım olmuştur.	1	1	22	16	7	3,58	0,85
17. Yazılım, derste yapılan uygulamalardaki veri setlerine erişimimi kolaylaştırmıştır.	1	3	19	17	7	3,55	0,90
18. Yazılım popüler uygulamaların belgelerini veri seti olarak kullanmamı kolaylaştırmıştır (Excel,Access,Metin Dosyaları.)	4	4	9	13	17	3,75	1,27
19. Yazılım derste yapılan uygulamalarda kullanılan veri setleri üzerinde değişiklikler yapmama olanak sağlamıştır.	1	2	15	18	11	3,77	0,94
20. Yazılım, kurulum gerektirmeden çalıştırılabilmektedir.	1	2	3	10	31	4,45	0,95
21. Yazılım, güncel veri madenciliği uygulamalarını içermektedir.	1	5	9	17	15	3,85	1,06
22. Yazılım, ek program gerektirmeden çalıştırılabilmektedir.	7	3	4	9	24	3,85	1,49

Anketin 2 numaralı maddesi olan “Yönlendirmeler kullanımı kolaylaştırıyor.” Maddesine ankete katılan 47 katılımcıdan 13’ü “Kesinlikle Katılıyorum”, 15’i ise “Katılıyorum” seçeneğini işaretlemişlerdir. Katılımcıların sadece 2 tanesi “Kısmen Katılmıyorum” ve “Katılmıyorum” seçeneğini işaretleyen hiçbir öğrencinin olmadığı görülmektedir. Bu sonuçlar doğrultusunda yapılan ölçme tekniği ile 2. Maddenin Ortalama değerine baktığımızda 3,83 ve standart sapma da 0,89 çıkmıştır. Standart sapmanın düşük olması, katılımcılar arasında verilen cevapların benzerliğinin yüksek olduğunu ve homojenliğe yakın bir cevap kümesi oluştuğunu göstermektedir. Bu maddeye kullanıcıların ortalama olarak 3,89 ile “Katılıyorum” düzeyine yakın görüş bildirdikleri gözlemlenmiştir. Elde edilen bu sonuç hazırlanmış olan yazılımda, kullanılan yönlendirmelerin program kullanımını kolaylaştırdığını göstermektedir.

7. maddede belirtilen “Her ekrandan ana menüye kolaylıkla gidilebilir” ifadesine toplamda 21 öğrenci katılmadıklarını, 20 öğrenci de katıldıklarını beyan etmişlerdir. Eşit dağılıma baktığımızda her ekrandan ana menüye dönüşün kullanıcılar tarafından farklı anlaşıldığı sonucu çıkarılabilir. Yazılımın tam olarak tüm aşamalarının detaylı olarak kullanılmadığı da yine göz önünde bulundurulabilecek bir sonuçtur.

Anketin “Yazılımın kullanımı esnasında çıkan ‘hata uyarıları’ yeterince anlaşılabilir.” olan 11. Maddesi, öğrencilerin yazılımda oluşan hata uyarılarının anlaşılmasının ölçülmesi amaçlanmıştır. Maddenin standart sapması 1.29 çıkmasından dolayı heterojen bir dağılımın olduğu ve yazılımın kullanımında oluşan hataların tam olarak anlaşılmadığı görülmektedir. Bu maddeye 5 öğrencinin “Kesinlikle Katılıyorum” seçeneğini işaretlemeleri olumlu bir sonuç çıkarsa da ortalamanın 2.83 çıkmasından dolayı; katılımcıların bu maddeye katılmadıklarını ve EVMA’nın hata uyarılarının yetersiz olduğu görülmektedir. Anket çalışmasının “Yazılım kesintiye uğramaksızın rahatça yönetilebilir.” şeklinde olan 12. Maddesi de bu sonuçlara yakın çıkmış ve toplamda 16 öğrencinin katılmadıklarını ifade etmişlerdir. Ortalamanın 3.02 çıkması bu maddeye çoğunlukla katıldıklarını ifade etmiştir.

13. Maddede belirtilen “Yazılım, veri madenciliği içeriğine uygundur” ifadesi birçok öğrenci tarafından “Katılıyorum” ve “Kesinlikle Katılıyorum” olarak işaretlemişlerdir. Toplamdan 41 öğrenci katıldığını ifade etmiş, buna karşın sadece 1 öğrenci kısmen katılmadığını belirtmiştir. Bu madde ile hazırlanan yazılımın eğitim içeriği ile örtüştüğü görülmektedir.

Çalışmada 18.maddede “ Yazılım popüler uygulamaların belgelerini veri seti olarak kullanmamı kolaylaştırmıştır(Excel, Access, Metin Dosyaları)” sorusuna verilen cevaplar neticesinde oluşan ortalamanın 3.75 ile yüksek olduğu görülmektedir. Bunun en önemli nedeni, eğitici yazılımın hazırlanması aşamasında veri setlerinin kullanımında popüler programların etkili bir şekilde yazılıma entegre edilmesidir.

20 numaralı maddede belirtilen “Yazılım, kurulum gerektirmeden çalıştırılabilmektedir” ifadesi öğrenciler tarafından iyi anlaşılmış ve öğrencilerin hemen hemen hepsi katıldıklarını ifade etmiştir.

22 maddelik hazırlanan anket çalışmasının yazılımı kullananlar arasında yapılan değerlendirme sonuçlarına bakıldığında, genel olarak hazırlanan eğitici yazılımın, bilgisayar destekli öğretimde kullanımının olumlu karşılandığı görülmektedir. Ayrıca kullanıcılardan alınan geri bildirimlere bakılarak yazılımın eğitimsel olarak daha da geliştirilebileceği sonucu da çıkarılabilmektedir. Hazırlanan veri madenciliği aracının, veri madenciliği dersi için kullanılabilecek etkili bir araç haline getirilmesi ve bu yazılımın aynı zamanda farklı alanlar için de geliştirebileceği görülmüştür. Aynı zamanda bu yazılım aracının ve bu alanda çalışma yapan kullanıcıların kullanımı için de uygun olacağı sonucu çıkarılabilmektedir.

6. SONUÇLAR VE ÖNERİLER

Yapılan tez çalışmasında Veri Madenciliği kullanıcıları için bir yazılım geliştirilmiş ve bu yazılımın öğrenciler üzerindeki katkısı incelenmiştir. Yazılım C# ortamında hazırlanmış olup birçok veri madenciliği algoritmasını barındırmakta ve kullanım için kolay bir arayüz içermektedir. Hazırlanan bu yazılımın, 2015-2016 Bahar yarıyılında Fırat Üniversitesi Teknoloji Fakültesi Yazılım Mühendisliği bölümünde YMT324 Veri Madenciliği dersini alan öğrencilere kullanımı sağlanarak değerlendirilmesi yapılmıştır. Yazılımın öğrencilerin akademik başarısına etkisine yönelik değerlendirmelerde yazılımı kullanan öğrencilerin sınavlarda daha başarılı olduğu görülmüştür. Ayrıca yazılımı kullanan öğrencilerden anket yolu ile alınan görüşlerde de geriye olumlu dönütler alınmıştır. Her iki değerlendirme yöntemiyle elde edilen sonuçlardan, geliştirilen veri madenciliği eğitsel yazılımını kullanan öğrenciler ile kullanmayanlar arasında anlamlı farklılıklar olduğu gözlemlenmiştir.

Yazılımın kurulum gerektirmeden çalışması ve veri seti olarak kullanılan ortamın basit olması, öğrenciler tarafından oldukça olumlu karşılanmıştır. Yazılımın ekran tasarımı alanında başarımının istenen düzeyde olmadığı ve bu konuda yazılımın tekrar güncellenmesi gerekliliği de yine elde edilen sonuçlarda ifade edilmiştir. Karmaşık veri gösterimlerinin olmadığı, yazılımın kullanımının profesyonellik istememesi, veri madenciliği dersinin anlaşılması noktasında yararlı olduğu görülmüştür. Tüm bunlar değerlendirildiğinde; yazılımın başarılı olduğu, istenen sonuçlara ulaşıldığı, eğitsel amaçlı olarak kullanılabileceği ve veri madenciliği dersi için olumlu sonuçlar verebileceği gözlenmiştir.

Her ne kadar elde edilen sonuçlar yeterli görünse de öğrencilerden elde edilen olumsuz görüşlerin değerlendirilmesi ve yazılımın bu görüşlere göre güncellenerek istenen düzeye erişmesi halinde, yazılım hem etkin bir veri madenciliği aracı olarak hem de etkin bir eğitsel yazılım olarak kullanılabilecek bir yazılım olacaktır.

KAYNAKLAR

- Akgöbek, Ö., ve Çakır, F.** (2009). Veri Madenciliğinde Bir Uzman Sistem Tasarımı. *Akademik Bilişim*, 9, 801-806.
- Akkoyunlu, B.** (1996). Bilgisayar okuryazarlığı yeterlilikleri ile mevcut ders programlarının kaynaştırılmasının öğrenci başarı ve tutumlarına etkisi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 12(12).
- Akpınar H.**, (2000). Veri Tabanlarında bilgi keşfi ve Veri Madenciliği, İstanbul Üniversitesi İşletme Fakültesi Dergisi, 29(1), 1-22.
- Akpınar, Y.** (1999). *Bilgisayar destekli Eğitimde uygulamalar*. Anı Yayınları.
- Alessi, S. M. ve Trollip, S. R.** (2001). *Multimedia for Learning: Methods and Development*. USA: Allyn and Bacon.
- Arslan, A.**, (2006). Bilgisayar destekli eğitim yapmaya ilişkin tutum ölçeği, Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi, Van.
- Ateş, A. ,** (2010). Eğitsel yazılımların niteliklerini nasıl ölçebiliriz?: Bir ölçek önerisi. *Proceedings of 10th International Educational Technology Conference (IETC2010), Volume I*, 473-476. Boğaziçi University, İstanbul-TURKEY
- Aydoğan, Ü.**, (2008). Bilişim teknolojilerinin temelleri eğitim içeriğinin interaktif yöntemlerle geliştirilmesi. *Y. Lisans Tezi*. İstanbul: Bahçeşehir Üniversitesi FBE.
- Bilgin, T.T.**, (2009). “Veri Akışı Diyagramları Tabanlı Veri Madenciliği Araçları ve Yazılım Geliştirme Ortamları”, Akademik Bilişim’09 - XI. Akademik Bilişim Konferansı Bildirileri, Şanlıurfa, 807-814.
- Çalışkan, S. Kırmızıgül,** (2008). “Kümeleme ve K En Yakın Komşu Yöntemi ile Ağlarda Nüfuz Tespiti”, Gebze İleri teknoloji Enstitüsü Mühendislik ve Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, GEBZE.
- Çankaya, S., Karamete, A.** (2008). Eğitsel bilgisayar oyunlarının öğrencilerin matematik dersine ve eğitsel bilgisayar oyunlarına yönelik tutumlarına etkisi. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 4(2).
- Demir, Ü.** (2004). *İlköğretim 7. Sınıf Öğrencilerinin Eğitsel Yazılım Ekran Tasarım Seçimlerinin ve Ekran Tasarımında Dikkat Ettikleri Noktaların Değerlendirilmesi*. Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi, İzmir.
- Dempster, A.P., Laird, N.M, Andrubin, D.**, (1977), Maximum Likelihood estimation from incomplete Data Via the EM Algorithm, Journal royal Statistical Soc. Vol: 39, 1–38.

- Delen, D., Walker, G., ve Kadam, A.** (2005). Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial intelligence in medicine*, 34(2), 113-127.
- Dener, M., Dörterler, M., ve Orman, A.** (2009). Açık Kaynak Kodlu Veri Madenciliği Programları: Weka’da Örnek Uygulama. *Akademik Bilişim*, 9, 11-13.
- Doğan B.** (2006). Zeki Öğretim Sistemlerinde Veri Madenciliği Kullanılması Doktora Tezi Marmara Üniversitesi Fen Bilimleri Enstitüsü, Elektronik – Bilgisayar Eğitimi Anabilim Dalı, İstanbul
- Erdemir, U., Tekin, U. ve Buzluca, F.** (2011). *Nesneye Dayalı Yazılım Metrikleri ve Yazılım Kalitesi*. 18th Asia Pacific Software Engineering Conference, Vietnam.
- Gül, Ş., ve Yeşilyurt, S.** (2011). Yapılandırmacı Öğrenme Yaklaşımına Dayalı Bir Ders Yazılımının Hazırlanması Ve Değerlendirilmesi. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi*, 1(40), 19-36.
- Güneş, A.** (2009). Bilgisayar I-II temel bilgisayar becerileri.
- Han, J. ve Kamber, M.** (2000) “Data Mining: Concepts and Techniques”, CA: Morgan Kaufmann, San Francisco, 312s
- Işık M.**, (2006). Bölünmeli Kümeleme Yöntemleri İle Veri Madenciliği Uygulamaları, Yüksek Lisans Tezi Marmara Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Kalkov, A.**, (2006). Veri Madenciliği Ve Bir E-Ticaret Uygulaması, Yüksek Lisans Tezi Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Karamustafaoğlu, O.** (2006). Fen ve teknoloji öğretmenlerinin öğretim materyallerini kullanma düzeyleri: Amasya ili örneği. *Atatürk Üniversitesi Bayburt Eğitim Fakültesi Dergisi*, 1(1), 90-101.
- Kelleci, Ö.** (2010). *Bir Eğitsel Yazılım Değerlendirme Formunun Geliştirilmesi ve Uygulanması*. Yayımlanmamış Yüksek Lisans Tezi, Marmara Üniversitesi, İstanbul.
- Kelleci, Ö., ve Şahinkaya, Y.**, (2012) Developing and evaluating the process of educational software through animated-narrative vignettes for values education. *The Special Issue on Computer and Instructional Technologies*, 130.
- Köksal, A.** (1981) Bilişim Terimleri Sözlüğü, Ankara: TDK Yayıncılık.
- Kurtel, K., ve Eren, Ş.** (2008). Yazılım Ölçümü: Genel Bir Bakış. *Yazılım Kalitesi ve Yazılım Geliştirme Süreçleri Sempozyumu, YKGS*.
- Kutluca, T., ve Birgin, O.** (2007). Doğru Denklemi Konusunda Geliştirilen Bilgisayar. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 27(2).

- Küpçüoğlu, E.**, (2008). Bilişim teknolojileri temelleri eğitiminin ortaöğretimde interaktif yöntemlerle verilmesi. *Y. Lisans Tezi*. İstanbul: Bahçeşehir Üniversitesi FBE.
- Mardani, A., Bagheri, M. M., Soltan, E. K. H., & Lari, M. K.** (2013). Relationship between National Culture and Soft Total Quality Management in Iranian Multinational Firms. *International Journal of Basic and Applied Science*,2(01), 160-172.
- McConnell, S.** (1996). *Rapid Development: Taming Wild Software Schedules*. Redmond: MS Press.
- McDaniel, J. G.** (2002). Improving System Quality Through Software Evaluation. *Computers in Biology and Medicine*, 32, 127-140.
- Mitra, S., ve Acharya, T.** (2005). *Data mining: multimedia, soft computing, and bioinformatics*. John Wiley ve Sons.
- Odabaşı, F.** (1998). Bilgisayar destekli eğitim. *Editör: Yaşar Hoşcan, Açıköğretim Fakültesi İlköğretim Öğretmenliği Lisans Tamamlama Programı, Eskişehir*.
- Özkan Y.**, (2013), “Veri Madenciliği Yöntemleri”, Papatya Yayınları, İstanbul 216s.
- Pressman, R. S.** (2001). Software Quality Assurance. In R. S. Pressman (Ed.), *Software Engineering: A Practitioner’s Approach*(Vol. 5, pp. 193-225). New York: McGraw-Hill.
- Ren, J., Lee, S. D., Chen, X., Kao, B., Cheng, R., & Cheung, D.** (2009, December). Naive bayes classification of uncertain data. In *2009 Ninth IEEE International Conference on Data Mining* (pp. 944-949). IEEE.
- Savran, Y.**, (2009) Çok Değişkenli Gauss Karışım Modelleme Yöntemiyle Görsel Doku Analizi Yüksek Lisan Tezi, İstanbul Teknik Üniversitesi İstanbul
- Seidman C.** (2001). "DataMiningwith Microsoft® SQL Server™ 2000 Technical Reference", *Microsoft Press*, Redmond, WA, USA.
- Shiri, M. E., Aïmeur, A. E., ve Frasson, C.** (1998, August). Student modelling by case based Reasoning. In *International Conference on Intelligent Tutoring Systems* (pp. 394-403). Springer Berlin Heidelberg.
- Şen, B.**, (2006). Meslek liselerinin atölye derslerindeki elektronik devre uygulamalarının bilgisayar destekli açık uçlu deney yazılımları kullanılarak gerçekleştirilmesinin akademik başarıya ve kalıcılığa etkisi. *Y. Lisans Tezi*. Adana: Çukurova Üniversitesi SBE.
- Thuraisingham, B.** (2003). *Web data mining and applications in business intelligence and counter-terrorism*. CRC Press.

- Türker, F. M., Cihan, A., Şahin, A.** (2015). Web Tabanlı Uzaktan Türkçe Öğretimi Uygulamalarının Yazılı Anlatım Dersinde Başarıya Etkisi. *İstanbul Açık ve Uzaktan Eğitim Dergisi (AUZED)*, 1(2).
- URL-1,** WEKA, <http://www.cs.waikato.ac.nz/ml/weka/>, (Erişim Tarihi: 10.07.2016).
- URL-2,** KNIME, <http://www.knime.org/>, (Erişim Tarihi: 10.07.2016).
- URL-3,** KEEL, <http://www.keel.es/>, (Erişim Tarihi: 11.07.2016).
- URL-4,** ORANGE, <http://orange.biolab.si/>, (Erişim Tarihi: 11.07.2016).
- Varol, N.** (1997). Bilgisayar Destekli Eğitimi. Türk Cumhuriyetleri ve Asya Pasifik Ülkeleri Uluslar arası Eğitim Sempozyumu, 138-145.
- Witten, I. H., Frank, E,** (2005). “Data Mining Practical Machine Learning Tools and Techniques,” Morgan Kaufmann, Second Edition, San Fransisco, (2005).
- Yeşilova, A., Kaki, B., Şen, C,** (2004). Poisson Regresyonunda meydana Gelen Aşırı Yayılımın İncelenmesi, IV. Ulusal Zootekni Bilim Kongresi, 633-637, Isparta, Türkiye
- Yıldırım Ö.,** (2010) “Veri Madenciliği Yöntemleriyle Depremlerin Analizi”, Yüksek Lisans Tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Elazığ 2010.
- Yiğit, N., ve Akdeniz, A. R.** (2003). Fizik Öğretiminde Bilgisayar Destekli Etkinliklerin Öğrenci Kazanımları Üzerine Etkisi Elektrik Devreleri Örneği. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 23(3).

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı: Yusuf ÇELİK

Doğum Yeri ve Tarihi: Bingöl-1983

Eğitim Durumu

Lisans Öğrenimi: Fırat Üniversitesi Bilgisayar Mühendisliği

Deneyim

Çalıştığı Kurumlar ve Yıl:

- Tunceli Üniversitesi-6 yıl (Devam)
- Hastane Bilgi Yönetim Sistemleri -3Yıl

İletişim

E-posta Adresi: yusufcel@gmail.com