

T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİMDALI
BİLGİSAYAR MÜHENDİSLİĞİ BİLİMDALI

***DERSHANE EĞİTİMİNİN ÜNİVERSİTEYE YERLEŞMEDEKİ ETKİSİNİN
VERİ MADENCİLİĞİ İLE İRDELENMESİ***

Yüksek Lisans Tezi

BÜNYAMİN HATİPOĞLU

Danışmanı

Prof. Dr. Zafer ASLAN

İstanbul, 2013



T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ
MÜDÜRLÜĞÜ'NE

Tez Onay Belgesi

Enstitümüz Bilgisayar Mühendisliği Anabilim Dalı, Bilgisayar Mühendisliği (Tezli) Yüksek Lisans Programı Y1113.010005 numaralı öğrencisi **Bünyamin HATİPOĞLU**' nun "**DERSHANE EĞİTİMİNİN ÜNİVERSİTEYE YERLEŞMEDEKİ ETKİSİNİN VERİ MADENCİLİĞİ İLE İRDELENMESİ**" adlı tez çalışması Enstitümüz Yönetim Kurulunun 18.02.2013 tarih ve 2013/05 sayılı kararıyla oluşturulan jüri tarafından **oybirliği / oyçokluğu** ile Yüksek Lisans Tezi olarak **kabul** edilmiştir.

Öğretim Üyesi Adı Soyadı

İmzası

Tez Savunma Tarihi : 18.06.2013

- 1) Tez Danışmanı : Prof. Dr. Zafer ASLAN
2) Jüri Üyesi : Prof. Dr. Ali GÜNEŞ
3) Jüri Üyesi : Yrd. Doç. Dr. Metin ZONTUL

Zafer Aslan
Ali Güneş
Metin Zontul

Not: Öğrencinin Tez savunmasında **Başarılı** olması halinde bu form **imzalanacaktır**. Aksi halde geçersizdir.

ÖNSÖZ

Kariyerimin gelişmesinde emeđi olan tüm hocalarıma, Kùltür Dershaneleri rehberlik koordinatörü ve tv24 kanalında eğitim atölyesi adlı programın yorumcusu Sn. Salim Ünsal'a, ayrıca verilerin sağlanmasında katkısı olan Kùltür Dershaneleri rehberlik personeli Sn. Fatih Yılmaz'a...

Tüm Kalbimle Teşekkürlerimi Sunuyorum.

ÖZET

Bu çalışmanın amacı farklı veri madenciliği modelleri kullanarak, öğrencinin dershanede aldığı eğitim programına göre katılım süresi, branş dersleri parametrelerine göre üniversiteye yerleşme durumlarını irdelemek, sonuç ve önerilerde bulunmaktır.

Bu amaçla, son yıllarda geniş kullanım alanı olan “veri madenciliği” yöntemlerinden yararlanılmaktadır. Veriler genel olarak ve ayrı ayrı bazı kümeler halinde incelenmekte ve gerekli sonuçlar çıkarılmaktadır.

Birinci bölümde tez’ in amacından ve literatürde bu tip bir çalışma üzerine yapılan araştırmalar hakkında bilgi verilmektedir.

İkinci bölümde dershanelerin eğitim öğretim sistemindeki yerinden bahsedilmektedir.

Üçüncü bölümde veri madenciliği hakkında detaylı bilgiye, kullanılan algoritmalara ve yapılan çalışmalara yer verilmiştir.

Dördüncü bölümde tez çalışmasında kullanılan uygulama araçlarından bahsedilmiştir.

Beşinci bölümde uygulama sonucu elde edilen bilgiler analiz edilmiş ve elde edilen bulgular, sonuç ve önerilerde bulunmaya önemli katkı sağlamıştır.

Tezin sonuç bölümünde veri madenciliğinin kullanım gereksiniminden, dershanelerin eğitim öğretim sistemindeki durumu ve öğrencilerin üniversiteye yerleşme durumuna etki eden faktörlerden bahsedilmekte ve hangi modellerin hangi alanlarda kullanılabileceği önerilmektedir.

Anahtar Sözcükler:

Veri Madenciliği, Algoritma, Dershane, Üniversiteye Yerleşme Durumu, Literatür

ABSTRACT

The purpose of this study, using different data mining models, according to the student's classroom participation in the training program period, to examine the status of industry to the university courses according to the parameters, results and make recommendations.

To this end, in recent years, the wide range of “data mining” methods is used. In general and some of the data into clusters and necessary results separately extracted.

In the first chapter the aim of the thesis and the literature provides information on the research on this type of study.

In the second part from tutoring, education and training system are discussed.

The third section, detailed information about the data mining, algorithms, and algorithms used in the studies are presented.

In this thesis discussed the application tools used in the fourth chapter.

Information obtained from the fifth chapter, the application has been analyzed and the findings, conclusions and offer suggestions, made a significant contribution.

The need to use data mining results section of the thesis, tutoring students to the university education system status, and which models are introduced and the factors that affect the areas in which the state proposed to be used.

Keywords:

Data Mining, Algorithm, Private/Foundation Education Institutions, University Enrollment, Literature

İÇİNDEKİLER

ÖNSÖZ	i
ÖZET	ii
ABSTRACT	iii
İÇİNDEKİLER	iv
TABLO LİSTESİ	vi
GRAFİK LİSTESİ	vii
ŞEKİL LİSTESİ	viii
KISALTMA LİSTESİ	x
1. GİRİŞ	1
1.1 TEZ AMACI	1
1.2 LİTERATÜR ARAŞTIRMASI	1
2. PROBLEMİM TANIMLANMASI	12
2.1 EĞİTİM VE ÖĞRETİM SİSTEMİNDE DERSHANELERİN YERİ	12
2.2 DERSHANELERİN KURULUŞ AMACI	12
2.3 DERSHANELERDE EĞİTİM ÖĞRETİM SÜRECİ	13
2.4 DERSHANELERDE TEKNOLOJİNİN YERİ	15
2.5 SINAV VE NOT KAYGISI	16
2.6 DERSHANELERDE VERİ MADENCİLİĞİ İHTİYACI	16
3. VERİ MADENCİLİĞİ	17
3.1 VERİ MADENCİLİĞİ TARİHİ	17
3.2 VERİ MADENCİLİĞİNİN KULLANIM GEREKSİNİMİ	18
3.3 VERİ MADENCİLİĞİNDE KULLANILAN ADIMLAR	18
3.4 VERİ AMBARI	20
3.5 VERİ MADENCİLİĞİ MODELLERİ	21
3.5.1 Sınıflama ve regresyon	22
3.5.1.1 Karar Ağaçları	22
3.5.1.2 Bayes Sınıflandırıcıları	25

3.5.1.3 Yapay Sinir Ağları	26
3.5.2 Association Rules – Birliktelik Kuralları	27
3.5.2.1 Apriori Algoritması	27
3.5.3 Kümeleme modelleri	28
3.5.3.1 K-Means (K- Ortalama) Algoritması	29
3.5.3.2 Benzerlik ve Mesafenin Ölçülmesi.....	30
3.5.3.3 Hiyerarşik Modeller	30
3.5.3.4 Bölümlenmeli Modeller	31
4. UYGULAMA ARAÇLARI	32
4.1 MICROSOFT SQL SERVER	32
4.2 MICROSOFT ANALYSSERVICES	33
4.3 BUSINESS INTELLIGENCE DEVELOPMENTSTUDIO	33
5. ANALİZ	34
5.1 KARAR AĞAÇLARI.....	34
5.2 BAYES SINIFLANDIRICILARI.....	35
5.3 BİRLİKTELİK KURALLARI.....	36
5.3.1 Apriori Algoritması	45
5.4 KÜMELEME MODELLERİ	47
5.4.1 K-Ortalama Algoritması	48
5.5 BIDS MODEL KARŞILAŞTIRMA.....	49
6. SONUÇ VE ÖNERİLER	60
KAYNAKÇA.....	63
Yapay Sinir Ağları	70
EKLER	71
ÖZGEÇMİŞ.....	78

TABLO LİSTESİ

Tablo 5.1 Lisans Bölümlerine Yerleşmede Birliktelik Kurallarının Etkisi	36
Tablo 5.2 Ön Lisans Bölümlerine Yerleşmede Birliktelik Kurallarının Etkisi ..	38
Tablo 5.3 Herhangi Bir Bölüme Yerleşememede Birliktelik Kurallarının Etkisi	39
Tablo 5.4 Herhangi Bir Tercih Yapmamada Birliktelik Kurallarının Etkisi	41

GRAFİK LİSTESİ

Grafik 3.1 Ayrık Veri Örneği	23
Grafik 3.2 Ayrık Veri Sonucu	24
Grafik 3.3 Sürekli Veri Örneği.....	24
Grafik 5.1 Katılım Sayısı Yoğunluğu	47
Grafik 5.2 Tahmini Lisans Yerleşme Grafiği (Kırmızı Çizgi: İdeal Durum,.....	58

ŞEKİL LİSTESİ

Şekil 3.1 Veri Bütünleştirmeye Ait Bir Örnek	19
Şekil 3.2 Veri Madenciliğinde Kullanılan Adımlar	20
Şekil 3.3 Yaş, Mesafe ve Eğitim Durumuna Göre Bisiklet Satın Alma ya da Satın Almama Durumuna Göre Sınıflandırma (0 – Satın Almama, 1 – Satın Alma).....	25
Şekil 3.4 YSA Katmanlarının Gösterimi	26
Şekil 3.5 Apriori Algoritması Pseudo-Code.....	28
Şekil 5.1 Dershaneye Katılım Süresinin Karar Ağacı Modeli İle Belirlenmesi	34
Şekil 5.2 “12. Sınıf” ve 2+ Yıl Dershaneye Katılan Öğrencilerin Durumu	35
Şekil 5.3 Katılım Süresi ve Grubun Bayes Sınıflandırma Sonucu Yerleşme Durumuna Etkisi	36
Şekil 5.4 Lisans Bölümlerine Yerleşmede Birliktelik Kurallarının Ağırlıklı Etkisi	37
Şekil 5.5 Ön Lisans Bölümlerine Yerleşmede Birliktelik Kurallarının Ağırlıklı Etkisi	39
Şekil 5.6 Herhangi Bir Bölüme Yerleşememede Birliktelik Kurallarının Ağırlıklı Etkisi	40
Şekil 5.7 Herhangi Bir Tercih Yapmamada Birliktelik Kurallarının Ağırlıklı Etkisi	41
Şekil 5.8 Yerleşme Durumuna Etki Eden Birliktelik Kurallarının Güven Seviyeleri	43
Şekil 5.9 Yerleşme Durumuna Etki Eden Kombinasyonların Destek Sayıları	44
Şekil 5.10 Apriori Algoritması Veri Destek Sayısı ve Güven Seviyeleri	45
Şekil 5.11 Öğrencilerin Sorumlu Olduğu Branş Ders ve Katılım Sayıları	46
Şekil 5.12 Katılım ve Gruba Göre Yerleşme Durumunu Tahmine Dayalı Kümeleme Modeli.....	47

Şekil 5.13 K-Ortalama Algoritması Küme Değerleri Ortalamaları	48
Şekil 5.14 VM Algoritma Belirlenmesi	49
Şekil 5.15 Veri Kaynağının Belirlenmesi	50
Şekil 5.16 Ana Tablo ya da Yardımcı Tabloların Belirlenmesi	51
Şekil 5.17 Algoritma İçin Sütunların Belirlenmesi	52
Şekil 5.18 Algoritma İçin Sütunların Tahmin Edilmesi.....	53
Şekil 5.19 Sütun Tiplerinin Belirlenmesi.....	54
Şekil 5.20 Test Oranının Belirlenmesi.....	55
Şekil 5.21 Yeni Model Ekleme.....	56
Şekil 5.22 Uygulamaya Yeni Model Ekleme	56
Şekil 5.23 Modelin Eğitilmesi ve Çalıştırılması	57
Şekil 5.24 Lisans Yerleşme Tahmin Oranları (Kırmızı Çizgi: İdeal Durum, ...	59

KISALTMA LİSTESİ

BIDS	:	Business Intelligence Development Studio
BT	:	Bilişim Teknolojileri
k	:	Küme Sayısı
KDD	:	Knowledge Discovery From Databases
MDT	:	Microsoft Decision Trees
MNB	:	Microsoft Naive Bayes
MS-SQL	:	Microsoft Structured Query Language
n	:	Eleman Sayısı
SQL	:	Structured Query Language
VBK	:	Veri Tabanlarında Bilgi Keşfi
VLDB	:	Very Large Databases
VM	:	Veri Madenciliği
YSA	:	Yapay Sinir Ağları

1. GİRİŞ

1.1 TEZ AMACI

Bu çalışmada amaç, son yıllarda geniş kullanım alanına sahip olan veri madenciliğinden ve veri madenciliği yöntemlerinden yararlanarak, öğrencilerin dershanede aldığı eğitim programına göre katılım süresi, branş dersleri parametrelerini baz alarak üniversiteye yerleşme durumlarını irdelemek, sonuç ve önerilerde bulunmaktır.

1.2 LİTERATÜR ARAŞTIRMASI

Veri Madenciliği (VM) yaygın bir kullanım alanına sahiptir. Literatürde VM kullanılarak müşteri davranışlarının incelendiği birçok çalışma mevcuttur. Ülkemizde de birçok alanda Veri Madenciliği uygulamaları mevcuttur. Eğitim sektöründe öğrenciler üzerine yapılmış birkaç VM çalışmasına bu tez çalışmasında yer verilmiştir.

Günümüzde devlet okullarında e-okul sistemi ile öğrenci performansları kayıt altına alınmaktadır. VM yöntemi kullanılarak bu alanda yaygın bir çalışma yapılabileceği gösterilmektedir. Bu çalışmanın yapılacak benzer çalışmalar arasında önemli bir kaynak oluşturması beklenmektedir.

Mağazaların alışveriş kayıtlarının tutulduğu veri tabanlarını kullanarak müşterilerin satın alma davranışlarını belirleyebilmek için Chen ve grubu 2005 yılında bir sepet analizi uygulaması yapmışlardır.

Analiz sonucunda elde edilen bilgiler pazarlama, satış ve operasyon stratejilerinin şekillendirilmesi açısından önem taşımaktadır. Burada sunulan yüksek lisans tezinde Chen ve grubunun çalışmalarında kullanılan sepet analizine benzer çalışmaya yer verilmiştir.

Vindervogel ve grubu (2005) çalışmalarında büyük miktardaki ürün çiftleri arasındaki bağıntıyı hesaplayabilmek için sepet analizi kullanmış, promosyon stratejilerini optimize etmek için yeni bir öneri getirmiştir. Araştırmada birlikte satın alınan ürünlerden sadece bir tanesinin fiyatının düşürülmesi ile etkin sonuçlar elde edileceği ileri sürülmüştür. Öğrencilerin gösterdiği performansların başarılı oldukları derslerle birlikte yorumlanmasında bu makaleden yararlanılmıştır.

Timor ve Şimşek “Veri Madenciliğinde Sepet Analizi İle Tüketici Davranışı Modellemesi” adlı çalışmada literatürden ve veri madenciliğinde kullanılan modellerden bahsetmektedir.

Burada sunulan yüksek lisans tezinde bu çalışmada kullanılan birliktelik kurallarına benzer olarak, veriler arasındaki potansiyel ilişkiler tespit edilmiştir.

Journal of Aeronautics and Space Technologies dergisinde yayınlanan “A Data Mining Application in a Student Database” adlı makalede öğrencilerin üniversite giriş sınav sonucu ile yerleştikleri bölümler arasındaki ilişki, kümeleme analizi tekniklerinden K-Means algoritması kullanılarak analiz edilmiştir (Erdoğan ve Timor, 2005).

Sonuç olarak VM tekniklerinin eğitimde kullanılmasının bize çok özel bulgular sağlayacağı ve eğitimde kalitenin artırılmasında kullanılabileceği anlaşılmıştır.

International Journal of Computer Science and Information Security adlı dergide yayınlanan “Application of K-Means Clustering algorithm for prediction of Students’ Academic Performance” adlı makalede akademik planlayıcılar tarafından efektif karar vermelerine yardımcı olmak amacıyla öğrencilerin akademik performansları üzerine K-Means kümeleme algoritması kullanılarak analiz çalışması yapılmıştır (Oyelade ve diğerleri, 2010).

Nijerya’da bir üniversitede bir dönemdeki 79 öğrencinin akademik performanslarından oluşan veriler üzerinde kümeleme yöntemi uygulanmıştır. Bu çalışma akademik planlayıcıların eğitim sürecinde öğrencilerin performanslarını izlemelerine yardımcı olmakta ve ona göre akademik planlama yapmalarına imkân sağlayacağı düşünülmektedir.

2009 yılında, Technical Education of Marmara University de yayınlanan “Data Mining Application on Students’ Data” adlı makalede İstanbul Eyüp İ.M.K.B. Ticaret Lisesinde eğitim gören öğrencilerin 9., 10. Ve 11. Sınıflarda başarısız oldukları dersler arasındaki ilişki VM yöntemi kullanılarak Apriori algoritması ile ortaya konulmuştur (Baldu ve Üçgün, 2009).

Bu çalışmada kullanılan Apriori algoritmasından faydalanılarak öğrencilerin sorumlu olduğu branş derslerindeki başarıları arasındaki ilişki ortaya çıkarılmıştır.

Savaş ve diğerleri tarafından 2012 yılında İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisinde, “Veri Madenciliği ve Türkiye’deki Uygulama Örnekleri” adlı makalede veri madenciliğinin günümüz disiplinleri arasında geldiği noktaya değinilmiş ve Türkiye’de veri madenciliği üzerine yapılan çalışmalar ve gerçekleştirilen uygulamalar incelenmiştir. İncelenen çalışmalar da göstermektedir ki kurum ve kuruluşların çoğu müşteri/kullanıcı analizlerine yönelmiştir.

2011 yılında, Australasian Journal of Educational Technology adlı dergide “Data Mining Techniques for Identifying Students at Risk of Failing a Computer Proficiency Test Required for Graduation” adlı makalede, Tayvan’da bir devlet üniversitesinde bilgisayar dersinden sınava giren öğrenciler üzerine veri madenciliği teknikleri uygulanmıştır (Tsai ve diğerleri, 2011). Sınavdan başarısız olan öğrencilerin yerleşim yerlerine göre analizler çıkarılmıştır. Bu yerleşim yerlerinin öğrencilerin başarısızlığına etkisinin olup olmadığı kümeleme yöntemi ile irdelenmiştir.

Özçınar tarafından 2006 yılında “KPSS Sonuçlarının Veri Madenciliği Yöntemleri ile Tahmin Edilmesi” adlı yüksek lisans tez çalışmasında, materyal olarak sınıf öğretmenliği A.B.D. öğrencilerinin lisans eğitimleri süresince bazı derslerden aldıkları ders geçme notları, genel not ortalamaları, öğretim türleri ve KPSS puanları kullanılmıştır.

Çalışmada ilk olarak toplanan veriler temizlenip birleştirilmiş ve veri madenciliği uygulamasında kullanılacak şekilde düzenlenmiştir. Modelleme aşamasında tahmin doğruluklarının karşılaştırılabilmesi için yapay sinir ağı ve regresyon modelleri oluşturulmuştur. Yapay sinir ağı modelini oluşturmak için öğrenme yöntemi olarak geriye yayılım algoritmasını kullanan çok katmanlı perseptron kullanılmıştır. Regresyon modelini oluşturmak için çoklu doğrusal regresyon yöntemi kullanılmıştır. Frekans analizi yöntemiyle veri kümesinin özellikleri belirlenmiştir. Oluşturulan regresyon modeli ile KPSS sonuçlarının değişimi üzerinde anlamlı katkısı olan değişkenler incelenmiş ve oluşturulan modellerin tahmin doğrulukları, ortalama mutlak hata ve ortalama hata kareler kökü değerleri kullanılarak karşılaştırılmıştır.

”İnternet Tabanlı Öğretimde Veri Madenciliği Tekniklerinin Uygulanması” adlı yüksek lisans tez çalışmasında internet tabanlı bir öğretim sistemi tasarlanmıştır (Altıntop, 2006). Uygulama kapsamında tasarlanan site, Kocaeli Üniversitesi Sağlık Yüksekokulu ve Arslanbey Meslek Yüksekokulu öğrencilerinin eğitiminde uygulamalı olarak kullanılmıştır.

Öğrencilerin site üzerindeki hareketleri bir veri tabanında tutulmuştur. Öğrencilerden toplanan bu veriler üzerinde, veri madenciliği modellerinden birliktelik kuralları ve kümeleme yöntemleri kullanılmıştır. Birliktelik kuralları algoritmalarından Apriori algoritması kullanılarak, tasarlanan site üzerinde en çok bağlanılan sayfa çiftleri ve yayınlanan değerlendirme sorularında en sık yanlış cevaplanan soru çiftleri keşfedilmektedir. Kümeleme tekniklerinden DBSCAN algoritması uygulanarak, öğrenciler ödev ve sınav notlarına göre gruplandırılmıştır.

Oluşturulan bu yapı sayesinde yayınlanan ders notlarının; öğrencinin amaçlarına, bilgi düzeyine ve öğrenme metoduna uyarlanmış bir düzene getirilebilmesi, öğrenci ve öğretim elemanı performansını arttırıcı yönde kullanabilmek amaçlanmıştır.

Sonuç olarak, veri madenciliği tekniklerinin, İnternet tabanlı öğretimin geliştirilmesinde kullanılabileceği görülmüştür.

“Veri Madenciliğinde K-Means Algoritmasının İncelenmesi ve Uygulanması” adlı lisans tez çalışmasında Matlab programı kullanılarak farklı sentetik veriler üzerinde K-Means algoritmasının incelenerek bu verilerin birbiriyle karşılaştırılmasını amaçlanmıştır (Kalaycı, 2008).

“Veri Madenciliği ile Birliktelik Kurallarının Bulunması” adlı yüksek lisans tez çalışmasında, veri tabanlarında bilgi keşfi süreçleri, veri madenciliği, veri madenciliğinde kullanılan birliktelik-ilişki kuralı ve Apriori algoritması hakkında bilgiler verilmiştir (Şen, 2008).

Uygulama bölümünde, gerçek veriler kullanarak Birliktelik Kuralları yöntemi ile Pazar Sepeti Çözümlemesi uygulaması yapılmış ve elde edilen sonuçlar tartışılmıştır.

“Veri Madenciliğinde Kullanılan Teknikler ve Bir Uygulama” adlı yüksek lisans tez çalışmasında, son yıllarda geniş kullanım alanı olan veri madenciliğinde kullanılan tekniklerini tanıtmak ve Satışlara yönelik analizler yapmaktır (Özdamar, 2002). Bu bağlamda kümeleme, yapay sinir ağları ve karar ağaçları kullanılarak sınıflandırma ve Treemap uygulaması yapılmıştır.

“Bölünmeli Kümeleme Yöntemleri ile Veri Madenciliği Uygulamaları” adlı yüksek lisans tez çalışmasında, öncelikle veri madenciliği ve kümeleme analizi hakkında genel bilgiler verildi (Işık, 2006). Daha sonra, bölünmeli kümeleme algoritmaları hakkında ayrıntılı teorik bilgiler verilip, bu bilgilerin ışığında kümeleme algoritmalarının kolaylıkla uygulanması ve görsel olarak yorumlanması için MATLAB ortamında program geliştirildi.

K-Means ve fuzzy c-means algoritmalarının web dokümanları üzerinde kümeleme başarıları karşılaştırılmalı olarak incelendi. Yapılan testlerde fuzzy c-means algoritmasının hem öteleme sayılarının fazla çıkması hem de daha karmaşık formüller içermesi nedeniyle k-means algoritmasından daha geç

sonuç ürettiği görülmüştür. Bununla birlikte, kümeleme işleminde daha düşük hata oranlarına neden olarak daha başarılı sonuçlar ürettiği gözlenmiştir.

Algoritmaların test aşamalarında veri kümelerinin büyüklüğünden ve algoritmaların hızlarından dolayı bellek ve zamanla ilgili çeşitli performans sorunları yaşanmıştır. Bu nedenle, bu gibi testleri yapmak için yüksek performanslı bilgisayarlar gerekmektedir. Ayrıca performans açısından MATLAB yerine daha hızlı programlama dilleri de seçilebilir.

“Veri Madenciliğinde K-Means Algoritması ve Tıp Alanında Uygulanması” adlı yüksek lisans tez çalışmasında, amaç veri madenciliğinde bir kümeleme tekniği olan k-means algoritmasını incelemek ve bu algoritmayı kullanarak geliştirilen bir yazılım aracılığı ile gırtlak kanseri ameliyat verilerinin analizini yapmaktır. Uygulamanın tıp doktorlarının kullanımına uygun şekilde verileri çeşitli açılardan analiz etmesi hedeflenmiştir (Dinçer, 2006).

Bu çalışma ile geçmiş verileri analiz ederken değişken parametreler kullanılarak değerlendirme yapılabilir. Tüm durumlar için mevcut ve gelecek vakalarla ilgili tahminde bulunulabilir, mevcut ve gelecek vakalar için ameliyat sonrasında tümörün nüksetme olasılığı ve hastanın hayatta kalma olasılığı değerlendirilebilir. Doğru öngörülen ameliyat öncesi evreler görüntülenerek incelenebilir ve bu şekilde ameliyat öncesi tahmin başarısı değerlendirilebilir, başarılı ameliyat bilgileri izlenerek, gelecek ameliyat tercihlerinde fikir alınabilir. Ayrıca yazılım, araştırma, denetim ve eğitim etkinliklerinde de kullanılabilir.

“Hiyerarşik Kümeleme Metotları ile Veri Madenciliği Uygulamaları” adlı yüksek lisans tez çalışmasında veri madenciliği ve veri madenciliğinde kullanılan kümeleme analizi metotları ve bu metotlardan birisi olan hiyerarşik kümeleme algoritmaları hakkında teorik bilgiler verilmiştir. Daha sonra hiyerarşik kümeleme algoritmalarından CURE (Clustering Using Representatives) ve AGNES (Agglomerative Nesting) ile bölümleyici kümeleme algoritmalarından k-means algoritmasının yapıları ayrıntılı olarak incelenmiş ve bu algoritmalar MATLAB 'da hazırlanmış bir program aracılığıyla sentetik veri setleri üzerinde uygulanmıştır (Demiralay, 2005).

Elde edilen sonuçların karşılaştırmaları yapılmıştır. Algoritmaların gerçek veri setleri üzerindeki sonuçlarının değerlendirilmesini sağlamak için süsen bitkisinin taç ve çanak yapraklarının büyüklükleri bilgilerini taşıyan iris veri setinde de uygulamalar gerçekleştirilmiştir.

“Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi” adlı doktora tez çalışmasında, Öncelikle veri madenciliği kavramı açıklanarak veri madenciliğinin kullanım amaçları ve kullanım alanları hakkında bilgiler verilmiş ve daha sonra veri madenciliğinde kullanılan kümeleme algoritmaları teorik bir çerçevede açıklanmaya çalışılmıştır (Akın, 2008).

Uygulamada TÜİK tarafından 2004 yılında yapılan ‘Hane halkı Bütçe Anketi’ çalışmasından derlenen veriler kullanılarak hem tüketici davranış kalıpları belirlenmeye çalışılmış hem de uygulamada kullanılan merkeze dayalı bölümleyici kümeleme algoritması ile yoğunluk tabanlı kümeleme algoritması sonuçları karşılaştırılmıştır.

“Veri Madenciliği ve Apriori Algoritması ile Süpermarket Analizi” adlı yüksek lisans tez çalışmasında, veri madenciliği hakkında temel bilgiler verilerek, birliktelik kuralları ve birliktelik kurallarının en temel algoritmalarından biri olan Apriori algoritması detaylı olarak incelenmiştir (Ay, 2009). Bir veri madenciliği programı aracılığıyla, Apriori algoritması kullanılarak Migros Türk A.S. verileri ile market sepet analizi yapılmıştır. Yapılan bu çalışma sonunda birlikte satılma eğilimi gösteren ürünler hakkında bilgiler verilerek, yeni bir market yerleşim düzeni önerilmiştir.

“Orta Öğretim Okulları İçin Öğrenci Otomasyonu Tasarımı ve Öğrenci Verileri Üzerine Veri Madenciliği Uygulamaları” adlı yüksek lisans tez çalışmasında, İ.M.K.B. Ticaret Meslek Lisesi öğrenci verileri üzerinde yapılan bir VM çalışması ile ilgili bilgiler verilmiştir (Üçgün, 2009). Veri tabanındaki veriler üzerinde Apriori algoritması uygulanarak her aşaması kullanıcı tarafından gözlemlenmiştir. Elde edilen sonuçlardan öğrencilerin başarısız olduğu dersler arasındaki ilişkiler ortaya çıkarılmıştır.

“Veri Ambarı ve Veri Madenciliği Teknikleri Kullanılarak Öğrenci Karar Destek Sistemi Oluşturma” adlı yüksek lisans tez çalışmasında, ilk olarak Pamukkale Üniversitesi öğrenci verileri üzerinde bir veri ambarı oluşturulmuştur (Gülçe, 2010). Ardından Tahmin işlemlerini yapabilmek için veri madenciliği algoritmaları olan karar ağaçları, yapay sinir ağları, naive bayes ve birliktelik kurallarından faydalanılmıştır.

Bu algoritmaların sonuçlarına göre bir raporlama ara yüzünden elde edilen sonuçlar karar vericilerin kullanımına sunulmuştur. Bu çalışmada kullanılan VM algoritmalarından ve “Business Intelligence Development Studio” adlı VM çalışmalarına imkân sağlayan bir Microsoft ürününden faydalanılmıştır.

“Kız Meslek Lisesi Öğrencilerinin Akademik Başarısızlık Nedenlerinin Veri Madenciliği Tekniği ile Analizi” adlı yüksek lisans tez çalışmasında, öğrencilerin başarısızlıklarına neden olan etkenler incelenmiştir. Başarısızlık nedenlerini belirlemek için öğrencilere bir anket uygulanmıştır (Birtıl, 2011).

Anketin değerlendirilmesi aşamasında, verilerin içeriği ve çalışmanın konusu göz önüne alınarak veri madenciliği yöntemlerinden kümeleme yöntemi uygulanmıştır ve veriler üç kümeye ayrılmıştır. Her kümenin incelenmesi sonucunda, öğrencileri başarısızlığa iten etkenlerin hangilerinin aynı anda görüldüğü ve aralarındaki ilişkiler tespit edilmiştir.

“Veri Madenciliğinde Regresyon Ağaçları ile Sınıflandırma ve Bir Uygulama” adlı doktora tez çalışmasında, veri madenciliği süreci, verilerin hazırlanması, nitelik seçimi, sınıflandırma, modelin değerlendirilmesi konuları açıklanmıştır (Dondurmacı, 2011).

Öğrenme kavramı, karar ağaçlarının elde edilme süreci, karar kuralları ve karar ağaçlarında entropiye dayalı bölünme, regresyon ağaçları ile sınıflandırma esasları ele alınmıştır. Son olarak IMKB 30 grubuna dahil menkul kıymetlere ilişkin günlük kapanış fiyatları, menkul kıymet teknik analizlerinde yaygın biçimde kullanılan teknik göstergeler, altın fiyatlarındaki değişimler, dolar kurundaki değişimler ve bazı yurtdışı borsa göstergeleri göz önüne alınarak karar ağaçlarının oluşturulması ve bu ağaçlara dayalı olarak karar kurallarının elde edilmesi sağlanmıştır.

“Mesleki Eğitimde Öğrenci Altyapısının Öğrenci Eğitim Başarısına Etkisinin Veri Madenciliği Yöntemleriyle Ortaya Çıkartılması” adlı yüksek lisans tez çalışmasında, meslek lisesinde eğitim gören öğrencilerin başarısına etki eden “öğrenci altyapısının” veri madenciliği yöntemleri kullanılarak tespit edilmesi hedeflenmiştir (Yelegin, 2012). Yapılan bu çalışmada öğrenci başarısının alt yapısına etki eden faktörlerden sosyal, ekonomik ve kültürel sorunları araştırılmıştır. Konuyla ilgili başarılı sonuçlara ulaşılmıştır.

“Analyzing Student Performance in Distance Learning With Genetic Algorithms and Decision Trees” adlı çalışmasında, Hellenic Open Üniversitesinde kayıtlı uzaktan eğitim öğrencilerinin final sınavlarındaki başarılarını karar ağaçları ve genetik algoritma kullanarak tahmin etmeye çalışmışlardır, öğrencilerin ev ödevleri ölçüt olarak kullanılmıştır (Kalles ve Pierrakeas, 2006).

Uluslararası İleri Teknolojiler Sempozyumu’nda sunulan “Öğrenci Seçme Sınavında(ÖSS) Öğrencilerin Başarısını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle Tespiti” adlı çalışmada, Karar ağaçları algoritması kullanılarak ÖSS sınavına giren öğrenciler üzerine yapılan sosyal, eğitim hayatları ve ebeveyn eğitimleri ile ilgili anket çalışması ile sınav başarıları üzerine etkisi belirlenmiştir (Bozkır ve diğerleri, 2009).

International Arab Conference on Information Technology konferansında sunulan “Mining Student Data Using Decision Trees” adlı çalışmada, karar ağaçları algoritmaları ile öğrencilerin üniversitede aldıkları derslerde başarıyı etkileyen faktörler tespit edilmeye çalışılmış ve bir dersin final notları tahmin edilmeye çalışılmıştır (Al-Radaideh ve diğerleri, 2006).

Elektrik Elektronik Bilgisayar Mühendisliği Sempozyumunda sunulan çalışmada, Fırat Üniversitesi Teknik Eğitim Fakültesi Elektronik ve Bilgisayar Eğitimi bölümünde eğitim gören öğrencilerin aldıkları matematik, fizik, kimya, Türk dili ve Atatürk İlkeleri Ve İnkılâp Tarihi dersleri arasındaki ilişkiler birliktelik kurallarından Apriori Algoritması ile gösterilmiştir (Karabatak ve İnce, 2004).

Ticaret ve Turizm Eğitim Fakültesi Dergisine yayınlanan, “Öğrenci Başarılarının Sınıflandırılmasında Lojistik Regresyon Analizi ve Sinir Ağları Yaklaşımı” çalışmasında yapay sinir ağları ve lojistik regresyon algoritmaları kullanılarak, Gazi Üniversitesi Ticaret ve Turizm Eğitim Fakültesinde eğitim gören öğrenciler üzerinde program, cinsiyet, lise ortalaması, ÖSS puanı gibi parametreler kullanılarak akademik başarı tahmini ve performans karşılaştırması yapılmıştır (Güneri ve Apaydın, 2004).

Journal of Data Science dergisinde yayınlanan, “A Data Mining Approach for Identifying Predictors of Student Retention from Sophomore to Junior Year” çalışmasında, Arizona State Üniversitesindeki öğrencilerin hangi nedenlerden dolayı okula gitmekten vazgeçtikleri sınıflama ağaçları, MARS (Multivariate adaptive regression splines) ve yapay sinir ağları yardımıyla tespit edilmeye çalışılmıştır (Ho Yu ve diğerleri, 2010).

Öğrencilerin Etnik köken ve ikametgâh adreslerinin lise başarılarından daha önemli olduğu bulgusuna varılmıştır.

Education Economics’te yayınlanan, “Predicting Academic Performance by Data Mining Methods” adlı çalışmada, üniversiteye yeni kayıt yaptıran öğrencilerin akademik başarılarını en çok hangi parametrelerin etkilediği tespit edilmeye çalışılmıştır (Vandamme ve diğerleri, 2007). Bu amaçla öğrenciler düşük, orta ve yüksek risk gruplarına ayrılmışlardır. Diskriminant analizi, sinir ağları ve karar ağaçları kullanarak akademik başarıları tahmin edilmeye çalışılmıştır.

Proceedings of Informing Science & IT Education Conference (InSITE) konferansında sunulan “Early Prediction of Student Success: Mining Students Enrolment Data” çalışmasında Yeni Zelanda Open Polytechnic ’de eğitim gören öğrencilerin yaş, cinsiyet, etnik kimlik, ders ve program gibi parametreler kullanılarak CART algoritması kullanılarak başarıyı tahmin etmeye yönelik çalışmada sınıflandırma ve tahmin işlemleri gerçekleştirilmiştir (Kovačić, 2010).

College of Arts and Sciences dergisinde yayınlanan, “Data Mining Techniques Applied to Texas Woman’s University’s Enrollment Data – What Can The Data Tell Us?” adlı alıřmada, Texas Woman’s University ’de kayıtlı ğrenci verileri incelenerek veriler arasında anlamlı iliřkiler ıkarılmaya alıřılmıř ve kmeleme analizi ile ğrenci srekliyi iincelenmiřtir (Yang, 2006).

2. PROBLEMİM TANIMLANMASI

Son dönemde oldukça gündemde olan dershaneler özellikle kayıt sayılarına baktığımızda oldukça revaçta oldukları gözlemlenmektedir. Öğrencilere sundukları eğitim programları ile öğrencinin üniversiteye yerleşmesine yardımcı olmaktadır. Hatta daha lise döneminde kayıt almakta ve öğrenciyi derslerindeki başarısına göre ve üniversitelerin bölümlerinin hangi branş derslerde daha başarılı olan öğrencileri kabul ettiğine göre öğrenciyi branş derslere yönlendirmektedirler.

Bu tez çalışmasında özellikle katılım süresinin ve branş derslerinin öğrencinin üniversiteye yerleşmesindeki katkısı irdelenmektedir.

2.1 EĞİTİM VE ÖĞRETİM SİSTEMİNDE DERSHANELERİN YERİ

Dershanelerin birinci önceliği, bir eğitim kurumu olmaktan önce öğretim faaliyetlerini sınırlı zaman diliminde gerçekleştirmeyi amaçlayan kurumlar olma şeklinde açıklanabilir.

Eğitim, genel anlamda öğretimi de kapsayan çok genel bir tanımlamadır. Öğretim ise daha sınırlı bir alanda öğrencilerde davranış değişikliğine yol açacak bilgilerin kazandırılmasıdır.

Dershaneler sınav odaklı kurumlar olduğu için öncelikle işin öğretim ayağını tamamlamaya yönelik faaliyetler yaparlar.

2.2 DERSHANELERİN KURULUŞ AMACI

Dershanelerin başlangıçta kuruluş amacı ile günümüzdeki işlevlerinde de bir takım değişiklikler yaşanmaya başlanmıştır.

Artık sadece sınavlara hazırlayan kurumlar değil, öğrencinin okul dışı zamanını yine öğretim ortamı ile değerlendirdiği ve onu zararlı alışkanlıklardan uzaklaştıran yerler olarak da değerlendirilebilir.

Son yıllarda dershaneler bir takım sosyal, kültürel ve sportif aktiviteler ile öğrencilerin ilgi ve yeteneklerini destekleyen, onları sosyalleştiren yerler olarak da faaliyet göstermeye başlamışlardır.

Bu açıdan bakıldığında kurumsallaşmış dershaneler sadece öğretim yapan yerler değil, aynı zamanda eğitim sürecini de devam ettiren kurumlar olma yolunda ve okullarda gerçekleştirilmeyen sosyal etkinliklerin yapıldığı yerler haline gelmektedir.

Dershaneler özel sektör girişimidir. Hizmet üreten kurumlardır ve hedef kitlesi insandır. Gerek özel sektör oluşu gerekse hizmet sektöründe olması müşteri memnuniyetini de beraberinde getirmektedir.

Okulda ötelenen, dışlanan, ilgi gösterilmeyen öğrencilerin kendilerini iyi hissettikleri kurumlar olarak hizmet standartlarını artırmaktadırlar. Ülkemizde çok sayıda dershanenin oluşu, hizmet kalitesi konusunda bir rekabet de yaratarak öğrencinin daha nitelikli bir eğitim almasını sağlamaktadır.

Dershanelerin değişen ve gelişen sisteme ve bilimsel yeniliklere ayak uydurması da son derece hızlıdır (Ünsal, 2013). Derslerin içerik değişimleri ve sınav sistemindeki değişimler, hemen algılanarak kurumsal bazda entegrasyon hızlı bir şekilde sağlanmaktadır. Hizmet içi eğitim programlarıyla eğitim hizmetini sunan öğretim kadrosu bilgilendirilmekte, ders içerikleri, dersin işleniş yöntemi vb. sürekli güncellemeler yapılmaktadır.

2.3 DERSHANELERDE EĞİTİM ÖĞRETİM SÜRECİ

Genelde ilk ve orta öğrenim kurumlarında temel kazanımları edinen bir öğrenci, hem genel tekrarlarla bilgisini pekiştirmek, hem de okulda süresi içinde edinemediği bilgiyi tamamlamak üzere dershaneye gelme ihtiyacı duymaktadır.

Öğrenciler bu kurumlarda uygulanan sınav odaklı eğitimle sınav içeriğine uygun olan ve soru çıkma olasılığı yüksek olan ünite ve konulardan bilgisini tazeler, ayrıca test ve sınav tekniği konusunda kendisini geliştirir. Bu süreçte kendisi ile aynı programı izleyen arkadaşları ile bir rekabet ve yarış havası içinde daha iyi bir motivasyon sağlar.

Dershanelerde düzey sınıflarının oluşturulması aynı başarı düzeyine sahip öğrencilerin bir sınıfta toplanmasına ve işlenen konuların sınıfın, dolaylı olarak da o sınıftaki her bir öğrencinin bilgi, başarı ve algı düzeyine uygun yöntemlerle anlatılmasını sağlar. Bu da hem öğrencinin anlamasını, hem de öğretmenin işini kolaylaştırır.

Yıl içinde çok sayıda uygulanan deneme sınavları ile öğrenci, hedef olarak belirlediği sınava girmeden önce önemli bir deneyim kazanmış olur. Bir taraftan soru çözme hızını artırırken, diğer taraftan önemli ölçüde zaman kullanımı, okuma ve anlama hızını geliştirme, kodlama disiplini gibi sınavın teknik ayrıntılarını öğrenir.

Merkezi sınavların en önemli olumsuzluklarının başında sınavın psikolojik yükü gelmektedir. Sonunda öğrencinin geleceği ve kariyer planlaması ile ilgili belirleyiciliği olan bir sınavın mutlaka kaygı ve heyecan yaratması beklenir. Bu her adayda aynı düzeyde olmasa da büyük ölçüde yaşanır.

Dershanelerde yapılan bu uygulamalar sınava sadece bilgi yönünden değil psikolojik yönden de hazırlanmayı kolaylaştırır. Sürekli deneme sınavına girerek hız ve deneyim kazanan bir aday zamanla gerçek sınavı da bir deneme sınavından farksız olarak algılamaya başlar ve sıradanlaştırır.

Dershanelerin sistem içindeki en vazgeçilmez yanlarından birisi de ders dışı uygulamaların yoğunluğudur. İsteyen adaylar ders dışında organize edilen etüt ve birebir çalışmalar ile kendini daha ayrıcalıklı hisseder.

Ülkemizde genç nüfus sayısındaki fazlalık ve okullaşmanın henüz ihtiyacı yeterince karşılayamaması okulların ve öğretmenlerin öğrenciye ilgi konusunda yeterince zaman ayıramaması öğrenme sürecini olumsuz etkiler. Günümüz okulu öğrenci zihninde devam edilip diploma alınan yer olarak algılanmaktadır.

Sınava hazırlık konusunda dershaneler okullara oranla daha güçlü bir algıya sahiptir. En yüksek puanlı ve en nitelikli okulların öğrencileri dahi bir dershanede eğitim hizmeti alarak bu yarışta geri kalmamayı amaçlamaktadır.

Toplumda dershaneye gitmeden sınavda başarılı olunamayacağına dair net bir algı oluşmaya başlamıştır (Ünsal, 2013). Bu algı tartışmaya açıktır ama dershanelerin sınava hazırlıkta öğrenciye kattığı bilgi, teknik ve motivasyon kesinlikle yadsınamaz.

Dershanelerin bir diğer ayrıcalığı sınava yönelik test ve doküman zenginliğidir. Okul kitaplarında kullanılan içerik ve yöntem doğrudan sınava yönelik bir hazırlığı destekler nitelikte değildir.

Okulda merkezi sınav tekniğine uygun olmayan yöntemlerle ölçme değerlendirme yapılması da öğrencilerin merkezi sınavda başarılı olabilmeleri için yeterli veri sunmamaktadır.

2.4 DERSHANELERDE TEKNOLOJİNİN YERİ

Teknolojik yenilikler artık eğitim öğretim ortamında da etkin ve verimli bir şekilde kullanılmaktadır. Dershaneler bu yeniliklere de hızlı bir geçiş yaparak sadece teknolojiyi sınıfta hazır bulundurmak değil, içeriğini de öğrenciye sunmak bakımından etkin çalışmalar yürütmektedirler.

Bugün kurumsal pek çok dershanenin sınıflarında yıllardır “akıllı tahta” ve “akıllı içerik” in kullanılması hem öğrencinin zamanı verimli kullanmasına olanak sunmakta, hem de öğrenmeyi kolaylaştırmaktadır.

Okullarda “FATİH Projesi” ne geçilmesine rağmen içerik konusunda bir netliğin olmaması sadece teknolojiyi kurmak değil, etkin bir şekilde kullanmanın da önemli olduğunu bize göstermektedir(Ünsal, 2013).

Fatih projesinin 2011-2014 yılları arasında gerçekleştirilmesi planlanmaktadır. Proje kapsamında her okula 1 adet yazıcı, 1 adet doküman kamera, her dersliğe etkileşimli tahta ve kablolu internet bağlantısı, her öğretmene ve öğrenciye tablet bilgisayar verilmesi planlanmaktadır. <http://fatihprojesi.meb.gov.tr> adresinden detaylı bilgiye erişilebilir.

2.5 SINAV VE NOT KAYGISI

Okulların elini güçlerinden en önemli unsur not olmasına ve dershanelerin not gibi bir kaygıları olmamasına rağmen öğrencinin boş zamanını dershanede değerlendirmesi ölçme değerlendirme yöntemi konusunda dershaneyi bir adım öne çıkarmaktadır.

Öğrenci okulu bitirecek kadar not almayı yeterli görmekte ama nihai sınavına daha ciddi bir tutum ile yaklaşmaktadır. Bu doğrultuda dershanede sunulan ölçme değerlendirme hizmeti öğrenciyi daha objektif kriterlere göre değerlendirir niteliktedir (Ünsal, 2013).

Öğrenci kaygısı üzerine “Multi-Resolution Wavelet Analyses of Two Different Perfectionism Scales: A University Sample” adlı çalışmada kişinin anlık tepkisi program vasıtasıyla kayıt altına alınmış ve verdiği tepkiler gözlemlenmiştir (Karaca ve Diğerleri, 2010).

2.6 DERSHANELERDE VERİ MADENCİLİĞİ İHTİYACI

Özellikle kurumsallaşan dershaneler çok sayıda öğrenciye hizmet vermekte, çok sayıda öğretmen ve görevli istihdam etmektedir. Öğrencilere yıl boyunca kitap ve sınav hizmeti sunmaktadır.

Öğrenci ödemelerinden maaş ve çeşitli giderlere, sınav hizmetlerine kadar her şey kayıt altına alınmaktadır. Kurumsallaşan dershaneler bu kayıtları her yıl düzenli olarak tutmaktadırlar. Bu verileri kullanarak anlamlı sonuç çıkarma ve yeniden yapılanma ve planlama ihtiyacı duymaktadırlar. Ancak tutulan veriler çok fazla ve birbirinden çok farklı olacağından bu verilerin ihtiyaca cevap verecek şekilde ayrılması gerekmektedir. Bunun için de veri madenciliği tercih edilebilmektedir.

3. VERİ MADENCİLİĞİ

Veri madenciliği sürekli güncellenen ve hareket gören deęişken veri tabanlarına uygulanmaz. Öncelikle veriler belirli aşamalardan geçirilerek oluşturulan veri ambarına aktarılır. Veri ambarındaki özet veriler üzerine VM algoritmaları uygulanır.

Veri madencilięi, veriler arasındaki ilişkilerden faydalanarak yeni, anlaşılır bilgi çıkarma işlemidir (Al-Hudairy ve Diğerleri, 2004).

Veri Madencilięi, sistemde bulunan ama bilinirlięi olmayan ancak potansiyel olarak yararlı olan verinin bilgiye dönüştürölme sürecidir (Shen, 2007).

Daha basit olarak veri madencilięi büyük ölçekli veriler arasından deęerli olan bilginin bulunup ortaya çıkarılması işlemidir (Silahtaroęlu, 2008).

VM Satış ve Pazarlama, Müşteri İlişkileri Yönetimi, Risk Analizi ve Yönetimi, Eğitimde Başarının Arttırılması alanlarında, Perakende, Telekomünikasyon, Bankacılık, Finans ve Portföy Yönetimi, Eğitim gibi sektörlerde kullanılmaktadır.

3.1 VERİ MADENCİLİĞİ TARİHİ

Veri madencilięi (Data Mining) – veri madencilięi olarak adlandırılmasa da - tarihi yaklaşık olarak 40 yıl kadar öncesine dayanır. SAS ve SPSS şirketleri tarafından istatistiksel analiz için kullanılmıştır.

Regresyon analizi ile istatistik, standart sapma/ dağıtım/ varyans, kümeleme analizi, güven aralıęı önemli yöntemler olsa da yeni teknikler bu yöntemlerin gücüne güç katmaktadır.

Bu yeni yöntemler 1980'lerde ortaya çıkan bulanık mantık, sezgisel ve sinir aęları gibi yöntemlerdir. Bu yöntemler yapay zekâ ve makine öğrenmesi gibi iki gruba ayrılırlar.

1990'lı yılların başında veri tabanlarında bilgi keşfi süreci başlamıştır. Günümüzde birçok sektör ve alanda veri madenciliği kullanılmaktadır.

3.2 VERİ MADENCİLİĞİNİN KULLANIM GEREKSİNİMİ

Kayıtların hızlı bir şekilde artması, rekabetin artması gibi faktörler özellikle pazarlama stratejilerinin de geliştirilmesini sağlamaktadır. Aşağıda sayabileceğimiz birkaç neden veri madenciliğinin kullanılma gereksinimini ortaya koymaktadır.

- Hızla artan veri kayıtları,
- Çok fazla ve çok farklı türde verinin, tek bir ortamda depolanması,
- Geleneksel eski tekniklerin ham verileri işlemede yetersiz kalması,
- Veri madenciliği bilim insanlarına verileri sınıflandırma ve gruplandırma, hipotez oluşturma, varsayımlar ve tahmin yapabilme konusunda yardımcı olur,
- Rekabet ve gücün işletmeler için büyük önem arz etmesi.

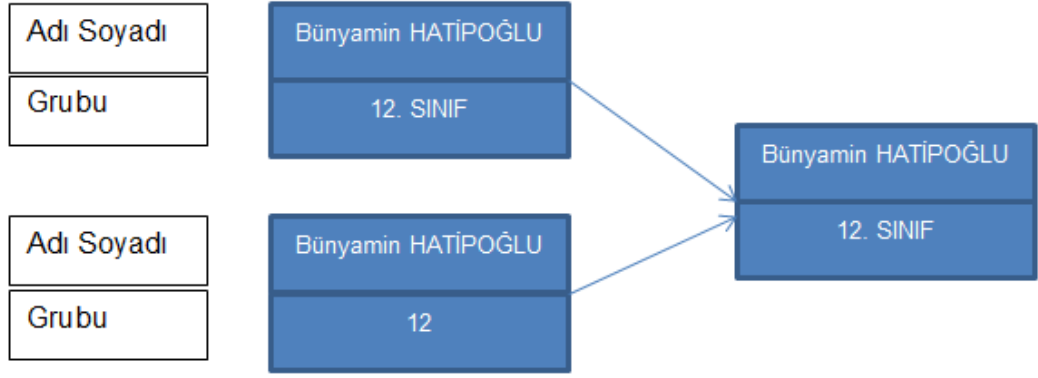
3.3 VERİ MADENCİLİĞİNDE KULLANILAN ADIMLAR

Veri temizleme: Tutarsız ve hatalı veriler, veri tabanı üzerinde yapılacak analizlerde yanlış sonuç verebileceğinden bu hatalı veriler veri ambarına aktarılmadan önce silinir. Örneğin dershanede eğitim gören bir öğrenci eğer eğitimi boyunca hiçbir deneme sınavına girmediyse yapılacak analizde bu öğrenciye yer verilmemelidir.

Veri bütünleştirme: Farklı veri kaynaklarında tutulan aynı türdeki veriler farklı isimlere sahip olabilir. Örneğin öğrenci verilerinin tutulduğu kaynakta öğrencinin grubu "12. Sınıf" ve "Mezun" diye tutulabilirken, sınav verilerinin

tutulduğu kaynakta “12” ve “MZ” adıyla tutulabilir. Bu yüzden veriler birleştirilirken aynı isimle birleştirilmelidirler.

Şekil 3.1’de veri bütünleştirmeye ait örnekte ad ve soyadı aynı olmasına rağmen grubu farklı gösterilen öğrencinin “Grubu” adlı verisi “12. SINIF” diye aynı adla isimlendirilmiştir.



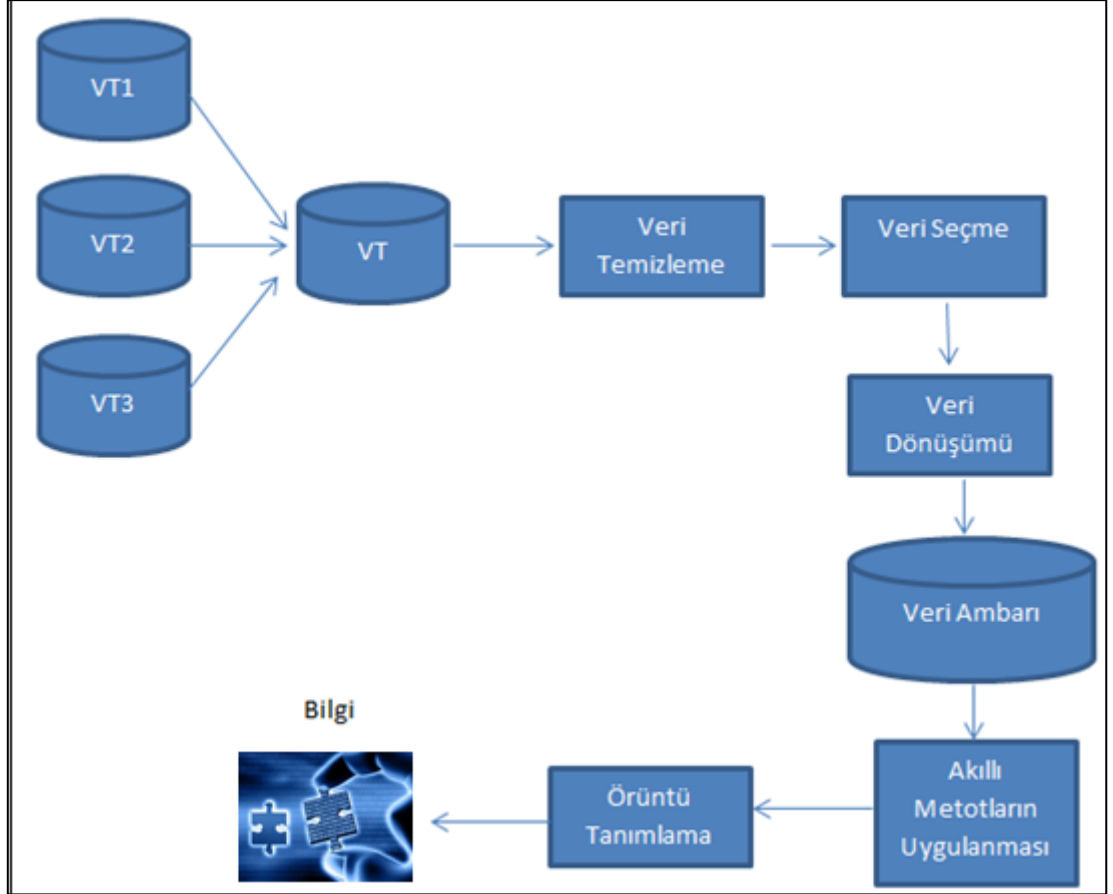
Şekil 3.1 Veri Bütünleştirmeye Ait Bir Örnek

Veri Seçme: Yapılacak olan analiz ile ilgili yalnızca ihtiyaç duyduğumuz verileri belirlemek gerekir. Örneğin öğrencinin telefon bilgisine ihtiyaç duymuyorsak telefon bilgilerini veri ambarına almaya gerek yoktur.

Veri Dönüşümü: Bazen özellikle sayısal değerleri aynen hesaplamaya katmamak gerekir. Örneğin “A” dersinden 20 soru üzerinden 10 soru çözen %50 başarılıdır. “B” dersinden 30 soru üzerinden 15 soru çözen yine %50 başarılıdır. Bundan dolayı bazı hesaplamalarda soru sayılarını birbirlerine endeksleyerek hesaplama yapılmalıdır.

Bu aşamadan sonra veri örüntüleri için akıllı metotlar uygulanabilir. Ardından bazı ölçümlere göre elde edilen bilgiyi gösteren örüntüler tanımlanır. Son olarak da elde edilen bilginin kullanıcıya sunumu gerçekleştirilir.

Şekil 3.2’de veri madenciliğinde kullanılan adımlara görsel olarak yer verilmiştir.



Şekil 3.2 Veri Madenciliğinde Kullanılan Adımlar

3.4 VERİ AMBARI

Veri Ambarı, bir veya birden fazla veri kaynağında bulunan verilerin, veri temizleme, veri bütünleştirme, veri dönüştürme gibi belirli evrelerden geçirilerek aktarıldığı özet bir veri tabanıdır.

Veri Ambarları, veri kaynaklarından aldığı verileri birleştirip, bunları karar destek uygulamalarında kullanılmasını sağlarlar (Seetharaman, 2008).

Veri ambarı, sorgulama ve analiz için kullanılmak üzere tasarlanmış ilişkisel bir veri tabanıdır. Veri ambarında veriler, genelde hareket verisinden elde edilmiş tarihsel bilgilerle beraber saklanırlar. Veri tabanı hareketlerinden kaynaklanan iş yüküyle analiz yükünü birbirinden ayırır. Bu sayede değişik veri kaynaklarından toplanan verilerin tek bir yerden analiz edilmesini sağlarlar (Düzgünoğlu, 2006).

Veri ambarında tutulan verilerin değişken olmayan yapıda olması gerekir. Veri tabanlarında tutulan veriler gerçek zamanlıdır ve değişken olabilirler ancak veri ambarına aktarıldıktan sonra veri ambarındaki veriler değiştirilmemelidir. Veri ambarındaki tarihsel veri genellikle gelecekte ne olabileceğini tahmin etmek için raporlama ve analiz amacıyla kullanılır (Düzgünoğlu, 2006).

Veri tabanları üzerinde sorgulama yapıp gelen sonuçların düzenlenmesi yerine, daha önceden hazırlanıp düzenlenen veriler üzerinde sorgulama yapılması amaçlanır. Bu işlemlerin önceden yapılıp veri ambarında hazır tutulması hem veri çekim işlemlerinde performansı artırır, hem de veri tabanlarındaki iş yükünü azaltır (İşli, 2009).

3.5 VERİ MADENCİLİĞİ MODELLERİ

Veri madenciliğinde kullanılan modeller, tahmin edici ve tanımlayıcı modeller olmak üzere ikiye ayrılır.

Tahmin edici modellerde, sonuçlardan faydalanılarak bir model geliştirilir ve bu model başka veriler üzerine kullanılarak o verilere ait sonuç tahmin edilmeye çalışılır.

Tanımlayıcı modellerde ise veriler üzerine örüntü tanımlanması amaçlanmaktadır.(Gülçe, 2010)

Veri madenciliği modelleri gördükleri işlevlere göre üçe ayrılmaktadır. Sınıflama ve Regresyon, Kümeleme ve Birliktelik Kuralları (Akpınar, 2000).

3.5.1 Sınıflama ve regresyon

Amaç, girdi olarak tahmin edici değişkenlerin yer aldığı ortamda, çıktı olarak bağımlı bir değişkenin değerinin bulunduğu bir model kurmaktır. Bağımlı değişken sayısal değil ise problem sınıflama problemidir. Eğer bağımlı değişken sayısal ise problem regresyon problemi olarak adlandırılır(Akpınar, 2000).

Sınıflama ve regresyon, verileri sınıflamak veya gelecek veri eğilimlerini tahmin etmek için kullanılan iki veri analiz yöntemidir. Sınıflama verileri kategorize etmek için kullanılırken, regresyon süreklilik gösteren değerlerin tahmin edilmesinde kullanılır. (Özekes ve Çamurcu, 2002)

Örneğin, sınıflama modeli öğrencilerin üniversiteye yerleşme durumlarını temel alarak, eğitim gördükleri branş derslerinin ne kadar isabetli olduğunu kategorize etmek için kullanılırken, regresyon modeli anket sonuçları verilen öğrencilerin üniversitede okumak isterken hangi programları tercih edeceklerini tahmin etmek için kullanılabilir.

Sınıflama ve regresyon modelinde kullanılan başlıca yöntemler aşağıda listelenmiştir.

- Karar Ağaçları
- Bayes Sınıflandırma
- Yapay Sinir Ağları

3.5.1.1 Karar Ağaçları

Karar ağaçları ile üretilen model tersine çevrilmiş bir ağaca benzemektedir. Bu ağaç karar verme noktaları olan düğümler ve bu düğümleri birbirine bağlayan dallardan oluşmaktadır. En tepede kök düğüm bulunmaktadır. Bu düğümde bir takım özellikler test edilmekte ve bu testin

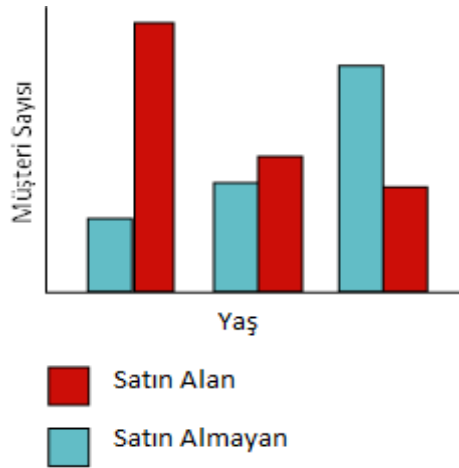
farklı sonuçlarına göre kök düğümden dallar türemektedir. Her bir dal yeni bir karar düğümüne bağlanmakta ve burada yeni birtakım özellikler test edilerek bu düğümlerden dallar türemektedir. Ağaç yapısının en altında ise artık kendisinden dal türemeyen yaprak düğümleri bulunmaktadır (Seyrek ve Ata, 2010)

Birbirinden farklı, ayrık (Ayrık Veri) özellikler için, algoritma, veri seti içinden giriş (input) sütunları arasındaki ilişkileri temel alarak bir tahmin gerçekleştirir. Özellikle algoritma, tahmin sütunu ile bağlantılı olan giriş sütunlarını tahmin eder.

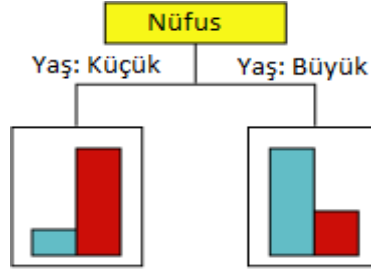
Örneğin, Bisiklet satın alan müşterilere bakacak olursak, 10 genç müşteriden 9'u bisiklet satın alırken yaşlı müşterilerden sadece 2'si bisiklet alıyorsa algoritma yaş faktörünün önemli bir tahmin edici unsur olduğu sonucuna varır.

Süreklilik arz eden (Sürekli Veri) özellikler için, algoritma lineer regresyon kullanarak, karar ağacının nerelerde dallanacağını belirler. Örneğin pazarlama bölümü, müşterilerin demografik yapılarını inceleyerek hangi ürünleri tercih ettiklerine göre demografik bir satış stratejisi belirleyebilir.

Grafik 3.1'de bisiklet satın alan ve almayan müşteriler yaş temelinde sınıflandırılmıştır. Grafik 3.2'de ise karar ağacı modelinin yaş faktörünü "Küçük" ve "Büyük" diye ikiye ayırdığını görüyoruz.



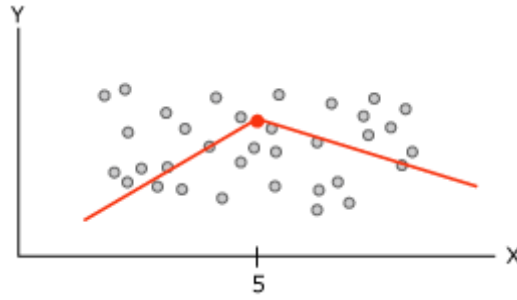
Grafik 3.1 Ayrık Veri Örneği



Grafik 3.2 Ayrık Veri Sonucu

Uygulanan algoritma sonucu modele yeni dallar (node) eklenmiştir ve kökte Nüfusun tamamı yer almış ve dallanarak devam etmiştir.

Grafik 3.3'de regresyon formülüyle lineer olmayan noktalar belirlenir.



Grafik 3.3 Sürekli Veri Örneği













Diyagram tek ya da iki birbirine bağlantılı çizgi kullanılarak modellenen veri içerir. Tek çizgi veri için yeterli değildir. İki çizginin kullanılması verinin sunumunu daha anlaşılır hale getirir.

3.5.1.2 Bayes Sınıflandırıcıları

Naive bayes algoritması Bayes teoremini temel alan bir sınıflama algoritmasıdır. Bayes algoritması diğer algoritmalara göre daha az hesaplama içerir ve giriş sütunu ile tahmin sütunu arasındaki ilişkilerin daha hızlı bir şekilde keşfedilmesini sağlar.

Örneğin; promosyon stratejisi geliştiren bir pazarlama bölümü hangi müşterilere el ilanı göndereceğine bu yolla karar verebilir. Maliyeti düşürmek için olumlu cevap alabilecek müşterileri seçmek ister. Yaş ve lokasyon bilgilerine bakarak hangi müşterilerin ürün satın almada benzer özellikler gösterdiğini keşfedebilir.

Şekil 3.3'de yaş, mesafe ve eğitim durumuna göre bisiklet satın alma ya da satın almama durumuna göre sınıflandırma sonucu gösterilmiştir.

Attributes	States	Population ... Size: 18484	0 Size: 9352	1 Size: 9132	missing Size: 0
Age	<ul style="list-style-type: none">38 - 4329 - 3443 - 48Other				
Commute Distance	<ul style="list-style-type: none">0-1 Miles2-5 Miles1-2 MilesOther				
Education	<ul style="list-style-type: none">BachelorsPartial CollegeHigh SchoolOther				

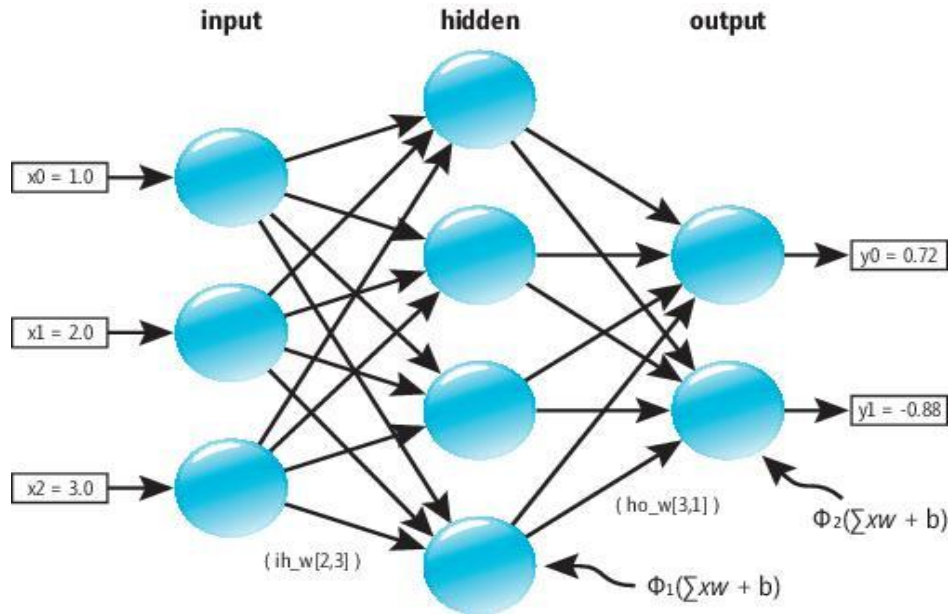
Şekil 3.3 Yaş, Mesafe ve Eğitim Durumuna Göre Bisiklet Satın Alma ya da Satın Almama Durumuna Göre Sınıflandırma (0 – Satın Almama, 1 – Satın Alma)

3.5.1.3 Yapay Sinir Ağları

Yapay sinir ağları insan beyninden esinlenerek geliştirilmiş bir yöntemdir ve mühendislik, finans, eğitim gibi birçok alanda kullanılmaktadır (Paliwal ve Kumar, 2009).

Yapay sinir ağları temelde 3 katmandan oluşur. Giriş (input) katmanı, gizli (hidden) katman ve çıkış (output) katmanı. Ayrıca gizli katman ihtiyaca göre artırılabilir. Ancak katman sayısının artması öğrenme sürecini arttırmakla birlikte performansı da düşürmektedir.

Şekil 3.4'de bu katmanlar bir örnekle gösterilmiştir (<http://msdn.microsoft.com/en-us/magazine/hh975375.aspx>, Ocak 2013).



Şekil 3.4 YSA Katmanlarının Gösterimi

3.5.2 Association Rules – Birliktelik Kuralları

Birliktelik algoritması özellikle market sepet analizi için kullanılan kullanışlı bir algoritmadır. Müşterilerin önceden satın aldığı ürünleri temel olarak müşteri davranışlarını tespit etmekte kullanılmaktadır. Birliktelik algoritması hangi ürünlerin birlikte gruplandığını tespit eder. Amaç müşterinin gelecekte hangi ürünleri satın alabileceğini tespit etmektir.

Ayrıca eğitimde çeşitli çalışmalarla örneğin öğrencilerin başarısız oldukları dersler arasındaki bağlantılar tespit edilebilmekte veya seçmeli derslerin belirlenmesinde öğrenci tercihleri temel alınarak seçmeli dersler arasındaki ilişkiler tespit edilebilmekte ve buna göre seçmeli dersler açılabilenkte veya kaldırılabilenktedir.

Bu tez çalışmasında öğrencilerin sorumlu oldukları dersler, dershanede eğitim alma süresi gibi gruplar arasındaki bağlantıların yerleşme durumuna etkisi incelenenktedir.

Destek sayısı (support) veri seti içerisindeki verilerin kombinasyonun kaç defa geçtiğini gösterenktedir. Güven seviyesi (confidence) ise bu kombinasyonla birlikte diğer verilen birlikte geçme olasılığını gösterir.

3.5.2.1 Apriori Algoritması

Veri madenciliğinde birliktelik kuralı modelinde kullanılan ve veri kümeleri arasındaki ilişkiyi çıkarmak için geliştirilen bir algoritmadır.

Apriori algoritması, özellikle çok büyük ölçekli veri tabanları üzerindeki veri madenciliği çalışmaları için geliştirilmiştir.

Algoritmanın amacı, veri tabanında bulunan satırlar arasındaki bağlantıyı ortaya çıkarmaktır. Algoritmanın ismi, Latince de önce anlamına gelen “prior” kelimesinden gelmektedir.

Algoritma yapı olarak, aşağıdan yukarıya (bottom-up) mantığını kullanır ve her seferinde tek bir elemanı inceleyerek bu elemanın diğer elemanlarla birlikteliğini ortaya çıkarmaya çalışmaktadır.

Ayrıca algoritmanın her eleman için çalışmasını, bir arama algoritmasına benzetmek mümkündür. Algoritma, bu anlamda sıg öncelikli arama (breadth first search) yapısında olup, elemanları birer ağaç (tree) gibi düşünerek bu ağaç üzerinde arama yapıyor kabul edilebilir. (http://en.wikipedia.org/wiki/Apriori_algorithm, 2013).

Apriori algoritması pseudo-code'u Şekil 3.5'de gösterilmiştir.

```
Apriori( $T, \epsilon$ )
 $L_1 \leftarrow \{\text{large 1 - itemsets}\}$ 
 $k \leftarrow 2$ 
while  $L_{k-1} \neq \emptyset$ 
     $C_k \leftarrow \{c | c = a \cup \{b\} \wedge a \in L_{k-1} \wedge b \in \bigcup L_{k-1} \wedge b \notin a\}$ 
    for transactions  $t \in T$ 
         $C_t \leftarrow \{c | c \in C_k \wedge c \subseteq t\}$ 
        for candidates  $c \in C_t$ 
             $count[c] \leftarrow count[c] + 1$ 
         $L_k \leftarrow \{c | c \in C_k \wedge count[c] \geq \epsilon\}$ 
     $k \leftarrow k + 1$ 
return  $\bigcup_k L_k$ 
```

Şekil 3.5 Apriori Algoritması Pseudo-Code

3.5.3 Kümeleme modelleri

Kümeleme algoritması bir bölümlenme algoritmasıdır. Tekrarlamalı tekniklerle veriler benzer karakteristik özelliklere göre gruplara ayrılırlar. Model belirlenirken tahmin edilecek sütun birden fazla olabilir.

Nesnelerin kendilerini ya da diğer nesnelere olan ilişkilerini tarif eden bilgileri kullanarak nesnelere gruplara ayırma işlemidir. Kümelemede amaç; grup içindeki nesnelere, diğer gruplardaki nesnelere olabildiğince

ayrı/bağımsız, kendi aralarında ise birbirine benzer/bağımlı olacak şekilde oluşturmaktır (Özdamar, 2002).

Kümeleme modellerinde amaç verilerin birbirlerine çok benzediği, ancak özellikleri birbirlerinden çok farklı olan kümelerin bulunması ve veri tabanındaki kayıtların bu farklı kümelere bölünmesidir (Ayık ve Diğerleri, 2007).

3.5.3.1 K-Means (K- Ortalama) Algoritması

K-Ortalama algoritması bölümlenmeli algoritmaların en çok bilinenidir (1967 yılında geliştirilmiştir). Bundan sonrasında geliştirilen bölümlenmeli algoritmalar, k-Ortalama algoritmasına çok benzer çalışma mantığına sahiptir.

K-Ortalama algoritması sayısal veriler üzerinde çalışan bir algoritmadır. Bu algoritma aşırı uç veya gürültülü verilerden etkilenir.

Algoritmada ilk olarak k adet küme oluşturulması hedefi ortaya konur ve sonrasında k tane ortalama değeri rastgele belirlenir.

Verilen bu ortalama değerlerine göre de bütün sayılar hangi ortalamaya yakınsa o kümeye dâhil edilir. Algoritma bütün sayıları kümeledikten sonra bir kez daha ortalama değerleri bulunur ve tekrar sayılar hangi ortalamaya yakınlarsa oraya dâhil edilirler. Bu işlemler son 2 işlemden önce aynı kümeler çıkana kadar devam eder

(<http://www.iszekam.net/?tag=/veri+madenciliği+algoritmaları>, Ocak 2013).

Verilerin belirli gruplara dâhil edilmesi için bazı hesaplamalara tabii tutulması gerekmektedir. Bunlar benzerlik ve mesafe ölçümleridir.

3.5.3.2 Benzerlik ve Mesafenin Ölçülmesi

Mesafenin ölçülmesi için euclid teoreminden yararlanılmaktadır. Euclid teoremine göre mesafe 3.1 deki formül yardımıyla hesaplanmaktadır (<http://www.iszekam.net/?tag=/veri+madenciliği+algoritmaları>, 2013).

$$d(x, y) = \sum_{i=1}^p (x_i - y_i)^2 \quad 3.1$$

Bir diğer hesaplanması gereken ise benzerliktir. Benzerlik kavramı mesafenin tersi bir anlam içerir ve iki veri arasındaki yakınlığı gösterir. Benzerlik hesabı genel olarak aşağıdaki genel formül yardımıyla hesaplanabilmektedir.

$$\text{ben}(X_m, X_j) = 1 / (1 + d(x, y))$$

d: Mesafe

x: Nokta

y: Nokta

p: Boyut Sayısı

ben: Benjamin –Kendi adını taşıyan formül-

3.5.3.3 Hiyerarşik Modeller

Hiyerarşik modeller bir ağaç yapısı oluşturarak kümeleme işlemini gerçekleştirmektedir. Oluşturulan kümeleme ağacının bütün ağaç yapılarında olduğu gibi bir root düğümü ve çocuk düğümleri mevcuttur. Aşağıdan yukarıya, toplasım kümeleme algoritmaları ve yukardan aşağıya kümeleme algoritmaları olarak iki grupta toplanabilir.

Toplaşım kümeleme algoritmaları, başlangıçta veri tabanındaki her bir noktayı bir küme olarak görür. Bu kümeleri birleştire birleştire birbirinden ayrı kümeler oluşturur.

Bölünür kümeleme algoritmaları ise başlangıçta veri tabanındaki tüm noktaları tek bir kümeymiş gibi görür. Veri tabanını taradıkça, birbirine benzemeyen noktaları kümeden dışarı atarak önceden verilmiş “k” kadar kümeye dağıtır (<http://www.iszekam.net/?tag=/veri+madenciliği+algoritmaları>, 2013).

3.5.3.4 Bölümlemeli Modeller

Bölümlemeli yöntemlerde n adet nokta önceden verilen k küme sayısına ($k < n$) göre kümelere ayrılır. Hiyerarşik yöntemlerin tersine kullanıcı tarafından verilen bazı kriterlere uygun kümeler yaratılırken, yaratılacak küme sayısı önceden belirlidir.

Bölümlemeli algoritmalar genel olarak hiyerarşik algoritmalarından daha hızlı çalışırlar; çünkü hiyerarşik algoritmalarındaki gibi benzerlik/mesafe matrisi kullanmak zorunda değillerdir.

4. UYGULAMA ARAÇLARI

Bu tez çalışmasında hem veriler üzerinde sorgulama yapılmasına hem de veri madenciliği yöntemlerinin uygulanmasına imkân sağlaması bakımından Microsoft SQL Server platformu kullanılmasına karar verilmiştir. Microsoft SQL Server platformu çatısı altında “Business Intelligence Development Studio” adlı program veri tabanı ile bağlantı kurabilmekte ve veriler üzerinde Veri Madenciliği uygulamalarına olanak sağlamaktadır.

4.1 MICROSOFT SQL SERVER

Microsoft SQL Server kuruluşların veriler üzerinde çeşitli sorgulama ve uygulamalar yapmasına olanak sağlayan bir bilgi platformudur.

Microsoft SQL Server ile birlikte gelen Business Intelligence Development Studio, Analiz Servis ile entegre çalışan derinlemesine raporlama ve analiz yapabilen ve ayrıca Veri Madenciliği uygulamaları geliştirilmesinde hazır algoritmalar yardımıyla görsel ve metinsel analiz sunabilen, görsel ara yüze sahip bir uygulamadır.

Business Intelligence Development Studio çatısı altında hazır algoritmalar yardımıyla örnek uygulamalar gerçekleştirilecektir.

Veriler Microsoft SQL Server veri tabanı sisteminde tutulacaktır. Veri Madenciliği yöntemleri kullanılarak veriler üzerinde çeşitli işlemler yapılarak ihtiyaç duyduğumuz veriler Veri Ambarına aktarılacaktır. Daha sonra Business Intelligence Development Studio adlı program kullanılarak hazır algoritmalar yardımıyla örnek uygulamalar gerçekleştirilecektir.

4.2 MICROSOFT ANALYSISSERVICES

SQL Server'ın en önemli servislerinden biri olan SQL Server Analysis Services, karar destek motorunun ve araçlarının yer aldığı ortamdır. Karar destek mekanizmasına ait iki içerik olan Veri Madenciliği ve OLAP bu ürün kapsamında desteklenmektedir.

4.3 BUSINESS INTELLIGENCE DEVELOPMENTSTUDIO

Business Intelligence Development Studio, Analysis Services, Integration Services ve Reporting Services projelerini içeren iş çözümleri geliştirmek için kullanılacak birincil ortamdır.

Her türlü proje ve karar destek çözümleri için gerekli nesnelerin oluşturulması için şablonlar sağlar ve nesnelerle çalışmaya yönelik çok çeşitli tasarımlar, araçlar ve sihirbazlar sunar.

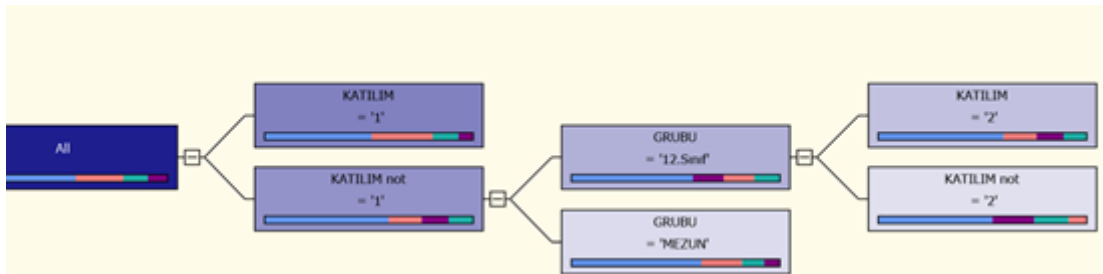
5. ANALİZ

VM bölümünde bahsedilen modellerden faydalanarak bu modeller dershanede eğitim programına katılan öğrenciler üzerinde BIDS programını kullanılarak uygulanacaktır.

Toplamda 4858 öğrenci üzerinde çalışılmıştır. Karşılaştırmalı modeller için test amacı ile rastgele 3000 kayıt üzerinde çalışma gerçekleştirilmiştir. Bunun için SQL üzerinde “SELECT TOP 3000 * FROM Tablo_Adı ORDER BY NEWID()” sorgusu ile birbirinden farklı kayıtlar elde edilmiştir.

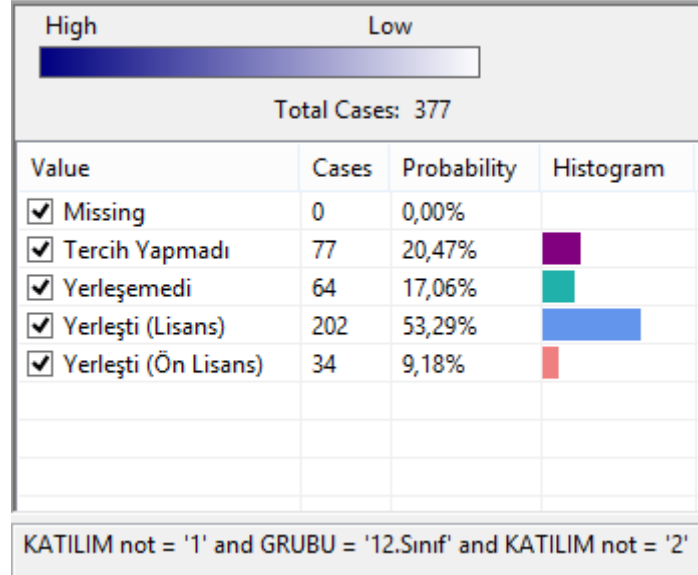
5.1 KARAR AĞAÇLARI

Şekil 5.1’de dershanede eğitim programına katılan öğrenciler üzerine karar ağacı modeli uygulanmasıyla ortaya çıkan sonuç gösterilmiştir. Karar ağacının katılım sayısını öncelikle 1 ve 1 olmayan şekilde ikiye ayırarak alt dallara doğru ilerlediğini görülmektedir. Burada dershaneye 1 yıllık katılımın önemli bir faktör teşkil ettiği görülmektedir.



Şekil 5.1 Dershaneye Katılım Süresinin Karar Ağacı Modeli İle Belirlenmesi

Ağaç yapısında alttaki dallara indikçe daha spesifik sonuçlara ulaşmak mümkündür. 12. Sınıf grubundan KATILIM değeri 1 ve 2 olmayan öğrencilerin yerleşme durumları Şekil 5.2’de verilmiştir.

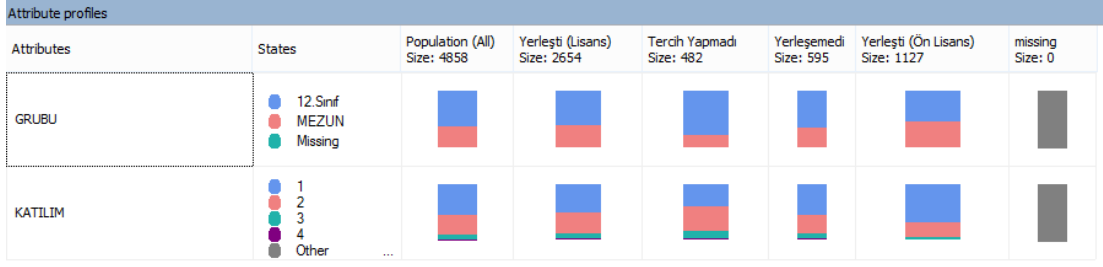


Şekil 5.2 “12. Sınıf” ve 2+ Yıl Dershaneye Katılan Öğrencilerin Durumu

Karar ağacı modeli katılım süresinin ağırlıklı etkisine bakarak ağaç yapısını “1” ve “1 olmayan” şeklinde ikiye ayırmıştır. Bu da eğitim programının değerlendirilmesi açısından eğitim süresinin 1 yıl olmasının etkisinin üzerinde durulabileceğini ifade etmektedir.

5.2 BAYES SINIFLANDIRICILARI

Belli kriterlere göre sınıflandırma yapmak hedef kitlenin eğilimleri hakkında bilgi verir. Şekil 5.3’de katılım süresi ve öğrencinin mezun ya da son sınıf olmasına göre gerçekleştirilen sınıflandırma, öğrenciye hangi eğitim programının önerilebileceğini göstermektedir.



Şekil 5.3 Katılım Süresi ve Grubun Bayes Sınıflandırma Sonucu Yerleşme Durumuna Etkisi

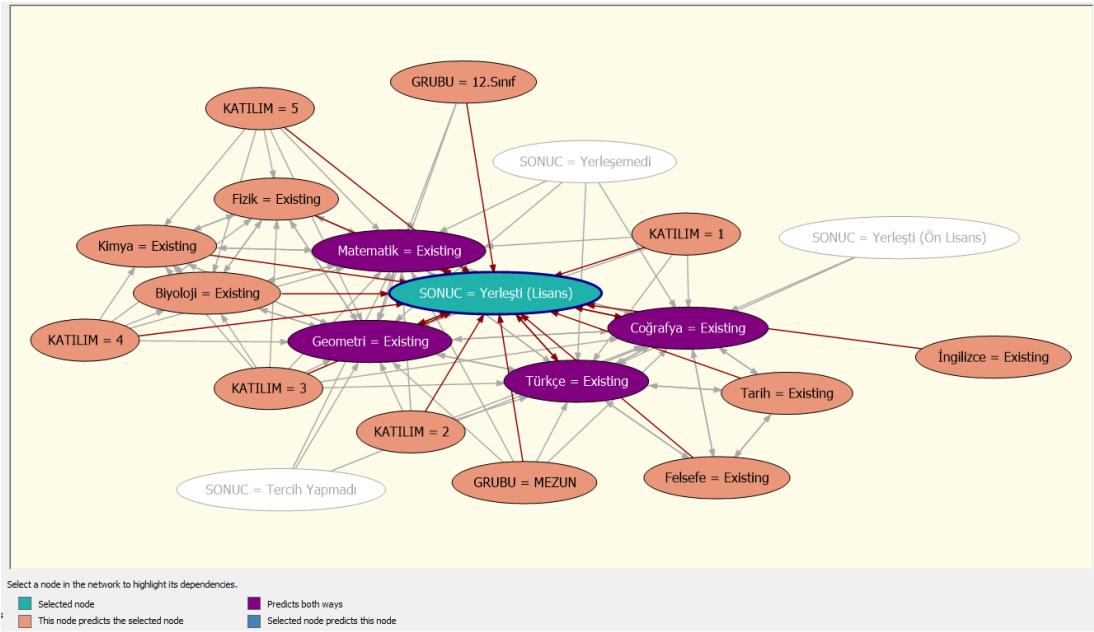
5.3 BİRLİKTELİK KURALLARI

Tablo 5.1’de lisans bölümlerine yerleşmede, destek sayısı ve güven seviyesi birliktelik kuralına göre hesaplandığında yerleşme durumuna etkisini görebilmekteyiz.

Tablo 5.1 Lisans Bölümlerine Yerleşmede Birliktelik Kurallarının Etkisi

Destek Sayısı	Güven Seviyesi	Birliktelik Kuralı
14	0,88	KATILIM = 5, Kimya -> SONUC = Yerleşti (Lisans)
14	0,88	KATILIM = 5, Biyoloji -> SONUC = Yerleşti (Lisans)
14	0,88	KATILIM = 5, Fizik -> SONUC = Yerleşti (Lisans)
47	0,81	KATILIM = 4, Fizik -> SONUC = Yerleşti (Lisans)
47	0,81	KATILIM = 4, Kimya -> SONUC = Yerleşti (Lisans)
47	0,81	KATILIM = 4, Biyoloji -> SONUC = Yerleşti (Lisans)
82	0,78	KATILIM = 1, İngilizce -> SONUC = Yerleşti (Lisans)
13	0,77	KATILIM = 5, Geometri -> SONUC = Yerleşti (Lisans)
13	0,77	KATILIM = 5, Matematik -> SONUC = Yerleşti (Lisans)
362	0,76	KATILIM = 2, Biyoloji -> SONUC = Yerleşti (Lisans)
285	0,45	KATILIM = 1, Felsefe -> SONUC = Yerleşti (Lisans)
285	0,45	KATILIM = 1, Tarih -> SONUC = Yerleşti (Lisans)
171	0,45	KATILIM = 2, Tarih -> SONUC = Yerleşti (Lisans)
171	0,45	KATILIM = 2, Felsefe -> SONUC = Yerleşti (Lisans)
30	0,42	Felsefe, Matematik -> SONUC = Yerleşti (Lisans)
30	0,42	Tarih, Geometri -> SONUC = Yerleşti (Lisans)
30	0,42	Felsefe, Geometri -> SONUC = Yerleşti (Lisans)
30	0,42	Tarih, Matematik -> SONUC = Yerleşti (Lisans)

1723	1,00	Matematik, SONUC = Yerleşti (Lisans) -> Geometri
1253	1,00	Türkçe, SONUC = Yerleşti (Lisans) -> Coğrafya
868	1,00	Fizik, SONUC = Yerleşti (Lisans) -> Biyoloji
868	1,00	Fizik, SONUC = Yerleşti (Lisans) -> Kimya
868	1,00	Biyoloji, SONUC = Yerleşti (Lisans) -> Kimya
493	1,00	Tarih, SONUC = Yerleşti (Lisans) -> Felsefe
493	1,00	Felsefe, SONUC = Yerleşti (Lisans) -> Coğrafya
493	1,00	Tarih, SONUC = Yerleşti (Lisans) -> Coğrafya



Şekil 5.4 Lisans Bölümlerine Yerleşmede Birlikte Kurallarının Ağırlıklı Etkisi

Şekil 5.4'de lisans bölümüne yerleşmede ağırlığı olan branş derslerin rolü görülmektedir. Mor renkte görülen matematik, geometri, türkçe ve coğrafya dersleri öğrencinin lisans bölümlerine yerleşmede daha fazla etkin rol oynamaktadır. Turuncu renkte görülen katılım süresi ve branş derslerin rolünün daha az etkin olduğu görülmektedir.

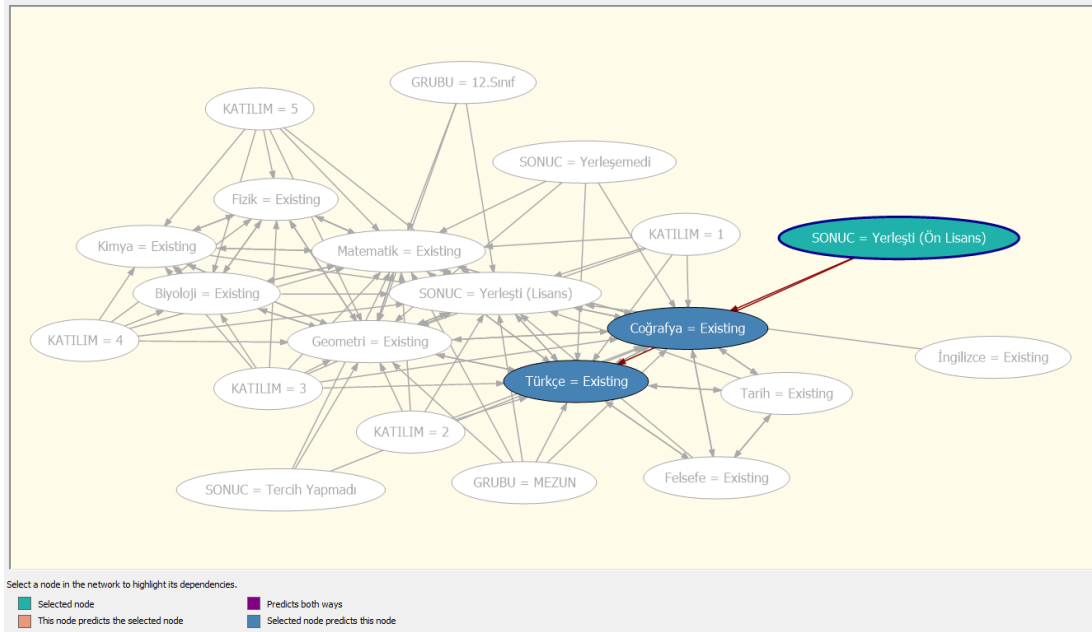
Lisans bölümüne yerleşmede güven seviyesine bakıldığında matematik dersinden başarılı olan bir öğrencinin geometri dersinde de başarı gösterdiği %100 kesindir.

Tablo 5.2’de Ön lisans bölümlerine yerleşmede, destek sayısı ve güven seviyesi birliktelik kuralına göre hesaplandığında yerleşme durumuna etkisini görebilmekteyiz.

Tablo 5.2 Ön Lisans Bölümlerine Yerleşmede Birliktelik Kurallarının Etkisi

Destek Sayısı	Güven Seviyesi	Birliktelik Kuralı
139	1,00	SONUC = Yerleşti (Ön Lisans), Kimya = Existing -> Fizik = Existing
338	1,00	Tarih = Existing, SONUC = Yerleşti (Ön Lisans) -> Coğrafya = Existing
340	1,00	SONUC = Yerleşti (Ön Lisans), Matematik = Existing -> Geometri = Existing
139	1,00	SONUC = Yerleşti (Ön Lisans), Biyoloji = Existing -> Fizik = Existing
139	1,00	SONUC = Yerleşti (Ön Lisans), Fizik = Existing -> Biyoloji = Existing
340	1,00	SONUC = Yerleşti (Ön Lisans), Geometri = Existing -> Matematik = Existing
338	1,00	Felsefe = Existing, SONUC = Yerleşti (Ön Lisans) -> Tarih = Existing
338	1,00	Tarih = Existing, SONUC = Yerleşti (Ön Lisans) -> Felsefe = Existing
139	1,00	SONUC = Yerleşti (Ön Lisans), Fizik = Existing -> Kimya = Existing
139	1,00	SONUC = Yerleşti (Ön Lisans), Kimya = Existing -> Biyoloji = Existing
139	1,00	SONUC = Yerleşti (Ön Lisans), Biyoloji = Existing -> Kimya = Existing
338	1,00	Felsefe = Existing, SONUC = Yerleşti (Ön Lisans) -> Coğrafya = Existing
497	1,00	SONUC = Yerleşti (Ön Lisans), Türkçe = Existing -> Coğrafya = Existing
57	0,52	KATILIM = 3, SONUC = Yerleşti (Ön Lisans) -> Coğrafya = Existing
57	0,47	KATILIM = 3, SONUC = Yerleşti (Ön Lisans) -> Türkçe = Existing
11	0,45	KATILIM = 4, SONUC = Yerleşti (Ön Lisans) -> Coğrafya = Existing
11	0,45	KATILIM = 4, SONUC = Yerleşti (Ön Lisans) -> Türkçe = Existing
765	0,43	SONUC = Yerleşti (Ön Lisans), KATILIM = 1 -> Türkçe = Existing
292	0,43	SONUC = Yerleşti (Ön Lisans), KATILIM = 2 -> Türkçe = Existing
513	0,41	SONUC = Yerleşti (Ön Lisans), GRUBU = MEZUN -> Felsefe = Existing
513	0,41	SONUC = Yerleşti (Ön Lisans), GRUBU = MEZUN -> Tarih = Existing

Şekil 5.5’de ön lisans bölümüne yerleşmede ağırlığı olan branş derslerin rolü görülmektedir. Mavi renkte görülen türkçe ve coğrafya dersleri öğrencinin ön lisans bölümlerine yerleşmede daha fazla etkin rol oynamaktadır.



Şekil 5.5 Ön Lisans Bölümlerine Yerleşmede Birliktelik Kurallarının Ağırlıklı Etkisi

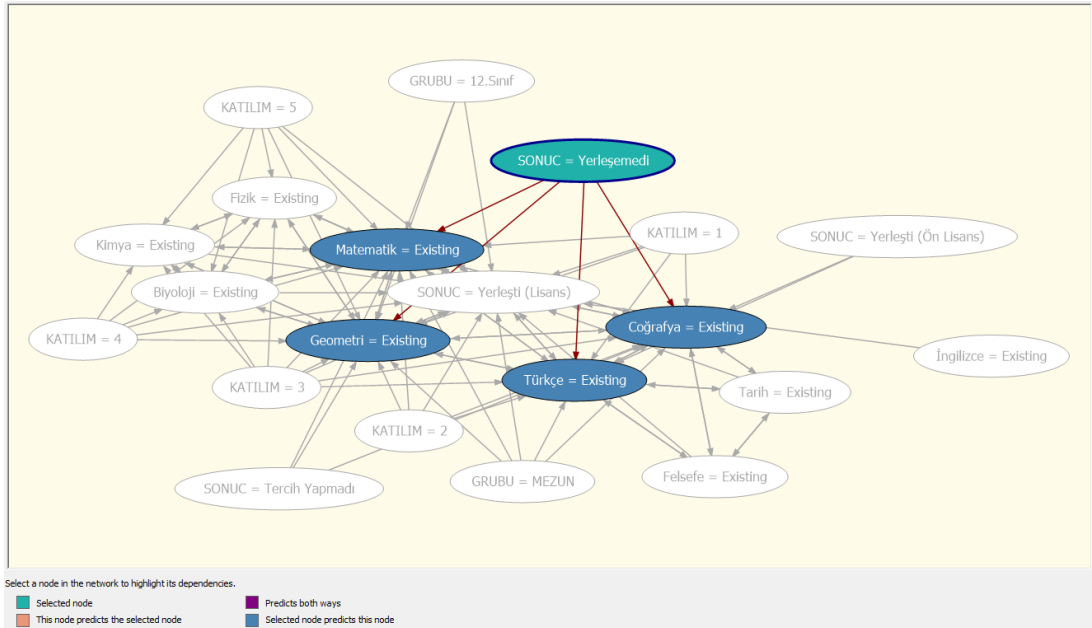
Tablo 5.3’de herhangi bir bölüme yerleşememede, destek sayısı ve güven seviyesi birliktelik kuralına göre hesaplandığında yerleşme durumuna etkisini görebilmekteyiz.

Tablo 5.3 Herhangi Bir Bölüme Yerleşememede Birliktelik Kurallarının Etkisi

Destek Sayısı	Güven Seviyesi	Birliktelik Kuralı
137	1,00	SONUC = Yerleşemedi, Tarih = Existing -> Coğrafya = Existing
319	1,00	SONUC = Yerleşemedi, Geometri = Existing -> Matematik = Existing
119	1,00	SONUC = Yerleşemedi, Fizik = Existing -> Kimya = Existing
137	1,00	SONUC = Yerleşemedi, Felsefe = Existing -> Tarih = Existing
137	1,00	SONUC = Yerleşemedi, Felsefe = Existing -> Coğrafya = Existing
119	1,00	SONUC = Yerleşemedi, Kimya = Existing -> Fizik = Existing
119	1,00	SONUC = Yerleşemedi, Kimya = Existing -> Biyoloji = Existing
119	1,00	SONUC = Yerleşemedi, Biyoloji = Existing -> Fizik = Existing
317	1,00	SONUC = Yerleşemedi, Türkçe = Existing -> Coğrafya = Existing
137	1,00	SONUC = Yerleşemedi, Tarih = Existing -> Felsefe = Existing
61	0,70	KATILIM = 3, SONUC = Yerleşemedi -> Matematik = Existing

61	0,70	KATILIM = 3, SONUC = Yerleşemedi -> Geometri = Existing
13	0,69	KATILIM = 4, SONUC = Yerleşemedi -> Matematik = Existing
13	0,69	KATILIM = 4, SONUC = Yerleşemedi -> Geometri = Existing
196	0,63	SONUC = Yerleşemedi, KATILIM = 2 -> Coğrafya = Existing
196	0,62	SONUC = Yerleşemedi, KATILIM = 2 -> Türkçe = Existing
386	0,54	SONUC = Yerleşemedi, GRUBU = 12.Sınıf -> Matematik = Existing
386	0,54	SONUC = Yerleşemedi, GRUBU = 12.Sınıf -> Geometri = Existing
209	0,51	SONUC = Yerleşemedi, GRUBU = MEZUN -> Geometri = Existing
209	0,51	SONUC = Yerleşemedi, GRUBU = MEZUN -> Matematik = Existing

Şekil 5.6'da herhangi bir bölüme yerleşememede ağırlığı olan branş derslerin rolü görülmektedir. Mavi renkte görülen matematik, geometri, türkçe ve coğrafya derslerinde başarısız olmak öğrencinin herhangi bir bölüme yerleşememede daha fazla etkin rol oynamaktadır.



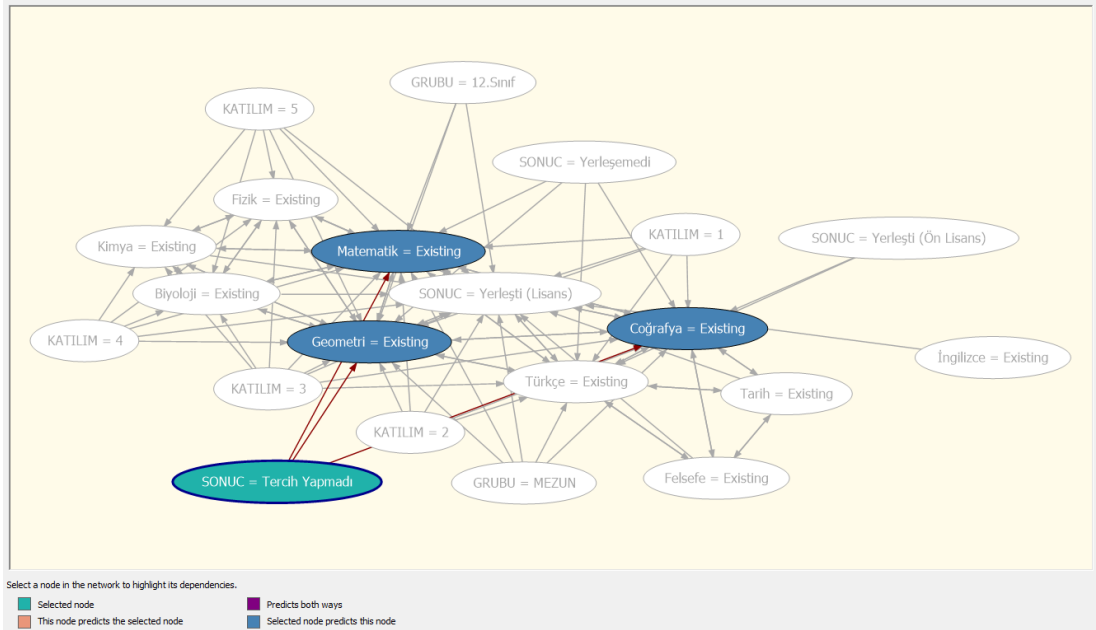
Şekil 5.6 Herhangi Bir Bölüme Yerleşememede Birliktelik Kurallarının Ağırlıklı Etkisi

Tablo 5.4'de herhangi bir tercih yapmamada, destek sayısı ve güven seviyesi birliktelik kuralına göre hesaplandığında yerleşme durumuna etkisi görülmektedir.

Tablo 5.4 Herhangi Bir Tercih Yapmamada Birliktelik Kurallarının Etkisi

Destek Sayısı	Güven Seviyesi	Birliktelik Kuralı
100	1	SONUC = Tercih Yapmadı, Biyoloji = Existing -> Fizik = Existing
100	1	SONUC = Tercih Yapmadı, Biyoloji = Existing -> Kimya = Existing
110	1	SONUC = Tercih Yapmadı, Felsefe = Existing -> Tarih = Existing
110	1	SONUC = Tercih Yapmadı, Felsefe = Existing -> Coğrafya = Existing
100	1	SONUC = Tercih Yapmadı, Fizik = Existing -> Kimya = Existing
100	0,9	SONUC = Tercih Yapmadı, Fizik = Existing -> Matematik = Existing
100	0,9	SONUC = Tercih Yapmadı, Fizik = Existing -> Geometri = Existing
14	0,786	KATILIM = 4, SONUC = Tercih Yapmadı -> Geometri = Existing
14	0,5	KATILIM = 4, SONUC = Tercih Yapmadı -> Coğrafya = Existing
212	0,434	SONUC = Tercih Yapmadı, KATILIM = 2 -> Matematik = Existing
67	0,418	KATILIM = 3, SONUC = Tercih Yapmadı -> Coğrafya = Existing

Şekil 5.7'de herhangi bir tercih yapmamada ağırlığı olan branş derslerin rolü görülmektedir. Mavi renkte görülen matematik, geometri ve coğrafya derslerinin tercih yapmamada daha fazla etkin rol oynamaktadır.



Şekil 5.7 Herhangi Bir Tercih Yapmamada Birliktelik Kurallarının Ağırlıklı Etkisi

Şekil 5.8'de Business Intelligence Development Studio programı kullanılarak belirlenen birliktelik kuralları ve güven seviyeleri gösterilmektedir.

“Yerleşti” kriterine göre filtrelediğimizde lisans bölümlerine yerleşen öğrencilerden fizik, kimya ve biyoloji derslerinde başarılı olan öğrencilerin yine bu derslerin kombinasyonunda da başarılı oldukları görünmektedir. Tarih ve felsefe derslerinde başarılı olan öğrencilerin türkçe dersinde de başarılı oldukları saptanmıştır.

Ön lisans bölümlerine yerleşmede Tarih ve felsefe derslerinde başarılı olan öğrencilerin ise coğrafya dersinde daha başarılı oldukları gözükmektedir.

Şekil 5.9'da veri seti içerisindeki değerlerin birbirleriyle olan kombinasyonlarını ve destek sayılarını görebilmekteyiz. Destek sayısı o kombinasyonun veri seti içerisinde ne kadar sıklıkla geçtiğini göstermektedir.

Bağımlılık ağı açısından incelendiğinde destek sayısı daha büyük olan branş derslerinin lisans bölümüne yerleşmede matematik, geometri, coğrafya ve Türkçe derslerinin daha etkili olduğu belirlenmiştir. Katılım sayısına baktığımızda dershaneye 1 ve 2 yıl devam etmenin üniversiteye yerleşmede daha etkili olduğu söylenebilir.

Rules		Itemsets	Dependency Network
Minimum probability:	0,40	Filter Rule:	SONUC = Yerleşti
Minimum importance:	-0,12	Show:	Show attribute name and value
<input type="checkbox"/> Show long name		Maximum rows:	2000
Probability	Rule		
1,000	Felsefe = Existing, SONUC = Yerleşti (Ön Lisans) -> Coğrafya = Existing		
1,000	SONUC = Yerleşti (Ön Lisans), Kimya = Existing -> Biyoloji = Existing		
1,000	Tarih = Existing, SONUC = Yerleşti (Ön Lisans) -> Coğrafya = Existing		
1,000	Tarih = Existing, SONUC = Yerleşti (Lisans) -> Coğrafya = Existing		
1,000	SONUC = Yerleşti (Ön Lisans), Fizik = Existing -> Kimya = Existing		
1,000	Tarih = Existing, SONUC = Yerleşti (Ön Lisans) -> Felsefe = Existing		
1,000	Türkçe = Existing, SONUC = Yerleşti (Lisans) -> Coğrafya = Existing		
1,000	Felsefe = Existing, SONUC = Yerleşti (Lisans) -> Coğrafya = Existing		
1,000	Geometri = Existing, SONUC = Yerleşti (Lisans) -> Matematik = Existing		
1,000	Fizik = Existing, SONUC = Yerleşti (Lisans) -> Biyoloji = Existing		
1,000	Biyoloji = Existing, SONUC = Yerleşti (Lisans) -> Kimya = Existing		
1,000	SONUC = Yerleşti (Ön Lisans), Kimya = Existing -> Fizik = Existing		
1,000	SONUC = Yerleşti (Ön Lisans), Biyoloji = Existing -> Kimya = Existing		
1,000	SONUC = Yerleşti (Ön Lisans), Biyoloji = Existing -> Fizik = Existing		
1,000	SONUC = Yerleşti (Ön Lisans), Fizik = Existing -> Biyoloji = Existing		
1,000	SONUC = Yerleşti (Ön Lisans), Matematik = Existing -> Geometri = Existing		
1,000	SONUC = Yerleşti (Ön Lisans), Geometri = Existing -> Matematik = Existing		
1,000	Matematik = Existing, SONUC = Yerleşti (Lisans) -> Geometri = Existing		
1,000	SONUC = Yerleşti (Ön Lisans), Türkçe = Existing -> Coğrafya = Existing		
1,000	Fizik = Existing, SONUC = Yerleşti (Lisans) -> Kimya = Existing		
1,000	Kimya = Existing, SONUC = Yerleşti (Lisans) -> Fizik = Existing		
1,000	Felsefe = Existing, SONUC = Yerleşti (Lisans) -> Tarih = Existing		
1,000	Biyoloji = Existing, SONUC = Yerleşti (Lisans) -> Fizik = Existing		
1,000	Kimya = Existing, SONUC = Yerleşti (Lisans) -> Biyoloji = Existing		
1,000	Felsefe = Existing, SONUC = Yerleşti (Ön Lisans) -> Tarih = Existing		
1,000	Tarih = Existing, SONUC = Yerleşti (Lisans) -> Felsefe = Existing		
0,981	Coğrafya = Existing, SONUC = Yerleşti (Lisans) -> Türkçe = Existing		
0,980	Biyoloji = Existing, SONUC = Yerleşti (Lisans) -> Matematik = Existing		
0,980	Fizik = Existing, SONUC = Yerleşti (Lisans) -> Matematik = Existing		
0,980	Kimya = Existing, SONUC = Yerleşti (Lisans) -> Geometri = Existing		
0,980	Biyoloji = Existing, SONUC = Yerleşti (Lisans) -> Geometri = Existing		
0,980	Fizik = Existing, SONUC = Yerleşti (Lisans) -> Geometri = Existing		
0,980	Kimya = Existing, SONUC = Yerleşti (Lisans) -> Matematik = Existing		
0,951	Tarih = Existing, SONUC = Yerleşti (Lisans) -> Türkçe = Existing		
0,951	Felsefe = Existing, SONUC = Yerleşti (Lisans) -> Türkçe = Existing		
0,941	SONUC = Yerleşti (Ön Lisans), Coğrafya = Existing -> Türkçe = Existing		
0,908	Tarih = Existing, SONUC = Yerleşti (Ön Lisans) -> Türkçe = Existing		
0,908	Felsefe = Existing, SONUC = Yerleşti (Ön Lisans) -> Türkçe = Existing		

Selected rules: 3/232

Şekil 5.8 Yerleşme Durumuna Etki Eden Birliktelik Kurallarının Güven Seviyeleri

Rules	Itemsets	Dependency Network
Minimum support:	4	Filter Itemset: SONUC
Minimum itemset size:	0	Show: Show attribute name and value
Maximum rows:	2000	<input type="checkbox"/> Show long name
Support	Size	Itemset
2654	1	SONUC = Yerleşti (Lisans)
1723	2	Geometri = Existing, SONUC = Yerleşti (Lisans)
1723	3	Matematik = Existing, Geometri = Existing, SONUC = Yerleşti (Lisans)
1723	2	Matematik = Existing, SONUC = Yerleşti (Lisans)
1637	2	SONUC = Yerleşti (Lisans), GRUBU = 12.Sınıf
1319	2	KATILIM = 1, SONUC = Yerleşti (Lisans)
1277	2	Coğrafya = Existing, SONUC = Yerleşti (Lisans)
1253	3	Türkçe = Existing, Coğrafya = Existing, SONUC = Yerleşti (Lisans)
1253	2	Türkçe = Existing, SONUC = Yerleşti (Lisans)
1127	1	SONUC = Yerleşti (Ön Lisans)
1096	3	Geometri = Existing, SONUC = Yerleşti (Lisans), GRUBU = 12.Sınıf
1096	3	Matematik = Existing, SONUC = Yerleşti (Lisans), GRUBU = 12.Sınıf
1017	2	GRUBU = MEZUN, SONUC = Yerleşti (Lisans)
1003	2	KATILIM = 2, SONUC = Yerleşti (Lisans)
868	3	Kimya = Existing, Biyoloji = Existing, SONUC = Yerleşti (Lisans)
868	3	Fizik = Existing, Biyoloji = Existing, SONUC = Yerleşti (Lisans)
868	2	Biyoloji = Existing, SONUC = Yerleşti (Lisans)
868	3	Kimya = Existing, Fizik = Existing, SONUC = Yerleşti (Lisans)
868	2	Kimya = Existing, SONUC = Yerleşti (Lisans)
868	2	Fizik = Existing, SONUC = Yerleşti (Lisans)
851	3	Biyoloji = Existing, Geometri = Existing, SONUC = Yerleşti (Lisans)
851	3	Fizik = Existing, Geometri = Existing, SONUC = Yerleşti (Lisans)
851	3	Fizik = Existing, Matematik = Existing, SONUC = Yerleşti (Lisans)
851	3	Kimya = Existing, Matematik = Existing, SONUC = Yerleşti (Lisans)
851	3	Kimya = Existing, Geometri = Existing, SONUC = Yerleşti (Lisans)
851	3	Biyoloji = Existing, Matematik = Existing, SONUC = Yerleşti (Lisans)
778	3	Matematik = Existing, KATILIM = 1, SONUC = Yerleşti (Lisans)
778	3	Geometri = Existing, KATILIM = 1, SONUC = Yerleşti (Lisans)
776	3	KATILIM = 2, SONUC = Yerleşti (Lisans), GRUBU = 12.Sınıf
765	2	SONUC = Yerleşti (Ön Lisans), KATILIM = 1
763	3	Coğrafya = Existing, Matematik = Existing, SONUC = Yerleşti (Lisans)
763	3	Türkçe = Existing, Matematik = Existing, SONUC = Yerleşti (Lisans)
763	3	Türkçe = Existing, Geometri = Existing, SONUC = Yerleşti (Lisans)
763	3	Coğrafya = Existing, Geometri = Existing, SONUC = Yerleşti (Lisans)
693	3	Coğrafya = Existing, KATILIM = 1, SONUC = Yerleşti (Lisans)
687	3	KATILIM = 2, Matematik = Existing, SONUC = Yerleşti (Lisans)
687	3	KATILIM = 2, Geometri = Existing, SONUC = Yerleşti (Lisans)

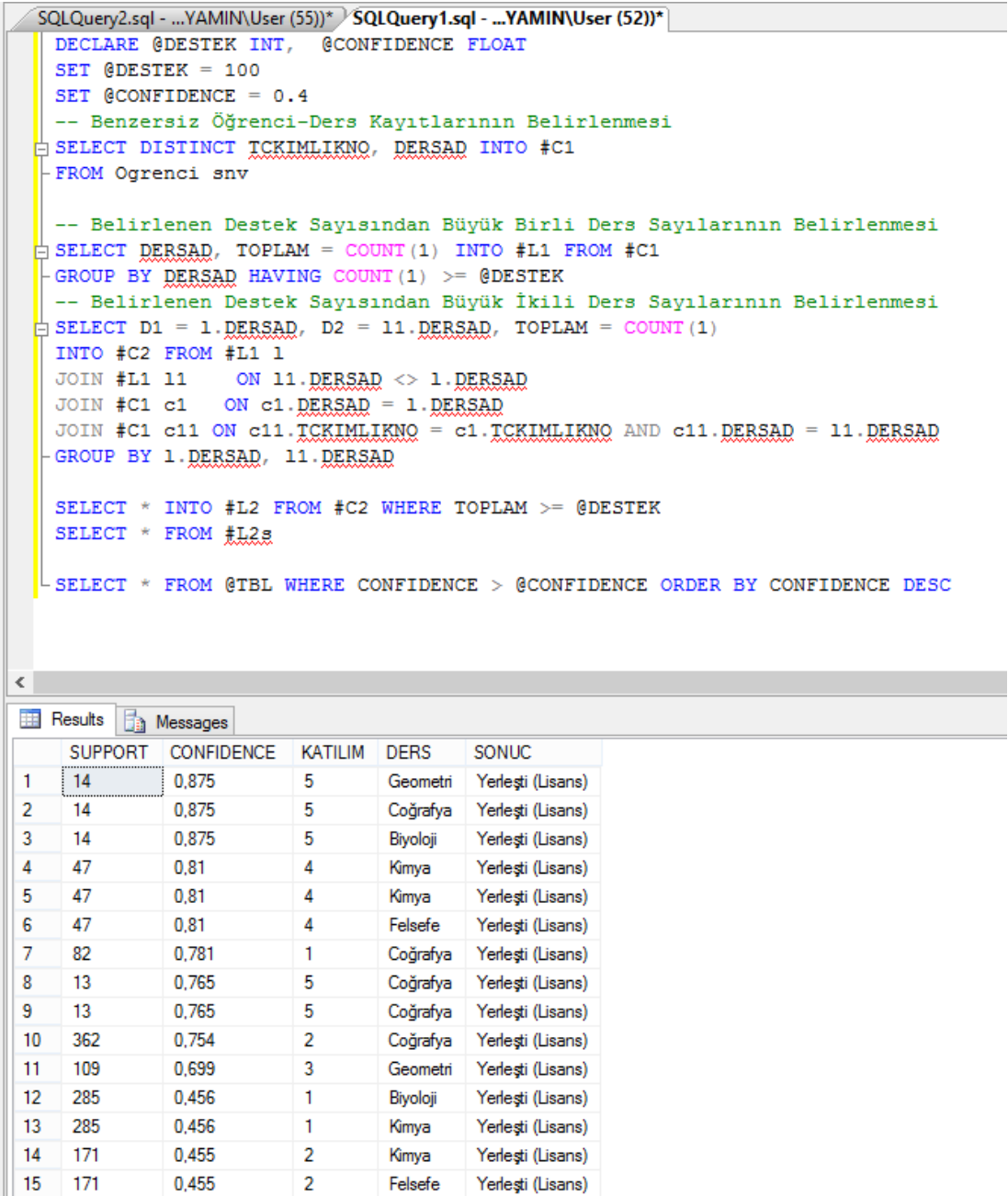
Itemsets: 429

Şekil 5.9 Yerleşme Durumuna Etki Eden Kombinasyonların Destek Sayıları

5.3.1 Apriori Algoritması

Veri tabanında Öğrenci Yerleşme Sonuçları, TC Kimlik Numarası ile Sınavda çözemediği ders ve katılım sayısı birlikte tutulmaktadır.

Apriori algoritmamızın ilk adımı, her dersin ve katılım sayısının destek sayısını, yani kaç kere listede geçtiğini saymak olacaktır.



```
SQLQuery2.sql - ...YAMIN\User (55))* SQLQuery1.sql - ...YAMIN\User (52))*
DECLARE @DESTEK INT, @CONFIDENCE FLOAT
SET @DESTEK = 100
SET @CONFIDENCE = 0.4
-- Benzersiz Öğrenci-Ders Kayıtlarının Belirlenmesi
SELECT DISTINCT TCKIMLIKNO, DERSAD INTO #C1
FROM Oğrenci snv

-- Belirlenen Destek Sayısından Büyük Birli Ders Sayılarının Belirlenmesi
SELECT DERSAD, TOPLAM = COUNT(1) INTO #L1 FROM #C1
GROUP BY DERSAD HAVING COUNT(1) >= @DESTEK
-- Belirlenen Destek Sayısından Büyük İkili Ders Sayılarının Belirlenmesi
SELECT D1 = 1.DERSAD, D2 = 11.DERSAD, TOPLAM = COUNT(1)
INTO #C2 FROM #L1 1
JOIN #L1 11 ON 11.DERSAD <> 1.DERSAD
JOIN #C1 c1 ON c1.DERSAD = 1.DERSAD
JOIN #C1 c11 ON c11.TCKIMLIKNO = c1.TCKIMLIKNO AND c11.DERSAD = 11.DERSAD
GROUP BY 1.DERSAD, 11.DERSAD

SELECT * INTO #L2 FROM #C2 WHERE TOPLAM >= @DESTEK
SELECT * FROM #L2s

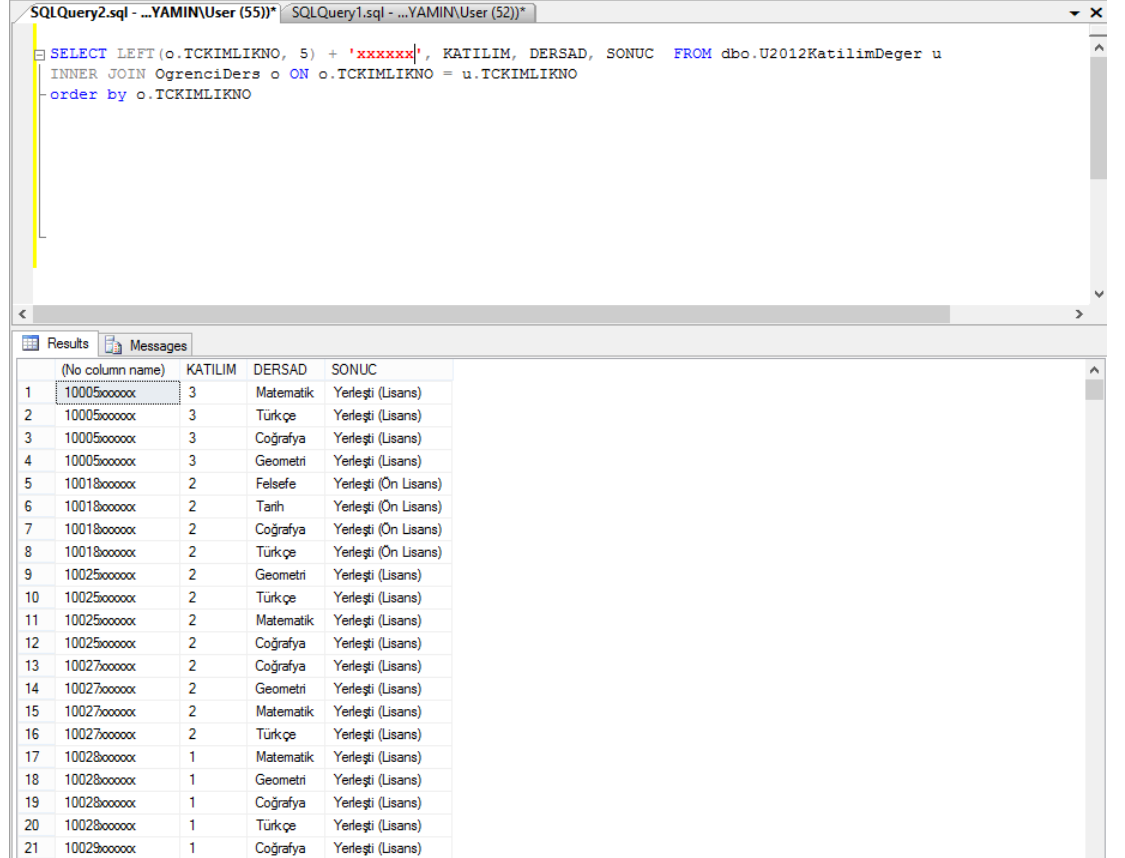
SELECT * FROM @TBL WHERE CONFIDENCE > @CONFIDENCE ORDER BY CONFIDENCE DESC
```

	SUPPORT	CONFIDENCE	KATILIM	DERS	SONUC
1	14	0,875	5	Geometri	Yerleşti (Lisans)
2	14	0,875	5	Coğrafya	Yerleşti (Lisans)
3	14	0,875	5	Biyoloji	Yerleşti (Lisans)
4	47	0,81	4	Kimya	Yerleşti (Lisans)
5	47	0,81	4	Kimya	Yerleşti (Lisans)
6	47	0,81	4	Felsefe	Yerleşti (Lisans)
7	82	0,781	1	Coğrafya	Yerleşti (Lisans)
8	13	0,765	5	Coğrafya	Yerleşti (Lisans)
9	13	0,765	5	Coğrafya	Yerleşti (Lisans)
10	362	0,754	2	Coğrafya	Yerleşti (Lisans)
11	109	0,699	3	Geometri	Yerleşti (Lisans)
12	285	0,456	1	Biyoloji	Yerleşti (Lisans)
13	285	0,456	1	Kimya	Yerleşti (Lisans)
14	171	0,455	2	Kimya	Yerleşti (Lisans)
15	171	0,455	2	Felsefe	Yerleşti (Lisans)

Şekil 5.10 Apriori Algoritması Veri Destek Sayısı ve Güven Seviyeleri

Şekil 5.10'da katılım sayısı ve derslerin, yerleşme durumuna etkisi incelenerek destek sayısı ve güven seviyeleri tespit edilmiştir.

Şekil 5.11'de ise öğrencilerin sorumlu olduğu branş dersler ve katılım sayılarının yer aldığı bir örnek liste gösterilmektedir.



```
SELECT LEFT(o.TCKIMLIKNO, 5) + 'xxxxxx', KATILIM, DERSAD, SONUC FROM dbo.U2012KatilimDeger u
INNER JOIN OgrenciDers o ON o.TCKIMLIKNO = u.TCKIMLIKNO
-order by o.TCKIMLIKNO
```

(No column name)	KATILIM	DERSAD	SONUC
1 1000500000	3	Matematik	Yerleşti (Lisans)
2 1000500000	3	Türkçe	Yerleşti (Lisans)
3 1000500000	3	Coğrafya	Yerleşti (Lisans)
4 1000500000	3	Geometri	Yerleşti (Lisans)
5 1001800000	2	Felsefe	Yerleşti (Ön Lisans)
6 1001800000	2	Tarih	Yerleşti (Ön Lisans)
7 1001800000	2	Coğrafya	Yerleşti (Ön Lisans)
8 1001800000	2	Türkçe	Yerleşti (Ön Lisans)
9 1002500000	2	Geometri	Yerleşti (Lisans)
10 1002500000	2	Türkçe	Yerleşti (Lisans)
11 1002500000	2	Matematik	Yerleşti (Lisans)
12 1002500000	2	Coğrafya	Yerleşti (Lisans)
13 1002700000	2	Coğrafya	Yerleşti (Lisans)
14 1002700000	2	Geometri	Yerleşti (Lisans)
15 1002700000	2	Matematik	Yerleşti (Lisans)
16 1002700000	2	Türkçe	Yerleşti (Lisans)
17 1002800000	1	Matematik	Yerleşti (Lisans)
18 1002800000	1	Geometri	Yerleşti (Lisans)
19 1002800000	1	Coğrafya	Yerleşti (Lisans)
20 1002800000	1	Türkçe	Yerleşti (Lisans)
21 1002900000	1	Coğrafya	Yerleşti (Lisans)

Şekil 5.11 Öğrencilerin Sorumlu Olduğu Branş Ders ve Katılım Sayıları

İlk adımdan sonra daha önceden belirlediğimiz destek sayısı ya da güven seviyesi büyük olan değerleri alarak bir sonraki adıma geçer ve mevcut veriler arasındaki kombinasyonları belirleriz. Bu adımlar destek sayısı ya da güven seviyesinden büyük veriler arasında yeni kombinasyonlar bulunmayıncaya kadar devam eder.

5.4 KÜMELEME MODELLERİ

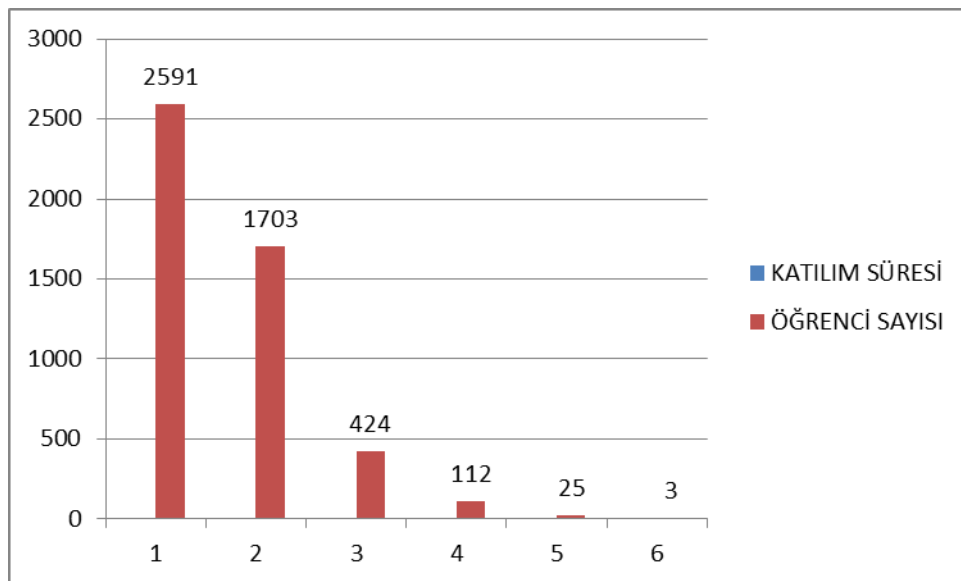
Şekil 5.12’de BIDS programının kümeleme modelinin kullanılmasıyla katılım ve grubunun input değişken olduğu ve yerleşme durumunu tahmin etmeye yönelik uygulama sonucu verilerin aktarıldığı kümeler gözükmemektedir.

Attributes		Cluster profiles								
Variables	States	Population (All) Size: 4858	Cluster 1 Size: 1313	Cluster 2 Size: 1267	Cluster 3 Size: 268	Cluster 4 Size: 106	Cluster 5 Size: 1324	Cluster 6 Size: 390	Cluster 7 Size: 190	
GRUBU	<ul style="list-style-type: none"> 12.Sinif MEZUN missing 									
KATILIM	<ul style="list-style-type: none"> 1 2 3 4 Other 									
SONUC	<ul style="list-style-type: none"> Yerleşti (Lisans) Yerleşti (On Lisans) Yerleşmedi Tercih Yapmadı Other 									

Şekil 5.12 Katılım ve Gruba Göre Yerleşme Durumunu Tahmine Dayalı Kümeleme Modeli

Dershaneye 1 veya 2 yıl devam eden öğrencilerin farklı kümelere aktarıldığı, katılım sayısı 3’den fazla olan öğrenciler olmasına rağmen yoğunluğu az olduğundan onlar için ayrı kümeler oluşturma ihtiyacının olmadığı saptanmıştır.

Grafik 5.1’de bu katılım sayılarının yoğunluğu görülebilmektedir.



Grafik 5.1 Katılım Sayısı Yoğunluğu

5.4.1 K-Ortalama Algoritması

K-Ortalama algoritması ayrıntılı olarak aşağıda anlatılmaktadır:

Adım 1	K = 7 K'nın 7 olması 7 kümeye ayrılacağını ifade eder
Adım 2	7 küme belirlendiği için yükseköğretim programlarını kazanan öğrenciler bazında rastgele 7 sayı belirlenerek 7 ayrı kümeye birer adet eklendi.
Adım 3	Birbirlerine en yakın olan sayılar aynı tabloda aktarılmıştır. Bu işlem ortalama değerleri aynı oluncaya kadar devam ettirilir.

```
--SELECT COUNT(1) FROM SKUME1
--SELECT COUNT(1) FROM SKUME2
--SELECT COUNT(1) FROM SKUME3
--SELECT COUNT(1) FROM SKUME4
--SELECT COUNT(1) FROM SKUME5
--SELECT COUNT(1) FROM SKUME6
--SELECT COUNT(1) FROM SKUME7

DROP TABLE #Tbl1
```

RN	DEGER	(No column name)	(No column name)	(No column name)	(No column name)	(No column name)	(No column name)	(No column name)	(No column name)	(No column name)	(No column name)	(No column name)	(No column name)	(No column name)	(No column name)	(No column name)	(No column name)	(No column name)								
1	1	55	5	5	10	11	15	28.4285714285714	75	72	75	305.4	13	13	53.1666666666667	5	6	11	12	28.4285714285714	28.4285714285714	74.6	305.4	454.666666666667	13	13
2	2	55	5	5	10	11	15	28.4285714285714	75	72	75	305.4	13	13	53.1666666666667	5	6	11	12	28.4285714285714	28.4285714285714	74.6	305.4	454.666666666667	13	13
3	3	55	5	5	10	11	15	28.4285714285714	75	72	75	305.4	13	13	53.1666666666667	5	6	11	12	28.4285714285714	28.4285714285714	74.6	305.4	454.666666666667	13	13
4	4	55	5	5	10	11	15	28.4285714285714	75	72	75	305.4	13	13	53.1666666666667	5	6	11	12	28.4285714285714	28.4285714285714	74.6	305.4	454.666666666667	13	13
5	5	55	5	5	10	11	15	28.4285714285714	75	72	75	305.4	13	13	53.1666666666667	5	6	11	12	28.4285714285714	28.4285714285714	74.6	305.4	454.666666666667	13	13
6	6	55	5	5	10	11	15	28.4285714285714	75	72	75	305.4	13	13	53.1666666666667	5	6	11	12	28.4285714285714	28.4285714285714	74.6	305.4	454.666666666667	13	13
7	7	55	5	5	10	11	15	28.4285714285714	75	72	75	305.4	13	13	53.1666666666667	5	6	11	12	28.4285714285714	28.4285714285714	74.6	305.4	454.666666666667	13	13
8	8	55	5	5	10	11	15	28.4285714285714	75	72	75	305.4	13	13	53.1666666666667	5	6	11	12	28.4285714285714	28.4285714285714	74.6	305.4	454.666666666667	13	13

Şekil 5.13 K-Ortalama Algoritması Küme Değerleri Ortalamaları

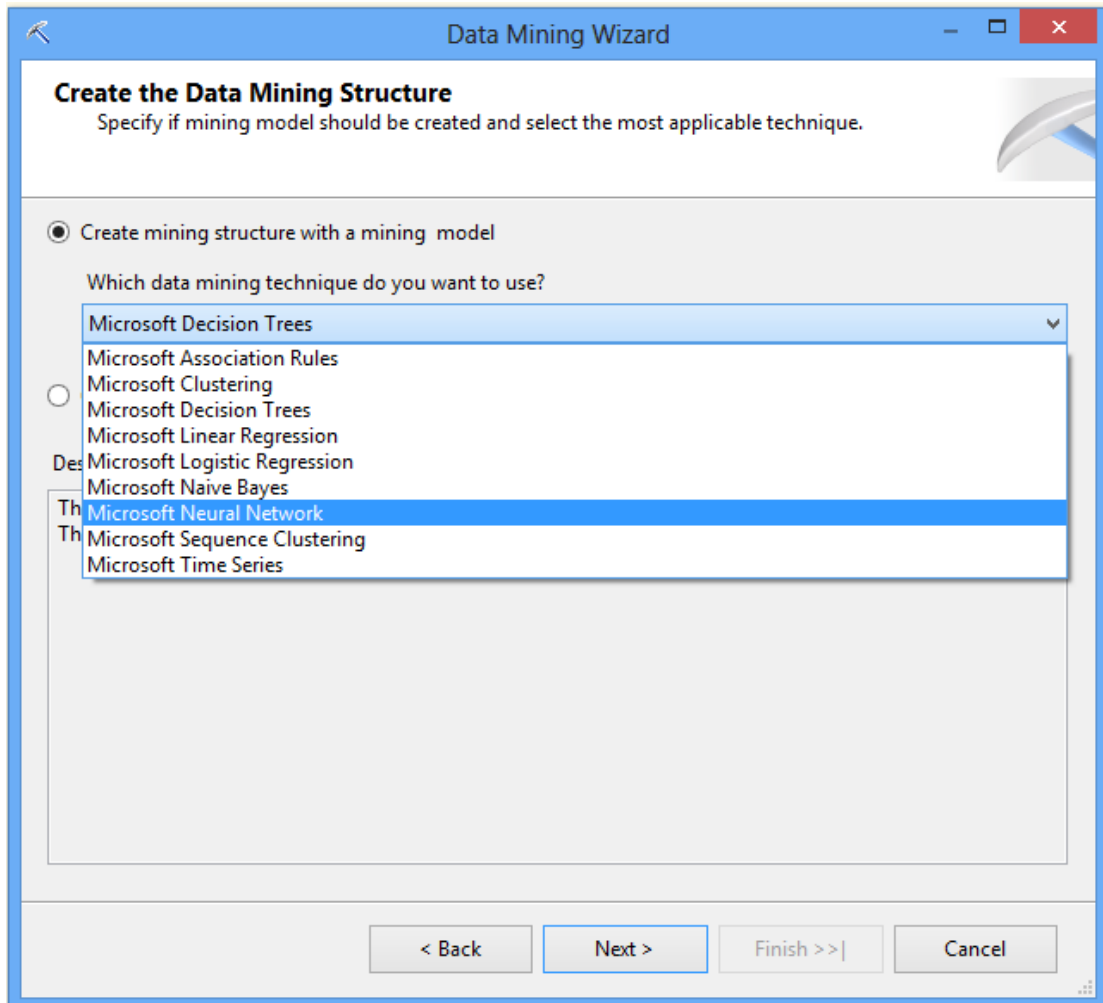
Şekil 5.13'de algoritmanın uygulanması sonucu oluşan ortalama değerlerinin son aşamada aynı olduğunu ve bu aşamadan sonra kümeleme işleminin bittiğini görüyoruz.

Algoritmanın uygulanmasından sonra 2 ve 3 olan katılım sayısına sahip öğrenciler ayrı kümelere aktarılmıştır. 2 ayrı kümeye sadece 2 ve 3 katılım sayısına sahip öğrenciler yer almıştır. Bu da katılım sayısı 2 ve 3 olan öğrencilerin benzer özellik taşıdığını göstermektedir.

BIDS programı ise dershaneye 1 ve 2 yıl devam eden öğrencilerin kendi aralarında benzer özelliklere sahip olduğunu tespit etmiştir.

5.5 BIDS MODEL KARŞILAŞTIRMA

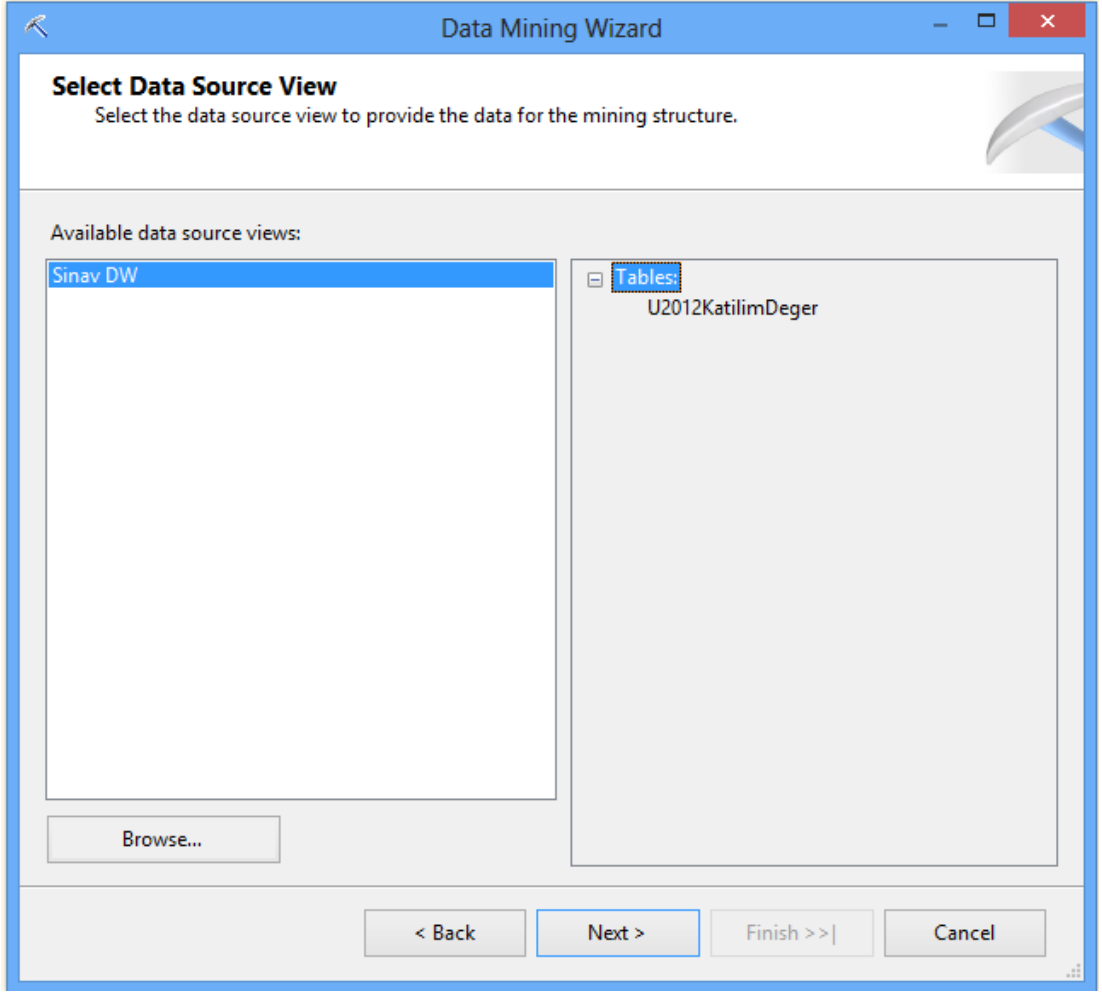
Bu bölümde Business Intelligence Development Studio adlı program kullanılarak veriler üzerine algoritmaların uygulanması şekillerle anlatılacaktır.



Şekil 5.14 VM Algoritma Belirlenmesi

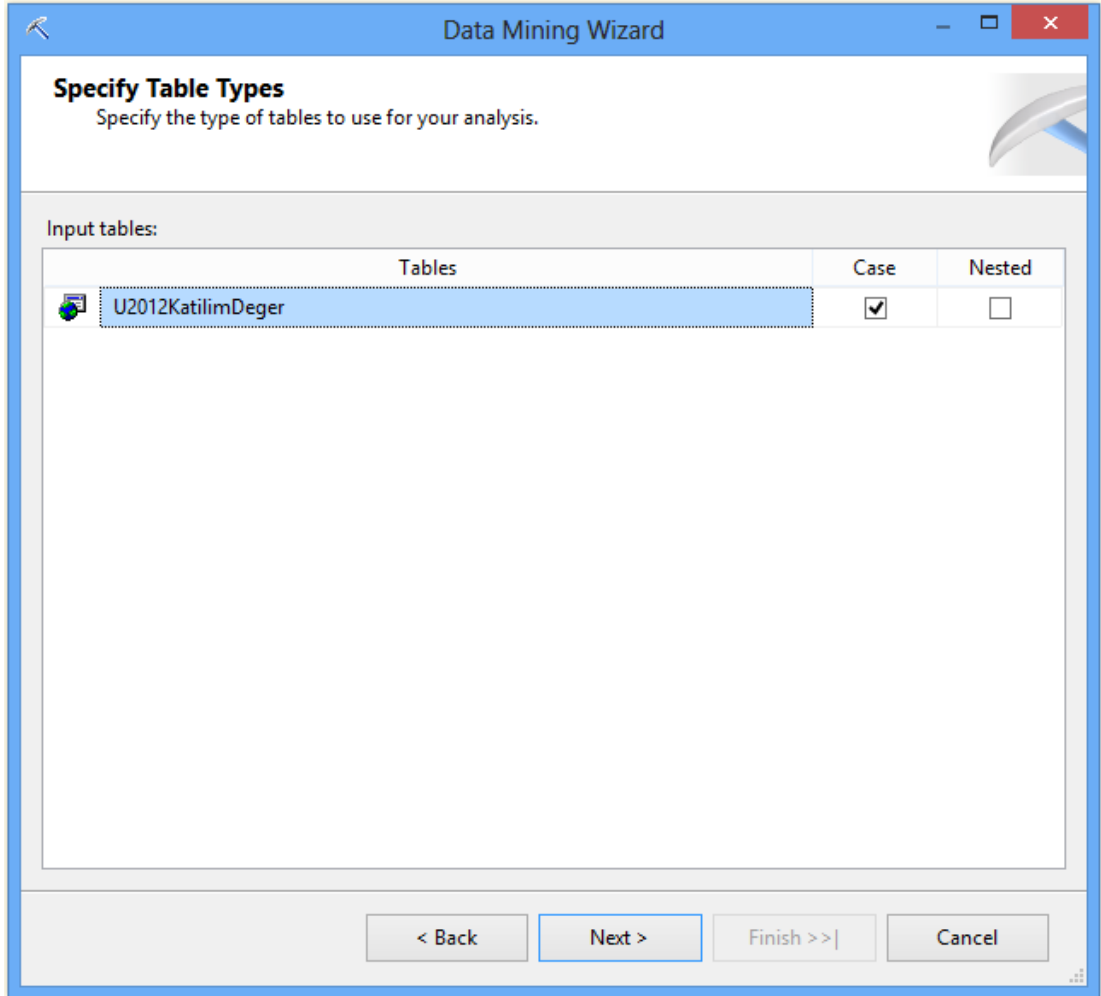
Şekil 5.14’de hangi VM algoritmasının uygulanacağı seçilir. Örneğin; “Microsoft Neural Network”.

“Data Mining Wizard” ekranından Şekil 5.15’de olduğu gibi veri kaynağının belirlenmesi sağlanır. Hangi tablolar üzerinde çalışılacağı buradan belirlenir.



Şekil 5.15 Veri Kaynağının Belirlenmesi

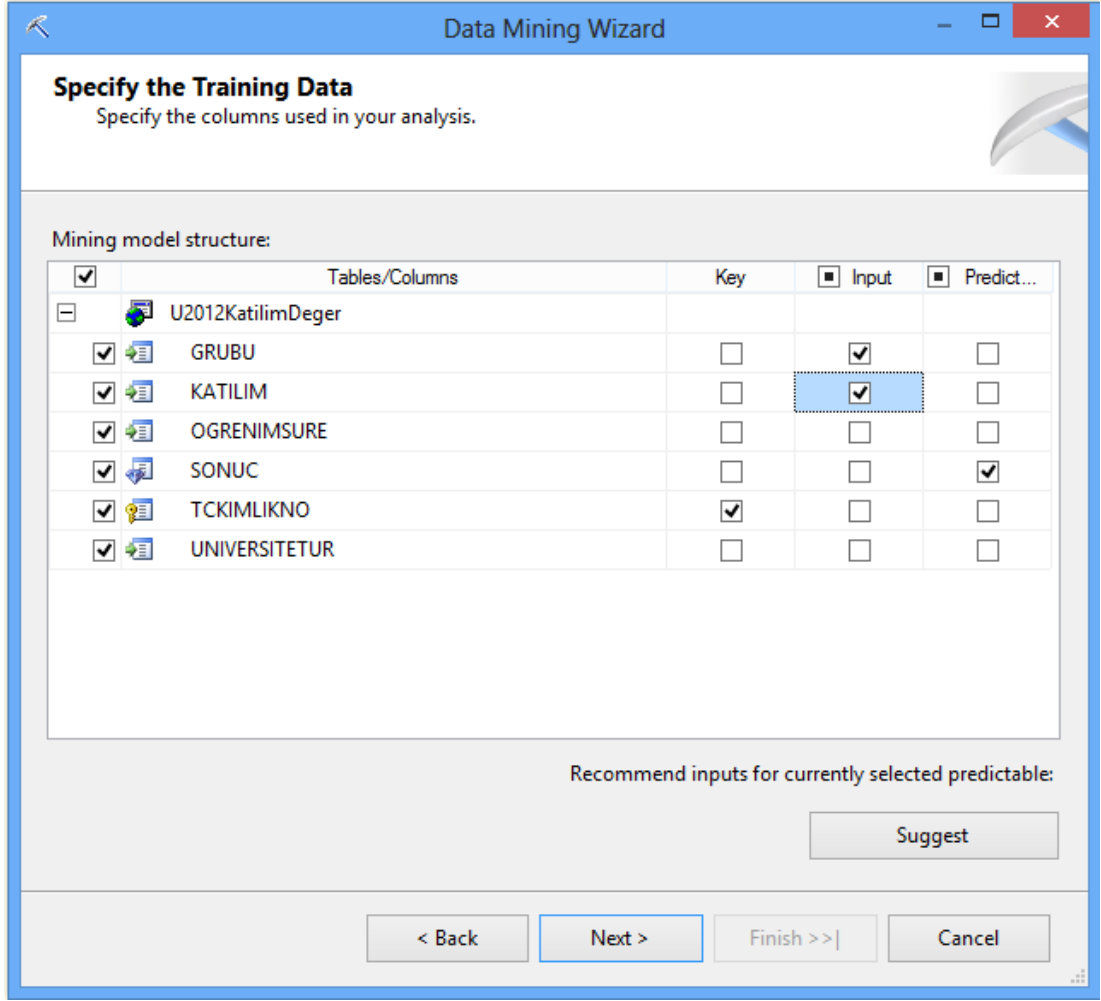
Bu ekranda kayıtlı tablo ya da daha önceden hazırladığımız sorgular bulunmaktadır. İstenilen tablo ya da hazır sorgular seçilebilir.



Şekil 5.16 Ana Tablo ya da Yardımcı Tabloların Belirlenmesi

Bu ekranda “Case” ana tabloyu “Nested” ise yardımcı tabloları belirlemek için seçim yapıyoruz. Tek bir tablo ya da sorgu üzerinde çalışacak ise “Nested” seçeneğini işaretlemeye gerek yoktur.

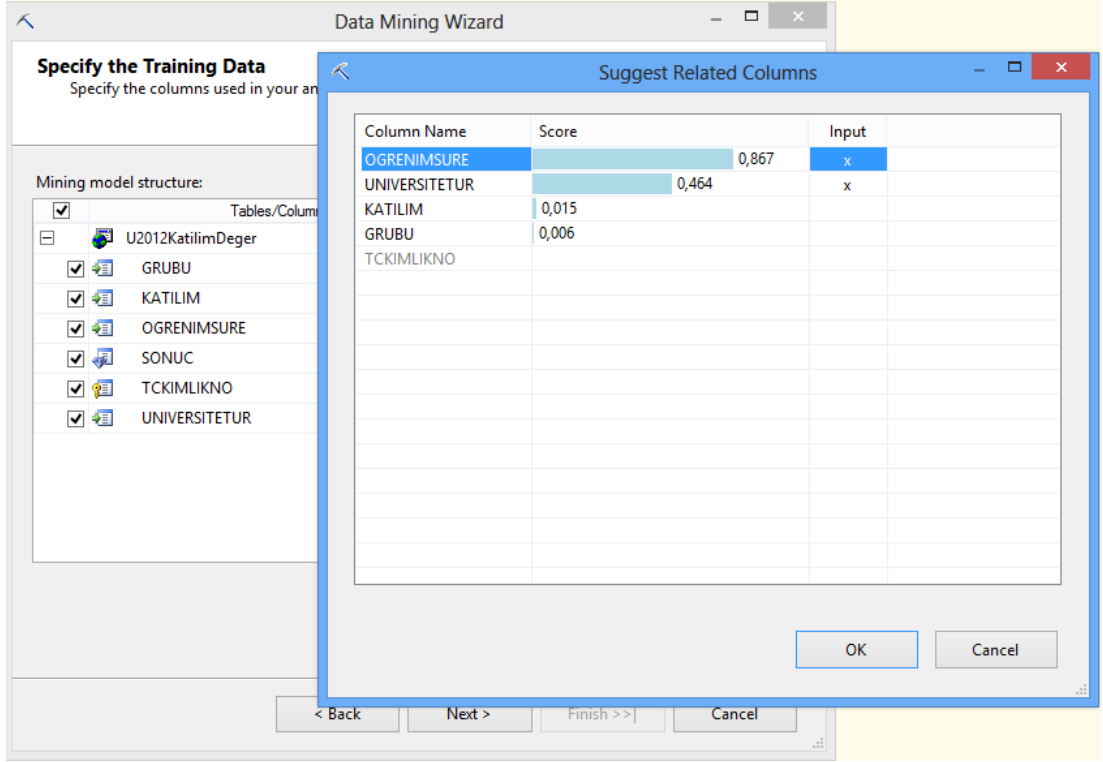
Şekil 5.16’da görüldüğü üzere “U2012KatilimDeger” tablosu tek/ana tablo olarak seçilmiştir. Çünkü sadece tek bir tablo üzerinde çalışılacaktır.



Şekil 5.17 Algoritma İçin Sütunların Belirlenmesi

Bu ekranda işlem yapmak istenilen sütunlar seçilir. En az 1 adet “key” tanımlanmalıdır. Giriş değerleri için “input” bölümünden istediğimiz giriş değerleri seçilir. Output (çıkıtı) için “Predictable” bölümünden elde edilmesi istenilen sütunlar seçilir. Çıkıtı için “SONUC” sütunu seçilir. Giriş ve çıkış değerleri için birden fazla sütün seçilebilir.

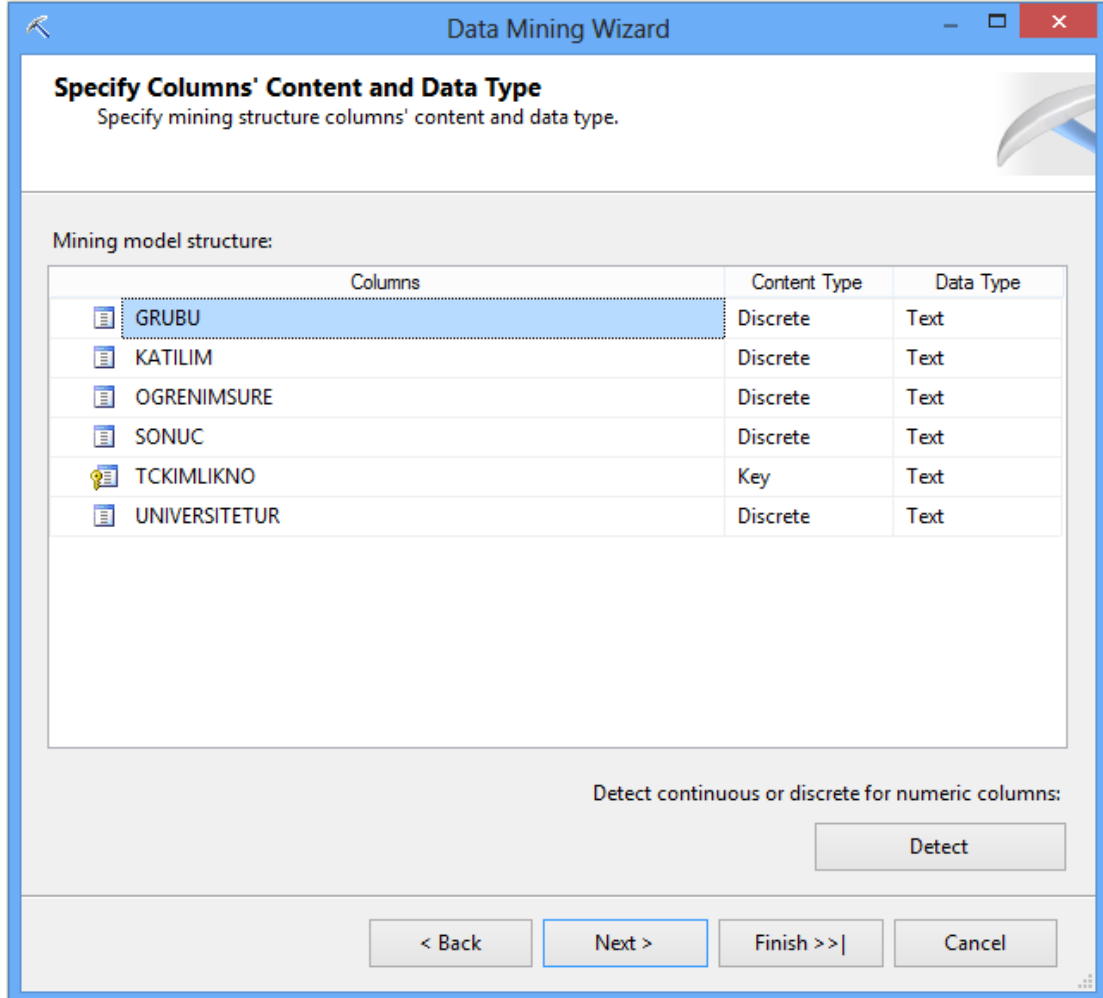
Şekil 5.17’de örnek olarak “KATILIM” ve “GRUBU” input değişken, “TCKIMLIKNO” key, “SONUC” ise predictable column olarak seçilmiştir.



Şekil 5.18 Algoritma İçin Sütunların Tahmin Edilmesi

“Suggest” butonuna tıktatıldığında sistemin daha iyi sonuç verebilmesi için hangi sütunların seçilebileceği program tarafından tavsiye edilir.

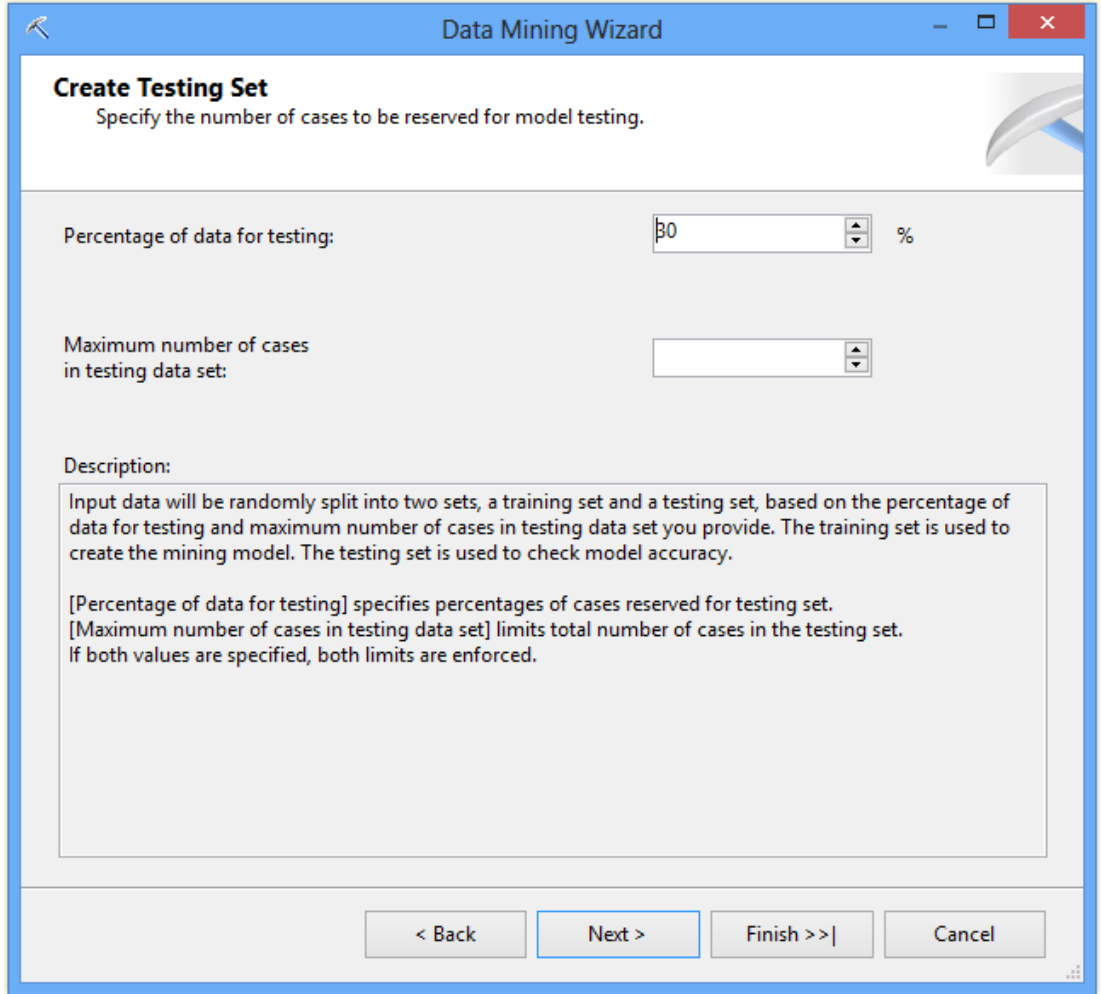
Şekil 5.18’de programın tavsiye ettiği sütunlar öğrenim süresi ve üniversite türüdür. İstenilirse bu sütunlar input değişken olarak da tanımlanabilirdi.



Şekil 5.19 Sütun Tiplerinin Belirlenmesi

Bu ekranda sütun tipleri üzerinde değişiklik yapabiliriz. Bazı değerleri sayısal olarak işleme sokabiliriz. Karakter tabanlı işlemlerde sütun tipi “Text” olabilir.

Şekil 5.19’de sütun tiplerinin tamamını “Text” olarak belirledik. Verilerimiz içerisinde sayısal hesaplamalar yapmaya gerek yoktur.

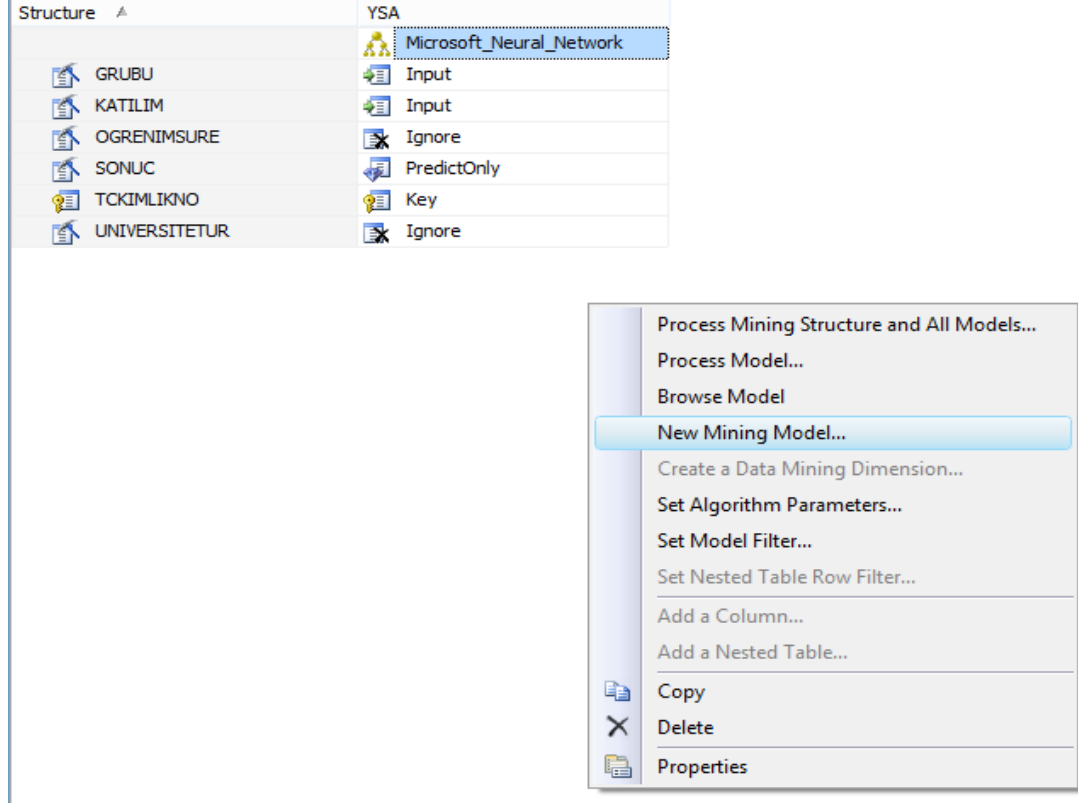


Şekil 5.20 Test Oranının Belirlenmesi

Bu ekranda ise ne kadar test verisi kullanacağımızı seçeriz. Hem yüzde olarak hem de sayı olarak seçim yapabiliriz.

Şekil 5.20'de test oranına 30 yazılması test amaçlı olarak verilerin %30'u üzerinde çalışılacağını gösterir. İstenildiği takdirde bir alttaki satıra sayı yazılarak belirli sayıda verinin test edilmesi de sağlanabilir.

Bu işlemlerin ardından veri modelimiz artık hazırdır.



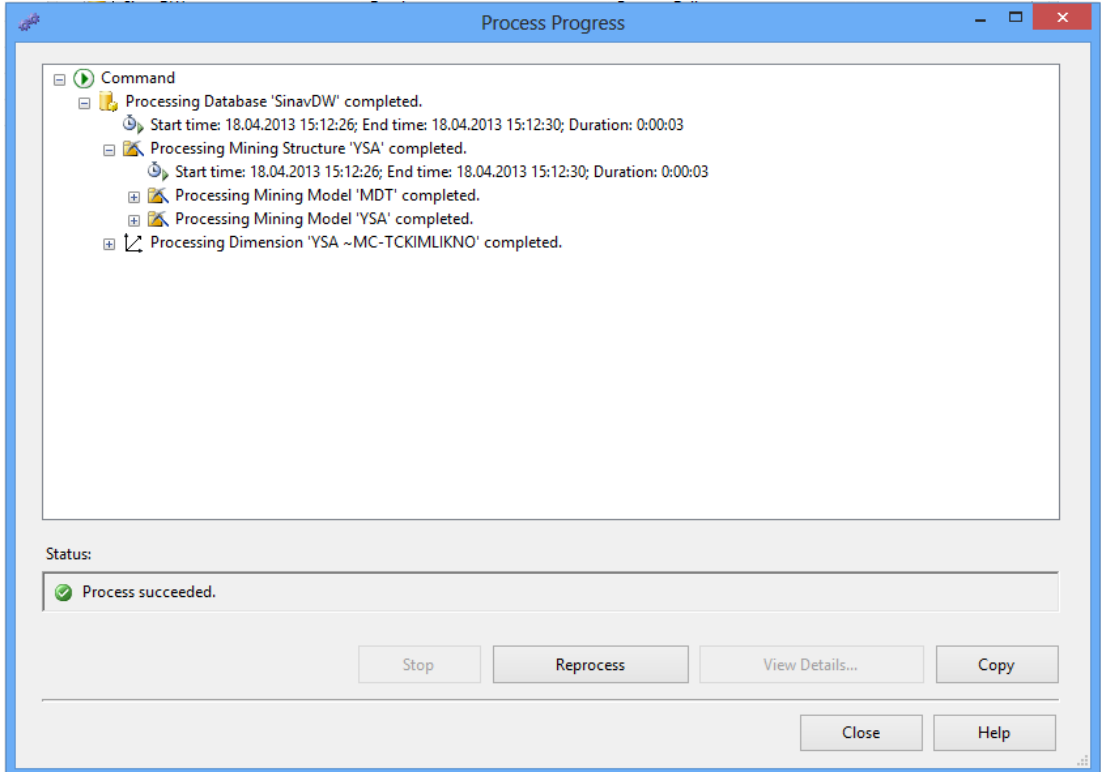
Şekil 5.21 Yeni Model Ekleme

Ayrıca veri modelimize 5.21’de gösterildiği gibi “New Minig Model...” bölümünden farklı bir VM tekniği ekleyebiliriz.

Structure	YSA	MDT
	Microsoft_Neural_Network	Microsoft_Decision_Trees
GRUBU	Input	Input
KATILIM	Input	Input
OGRENIMSURE	Ignore	Ignore
SONUC	PredictOnly	PredictOnly
TCKIMLIKNO	Key	Key
UNIVERSITETUR	Ignore	Ignore

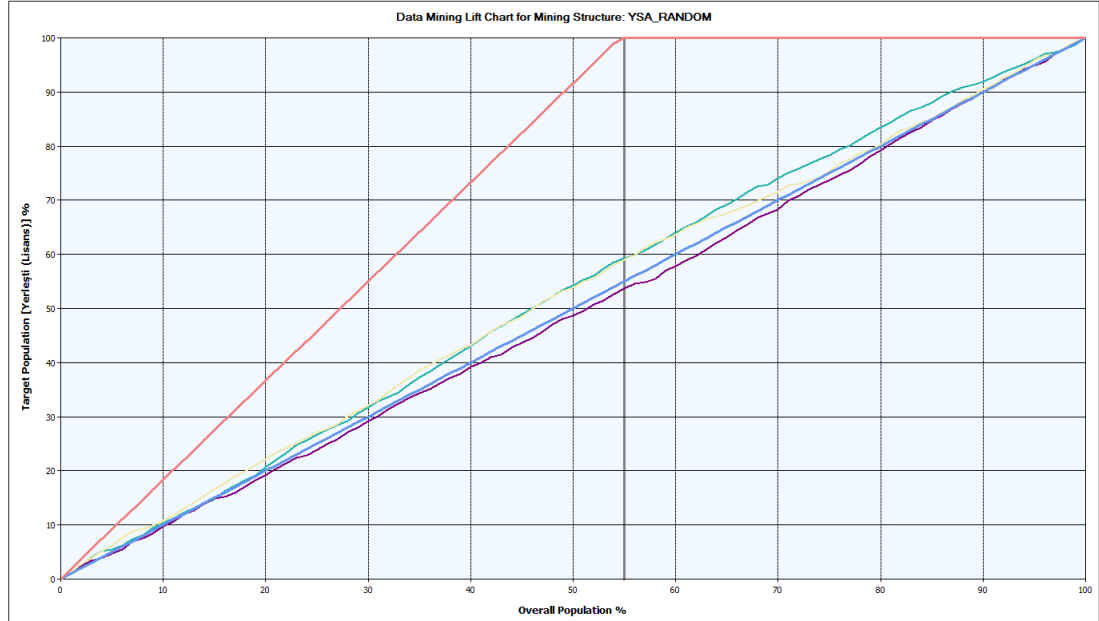
Şekil 5.22 Uygulamaya Yeni Model Ekleme

Şekil 5.22’de “Karar Ağaçları” adlı yeni bir model eklemiř olduk. Daha sonra “Naive Bayes” modeli eklenecektir ve bu 3 model karşılařtırılacaktır.



Şekil 5.23 Modelin Eğitilmesi ve Çalıştırılması

“Database” menüsünden Şekil 5.23’de gösterildiđi gibi “Process” seçeneđi ile sonuç alabilmemiz için veri modelimizi çalıştırıyoruz.



Grafik 5.2 Tahmini Lisans Yerleşme Grafiği (Kırmızı Çizgi: İdeal Durum, Mavi Çizgi: Rastgele Tahmin Değeri, Yeşil Çizgi: Yapay Sinir Ağları, Mor Çizgi: Karar Ağaçları, Sarı Çizgi: Naive Bayes)

Kırmızı çizgi sistemin en ideal durumunu belirtir. Bu kısım %55 civarı bir rakama denk gelmektedir. Bu demektir ki “Yerleşti (Lisans)” toplam küme içerisindeki oranı %55’dir. En ideal durumda tek seferde tüm hedef kitleyi bulmak için en iyi ihtimalle kayıtların sadece %55’lik kısmına bakmak yeterlidir. Mavi çizgi ise rastgele incelendiğinde istenilen kayıtlara ulaşılma durumunu belirtir.

Population percentage: 55,00%			
Series, Model	Score	Targe...	Predict probability
YSA_RANDOM	0,72	59,30%	77,55%
MDT_RANDOM	0,68	53,64%	79,56%
MNB_RANDOM	0,72	59,05%	77,67%
Random Guess M...		55,00%	
Ideal Model for: Y...		100,0...	

Şekil 5.24 Lisans Yerleşme Tahmin Oranları (Kırmızı Çizgi: İdeal Durum, Mavi Çizgi: Rastgele Tahmin Değeri, Yeşil Çizgi: Yapay Sinir Ağları, Mor Çizgi: Karar Ağaçları, Sarı Çizgi: Naive Bayes)

Şekil 5.24’de ideal duruma yakınlık ve tahmin değerleri yer almaktadır. “Score” değeri modelleri karşılaştırma açısından bize oldukça yardımcı olmaktadır. Yüksek “Score” değeri modelin daha iyi olduğu anlamına gelmektedir. “Target Population” hedef kitleyi göstermektedir. Önemli olan %100 değere en yakın olan sonucu elde edebilmektir. Nitekim skor ve hedef kitleye baktığımızda yapay sinir ağları ve naive bayes modelinin neredeyse aynı olduğu saptanmıştır.

Üniversiteye yerleşme durumunu tahminden daha çok hedef kitlenin tamamına ya da büyük bir kısmına ulaşmak daha önemlidir. Örneğin bu öğrencilere doğru e-posta göndermek için hedef kitleye yakınlığına bakılmalıdır. Lisans öğrencilerine ön lisans ile ilgili e-posta gönderilmesinin önüne geçilebilir.

6. SONUÇ VE ÖNERİLER

Dershaneler sadece sınavlara hazırlayan kurumlar değil, öğrencinin okul dışı zamanını yine öğretim ortamı ile değerlendirdiği ve onu zararlı alışkanlıklardan uzaklaştıran yerler olarak da değerlendirilebilir. Dershaneler sadece öğretim yapan yerler değil, aynı zamanda eğitim sürecini de devam ettiren kurumlar olma yolunda ve okullarda gerçekleştirilmeyen sosyal etkinliklerin yapıldığı yerler haline gelmektedir.

Dershanelerde düzey sınıflarının oluşturulması aynı başarı düzeyine sahip öğrencilerin bir sınıfta toplanmasına ve işlenen konuların sınıfın, dolaylı olarak da o sınıftaki her bir öğrencinin bilgi, başarı ve algı düzeyine uygun yöntemlerle anlatılmasını sağlar. Bu da hem öğrencinin anlamasını, hem de öğretmenin işini kolaylaştırır.

Sürekli deneme sınavına girerek hız ve deneyim kazanan bir aday zamanla gerçek sınavı da bir deneme sınavından farksız olarak algılamaya başlar ve sıradanlaştırır.

Dershanelerin bir diğer ayrıcalığı sınava yönelik test ve doküman zenginliğidir. Okul kitaplarında kullanılan içerik ve yöntem doğrudan sınava yönelik bir hazırlığı destekler nitelikte değildir. Öğrenciye daha geniş ve sınav müfredatının dışında da içerik sunulmaktadır.

Dershanelerin sistem içindeki en vazgeçilmez yanlarından birisi de ders dışı uygulamaların yoğunluğudur. İsteyen adaylar ders dışında organize edilen etüt ve birebir çalışmalar ile kendini daha ayrıcalıklı hisseder.

Özellikle rekabetin arttığı piyasada rakiplerle yarışacak büyük firmalar doğru kitleye ulaşabilecek pazarlama stratejileri de geliştirmek zorundadırlar. Bu stratejileri düşük maliyetle gerçekleştirmek isteyen firmalar mevcut kayıtlar arasından doğru müşterilere erişmek için veri madenciliği yöntemlerine başvurumaktadırlar.

Birliktelik kuralı ile öğrencilerin sorumlu oldukları branş dersler ve katılım süreleri arasındaki ilişkilerin öğrencinin üniversiteye yerleşme durumuna etkisi incelenmiştir. Bu sayede öğrencinin üniversiteye yerleşmesine katkı sağlamak amacı ile öğrenciler branş derslere yönlendirilebilmekte ve ona göre eğitim programı düzenlenebilmektedir.

Karar ağaçları gibi modellerle özellikle doğru yaş grubuna hitap edecek stratejiler belirlenmektedir. Doğru lokasyon, cinsiyet, yaş gibi faktörlere göre de stratejiler bu modeller sayesinde belirlenebilmektedir. Karar ağacının oluşturduğu dal yapısına göre farklı eğitim programları oluşturulabilir. Bu yapı aynı zamanda hangi eğitim programının öğrencinin üniversiteye yerleşmesine daha fazla katkı sağladığını belirleme açısından da önemlidir. Kümeleme modelleri ile de benzer karakteristik özelliklere sahip veriler gruplanarak farklı gruplara farklı stratejiler geliştirilebilir.

Dershaneler üniversite bölümlerini sürekli incelemede hangi branş derslerde başarılı olan öğrencileri kabul ettikleri izlenmektedir. Öğrencilerin kariyer planlamasına ve branş derslere olan eğilimlerine bakılarak lise dönemlerinde rehber öğretmenler tarafından yönlendirilmektedir. Bu sistem öğrencinin üniversitede istediği bölüme yerleşmesine katkı sağlamaktadır.

Matematik, Geometri, Türkçe ve Coğrafya derslerinden başarılı olan öğrencilerin lisans bölümlerine yerleşme oranlarının oldukça yüksek olduğu görülmüştür. Lisans bölümlerini kazanmada bu branşlarda başarılı olmanın oldukça etkisi vardır (Şekil 3.7).

Matematik ve Geometri derslerinden başarılı olmayıp, Türkçe ve Coğrafya derslerinde başarı gösteren öğrenciler ön lisans bölümlerini tercih etmektedirler (Şekil 3.8).

Matematik, Geometri, Türkçe ve Coğrafya derslerinde başarılı olamayan öğrenciler ya tercih yapmamakta ya da herhangi bir bölüme tercih ettiği halde yerleşmemektedirler (Şekil 3.9, Şekil 3.10).

Bugün bankacılık sektöründen telekomünikasyon sektörüne ve hatta perakende sektörüne kadar birçok büyük firma veri madenciliği yöntemlerini kullanmaktadır. Bu amaçla yazılım sektöründe veri madenciliği yöntemlerini uygulayabilen çalışana ihtiyaç duyulmaktadır. Üniversitelerde veri madenciliği dersleri uygulamalı olarak verilebilmeli ve piyasanın ihtiyaç duyduğu personel açığı giderilebilir.

Tez çalışması sonucunda veri madenciliği modelleri hakkında deneyim kazanılmıştır. Veri madenciliği ile aynı zamanda veri tabanı tasarımının daha iyi yapılabilmesi teşvik edilmiştir. Dershanelerin eğitim öğretim sistemindeki yeri hakkında bilgi edinilmiştir. Elde edilen bulgular sonucunda stratejik kararlar alma ve planlama yapılabileceği anlaşılmıştır.

Bu tez çalışmasında kullanılan veri madenciliği modelleri ile ilerde devlet-vakıf, kız-erkek gibi parametreler kullanılarak benzer çalışmalar yapılması planlanmaktadır. Özellikle birliktelik modelleri ile üniversitelerde öğrencilerin seçtiği seçmeli dersler arasındaki ilişkiler tespit edilerek hangi seçmeli derslerin açılıp açılmayacağına karar verilebilir veya hangi seçmeli ders üzerinde daha fazla durulacağına karar verilebilir.

Öğrencilerin belirli periyotlarla hangi derslerde başarısız oldukları kayıt altına alınarak bu dersler arasındaki ilişkiler tespit edilerek bu sorunun giderilmesi üzerinde durulabilir.

KAYNAKÇA

Aköz, E., "Otomotiv Sektöründe Veri Ambarları ve Bir Uygulama", Matematik Bilgisayar Ana Bilim Dalı, Fen Bilimleri Enstitüsü, Beykent Üniversitesi, 2007

Akpınar, H., "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği", İstanbul Üniversitesi, İşletme Fakültesi Dergisi, c:29, s:1-22, 2000

Al-Hudairy, "Data Mining and Decision Making Support in The Governmental Sector", Department of Computer Engineering, University of Louisville, 2004

Alpaydın, E., "Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri", Bilişim 2000 Eğitim Semineri, 2000

Al-Radaideh, Al-Shawakfa, Al-Nijjar, "Mining Student Data Using Decision Trees", International Arab Conference on Information Technology, 2006

Altıntop Ü., "İnternet Tabanlı Öğretimde Veri Madenciliği Tekniklerinin Uygulanması", Kocaeli Üniversitesi FBE, 2006

AY D., "Veri Madenciliği ve Apriori Algoritması İle Süpermarket Analizi", Sakarya Üniversitesi, FBE, 2009

Ayık, Y., Özdemir, A., Yavuz, U., "Lise Türü ve Lise Mezuniyet Başarısının, Kazanılan Fakülte ile İlişkisinin Veri Madenciliği Tekniği ile Analizi", Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, Cilt 10, Sayı 2, 2007

Bırtıl S., "Kız Meslek Lisesi Öğrencilerinin Akademik Başarısızlık Nedenlerinin Veri Madenciliği Tekniği İle Analizi", Afyon Kocatepe Üniversitesi, FBE, 2011

Bozkır, A., Sezer E., Gök B., “Öğrenci Seçme Sınavında(ÖSS) Öğrencilerin Başarımını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle Tespiti”, Uluslararası İleri Teknolojiler Sempozyumu (İATS’09), 2009

Bozkır A., Sezer E., Gök B., “Üniversite Öğrencilerinin İnterneti Eğitimsel Amaçlar İçin Kullanmalarını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle Tespiti”, Bilimde Modern Yöntemler Sempozyumu, 2008

Buldu A., Üçgün K., “Data mining application on students’ data” Technical Education of Marmara University, Istanbul, 34722, Turkey Received November 15, 2009; revised December 3, 2009; Accepted January 25, 2010

Chen Y., Tang K., Shen R. Ve Hu Y., ‘Market Basket Analysis in a Multiple Store Environment’, Decision Support Systems, Volume 40, Issue 2, s. 339-354, 2005

Chih-Fong Tsai, Ching-Tzu Tsai, Chia-Sheng Hung, Po-Sen Hwang. “Data mining techniques for identifying students at risk of failing a computer proficiency test required for graduation” Australasian Journal of Educational Technology, 27(3), 481-498, 2011

Cloyd, J.,D., “Data Mining With Newton’s Method”, Department of Computer and Information Sciences, East Tennessee StateUniversity, 2002

Çetinyokuş, T., Gökçen H., “Borsada Göstergelerle Teknik Analiz İçin Bir Karar Destek Sistemi”, Gazi Üniv. Müh. Mim. Fak. Der. Cilt 17, No 1, 43-58, 2002

Demiralay M., “Hiyerarşik Kümeleme Metotları İle Veri Madenciliği Uygulamaları”, Temmuz 2005

Dinçer E., “Veri Madenciliğinde K-Means Algoritması ve Tıp Alanında Uygulanması”, Kocaeli Üniversitesi FBE, 2006

Dondurmacı G. , “Veri Madenciliğinde Regresyon Ağaçları İle Sınıflandırma ve Bir Uygulama”, Mimar Sinan Güzel Sanatlar Üniversitesi, 2011

Döşlü, A., “Veri Madenciliğinde Market Sepet Analizi ve Birliktelik Kurallarının Belirlenmesi”, Bilgisayar Mühendisliği Anabilim Dalı, Fen Bilimleri Enstitüsü, Yıldız Teknik Üniversitesi, 2008

Düzgünoğlu, S., “Veri Ambarı ve OLAP Teknolojilerinden Yararlanılarak Karar Destek Amaçlı Raporlama Aracı Gerçekleştirimi”, Yüksek Lisans Tezi, Bilgisayar Mühendisliği Anabilim Dalı, Fen Bilimleri Enstitüsü, Hacettepe Üniversitesi, 2006

Erdoğan, Z., Timor, M., A Data Mining Application In a Student Database. Journal Of Aeronautics and Space Technologies, July 2005 Volume 2 Number 2 (53-57), 2005

Giran, Ö., “İnşaat İşletme Yönetiminde Bilgi Ambarlama”, İnşaat Mühendisliği Anabilim Dalı, Fen Bilimleri Enstitüsü, İstanbul Üniversitesi, 2008

Gülçe G. , “Veri Ambarı ve Veri Madenciliği Teknikleri Kullanılarak Öğrenci Karar Destek Sistemi Oluşturma”, Pamukkale Üniversitesi FBE, 2010

Güneri, N., Apaydın, A., “Öğrenci Başarılarının Sınıflandırılmasında 96 Lojistik Regresyon Analizi ve Sinir Ağları Yaklaşımı”, Ticaret ve Turizm Eğitim Fakültesi Dergisi, Yıl: 2004, Sayı: 1, ss. 170 – 188, 2004

Han, J., Kamber, Kamber, M., “Data Mining Concepts and Techniques”, 95 Morgan Kaufmann, 2006

Ho Yu, C., DiGangi, S., Jannasch-Pennell, A., Kaprolet, C., "A Data Mining Approach for Identifying Predictors of Student Retention from Sophomore to Junior Year", Journal of Data Science 8(2010), 307-325, 2010

Işık M., "Bölünmeli Kümeleme Yöntemleri İle Veri Madenciliği Uygulamaları", Marmara Üniversitesi FBE, 2006

Inmon, W.H., "Building the Data Warehouse", Wiley Computer Publishing, 2002

İç, Y., T., "İşleme Merkezlerinin Seçiminde Kullanılacak Bir Karar Destek 94 Sisteminin Geliştirilmesi", Makine Mühendisliği Anabilim Dalı, Fen Bilimleri Enstitüsü, Gazi Üniversitesi, 2006

İşli, D., "Veri Ambarı ve OLAP Teknolojilerinden Yararlanılarak Raporlama Aracı Gerçekleştirimi", Bilgisayar Mühendisliği Anabilim Dalı, Fen Bilimleri Enstitüsü, Pamukkale Üniversitesi, 2009

Kalaycı R., "Veri Madenciliğinde K-Means Algoritmasının İncelenmesi ve Uygulanması", Marmara Üniversitesi, 2008

Koldere Akın Y., "Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi", Marmara Üniversitesi, SBE, 2008

Kalles, D., Pierrakeas, C., "Analyzing Student Performance in Distance Learning With Genetic Algorithms and Decision Trees", Applied Artificial Intelligence, Volume 20, Issue, pages 655 – 674, 2006

Karabatak, M., İnce, M., C., "Apriori Algoritması ile Öğrenci Başarısı Analizi", Elektrik Elektronik Bilgisayar Mühendisliği Sempozyumu, ELECO, Elektronik-Bilgisayar sayfa 348-352, 2004

Karaca Y., Şahiner D. S., Aslan Z., “Multi-Resolution Wavelet Analyses of Two Different Perfectionism Scales: A University Sample”, İAÜ, İstanbul, 2010

Kovačić Z., “Early Prediction of Student Success: Mining Students Enrolment Data”, Proceedings of Informing Science & IT Education Conference (InSITE), 2010

MacLennan J., Tang Z., Crivat B., “Data Mining With Microsoft SQL Server 2008”, Wiley Publishing, 2009

Oyelade O. J, Oladipupo O. O, Obagbuwa I. C., “Application of K-Means Clustering Algorithm for Prediction of Students’ Academic Performance”, (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, _o. 1, 2010

Özdamar E. Ö., “Veri Madenciliğinde Kullanılan Teknikler ve Bir Uygulama”, İstatistik Anabilim Dalı, Fen Bilimleri Enstitüsü, Mimar Sinan Üniversitesi, 2002

Özçınar H., “KPSS Sonuçlarının Veri Madenciliği Yöntemleriyle Tahmin Edilmesi”, Bilgisayar Mühendisliği, Fen Bilimleri Enstitüsü, Pamukkale Üniversitesi, 2006

Özekes S., Çamurcu A. Y., “Veri Madenciliğinde Sınıflama ve Kestirim Uygulaması”, Marmara Üniversitesi Fen Bilimleri Dergisi, 18, 2002

Özkan Y., “Veri Madenciliği Yöntemleri”, Papatya Yayıncılık Eğitim, 2008

Paliwal M., Kumar U. A., “Neural Networks and Statistical Techniques: A Review of Applications”, Expert Systems With Applications, 36 (1), 2-17, 2009

Savaş S., Topalođlu N., Yılmaz M., “Veri Madenciliđi ve Türkiye’deki Uygulama Örnekleri” İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi Yıl:11 Sayı: 21 Bahar 2012 s. 1-23, 2012

Seetharaman M., “Data Warehousing: Case study in data quality improvement, School of Information Systems and Engineering Technology”, Institute of Technology, State University of New York, 2008

Seyrek İ. H., Ata H., “Veri Zarflama Analizi ve Veri Madenciliđi ile Mevduat Bankalarında Etkinlik Ölçümü“, BDDK Bankacılık ve Finansal Piyasalar, Cilt:4, Sayı:2, 2010

Şen F., “Veri Madenciliđi İle Birliktelik Kurallarının Bulunması”, Sakarya Üniversitesi, FBE, 2008

Sezer, Ü., “Karar Ağaçlarının Birliktelik Kuralları ile İyileştirilmesi”, Bilgisayar Mühendisliđi Anabilim Dalı, Fen Bilimleri Enstitüsü, Kocaeli Üniversitesi, 2008

Shen Y. , “A Formal Ontology for Data Mining: Principles, Design, and Evolution”, Département de mathématiques et d’informatique ,Université du Quebec a Trois-Rivieres, 2007

Silahtarođlu G., “Kavram ve Algoritmalarıyla Veri Madenciliđi”, Papatya Yayıncılık Eğitim, 2008

Tantuđ A., “Veri Madenciliđi ve Demetleme”, Bilgisayar Mühendisliđi Anabilim Dalı, Elektrik-Elektronik Fakültesi, İstanbul Teknik Üniversitesi, 2002

Timor M., Erdođan Ş. Z., “A Data Mining Application in a Student Database” Journal Of Aeronautics and Space Technologies, July 2005 Volume 2 Number 2 (53-57) , 2005

Timor M., Şimşek T., “Veri Madenciliğinde Sepet Analizi İle Tüketici Davranışı Modellemesi”, Yönetim, Yıl:19, Sayı:59, 2008

Turban E., “Decision Support and Expert Systems: Management Support Systems”, Englewood Cliffs, N.J., PrenticeHall, 1995

Türkmen, G., ”Developing A Data Warehouse for A University Decision Support System, Department of Computer Engineering”, Graduate School of Natural and Applied Sciences, Atılım University, 2007

Üçgün K., “Orta Öğretim Okulları İçin Öğrenci Otomasyonu Tasarımı ve Öğrenci Verileri Üzerine Veri Madenciliği Uygulamaları”, Marmara Üniversitesi, FBE, 2009

Vandamme J., Meskens N., Superby J., F., “Predicting Academic Performance by Data Mining Methods“, Education Economics , Vol. 15 , 405-419, 2007

Vindervogel B., Van Den Poel D. Ve Wets G., “Direct and Indirect Effects of Retail Promotions on Sales and Profits in the Do-It-Yourself Market”, Expert Systems With Applications, Volume 28, Issue 3, s. 583-590, 2005

Yang M., “Data Mining Techniques Applied to Texas Woman’s University’s Enrollment Data – What Can The Data Tell Us?”, College of Arts and Sciences, Texas Woman’sUniversity, 2006

Yelegin A., “Mesleki Eğitimde Öğrenci Altyapısının Öğrenci Eğitim Başarısına Etkisinin Veri Madenciliği Yöntemleriyle Ortaya Çıkartılması” Beykent Üniversitesi FBE, 2012

Zhu, J, Sampling- An Efficient Solution For Data Mining Of Association Rules, Department of Computer Science, Dalhousie University, 2003

Ünsal S. Kültür Dershaneleri Rehberlik Koordinatörü, Görüş alışverişi, (Ocak 2013)

'Fatih projesi', Detaylı Bilgi,
<http://fatihprojesi.meb.gov.tr>, Ocak 2013

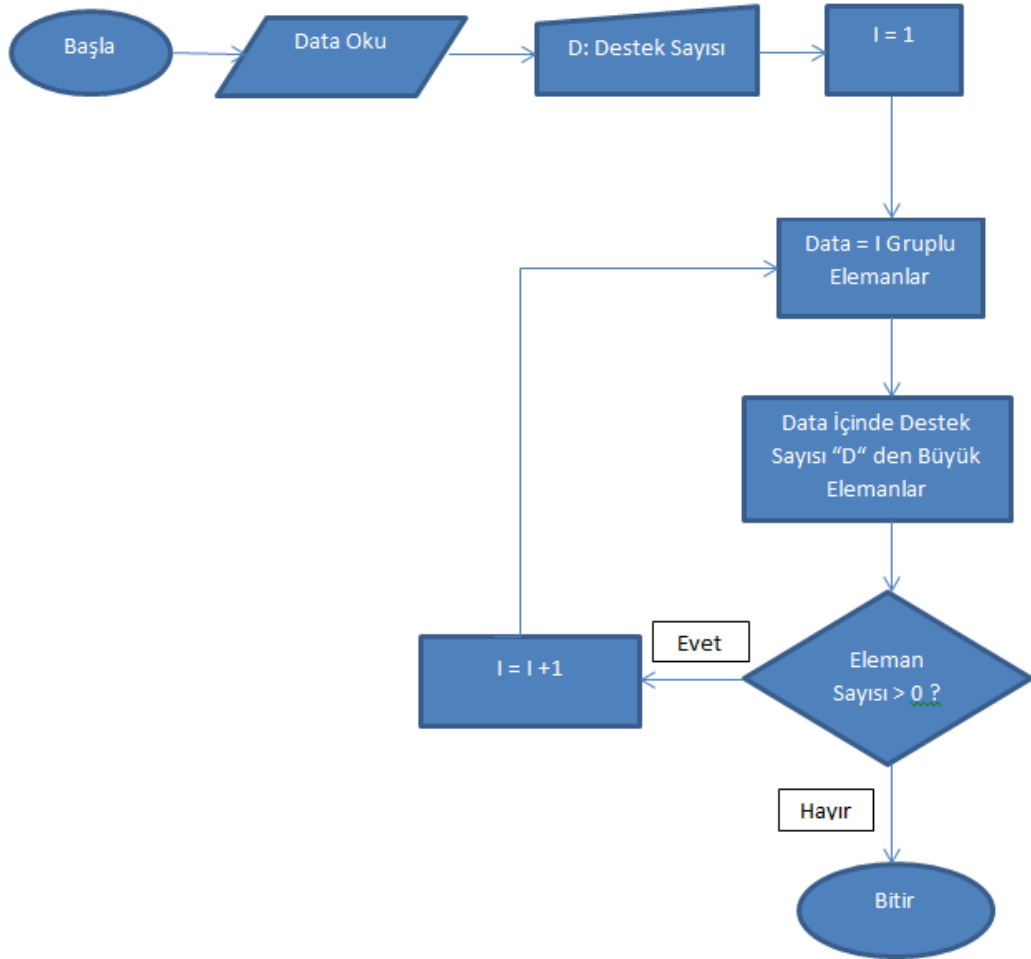
'K-Means (K – Ortalama) Algoritması'
<http://www.iszekam.net/?tag=/veri+madenciligi+algoritmaları>, Ocak 2013

Microsoft Data Mining
<http://msdn.microsoft.com/en-us/library/ms175595.aspx>, Ocak 2013

Yapay Sinir Ağları
<http://msdn.microsoft.com/en-us/magazine/hh975375.aspx>, Ocak 2013

EKLER

I Apriori Algoritması Akış Şeması



I Apriori Algoritması SQL Uyarlaması Devam

-- Destek Sayısının Belirlenmesi

```
DECLARE @DESTEK INT, @CONFIDENCE FLOAT
```

```
SET @DESTEK = 1200 SET @CONFIDENCE = 0.4
```

-- Benzersiz Öğrenci-Ders Kayıtlarının Belirlenmesi

```
SELECT DISTINCT TCKIMLIKNO, DERSAD INTO #C1
```

```
FROM Analiz snv
```

-- Belirlenen Destek Sayısından Büyük Birli Ders Sayılarının Belirlenmesi

```
SELECT DERSAD, TOPLAM = COUNT(1) INTO #L1 FROM #C1
```

```
GROUP BY DERSAD HAVING COUNT(1) >= @DESTEK
```

-- Belirlenen Destek Sayısından Büyük İkili Ders Sayılarının Belirlenmesi

```
SELECT D1 = I.DERSAD, D2 = I1.DERSAD, TOPLAM = COUNT(1)
```

```
INTO #C2 FROM #L1 I
```

```
JOIN #L1 I1 ON I1.DERSAD <> I.DERSAD
```

```
JOIN #C1 c1 ON c1.DERSAD = I.DERSAD
```

```
JOIN #C1 c11 ON c11.TCKIMLIKNO = c1.TCKIMLIKNO AND c11.DERSAD
```

```
= I1.DERSAD
```

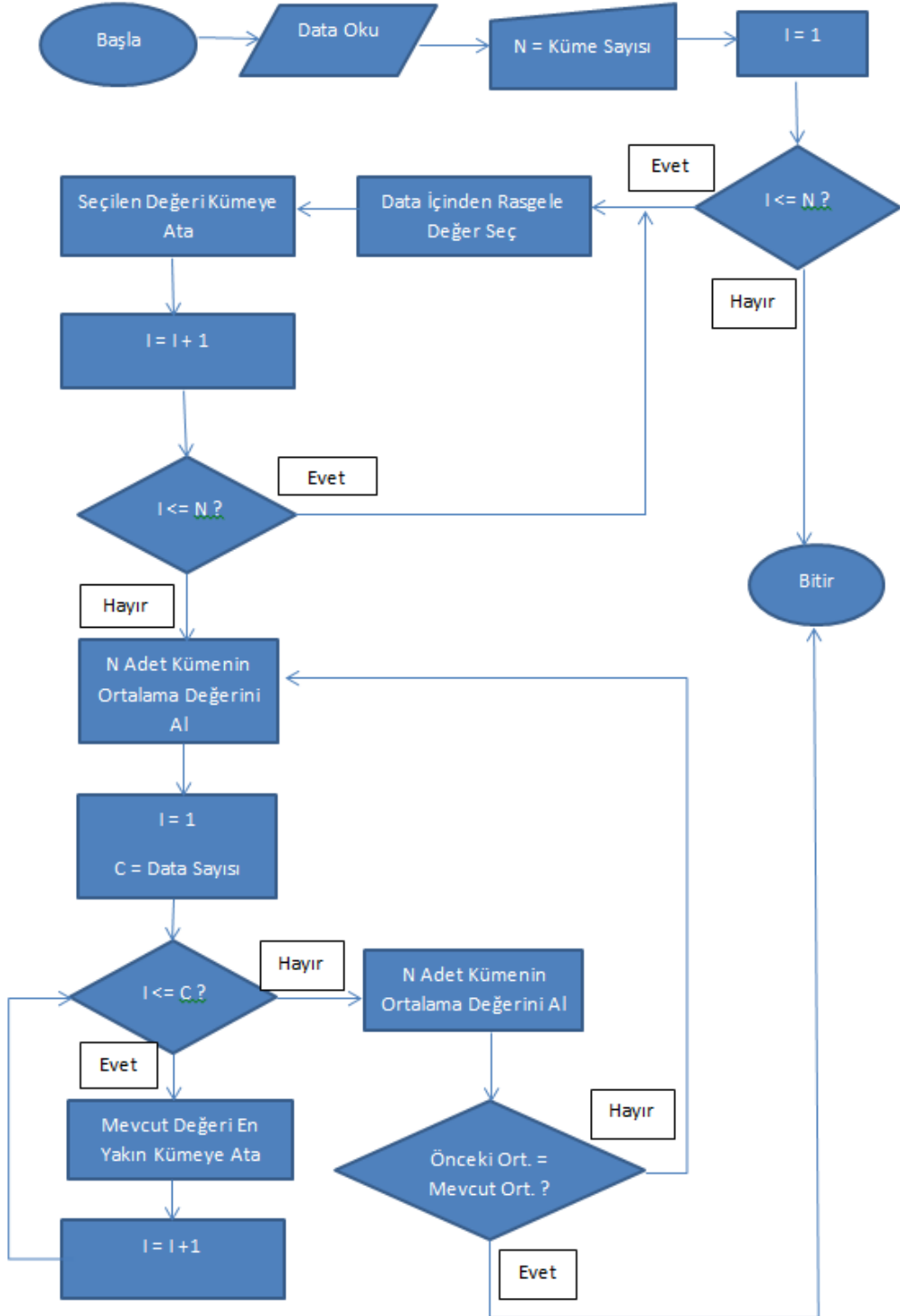
```
GROUP BY I.DERSAD, I1.DERSAD
```

```
SELECT * INTO #L2 FROM #C2 WHERE TOPLAM >= @DESTEK
```

```
SELECT * FROM #L2s
```

-- Bu Adımlar Destek Sayısından Büyük Kayıtlar Bulunduğu Sürece Devam Eder.

II K-Ortalama Algoritması Akış Şeması



II K-Ortalama Algoritması SQL Uyarlaması Devam

```
DECLARE @TBL TABLE (RN INT, DEGER INT)
-- Üniversiteye Yerleşen Aaylar İçin Yükseköğretim Programlarını Belirle
SELECT YOP, DEGER = COUNT(1) INTO #Temp FROM Universite
WHERE SONUC = 'Yerleşti.' GROUP BY YOP
SELECT DISTINCT DEGER INTO #Tbl FROM #Temp
orderby DEGER
INSERT INTO @TBL SELECT RN = ROW_NUMBER()
OVER (ORDER BY DEGER), DEGER FROM #Tbl
-- 3 Küme İçin 3 Adet Tablo Tanımla
DECLARE @KUME1 TABLE (DEGER INT)
DECLARE @KUME2 TABLE (DEGER INT)
DECLARE @KUME3 TABLE (DEGER INT)

DECLARE @Random INT; @Upper INT; @Lower INT,
@TBLCOUNT INT
-- Veri Sayısını Bul
SELECT @TBLCOUNT = COUNT(1) FROM @TBL
---- 1 ile Eleman Sayısı Arası Rasgele Sayı Üret
SET @Lower = 1 ---- Minimum Rakam
SET @Upper = @TBLCOUNT ---- Eleman Sayısı

-- 1 ile Veri Sayısı Arasında Rasgele Sayı Üret ve Kümelere Ata
SELECT @Random = ROUND((((@Upper - @Lower - 1) * RAND() +
@Lower), 0)
INSERT INTO @KUME1 SELECT DEGER FROM @TBL WHERE RN =
@Random
SELECT @Random = ROUND((((@Upper - @Lower - 1) * RAND() +
@Lower), 0)
INSERT INTO @KUME2 SELECT DEGER FROM @TBL WHERE RN =
@Random
```

II K-Ortalama Algoritması SQL Uyarlaması Devam

```
SELECT @Random = ROUND((((@Upper - @Lower -1) * RAND() +
@Lower), 0)
INSERT INTO @KUME3 SELECT DEGER FROM @TBL WHERE RN =
@Random
```

```
DECLARE @C INT, @I INT, @USTSAYAC INT
SET @USTSAYAC = 0 SET @I = 1
SELECT @C = COUNT(1) FROM @TBL
```

```
DECLARE @DEGER INT, @BIRINCIKUME INT, @IKINCIKUME INT,
@UCUNCUKUME INT
DECLARE @ONCEKIM1 FLOAT, @ONCEKIM2 FLOAT,
@ONCEKIM3 FLOAT, @M1 FLOAT, @M2 FLOAT, @M3 FLOAT
-- Ortalamalar Eşit Oluncaya Kadar Düngüyü Çalıştır.
WHILE @USTSAYAC = 0
BEGIN
-- Önceki Ortalama Değerlerini Bul
    SET @ONCEKIM1 = (SELECT AVG(CAST(DEGER AS FLOAT))
FROM @KUME1)
    SET @ONCEKIM2 = (SELECT AVG(CAST(DEGER AS FLOAT)) FROM
@KUME2)
    SET @ONCEKIM3 = (SELECT AVG(CAST(DEGER AS FLOAT)) FROM
@KUME3)
    SET @I = 1
    WHILE @I <= @C
    BEGIN
        SELECT @DEGER = DEGER FROM @TBL WHERE RN = @I
        IF (SELECT COUNT(1) FROM @KUME1) <> 1
            DELETE FROM @KUME1 WHERE DEGER = @DEGER
```

II K-Ortalama Algoritması SQL Uyarlaması Devam

```
IF (SELECT COUNT(1) FROM @KUME2) <> 1
    DELETE FROM @KUME2 WHERE DEGER = @DEGER
    IF (SELECT COUNT(1) FROM @KUME3) <> 1
        DELETE FROM @KUME3 WHERE DEGER = @DEGER
-- Kümedeki Elemanlara En Yakın Sayıları Bul
    SELECT @BIRINCIKUME = MIN(ABS(@DEGER - DEGER))
    FROM @KUME1
SELECT @IKINCIKUME = MIN(ABS(@DEGER - DEGER))
FROM @KUME2
SELECT @UCUNCUKUME = MIN(ABS(@DEGER - DEGER))
FROM @KUME3
-- Bulunan Sayıları Olması Gereken Kümelere Ata
    IF (@BIRINCIKUME <= @IKINCIKUME AND
        @BIRINCIKUME <= @UCUNCUKUME)
        BEGIN
            DELETE FROM @KUME1 WHERE DEGER = @DEGER
            INSERT INTO @KUME1 SELECT @DEGER
        END
    ELSE IF (@IKINCIKUME <= @BIRINCIKUME AND @IKINCIKUME
<= @UCUNCUKUME)
        BEGIN
            DELETE FROM @KUME2 WHERE DEGER = @DEGER
            INSERT INTO @KUME2 SELECT @DEGER
        END
    ELSE
        BEGIN
            DELETE FROM @KUME3 WHERE DEGER = @DEGER
            INSERT INTO @KUME3 SELECT @DEGER
        END
    SET @I = @I + 1          END
```


II K-Ortalama Algoritması SQL Uyarlaması Devam

```
-- Son Ortalamaları Bul
SET @M1 = (SELECT AVG(CAST(DEGER AS FLOAT)) FROM @KUME1)
SET @M2 = (SELECT AVG(CAST(DEGER AS FLOAT)) FROM @KUME2)
SET @M3 = (SELECT AVG(CAST(DEGER AS FLOAT)) FROM @KUME3)
-- Ortalamalar Eşit İse Döngüyü Bitir.
IF @ONCEKIM1 = @M1 AND @ONCEKIM2 = @M2 AND
@ONCEKIM3 = @M3
SET @USTSAYAC = 1
ELSE
    SET @USTSAYAC = 0
END
```

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

- **İsim** : Bünyamin HATİPOĞLU
- **Doğum Yeri** : Trabzon / Of
- **Doğum Tarihi** : 07.07.1983
- **Uyruğu** : TC
- **Askerlik Durumu** : Yapıldı (17.05.2008)
- **Telefon** : 0538 041 41 90
- **E-Mail** :
bunyaminhatipoglu@hotmail.com



İŞ TECRÜBELERİ

- Techlife Bilişim Çözümleri 01.01.2009 – Halen Çalışıyorum

ÖĞRENİM DURUMU

- Bilgisayar Mühendisliği – Yüksek Lisans – İstanbul Aydın Üniversitesi (Öğrenci – Tez Dönemi)
- Bilgisayar Mühendisliği – Lisans – Uluslararası Kıbrıs Üniversitesi (Mezun - 2007)

YABANCI DİL VE DÜZEYİ

- İngilizce -Uluslararası Kıbrıs Üniversitesi – İyi Seviye

BİLGİSAYAR BECERİLERİ

- C#
- MS-SQL (StoredProcedure – Transaction - Dynamic SQL)
- SILVERLIGHT
- WCF
- LINQ
- ASP.NET

ÖZGEÇMİŞ Devam

BİLGİSAYAR BECERİLERİ

- HTML
- JAVA SCRIPT - JQUERY
- JSON
- DevExpress Controls

KURS VE SERTİFİKALAR

- Design Patterns and OOP Prnciples – BT Akademi – 32 Saat - 26.03.2011

PROJELER

- <http://obs.kulturdersanesi.com.tr>
- <http://obs.okyanuskoleji.com>
- <http://www.tercichnet.com>
- <http://kurumsal.tercichnet.com>
- <http://sinav.kulturyayincilik.com>
- Optik Sınav Değerlendirme Sistemi
- Öğrenci, Personel, Finans, Stok, Rehberlik, Danışmanlık, İK, Sınav Değerlendirme vb. Modüler Sistem
- <http://xproje.techlife.com.tr>

REFERANSLAR

- Kültür Dershaneleri
- Fatih Dershaneleri
- Okyanus Kolejleri
- Kültür Yayıncılık
- Okyanus Yayıncılık
- English Time