

KOCAELİ ÜNİVERSİTESİ * FEN BİLİMLERİ ENSTİTÜSÜ

**İNTERNET TABANLI ÖĞRETİMDE VERİ MADENCİLİĞİ
TEKNİKLERİNİN UYGULANMASI**

YÜKSEK LİSANS

Bilgisayar Öğr. Ümmühan ALTINTOP

Anabilim Dalı: Elektronik ve Bilgisayar Eğitimi

Danışman: Yrd.Doç.Dr. Nevcihan DURU

KOCAELİ, 2006

ÖNSÖZ ve TEŞEKKÜR

Günümüzde internetin yaşantımızı adeta yeniden şekillendirmesi ile eğitim alanında da gözler web sitelerini etkin bir şekilde kullanımına çevrilmiştir. Geleneksel eğitimden uzaklaşarak internet tabanlı öğretim sistemleri geliştirilmeye başlanmıştır. İnternet tabanlı öğretimi amaçlayan web sitelerinin bulunduğu sunucular üzerindeki erişim dosyalarında öğrencilerin site içinde gezinirken yaptığı her hareket karşılığında veri tabanında birçok veri birikir. Bu veri miktarının günden güne hızla arttığı düşünüldüğünde, bu veriler içerisinde anlamlı bilgi örüntüleri çıkarmanın veri madenciliği alanındaki gelişmeler ile sağlanabileceği görülmektedir.

Birliktelik kuralları ve kümeleme teknikleri, günümüzde en sık uygulanan veri madenciliği tekniklerindedir.

Birliktelik kuralları, büyük ve anlamsız veri yığınları içerisinde ilginç ve anlamlı örüntüleri keşfetmek; kümeleme teknikleri ise bir veri kümesindeki bilgileri yakınlık kriterlerine göre gruplara ayırma işlemidir. Tezde, bu yaklaşımlardan yola çıkarak tasarlanan İnternet tabanlı bir öğretim sitesine bağlanan öğrencilerin verileri üzerinde bir analiz aracı geliştirilmiştir.

Bana bu konuda çalışma fikri veren, çalışmam süresince fikir ve yapıcı eleştirileriyle benden desteğini esirgemeyen değerli hocam Sayın Yrd. Doç Dr. Nevcihan DURU'ya teşekkür ederim. Ayrıca çalışmam boyunca hep yanımda olup beni destekleyen değerli eşime, çalışmamda yardımlarını esirgemeyen tüm arkadaşlarıma teşekkür ederim.

Mayıs 2006, KOCAELİ

Ümmühan ALTINTOP

İÇİNDEKİLER

| | |
|---|------|
| ÖNSÖZ ve TEŞEKKÜR | ii |
| İÇİNDEKİLER | iii |
| ŞEKİLLER LİSTESİ | vi |
| TABLolar DİZİNİ | viii |
| SİMGELEr DİZİNİ ve KISALTMALAR | ix |
| ÖZET | x |
| ABSTRACT | xi |
| | |
| BÖLÜM 1. GİRİŞ | 1 |
| | |
| BÖLÜM 2. VERİ MADENCİLİĞİ ve KULLANILAN TEKNİKLER | 10 |
| 2.1. Giriş 10 | 10 |
| 2.2. Veri Tabanlarında Bilgi Keşfi Süreci (VTBK) | 10 |
| 2.3. Veri Madenciliği | 13 |
| 2.4. Veri Ambarı | 13 |
| 2.5. Veri Madenciliği ve Diğer Disiplinler | 14 |
| 2.6. Veri Madenciliği Uygulama Alanları | 15 |
| 2.7. Veri Madenciliğinde Karşılaşılan Problemler | 17 |
| 2.7.1. Veritabanı boyutu | 17 |
| 2.7.2. Gürültülü veri | 18 |
| 2.7.3. Boş değerler | 18 |
| 2.7.4. Eksik veri | 18 |
| 2.7.5. Artık veri | 19 |
| 2.7.6. Dinamik veri | 19 |
| 2.7.7. Farklı tipteki verileri ele alma | 19 |
| 2.8. Veri Madenciliği Modelleri | 20 |
| 2.8.1. Sınıflama ve regresyon | 21 |
| 2.8.2. Kümeleme tekniği | 21 |
| 2.8.2.1. Bölümleme yöntemleri | 22 |
| 2.8.2.1.1. K-Means algoritması | 22 |
| 2.8.2.1.2. K-Medoids algoritması | 23 |
| 2.8.2.1.3. Clara-Clarans algoritmaları | 23 |
| 2.8.2.2. Hiyerarşik yöntemler | 23 |
| 2.8.2.3. Yoğunluk tabanlı yöntemler | 23 |
| 2.8.2.4. Izgara tabanlı yöntemler | 23 |
| 2.8.2.5. Model tabanlı yöntemler | 24 |
| 2.8.3. Birliktelik kuralları ve ardışık zamanlı örüntüler | 24 |
| 2.8.4. Bellek tabanlı yöntemler | 24 |
| 2.8.5. Yapay sinir ağları | 24 |
| 2.8.6. Karar ağaçları | 25 |

| | |
|---|----|
| BÖLÜM 3. APRİORİ ve DBSCAN ALGORİTMALARI | 26 |
| 3.1. Giriş | 26 |
| 3.2. Birliktelik Kuralları..... | 26 |
| 3.2.1. Birliktelik kurallarının türleri..... | 28 |
| 3.2.1.1. Hiyerarşik birliktelik kuralları | 28 |
| 3.2.1.2. Çok seviyeli birliktelik kuralları | 29 |
| 3.2.1.3. Negatif birliktelik kuralı..... | 30 |
| 3.2.1.4. Nicel birliktelik kuralı | 30 |
| 3.2.2. Birliktelik kuralları algoritmaları..... | 31 |
| 3.2.2.1. Apriori algoritmasının türleri..... | 35 |
| 3.3. DBSCAN Algoritması (Density Based Spatial Clustering of Applications with Noise) | 36 |
| 3.3.1. DBSCAN algoritmasının çalışması | 39 |
| | |
| BÖLÜM 4. İNTERNET TABANLI ÖĞRETİM | 42 |
| 4.1. Giriş | 42 |
| 4.2. İnternet Tabanlı Öğretim..... | 42 |
| 4.2.1. İnternet tabanlı öğretimin faydaları..... | 43 |
| 4.2.2. İnternet tabanlı öğretim modeli..... | 44 |
| 4.2.3. Web tabanlı uzaktan eğitim sistemleri'nin sahip olması gereken özellikler.... | 47 |
| 4.3. İnternet Tabanlı Öğretimde Kullanılan Teknolojiler | 49 |
| 4.3.1. Uyarlanır içerik sunumu | 49 |
| 4.3.2. Öğrenci çözümlerinin irdelenmesi ve hata kaynaklarının bulunması | 50 |
| 4.3.3. Etkileşimli problem çözüme desteği | 50 |
| 4.3.4. Uyarlanır gezinme desteği | 51 |
| 4.4. İnternet Tabanlı Öğretim Sistemlerinin Mimarisi..... | 51 |
| 4.4.1. Bilgi anabirimi | 51 |
| 4.4.2. Öğrenci modeli..... | 51 |
| 4.4.3. Eğitim birimi | 53 |
| 4.4.4. Kullanıcı arayüzü | 53 |
| 4.5. Kocaeli Üniversitesi Enformatik bölümü İnternet Tabanlı Öğretim Öğretim sitesi ve Hazırlanan Veri Tabanı | 54 |
| | |
| BÖLÜM 5. İNTERNET TABANLI ÖĞRETİMDE VERİ MADENCİLİĞİ TEKNİKLERİNİN UYGULANMASI | 63 |
| 5.1. Giriş | 63 |
| 5.2. Neden Birliktelik Kuralları ve DBSCAN Algoritması | 63 |
| 5.3. Uygulamanın Açıklaması..... | 65 |
| 5.3.1. Genel değerlendirme menüsü..... | 66 |
| 5.3.1.1. Sayfa bilgileri | 67 |
| 5.3.1.1.2. Girişlere göre..... | 67 |
| 5.3.1.1.3. Anket ve girişlere göre | 69 |
| 5.3.1.2. Quiz sonuçları | 73 |
| 5.3.1.2.1. Quiz değerlendirmesi | 73 |
| 5.3.1.2.2. Girişlere göre quiz değerlendirme | 74 |
| 5.3.1.2.3. Bölümlere göre quiz değerlendirme | 75 |
| 5.3.1.3. Ödevler..... | 77 |

| | |
|--|----|
| 5.3.2. Birliktelik kuralı çıkarma | 78 |
| 5.3.2.1. Birlikte sık girilen sayfalar..... | 79 |
| 5.3.2.2. Birlikte yanlış yapılan sorular | 83 |
| 5.3.3. Ağırlıklı notlar..... | 85 |
| | |
| SONUÇLAR VE ÖNERİLER | 89 |
| KAYNAKLAR | 92 |
| ÖZGEÇMİŞ | 96 |

ŞEKİLLER DİZİNİ

| | |
|---|----|
| Şekil 1.1. Web kullanım madenciliği | 3 |
| Şekil 2.1. Veri tabanlarında bilgi keşfi süreci | 11 |
| Şekil 2.2. Veri madencili ve diğer disiplinler | 14 |
| Şekil 3.1. Örnek bir sınıflandırma..... | 29 |
| Şekil 3.2. Apriori algoritmasının gösterimi | 33 |
| Şekil 3.3. FP-Ağaç Yapısı..... | 36 |
| Şekil 3.4. Doğrudan yoğunluk erişilebilir noktalar | 37 |
| Şekil 3.5. Yoğunluk erişilebilir noktalar | 37 |
| Şekil 3.6. Yoğunluk bağlı noktalar | 38 |
| Şekil 3.7. DBSCAN algoritması çalışma yapısı | 38 |
| Şekil 3.8. DBSCAN Uygulaması | 40 |
| Şekil 3.9. Eps=5 ve MinPts=7 için DBSCAN algoritması sonucu..... | 41 |
| Şekil 3.10. Eps=7 ve MinPts=5 için DBSCAN algoritması sonucu..... | 41 |
| Şekil 3.11. Eps=10 ve MinPts=5 için DBSCAN algoritması sonucu..... | 41 |
| Şekil 4.1. İnternet Tabanlı Öğretim Sistemlerinin Genel Mimarisi..... | 52 |
| Şekil 4.2. Şifre Giriş Ekranı..... | 56 |
| Şekil 4.3. Anket Formu..... | 57 |
| Şekil 4.4. Bölüm girişleri..... | 59 |
| Şekil 4.5. Ders İçerikleri..... | 59 |
| Şekil 4.6. Quiz Sınavı..... | 61 |
| Şekil 4.7. Ödevler..... | 62 |
| Şekil 4.8. Kaynak Linkler..... | 62 |
| Şekil 5.1. Açılış Ekranı | 65 |
| Şekil 5.2. Programın Açılış Ara yüzü..... | 66 |
| Şekil 5.3. Genel Değerlendirme Ekranı..... | 67 |
| Şekil 5.4. Sayfa Giriş Bilgileri..... | 67 |
| Şekil 5.5. Anket – Giriş Değerlendirilmesi..... | 70 |
| Şekil 5.6. Öğrencinin Kaldığı Yer Niteliğine göre Değerlendirme Sonuçları..... | 71 |
| Şekil 5.7. Öğrencinin Kaldığı Yer ve Bitirdiği Lise Niteliğine Göre Karşılaştırılması..... | 72 |
| Şekil 5.8. Quiz Sonuçları..... | 74 |
| Şekil 5.9. Girişlere Göre Quiz Değerlendirmesi..... | 75 |
| Şekil 5.10. Bölümlere Göre Quiz Değerlendirmesi..... | 76 |
| Şekil 5.11. Ödevler..... | 77 |
| Şekil 5.12. Birliktelik Kuralı Çıkarma Menüsü..... | 78 |
| Şekil 5.13. Birlikte En Sık Girilen Sayfalar..... | 79 |
| Şekil 5.14. Sık Tekrarlanan Öğeler..... | 81 |
| Şekil 5.15. Üretilen Birliktelik Kuralları..... | 82 |
| Şekil 5.16. Birlikte Yanlış Yapılan Sorular..... | 83 |
| Şekil 5.17. Birlikte Yanlış Yapılan Sorular..... | 84 |

| | |
|---|----|
| Şekil 5.18. Ağırlıklı Notlar Arayüz Ekranı. | 85 |
| Şekil 5.19. Not Yakınlığı=5 ve Minimum Öğrenci=7. | 87 |
| Şekil 5.20. Not Yakınlığı=3 ve Minimum Öğrenci=5. | 87 |
| Şekil 5.21. Not Yakınlığı=7 ve Minimum Öğrenci=10. | 88 |

TABLolar DİZİNİ

| | |
|---|----|
| Tablo 3.1. Marketten Yapılan Alışveriş Bilgilerini İçeren Veritabanı | 31 |
| Tablo 3.2. Veritabanı Örneği | 35 |
| Tablo 4.1. Türkiye’de ITÖ Programına Sahip Üniversiteler | 46 |
| Tablo 4.2. Türkiye’de ITÖ Hizmeti Veren Özel Sektör Faaliyetlerinden Bazıları. | 47 |
| Tablo 4.3. “departman” Tablosu. | 54 |
| Tablo 4.4. “ogrenci” Tablosu. | 54 |
| Tablo 4.5. “konu” Tablosu. | 55 |
| Tablo 4.6. “bolum” Tablosu. | 55 |
| Tablo 4.7. “quiz” Tablosu. | 55 |
| Tablo 4.8. “soru” Tablosu. | 55 |
| Tablo 4.9. “odevler” Tablosu. | 56 |
| Tablo 4.10. “anket” Tablosu. | 58 |
| Tablo 4.11. “giris” Tablosu. | 60 |
| Tablo 4.12. “cevap” Tablosu. | 60 |
| Tablo 4.13. “quiznotu” Tablosu. | 61 |
| Tablo 5.1. Öğrencilerin siteye yapmış oldukları toplam ziyaret sayısı. | 68 |
| Tablo 5.2. Parametreye Göre Filtreleme İşlemi. | 68 |
| Tablo 5.3. Öğrencilerin Sitede Kaldıkları Toplam Süre. | 69 |
| Tablo 5.4. Anket ve Giriş Tablolarının Birleştirilmesi | 69 |
| Tablo 5.5. Seçilen Niteliğin Sayı-Oran Değerlerinin Hesaplanması. | 71 |
| Tablo 5.6. Seçilen Quizin Sorularına Verilen Cevapların Doğruluk Oranlarını Hesaplama. | 73 |
| Tablo 5.7. Departmanlara Göre Ortalama Hesaplama. | 76 |
| Tablo 5.8. Ödevlerle İlgili İşlemleri Yapan SQL Komutları. | 78 |
| Tablo 5.9. Minimum Destek Eşik Değerini Geçen Sayfaları Bulan SQL Komutu | 79 |
| Tablo 5.10. Sık Girilen Sayfaların İkili Kombinasyonlarının Oluşturulması ve Veritabanında Taranması. | 80 |
| Tablo 5.11. Üçlü Öğeleri Veritabanında Arayan SQL Komutu. | 81 |
| Tablo 5.12. Birlikte Yanlış Yapılan Birli Öğe Kümesini Bulan SQL Komutu. | 83 |
| Tablo 5.13. İkili Öğeleri Bulan SQL Komutu. | 84 |
| Tablo 5.14. Üçlü Öğeleri Bulan SQL Komutu. | 84 |
| Tablo 5.15. DBSCAN Algoritmasını Çalıştıran Kodlar. | 86 |

SİMGELER DİZİNİ ve KISALTMALAR

VTBK : Veri Tabanlarında Bilgi Keşfi

VM : Veri Madenciliği

ITÖ : İnternet Tabanlı Öğretim.

KDD : Knowledge Discovery in Databases

D : Destek (Support).

C : Güven (Confidence).

DBSCAN : Density Based Spatial Clustering of Applications with Noise.

INTERNET TABANLI ÖĞRETİMDE VERİ MADENCİLİĞİ TEKNİKLERİNİN UYGULANMASI

Ümmühan ALTINTOP

Anahtar Kelimeler: İnternet Tabanlı Öğretim, Veri Madenciliği, Birliktelik Kuralları, Apriori Algoritması, Kümeleme Teknikleri, DBSCAN Algoritması.

Özet: Bu çalışmada, İnternet Tabanlı bir öğretim sistemi tasarlanmıştır. Uygulama kapsamında tasarlanan site, Kocaeli Üniversitesi Sağlık Yüksekokulu ve Arslanbey Meslek Yüksekokulu öğrencilerinin eğitiminde uygulamalı olarak kullanılmıştır. Öğrencilerin site üzerindeki hareketleri bir veritabanında 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 artırıcı yönde kullanabilmek amaçlanmıştır.

APPLICATION OF DATA MINING TECHNIQUES AT THE INTERNET BASED EDUCATION

Ümmühan ALTINTOP

Keywords: Internet Based Education, Data Mining, Association Rules, Apriori Algorithm, DBSCAN Algorithm, Clustering Techniques.

Abstract: In this study, an Internet Based Education System is projected. Within the context of application, projected site Health College of Kocaeli University and Arslanbey Professional College has been used practically for the student's education. Behaviors of the students in the site have been saved on a database. Association rules of data mining models has been applied on these data, which are collected from students. With using apriori algorithm, which is one of the associations rule technuques, connected page pair frequency and modal wrong answers, which were given to the published evaluation question pair on the designed site, are determined. Students were classified according to their assignments and results of their examinations by implementing of DBSCAN algorithm, that one of the clustering techniques. Through this structure, it is aimed that published class notes shall become a formation, which are adapted according to the student's aims, standard of attainments and learning methods and shall increase performances of students and lecturers.

BÖLÜM.1

1.1. Giriş

İnsanođlu ilk çağlarda mağara duvarlarını kazıyarak verileri saklamaya çalışmıştır. Daha sonra kağıdın icadıyla verileri kağıt üzerinde toplamaya başlamıştır. Matbaanın icadıyla ise bu veriler kitap haline getirilmiştir. Teknolojinin gelişme ve verilerin dijital ortamda saklanmaya başlanması ile birlikte, yeryüzündeki bilgi miktarı her geçen gün katlanarak artmıştır. Veri tabanlarının sayısı da benzer, hatta daha yüksek oranlarda artmıştır.

Yüksek kapasiteli işlem yapabilme gücünün ucuzlaması ile birlikte veri saklama işlemi kolaylaşmıştır. Fakat son yıllarda, veriyi toplama ve saklama kapasitesindeki çok ani büyüme, yeni arayışlara yol açmıştır. Bir bilgisayarın işleyebileceği veriden daha fazlası üretilmektedir. Verilerin bu hızla büyümesi, yorumlama ve özümsemede akıllı veritabanı analizi için, yeni nesil araçlara ve tekniklere olan ihtiyacı doğurmuştur. Geleneksel sorgu veya raporlama araçları veri yığınları karşısında yetersiz kalmıştır.

Büyük veritabanlarından değerli, ilginç ve önceden bilinmeyen bilgiyi keşfetmek için pratik uygulamalar ve olası çözümler için önemli ve aktif bir araştırma alanı olan, veritabanlarında bilgi keşfi (VTBK) ortaya çıkmıştır.

Veritabanlarında bilgi keşfi (VTBK, Knowledge Discovery in Databases, KDD), verideki geçerli, yeni, kullanışlı ve anlaşılır bilgiyi ortaya koyma işlemidir. Diğer bir deyişle uygun ölçütlere göre ilginç bilgiyi çekip çıkartma işlemidir. VTBK süreci içerisinde modelin kurulması ve değerlendirilmesi aşamalarından meydana gelen veri madenciliği (Data Mining) en önemli kesimi oluşturmaktadır. Bu önemden ötürü bir çok kaynakta VTBK ile veri madenciliği eş anlamlı olarak kullanılmaktadır. [1]

Veri madenciliđi, elektronik ticaret, bilim, tıp, iř ve eđitim alanlarında yeni ve temel bir arařtırma sahası olmuřtur. Eldeki anlamsız veriden, anlamlı ve kullanıřlı bilgiyi ıkarmaya yarayacak iřlemleri analiz ve uygulamaya ynelik alıřmalar yapılmıřtır. Geniř veri yuđunlarından desenleri, deđiřiklikleri ve iliřkileri ıkarmakta kullanılan veri madenciliđi, Internet üzerinde filtrelemeler, Dna ierisinde genlerin tespiti, ekonomideki eđilim ve dzensizliklerin tespiti, elektronik alıřveriř yapan mřterilerin alıřkanlıkları gibi karar verme mekanizmaları iin nemli bulgular elde etmede kullanılmıřtır.

Bunların yanında sayısal veri miktarının byk miktarda artıř gstermesi fakat, buna karřılık bilim adamlarının, mhendislerin ve analistlerin sayılarının bu oranla artamıyor olması, veri madenciliđi ile ilgili yeni arařtırma problemlerini ortaya ıkarmıřtır. Bu problemlerin zmlerini birkaç grupta toplanabilir.

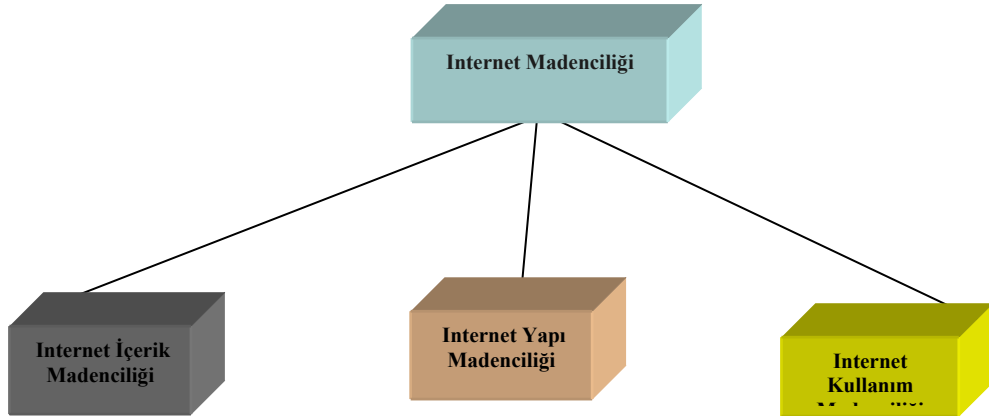
1. Yeni veri tiplerinin madenciliđi iin yeni algoritma, teknik ve sistemlerin geliřtirilmesi,
2. Geniř hacimli ve ok boyutlu veri madenciliđi iin yeni algoritma ve sistemlerin geliřtirilmesi,
3. Dađıtık veri madenciliđi iin algoritma, protokol ve alt yapılarının geliřtirilmesi,
4. Mevcut veri madenciliđi sistemlerinin kullanımının ilerletilip geliřtirilmesi,
5. Veri madenciliđi iin zel gizlilik ve gvenlik modellerinin geliřtirilmesi.

Gnmzde de veri madenciliđi teknikleri her alanda kullanılmaya ve geliřtirilmeye devam etmektedir. Internet üzerinde bilgi kaynaklarının byk oranda artmasıyla istenilen bilgi kaynaklarına ulařmada veri madenciliđi teknikleri kullanılarak Internet madenciliđi ortaya ıkmıřtır. Internet madenciliđi, web’de kullanıřlı bilgilerin keřfi ve analiz iin geliřtirilen bir yntemdir [2]. Web madenciliđi terimi ilk olarak Etzioni tarafından 1996’da ortaya ıkmıřtır. Etzioni, web madenciliđinin, veri madenciliđi tekniklerini kullanarak Worl Wide Web’de bulunan dosya ve servislerden otomatik olarak desenler bulmak ve ngrlmeyen bilgiye ulařmak olduđunu ileri srmektedir [3].

Web sitelerinin bulunduğu sunucular üzerindeki erişim ve hata kayıt dosyalarında kullanıcının site içinde gezinirken yaptığı her tıklama işlemi bir ya da birden çok hareket kaydı biriktirir. Bu veriler site içerik verisi ve kayıtlı kullanıcılara ait veri ile birleştirildiğinde fayda sağlanabilecek bir veritabanı oluşturmaktadır. [2]

Bu veritabanı sayesinde kullanıcıların profilleri çıkarılabilir. Zaman içinde değişimleri takip edilebilir, sitedeki beğenilen ya da beğenilmeyen köşeler tespit edilebilir, kullanıcın gezinti şekli, sitenin içerik yapılandırma ve altyapı açısından performansı hakkında bilgi verir.

Madria'ya göre ise Internet madenciliği, Internet'de bulunan veriden faydalı bilgiye ulaşmaktır. Internet 3 tip veri bulundurur: içerik (content), Internet log dosyaları (usage)ve Internet yapı (structure) verisidir.



Şekil 1.1. Internet Kullanım Madenciliği.

Internet sitelerinin bulunduğu sunucular üzerindeki erişim ve hata kayıt dosyalarında kullanıcının site içinde gezinirken yaptığı her tıklama işlemi bir ya da birden çok hareket kaydı biriktirir. Bu veriler site içerik verisi ve kayıtlı kullanıcılara ait veri ile birleştirildiğinde fayda sağlanabilecek bir veritabanı oluşturmaktadır [2] .

Bu yapıdaki bir veritabanından, kullanıcıların profilleri çıkarılabilir. Zaman içinde değişimleri takip edilebilir, sitedeki beğenilen ya da beğenilmeyen köşeler tespit

edilebilir, kullanıcın gezinti şekli, sitenin içeriğini yapılandırma ve altyapı açısından performansı hakkında bilgi verir.

İnternet İçerik Madenciliği: İnternet kaynaklarından otomatik bilgi arama tekniklerini içerir. Otomatik anahtar kelime arama ötesinde metinler içindeki bilinen yapıları bazı veri modellerine indirgeme yöntemidir [2].

İnternet Yapı Madenciliği: İnternet sitesi ve web sayfası hakkında bağlantı verisine bakarak bilgi üretmektir. İnternet içerik madenciliği doküman içeriğine, yapı madenciliği ise dokümanlar arası bağlantılara yoğunlaşır [2].

İnternet Kullanım Madenciliği: kullanıcıların İnternet’de dolaşırken yaptığı erişim hareketlerince oluşturulan veride, bilgi üretmeyi sağlar. Bu konuda bir çalışma alanı olan site güncelleştirme sistemleri, site içerik ve yapısında gerekli düzenlemeleri yapmaktadır. Diğer bir çalışma alanı olan sistem iyileştirme, İnternet kullanım verisini kullanarak trafiği etkinleştirmeyi amaçlar. Son olarak kişileştirme çalışmaları ise, kişisel taleplere göre düzenlenen siteler tasarlamaya çalışır [2].

Bu tez çalışmasında internet üzerinde yayınlanmış bir uygulamaya bağlanan öğrencilerin oluşturduğu kayıtlar alınarak, bu kayıtlara veri madenciliği birliktelik kurallarından apriori algoritması ve kümeleme modelinde DBSCAN algoritması uygulanmıştır. Tez çalışması hazırlanırken literatürde,İnternet tabanlı öğretim, Veri madenciliği, Apriori algoritması ve DBSCAN algoritması üzerinde yapılmış olan çalışmalar taranarak, doküman hazırlanmıştır. Bu tez çalışması, öğrenci performansını değerlendirmek, öğrenci örüntüleri çıkarmak amacıyla yapılan İnternet tabanlı öğretim uygulaması hakkında bilgilendirici bir kaynak olabilir.

Bidgoli ve diğ. (2002), İnternet tabanlı bir sistemde yapılan öğretimde alınan log kayıtları üzerinde veri madenciliği uygulayarak öğrenci performansını değerlendirmişlerdir. Çalışmada, öğrenciler için internet tabanlı fizik kursu açılmış ve 261 öğrenci katılmıştır. Kursta, 12 ödev ve final sonuçları değerlendirilerek öğrenciler kümelere (gruplara) ayrılmış, öğrenciler başarılarına göre 9 ayrı sınıfa

ayrılmıştır. Veri madenciliğinin Sınıflama modellerinden k-en yakın komşu (k-nearest neighbor) metodu kullanılarak LON-CAPA isimli programı yazmışlardır [4].

Chiang ve diğ.(2004), Veritabanlarında doğrusal korelasyon keşfi için bir veri madenciliği uygulaması yapmışlardır. Doğrusal (linear) korelasyon buluşu (VCD) için statik metotla otomatik birleştirmenin nasıl sağlanacağı hususunda çalışmışlardır. Veritabanında doğrusal korelasyonların buluşu için fonksiyonel ölçümler yaparak büyük bir veritabanı elde etmişler, elde ettikleri bu veritabanı içinde anlamlı veri üretebilmek için veri madenciliği modellerinden sınıflama modelini kullanmışlardır. Bu verileri sınıflama algoritması yardımıyla gruplara ayırmış ve doğrusal korelasyon ölçümlerini bulabilmek için uygun verilerin bulunduğu gruba karar vermişlerdir [5].

Tang ve McCalla (), Internet tabanlı öğretim sisteminde benzer örüntüdeki sayfalara ağırlık değerleri vererek, Internet madenciliği tekniklerini uygulamışlar, yakın filtreleme teknikleri kullanarak web sayfalarını gruplamışlardır. Çalışmada, öğrenciler belli kümelere ayrılarak sayfalara ulaşmaları sağlanmıştır. Yapılan çalışmada 250 sayfa ders notu ve 2 ek sözlük yayınlanmıştır. Internet üzerinden yapılan öğretimde her bir öğrencinin bağlandığı her sayfa için, sayfaya ziyaretinin toplam sayısı ve her bölümde harcamış olduğu zaman alınmıştır. Daha sonra toplanan bu veriler üzerinde Internet kullanım madenciliğini uygulamışlardır. Internet sitesini ziyaret eden öğrencilerle ilgili olarak, kullanıcıların en sık eriştiği sayfa çiftleri, okul içi ve dışı kullanıcı erişim dağılımı gibi tanımsal ilişkiler tespit etmişlerdir. Bu gibi ilişkiler aynı zamanda Internet sitesinin yapılandırılmasında da faydalı olacaktır [6].

Luan ve Ph (2002), yüksek eğitimde veri madenciliği ve bilgi keşfi isimli yaptığı çalışmada bilgi yönetimi bağlamında veri madenciliğinin nereye uygun olduğunu araştırmıştır. Yüksek eğitimde veri madenciliği tekniklerinden nasıl yararlanılabileceği, bir öğrencinin başarısı ve performansı en iyi şekilde veri madenciliği modelleri kullanılarak en iyi şekilde nasıl tahmin edilebileceği hakkında çalışma yapmışlardır [7].

Minaei ve diğ. (2004), Lon-Capa isimli yaptığı çalışmada Internet sunucusunun sistem log kayıtlarına Internet kullanım madenciliği sistemi, veritabanı yaklaşımı kullanılarak uygulamışlardır. Yapılan uygulamada ilişkisel kurallar keşfi için genetik algoritmalar kullanılmış ve web sitesini ziyaret eden kullanıcılarla ilgili olarak çeşitli desenler elde etmişlerdir. Web sitesine bağlanan öğrenciler gruplandırılarak toplanmış olan verilere göre final notları tahmin edilmeye çalışılmışlardır [8].

Kuo ve diğ. (2004), yapay sinir ağları ve genetik algoritmalar kullanarak elektronik alışverişte web tarayıcı yollarını analiz etmişlerdir. 2000 yıllarının sonlarına doğru Internet kullanıcılarının artmasıyla elektronik ticaret de artmıştır. Yapmış oldukları çalışmada kullanıcıların web sayfalarına tıklama frekansı ve izlenen yol verileri ele alınarak yapay sinir ağları ve genetik algoritmalar kullanarak web kullanım madenciliği yöntemi uygulanmıştır[9].

Tang ve McCalla (2000), AIED isimli web tabanlı bir öğretim sistemi hazırlamışlardır. Öğrenci profillerini değerlendirmede AIED sistemi yardımıyla bir çok veriyi toplamışlardır. Elde ettikleri veritabanı üzerinde verimadenciliğinin kümeleme modelini uygulayarak öğrenci desenleri oluşturmuş, öğrencileri gruplandırmışlardır. Kümelenen öğrenci grupları için, belirlenen profillere göre çeşitli kurs içerikleri belirlenmesini sağlamışlardır [6].

Agethe ve Kalina (2001), öğretim ve öğrenimi geliştirmek için veri madenciliği tekniklerini kullanarak web tabanlı bir araç (tool) geliştirmişlerdir. Sydney üniversitesinde yayınlamış oldukları Logic-ITA isimli web tabanlı akıllı öğretim sistemi yardımıyla öğrenci hakkında bilgileri toplama ve analiz etme imkanı doğmuştur. Alınan veritabanı sayesinde öğrenci performansını değerlendirmek için veri madenciliği tekniklerinden birliktelik kuralları uygulanarak yapılan hatalar değerlendirilmiş yeni araçlar geliştirilmiştir. Logic-ITA uygulamasına 721 öğrenci katılmış, 10 deneme sınavı yapılmıştır. En çok yanlış yapılan soru çiftleri veri madenciliği tekniklerinden apriori algoritması kullanılarak bulunmuştur [10].

Hang ve diğ. (1999), web madenciliği hakkında araştırma yapmışlardır. Web madenciliği tanımı ve web üzerinde bilgi madenciliği arasındaki ilişki araştırılmışlar,

veri madenciliği teknikleri kullanarak web verileri üzerinde bilgi keşfi ve analiz yapmışlardır. İnternet içerik madenciliği ile, dokümanların içerisinden bilgi çıkararak, kullanıcı davranışlarını bulabilmeye çalışmışlardır. WebTMS ismini verdikleri bu çalışmada, ilk olarak metinleri kategorilere, sonra kümelere ayırmışlardır. Düzenlenmiş olan küme yapılarına birliktelik kuralları uygulayarak birbirleriyle ilişkisi olan kelime ve metin gruplarının bulunduğu sayfaları bulmuşlardır [11].

Jea ve diğ.(2002), OLA (Online Combinatorial Approximation) isimli çalışmalarında, web üzerindeki geniş veri gruplarına online madencilik yaparak yararlı ve esnek algoritma bulmaya çalışmışlardır. Günlük işlemler dolayısıyla, toplanan ve üretilen büyük hacimli verileri, web sunucular aracılığıyla otomatik olarak sunucu veya erişim günlüklerinde toplamışlardır. Bu günlüklerdeki verileri ayrıştırarak ve analiz ederek değerli bilgiyi çıkarmak için apriori algoritmasına benzer yapıda olan OCA isimli algoritmayı oluşturmuşlardır [12].

Zaiane, O.R., (2004), daha iyi bir web tabanlı öğretim için web kullanım madenciliğini araştırmıştır. Web kullanım madenciliği bir veya birçok web sunucudan kullanıcı erişim desenlerinin otomatik olarak keşfinin ve analizinin yapıldığı veri madenciliği etkinliğidir. Yapmış olduğu çalışmada veri madenciliği ve makine öğrenimi tekniklerini kullanarak web tabanlı öğrenimi genişletmeye çalışmıştır. Web tabanlı bir kurs açılarak log kayıtları bir veritabanında tutulmuştur. Bu log kayıtları üzerinde gereksiz verileri kaldırarak, erişim oturumları tanımlamıştır. Kullanıcı erişim desenlerinin keşfini ve analizini yapmıştır. Log kayıtlarının analizi sonucunda bir ziyaretçinin sitede kalma süresi, hizmet stratejileri bulunarak web tabanlı bir öğretim programının genişletilmesi sağlanmıştır [13].

Agrawal ve Srikant (2002), çalışmalarında müşteri odaklı süpermarket tasarımına odaklanmıştır. Etkili bir tasarım için müşterilerin satın alma alışkanlıkları araştırılmıştır. Marketlerde her bir satış için verilen fişlerden ötürü büyük bir veri saklanmaktadır. Bilgisayar destekli veri madenciliği araçları kullanılarak yapılabilecek sepet analizinin, insanların hangi ürünleri birlikte alabileceğinin düşünülmesi yerine, satış verilerine bakılarak ve analiz edilerek bu ilişki sepet analizi

aracılıđıyla ortaya ıkarmıřlardır. Spermarketin verimliliđinin arttırılması aısından, mřteri hareketleri verilerine dayanarak sepet analizi kullanılarak reyon dzeni ve rn grupları, bir sonraki ayın satıř tahminlerinin ıkarılması, mřterileri satın aldıkları rnlere bađlı olarak gruplandırılması, yeni rnlerin potansiyel mřterilerinin belirlenmesi alıřmasını yapmıřlardır [14].

Tsay ve Chiang (2002), CBAR (Cluster-Based Association Rule) isimli alıřmalarında veri madenciliđinin birliktelik kurallarının uygulamasını gerekleřtirmiřtir. Veritabanında toplamıř oldukları veriler zerinde CBAR uygulayarak (Kmeleme tabanlı birliktelik Kuralı) yeni tablolar oluřturmuřlardır. Belirlenen bir k uzunluđuna gre her tablodaki verilere apriori algoritmasını uygulamıřlardır [15].

Tezde arařtırma ve geliřtirme sreci altı blmde ele alınmıřtır. Birinci blmde, bu tez alıřmasına niin gerek duyulduđu ve ne amala byle bir alıřmanın gerekleřtirildiđi aıklanmıřtır. Bu tezi meydana getiren blmler ve her blmde incelenen konular zet halinde birinci blmde aıklanmıřtır.

İkinci blmde, veri madenciliđinin tanımı, veri madenciliđine neden ihtiya duyulmuřtur, kullanılan teknikler, veri madenciliđi ve diđer disiplinler arasındaki iliřki, veri madenciliđinin hangi tr veriler zerinde uygulanabilir, veritabanlarında bilgi keřfi ařamaları incelenmiřtir.

nc blmde, veri madenciliđi tekniklerinden birliktelik kurallarına ait apriori algoritması ve kmeleme analiz tekniklerinde kullanılan DBScan algoritması geniř bir řekilde incelenmiřtir.

Drdnc blmde, İnternet tabanlı đretim, “Eđitim ve đretim”in insan hayatındaki yeri, tewknolojinin geliřmesiyle birlikte “Eđitim ve đretim” de klasik đretim řeklinin deđiřmesi incelenmiřtir. İnternet tabanlı đretim zerinde durulmuř Kocaeli niversitesi’nde yapılan bir alıřma ele alınmıřtır.

Beşinci bölümde, İnternet tabanlı öğretimde alınan veriler üzerinde veri madenciliği tekniklerinden apriori algoritması ve kümeleme analizi tekniklerinden DBSCAN algoritmasının uygulanmasına ilişkin Delphi programlama dilinde yapılan uygulama açıklanmıştır.

Altıncı bölümde, gerçekleştirilen çalışmadan çıkan sonuçlar açıklanmış ve bu sonuçlara bağlı kalınarak bazı öneriler sunulmuştur.

BÖLÜM 2. VERİ MADENCİLİĞİ ve KULLANILAN TEKNİKLER

2.1. Giriş

Üretilen sayısal bilginin sürekli arttığı buna paralel olarak veri tabanlarının daha fazla veriyi saklayabilecek boyutlara ulaştığı ve bilgisayar sistemlerindeki gelişme ile veriye ulaşmanın kolaylaştığı günümüzde, bilgisayar sistemleri her geçen gün ucuzlamakta ve aynı zamanda güçleri de artmaktadır. Bu sayede doğru ve daha detaylı bilgiye ulaşmamız mümkün hale gelmiş fakat başka bir sorunu ortaya çıkarmıştır. Bu sorun, oluşan bu büyük sayısal veri yığınlarının yönetilmesi ve anlamlı hale getirilmesidir.

Şirketlerin bilgi sistemleri üzerinden ürettiği bilgi miktarının büyük artış gösterdiği ve firmaların veritabanlarının boyutlarının yüksek hacimlere ulaştığını görmekteyiz. İşte veritabanındaki bu teknolojik gelişme ve veri miktarındaki bu artış, eldeki bu verilerden nasıl faydalanılacağını ve bu verilerin nasıl anlamlı hale getirileceği sorununu ortaya çıkarmıştır.

Bilgisayar sistemleri ile üretilen bu veriler, tek başlarına değersizdirler ve bir anlam ifade etmezler. Bu veriler, belli bir amaç doğrultusunda işlendiği zaman anlamlı hale gelmektedir. İşte ham veriyi, bilgiye veya anlamlı hale dönüştürme sürecine Veri Tabanlarında Bilgi Keşfi (VTBK) süreci denmektedir.

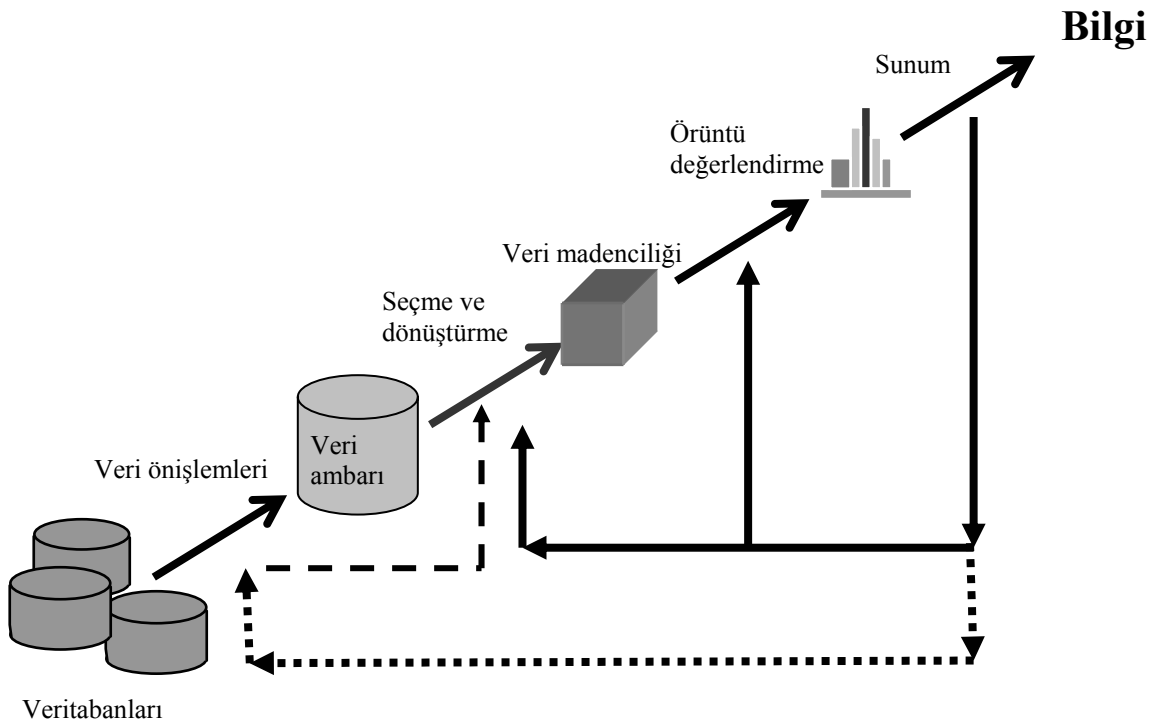
2.2. Veri Tabanlarında Bilgi Keşfi Süreci (VTBK)

Dijital dünyadaki teknolojik gelişmeler, kullanılan ve saklanması gereken veri miktarını her geçen gün arttırmaktadır. Boyutları hızla artan veriden anlamlı bilgiler çıkarmak için bilgisayar hızlarının ve güçlerinin artmasını sağlayacak yeni teoriler ve araçlar geliştirilmektedir. Bu teoriler ve araçlar, VTBK sürecinin konusunu oluşturmaktadır.

Veriden anlamlı örüntüler çıkarma sürecine literatürde, veri madenciliği, bilgi çıkarımı (knowledge Extraction), bilgi keşfi, veri arkeolojisi ve veri örüntü işleme (data pattern processing) gibi isimler verilmektedir. İlk olarak 1989 yılında yapılan bir atelyede veri işleme sürecinde bilginin son ürün olduğunu vurgulamak için “veri tabanlarında bilgi keşfi” tanımlaması yapılmıştır [16].

Veri tabanlarında bilgi keşfi, veriden anlamlı ve yararlı bilginin çıkarıldığı süreç olarak tanımlanmaktadır. Bu anlamda veri madenciliği (VM), bu sürecin sadece bir kısmını oluşturmaktadır. Bilgi keşfi sürecinde amaç, büyük veri kümelerindeki düşük seviyedeki veriden yüksek seviyede bilgi çıkarımını sağlamaktır.

VTBK, verinin nasıl saklanması ve algoritmaların büyük veri kümelerine nasıl uygulanması gerektiği, sonuçların nasıl yorumlanacağı sorularının cevabını arama aşamalarıdır.



Şekil 2.1. Veri Tabanlarında Bilgi Keşfi Süreci [17].

VTBK sürecini oluşturan aşamalar:

1. Veri Önışlemleri: Bu aşamada öncelikle veriler içindeki gürültüler, tutarsızlık ve düzensizlikler giderilir. Bu işleme veri temizleme denir. İkinci aşamada veri birleştirme işlemi uygulanır. Bu aşamada çeşitli kaynaklardan gelen verilerin tek bir veri ambarında toplanabilmesi için gerekli genelleme ve uyumluluk işlemleri yapılır.

2. Veri Seçme ve Dönüştürme (Data Selection): Bu aşamada, veri madenciliğinin sağlıklı yapılabilmesi için veriler üzerinde önışlemler yapılır. Bu önışlemler:

- Veri madenciliği konusu ile ilgili bilgi seçimi.
- Madencilik yapılacak veri türünün belirlenmesi.
- Veriler arasında hiyerarşik yapı ve genellemelerin belirlenmesi.
- Veri madenciliği sonunda bulunacak bilgi için yenilik ve ilginçlik ölçümü yöntemlerinin belirlenmesi.
- Veri madenciliği sonunda bulunacak veri için sunum ve görselleştirme araçlarının belirlenmesi.

Tüm bu önışlemleri gerçekleyebilmek için bir veri madenciliği sorgulama dili kullanılır.

3. Veri Madenciliği: İnsanoğlu için anlamlı veri örüntüleri ortaya çıkarmak için çeşitli algoritmaların kullanıldığı aşamadır. İlerleyen sayfalarda bu işlem detaylı olarak anlatılmıştır.

4. Örüntü Değerlendirme: İkinci aşamada belirlenen ilginçlik ölçüm yöntemleri kullanılarak veri madenciliği ile bulunan verilerin ne kadar ilginç ve yararlı olduğu tespit edilir.

5. Bilgi Sunumu: Çeşitli görselleştirme ve raporlaştırma araçları kullanılarak bulunmuş olan veriler ilgili kullanıcılara sunulur.

VTBK süreci defalarca tekrar ve aşamalar arası atlamalar ve ileri geri hareketler içerebilmektedir. Günümüzde çoğunlukla veri madenciliği aşamasına odaklanılmakta, fakat diğer tüm aşamalar VTBK işleminin bütünlüğü açısından en az veri madenciliği kadar önemlidir [17].

2.3. Veri Madenciliği

Veri kendi başına değersiz olduğundan, verinin amacımız doğrultusunda bilgiye çevrilmesine veri analizi denmektedir. Büyük miktarlardaki ve oldukça hızlı toplanan verilerin çeşitli analizler sonucunda anlamlı bilgilere dönüştürülmesi sürecine veri madenciliği denmektedir. Veri madenciliği tanımları incelendiğinde, ortak nokta “ çok büyük” miktarlarda veri ve bu verilerden “anlamlı” bilgiler elde edilmesidir.

Veri madenciliği, veri tabanlarında tutulan çok çeşitli verilerden, daha önce keşfedilememiş bilgileri oraya çıkarmaktır. Veri madenciliği, kendi başına bir çözüm değil, çözüme ulaşmak için verilecek karar sürecini destekleyen, problemi çözmek için gerekli olan bilgileri sağlamaya yarayan bir araçtır [18].

Veri madenciliği 1990’larda ortaya çıkan ve dünyada yeni yaygınlaşan bir kavramdır. Veri madenciliği uygulamalarında kullanılan ve büyük miktarlarda verilerin tutulduğu veri ambarları da son yıllarda önem kazanan bir yapı olmuştur. Bilişim sektöründe veri ambarı ve veri madenciliği ile ilgili donanım ve yazılım ürünleri üstünde çalışan firmalar son yıllarda artış göstermektedir.

2.4. Veri Ambarı

Karar verme sürecinde kullanılan, konu tabanlı, birleştirilmiş, zamana bağımlı, verilerin sabit olduğu veri topluluğuna veri ambarı denmektedir [19]. Veri ambarı olmayı gerektiren bu dört özelliği kısaca açıklamak gerekirse:

Konu tabanlı: Veri ambarları belirli bir konuda veriler içerir. Örneğin satış verileri, müşteri bilgileri vs..

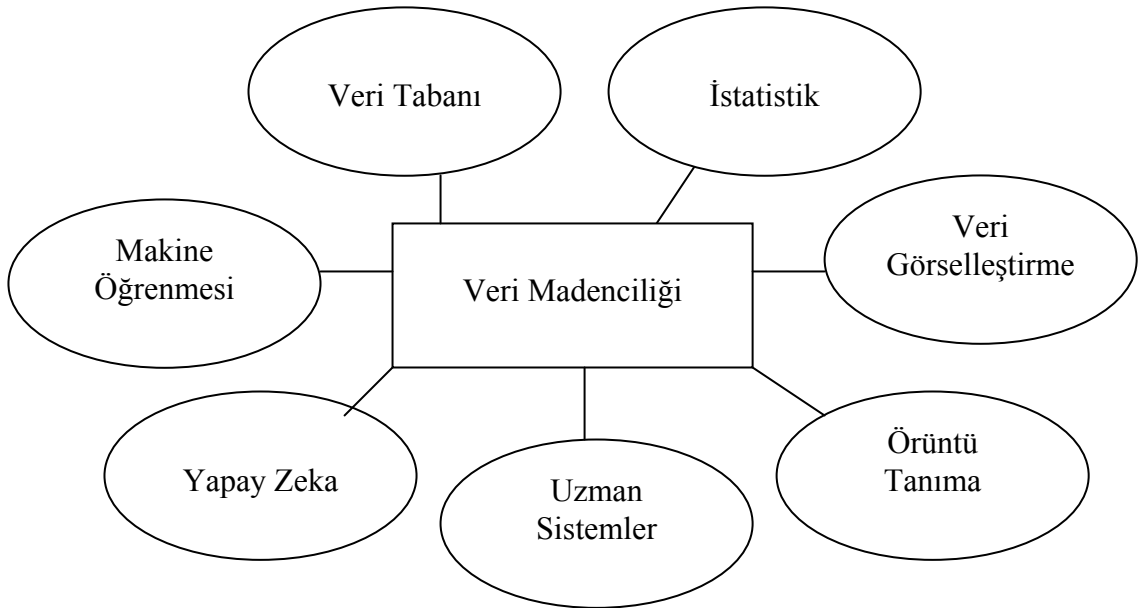
Birleştirilmiş : Veri ambarı birçok farklı kaynaktan gelen bilgilerin toplanması ile kurulur. Örneğin bir veri ambarı içinde ilişkisel veritabanları, düz metin dosyaları, işlemsel veritabanları bulunabilir.

Zamana bağımlı: Veri ambarlarında bilgiler periyodik aralıklarla eklenir. Veri ambarındaki her bir anahtar yapı, tarihsel olarak dizilmiş olmalıdır. Örneğin günlere göre son beş yılın satış rakamları.

Sabit: Veri ambarında veriler işlevsel veritabanlarında olduğu gibi sürekli güncellenmez. Veri ambarına eklendiği andan itibaren sabit olarak kaydedilir.

2.5. Veri Madenciliği ve Diğer Disiplinler

Veri madenciliği, makina öğrenmesi, örüntü tanıma, veritabanı teknolojileri, istatistik, yapay zeka, uzman sistemler, veri görselleştirme (data visualization) alanlarının bir kesişim noktası olarak doğmuştur ve bu yönde gelişmesini sürdürmektedir [20]. Bu yapı temel olarak Şekil 2.2’de görüldüğü gibi sembolize edilebilir.



Şekil 2.2. Veri Madenciliği ve Diğer Disiplinler [21]

Makine öğrenmesi, örüntü tanıma ve istatistik alanları veri madenciliğinde örüntü keşfetme aşamasında; yapay zeka teknolojileri bulunan örüntüleri yorumlama aşamasında; veritabanı teknolojileri, eldeki verileri depolama, süzme, temizleme, sorgulama işlemi aşamasında; veri görselleştirme ise raporlama ve insan beyni için anlamlı sembollere çevirme aşamasında yardımcı olmaktadır.

2.6. Veri Madenciliği Uygulama Alanları

Veri madenciliği her geçen gün yeni ve farklı alanlarda kullanılmaya başlanmıştır. Günümüzde yaygın olarak kullanıldığı alanlar birkaç başlık altında toplanabilir.

Pazarlama

- Müşterilerin satın alma örüntülerinin tespitinde,
- Kampanya ürünlerini belirlemede,
- Mevcut müşterilerin elde tutulması için geliştirilecek pazarlama stratejilerinin oluşturulmasında,
- Pazar sepeti analizinde,
- Çapraz satış analizleri,
- Müşteri değerlendirilmede,
- Müşteri ilişkileri yönetiminde,
- Çeşitli müşteri analizlerinde,
- Satış tahminlerinde,

Banka ve Sigortacılık

- Farklı finansal göstergeler arasındaki gizli korelasyonların bulunmasında,
- Kredi kartı ve sigorta dolandırıcılıklarının tespitinde,
- Kredi taleplerinin değerlendirilmesinde,
- Kredi kartı harcamalarına göre müşteri profili belirlenmesinde,
- Yeni Poliçe talep edecek müşterilerin tahmininde,
- Risk yönetimi konusunda,
- Riskli müşteri tipinin belirlenmesinde.

Borsa

- Hisse senedi fiyat tahmininde,
- Genel piyasa analizlerinde,
- Alım-satım stratejilerinin uygunluğunda.

Telekomünikasyon

- Kalite ve iyileştirme analizlerinde,
- Hisse tespitlerinde,
- Hatların yoğunluk tahminlerinde,

Sağlık ve İlaç

- Test sonuçlarının tahmininde,
- Ürün geliştirmede,
- Tıbbi teşhiste,
- Tedavi sürecinin belirlenmesinde ,
- Yeni ilaç türlerini keşfi ve sınıflandırılması,

Endüstri

- Kalite kontrol analizlerinde
- Lojistik uygulamalarda,
- Üretim süreçlerinin uygunluğunda,

Bilim ve Mühendislik

- Ampirik veriler üzerinde modeller kurarak bilimsel ve teknik problemlerin çözümlenmesinde,
- Yeni virüs türlerinin keşfi ve sınıflandırılmasında,
- Gen haritasının analizi ve genetik hastalıkların tespitinde,
- Kansersiz hücrelerin tespitinde,
- Gezegen yüzey şekillerinin, gezegen yerleşimlerinin ve yeni galaksilerin keşfinde,

2.7. Veri Madenciliğinde Karşılaşılan Problemler

Küçük veri kümelerinde hızlı ve doğru bir biçimde çalışan bir sistem çok büyük veritabanlarına uygulandığında sorun çıkabilir. Bir VM sistemi ayıklanmış veri üzerinde mükemmel çalışırken, aynı veriye gürültü eklendiğinde net olmayan sonuçlar oluşabilir. İzleyen kısımda günümüz VM sistemlerinin karşılaştığı problemler ele alınacaktır.

2.7.1. Veritabanı boyutu

VM sistemlerinin karşı karşıya olduğu en önemli sorunlardan biri veritabanı boyutunun çok büyük olmasıdır. Küçük test verilerini ele alabilecek bir biçimde geliştirilmiş bir algoritmanın, yüz binlerce kat büyük test verilerini kullanabilmesi azami dikkat gerektirmektedir. Dolayısıyla VM yöntemleri ya sezgisel bir yaklaşımla arama uzayını taramalıdır ya da test verileri en aza indirilmelidir [22].

Belirli bir niteliğin alan değerleri önceden sıradüzensel olarak kategorize edilir. Sonrasında ise, ilgili niteliğin değerleri aşağıdan yukarıya doğru seviye seviye güncellenir. Yani tekrarlı çokluklar çıkarılır. Oldukça sağlam bir test verisi kuramı kullanılarak çok büyük boyutlu veri öyle bir boyuta indirgenir ki, hem kaynak veri belirli bir güven aralığında temsil edilir hem de indirgenen veri kümesinin boyutu kullanılan algoritma tarafından işlenebilir hale gelir. Son aşamada ise sürekli değerlerin belirli aralık değerlerine dönüştürülmesi ile tekrarlılık gösteren çokluklar ortadan kaldırılır.

2.7.2. Gürültülü veri

Büyük veri tabanlarında pek çok alanın içerdiği değer yanlış olabilir. Veri girişi ya da veri toplanması sırasında oluşan sistem dışı hatalara gürültü adı verilir. Hatalı veri gerçek dünya veritabanlarında ciddi problemler oluşturabilir. Bu durum, bir VM yönteminin kullanılan veri kümesinde bulunan gürültülü verilere karşı daha az duyarlı olmasını gerektirir. Eğer veri kümesi gürültülü ise, sistem bozuk veriyi tanımalı ve ihmal etmelidir. Quinlan, 1986'da gürültünün sınıflama üzerindeki

etkisini arařtırmak için bir dizi deney yapmıřtır. Deneysel sonuçlar etiketli öğrenmede etiket üzerindeki gürültü öğrenme algoritmasının performansını doğrudan etkileyerek düşmesine sebep olmuřtur. Buna karřın eğitim kümesindeki nesnelere nitelikleri üzerindeki en çok %10'luk gürültü miktarı ayıklanabilmektedir [22].

2.7.3 Boř deęerler

Boř deęer tanımı gereęi kendisi de dahil olmak üzere hiçbir deęere eřit olmayan deęerdir. Bir çoklukta eęer bir nitelik deęeri boř ise o nitelik bilinmeyen ve uygulanamaz bir deęere sahiptir. Bu durum iliřkisel veritabanında sıkça karřımıza çıkmaktadır. Mevcut boř deęer taşıyan veri için herhangi bir çözüm sunmayan yaklaşımın dıřında bilinmeyen deęer üzerinde çalışmalar yapılmıřtır. Boř deęerli nitelikler veri kümesinde bulunuyorsa ya bu çoklular tamamıyla ihmal edilmeli ya da bu çoklularda nitelięe olası en yakın deęer atanmalıdır.

2.7.4. Eksik veri

Her nesnenin ayrıntılı bir biçimde tanımlandığı ve bu nesnelere alabileceęi deęerler kümesinin belirtili olduęu durumlarda her bir nesnenin tanımı kesin ve yeterli olsaydı, sınıflama iřlemi basitçe nesnelere alt kümelerinden faydalanılarak yapılabilirdi. Bununla birlikte veriler kurum ihtiyaçları göz önünde bulundurularak düzenlenip toplandıęında mevcut veri, gerçek hayatı yeterince yansıtmayabilir. Bu gibi kořullarda bilgi keřfi modeli belirli bir güvenlik derecesinde tahmini kararlar alabilmelidir [17].

2.7.5. Artık veri

Kullanılan veri kümesi eldeki probleme uygun olmayan veya iře yaramayan nitelikler içerebilir. Artık nitelikleri elemek için geliştirilmiř algoritmalar, özellik seçimi olarak adlandırılır.

Özellik seçimi, tümevarıma dayalı öğrenmede budama öncesi yapılan bir iřlemdir. Bařka bir deyiřle özellik seçimi, verilen bir iliřkinin içsel tanımını, dıřsal tanımın

taşıdığı bilgiyi bozmadan onu eldeki niteliklerden daha az sayıdaki niteliklerle ifade edebilmektir. Özellik seçimi yalnızca arama uzayını küçültmekle kalmayıp, sınıflama işleminin kalitesini de artırır.

2.7.6. Dinamik veri

Kurumsal çevrimiçi veritabanları dinamiktir, yani içeriği sürekli olarak değişir. Bu durum, bilgi keşfi metotları için önemli sakıncalar doğurmaktadır. İlk olarak sadece okuma yapan ve uzun süre çalışan bilgi keşfi metodu bir veritabanı uygulaması olarak mevcut veritabanı ile birlikte çalıştırıldığında mevcut uygulamanın da performansı ciddi ölçüde düşer. Diğer bir sakıncası ise veritabanında bulunan verilerin kalıcı olduğu varsayıp, çevrim-dışı veri üzerinde bilgi keşif metodu çalıştırıldığında, değişen verinin elde edilen örüntülere yansımaları gerekmektedir. Bu işlem, bilgi keşfi metodunun ürettiği örüntüleri zaman içinde değişen veriye göre sadece ilgili örüntüleri güncelleme yeteneğine sahip olmasını gerektirir.

2.7.7. Farklı tipteki verileri ele alma

Gerçek hayattaki uygulamalar makine öğrenmesinde olduğu gibi, yalnızca sembolik veya kategorik veri türleri değil aynı zamanda tamsayı, kesirli sayı, çoklu ortam verisi, coğrafi bilgi içeren veri gibi farklı tipteki veriler üzerinde işlem yapılmasını gerektirir. Kullanılan verinin saklandığı ortam düz bir kütük veya ilişkisel veritabanlarında yer alan tablolar olabileceği gibi nesneye yönelik veritabanları, çoklu ortam veritabanları, coğrafi veritabanları vs. olabilir. Bununla birlikte veri çeşitliliğinin fazla olması bir VM algoritmasının tüm veri tiplerini ele alabilmesini olanaksızlaştırmaktadır. Bu yüzden veri tipine özgü, VM algoritmaları geliştirilmektedir.

2.8. Veri madenciliği modelleri

Veri Madenciliğinde kullanılan modelleri tahmin edici (Predictive) ve tanımlayıcı (Descriptive) olmak üzere iki ana başlık altında toplayabiliriz.

Tahmin edici modellerde; sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesine çalışılmaktadır [23]. Örneğin bir sınıftaki öğrencilerin bir dersle ilgili almış oldukları vize ve ödev notları gibi veriler bir veritabanında toplanabilir. Bu verilere uygun olarak kurulan model öğrencilerin o dersin sonunda finalden alacağı notun tahmininde kullanılmaktadır.

Tanımlayıcı modellerde ise, karar vermeye rehberlik etmede kullanılacak mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır. 25 yaş altı bekar kişiler ile, 25 yaş üstü evli kişiler üzerinde yapılan ve ödeme performanslarını gösteren bir analiz tanımlayıcı modellere örnek olarak verilebilir [23].

Gerek tanımlayıcı gerekse tahmin edici modellerde yoğun olarak kullanılan belli başlı teknikler; Sınıflama ve Regresyon, Kümeleme, Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler, Bellek Tabanlı Yöntemler, Yapay sinir ağları ve karar ağaçları olarak sıralanabilir. Sınıflama ve regresyon modelleri tahmin edici, kümeleme, birliktelik kuralları ve ardışık zamanlı örüntü modelleri tanımlayıcı modellerdir.

2.8.1. Sınıflama ve regresyon

Sınıflama ve regresyon, eldeki mevcut verilerden hareket edilerek geleceğin tahmin edilmesinde kullanılan veri madenciliği tekniğidir. Sınıflama gruplanacak verileri tahmin ederken, regresyon süreklilik gösteren değerlerin tahmin edilmesinde kullanılır. Sınıflama ve regresyon modellerinde kullanılan başlıca teknikler,

- Genetik Algoritmalar (Genetic Algorithms),
- K-En Yakın Komşu (K-Nearest Neighbor),
- Naïve-Bayes,
- Çoklu Regresyon, Lojistik Regresyondur (Logistic Regression),
- Faktör ve Ayırma analizleri (factor and discriminant) olarak verilebilir.

2.8.2. Kümeleme tekniđi

Kümeleme tekniđinde amaç üyelerinin birbirlerine çok benzediđi, ancak özellikleri birbirlerinden çok farklı olan kümelerin bulunması ve veri tabanındaki kayıtların bu farklı kümelere bölünmesidir. Kümeleme analizinde; veri tabanındaki kayıtların hangi kümelere ayrılacağı veya kümelemenin hangi deđişken özelliklerine göre yapılacağı, konunun uzmanı olan bir kiři tarafından belirtilebileceđi gibi veri tabanındaki kayıtların hangi kümelere ayrılacağını geliştirilen yazılımlar da yapabilmektedir. Kümeleme; veri madenciliđi, istatistik, biyoloji ve makine öğrenmesi gibi pek çok alanda kullanılır. Kümeleme tekniđinde, sınıflama tekniđinde olan veri sınıfları yoktur. Sınıflama tekniđinde, verilerin sınıfları bilinmekte ve yeni bir veri geldiđinde bu verinin hangi sınıftan olabileceđi tahmin edilmektedir [17]. Oysa kümeleme modelinde, sınıfları bulunmayan veriler gruplara ayrılırlar.

Biyolojide bitki ve hayvan sınıflandırmaları ve işlevlerine göre benzer genlerin sınıflandırılması, marketlerde farklı müşteri gruplarının alışveriş örüntülerinin ortaya konması gibi uygulamalar tipik kümeleme uygulamalarıdır [17].

Literatürde pek çok kümeleme algoritması bulunmaktadır. Kullanılacak olan kümeleme algoritmasının seçimi, veri tipine ve amaca bađlıdır. Başlıca kümeleme yöntemleri şu şekilde sınıflandırılmaktadır [17]:

- 1- Bölümleme Yöntemleri
- 2- Hiyerarşik Yöntemler
- 3- Yođunluk tabanlı yöntemler
- 4- Izgara tabanlı yöntemler
- 5- Model tabanlı yöntemler

2.8.2.1. Bölümleme yöntemleri

Bölümleme metotları, n adet nesneden oluşan veritabanını, giriş parametresi olarak belirlenen k adet bölüme ($k \leq n$) ayırma temeline dayanır. Veritabanındaki her bir eleman bir farklılık fonksiyonuna göre k adet bölümden birine dahil edilir [24]. Bu

bölümlerden her biri bir küme olarak adlandırılır. Bölümleme metotları k-means, k-medoids ve CLARA-CLARANS olarak bilinen algoritmaları kullanır.

2.8.2.1.1. K-Means algoritması

Bu algoritmaya K-means adı verilmesinin nedeni k adet kümenin her birini, kümeyi oluşturan elemanların ağırlıklı ortalaması ile temsil etmesidir [25]. Kümenin ağırlıklı ortalama değerine sahip olan ya da bu değere en yakın olan nokta küme merkezi olarak adlandırılır.

K-means algoritması öncelikle k adet rastgele nokta belirler. Bu noktalar, ilk küme merkezleridir. Bundan sonraki her eleman merkez noktaya olan yakınlık derecesine göre k adet kümeden birine dahil edilir. Her bir kümeye yeni eleman eklendiğinde küme elemanlarının ağırlıklı ortalaması tekrar hesaplanarak yeni bir küme merkezi bulunur ve bundan sonraki seçim işlemlerinde kümeyi bu yeni nokta temsil eder. Kümeleme işlemi tüm noktalar bitene kadar bu şekilde devam eder [1].

2.8.2.1.2. K-Medoids algoritması

K-medoids algoritması kümeyi temsil edecek noktayı bulmak için küme elemanlarının ortalamasını almak yerine kümenin en merkez noktasındaki elemanı yeni küme merkezi olarak alır. Böylece istisna verilerin küme merkezini kenarlara doğru kaydırması problemi giderilmiş olur [1].

2.8.2.1.3. Clara-Clarans algoritmaları

CLARA, veritabanının tümünü almak yerine küçük bir örnekleme kümesini temsilcisi olarak alıp örnekleme üzerinde uygular. Veritabanında birden çok örnekleme seçerek en iyi sonuç veren örneklemeden elde ettiği sonucu çıktı olarak verir [1].

2.8.2.2. Hiyerarşik yöntemler

Hiyerarşik yöntemler nesnelere ağaç yapısı şeklinde gruplandırma temeline dayanır. Hiyerarşik yöntemler giriş parametresi olarak bulunacak küme sayısını belirten k değerine ihtiyaç duymazlar, fakat ağaç yapısı oluşturma işlemini ne zaman durdurulacağını belirten eşik değeri parametresine ihtiyaç duyarlar.

2.8.2.3. Yoğunluk tabanlı yöntemler

Yoğunluk tabanlı yöntemler, nesnelere doğal dağılımını bir yoğunluk fonksiyonu aracılığı ile tespit ederek bir eşik yoğunluğunu aşan bölgeleri küme olarak adlandırır. Yoğunluk tabanlı algoritmalar düzgün şekilli olmayan kümeleri bulma başarısı, gürültü ve istisnalardan etkilenmemeye ve tek tarama ile sonuca ulaşma avantajları ile en başarılı kümeleme yöntemleri arasındadır [1].

2.8.2.4. Izgara tabanlı yöntemler

Izgara tabanlı yöntemler veri uzayını incelemek için sonlu sayıda kare şeklinde hücrelerden oluşan ızgara yapıları kullanırlar. Kullandıkları ızgara yapısından dolayı veritabanındaki nesne sayısından bağımsızdırlar. Performanslarını etkileyen tek unsur kullandıkları kare sayısıdır, kare sayısı arttıkça hesaplama zamanı artacağından performans düşer. Izgara tabanlı yöntemlerin en önemli avantajları işlem yükü az olduğu için hızlı ve çabuk sonuca ulaşabilmeleridir [17].

2.8.2.5. Model tabanlı yöntemler

Model tabanlı yöntemler eldeki verileri bir matematiksel model ile ifade etmeye çalışırlar. Bu yöntem verilerin belirli bazı olasılık teorilerinin karışımından oluşan bir mantık ile veri uzayına yerleştiklerini farz ederler.

2.8.3. Birliktelik kuralları ve ardışık zamanlı örüntüler

Alışveriş sırasında veya birbirini izleyen alışverişlerde müşterinin hangi mal veya hizmetleri satın almaya eğilimli olduğunun belirlenmesi, müşteriye daha fazla ürünün satılmasını sağlama yollarından biridir. Satın alma eğilimlerinin tanımlanmasını sağlayan birliktelik kuralları ve ardışık zamanlı örüntüler, pazarlama amaçlı olarak pazar sepeti analizi (Market Basket Analysis) adı altında veri madenciliğinde yaygın olarak kullanılmaktadır. Bununla birlikte bu teknikler, tıp, finans ve farklı olayların birbirleri ile ilişkili olduğunun belirlenmesi sonucunda değerli bilgi kazanımının söz konusu olduğu ortamlarda da önem taşımaktadır. birliktelik kuralları ve ardışık zamanlı örüntüler 3. bölümde ayrıntılı olarak ele alınacaktır [1].

2.8.4. Bellek tabanlı yöntemler

Bellek tabanlı veya örnek tabanlı bu yöntemler istatistikte 1950’li yıllarda önerilmiş olmasına rağmen, o yıllarda gerektirdiği hesaplama ve bellek boyutları yüzünden kullanılamamış ama günümüzde bilgisayarların ucuzlaması ve kapasitelerinin artmasıyla, özellikle de çok işlemcili sistemlerin yaygınlaşmasıyla, kullanılabilir olmuştur. Bu yönteme en iyi örnek en yakın k komşu algoritmasıdır.

2.8.5. Yapay sinir ağları

1980’lerden sonra yaygınlaşan yapay sinir ağlarında (artificial neural networks) amaç fonksiyon birbirine bağlı basit işlemci ünitelerinden oluşan bir ağ üzerine dağıtılmıştır. Yapay sinir ağlarında kullanılan öğrenme algoritmaları veriden üniteler arasındaki bağlantı ağırlıklarını hesaplar. YSA istatistiksel yöntemler gibi veri hakkında parametrik bir model varsaymaz yani uygulama alanı daha geniştir, ve bellek tabanlı yöntemler kadar yüksek işlem ve bellek gerektirmez [26].

2.8.6. Karar ağaçları

İstatistiksel yöntemlerde veya yapay sinir ağlarında veriden bir fonksiyon öğrenildikten sonra bu fonksiyonun insanlar tarafından anlaşılabilir bir kural olarak yorumlanması zordur. Karar ağaçları ise veriden oluşturulduktan sonra ağaç kökten yaprığa doğru inilerek kurallar (IF-THEN rules) yazılabilir. Bu şekilde kural çıkarma (rule extraction), veri madenciliği çalışmasının sonucunun doğrulanmasını sağlar. Bu kurallar uygulama konusunda uzman bir kişiye gösterilerek sonucun anlamlı olup olmadığı denetlenebilir. Sonradan başka bir teknik kullanılacak bile olsa karar ağacı ile önce bir kısa çalışma yapmak, önemli değişkenler ve yaklaşık kurallar konusunda analizciye bilgi verir ve daha sonraki analizler için yol gösterici olabilir.

BÖLÜM 3. APRIORİ ve DBSCAN ALGORİTMALARI

3.1 Giriş

Veri tabanlarından birliktelik kurallarının bulunması, veri madenciliğinin en önemli konularından biridir. Birliktelik kuralları madenciliği, veri tabanlarında bir arada sık görülen ilişkileri ortaya çıkaran ve özetleyen kuralları bulmaktır. Birliktelik kuralları madenciliği; sepet analizi, çapraz pazarlama, yerleşim düzeni tasarımı, katalog tasarımı, promosyon analizleri gibi alanlarda uygulanır. Birliktelik kuralları madenciliğinde kullanılan birçok algoritma vardır. Bu çalışmada amaç; internet tabanlı öğretimde sık girilen sayfa çiftlerinin ve en sık yanlış yapılan soru çiftlerinin birliktelik kurallarından, Apriori Algoritması kullanılarak bulunması amaçlanmıştır.

Kümeleme analizi ise bir veri kümesindeki yoğunlukları bulmak amacıyla kullanılır. Kümeleme analizine kısaca kümeleme denir. Kümeleme işleminde küme içindeki elemanların benzerliği fazla, kümeler arası benzerlik ise az olmalıdır. Kümeleme işleminde temel amaç, yeni gelecek ve henüz hangi sınıfta olduğu bilinmeyen verilerin varola sınıflardan en uygun olanına yerleştirilmesidir. Kümeleme, istatistik, biyoloji, uzaysal veri madenciliği ve makine öğrenmesi, örüntü tanıma ve resim tanıma alanlarında yaygın olarak kullanılmaktadır. Bu çalışmada kümeleme tekniklerinde DBSCAN algoritması kullanılarak quiz ve ödev notlarının yoğunluk olarak toplandığı gruplar tespit edilmiştir.

3.2. Birliktelik Kuralları

Bilişim uygulamalarının yaygınlaşması ile, bilgisayarda büyük miktarda veri depolanmaya başlanılmıştır. Günümüz veritabanı sistemleri, kullanıcıya depolanan bütün bilgilere kolayca ulaşabileceği araçları ve fonksiyonları sunmamaktadır. Büyük veri tabanlarında saklı olan bu bilgilere ulaşmak ve bu bilgileri kullanmak

üzere, otomatik bilgi keşfetmeye yarayan teknikler geliştirilmektedir. Bu tekniklerden biri olan birliktelik kuralları, depolanan verilerden, ilginç ve sıklıkla rastlanan bilgileri ortaya çıkarma işlevidir. Birliktelik kuralları, büyük veri kümeleri arasında birliktelik ilişkileri bulurlar. Birliktelik kuralları, nesnelerin bir arada olma durumlarını belirlemeyi amaçlar ve bir çok alanda geniş bir şekilde kullanılabilirler. Birliktelik kuralları bulma tekniği, yoğun nesne kümelerinin hesaplanması esasına dayanır ve büyük veri tabanlarında bunun hesaplanması oldukça zor bir işlemdir.

Birliktelik kurallarının kullanıldığı en tipik örnek, market sepeti uygulamasıdır. Bu işlem, müşterilerin yaptıkları alışverişlerdeki ürünler arasındaki birliktelikleri bularak müşterilerin satın alma alışkanlıklarını analiz eder. Bu tip birlikteliklerin keşfedilmesi, müşterilerin hangi ürünleri bir arada aldıkları bilgisini ortaya çıkarır ve market yöneticileri de bu bilgi ışığında daha etkili satış stratejileri geliştirebilirler [40]. Örneğin bir müşteri süt satın alıyorsa, aynı alışverişte sütün yanında ekmek alma olasılığı nedir? Bu tip bir bilgi ışığında rafları düzenleyen market yöneticileri ürünlerindeki satış oranını arttırabilirler. Örneğin bir marketin müşterilerinin süt ile birlikte ekmek satın alma oranı yüksekse, market yöneticileri süt ile ekmek raflarını yan yana koyarak ekmek satışlarını arttırabilirler.

Örneğin bir A ürününü satın alan müşteriler aynı zamanda B ürününü de satın alıyorsa, bu durum birliktelik kuralı ile aşağıdaki gibigösterilir [27, 28].

$A \Rightarrow B$ [destek = 2%, güven = 60%]

Buradaki destek ve güven ifadeleri, keşfedilen kuralların kullanışlılığını ve doğruluğunu gösterirler. Birliktelik kuralı için 2% destek değeri, analiz edilen tüm alışverişlerden 2%'sinde A ile B ürünlerinin birlikte satıldığını belirtir. [29] 60% oranındaki güven değeri ise A ürününü satın alan müşterilerin 60%'ının aynı alışverişte B ürününü de satın aldığını gösterir. Kullanıcı tarafından minimum destek eşik değeri ve minimum güven eşik değeri belirlenir ve bu değerleri aşan birliktelik kuralları ele alınır [17].

Büyük veritabanlarından birliktelik kuralları bulunurken, ilk olarak sık tekrarlanan öğeler bulunur. Bu öğelerin her biri en az, önceden belirlenen minimum destek sayısı kadar sık tekrarlanır. Daha sonra sık tekrarlanan öğelerden birliktelik kuralları oluşturulur. Bu kurallar minimum destek ve minimum güven değerlerini karşılamalıdır.

Birliktelik kuralları tekniği, veri içinde görülen özelliklerin kural olarak ortaya çıkarılması işlemidir. Genel olarak $A_1 \wedge A_2 \dots \wedge A_k \Rightarrow B_1 \wedge B_2 \dots \wedge B_l$ olarak gösterilir ve her kural bir destek (suport) ve güven (confidence) değeri ile ifade edilir.

3.2.1. Birliktelik kurallarının türleri

3.2.1.1. Hiyerarşik birliktelik kuralları

Uygulamaların çoğunda özellikler kümesi, bir sınıf hiyerarşi içinde verilir. Sınıflar hiyerarşik özellikler kümesine göre tek veya toplu görünüm yansıtırlar.

Örneğin verilen bir x destek ve y güven eşik değerleri için aşağıdaki gibi bir x-y geçerli kurallar bulunmaya çalışılır.

- Alkolsüz içecek -> meze destek değeri = %50, güven değeri = % 90 veya
- Alkollü içecek -> yer fıstığı destek değeri = %30, güven değeri = %90

Büyük destek değerleri, yüksek sınıf seviyesinde daha çok mevcuttur. Bu nedenle;

- Alkollü içecek -> meze

Kuralı bir x-y geçerli kuraldır.

Çoklu sınıflar, farklı görünümler yansıtacak şekilde eş zamanlı olarak gösterilirler. Örneğin şekil 3.1'de maden suları basit içecekler olarak düşünülen fakat eş zamanlı olarak diyet yiyecekler sınıfında da özel olan parçalardır. Mesela maden suları bir taraftan içecekler sınıfına ait iken diğer taraftan da eş zamanlı olarak diyet yiyecekler sınıfına aittir.

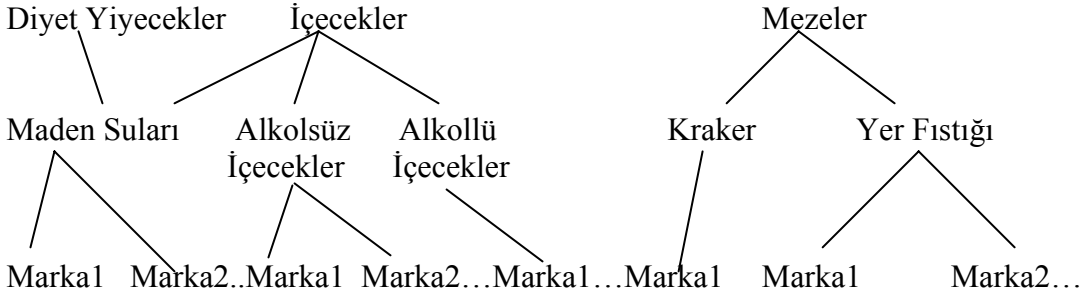
Olasılık mantığına göre sınıflandırma işlemi,

- İçecekler-> mezeler

Kuralı aynı zamanda

- (maden suları veya alkolsüz içecekler veya alkollü içecekler veya...) -> (yer fıstığı veya kraker veya...)

Olarak da okunabilir. Aynı zamanda bu, kesişimi ortaya koymayı kısıtlayan bir yoldur. Çünkü kesişen ifadeler birleşen ifadeler gibi kolayca bulunamazlar.



Şekil 3.1. Örnek bir sınıflandırma.

3.2.1.2. Çok seviyeli birliktelik kuralları

Bazı işlemsel veritabanları hiyerarşik yapıları içerirler ve bu hiyerarşinin farklı seviyelerine yayılan genelleştirilmiş ilginç birliktelik kuralları ortaya çıkar [48, 53].

Yüksek Seviyeli Kurallar

- “Yoğurt satın alan müşterilerin %80’i ekmek de satın alabilir” gibi kurallardır.

Düşük Seviyeli Kurallar

- “Eğer müşterilerin % 2’si yoğurt satın alırsa onların %70’i ekmek de alır”

şeklinde kurallardır.

Yüksek seviyeli kurallar, yüksek destek değerine sahip olabilir. Düşük seviyeli kurallar ise minimum desteğe sahip olmayabilir fakat daha bilgi verici olabilirler.

3.2.1.3. Negatif birliktelik kuralı

Geleneksel birliktelik kural madenciliği, aralarında sık ve yüksek korelasyon olan sıkıca bağlı kümeler arasındaki ilişkileri bulmaya çalışır. Birliktelik kuralı, bir uygulamada birlikte ortaya çıkan parçaların belirlenmesine yardımcı olurlar. Destek değeri-Güven değeri çatisından oluşan kurallar pozitif kurallardır. Bazı kümelerin varlığını, veri tabanının içinde başka kümelerinin de olabileceğini gösterir [48].

Uygulamalarda $A \rightarrow B$ şeklindeki bir kural “eğer A oluşursa B de genellikle oluşur” tahmininde kullanılır. Stok planlama, ürün yerleştirme, süpermarket gibi uygulamalarda bu kural, B’yi A’nın yanına yerleştirmek için kullanılabilir. Aşağıda bir market için birliktelik analizi önerilerine örnekler verilmiştir [48].

- Yüksek sınıfı ürünler veya ürün kümeleri satıcıya yakın yerleştirilebilir.
- A ve B ürünleri için $A \rightarrow B$ şeklinde bir kural var ise B, A’nın yanına yerleştirilebilir.

İlk örnek her gün sıkça satılan ürünlere satıcı tarafından kolay ulaşılabilmesini amaçlar ve çok etkili bir pazarlama imkanı sağlar. İkinci örnek ise, birlikte sıkça satın alınan bir grup parçanın seçilerek birlikte sergilenmesi nedeniyle satıcının zaman kaybını önlemektedir.

3.2.1.4. Nicel birliktelik kuralı

Birliktelik kuralı konusunda yapılan araştırmaların bir çoğu Boole birliktelik kuralları üzerinde yapılmıştır [48].

- “Eğer bir müşteri bir diş fırçası alırsa, o zaman diş macunu da alır”

Bu kuralın genel gösterimi

- “Diş fırçası \rightarrow Diş macunu”

Boole birliktelik kuralları bir işlemde satın alınan parçaların miktarı ile ilgilenmez.

3.2.2. Birliktelik kuralları algoritmaları

AIS: AIS algoritmasında aday kümeler veri tabanı taranırken anında bulunup sayılır. Bir işlem yapıldıktan sonra, bir önceki taramada sık olduğu görülen kümelerden hangilerinin bu işlemde olduğuna bakılır. Yeni aday kümeler bu sık rastlanan kümelerinin işlemdeki diğer kümelerle birleştirilmesiyle elde edilir [52, 51].

SETM: Bu algoritma sık kullanılan kümelerin hesaplanmasında SQL kullanılmasını amaçlar. AIS gibi bu algoritma da veri tabanından okunan işlemlere göre adayları anında oluşturur [49, 48].

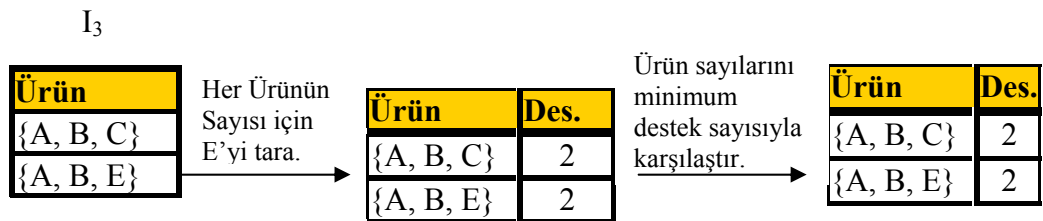
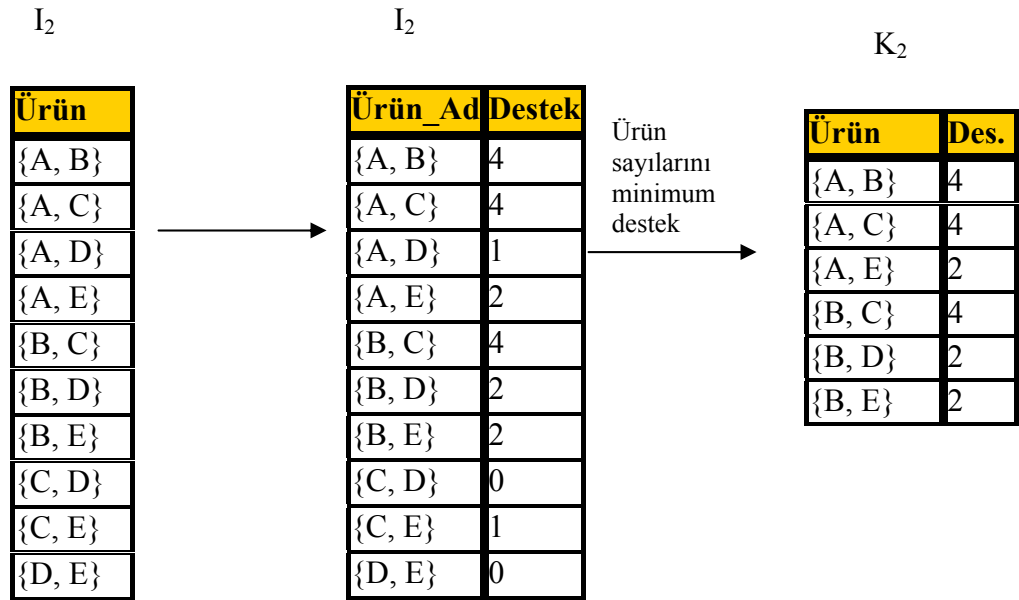
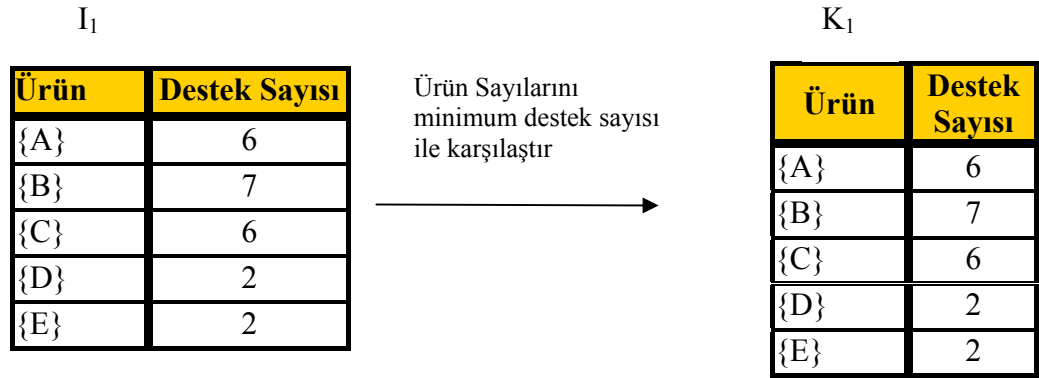
Apriori: Apriori algoritması yalnızca veri tabanının üstünden bir önceki geçişte elde edilen sık rastlanan kümeleri kullanarak veri tabanındaki işlemleri göz önüne almadan, sayılacak yeni aday kümeler oluşturur. Apriori algoritması, AIS ve SETM'e göre büyük veri kümeleri için daha iyi sonuç elde edebilmektedir. Sık tekrarlanan öğeleri bulmak için kullanılan en temel yöntem olan apriori aşağıda bir örnekle açıklanmaktadır [48].

Tablo 3.1. Marketten Yapılan Alışveriş Bilgilerini İçeren Veritabanı [29] .

| İşlem No | Ürün ad |
|----------|------------|
| I10 | A, B, E |
| I20 | B, D |
| I30 | B, C |
| I40 | A, B, D |
| I50 | A, C |
| I60 | B, C |
| I70 | A, C |
| I80 | A, B, C, E |
| I90 | A, B, C |

Tablo 3.1’de bir marketten yapılan alışverişlerin bilgilerini içeren E veritabanı görülmektedir. Bu veritabanında yapılan alışverişlerin numaraları İşlem_No sütununda görülmektedir. Her alışverişte satın alınan ürünler de Ürün_Ad sütununda görülmektedir. Apriori algoritmasını takip eden basamaklar şekil 3.2’de gösterilmektedir [49].

- 1- Algoritmanın ilk adımında, her ürün tek başına bulunduğu I_1 kümesinin elemanıdır. Algoritma, her ürünün sayısını bulmak için tüm alışverişleri tarar ve elde edilen sonuçlar şekil 3.2’de destek sayısı sütununda görülmektedir. Tablo 3.1’de görülebileceği gibi E’de A ürününden 6 adet, B ürününden 7 adet, C ürününden 6 adet, D ürününden 2 adet ve E ürününden de 2 adet görülmektedir.
- 2- Minimum alışveriş destek sayısının 2 olduğu varsayılırsa, tek başlarına sık tekrarlanan ürünler K_1 kümesinde görülmektedir. I_1 kümesindeki tüm ürünlerin destek sayısı, minimum destek eşik değeri olan 2’den fazla olduğu için I_1 ’in tüm ürünleri sık tekrarlanan ürün olarak değerlendirilir ve K_1 kümesine aktarılır.
- 3- Hangi ürünlerin ikili olarak sık tekrarlandığını belirlemek için K_1 kümesindeki ürünlerin ikili kombinasyonları bulunarak I_2 kümesi oluşturulur.
- 4- I_2 kümesindeki ürünlerin destek sayılarını bulmak için E taranır ve bulunan değerler destek sayısı sütununda belirtilir.
- 5- I_2 kümesindeki ürünlerden minimum destek eşik değerini aşan ürünler K_2 kümesine aktarılır.
- 6- Hangi ürünlerin üçlü olarak sık tekrarlandığını belirlemek için K_2 kümesindeki ürünlerin üçlü kombinasyonları bulunarak I_3 kümesi oluşturulur. Bu durumda $I_3 = \{\{A,B,C\}, \{A,B,E\}, \{A,C,E\}, \{B,C,D\}, \{B,C,E\}\}$ olması beklenir. Ancak apriori algoritmasına göre, sık tekrarlanan öğelerin alt kümelerinin de sık tekrarlanan öğe olması gerekmektedir. Buna göre yukarıdaki I_3 kümesindeki bazı elemanlar sık tekrarlanan olmadığı için, yeni I_3 kümesi $I_3 = \{\{A,B,C\}, \{A,B,E\}\}$ olur [39].



Şekil 3.2. Apriori Algoritmasının Gösterimi [17].

- 7- I_3 kümesindeki ürünlerin destek sayılarını bulmak amacıyla E taranır ve bulunan değerler destek sayısı sütununda belirtilir.
- 8- I_3 kümesindeki ürünlerden minimum destek eşik değerini aşan ürünler K_3 kümesine aktarılır.

9- Hangi ürünlerin dörtlü olarak sık tekrarlandığını belirlemek için K_3 kümesindeki ürünlerin dörtlü tek kombinasyonu {A, B, C, E} olarak belirlenir. Ancak bu kümenin alt kümelerinin tamamı sık tekrarlanan öğe olmadığı için I_4 kümesi boş küme olur ve apriori tüm sık tekrarlanan öğeleri bulunarak sonlanmış olur.

Sık tekrarlanan öğeleri bulduktan sonra, sıra birliktelik kurallarını oluşturmaya gelir. Birliktelik kurallarını oluştururken güven ve destek değerlerine ihtiyaç duyulur. Bu değerler kullanıcı tarafından girilir.

- $X \rightarrow Y$ Kural yapısı: Satın alınır(x, "SQLServer") \wedge Satın alınır(x, "Book")
Satın alınır(x, "DBMiner") SQLServer ve bir kitap alındığında büyük oranda DBMiner'da alındığını gösteren bir kuraldır [47].
- Destek (Support) ve Güven (confidence) değerleri:

Destek : İşlemlere başlamadan kuralların geçerliliğini belirlemek üzere kullanıcının belirlediği bir değerdir. $X \rightarrow Y$ X 'in ve Y 'nin bir arada tüm veritabanı içinde bulunma olasılığıdır.

$$(X \text{ sayısı} + Y \text{ sayısı}) / \text{Toplam İşlem sayısı}$$

Güven : Şarta bağlı olasılık olarak geçer. $X \rightarrow Y$ X ve Y nin bir arada X değerini içeren işlemlerin içinde bulunma olasılığıdır.

$$(X \text{ sayısı} + Y \text{ sayısı}) / X \text{'i içeren işlem sayısı}$$

Bu durumda Tablo 3.1'deki veri tabanına bakılarak şu birliktelik kuralları çıkartılabilir.

$$A \wedge B \Rightarrow E, \quad \text{güven} = 2/4 = 50\%$$

$$A \wedge E \Rightarrow B, \quad \text{güven} = 2/2 = 100\%$$

$$B \wedge E \Rightarrow A, \quad \text{güven} = 2/2 = 100\%$$

$$A \Rightarrow B \wedge E, \quad \text{güven} = 2/6 = 33\%$$

$$B \Rightarrow A^E, \quad \text{güven} = 2/7 = 29\%$$

$$E \Rightarrow A^B, \quad \text{güven} = 2/2 = 100\%$$

Eğer kullanıcı tarafından güven eşik değeri %80 olarak belirlenmişse birliktelik olarak ikinci, üçüncü ve altıncı kurallar dikkate alınır [17] .

3.2.2.1. Apriori algoritmasının türleri

AprioriTID ve AprioriHybrid algoritmaları apriori algoritmasının iki türüdür.

AprioriTID: Bu Algoritmanın en ilginç özelliği, ilk geçişten sonra sık tekrarlanan öge kümelerinin destek değerlerinin hesabı için E veritabanının kullanılmamasıdır. Onun yerine bir önceki geçişte kullanılan sık tekrarlanan öge kümelerinin kodlandığı bir küme kullanılır. Sonraki geçişlerde bu kodlanan kümenin boyutu veritabanından çok daha küçük hale gelebilir, böylece okuma gücünden kazanç sağlanır [44].

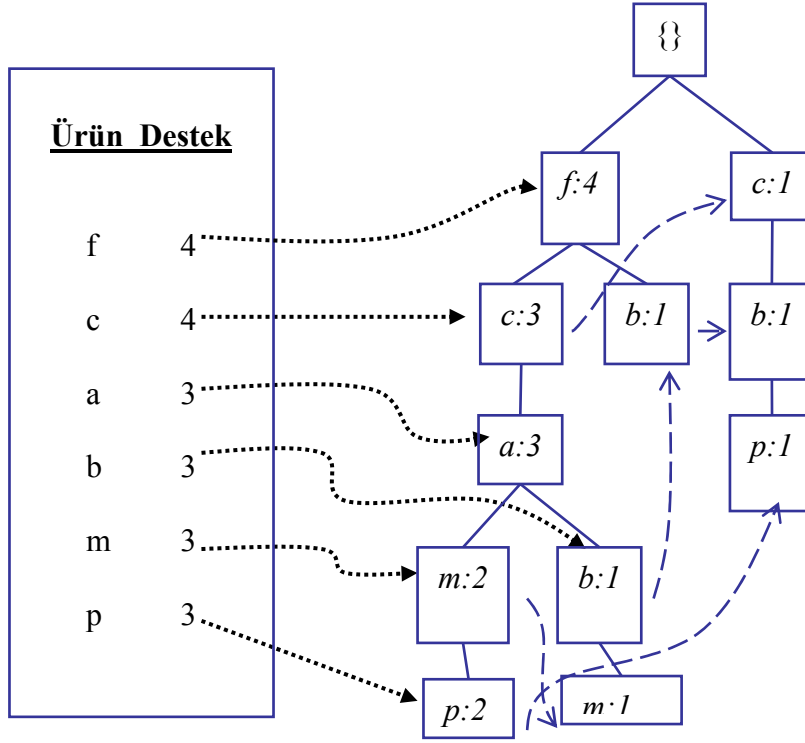
AprioriHybrid: Algoritmalarının çalışma zamanlarının performans analizine göre, önceki geçişlerde apriori daha iyi performans gösterir. Ama sonraki geçişlerde AprioriTID, aprioriyi daha iyi sonuç verir. Ayrıca her ikisi de aynı küme oluşturma prosedürünü kullanır. Bunları göz önünde bulundurarak AprioriHybrid algoritması, Apriori'nin ve AprioriTID'nin en iyi özelliklerini birleştirmeyi amaçlar [44].

DHP: Yine temel olarak Apriori algoritmasına bağlıdır. Ama geliştirilmiş bir biçimdir. Öncelikle ilk sık tekrarlanan öğeler kümesi oluşturulur. Bu küme üzerinden bir tablo belirlenir. Bundan sonra bütün adımlarda bu F-tablo temel alınır. Ve bir FP-ağaç yapısı çizilir [46].

Tablo 3.2. Veritabanı Örneği.

| İşlem_No | Satın Alınan Ürünler | Sık Tekrarlanan Öğeler |
|----------|--------------------------|------------------------|
| 100 | {f, a, c, d, g, i, m, p} | {f, c, a, m, p} |
| 200 | {a, b, c, f, l, m, o} | {f, c, a, b, m} |
| 300 | {b, f, h, j, o, w} | {f, b} |
| 400 | {b, c, k, s, p} | {c, b, p} |

Veritabanı bir kere taranır ve sık tekrarlanan tekli öge kümesi bulunur. Bu öge kümeleri azalan sırada sıralanır. Veritabanı tekrar taranır ve FP-ağaç yapısı oluşturulur [43].



Şekil 3.3. FP-Ağaç Yapısı [17].

3.3. DBSCAN Algoritması (Density Based Spatial Clustering of Applications with Noise)

Veri Madenciliği Modellerinden, kümeleme modelinin yoğunluk tabanlı yöntemleri içerisinde yer alan Dbscan algoritması, nesnelerin komşuları ile olan mesafelerini hesaplayarak belirli bir bölgede önceden belirlenmiş eşik değerden daha fazla nesne bulunan alanları gruplandırarak kümeleme işlemini gerçekleştirir. DBSCAN algoritması Veri madenciliğine birçok yeni terim ve yaklaşım getirmiştir.

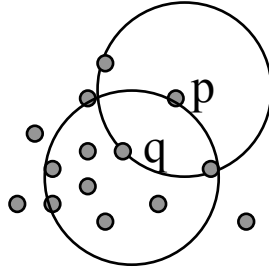
Dbscan algoritmasında kullanılan tanımlar:

Çekirdek nesne: Bir veri nesnesi önceden belirlenen bir eşik değerden daha çok nokta içeriyorsa bu nesne çekirdek nesnedir.

Eps: Bir veri nesnesinin komşularını belirlemek için gerekli olan yakınlık mesafesidir.

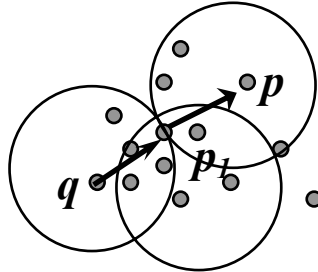
MinPts: Bir bölgenin yoğun olarak adlandırılabilmesi için Eps komşuluğunda bulunması gereken en az komşu sayısıdır.

Doğrudan Yoğunluk Erişilebilir Nokta : p noktası q nun Eps komşuluğunda ise ve q noktası p ye göre çekirdek nesne ise p noktası q'ya göre doğrudan yoğunluk erişilebilir noktadır. Şekil 3.4'de p ile q noktaları doğrudan yoğunluk erişilebilir noktalar.



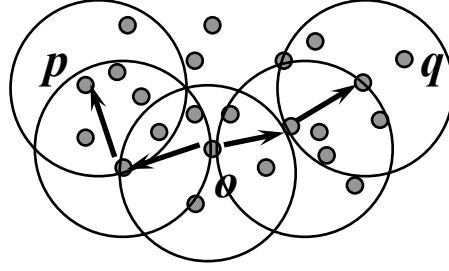
Şekil 3.4. Doğrudan yoğunluk erişilebilir noktalar [30].

Yoğunluk Erişilebilir Nokta : p_{i+1} noktası p ye göre doğrudan yoğunluk erişilebilir ise $p_1, p_2, p_3, \dots, p_n$ doğrudan yoğunluk erişilebilir noktalar olmak üzere $p_1=p$ ve $p_n=q$ ise q noktası Eps ve MinPts değerlerine göre p noktasına yoğunluk erişilebilirdir. Şekil 3.5'de p ile q noktaları yoğunluk erişilebilir noktalar.



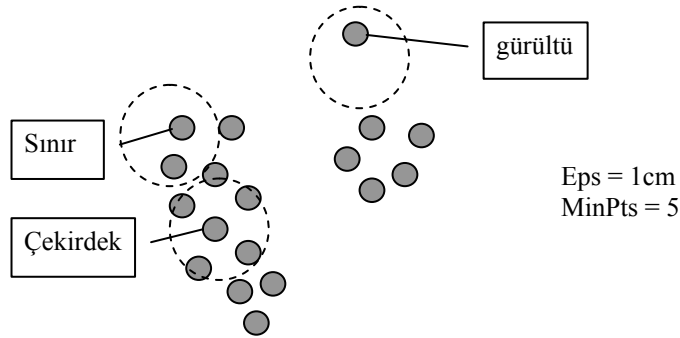
Şekil 3.5. Yoğunluk erişilebilir noktalar [51].

Yoğunluk bağlı noktalar : Herhangi p ve q noktalarının her ikisi de bir o noktaya yoğunluk erişilebilir durumda ise p ve q noktaları Eps ve MinPts değerine göre yoğunluk bağlı noktalar. Bu durum Şekil 3.6.'da görülmektedir.



Şekil 3.6. Yoğunluk bağlı noktalar [30].

DBSCAN algoritmasının çalışması için MinPts ve Eps parametreleri bildirilmelidir. Algoritma öncelikle rasgele bir p noktası seçer. p noktasına MinPts ve Eps değerlerine göre yoğunluk erişilebilir olan tüm noktaları bulur, eğer p çekirdek nokta koşulunu sağlıyor ise yeni bir küme keşfedilmiş olur. p noktasına yoğunluk erişilebilir olan tüm noktalara teker teker olarak aynı işlem uygulanır, eğer herhangi bir nokta çekirdek nokta koşulunu sağlamıyorsa bu nokta kümenin sınır noktasıdır. İncelenen tüm noktalardan hiçbiri çekirdek nokta koşulunu sağlamadığı zaman kümenin sınırları belirlenmiş olur. algoritma yeni bir rasgele nokta seçerek aynı işlemleri tekrar eder. Eğer rasgele seçilen nokta çekirdek nokta koşulunu sağlamıyorsa bu nokta gürültü ya da istisna olarak tanımlanır. Eps=1 ve MinPts=5 için DBSCAN algoritmasının çalışma yapısı Şekil 3.7’de görülmektedir.



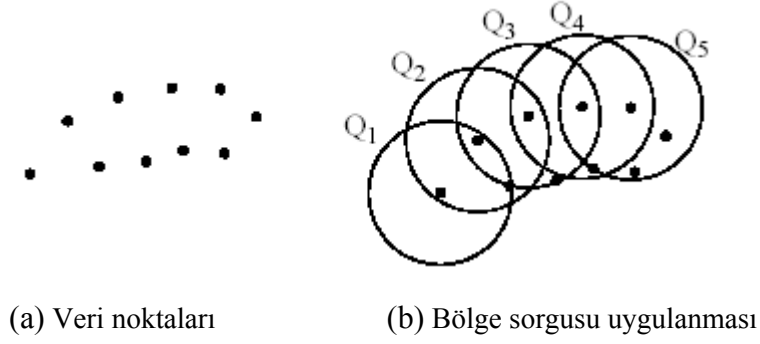
Şekil 3.7. DBSCAN algoritması çalışma yapısı [17].

3.3.1. DBSCAN algoritmasının çalışması

DBSCAN algoritması, veri noktalarının iki ya da çok boyutlu uzayda birbirleri ile olan komşuluklarını ortaya çıkarır [1]. DBSCAN algoritmasının amacı, Bazı bölgelerde kendiliğinden oluşan kümelenmeleri tespit etmek ve bu kümelenmelerden anlamlı sonuçlar çıkarmaktır. Bu tür kümelenmeler bazı noktaların ortak bir eğilimi hakkında önemli bilgiler verir. Örneğin deprem noktalarının kayıtlarının tutulduğu veritabanındaki kümelenmeler bir fay hattını belirtebilir [17].

DBSCAN algoritması için temel kavramlar olan, çekirdek nesne, Eps, Minpts, doğrudan yoğunluk erişebilir nokta, yoğunluk erişebilir nokta, yoğunluk bağlı nokta terimleri bölüm 3.2.' de açıklanmıştı. Öncelikle veritabanı, Eps ve MinPts değerlerini giriş parametresi olarak alır. Veritabanındaki herhangi bir nesneden başlayarak tüm nesnelere kontrol eder. Eğer kontrol edilen nesne daha önce bir kümeye dahil edilmişse işlem yapmadan diğer nesneye geçer. Eğer nesne daha önce kümelenmemiş ise, bir bölge sorgusu yaparak nesnenin Eps komşuluğundaki komşularını bulur. Komşu sayısı MinPts'den fazla ise, bu nesne ve komşularını yeni bir küme olarak adlandırır. Daha sonra, önceden kümelenmemiş her bir komşu için yeni bölge sorgusu yaparak yeni komşular bulur. Bölge sorgusu yapılan noktaların komşu sayıları MinPts'den fazla ise kümeye dahil eder. Bu işlem, incelenen noktaların komşu sayısı MinPts'den az çıkana kadar sürer. Bu durum kümenin sınırlarına gelindiğini işaret eder [17].

Bölge sorgusunun tekrarlı olarak uygulanması Şekil 3.8'de görülmektedir. Şekil 3.8 (a) bir küme olmak üzere Q1,Q2,Q3... bölge sorgularını temsil etmektedir. Rastgele seçilen ilk nokta Q1 sorgusunun merkezindeki nokta olsun. Örneğin MinPts=3 için bu nokta çevresinde Q1 bölge sorgusu gerçekleştirildiğinde iki yeni nokta bulunmuştur. Bu noktalar kümeye dahil edilmiştir. Noktalardan biri üzerinde Q2 sorgusu oluşturulmuş ve yeni bir nokta daha bulunmuştur. Bu şekilde bulunan her yeni nokta için Q3, Q4, Q5 sorguları oluşturularak kümenin tamamı bulunmuştur [1].

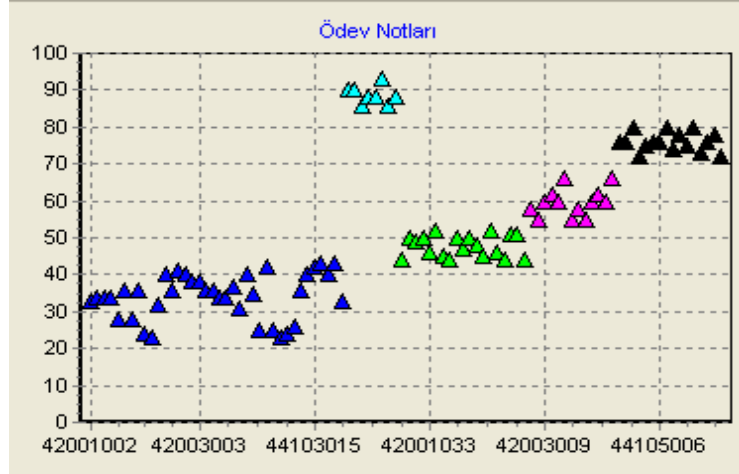


Şekil 3.8. DBSCAN Uygulaması [17].

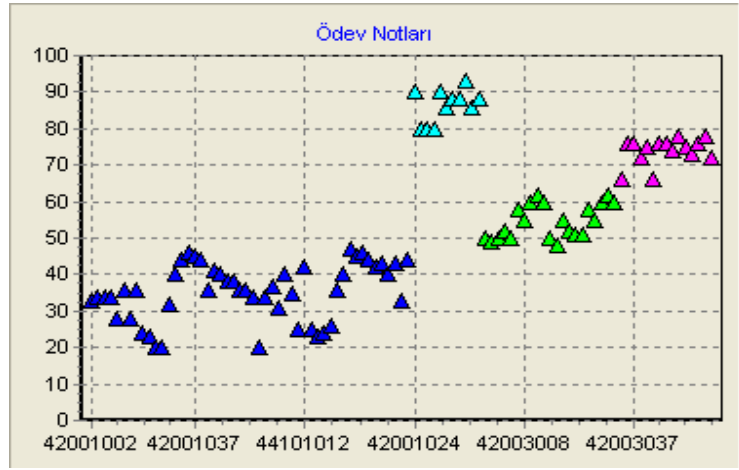
Bir kümenin tamamı bulunana kadar küme elemanları Şekil 3.8’de görüldüğü gibi bir yığılda (stack) tutulmaktadır. Herhangi bir anda ana bellekte veritabanının tamamı değil, yalnızca ilgili küme elemanlarının bulunduğu yığın tutulmaktadır. Bu durum, veritabanının tümünü ana bellekte tutarak işlem yapan kümeleme algoritmalarına göre bir miktar bellek tasarrufu sağlamaktadır [30].

DBSCAN algoritması Eps ve MinPts olmak üzere iki parametre aldığı için her iki parametrenin de kümeleme sonucuna etkisini görebilmek için farklı parametreler ile uygulanmıştır. DBSCAN algoritması veritabanının her elemanını bir kümeye dahil etmez, istisna verileri süzme yeteneğine sahiptir. Sonuç grafiklerinde algoritmanın gürültü (istisna) olarak belirlediği değerler gösterilmemiştir.

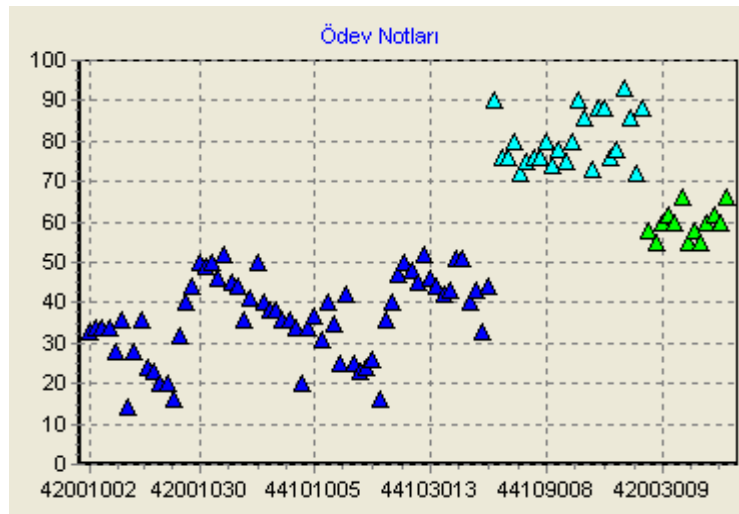
Bu tezde yapılan DBSCAN uygulamasında, Eps değeri arttıkça küme bulma hassasiyeti azalmakta, tüm veritabanını tek bir küme gibi görme eğilimi ağırlık kazanmaktadır. MinPts değeri arttıkça kümelerin hacmi daralmakta ve yalnızca yoğun iç bölgeleri tespit edilebilmekte, azaldıkça ise az yoğun olan gürültü alanları dahi küme olarak nitelendirilmektedir. Bu yüzden en uygun kümelenemeyi bulmak için farklı değerler ile işlem birkaç defa tekrar edilmelidir.



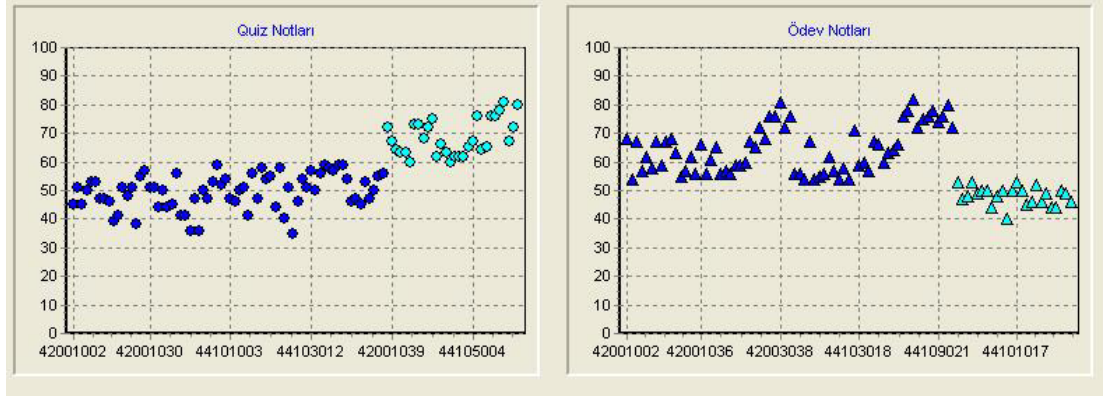
Şekil 3.9. Eps=5 ve MinPts=7 için DBSCAN algoritması sonucu.



Şekil 3.10. Eps=7 ve MinPts=5 için DBSCAN algoritması sonucu.



Şekil 3.11. Eps=10 ve MinPts=5 için DBSCAN algoritması sonucu.



Şekil 5.21. Not Yakınlığı=7 ve Minimum Öğrenci=10.

Bölüm 4. İnternet Tabanlı Öğretim

4.1. Giriş

“Eğitim ve Öğretim” insanoğlunun varoluşundan bugüne kadar değişerek varlığını sürdüren bir olgudur. İnsan okulda, işyerinde, günlük yaşamında çeşitli şekillerde “Eğitim ve Öğretim” ile karşı karşıya kalır. Eğitim, öğretime göre çok daha kapsamlı bir kavramdır. Bunun yanında öğretim kavramı, kişinin belli bir alanda sahip olduğu bilgi birikimini oluşturmasını sağlamak olarak tanımlanabilir. Bu tanımlamaya göre, insanın hayatının her aşamasında çeşitli durumlarla etkileşimde bulunması sonucu yaşadığı temel olgu öğretimdir [54].

Son yıllarda her alanda köklü değişimlere sahip olan teknolojik gelişmeler, öğretim alanında da etkisini göstermektedir. Bilgisayar teknolojisinin öğretim alanında kullanılmaya başlamasıyla beraber, klasik öğretim şekil değiştirmiştir. Bu iki şekilde incelenebilir. Birinci şekilde; bilgisayar öğretim alanında yardımcı araç olarak kullanılabilir. Örneğin, programlama dillerinin öğretilmesinde, bilgisayar bir uygulama aracı olarak kullanılabilir. İkinci şekilde ise bilgisayar, öğretim ortamını oluşturmada kullanılan bir kaynak olarak karşımıza çıkar [54].

Günümüzde öğretim ve bilgisayar teknolojilerinin bir araya getirilmesiyle, öğretim ortamını düzenlemede, klasik öğretim ortamlarından farklı etkin öğretim ortamlarının tasarımında bilgisayar teknolojileri kullanılmaya başlanmıştır.

4.2. İnternet Tabanlı Öğretim (İTÖ)

Öğretim sistemleri, gelişiminde çeşitli uygulama biçimleri bulmuş olmasına rağmen, İnternet’in gelişip güçlenmesi ile birlikte İnternet’e dayalı öğretim, öğretim sistemleri ile özdeşleşmiştir.

Internet'in kullanımı ve insan hayatındaki yeri sürekli olarak ve hızla gelişmektedir. Internet kişisel ve toplumsal iletişimin yanı sıra, bilgiye zaman ve mekandan bağımsız şekilde erişimi ve çok büyük boyutlardaki bilgiyi hızlı ve güvenilir bir şekilde elde etmeyi sağlamaktadır.

Internet üzerindeki bilgiler elektronik ortamda sayfalar halinde tutulur. Tıpkı bir kitap gibi her sayfa yazı, resim, çizim içerir. Bunlara ek olarak animasyonlar, sesli ve görüntülü anlatımlar, bilgi formları, veritabanı erişimi ve diğer sayfalara erişim için bağlantıları (links) içerebilir. Bu durumda Internet'in eğitim ve öğretim alanındaki gücü çok açık bir şekilde ortaya çıkmaktadır [31].

4.2.1. Internet tabanlı öğretimin faydaları

Günümüzde yüksek öğrenim gören öğrenci sayısı içerisinde, ağırlıklı yeri olmasa da Internet üzerinde öğrenim yapan ve öğrenme aktivitelerini bilgisayar karşısında gerçekleştiren öğrenci toplulukları bulunmaktadır. Bu toplulukları, öğrenme süreci içerisinde, geleneksel sınıf ortamında öğrenim yapandan farklı bir şekilde öğrenim deneyimleri geçirmektedir [54].

Internet Tabanlı Öğretim'in faydaları aşağıdaki gibi sıralanabilir:

- Birey bilgisayar ve internet aracılığıyla bilgiye ulaşır.
- Internet ortamından sanal sınıf adı verilen yapılarla öğrenciler gruplandırılır ve ders alabilir.
- Öğrenci diğer arkadaşlarıyla sesli ya da görüntülü iletişime geçebilir.
- Ders sonrası okul kantininde bir araya gelmek yerine sanal sohbet odalarında buluşabilirler.
- Ödevlerini elden teslim etmek yerine e-posta aracılığı ile öğretmenine ulaştırır.
- Ders tekrarı yapmak istediğinde masa başı yerine ekran başına geçer, online notlarını gözden geçirir.
- Evde öğrenim dikkate alınırsa ulaşım için geçirilecek zamandan ve harcanacak paradan tasarruf edilmiş olur.

- Öğrenciler bilgisayar bulunan ve internet erişimi olan her yerde öğrenim yapabilme olanağına sahiptir.
- Kendi hızında öğretim modülleri ile öğrencinin kendi öğrenme hızında ilerlemesi mümkün olmaktadır.
- Dersle ilgili tartışma forumlarına istediği anda erişme, izleme ve katılmanın yanı sıra canlı konuşma odaları ile diğer sınıf arkadaşlarına veya öğretmene erişme olanağı öğrencinin seçimine kalmıştır.
- Derslerini alması için günün belli bir saatinde hazır olması gerekmez dilerse gece yarısından sonra bile öğrenimini sürdürebilir.
- Yenilenen bilgiye hızla ulaşır, kitapların yeniden yazılmasını beklemesine gerek kalmaz.
- Basılı içeriklerde yer alan hareketsiz resimler, ekranda etkileşimli bir hale gelir. Birey güvenli bir ortamda deney yapabilir ve içerikle etkileşime girmeyi de öğrenebilir.
- Dersleri ile ilgili kendi istediği mekanda ve zamanda iletişime geçen bir öğrenci, belirli bir sınıf ortamında ve belirli saatlerde ders gören bir öğrenciye oranla daha rahattır.
- İnternet tabanlı öğretim çok farklı öğretim teknik ve araçlarını çeşitli yöntemleri bir araya getirmektedir.
- İnternet tabanlı öğretim ile geliştiren bilgisayar ve internet deneyimi, öğrencilere ileriki yaşantı ve iş çevrelerinde de yararlı olacaktır.
- Başarıyla tamamlanan bir internet tabanlı öğretim programı kişinin kendine olan güvenini arttıracak ve kendi öğreniminin sorumluluğunu üstlenmesine sebep olacaktır.
- Öğrenciler geliştirilmiş olan materyalleri test edeceklerdir.

4.2.2. İnternet tabanlı öğretim modeli

İnternet Tabanlı Öğretim Modeli'nde, "İnternet'e Dayalı Uzaktan Eğitim" adı altında kullanılan farklı tekniklerin hemen hemen tamamından yararlanılmaktadır. İçeriğe erişmek için HTML sayfa yapıları düzenlenmekte, iletişimin sağlanması ve sağlıklı olarak yürütülmesi için elektronik posta listelerinden faydalanılmakta, etkileşimin artırılabilmesi için tartışma listeleri ve sohbet programları kullanılmaktadır [40].

Internet Tabanlı Öğretimin ön plana çıkan zaman ve mekan avantajlarının yanında diğer önemli üstünlükleri de vardır. Bu üstünlükler sistemin tercih edilmesinde önemli roller oynayabilecek niteliktedir[24]. Bu nitelikler kişiselleştirilebilir eğitim, öğrenci merkezli eğitim, öğrenci yönetimli eğitim ve düşük maliyetli eğitim olarak sıralanabilir. Kişiselleştirilebilir eğitimde, verilen eğitimin şirket, bölüm, grup hatta kişiye göre özelleştirilmesi sağlanmaktadır. Öğrenciye verilecek eğitimin, öğreticinin kapasitesine göre değil, öğrencinin gereksinimlerine göre belirlenebilmesini öğrenci merkezli eğitim sağlamaktadır. Öğrenci yönetimli eğitimin getirisi ise öğrencinin çevrimiçi topluluklar oluşturmasına olanak tanıyarak, öğrencinin kendisi için içerik ya da program oluşturabileceği bir ortam hazırlayabilmesidir. Maliyet açısından bakıldığında Internet Tabanlı Eğitim'in maliyetinin geleneksel örgün eğitim maliyetinin ortalama yarısı kadar olduğu görülmektedir. Öte yandan ITÖ'nün etkileşimli eğitim özelliği ve güncel içerik sunması da göz ardı edilmemelidir[32].

ITÖ ile ilgili örnekleri incelediğimizde temel olarak iki farklı yaklaşım ortaya çıkmaktadır. Bu yaklaşımlar akademik programlar ve sertifika programları olarak ifade edilebilir. Üniversite ve özel sektör bünyesinde verilen çok çeşitli sertifika programları bulunmaktadır. Sayıca üstün olan bu sertifika programlarını lisans, ön lisans ve yüksek lisans programları izlerken, doktora programlarının sayısı diğerlerine göre bir hayli düşüktür.

Bilişimin hemen hemen her alanında olduğu gibi Internet Tabanlı Eğitim konusunda da ülkemizde içerik sıkıntısı bulunmaktadır. Bant genişliğinin yeni yeni belirli standartlara ulaşabildiği ülkemizde, ITÖ konusunda da yeterli ilerlemeler kaydedilememiştir. En büyük sıkıntılardan biri, içeriğin yetersiz oluşudur. Yurt dışında oluşturulan zengin içerik, dil sorunu ve müfredat farklılıkları yüzünden kullanılamamaktadır. Tüm bu olumsuz tabloya rağmen sayıları çok az da olsa ITÖ konusunda standartları yakalamış özel sektör kuruluşlarına ve üniversitelere rastlanmaktadır.

Türkiye’de Web sayfası aktif olarak çalışan toplam 79 üniversiteden sadece 5’i internet tabanlı uzaktan eğitim olarak tanımlayabileceğimiz bir programa sahiptir. Bu üniversiteler ve programları Tablo 4.1’de yer almaktadır.

Tablo 4.1. Türkiye’de ITÖ Programına Sahip Üniversiteler [24].

| Üniversite | Program Adı | URL |
|--|---|---|
| Ahmet Yesevi Üniv. | Türtep | http://www.yesevi.net |
| Anadolu Üniv. | E - MBA | http://emba.anadolu.edu.tr |
| Anadolu Üniv. Açık Öğretim Fakültesi | Bilgi Yönetimi Önlisans Programı | http://www.bilgi.aof.edu.tr |
| İTÜ | UZEM | http://www.uzem.itu.edu.tr |
| ODTÜ – IDEA | Asynchronous Internet Education | http://idea.metu.edu.tr |
| ODTÜ – Online | METU Online | http://online.metu.edu.tr |
| ODTÜ - Informatics | Informatics Online - Master of Science Program | http://ion.ii.metu.edu.tr |
| ODTÜ – DİL | diL (Distance Interactive Learning) | http://www.dil.metu.edu.tr |
| İstanbul Bilgi Üniv. | E - MBA | http://www.bilgiemba.net |

Üniversiteler haricinde özel sektör internet tabanlı uzaktan eğitim programlarından bazılarını Tablo 2’de yer verilmektedir.

Tablo 4.2. Türkiye’de ITÖ Hizmeti Veren Özel Sektör Faaliyetlerinden Bazıları.

| Kuruluş | Program Adı | URL |
|---------------------------|----------------------|---|
| IDEA e-learning solutions | Microsoft Eğitimleri | http://www.ideaegitim.com |
| Öğretmenler Sitesi | Teknoloji Eğitimleri | http://www.ogretmenlersitesi.com |
| Enocta | Mesleki Eğitim | http://www.meslekegitimleri.com |
| Netron | E-LearnCampus | http://www.netron.com.tr |

* 19 Mayıs 2004 tarihinde YÖK’ün web sitesinde yer alan liste dikkate alınmıştır.

4.2.3. İnternet tabanlı uzaktan eğitim sistemleri’nin sahip olması gereken özellikler

İnternet Tabanlı Öğretimin, basit eğitim içeriklerinden ayrılabilmesi için sahip olması gereken temel özellikler vardır. Bu özellikler eğitim sisteminin amacına ve hedef kitlesine göre kimi zaman değişiklikler gösterse de genel hatlarıyla aşağıdaki fonksiyonları içermelidir

- Kullanıcıların tanımlanması ve yönetilmesi: İnternet üzerinden yayın yapan Öğretim sistemleri genel erişime açık bir yapıya sahip olabilmektedir. Ancak eğitim içeriklerinin herkes tarafından görüntülenmesi istenmeyebilir. Belirli kullanıcı grup ve hakları doğrultusunda sisteme giriş yetkisi verilmek istendiği durumlarda, kullanıcı tanımlayabilir ve yönetebilir bir yapıda olması gerekmektedir [32].
- Ders içeriklerinin hazırlanması: ITÖ’nün temelini oluşturan ders içeriklerinin hazırlanması ya da hazırlanmış içeriklerin İnternet ortamına aktarılması sistem içerisinden yapılabilir. Hazır bir şablon kullanılabileceği gibi, içeriğin oluşturulmasında farklı programları da kullanmak mümkündür [32].
- Derslerin yönetilmesi: Öğrenci derslerinin kontrol edilmesi, hangi bölümü almaları gerektiği ya da hangi bölümde oldukları gibi bilgilerin takip edilebilmesi gerekmektedir. Tüm bu bilgiler ışığında öğrencinin belirli bir programı takip etmesi

ve bitirmesi sağlanabilir. Bu sayede sistem genelinde aktif olan derslerin kullanım yoğunluğu da takip edilmiş olmaktadır.

- Ödev ve proje verilmesi/teslimi: Öğrencilere ödev ve projelerin verilmesi, bu çalışmalar ile ilgili içerik ve açıklamaların öğrencilere aktarılması, tamamlanan çalışmaların toplanıp değerlendirilmesi gibi işlemlerin yapılabilmesi gerekmektedir. Tüm bu işlemlerin tek bir merkezden yapılması, sorumlu kişilerin üzerindeki iş yükünü azaltacağı gibi, sürece de hız kazandıracaktır [32].

- Sınav ve testlerin hazırlanması ve uygulanması: ITÖ uygulamalarında dönem içinde aktarılan bilginin öğrenci tarafından ne derecede alınabildiği ortaya konmalıdır. Bütün eğitim sistemlerinde olduğu gibi ITÖ’de de bu çalışma sınav ve testler yoluyla yapılmaktadır. Bu çalışmalarda iki farklı yöntem genel olarak tercih edilmektedir. Bunlardan biri dönem/eğitim sonunda öğrencilerin bir merkezde toplanarak sınava tabi tutulmalarıdır. Bu sistem farklı ülkelerden sisteme dahil olan kullanıcılar için uygun bir yöntem değildir. Bu durumda çevrimiçi sınavlar devreye girmektedir. Öğrenciler terminaller yardımıyla merkezden gelen soruları yanıtlamaktadırlar. İki yöntemin beraber kullanıldığı sistemler de mevcuttur. Her iki yöntemde de (ya da ikisini de uygulayan sistemlerde) eğitim süresince öğrencinin kendi bilgi düzeyini test etmesi gerekmektedir. Genel değerlendirmede kullanılacak testlerin yanı sıra, sadece deneme amaçlı olarak testlerin oluşturulabilmesi ve bu testlerin eğitim sistemi üzerinden öğrenciye sunulabilmesi de gerekmektedir [33].

- Öğrenci davranışlarının izlenmesi ve incelenmesi: ITÖ’leri başarıya taşıyacak en önemli çalışmalardan biri de sistemin ne derece etkin kullanıldığının gözlenebilmesidir. Bunun yolu kullanıcıların sistem içerisinde davranışlarının izlenebilmesinden geçer. Öğrencilerin günün hangi saatinde sistemden ne ölçüde yararlandıkları, hangi ders içeriklerinde ne kadar vakit geçirdikleri gibi bilgilerin sistem üzerinden takip edilebilmesi gerekmektedir. Elde edilen verilerin belirli istatistiksel bilgiler halinde sorumlu kişilere aktarılması yine sistemin sorumluluğunda olmalıdır.

- Öğrencilerin başarı durumlarının değerlendirilmesi: Eğitimin sonunda hem sistemin başarısını, hem de öğrencinin başarısını öğrenci başarı durum değerlendirmesi ortaya koyacaktır. Bu değerlendirme aynı zamanda, diploma, sertifikasyon ya da başarı belgesine öğrencinin hak sahibi olup olmadığını da belirleyecektir. Başarı durumlarının değerlendirilmesi eğitim programında daha

sonraki aşamalarda ön koşulun yerine getirilip getirilmediğinin de bir göstergesi olacaktır. Tüm bu çalışmalar sistemin sorumlulukları arasında yer almaktadır [33].

- Etkileşimli iletişim ortamlarının oluşturulması ve yönetilmesi: ITÖ'nün önemli avantajlarından birisi de birçok değişik Internet tabanlı iletişim sistemini kendi bünyesinde barındırıyor olmasıdır. Tartışma grupları, sohbet odaları, akışkan video ve ses aktarımı, Flash gibi kullanıcı etkileşimi sağlayabilecek ara yüz teknolojilerinden fayda sağlanması, sistemin sahip olması gereken özelliklerin başında gelmelidir [34].

4.3. Internet Tabanlı Öğretimde Kullanılan Teknolojiler

Internet Tabanlı Öğretim Sistemleri üzerinde yapılan bir araştırma sonucunda, Internet Tabanlı Öğretim Sistemlerinde dört temel teknolojinin kullanıldığı ortaya konulmuştur :

- Uyarlanır İçerik Sunumu (Adaptive Content Presentation)
- Öğrenci Çözümlerinin İrdelenmesi (Examining Student Solutions)
- Etkileşimli Problem Çözme Desteği(Interactive Problem Solution Support)
- Uyarlanır Gezinme Desteği (Adaptive Navigation Support)

4.3.1. Uyarlanır içerik sunumu

Uyarlanır içerik sunumunun hedefi, bilgi anabirimindeki bilgiyi öğrenciye sunmadan önce, konu ön gereksinim denetimini yapmak ve sunarken, öğrencinin amaçlarına, bilgi düzeyine ve öğrenme metoduna uyarlanmış bir düzende sunmaktır. Başka bir deyişle hedef, bilgi öğelerinin öğrenciye en uygun şekilde sunulmasını ve eğitsel aktivitelerin en uygun zamanda işleme konmasını sağlamaktır [40].

Uyarlanır içerik sunumu teknolojisi sistemden sisteme farklılık gösterir. Kimi sistemlerde sayfaların sıradüzeni devingen yapıda tasarlanırken, kimi sistemlerde sayfaların içeriği devingen yapıda tasarlanır. Bunlardan ikincisi uygulamada daha zordur. Her kullanıcı için sayfaların yeniden üretilmesini gerektirir.

Her iki yaklaşım için de öğrencinin bilgi düzeyi, amacı, öğrenme metodu, öğrenme yeteneği gibi bazı özelliklerinin bilinmesi büyük önem taşır. Bu bilgiler bir Öğrenci Modeli şeklinde toparlanmalıdır. En genel anlamda Öğrenci Modeli, sistemin öğrenciye ilişkin yargılarını ve aldığı kararları temsil eder. Sistem, uyarlanırlık görevini (adaptivity task) gerçekleştirmek için bu modeli kullanılır. Teorik olarak komple bir öğrenci modelinin, yukarıda sayılan öğrenci özelliklerinin yanı sıra, geçmişteki bilgi ve deneyimleri, grup çalışmasına yatkınlığı ve ilgili daha başka bilgileri de tutması öngörülür.

Değerlendirilecek öğrenci özelliklerine ilk değerlerin nasıl atanacağı konusunda iki ayrı görüş vardır. Bunlardan birisi, başlangıçta öğrenciye test uygulamak ve verdiği cevaplardan eğitim bilimleri ışığında yorumlar yapmak şeklindedir. Böylece sistem öğrencinin bilgi düzeyi dışındaki bazı özelliklerini tanımış olur. Diğer bir yöntem ise, başlangıçta belli öğrenci grupları tanımlamak, öğrencileri bu gruplara rasgele dağıtmak ve sonra öğrenci davranışlarını izledikçe öğrenci dağılımları üzerinde gerekli değişiklikleri yapmak şeklindedir.

4.3.2. Öğrenci çözümlerinin irdelenmesi ve hata kaynaklarının bulunması

Bu teknoloji de, sistemden sisteme farklılık gösterir. Kimi sistemlerde sadece öğrencinin verdiği yanıt ele alınırken, kimi sistemlerde, problem çözme adımları tek tek ele alınır. İdeal bir ITÖ Sisteminde, çözümün doğru olup olmadığının, yanlış varsa nerede olduğunun, hangi olası bilgi eksikliği ya da yanlış bilginin hataya sebep olduğunun bulunması gerekir [39].

4.3.3. Etkileşimli problem çözme desteği

Bu teknoloji daha yeni ve daha güçlü bir teknolojidir. Öğrencinin son yanıtının ne olduğunu beklemek yerine, öğrenciye problem çözümünün her aşamasında akıllı yardım desteği vermeyi hedefleyen bir teknolojidir.

4.3.4. Uyarlanır gezinme desteđi

Bu teknolojinin hedefi, i ve dıř bađlantılardan (internal-external links) hangilerinin đrenciye gsterileceđine ve nasıl gsterileceđine đrenci zellikleri dođrultusunda karar vermektir. Bylece đrencinin kendisine en uygun yolu bulması iin destek sađlanır.

4.4. Internet Tabanlı đretim Sistemlerinin Mimarisi

Bir IT Sisteminin drt temel bileřeni vardır:

- Bilgi Anabirimi
- đrenci Modeli
- Eđitim Birimi
- Kullanıcı Arayz

Sz edilen bileřenlerden Eđitim Birimi ile Kullanıcı Arayz Birimi iřlevsel birim; Bilgi Anabirimi bilginin saklandığı birim; đrenci Modeli ise hem iřlevsel birim, hem de bilginin saklandığı birim olarak dřnlmřtr ve sistemi oluřturan bileřenler bu řekilde deđerlendirilmiřlerdir.

Ařađıda bu bileřenlerin birbirleriyle iliřkilerinin gsterimi ve her birinin iřlevine iliřkin kısa aıklamalara yer verilmektedir [40].

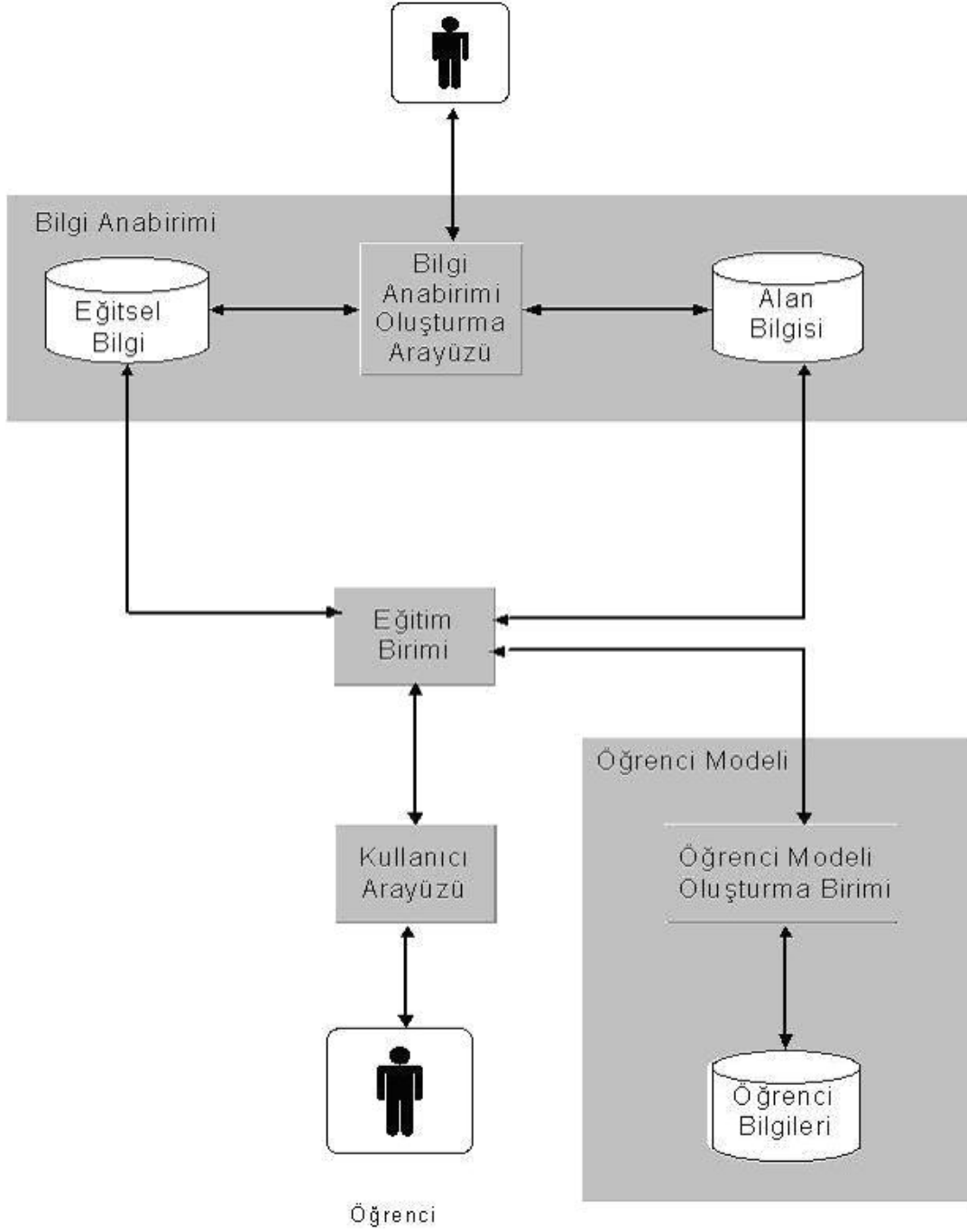
4.4.1. Bilgi anabirimi

đretilecek ana bilginin ve eđitsel bilgilerin tutulduđu birimdir. Bilgi Anabiriminin kalitesi deđiřkendir ve sistemin kalitesini dođrudan etkiler. Buradaki bilginin kalitesine ve tasarımına gre Internet Tabanlı đretim Sistemi problemler karřısında zm retebilir ya da sadece sunma amacını gerekleřtirebilir [43].

4.4.2. đrenci modeli

đrenci modeli, sistemde yer alan đrencilerin zelliklerinin tutulduđu birimdir. đrenciden srekli bilgi toplayan ve bir sonraki aktivitenin ne olacađına karar veren mekanizma; sistemlerin z niteliđindedir ve iyi bir đrenci modelinin varlığını gerektirir. Toplanan bilgiler, đrencinin uzun dnemli ve kısa dnemli bilgileri

olabilir. Uzun dönemli bilgiler arasında en önemlileri; öğrencinin amacı, bilgi seviyesi ve öğrenme metodu bilgileridir.



Şekil 4.1. İnternet Tabanlı Öğretim Sistemlerinin Genel Mimarisi [39].

Kısa dönemli bilgiler arasında en önemlileri; öğrencinin bir bilgi ögesine karşılık gösterdiği davranışı, hata kaynakları ve bilgi ögesi için harcadığı zaman bilgileridir. Öğrenci sistemden çıkarken öğrenci bilgileri veri tabanında saklanır, sisteme girerken bu bilgiler tekrar veri tabanından alınıp kullanılır [39].

4.4.3. Eğitim birimi

İnternet Tabanlı Öğretim Sistemlerinin en önemli birimlerinden biri olan Eğitim Birimi, bilgi anabiriminde yer alan bilgiyi ve öğrenci modelini kullanarak, sistemin nasıl davranacağına karar veren yapıyı içerir. Bu kararlar, bir sonraki konu seçimi, soru seçimi, hata mesajlarının uygun biçimde gösterimi, bir ipucunun gösterilip gösterilmeyeceği gibi konular olabilir. Örneğin bir sonraki konu seçimi sırasında, yeni bir konunun getirilmesi ile eski bir konunun belli bir bölümünün yeniden getirilmesi arasında karar vermek için eğitim birimi devreye girer [39].

4.4.4. Kullanıcı arayüzü

Kullanıcı arayüzü, İnternet Tabanlı Öğretim Sistemi ile öğrenci arasındaki iletişimin gerçekleşmesini sağlayan birimdir. Bu birim, öğrencinin devam etmesi için gerekli motivasyonu sağlayabilmelidir. Eğer öğrenci sistemi kullanma arzusunu yitirirse, sistemin hiçbir etkinliğinin olmayacağı açıktır [54].

Bir İnternet Tabanlı Öğretim Sisteminin bütün bileşenleri önemlidir ancak, Öğrenci Modelleme Birimi içlerinde en önemli olanıdır. Öğrenci modeli kötü ise, sistem öğrenciyi yeterince tanıyamaz ve diğer birimlerin işlevleri ne kadar kaliteli olursa olsun, alınan kararların kötü olmasına sebep olur. Bu nedenle Öğrenci Modelleme konusu, İnternet Tabanlı Öğretim Sistemlerinin geliştirilmesinde pek çok kayda değer çalışmanın bir konu olmuştur.

4.5. Kocaeli Üniversitesi Enformatik bölümü İnternet Tabanlı Öğretim Sitesi ve Hazırlanan Veri Tabanı

2004-2005 Öğretim yılı bahar döneminde, Kocaeli Üniversitesinin Sağlık Yüksekokulu'na ait iki bölüm ve aynı üniversitenin Meslek Yüksekokulu'na ait dört bölümün birinci sınıflarında ders içerikleri ve kredisi aynı olan Temel Bilgisayar Bilimleri zorunlu dersi İnternet tabanlı olarak verilmiştir. Derse toplam 100 öğrenci katılmıştır. Öğrenciler, öğretim elemanı tarafından belirlenen kullanıcı adı ve şifresini kullanarak ders notlarına ulaşmıştır. Dersle ilgili ders notları, ödevler, değerlendirme sınavları, ilişkili linkler, öğretim elemanın yapacağı duyurular hazırlanan sitede yer almaktadır. ColdFusion programlama diliyle tasarlanan site, Kocaeli Üniversitesi Enformatik bölümü'nün sunucusunda yayınlanmıştır. Veriler Access'te hazırlanan veri tabanında bulunan tabloların yapısı aşağıda verilmiştir.

Tablo 4.3. “bölüm” Tablosu.

| Alan Adı | Açıklaması |
|----------|--------------------------------|
| deptid | Bölüm kodu |
| deptad | Öğrencinin okuduğu bölümün adı |

“bölüm” tablosu, öğrencilerin okuduğu bölümlerin kodlarını tutar. 7 kayıt vardır.

Tablo 4.4. “öğrenci” Tablosu.

| Alan Adı | Açıklama |
|----------|--|
| kid | Öğrenci numarasıdır. Siteye girerken “kullanıcı adı” olarak kullanılır |
| ad | Öğrencinin adı. |
| soyad | Öğrencinin soyadı. |
| sifre | Siteye girerken kullanılan şifre. |
| dept | Öğrencinin okuduğu bölüm |

“öğrenci” tablosu, öğrenci kimlik bilgilerini tutar, 136 kayıt vardır.

Tablo 4.5. “konu” Tablosu.

| Alan Adı | Açıklaması |
|----------|---|
| konuid | Ders konularının anlatıldığı sayfalara verilen numaradır. |
| konuadi | Konu başlığı bilgisidir. |
| tarih | Konunun anlatıldığı tarih |
| bolum | Konunun ait olduğu bölümün numarası |

“Konu” Tablosu, sitede yer alan ders içerik sayfalarının bilgilerini tutar, 122 kayıt vardır.

Tablo 4.6. “bolum” Tablosu.

| Alan Adı | Açıklaması |
|----------|------------|
| bolumid | Bölüm kodu |
| bolumadi | Bölüm adı |

“bölüm” tablosu, dersin bölüm bilgilerini tutar, 5 kayıt vardır.

Tablo 4.7. “quiz” Tablosu.

| Alan Adı | Açıklaması |
|----------|--------------------------|
| quizno | Değerlendirme numarası |
| quizadi | Değerlendirme sınavı adı |

“quiz” tablosu, değerlendirme sınavı bilgilerini içerir, 5 kayıt vardır.

Tablo 4.8. “soru” Tablosu.

| Alan Adı | Açıklaması |
|----------|--|
| konuno | Sorunun ait olduğu konu numarası |
| quizno | Sorunun ait olduğu değerlendirme sınavı numarası |
| soruno | Soru numarası |
| soru | Sorunun metni |
| d_cevap | Doğru cevap şıkkı |

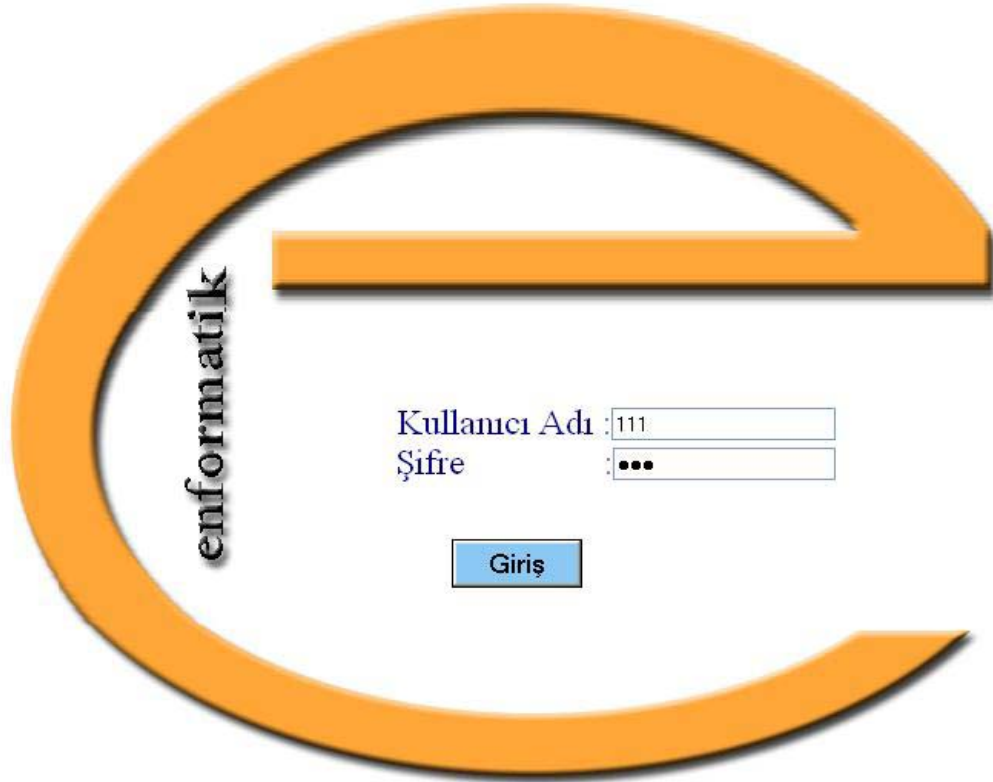
“soru” tablosu, değerlendirme sınavlarının sorularını tutar, 50 kayıt vardır.

Tablo 4.9. “odevler” Tablosu.

| Alan Adı | Açıklaması |
|----------------|--------------------------------|
| ogrid | Öğrenci numarası |
| odevno | Odev numarası |
| bolumno | Ödev ait olduğu bölüm numarası |
| verilis_tarihi | Ödevin verilmiş tarihi |
| teslim_tarihi | Ödevin teslim edildiği tarih |
| notu | Öğrencinin ödevden aldığı not |

“ödevler” tablosu, öğrencilerin ödev bilgilerini tutar, 1000 kayıt vardır.

<http://enformatik.kou.edu.tr/verimadeni/index.cfm> adresine bağlanan öğrenciler şekil 4.1’deki giriş ekranıyla karşılaşmıştır.



enformatik

Kullanıcı Adı : 111

Şifre : ●●●

Giriş

Şekil 4.1. Şifre Giriş Ekranı.

Siteye bağlanan öğrenciler şekil 4.1. ekranında kullanıcı adı ve şifrelerini girerek bir defaya mahsus olarak şekil 4.2'deki anket bilgilerini doldurmuşlardır. Formdaki veriler veri tabanına kodlanarak kaydedilmiştir. Kodlamalar:

Yaş aralığı: 1-[17-18], 2-[19-20], 3-[21-22], 4-[23-24], 5-[25-26], 6-[27 ve üstü]

Medeni Durum: 1-Bekar, 2-Evli

Kaldığı yer: 1-Yurt, 2-Ev, 3-Akraba, 4-Tanıdık, 5-Diğer

Baba Mesleği: 1-Memur, 2-İşsiz, 3-Yönetici, 4-Emekli, 5-Serbest Meslek, 6-İşçi

Anne Mesleği: 1-Memur, 2-Yönetici, 3-Ev Hanımı, 4-Serbest Meslek, 5-İşçi,

6-Emekli

Baba Sağ/Ölü: 1-Sağ, 2-Ölü

Anne Sağ/Ölü: 1-Sağ, 2-Ölü

Anne-Baba Birlikte/Ayrı: 1-Birlikte, 2-Ayrı

Bitirdiği Lise: 1-Düz Lise, 2-Meslek Lisesi, 3-Teknik Lise, 4-Anadolu Teknik Lisesi,

5-Anadolu Lisesi, 6-Fen Lisesi, 7-Süper Lise, 8-Açıköğretim Lisesi

ANKET FORMU

| | | |
|--|--|--|
| Doğum Yeri: | <input type="text"/> | |
| Doğum Tarihi: | <input type="text"/> | |
| Medeni Durumu: | BEKAR <input type="button" value="v"/> | |
| Kardeş Sayısı: | <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 ve üstü | |
| Kaldığı Yer: | <input type="radio"/> Yurt <input type="radio"/> Ev <input type="radio"/> Akraba <input type="radio"/> Tanıdık <input type="radio"/> Diğer | |
| Baba Mesleği: | Memur <input type="button" value="v"/> | |
| Baba Sağ/Ölü: | <input type="radio"/> Sağ <input type="radio"/> Ölü | |
| Anne Mesleği: | Memur <input type="button" value="v"/> | |
| Anne Sağ/Ölü: | <input type="radio"/> Sağ <input type="radio"/> Ölü | |
| Baba-Anne : | <input type="radio"/> Birlikte <input type="radio"/> Ayrı | |
| Bitirdiğiniz Lise: | Düz Lise <input type="button" value="v"/> | |
| Hobileriniz: | | |
| <input type="checkbox"/> Kitap Okumak | <input type="checkbox"/> Müzik Dinlemek | <input type="checkbox"/> Sinemaya Gitmek |
| <input type="checkbox"/> İnternet | <input type="checkbox"/> El Sanatları | <input type="checkbox"/> Spor Yapmak |
| <input type="checkbox"/> Seyahat Etmek | <input type="checkbox"/> Televizyon Seyretmek | <input type="checkbox"/> Diğer |
| <input type="button" value="GÖNDER"/> | | |

Şekil 4.2. Anket Formu.

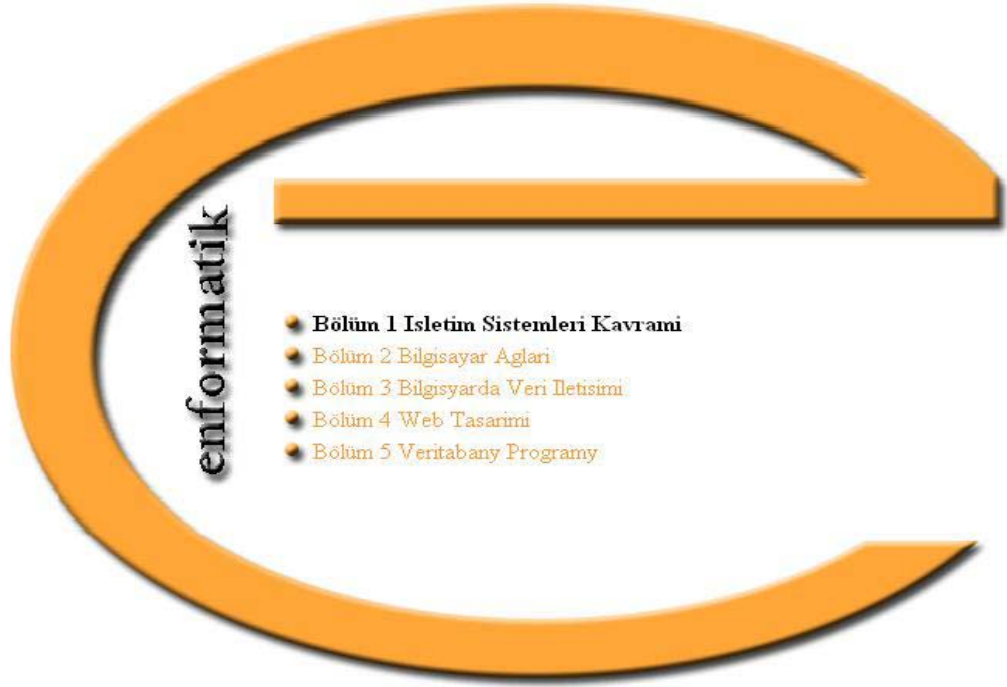
Şekil 4.2’deki anket formundan girilen veriler anket tablosuna kaydedilir.

Tablo 4.10. “anket” Tablosu.

| Alan Adı | Açıklaması |
|----------|------------------------------|
| ogrid | Öğrenci numarası |
| d_yeri | Doğum yerinin trafik il kodu |
| yas | Yaşı |
| medenid | Medeni durumu |
| kardes | Kardeş sayısı |
| kalyer | Kaldığı yer |
| babames | Babasının mesleği |
| annemes | Annesinin mesleği |
| baba | Baba sağ/ölu bilgisi |
| Anne | Anne sağ/ölu bilgisi |
| Birlikte | Anne ve baba beraber mi |
| hobi | Öğrencinin hobileri |

“anket” tablosunda 122 kayıt vardır.

Anket formunu dolduran öğrencilerin karşına ilgili dersin bölüm konuları gelecektir. Şekil 4.3’deki bölümlerden yayına açık olan, yani aktif olan bölümü tıklayarak ders notlarına ulaşmaktadır. Her bölüm sırayla yayınlanmaktadır.



Kocaeli Üniversitesi Enformatik Bölümü

Şekil 4.3. Bölüm girişleri.

Şekil 4.3'den ilgili bölüm tıklanarak şekil 4.4'deki ders içeriklerine ulaşılır.

Şekil 4.4. Ders İçerikleri.

Şekil 4.4’de bir örneği görülen ders içerik sayfaları açıldığında sistem tarihi alınır ve öğrencinin sayfada kaldığı süre hesaplanır. Veri tabanında öğrencinin o sayfaya yaptığı girişler araştırılıp süre ve giriş sayısı bilgileri güncellenir. Veriler giriş tablosunda tutulur. Bu tabloda toplam 2200 kayıt vardır.

Tablo 4.11. “giris” Tablosu.

| Alan Adı | Açıklaması |
|-------------|--|
| kullanici | Öğrenci numarası |
| konu | Girilen sayfa(konu) numarası |
| girissayisi | Sayfaya kaç kere girildiğini gösterir |
| sonsure | Sayfada kaç dakika kalındığını gösterir. |
| oturum | Öğrencinin kaçınıcı oturumu olduğunu gösterir. |

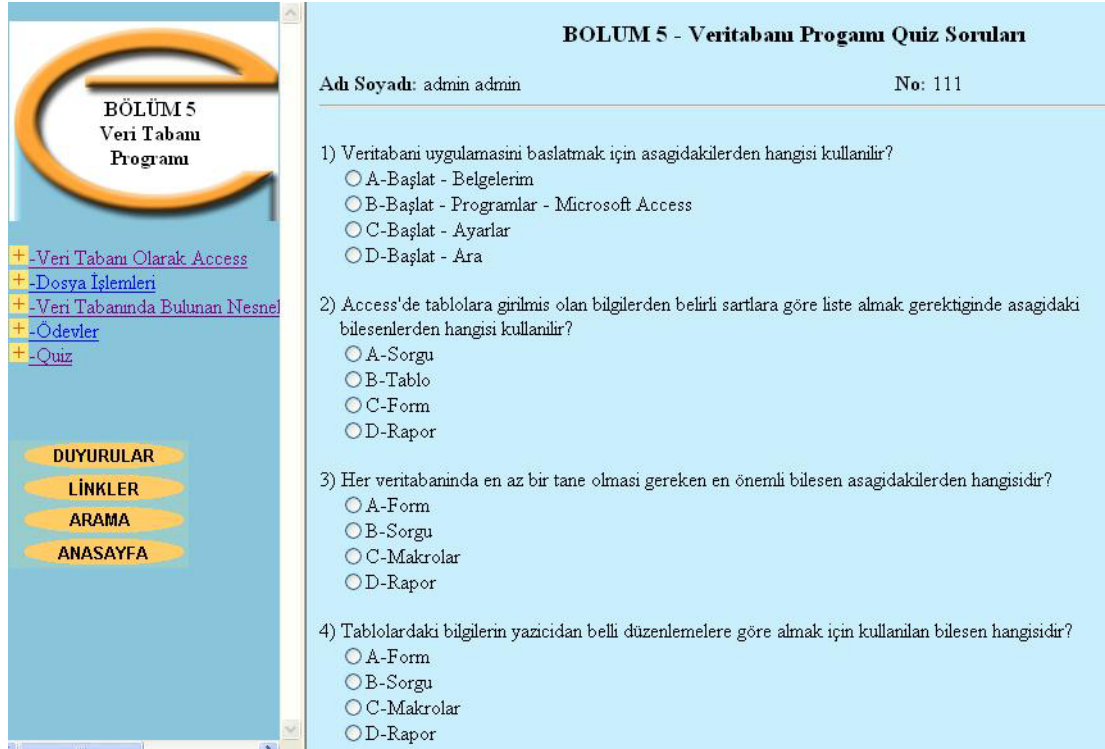
Şekil 4.4’de açılan sayfada öğrenci ders notlarına ulaşabildiği gibi bölümle ilgili duyurulara, ilgili bağlantılara ulaşabilmiştir. Buradaki ARAMA butonu ile de herhangi bir kelime ya da cümleyi de arayabilmektedirler. Ödevleri de bağlandıkları bölüm sayfasından alarak son teslim tarihine kadar belirtilen mail adresine gönderebilmeleri sağlanmıştır.

Öğrenci her bölüm sonunda değerlendirme sınavı olmaktadır. Değerlendirme sınavı sonuçlarını ise değerlendirme sınavı sorularını bitirdikten sonra Gönder butonuna tıkladıktan sonra alacaktır. Öğrencinin verdiği cevaplar “cevap”, aldığı not “quiznotu” tablolarına kaydedilir.

Tablo 4.12. “cevap” Tablosu.

| Alan Adı | Açıklaması |
|-----------|--------------------------------|
| kullanici | Öğrenci numarası |
| quizno | Değerlendirme sınavı numarası |
| soruno | Soru numarası |
| cevap | Öğrencinin verdiği cevap şıkkı |

“cevap” tablosunda 5000 kayıt vardır.



Şekil 4.5. Değerlendirme Sınavı.

Tablo 4.13. “quiznotu” Tablosu.

| Alan Adı | Açıklaması |
|-----------|-------------------------------|
| kullanici | Öğrenci numarası |
| quizno | Değerlendirme sınavı numarası |
| notu | Aldığı not |

“quiznotu” tablosunda 500 kayıt vardır.

Yapılan sitede, ödevler de sayfa üzerinden yayınlanmaktadır. Öğrenciler internet üzerinden ödevlerini alarak son teslim tarihine kadar eğitmeninin mail adresine göndermektedir.

BÖLÜM 5
Veri Tabanı
Programı

- + [-Veri Tabanı Olarak Access](#)
- + [-Dosya İşlemleri](#)
- + [-Veri Tabanında Bulunan Nesnel](#)
- + [-Ödevler](#)
- + [-Quiz](#)

DUYURULAR

LİNKLER

ARAMA

ANASAYFA

ÖDEVLER

»Veritabanı nedir? Veritabanı programları hakkında bilgi toplayın.

»Sınıfınızda bulunan öğrencilerin ENF101 Temel Bilgi Teknolojileri Kullanımı dersinden almış olduğu vize ve final notlarının bulunduran, ayrıca bu kişilerin ortalamasını hesaplatıp, bütçe kalıp kalmadığını bildiren bir veritabanı oluşturunuz.

Şekil 4.6. Ödevler.

Her bölüme girildiğinde ilgili bağlantı adresleri linkler butonuna tıklanarak ulaşılabilir. Şekil 4.7’de görülen ekrandan Öğrenci buradan istediği bağlantı adresini tıklayarak ilgili web sitesine ulaşıp faydalanabilir.

BÖLÜM 5
Veri Tabanı
Programı

- + [-Veri Tabanı Olarak Access](#)
- + [-Dosya İşlemleri](#)
- + [-Veri Tabanında Bulunan Nesnel](#)
- + [-Ödevler](#)
- + [-Quiz](#)

DUYURULAR

LİNKLER

ARAMA

ANASAYFA

KAYNAK LİNKLER

www.freecode.com

www.programmerkezi.com

www.perfect.com.uk

Şekil 4.7. Kaynak Linkler.

BÖLÜM 5. İNTERNET TABANLI ÖĞRETİMDE VERİ MADENCİLİĞİ TEKNİKLERİNİN UYGULANMASI

5.1. Giriş

Veri madenciliği teknikleri, büyük ölçekli veritabanlarında veriler arasındaki ilginç desenleri bulmaya yarayan yöntemlerdir.

Bu tezde, veri madenciliği birliktelik kuralları tekniklerinden apriori algoritmasıyla, kümeleme tekniklerinden DBSCAN algoritması uygulanarak internet tabanlı bir öğretim sisteminde öğrenci davranışları analiz aracı geliştirilmiştir. Uygulamada ilk olarak bir web sayfası tasarlanmıştır. Tasarlanan siteye, Kocaeli Üniversitesi Sağlık Yüksekokulu ve Arslanbey Meslek Yüksekokulu'nda bulunan bölümlerin öğrencileri 2003-2004 eğitim-öğretim yılında kendilerine sağlanan kullanıcı adı ve şifresiyle bağlanmışlardır. Site içerisindeki öğrenci hareketleri Microsoft Access veritabanına kaydedilmiştir. Tasarlanan internet tabanlı öğretim sitesinden elde edilen bu veriler, VTBK adımlarından biri olan veri ön işleme adımı uygulanarak gereksiz yinelenmelerden ve null değerlerinden temizlenmiştir. MS Access veritabanındaki tablolar, sorgular yardımıyla birliktelik kurallarının bulunabileceği ve DBSCAN'ın uygulanabileceği koşullara göre düzenlenerek, bir arayüz ile uygulama tasarlanmıştır. Bu bölümde uygulamanın adımları ve çalışma prensibi örnekler üzerinde detaylı bir şekilde anlatılacaktır. Uygulama Delphi 6.0 ile geliştirilmiş, veritabanı olarak Microsoft Access kullanılmıştır.

5.2. Neden Birliktelik Kuralları ve DBSCAN Algoritması

Birliktelik kuralları genellikle müşterilerin alışveriş desenlerini bulmak amacıyla uygulanmıştır. Ancak bu tekniğin bu tezdeki kullanım amacı, İnternet tabanlı bir öğretim sitesine bağlanan ve yine aynı sitedeki değerlendirme sınavı sınavlarına katılan öğrencilerin, birlikte giriş yaptığı sayfalar ve birlikte yanlış yaptığı

değerlendirme sınavı sorularını bulmaktır. Bu sayede geniş ve karmaşık giriş ve not tabloları arasındaki birliktelikler tespit edilebilecek ve bu birliktelikler; yayınlanan ders içeriği, bir internet tabanlı öğretim sitesinde öğrenci davranışları ve dersi veren öğretim elemanı hakkında yorumlar yapmak amacıyla kullanılmıştır.

Kocaeli Üniversitesi Sağlık Yüksekokulu ve Meslek Yüksekokulu bölümlerinin birinci sınıflarında okutulmakta olan ENF 102 Temel Bilgisayar Bilimleri dersi uygulamada örnek ders olarak alınmıştır. İnternet tabanlı bir öğretim sistemi tasarlanarak ENF 102 Temel Bilgisayar Bilimleri dersi internet üzerinden yayınlanmıştır. Yayınlanan sitede ders notları, ödevler, değerlendirme sınavları yer almaktadır. Ders notları, beş bölümden oluşmaktadır ve, her bölüm 2004-2005 bahar döneminde sırasıyla yayınlanmıştır. Her bölüm sonunda 10 sorudan oluşan bir değerlendirme sınavı yapılmış ve her bölüm için 2 ödev verilmiştir. Siteye bağlanan öğrencilerin davranışları ve değerlendirme sınavı sonuçları bir Microsoft Access veritabanında toplanmıştır. Delphi 6.0 ile yapılan bir arayüz ile veritabanındaki toplam giriş ve not tabloları üzerinde apriori algoritması ile baştan belirlenen destek ve güven değerleri de kullanılarak ilgili derse ait giriş ve yanlış yapılan soruların birliktelikleri bulunmaktadır.

Bulunan birliktelikler ile, Temel Bilgisayar Bilimleri dersinde yapılan öğretim sitesine bağlanan öğrencilerin girdikleri sayfalar ve değerlendirme sınavlarına vermiş oldukları yanlış cevap birliktelikleri ortaya çıkarılmaktadır. Bulunan birliktelikler öğrencilerin derse olan ilgilerini, ders içeriklerinin yeterliliğini, sınav sorularının zorluk derecelerini ortaya koymak için gerekli bilgileri sağlamaktadır.

Tezde kullanılan diğer bir metot ise, veri madenciliği kümeleme tekniklerinden DBSCAN algoritmasıdır. DBSCAN algoritması, yoğunluk tabanlı bir algoritma olup, daha çok meteoroloji ve deprem fay hatlarını belirlemede uygulanmıştır. Bu tezde DBSCAN algoritmasının kullanım amacı, önceden belirlenen not yakınlığı ve kümeyi oluşturacak öğrenci sayısı kriterleri dikkate alınarak veritabanında bulunan tablolardaki yoğunluklar, tespit edilmektir. Veriler doğrusal olduğu için DBSCAN algoritması kullanılmıştır.

DBSCAN algoritması ile bulunan kümelerle öğrencilerin ödev ve değerlendirme sınavı notlarının hangi not aralıklarında yığılma gösterdiği tespit edilmiştir.

5.3. Uygulamanın Açıklaması

Program çalıştırıldığında ilk olarak Şekil 5.1'deki açılış ekranı ekrana gelir.



Şekil 5.1. Açılış Ekranı

Şekil 5.1'deki açılış ekranı bir defa tıklanınca şekil 5.2'deki arayüz akranına gelir. Bu arayüzde;

- 1- Genel Değerlendirme,
- 2- Birliktelik Kuralı Çıkarma,
- 3- Ağırlıklı Notlar,
- 4- Çıkış

menüleri tasarlanmıştır. Genel Değerlendirme menüsünde; öğrencilerin sayfaya giriş hareketleri, değerlendirme sınavı sonuçları ve ödev bilgileri incelenmektedir.

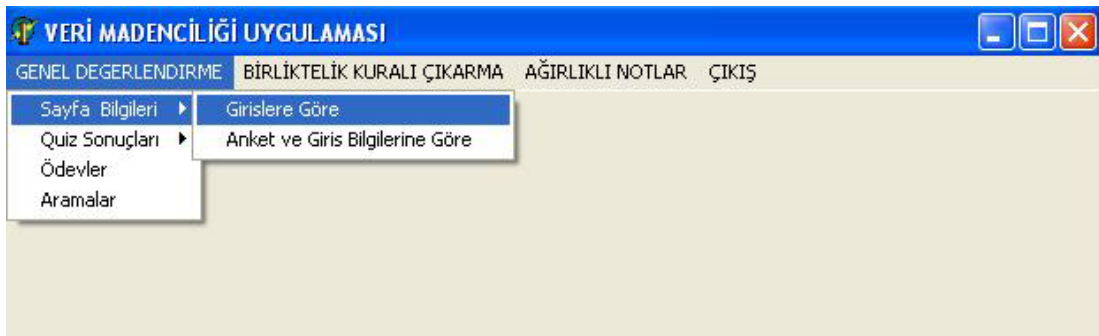
Birliktelik Kuralı Çıkarma menüsünde; Apriori algoritması ile birlikte girilen sayfalar ve birlikte yanlış yapılan sorular değerlendirilmektedir. Ağırlıklı Notlar menüsünde ise değerlendirme sınavı ve ödev notlarının hangi not aralıklarında yığılma gösterdiği DBSCAN algoritmasıyla bulunmuştur.



Şekil 5.2. Programın Açılış Ara yüzü.

5.3.1. Genel değerlendirme menüsü

Şekil 5.2deki arayüzden Genel Değerlendirme tıkladığında, şekil 5.3'deki seçenekler ekrana gelir.

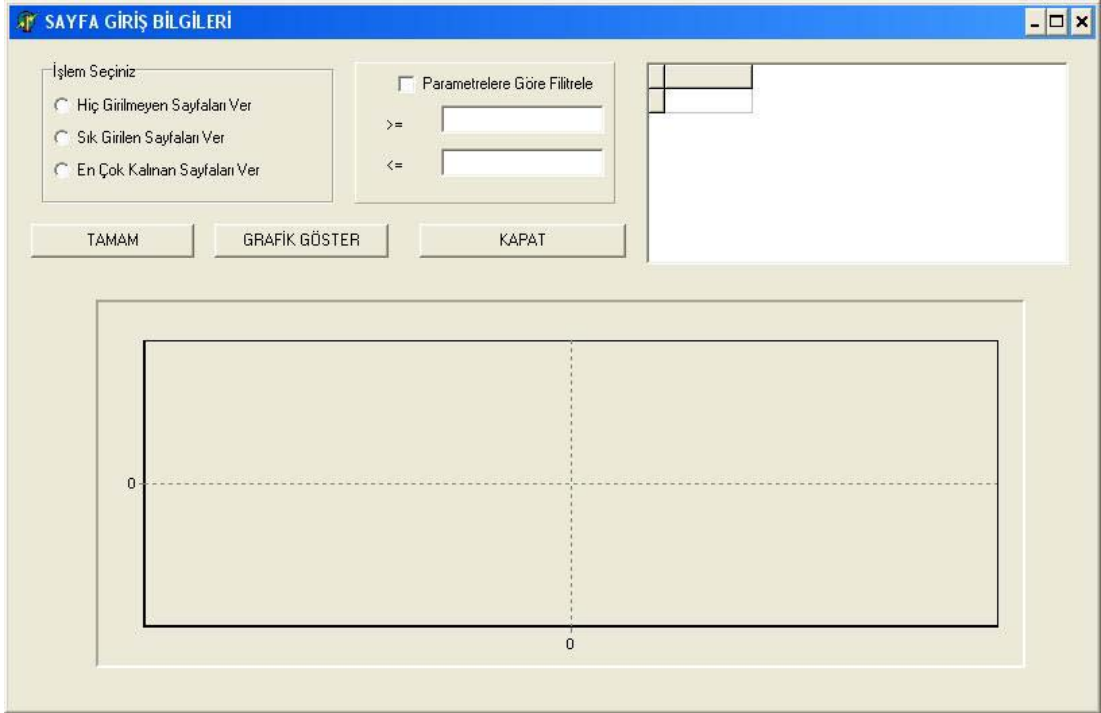


Şekil 5.3. Genel Değerlendirme Ekranı.

5.3.1.1. Sayfa bilgileri

5.3.1.1.2. Girişlere göre

Şekil 5.3'den Sayfa Bilgileri → Girişlere Göre seçeneği tıklandığında, öğrencinin sayfaya giriş sayıları ve kalma süreleri kullanılarak istatistik bilgileri ortaya çıkarılmıştır.



Şekil 5.4. Sayfa Giriş Bilgileri.

Şekil 5.4.'de, Hiç Girilmeyen Sayfalar, Sık Girilen Sayfalar, En çok Kalınan Sayfalar şeklinde üç işlem seçeneği vardır.

- 1- Hiç Girilmeyen Sayfalar: Ziyaret edilmeyen sayfaları verir. Herhangi bir parametre girişine ihtiyaç duyulmayacağından; Parametrelere Göre Filtrele seçeneği pasif duruma gelir.
- 2- Sık Girilen Sayfalar: Öğrencilerin en sık girdiği sayfaları verir. Bu, sayfalara girilme oranlarını gösteren tanımlayıcı bir bilgidir. Bütün sayfalar kontrol

edilebileceği gibi parametreye göre de (örn, 10'dan fazla girilen sayfalar) incelenebilir.

Tablo 5.1'deki kod, öğrencilerin, hazırlanmış olan siteye yapılan toplam ziyaret sayısını bulmak için kullanılmıştır.

Tablo 5.1. Öğrencilerin siteye yapmış oldukları toplam ziyaret sayısı.

```
adotable1.TableName:='giris';
adotable1.Open;
adotable1.First;
toplam:=0;
while not adotable1.Eof do
begin
toplam:=toplam + adotable1.FieldValues['girissayisi'];
adotable1.Next;
end;
```

Toplam ziyaret sayısını bulduktan sonra sayfa giriş oranlarını bulmak için, girilen parametreye göre uygun SQL ifadesi çalıştırılır. Örneğin; iki parametre girildiği zaman tablo 5.2'deki SQL ifadesi çalışır. (“Parametreye Göre Filtrele” seçeneği işaretli olmalıdır. Aksi halde girilen parametreler dikkate alınmaz.)

Tablo 5.2. Parametreye Göre Filtreleme İşlemi.

```
'select konu as Konular, sum(girissayisi) as Giriş_Sayisi,
sum(girissayisi)*100)' + floattostr(toplam) + ' as Oran from giris
group by konu having sum(girissayisi)>=' + edit1.text + '
and sum(girissayisi)<=' + edit2.text + ' order by sum(girissayisi)
desc'
```

3- En Çok Kalınan Sayfaları Ver: Öğrencilerin uzun süre kaldığı sayfaları verir.

Tablo 5.3'de öğrencilerin bütün ziyaretleri boyunca sayfalarda kaldığı süreler toplanarak “sure” değeri bulunur. Yine girilen parametreye uygun SQL ifadesi çalıştırılır. Sonuç olarak her hangi bir sayfada kalınan sürenin toplam sureye oranı

elde edilir. Grafiğini görüntülemek için grafiğin Y ekseninde gösterilecek alan ismi “Kalinan_Sure” olarak ayarlanır.

Tablo 5.3. Öğrencilerin Sitede Kaldıkları Toplam Süre.

```
adotable1.TableName:='giris';
adotable1.Open;
adotable1.First;
toplam:=0;
while not adotable1.Eof do
begin
sure:=sure + adotable1.FieldValues['sonsure'];
adotable1.Next;
end;
```

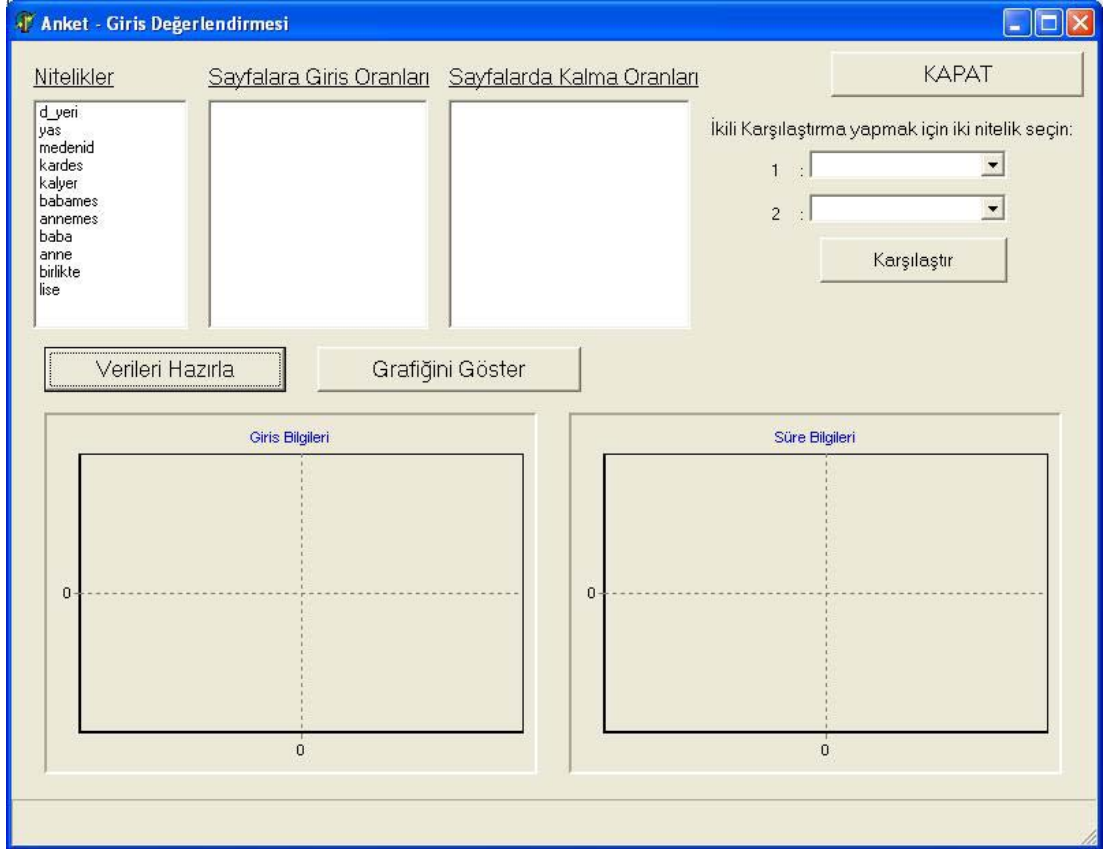
5.3.1.1.3. Anket ve girişlere göre

Öğrencilerin doldurdıkları ankete göre sayfa ziyaretlerindeki farklılıkları gösteren işlem seçeneğidir. İşlemlere başlamadan önce “anket” ve “giris” tabloları tablo 5.4’deki SQL ile birleştirilerek veriye ön işlem uygulanmıştır.

Tablo 5.4. Anket ve Giriş Tablolarının Birleştirilmesi.

```
'select g.kullanici, a.d_yeri, a.yas, a.medenid, a.kardes, a.kalyer, a.babames,
a.annemes, a.lise, a.baba, a.anne, a.birlikte, sum(g.girissayisi) as
Toplam_Giris, sum(g.sonsure) as Toplam_Sure into anket_giris from
anket a, giris g where a.ogrid=g.kullanici
group by g.kullanici, a.d_yeri, a.yas, a.medenid, a.kardes, a.kalyer,
a.babames, a.annemes, a.lise, a.baba, a.anne, a.birlikte
```

İşlemler şekil 5.5’deki ara yüz ekranı kullanılarak yapılır. Bu ekran ilk açıldığında verileri hazırla butonu kullanılarak anket giriş tablosundaki alanlar listbox’ta listelenir. Listedenden seçilen niteliğe göre Sayfalara Giriş Oranları ve Sayfalarda Kalma Süreleri hesaplanır. Elde edilen sonuçlar [Bilgi], [Sayı], [Oran] başlıkları altında ifade edilir. Bilgi; niteliklerin içerdiği verilerin sayılarla sembolize edildiği kısımdır. Sembollerin anlamı durum çubuğunda, ekranın alt kısmında görülmektedir. Sayı; seçilen niteliğin veritabanındaki toplam kayıt sayısıdır. Oran; seçilen niteliğin sayısının veritabanındaki toplam veri sayısına oranıdır.



Şekil 5.5. Anket – Giriş Değerlendirilmesi.

Örneğin nitelikler kısmından Kalyer seçildiğinde, öğrencilerin kaldığı yerlere göre (1-Yurt, 2-Ev, 3-Akraba, 4-Tanıdık, 5-Diğer), sayfaya giriş oranları ve sayfalarda kalınan süre oranları bulunmaktadır. Tablo 5.5’deki kod ile; ilk önce seçilen niteliğin tekrarlı kayıtları kaldırılır. Bulunan verilerin; iç içe iki döngüyle Sayı ve Oran değerleri hesaplanıp listbox’lara yazdırılır.

Tablo 5.5. Seçilen Niteliğin Sayı-Oran Değerlerinin Hesaplanması.

```

y:=listbox1.itemindex;
sec:=listbox1.Items.Strings[y];
ydeg:='Toplam_Giris';
adoquery1.SQL.Add('select distinct ('+ sec + ') from anket_giris');
for j:=11 to adodataset1.FieldCount-1 do
begin
for x:=0 to adoquery1.RecordCount-1 do
begin
sql2:='select '+
sec+'sum('+adodataset1.FieldList.Fields[j].FieldName+')
as alt_top from anket_giris where '+ sec + '=' +
inttostr(adoquery1.Fields[0].Value) + ' group by '+sec;
adoquery2.SQL.Add(sql2);

```

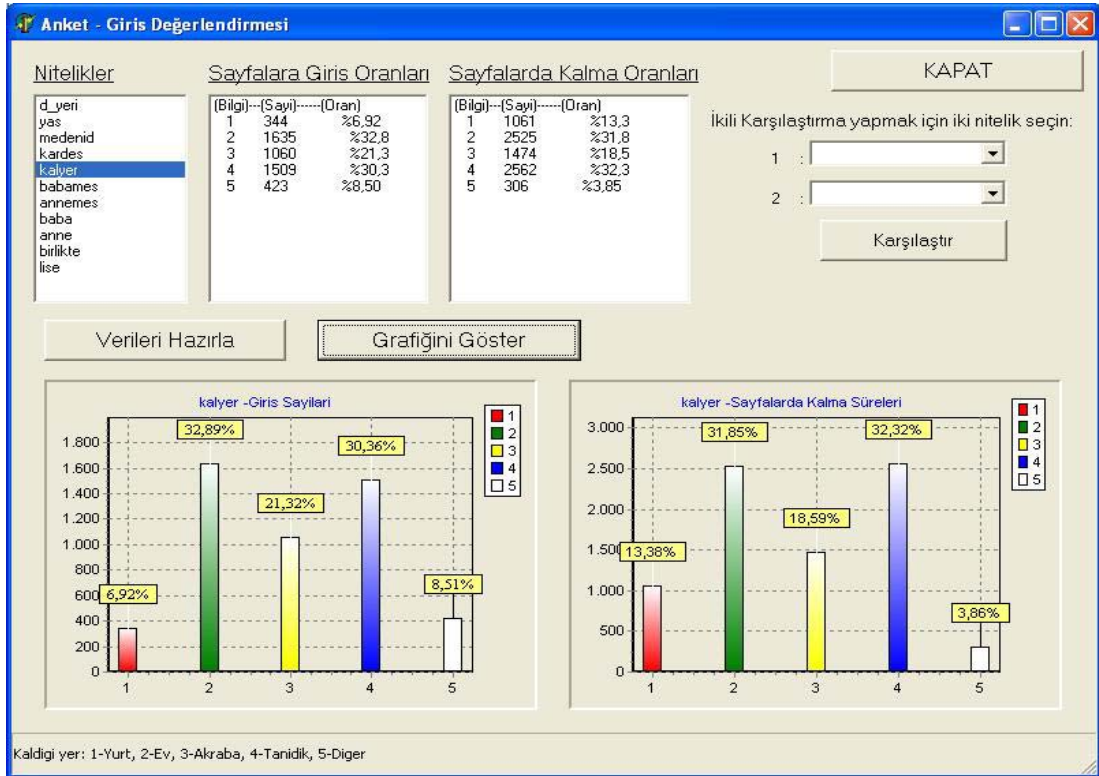
Tablo 5.5 (devam). Seçilen Niteliğin Sayı-Oran Değerlerinin Hesaplanması.

```

if adoquery2.RecordCount > 0 then
begin
  if j=11 then
    oran:=(adoquery2.FieldValues['alt_top']*100)/toplam;
  if j=12 then begin
    oran:=(adoquery2.FieldValues['alt_top']*100)/sure;
  end;
  adoquery1.Next;
end;
end;

adoquery1.First;
adoquery2.Next;
end;

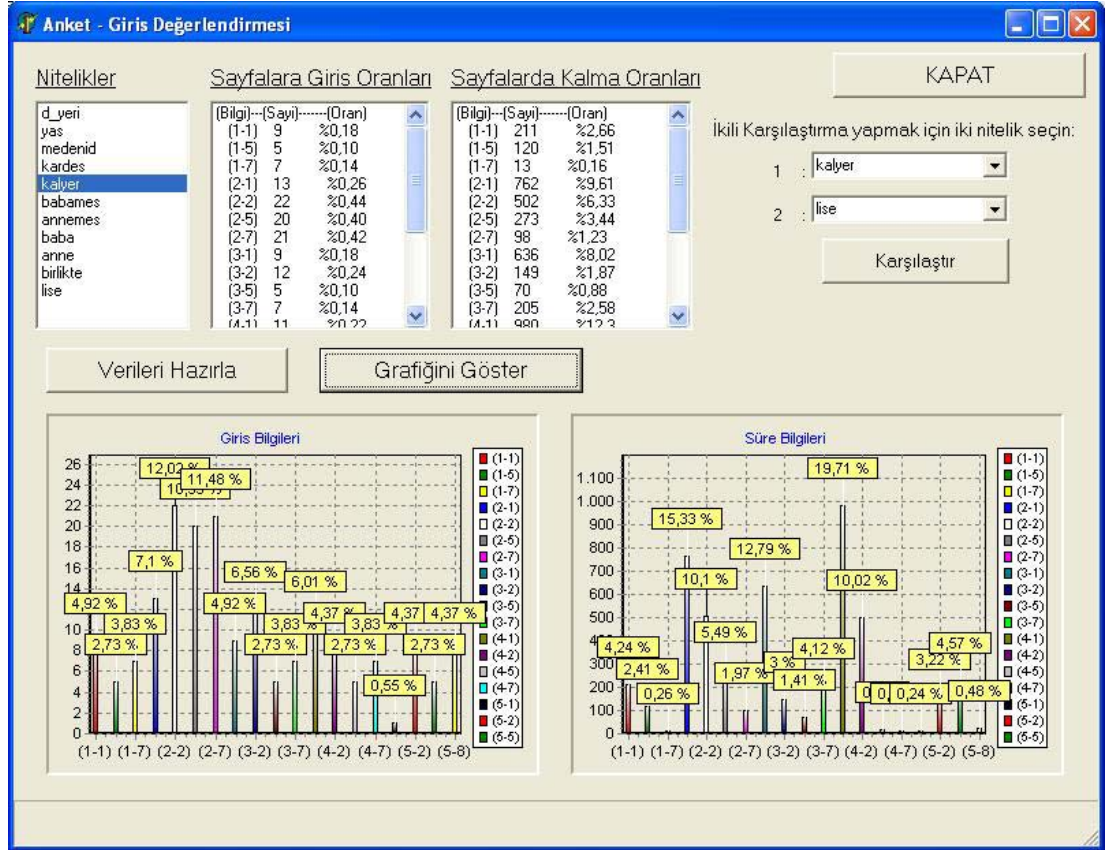
```



Şekil 5.6. Öğrencinin Kaldığı Yer Niteliğine göre Değerlendirme Sonuçları.

Şekil 5.5'deki arayüz aynı zamanda öğrencilerin, sayfalara giriş oranı ve kalma süreleri oranını iki niteliğe göre karşılaştırmalı olarak değerlendirme olanağı verir. Bunun için Şekil 5.5.'de sağ tarafta bulunan açılır liste kutularından iki nitelik seçilir. Karşılaştır butonuna tıklandığında, bu iki nitelik veritabanında birlikte değerlendirilerek sonuçlar alınır. Örneğin şekil 5.7'deki gibi öğrencinin kaldığı yer ve bitirdiği lise kriterleri ele alınarak incelenebilir. Yurtta kalıp meslek lisesi

mezunu öğrencilerin sayfalara yaptığı giriş ve sayfalarda kalma süreleri oranları alınabilir.



Şekil 5.7. Öğrencinin Kaldığı Yer ve Bitirdiği Lise Niteliğine Göre Karşılaştırılması.

5.3.1.2. Değerlendirme sınavı sonuçları

Şekil 5.3'de gösterildiği gibi, Genel Değerlendirme altında bulunur. Girişlere göre ve bölümlere göre sınav değerlendirilmesi üzere üç başlık altında incelenmektedir.

5.3.1.2.1. Sınav değerlendirilmesi

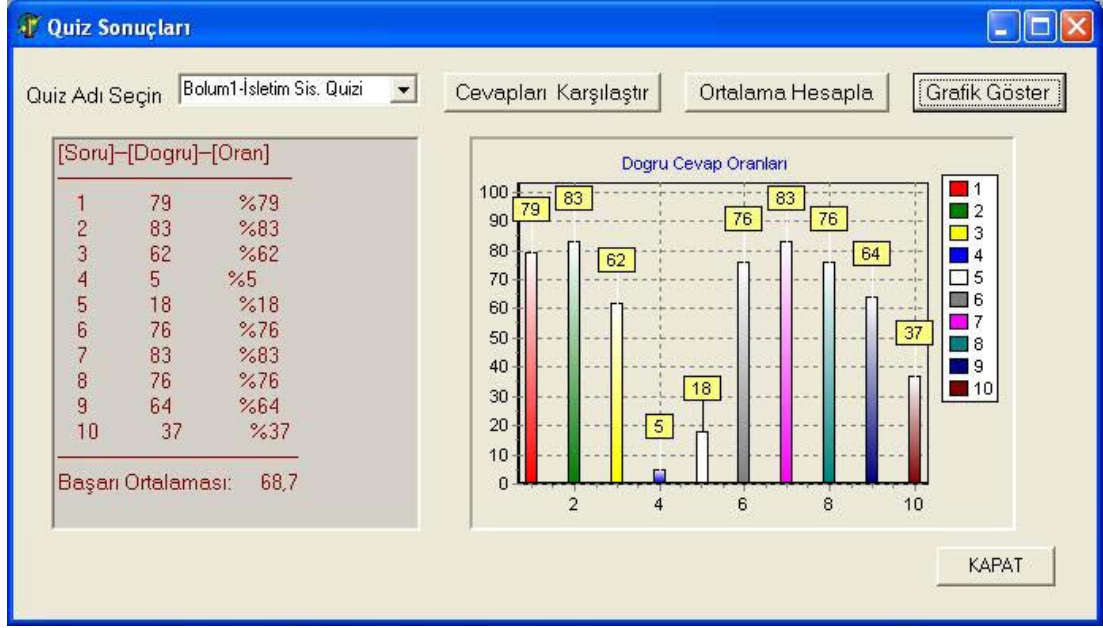
Sınav değerlendirilmesi ekranında dönem içerisinde yapılan sınavlara verilen cevapların doğruluk oranlarını bulur. Şekil 5.8'de açılır liste kutusundan incelenmek istenen sınav ismi seçilir. Cevapları karşılaştırmaya butonuna tıklanarak soruların doğru cevapları ile öğrencilerin verdiği cevaplar karşılaştırılır. Doğru cevap sayıları ve

oranlar hesaplanır. Ortalama hesaplama butonu ile seçilen sınavın başarı ortalaması bulunur. Grafik göster butonu yardımıyla oranların grafiği çizdirilir.

Tablo 5.6. Seçilen Sınavın Sorularına Verilen Cevapların Doğruluk Oranlarını Hesaplama.

```
secim:=combobox1.Itemindex+1;
  ///öğrenci sayisini bul (adoquery3)
sql:='select distinct(kullanici) from cevap where quizno=' + inttostr(secim);
  ///seçilen bölümün sorularını getir (adoquery1)
sqlstr:='select * from soru where quizno=' + inttostr(secim);
  ///sorulara verilen cevapları getir (adoquery2)
sqlstr1:='select * from cevap where quizno=' + inttostr(secim);

//-----cevapları karşılaştır-----
for i:=0 to adoquery1.recordcount-1 do begin
  for j:=0 to adoquery2.RecordCount-1 do begin
    if adoquery1.( 'soruno' ).asinteger=adoquery2.( 'soruno' ).asinteger then begin
      if adoquery1.( 'd_cevap' ).asstring=adoquery2.( 'cevap' ).asstring then begin
        adoquery2.FieldValues['sonuc']:=1;
        dogru:=dogru+1;
      end
    else
      begin
        adoquery2.FieldValues['sonuc']:=0;
        yanlis:=yanlis+1;
      end;
    end;
    adoquery2.Next;
  end;
  oran:=copy(floattostr((dogru*100)/adoquery3.recordcount),1,4);
  .
  .
```



Şekil 5.8. Sınav Sonuçları.

5.3.1.2.2. Girişlere göre sınav değerlendirme

Girişlere göre sınav değerlendirme bölümünde seçilen sınavın sorularına verilen cevapların doğruluk oranlarıyla, sorunun ait olduğu konuya yapılan giriş sayılarını ve kalma sürelerinin ilişkisini inceler.

Şekil 5.9. ekranındaki açılır liste kutusundan sınav seçildiğinde seçilen sınavın soru numaraları ve soruların ilişkili olduğu konu numaraları listelenmektedir.

Örneğin 5 nolu soruya 18 kişi doğru yanıt vermiş, toplam öğrenci sayısı içerisinde genel oranı %18 olarak hesaplanmıştır.

Giriş ve Süre Oranlarını Hesapla butonu tıklanarak konulara giriş sayılarının ve kalma sürelerinin oranı hesaplanır.

Soruların Doğruluk Oranlarını Bul butonu yardımıyla, sorulara verilen cevapların doğruluk oranları hesaplanır.

Başarı Oranını Hesapla butonu ise, sınavın genel başarı ortalamasını bulur.

Listeden Bölüm1 sınavı seçildiğinde Şekil 5.8.'deki sonuçlar elde edilmiştir.

| Soru No | Konu No | Giris Oranı | Süre Oranı | Doğruluk Oranı |
|---------|---------|-------------|------------|----------------|
| 1 | 3 | %3.63 | %3.59 | %79 |
| 2 | 4 | %4.91 | %4.84 | %83 |
| 3 | 6 | %3.03 | %3.43 | %62 |
| 4 | 7 | %0.14 | %0.17 | %5 |
| 5 | 9 | %0.43 | %0.61 | %18 |
| 6 | 8 | %4.89 | %4.65 | %76 |
| 7 | 8 | %4.89 | %4.65 | %83 |
| 8 | 1 | %4.26 | %3.91 | %76 |
| 9 | 2 | %3.59 | %3.22 | %64 |
| 10 | 1 | %4.26 | %3.91 | %37 |

Başarı Ortalaması: 68.7

Şekil 5.9. Girişlere Göre Sınav Değerlendirmesi.

Şekil 5.9' da elde edilen sonuçlarda görüldüğü gibi, 4 numaralı konuyla ilgili olan 2. soru %83 doğruluk oranıyla en fazla doğru cevaplanan sorudur. Bu konunun anlatıldığı sayfaya yapılan girişlerin oranı (%4,91) ve sayfada kalınan sürenin oranı (%84) da diğerlerinden fazla olduğu görülmektedir.

5.3.1.2.3. Bölümlere göre sınav değerlendirme

Öğrencilerin öğrenimlerine devam ettikleri bölümlere göre sınav ortalamalarını hesaplar. Form açıldığında bölüm isimleri listelenmiş olarak gelir.

Şekil 5.10'daki elde edilen sonuçlar incelendiğinde Seracılık bölümünün 1.Sınav ortalamasının 81, 2.Sınav ortalamasının 73.2, 3.Sınav ortalamasının 71, 4.Sınav ortalamasının 22.7, 5.Sınav ortalamasının 31.2 olduğu görülmektedir. Bu da dönem sonuna doğru bölümün başarısının düştüğünü göstermektedir.

| BÖLÜMLERE GÖRE QUIZ DEĞERLENDİRMESİ | | | | | |
|-------------------------------------|--------|--------|--------|--------|--------|
| Ortalamaları Hesapla | | KAPAT | | | |
| Bölüm Adı | 1.Quiz | 2.Quiz | 3.Quiz | 4.Quiz | 5.Quiz |
| Hasat Sonrası Teknolojisi | 65,8 | 58,8 | 37,4 | 41,6 | 47,6 |
| Mantarçılık | 70 | 64,3 | 66,6 | 49 | 56,8 |
| Fidan ve Fidecilik | 72 | 59,4 | 53,9 | 39 | 57,2 |
| Seraçılık | 81 | 73,2 | 71 | 22,7 | 31,2 |
| Bitki Koruma | 82,1 | 82,1 | 80,8 | 62,3 | 70,6 |
| Hemşirelik | 66,4 | 66,0 | 58,0 | 59,5 | 50,8 |
| Ebelik | 62,1 | 59,6 | 53,3 | 31,6 | 36,6 |

Şekil 5.10. Bölümlere Göre Sınav Değerlendirmesi.

Şekil 5.10'da Ortalama Hesapla butonu tıklandığında Tablo 5.7.'deki kodlarla her bölümün beş sınav ortalaması hesaplanır.

Tablo 5.7. Departmanlara Göre Ortalama Hesaplama.

```

for j:=1 to 5 do
  for i:=1 to 7 do
    begin
      adoquery1.Close;
      adoquery1.SQL.Clear;
      sql:='select count(kullanici), quizno, sum(notu),sum(notu)/count(kullanici) as
        ortalama from quiznotu where quizno='+inttostr(j);
      sql:=sql+' and kullanici in (select kid from ogrenci where dept='+inttostr(i)+'
        group by quizno';
      adoquery1.SQL.Add(sql);
      adoquery1.Open;
    End;
  
```

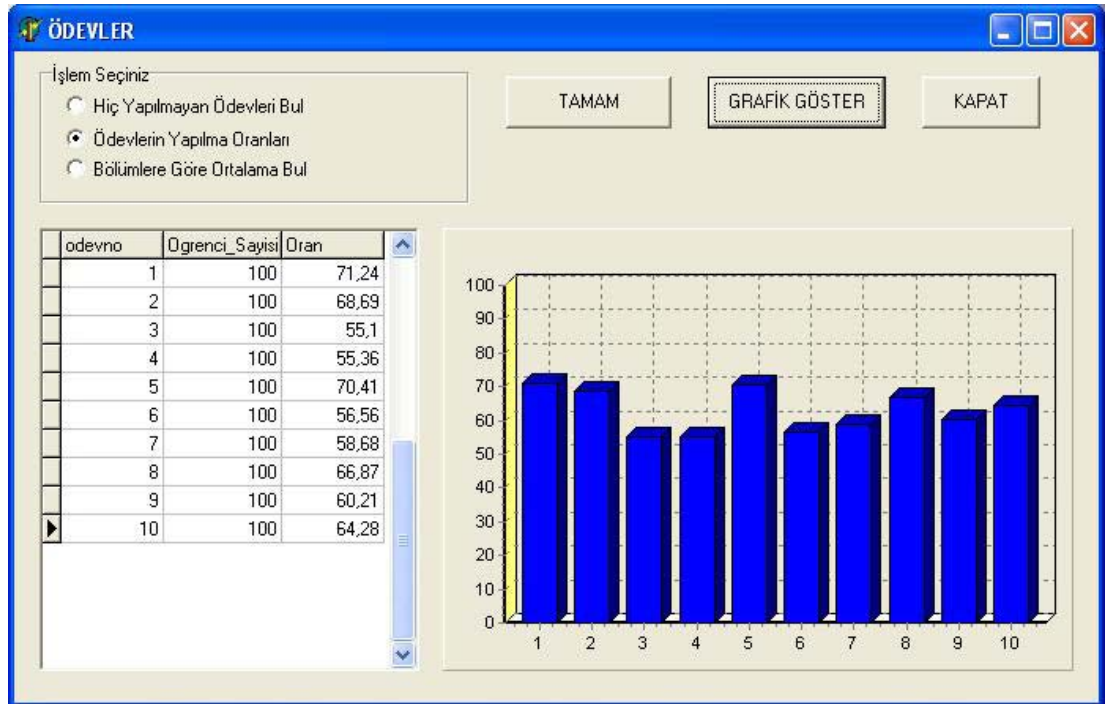

5.3.1.3. Ödevler

Şekil 5.3'deki Genel Değerlendirme menüsünün altında bulunmaktadır. Ödevler ekranında hiç yapılmayan ödevler, ödevlerin yapılma oranları ve bölümlere göre ödev ortalaması olmak üzere üç işlem yapılabilir.

Hiç Yapılmayan Ödevler seçeneği seçildiğinde, dönem içerisinde teslim edilmeyen ödevleri verir.

Ödevlerin Yapılma Oranları seçildiğinde, öğrencilerin ödevlerden aldığı notların ortalamasını hesaplar.

Bölmelere Göre Ödev Ortalaması ise, öğrencilerin öğrenim gördükleri bölümlere göre ödevlerden aldıkları notların ortalamasını bulur.



Şekil 5.11. Ödevler.

Şekil 5.11'deki sonuçlarda ödevlerin ortalamaları görülmektedir. En yüksek ortalama 71.24'lük oranla 1. ödeve aittir.

Tamam butonu tıklandığında; Tablo 5.7.'deki yapılan seçime uygun olan SQL komutları çalışır.

Tablo 5.8. Ödevlerle İlgili İşlemleri Yapan SQL Komutları.

```
if radiobutton1.Checked then  ///Hiç yapılmayan ödevler
    sql:='select odevno,bolumno from odevler where odevno not
        in(1,2,3,4,5,6,7,8,9,10)';
if radiobutton2.Checked then  ///Ödevlerin yapılma oranları
    sql:='select odevno, count(ogrid) as Ogrenci_Sayisi, (sum(notu)/100) as Oran
        from odevler group by odevno';
if radiobutton3.Checked then  ///Bölgümlere göre ödev ortalamaları
    begin
        sql:='select bolumno as Bolum_No, avg(notu) as Odv_Ort from odevler ';
        sql:=sql+' group by bolumno' ;
    end;
adoquery1.Close;
adoquery1.SQL.Clear;
adoquery1.SQL.add(sql);
adoquery1.Open;
```

5.3.2. Birliktelik kuralı çıkarma

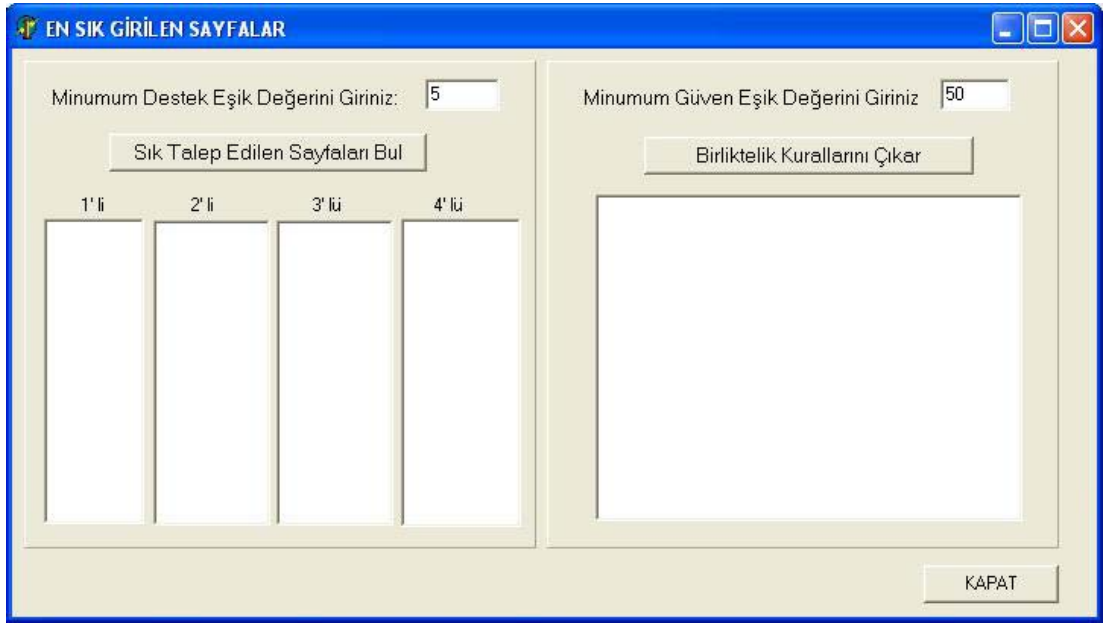
Şekil 5.3'de gösterilmektedir. Birliktelik Kuralı Çıkarma menüsünde, öğrencinin herhangi bir oturumunda birlikte girdiği sayfaları ve bir sınavda aynı anda yanlış cevapladığı soruları bulup birliktelik kuralı çıkarmayı amaçlar. Şekil 5.12'de Birliktelik Kuralı Çıkarma menüsü görülmektedir.



Şekil 5.12. Birliktelik Kuralı Çıkarma Menüsü

5.3.2.1. Birlikte sık girilen sayfalar

Öğrenciyi müşteri, sayfaları ürün, sayfa taleplerini alışveriş gibi düşünersek, Apriori algoritmasını kullanarak; öğrencilerin oturumlarında hangi sayfalara birlikte kaç kez girildiğini bulup birliktelik kuralı çıkarmak mümkündür. Şekil 5.13’de işlemleri yapıldığı ara yüz ekranı görünmektedir.



Şekil 5.13. Birlikte En Sık Girilen Sayfalar.

Sık Talep Edilen Sayfaları Bul, butonu tıklandığında yapılan işlem basamakları şöyledir:

- Tablo 5.9’deki SQL komutu ile minimum destek eşik değerini geçen sayfalar bulunur.

Tablo 5.9. Minimum Destek Eşik Değerini Geçen Sayfaları Bulan SQL Komutu.

```
'select konu as Konular, count(girissayisi) as Giriş_Sayisi  
from giris group by konu having count(girissayisi)>=' + edit1.text
```

- Tablo 5.9'daki sorgudan dönen sayfa numaraları ve destek değerleri bir döngü ile hem *kume* isimli dizi değişkenine kaydedilir hem de *listbox1*'e yazdırılır.
- Tablo 5.10'daki komutlarla *kume* dizisindeki elemanların ikili kombinasyonları oluşturulur. Bu ikili sayfa numaraları SQL ifadesi ile veritabanı taranır. Minimum destek eşik değerini geçen ikili sık girilen sayfalar *kume2* dizisine kaydedilir ve *listbox2*'de görüntülenir.

Tablo 5.10. Sık Girilen Sayfaların İkili Kombinasyonlarının Oluşturulması ve Veritabanında Taranması.

```

for i:=0 to adoquery1.RecordCount do  ///ikili konbinasyonları oluştur
  for j:=i+1 to adoquery1.RecordCount-1 do
    begin  /// veritabanını tara
      sqlstr2:='select a.konu,a.kullanici,a.oturum from giris a where
        a.konu='+inttostr(kume[i,0])+' and a.kullanici in
        (select b.kullanici from giris b where konu='+inttostr(kume[j,0])+')
        and a.oturum in (select b.oturum from giris b where
        b.kullanici=a.kullanici and b.konu='+inttostr(kume[j,0])+')';
      adoquery2.SQL.Add(sqlstr2);
      /// minimum destek eşik değerini geçenleri kume2 dizisine at
      if adoquery2.RecordCount>=strtoint(edit1.text) then
        begin
          listBox2.Items.Add(inttostr(kume[i,0])+'-'+inttostr(kume[j,0])+
            '+inttostr(adoquery2.recordcount));
          kume2[x,0]:=kume[i,0];
          kume2[x,1]:=kume[j,0];
          kume2[x,2]:=adoquery2.RecordCount;
          kume2[x,3]:=kume[i,1];
          kume2[x,4]:=kume[j,1];
          x:=x+1;
        end ;
    end;
end;

```

- *kume2* dizisindeki eleman sayısı ikiden fazla ise, *kume2* deki elemanlar kullanılarak sayfa numaralarının üçlü kombinasyonları oluşturulur. Oluşturulan bu üçlü öğeler veritabanında taranır ve minimum destek eşik değerini geçen öğeler *kume3* dizisine aktarılır.

- Apriori algoritmasına göre sık tekrarlanan bir ögenin alt kümeleri de sık tekrarlanan öge olmalıdır. Bu yüzden *kume3* dizisinin elemanlarının alt kümeleri de araştırılır. Tablo 5.11'deki SQL komutu ile veritabanı taranır.

Tablo 5.11. Üçlü Öğeleri Veritabanında Arayan SQL Komutu.

```
'select a.konu,a.kullanici,a.oturum from giris a where
a.konu=' + inttostr(kume2[i,0]) + ' and a.kullanici in
(select b.kullanici from giris b where b.konu= '+ inttostr(kume2[i,1]) +
' and b.kullanici in (select c.kullanici from giris c where
c.konu=' + inttostr(oge_3) + ')
and a.oturum in(select b.oturum from giris b
where b.kullanici=a.kullanici and b.konu=' + inttostr(kume2[i,1]) +
' and b.oturum in (select c.oturum from giris c where
b.kullanici=c.kullanici and c.konu=' + inttostr(oge_3) + ')'
```

- Bu şekilde sık tekrarlanan öğeler bulunarak Apriori Algoritmasının 1. aşaması tamamlanır. Minimum Destek Eşik Değeri=50 için algoritmanın ürettiği sonuçlar Şekil 5.14'de görülmektedir.

Minumum Destek Eşik Değerini Giriniz:

Sık Talep Edilen Sayfaları Bul

| Tekli | | İkili | | Üçlü | | Dörtlü | |
|--------|----------|---------|----------|----------|----------|--------|----------|
| [S.No] | [Destek] | [S.No] | [Destek] | [S.No] | [Destek] | [S.No] | [Destek] |
| 1 | 70 | 1-3 | 69 | 1-3-4 | 66 | | |
| 2 | 65 | 1-4 | 67 | 1-3-6 | 50 | | |
| 3 | 79 | 1-6 | 50 | 1-3-8 | 61 | | |
| 4 | 87 | 1-8 | 62 | 1-4-8 | 62 | | |
| 6 | 62 | 2-4 | 53 | 2-4-8 | 51 | | |
| 8 | 82 | 2-8 | 51 | 3-4-6 | 58 | | |
| 10 | 76 | 3-4 | 76 | 3-4-8 | 71 | | |
| 17 | 58 | 3-6 | 61 | 3-6-8 | 53 | | |
| 18 | 60 | 3-8 | 71 | 38-43-44 | 58 | | |
| 20 | 70 | 4-8 | 82 | | | | |
| 22 | 64 | 17-35 | 57 | | | | |
| 30 | 60 | 30-33 | 60 | | | | |
| 33 | 67 | 38-43 | 58 | | | | |
| 35 | 57 | 38-44 | 58 | | | | |
| 38 | 59 | 41-42 | 65 | | | | |
| 41 | 66 | 43-44 | 61 | | | | |
| 42 | 65 | 110-119 | 75 | | | | |
| 43 | 62 | | | | | | |
| 44 | 63 | | | | | | |
| 46 | 71 | | | | | | |
| 49 | 75 | | | | | | |

Şekil 5.14. Sık Tekrarlanan Öğeler.

Şekil 5.12'deki elde sonuçlarda görüldüğü gibi 9 adet üçlü sık tekrarlanan öge bulunmuştur. Örneğin; 66 öğrenci 1, 3, 4 numaralı sayfalara aynı oturumda girmiştir. Şekil 5.12'deki Birliktelik Kuralı Çıkar butonu tıklandığında birliktelik kuralı çıkarma işlemleri yapılır. Birliktelik kuralı çıkarmak için sık tekrarlanan öğelerin güven değerleri hesaplanır. Minimum güven eşik değerini geçen kurallar birliktelik kuralı olarak kabul edilir. I sık tekrarlanan bir öge kümesi, a bir alt küme olsun.

$$(I-a) \rightarrow a \text{ kuralı için güven} = \frac{I}{(I-a)} \text{ formülüyle hesaplanır.}$$

Şekil 5.14'deki sık tekrarlanan öğelerden; $I=\{38-43-44\}$ un güven değerleri hesaplandığında aşağıdaki kurallar elde edilir.

- 1) 38-44→43 = 58/58 = %100
- 2) 43-44→38 = 58/61 = %95
- 3) 38-43→44 = 58/58 = %100
- 4) 38→43-44 = 58/59 = %98,3
- 5) 43→38-44 = 58/62 = %93,5
- 6) 44→38-43 = 58/63 = %92,5

Eğer minimum güven eşik değeri %100 kabul edilirse yukarıdaki kurallardan 1. ve 2. kurallar dikkate alınır. Üretilen bütün kurallar Şekil 5.15'de görülmektedir.

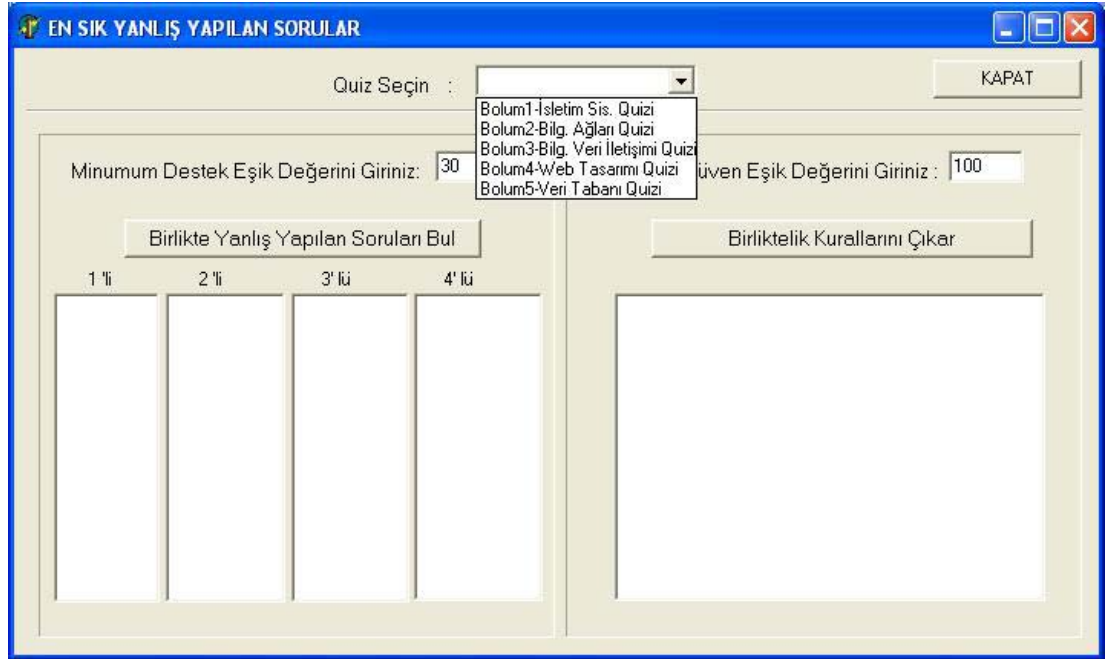
| [Kural] | [Güven] |
|-------------|---------|
| 8--->4 | ≈100 |
| 35--->17 | ≈100 |
| 30--->33 | ≈100 |
| 42--->41 | ≈100 |
| 1-6--->3 | ≈100 |
| 1-8--->4 | ≈100 |
| 2-8--->4 | ≈100 |
| 3-8--->4 | ≈100 |
| 38-44--->43 | ≈100 |
| 38-43--->44 | ≈100 |

Şekil 5.15. Üretilen Birliktelik Kuralları.

Şekil 5.15'deki elde edilen kurallardan örneğin 5. kurala göre, 1 ve 6 numaralı sayfalara giren bir öğrenci aynı oturumda kesinlikle 3 numaralı sayfaya da girecektir

5.3.2.2. Birlikte yanlış yapılan sorular

Öğrencilerin seçilen sınavda hangi soruları birlikte yanlış yaptığını bulan işlem seçeneğidir.



Şekil 5.16. Birlikte Yanlış Yapılan Sorular.

Şekil 5.16'deki açılır liste kutusundan inceleme yapacağımız sınav seçilir. Bir sınav seçildiğinde o sınavın sorularına verilen cevaplar taranır. Veri tabanı Tablo5.12'deki SQL komutu ile taranır. Yanlış yapılan soruların sayısı bulunur ve minimum destek eşik değerini geçenler sık tekrarlanan 1'li öge kümesini oluşturur.

Tablo 5.12. Birlikte Yanlış Yapılan Birli Öge Kümesini Bulan SQL Komutu.

```
'select soruno,count(sonuc) as Yanlis_Sayisi from cevap where sonuc=0 and  
quizno=' + inttostr(secim) + ' group by soruno having  
count(soruno)>='+edit1.text;
```

SQL den dönen değer *kume* dizisine kaydedilir. İkili kombinasyonlar oluşturulur ve oluşan ikili öğeler Tablo 5.13'deki SQL komutu ile veritabanında taranır.

Tablo 5.13. İkili Öğeleri Bulan SQL Komutu.

```
'select a.kullanici,a.soruno,a.sonuc from cevap a where a.sonuc=0 and
a.soruno=' + inttostr(kume[i,0]) + 'and quizno=' + inttostr(secim) +
' and a.kullanici in (select b.kullanici from cevap b where b.sonuc=0
and b.soruno=' + inttostr(kume[j,0]) + ' and quizno=' + inttostr(secim)+')
```

Tablo 5.14. Üçlü Öğeleri Bulan SQL Komutu.

```
'select a.kullanici,a.soruno,a.sonuc from cevap a where a.sonuc=0 and
a.soruno='+inttostr(kume2[i,0])+'and quizno=' + inttostr(secim)
+' and a.kullanici in (select b.kullanici from cevap b where b.sonuc=0 and
b.soruno='+inttostr(kume2[i,1])+' and quizno=' + inttostr(secim)
+' and b.kullanici in (select c.kullanici from cevap c where c.sonuc=0 and
c.soruno='+inttostr(oge_3)+' and quizno='+ inttostr(secim)+')
```

Bölüm1-İşletim Sis. Sınav'ı için minimum destek eşik değeri 50, minimum güven eşik değeri 100 olarak verildiğinde algoritma Şekil 5.17'deki sonuçları üretmiştir.

| 1'li | 2'li | 3'lü | 4'lü |
|--------------|--------------|--------------|--------------|
| [Soru No][D] | [Soru No][D] | [Soru No][D] | [Soru No][D] |
| 4 95 | 4-5 82 | 4-5-10 61 | |
| 5 82 | 4-10 62 | | |
| 10 63 | 5-10 61 | | |

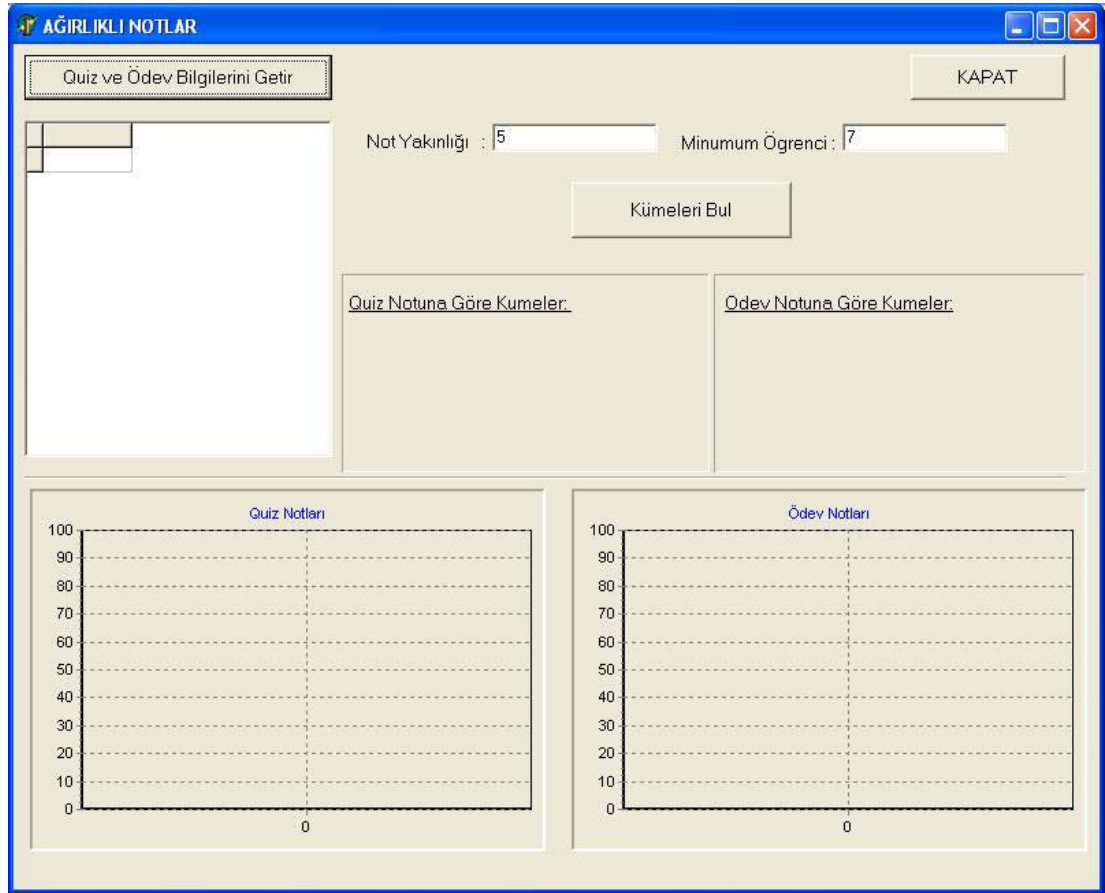
| [Kural] | [Güven] |
|----------|---------|
| 5-->4 | %100 |
| 5-10-->4 | %100 |

Şekil 5.17. Birlikte Yanlış Yapılan Sorular.

Şekil 5.17'deki elde edilen birliktelik sonuçlarına göre, 4., 5. ve 10. soruların üçünü birden 61 kişi yanlış cevaplamış. Bu birliktelikten üretilen kurala göre, 5. ve 10. soruyu yanlış yapan bir öğrenci kesinlikle 4. soruya da yanlış cevap verecektir.

5.3.3. Ağırlıklı notlar

Şekil 5.3'de görülen, Ağırlıklı Notlar menüsünde öğrencilerin ödev ve sınav notlarının hangi not aralığında yoğunlukta olduğunu bulmak amaçlanmıştır. Verilerimiz doğrusal olduğu için küresel olmayan kümeleri başarıyla bulabilen DBSCAN algoritması kullanılmıştır. Şekil 5.18'de işlemlerin yapıldığı arayüz ekranı görülmektedir.



Şekil 5.18. Ağırlıklı Notlar Arayüz Ekranı.

Sınav ve Ödev Bilgilerini Getir butonu tıklandığında, öğrencilerin ödev ve sınav ortalamaları SQL komutuyla hesaplanıp *kume* ve *kume2* dizilerine kaydedilir. Başlangıçta bütün verilerin küme numaraları “0” olarak verilir.

Not Yakınlığı ve Minimum Öğrenci parametreleri kullanılarak oluşabilecek farklı kümeler bulunur. Her bir veri sırayla ele alınır. Küme numarasına bakılıp işlem yapılır. Daha önce küme numarası verilmişse başka bir veri incelenir.

İncelemeye alınan veri küme merkezi kabul edilir. Not yakınlığı parametresi ile küme sınırları belirlenir. Örneğin ele alınan veri 60, not yakınlığı 5 ise kümenin alt sınırı 55 üst sınırı 65 kabul edilir. Belirlenen bu sınırlar arasındaki veriler bulunur. Eğer bulunan veri sayısı Minimum Öğrenci değerine eşit veya daha fazlaysa bu verilerden bir küme oluşturulur ve bir küme numarası verilir. DBSCAN algoritmasına göre bir küme oluşturan öğelerin de yakın değerleri araştırılır ve herhangi bir kümeye dahil olmayan elemanlar bu kümeye dahil edilir. İşlemleri yapan kodlar Tablo 5.15’de kısaltılmış haliyle verilmektedir.

Tablo 5.15. DBSCAN Algoritmasını Çalıştıran Kodlar.

```
for i:=0 to toplam do kume[i,2]:=0; ///Sınav verilerinin küme numaralarını sıfırla

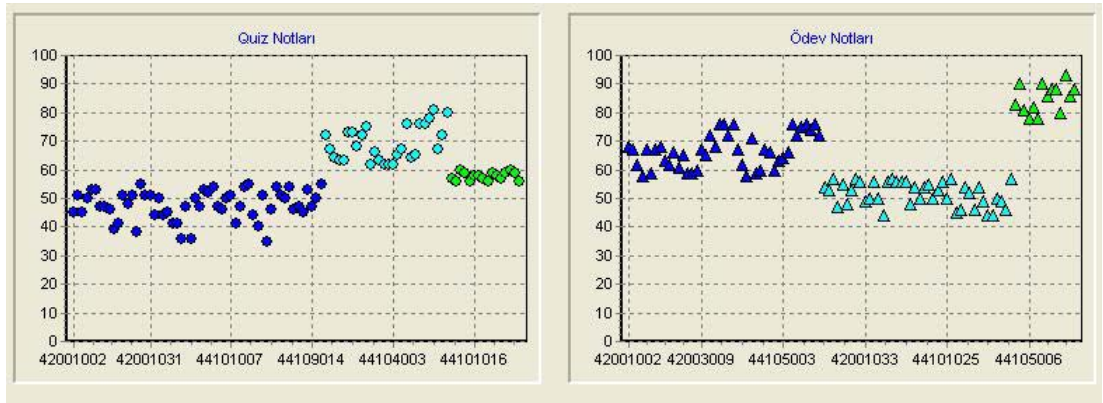
for i:=0 to toplam do kume2[i,2]:=0; /// Ödev verilerinin küme numaralarını sıfırla

for i:=0 to toplam do //bir veri(sınav notu) al
  if kume[i,2]=0 then // herhangi bir kumeye dahil değilse işlem yap
    begin
      //alınan veriye göre kumenin alt ve üst sınırını belirle
      alt:=kume[i,1]-strtoint(edit1.Text);
      ust:=kume[i,1]+strtoint(edit1.Text);
      for j:=0 to toplam do //bu aralıktaki notları bul
        if (kume[j,1]>=alt) and (kume[j,1]<=ust) and (kume[j,2]=0) then
          begin
            gecici_kume[ogr,0]:=kume[j,0];
            gecici_kume[ogr,1]:=kume[j,1];
            ogr:=ogr+1;
          end;
      if ogr>=strtoint(edit2.text) then //eger minimum öğrenci sayısını karşılıyorsa
        begin
          for f:=0 to ogr-1 do
            for c:=0 to toplam do
              if (gecici_kume[f,0]=kume[c,0]) then
                begin
                  kume[c,2]:=k; //bulunan sınav notuna kume numarası ver
                  // bu nota yakın notları da bul
```

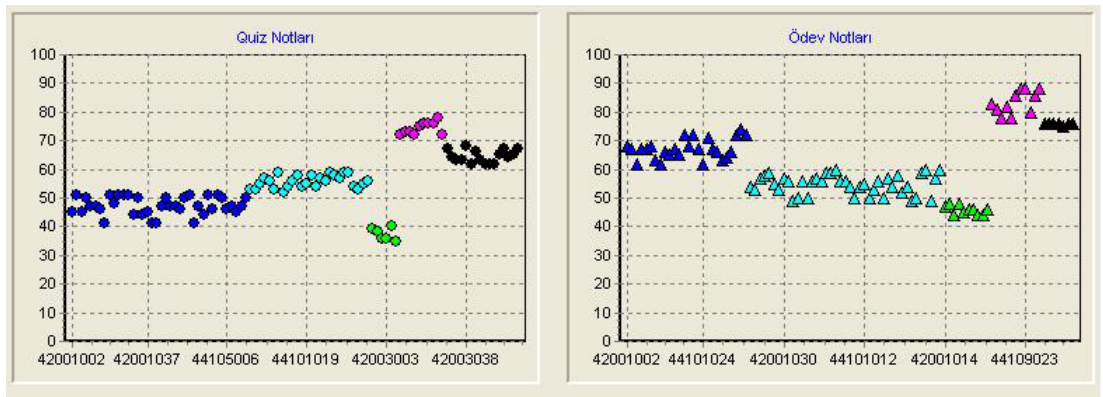
Tablo 5.1(devam). DBSCAN Algoritmasını Çalıştıran Kodlar.

```
alt1:=kume[c,1]-strtoint(edit1.text);
ust1:=kume[c,1]+strtoint(edit1.text);
for x:=0 to toplam do
  if ((kume[x,1]>=alt1) and (kume[x,1]<=ust1)) then
    begin
      gecici_1[R,0]:=kume[x,0];  gecici_1[R,1]:=kume[x,1];  R:=R+1;
    end;
  if R>=strtoint(Edit2.text) then
    for y:=0 to R-1 do
      for z:=0 to toplam do
        if (kume[z,0]=gecici_1[y,0]) and (kume[z,2]=0) then
          kume[z,2]:=k;
        :
        :
```

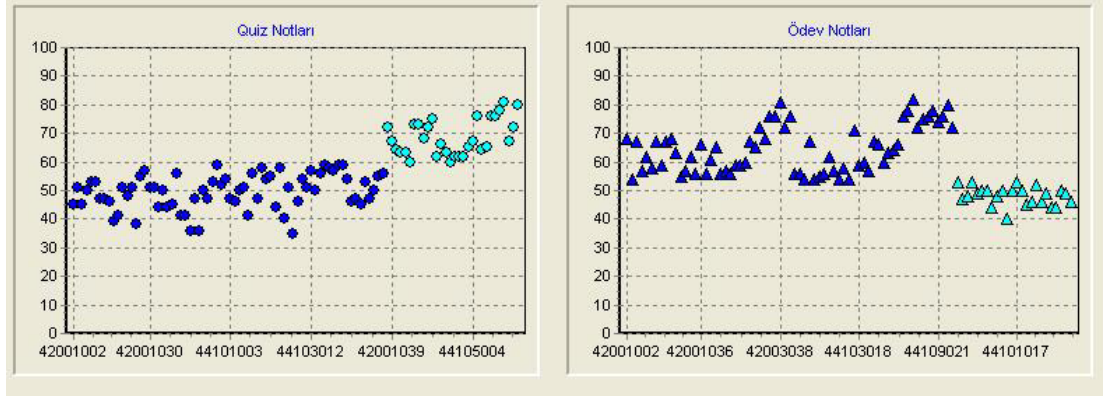
Farklı parametrelere göre oluşan kümeler:



Şekil 5.19. Not Yakınlığı=5 ve Minimum Öğrenci=7.



Şekil 5.20. Not Yakınlığı=3 ve Minimum Öğrenci=5.



Şekil 5.21. Not Yakınlığı=7 ve Minimum Öğrenci=10.

Bu farklı parametrelerle oluşan kümeler incelendiğinde Not Yakınlığı değeri arttıkça oluşan küme sayısı azalmakta ve oluşan kümelerin boyutu büyümektedir. Minimum Öğrenci değeri arttıkça kümelerin hacmi daralmakta ve yalnızca yoğun iç bölgeleri tespit edilebilmekte, azaldıkça ise az yoğun olan gürültü alanları dahi küme olarak nitelendirilmektedir.

SONUÇLAR ve ÖNERİLER

Günümüz dünyasında, insanlar alışverişlerde, bankacılık işlemlerinde işletmelerin günlük işlemlerinde, transfer edilen verilerin sayısı her geçen gün artmaktadır. Veri sayısının inanılmaz boyutlarda artış problemi, bilgisayar sistemlerindeki teknolojinin her geçen gün büyümesiyle aşılmaktadır. İşlemcilerin gittikçe hızlanması, bilgisayarların daha büyük miktarlardaki veriyi saklamasına olanak tanımaktadır. Bunun yanında bilgisayar ağlarındaki ilerleme ve bu veriye başka bilgisayarlardan da hızla ulaşabilmek olanaklı duruma gelmiştir.

Veri sayısının büyük miktarlara ulaşması, bu verilerden anlamlı bilgi çıkarabilmek amacıyla kullanılan istatistik yöntemleri yetersiz duruma gelmiştir. Bu amaçla veri madenciliği ve bilgi keşfi, büyük miktarlardaki veri içerisinde anlamlı sonuçlar çıkarmada kullanılmaktadır.

Veri madenciliğinde amaç, kolaylıkla mantıksal kurallara ve görsel sunumlara çevrilebilecek modellerin çıkarılmasıdır. Veri madenciliği, verilerin içindeki örüntüyü, ilişkiyi ve kuralları yarı otomatik olarak keşfetmektedir.

Yapılan bu çalışmada, internet tabanlı bir öğretim sitesinde alınan veriler üzerinde birliktelik kuralları madenciliği ve kümeleme tekniklerinden DBSCAN algoritması kullanılarak öğrenci performansı değerlendirilmiştir. Uygulamada yapılan siteye öğrencilerin giriş yapması sağlanmıştır. Site içerisindeki öğrenci hareketleri bir veritabanında toplanmıştır. Uygulamada öğrencilerin siteye girişleri ve değerlendirme sınavlarına verdiği yanıtlar, birliktelik kuralları algoritmalarından apriori algoritması ile analiz edilerek, en sık girilen sayfa birliktelikleri ve en sık yanlış yapılan soru birliktelikleri ortaya çıkarılmıştır. Oluşturulan sayfa birliktelikleri sitenin güncelleştirilmesinde kullanılarak sitenin yenilenmesi sağlanmıştır. Ayrıca, öğrenci ödev ve sınav notları, kümeleme tekniklerinden DBSCAN algoritması kullanılarak, öğrenciler gruplandırılmıştır. Öğrenciler gruplandırılırken, kümeleme,

ödev ve sınav not yakınlığına ve gruba girecek minimum öğrenci sayısına göre yapılmıştır. En uygun kümelenebilirliği bulmak için not yakınlığı ve minimum öğrenci sayısı değerleri değiştirilerek birden fazla küme oluşturulmuştur.

Oluşturulan bu yapı sayesinde, internet tabanlı öğretim uygulamasına katılan öğrencilerin başarılarının değerlendirilmesi, dersin verimli ve yararlı olabilmesi için daha kapsamlı tasarım aşamalarına geçilmesi ve yürütülmesi için, birliktelik kuralları keşfedilmektedir. Keşfedilen bu kurallar gereksinimlere göre farklı amaçlar için kullanılabilirler.

Veri madenciliği modellerinden, kümeleme modelinin yoğunluk tabanlı yöntemleri içerisinde yer alan, DBSCAN algoritması kullanılıp öğrenci grupları keşfedilerek öğrenci başarısını artırıcı çalışmalar yapılabilir. Eldeki veri sayısını artırılarak, DBSCAN algoritması uygulamasında kullanılan MinPts (Minimum öğrenci sayısı) ve Eps (Not Yakınlığı) parametreleri ile kümeler daha net bir şekilde çıkarılabilir. Eps değeri arttıkça, küme bulma hassasiyeti azalmakta, tüm veritabanını tek bir küme gibi görme eğilimi ağırlık kazanmaktadır. MinPts değeri arttıkça kümelerin hacmi daralmakta ve yalnızca yoğun iç bölgeler tespit edilerek en uygun kümelemeyi bulmak mümkün olmaktadır.

Yapılan çalışmada internet tabanlı öğretimde öğrenci hareketlerinden oluşan veri tabanı üzerinde keşfedilen birliktelikler ve yoğunluk tabanlı öğrenci kümeleri özellikleri değerlendirilerek, site tasarımı daha da geliştirilebilir, ve bu şekilde öğrenci ve öğretim elemanı performansını artırıcı yönde kullanılabilir.

Uygulamada, internet tabanlı bir öğretim sitesi tasarlamak için ColdFusion programı kullanılmıştır. Tasarlanan site Kocaeli Üniversitesi sunucusunda yayınlanmıştır. Öğrencilerin siteye giriş yapması sağlanmış, öğrenci hareketleri Microsoft Access veri tabanına kaydedilmiştir. Delphi 6.0 programlama dili ile bir ara yüz programı geliştirilerek, veri madenciliği teknikleri Microsoft Access veritabanındaki veriler üzerinde uygulanmıştır.

Uygulamada kullanılan Apriori ve DBSCAN algoritmalarından elde edilen sonuçlarda, öğrencilerin sayfalara giriş sayılarının ve sayfalarda kalma sürelerinin, sınavlarda ki başarı oranını etkilediği izlenmiştir. Aynı şekilde giriş ve süre oranlarının yeterli olup da başarı oranlarının düşük olduğu durumlar gözlenmiştir. Bu da tasarlanan sitedeki ilgili sayfanın iyileştirilmesi gerektiği sonucunu ortaya çıkarmaktadır. Aynı zamanda öğrencilerin kişisel bilgilerine göre, sitedeki sayfa hareketleri incelenmiştir. Örneğin yurttan kalan öğrencilerin, diğerlerine oranla sayfaya giriş oranlarının daha az sayıda olduğu gözlenmiştir.

Sonuç olarak, veri madenciliği tekniklerinin, İnternet tabanlı öğretimin geliştirilmesinde kullanılabileceği görülmüştür. Ancak mevcut öğrenci sayısı ve veri miktarı veri madenciliği tekniklerinin uygulamasında önemli birer parametre olduğundan, sonraki çalışmada önceden oluşturulan yoğun bir bilgi bankası ile veri miktarı artırılabilir. Böylece keşfedilen kuralların ve oluşturulan yoğunluk tabanlı grupları güvenilirliğinin de artacağı öngörülebilir.

KAYNAKLAR

- [1] Bilgin, T.T., “Veri Madenciliğinde Kümeleme Yöntemi”, Yüksek Lisans Tezi, **Marmara Üniversitesi Fen Bilimleri Enstitüsü**, İstanbul, (2003).
- [2] Özakar, B., Püskülcü, H., “Tıkların Dili: Web İçerik ve Web Kullanım Madenciliği Tekniklerinin Entegrasyonu ile Oluşmuş bir Veritabanından Nasıl Yararlanılabilir?”, <http://www.teknoturk.org>, (**Ziyaret Tarihi: 05.06.2003**).
- [3] Etzioni, O., “The World Wide Web: Quagmire or Goldmine”, **Communications of the ACM**, 65-68, 39(11), (1996).
- [4] Bidgoli, B.M., Deborah, A., “Predicting Student Performance: An Application of Data Mining Methods with An Educational Web-Based System”, **ASEE / IEEE**, 13-18, 0-7803-7961-6/03 (2003).
- [5] Chiang, R.H.L., Cecil, C.E.H., Lim, E. “Linear Correlation Discovery in Databases A Data Mining Approach”, **Elsevier**, 311-337, 0169-023X, (2004).
- [6] Tang, T., McCalla, G., “Mining Implicit Ratings for Focused Collaborative Filtering for Paper Recommendations”, **Naturel Sciences and Engineering Research Council of Canada**, 1-12, (2004)
- [7] Luan, J., Ph, D., “Data Mining and Knowledge Management in Higher Education”, **Presentation at AIR Forum, Toronto, Canada**, 1-17, (2002).
- [8] Bidgoli, B.M., Punch, W.F., “Using Genetic Algorithms for Data Mining Optimization in an Educational Web-Based System”, **Genetic Algorithms Research and Applications Group (Garage)**, (2003)
- [9] Kuo, R.J., Lio, J.L., Tu, C., “Integration of ART2 Neural Network and Genetic K-Means Algorithm for Analyzing Web Browsing Paths in Electronic Commerce”, **Elsevier**, 355-374, 0167-9236 doi:10.1016/j.dss, (2004).
- [10] Zaiane, O.R., “Web Usage Mining for a Better Web-Based Learning Environment”, **IEEE International Conference on Advanced Learning Technologies**, WI, USA, 6-8 August (2001)
- [11] Jicheng, W., Yuan, H., Gangshan, W., Fuyan, Z., “Web Mining: Knowledge Discovery on the Web”, **IEEE**, II-137-141, 0-7803-5731-0/99/, (1999).
- [12] Tang, T.Y., McCalla, G., “Student Modeling for A Web-Based Learning Environment: A Data Mining Approach”, **American Association for Artificial Intelligence**, (2002).

- [13] He, Z., Xu, X., Huang, J.Z., Deng, S., “Mining class Outliers: Concepts, Algorithms and Applications in CRM”, *Elsevier*, 681-697, 0957-4174 doi:10.1016, (2004).
- [14] Evfimievski, A., Srikant, R., Agrawal, R., Gehrke, J., “Privacy Preserving Mining of Association Rules”, *Sigkdd*, 1-58113-567-X, (2002).
- [15] Merceron, A., Yacef, C., “A Web-Based Tutoring Tool with Mining Facilities to Improve Learning and Teaching”, *Esilv – Pôle Universitaire University of Sydney*, Australia, (2002).
- [16] Biçen, P., “Veri Madenciliği: Sınıflandırma Ve Tahmin Yöntemlerini Kullanarak Bir Uygulama”, Yüksek Lisans Tez, *Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü*, İstanbul, (2002).
- [17] Han, J., Kamber, S.F., “Data Mining: Concepts and Techniques”, *Morgan Kaufmann Publishers*, (2001).
- [18] Madria, S.K., Bhowmick, W.K., Ng, E.P., “Research issues in web data mining, In Proceedings of Data Warehousing and Knowledge Discovery”, *First International Conference, DaWak’99*, 303-312, (1999).
- [19] Agrawal, R., Srikant, R., “Fast Algorithms for Mining Association Rules”, *Proceedings of the 20th VLDB Conference Santiago, Chile*, (1994).
- [20] Fayyad, U.M., Piatesky-Shapiro, G., Smyth, P., Uthurusamy, R., “Advances in Data Mining and Knowledge Discovery”, *AAAI Pres*, USA (1994).
- [21] Hinneburg, A., Keim, D.A., “Clustering Techniques for Large Data Sets from Past to Future”, Powerpoint Slayt, www.ece.northwestern.edu/~harsha/clustering/keim_slides.pdf (*Ziyaret Tarihi: Aralık 2004*).
- [22] Sever, H., Oğuz, B., “Veritabanlarında Bilgi Keşfine Formal Bir Yaklaşım”, (2002)
- [23] Zhang, N., Zhou, L., “Methodologies for knowledge Discovery and Data Mining”, Third Pacific-Asia Conference, *Pak-dd99, China*, April 26-28, (1999).
- [24] Hou, J.F., “Clustering with Obstacle Entities”, Yüksek Lisans Tezi, *Simon Fraser University Computing Science*, Kanada, (1999).
- [25] Tsay, Y.J., Chiang, J.y., “CBAR: An Efficient Method for Mining Association Rules”, *Elsevier*, 99-105, 0950-7051 doi:10.1016/j.knosys, (2004).
- [26] Uğuz, H., Kodaz, H., Çomak, E., Baykan, Ö.K., “Apriori Algoritması Kullanılarak Web Kulolanım Madenciliği Yönteminin Web Log Kayıtlarına Uygulanması”, *IJCI Proceedings of International Conference on Signal Processing, ISNN*, 499-501, 1304-20386, Volume:1, Number:2, (2003).

- [27] Chen, M., Han, J., “Data Mining: An Overview from Database Perspective”, *IEEE Transactions on Knowledge and Data Engineering*, Vol.8 No.6, (1996).
- [28] Zaki, M.J., “Paralel and Distributed Association Mining:A Survey”, *IEEE Concurrency, special issue on paralel Mechanisms for Data Mining*, 14-25, Vol.7, No.5, (1999).
- [29] Srikant, R., Yu, Q., “Mining Association Rules with Item Constraints”, *Proceedings of the Third International Conference of Knowledge Discovery and Data Mining*, 67-73, (1997).
- [30] Ester, M., Kriegel, H.P., Sander, J., “A DEensity Based Algorithm for Discovering Clusters in Large Spatial Databases”, *Conference of Knowledge Discovery and Data Mining (KDD’96), Portland, USA*, 226-231, (1996).
- [31] Mutlu, M.E., Öztürk, M.C., “Internet Üzerinde Bilgisayar Destekli Eğitim Yazılımı Geliştirme ve Sunum Araçlarının Gereksinimleri Karşılama Düzeyleri”, *BTIE’99 Bilişim Teknolojileri Işığında Eğitim 1999*, 13-15 Mayıs (1999).
- [32] Al, U., Mardan, R.O., “Web Tabanlı Uzaktan Eğitim Sistemleri: Sahip Olması Gereken Özellikler ve Standartlar”, *Bilgi Dünyası 5(2,)*1-9, (2004)
- [33] Özbek, M. Kuzucuoğlu, A.E., Gürbüz, A., “Student Assessment in an Intelligent System”, *International XII. Turkish Symposium on Artificial Intelligence and Neural Networks*, TAINN (2003).
- [34] Yiğit, Y., Özden, M.Y., “Web Tabanlı Materyali İçerisinde Internet Üzerinden Görüntü Aktarımı”, *Milli Eğitim Dergisi*, Sayı:150, Mayıs (2001).
- [35] Jea, K.F., Chang, M.Y., Lin, K., “An Efficient and Flexible Algorithm for Online Mining of Large Itemsets”, *Elsevier*, 311-316, 10.1016/j.ipl.2004.09.008, (2004).
- [36] Takcı, H., Soğukpınar, İ., “Kullanıcı Erişim Desenleriyle Saldırı Tespiti”, <http://www.bilmuh.gyte.edu.tr/datamining/files/saldiri-tespiti-h-takci.doc>, (**Ziyaret Tarihi: 15.05.2005**).
- [37] Takcı, H., Soğukpınar, İ., “Saldırı Tespitinde En Yakın K Komuşu Uygulaması”, <http://www.bilmuh.gyte.edu.tr/datamining/files/saldiri-tespiti-h-takci.doc>, (**Ziyaret Tarihi: 10.05.2005**).
- [39] Çakır, H., “Web Destekli Öğretimin Cobol Programlama Dili Dersindeki Öğrenci Başarısına Etkisi”, *2003 Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi*, 44-55, Y.11, (2003).
- [40] Burgu, O.B., http://bm-dergi.emo.org.tr/index.php?option=com_content&task=view&id=34&Itemid=74, (**Ziyaret tarihi: 17 Aralık 2004**).

- [41] Uğuz, H., Kodaz, H., Çomak, E., Baykan, Ö.K., “ Genetik Algoritmalar Kullanılarak Web kullanım Madenciliği Yönteminin Sistem Log Kayıtlarına Uygulanması”, *International XII. Turkish Symposium on Artificial Intelligence and Neural Network*, TAINN (2003).
- [42] Ayanoğlu, M., Mert, K., “Peraknede Sektöründe Veri Madenciliği Vazgeçilmez mi? Alternatifi Crm mi?”, *Yöneylem Araştırması/Endüstri Mühendisliği – XXIV Ulusal Kongresi, Gaziantep*, 15-18 Haziran (2004).
- [43] Fox, A., MAzer, M.S., “A Characterization of Interactive Internet Applications”, <http://citeseer.ist.psu.edu/19812.html>, (*Ziyaret Tarihi:10.052005*).
- [44] Faloutsos, C., Lin, D., “Fast Map: A Fast Algorithm for Indexing, Data Mining and Visualization of Traditional and Multimedia Datasets”, *National Science Foundation*, 1-25, IRI-8958546 ve IRI-9205273, (1996).
- [45] Menon, R., Tong, L.H., Sathiyakeerthi, S., “Analyzing Textual Databases Using Data Mining to Enable Fast Product Development Processes”, *Elsevier*, 171-180, 0951-8320, (2004).
- [46] Two Crows Corporation, “Introduction to Data Mining and Knowledge Discovery”, *Third Edition, Two Crows Corporation*, 1-36, 1-892095-02-5, (1999).
- [47] Özkes, S., “Veri Madenciliği Uygulaması”, Yüksek Lisans Tezi, *Marmara Üniversitesi Fen Bilimleri Enstitüsü*, İstanbul, (2002).
- [48] Ünal, O.O., “Data Mining on the Internet”, Yüksek Lisans Tezi, *Boğaziçi Üniversitesi Fen Bilimleri Enstitüsü*, İstanbul, (2003).
- [49] Koyuncu, F., “Veri Madenciliğinde Apriori Temelli İlişkilendirme Kuralı Algoritmalarının Uygulama ve Karşılaştırılması”, Lisans Tezi, *Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü*, İstanbul, (2004).
- [50] Berry, G.S., Linoff, G.S., “Data Mining Techniques For Marketing, Sales, and Customer Relationship Management”, *Wiley Publishing, Inc.*, (2002).
- [51] Roiger, R.J., Geatz, M.W., “Data Mining”, *Pearson Education.Inc.*, (2003).
- [52] Ulaş, M.A., “Market Basket Analysis for Data Mining”, Yüksek Lisans Tez, *Boğaziçi Üniversitesi Fen Bilimleri Enstitüsü*, İstanbul, (2001).
- [53] Özmen, Ş., 2004 “İş Hayatı Veri Madenciliği ile İstatistik Uygulamaları”, (*Ziyaret Tarihi: 10 Aralık 2004*)
- [54] Doruk, Z. “e-öğrenme Standartlarına Genel Bakış” <http://www.dso.iastate.edu/dept/asc/alearner/advantage.html>, (*Ziyaret Tarihi: 25 Nisan 2005*)

ÖZGEÇMİŞ

1978 yılında Mersin’de doğdu. İlk, orta ve lise öğrenimini Mersin’de tamamladı. 1996 yılında girdiği Kocaeli Üniversitesi Teknik Eğitim Fakültesi Elektronik ve Bilgisayar Eğitimi Bölümü’nden 2001 yılında Bilgisayar Öğretmeni olarak mezun oldu. 2001 yılından beri Kocaeli Üniversitesi Enformatik Bölümü’nde Öğretim Görevlisi olarak görev yapmaktadır.