

OKAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



SPOR VE YAŞAM MERKEZLERİ ÜZERİNE
VERİ MADENCİLİĞİ ÇALIŞMASI

YÜKSEK LİSANS TEZİ

Melih Keskin

Tez Danışmanı: Yrd.Doc.Dr.Pınar Yıldırım

AĞUSTOS 2013

Program: Bilgisayar Mühendisliği

SPOR VE YAŐAM MERKEZLERİ ÜZERİNE
VERİ MADENCİLİĐİ ÇALIŐMASI

YÜKSEK LİSANS TEZİ

MELİH KESKİN

tarafından

OKAN ÜNİVERSİTESİ

Bilgisayar MühendisliĐi Bölümüne

Yüksek Lisans

derecesi Őartını saĐlamak için sunulmuŐtur.

Onaylayan:

Chair
Name
DanıŐman
Yrd.Doç.Dr.Pınar Yıldırım

Member's
Name
Üye
Prof.Dr.Bekir Tefvik Akgün

Member's
Name
Üye
Prof.Dr.ÇoŐkun Sönmez

AĐUSTOS 2013

Program: Bilgisayar MühendisliĐi

DATA MINING ON SPORTS AND WELLNESS CENTRES

ABSTRACT

In this study, the customer's attribute information of a sports and life center was prepared by examining sold services and issues of customer loyalty, reliability and new marketing strategies were tried to discover. The data was analyzed by using data mining techniques such as classification and clustering algorithms. Weka and Microsoft SQL Business Intelligence were used. The results of the analyses were examined and evaluated in detail.

Apart from utilizing data mining softwares(e.g.Weka), a software development was conducted. Decision tree algorithm (ID3) was adapted to the operation of this application form, rules and the drawing of decision tree were carried out.

The approach and the method of data mining modeling in the study was designed to be easily changed and extended in the future. In addition , the information obtained from the the results of data mining will provide significant contributions to the studies in the field of sports and life centers.

TEŐEKKÜR

Bu alıőmanın gerekleőmesine katkılarından dolayı ve danıőmanım olarak tezin yazılmasında yol gősterdiėi iin Sayın Pınar Yıldırım'a itenlikle teőekkür ederim. Yüksek lisansın baőlangıcından itibaren iő yerinde, gerek anlayıőlı olmaları gerekse duyarlılıklarından ötürü müdürlerim Sayın Erdal Özkan ve Sayın Mehmet Yıldızdoėan'a ve son olarak bana hibir zaman desteėini esirgemeyen sevgili eőim Sayın İpek Keskin'e teőekkürleri bir bor bilirim.

İÇİNDEKİLER

| | |
|---|-----|
| İÇİNDEKİLER | I |
| TABLO LİSTESİ..... | III |
| ŞEKİL LİSTESİ..... | IV |
| KISALTMALAR..... | V |
| I. GİRİŞ..... | 1 |
| II. VERİ MADENCİLİĞİ VE ALGORİTMALAR..... | 3 |
| 2.1. VERİ MADENCİLİĞİ..... | 3 |
| 2.1.1. Veri Madenciliğinin Tarihçesi | 7 |
| 2.1.2. Veritabanı Teknolojilerinin Gelişimi | 9 |
| 2.2. VERİ MADENCİLİĞİ SÜRECİ..... | 13 |
| 2.2.1. Veri Madenciliği Standart Süreci | 15 |
| 2.2.2. İş Tanımı | 16 |
| 2.2.3. Verileri Anlama..... | 18 |
| 2.2.4. Verileri Hazırlama..... | 20 |
| 2.2.5. Model Oluşturma..... | 23 |
| 2.2.6. Değerlendirme | 24 |
| 2.2.7. Sonuç..... | 26 |
| 2.3. VERİ MADENCİLİĞİ PROBLEM TİPLERİ | 28 |
| 2.3.1. Veri Tanımlama Ve Özetleme..... | 28 |
| 2.3.2. Bölümlenme | 29 |
| 2.3.4. Kavram Tanımlama..... | 31 |
| 2.3.5. Sınıflandırma | 31 |
| 2.3.6. Öngörü..... | 32 |
| 2.3.7. Bağımlılık Analizi | 32 |
| 2.4. ALGORİTMALAR..... | 33 |
| 2.4.1. Sınıflandırma Algoritmaları | 33 |
| 2.4.2. Karar Ağaçları | 34 |
| 2.4.3. Örnek Tabanlı Yöntemler..... | 34 |

| | |
|---|-----|
| 2.4.4. İstatistiksel Sınıflandırıcılar | 35 |
| 2.4.5. Kümeleme Analizi..... | 35 |
| 2.4.6. K-Ortalama Kümeleme Algoritması | 36 |
| III. SPOR VE YAŞAM MERKEZLERİ ÜZERİNE | |
| VERİ MADENCİLİĞİ ÇALIŞMASI..... | 37 |
| 3.1. AMAÇ, KAPSAM VE YÖNTEM..... | 37 |
| 3.2. İŞ TANIMI..... | 38 |
| 3.3. VERİLERİ ANLAMA | 41 |
| 3.4. VERİLERİ HAZIRLAMA..... | 45 |
| 3.5. MODEL OLUŞTURMA..... | 64 |
| 3.5.1. Sınıflandırma Ve Kümeleme Teknikleri | 65 |
| 3.5.1.1. Weka İle Sınıflandırma Ve Kümeleme Analizi | 65 |
| 3.5.1.2. Microsoft Sql Business Intelligence İle Kümeleme Algoritması..... | 69 |
| 3.5.1.3. Id3 Algoritması Kullanılarak Geliştirilen Uygulama | 73 |
| 3.6. DEĞERLENDİRME | 83 |
| 3.7. ALGORİTMA SONUÇ RAPORU | 85 |
| IV. SONUÇ..... | 88 |
| KAYNAKÇA..... | 91 |
| EKLER..... | 94 |
| EK-1 | 94 |
| EK-2..... | 105 |
| ÖZGEÇMİŞ | 106 |

TABLO LİSTESİ

| | |
|---|----|
| 2.1. İş Tanımı Aşamaları..... | 17 |
| 2.2. Verileri Anlama Aşamaları | 19 |
| 2.3. Verileri Hazırlama Aşamaları | 21 |
| 2.4. Model Oluşturma Aşamaları..... | 23 |
| 2.5. Değerlendirme Aşamaları | 25 |
| 3.1. Değişkenlere İlişkin Bilgilendirme | 42 |
| 3.2. Veri Seti Nitelik Değer Aralıkları..... | 43 |
| 3.3. Müşteri Veri Seti..... | 46 |
| 3.4. Müşteri Kartı: Müşterilere Dair Detayların Yer Aldığı Tablo Bilgisi..... | 47 |
| 3.5. Üyelik Sözleşmesi Kartı: Üyelik Sözleşmelerine Dair Detayların Yer Aldığı Tablo Bilgisi | 47 |
| 3.6. Giriş-Çıkış Tablosu: Tesis Ve Havuz Kullanımlarının Yer Aldığı Tablo Bilgisi | 48 |
| 3.7. Plan Rezervasyonları: Spa Ve Özel Ders Hizmeti Alan Müşteri Bilgileriyle Beraber Kullanılan Tablo Bilgisi | 48 |
| 3.8. Plan Satışlarına Dair Detayların Yer Aldığı Tablo Bilgisi | 48 |
| 3.9. Grup Ders Seansları: Grup Derslerinin Takip Edildiği Tablo Bilgisi | 49 |
| 3.10. Grup Ders Seans Katılımcılarına Dair Detayların Yer Aldığı Tablo Bilgisi | 49 |
| 3.11. Weka Sınıflandırma Algoritma Değerlendirme Sonuçları | 67 |
| 3.12. K Ortalamalar Algoritması İçin K Değerleri İle Küme İçindeki Hata Payı Değişim Tablosu | 68 |

ŞEKİL LİSTESİ

| | |
|--|----|
| 2.1. Crisp – Dm İşlem Basamakları | 15 |
| 2.2. Veri Madenciliği Uygulama Alanı..... | 15 |
| 3.1. Müşteri Veri Seti İle Tablolar Arası İlişki Diyagramı | 50 |
| 3.2. Weka J48 (C4.5) Algoritma Sonucu Oluşan Kurallar | 67 |
| 3.3. Weka J48 (C4.5) Algoritma Doğruluk Oranları | 68 |
| 3.4. Weka J48 (C4.5) Algoritma Değerlendirme Sonuçları | 68 |
| 3.5. K = 5 İçin Oluşturulan Küme Tablosu | 69 |
| 3.6. K Ortalamalar Algoritması İçin K Değerleri İle Küme İçindeki Hata Payı | 69 |
| 3.7. Kümelerin Birbirleriyle Bağlantıları..... | 71 |
| 3.8. Microsoft Sql Business Intelligence Küme Profili | 72 |
| 3.9. Microsoft Sql Business Intelligence Kümeleme Algoritma Sonuçları | 72 |

KISALTMALAR

- PR: Public Relations
- WWW: World Wide Web
- Gsm: Global System For Mobile Communications
- Olap: Online Analytical Processing
- CRISP–DM: Cross Industry Standard Process for Data Mining
- Sig: Special Interest Group
- Arff: Attribute Relation File Format
- Dbms: Database Management System

I. GİRİŞ

Bilgi ve iletişim teknolojilerindeki gelişmeler, karar verme sürecine etki eden veri miktarının artmasını sağlamış ancak veri toplanan ortamların çoğalması, toplanan bilginin depolanması ve yönetilmesi konularında bir takım zorlukları da beraberinde getirmiştir. Bu durum, basit veritabanı sistemlerini yetersiz kılmış ve yeni bir takım veritabanı mimarilerinin geliştirilmesine sebep olmuştur. Karar verme sürecinin bir bileşeni olarak ilişkisel veritabanları ve veri ambarları kullanımında önemli bir artış ortaya çıkmıştır.

Verilerin elde edilmesi, depolanması ve yönetilmesi tek başına yeterli olmamıştır. Veri miktarındaki artışla beraber verilerden kullanışlı bilgi ortaya çıkarılması ve bilginin karar verme sürecine katkısının ortaya çıkarılması sektörler için yeni fırsatların ortaya çıkarılmasına konu olmuştur.

Karar verme süreçlerinde halen kullanımda olan geleneksel analiz yöntemleriyle beraber özellikle veritabanı teknolojileri ve yapay zekâ uygulamalarında yaşanan gelişmelerle bilgi keşfi önemli bir konu haline gelmiştir. Bu çalışmaların son dönem de bir araya toplandığı yaklaşım veri madenciliğidir.

Veri madenciliği, verilerin toplanmaya başlanmasından itibaren çeşitli analiz yöntemleri ile harmanlanarak bilgidan anlam çıkarma ve çıkarılan anlamı karar mekanizmasının aksiyon planı içerisinde kullanılmasına kadar olan bütün süreci kapsamaktadır.

Günümüzde çok miktarda veri ve anlamlı bilgi çıkarımı konuları astrolojiden tıpa, gsm şirketlerinde lojistik hizmeti veren firmalara, meteorolojiden spora kadar bir çok alanda veri madenciliği uygulamalarına rastlamak mümkündür. Bu çalışmada uygulama alanı olarak spor ve yaşam merkezlerinin kullandığı bir içerik yönetim sisteminin veritabanı kullanılacak ve veri madenciliğinin kümeleme ve sınıflandırma fonksiyonuna ilişkin teknikler üzerinde durulacaktır.

Bu çerçevede, spor ve yaşam merkezi sektöründe yaşanan problemler ve karar alma süreçleri için veri madenciliği yöntemleri ile çözüm sunmaktır. Çalışmada, sınıflandırma algoritmaları ve kümeleme yöntemleri kullanılmıştır. İnceleme sırasında müşterilerin tesis kullanımı, üyelik bilgileri, özel ders hizmetleri, grup ders hizmetleri ve havuz kullanımları teker teker hesaplanmış ve birleştirilen bir tablo üzerinde bu yöntemlerle çalışmalar yapılmıştır. Bu yöntemle ortaya çıkarılan kurallar test edilerek doğruluk oranları ortaya çıkarılmış, sonuçlar tablolarla göz önüne sunulmuştur. Çalışmayı destekleyecek ve kullanılan içerik yönetim sistemi ile beraber çalışabilecek bir web uygulamasının prototipinden de bahsedilmiştir.

Bu çalışmada gözden kaçan ayrıntılar ortaya çıkarılmaya çalışılmıştır. Çalışmanın amacı, veri madenciliği yöntemleri kullanılarak üyelik yenileme ve müşteri sadakatini arttırma, fiyatlandırma politikası, pr yönetimi ve buna bağlı pazarlama stratejilerine yardımcı olacak kararların çıkarılması hedeflenmiştir.

II. VERİ MADENCİLİĞİ VE ALGORİTMALAR

2.1. VERİ MADENCİLİĞİ

Karar verme süreçlerinde personel düzeyinde uzmanlık ile ilişkisel veri yönetim sistemlerini birleştirme gayreti içerisinde olan geleneksel veri analizinin yerine son yıllarda yaşanan teknolojik gelişmelerle ve yapay zekâ uygulamalarında geline seviye mevcut sistemlerin gelişimine ve değişimine önemli bir katkısı olmuştur. Bu değişimin oluşmasında veri kaynaklarının çoğalması, anlamlı verinin toplanması ve depolama alanında yaşanan teknolojik gelişme, mevcut sistem sonuçlarının karar verici açısından kolayca anlaşılması ve uygulanması ihtiyacı en önemli etkenleri oluşturmaktadır^[1].

Veri toplama ve depolama teknolojilerindeki hızlı gelişme veritabanı, veri ambarı ve www alanları gibi birçok veri deposunda her geçen gün daha fazla verinin birikmesine neden olmaktadır. Uzun olmayan bir dönem öncesine kadar karar vericilerin, yöneticilerin karşılaştığı temel problemlerden biri olan veri kıtlığı şimdilerde yaşanan olağanüstü gelişmeyi güçlü veri analizi araçları olmaksızın kavramak insanoğlunun yeteneklerini aşan bir durum haline gelmiş ve bu durum karar vericiyi veri zengini fakat bilgi fakiri konumuna sokmuştur^[2].

Günümüzde birçok işletme, faaliyet döngüsünde yeni müşterilerin kazanılması, mevcut müşterilerden varolan kazancın arttırılması gibi çalışmalarda veri madenciliği kullanılmaktadır. Müşteri analiz yöntemleri ile

¹ C. Apte, B. Liu, E. P. D. Pednault, and P. Smyth. «Business Applications of Data Mining.» Communications of the ACM, 2002: 49-53.

² Han, J., ve B. Kamber. «Data Mining : Concepts and Techniques.» AP Magazine, 2001: 4.

oluşturulan bilgi bankası yeni müşterilerin bünyesine kazandırdığı ilk dakikadan itibaren işlemeye başlar ve sistemin yol gösterdiği öngörüler ile müşteri kazanç terazisini arttırmaya çalışır.

Veri madenciliği, tüm dünyada bilinen adı ile data mining, konu kapsamı açısından herkesin üzerinde uzlaştığı tek bir ifade bulunmamaktadır. Tercih edilen biçimlerine öne çıkan tanımları aşağıdaki gibidir.

- Veri madenciliği, veri içerisinde gizli kalmış, önceden bilinmeyen ve potansiyel olarak kullanışlı olan anlamlı bilginin çıkarımıdır^[3].

- Veri madenciliği, içerisinde varolan anlamlı örüntü ve kuralları ortaya çıkarmak amacıyla, büyük miktarlardaki verinin otomatik ve yarı otomatik araçlar yardımıyla incelenmesi ve analiz edilmesi sürecidir^[4].

- Veri madenciliği, çeşitli mimarilerde depolanmış olan büyük miktarlardaki verilerden ilgi çekici bilginin keşfedilmesi sürecidir^[5].

- Veri madenciliği, veri ambarında tutulan çok çeşitli verilere dayanarak daha önce keşfedilmemiş bilgileri ortaya çıkarma ve bu bilgileri, karar vermek ve eylem planını gerçekleştirmek için kullanma sürecidir^[6].

- Veri madenciliği, veriye sahip olan kişi ya da kurum için, kuralları ve ilişkileri keşfederek, önceden bilinmeyen açık ve yararlı sonuçlar elde etmek

³ Clifton C., Thuraisingham B. «Emerging Standards for Data Mining.» Computer Standards & Interfaces, 2001: 187-193.

⁴ Flawley, W. J.; G., Piatetsky-Shapiro; Matheus C.J. «Knowledge Discovery in Databases : An Overview.» AI Magazine, 1992: 57-70.

⁵ Berry M.J.A.; Linoff G.S. «Mastering Data Mining : The Art and Science of CRM.» The Art and Science of CRM, 2000: 7.

⁶ Han J., Kamber M., a.g.e., sf.7

amacıyla, çok miktardaki verinin seçilmesi, incelenmesi ve modellenmesi sürecidir^[7].

Tüm bu tanımlamaları göz önüne aldığımızda, “çok fazla miktarda veri” ve “anlamlı bilgi çıkarımı“ ifadeleri ön plana çıkmaktadır^[8]. Yine bu tanımlamalardan ortak hareketle kolayca görülmektedir ki, veri madenciliği belirli bir teknik, kural ya da algoritmalar bütünlüğünün dışında bir süreçtir. Bu süreç, verilerin elde edilmesi ile başlayan ve bulguların gerçek yaşama aktarılmasına kadar süren tüm döngüyü kapsamaktadır^[9].

Veri madenciliğinin en önemli üstünlüğü yapılmış analiz sonrası öngörülen bulguları bir araya getirerek verilerin analizini gerçekleştirilmesi ve işletme aksiyonlarını ya da çeşitli faaliyet girişimlerine destek olacak bilgilerin çıkarımını sağlamasıdır. Bu destekle beraber, birikmiş ve faydasız veri bütünlüğünden toprak altında yatan hazine değerindeki bilgiye ulaşmak mümkündür.

Ancak bu bakış açısı, yaygın olarak bazı yanlışları da beraberinde getirmiştir. Bu yanlışlardan biri veri madenciliğinin olduğundan daha farklı bir araç haline getirmiştir. Veri madenciliğini bir yöntem olarak seçen tüm uygulamacıların veri madenciliğini doğru bir şekilde anlaması ve beklentilerini bu doğrultuda belirlemesi son derece önemlidir. Doğru şekilde anlaşılmamış bir uygulama ve veri kümesinin karar alma sürecine yarardan çok zararı dokunacaktır.

⁷ Swift R. «Accelerating Customer Relationship.» Prentice Hall PTR, 2001:93

⁸ Edelstein H. Data Mining in Depth: Description is not Prediction. Mart 2003. <http://www.dmreview.com/issues/20030301/6388-1.html> (Mart 8, 2013 tarihinde erişilmiştir).

⁹ Edelstein H. Data Mining in Depth: Using Data Mining to Find Terrorists. Mart 2003. <http://www.dmreview.com/issues/20030501/6655-1.html> (Mart 8, 2013 tarihinde erişilmiştir).

Veri madenciliği uygulamalarının başarılı olabilmesi için mutlaka göz önünde bulundurulması gereken bazı önemli noktalar mevcuttur^[10].

Veri madenciliği araçları problemi kendiliğinden çözemez. Mutlaka sürecin her aşaması özenli ve dikkatli bir şekilde ele almak gerekmektedir.

Veri madenciliğinde bilgi ne kadar önemliyse bilginin doğru şekilde yönlendirilmesi de bir o kadar önemlidir. Yapılacak araştırma konusunda deneyim ve bilgi birikiminin sürecin sağlıklı işlemesi konusunda anahtar rol oynayacaktır. Bu süreçte tecrübe verilerden daha önemlidir.

Veri madenciliği çalışması için yapılan çalışma ciddi bir maliyet getirebilir. Bu sebepten ötürü çalışmanın sağlayacağı artı ve eksi yönler çıkarmalıdır. Titizlikle kontrol edilmeli ve buna göre çalışmaya başlanmalıdır.

Veri madenciliği yazılımları kullanıcı odaklı tasarlanmış olsa da uygulanan teknikler açısından bilgi birikimi ve deneyim gerektirir.

Veri madenciliği analizi sonucu belli kurallar ya da kalıplar ortaya çıkarılsa da bu işlemlerin nedenleri ile ilgilenmezler. Sadece gözden kaçmış gizli verileri ve ilişkileri ortaya çıkarırlar. Ortaya çıkarılan sonuçlar sadece işin uzmanları tarafından değerlendirilebilir.

Veri madenciliğinin en önemli aşaması verilerin hazırlanma sürecidir. Verilerin kalitesi uygulama sonuçlarının kalitesi demektir.

Bu önemli noktaların yanı sıra uygulamanın olmazsa olmaz diğer bir koşulu veri madenciliği araçlarıdır. Veri madenciliği projelerinin süreçlerini göz önüne aldığımızda yapı bu araçlar yoluyla yürütüldüğünden, projede

¹⁰ Co., Two Crows. Introduction to Data Mining and Knowledge Discovery. <http://www.twocrows.com/intro-dm.pdf> (5 8, 2013 tarihinde erişilmiştir)

kullanılacak yazılımın seçilmesi de çok önemlidir. Birçok yazılım, aynı içeriğe sahip gözüken farklı uygulamaya sahip algoritma ve komutları içermektedir. Bundan yola çıkarak hafıza kullanımı, depolama ve hız gibi nitelikler açısından da yazılım seçimi önemli bir noktadır^[11].

En yaygın kullanımlı veri madenciliği yazılımları:

- SPSS Clementine,
- SAS Enterprise Miner,
- IBM Intelligent Miner for Data,
- DBMiner,
- Statistica Data Miner

ile Weka, Yale, Rapidminer ve Orange gibi özgür yazılımlar yer almaktadır.

2.1.1. Veri Madenciliğinin Tarihçesi

Veri madenciliği bilişim sektöründe yeni bir çalışma olduğu ne kadar doğru ise kökeni bir o kadar da eskidir. Veri madenciliği, gerçekte çok uzun yıllara dayanan bir evrimin sonucu olmakla birlikte, bütün çalışmaların başlangıcı olarak 1990'lar kabul edilmektedir. 90'lı yıllardaki yapılan bilimsel araştırma, geliştirme ve çalışmaların tartışıldığı ilk konferans^[12] 1995 yılında Bilgi Keşfi ve Veri Madenciliği adı altında Montreal 'de yapılmıştır. Daha sonrasında buradan hareketle sürekli yayın haline gelmiştir^[13].

Veri madenciliğinin birden çok disiplini bir araya getirildiği bir çalışma olduğunu varsayarsak, köklerinin temeli yüz yıllarca öncesine dayanmaktadır. Çünkü veri madenciliği disiplininin içinde barındırdığı

¹¹ Co., Two Crows. a.g.e

¹² The First International Conference on Knowledge Discovery and Data Mining. Montreal: Springer US.

¹³ Journal of Data Mining and Knowledge Discovery, Springer US.

istatistik, yapay zekâ ve veritabanı teknolojilerinin gelişimiyle ortaya çıkmıştır. Bütün bu disiplinler ayırmak gerekirse istatistik analizi, veritabanı teknolojileri bilgi ve verimi, yapay zekâ ise kazancı ortaya çıkarmaktadır^[14].

Bütün gelişim sürecini göz önüne aldığımız da istatistik ile teorik bilgiyi ortaya çıkarır ve pratik uygulamaları tetikler, yapay zekâ ise neden sonuç ilişkisinin dışında faydayı dikkate alır.

Sonuç olarak istatistik modelleme, yapay zekâ ise algoritmaları kapsamaktadır^[15].

Bilgi keşfinin her aşamasında gelişmelerin hızlanacağından kimselerin kuşkusu yoktur. Algoritma kullanımının öneminin daha da artacağı için önemli seviyede gelişeceği, keşfedilen bilginin yeniden kullanılabilen yöntemlere ihtiyaç duyulacağı düşünülmektedir.

Geleceğin en önemli keşiflerinden bir tanesi de bilgi seviyesinin geliştirilmesi ve her seviyede kolaylıkla entegre edilebilir hale getirilip bilginin istenilen çözümlenmeye ulaştırılmasının sağlanması isteğidir.

Veri madenciliği, köken ve gelişim süreci incelendiğinde temelini insanoğlunun gelişiminin sağladığı ihtiyaçlardan almaktadır. Bu sebepten gelecek için insanların oluşturduğu teknolojilerle elde edilen veri birikimi arttıkça önemini kaybetmeden kullanımının yoğunlaşacağı ve gelişime katkı sağlayacağı söylenebilmektedir.

¹⁴ Zhou Z. «Three Perspectives of Data Mining.» Artificial Intelligence 2003: 139-146.

¹⁵ Kantardzic M. «Data Mining: Concepts, Models, Methods and Algorithms.» IEEE, 2003: 4.

2.1.2. Veritabanı Teknolojilerinin Gelişimi

Bilgi edinilecek veri kaynaklarının artması ile birlikte muhafaza edilmesine yönelik çalışmalarında artmasına neden olmuştur. Veri çeşitliliği göz önüne alındığında basit saklama yöntemleri ile gelişmiş veritabanı mimarileri de ortaya çıkmıştır. Çeşitlilikle birlikte uygulamalar farklı ve esnek bir şekilde bir arada uygulanması söz konusu olmuştur.

Veritabanı basit olarak düzenli bilgilerin bir arada toplandığı bir topluluktur. Biraz detaylandırmak gerekirse verilerin depolandırılması, erişilmesi, değiştirilmesini tetikleyecek şekilde organize edilmiş ve birden fazla veri kaynağının bir araya getirildiği topluluktur^[16].

Günümüzdeki veritabanı mimarilerini incelediğimizde de yukarıda bahsedilen çeşitlilik karşımıza çıkmaktadır. Bunlar:

- İlişkisel veritabanları
- İşlemsel veritabanları
- Veri ambarları
- Düz dosyalar (Flat Files)
- WWW (World Wide Web)
- Nesne merkezli veritabanları
- Nesne ilişkili veritabanları
- Özel uygulama merkezli veritabanlarıdır^[17].

Bir veritabanının gereken temel özellikleri ölçeklenebilir ve esnek olması, güvenilir olması, yüksek performanslı ve yönetilebilir olmasıdır.

¹⁶ Flawley W.J., Piatetsky-Shapiro G., Matheus C.J., a.g.e.

¹⁷ Han J., Kamber M., a.g.e., sf 10-21

Ölçeklenebilirlik ve esneklik, kullanıcı ve uygulama bazında anlaşılrsa da veritabanının gerek bellek gerekse işlemci için ihtiyaçlarına da karşılık gelmektedir. Veri boyutu ve kaynağının gün geçtikçe büyümesi sistem bazında bu büyümenin veritabanı sisteminin de karşılayabilmesi gerekmektedir. Burada yaşanacak bir sorun bütün uygulamanın etkinliğini ve maliyetine doğrudan etki etmektedir.

Güvenilirlik süreci her aşamasında oldukça önem taşıyan bir konudur. Çünkü veritabanı üzerinde koşan uygulamalar sadece şahısa özel yazılımlar değil gerek kamu sektörü gerekse askeri alanlarda da uygulamalar kullanılmaktadır. Güvenirlik konusunda oluşacak bir aksaklık bilgi sürecinin istenilen ulaşma konusunda sağlıklı veriye ulaşımını engelleyecektir.

Kullanılacak verilerin süreci göz önüne aldığımızda sadece depolanması yeterli olmayacaktır. Bu yüzden süreçte veriye erişim ve yönetim konusu oldukça önemlidir. Bu sebepten veri işleme performansı ve yönetebilirlik, veritabanı teknolojilerinin performans değerlendirmeleridir. Veri madenciliği uygulamasının temeli veritabanıdır. Burada veritabanı konusunda işlem sırasında bazı sorunlar ile karşılaşabiliriz. Bunlar:

- Dinamik veri yapısı
- Geçersiz veri alanları
- Kayıp veri alanları
- Kayıp veri değerleri
- Gürültü ve belirsizliktir^[18].

¹⁸ Flawley W.J., Piatetsky-Shapiro G., Matheus C.J., a.g.e.

Dinamik veri yapısı, verilerin zaman duyarlı olmasına rağmen gözlemlerin zamandan bağımsız olarak elde edilmesinden kaynaklanan bir sorunu ifade eder. Bu durum sürecin gelişimine olumsuz yönde etki eder.

Geçersiz veri alanları, veri kümesinde yer alan veri alanlarının veri madenciliği çalışmasına karşılık uygun olmamasından kaynaklanmaktadır. Bu sürecin önemli basamaklarından biridir. Uygun alan seçimi hedefe adım adım ilerlemesinde önemli bir kriterdir.

Kayıp veri alanları, birden fazla veri kaynağından çalışıldığı varsayılırsa bir kaynakta olan verinin diğer kaynakta karşılığının olmaması durumudur. Bu durum aşmak için verileri göz önüne alınmaması yapılan ve yapılacak çalışmalar için geçerli bir çözüm olmaktadır.

Kayıp veri değerleri, etkileşimli sistemlerde değer ataması bölümlerinde ilgili alana karşılık gelmeyen verilerin oluşmasından kaynaklanır. Bu durum dikkatle incelenmeli ve göz önünden kaçırılmaması gereken bir durumdur.

Gürültü ve belirsizlik sorunları, verilerin doğasından kaynaklanmaktadır. Burada karşılaşılan hata veri alanının belirlenen değerler üzerinde yer alması, sayısal alanlara metin değerlerin atanması, sıralı ilerleyen bir alanının sırasının bozulması da bu sorunumuza örnektir. Bu durum her zaman oluşabileceği gibi daima dikkat edilmesi gereken bir noktaldan biridir.

Bir veritabanı sistemi, ya da veritabanı yönetim sistemi (DBMS), veritabanını oluşturan ve karşılıklı ilişki halindeki verilerin bir topluluğu ile

verilere erişimi ve verilerin yönetilmesini sağlayan bir grup yazılım paketidir^[19].

Veritabanı yönetimi sistemlerinden bazıları:

- Microsoft Access
- dBase
- FoxPro
- Paradox
- Microsoft Excel vb. kişisel veritabanı yazılımları
- Oracle
- DB2
- Sysbase
- Informix
- Progress
- MS SQL Server 'dir.

Yukarıda belirtilmiş veritabanı yönetim sistemlerini kullanarak veri ambarları üzerinde karmaşık analizlere imkân veren OLAP çözümlerini örnek vermek mümkündür^[20].

OLAP çözümlerinin veritabanında ortaya çıkan örüntülerin nedenlerini de açıklayabilir. Bu özellik geleneksel sorgulama ve raporlama araçlarından farkını ortaya çıkarır.

OLAP uzmanı veri üzerinde ilişki üretmekte ve bunları kanıtlamak veya reddetmek için veritabanı üzerinde bu sorguları kullanmaktadır. OLAP

¹⁹ Han J., Kamber M., a.g.e., sf.10

²⁰ Bayraktar R., a.g.e.

analizi veri madenciliği anlamına gelmemektedir. Çünkü OLAP analizi tündengelim, veri madenciliği ise tümevarım süreci ifade eder^[21].

Veri ambarları, veri madenciliği söz konusu olduğunda akla gelen en önemli veritabanı mimarisi olup, yönetim karar verme sürecini kolaylaştırmak amacıyla çok sayıda heterojen veri kaynağının tek bir düzen altında organize edilmesiyle oluşturulan büyük bir veri deposu olarak tanımlanmaktadır^[22].

Bir başka ifade ile veri ambarları, bir kurumda gerçekleşen tüm operasyonel işlemlerin en alt düzeydeki verilerine kadar inebilen, etkili analiz yapılabilmesi için özel olarak modellenen, tarihsel derinliği olan, fiziksel olarak operasyonel verilerden farklı ortamdaki yapılardır.

2.2. VERİ MADENCİLİĞİ SÜRECİ

Veri madenciliği, işe ilişkin ihtiyacın ortaya çıkmasından, oluşturulacak modelin gerçek yaşama aktarılmasına kadar yürütülecek tüm faaliyetleri içeren bir süreç olarak düşünülmelidir^[23].

Veri madenciliği sürecinin bir diğer önemli boyutu ise yinelemeli bir süreç olmasıdır. Veri madenciliği uygulamalarında genellikle birden çok model oluşturulmakta ve en uygun modelin elde edilmesi hedeflenmektedir. Eğer elde edilen en uygun model de yeterli görülmez ise süreç yeni kayıtların veri setine dâhil edilmesi ya da kayıtlardaki sorunların giderilmesi yoluyla yeniden gerçekleştirilir^[24].

²¹ Alpaydın E. «Zeki Veri Madenciliği : Ham Veriden Altın Bilgiye Ulaşma Yöntemleri.» *Bilişim 2000 Eğitim Semineri*. İstanbul, 2000.

²² Man D. «Answering Some Common Data Warehousing Questions.» *Direct Marketing*, 1996: 12-15.

²³ CRISP – DM Konsorsiyumu; CRISP – DM 1.0 Step – by – Step Data Mining Guide. <http://asaha.com/ebook/UNjcxNzA-/CRISPWP-0800.pdf> (1 21, 2013 tarihinde erişilmiştir).

²⁴ Fayyad U., Uthurusamy R. «Data Mining and Knowledge Discovery in Databases.» *Communications of the ACM*, 1996: 24-26.

Veri madenciliği süreci, temel bir takım ortak faaliyetleri kapsamakla birlikte, başlangıçta her uygulamacı tarafından farklı şekilde yürütülmekteydi. Bu nedenle, uygulamada ortaya çıkan farklılıkları en aza indirecek bir standart süreç geliştirilmesi ihtiyacı ortaya çıkmıştır. Bu ihtiyaç doğrultusunda, veri madenciliği uygulamalarının dört önemli lideri^[25] ile iki yüzden fazla araştırmacıyı bir araya getiren CRISP–DM konsorsiyumu^[26], düzenlediği çeşitli çalıştaylarla veri madenciliği uygulamalarının olgunluk kazanması ve tüm uygulamacılara yol gösterecek bir standart süreç modelinin oluşturulması için 1996 yılının sonlarında çalışmalarına başlamıştır^[27].

Bu konsorsiyum, bir yıl sonra Avrupa Komisyonu'nun sağladığı fon ile çalışmalarını hızlandırmış ve 1990'lı yılların sonuna gelindiğinde çalışma taslağı büyük oranda oluşturulmuştur.

Güncellenmiş şekli ile CRISP–DM 1.0 dokümanı ise 2000 yılında yayımlanmıştır^[28]. CRISP–DM 1.0, sadece akademik bir çalışma olmayıp veri madenciliği konusunda pratik, gerçek yaşam deneyimleri üzerine kurgulanmış bir süreç modelidir. Bu süreç modeli, temel hedefine ulaşmış olmakla birlikte nihai nitelik taşımamakta ve güncelleştirme ihtiyaçları doğrultusunda CRISP–DM 2.0 üzerindeki çalışmalar yine CRISP–DM SIG tarafından sürdürülmektedir.

Veri madenciliği süreci dışında veri madenciliği araçları ve mimarisine dönük standartlaştırma çabaları günümüzde de devam etmektedir.

²⁵ (DaimlerChrysler), Thomas Reinartz. SPSS, NCR Systems Engineering, OHRA Verzekeringen en Bank Groep.

²⁶ Cross Industry Standard Process for Data Mining. CRISP-DM.

²⁷ Shearer C. «The CRISP-DM Model: The New Blueprint for Data Mining.» Journal of Data, 2000: 13-22.

²⁸ CRISP – DM Konsorsiyumu, a.g.e

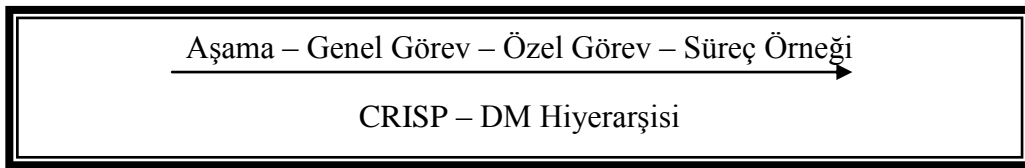
2.2.1. Veri Madenciliği Standart Süreci

Veri madenciliği uygulama sürecinin standardize edilmesine yönelik ilk çalışma olan CRISP – DM 1.0, veri madenciliği uygulamalarında izlenmesi gereken basamakları belli bir disipline sokan bir kılavuz niteliğindedir. Uygulamamın aşamaları da bu standart süreci temel alarak hazırlanacaktır.

CRISP-DM veri madenciliği aşamaları (Şekil 2.1), genelden özele olmak üzere dört farklı seviyeden oluşup belirli sıraya sahip bir yapıdır. Bu sıralama altı farklı aşama, her aşamaya özgü görevler, tüm görevlere ilişkin bir takım farklı özel görevler ile somut faaliyetler süreci tanımlayacaktır.

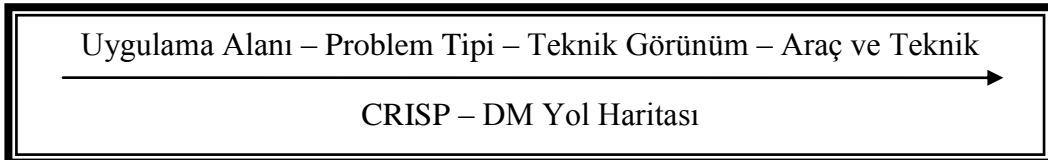
Bu sıralamanın ilk basamağında iş tanımı, verileri anlama, verileri hazırlama, model oluşturma, değerlendirme ve sonuç yer almaktadır. İkinci adımda, ilgili aşamada yapılması gereken genel görevler sıralanmıştır. Üçüncü adım görevlerin işleyişini belirleyecek özel görevler yer almaktadır. Son adımda süreç örneği, gerçek bir veri madenciliği uygulamasının sonuçlarını, kararlarını ve eylemlerini ifade etmektedir.

Şekil 2.1 CRISP – DM İşlem Basamakları



Veri madenciliği çerçevesi (Şekil 2.2) olarak ifade edilen yol haritası ise genelden özele giden farklı seviyenin değerlendirmesini oluşturmaktadır.

Şekil 2.2 Veri Madenciliği Uygulama Alanı



Uygulama alanı: Projenin yönetileceği alanı ifade eder. (1.Adım)

Problem Tipi: Projede kullanılan veri madenciliği fonksiyonlarını niteler. (2.Adım)

Teknik Görünüm: Teknik anlamda yaşanan sorunları tanımlar. (3.Adım)

Araç ve Teknik: Kullanılan yazılımları ve uygulama teknikleri gösterir. (4.Adım)

CRISP-DM modeli, veri madenciliği çalışma sürecini altı basamaktan oluşan bir döngü olarak göstermektedir. Bu proje süreçlerinin aşamalarını ve işlem döngülerini göstermektedir. Model, her aşamanın teker teker geçilmesi, gerektiğinde bir önceki aşamaya geri dönülmesi ve döngünün kendi için bu süreci takip etmesini önermektedir.

Veri madenciliği projelerinin sonucu aslında projenin bittiği anlamına gelmemektedir. Projeye eklenecek verilerle, karşılaşılabilecek sorunlarla veya yeni oluşabilecek ihtiyaçlarla sürecin yaşam döngüsünün devam etmesi mümkündür.

2.2.2. İş Tanımı

Veri madenciliği projesinin tartışmasız en önemli aşaması proje hedeflerini anlama, eldeki veriyi veri madenciliği problem tanımına dönüştürme ve bundan yola çıkarak proje planı oluşturmaya kadar uzanan bir süreci kapsamaktadır.

İş tanımı, başlangıç aşaması projenin hedef ve ihtiyaçlarına odaklanmayı ve buradan edinilen bilgi ile de problem tanımı oluşturmayı sağlar. Sonraki aşaması da projeye planını temelini oluşturmaktır (Tablo 2.1).

Tablo 2.1 İş Tanımı Aşamaları

| Aşama | İş Tanımı |
|--------------|---|
| Görev | İş Hedeflerini Belirlemek |
| Çıktı | İş Geçmişi |
| Çıktı | İş Hedefleri |
| Çıktı | İş Başarı Kriterleri |
| Görev | Durumu Değerlemek |
| Çıktı | Kaynakların Envanteri |
| Çıktı | Gereklilikler, Varsayımlar ve Kısıtlar |
| Çıktı | Riskler ve Çözüm Alternatifleri |
| Çıktı | Terminoloji |
| Çıktı | Maliyetler ve Faydalar |
| Görev | Veri Madenciliği Başarısını Belirlemek |
| Çıktı | Veri Madenciliği Amaçları |
| Çıktı | Veri Madenciliği Başarı Kriteri |
| Görev | Proje Planını Hazırlamak |
| Çıktı | Proje Planı |
| Çıktı | Araç ve Tekniklerin Başlangıç Değerlemesi |

İşin hedefinin çıkarılması, veri madenciliği sürecinin tam olarak ne elde edilmek istendiğinin cevabını ortaya çıkarmaktadır. Bu doğrultuda, verinin bütünlüğü çerçevesinde işin durumu hakkında bilgi edinilecek, proje organizasyonunun temel beklentisi ortaya çıkacak ve projenin başarıya ulaşmasındaki kısıtlar belirlenecektir.

Durum değerlendirmesi, veri madenciliği projesinin başarısı ve projenin planında ortaya çıkarılacak tüm yeterlilikler, sınırlamalar ve varsayımların tüm ayrıntılarıyla ortaya çıkarılmasını sağlamaktadır. Bu doğrultuda projeye katılabilecek personel sayısı, veri ihtiyaçları, donanım ve yazılım olanakları ortaya çıkarılmış, bir proje iş takvimi oluşturulmuş, kullanılacak teknoloji ve veri güvenliği konusunda tüm sınırlamalar ortaya çıkması sağlanacaktır.

Yine bu görev sırasında projede oluşacak aksaklıklar, başarısızlıklar veya sonucu etkileyecek tehditlere karşı alternatif plan da oluşturulacaktır.

Görev sonucunda da yapılan fayda maliyet analizi ile üretilecek fayda maliyeti kıyaslanabilecektir.

Veri madenciliği başarısını belirleme görevi, veri madenciliği projesi ile neye ulaşılmak istendiğinin ve ulaşılabilecek sonuç değerlendirmelerinin proje bütününe de genel ve ya nesnel anlamda proje başarı kriterinin ne olabileceğini ifade etmektedir.

Başarı kriteri kavramını hem iş hedeflerine göre hem de veri madenciliği hedeflerine göre ayrı ayrı değerlendirmek mümkündür. İş hedefi proje beklentilerini, veri madenciliği hedefi ise teknik bakımda uygulanan işlemlerle veri madenciliği algoritmalarının ortaya çıkaracağı analizdir. Bu iki hedefte birbirine sıkıca bağlı gözükse de projeden projeye birbiri ile aynı doğrultu da sonuç vermeyebilir.

2.2.3. Verileri Anlama

Başlangıç verilerinin elde edilmesi, verileri tanımaya dönük analizleri ve verilerdeki saklı örüntüleri bulabilmek adına ilk izlenimlerin ortaya çıkarıldığı veri madenciliği aşamasıdır.

Bu aşamada, verilerin elde edilmesinden başlayarak veriyi tanıyabilme adına temelden incelenip genel yapı itibarıyla fikir edinebilme, veri kalitesi, veri problemleri ve verilerin kümeleme işlemleri gibi birçok çalışmayı kapsamaktadır (Tablo 2.2). Başlangıç verilerini elde etme görevi, proje çerçevesinde tanımlanan verilerin elde edilmesi ve veri kümesinin oluşturulması çalışmalarının yapılabilmesi adına yazılımla işlenmesi sürecini kapsamaktadır.

Tablo 2.2. Verileri Anlama Aşamaları

| | |
|--------------|---------------------------------------|
| Aşama | Verileri Anlama |
| Görev | Başlangıç Verilerini Elde Etmek |
| Çıktı | Başlangıç Verilerini Elde Etme Raporu |
| Görev | Verileri Tanımlamak |
| Çıktı | Veri Tanımlama Raporu |
| Görev | Verileri İncelemek |
| Çıktı | Veri İnceleme Raporu |
| Görev | Veri Kalitesini Doğrulamak |
| Çıktı | Veri Kalitesi Raporu |

Bu görev veri kümesinin listelenmesi, verinin elde edilmesi ve yüklenmesi sırasındaki zorlukların ortaya çıkarılması ve projenin ilerleyen safhalarında çıkacak sorunlara karşı başarılı müdahalelerin yapılabilmesi adına yapılan bir birleşimdir.

Veri tanımlama görevi, yapılan ilk analiz sonrası veri kümelerinden ön plana çıkan özelliklerin incelendiği kısımdır. Bu görev ile beraber, verilerin genel yapısı, alan ve kayıt sayıları, veri alanı özellikleri ve proje için verinin yeterliliklerini ortaya çıkaran önemli bir adımdır.

Veri inceleme görevi, veri sorgulama, görselleştirme ve raporlama gibi yöntemlerle sorulara cevap aranacak kısımdır. Bu kısım, ilişki analizleri ve birleştirme işlemleri, karakteristik özellikler ortaya çıkarılması gibi faaliyetleri içermektedir. Bu faaliyetlerin sonucunda veri tanımlama, veri kalitesi raporlamaları ya da ileri de analiz sonuçlarını destekleyecek çıkarımlar ortaya çıkarılması istenmektedir. Gerçek yaşam verileri ile yapılan bir çalışmada verilerin genel özellikleri eksik, gürültülü ve büyük miktarda olduğundan veri kalitesinin sağlanması dikkat gerektiren, uzun ve yorucu bir görevdir^[29].

²⁹ Fayyad U., Piatetsky-Shapiro G., Smyth P. «The KDD Process for Extracting Useful.» Communications of the ACM, 1996: 27-34.

Veri kalitesini doğrulama görevi, geçerli bir veri kümesine sahip olup olunmadığı, veri kümesinin hata içerip içermediği, varsa hataların ne derecede ve hangi alanlarda olduğu, kayıp veya gürültünün olup olmadığı gibi işlemlerle veri kalitesi hangi derecede olduğu sorusuna cevap arandığı kısımdır.

2.2.4. Verileri Hazırlama

Hazırlama sürecinde, modelleme amacıyla kullanılacak veri kümesinin oluşturulmasındaki bütün faaliyetleri içermektedir. Tablo, değişken ve kayıt seçimi, birleştirme ve temizleme işlemleri bu aşamada yapılacak başlıca işlemlerdir.

Bu aşama, model geliştirme sürecinde kullanılacak olan veri kümesinin oluşumundaki tüm çalışmaları kapsamaktadır (Tablo 2.3). Önceden belirlenmiş bir kalıp bulunmamakta, veri kümesi üzerinde tablo, alan ve kayıt seçimlerinin yanı sıra modelleme araçları için temizleme ve veri dönüşümü gibi işlemler yapılmaktadır. Bu aşama sonucunda modelleme ya da temel analiz çalışmaları için gerekli tanımlamaların yapılması mümkündür.

Verileri seçme görevi, veri madenciliği hedefi ve veri kalitesini hakkında ilk bilgiler ışığında analiz aşamasında kullanılacak olan veri setine karar verilir. Verilen karar ile oluşturulan veri kümesine yapılacak ekleme ve çıkarma işlemleri nedenleri ile belirtilmelidir. Bunun nedeni proje gidişatını ciddi şekilde etkileyecek bir durumda olmasıdır. Ayrıca burada yapılacak işlemlerde kayıt ve alan seçimi yapılmaktadır.

Tablo 2.3. Verileri Hazırlama Aşamaları

| | |
|--------------|------------------------------------|
| Aşama | Verileri Hazırlama |
| Çıktı | Veri Kümesi |
| Çıktı | Veri Kümesi Tanımlanması |
| Görev | Verileri Seçmek |
| Çıktı | Veri Seçiminin Mantığı |
| Görev | Verileri Temizlemek |
| Çıktı | Veri Temizleme Raporu |
| Görev | Veri Yapılandırmak |
| Çıktı | Türetilmiş Değişkenler |
| Çıktı | Üretilmiş Kayıtlar |
| Görev | Verileri Birleştirmek |
| Çıktı | Birleştirilmiş Veri Seti |
| Görev | Verileri Biçimlendirmek |
| Çıktı | Yeniden Biçimlendirilmiş Veri Seti |

Veri temizleme görevi, veri kalitesinin analiz için istenilen seviyeye çıkarma işlemidir. Bu kısımda, veri kalitesini üst safhaya çekebilmek için gerekirse alt kümeleme işlemleri ya da uygulanacak modellemeler ile kayıp veri kestirimi yapılmaktadır. Bu kısmın sonunda oluşturulacak raporda veri kalitesini arttırmak ve uygulanacak teknikler yardımıyla daha yüksek ihtimalli sonuçlar üretmek adına neler yapılabileceği belirtilmektedir.

Veri yapılandırma görevi, veri kümesi üzerinde bir ya da birden fazla yeni değişkenin türetilmesi ve yeni kayıtların ya da mevcut veri alanlarının dönüştürülmüş değerler üretebilmesi için çalışma yapılmaktadır.

Verileri birleştirme, farklı ortamlarda bulunan veri kaynaklarının kullanılması durumunda oluşturulan yapı ile tek bir merkezi komutadan idare edilmesi ve bütünlüğün sağlanarak erişimlerle yeni kayıtların oluşturulmasına olanak sağlayan basamaktır.

Ancak her veritabanı, kayıp verileri ve doğru olmayan eksik değerli kayıtlar içerecektir. Burada yapılacak birleştirme işlemi yapıldıktan sonra

mutlaka dikkatlice veri kaynağı incelenmelidir. Tek bir kaynaktan veri kalitesini yakalamak ne kadar sorunlu bir süreçse burada oluşturulan farklı kaynakların entegrasyonu ve kalite işlemleri için bir o kadar sorun ve zorluk çıkaracaktır^[30].

Verilerin biçimlendirme görevi, verilen anlamlarını değiştirmeden modelleme gereği belli bir formata sokulma işlemleridir. Bu görev sonucunda, veri kümesinde bulunan değişkenlerin ve kayıtların belli bir düzende, belli bir standartta ve sıralı bir şekilde modellemeye uygun olarak kullanıma uygun hale getirilmesi amaç haline getirilmiş olmaktadır.

Veri madenciliği başarısında kaliteli verinin ne kadar önemli olduğu dikkate alınırsa veri ön işleme (Data Preprocessing) faaliyetlerinin önemi de bir o kadar öne çıkmaktadır.

Modelleme öncesinde:

- Veri Temizleme (Data Cleaning)
- Veri Birleştirme (Data Integration)
- Veri Dönüştürme (Data Transformation)
- Veri İndirgeme (Data Reduction) faaliyetleri kaliteli veri için altın değerinde önem taşımaktadır^[31].

Veri ön işleme sürecisinin sınıflandırma kesinliği açısından olumlu sonuçlara neden olduğu çeşitli tekniklerin kullanımında deneysel olarak kanıtlanmıştır^[32].

³⁰ Edelstein H. Data Mining in Depth: TI Ain't. Nisan 2003. <http://www.dmreview.com/issues/20030401/6512-1.html> (5 8, 2013 tarihinde erişilmiştir).

³¹ Oğuzlar A. «Veri Ön İşleme.» Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 2003: 67-76.

³² Crone S.F. «The Impact of Preprocessing on Data Mining: An Evaluation of Classifier Sensitivity in Direct Marketing.» European Journal of Operational Research, 2006: 781-800.

Veri madenciliği projesi için harcanan emek, zaman ve maliyetin büyük bir bölümü verilerin modelleme için hazır hale getirilme sürecinde harcanmaktadır. Net bir oran vermek gerekirse %50 - %70 aralığında bu faaliyet oluşumu için zaman harcanmaktadır.

2.2.5. Model Oluşturma

Model oluşturma aşaması, modelleme tekniğinin seçilmesi ve seçilen modele ilişkin parametrelerin ayarlanması gibi çalışmalar yer almaktadır (Tablo 2.4). Ayrıca veri madenciliği problemi için kullanılacak birden çok tekniğin bulunması ve bunların uygulamasında farklı veri kümelerine ihtiyaç duyulması veri madenciliği yaşam döngüsünde de belirttiğimiz gibi tekrar veri hazırlama aşamasına dönmeyi gerektirmektedir.

Tablo 2.4. Model Oluşturma Aşamaları

| | |
|--------------|-----------------------------------|
| Aşama | Model Oluşturma |
| Görev | Modelleme Tekniğini Seçmek |
| Çıktı | Modelleme Tekniği |
| Çıktı | Modelleme Varsayımları |
| Görev | Sınama Tasarımı Üretmek |
| Çıktı | Sınama Tasarımı |
| Görev | Model Kurmak |
| Çıktı | Parametre Ayarları |
| Çıktı | Modeller |
| Çıktı | Model Tanımlaması |
| Görev | Model Değerlemek |
| Çıktı | Model Değerlemesi |
| Görev | Revize Edilmiş Parametre Ayarları |

Model kurma süreci denetimli (supervised) ve denetimsiz (unsupervised) öğrenmenin kullanılmasına göre farklılıklar sağlayacaktır. Denetimli öğrenme sürecinde önceden tanımlanmış sınıfların özelliklerinin belirlenmesi ve bu özelliklerden yola çıkarak kuralların oluşturulması hedeflenmiştir. Kurallar ortaya çıktıktan sonra yeni oluşturulan veri kümeleri

bu kurallar ile analiz edilip öngörüler ortaya çıkarılmaktadır. Bu da verilerin hangi sınıflara ait olduğu ortaya çıkarılmaktadır. Denetimsiz öğrenmede ise verilerin benzerliklerine ya da birbirine olan uzaklıklarından hareketle sınıfların üretilmesi amaçlanmaktadır^[33].

Modelleme tekniğinin seçilmesi görevi, kullanılacak olan tekniği net ve detaylandırılmıştır. Birden fazla teknik kullanılmak istenirse bu görev teknik sayısı kadar oluşturulmalıdır. Bu görev sonrası kullanılacak teknik ve varsayımlar belgelenmiş olacaktır.

2.2.6. Değerlendirme

Modelin sonuçlandırılmasından önceki modelin oluşum sürecinin dikkate alındığı kısımdır. Burada süreç gözden geçirilir, hedeflerin yerine getirilme durumları incelenir ve son olarak oluşumun başarı değerlendirilmesinin yapıldığı aşamadır. Buradan elde edilen verilerle sonuçların nasıl kullanılması gerektiğine karar verilmektedir.

Proje bu aşamaya geldiğinde veri analizi ortaya çıkmış, veri belirli bir kaliteye ulaşmış olduğu görülmekte ve uygulanacak model ya da modellerin elde edilmiş olduğu aşamadır (Tablo 2.5). Bu kısımda modelin iş hedefine dönük olarak başarılı olup olmadığı veya atlanılan bir kısmın olup olmadığı kontrol edilmektedir.

Bu aşama önceki aşamalara göre daha çok kesinlik ve bu kesinliklerden oluşan genel koşullarla ilgili faktörlerle ilgilidir.

³³ Hegland M. Data Mining Techniques, 2001: 313-315.

Tablo 2.5. Değerlendirme Aşamaları

| | |
|--------------|---|
| Aşama | Değerlendirme |
| Görev | Sonuçları Değerlendirmek |
| Çıktı | İş Başarısı Açısından Veri Madenciliği Sonuçlarının Değerlemesi |
| Çıktı | Onaylanmış Modeller |
| Görev | Süreci Gözden Geçirmek |
| Çıktı | Sürecin Gözden Geçirilmesi |
| Görev | Sonraki Adımları Belirlemek |
| Çıktı | Olası Eylemlerin Listesi |
| Çıktı | Karar |

Sonuçların değerlendirme görevi ise modelin hedefleri karşılayıp karşılamaması ve yeterlilik durumlarının incelenmesine yöneliktir. Değerlendirmeye beraber gerçek ortam uygulaması ile maliyet ve zaman faktörleri üzerinden kıyaslama ile bir sınama yapılabilmektedir.

Belirttiğimiz değerlendirme unsurları ile beraber veri madenciliği sonuçları da incelenip değerlendirilmektedir. Veri madenciliği sonuçlarının burada değerlendirilmesinin sebebi veri madenciliği sonuçlarının iş hedefi ile doğrudan bir ilişki ve hedefteki ilişkilerini tüm boyutları ile ortaya çıkarması zorlukları ve faydalı bilgi ve belirtileri içermektedir. Görevle beraber, projede iş hedeflerinin doğruluğu başarı kriterlerine göre ele alınır. Elde edinilen sonuçlar değerlendirilip özetlenir ve son kısımda elde edilen başarı kriterlerini sağlayan model yada modeller ortaya çıkmış olacaktır.

Sürecin gözden geçirilmesi görevi, hedefin gerçekle kıyaslanarak umulan sonuç modelin elde edilme süreci gözden geçirilir. Bu görev gelinen durumda hata olup olmadığı, atlanan bir kısım olup olmadığı, veri oluşumunu sağlayan tercihlerin bir kez daha incelenerek proje başarı ve kalitesinin garanti altına alınması hedeflenmektedir. Bu görev bize gözden kaçan hataların yada

incelenmesi gereken detayların bilgisini geçerek bu işlemlerin kayıt altına alınmasını sağlayacaktır.

Değerlendirme sürecinin en önemli kısmı bu sürecin değerlendirilmesi ve gözden geçirilmesidir. Bu kısım sonuçların değerlendirilmesi ve projeyi bitirmek üzere bir sonraki aşamaya geçişi sağladığı gibi buradan çıkabilecek yeni projelere de referans olmaktadır. Sonraki aşamalara geçiş adımlarının belirlenmesinde yeterlilik ve bütçe açısından karar vermeyi sağlamaktadır. Görevin sonucunda potansiyel eylemler ortaya çıkarılmakta ve bununla beraber gerekçelerin ortaya çıkarılacağı ve bütünü içeren açıklamalı bir rapor hazırlanarak buradan bir karar alınacaktır.

2.2.7. Sonuç

Bu aşama döngü içerisinde yer alan aşamaların sonuncu halkası olarak yer almaktadır. Burada analizciden çok ortaya çıkan sonuçla proje liderlerinin incelemesi, anlaması ve kararlara dâhil olması gereken bir aşamadır (Tablo 2.6). Bu aşamada bilgi açısından oluşturulan yapının başarısından çok kullanıcılar için organize edilmesi ve sunumunun yapılması gerekmektedir.

Tablo 2.6. Sonuç Aşamaları

| Aşama | Sonuç |
|--------------|------------------------------|
| Görev | Sonuç Üzerine Planlama |
| Çıktı | Sonuç Planı |
| Görev | İzlemeyi ve Bakımı Planlamak |
| Çıktı | İzleme ve Bakım Planı |
| Görev | Sonuç Raporu Oluşturmak |
| Çıktı | Sonuç Raporu |
| Çıktı | Sonuç Sunumu |
| Görev | Projeyi Gözden Geçirmek |
| Çıktı | Deneyim Dokümantasyon |

Bu kısım ortaya çıkan raporlamaların sunulması gibi basit bir aşama olsa da projenin canlı sistemler ile uygulanmasına olanak sağlanabilmesi zorluk seviyesini bir hayli yükseltmektedir.

Planlama görevi, veri madenciliği sonuçlarının uygulanabilir hale gelebilmesi için, değerlendirme sonuçlarını alarak bir strateji geliştirmesini ifade eder. Bu görev alınan sonucun uygulanabilirliği için gerekli adımlarını ve nasıl yol alınacağını belirten bir stratejik plan ortaya çıkarmaktadır.

İzleme ve bakım, veri madenciliği ile ortaya çıkarılmış bir projenin canlı bir sistem veya canlı bir sistemin parçası haline getirilmesini sağlamaktadır. İzleme kısmı ayrıntılı bir plan oluşturularak sonucun uygulamaya konulması işleminin oluşumunu sağlar. Bakım kısmı ise oluşabilecek yanlışlıkları ve atıl işlem yükünü ortadan kaldırmak için tasarlanmaktadır. Bu görev projenin yol haritasını ortaya çıkarmaktadır.

Proje sonunda, proje lideri ve takım liderleri tarafından bir sonuç raporu hazırlanır ve sonuç planına proje özetinin genel ve ayrıntılı sunumunu gerçekleştirir. Ayrıca sonuç raporunun uygulamacıya aktarılması da yine bu kısımda gerçekleştirilir.

Projenin gözden geçirilmesi nelerin doğru yapıldığı nelerin yapıldığı, nerelerde düzeltme gerektiği gerekmediği veya ihtiyaçların oluşup oluşamayacağı konusunda son kez değerlendirilir. Bu görev proje sırasında edinilen deneyimlerin özetlenmesiyle sonlanmaktadır.

Deneyim dokümantasyonu projenin her aşamasında, görevler sırasında proje üyelerinin bireysel olarak hazırladığı raporları kapsamaktadır.

Bu işlemler çalışma sonrası yaşanılacak problem tipleri ile ilgili maddeler halinde bilgilendirmeler yapılacaktır. Amaç bütünlük olarak sistemin her koşulu hakkında bilgi sahibi olunması gereğidir.

2.3. VERİ MADENCİLİĞİ PROBLEM TIPLERİ

Bu çalışmada, veri madenciliği problem tiplerine ilişkin olarak CRISP – DM dokümanında altı alt başlık altında incelenmiştir. Buna göre, veri madenciliğinin problem tipleri, veri tanımlama ve özetleme, bölümlenme, kavram tanımlama, sınıflandırma, öngörü ve bağımlılık analizi şeklinde sıralanmıştır.

Bu sıralamayı yapmakla birlikte, veri madenciliği uygulamalarında iş hedeflerine ulaşmak için genellikle birden çok problem türü ile ilgilenilmesinin gerekeceğini vurgulamakta ve birden çok tekniğin kullanılmasını önermektedir.

2.3.1. Veri Tanımlama Ve Özetleme

Veri tanımlama ve özetleme, verilerin genel yapısını ortaya çıkarmaya ve karakteristiklerini tanımlamaya yönelik faaliyetlerin yürütüldüğü veri madenciliği çalışmalarını ifade etmektedir. Bu faaliyetler, veri madenciliği yelpazesindeki en düşük seviye olarak kabul edilmektedir. Bu açıdan, veri tanımlama ve özetleme faaliyetleri tanımlayıcı veri madenciliği çalışmalarının en belirgin örneğini teşkil etmektedir.

Bu açıdan, veri tanımlama ve özetleme bir veri madenciliği projesinin nihai hedefi olabileceği gibi, hemen her veri madenciliği projesinin alt hedeflerinden biri olarak da kabul edilmektedir.

Veri madenciliği sürecinin başlangıcında uygulamacı, verilerin yapısını ve veri madenciliği hedefini kesin olarak bilmemekte, veri içerisinde gizlenmiş potansiyel nitelikteki hipotezleri kurgulamak için veriyi anlamaya dönük başlangıç düzeyinde çalışmalara ihtiyaç duymaktadır. Bu nedenle, basit tanımsal istatistik ve görselleştirme teknikleri verinin içyapısına ilişkin ilk bilgileri sağlayan önemli bir araç olmaktadır.

Veri tanımlama ve özetleme, genellikle diğer veri madenciliği sorunlarıyla birlikte düşünülmekte ve hatta ürettiği sonuçlarla diğer veri madenciliği sorunlarının neler olduğunu da ortaya koymaktadır. Bu nedenle, her veri madenciliği projesi için ilk başta dikkate alınması önerilen faaliyetleri kapsamaktadır.

Yüksek kalitede tanımlama genellikle grafik analizi de içeren keşifsel veri analizi ile mümkün olmaktadır. Keşifsel veri analizi ile verileri derinlemesine araştırmak, değişkenler arası ilişkileri gözden geçirmek, verilerdeki ilginç alt kümeleri belirlemek ve değişkenlerin kendi aralarında ve bağımlı değişkenle birliktelikleri konusunda düşünce üretmek mümkün olmaktadır^[34].

2.3.2. Bölümleme

Bölümleme, verilerin ortak karakteristikleri paylaşan ilgi çekici ve anlamlı alt gruplara ayrılmasını amaçlayan veri madenciliği çalışmalarını ifade etmektedir.

Bu çalışmalarda uygulamacı, kendi bilgi birikimi ya da veri tanımlama ve özetleme çalışmalarında öğrendiklerine dayanarak belirli alt grupları

³⁴ Giudici P. «Applied Data Mining: Statistical Methods for Business and Industry.» John Wiley and Sons, 2003.

tanımlayabileceği gibi, veri içerisinde saklı bulunan ve öngörülemeyen alt grupların tespit edilmesine dayanan kümeleme yöntemlerini de kullanabilmektedir.

Kümeleme analizi veri içerisinde saklı bulunan ancak önceden tanımlanmamış sınıfların üretilmesini amaçlamakta ve sınıflar arası benzerliğin en büyüleştirilmesi ya da sınıf içi benzerliğin en küçükleştirilmesi prensiplerini kullanmaktadır^[35].

Kümeleme, kimi zaman yanlış bir şekilde sınıflandırma ile aynı anlamda kullanılmaktadır. Ancak bu iki yöntem temelde alt veri kümelerinin verilerin kendisinden türetilmesi ya da önceden tanımlanmış olması açısından farklılık göstermektedir.

Kümelemede ortaya çıkan grupların anlamlılığı ancak konunun uzmanı tarafından açıklanabilir. Bu nedenle hangi değişkenlerin dikkate alınarak kümeleme yapılacağı uygulama alanında uzmanlığı gerektirir.

Bölümleme, bir veri madenciliği projesinin temel sorunu olabileceği gibi diğer veri madenciliği sorunlarının çözülmesi için bir adım olarak da düşünülebilir. Bu çalışmalarda amaç, verileri yönetilebilir büyüklükte tutmak ya da analizi kolaylaştırmak açısından daha homojen alt veri kümeleri oluşturmak olabilir.

Bölümleme, tanımlayıcı veri madenciliği çalışmaları için çok bilinen bir örnek olup, sonrasında sınıflandırma çalışmaları da uygulanabilmektedir.

³⁵ Han J., Kamber M., a.g.e.:335

2.3.3. Kavram Tanımlama

Kavram tanımlama, verilerde mevcut bulunan kavramların anlaşılır bir tanımını amaçlayan faaliyetleri ifade etmekte ve model oluşturmaktan çok kavrayışı güçlendirmeye dönük olarak gerçekleştirilmektedir. Bu faaliyetleri, tanımlama ya da öngörü modelleri olarak gerçekleştirmek mümkündür.

Kavram tanımlama, hem bölümlene hem de sınıflandırma ile yakın anlam ifade etmektedir. Ancak bölümlene faaliyetinde ulaşılan grupların anlamlandırılması gibi bir hedef bulunmamaktadır. Ayrıca bölümlene, kavram tanımlamadan önce gerçekleşecek bir faaliyettir. Sınıflandırma ile arasındaki en temel fark ise kavram tanımlamanın tüm veri kümesi üzerinde değil sadece ilgili kavram açısından önem arz eden kayıtlarda uygulanmasının yeterli görülmesidir.

Kavram tanımlama, üzerinde durulan kavrama ilişkin verilerle gerçekleştirilecek tanımlayıcı çalışmaları, bu verilerin yapısı ile geriye kalan verilerin yapısının karşılaştırılmasını ya da bu iki faaliyeti birden içerebilir^[36].

2.3.4. Sınıflandırma

Sınıflandırma, veri kümesindeki kayıtları önceden belirlenmiş alt gruplara ait varsayarak tüm kayıtlara doğru sınıf etiketlerinin atanmasını amaçlayan veri madenciliği çalışmalarını ifade etmektedir. Diğer yandan, sınıflandırma modelleri değişkenlikleri ve veri içerisindeki çeşitli problemleri tanımlamak için de kullanılabilirler.

Sınıflandırma modellerinde bağımlı değişken kesikli olmakta ve bu değişkenin her bir değeri birer sınıf etiketini ifade etmektedir. Sınıflandırma

³⁶ Han J., Kamber M., a.g.e.,:179

modeli, bağımlı değişkenin de değerlerini içeren veri kümesi üzerinden sınıflayıcı kuralları türetmekte ve bu kuralları yeni kayıtlara uygulayarak öngörülerde bulunmaktadır.

2.3.5. Öngörü

Öngörü, hedef değişkenin her bir kayıttaki değerini veri kümesinden edinilen bilgi doğrultusunda doğru olarak belirlemeyi amaçlayan veri madenciliği çalışmalarını ifade etmektedir.

Öngörü ile sınıflandırma arasında bir takım benzerliklerin varlığından söz edilebilse de, iki yaklaşım arasındaki en büyük farklılık hedef değişkenin ölçeği tarafından belirlenmektedir. Çünkü öngörü çalışmasında hedef değişken sınıflandırmada olduğu gibi kesikli değil sürekli ölçektir. Bu nedenle, öngörü modellerine kimi kaynaklarda regresyon modeli adı da verilmektedir. Ayrıca, öngörü bir zaman serisi ile ilgilenmekte ise tahmin söz konusu olacaktır.

2.3.6. Bağımlılık Analizi

Bağımlılık analizi, veri kayıtları veya veri alanları arasında anlamlı bağımlılıkların tespit edilmesine dayanan çalışmaları içermektedir. Bu yöntemlerle, bilgisi verilen kayıtlardan hareketle herhangi bir veri kaydının değerini kestirmek mümkün olmaktadır. Birliktelik tespiti ve ardışık örüntü tespiti, bağımlılık analizinde kullanılan iki yaygın yaklaşımdır. Birliktelik tespiti, veri kayıtları arasındaki kuralları, yani sıklıkla birlikte ortaya çıkan veri parçalarını tespit etmeye çalışır. Bu birliktelikleri tarayan algoritmalar son derece hızlı olup, çok sayıda birliktelik üretebilirler. Buradaki zorluk, en

önemli birlikteliklerin tespit edilmesidir. Benzer şekilde, kuralları zamana bağlı olarak tespit etmeye dayanan analizlere de ardışık örüntü tespiti modelleri adı verilir^[37].

Uygulamada, bağımlılık analizi ile bölümlene sıklıkla birlikte gerçekleşirler. Büyük veri setlerinde, aralarında çok fazla etki yer alması nedeniyle anlamlı bağımlılıklara nadiren rastlanır. Bu tür durumlarda, bağımlılıkların daha homojen alt kümelerde aranması önerilmektedir.

2.4. ALGORİTMALAR

Bu bölümde çalışmada kullandığım veri madenciliği algoritmalarının kısaca tanımları yapılmaktadır. Bu algoritmaların seçimi, veri madenciliği için kullanılan yardımcı programların yeteneklerine ve tez danışmanımın yönlendirmesi sonucu ortak bir karar ile oluşturulmuştur. Bu bölüm detaylı incelemesinden çok tanım niteliğinde olacaktır.

2.4.1. Sınıflandırma Algoritmaları

Sınıflandırma, yeni bir nesnenin niteliklerini inceleme ve bu nesneyi önceden tanımlanmış bir sınıfa atamaktır. Burada önemli olan, her bir sınıfın özelliklerinin önceden net bir şekilde belirlenmiş olmasıdır. Sınıflandırmaya örnek olarak kredi kartı başvurularını düşük, orta ve yüksek risk grubu olarak ayırmak gösterilebilir^[38]. Aşağıda kısaca çalışmada kullanılan algoritmalar tanıtılmıştır.

³⁷ Two Crows Co., a.g.e.

³⁸ Özkan, Dr.Yalçın. «Veri Madenciliği Yöntemleri.»,Papatya Yayıncılık, 2012:40-45

2.4.2. Karar Ağaçları

Karar ağaçları veri oluşturulduktan sonra ağaç kökten yaprağa doğru inilerek kurallar (“Eğer ise kural”) şeklinde yazılabilir^[39]. Karar ağaçlarında kök ve her düğüm bir soruyla etiketlenir. Düğümlerden ayrılan dallar ise ilgili sorunun olası yanıtlarını belirtir. Her dal düğümü de söz konusu sorunun çözümüne yönelik bir tahmini temsil eder. Kural çıkarma, veri madenciliği çalışmasının sonucunu doğrulamak için kullanılır^[40].

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 analiz bilgi verir ve daha sonraki analizler için yol gösterici olabilir^[41].

Karar ağaçları (decision trees), bir tahmin tekniğidir. Genelde sınıflandırma, kümeleme ve tahmin modellerinde kullanılır ve sorunla ilgili araştırma alanını alt gruplara ayırmak için kullanılır. Karar ağaçları eğitici örnekteki veriyi sınavan bir algoritma aracılığıyla gerçekleştirilir ya da sektörün bir uzmanı tarafından oluşturulur. Oluşturma tekniğine bağlı olarak karar ağaçları birbirlerinden ayrılır^[42].

2.4.3. Örnek Tabanlı Yöntemler (En Yakın Komşu)

Sınıflandırma yapılırken eldeki verilerin birbirlerine olan uzaklığı veya benzerlikleri kullanarak sınıflandırmanın gerçekleştirildiği tekniktir. Uzaklık ölçümünde en çok Öklid bağıntısı kullanılır. Çalışma ilkesi; uygun bir mesafe

³⁹ Han J., Kamber M., a.g.e.,:4

⁴⁰ Flawley W.J. , Piatetsky - Shapiro G. , Matheus C.J.,a.g.e:57-70

⁴¹ Clifton C., Thuraisingham B.,a.g.e

⁴² Gökhan Silaharoğlu. «Veri Madenciliği Kavram ve Algoritmaları.» Papatya Yayıncılık, 2013:20-25

ölçüm uzayı belirlenir. Birbirine yakın olan noktalar işaretlenir. Belirlenen grubun en çok rastlandığı sınıf belirlenir ve bu gruplar isimlendirilerek çalışma tamamlanır.

2.4.4. İstatistiksel Sınıflandırıcılar (Naive Bayes)

Bu sınıflandırma tekniği hali hazırda sınıflandırılmış olan verileri kullanarak yeni bir verinin mevcut sınıflardan herhangi birine girme olasılığını hesaplayan bir yöntemdir. Bayes teoremini şu şekilde anlatılabilir:

$$P(C_j | \mathbf{x}) = \frac{p(\mathbf{x} | C_j)P(C_j)}{p(\mathbf{x})} = \frac{p(\mathbf{x} | C_j)P(C_j)}{\sum_k p(\mathbf{x} | C_k)P(C_k)} \quad (41)$$

Burada C_j ve C_k olarak gösterilen iki ayrı hipotezin, başka bir deyişle iki ayrı sınıfın olduğu kabul edilmiştir. $P(C_j|X)$ x 'in C sınıfında olma ihtimalini belirtmektedir. $P(x)$, x değerinin veritabanındaki bulunma sıklığı/sayıdır. $P(C_j)$ ve $P(C_k)$ ise C_j ve C_k sınıflarının veritabanında bulunma sıklığıdır.

Bayes algoritması, öncelikle kendisine verilen öğrenme kümesinde $P(C_j)$ değerini, her sınıfın verilen öğrenme kümesi içinde bulunma olasılığını hesaplar. Daha sonra x 'ler sayılarak $P(x)$ bulunur. Benzer şekilde her bir sınıfta, her bir x değerinin bulunma olasılığı, $P(x|C_j)$, C_j ler içinde x 'lerin sayısı ile elde edilir^[43].

2.4.5. Kümeleme Analizi

Kümeleme analizinde, verilerin birbirlerinden farklı olan kümelerin bulunması ve veritabanındaki kayıtların bu farklı kümelere bölünmesidir.

⁴³ Gökhan Silahtaroğlu, a.g.e: 97-103

Başlangıç aşamasında veritabanındaki kayıtların hangi kümelere ayrılacağı veya kümelemenin hangi değişken özelliklerine göre yapılacağı bilinmemektedir. Kümeleme analizi sonrasında kümeler arasında önceden belirlenemeyen özelliklerin ortaya çıkarılması amaçlanır^[44]. Aşağıda kısaca çalışmada kullanılan algoritmalar tanıtılmıştır.

2.4.6. K-Ortalama Kümeleme Algoritması (K-Means)

K ortalama algoritması adından da anlaşılabilceği gibi giriş uzayını k adet merkezle ifade etmeye çalışan bir yöntemdir. Merkezlere ilk değer ataması rastgele olarak yapıldıktan sonra merkez değerlerinin güncellenmesi için iki farklı yöntem kullanılır. Birinci yöntemde (batch metodu) giriş kümesindeki her bir örneğin hangi merkeze yakın olduğu hesaplanır. Aynı merkeze yakın olan örneklerin ortalaması alınarak merkezin değeri güncellenmiş olur. Durma koşulu sağlanana kadar bu işlem tekrar edilir. İkinci yöntemde (online metodu) giriş kümesinden bir örnek seçilir ve bu örneğin merkezlere olan uzaklığına bakılır. Örneğin en yakın olduğu merkez bulunarak bu merkezin değeri güncellenir. Her bir örnek için bu işlem tekrarlanır. Merkez değeri güncellenirken merkezle örnek arasındaki mesafe değeri her adımda azalan bir öğrenme katsayısıyla çarpılarak kullanılır. Bu sayede ilk adımlarda merkezlerin yer değiştirmesi büyük miktarlarda olurken zamanla yer değiştirme azalır ve merkezler yakınlaşır. Durma şartı sağlanıncaya kadar her örnek için bu işlem tekrarlanır. Örneklerin işleme sırası her adımda aynı olacağı gibi rastgele sırada da olabilir^[45].

⁴⁴ Gökhan Silahtaroglu, a.g.e: 155-163

⁴⁵ Berry M.J.A. , Linoff G.S., a.g.e

III. SPOR VE YAŞAM MERKEZLERİ ÜZERİNE VERİ MADENCİLİĞİ ÇALIŞMASI

Bu bölümde, veri madenciliğinin sınıflandırma ve kümeleme algoritmalarına ilişkin teknikleri spor ve yaşam merkezlerinde kullanılan bir içerik yönetim sisteminin kullandığı Ms Sql veritabanı içerisinde oluşturulan bir veri seti üzerinde uygulanacaktır. Bu şekilde, sektör içerisinde veri madenciliği üzerine çalışması bulunmayan bir dalda çalışma yapılmış olacak hem veri madenciliği süreci gerçekleştirilmiş ve veri madenciliğinin tekniklerinin karşılaştırması yapılmış olacak hem de uygulamada ortaya çıkan sorunlar, değerlendirmeler ve sonuçlar üzerinde yorumlama yapılacaktır.

3.1. AMAÇ, KAPSAM VE YÖNTEM

Çalışmanın amacı, veri madenciliğinin sınıflandırma ve kümeleme fonksiyonuna ilişkin tekniklerini gerçek yaşam verileri üzerinde uygulamak, veri madenciliği sürecini tüm aşamaları ile gerçekleştirmek, uygulamada elde edilen sonuçlarla tekniklere ilişkin farklılıkları tartışmak ve hangi tekniğin hangi koşullarda uygulanmasının daha uygun olacağına yönelik önerilerde bulunmaktır.

Çalışma, spor ve yaşam merkezlerine ait bir veritabanından oluşturulan veri seti kümeleme ve sınıflandırma fonksiyonuna ilişkin tekniklerle veri kümesi üzerinde uygulanacaktır. Bu kapsamda, veri madenciliğinin sınıflandırma teknikleri arasında ölçeklenebilirlikleri ve yaygınlıkları dikkate

alınarak J48 (C4.5), ID3, IB1, IBk, Naive Bayes, kümeleme algoritmasında ise Simple K-Means, Scalable K-Means algoritmaları uygulanmıştır.

Uygulamaya konu olan veri kümesi, bir spor merkezi içerik yönetim sistemi veritabanından belirli kıstaslara göre yedi adet tablo birleşimi ile oluşturulmuştur. Bu oluşturma T-Sql scriptleri ile yapılmıştır. Uygulamada CRISP–DM konsorsiyumu tarafından önerilen veri madenciliği standart süreci esas alınacaktır. Söz konusu veri kümesi üzerinde Weka, Sql Business Intelligence ve ID3 algoritması üzerine kurgulanan bir kodlama uygulaması kullanılarak veri madenciliğinde sınıflandırma ve kümeleme algoritmalarını kapsayan örnek bir uygulama ortaya konacaktır.

3.2. İŞ TANIMI

Bu aşamada, uygulamaya başlamadan önce iş perspektifi açısından sürecin ana hatlarıyla tanımlanması için ulaşılmak istenen iş hedeflerinin belirlenmesi, mevcut durumun değerlemesi, veri madenciliği amaçlarının belirlenmesi ve proje planının hazırlanması görevleri üzerinde durulacaktır.

Son yıllarda spor ve yaşam merkezlerinin yurt dışında olduğu gibi ülkemizde de ticari sektör haline gelmesiyle ciddi anlamda maddi gelirlerin sağlandığı ve rekabetin bol olduğu bir yarışın yaşanıldığı bir sektör haline gelmiştir. Hizmet sektörü olması itibariyle müşteri memnuniyeti, müşteri sadakati ve mevcut kazançların arttırılması sektörün önemli konuları haline gelmiştir.

Bu rekabet ortamında başarılı olmak, mevcut müşterileri tanımak, onlara istedikleri düzeyde hizmet sunmak ve mevcut müşterilerden öğrenilen

bilgi ışığında potansiyel müşteriler için doğru yaklaşımlar geliştirmek ile mümkün olacaktır.

Bu bakış açısıyla, spor ve yaşam merkezinde kullanılan içerik yönetim sisteminin veritabanı üzerinde gerçekleştirilecek veri madenciliği uygulaması ile ulaşılmak istenen iş hedefi, mevcut müşterilerin üyelik yenileme durumları üzerine yoğunlaşacak ve yenileme işlemlerini tetikleyen ve gözden kaçan ayrıntıların bulunmasını sağlamak olarak belirlenmiştir.

Böyle bir çalışmadan sektörün beklentisi, müşterilerin üyelik yenileme süreçlerini kontrol altına alma, üyelik sahibi bireylerin üyeliklerini yenilemeleri adına karar destek sistemlerinin yönlendirme işlemlerini ya da yeni üyelik satışlarında ortaya çıkan potansiyelin doğru planlama ile gelire dönüştürülmesi olarak modellenmektedir.

Ancak bu çalışmanın temel amacı müşterilerin tercih ve davranışlarını temsil kabiliyeti yüksek modellerin elde edilebilmesini ve bu modelin potansiyel kazancın artırılmasına yönelik sektöre artı bir dinamizm getireceği konusunda beklentilere cevap verebilecek nitelikte olması temenni edilmektedir.

Bu çalışma herhangi bir ekip çalışması olmayıp kısıtlı olanaklarla gerçekleştirildiğinden ideal bir veri madenciliği projesi olmaktan çok, daha önce de belirtildiği gibi bir örnek uygulama niteliğindedir. Benzer şekilde, üzerinde çalışılan veri kümesi de büyüklük açısından ideal olmamakla birlikte asgari yeterlilikleri sağlamaktadır.

Uygulamada kullanılacak olan veri kümesi an itibari ile hala faal bir işletme olması ve müşterilerin kişisel bilgileri ve işletmenin kazanç bilgilerini içermesi nedeniyle gizlilik içermektedir. Bu açıdan ilgili işletme ve müşterilerine ilişkin bilgiler paylaşılmayacaktır.

Her veri madenciliği uygulamasında olduğu gibi başlangıç düzeyinde varsayılan süreç ve iş hedeflerinin gerçekleşmesi bu çalışmada da çeşitli nedenlerle mümkün olmayabilir. Çalışmanın başarısı üzerindeki en önemli tehdit veri kümesinden kaynaklanabilecek sorunlardır. Bu durum tekniklerin uygulanmasına engel oluşturmaya da sonuçların elde edilmesinde problemlere neden olabilir.

Veri kümesinden kaynaklanabilecek bu türden sorunların yaşanması durumunda alternatif olarak uygulama daha fazla kaydın elde edilmesi ya da yeni bir veri kümesi edinilmesi yoluyla oluşabilecek sorunların önüne geçilmesi sağlanabilir.

Yine uygulamanın başarıya ulaşmasını geciktirebilecek unsurlardan biri de teknik bir takım sınırlamalardır. Uygulanmada mevcut donanım ve yazılımların kullanılması yeterli olmaz ise alternatif yazılım ve donanımların kullanılacaktır.

Veri kümesi üzerinde veri madenciliğinin sınıflandırma teknikleri ve kümeleme algoritmaları için tek tek gerçekleşmesi, geliştirilen öngörü modelleri ışığında bu tekniklerin ilgili amaç doğrultusunda kullanılabilirliğinin sınanması ve tekniklerin karşılaştırılması imkânının yaratılması çalışmanın veri madenciliği açısından diğer amaçlarını oluşturmaktadır.

Veri kümesi üzerinde ilgilenilen tüm tekniklerin başarı ile uygulanması ve sonuçları açısından bu tekniklerin karşılaştırılabilmesini sağlayacak ölçülerin elde edilmesi durumunda arzulanan veri madenciliği başarısı elde edilmiş olacaktır.

3.3. VERİLERİ ANLAMA

Bu aşamada, belirlenen iş ve veri madenciliği hedefleri doğrultusunda veri kümesinin incelenmesi, sorunlarından arındırılması ve analize hazır hale getirilmesi için başlangıç verilerini elde etme, verileri tanımlama, verileri inceleme ve veri kalitesini sınaama görevleri üzerinde durulacaktır.

Uygulamada kullanılacak olan ham veri kümesi Microsoft Sql Server veritabanında bulunmaktadır. Veri hazırlama işlemlerinin hepsi bu aşamada yapılmıştır. Burada hazırlanan veri kümesi veri madenciliği yazılımlarında kullanabilmesi adına Microsoft Excel çalışma sayfasına aktarılmıştır. Bu sorgulama işlemleri veriyi hazırlama aşamasında detaylı olarak gösterilecektir. Burada Excel'in seçilmesi, ilerleyen görevlerde ortaya çıkacak veri işleme ihtiyaçları doğrultusunda süzme, sıralama, özetleme gibi pratik fonksiyonları desteklemesinden kaynaklanmaktadır. Ayrıca Weka programının veri seti olarak kabul ettiği dosya formatı olan arff için dönüştürme işlemleri Excel dosyasının başka bir tarafından kullanılarak oluşturulmasından dolayı yapılmaktadır.

Satırlarda yer alan her bir kayıt, bir müşterinin ilgili sütunlardaki değişkenlere karşılık gelen değerlerinin birleşimini ifade etmektedir. Sütunlarda yer alan her bir değişken ise tüm müşterilerin aynı ölçüme karşılık gelen değerlerinin topluluğunu ifade etmektedir.

Değişkenlerin büyük bir bölümü nümerik olarak gruplandırılmış kategorik veriler içerirken, az sayıda değişken ise metin verileri içermektedir. Daha önce belirtildiği üzere müşterilerin kişisel bilgi gizliliği ve veri kümesinin ilgili işletmenin kazanç bilgilerini içermesi nedeniyle müşterilerin kimlik bilgileri yer almayacaktır. Ancak değişkenlerin bilgi düzeyleri hakkında kısaca bilgi verilecektir. Çalışmada kullanılacak olan veri seti (Tablo 3.1) niteliklerin kısaca açıklamaları ile beraber verilmiştir. Ayrıca niteliklerin alabilecekleri değerlerde ayrı olarak (Tablo 3.2) gösterilmiştir.

Tablo 3.1. Değişkenlere İlişkin Bilgilendirme

| Nitelik Adı | Açıklama |
|--------------------------|--|
| S.F. | Üyelik sözleşme fiyatı |
| Asil Üye T.G.S. | Asil üyenin tesise geliş sayıları |
| Aile Üyesi T.G.S. | Aile üyesinin tesise geliş sayıları |
| Asil Üye S.K.S | Asil üyenin Spa kullanım sayısı |
| Aile Üyesi S.K.S | Aile üyesinin Spa kullanım sayısı |
| Asil Üye H.K.S | Asil üyenin havuz kullanım sayısı |
| Aile Üyesi H.K.S | Aile üyesinin havuz kullanım sayısı |
| Asil Üye G.K.S | Asil üyenin grup ders kullanım sayısı |
| Aile Üyesi G.K.S | Aile üyesinin grup ders kullanım sayısı |
| Asil Üye O.K.S | Asil üyenin özel ders kullanım sayısı |
| Aile Üyesi O.K.S | Aile üyesinin özel ders kullanım sayısı |
| Kaynak | Üyelerin hangi kaynaktan tesise geldiğinin bilgisi |
| Lokasyon | Üyelerin ikamet ettiği ilçe bilgisi |
| Üyelik Yenilendi | Üyenin üyeliğini yenileyip yenilenmediği bilgisi |

Tablo 3.2. Veri Seti Nitelik Değer Aralıkları

| NİTELİK ADI | DEĞER ARALIKLARI | AÇIKLAMA |
|--------------------------|---|--|
| S.F. | (KATEGORİK) 1,2,3,4,5,6,7,8,9,10 | Nitelik Değeri / 1000 |
| Asil Üye T.G.S. | (KATEGORİK) 1,2,3,4,5,6,7,8,9,10 | Nitelik Değeri / 5 |
| Aile Üyesi T.G.S. | (KATEGORİK) 1,2,3,4,5,6,7,8,9,10 | Nitelik Değeri / 5 |
| Asil Üye S.K.S | (KATEGORİK) 1,2,3,4,5,6,7,8,9,10 | Nitelik Değeri / 5 |
| Aile Üyesi S.K.S | (KATEGORİK) 1,2,3,4,5,6,7,8,9,10 | Nitelik Değeri / 5 |
| Asil Üye H.K.S | (KATEGORİK) 1,2,3,4,5,6,7,8,9,10 | Nitelik Değeri / 5 |
| Aile Üyesi H.K.S | (KATEGORİK) 1,2,3,4,5,6,7,8,9,10 | Nitelik Değeri / 5 |
| Asil Üye G.K.S | (KATEGORİK) 1,2,3,4,5,6,7,8,9,10 | Nitelik Değeri / 5 |
| Aile Üyesi G.K.S | (KATEGORİK) 1,2,3,4,5,6,7,8,9,10 | Nitelik Değeri / 5 |
| Asil Üye O.K.S | (KATEGORİK) 1,2,3,4,5,6,7,8,9,10 | Nitelik Değeri / 5 |
| Aile Üyesi O.K.S | (KATEGORİK) 1,2,3,4,5,6,7,8,9,10 | Nitelik Değeri / 5 |
| Kaynak | (KATEGORİK) BASIN,BILLBOARD,SMS,TA VSIYE BROSUR,INTERNET,STAND, TOPLUMAIL,TVREKLAMI | Müşteri kartındaki kaynak alanı temel alınarak gruplandırılmıştır. |
| Lokasyon | (KATEGORİK) ATASEHIR,BATIATASEHIR, BOSTANCI,GOZTEPE,ICERE NKOY, KADIKOY,KOZYATAGI,KU CUKYALI, MALTEPE,USKUDAR | Müşteri kartındaki semt alanı temel alınarak gruplandırılmıştır. |
| Üyelik Yenilendi | 1,0 | - |

Nitelikler oluşturulurken değerlerin ölçülebilir hale gelmesi ve veri madenciliği tekniklerinin başarı oranını arttırabilmek adına gruplandırılmıştır. Nümerik alanlar için nitelik değerleri bölünerek grup oluşturulmaya çalışılmıştır. Yukarıdaki tablo (Tablo 3.2) çalışmanın daha anlaşılır olabilmesi açısından değişken grupları hakkında bazı özet bilgileri içermektedir.

Bu değişkenler hedef değişken olarak seçilmiş olmakla birlikte sınıflandırmada üyelik yenileme alanı sınıf olarak kullanılacaktır.

Veri kümesi üzerinde yapılan ilk inceleme sonucunda bir takım veri kalitesi sorunları tespit edilmiştir. Kayıp veri değerleri sorununun hem değişken bazında (örn.bazı değişkenlerin dinamik olarak oluşturulması) hem de kayıt bazında (örn. bazı müşterilerin ikamet bilgisi içeren lokasyon değişkenine ait alanı bilinmeyen kayıtların olması) yaygın olduğu görülmüştür. Ayrıca bazı çelişkili veri değerleri olduğu (örn. Aynı kişiye ait birden fazla müşteri kartının olması ve farklı üyelik tiplerine ait sözleşmesi bulunması vb.), bazı veri değerlerinin yanlış kodlandığı (örn. stanbul ya da İstanbul, Erkek ya da Bay vb.) tespit edilmiştir.

Bilindiği gibi birçok veritabanı güncel olmayan verileri de içermektedir. Özellikle yeni gelişen bir sektör olması verilerin bilgisayar ortamında tutulması ile beraber yazılı olarak tutulması veri kaynağı açısından eksik bilgilerinde var olabileceği gerçeğini ortaya çıkarmaktadır. Ayrıca bu ortamlardan verilerin aktarılması, farklı içerik yönetim sistemleri ile olabilecek entegrasyonlar ve veri yönetimi konusunda kullanılan yazılımların güncellenmesi gibi durumlarda da veri kalitesi sorunları oluşabilmektedir. Bu türden sorunların üzerinde çalışılan veri kümesi için de geçerli olduğu görülmüştür.

Veri kalitesini yükseltmek açısından yapılacak ön çalışma ile hem kayıt bazında hem de değişken bazında bazı çözümler öngörülmektedir. Bu çözümlerden biri eleme yöntemine başvurmak olacaktır. Değişken bazında düşünüldüğünde bilgi düzeyi çok düşük ve kayıp veri değerleri sorunu çok yüksek olan değişkenler tespit edilerek elenecektir. Eleme işlemi sonucunda ortaya çıkan veri kümesinin kalitesi görece daha yüksek düzeyde olacak ve bu aşamadan sonra veri kalitesi sorunlarının veri kaybına neden olmadan çözülmesine çalışılacaktır. Bu amaçla değişken bazında öngörülen ilk çözüm kayıp veri değerleri yerine kestirimde bulunulması olacaktır. Ayrıca veri kalitesini yükseltme çalışmalarında hedef değişkenlerin sınıflandırılması ile kalitenin istenilen seviyede tutulmasına çalışılmıştır.

3.4. VERİLERİ HAZIRLAMA

Bu aşamada, üzerinde veri madenciliği tekniklerinin uygulanacağı nihai veri kümesinin oluşturulması amacıyla verileri seçme, verileri temizleme, verileri yapılandırma, verileri birleştirme ve verileri biçimlendirme görevleri üzerinde durulacaktır.

Ayrıca bu aşama sonunda, elde edilecek nihai veri kümesine ilişkin tanımlayıcı nitelikte istatistiklerin sunulması ile modelleme aşaması öncesinde eldeki veri kümesinin daha detaylı olarak anlaşılması sağlanacaktır. Çalışmada kullanacağımız müşteri veri setimiz aşağıdaki niteliklerin veri setine dahil değeri evet olanlardan oluşmaktadır (Tablo 3.3).

Tablo 3.3.Müşteri Veri Seti

| ALAN ADI | AÇIKLAMASI | Veri Setine Dâhil |
|-------------------------------------|--|-------------------|
| SozlesmeFiyati | Üyelik sözleşmesi fiyatı temsil etmektedir. | Evet |
| AsilUyelikSatisId | Üyelik tablosunda asil üyeyi belirten satırı temsil etmektedir. | Hayır |
| AsilUyeId | Üyelik tablosunda asil üye müşteri kartı bağlantısını belirten satırı temsil etmektedir. | Hayır |
| SonrakiUyelikSatisId | Üyelik yenilendiyse, yeni oluşturulan sözleşmenin bağlantısını temsil etmektedir. | Hayır |
| BaslangicTarihi | Üyeliğin başlangıç tarihidir. | Hayır |
| BitisTarihi | Üyeliğin bitiş tarihidir. | Hayır |
| TesiseGelisAsilUye | Asil üyenin tesise geliş sayılarıdır. | Evet |
| TesiseGelisSayisiAilesi | Aile üyesinin tesise geliş sayılarıdır. | Evet |
| SpaKullanimiAsilUye | Asil üyenin spa kullanım sayısıdır. | Evet |
| SpaKullanimiAileUyesi | Aile üyesinin spa kullanım sayısıdır. | Evet |
| HavuzKullanimSayisiAsilUye | Asil üyenin havuz kullanım sayısıdır. | Evet |
| HavuzKullanimSayisiAileUyesi | Aile üyesinin havuz kullanım sayısıdır. | Evet |
| GrupDersKullanimiAsilUye | Asil üyenin grup ders kullanım sayısıdır. | Evet |
| GrupDersKullanimiAileUyesi | Aile üyesinin grup ders kullanım sayısıdır. | Evet |
| OzelDersKullanimiAsilUye | Asil üyenin özel ders kullanım sayısıdır. | Evet |
| OzelDersKullanimiAileUyesi | Aile üyesinin özel ders kullanım sayısıdır. | Evet |
| Kaynak | Üyelerin hangi kaynaktan tesise geldiğinin bilgisidir. | Evet |
| Lokasyon | Üyelerin ikamet ettiği ilçe bilgisidir. | Evet |
| UyelikYenileme | Üyenin üyeliğini yenileyip yenilenmediği bilgisidir. | Evet |

Tablo 3.4. Müşteri Kartı : Müşterilere Dair Detayların Yer Aldığı Tablo Bilgisi

| ALAN ADI | AÇIKLAMASI |
|----------------------|---|
| FirmaID | Müşterinin tabloda temsil ettiği satır bilgisidir. |
| Semt | Müşterinin hangi semtte ikamet ettiği bilgisidir. |
| MusteriAdi | Müşteri adı alanıdır. |
| MusteriSoyadi | Müşteri soyadı alanıdır. |
| KaynakId | Müşterinin hangi kaynaktan tesise geldiğinin bilgisini temsil etmektedir. |

Tablo 3.5. Üyelik Sözleşmesi Kartı: Üyelik Sözleşmelerine Dair Detayların Yer Aldığı Tablo Bilgisi

| ALAN ADI | AÇIKLAMASI |
|---------------------------------|---|
| UyelikSatisID | Sözleşmenin tabloda temsil ettiği satır bilgisidir. |
| AnaUyelikSatisID | Sözleşmedeki aile üyelerinin bireysel sözleşmeler ile ayrımının yapıldığı alanıdır. |
| SozlesmeKodu | Sözleşmeye ait numara alanıdır. |
| MusteriId | Sözleşmenin müşteri bağlantısıdır. |
| HizmetID | Sözleşmenin hangi hizmet üzerinden yapıldığının bilgisidir. |
| Tutari | Sözleşmeye ait tutar bilgisi |
| BaslangicTarihi | Sözleşmenin başlangıç tarihidir. |
| BitisTarihi | Sözleşmenin bitiş tarihidir. |
| UyelikSozlesmesiDurumuID | Sözleşmenin mevcut durumunu ifade etmektedir. |
| OnceliUyelikSatisID | Mevcut sözleşme bir başka sözleşmeden yükseltilecek yapıldıysa bu alan sözleşmeler arası bağlantıyı yapmaktadır.. |
| SonrakiSozlesmeId | Mevcut sözleşmeden başka bir sözleşmeye yükseltiliyorsa sözleşmeler arası bağlantıyı bu alan ile yapılmaktadır. |

Yukarıda belirtilen tablolar çalışmanın temelini oluşturan tablolardır.

Bu tablolar veri seti oluşturduğum scriptlerden de takip edileceği gibi sınıf niteliğini de oluşturmaktadır. Tablolarda fazladan olan alanlar hesaplamalar için eklenmiştir.

Tablo 3.6. Giriş – Çıkış Tablosu : Tesis ve Havuz Kullanımlarının Yer Aldığı Tablo Bilgisi

| ALAN ADI | AÇIKLAMASI |
|---------------------------|---|
| KisiGuid | Üyeliğe sahip kişinin müşteri kartını temsil etmektedir. |
| GirisSaati | Kişinin tesise veya havuz alanına girişini temsil etmektedir. |
| CikisSaati | Kişinin tesise veya havuz alanına çıkışını temsil etmektedir. |
| MekanID | Kişinin tesise mi yoksa havuza mı giriş yaptığının bilgisidir. |
| GirisCihazID | Kişinin tesise veya havuz alanına girişini yaptığı cihazı belirtmektedir. |
| CikisCihazID | Kişinin tesise veya havuz alanından çıkış yaptığı cihazı belirtmektedir. |
| UyelikSozlesmesiID | Kişinin geçişlerde kullandığı üyeliği temsil etmektedir. |

Tablo 3.7. Plan Rezervasyonları: Spa ve Özel Ders Hizmeti Alan Müşteri Bilgileri İçin
Kullanılan Tablo Bilgisi

| ALAN ADI | AÇIKLAMASI |
|-----------------------|--|
| PlanID | Plan tablosundaki rezervasyonu temsil etmektedir. |
| MekanID | Rezervasyonun hangi mekânda yapıldığını belirtmektedir. |
| HizmetID | Rezervasyon içeriğinde hangi servis olduğunu belirtmektedir. |
| BaslangicSaati | Rezervasyonun başlangıç saatidir. |
| BitisSaati | Rezervasyonun bitiş saatidir. |

Tablo 3.8. Plan Satışlarına Dair Detayların Yer Aldığı Tablo Bilgisi

| ALAN ADI | AÇIKLAMASI |
|--------------------------|--|
| PlanSatisId | Satın alınmış planı temsil etmektedir. |
| PlanId | Kişinin rezervasyon planını temsil etmektedir. |
| MusteriID | Satın alınan hizmetin müşterisini temsil etmektedir. |
| HizmetID | Satın alınan hizmeti temsil etmektedir. |
| PlanSatisDurumuID | Planın yapılıp yapılmadığına ait durum bilgisidir. |
| ServisTarihi | Hizmetin verilme tarihini belirtmektedir. |

Tablo 3.9. Grup Ders Seansları: Grup Derslerinin Takip Edildiği Tablo Bilgisi

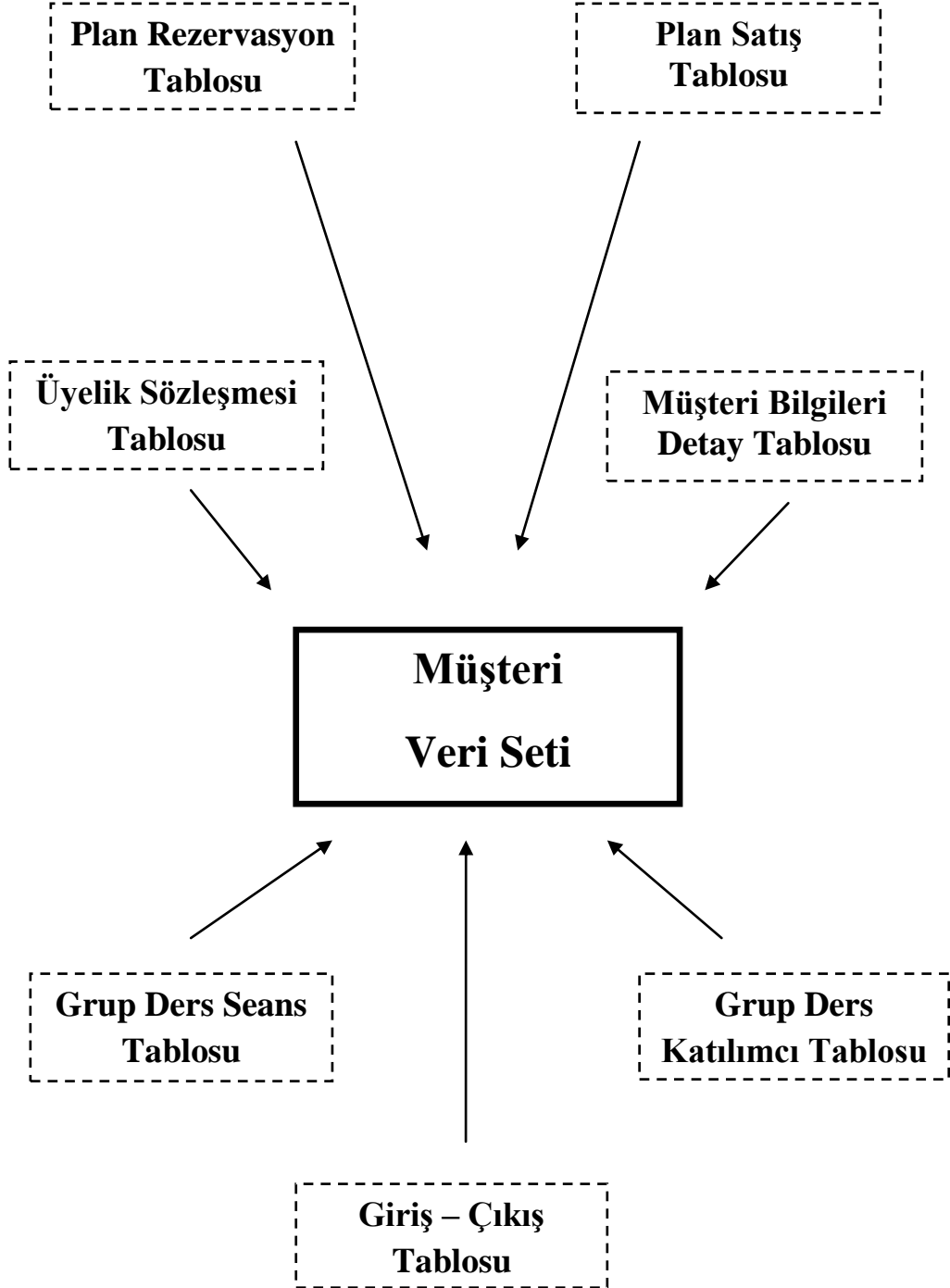
| ALAN ADI | AÇIKLAMASI |
|--------------------------------|--|
| GrupDersİd | Hangi dersin verildiğini belirtmektedir. |
| Seansİd | Hangi sınıfa dahil olduğunu bilgisidir. |
| SeansVarsayılanBasSaati | Grup dersine ait sınıfın başlangıç saatidir. |
| SeansVarsayılanBitSaati | Grup dersine ait sınıfın bitiş saatidir. |
| SeansDurumuID | Sınıfın tamamlanma durumunu temsil eder. |
| SeansTarihi | Sınıfın hangi tarihte işlem yapacağını belirtir. |

Tablo 3.10. Grup Ders Seans Katılımcılarına Dair Detayların Yer Aldığı Tablo Bilgisi

| ALAN ADI | AÇIKLAMASI |
|----------------------|---|
| KatilmciID | Seansa katılan kişiyi temsil eder. |
| SeansID | Hangi seans üzerinde işlem yapılacağını belirtir. |
| DersID | Hangi dersin verildiğini belirtmektedir. |
| MusteriID | Kişinin müşteri bilgilerini temsil eder. |
| UyelikSatisID | Kişinin üyelik bilgilerini temsil eder. |
| Katildi | Kişinin seansa katılıp katılmadığı bilgisidir. |

Çalışmada kullanılan tabloların birçoğu birbiri ile bağlantılı ve hepside veri seti ile doğrudan bağlantılı olacak bir yapıdadır (Şekil 2.5). Bu çalışma yapısının bütün yeni eklentileri ve düzeltmeleri uyumlu halde olmasının sebebi budur. Yeni eklenecek verilerin oryantasyonu çalışmada bulunan scriptlerin uyumlu şekilde yeni veri setini oluşturmasını sağlar. Böylelikle çok fazla zaman harcanmadan yeni oluşum tamamlanmış olmaktadır.

Şekil 3.1. Müşteri Veri seti ile Tablolar Arası İlişki Diyagramı



Asil Üye İçin Müşteri Listesinin İçeriğini Oluşturan Sorgu

- Sql cursor için kullanılacak olan değişken tanımları

```
Declare @UyelikSatisID Int
Declare @AsilMusteriID Int
Declare @BaslangicTarihi Datetime
Declare @BitisTarihi Datetime
Declare @SonrakiSozlesmeID Int
Declare Curl Cursor For Select UyelikSatisId, MusteriID,
BaslangicTarihi, BitisTarihi, SonrakiSozlesmeId From
LP_UyelikSatislari
Where AnaUyelikSatisID is null And UyelikSozlesmesiDurumuID
not in (16,32,512)
```

- Sql cursor için sorguya sadece asil üyeler dahil edilmiştir.

“ AnaUyelikSatisID is null ”

- Sorgu ile oluşturulacak sözleşme listesine sözleşme durumu olarak üyeliği başlamayanlar , iptal edilenler ve onaylanmamış sözleşmeler dahil edilmemiştir.

“ UyelikSozlesmesiDurumuID Not In (16,32,512) ”

```
Open Curl
Fetch Next From Curl Into @UyelikSatisID, @AsilMusteriID,
@BaslangicTarihi, @BitisTarihi, @SonrakiSozlesmeID
While (@@Fetch_Status = 0)
Begin
```

Tesis Kullanımlarının Hesaplanması:

- Aynı gün içerisinde birden fazla girişler göz ardı edilmiştir.
- Kullanım sayıları oluşturulurken kategorizasyon için beşe bölünmüştür.“ Count(GecisId)/5 ”

- Tesis kullanımı, havuz kullanımı ile aynı tablo içerisinde hesaplandığı için mekana göre ayırım yapılmıştır.“ GK_Gecisler.MekanId = 1 ”

Sorgu :

```
Insert Into MusteriListesi ( TesiseGelisAsilUye,
AsilUyelikSatisId, AsilUyeId )
Select Count(GecisId)/5, @UyelikSatisID, @AsilMusteriID
From GK_Gecisler
Where (GK_Gecisler.UyelikSozlesmesiId = @UyelikSatisID) And
(GK_Gecisler.MekanID = 1)
```

Spa Kullanımlarının Hesaplanması :

- Spa kullanımı hesaplamalarında hizmet gruplarına göre hizmetler baz alınmıştır.“ (LP_HizmetGruplari.HizmetGrupID = 1) ”
- Her hizmet kullanımı dikkate alınmıştır.
- Hizmet kullanım sayısını oluşturabilmek için hizmetin müşteri tarafından kullanılmış olması gerekmektedir.Bu kıstas plan durumu kontrolü ile sağlanmıştır.“ PlanSatisDurumuID = 8 ”
- Kullanım sayıları oluşturulurken kategorizasyon için beşe bölünmüştür.“ Count (PlanSatisId)/5 ”
- Hizmetin üyelik süresi içerisinde kullanılmış olmasına dikkat edilmiştir.“ (LP_PlanSatislar.ServisTarihi >= @BaslangicTarihi And LP_PlanSatislar.ServisTarihi <= @BitisTarihi) ”

Sorgu :

```
Update MusteriListesi
Set SpaKullanimiAsilUye = SpaPlanSayilari.ToplamSpaHizmeti
From MusteriListesi
Inner Join (
```

```

Select      Count      (PlanSatisId)/5      ToplamSpaHizmeti      From
LP_PlanSatislar      Inner      Join      LP_Hizmetler      On
LP_PlanSatislar.HizmetId = LP_Hizmetler.HizmetId Inner join
LP_HizmetGruplari      On      LP_Hizmetler.HizmetGrupId      =
LP_HizmetGruplari .HizmetGrupId
Where
(LP_PlanSatislar.MusteriId = @AsilMusteriID) And
(LP_HizmetGruplari.HizmetGrupID = 1) And
(LP_PlanSatislar.ServisTarihi >= @BaslangicTarihi And
LP_PlanSatislar. ServisTarihi <= @BitisTarihi) And
PlanSatisDurumuID = 8) SpaPlanSayilari On
MusteriListesi.AsilUyeID = SpaPlanSayilari.MusteriId
Where AsilUyelikSatisId = @UyelikSatisID

```

Havuz Kullanımlarının Hesaplanması :

- Kullanım sayıları oluşturulurken kategorizasyon için beşe bölünmüştür.“ Count(GecisId)/5 “
- Havuz kullanımı, tesis kullanımı ile aynı tablo içerisinde hesaplandığı için mekana göre ayırım yapılmıştır. “ GK_Gecisler.MekanID in (19,20) ”

Sorgu :

```

Update MusteriListesi
Set      HavuzKullanımSayisiAsilUye      =      HavuzKullanımı.
ToplamHavuzKullanımı
From      (Select      Count (GecisId)/5      ToplamHavuzKullanımı,
@UyelikSatisID, @AsilMusteriID From GK_Gecisler
Where GK_Gecisler.UyelikSozlesmesiId = @UyelikSatisID And
GK_Gecisler.MekanID in (19,20) ) HavuzKullanımı
On MusteriListesi.AsilUyeID = HavuzKullanımı.MusteriId
Where AsilUyelikSatisId = @UyelikSatisID

```

Grup Ders Kullanımlarının Hesaplanması :

- Kullanım sayıları oluşturulurken kategorizasyon için beşe bölünmüştür.“ Count(KatilimciId)/5 ”
- Hesaplamanın yapılabilmesi için grup dersine ait sınıfın durumunun tamamlandı olması gerekmektedir.“ LP_GrupDersleriSeanslari.SeansDurumuId = 4 ”
- Hesaplamada diğer kıstas kişinin sınıfa katılmış olması gerekmektedir.“ Katildi = 1 ”

Sorgu :

```
Update MusteriListesi
Set GrupDersKullanimiAsilUye =
(
    Select
    Count
    (KatilimciId)/5
    From
    LP_GrupDersleriKatilimcilar
    Inner
    Join
    LP_GrupDersleriSeanslari
    On
    LP_GrupDersleriKatilimcilar.SeansId
    =
    LP_GrupDersleriSeanslari.SeansId
    Where UyelikSatisID = @UyelikSatisID And
    LP_GrupDersleriSeanslari.SeansDurumuId = 4 And Katildi = 1 )
Where AsilUyelikSatisId = @UyelikSatisID
```

Özel Ders Kullanımlarının Hesaplanması :

- Kullanım sayıları oluşturulurken kategorizasyon için beşe bölünmüştür.“ Count (PlanSatisId)/5 ”
- Özel ders hesaplamalarında hizmet gruplarına göre hizmetler baz alınmıştır.“ (LP_HizmetGruplari.HizmetGrupID in (2,3,4)) -- Pilates&Yoga, Cilt Bakımı, Pt “

- Hizmetin üyelik süresi içerisinde kullanılmış olmasına dikkat edilmiştir.“ (LP_PlanSatislar.ServisTarihi >= @BaslangicTarihi And LP_PlanSatislar.ServisTarihi <= @BitisTarihi) ”
- Her hizmet kullanımı dikkate alınmıştır.
- Hizmetin kullanım sayısını oluşturabilmek için hizmetin müşteri tarafından kullanılmış olması gerekmektedir.Bu kıstas plan durumu kontrolü ile sağlanmıştır.“ PlanSatisDurumuID = 8 ”

Sorgu :

```

Update MusteriListesi
Set      OzelDersKullanimiAsilUye      =      OzelDersPlanSayilari.
ToplamOzelDersHizmeti From MusteriListesi
Inner Join (
Select Count (PlanSatisId)/5      ToplamOzelDersHizmeti From
LP_PlanSatislar
Inner  Join  LP_Hizmetler  On  LP_PlanSatislar.HizmetId  =
LP_Hizmetler.HizmetId
Inner  Join  LP_HizmetGruplari  On  LP_Hizmetler.HizmetGrupId  =
LP_HizmetGruplari.HizmetGrupId
Where
(LP_PlanSatislar.MusteriId = @AsilMusteriID) And
(LP_HizmetGruplari.HizmetGrupID in (2,3,4)) And
(LP_PlanSatislar.ServisTarihi      >=      @BaslangicTarihi      And
LP_PlanSatislar.ServisTarihi      <      =      @BitisTarihi)      And
PlanSatisDurumuID = 8 ) OzelDersPlanSayilari
On MusteriListesi.AsilUyeID = OzelDersPlanSayilari.MusteriId
Where AsilUyelikSatisId = @UyelikSatisID

```

Kaynak & Lokasyon Niteliklerine Değer Ataması:

```

Update MusteriListesi
Set Kaynak = MI_Kaynaklar.Kaynak, Lokasyon = FN_Firmalar.Semt

```



```

From MusteriListesi
Inner Join FN_Firmalar On MusteriListesi.AsilUyeId =
FN_Firmalar.FirmaId
Inner Join MI_Kaynaklar On FN_Firmalar.KaynakId =
MI_Kaynaklar.KaynakID
Where AsilUyeId = @MusteriID

```

Üyelik Yenileme Niteliğine Değer Ataması:

```

Update MusteriListesi
Set UyelikYenileme = (case when @SonrakiSozlesmeID <> 0 then 1
else 0 end)Where AsilUyelikSatisId = @UyelikSatisID
Fetch next From curl Into @UyelikSatisID, @AsilMusteriID,
@BaslangicTarihi, @BitisTarihi, @SonrakiSozlesmeID
End
close curl
Deallocate curl

```

Aile Üyesi İçin Müşteri Listesinin İçeriğini Oluşturan Sorgu

- Sql cursor için kullanılacak olan değişken tanımları

```

Declare @UyelikSatisID Int
Declare @BaslangicTarihi Datetime
Declare @BitisTarihi Datetime
Declare @SonrakiSozlesmeID Int

```

- Cursor içerisinde işlemleri kolaylaştırmak adına aşağıdaki aile üyesi tablosu oluşturulmuştur.

```

Create Table #AileUyeleri
(
AsilUyesiUyelikSatisId int,
AileUyesiID int,
AileUyesiUyelikSatisId int,
BaslangicTarihi datetime,

```

```

BitisTarihi datetime )
Declare curl Cursor For
Select AsilUyelikSatisId From MusteriListesi

```

- Asil üye için kayıtlar müşteri tablosuna aktarıldıktan sonra aile üyesi için işlemleri kısaltmak adına direk müşteri tablosu ile cursor sorgusu oluşturulmuştur.

```

Open Curl
Fetch Next From Curl Into @Uyeliksatisid
While (@@Fetch_Status = 0)
Begin

```

- Yukarıda Oluşturulan Aile Üyeleri Tablosuna Asil Üyelerine Göre Kayıtlar Atılmaktadır.

```

Insert Into #Aileuyeleri
Select      @Uyeliksatisid,      Musteriid,      Uyeliksatisid,
Baslangictarihi, Bitistarihi From Lp_Uyeliksatislari
Where AnaUyeliksatisid = @Uyeliksatisid

```

Tesis Kullanımlarının Hesaplanması :

- Aile üyelerinin tesise gelişlerinin hesaplanması için her bir aile üyesinin geliş sayıları teker teker hesaplanmıştır. Yapılan hesaplamanın sisteme uyumluluğu için ortalaması alınmıştır.“ Avg(AileKullanımSayisi) ”
- Aynı gün içerisinde birden fazla girişler göz ardı edilmiştir.
- Kullanım sayıları oluşturulurken kategorizasyon için beşe bölünmüştür.“ Avg(AileKullanımSayisi)/5 ”
- Tesis kullanımı, havuz kullanımı ile aynı tablo içerisinden hesaplandığı için mekan göre ayırım yapılmıştır.“ GK_Gecisler.MekanID = 1 ”

Sorgu :

```

Update MusteriListesi
Set TesiseGelisSayisiAilesi = AileUyesiTesisKullanimlari.
OrtAileKullanimi
From MusteriListesi
Inner Join (
Select Avg(AileKullanimSayisi)/5 OrtAileKullanimi, t.
AsilUyesiUyelikSatisId
From (select count(GecisId) AileKullanimSayisi,
UyelikSozlesmesiId, #AileUyeleri.AsilUyesiUyelikSatisId
From GK_Gecisler
Inner Join #AileUyeleri On GK_Gecisler.UyelikSozlesmesiID =
#AileUyeleri.AileUyesiUyelikSatisId
Where
(GK_Gecisler.MekanID = 1)
Group By UyelikSozlesmesiId, #AileUyeleri.
AsilUyesiUyelikSatisId ) t
Group By t.AsilUyesiUyelikSatisId
) AileUyesiTesisKullanimlari
On MusteriListesi.AsilUyelikSatisId =
AileUyesiTesisKullanimlari.AsilUyesiUyelikSatisId

```

Spa Kullanımının Hesaplanması :

- Kullanım sayıları oluşturulurken kategorizasyon için beşe bölünmüştür.“ Avg(ToplamaSpaHizmeti)/5 ”
- Aile üyelerinin spa kullanımlarının hesaplanması için her bir aile üyesinin spa kullanım sayıları teker teker hesaplanmıştır. Yapılan hesaplamanın sisteme uyumluluğu için ortalaması alınmıştır.“ Avg(ToplamaSpaHizmeti) ”

- Spa kullanımını hesaplamalarında hizmet gruplarına göre hizmetler baz alınmıştır.“ (LP_HizmetGruplari.HizmetGrupID = 1) ”
- Hizmetin üyelik süresi içerisinde kullanılmış olmasına dikkat edilmiştir.“ (LP_PlanSatislar.ServisTarihi >= @BaslangicTarihi And LP_PlanSatislar.ServisTarihi <= @BitisTarihi) ”
- Her hizmet kullanımını dikkate alınmıştır.
- Hizmetin kullanım sayısını oluşturabilmek için hizmetin müşteri tarafından kullanılmış olması gerekmektedir.Bu kıstas plan durum kontrolü ile yapılmaktadır.“PlanSatisDurumuID = 8”

Sorgu :

```

Update MusteriListesi
Set
SpaKullanimiAileUyesi = SpaPlanSayilari.ToplamSpaHizmeti
From MusteriListesi
Inner Join (Select Avg(ToplamSpaHizmeti)/5 ToplamSpaHizmeti,
t.AsilUyesiUyelikSatisId From
(
Select          Count          (PlanSatisId)          ToplamSpaHizmeti,
#AileUyeleri.AsilUyesiUyelikSatisId
From LP_PlanSatislar
Inner  Join  LP_Hizmetler  On  LP_PlanSatislar.HizmetId  =
LP_Hizmetler.HizmetId
Inner  Join  LP_HizmetGruplari  On  LP_Hizmetler.HizmetGrupId  =
LP_HizmetGruplari.HizmetGrupId
Inner  Join  #AileUyeleri  on  LP_PlanSatislar.MusteriID  =
#AileUyeleri.AileUyesiID
Where
(LP_HizmetGruplari.HizmetGrupID = 1) And

```

```

(LP_PlanSatislar.ServisTarihi >= #AileUyeleri.BaslangicTarihi
And
LP_PlanSatislar.ServisTarihi <= #AileUyeleri.BitisTarihi) And
PlanSatisDurumuID = 8
Group By #AileUyeleri.AsilUyesiUyelikSatisId ) t
Group By t.AsilUyesiUyelikSatisId ) SpaPlanSayilari
On MusteriListesi.AsilUyesiUyelikSatisId =
SpaPlanSayilari.AsilUyesiUyelikSatisId

```

Havuz Kullanımlarının Hesaplanması :

- Kullanım sayıları oluşturulurken kategorizasyon için beşe bölünmüştür.“ Avg(AileHavuzKullanimi)/5 ”
- Havuz kullanımı, tesis kullanımı ile aynı tablo içerisinde hesaplandığı için mekana göre ayırım yapılmıştır. “GK_Gecisler.MekanID in(19,20)”
- Aile üyelerinin havuz kullanımlarının hesaplanması için her bir aile üyesinin havuz kullanım sayıları teker teker hesaplanmıştır. Yapılan hesaplamanın sisteme uyumluluğu için ortalaması alınmıştır.“Avg(AileHavuzKullanimi) ”

Sorgu :

```

Update MusteriListesi
Set HavuzKullanimSayisiAileUyesi =
AileUyesiHavuzKullanimlari.OrtAileKullanimi
From MusteriListesi
Inner Join (Select Avg(AileHavuzKullanimi)/5 OrtAileKullanimi,
t.AsilUyesiUyelikSatisId From (
Select Count(GecisId) AileHavuzKullanimi, UyelikSozlesmesiId,
#AileUyeleri.AsilUyesiUyelikSatisId From GK_Gecisler

```

```

Inner Join #AileUyeleri On GK_Gecisler.UyelikSozlesmesiID =
#AileUyeleri.AileUyesiUyelikSatisId
Where      (GK_Gecisler.MekanID In (19,20))
Group      By      UyelikSozlesmesiId,
#AileUyeleri.AsilUyesiUyelikSatisId ) t
Group By t.AsilUyesiUyelikSatisId ) AileUyesiHavuzKullanimlari
On
MusteriListesi.AsilUyelikSatisId =
AileUyesiTesisKullanimlari.AsilUyesiUyelikSatisId

```

Grup Dersi Kullanım Hesaplamaları :

- Aile üyelerinin grup ders kullanımalarının hesaplanması için her bir aile üyesinin grup ders kullanım sayıları teker teker hesaplanmıştır. Yapılan hesaplamanın sisteme uyumluluğu için ortalaması alınmıştır. “ $Avg(AileGrupDersKullanimi)$ ”
- Kullanım sayıları oluşturulurken kategorizasyon için beşe bölünmüştür. “ $Avg(AileGrupDersKullanimi)/5$ ”
- Hesaplamanın yapılabilmesi için grup dersine ait sınıfın durumunun tamamlandı olması gerekmektedir. “ $LP_GrupDersleriSeanslari.SeansDurumuId = 4$ ”
- Hesaplamada diğer kıstas kişinin sınıfa katılmış olması gerekmektedir. “ $Katildi = 1$ ”

Sorgu :

```

Update MusteriListesi
Set
GrupDersKullanimiAsilUye =
OrtGrupDersKullanimi.OrtAileKullanimi
From MusteriListesi
Inner Join (

```

```

Select      Avg(AileGrupDersKullanimi)/5      OrtAileKullanimi,
t.AsilUyesiUyelikSatisId From ( Select Count(KatilimciId)
AileGrupDersKullanimi,      #AileUyeleri.AsilUyesiUyelikSatisId
from LP_GrupDersleriKatilimcilar
Inner      Join      LP_GrupDersleriSeanslari      on
LP_GrupDersleriKatilimcilar.SeansId      =
LP_GrupDersleriSeanslari.SeansId
Inner Join #AileUyeleri on GK_Gecisler.UyelikSozlesmesiID =
#AileUyeleri.AsilUyesiUyelikSatisId
Where LP_GrupDersleriSeanslari.SeansDurumuId = 4
Group By #AileUyeleri.AsilUyesiUyelikSatisId ) t
Group By t.AsilUyesiUyelikSatisId ) OrtGrupDersKullanimi
On      MusteriListesi.AsilUyelikSatisId      =
AileUyesiTesisKullanimlari.AsilUyesiUyelikSatisId

```

Özel Ders Kullanımlarının Hesaplanması :

- Aile üyelerinin özel ders kullanımlarının hesaplanması için her bir aile üyesinin özel ders kullanım sayıları teker teker hesaplanmıştır. Yapılan hesaplamaların sisteme uyumluluğu için ortalaması alınmıştır.“ Avg(ToplamSpaHizmeti) ”
- Kullanım sayıları oluşturulurken kategorizasyon için beşe bölünmüştür.“ Avg(ToplamSpaHizmeti)/5 ”
- Özel ders hesaplamalarında hizmet gruplarına göre hizmetler baz alınmıştır.“ (LP_HizmetGruplari.HizmetGrupID in (2,3,4)) -- Pilates&Yoga, Cilt Bakımı, Pt “
- Hizmetin üyelik süresi içerisinde kullanılmış olmasına dikkat edilmiştir.“(LP_PlanSatislar.ServisTarihi >= @BaslangicTarihi And LP_PlanSatislar.ServisTarihi <= @BitisTarihi)”
- Her hizmet kullanımı dikkate alınmıştır.

- Hizmetin kullanım sayısını oluşturabilmek için hizmetin müşteri tarafından kullanılmış olması gerekmektedir. Bu kıstas plan durumu kontrolü ile sağlanmıştır. “PlanSatisDurumuID = 8”

Sorgu :

```

Update MusteriListesi
Set SpaKullanimiAileUyesi = SpaPlanSayilari.ToplamSpaHizmeti
From MusteriListesi
Inner Join (
Select          Avg(ToplamSpaHizmeti)/5          ToplamSpaHizmeti,
t.AsilUyesiUyelikSatisId
From (
Select          Count          (PlanSatisId)          ToplamSpaHizmeti,
#AileUyeleri.AsilUyesiUyelikSatisId From
LP_PlanSatislar
Inner  Join  LP_Hizmetler  on  LP_PlanSatislar.HizmetId  =
LP_Hizmetler.HizmetId
Inner  Join  LP_HizmetGruplari  On  LP_Hizmetler.HizmetGrupId  =
LP_HizmetGruplari.HizmetGrupId
Inner  Join  #AileUyeleri  On  LP_PlanSatislar.MusteriID  =
#AileUyeleri.AileUyesiID
Where
(LP_HizmetGruplari.HizmetGrupID in (2,3,4)) And
(LP_PlanSatislar.ServisTarihi >= #AileUyeleri.BaslangicTarihi
And
LP_PlanSatislar.ServisTarihi < = #AileUyeleri.BitisTarihi) And
PlanSatisDurumuID = 8
Group by #AileUyeleri.AsilUyesiUyelikSatisId ) t
Group by t.AsilUyesiUyelikSatisId ) SpaPlanSayilari
On          MusteriListesi.AsilUyesiUyelikSatisId          =
SpaPlanSayilari.AsilUyesiUyelikSatisId

```



```

Delete Table #AileUyeleri
Fetch Next From Curl Into @UyelikSatisID
End
Close Curl Deallocate Curl
Alter Table MusteriListesi
Drop Column AsilUyelikSatisId, AsilUyeID, SonrakiUyelikSatisId,
BaslangicTarihi, BitisTarihi

```

3.5. MODEL OLUŞTURMA

Bu aşamada, önceki aşamalarda iş hedefleri ve veri madenciliği amacı doğrultusunda biçimlendirilen spor ve yaşam merkezine ait müşteri verileri üzerinde sınıflandırma modelleri ve kümeleme algoritmalarıyla geliştirilecektir. Bu kapsamda modelleme tekniğinin seçimi, sınama tasarımının üretilmesi, modelin kurulması ve model değerlemesi üzerinde ayrıntılarıyla durulacaktır. Bu çalışmada, daha önce belirtildiği gibi sınıflandırma modelleri ve kümeleme algoritmaları ile bunlara ilişkin fonksiyonlar arasında seçim yapılmıştır. Bu seçimde gerek literatürdeki yaygınlık gerekse mevcut yazılım ve donanım olanakları önemli rol oynamıştır.

Veri madenciliğinin en önemli bileşenlerinin istatistik ve veritabanı teknolojileri olduğundan önceki bölümde söz edilmişti. Çalışmada kullanılacak yöntemlerin de bu alanları temsil etmelerine özen gösterilmiştir. Sonuç olarak, yazılım ve donanım olanakları doğrultusunda sınıflandırma için J48 (C4.5), ID3, IB1, IBk, Naive Bayes, kümeleme algoritmasında ise Simple K-Means, Scalable K-Means algoritmaları uygulanmasına karar verilmiştir.

Seçilen bu tekniklerin hem hedef değişkenlerin sınıflandırılmasına uygun olmaları hem de her ölçekte değişken tipinin kümeleme algoritması için bağımsız değişken olarak kabul edebilmeleri nedeniyle hazırlanan veri kümeleri üzerinde uygulanmalarında herhangi bir sakınca bulunmamaktadır.

Yapılacak uygulama sonucu elde edilecek sınıflandırma modelleri ve kümeleme algoritmaları için iki temel başarı kriteri söz konusudur. Öncelikle geliştirilen modellerin öğrenme kümesi üzerinde yüksek sınıflandırma ve kümeleme başarısı göstermesi gerekir. Ancak veri madenciliği amacı doğrultusunda asıl olan geliştirilen modellerin öngörü amaçlı olarak kullanılabilmesidir. Bu açıdan ikinci başarı kriteri, sınıflandırma ve kümeleme algoritmaları için öğrenme kümesinden tamamen farklı bir veri kümesi üzerinde de yüksek sınıflandırma başarısı göstermesi olacaktır.

Özetlemek gerekirse, her modelleme işlemi için veri kümesi yazılıma yüklenecek, ilgili veri kümesi üzerinde değişken ya da kayıt bazında gereken işlemler yürütülecek ve sürecin sonunda model ya da çıktılar elde edilecektir.

3.5.1. SINIFLANDIRMA VE KÜMELEME TEKNİKLERİNİN KULLANIMI

3.5.1.1. WEKA İLE SINIFLANDIRMA VE KÜMELEME ANALİZİ

Veri madenciliği analizlerini oluşturacağımız ilk yardımcı program Weka 'dır. Weka, Yeni Zelanda'daki Waikato Üniversitesi tarafından geliştirilmiş, makine öğrenimi algoritmalarının bir arada barındıran, işlevsel bir grafik arabirimine sahip, açık kaynak kodlu bir veri madenciliği

programıdır ^[46]. Weka çeşitli veri ön işleme, sınıflandırma, regresyon, kümeleme, ilişkilendirme kuralları ve görselleştirme araçları içerir. Algoritmalar veri kümesine doğrudan veya Java kodundan çağrılarak uygulanabilir ^[47]. Aynı zamanda yeni makine öğrenme algoritmaları geliştirmek için de uygundur. Veri madenciliği çalışmanın bu aşamasında veri setimize uygunluğu tez danışmanımın yönlendirmesi ile sınıflandırma algoritması için J48 (C4.5), IB1, IBk, Naive Bayes algoritmaları ile veri seti üzerinde çalışma yapılmıştır. Çalışma sonrası kullandığımız algoritmalar karar ağaçları için bazı kurallar oluşturulmuştur. Üyelik yenileme alanı sınıflarını oluşturmaktadır. Bu bağlamda oluşturulan bazı kurallar aşağıdaki gibidir:

1. Eğer Aile Üyesinin Tesise Geliş Sayıları ≥ 5 ise Yenileme = 1
2. Eğer Aile Üyesinin Tesise Geliş Sayıları ≤ 1 ise Yenileme = 0
3. Eğer Asil Üyesinin Havuz Kullanımı > 2 ve Aile Üyesinin Tesise Geliş Sayıları > 4 ise Yenileme = 1
4. Eğer Asil Üyesinin Tesise Geliş Sayıları < 2 ve Aile Üyesinin Tesise Geliş Sayıları < 3 ise Yenileme = 0
5. Eğer Asil Üyesinin Havuz Kullanımı < 2 ise Yenileme = 0

Karar ağaçları ile çıkan sonucu incelediğimizde aile üyelerinin tesise geliş sayıları kuralların oluşumunda baskın nitelik olarak görmekteyiz. Bununla birlikte havuz kullanımının da üyelik yenileme de bir kriter olduğunu söyleyebiliriz. Asil ve aile üyelerinin tesise giriş sayılarını beraber incelediğimizde kritik bir noktası olduğunu görmekteyiz. Bu sebeple asil ve

⁴⁶ Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten. «The WEKA Data Mining Software: An Update.» ACM SIGKDD Explorations, 2009: 10-18.

⁴⁷ Patterson, D., Liu, F., Turner, D., Concepcion, A., Lynch, R. «Performance Comparison of the Data Reduction System.» Proceedings of the SPIE Symposium on Defense and Security. Orlando, FL., 2008.

aile üyelerinin beraber giriş sayıları da bir kriter olarak gözümüze çarpmaktadır.

Karar ağaçları dışında diğer algoritmaları incelediğimiz aşağıdaki tablodaki sonuçları elde etmekteyiz.

Tablo 3.11. Weka Sınıflandırma Algoritma Değerlendirme Sonuçları

| | Tp Rate | Fp Rate | Precision | Recall | F-Measure | Roc Area |
|--------------------|----------------|----------------|------------------|---------------|------------------|-----------------|
| J48 (C4.5) | 0.966 | 0.063 | 0.966 | 0.966 | 0.966 | 0.978 |
| Naive Bayes | 0.812 | 0.08 | 0.812 | 0.812 | 0.875 | 0.946 |
| IB1 | 0.735 | 0.33 | 0.804 | 0.735 | 0.768 | 0.703 |
| IBk | 0.735 | 0.33 | 0.804 | 0.735 | 0.768 | 0.703 |

Çıkan sonuçlarda J48 (C4.5) algoritmasının başarı oranının oldukça yüksek olduğunu görmekteyiz. Öne çıkan nitelikler, tesise geliş sayıları ile havuz girişleri olmuştur. Bu nitelikler üyelik yenilenmesine doğrudan etki eden faktörlerdir. **TP**; doğru sınıflandırılmış pozitif örnek sayısı, **FP**; yanlış sınıflandırılan pozitif örnek sayısını, **Precision**; Pozitif örnek sayısının, sınıfı 1 olarak tahminlenmiş tüm örnek sayısına oranı, **F-Measure**; Precision ve Tp ile formülize edilerek hesaplanmıştır.

Şekil 3.2. Weka J48 (C4.5) Algoritma Sonucu Oluşan Kurallar

```
J48 pruned tree
-----
TesiseGelisSayisiAilesiGrup <= 5
| TesiseGelisAsilUyeGrup <= 4
| | TesiseGelisSayisiAilesiGrup <= 4: NO (124.0)
| | TesiseGelisSayisiAilesiGrup > 4
| | | TesiseGelisAsilUyeGrup <= 3: NO (7.0/1.0)
| | | TesiseGelisAsilUyeGrup > 3: YES (2.0)
| TesiseGelisAsilUyeGrup > 4
| | HavuzKullanimSayisiAsilUyeGrup <= 2
| | | TesiseGelisSayisiAilesiGrup <= 1: YES (65.0/10.0)
| | | TesiseGelisSayisiAilesiGrup > 1: NO (44.0/8.0)
| | HavuzKullanimSayisiAsilUyeGrup > 2: YES (45.0)
TesiseGelisSayisiAilesiGrup > 5: YES (213.0)

Number of Leaves :      7
Size of the tree :     13
```

Şekil 3.3. Weka J48 (C4.5) Algoritma Sonucu Doğruluk Oranları

| | | | |
|----------------------------------|---------|------|---|
| Correctly Classified Instances | 478 | 95.6 | % |
| Incorrectly Classified Instances | 22 | 4.4 | % |
| Kappa statistic | 0.9035 | | |
| Mean absolute error | 0.0721 | | |
| Root mean squared error | 0.1989 | | |
| Relative absolute error | 15.7985 | % | |
| Root relative squared error | 41.6479 | % | |
| Total Number of Instances | 500 | | |

Şekil 3.4. Weka J48 (C4.5) Algoritma Değerlendirme Sonuçları

```

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
          0.938    0.034    0.938     0.938    0.938     0.978    NO
          0.966    0.063    0.966     0.966    0.966     0.978    YES
Weighted Avg.   0.956    0.052    0.956     0.956    0.956     0.978

=== Confusion Matrix ===
  a  b  <-- classified as
165 11 |  a = NO
 11 313 |  b = YES

```

Kümeleme analizi için Simple K-Means algoritması kullanılmıştır. Algoritma çalıştırılmadan önce küme sayısı olarak beşe kümelenmesi için parametre verilmiştir.

Tablo 3.12. K Ortalamalar Algoritması İçin K Değerleri İle Küme İçindeki

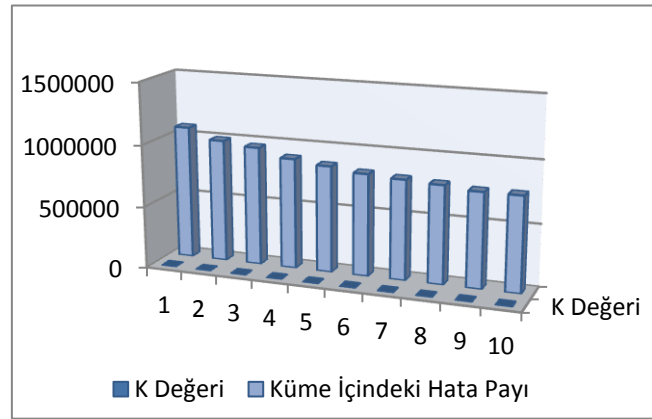
Hata Payı Değişim Tablosu

| K Değeri | Tekrarlanma Sayısı | Küme İçindeki Hata Payı | Çalışma Süresi (Saniye) |
|----------|--------------------|-------------------------|-------------------------|
| 2 | 12 | 1275.532 | 0.05 |
| 3 | 7 | 1095.624 | 0.04 |
| 4 | 11 | 1050.133 | 0.06 |
| 5 | 9 | 997.449 | 0.07 |
| 6 | 14 | 948.265 | 0.09 |
| 7 | 9 | 912.846 | 0.07 |
| 8 | 12 | 892.867 | 0.12 |
| 9 | 16 | 870.296 | 0.14 |
| 10 | 13 | 841.775 | 0.13 |

Şekil 3.5. K = 5 İçin Oluşturulan Küme Tablosu

| Attribute | Full Data (500) | Cluster# | | | | | |
|------------------------------|--------------------|--------------|-----------|-----------|-----------|------------|---------------------|
| | | 0 (110) | 1 (71) | 2 (86) | 3 (72) | 4 (161) | |
| SozlesmeFiyati | 2.384 | 2.5545 | 2.9296 | 2.2748 | 3.1111 | 2.4348 | |
| TesiseGelisAsilUye | 5.058 | 5.8273 | 7.3662 | 6.314 | 3.9306 | 3.3478 | |
| TesiseGelisSayisiAilesi | 5.266 | 5.4909 | 3.6056 | 8.1279 | 10.375 | 2.0311 | |
| SpaKullanimiAsilUye | 1.096 | 1.0273 | 1.2254 | 1.2558 | 1.0139 | 1.0373 | |
| SpaKullanimiAileUyesi | 1.212 | 1.1727 | 1.0282 | 1.3953 | 1.7083 | 1 | |
| HavuzKullanimSayisiAsilUye | 1.638 | 1.8 | 2.2676 | 2.0116 | 1.4167 | 1.1491 | |
| HavuzKullanimSayisiAileUyesi | 1.738 | 1.7727 | 1.3944 | 2.4767 | 2.6389 | 1.0683 | |
| GrupDersKullanimiAsilUye | 1.244 | 1.2636 | 1.6338 | 1.3488 | 1.0833 | 1.0745 | |
| GrupDersKullanimiAileUyesi | 1.372 | 1.3182 | 1.169 | 1.7907 | 1.9722 | 1.0062 | |
| OzelDersKullanimiAsilUye | 1.128 | 1.1 | 1.3521 | 1.186 | 1.0694 | 1.0435 | |
| OzelDersKullanimiAileUyesi | 1.224 | 1.1545 | 1.0704 | 1.4884 | 1.6667 | 1 | |
| Kaynak | TAVSIYE | STAND | TAVSIYE | BILLBOARD | TAVSIYE | STAND | |
| Lokasyon | BATIATASEHIR | BATIATASEHIR | MALTEPE | BOSTANCI | KUCUKYALI | KUCUKYALI | |
| UyelikYenileme | YES | YES | YES | YES | YES | NO | |
| | | | | | | | Clustered Instances |
| | | | | | | | 0 110 (22%) |
| | | | | | | | 1 71 (14%) |
| | | | | | | | 2 86 (17%) |
| | | | | | | | 3 72 (14%) |
| | | | | | | | 4 161 (32%) |

Şekil 3.6. K Ortalamalar Algoritması İçin K Değerleri İle Küme İçindeki Hata Payı



Algoritma çalıştırdıktan sonra alınan sonuçlar şekil ve tablolar ile gösterilmiştir. Kümeleme algoritmalarında ise veri seti içinde doğrudan görülemeyen benzer özelliklere sahip müşteri kümeleri ortaya çıkarılmıştır. Bunun sonuçlarında lokasyon niteliği baskın çıkmıştır.

3.5.1.2. SQL BUSINESS INTELLIGENCE İLE KÜMELEME ALGORİTMASI

Veri madenciliği, verilerden çeşitli teknikler, algoritmalar ve sorgulamalarla anlamlı bilgiler keşfetmektir. Veriye dayalı stratejik karar destek sistemlerinin tümü sonuçta "**iş zekası (business intelligence) çözümleri**" olarak isimlendirilmektedir. **OLAP** sistemleri; kullanıcıların hızlı analiz yapmalarını sağlayan iş zekası sistemleridir.

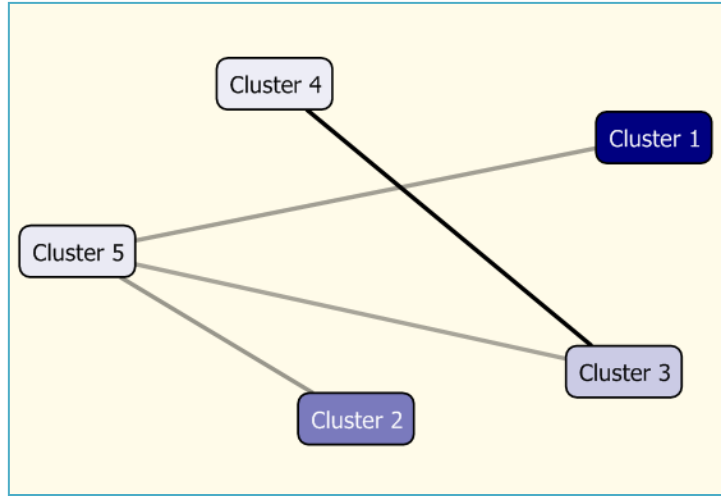
Microsoft İş Zekası (BI) Microsoft SQL Server özelliklerinin üzerine kurulmuştur; Microsoft® SharePoint® ve Microsoft® Office ile birlikte tam donanımlı bir bilgi platformu sunar. Kuruluşların değişen ihtiyaçlarına hızlı cevap verebilmek için esneklik sağlamaktadır. Microsoft İş zekası çözümleri kuruluşlarda IT verimliliğini düşük maliyetle artırmak mümkündür. Bugünü ve yarını görerek analizler ve raporlamalar yapmayı sağlar. Bunun için Microsoft SQL Server 'ın altyapısını kullanır. SQL Server Analysis Services, OLAP küplerinden anlamlı verilerin elde edilmesine ve analizlerde kullanılmasına olanak sağlamaktadır.

Microsoft iş zekası; kuruluşlar ile ilgili karmaşık veri setlerinden raporlamalar yaparak oluşturulan pano (dashboard) ile yöneticilerin işleriyle alakalı bilgileri analiz etmesine olanak tanır. Yöneticiler analiz edilmiş verileri panolar sayesinde kolaylıkla yorumlayabilirler. Kuruluşun bugünü ve yarını ile ilgili öngöründe bulunma konusunda Microsoft BI teknolojisinin performansı oldukça yüksektir. SQL Server Analysis Services ile OLAP datalarından anlamlı veriler çekerek veri madenciliği yapmak mümkündür. Microsoft veri madenciliği çözümleri, çalışanların basit veri analizinin ötesine geçerek bilgi dahilindeki gizli eğilimleri, sorunları veya ilişkileri tanımlamalarını kolaylaştırır.

Bu aşamada veri setimizi kümeleme algoritması tekniklerini kullanacağımız yardımcı programımız Microsoft Sql Business Intelligence olacaktır. Burada kullanacağımız algoritmalar Scalable K-Means olacaktır. Burada kullanacağımız veri seti ile bağlantı direk Microsoft Sql Server üzerinden yapılacaktır. Veri hazırlama işlemleri bittikten sonra veri seti için oluşturulan veri tablosu için sunucu ile bağlantı sağlanacaktır. Bağlantı

kurulduktan sonra sırayla veri seti seçimi, hedef niteliklerin belirlenmesi ile devam edecektir. Bu işlemlerin sonunda kullanacağımız kümeleme algoritmasını ve işlem sırasında algoritmaya ait analiz için parametreler verilecektir. Burada kümeleme işlemlerinde sayısı olarak beş olarak belirlenmiştir (Şekil 3.7).

Şekil 3.7.Kümelerin Birbiriyle Bağlantıları

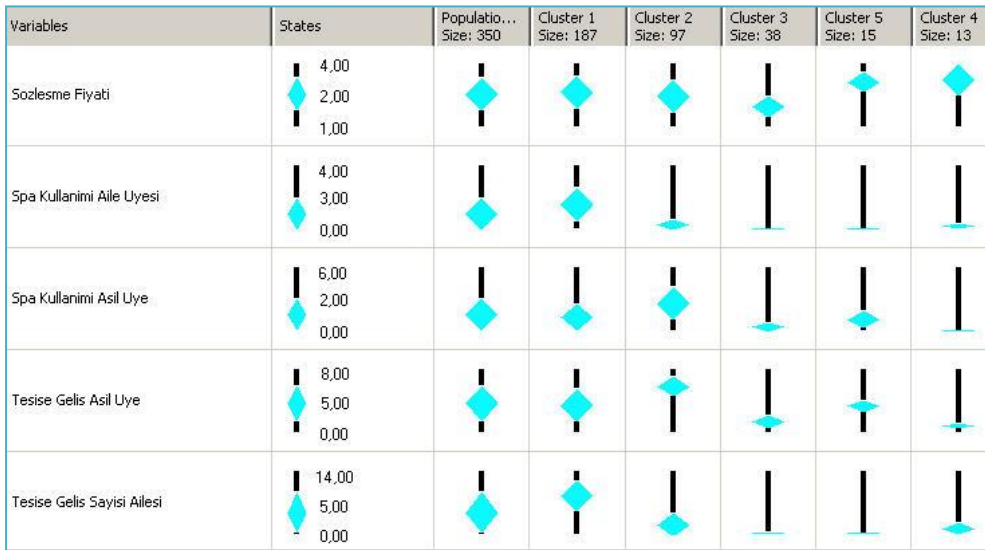


Yukarıdaki küme renkleri kümedeki veri yoğunluğunu temsil etmektedir.

Kümelerin birbiriyle olan benzerliklerini göstermek adına çizgiler çizilmiştir. Burada renklerin açıktan koyuya gitmesi veri kümeleri arası benzerliklerin yoğunluğu belirtmektedir. Analiz resmi küme yoğunluğu üzerine yorumlanmıştır.

Bir sonraki aşama kümelerde bulunan nitelikleri değerlerinin minimum, maksimum ve ortalama değerlerinin hesaplanmış halini görerek yapılmış küme analizinin kümeleri neye göre oluşturduğuna dair fikir sahibi olunmaktadır.

Şekil 3.8. Microsoft Sql Business Intelligence Küme Profili



Son aşamada kümeleme algoritmalarının oluşturmuş olduğu kümelerin birbirleri ile kıyaslanma safhasına gelinmiştir. Kümelerde yer alan değerler birbirleri ile kıyaslanabilir hale getirilmiştir. Kümeleme incelemesi sonrası kümelerin değer aralıkları ve farklılıkları incelenmiştir. Bu bağlamda bütün kümelerin ortak oluşturacağı değerler baz alınarak aşağıdaki şekilde yer alan sonuçlara ulaşılmıştır.

Şekil 3.9. Microsoft Sql Business Intelligence

Kümeleme Algoritma Sonuçları (Scalable K-Means)

| Variables | Values | Probability |
|----------------------------------|--------|-------------|
| Tesise Gelis Sayisi Ailesi | 5 - 8 | |
| Ozel Ders Kullanimi Aile Uyesi | 1 - 3 | |
| Spa Kullanimi Aile Uyesi | 1 - 3 | |
| Grup Ders Kullanimi Aile Uyesi | 2 - 4 | |
| Havuz Kullanim Sayisi Aile Uyesi | 2 - 6 | |
| Uyelik Yenileme | 1 - 0 | |

Kümeleme algoritmalarında veri seti içinde doğrudan görülemeyen benzer özelliklere sahip müşteri kümeleri ortaya çıkarılmıştır. Bunun sonuçlarında lokasyon niteliği baskın çıkmıştır. Sql Business Intelligence ile yapılan çalışmada bütün kümeler temel alınarak sonuçlar (Şekil 3.9) ortaya

çıkarılmıştır. İşlemleri detaylandırmak ve bütün kümeleri teker teker analiz etmek ve buradan gizli örüntüleri yakalamak mümkündür.

3.5.1.3. ID3 ALGORİTMASI KULLANILARAK GELİŞTİRİLEN UYGULAMA

Algoritma kullanımı için geliştirilen uygulamanın etkin bir şekilde çalışabilmesi için elimizde bulunan veri setinin uygulamanın yorumlayacağı şekilde oluşturulması gerekmektedir. Uygulama web tabanlıdır. Bu uygulama javascript kodları ile yönetilmektedir. Veri seti “example.js” adındaki script dosyasında bulunmaktadır. Ayrıca geliştirilen yazılım içerisinde kullanılacak olan niteliklerin listesi de aynı dosyada yer almaktadır. ID3 algoritmasının oluşturulmasında kullanılan bütün fonksiyonlar “id3.js” adlı script dosyasında yer almaktadır. Uygulama çalışmaya başlamasından itibaren “id3.js” dosyasında bulunan fonksiyonlar ile veri seti değerlendirmelere alınmaya başlanacak, entropi ve kazanç hesaplamaları ile dallanmalar yaparak uygulamanın oluşturacağı karar ağacını şekillendirmeye başlayacaktır. En sonda program bulduğu sonuçlara göre karar ağacını çizip ve kuralları oluşturmuştur. Dördüncü aşamasında ”id3.js” dosyası detaylı şekilde açıklanmıştır.

Çalışmanın adımları sırayla:

1-Veri setinin programın yorumlayabileceği şekilde düzenlenmesi:

Program veri setinin her satırı için belli bir formata bağlı olarak veri analizi yapmaktadır. Format şu şekildedir:

```
{Sozlesme:'D1',    SozlesmeFiyati:'4',    TesiseGelisAsilUye:'3',
TesiseGelisSayisiAilesi:'4',                SpaKullanimiAsilUye:'1',
SpaKullanimiAile    Uyesi:'1',    HavuzKullanimSayisiAsilUye:'1',
```

```
HavuzKullanımSayısıAileÜyesi: '2', GrupDersKullanımıAsilUye:'1'
,GrupDersKullanımıAileÜyesi:'1', ÖzelDers KullanımıAsilUye:'1',
ÖzelDersKullanımıAileÜyesi:'1', Kaynak:'STAND', Lokasyon:
'KÜÇÜKYALI', ÜyelikYenileme:'NO'}
```

Yukarıdaki formatı oluşturabilmek için Sql Cursor kullanılmıştır. Formata uygun veri satırları oluşturulan bir temp tabloya atılmıştır ve Sql Cursor işlemini bitirdikten sonra programın veri seti hazır hale gelecektir.

- Sql Cursor için tanımlanan değişkenlerdir.

```
Declare @Sayac Int
Declare @SozlesmeFiyati NVarchar(50)
Declare @TesisGelisAsilUye NVarchar(50)
Declare @TesisGelisSayisiAilesi NVarchar(50)
Declare @SpaKullanımıAsilUye NVarchar(50)
Declare @SpaKullanımıAileÜyesi NVarchar(50)
Declare @HavuzKullanımSayısıAsilUye NVarchar(50)
Declare @HavuzKullanımSayısıAileÜyesi NVarchar(50)
Declare @GrupDersKullanımıAsilUye NVarchar(50)
Declare @GrupDersKullanımıAileÜyesi NVarchar(50)
Declare @ÖzelDersKullanımıAsilUye NVarchar(50)
Declare @ÖzelDersKullanımıAileÜyesi NVarchar(50)
Declare @Kaynak NVarchar(50)
Declare @Lokasyon NVarchar(50)
Declare @ÜyelikYenileme NVarchar(50)
Declare @Sql NVarchar(400)
Set @Sayac = 1
```

- Niteliklerin uygulamayan eklenmesi için kullanılan tablodur.

```
Create Table #NitelikSeti (VeriSetiSatiri NVarchar(4000))
Declare Cur1 Cursor For
Select SozlesmeFiyati, TesisGelisAsilUye
, TesisGelisSayisiAilesi, SpaKullanımıAsilUye
```

```

, SpaKullanimiAileUyesi, HavuzKullanimSayisiAsilUye
, HavuzKullanimSayisiAileUyesi, GrupDersKullanimiAsilUye
, GrupDersKullanimiAileUyesi, OzelDersKullanimiAsilUye
, OzelDersKullanimiAileUyesi, Kaynak, Lokasyon, UyelikYenileme
From MusteriListesi
Open Curl
Fetch Next From Curl Into
@SozlesmeFiyati, @TeseGelisAsilUye, @TeseGelisSayisiAilesi
, @SpaKullanimiAsilUye, @SpaKullanimiAileUyesi
, @HavuzKullanimSayisiAsilUye, @HavuzKullanimSayisiAileUyesi
, @GrupDersKullanimiAsilUye, @GrupDersKullanimiAileUyesi
, @OzelDersKullanimiAsilUye, @OzelDersKullanimiAileUyesi
, @Kaynak, @Lokasyon, @UyelikYenileme
While (@@Fetch_Status = 0)
Begin

```

- Uygulamanın istediği formatı oluşturabilmek için string birleştirme işlemi yapılmıştır.

```

Set @Sql = '{Sozlesme:''S'+Cast(@Sayac as NVarchar(10))+''',
SozlesmeFiyati:'''+Cast(@SozlesmeFiyati as NVarchar(50))+''',
TeseGelisAsilUye:'''+Cast(@TeseGelisAsilUye as NVarchar
(50))+''', TeseGelisSayisiAilesi:'''+Cast(@TeseGelisSayisi
Ailesi as NVarchar(50))+''',SpaKullanimiAsilUye:'''+Cast
(@SpaKullanimiAsilUye as NVarchar(50))+''',
SpaKullanimiAileUyesi:'''+Cast(@SpaKullanimiAileUyesi as
NVarchar(50))+''',HavuzKullanimSayisiAsilUye:'''+Cast(@HavuzKul
lanimSayisiAsilUye as NVarchar(50))+''',HavuzKullanimSayisi
AileUyesi:'''+Cast (@HavuzKullanimSayisiAileUyesi as NVarchar
(50))+''',GrupDersKullanimi AsilUye:'''+Cast(@GrupDersKullanimi
AsilUye as NVarchar(50))+''',GrupDers KullanimiAileUyesi:'''+
Cast(@GrupDersKullanimiAileUyesi as NVarchar(50))+''',

```

```
OzelDersKullanimiAsilUye:'''+Cast (@OzelDersKullanimiAsilUye as
NVarcha r(50))+'',OzelDersKullanimiAileUyesi:'''+
Cast (@OzelDersKullanimiAileUyesi as NVarcha r(50))+'',
Kaynak:'''+replace (Cast (@Kaynak as NVarcha r(50)), '
','')+''',Lokasyon:'''+Cast (@Lokasyon as NVarcha r(50))+'',
UyelikYenileme:'''+Cast (@UyelikYenileme as NVarcha r(50))+''},'
```

- Oluşturulan format temp tabloya insert edilmektedir.

```
Insert Into #NitelikSeti Values (@Sql)
Set @Sayac = @Sayac +1
Fetch Next From Curl Into
@SozlesmeFiyati
, @TesiseGelisAsilUye, @TesiseGelisSayisiAilesi
, @SpaKullanimiAsilUye, @SpaKullanimiAileUyesi
, @HavuzKullanimSayisiAsilUye, @HavuzKullanimSayisiAileUyesi
, @GrupDersKullanimiAsilUye, @GrupDersKullanimiAileUyesi
, @OzelDersKullanimiAsilUye, @OzelDersKullanimiAileUyesi
, @Kaynak, @Lokasyon, @UyelikYenileme
End
Close Curl
Deallocate Curl
```

- Sql Cursor sayacını bitirdikten sonra programın kullanacağı formatta veri setimiz oluşmuştur.

```
Select * From #NitelikSeti Drop Table #NitelikSeti
```

- Sql Cursor ile belli bir formata göre oluşturulan veri setinin geliştirilen uygulama içerisinde kullanılması için programın “Examples.js” dosyası içerisine eklenmesi gerekmektedir. Kopyalama ile program içerisine eklenmektedir.

2-Veri setinin geliştirilen uygulamaya uyarlanması:

Veri setinin uygulama içerisinde örnek olarak gösterimi şu şekildedir:

```
Var Examples = [Sozlesme:'D1', SozlesmeFiyati:'4',
TesiseGelisAsilUye:'3', TesiseGelisSayisiAilesi:'4', SpaKull
animiAsilUye:'1', SpaKullanimiAileUyesi:'1', HavuzKullani
mSayisiAsilUye:'1', HavuzKullanimSayisiAileUyesi:'2', Grup
DersKullanimiAsilUye:'1', GrupDersKullanimiAileUyesi:'1', Ozel
DersKullanimiAsilUye:'1', OzelDersKullanimiAileUyesi:'1',
Kaynak:'STAND', Lokasyon:'KÜÇÜKYALI', UyelikYenileme:'NO']
```

Yukarıda tanımlanan Examples bir diziyi temsil etmektedir.İçerisinde belirtilen örnek sadece bir veri seti satırını temsil etmektedir.Uygulamanın çalışabilmesi için veri setine ait bütün satırların yukarıdaki formata göre diziyeye tanımlanması gerekmektedir.

3-Veri setinde kullanılan niteliklerin uygulamaya eklenmesi:

Veri setinde kullanılan niteliklerin uygulama içerisinde örnek olarak gösterimi şu şekildedir:

```
Var Features = ['SozlesmeFiyati', 'TesiseGelisAsilUye',
TesiseGelisSayisiAilesi', 'SpaKullanimiAsilUye', 'SpaKulla
nimiAileUyesi', 'HavuzKullanimSayisiAsilUye', 'HavuzKullanim
SayisiAileUyesi', 'GrupDersKullanimiAsilUye', 'GrupDersKulla
nimiAileUyesi', 'OzelDersKullanimiAsilUye', 'OzelDersKullan
imiAileUyesi', 'Kaynak', 'Lokasyon'];
```

Yukarıda belirtilen Features, Examples gibi bir diziyi temsil etmektedir. Burada sınıf harici kalan bütün nitelikler bu alana eklenmesi gerekmektedir.

4-ID3 Algoritması İçin Kullanılan Fonksiyon Tanımları ve Kullanımları:

Bu aşamada uygulamanın aksiyon planını oluşturan fonksiyonlar yer almaktadır.Sırayla fonksiyonlar incelemek gerekirse:

- Aşağıdaki kod bloğu ilk olarak başlangıçta karar ağacını oluşturacak ilk niteliği buluyor ve bundan sonraki aşamalarında dallanma

işlemlerine başlamaktadır. Her kazanç değeri yüksek olan nitelik fonksiyonun kullandığı nitelik listesinden çıkarılarak temel alınmaktadır.

- Fonksiyon işlemlerde oluşan dallanmalara göre kendi kendini çağırarak tekrar tekrar kontrol, hesaplama ve kural oluşturma için çalışmalar yapmaktadır.
- `Function(_s,target,Features)` - Fonksiyon parametre olarak veri seti,sınıf ve nitelikleri almaktadır.

```
Var id3 = Function(_s,target,Features)
```

```
{
```

- Niteliği ait sınıf için değerler pluck fonksiyonu elde edilip benzersizliği kontrol edilerek targets dizisine atılmaktadır.(Sınıf UyelikYenileme alabileceği değerler : 1,0 gibi.)

```
Var targets = _.unique(_s.pluck(Target));
```

- Düğüm Sonu kontrolüdür.

```
if (targets.length == 1)
```

```
{
```

```
    Return {type:"result", val: targets[0], name:
    targets[0],alias:targets[0]+randomTag()
};
```

```
};
```

```
}
```

- Nitelik değeri için kontrol yapılıp yapılmamasına karar verilmektedir.

```
if(features.length == 0)
```

```
{
```

```
Var topTarget = mostCommon(targets);
```

```
Return {type:"result", val: topTarget, name: topTarget,
alias: topTarget+randomTag()};
```

```
}
```

- Nitelik için kazancı en yüksek olan nitelik bulunmaktadır.

```
Var bestFeature = maxGain(_s,target,Features);
```

- Kazancı yüksek olan nitelik dışında kalan nitelikler belirlenmektedir.

```
Var remainingFeatures = _.without(Features,bestFeature);
```

- Kazancı yüksek olan niteliğin alabileceği değerler belirlenmektedir.

```
Var possibleValues = _.unique(_s.pluck(bestFeature));
```

- Karar ağacı için ilk basamak oluşturulmaktadır.

```
Var node = {name: bestFeature,alias: bestFeature+randomTag()};
```

```
node.type = "feature";
```

```
node.vals = _.map(possibleValues,Function(v){
```

- İlk basamaktaki niteliğin dallanacağı değer seçilmektedir.

```
Var _newS = _(_s.filter(Function(x) {Return x[bestFeature] ==  
v})));
```

```
Var child_node = {name:v,alias:v+randomTag(),type: "featu  
re_value"};
```

```
child_node.child = id3(_newS,target,remainingFeatures);
```

```
Return child_node;
```

```
});Return node;
```

```
}
```

- Karar ağacı çiziminde kuralların temsil etmektedir.

```
Var predict = Function(id3Model,sample)
```

```
{
```

```
    root = id3Model;
```

```
    While(root.type != "result"){
```

```
        Var attr = root.name;
```

```
        Var sampleVal = sample[attr];
```

```
        Var childNode = _.detect(root.vals,Function(x){Return  
x.name == sampleVal});
```

```
        root = childNode.child;
```



```

    } Return root.val;
}

```

- Nitelikler için entropi hesaplamalarında kullanılmaktadır.

```

Var entropy = Function(vals)
{
    Var uniqueVals = _.unique(vals);
    Var probs = uniqueVals.map(Function(x) {Return prob
(x, vals)});
    Var logVals = probs.map(Function(p) {Return -p*log2(p) });
    Return logVals.reduce(Function(a,b) {Return a+b}, 0);
}

```

- Niteliklerin kazanç hesaplamasında kullanılmaktadır.

```

Var gain = Function(_s, target, feature)
{
    Var attrVals = _.unique(_s.pluck(feature));
    Var setEntropy = entropy(_s.pluck(target));
    Var setSize = _s.size();
    Var entropies = attrVals.map(Function(n) {
    Var subset = _s.filter(Function(x) {Return x[feature] ===
n});
    Return
(subset.length/setSize)*entropy(_.pluck(subset, target));
});
    Var sumOfEntropies = entropies.reduce(Function(a,b) {Return
a+b}, 0);
    Return setEntropy - sumOfEntropies;
}

```

- Kazançlar arasında en yüksek değere sahip niteliği bulunmasını sağlamaktadır.

```

Var maxGain = Function(_s, target, Features)

```

```
{
    Return _.max(Features,Function(e){Return gain (_s,target
,e)});
}
```

- Kazanç hesaplamalarına yardımcı fonksiyon olarak görev almaktadır.

```
Var prob = Function(val,vals)
{
    Var instances = _.filter(vals,Function(x) {Return x ===
val}).length;
    Var total = vals.length;
    Return instances/total;
}
```

- Logaritma hesaplamalarının yapılmasını sağlamaktadır.

```
Var log2 = Function(n)
{
    Return Math.log(n)/Math.log(2);
}
```

- Kural kontrollerinde nitelik geçişlerinin yapılmasını sağlamaktadır. Basit olarak işlemi sıralama yaparak nitelik değerleri içerisinde sıkça kullanılmış olan değeri hesaplamaktadır.

```
Var mostCommon = Function(l)
{
    Return _.sortBy(l,Function(a){
        Return count(a,l);
    }).reverse()[0];
}
```

- Değer hesaplaması için yardımcı fonksiyon görevi görmektedir. Kullanıldığı yer mostcommon değerinin oluşumunda kullanılan fonksiyondur.

```

Var count = Function(a,l){
    Return _.filter(l,Function(b) { Return b === a}).length
}

```

- Çizim için kullanılan bir fonksiyondur.Görsel oluşumun sağlanmasında kullanılmıştır.

```

Var randomTag = Function()
{
    Return "_r"+Math.round(Math.random()*1000000).toString();
}

```

5-ID3 Algoritması İle Oluşturulan Karar Ağacı Kuralları:

Uygulama hesaplamalarının sonucunda oluşturulan karar ağacı kuralları aşağıdaki gibi oluşmaktadır.Kurallar belli bir gruplandırma tabii tutulmamıştır.Bütün olası sonuçlar aşağıda yer almaktadır.Oluşan kurallar için kısaca açıklamalar eklenerek anlaşılabilirlik seviyesi yükseltilmeye çalışılmıştır.

Uygulamanın Kullandığı Veri Setindeki Örnek Sayısı: 500

Algoritmanın Sınıfı: UyelikYenileme

Algoritmadaki Nitelik Sayısı: 13

Uygulamanın Oluşturduğu Kural Sayısı: 41

Id3 algoritması için geliştirilen uygulama çalıştırıldıktan sonra ortaya çıkan kural maddeler halinde ortaya çıkarılmıştır. EK-1 içerisinde oluşturulmuş olan kurallar detaylı şekilde belirtilmiştir. Çıkarılan kurallar detaylı bir şekilde analiz edildiğinde tesise geliş sayıları, havuz kullanım sayıları, lokasyon ve kaynak nitelikleri ön plana çıkmıştır.En çarpıcı kural aile üyelerinin tesise geliş sayıları ile ilgilidir.Bireylerin tesise yirmi beş ve üzerinde girişlerinin yapılması halinde başka hiçbir niteliğe bağlı kalınmadan üyeliklerini yeniledikleri ortaya çıkarılmıştır.Bu örneği göz önüne aldığımız

vakit bireyleri tesise getirebilmek adına yapılacak bütün aktiviteler mevcut üyeler için üyelik yenileme potansiyelini arttırma sürecine katkıda bulunabilecek sonuçlar ortaya çıkarılmıştır.Uygulama sonrası çizilen karar ağacı şekli Ek-2 içerisinde gösterilmiştir.

Buradan birçok sonuç ve bu sonuçlara göre karar destek sistemlerine katkı sağlayacak sonuçlar ortaya çıkarılabilir.Satış aktiviteleri ve pr çalışmaları buna en uygun örnekler arasındadır.Ön plana çıkan nitelikler veri setinin tekrar düzenlenmesi ile detaylandırılabilir.Performans değerlendirmesi, satış fırsatı yaratımı ve üyelik yenileme analizi detaylı bir şekilde incelenerek karar destek sistemlerinin doğru planlama yapmasında yardım olabilecek kararlar çıkarılmaktadır.

3.6. DEĞERLENDİRME

Bu aşamaya kadar yapılan değerlendirmelerde modellerin veri madenciliği açısından başarısı dikkate alınmıştı. Bu aşamada ise elde edilen modellerin iş hedefleri açısından değerlendirmesi yapılacaktır. Bu kapsamda sonuçları değerlendirme, süreci gözden geçirme ve bir sonraki adımı belirleme görevleri üzerinde durulacaktır.

Daha önce belirtildiği gibi, bu çalışmada iş hedeflerine yönelik bir başarı kriteri belirlenmemiştir. Ancak kullanılan tekniklerle üye müşterilerin tercih ve davranışlarını temsil kabiliyeti yüksek modellerin elde edilebilmesini ve bu modelin potansiyel üyeliğini yenileyecek olan müşterilerin davranışlarını öngörebilme konusundaki beklentilere cevap verebilecek nitelikte olması beklentisinden söz edilmişti. Geliştirilen modellerin tamamı için her iki beklentinin de karşılandığını söylemek mümkündür. Veri

madenciliği amacına dönük değerlendirmede belirtildiği gibi her bir model hem öğrenme kümesi üzerinde sınıflandırma görevini hem de sınıma kümesi üzerinde öngörü görevini yaklaşık olarak aynı düzeylerde ve başarıyla gerçekleştirmiştir.

Üyelik yenileme dönemlerinde hedefte olan müşteri profilleri arasında yakalanacak %80-90 oranında bir başarı müşteri sadakati ve işletme gelirleri açısından önemli düzeyde işletmeye yarar sağlayacağı anlamına gelir. Özellikle başarı kriterinin %70 – 96 arasında yüksek olması bu modellerle alınacak işe yönelik kararların %4 – 30 arasında olması daha az risk içereceğini göstermektedir.

İşe yönelik uygulamalarda yapılacak durum değerlendirmesi doğrultusunda bu çalışmada uygulanan tekniklerin kullanılması söz konusu olabilir. Yapılan çalışma mümkün olduğunca titizlikle gerçekleştirilmiş, her aşamada ve her görevde alınan kararlar veri madenciliğinin yinelemeli doğası gereği tekrar tekrar değerlendirilmiştir. Başlangıç verilerinden model kurma görevine kadar her adım defalarca tekrarlanarak en uygun sonuçlara ulaşılmasına çaba harcanmıştır.

Gelinen noktada çalışma detaylandırılarak daha önemli bilgilere ulaşılması mümkündür. Geride bırakılan sürecin başarısının ancak bir sonraki aşama olan sonuç ile gerçek anlamda ortaya çıkarılması mümkün olacaktır.

Bununla birlikte, kontrol edilemeyen nedenlerle modellerin ideal düzeyde gerçekleşmiş olmama olasılığı halen mümkündür. Hedef değişkenlerin sınıflandırılmasında ve kümeleme işlemlerinde önem arz edebilecek ancak veri kümesinde mevcut olmayan değişken ya da

değişkenlerin olması söz konusu olabilir. Aranılan örüntülerin tespitinde, ilgili örüntüyü temsil eden kayıtların sayısı yetersiz kalmış olabilir. Böylesine sorunları aşmak ancak veri kümesinin güncellenmesi veya yeniden edinilmesi ile aşılabılır.

Bunun dışındaki sorunların tümü ise sürecin yenilenmesi ile aşılabılır. Bu çalışma için bir sonraki adımda yapılması gereken, sonuç aşamasına geçmeden önce elde edilen modellerin sınırlı sürede ve sınırlı sayıda müşteri üzerinde denenmesidir. Bu kısıtlı uygulamadan alınacak derslerle veri madenciliği sürecinin yenilenmesi ve sorunlardan arındırılması, algoritma sonucunun başarı düzeyini arttıracaktır.

3.7. ALGORİTMA SONUÇ RAPORU

Sonuç aşamasının gerçekleştirilmesi için veri madenciliği projesinin veriye sahip olan kurum tarafından ya da bu kurumla işbirliği içerisinde yapılabilmesi gerekmektedir. Bu çalışmanın gerçekleştirilmesi için bu türden bir işbirliği olanağının oluşturulmasına yönelik çalışmalar devam etmektedir. Özellikle ticari bilgi içeren verilerin kullanılmasında işbirliği yapmak bir yana bu verilerin paylaşılmasında dahi önemli engellerle karşılaşmaktadır. Bu sebeple uygulamanın gerçek hayat dönüştürülmesi süreci biraz zaman almaktadır.

Çalışmanın sonuç aşaması bir uygulama geliştirimi, öneri ve temennilerden oluşacaktır. Bu kapsamda sonuç planlama, izleme ve bakımın planlanması, sonuç raporu oluşturma ve projenin gözden geçirilmesi görevlerinden söz edilecektir.

Sonuç planlama, geliştirilen model ya da modellerin iş hedefleri doğrultusunda uygulanmasının planlanmasıdır. Bu uygulamanın nasıl yapılacağına, çalışmanın iş hedefleri perspektifinden bakmak gerekir. Hedef uygulamayı yönlendirecek gerçekliktir.

Sonuç planı, geliştirilen modellerin ilgili koşullara sahip olmayan müşterilerin oluşturduğu bir veri kümesi üzerinde uygulanması sonucu, potansiyel üyelik yenilemelerin tespit edilmesi üzerine kurgulanmalıdır. Bu kurgulama pazarlama faaliyeti açısından potansiyel üyelik yenileyecek müşteri kitlesi olarak kabul edilmelidir. Bu müşterilerin üyeliklerini yenileme olasılıkları, uygulanan modelin öngörü kesinliği ile orantılı olacaktır.

Sonucun gerçekleştirilmesi aynı zamanda model ya da modellerin gerçek yaşam üzerinde test edilmesi anlamını taşımaktadır. Bu aşamanın her bir adımında, veri madenciliğinin yinelenmeli doğası gereği, ortaya çıkan sonuçların gözden geçirilmesi ve gerektiğinde sürecin bazı adımlarının yinelenmesi gerekecektir.

Projenin izlenmesi ve bakımı, veri madenciliği sürecinin sonuç aşamasında da devam ettiği anlamına gelmektedir.

Bu çalışma, bir spor ve yaşam merkezine ait içerik yönetim sistemi veritabanından belirli kısıtlara göre elde edilen veri kümesi üzerinde başlatılmıştır. Veri kümesinin analize hazır hale getirilmesi için değişken seçimi, kayıtların seçimi, verilerin sorunlarından arındırılması, hedef değişkenlerin belirlenmesi ve veri kümesinin hedef değişkenlere uygun olarak yapılandırılması gibi faaliyetler yürütülmüştür.

Veri madenciliği sürecinde yürütülen çalışmaların tümü dikkate alındığında en fazla çabanın ve zamanın beklendiği gibi veri kümesinin anlaşılması ve analize hazır hale getirilmesi aşamasında tüketildiğini söylemek mümkündür. Çok fazla değişken ve kayıt içeren veri kümesini anlamak ve analize hazırlamak, oldukça karmaşık ve özenle çalışmayı gerektiren bir süreç olmuştur.

Modelleme aşamasında ise kullanılan yazılımın öğrenilmesi ve en uygun modelin üretilmesi, oldukça fazla zaman ve çaba gerektiren çalışmalar olmuştur. Yazılımın doğru bir şekilde kullanılabilmesi ve en uygun modelin üretilmesi için her bir uygulamanın defalarca tekrarlanması gerekmiştir.

Bu çalışmada uygulanan veri madenciliği sürecinin tamamı dikkate alındığında, verilerin analize hazırlanmasının veri madenciliğinde en fazla zamana ihtiyaç duyulan, en fazla çaba harcanan, en karmaşık ve en hassas kararların alındığı işlemleri kapsadığını edindiğim deneyim doğrultusunda kolaylıkla söylenebilir.

Ayrıca, projenin gerçekleştirilmesinde gelişmiş yazılım ve donanım olanaklarına sahip olmanın ne derece önemli olduğunu defalarca gözlemlenmiştir ve son olarak, veri madenciliği sürecinin doğru bir şekilde yürütülmesi ve uygulanabilmesi için hem işe yönelik hem de uygulanan tekniklere yönelik güçlü bir bilgi birikiminin gerektiğini uygulama boyunca öğrenilmiştir.

IV. SONUÇ

Bu çalışmada, veri madenciliği standart sürecinin tüm aşamaları spor ve yaşam merkezine ait üyelik sahibi müşteri bilgisi içeren veritabanından belirli kıstaslar temel alınarak seçilmiş ve çok geniş bir çalışma alanı olan veri madenciliğinin sınıflandırma ve kümeleme algoritmaları üzerinde durulmuştur.

Uygulama, birden çok bağımlı değişken üzerinde birden çok sınıflandırma tekniğini kullanılmış ve kümeleme algoritmalarında da gerekli tekniklerin karşılaştırılması üzerine kurgulanmıştır. Bu nedenle, veri madenciliğinin önemli bileşeni olan istatistik alanını temsil edecek J48 (C4.5), ID3, IB1, IBk, Naive Bayes, kümeleme algoritmasında ise Simple K-Means, Scalable K-Means algoritmaları uygulanmıştır.

Modellerin tarafsız bir şekilde karşılaştırılması için her bağımlı değişkene ilişkin tek bir veri kümesi kullanılmış ve karşılaştırma ölçütleri olarak hız, ölçeklenebilirlik, sınıflandırma kesinliği ve öngörü kesinliği kullanılmıştır.

Hız ölçütü açısından yapılan değerlendirmede, J48 (C4.5) algoritmasının tartışmasız bir şekilde avantaj sağladığı görülmüştür. Verilerin sıkça güncellendiği, modelin sürekli yinelenmesi ve karar sürecine hızla dâhil edilmesinin gerektiği durumlarda J48 (C4.5) ve ID3 algoritmalarının önerilebileceği sonucuna varılmıştır.

Modellerin, geliştirildikleri veri kümesi üzerinde gösterdikleri başarı ölçüsü olan, sınıflandırma kesinliği açısından anlamlı bir farklılık göstermedikleri görülmüştür. Tüm modellerin %70 veya daha fazla sınıflandırma başarısı gösterdiği dikkat çekicidir.

Bir diğer karşılaştırma kriteri olarak ise modellerin öğrenme kümesinden tamamen bağımsız bir veri kümesi üzerinde gösterdikleri sınıflandırma başarısını ifade eden öngörü kesinliği kullanılmıştır. Bu kritere göre de modellerin anlamlı bir farklılık göstermedikleri tespit edilmiştir.

Buradan hareketle, zaman sınırlamasının esnek olduğu, yeterli yazılım ve donanım olanaklarının mevcut olduğu ve hata riskinin en küçükleştirilmesinin esas alındığı çalışmalarda karar ağacı algoritmalarının kullanılması uygun bulunmuştur.

Ayrıca, veri sayısı arttıkça sınıflandırma ve kümeleme algoritmaları için öngörü kesinliğinin arttığı görülmüş ve bu durum, veri madenciliğinde veri sayısı arttıkça örüntü ve ilişkilerin tespit edilmesinin kolaylaşacağı şeklindeki yaklaşımın doğrulanması olarak algılanmıştır.

Bununla birlikte, az sayıda veri ile gerçekleştirilen uygulamaların beklenti seviyesinin düşük olduğu görülmüş ve öngörü başarısını göstermemiş olması, örüntü ve ilişkilerin tespitinde veri kalitesinin en önemli rolü oynayacağı gerçeğini hatırlatan nitelikte bir bulgu olarak görülmüştür.

Sonuç olarak, gerçekleştirilen standart veri madenciliği süreci ve uygulanan sınıflandırma teknikleri ve kümeleme algoritmaları ışığında, sürecin en zorlu kısmının veri hazırlama aşaması olduğu kısım. Sürecin her aşamasında iş ve tekniklere ilişkin güçlü bilgi birikimine ihtiyaç duyulduğu

görüşü üzerinde mutabık kalınmıştır. Veri sayısının ve veri kalitesinin uygulamaların başarısında önemli birer faktör olduğu, güncel ve hızlı karar verme ihtiyaçları doğrultusunda en uygun seçimin J48 (C4.5) algoritması olacaktır. Id3 algoritması ile geliştirilen uygulamanın başarısının artırılması konusunda çalışmalara devam edilmektedir. Geliştirilen uygulamanın ticari bir uygulamaya dönüştürülmesi konusunda adımlar atılmaktadır. Uygulamanın gerçek hayata sunulması sektördeki ihtiyaçlara yönelik büyük bir katkı sağlayacağı kuşkusuzdur.

KAYNAKÇA

1. Schumaker, Robert P., Solieman, Osama K. ve Chen, Hsinchun. Sports Knowledge Management and Data Mining. 2010, s. 4-10.
2. Data Mining : Concepts and Techniques. Han, J. ve Kamber, B. 2001, AP Magazine, s. 4.
3. Knowledge Discovery in Databases : An Overview. Flawley, W. J.; G., Piatetsky-Shapiro; Matheus C.J. 1992, AI Magazine, s. 57-70.
4. Mastering Data Mining : The Art and Science of CRM. Berry M.J.A.; Linoff G.S. 2000, The Art and Science of CRM, s. 7.
5. CRISP – DM Konsorsiyumu;CRISP – DM 1.0 Step – by – Step Data Mining Guide. [Çevrimiçi] [Alıntı Tarihi: 21 1 2013.] <http://asaha.com/ebook/UNjcxNzA-/CRISPWP-0800.pdf>.
6. Data Mining and Knowledge Discovery in Databases. Fayyad U., Uthurusamy R. 1996, Communications of the ACM, s. 24-26.
7. (DaimlerChrysler), Thomas Reinartz. s.l. : SPSS, NCR Systems Engineering, OHRA Verzekeringen en Bank Groep.1996
8. Cross Industry Standard Process for Data Mining. s.l. : CRISP-DM.2000
9. Edelstein H. Data Mining in Depth: Description is not Prediction. [Çevrimiçi] Mart 2003. [Alıntı Tarihi: 8 Mart 2013.] <http://www.dmreview.com/issues/20030301/6388-1.html>.
10. —. Data Mining in Depth: TI Ain't. [Çevrimiçi] Nisan 2003. [Alıntı Tarihi: 8 5 2013.] <http://www.dmreview.com/issues/20030401/6512-1.html>.
11. —. Data Mining in Depth: Using Data Mining to Find Terrorists. [Çevrimiçi] Mart 2003. [Alıntı Tarihi: 8 Mart 2013.] <http://www.dmreview.com/issues/20030501/6655-1.html>.
12. The CRISP-DM Model: The New Blueprint for Data Mining. Shearer C. 2000, Journal of Data, s. 13-22.
13. Answering Some Common Data Warehousing Questions. Man D. 1996, Direct Marketing, s. 12-15.
14. Data Mining: Concepts, Models, Methods and Algorithms. Kantardzic M. 2003, IEEE, s. 4.

15. Zhou Z. Three Perspectives of Data Mining. Artificial Intelligence. 2003, s. 139-146.
16. Giudici P. Applied Data Mining: Statistical Methods for Business and Industry. s.l. : John Wiley and Sons, 2003, s. 2.
17. The Impact of Preprocessing on Data Mining: An Evaluation of Classifier Sensitivity in Direct Marketing. Crone S.F. 2006, European Journal of Operational Research, s. 781-800.
18. Hegland M. 2001, Data Mining Techniques, s. 313-315.
19. The WEKA Data Mining Software: An Update. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten. 2009, ACM SIGKDD Explorations, s. 10-18.
20. Performance Comparison of the Data Reduction System. Patterson, D., Liu, F., Turner, D., Concepcion, A., Lynch, R. Orlando, FL. : s.n., 2008. Proceedings of the SPIE Symposium on Defense and Security.
21. Business Applications of Data Mining. C. Apte, B. Liu, E. P. D. Pednault, and P. Smyth. 2002, Communications of the ACM, s. 49-53.
22. Emerging Standards for Data Mining. Clifton C., Thuraisingham B. 2001, Computer Standards & Interfaces, s. 187-193.
23. Accelerating Customer Relationship. Swift R. 2001 : Prentice Hall PTR, 2001, s. 93.
24. Co., Two Crows. Introduction to Data Mining and Knowledge Discovery. [Çevrimiçi] [Alıntı Tarihi: 8 5 2013.] <http://www.twocrows.com/intro-dm.pdf>.
25. Zeki Veri Madenciliği : Ham Veriden Altın Bilgiye Ulaşma Yöntemleri. Alpaydın E. İstanbul : s.n., 2000. Bilisim 2000 Eğitim Semineri.
26. The KDD Process for Extracting Useful. Fayyad U., Piatetsky-Shapiro G., Smyth P. 1996, Communications of the ACM, s. 27-34.
27. Veri Ön İşleme. Oğuzlar A. 2003, Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, s. 67-76.
28. Larose D.T. Discovering Knowledge in Data: An Introduction to Data Mining. s.l. : John Wiley & Sons, 2005, s. 41-66.
29. Özkan, Dr.Yalçın. Veri Madenciliği Yöntemleri. s.l. : Papatya Yayıncılık, 2012, s. 40-45.
30. Gökhan Silahtaroğlu. Veri Madenciliği Kavram ve Algoritmaları. s.l. : Papatya Yayıncılık, 2013.

31. Montreal : Springer US, 1995. The First International Conference on Knowledge Discovery and Data Mining.
32. Journal of Data Mining and Knowledge Discovery, Springer US. 1997.

EKLER

EK-1

1. Eğer $TesiseGelisAileUyesi = 1$ ve $TesiseGelisAsilUye$ in (1, 2, 3, 4) ise $UyelikYenileme = 0$
 - *Aile üyelerinin tesise geliş sayıları (0-5), asil üyelerin tesise geliş sayıları (0-5,5-10,10-15,15-20) arasında ise üyelerin üyeliklerini yenilemedikleri gözlenmiştir.*
- 2- Eğer $TesiseGelisAileUyesi = 1$ ve $TesiseGelisAsilUye = 5$ ve **Lokasyon** in (“Kucukyali”, “Kadikoy”, “Kozyatagi”, “Uskudar”, “Goztepe”, “Icerenkoy”, “BatiAtasehir”) ise $UyelikYenileme = 1$
 - *Aile üyelerinin tesise geliş sayıları (0-5), asil üyelerinse toplam tesise geliş sayısı (20-25) arasında ve ikamet bölgesi (“Küçükalyalı”, “Kadıköy”, “Kozyatağı”, “Üsküdar”, “Göztepe”, “İçerenköy”, “Bati Ataşehir”) ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*
- 3- Eğer $TesiseGelisAileUyesi = 1$ ve $TesiseGelisAsilUye = 5$ ve **Lokasyon = ”Bostanci”** ve **Kaynak = ”TvReklamı”** ise $UyelikYenileme = 1$
 - *Aile üyelerinin tesise geliş sayıları (0-5), asil üyelerinse toplam tesise geliş sayısı (20-25) arasında ve ikamet bölgesi (“Bostancı”) ve tesise ulaşım kaynağı (“Tv Reklamı”) ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*
- 4- Eğer $TesiseGelisAileUyesi = 1$ ve $TesiseGelisAsilUye = 5$ ve **Lokasyon = ”Bostanci”** ve **Kaynak = ”Tavsiye”** ise $UyelikYenileme = 0$

- *Aile üyelerinin tesise geliş sayıları (0-5), asil üyelerinse toplam tesise geliş sayısı (20-25) arasında ve ikamet bölgesi (“Bostancı”) ve tesise ulaşım kaynağı (“Tavsiye”) ise üyelerin üyeliklerini yenilemedikleri gözlenmiştir.*
- 5- Eğer TesiseGelisAileUyesi = 1 ve TesiseGelisAsilUye = 6 ve Kaynak in (“Tavsiye”, ”Billboard”, “Brosur”, “TopluMail”, “Sms”) ise UyelikYenileme = 1**
- *Aile üyelerinin tesise geliş sayıları (0-5), asil üyelerinse toplam tesise geliş sayısı (25-30) arasında ve tesise ulaşım kaynağı (“Tavsiye”, ”Billboard”, “Broşür”, “Toplu Mail”, “Sms”) ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*
- 6- Eğer TesiseGelisAileUyesi = 1 ve TesiseGelisAsilUye = 6 ve Kaynak = ”Stand” ve Lokasyon in (“Goztepe”, “Atasehir”, “Kucukyali”, “BatiAtasehir”) ise UyelikYenileme = 1**
- *Aile üyelerinin tesise geliş sayıları (0-5), asil üyelerinse toplam tesise geliş sayısı (25-30) arasında ve tesise ulaşım kaynağı (“Stand”) ve ikamet adresi (“Göztepe”, “Ataşehir”, “Küçükyalı”, “Batı Ataşehir”) ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*
- 7- Eğer TesiseGelisAileUyesi = 1 ve TesiseGelisAsilUye = 7 ise UyelikYenileme = 1**
- *Aile üyelerinin tesise geliş sayıları (0-5), asil üyelerinse toplam tesise geliş sayısı (25-30) arasında ve tesise ulaşım kaynağı (“Stand”) ve ikamet adresi (“Bostancı”, “Ataşehir”, “Küçükyalı”) ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

8- Eğer TesiseGelisAileUyesi = 1 ve TesiseGelisAsilUye = 8 ve Kaynak in ("Tavsiye", "Billboard", "Sms", "TvReklami", "Internet") ise UyelikYenileme = 1

- *Aile üyelerinin tesise geliş sayıları (0-5), asil üyelerinse toplam tesise geliş sayısı (35-40) arasında ve tesise ulaşım kaynağı ("Tavisye", "Billboard", "Sms", "Tv Reklami", "Internet") ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

9- Eğer TesiseGelisAileUyesi = 1 ve TesiseGelisAsilUye = 8 ve Kaynak in ("Basin", "TopluMail") ise UyelikYenileme = 0

- *Aile üyelerinin tesise geliş sayıları (0-5), asil üyelerinse toplam tesise geliş sayısı (35-40) arasında ve tesise ulaşım kaynağı ("Basın", "Toplu Mail") ise üyelerin üyeliklerini yenilemedikleri gözlenmiştir.*

10- Eğer TesiseGelisAileUyesi = 1 ve TesiseGelisAsilUye = 8 ve Kaynak = "Stand" ve Lokasyon = "BatiAtasehir" ise UyelikYenileme = 1

- *Aile üyelerinin tesise geliş sayıları (0-5), asil üyelerinse toplam tesise geliş sayısı (35-40) arasında ve tesise ulaşım kaynağı ("Stand") ve ikamet adresi ("Batı Ataşehir") ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

11- Eğer TesiseGelisAileUyesi = 1 ve TesiseGelisAsilUye = 8 ve Kaynak = "Stand" ve Lokasyon = "Uskudar" ise UyelikYenileme = 0

- *Aile üyelerinin tesise geliş sayıları (0-5), asil üyelerinse toplam tesise geliş sayısı (35-40) arasında ve tesise ulaşım kaynağı ("Stand") ve*

ikamet adresi (“Üsküdar”) ise üyelerin üyeliklerini yenilemedikleri gözlenmiştir.

12- Eğer TesiseGelisAileUyesi = 1 ve TesiseGelisAsilUye = 8 ve Kaynak = ”Brosur” ve Lokasyon = ”Kozyatagi” ise UyelikYenileme = 0

- *Aile üyelerinin tesise geliş sayıları (0-5), asil üyelerinse toplam tesise geliş sayısı (35-40) arasında ve tesise ulaşım kaynağı (“Broşür”) ve ikamet adresi (“Kozyatağı”) ise üyelerin üyeliklerini yenilemedikleri gözlenmiştir.*

13- Eğer TesiseGelisAileUyesi = 1 ve TesiseGelisAsilUye = 8 ve Kaynak = ”Brosur” ve Lokasyon in (“Icerenkoy”, “BatiAtasehir”) ise UyelikYenileme = 1

- *Aile üyelerinin tesise geliş sayıları (0-5), asil üyelerinse toplam tesise geliş sayısı (35-40) arasında ve tesise ulaşım kaynağı (“Broşür”) ve ikamet adresi (“İçerenköy”, “Batı Ataşehir”) ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

14- Eğer TesiseGelisAileUyesi = 1 ve TesiseGelisAsilUye = 9 ve HavuzKullanimSayisiAsilUye in (1, 3, 4, 5) ise UyelikYenileme = 1

- *Aile üyelerinin tesise geliş sayıları (0-5), asil üyelerinse toplam tesise geliş sayısı (40-45) arasında ve asil üyelerin havuz kullanım sayıları (0-5, 10-15, 15-20, 20-25) ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

15- Eğer TesiseGelisAileUyesi = 1 ve TesiseGelisAsilUye = 9 ve HavuzKullanimSayisiAsilUye = 2 ve Lokasyon = “Kucukyali“ ise UyelikYenileme = 1

- *Aile üyelerinin tesise geliş sayıları (0-5), asil üyelerinse toplam tesise geliş sayısı (40-45) arasında ve asil üyelerin havuz kullanım sayıları (5-10) ve ikamet adresi (“Küçükyalı”) ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

16- Eğer $TesiseGelisAileUyesi = 1$ ve $TesiseGelisAsilUye = 9$ ve $HavuzKullanımSayisiAsilUye = 2$ ve $Lokasyon = “Atasehir“$ ise $UyelikYenileme = 0$

- *Aile üyelerinin tesise geliş sayıları (0-5), asil üyelerinse toplam tesise geliş sayısı (40-45) arasında ve asil üyelerin havuz kullanım sayıları (5-10) ve ikamet adresi (“Atasehir”) ise üyelerin üyeliklerini yenilemedikleri gözlenmiştir.*

17- Eğer $TesiseGelisAileUyesi = 1$ ve $TesiseGelisAsilUye = 10$ ise $UyelikYenileme = 1$

- *Aile üyelerinin tesise geliş sayıları (0-5), asil üyelerinse toplam tesise geliş sayısı (45-50) arasında ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

18- Eğer $TesiseGelisAileUyesi = 2$ ve $HavuzKullanımSayisiAsilUye = 1$ ise $UyelikYenileme = 0$

- *Aile üyelerinin tesise geliş sayıları (5-10), asil üyelerin havuz kullanım sayıları (0-5) arasında ise üyelerin üyeliklerini yenilemedikleri gözlenmiştir.*

19- Eğer $TesiseGelisAileUyesi = 2$ ve $HavuzKullanımSayisiAsilUye in (3, 5)$ ise $UyelikYenileme = 1$

- *Aile üyelerinin tesise geliş sayıları (5-10), asil üyelerin havuz kullanım sayıları (10-15, 20-25) arasında ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

20- Eğer TesiseGelisAileUyesi = 2 ve HavuzKullanımSayisiAsilUye = 2 ve Lokasyon in (“BatiAtasehir”, “Goztepe”, “Icerenkoy”) ise UyelikYenileme = 1

- *Aile üyelerinin tesise geliş sayıları (5-10), asil üyelerin havuz kullanım sayıları (5-10) arasında ve ikamet adresi (“Batı Ataşehir”, “Göztepe”, “İçerenköy”) ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

21- Eğer TesiseGelisAileUyesi = 2 ve HavuzKullanımSayisiAsilUye = 2 ve Lokasyon = “Kucukyali” ve Kaynak = “Stand” ise UyelikYenileme = 0

- *Aile üyelerinin tesise geliş sayıları (5-10), asil üyelerin havuz kullanım sayıları (5-10) arasında ve ikamet adresi (“Kucukyali”) ve tesise ulaşım kaynağı (“Stand”) ise üyelerin üyeliklerini yenilemedikleri gözlenmiştir.*

22- Eğer TesiseGelisAileUyesi = 2 ve HavuzKullanımSayisiAsilUye = 2 ve Lokasyon = “Kucukyali” ve Kaynak = “Billboard” ise UyelikYenileme = 1

- *Aile üyelerinin tesise geliş sayıları (5-10), asil üyelerin havuz kullanım sayıları (5-10) arasında ve ikamet adresi (“Küçükaly”) ve tesise ulaşım kaynağı (“Billboard”) ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

23- Eğer TesiseGelisAileUyesi = 3 ve HavuzKullanımSayisiAsilUye in (1,2) ise UyelikYenileme = 0

- *Aile üyelerinin tesise geliş sayıları (10-15), asil üyelerin havuz kullanım sayıları (0-5,5-10) arasında ise üyelerin üyeliklerini yenilemedikleri gözlenmiştir.*

24- Eğer $TesiseGelisAileUyesi = 3$ ve $HavuzKullanimSayisiAsilUye$ in (3,4) ise $UyelikYenileme = 1$

- *Aile üyelerinin tesise geliş sayıları (10-15), asil üyelerin havuz kullanım sayıları (10-15,15-20) arasında ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

25- Eğer $TesiseGelisAileUyesi = 4$ ve $HavuzKullanimSayisiAsilUye$ in (1,2) ise $UyelikYenileme = 0$

- *Aile üyelerinin tesise geliş sayıları (15-20), asil üyelerin havuz kullanım sayıları (0-5,5-10) arasında ise üyelerin üyeliklerini yenilemedikleri gözlenmiştir.*

26- Eğer $TesiseGelisAileUyesi = 4$ ve $HavuzKullanimSayisiAsilUye = 4$ ise $UyelikYenileme = 1$

- *Aile üyelerinin tesise geliş sayıları (15-20), asil üyelerin havuz kullanım sayıları (15-20) arasında ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

27- Eğer $TesiseGelisAileUyesi = 5$ ve $TesiseGelisAsilUye$ in (1, 2, 10) ise $UyelikYenileme = 0$

- *Aile üyelerinin tesise geliş sayıları (20-25), asil üyelerinse toplam tesise geliş sayısı (0-5,5-10,45-50) arasında ise üyelerin üyeliklerini yenilemedikleri gözlenmiştir.*

28- Eğer TesiseGelisAileUyesi = 5 ve TesiseGelisAsilUye = 3 ve Lokasyon in (“BatiAtasehir”, “Kozyatagi”, “Maltepe”) ise UyelikYenileme = 0

- *Aile üyelerinin tesise geliş sayıları (20-25), asil üyelerinse toplam tesise geliş sayısı (10-15) arasında ve ikamet adresi (“Batı Ataşehir”, “Kozyatağı”, “Maltepe”) ise üyelerin üyeliklerini yenilemedikleri gözlenmiştir.*

29- Eğer TesiseGelisAileUyesi = 5 ve TesiseGelisAsilUye = 3 ve Lokasyon = ”Icerenkoy” ve Kaynak = “Brosur” ve OzelDersKullanimiAileUyesi = 1 ve OzelDersKullanimiAsilUye =1 ve GrupDersKullanimiAileUyesi = 1 ve GrupDersKullanimiAsilUye = 1 ve HavuzKullanımSayisiAileUyesi = 1 ve HavuzKullanımSayisiAsilUye = 1 ve SpaKullanimiAileUyesi = 1 ve SpaKullanimiAsilUye = 1 ve SozlesmeFiyati = 5 ise UyelikYenileme = 0

- *Aile üyelerinin tesise geliş sayıları (20-25), asil üyelerinse toplam tesise geliş sayısı (10-15) arasında ve ikamet adresi (“İçerenköy”) ve tesise ulaşım kaynağı (“Brosur”) ve asil-aile üyelerinin grup ders kullanımı (0-5) arasında ve asil-aile üyelerinin havuz kullanım sayısı (0-5) arasında ve asil-aile üyelerinin spa kullanım sayıları (0-5) arasında ve sözleşme fiyatı (4000-5000) arasında ise üyelerin üyeliklerini yenilemedikleri gözlenmiştir.*

30- Eğer TesiseGelisAileUyesi = 5 ve TesiseGelisAsilUye in (4, 5) ise UyelikYenileme = 1

- *Aile üyelerinin tesise geliş sayıları (20-25), asil üyelerinse toplam tesise geliş sayısı (15-20, 20-25) arasında ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

31- Eğer TesiseGelisAileUyesi = 5 ve TesiseGelisAsilUye = 6 ve Lokasyon in (“Kozyatağı”, “Atasehir”) ise UyelikYenileme = 1

- *Aile üyelerinin tesise geliş sayıları (20-25), asil üyelerinse toplam tesise geliş sayısı (25-30) arasında ve ikamet adresi (“Kozyatağı”, “Atasehir”) ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

32- Eğer TesiseGelisAileUyesi = 5 ve TesiseGelisAsilUye = 6 ve Lokasyon in (“BatiAtasehir”, “Kucukyali”) ise UyelikYenileme = 0

- *Aile üyelerinin tesise geliş sayıları (20-25), asil üyelerinse toplam tesise geliş sayısı (25-30) arasında ve ikamet adresi (“Batı Ataşehir”, “Küçükalyalı”) ise üyelerin üyeliklerini yenilemedikleri gözlenmiştir.*

33- Eğer TesiseGelisAileUyesi = 5 ve TesiseGelisAsilUye = 6 ve Lokasyon = “Uskudar” ve Kaynak = “Brosur” ise UyelikYenileme = 1

- *Aile üyelerinin tesise geliş sayıları (20-25), asil üyelerinse toplam tesise geliş sayısı (25-30) arasında ve ikamet adresi (“Üsküdar”) ve tesise ulaşım kaynağı (“Broşür”) ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

34- Eğer TesiseGelisAileUyesi = 5 ve TesiseGelisAsilUye = 6 ve Lokasyon = “Uskudar” ve Kaynak = “Tavsiye” ise UyelikYenileme = 0

- *Aile üyelerinin tesise geliş sayıları (20-25), asil üyelerinse toplam tesise geliş sayısı (25-30) arasında ve ikamet adresi (“Üsküdar”) ve*

tesise ulaşım kaynağı (“Tavsiye”) ise üyelerin üyeliklerini yenilemedikleri gözlenmiştir.

35- Eğer TesiseGelisAileUyesi = 5 ve TesiseGelisAsilUye = 7 ve Lokasyon = “Goztepe” ise UyelikYenileme = 1

- *Aile üyelerinin tesise geliş sayıları (20-25), asil üyelerinse toplam tesise geliş sayısı (30-35) arasında ve ikamet adresi (“Göztepe”) ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

36- Eğer TesiseGelisAileUyesi = 5 ve TesiseGelisAsilUye = 7 ve Lokasyon in (“Atasehir”, “BatiAtasehir”) ise UyelikYenileme = 0

- *Aile üyelerinin tesise geliş sayıları (20-25), asil üyelerinse toplam tesise geliş sayısı (30-35) arasında ve ikamet adresi (“Ataşehir”, “Bati Ataşehir”) ise üyelerin üyeliklerini yenilemedikleri gözlenmiştir.*

37- Eğer TesiseGelisAileUyesi = 5 ve TesiseGelisAsilUye = 8 ve Lokasyon in (“Atasehir”, “Icerenkoy”) ise UyelikYenileme = 0

- *Aile üyelerinin tesise geliş sayıları (20-25), asil üyelerinse toplam tesise geliş sayısı (35-40) arasında ve ikamet adresi (“Ataşehir”, “İçerenköy”) ise üyelerin üyeliklerini yenilemedikleri gözlenmiştir.*

38- Eğer TesiseGelisAileUyesi = 5 ve TesiseGelisAsilUye = 8 ve Lokasyon = “Kadikoy” ise UyelikYenileme = 1

- *Aile üyelerinin tesise geliş sayıları (20-25), asil üyelerinse toplam tesise geliş sayısı (35-40) arasında ve ikamet adresi (“Kadıköy”) ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

39- Eğer TesiseGelisAileUyesi = 5 ve TesiseGelisAsilUye = 9 ve Lokasyon = “Icerenkoy” ise UyelikYenileme = 0

- *Aile üyelerinin tesise geliş sayıları (20-25), asil üyelerinse toplam tesise geliş sayısı (40-45) arasında ve ikamet adresi (“İçerenköy”) ise üyelerin üyeliklerini yenilemedikleri gözlenmiştir.*

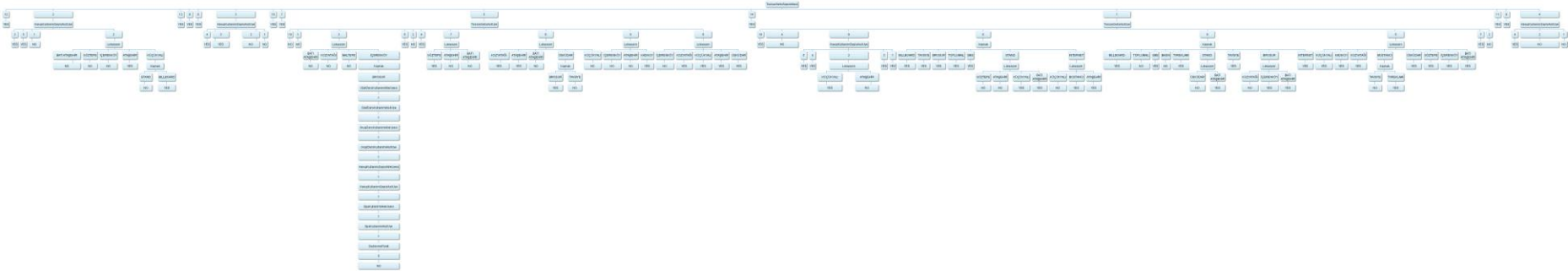
40- Eğer $TesiseGelisAileUyesi = 5$ ve $TesiseGelisAsilUye = 9$ ve Lokasyon in (“Kozyatagi”, “Kucukyali”, “Atasehir”, “Uskudar”) ise $UyelikYenileme = 1$

- *Aile üyelerinin tesise geliş sayıları (20-25), asil üyelerinse toplam tesise geliş sayısı (40-45) arasında ve ikamet adresi (“Kozyatağı”, “Küçükyalı”, “Ataşehir”, “Üsküdar”) ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

41- Eğer $TesiseGelisAileUyesi$ in (6, 7, 8, 9, 10, 11, 12, 13, 14) ise $UyelikYenileme = 1$

- *Aile üyelerinin tesise geliş sayıları (25-30,30-35,35-40,40-45,45-50,50-55,55-60,60-65,65-70) aralıklarında ise üyelerin üyeliklerini yeniledikleri gözlenmiştir.*

EK-2



ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı : Melih Keskin
Uyruğu : Türkiye
Doğum Yeri / Tarihi : 11.09.1987 / Fatih
Medeni Hal : Evli
Telefon : 0 (542) 581 44 41
E-mail : melihkeskinn@gmail.com

İş Deneyimleri

| Firma | Departman / Pozisyon | Tarih |
|-----------------------|--|-------------|
| Senkron Yazılım | Yazılım Geliştirme / Proje Yöneticiliği | 2012 - .. |
| | Yazılım Geliştirme / Proje Sorumlusu | 2010 - 2012 |
| Ekin Proje | Uygulama Geliştirme / Bilgi Yönetimi Uzmanlığı | 2009 - 2010 |
| İstanbul Üniversitesi | Bilgisayar Asistanlığı | 2007 - 2009 |

Eğitim

| Derece | Eğitim Birimi | Tarih |
|---------------|---|-----------|
| Yüksek Lisans | Okan Üniversitesi /Bilgisayar Mühendisliği | 2011 - .. |
| Lisans | Anadolu Üniversitesi / İşletme | 2011 |
| Ön Lisans | İstanbul Üniversitesi / Bilgi Tek. ve Programlama | 2009 |

Kurs

| Adı | Süresi | Tarih |
|--|---------------|--------------|
| Bilge Adam Yazılım Geliştirme Ve Veritabanı Uzmanlığı | 10 Ay | 2010 |

Sertifika

| Adı | Tarih |
|--|--------------|
| Microsoft Certified Software Developer (MCPD) | 2011 |

Yabancı Dil

İngilizce

Hobiler

Basketbol, Kitap Okuma, Film İzlemek, Müzik Dinlemek



ULUSLARARASI KATILIMLI
15. ULUSAL BİYOİSTATİSTİK KONGRESİ
20-23 AĞUSTOS 2013



Katılım Belgesi

Sayın Melih KESKİN

*Adnan Menderes Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı ile Biyoistatistik Derneği tarafından
20-23 Ağustos 2013 tarihleri arasında Aydın'da düzenlenen
"Uluslararası Katımlı 15. Ulusal Biyoistatistik Kongresi"ne katılmıştır.*


Prof. Dr. Mevlüt TÜRE
Düzenleme Kurulu Başkanı


Prof. Dr. Kadir SÜMBÜLOĞLU
Biyostatistik Derneği Başkanı



ULUSLARARASI KATILIMLI
15. ULUSAL BİYOİSTATİSTİK KONGRESİ
20-23 AĞUSTOS 2013



ADNAN MENDERES ÜNİVERSİTESİ
DİDİM SOSYAL TESİSLERİ KÜLTÜR VE
KONGRE MERKEZİ
DİDİM, AYDIN

03.06.2013

Sayın

Melih KESKİN

Pınar YILDIRIM

Biyoistatistik Derneği ve Adnan Menderes Üniversitesi tarafından Didim, Aydın'da düzenlenecek olan "Uluslararası Katılımlı 15. Ulusal Biyoistatistik Kongresi"nde sunulması için göndermiş olduğunuz "Spor ve Yaşam Merkezlerinde Veri Madenciliği Çalışması" isimli bildiriniz sözlü bildiri olarak kabul edilmiştir.

Bilgilerinize sunar, kongremizde görüşmek ümidi ile çalışmalarınızda başarılar dileriz.

Doç. Dr. İmran KURT ÖMÜRLÜ

Prof. Dr. Mevlüt TÜRE

Kongre Sekreteri

Kongre Başkanı



01.01.2013

İlgili Makamın Dikkatine,

09.03.2010 tarihinden bu yana bizimle birlikte çalışmakta olan 14408316748 TC Kimlik numaralı personelimiz Melih Keskin'e, yüksek lisans sebebiyle tez konusu sektörde çalışmalar yapmak üzere şirketimizin veri tabanı kullanma onayı verilmiştir.

Senkron Bilgisayar Otomasyon Sistemleri

Eğitim Dan. San. Tic. Ltd. Şti.

**SENKRON BİLGİSAYAR OTOMASYON
SİSTEMLERİ EĞİTİMİ DANIŞMANLIK
SAN. VE TİC. LTD. ŞTİ.**
Yıldız Teknik Üniversitesi Orduyeri Kampüsü
Teknoloji Geliştirme Bölgesi B2 Blok No 301
34220 Esentepe / BEŞİKTAŞ/İSTANBUL Tel:0212 329 20 50
E-Posta: info@senkron.com.tr