

**YILDIZ TEKNİK ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**DİNAMİK VERİTABANLARINDA VERİ MADENCİLİĞİ  
VE ANLAMSAL SORGU ENİYİLEMESİ**

**Matematik Müh. Özgür SARIHAN**

**F.B.E. Matematik Mühendisliği Anabilim Dalı Matematik Mühendisliği Programında  
Hazırlanan**

**YÜKSEK LİSANS TEZİ**

**Tez Danışmanı : Doç. Dr. Ayla ŞAYLI (YTÜ)**

**İSTANBUL, 2011**

# İÇİNDEKİLER

Sayfa

|                                                          |      |
|----------------------------------------------------------|------|
| İÇİNDEKİLER.....                                         | ii   |
| SİMGE LİSTESİ .....                                      | iv   |
| KISALTMA LİSTESİ .....                                   | v    |
| ŞEKİL LİSTESİ .....                                      | vi   |
| ÇİZELGE LİSTESİ .....                                    | vii  |
| ÖNSÖZ .....                                              | viii |
| ÖZET .....                                               | ix   |
| ABSTRACT .....                                           | x    |
| 1. GİRİŞ .....                                           | 11   |
| 1.1 SQL Veri Tabanı Genel Özellikleri .....              | 12   |
| 1.2 Programlama Dillerinde Veritabanı Bağlantıları ..... | 14   |
| 1.3 Visual Basic.Net Programlama Dili .....              | 14   |
| 2. VERİ MADENCİLİĞİ VE VERİ MADENCİLİĞİ SÜRECİ.....      | 16   |
| 2.1 Veri Madenciliği .....                               | 16   |
| 2.2 Veri Madenciliği Süreci.....                         | 18   |
| 2.2.1 Problemin Tanımlanması.....                        | 19   |
| 2.2.2 Verilerin Hazırlanması.....                        | 19   |
| 2.2.3 Modelin Kurulması ve Değerlendirilmesi .....       | 20   |
| 2.2.4 Modelin Kullanılması .....                         | 20   |
| 2.2.5 Modelin İzlenmesi .....                            | 21   |
| 3. VERİ MADENCİLİĞİ YÖNTEMLERİ .....                     | 21   |
| 3.1 Karar Ağaçları .....                                 | 21   |
| 3.1.1 Ağaç Oluşturma .....                               | 21   |
| 3.1.2 Ağaç Budama.....                                   | 22   |
| 3.2 Karar Ağacı Oluşturma.....                           | 22   |
| 3.3 En İyi Bölen Nitelik Nasıl Belirlenir? .....         | 22   |
| 3.4 Bilgi Kazancı (ID3 / C4.5).....                      | 23   |
| 3.5 Örnek (Entropi) Algoritma Hesaplaması.....           | 24   |
| 4. VERİ MADENCİLİĞİNİ KULLANAN SİSTEMLER .....           | 33   |
| 4.1 Statistica.....                                      | 33   |
| 4.2 Dtree .....                                          | 36   |
| 4.3 Cart .....                                           | 39   |
| 5. SQL – YAPISAL SORGU DİLİ .....                        | 44   |
| 5.1 SQL Server Tabloları.....                            | 44   |
| 5.1.1 Tablo Oluşturulması (CREATE TABLE) .....           | 44   |
| 5.2 Sorgu.....                                           | 46   |
| 5.2.1 SELECT .....                                       | 47   |
| 6. ÖĞRENEN SİSTEMLER .....                               | 53   |
| 6.1 Bilgiden Bilgi Öğrenme.....                          | 53   |

|       |                                                                                      |    |
|-------|--------------------------------------------------------------------------------------|----|
| 6.2   | Öğrenme Metotları.....                                                               | 54 |
| 6.2.1 | Siegel Yöntemi .....                                                                 | 55 |
| 6.2.2 | Shekhar Yöntemi .....                                                                | 57 |
| 6.2.3 | İndeksli Arama Yöntemi.....                                                          | 62 |
| 7.    | UYGULAMA .....                                                                       | 62 |
| 7.1   | Uygulamanın Yapısı .....                                                             | 64 |
| 7.2   | Uygulamanın Çalıştırılması .....                                                     | 67 |
| 7.3   | Uygulamanın Sonuçları .....                                                          | 68 |
| 7.4   | Sorguların İyileştirmesi .....                                                       | 71 |
| 8.    | SONUÇLAR.....                                                                        | 73 |
|       | KAYNAKLAR.....                                                                       | 74 |
|       | EKLER .....                                                                          | 75 |
| Ek 1  | Araba Tablosu veri bilgileri (İlk 100 Kayıtlık).....                                 | 76 |
| Ek 2  | Araba Tablosu için sorgu metodu ile öğrenilen kurallar (İlk 50 kural).....           | 83 |
| Ek 3  | Araba Tablosu için tam arama metoduyla öğrenilen kurallar (İlk 50 kural) .....       | 84 |
| Ek 4  | Araba Tablosu için indeksli arama metoduyla öğrenilen kurallar (İlk 50 kural) ..     | 86 |
| Ek 5  | Uygulama algoritması kod örneği .....                                                | 87 |
|       | Uygulamanın ilk aşamasında kullanılan var olan kuralların bulunması (sol taraf)..... | 87 |
|       | Uygulamanın elde edilen kurallardan yeni kural öğrenmesi (sağ taraf).....            | 88 |
|       | ÖZGEÇMİŞ.....                                                                        | 90 |

## SİMGE LİSTESİ

|     |                                                     |
|-----|-----------------------------------------------------|
| @   | Karşılaştırma operatör göstergesi                   |
| --> | Kurallarda sol ve sağ tarafların ayrıldığı gösterge |
| (?) | Bilinmeyen değer                                    |
| Op  | Operatör                                            |
| P   | Bağlantılar ve ilişkiler                            |
| not | Değil bağlantısı                                    |
| or  | Veya bağlantısı                                     |

## KISALTMA LİSTESİ

|          |                                                            |
|----------|------------------------------------------------------------|
| ADO      | Aktif Veri Nesnesi (Activex Data Objects)                  |
| ANSI     | American National Standards Institute                      |
| CRISP-DM | Cross Industry Process for Data Mining                     |
| DDL      | Veri Tanımlama Dili Cümleleri                              |
| DML      | Veri Değiştirme Dili Cümleleri                             |
| InSeQt   | Index-Search Query optimization                            |
| ISO      | International Organization of Standardization              |
| ODBC     | Açık Veritabanı Bağlantısı (Open Database Connectivity)    |
| QE       | Sorgu Çalıştırma (Query Execute)                           |
| RDBMS    | Relational DataBase Management System                      |
| SQL      | Yapısal Sorgu Dili (Structured Query Language)             |
| SQO      | Anlamsal Sorgu İyileştirmesi (Semantic Query Optimization) |

## ŞEKİL LİSTESİ

|            |                                                                              |    |
|------------|------------------------------------------------------------------------------|----|
| Şekil 2.1  | Veri madenciliği standart süreci.....                                        | 18 |
| Şekil 3.1  | Örnek karar ağacı .....                                                      | 22 |
| Şekil 3.3  | Oluşturulan karar ağacı .....                                                | 31 |
| Şekil 3.4  | Öğrenim seviyesine göre çocuk sayısı .....                                   | 32 |
| Şekil 4.1  | Statistica veritabanı tanıma .....                                           | 33 |
| Şekil 4.2  | Statistica veri madenciliği yöntemi seçimi.....                              | 34 |
| Şekil 4.3  | Statistica bağımlı-bağımsız değişken seçimi.....                             | 34 |
| Şekil 4.4  | Statistica değişken dönüşümü .....                                           | 35 |
| Şekil 4.5  | Statistica sonuç .....                                                       | 35 |
| Şekil 4.6  | Dtreg hedef ve belirleyici seçimi.....                                       | 37 |
| Şekil 4.7  | Dtreg metod seçimi .....                                                     | 37 |
| Şekil 4.8  | Dtreg analizin başlatılması .....                                            | 38 |
| Şekil 4.9  | Dtreg analiz sonucu .....                                                    | 38 |
| Şekil 4.10 | Dtreg karar ağacı .....                                                      | 39 |
| Şekil 4.11 | Cart kullanılabilir veri alanları .....                                      | 40 |
| Şekil 4.12 | Cart hedef ve belirleyici seçimi.....                                        | 40 |
| Şekil 4.13 | Cart yöntemi seçimi .....                                                    | 41 |
| Şekil 4.14 | Ağaç özelliği seçimi .....                                                   | 41 |
| Şekil 4.15 | Sınıflama metodu seçimi .....                                                | 41 |
| Şekil 4.16 | Cart analiz sonucu .....                                                     | 42 |
| Şekil 4.17 | Programın çıktısı olan karar ağacı.....                                      | 42 |
| Şekil 4.18 | Cart, karar ağacı yaprakları detay 1.....                                    | 43 |
| Şekil 4.19 | Cart, karar ağacı yaprakları detay 2.....                                    | 43 |
| Şekil 6.1  | Data dağılımı tablosu.....                                                   | 58 |
| Şekil 6.2  | Bağlantı tanımlamaları .....                                                 | 59 |
| Şekil 6.3  | Operatörlerin zıtlığı .....                                                  | 60 |
| Şekil 6.4  | A1,A2 birleşim operatörü (Union(A1,A2)) ve ortak a değerinin kullanımı ..... | 60 |
| Şekil 6.5  | A1,A2 kesişim operatörü (intersection(A1,A2)) ve ortak a değerinin kullanımı | 61 |
| Şekil 7.1  | Uygulama ana ekranı.....                                                     | 64 |
| Şekil 7.2  | Veritabanının tanıtılması ve metod seçimi.....                               | 65 |
| Şekil 7.3  | Bekleme ekranı.....                                                          | 66 |
| Şekil 7.4  | Çıkış ekranı .....                                                           | 67 |
| Şekil 7.5  | Sınırlandırma ekranı .....                                                   | 68 |
| Şekil 7.6  | 5000 kayıta süre / 100 kural .....                                           | 70 |
| Şekil 7.7  | 5000 kayıta bir sn de öğrenilen kural sayısı.....                            | 70 |
| Şekil 7.8  | 25000 kayıta süre / 100 kural .....                                          | 70 |
| Şekil 7.9  | 25000 kayıta bir sn de öğrenilen kural sayısı.....                           | 70 |
| Şekil 7.10 | Zaman-Kural sayısı ölçümü .....                                              | 71 |

## ÇİZELGE LİSTESİ

|                |                                                                             |    |
|----------------|-----------------------------------------------------------------------------|----|
| Çizelge 3.1    | Entropi Algoritması için örnek tablo.....                                   | 24 |
| Çizelge 3.2    | Tablonun toplam entropisi .....                                             | 24 |
| Çizelge 3.3    | Tablonun cinsiyete göre dağılımı .....                                      | 24 |
| Çizelge 5.1    | TableEx tablosu ve veri bilgileri .....                                     | 48 |
| Çizelge 5.2    | Sütundaki tüm kayıtları getiren sorgu sonucu.....                           | 48 |
| Çizelge 5.3    | Sütundaki tek kayıtları getiren sorgu sonucu .....                          | 48 |
| Çizelge 5.4    | Sütunda en üstteki iki kaydı getiren sorgu sonucu .....                     | 49 |
| Çizelge 5.5    | Seçilen sütunları getiren sorgu sonucu.....                                 | 49 |
| Çizelge 5.6    | Sütun başlığı değiştirilmiş sorgu sonucu.....                               | 50 |
| Çizelge 5.7    | Koşul kullanılmış sorgu sonucu .....                                        | 51 |
| Çizelge 5.8    | Grup yapılmış sorgu sonucu.....                                             | 52 |
| Çizelge 5.9    | Grup koşulu kullanılmış sorgu sonucu.....                                   | 52 |
| Çizelge 5.10   | Sıralama kullanılmış sorgu sonucu .....                                     | 53 |
| Çizelge 7.1    | Sorgu tablosu.....                                                          | 63 |
| Çizelge 7.2    | Sonuç tablosu .....                                                         | 69 |
| Çizelge 7.3    | Sorgu süreleri karşılaştırması .....                                        | 72 |
| Çizelge Ek 1.1 | Araba Tablosu veri bilgileri (İlk 100 Kayıtlık).....                        | 76 |
| Çizelge Ek 2.1 | Araba Tablosu için sorgu metodu ile öğrenilen kurallar (İlk 50 kural) ..... | 83 |
| Çizelge Ek 3.1 | Araba Tablosu için tam arama metoduyla öğrenilen kurallar (İlk 50 kural) .. | 84 |
| Çizelge Ek 4.1 | Araba Tablosu için indeks aramayla öğrenilen kurallar (İlk 50 kural).....   | 86 |

## ÖNSÖZ

Geçtiğimiz her gün hayatlarımızda gerek iş gerekse özel kullanım amacıyla bilgisayarlar daha fazla yer etmektedirler. Artık bilgi sistemlerinin kullanılmadığı bir alan neredeyse kalmamıştır. Bilgisayar kullanımı ise an ve an kendine gelişecek yeni alanlar yaratmakta ve hızla gelişimini sürdürmektedir. Her kullanım alanında gelişimi, tutulmakta olan verilerin sağlıklı olarak işlenip yorumlanması ile gerçekleştirilir. Bu sebepten dolayı büyük verileri depolayacak ve analiz için ulaşımının kolay olacağı veritabanı sistemleri kurulmuştur.

Bu tezde, bilgiden bilgi çıkarma metotları kullanarak dinamik bir veritabanı üzerinde analizler yapılmıştır. Veri madenciliği metotlarının veri analizine katkısı araştırılmıştır. Sorgu iyileştirilmesi için kendinden öğrenen dinamik bir sistem kurulmuş ve test edilmiştir.

Bu tezin hazırlanmasında yardımlarını ve desteğini esirgemeyen değerli hocam Sayın Doç. Dr. Ayla Şaylı'ya çok teşekkür ederim.

## ÖZET

Bilgi çağı diye anılan çağımızda veri ve bunun kullanımı tüm dünyada yönetilmesi gereken bir konu olmuştur. Bu konu her geçen gün kendine geliştirecek yeni alanlar yaratmaktadır. Bu sayede insan hatalarını en aza indirgeyerek aynı işleri otomatik olarak yapacak sistemler yaygınlaşmaktadır. Böylece insan üzerindeki yük giderek hafiflemektedir.

Bilgisayar sistemlerini insan işleyişine benzetecek olursak, insan beyni gibi verileri tutacağı bir hafızaya, bu verileri doğru işleyerek çıkarımlar yapacak bir zekaya ve nihayetinde yaptığı çıkarımları yorumlayarak kendini geliştirecek bir mantığa ihtiyacı vardır. Bu durumda verilerin tutulduğu veritabanlarını hafızaya, onları işleyip veriden veri çıkaracak zekayı veri madenciliği algoritmalarına, kendini geliştirecek mantığı ise eniyileştirmeye benzetebiliriz.

Geçmişte çok küçük veri kümeleri ile yapılan analizler gelişen veri madenciliği algoritmaları ve eniyileştirme yöntemleri ile çok büyük veri gruplarına uygulanabilir hale gelmiştir. Nihayetinde algoritmalar sayesinde verilerden çıkarılan bilgi, veriye daha hızlı ulaşmamızı ve veritabanının performansını yükseltmesini sağlayabilmektedir.

Bu çalışmada dinamik veritabanı üzerinde veri madenciliği algoritmaları araştırılmış ve çıkarılan kurallarla sorgu eniyilemesi yapacak bir uygulama geliştirilmiştir. Nihayetinde dinamik veritabanı üzerinde kendinden öğrenen bir sistem yapısının kurulması araştırılmış ve bu sistemin yararı ispatlanmıştır.

**Anahtar Kelimeler:** Veri madenciliği, Veritabanı, SQL, ADO, Shekhar, Cart, Statistica, sorgu eniyileştirmesi, karar ağacı.

## ABSTRACT

In this century, known as Information Age, data and usability of data become an important topic that should be manage. This topic finds new areas to develop itself every day. Therefore using of automatic systems that minimize human mistakes are growing and work load over human decreases.

If we match working of both system Computer and Human, computer needs to have a memory to keep the data, a processor to process the data and a logic to gain experience of usability of data. So we can say databases acting like memory, working of data mining algorithms acting like processor and finally optimizing technics like logic.

Analysis which use to be made via very little datasets may be done on huge datawarehouses due to improved data mining algorithms. Through these algorithms we can reach the data quicker than we use to do and that ensure improvement of the performance of databases.

In this thesis, data mining algorithms on dynamic databases are researched and the application is developed having semantic query optimisation uses the learned rules. In final, the self-learning system structure on a dynamic database is done and the profit of this system is proved.

**Keywords:** Datamining, Database, SQL, ADO, Shekhar, Cart, Statistica, query optimization, desicion trees.

## 1. GİRİŞ

Veri her geçen gün yeniden eklenen bilgi kümeleri ile sürekli çoğalan bir yapıdır. Bu sebeple Veritabanları içerisinde çok fazla miktarda veri barındıracak duruma gelmişlerdir.

Bu verilerin işlenmesi gün geçtikçe ulaşılabilirlik ve ulaşım hızı arttıkça daha kolay hale gelmiştir. Ancak gene de veri sayısının giderek artması verileri çok büyük oranda saklayan sistemlerin giderek işlem yapamaz ve sorguları çok yavaşlayan, hantal sistemler haline dönüşmesine yol açmaktadır.

Bu çalışmada veri madenciliğinin ne olduğu ve bazı metodları incelenmiş, metodlar kullanılarak eldeki veriler yorumlanmıştır. Bu yorumlama sırasında kullanılan hesaplama yöntemi ve algoritması, sorguların sonuçlarıyla beraber adım adım gösterilmiştir. Verilerin elde edilmesi esnasında kullanılan sorgular için uygulanabilecek anlamsal öğrenme metodlarından bazıları incelenmiş içlerinden biri üzerine bir uygulama geliştirilmiştir.

Öncelikle veri madenciliği yöntemlerinden ve anlamsal öğrenme metodlarından bahsedilecektir. Bir sonraki bölümde ise bu metodlardan ikisi kullanılarak geliştirilen uygulama anlatılmaya çalışılacaktır. Bu uygulamada çıkarılan bilgiler ile veritabanının kriterlerine uygun yeni kural kümeleri üretilmektedir. Üretilen bu yeni verilerin zaman-performans açısından karlılığı ayrıca hesaplanmaktadır. Son olarak bu kuralların iyileştirmeye göre performansları karşılaştırılmaktadır.

Bu çalışma SQL Server veritabanı kullanılarak, Visual Basic .Net programlama ortamında, ADO.Net veri iletişim protokolü kullanılarak, anlamsal öğrenme metodları yardımıyla geliştirilmiştir.

### 1.1 SQL Veri Tabanı Genel Özellikleri

Bir veritabanı yönetim sistemi (DBMS: DataBase Management System), kütüphaneler uygulamalar ve yardımcı programların birleşmesinden oluşur ve verilerin saklanması ve yönetilmesi ile ilgili konulardaki ayrıntılardan veritabanı yöneticilerini kurtarır. Aynı zamanda, kayıtların güncellenmesi ve kayıtlar üzerinde araştırma yapılması da mümkündür. Veritabanı yönetim sistemleri zamanla çok daha fazla özellikler kazanmaktadırlar. Bu özellikler maddeler halinde aşağıda verilmiştir:

Belirli bir tarzda organize edilmiş bilgi “koleksiyon”udur.

Büyük miktarda veri tutabilme ve verilerin depolandığı alanları ayarlayabilme imkanı sağlar.

Aynı anda çok sayıda kullanıcıya birden verilerin bütünlüğünü bozmadan hizmet verebilmektedir.

7 / 24 , haftalar boyu kapatılmadan çalışabilmektedir.

İşletim sistemi, veri erişim dilleri ve ağ iletişim protokolleri standartlarıyla uyumludur.

Yetkilendirme sayesinde erişimleri engelleme ve kontrol edebilme imkanı sağlamaktadır.

Bütünlüğü veritabanı düzeyinde sağlayabilmektedir, böylece daha az kod yazılmaktadır.

İstemci/Sunucu mimarisinin bütün avantajlarını kullanabilmektedir.

İlişkisel veritabanı yönetim sistemi (RDBMS: Relational DataBase Management System) bir programlama dili değildir. Bunun için Yapılandırılmış sorgulama dili (SQL: Structured Query Language) tarafından geliştirilen ve SQL'in kendi uygulama geliştirme araçları içerisinde kullanılan bazı programlama dilleri vardır. SQL ürünleri genellikle büyük çaplı veri kontrolünü gerektiren uygulamalarda kullanılır.

SQL veritabanının, işletim sistemi tarafından bakıldığında, biri fiziksel diğeri mantıksal olmak üzere iki bölümü vardır. Fiziksel bölüm, işletim sisteminden görünen kısımdır. Bunlar Veri(Data) dosyası, Kontrol(Control) Dosyası ve Kütük(Log) dosyalarından oluşmaktadır. Mantıksal Bölüm, bir ya da daha fazla tablo alanı(Tablespace) ve tablolar(table), görüntüler(view), sıralar(sequence), eşanlamlar(synonym), dizinler(index), kümeler(cluster), veritabanı bağlantıları (database link), prosedürler(procedure), fonksiyonlar(function), ve paketlerden(package) oluşan şema nesnelerinden oluşmaktadır. Fiziksel bölüm işletim sistemi tarafından görülebilmesine rağmen, mantıksal bölüm ancak SQL'a bağlanıp, SQL komutları çalıştırılarak görülebilmektedir. Yani, SQL kurulu herhangi bir makinede, SQL bilgisi olmayan bir insan, SQL'in sadece fiziksel bölümünü görebilmektedir.

SQL, Amerikan ulusal standartlar enstitüsü (ANSI: American National Standards Institute) ve Uluslararası standartlar teşkilatı (ISO: International Organization of Standardization) tarafından bir standart olarak tanımlanmış olsada, bu standartlar üzerinden yola çıkarak birçok değişiklikler meydana getirilmiş ve çeşitlenmeler oluşturulmuştur. Bunlar Oracle Şirketi'nin yapmış olduğu PL/SQL veya Sybase, IBM firmasının geliştirdiği SQL PL (SQL Procedural Language) ve Microsoft şirketinin geliştirdiği Transact-SQL dilleridir. Ayrıca şirketler ticari ürünlerinde SQL standartlarının temel özelliklerini nadir olarak desteklemekte, onun yerine kendi oluşturmuş olduğu türleri tercih etmektedirler.

Şirketlerin standartların dışında kendi türlerini geliştiriyor olmaları, dilin taşınabilirliğini zorlaştırmaktadır. Kısacası bir ortamdan başka bir ortama küçük değişikliklerle birlikte kolayca taşıyabileceğiniz ANSI C ve ya ANSI Fortran'ın aksine, SQL diliyle yazılmış kodların, büyük yapısal değişiklikler yapmadan bir ortamdan başka bir ortama aktarılması hayli zordur. Bunun nedenlerini ise şöyle açıklayabiliriz : SQL standartlarının karmaşıklığı, büyüklüğü ve bu standartların bir çok veritabanı tarafından uygulanmıyor olması ve standartların bir çok alanda, veritabanının nasıl davranması gerektiğini belirleyememiş olması. SQL standartlarının oluşmasıyla birlikte veritabanı uygulaması için kullanılacak sözdizimi kesin bir şekilde belirlenmiştir. Standartlar dilin yapısını anlam bilimi açısından iyi tanımlamamış ve bir çok alanda anlam karmaşasının oluşmasına neden olmuştur. SQL özel bir amaç (İlişkisel bir veritabanında bulunan bilgilerin sorgulanmasını sağlamak) için üretilmiştir. Bu yönüyle SQL, geniş kapsamlı problemleri çözmek için tasarlanmış olan C veya BASIC gibi programlama dilleri gibi imperative bir dil olmaktan çok, bir set-based, declarative programlama dilidir. SQL'in tam anlamıyla bir programlama diline dönüşebilmesini sağlamak amacıyla, ilk yöntem olarak PL/SQL gibi dil uzantıları, SQL'in getirmiş olduğu avantajlar gözönünde bulundurularak tekrardan tasarlanmıştır. İkinci yöntem ise, kodun programın içerisinde gömülü olması ve veritabanı ile etkileşimli çalışmasını sağlamak olmuştur. Örneğin, Oracle ve diğer diller veritabanlarında Java'yı barındırırken, PostgreSQL dili, Perl, Tcl ve C dilini de kapsayan büyük çaptaki bir dil grubu kullanılarak, kendi içerisinde fonksiyon yazılmasına izin vermektedir.

SQL veritabanındaki her nesnenin bir sahibi vardır. Her kullanıcı bir veya daha fazla tablo uzayına sahip olabilir. Her nesne, ait olduğu kullanıcının herhangi bir nesne uzayında (mantıksal olarak) bulunur. Her nesne uzayı da, kendisine sahip olan kullanıcının nesnelerini tutmak için işletim sisteminde bir veya daha fazla veri dosyasına sahiptir.

Sonuç itibariyle, veritabanındaki her nesnenin bir kullanıcısı vardır ve bu nesneler mantıksal olarak o kullanıcının sahip olduğu nesne uzaylarının herhangi birinin içerisinde, fiziksel olarak da o kullanıcının sahip olduğu nesne herhangi bir veri dosyasında bulunur. Fakat, o veri dosyasının içerisine işletim sistemi üzerinden bulunamaz. Bu nesnenin sahibi ve mantıksal yeri sadece Veri güncelleme dili (DML: Data Manipulation Language) komutları ile bulunabilmektedir.

## 1.2 Programlama Dillerinde Veritabanı Bağlantıları

Programlama dillerinden veritabanlarına erişim için iki ayrı yol bulunmamaktadır. Bunlar Gömülü SQL üzerinden (Embedded SQL) ve Aktif Veri Objesi üzerinden (ADO - Activex Data Objects) kullanılmaktadır.

Gömülü SQL ; DML, Veri tanımlama dili (DDL: Data Definition Language) ve veritabanı hareketi kontrol cümlelerinin prosedürel programlama dillerinin içinde gömülü olarak kullanılmasıdır.

ADO.Net ; Eski teknoloji ADO ‘nun geliştirilmiş versiyonudur. En önemli özelliği eskisi gibi her zaman açık bir bağlantıya ihtiyaç duymamasıdır. Bir defa bağlantıyı kurup veriyi aldıktan sonra bağlantıyı koparmakta ve böylece çok hızlı hareket imkanına olanak vermektedir. Ayrıca .Net mimarisi ile SQL Server desteği eklenen ADO.Net SQL server için farklılaştırılmış ve ayrıca hız kazandırılmıştır.

## 1.3 Visual Basic.Net Programlama Dili

Visual Basic .NET, Visual Basic’in kapsamlı yeniden tasarım işleminden geçirilmiş halidir. Visual Basic .NET’i daha sağlam bir yapılandırılmış programlama dili haline getiren pek çok yeni özellik eklenmiştir. En büyük değişiklik, Visual Basic .NET’in artık yönetilen bir dil olmasıdır. Visual Basic .NET artık kendi yerel derleyicisine sahip değildir, ama “common language runtime” ögesine derlenmektedir. Visual Basic .NET’in, “runtime” ile çalışabilmesi ve yeni programlama modelini kullanabilmesi için kapsamlı değişiklikler yapılmıştır. Visual Basic .NET, daha yüksek nesne yönelimi düzeyine sahiptir ve önceki sürümlere oranla daha yüksek “type safety” düzeyi sunmaktadır. Bu değişiklikler sayesinde, Visual Basic 6.0 projelerini .NET Framework sistemine taşıırken kodları yeniden yazmak yerine, basit bir bağlantı noktası yeterli olacaktır. Yeni özelliklerden bazıları aşağıda sunulmaktadır:

- Common Language Runtime Desteđi ; Visual Basic .NET common language runtime yaklaşımını temel almaktadır; COM - tabanlı bir Visual Basic .NET yoktur. Visual Basic 6.0 uygulamalarınızı Visual Basic.NET ortamına taşımak için küçük bir güncelleme çalışması yapılabilir ve runtime kazançlarından yararlanabilirsiniz.
- Nesne Yönelimli Tasarım ; Dil açısından en çok istenen özellik eski bileşenlerin kullanılabilmesidir. .NET sayesinde, artık Visual Basic programcıları eski uygulamaları kullanabilecek ve yeni Inherits anahtar sözcüğü ile mevcut sınıfları alabilecek ve Overrides ile taban sınıfı işlevini silecektir. Eski öğeleri kullanma özelliđi, yönetilen bir dille oluşturulan tüm sınıflarda çalışacaktır.
- Visual Basic .NET'e eklenen bir başka özellik de işlev aşırı yüklemesidir. Uygulama geliştiriciler artık aynı isme sahip olan, ama farklı argüman türleri ve dönüş türleri içeren işlevleri oluşturabileceklerdir.
- Visual Basic .NET içindeki oluşturucuları kullanan uygulama geliştiriciler, bir sınıfın yeni örneklerini oluştururken, argümanları sınıfa eşzamanlı olarak iletebilecektir.
- Boş Düğüm ; Serbest düğümleme sayesinde, uygulama geliştiriciler zaman-uyumsuz (asynchronous) yürütme özelliđini kullanarak, daha ölçeklenebilir ve daha güvenli uygulamalar geliştirecek. Yürütülmesi uzun süren veya harici kaynaklara gerek duyan işlev çağrıları, artık işlem için ikincil düğüm oluşturarak, uygulamanın geri kalan kısmının zamanuyumsuz olarak çalışmaya devam etmesini sağlayacaktır.
- Diğer Yeni Özellikler ; Visual Basic .NET içindeki kesin tür denetimi, Visual basic 6.0 içindeki gizli tür zorlamayı sıkılaştırmaktadır. Visual Basic 6.0'da, hemen hemen her tür başka bir türe gizli olarak çevrilebilmekte, ama tür sınırları aşıldığında çalışma zamanı hatası oluşturmaktadır. Visual Basic .NET, artık çalışma zamanında hataya neden olabilen tüm çevrimler için derleme zamanı hataları oluşturabilmektedir. Yapılandırılmış istisna işleme özelliđi, Visual Basic 6.0'daki On Error GoTo veya Resume Next hata işleme özelliđinin yerini almıştır. Yapılandırılmış istisna işleme sayesinde uygulama geliştiriciler, Try...Catch deyimlerini kullanarak normal koşullarda ve istisna koşullarında çalışan kodlar yazabilmektedir.
- Windows Forms, Visual Basic .NET'te uygulama geliştirme için yeni teknolojidir ve diller arasında uyumludur.

## 2. VERİ MADENCİLİĞİ VE VERİ MADENCİLİĞİ SÜRECİ

### 2.1 Veri Madenciliği

Veri madenciliği; önceden bilinmeyen, geçerli ve uygulanabilir bilginin veri yığınlarından dinamik bir süreç ile elde edilmesi olarak tanımlanabilir. Bu süreçte kümeleme, veri özetleme sınıflama kurallarının öğrenilmesi, bağımlılık ağlarının bulunması, değişkenlik analizi ve anomali tespiti gibi farklı birçok teknik kullanılmaktadır.

Veri madenciliği ile büyük veri yığınlarından oluşan database sistemleri içerisinde gizli kalmış bilgilerin çekilmesi sağlanır. Bu işlem, istatistik, matematik disiplinleri, modelleme teknikleri, database teknolojisi ve çeşitli bilgisayar programları kullanılarak yapılır.

Örneğin, Gartner Group [1] tarafından yapılan tanımda, veri madenciliği, istatistik ve matematik tekniklerle birlikte ilişki tanıma teknolojilerini kullanarak, depolama ortamlarında saklanmış bulunan veri yığınlarının elenmesi ile anlamlı yeni ilişki ve eğilimlerin keşfedilmesi sürecidir.

Veri madenciliği kendi başına bir çözüm değil çözüme ulaşmak için verilecek karar sürecini destekleyen, problemi çözmek için gerekli bilgileri sağlamaya yarayan bir araçtır. Veri madenciliği; analistin'e, iş yapma aşamasında oluşan veriler arasındaki şablonları ve ilişkileri bulması konusunda yardım etmektedir. Dilly, R. (1995)

Günümüzde veri madenciliğinin kullanılabildiği genel ilgi alanları olarak aşağıdakiler sayılabilir;

Pazarlama

- Müşteri segmentasyonu,
- Müşterilerin bazı özellikleri arasındaki bağlantıların kurulması,
- Pazarlama kampanyaları,
- Varolan müşterilerin elde tutulması için geliştirilecek pazarlama stratejilerinin oluşturulması,
- Pazar sepeti analizi,
- Çapraz satış analizleri,
- Müşteri değerlendirme,
- Müşteri ilişkileri yönetimi,
- Satış tahminlerinde,

Bankacılık

- Farklı finansal göstergeler arasındaki gizli korelasyonların bulunması,
- Kredi kartı dolandırıcılıklarının tespiti,
- Müşteri segmentasyonu,
- Kredi taleplerinin değerlendirilmesi,
- Usulsüzlük tespiti,
- Risk analizi,
- Risk yönetimi,

#### Sigortacılık

- Poliçe talep edecek müşterilerin tahmin edilmesi,
- Sigorta dolandırıcılıklarının tespiti,
- Müşteri risklerinin belirlenmesi.

#### Perakendecilik

- Satış noktası verilerinin analizleri,
- Alış-veriş sepetinin analizleri,
- Tedarik ve mağaza yerleşim eniyilemeleri,

#### Borsa

- Hisse senedi değer tahmini,
- Genel piyasa analizleri,
- Alım-satım stratejilerinin eniyilemesi.

#### Telekomünikasyon

- Kalite ve iyileştirme analizleri,
- Hisse tespitleri,
- Hatların yoğunluk tahminleri,

#### Sağlık ve İlaç

- Test sonuçları analizi,
- Ürün geliştirme,
- Tıbbi teşhis,
- Tedavi sürecinin belirlenmesi

Endüstri

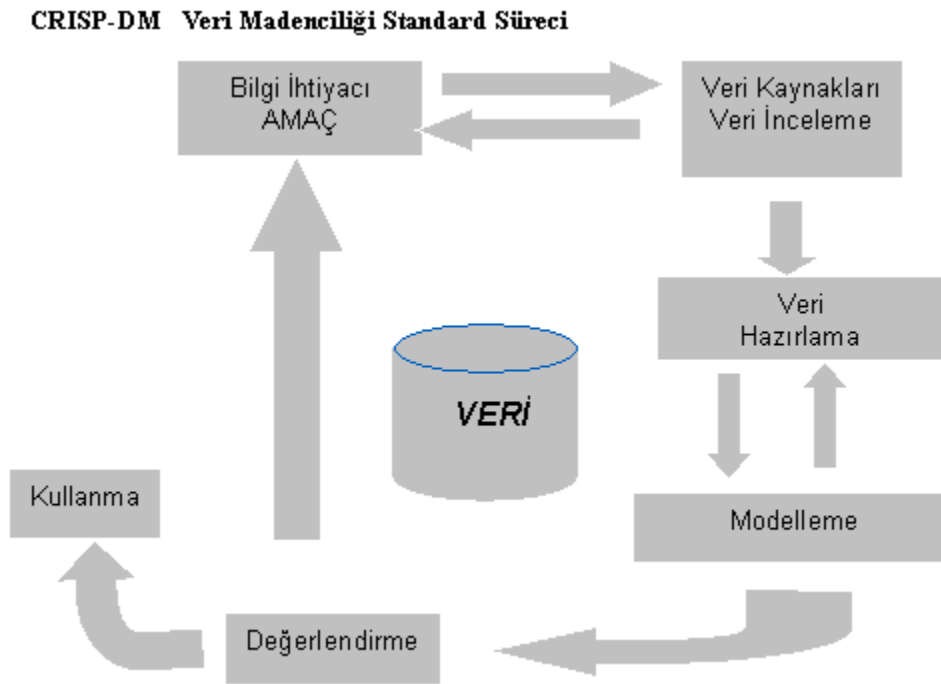
- Kalite kontrol analizleri,
- Lojistik,
- Üretim süreçlerinin eniyilemeleri,

Bilim ve Mühendislik

Bilimsel ve teknik problemlerin çözülmesi.

## 2.2 Veri Madenciliği Süreci

Veri Madenciliği Standard Süreci (CRISP-DM)



Şekil 2.1 Veri madenciliği standart süreci

Ne kadar etkin olursa olsun hiç bir veri madenciliği algoritmasının üzerinde inceleme yapılan işin ve verilerin özelliklerinin bilinmemesi durumunda fayda sağlaması mümkün değildir. Bu nedenle yukarıda tanımlanan tüm aşamalardan önce, iş ve veri özelliklerinin öğrenilmesi / anlaşılması başarının ilk şartı olacaktır. Başarılı bir Veri madenciliği projelerinde izlenmesi gereken yol ;

- I. Problemin Tanımlanması,
- II. Verilerin Hazırlanması,

- III. Modelin Kurulması ve Değerlendirilmesi,
- IV. Modelin Kullanılması,
- V. Modelin İzlenmesi

### 2.2.1 Problemin Tanımlanması

Veri madenciliği çalışmalarında başarılı olmanın en önemli şartı, projenin hangi işletme amacı için yapılacağını açık bir şekilde tanımlanmasıdır. İlgili işletme amacı işletme problemi üzerine odaklanmış ve açık bir dille ifade edilmiş olmalı, elde edilecek sonuçların başarı düzeylerinin nasıl ölçüleceği tanımlanmalıdır. Ayrıca yanlış tahminlerde katlanılacak olan maliyetlere ve doğru tahminlerde kazanılacak faydalara ilişkin tahminlere de bu aşamada yer verilmelidir.

Bu aşamada mevcut iş probleminin nasıl bir sonuç üretilmesi durumunda çözüleceğinin, üretilen sonucun fayda - maliyet analizinin başka bir değişle üretilen bilginin işletme için değerinin doğru analiz edilmesi gerekmektedir. Analistin işletmede üretilen sayısal verilerin boyutlarını, proje için yeterlilik düzeyinin iyi analiz edilmesi gerekmektedir. Ayrıca analistin işletme konusu hakkındaki iş süreçlerinin de iyi analiz edilmesi gerekmektedir.

### 2.2.2 Verilerin Hazırlanması

Veri madenciliğinin en önemli aşamalarından bir tanesi olan verinin hazırlanması aşaması analistin toplam zaman ve enerjisinin %50 - %75 ini harcamasına neden olmaktadır. Bu aşamada firmanın mevcut bilgi sistemleri üzerinde ürettiği sayısal bilginin iyi analiz edilmesi, veriler ile mevcut iş problemi arasında ilişki olması gerektiği unutulmamalıdır. Proje kapsamında kullanılacak sayısal verilerin, hangi iş süreçleri ile yaratıldığı da bu veriler kullanılmadan analiz edilmelidir, bu sayada analist veri kalitesi hakkında fikir sahibi olabilir.

Verilerin hazırlanması aşaması kendi içerisinde toplama, birleştirme ve temizleme, dönüştürme adımlarından meydana gelmektedir.

#### 2.2.2.1 Toplama

Tanımlanan problem için gerekli olduğu düşünülen verilerin ve bu verilerin toplanacağı veri kaynaklarının belirlenmesi adımdır. Verilerin toplanmasında kuruluşun kendi veri kaynaklarının dışında, nüfus sayımı, hava durumu, merkez bankası kara listesi gibi veri tabanlarından veya veri pazarlayan kuruluşların veri tabanlarından faydalanılabilir.

### 2.2.2.2 Birleştirme ve Temizleme

Bu adımda toplanan verilerde bulunan farklılıklar giderilmeye çalışılır. Hatalı veya analizin yanlış yönlendirilmesine sebep olabilecek verilerin temizlenmesine çalışılır. Genellikle yanlış veri girişinden veya bir kereye özgü bir olayın gerçekleşmesinden kaynaklanan verilerin, önemli bir uyarıcı enformasyon içerip içermediği kontrol edildikten sonra veri kümesinden atılması tercih edilir. Ancak basit yöntemlerle ve baştan savma olarak yapılacak sorun giderme işlemlerinin, ileriki aşamalarda daha büyük sorunların kaynağı olacağı unutulmamalıdır.

### 2.2.2.3 Dönüştürme

Kullanılacak model ve algoritma çerçevesinde verilerin tanımlama veya gösterim şeklinde değiştirilmesi gerekebilir.

## 2.2.3 Modelin Kurulması ve Değerlendirilmesi

Tanımlanan problem için en uygun modelin bulunabilmesi, olabildiğince çok sayıda modelin kurularak denenmesi ile mümkündür. Bu nedenle veri hazırlama ve model kurma aşamaları, en iyi olduğu düşünülen modele varılıncaya kadar yinelenen bir süreçtir.

Bir modelin doğruluğunun test edilmesinde kullanılan en basit yöntem basit geçerlilik (Simple Validation) testidir. Bu yöntemde tipik olarak verilerin % 5 ile % 33 arasındaki bir kısmı test verileri olarak ayrılır ve kalan kısım üzerinde modelin öğrenimi gerçekleştirildikten sonra, bu veriler üzerinde test işlemi yapılır. Bir sınıflama modelinde yanlış olarak sınıflanan olay sayısının, tüm olay sayısına bölünmesi ile hata oranı, doğru olarak sınıflanan olay sayısının tüm olay sayısına bölünmesi ile ise doğruluk oranı hesaplanır. (Doğruluk Oranı = 1 - Hata Oranı)

Önemli diğer bir değerlendirme kriteri modelin anlaşılabilirliğidir. Bazı uygulamalarda doğruluk oranlarındaki küçük artışlar çok önemli olsa da, bir çok işletme uygulamasında ilgili kararın niçin verildiğinin yorumlanabilmesi çok daha büyük önem taşıyabilir. Çok ender olarak yorumlanamayacak kadar karmaşıklaşırsalar da, genel olarak karar ağacı ve kural temelli sistemler model tahmininin altında yatan nedenleri çok iyi ortaya koyabilmektedir.

### 2.2.4 Modelin Kullanılması

Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak kullanılabilir. Kurulan modeller risk analizi, kredi değerlendirme, dolandırıcılık tespiti gibi işletme uygulamalarında doğrudan kullanılabilmesi gibi, promosyon planlaması simülasyonuna entegre edilebilir veya tahmin edilen envanter

düzeyleri yeniden sipariş noktasının altına düştüğünde, otomatik olarak sipariş verilmesini sağlayacak bir uygulamanın içine gömülebilir.

### 2.2.5 Modelin İzlenmesi

Zaman içerisinde bütün sistemlerin özelliklerinde ve dolayısıyla ürettikleri verilerde ortaya çıkan değişiklikler, kurulan modellerin sürekli olarak izlenmesini ve gerekiyorsa yeniden düzenlenmesini gerektirecektir. Tahmin edilen ve gözlenen değişkenler arasındaki farklılığı gösteren grafikler model sonuçlarının izlenmesinde kullanılan yararlı bir yöntemdir.

## 3. VERİ MADENCİLİĞİ YÖNTEMLERİ

Günümüzde birçok veri madenciliği yöntemi bilinmektedir. Bunlardan başlıcaları;

- Karar Ağaçları
- Yapay Sinir Ağları
- Zaman Serileri
- Bayes Sınıflandırma
- K-En Yakın Komşu
- Karar Destek Makineleri
- Eğri Uydurma
- Diğer Yöntemler

olarak sınıflandırılabilir. Bu yöntemler arasından en çok kullanılanı Karar Ağaçları'dır.

Biz bu noktada piyasadaki veri madenciliği uygulamalarının karar ağaçları yöntemi ile entropi algoritması kullanarak elimizdeki verilerden çıkardıkları sonuçları inceleyeceğiz.

### 3.1 Karar Ağaçları

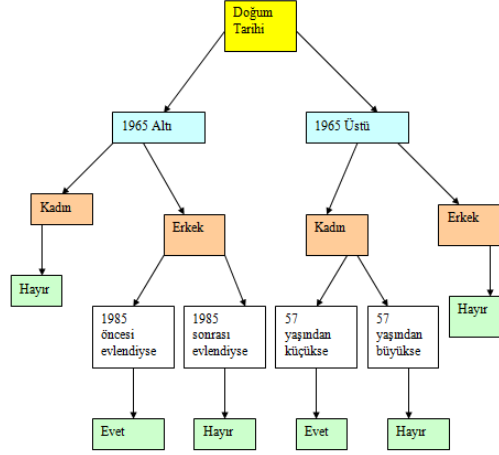
Karar ağacı oluşturma yöntemleri genel olarak iki aşamadan oluşur:

#### 3.1.1 Ağaç Oluşturma

En başta bütün öğrenme kümesi örnekleri kökte seçilen niteliklere bağlı olarak örnek yinelemeli olarak bölünüyor

### 3.1.2 Ağaç Budama

öğrenme kümesindeki gürültülü verilerden oluşan ve sınıma kümesinde hataya neden olan dalları silme (sınıflandırma başarımını artırır)



Şekil 3.1 Örnek karar ağacı

### 3.2 Karar Ağacı Oluşturma

- Yinelemeli işlem
- Ağaç bütün verinin oluşturduğu tek bir düğümle başlıyor
- Eğer örnekleri hepsi aynı sınıfa aitse düğüm yaprak olarak sonlanıyor ve sınıf etiketini alıyor
- Eğer değilse örnekleri sınıflara en iyi bölecek olan nitelik seçiliyor
- İşlem sona eriyor
- Örneklerin hepsi (çoğunluğu) aynı sınıfa ait
- Örnekleri bölecek nitelik kalmamış
- Kalan niteliklerin değerini taşıyan örnek yok

### 3.3 En İyi Bölen Nitelik Nasıl Belirlenir?

- İyilik Fonksiyonu (Goodness Function)
- Farklı algoritmalar farklı iyilik fonksiyonları kullanabilir:
- Bilgi kazancı (information gain): ID3, C4.5
- Bütün niteliklerin ayrıık değerler aldığı varsayılıyor

- Sürekli değişkenlere uygulamak için değişiklik yapılabilir
- Gini index (IBM IntelligentMiner)
- Bütün niteliklerin sürekli değerler aldığı varsayılıyor
- Her nitelik için farklı bölme değerleri olduğu varsayılıyor
- Bölme değerlerini belirlemek için başka yöntemlere (demetleme gibi) ihtiyaç var
- Ayrık değişkenlere uygulamak için değişiklik yapılabilir

### 3.4 Bilgi Kazancı (ID3 / C4.5)

Karar ağacı kullanan C4.5 ve ID3 algoritmaları için bilgi kazancı aşağıdaki gibi hesaplanır.

- Bilgi kuramı kavramlarını kullanarak karar ağacı oluşturulur. Sınıflandırma sonucu için en az sayıda karşılaştırma yapmayı hedefler.
- Ağaç bir niteliğe göre dallandığında entropi ne kadar düşer?
- A niteliğinin S veri kümesindeki bilgi kazancı  $Values(A)$ , A niteliğinin alabileceği değerler,  $S_v$ ,  $A=v$  olduğu durumda S'nin altkümesi.

**Entropi:** Sistemlerdeki düzensizlik arttıkça, entropi de artar. Bu durum da faydalı (iş yapılabilir) enerji miktarını azaltır. Faydasız enerjiyi (entropi) artırır. Eğer bir sistem tamamı ile düzenli ise entropisi sıfır olabilir. Entropi, enerji gibi korunan bir özellik değildir. Bütün enerji değişimlerinde çevre ile sistemin entropi değişimlerinin toplamı daima pozitifdir. Bu da evrendeki toplam entropinin sürekli artmasına sebep olur. Çoğunlukla bir sistemdeki rastgelelik ve düzensizlik olarak tanımlanır ve istatistikten teolojiye birçok alanda yararlanır. [5]

$$H(p_1, p_2, \dots, p_s) = - \sum_{i=1}^s p_i \log p_i$$

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

### 3.5 Örnek (Entropi) Algoritma Hesaplaması

Aşağıda “çocuk sayısı 2 den fazla” olan kayıtlar için örnek bir veri topluluğu sorgusu bulunmaktadır. Bu örnek topluluk entropi algoritmasıyla karar ağacı oluşturacaktır.

Çizelge 3.1 Entropi Algoritması için örnek tablo

| customer_id | adı      | soyadı       | cinsiyet | egitim_durumu | medeni_durumu | evlilik_yili | dogum_yili | cocuk_sayisi |       |
|-------------|----------|--------------|----------|---------------|---------------|--------------|------------|--------------|-------|
| 131127      | AYŞE     | ÖZDAL        | k        | Üniversite    | bekar         | 0            | 1980       | 0            | Hayır |
| 131128      | Mete     | Ellialtioglu | e        | Lisansüstü    | bekar         | 0            | 1970       | 0            | Hayır |
| 131129      | ÖNDER    | ELBEYLİOĞLU  | e        | Lise          | evli          | 2004         | 1979       | 0            | Hayır |
| 131130      | KAAN     | COŞKUN       | e        | Lise          | bekar         | 0            | 1984       | 0            | Hayır |
| 131155      | SEDAT    | GÖKPINAR     | e        | Üniversite    | evli          | 2006         | 1977       | 0            | Hayır |
| 131156      | EMİNE    | ÇUHA         | k        | İlkokul       | evli          | 1989         | 1972       | 3            | Evet  |
| 134089      | EMİNE    | ŞİMŞEK       | k        | Üniversite    | evli          | 1989         | 1966       | 1            | Hayır |
| 134090      | Gülseren | Güven        | k        | Lise          | evli          | 1977         | 1956       | 2            | Hayır |
| 134091      | hulusi   | canbay       | e        | Lise          | evli          | 1994         | 1968       | 1            | Hayır |
| 134092      | CEMAL    | ÖZKAN        | e        | İlkokul       | evli          | 1987         | 1963       | 4            | Evet  |
| 157219      | tahir    | bal          | e        | Ortaokul      | evli          | 1980         | 1958       | 3            | Evet  |
| 157220      | AYNUR    | YILMAZ       | k        | Önlisans      | evli          | 1997         | 1972       | 1            | Hayır |
| 157221      | HASAN    | SEYMEN       | e        | Üniversite    | evli          | 1996         | 1966       | 1            | Hayır |
| 157222      | HÜSEYİN  | ŞAHİN        | e        | Ortaokul      | evli          | 1977         | 1958       | 3            | Evet  |
| 157223      | YAKUP    | ALTINKAYNAK  | e        | Üniversite    | evli          | 1989         | 1967       | 2            | Hayır |
| 157284      | AHMET    | MİRZAOĞLU    | e        | Üniversite    | evli          | 1977         | 1954       | 2            | Hayır |
| 157285      | Orhan    | Yıldırım     | e        | Önlisans      | evli          | 1989         | 1962       | 2            | Hayır |
| 180987      | ERKAN    | AYDIN        | e        | Üniversite    | evli          | 2004         | 1974       | 0            | Hayır |
| 131424      | Bahri    | Emekli       | e        | İlkokul       | evli          | 1980         | 1962       | 5            | Evet  |
| 182291      | meral    | ayan         | k        | Üniversite    | evli          | 1987         | 1956       | 1            | Hayır |

Çizelge 3.2 Tablonun toplam entropisi

|                |    |   |
|----------------|----|---|
| Fazla =        | 5  | A |
| Az =           | 15 | B |
| toplam örnek = | 20 | T |

Tablo sorgusuna göre toplam 20 kaydın 5 tanesi iki çocuktan fazlasına sahipken, kalan 15 kaydın çocuk sayısı ikiden azdır.

Çizelge 3.3 Tablonun cinsiyete göre dağılımı

|            |    |       |
|------------|----|-------|
| Cinsiyet   | 6  | K     |
|            | 14 | E     |
| Kadınların | 1  | Evet  |
|            | 5  | Hayır |

|            |    |       |
|------------|----|-------|
| Erkeklerin | 4  | Evet  |
|            | 10 | Hayır |

Çizelge 3.4 Tablonun genel entropisi

|       | a/t  | log2(a/t) | b/t  | log2(b/t)   | a/t *<br>log2(a/t) | b/t *<br>log2(b/t) | Entropi [-<br>a/t]log2[a/t]-<br>[b/t]log2[b/t] |
|-------|------|-----------|------|-------------|--------------------|--------------------|------------------------------------------------|
| Genel | 0,25 | -2        | 0,75 | 0,415037499 | -0,5               | 0,311278124        | 0,811                                          |

Bu hesaplamada ilk değer (5) a olarak ikinci değer (15) b ve toplam (20) t olarak alındığında, aranan sonucun entropisine Genel Entropi E(T) denirse, hesabı  $[-a/t]\log_2[a/t]-[b/t]\log_2[b/t]$  olarak formüle edilebilir. Çizelge 3.4’de bu çarpım değerleri, her biri ayrıca hesaplanarak biraraya getirilmiştir. Genel Entropi bulunduğundan sonra ortaya çıkan tablodan tüm değerlere tek tek bakılarak kazanç (GAIN) hesaplanır ve bir sonraki bölünecek veri kümesi seçilir. Böleceğimiz veriye göre kazanç en yüksek olmalıdır.

Çizelge 3.5 Cinsiyet için Entropi Hesabı

|           | a/t   | log2(a/t) | b/t   | log2(b/t)    | a/t * log2(a/t) | b/t * log2(b/t) | Entropi [-a/t]log2[a/t]-<br>[b/t]log2[b/t] |
|-----------|-------|-----------|-------|--------------|-----------------|-----------------|--------------------------------------------|
| Cins(Er)  | 0,286 | -1,807355 | 0,714 | -0,485426827 | -0,516387121    | -0,346733448    | 0,863                                      |
| Cins(Kad) | 0,167 | -2,584963 | 0,833 | -0,263034406 | -0,430827083    | -0,219195338    | 0,650                                      |

**GAIN**

$$E(T)-x/t*e(x)-$$

| x/t | e(x)  | y/t | e(y)  | y/t*e(y) |
|-----|-------|-----|-------|----------|
| 0,7 | 0,863 | 0,3 | 0,650 | 0,012087 |

Çizelge 3.6 Medeni Durum için Entropi Hesabı

|              |       |           |       |              |              |              |       |
|--------------|-------|-----------|-------|--------------|--------------|--------------|-------|
| MedDur(Bek)  | 0     | 1         | 1     | 0            | 0            | 0            | 0,000 |
| MedDur(Evli) | 0,294 | -1,765535 | 0,706 | -0,502500341 | -0,519274925 | -0,354706123 | 0,874 |

| GAIN           |       |      |       |                    |
|----------------|-------|------|-------|--------------------|
| E(T)-x/t*e(x)- |       |      |       |                    |
| x/t            | e(x)  | y/t  | e(y)  | y/t*e(y)           |
| 0,15           | 0,000 | 0,85 | 0,874 | <b>0,068394234</b> |

Çizelge 3.7 Doğum tarihi (Yaş) için Entropi Hesabı

|           |       |           |       |              |              |              |       |
|-----------|-------|-----------|-------|--------------|--------------|--------------|-------|
| 1965 üstü | 0,083 | -3,584963 | 0,917 | -0,125530882 | -0,298746875 | -0,115069975 | 0,414 |
| 1965 altı | 0,5   | -1        | 0,5   | -1           | -0,5         | -0,5         | 1,000 |

| GAIN           |       |     |       |                    |
|----------------|-------|-----|-------|--------------------|
| E(T)-x/t*e(x)- |       |     |       |                    |
| x/t            | e(x)  | y/t | e(y)  | y/t*e(y)           |
| 0,6            | 0,414 | 0,4 | 1,000 | <b>0,162988014</b> |

Çizelge 3.8 Eğitim Durumu için Entropi Hesabı

hepsi tek değere bağlı olduğundan sonuç hepsinde 0 olacaktır.

Çizelge 3.9 Evlilik Yılı için Entropi Hesabı

|              |       |           |       |              |              |              |       |
|--------------|-------|-----------|-------|--------------|--------------|--------------|-------|
| 1985 öncesi  | 0,6   | -0,736966 | 0,4   | -1,321928095 | -0,442179356 | -0,528771238 | 0,971 |
| 1985 sonrası | 0,133 | -2,906891 | 0,867 | -0,206450877 | -0,387585413 | -0,178924094 | 0,567 |

| GAIN           |       |      |       |                    |
|----------------|-------|------|-------|--------------------|
| E(T)-x/t*e(x)- |       |      |       |                    |
| x/t            | e(x)  | y/t  | e(y)  | y/t*e(y)           |
| 0,25           | 0,971 | 0,75 | 0,567 | <b>0,143658346</b> |

Entropi hesapları sonunda çıkan sonuç;

Max Gain = Yaş

Ağaç YAŞ'a göre dallanacaktır. Daha sonra çıkan tablo için yeni bir entropi kazanç hesabı yapılacaktır. Bu hesaplama son bölünmeye kadar devam eder.

Çizelge 3.10 Birinci dal için Entropi tablosu

| customer_id | adı      | soyadı    | cinsiyet | egitim_durumu | medeni_durumu | evlilik_yili | dogum_yili | cocuk_sayisi |       |
|-------------|----------|-----------|----------|---------------|---------------|--------------|------------|--------------|-------|
| 134090      | Gülseren | Güven     | k        | Lise          | evli          | 1977         | 1956       | 2            | Hayır |
| 134092      | CEMAL    | ÖZKAN     | e        | İlkokul       | evli          | 1987         | 1963       | 4            | Evet  |
| 157219      | tahir    | bal       | e        | Ortaokul      | evli          | 1980         | 1958       | 3            | Evet  |
| 157222      | HÜSEYİN  | ŞAHİN     | e        | Ortaokul      | evli          | 1977         | 1958       | 3            | Evet  |
| 157284      | AHMET    | MİRZAOĞLU | e        | Üniversite    | evli          | 1977         | 1954       | 2            | Hayır |
| 157285      | Orhan    | Yıldırım  | e        | Önlisans      | evli          | 1989         | 1962       | 2            | Hayır |
| 131424      | Bahri    | Emekli    | e        | İlkokul       | evli          | 1980         | 1962       | 5            | Evet  |
| 182291      | meral    | ayan      | k        | Üniversite    | evli          | 1987         | 1956       | 1            | Hayır |

Çizelge 3.11 Evlilik tarihi için dağılım tablosu

|                     |   |             |
|---------------------|---|-------------|
| Evlilik Tarihi için | 5 | 1985öncesi  |
|                     | 3 | 1985sonrası |
| 1985 öncesi         | 3 | evet        |
|                     | 2 | hayır       |
| 1985 sonrası        | 1 | evet        |
|                     | 2 | hayır       |

Çizelge 3.12 Evlilik Yılı için Entropi Hesabı

|              |       |           |       |              |              |              |       |
|--------------|-------|-----------|-------|--------------|--------------|--------------|-------|
| 1985 öncesi  | 0,6   | -0,736966 | 0,4   | -1,321928095 | -0,442179356 | -0,528771238 | 0,971 |
| 1985 sonrası | 0,333 | -1,584963 | 0,667 | -0,584962501 | -0,528320834 | -0,389975    | 0,918 |

GAIN

 $E(T)-x/t \cdot e(x)-$ 

| x/t   | e(x)  | y/t   | e(y)  | y/t * e(y)  |
|-------|-------|-------|-------|-------------|
| 0,625 | 0,971 | 0,375 | 0,918 | 0,048794941 |

Çizelge 3.13 Cinsiyet için dağılım tablosu

|            |   |       |
|------------|---|-------|
| Cinsiyet   | 2 | kadın |
|            | 6 | erkek |
| kadınların | 0 | evet  |
|            | 2 | hayır |
| erkeklerin | 4 | evet  |
|            | 2 | hayır |

Çizelge 3.14 Evlilik Yılı için Entropi Hesabı

|       |       |           |       |              |           |              |       |
|-------|-------|-----------|-------|--------------|-----------|--------------|-------|
| kadın | 0     | 1         | 1     | 0            | 0         | 0            | 0,000 |
| erkek | 0,667 | -0,584963 | 0,333 | -1,584962501 | -0,389975 | -0,528320834 | 0,918 |

**GAIN**

$$E(T)-x/t * e(x) -$$

| x/t  | e(x)  | y/t  | e(y)  | y/t * e(y)         |
|------|-------|------|-------|--------------------|
| 0,25 | 0,000 | 0,75 | 0,918 | <b>0,311278124</b> |

**Max Gain = Cinsiyet**

Ağaç CİNSİYET'e göre dallanacaktır. Medeni durum ağacı tek değere sahip olduğundan bu aşamada budanmıştır.

1965 altı kadınlar için                      tek değer var entropi 0                      sonuç HAYIR

1965 altı erkekler için                      1985 üstü evliler HAYIR  
1985 altı evliler EVET

Daha sonra çıkan tabloda tekrar Entropi kazanç hesabı yaparsak.

Çizelge 3.15 İkinci dal için Entropi tablosu

| customer_id | adı    | soyadı       | cinsiyet | egitim_durumu | medeni_durumu | evlilik_yili | dogum_yili | cocuk_sayisi |       |
|-------------|--------|--------------|----------|---------------|---------------|--------------|------------|--------------|-------|
| 131127      | AYŞE   | ÖZDAL        | k        | Üniversite    | bekar         | 0            | 1980       | 0            | Hayır |
| 131128      | Mete   | Ellialtioglu | e        | Lisansüstü    | bekar         | 0            | 1970       | 0            | Hayır |
| 131129      | ÖNDER  | ELBEYLİOĞLU  | e        | Lise          | evli          | 2004         | 1979       | 0            | Hayır |
| 131130      | KAAN   | COŞKUN       | e        | Lise          | bekar         | 0            | 1984       | 0            | Hayır |
| 131155      | SEDAT  | GÖKPINAR     | e        | Üniversite    | evli          | 2006         | 1977       | 0            | Hayır |
| 131156      | EMİNE  | ÇUHA         | k        | İlkokul       | evli          | 1989         | 1972       | 3            | Evet  |
| 134089      | EMİNE  | ŞİMŞEK       | k        | Üniversite    | evli          | 1989         | 1966       | 1            | Hayır |
| 134091      | hulusi | canbay       | e        | Lise          | evli          | 1994         | 1968       | 1            | Hayır |
| 157220      | AYNUR  | YILMAZ       | k        | Önlisans      | evli          | 1997         | 1972       | 1            | Hayır |
| 157221      | HASAN  | SEYMEN       | e        | Üniversite    | evli          | 1996         | 1966       | 1            | Hayır |
| 157223      | YAKUP  | ALTINKAYNAK  | e        | Üniversite    | evli          | 1989         | 1967       | 2            | Hayır |
| 180987      | ERKAN  | AYDIN        | e        | Üniversite    | evli          | 2004         | 1974       | 0            | Hayır |

Çizelge 3.16 Cinsiyet için dağılım tablosu

|            |   |       |
|------------|---|-------|
| Cinsiyet   | 4 | Kadın |
|            | 8 | Erkek |
| kadınların | 1 | Evet  |
|            | 3 | Hayır |
| erkeklerin | 0 | Evet  |
|            | 8 | Hayır |

Çizelge 3.17 Cinsiyet için Entropi Hesabı

|       |      |    |      |              |      |              |       |
|-------|------|----|------|--------------|------|--------------|-------|
| kadın | 0,25 | -2 | 0,75 | -0,415037499 | -0,5 | -0,311278124 | 0,811 |
| erkek | 0    | 1  | 1    | 0            | 0    | 0            | 0,000 |

**GAIN** **$E(T)-x/t \cdot e(x)-$** 

| <b>x/t</b> | <b>e(x)</b> | <b>y/t</b> | <b>e(y)</b> | <b>y/t * e(y)</b>  |
|------------|-------------|------------|-------------|--------------------|
| 0,333333   | 0,811       | 0,666667   | 0,000       | <b>0,143573959</b> |

Çizelge 3.18 Evlilik Tarihi için dağılım tablosu

|              |    |              |
|--------------|----|--------------|
|              | 0  | 1985 öncesi  |
|              | 12 | 1985 sonrası |
| 1985 öncesi  | 0  | evet         |
|              | 0  | hayır        |
| 1985 sonrası | 1  | evet         |
|              | 11 | hayır        |

Çizelge 3.19 Evlilik Tarihi için Entropi Hesabı

|              |       |           |       |              |              |              |       |
|--------------|-------|-----------|-------|--------------|--------------|--------------|-------|
| 1985 öncesi  | 0     | 1         | 0     | 1            | 0            | 0            | 0,000 |
| 1985 sonrası | 0,083 | -3,584963 | 0,917 | -0,125530882 | -0,298746875 | -0,115069975 | 0,414 |

| GAIN                   |        |       |        |                  |
|------------------------|--------|-------|--------|------------------|
| $E(T)-x/t \cdot e(x)-$ |        |       |        |                  |
| $x/t$                  | $e(x)$ | $y/t$ | $e(y)$ | $y/t \cdot e(y)$ |
| 0                      | 0,000  | 1     | 0,414  | 0,00018315       |

### Max Gain = Cinsiyet

Ağaç CİNSİYET'e göre dallanacaktır.

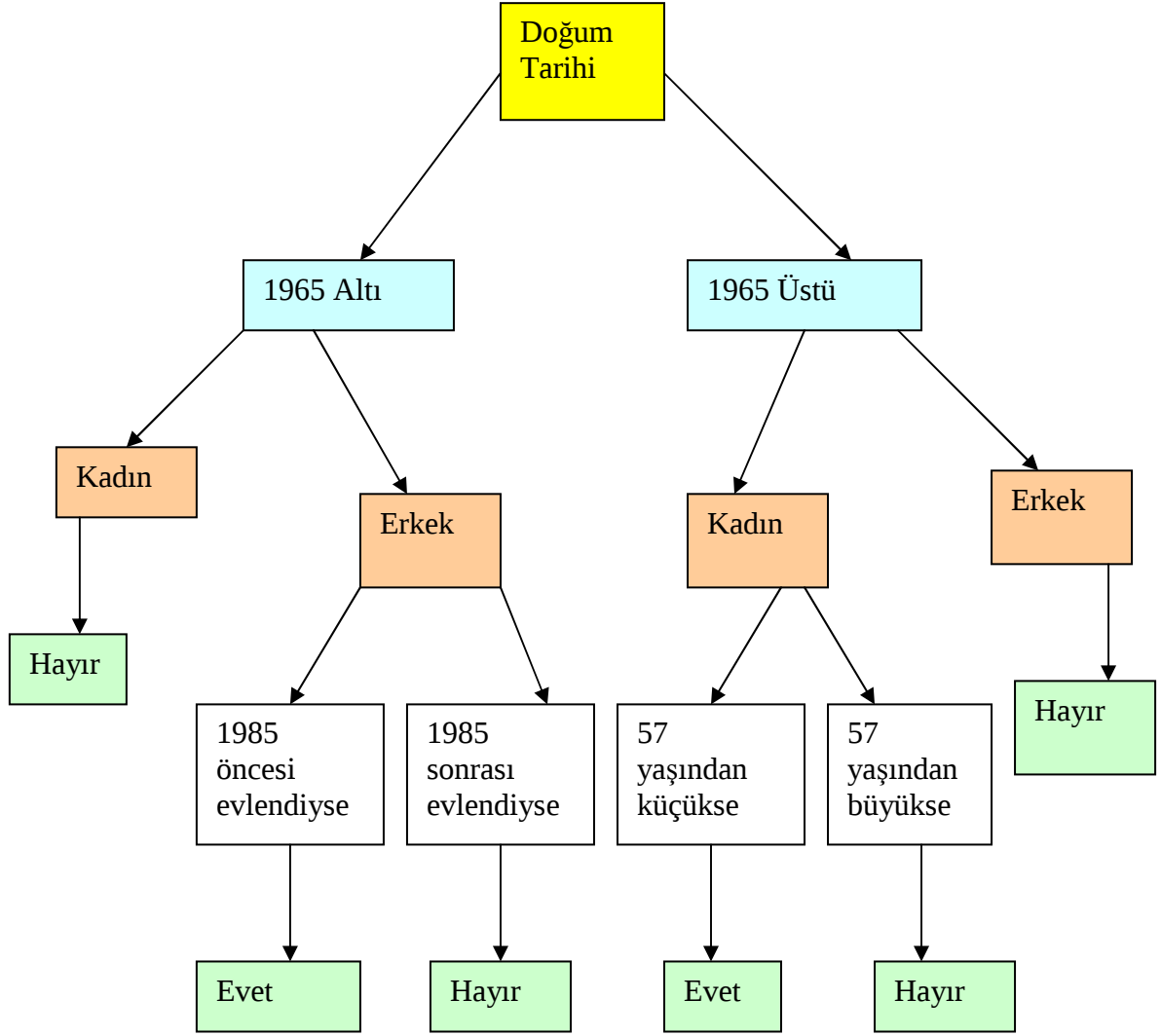
Buna göre Karar Ağacı şu şekilde oluşur.

Çizelge 3.20 Karar Ağacı Sonuç Tablosu

| Doğum Tarihi | Cinsiyet | Evlilik Tarihi | Yaş Kontr.  | Sonuç |
|--------------|----------|----------------|-------------|-------|
| 1965 altı    | Kadın    | -----          | -----       | Hayır |
| 1965 altı    | Erkek    | 1985 öncesi    | -----       | Evet  |
| 1965 altı    | Erkek    | 1985 sonrası   | -----       | Hayır |
| 1965 üstü    | Erkek    | -----          | -----       | Hayır |
| 1965 üstü    | Kadın    | -----          | 57 Yaş Üstü | Hayır |
| 1965 üstü    | Kadın    | -----          | 57 Yaş Altı | Evet  |

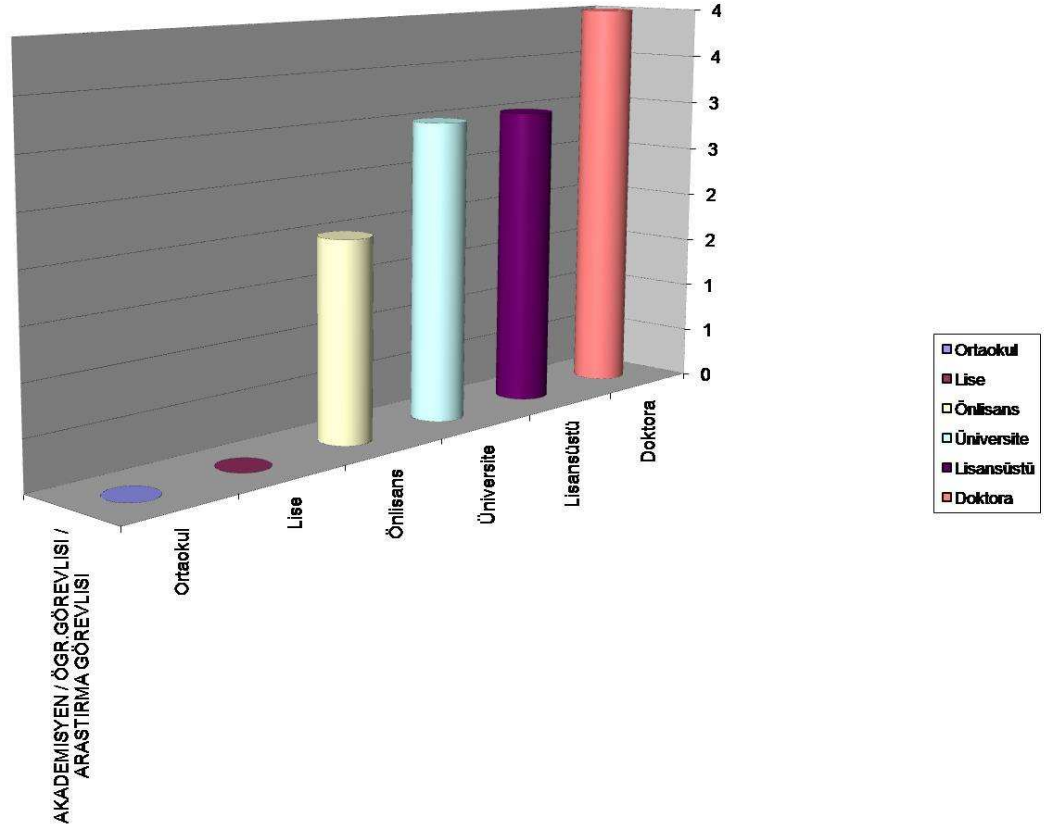
Değerleri ile karar ağacı oluşturulmuş olur.

Bu karar ağacı farklı sektörlerde elde edilmiş farklı amaçlar için kullanılabilir. Örneğin bu örnek tablosunu kullanan bir mobilya mağazası kişinin yaşı, cinsiyeti ve evlilik tarihi bilgilerinden yola çıkarak o anki satış stratejisini değiştirebilir. Bir başka firmanın insan kaynakları ise işe alım sürecinde bu bilgilerle kişinin yaklaşık sigorta primini hesaplayabilir.



Şekil 3.3 Oluşturulan karar ağacı

Aşağıda bu metodla elde edilmiş bir öğrenim seviyesine göre en fazla çocuk sayısı analizi sonucunu gösterilmektedir.



Şekil 3.4 Öğrenim seviyesine göre çocuk sayısı

Görüldüğü üzere yaklaşık 25000 kayıtlık bir veri içerisinde çekilen analize göre eğitim seviyesi arttıkça çocuk sayısında da artış söz konusudur.

#### 4. VERİ MADENCİLİĞİNİ KULLANAN SİSTEMLER

Bu bölümde bazı veri madenciliği yöntemlerini kullanan, piyasaya en hakim 3 uygulama incelenmiştir. Bu uygulamalarla varolan bir veri tabanı üzerinde veri madenciliği çalışmaları yapılmıştır. Bu programlar sırasıyla STATISTICA, DTREG ve CART dır.

##### 4.1 Statistica

Statistica birçok yöntem ile veri madenciliği yapan bir programdır. İş zekası , kalite kontrol ve araştırma & geliştirme alanlarında kullanılan, çok kapsamlı analizlere olanak tanır. Üniversitelerden kimya araştırma şirketlerine, banka kredi kartı üreticilerinden botanik üzerine polimer üreticilerine kadar birçok sektörde kullanılmaktadır. [2]

Statistica birçok algoritma ile çalışabilmektedir. Bu algoritmalar şunlardır;

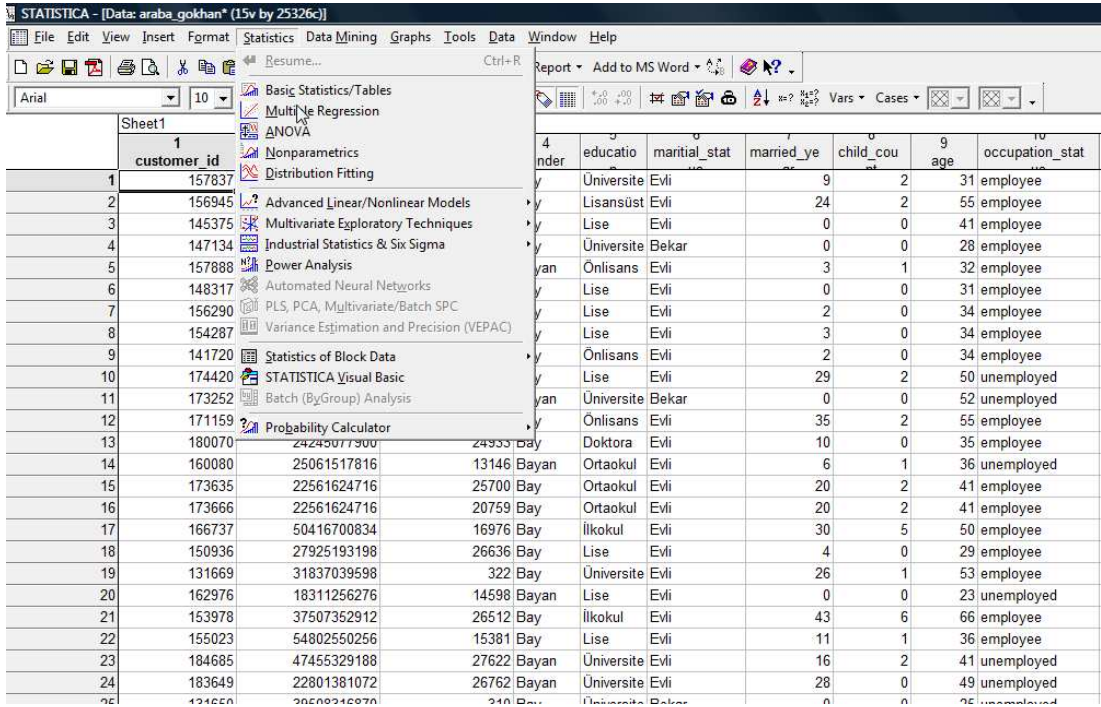
- Gradient Descent ,Conjugate Gradient Descent
- BFGS
- Kohonen training
- k-Means Center Assignment for Radial Basis networks

Varolan veritabanımızı Statistica ile tanıttınca program veri yöneticisi gibi davranarak verileri sıralamaktadır.

|    | 1           | 2           | 3                | 4      | 5          | 6            | 7          | 8         | 9   | 10              | 11   |
|----|-------------|-------------|------------------|--------|------------|--------------|------------|-----------|-----|-----------------|------|
|    | customer_id | memis_no    | customer_vehicle | gender | education  | marital_stat | married_ye | child_cou | age | occupation_stat | corp |
| 1  | 157837      | 13192763080 | 12178            | Bay    | Universite | Evli         |            | 9         | 2   | 31 employee     | EGİ  |
| 2  | 156945      | 12544889720 | 13431            | Bay    | Lisansüst  | Evli         |            | 24        | 2   | 55 employee     | MİM  |
| 3  | 145375      | 34444526338 | 6731             | Bay    | Lise       | Evli         |            | 0         | 0   | 41 employee     | OTO  |
| 4  | 147134      | 54130596766 | 22824            | Bay    | Universite | Bekar        |            | 0         | 0   | 28 employee     | DEN  |
| 5  | 157888      | 12389087008 | 12194            | Bayan  | Önlisans   | Evli         |            | 3         | 1   | 32 employee     | DEV  |
| 6  | 148317      | 13240017042 | 9095             | Bay    | Lise       | Evli         |            | 0         | 0   | 31 employee     | SAN  |
| 7  | 156290      | 21343893654 | 11389            | Bay    | Lise       | Evli         |            | 2         | 0   | 34 employee     | GİD/ |
| 8  | 154287      | 21343893654 | 26508            | Bay    | Lise       | Evli         |            | 3         | 0   | 34 employee     | GİD/ |
| 9  | 141720      | 21343893654 | 22213            | Bay    | Önlisans   | Evli         |            | 2         | 0   | 34 employee     | GİD/ |
| 10 | 174420      | 23023560078 | 21186            | Bayan  | Lise       | Evli         |            | 29        | 2   | 50 unemployed   | NUL  |
| 11 | 173252      | 16847349226 | 21942            | Bayan  | Universite | Bekar        |            | 0         | 0   | 52 unemployed   | NUL  |
| 12 | 171159      | 19777593042 | 19811            | Bay    | Önlisans   | Evli         |            | 35        | 2   | 55 employee     | SER  |
| 13 | 180070      | 24245077900 | 24933            | Bayan  | Doktora    | Evli         |            | 10        | 0   | 35 employee     | ASK  |
| 14 | 160080      | 25061517816 | 13148            | Bayan  | Ortaokul   | Evli         |            | 6         | 1   | 36 unemployed   | NUL  |
| 15 | 173635      | 22561624716 | 25700            | Bay    | Ortaokul   | Evli         |            | 20        | 2   | 41 employee     | ASK  |
| 16 | 173666      | 22561624716 | 20759            | Bay    | Ortaokul   | Evli         |            | 20        | 2   | 41 employee     | ASK  |
| 17 | 166737      | 50416700834 | 16976            | Bay    | İlkokul    | Evli         |            | 30        | 5   | 50 employee     | SER  |

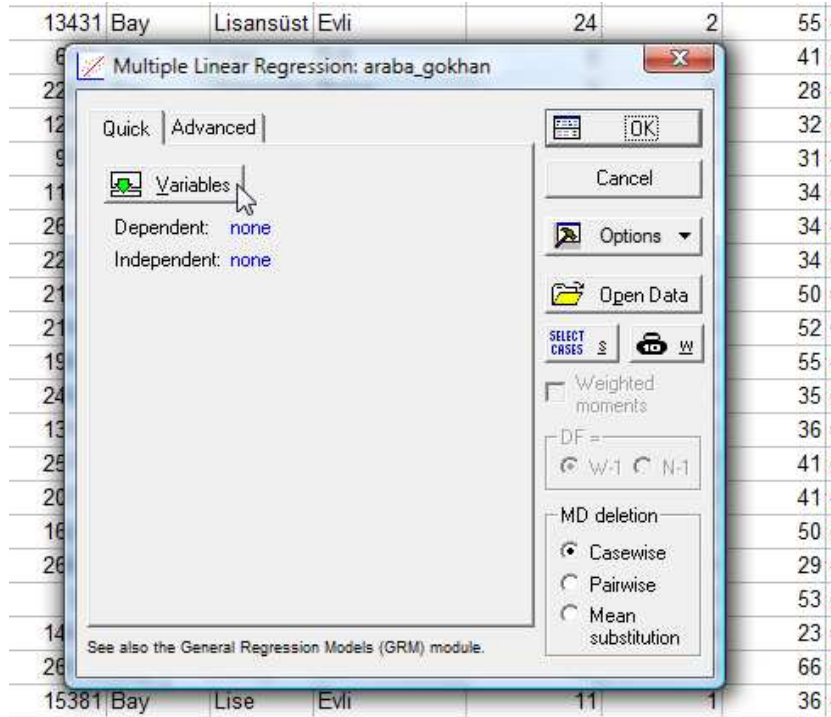
Şekil 4.1 Statistica veritabanı tanıma

Veritabanı programa tanıtılınca program kullanıcıdan veri madenciliği yöntemini ister,



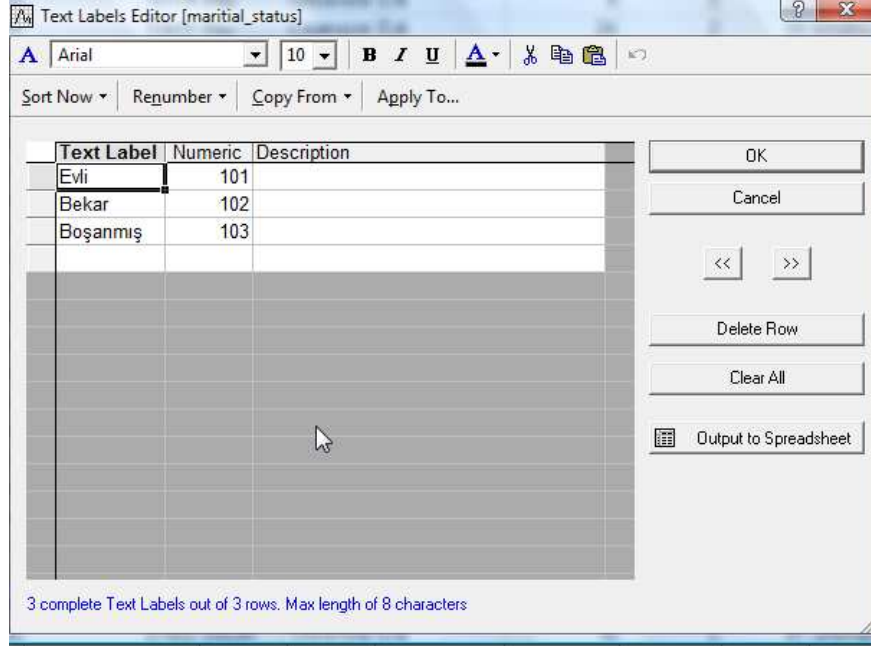
Şekil 4.2 Statistica veri madenciliği yöntemi seçimi

Daha sonra program kullanıcıdan bağımlı ve bağımsız değişkenleri seçmesini isteyecektir,



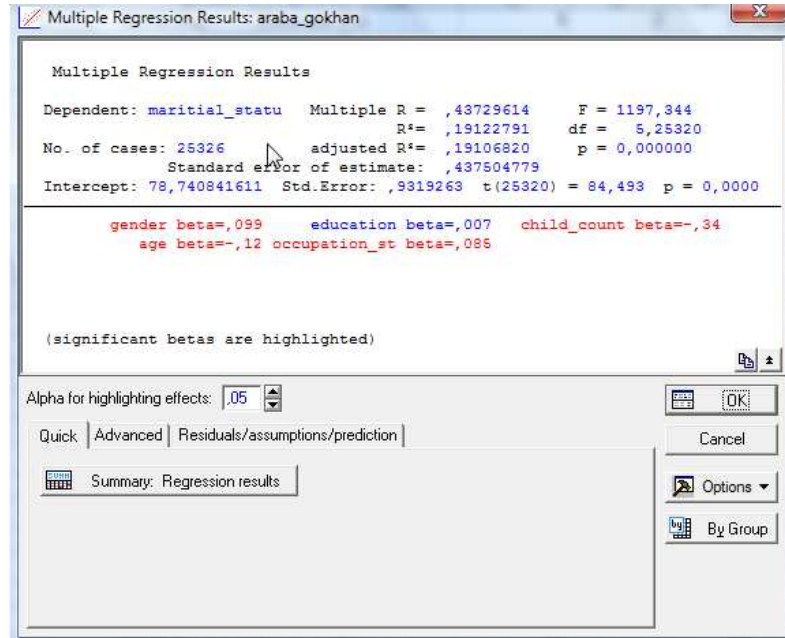
Şekil 4.3 Statistica bağımlı-bağımsız değişken seçimi

Program analizi yapabilmek için text formatındaki verileri nümerik değerlere eşitleyerek etiketler,



Şekil 4.4 Statistica değişken dönüşümü

Ve bu bilgiler ışığında regresyon analizi yaparak sonuçlarını sunar,



Şekil 4.5 Statistica sonuç

## 4.2 Dtreg

DTREG zaman serilerinden , basit ağaçlara birçok modele göre Veri Madenciliği yapabilen bir programdır. Hastanelerde ağrı nedenlerinin araştırmasından bankalarda çalışanlara verilebilecek kredi oranları hesaplamalarına, seçimlerde politik kampanyaların başarı trendlerine kadar birçok alanda kullanılmaktadır. [3]

DTREG programının kullandığı algoritma yelpazesi Statistica'ya göre daha geniştir;

- Decision Trees
- TreeBoost — Boosted Decision Trees
- Decision Tree Forests
- Multilayer Perceptron Neural Networks
- RBF Neural Networks
- Cascade Correlation Neural Networks
- Probabilistic Neural Networks
- General Regression Neural Networks
- Support Vector Machine (SVM)
- Gene Expression Programming - Symbolic Regression
- Linear Discriminant Analysis (LDA)
- Linear Regression , Logistic Regression

Veritabanı programa tanıtılınca program çalışabilmek için kullanıcıdan hedef (target) ve belirleyici (predictor) değerlerini ister,

| Variable                | Target                              | Predictor                           | Weight                   | Categorical                         | Character                           |
|-------------------------|-------------------------------------|-------------------------------------|--------------------------|-------------------------------------|-------------------------------------|
| customer_id             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            |
| memis_no                | <input type="checkbox"/>            | <input type="checkbox"/>            | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| customer_vehicle_id     | <input type="checkbox"/>            | <input type="checkbox"/>            | <input type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            |
| gender                  | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| education               | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| marital_status          | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| married_year            | <input type="checkbox"/>            | <input type="checkbox"/>            | <input type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            |
| child_count             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            |
| age                     | <input type="checkbox"/>            | <input type="checkbox"/>            | <input type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            |
| occupation_status       | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| corporation_sector      | <input type="checkbox"/>            | <input type="checkbox"/>            | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| corporation_position    | <input type="checkbox"/>            | <input type="checkbox"/>            | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| occupation              | <input type="checkbox"/>            | <input type="checkbox"/>            | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| countrymodel_definition | <input type="checkbox"/>            | <input type="checkbox"/>            | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| top_model_definition    | <input type="checkbox"/>            | <input type="checkbox"/>            | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |

Report options

☒ Report summary of variables

☐ Report category statistics for categorical variables

☐ Report category statistics for continuous variables

Custom cross-validation fold control

☐ Use a variable to control cross validation folds

customer\_id

OK Cancel Apply

Şekil 4.6 Dtreg hedef ve belirleyici seçimi

Daha sonra kullanılacak metod ve metodun özellikleri belirlenir,

Single tree model

Properties on this page control the generation of single tree models.

Type of model to build

Single tree

Tree size controls

Minimum size node to split: 10

Maximum tree levels: 10

☐ Generate report of tree splits

Method for validating and pruning the tree

☐ No validation, use full tree

☒ V-fold cross-validation trees: 10

☐ Random percent of rows: 20

☐ Fixed number of terminal nodes: 20

☒ Smooth minimum spikes: 3

Tree pruning control

☒ Prune to minimum cross-validated error

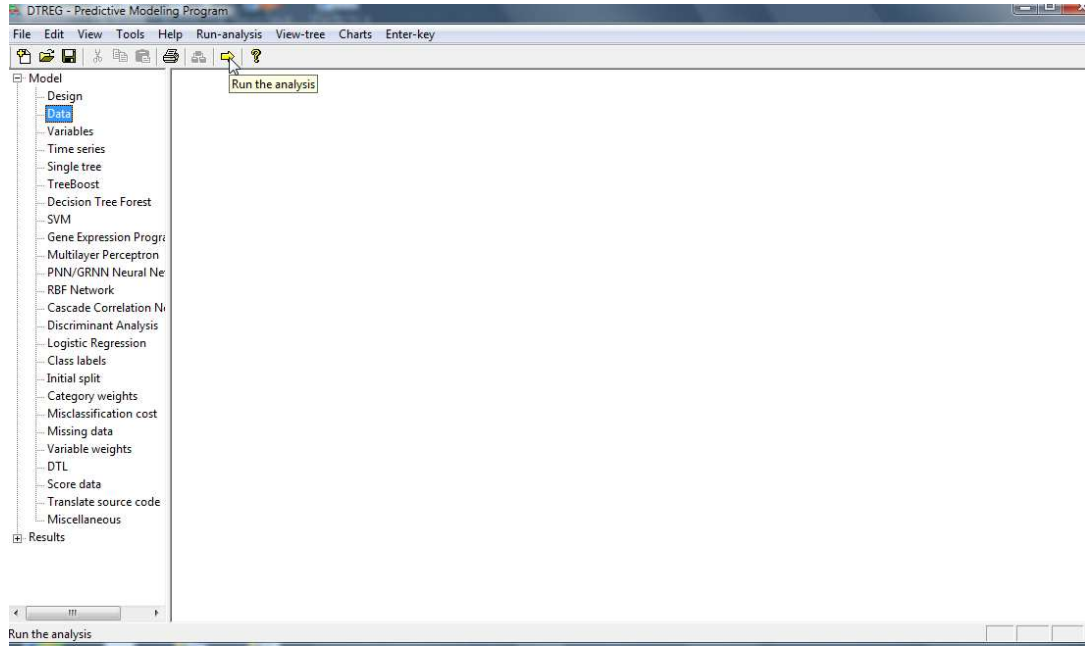
☐ Allow 1 standard error from minimum

☐ Allow this many S.E. from min.: 0.500

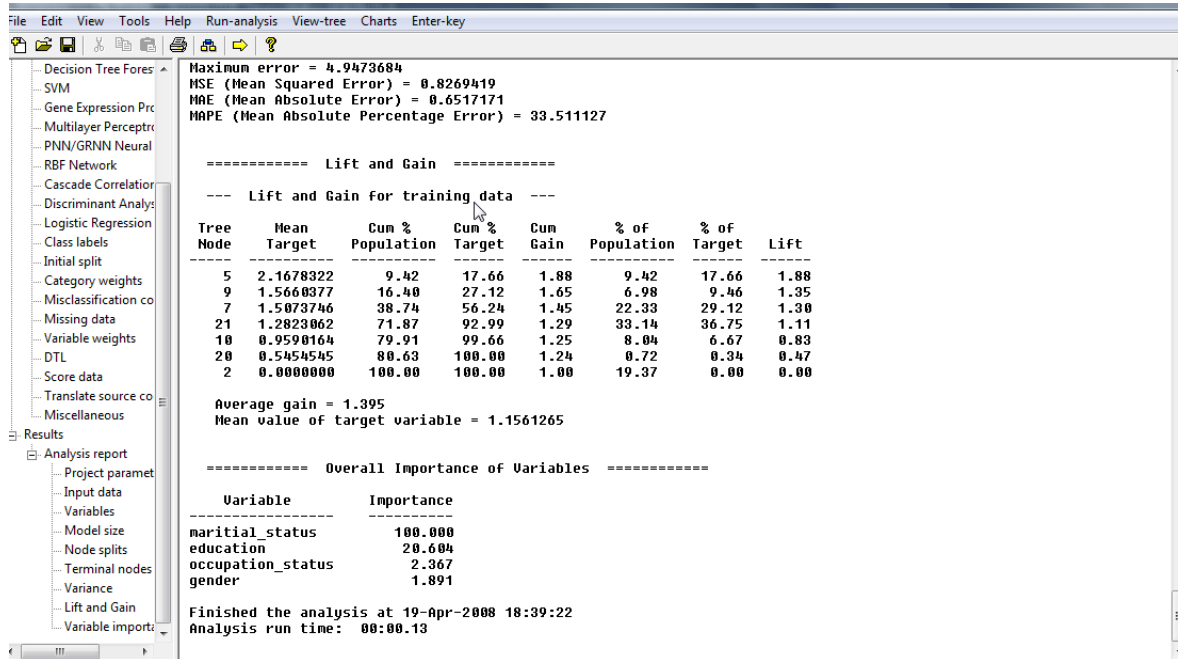
☐ Do not prune the tree

Şekil 4.7 Dtreg metod seçimi

Analizin başlatılmasıyla program seçilen yöntemler için analiz yapar,

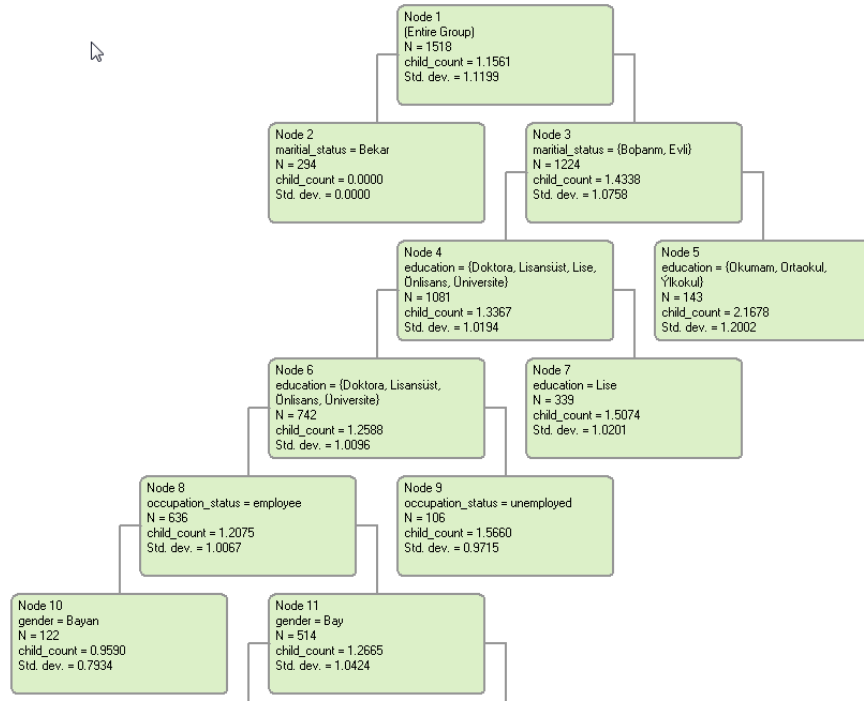


Şekil 4.8 Dtreg analiz başlatılması



Şekil 4.9 Dtreg analiz sonucu

Programın karar ağaçları yöntemini kullanarak ürettiği ağaç yapısı şu şekildedir;



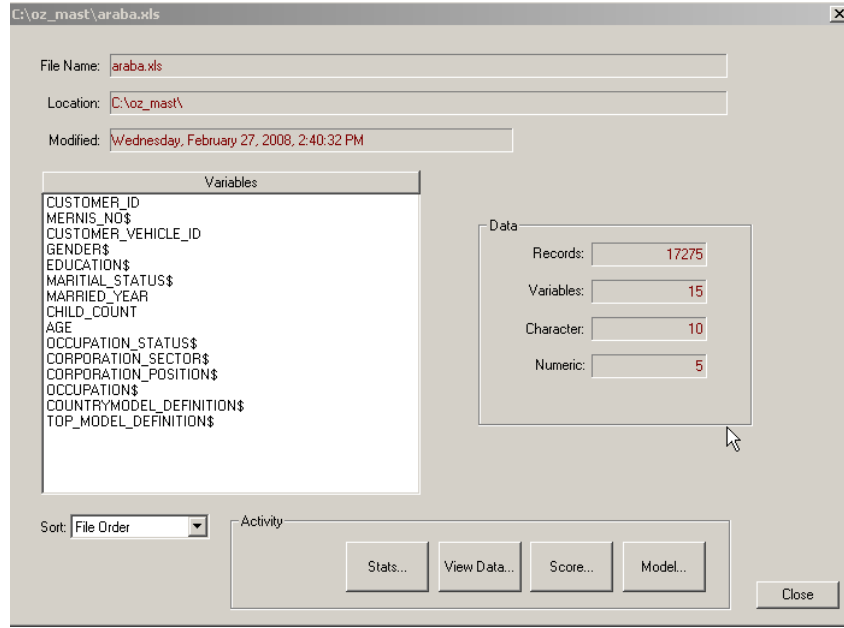
Şekil 4.10 Dtree karar ağacı

### 4.3 Cart

Cart en büyük özelliği olan kolay kullanımı ve kullanıcı dostu yapısı sayesinde en çok tercih edilen veri madenciliği uygulamalarından biridir. Dünyaca ünlü Silikon vadisi olarak bilinen bilişim bölgesinin kurucusu Kaliforniya’ daki Stanford Üniversitesi’ nde geliştirilmiş olan program, veri madenciliğinde karar ağaçlarını tercih eden insanlar için ihtiyaçlarına tam anlamıyla ve kolay yoldan cevap verebilmektedir. [4]

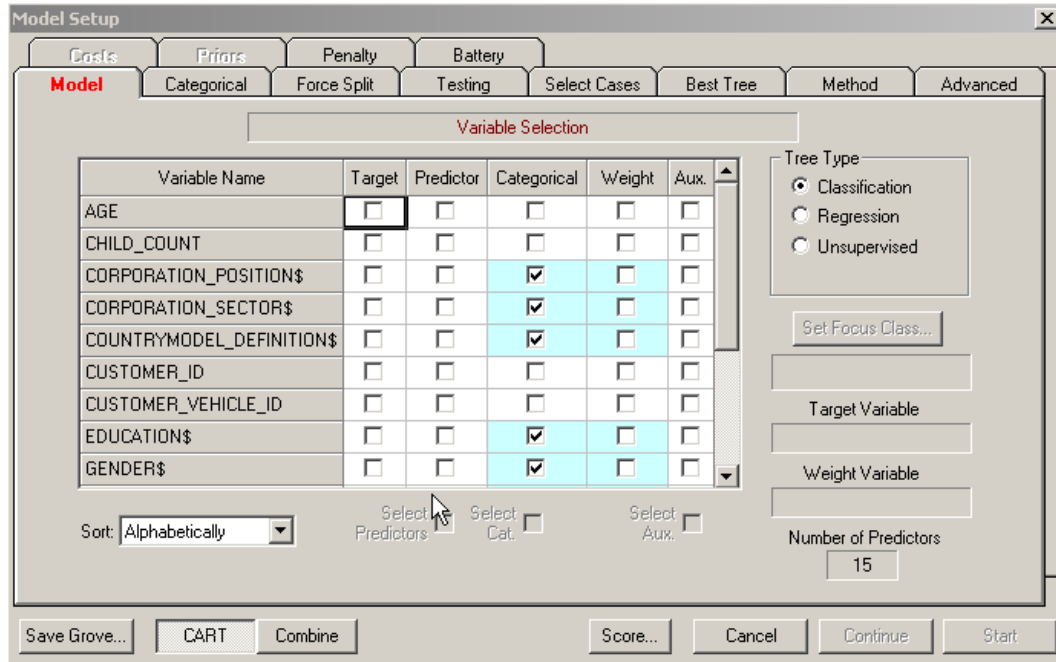
Cart, inşaat, otomotiv den ilaç sektörüne kadar birçok alanda kullanımdadır. Örneğin dünya ilaç devlerinden Pfizer firması veri madenciliği uygulamalarını CART ile gerçekleştirmektedir. Hatta hepimizin bildiği Viagra ilacının kullanım alışkanlıklarından dolayı oluşabilecek sorunları bu programla analiz etmiştir. [4]

Programa veri tabanı tanıtılınca program kullanılabilir veri alanlarını gösterir,



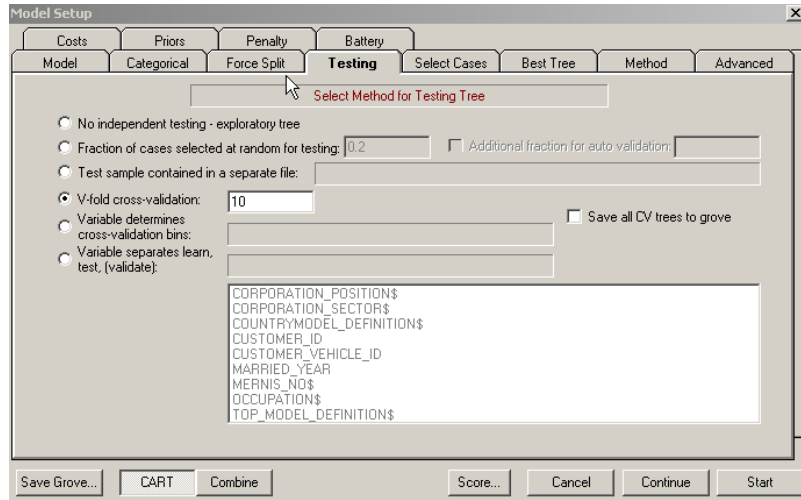
Şekil 4.11 Cart kullanılabilir veri alanları

Kategori olarak eklenecek alanlar seçildikten sonra hedef ve belirleyiciler seçilir,

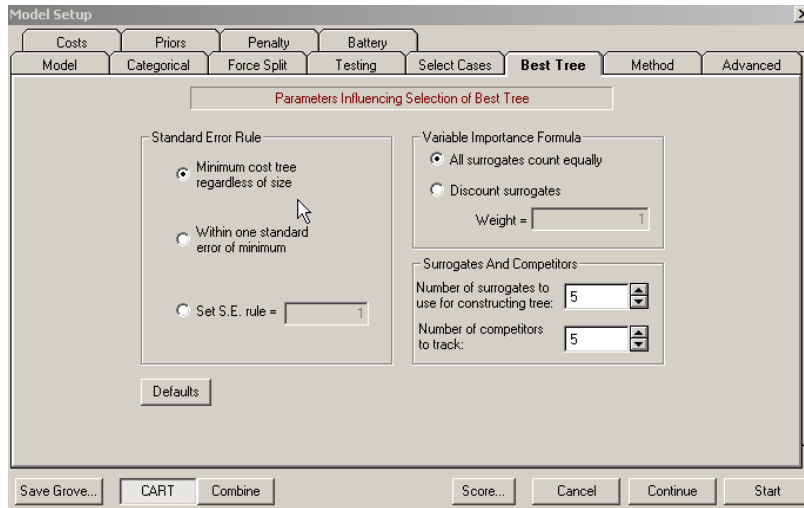


Şekil 4.12 Cart hedef ve belirleyici seçimi

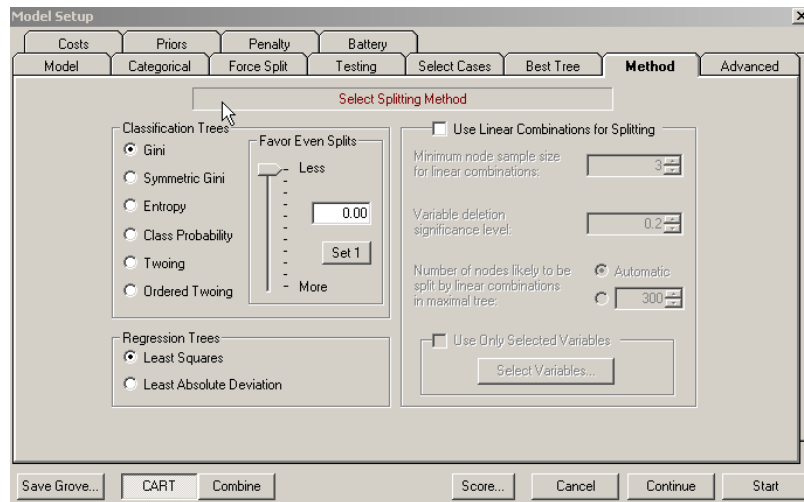
Test yöntemi, ağaç özellikleri ve sınıflama metodu gelen ekranlar yardımıyla belirlenir,



Şekil 4.13 Cart yöntemi seçimi

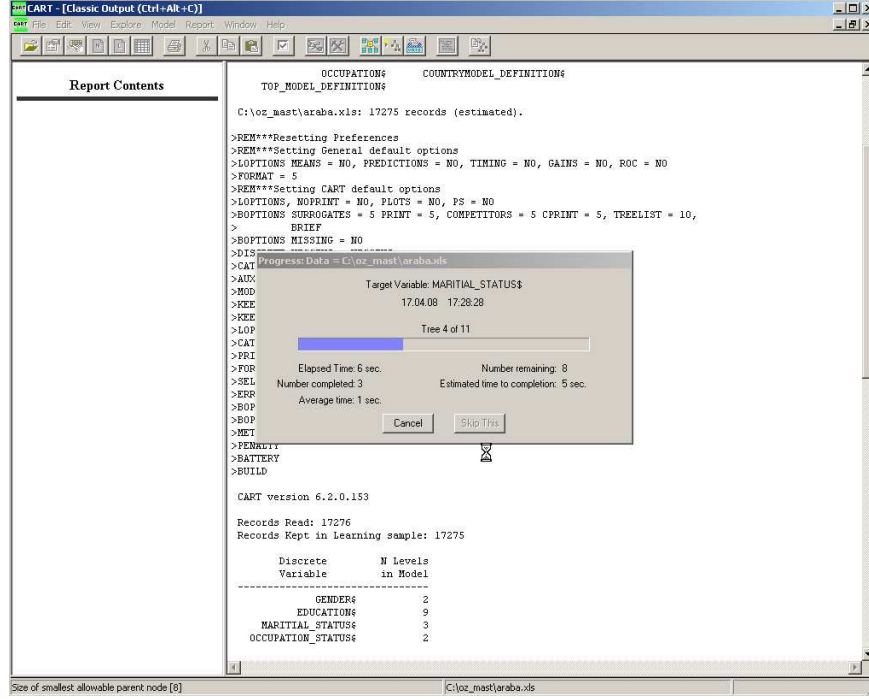


Şekil 4.14 Ağaç özelliği seçimi

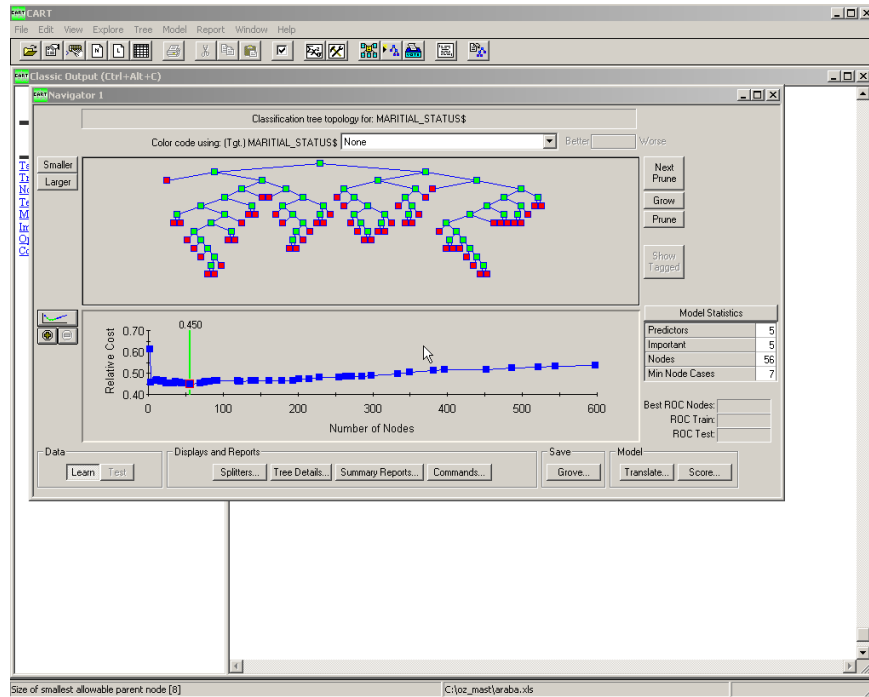


Şekil 4.15 Sınıflama metodu seçimi

Program analizi başlattığımızda 25000 kayıt için analizi 5-6 saniye aralığında tamamlamaktadır. Sonuçları ağaç formunda çıkartır.

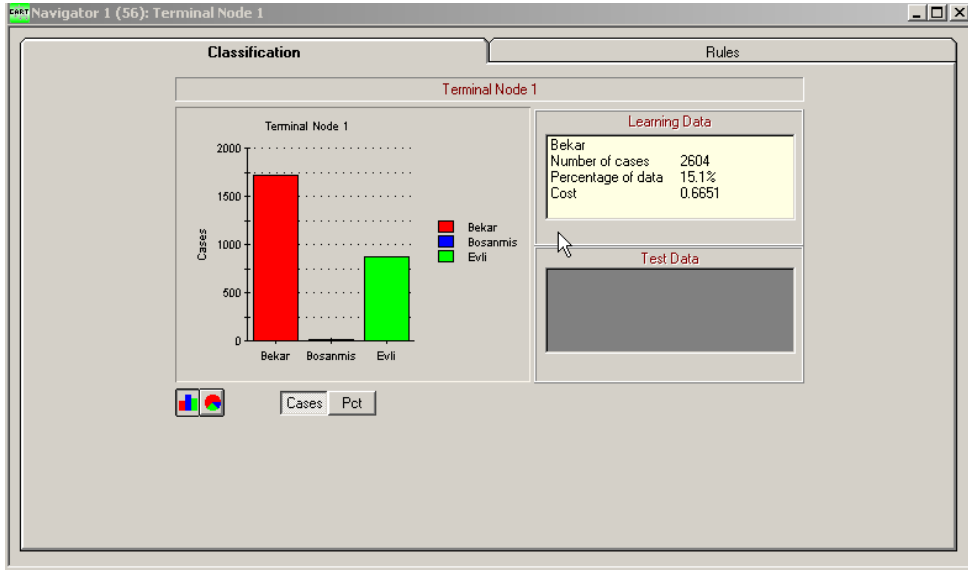


Şekil 4.16 Cart analiz sonucu

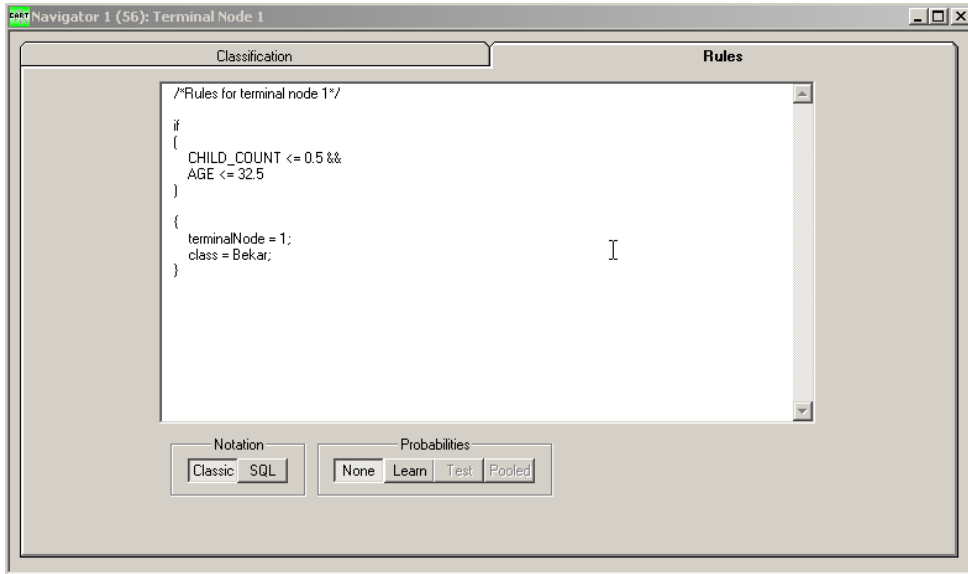


Şekil 4.17 Programın çıktısı olan karar ağacı

Ağacın yapraklarını incelediğimizde, her bir yaprak için detaylı analiz verilmektedir,



Şekil 4.18 Cart, karar ağacı yaprakları detay 1



Şekil 4.19 Cart, karar ağacı yaprakları detay 2

Görüldüğü üzere, farklı algoritmalar ve metodlar kullanarak çok hızlı ve kolay şekilde verileri analiz eden Veri Madenciliği Programları piyasada bir çok yerde ihtiyaçlara göre seçilerek kullanılabilmektedirler. Bu programlar çok farklı metod ve algoritma yapılarına rağmen ortak olarak yüksek ücret zemininde buluşurlar.

## 5. SQL – YAPISAL SORGU DİLİ

Sql "Structured Query Language" kelimelerinin baş harflerinin kısaltılmasıyla oluşan ve veri tabanı sistemleri üzerinde çeşitli sorgu işlevleri için kullanılan bir alt dildir. Bu çalışma içinde SQL Server içindeki verilere ulaşmak için ADO.Net veri erişim yöntemi kullanılmıştır. ADO.Net veri erişim yöntemiyle erişilen tablo verileri Visual Basic.Net programlama ortamında analiz edilmiş ve sonuçlar alınmıştır. Çıkan sonuçlara göre elde edilen verilerin analizini yapmak ve programlama ortamında kullanabilmek için SQL sorgulama dili ve SQL Server tablo yapısı bilinmelidir. Bu bölümün amacı tablo yapısını, SQL sorgulama dilini açıkça ifade etmek ve sorguların anlamlarını anlatmaktır.

### 5.1 SQL Server Tabloları

SQL Server birçok verinin saklı tutulduğu, depolandığı, birden fazla kullanıcıya aynı anda bilgiye erişim imkanının sağlandığı yapıdır. SQL Server'da veri depolamanın temel birimi tablolarıdır. SQL Server tabloları, veritabanında yaratılır ve devamlılıkları burada sağlanır. Bir SQL Server tablosu satır ve sütunlardan oluşur. Bir tablo içindeki satırlar verilerin eksiksiz bir kaydını temsil eder. (Shapiro 2001)

Tabloların Genel Özellikleri

- Bir tablo, SQL Server veritabanında bulunan bir depolama birimidir.
- Tablo bir veritabanı nesnesidir ve başka bir ortama taşınamaz. Ancak verileri bir sonuç kümesi içinde bir kopya olarak taşınabilir.
- Veriler tablodan, bir tablolu datagram olarak taşınabilirler.
- İstemci bileşenleri(Ado gibi) tablolu verileri protokol yığınının nasıl okuyacaklarını ve onları istemcide nasıl sunacaklarını bilirler.
- Değiştirilmiş olana veriler atılır. Değiştirilmesi gereken veriler sunucuya bir güncelleme olarak geri döndürülür.

#### 5.1.1 Tablo Oluşturulması (CREATE TABLE)

Veritabanı üzerinde tablo oluşturma ve tablo ile ilgili özelliklerinin tanımlanması için kullanılır.

CREATE TABLE Tablo\_Adi (

Alan\_Isimleri Veri\_Tipleri [Default ifade] [Alan Kısıtlaması]

[Tablo\_Kısıtlamalar ]

)

Alan Ismi : Oluşturulacak kolona verilecek isim .

Veri\_Tipleri : Oluşturduğumuz alanın veri tipi char, varchar, nvarchar, integer, decimal, double, datetime, text ,vb. Bazı veri tipleri veritabanından veritabanına değişmektedir. Eğer veri karakter ise char , varchar(karakter uzunluğu) vb, çok uzun karakterler için text , sayı ise int, küsuratlı sayı ise decimal yada double, tarih alanı için datetime gibi tipler kullanılabilir.

Alan Kısıtlaması : İlgili alanın boş geçilip geçilemeyeceği durumudur. NULL boş geçilebilir, NOT NULL boş geçilemez, ilgili alana kayıt girmek zorundalığı anlamındadır.

Tablo\_Kısıtlamaları : Tabloda yapılacak kısıtlamalardır. Primary Key, Foreign Key, Check lerdir.

Örnek : Departman isimli bir tablonun oluşturulması.

CREATE TABLE Departman

( DepartmanID int IDENTITY (1, 1) NOT NULL , //DepartmanId birden başlayıp birer birer otomatik olarak artacaktır.

DAdi nvarchar (20) DEFAULT('') , //DAdi alanına değer atanmaması durumunda NULL değeri yerine '' (boşluk) atanması

KurulusTarihi smalldatetime NULL , //smalldatetime t-sql özgu bir veri tipidir.Standart DateTime

CONSTRAINT PK\_Departman PRIMARY KEY CLUSTERED ( DepartmanID ) //İlgili primary key tanımı

)

Bir alanın otomatik olarak artması istiyorsa, ilgili alan tanımlanırken alan kısıtlamasına IDENTITY (Başlangıç değeri , Kaçar kaçır artacağı) ifadesinin eklenmesi gerekmektedir. Ayrıca, eğer istenirse, DEFAULT () ifadesi ile alana ilk değer verilir. Eğer ilgili alan boş geçilirse yani değer atanmazsa otomatik olarak ilk değeri atanacaktır.

**Sütun Kısıtlama (Column Constraint) :** Sütunlara bir kısıtlama koymak için kullanılır. Bu kısıtlama birincil anahtarları, sütunların tekiliğini, sütunlarda boş veri olup olmayacağı veya

başka tablo üzerindeki alanlarla verinin bütünlüğünü korumak amacıyla kullanılır. (Codd 1970)

- i. **Birincil Anahtar (Primary Key)** : PRIMARY KEY kısıtlaması bir tabloda bir birincil anahtar tanımlaması için kullanılır. Bir tabloda sadece bir tane birincil anahtar olabilir. Birincil anahtar (primary key) olarak tanımlanan o alanda girilen her değer birbirinden farklı olmak zorundadır. Null değeri tanımlanmasına izin verilmez ve otomatik olarak belirlenen o alanda bir içerik yaratılır.
- ii. **Yabancı Anahtar (Foreign Key)** : Birincil anahtar ya da UNIQUE olarak tanımlanmış (aynı tabloda ya da farklı tablolarda) iki alanın birbiri ile olan ilişkisini gösterir. Yabancı anahtar kısıtlaması yaratılan alanda diğer alanda belirtilen değerler dışında başka bir değer girilemez. Bu kısıtlama yaratıldığında otomatik olarak dizin oluşmaz.
- iii. **Tekil Anahtar (Unique Key)** : Bir tabloda bir alanda aynı değerden tekrar girilmesini önler. Null değeri tanımlanmasına izin verir. Bir tabloda birden çok alan için UNIQUE değeri tanımlanabildiği için bir tabloda birden çok UNIQUE olabilir. Bir alanda UNIQUE değeri tanımlanınca, otomatik olarak Unique Index (Özel içerik) oluşur.
- iv. **Boş Olmayan (Not Null)** : Tabloda belirtilen alanın boş olmayacağını verilmesidir. Bu alan içerisinde bilgi bulunması zorunludur. Yani bu alan hükümsüz olamaz.
- v. **Veri Sınırı Koymak (Check)** : Tablodaki bir alana verilecek değerlere bir sınır verme işlemidir. Bir tabloda bir alana girilen değerleri, diğer alanlardaki değerleri göz önüne alarak geçerliliğine bakar.

## 5.2 Sorgu

Sorgulama, veritabanlarında kullanılan bir veri işleme şeklidir. Sorgulamada SQL sorgulama dili kullanılır Birçok veri erişim erişim yöntemi SQL deyimlerini destekler.

Yaygın Kullanılan SQL deyimleri

SELECT Bir ya da birden fazla tablodan istenen verilerin alınmasını sağlar.

ALTER Mevcut bir tabloyu değiştirir

CREATE TABLE Yeni bir tablo yaratır

|         |                                             |
|---------|---------------------------------------------|
| DELETE  | Kayıtların silinmesini sağlar               |
| INSERT  | Mevcut bir tabloya kayıt eklenmesini sağlar |
| UPDATE  | Bir tablodaki kayıtları günceller           |
| EXECUTE | Bir saklı yordam ya da sorguyu çalıştırma   |

### 5.2.1 SELECT

SELECT deyimi SQL dilinin en yaygın deyimidir. Bu deyim veritabanından veri almayı, veritabanındaki veriler üzerinde sorgu oluşturmayı sağlar. Bu deyim bir veri işleme deyimidir.

#### **SELECT Deyiminin Yapısı:**

SELECT (seçim listesi )

INTO (yeni tablo)

FROM (tablo )

WHERE (ifade)

GROUP BY (ifade)

HAVING (ifade )

ORDER BY(ifade )

En basit SELECT deyimi aşağıdaki gibidir.

SELECT 'OZGUR'

Bu deyim herhangi bir tablo ya da veriye gerek duymaz. Bu deyim tek bir sonuç kümesi içerisinde tek değer olarak 'OZGUR' geri döndürür.

Sonuç kümesinde kaç veri satırının geri döndürülmesi gerektiğinin belirlenmesine yardımcı argümanlar mevcuttur. Bunlar aşağıda şöyle açıklanmıştır:

**ALL:** ALL(SELECT ALL) argümanı bütün satırların sonuç listesinde görülmesini sağlar. Bu argüman varsayılan olup ayrıca belirtilmeyebilir. Sorgu kriterini sağlayan bütün satırlar, tekrar edilenler ve özelliiksizler(nullar) dahil olmak üzere geri döndürülür.

Örnek:

Çizelge 5.1 TableEx tablosu ve veri bilgileri

| Numarası | Adı   | Soyadı  |
|----------|-------|---------|
| 1        | Özgür | Sarıhan |
| 2        | Arda  | Anıl    |
| 3        | Emrah | Çuvak   |
| 4        | Emrah | Kaçmaz  |

Select ALL(Adi)

from TableEx

Sorgu çalıştığında döndürülen sonuç çizelge 5.2’de gösterilmiştir.

Çizelge 5.2 Sütundaki tüm kayıtları getiren sorgu sonucu

| Adi   |
|-------|
| Özgür |
| Arda  |
| Emrah |
| Emrah |

**DISTINCT:** DISTINCT(SELECT DISTINCT) argümanı sadece tek olan kayıtların sonuç listesinde görülmesini sağlar.

Örnek: Çizelge 5.1 için aşağıdaki sorgu çalıştırıldığında sonuç seti çizelge 5.3’deki gibi olmaktadır.

Select DISTINCT(ADI)

From TableEx

Çizelge 5.3 Sütundaki tek kayıtları getiren sorgu sonucu

| Adi   |
|-------|
| Özgür |
| Arda  |
| Emrah |

Çizelgeden de görüldüğü üzere tablodaki birbirinin aynı olan kayıtlar geri döndürülmemiştir.

**TOP:** Top(SELECT TOP) argümanı sorguyu, yalnızca onu karşılayan ilk n satırı, tabloya daha önce eklendikleri sırada geri döndürmeye zorlar.

Örnek: Çizelge 5.1 için aşağıdaki sorgu çalıştırıldığında sonuç seti çizelge 5.4'deki gibi olmaktadır.

Select top 2 Adi

From TableEx

Çizelge 5.4 Sütunda en üstteki iki kaydı getiren sorgu sonucu

| Adi   |
|-------|
| Özgür |
| Arda  |

#### 5.2.1.1 Seçim

Seçim listesi veriden seçilecek sütunu belirtir. Virgüllerle ayrılmış olarak seçim listesinde belirtilen öğeler; verilerin sonuç kümesi için seçileceği sütunları temsil eder.

Örnek:Aşağıdaki ifade çizelge 5.1 için çalıştırıldığında, çizelge 5.5 oluşmaktadır.

Select adi,soyadi

From TableEx

Çizelge 5.5 Seçilen sütunları getiren sorgu sonucu

| Adi   | Soyadı  |
|-------|---------|
| Özgür | Sarıhan |
| Arda  | Anıl    |
| Emrah | Çuvak   |
| Emrah | Kaçmaz  |

**Bütün Sütunlar(\*):** Tüm sütunların getirilmesidir. Yapılan sorguda tablodan veya görüntüden tüm sütunların gelmesi demektir. Sütun ismini vermeden veritabanı üzerinde sorgulanan tablo veya görüntüye ait tüm sütunları getirir.

Örnek: Aşağıdaki ifadede \*, TableEx tablosundaki bütün sütunların getirilmesini sağlar. Çizelge 5.1'de bu sorgu çalıştırıldığında aynı sonuç bulunmaktadır.

Select \*

From TableEx

**Adlandırma (c alias)** : Sütunlara yeniden isim verme veya başlık vermektir. Yapılan sorguda gelen verilerin başlığının değiştirilerek istenilen bir isimle gösterilmesidir. Takma isim verilirken AS kelimesi isteğe bağlı olarak verilebilir. Verilen takma isim sorguda sadece gelen verilerin sütun sıralanmasında kullanılabilir.

Örnek: Aşağıdaki ifadede “Adı” alanının adlandırma ile sonucunda başlığın ‘Ad’ olarak gösterilmesini sağlar. Çizelge 5.1’de bu sorgu çalıştırıldığında, çizelge 5.6 oluşmaktadır.

Select Adı Ad

From TableEx

veya

Select Adı As Ad

From TableEx

Çizelge 5.6 Sütun başlığı değiştirilmiş sorgu sonucu

| Ad    |
|-------|
| Özgür |
| Arda  |
| Emrah |
| Emrah |

### 5.2.1.2 FROM

FROM cümlecisi SELECT sorgusunu karşılayacak olan sütunların hangi tablolardan alınacağını belirler. SELECT deyimi içerisinde tablolara, görünümlere, türetilmiş tablolara ve tablo değişkenlerine her başvurduğumuzda FROM anahtar sözcüğüne ihtiyaç duyulur.

SELECT \* FROM tablo1

Yukarıdaki ifadede sorguya, tablodan bütün satır ve sütunları geri döndürmesi talimatını veren \*(yıldız), operatöründen faydalanılmıştır. Ayrıca sorguda aşağıdaki birden fazla tablo adı belirtilebilir.

```
SELECT * FROM tablo1,tablo2,tablo3
```

### 5.2.1.3 WHERE

WHERE anahtar bir koşul belirterek sadece o koşula uyan kayıtların getirilmesini sağlar. Sorguda kayıtların verilen koşul veya koşullarla sınırlandırılarak getirilmesidir. Verilen koşul doğru ise verileri getirir. Eğer bir koşul verilmezse tablo veya görüntü üzerinde tüm kayıtlar getirilir.

Örnek:Aşağıdaki ifade TableEx tablosunda adı 'Emrah' olan kayıtların getirilmesini sağlar. Çizelge 5.1'de bu sorgu çalıştırıldığında çizelge 5.7 oluşmaktadır.

```
SELECT *
```

```
FROM TableEx
```

```
WHERE adi='Emrah'
```

Çizelge 5.7 Koşul kullanılmış sorgu sonucu

| Numarası | Adi   | Soyadı |
|----------|-------|--------|
| 1        | Emrah | Çuvak  |
| 4        | Emrah | Kaçmaz |

### 5.2.1.4 GROUP BY

GROUP BY kayıtların gruplamasını sağlar. Sonuç kümesini alır ve onu gruplandırır. Gruplama yapılırken grup içerisinde istenilen hesaplamalar yapılabilir.

Örnek: Aşağıdaki ifadede aynı isme sahip olan kayıtları gruplayarak gösterir ve Count(\*) komutuyla kaçar tane oldukları gösterilir. Çizelge 5.1'de bu sorgu çalıştırıldığında sonuç çizelge 5.8'de gösterilmiştir.

```
SELECT Adi, Count(*) Adet
```

```
FROM TableEx
```

```
GROUP BY Adi
```

Çizelge 5.8 Grup yapılmış sorgu sonucu

| Adi   | Adet |
|-------|------|
| Özgür | 1    |
| Arda  | 1    |
| Emrah | 2    |

#### 5.2.1.5 HAVING

Sorguda gruplanmış kayıtların istenilen şekilde sınırlandırılmasıdır. İstenen koşul doğru ise verileri getirir. Buradaki amaç gruplamada ki hesaplamalar üzerinde sınırlama yapmaktır. Eğer bir koşul verilmezse tablo veya görüntü üzerinde tüm kayıtlar gruplanarak getirilir.

Örnek:Aşağıdaki sorguda HAVING deyimi GROUP BY ifadesiyle kullanılmaktadır.İfadede TableEx tablosunda ismi en az 2 defa geçen öğrenci isimlerini gruplanarak sayıları birlikte getirir. Çizelge 5.1’de bu sorgu çalıştırıldığında sonuç çizelge 5.9’da verilmiştir.

```
SELECT Adi, Count(*)
```

```
FROM TableEx
```

```
GROUP BY Adi
```

```
HAVING Count(*) >= 2;
```

Çizelge 5.9 Grup koşulu kullanılmış sorgu sonucu

| Adi   |
|-------|
| Emrah |

#### 5.2.1.6 ORDER BY

Yapılan sorguda gelen kayıtların istenilen sütuna göre sıralanarak getirilmesidir. ASC ve DESC sıralama şekillerini belirler. Eğer sıralama şekli belirtilmezse ASC olarak standart şekli otomatik olarak alınır.

Sıralama şekli:

ASC, küçükten büyüğe sıralar,

DESC, büyükten küçüğe sıralar.

Örnek: Aşağıdaki ifadede TableEx tablosu getirilirken, Adi sütununa göre alfabetik olarak sıralayarak getirir. Çizelge 5.1’de hasta adına göre büyükten küçüğe sıralama sorgusu çalıştırıldığında çizelge 5.10 sonucu elde edilir.

Select \*

From Prst0100

Order By Hasta\_Adi Desc;

Çizelge 5.10 Sıralama kullanılmış sorgu sonucu

| Numarası | Adi   | Soyadı  |
|----------|-------|---------|
| 2        | Arda  | Anıl    |
| 3        | Emrah | Çuvak   |
| 1        | Emrah | Kaçmaz  |
| 4        | Özgür | Sarıhan |

## 6. ÖĞRENEN SİSTEMLER

Bu bölümde SQL tabanlı yapılarda bulunan öğrenme metotları anlatılacaktır. SQO yaklaşımının, daha performanslı çalışması için sorgu iyileştirmesi yapılır. Bu işlem verilen sorguda bulunan koşula göre gerçekleştirilir. Sorgunun koşulu alınır ve sorgu sonuç setiyle öğrenme metotlarından biri kullanılarak sonuç verileriyle kurallar türetilir. Bu öğrenilen kurallar araştırmada DML ifadelerine uygulanır.

### 6.1 Bilgiden Bilgi Öğrenme

Sorgu iyileştirme yöntemleri, veri modelleri ve dillerinin (Ullman 88) ortaya çıkışından beri üzerinde çalışılan bir konudur. Çünkü genellikle, sorguların verimli uygulanması zordur. Sorgunun yeniden formüle edilmesi yaklaşımı, aynı zamanda anlamsal sorgu iyileştirme yaklaşımı (SQO) olarak da bilinir. İyileştirme yaklaşımına geleneksel sözdizimsel yaklaşımdan daha farklı yaklaşılr, şöyle ki sorguları iyileştirmek için veritabanlarının içerikleri hakkında daha zengin bir bilgi kümesi oluşumunu gerektirir. Anlamsal bilginin kullanımı maliyet indirimi sağlar. Verilerden elde edilen sorgunun sonuç ve koşulu birlikte kullanılarak kurallar üretilir. Üretilen bu kurallar yeni sorgulara uygulanır. Kuralların içerdiği özelliklere göre çalışma performansı artırılır. Bu nedenle SQL, içinde bilgiden bilgi öğrenme sistemini içerir.

Yeniden formüle etmek için otomatik kural türetme fikri (Siegel 1988) tarafından önerilmiştir. Siegel'in yaklaşımında her ne kadar örnek sorgular kullanıyorsa da öğrenme aslında sabit bir Heuristic'ler kümesinden yürütülür. Heuristic'ler veritabanı yapısına ve uygulamasına göre tasarlanırlar. Daha sonradan bu fikir (Shekhar, S 1993) tarafından varolan veritabanı için sorgular gelmeden önce olası sorguların tamamına uygulanacak kuralları baştan öğrenme olarak geliştirilmiştir.

Bu çalışmada ek olarak kullanılan yöntemde Shekhar algoritmasına ek olarak veritabanlarında öğrenilebilecek kuralların gereksiz olma ihtimaline karşılık indeksli kolonlara atılan sorgular için kural öğrenme ihmal edilmektedir. Bu yaklaşımın amacı, veritabanlarında öğrenilecek kuralların gerekliliği ile varolan indeksli kolonlara eklenecek index içermeyen kural öğrenmelerinin sorguyu olduğundan daha da yavaşlattığının ispatına dayanır.

Sonuç olarak elimizdeki kurallardan yeni kurallar çıkarabilme ve bu kuralları kullanabilme performans için çok etkilidir. Kurallardan kural öğrenme işleminde oluşan kuralın veritabanı üzerinde doğruluğunun kesin olması gerekir, bu işlem için veritabanına erişerek veya elimizdeki kuralların veritabanı üzerinde kaç tane satırda bulunduğu bilgisini de kullanabiliriz.

Örneğin : Elimizde aşağıdaki gibi iki kural olsun

R1) Pers\_No = '00000006' --> Personel\_Ad = 'Süleyman'

R1) SSN\_No = '2121234567' --> Personel\_Ad = 'Süleyman'

Burada personel numarası '00000006' olan çalışanın adı 'Süleyman' ve veritabanında 'Süleyman' ismi tek ise çalışan numarası '00000006' olan personelin sosyal güvenlik numarası '2121234567' olarak bulunur.

R3) Pers\_No = '00000006' --> SSN\_No = '2121234567'

Bu şekilde R1 ve R2 kurallarından R3 kuralı elde edilmiş olur.

## 6.2 Öğrenme Metotları

Anlamsal sorgu eniyileştirme (SQO: Semantic Query Optimization) yaklaşımında, daha iyi bir sorgu uygulanması için, sorgu iyileştirme yapılır ve bunun için kurallar kullanılır. Bununla birlikte kurallar sorgu çalıştırılmadan önce türetilmelidirler. Kural türetmede ki ana düşünce, o anki sorgu ve verilen veritabanına bağlı olarak kuralları otomatik olarak öğrenmektedir. Kural türetmede, SQO yaklaşımının performansını etkileyen bir faktördür. Eğer bir kural türetilmediyse, sorgu iyileştirmede kullanılacak bir kural olmayacaktır ve iyileştirme

yapılamayacaktır. Bu nedenle kuralların öğrenilmesi SQO'da çok önemlidir ve bir çok araştırmacı tarafından sorgu iyileştirmesinde kural kullanımının avantajları incelendiği gibi öğrenmede araştırılmıştır. (Sayli ve Lowden 1996, 1997a-b ).

Kural öğrenme farklı sınırlandırmalar ile yapılabilir. Sadece gerekli olan kuralları öğrenmek zaman maliyeti açısından önemlidir. Bu çalışmada zaman ölçümlerini daha net yapabilmek için eşitlik (=), küçüklük () ve büyüklük karşılaştırmaları yapılmış, kurallar buna göre öğrenilmiştir.

### **6.2.1 Siegel Yöntemi**

(Siegel 1988) tarafından, SQL yaklaşımına otomatik kural türetme için temel bir yöntem olarak sunulmuştur. Bu yöntem iki bölümlü bir işlem olarak görülebilir. İlk bölüm kuralların karakteristiklerini belirler. İkinci bölümde ise bu karakteristiklere sahip yeni kuralları türetmek için veritabanından sorgulama yapılır. Bu bölümler SQO yaklaşımında yararlı kurallar bulunmasını sağlar. Bu yöntemle öncelikle bir 'önerilen kural' fikri oluşturulur ve önerilen kuralın önceki durumu, sorgunun herhangi bir durumunu alarak oluşturulur. Sonuç sütunu, heuristics tarafından bulunur ve sonra yeni bir kural türetmek için veritabanına girilerek uygun bir koşul bulunabilir. Bu yöntem dört modülde açıklanmıştır: kural karakteristiklerini tanımlama, önerilen kuralların seçimi, sorgu üretilmesi ve kural yönetimidir. Bunlar bu bölümde açıklanmaktadır.

#### **6.2.1.1 Kural Karakteristiklerini Tanımlama**

Kural karakteristiklerini tanımlama yöntemin ilk modülüdür. Bu modülün amacı, işe yarar kurallar türetmek için sorgu iyileştiricisi tarafından kural türetme işleminin kontrol edilmesidir. Kuralları var olan kurallardan veya doğrudan veritabanından türetebilmek mümkündür. Yinede, eğer türetme işlemi sorgu iyileştiricisine bağlı değilse bu işlem SQO yaklaşımı için gereksiz kurallar bulmamıza neden olabilir. Bu modülde, SQO yaklaşımı beklenen maliyet tasarrufu ile birlikte bütün uyumsuz önerilen kuralları, kural türetme modülüne gönderir ve sonra bu modül tarafından incelenen önerilen kurallar, hesaplanan potansiyel maliyet tasarrufları ile birlikte önerilen kural listesine girer. Eğer bu kural listede mevcutsa birikmiş potansiyel maliyet tasarrufu bu önerilen kural için hesaplanır. Bundan sonra, belli bir eşiğin altındaki birikmiş potansiyel maliyet tasarruflu önerilen kurallar, önerilen kural listesinden silinmelidir. Aksi takdirde, bu önerilen kurallar yeni kurallar gibi algılanır.

### 6.2.1.2 Önerilen Kuralları Seçme

Yöntemin ikinci modülü önerilen kuralları seçmektir. Bu modül türetilen en iyi kuralları belirlemede kullanılır. Önerilen kuralların belirlenmesi uzmanlar için çok zor bir iştir. Bu sebeple, gittikçe artan beklenen maliyet tasarruflarını hesaplama, önceki tecrübelerle dayandırılır. Potansiyel değeri yeterince yüksek olan uyumlu önerilen kurallardan biri listede yer aldığı anda, önerilen kurallar işlem için seçilebilir. Bu işlem boyunca, daha düşük potansiyel değeri var olan kurallar, kurallar kümesinden silinecektir. Bu işlem kurallar kümesinin büyüklüğü ve kalitesinin belirlenmesinde kullanılabilir. Aynı amaçla, kural kümesinin en fazla değerine sınır koymak ve bir kuralın en az değerine sınır koymakta yararlı olabilir. Eğer kurallar bu en az değer altında bir değere sahipse kurallar kümesinden silinecektir. (Gökçe 2003)

### 6.2.1.3 Sorgu Üretme

Sorgu üretme yöntemin üçüncü modülüdür. Bu modül, ikinci modül tarafından seçilen bir önerilen kural alır. Temel olarak, modül önerilen kuralı kullanarak bir sorgu şablonu üretir. Bundan sonra veritabanı üzerinde sorgu çalıştırılır. Bu işlem boyunca ortaya çıkabilecek birkaç durum söz konusudur:

- i. Eğer bir önerilen kural bir koşulun çıkarılması ile bulunduysa sonucu ile ilgili bir koşul olacaktır. Bu koşul üretilen sorguda kullanılmaz.
- ii. Eğer önerilen kuralda, iki nesne ilişkisi arasında bir bağlantı her hangi bir sütun ile geliyorsa, bu ilişki koşulu kural oluşturulmasında kullanılmaz.

### 6.2.1.4 Kural Yönetimi

Kural yönetimi yöntemin son modülüdür. Bu modül, kurallar kümesine yeni bir kuralın eklenip eklenmeyeceğine karar verir. Modülde, türetilmiş kural, üretilen sorgunun cevabından ve önerilen kuraldan oluşturulur. Eğer cevap hükümsüzse, yeni kural olamaz ve önerilen kural çıkarılır. Eğer önerilen kural bir koşul çıkarılması ile oluşuyorsa, önerilen kuralın sonucu sorgunun cevabına göre kontrol edilmelidir. Bu durumda türetilmiş kuralın sonucu sorgunun olabilecek tüm cevaplarını kullanarak bulunabilir. Bu yöntem, olabilecek tüm kuralların türetilmesinde kullanılabilir. Yinede sadece gerekli kuralları bulabilmek, otomatik kural türetmedeki en önemli faktördür.

### 6.2.1.5 Algoritma (Sorgu Tabanlı)

Adım 1: Gelen sorgunun var olan kurallarını bulmak

Adım 2: Var olan kuralları taratarak aynı sonucu döndürecek kuralları bulmak

Adım 3: Öğrenilen kuralları sorguya eklemek

Adım 4: Optimize edilmiş sorguyu çalıştırmak

Adım 5: Sorgunun sonucunu döndürmek

### 6.2.2 Shekhar Yöntemi

Shekhar kural öğrenmesinde , öncekilere oranla tekrarı önlediği ve kaybı azalttığı için Veri – Tabanlı Niteliksel kural öğrenmesini önermektedir. Veri dağılımlarını tabloların içeriği, koordinatları ve aralığını belirleyerek formülize etmiştir. Sorgu dönüşümlerini ve data dağılımları arasındaki ilişkiyi karakterize ederek sorguların eşitliğini, kuralların doğruluğunu ve sorgu dönüşümlerini sağlamıştır. Shekhar vd. (1993)

#### 6.2.2.1 Anlamsal Sorgu Eniyilemesi için Kural Dönüşümleri Öğrenmeleri

Bu konuda öncelikle sorgular için kullanılacak sorgu dili belirlenir. Sonra öğrenilecek sorgu kurallarının sınıfları kategorize edilir. Bundan sonra temel dönüşüm kurallarının veri – tabanlı desene uygunluğu ortaya koyulur. Bu desen veri – tabanlı kural keşfi algoritmasıyla olacaktır.

#### 6.2.2.2 Sunum Dili

Shekhar mantık tabanlı bir sunum dili kullanmıştır.

Bağlantılar ve ilişkiler için P

Operatörler (>,<=) için op

Bağlantılar için 'not' , 'or' , '->'

Kullanılmıştır. '?' ise sorguları bildirmektedir.

Örneğin; Maksimum ağırlığı 400 grt den büyük ve Kargo Değeri 4000 Amerikan Dolarından fazla olan gemilerin varış noktaları nereleridir?

İfade şu şekilde sunulabilir;

$$Q = \text{SHIPS}(\text{DeadWt} > 400) \text{ and } \text{CARGOES}(\text{DollarValue} > 4000), ? \text{ Destination}$$

$$= (\text{DeadWt} > 400) \text{ and } (\text{DollarValue} > 4000) : (? \text{ Destination})$$



Şekil 6.1’de gemi veritabanına bağlı data dağılımını gösteren bir tablo verilmiştir. Tablo kesin ayrımlı koordinatlarla birbirinden bağımsız hücrelerden oluşmaktadır. Bu tabloda birleşme (union), kesişme (intersection) ve fark (negation) operatörleriyle işlemler aralığın doğasına bağlı olarak yada hücrede bulunan değerin doğasına göre gerçekleştirilebilir.

Aralığın doğasına göre olabilecek bağlantılar şöyle olabilir.

$$\begin{aligned}
 \text{birleşim } (R1,R2) &= \begin{cases} \text{Aralıksal birleşim } (R1,R2) & \rightarrow \text{eğer } R1,R2 \text{ ortak koordinatlarda bir aralıksa} \\ \text{Mantıksal ayrılık } (R1,R2) & \rightarrow \text{diğer} \end{cases} \\
 \text{kesişim } (R1,R2) &= \begin{cases} \text{Aralıksal kesişim } (R1,R2) & \rightarrow \text{eğer } R1,R2 \text{ ortak koordinatlarda bir aralıksa} \\ \text{Mantıksal birleşim } (R1,R2) & \rightarrow \text{diğer} \end{cases} \\
 \text{fark } (R1,R2) &= \begin{cases} \text{Aralıksal zıtlık } (R) & \rightarrow \text{eğer } R \text{ bir koordinat aralığıysa} \\ \text{Mantıksal zıtlık } (R) & \rightarrow \text{diğer} \end{cases}
 \end{aligned}$$

Şekil 6.2 Bağlantı tanımlamaları

Burada Mantıksal – ayrılık (OR), Mantıksal – birleşim (AND) ve Mantıksal – zıtlık (NOT) operatör olarak “or” , “and” ve “not” ifadelerini temsil eder.

Mantıksal ayrılık; iki ifadeden herhangi birinin (ya da ikisinin birden) doğruluğunu sağlama,

Mantıksal birleşim; iki ifadenin aynı anda doğruluğunu sağlama,

Mantıksal zıtlık; ifadenin doğruluğunu reddetme (sağlamama) durumudur.

Bu bağlamda veritabanındaki veriler ikiye ayrıştırılabilir; sıralı veriler ve sırasız veriler.

Sırasız veri kümeleri aralarında herhangi sıraya dayalı bir ilişki olmayan veriler kümesidir. Örneğin renkler kümesi (kırmızı, yeşil, siyah ..vb..) verilerinin aralarında sıralamaya dair herhangi bir bağlantı olmayan kümedir. Sadece “=” veya “≠” ifadeleri ile karşılaştırılabilir ve şöyle ifade edilir. “A = B” ya da “A != B”

Sıralı veri kümeleri ise aralarında bir sıralamaya dayalı ilişkiler bulunan veriler kümesidir. Ve birbirleriyle ilişkileri arada artan ve ya azalan matematiksel bir ifadeye dayanır. Bu kümelerde “=” ve “≠” ifadelerinin yanında “<”, “>”, “≤” ve “≥” operatörlerini kullanmak ta mümkündür.

Sonucu A olan ve A olmayan sorgular için bir ayrım yapacak olursak şöyle bir tablo ortaya çıkar.

| A     | Not(A)                    |
|-------|---------------------------|
| $a=x$ | $(a>x) \text{ OR } (a<x)$ |
| $a>x$ | $a\leq x$                 |
| $a<x$ | $a\geq x$                 |

Şekil 6.3 Operatörlerin zıtlığı

| Clauses              | $X = Y$                                         | $X > Y$                       | $X < Y$                       |
|----------------------|-------------------------------------------------|-------------------------------|-------------------------------|
| $A1:(a=x), A2:(a=y)$ | $a=x$                                           | $(a=x) \text{ or } (a=y)$     | $(a=x) \text{ or } (a\leq y)$ |
| $A1:(a>x), A2:(a=y)$ | $a \geq x$                                      | $(a>x) \text{ or } (a\leq y)$ | $a>x$                         |
| $A1:(a<x), A2:(a=y)$ | $a\leq x$                                       | $a<x$                         | $(a<x) \text{ or } (a\leq y)$ |
| $A1:(a=x), A2:(a>y)$ | $a\geq x$                                       | $a>y$                         | $(a=x) \text{ or } (a>y)$     |
| $A1:(a=x), A2:(a<y)$ | $a\leq x$                                       | $(a=x) \text{ or } (a<y)$     | $a<y$                         |
| $A1:(a>x), A2:(a>y)$ | $a>x$                                           | $a>y$                         | $a>x$                         |
| $A1:(a>x), A2:(a<y)$ | $(a\neq x) \text{ or } (a<x) \text{ or } (a>x)$ | $(a>x) \text{ or } (a<y)$     | $-\infty < a < +\infty$       |
| $A1:(a<x), A2:(a<y)$ | $a<x$                                           | $a<x$                         | $a<y$                         |
| $A1:(a<x), A2:(a>y)$ | $(a\neq x) \text{ or } (a<x) \text{ or } (a>x)$ | $-\infty < a < +\infty$       | $(a<x) \text{ or } (a>y)$     |

Şekil 6.4 A1,A2 birleşim operatörü (Union(A1,A2)) ve ortak a değerinin kullanımı

RC(S) kapalı aralığını S in aralıklarının kümesi olarak tanımlayacak olursak, aralık operatörleri şu kurallara uymak durumundadır;

1. S in ait olduğu aralıkların her biri RC(S) kapalı aralığının içerisinde.
2. Eğer  $r1 \in S$  ,  $r2 \in S$  ise birleşim( $r1,r2$ ) , kesişim( $r1,r2$ ) , zıt( $r1$ ) ve zıt( $r2$ ) RC(S) kapalı aralığındadır.

| Clauses              | $X = Y$ | $X > Y$ | $X < Y$ |
|----------------------|---------|---------|---------|
| $A1:(a=x), A2:(a=y)$ | $a=x$   | NULL    | NULL    |
| $A1:(a>x), A2:(a=y)$ | NULL    | NULL    | $a=y$   |
| $A1:(a<x), A2:(a=y)$ | NULL    | $a=y$   | NULL    |
| $A1:(a=x), A2:(a>y)$ | NULL    | $a=x$   | NULL    |
| $A1:(a=x), A2:(a<y)$ | NULL    | NULL    | $a=x$   |
| $A1:(a>x), A2:(a>y)$ | $a>x$   | $a>x$   | $a>y$   |
| $A1:(a>x), A2:(a<y)$ | NULL    | NULL    | $x<a<y$ |
| $A1:(a<x), A2:(a<y)$ | $a<x$   | $a<y$   | $a<x$   |
| $A1:(a<x), A2:(a>y)$ | NULL    | $y<a<x$ | NULL    |

Şekil 6.5 A1,A2 kesişim operatörü (intersection(A1,A2)) ve ortak a değerinin kullanımı

#### 6.2.2.5 Sorgu Dönüşüm Kurallarının Bulunması

Önceden tanımlı kullanıcı tabanlı dönüşüm kurallarıyla bir sorgu dönüşüm deseni oluşturulur. Örneğin bir sorguda seçilen verinin a iken sonucun b olması yada olmaması gibi yani (Attribute1 = "a") -> (Attribute2 = "b") ya da (Attribute1 = "a") -> (Attribute2  $\neq$  "b")

Burada bulunan bağıntı sorgu sonuçlarının aynı olması durumudur. (Attribute1 = "a") -> (Q = "c") ise ve (Attribute2 = "b") -> (Q = "c") ise

(Attribute1 = "a") : ?Q sorgusunda

(Attribute1 = "a") and (Attribute = "b") : ?Q yazılabilir ve (Q = "c") sonucu döner. Kısıtlama fazla olduğundan süre kısa ama sonuç aynıdır.

#### 6.2.2.6 Algoritma (Tam Arama)

Adım 1: Tüm tabloda var olası sorgu kurallarını bulmak

Adım 2: Bulunan kuralları taratarak aynı sonucu döndürecek yeni kuralları bulmak

Adım 3: Gelen sorguya göre öğrenilen kuralları eklemek

Adım 4: Optimize edilmiş sorguyu çalıştırmak

Adım 5: Sorgunun sonucunu döndürmek

### 6.2.3 İndeksli Arama Yöntemi

Bu çalışmada anlatılan metodlardan ayrı olarak, geliştirilen uygulamada sorguları sadece indexli kolonlara göre arayan ve kuralları sadece bu kolonlardan öğrenen indeksli arama metodu kullanılmıştır (InSeQT: Index-Search Query opTimization). Bu arama metoduna göre öğrenilecek kurallar sadece indeks kayıtları bulunan değerler üzerinden yeni kurallar öğrenmektedir. Bu sayede sorguya katkısı olmayan gereksiz kayıtlar öğrenmeye başlanmadan elenmiş olmaktadır.

#### 6.2.3.1 Algoritma (İndex Arama)

Adım 1: Tüm tabloda var olası sorgu kurallarını bulmak

Adım 2: Bulunan kuralları sadece indexleri var olan kolonlarda taratarak aynı sonucu döndürecek yeni kuralları bulmak

Adım 3: Gelen sorguya göre öğrenilen kuralları eklemek

Adım 4: Optimize edilmiş sorguyu çalıştırmak

Adım 5: Sorgunun sonucunu döndürmek

## 7. UYGULAMA

### **Bu uygulamada kullanılan:**

Program Geliştirme Ortamı: Visual Basic.Net

Veritabanı: SQL Server 2005

Veri Erişim Yöntemi:ADO

Veri Tablosu:Araba

Veri Sayısı:25000

Sütun Sayısı:16

Tablo Anahtar Değeri:id

Tablo dizin(index) değeri: customer\_id, customer\_vehicle\_id, married\_year, Age, mernis\_no, child\_count

|               |             |    |       |
|---------------|-------------|----|-------|
| Tablo Şeması: | ID          | -> | int   |
|               | customer_id | -> | float |

|                         |    |               |
|-------------------------|----|---------------|
| customer_vehicle_id     | -> | float         |
| mernis_no               | -> | float         |
| education               | -> | nvarchar(255) |
| gender                  | -> | nvarchar(255) |
| marital_status          | -> | nvarchar(255) |
| child_count             | -> | float         |
| married_year            | -> | float         |
| age                     | -> | float         |
| corporation_sector      | -> | nvarchar(255) |
| occupation_status       | -> | nvarchar(255) |
| corporation_position    | -> | nvarchar(255) |
| countrymodel_definition | -> | nvarchar(255) |
| occupation              | -> | nvarchar(255) |
| top_model_definition    | -> | nvarchar(255) |

Sisteme gönderilen öntanımlı sorgu tablosu şöyledir;

Çizelge 7.1 Sorgu tablosu

|     |                                                                |
|-----|----------------------------------------------------------------|
| S1  | select * from Araba where ID = 2225;                           |
| S2  | select * from Araba where ID <= 3000;                          |
| S3  | select * from Araba where ID >= 20000;                         |
| S4  | select * from Araba where marital_status = 'Bekar';            |
| S5  | select * from Araba where countrymodel_definition = 'AUDI A4'; |
| S6  | select * from Araba where occupation = 'SENDİKACI';            |
| S7  | select * from Araba where mernis_no <= '10009835602';          |
| S8  | select * from Araba where mernis_no = '74923039232';           |
| S9  | select * from Araba where child_count = '3';                   |
| S10 | select * from Araba where child_count <= '3';                  |
| S11 | select * from Araba where child_count >= '1';                  |

Bu çalışmada Siegel ve Shekhar'ın öngörülerine göre dinamik bir veri tabanının üzerinde kural öğrenmesi yapan bir uygulama geliştirilmiştir. Uygulama, sorgu sonucuna göre, olası bütün sorgular için ve olası sorgulardaki index içermeyen kolonların indexli karşılıkları için olmak üzere üç farklı metod için çalıştırılabilmekte ve hepsi için kuralları üretmektedir.

Çalışmada her üç metod için program çalıştırılıp kurallar öğrenilmiş, çıkan kural tabloları sorgulara uygulanmış ve sorguların iyileştirilmesi ölçülüp karşılaştırılmıştır. Uygulamada kullanılan veri tabanı SQL Server 2005 üzerindedir. Programlama ortamı VS2008 ve Programlama dili VB.Net dir. Veri erişim yöntemi Ado.net dir.

### 7.1 Uygulamanın Yapısı

Program ilk çalıştırıldığında ana ekran kullanıcıyı karşılar. Ana ekranda program kullanıcıdan birtakım seçimler yapmasını istemektedir. Kullanıcı bu ekrandan kullanacağı MSSQL programının sunucu (server) ismini seçecek ve o sunucu ismine ait Veritabanları arasından kullanacağı veritabanı ile ilgili tablosunu belirleyecektir. Program Veritabanı ve tablo seçildikten sonra yapacağı işlemleri bu seçime göre düzenleyecek ve uygulayacaktır. Çalışmada lokal sunucu üzerindeki “ozgur” veritabanı altından, 25000 kayıtlık “Araba” tablosu seçilmiştir. Kurallar bu tabloya göre oluşturulmuştur. Uygulamanın ana ekranında kullanıcı, çalıştıracağı kural metodunu da tek seçimlik düğmeler (Radiobutton) yardımıyla seçer.



Şekil 7.1 Uygulama ana ekranı

Server Name: (local) 

Database Name: ozgur

Table Name: Araba

Method:

☒ Sorgu

☐ Tam Arama

☐ İndeks Arama

START PAUSE EXIT



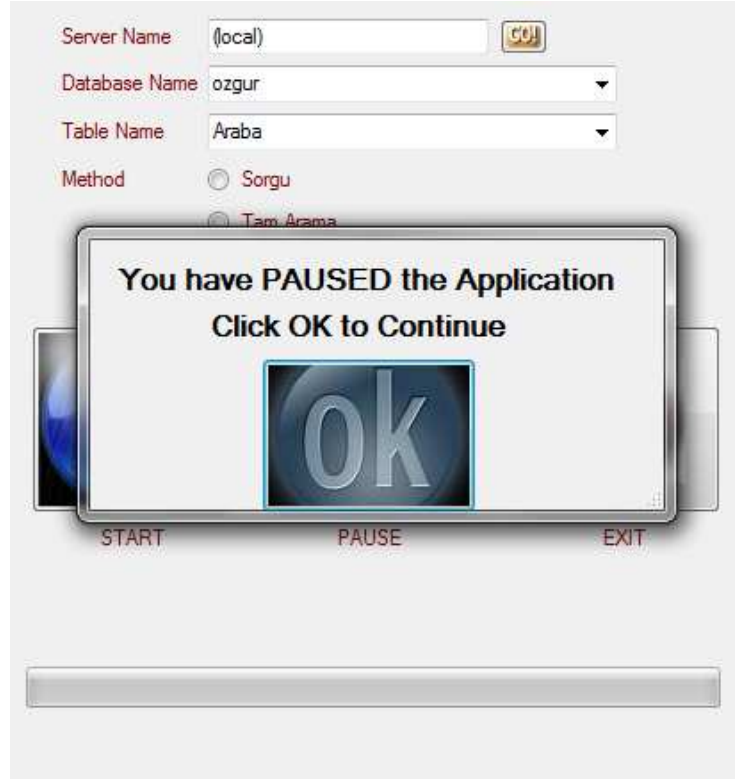
Şekil 7.2 Veritabanının tanıtılması ve metod seçimi

Programda Sorgu bazlı kural öğrenmesi yaptırılabilmesi için seçilecek veritabanında daha önceden tanımlanmış bir Queries tablosu ve içerisinde sorguların bilgileri bulunmalıdır. Çalışmada kullanılan Queries tablosu için örnek sorgular Çizelge 7.1’de gösterilmiştir.

Program yapısı itibariyle 4 tane tuştan oluşur. Bunlardan birincisi sunucu seçiminde kullanılan “GO” tuşudur. Söz konusu sunucu adı yazıldıktan sonra bu tuşa basılarak uygulamanın yazılan sunucu ismiyle bağlantı kurması sağlanır. Uygulama otomatik olarak sunucu üzerindeki veritabanı isimlerini ve bunlara bağlı tablo adlarını ilgili seçim bölgelerine getirecektir.

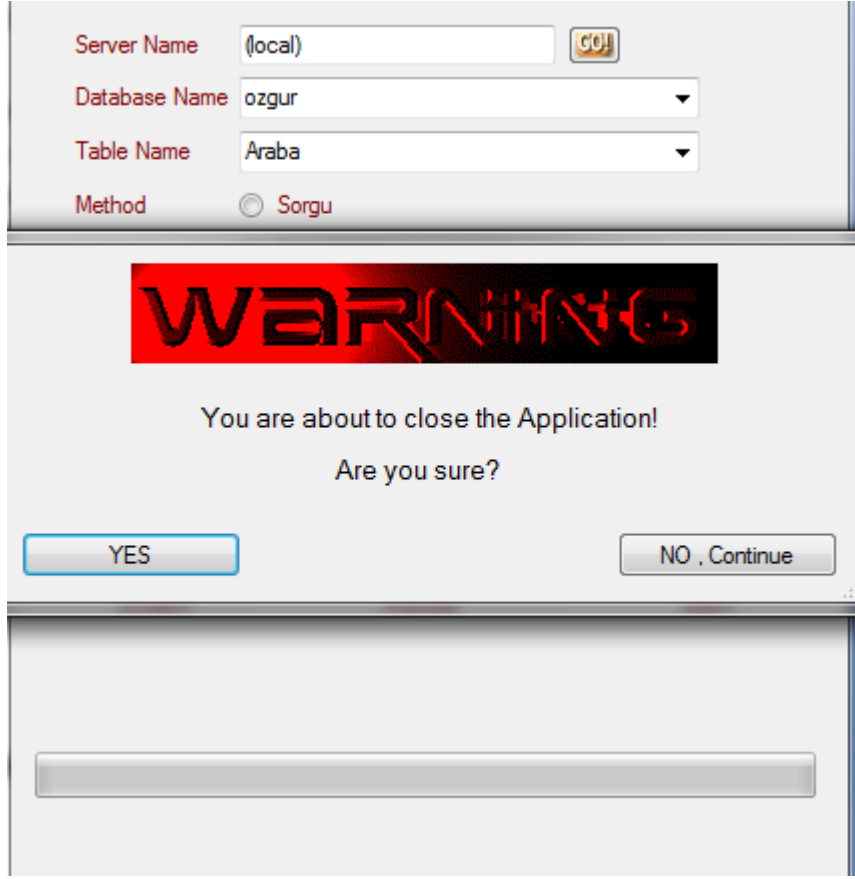
Programın START tuşu, gerekli seçimler yapıldıktan sonra kural öğrenmenin başlaması için kullanılır. Bu tuş uygulamayı koşturan tuştur.

PAUSE tuşu , uygulamanın herhangi bir anda duraklayarak tekrardan devam etme komutu alana kadar beklemesini sağlar. Bu tuşa basıldığında uygulamada, Şekil 7.3 teki ekran açılacak ve kaldığı yerden devam edebilmek için kullanıcının OK tuşuna basması beklenecektir.



Şekil 7.3 Bekleme ekranı

EXIT tuşu adından da anlaşılacağı üzere çıkış tuşudur. Uygulama bu tuşa basıldığında işlemi bekleterek kullanıcıya Şekil 7.4’de ki son karar ekranını getirir. Buradaki seçime göre program ya işlemi kaldığı yerde keserek çıkış yapıp kapanacak, ya da kaldığı yerden koşturmaya devam edecektir.



Şekil 7.4 Çıkış ekranı

## 7.2 Uygulamanın Çalıştırılması

Uygulama seçimler yapıldıktan sonra START tuşuna basılarak çalıştırılır. Uygulama ilk olarak çalıştırıldıktan sonra kullanıcıya kullanılacak tablonun sınırlandırılmasını isteyip istemediğini sorar. Bu işlem büyük veritabanlarında uygulamanın test amaçlı çalıştırılabilmesi ve aynı zamanda tabloyu sınırlandırarak yeni bir tablo gibi hareket etmesini, böylece ölçümlerin sağlıklı olup olmadığının kontrolünü sağlar.



Şekil 7.5 Sınırlandırma ekranı

Seçilen metoda göre program üç farklı şekilde kural öğrenmesi yapacaktır.

Birincisi Sorgu koşulunda belirtilen değerlerin tablo içerisinde başka değerlerce sağlanıp sağlanmadığına, ikincisi sorgudan bağımsız kaynak tablodaki olası tüm koşulları sağlayan birlikteliklere, üçüncü olarak da yine sorgudan bağımsız, kaynak tablodaki olası tüm indeks olmayan kolonlar için indeksli bir koşul eklenip eklenemeyeceğine göre kuralları öğrenir.

Öğrenilen kurallar RULESET denilen kural tablosuna yazılır. Sorgular çalıştığında bu tablodan öğrenilen kuralları da kullanacaklardır.

Uygulamanın kuralları öğrenme kodlaması Ek 5’ de verilmiştir.

### 7.3 Uygulamanın Sonuçları

Çalışmada uygulama önce 5000 kayıda sınırlandırılarak sonra da 25000 kayıtlık Araba isimli tüm tablo (Ek1) için üç farklı metod ile çalıştırılmıştır. Elde edilen sonuçlar EK-2,EK-3 ve EK-4 de örnek olarak verilmektedir. Uygulamadan şu sonuçlar elde edilmiştir;

Çizelge 7.2 Sonuç tablosu

**Sorgu (Siegel) 5000 kayıt**

| Süre (sn) | Kural | Kural/Süre<br>(krl/sn) | Süre/Kural<br>(sn/100 krl) |
|-----------|-------|------------------------|----------------------------|
| 1,2       | 72    | 60                     | 1,666666667                |

**Sorgu (Siegel) 25000 kayıt**

| Süre (sn) | Kural | Kural/Süre<br>(krl/sn) | Süre/Kural<br>(sn/100 krl) |
|-----------|-------|------------------------|----------------------------|
| 2,3       | 280   | 121,7391304            | 0,821428571                |

**Tam Arama (Shekhar) 5000 kayıt**

| Süre        | Süre(sn) | Kural  | Kural/Süre<br>(krl/sn) | Süre/Kural<br>(sn/100 krl) |
|-------------|----------|--------|------------------------|----------------------------|
| 33 dk:34 sn | 2014     | 113984 | 56,5958292             | 1,766914655                |

**Tam Arama (Shekhar)****25000 kayıt**

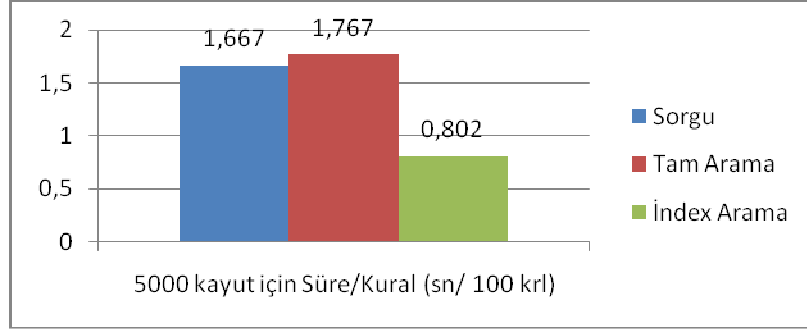
| Süre           | Süre(sn) | Kural   | Kural/Süre<br>(krl/sn) | Süre/Kural<br>(sn/100 krl) |
|----------------|----------|---------|------------------------|----------------------------|
| 155 dk:37,5 sn | 9337,5   | 1617670 | 173,2444444            | 0,577219087                |

**İndex Arama (InSeQt) 5000 kayıt**

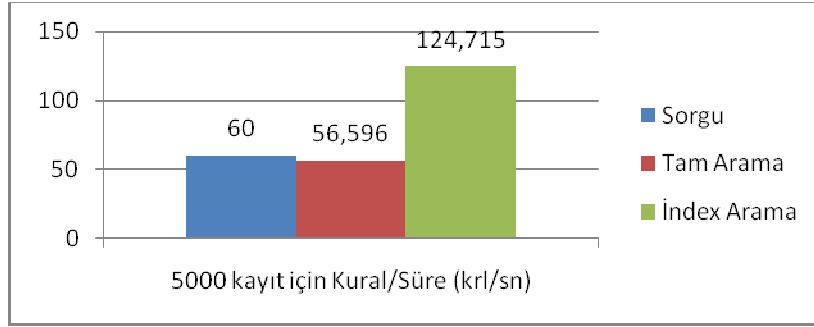
| Süre    | Süre(sn) | Kural | Kural/Süre<br>(krl/sn) | Süre/Kural<br>(sn/100 krl) |
|---------|----------|-------|------------------------|----------------------------|
| 36,9 sn | 36,9     | 4602  | 124,7154472            | 0,801825293                |

**İndex Arama (InSeQt) 25000 kayıt**

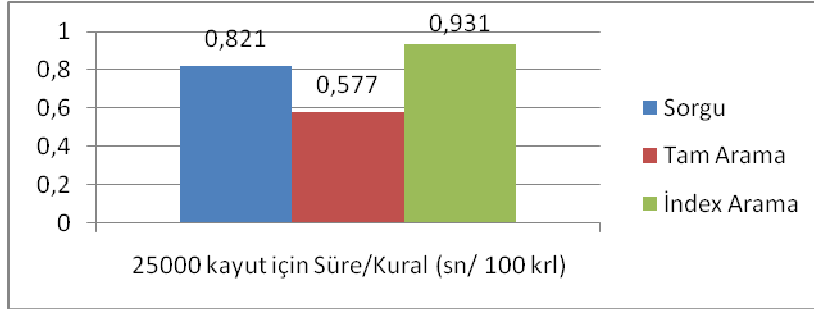
| Süre         | Süre(sn) | Kural | Kural/Süre<br>(krl/sn) | Süre/Kural<br>(sn/100 krl) |
|--------------|----------|-------|------------------------|----------------------------|
| 2 dk 01.9 sn | 121,9    | 13100 | 107,4651354            | 0,930534351                |



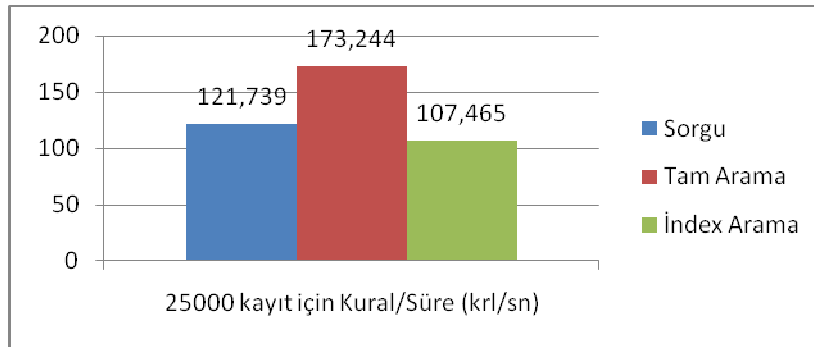
Şekil 7.6 5000 kayıta süre / 100 kural



Şekil 7.7 5000 kayıta bir sn de öğrenilen kural sayısı



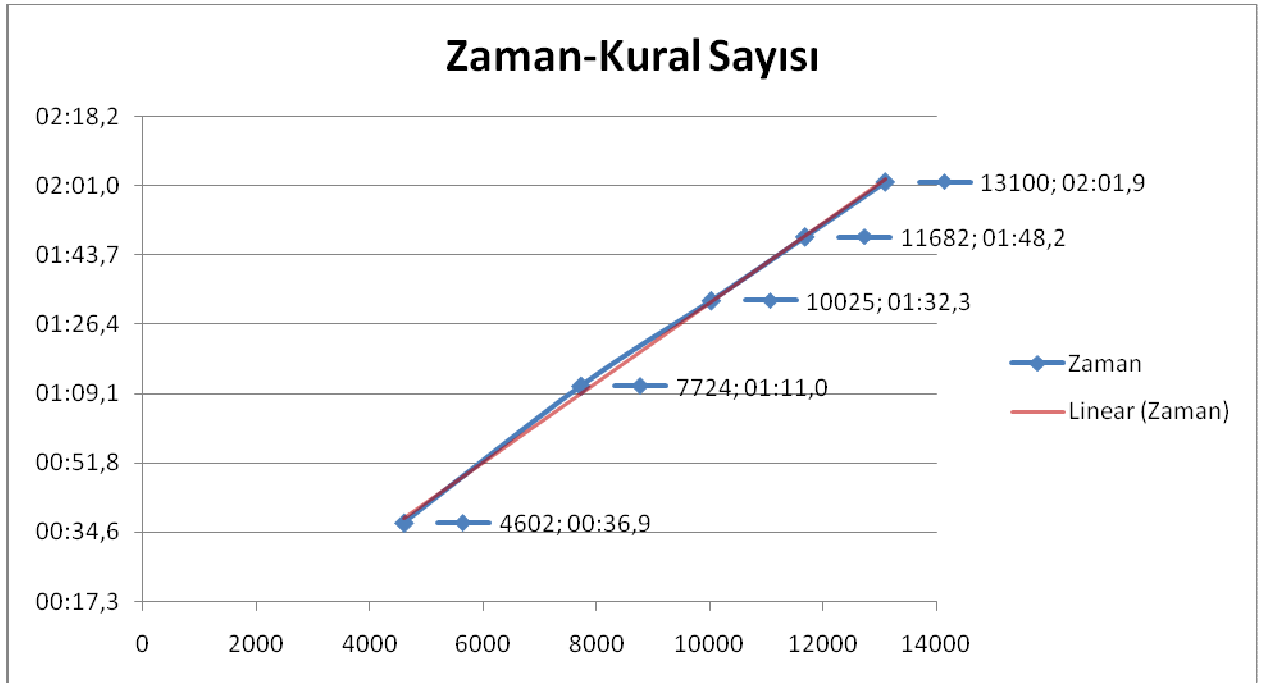
Şekil 7.8 25000 kayıta süre / 100 kural



Şekil 7.9 25000 kayıta bir sn de öğrenilen kural sayısı

Uygulama ayrıca indeks arama metoduyla sonuçların güvenilirliği açısından 5000 – 10000 – 15000 – 20000 ve 25000 kayıt için ayrı ayrı çalıştırılmıştır. Şekil 7.10’da zaman – öğrenilen kural sayısı bağıntısı elde edilen sonuçlara göre çizilmiştir. Buna göre öğrenilen kural sayısı tablonun büyüklüğüne göre lineer bir artış göstermektedir.

Şekil 7.7 ve Şekil 7.9’da görüldüğü üzere kayıt sayılarındaki artış sonucu öğrenilen kayıt sayısı da artmaktadır. Şekil 7.6 ve Şekil 7.8 karşılaştırıldığında her 100 kural başına düşen ortalama süre sadece indexli arama metodu süresinde artış vardır. Normalde kayıt sayısı bu metodda olduğu gibi kayıt sayısı arttıkça yavaşlamaktadır. Diğer metodlarda artış gibi görülen süre kısılması ise gereksiz kuralların bulunması yüzünden kayıt başına düşen sürenin azalmasıdır.



Şekil 7.10 Zaman-Kural sayısı ölçümü

#### 7.4 Sorguların İyileştirmesi

Uygulamanın çalıştırılmasıyla elde edilen kural tablolarına göre sorgular tek tek sisteme gönderilmiş ve sorguların çalışma zamanları Çizelge 7.3 ‘ de karşılaştırılmıştır.

Çizelge 7.3 Sorgu süreleri karşılaştırması

|     | Orjinal<br>Süre | Sorgu<br>Süre | Tam Arama<br>Süre | İndex Arama<br>Süre |
|-----|-----------------|---------------|-------------------|---------------------|
| S1  | 16,771          | 22,333        | 17,457            | 16,771              |
| S2  | 88,73           | 98,662        | 97,328            | 88,73               |
| S3  | 144,512         | 153,256       | 151,071           | 144,512             |
| S4  | 134,104         | 134,995       | 132,971           | 132,885             |
| S5  | 17,747          | 17,577        | 17,301            | 16,571              |
| S6  | 21,642          | 29,764        | 29,442            | 23,485              |
| S7  | 19,657          | 89,184        | 88,885            | 19,657              |
| S8  | 15,914          | 20,944        | 20,714            | 15,914              |
| S9  | 63,838          | 373,137       | 372,171           | 63,838              |
| S10 | 605,971         | 2395,008      | 2385,9            | 605,971             |
| S11 | 399,428         | 3641,303      | 3635,3            | 399,428             |

Çizelge 7.3’de her üç metodla çalıştırılan sorguların kural öğrenilmeden ve öğrenildikten sonra döndükleri süreler verilmiştir. Bazı sorgular için orjinal süre ve kural öğrenilmiş süre eşit olarak gözükmemektedir. Bu durum sorgunun zaten indeksli bir kural bulundurması ve öğrenilecek yeni kuralların sorguya fayda etmeyeceğinden kaynaklanmaktadır.

Çizelgeden anlaşılacağı üzere büyük boyutlu veritabanlarında indeks aramayı kullanmak diğer metodlara göre daha mantıklı olacaktır. Böylece zaten indeksli olan kurallar öğrenmeye dahil edilmeyeceğinden ve gereksiz kuralların öğrenilme süresinden de kazanç sağlanacağından öğrenme süresi daha kısa olacaktır.

## 8. SONUÇLAR

- 1) Veri madenciliği ve uygulama alanlarının incelenmesi.
- 2) Veri madenciliğinde kullanılan metodlardan bazılarının araştırılması.
- 3) Veri madenciliği uygulamalarının çalışma prensiplerinin araştırılması.
- 4) Veri tabanları üzerinde çalışan anlamsal sorgu eniyilemesinin araştırılması.
- 5) Siegel ve Shekhar metodlarının incelenmesi
- 6) Anlamsal sorgu eniyileme metodları ile uygulama geliştirilmesi.
- 7) Geliştirilen uygulama ile farklı metodların birbiriyle karşılaştırılması.
- 8) Öğrenilen kurallar ile sorgu iyileştirmelerinin sağlanması.
- 9) Metod bazlı iyileştirmelerin süre tabanlı karşılaştırılması.

## KAYNAKLAR

- Alpaydın E. (2000), Zeki veri madenciliği; Ham veriden altın bilgiye ulaşmanın yöntemleri,
- Codd E.F. (1970), A. relational model of data for large shared data banks, Comm. of the ACM, 13:377-387.
- Dilly, R. (1995), Data Mining: An Introduction.
- Elmasri R. ve Navathe S.B. (1994), Fundamentals of Database Systems, The Benjamin/Cummings Publishing Company, 1994
- Gökçe B.(2003), “Visual C++’Ta Kurallara Dayalı Oracle Veri Tabanlı Sorgulamalar”, YTÜ, Fen Bilimleri En.
- Korth H.F. ve Silberschartz A.(1991), Database System Concepts. McGraw-Hill Book Company.
- Sayli A. ve Lowden B.G.T. (1996), The use of statistics in semantic query optimization. Thirteenth European Meeting on Cybernetics and Systems Research, Vienna, 991-996
- Sayli A. ve Lowden B.G.T. (1997), A fast transformation method to semantic query optimization, International Database Engineering and Application Symposium, Canada.
- Siegel M.D. (1988), Automatic rule derivation for semantic query optimization, 2. International Conference on Expert Database Systems, 669-698.
- Shapiro J.R. (2001), SQL Server 2000 The Complete Reference
- Shekhar, S., Hamidzadeh, B., Kohli, A., Coyle, M. (1993), Learning Transformation Rules for Semantic Query Optimization: A Data-Driven Approach. IEEE, 949-964

### İnternet Kaynakları:

- [1] [www.gartner.com](http://www.gartner.com)
- [2] [www.statsoft.com](http://www.statsoft.com)
- [3] [www.dtreg.com](http://www.dtreg.com)
- [4] [www.salford-systems.com](http://www.salford-systems.com)
- [5] [tr.wikipedia.org](http://tr.wikipedia.org)

**EKLER**

- Ek 1 Araba Tablosu veri bilgileri (İlk 100 Kayıtlık)
- Ek 2 Araba Tablosu için sorgu metodu ile öğrenilen kurallar (İlk 50 kural)
- Ek 3 Araba Tablosu için tam arama metoduyla öğrenilen kurallar (İlk 50 kural)
- Ek 4 Araba Tablosu için indeksli arama metoduyla öğrenilen kurallar (İlk 50 kural)
- Ek 5 Uygulama algoritması kod örneği

# Ek 1 Araba Tablosu veri bilgileri (İlk 100 Kayıtlık)

Çizelge Ek 1.1 Araba Tablosu veri bilgileri (İlk 100 Kayıtlık)

| Customer_id | Customer vehicle_id | Mernis_no   | education  | gender | Marital_status | Child Count | Married Year | age | Corporation sector     | Occupation status | Corporation position                          | Country model definition | occupation                                         | ID | Top model definition |
|-------------|---------------------|-------------|------------|--------|----------------|-------------|--------------|-----|------------------------|-------------------|-----------------------------------------------|--------------------------|----------------------------------------------------|----|----------------------|
| 157837      | 12178               | 13192763080 | Üniversite | Bay    | Evli           | 2           | 9            | 31  | EĞİTİM                 | employee          | MÜDÜR / MÜDÜR YRD / YÖNETİCİ                  | GOLF GL                  | AKADEMİSYEN / ÖĞR.GÖR EVLİSİ / ARAŞTIRMA GÖREVLİSİ | 1  | GOLF                 |
| 156945      | 13431               | 12544889720 | Lisansüstü | Bay    | Evli           | 2           | 24           | 55  | MİMARLIK / MÜHENDİSLİK | employee          | MÜDÜR / MÜDÜR YRD / YÖNETİCİ                  | AUDI A4                  | MÜHENDİS                                           | 2  | AUDI A4              |
| 145375      | 6731                | 34444526338 | Lise       | Bay    | Evli           | 0           | 0            | 41  | OTOMOTİV / YEDEK PARÇA | employee          | FİRMA SAHİBİ / ORTAĞI                         | A4 LIM V62.8L41 28EGII   | OTO. BAYISI / OTO. GALERİSİ                        | 3  | AUDI A4              |
| 147134      | 22824               | 54130596766 | Üniversite | Bay    | Bekar          | 0           | 0            | 28  | DENİZCİLİK             | employee          | FİRMA TEMSİLCİSİ / YETKİLİSİ / SATIN ALMACISI | AUDI A4 QUATTRO          | GÜMRÜK KOMİSYONCUSU                                | 4  | AUDI A4              |
| 157888      | 12194               | 12389087008 | Önlisans   | Bayan  | Evli           | 1           | 3            | 32  | DEVLET / KAMU          | employee          | MEMUR / KAMU / DEVLET GÖREVLİSİ               | GOLF CL                  | MİMAR / İÇ MİMAR                                   | 5  | GOLF                 |
| 148317      | 9095                | 13240017042 | Lise       | Bay    | Evli           | 0           | 0            | 31  | SANAYİ / ENDÜSTRİ      | employee          | ASİSTAN / SEKRETER / YARDIMCI                 | GOLF GL TDI              | TEKNİSYEN / MEKANİKER / TEKNİKER                   | 6  | GOLF                 |
| 156290      | 11389               | 21343893654 | Lise       | Bay    | Evli           | 0           | 2            | 34  | GIDA-ÜRETİM            | employee          | YETKİLİ / SORUMLU / UZMAN                     | POLO 1.6                 | DİĞER                                              | 7  | POLO                 |
| 154287      | 26508               | 21343893654 | Lise       | Bay    | Evli           | 0           | 3            | 34  | GIDA-ÜRETİM            | employee          | YETKİLİ / SORUMLU / UZMAN                     | POLO 1.6                 | DİĞER                                              | 8  | POLO                 |
| 141720      | 22213               | 21343893654 | Önlisans   | Bay    | Evli           | 0           | 2            | 34  | GIDA-ÜRETİM            | employee          | YETKİLİ / SORUMLU / UZMAN                     | POLO 1.6                 | DİĞER                                              | 9  | POLO                 |
| 174420      | 21186               | 23023560078 | Lise       | Bay    | Evli           | 2           | 29           | 50  | NULL                   | unemployed        | NULL                                          | POLO 1.6                 | NULL                                               | 10 | POLO                 |
| 173252      | 21942               | 16847349226 | Üniversite | Bayan  | Bekar          | 0           | 0            | 52  | NULL                   | unemployed        | NULL                                          | POLO 1.6                 | NULL                                               | 11 | POLO                 |
| 171159      | 19811               | 19777593042 | Önlisans   | Bay    | Evli           | 2           | 35           | 55  | SERBEST TİCARET        | employee          | DİĞER                                         | POLO 1.6                 | SERBEST                                            | 12 | POLO                 |
| 180070      | 24933               | 24245077900 | Doktora    | Bay    | Evli           | 0           | 10           | 35  | ASKERİ / EMNİYET       | employee          | SUBAY / ASTSUB                                | POLO 1.6                 | ASKER (SUBAY /                                     | 13 | POLO                 |

|        |       |                 |                |           |           |   |    |    |                                                           |                |                                                                       |                               |                                                                      |    |         |
|--------|-------|-----------------|----------------|-----------|-----------|---|----|----|-----------------------------------------------------------|----------------|-----------------------------------------------------------------------|-------------------------------|----------------------------------------------------------------------|----|---------|
|        |       |                 |                |           |           |   |    |    | T                                                         |                | AY                                                                    |                               | ASTSUBA<br>Y)                                                        |    |         |
| 160080 | 13146 | 250615<br>17816 | Ortao<br>kul   | Bay<br>an | Evli      | 1 | 6  | 36 | NULL                                                      | unemp<br>loyed | NULL                                                                  | POLO 1.6                      | NULL                                                                 | 14 | POLO    |
| 173635 | 25700 | 225616<br>24716 | Ortao<br>kul   | Bay       | Evli      | 2 | 20 | 41 | ASKERİ /<br>EMNİYE<br>T                                   | emplo<br>yee   | DİĞER                                                                 | AUDI A4<br>1.8 125<br>hp Oto. | İŞÇİ                                                                 | 15 | AUDI A4 |
| 173666 | 20759 | 225616<br>24716 | Ortao<br>kul   | Bay       | Evli      | 2 | 20 | 41 | ASKERİ /<br>EMNİYE<br>T                                   | emplo<br>yee   | YÖNETM<br>EN /<br>ŞEF                                                 | AUDI A4<br>1.8 125<br>hp Oto. | İŞÇİ                                                                 | 16 | AUDI A4 |
| 166737 | 16976 | 504167<br>00834 | İlkok<br>ul    | Bay       | Evli      | 5 | 30 | 50 | SERBEST<br>TİCARET                                        | emplo<br>yee   | FİRMA /<br>YÖNETİ<br>Cİ<br>ŞÖFÖRÜ                                     | AUDI A4<br>1.6                | EMEKLİ                                                               | 17 | AUDI A4 |
| 150936 | 26636 | 279251<br>93198 | Lise           | Bay       | Evli      | 0 | 4  | 29 | HALKLA<br>İLİŞKİLE<br>R /<br>REKLAM<br>/<br>PAZARL<br>AMA | emplo<br>yee   | FİRMA<br>SAHİBİ /<br>ORTAĞI                                           | AUDI A4<br>1.8 TQ             | BİLİŞİMC<br>İ /<br>BİLGİSAY<br>AR                                    | 18 | AUDI A4 |
| 131669 | 322   | 318370<br>39598 | Ünive<br>rsite | Bay       | Evli      | 1 | 26 | 53 | GIDA-<br>ÜRETİM                                           | emplo<br>yee   | MÜDÜR<br>/<br>MÜDÜR<br>YRD /<br>YÖNETİ<br>Cİ                          | A4 LİM<br>1.8 T               | TOPTANC<br>I                                                         | 19 | AUDI A4 |
| 162976 | 14598 | 183112<br>56276 | Lise           | Bay<br>an | Evli      | 0 | 0  | 23 | NULL                                                      | unemp<br>loyed | NULL                                                                  | A4 LİM<br>1.8 T               | NULL                                                                 | 20 | AUDI A4 |
| 153978 | 26512 | 375073<br>52912 | İlkok<br>ul    | Bay       | Evli      | 6 | 43 | 66 | TARIM /<br>ORMAN<br>/<br>HAYVAN<br>CILIK                  | emplo<br>yee   | FİRMA<br>SAHİBİ /<br>ORTAĞI                                           | CARAVE<br>LLE GL<br>2.5 S/W   | ESNAF                                                                | 21 | T4      |
| 155023 | 15381 | 548025<br>50256 | Lise           | Bay       | Evli      | 1 | 11 | 36 | HIZLI<br>TÜKETİ<br>M                                      | emplo<br>yee   | FİRMA<br>TEMSİLC<br>İŞİ /<br>YETKİLİ<br>Sİ /<br>SATIN<br>ALMACI<br>Sİ | GOLF<br>GTI/4<br>KAPI         | İŞÇİ                                                                 | 22 | GOLF    |
| 184685 | 27622 | 474553<br>29188 | Ünive<br>rsite | Bay<br>an | Evli      | 2 | 16 | 41 | NULL                                                      | unemp<br>loyed | NULL                                                                  | GOLF<br>GL AUT.               | NULL                                                                 | 23 | GOLF    |
| 183649 | 26762 | 228013<br>81072 | Ünive<br>rsite | Bay<br>an | Evli      | 0 | 28 | 49 | NULL                                                      | unemp<br>loyed | NULL                                                                  | GOLF CL<br>1.6                | NULL                                                                 | 24 | GOLF    |
| 131650 | 310   | 395083<br>16870 | Ünive<br>rsite | Bay       | Bek<br>ar | 0 | 0  | 25 | NULL                                                      | unemp<br>loyed | NULL                                                                  | GOLF CL                       | NULL                                                                 | 25 | GOLF    |
| 131756 | 22829 | 450973<br>95776 | Ünive<br>rsite | Bay       | Evli      | 2 | 21 | 58 | BANKA /<br>FİNANS /<br>SİGORTA<br>/<br>YATIRIM            | emplo<br>yee   | FİRMA<br>SAHİBİ /<br>ORTAĞI                                           | VENTO<br>GL 1.6               | İŞLETME<br>Cİ /<br>İMALATÇ<br>I (KOBİ -<br>250<br>ÇALIŞAN<br>DAN AZ) | 26 | VENTO   |
| 168028 | 17666 | 400011<br>26958 | Lise           | Bay       | Evli      | 2 | 15 | 37 | HAVACI<br>LIK                                             | emplo<br>yee   | GENEL<br>MÜDÜR<br>YRD /<br>KOORDİ<br>NATÖR /<br>DİREKT<br>ÖR          | VENTO<br>GL 1.8               | TEKNİSY<br>EN /<br>MEKANİ<br>KER /<br>TEKNİKE<br>R                   | 27 | VENTO   |
| 183487 | 26664 | 400011<br>26958 | Lise           | Bay       | Evli      | 2 | 15 | 37 | HAVACI<br>LIK                                             | emplo<br>yee   | MÜDÜR<br>/<br>MÜDÜR<br>YRD /<br>YÖNETİ<br>Cİ                          | VENTO<br>GL 1.8               | NAKLIYE<br>Cİ /<br>LOJİSTİK<br>/<br>TAŞIMAC<br>ILIK                  | 28 | VENTO   |

|        |       |                 |                |           |                  |   |    |    |                                                |                |                                                                              |                            |                                                                        |    |      |
|--------|-------|-----------------|----------------|-----------|------------------|---|----|----|------------------------------------------------|----------------|------------------------------------------------------------------------------|----------------------------|------------------------------------------------------------------------|----|------|
| 166185 | 16543 | 267678<br>76290 | Ünive<br>rsite | Bay<br>an | Evli             | 1 | 11 | 37 | DEVLET<br>/ KAMU                               | emplo<br>yee   | MEMUR /<br>KAMU /<br>DEVLET<br>GÖREVLİ<br>İŞİ                                | POLO 1.4<br>4 KAPI         | POLİS                                                                  | 29 | POLO |
| 131246 | 111   | 532092<br>83484 | Ünive<br>rsite | Bay<br>an | Boş<br>an<br>mış | 0 | 0  | 42 | NULL                                           | unemp<br>loyed | NULL                                                                         | POLO 1.3<br>4 KAPI         | NULL                                                                   | 30 | POLO |
| 164323 | 15219 | 133930<br>76316 | Lise           | Bay       | Evli             | 2 | 24 | 51 | NULL                                           | unemp<br>loyed | NULL                                                                         | POLO 1.6                   | NULL                                                                   | 31 | POLO |
| 183430 | 26601 | 351640<br>65336 | Ünive<br>rsite | Bay<br>an | Bek<br>ar        | 0 | 0  | 43 | İNŞAAT /<br>ÇİMENT<br>O                        | emplo<br>yee   | MÜDÜR<br>/<br>MÜDÜR<br>YRD /<br>YÖNETİ<br>Cİ                                 | POLO 1.6                   | MÜHENDİS                                                               | 32 | POLO |
| 133887 | 1386  | 351981<br>58202 | Lise           | Bay       | Bek<br>ar        | 0 | 0  | 26 | ÜRETİM<br>/<br>SANAYİ<br>(DİĞER)               | emplo<br>yee   | DİĞER                                                                        | POLO 1.3<br>4 KAPI         | İŞÇİ                                                                   | 33 | POLO |
| 132883 | 880   | 265991<br>80268 | Ünive<br>rsite | Bay       | Evli             | 2 | 37 | 62 | NULL                                           | unemp<br>loyed | NULL                                                                         | POLO 1.3<br>4 KAPI         | NULL                                                                   | 34 | POLO |
| 133861 | 1380  | 351981<br>58202 | Lise           | Bay       | Bek<br>ar        | 0 | 0  | 26 | ÜRETİM<br>/<br>SANAYİ<br>(DİĞER)               | emplo<br>yee   | DİĞER                                                                        | POLO 1.3<br>4 KAPI         | İŞÇİ                                                                   | 35 | POLO |
| 173569 | 20700 | 182446<br>54958 | Dokto<br>ra    | Bay       | Evli             | 2 | 10 | 39 | EĞİTİM                                         | emplo<br>yee   | DOKTOR<br>/<br>DOÇENT<br>/ YRD.<br>DOÇENT<br>(<br>AKADE<br>MİK /<br>SAĞLIK ) | POLO 1.6                   | AKADEMİSYEN /<br>ÖĞR.GÖR<br>EVLİSİ /<br>ARAŞTIR<br>MA<br>GÖREVLİ<br>Sİ | 36 | POLO |
| 184997 | 27884 | 419745<br>11956 | İlkok<br>ul    | Bay<br>an | Evli             | 3 | 0  | 43 | NULL                                           | unemp<br>loyed | NULL                                                                         | POLO 1.6                   | NULL                                                                   | 37 | POLO |
| 133118 | 23708 | 345317<br>87222 | Lise           | Bay<br>an | Evli             | 2 | 32 | 54 | NULL                                           | unemp<br>loyed | NULL                                                                         | POLO 1.6<br>AUT. 4<br>KAPI | NULL                                                                   | 38 | POLO |
| 170294 | 18908 | 184840<br>51012 | Ünive<br>rsite | Bay       | Bek<br>ar        | 0 | 0  | 34 | PETROL<br>/<br>PLASTİK                         | emplo<br>yee   | YETKİLİ<br>/<br>SORUML<br>U /<br>UZMAN                                       | POLO 1.6<br>AUT. 4<br>KAPI | MUHASEBECİ /<br>MALİ<br>MÜŞAVİR<br>/<br>DENETÇİ<br>/<br>MÜFETTİŞ       | 39 | POLO |
| 158620 | 12535 | 305302<br>20814 | Ünive<br>rsite | Bay       | Evli             | 1 | 0  | 33 | OTOMOTİV /<br>YEDEK<br>PARÇA                   | emplo<br>yee   | DİĞER                                                                        | POLO 1.6<br>AUT. 4<br>KAPI | MÜHENDİS                                                               | 40 | POLO |
| 135644 | 2314  | 487872<br>72050 | Lise           | Bay<br>an | Evli             | 1 | 25 | 51 | NULL                                           | unemp<br>loyed | NULL                                                                         | POLO 1.6<br>AUT. 4<br>KAPI | NULL                                                                   | 41 | POLO |
| 134465 | 1702  | 343307<br>16782 | Önlis<br>ans   | Bay       | Evli             | 0 | 3  | 31 | BANKA /<br>FİNANS /<br>SIGORTA<br>/<br>YATIRIM | emplo<br>yee   | DİĞER                                                                        | POLO 1.6<br>AUT. 4<br>KAPI | AVUKAT<br>/<br>HUKUKÇU /<br>HAKİM /<br>SAVCI                           | 42 | POLO |
| 168942 | 22047 | 230516<br>24902 | Ünive<br>rsite | Bay       | Evli             | 1 | 8  | 36 | DEVLET<br>/ KAMU                               | emplo<br>yee   | MÜDÜR<br>/<br>MÜDÜR<br>YRD /<br>YÖNETİ<br>Cİ                                 | POLO 1.6<br>AUT. 4<br>KAPI | HALKLA<br>İLİŞKİLER /<br>REKLAM<br>CI/<br>PAZARLA<br>MA /<br>İLETİŞİM  | 43 | POLO |

|        |       |                 |                |           |                  |   |    |    |                  |                |                                                                              |                                        |                                                                            |    |                 |
|--------|-------|-----------------|----------------|-----------|------------------|---|----|----|------------------|----------------|------------------------------------------------------------------------------|----------------------------------------|----------------------------------------------------------------------------|----|-----------------|
| 134245 | 1598  | 358246<br>91660 | Ünive<br>rsite | Bay<br>an | Evli             | 1 | 12 | 37 | NULL             | unemp<br>loyed | NULL                                                                         | POLO<br>OPEN<br>AIR/4KA<br>PI          | NULL                                                                       | 44 | POLO            |
| 173598 | 20714 | 316393<br>04122 | İlkok<br>ul    | Bay       | Bek<br>ar        | 0 | 0  | 75 | NULL             | unemp<br>loyed | NULL                                                                         | POLO 1.4<br>2 KAPI                     | NULL                                                                       | 45 | POLO            |
| 174250 | 21098 | 264341<br>56090 | Ünive<br>rsite | Bay<br>an | Bek<br>ar        | 0 | 0  | 28 | SAĞLIK /<br>İLAÇ | emplo<br>yee   | DOKTOR<br>/<br>DOÇENT<br>/ YRD.<br>DOÇENT<br>(<br>AKADE<br>MİK /<br>SAĞLIK ) | POLO<br>CLASSIC                        | DOKTOR /<br>DIŞ<br>HEKİMİ /<br>VETERİN<br>ER                               | 46 | POLO<br>CLASSIC |
| 159929 | 13089 | 485624<br>26388 | Lise           | Bay<br>an | Evli             | 1 | 5  | 27 | NULL             | unemp<br>loyed | NULL                                                                         | POLO<br>CLASSIC                        | NULL                                                                       | 47 | POLO<br>CLASSIC |
| 155195 | 10846 | 156204<br>18074 | Lise           | Bay<br>an | Evli             | 1 | 3  | 32 | NULL             | unemp<br>loyed | NULL                                                                         | POLO<br>CLASSIC                        | NULL                                                                       | 48 | POLO<br>CLASSIC |
| 158723 | 26709 | 104516<br>12202 | Ünive<br>rsite | Bay<br>an | Bek<br>ar        | 0 | 0  | 30 | EĞİTİM           | emplo<br>yee   | ASİSTAN<br>/<br>SEKRET<br>ER /<br>YARDIM<br>CI                               | POLO<br>CLASSIC                        | AKADEM<br>İSYEN /<br>ÖĞR.GÖR<br>EVLİSİ /<br>ARAŞTIR<br>MA<br>GÖREVLİ<br>Sİ | 49 | POLO<br>CLASSIC |
| 184930 | 27835 | 439934<br>45090 | Ünive<br>rsite | Bay<br>an | Boş<br>an<br>mış | 1 | 0  | 58 | NULL             | unemp<br>loyed | NULL                                                                         | A4 LİM<br>1.6<br>AUTO                  | NULL                                                                       | 50 | AUDI A4         |
| 173173 | 20477 | 486797<br>76116 | Lisan<br>süstü | Bay       | Evli             | 0 | 0  | 32 | EĞİTİM           | emplo<br>yee   | MÜDÜR<br>/<br>MÜDÜR<br>YRD /<br>YÖNETİ<br>Cİ                                 | A4 LİM<br>1.8/M5                       | ÖĞRETM<br>EN /<br>EĞİTMEN                                                  | 51 | AUDI A4         |
| 163079 | 14648 | 108414<br>99680 | Dokto<br>ra    | Bay<br>an | Evli             | 0 | 4  | 33 | NULL             | unemp<br>loyed | NULL                                                                         | A4 LİM<br>1.8/M5                       | NULL                                                                       | 52 | AUDI A4         |
| 149690 | 8710  | 463753<br>39796 | Lise           | Bay       | Evli             | 1 | 30 | 71 | NULL             | unemp<br>loyed | NULL                                                                         | A4 LİM<br>1.8/M5                       | NULL                                                                       | 53 | AUDI A4         |
| 154793 | 14268 | 348347<br>78330 | Lise           | Bay       | Evli             | 3 | 22 | 49 | HİZMET           | emplo<br>yee   | DİĞER                                                                        | A4 LİM<br>1.8<br>AUTO                  | TURİZMC<br>İ                                                               | 54 | AUDI A4         |
| 138870 | 3805  | 264015<br>17398 | Ünive<br>rsite | Bay       | Evli             | 4 | 22 | 47 | NULL             | unemp<br>loyed | NULL                                                                         | A4 LİM<br>1.8 T                        | NULL                                                                       | 55 | AUDI A4         |
| 177544 | 28038 | 259461<br>84874 | Ortao<br>kul   | Bay       | Evli             | 2 | 12 | 36 | MOBİLY<br>ACILIK | emplo<br>yee   | FİRMA<br>SAHİBİ /<br>ORTAĞI                                                  | A4 LİM<br>1.8 T                        | MOBİLYA<br>CI                                                              | 56 | AUDI A4         |
| 174837 | 21385 | 124756<br>65264 | Ünive<br>rsite | Bay       | Evli             | 1 | 7  | 27 | SAĞLIK /<br>İLAÇ | emplo<br>yee   | MÜDÜR<br>/<br>MÜDÜR<br>YRD /<br>YÖNETİ<br>Cİ                                 | AUDI A4<br>1.8<br>TURBO<br>QUATTR<br>O | ÜST<br>DÜZEY<br>BÜROKR<br>AT /<br>KAMU /<br>BELEDİ<br>YE /<br>EMNİYET      | 57 | AUDI A4         |
| 144899 | 6507  | 102105<br>66326 | Lise           | Bay       | Evli             | 2 | 34 | 62 | NULL             | unemp<br>loyed | NULL                                                                         | AUDI A4<br>1.8<br>TURBO<br>QUATTR<br>O | NULL                                                                       | 58 | AUDI A4         |

|        |       |                 |                |           |           |   |    |    |                                                   |              |                                                                       |                                         |                                                                          |    |                  |
|--------|-------|-----------------|----------------|-----------|-----------|---|----|----|---------------------------------------------------|--------------|-----------------------------------------------------------------------|-----------------------------------------|--------------------------------------------------------------------------|----|------------------|
| 159339 | 12795 | 150586<br>84928 | Lise           | Bay       | Evli      | 2 | 0  | 31 | TEKSTİL                                           | emplo<br>yee | MÜDÜR<br>/<br>MÜDÜR<br>YRD /<br>YÖNETİ<br>Cİ                          | A4 LİM<br>1.8 T                         | MUHASE<br>BECİ /<br>MALİ<br>MÜŞAVİR<br>/<br>DENETÇİ<br>/<br>MÜFETTİ<br>Ş | 59 | AUDI A4          |
| 142867 | 5592  | 280255<br>35246 | Lise           | Bay       | Evli      | 0 | 12 | 37 | MÜŞAVİ<br>RLİK /<br>DENETİ<br>M /<br>MUHASE<br>BE | emplo<br>yee | FİRMA<br>SAHİBİ /<br>ORTAĞI                                           | A4 LİM<br>1.8 T                         | MUHASE<br>BECİ /<br>MALİ<br>MÜŞAVİR<br>/<br>DENETÇİ<br>/<br>MÜFETTİ<br>Ş | 60 | AUDI A4          |
| 171802 | 19745 | 460127<br>35110 | Ünive<br>rsite | Bay       | Evli      | 1 | 0  | 44 | DEVLET<br>/<br>KAMU                               | emplo<br>yee | MEMUR /<br>KAMU /<br>DEVLET<br>GÖREVL<br>İŞİ                          | A4 LİM<br>1.8 T                         | TEKNİSY<br>EN /<br>MEKANİ<br>KER /<br>TEKNİKE<br>R                       | 61 | AUDI A4          |
| 180720 | 25351 | 358431<br>54924 | Lisan<br>üstü  | Bay<br>an | Bek<br>ar | 0 | 0  | 31 | SAĞLIK /<br>İLAÇ                                  | emplo<br>yee | YETKİLİ<br>/<br>SORUML<br>U /<br>UZMAN                                | A4 LİM<br>1.8 T                         | MUHASE<br>BECİ /<br>MALİ<br>MÜŞAVİR<br>/<br>DENETÇİ<br>/<br>MÜFETTİ<br>Ş | 62 | AUDI A4          |
| 136065 | 2495  | 514840<br>94016 | Ünive<br>rsite | Bay<br>an | Evli      | 2 | 14 | 37 | MİMARLI<br>K /<br>MÜHEND<br>İSLİK                 | emplo<br>yee | FİRMA<br>SAHİBİ /<br>ORTAĞI                                           | AUDI A4<br>1.8<br>TURBO<br>QUATTR<br>O  | MİMAR /<br>İÇ<br>MİMAR                                                   | 63 | AUDI A4          |
| 170694 | 19118 | 311958<br>58190 | Lise           | Bay       | Evli      | 0 | 0  | 37 | OTOMOT<br>İV /<br>YEDEK<br>PARÇA                  | emplo<br>yee | FİRMA<br>TEMSİLC<br>İŞİ /<br>YETKİLİ<br>Sİ /<br>SATIN<br>ALMACI<br>SI | A4 LİM<br>1.8 T<br>AUT                  | OTO.<br>TAMİR /<br>SERVİS /<br>YEDEK<br>PARÇA /<br>TEKNİK<br>DESTEK      | 64 | AUDI A4          |
| 166484 | 16742 | 343821<br>51692 | Ünive<br>rsite | Bay       | Evli      | 0 | 8  | 32 | DİĞER                                             | emplo<br>yee | MÜDÜR<br>/<br>MÜDÜR<br>YRD /<br>YÖNETİ<br>Cİ                          | A4 LİM<br>1.6<br>AUTO                   | MÜHEND<br>İS                                                             | 65 | AUDI A4          |
| 164978 | 15541 | 411371<br>22708 | Ünive<br>rsite | Bay       | Evli      | 1 | 8  | 31 | DEVLET<br>/<br>KAMU                               | emplo<br>yee | MEMUR /<br>KAMU /<br>DEVLET<br>GÖREVL<br>İŞİ                          | A4 LİM<br>1.8/M5                        | ÖĞRETM<br>EN /<br>EĞİTMEN                                                | 66 | AUDI A4          |
| 176316 | 22551 | 105267<br>74866 | Ünive<br>rsite | Bay       | Evli      | 2 | 10 | 42 | DEVLET<br>/<br>KAMU                               | emplo<br>yee | ÜST<br>DÜZEY<br>KAMU /<br>DEVLET<br>GÖREVL<br>İŞİ                     | AUDI A4<br>AVANT<br>1.8 150<br>hp Conv. | ÜST<br>DÜZEY<br>BÜROKR<br>AT /<br>KAMU /<br>BELEDİYE<br>/<br>EMNİYET     | 67 | AUDI A4<br>AVANT |
| 184445 | 27423 | 344931<br>77154 | Ünive<br>rsite | Bay<br>an | Bek<br>ar | 0 | 0  | 29 | EĞİTİM                                            | emplo<br>yee | DANIŞM<br>AN                                                          | A4 1.8<br>TQ<br>AVANT                   | PSİKOLO<br>G /<br>SOSYOLO<br>G /<br>SOSYAL<br>BİLİMLE<br>R               | 68 | AUDI A4<br>AVANT |

|        |       |                 |                |           |                  |   |    |    |                                  |                |                                                                       |                                         |                                                                            |    |         |
|--------|-------|-----------------|----------------|-----------|------------------|---|----|----|----------------------------------|----------------|-----------------------------------------------------------------------|-----------------------------------------|----------------------------------------------------------------------------|----|---------|
| 158026 | 12240 | 362511<br>91020 | Ortao<br>kul   | Bay       | Bek<br>ar        | 0 | 0  | 26 | TEKSTİL                          | emplye         | FİRMA<br>TEMSİLC<br>İŞİ /<br>YETKİLİ<br>Sİ /<br>SATIN<br>ALMACI<br>Sİ | AUDI A3<br>1.8<br>TURBO<br>AMBITİ<br>ON | TEKSTİL<br>Cİ                                                              | 69 | AUDI A3 |
| 136279 | 2618  | 467263<br>69946 | Ünive<br>rsite | Bay       | Evli             | 0 | 5  | 34 | SANAYİ<br>/<br>ENDÜST<br>Rİ      | emplye         | FİRMA<br>SAHİBİ /<br>ORTAĞI                                           | AUDI A3<br>1.6<br>AMBITİ<br>ON          | MÜHEND<br>İS                                                               | 70 | AUDI A3 |
| 131973 | 443   | 525641<br>35124 | Dokto<br>ra    | Bay<br>an | Evli             | 1 | 18 | 45 | GIDA-<br>ÜRETİM                  | emplye         | FİRMA<br>SAHİBİ /<br>ORTAĞI                                           | A3 1.6<br>AUT<br>AMBIEN<br>TE           | İŞLETME<br>Cİ /<br>İMALATÇ<br>I (KOBİ -<br>250<br>ÇALIŞAN<br>DAN AZ)       | 71 | AUDI A3 |
| 158877 | 12627 | 174670<br>57850 | Lise           | Bay       | Boş<br>an<br>mış | 0 | 0  | 48 | MOBİLY<br>ACILIK                 | emplye         | FİRMA<br>SAHİBİ /<br>ORTAĞI                                           | CARAVE<br>LLE 2.5                       | MOBİLYA<br>Cİ                                                              | 72 | T4      |
| 148040 | 7940  | 348131<br>51376 | Lise           | Bay<br>an | Evli             | 1 | 0  | 57 | NULL                             | unemp<br>loyed | NULL                                                                  | CARAVE<br>LLE 2.5                       | NULL                                                                       | 73 | T4      |
| 151555 | 9504  | 240597<br>29828 | Lise           | Bay       | Evli             | 3 | 2  | 47 | GIDA-<br>ÜRETİM                  | emplye         | FİRMA<br>TEMSİLC<br>İŞİ /<br>YETKİLİ<br>Sİ /<br>SATIN<br>ALMACI<br>Sİ | CARAVE<br>LLE 2.5                       | ESNAF                                                                      | 74 | T4      |
| 154203 | 10515 | 554536<br>71758 | Lise           | Bay       | Evli             | 3 | 27 | 52 | NULL                             | unemp<br>loyed | NULL                                                                  | CARAVE<br>LLE 2.5                       | NULL                                                                       | 75 | T4      |
| 143532 | 5903  | 181880<br>18342 | Lise           | Bay       | Evli             | 4 | 25 | 49 | DEVLET<br>/ KAMU                 | emplye         | MEMUR /<br>KAMU /<br>DEVLET<br>GÖREVL<br>İŞİ                          | CARAVE<br>LLE 2.5                       | ŞÖFÖR /<br>TAKSİCİ /<br>TRANSFE<br>R                                       | 76 | T4      |
| 177943 | 24574 | 273851<br>98788 | Lise           | Bay       | Evli             | 3 | 24 | 51 | ELEKTRİ<br>K /<br>ELEKTR<br>ONİK | emplye         | FİRMA<br>SAHİBİ /<br>ORTAĞI                                           | CARAVE<br>LLE 2.5                       | ESNAF                                                                      | 77 | T4      |
| 174996 | 21451 | 236442<br>57586 | İlkok<br>ul    | Bay       | Evli             | 0 | 33 | 57 | NULL                             | unemp<br>loyed | NULL                                                                  | CARAVE<br>LLE 2.5                       | NULL                                                                       | 78 | T4      |
| 159597 | 12954 | 397130<br>65914 | İlkok<br>ul    | Bay<br>an | Evli             | 4 | 32 | 47 | NULL                             | unemp<br>loyed | NULL                                                                  | CARAVE<br>LLE 2.5                       | NULL                                                                       | 79 | T4      |
| 154678 | 10673 | 405166<br>47924 | Ortao<br>kul   | Bay<br>an | Evli             | 2 | 24 | 49 | NULL                             | unemp<br>loyed | NULL                                                                  | CARAVE<br>LLE 2.5                       | NULL                                                                       | 80 | T4      |
| 162861 | 14543 | 134053<br>14764 | Lise           | Bay       | Evli             | 2 | 15 | 46 | ELEKTRİ<br>K /<br>ELEKTR<br>ONİK | emplye         | FİRMA<br>SAHİBİ /<br>ORTAĞI                                           | CARAVE<br>LLE 2.5                       | ESNAF                                                                      | 81 | T4      |
| 164994 | 15548 | 255112<br>32528 | İlkok<br>ul    | Bay       | Evli             | 3 | 2  | 51 | DIĞER                            | emplye         | YÖNETİ<br>M<br>KURULU<br>BAŞKAN<br>I /<br>ÜYESİ                       | CARAVE<br>LLE 2.5                       | SERBEST                                                                    | 82 | T4      |
| 142998 | 5672  | 360855<br>38008 | Ünive<br>rsite | Bay       | Evli             | 1 | 20 | 50 | HİZMET                           | emplye         | MÜDÜR<br>/<br>MÜDÜR<br>YRD /<br>YÖNETİ<br>Cİ                          | CARAVE<br>LLE 2.5                       | ORTA /<br>ALT<br>SEVİYE<br>KAMU<br>GÖREVLİ<br>/ MEMUR<br>/<br>BÜROKR<br>AT | 83 | T4      |

|        |       |                 |                |           |           |   |    |    |                                                           |                |                                              |                                   |                                                                            |    |       |
|--------|-------|-----------------|----------------|-----------|-----------|---|----|----|-----------------------------------------------------------|----------------|----------------------------------------------|-----------------------------------|----------------------------------------------------------------------------|----|-------|
| 148526 | 8161  | 245515<br>20060 | Lise           | Bay       | Evli      | 3 | 0  | 49 | DİĞER                                                     | emplyee        | FİRMA<br>SAHİBİ /<br>ORTAĞI                  | CARAVE<br>LLE 2.5                 | DİĞER                                                                      | 84 | T4    |
| 181213 | 25640 | 131276<br>58474 | Lise           | Bay       | Evli      | 2 | 2  | 36 | ÜRETİM<br>/<br>SANAYİ<br>(DİĞER)                          | emplyee        | FİRMA<br>SAHİBİ /<br>ORTAĞI                  | CARAVE<br>LLE 2.5                 | İŞLETME<br>Cİ /<br>İMALATÇ<br>I (KOBİ -<br>250<br>ÇALIŞAN<br>DAN AZ)       | 85 | T4    |
| 179301 | 24527 | 256759<br>10768 | Lise           | Bay       | Evli      | 2 | 11 | 39 | ELEKTRİK<br>/<br>ELEKTR<br>ONİK                           | emplyee        | FİRMA<br>SAHİBİ /<br>ORTAĞI                  | CADDY<br>1.9 SDI                  | ESNAF                                                                      | 86 | CADDY |
| 182312 | 26100 | 581624<br>52496 | Lise           | Bay       | Evli      | 0 | 10 | 39 | SERBEST<br>TİCARET                                        | emplyee        | FİRMA<br>SAHİBİ /<br>ORTAĞI                  | GOLF<br>CABRIO<br>LET 2.0<br>AUT. | ESNAF                                                                      | 87 | GOLF  |
| 180427 | 25182 | 339472<br>22518 | Ünive<br>rsite | Bay       | Evli      | 0 | 2  | 35 | TEKSTİL                                                   | emplyee        | YÖNETM<br>EN /<br>ŞEF                        | GOLF<br>CABRIO<br>LET 2.0<br>AUT. | DANIŞMA<br>N                                                               | 88 | GOLF  |
| 181255 | 25655 | 113844<br>90202 | Ünive<br>rsite | Bay       | Evli      | 0 | 4  | 34 | OTOMOT<br>İV /<br>YEDEK<br>PARÇA                          | emplyee        | MEMUR /<br>KAMU /<br>DEVLET<br>GÖREVL<br>İŞİ | GOLF<br>CABRIO<br>LET 2.0<br>AUT. | MÜHEND<br>İS                                                               | 89 | GOLF  |
| 178128 | 23908 | 101634<br>57726 | Önlis<br>ans   | Bay       | Evli      | 2 | 16 | 40 | EĞİTİM                                                    | emplyee        | DİĞER                                        | GOLF<br>CABRIO<br>LET 2.0<br>AUT. | ÖĞRETM<br>EN /<br>EĞİTMEN                                                  | 90 | GOLF  |
| 179606 | 24677 | 115371<br>53848 | İlkok<br>ul    | Bay       | Evli      | 2 | 17 | 57 | PERAKE<br>NDE /<br>MAĞAZA<br>CILIK                        | emplyee        | FİRMA<br>SAHİBİ /<br>ORTAĞI                  | GOLF<br>CABRIO<br>LET 2.0<br>AUT. | DİĞER                                                                      | 91 | GOLF  |
| 185042 | 27934 | 576880<br>53484 | Ünive<br>rsite | Bay       | Evli      | 0 | 13 | 43 | DAYANI<br>KLI<br>TÜKETİ<br>M                              | emplyee        | FİRMA<br>SAHİBİ /<br>ORTAĞI                  | GOLF<br>CABRIO<br>LET 2.0<br>AUT. | ESNAF                                                                      | 92 | GOLF  |
| 172637 | 20172 | 514302<br>04278 | Ünive<br>rsite | Bay<br>an | Evli      | 2 | 0  | 49 | NULL                                                      | unemp<br>loyed | NULL                                         | GOLF<br>CABRIO<br>LET 2.0<br>AUT. | NULL                                                                       | 93 | GOLF  |
| 156014 | 23128 | 263359<br>96238 | Ünive<br>rsite | Bay       | Evli      | 0 | 6  | 31 | DİĞER                                                     | emplyee        | DİĞER                                        | GOLF<br>CABRIO<br>LET 2.0<br>AUT. | MÜŞTERİ<br>İLİŞKİLE<br>Rİ                                                  | 94 | GOLF  |
| 160384 | 13315 | 557351<br>70788 | Lisan<br>üstü  | Bay       | Evli      | 0 | 0  | 29 | TELEKO<br>MİNİKAS<br>YON /<br>İLETİŞİM                    | emplyee        | DİĞER                                        | GOLF<br>CABRIO<br>LET 2.0<br>AUT. | GAZETEC<br>İ / BASIN /<br>YAYIN /<br>MUHABİ<br>R /<br>RADYO<br>DJ          | 95 | GOLF  |
| 162368 | 26618 | 299754<br>82130 | Lise           | Bay       | Bek<br>ar | 0 | 0  | 40 | HALKLA<br>İLİŞKİLE<br>R /<br>REKLAM<br>/<br>PAZARL<br>AMA | emplyee        | YETKİLİ<br>/<br>SORUML<br>U /<br>UZMAN       | GOLF<br>CABRIO<br>LET 2.0<br>AUT. | HALKLA<br>İLİŞKİLE<br>R /<br>REKLAM<br>CI /<br>PAZARLA<br>MA /<br>İLETİŞİM | 96 | GOLF  |
| 137258 | 3025  | 450949<br>67532 | Ünive<br>rsite | Bay       | Evli      | 1 | 14 | 46 | PETROL<br>/<br>PLASTİK                                    | emplyee        | YÖNETM<br>EN /<br>ŞEF                        | GOLF<br>CABRIO<br>LET 2.0<br>AUT. | MUHASE<br>BECİ /<br>MALİ<br>MÜŞAVİR<br>/                                   | 97 | GOLF  |

|        |       |             |            |     |      |   |    |    |                  |          |                                 |                         |                                      |     |      |
|--------|-------|-------------|------------|-----|------|---|----|----|------------------|----------|---------------------------------|-------------------------|--------------------------------------|-----|------|
|        |       |             |            |     |      |   |    |    |                  |          |                                 |                         | DENETÇİ / MÜFETTİŞ                   |     |      |
| 140472 | 27870 | 32822257652 | Üniversite | Bay | Evli | 1 | 19 | 42 | BASIN / MEDYA    | empleado | DİĞER                           | GOLF CABRIOLET 2.0 AUT. | TASARIMCI (GRAFİK, MODA)             | 98  | GOLF |
| 136110 | 2534  | 17302078378 | Üniversite | Bay | Evli | 2 | 9  | 34 | ASKERİ / EMNİYET | empleado | MEMUR / KAMU / DEVLET GÖREVLİSİ | GOLF CABRIOLET 2.0 AUT. | POLİS                                | 99  | GOLF |
| 134812 | 26026 | 55714336150 | Lise       | Bay | Evli | 1 | 2  | 54 | TEKSTİL          | empleado | FİRMA SAHİBİ / ORTAĞI           | GOLF CABRIOLET 2.0 AUT. | İŞLETMECİ / İMALATÇI (BÜYÜK İŞLETME) | 100 | GOLF |

## Ek 2 Araba Tablosu için sorgu metodu ile öğrenilen kurallar (İlk 50 kural)

Çizelge Ek 2.1 Araba Tablosu için sorgu metodu ile öğrenilen kurallar (İlk 50 kural)

| ID | Left column | Operator | Left Value | Right column            | Right Operator | Right Value                   |
|----|-------------|----------|------------|-------------------------|----------------|-------------------------------|
| 1  | ID          | =        | 2225       | customer_id             | =              | 152769                        |
| 2  | ID          | =        | 2225       | customer_vehicle_id     | =              | 10015                         |
| 3  | ID          | =        | 2225       | mernis_no               | =              | 21824183380                   |
| 4  | ID          | =        | 2225       | education               | =              | Üniversite                    |
| 5  | ID          | =        | 2225       | gender                  | =              | Bayan                         |
| 6  | ID          | =        | 2225       | marital_status          | =              | Evli                          |
| 7  | ID          | =        | 2225       | child_count             | =              | 0                             |
| 8  | ID          | =        | 2225       | married_year            | =              | 4                             |
| 9  | ID          | =        | 2225       | age                     | =              | 34                            |
| 10 | ID          | =        | 2225       | corporation_sector      | =              | DEVLET / KAMU                 |
| 11 | ID          | =        | 2225       | occupation_status       | =              | employee                      |
| 12 | ID          | =        | 2225       | corporation_position    | =              | ASİSTAN / SEKRETER / YARDIMCI |
| 13 | ID          | =        | 2225       | countrymodel_definition | =              | A3 1.8 AUT AMBIENTE           |
| 14 | ID          | =        | 2225       | occupation              | =              | SEKRETER / ASİSTAN            |
| 15 | ID          | =        | 2225       | ID                      | =              | 2225                          |
| 16 | ID          | =        | 2225       | top_model_definition    | =              | AUDI A3                       |
| 17 | ID          | <=       | 3000       | customer_id             | <=             | 185098                        |
| 18 | ID          | <=       | 3000       | customer_id             | >=             | 146995                        |
| 19 | ID          | <=       | 3000       | customer_vehicle_id     | <=             | 28054                         |
| 20 | ID          | <=       | 3000       | customer_vehicle_id     | >=             | 7461                          |
| 21 | ID          | <=       | 3000       | mernis_no               | <=             | 74941009408                   |
| 22 | ID          | <=       | 3000       | mernis_no               | >=             | 50497749296                   |
| 23 | ID          | <=       | 3000       | education               | <=             | Üniversite                    |
| 24 | ID          | <=       | 3000       | education               | >=             | İlkokul                       |
| 25 | ID          | <=       | 3000       | gender                  | <=             | Bayan                         |
| 26 | ID          | <=       | 3000       | gender                  | >=             | Bay                           |

|    |    |    |       |                         |    |                                                       |
|----|----|----|-------|-------------------------|----|-------------------------------------------------------|
| 27 | ID | <= | 3000  | marital_status          | <= | Evli                                                  |
| 28 | ID | <= | 3000  | marital_status          | >= | Evli                                                  |
| 29 | ID | <= | 3000  | child_count             | <= | 10                                                    |
| 30 | ID | <= | 3000  | child_count             | >= | 2                                                     |
| 31 | ID | <= | 3000  | married_year            | <= | 62                                                    |
| 32 | ID | <= | 3000  | married_year            | >= | 2                                                     |
| 33 | ID | <= | 3000  | age                     | <= | 84                                                    |
| 34 | ID | <= | 3000  | age                     | >= | 38                                                    |
| 35 | ID | <= | 3000  | corporation_sector      | <= | VAKIF / DERNEK / SENDİKA                              |
| 36 | ID | <= | 3000  | corporation_sector      | >= | HİZMET                                                |
| 37 | ID | <= | 3000  | occupation_status       | <= | unemployed                                            |
| 38 | ID | <= | 3000  | occupation_status       | >= | employee                                              |
| 39 | ID | <= | 3000  | corporation_position    | <= | YÖNETMEN / ŞEF                                        |
| 40 | ID | <= | 3000  | corporation_position    | >= | YETKİLİ / SORUMLU / UZMAN                             |
| 41 | ID | <= | 3000  | countrymodel_definition | <= | VENTO GL 1.8                                          |
| 42 | ID | <= | 3000  | countrymodel_definition | >= | BORA 1.6 TRENDLINE                                    |
| 43 | ID | <= | 3000  | occupation              | <= | YAZAR / RESSAM / TİYATROCU /<br>BESTECİ / ŞARKICI VS. |
| 44 | ID | <= | 3000  | occupation              | >= | ESNAF                                                 |
| 45 | ID | <= | 3000  | top_model_definition    | <= | VENTO                                                 |
| 46 | ID | <= | 3000  | top_model_definition    | >= | BORA                                                  |
| 47 | ID | >= | 20000 | customer_id             | <= | 185106                                                |
| 48 | ID | >= | 20000 | customer_id             | >= | 167351                                                |
| 49 | ID | >= | 20000 | customer_vehicle_id     | <= | 28060                                                 |
| 50 | ID | >= | 20000 | customer_vehicle_id     | >= | 17244                                                 |

### Ek 3 Araba Tablosu için tam arama metoduyla öğrenilen kurallar (İlk 50 kural)

Çizelge Ek 3.1 Araba Tablosu için tam arama metoduyla öğrenilen kurallar (İlk 50 kural)

| ID | Left column | Operator | Left Value | Right column            | Right Operator | Right Value |
|----|-------------|----------|------------|-------------------------|----------------|-------------|
| 1  | customer_id | =        | 170735     | customer_id             | =              | 170735      |
| 2  | customer_id | =        | 170735     | customer_vehicle_id     | =              | 19146       |
| 3  | customer_id | =        | 170735     | mernis_no               | =              | 13237078742 |
| 4  | customer_id | =        | 170735     | education               | =              | Lisansüstü  |
| 5  | customer_id | =        | 170735     | gender                  | =              | Bayan       |
| 6  | customer_id | =        | 170735     | marital_status          | =              | Bekar       |
| 7  | customer_id | =        | 170735     | child_count             | =              | 0           |
| 8  | customer_id | =        | 170735     | married_year            | =              | 0           |
| 9  | customer_id | =        | 170735     | age                     | =              | 58          |
| 10 | customer_id | =        | 170735     | corporation_sector      | =              | NULL        |
| 11 | customer_id | =        | 170735     | occupation_status       | =              | unemployed  |
| 12 | customer_id | =        | 170735     | corporation_position    | =              | NULL        |
| 13 | customer_id | =        | 170735     | countrymodel_definition | =              | POLO 1.6    |

|    |             |   |        |                         |   |                                    |
|----|-------------|---|--------|-------------------------|---|------------------------------------|
| 14 | customer_id | = | 170735 | occupation              | = | NULL                               |
| 15 | customer_id | = | 170735 | ID                      | = | 952                                |
| 16 | customer_id | = | 170735 | top_model_definition    | = | POLO                               |
| 17 | customer_id | = | 137771 | customer_id             | = | 137771                             |
| 18 | customer_id | = | 137771 | customer_vehicle_id     | = | 3273                               |
| 19 | customer_id | = | 137771 | mernis_no               | = | 10585263752                        |
| 20 | customer_id | = | 137771 | education               | = | Lise                               |
| 21 | customer_id | = | 137771 | gender                  | = | Bay                                |
| 22 | customer_id | = | 137771 | marital_status          | = | Evli                               |
| 23 | customer_id | = | 137771 | child_count             | = | 1                                  |
| 24 | customer_id | = | 137771 | married_year            | = | 37                                 |
| 25 | customer_id | = | 137771 | age                     | = | 56                                 |
| 26 | customer_id | = | 137771 | corporation_sector      | = | NULL                               |
| 27 | customer_id | = | 137771 | occupation_status       | = | unemployed                         |
| 28 | customer_id | = | 137771 | corporation_position    | = | NULL                               |
| 29 | customer_id | = | 137771 | countrymodel_definition | = | CORDOBA SE 1.6I                    |
| 30 | customer_id | = | 137771 | occupation              | = | NULL                               |
| 31 | customer_id | = | 137771 | ID                      | = | 2626                               |
| 32 | customer_id | = | 137771 | top_model_definition    | = | CORDOBA                            |
| 33 | customer_id | = | 158332 | customer_id             | = | 158332                             |
| 34 | customer_id | = | 158332 | customer_vehicle_id     | = | 12375                              |
| 35 | customer_id | = | 158332 | mernis_no               | = | 39133664094                        |
| 36 | customer_id | = | 158332 | education               | = | Üniversite                         |
| 37 | customer_id | = | 158332 | gender                  | = | Bay                                |
| 38 | customer_id | = | 158332 | marital_status          | = | Evli                               |
| 39 | customer_id | = | 158332 | child_count             | = | 1                                  |
| 40 | customer_id | = | 158332 | married_year            | = | 10                                 |
| 41 | customer_id | = | 158332 | age                     | = | 43                                 |
| 42 | customer_id | = | 158332 | corporation_sector      | = | DEVLET / KAMU                      |
| 43 | customer_id | = | 158332 | occupation_status       | = | employee                           |
| 44 | customer_id | = | 158332 | corporation_position    | = | MEMUR / KAMU / DEVLET<br>GÖREVLİSİ |
| 45 | customer_id | = | 158332 | countrymodel_definition | = | POLO 1.4 75 HP COMFORTLINE         |
| 46 | customer_id | = | 158332 | occupation              | = | DİĞER                              |
| 47 | customer_id | = | 158332 | ID                      | = | 5604                               |
| 48 | customer_id | = | 158332 | top_model_definition    | = | POLO                               |
| 49 | customer_id | = | 174648 | customer_id             | = | 174648                             |
| 50 | customer_id | = | 174648 | customer_vehicle_id     | = | 21294                              |

#### Ek 4 Araba Tablosu için indeksli arama metoduyla öğrenilen kurallar (İlk 50 kural)

Çizelge Ek 4.1 Araba Tablosu için indeks aramayla öğrenilen kurallar (İlk 50 kural)

| ID | Left column | Operator | Left Value | Right column        | Right Operator | Right Value |
|----|-------------|----------|------------|---------------------|----------------|-------------|
| 1  | education   | =        | Diğer      | customer_id         | <=             | 184823      |
| 2  | education   | =        | Diğer      | customer_id         | >=             | 132028      |
| 3  | education   | =        | Diğer      | customer_vehicle_id | <=             | 27769       |
| 4  | education   | =        | Diğer      | customer_vehicle_id | >=             | 478         |
| 5  | education   | =        | Diğer      | mernis_no           | <=             | 9,42E+10    |
| 6  | education   | =        | Diğer      | mernis_no           | >=             | 1,01E+10    |
| 7  | education   | =        | Diğer      | child_count         | <=             | 6           |
| 8  | education   | =        | Diğer      | child_count         | >=             | 0           |
| 9  | education   | =        | Diğer      | married_year        | <=             | 44          |
| 10 | education   | =        | Diğer      | married_year        | >=             | 0           |
| 11 | education   | =        | Diğer      | age                 | <=             | 77          |
| 12 | education   | =        | Diğer      | age                 | >=             | 20          |
| 13 | education   | =        | Doktora    | customer_id         | <=             | 185105      |
| 14 | education   | =        | Doktora    | customer_id         | >=             | 131223      |
| 15 | education   | =        | Doktora    | customer_vehicle_id | <=             | 28060       |
| 16 | education   | =        | Doktora    | customer_vehicle_id | >=             | 93          |
| 17 | education   | =        | Doktora    | mernis_no           | <=             | 7,34E+10    |
| 18 | education   | =        | Doktora    | mernis_no           | >=             | 1E+10       |
| 19 | education   | =        | Doktora    | child_count         | <=             | 5           |
| 20 | education   | =        | Doktora    | child_count         | >=             | 0           |
| 21 | education   | =        | Doktora    | married_year        | <=             | 44          |
| 22 | education   | =        | Doktora    | married_year        | >=             | 0           |
| 23 | education   | =        | Doktora    | age                 | <=             | 78          |
| 24 | education   | =        | Doktora    | age                 | >=             | 26          |
| 25 | education   | =        | İlkokul    | customer_id         | <=             | 185106      |
| 26 | education   | =        | İlkokul    | customer_id         | >=             | 131116      |
| 27 | education   | =        | İlkokul    | customer_vehicle_id | <=             | 28061       |
| 28 | education   | =        | İlkokul    | customer_vehicle_id | >=             | 32          |
| 29 | education   | =        | İlkokul    | mernis_no           | <=             | 7,25E+10    |
| 30 | education   | =        | İlkokul    | mernis_no           | >=             | 1E+10       |
| 31 | education   | =        | İlkokul    | child_count         | <=             | 11          |
| 32 | education   | =        | İlkokul    | child_count         | >=             | 0           |
| 33 | education   | =        | İlkokul    | married_year        | <=             | 66          |
| 34 | education   | =        | İlkokul    | married_year        | >=             | 0           |
| 35 | education   | =        | İlkokul    | age                 | <=             | 80          |
| 36 | education   | =        | İlkokul    | age                 | >=             | 20          |
| 37 | education   | =        | Okumamış   | customer_id         | <=             | 174567      |
| 38 | education   | =        | Okumamış   | customer_id         | >=             | 134358      |
| 39 | education   | =        | Okumamış   | customer_vehicle_id | <=             | 26708       |
| 40 | education   | =        | Okumamış   | customer_vehicle_id | >=             | 1649        |

|    |           |   |          |              |    |          |
|----|-----------|---|----------|--------------|----|----------|
| 41 | education | = | Okumamış | mernis_no    | <= | 6,85E+10 |
| 42 | education | = | Okumamış | mernis_no    | >= | 1,24E+10 |
| 43 | education | = | Okumamış | child_count  | <= | 4        |
| 44 | education | = | Okumamış | child_count  | >= | 0        |
| 45 | education | = | Okumamış | married_year | <= | 44       |
| 46 | education | = | Okumamış | married_year | >= | 0        |
| 47 | Education | = | Okumamış | age          | <= | 88       |
| 48 | Education | = | Okumamış | age          | >= | 9        |
| 49 | Education | = | Önlisans | customer_id  | <= | 185053   |
| 50 | Education | = | Önlisans | customer_id  | >= | 131123   |

### Ek 5 Uygulama algoritması kod örneği

#### Uygulamanın ilk aşamasında kullanılan var olan kuralların bulunması (sol taraf)

```
#Region "Create Left Rule Side"
Function CreateLeftSide()

    Dim sqlConnection As New SqlConnection

    Dim strConnection = "Server=" & txtServerName.Text & ";"
    strConnection = strConnection + "Database=" & ComboBox1.Text &
    ";Integrated Security=SSPI"
    CreateTable(strConnection)
    Dim strSQL = "select name from sys.columns c "
    strSQL = strSQL + "where c.object_id = '" & ComboBox2.SelectedValue
    & "' and c.column_id not in "
    strSQL = strSQL + "(select Column_id from sys.index_columns i "
    strSQL = strSQL + "Where i.object_id = '" & ComboBox2.SelectedValue
    & "' )"

    Try

        Dim Data = ExecSQL(strSQL, strConnection)
        Dim DataKolon = New DataSet
        Dim Command As SqlCommand

        For Each fields In Data.Tables("Table").Rows
            Dim strKolon = fields("name").ToString()

            strSQL = "select " & strKolon & ",Count(" & strKolon & ")
            as LeftCount"
            If TotalNumberOfTest = 0 Then
                strSQL = strSQL + " from " & ComboBox2.Text
            Else
                strSQL = strSQL + " from (Select TOP " &
                TotalNumberOfTest & " * from " & ComboBox2.Text & ") as Table1"
            End If
            strSQL = strSQL + " Group by " & strKolon
            DataKolon = ExecSQL(strSQL, strConnection)

            Dim total = DataKolon.Tables("Table").Rows.Count
            Dim justCount = 0
            lblProgress1.Text = "Progressing " & strKolon
            For Each rows In DataKolon.Tables(0).Rows
```

```

        Dim strInsert = "Insert into RuleLeft (leftcolumn,
lefttop, leftvalue, leftcount, leftIND ) VALUES ('"
        strInsert = strInsert + strKolon & "','=','" &
rows(0).ToString() & "',''" & rows(1).ToString() & "','0')"
```

```

        sqlConnection.ConnectionString = strConnection
        Command = New SqlCommand(strInsert, sqlConnection)
        sqlConnection.Open()
        Command.ExecuteNonQuery()
        sqlConnection.Close()
        Application.DoEvents()

    Next
    justCount = justCount + 1
    lblCounter.Text = "%" & CInt((justCount / total) * 100)
    ProgressBar1.Value = CInt((justCount / total) * 100)

    Next
    ProgressBar1.Value = Nothing
    lblCounter.Text = "%100"
    lblProgress1.Text = "Finish"
Catch ex As Exception
    'log
End Try
Return Nothing
End Function
```

### Uygulamanın elde edilen kurallardan yeni kural öğrenmesi (sağ taraf)

```

Dim strSQL = "select name, is_identity from sys.columns c "
"where c.object_id = '" & ComboBox2.SelectedValue & "' and c.column_id in "
"(select Column_id from sys.index_columns i "
"where i.object_id = '" & ComboBox2.SelectedValue & "' )"

Dim strLeftSql = "SELECT leftcolumn, lefttop, leftvalue, leftcount, leftIND
FROM RuleLeft"

For Each fields In LeftData.Tables("Table").Rows
    strLeftColumn = fields("leftcolumn").ToString()
    strLeftOp = fields("lefttop").ToString
    strLeftVal = fields("leftvalue").ToString
    strLeftCount = fields("leftcount").ToString
    strLeftIND = fields("leftIND").ToString

    strselectsql = "SELECT Count(DISTINCT " & strRightName & ") "
    strselectsql = strselectsql + " WHERE " & strLeftColumn & "=" & strLeftVal
    & """"

    strselectsql = "SELECT Max(" & strRightName & ") "
    strselectsql = strselectsql + " WHERE " & strLeftColumn & "=" & strLeftVal
    & """"

    strInsert = "Insert into RuleSet (leftcolumn, lefttop, leftvalue, leftcount,
leftIND, rightcolumn, righttop, rightvalue, rightcount, rightIND) VALUES ('"
    strInsert = strInsert + strLeftColumn & "',''" & strLeftOp & "',''" &
    strLeftVal & "',''" & strLeftCount

    strInsert = strInsert + "',''" & strLeftIND & "',''" & strRightName &
    "','<=','" & strRightValue & "',''" & strRightCount & "',''" & strRightIND
    & """)"

    strselectsql = "SELECT Min(" & strRightName & ") "
```

```

strselectsql = strselectsql + " from (Select TOP " & TotalNumberOfTest & "
* from " & ComboBox2.Text & ") as Table1"
strselectsql = strselectsql + " WHERE " & strLeftColumn & "=" & strLeftVal
& ""

```

```

strInsert = "Insert into RuleSet (leftcolumn, lefttop, leftvalue, leftcount,
leftIND, rightcolumn, righttop, rightvalue, rightcount, rightIND) VALUES ("
strInsert = strInsert + strLeftColumn & "', '" & strLeftOp & "', '" &
strLeftVal & "', '" & strLeftCount
strInsert = strInsert + "', '" & strLeftIND & "' ,'" & strRightName &
"', '>=', '" & strRightValue & "', '" & strRightCount & "', '" & strRightIND
& "')"

```

```

strselectsql = "SELECT DISTINCT " & strRightName & " FROM " & ComboBox2.Text
strselectsql = strselectsql + " WHERE " & strLeftColumn & "=" & strLeftVal
& ""
strselectsql = strselectsql + " and " & strRightName & " is Not Null"

```

```

strInsert = "Insert into RuleSet (leftcolumn, lefttop, leftvalue, leftcount,
leftIND, rightcolumn, righttop, rightvalue, rightcount, rightIND) VALUES ("
strInsert = strInsert + strLeftColumn & "', '" & strLeftOp & "', '" &
strLeftVal & "', '" & strLeftCount
strInsert = strInsert + "', '" & strLeftIND & "' ,'" & strRightName &
"', '<=', '" & strRightValue & "', '" & strRightCount & "', '" & strRightIND &
"')"

```

## ÖZGEÇMİŞ

|               |            |                                                                                                              |
|---------------|------------|--------------------------------------------------------------------------------------------------------------|
| Doğum Tarihi  | 30.09.1981 |                                                                                                              |
| Doğum Yeri    | İstanbul   |                                                                                                              |
| Lise          | 1996-1999  | Üsküdar Anadolu Lisesi                                                                                       |
| Lisans        | 2000-2006  | Yıldız Teknik Üniversitesi<br>Matematik Mühendisliği Anabilim Dalı<br>Matematik Mühendisliği Programı        |
| Yüksek Lisans | 2007-Devam | Yıldız Teknik Üniversitesi<br>Ediyor Matematik Mühendisliği Anabilim Dalı<br>Matematik Mühendisliği Programı |

## Çalıştığı Kurumlar

|            |                                 |
|------------|---------------------------------|
| 2006-2008  | Doğuş Otomotiv A.Ş.             |
| 2008-Devam | Ediyor Roche Müstehzarları A.Ş. |