



MARMARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



**MÜŞTERİ İLİŞKİLERİ YÖNETİMİ İÇİN VERİ
MADENCİLİĞİ KULLANILMASI VE
SİGORTACILIK SEKTÖRÜ ÜZERİNE BİR
UYGULAMA**

BAHAR EROL

YÜKSEK LİSANS TEZİ

Bilgisayar ve Elektronik Anabilim Dalı
Elektronik ve Haberleşme Eğitimi Programı

DANIŞMAN

Yrd. Doç. Dr. Buket DOĞAN

EŞ-DANIŞMAN

Doç. Dr. Ali BULDU

İSTANBUL, 2013

ÖNSÖZ VE TEŞEKKÜR

Ülkemizde rekabetin hızla artmasıyla müşteri ilişkileri yönetimi ön plana çıkmış ve artan rekabetle birlikte işletmeler müşterilerini daha iyi anlayabilmek, değerli müşterileri açığa çıkartmak, mevcut müşterileri elde tutmak ve yenilerini kazanmayı hedeflemektedirler. Bu ihtiyaçlar üzerine mevcut müşteri verilerinden faydalanılarak veri madenciliği yöntemleri kullanılmış ve veriden bilgi keşfi gerçekleştirilmiştir.

Tez çalışmasında, Türkiye'nin önde gelen sigorta şirketlerinden birinin veri setlerini incelemiş, SPSS Clementine yazılımı ile veriler analiz edilerek, sigorta sektörü için yararlı olacak bilgi keşfi yapılması hedeflenmiştir.

Her konuda yardımlarını benden esirgemeyen, her türlü desteğini aldığım kıymetli eğitimcim Yrd. Doç. Dr. Buket DOĞAN'a ve Doç. Dr. Ali BULDU'ya ayrıca her konuda desteğini aldığım, tecrübelerini benden esirgemeyen Harun CEREN'e sonsuz teşekkürlerimi sunarım.

Bu alanda çalışanlara faydalı olmasını diliyorum.

Saygılarımla,

Mayıs,2013

Bahar EROL

İÇİNDEKİLER

	SAYFA
ÖNSÖZ VE TEŞEKKÜR	i
İÇİNDEKİLER	ii
ÖZET	vi
ABSTRACT	vii
SEMBOLLER	viii
KISALTMALAR	ix
ŞEKİL LİSTESİ	x
TABLO LİSTESİ	xiv
1. GİRİŞ	1
1.1. Tezin Amacı	1
1.2. MİY Yaklaşımıyla Veri Madenciliği	2
1.2.1. MİY kavramının ortaya çıkışı	2
1.2.2. MİY'nin tanımı	3
1.2.3. MİY'nin mimarisi	5
1.2.4. MİY'nin bileşenleri	6
1.2.4.1. MİY insan bileşeni	6
1.2.4.2. MİY iş süreçleri bileşeni	7

1.2.4.3. MİY’de teknoloji bileşeni	8
1.2.5. MİY’nin süreçleri	9
1.2.5.1. Müşteri seçimi	9
1.2.5.2. Müşteriyi elde tutma	10
1.2.5.3. Müşteriyi edinme	10
1.2.5.4. Müşteri büyütme ve derinleştirme	10
1.2.6. MİY’nin amaçları	11
1.2.7. MİY’nin sağladığı avantajlar	12
1.3. MİY’de Veri Madenciliği	13
1.4. Literatürde Veri Madenciliği ve MİY	15
2. MATERYAL VE YÖNTEM	19
2.1. Veritabanı ve Veri Ambarı	19
2.1.1. Veritabanı sistemleri	19
2.1.2. Veri ambarı	19
2.1.3. OLTP(On-line Transaction Process)	19
2.1.4. OLAP(On-line Analytic Processing)	20
2.2. Veri Madenciliği	22
2.2.1. Veri madenciliği kavramı ve gelişimi	22
2.2.2. Veri madenciliğinin uygulama alanları	23
2.2.3. Veri madenciliği sürecinin aşamaları	24

2.3. Veri Madenciliği Teknikleri	25
2.3.1. Tahmin edici model	26
2.3.2. Tanımlayıcı model	26
2.4. Veri Madenciliği İşlevleri	26
2.4.1. Sınıflama ve Regresyon	27
2.4.1.1. Karar ağaçları (Decision Trees)	28
2.4.2. Birliktelik kuralları	33
2.4.2.1. AIS algoritması	35
2.4.2.2. Apriori algoritması	36
2.4.2.3. DHP algoritması	40
2.4.2.4. Partition algoritması	40
2.4.3. Kümeleme	40
2.4.3.1. K-ortalamlar algoritması	44
2.4.3.2. Kohonen ağları	45
3. BULGULAR VE TARTIŞMA	51
3.1. Sigortacılık Sektörü Üzerine Uygulama	51
3.1.1. Araştırma probleminin tespiti (Business Understanding)	52
3.1.2. Verileri tanıma (Data Understanding)	52
3.1.3. Veri hazırlama (Data Preparation)	53
3.1.3.1. Kurulum aşaması	53

3.1.3.2. Verilerin temizlemesi ve düzenlenmesi aşaması	56
3.1.3.3. Verinin dönüştürülmesi ve veri setlerinin hazırlanması	59
3.1.3.4. Veri setlerinin analizi	66
3.1.4. Modelleme aşaması (Modeling)	73
3.1.4.1. Birliktelik analizinin uygulanması	74
3.1.4.2. Kümeleme analizinin uygulanması	76
3.1.5. Değerlendirme (Evaluation)	102
3.1.6. Konuşlandırma (Deployment)	103
4. SONUÇLAR	104
KAYNAKLAR	108
EKLER	117
ÖZGEÇMİŞ	

ÖZET

Bilgi teknolojilerinin geliřimi, rekabetin artması, müşteri profilinin deęiřmesi ve hızlı deęiřimler günümüzde işletmelerin pazarlama anlayışını da köklü deęiřimler yapmaya yöneltmiştir. MİY(Müşteri İliřkileri Yönetimi), müşterinin sadakatini ve memnuniyetini arttırmak amacıyla müşterilerle sürekli ilişki içerisinde olmayı ve müşterilerin beklentilerine uygun ihtiyaç ve istekleri belirleyip müşteriye sunabilmelidir. Müşteri ilişkileri yönetiminde süreç, ilk olarak müşteriler hakkında olabildiğince veri toplamaya dayanır. Bu nedenle şirketlerin toplamış olduęu veri ambarlarındaki veriler bir araya getirilerek, bunların gruplandırılıp müşterilerin karakteristik özelliklerinden yola çıkarak ne tür ürünleri tercih ettiklerini bulmak ve bunların ışığında müşteriye yön vermek son derece önemlidir. VM(Veri Madencilięi) bu aşamada fayda sağlamakta olup veri ambarlarında tutulan verilerden analizler sayesinde anlamlı bilgi ve ilişkileri açığa çıkarma sürecini kapsamaktadır.

“Müşteri İliřkileri Yönetimi İçin Veri Madencilięi Kullanılması Ve Sigortacılık Sektörü Üzerine Bir Uygulama” adlı çalışmada, müşteri ilişkileri yönetimi için sigortacılık sektöründeki bir veri ile bilgi keşfi sürecinin hangi aşamalardan oluştuęunu örnek çalışmalarla ortaya koymaktadır. Türkiye’de faaliyet gösteren Sektöründe öncü bir sigorta şirketinin veri tabanlarından veriler alınarak, bu veriler üzerinden VM’nin başlıca algoritmalarından olan Apriori, K-Ortalamalar ve Kohonen Ağları algoritmaları veri setlerimize uygulanmıştır. Uygulamanın ardından Müşteri İliřkileri Yönetimine ilişkin bilgiler elde edilmiştir.

ABSTRACT

Development in information Technologies, increased competition, change in customer profile and rapid changes have made today's enterprises introduce radical changes in their marketing understanding. CRM(Customer Relations Management) should be able in a constant relation with the customers in order to increase loyalty and satisfaction of the customers and establish requirements and demands suitable for customer expectations and offer the same to the customers. Process in customer relations management is based on collecting as much data as possible about the customers. Therefore, by bringing together the data compiled by the companies in data warehouses, it is extremely important to group such data and find out what types of products customers prefer based on the characteristics of the customers and direct the customers under the light of these. DM (Data Mining) is useful at this stage and covers process of revealing meaningful information and relation from data kept in data warehouses owing to analyses.

“Using Data Mining for Customer Relationship Management and An Application on the Insurance Sector” named study shows what components are involved in the information discovery process with data related to insurance industry through case studies for management of customer relations. After taking data from databases of a market leading insurance company running operations in Turkey, algorithms such as Apriori, K-Means and Kohonen Networks, main algorithms of DM via such data, have been applied to our datasets. Following the application, information concerning CRM has been gained.

May, 2013

Bahar,EROL

SEMBOLLER

- $A \Rightarrow B$: Birliktelik Kuralı
- C_k : k deęerindeki aday kümeler
- $d(x,y)$: Öklit mesafe ölçüsü
- k : Küme sayısı
- L_k : k deęerindeki büyük ürün kümelerinin kümesi
- M : Marriot tarafından önerilen küme sayısı
- m_i : C_i kümesinin merkez noktası
- n : Gözlem sayısı
- p : Bölümlerin sayısı
- SSE : Karesel hata kriteri
- W : Grup içi kareler toplamı matrisi
- W_{ki} : İlgili girişin ağırlık deęeri
- x : C_i kümesinde bulunan bir nesne
- X_i : Birden fazla girişin temsili
- Y_k : Nöronun sınırlandırılmış genlikli çıkışı (Aksiyon Fonksiyonu)

KISALTMALAR

BPR	: Business Process Re-Engineering
CPR	: Customer Process Re-Engineering
CRM	: Customer Relationship Management
MİY	: Müşteri İlişkileri Yönetimi
OLAP	: On-line Analytic Process
OLTP	: On-line Transaction Process
SOM	: Özdüzenlemeli Harita
SQL	: Structured Query Language
SSE	: Sum of Squared Error
VM	: Veri Madenciliği
VTBK	: Veri Tabanlarında Bilgi Keşfi
VYS	: Veritabanı Yönetimi Sistemleri

ŞEKİL LİSTESİ

	SAYFA
Şekil 1.1. MİY Mimarisi	5
Şekil 2.1. Bilgisayar Satın Alma Kavramı Üzerine Karar Ağaçlarının Uygulanması	29
Şekil 2.2. SOM Temel Mimarisi	46
Şekil 2.3. SOM'un Geliştirilmesi	46
Şekil 2.4. SOM'un Çıkış Ağırlığının Belirlenmesi	47
Şekil 2.5. Kohonen Geri Bildirim Grafiği	48
Şekil 3.1. CRISP-DM Modeli	51
Şekil 3.2. Oracle Enterprise Edition Database 11g	52
Şekil 3.3. Oracle Veritabanı Kurulum Aşaması	53
Şekil 3.4. Oracle Client Kurulum Aşaması	54
Şekil 3.5. ODBC Veri Kaynağı Yöneticisine Sürücünün Eklenmesi	55
Şekil 3.6. Clementine Data Source Sekmesi	55
Şekil 3.7. Veri Tabanı Bağlantısının Kurulması Aşaması	56
Şekil 3.8. Clementine Arayüzünde Örnek Veri Setinin Görüntülenmesi	56
Şekil 3.9. Veri Setinin Data Audit Aracından Geçirilmesi	58
Şekil 3.10. Data Audit Analiz Sonuçları	58
Şekil 3.11. Veri Setinin Filtreden Geçirilmesi	59
Şekil 3.12. Açıklama Değişkeninin İncelenmesi	60

Şekil 3.13. İl değişkeninin İncelenmesi	60
Şekil 3.14. Birliktelik Veri Setinden Kesit	63
Şekil 3.15. Birliktelik Analizi Veri Setinin Alanlarının Türleri	64
Şekil 3.16. Kümeleme Veri Setinin Veri Tabanından Görüntüsü	65
Şekil 3.17. Web Grafiği Aracı	66
Şekil 3.18. Gruplanmış Sigorta Türleri Arasındaki Bağlantılar	67
Şekil 3.19. Sıklığı Yüksek Olan Bağlantılar	68
Şekil 3.20. Kümeleme Veri Seti Analizi	69
Şekil 3.21. Cinsiyet Değişkeninin Veriler Üzerindeki Etkisi	69
Şekil 3.22. Teminatlara Göre Dağılım İstatistikliği	70
Şekil 3.23. Meslek Değişkeninin Veri Setine Etkisi	71
Şekil 3.24. İllere göre dağılım grafiği	72
Şekil 3.25. Yaş Dağılımı Grafiği	73
Şekil 3.26. Apriori Algoritmasının Uygulanması	74
Şekil 3.27. Apriori Analiz Sonuçları	75
Şekil 3.28. K-Ortalamlar Analizinin Uygulanması	76
Şekil 3.29. Veritabanı Aracında Sorguların Düzenlenmesi	77
Şekil 3.30. K-Ortalamlar Modelinin Kurulma Aşaması	78
Şekil 3.31. K-Ortalamlar Yöntemi ile Elde Edilen Kümeler	79
Şekil 3.32. Müşteri Cinsiyetinin Birinci Küme Üzerindeki Etkisi	80

Şekil 3.33. Müşteri İl Değişkeninin Birinci Küme Üzerindeki Etkisi	80
Şekil 3.34. Satın Alınan Teminat Değişkeninin Birinci Küme Üzerindeki Etkisi	81
Şekil 3.35. Müşteri İl'i değişkenin İkinci Küme Üzerine Etkisi	82
Şekil 3.36. Müşteri Cinsiyetinin İkinci Küme Üzerindeki Etkisi	83
Şekil 3.37. Meslek Değişkeninin İkinci Küme Üzerine Etkisi	84
Şekil 3.38. Satın Alınan Teminatların İkinci Küme Üzerindeki Etkisi	85
Şekil 3.39. Müşteri Cinsiyetinin Üçüncü Küme Üzerine Etkisi	86
Şekil 3.40. Müşteri İl Değişkeninin Üçüncü Küme Üzerine Etkisi	86
Şekil 3.41. Satın Alınan Teminatların Üçüncü Küme Üzerine Etkisi	87
Şekil 3.42. Meslek Değişkeninin Üçüncü Küme Üzerine Etkisi	88
Şekil 3.43. Yaş Değişkeninin Üçüncü Küme Üzerine Etkisi	89
Şekil 3.44. K-Ortalamlar Analizi Model Özeti	89
Şekil 3.45. Müşteri Cinsiyeti Değişkeninin Kümeler Üzerindeki Etkisi	90
Şekil 3.46. Müşteri İl'i Değişkeninin Kümeler Üzerindeki Etkisi	91
Şekil 3.47. Satın Alınan Teminatların Kümeler Üzerindeki Etkisi	92
Şekil 3.48. Yaş Değişkeninin Kümeler Üzerine Etkisi	93
Şekil 3.49. Meslek Değişkeninin Kümeler Üzerindeki Etkisi	94
Şekil 3.50. Veri Setine Kohonen Ağlarının Uygulanması	95
Şekil 3.51. Kohonen Ağları Boyut Seçimi	96
Şekil 3.52. Clementine Kohonen Modeli Aracı	96

Şekil 3.53. Kohonen Ağları Geri Besleme Grafiği	97
Şekil 3.54. Kohonen Ağları ile Kümeleme Analizi Sonuçları	98
Şekil 3.55. Analiz Sonucu Oluşan Kümelerin Eleman Sayıları	99
Şekil 3.56. $(X=0, Y=0)$ Koordinatlı Kümeye Satın Alınan Teminatın Etkisi	101
Şekil 3.57. $(X=0, Y=0)$ Koordinatlı Küme Analizi	102

TABLO LİSTESİ

	SAYFA
Tablo 2.1. Tek Elemanlı Sık Tekrar Eden Ürün Kümesi	39
Tablo 2.2. İki Elemanlı Sık Tekrar Eden Ürün Kümesi	39
Tablo 3.1. Tabular Data Formatı	61
Tablo 3.2. Birliktelik Analizi Veri Seti	62
Tablo 3.3. Birliktelik Veri Setindeki Alan ve Açıklamalar	64
Tablo 3.4. Kümeleme Veri Seti	65
Tablo 3.5. Meslek Grupları Dönüşümü	77

1. GİRİŞ

Günümüzde, teknolojinin hızla ilerlemesiyle birlikte insanların kendi aralarındaki ilişkiler kadar şirketlerin müşterileriyle olan ilişkileri de farklılaşmıştır. Bu yeni dönemde, özellikle şirketler için rekabet daha yoğun hale gelmiş, coğrafi sınırlar anlamını yitirmiş, müşterilerin kişisel tercihlerine daha fazla önem vermeye başlanmış ve birebir pazarlama stratejileri ön plana çıkmaya başlamıştır. Bu şartlarda şirketler müşterilerinin sadakatini garanti edemez hale gelmiş ve müşterisine daha yakın olma ihtiyacı duymaya başlamıştır (Odabaş, 2005). MİY(Müşteri İlişkileri Yönetimi), işletmenin herhangi bir bölümünün tek başına öne çıkmasını engelleyen ve koordineli olarak uygulanan bütünleşik bir pazarlama, satış ve servis stratejisidir (Dereli, 2004).

İşletmelerde başarılı bir MİY ile küresel bir dünyada ve giderek artan rekabetçi piyasalarda işletmeler için yaşamsal önem taşıyan, müşteri için değer yaratmak, müşteri sadakatini sağlamak ve bu konularda kurumsallaşmayı gerçekleştirmek mümkün olabilmektedir. Yazılım ve donanım alanındaki teknolojik gelişmeler, işletmelerin çok miktarda veriyi saklayabilmesini ve işleyerek anlamlı bilgilere dönüştürmesini mümkün hale getirmiştir. Teknolojik gelişmelerin özellikle iş hayatına yansımaları sonucunda satış ve pazarlama alanlarındaki gelişmelerin bir sonucu olarak, ortaya çıkan müşteri ilişkileri ve VM(Veri Madenciliği) yöntemleri çok sayıda müşteri verilerini analiz ederek satışların artmasını sağlamaktadır.

1.1. Tezin Amacı

Bu tez çalışmasının amacı veri madenciliği yöntemlerinin sigortacılık sektöründe müşteri ilişkileri yönetimine yönelik kullanılmasına ilişkin uygulama gerçekleştirilmesi amaçlanmıştır. Bu amaçla bir sigorta şirketinin müşterine ait veri seti kullanmış ve buradaki verilerinden yararlanılarak birliktelik kuralı analizi ve kümeleme analizi gerçekleştirilmiştir. Bu analizlerin amacı, müşterilerin ürün satın alma alışkanlıklarının bulunması ve benzer müşteri kümelerinin ortak özelliklerinin elde edilmesidir. Bu sonuçlar yardımıyla, şirketlerin müşterilerini tanıması ve yeni pazarlama stratejileri geliştirmeleri hedeflenmektedir.

1.2. MİY Yaklaşımıyla Veri Madenciliği

1.2.1. MİY kavramının ortaya çıkışı

Don Peppers tarafından ilk ileri sürülen MİY ya da İngilizce adıyla CRM(Customer Relationship Management) aslında işletmede yeni bir kavram değildir (Payne, 2005).

Bill Gates 1995 yılında yazdığı “The Road Ahead” adlı kitabında “İletişimdeki devrim henüz yeni başlıyor” demiştir(Gates, 1995). Zaman içinde iletişimde yaşanan devrim pazarlama stratejilerini değiştirmiş ve bilgi çağı işletmelerin yalnızca satış ve pazarlama stratejilerini değiştirmiş ve bilgi çağı işletmelerin sadece satış ve pazarlama sistemleri değil bunun yanı sıra tüm organizasyon yapısını etkilemiştir.

MİY, bir şirketin müşterisiyle yakınlaşmak, onlar üzerinde olumlu etkiler yaratmak suretiyle tercih edilme oranını yükseltilmesini amaçlayan bir süreçtir(Payne, 2005). Şirketler yoğun rekabet koşulları altında, hem müşterisini elinde tutabilmek hem de onu memnun etmek aynı zamanda da yeni müşteriler elde etmek için çabalarken, kar ve zarar hesabını doğru yapabilmek için yeniden yapılanmaktaydılar. Gelişen bilgi teknolojileri sayesinde, insan gücüne destek olacak yeni uygulamalar yaygınlaştı. Önceleri sadece veri bankası olarak kullanılan sistemlerin, işlem güçlerini arttırmasıyla şirketler tüm süreçlerde daha kontrollü ve daha verimli çalışabilme yeteneğini kavuştu (Payne, 2005).

MİY’in ortaya çıkış nedenleri

- Kitlesel pazarlamanın gittikçe pahalı bir müşteri kazanma yolu olması
- Pazar payının değil müşteri payının önemli hale gelmesi
- Var olan müşterinin değerinin anlaşılması ve bu müşteriyi elde tutma çabalarına gerek duyulması
- Yoğun rekabet ortamı
- İletişim teknolojileri ve veri tabanı sistemlerinde yaşanan gelişmeler

MİY, henüz gelişimini tam olarak tamamlamamış olmakla birlikte, uygulama alanlarının yaygınlaşması ve teknoloji ile bütünleşmiş bir felsefe olduğunun bilincine

varılmasıyla MİY, yeni yüzyılda firmaların hayatlarını sürdürebilmeleri için benimsenmesi ve uygulanması gereken anlayış olarak ortaya çıkmıştır (Gel, 2003).

MİY'in temelleri özellikle rekabetin kendini açıkça göstermeye başlamasından sonra atılmıştır. Çok sayıda müşterisi olan veya çok sayıda ürün/hizmet sunan ancak daha da önemlisi rekabeti her an yaşayan ve çoğu serbest piyasa ekonomisinin olabildiğince doğru uygulandığı Amerika'da faaliyet gösteren büyük şirketler, 1970'li yılların sonunda MİY kavramına önem vermeye başlamışlardır. Bu dönemde şirketler farklı bir bakış açısıyla hareket edip, stratejilerini oluşturmak için müşterilerin gerçek davranışlarını da anlamaya çalışmışlardır (Gel, 2003).

1.2.2. MİY'nin tanımı

MİY'in birden fazla tanımı yapılmıştır. Bunlar şu şekilde ifade edilebilir:

- MİY'in satışta, pazarlamada ve hizmet süreçlerinin müşteri odaklı bir felsefe etrafında yeniden tanımlanmasını gerektiren bir süreçtir (Odabaş, 2005).
- MİY, işletmenin tümüne müşteri kavramını yerleştiren müşteri odaklı olma kültürünü benimseyen bir strateji olarak tanımlanabilir. Uzun dönemli müşteri ilişkileri kurmayı hedefleyen, buna bağlı olarak karlılığı amaç edinen kavramdır. Bu kavram pazarlama stratejileri ile bilgi teknolojileri arasında köprü vazifesi görür (Ersoy, 2002).
- MİY, müşteri memnuniyetini kâra dönüştürmek amacı taşıyan, istediği müşteriye istediği deneyimi yaşatabilecek kabiliyette bir kurum felsefesi ve bu hedefe ulaşmak için gerekli insan, süreç, teknoloji yapılanmasıdır (Dereli, 2004).
- MİY, farklı müşterilere farklı davranmak ilkesi üzerine kurulan bir pazarlama yaklaşımıdır. Eğer işletme, müşteriye, işletmeye sağladığı mevcut ve potansiyel değere uygun olarak farklı davranış sergilerse, bu durum işletmeyi, müşteri

gözünde farklılaştıran bir konuma getirir. Farklı müşterilere farklı davranışta bulunabilmek için, öncelikle bu müşterilerin kimliklerini belirlemek, diğer bir ifade ile müşteriye tanımlamak gerekmektedir. Bu nedenle, işletmeler müşterilerini kişi olarak tanıdıkları, ya da bir diğer anlatımla kârlı ve kârsız müşterilerini ayırabildikleri ölçüde müşterilerine beklendikleri pazarlama hizmetini verebileceklerdir. Bunun için işletme, müşteri kimlik bilgisi şeklinde bir müşteri tanımlama sistemi oluşturabilir. Bu sistem, bir müşteriye, diğer müşteriden ayıran, işletmenin bu müşterisi ile zaman içindeki ilişkilerini izlemesine olanak sağlayan ya da müşteri ile bire-bir temasa geçmesine yardımcı olan bilgilerden oluşmaktadır. Bu çerçevede işletmenin aşağıda yer alan sorulara yanıt bulması gerekmektedir (Kırım, 2001: 60-63).

-İşletmeniz kaç tane müşteriye kişisel olarak tanıyor?

-İşletmenin tüm müşterilerini içeren bir veritabanı (database) mevcut mu?

-İşletmenin bu veri tabanı ne kadar güncel tutuluyor?

-Veri tabanında müşteri hakkında hangi ayrıntıda bilgi bulunuyor?

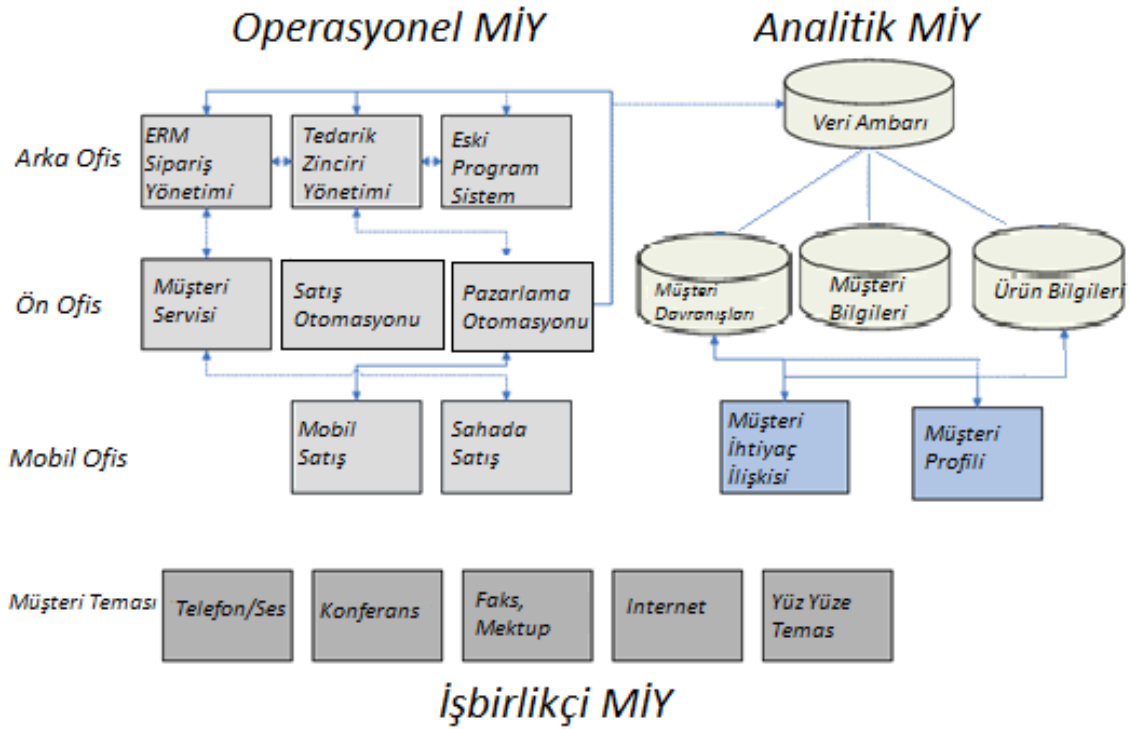
-Müşteri kimlik bilgilerini edinebileceğiniz diğer bilgi kaynakları mevcut mu?

-Elde edebileceğiniz müşteri bilgisini artırmanın değişik yöntemleri mevcut mu?

- MİY, herhangi bir bölümün tek başına öne çıkmasını engelleyen ve koordineli olarak uygulanan bütünleşik bir pazarlama servis ve satış stratejisidir (Kallkota, 1999).
- MİY, müşteri değerinin artışını ve değerli müşterilerin yaşam boyu sadık kalmasını sağlayarak satın almaya yönelmesini sağlayan doğru araçları kullanmayı mümkün kılan bir alt yapıdır (Dyche, 2012).
- Philip Kotler'e göre MİY, müşterilerle etkin veritabanları yardımıyla bireysel ilişki kurarak mükemmel gerçek zamanlı hizmet sağlayan iş stratejisidir (Kotler, 2003).

1.2.3. MİY'nin mimarisi

MİY mimarisi Şekil 1.1'de görüleceği gibi; Operasyonel MİY, Analitik MİY, İşbirlikçi MİY olmak üzere üç unsurdan oluşmaktadır. Operasyonel MİY; Ön Ofis ve Arka Ofisten bölümlerinden oluşmaktadır. Operasyonel MİY sürecinde; müşteri veri tabanı, Analitik MİY de ise müşteri işlemleri bazında veri ambarı oluşturulur. İşbirlikçi MİY ise müşteri ile ilişkilerin kurulduğu arayüzdür.



Şekil 1.1. MİY Mimarisi (Ezerce, 2008)

Operasyonel MİY; işletmelerin sipariş yönetimi, tedarik zinciri yönetimi, servis satış otomasyonu, pazarlama otomasyonu, saha servisi, mobil satış gibi operasyonel işleyişini içermektedir. Pazarlama satış faaliyetleri ile başlayan süreçte siparişlerin alınması ve yerine getirilmesi, satış sonrası destek hizmeti kapsam içerisindedir (Ezerce, 2008).

Analitik MİY, işletme performans yönetimini içermektedir. Operasyonel kısım ve temas noktalarında oluşan tüm bilgilerin derlenmesini ve detaylı analizini içermektedir. Analitik MİY'de müşteri işlemlerinin detaylı analizi ile müşteri davranışlarının anlaşılmasını sağlayan anlamlı istatistikler elde edilmektedir. Elde edilen bu raporlara

göre uygun MİY satış ve pazarlama stratejileri geliştirilmektedir(Taşpınar, 2006). Analitik MİY, uygulamaların operasyonel tarafı ile yaratılan verilerin incelenmesi neticesinde, performansın yönetimine dair kararların alınmasını sağlamasıdır. Bu yüzden veri ambarlarının alt yapısının kurulması esastır. Analitik MİY, işletmelerin mevcut ve potansiyel müşteriye yönelik ihtiyaçlarının, fırsatların kaçırılmaması, risk ve maliyet unsurlarının bir arada düşünülmesini sağlayan bir çalışmadır (Sözügüzel, 2004).

Analitik MİY, müşteri odaklı bir kurumun en önemli uygulamalarından birisidir. Kurum ve müşteri arasında adeta bir köprü vazifesi görmektedir (Mamaklı, 2005).

1.2.4. MİY'nin bileşenleri

MİY standart olarak İnsan, İş Süreçleri ve Teknoloji bileşenlerinden oluşmaktadır.

MİY bileşenlerine, MİY'nin kapsamı açısından bakıldığında, bileşenlerin yapıdaki yerleri aşağıdaki gibidir:

İnsan → Operasyonel MİY, işbirlikçi MİY

Süreç → Operasyonel MİY, işbirlikçi MİY

Teknoloji → Operasyonel MİY, Analitik MİY, işbirlikçi MİY'dir (Akça, 2007).

MİY'nin bileşenleri aşağıda ayrıntılı olarak anlatılmıştır.

1.2.4.1. MİY insan bileşeni

İnsan bileşeni MİY'nin ve müşteri odaklı yeniden yapılanma projelerinde en temel unsurdur. MİY yolculuğuna çıkan işletme ciddi bir değişim içerisine girecektir ve işletme kültürü bu değişim programının başarısında en belirleyici rolü oynamaktadır (Taşpınar, 2006: 38). Unutulmaması gereken bir konu vardır ki, o da MİY projeler, yazılımlar ve danışmanlar ile değil o işletmenin insan kaynağı ile yapılır (Oğuz, 2004).

1.2.4.2. MİY iş süreçleri bileşeni:

Üretim işletmelerinde, 1995-2000 yılları arasında yaşanan BPR (Business Process Re-Engineering) çalışmaları işletmenin tüm işlemlerinin sıfırdan masaya yatırılıp gözden geçirilmesi ve yeniden kurulmasını içermekteydi ve ne yazık ki çok başarısızlıkla sonuçlandı. Başarılı olmamasının temel sebebi çalışmaların vizyon ve stratejiden yoksun olmasıydı. Süreçler gözden geçirilecekti ancak ne için geçirileceği bilinmiyordu. MİY ile birlikte, BPR (Business Process Re-Engineering) türkçe karşılığı olan iş süreçleri yapılandırma, biraz değişerek müşteri odaklı yapılanma olarak CPR (Customer Process Re-Engineering) yeniden karşımıza çıkmıştır. Tüm süreçler artık müşteriye en fazla değer sağlayacak şekilde süzgeçten geçirilmelidir (Online ERP, 2013).

MİY’de süreç yönetimi, öncelikle müşterileri dinlemekle, müşterilerin şirketten beklentilerini ve bu şirketten nasıl bir ürün beklediklerini ve bu ürünün kendilerine hangi şekilde ulaştırılmasını istediklerini anlamakla başlar. Daha sonra karlı olma potansiyeline sahip olanlar ve olmayan müşteriler tespit edilerek belirlenmelidir. Bundan sonra bölümlere ayrılmış müşterilerin, her kademedeki müşterilerin ihtiyaçları belirlenip ihtiyaca yönelik beklentilerini karşılayacak konularda bilgiler toplanmalıdır (Kırım, 2001: 52).

Dünyada kabul görmüş MİY uygulama aşamaları aşağıdaki gibidir. Öncelikle işletmenin MİY projesine ihtiyaç duyup duymadığı araştırılmalı ve işletmenin MİY projesine ihtiyaç duyduğuna karar verildikten sonra aşamalar izlenmelidir.

i. Hazırlık Aşaması

MİY çalışmalarının en önemli aşaması hazırlık aşamasıdır. Başarı veya başarısızlık büyük oranda bu aşamada belli olmaktadır. Bu nedenle önce MİY projesinin boyutuna karar verilmeli ve karar verilen uygulamaya inanılmalıdır. İşletmelerde MİY projelerini uygulayanlar, sadece konu ile ilgilenmiş olmak, MİY’i yalnız gücünü artırma amacı ile yeniden yapılanma aracı olarak görmelidirler (Mamaklı, 2006).

ii. Planlama Aşaması

MİY projelerinde planlama aşaması doğru iş stratejilerinin belirlenmesi gereken bölümdür. Firma dışı unsurlar; Pazar, müşteri, rakipler ve firma içi unsurlar; iş yapış biçimi, ürünler incelenerek iş stratejilerine ulaşılmalıdır (Mamaklı, 2006).

iii. Tasarım Aşaması

Bu aşamada müşteri odaklı yeniden yapılanma için önce müşteriler tanımlanmalıdır. Daha sonra ihtiyaçları ve değerlerine göre sınıflandırılmalı ve onlara en doğru ortamlarda müşterilerle geri beslemeli ilişkiler kurulmalıdır. Bu ilişkiler her seferinde işletme hizmet ve ürünlerinin bu ihtiyacı ve talebi doğrultusunda değiştirilmesini kapsayacaktır (Kırım, 2001: 70).

iv. Uygulama Aşaması

Uygulama aşaması tasarlanan sistemin hayata geçirileceği aşamadır. Bu aşamada mevcut sistem ile tasarlanan sistemin farkları çıkarılarak yeni sisteme hangi süreç içerisinde geçirilebileceği analiz edilmelidir. İşletmelerin MİY uygulamalarını hayata geçirebilmeleri için aşağıdaki aşamaları gerçekleştirmeleri gerekmektedir.

- İlişki Bazlı Pazarlama Stratejisi
- İşlevsel Rollerin Yeniden Tasarlanması
- İş Süreçlerinin Yeniden Tasarlanması

1.2.4.3. MİY’de teknoloji bileşeni:

MİY kavramının ve uygulamalarının bu derece popülerlik kazanmasını temel yazılım ve teknolojiye bağlı gelişmeler sayesinde olmuştur. Ne var ki bu gelişim süreci, potansiyel kullanıcılar üzerinde MİY’de sadece yazılımdan ibaret olduğu gibi yanlış bir düşünceye sebep olmuştur. Oysaki çoğu yazılımda olduğu gibi projelerin başarısında insan faktörü çok önemlidir.

Ancak yine de işletmeler deęişime ayak uydurabilmek, yeni stratejiler geliştirebilmek ve bunları uygulayabilmek için alt yapılarını, teknolojilerini ve enformasyon sistemlerini bütünleştirebilen sistemlerle donatmak zorundadırlar (Taşpınar, 2006: 24-36).

1.2.5. MİY'nin süreçleri

MİY'de süreç, ilk olarak müşteriler hakkında olabildiğince bilgi toplamaya dayanır. Bu anlamda daha önce belirtildiği gibi bu sistem (MİY), yaratılan felsefenin arkasında yüksek bir teknolojiye ihtiyaç duyulmaktadır. Elde edilen verilerin ışığında müşterilerin gruplandırılması, karakteristik özelliklerinin çıkarılması ve son olarak doğru ürünün doğru müşteriye yönlendirilmesini sağlar (Güldür, 2007). Bu süreç aşağıdaki şekilde açıklanmıştır.

1.2.5.1. Müşteri seçimi

MİY sürecinin bu aşamasında özellikle hedef kitlesinin belirlenmesi, müşterilerin sınıflandırılması ve ana amacı işletmeye kar sağlayacak müşterinin bulunmasıdır. Özellikle MİY kampanyasının, temel planının oluşturulduğu ilk aşama olması açısından çok önemlidir. Müşteri sınıflandırma süreçlerinde amaç en karlı müşteri grubunu tespit edip bu gruba yönelik kampanyaların planlandığı aşamadır. Bu aşamada:

- Hedef Kitlenin Belirlenmesi: işletmenin hitap etmek istediği müşteri yaş ve gelir aralığı belirlenmektedir.
- Bölümlendirme: Belirlenen hedef kitlenin müşteri profili belirlenerek, kitle yaş, gelir ve müşteri profiline göre bölümlere ayrılmaktadır.
- Konumlandırma ve Marka/Müşteri Planlamaları: Müşteri profiline uygun marka ve ürün konumlandırmaları yapılmaktadır.
- Müşteri kazanılmasını sağlayacak çeşitli kampanyalar hazırlanmaktadır.
- Yeni Ürün Lansmanları: Belirlenen müşteri ihtiyaçları doğrultusunda yeni ürünler tasarlanarak, pazara sunulmaktadır.

1.2.5.2. Müşteriyi elde tutma

Müşteriyi elde tutma aşamasında önceden ürün ve hizmet satın almış müşterinin elde edilmesi aşamasıdır. Bu aşama müşteriyi kaybetmemek ve dolayısıyla onun için veya onun yerine gelecek müşteriyi kazanmak için yapılacak olan işlemlerin ve maliyetlerin önlenmesini sağlar. Bu amaç için, müşterinin beklentileri iyi takip edilmeli ve tekrar benzer ürünlere yönelmesi sağlanmalıdır. Müşterinin şirket bazında algılanması ve tanınması, kişisel bilgilerinin defalarca sorulmasının engellenmesi, müşteriye doğru zamanda ve doğru kanaldan ulaşılmasının sağlanması bu aşamada dikkat edilmesi gereken diğer hususlardır.

1.2.5.3. Müşteriyi edinme

Müşteri edinme evresinin amacı ise satış gerçekleştirmektir. Müşteri seçimi evresinde belirlenen işletmeye kar sağlayacak müşterilere satış yapma stratejileri oluşturulmalıdır. Bu çerçevede ihtiyaç analizleri yapılarak, satış sağlayacak teklifler müşterilere sunulmaktadır (Güldür, 2007).

Seçilen müşterinin kazanılması için yapılan tüm çalışmalar, özellikle satış gayretleri bu eksene yerleşmektedir. Bu evrenin asıl amacı satışı gerçekleştirmektir. “Belirli bir müşteriye en etkili yoldan satış nasıl yapılır?” sorusunun yanıtı bu evrenin amacını oluşturmaktadır. İhtiyaç analizleri, teklif oluşturma gibi adımlar bu evrede yapılan eylemler olarak tanımlanabilir (Aksoy, 2002).

1.2.5.4. Müşteri büyütme ve derinleştirme

MİY'nin son evresi “Müşteri Derinleştirme” evresidir. Pazar payı anlayışı yerine, artık sadık hale getirdiğiniz müşterinin cüzdanındaki payını arttırmak için yapılan çalışmalar da bu evrede odaklanmaktadır. Kazanılmış bir müşterinin sadakat ve karlılığının uzun süre korunması ve müşteri harcamalarındaki payın yükseltilmesi için gereken adımları içerir.

Bu süreçte var olan müşterilere, ürün satışını arttırmak için çapraz satış kampanyaları kullanılır. Bu kampanyalar, müşterinin veri tabanındaki bilgileri kullanılarak müşteri ihtiyaçlarını tespit edip, ona uygun ürünü müşteriye sunma sürecidir.

1.2.6. MİY'nin amaçları

MİY tanımında da belirtildiği gibi MİY pazarlama stratejisidir. Strateji ile anlatılmak istenen rakiplerden farklı olabilmek için uygulanması gereken yöntemler bütünü ifade etmektedir. MİY'nin amaçlarını şöyle sıralamak mümkündür: (Kırım, 2001: 104)

- Müşteri ilişkilerini karlı hale getirmek
- Farklılaşma sağlamak
- Maliyetin azalmasını sağlamak
- İşletmenin verimini sağlamak
- İşletmenin verimini arttırmak
- Uyumlu faaliyetler sağlamak
- Müşteri taleplerini karşılamak

Bu amaçlara istinaden firmaların amaçları ise iki türdür; şirketin ulaşabileceği en çok sayıda karlı müşteriye ulaşması, diğeri ise daha az sayıda karlı müşteriye ulaşmasıdır. İkinci amaç firmaların karlarını maksimize etmek istediklerinde ortaya çıkar. Dell Bilgisayar, Chevrolet ve Docker's ulaşabildikleri en çok sayıda karlı müşteriye ulaşma amacını benimserken; Apple Bilgisayar, Porsche ve Armani daha az sayıda karlı müşteriye ulaşma amacı benimsemektedirler (Bergeron, 2002).

1.2.7. MİY'nin sağladığı avantajlar

MİY işletme ve müşterinin ilişkisi üzerine kurulmuştur. Etkili olabilmesi için müşteri merkezli ve müşteri ile iç içe olması gerekmektedir. MİY müşterinin ne zaman nerede ve nasıl bir şekilde işletme ile bağlantıya geçildiğini göstermelidir (Özgener ve İraz, 2006). Müşteri ilişkilerinin sağladığı avantajlar şu şekilde sıralanabilir:

- **Müşteri Tatmini:** Müşterinin tatminini sağlamak ve sadakatini arttırmayı amaçlamaktadır. İşletmelerin ana amacı, müşterinin sadakatini yukarıya taşımayı amaçlamaktadır. Müşteriye daha iyi hizmet vermek için müşteriyi tanımak, ona özel hizmetler sunmak, şikayetleri dinlemek ve onları çözüme ulaştırmak, sadakatini ödüllendirmek ve bunların dahilinde müşteri tatminini arttırmak MİY'nin başlıca görevlerindedir(Uçkan ve ark., 2003). Ayrıca MİY'nin bir diğer görevi de işletmeyi terk etmiş eski müşterilerin yeniden kazandırılmasıdır. MİY'nin bu duruma katkısı %10 ile %20 arasında değişmektedir (Duran, 2006).
- **Maliyetleri Düşürme:** Mevcut müşterilere ek yapılacak ilave satışlar, müşteriyi muhafaza etmekle birlikte müşteri payını artırma olanakları toplamı; bu toplamdan satış maliyetlerinden ve pazarlama iletişimi faaliyetlerinden sağlanacak tasarruf ve işletme içi maliyetlerin toplamı çıkarıldığında maliyetlerdeki azalış ortaya çıkacaktır (Acuner, 2005).
- **Karlılığı ve Gelirleri Arttırmak:** MİY işletmelerin değerli müşteri oranını yükselterek, kar oranına direkt katkı sağlanmaktadır. MİY'nin en temel amaçlarından birisi, müşteriyi yaşam boyu elde tutmak, onları tanıyarak değişen ihtiyaçlarına göre daha farklı veya fazla nasıl satabileceğini düşünmek. Çapraz satış olanaklarını arttırmaktır (Acuner, 2005).
- **Daha iyi müşteri hizmeti:** MİY uygulamaları ile müşterileri bölümlere ayırıp bu yolla onları yakından takip etmek mümkündür. Hedef müşterilere daha iyi hizmet sunma ve müşteri sadakati sağlama imkânı elde edilmiş olunacaktır (Acuner, 2005).

Sonuç olarak MİY hem işletmeler açısından hem de müşteriler açısından birçok avantaj sağlamaktadır. Müşterilerini tanıyabilen, müşteri ilişkilerini yönetebilen işletmeleri MİY'den sağladıkları kazançlar ellerindeki somut verilere de bakılarak anlaşılabilir. MİY uygulayan işletmelerde ciro, maliyet ve kalite durumu şu şekilde olmaktadır (Aydın, 2006).

Ciro

- ✓ Çapraz satışlarda %0,5 ile %5 arasında artış
- ✓ Müşteri başına ciro artışı %20 kadar

Maliyet İyileşmesi

- ✓ Pazarlama kampanyalarını geliştirme maliyetlerinde %10'a kadar iyileşme

Kalitede İyileşme

- ✓ Tekliflere geri dönüş oranlarında %10'dan %100'ün üzerine çıkan oranlarda artış
- ✓ Müşterilerin tutulması : %5 ile %25 oranında artış
- ✓ Müşteri tatmini :%10'dan %50'nin üzerine çıkan oranlarda artış görülmektedir.

1.3. MİY'de Veri Madenciliği

MİY, müşteriler hakkında detaylı bilgiye ulaşmayı zorunlu kılan ve bu sayede müşteriler ile sürekli ve birebir ilişki kurmayı sağlayan stratejik bir yaklaşımdır (Kırım, 2001: 82). İyi bir MİY sistemi müşterileri en iyi şekilde tanımayı hedefleyenler. MİY, müşterilerin dilek ve beklentilerini belirledikten sonra, buna uygun strateji geliştirmeyi de kapsamaktadır. Günümüzde VM işletmeler için çok önemli bir unsurdur. Doğru bilgiye ulaşmak, doğru raporlamalara ve doğru sonuçlara varılmak isteniyorsa, kaliteli veriye sahip olmak gereklidir. Ancak VM uygulamaları ile büyük şirketlerin veri tabanlarında yer alan kayıtlar anlamlı bilgilere dönüştürülebilmektedir (Michael ve Gordon, 2000).

Günümüzde MİY oldukça yaygın olarak kullanılan bir kavram haline gelmiştir. İyi bir MİY sistemi müşterileri en iyi şekilde tanıyan ve nelerden hoşlanıp hoşlanmadıklarını

anlamayı gerektirmektedir. MİY müşterilerin isteklerini ve beklentilerini belirledikten sonra buna uygun strateji geliştirmeyi de kapsamaktadır. VM, MİY uygulamalarında oldukça önemli bir role sahiptir. VM uygulamaları kullanarak büyük şirketlerin veri tabanlarında yer alan kayıtlarla anlamlı bilgilere dönüştürülebilmektedir (Michael ve Gordon, 2000).

VM büyük verilerden anlamlı bilgiler edinilmesini ve veriler arasındaki ilişki belirlenmesini sağlamaktadır. Örneğin sepet analizi ile hangi müşterilerin hangi ürün birleşimlerini satın aldıklarının belirlenmesi bu konuda yapılan önemli bir çalışmadır. Bu analiz sonucu elde edilen sonuçlar kampanyalarda hedef kitlenin daha doğru belirlenmesi ve ürünlerin raflara yerleştirilmesi gibi kararlarda, önemli karar desteği sağlamaktadır. Pazarlama ve perakende sektöründe satış terminalleri ve kodlama sistemleri sayesinde veriler sistematik olarak toplanmaktadır. Mağaza alışveriş kartları ve kredi kartları sayesinde yapılan alışverişin müşteri ile ilişkilendirilmesi sağlanmaktadır.

İşletme için kaliteli bilginin tanınım belirlenmesi ve sistemin uygulayıcılarına iyi anlatılması gerekmektedir. Veri aktarma sürecinin işletmenin kurumsallığı ve devamlılığı için temel olduğu kullanıcılara çok iyi izah edilerek sistemde uygulanması gereken kurallar ve beklentiler iyi tanımlanmalıdır (Çiçek, 2007).

Bu uygulamaların gerçekleştirilebilmesi için özellikle verilerin toplanmasında pazara yönelik stratejik bilgilerin oluşturulmasında ve pazarlama planlamasında VM önemli bir araç olarak kullanılmaktadır (Şimşek, 2006). İşletmeler VM araçlarını kullanarak müşterilerinin tüketim davranışlarını, harcama kalıplarını ortaya çıkarırlar ve bu bilgileri geleceğe yönelik kararlarında ve stratejilerinde kullanırlar (Tezcanlar, 2007).

1.4. Literatürde Veri Madenciliği ve MİY

VM 1990'lı yıllarda ortaya çıkmıştır. Online veri tabanı olan Science Direct'te 1960'dan günümüze kadar bir literatür taraması yapıldığında VM ile ilgili olarak 1240'a yakın makale olduğu görülmektedir.

Aşağıda VM konusunda ve VM'nin MİY'de uygulamalarına ilişkin örnekler yer almaktadır.

2012

Rodpysh ve arkadaşları, tüm müşteri ilişkileri yönetimlerini ve veri madenciliğini tanımlamışlardır. Ayrıca teknik ve MİY'deki veri madenciliğini uygulamalarının arasındaki etkileşim faktörlerini araştırmışlardır. Bu amaçla VM ve MİY üzerine eski çalışmalar ve raporları baz alarak; VM'ne etkileri ve büyük veri tabanlarındaki değerli olabilecek müşterileri tespit edip ayrıca değerli olabilecek müşterileri çekerek geliştirmek ve daha ileri götürmek için değerli müşterilerdeki gizli verileri açığa çıkarmanın ve daha ileriye götürmenin yöntemlerini incelemişlerdir (Rodpysh ve ark., 2012)

Babu ve Bhuvanewari, VM ve müşteri ilişkilerinde VM tekniklerini kullanarak MİY sistemini verimliliğini geliştirmişlerdir. Temel olarak müşterilerin geçmişe dönük veri tabındaki bilgilerine dayanmaktadır, bu çalışmada müşterilerin şirkete getirdikleri karlılık ve işlem sıklıkları ele alınarak en uygun müşteriler seçilmiştir ve seçilen uygun müşteriler veri tabanındaki diğer müşterilerle karşılaştırmışlardır. Bunun sonucunda etkin bir verimli müşteriler etkin bir şekilde elde tutulmaya devam etmiş olduğunu ortaya koymuşlardır (Babu ve Bhuvanewari, 2012).

Wang ve Pang, VM'de müşteri ilişkilerini incelemek üzere bir alışveriş merkezinin işlem verilerinden yararlanarak VM algoritmalarından olan Apriori ve Fuzzy algoritmalarına uygulamışlardır. Bu çalışmanın sonucunda eski müşteri bilgilerinden faydalanarak yeni müşterileri çekmeye yönelik kriterlerini ve değerli müşterileri açığa çıkarmışlardır (Wang ve Pang, 2012).

Peyman ve Safanaz, BY (Bilgi Yönetimi), MİY ve VM aralarındaki ilişkiyi açığa çıkarmışlardır. Demografik veriler ışığında bu verileri istatistik uygulayarak aralarındaki ilişkileri gözlemlemişlerdir (Peyman ve Safanaz, 2012).

2011

Khajvand ve Tarokh, Müşterilere ait değerli bileşenleri bölümlere ayırarak analiz etmişler, müşterilerin daha çok son yıllardaki elde tutulma yıllarındaki artış sıralamasına göre gizli örüntüler keşfedilerek müşterilerin demografik özellikleri kümeleme algoritması kullanılarak saptanmıştır. Müşteriyi elde tutma yılını arttırmayı hedeflemişlerdir (Khajvand ve Tarokh, 2011).

Wang ve Wu, VM tabanlı Banka sektöründe müşterileri ilişkileri araştırmasında müşterilerin demografik özelliklerinden faydalanarak aylık kredi kartı alışverişlerinden finansal ürünler alınma sıklığı karar ağaçları algoritmalarından faydalanılarak keşfedilmeye çalışılmıştır (Wang ve Wu, 2011).

2006

Boginski ve Butensko, çapraz korelasyon ve kümeleme tekniklerini kullanarak hisse senedi piyasalarının yapısal özelliklerini ortaya koymuşlardır (Boginski ve ark., 2006).

Liu, Shue ve Lee, IPS (Internet Servis Sağlayıcı) endüstrisinde kullanıcıların genel ağ kullanma örüntülerini ortaya çıkarmak ve müşteri ilişkilerini güçlendirecek stratejiler geliştirmek amacıyla nitelik odaklı indirgeme yöntemini(AOI) kullanmışlardır (Liu ve ark., 2006).

2005

Liu ve Shih, AHP ve birliktelik kuralı yöntemlerini birleştiren bir öneri metodolojisi geliştirerek müşterilerin yaşam ömrü değerlerine göre yani müşterinin elde tutulma ömrüne göre bölümlere ayırmışlardır (Liu ve Shih,2005).

Huang, Chen ve Wu, kümeleme tekniklerini kullanarak dağıtım merkezleri için bir sipariş yönetim sistemi geliştirmişlerdir (Chen ve ark., 2005).

Chen, Chi ve Hwa, perakende endüstrisinde etkin pazarlama kampanyaları düzenlemek amacıyla davranışsal müşteri verileri madenciliği çalışmasını gerçekleştirmişlerdir (Lin ve McClean, 2001).

Wang ve Hung, veri madenciliği teknikleri ile bir telekomünikasyon şirketi için müşteri karlılık yönetim sistemini geliştirmişlerdir (Hung ve ark., 2005).

2004

Last ve Kandel, karar ağaçları algoritmasını kullanarak yarı iletken endüstrideki bir fabrikada üretim planlama uygulaması gerçekleştirmişlerdir (Last ve Kandel, 2004).

Hsieh, bir bankanın kredi kartı müşterileri için sinir ağları ve birliktelik kurallarını kullanarak bir davranışsal skorlama modeli yaratmış; müşterileri davranışlarını ve karakteristiklerine göre farklı gruplara ayırarak her grubun karakteristiklerine uygun yönetim stratejileri önermek suretiyle müşteri sadakatini artırmayı hedeflemiştir (Hsieh, 2004).

Hwang, Jung ve Suh, müşterinin geçmişte yarattığı potansiyel değeri ve rakip şirketi tercih etme olasılığını dikkate alarak, lojistik regresyon, sinir ağları ve karar ağacı tekniklerini kullanmak suretiyle bir müşteri yaşam ömrünü yani elde tutulduğu süreyi kullanarak modeli geliştirmişlerdir (Hwang ve ark., 2004).

2002

Cox ve Lewis, çelik endüstrisindeki bir fabrikada yaptıkları uygulamada yapay sinir ağları yönetimini kullanarak ürünün istenilen kalite standartlarını sağlamanın için gerekli girdi miktarlarını saptamışlardır (Cox ve ark., 2002).

Kim ve Lee, genetik algoritmaya dayalı çoklu sınıflayıcıları birleştirerek e-ticaret müşterilerinin satın alma davranışlarının tahmini için bir model geliştirmişlerdir (Kim ve Lee, 2002).

2001

Caskey, genetik algoritma ve sinir ađları teknikleri ile bir fabrikadaki alıřma kořullarını ortaya koymuř ve bu kořulları iyileřtirici iřletme stratejileri nermiřtir (Caskey, 2001).

Bayram (2001), veri madenciliđi tekniklerinden karar ađaları, kmeleme analizleri ve lojistik regresyon analizlerini kullanarak kablosuz iletiřim endstrisinde mřterileri segmentasyonunu ve ayrılacak mřterilerin tahmin modelini ortaya koymuřtur (Bayram, 2001).

2. MATERİYAL VE YÖNTEM

2.1. Veritabanı ve Veri Ambarı

2.1.1. Veritabanı sistemleri

Günlük hayatta veri ile bilgi eş anlamlı kullanılmaktadır. Ancak düzenlenmiş bir veri bilgiye dönüşmektedir. Veriler kendi başına değersizdir. Bilgi, bir amaca yönelik işlenmiş bir veridir. Veriler eskiden dosyalar halinde tutulmaktaydı ve karmaşık dosya yapıları, çok sayıda dosyalar arasındaki ilişki ve kullanıcıların dosyalara erişimi söz konusu olduğunda geleneksel dosya sisteminin yetersiz kaldığı görülmüştür. Bu sorunu çözmek için günümüzde veriyi saklama ve erişim konusunda yeni yazılım teknolojilerine yönelme başlamış ve veritabanı sistemlerini oluşturmak ve veriyi yönetmek üzere VYS (Veritabanı Yönetimi Sistemleri) ortaya çıkmıştır. (Özkan, 2008: 14,20).

2.1.2. Veri ambarı

Veritabanı sistemlerinin karar destek uygulamalarında gereksinimleri karşılamakta yeterli olmaması nedeniyle, gereksinime bağlı olarak verinin farklı biçimde saklanması ve veriye hızlı erişimin sağlanabilmesi için farklı yöntemler aranmaktadır. Söz konusu veriler bazen o kadar büyüktür ki verileri geleneksel veri tabanı yöntemleri ile işlemek olanaksızdır. Bunun sonucu olarak “Veri Ambarı” kavramı ortaya çıkmıştır. Belirli bir döneme ait çalışmaya yönelik konu odaklı olarak düzenlenmiş ve birleştirilmiş ve sabitlenmiş işletmelere ait veri tabanlarına veri ambarı denilir. Veri ambarı (data warehouse) ilişkili verilerin sorgulanabildiği ve analizlerinin yapılabildiği bütünleşik bir bilgi deposudur (Özkan,2008; Döşlü, 2008; Silahtaroglu, 2008:14,20).

2.1.3. OLTP(On-line Transaction Process)

Bir kurumun günlük verilerinin işlendiği ortamlara OLTP (On-line Transaction Process) sistemler adı verilmektedir. Örneğin bir işletmenin sahip olduğu stok sistemi ile depoya giren çıkan ürünler ve ödemeleri izlenebilir. Stoklarla ilgili tüm işlemler veri tabanına kaydedilir. Bu kayıtlara dayalı olarak çeşitli belge ve raporlar üretilir. OLTP

sistemlerine ilişkin veri tabanlarına veri kaydedilebilir, veriye erişilip raporlanabilir veya istendiğinde silinebilir. Veri ambarları, veri tabanlarından daha fazla veriyi saklama kapasitesine sahiptir. OLTP uygulamalarına olanak sağlamıştır. OLTP uygulamaları veri tabanlarından veri silme, değiştirme ya da veri ekleme gibi işlemleri içermektedir. Bu işlemler SQL(Structured Query Language) sorgulama dili ile gerçekleştirilmektedir.

OLTP, operasyonel verinin veri tabanlarına aktarılmasına yardımcı olan bir işlem sürecidir. Bu süreç verilerin kayıt edilmesinden güncellenmesine kadar uzanan bir süreçtir. Veri ambarından ziyade veri tabanı üzerinde gerçekleşen işlemler, genel olarak veri yüklenmesi aşamasında kullanılmaktadır. OLTP'nin amacı veri setinin en az hata ile veri tabanına aktarılmasını amaçlamaktadır.

2.1.4. OLAP(On-line Analytic Processing)

Veri ambarları daha çok karar destek sistemlerine yönelik veri saklama araçlarıdır. VM, OLAP (On-Line Analytical process) gibi uygulamalara olanak sağlamaktadır. OLAP verilere analiz yapılmasına, raporlanmasına olanak sağlayan işlemler bütünüdür. Bu süreç tamamen veritabanlarından bağımsız olarak gerçekleşir. Bütün analizler veri ambarı üzerinde yapılarak çevrimiçi işlem sürecini oluşturur. Bu süreç sadece karar vericiye daha sağlıklı karar vermesine olanak sağlayan tüm analizleri veri setine uygulama olanağı vermektedir (Kasap, 2007).

OLAP hepimizin sıklıkla kullanmış olduğu ilişkisel veritabanlarında depolanan enformasyondan çok boyutlu veri küpü kurmaya olanak sağlayan bir teknolojidir. Kullanıcılar verileri karmaşık problemlerde cevap vermek üzere kullanmaktadır. OLAP'ın kullanım alanına dair bir örnek verecek olursak; örneğin bir şirket kendi ürünlerini belirlediği periyotlarda (hafta, onbeş gün, ay, mevsim, yıl vs.) hangi şubelerin, bayilerin ürünlerinden ne kadar sipariş ettiğini, sattığını ve bunun gibi birçok sorulara OLAP yardımıyla cevap verebilmektedir (Kocabaş, 2012).

OLAP Veritabanlarının Özellikleri :

1. Çok kullanıcı desteği
2. Performans

3. Esnek raporlama
4. Çapraz rapor olanağı
5. Sorgulamalarda aynı performansı gösterebilme
6. Boyutlarda sınır olmaması

Veri Martı: Veri ambarlarının konulara göre düzenlenmiş ve işletmenin satış veya pazarlama gibi sadece belirli bir bölümünü ilgilendiren, başka bir deyişle tek bir konu ya da bölüme odaklanan parçasına veri martı denir. Veri martı özel amaçlar için üretilir ve kullanılır (Silahtaroglu, 2008: 18).

Han ve Kamber'e göre veri ambarları ve OLAP sistemi çok boyutlu veri modelleri üzerine kurulmuştur. OLAP işlemleri arka planda bilgi ile ilgili öğrenilebilir verilerin çıkarılması ve verilerin farklı seviyelerde özetlenmesine olanak sağlar. Çok boyutlu veri modelleri veriyi veri küpü(data cube) formunda incelemektedir. Veri küpü, verinin çok boyutta modellenmesini ve incelenmesini sağlar ve olgu tablosundan beslenen boyutlar(dimension) ve ölçümler(measure)'den oluşur. Boyutlar organizasyonun kayıtlarını tutmak istediği perspektifler veya varlıklar ile ilgili iken ölçümler rakamsal değerler, elde etmek istenen sonuçlardır (Han ve Kamber, 2006: 14).

OLAP işlemleri; veri seçimi, veri küpü kurma ve veri küpü üzerinde çevrimiçi analizin yapılacağı uygulama olmak üzere üç aşamadan oluşmaktadır. Çok boyutlu veri küpleri üzerinde birçok OLAP işlemi uygulanmaktadır. Bu işlemler aşağıdaki gibidir:

- Detay azaltma (Drill-down)
- Detaylandırma (Roll-Up)
- Dilimleme ve parçalama (Slice and dice)
- Çevirim (Pivot) vb.

Bir veri ambarının olması, OLAP'a ihtiyaç olmadığı anlamına gelmez. Veri ambarları ve OLAP birbirini tamamlar. Veri ambarı verileri uygun şekilde tutmaya ve kontrol etmeye yarar. OLAP ise, veri ambarı verilerini stratejik bilgilere dönüştürmeye yarar.

2.2. Veri Madenciliği

2.2.1. Veri madenciliği kavramı ve gelişimi

VM, büyük veri ambarlarında tutulan verilerden otomatikleşmiş modeller sayesinde anlamlı bilgileri, ilişkileri ve davranışları ortaya çıkarma süreci olarak tanımlanmaktadır. Bu süreçte, veri içinde önceden pek fazla bilinmeyen veya görünmeyen desenler(pattern) öncelikle ortaya çıkarılmaktadır. Bu desenler genellikle bilgiler arasındaki ilişkilerin, sınıflandırmanın, veri birlikteliğinin ve tahminlemenin sonucunda elde edilmektedir (Mitra ve Archarya, 2003: 24).

Özkan'a göre VM, kurumlarda biriken veri içerisinden kurum için yararlı olanları bulup ortaya çıkarma işine denir (Özkan, 2008: 37).

VM, verileri açığa çıkarma sürecidir. Bu süreçte, veri içinde önceden bilinmeyen veya görülemeyen desenleri ortaya çıkarılmaktadır. Bu desenler genellikle bilgiler arasındaki ilişkilerin, sıralamanın, sınıflandırmanın, veri birlikteliğinin ve tahminlemenin sonucunda elde edilmektedir (Chu, 2005).

İşlenmemiş bir verinin bilgiye çevrilmesi zamanımıza kadar uzanan bir problemdir. Günümüzde hızla gelişen teknoloji ile birlikte yazılımlar sayesinde veriler çok hızlı bir şekilde depolanabilmektedir. Bu depolanan verilerden anlamlı verilerin çıkarılması da VM'nin görevidir. VM, gelecekle ilgili tahminin yapılmasını sağlayan bağlantı ve kuralların yazılımlar kullanılarak bulunmasıdır. Yakın geleceğin geçmişten çok fazla farklı olmayacağı varsayılırsa, geçmiş veriden çıkartılmış olan kurallar gelecekte de geçerli olmaya devam edecek ve ileriye dönük tahminlerin yapılmasına olanak sağlayacaktır (Alpaydın, 2000).

VM kullanılarak büyük veri yığınlarından oluşan veritabanı sistemleri içerisinde gizli kalmış bilgileri açığa çıkarmak mümkündür. Bu işlem, istatistik, matematik disiplinleri, modelleme teknikleri, veri tabanı teknolojisi ve çeşitli bilgisayar programları kullanılarak yapılır. VM büyük miktarda veriyi incelediğinden dolayı veri tabanları ile yakından ilişkilidir. Gerekli verilerin gerektiği zamanda hızla erişilebilmesi gerekmektedir. Günümüzde yaygın olarak kullanılan veri ambarları, günlük kullanılan

verilerin işlemeye uygun hale getirilmiş özetini veri tabanlarında saklamayı amaçlar (Baykal, 2006: 96).

VM, verilerin içerisindeki desenlerin, ilişkilerin, değişimlerin, düzensizliklerin, kuralların ve istatistiksel olarak önemli olan yapıların yarı otomatik olarak keşfedilmesidir. Veriler arasındaki ilişkiyi, kuralları ve özellikleri belirlemekten bilgisayar sorumludur. Amaç daha önceden fark edilmemiş veri desenlerini tespit edebilmektir (Dener ve ark., 2009).

VM'den kazanılan bilgi çeşitli market analizleri, sahtekârlık algılama, müşteri tutma, üretim kontrolü için ve bilim keşif uygulamaları için kullanılabilir. VM, bilgi teknolojisinin doğal bir evriminin sonucu olarak görülebilir.

2.2.2. Veri madenciliğinin uygulama alanları

VM farklı iş sektörlerinde yaygın olarak kullanılmaktadır. Örneğin pazarlama, bankacılık, sigortacılık ve elektronik ticaretle ilgili alanlardır. VM'nin uygulama alanları aşağıdaki gibi sıralanmıştır (Baykal, 2006: 95-107; Özkan, 2008: 37-39).

Pazarlama

- Müşterilerin satın alma örüntülerinin tespit edilmesi
- Müşterilerin demografik özellikleri arasındaki bağlantıların ortaya çıkarılması
- Posta kampanyalarında cevap verme oranının artırılması
- Mevcut müşterilerin elde tutulması, yeni müşteri kazanılması
- Pazar sepeti analizi
- MIY
- Müşteri değerlendirme
- Satış tahmini

Bankacılık

- Farklı finansal göstergeler arasında gizli ilişkilerin bulunması
- Kredi kartı dolandırıcılıklarının tespiti
- Kredi kartı harcamalarına göre müşteri gruplarının belirlenmesi

- Kredi taleplerinin deęerlendirilmesi

Sigortacılık

- Yeni poliçe talep edecek müşterilerin tahmin edilmesi
- Sigorta dolandırıcılıklarının tespiti
- Riskli müşterilerin örüntülerinin belirlenmesi

Elektronik Ticaret

- Saldırıların çözümlenmesi
- E-MİY uygulamalarının yönetimi
- Web sayfalarına yapılan ziyaretlerin çözümlenmesi

2.2.3. Veri madencilięi sürecinin aşamaları

VM altı adımlı fazda incelenebilir (Larose, 2005).

➤ Araştırma Probleminin Tespit Fazı (Business Understanding Phase):

Bu aşama VM'nin en önemli aşamasıdır. Araştırma probleminin konusunun tanımlanması aşamasını, araştırmanın amacını, mevcut durumunun deęerlendirilip veri madencilięinin amaçlarını ve proje planlanma sürecinin belirlenmesini kapsar.

➤ Verileri Tanıma Fazı (Data Understanding Phase):

Veri anlama aşaması veri toplamakla başlamaktadır daha sonra bezer verileri bir araya getirip veri niteliklerini tanımlama, verileri keşfetme gizli bilgileri sınıflandırma ile sürece devam edilmektedir.

➤ Veri Hazırlama Fazı (Data Preperation Phase):

Veri hazırlama aşaması, ham veriden başlayarak son veriye kadar yapılması gereken düzenlemeleri içermektedir. Veri hazırlama, tablo, kayıt, veri dönüşümü ve modelleme araçları için temizleme gibi özellikleri içermektedir.

➤ Modelleme Aşaması Fazı (Modelling Phase):

Bu aşamada, verilerden bilgileri çekmek için ileri çözümlenme yöntemleri kullanıldığından VM sürecinin en karmaşık aşamasıdır. Bu aşama uygun modelleme

tekniklerinin seçimi, test tasarımının üretimi, model geliştirme ve tahmin işlemlerini içermektedir. Uygun modellerin seçilip uygulanmasıyla birlikte parametreler en uygun değişkenlere dönüştürülmektedir. Bazı yöntemler verilerin tipi için uygun değildir ya da özel tanımlamalar gerektirmektedir. Bu nedenle gerekli olduğunda 3.faz olan veri hazırlama fazına geri dönlür.

➤ **Değerlendirme Fazı (Evaluation Phase):**

Değerlendirme aşamasında, uygun model ya da modeller kurduktan sonra, VM sonuçlarının araştırma probleminin amaçlarını gerçekleştirip gerçekleştirilmediği gözden geçirilir. Bu aşamada sonuçların değerlendirilmesi, VM sürecinin gözden geçirilmesi ve sonraki adımların ne olacağını içermektedir. Bu aşamanın sonunda VM sonuçlarının kullanımı üzerindeki karara varılmaktadır.

➤ **Uygulama Fazı (Deployment Phase):**

Son aşama olan uygulama aşaması, araştırmacının tüm emeklerinin karşılığını aldığı aşamadır. Bu aşamada VM süreciyle üretilen bilgiler, pratik işletme problemlerinin çözümünde kullanılmaktadır. Bu aşamada elde edilen bilgilerin uygulanabilmesine yönelik bir plan hazırlama, gözden geçirme ve bakım faaliyetlerini içerir. Ayrıca bu aşamada araştırma raporunun yazılması ve projenin gözden geçirilmesi işlemleri yer almaktadır.

2.3. Veri Madenciliği Teknikleri

VM teknikleri, verilerin kullanılma amaçları ve veri yapısına göre farklılaşmaktadır. Han ve Kamber'e göre veri madencilik görevleri ikiye ayrılır: Tahmin ve Tanımlayıcı olarak (Han ve Kamber, 2001).

Bu teknikler temel olarak iki ana grupta toplanmaktadır: (Joseph, 1996)

- Tahmin Edici(Predictive) Model
- Tanımlayıcı(Descriptive) Model

Uygulamalarda hangi modelin kullanılacağına veriye bakılarak karar verilir. Kullanılan model tanımlayıcı, tahmin edici veya ikisi birden olabilmektedir.

Tahmin edici model, verilerin içinden bulunmuş olan desenden yola çıkarak gelecekteki değerini tahmin eder (Levin ve Zahavi, 1999).

Tanımlayıcı model, denetimsiz olarak öğrenme yöntemine dayanmaktadır. Bu tür modeller veri tabanını sorgulayarak, tabanındaki veri ilişkileri ve modellerini bulmaya dayanmaktadır (Han ve Kamber, 2001).

2.3.1. Tahmin edici model

Tahmin edici model, yeterince verinin varlığında veri kalitesiyle ilgilenmeden, tanımlayıcı modellerin keşfedilmesi için olanak sağlar. Kötü veriler tahmin edici model ne kadar iyi olursa olsun verimsiz modeller üretilmesine sebep olur (Guazelli, 2012).

2.3.2. Tanımlayıcı model

Tanımlayıcı model tipik denetimsiz öğrenmeye dayanan VM modelidir. İşaretlenmemiş verilerden örüntü ilişkileri çıkarmaya dayanır. Veri analizinden çıkarılan ilişkiler araştırmalar ile ilgi keşif yapılmasını sağlar (Fürnkranz ve ark., 2012).

Tanımlayıcı modellerde karar vermeye rehberlik etmede kullanılacak mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır. X/Y aralığında geliri ve iki veya daha fazla arabası olan çocuklu ailelerle, çocuğu olmayan ve geliri X/Y aralığından düşük olan ailelerin satın alma örüntülerinin birbirlerine benzerlik gösterdiğinin belirlenmesi tanımlayıcı modellere bir örnektir (Akpınar, 2000: 2).

2.4. Veri Madenciliği İşlevleri

VM modellerini işlevlerine göre (Tsiptsis ve Chorianopoulos, 2009)

- Sınıflama(Classification) ve Regresyon(Regression)
- Kümeleme(Clustering)
- Birliktelik Kuralları(Association Rules)

Olmak üzere üç ana başlık altında inceleyebiliriz. Sınıflama ve regresyon modelleri tahmin edici, kümeleme ve birliktelik kuralları tanımlayıcı modellerdir (Akpınar, 2000).

Veri madenciliğinde sınıflandırma, kümeleme ve birliktelik kuralları analizi gibi konularda kullanılmaktadır üzere birçok algoritma geliştirilmiştir. Bu yöntemlerin birçoğu istatistiksel tabanlıdır. Bu algoritmaların bazıları sadece sınıflandırma ya da kümeleme gibi konuları ilgilendirirken bazıları ise değişik varyasyonlarla birden fazla konuda kullanılabilir. Örneğin, genetik algoritmalar, yapay sinir ağları gerek sınıflandırma ve gerekse de kümeleme modellerinde kullanılabilirler; Oysa Apriori algoritması sadece birliktelik kurallarının belirlenmesinde kullanılan bir algoritmadır (Agrawal ve Srikant, 1994).

2.4.1. Sınıflama ve Regresyon

Sınıflama ve regresyon, önemli veri sınıflarını ortaya koyan veya gelecek veri eğilimlerini tahmin eden modelleri kurabilen iki veri analiz yöntemidir (Han ve Kamber, 2000). Sınıflama, verinin önceden belirlenen çıktılarına uygun olarak ayrıştırılmasını sağlayan bir tekniktir. Çıktılar önceden bilindiği için sınıflama veri kümesini denetimli olarak öğrenir (Giudici, 2000). Örneğin, bir A sigorta şirketi; müşterilerinin yeni bir sağlık sigortasının ürünü ile ilgilenip ilgilenmediğini öğrenmek istemektedir. Daha önceden benzer bir ürünü satmış ve geçmiş veriler hangi müşterilerin önceki teklife cevap verdiklerini göstermektedir. Amaç; bu teklife cevap veren müşterilerin özelliklerini belirlemek ve böylece pazarlama ve satış çalışmalarını daha etkin yürütmektir.

Sınıflama kategorik değerleri tahmin ederken, regresyon süreklilik gösteren değerlerin tahmin edilmesinde kullanılır (Han ve Kamber, 2000). Belirli kriterlere göre önceden tanımlanmış olan sınıfların, veritabanı kayıtları incelenerek sınıflandırılma ile gelecekteki müşterilerin davranışlarını tahmin etmek için model oluşturması hedeflenir. Kullanılan ortak araçları sınıflandırılmak için yapay sinir ağları, karar ağaçları kullanılabilir. Sınıflama modeli, geleceğe dönük müşteri davranışlarını tahmin için MİY de en çok kullanılan bir modeldir (Ngai ve ark., 2009).

Sürekli değişkenin tahmini Regresyonun istatistiksel tekniği ile modellenebilir. Birçok problem lineer regresyon ile çözümlenmiştir. Daha fazla değişkenlere dönüşümleri

uygulana bilir, böylece doğrusal olmayan bir problem doğrusal hale dönüştürülür (Han ve Kamber, 2001).

Regresyon analizi herhangi bir bağımlı değişkenin birden fazla değişkenle arasındaki ilişkinin matematik bir fonksiyon şeklinde yazılması ve bu fonksiyon yardımı ile bağımlı değişkenin ulaşacağı değerin tahmin edilmesidir (Sykes, 2013).

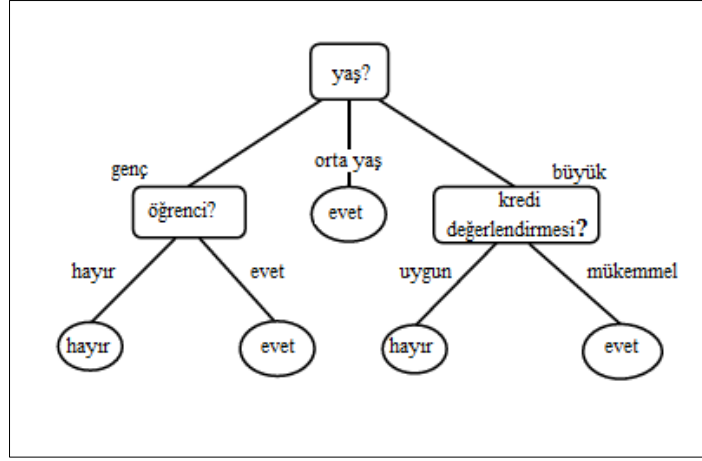
Sınıflama ve regresyon modellerinde kullanılan başlıca teknikler şunlardır: (Akpınar, 200: 1-22)

1. Karar Ağaçları (Decision Trees)
2. Yapay Sinir Ağları (Artificial Neural Networks)
3. Genetik Algoritmalar (Genetic Algorithms)
4. Bellek Temelli Nedenleme (Memory Based Reasoning)
5. K-En Yakın Komşu (K-Nearest Neighbor)
6. Naive-Bayes
7. Diskriminant Analizi
8. Regresyon Analizi

2.4.1.1. Karar ağaçları (Decision Trees)

Karar ağaçları yorumlarının kolay olması, veri tabanı sistemleri ile kolayca entegre edilebilmeleri ve güvenilirliklerinin iyi olması nedenleri ile sınıflama modelleri içerisinde en yaygın kullanıma sahip tekniktir. Karar ağaçları, bir ağaç görünümünde, tahmin edici bir tekniktir (Berry ve Linoff, 1999).

Şekil 2.1’de örnek bir karar ağacı görülmektedir. Bu karar ağacı bilgisayar satın alınma ya da almama durumunu temsil etmektedir. Yani tüm elektronik eşya alan müşterilerin bir bilgisayar satın alma olasılığı olup olmadığını tahmin amaçlı kurulmuştur. Karar ağacında Şekil 2.1’de görüldüğü üzere iki sınıf bulunmaktadır. Her yaprak düğümü bir sınıfı oluşturmaktadır. Bilgisayar satın alma durumu “evet”, bilgisayar satın almama durumu “hayır” ile temsil etmektedir. Müşterinin özelliklerine göre, hangi sınıfa dahil olduğu karar ağacında kök ve yaprak düğümler takip edilerek elde edilebilir (Han ve Kamber, 2000).



Şekil 2.1. Bilgisayar Satın Alma Kavramı Üzerine Karar Ağaçlarının Uygulanması

Karar ağacı, karar vericinin en uygun olan karara ulaşması yolunda yapılan gerek olasılık, gerekse maksimum faydanın sağlanmasına esas alınarak düzenlenen bir tekniktir. Karar ağaçları karar düğümleri, dallar ve yapraklardan oluşur. Karar düğümü, gerçekleştirilecek testi belirtir. Bu test sonucunda ağacın veri kaybetmeden dallara ayrılmasına neden olur. Her düğümden test ve dallara ayrılma işlemi ardışık olarak devam eder ve ayrılma işlemi üst seviyede bulunan ayrılmalara bağlıdır. Eğer bir dalın ucunda sınıflama işlemi gerçekleşmiyorsa, o dalın sonucunda bir karar düğümü oluşur. Ancak dalın sonunda belirli bir sınıf oluşursa, o dalın sonunda yaprak vardır. Bu yaprak, veri üzerinde tespit edilmek istenen sınıflardan biridir. Karar ağacı işlemi kök düğümden başlar ve yukarıdan aşağıya doğru olarak ardışık düğümleri takip ederek devam eder (Han ve Kamber, 2000).

Örneğin eğitim verisi incelenerek kredi durumu sınıfını tahmin edecek bir model oluşturulmak istendiğinde, bu modeli oluşturan bir sınıflama kuralı: **IF yas =“41..50” and gelir =yüksek then kredi_durumu=mükemmel** şeklinde olabilmektedir. Bu kural gereğince yaşı “41...50” yaş aralığında olan ve gelir düzeyi yüksek kişinin kredi durumunun “mükemmel” olduğu, yani mükemmel olarak adlandırılan sınıfa dahil olduğu görülür.

Oluşan bu modelin doğruluğunu test verisi aracılığı ile onayladıktan sonra, modeli sınıfı belli olmayan yeni veriye uygulanabilir ve sınıflama kuralı gereği yeni verinin sınıfı “mükemmel” olarak belirlenebilir. Özetlemek gerekirse karar ağacı, bir alandaki testi

belirten karar düğümlerinden, testteki deęerleri belirten dallardan ve sınıfı belirten karar düğümlerinden, akış diyagramı şeklindeki ağaç yapısıdır (Özekes, 2003: 65-71).

Karar ağacı analizi, genellikle seçenekler üzerinde yapılan bir analiz türüdür. Bu analizin veri madenciliğinde kullanılma sebepleri şunlardır (Chu, 2005):

- Maliyeti azdır.
- Anlaşılması ve yorumlanması kolaydır.
- Veri tabanlarına kolay entegre olabilir.
- Güvenilirliği yüksektir.

Karar ağacı temelleri analizlerinin yaygın olarak kullanıldığı alanlar şunlardır;

- Belirli bir sınıfın olası üyesi olacak elemanların belirlenmesi
- Parametrik modellerin kurulmasında kullanılmak üzere çok sayıda deęişkenlerden en önemlilerinin seçilmesi
- Sadece belirli alt gruplara özgü olan ilişkilerin tanımlanması
- Çeşitli deęişkenlerin yüksek, orta, düşük risk grupları gibi çeşitli kategorilerine ayrılması
- Kategorilerin birleştirilmesi ve sürekli olan deęişkenlerin kesikli deęişkenlere dönüştürülmesidir.

Karar ağacı oluşturmak için geliştirilen bu algoritmalar içerisinde CHAID (Chi-Squared Automatic Interaction Detector), Exhaustive CHAID, MARS (Multivariate Adaptive ve Regression Splines) ID3, CART, C4.5, QUEST(Quick, Unbiased, Effecient Statistical Tree),SLIQ (Supervised Learning in Quest), SPRINT (Scalable Paralleizable Induction of Decision Trees), CRT (Classification and Regression Trees) karar ağacı algoritmalarıdır. Karar ağacı algoritmaları yukarıdan aşağıya doğru böl ve yönet şeklindedir. (Han ve Kamber, 2000 ve Altay, 2005)

ID3

ID3 algoritması ilk olarak J.Ross tarafından Sydney Üniversitesinde geliştirilmiştir.ID3 makine öğrenmesi ve bilişim teknolojisine dayanarak, verilen örnekler içinde en ayırt edici olan değişkeni bulan bir algoritmadır. ID3, diğer değişkenler içinden sınıflamada en ayırt edici olan değişkeni bulurken entropiden yararlanır. Entropi eldeki bilginin sayısallaştırılmasıdır. Entropi belirsizliği ve beklenmeyen durumun ortaya çıkma olasılığını gösterir. Entropi bir veri kümesi içerisinde belirsizlik, şaşkınlık ve rastgeleliği ölçmek için kullanılmaktadır (Quinlan, 1986).

Bir veri tabanının tamamının entropisinin hesaplanabileceği çeşitli bölümlere ayrıldıktan sonrada hesaplanabilir. ID3 algoritması veri tabanı bölünmeden önce doğru sınıflandırma yapmak için gelen bilgiyle, veri bölündükten sonra, doğru sınıflandırma için gelen bilgi arasındaki farkı kullanarak, öncelikle düğüme ve dallanmalara karar verir. Bu aradaki fark ise kazanım olarak adlandırılır (Silahtaroglu, 2008: 51-52).

CART(C&RT)

CART veya C&RT (Classification and Regression Trees) Breiman, Friedman, Olshen ve Stone tarafından 1984 yılında geliştirilmiş binary (ikili) ağaç sistemi olarak büyüyen bir algoritmadır. C&RT en iyi dallara ayırma yöntemi için entropiyi kullanır. C & RT veriyi iki alt kümeye ayırır. Böylece bir sonraki adımda oluşacak olan alt küme, bir öncekinden daha homojen olur. Bu süreç sonuç bulunana kadar devam edecektir. Kendini tekrarlayan bir algoritmadır. Büyük verilerle çalışıldığında sonuç bulunması daha uzun sürmektedir. C&RT sınıflandırma ve regresyon analizi için kullanılan bir algoritmadır (SPSS, 2011).

CHAID

Chaid (Chi-Squared Automatic Interaction Detector) algoritması 1980 yılında Kaas tarafından geliştirilmiş bir karar ağacı tekniğidir(Kaas, 1980). Chaid algoritması ayırma kriteri olarak ki-kare'yi kullanır. Chaid algoritması, tahmin edici değişkenin tüm değerlerini dikkate alarak analiz eder. Hedef değişkeni dikkate alarak istatistiksel olarak benzer olan değişkenleri birleştir ve farklı olan değişkenle işlemini sürdürür. Daha sonra

karar ağacının ilk dalını oluşturmak için en iyi tahmin edici değişkeni seçer. Her bir değer seçilen değişkenin benzer değerlerinden oluşur. Bu süreç ağaç tamamıyla büyüyene kadar tekrarlanarak devam eder.

Chaid algoritması en popüler karar ağacı yöntemlerinden birisidir. CHAID algoritması ikili bir algoritma değildir. Bu yüzden ağaçta herhangi bir seviyede ikiden çok kategori üretilmesi anlamına gelir. Bu nedenle daha geniş ağaç yaratmaya eğilimlidir. Her tür değişken için kullanılan bir tekniktir (SPSS, 2011).

C4.5

C4.5, 1993 yılında Quinlan tarafından ortaya geliştirilmiştir (Quinlan, 1993). Karar ağacı oluşturulurken kayıp veriler hesaba katılmazlar. Kazanım oranı hesaplanırken sadece eksik olmayan kayıtlar kullanılır. Kayıp verileri diğer veri ve değişkenler yardımıyla öngörerek kazanım oranının hesaplanmasında kullanılır. Böylece daha anlamlı kurallar çıkarabilir

C4.5 algoritması, kalitatif (nitel analiz) değişkenleri dikkate alır. Ayrıca kayıp verileri diğer veri ve değişkenler yardımı ile tahmin ederek daha hassas ve daha anlamlı kurallar çıkartabilen bir algoritmadır (Doumpos ve Zopounidis, 2002).

QUEST

Quest (Quick, Unbiased, Efficient, Statistical Tree) 1997 yılında Loh ve Shih tarafından geliştirilmiş olan yeni bir tekniktir (Loh ve Shih, 1997). İkili büyüyen bir ağaç algoritmasıdır. Ayrı ayrı değişken seçimi ile ilgilenir. Quest'deki birim değişken aracı, tahmini olarak tarafsız değişken seçimini gerçekleştirir. QUEST algoritmasının C&RT algoritmasına benzer avantajları vardır, ancak ağaçlar yavaş büyüyebilir ve ikili olduğu için karar ağacı çok geniş olabilir (SPSS, 2011).

2.4.2. Birliktelik kuralları

Birliktelik kuralları, veri tabanlarındaki bilgiler arasındaki ilişkileri arayıp bulur ve veri tabanındaki her bilginin diğer kayıtlarla olan bağlantısını açıklar. Birliktelik kuralları veri madenciliğinin en önemli tekniklerindedir. Birliktelik analizi, bir veri kümesindeki kayıtlar arasındaki bağlantıları arayan denetimsiz(unsupervised) VM şeklindedir. Veri setinde birliktelik analizi eş zamanlı oluşum, olay ve durumların tespiti için kullanılan bir analizdir. VM sürecinde en olası modeli ortaya çıkarması açısından önemli analizlerden birisidir. Bu analizler genel olarak çıkardıkları kurallardan dolayı, müşteri davranışlarını tanımlama olanağı sağlamaktadır (Kantardzic, 2003:114).

Örneğin bir birliktelik kuralı analizi, müşterilerin satın aldıkları ürünler arasındaki birliktelik ilişki bağlarını bularak müşterilerin satın alma alışkanlıklarının tespit edilmesini sağlayabilir. Bu durumda oluşabilecek birliktelik kurallarına ait örnekler aşağıda yer almaktadır:

- Müşteriler bira satın aldıklarında %75 olasılıkla çocuk bezi de satın almaktadırlar.
- Düşük yağlı peynir ve yağsız süt alan müşteriler %85 oranla yağsız süt alırlar.
- Çadır alan müşterilerin %10'u bir ay içerisinde sırt çantası almaktadırlar.

Birliktelik kurallarının örnek uygulamaları aşağıdaki gibidir (Mitra ve Archarya, 2003: 267-290);

- Market - sepet analizi (market –basket analysis),
- Direkt satışta başka bir ürün önermek,
- Kredi kartı sahtekarlığını ortaya çıkarmak,
- Standların nasıl dizilmesi gerektiğini ortaya koymak

Birliktelik kurallarının matematiksel modeli 1993 yılında Agrawal, Imielinski ve Swami tarafından sunulmuştur (Agrawal ve ark., 1993).

Bu modele göre $I = \{i_1, i_2, \dots, i_n\}$ nesnelerin kümesi, D işlemler kümesi olarak ifade edilir. Her bir i , bir nesne olarak adlandırılır. D , veritabanındaki tüm hareketleri, T ise I kümesinde tanımlanan nesnelerin her bir hareketini temsil eder. TID ise, her harekete ait olan tek belirteçtir. Bir birliktelik kuralı $A \Rightarrow B$ formunda ifade edilir. Burada;

$$A \subset I, B \subset I \text{ ve } A \cap B = \emptyset \quad (2.1)$$

Birliktelik kuralı şu şekilde tanımlanabilir;

$$A_1, A_2, \dots, A_m \Rightarrow B_1, B_2, \dots, B_n \quad (2.2)$$

Öncelikle $A \Rightarrow B$ kuralı için d olasılığı ile kuralın destek değeri tanımlanır. Destek, T işleminin $A \cup B$ 'yi içerme olasılığı olasılığıdır. İkinci olarak, $Güven(A \Rightarrow B)$ kuralı için güven değerinin hesaplanması gerekmektedir. Bu olasılık, T işleminin A 'yı ve aynı zamanda B 'yi içermesidir.

Matematiksel ifade olarak kuralın destek ve güven değerleri;

$$Destek(A \Rightarrow B) = P(A \cup B) \quad (2.3)$$

$$Güven(A \Rightarrow B) = P(B/A) \quad (2.4)$$

$$Güven(A \Rightarrow B) = Destek(A \Rightarrow B) / Destek(A) \quad (2.5)$$

Şeklinde ifade edilir.

Burada destek ve güven değeri;

$$Destek(A) = |A|/|D| \quad (2.6)$$

$$Destek(A \Rightarrow B) = |A.B|/|D| \quad (2.7)$$

$$Güven(A \Rightarrow B) = \frac{Destek(A \Rightarrow B)}{Destek(A)} = P(A \cup B) / P(B) \quad (2.8)$$

olarak tanımlanır.

Burada;

|A| : İncelenen kayıtlardaki A ürününü içeren işlemlerin sayısı,

|A.B| : İncelenen kayıtlardaki A ve B ürünlerini birlikte içeren işlemlerin sayısını,

|D| : Veri tabanındaki bütün işlemlerin sayısını ifade etmektedir.

Kuralın destek ve güven değerleri, kuralın ilginçliğini ifade eden iki ölçüdür. Bu değerler sırasıyla keşfedilen kuralların yararlılığını ve kesinliğini ifade eder. Güven değeri, A'yı içeren kayıtların B'yi de içereceğini ifade eder (Ngai ve ark., 2009).

Bu durumu çok bilinen bir örnek ile açıklamak gerekirse: “Toplam kayıtlar içerisinde %5'i bira ve çocuk bezi ürünlerini bir arada aldığı, bira alan müşterilerin %30'unun çocuk bezi aldığı tespit edilmiştir”. Bu örnekte %30 güven seviyesini, %5 ise desteği belirtmektedir.

2.4.2.1. AIS algoritması

AIS (Agrawal, Imielinski ve Swami) algoritmasında, 1993 yılında geniş nesne kümeleri üretmek için geliştirilmiş bir algoritmadır. Veri tabanlarındaki nesne isimlerinin A'dan Z'ye sıralanması kısıtını taşır. Bu algoritma ile üretilen eşleştirme kurallarının sağ kesiminde sadece bir elemanlı ürünler kümesi yer alabilmektedir; $X \Rightarrow I_k$ biçimindedir (Agrawal ve ark., 1993).

AIS algoritması veri tabanını birçok kez tarar ve her tarama işleminde tüm işlemleri okur. İlk tarama esnasında veritabanındaki nesnelere tek tek sayarak hangilerinin geniş nesnelere olduğu belirlenir. Bu işlem tamamlandıktan sonra bir önceki işlemde geniş nesnelere olduğu belirlenen nesnenin kümeleriyle, o işlemin nesnelere arasındaki ortak nesnelere birleştirilerek yeni aday kümeler oluşturulur. Herhangi bir nesne kümesi bir işlemdeki nesnelere birleşip aday kümelerinden birini oluşturabilmesi için birleşebileceği nesnenin hem geniş olması hem de harf sırası açısından nesne kümesi içindeki tüm nesnelere sonra geliyor olması gerekir.

2.4.2.2. Apriori algoritması

Agrawal ve diğerleri tarafından 1994 yılında geliştirilen Apriori algoritmasıdır. Birliktelik kurallarının çıkarılması konusunda elde edilmiş büyük bir başarıdır. Birliktelik kuralları çıkarımında en çok bilinen algoritma olmuştur. Apriori boolean ilişki kuralları için geçerli bir VM algoritmasıdır. Bu algoritma özünde tekrarlayan bir algoritmadır. Algoritmanın ismi, sık tekrar eden nesnelere önsel bilgilerini kullanmasından dolayı bir önceki adımdan almasından (prior) anlamında aprioridir (Agrawal ve ark., 1994).

Bu teknik, sık tekrar eden nesne kümesinin tüm alt kümeleri de sık tekrar eden kümelerden oluşmalıdır kuralına dayanmaktadır. Apriori algoritması önceki geçişte oluşan sık tekrar eden nesne kümelerini birleştirerek aday nesne kümelerini oluşturur ve veritabanındaki hareketlerle ilgilenmeden, önceki geçişte oluşan alt kümelerden küçük olanlarını siler, önceki geçişte anlamlı bir azalma olur. Apriori seviye mantığı arama olarak bilinen yinelemeli bir yaklaşım kullanır. İlk olarak bir elemanlı sık tekrar eden kümeler bulunur. Bulunan bu küme L_1 (sık tekrar eden 1 elemanlı küme) olarak adlandırılır. L_1 , L_2 'nin (sık tekrar eden 2 elemanlı küme) elde edilmesinde kullanılır. Algoritma tekrarlı bir şekilde çalışarak, elde edilebilecek sık tekrar eden kümeleri bulur. Her L_K 'nin bulunması bütün veri tabanının taranması anlamına gelmektedir. Sık geçen öğeleri bulmak için birçok kez veri tabanını taranır ve bu taramalar aşamasında Apriori algoritmasının birleştirme, budama işlemleri ve minimum destek ölçütü yardımı ile birliktelik ilişkisi olan öğeler bulunur (Han ve Kamber, 2000).

Algoritmanın Kısa Kodu

Konu üzerindeki ilk yayını yapan Agrawal ve Srikat, tarafınan algoritmanın kodu apriori algoritması aşağıdaki şekilde verilmiştir(Agrawal ve Srikat, 1994: 5; Han ve Kamber, 2006: 239).

Giriş:

D: veri tabanındaki bir işlem,

min_sup: eşikteki minimum destek sayısı,

Çıkış:

L: Veri tabanındaki sık geçen öge setleridir.

Metod:

- 1) $L_1 = \text{find_frequent_1-itemsets}(D)$;
- 2) for ($k = 2$; $L_{k-1} \neq \emptyset$; $k++$) {
- 3) $C_k = \text{apriori_gen}(L_{k-1})$;
- 4) For each transaction $t \in D$
- 5) $C_t = \text{subset}(C_k, t)$; //Tüm alt kümeleri
- 6) For each candidate $c \in C_t$
- 7) $c.\text{count} ++$;}
- 8) $L_k = \{ c \in C_k \mid c.\text{count} \geq \text{min_sup} \}$
- 9) return $L = \bigcup_k L_k$;

Öncelikle, sık tekrar eden 1-nesne kümelerinin kümesi bulunur. Bulunan bu küme L_1 olarak adlandırılır (1.adım). $k=2$ için L_1 , L_2 'nin sık tekrar eden 2-nesne kümelerinin kümesi kullanılır. k birer arttırılarak L_2 ve L_3 ün bulunmasında kullanılır ve algoritma bu şekilde daha fazla sık tekrar eden k -nesneler kümesi bulamayınca kadar döngü şeklinde devam eder (2.adım). C_k kümesi $\text{apriori_gen}()$ prosedüründen elde edilmiş (3.adım), t bir işlem (transaction) olmak üzere veri tabanındaki (4.adım), C_t kümesi her işlem için $\text{subset}()$ prosedüründen elde edilmiştir. Küme işlemlerinde kullanılan sayaç, her işlemde bir arttırılır (7.adım). Ardından sıradaki L değerini alarak algoritma çalışmaya devam eder. Sonuç olarak da üretilen bu L değerlerinin birleşim kümesini döndürülür (9.adım) (Han ve Kamber, 2006: 238).

procedure **apriori_gen**(L_{k-1} :frequent($k-1$)- itemsets)

- 1) for each itemset $l_1 \in L_{k-1}$
- 2) for each itemset $l_2 \in L_{k-1}$
- 3) if ($l_1[1] = l_2[1] \wedge l_1[2] = l_2[2] \wedge \dots \wedge l_1[k-2] = l_2[k-2] \wedge l_1[k-1] < l_2[k-1]$) then{
- 4) $c = l_1 \times l_2$;

- 5) *if has_infrequent_subset(c, L_{k-1}) then*
- 6) delete c;
- 7) else add c to C_k;
- 8) }
- 9) return C_k;

Apriori_gen alt programı iki görevi yerine getirmektedir. Bunlar, birleştirme ve budama işlemleridir. L_{k-1} ile L_{k-1} birleştirilerek aday kümeler oluşturulmaktadır. Birleştirme işleminde ilk (k-2) elemanı eşit olan elemanların birleştirilmesi gerekmektedir. Birleştirme işleminde, l_1 ve l_2 , L_{k-1} 'in herhangi bir elemanını temsil etmektedir. Her L_{k-1} elemanı iç içe döngü yapısında kontrol edilerek, l_1 ve l_2 elemanlarından sadece ilk (k-2) elemanı eşit olanlar birleştirilmektedir. Daha sonra ise, aday kümeler içinden alt kümesi sık tekrar edilmeyenler atılarak budama işlemi yapılmaktadır. Bu işlemi, *if has_infrequent_subset()* alt programı gerçekleştirmektedir.

procedure **has_infrequent_subset**(c: candidate k-itemset;

L_{k-1} : frequent(c: candidate k - itemset;

- 1) for each(k-1) – subset s of c
- 2) if $s \notin L_{k-1}$ then
- 3) return TRUE;
- 4) return FALSE;

c: k-nesne kümesindeki adayları temsil etmektedir.

L_{k-1} : Her bir aday için sıklığı ifade etmektedir.

Her (k-1) alt kümelerinden bir veya bir kaç L_{k-1} 'de olmayan tüm nesne alt kümeleri bulunur(adım 1- adım 4) ve bir sonraki *apriori_gen()* procedüründeki 6. adımda silme işlemi gerçekleşir (Kan ve Hamber, 2006: 240).

Apriori algoritmasının çalışmasını bir örnek ile açıklamak için, her farklı sayı bir poliçe örneği ifade etmek üzere aşağıdaki veri seti kullanılacaktır. Örneğin, bu örnek veri kümesinde yer alan {1,2,3,4} elemanı 1,2,3 ve 4 numaralı poliçelerin aynı müşteri tarafından satın alındığını ifade etmektedir.

Örnek Veri= {1, 2, 3, 4}, {1,2}, {2, 3, 4}, {2,3}, {1, 2, 4}, {3,4},{2, 4}

Apriori algoritmasının ilk adımında, her ürünün destek değeri, yani kaç kere veri kümesinde tekrar ettiği hesaplanmaktadır

Tablo 2.1. Tek Elemanlı Sık Tekrar Eden Ürün Kümesi

Ürün	Destek
1	3
2	6
3	4
4	5

Tablo 2.1’de her ürünün toplam satış sayısı bulunmaktadır. Bu değere frekans veya destek (support) ismi verilmektedir. Algoritmaya göre, minimum destek değerine sahip elemanlar sık tekrar eden olarak kabul edilmektedir. Tablo 2.1’de destek değeri=3 için tüm elemanların sağladığı, yani destek değerinin üç ve üzerinde olduğu görülmektedir.

Algoritma bundan sonraki adımda iki elemanlı sık tekrar eden kümeleri bulacaktır. Bunun için birleştirme işlemi yapılacak, ardından da birleştirme işleminde ortaya çıkan her aday kümenin destek değeri hesaplanacaktır. Örneğin {1,2} aday kümesi örnek verisetinde üç kez tekrar etmektedir ve bunlar {1, 2, 3, 4}, {1,2}, {1, 2, 4}’dür. Benzer şekilde diğer aday kümelerin de destek değeri hesaplanmak ve üç ve üzerinde destek değerine sahip olanlar sık tekrar eden olarak kabul edilecektir. Bu aşama sonrasında iki elemanlı sık tekrar eden elemanlar kümesi Tablo 2.2’deki gibi olmaktadır.

Tablo 2.2. İki elemanlı sık tekrar eden ürün kümesi

Ürün	Destek
{1,2}	3
{2,3}	3
{2,4}	4
{3,4}	3

Apriori algoritması, bir sonraki aşamada üç elemanlı sık tekrar eden elemanları bulmaya çalışmaktadır. Fakat örnekteki veri setiyle ve destek değeri üç kabul edilerek, üç elemanlı sık tekrar eden ürün kümesini elde edilememektedir. Bunun sebebi minimum destek değerimizi 3 kabul etmiş olmamızdır. Örneğin {1, 2, 4} şeklindeki üç elemanlı aday kümeyi ele aldığımızda bu değer dizisinin destek değerinin iki olduğu görülmektedir. Bu destek değeri, minimum destekten küçük olduğu için kabul edilmemektedir (Şeker, 2012).

2.4.2.3. DHP algoritması

DHP(Direct Hashing and Pruning) algoritması da k-öğekümesi adaylarını k-1 elemanlı sık geçen öge kümelerinden elde eder. Apriori algoritmasından farklı olarak sık geçen küme adaylarını azaltır. DHP algoritması da veritabanının birçok kere tekrarlanmasını gerektirir (Zaki ve Ogihara, 1998).

2.4.2.4. Partition algoritması

Partition algoritması giriş/çıkış işlemlerini, veritabanını sadece iki kez okuyarak en aza indirger. Bu algoritma veritabanını bellekte ele alınabilecek küçük parçalara böler. İlk geçişte potansiyel olarak sık geçen öge kümelerini bulur, ikinci geçişte ise öge kümelerinin destek değerleri hesaplanır (Chen ve ark., 1996).

2.4.3. Kümeleme

Kümeleme, verideki benzer kayıtların gruplandırılmasını sağlayan bir tekniktir. Kümeleme yönteminde amaç; küme üyelerinin birbirine çok benzediği ancak özellikleri birbirinden farklı olan kümelerin elde edilip, veritabanındaki kayıtların bu farklı kümelere bölünmesidir(Akpınar, 2000: 6). Kümeleme analizinin sınıflandırmadan farkı, sınıflandırmada sınıflar önceden belli iken kümelemede sınıfların önceden belli olmamasıdır (Silahtaroglu, 2008: 99).

Kümeleme, bir dizi örüntüyü ayrık ve homojen guruplar oluşturacak şekilde gruplandırma işlemidir. Bu işlem örüntülerin benzerlik derecelerine göre sınıflara veya kümelere ayrılmasıyla gerçekleştirilir. Kümeleme işleminin sonunda elde edilen her başarılı kümede yer alan nesnelere arasında maksimum benzerlik gösteren ve kümeleme

sonucu elde edilen her kümedeki maksimum farklılık oluşması sağlanır Karypis'e göre kümeleme işlemi nesnelere arasındaki benzerliğe bağlı olarak gruplara ayrılır. Böylece benzer nesnelere aynı grupta olurken, farklı nesnelere başka gruplarda olur (Karypis ve ark., 1999).

Kısaca kümeleme analizinde;

- Her bir küme homojendir veya belirli özellikler bakımından benzerdir. Dolayısıyla her bir kümedeki gözlemler birbirine benzerdir.
- Her bir kümenin aynı özellik açısından diğer kümelerden farklı olması gerekmektedir. Böylelikle bir kümedeki gözlemler, diğer kümedeki gözlemlerden farklı olmaktadır.
- Benzerlik veya homojenlik kavramları, analizden analize farklılık göstermektedir ve yapılan çalışmaların amaçlarına bağlıdır. Amaca göre farklı sayıda ve özellikte kümeler oluşturulabilir (Sharma, 1996).

Benzerlik Ölçüsü

Veri tabanındaki veriler kümelere ayrılırken, benzerlik ve uzaklık kavramlarından yararlanılırlar. Bu veri tabanındaki her bir kaydın veri tabanındaki diğer kayıtlara olan benzerliği veya uzaklığı olduğu gibi oluşturulan gerçek ve aday kümeler arasındaki mesafe ve benzerliği de içerir. Birbirine olan uzaklığa göre ayrılmış kümelerin gerçekten farklı özelliklere sahip olup olmadıklarının belirlenmesi gerekir. Birbirinden çok farklı olmayan kümeler birleştirilerek tek bir küme haline dönüştürülebilir. Kümeler arasındaki mesafenin ölçülmesiyle iki veya daha fazla kümenin birleştirilmesi söz konusu olduğu gibi aynı zamanda, bir kümeden daha fazla küme üretilmesi de söz konusu olabilir. Bu nedenle sürekli olarak kümelerin büyüklüğü ve çapı ölçülmelidir.

Sürekli veriler için; her biri p tane sürekli değişken içeren x_i ve y_j gözlem çifti arasındaki uzaklık $d_{ij} = d(x_i, y_j)$ olsun.

p sayıda değişkene göre birimler ya da değişkenler arasındaki uzaklıkları hesaplamak için kullanılan uzaklık ölçülerine genel olarak Minkowski Uzaklık Ölçüsü adı verilir.

Minkowski uzaklığı aşağıdaki gibi hesaplanır; (Merigo ve Casanovas, 2008)

$$d_{\lambda}(x, y) = \left[\sum_{i=1}^p |x_{ik} - y_{ik}|^{\lambda} \right]^{1/\lambda} ; \lambda \geq 1 \quad (2.9)$$

Burada $\lambda=1$ durumu için **Manhattan (City-Block Distance) Uzaklığı** olarak bilinir (Shih and Wu, 2004).

$$d(x, y) = \sum_{i=1}^p |x_{ik} - y_{ik}| ; \lambda = 1 \quad (2.10)$$

Burada λ üs değerini göstermektedir. Minkowski uzaklığının $\lambda = 2$ için özel hesaplama formülü **Öklit Uzaklığı** olarak bilinmektedir (Shih and Wu, 2004).

$$d_2(x, y) = \left[\sum_{i=1}^p |x_{ik} - y_{ik}|^2 \right]^{1/2} ; \lambda = 2 \quad (2.11)$$

Ölçekli Öklid Uzaklığı:

$$d_2(x_i, y_j) = \left[\sum_{i=1}^p w_k^2 |x_{ik} - y_{ik}|^2 \right]^{1/2} \quad (2.12)$$

Burada w_k k'nıncı değişkenin standart sapma değerinin (s_k) ya da dağılım aralığı değerinin tersidir. w_k 'nin standart sapma değerinin tersi olması durumunda bu uzaklığa Karl-Pearson uzaklığı da denir.

Mahalanobis Uzaklığı:

$$d(x_i, y_i) = (x_{ik} - y_{ik})^T S^{-1} (x_{ik} - y_{ik}) \quad (2.13)$$

Burada S örnek ya da küme içi kovaryans matrisidir (McLachlan, 1999).

Kümeleme Teknikleri

Literatürde kullanılan birçok kümeleme tekniği mevcuttur. Teknikler birbirinden kümelemenin oluşturuluş şekline, kullanılan veri tipine ve yapılacak çalışmanın amacına göre de farklılık gösterirler. Kümeleme teknikleri genel olarak hiyerarşik ve hiyerarşik olmayan yöntemler olarak ikiye ayrılmaktadır (Dillon ve Matthew, 1984).

i. Hiyerarşik Yöntemler

a) Toplaşım(Agglomerative) Kümeleme Algoritmaları

- Merkezi Kümeleme(Centroid) Yöntemi
- Tek Bağlantı(En Yakın Komşu – Single Linkage) Yöntemi
- Tam Bağlantı(En Uzak Komşu-Complete Linkage) Yöntemi
- Ortalama Bağlantı (Average Linkage) Yöntemi
- Ward Yöntemi

b) Bölünür(Divisive) Kümeleme Algoritmaları

- Bölünmüş Ortalamalar(Splinter –Average Distance) Yöntemi
- Otomatik Etkileşim Dedektörü(Automatic Interaction Detection–AID) Yöntemi

ii. Hiyerarşik Olmayan Yöntemler

- K-Ortalamlar(K-Means) Yöntemi
- Medoid Parçalama Yöntemi
- Yığma Kümeleme Yöntemi
- Bulanık Kümeleme Yöntemi

Kümelemede genellikle k-ortalamlar algoritması ya da Kohonen şebekesi gibi istatistiksel yöntemler kullanılmaktadır. Hangi yöntem kullanılırsa kullanılsın, süreç aynı şekilde işler. Her kayıt var olan kümelerle karşılaştırılır. Bir kayıt kendisine en yakın kümeye atanır ve bu kümeyi tanımlayan değeri değiştirir. Optimum çözüm buluncaya kadar kayıtlar yeniden atanır ve küme merkezleri ayarlanır (Hui ve Jha, 2000).

2.4.3.1. K-ortalamlar algoritması

K-ortalamlar algoritması 1967 yılında J.B. MacQueen tarafından geliştirilmiştir (MacQueen, 1967). K ortalamlar algoritması büyük veritabanlarının kümelenmesinde oldukça etkin bir algoritmadır. Algoritma n boyutlu bir kümedeki tüm veriyi inceleyerek k bir örnek kümesi ayarlanır. En çok benzeyen değişkenler bir alt kümesini oluşturur onlar çıktıyı tahmin etme aşamasında kullanılır (MacQueen, 1967; Anand, 2003).

Bu algoritmanın adımları şu şekildedir:

1. Veri seti raslantısal olarak k adet başlangıç kümesine ayrılır.
2. Veri setinde yer alan örnekler yani değişkenler; merkezi kendisine en yakın olan kümeye atanır.
3. Her atamanın sonunda küme merkezi yeniden hesaplanır.
4. Veri setindeki tüm örneklerin ataması yapılana kadar 2. ve 3. adımlar tekrarlanır.

K-ortalamlar algoritması kümeleme yönteminin değerlendirilmesinde en yaygın olarak karesel hata kriteri SSE(sum of squared error) kullanılır. En düşük SSE değerine sahip kümeleme sonucu en iyi sonucu verir. Nesnelerin buldukları demetin merkez noktalarına olan uzaklıklarının karelerinin toplamı aşağıdaki formülle hesaplanmaktadır (Ögüdücü, 2011).

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} dist^2(m_i, x) \quad (2.14)$$

x : C_i kümesinde bulunan bir nesne

m_i : C_i kümesinin merkez noktası

Bu kriterler sonucu k tane kümenin olabildiğince yoğun ve birbirinden ayrı sonuçlanması hedeflenmektedir. Algoritma, karesel-hata fonksiyonunu azaltacak k parçayı belirlemektedir (Kumar, 2000).

Yeni bir değişkenin ait olduğu kümeyi belirlemek için algoritma yeni değişkenin öğrenme verisindeki her bir değişkenden uzaklığı hesaplar. k değerinin ve uzaklık

ölçüsünün modelin kalitesi üzerinde büyük oranda etkisi vardır. Bu nedenle onları dikkatle seçmek çok önemlidir. K ortalamalar algoritması oldukça etkin bir algoritma olup sadece nümerik veri ile çalışır. Fakat VM uygulamaları sıklıkla kategorik verileri de içermektedir. K-ortalamalar algoritmasının geliştirilmesi ile elde edilen k modlar algoritması ise kategorik veriler üzerinde çalışabilen bir algoritmadır. K-ortalamalar algoritmasında küme merkezleri, küme ortalaması alınarak hesaplanırken; k modlar algoritması küme merkezlerinin belirlenmesinde kümede en sık tekrarlanan değerler(mod) dikkate alınır (San ve ark., 2004).

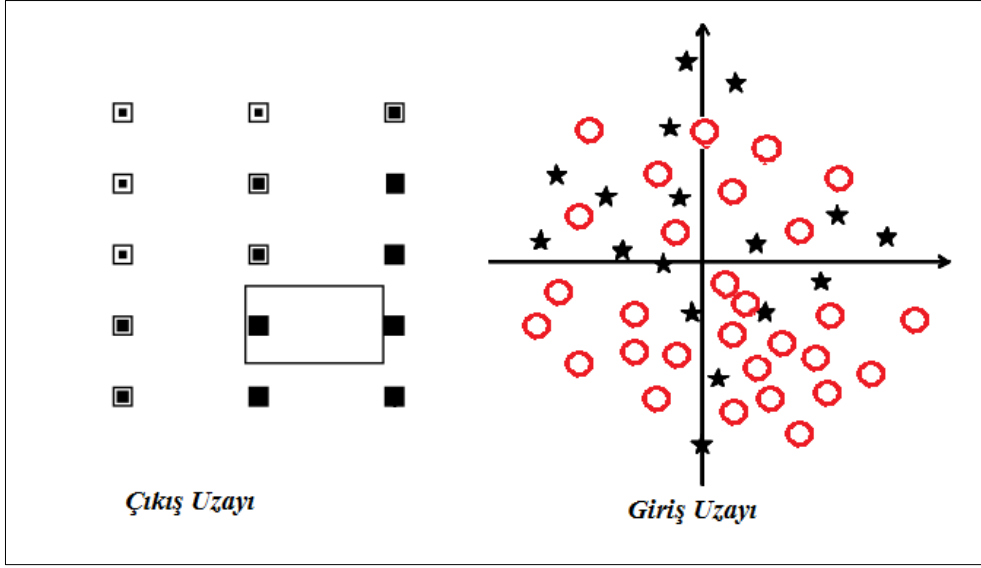
2.4.3.2. Kohonen ağırları

Kohonen ağırları diğer bir adıyla SOM(Self Organizing Map) kendi kendini organize eden harita yöntemi olarak adlandırılmaktadır. Bu yöntem ile; yüksek boyutlu vektörler, iki boyutlu ızgara şeklindeki haritada temsil edilebilmektedir (Kohonen, 1982).

Kohonen ağırları, Giriş Uzayı ve Çıkış Uzayı olarak iki katmana ayrılır. Başlangıçta grupların bilinmediği durumlarda, verilerin kümelendirilmesini amaçlamaktadır. Kümeleme amacıyla kullanılmasının yanında, bir “veri görselleştirme aracı”dır (Flexer, 2001).

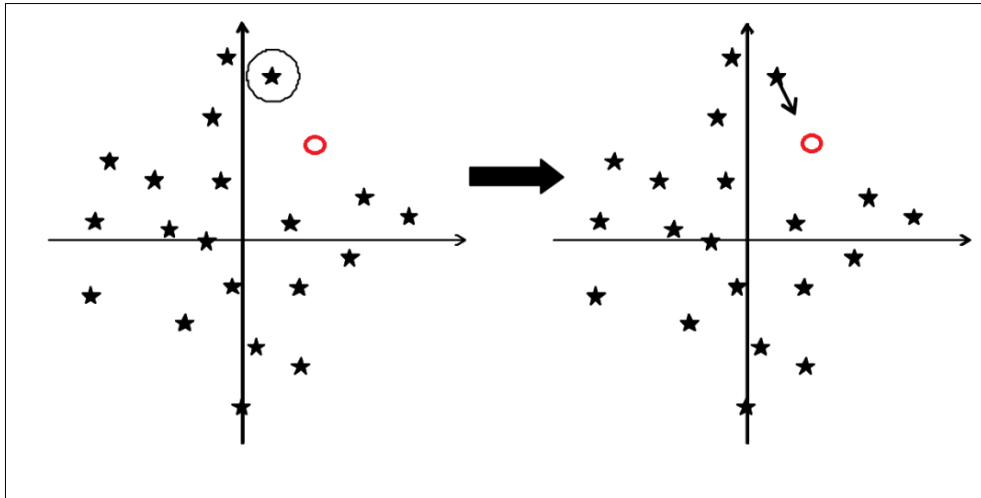
Denetimsiz öğrenim gerçekleştiren bir sinir ağı türüdür. SOM’un temel fikri insan beyninin çalışmasına benzer bir şekildedir.

SOM, tek katmanlı bir nöron ağıdır ve nöronlar n boyutlu olarak grid (ızgara)’e yerleştirilmiştir. Birçok uygulamada grid 2-boyutlu ve dikdörtgen şeklindedir ve birçok uygulamada altıgen grid kullanılıyor olsa bile, bunun yanı sıra 1,3 ve daha fazla boyutlu uzay da kullanılmaktadır. Grid’in içindeki veri ilişkilerinin giriş uzayındakine karşılık gelen çıkış uzayını tanımlamak mümkündür (Baçao ve Lobo, 2013).



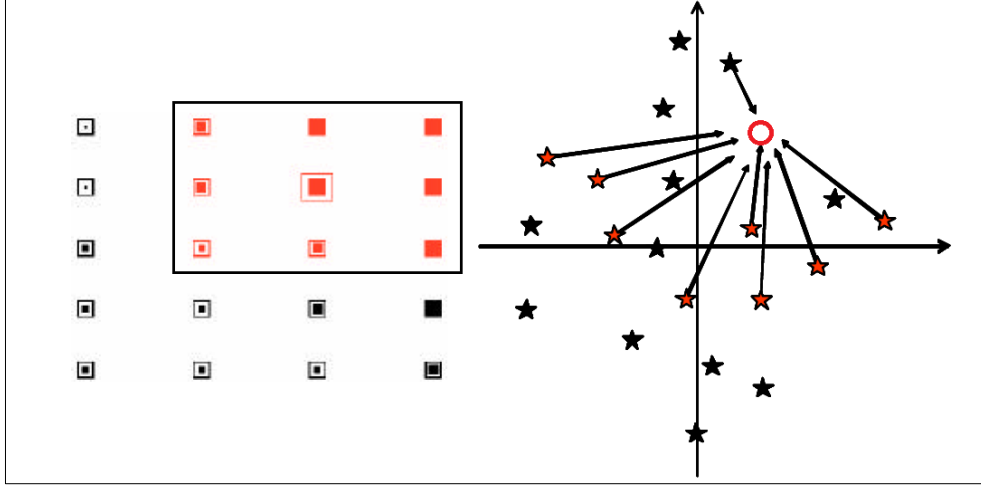
Şekil 2.2. SOM Temel Mimarisi

Şekil 2.2’de her iki uzayın mimarisi sunulmuştur. Şeklin sağ tarafında giriş uzayı, solunda ise çıkış uzayı görülmektedir. Giriş uzayındaki yuvarlak şekiller örüntüleri, yıldızlar ise her bir nöronu ifade etmektedir. Görselleştirme amacıyla her vektör sadece iki koordinatlı olarak tanımlanmıştır. Şeklin sol tarafındaki çıkış uzayında ise komşuluk bağlantıları kurulmuştur. Desenler birçok ağırlıklara sahiptir. Örüntüleri aynı uzaydaki vektörler gibi kabul görülebilir. Daha sonra seçilmiş olan nöron kazanan nöron’a yakın olanı güçlendirir. Daha sonra desen, kazanan nöronla eşlenir. Bunun bir sonucu olarak nöron, giriş deseninin pozisyonuna doğru taşınacaktır (Baçao ve Lobo, 2013).



Şekil 2.3. SOM’un Geliştirilmesi

Giriş desenleri kullanılarak SOM'a uygulamak veya geliştirmek istendiğinde, ağdaki mevcut desenler ve her nöron arasındaki yoğunluk hesaplanır. Şekil 2.3'de görüldüğü gibi kendine en yakın desen seçilir, onun yerine taşınır (Baçao ve Lobo, 2013).

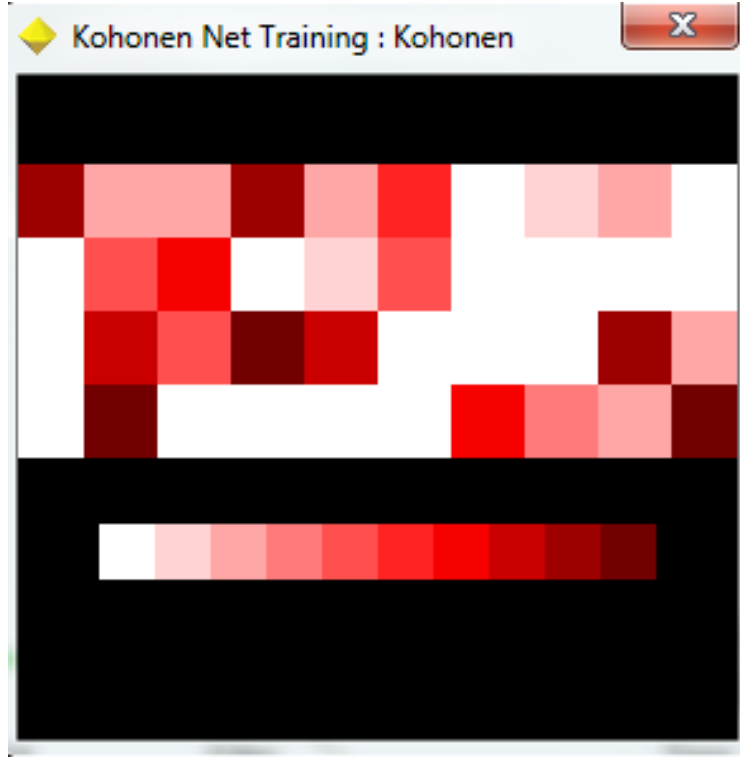


Şekil 2.4. SOM'un Çıkış Ağırlığının Belirlenmesi

Şekil 2.4'de çıkış katmanında pozisyonu korumak yerine kazanan nöronun pozisyonunu doğrulamak daha önemlidir. Fakat bunun yanı sıra diğer komşu nöronların pozisyonları da önemlidir. Bu yolla kesin olan alt kümeler belirlenerek geliştirilir (Baçao ve Lobo, 2013).

Kohonen ağları, girdi (bağımsız) değişken kümesindeki örüntülerin açığa çıkarılması amacıyla kullanılmaktadır. Bu ağın çıktısında, gözlemler gruplandırılmış olarak elde edilmektedir. Bir grup veya kümenin içindeki gözlemlerin birbirine benzer olduğu, farklı gruplarda yer alan gözlemlerin ise birbirine benzer olmadığını söylemek mümkündür (Kiang, 2001).

Her bir nöron, girdi değişkenleri veya aynı anlamda girdi alanlarının her biri ile bağlantılıdır ve tekrar ağırlıkları (önemleri) bu bağlantıların her birinin üzerinde yer almaktadır. Genellikle bir Kohonen ağı, az sayıda birim, çok sayıda gözlemi özetlediğinde (güçlü birimler) veya çok sayıda birim gözlemlerin her hangi birine karşılık gelemediğinde (zayıf birimler) son bulacaktır (SPSS Clementine, 2007).



Şekil 2.5. Kohonen Geri Bildirim Grafiği

Kohonen ağlarının eğitimi boyunca, istenildiği takdirde Clementine programı üzerinde görselleştirilmiş bir geri bildirim grafiği (feedback graph) görüntülenebilir. Bir Kohonen Geri Bildirim grafiği örneği Şekil 2.5'deki gibidir. Kohonen geri bildirim grafiği, eğitim süresince görüntülenmektedir. Her bir düğümün gücü, kırmızıdan beyaza doğru değişen bir renk ile temsil edilmektedir. Kırmızı renk güçlü birimler, beyaz renk ise birkaç veya hiç kayıt kazanamayan birimleri, yani zayıf birimleri temsil etmektedirler. Kırmızının rengi koyulaştıkça daha da kuvvetli bir birim olduğunu ifade eder (SPSS Clementine, 2007).

Algoritma

SOM'un birçok çeşitli tanımlaması mevcuttur. SOM haritalama en temel olarak, \mathcal{R}^n giriş uzayında, rasgele iki boyutlu dizi düğümlerinin üstüne haritalanır. Her düğüm i, a parametrik referans vektörü ile ilişkilidir. Algoritmanın gelişimi aşağıda açıklanmaktadır: (Kohonen ve Kangas, 1996; Baçao ve ark., 2005)

- 1) Kohonen ağındaki ağırlık değerleri rastgele seçilir.
- 2) Desenler ile her bir nöron arasındaki yoğunluk hesaplanır.

$$w_{kazanan}(w_{ij}: d_{ij} = \min(d_{mn})) \quad (2.15)$$

- 3) Giriş uzayındaki giriş vektörü $X = [x_1, x_2, \dots, x_n]^T \in \mathfrak{R}^n$ ve her değer için kendine ait ağırlık vektörü $W_i = [w_1, w_2, \dots, w_n]^T \in \mathfrak{R}^n$ vardır.
- 4) Giriş vektörleri $x \in \mathfrak{R}^n$ her düğüm m_i ile karşılaştırılır ve pratik uygulamalarda öklid mesafesi, $\|x - m_i\|$ olarak hesaplanır. Hesaplanan mesafelere göre en küçük uzaklığa sahip olan düğüm bulunur. Bunun sonucunda en iyi eşleşen düğüm yani m_c bulur. Giriş böylece bu konum üzerine haritalanır.
 - a) Her nöron 2.16'deki kurala göre güncellenir.

$$w_{ij} = w_{ij} + \alpha \cdot h(w_{kazanan}, w_{ij}) \|x - w_i\| \quad (2.16)$$

- b) En iyi eşleşen düğümün bulunması için 2.17'deki formül kullanılır.

$$\|x - m_c\| = \min_i \{\|x - m_i\|\} \quad (2.17)$$

- c) c sinir hücresi eğitim aşamasında bulunurken formül 2.18'deki gibi hesaplanır.

$$c = \arg \min_i \{\|x - m_i\|\} \quad (2.18)$$

- d) Öğrenme aşamasında, benzer giriş dizisindekiler; bunlar dizinin içinde birbirlerine daha yakın olacaklardır. Yakınlaştırma işlemi için 2.19'daki formül kullanılır.

$$m_i(t + 1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)] \quad (2.19)$$

e) Burada t , koordinattaki farklı bir zaman(discrete-time) ve $h_{ci}(t)$ çekirdek üzerindeki tanımlı örgü noktalarındaki komşuluklardır. Komşulukların hesaplanması için formül 2.20'deki gibidir.

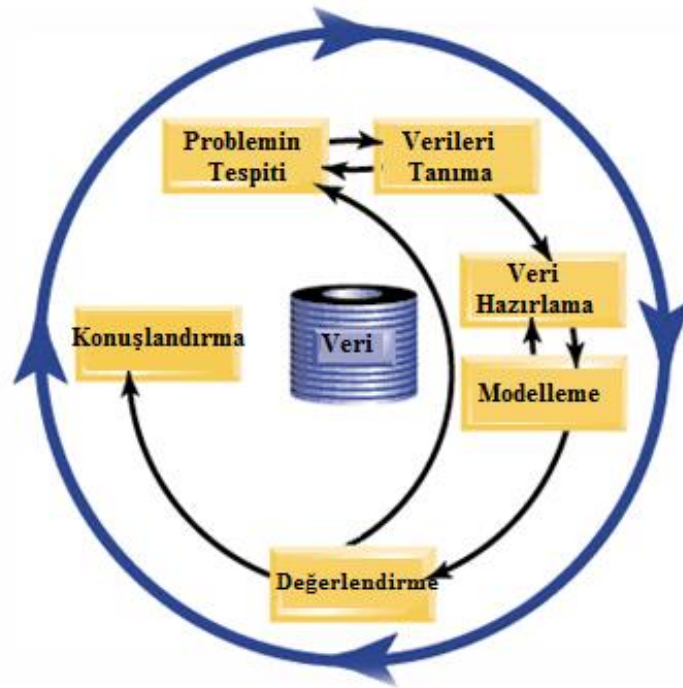
$$h_{ci}(t) = \alpha(t) \cdot \exp\left(-\frac{(\|r_c - r_i\|)^2}{2\sigma^2(t)}\right), \quad (0 < \alpha(t) < 1) \quad (2.20)$$

Nöronların pozisyonlarına (i, j) dersek, $(0 \leq \alpha(t) \leq 1)$ öğrenme oranı, $h(w_{ij} - w_{mn})$ komşu fonksiyonu olmaktadır. Komşu fonksiyonu $[0,1]$ içerdiği, yüksek olanlar yani çıkış katmanında yakın olanlar ve küçük(veya 0) olan nöronlar uzaklardakiler olarak farz edilir. w_i giriş desenindeki kazanan nöron olarak ifade edebiliriz (Kohonen, 1988).

3. BULGULAR VE TARTIŞMA

3.1. Sigortacılık Sektörü Üzerine Uygulama

Sigortacılık Sektörü üzerine gerçekleştirilen uygulamada farklı sektörlerle ve iş problemlerine uygulanabilen CRISP-DM(Cross-Industry Standard Process for Data Mining) modelindeki adımlar izlenmiştir. CRIPS-DM modeli VM sürecinde izlenecek adımları sunmaktadır. CRIPS-DM modeli, Şekil 3.1’de görülebileceği gibi VM sürecinde işlenecek olan adımların sunulduğu bir çevrimden oluşmaktadır.



Şekil 3.1. CRISP-DM Modeli (CRISP-DM; Chapman ve ark., 2000)

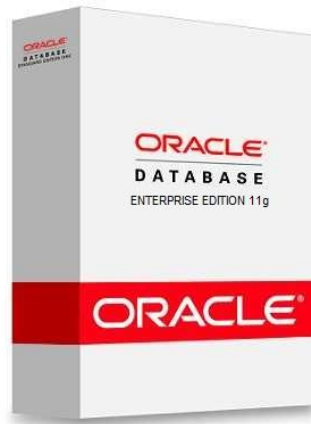
CRISP-DM referans modelinin fazları altı aşamadan meydana gelmektedir. Bu aşamalar araştırma sürecinde başlıklar halinde incelenmiştir. Bunlar;

3.1.1. Arařtırma probleminin tespiti (Business Understanding)

Bu ařama veri madenciliđinin en önemli ařamasıdır. Bu ařamada iři anlamak ve veri madenciliđi sũrecinde tam olarak ne elde edilmek istendiđinin belirlenmesi anlamına gelmektedir. Bu gœrev dođrultusunda, veriye sahip olan sigorta řirketin iři durumu hakkında bilgi edinilmiř ve temel beklentinin ne olduđu tespit edilmiřtir. Ardından ihtiyaçlar belirlenmiřtir. Bu adımda Sigortacılık sektœrũnde veri madenciliđine dayalı arařtırmalar yapılmıř olup, MİY ve Veri Madenciliđini bir araya getiren arařtırmalarla ilgili olarak literatũr taraması yapılmıřtır.

3.1.2. Verileri tanıma (Data Understanding)

Verileri tanıma ařamasında bařlangıç verilerinin elde edilmesini ve verileri tanımaya yœnelik analizleri ve verilerde saklı olabilecek bilgiler iœin ilk izlenimlerin oluřturulmasını kapsayan bir ařamadır. Bu ařamada verilerin elde edilmesinden bařlanarak verinin genel yapısı hakkında fikir geliřtirmeye, problemi tespit etmeye ve hatta veri iœerisinde ilginç olabilecek alt kũmeleri tespit etmeye kadar birœok çalıřmayı kapsamaktadır. Bu ařamada Sigorta řirketinden veriler sađlanmıřtır ve kullanılan VM yazılımları arařtırılmıřtır. Verilerin saklanması iœin bir veritabanına ihtiyaç duyulmuř ve bunun iœin Őekil 3.2’de gœrũldũđũ gibi en popũler ve en fonksiyonel veritabanı yœnetim sistemlerinden birisi olan Oracle Veritabanı (Oracle Database Enterprise Edition 11g) kullanımına karar verilmiřtir.



Őekil 3.2. Oracle Enterprise Edition Database 11g

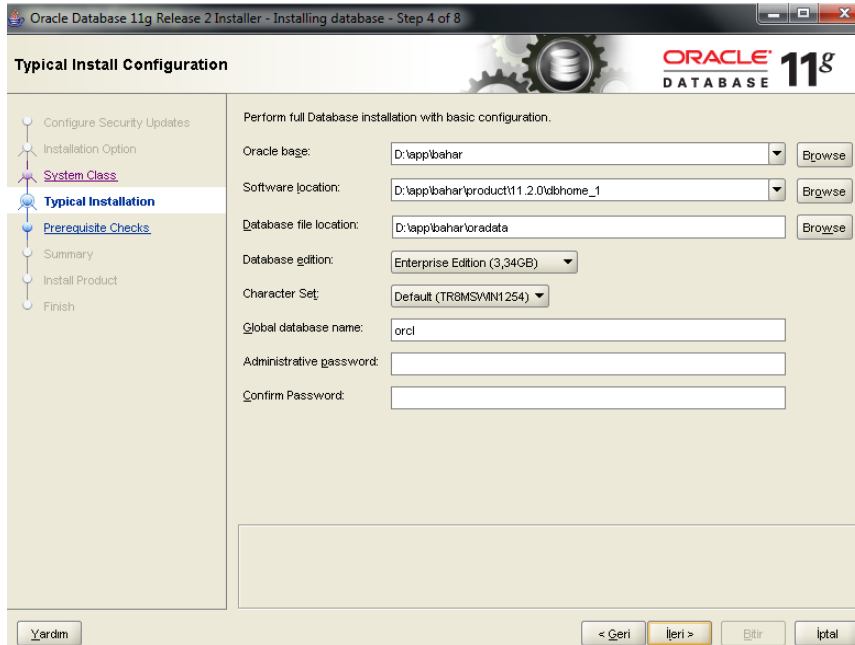
Ardından analizlerde kullanılmak üzere VM yazılımları araştırılmıştır. Veri madenciliğinde kullanılan birçok program vardır. Bunlar; WEKA, SAS, SPSS vd. olmak üzere çeşitlilik göstermektedir. Fakat son zamanlarda IBM'in sunmuş olduğu SPSS Clementine programı kullanıcı arayüzü ve kullanım kolaylığı sağlamakla birlikte, görsel tasarımıyla da birçok sektörde yaygın olarak kullanılmaktadır. Bundan dolayı bu çalışmada Clementine 12.0 versiyonu tercih edilmiştir.

3.1.3. Veri Hazırlama (Data Preparation)

Veri hazırlama aşaması, ham veriden başlayarak son veriye kadar yapılması gereken düzenlemeleri içermektedir. Veri hazırlama, tablo, kayıt, veri dönüşümü ve modelleme araçları için temizleme gibi özellikleri içermektedir. Bu aşama da aşağıdaki adımlar sırasıyla gerçekleştirilmiştir.

3.1.3.1. Kurulum aşaması

Bu aşamada kullanılacak veriler sigorta şirketinin veritabanından veriler excel aracılığı ile csv formatında alınmıştır. Ardından bu verilerin kontrolünün rahatlığını sağlamak, filtrelemek ve amaca uygun veri setleri elde edebilmek için Şekil 3.3'de görüleceği üzere kişisel bir bilgisayara Oracle veritabanı kurulumu gerçekleştirilmiştir.



Şekil 3.3. Oracle Veritabanı Kurulum Aşaması

Kuruluma kolaylık ve performans için güncel sürümü olan “Oracle 11g Enterprise Edition Database” kurulumu tercih edilmiştir. Bu sayede verileri daha kolay işlenmesine ve dinamik olarak sorgu çekilmesine olanak sağlanmış olunur. Ardından veritabanına Excel’den elde edilen verileri ekleme(insert) işlemi ile verilerin ham halinin veri tabanına yüklenme işlemi gerçekleştirilir. Oracle veritabanının kurulumunun ardından tez çalışmasında kullanılan Clementine programı ile haberleşebilmesi için “Oracle Client” kurulumunun yapılmış olması gerekmektedir. Şekil 3.4 de görüldüğü gibi Oracle Client kurulumunun seçim aşaması gözlenmektedir.

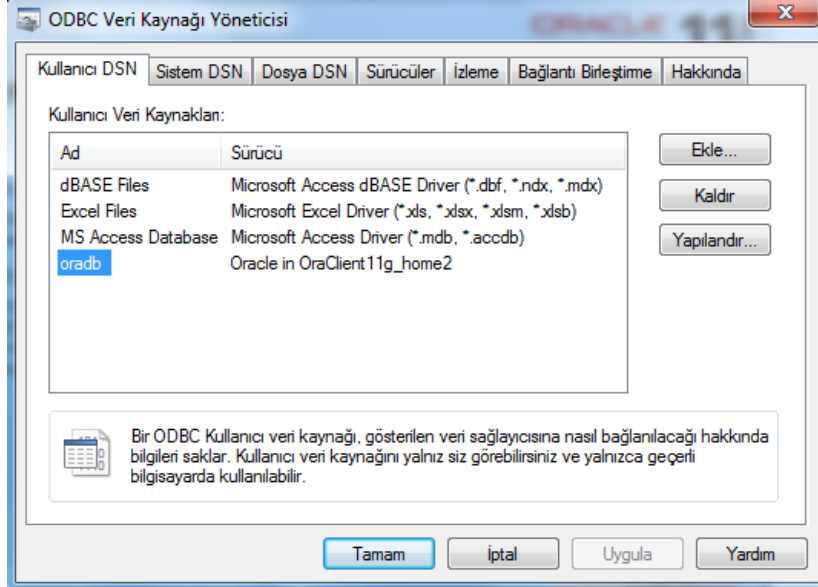
Gerekli tüm dökümanlara ve kurulumlara Oracle’ın resmi sitesi olan “www.oracle.com” adresinden erişilebilir.



Şekil 3.4. Oracle Client Kurulum Aşaması

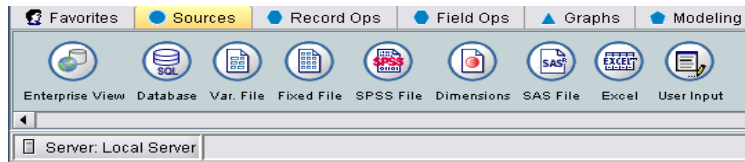
Clementine veri madenciliği programına Oracle Veritabanını ile iletişime geçebilmesi için, öncelikli olarak Oracle Client kurulumunun gerçekleşmiş olması gerekmektedir. Bu kurulumun ardından ODBC Veri Kaynağı yöneticisinden uygulamayı tanıtmak gerekmektedir. Clementine ile Oracle Veritabanını tanıtabilmek(bağlantıyı sağlayabilmek) için Microsoft’un ODBC Veri Kaynağı Sürücüsü olarak Oracle Client’i

eklenmesi gerekmektedir. Şekil 3.5’de görüleceği gibi ODBC Veri Kaynağı Yöneticisi aracılığı ile ekle seçeneği kullanılarak sürücü olarak Oracle’ın eklenmesi gerçekleştirilmiştir.



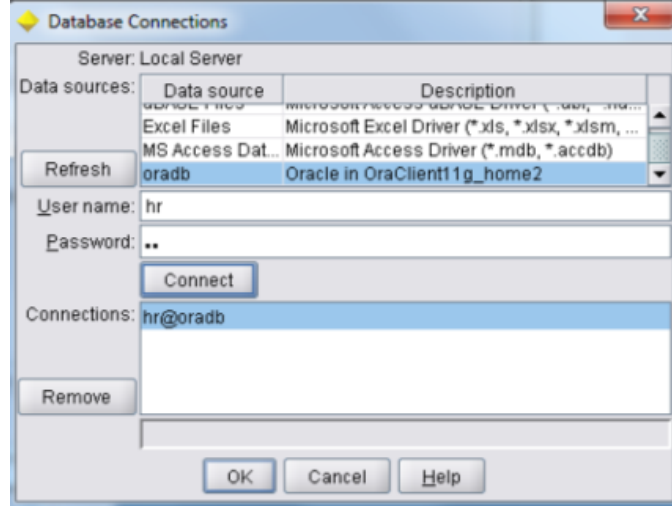
Şekil 3.5. ODBC Veri Kaynağı Yöneticisine Sürücünün Eklenmesi

Bu işlem yapıldıktan sonra Clementine programına Şekil 3.6’da görüleceği gibi Sources(Kaynaklar) sekmesinden Database(Veritabanı) aracını(node) seçip alana(stream) sürüklenmiştir.



Şekil 3.6. Clementine Data Source Sekmesi

Ardından, Database sembolüne çift tıklanarak, Şekil 3.7’deki gösterilen Veritabanı(Database Connections) bağlantısı kısmından, veritabanı kullanıcı adı(username) ve şifre(password) ile giriş yapılmıştır. Örneğin; hr@oradb olarak kullanıcı adı(username) ve şifre(password) girilerek bağlanılmaktadır. Burada “hr” kullanıcı adı ve “oradb” veritabanının adı’dır. Bu işlemin ardından tablo adı(table name) kısmına tablo adı yazılarak veya sql sorgusu(sql query) ile istenilen sorgu cümlecikleri çalıştırılabilir.



Şekil 3.7. Veri Tabanı Bağlantısının Kurulması Aşaması

Clementine programına Oracle Veritabanı tanıtıldıktan sonra, veritabanına bağlanılarak SQL sorgularına Clementine ile erişilebilir duruma gelmiştir. Örnek veri seti ekran görüntüsü şekil 3.8’deki gibidir.

SEQ	BRANSKOD	GF	KOD	ACIKLAMA	TARIFENO	STATUS	STAT	TUR	POLID	POLTANTAR
1	1224....K	F	FK	FERDİ KAZA SİGORTALARI	DENTA FK	M	MM	MERİ	1784663.....	2010-01-12 00:00
2	1225....K	F	FK	FERDİ KAZA SİGORTALARI	DENTA FK	M	MM	MERİ	1784663.....	2010-01-12 00:00
3	1226....K	F	FK	FERDİ KAZA SİGORTALARI	DENTA FK	M	MM	MERİ	1784663.....	2010-01-12 00:00
4	1227....H	F	YH	YILLIK HAYAT SİGORTALARI	749	M	MM	MERİ	1785388.....	2010-01-05 00:00
5	1228....K	F	FK	FERDİ KAZA SİGORTALARI	DENTA FK	M	MM	MERİ	1798551.....	2010-01-05 00:00
6	1229....K	F	FK	FERDİ KAZA SİGORTALARI	DENTA FK	M	MM	MERİ	1798551.....	2010-01-05 00:00
7	1230....K	F	FK	FERDİ KAZA SİGORTALARI	DENTA FK	M	MM	MERİ	1798551.....	2010-01-05 00:00
8	1231....H	F	YH	YILLIK HAYAT SİGORTALARI	750	M	MM	MERİ	1785136.....	2010-01-06 00:00
9	1232....H	F	YH	YILLIK HAYAT SİGORTALARI	750	M	MM	MERİ	1785136.....	2010-01-06 00:00
10	1233....H	F	YH	YILLIK HAYAT SİGORTALARI	750	M	MM	MERİ	1785499.....	2010-01-04 00:00
11	1234....H	F	YH	YILLIK HAYAT SİGORTALARI	750	M	MM	MERİ	1785500.....	2010-01-04 00:00
12	1235....H	F	YH	YILLIK HAYAT SİGORTALARI	749	M	MM	MERİ	1798577.....	2010-01-06 00:00
13	1236....H	F	YH	YILLIK HAYAT SİGORTALARI	749	M	MM	MERİ	1798577.....	2010-01-06 00:00
14	1237....K	F	FK	FERDİ KAZA SİGORTALARI	FKZ_TL	M	MM	MERİ	1785374.....	2010-01-14 00:00
15	1238....K	F	FK	FERDİ KAZA SİGORTALARI	FKZ_TL	M	MM	MERİ	1785374.....	2010-01-14 00:00
16	1239....K	F	FK	FERDİ KAZA SİGORTALARI	FKZ_TL	M	MM	MERİ	1785374.....	2010-01-14 00:00
17	1240....K	F	FK	FERDİ KAZA SİGORTALARI	DENTA FK	M	MM	MERİ	1830161.....	2010-02-02 00:00
18	1241....K	F	FK	FERDİ KAZA SİGORTALARI	DENTA FK	M	MM	MERİ	1830161.....	2010-02-02 00:00
19	1242....K	F	FK	FERDİ KAZA SİGORTALARI	DENTA FK	M	MM	MERİ	1830161.....	2010-02-02 00:00
20	1243....K	F	FK	FERDİ KAZA SİGORTALARI	DENTA FK	M	MM	MERİ	1830161.....	2010-02-02 00:00

Şekil 3.8. Clementine Arayüzünde Örnek Veri Setinin Görüntülenmesi

3.1.3.2. Verilerin temizlemesi ve düzenlenmesi aşaması

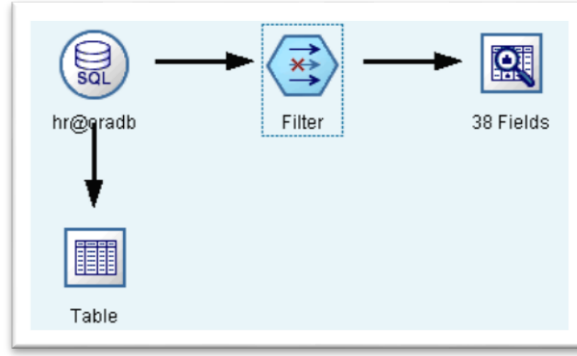
Verilerin kontrolünün kolaylığı için Oracle Veritabanında “MİY”, “MİY_FK” ve “GENELHAYAT” adı altında kullanılacak üç ayrı tablo oluşturulmuştur. Bu tablolardan “MİY” ve “MİY_FK” tablolarında birliktelik analizinde kullanılacak olan veriler bulunmaktadır. Bu tablolarda müşterilerin hangi ürünü tercih ettikleri bilgisi tutulmaktadır. “GENELHAYAT” tablosunda ise Hayat Poliçeleri satın alan müşterilerin demografik özellikleri tutulmaktadır, ayrıca “GENELHAYAT” tablosu Kümeleme Analizlerinden Kohonen ve K-Ortalamalar algoritmalarında kullanılmak üzere yaratılmıştır.

Uygulamada Türkiye'nin önemli sigorta şirketlerinden birinin farklı tarih aralıklarında satış gerçekleştirilmiş olan Bireysel Poliçe verilerinden oluşmaktadır. Müşteri gizlilik prensiplerine bağlı kalınarak müşteri kimlik bilgileri ve poliçe numaraları belirli algoritmalar kullanılarak değiştirilmiştir.

Bu uygulamada kullanılan veri setleri Birliktelik algoritmalarından olan Apriori algoritması ve Kümeleme Analizinden olan K-Ortalamalar Analizi ile Kohonen Ağları analizlerine uygulanması ve sonuçları incelenmiştir.

Verilerin temizlenmesi, veri kalitesi için önemlidir. İl bazında ilçe ayrımı yapabilmek için ilçeleri boş olan alanları adresler kolonundaki adres bilgileri ile düzenleyerek elle veri temizlemesi yapılmıştır. Bunun devamında bazı durumlarda belirli kolonları görmek isteyip belirli durumlarda ise kolonları kullanmak istenmediği durumların ortaya çıkmasından dolayı, tabloları istenilen şekilde gibi şekillendirebilmek için, SQL ile görüntü(view)'ler oluşturulmuştur. Müşterilerin kişisel bilgilerini saklamak için, ad ve soyad alanını gizlenmiş ve “müşteri_id” ile analizler gerçekleştirilmiştir.

Veri setindeki sütunlarda verilerin eksik değerlerinin bulunup bulunmadığının tespiti için Clementine Programının “Data Audit” aracı kullanılmıştır. Bu araç, veri kalitesinin belirlenmesi ve eksik değerlerin tespitine olanak sağlar. Şekil 3.9'da Data Audit aracının kullanılmasına ait ekran görüntüsü görülmektedir.



Şekil 3.9. Veri Setinin Data Audit Aracından Geçirilmesi

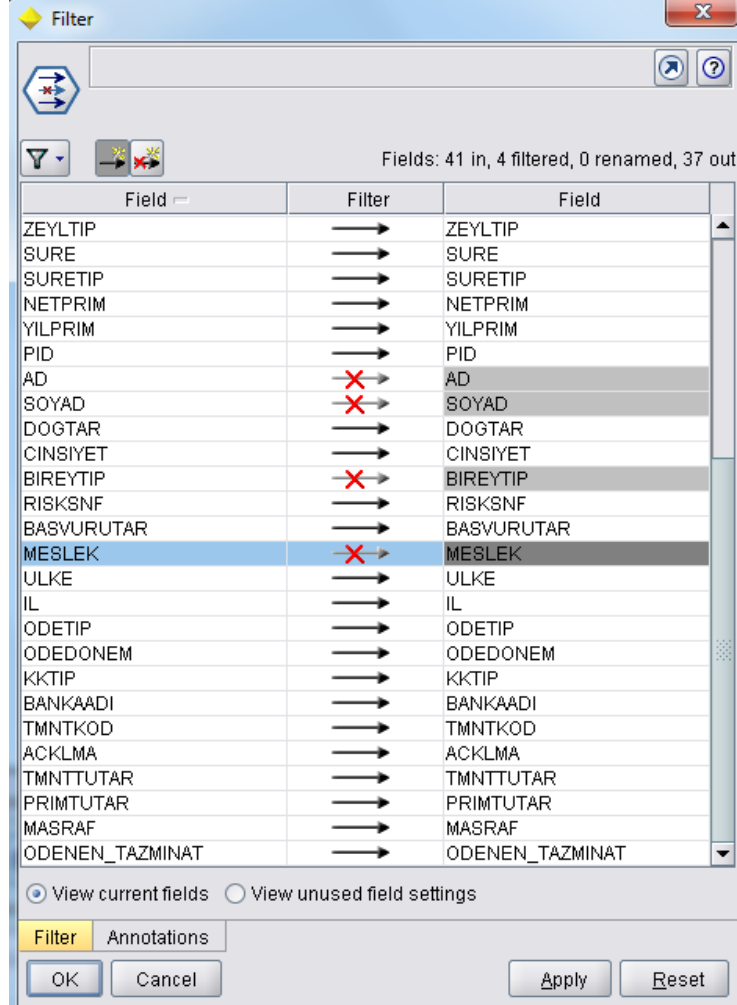
Şekil 3.10'da Data Audit aracının GENELHAYAT tablosuna ait analiz sonuçları görülmektedir. İlerleyen aşamalarda bu analizin sonuçlarının değişkenler üzerindeki etkisi incelenmiştir. Bu sonuçlarda görüleceği gibi verilerin %94,74 ünün dolu olduğu ayrıca meslek alanına bakıldığında %58,557'sinin dolu olduğunu ve 1518 alanın girilmemiş (Null) değerlerden olduğunu gözlemlenebilmektedir.

Field	Type	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value
POLBITTAR	Range	0	0 None	Never	Fixed	100	3662	0	
ZEYLNO	Range	38	47 None	Never	Fixed	100	3662	0	
SURE	Range	52	20 None	Never	Fixed	100	3662	0	
NETPRIM	Range	0	33 None	Never	Fixed	100	3662	0	
YILPRIM	Range	0	29 None	Never	Fixed	100	3662	0	
PID	Range	132	0 None	Never	Fixed	100	3662	0	
DOG TAR	Range	0	0 None	Never	Fixed	100	3662	0	
BASVURUTAR	Range	0	0 None	Never	Fixed	100	3662	0	
ODEDONEM	Range	0	0 None	Never	Fixed	100	3662	0	
TMINTTUTAR	Range	64	28 None	Never	Fixed	100	3662	0	
FRIMTUTAR	Range	0	4 None	Never	Fixed	100	3662	0	
MASRAFI	Range	0	4 None	Never	Fixed	100	3662	0	
ODENEN_TA...	Range	7	21 None	Never	Fixed	100	3662	0	
BRANSKOD	Discrete	--	--	Never	Fixed	100	3662	0	
OF	Discrete	--	--	Never	Fixed	100	3662	0	
KOD	Discrete	--	--	Never	Fixed	100	3662	0	
ACIKLAMA	Discrete	--	--	Never	Fixed	100	3662	0	
TARIFENO	Discrete	--	--	Never	Fixed	100	3662	0	
STATUS	Discrete	--	--	Never	Fixed	100	3662	0	
STAT	Discrete	--	--	Never	Fixed	100	3662	0	
TUR	Discrete	--	--	Never	Fixed	100	3662	0	
DOVKOD	Discrete	--	--	Never	Fixed	100	3662	0	
ZEYLACIKLA...	Discrete	--	--	Never	Fixed	100	3662	0	
ZEYLTIP	Discrete	--	--	Never	Fixed	100	3662	0	
SURETIP	Discrete	--	--	Never	Fixed	100	3662	0	
CINSİYET	Discrete	--	--	Never	Fixed	100	3662	0	
RISKSNF	Discrete	--	--	Never	Fixed	100	3662	0	
MESLEK	Discrete	--	--	Never	Fixed	58,547	2144	1518	
ULKE	Discrete	--	--	Never	Fixed	99,672	3650	12	
IL	Discrete	--	--	Never	Fixed	100	3662	0	
ODETIP	Discrete	--	--	Never	Fixed	100	3662	0	
KKTIP	Discrete	--	--	Never	Fixed	100	3662	0	
BANKAADI	Discrete	--	--	Never	Fixed	100	3662	0	
TMINTRKOD	Discrete	--	--	Never	Fixed	100	3662	0	
ACKLMA	Discrete	--	--	Never	Fixed	100	3662	0	

Şekil 3.10. Data Audit Analiz Sonuçları

Bu analizin sonucunda, meslek değişkeninin eksik verilerden dolayı sonucu etkilememesi için çıkarılması gerektiğine karar verilmiştir. İstenilmeyen alanların Fitre Aracını (Filter Node) kullanarak şekil 3.9'da görüldüğü gibi filtreden geçirilerek Data

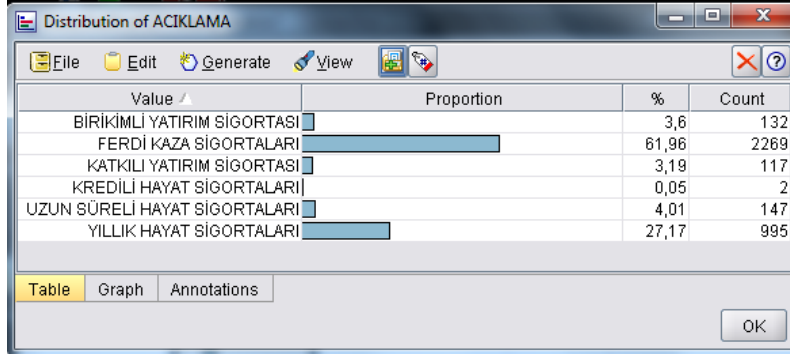
Audit aracına uygulanmıştır. Filtrenin üstüne çift tıklanarak çıkan ekranda Şekil 3.11’de görüldüğü gibi gizlenmek istenen alanları işaretleyerek (çarpı sembolü ile gösterilmiştir) kolonları filtrelenebilmektedir.



Şekil 3.11. Veri Setinin Filtreden Geçirilmesi

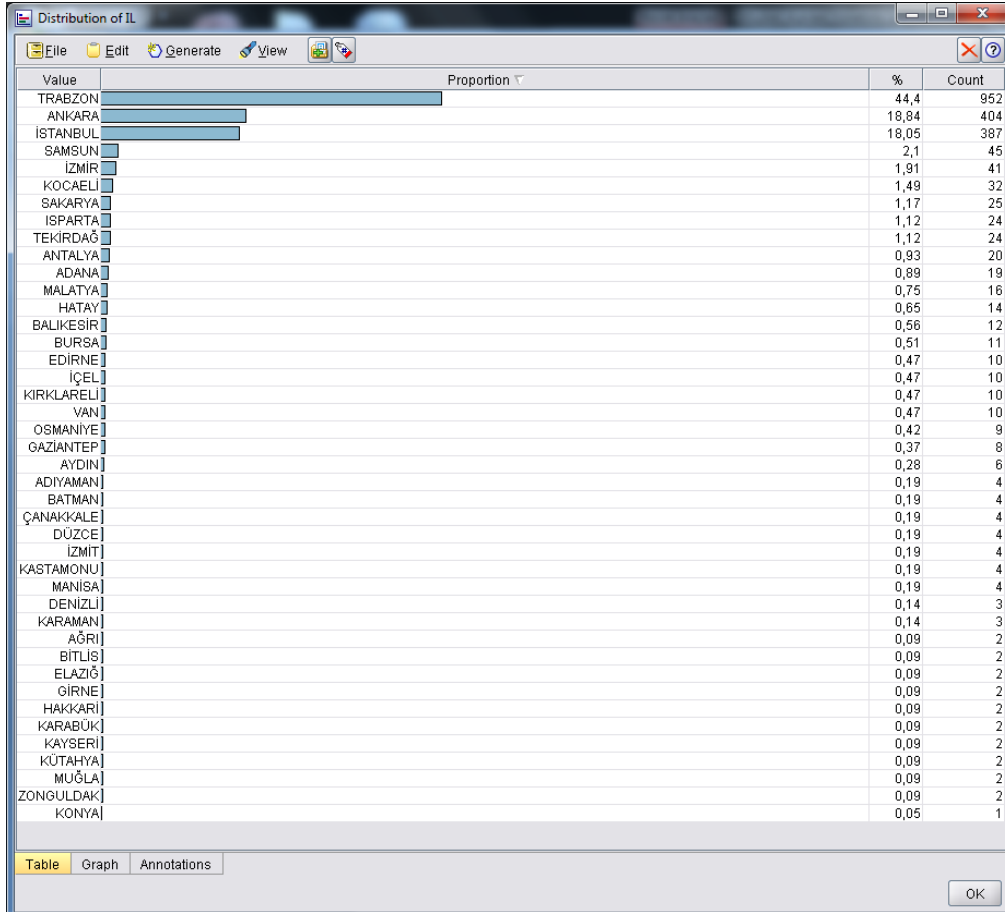
3.1.3.3. Verinin dönüştürülmesi ve veri setlerinin hazırlanması

Verilerin dönüştürülmesi aşamasından Clementine’da “Derive” aracı kullanılarak başvuru tarihinden, doğum tarihi çıkartılarak yaş olarak yeni kolon eklenmiştir. Aynı zamanda Derive kolonu kullanarak yaşlar belli aralıklarla dönüştürülmüştür.



Şekil 3.12. Açıklama Değişkeninin İncelenmesi

Şekil 3.12 de Hayat Sigortalarının ürün bazında ayrımı görülmektedir. %61,96 oranla en çok Ferdi Kaza Sigortaları tercih ediliyor olup, daha sonra Yıllık Hayat sigortaları %27,17 ile ardından %4,01 ile Uzun Süreli Hayat Sigortaları olup %3,6 ile Birikimli Yatırım Sigortası ve son olarak en az oranla %0,05 ile Kredili Hayat Sigortaları tercih edildiği yukarıdaki grafikte gözlenmektedir.



Şekil 3.13. İl Değişkeninin İncelenmesi

Şekil 3.13’de İl değişkeni bazında incelendiğinde bazı iller’in sayısının çok az olmasından dolayı en çok sayıda olanlar seçilerek ana gruplara bölünmüş ve diğer ilçeleri tek bir grupta toplayarak diğer olarak isimlendirilmiştir. Sonuç olarak Trabzon, Ankara, İstanbul, Diğer ilçelerden oluşmaktadır. Diğer ilçeler ortalama %3’ün altında ilçelerden oluşmaktadır.

Oracle SQL’de oluşturulan orijinal veri setlerini kullanarak algoritmalara uygun olarak sorgularla farklı veri setleri oluşturulmuştur. Görüntüler(View) haline yani, veri tabanındaki tablonun istenilen kolonlarını istenilen sınırlama ve koşullarla alınmasını sağlayacak biçime dönüştürülmüştür. Oluşturulan bu görüntüler(view) veri setlerini oluşturmaktadır.

Birliktelik analizi yapabilmek için verilerin Tablo 3.1’de görüldüğü gibi Tabular Data olarak adlandırılan biçime dönüştürülmesi gerekmektedir.

Tablo 3.1’deki gibi müşteri bazında incelendiğinde ürünlerin satın alınıp alınmadıklarına yönelik oluşturulmuş bir tablo yapısına sahipse Apriori algoritmalarına uygulanabilecek veri yapısına sahiptir. Örneğin, Tablo 3.1’deki gibi birliktelik kuralı algoritmasına uygulanacak olan veri setinin farklı poliçeler üzerinde satın alınıp alınmadığını ifade etmek amacıyla T(True) ve F(False) ile işaretlenmiştir.

Tablo 3.1. Tabular Data Formatı

Müşteri	X Poliçesi	Y Poliçesi	Z Poliçesi
1	T	F	F
2	F	F	T
3	T	T	F
4	T	T	T
5	F	T	T
6	F	F	T
7	T	F	T

Bu yüzden birliktelik analizi yapmadan önce verileri “musteri_id” bazında, satın alınan poliçe türleri 1 ve 1’den fazla ise “T” değilse “F” şeklinde kodlanmıştır. Tablo 3.2’de “Birliktelik Veri Seti” için kullanılacak olan veri seti “MİY_FK” ve “MİY” tabloları birleştirilerek hazırlanmış olan sorgu cümlesi gözükmetedir.

Tablo 3.2. Birliktelik Analizi Veri Seti

Birliktelik Analizi Veri Seti

```
CREATE OR REPLACE VIEW v_birliktelik as
```

```
select sk_aseg, (case when sum(kaza_kasko)>=1 then 'T' else 'F' end)
kaza_kasko, (case when sum(kaza_trafik)>=1 then 'T' else 'F' end)
kaza_trafik, (case when sum(yangin)>=1 then 'T' else 'F' end) yangin, (case
when sum(muhtelif_kaza)>=1 then 'T' else 'F' end) muhtelif_kaza, (case when
sum(nakliyat)>=1 then 'T' else 'F' end) nakliyat, (case when
sum(muhendislik)>=1 then 'T' else 'F' end) muhendislik, (case when
sum(destekli_tarim_sigortasi)>=1 then 'T' else 'F' end)
destekli_tarim_sigortasi, (case when sum(saglik)>=1 then 'T' else 'F' end)
saglik, (case when sum(dask)>=1 then 'T' else 'F' end) dask
from MİY_fk where police_tipi='BIREYSEL'
group by sk_aseg
```

```
union
```

```
select sk_aseg,(case when sum(kaza_kasko)>=1 then 'T' else 'F' end)
kaza_kasko, (case when sum(kaza_trafik)>=1 then 'T' else 'F' end)
kaza_trafik, (case when sum(yangin)>=1 then 'T' else 'F' end) yangin, (case
when sum(muhtelif_kaza)>=1 then 'T' else 'F' end) muhtelif_kaza, (case when
sum(nakliyat)>=1 then 'T' else 'F' end) nakliyat, (case when
sum(muhendislik)>=1 then 'T' else 'F' end) muhendislik, (case when
sum(destekli_tarim_sigortasi)>=1 then 'T' else 'F' end)
destekli_tarim_sigortasi, (case when sum(saglik)>=1 then 'T' else 'F' end)
saglik, (case when sum(dask)>=1 then 'T' else 'F' end) dask
from MİY where police_tipi= 'BIREYSEL'
group by sk_aseg;
```

Tablo 3.2’deki sorguda bir müşterinin birden fazla poliçesi olduğundan dolayı hangi poliçe türlerini tercih ettiğini tespit edebilmek için “v_birliktelik” adında bir görüntü(view) oluşturulmuştur. Oluşturulan v_birliktelik görüntüsü bundan sonraki aşamalarda anlaşılabilirliği açısından Birliktelik Veri Seti olarak adlandırılacaktır.

Birliktelik analizi için kullanılacak olan verilerin son hali Şekil 3.14’de görüldüğü gibidir.

	SK_ASEG	KAZA_KASKO	KAZA_TRAFIK	YANGIN	MUHTELIF_KAZA	NAKLIYAT	MUHENDISLIK	DESTEKLI_TARIM_SIGORTASI	SAGLIK	DASK	FERDI_KF
1	7890.000	T	T	F	F	F	F	F	F	F	F
2	8209.000	T	T	F	F	F	F	F	F	F	F
3	8827.000	F	F	F	F	F	F	F	F	T	F
4	15680.000	F	T	F	F	F	F	F	F	F	F
5	17154.000	F	T	F	F	F	F	F	F	F	F
6	8893.000	T	T	F	F	F	F	F	F	T	F
7	13743.000	F	T	F	F	F	F	F	F	F	F
8	245.000	T	T	F	F	F	F	F	F	F	F
9	18321.000	F	T	F	F	F	F	F	F	F	F
10	17102.000	T	T	F	F	F	F	F	F	F	F
11	14669.000	F	T	F	F	F	F	F	F	F	F
12	17824.000	T	T	T	F	F	F	F	F	T	F
13	10107.000	F	F	F	F	F	F	F	F	T	F
14	10860.000	T	T	F	F	F	F	F	F	F	F
15	17949.000	F	T	F	F	F	F	F	F	T	F
16	14865.000	T	F	F	F	F	F	F	F	F	F
17	13082.000	F	F	T	F	F	F	F	F	T	F
18	11115.000	F	T	F	F	F	F	F	F	F	F
19	11743.000	F	T	F	F	F	F	F	F	F	F
20	11856.000	F	F	F	F	F	F	F	F	T	F
21	14720.000	F	F	F	F	F	F	F	F	T	F
22	19139.000	F	T	F	F	F	F	F	F	F	F
23	18581.000	F	T	F	F	F	F	F	F	F	F
24	10379.000	F	T	F	F	F	F	F	F	F	F
25	18962.000	T	T	F	F	F	F	F	F	F	F
26	674.000	F	F	T	F	F	F	F	F	F	F

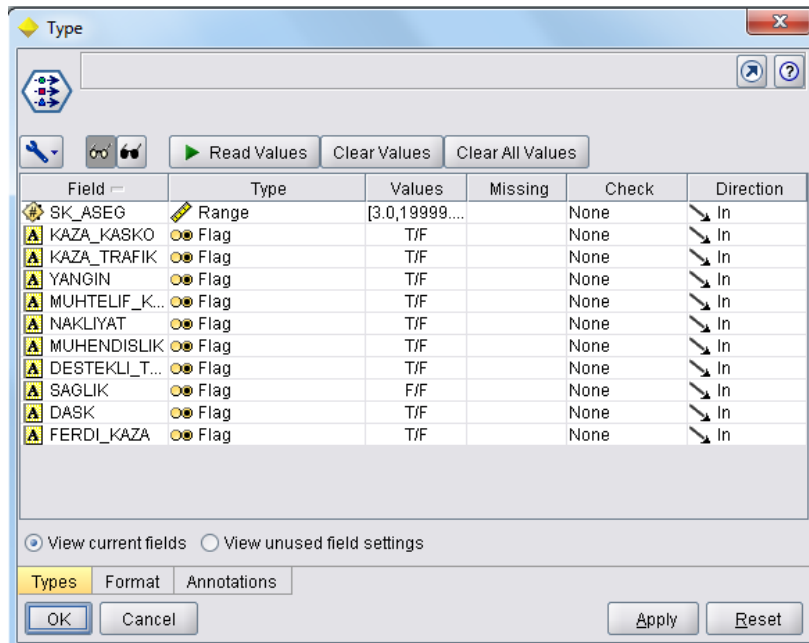
Şekil 3.14. Birliktelik Veri Setinden Kesit

Tablo 3.3’de Birliktelik Analizinde kullanılan veri setinin alanları ve açıklamaları bulunmaktadır.

Tablo 3.3. Birliktelik Veri Setindeki Alan ve Açıklamalar

Kolon Adı	Açıklama
SK_ASEG	Müşteri numarasının tutulduğu alan
KAZA_KASKO	Oto Sigortaları
KAZA_TRAFIK	Trafik Sigortaları
DESTEKLI_TARIM_SIGORTASI	Destekli Tarım Sigortası
YANGIN	Yangın
NAKLIYAT	Nakliyat
MUHTELIF_KAZA	Muhtelif Kaza Sigortası
MUHENDISLIK	Mühendislik Sigortaları
DASK	Konut Sigortaları(Zorunlu Deprem Sigortası)

Şekil 3.15’de görülebileceği üzere birliktelik analizinde kullanılacak olan verilerin son hali Clementine programı üzerinden değişkenlerin tipleri ve verilerin değerleri gözlenmektedir.



Şekil 3.15. Birliktelik Analizi Veri Setinin Alanlarının Türleri

Tablo 3.4’de kümeleme analizinde kullanılmak üzere müşterilerin demografik özelliklerini içeren “kumeleme_analizi” adı altında veri seti oluşturulmuştur. Kümeleme veri seti oluşturulurken “GENELHAYAT” tablosundan yararlanılmıştır.

Tablo 3.4. Kümeleme Veri Seti

Kümeleme Veri Seti

```
CREATE OR REPLACE VIEW kumeleme_analizi as
select h.pid, h.cinsiyet, h.acklma, h.meslek,
substr(meslek,1,(instr(h.meslek,')-1)) meslekler, netprim, (case when
h.il='ANKARA' then 'ANKARA' when h.il='İSTANBUL' then
'İSTANBUL' when h.il='TRABZON' then 'TRABZON' else 'DİĞER' end)
iller,round((basvurutar-dogtar)/365) yas
from genelhayat h;
```

Kümeleme Veri Setinden örnek bir ekran görüntü aşağıdaki şekil 3.16’deki gibidir.

	PID	CINSİYET	ACKLMA	MESLEK	İL	DOGTAR	BASVURUTAR	YAS
▶ 1	1109863	K	Kaza Sonucu Vefat	DİĞER	KOCAELİ	18.01.1960	11.01.2010	50
2	1109863	K	Kaza Sonucu Maluliyet	DİĞER	KOCAELİ	18.01.1960	11.01.2010	50
3	1109863	K	Kaza Sonucu Vefat	DİĞER	KOCAELİ	18.01.1960	11.01.2010	50
4	1110158	E	Vefat Teminatı	MÜDÜR	İZMİR	19.04.1970	05.01.2010	40
5	1117668	E	Kaza Sonucu Vefat	DİĞER	İSTANBUL	16.03.1987	05.01.2010	23
6	1117668	E	Kaza Sonucu Maluliyet	DİĞER	İSTANBUL	16.03.1987	05.01.2010	23
7	1117668	E	Kaza Sonucu Vefat	DİĞER	İSTANBUL	16.03.1987	05.01.2010	23
8	1110054	E	Vefat Teminatı	SİGORTACI	BURSA	25.03.1971	06.01.2010	39
9	1110054	E	Vefat Teminatı	SİGORTACI	BURSA	25.03.1971	06.01.2010	39
10	1110058	K	Vefat Teminatı	SİGORTACI	İSTANBUL	05.06.1971	04.01.2010	39
11	1110059	E	Vefat Teminatı	SİGORTACI	İSTANBUL	10.02.1971	04.01.2010	39
12	1117671	E	Deprem Dahil Kaza Sonucu Vefat	DİĞER	ANKARA	14.03.1983	06.01.2010	27
13	1117671	E	Vefat Teminatı	DİĞER	ANKARA	14.03.1983	06.01.2010	27
14	1110126	E	Deprem Dahil Hast. Tedavi Masrafları	REKLAMCI	İSTANBUL	07.06.1963	14.01.2010	47
15	1110126	E	Deprem Dahil Kaza Sonucu Maluliyet	REKLAMCI	İSTANBUL	07.06.1963	14.01.2010	47
16	1110126	E	Deprem Dahil Kaza Sonucu Vefat	REKLAMCI	İSTANBUL	07.06.1963	14.01.2010	47
17	1129227	E	Kaza Sonucu Vefat	DİĞER	İSTANBUL	07.04.1966	02.02.2010	44
18	1129227	E	Kaza Sonucu Maluliyet	DİĞER	İSTANBUL	07.04.1966	02.02.2010	44
19	1129227	E	Kaza Sonucu Vefat	DİĞER	İSTANBUL	07.04.1966	02.02.2010	44
20	1129227	E	Deprem Dahil Kaza Sonucu Vefat	DİĞER	İSTANBUL	07.04.1966	02.02.2010	44
21	1129227	E	Kaza Sonucu Maluliyet	DİĞER	İSTANBUL	07.04.1966	02.02.2010	44
22	1129227	E	Kaza Sonucu Vefat	DİĞER	İSTANBUL	07.04.1966	02.02.2010	44
23	1055471	E	Deprem Dahil Kaza Sonucu Vefat	MÜHENDİS	KOCAELİ	02.02.1967	10.02.2010	43

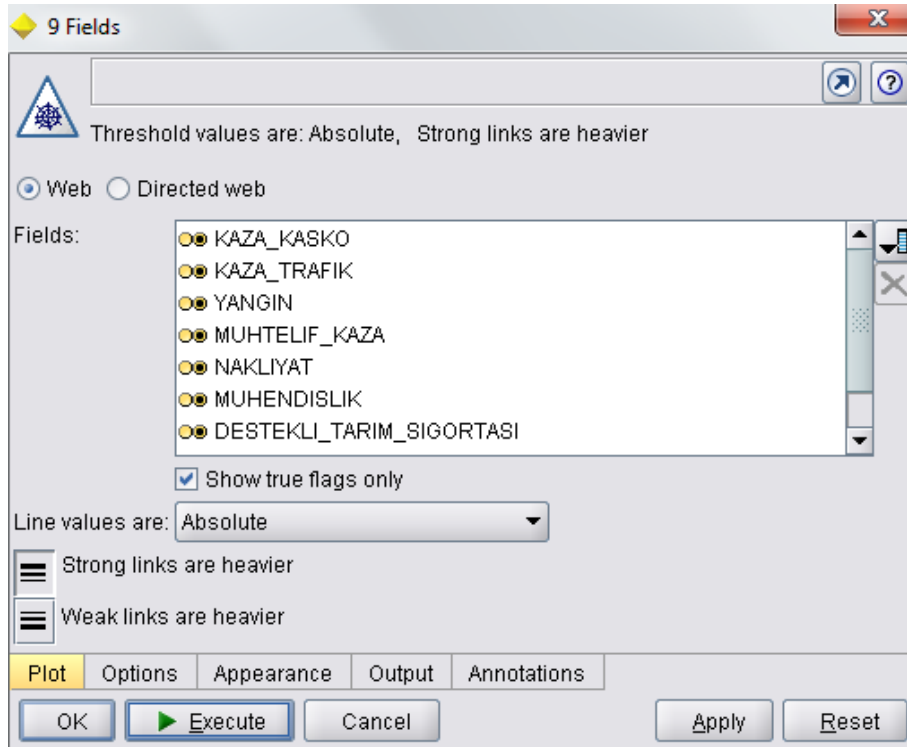
21 hr@ORADB 23 rows selected in 0.031 seconds (more...)

Şekil 3.16. Kümeleme Veri Setinin Veri Tabanından Görüntüsü

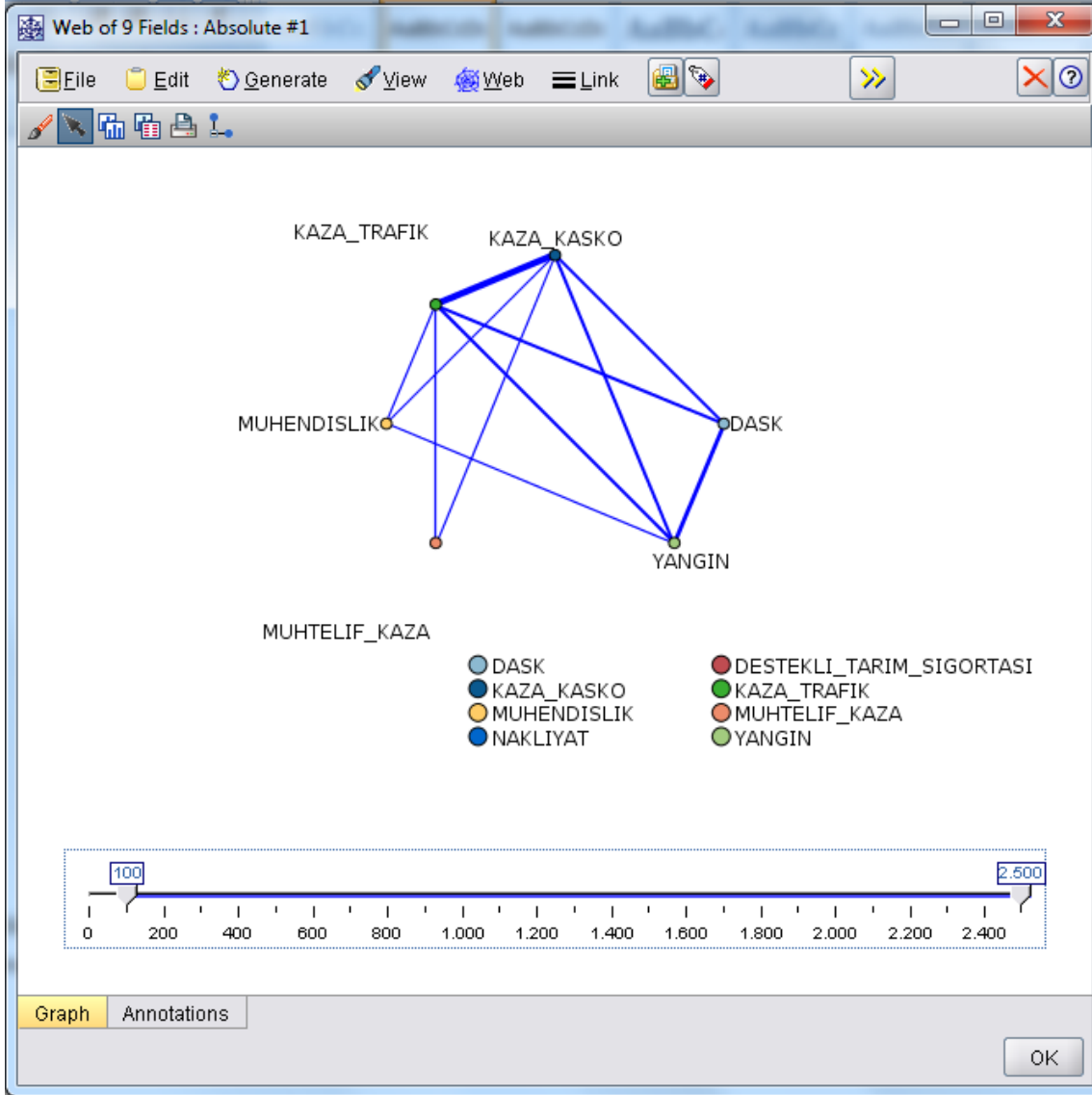
Uygulama aşamalarında veri setlerinin takibi açısından isimlendirme yapılmıştır. Birliktelik analizi için kullanılan v_birliktelik veri setine **birliktelik veriseti**, kümeleme analizi için kullanılacak olan kumeleme_analizi veri seti **kümeleme veri seti** olarak isimlendirilmiştir.

3.1.3.4. Veri setlerinin analizi

Birliktelik Kuralları Analizinde kullanılacak veri setine modeli kurmadan önce verileri daha iyi anlayabilmek için analize tabi tutulmuştur. Clementine Programının Web Grafiği aracı kullanılarak ürün gruplarının arasındaki ilişkiler incelenmiştir. Şekil 3.17’de görüleceği gibi birliktelik veri setinde bulunan tüm alanlar analize dahil edilmiştir.



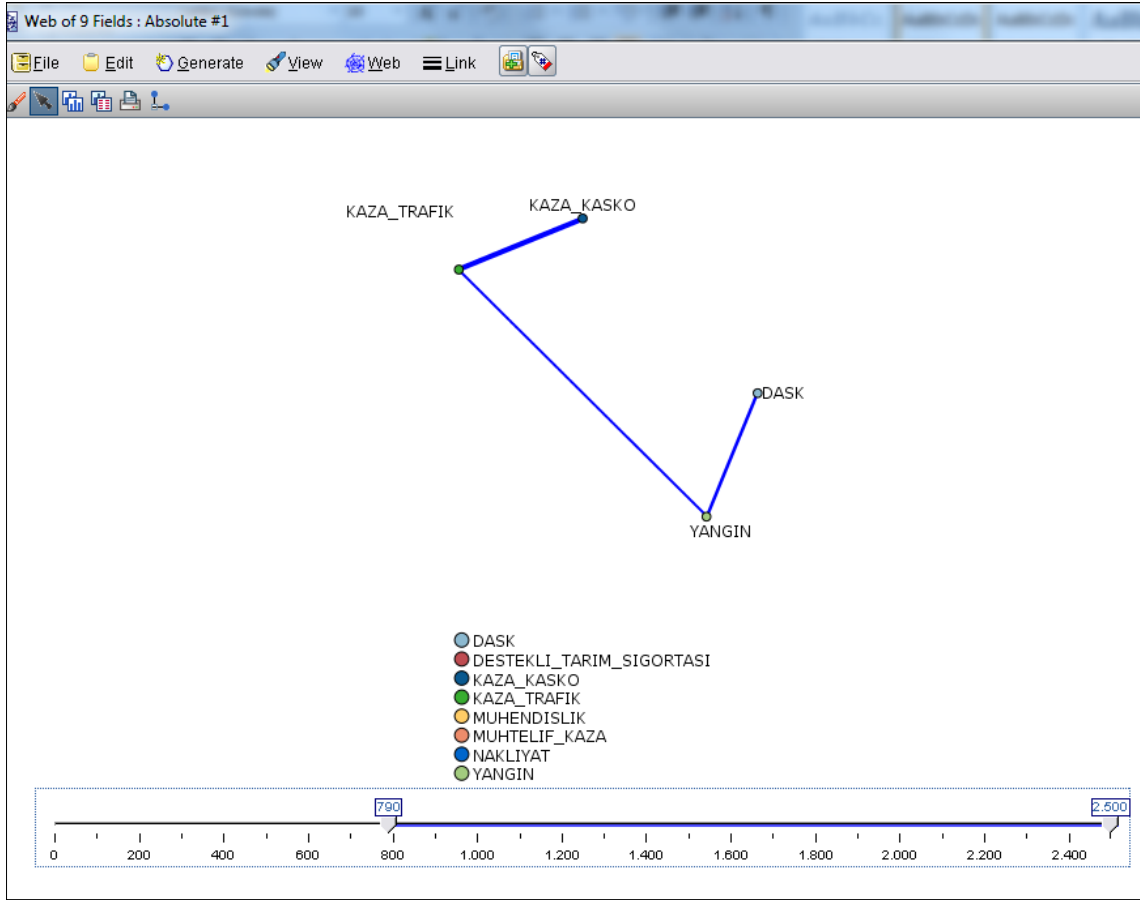
Şekil 3.17. Web Grafiği Aracı



Şekil 3.18. Gruplanmış Sigorta Türleri Arasındaki Bağlantılar

Sigorta türlerinin birlikte alınmasına dair bağlantıları gösteren grafik Şekil 3.18'deki gibidir. Çizgilerin kalınlığı birlikte bulunma sıklıklarını ifade etmektedir.

Şekil 3.19'da görülebileceği gibi minimum birliktelik sayısı 100'den 790'a yükseltildiğinde, 790 ile 2.500 aralığındaki bir arada bulunma durumu şekildeki gibi olacaktır.



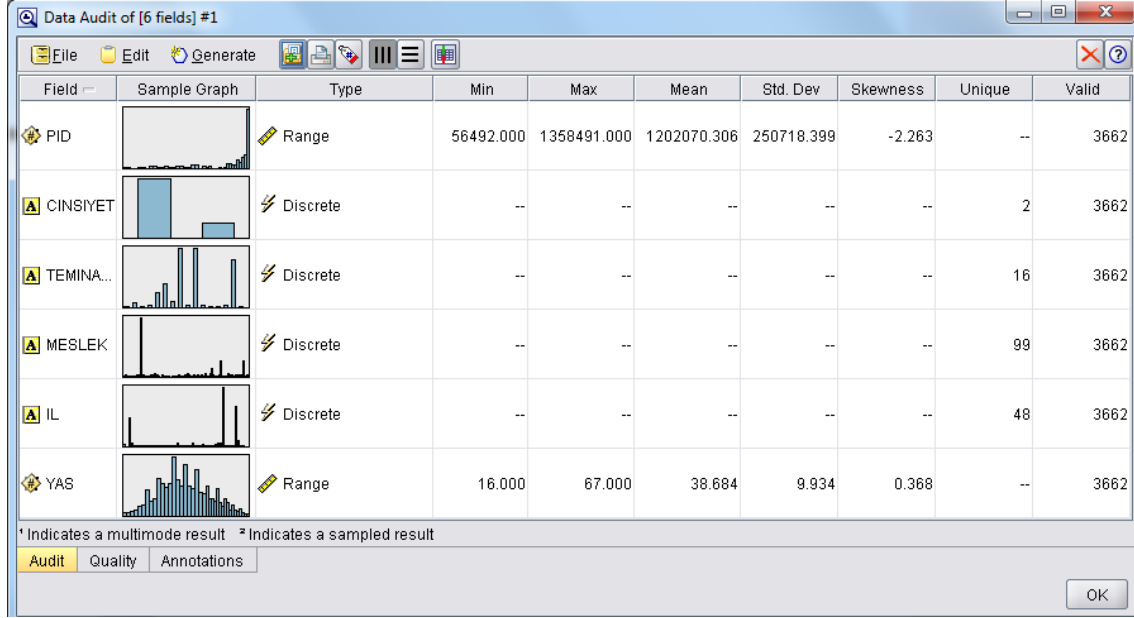
Şekil 3.19. Sıklığı Yüksek Olan Bağlantılar

Bu analizlerin sonucunda en yüksek birliktelikler şu şekildedir:

- KAZA_TRAFIK ve KAZA_KASKO 24.765'nin 3.654'ünde birlikte satın alınmışlardır. Bu değer yaklaşık olarak %14,75'ini oluşturmaktadır.
- DASK ve YANGIN 1.875'inde birlikte satın alınmıştır. Bu sayı %7,57'sini oluşturmaktadır.
- KAZA_TRAFIK ve YANGIN 1.042'sinde birlikte satın alınmış olup, %4,02'sini oluşturmaktadır.

Bu analiz sonucunda veri setinin birliktelik analizine uygun yapıda olduğu saptanmıştır. Modelin kurulması aşamasında bu değişkenler üzerinde sonuca yönelik bilgi aranmıştır.

Aynı şekilde Kümeleme veri setini daha iyi anlayabilmek için modellemeden önce Data Audit aracına uygulayarak veriler incelenmektedir. Şekil 3.20’de kümeleme veri setinin Data Audit aracının analiz sonuçları verilmiştir.



Şekil 3.20. Kümeleme Veri Seti Analizi

Şekil 3.20’deki Data Audit ekranında her bir değişkenin adı, türü, ortalaması, standart sapması, en büyük ve en küçük değerleri gibi istatistiksel sonuçlar görülmektedir. PID(Person Id) alanı müşteri id’sini temsil etmektedir. Bu yüzden aşağıda analiz sonuçlarını incelerken müşteri id alanı anlamlı değer ifade etmeyeceğinden dolayı dikkate alınmamıştır.

Analiz sonucundaki değişkenlerin hayat poliçelerindeki dağılımları şu şekildedir:

Cinsiyet değişkeni;

Value	Proportion	%	Count
E		81,19	2973
K		18,81	689

Şekil 3.21. Cinsiyet Değişkeninin Veriler Üzerindeki Etkisi

Şekil 3.21’de görülebileceği gibi cinsiyet değişkeninin hayat sigortalarına etkisi erkeklerde, kadınlara oranla çoğunluk göstermektedir. %81,19 u erkek müşterilerden oluşurken %18’i bayan müşterilerden oluşmaktadır.

Teminat deęiřkeni Őekil 3.22’de grleceęi zere en fazla Kaza Sonucu Vefat Teminatı %27,88 talep grrken, ardından Kaza Sonucu Maluliyet %25,55 ile talep grmektedir.

Value /	Proportion	%	Count
Asistans Hizmetleri		0,05	2
Asist.		1,56	57
Deprem Dahil Gndelik Tazminat		0,08	3
Deprem Dahil Hast. Tedavi Masrafları		0,27	10
Deprem Dahil Kaza Sonucu Maluliyet		5,68	208
Deprem Dahil Kaza Sonucu Vefat		10,49	384
Hastalık Sonucu Maluliyet		2,32	85
Kaza Sonucu Maluliyet		27,55	1009
Kaza Sonucu Tedavi Masrafları		0,41	15
Kaza Sonucu Vefat		27,88	1021
Mezuniyet Hastalık Sonucu Maluliyet		0,25	9
Mezuniyet İř Kaybı		0,25	9
Mezuniyet Kazaen Maluliyet		0,25	9
Mezuniyet Kazen Vefat		0,25	9
Vefat Teminatı		22,64	829
Yařam Destek		0,08	3

Őekil 3.22. Teminatlara Gre Daęılım İstatistiklięi

Őekil 3.23’de hayat polielerinin satıřlarının mesleklere gre daęılımı gzlemlenmektedir. Analiz sonularına Hayat Poliesi tercih eden en ok meslek grubu %17,17 ile İřilerden oluřmaktadır. Daha sonrasında %16,75 ile Tccar, % 9,63 Sigortacı, %3,13 Ev Hanımı, %3.08 Serbest Meslek, %2,99 Ynetici,%2,75 Mhendis ve %2,18 Memur yzdelik oranları azalarak devam etmektedir.

Őekil 3.23’de grlen meslek grupları daęılımı analizde kullanılmadan nce, %1,85’den kk olanlar birleřtirilip “Dięer” adı altında toplanmıřtır. Bunun amacı meslek deęiřkeninin daha anlamlı bir sonu ıkarmasına olanak saęlamaktadır.

Value	Proportion	%	Count
İŞÇİ		17,17	362
TÜCCAR		16,75	353
SİGORTACI		9,63	203
EV HANIMI		3,13	66
SERBEST MESLEK		3,08	65
YÖNETİCİ		2,99	63
MÜHENDİS		2,75	58
MEMUR		2,18	46
SATIŞ ELEMANI		2,18	46
AVUKAT		2,04	43
MUHASEBECİ		2,04	43
ÖĞRETMEN		2,04	43
ŞÖFÖR		1,85	39
OTO TAMIRCISI		1,61	34
FİRMA SAHİBİ		1,38	29
PAZARLAMACI		1,28	27
İNŞAAT İŞÇİSİ		1,14	24
DOKTOR		1,09	23
MÜŞAVİR		1,09	23
GENEL MÜDÜR		1,04	22
TAMIRCI		1,04	22
TURİZMCİ		1,04	22
ECZACI		1,0	21
MALİ MÜŞAVİR		1,0	21
DİŞ HEKİMİ		0,95	20
ÖĞRENCİ		0,95	20
EMEKLİ		0,9	19
İŞLETMECİ		0,9	19
DANIŞMAN		0,76	16
ESNAF		0,76	16
İMALATÇI		0,76	16
MÜDÜR		0,71	15
MİMAR		0,62	13
REKLAMCI		0,62	13
PAZARCI		0,57	12
ASİSTAN		0,52	11
BALIKÇI		0,47	10
İŞVEREN		0,47	10
MONTAJ ELEMANI		0,47	10
ŞEF		0,47	10
MUHASEBE MÜD.		0,43	9
EKONOMİST		0,38	8
BOYACI		0,33	7
MONTAJ SORUMLU.		0,33	7
TEKNİSYEN		0,33	7

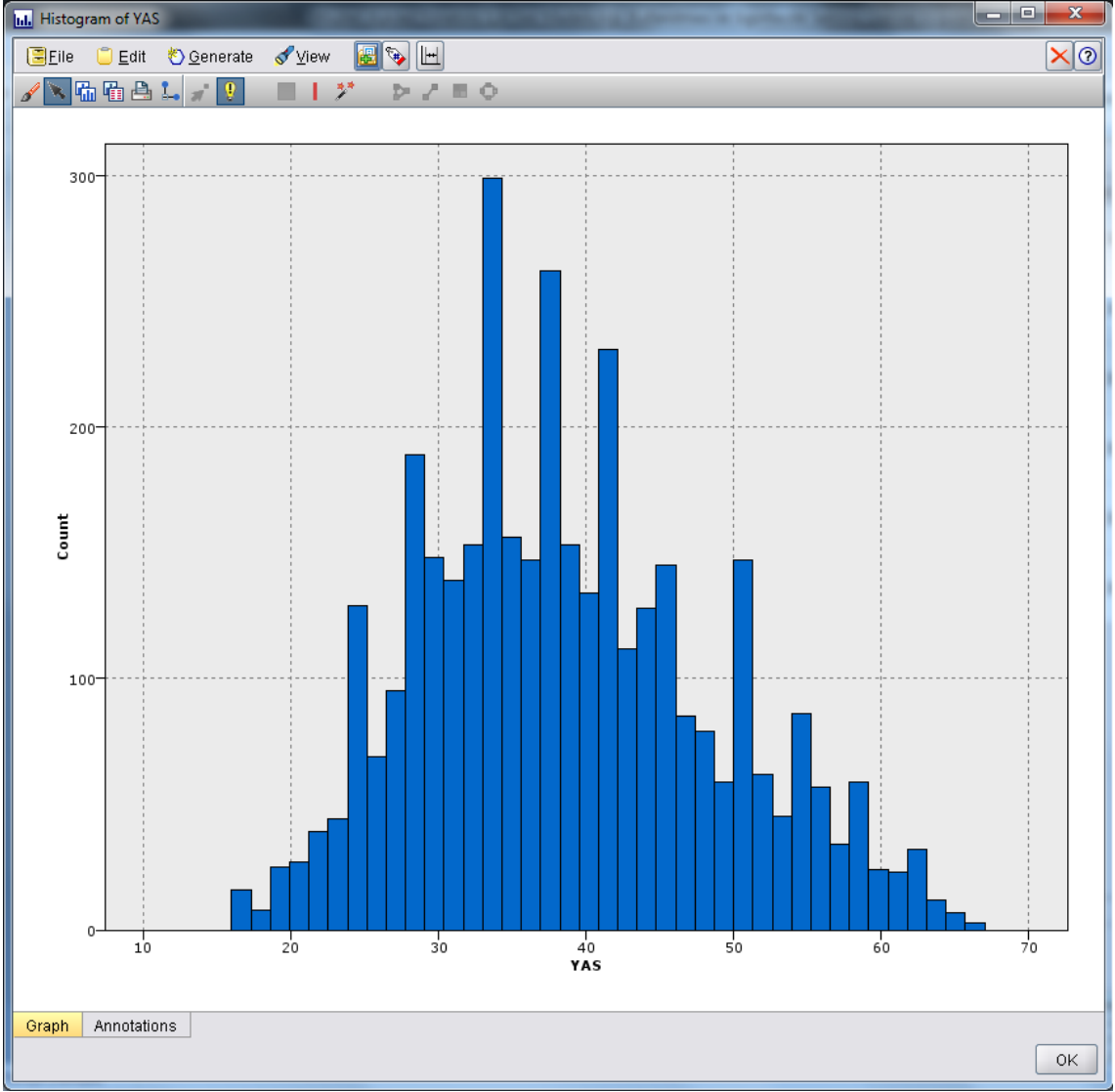
Şekil 3.23. Meslek Değişkeninin Veri Setine Etkisi

İllere göre dağılım grafiği Şekil 3.24'deki gibidir. Meslek gruplarında olduğu gibi %17'den küçük olanlar birleştirilip "Diğer" adı altında toplanmıştır. %37 oranla Trabzon, % 27,53 oranla İstanbul ve %17,64 oranla Ankara şehirleri analize dahil edilmiş olup, bu illerin haricindeki şehirler "Diğer" olarak adlandırılmıştır.

Value	Proportion	%	Count
TRABZON		37,85	1386
İSTANBUL		25,53	935
ANKARA		17,64	646
İZMİR		3,36	123
KOCAELİ		1,91	70
SAMSUN		1,39	51
ANTALYA		1,37	50
ISPARTA		1,2	44
TEKİRDAĞ		1,06	39
SAKARYA		0,85	31
İÇEL		0,6	22
BURSA		0,57	21
ADANA		0,55	20
MALATYA		0,55	20
GAZİANTEP		0,46	17
HATAY		0,44	16
MANİSA		0,44	16
BALIKESİR		0,38	14
BATMAN		0,38	14
KIRKLARELİ		0,33	12
VAN		0,33	12
AYDIN		0,27	10
EDİRNE		0,27	10
OSMANİYE		0,27	10
MUĞLA		0,22	8
ÇANAKKALE		0,16	6
ADİYAMAN		0,11	4
DÜZCE		0,11	4
GİRESUN		0,11	4
İZMİT		0,11	4
KASTAMONU		0,11	4
KAYSERİ		0,11	4
DENİZLİ		0,08	3
DİYARBAKIR		0,08	3
KARAMAN		0,08	3
AĞRI		0,05	2
BITLİS		0,05	2
ÇORUM		0,05	2
ELAZIĞ		0,05	2
ESKİŞEHİR		0,05	2
GİRNE		0,05	2
HAKKARİ		0,05	2
KARABÜK		0,05	2
KONYA		0,05	2

Şekil 3.24. İllere göre dağılım grafiği

Şekil 3.25’de ise Yaş Dağılımının histogram grafiği görülmektedir. Maksimum değere yaklaşık 34 yaşında ulaşıldığı gözlenmektedir.



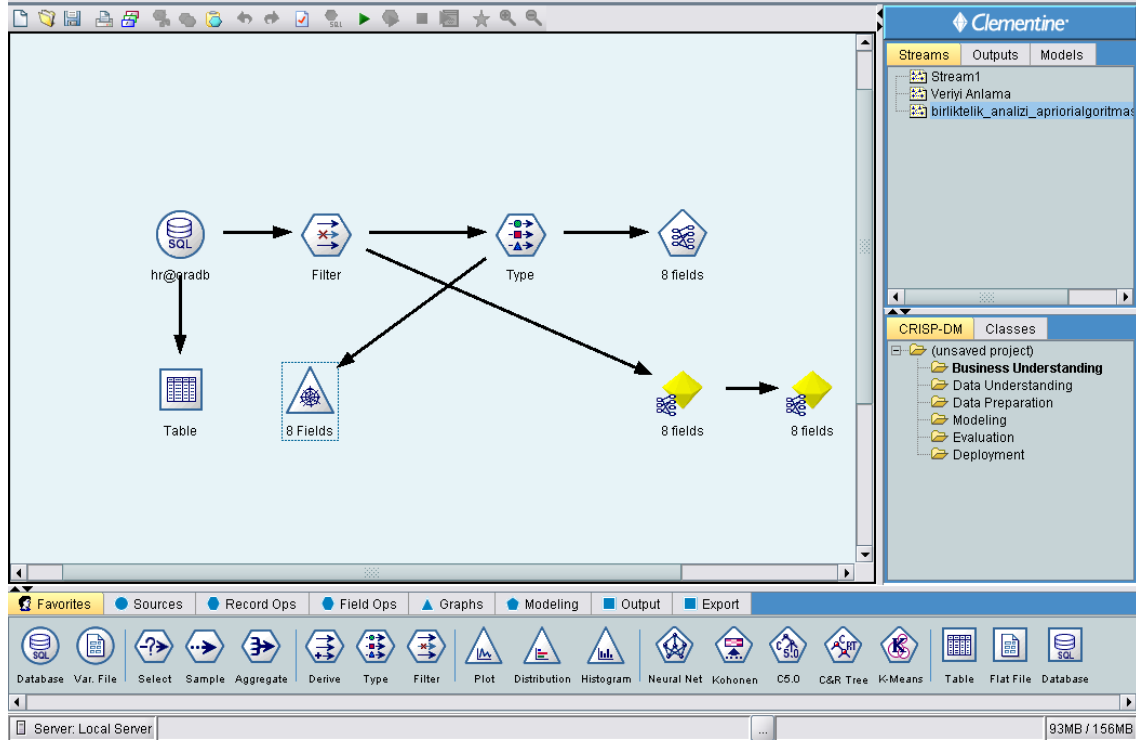
Şekil 3.25. Yaş Dağılımı Grafiği

3.1.4. Modelleme aşaması (Modeling)

Bu aşamada, modelleme tekniğinin seçilmesi ve bu modellere ilişkin parametrelerin optimal değerlere ayarlanması gibi çalışmaların tümü yer almaktadır. Uygun modelleme tekniğinin seçimi, test tasarımının üretimi, model geliştirme ve tahmin işlemlerini içermektedir.

3.1.4.1. Birliktelik analizinin uygulanması

Apriori Algoritması



Şekil 3.26. Apriori Algoritmasının Uygulanması

Uygulamanın bu aşamasında, mantıksal ilişki kuralları içinde en çok kullanılan veri madenciliği modellerinden birisi olan Apriori algoritması kullanılmıştır. Şekil 3.26’da Tablo 3.2’de oluşturulmuş olan “**birliktelik veri setine**” apriori algoritması uygulandığındaki model yapısı görülmektedir.

Şekil 3.27’de apriori algoritmasının sonuçları görülmektedir. Şekilde “Confidence” güven, “Rule Support” destek, “Consequent” sonuçta birlikte alınmış olan ürünü, “Antecedent” ise hangi ürün veya ürünlerle birlikte alındığını temsil etmektedir. Gözlenen sonuçlar şu şekildedir:

Bu sonuçlara göre en yüksek oranlara sahip olan ilk üç birliktelik sonucu ele alınmış olup bunlar şu şekildedir:

- KAZA_KASKO'su satın alan müşterilerin %63,58'i KAZA_TRAFİK yani "Trafik Kazası Durumunda Kasko Sigortası" da satın almıştır ve bu kişiler toplam müşterilerin %17,342'sini oluşturmaktadır.
- YANGIN sigortası alanların %54,96'sı DASK sigortası da satın almıştır ve bu kişiler toplam kişilerin %8,95'ini oluşturmaktadır.
- DASK sigortası alanların %47,38'i YANGIN sigortası da satın almıştır ve bu kişiler toplam kişilerin %8,95'ini oluşturmaktadır.

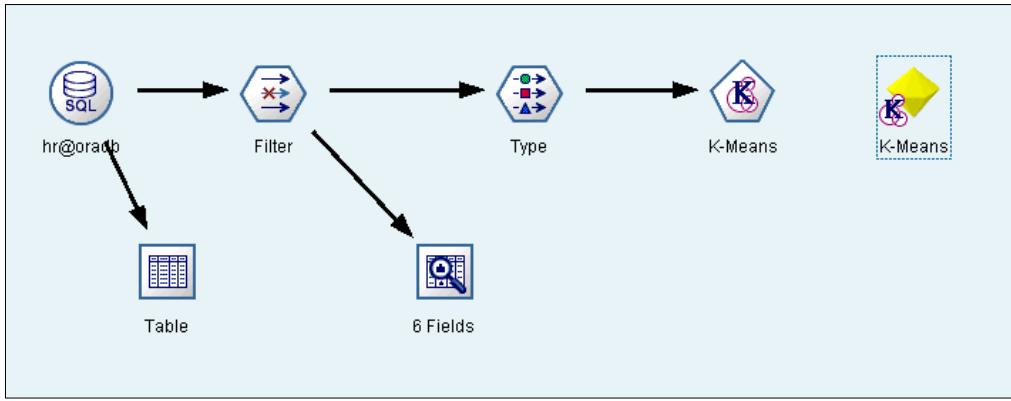
Consequent	Antecedent	Confidence %	Rule Support %
KAZA_TRAFIK	KAZA_KASKO	63,585	17,342
DASK	YANGIN	54,962	8,95
YANGIN	DASK	47,382	8,95
KAZA_TRAFIK	YANGIN	33,832	5,509
KAZA_KASKO	YANGIN	32,293	5,258
KAZA_TRAFIK	DASK	27,765	5,244
KAZA_KASKO	KAZA_TRAFIK	24,077	17,342
YANGIN	KAZA_KASKO KAZA_TRAFIK	23,775	4,123
KAZA_KASKO	DASK	22,308	4,214
YANGIN	KAZA_KASKO	19,28	5,258
DASK	KAZA_KASKO KAZA_TRAFIK	18,635	3,232
DASK	KAZA_KASKO	15,449	4,214

Şekil 3.27. Apriori Analiz Sonuçları

3.1.4.2. Kümeleme analizinin uygulanması

K-Ortalamlar Analizi

En popüler kümeleme analizlerinden birisi olan K-Ortalamlar analizi Tablo 3.4’de oluşturulmuş olan “**kümeleme veri setine**” Şekil 3.28’de görüleceği gibi uygulanmıştır. K-Ortalamlar tekniği ile yapılan analiz sonucu üç küme elde edilmiştir.

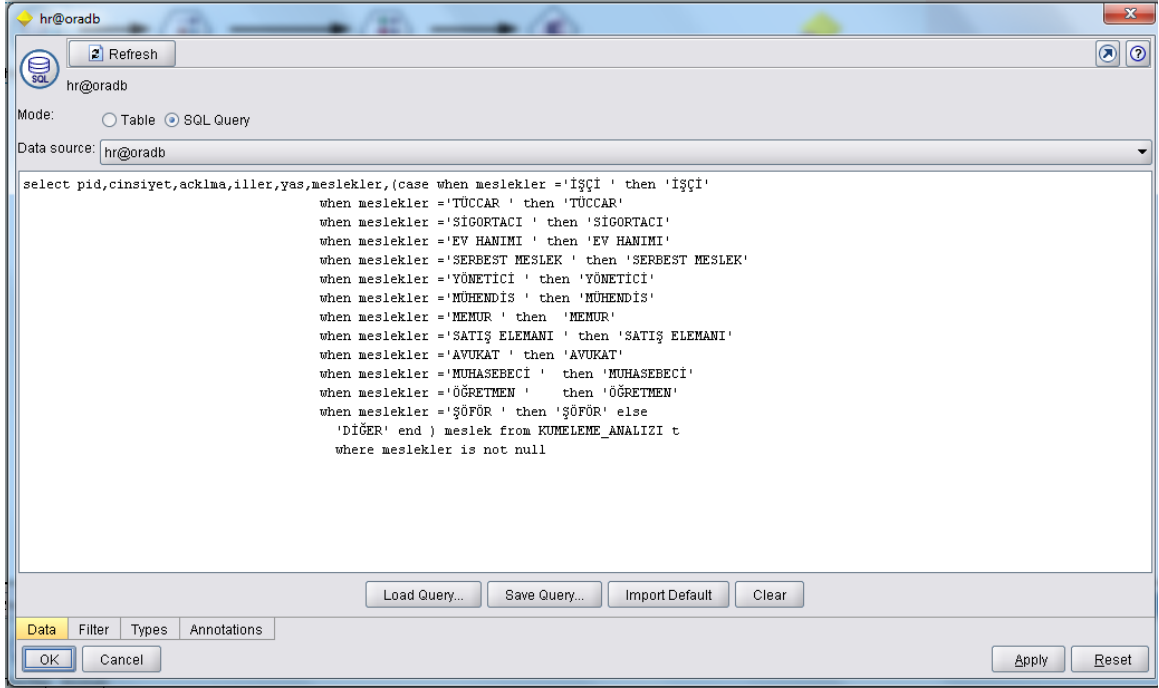


Şekil 3.28. K-Ortalamlar Analizinin Uygulanması

Kümeleme analizi yapılmadan önce SQL sorgusu oluşturulurken kümeleme veri setindeki meslekler alanı alınırken daha önceki yapılmış olan analizlere göre (bkz: Şekil 3.23) %1,85’den az olan meslek grupları seçime dâhil edilmemiş “Diğer” olarak isimlendirilmiştir.

Şekil 3.29’da görüleceği gibi dönüşümü veritabanı aracının(database node) içine sorgu yazılırken, az sayıdaki değerler DİĞER adı altındaki başlıkta toplanmıştır.

Tablo 3.5’de veritabanı aracının içine yazılmış olan SQL deyimini görülmektedir.



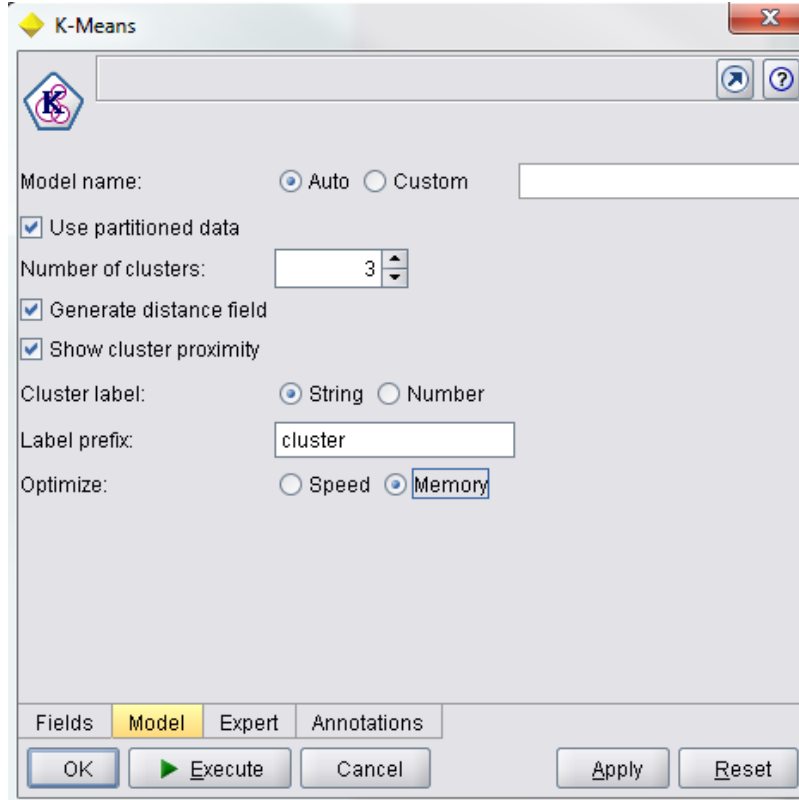
Şekil 3.29. Veritabanı Aracında Sorguların Düzenlenmesi

Tablo 3.5. Meslek Grupları Dönüşümü

```

select pid, cinsiyet, acklma, iller, yas, meslekler, (case when meslekler ='İŞÇİ '
then 'İŞÇİ'
when meslekler ='TÜCCAR ' then 'TÜCCAR'
when meslekler ='SİGORTACI ' then 'SİGORTACI'
when meslekler ='EV HANIMI ' then 'EV HANIMI'
when meslekler ='SERBEST MESLEK ' then 'SERBEST MESLEK'
when meslekler ='YÖNETİCİ ' then 'YÖNETİCİ'
when meslekler ='MÜHENDİS ' then 'MÜHENDİS'
when meslekler ='MEMUR ' then 'MEMUR'
when meslekler ='SATIŞ ELEMANI ' then 'SATIŞ ELEMANI'
when meslekler ='AVUKAT ' then 'AVUKAT'
when meslekler ='MUHASEBECİ ' then 'MUHASEBECİ'
when meslekler ='ÖĞRETMEN ' then 'ÖĞRETMEN'
when meslekler ='ŞÖFÖR ' then 'ŞÖFÖR' else 'DİĞER' end ) meslek
from KUMELEME_ANALIZI t
where meslekler is not null

```



Şekil 3.30. K-Ortalamlar Modelinin Kurulma Aşaması

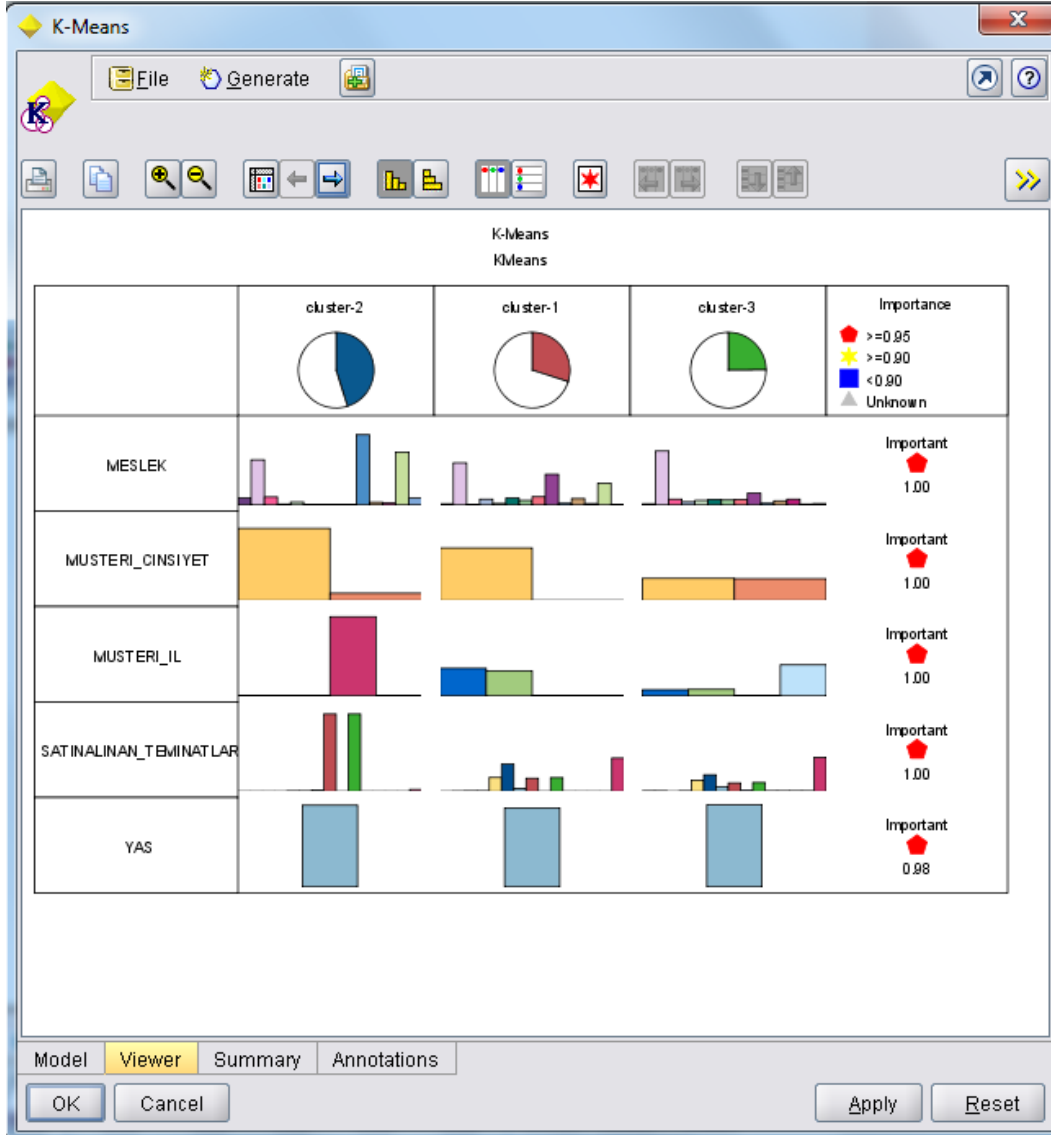
Şekil 3.30'da görüleceği üzere K-Ortalamlar algoritması seçilirken küme sayısı(Number of clusters) üç olarak seçilmiştir.

Kümeleme analizinde kullanılan değişkenler aşağıdaki gibidir:

- Müşteri Cinsiyeti
- Müşteri İli
- Meslek
- Yaş
- Satın Alınan Teminatlar

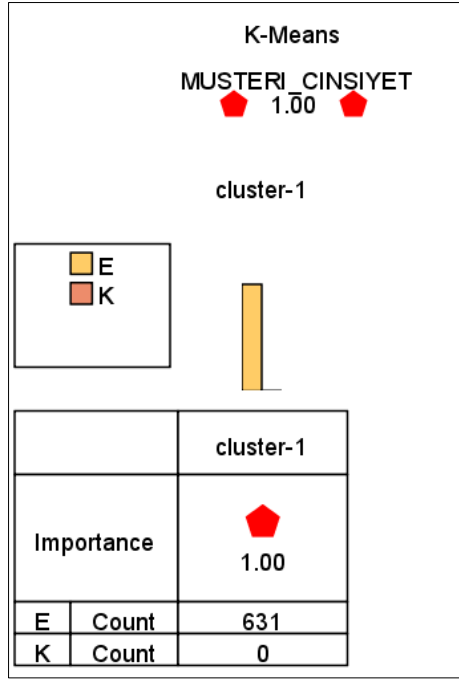
Clementine programı tarafından K-Ortalamlar algoritmasının sonuçları incelendiğinde önem derecesi %95'den büyük olan değişkenler önemli(important) değişken adı altında gözükmektedir. Değişkenin önemli(important) veya önemsiz(unimportant) olmasını gösteren önem değeri, değişkenin kümeler arasındaki olasılığındaki varyasyonuna göre değişiklik gösterir. Önem değeri, 1 değerinden, istatistiksel testin anlamlılık değeri

çıkartılarak bulunmaktadır. Anlamlılık değeri, kategorik değerler için ki-kare testi(chi-squares test), sürekli değişkenlerde t-testi (t-test or Student's t test) değeri ile belirlenmektedir. Şekil 3.31'de K-ortalamlar sonucunda elde edilen sonuçların değişkenler üzerindeki önem değerleri görülmektedir.



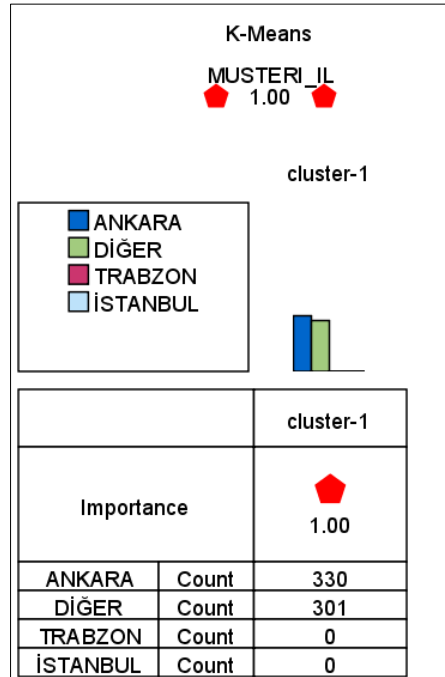
Şekil 3.31. K-Ortalamlar Yöntemi ile Elde Edilen Kümeler

Şekil 3.31'de K-ortalamlar yöntemi ile elde edilen kümeler görülmektedir. Bu sonuçlara göre, birinci küme(Cluster-1)'de meslek, müşteri cinsiyeti, müşteri il, satın alınan teminatlar ve yaş değişkenlerinin önemli(important) değişkenler olduğu görülmektedir.



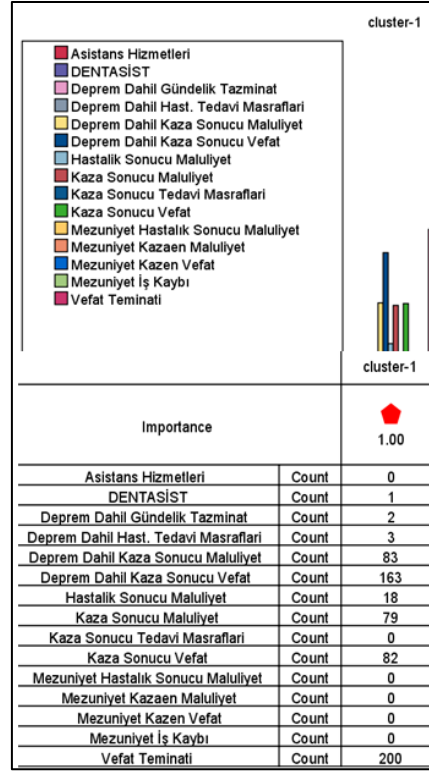
Şekil 3.32. Müşteri Cinsiyetinin Birinci Küme Üzerindeki Etkisi

Şekil 3.32’de görüleceği gibi MUSTERI_CINSIYET değişkeninin birinci küme-1(Cluster-1) üzerine etkisi incelendiğinde erkek müşterilerden oluştuğu gözlenmektedir. Şekil 3.2’deki count ifadesi kümenin içinde bulunma sayısını ifade etmektedir.



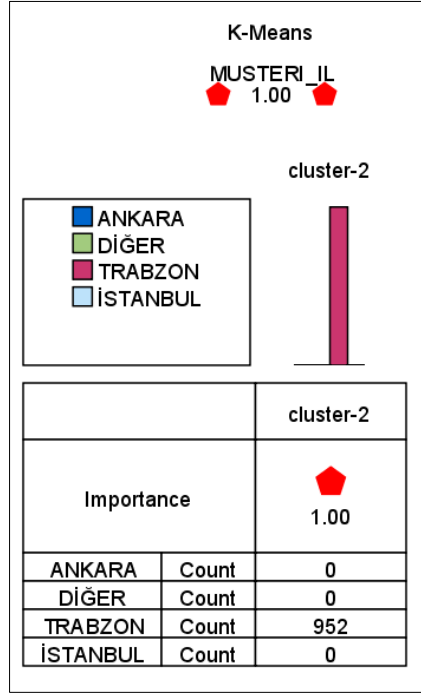
Şekil 3.33. Müşteri İl Değişkeninin Birinci Küme Üzerindeki Etkisi

Şekil 3.33'deki MUSTERI_IL değişkeni incelendiğinde Ankara il'i çoğunlukta gözükmektedir.



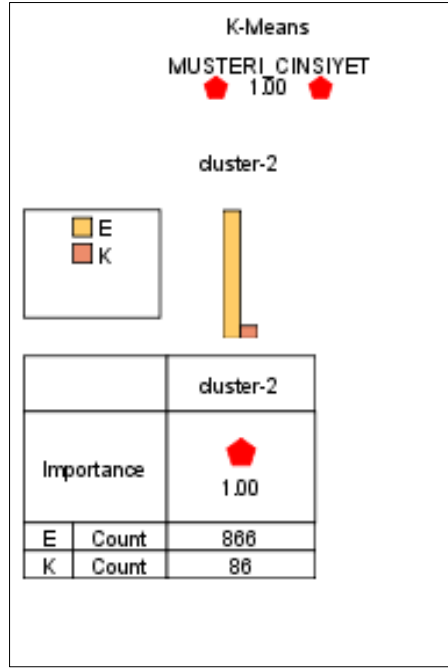
Şekil 3.34. Satın Alınan Teminat Değişkeninin Birinci Küme Üzerindeki Etkisi

Şekil 3.34'de görüldüğü üzere, Satın alınan teminatların birinci küme üzerine etkisi incelendiğinde en fazla satın alınan teminat Vefat Teminatı tercih edenlerden oluşmaktadır. İkinci en çok tercih edilen teminat ise Deprem Dahil Kaza Sonucu Vefat teminatıdır.



Şekil 3.35. Müşteri İl'i değişkenin İkinci Küme Üzerine Etkisi

K-ortalamlar yöntemi ile ikinci küme-2(cluster-2)'ye etki eden MUSTERI_IL, MUSTERI_CINSIYETI, MESLEK ve SATINALINAN_TEMINATLAR değişkenlerinden oluşmaktadır. Şekil 3.35'de görüldüğü gibi MUSTERI_IL değişkenin Küme-2(Cluster-2) üzerindeki etkisi incelenecek olunursa "Trabzon" şehrinde yaşayanlardan oluşmakta olduğu gözlenmektedir.



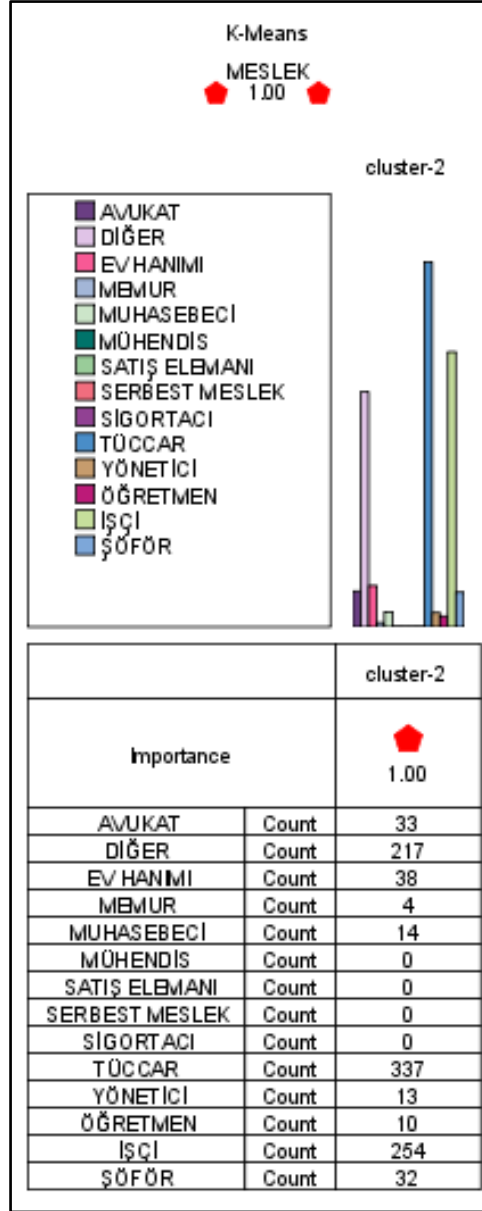
Şekil 3.36. Müşteri Cinsiyetinin İkinci Küme Üzerindeki Etkisi

MUSTERI_CINSIYETI değişkeninin Küme-2(Cluster-2) üzerine etkisi incelendiğinde Şekil 3.36'de görüleceği üzere Erkek(E) müşterilerin çoğunlukta olduğu görülmektedir.

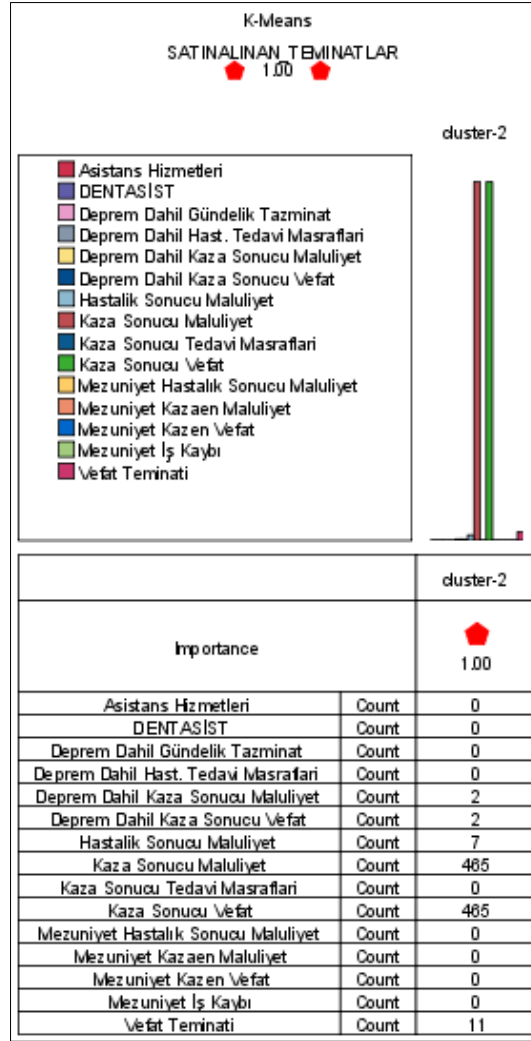
MESLEK değişkeni incelenecek olursa, Şekil 3.37'de görüleceği üzere çoğunluk Tüccar'lardan oluştuğu gözlenmektedir.

SATINALINAN_TEMINATLAR değişkeninin Küme-2(Cluster-2) üzerine etkisi incelenecek olursa, Şekil 3.38'de görüldüğü üzere teminatlarda Kaza Sonucu Maluliyet ve Kaza Sonucu Vefat teminatı rekabet görmektedir.

Tüm değişkenlerin Küme-2(Cluster-2) üzerindeki etkileri incelendiğinde, Trabzon ilinde yaşayan erkek müşterilerin çoğunlukta olduğu, ayrıca Satın alınan teminat bazında inceleme yapıldığında en çok Kaza Sonucu Vefat Teminatı ve Kaza Sonucu Maluliyet teminatları rekabet görmesinin yanı sıra Tüccar sınıfı meslek grubundaki müşterilerin çoğunlukta olduğu görülmektedir.

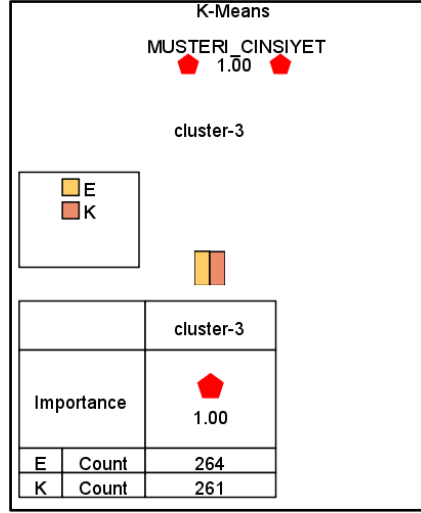


Şekil 3.37. Meslek Değişkeninin İkinci Küme Üzerine Etkisi

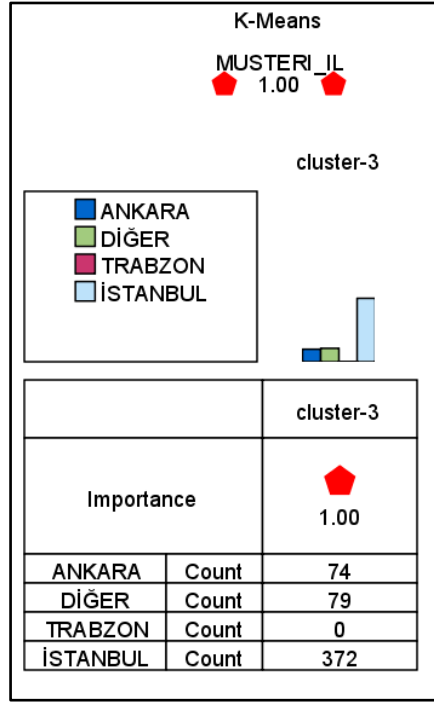


Şekil 3.38.Satın Alınan Teminatların İkinci Küme Üzerindeki Etkisi

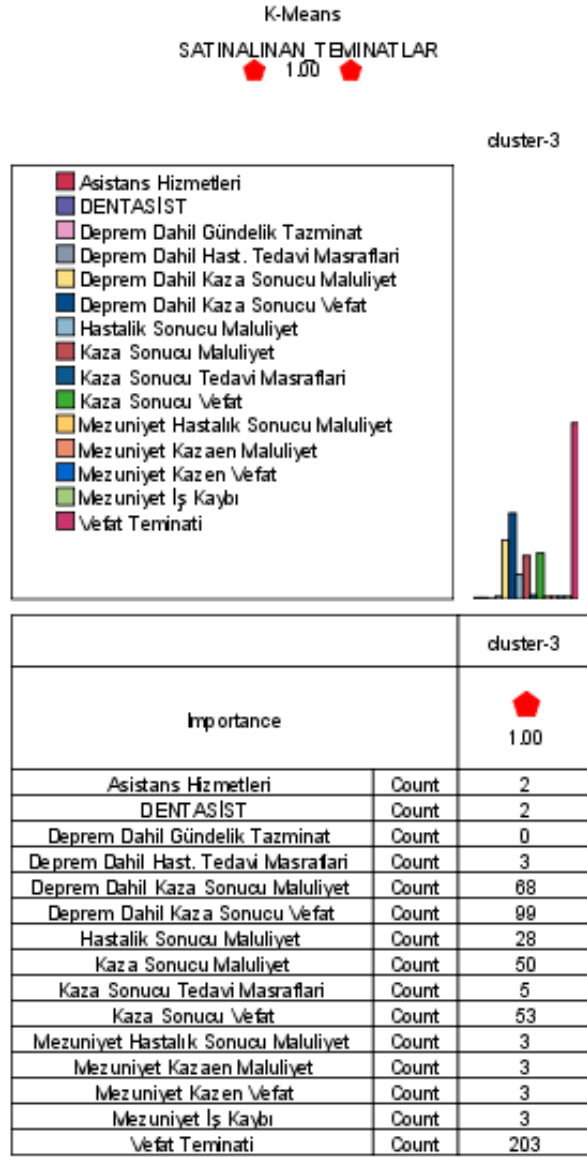
K-ortalamlar yöntemi ile elde edilen üçüncü küme-3(cluster-3)'ye etki eden değişkenler incelendiğinde, Şekil 3.39'da CINSIYET değişkeninin küme-3(cluster-3) üzerine etkisi, Erkekler ile Kadınların sayıları yaklaşık olarak birbirine eşittir. Şekil 3.40'da MUSTERI_IL değişkeni incelendiğinde, İstanbul'da yaşayanlar çoğunluğu oluşturmaktadır. Şekil 3.41'de SATINALINAN_TEMINATLAR değişkeni bazında incelendiğinde, tercih edilen teminat türü Vefat Teminatının çoğunlukta olduğu görülmektedir.



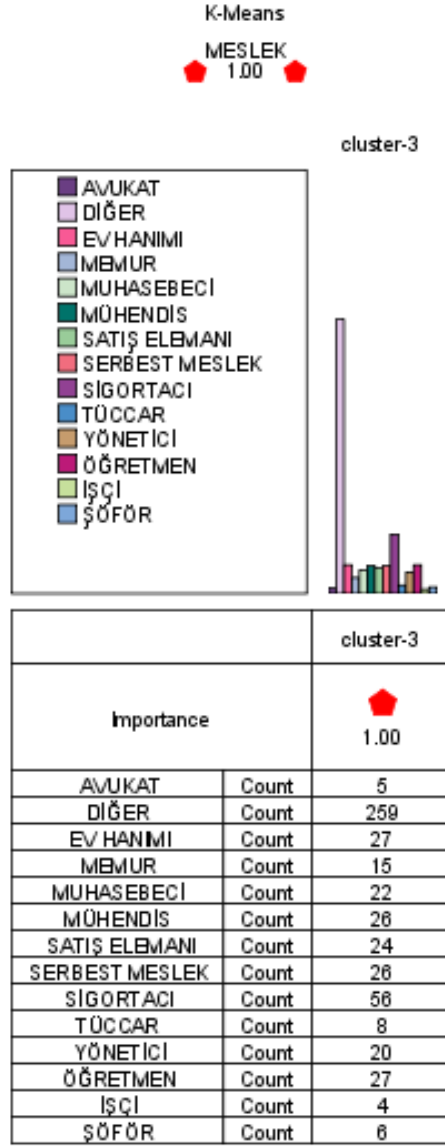
Şekil 3.39. Müşteri Cinsiyetinin Üçüncü Küme Üzerine Etkisi



Şekil 3.40. Müşteri İl Değişkeninin Üçüncü Küme Üzerine Etkisi

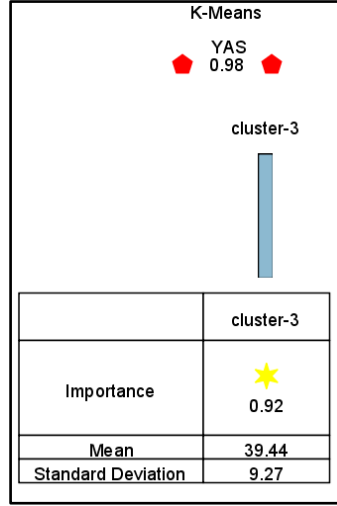


Şekil 3.41. Satın Alınan Teminatların Üçüncü Küme Üzerine Etkisi



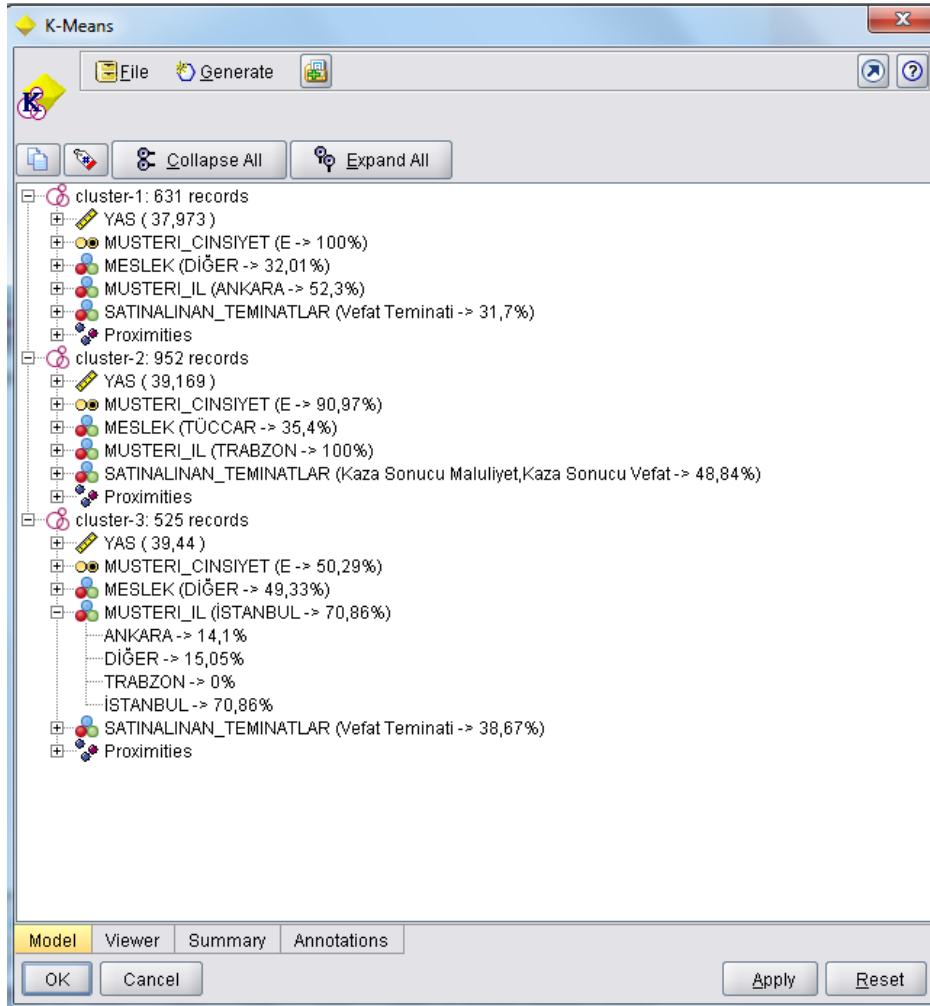
Şekil 3.42. Meslek Değişkeninin Üçüncü Küme Üzerine Etkisi

Şekil 3.42’de MESLEK değişkeninin Küme-3(Cluster-3) üzerine etkisi incelendiğinde En çok DIĞER olan meslek grubu görülmektedir. Diğer meslek grubu, çalışmamızda sayıları az olan farklı mesleklerden oluşmaktadır. Bu yüzden uygulamanın sonucunda anlamlı gibi gözükse de, aslında anlamlı bir ifade değildir. Şekil 3.43’de YAS değişkeni Küme-3(Cluster-3) üzerine etkisi incelendiğinde ortalama 39 yaş olan müşterilerden oluşturduğu gözlenmektedir.

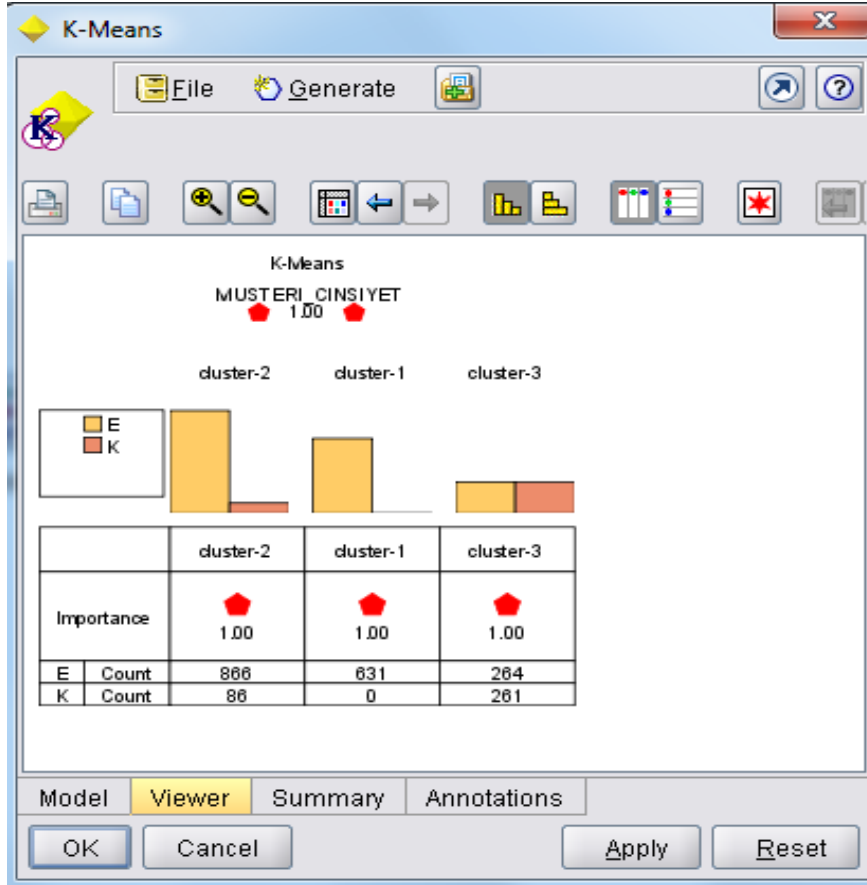


Şekil 3.43. Yaş Değişkeninin Üçüncü Küme Üzerine Etkisi

K-ortalamalar yönteminde elde edilen üç küme için özet modeli Şekil 3.44'deki gibidir.

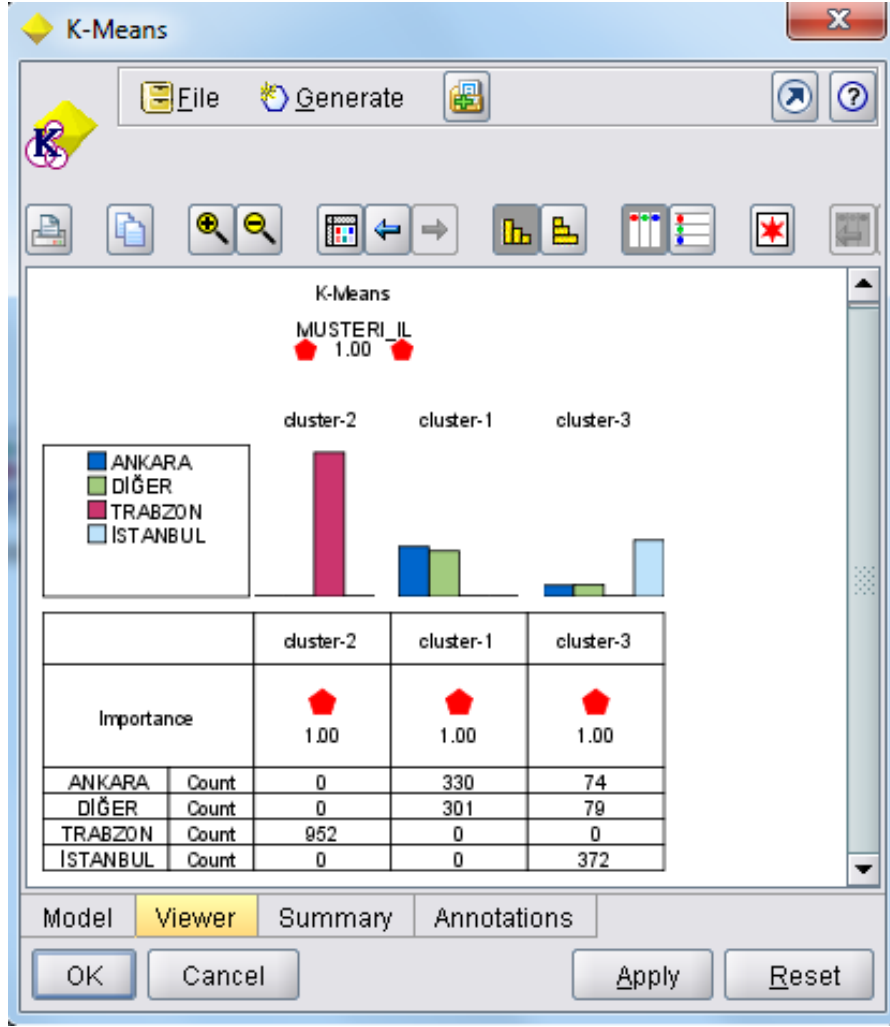


Şekil 3.44. K-Ortalamalar Analizi Model Özeti



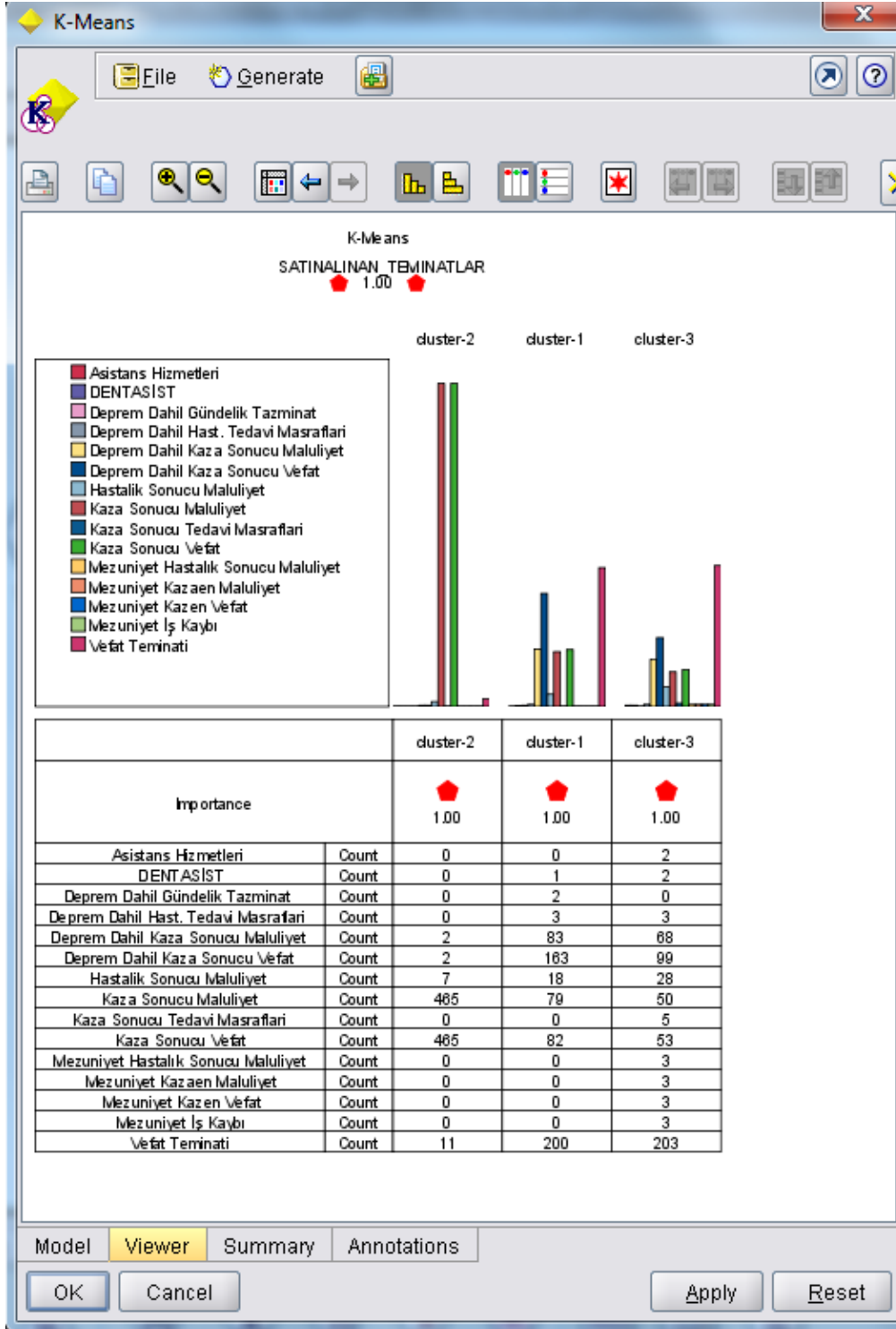
Şekil 3.45. Müşteri Cinsiyeti Değişkeninin Kümeler Üzerindeki Etkisi

Şekil 3.45’de MUSTERI_CINSIYET değişkeninin kümeler üzerine etkisi incelendiğinde, bu değişkenin her üç küme için de önemli olduğu görülmektedir. Her üç kümede de erkek müşterilerin çoğunlukta olduğu görülmektedir.



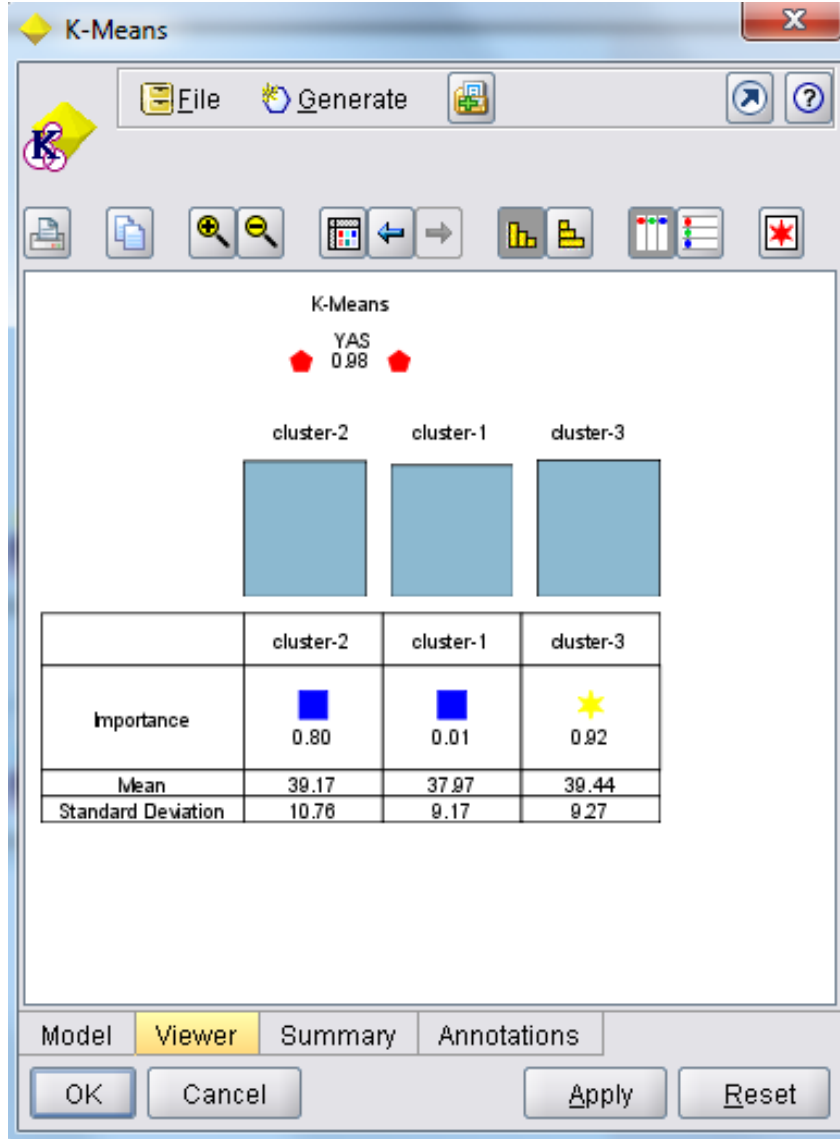
Şekil 3.46. Müşteri İl'i Değişkeninin Kümeler Üzerindeki Etkisi

Şekil 3.46'da MUSTERI_IL değişkeni incelendiğinde, bu değişkenin her üç küme içinde önemli bir değişken olduğu gözlenmektedir. Birinci küme(Cluster-1)'de Ankara, İkinci küme(Cluster-2)'de Trabzon şehir grubu, Üçüncü küme(Cluster-3)'de ise İstanbul'da yaşayan müşterilerin çoğunlukta olduğu gözlenmektedir.



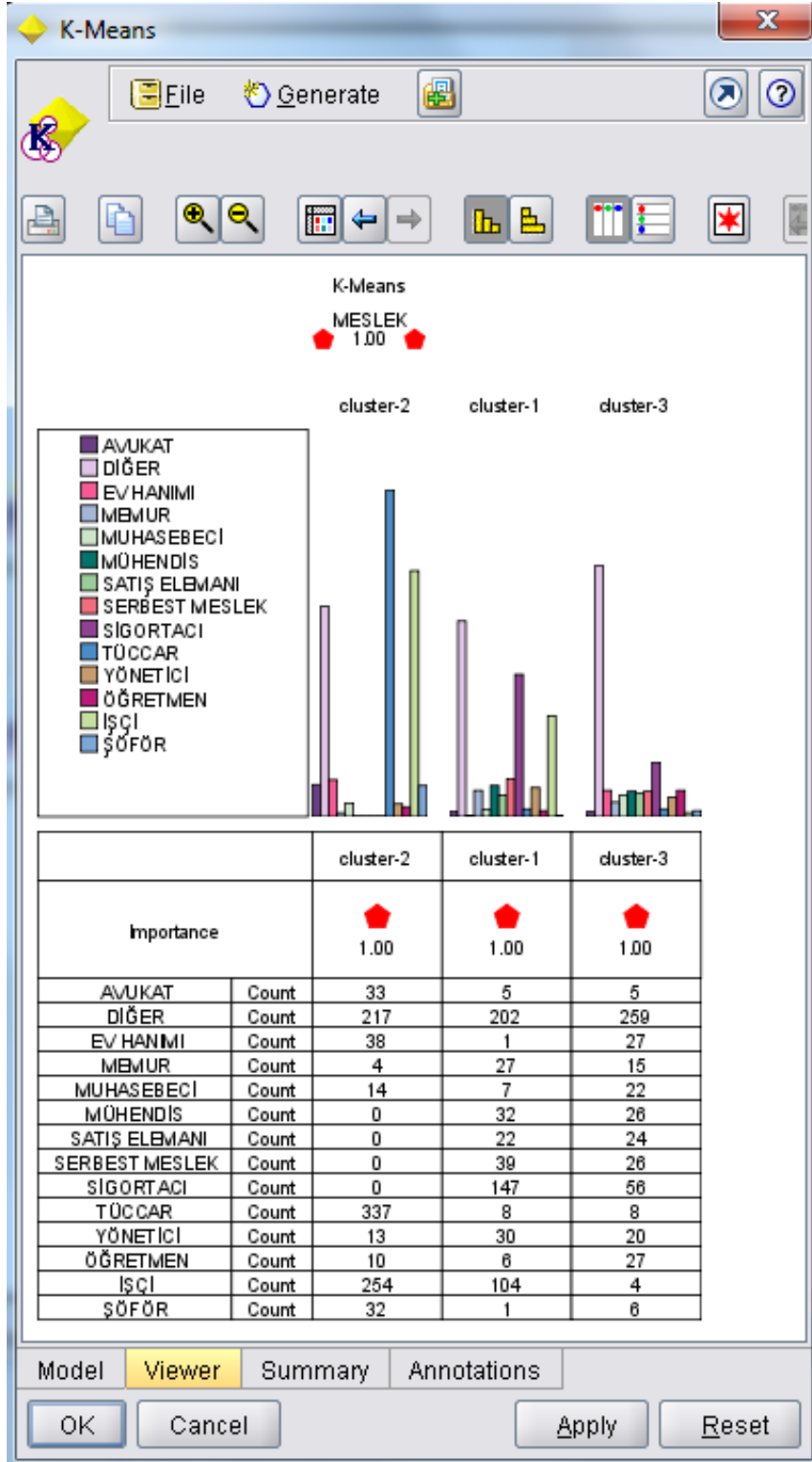
Şekil 3.47. Satın Alınan Teminatların Kümeler Üzerindeki Etkisi

Şekil 3.47’de SATINALINAN_TEMINATLAR değişkeninin kümeler üzerine etkisi incelendiğinde birinci ve ikinci kümede Vefat teminatı, üçüncü kümede Kaza Sonucu Maluliyet ve Kaza Sonucu Vefat Teminatları en çok tercih edilen olduğu görülmektedir.



Şekil 3.48. Yaş Değişkeninin Kümeler Üzerine Etkisi

Şekil 3.48’de YAS değişkeninin tüm kümeler üzerine etkisi incelenecek olursa, birinci ve ikinci küme için yaş değişkeni herhangi bir önem ifade etmemektedir. Üçüncü kümeye bakılacak olursa, 0,92 önem değeri ile diğer kümelere göre daha büyük önem değerinin sahiptir.



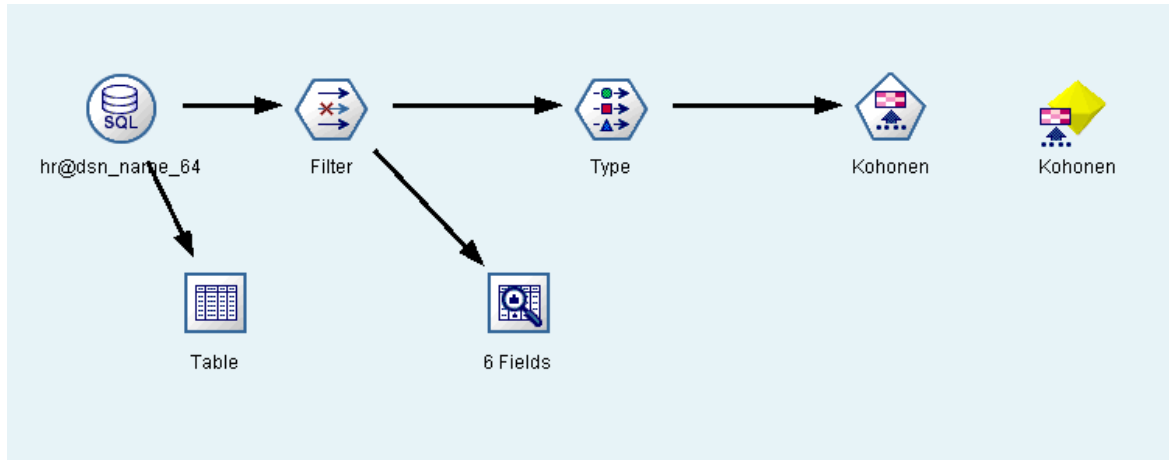
Şekil 3.49. Meslek Değişkeninin Kümeler Üzerindeki Etkisi

Şekil 3.49’da MESLEK değişkenini incelendiğinde birinci küme-1(cluster-1)’de Diğer adı altına toplanan meslek grupları sonucu çıkmıştır. Fakat diğer meslek grupları az

olan, birden fazla meslekte meydana geldiği için değerli gibi gözükse de aslında değersiz bir ifadedir. Küme-1’de yer alan ikinci en büyük meslek grubunda ise, 147 değeri ile Sigortacılık mesleği yer almaktadır. İkinci küme-2(cluster-2)’de ise Tüccarlar çoğunlukta olup, son olarak üçüncü küme(cluster-3)’de Diğer meslek grubu ön plana çıkmıştır. Bu sonuçlara oranla Birinci küme(cluster-1)’de bulunan en çok sayıda meslek grubu Sigortacılık Sektöründe çalışanlardan, ikinci küme(cluster-2)’de ise Tüccar meslek grubunda olanlar anlamlı bir sonuç ifade etmektedir.

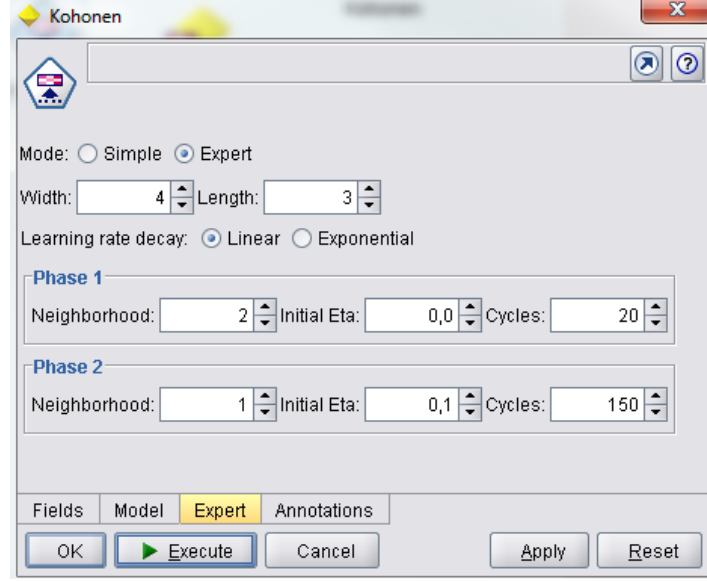
Kohonen Ağları Analizi

Tezin bu bölümünde en önemli ağ yapılarından birisi olan Kohonen Ağları, diğer bir ifadeyle Kendini Düzenleyen Haritalar (Self-Organizing Map-SOM) kümeleme veri setine uygulanmıştır.1980’lerde geliştirilen Kohonen Ağlarının ana fonksiyonu, n-boyutlu uzaydaki girdi verilerini daha düşük boyutlu çizimle haritalaştırmaktadır. Şekil 3.50’de Clementine programında oluşturulan yapı gözükmektedir.



Şekil 3.50. Veri Setine Kohonen Ağlarının Uygulanması

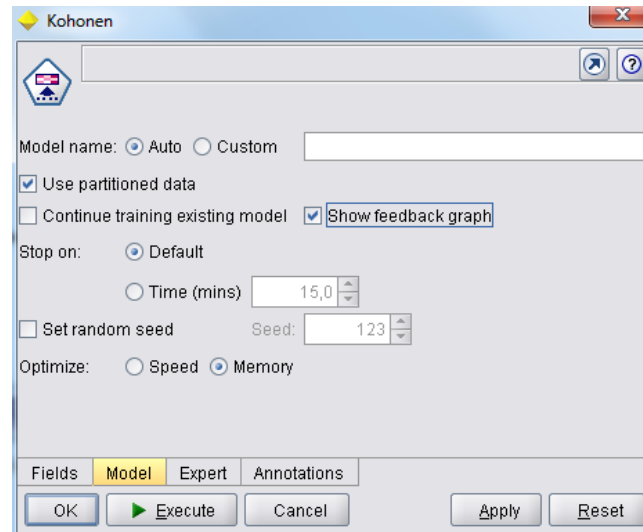
Kohonen ağları analizinde K-ortalamlar analizinde kullanılan veri seti olan Tablo 3.4’deki “kümeleme veri seti” kullanılmıştır.



Şekil 3.51. Kohonen Ağları Boyut Seçimi

Şekil 3.51’de kümeleme veri setine kohonen ağları analizini uygulamadan önce koordinat sisteminde boyutlarını seçebilme olanağı sağlanmaktadır. Burada Genişlik(Width) ve Yükseklik(Length) değerlerine karşılık gelmektedir. Bu değerleri arttırarak küme sayısı arttırılabilir.

Kohonen ağları modeli uygularken Kohonen aracında Şekil 3.52’de görülen Geribesleme grafiği (Show feedback graph) alanı işaretlendiği zaman analizin hazırlanması aşamasında Şekil 3.53’deki gibi Geri besleme grafiği etkin hale getirilmiş olunur. Aksi durumda geri besleme grafiği etkin olmayacaktır.



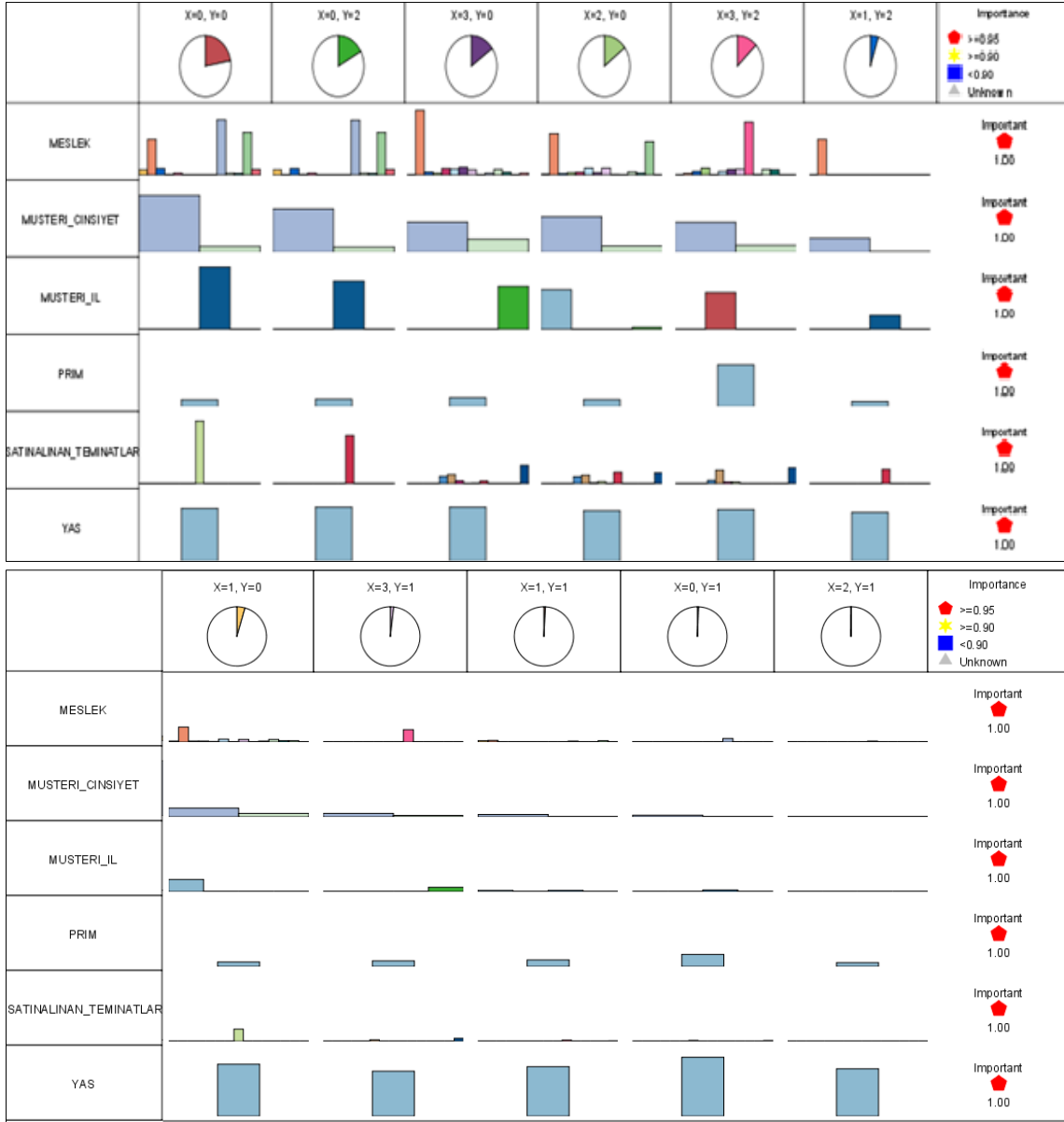
Şekil 3.52. Clementine Kohonen Modeli Aracı

Şekil 3.53'de model oluşturulma aşamasında herhangi bir zamanda alınan geri besleme grafiği görülmektedir. Bu grafiğin ifade ettiği iki boyutlu bir dizinin gelişimi esnasındaki bir görüntü olmakla birlikte, her renk farklı bir direnç ifade etmektedir. Kırmızı birimler kazanan birçok kayıtları, beyaz birimler kazanılan birkaç veya kayıt olamayanları ifade etmektedir. Şekil 3.53'de görülen geri besleme grafiği gelişim esnasında ekran görüntüsü alınmış olup son durumu temsil etmemektedir.



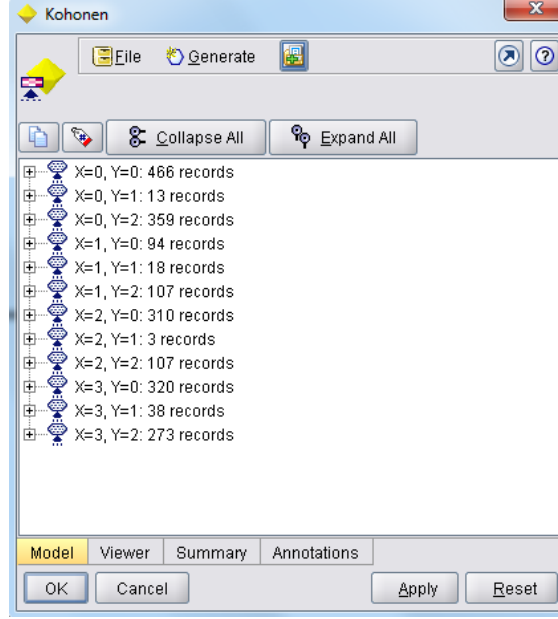
Şekil 3.53. Kohonen Ağları Geri Besleme Grafiği

Şekil 3.54'da görüldüğü gibi $X=3$, $Y=2$ boyutlu analiz yapılmıştır. Boyutları isteğe göre yükseltilebilmektedir. Boyutları arttırdıkça daha çok küme oluşturulacak ve küme başına düşen müşteri sayısı azalacağından dolayı bu boyutlar tercih edilmiştir. Bunun sonucunda 12 tane küme oluşmuştur. Bunlar; $(X=0, Y=0)$, $(X=0, Y=1)$, $(X=0, Y=2)$, $(X=1, Y=0)$, $(X=1, Y=1)$, $(X=1, Y=2)$, $(X=2, Y=0)$, $(X=2, Y=1)$, $(X=2, Y=2)$, $(X=3, Y=0)$, $(X=3, Y=1)$, $(X=3, Y=2)$ 'dir.



Şekil 3.54. Kohonen Ağları ile Kümeleme Analizi Sonuçları

Kümeler üzerinde Şekil 3.54’de görüldüğü gibi Meslek, Müşteri Cinsiyeti, Müşteri ili, Prim, Satın Alınan Teminatlar ve Yaş Değişkeni ayırt edici özelliğe sahiptirler.



Şekil 3.55. Analiz Sonucu Oluşan Kümelerin Eleman Sayıları

Şekil 3.55’de görüleceği üzere Kohonen Analizi sonucunda oluşan kümelerin içerdikleri müşteri sayılarını gözükmemektedir. En büyük küme 466 kayıt ile (X=0,Y=0) kümesidir.

- X=0,Y=0: 466 kayıt(records)
- X=0,Y=1: 13 kayıt(records)
- X=0,Y=2: 359 kayıt(records)
- X=1,Y=0: 94 kayıt(records)
- X=1,Y=1: 18 kayıt(records)
- X=1,Y=2: 107 kayıt(records)
- X=2,Y=0: 310 kayıt(records)
- X=2,Y=1: 3 kayıt(records)
- X=2,Y=2: 107 kayıt(records)
- X=3,Y=0: 320 kayıt(records)
- X=3,Y=1: 38 kayıt(records)
- X=3,Y=2: 273 kayıt(records)

Meslek deęişkeni tüm kümeler için edici özellięe sahiptir. Tablodan da görülebildięi gibi satın alınan poliçelerde $X=0, Y=0$ koordinatlı kümede 164 kiři tüccar, $X=3, Y=2$ koordinatında bulunan kümede 157 kiři sigortacı meslek grubundan oluşmaktadır (bkz: Ek 3-Şekil 1).

$(X=0, Y=0)$, $(X=0, Y=2)$, $(X=3, Y=0)$, $(X=1, Y=2)$, $(X=2, Y=2)$, $(X=1, Y=0)$, $(X=3, Y=1)$ koordinatlı kümelerde, Müşteri Cinsiyeti deęişkeni ayırt edici özellik olup. Kümelerde ortalama olarak erkek müşteriler çoğunluktadır (bkz: Ek 3-Şekil 2).

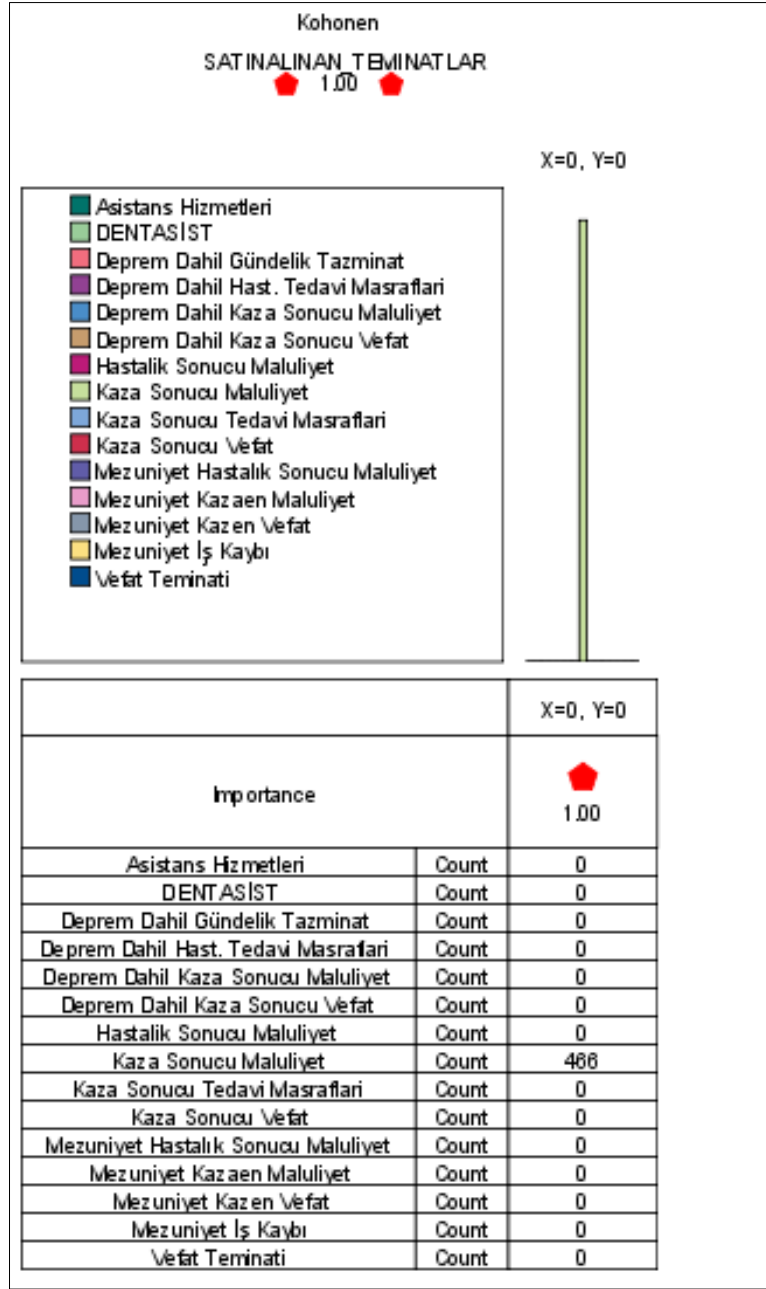
Bütün kümelerde İl deęişkeni ayırt edici deęişken olup, $(X=0, Y=0)$ ve $(X=0, Y=2)$ koordinatlı kümelerde en yüksek müşteri sayısı Trabzon ilçesinden çıkmıştır. $(X=3, Y=0)$ koordinatlı kümede ise İstanbul şehrinde yaşayan müşterilerden oluşmaktadır (bkz: Ek 3-Şekil 3).

Kazanılan prim deęişkeni göz önüne alındığı zaman $(X=3, Y=2)$ koordinatlı küme ve $(X=2, Y=2)$ koordinatlı kümelerde ayırt edici özellięe sahiptir. $(X=3, Y=2)$ koordinatlı kümede ortalama prim 2472,43 TL ve yedinci kümede $(X=2, Y=2)$ koordinatlı kümelerde ortalama tutar 3388,29 TL şeklinde gözlenmektedir (bkz: Ek 3-Şekil 4).

Satın alınan teminatlar bazında birkaç küme incelenecek olursa, $(X=0, Y=0)$ koordinatlı kümede “Kaza Sonucu Maluliyet”, $(X=0, Y=2)$ ve $(X=1, Y=2)$ koordinatlı kümelerde “Kaza Sonucu Vefat Teminatı” tercih edilen çoğunluklu müşterilerden oluşmaktadır. $(X=3, Y=0)$ koordinatlı kümede ise “Vefat Teminatı” tercih edenler çoğunluktadır (bkz: Ek 3-Şekil 5).

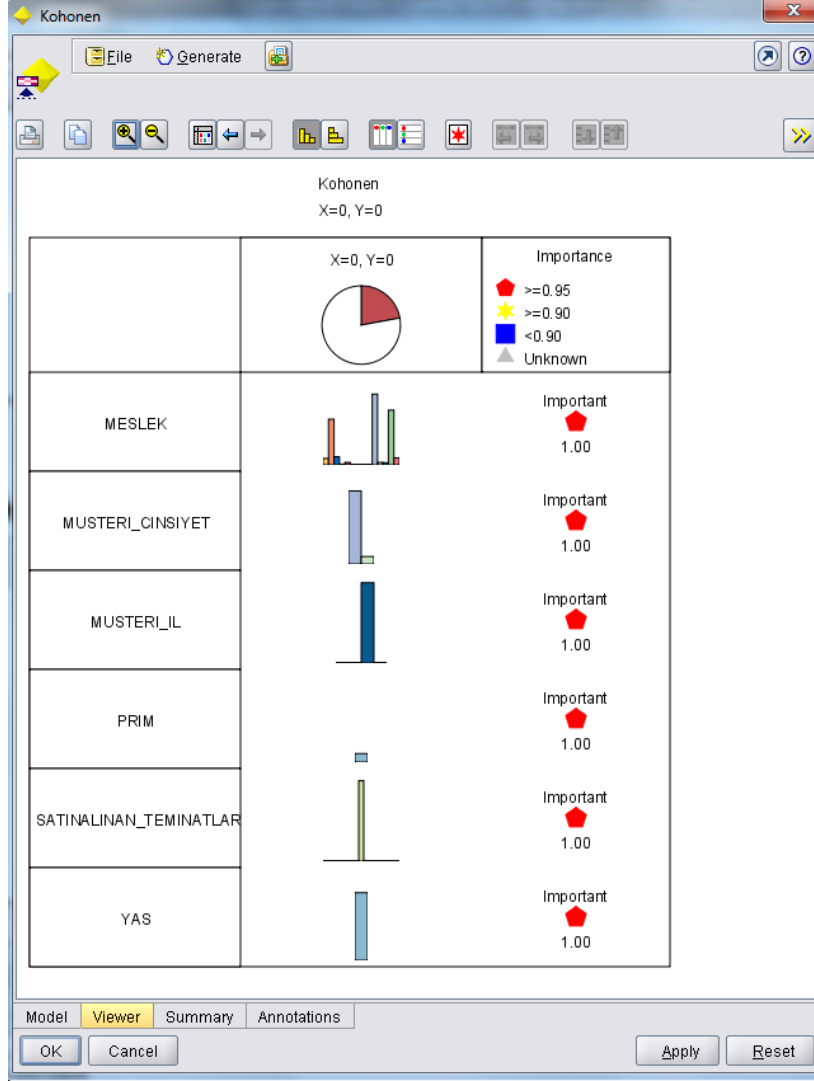
Yaş deęişkeni göz önüne alındığında $(X=0, Y=2)$, $(X=3, Y=0)$, $(X=0, Y=1)$ koordinatlı kümelerde yaş deęişkeni ayırt edici önemli bir deęişken olup bu kümelerden den yaş deęişkeni ayırt edici deęişkendir. $(X=0, Y=2)$, $(X=3, Y=0)$ koordinatlı kümelerde ortalama olarak 40 yaş grubu müşterilerin tercihlerinden oluşmaktadır. $(X=0, Y=1)$ koordinatlı kümede ortalama yaş grubu 45 olan müşteriler bulunmaktadır (bkz: Ek 3-Şekil 6).

Şekil 3.56’da görüldüğü gibi Teminat bazında hedef deęişken olarak incelediğinde $(X=0, Y=0)$ koordinatlı kümede “Kaza Sonucu Maluliyet” Teminatı en yüksek teminat müşteri sayısına sahip teminat arasındadır.



Şekil 3.56. (X=0,Y=0) Koordinatlı Kümeye Satın Alınan Teminatın Etkisi

Şekil 3.57’de (X=0,Y=0) koordinatlı küme üzerine değişkenlerin etkisi görülmektedir. Oranlar itibariyle en önemli küme olan X=0,Y=0 koordinatlarındaki küme incelendiğinde, Meslek grubu Tüccar, sıralamanın en üstündedir (bkz:Ek 3-Şekil 1). Müşteri Cinsiyeti erkekler çoğunluk olmakla birlikte, Müşteri Şehri Trabzon, Prim ortalaması 397 TL, Yaş grubu ortalama olarak 39 yaşındakilerden oluşmaktadır. Diğer koordinatların analiz raporları Ek 3-Şekil 7-13 arasında yer almaktadır.



Şekil 3.57. (X=0,Y=0) Koordinatlı Küme Analizi

3.1.5. Değerlendirme (Evaluation)

Değerlendirme aşamasında elde edilen modelin konuşlandırılması öncesinde model oluşum sürecinin dikkatlice gözden geçirildiği ve modelin iş hedeflerini başarma konusundaki yeterliliğinin kararlaştırıldığı ve veri madenciliği sonuçlarının nasıl kullanılacağına karar verileceği aşamadır.

3.1.6. Konuşlandırma (Deployment)

Bu aşama analizciden çok organizasyonun kendisi tarafından anlaşılması gereken bir aşamadır. Geçerliliği sağlanan model iki farklı şekilde kullanılabilir. Birinci yol analizci tarafından modelin ve sonuçlarının sunulmasıdır. İkinci yol ise modelin yeni veri kümeleri üzerinde kullanılmasıdır.

Model sonuçlarına göre örnek verecek olursak;

- Birliktelik kuralı analizinden çıkarılan sonuçlardan birisi “Kaza Trafik Sigortası satın alan müşterilerin çoğunluğu Kaza Kaskosu da satın alacaktır.” şeklindedir. Bu bilgiden yola çıkarak satış kampanyaları düzenlenebilir. Örneğin, Kaza Trafik Sigortası satın almış fakat Kaza Kaskosu satın almamış müşterilere indirim paketi öngörülebilir.
- Kümeleme Analizlerinden olan K-Ortalamlar Analizi kullanılarak elde edilen birinci küme sonuçlarını inceleyecek olursak; müşterilerin cinsiyeti Erkekler olan, Ankara’da yaşayanların çoğunlukta olduğu ayrıca en çok satın alınan teminatları Vefat Teminatı ve Deprem Dahili Kaza Sonucu Maluliyet teminatlarının tercih edildiği görülmektedir. Bu sonuçlara göre örneğin, Ankara şehrindeki acenteliğinde daha çok müşteri çekmek amacıyla en çok tercih edilen ürünlerde kampanyalar düzenlenerek daha çok müşteriye ulaşılması ve satışların artırılması sağlanabilir.

Örneklerden de görüldüğü gibi daha çok müşteri çekmek amacıyla elde edilen bilgilerden yararlanarak çeşitli satış kampanyaları düzenlenebilir. Eldeki müşterilerin devamlılığını sağlamanın yanı sıra, daha çok müşteriye ulaşılması sağlanmış olur.

4. SONUÇLAR

Bu çalışmada Sigorta Sektöründe MİY’i gerçekleştirme için VM analizlerinden olan Kümeleme ve Birliktelik kuralı analizlerinden yararlanılmıştır. Birliktelik Analizlerinden olan Apriori Algoritması kullanılarak elde edilen sonuçlar şu şekildedir:

- Kaza Trafik Sigortası satın alan müşterilerin %67,58’i Kaza Kaskosu da satın almaktadır. Bu müşteriler toplam müşterilerin yaklaşık %27,27’lik bir kısmını oluşturmaktadır.
- Dask Sigortası satın alanların %54,96’sı Yangın Sigortası da satın almaktadır. Bu müşteriler toplam müşterilerin yaklaşık %16,28’ini oluşturmaktadır.
- Yangın Sigortası satın alan müşterilerin %47,38’i Dask Sigortası da satın almıştır. Bu müşteriler, toplam müşterilerin %18,88’ini oluşturmaktadır.

Kümeleme Analizlerinden olan K-Ortalamalar Analizi kullanılarak üç küme elde edilmiş olup sonuçlar şu şekildedir:

- Birinci Kümede ki önemli değişkenler incelendiğinde, müşterilerin cinsiyeti Erkek, Ankara’da yaşayanların çoğunlukta olduğu ayrıca satın alınan teminat değişkeni incelendiğinde Vefat Teminatı ve Deprem Dahili Kaza Sonucu Maluliyet teminatlarının tercih edildiği görülmektedir.
- İkinci Kümedeki önemli değişkenler incelendiğinde, Cinsiyeti Erkek olanlar kümedeki müşterilerin %90’nını oluşturduğu, Trabzon’da yaşayanlardan oluştuğu, Meslek değişkeni çoğunlukla Tüccar ve İşçi’lerden oluştuğu gözlenmektedir. Ayrıca Teminat bazında incelenecek olunursa en çok Kaza Sonucu Maluliyet ile Kaza Sonucu Vefat Teminatı tercih edilmektedir.
- Üçüncü Kümedeki önemli değişkenler incelendiğinde, Cinsiyet değişkeni incelendiğinde Erkek ve Kadın sayılarının birbirlerine çok yakın olduğu ve 34 yaşındakilerin çoğunlukta olduğu görülmektedir. Ayrıca Müşterinin İl’i incelendiğinde İstanbul’da yaşayanların çoğunlukta olduğu ve Satın Alınan Teminat bazında Vefat Teminatı en çok tercih edilen teminatıdır.

Değişkenlerin kendi içinde kümeler üzerine etkisi incelenecek olunursa;

- Müşteri Cinsiyet değişkeninin Kümeler Üzerine Etkisi:

Elde edilen üç küme içinde müşteri cinsiyeti önemli değişkendir. Her üç kümede de Erkek sayısı çoğunlukta olmakla birlikte, birinci kümenin tamamı erkeklerden oluşmaktadır. Üçüncü kümede ise Erkekler ile Kadınların sayıları birbirine yakındır. Cinsiyet göz önüne alınarak bir satış kampanyası yapılacağı zaman ikinci kümedeki müşterilerin demografik özelliklerinden faydalanılabilir.

- Müşteri İl değişkeninin Kümeler Üzerine Etkisi:

Bu değişken her üç küme içinde önemli bir değişkendir. Birinci kümenin çoğunluğu Ankara İlinde yaşayanlardan, İkinci kümenin tamamı Trabzon ilinde yaşayanlardan, Üçüncü kümenin çoğunluğu ise İstanbul'da yaşayanlardan oluşmaktadır. Bu sonuçlara istinaden şehir bazında kampanya düzenlenmek istendiğinde gruplarındaki müşterilerin özellikleri göz önüne alınarak daha çok müşteri çekilmesi sağlanabilmektedir.

- Satın Alınan Teminatların Kümeler Üzerine Etkisi:

Birinci kümede en çok tercih edilen teminatların başında Vefat Teminatı ve Deprem Dahil Kaza Sonucu Vefat Teminatı gelmektedir. İkinci kümede çoğunluk Kaza Sonucu Maluliyet ve Kaza Sonucu Vefat Teminatlarını satın alınan müşterilerden oluşmaktadır. Üçüncü Kümede ise Vefat Teminatı en çok tercih edilen üründür.

- Yaş Değişkeninin Kümeler Üzerine Etkisi:

Yaş değişkeni birinci ve ikinci küme için az oranda önemli bir değişken iken üçüncü kümede ise daha az oranda önemli bir değişkendir. Birinci küme yaklaşık 38 yaşındakilerden, ikinci küme ise yaklaşık 39 yaşındakilerden oluşmaktadır.

- Meslek deęişkeninin Kümeler Üzerine Etkisi:

Birinci ve İkinci küme için Meslek deęişkeni önemli bir deęişkendir. Birinci kümede Sigortacılar çoęunluğu oluştururken, İkinci kümede ise Tüccar ve İşçilerden oluşmaktadır.

Kümeleme Analizlerinden olan Kohonen Ağları Analizi kullanılarak elde edilmiş olan oranlar itibariyle en yüksek beş koordinat incelenecek olunursa sonuçlar şu şekildedir:

- (X=0,Y=0) koordinatlı küme sonuçlarına göre, Meslek, İl, Cinsiyet, Satın Alınan Teminat ve Yaş deęişkeleri önem arz etmekle birlikte, Meslek grubu Tüccarlar çoęunluktur. İl deęişkeni incelendiğinde Trabzon'da yaşayanlardan oluşmaktadır. Cinsiyet deęişkenine bakılacak olunursa Erkek cinsiyetine sahip olanlar çoęunluktur. En çok tercih edilmiş olan teminat Kaza Sonucu Maluliyet Teminatıdır. Son olarak Yaş deęişkeni incelenecek olunursa 39 yaşındakiler çoęunluktur.
- (X=0,Y=2) koordinatlı küme sonuçlarına göre, Meslek, İl, Cinsiyet, Satın Alınan Teminat ve Yaş deęişkeleri önemli deęişkenlerdir. Deęişkenler incelendiğinde Tüccarların çoęunlukta olduğu, Trabzon'da yaşayanlardan oluştuęu, cinsiyeti Erkek olanların ve ortalama 40 yaşına sahip bireylerin çoęunlukta olduğu görülmektedir. Satın Alınan Teminatlar incelendiğinde en çok Kaza Sonucu Vefat Teminatı tercih edilmektedir.
- (X=3,Y=0) koordinatlı küme sonuçlarına göre, İl, Cinsiyet, Satın Alınan Teminat ve Yaş deęişkeleri önemli deęişkenlerdir. Bu deęişkenler incelendiğinde; İstanbul'da yaşayanlar çoęunlukta olmakla birlikte, cinsiyeti Erkek olanlar çoęunluktur. Ortalama yaş grubu 40 yaşındakilerden oluşmaktadır ve Vefat Teminatı en çok tercih edilen teminat türüdür.
- (X=2,Y=0) koordinatlı küme sonuçlarına göre, Meslek, İl, Cinsiyet, Satın Alınan Teminat ve Yaş deęişkeleri önemli deęişkenlerdir. Bu deęişkenler incelendiğinde Ankara ilinde yaşayanların çoęunlukta olduğu, Meslek deęişkenine göre İşçilerin çoęunlukta olduğu, 37 yaş ortalamasında, çoęunluğun Erkek müşterilerin oluştuęu gözlenmektedir. Ayrıca bu kümede Kaza Sonucu Vefat ve Vefat Teminatı en çok tercih edilen iki teminat türüdür.

- (X=3,Y=2) koordinatlı küme sonuçlarına göre, , Meslek, Cinsiyet, Satın Alınan Teminat ve Yaş değişkeleri önemli değişkenlerdir. Sigortacı mesleğine sahipler çoğunlukta olmakla birlikte, Cinsiyet değişkenine bakıldığında Erkek müşteriler çoğunluktadır. Satın Alınan Teminat bazında incelendiğinde, Deprem Dahil Kaza Sonucu Vefat ve Vefat Teminatını Tercih edenler çoğunluktadır. Kümenin Yaş değişkeni incelendiğinde yaklaşık olarak 38 yaşındakilerden oluşmaktadır.

Elde edilen sonuçlara göre anlaşılmaktadır ki, Sigortacılık Sektöründe Veri Madenciliği analizlerinden birçoğu veri yapısına göre incelenerek istenilen bilgiye uygun olarak analizlere uygulanabilmektedir. Bunun yanı sıra şirketlerin veritabanlarında tutulan verilerden yola çıkılmış ve bilgi keşfi yapılmış olup, veritabanında gizli kalmış verilerden bilgi açığa çıkartılmaktadır. Bunun sonucunda değerli ve değersiz müşterilerin farkına varılmaktadır.

Bunların sonucunda bir sigorta şirketinde, elde tutulan müşterilerin devamlılığının sağlanması çalışılırken, bunun yanı sıra değerli müşterileri özelliklerini keşfederek daha hangi türdeki bireylere satış kampanyalarının düzenlenmesi gerektiği hakkında bilgi sahibi olunmuş olur. Hangi müşterilerin ne tür poliçe tercih ettiği sonucundan yola çıkılarak, o tür müşterilere odaklanılarak daha çok müşteri elde edilebilir. Ürünlerin bir arada alınma sıklıklarından faydalanılarak, bir arada alınan ürünlere kampanyalar düzenlenerek müşterilerin daha çok dikkatinin çekilmesi sağlanabilir.

KAYNAKLAR

Acuner, Ş. (2005) Günümüzde pazarlama anlayışında MİY Uygulamalarının Yeri ve Önemi, Pazarlama Dünyası, Mayıs-Haziran, Sayı: 3, 52.

Agrawal, R., Imielinski, T., Swami, A. (1993) Mining Association Rules Between Sets of Items in Large Databases, In Proceedings of the ACM SIGMOD International Conference on Management of Data, Washington, USA., 207-216.

Agrawal, R., Srikant, R. (1994) Fast Algorithms for Mining Association Rules, Proc. of the 20th Int'l Conference on Very Large Databases, Santiago.

Akça, U. (2007) MİY Bileşenleri, <http://www.erpMIY.com/>, (01.01.2007).

Akpınar, H. (2000) Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği, İstanbul Üniversitesi İşletme Fakültesi Dergisi,29(1).

Aksoy, N. (2002) MİY El Kitapçığı, Dohaş Müsteri Hizmetleri, İstanbul, 32.

Alpaydın, E. (2000) Zeki Veri Madenciliği: Ham veriden altın veriye ulaşma yöntemleri, Bilişim Eğitim Semineri,1-10.

Altay, T. (2005) Knowledge discovery in databases and data mining techniques: an applied study, Thesis, 25.

Anand, A. (2003) A study and comparison of data clustering techniques, Master Thesis, Faculty of the Graduate School of The University of Texas El Paso, Texas, 21-22.

Arda, M.O. (2003) Veri Madenciliği Aracı Olarak Genetik Algoritmalar ile Yapay Sinir Ağları ve Genetik Algoritmalar Genetik Algoritma-Yapay Sinir Ağı Melez Modelinin Müşteri Değerlendirilmesinde Kullanılması, Doktora Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, Sayısal Yöntemler, 43-46.

Aydın, T. (2006) Müşteri ilişkileri Yönetimi hakkında MİY Bilgilendirme Semineri, United Nations Development Programme, Kıbrıs, http://65.110.73.19/UploadsNew/Gallery/Seminars/ICT3/III_MİY_Awareness-Handouts.pdf

- Babu, G., Bhuvanewari, T.A. (2012) Data Mining Technique to Find Optimal Customers for Beneficial Customer Relationship Management, Journal of Computer Science, 8(1), ISSN 1549-3636,89-98.
- Baao, F., Lobo, V. (2013) Introduction to Kohonen's Self Organizing Maps, <http://edugi.uji.es/Bacao/SOM%20Tutorial.pdf>, (27.01.2013).
- Baao, F., Lobo, V., Painho, M. (2005) The Self-organizing map, the Geo-SOM, and relevant variants for geosciences, Computer & Geosciences, Elsevier, 155-163.
- Baykal, A. (2006) Veri madencilięi uygulama alanları Application Fields of Data Mining, Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi, 7.
- Bayram, E. (2001) Customer segmentation and churn modeling in wireless communication, Yüksek Lisans Tezi, Bogaziçi Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 1-5.
- Bergeron, B. (2002) Essentials of MIY, Willey, USA, 1-6.
- Berry, M.J.A., Linoff, G.S. (1999) Mastering Data Mining: The Art and Science of Customer Relationship Management, John Wiley & Sons, 1st Edition
- Bilgin, T.T. (2003) Veri Madencilięinde Kavramı Ve Analiz Yöntemi Uygulamaları, Marmara Üniversitesi, Yüksek Lisans Tezi, İstanbul
- Boginski, V., Butenko, S., Pardalos, P. (2006) Mining market data: A network approach, Computers & Operations Research, 33(11), 3171-3184.
- Caskey, K. (2001) A manufacturing problem solving environment combining evaluation, search, and generalisation methods, Computers in Industry, 44(2), 175-187.
- Chen, M., Huang, C., Chen, K., Wu, H. (2005) Aggregation of orders in distribution centers using data mining, Expert Systems with Applications, 28(3), 453-460.
- Chen, M.S., Han, J., Yu, P.S. (1996) Data Mining: An Overview from a Database Perspective, IEEE Transactions on Knowledge and Data Engineering, 8(6), 866-883.
- Chu, W. (2005) Foundations and Advances in Data Mining, Springer Publishers, 25, 100.

Cox, I., Lewis, R., Ransing, R., Laszczewski, H., Berni, G. (2002) Application of neural computing in basic oxygen steelmaking, Journal of Materials Processing Technology, 120(1), 310-315.

CRISP-DM; Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Rüdiger, W. (2000) SPSS CRISP-DM 1.0, Spss Inc., U.S.A., 10.

Çiçek, G. (2007) Kısaca MİY Eğitimi,

http://www.gsuik.org/index.php?option=com_content&task=view&id=509&Itemid=119, (11.10.2010)

Dener, M., Dörtler, M., Orman, A. (2009) Akademik Bilişim 09 XI. Akademik Bilişim Konferansı Bildirileri, Açık Kaynak Kodlu Veri Madenciliği Programları WEKA'da örnek uygulama, Harran Üniversitesi, 11-13 Şubat, Şanlıurfa.

Dereli, F.Z. (2004) CRM Nedir?,

<http://crmlab.clubs.bilkent.edu.tr/crm/crmnediryazi.html>, (28.01.2013).

Dillon, W.R., Matthew, G. (1984) Multivariate Analysis, John Wiley & Sons, USA, 167.

Doumpos, M., Zopounidis, C. (2002) Multi-Criteria Decision Aid Classification Methods, Holland, Academic Publishers, 28.

Döşlü, A. (2008) Veri Madenciliğinde Market Sepet Analizi ve Birliktelik Kurallarının Belirlenmesi, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, İstanbul.

Duran, M. (2006) MİY: Çok Konuşulan Ama Az Bilinen Bir Kavram.

Dyche, J. (2012) The MİY Handbook, Addison-Wesley.

Ersoy, F. N. (2002) Finansal Hizmetlerde MİY Kavramı, Pazarlama Dünyası Dergisi, Mart-Nisan, İstanbul, (15), 5.

Ezerce, A. (2008) Müşteri ilişkileri Yönetimi ve Veri Madenciliği: Tekstil Sektörü üzerine bir uygulama, Yüksek Lisans Tezi, İstanbul

Flexer, A. (2001) On the Use of Self-Organizing Maps for Clustering and Visualization, Intelligent Data Analysis 5, 382.

Fürnkranz, J., Gamberger, D., Lavrac, N. (2012) Foundations of Rule Learning, Chapter 1 Machine Learning and Data Mining, Springer, XVII, 8.

- Gates, B. (1995) The Road ahead, 1 st Edition, Penguin group, USA, 12.
- Gel, O.C. (2003) MİY Yolculuğu, Sistem Yayıncılık, 2. Baskı, İstanbul, 12.
- Giudici, P. (2000) Applied Data Mining Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann Publishers, San Fransisco, 120-129, 267-277.
- Güldür, G. (2007) Bilgi Ekonomisinde Müşteri Yönetimi, MİY Institute Turkey, <http://www.MIYinturkey.org>, (01.01.2007).
- Han, J., Kamber, M. (2000) Simon Fraser University, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, Second Edition
- Han, J., Kamber, M. (2001) Data Mining Concepts and Techniques, Morgan Kaufmann Publisher, San Francisco, USA.
- Han, J., Kamber, M. (2006) Morgan Kaufmann Publisher, Data Mining Concepts and Techniques, 2 st Edition, San Francisco, USA.
- Hsieh, N. (2004) An integrated datamining and behavioral scoring model for analyzing bank customers, Expert Systems with Applications, 27(4), 623-633.
- Hui, S., Jha, G. (2000) Application data mining for customer service support, Information and Management , 38,1-13.
- Hung, S., Yen, D., Wang, H. (2005) Applying data mining to telecom churn management, Expert Systems with Applications, In Press
- Hwang, H., Jung, T., Suh, E. (2004) An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry, Expert System with Applications, 26, 181-188.
- Joseph, P.B. (1996) Data Mining With Neural Networks: Solving Business Problems from Application Development to Decision Support, McGrawHill
- Kaas, G.V. (1980) An exploratory technique for investigating large quantities of categorical data, Journal of Applied Statistics, 29(2), 119-127.
- Kallkota, R., Robinson, M. (1999) Roadmap for Success, Addison Wesley, 41.

- Kantardzic, M. (2003) Data Mining Concepts Models, Methods and Algorithms, John Wiley & Sons
- Karypis, G., Han, E.H., Kumar, V. (1999) CHAMELEON :A hierarchical clustering algorithm using dynamic modeling ,IEEE Computer
- Kasap, E. (2007) Sigortacılık Sektöründe Müşteri ilişkileri Yaklaşımıyla Veri Madenciliği Teknikleri ve Bir Uygulama, Marmara Üniversitesi Bankacılık ve Sigortacılık Enstitüsü Sigortacılık Bölümü, Yüksek Lisans Tezi, İstanbul.
- Khajvand, M., Tarokh, M., J. (2011) Analyzing Customer Segmentation Based on Customer Value Components (Case Study: A Private Bank), Journal of Industrial Engineering, University of Tehran, Special Issue , 79-93.
- Kırım, A. (2001) Strateji ve Birebir Pazarlama MİY, Sistem Yayıncılık, İstanbul
- Kiang, M.Y. (2001) Extending the Kohonen Self-organizing Map Networks for Clustering Analysis, Computational Statistics&Data Analysis, 163.
- Kim, E., Lee, Y. (2002) Combination of multiple classifiers for the customer's purchase behavior prediction, Decision Support Systems, 34 (2), 167-175.
- Kocabaş, K. (2012) Veri Madenciliği İş Zekası, <http://www.misjournal.com/?p=724> (30.12.2012)
- Kohonen, T. (1982) Self-Organized Formation of Topologically Correct Feature Maps, Biological Cybernetics 43(1), 59–69.
- Kohonen, T. (1988) The Neural Phonetic Typewriter, Volume 21, Issue 3, IEEE Computer
- Kohonen, T., Kangas, J. (1996) Developments and Applications of the Self-Organizing Map and Related Algorithms, Helsinki University of Technology Neural Networks Research Centre, Accepted for Mathematics and Computers in Simulation. Predicted publication: MACTOM, 41(5-6) July, Finland.
- Kotler, P. (2003) Marketing Management, Pearson Education, 52.
- Kumar, V. (2000) University of Minnesota, An Introduction to Cluster Analysis for Data Mining, http://www-users.cs.umn.edu/~han/dmclass/cluster_survey_10_02_00.pdf, (18.12.2012)

- Larose, D.T. (2005) *Discovering Knowledge in Data An Introduction to Data Mining*, John and Willey Sons Incorporated, USA, 8-10.
- Lee, T., Chiu, C., Chou, Y., Lu, C. (2006) Mining the customer credit using classification and regression tree and multivariate adaptive regression splines, *Computational Statistics & Data Analysis*, 50(4), 1113-1130.
- Levin N., Zahavi, J. (1999) *Predictive modeling using segmentation*, John Wiley & Sons, Inc. and Direct Marketing Educational Foundation
- Lin, F., McClean, S. (2001) A data mining approach to the prediction of corporate failure, *Knowledge-Based Systems*, 14(3), 189-195.
- Liu, D., Shih, Y. (2005) Integrating AHP and data mining for product recommendation based on customer lifetime value, *Information & Management*, 42(3): 387-400.
- Loh, W.Y., Shih, Y.S. (1997) Split selection methods for classification trees, *Statistica Sinica*, Vol 7, 815-840.
- MacQueen, J. (1967) Some methods for classification and analysis of multivariate observations. *Fifth Berkeley Sympos. Math. Statist. and Probability Vol. I: Statistics*, Univ. California Press, 281–297, Berkeley, California.
- Mamaklı, S. (2005) MİY'in Ortaya Çıkışı, <http://www.msakademik.net/makaleler.aspx?grup=MBS>, (08.09.2005).
- Mamaklı, S. (2006) MİY ve Bire-Bir Pazarlama-2, <http://www.msakademik.net/makaleler.aspx?grup=MBS>, (07.11.2006).
- McLachlan, G. J. (1999) Mahalanobis Distance, *Resonance* June 1999, <http://www.ias.ac.in/resonance/June1999/pdf/June1999p20-26>
- Merigo, J. M., Casanovas, M. (2008) The Induced Minkowski Ordered Weighted Averaging Distance Operator, *XIV Congreso Español sobre Tecnologías y Lógica fuzzy*, Mieres – Langreo.
- Michael, J.B., Gordon, S.L. (2000) *Mastering Data Mining The Art Science of CM*, Wiley, USA, 14.
- Mitra, S., Acharya, T. (2003) *Data Mining :Multiple Soft Computing and Bioinformatics* John Wiley & Sons Publisher

- Ngai, E.W.T., Xiu, L., Chau, D.C.K. (2009) Application of data mining techniques in customer relationship management: A literature review an classification, *Expert Systems with Applications* 36, 2592–2602.
- Odabaş, Y. (2001) Satışta ve Pazarlamada Müşteri İlişkileri Yönetimi, Sistem Yayıncılık, Sayı 1, Ağustos, İstanbul.
- Odabaş, Y. (2005) Satışta ve Pazarlamada Müşteri ilişkileri Yönetimi, Sistem Yayınları, 3.
- Oğuz, A.Ç. (2004) Müşteri ilişkileri Yönetiminde Kritik Başarı Faktörleri, Yüksek Lisans Tezi, Muğla.
- Online ERP, (2013) CRM-Müşteri İlişkileri Yönetimi, <http://erp.karmabilgi.net/musteriiliskileri-yonetimi/>, (24.01.2013).
- Ögüdücü, S. (2001) ,Veri Madenciliği, Demetleme, Yöntemleri, <http://www.cs.itu.edu.tr/~gunduz/courses/verimaden/slides/d5.pdf>, (23 Haziran 2011).
- Özekes, S. (2003) Veri Madenciliği Modelleri ve Uygulama Alanları, Ticaret Üniversitesi Dergisi, No: 3, İstanbul.
- Özgener, Ş., İraz, R. (2006) Customer Relationship Management in Small-Medium Enterprises: The case of Turkish Tourism İndustry, *Tourism Management*, Vol 27.
- Özkan, Y. (2008) Veri Madenciliği Yöntemleri, Papatya Yayıncılık,1.Basım, Mayıs, İstanbul.
- Payne, A. (2005) *Handbook of MIY: Achieving Excellence in Customer Management*, Butterworth-Heinemann Publidhers, 10.
- Peppers, D., Rogers M. (1999) *Enterprise One to One*, Doubleday, Newyork, 67.
- Peyman, A., Safanaz, H. (2012) Application of Knowledge management in Customer relationship management: a data mining approach, *Iran University of Science and Technology*, Available at SSRN: <http://ssrn.com/abstract=2188230>, (07.02.2012)
- Quinlan, J.R. (1986) *Induction of Decision Trees*, Kluwer Academic Publishers, *Machine Learning* 1, Boston - Manufactured in The Netherlands, 81-106.
- Quinlan, J.R. (1993) *Programs for Machine Learning*, Morgan Kaufman Publishers

- Rodpysh, K.V., Aghai, A., Majdi, M. (2012) Applying Data Mining in Customer Relationship Management, International Journal of Information Technology, Control and Automation(IJITCA), Vol2, No.3, July.
- San, O., Huynh,V., Nakamori, Y. (2004) An alternative extention of the k means algorithm for clustering categorical data , Applied Math. And Computer Science, 14(2), 241-247.
- Sharma, S. (1996) Applied Multivariate Techniques, John Wiley & Sons, USA, 185.
- Shih, F.Y., Wu, Y. (2004) Fast Euclidean distance transformation in two scans using a 3x3 neighborhood, Computer Vision and Image Understanding 93, Elsevier, 195–205.
- Silahtaroglu, G. (2008) Kavram ve Algoritmalarıyla Temel Veri Madenciliği, Papatya Yayıncılık, 1.Basım, Mayıs, İstanbul.
- Sözügüzel, B. (2004) Türk Hava Yollarında MİY Uygulamaları, Yayınlanmamış Yüksek Lisans Tezi, Marmara Üniversitesi, 18.
- SPSS Clementine (2007) Clementine 12.0 Algorithms Guide, USA, 67-71.
- SPSS Clementine (2011) Statistics and Algorithms, Answer Tree 3.1 User's Guide, 204-205.
- Sykes, A.O. (2013) An Introduction to Regression Analysis, http://www.law.uchicago.edu/files/files/20.Sykes_.Regression.pdf , (24.01.2013).
- Şeker, Ş.E. (2012) <http://www.bilgisayarkavramlari.com/2011/09/07/apriori-algoritmasi>, (18.12.2012).
- Şimşek, U.T. (2006) Veri Madenciliği ve Müşteri ilişkileri Yönetiminde Bir Uygulama, Doktora Tezi, Sayısal Yöntemler, İstanbul Üniversitesi, 108.
- Taşpınar, H. (2006) Bilişim Alt Yapısıyla MİY Teknik Alt Yapısı & İşlevlikleri, Seçkin Yayınları, 2. Basım, Ankara.
- Tezcanlar, P. (2007) Müşteri İlişkileri Yönetimi, Veri Madenciliği ve Bir Uygulama, Yüksek Lisans Tezi, Üniversitesi Sosyal Bilimler Enstitüsü, Sayısal Yöntemler Anabilim Dalı, İstanbul.
- Tsiptsis, K., Chorianopoulos, A. (2009) Data Mining Techniques in CRM: Inside Customer Segmentation, Willey, 1st Edition,14.

Uçkan, C.G., Latif, H., Çeltek, E. (2003) Turistik Ürün Oluşturma sürecinde Yeni bir yaklaşım MİY, Pazarlama Dünyası, Sayı:4.

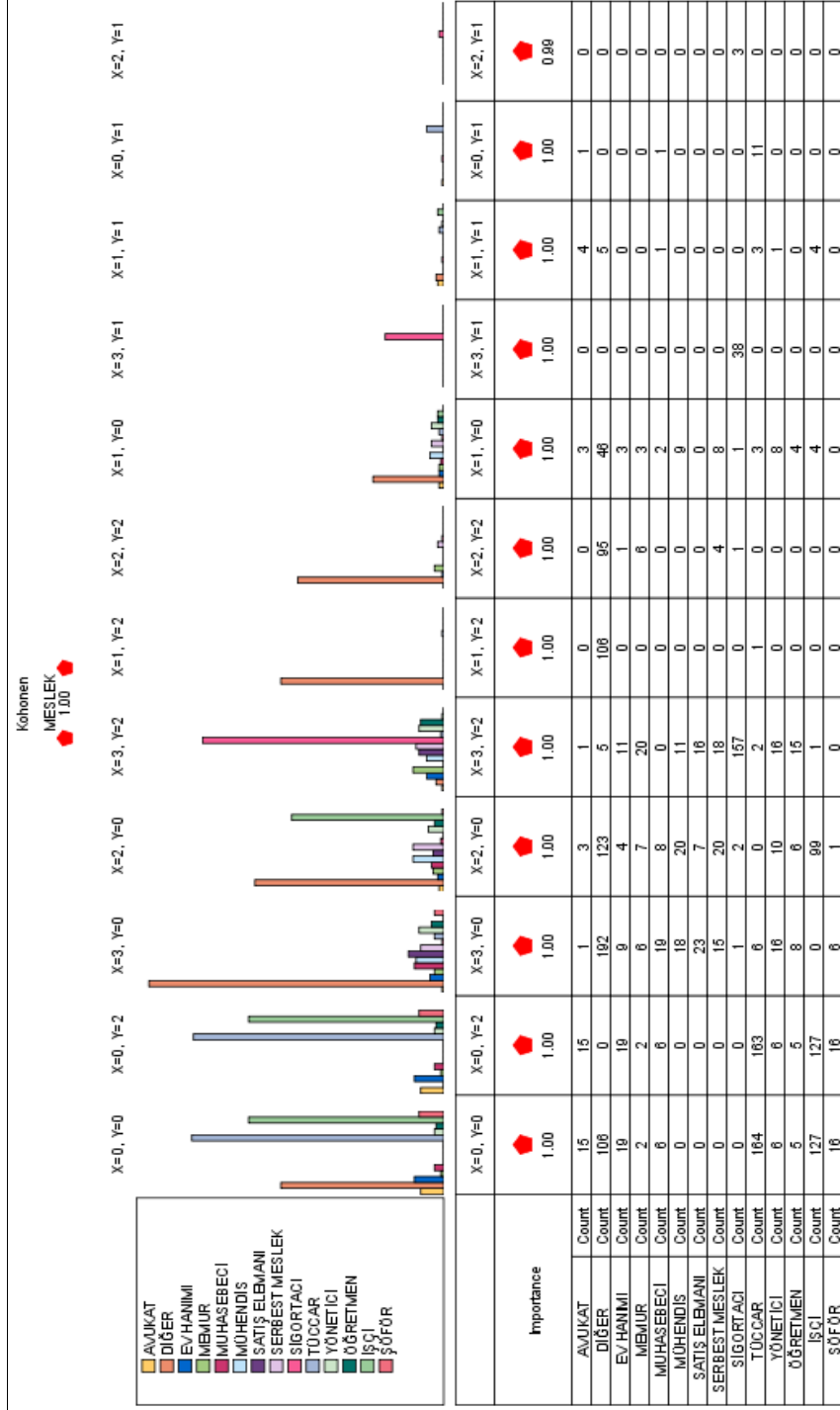
Wang Y., Wu, D., S. (2011) Research of the Bank's CRM Based on Data Mining Technology, Communications in Information Science and Management Engineering, CISME Vol.1 No.4, 30-35.

Wang, C., Pang, C. (2012) Applying Fuzzy Data Mining for an Application CRM, Bulletin of Networking, Computing, Systems, and Software, ISSN 2186-5140, Volume 1, Number 1, January 2012, 46-51.

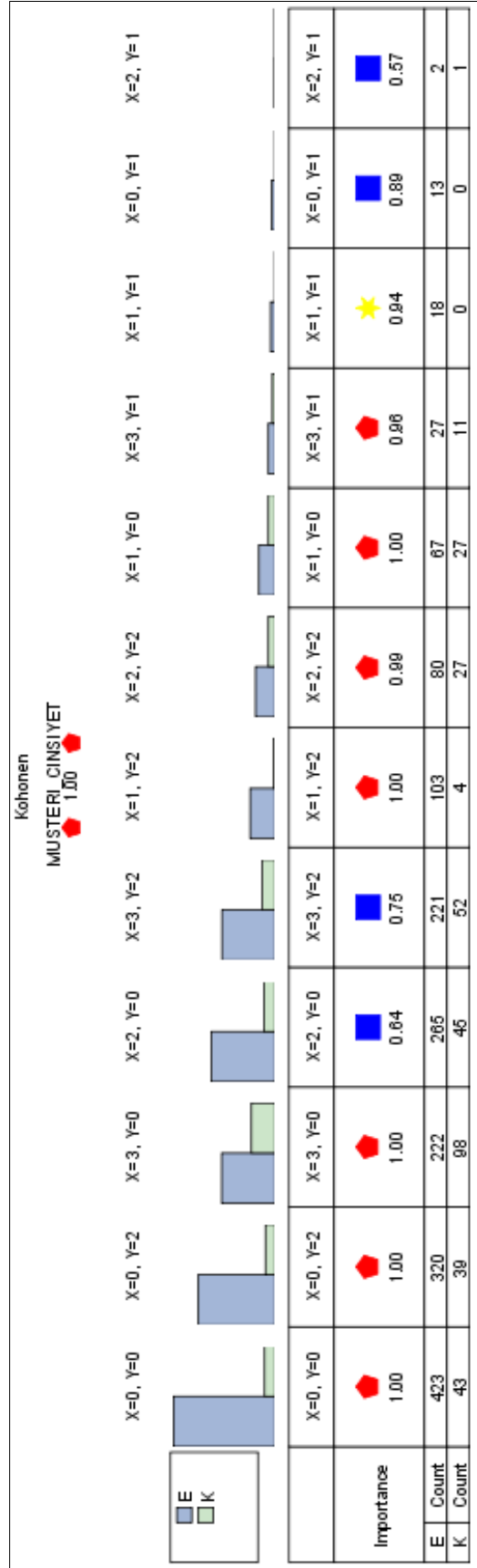
Zaki, M.J., Ogihara, M. (1998) Theoretical foundations of association rules, 3rd SIGMOD'98 Workshop on Research Issues in Data Mining and Knowledge Discovery(DMKD) , Seattle, WA.

EKLER

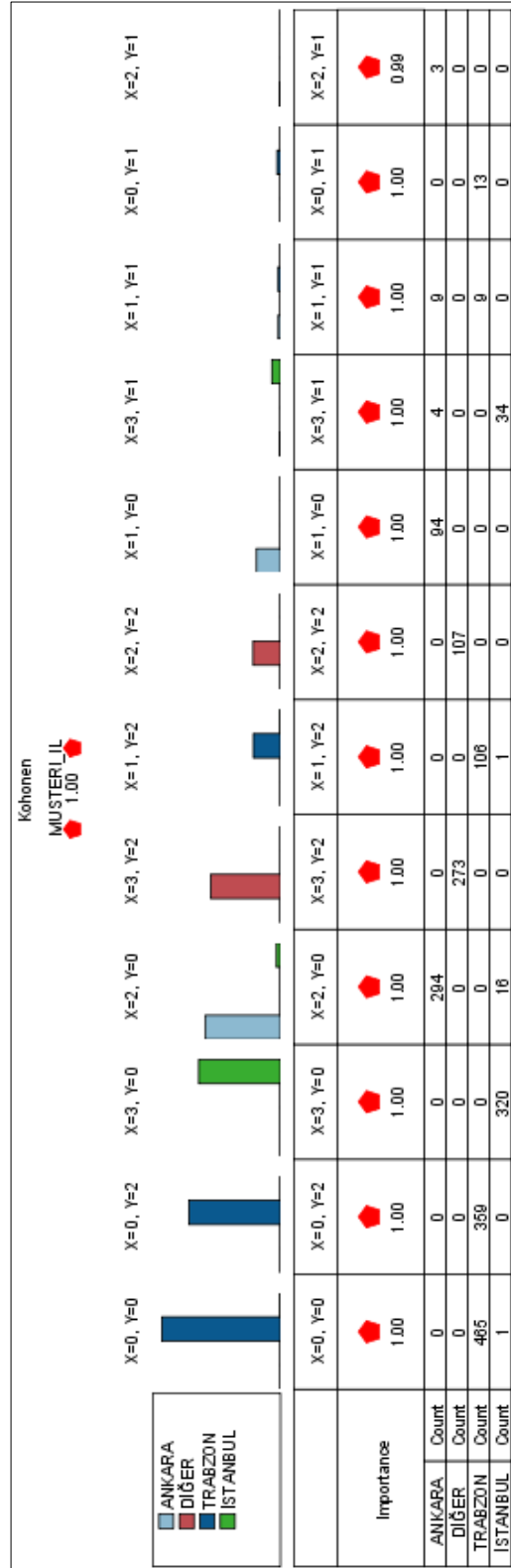
EK-3



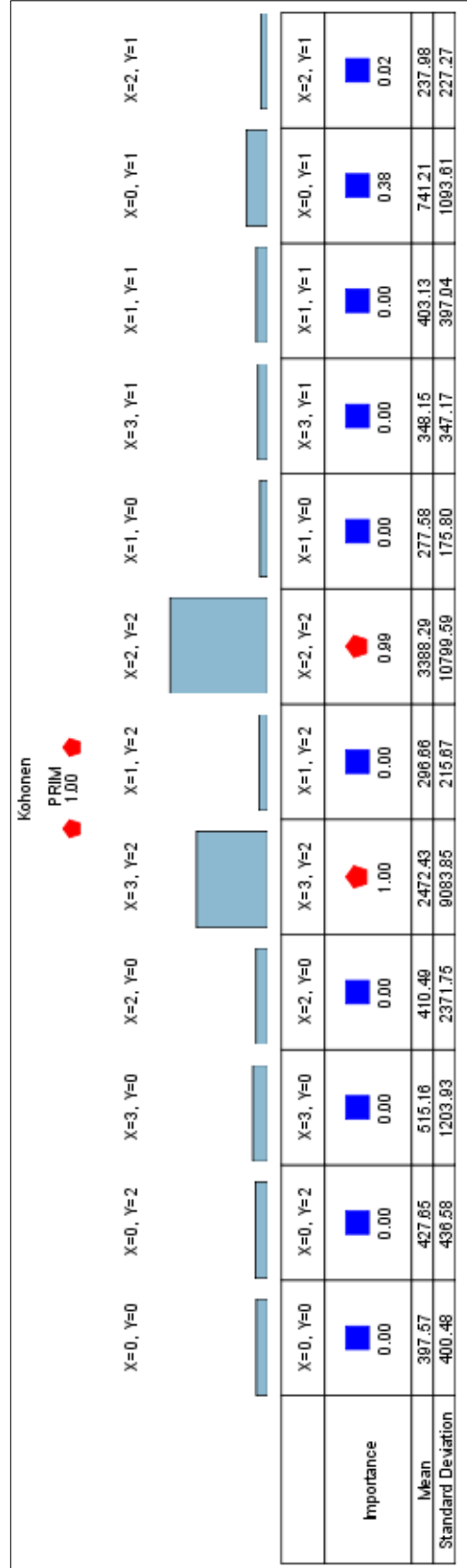
Şekil 1. Kohonen Ağları Analizinde Meslek Değişkeninin Kümeler Üzerindeki Etkisi



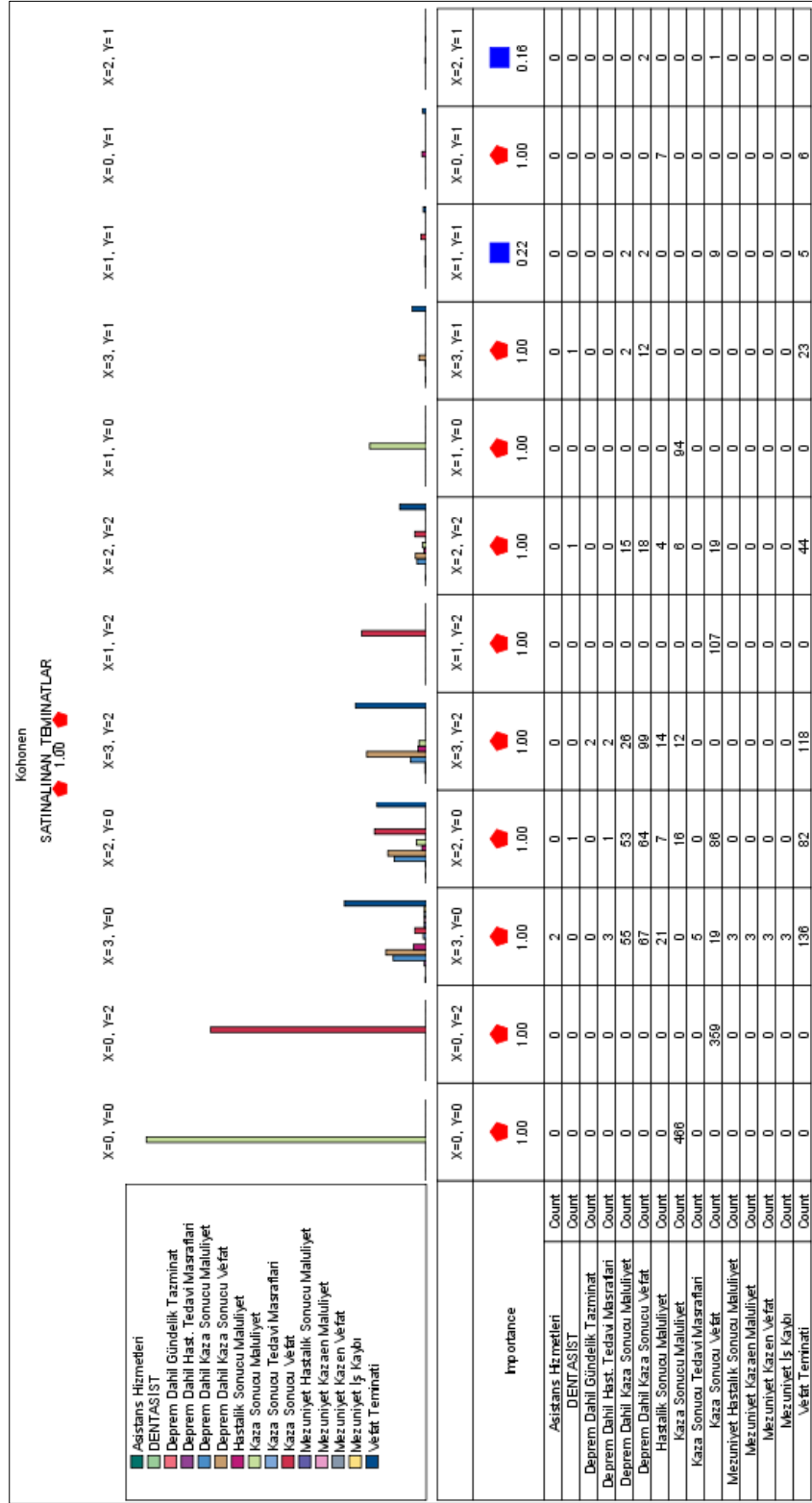
Şekil 2. Kohonen Ağları Analizinde Müşteri Cinsiyetinin Kümeler Üzerindeki Etkisi



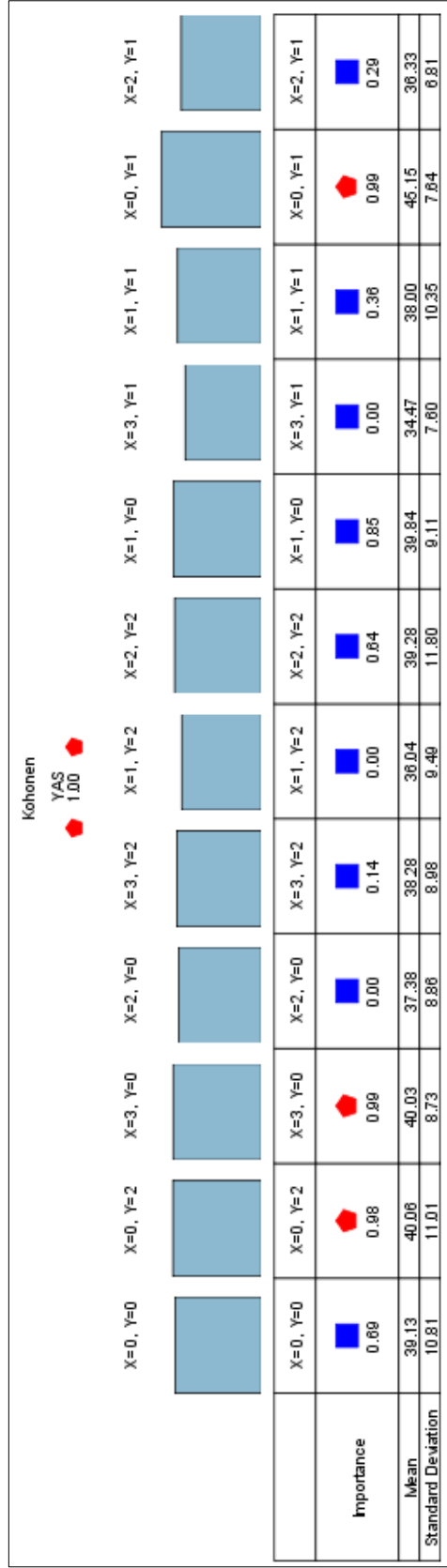
Şekil 3. Kohonen Ağları Analizinde Müşteri İl'i Değişkeninin Kümeler Üzerindeki Etkisi



Şekil 4. Kohonen Ağları Analizinde Prim Değişkeninin Kümeler Üzerindeki Etkisi

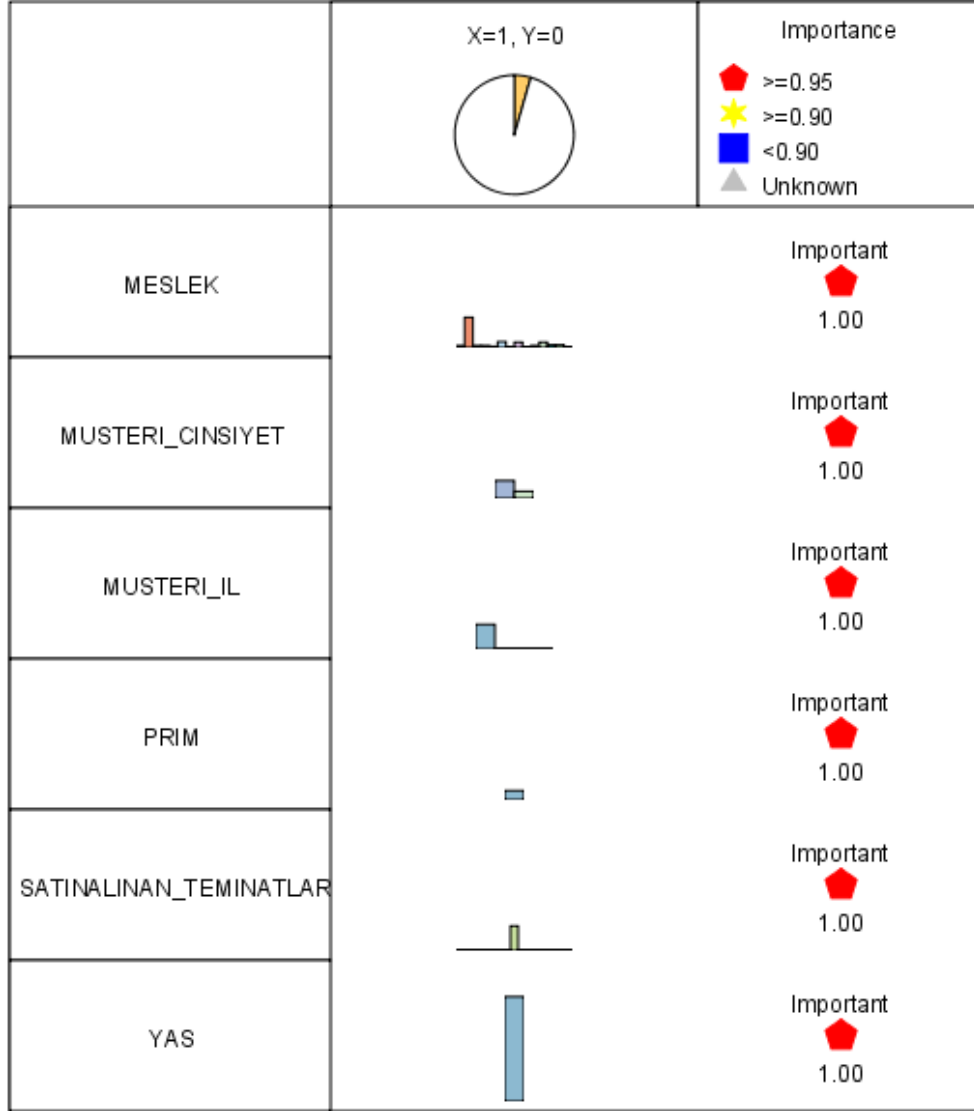


Şekil 5. Kohonen Ağları Analizinde Satın Alınan Teminatların Kümeler Üzerine Etkisi



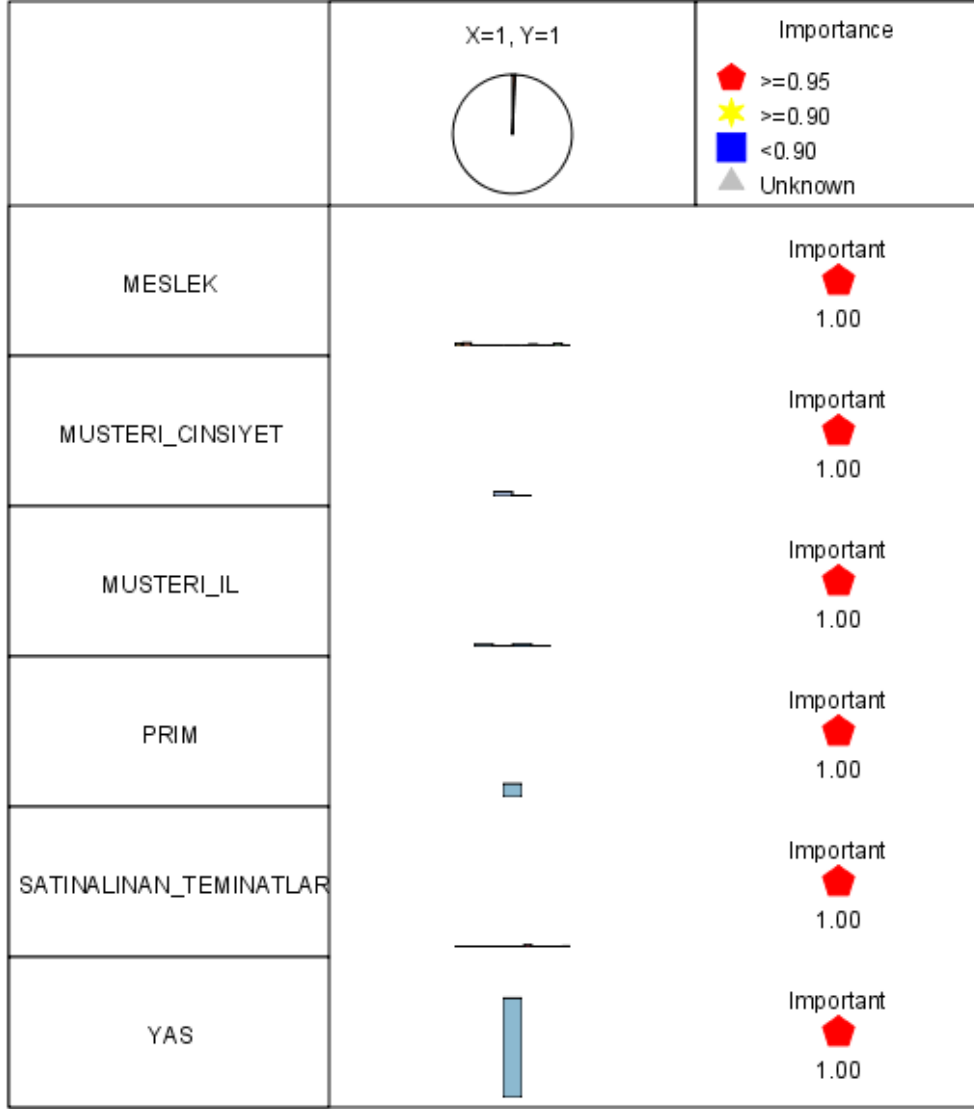
Şekil 6. Kohonen Ağları Analizinde Yaş Değişkeninin Kümeler Üzerine Etkisi

Kohonen
X=1, Y=0



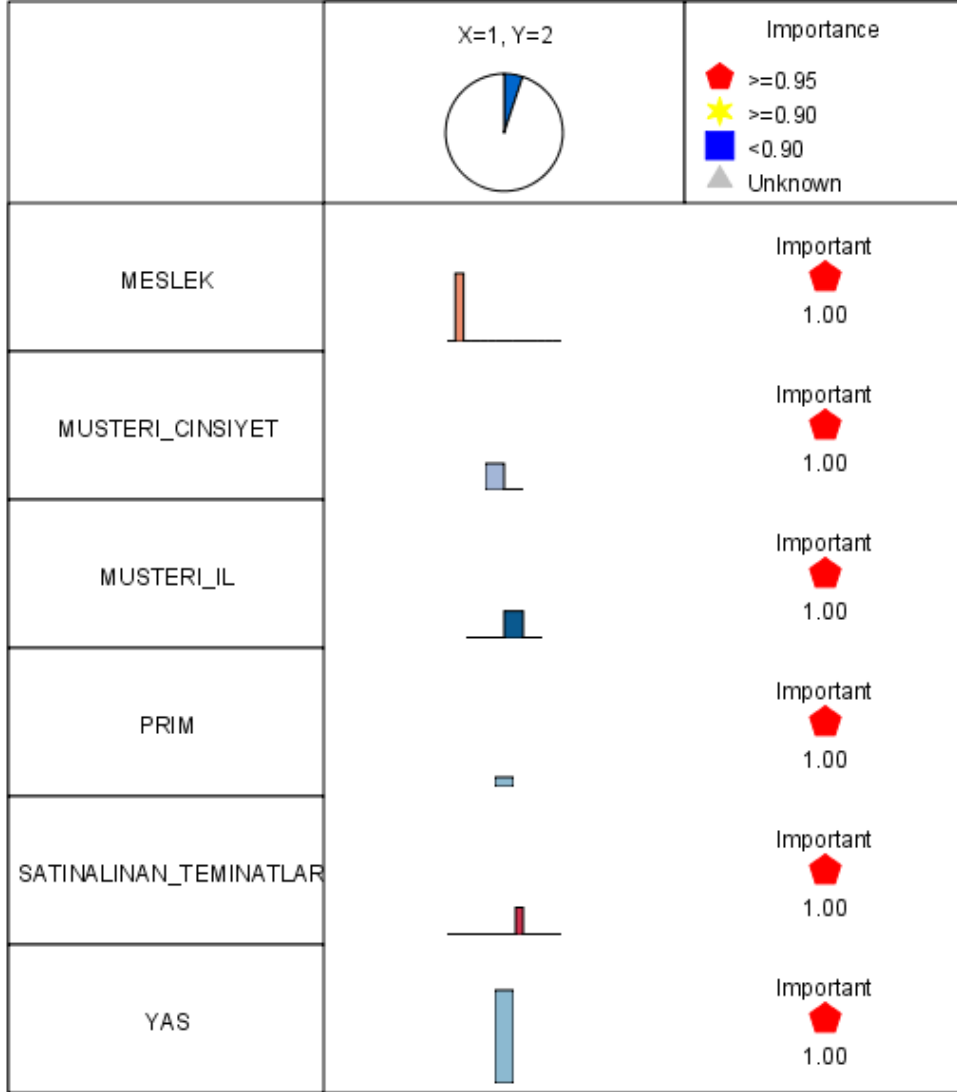
Şekil 7. (X=1,Y=0) Kümesi İçin Kohonen Analizi Sonuçları

Kohonen
X=1, Y=1





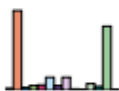











Şekil 8. (X=1,Y=1) Kümesi İçin Kohonen Analizi Sonuçları

Kohonen
X=1, Y=2



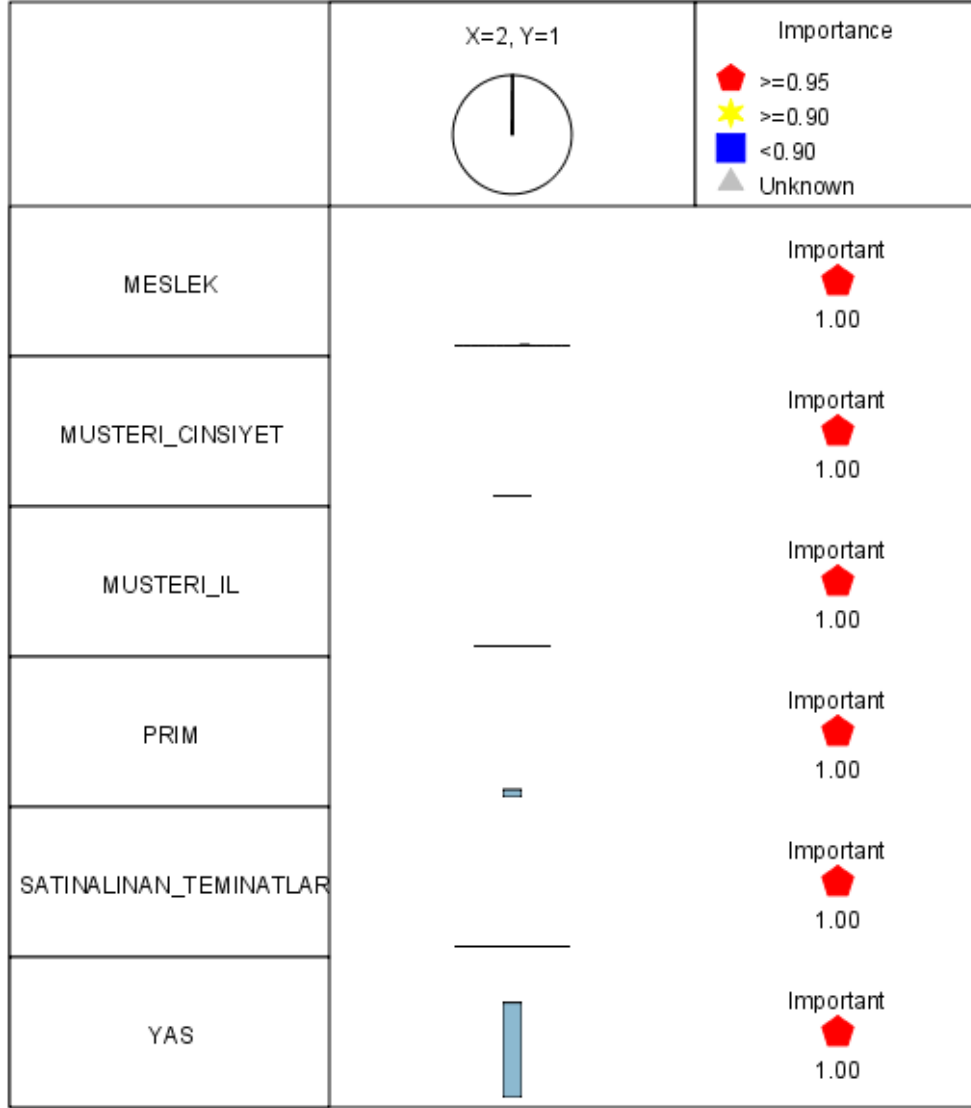
Şekil 9. (X=1,Y=2) Kümesi İçin Kohonen Analizi Sonuçları

Kohonen
X=2, Y=0

	X=2, Y=0	Importance
		
MESLEK		Important  1.00
MUSTERI_CINSIYET		Important  1.00
MUSTERI_IL		Important  1.00
PRIM		Important  1.00
SATINALINAN_TEMINATLAR		Important  1.00
YAS		Important  1.00

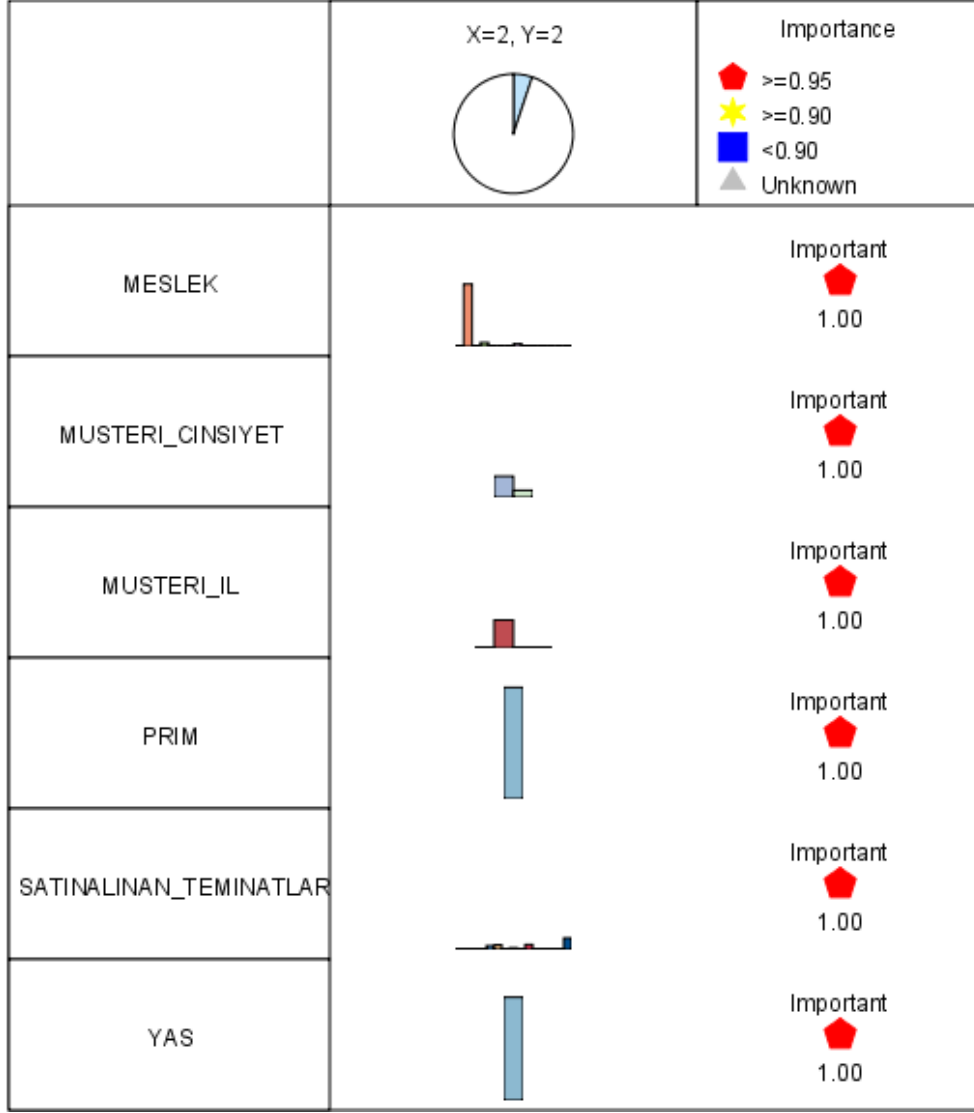
Şekil 10. (X=2,Y=0) Kümesi İçin Kohonen Analizi Sonuçları

Kohonen
X=2, Y=1



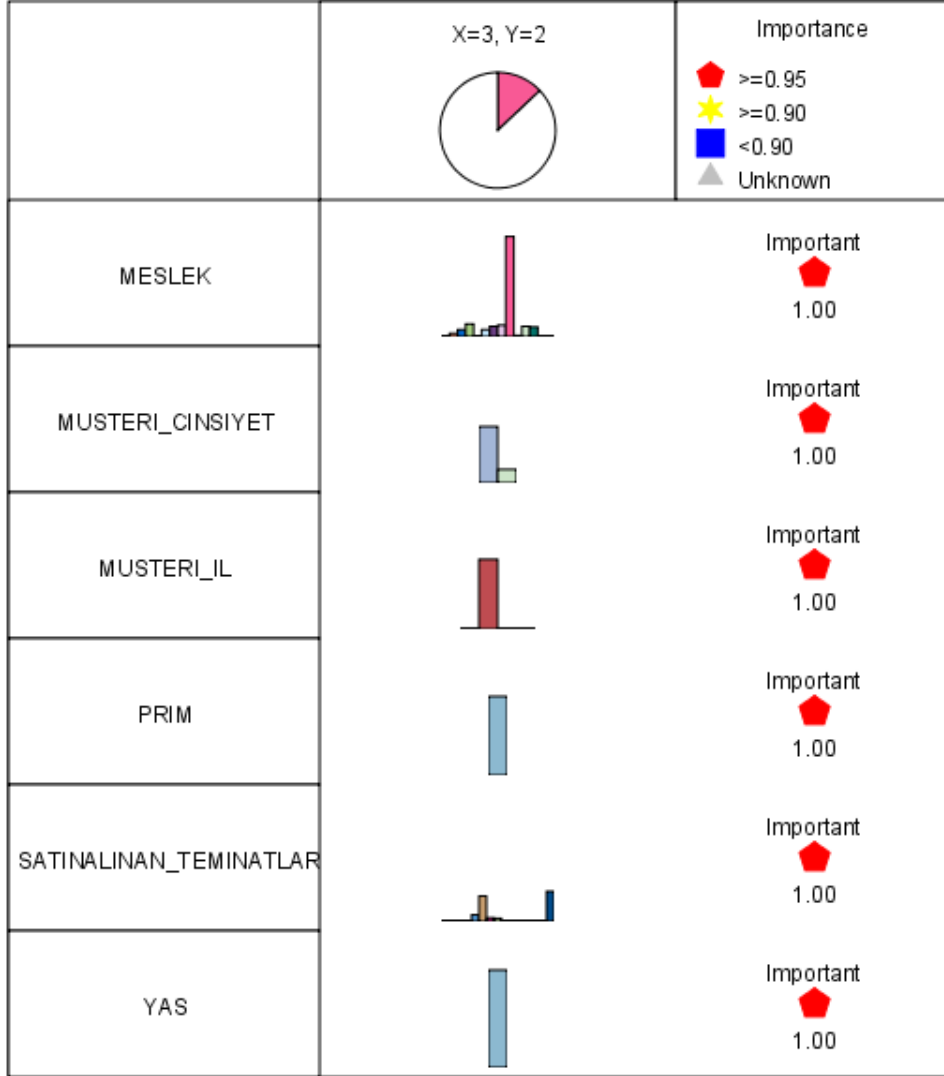
Şekil 11. (X=2,Y=1) Kümesi İçin Kohonen Analizi Sonuçları

Kohonen
X=2, Y=2



Şekil 12. (X=2,Y=2) Kümesi İçin Kohonen Analizi Sonuçları

Kohonen
X=3, Y=2



Şekil 13. (X=3,Y=2) Kümesi İçin Kohonen Analizi Sonuçları

ÖZGEÇMİŞ

Adı Soyadı : Bahar Erol
Doğum Yeri ve Tarihi : Tarsus,1987
Yabancı Dili : İngilizce
E-Posta : bhrerol@gmail.com

Öğrenim Durumu

Derece	Bölüm/Program	Üniversite/Lise	Mezuniyet Yılı
Lise	Sayısal Ağırlıklı	Cengiz Topel Lisesi (Tarsus)	2004
Üniversite	Fizik Bölümü	Marmara Üniversitesi	2009

İş Deneyimi

Yıl	Firma/Kurum	Görevi	İş Veren
2011-2012	Mapfre Genel Yaşam Sigorta	Yazılım Uzmanı	ORATURK
2012-2013	Anadolu Hayat Emeklilik	Veri Uzmanı	ORATURK