

T.C.
İSTANBUL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİM DALI
SAYISAL YÖNTEMLER BİLİM DALI

146871

DOKTORA TEZİ

146871

VERİ MADENCİLİĞİ ARACI OLARAK GENETİK ALGORİTMALAR
İLE YAPAY SİNİR AĞLARI
VE
GENETİK ALGORİTMA-YAPAY SİNİR AĞI MELEZ MODELİNİN
MÜŞTERİ DEĞERLENDİRİLMESİNDE KULLANILMASI

OSMAN ARDA MERİÇ
2502980215

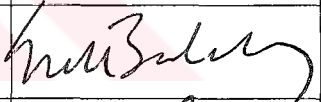
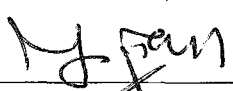
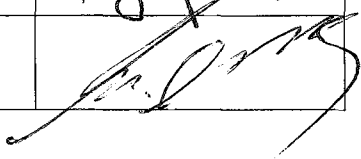
TEZ DANIŞMANI:
PROF. DR. YILMAZ TULUNAY

İSTANBUL, 2004

T.C
İSTANBUL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

TEZ ONAYI

Enstitümüz Sayısal Yöntemler Bilim Dalında 2502980215 numaralı hazırladığı “Veri Madenciliği Aracı Olarak Genetik Algoritma-Yapay Sinir Ağı Melez Modelinin Müşteri Değerlendirilmesinde Kullanılması” konulu ~~YÜKSEK LİSANS~~ / **DOKTORA** TEZİ ile ilgili **TEZ SAVUNMA SINAVI**, Lisansüstü Öğretim Yönetmeliği’nin 28.Maddesi uyarınca **25/06/2004 Cuma** günü saat:10.00’da yapılmış, sorulan sorulara alınan cevaplar sonunda adayın tezinin ...~~KABULÜ~~.....’ne* **OYBİRLİĞİ / OYÇOKLUGUYLA** karar verilmiştir.

JÜRİ ÜYESİ	KANAATİ(*)	İMZA
PROF.DR. YILMAZ TULUNAY	KABULÜ	
PROF.DR. ERDAL BALABAN	KABULÜ	
PROF.DR. ÖNER ESEN	Kabulu	
PROF.DR.İBRAHİM DOĞAN	Kabulu	
DOÇ.DR.ERHAN ÖZDEMİR	Kabulu	

ÖZ

Bu tez çalışmasının temel amacı, ham veriden yola çıkarak işlenmiş bilginin elde edilmesini hedefleyen önemli metodoloji olan Veri Madenciliği(VM) kavramını işlemektir. Bu çalışma VM ile ilgili tüm ana konulara kapsamlı bir bakış açısı sunmaktadır. VM aracı olarak kullanılan çok çeşitli teknikler vardır. Bunlardan bazıları, Bulanık mantık(Fuzzy Logic), istatistiksel teknikler, genetik algoritmalar(GA), yapay sinir ağları (YSA) ve karar ağaçlarıdır.

Bu çalışma, özellikle iki temel VM aracı olan GA'ın ve YSA'ların temellerine, prensiplerine ve uygulamalarına odaklanmıştır. Bunun başlıca sebebi, GA'ın ve YSA'ların esnek ve kullanışlı doğalarıdır. Çalışmamızın uygulama kısmında, GA ve YSA tekniklerini birlikte kullanan melez bir model geliştirdik. Bu modelin temel hedefi, var olan verileri kullanarak bazı yapısal özelliklerin ortaya çıkarılmasıdır. Bu melez model dört temel uygulama aşaması önermektedir:

- A. GA ile birincil eğitim,
- B. Genelleştirilmiş delta kuralı ile birincil eğitim sonuçlarının geliştirilmesi,
- C. Eğitim süreci sonuçlarının test edilmesi,
- D. Modelin tahminde kullanılmasıdır.

Bu çalışmada geliştirilen model, müşteri değerlendirilmesinde kullanılmıştır. Uygulamaya konu olan veriler internet servis sağlayıcısı olan bir Türk firmadan elde edilmiştir. Firma, müşterilerini, grupta ve sınıflandırma yoluyla değerlendirmeyi hedeflemektedir. Geliştirdiğimiz modeli, müşteri değerlendirmeye yönelik olarak eğittik, gerekli testleri yaparak genelleştirilebilirliğini sınadık ve tahminde kullanılmaya hazır hale getirdik.

ABSTRACT

The main purpose of this dissertation is to cover Data Mining (DM) concept, which is an essential methodology to discover knowledge from data. This study presents a comprehensive view of all major issues related to DM. There are various techniques used for DM process such as fuzzy logic, statistical techniques, genetic algorithms (GA), artificial neural networks (ANN), and decision trees.

The dissertation focuses particularly on the fundamentals, principles, and applications of two major DM techniques; GA and ANN, because of their flexible and practical nature. We have constructed a hybrid model, which uses GA and ANN together, for the application section of the study. The main aim of the model is to recognize patterns from the historical data set. The hybrid model consists of four main stages;

- A. The initial training of the ANN with GA
- B. Improving the performance of the initial training stage with the generalized delta rule,
- C. Testing the results,
- D. Using the model for prediction.

In this study, the hybrid model was used for customer evaluation. Customer records were obtained from a Turkish Internet service provider company. The company would like to evaluate their customers by grouping and scoring them. We have used our model for the evaluation. We have trained the ANN-GA model, tested the results, and generalized the model for further forecasts.

ÖNSÖZ

Bu çalışmada, VM kavramı merkez alınarak, GA ve YSA incelenmiş ve özgün bir GA-YSA melez modeli yaratılmaya çalışılmıştır. Bu konunun seçilmesinde temel güdülerden birisi VM kavramının dünyada ve Türkiye’de öneminin her geçen gün daha çok artması ve geleneksel yöntemlere paralel olarak YSA ve GA gibi modern yöntemlerin de uygulama alanlarının genişlemesidir.

Başlangıçta çok karmaşık gibi görünen GA ve YSA, canlıların özelliklerini ve yaşamlarını basit bir benzeşimle bilgisayar ortamına taşımayı hedefler. Çalışmada geliştirdiğimiz melez model de aynı amacı taşımaktadır.

Geliştirdiğimiz model, müşteri değerlendirilmesi uygulamasında kullanılmıştır, ancak modelde çok basit değişiklikler yapılarak model, çok sayıda bağımlı ve çok sayıda bağımsız değişken içeren her türlü veri setlerinin, doğrusal veya doğrusal olmayan ilişkilerinin analizinde kullanılabilir. Ayrıca GA modülü, yine basit düzenlemelerle, her türlü doğrusal olmayan optimizasyon probleminin çözümünde bir alternatif olarak düşünülebilir.

Bu çalışmanın her aşamasında yanımda olan ve bana sabır gösteren eşim Günseli Meriç’e, kendisine çok zaman ayıramadığım için bana kırılmayan minik kızım Perisu Meriç’e, anneme, babama ve tüm dostlarıma teşekkür ederim.

Ayrıca bu çalışmanın her aşamasını titizlikle inceleyen, fikir ve eleştirileriyle daha iyiye yol almamı amaçlayan ve beni sürekli olarak destekleyen sayın hocam Prof. Dr. Yılmaz Tulunay başta olmak üzere sayın Prof Dr. Erdal Balaban hocama ve diğer tüm hocalarıma en içten teşekkürlerimi sunarım.

İÇİNDEKİLER

ÖZET	iii
ABSTRACT	iv
ÖNSÖZ	v
İÇİNDEKİLER	vi
ŞEKİLLER LİSTESİ	xii
TABLolar LİSTESİ	xiv
GİRİŞ	1
Çalışmanın Genel Amaçları ve Hedefleri.....	3
Çalışmanın Yapısı	4
Konunun Dünya Literatüründeki Yeri.....	7
Çalışmanın Kapsamı	8
BÖLÜM 1. VERİ AMBARLARI (VA)	10
1.1 Tanım	12
1.2 VA'ların Karakteristikleri	12
1.3 Veri Küpleri.....	14
1.4 Operasyonel Veri Tabanı ile VA'nın Karşılaştırılması	15
1.5 Normalizasyon.....	16
1.5.1 Birinci Normal Form	16
1.5.2 İkinci Normal Form.....	17
1.5.3 Üçüncü Normal Form.....	18
1.5.4 Dördüncü Normal Form	19
1.5.5 Beşinci Normal Form	21
1.5.6 Normal Formların Kullanımı	21
1.6 VA'larda Yüksek Seviyeli Normal Formların Sonuçları	22
1.7 VA Organize Ederken Kontrol Listelerinin Kullanımı	22
1.8 VA Oluşturma Aşamaları	24

1.9 VA Mimarileri	25
1.9.1 İki Kademeli Mimari.....	26
1.9.2 Üç Kademeli Mimari.....	27
1.9.3 Aşağıdan Yukarı Mimari.....	28
1.10 Dünyada VA Sektöründe Pazar Payları:.....	29
1.11 OLAP Pazarının Büyüme Seyri	30
BÖLÜM 2. VERİ MADENCİLİĞİ.....	32
2.1 Tanım	32
2.2 VM'ni Gerekli ve Mümkün Kılan Etkenler	34
2.2.1 VM Uygulamalarının Başarıya Ulaşmasını Sağlayan Alt Yapı Çalışmaları.....	34
2.2.2 VM Süreci	36
2.3 VM ile Yapılan Temel Görevler ve VM Kullanım Alanları.....	37
2.3.1 Sınıflandırma.....	38
2.3.2 Gruplandırma	39
2.3.3 Görselleştirme.....	39
2.3.4 Bağintı Kurma (Sepet Analizi).....	40
2.3.5 İstisnaların Tespiti	40
2.3.6 Doküman Madenciliği.....	41
2.3.7 Tahmin	41
2.5 Temel VM Yöntemleri	42
2.5.1 Karar Ağaçları	42
2.5.2 İstatistikî Teknikler.....	44
2.5.3 Bulanık Mantık (Fuzzy Logic).....	47
2.5.4 Genetik Algoritmalar(GA).....	47
2.5.5 Yapay Sinir Ağları	48
2.6 VM ve Ahlakî Konular	48
2.7 VM Araçları Geliştiren Başlıca Firmalar	49
BÖLÜM 3. GENETİK ALGORİTMALAR.....	51
3.1 Tanım	51
3.2 Tarihçe ve Genel Prensipler	51

3.3 Genetik Algoritmaların Temel Adımları.....	52
3.4 Algoritmayı Açıklayıcı Bir Örnek	53
3.4.1 Başlangıç Kümesinin Belirlenmesi.....	53
3.4.2 Göreceli Uygunluk Değerlerinin Belirlenmesi.....	54
3.4.3 Yeni Kümeye Girecek Üyelerin Tespiti.....	55
3.4.4 Yeni Kümeye Girecek Üyeler Arasında Eşleşme.....	57
3.4.5 Mutasyon	60
3.5 Yeni Kümeye Girecek Üyelerin Belirlenmesinde Uygulanabilecek Farklı Yöntemler	61
3.5.1 Uygunluk Değerleri Temel Alınarak Rasgele Seçim	61
3.5.2 Seçkincilik	61
3.5.3 Deterministik Seçim	61
3.5.4 Kalanların Tesadüfi Olarak Tamamlandığı Deterministik Seçim	62
3.6 Eşleşme(Crossover) İçin Uygulanabilecek Farklı Yöntemler	62
3.6.1 Tek Noktadan Kesişim	62
3.6.2 Çok Noktadan Kesişim.....	63
3.6.3 Çok Noktadan Tesadüfi Kesişim	63
3.7 Genetik Algoritmaların Karmaşık Bir Örnekle İncelenmesi	64
3.7.1 Uygulamada Kullanılan Parametreler	65
3.7.2 Başlangıç Kümesi	66
3.7.3 Algoritmanın Çalışması İle Elde Edilen Sonuçlar	73
3.7.4 Sonuçların Yorumlanması.....	78
3.8 Genetik Algoritmaların Güçlü ve Zayıf Yanları.....	79
3.8.1 Genetik Algoritmaların Güçlü Yanları	80
3.8.2 Genetik Algoritmaların Zayıf Yanları	80
BÖLÜM 4. YAPAY SİNİR AĞLARI.....	81
4.5 Sinir Fizyolojisi ve Yapay Sinir Ağları	81
4.1.1 Nöronların Yapıları.....	82
4.1.2 Bilgisayarlarla Beyin Özelliklerinin Karşılaştırılması.....	85
4.2 Tarihsel Gelişim.....	86
4.3 Yapay Sinir Ağlarının Temel Elemanları.....	87

4.4 Yapay Sinir Ağlarında Öğrenme	88
4.4.1 Hebb Kuralı	89
4.4.2 Geri ve İleri Beslemeli Öğrenme	90
4.4.3 En Küçük Kareler, Delta Kuralı ve Geri Yayılım	90
4.5 Çok Katmanlı İleri Beslemeli Geri Yayılımlı Yapay Sinir Ağları.....	91
4.5.1. ÇKİBGYYSA'nın Genel Yapısı.....	91
4.5.2 ÇKİBGYYSA'nda Kullanılan Temel Fonksiyonlar.....	93
4.5.2.1 Sigmoid Fonksiyonu	94
4.5.3 ÇKİBGYYSA'nda Kullanılan Yöntemin Açıklanması	95
4.5.4 ÇKİBGYYSA'nda Öğrenme	97
4.5.5 ÇKİBGYYSA Verilerin Kullanımı	102
4.5.6 ÇKİBGYYSA Parametrelerinin Sonuca Etkileri	102
4.5.6.1 Öğrenme Oranı	103
4.5.6.2 Momentum Katsayısı.....	106
4.5.6.3 İterasyon Sayısı ve Aşırı Eğitim.....	109
4.5.6.4 Girdi Katman Eleman Sayısı	110
4.5.6.5 Gizli Katman Sayısı	110
4.5.6.6 Gizli Katman Eleman Sayısı	110
4.5.6.7 Çıktı Katman Eleman Sayısı.....	113
4.5.6.8 Kabul Edilebilir En Büyük Hata Değeri	113
4.5.6.9 Eğitim Veri Seti Büyüklüğü	113
4.5.6.10 Test Veri Seti Büyüklüğü	114
4.5.6.11 Başlangıç Ağırlık Değerlerinin Öğrenmeye Etkisi	114
4.6 Genetik Algoritmaların Yapay Sinir Ağı Modellerinde Kullanımı	115
4.6.1 GA İle Optimum Parametrelerin Tahmini	116
4.6.2 GA İle Optimum Ağırlık Değerlerinin Elde Edilmeye Çalışılması.....	116
4.7 Yapay Sinir Ağlarının Kullanım Alanları.....	117
4.7.1 Mühendislik/Teknolojik Uygulamalar.....	117
4.7.2. Sosyal Bilimler/İşletmecilik Uygulamaları	118
4.7.2.1 Finans Uygulamaları	118
4.7.2.2 İnsan Kaynakları Uygulamaları.....	119

4.7.2.3 Muhasebe Uygulamaları.....	119
4.7.2.4 Pazarlama ve Müşteri İlişkileri Yönetimi Uygulamaları	119
4.7.2.5 Üretim Yönetimi Uygulamaları.....	120
4.8 Yapay Sinir Ağı Algoritmalarının Güçlü ve Zayıf Yönleri.....	120
4.8.1 YSA'nın Güçlü Yönleri	120
4.8.2 YSA'nın Zayıf Yönleri.....	121
BÖLÜM 5. UYGULAMA	122
(ÇKİBGYYSA'NIN GA İLE MELEZ KULLANIMI VE MÜŞTERİ GRUPLANDIRMASI VE SINIFLANDIRMASI)	122
5.1 Problemin Tanımlanması.....	122
5.2 Veri Seti ve Özellikleri.....	123
5.3 Veri Hazırlanması	124
5.3.1 Veri Dönüştürme	125
5.3.2 Verilerin Temizlenmesi.....	125
5.3.3 Verilerin Süzülerek Kullanışlı Hale Getirilmeleri	126
5.4 Modelin Oluşturulması	126
5.4.1 Modelin Algoritmik Yapısı	126
5.4.2 Uygulamanın Hedefleri	129
5.4.3 Uygulama Yöntemi ve Şeması.....	129
5.5 Modelin Uygulanması	131
5.5.1 Uygulama Veri Seti ve Aralıkları	131
5.5.2 Modelin, Müşteri Gruplandırmada Kullanılması	131
5.5.2.1 Girdi Katman Değer Aralığı	133
5.5.2.2 Çıktı Katman Değer Aralığı.....	133
5.5.2.3 Sabit Olmayan Öğrenme Oranı	134
5.5.2.4 Model Tarafından Elde Edilen Tahmini Çıktı Değerlerinin Kullanımı	134
5.5.3 Modelin, Müşteri Sınıflandırmada Kullanılması	139
5.5.3.1 Müşteri Sınıflandırması Çalışmasında Hata Teriminin İterasyona Bağlı Değişimi	141
5.5.3.2 Müşteri Sınıflandırması Çalışmasında Elde Edilen Ağırlık Matrisi	

Değerleri.....	142
5.5.3.4 Müşteri Sınıflandırması Çalışmasında Elde Edilen Sonuçların Kullanımı	144
5.6 Modelin Performansının Test Edilmesi.....	145
5.6.1 Model Performansını Geliştirici Test	146
5.6.2 Aşırı Öğrenmeyi Engelleyici Test	146
5.6.3 Algoritmanın Geçerliliğini Araştıran Test.....	147
5.6.3.1 Müşteri Gruplama Uygulaması İçin GA-YSA Melez Modeli ile Mat-Lab Programının Sonuçlarının Kıyaslanması	147
5.6.3.2 Müşteri Sınıflandırma Uygulaması İçin GA-YSA Melez Modeli ile Mat-Lab Programının Sonuçlarının Kıyaslanması	148
5.7 Model İçerisindeki Kısıtlar ve Kabuller.....	149
5.7.1 Kullanılan Bilgisayar ve Programlama İle İlgili Parametrik Kısıtlar ..	150
5.7.1.1 Çıktı Katman Eleman Sayısı.....	150
5.7.1.2 Gizli Katman Sayısı	150
5.7.1.3 Aktivasyon Fonksiyonu.....	151
5.7.2 Uygulama İle İlgili Kısıtlar ve Kabuller.....	151
SONUÇ	152
KAYNAKLAR	156
EK 1. GENETİK ALGORİTMA PROGRAMI	165
EK 2. GA YSA MELEZ MODELİNE AİT PROGRAM	171
EK 3. MÜŞTERİ GRUPLAMA EĞİTİM VE TEST VERİ SETİ.....	183
EK 4. MÜŞTERİ SINIFLANDIRMA EĞİTİM VE TEST VERİ SETİ	190
ÖZGEÇMİŞ.....	194

ŞEKİLLER LİSTESİ

Şekil 1.1	VA Mimarisi	14
Şekil 1.2	Veri Küpü	15
Şekil 1.3	İki Kademeli Mimari	26
Şekil 1.4	Üç Kademeli Mimari	27
Şekil 1.5	Aşağıdan Yukarıya Mimari	28
Şekil 1.6	OLAP Pazarının Büyüme Seyri	30
Şekil 2.1	Veri Bilgi Dönüşümü	33
Şekil 2.2	Karar Ağacı Örneği	43
Şekil 2.3	Karar Ağaçları ile Kural Üretilmesi	44
Şekil 3.1	Rulet Usulü ile Seçim	57
Şekil 3.3	Tek Noktadan Kesişim	63
Şekil 3.4	Çok Noktadan Kesişim	63
Şekil 3.5	Seçilen Fonksiyonun Üç Boyutlu Grafiği	64
Şekil 3.6	Ortalama Fonksiyon Değerinin Nesle Bağlı Değişimi	78
Şekil 3.7	Maksimum Fonksiyon Değerinin Nesle Bağlı Değişimi	79
Şekil 4.1	Nöron Yapısı	82
Şekil 4.2	Sinapsın Yapısı	83
Şekil 4.3	Nöronlar Arası İlişkinin Şematik Gösterimi	84
Şekil 4.4	Yapay Sinir Ağlarının Temel Elemanları	87
Şekil 4.5	ÇKİBGYYSA Genel Yapısı	92
Şekil 4.6	YSA Katmanları ve Fonksiyonlar	93
Şekil 4.7	Sigmoid Fonksiyonun Grafiği	95
Şekil 4.8	Örnek Uygulama	95
Şekil 4.9	Değişik Öğrenme Oranlarının Performansı Grafiği	104
Şekil 4.10	Değişik Öğrenme Oranlarının Performansının Karşılaştırılması	105
Şekil 4.11	Eğitim Süresi, Hata Grafiği ve Yerel Minimum	107
Şekil 4.12	Momentum Katsayısının Öğrenmeye Etkisi	108

Şekil 4.13	Momentum Katsayısının Öğrenmenin İlk Aşamasına Etkisi.....	109
Şekil 4.14	Gizli Katman Eleman Sayısı ve Toplam İşlem Süresi	111
Şekil 4.15	Gizli Katman Eleman Sayısının YSA Performansına Etkisi.....	112
Şekil 4.16	Başlangıç Ağırlık Değerlerinin Seçimi ve Öğrenme.....	115
Şekil 5.1	Geliştirilen Melez Yapının Şematik Gösterimi	128
Şekil 5.2	Geliştirilen Melez Modelin Uygulanması.....	130
Şekil 5.3	Müşteri Gruplamasına Ait Öğrenme Süreci ve Hata Değerleri.....	135
Şekil 5.4	Müşteri Gruplamasına Ait Öğrenme Süreci ve Hata Değerleri.....	141



TABLÖLAR LİSTESİ

Tablo 1.1	VA ve Operasyonel Veri Tabanı Kıyaslaması	15
Tablo 1.2	Birinci Normal Forma Aykırı Tabloya Örnek	16
Tablo 1.3	İkinci Normal Forma Aykırı Tabloya Örnek.....	17
Tablo 1.4	Üçüncü Normal Forma Aykırı Tabloya Örnek.....	18
Tablo 1.5	Dördüncü Normal Forma Aykırı Tabloya Örnek	19
Tablo 1.6	Dördüncü Normal Forma Uygun İlk Tablo	20
Tablo 1.7	Dördüncü Normal Forma Uygun İkinci Tablo	20
Tablo 1.8	Beşinci Normal Forma Aykırı Tabloya Örnek	21
Tablo 1.9	Kontrol Listesi.....	22
Tablo 1.10	Dünyada VA Pazarı ve Pazar Payları.....	29
Tablo 2.1	VM Hizmet ve/veya Ürün Sağlayıcıları	50
Tablo 3.1	Başlangıç Kümesi.....	54
Tablo 3.2	Uygunluk Değerleri.....	55
Tablo 3.3	Yeni Kümede Yer Alma Sayıları	56
Tablo 3.4	Eşleşmeye Tabi Tutulacak Üyeler	57
Tablo 3.5	Eşleşme(Crossover) İşlemi	58
Tablo 3.6	Eşleşme Sonrası Yeni Üyeler.....	59
Tablo 3.7	GA Süreci Ortalama Fonksiyon Değerleri	59
Tablo 3.8	Uygulamada Kullanılan Parametreler.....	65
Tablo 3.9	x ve y Değerleri	66
Tablo 3.10	Başlangıç Kümesi ve Elde Edilen Değerler	66
Tablo 3.11	Başlangıç Kümesi ve Küme Üyelerinin Yeni Nesilde Yer Alma Sayıları	69
Tablo 3.12	200 Kuşak Sonra Elde Edilen Sonuçlar.....	73
Tablo 3.13	Microsoft Excel Çözücü Aracı İle Elde Edilen Sonuçlar.....	76
Tablo 4.1	Bilgisayarla Beyin Özelliklerinin Karşılaştırılması	85
Tablo 4.2	YSA Performans Testi Modelinin Özellikleri	104

Tablo 5.1	Uygulamada Kullanılan Veri Setinin Özellikleri.....	124
Tablo 5.2	Müşteri Gruplamasına Ait Parametreler	132
Tablo 5.3	Müşteri Gruplama Sonuçları.....	135
Tablo 5.4	Müşteri Gruplamada İterasyon Sayısı ve Hata Kareleri Ortalaması Değerlerinin, Eğitim ve Test Verileri İçin Değişimi.....	136
Tablo 5.5	Müşteri Gruplamada, Eğitim Sonrası, Girdi Katman Ve Gizli Katman Arası Ağırlık Değerleri	137
Tablo 5.6	Müşteri Gruplamada, Eğitim Sonrası Gizli Katman, Çıktı Katman Arası Ağırlık Değerleri	138
Tablo 5.7	Müşteri Sınıflandırmasına Ait Parametreler.....	140
Tablo 5.8	Müşteri Sınıflandırmada İterasyon Sayısı ve Hata Kareleri Ortalaması Değerlerinin, Eğitim ve Test Verileri İçin Değişimi	142
Tablo 5.9	Müşteri Sınıflandırmada, Eğitim Sonrası Girdi Katman, Gizli Katman Arası Ağırlık Değerleri	143
Tablo 5.10	Müşteri Sınıflandırmada, Eğitim Sonrası Gizli Katman, Çıktı Katman Arası Ağırlık Değerleri	144
Tablo 5.11	Müşteri Gruplamaya Ait Çeşitli Denemelere Ait Sonuçlar	148
Tablo 5.12	Müşteri Sınıflandırmaya Ait Çeşitli Denemelere Ait Sonuçlar	149

GİRİŞ

20. yüzyılın ikinci yarısında bilgisayar ve uygulamaları ile tanışan iş hayatı, bu aracı, önceleri elle yapılan ve çok büyük zaman ve iş gücü kaybına yol açan uygulamaları otomatikleştirmek, bu sayede zaman ve kaynak tasarrufu yapmak amacıyla kullanmaya başlamıştır. Bilgisayarların iş hayatında yoğun bir şekilde yer aldıkları ilk uygulamalar muhasebe kayıtlarının tutulması ve bilanço, gelir tablosu gibi standart formların ve raporların elde edilmesi ile başlamıştır. Bu görüşü destekleyen bir gerçek de, başlarda bir çok kuruluşun bilgi işlem departmanlarını muhasebe veya finans bölümlerine bağlı bir destek ünitesi olarak organize etmeleridir.

İş dünyasında bilgi teknolojilerinin yeni yeni kullanılmaya başlandığı ilk aşamada veri büyüklükleri görece olarak az ve verilerin özellikleri basit sayılabilecek durumda idi. Bu süreçte temel amaç verilerin derlenerek elektronik ortama taşınması idi. Verilerin işlenerek, kullanılabilir bilgi kümeleri haline getirilmeleri en önemli problemlerden biri sayılmayabilirdi.

İlk aşaması yukarıda özetlenen bilgi teknolojilerinin gelişimi, Richard L. Nolan'a ait, "Bilgi işlemin altı aşaması modeli" ile etraflıca incelenmiştir. Nolan, modelinde bilgi işlemin gelişimini altı aşamada değerlendirmiştir:

1. Aşama: Başlangıç
2. Aşama: Yayılma
3. Aşama: Kontrol
4. Aşama: Bütünleşme
5. Aşama: Veri Yönetimi
6. Aşama: Olgunluk

Sunulan gelişmişlik aşamaları, hem kronolojik olarak değerlendirilebilir, hem de kuruluşlar bazında ilgili kuruluşun bilgi teknolojileri kullanımında ulaştığı olgunluk seviyesini değerlendirmede kullanılabilir.

Tarihsel olarak ve/veya kuruluş bazında olgunluk seviyesi olarak beş ve altıncı seviyelere ulaşıldığında organizasyonlar için temel sorun, veri elde edilmesi veya verilerin elektronik ortama taşınması değil bu veri dağlarının anlamlı, kullanılabilir ve istenildiği zaman istenildiği alanda, doğruluğuna, tamlığına ve tutarlılığına güvenilir bir şekilde kullanılması şekline dönüşür.

Bu aşamada, optimal ve bilimsel veri tabanı yönetimi, donanım ve yazılımların eş uyumunun sağlanması, bütünleşik teknolojilerin uygulanması, bilgi teknolojilerinin stratejik bir rekabet aracı olarak kullanılması gibi parametreler öncelikli başlıklar halini alır.¹

1950'li yıllarda gelişmeye başlayan bilişim teknolojileri sektörü, başlangıç aşamasından bu yana büyük değişiklikler yaşamıştır. Hızlı gelişen bilgisayar teknolojileri ve ani değişen işletme istekleri, bilişim sistemlerinin kurumsal rollerini de yakından etkilemiştir. Bir kurumsal bölüm olarak görülen bilişim sisteminin, güncel konular ve problemler ele alındığında, tüm kurumu etkileyebilecek bir yaygınlığa sahip olmaya başladığı görülmektedir. Buna verilebilecek en güzel örneklerden birisi "Yıl 2000" problemidir. Bu problem,

¹ HOOFFER, J.A., GEORGE, J.F., VALACICH, J.S.: " **Modern System Analysis and Design**" İkinci Baskı, New York, Addison Wesley Longman Inc., 1999, s.12.

bilgi veya bilişim sisteminin, olası kurumsal etkilerinin derecesi ile ilgili önemli bilgiler sunmuştur.²

Açıklanan temel gerçekler, ayrıntıları bu tez içerisinde tartışılacak veri ambarlarının(Data Warehouse) (VA) oluşturulması ve bu ambarlar kullanılarak veri madenciliği(Data Mining) (VM) faaliyetlerinin devreye girmesini zorunlu kılar.

Çalışmanın Genel Amaçları ve Hedefleri

Bu çalışmanın genel amaçları:

- A. VM konusunun önemini açıklamak, VM'nin sadece bir teknik, bir paket program ya da istatistiki bir çalışma olmadığını bunların hepsini kapsayan bir metodoloji olduğu işlemek,
- B. Veri bankası ve VA arasındaki benzerlik ve farkların açıklanması sayesinde, VM'nin kurumsal altyapısını değerlendirmek ve kullanılan yöntemler ve kullanım alanlarını genel olarak irdelemek,
- C. Çalışmamızın temel odak konusu olan ve VM araçları olarak yaygın bir şekilde kullanılan Genetik Algoritmalar(GA) ve Yapay Sinir Ağları(YSA) yöntemlerini ayrıntılı incelemek,
- D. YSA eğitiminin performansını GA kullanarak arttırmaya çalışmak ve bu amaçla melez ve kademeli bir modelle öğrenme algoritması geliştirmek,

² HOOFFER, J.A., GEORGE, J.F., VALACICH, J.S.: a.g.e. s.11

E. Geliştirilen GA ve YSA melez modeli ile pratik bir işletme uygulamasına ışık tutmaktır.

Çalışmanın Yapısı

Bu çalışma, amaçlar kısmında özetlenen hedefleri göz önünde bulundurularak yapılandırılmıştır. Her bir bölümün mimarisi, amaçlar kısmında özetlenen belirli konuları işlemek üzere oluşturulmuştur:

İlk bölüm olan giriş bölümünde, böyle bir çalışmanın neden yapılması gerektiği anlatılarak, okuyuculara genel bilgiler sunulmuştur. Bu bölüm, okuyucuya çalışmanın bölümleri arasındaki ilişkinin ve genel yapının sunulmasını hedeflemiştir.

Birinci bölümde, VM için gerekli olan altyapı hazırlıklarından birisi, belki de en önemlisi olan veri alt yapısının hazırlanması ve VA hakkında bilgiler sunulmuştur. Büyük veri bankalarında, gündelik veri akışı döngüsünden organizasyonel ve de teknolojik olarak kurtulmadan, VM çalışması yapmak çok zor bir uğraş olacağı için VA oluşturulması, VM çalışmalarının başarı oranını doğrudan etkilemektedir. Örneğin herhangi bir şirket için, bir periyoda ait cironun bilinmesi önemli ve gereklidir. Ancak bir zaman dilimi içerisinde cironun seyrinin takip edilmesi ve daha da önemlisi buradan çıkarak geleceğe yönelik tahminler yapmak ve bu tahminler ışığında riskleri azaltmaya çalışarak potansiyel faydaları arttırmayı sağlamak çok daha önemlidir. Bu ve buna benzer çalışmalar, VM kapsamı içerisinde değerlendirilmektedir. Buradaki temel sorun, veri yönetim alt yapısının benzer ihtiyaçlara hitap edecek şekilde yapılandırılmasıdır.

VM metodolojisi izleyecek firmaların, veri yapılarını iki ana gruba ayırmaları gerekmektedir. Bunlardan birincisi gündelik(operasyonel) verilerin

toplanması, saklanması, kullanımı ve korunmasıdır. İkincisi ise bu gündelik veri yapısından elde edilen verilerin VA alanında depolanması ile VM'ne olarak sağlanmasıdır.

İkinci bölüm, VM'ni incelemek üzere oluşturulmuştur. Bu bölümde VM için, VA oluşturulması dışındaki alt yapı çalışmaları, VM'ni gerekli ve mümkün kılan sebepler, VM kullanım alanları, VM yöntemleri ve kurumlara sağlanacak yararlar işlenmiştir. Üzerinde önemle durulması gereken konulardan birisi, VM'nin istatistiki uygulama, bir yazılım veya bir teknik olarak değerlendirilmesinin yanlış olacağıdır. VM sözü geçen parçaların hepsine ihtiyaç duyan bir metodoloji olarak değerlendirilmelidir.

Üçüncü bölümde, çalışma için temel alınan VM araçlarından birisi olan GA işlenmiştir. GA, canlıların kalıtsal özelliklerinin kuşaktan kuşağa aktarımını ve evrim süreci içerisinde, canlı özelliklerinin değişime uğrayarak, gelişmenin sağlanmasını temel alan süreçlerdir. GA'nın yapıları, temel prensipleri, uygulama alanları, güçlü ve zayıf yönleri gibi konular işlenmiştir. Bu bölümde işlenen konular, çalışmanın model oluşturulması ve modelin uygulanması aşamalarında da kullanıldığı için bu bölüm oldukça ayrıntılı olarak ele alınmıştır. Özellikle GA'nın kullanımı ile ilgili örnekler sunulmuş ve bu örnekler yardımı ile algoritmanın daha iyi anlatılabilmesi hedeflenmiştir.

Dördüncü bölüm, çalışmada oluşturulan modelin diğer temel elemanı olan YSA'ları ayrıntılı olarak incelemektedir. Canlıların sinir fizyolojilerini, benzeşim yoluyla bilgisayar ortamına taşıyan YSA, VM alanında, mühendislikte, tıpta ve bir çok sosyal alanda yaygınlıkla kullanılan araçlardandır. YSA, esnek yapıları ve geliştirilen çok sayıda algoritma ile yaygın bir kullanım alanına sahiptir. Bu bölümde konu teorik ve pratik boyutlarıyla ele alınarak, okuyuculara YSA hakkında ayrıntılı bilgiler sunulmuştur.

Beşinci bölüm, GA ve YSA tekniklerinin, melez ve iki kademeli bir yapı ile modellenmesinin ve bu modelin uygulamada kullanılmasının işlendiği bölümdür. YSA için temel kavramlardan birisi “öğrenme” (Training) sürecidir. Var olan yapıdan öğrenme veya eğitim süreci sonrası gerekli bilgileri edinen ağlar sayesinde bu bilgiler tahmin aracı olarak kullanılabilir.

Literatürde çok sıklıkla karşılaşılan öğrenme yöntemlerinden birisi, geri yayılım ile hata kareleri toplamını minimuma indirmeye çalışan analitik yöntemdir. Bu yöntemin çok çeşitli varyasyonları ile karşılaşılabilmektedir.

Diğer popüler yöntem ise öğrenme sürecinin GA kullanarak gerçekleştirilmesidir. GA pratik ve hızlı sonuç elde edilen bir yöntem olmasına rağmen matematiksel olarak performansı garanti edilebilecek bir teknik değildir. Ancak pratik faydaları nedeniyle de GA göz ardı edilemez. Bu çalışmada izlenen yöntem, GA ile ön eğitime tutulan ağların, geri yayılım algoritmalarından birisi ile ikincil eğitime tabi tutulması ve daha hassas sonuçların elde edilmesidir. Böylelikle iki farklı yaklaşımın, güçlü olduğu yanların bir model içerisinde birleştirilmesi amaçlanmıştır.

Bu bölümün diğer önemli başlığı ise oluşturulan modelin pratik bir işletme uygulamasında kullanılmasıdır. Buradaki işletme problemi, müşterilerin firmaya olan katkılarının geliştirilen modelle öğrenilmesi ve bu sayede potansiyel müşterilere yaklaşım ile ilgili politikaların geliştirilmesine destek verilmesidir.

Sonuç bölümünde ise oluşturulan modelin performansı ile ilgili sonuçların sunulduğu, uygulamanın değerlendirildiği ve benzer konularda araştırma yapacak okuyuculara önerilerde bulunduğu bölümdür.

Konunun Dünya Literatüründeki Yeri

VM konusu, 20.yy'ın sonlarında popüler olmaya başlamıştır. Popülerliği devam eden bir konu olduğu için, konu ile ilgili çok sayıda kitap, makale, araştırma ve uygulama mevcuttur. İncelenen kaynakların çoğu VM ile ilgili olarak genel yaklaşımları sunarak pratik çözümler hedeflemektedir. VM konu başlığı akademik çalışmalarla birlikte endüstri merkezli olarak da ele alınmaktadır. Bununla birlikte, VM aracı olarak kullanılan yöntemler ve metodolojik yaklaşımlar, VM için gerekli altyapı çalışmaları gibi konular akademisyenlerin daha sıklıkla inceledikleri alanlardır.

Bu yöntemler arasında yer alan GA ve YSA ile ilgili olarak, çok sayıda yayınlara karşılaşılmıştır. Bu durum da doğal karşılanmalıdır. Çok sayıda bilimsel alanın kullanımına açık olan ve çok sayıda bilim alanının desteğini gerektiren bu ve benzeri teknikler ayrıntılarıyla incelenmelidirler.

Önceleri kuşku ile bakılan ve bilim kurgu penceresinden izlenen GA ve YSA çalışmaları, pratik başarıları, teorik temellerinin derinleştirilmesi ve en önemlisi de geleneksel yöntemlere göre bazı bariz üstünlükleri nedeniyle kabul görmektedirler.

YSA ile ilgili en önemli eleştiri ise, mekanizmanın içerisinde gerçekleşenlerin, uygulayıcılar tarafından bilinmemesi ve buna bağlı olarak da güvenilirliğinin bilimsel anlamda ispatlanmasındaki zorluklardır. Örneğin basit doğrusal regresyon modellerinde bir fonksiyon elde edilebilir. Bu fonksiyona ait katsayılar her bir bağımsız değişkende gözlenecek bir birimlik değişimin, bağımlı değişkene etkisini gösterir. Ayrıca regresyon modellerinin tutarlılığını ve istatistiksel anlamlılığını test eden yöntemler mevcuttur (t, F testi vb.). Ancak YSA için bu kadar kesin yargılar üretilemeyebilir.

Çalışmanın Kapsamı

Bu çalışmada, VA, VM ve VM için kullanılan yöntemler incelenmiş ve bu yöntemler içerisinde GA ve YSA teknikleri, ayrıntılı bir şekilde ele alınmıştır. Ancak hem GA hem de YSA üzerine yapılmış çok sayıda çalışma olması nedenleriyle, konu oldukça geniş bir alana yayılmıştır.

Bu çalışmada temel sayılabilecek ve genel kabul görmüş GA yaklaşımı benimsenmiştir. Bir çok varyantı olan bu yöntem için seçilen yapı, başka araştırmacılar tarafından değiştirilerek, farklı ve performansı daha da yüksek olabilecek modellerin oluşturulmasına olanak sağlayabilir.

YSA, için alternatifler daha da fazladır. Karmaşıklık, topolojik yapı, öğrenme şekli (denetimli/denetimsiz öğrenme), öğrenme mekanizması, kullanım alanı gibi değişkenler gözönünde bulundurulduğunda uygulamada kullanılabilecek çok sayıda benzer yöntemle rastlanmıştır.

Bu çalışma kapsamında, yaygın kullanım alanı olan ve görece olarak da basit sayılabilecek, denetimli öğrenmeyi temel alan çok katmanlı, ileri beslemeli (asimetrik), geri yayımlı ağ modeli ile delta kuralı ile eğitimi baz alan metod kullanılmıştır.

Çalışmanın uygulama kısmında temel alınan amaç, geliştirilen modelin tahmin aracı olarak kullanılmasıdır. Karar destek sistemi bileşeni olarak kullanılan en önemli çalışmalardan birisi, tahmindir. Tam bilginin olmadığı her durumda, karar vericiler, belirli şekillerde tahminlerle yol almak zorundadırlar. Geleceğe yönelik, belirsizlik ve risk altındaki durumlarda karar verebilmek için sezgisel ve sayısal metotlar kullanılarak objektif kriterlerle karar verme süreci kurumlara önemli faydalar sağlamaktadır. Alınacak kararların stratejik öneminin artmasına paralel olarak ihtiyaç duyulan bilgi, tecrübe ve yöntemsel yaklaşım daha da önemli bir hal alır.

VM'nin de temel amaçlarından birisi, var olan veri ve bilgileri işleyerek, rafine bilgiye ulaşmaktır. Bu da çoğunlukla çeşitli varsayımların kabulünü ve tahmin sürecini gerekli kılar.



BÖLÜM 1. VERİ AMBARLARI (VA)

Giriş bölümünde de kısaca açıklandığı gibi VA çok büyük hacimlerdeki veri kütlelerinin etkin kullanımına olanak vermek amacı ile sistematik bir mimari ile yapılandırılmalarıdır.

Dünyanın en büyük veri bankalarından birisine sahip olan Amerikan merkezli Wall-Mart mağazalar zincirinin, 1997 yılı değerleriyle, yönetmek zorunda olduğu veri büyüklüğü yaklaşık olarak 24 tetra byte yani 24,000,000,000 Kb'dir.³ Bu kadar büyük bir veri hacminde, marketler zincirine ait herhangi bir mağazada, herhangi bir raftaki, herhangi bir ürünün ayın hangi periyodunda, hangi gelir düzeyindeki tüketiciler tarafından hangi miktarlarda alındığının sorgulanması başlı başına bir problemdir.

Bu tür firmalar veri kaynaklarını temel olarak iki amaçla kullanırlar:

- i. Operasyonel amaçlar: Bu tür amaçlara verilebilecek en yaygın örnekler;
 - Bir ürünün stok değerinin belirli bir seviyeye düştüğünde siparişinin verilmesi,

³ ANONİM: "Wal-Mart Deploys Data Mining Software Into Its Production Support Environment", 24 Mayıs 1997, [Çevrimiçi], <http://www.walmartstores.com/wmstore/wmstores/>, 15 Mayıs 2003

- Satılan bir ürünün fiyatının otomatik olarak bar kod okuyucusundan gelen sinyal ve veri tabanındaki fiyat verilerinin karşılaştırılmasıyla kasiyer tarafından elde edilmesi.,
- ii. Stratejik amaçlı kullanımlar: Bu tür uygulamalar genelde büyük veri tabanlarında stratejik önemi olan kararların alınmasında kullanılırlar. Wall-Mart firması ile ilgili olarak yukarıda verilen örnek bu tür uygulamaların zorluklarını göstermek amacı ile sunulmuştur.

Her iki durumda da amaçlar ayrı olduğu gibi araçlar da ayrıştırılabilir. Örneğin birinci grup işlemlerde sürekli olarak basit sayılabilecek sorgular kullanıldığı için normal formlardan en az üçüne sahip ilişkisel veri tabanları tercih edilir. İkinci grup uygulamalarda daha sonra sebepleri açıklanacağı üzere yüksek dereceli normal formlar yerine sorgulamayı kolaylaştırma ve gizli veya çok açık olmayan yapıların (pattern) elde edilmesini sağlayan bir yapı tercih edilir.

Bir ürüne ait fiyat bilgisi gayet net ve yoruma açık olmayan bir bilgidir. Ancak tüketicilerin davranışları ile ilgili ipuçlarının yakalanması ve buna uygun stratejilerin belirlenmesi çok daha karmaşık bir durumdur. Bu durum VM'nin önemini ortaya koyarken VA kullanımını da ön görür.

Bu çalışmamızda, sadece konunun anlaşılması amacı ile ilgili olarak ilişkisel veri tabanlarından bahsedilecektir. Derinlemesine bilgi almak isteyen okuyucular, veri tabanları ile ilgili herhangi bir kaynak kullanarak daha ayrıntılı bilgi edinebilirler.

1.1 Tanım

VA, William Inmon tarafından 1990 yılında ortaya atılan bir tanımdır. “ VA, operasyonel veri tabanından, içsel ve dışsal kaynaklardan gelen, entegre edilmiş, temiz, arşivlenmiş, büyük hacimli verilerin, yönetim tarafından karar vermeyi destekleyecek şekilde kullanımına olanak sağlayacak şekilde derlendiği depolama alanlarıdır.”⁴

1.2 VA'ların Karakteristikleri

VA'ların dört temel özellikleri vardır, özne odaklılık, bütünlüklük yapı, zamana bağıllık, kalıcılık ve tutarlılık.⁵

Özne-odaklılık(Subject oriented): VA genellikle, kuruluşun ilişki içinde bulunduğu özneler veya parametrelerle ilgilidir. Tüketici, sipariş, ürünler gibi temel başlıklar VA için önemli parametrelerdir. Operasyonel veri tabanları ise buna karşılık olarak daha iş süreçleri odaklıdır. İş süreci odaklı veri tabanları söz konusu olan süreci yönetmekle yükümlüdürler. Otomatik para çekme makineleri ile yapılan bankacılık işlemleri buna en güzel örneklerden birisidir.

Bütünlüklük Yapı (Ingerated): Birden çok sayıda olabilen operasyonel veri kaynakları ve diğer dışsal kaynaklar birleşerek VA için bütünlüklük bir zemin oluştururlar. Bu çok farklı kaynaklı veriler, birbirleri ile uyumlu hale getirilerek tek bir amaca, karar desteğine

⁴ MANNINO, M. V.: “Database Application Development & Design”, New York, NY: McGraw- Hill, 2001, s.455.

⁵ CABENA, P., HADJINIAN, P., STADLER R., VERHEES, J., ZANASI, A.: “Discovering Data Mining From Concept to Implementation”, New Jersey, NY: Prentice Hall., 1997, s.19.

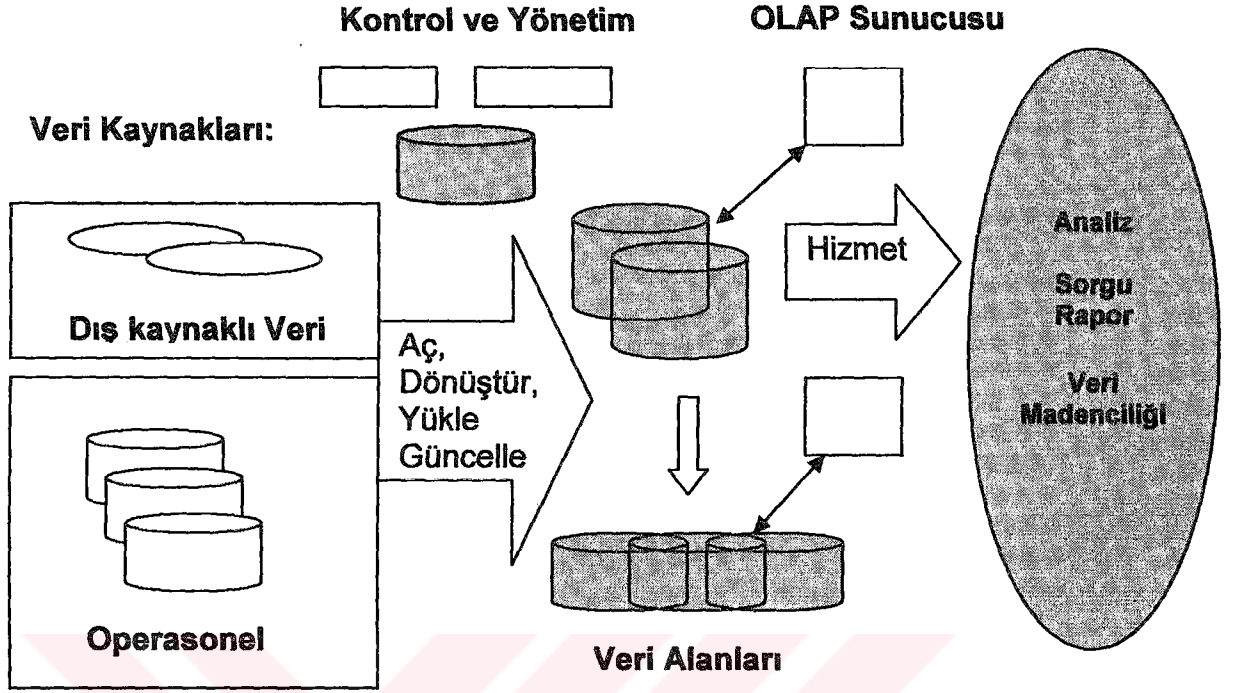
yönelik olarak konsolide edilirler. Burada veri türlerinin uyumu, birlikte çalışabilmeleri, ölçüm skalalarının uygunluğu gibi sorunların çözümü için değişik metot ve araçlar kullanılmalıdır.

Zamana Bağlılık (Time Variant): VA'lar tarihsel verileri işlemek için zaman parametresini kullanırlar. Zaman boyutu trend gibi zaman serileri analizlerinin yapılabilmesi için kritiktir. VA bir video filmine benzetilirse, çok sayıdaki operasyonel veri tabanı çıktılarından oluşan film karelerinden meydana gelmişlerdir.

Kalıcılık (Non-volatile): VA içine giren yeni veriler, başka verilerin güncellenmesi sonucunu doğurmazlar. Arşivde kronolojik olarak yerlerini alırlar. Bu sırada daha eski tarihli veriler de aynen muhafaza edilir. Operasyonel veri tabanındaki veriler güncellenip son şekillerini almadan VA içerisinde yer almazlar.

Sürekli olarak çok sayıda veri tabanından gelen ve zaman parametresine bağlı olarak saklanan veriler hem VA yönetimini hem de buradan anlamlı bilgilerin elde edilmesini zorlaştırırlar. Bu zor uğraşlar karşılığında kurumsal bir hafıza sağlanmış olur. Bu hafıza kurumun hayati kararlarının alınmasında dahi temel destek elemanlardan birisi olarak görev alır. Akıllı ve öğrenen organizasyon çalışmalarının temel taşlarından birisi olan VA'ların gelecekte önemlerini daha da artacaktır.

Şekil 1.1'den de anlaşılacağı üzere VA, çeşitli veri kaynaklarından gelen verileri açmak, dönüştürmek, yüklemek ve güncellemek yolu ile kullanır. Bu işlemler sonunda ise hazırlanan veriler analiz yöntemleri, sorgu-rapor süreçleri, veri madenciliği çalışmaları gibi çıktı yönelimli çalışmalarda kullanılır. OLAP(On-line Analytical Processing), Data Mart gibi alanlar ise çıktı bilgilere ulaşılabilecek diğer alanlardır.

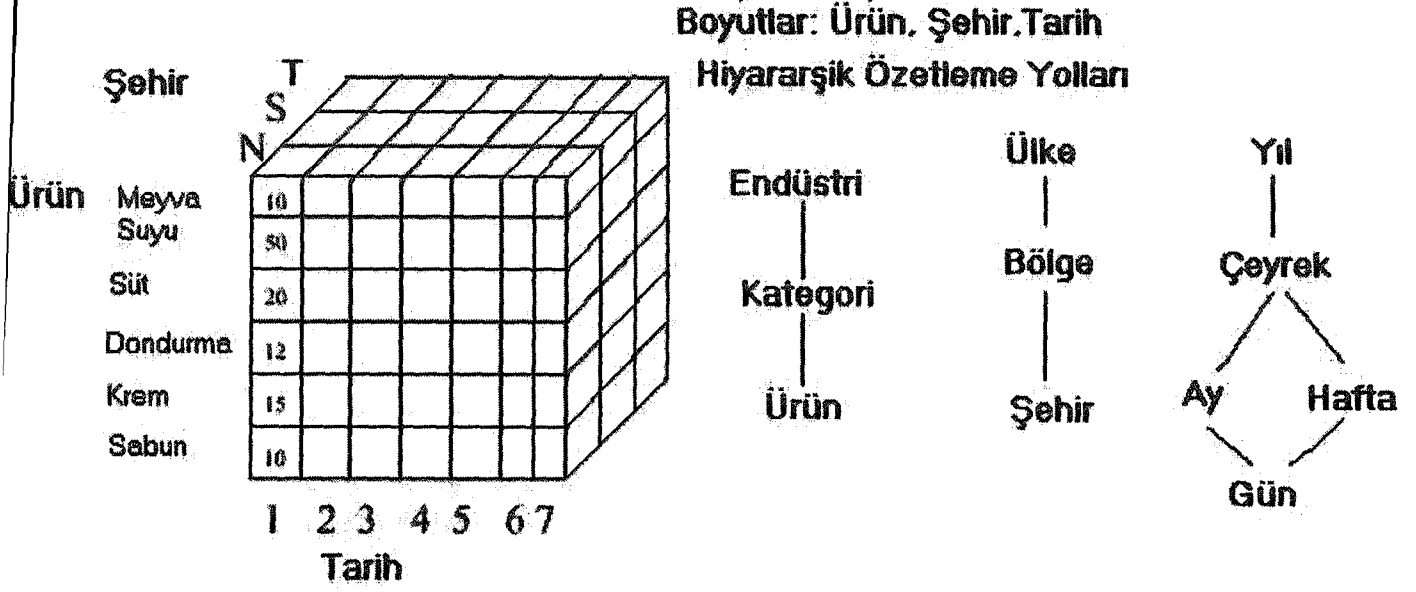


Şekil 1.1 VA Mimarisi ⁶

1.3 Veri Küpleri

Veri küpü olarak adlandırılabilen n boyutlu matrisler VA teknolojilerinin kullandıkları önemli araçlardan bir grubu temsil ederler. Bu matrislerin de kolay olarak canlandırılabilirleri durum $n=3$ olduğu haldir yani kübik üç boyutlu hal. Klasik kayıt mantığındaki satır ve sütunlardan oluşan iki boyutlu yapıya üçüncü bir boyut ilave edildiğinde veri küpü elde edilmiş olur. Şekil 1.2 veri küplerinde bilgi depolanmasını basitçe göstermektedir.

⁶ CHAUDHURI, S. AND DAYAL, U.: "An Overview of Data Warehousing and OLAP Technology", *ACM Sigmod Record*, Mart 1997 s.518.



Şekil 1.2 Veri Küpü⁷

1.4 Operasyonel Veri Tabanı ile VA'nın Karşılaştırılması:

Tablo 1.1 iki veri modelinin özelliklerini karşılaştırmalı olarak özetlemektedir.

Tablo 1.1 VA ve Operasyonel Veri Tabanı Kıyaslaması⁸

Özellik:	Operasyonel Veri Tabanı	VA
Güncellik	Güncel	Tarihsel
Ayrıntı Seviyesi	Tekil	Özet
Odak	Proses odaklı	Özne odaklı
Proses Tabi Tutulan Kayıtlar	Birkaç	Binlerce
Normalizasyon Düzeyi	Yüksek	De-normalize Edilmiş
Veri Modeli	İlişkisel	Çok boyutlu

⁷ CHAUDHURI, S., DAYAL, U.: a.g.e., s.520.

⁸ MANNINO, M. V. : a.g.e., s.456.

1.5 Normalizasyon

Normalizasyon, günümüzde operasyonel veri tabanlarında en yaygın kullanılan ilişkisel veri tabanlarının optimizasyonunda kullanılan önemli bir araçtır. Oluşturulan kurumsal veri bankasındaki verilerin etkin bir şekilde kullanımı, veri tutarlılığının ve bütünlüğünün sağlanması ancak normalizasyon ile sağlanabilir. Yaygın olarak kullanılan 5 normal form vardır. Kurumlar, pratik sebeplerden dolayı, bu beş norm seviyesinin tamamını gerçekleştirmeyebilirler. Ayrıca VA'lar bazı geçerli sebeplerden dolayı genellikle çok yüksek seviyeli normal formlara sahip değildirler. Bu durumu ve sebeplerini anlatabilmek için her bir form kuralı ve ilişkisel veri tabanlarında önledikleri anormallikler açıklanarak VA'lar ile karşılaştırma yapılmalıdır:

1.5.1 Birinci Normal Form:

Satır ve sütunlardan oluşan tasarımda iki satır tamamen aynı olamaz.

Tablo 1.2 Birinci Normal Forma Aykırı Tabloya Örnek

Çalışan İsmi	Semt	Eğitim	Ünvan
Ali Tanır	Erenköy	Lise	Üretim Md.
Veli Tanırmaz	Bakırköy	Üniversite	Satın alma Md.
Ali Tanır	Erenköy	Lise	Üretim Md.
Ayşe Tanıyabilir	Bostancı	Lise	Muhasebe Md.

Yukarıdaki tabloda birinci ve üçüncü satırlar tamamen aynıdır, ve bu durum yapılacak sorgularda karışıklığa yol açacaktır. Örneğin bilgi sistemi dört çalışan olduğu raporunu üretirken gerçekte sadece üç çalışan vardır.

1.5.2 İkinci Normal Form:

İkinci normal formun sağlanabilmesi için, öncelikle birinci normal form sağlanmalıdır ve bir kayıt için, tüm anahtar olmayan özellikler, anahtar özelliklerin tamamına bağımlı olmalıdır.

Burada *anahtar özellik veya özellikler(AÖ)*, bir kaydı tanımlamak için gerekli minimum özellikler olarak özetlenebilir. Bir anahtar özellik ailesi bir ve yalnız bir kayda aittir. Öğrenci numarası okul sistemi için anahtar özellik sayılabilir, çünkü boş bırakılamaz ve her öğrencinin tek ve başkasında bulunmayan bir numarası vardır. Anahtar özellik kompozit yani birden çok özellikten oluşmuş da olabilir. Örneğin bir satın alma kaydı için ürünün kodu ve satın alınan firmanın kodu birlikte anahtar özellik görevi görürler. Her bir özellik tek başına yeterli değildir, çünkü bir satıcı birden fazla ürün satabilir ve ürün birden fazla satıcı tarafından satılabilir.

Tablo 1.3 ikinci normal forma aykırı bir duruma örnek oluşturmak amacı ile verilmiştir.

Tablo 1.3 İkinci Normal Forma Aykırı Tabloya Örnek

Satıcı Kodu(AÖ)	Ürün Kodu(AÖ)	Miktar	Satıcının semti
1010	2255	100	Erenköy
1011	2350	450	Kadıköy
1011	4520	50	Şişli
1013	4781	10	Taksim

Satıcının Semti Anahtar özellik ailesinin sadece bir ferdi ile ilgilidir

Bu tabloda amaç satın alma prosesini yönetmektir. Satıcı bilgileri ayrı bir tablo olan satıcı bilgileri tablosunda tutulmalıdır. 1010 nolu satıcı 120 kez satış yapmışsa o satıcının semti 120 kez girilmek zorunda kalınacaktır. Diğer

bir olası problem ise veri kaybıdır. Tablonun sadece dört kayıttan oluştuğu var sayılırsa, 1010 nolu satıcı ve 2255 nolu ürünün olduğu satış satırı bir sebeple silinirse satıcının bulunduğu semte ait bilgi de silinmiş olacaktır ki bu bilgi gelecekte kullanılabilir.

1.5.3 Üçüncü Normal Form:

Üçüncü normal formun sağlanması için, ilgili tablo ikinci normal formu sağlamalıdır ve anahtar olmayan bir özellik anahtar olmayan başka bir özelliğe bağlı olmamalıdır.

Tablo 1.4 Üçüncü normal forma aykırı bir duruma örnek oluşturmak amacı ile verilmiştir.

Tablo 1.4 Üçüncü Normal Forma Aykırı Tabloya Örnek

Ürün Kodu(AÖ)	İsmi	Satış Sorumlusu	İş Deneyimi(Yıl)
2255	Enjektör	Ali Tanır	2
2350	Valf	Veli Tanımaz	1
4520	Silindir Kapağı	Ayşe Tanıyabilir	4
4781	Valf	Ali Tanır	2

İş Deneyimi, Anahtar olmayan bir sütuna bağımlıdır

Buradaki sorunlar ikinci normal form ile ilgili olanlara benzerdir. Veri bütünlüğünün sağlanması ve tablo yapılarının geliştirilmesi zordur. İş deneyimi personel dosyasının bir parçasıdır, ürünle ilgisi yoktur. İkinci formdaki problemler burada da geçerlidir. Bunlara ilave başka bir sorun ise, veri giren kişi yanlışlıkla son satırda iki değeri yerine üç değeri girerse veri

tutarlılığı da ortadan kalkacaktır, çünkü aynı kişi iki ayrı deneyim seviyesinde olacaktır.

1.5.4 Dördüncü Normal Form:

Dördüncü normal formun sağlanması için, ilgili tablo üçüncü normal formu sağlamalıdır ve tablo, bir başka özellik (attribute) ilgili olan bağımsız ve çok değerli olabilen birden çok özelliğe sahip olmamalıdır.

Tablo 1.5 Dördüncü normal forma aykırı bir duruma örnek oluşturmak amacı ile verilmiştir.

Tablo 1.5 Dördüncü Normal Forma Aykırı Tabloya Örnek⁹

Kitap Başlığı(AÖ)	Konu (AÖ)	Dil (AÖ)
MacBeth	Politika	İngilizce
MacBeth	Aşk	Fransızca
Prolog	Bilgisayar	İngilizce
Prolog	Mantık	İtalyanca
Prolog	Mantık	İngilizce

Burada üç özellik birden, bir satırı diğerlerinden ayıracak şekilde tanımladıkları için her üçü birlikte anahtar görevini üstlenmişlerdir. Burada da gereksiz veri saklanması sorunu ile birlikte tablo revize sorunları yaşanabilmektedir. Çünkü MacBeth hem politika konusu hem de aşk konusunda yer alabilir ve iki ayrı dilde olabilir. İki tanım için üç satır kullanılmıştır. Bunun yerine tablonun ikiye bölünmesi dördüncü normal forma uygunluğu sağlar:

⁹ KOSTER, A.: "Supplementary Material For IDS 480/686 ,Database Management Systems", Montezuma Publishing, San Diego Ca, 2002, s.105.

Tablo 1.6 Dördüncü Normal Forma Uygun İlk Tablo¹⁰

Kitap Başlığı(AÖ)	Konu (AÖ)
MacBeth	Politika
MacBeth	Aşk
Prolog	Bilgisayar
Prolog	Mantık

Tablo 1.6, kitap başlığı ve konu bilgilerini barındırmaktadır. Bu tablonun kitabın hangi dilde yazıldığı ile ilgisi bulunmamaktadır. Benzer şekilde Tablo 1.7’de ise sadece kitap başlığı ile kitabın çevrildiği diller bilgileri yer almaktadır. Eğer uygulama gereği olarak bu üç bilgi aynı anda sorgulanmak istenirse, tablolar arası ilişki kurularak, kullanılan sorgulama dili ile iki tablodan gerekli bilgiler çağrılarak sorgu gerçekleştirilebilir.

Tablo 1.7 Dördüncü Normal Forma Uygun İkinci Tablo¹¹

Kitap Başlığı(AÖ)	Dil (AÖ)
MacBeth	İngilizce
MacBeth	Fransızca
Prolog	İngilizce
Prolog	İtalyanca

Her iki tablo da bir satır eksikle gerekli tüm veriyi saklayabilmektedir. Burada sağlanan bir diğer fayda ise aynı verileri daha basit olan bir yerine iki tabloda göstermektir. Bu da, veri tabanında ki tasarım değişikliklerini kolaylaştırmaktadır.

¹⁰ KOSTER A.: a.g.e., s.106.

¹¹ KOSTER A: a.g.e., s.106.

Tablo 1.6 Dördüncü Normal Forma Uygun İlk Tablo¹⁰

Kitap Başlığı(AÖ)	Konu (AÖ)
MacBeth	Politika
MacBeth	Aşk
Prolog	Bilgisayar
Prolog	Mantık

Tablo 1.6, kitap başlığı ve konu bilgilerini barındırmaktadır. Bu tablonun kitabın hangi dilde yazıldığı ile ilgisi bulunmamaktadır. Benzer şekilde Tablo 1.7’de ise sadece kitap başlığı ile kitabın çevrildiği diller bilgileri yer almaktadır. Eğer uygulama gereği olarak bu üç bilgi aynı anda sorgulanmak istenirse, tablolar arası ilişki kurularak, kullanılan sorgulama dili ile iki tablodan gerekli bilgiler çağrılarak sorgu gerçekleştirilebilir.

Tablo 1.7 Dördüncü Normal Forma Uygun İkinci Tablo¹¹

Kitap Başlığı(AÖ)	Dil (AÖ)
MacBeth	İngilizce
MacBeth	Fransızca
Prolog	İngilizce
Prolog	İtalyanca

Her iki tablo da bir satır eksikle gerekli tüm veriyi saklayabilmektedir. Burada sağlanan bir diğer fayda ise aynı verileri daha basit olan bir yerine iki tabloda göstermektir. Bu da, veri tabanında ki tasarım değişikliklerini kolaylaştırmaktadır.

¹⁰ KOSTER A.: a.g.e., s.106.

¹¹ KOSTER A: a.g.e., s.106.

1.5.5 Beşinci Normal Form:

Beşinci normal formun sağlanması için, ilgili tablo dördüncü normal formu sağlamalıdır ve tablo daha küçük tablolara başka özelliklerle de olsa veri kaybetmeden ayrılmalıdır.

Tablo 1.8 beşinci normal forma aykırı bir duruma örnek oluşturmak amacı ile verilmiştir.

Tablo 1.8 Beşinci Normal Forma Aykırı Tabloya Örnek

Satış Elemanı(AÖ)	Marka (AÖ)	Üretim Yılı (AÖ)
Ali Tanır	A	2000
Ali Tanır	A	2001
Ali Tanır	B	2000
Veli Tanımaz	A	1999
Veli Tanımaz	B	2000
Ayşe Tanıyabilir	B	1999
Ayşe Tanıyabilir	A	2001

Burada dördüncü normal form sağlanmaktadır ama tabloda çok sayıda tekrar vardır. Bu nedenle tablo daha basit tablolara veri kaybetmeden dönüştürülebilir.

1.5.6 Normal Formların Kullanımı

Yukarıda sıralanan ve ayrıntıları anlatılan tüm normal formlar uygulanmalıdır şeklinde genel kabul görmüş bir kural yoktur. Uygulama şekli, öncelikler, performans gerekleri ve veri tabanının revize edilme sıklığı uygun normal form seviyesinin seçilmesine rehberlik etmelidir.

1.6 VA'larda Yüksek Seviyeli Normal Formların Sonuçları

Yüksek normal forma sahip veri mimarilerinin revize edilmeleri genelde daha kolaydır. Çünkü olabildiğince modüler alt gruplara ayrılmışlardır. Banka işlemleri gibi sürekli veri girişi gerektiren uygulamalar için de uygundur. Ancak çok sayıda tablodan veri alan sorguların bulunduğu, sürekli sadece veri okumaya dayanan ve veri girişinin nispeten az olduğu Veri madenciliği gibi uygulamalarda VA'lar ve nispi olarak düşük normal form seviyeleri tercih edilebilmektedir.¹²

1.7 VA Organize Ederken Kontrol Listelerinin Kullanımı

VA mimarisinin oluşturulması sırasında çeşitli ihtiyaçlardan kaynaklanan kurumsal isteklerin değerlendirilmesi gereklidir. Bu değerlendirmeye yol göstermesi açısından çeşitli kontrol listeleri kullanılabilir. Bunlara örnek olarak aşağıdaki liste sunulmuştur.

Tablo 1.9 Kontrol Listesi¹³

İşletme İhtiyaçları	Evet	Hayır
Neden VA tasarımı yapıldığı ile ilgili açık bir fikir var mı?		
İsteklere ulaşıldığı zaman farkına varılabilecek mi?		
İlk aşamada yapılması gerekenler biliniyor mu?		
VA tasarımının, diğer tüm kurumsal faaliyetlerle ilişkisi anlaşılmış durumda mı?		
Tanımlı ve aktif yönetsel destek var mı?		

¹² KOSTER A: a.g.e., s.106.

¹³ MOORE, A. AND WELLS, D.: "Discussion the Questions to Ask When Performing a High-level Post-Implementation Data Warehouse Audit", **The Data Administration Newsletter**, Haziran 2000

(Önceki Sayfadan Devam)		
VA kullanıcıları ve sistemi paylaşılanlar tanımlı mı?		
VA'ya ulaşımın nasıl sağlanacağı açıkça tanımlı mı?		
Bilginin nasıl kullanılacağı tanımlı mı?		
Bilgi Mimarisi	Evet	Hayır
Veri kalitesi parametreleri belirli mi?		
Bu parametreler nasıl sağlanabilir?		
Verilerden doğru bilgiler türetililecek mi?		
Siteme yeni giren veriler VA ile entegre hale getirilebilecek mi?		
Kaynak sistemlerdeki değişiklikler izlenerek bunlara uyum sağlanabilecek mi?		
Teknik Mimari	Evet	Hayır
Mevcut teknoloji Yeterli mi?		
Performansı istekleri karşılayabilir mi?		
Kullanılan teknoloji gelişmeye ve hacim büyümesine izin verebilecek mi?		
Sistemin kullanım süresi boyunca etkin işletim için gerekli kaynaklar var mı?		
Fayda maliyet oranı kabul edilebilir bir düzeyde mi?		
Kullanılacak araçlar, kullanıcıların ihtiyaç ve becerilerine uygun mu?		
Sisteme, bütün kullanıcıların ulaşması ve etkin kullanımı sağlanabilecek mi?		
Organizasyon	Evet	Hayır
VA'nın bir sorumlusu var mı?		
Sorumluların, organizasyon içerisinde yeri belirli mi?		
VA'nın bilgi işlem yapısı içerisinde ki yeri nedir?		
VA geliştirme, yenileme, bakım hizmetleri için yeterli personel var mı?		
Proje Planlama ve Metodoloji	Evet	Hayır

(Önceki Sayfadan Devam)		
Proje yönetiminde karşılaşılabilecek zorluklar belirlendi mi?		
Proje, belirlenen zaman ve kaynaklarla yürütülebilecek mi?		
Proje işletme ihtiyaçlarını tam olarak karşılayabilecek mi?		
Proje faaliyetlerin gidişatı sürekli olarak denetlenebilecek mi?		
IT bölümü ile fonksiyonel bölümlerin işbirliği sağlanabilecek mi?		

Yukarıda sadece örnek olarak verilen kontrol listesi, çok farklı kurumsal ihtiyaçlar ve şartlar göz önünde bulundurularak değişik şekillerde yaratılabilir ve kullanılabilir. Burada ki temel prensip sistematik bir şekilde ihtiyaçların, kaynakların, yöntemlerin ve sorumlulukların izlenmesi ve gerekli değişikliklerin hızlı bir şekilde yapılması ile VA'nın kontrollü bir şekilde yaratılması ve kullanılmasının sağlanmasıdır.

1.8 VA Oluşturma Aşamaları

VA oluşturulması ile ilgili aşamalar, genel proje yürütme mantığını ve metodolojisini temel alır. Faaliyetler genel olarak dokuz aşamada toplanabilir, bunlar;¹⁴

1. Planlama,
2. Veri gereklerinin ve modelinin elde edilmesi,
3. Fiziksel veri tabanı tasarımı ve geliştirilmesi,
4. Veri kaynaklarının belirlenmesi, entegrasyonu ve haritalanması,
5. Gerçek verilerin yerleştirilmesi,

¹⁴ ANONİM: "Datawarehousing-An Approach Document", Goldstone Technologies, 2000, s.6. [Çevrimiçi]<<http://www.goldstonetech.com/downloads/Datawarehousing.pdf>>, 10 Şubat, 2003

6. Otomatik veri yönetim prosesinin oluşturulması,
7. Raporların oluşturulması,
8. Veri geçerliliğinin sağlanması ve test edilmesi,
9. Uygulamaya geçilmesi.

1.9 VA Mimari

VA oluşturmanın en önemli aşamalarından birisi planlama ve ihtiyaçlara en uygun mimari yapının belirlenmesi aşamasıdır. Her projede olduğu gibi VA geliştirme projelerinde de planlama aşamasında yapılabilecek yanlışlar uygulama aşamasında önemli problemlere yol açabilmektedir. VA oluştururken kullanılabilecek çok sayıda mimari yapı mevcuttur. Bunlardan bazıları:

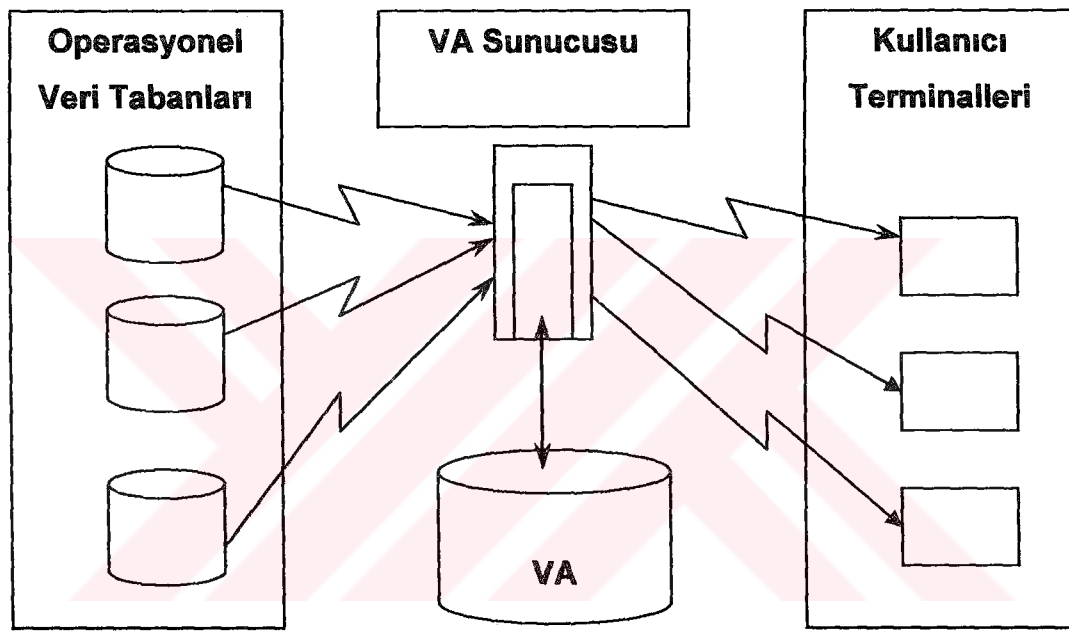
- İki kademeli mimari,
- Üç kademeli mimari,
- Aşağıdan yukarı doğru mimaridir.

Kurumların içinde bulundukları şartlar, ihtiyaçları ve öncelikleri, hangi yapının seçileceğine rehberlik etmelidirler. Veri yönetimi açısından ele alınabilecek temel kriterler: Veri hacmi, verilerin özellikleri, veriye ihtiyaç duyulma sıklığı, veriye ulaşım hızı istekleri, kullanıcı sayısı ve benzeri temel teknik özelliklerdir.

Kurumsal kaynakları en uygun şekilde kullanacak olan yapının seçilmesinin ardından, o yapının sistematik bir şekilde uygulanması aşaması gelmektedir.

1.9.1 İki Kademeli Mimari

İki kademeli yapı genel olarak, operasyonel veri tabanında saklanan verilerin VA yapısına uygun hale getirilmesi sürecinin ardından VA'da saklanmasını ve ardından kullanıcıların ihtiyaç duydukları anda VA sunucusu aracılığı ile veriye ulaşabilmeleri üzerine kurulmaktadır. Bu yapı Şekil 1.3'te şematik olarak sunulmuştur:



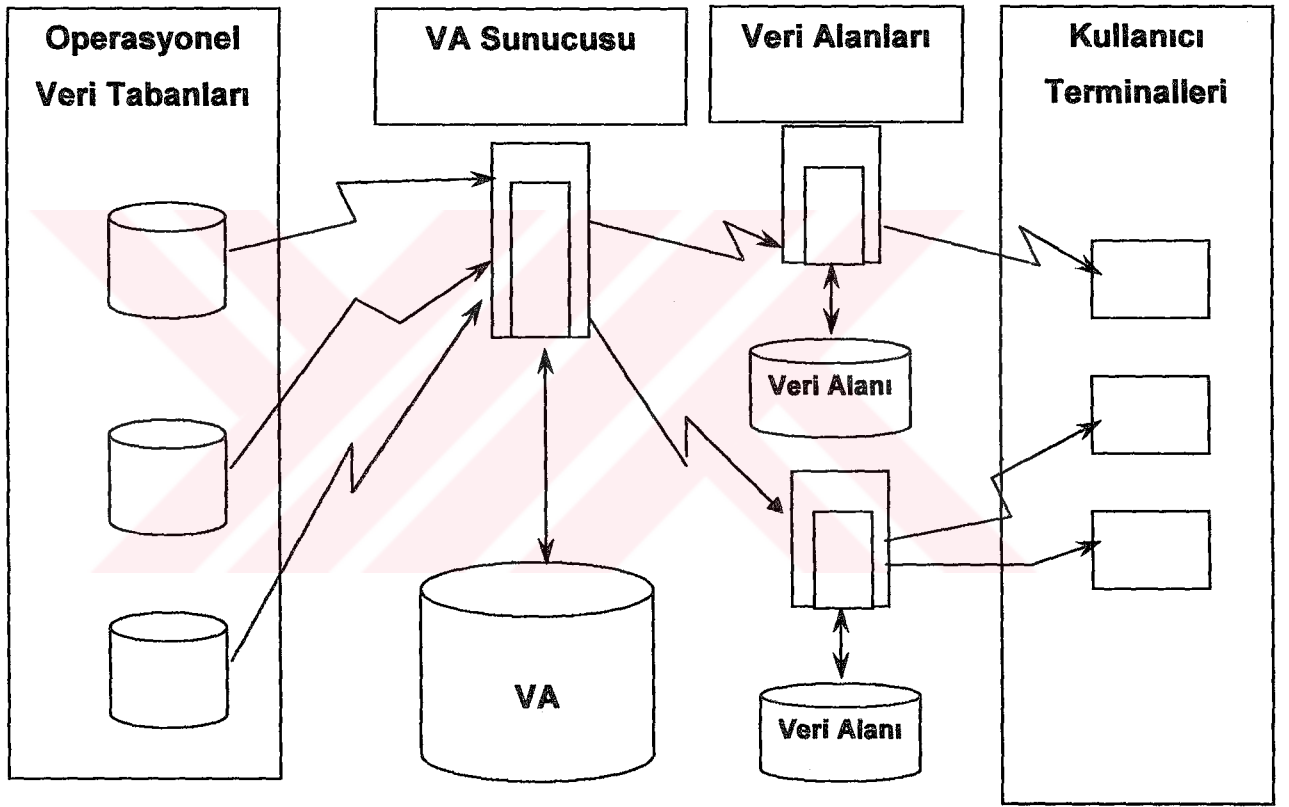
Şekil 1.3 İki Kademeli Mimari¹⁵

İki kademeli yapı, veri hacimlerinin çok büyük olduğu ve veriye ulaşma ihtiyaçlarının sık olduğu durumlarda problemler yaratmaktadır. Bu problemleri ortadan kaldırmaya yönelik olarak üç kademeli yapı kullanılabilir.

¹⁵ MANNINO M. V.: a.g.e., s.457.

1.9.2 Üç Kademeli Mimari

İki kademeli yapının gereği olarak kullanıcılar VA'nın tümü içerisinde bilgi ihtiyaçlarına ulaşmaya çalışmak zorundadırlar. Ancak veri hacimlerinin çok büyük olduğu ve veriye ulaşma ihtiyaçlarının sık olduğu durumlarda kullanıcılar VA'nın tamamı yerine kendileri ile ilgili küçük bir kısmı ile bağlantı kurmak istediklerinde üç kademeli yapıda oluşturulan veri alanları onlara hizmet eder. Böylelikle kaynakların verimli kullanımı sağlanmış olur.



Şekil 1.4 Üç Kademeli Mimari¹⁶

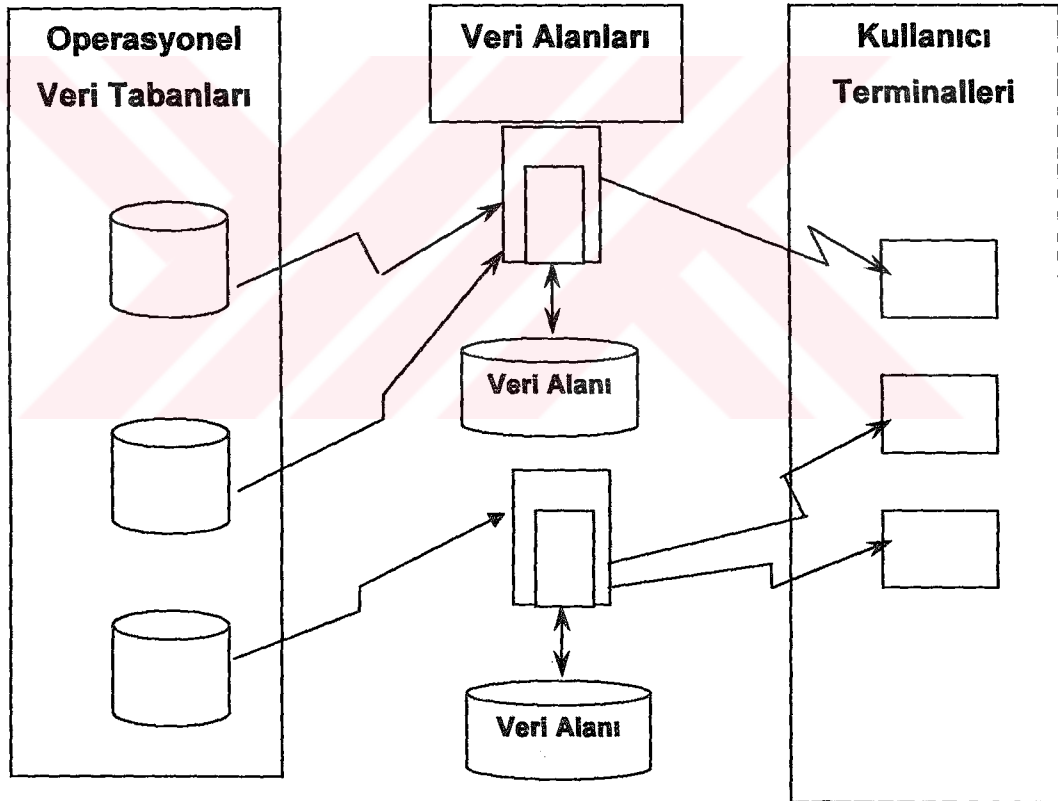
Şekil 1.4'te şematik olarak sunulan yapıda VA ile kullanıcı ara birimi arasında veri alanları oluşturulmuştur ve kullanıcılar VA'nın sadece kendi ihtiyaçlarına

¹⁶ MANNINO M. V.: a.g.e., s.458.

uygun olan kısmına giriş yapmaktadırlar. Burada sözü geçen veri alanları İngilizce'de DATA MART olarak adlandırılmaktadır.

1.9.3 Aşağıdan Yukarı Mimari

Aşağıdan yukarıya doğru yapılandırılan modelde ise büyük bir VA yerine daha küçük ve görece olarak daha az karmaşık veri alanları kullanılmaktadır. Oluşturulan veri alanının tamamına erişimin hemen hemen hiç gerekli olmadığı durumlarda uygulanabilecek bu yapı bilgiye hızlı ulaşımın sağlanması amacıyla kullanılabilir.



Şekil 1.5 Aşağıdan Yukarıya Mimari¹⁷

¹⁷ MANNINO M. V.: a.g.e., s.459.

Aşağıdan yukarı doğru yapı, çok kaynaklı bir mimari oluşturduğundan bazı durumlarda veri bütünlüğünün sağlanması problemini yaratabilir ve bunun sonucunda da kurumsal bilgi üretiminin karmaşıklaşması sonucunu doğurabilir.

1.10 Dünyada VA Sektöründe Pazar Payları:

VA sektöründe faaliyet gösteren irili ufaklı bir çok şirket mevcuttur. Bunlardan bazıları VA çözümleri üretme konusunda uzmanlaşmış firmalardır. Tablo 1.10 VA pazarında dünya çapında faaliyet gösteren firmaları ve bu firmaların Pazar paylarını özetlemek üzere sunulmuştur.

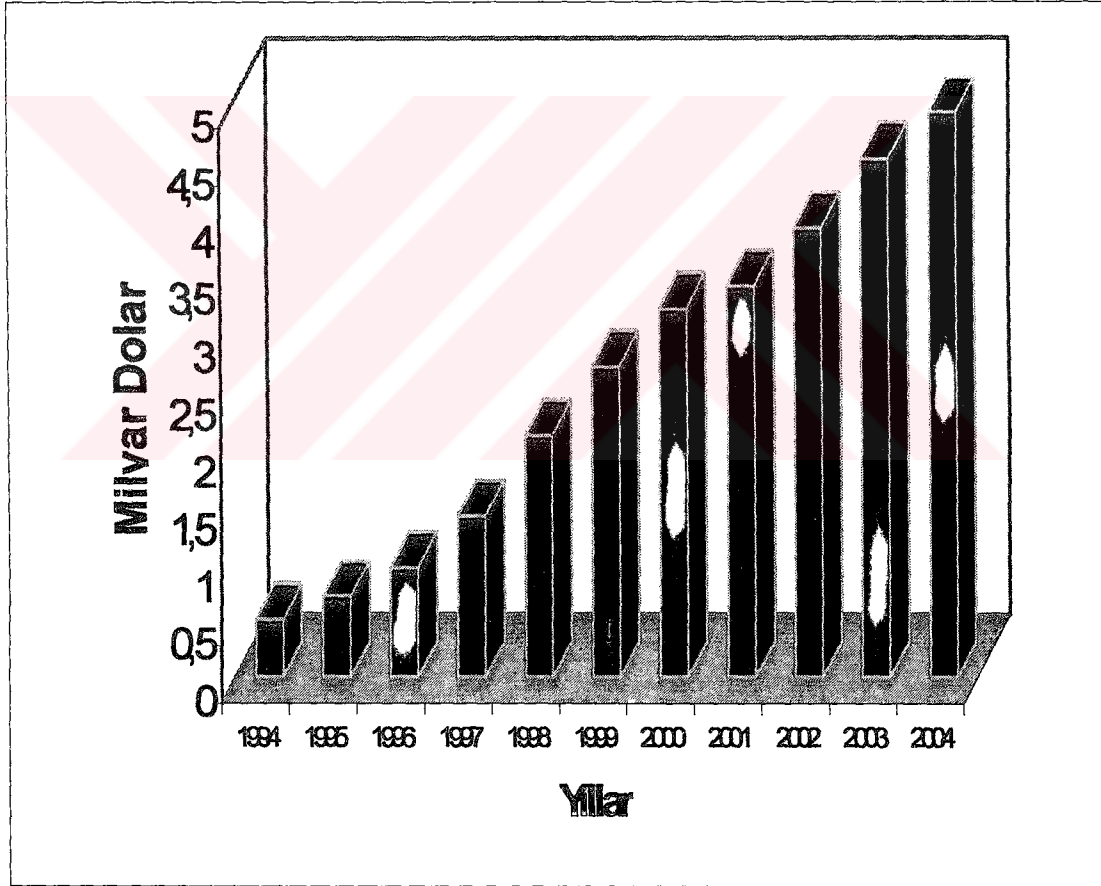
Tablo 1.10 Dünyada VA Pazarı ve Pazar Payları¹⁸

Ürünler:	Yıllar						
	2001	2000	1999	1998	1997	1996	1995
Hyperion Solutions	21.3%	22.3%	23.0%	28.7%	24.7%		
Microsoft	21.3%	11.7%	7.6%				
Cognos	12.1%	12.6%	11.1%	9.6%	10.8%	9.0%	5.0%
Oracle	7.1%	10.1%	11.4%	17.0%	20.7%	18.7%	20.2%
MicroStrategy	6.9%	9.2%	7.9%	6.5%	4.8%	3.0%	1.9%
Business Objects	6.6%	6.2%	5.3%	4.4%	3.6%	1.0%	-
Applix	2.5%	3.0%	3.1%	2.5%	2.5%	1.7%	1.4%
Cartesis/PwC	2.5%	2.3%	2.5%	2.1%	2.7%		
Comshare	2.4%	2.5%	3.2%	4.8%	7.5%	11.8%	%16
IBM(DB2 OLAP Server)	2.1%	2.6%	3.0%	1.9%			

¹⁸ PENDSE, N.: "The OLAP Report", **Business Intelligence Ltd. Group**, 30 Mart 2003, [Çevrimiçi], <<http://www.olapreport.com/Market.htm>>, 15 Nisan 2003

(Önceki Sayfadan Devam)							
Adaytum	1.5%	1.1%	1.0%	0.8%			
Brio Technology	1.3%	1.5%	1.5%	1.1%	0.8%	0.5%	0.3%
SAS Institute	1.2%	1.6%	2.2%	3.6%	2.1%	1.5%	0.7%
Crystal Decisions	1.0%	1.4%	1.7%	2.9%	4.1%	4.2%	4.7%
CA	0.3%	0.9%	2.8%	2.9%	1.9%	1.4%	0.9%

1.11 OLAP Pazarının Büyüme Seyri



Şekil 1. 6 OLAP Pazarının Büyüme Seyri¹⁹

¹⁹ PENDSE N.: a.g.e., [Çevrimiçi], <<http://www.olapreport.com/Market.htm>>, 15 Nisan 2003

Firmaların zaman içerisinde ürettikleri verilerin ve bu veri yığınlarından bilgi elde etme ihtiyaçlarının artmasına paralel olarak dünyadaki OLAP sektörü de büyümeye devam etmektedir. Şekil 1.6 geleceğe yönelik tahminleri de gözönünde bulundurarak, piyasa büyüme seyrini göstermek üzere sunulmuştur.



BÖLÜM 2. VERİ MADENCİLİĞİ

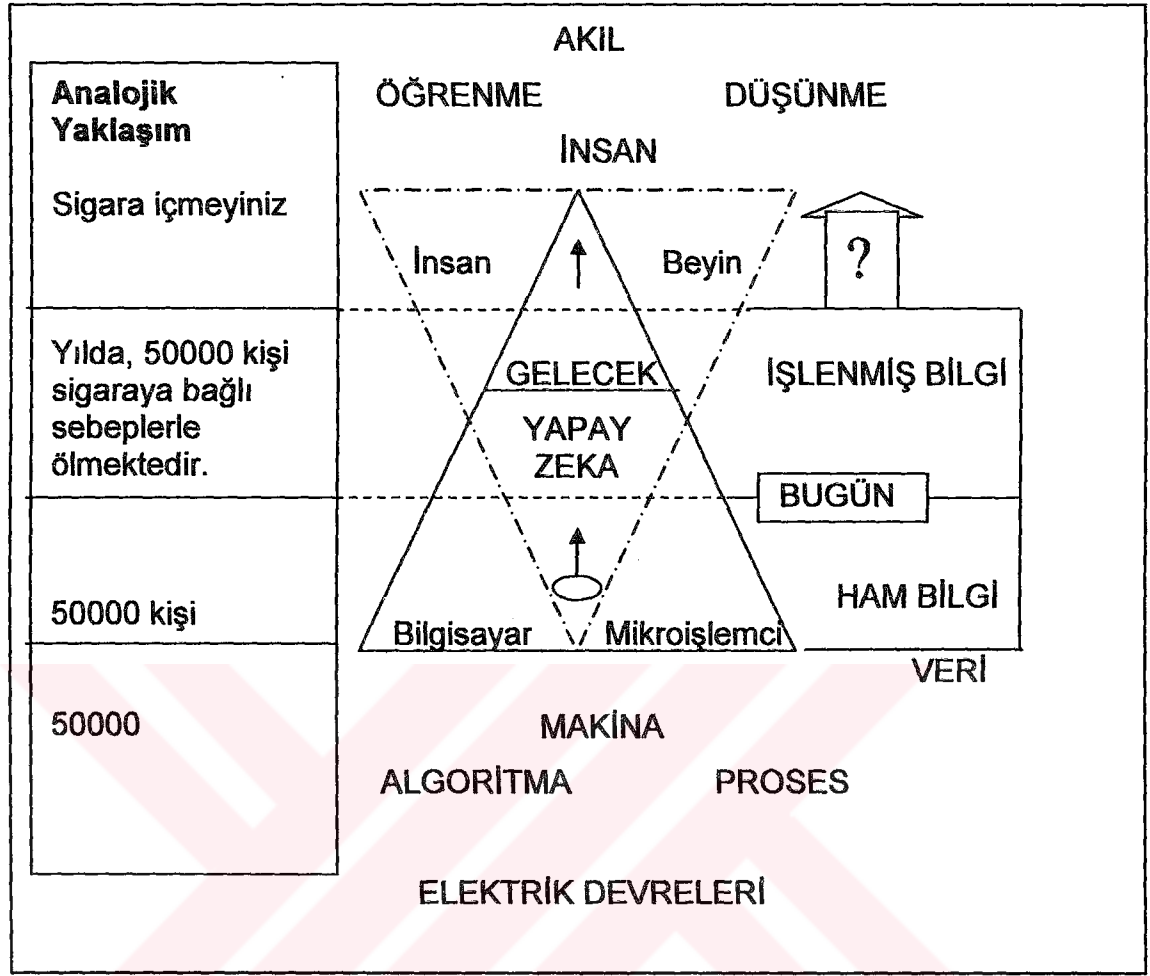
Organizasyonlar ihtiyaç duydukları veri zenginliğine ulaştıktan sonra çeşitli kaynaklarla elde edilen verilerin işlenmesini ve organizasyona stratejik avantajlar sağlamasını amaçlarlar. Daha önce de belirtildiği gibi büyük veri kitleleri ve çok açık olmayan bilgi ihtiyaçları, kurumları konu üstünde daha çok çalışmaya yönlendirmiştir. Bunun neticesi olarak da VM kavramı ortaya atılmıştır.

2.1 Tanım

VM, büyük veri kümelerinden, otomatik ve yarı otomatik yöntemlerle anlamlı ve kullanılabilir bilgilerin, çeşitli analiz ve araştırmalarla çıkarılmasıdır.²⁰

Yukarıdaki tanım net olarak anlamlı ve kullanılabilir bilgiye odaklanmıştır. İngilizce’de yer alan data, information ve knowledge terimleri veri, ham bilgi ve kullanılabilir işlenmiş bilgi kavramlarının altını çizmektedir.

²⁰ BERRY, M. A., LINOFF, G. S.: “**Mastering Data Mining: The Art And Science Of Customer Relationship Management**”, New York, NY: John Willey & Sons, Inc., 2000, s.7



Şekil 2.1 Veri Bilgi Dönüşümü

Yukarıdaki model sadece açıklayıcı olması amacı ile geliştirdiğimiz şematik bir gösterimdir. Bilgisayarların algoritma ve prosese dayalı çalışma prensipleri ile akıl sahibi insanoğlunun düşünme ve öğrenme süreçleri arasındaki açık farkı göstermek üzere sunulmuştur.

VM'nin temel amacı, veriden yola çıkarak bilgiye doğru olabildiğince uzun yol kat etmektir. Yapılan tüm çalışmalar ve geliştirilen yöntemler bu amaca hizmet etmeye odaklanmıştır. Sözü geçen yöntemler, istatistiki bir çok analizi,

genetik algoritmaları, yapay sinir ağlarını, bulanık kümeleri ve benzeri bir çok yaklaşımı barındırır.

En basitinden basit regresyon analizi bile biri bağımsız diğeri ise bağımlı olan iki değişken arasındaki ilişkiyi araştırır ve buradan kullanılabilir bir bilgi üretmek için kullanılır.

2.2 VM'ni Gerekli ve Mümkün Kılan Etkenler

Şekil 2.1'de şematik olarak gösterilen veriden işlenmiş bilgiye doğru yolda atılan adımlar teknolojik gelişmenin sunduğu yazılım ve donanım gücü ile ulaşılan bilgi birikiminin kullanılması ile mümkün olmuştur. Yüzlerce megabayt verinin, binlerce iterasyona ve benzeri işlemlere tabi tutulması kullanılabilir veri ile parazit olarak sistemde bulunan anlamsız veri parçalarının ayıklanması ve kullanılacak verilerin birçok karmaşık sürece tabi tutulması ancak güçlü yazılım donanım ve uygun yöntemlerin kullanılmasıyla mümkün olabilmektedir.

2.2.1 VM uygulamalarının Başarıya Ulaşmasını Sağlayan Alt Yapı Çalışmaları

VM ile ilgilenen kişi ve kurumların üzerinde durmaları gereken bazı temel ve uygulamanın başarısı şansını en baştan belirleyen bazı temel faaliyetler vardır bunlar;²¹

- Verilerin açılması(extract): Bir çok uygulamada veriler ham vaziyette depolanarak bir çok pratik fayda sağlanır ancak bu verilerin kullanılabilmesi için sözü geçen ön işlem birçok durumda gerekli

²¹ GROTH, R.: "Data Mining: Building Competitive Advantage", New Jersey, NY: Prentice Hall., 2000, s.46-55

olmaktadır. Örneğin bir çalışana ödenecek ücret, birim çalışma ücreti ile çalışma saatinin çarpımı ile hesaplanır ve genellikle böyle bir işlem sonucu elde edilen veri, veri tabanında saklanmaz sadece bu sonuca ulaşmak isteyen kullanıcı gerek duyduğunda hesaplanarak uygulama programı ile bir çıktı olarak kullanıcıya sunulur. Böyle bir grup veri ile işlem yapabilmek için bu tür verilerin öncelikle açılarak kullanılabilir hale getirilmesi gerekir.

- Verilerin bütünleşik hale getirilmesi: VM çalışması bir çok kaynaktan elde edilen verilerle çalışılmasını gerekli kıldığı için bu tür farklı kaynaklı verilerin gerek format ve gerekse içerik olarak uyumlu hale getirilmesi gerekmektedir. Ancak bu şekilde tutarlı bir veri bütünü elde edilerek anlamlı bilgiler elde edilebilir.
- Verilerin temizlenmesi: Buradaki temizlik sözcüğü anlam ifade etmesi mümkün olmayan ve sisteme bir şekilde girmiş verilerin ayıklanmasını vurgular. Örneğin hatalı bir tasarım sonucu, sistemde sayısal olması gereken ve üzerinde matematiksel işlem yapılacak alanların bazılarında sayısal olmayan değerler bulunabilir veya yanlışlıkla bir müşterinin yaşı 1009 olarak girilmiş olabilir. Bu tür beklenmeyen ve önceden önlenmemiş anlamsız verilerin ayıklanması sonucun gerçekçi olabilmesi için çok önemli bir aşamadır.
- Veri tutarlılığının devamını sağlayacak güvenlik önlemlerinin alınması: Bir çok firma için elde edilmiş olan veriler bir tür aktif olarak görülmektedir ve firmanın önemli bir kaynağı olan veri bankasının hatalı işlemler sonucu hasar görmesi önlenmeye çalışılmaktadır. VM çalışmaları ile aynı şekilde veri bütünlüğünün, tutarlılığını ve tamlığının bir hata sonucu zedelemesi riski engellenmelidir.

2.2.2 VM Süreci

VM süreci, bilimsel problem çözme sürecine benzer şekilde ele alınmalı ve yapılacak faaliyetler bir mantık sırası içerisinde düzenlenmelidir. Buradaki temel sorun, yapılacak VM çalışmasının kuruma en fazla faydayı sağlamasını planlamak ve projelendirmektir.

VM çalışması bir proje faaliyeti olarak algılandığında gerçekleştirilmesi gereken süreçler aşağıda özetlenmiştir:²²

1. Problemin tanımlanması,
2. Keşif süreci,
3. Uyarılama,
4. Uygulama,
5. Sonuçların gözlemlenmesi,
6. Tüm süreçlerin gözden geçirilmesidir.

VM uygulamalarında izlenebilecek beş temel yaklaşım vardır:²³

1. *Sonuçların Satın Alınması*: Bunun en güzel örneklerinden birisi, bankalara hizmet veren, bağımsız firmalardır. Bir kredi talebi incelenirken, çok sayıda bankanın üyesi olduğu kuruluş bütün finansal kurumlardan derlenmiş verileri işleyerek kredi puanlaması yapar. Burada müşteri konumundaki banka sadece modelin sonuçlarını kullanır. Buna benzer uygulamalar Türkiye’de başlatılmıştır.

²² GROTH, R.: a.g.e., s.13-17.

²³ BERRY, M. A., LINOFF, G. S.: “Mastering Data Mining: The Art And Science Of Customer Relationship Management”, New York, NY: John Wiley & Sons, Inc., 2000, s.24-36.

2. *Yazılım Satın alınması:* VM pazarın çok sayıda bulunan paket programlardan bir veya bir kaçını satın alınarak VM süreci gerçekleştirilebilir. Paket programların odakları alanlar farklılık gösterdiği için farklı işlerde farklı programlar kullanılması zorunlu olabilir.
3. *Model Satın Alınması:* Buradaki temel amaç, daha önce başka bir kurumda uygulanmış bir modelin top yekûn satın alınması ve model oluşturulması, verilerin derlenmesi, uygulamanın performansının test edilmesi gibi zahmetli uğraşlardan kurtulmaktır.
4. *Danışmanlardan veya Danışmanlık Firmalarından Destek Alınması:* Daha önce bir çok benzer projelerde yer almış ve uzmanlık alanı VM olan bir kişi ve veya grup ile, VM projesinin yürütülmesi için anlaşılabilir.
5. *Kurum İçi Uzmanlığın Oluşturulması:* Eğer bir firmada bu tür bir faaliyeti yürütebilecek bilgi ve deneyime sahip bir kadro oluşturulabiliyorsa ve projenin firma içinde yürütülmesinin güvenlik ve/veya gizlilik gibi pratik faydaları varsa izlenebilecek bir yoldur.

2.3 VM ile Yapılan Temel Görevler ve VM Kullanım Alanları

VM, veriden rafine bilgiye ulaşma süreci olduğuna göre, VM sürecinden beklenen görevler bilgi keşfi ve elde edilen bilginin anlaşılabilir bir biçimde sunulmasıdır.

VM, verinin bulunduğu ve anlamlı bilginin kolayca elde edilemediği her işletme probleminin çözülmesinde kullanılabilen bir yaklaşımdır. VM, sadece bir analiz yöntemi veya bir paket program olmadığı, geleneksel ve modern bir

çok tekniğin aynı anda kullanımına olanak sağladığı için kullanım alanları çok geniştir. Her türlü karar destek ihtiyacı ile karşılaşılan durumlar, VM için uygulama alanlarıdır. Özellikle, pazarlama, müşteri ilişkileri yönetimi, finansal yönetim, bilişim sisteminin yönetimi, araştırma ve geliştirme faaliyetleri, üretim planlaması ve kontrolü, proses ve kalite kontrolü, VM'den yararlanılabilecek alanlardandır.

Bu amaçla, VM algoritmaları tarafından yapılan işlevlerin başlıcaları şunlardır:^{24 25}

Sınıflandırma,
Gruplandırma,
Görselleştirme,
Bağıntı kurma (Sepet Analizi),
İstisnaların tespiti,
Doküman madenciliği,
Tahmindir.

2.3.1 Sınıflandırma

Sınıflandırma çalışmalarında uygulamacılar, bir olgu ile ilgili var olan verilerden yola çıkarak, etkileşimde bulunulan öğeleri sınıflandırmaya çalışırlar. Örneğin kredi kartı almak için bir bankaya başvuran müşterilere tahsis edilecek kredi miktarının belirlenmesi bir sınıflandırma işlemidir.

Sınıflandırma, mevcut veriden hareket eden ve bilgi keşfinde yararlanılan VM modelleri içerisinde en yaygın kullanıma sahip modeldir. Sınıflandırma işlevi

²⁴ AKPINAR, H. : "Veri Tabanlarında Bilgi Keşfi Ve Veri Madenciliği", İ.Ü. İşletme Fakültesi Dergisi, Sayı :1, Nisan 2000, s.5

²⁵ GROTH, R.: a.g.e., s.22-39.

için kullanılan başlıca teknikler; Karar ağaçları, YSA, GA, K-en yakın komşu, bellek temelli nedenleme, lojistik regresyondur.²⁶

2.3.2 Gruplandırma

Gruplandırma veya kümeleme çalışmalarında elde edilmeye çalışılan bilgi, ögeleri gruplar halinde ifade ederek ortak özelliklere sahip gruplara ortak davranışlarla yaklaşmaktır. Kredi kartı başvurusu örneğine geri dönülecek olursa, kredi kartı talep eden müşteriler iki temel gruba ayrılabilir. Bunlar; kredi kartı borcunu düzenli ödeyen müşteriler ve borçları yasal takip yoluyla tahsil edilmeye çalışılacak müşterilerdir. Böylesi gruplama ile potansiyel müşterilerin kredi kartı borçlarını zamanında ödeyip ödemeyeceklerinin başarılı bir şekilde tahmin edilebilmesinin, o bankaya önemli avantajlar sağlayabileceği açıktır.

2.3.3 Görselleştirme

Görselleştirme, verilerin grafiksel ögelerle ifade edilmesi yöntemidir. Basit bir yöntem olmasına rağmen görselleştirme bazı durumlarda verilerin en iyi şekilde anlaşılabilmesi yöntemi.²⁷ Belki de bu nedendendir ki, çoğu makalelerde ve tezlerde grafikler ve şekiller sıklıkla kullanılmaktadır.

Özellikle, üst düzey yöneticilere sunulan raporlardaki verilerin grafiksel biçimde sunulması, hem anlaşılabilirliği artırırken hem de yöneticilerin dikkatlerini dağıtmadan hızlı karar verebilmelerini sağlayabilmektedir.

²⁶ AKPINAR, H. : a.g.e., s.5

²⁷ GROTH, R.: a.g.e., s.26.

2.3.4 Bağıntı Kurma (Sepet Analizi)

Sepet analizi temelinde süper marketlerde müşterilerin alışveriş davranışlarının belirlenmesi ve buna göre düzenlemelerin yapılması prensibine dayanır. Örneğin, bir marketten bebek bezi alan insanların büyük bir kısmı bebek maması da alırlar önermesi, ispatlandığında market yöneticileri bebek bezi ve mamasını aynı bölümde sergileyerek satışlarını arttırma yoluna gidebilirler. Buna benzer örnekler birçok kaynaktan rastlanmaktadır. Buradaki temel mantık gruplandırma veya kümelendirmeye benzemektedir.

2.3.5 İstisnaların Tespiti

Buradaki amaç, diğer birçok uygulamanın aksine kural bulmak değil, kuralları uymayan istisnai hareketleri bulmaktır.²⁸ Bir işletmenin müşterilerine karşı davranışları, sınıflama ve gruplama prensiplerine uygun bir şekilde düzenlendiği uygulamalarda, istisnalar sürekli olarak değerlendirilmeli ve bu örneklerle ayrıcalıklı yaklaşılmalıdır. Aksi halde fırsat kaybı veya gereksiz risk alma durumlarıyla karşılaşılabilir. Örneğin lüks otomobillerin satıldığı bir mağazaya müşteri olarak gelen, giyim tarzı orta düzeyli olan birçok insan, mağazadaki lüks otomobillerden birisini alabilecek ekonomik güce sahip değildir. Ancak bu ön yargı ile tüm, bu tür insanlara ilgisiz kalmak, istisnai olarak giysisi vasat olan ancak gerçek alıcı olan bir müşterinin de elden geçirilmesine yol açabilir. Benzer şekilde, çok iyi giyimli bir kişi ise sadece zaman geçirmek için o mağazada bulunuyor olabilir. Burada istisnaların tespit edilmesi önemli bir fayda sağlayabilmektedir.

²⁸ APAYDIN, E. : "ZEKİ VERİ MADENCİLİĞİ:Ham Veriden Altın Bilgiye Ulaşma Yöntemleri",2000, **Bilişim 2000 Eğitim Semineri**, İstanbul, [Çevrimiçi]
http://www.cmpe.boun.edu.tr/~ethem/files/papers/veri-maden_2k-notlar.doc, 5. Mayıs 2003

2.3.6 Doküman Madenciliği

Buradaki amaç, dokümanlar arasında ayrıca elle bir tasnif gerekmeden benzerlik hesaplayabilmektir (*text mining*). Bu genelde otomatik olarak çıkarılan anahtar sözcüklerin tekrar sayısı sayesinde yapılır.²⁹

İnternet protokolü üzerinde faaliyet gösteren arama motorları, belki de bu tür uygulamalara en güzel örneklerdir. Milyonlarca megabyte verinin bulunduğu internet bulutu içerisinde anahtar kelimeler yardımı ile, bilgilerin saniyeler mertebesinde sorgulanabilmesi önemli sayılacak bir algoritmik güç gerektirir.

2.3.7 Tahmin

Tahmin terimi önemi gereği ayrı bir başlık olarak yer almasına rağmen, diğer birçok işlevde gizli olarak yer almaktadır. Örneğin sınıflandırma metodolojisinde sınıflandırmanın yapılması ya şu ana kadar karşılaşılmamış bir durumun yada gelecekte olabilecek bir durumun sınıflandırma yöntemi ile tahmin edilmesi prensibine dayanır.

Deterministik olan, yaşanmış, sebepleri ve sonuçları kolaylıkla belirlenebilir olaylar için VM çalışmasına gerek dahi yoktur, bu gibi durumlar kolaylıkla algılanabilir. Örneğin Türkiye'nin en büyük sanayi kuruluşunun yönetim kurulu başkanı, bir bankadan kredi kartı talep ettiğinde değerlendirme yapılması gereken fazla bir konu olmayacağı açıktır. (İstisnalar bu ifade dışında değerlendirilmelidir.)

Tahmin, belirsizlik ve/veya risk durumunda çoğunlukla mecburen izlenen bir yoldur. Çünkü kesin olarak bilinebilen bir olgu ile ilgili tahmin yapılması mantıksal bir durum sayılamaz. Tahmin, zaman serileri analizi, ilişki analizi

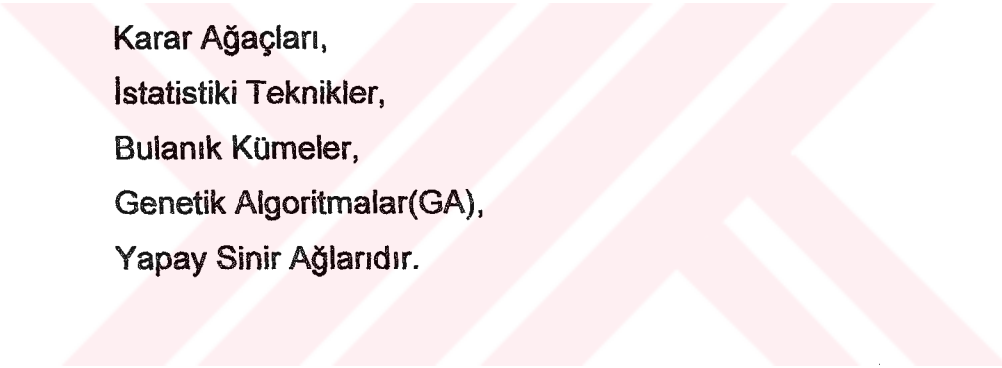
²⁹ APAYDIN, E. : a.g.e., s.5

veya başka benzer analizlerle yapılabilir ve her koşulda kesinliği olmayan bir durumun başarı ile önceden kestirilmesini hedefler.

2.5 Temel VM Yöntemleri

VM , bir çok yöntemi barındıran bir metodolojidir. VM bir algoritma veya bir yazılım olarak değerlendirilmemelidir. Bütünleşik bir yapı içerisinde, pratik sorunların ve uygulamaların gerektirdiği işlevler ve yöntemlerle kurulan VM metodolojilerinin başarı şansı daha yüksektir.

VM yöntemlerinin başlıcaları:

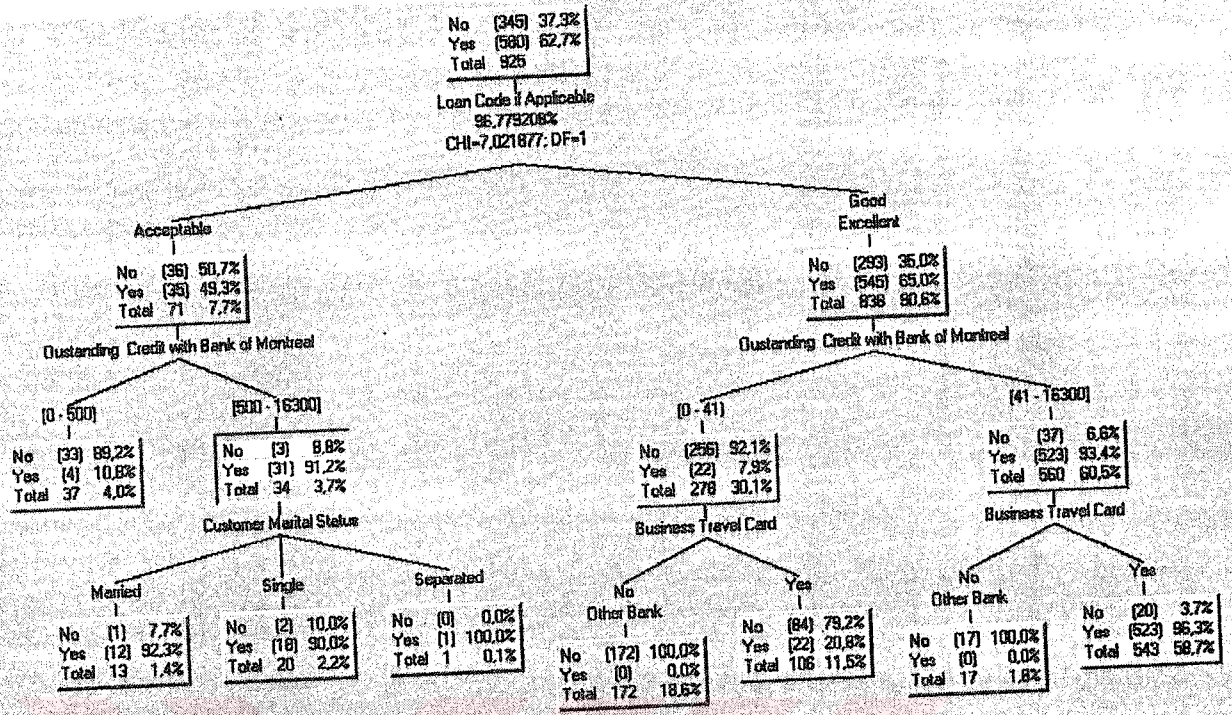


Karar Ağaçları,
İstatistiki Teknikler,
Bulanık Kümeler,
Genetik Algoritmalar(GA),
Yapay Sinir Ağlarıdır.

2.5.1 Karar Ağaçları

Karar ağaçları genel olarak çok basit bir altyapıya sahip olmalarına rağmen uygulayıcılara önemli bilgiler sunabilmektedirler. Genellikle Gruplama uygulamalarında kullanılan karar ağaçları, belirli bir kümeyi katmanlı olarak alt kümelere ayırarak uygulamacılara yol gösterir.

Karar ağaçlarının kullanımına örnek olarak "Knowledge-Seeker IV Version 4.4" adlı program ile elde edilen bir karar ağacı Şekil 2.2'de sunulmuştur.



Şekil 2.2 Karar Ağacı Örneği

Şekil 2.2'de örnek olarak sunulan ağaç, çok boyutlu bir veri grubu içerisinde anlamlı alt gruplar oluşturarak oluşturulmaktadır. Alt grupların oluşturulmasında izlenen yol Ki-Kare değerlerinin hesaplanmasıdır.

Karar ağaçlarının kullanılabileceği alanlardan birisi de, kurallar geliştirmektir.³⁰ Sözü edilen kurallar iş süreçlerinin yönetiminde yol gösterici olarak kullanılabilirler. Şekil 2.3'te Knowledge Seeker programı ile otomatik olarak oluşturulan kurallara örnekler sunulmuştur.

³⁰ BERRY, M. A. and LINOFF, G. L.: a.g.e., s.120.

```
RULE_1 IF Bank of Montreal Savings Account = Yes
        Business Travel Card = Yes
        Outstanding Credit with Bank of Montreal = [41,16300]
THEN Bank of Montreal Mastercard? = No 3.0%
        Bank of Montreal Mastercard? = Yes 97.0%

RULE_2 IF Bank of Montreal Savings Account = Other Bank
        Business Travel Card = Yes
        Outstanding Credit with Bank of Montreal = [41,16300]
THEN Bank of Montreal Mastercard? = No 14.9%
        Bank of Montreal Mastercard? = Yes 85.1%
```

Şekil 2.3. Karar Ağaçları ile Kural Üretilmesi

Yukarıda örneği gösterilen karar ağaçlarının ve oluşturulan kuralların geçerliliğinin test edilmesi önemlidir. Elde edilen sonuçlar istatistiki analizler kullanılarak elde edilmektedir. Bu nedenle güvenilirlik testi gerçekleştirilmelidir.

2.5.2 İstatistiki Teknikler

VM, çalışmaları sadece istatistiki tekniklerin kullanıldığı alanlar olarak değerlendirilmemelidir. Ancak VM, istatistiki teknikleri dışlayan, o tekniklerin sağladığı pratik faydalardan yararlanmayan bir alan hiç değildir. En basitinden, en karmaşığına kadar bir çok istatistiki yöntem, VM kapsamında uygulama alanı bulabilmektedir. Hemen hemen tüm istatistiki teknikler VM sürecinde kullanılabilir. Bunlardan bazıları şunlardır:

Regresyon Analizi: Bir bağımlı ve bir veya birden çok bağımsız değişken arasındaki ilişkiyi matematiksel bir fonksiyon şeklinde modellemeyi amaçlayan regresyon analizi, pratikte yaygınlıkla kullanılmaktadır.

Değerli bir teorik altyapısı olan en küçük kareler yöntemi ile hata kareleri toplamını minimum kılan regresyon denklemi oluşturulur.³¹ Bu denklem ve denkleme ait katsayılar sayesinde, kurulan model tahminde ve duyarlılık analizinde rahatlıkla kullanılabilir.

Geleneksel regresyon modellerinde sürekli değişkenler kullanılmak zorundadır. Bu dezavantajı ortadan kaldırmak üzere, hem kategorik hem de sürekli değişkenlerin aynı modelde yer alabileceği "Lojistik Regresyon" yöntemi geliştirilmiştir.³²

Diskriminant Analizi: Bu analiz, katsayılar ve ağırlıklar setini bularak doğrusal sınıflandırma fonksiyonunu elde etmeye çalışır. Doğrusal sınıflandırma ise değişken gruplarını maksimum bir şekilde ayırır.³³

Diskriminant analizi, VM'nin görevlerinden olan gruplandırma ve kümeleme çalışmalarında rahatlıkla kullanılabilir.

Faktör Analizi: Bu analizin amacı, çok sayıdaki değişkenlerden oluşan bir veri setinin özelliklerini çok daha az sayıdaki değişkenle ifade edebilmektir. Bu az sayıdaki değişkene de faktör denilmektedir. Bu analiz, çok sayıdaki bağımlı değişken arasındaki ilişkinin yapısının belirlenmesi, bir olaya etki eden faktörlerin gruplanmasını sağlamaktadır.³⁴

³¹ NEWBOLD, P.: "Statistics For Business And Economics", New Jersey, NY: Prentice Hall., 1995, s.446.

³² GROTH: a.g.e., s.84.

³³ GROTH: a.g.e., s.83.

³⁴ DARLINGTON, R.B.: "Factor Analysis", [Çevrimiçi], <http://comp9.psych.cornell.edu/Darlington/factor.htm>, 15.Mayıs.2003.

Diğer analizler gibi faktör analizi de karmaşık veri grupları içerisinde anlamlı yapıların çıkarılması yoluyla rafine bilgiye ulaşılmasına yardımcı olur.

Ki-Kare Testi/Analizi: Ki-kare, parametrik olmayan analizlerden birisidir. Bu analizin kullanılabileceği alanlardan birisi karar ağaçları ve kural üretme mekanizmalarıdır. Şekil 2.2'de örneği verilen karar ağacı, ki kare değerleri referans alınarak üretilmiştir.

Zaman Serileri Analizi: "Herhangi bir olayın gözlenmesi neticesinde ortaya çıkan değerleri zamana ait bir unsur (ay, gün, yıl, vb.) ile açıklayan serilere zaman serileri denir."³⁵ Bu tür serilerin analiz edilmesi ile geleceğe yönelik tahminler yapılabilir. Örneğin bir firmanın, belirli dönemlere ait geçmiş satış miktarları bilindiği takdirde, bu verilerden yola çıkarak, gelecekteki satış rakamları tahmin edilebilir.

Temel olarak bir ilişki olan zaman serileri analizinde, bağımsız değişken zaman, bağımlı değişken ise geleceğe yönelik tahmin yapılmak istenen özelliktir. Bu tür analizler yöneticilerin geleceğe dönük plan yapmalarına ve nesnel veriler yardımı ile karar verebilmelerine yol açar.

Burada, sadece bazı istatistiki tekniklerden kısaca söz edilmiştir. Daha ayrıntılı bilgi edinmek isteyen okuyucular, istatistik ile ilgili yayınlardan yararlanabilirler. Ayrıca, burada isminden bahsedilmemiş tekniklerin daha az önemli oldukları veya daha az kullanışlı oldukları gibi bir yargıya varılmamalıdır.

³⁵ ÇAĞIL, G.: "Mevsimlik Olmayan Box-Jenkins Modellerinde İki Aşamalı Yapay Sinir Ağlarının Kullanılması ve Bir Uygulama", Doktora Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, 1997, s.2.

2.5.3 Bulanık Mantık (Fuzzy Logic)

Bulanık mantık, insanların düşünme sistemine yakın bir algılama biçimi öngören bilgisayar destekli algoritmalarından birisidir. Bulanık mantık temel olarak, tam verinin bulunmadığı durumlarda, problemlere insan beyni gibi yaklaşmayı hedeflemektedir.³⁶

Bilgisayarlarda kullanılan geleneksel mantık sistemi, kümelerin sınırlarını, kural gerçekleştiğinde bir, gerçekleşmediğinde ise sıfır değeri atayarak gerçekleştirir.³⁷

Eğitim sisteminden basit bir örnek ile durum daha iyi anlaşılabilir; Eğer geçer not "50" olarak belirlenmişse ve öğretmenler hiç inisiyatif kullanmıyorlarsa "49,4" alan bir öğrenci başarısız, "49,6" alan öğrenci ise başarılı sayılmalıdır. Ancak iki kümenin yani başarılı ve başarısız setinin çok net tanımlanamadığı durumlarda bu mekanik mantık çok başarılı sonuçlar vermez. Başarılı olduğu veya başarısız olduğu kesin ve net olan öğrenciler için "sıfır, bir" mantığı, geçerli, tutarlı ve kullanışlıdır. Ancak sınıra çok yakın değerler için daha esnek bir yöntem pratikte sıklıkla kullanılır. Bu mantığın matematiksel bir temele dayandırılarak ve algoritmik olarak uygulandığı yöntemler, bulanık mantık sistemleri olarak adlandırılmaktadır.

2.5.4 Genetik Algoritmalar(GA)

GA, evrim sürecini benzeşim yoluyla bilgisayar ortamına taşıyan önemli bir VM aracı olarak bir çok uygulamaya konu olagelmıştır. GA çalışmalarının

³⁶ LIN, T. C., LEE, C.S.G.: "Neural Fuzzy Systems, A Neuro-Fuzzy Synergism to Intelligent Systems", New Jersey, NY: Prentice Hall, 1996, s.1.

³⁷ YÜREKLİ, H.: "Bulanık Uzman Sistemler", Yüksek Lisans Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, 1999, s.2.

temel prensibi olan güçlü bireylerin hayatta kalabilmeleri ve genetik özellikleri sonraki nesillere taşıyabilmeleri yaklaşımı, GA'nın bir çok pratik uygulamada rol oynayabilmeleri şansını doğurmuştur. GA, bu çalışmanın temel elemanlarından birisi olduğu için Bölüm 3.2'de ayrıntılı olarak ele alınmıştır. Özellikle doğrusal olmayan optimizasyon problemlerinde sıklıkla kullanılan GA, çalışmamız kapsamında YSA öğrenmesinin bir aşamasını yönetme görevi üstlenmişlerdir.

2.5.5 Yapay Sinir Ağları

YSA metodolojisi, yöntemin ayrıntıları ve uygulamalar ayrıntılı olarak Bölüm 4'te açıklanmıştır. Burada sadece YSA'nın VM açısından önemine kısaca değinilmektedir. YSA, doğrusal olmayan ve belirsiz ilişkiler ağının var olduğu durumlarda, sinir fizyolojisi temellerini kullanarak, anlamlı ve kullanışlı bilgilerin elde edilmesine olanak veren çok esnek, bir çok problemde kullanılabilen ve de pratik sonuçlar üreten önemli bir VM algoritmasıdır. Sözü geçen önemli faydaları gereği, YSA bu çalışmanın temel parçalarından birisi olarak ele alınmıştır.

2.6 VM ve Ahlaki Konular

VM, kurumlar için çok önemli faydalar sağlamakla beraber bazı sorunlarda beraberinde getirmektedir. Veri halinde kalmış girdiler işlenemedikleri sürece, bireylere çok yararlı veya zararlı olma olasılıkları yoktur. Ancak herhangi bir kurum veriden yola çıkarak bilgiye ulaştığı anda bu bilgileri hukuki ve ahlaki kuralları gözетerek kullanarak fayda elde etmelidir.

Problem, bireylerden alınan verilerin işlenmesi sonrasında elde edilen bilginin, uygun kullanılıp kullanılmaması noktasında başlamaktadır. Çok güçlü bir rekabet silahı olarak kullanılabilecek olan VM, aynı zamanda kişilik

haklarına tecavüz eden bir araç olarak da kullanılabilir. Burada; güvenlik, kişisel haklara saygı, mesleki sırların saklı tutulması ve elde edilen bilgilerin üçüncü taraflara sunulmaması gibi ögeler gözönünde bulundurulmalıdır.

Türkiye'de de yaygınlaşmaya başlayan bir uygulama olan büyük marketlerin firma kredi kartları dağıtması uygulamasında, tüm kişisel bilgiler müşterilere verilen form sayesinde elde edilmektedir. Bu bilgiler gönüllü olarak bir promosyondan yararlanmak amacı ile birey tarafından verildiği için ahlaki veya hukuki olmayan bir durumdan söz edilemeyeceği basit bir gerçektir. Burada sunulan bilgilerin bir hasta doktor ilişkisinde olduğu gibi bir sır olarak kalması ve kötüye kullanılmamasının sağlanması ilişkide bulunulan kuruma aittir. Bilgiyi toplayan kurum, kişisel bilgilerin korunması ve gizliliğinin saptanması konularında önlem almak zorundadır. Aksi durumda ise yasalar gerekli şekilde düzenlenerek uygun yaptırımlar gerçekleştirilmelidir.

2.7 VM Araçları Geliştiren Başlıca Firmalar³⁸

Dünya çapında yerel, bölgesel ve küresel düzlemde faaliyet gösteren çok sayıda VM hizmet sağlayıcısı olmasına rağmen, bu çalışmamız kapsamında sadece belli başlı firmalar gerektiğinde yararlanılmak üzere sunulmuştur:

³⁸ GROTH, R.: a.g.e., s.246-248.

Tablo 2.1 VM Hizmet ve/veya Ürün Sağlayıcıları

Firma İsmi	Çevrimiçi Adresi
Angoss Software International LTC.(Knowledge SEEKER)	http://www.angoss.com
Attar Software USA (XpertRule Miner)	http://www.attar.com
Business Objects Inc (Business Miner)	http://www.businessobjects.com
Cognos Corporation (Scenario)	http://www.congos.com
HNC Software Inc.	http://www.hnc.com
IBM Corporation	http://www.ibm.com
Information Discovery Inc. (The Data Mining Suite, The Knowledge Access Suite)	http://www.datamining.com
Isoft (AC2, Alice d'Isoft)	http://www.alice-soft.com
Neovista Solutions Inc	http://www.neovista.com
Oracle Corporation (Darwin)	http://www.oracle.com
RighPoint Software (DataCruncher)	http://www.rightpoint.com
SPSS Inc. (SPSS Clementine)	http://www.spss.com
SAS Institute Inc (Enterprise Miner)	http://www.sas.com
Silicon Graphics Computer Systems (MineSet)	http://www.sgi.com

BÖLÜM 3. GENETİK ALGORİTMALAR

GA, veri madenciliği ve karar destek sistemlerinde yaygın olarak kullanılan yöntemlerden birisidir. Bu çalışmamız kapsamı içerisinde oluşturulan model, YSA, GA melez yapısına dayandığı için GA, ayrıntılı bir şekilde işlenmiştir:

3.1 Tanım

Genetik algoritmalar, biyolojik süreçlerin örnek alındığı yöntemlerdendir. Evrim, doğal eliminasyon ve mutasyon gibi temel kavramlar algoritmik olarak benzeşim kurulması yoluyla matematiksel forma dönüştürülmektedir. Başlangıçta bulunan genetik özelliklerin nesilden nesle geçerken değişimini temel prensip olarak kabul eden genetik algoritmalar, görece güçlü ve amaca daha uygun özelliklerin korunması, aynı şekilde de zayıf ve amaca uygun olmayan özelliklerin ise nesiller arası geçiş sırasında elimine edilmesini basit formda temel alırlar.

3.2 Tarihçe ve Genel Prensipler

Genetik algoritmalarla ilgili ilk çalışmalar 1950'li yıllarda biyoloji ve bilgisayar uzmanlarından oluşan bir grubun ortak projesi olarak başlamıştır. 1960'larda ise John Holland ve bir grup meslektaşı Michigan Üniversitesinde genetik algoritmalar ve optimizasyon konularında çalışmaya başlamışlardır. John

Holland genetik algoritmalarla hesaplama ile ilgili teoriyi ve çalışma prensiplerini ortaya atan bilim adamıdır.³⁹

Evrimin mekanizması ve temel kavramlarını şu şekilde özetlenebilir: Evrim, amacı belli olan ya da kontrollü bir süreç değildir; en azından bu şekilde olduğuna dair elde edilmiş geçerli bir kanıt bulunmamaktadır. Farklı genetik bilgilere sahip canlılar yarışır ve güçlü (FITNESS) olanlar hayatta kalıp çoğalarak/üreyerek genetik bilgilerini bir sonraki kuşağa (GENERATION) geçirme şansını yakalarlar. Evrimin motoru üreme sonucu oluşan farklılıktır. Her iki canlının genetik olarak farklı ama işlevsel olarak eş kromozomları üst üste gelerek bilgi değiş tokuşu yaparlar "CROSS-OVER". Oluşan farklı canlılardan bir kısmı, bir önceki kuşağa göre değişen çevreye daha iyi uyum gösterip (ADAPTASYON) daha güçlü olabilir. Bunların yaşaması ve üreyerek soyunu devam ettirmesine DOĞAL SELEKSİYON denir.⁴⁰

Yukarıda açıklanan genel ve temel bilgilerin bilgisayar ortamında basitleştirilmiş şekilde uygulanması sonucunda genetik algoritmalar pratik sorunların çözümünde kullanılmaktadırlar.

3.3 Genetik Algoritmaların Temel Adımları

Genetik algoritmalarla ilgili bir çok yöntem bulunmasına rağmen tüm yöntemler aşağıdaki temel adımlardan oluşmaktadır.⁴¹

³⁹ BERRY, M. A. and LINOFF, G. L.: "Data Mining Techniques For Marketing, Sales, and Customer Support", New York, NY: John Willey & Sons, Inc., 1997, s.335-336.

⁴⁰ ATALAĞ, K.: "E-tip", [Çevrimiçi], <http://www.tbd.org.tr/sayi77_html/e_tip.htm>, 10 ocak 2003

⁴¹ LIN, T. C., LEE C.S.G.: a.g.e., 1996, s.384.

- Adım 1: Başlangıç kümesinin seçilmesi,
Adım 2: Küme içerisindeki her bir kromozomun incelenmesi ve uygunluk değerlerinin belirlenmesi,
Adım 3: Kümenin bazı üyelerinin(görece zayıf üyeler), yeni üyelere yer açmak amacıyla elimine edilmesi,
Adım 4: Eşleşme yoluyla yeni kromozomların yaratılması,
Adım 5: Mutasyonun uygulanması,
Adım 6: Yeni kromozomların incelemesi ve kümeye dahil edilmeleri,
Adım 7: Başarıya ulaşma kriterleri yerine getirilene kadar Adım 2'ye geri dönülmesi,
Adım 8: Algoritmanın sonlandırılması ve sonuçların kullanılması.

3.4 Algoritmayı Açıklayıcı Bir Örnek

Bir çok kaynakta işlenen en temel ve basit bir uygulama olan ikinci dereceden bir fonksiyonun maksimum noktasının belirlenmesi problemi örnek olarak konuyu açıklamak amacıyla işlenmiştir.

$$\text{Seçilen Fonksiyon: } y = -x^2 + 64x$$

Fonksiyonun maksimum noktası, kolaylıkla bulunabilir:

$$x=32 \text{ ve } y= 1024 \text{ noktası fonksiyonun mutlak maksimum noktasıdır.}$$

3.4.1 Başlangıç Kümesinin Belirlenmesi

Genel olarak başlangıç kümesi rasgele seçilmektedir. Tamamen rastlantısal seçilmiş bir kümenin doğal süreci en iyi yansıtan yaklaşım olduğu düşünülmektedir.

Güncel bilgisayar programlama dillerinin hemen hemen hepsinde bulunan Rasgele sayı üretme fonksiyonları kullanılarak başlangıç kümesi rahatlıkla oluşturulabilir. Oluşturulan küme ikili sayma düzeninde sembolize edilmektedir.

Bu bölümdeki örnek için altı üyeli rasgele bir küme oluşturulmuştur. Her üye beş haneli ikili sayma düzeninde verilmiştir:

Tablo 3.1 Başlangıç Kümesi

i : j	b4	b3	b2	b1	b0	x_i	y_i
1	1	1	0	0	1	25	975
2	1	1	1	0	0	28	1008
3	0	0	1	0	1	5	295
4	1	0	1	0	1	21	903
5	0	1	0	1	0	10	540
6	1	0	0	1	1	19	855

$$x_i = \sum_{j=0}^4 b_{ij} \cdot 2^j \quad (3.1)$$

$$y_i = -x_i^2 + 64x_i \quad (3.2)$$

3.4.2 Göreceli Uygunluk Değerlerinin Belirlenmesi

Amaç fonksiyon değerinin maksimize edilmesi olduğuna göre yukarıdaki altı üyeden y_i değeri en yüksek olanı en uygun olan üyedir. u_i değerleri ise oransal olarak uygunluk değerlerini vermektedir. Birinci üye toplam uygunluğun %21,3'ünü sağlarken üçüncü üye sadece %6,4'ünü sağlayabilmektedir.

Tablo 3.2 Uygunluk Değerleri

i : j	b4	b3	b2	b1	B0	x_i	y_i	u_i
1	1	1	0	0	1	25	975	0,213
2	1	1	1	0	0	28	1008	0,220
3	0	0	1	0	1	5	295	0,064
4	1	0	1	0	1	21	903	0,197
5	0	1	0	1	0	10	540	0,118
6	1	0	0	1	1	19	855	0,187
						Toplam(t)	4576	

Her bir elemanın yüzde olarak görece uygunluk değeri:

$$u_i = y_i / t \quad (3.3)$$

Bir sonraki nesilde uygunluk değeri yüksek olan üyenin yeni kümede bulunma olasılığı ve/veya bulunma oranı, uygunluk değeri daha düşük olan elemana göre daha fazla olacaktır.

Yukarıdaki örnekte uygunluk değerlerinin hesabı pozitif olan değerlerin büyüklüğü ile ilintilidir. Ancak negatif değerlerinde yer aldığı durumlarda, yukarıdaki t ve u_i değerleri hesaplama yöntemleri geçersizdir. Böylesi durumlarda farklı yöntemler izlenmektedir.

3.4.3 Yeni Kümeye Girecek Üyelerin Tespiti

Bu aşamada uygulanabilecek çok sayıda yöntemden en basiti olan, uygunluk oranını ağırlık olarak kabul ederek rasgele seçim yapılmasıdır. Daha önce de belirtildiği gibi uygunluğu yüksek olan elemanlar bir sonraki nesilde daha fazla yer alabileceklerdir. Yöntemin uygulanmasında uygunluk değerlerinin birikimli toplamaları kullanılacaktır.

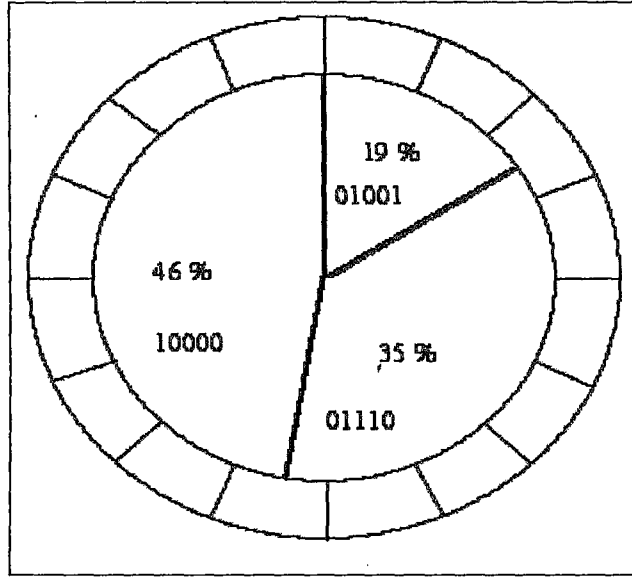
Rasgele belirlenecek altı eleman, 0-1 aralığındaki rasgele altı sayının hangi birikimli toplam değerine karşılık geldiği belirlenerek seçilecektir. Örneğin ilk rasgele sayı 0,554 olduğunda yeni kümenin ilk elemanı dördüncü eleman olacaktır. Çünkü rasgele sayı üçüncü üyenin k_{ui} değerinden büyük dördüncü elemanın k_{ui} değerinden ise küçüktür. Birikimli toplam değerleri sütununda 0,498 ile 0,695 aralığı dördüncü elemana ait alandır.

Tablo 3.3 Yeni Kümede Yer Alma Sayıları

i : j	b4	b3	b2	b1	b0	x_i	y_i	u_i	ku_i	Yeni Kümede Yer Alma Sayısı
1	1	1	0	0	1	25	975	0,213	0,213	1
2	1	1	1	0	0	28	1008	0,220	0,433	2
3	0	0	1	0	1	5	295	0,064	0,498	0
4	1	0	1	0	1	21	903	0,197	0,695	2
5	0	1	0	1	0	10	540	0,118	0,813	0
6	1	0	0	1	1	19	855	0,187	1,000	1
						Toplam	4576			

Kümülatif uygunluk değeri:

$$k_{ui} = k_{u(i-1)} + u_i \quad (k_{u0} = 0) \quad (3.4)$$



Şekil 3.1 Rulet Usulü İle Seçim⁴²

Şekil 3.1'de uygunluk değerleri referans alınarak tesadüfi seçim yapılması yöntemi görsel olarak ifade edilmeye çalışılmıştır.

3.4.4 Yeni Kümeye Girecek Üyeler Arasında Eşleşme

Yeni kümenin üyeleri Tablo 3.4'te belirtilmiştir:

Tablo 3.4 Eşleşmeye Tabi Tutulacak Üyeler

i : j	b4	b3	b2	b1	b0	x_i	y_i	u_i
1	1	1	0	0	1	25	975	0,213
2	1	1	1	0	0	28	1008	0,220
2	1	1	1	0	0	28	1008	0,220
4	1	0	1	0	1	21	903	0,197
4	1	0	1	0	1	21	903	0,197
6	1	0	0	1	1	19	855	0,187
						Toplam(t)	5652	

⁴² ANONİM: "Basic Genetic Algorithm Operations" [Çevrimiçi],
<<http://www.cs.qub.ac.uk/~M.Sullivan/ga/ga4.html>>, 22 Ocak 2003.

Tablo 3.4'te de gösterildiği gibi üç ve beş numaralı üyeler eşleşme işlemine girememişleridir, buna karşılık iki ve dört numaralı üyeler iki kez yer almışlardır. Açıkça anlaşılabacağı üzere uygunluğu yüksek olan üyelerin yeni kümede bulunma oranları ve/veya sayıları göreceli olarak daha fazladır.

Eşleşme işlemi belirli oranda üye çiftinin, rasgele olarak bir araya getirilmesi ve çocukların üretilmesidir. Çocuk üyeler genlerinin bir kısmını bir ebeveynden alırken diğer kısmını ise ikinci ebeveynden almaktadırlar. Yukarıdaki örnekte eşleşme oranı(Pe) 1.00 olarak alınacaktır, yani tüm üyeler eşleşme işlemine rastlantısal olarak tabi tutulacaklardır. Pe değeri [1.00, 0.00] aralığında seçilebilir.

Rasgele yapılan seçim sonrası eşleşme aşağıdaki gibi olmuştur:

1-4; 2-4; 2-6

Yine rasgele olarak seçilen kesim noktası (kn) belirlenerek her bir eşleşmenin yapılacağı j noktaları belirlenmelidir.

Verilen örnek için $0 < kn < 5$ sınırlandırması mevcuttur.

Örneğin 1-4 eşleşmesinde kesim noktasının 2 olduğu belirlendiğinde eşleşme sonrası durum Tablo 3.5'teki gibi olacaktır:

Tablo 3.5 Eşleşme(Crossover) İşlemi

Ebeveyn 1	1	1	0	0	1
Ebeveyn 4	1	0	1	0	1
Çocuk 14-1	1	1	1	0	1
Çocuk 14-2	1	0	0	0	1

Kümeden bir önceki nesil üyeler çıkarken o nesil tarafından üretilen yeni kuşak üyeler, kümedeki yerlerini alacaklardır.

Kümede daha yoğun olarak yer alan ve görece başarılı olan üyelerin genleri içerisindeki kromozom gruplarının bozulmadan bir sonraki nesle aktarılmalari olasılığı her zaman daha yüksektir ki bu durum nesiler arası geçişle daha iyi sonuçların elde edilmesini sağlar.

1-4; 2-4; 2-6;

Tablo 3.6 Eşleşme sonrası Yeni Üyeler

I : j	b4	b3	b2	b1	b0	xi	Yi	Ui
Çocuk 14-1	1	1	1	0	1	29	2697	0,21425
Çocuk 14-2	1	0	0	0	1	17	1377	0,10940
Çocuk 24-1	1	1	1	0	1	29	2697	0,21425
Çocuk 24-2	1	0	1	0	0	20	1680	0,13350
Çocuk 26-1	1	1	0	1	1	27	2457	0,19519
Çocuk 26-2	1	0	1	0	0	20	1680	0,13350
	16	8	4	2	1	Toplam(t)	12588	

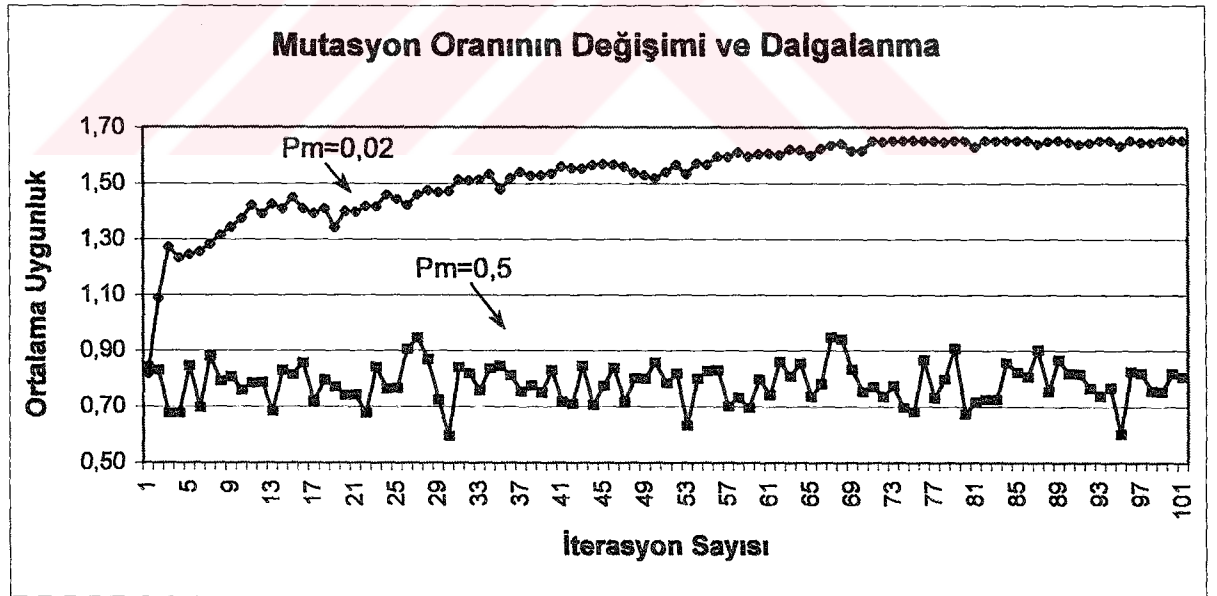
Tablo 3.2 Tablo 3.4 ve Tablo 3.6 incelendiğinde toplam Yi değerlerinin sürekli olarak arttığı gözlenebilir:

Tablo 3.7 GA Süreci Ortalama Fonksiyon Değerleri

Tablo Numarası	Ortalama Fonksiyon Değeri
Tablo 3.2: Başlangıç Kümesi	762,67
Tablo 3.4: Eşleşmeye Tabi Tutulacak Üyeler	942,00
Tablo 3.6: Eşleşme Sonrası Durum	2098,00

3.4.5 Mutasyon

Mutasyon, nesiller arası geçişte yaşanan ve kromozomların deformasyona uğramasıyla oluşan bir durumdur. Genetik algoritmalarda bu durum rasgele seçilen genlerin değerinin "0" ise "1" değerini alması "1" ise "0" değerini alması şeklinde uygulanır. Rastlantısal olarak mutasyona uğrayacak genlerin oranı (P_m) yine $[1.00, 0.00]$ aralığında olacaktır. P_m değerinin 1.00'a yaklaşması rastlantısallığı arttıracığı için çözümün bulunmasını zorlaştıracaktır. $P_m=1$ olduğunda ise tamamen şansa bırakılmış bir prosesle karşı karşıya kalınır. P_m 'in "0" seçilmesi ise yerel maksimum veya minimum noktalarının algoritmayı yanlış yönlendirmesinin engellenememesine yol açacaktır. 0.001- 0.01 aralığında seçilecek P_m değerinin optimuma daha yakın sonuçlar vereceği hem kullanılan kaynaklarda yer almaktadır hem de tarafımızdan yapılan 100 deneme ile gözlenmiştir.



Şekil 3.2 Mutasyon Oranının Dalgalanmaya Etkisi

Şekil 3.2'de de açıkça görüldüğü gibi mutasyon oranının büyük seçilmesi sonucunda ortalama uygunluk oranları daha değişken bir hal alırken aynı zamanda istenilen sonuca ulaşılması şansını da azaltmaktadır.

3.5 Yeni Kümeye Girecek Üyelerin Belirlenmesinde Uygulanabilecek Farklı Yöntemler

3.5.1 Uygunluk Değerleri Temel Alınarak Rasgele Seçim

Bu yöntemde uygunluk değerleri referans alınarak rulet usulü ile seçim yapılmaktadır. Bölüm 3.4.3'te bu yöntem ayrıntılı olarak açıklanmıştır.

3.5.2 Seçkincilik

Bu yöntem, kümedeki en iyi bireyin doğrudan bir şekilde yeni kuşakta yer almak üzere seçilmesini ön görür. Başlangıçta tesadüfi olarak seçilen üyelerden en iyisinin algoritmayı yanlış yönlendirerek yerel maksimum veya minimum noktaya yönlendirmesi riski bulunmasına rağmen bu yaklaşım, sonuca ulaşmayı genelde kolaylaştırır.⁴³

3.5.3 Deterministik Seçim

Bu tür seçimde, küme elemanları yüzde olarak belirlenen uygunluk değerleri oranında yeni kuşakta yer alırlar. Kusurlarlardan dolayı eksik kalan kısım olursa da eksiklik uygunluk oranı en büyük üyeden başlanarak üye sayısının tamamlanmasıyla tamamlanır.

⁴³ LIN T.C., LEE C. S. G.: a.g.e., s.393.

3.5.4 Kalanların Tesadüfi Olarak Tamamlandığı Deterministik Seçim

Bir önceki yönteme benzer şekilde, yeni üyeler % uygunluk değerlerinin tam kısımları esas alınarak seçilir. Kalan üyeler ise ondalık kısımları ağırlında rasgele seçilir.

Yukarıda anlatılan yöntemlere benzer veya bu yöntemlerin ortaklaşa kullanıldığı bir çok yöntem mevcuttur. Ancak tüm yöntemlerin amacı yeni nesle girecek üyeleri sağlıklı seçmektir.

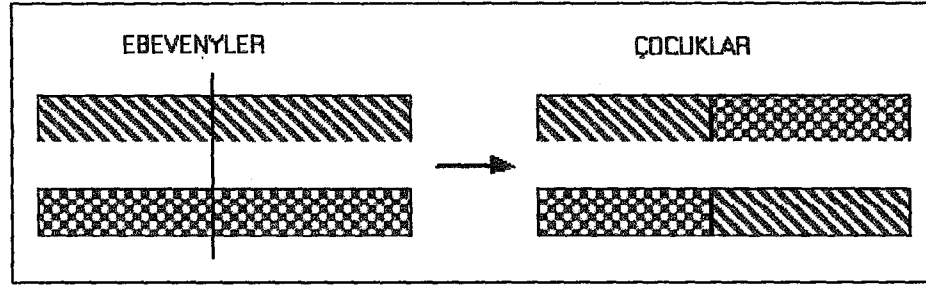
Yaratılan program kalanların tesadüfi olarak tamamlandığı deterministik seçim ve seçkinci seçimin birlikte kullanılması esasıyla çalışmaktadır. En iyi üye en kötü üyenin yerine geçtikten sonra, 100 üyeli bir grupta genellikle 50-75 arası üye uygunluk yüzdesinin tam kısmı ile yeni nesle geçerken 40-25 üye ise ondalık kısımları oranında tesadüfi olarak seçilmektedir.

3.6 Eşleşme(Crossover) İçin Uygulanabilecek Farklı Yöntemler⁴⁴

3.6.1 Tek Noktadan Kesişim

Bu yöntemde belirlenen bir kesim noktasına kadar ki genler bir ebeveynden diğer kısım ise diğer ebeveynden alınır. Şekil 3.3 bu durumu izah etmek üzere sunulmuştur.

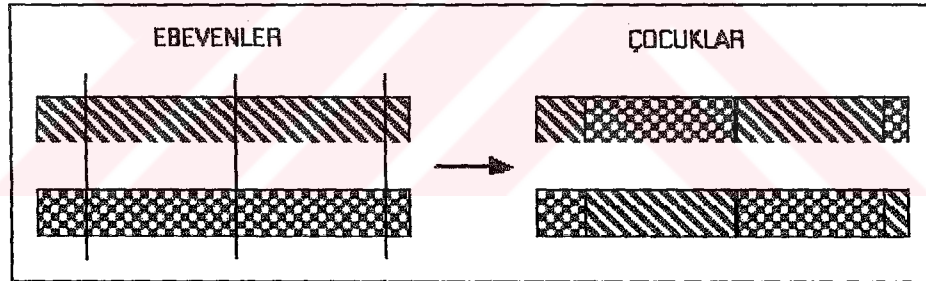
⁴⁴ POHLHEIM, H.: "Genetic and Evolutionary Algorithm Toolbox for use with *Matlab* (GEATbx)", [Çevrimiçi], <<http://www.geatbx.com/docu/algrecom.html#namerecombinationbinary>>, 22 Ocak 2003



Şekil 3.3 Tek Noktadan Kesişim

3.6.2 Çok Noktadan Kesişim

Tek noktadan kesişime benzer olan bu yöntemde birden çok nokta belirlenerek nesiller arası gen aktarımı sağlanır. Şekil 3.4 bu durumu ifade etmek üzere sunulmuştur.



Şekil 3.4 Çok Noktadan Kesişim

3.6.3 Çok Noktadan Tesadüfi Kesişim

Bu yöntemde her genin hangi ebeveynden geleceği rastlantısal olarak seçilir. Burada Kesişim sayısı gen sayısı kadardır.

Yukarıda açıklanan yöntemlere benzer çok sayıda eşleştirme operatörü belirlenebilir, ancak bütün yöntemlerin amacı, başarılı genlerin bir sonraki nesle olabildiğince bozulmadan aktarılmasının sağlanmasıdır.

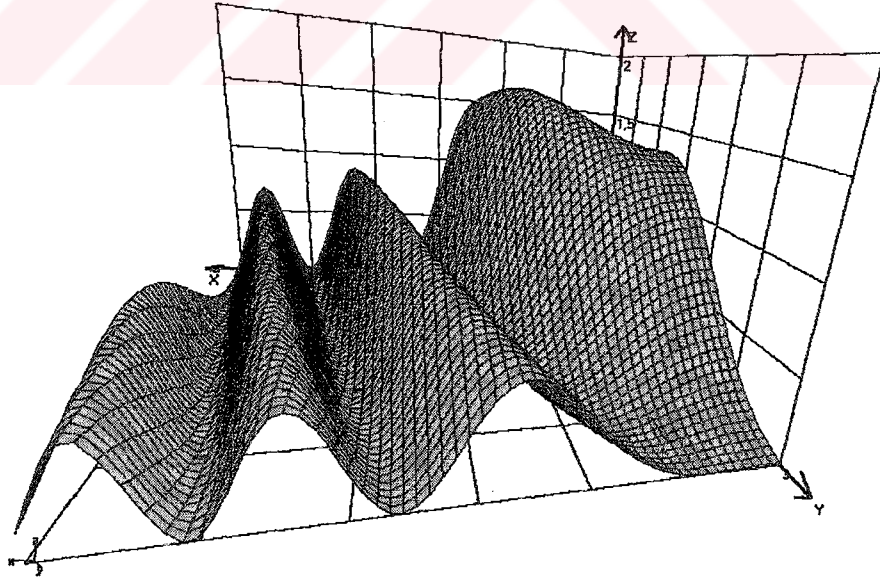
3.7 Genetik Algoritmaların Karmaşık Bir Örnekle İncelenmesi

Bu kısımda, iki bağımsız değişkenli doğrusal olmayan bir fonksiyonun maksimum değerinin bulunması problemi incelenecektir. İncelenecek fonksiyon aşağıda sunulmuştur:

$$f(x,y) = \frac{2 \cdot \sin^2(x^2 + y)}{\sqrt{x^2 + y^2} + 0,1} \quad (3.5)$$

[0.00 ,4.00] aralığında seçilecek olan x ve y değerleri ile seçilen fonksiyonun maksimum değeri bulunmaya çalışılmıştır. Genetik algoritmaların uygulamasını göstermek amacıyla Microsoft Excel Makro aracı kullanılarak Visual Basic ile geliştirilen program kullanılmıştır. Program ile elde edilen sonuçlar bu bölümde sunulmuştur. Programa ait kodlar da okuyucuların incelemesine olanak sağlamak amacıyla EK1'de listelenmiştir.

Demo



Şekil 3.5 Seçilen Fonksiyonun Üç Boyutlu Grafiği

3.7.1 Uygulamada Kullanılan Parametreler

Defalarca tekrarlanan denemeler sonucunda aşağıdaki parametrelerle genetik algoritmaların uygulanmasına karar verilmiştir.

Tablo 3.8 Uygulamada Kullanılan Parametreler

Üye Sayısı	100
Her Değişkenin İfade Edileceği Bit Sayısı	24
Değişken Sayısı:	2
Toplam Nesil Sayısı (İterasyon)	100
Eşleşme Oranı (Pn)	%95
Mutasyon Oranı (Pm)	%0,1
x,y aralığı	0,00-4,00

24 haneli ikili sayma düzenindeki bir sayı $[0 - (2^{24}-1)]$ aralığında değerler alır. Bu değerlerin $[0,00-4,00]$ aralığında ifade edilebilmesi için skala değişimi yapılmalıdır. İkili sayma düzenindeki bir sayı onluk düzene çevrildikten sonra aşağıdaki formülle kolaylık skala değişimine tabi tutulabilir:

Skala Değişimi Sonrası onlu değer:

$$od_{skala} = od \cdot \frac{U_{maks} - U_{min}}{2^b - 1} + U_{min} \quad (3.6)$$

- od : Skala değiştirme öncesi onluk sayma düzenindeki değer
- U_{maks} : Üst sınır değer
- U_{min} : Alt sınır değer
- od_{skala} : Alt ve üst sınırlar arasında skala edilmiş onlu sayma düzenindeki değer
- b : Bir gerçek sayıyı ifade etmek için kullanılacak ikili sayma düzeni hane sayısı

Bu işlem sonucu x ve y değerleri çok hassas bir şekilde ifade edilebilmektedir.

Tablo 3.9 x ve y Değerleri

	İkili Sayma Sistemi	Onlu Sayma Sistemi	0,00-4,00 Aralığı
x	10111010111001100011	6.008.419	1,43252
y	01101010111101110011	5.681.011	1,35446

3.7.2 Başlangıç Kümesi

Tamamen tesadüfi olarak seçilen başlangıç kümesi ve bu kümeye ait elde edilen değerler Tablo 3.10'da sunulmuştur.

Tablo 3.10 Başlangıç Kümesi ve Elde Edilen Değerler

İkili x	ikili y	x	y	z	Uygunluk
011000011011110001000000	1111111110000101010010100	1,5271	3,9850	0,0005	0,001
001100011010111000010111	010100101000111010110110	0,7763	1,2900	1,1212	2,244
001110100000000110101110	001111001100101010000011	0,9064	0,9499	1,3594	2,721
101110110010110100001101	100010000100110011100110	2,9246	2,1297	0,4871	0,975
100100011111100000000101	100100100010101010101110	2,2808	2,2839	0,5231	1,047
101001010000010000101001	110001101000011111100100	2,5784	3,1020	0,0494	0,099
001000101101110000001100	001010110000100111000111	0,5447	0,6725	1,4080	2,818
101000101101010000011101	011110010000101111010111	2,5442	1,8913	0,4657	0,932
111100010100000101110010	101000010000110001000100	3,7696	2,5164	0,3129	0,626
010110110010100001100110	010101010000100111110111	1,4243	1,3287	0,0448	0,090
000001000111011101101000	111110100000001011010101	0,0698	3,9064	0,2417	0,484
111110010000001000110010	101101111011110111101111	3,8908	2,8710	0,2249	0,450
101111111011111010111011	111110101011111001000000	2,9960	3,9179	0,0412	0,082
111000001011101010100110	110101110110010100000011	3,5114	3,3655	0,0001	0,000
110011011011110100101000	001000111011000000001000	3,2147	0,5576	0,5884	1,178
011110101000001111010111	010011001010010111011011	1,9143	1,1976	0,8293	1,660
110110011111001111100110	101100111100100010101110	3,4055	2,8091	0,4116	0,824

(Önceki Sayfadan Devam)

11100100101111110100000	111100110100001010000000	3,5742	3,8009	0,2189	0,438
001011001101101000011111	101111000010111101100111	0,7008	2,9404	0,0523	0,105
011111001110110101111000	110000111100100001010001	1,9520	3,0591	0,1641	0,329
111101000001101110100101	111001000101110000010110	3,8142	3,5681	0,1684	0,337
010110000101101001101101	001000111101110010011010	1,3805	0,5603	0,4917	0,984
101010000111010111000100	000010100100001111111100	2,6322	0,1604	0,3801	0,761
001010000010100001110100	101000000111011100010011	0,6275	2,5073	0,0423	0,085
100000100001011011110011	111100100011010110111000	2,0327	3,7845	0,4532	0,907
010000111001100001000011	101001100100010101000110	1,0562	2,5980	0,2017	0,404
011001101110110101001000	101000101001001111000111	1,6082	2,5403	0,5395	1,080
011110011101101110100101	100011000010100110100100	1,9040	2,1900	0,1355	0,271
000101000000011110110000	101001010011100011101110	0,3130	2,5816	0,1472	0,295
110011001000110010011110	101100100000110111001100	3,1961	2,7821	0,0804	0,161
000010011110000001101101	001101011001010111011111	0,1543	0,8373	1,2096	2,421
010010110000100011101111	000101011111000001110011	1,1724	0,3428	1,4811	2,965
010111101100111100011010	011010001000001110010010	1,4814	1,6330	0,3481	0,697
110001010101011101000001	110010000001000110011101	3,0835	3,1261	0,0020	0,004
001000001110011011000110	010010010000101100011010	0,5141	1,1413	1,4396	2,882
011110000001100010100000	111110101110010100001101	1,8765	3,9202	0,3775	0,756
111000100100000011001010	000010110011111101110010	3,5352	0,1757	0,0063	0,013
010100110111000100110100	101110100111110011110010	1,3038	2,9139	0,6016	1,204
000011101010001101111101	011001110110111011110001	0,2287	1,6161	1,1436	2,289
000000111101100000010010	010101110011100000010100	0,0601	1,3628	1,3097	2,622
10010111111110000010110	010101000110100000011110	2,3748	1,3189	0,2774	0,555
000101110110001110101000	110010010011100101001100	0,3655	3,1441	0,0113	0,023
111101100111101010111101	011000100011101101101000	3,8512	1,5349	0,1766	0,354
011001000110000101111111	010111111010110111000001	1,5685	1,4950	0,4659	0,933
001111010100100001110110	000100111001101000011011	0,9575	0,3063	1,5994	3,202
011001001010100111010011	101010100010000011011000	1,5729	2,6583	0,5230	1,047
010010100111010011101111	001001010110011011000001	1,1634	0,5844	1,2429	2,488
111010101000110111101011	110001001101110011001010	3,6649	3,0760	0,2105	0,421
000001101000111101111010	000000110001000001010110	0,1025	0,0479	0,0319	0,064
100111111001100110010101	100100001000111110101100	2,4937	2,2588	0,3804	0,761
010011110011011001101000	001001101110111101111100	1,2377	0,6084	0,9591	1,920

(Önceki Sayfadan Devam)

011111100001001110001000	111010011110101100111011	1,9699	3,6550	0,4243	0,849
110111011011100110010011	110001010100001101001101	3,4645	3,0822	0,1439	0,288
010110100010110001111011	111110101011001010000110	1,4090	3,9171	0,0648	0,130
100001011111100110111010	111010011000011101101001	2,0934	3,6489	0,4500	0,901
000100100010101110011000	010011000111010111001100	0,2839	1,1947	1,3783	2,759
100111011111101011011100	011011010011111000000111	2,4684	1,7069	0,6431	1,287
010001011101010110001010	101100110001001001100111	1,0912	2,7980	0,3619	0,724
110111110110101011001110	010111001010000110001101	3,4909	1,4474	0,3956	0,792
011010100110011100101010	101011100001001011101010	1,6625	2,7199	0,3126	0,626
110001000111001000111011	110001110001110111101010	3,0695	3,1112	0,0005	0,001
011011111111101101110000	001001111001001000011110	1,7497	0,6183	0,2687	0,538
100010011100000010001000	101100011011101101011011	2,1524	2,7771	0,4513	0,903
100110111100011100010111	011101011000001100101000	2,4340	1,8361	0,6296	1,260
010010000100000000000001	100100110010100001110001	1,1289	2,2993	0,1318	0,264
101111011001000011110000	001110011111101111111010	2,9620	0,9060	0,0396	0,079
001100000110010010111010	011100000100000000010010	0,7561	1,7539	0,5279	1,057
100100011110000011011111	101111110000010100101111	2,2794	2,9847	0,4655	0,932
000010100011001011110111	001111000000000100111010	0,1594	0,9376	1,2823	2,567
110100100100101101000010	111100000001100111100000	3,2858	3,7516	0,3303	0,661
010110011110111100100000	110000110011101101110010	1,4052	3,0505	0,5235	1,048
011000110101111000110001	011100001111000100011001	1,5526	1,7647	0,6026	1,206
000111110101010110101111	100010010111010011101010	0,4896	2,1478	0,4071	0,815
101101101110100111100011	010000011010111010011101	2,8580	1,0263	0,0332	0,066
110000011101111100111001	110111011101101100100111	3,0292	3,4665	0,0025	0,005
000010000100010010101011	111110101111110111101000	0,1292	3,9217	0,2542	0,509
001001110110111111010110	000101111111100000111101	0,6162	0,3745	1,1420	2,286
111111100101011010111110	101001000011110001000000	3,9740	2,5662	0,0918	0,184
101100110101011111001011	000001001100011001011100	2,8022	0,0746	0,6852	1,372
010100101000001101100000	101000111000011010001010	1,2893	2,5551	0,5228	1,047
001101110110000101011011	100100010011001110000111	0,8653	2,2688	0,0121	0,024
111001000100110100101111	100001101000100000100001	3,5672	2,1021	0,2806	0,562
100010001111000111000111	110011000111010101010011	2,1398	3,1947	0,5037	1,008
010001111111100001100110	110010110100100001110110	1,1245	3,1763	0,5350	1,071
101110101000010011101010	110100100000001001110111	2,9144	3,2814	0,2255	0,451

(Önceki Sayfadan Devam)

101010011100110111101101	111100101000111000010111	2,6532	3,7899	0,4116	0,824
000101111001000011101101	001001001111000011100010	0,3682	0,5772	1,0900	2,182
001110100010101110000101	011000000101001000010101	0,9089	1,5050	0,5651	1,131
001111100101000000101100	101100011011001111011110	0,9736	2,7766	0,1992	0,399
000001110000100100001010	010110001010110010011110	0,1099	1,3855	1,3025	2,607
011010111000110011011101	010110001000101000001100	1,6805	1,3834	0,6729	1,347
011001010001001000110101	011011010000010110010000	1,5792	1,7035	0,6253	1,252
010010010001101011110011	101011001011000111011111	1,1423	2,6984	0,3801	0,761
001011100100111110000011	001011001000010011011011	0,7236	0,6956	1,5971	3,197
101010101100010001001111	110110100011011101000001	2,6682	3,4096	0,3602	0,721
100011011000011010010010	000110100001111000111101	2,2113	0,4081	0,5913	1,184
000000001100110001001010	000101111101010001111110	0,0125	0,3723	0,5606	1,122
010111000000100100001110	101101000010000101011110	1,4381	2,8145	0,5958	1,193
001101110000100001011011	011011001101010110111110	0,8599	1,7005	0,4155	0,832
		0,6216	1,0557	1,4844	2,971

Tablo 3.11 Başlangıç Kümesi ve Küme Üyelerinin Yeni Nesilde Yer Alma Sayıları

x	y	z	Uygunluk	Deterministik Seçim	Ondalık Kısımlarla Tesadüfi Seçim
1,5271	3,9850	0,0005	0,001	0	0
0,7763	1,2900	1,1212	2,244	2	0
0,9064	0,9499	1,3594	2,721	2	0
2,9246	2,1297	0,4871	0,975	0	0
2,2808	2,2839	0,5231	1,047	1	0
2,5784	3,1020	0,0494	0,099	0	0
0,5447	0,6725	1,4080	2,818	2	1
2,5442	1,8913	0,4657	0,932	0	0
3,7696	2,5164	0,3129	0,626	0	1
1,4243	1,3287	0,0448	0,090	0	0
0,0698	3,9064	0,2417	0,484	0	1

(Önceki Sayfadan Devam)

3,8908	2,8710	0,2249	0,450	0	0
2,9960	3,9179	0,0412	0,082	0	0
3,5114	3,3655	0,0001	0,000	0	0
3,2147	0,5576	0,5884	1,178	1	1
1,9143	1,1976	0,8293	1,660	1	1
3,4055	2,8091	0,4116	0,824	0	2
3,5742	3,8009	0,2189	0,438	0	1
0,7008	2,9404	0,0523	0,105	0	0
1,9520	3,0591	0,1641	0,329	0	0
3,8142	3,5681	0,1684	0,337	0	0
1,3805	0,5603	0,4917	0,984	0	2
2,6322	0,1604	0,3801	0,761	0	1
0,6275	2,5073	0,0423	0,085	0	0
2,0327	3,7845	0,4532	0,907	0	0
1,0562	2,5980	0,2017	0,404	0	0
1,6082	2,5403	0,5395	1,080	1	0
1,9040	2,1900	0,1355	0,271	0	0
0,3130	2,5816	0,1472	0,295	0	0
3,1961	2,7821	0,0804	0,161	0	0
0,1543	0,8373	1,2096	2,421	2	1
1,1724	0,3428	1,4811	2,965	2	2
1,4814	1,6330	0,3481	0,697	0	0
3,0835	3,1261	0,0020	0,004	0	0
0,5141	1,1413	1,4396	2,882	2	1
1,8765	3,9202	0,3775	0,756	0	0
3,5352	0,1757	0,0063	0,013	0	0
1,3038	2,9139	0,6016	1,204	1	1
0,2287	1,6161	1,1436	2,289	2	0
0,0601	1,3628	1,3097	2,622	2	1
2,3748	1,3189	0,2774	0,555	0	2
0,3655	3,1441	0,0113	0,023	0	0
3,8512	1,5349	0,1766	0,354	0	2
1,5685	1,4950	0,4659	0,933	0	1
0,9575	0,3063	1,5994	3,202	3	0

(Önceki Sayfadan Devam)

1,5729	2,6583	0,5230	1,047	1	0
1,1634	0,5844	1,2429	2,488	2	1
3,6649	3,0760	0,2105	0,421	0	2
0,1025	0,0479	0,0319	0,064	0	0
2,4937	2,2588	0,3804	0,761	0	0
1,2377	0,6084	0,9591	1,920	1	0
1,9699	3,6550	0,4243	0,849	0	1
3,4645	3,0822	0,1439	0,288	0	0
1,4090	3,9171	0,0648	0,130	0	0
2,0934	3,6489	0,4500	0,901	0	0
0,2839	1,1947	1,3783	2,759	2	0
2,4684	1,7069	0,6431	1,287	1	0
1,0912	2,7980	0,3619	0,724	0	0
3,4909	1,4474	0,3956	0,792	0	0
1,6625	2,7199	0,3126	0,626	0	1
3,0695	3,1112	0,0005	0,001	0	0
1,7497	0,6183	0,2687	0,538	0	1
2,1524	2,7771	0,4513	0,903	0	2
2,4340	1,8361	0,6296	1,260	1	0
1,1289	2,2993	0,1318	0,264	0	0
2,9620	0,9060	0,0396	0,079	0	0
0,7561	1,7539	0,5279	1,057	1	0
2,2794	2,9847	0,4655	0,932	0	0
0,1594	0,9376	1,2823	2,567	2	2
3,2858	3,7516	0,3303	0,661	0	0
1,4052	3,0505	0,5235	1,048	1	0
1,5526	1,7647	0,6026	1,206	1	0
0,4896	2,1478	0,4071	0,815	0	2
2,8580	1,0263	0,0332	0,066	0	0
3,0292	3,4665	0,0025	0,005	0	0
0,1292	3,9217	0,2542	0,509	0	0
0,6162	0,3745	1,1420	2,286	2	1
3,9740	2,5662	0,0918	0,184	0	0
2,8022	0,0746	0,6852	1,372	1	0

(Önceki Sayfadan Devam)					
1,2893	2,5551	0,5228	1,047	1	0
0,8653	2,2688	0,0121	0,024	0	0
3,5672	2,1021	0,2806	0,562	0	1
2,1398	3,1947	0,5037	1,008	1	0
1,1245	3,1763	0,5350	1,071	1	0
2,9144	3,2814	0,2255	0,451	0	0
2,6532	3,7899	0,4116	0,824	0	1
0,3682	0,5772	1,0900	2,182	2	0
0,9089	1,5050	0,5651	1,131	1	0
0,9736	2,7766	0,1992	0,399	0	0
0,1099	1,3855	1,3025	2,607	2	0
1,6805	1,3834	0,6729	1,347	1	0
1,5792	1,7035	0,6253	1,252	1	0
1,1423	2,6984	0,3801	0,761	0	1
0,7236	0,6956	1,5971	3,197	3	2
2,6682	3,4096	0,3602	0,721	0	1
2,2113	0,4081	0,5913	1,184	1	0
0,0125	0,3723	0,5606	1,122	1	0
1,4381	2,8145	0,5958	1,193	1	0
0,8599	1,7005	0,4155	0,832	0	1
0,6216	1,0557	1,4844	2,971	2	0
Toplam:	49,9576	100		58	42

Bu aşamadan sonra eşleşme, mutasyon, fonksiyona ait değerlerin hesaplanması işlemleri döngüsel bir şekil belirtilen kriter gerçekleştirilene kadar devam ettirilmiş ve bölüm 3.7.3'te sunulan sonuçlar elde edilmiştir.

3.7.3 Algoritmanın Çalışması İle Elde Edilen Sonuçlar

Belirtilen fonksiyonun maksimum değerinin bulunmasına yönelik olarak genetik algoritmalar kullanılarak 200 kuşak sonra elde edilmiş sonuçlar Tablo 3.12'de sunulmuştur.

Tablo 3.12 200 Kuşak Sonra Elde Edilen Sonuçlar

Xikili	yikili	x	y	z
001110111010110010100101	001000000000010011111111	0,9324127	0,5003051	1,6579807
001110111000111100101011	001000000010011100110011	0,9306135	0,5023927	1,6579751
001110111001100010011101	001000000000011011111010	0,9311901	0,5004260	1,6579749
001110111001010011110101	001000000011001000111110	0,9309669	0,5030667	1,6579743
001110111001111001111001	001000000010110011111111	0,9315477	0,5027465	1,6579742
001110111000110011100110	001000000010001011111101	0,9304750	0,5021357	1,6579739
001110111000111011111000	001000000001101011111011	0,9306014	0,5016469	1,6579739
001110111000111011111001	001000000001010010011101	0,9306016	0,5012583	1,6579724
001110111000100000110101	001000000001110100110111	0,9301887	0,5035667	1,6579714
001110111100001001110001	001000000000010011111110	0,9337430	0,5003049	1,6579706
001110111011010010110001	001000000000110110111001	0,9329038	0,5016755	1,6579702
001110111000011001000001	001000000001111101111111	0,9300695	0,5038451	1,6579701
001110111010001001101101	001000000001101001011101	0,9317890	0,5032186	1,6579685
001110111100011000110100	001000000000011010110111	0,9339726	0,5004100	1,6579661
001110111000111010100101	001000000000010011111111	0,9305816	0,5003051	1,6579655
001110111000111001100001	001000000000010011111111	0,9305654	0,5003051	1,6579653
001110111010010011110011	001000000001111000111011	0,9319430	0,5036906	1,6579613
001110111100110000110001	001000000000011011110110	0,9343381	0,5004250	1,6579590
001110111000011000100001	001000000000011101111010	0,9300619	0,5004565	1,6579569
001110111010111000100101	0010000000011011110011100	0,9325042	0,5033943	1,6579561
001110111000011000110101	001000000000010011110011	0,9300666	0,5003022	1,6579552
001110111011111011110001	001000000010001101111100	0,9335294	0,5021659	1,6579533
001110111011111010100111	001000000010010000111011	0,9335118	0,5022115	1,6579530
001110111000010000111111	001000000011001010111010	0,9299469	0,5061928	1,6579518
001110111010111000100001	001000000001111110111101	0,9325033	0,5038747	1,6579487
001110111010010010100000	001000000010011101111101	0,9319232	0,5048207	1,6579447

(Önceki Sayfadan Devam)

001110111011001000110010	001000000011111000110110	0,9327515	0,5037972	1,6579446
001110111000110010110101	001000000111100011111100	0,9304638	0,5073845	1,6579252
001110111110011011110001	001000000000110111111011	0,9359708	0,5008535	1,6579024
001110111010110001101111	001000000110010111111111	0,9323998	0,5062255	1,6579019
001110111110001000101111	001000000010001111111101	0,9356804	0,5021967	1,6578818
001110111010101000100101	001000000111101011111010	0,9322601	0,5075060	1,6578708
001110111010111001110101	001000001000000011110111	0,9325233	0,5078715	1,6578486
001110111001100100101101	001000001010010011101111	0,9312244	0,5100669	1,6578272
001110110010110000110101	001000000100010001110111	0,9245735	0,5041789	1,6578097
001110110000111000111101	001000010000011011101100	0,9227443	0,5160476	1,6578003
001110110011001001100100	001000000010100001110111	0,9249509	0,5024699	1,6577853
001110110010110010100101	001000000010010001011111	0,9246002	0,5022201	1,6577552
001110110010101011100111	001000000010010001111111	0,9244938	0,5022277	1,6577483
001110110001110011100101	001000000011110101110011	0,9236389	0,5037507	1,6577400
001110110001100010100101	001000000010011010110111	0,9233795	0,5023631	1,6576727
001110111100110011100101	001000001001111001111111	0,9343810	0,5096740	1,6576675
001110110100000001100100	001000010100000000010100	0,9258054	0,5195362	1,6576514
001110111000011001101111	001000010000010011111010	0,9300804	0,5159289	1,6576226
001110101010111010110101	001000010100010000111011	0,9169135	0,5197896	1,6575970
001110110001100011110001	001000000000010011111110	0,9233976	0,5003048	1,6575858
001110101010011011110001	001000011000011001111110	0,9164396	0,5238339	1,6575456
001110110000010000100101	001000000001000011111101	0,9221282	0,5010370	1,6575046
001110111000010000100001	001000010010110011111011	0,9299398	0,5183705	1,6574884
001110110000011000110101	001000000000010011111111	0,9222541	0,5003051	1,6574788
001110101000111010110001	001000010010011111111111	0,9149595	0,5180663	1,6574624
001110101110101111100101	001000000010110011111100	0,9206481	0,5027458	1,6574454
001110101000110011100111	001000010000010001011010	0,9148503	0,5158907	1,6574066
001110111011001000100011	001000010001010011111010	0,9327479	0,5169055	1,6573559
001110101101111010101101	001000000001110001111111	0,9198413	0,5017394	1,6573018
001110111010111001110110	001000010010100111111111	0,9325235	0,5181884	1,6572818
001110111010010010110001	001000010011110001110110	0,9319272	0,5193154	1,6572529
001110111100010100110111	001000010001011111110110	0,9339123	0,5170876	1,6572265
001110101000110001110101	001000001011001010110001	0,9148231	0,5109066	1,6572195
001110101000011000100011	001000001011110001110110	0,9144373	0,5115029	1,6572034

(Önceki Sayfadan Devam)

001110111010011010001101	001000010100010001111111	0,9320407	0,5198058	1,6572027
001110111101011011101101	001000010000100111110101	0,9349933	0,5162329	1,6571776
001110111010110010100101	001000010100001001110110	0,9324127	0,5196816	1,6571746
001110101011101001100111	001000000011110011111101	0,9176274	0,5037226	1,6571579
001110111000111000110110	001000010110111010111111	0,9305551	0,5223846	1,6571405
001110101010111010110101	001000000100010000111011	0,9169135	0,5041646	1,6570950
001110101011111010100101	001000000000010011111010	0,9178863	0,5003039	1,6569534
001110101011111000111101	001000000000010000011100	0,9178615	0,5002510	1,6569458
001110110000110011100010	001000100010011101110111	0,9226616	0,5336589	1,6568829
001110101000110010110001	001000000100010011110000	0,9148374	0,5042078	1,6568076
001110101000011010101111	001000000011101101110101	0,9144707	0,5036291	1,6567062
001110111011111000111110	001000011000011011111101	0,9334867	0,5238642	1,6566537
001110101000111010110001	001000000000011011100111	0,9149595	0,5004214	1,6565100
001110011000111010100001	001000010111110011111001	0,8993307	0,5232529	1,6555802
001110111010011001100001	001000100011111011110110	0,9320302	0,5350930	1,6554903
001110111000010011100101	001000101000000011001100	0,9299865	0,5391113	1,6552741
001110011000110011110011	001000010010101101100100	0,8992281	0,5182735	1,6551273
001110111101010011111111	001000100010110011111101	0,9348755	0,5339960	1,6551037
001110011000000010111101	001000010010011011110110	0,8984828	0,5180031	1,6549323
001110011100110010100100	001000000010111101111101	0,9031155	0,5028986	1,6542260
001110101000001010100101	001001000001010011010001	0,9142242	0,5637707	1,6540115
001110010011110100110010	001000010000000011110110	0,8943603	0,5156838	1,6536113
001110010011001011100001	001000010001001001111110	0,8937307	0,5167538	1,6535781
001110011010111101111101	001000000000010101110111	0,9013362	0,5003337	1,6533346
001110011010110010110101	001000000000011011111110	0,9011665	0,5004269	1,6532996
001110011000010010100101	001000000011001011111111	0,8987212	0,5031127	1,6530068
001110011000111010110101	001000000000011011111010	0,8993354	0,5004260	1,6527422
001110011000111011111000	001000000000010011101001	0,8993514	0,5002998	1,6527253
001110010010111001110011	001000000111010011011011	0,8934603	0,5071324	1,6520199
001110111000110010111100	001000111011001110110101	0,9304650	0,5578436	1,6516366
001110111001110000000000	001000111011110010111101	0,9313968	0,5583948	1,6512391
001110010000110010101001	001000000011001011111000	0,8913980	0,5031110	1,6505401
001110111000100011100011	001001000010110111111111	0,9302302	0,5653075	1,6499072
001110010000010110110111	001000000000011011100001	0,8909741	0,5004200	1,6498163

(Önceki Sayfadan Devam)

001110111011001011100001	001001000000010011110100	0,9327932	0,5628025	1,6497135
001110111010010000100011	001001000010110111111001	0,9318934	0,5653061	1,6493610
001110001010000011110101	001000000100011011111101	0,8848243	0,5043329	1,6482119
001110111011001000101011	001001001011010101111100	0,9327498	0,5735771	1,6467156
001110101010001001100101	001001100011011011110111	0,9161620	0,5971049	1,6455919
001111110000100001101111	001000000000011010000001	0,9848900	0,5003971	1,6434619

Tablo 3.13'de ise Microsoft Excel araçlarından olan Solver (Çözücü) ile elde edilmiş sonuçlar listelenmiştir. Tablodan da görüleceği üzere [0,4] aralığında tesadüfi olarak seçilen x ve y değerlerinden çözücü programının bulduğu sonuçlar çok büyük oranda etkilenmektedir. Bulunan altı sonuçtan beşi tamamen anlamsızdır. Buna karşılık geliştirilen genetik algoritmalarla hemen hemen tüm denemelerde maksimum sonuca çok yakın değerler elde edilmiştir.

Çözücü aracı ile elde edilen en iyi değerle genetik algoritmalarla elde edilen kümenin en iyi 30 üyesi arasında kıyaslama yapıldığında maksimum fonksiyon değerleri arasındaki mutlak farkın 0.0001'den küçük olduğu görülmektedir.

Tablo 3.13 Microsoft Excel Çözücü Aracı İle Elde Edilen Sonuçlar

Deneme:	1	2	3	4	5	6
X Çözüm	3,236194	2,039187	3,690467	3,236162	0,932488	2,039192
Y Çözüm	0,499999	0,500025	0,500040	0,500214	0,500001	0,500007
Z Çözüm	0,592361	0,906601	0,522825	0,592361	1,657981	0,906601
X Başlangıç	2,387967	2,159959	3,565578	2,857684	1,253112	1,807425
Y Başlangıç	3,931955	0,571627	2,702619	3,270340	0,568237	2,793387

Ayrıca hem, MS Excel çözücü aracının, hem de GA ile elde edilen sonucun anlamlılığını test etmek amacıyla fonksiyonun kısmi türevleri alınmıştır ve çözüm olarak önerdiğimiz değişken değerleri, kısmi türev fonksiyonlarında kullanılarak türev değerlerinin sıfıra yakın olup olmadığı incelenmiştir:

$$\frac{\partial f(x,y)}{\partial x} = \frac{8x \sin(x^2 + y) \cos(x^2 + y) (\sqrt{x^2 + y^2} + 0,1) - \frac{2x \sin^2(x^2 + y)}{\sqrt{x^2 + y^2}}}{(\sqrt{x^2 + y^2} + 0,1)^2} \quad (3.7)$$

$$\frac{\partial f(x,y)}{\partial y} = \frac{4 \sin(x^2 + y) \cos(x^2 + y) (\sqrt{x^2 + y^2} + 0,1) - \frac{2y \sin^2(x^2 + y)}{\sqrt{x^2 + y^2}}}{(\sqrt{x^2 + y^2} + 0,1)^2} \quad (3.8)$$

Çözüm olarak bulunan değerler:

$$x=0,932488$$

$$y=0,500001$$

x ve y çözüm değerleri kısmi türev fonksiyonlarında yerine konduğunda türev değerleri aşağıdaki gibi bulunmuştur:

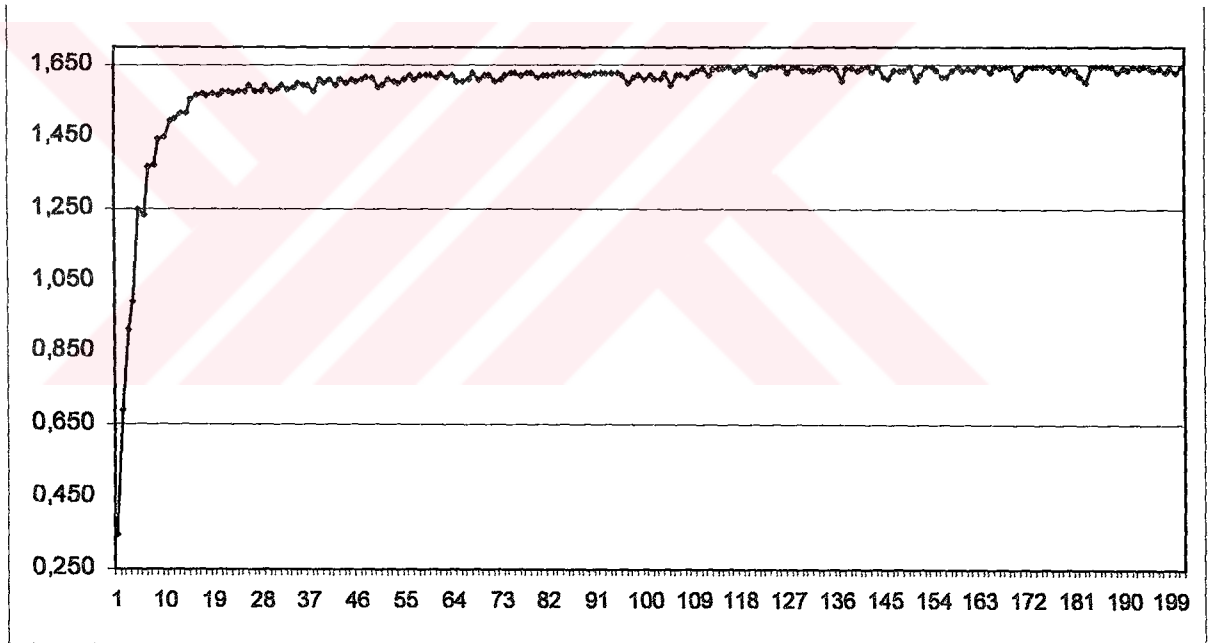
$$\frac{\partial f(x,y)}{\partial x} = 1,14 \cdot 10^{-7}$$

$$\frac{\partial f(x,y)}{\partial y} = 6,16 \cdot 10^{-8}$$

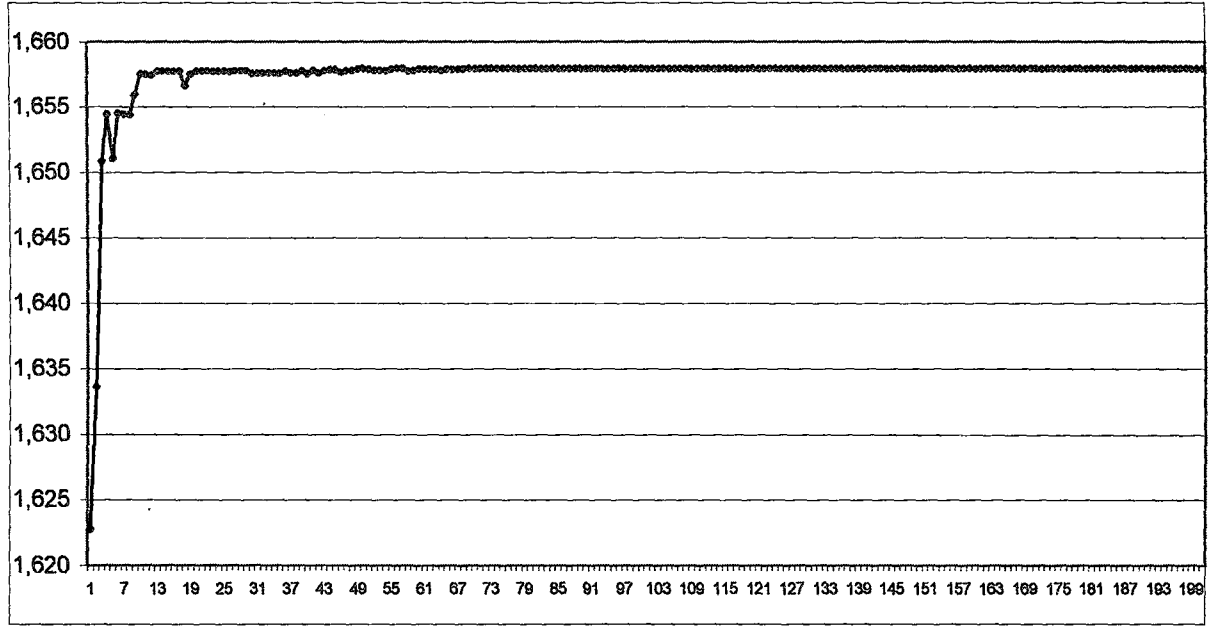
Görüldüğü gibi iki kısmi türev değeri de sıfıra çok yakındır. Yani çözüm olarak önerdiğimiz değerler, Şekil 3.5 ve kısmi türev değerleri göz önüne alındığında, fonksiyonu, verilen aralıkta maksimum yapan değerlerdir.

3.7.4 Sonuçların Yorumlanması

Genetik algoritmaların performansı tesadüfi olarak seçilen başlangıç kümesi, yeni kuşağa geçiş seçimi, eşleşme, mutasyon gibi faktörlerle değişmekle birlikte hemen hemen her denemede en iyi çözüme yakın sonuçlar elde edilebilmektedir. Bu problem için eklenebilecek bir modülle genetik algoritmanın elde ettiği sonuç kümesine çok küçük farklarla daha iyi sonuçlar arayan bir modülün eklenmesi de düşünülebilir. Böylelikle bulunan sonuca çok ince ayar yapılarak hata payının sıfıra daha da yakın olduğu sonuçlar elde edilebilir.



Şekil 3.6 Ortalama Fonksiyon Değerinin Nesle Bağlı Değişimi



Şekil 3.7 Maksimum Fonksiyon Değeri Nesle Bağlı Değişimi

Şekil 3.6 ve Şekil 3.7'den de görülebileceği gibi iterasyon sayısının çok yüksek değerlere çıkarılması, elde edilen ortalama fonksiyon değerini ve maksimum fonksiyon değerini fazlaca etkilememektedir. Yaklaşık 70. kuşaktan sonra sonuçların çok da geliştirilemediği görülmektedir. Bu durum genetik algoritmaların kısa sürelerde pratik anlamı olan sonuçların elde edilmesinde ne kadar etkili olabileceklerinin bir göstergesidir.

3.8 Genetik Algoritmaların Güçlü ve Zayıf Yanları

Her yöntem gibi GA da bazı yönleri ile çok kullanışlı iken bazı yönleriyle de uygulamada problemler yaratmaktadırlar:

3.8.1 Genetik Algoritmaların Güçlü Yanları⁴⁵

- i. Açıklanabilir sonuçlar elde edilebilir,
- ii. Sonuçların uygulaması kolaydır,
- iii. Çok çeşitli değişken tiplerine uygundurlar,
- iv. Optimizasyon problemlerinin çözümüne uygundurlar,
- v. YSA ile uyumlu çalışabilirler.

3.8.2 Genetik Algoritmaların Zayıf Yanları⁴⁶

- i. Bazı problemlerin modellenmesi zordur,
- ii. Optimum çözümün bulunması garanti edilemez,
- iii. Bilgisayar hesaplamaları çok fazla zaman ve kaynak kullanımı gerektirebilir,
- iv. Ticari paket programların sayısı azdır.

⁴⁵ BERRY, M. A., LINOFF, G. L.: "Data Mining Techniques For Marketing, Sales, and Customer Support", New York, NY: John Wiley & Sons, Inc., 1997, s.356.

⁴⁶ BERRY, M. A., LINOFF, G. L.: a.g.e., s.358.

BÖLÜM 4. YAPAY SİNİR AĞLARI

Yapay sinir ağları, canlıların sinir sistemi referans alınarak geliştirilen bilgisayar sistemi destekli algoritmalarıdır. Beynin ve sinir sisteminin çok karmaşık olması ve henüz tam anlaşılamamış olması nedeniyle geliştirilen bilgisayar destekli sistemler fizyolojik sinir sisteminin çok basit modelleri olarak algılanmaktadır.

4.1 Sinir Fizyolojisi ve Yapay Sinir Ağları

Yapay sinir ağlarının prensiplerini anlayabilmek için bu prensiplerin örnek alındığı fizyolojik sinir sistemi elemanlarının ve faaliyetlerinin kısaca özetlenmesi gerekmektedir.

Sinir sistemi, değişik duyu organlarından binlerce bilgi parçasığını alarak bütün bu verileri işler ve vücudun cevabını oluşturacak şekilde entegre eder.

Sinir sisteminin bölümleri:⁴⁷

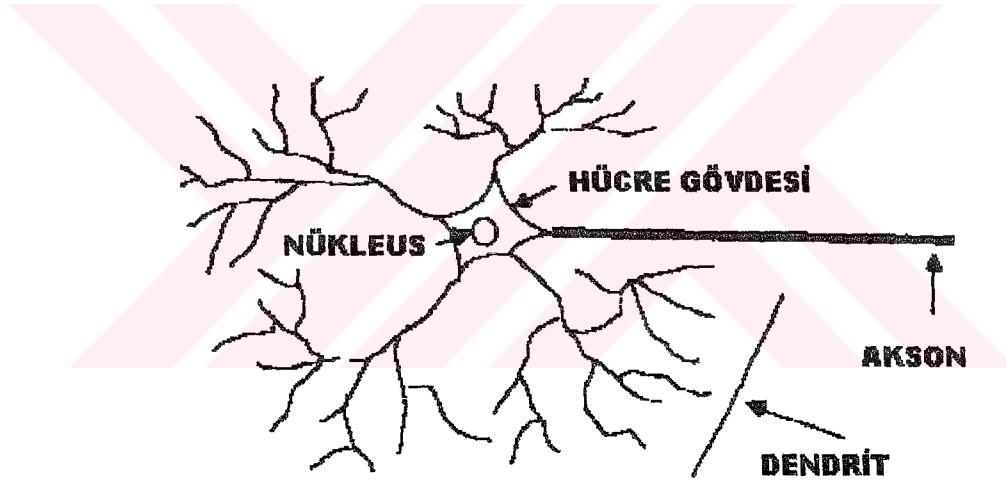
- Duyu Bölümü
- Motor Bölüm
- Enformasyon Hazırlanması
- Enformasyon Saklanması (Hafıza)

⁴⁷ GUYTON, A. C. : “Fizyoloji Cilt 2”, İngilizce 5. Baskıdan Türkçe'ye Çevrilmiş 1. Baskı, Güven Kitabevi Yayınları, Ankara, 1978, s. 323-325.

Beyinsel faaliyetlerin temelinde yer alan yukarıdaki dört faaliyet türünün gerçekleştirilmesinde nöronlar baş rolü oynamaktadırlar.

4.1.1 Nöronların Yapıları

Yapılan çalışmalar insan beyinde 10^{10} ile 10^{20} adet nöronun sinyal üreterek faaliyet gösterdiklerini saptamışlardır.⁴⁸ Sözü geçen nöronlar elektronik devre elemanları gibi davranarak bazı sinyalleri sonraki nörona aktarırlar. Bazılarını ise görmezden gelerek yok ederler. Şekil 4.1 çok basit bir şekilde nöronların yapısını anlatmak üzere sunulmuştur.



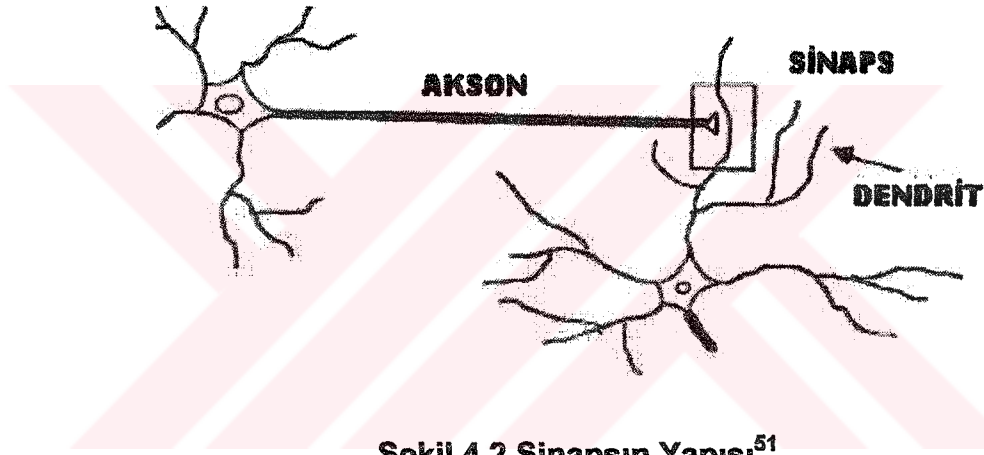
Şekil 4.1 Nöron Yapısı⁴⁹

⁴⁸ HARVEY, R. L.: "Neural Network Principles", New Jersey, NY: Prentice Hall., 1994, s.14.

⁴⁹ STERGIOU, C., SIGANOS, D.: "Neural Networks", [Çevrimiçi],
<http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html>, 11 Ağustos 2002

Klasik bir nöronda bulunan temel elemanlar:

1. Soma: Nöronun gövdesine verilen addır
2. Dendritler: Bir nöronda çok sayıda bulunan dendritler, nöronların alıcı uçlarıdır.
3. Akson: Gövdeden uzanan kısımdır. Aynı zamanda elektromekanik tepkimelerin gerçekleştiği alandır.
4. Snaps: Bir nöron ile bitişindeki nöron arasındaki bağlantı sinaps olarak tanımlanmaktadır.⁵⁰



Şekil 4.2 Sinapsın Yapısı⁵¹

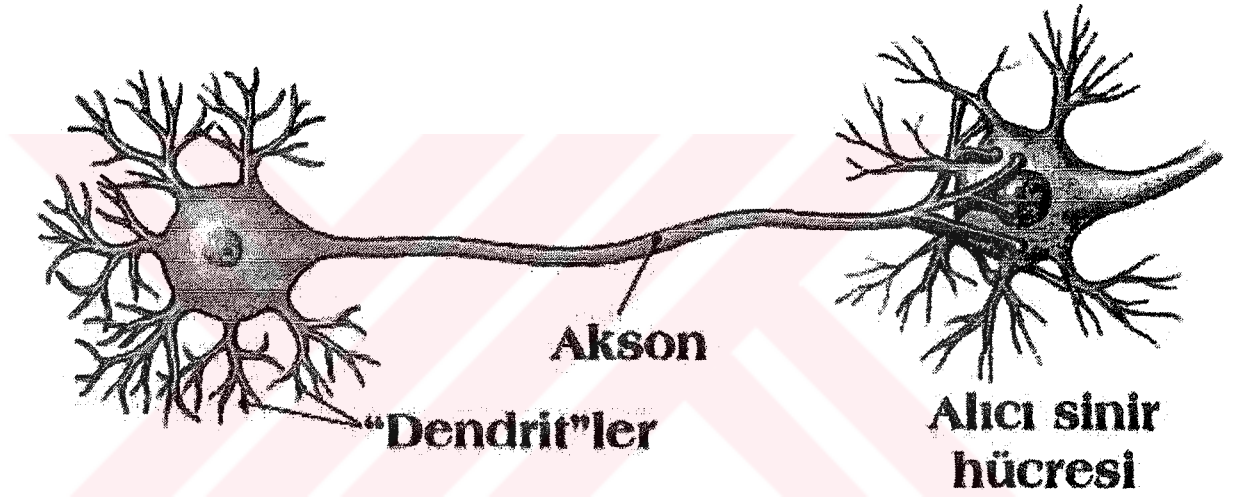
Bu iletişim **aksiyon potansiyelleri** denilen elektriksel sinyallerle sağlanır. Her nöron, **dendritleri** aracılığı ile diğer nöronlardan gelen uyarıları alır ve nöron gövdesine iletir. Bu uyarılar ya inhibitör ya da eksitatör niteliktedir. Değişik kaynaklardan gelen zıt yönlü uyarıların nöron gövdesindeki toplamına göre, nöron eksite ya da inhibe olur ve bu enformasyon nöronun aksonu

⁵⁰ GUYTON: a.g.e., s.330.

⁵¹ STERGIOU AND SIGANOS : a.g.e., [Çevrimiçi],
<http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html>, 11 Ağustos 2002

aracılığıyla bağlı olduğu nöronlara ya da efektör organa (kas, salgı bezi) iletilir.⁵²

Nöronlar sadece gerekler yerine geldiğinde aldıkları bir sinyali diğer nörona aktarırlar. Çalışma prensipleri 0-1 şeklindedir. Yani bir sinyalin gücü eşik değere eşit veya büyükse sinyal iletilir, aksi halde iletilmez. Sinyal gücü ise mili volt mertebelerindeki elektriksel potansiyel farkıdır. Bir nörona birden çok sinyal gelir ve aynı bilgi için gelen çok sayıda olabilen sinyaller toplam olarak değerlendirilir.



Şekil 4.3 Nöronlar arası ilişkinin Şematik Gösterimi⁵³

İstirahat halindeki bir nöronun sahip olduğu elektriksel potansiyel yaklaşık – 70 mV'tur.⁵⁴ Bu mertebedeki potansiyel, eşik değere kadar yükseldiğinde nöron uyarılmış olmaktadır. Örneğin –59mV değerine ulaşmış bir potansiyel,

⁵² EVLİCE, Y.E.: "Davranışın Nörobiyolojik Temelleri", [Çevrimiçi], <<http://lokman.cu.edu.tr/psychiatry/DERSNOT/1/1davrnbn.htm>>, 2 Şubat 2003

⁵³ ANONİM: "Sinir sistemi, insan sinir sisteminin yapısı ve işleyişi", [Çevrimiçi], <http://sinancanan.tripod.com/mss.htm>, 5 Temmuz 2002

⁵⁴ GUYTON, A. C.: a.g.e., s.336.

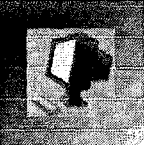
nöronu uyarılmış hale getirmektedir. Bu tür bir potansiyel değişimi çok sayıda uyarıcı nöronun sinyal göndermesiyle sağlanabilmektedir.

Uyarılmış bir nöron aldığı sinyali ilettikten sonra tekrar istirahat haline geri döner. Bu elektro kimyasal olaylar akson içi ve dışındaki soydum ve potasyum iyonlarının derişiminin değişimi ile sağlanır. Potansiyel farkını eşik değerin altında etkileyen sinyaller ise nöron tarafından yok sayılır ve sonraki nörona iletilmez.

4.1.2 Bilgisayarlarla Beyin Özelliklerinin Karşılaştırılması

Tablo 4.1 elektronik sistemlerle beyin yapısının özelliklerini karşılamak amacıyla sunulmuştur:

Tablo 4.1 Bilgisayarla Beyin Özelliklerinin Karşılaştırılması⁵⁵

	Proses Elemanı	Eleman Boyutu	Enerji Kullanımı	Proses Hızı	Hesap Yöntemi	Hata Toleransı	Öğrenme	Zekilik, Ruhsal Gelişmişlik
	10 ¹⁴ sinapsis	10 ⁻⁶ m	30 W	100 Hz	Paralel, dağıtılmış	Evet	Evet	Genellikle
	10 ⁸ transistor	10 ⁻⁸ m	30 W (CPU)	10 ⁹ Hz	Seri, Merkezi	Hayır	Az Miktarla	Hayır (Şimdilik)

Bilgisayarla ile Beyin yapısı sistem mantığı ile kıyaslandığında, bilgisayarların üstün oldukları belki de tek özellik işlem hızıdır. Yapay zeka çalışmalarında ağırlıkla üzerinde durulan bu üstünlük sayesinde, çok basit de olsa beynin

⁵⁵ ORR, G., SCHRAUDOLPH, N., CUMMINS, F.: "Computation in the brain, The Brain as an Information Processing System", [Çevrimiçi], <<http://www.willamette.edu/~gorr/classes/cs449/brain.html>>, 6 şubat 2003

alışmasının ve zekanın ok basit modellerle taklit edilmesi saėlanabilmiřtir. Geline gelişme dzeyi beyin ile yarıřabilecek noktadan hala ok uzaktadır.

4.2 Tarihsel Geliřim

Yapay sinir aėlarının tarihi, doėal olarak modern bilgisayar sistemlerinin geliřtirilmesine paraleldir.

YSA'larla ilgili ilk alışmalar 1943 yılında McCulloch ve Walter Pits tarafından gerekleřtirilmiřtir. Yapılan alışmanın amacı, beyin anatomisinin anlařılması ve bazı problemlerin özmnde yeni bir yaklařım geliřtirilmesidir.⁵⁶

Bu alışma ile nron fizyolojisini rnek alan sistemlerin eřitli problemlerin özmnde kullanılabilecekleri gsterilmiřtir.

1959 yılında Donald Hebb tarafından YSA'ların sadece nceden programlama ile deėil aynı zamanda ėrenme mekanizmasıyla donatılmıř olması mantıėı ortaya atılmıřtır.⁵⁷ Bu mantık, YSA algoritmalarının kullanım alanlarını ve glerini nemli oranda geliřmesine yol aan bir yaklařımdır.

Hebb'in ėrenme modelini takip eden alışmalar, Rosenblatt tarafından ortaya atılan Perception fikridir. Arkasından Widrow tarafından geliřtirilen Adaline model nemli kře tařlarından birisi olmuřtur. Ancak Perceptionun XOR iřlemini gerekleřtiremediėi ispatlandıktan sonra 1980'li yıllara kadar bir

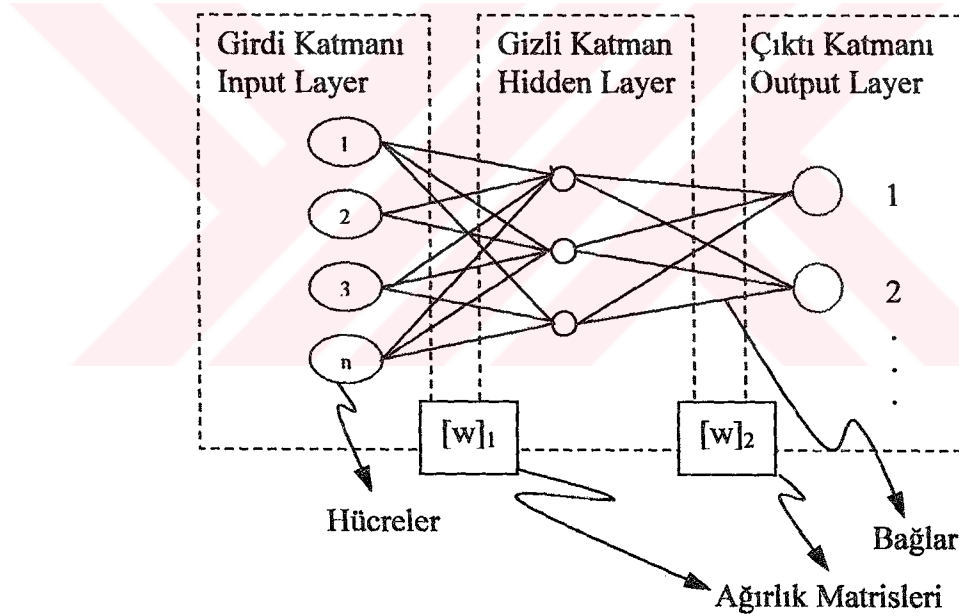
⁵⁶ BERRY, M. A., LINOFF, G. L.: a.g.e., 287.

⁵⁷ AKPINAR, H. : "Yapay Sinir Aėları ve Kredi Taleplerinin Deėerlendirilmesinde Bir Uygulama nerisi", İ İřletme Fakltesi, İstanbul, Mayıs 1993, [evrimii], <<http://www.isletme.istanbul.edu.tr/akpinar/content/Yapay%20Sinir%20Aglar.pdf>>, 20 Nisan 2003, s.13.

duraklama dönemi yaşanmıştır. Bu tarihten sonra John Hopfield tarafından gerçekleştirilen çalışmalar YSA'ların popülerliğini tekrar artırmıştır. Hopfield'in ortaya attığı geri yayılım prensibi, bir çok modelin temelini oluşturmuştur.⁵⁸

4.3 Yapay Sinir Ağlarının Temel Elemanları

Çok çeşitli topolojilere sahip YSA'lar mevcuttur. Ancak tüm modellerin esin kaynağı olan fizyolojik sinir ağları ile benzerliği açıklamak amacıyla basit bir YSA'nın temel yapısı Şekil 4.4'te sunulmuştur.



Şekil 4.4 Yapay Sinir Ağlarının Temel Elemanları⁵⁹

⁵⁸ BERRY, M. A., LINOFF, G. L.: a.g.e., s. 287.

⁵⁹ BIGUS J. P.: "Data Mining With Neural Networks", New York, NY: McGraw- Hill., 1996, s.63.

Şekil 4.4'ten de görülebileceği gibi YSA'larını oluşturan elemanlar: Girdi katmanı, Gizli katman (Gizli katman Sayısı sıfır, bir veya daha fazla olabilir), Çıktı katmanı, katmanlar arası bağlar ve bu bağları oluşturan ağırlık değerleridir. Son tahlilde hemen hemen tüm sinir ağı modelleri yukarıdaki şematik gösterime benzer yapılarla değişik öğrenme ve ağ oluşturma mantıklarıyla çalışırlar. Burada temel mantık; girdi katmanı verilerinden yola çıkarak katmanlar arası ilişkilerle ağırlık değerlerini elde etmek yoluyla çıktı değerlerine ulaşmak, böylelikle ağın öğrenmesini sağlamaktır. Burada her hücreye(nörona) birden çok sinyal gelir ve bu sinyaller birleştirilerek sonraki katmana tekil olarak gönderilir.

4.4 Yapay Sinir Ağlarında Öğrenme

Daha önce de belirtildiği gibi YSA'ların temel özelliklerinden birisi, öğrenmenin gerçekleşebilmesidir. Gerçek veriler kullanılarak Eğitilen (Train) modeller, modelin geçerliliğinin testinden sonra öğrenme sürecinde elde edilen bilgileri kullanarak, verilen girdilerle sonuç yada çıktı oluşturabilmektedirler.

Öğrenme modelleri temelde ikiye ayrılabilir:

Denetimli Öğrenme

Denetimsiz Öğrenme

Denetimli öğrenme modelleri nispeten kolay ve anlaşılabilir olanlarıdır. Yukarıda sözü geçen eğitim işlevi sırasında seçilecek örnekler ve diğer tüm parametreler insanlar tarafından sisteme verilir. Model aldığı kurallar ışığında öğrenme sürecini denetimli olarak gerçekleştirir.

Denetimsiz öğrenme modellerinde ise modele sadece girdiler verilir, bu girdilerden hangilerinin ne şekilde kullanılacağı ve diğer parametreler öğrenme süreci içerisinde model tarafından keşfedilir. Her iki öğrenme türünün arasında kalan denetimsiz fakat yönlendirilmiş öğrenmede ise modele öğrenme sırasında çıktıların beklenen değerleri hakkında bilgi verilmektedir.⁶⁰

Çeşitli öğrenme mekanizmaları mevcuttur bunlardan bazıları:

Hebb Kuralı

Hopfield Ağı / Geri Beslemeli Öğrenme

İleri Beslemeli Öğrenme

En Küçük Kareler ve Delta Kuralı

4.4.1 Hebb Kuralı

Hebb öğrenme kuralı, ilk ve en bilinen kuraldır. Güncel olarak kullanılan tüm kurallar Hebb'in ortaya attığı ilkeleri temel almaktadırlar. Donald O. Hebb bu kuramı, 1949 yılında yazdığı Davranışın Organizasyonu (Organization of Behavior) isimli kitabında açıklamıştır.⁶¹ Öğrenmenin temel kuralı olarak Hebb bir sinir hücresine gelen uyarıcı sinyallerin birlikte çıktı oluşturdıklarını ve yeterli tekrarın yapılması halinde uyarıcı hücrelerden bir kısmının ağırlık değerinin yeterli hale gelmesiyle uyarmayı gerçekleştirebileceklerini savunmuştur. Haldun Akpınar bu konuyu, "Yapay sinir Ağları ve Kredi Taleplerinin Değerlendirilmesinde Bir Uygulama Önerisi" isimli eserinde şartlı refleksle açıklamıştır. Deneklere zil sesi ile birlikte et gösterilmesi ile saldı

⁶⁰ BOSE, N. K., LIANG, P.: "Neural Network Fundamentals With Graphs, Algorithms, And Applications", New York, NY: McGraw- Hill., 1996, s.192, 193.

⁶¹ HOU, C.: "Digital Simulation of Artificial Neural Networks for Business Applications", Yüksek Lisans Tezi, San Diego State University, Sandiego, 1989, s. 29.

oluşması deneyinin sürekli tekrarlanması sonucu gerçek uyarıcının(et) olmadığı durumda dahi deneklerin, zil sesi ile tepki vermeleri örneği ile konuyu açıklamıştır.

4.4.2 Geri ve İleri Beslemeli Öğrenme

Geri beslemede her nöron herhangi başka nöronla simetrik bağlantıda bulunabilir. Adaline ve Hopfield ağları buna temel örneklerdir. Temelde geri beslemeli ağlar tek katmanlıdır. İleri beslemeli ağlarda ise bağlar sınırlandırılmıştır. Her nöron sadece sonraki katman nöronları ile bağlıdır ve bağ asimetriktir. Yani girdi katmanı hücreleri bir sonraki gizli katmanla bağlıdır. O gizli katman elemanları da varsa bir sonraki gizli katmanla bağlıdır. Bu durum son gizli katman nöronlarının çıktı katmanla bağlanması ile sonuçlanır. Yani bağlar ileri doğru gerçekleşir. Bu çalışmamızda kullanılan model de çok katmanlı Yani gizli katman içeren ileri beslemeli ve geri yayımlı ağ modelidir. Geri yayılım bölüm 4.4.4'de açıklanacaktır.

4.4.3 En Küçük Kareler, Delta Kuralı ve Geri Yayılım

En küçük kareler yöntemi, herhangi bir ağ ile elde edilen çıktı ile olması gereken çıktı arasındaki farkın, hatanın en aza indirilmesi için kullanılan bir yöntemdir. Regresyon analizinin temelinde yatan mantığa benzer şekilde hataların kareleri toplamını minimum yapacak ağırlık matrisleri elde edilmeye çalışılır. Delta kuralı olarak ve çok yaygın bilinen yöntem de yukarıda açıklanan en küçük kareler yöntemini referans alan bir yöntemdir ve Bernard Widrow ve Ted Hoff tarafından Stanford Üniversitesinde 1960 yılında geliştirilmiştir.⁶²

⁶² HOU, C.: a.g.e., s.37,38.

Geri yayılım, bir çok modelde kullanılan pratik ve popüler bir yöntemdir. Elde edilen ve gerçek çıktı değerleri arasındaki hatanın geriye doğru girdi katmanına kadar yayılımı prensibine dayanır. Son katman olan çıktı katmanında oluşan hatanın her bağa yada düğüme karşılık gelen kısmı geriye doğru girdi katmanına kadar dağıtılarak, hatayı minimize edecek bir yöntem izlenerek ağırlık değerleri tekrar hesaplanır. Bu işlem yeterince tekrarlandığında artık ağ öğrenme sürecini tamamlamış olur. Burada ki yeterli tekrar, baştan tanımlanan bir maksimum iterasyon sayısı olabilir ya da kabul edilebilir maksimum hata kareleri toplamı olabilir. En küçük kareler, delta kuralı ve geri yayılım ile ilgili ayrıntılı bilgi Bölüm 4.5'te verilmiştir

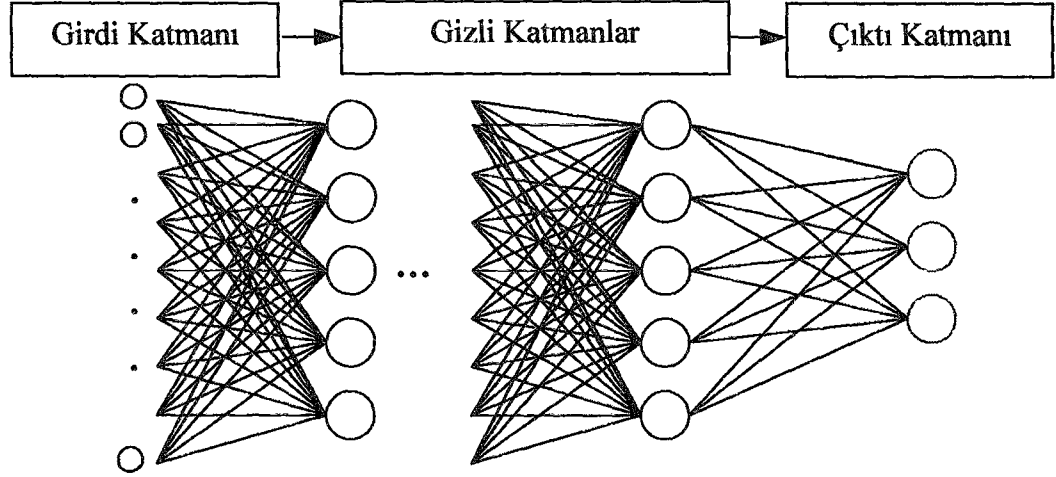
4.5 Çok Katmanlı İleri Beslemeli Geri Yayılımlı Yapay Sinir Ağları

En yaygın kullanılan YSA modellerinden birisi olan İleri beslemeli geri yayılımlı ağlar (Feedforward- Back propagation), denetimli öğrenme prensibi ile çalışan, genelde çok katmanlı ve asimetrik modellerdir.

4.5.1. ÇKİBGYYSA'nın Genel Yapısı

Şekil 4.5, sözü geçen YSA modellerini ve genel yapılarını açıklamak üzere sunulmuştur. Şekilden de anlaşılacağı üzere, Girdi katmanından başlayan süreç, gizli katman(lar) üzerinden tek yönlü olarak ilerleyerek çıktı katmanına kadar ulaşır.

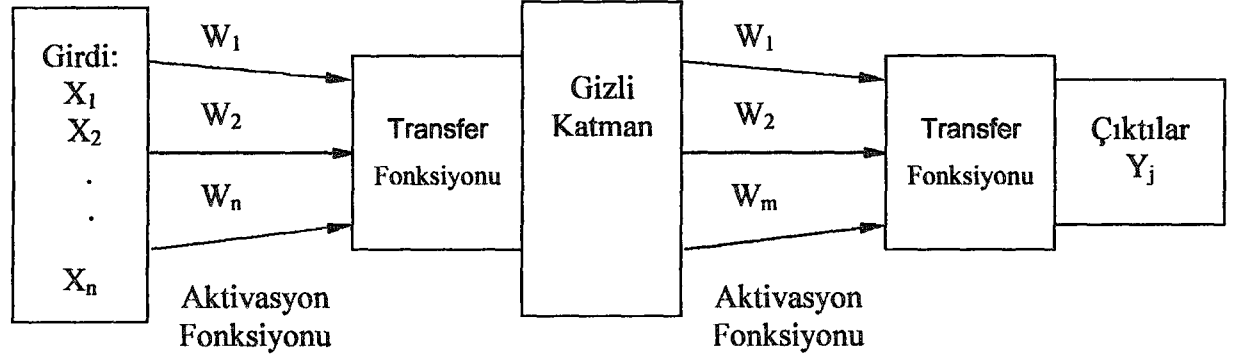
Buradaki işleyiş Şekil 4.4'de sunulan yapıya benzerdir ancak buradaki özel durum tek yönlü, asimetrik süreçtir.



Şekil 4.5 ÇKİBGYYSA Genel Yapısı⁶³

Bu tür ağların faaliyetleri sırasında gerçekleştirilen işlemler, daha öncede anlatıldığı gibi; ağırlık matrisi değerleri ile girdi katmanından başlanarak çıktı katmanına doğru tek yönlü ilerlenerek, çıktı değerlerinin hesaplanması ve öğrenme sürecinde de ağırlık değerlerinin geriye yayılım şekliyle tekrar düzenlenmesidir. Bu bölümde, sözü geçen ağların prensipleri ve ayrıntıları incelenmiştir.

⁶³ ROGERS, J.: "Object Oriented Neural Networks in C++", San Diego, CA: Academic Press Inc., 1997, s.29.



Şekil 4.6 YSA Katmanları ve Fonksiyonlar

4.5.2 ÇKİBGYYSA'nda Kullanılan Temel Fonksiyonlar

Çeşitli formlarda olabilen transfer ve aktivasyon fonksiyonlarından bu çalışma kapsamında yaygın kullanılan fonksiyonlar ele alınacaktır:

Aktivasyon Fonksiyonu (AF): Toplam Fonksiyonu⁶⁴

$$AF = \sum_{i=0, j=1}^{n, m} x_i \cdot w_{ij} \quad (4.1)$$

⁶⁴ NILSSON, N. J.: "Artificial Intelligence A New Synthesis", San Francisco, CA, Morgan Kaufmann, 1998, s.47.

Transfer Fonksiyonu(TF): Sigmoid Fonksiyon⁶⁵

$$TF = \frac{1}{1 + e^{-\lambda \cdot AFn,m}} \quad (4.2)$$

n: Toplam Girdi sayısı

m: Herhangi Gizli Katmanın Eleman Sayısı

x: Girdi Değeri

w: Ağırlık Değeri

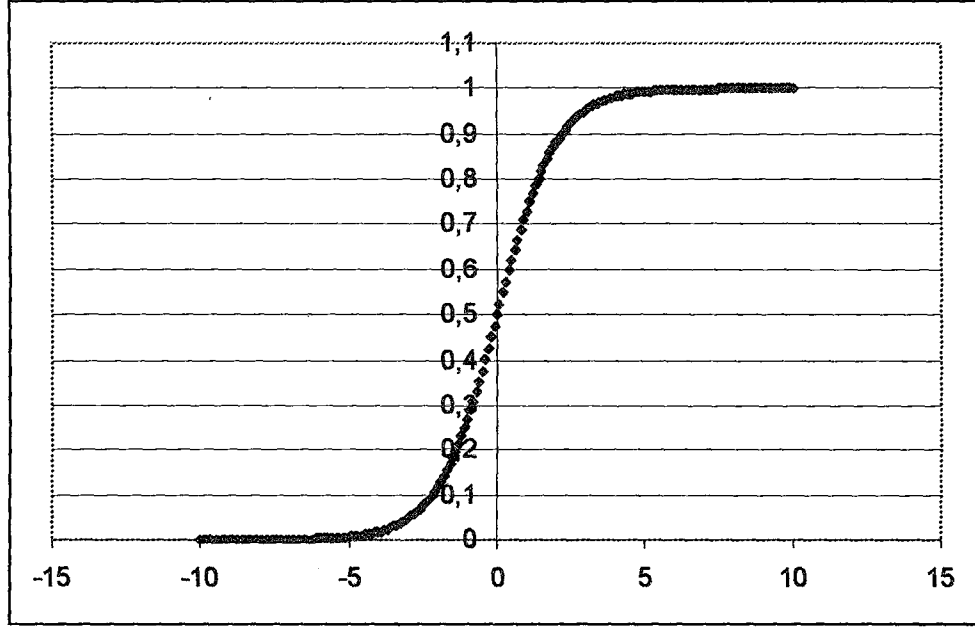
$x_0 = -1$

λ : Bu çalışma kapsamında 1 değeri kullanılmıştır.

4.5.2.1 Sigmoid Fonksiyonu

Şekil 4.7, YSA uygulamalarında sıklıkla kullanılan sigmoid fonksiyonun değerinin değişimini, $(-\infty, +\infty)$ tanım aralığında göstermek üzere sunulmuştur. Fonksiyonun değeri değerinin $[0,1]$ aralığında değişmektedir.

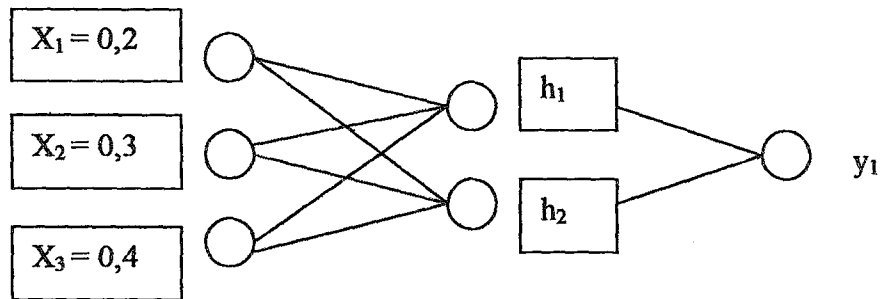
⁶⁵ ÇAĞIL, G.: a.g.e., s.50.



Şekil 4.7 Sigmoid Fonksiyonunun Grafiği

4.5.3 ÇKİBGYYSA'nda Kullanılan Yöntemin Açıklanması

Şekil 4.8, daha önce Şekil 4.4 ve 4.5'te gösterilen yapıya ilişkin bir örnek olarak sunulmuştur. Bu tür ağların parametrik değerlerini hesaplamak için yapılan işlemler, bu basit uygulama ile gösterilmiştir:



Şekil 4.8 Örnek Uygulama

w_{ij} : i numaralı girdi elemanından j numaralı gizli katmana doğru olan ağırlık değeri

$$\begin{array}{lll} w_{11}= 0,1 & w_{21}= 0,2 & w_{31}= 0,3 \\ w_{12}= 0,6 & w_{22}= 0,7 & w_{32}= 0,8 \end{array}$$

$$AF_1 = (0,2 \cdot 0,1 + 0,3 \cdot 0,2 + 0,4 \cdot 0,3) = 0,20$$

$$AF_2 = (0,2 \cdot 0,6 + 0,3 \cdot 0,7 + 0,4 \cdot 0,8) = 0,65$$

$$y_{h1} = 1 / [1 + \text{Exp}(-AF_1)] = 1 / [1 + \text{Exp} (-0,20)] = 0,5498$$

$$y_{h2} = 1 / [1 + \text{Exp}(-AF_2)] = 1 / [1 + \text{Exp} (-0,65)] = 0,6570$$

wh_{ij} = i gizli katman elemanından j numaralı çıktıya doğru olan ağırlık değeri

$$wh_{11}= 0,3 \quad wh_{21}= 0,4$$

$$AF'_1 = (y_{h1} \cdot wh_{11} + y_{h2} \cdot wh_{21}) = (0,5498 \cdot 0,3 + 0,6570 \cdot 0,4) = 0,4277$$

$$y_1 = 1 / [1 + \text{Exp}(-AF'_1)] = 1 / [1 + \text{Exp} (-0,4277)] = 0,6053$$

y_1 : Beklenen veya gerçek çıktı değeri

Başlangıç ağırlık değerleri genellikle rastlantısal olarak seçilen değerlerdir. Başlangıç değerlerle elde edilen çıktı değeri ve gerçek değer arasındaki fark kullanılarak geriye yayılımı hata kareleri toplamı, iterasyon sayısı arttıkça azaltılmaya çalışılmaktadır.

Hatanın süreç içerisinde azaltılabilmesi için hata kareleri toplamı elde edilerek, bu hata parametresinin geriye yayılımı ağırlık değerlerini revize etmede kullanılması gerekmektedir.

Öğrenme setinde bulunan N adet elemandan 1 tanesinin hata kareleri toplamı aşağıdaki gibi hesaplanabilir:⁶⁶

$$E(k) = \frac{1}{2} \sum_{j=1}^r [y_j(k) - Y_j(k)]^2 \quad r: \text{Çıktı sayısı} \quad (4.3)$$

E(k) değerlerinin hesaplandığı formülde ½ katsayısı sadece türev işlemleri sırasında sadeleştirme yapabilmek amacıyla konulmuştur. Hesaplama sonuçlarını etkilememektedir.

Öğrenme setindeki tüm elemanlara ait hata kareleri toplamı ise aşağıdaki gibidir:

$$E(T) = \sum_{k=1}^N E(k) \quad (4.4)$$

4.5.4 ÇKİBGYYSA'nda Öğrenme

Öğrenme kavramı, 4.3 numaralı formülle gösterilen hata kareleri toplamının minimuma indirilmeye çalışılması süreci olarak özetlenebilir.

Hata kareleri toplamını sürekli tekrarlanan işlemler sonucu en küçük değere doğru yaklaştıracak Delta kuralı 4.5 numaralı formülle verilmiştir:⁶⁷

$$w_{yeni} = w_{eski} - \eta \frac{\partial E(k)}{\partial w} \quad (4.5)$$

η : Öğrenme Oranı

⁶⁶ BOSE, N.K., LIANG, P.: a.g.e., s. 162.

⁶⁷ BOSE, N.K., LIANG, P.: a.g.e., s. 166.

Buradaki mantık, ağırlık değerlerindeki değişimin hata parametresinin ağırlığa göre kısmi türevine bağlı olduğuna dayanır. Bilgisayar uygulamalarında, ağırlık değerlerinin yenilenerek daha iyi(hata kareleri toplamı daha az) sonuçlara ulaşabilmesi için yukarıda belirtilen kısmi türevin hesaplanması gerekmektedir.

Her bir katmana ait çıktılar, 4.2 numaralı ifadeye benzer şekildedir:

$$y_j = \frac{1}{1 + e^{-s_j}} \quad (4.6)$$

Burada s_j değerleri 4.7 (Burada 4.1 numaralı ifade numaralı ifade temel alınmıştır) formülle hesaplanabilir:

$$s_j^l = \sum_{i=0}^n w_{ij}^l \cdot y^{l+1}_i(k) \quad (4.7)$$

l : Girdi katmandan başlayarak katman numarası

4.5 numaralı formülle elde edilmeye çalışılan yeni ağırlık değerlerinin hesaplanması için iki aşamalı bir yol izlenmelidir. Geriye yayılımın temel prensibi gereği hataların minimuma indirilmeye çalışılması, çıktı katmanından başlayarak geriye doğru yayılma ilkesi ile girdi katmanına kadar ilerlenmesiyle sağlanmaktadır. Burada hatanın geriye yayılımla azaltılmasında ilk aşama çıktı katmanı ile o katmana en yakın gizli katman arasında yayılımın yapılması gerekmektedir. Bu bağlantı için 4.5 numaralı formül değerleri aşağıdaki gibi hesaplanabilir:

$$\frac{\partial E(k)}{\partial w_{ij}} = \frac{\partial E(k)}{\partial s_j} \cdot \frac{\partial s_j}{\partial w_{ij}} \quad (4.8)$$

4.8 numaralı ifadenin daha açık yazılmış şekli 4.9 numaralı ifade ile gösterilmiştir. Burada eşitliğin sağ tarafındaki üç türev ifadesinin elde edilmesi yeterli olacaktır.

$$\frac{\partial E(k)}{\partial w} = \frac{\partial E(k)}{\partial y_j(k)} \cdot \frac{dy_j(k)}{ds_j} \cdot \frac{\partial s_j}{\partial w_{ij}} \quad (4.9)$$

4.9 numaralı formül, hata kareleri toplamının ağırlık değerleri setine göre kısmi türevini, zincir kuralı ile hesaplayabilmek üzere sunulmuştur.

$$\frac{\partial E(k)}{\partial y_j(k)} = \frac{\partial}{\partial y_j(k)} \left[\frac{1}{2} \sum_{j=1}^r [y_j(k) - Y_j(k)]^2 \right] = y_j(k) - Y_j(k) \quad (4.10)$$

$$\frac{dy_j}{ds_j} = \frac{d}{ds_j} \left(\frac{1}{1 + e^{-s_j}} \right) = y_j(k)(1 - y_j(k)) \quad (5.11)$$

$$\frac{\partial s_j}{\partial w_{ij}} = y_i(k) \quad (4.12)$$

4.10, 4.11 ve 4.12 numaralı ifadeler kullanılarak elde edilen sonuç 4.13 numaralı ifade ile verilmiştir.

$$\frac{\partial E(k)}{\partial w_{ij}} = [y_j(k) - Y_j(k)] \cdot [y_j(k) \cdot (1 - y_j(k))] \cdot y_i(k) \quad (4.13)$$

y_i : Çıktı katmanına en yakın gizli katmana ait değerler(çıktılar).

Bu aşamadan sonra, çıktı katmanından bir önceki katmandan başlayarak geriye doğru girdi katmanına kadar hatanın yayılımının sağlanması gerekmektedir. Burada temel mantık, yine hata kareleri toplamını azaltacak türevsel eşitliklerin yazılması ve her kademeye ait yeni ağırlık değerlerinin bu esasla hesaplanmasıdır.

$$\frac{\partial E(k)}{\partial w_{ai}} = \frac{\partial E(k)}{\partial y^{(1)}_j} \cdot \frac{\partial y^{(1)}_j}{\partial s^{(1)}_j} \cdot \frac{\partial s^{(1)}_j}{\partial w^{(1)}_{ai}} \quad (4.14)$$

4.14 nolu ifadede, terimler üzerinde yer alan 1 indeksi, çıktı kademesinin ardındaki ilk gizli katmanı ifade eder.

Çıktı katmanındaki hesaplama tarzına benzer şekilde geriye doğru diğer tüm katmanlara ait yeni ağırlık değerlerini hesaplamada kullanılacak formül aşağıda açıklanmıştır.

$$\frac{\partial E(k)}{\partial y^{(1)}_j} = \left[\sum_j \frac{\partial E(k)}{\partial s^{(1)}_j} \right] \cdot \frac{\partial s^{(1)}_j}{\partial y^{(1)}_i} \quad (4.15)$$

4.10 ve 4.11 kullanılarak 4.16 numaralı ifade elde edilmiştir.

$$\sum_j \frac{\partial E(k)}{\partial s^{(1)}_j} = \sum_j (y_i - Y_j) \cdot y_j \cdot (1 - y_j) \quad (4.16)$$

$$\frac{\partial s^{(1)}_j}{\partial y^{(1)}_i} = w_{ij} \quad (4.17)$$

4.16 ve 4.17 numaralı ifadeler kullanılarak 4.15 numaralı ifadenin daha açık yazılmış şekli 4.18 ile gösterilmiştir.

$$\frac{\partial E(k)}{\partial y_j^{(1)}} = \left[\sum_j (y_i - Y_j) \cdot y_j \cdot (1 - y_j) \right] \cdot w_{ij} \quad (4.18)$$

$$\frac{\partial y_j^{(1)}}{\partial s_j^{(1)}} = y_j^{(1)} \cdot (1 - y_j^{(1)}) \quad (4.19)$$

$$\frac{\partial s_j^{(1)}}{\partial w_{ai}^{(1)}} = y_a^{(2)} \quad (4.20)$$

4.14 numaralı ifade, 4.18, 4.19 ve 4.20 kullanılarak 4.21'de açık halde yazılmıştır.

$$\frac{\partial E(k)}{\partial w_{ai}} = \left[\sum_j (y_i - Y_j) \cdot y_j \cdot (1 - y_j) \right] \cdot w_{ij} \cdot [y_j^{(1)} \cdot (1 - y_j^{(1)})] \cdot [y_a^{(2)}] \quad (4.21)$$

4.13 ve 4.21 numaralı ifadeler, 4.5 numaralı ifadede kullanılarak yeni ağırlık değerleri hesaplanabilir⁶⁸. Bu işlemler, belirlenen maksimum iterasyon sayısı veya minimum hata kareleri toplamı elde edilinceye kadar devam ettirilerek, öğrenme süreci gerçekleştirilmiş olur.

Eğitim aşamasında elde edilen en iyi ağırlık değerleri, test ve modelin kullanımı sırasında da kullanılmaktadır.

⁶⁸ BOSE, N.K., LIANG, P.: a.g.e., s. 164-165.

4.5.5 ÇKİBGYYSA Verilerin Kullanımı

Sözü geçen YSA modellerinde kullanılacak olan veri setleri, ya $[0,1]$ aralığında olmalıdır, ya da bu aralıkta ifade edilecek bir dönüşüm uygulanmalıdır. Bu durum 4.22 numaralı ifade ile gösterilmiştir.⁶⁹

$$[x] \in (0,1)^n \Rightarrow [y] \in (0,1)^n \quad (4.22)$$

Girdi katmanda yer alan bağımsız değişkenlerle, çıktı katmanda yer alan bağımlı değişkenlerin $(0,1)$ aralığına dönüştürülmesinin belki de en kolay ve en yaygın yöntemi, her bir veri setinin en küçük değerini sıfır, en büyük değerini de bir olarak ifade eden doğrusal dönüşümdür:

$$x_i = (xb_i - mi_i) / (ma_i - mi_i) \quad (4.23)$$

$$Y_j = (Yb_j - mi_j) / (ma_j - mi_j) \quad (4.24)$$

Algoritmanın çalışması sonucunda elde edilen tahmini çıktı değerleri de sözü geçen aralıktadır, yani kullanıcıların istedikleri şekilde ifade edilebilmeleri için bu sefer de basit bir ters dönüşüm uygulanmalıdır:

$$yb_j = y_j \cdot (ma_j - mi_j) - mi_j \quad (4.25)$$

4.5.6 ÇKİBGYYSA Parametrelerinin Sonuca Etkileri

YSA'nın kullanılması sırasında kullanılacak çeşitli parametreler mevcuttur ve bu parametreler hem başarılı tahmin derecesini etkilerken hem de işlem süresini etkilemektedir.

⁶⁹ HASSOUN, M.H.: "Fundamentals of Artificial Neural Networks", The MIT Press , Massachusetts, 1995, s. 3.

Uygulayıcı tarafından seçilen parametrelerin başlıcaları aşağıda sunulmuştur:

Öğrenme Oranı,
Momentum Katsayısı,
İterasyon Sayısı,
Girdi Katman Eleman Sayısı,
Gizli Katman Sayısı,
Gizli Katman Eleman Sayısı,
Çıktı Katman Eleman Sayısı,
Kabul Edilebilir En Büyük Hata Değeri,
Eğitim Veri Seti Büyüklüğü,
Test Veri Seti Büyüklüğü.

4.5.6.1 Öğrenme Oranı

YSA parametrelerinden öğrenme oranı algoritmanın performansını etkileyen en önemli değişkenlerden birisidir. Çok küçük seçilen öğrenme oranı, her iterasyonda kaydedilen gelişmenin çok küçük olmasına ve bunun sonucunda da kabul edilebilir hata seviyesinin elde edilmesinin uzamasına yol açar. Buna karşılık çok büyük seçilen öğrenme oranı ise öğrenme hızını artırırken dalgalanmalara yol açmaktadır.

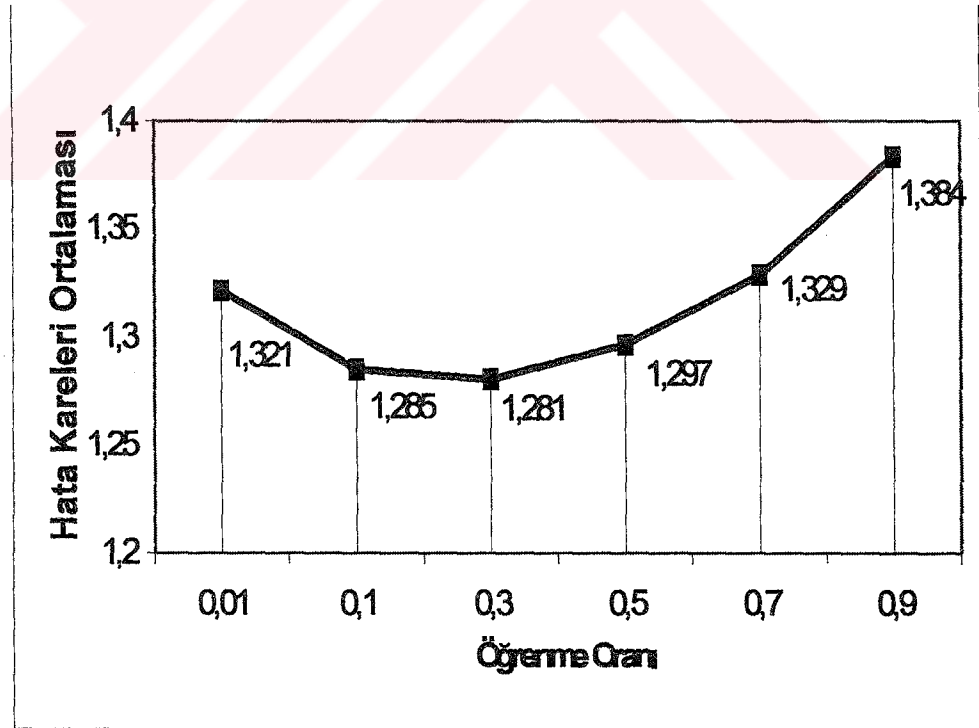
Öğrenme oranındaki değişimin, YSA performansına etkilerini araştırmak üzere örnek bir veri seti oluşturulmuş ve diğer tüm şartlar aynı tutularak değişik öğrenme oranları ile elde edilen sonuçlar karşılaştırılmıştır.

Kurulan modelin temel özellikleri Tablo 4.2 ile sunulmuştur:

Tablo 4.2 YSA Performans Testi Modelinin Özellikleri

Momentum Katsayısı : 0
İterasyon Sayısı : 50.000
Girdi Katman Eleman Sayısı : 5
Gizli Katman Sayısı : 1
Gizli Katman Eleman Sayısı : 4
Çıktı Katman Eleman Sayısı : 1
Tüm Başlangıç Ağırlık Değerleri : 0,5
Öğrenme Oranları: 0,01 – 0,1 – 0,3 – 0,5 – 0,7 – 0,9 – 1,5

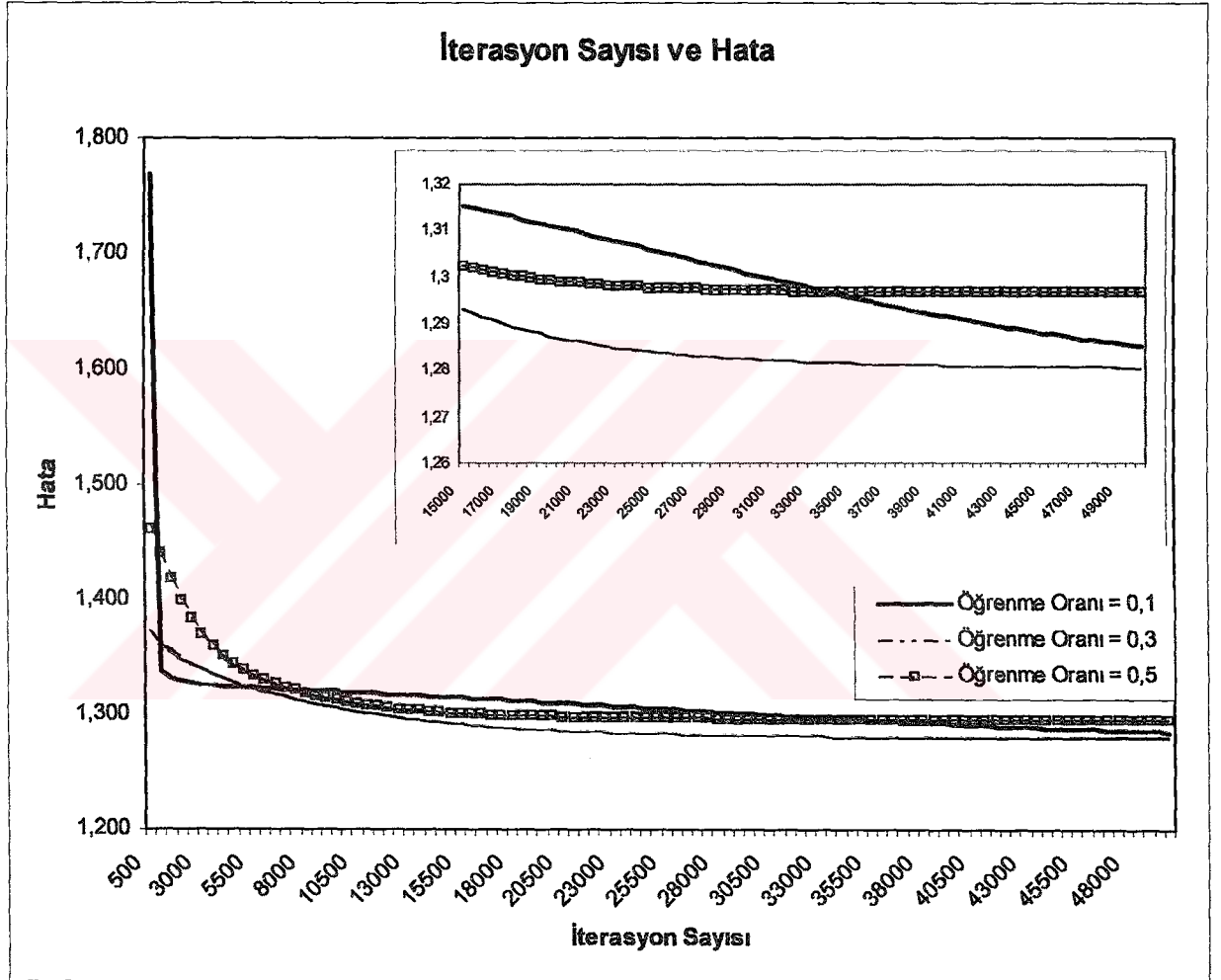
Şekil 4.9, çeşitli öğrenme oranları ile yapılan testin sonucunda elde edilen hata kareleri ortalamaları değerlerini grafikte göstermek üzere sunulmuştur.



Şekil 4.9 Değişik Öğrenme Oranlarının Performansı Grafiği

Şekil 4.9'dan da görüleceği üzere ne çok büyük ne de çok küçük öğrenme oranları belirlenen şartlarda iyi sonuçlar verememişlerdir.

Şekil 4.10'da en iyi sonuç veren üç öğrenme oranı değeri olan 0,1 , 0,3 ve 0,5 değerlerinin, iterasyon sayısına bağlı olarak performansları incelenmiştir.



Şekil 4.10 Değişik Öğrenme Oranlarının Performansının Karşılaştırılması

Şekil 4.10'dan da gözlenebileceği üzere; görece olarak daha büyük olan öğrenme oranları, başlarda hataları hızla düşürmelerine rağmen sonlara doğru, eğimleri azalmıştır.

Konuya ilişkin kaynakların bir çoğunda, öğrenme oranının tüm süreç boyunca sabit tutulması yerine başlarda nispeten büyük değerlerle başlamanın ve süreç ilerledikçe öğrenme oranının algoritma tarafından düşürülmesinin yararlı olacağını savunulmaktadır.

4.5.6.2 Momentum Katsayısı

Bu dalgalanmaların ortadan kaldırılabilmesi için öğrenme mekanizmasına momentum terimi eklenmesi önerilen bir yöntemdir. Momentum teriminin öğrenme mekanizmasına eklenmesi 4.26 numaralı ifade ile gösterilmiştir.⁷⁰

$$w_{(k+1)} = w_{(k)} - \left[\eta \frac{\partial E(k)}{\partial w} + \beta \Delta w(k-1) \right] \quad (4.26)$$

β : Momentum katsayısı

Momentum parametresinin eklenmesi, daha önceki eğitim verisinden gelen ağırlık değeri değişiminin de güncel ağırlık değişimine katılmasını sağlar böylelikle normal dışı bir verinin ağırlık değerini çok büyük ölçüde etkilemesini önler ve böylelikle de dalgalanma riskini azaltır.⁷¹

Momentum katsayısı için uygulamada genelde kullanılan değer 0,9'dur.⁷² Momentum katsayısının kullanılması öğrenme hızının yavaşlamasına yol açabilir ancak aynı zamanda da stabil bir proses gerçekleşmesi şansını da

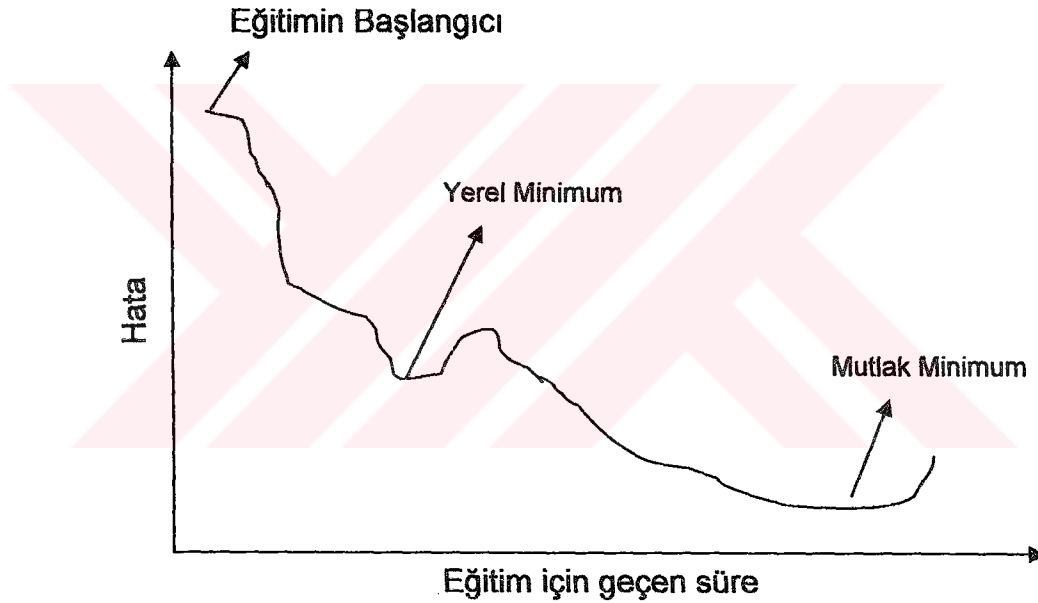
⁷⁰ BOSE, N.K., LIANG, P.: a.g.e., s. 167.

⁷¹ BIGUS J. P.: "Data Mining With Neural Networks", New York, NY: McGraw- Hill., 1996, s.89.

⁷² LIN T.C., LEE C.S.G.: a.g.e., s.247.

arttırmaktadır. Özellikle görece olarak büyük seçilen öğrenme oranları kullanıldığında bu fayda daha da belirgin hale gelmektedir.

Öğrenme sürecinde yaşanabilecek problemlerden birisi de hata kareleri toplamının en aza indirilmeye çalışılması sırasında bir yerel minimum noktanın algoritmayı yanıltarak o noktaya yaklaştırmaya çalışılmasını sağlamasıdır. Yerel minimumun hedef saptırdığı durumlarda, iterasyon sayısının artmasına rağmen hata değerinde istenen düşüş sağlanamamaktadır ve eğitim süresi uzamaktadır. Bu durum Şekil 4.11'deki grafikte açıklanmıştır.

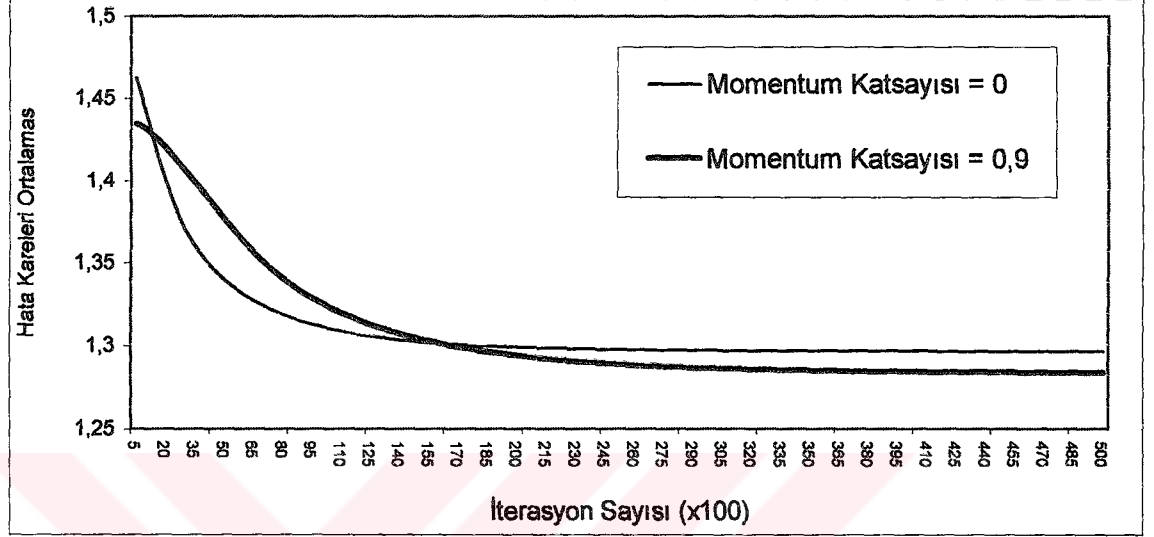


Şekil 4.11 Eğitim Süresi ve Hata Grafiği ve Yerel Minimum⁷³

Momentum katsayısının öğrenmeye etkisini incelemek üzere örnek veri kümesi ile deneyler yapılmıştır. Momentum katsayısı hariç tüm koşullar aynı tutularak sabit tutularak yapılan deneylerde momentum katsayısının gerekli

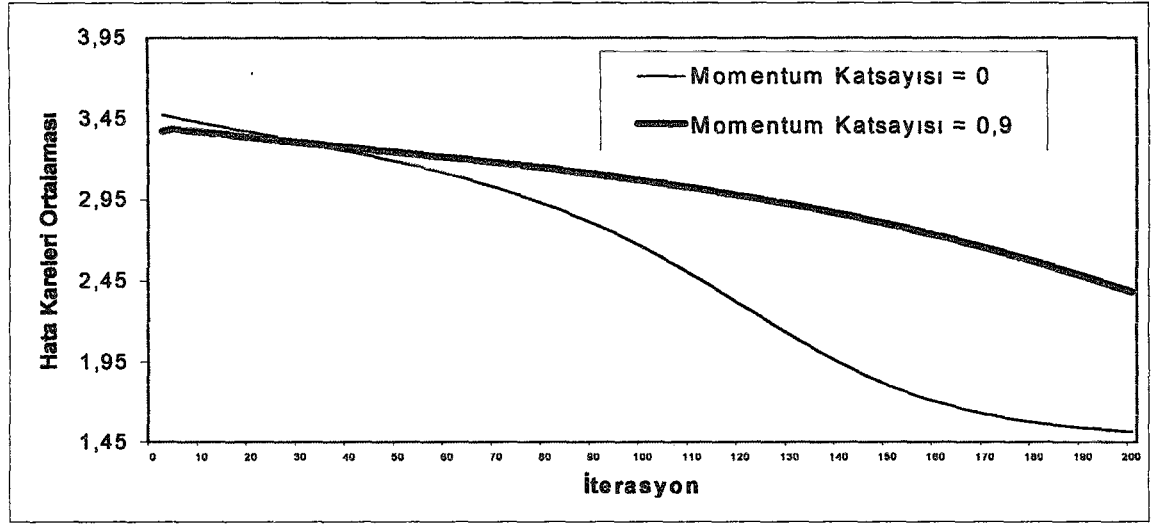
⁷³ ROGERS, J.: "Object Oriented Neural Networks in C++", San Diego, CA: Academic Press Inc., 1997, s.85.

ve kullanışlı olduğu ispat edilmeye çalışılmıştır. Momentum parametresinin modele dahil edilmesinin, genellikle daha iyi sonuçların elde edilmesine yol açtığı gözlemlenmiştir. Şekil 4.12'de, örnek veri seti ile 50,000 iterasyon ve 0,5'lik öğrenme oranı ile elde edilen sonuçlar sunulmuştur.



Şekil 4.12 Momentum Katsayısının Öğrenmeye Etkisi

Momentum parametresi başlarda daha büyük hatalara yol açmaktadır ve öğrenme hızı başlarda düşüktür. İterasyon sayısının artması ile birlikte momentum katsayısının kullanılmadığı durumda öğrenme eğrisi yataylaşırken momentum katsayısının 0,9 aldığı durumda öğrenme devam etmektedir.



Şekil 4.13 Momentum Katsayısının Öğrenmenin İlk Aşamasına Etkisi

Şekil 4.13, momentum katsayısının öğrenmenin ilk aşamasındaki etkisini göstermek üzere sunulmuştur. Şekilden de görülebileceği üzere, ilk 200 iterasyon için momentum kullanılmayan model daha iyi performans göstermiştir. Ancak sonraki aşamalarda bu fark giderek kapanmış ve öğrenme sürecinin sonunda momentum kullanılan model ile daha iyi sonuçlar elde edilmiştir.

4.5.6.3 İterasyon Sayısı ve Aşırı Eğitim

Aşırı eğitim sorunu, iterasyon sayısı arttıkça yaşanabilecek bir problemdir. Algoritma verilen bilgilerden yola çıkarak eğitilirken bir süre sonra, genel ilişkileri öğrenmekten daha çok veri setindeki parazitleri ve verinin kendisini öğrenmeye başlayabilir. Bu gibi durumlarda hata değeri düşerken ve o veri seti için daha iyi sonuçlar elde edilirken, modelin genele uygunluğu azalmaktadır.

Bu durumun kontrol altında tutulabilmesi ve aşırı eğitimden kaçınılabilmesi için algoritmanın çalışması sırasında belirli aralıklarla oluşturulan test verileri ile de denenmesi ve eğitim veri setinin hata mertebesi ile test veri setinin hata mertebesinin kıyaslanması gerekmektedir. İki veri setinin hata düzeyleri arasında önemli farklar oluşmaya başladığı gözlemlendiği anda eğitim sonlandırılmalıdır.⁷⁴

4.5.6.4 Girdi Katman Eleman Sayısı

Girdi katman eleman sayısı yada bağımsız değişken sayısı tamamen uygulamaya dayalı bir parametredir. Bir olayı açıklayabilmek için uygulamacının belirleyeceği sayıda tanımlayıcı değişken modelde yer alabilir. Burada yaşanacak problem, girdi katman eleman sayısına bağlı olarak işlem süresinin uzamasıdır.

4.5.6.5 Gizli Katman Sayısı

Gizli katman sayısı, kaç adet gizli katmanın modelde yer alacağını belirler. Bu çalışma kapsamında bir gizli katmanlı model geliştirilmiştir. Gizli katmanlar arttıkça ağırlık değerleri ve yapılacak hesaplamalar da geometrik olarak artacağı için her uygulama için optimum sayının belirlenmesi gerekmektedir. Aksi halde işlem süresi çok uzayabilmektedir.

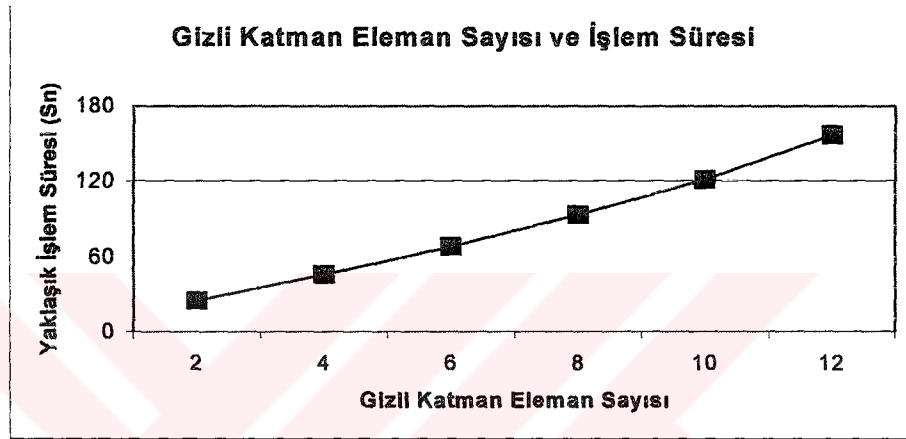
4.5.6.6 Gizli Katman Eleman Sayısı

Gizli katman eleman sayısı, her bir gizli katmanın kaç elemandan oluşacağını belirler. Gizli katmanların eleman sayısı arttıkça ağırlık değerleri ve yapılacak

⁷⁴ BOSE, N.K., LIANG, P.: a.g.e., s. 242-243.

hesaplamalar yine geometrik olarak artar ve bu durum işlem süresi çok uzayabilir.

İşlem Süresinin, gizli katman eleman sayısı ile ilişkisini göstermek üzere bir dizi deney yapılmıştır. Bu deneyde tüm çevre şartları aynı tutularak, aynı mikro işlemci ve aynı algoritma kullanılarak yapılan bu deney sonuçları Şekil 4.14'de sunulmuştur.



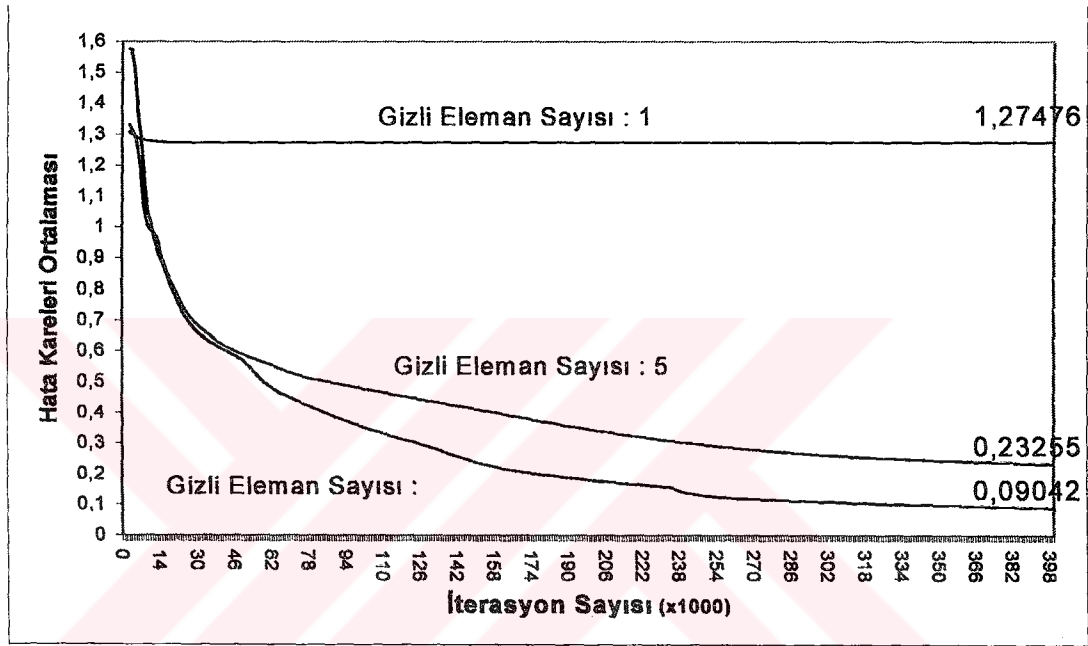
Şekil 4.14 Gizli Katman Eleman Sayısı ve Toplam İşlem Süresi

Şekil 4.14'de işlem süresi ile gizli eleman sayısı arasındaki ilişkiyi gösteren deney, öğrenme oranının 0,1 ve momentum katsayısının sıfır olduğu durumda 1.000 iterasyonun gerçekleştirilmesi için geçen süreyi göstermektedir. Mikro işlemci ve algoritmada yapılacak geliştirmeler tüm grafik değerlerine aynı oranda etki edeceğinden süreler mutlak değerleri ile değerlendirilmemelidir. Veriler sadece karşılaştırmalı olarak değerlendirildiğinde anlamlı yorumlar yapılabilir.

Daha hızlı bir işlemci ve daha geliştirilmiş bir bilgisayar programı oluşturularak tüm deney süreleri aşağıya çekilebilir. Bu durum da süre açısından gizli katman eleman sayısının daha büyük seçilebilmesine olanak sağlayabilir. Ancak her koşulda uygulayıcılar, toplam işlem süresi ve

performans kriterlerini optimum edecek bir gizli katman eleman değeri belirlemelidirler.

Gizli katman eleman sayısının artırılması, işlem süresini uzatırken aynı zamanda da süreç performansını da arttırdığını ispatlamak amacıyla gerçekleştirilen deneyin sonuçları Şekil 4.15'te sunulmuştur.



Şekil 4.15 Gizli Katman Eleman Sayısının YSA Performansına Etkisi

Şekil 4.14 ve Şekil 4.15'ten açıkça görüleceği üzere gizli katman eleman sayısının belirlenmesinde, işlem süresi ve performans arasında bir optimizasyon yapılması gerekmektedir. Ayrıca çalışmamız sırasında bizzat yaptığımız çok sayıda deneme sonrasında, gizli katman sayısının aşırı artırılmasının hata parametresi üzerinde anlamlı bir etki yapmadığı halde işlem süresini uzattığı gözlenmiştir.

4.5.6.7 Çıktı Katman Eleman Sayısı

Çıktı katman eleman sayısı yada bağımlı değişken sayısı benzer şekilde tamamen uygulamaya dayalı bir parametredir. Bir olayın sonucu olan değişkenler uygulamacının belirleyeceği sayıda modelde yer alabilir.

4.5.6.8 Kabul Edilebilir En Büyük Hata Değeri

Bu değişken iterasyon sayısı ile birlikte incelenmelidir. Teorik olarak iterasyon sayısı arttıkça hata düzeyi azalır. Uygulamaya son verecek kriterlerden birisi de kabul edilebilir hata düzeyine ulaşıldığına karar verilmesidir. Yine uygulamacı tarafından verilecek bu karar, uygun olandan daha büyük veya küçük seçilirse başarı oranı düşer. Uygun değerden daha küçük seçilen kabul edilebilir hata düzeyi iterasyon sayısının çok artmasına ve aşırı eğitim riskinin oluşmasına yol açabilir. Gereğinden büyük seçilen kabul edilebilir limit ise fırsat kaybına ve daha iyi sonuçların elde edilebileceği durumda bu şansın kaçırılmasına ve tahminin kalitesinin düşmesine yol açar.

4.5.6.9 Eğitim Veri Seti Büyüklüğü

Veri seti büyüklüğü seçilirken yine bir optimizasyon mantığı kullanılmalıdır. Veri seti genel kümeyi yansıtacak kadar büyük olmalıdır ama işlem süresini çok uzatmayacak kadar da sınırlı tutulmalıdır. Küçük veri setleri genel yapının algılanması için yeterli bilgi barındırmadığı için, model bir süre sonra veri setinin öznel yapısını öğrenmeye başlar ki bu durum aşırı öğrenmenin sebeplerinden birisidir.

4.5.6.10 Test Veri Seti Büyüklüğü

Aşırı öğrenmeyi engelleyecek ve modelin geçerliliği ile ilgili bilgiyi verecek olan yapı test veri seti ile sağlanır. Test verileri öğrenmenin ne kadar genelleştirilebilir olduğu ile ilgili önemli bilgiler verir.

Test verisinin, eğitim verisinin yaklaşık %10'u düzeylerinde olabileceği bir yaklaşım olarak düşünülebilir.⁷⁵

4.5.6.11 Başlangıç Ağırlık Değerlerinin Öğrenmeye Etkisi

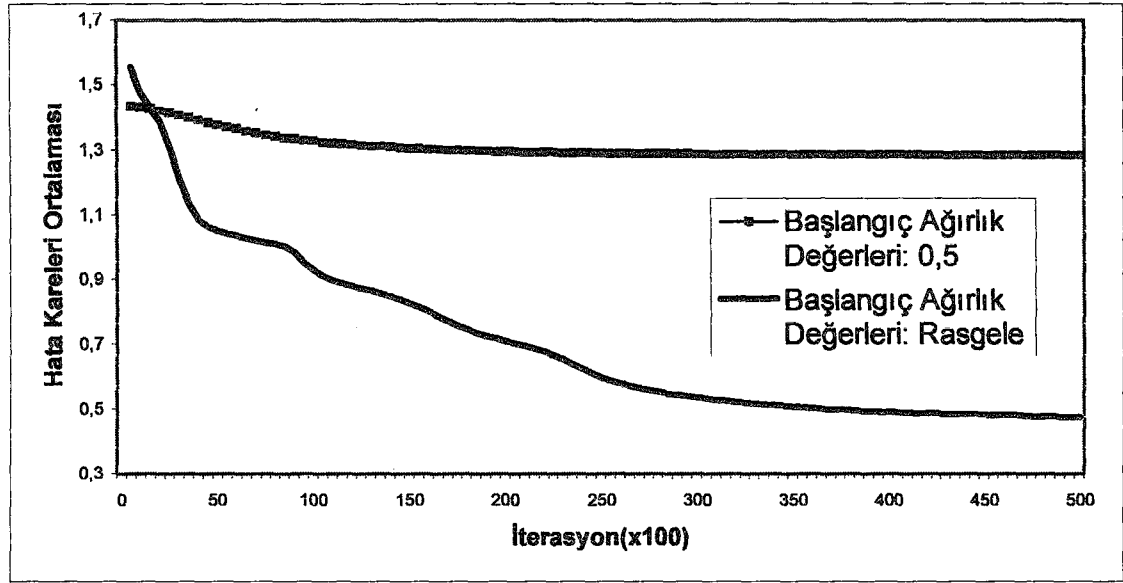
Başlangıç ağırlık değerlerinin belirlenmesi, öğrenme sürecinin başarı oranını etkileyen faktörlerden birisidir.

Ağırlık değerleri başlangıçta küçük rasgele sayılar üretilerek seçilmelidir. Eğer sonuçta ağırlıkların farklı değerlerde olmaları gerekiyorsa, başlangıç değerlerinin eşit seçilmesi iyi sonuçlar vermez.⁷⁶

Başlangıç ağırlık değerlerinin seçiminin öğrenmeye etkisini araştırmak amacıyla yaptığımız deneyde, diğer tüm şartlar aynı kalmak koşuluyla, birinci durumda tüm başlangıç ağırlık değerleri 0,5 seçilmiştir. İkinci durumda ise $[0, 1]$ tesadüfi sayılar üretilerek başlangıç değerleri belirlenmiştir. Sonuçlar Şekil 4.16'da sunulmuştur.

⁷⁵ BOSE, N.K., LIANG, P.: a.g.e., s. 242.

⁷⁶ LIN T.C., LEE C.S.G.: a.g.e., s.245.



Şekil 4.16 Başlangıç Ağırlık Değerlerinin Seçimi ve Öğrenme

Şekil 4.16'dan da açıkça görüleceği üzere rasgele seçilen başlangıç değerleri, tüm ağırlık değerleri için atanan sabit bir değere oranla çok daha iyi sonuçlar vermiştir.

4.6 Genetik Algoritmaların Yapay Sinir Ağı Modellerinde Kullanımı

Genetik Algoritmalar, yapay sinir ağlarında öğrenmenin gerçekleştirilmesinde kullanılan yöntemlerden birisidir. GA iki temel şekilde kullanılabilir:

- i. Optimum parametrelerin tahmini,
- ii. Optimum ağırlık değerlerinin elde edilmeye çalışılması.

4.6.1 GA İle Optimum parametrelerin tahmini

Denetimli öğrenme modellerinde karşılaşılan en önemli sorunlardan birisi parametrelerin tahminidir. Yani öğrenme oranı, momentum oranı, gizli katman sayısı, gizli katman nöronlarının sayısı gibi bir çok parametre, sonuçların hassasiyetini etkilemektedir. GA bu parametrelerin daha duyarlı sonuçlara ulaşmayı sağlayacak şekilde belirlenmesi için kullanılabilir. Ayrıntıları bölüm 3.5.2'de verildiği şekilde ikili sistemde ifade edilen parametrelerin nesiller arası geçişle daha uygun hale getirilmesi sağlanır.

4.6.2 GA İle Optimum ağırlık değerlerinin elde edilmeye çalışılması

GA'ın YSA uygulamalarında kullanım alanlarından birisi de optimum ağırlık değerlerinin elde edilmeye çalışılması yani öğrenme sürecinde kullanılmalarıdır. Bu çalışmada, GA'ın YSA'ında kullanımı ile ilgili olarak üzerinde durulacak uygulama öğrenme sürecindeki kullanımdır.

Bu uygulamada öğrenme, en popüler yöntemlerden olan delta kuralındaki öğrenmeye alternatif olarak GA kullanılmasıyla gerçekleşir. Rastlantısal olarak seçilen ağırlık değerleri ikili sisteme çevrilerek ifade edilir ve öğrenme nesiller boyunca kalıtsal olarak devam eden genetik geçiş sonrası elde edilir.

Burada problem bir minimizasyon problemi olarak ifade edilir ve hata kareleri toplamını minimum eden nesil yakalanmaya çalışılır. Bölüm 3.5.2'de ayrıntılı olarak açıklandığı gibi güçlü bireyler bir sonraki nesilde daha büyük oranla ve/veya şansla yer alırlar. Burada güçlü bireyden kastedilen hata kareleri toplamı görece küçük olan bireylerdir. Yine daha önce anlatıldığı gibi bu bireyler arası eşleşme sonucu belirli bir nesil sayısına ulaşıldığında elde

edilen kümedeki en iyi bireyin verdiği ağırlık değerleri de en iyi öğrenme sonucunu vermektedir.

4.7 Yapay Sinir Ağlarının Kullanım Alanları

Çok sayıda kullanım alanı olan YSA, her türlü girdi ve çıktı bulunduran problemin çözümünde kullanılabilir.

YSA'nın bir çok alanda rahatça kullanılabilmeleri, sözü geçen algoritmaların güçlü yanlarının çokluğundan kaynaklanmaktadır. Uygulamalar genel olarak iki kısma ayrılabilir:

Mühendislik/ teknolojik uygulamalar,
Sosyal bilimler/işletmecilik uygulamaları

4.7.1 Mühendislik/Teknolojik Uygulamalar

YSA çok çeşitli mühendislik ve teknolojik uygulamada kullanılmaktadırlar. Bunlardan bazıları:

Yapay görme sistemleri
Kişi Tanıma Sistemleri
Bilgisayar dosyalarının Sıkıştırılması
Bilgisayarlarda Virüs Taraması
Bilgisayar Ağ Güvenlik Sistemleri
Otomatik Kontrol ve Robot Kontrol
Kalite ve Proses Kontrolü
Hava kontrol sistemleri
Karayolu trafik yoğunluğu optimizasyonu

El Yazısı algılama sistemleri
Konuşmayı metne dönüştüren sistemler

4.7.2. Sosyal Bilimler/İşletmecilik Uygulamaları

Önceleri genellikle endüstriyel ve teknolojik uygulamalarda kullanılan YSA, esnek yapıları nedeniyle sosyal bilimlerde ve özellikle de işletmecilik alanında yaygın olarak kullanılmaya başlanmıştır.

Bo Wong, Thomas A. Bodnovich ve Yakup Selvi tarafından yapılan ve I.Ü. İşletme Fakülte Dergisinde Nisan 1997'de yayınlanan "Sinir Ağlarının İşletmecilik Alanındaki Uygulama Araştırmalarının Bibliyografyası 1988-Eylül 1994" isimli çalışmada, Dünya literatüründe bilişim sistemleri, finans, insan kaynakları, muhasebe / denetim, pazarlama, üretim yönetimi gibi çeşitli alanlarda çok sayıda makale yazıldığı belirlenmiştir.⁷⁷

İşletmecilik alanındaki uygulamaları giderek yaygınlaşan YSA'nın belli başlı kullanım alanları aşağıdaki gibidir:^{78 79 80}

4.7.2.1 Finans Uygulamaları

- Bankalar menkul kıymetler kurumları ve diğer şirketlere yönelik iflas tamini,

⁷⁷ WONG, B. K., BODNOVICH T. A., SELVI, Y.: "Neural Network Applications in Business: A review and Analysis of the Literature (1988-95)", **Decision Support Systems**, 19, 1997

⁷⁸ WONG, B. K., BODNOVICH T. A., SELVI, Y.: **a.g.e.**, s.119,120.

⁷⁹ GOONATILAKE S., TRELEAVEN, P.: "Intelligent Systems for Finance and Business", New York, NY: John Wiley & Sons, Inc., 1995, s. 61.

⁸⁰ BIGUS J. P.: **a.g.e.**, s. 29.

- Tahvil deęerlendirmesi,
- Ticari ve bireysel kredi taleplerinin deęerlendirilmesi,
- Gayrimenkul deęerlendirmesi,
- Hisse senedi ve dięer yatırım aralarının deęerlendirilmesi ve gelecek performanslarının tahmini,
- Sahtekarlık olasılıęının nceden tahmin edilmesi,
- Ekonomik gstergelerin tahmini,
- Sigortacılık prim, risk ve sahtekarlık tahminidir.

4.7.2.2 İnsan Kaynakları Uygulamaları

- İşyeri davranışı tahmini,
- Personel seçimidir.

4.7.2.3 Muhasebe Uygulamaları

- Denetim kararının verilmesi,
- LIFO/FIFO sınıflandırması,
- Vergi beyannamesi prosesinin yönetimidir.

4.7.2.4 Pazarlama ve Müşteri İlişkileri Yönetimi Uygulamaları

- Satış tahmini,
- Fiyat duyarlılığının tespiti,
- Müşteri deęerlendirme,
- Optimum fiyat belirleme,
- Müşteri eğilimlerinin tespiti,
- Pazar bölümlerinin tanımlanmasıdır.

4.7.2.5 Üretim Yönetimi Uygulamaları

- Araştırma ve geliştirme faaliyetleri,
- Maliyet mühendisliği,
- Kaynak planlaması,
- Proses kontrol,
- Kontrol şeması tanımlama ve yorumlama,
- Kaliteye etki eden parametrelerin düzenlenmesi ve makine/teçhizat ayarlarının yapılması,
- Araç dağıtım otomasyonu ve optimizasyonu,
- Esnek imalat sistemlerinin planlanması,
- Kayıp zaman minimizasyonudur.

Yapay sinir ağları ve genel olarak da yapay zeka, özellikle objektif karar vermenin önemli olduğu uygulamalar ve karmaşık veri kümeleri ile karar sisteminin karşılaştığı durumlarda geleneksel yöntemler yanında tercih edilen alternatiflerdendir.

4.8 Yapay Sinir Ağı Algoritmalarının Güçlü ve Zayıf Yönleri

Önemli avantajlara sahip YSA, bu özellikleriyle çok geniş bir kullanım alanı elde etmişlerdir. Her yöntem gibi YSA'nın da güçlü ve zayıf yönleri bulunmaktadır.

4.8.1 YSA'ın Güçlü Yönleri⁸¹

- YSA çok çeşitli gerçek hayat probleminin çözümünde kullanılabilirler,

⁸¹ BERRY M. A., LINO G.L.: a.g.e., s. 331.

- ii. Karmaşık durumlarda bile iyi sonuçlar elde edilebilir,
- iii. YSA kategorik ve sürekli değer içeren girdi ve çıktılarla çalışılabilir,
- iv. Geliştirilmiş bir çok paket program mevcuttur.

4.8.2 YSA'ın Zayıf Yönleri⁸²

- i. Girdilerin $[0,1]$ aralığında tanımlanması veya bu skalaya dönüştürülmesi gerekmektedir,
- ii. Sonuçların duyarlılık analizine benzer şekilde yorumlanması olanaksızdır,
- iii. Algoritmalar erken bir şekilde yerel bir çözüme doğru yönelebilirler.

⁸² BERRY M. A., LINFO G.L.: a.g.e., s. 333.

BÖLÜM 5. UYGULAMA

(ÇKİBGYYSA'NIN GA İLE MELEZ KULLANIMI VE MÜŞTERİ GRUPLANDIRMASI VE SINIFLANDIRMASI)

Bu çalışma kapsamında, yaygın olarak kullanılan ileri beslemeli, geri yayımlı ve çok katmanlı sinir ağlarının eğitim sürecinde, GA melez bir yapı ile kullanılmıştır. İlk aşamada GA ile eğitilen sinir ağı elde edilen sonucu daha iyi bir noktaya taşımak amacı ile delta kuralıyla eğitime devam edilmiştir. GA'nın hızlı ve çok iyi sayılabilecek sonuçlara ulaşabilme kabiliyetleri ile delta kuralının ve geri yayılımın matematiksel gücü aynı algoritma üzerinde birleştirilmeye çalışılmıştır. Melez uygulama ile yapılan tüm çalışmalar, algoritmaların ayrı ayrı kullanılmasına göre daha iyi sonuçlar elde edildiğini göstermiştir.

Bölüm 1'de de açıklanan VM işlevlerinden kümeleme ve sınıflandırmanın gerçek bir işletme problemini çözmek üzere bu bölümde kullanılmıştır.

5.1 Problemin Tanımlanması

Geliştirilen modelin uygulandığı işletme problemi, Türkiye'de faaliyet gösteren bir internet temelli hizmet sağlayıcısı kuruma aittir. Sözü geçen kurum internet ve kablosuz iletişim merkezli çok sayıda hizmet sunan sektöründe lider firmalardan birisi olan kurumlardandır.

Yapılan çalışma temel olarak, varolan müşterilerin özelliklerinden yola çıkarak, muhtemel müşterilerin firmaya katkılarını tahmin etmek yolu ile bir eylem şekli belirlemektir.

Müşterilerine objektif kriterler içerisinde makul bir şekilde yaklaşmayı hedefleyen firma yöneticileri, bu amaca ulaşmak için potansiyel müşterilerinin, firmalarına olası katkılarını tahmin ederek, bu doğrultuda bir promosyon ve fiyatlandırma stratejisi oluşturmayı hedeflemektedirler.

Buradaki iki temel amaç;

1. Promosyona ayrılan kaynakları eksik kullanarak müşteri memnuniyetinin istenen düzeyin altında gerçekleşmesini engellemek,
2. firmaya katkısı görece olarak az olacak bir müşteriye gereksiz yere promosyon yaparak gereksiz promosyon maliyetine yol açılmasını engellemektir.

5.2 Veri Seti ve Özellikleri

Uygulama için kullanılan veri seti, firma yöneticileri tarafından sağlanmıştır. Müşterilerin kişilik haklarına saygı prensibi gereği, veri seti tüm isim, adres ve telefon gibi bilgiler süzülerek firmadan alınmıştır.

Firmadan elde edilen veri setinin özellikleri Tablo 5.1'de sunulmuştur:

Tablo 5.1 Uygulamada Kullanılan Veri Setinin Özellikleri

Müşteriye Ait Özellik	Veri Türü	Veri Özelliği (Seçenekler)
Dağıtıcı	Metin	Belirli Sayıda, Firma Tarafından Önceden Belirlenmiş.
Ödeme Planı	Metin	Belirli Sayıda, Firma Tarafından Önceden Belirlenmiş.
Medeni Durumu	Metin	Seçenek: Evli, Bekar, Dul, Boşanmış
Promosyon (e-mail almak isteyen)	Sayı	Seçenek: 0-1
E-bilgilendirme (maili almak ister)	Sayı	Seçenek: 0-1
SMS almak istiyor mu?	Sayı	Seçenek: 0-1
Şehir	Metin	Belirli Sayıda, Firma Tarafından Önceden Belirlenmiş.
Meslek	Metin	Belirli Sayıda, Firma Tarafından Önceden Belirlenmiş.
Cinsiyet	Metin	Seçenek: Bay, Bayan
Müşteri Adedi	Sayı	Veri Aralığı: 1-n
Kullanıcı Sayısı	Sayı	Veri Aralığı: 1-n
Yaş	Sayı	Veri Aralığı: 0-100
Sms Adet	Sayı	Veri Aralığı: 0-n

5.3 Veri Hazırlanması

Uygulamaya konu olan firmadan ham halde alınan veriler, modelde kullanılabilecek şekilde hazırlanmıştır. Bu aşamada veriler dönüştürme, temizleme ve süzme işlemlerine tabi tutulmuştur.

5.3.1 Veri Dönüştürme

Burada yapılan işlem, verilerin türünün, veri özelliğini bozmadan dönüştürülmesidir. Metin şeklinde olan tüm veriler sayısal değer atanması yoluyla model tarafından kullanılabilir hale çevrilmiştir. Örneğin orijinal veri setinde “bay” özelliğine karşılık, sıfır ve “bayan” özelliğine karşılık olarak da bir değeri atanmıştır. Bu işlem tüm metinsel özellikler için ayrı ayrı uygulanmıştır.

5.3.2 Verilerin Temizlenmesi

Burada temizleme kelimesi ile anlatılmaya çalışılan durum: Gerçekleşmesi mümkün olmayan yada mümkün olduğu halde muhtemel olmayan kayıtların uygulamada kullanılmasının engellenmesidir.

Temizlemeye tabi tutulan kayıtlar ve temizlenme sebepleri ile ilgili bazı örnekler şunlardır:

- a. Mümkün olmayan kayıtlar: Bir müşteri yaş kısmına bir değeri girip, meslek kısmına da “Avukat” değerini girmiştir. Bu durum mümkün olmadığı için o kayıt tamamen silinmiştir.
- b. Mümkün olan ama muhtemel olmayan Kayıtlar: Bu tür kayıtların, gerçek olmadığı ispatlanamaz, ancak şüpheli oldukları için ihmal edilebilirler: Bir müşteri, yaşına “115” değerini girip, meslek kısmına da “öğrenci” değerini girmiştir. O kayıt muhtemel olmadığı için uygulamada kullanılmamıştır.

Mümkün veya muhtemel olmayan kayıtların silinmemesi veya buna benzer durumların gözden kaçması algoritmanın yanlış yönde etkilenmesine ve

eğitim sürecinin etkinliğinin düşmesine yol açabileceği için güven vermeyen tüm kayıtlar dikkatlice temizlenmeye edilmeye çalışılmıştır.

5.3.3 Verileri Süzülerek Kullanışlı Hale Getirilmeleri

Süzme işlemi sistematik ve tesadüfi olarak gerçekleştirilmiştir. Sistematik süzmede, belirli özelliklere sahip kayıtlar modele dahil edilmemiştir. Örneğin şehir alanı sadece yedi büyük şehri kapsayan kayıtlar alınmış diğer kayıtlar ise göz ardı edilmiştir. Benzer şekil üyelik durumu aktif olmayan müşterilere ait kayıtlar ihmal edilmiştir.

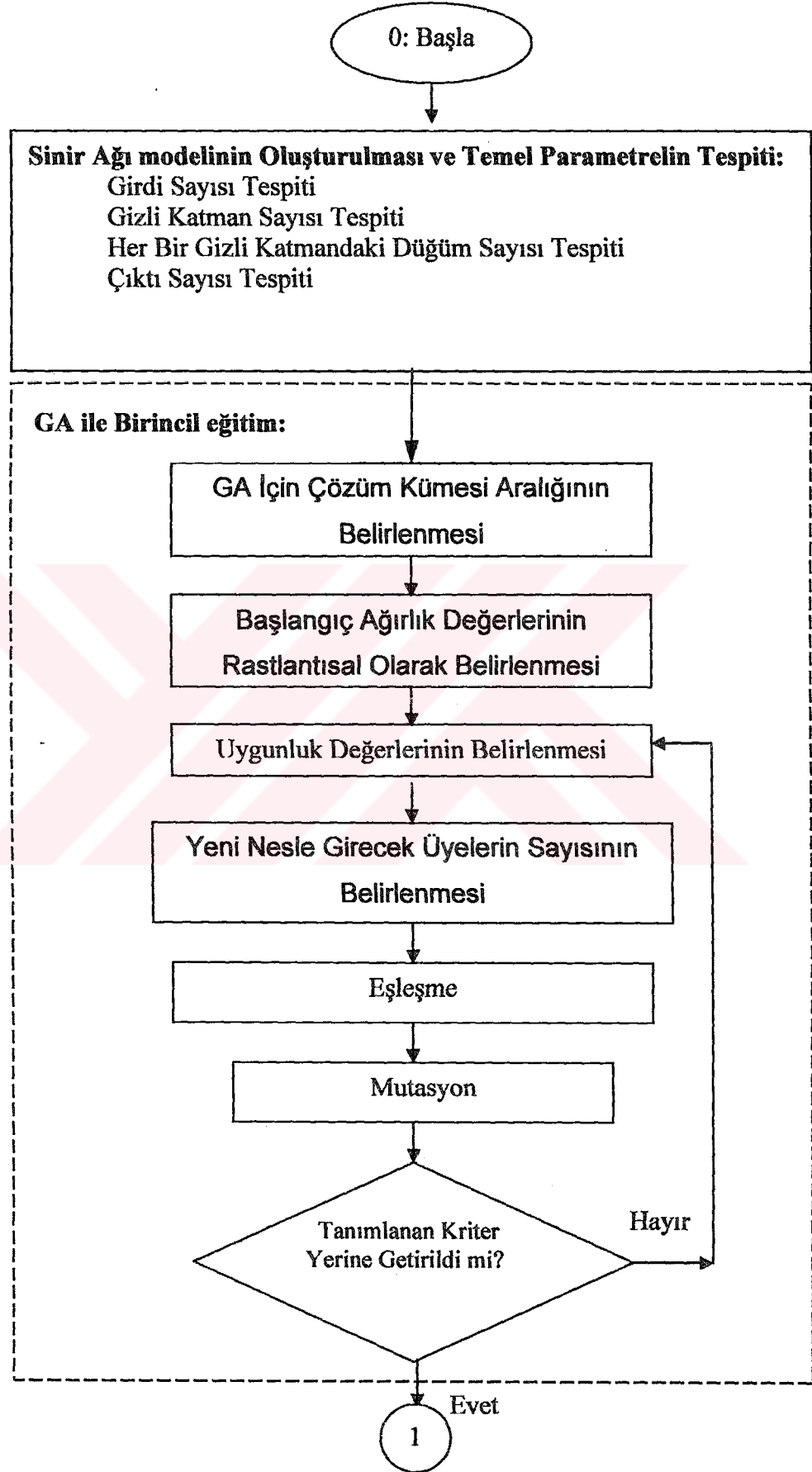
Uygulamada kullanılan veri setinin hacminin çok büyük seçilmesi işlem süresini uzatacağı için, gruplama çalışmasında rasgele seçim yöntemi ile belirli sayıda kayıt ile model uygulanmıştır. Sınıflandırma çalışmasında ise kullanılabilir olarak değerlendirilen kayıtların tamamı değerlendirmeye tabi tutulmuştur. Sınıflandırmada kullanılan toplam kayıt sayısı 136'dır. Gruplama çalışmasında da 275 kayıtlı işlem yapılmıştır.

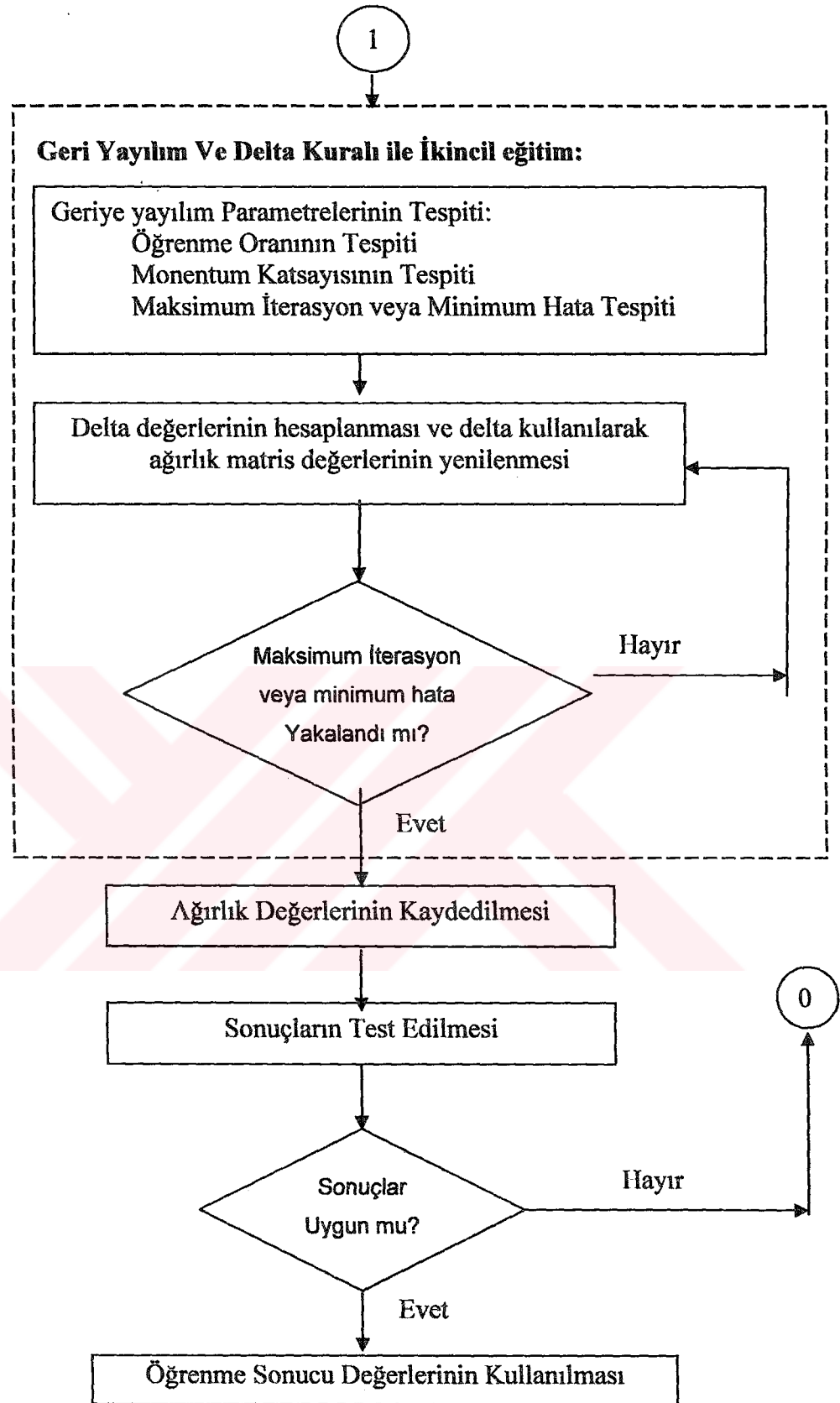
5.4 Modelin Oluşturulması

Daha önce de vurgulandığı gibi GA ve YSA tekniklerini iki kademeli bir melez yapı içerisinde kullanmak, bu çalışmanın temel amaçlarından birisidir.

5.4.1 Modelin Algoritmik Yapısı

GA ve YSA tekniklerini iki kademeli bir melez yapı içerisinde kullanma amacına yönelik olarak geliştirilen melez yapı şematik olarak Şekil 5.1'de sunulmuştur.





Şekil 5.1 Geliştirilen Melez Yapının Şematik Gösterimi

5.4.2 Uygulamanın Hedefleri

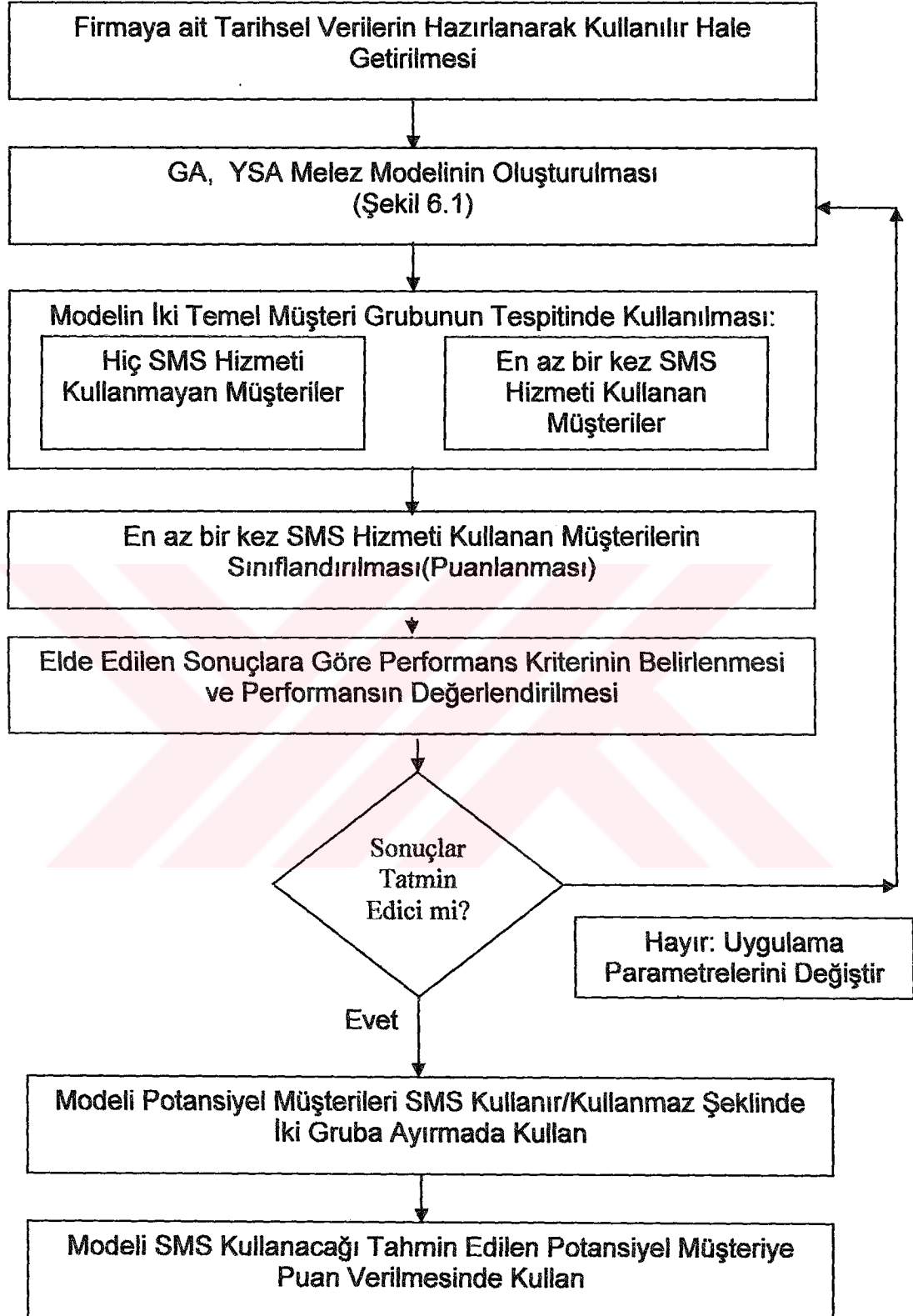
Şekil 5.1'de algoritmik temelleri anlatılmaya çalışılan melez model uygulamada da problemi iki aşamalı olarak çözmeyi hedeflemektedir:

1. Var olan müşterilerin gruplandırılması(Kümelendirilmesi): Uygulama verisi seti incelendiğinde azımsanmayacak sayıda SMS hizmetini hiç kullanmamış müşteri ile karşılaşmıştır. Bu durum öncelikle bir potansiyel müşterinin sözü geçen hizmeti kullanıp kullanmayacağını tespit edilmesi isteğini doğurmuştur. Çünkü bu tür müşteriler promosyondan bağımsız olarak bu konuya ilgi duymuyor olabilirler ve o müşterilere yönelik bambaşka hizmetlerin sunulması gerekebilir.
2. Var olan müşterilerin sınıflandırılması: İkinci aşama olan sınıflandırma çalışmasında ise SMS hizmetine ilgi duyan müşteriler hedef alınarak müşteri puanlama modeli ve bu modelin potansiyel müşteriler için tahmin aracı olarak kullanılması amacı güdülmüştür.

Her iki durumda da temel hedef minimum maliyetle müşteri memnuniyeti yaratmak yoluyla kurumun faydasını maksimize etmektir.

5.4.3 Uygulama Yöntemi ve Şeması

Geliştirilen modelin, uygulamanın amaçlarına göre organize edilmesi gerekmektedir. Buna yönelik olarak sistematik bir şekilde gerçekleştirilen uygulama; sonuçların elde edilmesi, değerlendirilmesi, gerekli düzenlemelerin yapılması ve elde edilen yapının kullanılması aşamalarında başarı şansını arttırabilir. Şekil 5.2 uygulamanın pratik akışını sunmak üzere sunulmuştur:



Şekil 5.2 Geliştirilen Melez Modelin Uygulanması

5.5 Modelin Uygulanması

Modelin uygulanması Şekil 5.2'de sunulan esaslara uygun şekilde gerçekleştirilmiştir. Daha önce hazırlanmış ve program kodları yazılmış algoritmanın, uygulamanın gereklerine göre kullanılması, bu aşamanın temel amacıdır.

5.5.1 Uygulama Veri Seti ve Aralıkları

Girdi katmana ait 12 değişken, müşteri gruplandırma ve sınıflandırma aynen kullanılmaktadır. Ancak çıktı katman değişkeni olan SMS kullanım adeti, grupta, kullanır(1) ve kullanmaz(0) şeklinde ikili değer almaktadır, sınıflandırmada ise bu değişken, [0,1] aralığında sürekli bir şekilde ifade edilmektedir.

Tablo 5.1 ile sunulan veriler, model tarafından otomatik olarak [0.1 , 1] aralığına dönüştürülmektedir.

5.5.2 Modelin, Müşteri Gruplandırmada Kullanılması

Müşteri gruplama çalışması için toplam 274 kayıt kullanılmıştır. Bu kayıtlardan 251 tanesi modelin eğitilmesinde, 23 tanesi de eğitimin genelleştirilebilirliğinin test edilmesinde kullanılmıştır.

YSA uygulamalarının en zor aşamalarından birisi olan uygun parametrelerin seçimi için çok sayıda deneme yapılarak, optimum sonuca ulaşılan değerler elde edilmiştir. Gruplamada kullanılan başlıca parametre değerleri Tablo 5.2'de özetlenmiştir:

Tablo 5.2 Müşteri Gruplamasına Ait Parametreler

Özellik	Değer / İsim	Açıklama
Girdi Katman Eleman Sayısı	12	Girdi katman elemanları, Tablo 5.1'de sunulan değişkenlerdir.
Girdi Katman Değişken Aralığı	[0.1 , 1]	
Gizli Katman Sayısı	1	
Gizli Katman Eleman Sayısı	28	Çok sayıda değer denenmiştir. Performans/süre optimizasyonu açısından en uygun değer seçilmiştir.
Çıktı Katman Eleman Sayısı	1	Çıktı katman, müşterinin SMS kullanıp kullanmadığını gösterir
Çıktı Katman Değer Aralığı	0,0001 ve 0,9999	SMS kullanan Müşteriler: 0,9999 SMS kullanmayan müşteriler: 0,0001
Aktivasyon Fonksiyonu	Toplam Fonksiyonu	Formül 5.1
Transfer Fonksiyonu	Sigmoid Fonksiyon	Formül 5.2
GA iterasyon Sayısı	30	İlk aşamada kullanılan GA ile 30 iterasyon sonrası elde edilen değerler delta kuralı ile geliştirilmiştir.
Toplam İterasyon Sayısı	2200	
Öğrenme Oranı	0,002 ve 0,001	İlk aşamada 0,002 değeri ile öğrenme gerçekleştirilmiştir (ilk 1000 iterasyon). Sonrasında ise 0,001 kullanılmıştır.
Momentum Katsayısı	0,9	Momentum katsayısı, öğrenme süreci boyunca değiştirilmemiştir.

Müşteri gruplaması sırasında bazı yaklaşımlar kullanılmıştır. Örneğin girdi katman değerleri [0 , 1] aralığında değil de [0.1 , 1] aralığında ifade edilmiştir. Bu ve buna benzer temel yaklaşımlar ve sebepleri açıklanmalıdır:

5.5.2.1 Girdi Katman Değer Aralığı

Genel olarak sıfır, bir aralığında değerler alan değişkenler için, modelde en düşük değer olarak sıfır yerine 0.1 değeri kullanılmıştır. Bu yaklaşımın basit bir sebebi vardır; sürekli olmayan değişkenlerin kullanıldığı ve değişken alabilecekleri değerlerin az sayıda olduğu durumlarda bu uygulama, bir çok durumda daha başarılı sonuçlar verebilmektedir.⁸³

Bu duruma en güzel örnek, cinsiyet değişkenidir. Cinsiyet, bay ve bayan değerlerini alabilir. Eğer modelde bay için *sıfır* ve bayan içinse *bir* değerleri atanırsa, kullanılan kayıtların hemen hemen yarısında bu değer *sıfır* olacaktır. Formül 5.1'de ifade edilen aktivasyon fonksiyonu değerleri sözü geçen tüm kayıtlar için sıfır olacaktır, bu durumda öğrenme sürecinde ağırlık değerlerinin değişiminin, sözü geçen değişkenler ve kayıtlar için hiçbir etkisi olmayacaktır. Yani öğrenme süreci bu değerlerin özelliklerinden yararlanamayacaktır.

5.5.2.2 Çıktı Katman Değer Aralığı

Çıktı Katman değer aralığında da girdi katman değer aralığı seçimine benzer bir problem vardır. Buradaki problem ise, Formül 5.2 ile ifade edilen transfer fonksiyonu ile ilgilidir. Eğer SMS kullanan müşterilere *bir*, kullanmayanlara *sıfır* değerleri atanırsa, transfer fonksiyonu ağırlık değerlerini, $\pm\infty$ 'a doğru taşımaya çalışacaktır.⁸⁴ Bu durum aşağıdaki ifadelerle açıklanmıştır.

⁸³ BOSE N. K, LIANG P.: a.g.e., s. 176.

⁸⁴ BOSE N. K, LIANG P.: a.g.e., s. 175.

$$\lim_{x \rightarrow \alpha} \left[\frac{1}{(1 + e^{-\lambda x})} \right] = 1 \quad (5.1)$$

$$\lim_{x \rightarrow -\alpha} \left[\frac{1}{(1 + e^{-\lambda x})} \right] = 0 \quad (5.2)$$

Bu durumu ortadan kaldırmak için SMS kullanan müşterilere 0,999 değeri, kullanmayanlara ise 0,0001 değeri atanmıştır.

5.5.2.3 Sabit Olmayan Öğrenme Oranı

Öğrenme oranını tüm süreç boyunca sabit bir değer olarak seçmemeye, çalışmamızda yaptığımız denemeler sonucunda karar verilmiştir. Küçük değerlerde seçilen öğrenme oranı, öğrenmeyi yavaşlatarak süreci uzatmaktadır. Buna karşılı büyük seçilen değerler ise öğrenme ve test verilerinin hata değerleri arasında çabuk bir şekilde fark oluşmasına yol açarak öğrenmenin etkinliğini azaltmaktadır. Bu sorunu çözmek için öğrenmenin başlarında, daha büyük öğrenme oranı kullanılmıştır. Sonraki aşamada ise hassas ilerleme sağlamak için daha küçük öğrenme oranı kullanılmıştır.

5.5.2.4 Model Tarafından Elde Edilen Tahmini Çıktı Değerlerinin Kullanımı

Geliştirdiğimiz algoritmayı kullanarak elde ettiğimiz tahmini çıktı değerleri, aşağıdaki gibi değerlendirilmiştir:

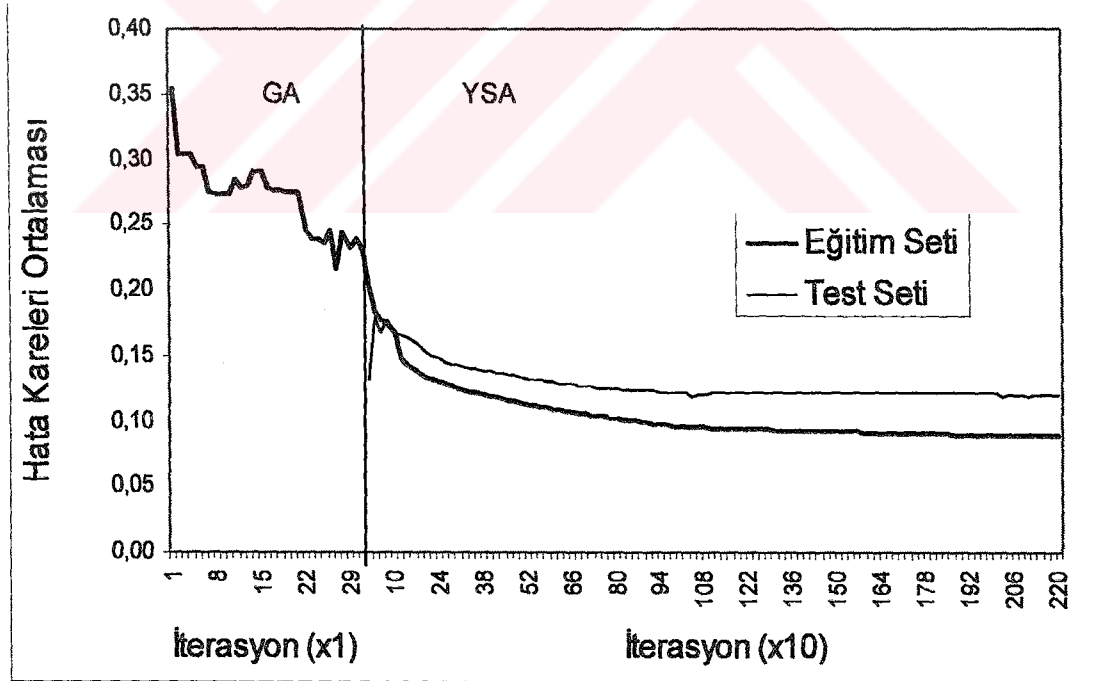
$$\text{Eğer} \begin{cases} \text{Tahmini Çıktı Değeri} \geq 0,5 \Rightarrow 1 \text{ (Müşteri SMS Kullanır.)} \\ \text{Tahmini Çıktı Değeri} < 0,5 \Rightarrow 0 \text{ (Müşteri SMS Kullanmaz.)} \end{cases}$$

Gerçek çıktı değerleri de benzer şekilde değerlendirilerek yapılan inceleme sonucunda elde edilen sonuçlar Tablo 5.3'te sunulmuştur.

Tablo 5.3 Müşteri Gruplama Sonuçları

	Doğru Tahmin Yapılan Müşteri Sayısı	Doğru Tahmin Yapılan Müşteri Sayısı	% Hata
Eğitim Veri Seti	232	19	%7,6
Test Veri Seti	19	4	%17

Şekil 5.3, GA, YSA melez modelinin, öğrenme sürecini ve iterasyon sayısına bağlı olarak hata kareleri ortalamasının değişimini göstermektedir.



Şekil 5.3. Müşteri Gruplamasına Ait Öğrenme Süreci ve Hata Değerleri

Burada modelin alışmasının 2200 iterasyon ile sınırlı tutulmasının sebebi, bu noktadan sonra eğitim veri setinin mutlak hata deęerleri azalırken, test veri setine ait mutlak hataların artmasıdır. Yani model fazla eğitime doęru yol almaya başladığı için öğrenme işlemi bu noktada kesilmiştir.

Tablo 5.4 Müşteri Gruplamada İterasyon Sayısı ve Hata Kareleri Ortalaması Deęerlerinin, Eğitim ve Test Verileri İçin Deęişimi

İterasyon Sayısı	Hata Kareleri Ortalamaları	
	Eğitim Veri Seti	Test Veri Seti
GA Başlangıç	0,35444	-
GA Bitiş	0,23019	-
0	0,20527	0,13218
200	0,13413	0,15327
400	0,11904	0,13776
600	0,10933	0,12962
800	0,10201	0,12466
1.000	0,09608	0,12187
1.200	0,09417	0,12239
1.400	0,09285	0,12245
1.600	0,09167	0,12237
1.800	0,09056	0,12240
2.000	0,08955	0,12179
2.200	0,08924	0,11995
<i>Aşırı Eğitim Sonrası Deęerler:</i>		
3.000	0,08848	0,12355

Yukarıda öğrenme süreci ve sonuçları sunulan model, ağırlık matrisi deęerleri kullanılarak, daha sonra karşılaşılan müşterilere uygulanacaktır. Bunun için ağırlık matrisi deęerlerinin saklanması gerekmektedir. Ağırlık matrisi deęerleri, Tablo 5.5 ve Tablo 5.6'da sunulmuştur.

Tablo 5.5 Müşteri Gruplamada, Eğitim Sonrası Girdi Katman, Gizli Katman Arası Ağırlık Değerleri

		Girdi Katmanı Eleman Numarası												
		0	1	2	3	4	5	6	7	8	9	10	11	12
Gizli Katman Eleman Numarası	1	3,978	4,780	1,701	-6,859	3,917	-5,006	3,334	6,478	3,118	-1,586	3,495	2,626	3,351
	2	6,590	0,405	-1,356	-1,106	5,024	-1,591	-0,292	4,655	4,761	5,639	2,891	-1,817	4,675
	3	2,315	-0,969	0,791	-4,505	2,847	3,216	-1,658	-0,389	-3,605	-4,530	4,880	6,744	-0,340
	4	-4,801	0,673	5,680	-1,532	6,392	2,607	-6,485	-0,817	3,093	-5,365	4,036	1,819	6,019
	5	-4,281	-5,083	-0,229	1,129	-3,311	0,811	8,689	-1,956	5,373	6,798	1,348	4,189	3,530
	6	-4,877	1,223	-3,493	2,197	-5,339	6,642	-0,008	-2,499	4,372	6,004	3,325	4,721	0,459
	7	-2,281	-1,360	-3,964	-0,151	-6,175	-5,673	0,115	-6,052	2,098	6,584	-0,657	1,780	-4,619
	8	-2,436	-0,384	-5,060	0,873	-1,939	6,861	4,540	2,153	-2,572	-6,782	6,470	-1,153	1,837
	9	0,364	6,413	-5,462	-3,565	-0,478	4,529	5,576	-5,317	-3,438	0,864	6,347	3,849	-0,638
	10	-0,194	0,435	0,958	-1,610	-1,776	-3,519	0,192	0,351	-6,637	5,479	-6,161	4,898	0,211
	11	-1,581	2,779	2,287	3,413	-2,068	-0,220	-4,796	3,008	2,754	5,931	-0,400	7,042	-1,168
	12	-4,730	2,862	-5,176	-5,132	2,332	-0,427	-4,074	-6,236	-2,473	4,627	0,275	-0,669	-5,946
	13	-6,850	-5,253	-6,633	-3,336	-0,469	-5,142	-1,373	3,596	4,875	-7,129	0,224	0,390	1,602
	14	-1,327	0,268	-1,667	-5,115	3,866	-2,910	-5,465	-6,943	-3,177	3,824	3,120	0,281	-3,487
	15	1,850	-0,901	3,511	-1,054	3,347	-2,781	-6,188	-5,355	-7,390	-5,886	4,214	2,452	1,738
	16	-6,182	-2,765	0,407	1,196	-4,273	5,803	7,468	-3,637	3,415	-5,502	1,644	4,432	-3,932
	17	-6,101	-3,400	2,932	5,756	-6,938	-2,521	3,345	-5,318	-6,224	-5,568	2,860	1,051	-3,403
	18	3,555	-3,886	-3,349	-2,589	0,673	1,182	-4,027	5,671	-2,665	-0,950	3,326	2,133	5,827
	19	1,772	-4,653	-2,224	6,933	-3,623	-2,773	6,483	2,433	1,531	8,438	1,442	-0,090	-1,111
	20	0,257	1,455	-8,748	0,038	-2,958	5,686	1,044	-2,288	6,641	6,871	-5,244	1,919	-3,381
	21	1,447	-6,819	-5,425	5,444	1,775	3,416	-0,764	-3,510	-1,796	3,217	-3,566	1,629	-3,417
	22	-6,571	-1,526	-1,288	-0,047	-4,305	-1,522	3,634	-2,129	-2,604	4,507	5,545	-4,094	6,380
	23	6,295	-0,855	-4,182	6,029	3,765	5,324	6,244	-1,503	4,527	-6,150	2,122	2,263	-4,518
	24	-3,873	0,575	-3,312	5,453	-0,093	1,835	-3,695	6,303	4,191	0,765	4,919	6,639	4,689
	25	4,531	6,464	3,860	4,351	1,790	-2,628	-1,862	2,663	-4,420	-0,945	-6,657	-0,949	-4,093
	26	-0,667	5,985	-4,625	-0,495	4,483	6,727	-5,474	4,048	4,940	-7,681	-0,946	5,876	0,246
	27	-4,147	-5,270	2,901	-4,347	-0,748	4,516	4,097	-5,222	2,971	-6,526	-0,578	-2,867	-4,660
	28	-6,900	5,860	0,487	-2,454	2,137	-5,289	1,930	5,052	-3,951	-6,064	4,610	-6,692	-4,369

**Tablo 5.6 Müşteri Gruplamada, Eğitim Sonrası Gizli Katman, Çıktı
Katman Arası Ağırlık Değerleri**

Gizli Katman Eleman Numarası	Ağırlık Değeri	Gizli Katman Eleman Numarası	Ağırlık Değeri
0	-4,057	15	-5,413
1	5,930	16	-4,183
2	-2,094	17	-5,475
3	-3,304	18	-3,647
4	-4,426	19	3,095
5	-5,180	20	-3,578
6	5,448	21	3,391
7	2,167	22	2,942
8	5,706	23	5,522
9	-1,635	24	6,590
10	-4,645	25	3,778
11	-3,650	26	-2,652
12	-4,248	27	6,176
13	2,579	28	6,426
14	5,780		

Modelin uygulamada kullanımı sırasında, Tablo 5.5 ve Tablo 5.6'da verilen ağırlık matrisi değerleri ve bir potansiyel müşterinin kişisel bilgileri ile birlikte kullanılarak tahmini olarak o potansiyel müşterinin SMS hizmetinden yararlanıp yararlanmayacağı tahmin edilir. Eğer bu analiz sonucunda tahmin kullanır şeklinde olursa, bu durumda da kullanıp oranını tahmin etmek amacıyla, müşteri sınıflandırma analizi uygulanır.

5.5.3 Modelin, Müşteri Sınıflandırmada Kullanılması

Müşteri sınıflandırmada, toplam 136 kayıt kullanılmıştır. Bunlardan 124'ü eğitim seti olarak değerlendirilirken bunun yaklaşık %10'u yani 12 kayıt da test seti olarak değerlendirilmiştir.

Sınıflandırmada, kullanılabilir nitelikte olan 456 kayıtlı analiz yapılmıştır ancak yapılan tüm eğitimlerde istenen sonuçlar elde edilememiştir. Bunun sebebi araştırıldığında, veri setinin çıktı katman değerlerinin çok kuvvetli asimetriye sahip olduğu için eğitimi olumsuz yönde etkilediği gözlenmiştir.

Sıfır, bir aralığında değişen SMS kullanım oranı değerinin ortalamasının 0.09 olduğu tespit edilmiştir. Bu da algoritmanın tüm tahmin değerlerini sıfıra yaklaştırma eğilimine girmesine yol açmıştır.

Canlılardaki öğrenme mekanizmasına benzer bir şekilde, öğrenme ortamının (canlılar için çevresel koşulların) performansı etkilediği değerlendirilmiştir. Bu nedenle asimetriyi azaltmak ve çıktı katman ortalamasını değer aralığının ortasına yaklaştırmak amacıyla çok küçük çıktı değerine sahip çok sayıdaki kayıttan bazıları rasgele olarak çıkarılmıştır. Bu işlem sonucunda elde edilen veri setinin çıktı değerleri ortalaması 0.33 değerine yükselmiştir.

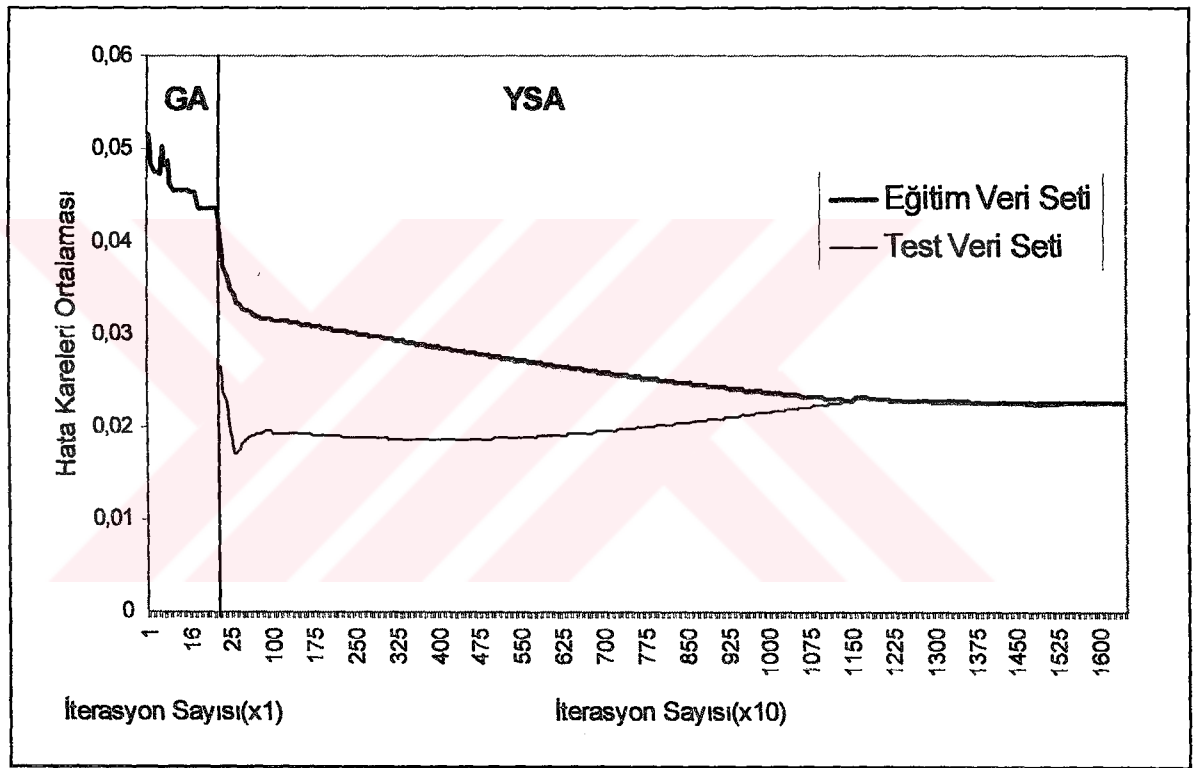
Eğitim sürecinde kullanılan parametrelerin belli başlıcaları Tablo 5.7'de özetlenmiştir. Girdi katman değişken aralığının [0.1 , 1] aralığında alınmasının ve değişken öğrenme oranının kullanılmasının sebepleri, müşteri gruplama çalışmasındaki sebeplerin aynısıdır.

Tablo 5.7 Müşteri Sınıflandırmasına Ait Parametreler

Özellik	Değer / İsim	Açıklama
Girdi Katman Eleman Sayısı	12	Girdi katman elemanları, Tablo 5.1'de sunulan değişkenlerdir.
Girdi Katman Değişken Aralığı	[0.1 , 1]	
Gizli Katman Sayısı	1	
Gizli Katman Eleman Sayısı	42	Çok sayıda değer denenmiştir. Performans/süre optimizasyonu açısından en uygun değer seçilmiştir.
Çıktı Katman Eleman Sayısı	1	Çıktı katman, müşterinin SMS oranını gösterir.
Çıktı Katman Değer Aralığı	[0 , 1]	En çok SMS kullanan müşteri: 1 En az SMS kullanan müşteri: 0
Aktivasyon Fonksiyonu	Toplam Fonksiyonu	Formül 5.1
Transfer Fonksiyonu	Sigmoid Fonksiyon	Formül 5.2
GA iteraasyon Sayısı	25	İlk aşamada kullanılan GA ile 25 iterasyon sonrası elde edilen değerler delta kuralı ile geliştirilmiştir.
Toplam İterasyon Sayısı	11.500	
Öğrenme Oranı	0,01 ve 0,006	İlk aşamada 0,01 değeri ile öğrenme gerçekleştirilmiştir(ilk 2500 iterasyon). Sonrasında ise 0,006 kullanılmıştır.
Momentum Katsayısı	0,9	Momentum katsayısı, öğrenme süreci boyunca değiştirilmemiştir.

5.5.3.1 Müşteri Sınıflandırması Çalışmasında Hata Teriminin İterasyona Bağlı Değişimi

Başlangıç eğitimi, 25 iterasyon ile GA prensibi ile yapmış ve bu ilk eğitimle elde edilen değerler, 16.450 iterasyonla geliştirilmiştir. Şekil 5.4, hata kareleri ortalamasının iterasyona bağlı olarak değişimini göstermek üzere sunulmuştur:



Şekil 5.4. Müşteri Gruplamasına Ait Öğrenme Süreci ve Hata Değerleri

Eğitim sırasında elde edilen değerler, Tablo 5.8'de sayısal olarak sunulmuştur.

Tablo 5.8 Müşteri Sınıflandırmada İterasyon Sayısı ve Hata Kareleri Ortalaması Değerlerinin, Eğitim ve Test Verileri İçin Değişimi

İterasyon Sayısı	Hata Kareleri Ortalamaları	
	Eğitim Veri Seti	Test Veri Seti
GA Başlangıç	0,05143	-
GA Bitiş	0,04376	-
0	0,03929	0,02642
1.000	0,03162	0,01945
2.000	0,03064	0,01911
3.000	0,02963	0,01882
4.000	0,02860	0,01871
5.000	0,02763	0,01881
6.000	0,02672	0,01910
7.000	0,02587	0,01956
8.000	0,02510	0,02015
9.000	0,02438	0,02085
10.000	0,02373	0,02162
11.000	0,02312	0,02245
12.000	0,02301	0,02300
13.000	0,02278	0,02269
14.000	0,02269	0,02253
15.000	0,02263	0,02249
16.000	0,02257	0,02253
16.450	0,02254	0,02255

5.5.3.2 Müşteri Sınıflandırması Çalışmasında Elde Edilen Ağırlık Matrisi Değerleri

Tablo 5.9 ve Tablo 5.10, müşteri sınıflandırma ağırlık matrisi değerlerini sunmak üzere hazırlanmıştır:

Tablo 5.9 Müşteri Sınıflandırmada, Eğitim Sonrası Girdi Katman, Gizli Katman Arası Ağırlık Değerleri

		Girdi Katmanı Eleman Numarası												
		0	1	2	3	4	5	6	7	8	9	10	11	12
Gizli Katman Eleman Numarası	1	1,321	-1,260	0,997	-0,971	1,425	0,312	-0,314	1,957	-0,537	-1,430	0,237	0,232	-0,620
	2	-1,070	-0,230	0,409	1,586	-0,476	1,828	1,821	-0,784	-2,924	-0,142	-1,325	-1,063	1,137
	3	-0,161	1,888	0,529	1,672	-1,198	1,127	-0,645	-1,418	1,011	-0,328	0,565	0,189	0,829
	4	1,444	-1,044	1,819	0,907	1,891	-1,235	0,694	-0,796	1,290	0,016	-0,417	1,441	-1,011
	5	1,401	1,443	-0,391	1,008	0,176	1,512	-2,255	2,337	-1,960	-2,021	1,950	0,514	-1,691
	6	0,493	0,007	-0,886	0,670	-1,915	1,018	-0,732	1,427	0,524	0,799	0,290	0,666	0,452
	7	0,407	0,924	-1,143	0,856	-0,110	-1,728	-1,111	1,128	-0,505	-0,122	-1,966	1,009	1,189
	8	-0,188	0,156	-1,990	-1,151	-0,809	-1,956	1,438	-1,210	-1,653	-1,090	1,236	-1,425	-1,441
	9	0,643	1,966	0,409	-0,947	0,650	0,888	-0,804	1,697	1,861	0,841	-0,768	-0,140	-0,951
	10	0,681	1,581	1,158	0,917	-0,992	0,464	-1,003	1,508	0,152	0,489	-1,595	0,955	-0,430
	11	-1,713	-0,066	0,700	1,242	1,162	0,444	-1,021	-0,861	-0,460	-1,908	-0,254	-1,264	0,708
	12	1,475	-1,375	-1,473	1,191	1,772	-0,906	-1,117	-0,069	-0,022	1,624	2,024	-0,790	-1,766
	13	1,754	1,835	1,857	0,080	1,428	-1,400	-0,033	0,231	0,748	-0,972	-1,275	0,996	1,501
	14	-0,967	-1,920	-1,056	1,791	2,060	-0,432	-0,138	-1,001	0,593	1,323	1,548	-1,428	-1,478
	15	0,127	0,267	-1,933	-1,733	0,979	0,838	-0,361	0,046	-0,260	-0,163	1,580	-0,027	-0,066
	16	0,863	-1,108	-1,407	-1,699	2,512	0,902	-0,751	1,638	-1,589	-0,823	0,380	-1,771	-1,954
	17	0,166	1,458	-0,103	0,310	1,321	-0,283	0,086	0,928	-0,419	1,510	0,170	-0,227	2,022
	18	-0,829	0,142	-0,313	-1,594	0,649	-1,714	0,473	-1,709	-0,629	-1,371	1,519	0,356	0,715
	19	-0,861	0,813	-1,361	-1,918	1,643	-1,223	-1,236	-1,113	0,282	-1,037	0,825	0,847	-0,590
	20	-1,378	-1,333	-1,773	0,940	-1,248	-1,313	-0,209	0,997	-0,755	1,359	0,618	1,070	-1,957
	21	-0,962	0,446	-0,800	-0,203	-0,389	1,640	-1,640	-0,842	0,867	0,649	0,368	-0,225	0,387
	22	-0,497	-2,086	1,659	1,345	0,242	1,055	1,241	-1,031	1,814	-1,631	-0,688	0,256	-1,434
	23	1,406	-2,170	0,015	0,041	-0,449	-0,109	2,047	0,016	0,545	-1,583	-1,771	-0,802	-1,595
	24	0,603	-2,665	3,300	-0,577	-0,108	0,020	-1,791	0,071	1,891	-0,467	-1,520	1,325	0,590
	25	1,765	-0,404	1,273	-1,389	-0,287	0,683	1,180	0,956	-0,632	-0,951	1,620	-0,455	1,085
	26	0,654	-1,849	1,219	0,154	1,506	-1,410	0,191	0,578	1,429	0,059	1,013	1,686	2,039
	27	0,634	0,436	-1,217	1,055	0,136	1,321	-0,056	-0,704	-0,215	-0,474	1,527	1,481	-0,807
	28	-0,661	1,710	0,872	-2,017	-1,746	-1,322	1,772	0,882	0,492	-1,239	-1,465	-0,586	-0,923
	29	0,421	-1,392	1,154	0,970	-1,384	0,829	-1,209	1,080	-2,041	-1,196	1,880	1,492	2,128
	30	-0,615	-1,682	-1,629	0,105	-0,212	-0,825	1,649	0,711	-0,243	0,212	1,077	-0,723	0,010
	31	1,191	-1,000	-1,040	-0,728	1,321	1,277	-0,104	-0,582	1,393	1,704	1,260	-1,872	-1,556
	32	-1,965	1,108	-0,595	0,739	0,465	1,404	-0,568	-1,536	-0,020	-1,936	1,935	-1,905	-1,728
	33	-1,383	-1,543	2,055	0,660	-2,126	-0,677	0,553	-0,438	1,738	1,308	1,681	0,420	1,707
	34	1,576	0,344	0,426	0,970	0,368	1,540	-0,304	0,697	-0,916	-0,689	1,775	1,246	0,691
	35	1,806	-1,944	2,040	-1,125	0,060	-1,503	-1,056	0,147	1,864	0,549	-1,826	0,609	0,637
	36	0,845	-1,547	-0,444	-2,023	1,664	0,802	1,342	-2,280	-0,800	-1,583	-0,161	0,801	1,009
	37	1,349	-0,873	-2,075	1,465	0,701	1,402	-1,513	1,580	1,515	1,527	-1,438	-1,260	-0,446
	38	-1,219	-1,845	2,004	1,195	0,843	-0,896	-0,112	0,040	1,759	0,901	-0,614	-1,600	-0,535
	39	0,492	-1,657	-0,610	-1,767	0,908	-0,773	-0,216	1,283	-1,433	-0,207	1,295	-0,363	-1,481
	40	-0,459	0,255	0,830	-0,481	-1,639	-1,951	-1,704	-1,969	0,759	-2,016	0,844	-0,010	1,614
	41	0,847	1,071	1,020	-1,612	-0,282	0,633	0,187	-1,861	1,132	-2,265	1,533	-1,461	0,687
	42	-0,188	1,534	-0,704	0,620	-1,223	0,653	-1,666	-0,375	-0,331	-0,528	-0,756	0,347	1,519

Tablo 5.10 Müşteri Sınıflandırmada, Eğitim Sonrası Gizli Katman, Çıktı Katman Arası Ağırlık Değerleri

Gizli Katman Eleman Numarası	Ağırlık Değeri	Gizli Katman Eleman Numarası	Ağırlık Değeri
0	0,805	22	1,317
1	-0,183	23	1,317
2	-2,000	24	-2,994
3	-0,610	25	-0,638
4	-1,343	26	0,793
5	2,287	27	-2,280
6	0,237	28	-2,478
7	-0,918	29	2,652
8	1,937	30	-0,864
9	1,105	31	2,025
10	1,270	32	-0,631
11	-1,604	33	0,777
12	-2,109	34	0,219
13	-1,700	35	-1,365
14	-1,462	36	2,298
15	-0,276	37	2,336
16	-2,037	38	1,644
17	-0,982	39	1,001
18	-1,316	40	-0,760
19	0,046	41	-2,029
20	-0,962	42	0,991
21	0,762		

5.5.3.4 Müşteri Sınıflandırması Çalışmasında Elde Edilen Sonuçların Kullanımı

Sınıflandırma çalışmasında, öncelikle iki ondalık hassasiyetle [0,1] aralığında ifade edilmiş gerçek çıktı değerleri tahmin edilmeye çalışılmaktadır. Bunun arkasından elde edilen değerler, On ana grupta puanlanmaktadır. Yani müşterilere bir ile on değerleri arasında tamsayı olan bir puan verilir.

Elde edilen sonuçlarda, bir ve on arası verilen notlarda ortalama mutlak hata bir seviyesinde gerçekleşmiştir. Veri seti büyüklüğü ve veri özellikleri dikkate alındığında, bu koşullarda daha iyi sonuçlar elde edilememiştir. Ancak bilişim sisteminin önemli bir ögesi olan insan faktörü burada devreye girmelidir;

Yöneticiler, modelle elde edilen puanları nokta olarak değil de bir aralık olarak değerlendirmelidirler. Model tarafından, potansiyel bir müşteriye [1,10] aralığında verilen tahmini bir puan bir aralık olarak değerlendirilmelidir:

$$\text{Tahmini Puan} = \text{Nokta Puan} \pm 1 \quad (5.3)$$

Burada yöneticinin tecrübesi ve öznel kanaati çerçevesinde, kararında, elde edilen herhangi bir nokta sonucun bir puan üstünü, bir puan altını veya aynısını verebilmesi esnekliği sağlanmıştır.

5.6 Modelin Performansının Test Edilmesi

Modelin performansının test edilmesi için üç ayrı yol izlenmiştir:

1. Algoritmanın çalışması sırasında, öğrenme sürecini yönlendiren test,
2. Algoritmanın performansını ve sonuçların genelleştirilebilirliğini gösteren test,
3. Algoritmanın uygun sonuçlar üretebildiğini araştırmak amacıyla "Mat Lab" isimli paket programın sonuçları ile kıyaslama prensibine dayanan test.

İlk iki test her uygulamada tekrarlanması gereken testlerdir. Bu testler algoritmanın başarısını değil, eğitim sürecinin etkinliğini test etmeyi

amaçlarlar. Üçüncü test ise, algoritmanın kullanılabilirliğini araştırmak amacıyla, deneme sürecinde yeterli miktarda gerçekleştirilmiş ve algoritmanın performansı incelenmiştir.

5.6.1 Model Performansını Geliştirici Test

Modelin performansını geliştirmeye yönelik olan test, GA ile eğitim yapılan aşamada kullanılmıştır. Bu aşamada, yerel olarak hata terimini en aza indiren değerlerin, sürekli olarak kabul görmesini engellemek amacıyla GA ile yapılan optimizasyona hataları en aza indirmenin yanı sıra bir amaç daha eklemiştir:

GA hata kareleri ortalamasını minimum yapmaya yönelik olarak çalışırken aynı zamanda da eğitim seti içinden rasgele seçilmiş bir alt ile o set dışında kalan eğitim seti verilerinin hata kareleri ortalamaları arasındaki farkı sıfıra yaklaştırmaya çalışmaktadır. Bu durum GA ile daha düşük hata değerlerinin elde edilmesini zorlaştırırken sonuçların genelleştirilebilirliğini arttırmaktadır. Bu durum Ek 2'de sunulan ve tarafımızdan geliştirilen programın 13 numaralı satırında gözlenebilir.

5.6.2 Aşırı Öğrenmeyi Engelleyici Test

Aşırı öğrenme, oluşturulan modelle otomatik olarak gözlemlenmektedir. Toplam veri setinden yaklaşık olarak %10 oranında rasgele olarak seçilen test veri seti, öğrenmenin genele uygulanabilirliğinin incelemektedir.

Çalışmamız sırasında geliştirdiğimiz modelde, GA ile eğitim kısmında zaten eğitim seti içerisinde bir miktar veri seçilerek model performansını geliştirici

test uygulandığı ve daha sürecin başında olduğu için bu aşamada aşırı öğrenmeyi engelleyici teste gereksinim duyulmamıştır.

Geleneksel YSA algoritması ile öğrenmeye geçildikten sonra ise eğitim ve test verilerinin hata kareleri ortalamaları rutin olarak hesaplanmaktadır. Test seti hata seviyesinin, eğitim seti hata seviyesinden anlamlı bir şekilde farklılaşmaya başladığı anda eğitim otomatik olarak sonlandırılmaktadır. Bu durum Ek 2'de sunulan program kodlarından 12 numaralı satırda görülmektedir.

5.6.3 Algoritmanın Geçerliliğini Araştıran Test

Geliştirilen modelin ve algoritmanın geçerli ve kullanışlı olduğunu ispatlamak amacıyla, çalışmamızda geliştirdiğimiz modelin sonuçlarını "Mat Lab" isimli paket programın YSA modülünün sonuçlarıyla kıyasladık.

Kıyaslama hem gruplama çalışması, hem de sınıflandırma çalışması için gerçekleştirildi.

5.6.3.1 Müşteri Gruplama Uygulaması İçin GA-YSA Melez Modeli ile Mat-Lab Programının Sonuçlarının Kıyaslanması

Mat Lab programının sunduğu çok sayıda, öğrenme algoritması, transfer fonksiyonu ve benzer parametreler mevcut olduğu için öncelikle modelimizin ikinci aşamasında kullanılan genelleştirilmiş delta kuralı temel alınarak Mat Lab programı ile YSA eğitilmiştir. Buradan elde edilen sonuçlar gerçekçi olmaktan çok uzaktır. Eğitim sonrası hata kareleri ortalaması 2230 iterasyon için 0,1435 bulunmuştur. Bu sonuç, bizim modelimizle elde edilen sonuçtan daha kötüdür.

Aynı veriler, daha gelişmiş ve fonksiyonların ikinci türevlerini de barındıran diğer bir algoritma olan “Levenberg-Marquardt” (Mat Lab’te kullanılan ismi ile: trainlm) ile tekrarlanmıştır. 100 iterasyon sonrası eğitim seti hata değerleri çok küçük bir değer olan 0,008 seviyesine indirilmiştir. Ancak aynı sonuçlar test veri setinde denendiğinde hata kareleri ortalamasının 0,74 gibi çok büyük bir değerde olduğu gözlenmiştir.

Tablo 5.11 Müşteri Gruplamaya Ait Çeşitli Denemelere Ait Sonuçlar

	İterasyon Sayısı	Hata Kareleri Ortalaması	
		Eğitim Seti	Test Seti
Bu Çalışma Kapsamında Geliştirilen Modelle Elde Edilen Sonuçlar	30+2,200	0,08924	0,11995
Mat Lab, Genelleştirilmiş Delta Kuralı ile Elde Edilen Sonuçlar	2,230	0,1435	0,328
Mat Lab, Genelleştirilmiş Delta Kuralı ile Elde Edilen Sonuçlar	10,000	0,1082	0,143
Mat Lab, Levenberg-Marquardt” (trainlm) ile Elde Edilen Sonuçlar	100	0,008	0,74
Mat Lab, Levenberg-Marquardt” (trainlm) ile Elde Edilen Sonuçlar	9	0,089162	0,218

5.6.3.2 Müşteri Sınıflandırma Uygulaması İçin GA-YSA Melez Modeli ile Mat-Lab Programının Sonuçlarının Kıyaslanması

Sınıflandırma çalışması ile ilgili Mat Lab YSA aracı ve çalışmamızda geliştirdiğimiz modeli kıyaslarken, Mat Lab aracının her seferinde aynı veriler ve ağ yapısı ile aynı şekilde eğitim performansı gösteremediği fark edilmiştir. Tesadüfi olarak seçilen başlangıç ağırlık değerleri bazı durumlarda yerel

minimuma yönelmeye yol açabilmektedir. Çalışmamızda geliştirilen model bu durumu engellemek amacıyla başlangıçta GA ile eğitim yapmaktadır. GA'nın çok sayıda popülasyon üyesi barındırması ve mutasyon mekanizması yerel minimuma ulaşılma riskini bir ölçüde azaltmaktadır. Çünkü bir gen ne kadar görece iyi bir küme elemanına ait olursa olsun mutasyona uğrama şansı vardır ve bu durum, eğer algoritma yerel minimum etrafında sıkışıp kaldıysa, için bir kurtuluş imkanı yaratır. Bu deterministik bir kural olmamakla birlikte, yaptığımız deneylerde çoğunlukla pratik sonuçlar doğurduğu gözlenmiştir.

Tablo 5.12 Müşteri Sınıflandırmaya Ait Çeşitli Denemelere Ait Sonuçlar

	İterasyon Sayısı	Hata Kareleri Ortalaması	
		Eğitim Seti	Test Seti
Bu Çalışma Kapsamında Geliştirilen Modelle Elde Edilen Sonuçlar	25+16.450	0,02254	0,2255
Mat Lab, Genelleştirilmiş Delta Kuralı ile Elde Edilen Sonuçlar	16.475	0,2105	0,2098

Mat-Lab ile elde edilen sonuçlar bu çalışmada, gruplamaya oranda daha iyi olmasına rağmen değerlerin birbirine yakınlığı geliştirdiğimiz modelin güvenilirliğini göstermiştir.

5.7 Model İçerisindeki Kısıtlar ve Kabuller

Model oluşturulması ve uygulama aşamalarındaki kısıtlar ve kabuller ikiye ayrılabilir:

1. Kullanılan Bilgisayar ve Programlama İle ilgili Parametrik Kısıtlar
2. Uygulama İle İlgili Kısıtlar ve Kabuller

5.7.1 Kullanılan Bilgisayar ve Programlama İle ilgili Parametrik Kısıtlar

Model, elimizde var olan yazılım ve donanım olanakları ile kullanılmıştır. Daha gelişmiş olanaklar ile, elde edilen sonuçların geliştirilebilmesi mümkündür. Kısıtların bazıları da elde var olan olanakların çerçevesinde, isteğe bağlı olarak yaratılmıştır.

Uygulama, 64Mb belleğe sahip ve Intel marka "Pentium 1" model, 200MHz. hıza sahip işlemcinin bulunduğu bir bilgisayarla gerçekleştirilmiştir. Uygulamada "Pentium 4" model işlemci ve 512Mb bellek kullanımının çok büyük işlem hızı farkları yaratacağı gözlenmiştir.

5.7.1.1 Çıktı Katman Eleman Sayısı

Çıktı katman eleman sayısı bir ile sınırlıdır yani sadece tek bağımlı değişken kullanılmıştır. Algoritmada yapılacak basit bir düzenleme ile bu kısıtın ortadan kaldırılması mümkündür ancak bu çalışma kapsamında pratik bir fayda sağlayacağı düşünülmendiğinden gerek duyulmamıştır.

5.7.1.2 Gizli Katman Sayısı

Gizli katman eleman sayısı değişken olmakla beraber gizli katman sayısı tek olarak alınmıştır. Yani modelde bir girdi katman, bir gizli katman(çoğul elemanlı) ve tekil çıktı katman oluşturulmuştur. Bu kısıt, programın çalışması için geçen sürenin çok artmasını ve problemin çok karmaşıklaşmasını

engellemek için getirilmiştir. Ancak benzer şekilde basit bir düzenleme ile bu kısıt da ortadan kaldırılabilir.

5.7.1.3 Aktivasyon Fonksiyonu

Aktivasyon fonksiyonu olarak seçilen fonksiyon, sigmoid fonksiyondur ki benzer karaktere sahip diğer fonksiyonlar da seçilebilir. Bu durumda program kodlarına müdahale edilerek fonksiyonun değiştirilmesi gerekmektedir.

5.7.2 Uygulama İle İlgili Kısıtlar ve Kabuller

Uygulama kısmı, sadece konu olan şirkete aittir, sektör içi ve veya dışı diğer şirketler için genelleştirilemez.

Modelde kullanılan veriler süzülürken bazı kayıtlar ihmal edilmiştir. Örneğin çok sayıdaki şehir seçeneği içerisinde sadece yedi büyük şehre ait kayıtlar esas alınmıştır.

Uygulama elde var olan veriler çerçevesinde yapılmıştır. Daha geniş ve tutarlı veri kümeleri ile daha da anlamlı sonuçlar elde edilebilir.

SONUÇ

Yapılan çalışmanın uygulama kısmı, genel pazarlama politikasına yön verilmesi konusuna yardımcı olmak amacıyla, karar destek sistemi olarak önerilmiştir. Böylesi çalışmaların tek başlarına işletmelere anlamlı katkılar sağlayamayacak-ları açıktır. Bu ve buna benzer çalışmalar bütünleşik bir yapı içerisinde yer aldıkları taktirde başarı şansının artacağı kesindir.

VM çalışmaları için önemli tehlikelerden birisi bir çok güçlü algoritmanın adacıklar olarak organize edilmesi ve elde edilebilecek faydaların lokal çalışmalarla sınırlandırılmasıdır. Örneğin işletmecilik açısından pazarlama faaliyetlerinin sonuçları sadece o departmana değil çoğunlukla, üretim, satın alma, finans ve benzeri bölümlere de önemli bilgiler verebilir. Benzer şekilde YSA ile elde edilen müşteri davranışı belirleme çalışmaları, Ar-Ge bölümü gibi bir çok departmana da ışık tutabilir. Üst yönetim, bütünleşik bir yapı içerisinde bilginin yatay , dikey ve çapraz akışını sağlamalı ve bilgi akışından kurumun maksimum faydayı sağlamasına yol açmalıdır.

Uygulamaya konu olan firma, bir çok örnekte de görülebileceği gibi, müşteri memnuniyetini minimum maliyetle maksimize etmeye çalışmaktadır.

Çalışmamıza konu olan uygulama geniş kapsamlı ele alının müşteri ilişkileri yönetiminin bir parçası olarak tasarlanmıştır. Örneğin, bu model her bir potansiyel müşteriye ne kadar promosyon maliyeti ile yaklaşılabileceğini belirlemeye çalışırken, benzer bir modelin de her bir müşterinin profiline eş maliyetli en uygun promosyon planının seçilmesine destek vermesi planlanmıştır.

Uygulanan modelde temel alınan algoritmaların verimliliğini arttırmak üzere bir çok teorik ve pratik çalışma mevcuttur.

Uygulamada elde edilen sonuçlar, gruplama çalışması için her 100 müşteriden yaklaşık 92'sinde o müşterinin SMS kullanıp kullanmayacağını belirlene-bildiğini göstermiştir. Benzer şekilde, her müşteriye bir ile on arasında puan veren sınıflandırma çalışmasında ise ortalama mutlak puanlama hatası bir seviyesinde belirlenmiştir. Yani model tarafından verilen bir nokta puan, ± 1 aralığında değerlendirilmelidir.

Konu üzerine çalışmak isteyen araştırmacılar için, çalışmamız kapsamı dışında kalan ve işletmecilik konusunda çalışan diğer araştırmacıların ilgilerini çekebilecek konulardan bazıları aşağıdaki sunulmuştur:

Denetimsiz öğrenme ile modele bir çıktı değişkeni sunulmadan oluşturulan YSA: Bu çalışmamamız kapsamında yer alan denetimli öğrenme modelleri, her ne kadar doğrusal olmayan ilişkilerin önceden kestirilemediği durumlara yönelik olsalar da yapısal modelin daha da belirsiz olduğu durumlarda denetimsiz öğrenme faydalı bir araç olarak kullanılabilir.

Geri yayılım algoritmalarının etkinliğini, hızını ve sonuca ulaşma performansını arttırmaya yönelik algoritmalar: Bu çalışma sırasında yapılan literatür araştırmasında, her durumda üstünlük sağlayan bir algoritmanın olmadığı saptanmış, benzer algoritmaların şartlara ve özellikle de problemin yapısına göre birbirlerine üstünlük sağlayabildikleri görülmüştür. Örneğin geri yayılım algoritmalarında kullanılan Levenberg-Marquardt modeli, sadece ilk türevlere bağlı kalmayarak aynı zamanda doğrusal olmayan optimizasyonun temel prensibi olan Heissien matris değerlerini de elde etmek yoluyla çok

hızlı öğrenmeyi sağlayabilmektedir. Ancak burada da problemin hacminin genişlemesiyle birlikte çok büyük matrislerin, o matrislere ait transpoze değerlerinin ve matrislerin testlerinin hesaplanması ve daha da önemlisi bu değerlerin sürekli olarak yenilenerek hafızada saklanması gerekmektedir.

Modelde kullanılan programlama dili temel olan Visual Basic yerine algoritmaların daha hızlı çalışmasını sağlayan dillerin kullanımı: Daha verimli diller olarak C, C++ veya Java seçilebilir. Ayrıca programlama yöntemi olarak nesne tabanlı metodolojinin seçilmesi, hem verimliliği artırırken hem de geliştirilen nesneler sayesinde gelecekte kullanılmak üzere bir nesne arşivinin yaratılmasına olanak sağlar.

Geliştirilen algoritmanın kurumsal bilgi sistemi ile bütünleşik çalışmasına yönelik öneri ve yöntem araştırmaları: VM araçları için belki de en önemli problem çok sayıdaki tekniklerin yerinde ve verimli kullanılmasıdır. Benzer şekilde, herhangi bir kuruma ait veri bankasında bulunan verilerin otomatik veya yarı otomatik araçlarla daha zeki bir şekilde işlenmesi ve faydalı bilginin rafine edilmesi ile ilgili çalışmalar, VM araçlarının kullanımını ve elde edilen faydanın artırılmasını sağlayabilir.

Verileri analiz edip bilgiye ulaşmaya çalışırken, tekrar gördüğümüz bir gerçek, hangi model veya yöntem izlenirse izlensin hala insan zekası, sezgileri ve tecrübesinin bu tür çalışmalarda önemli yeri olduğudur. Modeli eğitmek için uygun parametrelerin seçimi, çok büyük zaman almıştır ve düşünce gücü gerektirmiştir. Bu konuda, çalışma yapacak araştırmacılara önerilebilecek bir konu da, YSA eğitimi sırasında kullanılacak parametrelerin tahmini için başka yapay zeka uygulamasının kullanılması olabilir.

Bir paket program geliştirilmesi bu çalışmamız kapsamında değerlendirilmektedir. Bir yazılım geliştirmek, önemli ve aynı zamanda akademik bir değere sahip bir çalışma olabilir, ancak bu çalışma kapsamında oluşturulan modeli uygulamak amacı ile geliştirilen programlar, çalışmanın temel bölümlerinden birisi olarak değerlendirilmemelidir. Çalışmada yer alan programlar, melez modele, ihtiyaç duyduğumuz şekilde bir yazılımla karşılaşmadığı için, deneysel amaçla oluşturuldular.

Yaptığımız tüm çalışmalar, bir problemin çözümü için geçerli ve tutarlı daha basit bir yöntemin olduğu durumlarda o yöntemin uygulanmasının daha faydalı olacağı gibi pratik bir sonucun tekrar vurgulaması gereğini doğurmuştur. Örneğin doğrusal programlama ile kusursuz modellenen bir problemin GA ile çözülmesi anlamsız olduğu gibi, basit doğrusal regresyon analizini yeterli sonuçlar üretebileceği bir uygulamanın YSA ile incelenmesi de gereksizdir. VM aracı olarak kullanılan tüm karmaşık yöntemlere, geleneksel yöntemlerin sonuç vermede yetersiz kaldığı durumlarda başvurulması gerektiğini düşünmekteyiz.

Bu çalışma kapsamında geliştirilen algoritma ve program kodları, daha kullanıcılara yönelik ve parametrik hale getirilebilir. Ayrıca daha önce de belirtildiği gibi, daha optimize edilmiş bir programlama üslubu ve yöntemi ile bütünsel bir uygulama felsefesi uygulamalarda çalışmanın pratik faydalarını arttırabilir.

KAYNAKLAR

- AKPINAR, H. : "Veri Tabanlarında Bilgi Keşfi Ve Veri Madenciliği", İ.Ü. İşletme Fakültesi Dergisi, Sayı :1, Nisan 2000, s.1 - 22
- AKPINAR, H. : "Yapay Sinir Ağları ve Kredi Taleplerinin Değerlendirilmesinde Bir Uygulama Önerisi", İÜ İşletme Fakültesi, İstanbul, Mayıs 1993, [Çevrimiçi], <<http://www.isletme.istanbul.edu.tr/akpinar/content/Yapay%20Sinir%20Aglar.pdf>>, 20 Nisan 2003
- AKPINAR, H. : "Yapay Zeka ve Uzman Sistemler", İÜ İşletme Fakültesi, İstanbul, Şubat 1994
- ANONİM : "Basic Genetic AlgorithmOperations" [Çevrimiçi], <<http://www.cs.qub.ac.uk/~M.Sullivan/ga/ga4.html>>, 22 Ocak 2003.
- ANONİM : "Datawarehousing-An Approach Document", Goldstone Technologies, 2000, [Çevrimiçi] <<http://www.goldstonetech.com/downloads/Datawarehousing.pdf>>, 10 Şubat, 2003
- ANONİM : "Sinir sistemi, insan sinir sisteminin yapısı ve işleyişi", [Çevrimiçi], <http://sinancanan.tripod.com/mss.htm>>, 5 Temmuz 2002

- ANONİM : "Wal-Mart Deploys Data Mining Software Into Its Production Support Environment", 24 Mayıs 1997, [Çevrimiçi], <http://www.walmartstores.com/wmstore/wmstores/>, 15 Mayıs 2003
- ANTONY, M., : "Neural Network Learning: Theoretical
BARTLETT, P.L. Foundations", Cambridge, Cambridge University Press, 1999
- APAYDIN, E. : "ZEKİ VERİ MADENCİLİĞİ:Ham Veriden Altın Bilgiye Ulaşma Yöntemleri",2000, **Bilişim 2000 Eğitim Semineri**, İstanbul, [Çevrimiçi] http://www.cmpe.boun.edu.tr/~ethem/files/papers/veri-maden_2k-notlar.doc, 5. Mayıs 2003
- ATALAĞ K. : "E-tıp", [Çevrimiçi], <http://www.tbd.org.tr/sayi77_html/e_tip.htm>,10 Ocak 2003
- BADDELEY, R., : "Information Theory and the Brain", England,
HANCOCK, P., Cambridge University Press, 2000
FOLDIAK P.
- BARAY, Ş. A., : "Yöneylem Araştırması, Hamdy, A.Taha, Altıncı
ESNAF, Ş. Basımdan Çeviri", İstanbul, Eylül 2000
- BERRY, M. A., : "Mastering Data Mining: The Art And Science Of
LINOFF, G. L. Customer Relationship Management", New York, NY, John Willey & Sons, Inc., 2000

- BERRY, M. A., : **"Data Mining Techniques For Marketing, Sales, and**
LINOFF, G. L. **Customer Support"**, New York, NY: John Willey &
Sons, Inc., 1997
- BERSON, A., : **"Building Data Mining Applications for CRM"**, New
SMITH, S. York, NY: McGraw- Hill., 1999
- BHARATH, R., : **"Neural Network Computing"**, New York, NY:
DROSEN J. McGraw-Hill., 1994
- BIGUS J. P. : **"Data Mining With Neural Networks"**, New York, NY:
McGraw- Hill., 1996
- BOSE, N. K., : **"Neural Network Fundamentals With Graphs,**
LIANG, P. **Algorithms, And Applications"**, New York, NY:
McGraw- Hill., 1996
- BRYMAN, A., : **"Quantitative Data Analysis with SPSS for Windows**
CRAMER, D. **Guide for Social Sciences"**, Routledge, London And
New York, 1997
- CABENA, P., : **"Discovering Data Mining From Concept to**
HADJINIAN, P., **Implementation"**, New Jersey, NY: Prentice Hall.,
STADLER, R. 1997
- CHAUDHURI, S., : **"An Overview of Data Warehousing and OLAP**
DAYAL, U. **Technology"**, **ACM Sigmod Record**, Mart 1997.

- ÇAĞIL, G. : "Mevsimlik Olmayan Box-Jenkins Modellerinde İki Aşamalı Yapay Sinir Ağlarının Kullanılması ve Bir Uygulama", Doktora Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, 1997
- DARLINGTON, R.B. : Factor Analysis", [Çevrimiçi],
"http://comp9.psych.cornell.edu/Darlington/factor.htm,
15.Mayıs.2003.
- DECO, G., OBRADOVIC, D. : "An Information-Theoric Approach to Neural Computing", New York, NY, Springer-Verlag New York, Inc., 1996
- DE WILDE, P. : "Neural Network Models An Analysis", New York, NY: Springer., 1995
- EVLİCE, Y.E. : "Davranışın Nörobiyolojik Temelleri", [Çevrimiçi],
<http://lokman.cu.edu.tr/psychiatry/DERSNOT/1/1davrnbn.htm>, 2 Şubat 2003.
- FINE, T. L. : "Feedforward Neural Network Methodology", New York, NY, Springer-Verlag New York, Inc., 1999
- GAY, L.R., DIEHL P.L. : "Research Methods for Business and Management", Maxwell McMillan International Edition, McMillan 1992
- GOONATILAKE S., TRELEAVEN, P. : "Intelligent Systems for Finance and Business", New York, NY: John Willey & Sons, Inc., 1995

- GROSSBERG, S. : **" Neural Networks and Natural Intellience",**
Massachusetts, The MIT Press, 1988
- GROTH, R. : **"Data Mining: Building Competitive Advantage",**
New Jersey, NY: Prentice Hall., 2000
- GUYTON, A. C. : **"Fizyoloji Cilt 2",** İngilizce 5. Baskıdan Türkçeye
Çevrilmiş 1. Baskı, Güven Kitabevi Yayınları, Ankara,
1978
- HARVEY, R. L. : **" Neural Network Principles",** New Jersey, NY,
Prentice Hall., 1994
- HOOFER, J.A., : **"Modern System Analysis and Design",** İkinci Baskı,
GEORGE, J.F., New York, Addison Wesley Longman Inc., 1999
VALACICH, J.S.
- HOSSOUN, M. H. : **"Fundamentals Of Artificial Neural Networks",**
Massachusetts, The MIT Press, 1995
- HOU, C. : **"Digital Simulation of Artificial Neural Networks for
Business Applications",** Yüksek Lisans Tezi, San Diego
State University, San Diego, 1989
- İDİL, O. : **"Örnekleme Teorisi ve İşletme Yönetiminde
Uygulanması",** İÜ Yayın No: 2708, İşletme Fakültesi
Yayın No: 107, Fatih Yayınevi Matbaası, İstanbul 1980
- JAIN, L.C., : **"Fusion of Neural Networks, Fuzzy Sets, and
MARTIN, N. M. Genetic Algorithms",** New York, CRC Press, 1999

- KOCH, C., : **"Methods in Neural Modeling"**, Massachusetts, The
SEGEV, I. MIT Press, 1998
- KOSTER, A. : "Supplementary Material For IDS 480/686 ,Database
Management Systems", Montezuma Publishing, San
Diego Ca, 2002
- KYU, L. J. , : "A Two-Stage Neural Network Approach For ARMA
CHUL, W. Model Identification With ESACF.", **Decision Support
Systems**, Amsterdam, Haziran 1994, 11(5) sf:461
- LEE, T. H., : "Analyzes Logistic Regression Forecasts Of
JUNG, S. C. Creditworthiness.", **Journal of Business Forecasting
Methods & Systems**, Winter99/2000, 18(4), sf:28-31.
- LIN, T. C., : **"Neural Fuzzy Systems, A Neuro-Fuzzy Synergism
LEE, C.S.G. to Intelligent Systems"**, New Jersey, NY: Prentice
Hall, 1996
- MANNINO, M. V. : **"Database Application Development & Design"**,
New York, NY: McGraw- Hill, 2001
- MICHALSKI, R. S., : **"Machine Learning And Data Mining Methodologies
BRATKO, I. And Applications"**, New York, NY: John Willey &
Sons, Inc., 1998
- MOORE, A., : "Discussion the Questions to Ask When Performing a
WELLS, D High-level Post-Implementation Data Warehouse
Audit", **The Data Administration Newsletter**, Haziran
2000

- MOSHIRI, S. : "Neural Network Versus Econometric Models In Forecasting Inflation.", **Journal of Forecasting**, Nisan 2000, 19(3), sf: 201
- NEWBOLD, P. : "**Statistics For Business And Economics**", New Jersey, NY: Prentice Hall., 1995
- NILSSON, N. J. : "**Artificial Intelligence A New Synthesis**", San Francisco, CA, Morgan Kaukmann, 1998
- NOLAN, R.L. : "Managing the Crises in Data Processing", **Harvard Business Review**, California, Mart - Nisan 1979
- ORHUNBİLGE, N. : "**Örnekleme Yöntemleri ve Hipotez Testleri**", İÜ İşletme Fakültesi, Avcıol Basım Yayın, İstanbul 1999
- ORHUNBİLGE, N. : "**Tanımsal İstatistik, Olasılık ve Olasılık Dağılımları**", İÜ İşletme Fakültesi, Avcıol Basım Yayın, İstanbul 2000
- ORHUNBİLGE, N. : "**Uygulamalı Regresyon ve Korelasyon Analizi**", İÜ İşletme Fakültesi, Avcıol Basım Yayın, İstanbul 1996
- ORR, G.,
SCHRAUDOLPH,
N., CUMMINS, F. : "Computation in The Brain, The Brain as an Information Processing System", [Çevrimiçi], <<http://www.willamette.edu/~gorr/classes/cs449/brain.html>>, 6 şubat 2003
- PARKS, R. W.,
LEVINE, D. S.,
LONG, D. : "**Fundamentals of Neural Network Modeling**", Massachusetts, The MIT Press, 1998

- PENDSE, N. : "The OLAP Report", **Business Intelligence Ltd. Group**, 30 Mart 2003, [Çevrimiçi],
<<http://www.olapreport.com/Market.htm>>, 15 Nisan 2003
- POHLHEIM, H. : "Genetic and Evolutionary Algorithm Toolbox for use with *Matlab* (GEATbx)", [Çevrimiçi],
<<http://www.geatbx.com/docu/algrecom.html#namerecombinationbinary>>, 22 Ocak 2003
- ROGERS, J. : "**Object Oriented Neural Networks in C++**", San Diego, Academic Press Inc., 1997
- STERGIOU C. , : "Neural Networks", [Çevrimiçi],
SIGANOS D. <http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html>, 11 Ağustos 2002
- TATLIDİL, H. : "**Uygulamalı Çok değişkenli İstatistiksel Analiz**", Hacettepe Üniversitesi, Cem Web Ofset, Ankara 1996
- TULUNAY, Y. : "**Matematik Programlama ve İşletme Uygulamaları**", İÜ İşletme Fakültesi, Üçüncü Baskı, İstanbul 1988
- WONG, B. K., : "Neural Network Applications in Business: A review
BODNOVICH T. A., and Analysis of the Literature (1988-95)", **Decision
SELVİ, Y. Support Systems**, 1997, s. 301-320

- YÜREKLİ, H. : "Bulanık Uzman Sistemler", Yüksek Lisans Tezi,
İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, 1999
- ZHONG, T., : "Determinants of Consumer's Debt Repayment
DEVENEY, S. A. Patterns.", **Consumer Interests Annual**, 1999,45(5),
s. 65-71



EK 1. GENETİK ALGORİTMA PROGRAMI

(Bu program Tarafımızdan Geliştirilmiştir.)

Dim iterasyon, gensayısı, bitsayısı, degisken_sayısı, yenilei, yenilej As
Integer

Dim x(200, 64), xyeni(200, 64), xcross(200, 64), deneme(2000, 64)

Dim onlu(200, 10)

Dim deger(200) As Double

Dim uygunluk(200), kumulatif(200), secme_sayısı(200)

Dim anano(200)

Dim iterasuygun(10000), ortiterasuygun(10000)

Dim maksigen, minigen, maksix, maksiy As Double

Dim artık(200), kümülatifartık(200)

Sub Genetik_Algoritma()

'Makro Arda Meric tarafından 06.01.2003 tarihinde kaydedildi.

gensayısı = 100 *'parametreler tanımlanıyor*

bitsayısı = 48

degisken_sayısı = 2

maksiterasyon = 200

BASLANGIC_GENERASYON

For iterasyon = 1 To maksiterasyon

 MUTASYON

 FONKSIYON_DEGERLERI

 YENI_GENERASYON_SECME

 CROSSOVER

Next iterasyon

```

For gi = 1 To gensayısı
    For gj = 1 To bitsayısı
        Cells(gi, gj) = x(gi, gj)
    Next gj
Next gi
FONKSIYON_DEGERLERI
For sonuclar = 1 To gensayısı
For k = 1 To degisken_sayısı
Cells(sonuclar, bitsayısı + k) = onlu(sonuclar, k) 'x,y değerlerini onluk sayma
düzeninde yazdırır.

Cells(sonuclar, bitsayısı + 3) = deger(sonuclar) 'fonksiyon değerini yazdırır.
Next k
Next sonuclar
End Sub

```

```

Function BASLANGIC_GENERASYON()
For i = 1 To gensayısı 'rasgele popülasyon oluşturuluyor.
    For j = 1 To bitsayısı
        x(i, j) = Int((1 - 0 + 1) * Rnd + 0)
    Next j
Next i
End Function

```

```

Function FONKSIYON_DEGERLERI()
maksxy = 4
Top = maksxy = 0
mini = 9999999999
For i = 1 To gensayısı
    For k = 1 To degisken_sayısı
        For j = 1 To bitsayısı / degisken_sayısı
            onlu(i, k) = onlu(i, k) + x(i, j + (k - 1) * (bitsayısı / degisken_sayısı)) * (2 ^
                (bitsayısı / degisken_sayısı - j))
        Next j
        onlu(i, k) = (onlu(i, k) * (maksxy / ((2 ^ (bitsayısı / degisken_sayısı) - 1))))
    Next k
Next i

```

'Skala değışimi

```
Next k
üst = 2 * (Sin(onlu(i, 1) ^ 2 + onlu(i, 2))) ^ 2 ' Fonksyon Tanımlanıyor.
alt = (onlu(i, 1) ^ 2 + onlu(i, 2) ^ 2) ^ (1 / 2) + 0.1
deger(i) = üst / alt
Top = Top + deger(i)
If maksi < deger(i) Then
    maksi = deger(i)
    maksix = onlu(i, 1)
    maksiy = onlu(i, 2)
    maksigen = i
End If
If deger(i) < mini Then
    mini = deger(i)
    minigen = i
End If
Next i
iterasuygun(iterasyon) = Top
ortiterasuygun(iterasyon) = Top / gensayısı
Cells(iterasyon, bitsayısı + 7) = maksi
Cells(iterasyon, bitsayısı + 8) = maksix
Cells(iterasyon, bitsayısı + 9) = maksiy
Cells(iterasyon, bitsayısı + 6) = ortiterasuygun(iterasyon)
uygunluk(1) = deger(1) / Top
kumulatif(1) = uygunluk(1)
For k = 2 To gensayısı
    uygunluk(k) = deger(k) / Top
    kumulatif(k) = kumulatif(k - 1) + uygunluk(k)
Next k
End Function
```

```
Function YENI_GENERASYON_SECME()
For eniyi = 1 To bitsayısı
```

```

    x(minigen, eniyi) = x(maksigen, eniyi)
Next eniyi
For resetle = 1 To gensayısı
    secme_sayısı(resetle) = 0
Next resetle
topsecme = 0
For j = 1 To gensayısı
    secme_sayısı(j) = Int(uygunluk(j) * 100)
    artık(j) = (uygunluk(j) * 100) - Int(uygunluk(j) * 100)
    topsecme = topsecme + secme_sayısı(j)
    If j = 1 Then
        kümülatifartık(j) = artık(j)
    Else
        kümülatifartık(j) = kümülatifartık(j - 1) + artık(j)
    End If
Next j
kalanlar = gensayısı - topsecme
For tarama = 1 To kalanlar
    rasgele = kalanlar * Rnd()
    For i = 1 To gensayısı
        If kümülatifartık(i) > rasgele Then
            secme_sayısı(i) = secme_sayısı(i) + 1
            GoTo 10
        End If
    Next i
10 Next tarama
r = 1
For i = 1 To gensayısı
    If secme_sayısı(i) <> 0 Then
        For j = 1 To secme_sayısı(i)
            For strno = 1 To bitsayısı

```

```

        xyeni(r, strno) = x(i, strno)
    Next strno
    r = r + 1
Next j
End If
Next i
End Function

```

```

Function CROSSOVER()
pc = 0.95
eslesme = 1
Do While eslesme <= gensayısı
    anano(eslesme) = Int((gensayısı - 1 + 1) * Rnd + 1)
    flag = 0
    If eslesme > 1 Then
        For tekil = 1 To (eslesme - 1)
            If anano(eslesme) = anano(tekil) Then flag = 1
        Next tekil
    End If
    If flag = 0 Then
        eslesme = eslesme + 1
    End If
Loop
For i = 1 To gensayısı Step 2
    bolme = Int((bitsayısı - 1 - 1 + 1) * Rnd + 1)
    p = Rnd
    For gecis1 = 1 To bolme
        If p > 1 - pc Then
            xcross(i, gecis1) = xyeni(anano(i + 1), gecis1)
            xcross((i + 1), gecis1) = xyeni(anano(i), gecis1)
        Else
            xcross(i, gecis1) = xyeni(anano(i), gecis1)
        End If
    Next gecis1
Next i
End Function

```

```

        xcross((i + 1), gecis1) = xyeni(anano(i + 1), gecis1)
    End If
Next gecis1
For gecis2 = bolme + 1 To bitsayısı
    xcross(i, gecis2) = xyeni(anano(i), gecis2)
    xcross((i + 1), gecis2) = xyeni(anano(i + 1), gecis2)
Next gecis2
For yenilej = 1 To bitsayısı
    x(i, yenilej) = xcross(i, yenilej)
    x(i + 1, yenilej) = xcross(i + 1, yenilej)
Next yenilej
Next i
End Function

```

```

Function MUTASYON()
pm = 0.001
For mut1 = 1 To gensayısı
    For mut2 = 1 To bitsayısı
        mut = Rnd
        If mut > (1 - pm) Then
            If x(mut1, mut2) = 0 Then
                x(mut1, mut2) = 1
            Else
                x(mut1, mut2) = 0
            End If
        End If
    Next mut2
Next mut1
End Function

```

EK 2. GA YSA MELEZ MODELİNE AİT PROGRAM

(Bu program Tarafımızdan Geliştirilmiştir.)

Dim x(50, 400) As Single

Dim yhid(65), youtp As Single

Dim mi(50) As Single

Dim ma(50) As Single

Dim sh1 As Single

Dim onlu(101, 1000) As Single

Dim input_num, Hidden, total_examples_egitim, degisken_sayısı, bit,
Bitsayısı, gensayısı As Integer

Dim xb(101, 11056), xiyi(101, 11056), xyeni(101, 11056), xcross(101, 11056)
As Byte

Dim uygunluk(101) As Double

Dim maksigen, minigen, minhata As Single

Sub Egitim()

' Makro1 Makro

' Makro arda tarafından 12.02.2003 tarihinde kaydedildi.

Dim w(1000) As Single

Dim deltaold(1000), delta As Single

Dim it As Long

input_num = 12 'parametre tanımı

Hidden = 42 'parametre tanımı

```

total_examples_egitim = 124 'parametre tanımı
Total_test = 12 'parametre tanımı
Dim m, yem As Integer
For m = 1 To input_num + 1 ' max ve min sutun degerlerini hesaplar
    mi(m) = Sheets("egitim").Cells(2, m)
    ma(m) = 0
    For yem = 2 To total_examples_egitim + Total_test + 1

        If Sheets("egitim").Cells(yem, m) > ma(m) Then ma(m) =
            Sheets("egitim").Cells(yem, m)

        If Sheets("egitim").Cells(yem, m) < mi(m) Then mi(m) =
            Sheets("egitim").Cells(yem, m)

    Next yem
Next m
Dim i As Integer
For i = 1 To input_num
    For layer0 = 2 To total_examples_egitim + Total_test + 1
        x(i, layer0) = (0.9 * (Sheets("egitim").Cells(layer0, i) - mi(i)) / (ma(i) - mi(i))) +
            0.1
    Next layer0
Next i
For layer0 = 2 To total_examples_egitim + Total_test + 1
    x(input_num + 1, layer0) = Sheets("egitim").Cells(layer0, input_num + 1)
Next layer0
degisken_sayısı = (input_num + 1) * Hidden + Hidden + 1
gensayısı = 100 'parametre tanımı
bit = 10 'parametre tanımı
Bitsayısı = bit * degisken_sayısı
maksiterasyon = 25 'parametre tanımı

```

```

Dim tot, whid, winp, hid, inp As Integer
For tot = 2 To total_examples_egitim + Total_test + 1
    x(0, tot) = 1
Next tot
BASLANGIC_GENERASYON
For iterasyon = 1 To maksiterasyon
    MUTASYON
    FONKSIYON_DEGERLERI
    YENI_GENERASYON_SECME
    CROSSOVER
    Sheets("egitim").Cells(iterasyon, input_num + 4) = minhata
Next iterasyon
FONKSIYON_DEGERLERI
    Sheets("egitim").Cells(iterasyon, input_num + 4) = minhata
For sil = 0 To degisken_sayisi
    deltaold(sil) = 0
Next sil
learn = 0.01 'parametre tanımı
For atama = 1 To degisken_sayisi
    w(atama) = onlu(maksigen, atama)
Next atama
maksit = 20000 'parametre tanımı
For it = 1 To maksit
    For tot = 2 To total_examples_egitim + 1
        flag = 1
        Difx = 0
5        For hid = 1 To Hidden
            sh1 = 0
            For inp = 0 To input_num
                ifadebasit = w((inp + 1) + (hid - 1) * (input_num + 1))
                sh1 = sh1 + x(inp, tot) * ifadebasit
            
```

```

        Next inp
        yhid(hid) = 1 / (1 + Exp(-sh1))
    Next hid
    yhid(0) = 1
    sh2 = 0
    For hid = 0 To Hidden
        sh2 = sh2 + yhid(hid) * w((input_num + 1) * Hidden + (hid + 1))
    Next hid
    youtp = 1 / (1 + Exp(-sh2))
    dif = youtp - (x(input_num + 1, tot))
    For düzel = degisken_sayısı To 1 Step -1
        If düzel > Hidden * (input_num + 1) Then

            delta = learn * dif * youtp * (1 - youtp) * yhid(düzel - 1 - Hidden * (input_num +
            1)) + 0.9 * deltaold(düzel)

            w(düzel) = w(düzel) - delta
            deltaold(düzel) = -delta
        Else
            tam = Int((düzel - 1) / (input_num + 1)) + 1

            delta = learn * dif * youtp * (1 - youtp) * yhid(tam) * w((input_num + 1) *
            Hidden + (tam + 1)) * (1 - yhid(tam)) * x(düzel - 1 - (tam - 1) * (input_num +
            1), tot) + 0.9 * deltaold(düzel)

            w(düzel) = w(düzel) - delta
            deltaold(düzel) = -delta
        End If
    Next düzel
    If Abs(dif) > 0.15 And it < 50 Then 'parametre tanımı
    flag = flag + 1

```

```

Cells(1, 19) = Abs(dif)
Cells(3, 19) = tot
If Difx <> 0 Then
    If (Difx - dif) / Difx < 0.00001 Then flag = 100
End If
Difx = dif
If flag < 100 Then GoTo 5
End If
Cells(2, 19) = it
10 Next tot
    If (it / 50) = Int(it / 50) Then
        totsqdifg = 0
        totsqdifgtest = 0
        For tot = 2 To total_examples_egitim + Total_test + 1
            For hid = 1 To Hidden
                sh1 = 0
                For inp = 0 To input_num
                    ifade_basitle = w((inp + 1) + (hid - 1) * (input_num + 1))
                    sh1 = sh1 + x(inp, tot) * ifade_basitle
                Next inp
                yhid(hid) = 1 / (1 + Exp(-sh1))
            Next hid
            yhid(0) = 1
            sh2 = 0
            For hid = 0 To Hidden
                sh2 = sh2 + yhid(hid) * w((input_num + 1) * Hidden + (hid + 1))
            Next hid
            youtp = 1 / (1 + Exp(-sh2))
            Cells(tot, input_num + 2) = youtp
            If tot <= total_examples_egitim + 1 Then

```

```

    difg = youtp - Sheets("egitim").Cells(tot, input_num + 1)
    totsqudifg = totsqudifg + (difg ^ 2)
Else
    difg = youtp - Sheets("egitim").Cells(tot, input_num + 1)
    totsqudifgtest = totsqudifgtest + (difg ^ 2)
End If
Next tot
Sheets("egitim").Cells(it / 50, input_num + 5) = totsqudifg /
total_examples_egitim
Sheets("egitim").Cells(it / 50, input_num + 6) = totsqudifgtest / Total_test
For hid = 1 To Hidden
    For inp = 0 To input_num
        Sheets("cozum").Cells(hid, inp + 1) = w((inp + 1) + (hid - 1) * (input_num +
1))
    Next inp
Next hid
For hid = 0 To Hidden

    Sheets("cozum").Cells(Hidden + 2, hid + 1) = w((input_num + 1) *
Hidden + (hid + 1))

Next hid

12 If (totsqudifgtest / Total_test - totsqudifg / total_examples_egitim) /
(totsqudifgtest / Total_test) > 0.1 Then GoTo 100

End If
Next it
100 End Sub

```

```

Function BASLANGIC_GENERASYON()
For i = 1 To gensayısı 'rasgele popülasyon oluşturuluyor.

```

```

    For j = 1 To Bitsayısı
        xb(i, j) = Int((1 - 0 + 1) * Rnd + 0)
    Next j
Next i
End Function
Function FONKSIYON_DEGERLERI()
    Dim deger(101) As Double
    maksxy = 2
    minixy = -2
    Top = 0
    maks = 0
    mini = 999999999999999#
    For i = 1 To gensayısı
        For k = 1 To degisken_sayısı
            onlu(i, k) = 0
            For j = 1 To bit
                onlu(i, k) = onlu(i, k) + xb(i, j + (k - 1) * (bit)) * (2 ^ (bit - j))
            Next j
            onlu(i, k) = onlu(i, k) * (maksxy - minixy) / (2 ^ bit - 1) + minixy 'Skala
            değışimi
        Next k
        yhid(0) = 1
        totsqrif = 0
        Totsqriftest = 0
        For tot = 2 To total_examples_egitim - 49
            For hid = 1 To Hidden
                sh1 = 0
                For inp = 0 To input_num
                    ifade = onlu(i, (inp + 1) + (hid - 1) * (input_num + 1))
                    sh1 = sh1 + x(inp, tot) * ifade
                Next inp
            Next hid
        Next tot
    Next i
End Function

```

```

        yhid(hid) = 1 / (1 + Exp(-sh1))
    Next hid
    sh2 = 0
    For hid = 0 To Hidden
        sh2 = sh2 + yhid(hid) * onlu(i, (input_num + 1) * Hidden + (hid + 1))
    Next hid
    youtp = 1 / (1 + Exp(-sh2))
    dif = youtp - Sheets("egitim").Cells(tot, input_num + 1)
    totsqdif = totsqdif + (dif ^ 2)
Next tot
For tot = (total_examples_egitim - 48) To total_examples_egitim + 1
    For hid = 1 To Hidden
        sh1 = 0
        For inp = 0 To input_num
            ifade = onlu(i, (inp + 1) + (hid - 1) * (input_num + 1))
            sh1 = sh1 + x(inp, tot) * ifade
        Next inp
        yhid(hid) = 1 / (1 + Exp(-sh1))
    Next hid
    sh2 = 0
    For hid = 0 To Hidden
        sh2 = sh2 + yhid(hid) * onlu(i, (input_num + 1) * Hidden + (hid + 1))
    Next hid
    youtp = 1 / (1 + Exp(-sh2))
    dif = youtp - Sheets("egitim").Cells(tot, input_num + 1)
    Totsqdiftest = Totsqdiftest + (dif ^ 2)
Next tot

13 deger(i) = 1 / ((Totsqdiftest + totsqdif) / total_examples_egitim) +
Abs(Totsqdiftest / 50 - totsqdif / (total_examples_egitim - 50))

```

```

Top = Top + deger(i)
If maksı < deger(i) Then
    maksı = deger(i)
    minhata = (Totsqdiftest + totsqdif) / total_examples_egitim
    maksigen = i
End If
If deger(i) < mini Then
    mini = deger(i)
    minigen = i
End If
Next i
For k = 1 To gensayısı
    uygunluk(k) = deger(k) / Top
Next k
End Function

```

```

Function YENI_GENERASYON_SECME()
Dim artık(101), kümülatifartık(101) As Single
Dim secme_sayısı(101) As Integer
For eniyi = 1 To Bitsayısı
    xiyi(minigen, eniyi) = xb(maksigen, eniyi)
    xb(minigen, eniyi) = xb(maksigen, eniyi)
Next eniyi
For resetle = 1 To gensayısı
    secme_sayısı(resetle) = 0
Next resetle
topsecme = 0
For j = 1 To gensayısı
    secme_sayısı(j) = Int((uygunluk(j) * 100) * (gensayısı / 100))
    artık(j) = ((uygunluk(j) * 100)) - Int((uygunluk(j) * 100))
    topsecme = topsecme + secme_sayısı(j)
    If j = 1 Then

```

```

        kümülatifartık(j) = artık(j)
    Else
        kümülatifartık(j) = kümülatifartık(j - 1) + artık(j)
    End If
Next j
kalanlar = gensayısı - topsecme
For tarama = 1 To kalanlar
    rasgele = kalanlar * Rnd()
    For i = 1 To gensayısı
        If kümülatifartık(i) > rasgele Then
            secme_sayısı(i) = secme_sayısı(i) + 1
            GoTo 10
        End If
    Next i
10 Next tarama
r = 1
For i = 1 To gensayısı
    If secme_sayısı(i) <> 0 Then
        For j = 1 To secme_sayısı(i)
            For strno = 1 To Bitsayısı
                xyeni(r, strno) = xb(i, strno)
            Next strno
            r = r + 1
        Next j
    End If
Next i
End Function

```

Function CROSSOVER()

Dim anano(101) As Integer

pc = 0.85

eslesme = 1

```

Do While eslesme <= gensayısı
    anano(eslesme) = Int((gensayısı - 1 + 1) * Rnd + 1)
    flag = 0
    If eslesme > 1 Then
        For tekil = 1 To (eslesme - 1)
            If anano(eslesme) = anano(tekil) Then flag = 1
        Next tekil
    End If
    If flag = 0 Then
        eslesme = eslesme + 1
    End If
Loop
For i = 1 To gensayısı Step 2
    bolme = Int((Bitsayısı - 1 - 1 + 1) * Rnd + 1)
    p = Rnd
    For gecis1 = 1 To bolme
        If p > 1 - pc Then
            xcross(i, gecis1) = xyeni(anano(i + 1), gecis1)
            xcross((i + 1), gecis1) = xyeni(anano(i), gecis1)
        Else
            xcross(i, gecis1) = xyeni(anano(i), gecis1)
            xcross((i + 1), gecis1) = xyeni(anano(i + 1), gecis1)
        End If
    Next gecis1
    For gecis2 = bolme + 1 To Bitsayısı
        xcross(i, gecis2) = xyeni(anano(i), gecis2)
        xcross((i + 1), gecis2) = xyeni(anano(i + 1), gecis2)
    Next gecis2
    For yenilej = 1 To Bitsayısı
        xb(i, yenilej) = xcross(i, yenilej)
        xb(i + 1, yenilej) = xcross(i + 1, yenilej)
    
```

```
Next yenilej
Next i

For eniyi = 1 To Bitsayısı
    xb(minigen, eniyi) = xiyi(minigen, eniyi)
Next eniyi
End Function
```

```
Function MUTASYON()
    pm = 0.0005
    For mut1 = 1 To gensayısı
        For mut2 = 1 To Bitsayısı
            mut = Rnd
            If mut > (1 - pm) Then
                If xb(mut1, mut2) = 0 Then
                    xb(mut1, mut2) = 1
                Else
                    xb(mut1, mut2) = 0
                End If
            End If
        Next mut2
    Next mut1
End Function
```

EK 3. MÜŞTERİ GRUPLAMA EĞİTİM VE TEST VERİ SETİ

Alan 1: Dağıtıcı

Alan 2: Ödeme Planı

Alan 3: Medeni Durum

Alan 4: Promosyon (email almak ister mi?)

Alan 5: E-bilgilendirme (maili almak ister mi?)

Alan 6: SMS almak isityor mu?

Alan 7: Şehir

Alan 8: Meslek

Alan 9: Cinsiyet

Alan 10: Musteri Adedi

Alan 11: Kullanıcı Sayısı

Alan 12: Yaş Grubu

Alan 13: Sms Kullanımı

Alan 14: Tahmini Sms Kullanımı

1	2	3	4	5	6	7	8	9	10	11	12	13	14
31	7	1	1	1	0	6	30	0	1	1	3	0,9999	0,55
45	11	0	1	1	0	6	1	1	1	1	3	0,0001	0,02
15	3	1	1	1	0	6	6	0	1	1	4	0,9999	0,99
27	3	0	1	1	1	7	6	1	2	1	1	0,9999	1,00
39	6	0	1	1	0	6	6	1	2	1	3	0,0001	0,22
9	7	0	1	1	0	6	6	0	2	2	4	0,0001	0,03
55	3	0	1	1	0	6	21	0	1	1	3	0,0001	0,54
36	3	0	1	1	0	6	7	0	1	1	3	0,9999	0,67
46	3	0	1	1	0	6	24	0	1	1	3	0,9999	0,60
48	3	0	1	1	0	6	13	0	1	1	4	0,0001	0,62
18	6	0	1	1	0	4	30	0	1	1	4	0,0001	0,04
42	3	0	1	1	0	3	27	0	2	1	3	0,9999	0,90
55	6	0	1	1	0	6	1	1	2	1	2	0,0001	0,37
18	6	0	1	1	0	5	3	0	1	1	2	0,0001	0,72

(Bir Önceki Sayfadan Devam)													
12	11	0	0	0	0	6	31	0	1	1	3	0,0001	0,22
18	6	1	1	1	0	6	7	0	1	1	3	0,9999	0,98
2	11	0	1	1	0	2	22	0	2	1	3	0,0001	0,01
18	6	1	1	1	0	7	6	1	1	1	4	0,0001	0,23
27	6	0	1	1	1	5	14	0	1	1	3	0,0001	0,57
16	3	0	1	1	1	6	2	0	1	1	3	0,9999	0,94
55	6	0	1	1	0	6	13	1	1	1	3	0,9999	0,26
30	9	1	1	1	0	6	31	0	3	1	3	0,9999	0,99
53	6	0	1	1	1	2	1	0	1	1	3	0,0001	0,19
33	2	0	1	1	1	5	30	1	1	1	2	0,9999	0,99
16	3	0	1	1	0	7	14	0	3	2	5	0,0001	0,10
15	1	0	1	1	0	2	6	0	1	1	3	0,9999	0,93
17	11	1	1	1	0	6	10	1	1	1	3	0,0001	0,22
17	11	0	1	1	0	6	3	0	1	1	2	0,0001	0,11
30	6	0	1	0	0	6	6	0	1	1	3	0,0001	0,02
56	5	0	1	1	0	6	1	0	2	1	5	0,0001	0,52
54	7	1	1	1	1	5	6	0	1	1	3	0,9999	0,96
6	3	1	1	1	0	6	27	0	1	1	3	0,9999	0,97
39	6	0	1	1	0	5	1	0	1	1	4	0,0001	0,40
53	6	0	1	1	1	6	16	0	1	1	3	0,0001	0,34
43	11	0	1	1	0	5	13	0	1	1	4	0,0001	0,13
4	7	0	1	1	0	6	22	0	1	1	3	0,9999	0,58
55	3	0	1	1	1	4	23	0	1	1	2	0,9999	0,79
42	7	1	1	1	0	5	6	0	1	1	4	0,9999	0,71
38	3	3	1	1	0	6	6	0	3	1	4	0,9999	1,00
6	3	1	1	1	0	1	27	0	1	1	3	0,9999	0,81
27	6	1	1	1	0	7	6	0	1	1	4	0,9999	0,97
40	2	0	1	1	0	5	6	0	1	1	3	0,9999	0,54
26	9	1	1	1	1	6	25	0	1	1	4	0,9999	0,65
23	6	1	1	1	0	5	2	0	1	1	3	0,9999	0,84
27	6	0	1	1	0	6	25	1	2	1	4	0,0001	0,04
16	3	0	1	1	1	5	2	0	1	1	3	0,9999	0,89
1	6	1	1	1	0	6	2	0	1	1	3	0,9999	1,00
31	6	0	1	1	0	6	25	0	2	1	3	0,9999	0,88
34	11	0	1	1	0	5	22	0	1	1	3	0,0001	0,19
55	6	0	1	1	0	5	22	0	1	1	4	0,0001	0,41
10	6	1	1	1	1	5	27	0	1	1	3	0,0001	0,23
11	3	0	1	1	0	6	28	0	1	1	3	0,9999	0,57
57	8	0	1	1	0	5	6	0	1	1	3	0,0001	0,28
18	6	0	1	1	0	2	1	0	2	2	5	0,0001	0,02
19	4	1	1	1	0	5	6	1	2	1	3	0,9999	0,61
27	6	1	1	1	1	6	7	0	1	1	4	0,9999	0,93
55	6	0	1	1	0	3	6	0	2	1	3	0,0001	0,24
17	1	1	1	1	1	6	22	0	1	1	3	0,9999	0,80
52	3	0	1	1	1	5	6	1	1	1	2	0,9999	0,98
12	11	0	1	1	0	6	4	0	1	1	2	0,0001	0,14
15	10	2	1	1	0	5	6	1	1	1	4	0,9999	0,86

(Bir Önceki Sayfadan Devam)													
46	3	0	1	1	1	7	25	1	1	1	4	0,9999	0,96
16	7	1	1	1	0	2	22	0	1	1	4	0,0001	0,19
17	5	1	1	1	0	5	13	0	3	1	3	0,9999	0,99
52	6	0	1	1	0	6	2	0	1	1	4	0,9999	0,69
21	6	0	1	1	1	4	23	1	2	1	3	0,0001	0,17
35	11	0	0	0	0	2	23	1	1	1	3	0,0001	0,16
51	3	0	1	1	0	5	6	0	1	1	3	0,9999	0,51
40	10	0	1	1	0	2	31	0	3	1	3	0,9999	0,62
56	6	0	1	1	0	6	30	0	1	1	3	0,0001	0,23
44	5	0	1	1	1	5	2	0	1	1	4	0,9999	0,91
56	6	0	1	1	0	6	6	0	1	1	3	0,9999	0,73
17	10	0	1	1	0	5	23	0	2	1	2	0,9999	0,57
32	6	0	1	1	0	6	6	1	1	1	3	0,0001	0,05
55	2	0	1	1	1	3	31	0	1	1	3	0,9999	0,91
44	11	0	1	1	0	6	22	0	1	1	3	0,0001	0,24
47	10	0	0	1	0	6	4	0	2	1	2	0,9999	0,99
15	7	0	1	1	0	7	6	0	1	1	3	0,9999	0,94
11	3	1	1	1	0	5	6	0	1	1	3	0,9999	0,99
16	6	0	1	1	0	6	22	1	1	1	4	0,0001	0,14
15	11	0	1	1	0	2	23	0	1	1	2	0,0001	0,13
11	6	0	1	1	0	6	6	0	2	1	3	0,9999	0,89
15	3	0	0	0	0	6	23	0	1	1	2	0,9999	0,62
11	9	0	1	1	0	6	28	0	1	1	4	0,9999	0,11
26	3	0	1	1	0	2	30	0	1	1	3	0,0001	0,28
30	9	1	1	1	1	6	1	1	2	1	3	0,9999	0,89
27	6	0	1	1	0	3	1	0	1	1	3	0,0001	0,10
27	3	0	1	1	0	7	6	1	1	1	1	0,9999	0,91
28	2	1	1	1	0	2	1	0	1	1	3	0,9999	0,78
52	11	0	1	1	0	5	6	0	1	1	3	0,0001	0,02
42	6	0	1	1	0	6	4	1	2	1	2	0,9999	0,42
22	11	0	1	1	0	6	6	0	1	1	3	0,0001	0,21
3	6	1	1	1	0	3	30	0	1	1	3	0,0001	0,29
50	3	0	1	1	0	7	28	0	1	1	3	0,9999	0,71
17	11	0	1	1	0	7	22	0	1	1	3	0,0001	0,33
18	6	0	1	1	0	6	21	0	1	1	3	0,0001	0,65
26	3	0	1	1	0	6	6	0	1	1	3	0,0001	0,81
54	2	0	1	1	1	7	29	0	1	1	4	0,9999	0,67
54	10	0	0	0	0	2	22	0	1	1	3	0,9999	0,62
54	7	0	1	1	0	4	22	0	1	1	4	0,9999	0,25
30	6	0	1	1	0	5	19	0	2	2	4	0,0001	0,17
42	3	0	1	1	0	7	6	0	2	1	3	0,9999	0,89
27	10	1	1	1	0	7	1	0	3	1	4	0,0001	0,29
11	7	1	1	1	0	7	6	0	1	1	5	0,9999	0,97
41	3	1	1	1	1	6	19	1	1	1	3	0,9999	1,00
4	3	0	1	1	0	2	18	0	1	1	3	0,0001	0,74
5	7	0	1	1	0	7	2	0	3	1	3	0,9999	0,63
4	2	0	1	1	0	5	21	0	2	1	4	0,0001	0,27

(Bir Önceki Sayfadan Devam)													
11	2	1	1	1	0	6	6	0	1	1	3	0,9999	1,00
53	6	0	1	1	0	6	1	1	1	1	3	0,0001	0,04
30	6	0	1	1	0	6	6	0	1	1	3	0,9999	0,84
5	5	0	1	1	0	7	22	0	1	1	3	0,0001	0,77
45	11	0	0	0	0	2	1	1	1	1	3	0,0001	0,00
56	3	0	1	1	0	5	22	1	1	1	2	0,0001	0,42
15	3	0	1	1	1	4	6	0	1	1	4	0,9999	0,73
24	2	0	1	1	1	6	6	0	1	1	3	0,9999	0,79
33	2	0	1	1	1	6	30	1	1	1	2	0,9999	0,99
18	6	2	1	0	0	2	14	0	1	1	3	0,0001	0,12
55	6	0	1	1	0	3	1	0	1	1	3	0,0001	0,05
9	3	0	1	1	0	6	2	0	1	1	3	0,9999	0,98
11	7	0	1	1	0	7	6	0	1	1	3	0,9999	0,95
50	10	1	1	1	0	2	7	0	1	1	3	0,0001	0,00
17	11	0	1	1	0	6	31	0	1	1	2	0,0001	0,08
27	6	0	1	1	0	1	26	0	1	1	3	0,0001	0,05
26	2	0	1	1	0	2	1	1	3	1	3	0,9999	0,97
8	11	0	0	0	0	5	11	0	1	1	3	0,0001	0,08
57	11	0	1	1	0	6	30	0	1	1	3	0,0001	0,15
30	6	0	1	1	0	6	29	0	1	1	3	0,0001	0,29
51	11	0	1	1	1	2	30	0	1	1	3	0,0001	0,10
47	3	0	1	1	0	5	21	0	1	1	4	0,9999	0,67
55	6	0	1	1	0	6	6	0	2	2	5	0,9999	0,76
18	6	0	1	1	1	5	30	0	1	1	3	0,0001	0,18
16	10	1	1	1	0	6	2	0	1	1	3	0,9999	0,52
49	3	0	1	1	0	6	6	0	3	1	3	0,9999	0,94
12	11	0	1	1	0	5	4	0	1	1	2	0,0001	0,03
6	11	0	0	0	0	7	22	0	1	1	3	0,0001	0,25
27	6	0	1	1	0	4	20	0	2	1	3	0,9999	0,81
57	11	0	0	0	0	6	1	1	1	1	3	0,0001	0,09
44	2	0	1	1	0	6	30	0	1	1	3	0,9999	0,85
7	11	0	1	1	0	5	15	0	1	1	3	0,0001	0,17
4	7	0	1	1	0	5	23	1	1	1	1	0,9999	0,68
27	3	0	1	1	0	5	20	0	1	1	3	0,0001	0,56
27	6	1	1	1	0	2	29	0	1	1	3	0,0001	0,21
11	11	1	0	0	0	5	28	0	3	1	3	0,0001	0,15
30	4	0	1	1	0	6	30	0	1	1	4	0,9999	0,60
24	2	1	1	1	0	5	2	1	2	1	3	0,9999	0,93
27	6	0	1	1	1	1	25	0	1	1	3	0,9999	0,89
10	6	0	1	1	0	5	30	0	1	1	3	0,0001	0,06
10	6	0	1	1	0	5	19	0	1	1	2	0,9999	0,49
47	10	1	1	1	0	6	6	0	1	1	3	0,9999	0,52
28	7	1	1	1	0	6	30	0	2	1	4	0,9999	0,97
30	6	0	1	1	0	6	12	0	2	1	3	0,9999	0,90
26	11	1	1	1	0	6	6	0	1	1	3	0,0001	0,40
17	11	0	1	1	0	6	1	0	1	1	3	0,0001	0,09
42	11	1	1	1	0	4	31	0	1	1	3	0,0001	0,29

(Bir Önceki Sayfadan Devam)													
48	3	0	1	1	0	4	7	0	1	1	3	0,0001	0,43
16	3	0	1	1	0	2	14	0	1	1	4	0,9999	0,59
56	6	0	1	1	1	6	6	0	1	1	3	0,9999	0,96
15	11	0	1	1	0	6	6	0	2	1	4	0,0001	0,27
37	2	0	1	1	1	6	6	1	1	1	2	0,9999	1,00
4	8	0	1	1	0	6	23	0	2	1	1	0,9999	0,88
5	11	0	1	1	0	5	18	0	1	1	4	0,0001	0,10
47	10	0	1	1	1	6	6	0	1	1	2	0,9999	0,94
17	3	0	1	1	0	5	22	0	4	1	3	0,9999	0,82
5	3	0	1	1	0	2	22	0	1	1	3	0,9999	0,54
52	9	0	1	1	1	5	23	1	1	1	2	0,9999	0,83
37	7	0	1	1	0	7	4	0	2	1	2	0,9999	0,87
4	2	1	0	0	0	6	28	0	1	1	3	0,0001	0,36
27	6	0	1	1	0	6	20	1	1	1	3	0,0001	0,12
27	6	0	1	1	0	3	14	1	1	1	3	0,0001	0,08
44	6	0	1	1	0	2	6	0	1	1	3	0,0001	0,03
17	11	0	1	1	0	3	23	0	2	2	3	0,0001	0,00
6	11	0	1	1	0	2	4	1	1	1	3	0,0001	0,02
10	5	0	1	1	0	6	23	1	1	1	2	0,9999	0,54
16	11	0	1	1	0	6	2	0	1	1	3	0,0001	0,11
18	6	0	1	1	0	4	9	1	1	1	4	0,9999	0,04
43	11	1	1	1	0	1	6	0	1	1	3	0,0001	0,00
54	11	1	1	1	0	4	6	0	1	1	3	0,0001	0,02
30	6	1	1	1	0	7	14	0	2	1	3	0,9999	0,98
27	3	0	1	1	0	7	1	0	2	1	3	0,0001	0,90
4	8	0	1	1	0	5	14	0	1	1	3	0,9999	0,72
56	6	0	1	1	0	2	30	0	1	1	3	0,0001	0,23
17	11	0	0	0	0	5	2	0	1	1	2	0,0001	0,00
27	6	0	1	1	0	5	22	0	1	1	2	0,0001	0,17
15	8	0	1	1	0	6	6	0	2	1	3	0,9999	0,75
5	7	0	1	1	0	6	6	0	2	1	4	0,9999	0,80
21	3	1	1	1	0	2	25	0	1	1	3	0,9999	0,92
50	10	0	0	0	0	7	27	0	1	1	3	0,9999	0,88
25	3	0	1	1	1	6	26	1	1	1	3	0,9999	0,96
53	6	0	0	0	0	3	22	0	1	1	3	0,0001	0,44
15	10	0	1	1	0	5	2	0	1	1	3	0,9999	0,08
50	7	1	1	1	0	6	6	0	1	1	3	0,9999	0,80
7	2	1	1	1	0	5	28	0	1	1	3	0,9999	0,99
19	11	0	0	0	0	6	2	0	1	1	3	0,0001	0,00
12	8	0	1	1	1	6	20	0	1	1	3	0,9999	0,70
27	6	1	1	1	0	6	6	0	1	1	3	0,9999	0,95
51	3	0	1	1	0	6	6	0	2	1	3	0,9999	0,86
54	7	1	1	1	0	5	4	0	1	1	3	0,9999	0,57
6	3	0	1	1	0	6	1	1	1	1	4	0,0001	0,09
47	11	0	1	1	0	2	19	0	2	2	5	0,0001	0,00
54	11	0	0	0	0	6	22	1	1	1	3	0,0001	0,00
16	3	0	1	1	0	6	2	0	1	1	3	0,9999	0,94

(Bir Önceki Sayfadan Devam)													
11	2	1	1	1	0	1	30	0	3	1	4	0,0001	0,17
14	7	3	1	1	0	6	7	1	1	1	3	0,9999	0,97
17	11	0	1	1	0	6	1	0	2	2	4	0,0001	0,00
27	6	0	1	1	0	6	6	1	2	1	3	0,0001	0,11
16	3	1	1	1	0	7	2	0	3	1	4	0,9999	0,96
45	5	1	1	1	1	3	22	0	1	1	4	0,9999	0,88
46	10	1	1	1	0	5	30	0	1	1	3	0,9999	0,30
18	6	0	1	1	0	6	30	0	2	1	3	0,0001	0,71
23	7	0	0	0	0	5	23	1	1	1	2	0,9999	0,83
30	6	0	1	1	0	3	22	1	1	1	3	0,0001	0,08
55	6	3	1	1	0	5	2	0	2	1	4	0,9999	0,96
47	5	0	1	1	0	6	6	0	1	1	3	0,9999	0,75
6	11	0	1	1	0	4	3	0	1	1	3	0,0001	0,01
28	11	0	0	0	0	2	22	1	1	1	3	0,0001	0,11
15	11	0	0	0	0	5	22	0	1	1	3	0,0001	0,23
23	2	0	1	1	0	5	6	0	2	1	3	0,9999	0,88
15	5	1	1	1	0	5	9	0	1	1	4	0,9999	0,97
7	11	0	1	1	0	6	28	0	1	1	3	0,0001	0,07
17	11	0	0	0	0	6	12	1	1	1	3	0,0001	0,01
20	6	0	1	1	0	6	6	0	1	1	4	0,9999	0,91
2	3	0	1	1	0	5	6	1	1	1	3	0,0001	0,26
45	11	0	0	0	0	1	6	1	1	1	2	0,0001	0,01
7	11	0	1	1	0	6	5	0	1	1	3	0,0001	0,19
3	6	0	1	1	0	6	1	1	1	1	3	0,0001	0,04
6	11	0	1	1	0	4	6	0	1	1	3	0,0001	0,02
16	11	0	1	1	0	6	2	1	1	1	3	0,0001	0,01
16	3	0	1	1	0	5	22	0	1	1	3	0,9999	0,64
28	11	0	1	1	0	3	1	0	1	1	3	0,0001	0,00
30	6	0	1	1	0	5	12	1	2	1	3	0,0001	0,05
19	6	1	0	0	0	5	30	0	1	1	3	0,9999	0,40
42	7	1	1	1	0	6	6	0	1	1	4	0,9999	0,87
18	6	1	1	1	0	3	6	1	1	1	4	0,0001	0,28
40	11	1	1	1	0	6	11	0	1	1	3	0,0001	0,46
42	5	1	1	1	1	6	1	0	1	1	3	0,9999	0,98
47	3	0	1	1	0	6	2	0	1	1	2	0,9999	0,60
16	11	0	1	1	0	6	31	0	2	1	3	0,0001	0,25
24	2	0	1	1	0	6	6	0	2	1	3	0,9999	0,89
55	6	1	1	1	0	5	22	0	1	1	4	0,0001	0,75
16	11	1	1	1	0	5	28	0	1	1	3	0,0001	0,13
54	7	3	1	1	0	7	22	1	2	1	3	0,9999	0,83
26	1	0	1	1	0	6	23	1	2	1	1	0,0001	0,49
27	6	0	1	1	0	7	23	1	2	1	1	0,0001	0,15
26	1	1	1	1	1	6	20	0	1	1	3	0,0001	0,80
15	11	1	1	1	0	6	6	0	5	1	3	0,0001	0,20
26	7	1	1	1	0	2	22	0	1	1	4	0,0001	0,22
17	11	0	1	1	1	6	1	0	1	1	3	0,9999	0,79
27	6	0	1	1	0	3	6	1	1	1	3	0,0001	0,06

46	10	1	1	1	0	6	30	0	2	1	3	0,9999	0,92
21	3	1	1	1	0	5	6	0	1	1	3	0,9999	0,96
RASGELE SEÇİLMİŞ 23 ADET TEST VERİ SETİ:													
27	3	0	1	1	0	4	6	0	1	1	3	0,0001	0,64
1	6	1	1	1	0	3	31	0	1	1	3	0,0001	0,27
42	3	0	0	0	1	6	4	0	1	1	2	0,9999	1,00
47	3	1	1	1	0	6	15	0	1	1	4	0,9999	0,81
52	3	0	1	1	1	6	6	1	1	1	2	0,9999	0,98
27	6	0	1	1	0	6	26	0	1	1	3	0,0001	0,37
50	8	1	1	1	0	5	22	0	1	1	3	0,9999	0,43
16	3	0	1	1	1	5	22	0	1	1	3	0,0001	0,26
42	10	0	1	1	0	5	22	0	1	1	3	0,0001	0,20
9	6	0	1	1	0	5	19	0	1	1	3	0,9999	0,65
7	5	2	1	1	0	2	6	0	1	1	3	0,9999	0,99
1	6	3	1	1	0	2	8	0	1	1	3	0,9999	1,00
15	3	0	1	1	0	5	6	0	1	1	4	0,9999	0,92
44	11	0	1	1	0	2	23	0	1	1	2	0,0001	0,08
47	3	1	1	1	0	6	6	0	1	1	4	0,9999	0,84
27	6	0	1	1	0	2	23	1	1	1	1	0,0001	0,75
3	6	0	0	0	1	2	1	1	1	1	3	0,9999	0,19
45	11	0	0	0	0	6	1	0	1	1	3	0,0001	0,00
48	3	0	1	1	0	4	23	0	1	1	2	0,0001	0,54
17	11	0	0	0	0	6	1	0	1	1	3	0,0001	0,00
38	11	0	1	1	0	2	28	0	1	1	3	0,0001	0,08
55	6	0	1	1	0	4	25	0	1	1	3	0,0001	0,15
16	5	0	1	1	1	5	19	1	1	1	3	0,9999	0,82

EK 4. MÜŞTERİ SINIFLANDIRMA EĞİTİM VE TEST VERİ SETİ

Alan 1: Dağıtıcı

Alan 2: Ödeme Planı

Alan 3: Medeni Durum

Alan 4: Promosyon (email almak ister mi?)

Alan 5: E-bilgilendirme (maili almak ister mi?)

Alan 6: SMS almak isityor mu?

Alan 7: Şehir

Alan 8: Meslek

Alan 9: Cinsiyet

Alan 10: Musteri Adedi

Alan 11: Kullanıcı Sayısı

Alan 12: Yaş Grubu

Alan 13: Sms Kullanımı

Alan 14: Tahmini Sms Kullanımı

1	2	3	4	5	6	7	8	9	10	11	12	13	14
27	10	0	0	0	1	7	5	0	3	1	3	0,05	0,12
37	2	0	1	1	1	6	6	1	1	1	2	0,24	0,30
17	11	1	1	1	0	6	30	0	2	1	3	0,41	0,61
49	7	0	1	1	0	5	30	0	1	1	4	0,21	0,37
32	3	0	1	1	1	2	3	0	2	1	2	0,15	0,21
27	3	0	1	1	1	7	6	1	2	1	1	0,41	0,41
44	11	0	1	1	0	7	28	0	2	1	3	0,18	0,40
10	5	0	1	1	0	6	23	1	1	1	2	0,27	0,23
16	3	1	1	1	0	7	2	0	3	1	4	0,32	0,36
11	9	0	1	1	0	6	28	0	1	1	4	0,98	0,80
57	11	0	1	1	0	4	4	0	1	1	3	0,15	0,12
24	2	0	1	1	0	6	6	0	2	1	3	0,67	0,49
7	2	0	1	1	0	5	4	0	1	1	3	0,25	0,43
4	8	0	1	1	0	5	14	0	1	1	3	0,74	0,62

(Bir Önceki Sayfadan Devam)													
11	3	0	1	1	0	6	28	0	1	1	3	0,66	0,41
6	3	1	1	1	0	1	27	0	1	1	3	0,26	0,20
16	3	0	1	1	0	7	14	0	3	2	5	0,55	0,55
52	6	0	1	1	0	6	2	0	1	1	4	0,22	0,30
27	3	0	1	1	0	7	6	1	1	1	1	0,26	0,29
31	6	0	1	1	0	6	25	0	2	1	3	0,25	0,31
47	3	0	0	0	0	7	22	0	1	1	3	0,31	0,17
17	3	0	1	1	0	5	22	0	4	1	3	0,31	0,29
18	6	0	1	1	0	1	6	0	1	1	2	0,23	0,15
18	6	0	1	1	0	4	9	1	1	1	4	0,27	0,33
17	5	1	1	1	0	5	13	0	3	1	3	0,29	0,27
15	3	0	1	1	0	5	6	0	1	1	4	0,27	0,38
7	2	1	1	1	0	5	28	0	1	1	3	0,23	0,33
50	8	1	1	1	0	5	22	0	1	1	3	0,24	0,25
46	3	0	1	1	1	7	25	1	1	1	4	0,24	0,31
52	9	0	1	1	1	5	23	1	1	1	2	0,24	0,18
37	7	0	1	1	0	7	4	0	2	1	2	0,30	0,29
38	3	0	1	1	0	6	23	0	1	1	3	0,31	0,24
16	3	0	1	1	0	5	22	0	1	1	3	0,28	0,35
27	6	0	1	1	0	2	6	0	3	2	5	0,26	0,27
11	6	0	1	1	0	4	6	0	1	1	4	0,24	0,36
5	3	0	1	1	0	2	22	0	1	1	3	0,22	0,23
47	10	0	0	1	0	6	4	0	2	1	2	0,29	0,26
20	6	0	1	1	0	6	6	0	1	1	4	0,26	0,39
10	6	1	1	1	0	2	20	0	1	1	3	0,26	0,33
42	3	0	1	1	0	7	6	0	2	1	3	0,29	0,49
30	6	1	1	1	0	7	22	0	1	1	3	0,27	0,32
26	3	0	1	1	0	2	6	0	1	1	3	0,31	0,21
16	10	1	1	1	0	6	2	0	1	1	3	0,31	0,32
45	5	1	1	1	1	3	22	0	1	1	4	0,25	0,45
16	3	0	1	1	0	2	14	0	1	1	4	0,23	0,24
33	2	0	1	1	1	5	30	1	1	1	2	0,22	0,19
46	7	1	1	1	0	7	6	0	1	1	3	0,25	0,20
15	5	1	1	1	0	5	9	0	1	1	4	0,22	0,33
7	5	2	1	1	0	2	6	0	1	1	3	0,24	0,20
55	3	0	1	1	0	2	6	0	1	1	4	0,31	0,35
5	7	0	1	1	0	6	6	0	2	1	4	0,29	0,41
23	6	1	1	1	0	5	2	0	1	1	3	0,27	0,25
6	3	1	1	1	0	6	27	0	1	1	3	0,25	0,41
16	3	0	1	1	1	5	2	0	1	1	3	0,72	0,57
33	2	0	1	1	1	6	30	1	1	1	2	0,23	0,20
54	7	3	1	1	0	7	22	1	2	1	3	0,71	0,64
27	6	0	1	1	0	3	14	1	1	1	3	0,19	0,18
26	9	1	1	1	1	6	25	0	1	1	4	0,75	0,61
54	7	0	1	1	0	4	22	0	1	1	4	0,50	0,32

(Bir Önceki Sayfadan Devam)													
12	8	0	1	1	1	6	20	0	1	1	3	0,51	0,51
23	2	0	1	1	0	5	6	0	2	1	3	0,21	0,44
46	10	1	1	1	0	5	30	0	1	1	3	0,44	0,42
16	3	0	1	1	1	5	22	0	1	1	3	0,28	0,34
11	2	1	1	1	0	6	6	0	1	1	3	0,73	0,38
15	8	0	1	1	0	6	6	0	2	1	3	0,19	0,37
46	3	0	1	1	0	6	24	0	1	1	3	0,25	0,21
16	3	0	1	1	1	6	2	0	1	1	3	0,72	0,66
9	3	0	1	1	0	6	2	0	1	1	3	0,25	0,47
15	3	1	1	1	0	5	6	0	1	1	4	0,20	0,31
44	5	0	1	1	1	5	2	0	1	1	4	0,65	0,43
17	1	1	1	1	1	6	22	0	1	1	3	0,61	0,51
30	9	1	1	1	0	6	31	0	3	1	3	0,28	0,36
55	6	0	1	1	0	3	6	0	2	1	3	0,26	0,28
55	6	0	1	1	0	6	13	1	1	1	3	0,30	0,29
9	7	1	1	1	0	6	1	0	1	1	3	0,15	0,33
18	6	1	1	1	0	6	30	0	1	1	3	0,24	0,44
11	6	0	1	1	0	6	6	0	2	1	3	0,77	0,38
54	7	1	1	1	0	5	4	0	1	1	3	0,05	0,18
28	7	1	1	1	0	6	30	0	2	1	4	0,87	0,48
16	3	0	1	1	0	6	2	0	1	1	3	0,72	0,45
30	9	1	1	1	1	6	1	1	2	1	3	0,64	0,59
4	8	0	1	1	0	6	23	0	2	1	1	0,62	0,46
47	3	0	1	1	0	6	2	0	1	1	2	0,35	0,36
15	10	2	1	1	0	5	6	1	1	1	4	0,49	0,44
15	3	0	1	1	1	4	6	0	1	1	4	0,40	0,41
15	3	1	1	1	0	6	6	0	1	1	4	0,28	0,34
42	6	0	1	1	0	6	4	1	2	1	2	0,41	0,37
55	2	0	1	1	1	3	31	0	1	1	3	0,47	0,37
15	3	0	0	0	0	6	23	0	1	1	2	0,20	0,29
55	3	0	1	1	1	4	23	0	1	1	2	0,41	0,32
15	7	0	1	1	0	7	6	0	1	1	3	1,00	0,47
54	2	0	1	1	1	7	29	0	1	1	4	0,39	0,43
24	2	0	1	1	1	6	6	0	1	1	3	0,42	0,61
50	7	1	1	1	0	6	6	0	1	1	3	0,24	0,19
11	7	1	1	1	0	7	6	0	1	1	5	0,42	0,38
11	3	1	1	1	0	5	6	0	1	1	3	0,59	0,33
54	7	0	0	0	1	3	6	1	1	1	3	0,05	0,04
56	6	0	1	1	1	6	6	0	1	1	3	0,10	0,30
39	6	0	1	1	0	6	23	1	2	1	1	0,05	0,12
5	7	0	1	1	0	7	2	0	3	1	3	0,19	0,35
51	3	0	1	1	0	6	6	0	2	1	3	1,00	0,47
37	7	0	1	1	0	7	6	0	1	1	3	0,05	0,28
11	7	0	1	1	0	7	6	0	1	1	3	0,49	0,51
10	6	1	1	1	1	2	20	0	1	1	3	0,10	0,20
51	3	0	1	1	0	5	6	0	1	1	3	0,55	0,35
27	6	0	1	1	1	1	25	0	1	1	3	0,20	0,13

(Bir Önceki Sayfadan Devam)													
27	6	0	1	1	0	6	1	0	1	1	3	0,05	0,33
56	3	0	1	1	0	6	2	0	1	1	2	0,05	0,33
30	6	0	1	1	1	6	30	0	1	1	2	0,22	0,30
30	6	1	1	1	0	7	14	0	2	1	3	0,05	0,27
14	7	3	1	1	0	6	7	1	1	1	3	0,21	0,31
27	3	0	1	1	0	7	30	0	1	1	3	0,10	0,27
47	6	1	0	0	1	7	27	0	1	1	4	0,05	0,13
46	10	1	1	1	0	6	30	0	2	1	3	0,69	0,39
54	2	1	1	1	0	5	28	0	2	1	3	0,05	0,25
42	3	0	1	1	0	3	27	0	2	1	3	0,24	0,22
18	6	0	1	1	0	1	16	0	1	1	2	0,05	0,18
16	7	0	1	1	1	2	2	1	1	1	3	0,10	0,18
50	10	0	0	0	0	7	27	0	1	1	3	0,20	0,15
47	3	1	1	1	0	6	15	0	1	1	4	0,39	0,33
3	6	0	0	0	1	2	1	1	1	1	3	0,19	0,17
17	10	0	1	1	0	5	23	0	2	1	2	0,33	0,47
38	3	3	1	1	0	6	6	0	3	1	4	0,21	0,25
15	10	0	1	1	0	5	2	0	1	1	3	0,51	0,40
RASGELE SEÇİLMİŞ 12 ADET TEST VERİ SETİ:													
15	1	0	1	1	0	2	6	0	1	1	3	0,32	0,22
23	7	0	0	0	0	5	23	1	1	1	2	0,23	0,29
4	7	0	1	1	0	5	23	1	1	1	1	0,45	0,26
26	2	0	1	1	0	2	1	1	3	1	3	0,31	0,46
42	3	0	0	0	1	6	4	0	1	1	2	0,15	0,16
47	10	1	1	1	0	6	6	0	1	1	3	0,20	0,16
47	5	1	1	1	0	7	6	0	1	1	3	0,20	0,25
49	11	0	1	1	0	6	28	0	1	1	4	0,22	0,51
13	11	0	1	1	0	6	2	0	3	1	3	0,02	0,26
25	11	0	0	0	0	6	31	0	2	1	3	0,15	0,38
27	6	0	1	1	0	6	16	0	1	1	3	0,35	0,36
47	3	1	1	1	0	6	15	0	1	1	4	0,39	0,33
Ortamala Değerler:												0,328	0,336

ÖZGEÇMİŞ

1973 yılında Sivas'ta doğdu. İlköğrenimini Trabzon ve Hatay'da tamamladı. 1990 yılında Haydarpaşa Lisesi'nden mezun oldu ve aynı yıl İstanbul Teknik Üniversitesi Makina Fakültesi, Makina Mühendisliği bölümünü kazandı. İngilizce hazırlık sınıfını başarıyla tamamladıktan sonra mühendislik eğitimini de tamamlayıp 1996 yılında makine mühendisi oldu. Aynı yıl evlendi ve İstanbul Üniversitesi, İşletme Fakültesi, Sayısal Yöntemler Ana Bilim Dalı'nda araştırma görevlisi olarak çalışmaya başladı ve bu görevini 2000 yılına kadar devam ettirdi. 1996 yılında aynı üniversitede başladığı yüksek lisans eğitimini 1998 yılında başarıyla tamamladı ve ardından aynı kurumda doktora programına başladı.

Doktora çalışmaları devam ederken burslu olarak katıldığı, I.Ü. İşletme Fakültesi İşleme İktisadi Enstitüsü, İşletmecilik İhtisas Gece programını birincilikle tamamladı. Yine doktora çalışmalarını devam ettirmek ve bilişim sistemleri üzerine bilgi ve birikimini geliştirmek üzere gittiği A.B.D.'de San Diego State Üniversite'sinde (SDSU) Information and Decision Systems Master of Science in Business Administration (MSBA) programına katıldı. 2001 yılında SDSU'dan da yüksek lisans derecesi almaya hak kazandı. Aynı yıl Perisu Meriç isimli kızı dünyaya geldi.

2002 ve 2003 yıllarında, özel firmalarda dış ticaret ve bilgi işlem sorumlusu olarak görev yaptı.