T.C.
GEBZE YÜKSEK TEKNOLOJİ ENSTİTÜSÜ
SOSYAL BİLİMLER ENSTİTÜSÜ

KREDİ KARTI KULLANIMINDA SAHTECİLİK TESPİT SİSTEMLERİ

Yavuz Selim KERESTECİ

YÜKSEK LİSANS TEZİ
STRATEJİ BİLİMİ ANABİLİM DALI

GEBZE
2008

T.C.

GEBZE YÜKSEK TEKNOLOJİ ENSTİTÜSÜ

SOSYAL BİLİMLER ENSTİTÜSÜ

KREDİ KARTI KULLANIMINDA SAHTECİLİK TESPİT SİSTEMLERİ

Yavuz Selim KERESTECİ

STRATEJİ BİLİMİ ANABİLİM DALI

TEZ DANIŞMANI

Yrd.Doç.Dr. Hüseyin İNCE

GEBZE

2008

ÖZET

TEZİN BAŞLIĞI : KREDİ KARTI KULLANIMINDA SAHTECİLİK TESPİT SİSTEMLERİ

YAZAR ADI : YAVUZ SELİM KERESTECİ

Gelişen teknolojiler her geçen gün büyük bir hızla hayatımıza girerek yaşamın her alanında insan hayatını kolaylaştırmaktadır. Bilgisayar ve iletişim alanındaki teknolojik gelişmeler günümüzde insanlık tarihi acısından çok önemli bir devrim olarak kabul edilmektedir. Bu teknolojik gelişmelerle beraber insanların harcamalarının artması paranın yerini alan kredi kartı kullanımınında çok büyük bir oranda artmasına sebep olmakta, bütün bunlarla beraber bu kadar çok artan kredi kartı kullanımıyla birlikte kredi kartı sahteciliği de doğru orantılı olarak artmaktadır. Kredi kartı sahteciliğinin artması etkin ve verimli şekilde kullanabilen sahtecilik sistemlerini gündeme getirmektedir.

Çalışmanın birinci bölümünde kredi kartı tanımı, kredi kartı sahteciliği, yapılmış çalışmalar, sahtecilik ve sahtecilik tespit yöntemlerinde kullanılan teknikler hakkında genel bilgiler verilmeye çalışılmıştır. İkinci bölümde genel olarak sahtecilik terimi üzerinde durulmuş olup sahtecilik tanımı, sahtecilik ile ilgili kavramlar, yıllık kredi kartı rakamları, kredi kartı sahtecilik türleri, sahtecilik metodları, sahtecilik çeşitleri ve sahte bilgi elde etme yöntemleri hakkında geniş bilgi sunulmuştur. Üçüncü bölümde sahtecilik tespitinde kullanılan veri madenciliği, yapay zeka ve istatistiksel teknikler ayrıntılı bir şekilde incelenmiştir. Dördüncü bölümde bir finans kurumdan elde ettiğimiz kredi kartı harcamalarının olduğu veri seti, yapay zeka ve istatistiksel teknikler ile incelenerek test edilmiş ve yorumlanmıştır.

Sonuç olarak, kullanmış olduğumuz veri madenciliği teknikleri (Yapay Zeka, Uyarlanabilir Ağ-Tabanlı Bulanık Çıkarım Sistemleri, Destek Vektör Makineleri, Kural Tabanlı Öğrenme, Lojistik Regresyon ve Diskriminant Analizi) ile elde ettiğimiz sonuçlar doğru sınıflandırma oranı, birinci tip hata ve ikinci tip hata kriterleri ile yorumlanmıştır.

SUMMARY

TITLE **:** CREDIT CARD FRAUD DETECTION SYSTEMS

AUTHOR NAME **:** YAVUZ SELİM KERESTECİ

Technological evolutions, spreading into our life on an increasing scale with each passing day, are making human life simpler in many aspects. Technological improvements in computer and communication areas are currently being considered as the most essential revolution in history of civilization. The increase in expenditures, a by-product of these technological developments generalizes the use of credit cards as a medium of payment, superseding the cash money, also causes a proportionally expansion in number of credit cards fraud cases. Hence, an increase in credit cards fraud cases brings efficient and productive falsification systems into sharp relief.

In the first section, the definition and falsification of credit cards, literature review, fraud and falsification detection methods are explained by giving general information. Second section, by and large, concentrates on the term of fraud and highlights the definition of frauds as well as concepts, annual credit cards figures, fraud types of credit cards, methods and sorts of frauds in addition to the approaches of gaining fake information. In the third section, data mining methods used in fraud detection, artificial intelligence and statistical techniques are examined in detail. In the fourt section, a database regarding a set of credit cards expenditures, compiled by a financial institution, are submitted to an examination and interpreted through artificial intelligence and statistical techniques.

As a consequence, results derived through data mining techniques (Neural Network, Adaptive Network Based Fuzzy Inference System, Support Vector Machine, Rule Based Learning, Logistic Regression, Discriminant Analysis) are commented with the criteria of correct classification rate, Type I Error and Type II Error.

TEŞEKKÜR

Kredi kartı kullanımında sahtecilik tespit sistemleri isimli bu tez çalışmamı hazırlamamda desteğini esirgemeyen çok değerli eşime ve aileme, eleştirileri, önerileri, yol göstericiliği ile birlikte tezimi hazırlamama olanak sağlayan sayın hocam Yrd. Doç. Dr. Hüseyin İNCE'ye teşekkürlerimi sunarım.

Yüksek lisans öğrenimim boyunca desteklerini benden esirgemeyen Mehmet Fatih Keresteci ve Mehmet Tahir Zazaoğlu'na teşekkürü bir borç bilirim. Verdiği destekten dolayı TUBİTAK'a da teşekkür ederim.

İÇİNDEKİLER DİZİNİ

SİMGELER VE KISALTMALAR DİZİNİ

ADALINE : Adaptif Lineer Neuron.

ANFIS : Adaptive Neural Fuzzy Inference Systems.

ANOVA : Analysis Of Variance.

ART : Adaptive Resonance Theory.

ATM : Automatic Teller Machine.

BKM : Bankalararası Kart Merkezi.

BM : Bulanık Mantık.

GA : Genetik Algoritma.

GSM : Global System for Mobile.

GUI : Graphical User Interface.

IBM : International Business Machines.

LVQ : Learning Vector Quantization.

MANOVA : Multivariate Analysis Of Variance.

MLP : Multilayer Perceptron.

MOGP : Multi Objective Genetic Programming.

POS : Point Of Sale.

SMO : Sequential Minimal Optimization.

SOM : Self-Organized Maps.

SPSS : Statistical Package for the Social Sciences.

SVM : Support Vector Machine.

US : Uzman Sistemler.

WEKA : Waikato Environment for Knowledge Analysis.

YSA : Yapay Sinir Ağları.

ŞEKİLLER DİZİNİ

TABLOLAR DİZİNİ

1. GİRİŞ

Kredi kartı sahteciliği, kredi kartı endüstrisi içinde büyüyen bir problemi teşkil etmektedir. Dünya çapında her yıl kredi kartı sahteciliğinin etkileyici olarak büyümesinden dolayı milyonlarca dolar kayıp meydana gelmekte, sahteciliği önlemek için birçok iş alanında modern teknikler kullanılarak devamlı olarak sahtecilik önleme sistemleri geliştirilmekte ve uygulanmaktadır. Teknolojik sistemler içinde günlük yaşantımızda kullandığımız, örneğin telekomünikasyon ağları, mobil iletişim, online bankacılık ve e-ticaret gibi birçok alanda sahtecilik faaliyetleri meydana gelmektedir. Buna bağlı olarak günümüzde sahtecilik tespit sistemlerinin araştırılması önemli bir konu haline gelmiştir.

Kredi kartı, bankaların ve bazı finansman kuruluşlarının müşterilerine verdiği, anlaşmalı POS cihazı bulunan alışveriş noktalarında ödeme amaçlı veya banka ATM'lerinden nakit avans çekmek amaçlı kullanılabilen, yapılan harcamaların aylık olarak bankaya tek seferde ya da taksitlerle ödenmek zorunda olunduğu, nakit paraya alternatif bir ödeme aracıdır (Wikipedia, 2008).

Kredi kartları, bankacılık sektörü içerisinde pazarı hızla genişleyen hizmetlerden biridir. Gelişen teknolojiler ile birlikte kredi kartının bu denli yoğun kullanıldığı bir ortamda kredi kartı sahtecilik tespit sistemleri gündeme gelmekte ve pazarda önemli bir rol üstlenmektedir. Bugün hizmetlerin sağlıklı yürütülebilmesi için hem kredi kartı başvuru aşamasında hemde işlem aşamasında kredi kartı sahtekarlıklarını önlemek tüm bankalar için öncelikli bir görev haline gelmiştir. Böylelikle kredi kartının her aşamasında etkili sahtecilik tespit sistemlerinin olması sahtecilik ve dolandırıcılık eylemlerine karşı bir önlem olmaktadır.

Etkin ve verimli bir sahtecilik tespit sistemi oluşturabilmek için öncelikle sahteciliğin ne olduğu ve sahtecilik çeşitleri hakkında bilgi sahibi olmak gerekmektedir. Elde edilen bu bilgiler ışığında sahteciliğe karşı ne tür bir önlem almamız gerektiği konusunda da daha net fikirler ortaya çıkacaktır. Sahtecilik tespit sistemlerinin oluşturulmasının bir faydası da sahtecilik sonucu meydana gelen kayıplarda gözlenen artışı düşürmek, en aza indirmektir. Sahtecilik tespit sistemleri

oluşturulurken veri madenciliği, yapay zeka ve istatistiksel teknikler kullanılmaktadır. Kullanılan bu teknikler sayesinde kişinin harcama alışkanlıkları, kişi tipleri, işlem tipleri, profilleri ve daha birçok kategoriye bakılarak, kişi yada kredi kartı hareketleri birçok kıstasa göre değerlendirip, sistemin kendi kurduğu işlem mekanizmasından geçirildikten sonra sistemin yaptığı değerlendirme sonucuna göre çıktıları yorumlayıp, sonuç çıkaran etkin sistemler oluşturmak hedeflenmektedir.

Çalışmamızı yaparken veri madenciliği, yapay zeka ve istatistiksel teknikleri etkin bir şekilde kullanarak, araştırmamızın sonucu için en iyi model belirlenmiştir. Bu teknikler araştırmaların önemli bileşenlerinden birini oluşturarak birçok çalışmanın hedefine ulaşmasına öncülük etmiştir.

Kredi kartı sahtecilik tespiti ve önleme alanında yapılan çalışmada geri yayılımlı ileri beslemeli sinir ağları çok büyük uygulamalarda başarıyla kullanılmıştır (Quah ve diğ., 2007). Yapay sinir ağları ile kredi kartı sahtecilik tespit sistemi oluşturulan bir çalışmada, kredi kartını çıkaran kurumdan sağlanan kredi kartı hesap hareketleri geniş örneklerle sınıflandırılarak, eğitilmiştir (Srivastava ve diğ., 2008). Yapılan bir başka çalışmada sinir ağları ile sahtecilik sınıflandırma modeli geliştirilmiştir. Modelin sonuçlarına bakıldığında sahtecilik tespit araçlarını kullanırken sinir ağlarının büyük kapasiteye sahip olduğu görülmüştür (Kotsiantis ve diğ., 2006). Yapay sinir ağları kullanılarak para aklama sahteciliğini tespit etmek amacıyla etkin ve verimli uygulamalar geliştirilmiştir (Bolton ve diğ., 2002).

1908 kredi kartı başvurusu, lojistik regresyon ve diskriminant analizi ile değerlendirildiğinde lojistik regresyonun diskriminant analizine göre daha iyi sonuç verdiği görülmüştür (Çinko, 2006). Geçmiş kredi kartı bilgilerini kullanan lojistik regresyon ve diskriminant analizi teknikleri ile sahtecilik puanlama modeli gerçekleştirilmiştir (Shen ve diğ., 2007). Sahte finansal raporları kullanan bir başka çalışmada 77 şüpheli ve 305 normal kaydı lojistik regresyon tekniği kullanarak olasılıklarını hesap eden bir sistem geliştirilmiştir (Kotsiantis ve diğ., 2006). Yapay sinir ağları, lojistik regresyon ve diskriminant analizinin karşılaştırılmasının yapıldığı çalışmada 6000 gözlem yer almıştır, bunlardan 4000 tanesini modelleri kurmak ve kalan 2000 tanesini de modelleri test etmek için kullanılmıştır. Karşılaştırma doğru

sınıflama oranlarına göre yapıldığında yapay sinir ağlarının ilk sırada, lojistik regresyonun ikinci sırada ve diskriminant analizinin son sırada yer aldığı görülmüştür (Çinko, 2006). Yunan data seti kullanılarak tahrif edilmiş bilançoları ortaya çıkarmak için lojistik regresyon tekniği ile bir model inşaa edilmiştir. Bu model ile % 84 doğruluk oranı elde edildiği rapor edilmiştir (Kotsiantis ve diğ., 2006).

Destek vektör makineleri desen tanıma, bioinformatik ve metin sınıflandırma uygulamalarını içermekte, ABD ve Tayvan piyasalarında % 80'e yakın doğruluk oranı ile tahmin gerçekleştirmektedir (Huang ve diğ., 2007). Makine öğrenme tekniklerinden olan destek vektör makineleri, kurumsal kredi derecelendirme problemlerine ve daha doğru tahmin için yeni modeller geliştirmeyi sağlar (Lee, 2007).

Karar ağaçları, yapay sinir ağları ve bayes inanç ağaçları tekniklerini kullanarak sahte bilançoları ortaya çıkarmak için yapılan araştırmada, performans açısından en iyi performansı % 90.3 doğru sınıflandırma oranı ile bayes ağaçları ortaya koymuştur (Kotsiantis ve diğ., 2006). Bunun yanında karar ağaçları, kredi kartı sahtecilik tespit ve kredi kartı puanlama modellerinde sıkça kullanılan bir sınıflandırma tekniğidir.

Bu çalışmayı yapmamızdaki amaç, kredi kartı kullanımının hayatımıza bu denli yoğun girdiği bir ortamda meydana gelen kredi kartı sahteciliklerine karşı etkin modeller geliştirmek ve geliştirilen bu modellerin gerçek dünya problemlerine karşı başarılı bir şekilde uygulayarak kredi kartı sahteciliğini minimum düzeye çekip daha güvenli bir ortam meydana getirmektir.

2. SAHTECİLİK İLE İLGİLİ KAVRAMLAR

Genel olarak sahtecilik, başkasını aldatmak için kullanılan her türlü hile ve yöntemlerdir. Her hile ve aldatmaya elverişli şey sahtedir. Kanıtlama araçlarında, gerçekliğine inanılması gereken biçim ve alametler; belirti, iz, işaret, nişan, simgede gerçeğin bozulması, değiştirilmesi sahteciliktir. Hile, sahteciliğin ayrılmaz bir öğesidir. Sahtecilik amaç değil, amacı gerçekleştiren; haksız çıkar elde etmek için amaçlanan bir başka suça ulaşmak için kullanılan araçtır. Bir hile suçu olan sahtecilik; yazıda, sözde ve eylemde sahtecilik olarak sınıflandırılabilmektedir. Gerçeğin herhangi bir biçimde değiştirilmesi, gerçeğe aykırı yazı, söz veya eylemin doğru diye ileri sürülmesi sahteciliktir (Kaylan, 2008).

Benzer bir tanımla orijinal belgelerin içeriğinin değiştirilerek bu belgelere benzer yeni belgeler imal edilerek, şahıs veya tüzel kişilerin dolandırılması yoluyla sabıka ve gerçek kimliklerin gizlenerek menfaat sağlanılması veya bir yasak ya da sınırlamanın aşılması sahtecilik suçunu oluşturur (Kaylan, 2008).

Günümüzde sahtecilik tespiti birçok farklı metodla gerçekleştirilmektedir. Sahtecilik çeşitleri arasında kredi kartı sahteciliği, telekomünikasyon alanındaki sahtecilik ve bilgisayarlara karşı saldırı sayılabilir. Kredi kartı sahteciliğini iki tipe ayırabiliriz: Offline sahtecilik ve online sahtecilik (çevrimdışı ve çevrimiçi de diyebiliriz). Offline sahtecilik genellikle kartın fiziksel olarak çalınarak kullanılmasıyla gerçekleştirilir. Bunun önüne geçilmesi, kartın sahtecilik anlamda kullanılmadan önce kartı çıkaran kuruluşun kartı bloke etmesiyle önlenebilmektedir. İkinci tip olan online sahtecilik ise web, telefon üzerinden alışverişlerde ve kartın sahibinin gerek olmadığı durumlarda gerçekleşmektedir. Bu tür sahtecilikte ürün satınalmak için herhangi bir imza yada damgaya gerek olmaksızın sadece kartın bilgileri gerekmektedir (Kou ve diğ., 2004). Kredi kartı sahtecilik tespiti için yapılan bir başka çalışmada ise kurulan yapay zeka tabanlı veritabanı madenciliği uygulaması ile yapılan testlerde umut veren sonuçlar elde edilmiştir. Sonuçlara bakıldığında sahtecilik tespit oranı yüksek bir model geliştirilmiş olduğu görülmektedir (Aleskerov ve diğ., 1997). Potansiyel sahtecilik durumlarını yorumlamaya yarayan harcama modellerini anlamak için yeni, inovatif ve gerçek

zamanlı sahtecilik tespit yaklaşımları kullanılmıştır. Sahteciliği tespit için müşteri harcama davranışlarını kendi düzenleme haritaları (SOM) ile yorumlayıp süzerek analiz edilmiştir (Quah ve diğ., 2007). Bir başka yöntem olarak web servis tabanlı ortak şema kullanarak kredi kartı sahtecilik tespiti için farklı bir sistem kullanılmıştır. Buradaki amaç sahtecilik tespit sistemi olan bankaların sistemlerini diğer bankaların ortak kullanıma açarak onlarında kendi sistemlerini oluşturmasını sağlamaktır (Chiu ve diğ., 2004).

Kredi kartı sahtecilik kavramını, bu kavramı oluşturan öğelerden kredi kartı ve sahtecilik terimlerini ayrı ayrı inceleyerek, açıklayabiliriz.

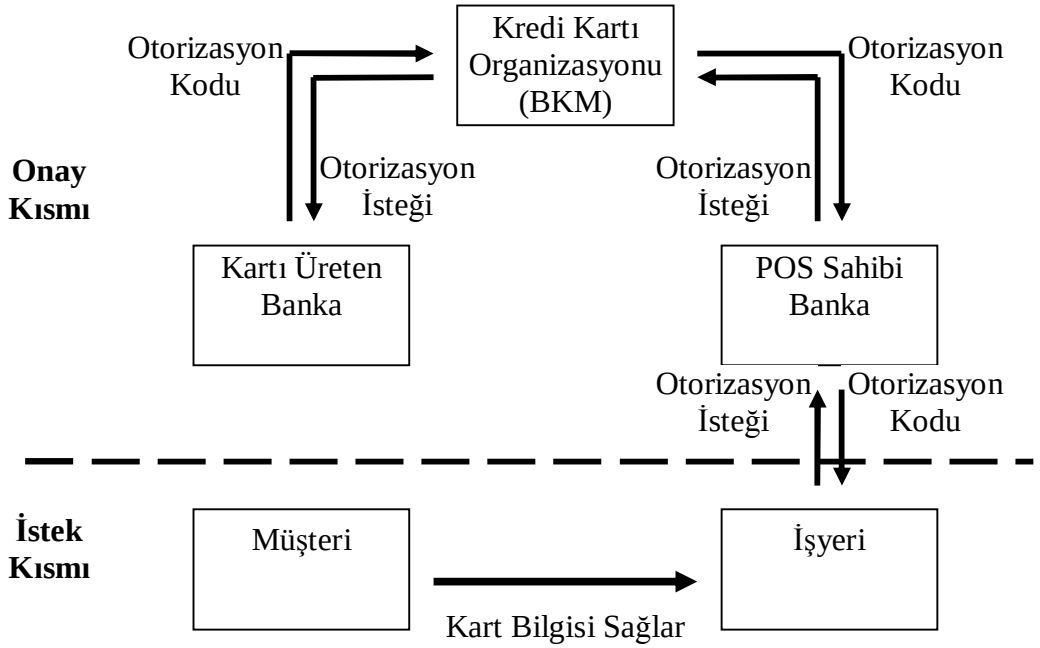
2.1 Kredi Kartı

Kredi kartı ülkemizde kullanılan en yaygın tüketici finansman araçlarının başında gelir. Kredi kartı ürün olarak bir nakit kullanım aracıdır ve kart sahibinin nakit para taşımaksızın alışveriş yapmasına olanak tanır. Bir başka deyişle cebinizde nakit para taşımadan nakit para taşımanın bütün avantajlarını sağlar ve para harcamanıza yardımcı olur. Tüketiciler sahip oldukları kredi kartları ile satış, nakit avans, taksitli satış, elektronik satış, telefon ve posta ile satış veya ödeme işlemlerini gerçekleştirebilirler (Tüketici Finansman Rehberi, 2008)

Kredi kartı, kartı veren finans kuruluşu tarafından tanınan limit dahilinde kredi kullanmak suretiyle harcama yapılmasına olanak veren bir kart türüdür. Bir başka deyişle; kredi kartı, kart hamilinin nakit para taşımaksızın alışveriş yapmasına olanak tanıyan ya da nakit para çekme (kredi) kolaylığı sağlayan bir çeşit ödeme aracıdır. Kart hamili satın aldığı mal ve hizmet bedellerinin tamamını belirli bir süre içerisinde ödeyebileceği gibi belirli bir faiz karşılığında taksitlendirerek de ödeme yapabilir (Çavuş, 2006).

Türkiye'de kredi kartı uygulaması 1968 yılında Diners kart ile başlamış ve sınırlı bir kitleye hitap etmiştir. Yalnızca yurt içinde kullanılan bu kartla, sadece alışveriş yapılabilmekte ve kısa bir süre sonra borcun tamamı ödenmek zorundaydı. Gerçek anlamda kredi kartı olgusu 1980'li yıllarda başlamakla birlikte, sistemin gelişimi ve aktivite kazanması 1983 yılının sonlarındadır. 1983-1988 yılları arasında

kredi kartı sisteminde yer alan bankaların küçük ve az şubeli bankalar olması nedeni ile kredi kartı ve kullanıcı sayısı oldukça azdır. 1990 yılından sonra kredi kartı sayısında gözle görülür artışlar meydana gelmiştir (Çavuş, 2006).



Şekil 2.1 Kredi Kartı Otorizasyon Süreci

Kredi kartı sisteminde üçlü bir ilişki vardır. İlişkinin bir tarafında kartı çıkaran banka, bir tarafta üye işyeri ve diğer tarafta ise kart sahibi vardır. Banka tarafından verilen kredi kartının ön tarafında bankanın logosu ve isminin yanı sıra kart hamilinin adı soyadı, kartın son kullanma tarihi, yeni uygulanmaya başlanan chipli sistemin chip cihazı ve sadece o karta ait olan 16 rakamlı kart numarası bulunur. Arka yüzünde ise bilgileri saklayan manyetik şerit, imza bandı ve bu bandın üzerinde ikincil kart numarası bulunur. Bir kart sahibi üye işyerinden alışveriş yapmak istediği zaman kartı üye işyerine banka tarafından tahsis edilen ve POS makinesi denilen bir cihazdan geçirilir. Kartın sahibine ait bilgiler, kart numaraları ve diğer güvenlik numaraları manyetik şeritte ve kredi kartında bulunan chip'te saklanmaktadır. Kartın POS makinesinden geçirilmesi sırasında okunan bu bilgiler özel telefon hatları aracılığıyla banka kredi veritabanında kart sahibini sorgular. Kart sahibinin bilgileri ile banka veritabanındaki bilgiler uyarsa ve kart sahibinin kredi limiti yeterliyse, kart sistemi otorizasyon yani satışı tamamlamaya yetki verir. POS makinesinden satışın yapıldığını belgeleyen bir slip çıkar. Aslında otorizasyon

işleminden sonra satış tamamlanmış ve satış için gerekli tutar kart sahibinin kredi hesabından düşülmüştür. Fakat daha sonra ortaya çıkacak uyuşmazlıkları ortadan kaldırmak ve gerektiğinde imza incelenmesi yapmak üzere bu imzalı slipler üye işyeri tarafından saklanır (Ahi, 2006).

Banka tarafını açacak olursak; kartı çıkaran banka, pos sahibi banka ve kredi kartı organizasyonu olmak üzere farklı birimler bulunmaktadır. Ülkemizde kredi kartı organizasyonu işlevini Bankalararası Kart Merkezi (BKM) gerçekleştirmektedir. Bankalararası Kart Merkezi, kartlı ödeme sistemi içerisinde ortak sorunlara çözüm bulmak, ülkemizdeki banka ve kredi kartları kural ve standartlarını geliştirmek amacıyla 1990 yılında, 13 kamu ve özel Türk bankasının ortaklığı ile kurulan bir organizasyondur (Bkm, 2008).

Şekil 2.1'de görülen kredi kartı otorizasyon sürecinde, kredi kartı üye işyeri POS cihazından geçirildikten sonra 2 türlü işlem gerçekleşebilir. Kredi kartını çıkaran banka kendi POS cihazından çekim yapıyorsa kredi kartı organizasyonuna (BKM) uğramadan işlemler kontrol edilir ve onay verilir. Bu tip işlemlere OnUs işlemler denir. Diğer şekilde kredi kartını çıkaran banka ile POS cihazı sahibi farklı bankalar ise çekilen tutar kredi kartı organizasyonunu (BKM) üzerinden geçerek kartın sahibi olduğu bankanın veritabanından sorgulandıktan sonra aynı yolu izleyerek işlem tamamlanır. Bu tip işlemlere ise notOnUs işlemler denir.

Günümüzde insanların nakit para taşıma yerine kredi kartına yönelmesi kredi kartı sayısının her geçen yıl daha da artmasına sebep olmaktadır. Kredi kartı sayısındaki bu artış, beraberinde işlem adetlerinin ve işlem tutarlarının da artmasına sebep olmaktadır. Ülkemizdeki 2001 – 2007 yılları arasında kullanılan kredi kartı sayısı, işlem adetleri ve işlem tutarları Tablo 2.1'de verilmektedir.

Tablo 2.1 2001–2007 Tarihleri Arasında Türkiye'de Kullanılan Kredi Kartı (Visa, Mastercard ve Diğer) Bilgileri (Bkm, 2008)

Yıl	Kredi Kartı Sayısı	İşlem Adetleri (Milyon)	İşlem Tutarı (Milyon YTL)
2001	13.996.806	512	15.128
2002	15.705.370	639	25.661
2003	19.863.167	833	40.336
2004	26.681.128	1.136	65.688
2005	29.978.243	1.302	86.494
2006	32.433.333	1.333	109.159
2007	37.335.179	1.441	142.787

Son yıllardaki rakamları yorumlayacak olursak; Türkiye'deki kredi kartları 2006 yılı sonu itibariyle bir önceki yıla göre % 8 artarak 32.4 milyona ulaşmıştır. Avrupa sıralamasında kart adedi ile 3. olan Türkiye son derece önemli bir kart pazarıdır (Bkm, 2008). 2007 yıl sonu itibariyle ise bir önceki yıla göre % 15 artarak 37.3 milyona ulaşmıştır. Kredi kartı kullanımının artması kayıtdışı ekonomi ile mücadele anlamında da önemli katkı sağlamaktadır (Bkm, 2008).

Kredi kartı sayısındaki bu artışın kredi kartını çıkaran, kullandıran ve kullanan lehine olumlu yanları olduğu kadar olumsuz yanları da olmaktadır. Bu artışa paralel olarak işlem adetleri ve işlem tutarları da artmakta ve buna bağlı olarak sahtecilik rakamlarında da gözle görülür bir artış meydana gelmektedir. 2004 sonu itibariyle sahteciliğin toplam harcamalar içindeki payının dünya genelinde onbinde 7.3 olduğunu, bu rakamın Avrupa'da onbinde 6.7, Türkiye'de ise onbinde 4 olduğunu belirtilmektedir (Bkm, 2008).

Tablo 2.1'de görüldüğü gibi bu kadar çok kredi kartı işleminin gerçekleştiği, harcamanın yapıldığı bir ortam beraberinde bazı güvenlik sorunlarını meydana getirmekte ve kredi kartı sahteciliğinin meydana gelmesine yol açmaktadır. Bu probleme karşı kurumların kendi geliştirdikleri önleme mekanizmalarının yanında fiziksel olarak kredi kartlarının yapılarında da geliştirmeler yapılmaktadır. Son geliştirilen chip&pin sistemi ile birlikte sahtecilik rakamlarının gözle görülür derecede düştüğü görülmektedir. Ancak kredi kartlarında bir yandan sahtekarlık ve

dolandırıcılık işlemlerine karşı teknik önlemler alınırken kartın kullanım alanlarının artmasından dolayı sahtecilik tehlikesi devam etmektedir.

Kartlı ödeme sistemlerinde en riskli grubu oluşturan kredi kartları sahteciliği ülkemizde de rastlanan bir sahtecilik türüdür. Kredi kartı sahteciliğinin önüne geçebilmek ancak kart sahiplerinin alacakları tedbirler ile mümkün olabilir. Kart sahiplerinin alması gereken tedbirler kısaca şunlardır; kredi kartı numarasının başkasının eline geçmesini önlemek, müracaat sonrası banka tarafından gönderilen kartın size teslim edildiğinden emin olmak, kartı kullanırken satıcının yaptığı işlemlere dikkat etmek, internet üzerinden yapılan alışverişlerde internet sitesinin güvenliğine dikkat etmek ve kullanılan bilgisayarın güvenliğinden emin olmak gibi biz kart kullanıcılarının dikkat etmesi gereken önemli hususlardır (Yetgin, 2008).

Kredi kartları yaygınlığı ve güçlü altyapısı sayesinde gerçek hayatta olduğu gibi sanal dünyada da en çok tercih edilen ödeme aracı olmuştur. Telefon/Mail Order ve internet üzerinden sipariş gibi fiziki olarak kredi kartının pos cihazından geçmesini gerektirmeyen işlemlerde kredi kart numarası ve kartın son kullanma tarihi bilgilerinin alışveriş için yeterli olması kredi kartının sahibinden başka kimseler tarafından da sahtecilik amaçlı olarak kolayca kullanabilmesine yol açmaktadır (Garanti, 2008)

Tablo 2.2 Kredi Kartı Sahtecilik Metodları ve Gerçekleşen Yüzdeleri (Bhatla ve diğ., 2003)

Metod	**Yüzde**
Kayıp veya Çalıntı Kart	% 48
Kimlik Hırsızlığı	% 15
Kartın Klonlanması	% 14
Sahte Kart	%12
Posta Sahteciliği	% 6
Diğer	% 5

Küresel olarak sahteciliğin hangi oranlarda gerçekleştiği Tablo 2.2'de gösterilmektedir. En çok gerçekleşen sahtecilik tipi kayıp ve çalıntı kartlar iken kimlik hırsızlığı, kartın kopyalanması, sahte kartlar, posta sahteciliği ve diğerleri tabloyu tamamlayan öğelerdir. Bunların yanında kredi kartı sahteciliği alanında % 19

ile Ukrayna başı çekerken Endonezya % 18.3 ile Ukrayna'yı takip etmektedir (Bhatla ve diğ., 2003).

2.2 Kredi Kartı Sahtecilik Çeşitleri

Bu bölümde kredi kartı sahtecilik çeşitleri üzerinde durup, birçok sahtecilik yöntemini anlatarak, sahteciliğin nasıl işlendiğini ve önlemek için ne gibi tedbirler alınması gerektiğinden de bahsedeceğiz. Sahtecilik çeşitleri olarak kayıp kartlar, çalıntı kartlar, sahte başvuru, sahte basılmış kartlar, internet ve internetten alışveriş sahteciliği ve kartın sahibine ulaşım gibi yöntemlerle kredi kartı kayıtları elde edilmektedir (Srivastava ve diğ., 2008). Michael Alliston'a göre ise sahtecilik çeşitleri; sahte kart, fiziksel olarak kartın olmadığı (mail, telefon ve internet) durumlarda sahtecilik, hesap oluşturma, kartı kötüye kullanma ve bilindik diğer sahtecilik tipleri sayabiliriz (Alliston, 2002). Yapılan diğer bir araştırmaya göre sahtecilik çeşitleri; kayıp çalıntı kart, sahte kart, internetten yapılan alışverişler, kartın usulsüz kullanımı, kartla yapılan harcama şekilleri ve kartın sahibine ulaştırılması sırasında meydana gelen yöntemler olarak sınıflandırılmıştır (Ahi, 2006).

2.2.1 Kayıp / Çalıntı Kart Yöntemi

Kimlik ve imza kontrolü yapılmayan iş yerlerinde yapılan alışverişlerde kayıp ya da çalıntı kartların kullanılması şeklinde yapılan dolandırıcılık türüdür. Bu yöntemde kullanılan kredi kartları gerçektir. Kartın sahibinden hırsızlık, yankesicilik veya gasp şeklinde elde edilerek usulsüzce kullanılmasıyla meydana gelen bir yöntemdir. Kart güvenliği zayıf olan işyerlerinde direkt kullanıldığı gibi güvenliği iyi olan işyerlerinde sahte kimliklerle kullanılmaktadır (Ahi, 2006).

Bu kapsama giren bir diğer yöntem; kredi kartı başvurusu yapıldıktan sonra, kredi kartının kurye veya dağıtıcılar vasıtası ile kartı teslim edecekleri doğru kişiye ulaşamayarak kartın yanlış kişilere teslim edilmesi, kurye veya dağıtıcıların kartın sahibi olduğu kişiye başka bir ürün getirmiş gibi kişinin özel bilgilerini ele geçirerek, ele geçirdiği bu bilgiler ile sahte işlem gerçekleştirmeleridir (Ahi, 2006).

Kredi kartı sahtecilik tespiti için kullanılan saklı markov yönteminde sistemin eğitimi için kullanılan kredi kartı kayıtları, kayıp çalıntı kart yöntemi ile elde edilen veriler kullanılarak yapılmıştır (Srivastava ve diğ., 2008). Sahtecilik tespitinin bir başka yöntemi olan harici kart sahteciliğinde kayıp ve çalıntı kart yöntemi kullanılarak nakite kolayca dönüştürülebilen eşyalar alınmıştır (Shen ve diğ., 2007).

2.2.2 Kredi Kartını Usulsüz Kullanma / Kullandırma Yöntemi

Bu yöntemde sahtecilik işlemi kart sahibi tarafından yapıldığından diğer yöntemlere göre farklı bir sahtecilik çeşididir. Kart sahibi kendine ait kredi kartıyla alışveriş yaptıktan sonra bankayı arayıp kartını düşürdüğünü veya çaldırdığını beyan ederek işlemleri geçersiz kılmaya çalışmaktadır. Aynı şekilde kart sahibi kendi kartını başka kişilere kullandırtabilmektedir. Bu şekilde kredi kartıyla kendi işlem yapmamış gibi gösterip bankaya işlemlerin sahte olduğunu iddia ederek mağduriyetinin giderilmesini talep eder (Ahi, 2006).

Sahte harcama belgesi oluşturma kredi kartının usulsüz kullanım alanına giren bir diğer sahtecilik yöntemidir. Bu yöntemde POS makineleri çok yaygın değilken, basit mekanik bir alet yardımıyla kredi kartı ile alışveriş yapılabiliyordu ve bu belgelerin kopyasını oluşturmak suretiyle sahte harcama belgeleri oluşturarak, tahsil yoluna gidilmekteydi. Ancak POS makinelerinin yaygınlaşması sonrasında bu sahtecilik yöntemi ortadan kaybolmuştur (Ahi, 2006).

Kredi kartının bir diğer usulsüz kullanım alanı sahte kart olarak kullanılmasıdır. Bu yöntemde gerçek kredi kartının varlığına ihtiyaç yoktur. Gerçek kredi kartı bilgileri bir şekilde elde edilerek sahte bir kartın üzerine kopyalanmakta ve daha sonra sahte kart ile işlem yapılmaktadır. Son yıllarda geliştirilen ve günümüzde kullanılan chip&pin uygulaması ile bu tip sahteciliğin önüne geçmeye çalışılmaktadır (Ahi, 2006).

Sahte kartın kullanıldığı bir diğer sahtecilik yöntemi üye işyeri ile işbirliği yöntemidir. Üye işyerleri, kredi kartını sağlayan banka ile ortak çalışan ve kart sahibinin kartla alışveriş yapılmasına olanak tanıyan işyerleridir. Sahte kart üye işyerinin POS makinesinden geçirilerek aslında olmayan sahte bir işlem yapılmış

gibi gösterilmektedir ve bu şekilde sahtecilik gerçekleştirilmektedir. Bu çeşit sahteciliğin önüne geçebilmek için bankalar kredi risk izleme ve kredi risk takip birimlerini oluşturmuş, hesapları titizlikle inceleyerek sahteciliği önlenmeye çalışmaktadırlar (Ahi, 2006).

Günümüzde oldukça sık kullanılan sahtecilik yöntemlerinden biri de kredi kartlarında manyetik şerit sahteciliği yöntemidir. Bu yöntem, kredi kartının arkasındaki manyetik şerit bilgileri silinerek bir cihaz yardımıyla başkasına ait kredi kartı bilgilerini yüklenmek suretiyle gerçekleştirilen sahtecilik yöntemidir. Bu tip sahtecilik yönteminin önüne geçebilmek için etkin teknikler geliştirilmektedir (Ahi, 2006).

Kartın usulsüz kullanımı kapsamına giren bir diğer sahtecilik yöntemi yapılan bir işlemin iptal edilmesi gibi gösterilerek aslında hiç yapılmayan bir alışveriş sonrasında para iadesi alınması şeklinde gerçekleşen bir sahtecilik yöntemidir. Bu yöntemde sahtecilik işyeri ve çalışanları tarafından işlenmektedir (Ahi, 2006).

Kartların usulsüz olarak kullanılması yada sahibinin bilgisi olmaksızın kart bilgilerinin kullanılması sonucu elde edilen kart verileri kriminal bir aldatma yöntemidir. Sistem genellikle kredi kartının usulsüz kullanılması yoluyla elde edilen kart bilgileri ile analizleri gerçekleştirmektedir (Quah ve diğ., 2007). Sahte kart, klonlama işlemi sonrasında meydana gelmiş olabilir. Klonlama işleminin önüne geçmek için kullanılan metodta kart numaralarının yerine alfanümerik kararkterler kullanılmıştır ve kart bilgileri saklanmamaktadır (Kou ve diğ., 2004).

2.2.3 Sahte Başvuru Yöntemi

Kredi kartını çıkaran kuruluşların sokaklarda stand açtığı dönemlerde çok sık kullanılmış olan bir yöntemdir. Bu sahtecilik yönteminde, kişinin başkasının kimliği üzerinden kendi fotoğrafı ve sahte belgeleri gibi bilgilerle kredi kartı başvurusu yapma işlemidir (Ahi, 2006).

Burada verilen belgelerin elde edilmesi işlemi olayın bir başka sahtecilik boyutudur. Çünkü bu belgeler kişi tarafından bir amaç için yasal olarak başka bir

kuruma verilirken diğer yandan belgelerin elde edilerek kötü amaçlar için kullanılması söz konusudur. Kart başvurusu sonucunda yanlış beyan edilen adrese veya kişiye gönderilen kredi kartları rahatlıkla şüphe çekmeden kullanılabilmektedir.

2.2.4 İnternetten Alışveriş Yöntemi

İnternetten alışveriş son yıllarda gittikçe artan ve popüler hale gelen bir alışveriş şeklidir. Günümüzde bankaların verdiği kredi kartlarını mağazalarda, restoranlarda kullanabildiğimiz gibi internet üzerinden alışveriş yapmak için kullanabiliyoruz. Bu genel bilgiden sonra bahsetmek istediğimiz sahtecilik yöntemi internetten yapılan işlemlerde kullandığımız kredi kartı ile ilgilidir. Bu yöntemde internetten herhangi bir mal veya hizmet satın almak için kredi kartı numarası, kartın geçerlilik tarihi, kart sahibinin adı ve imza bandında bulunan güvenlik numarasının son 3 rakamı dışında başka bilgiye ihtiyaç yoktur. Bu sebeple çok daha kolay işlenebilen bir dolandırıcılık türüdür (Ahi, 2006).

Bu tip işlemleri önlemek için bankaların ilgili birimlerinin internetten yapılan işlemlerde çeşitli kontroller uygulamaları ve bankaların internetten yapılan işlemlere karşı yeni uygulamalar geliştirerek bu tip sahteciliği bertaraf etmeyi hedeflemişlerdir.

Ödeme sistemlerini e-ticaret sitesi gibi oluşturarak kredi kartı kullanımıyla sahte kayıtları meydana gelebilir. Güvenilir email sunucu çözümü ile sahtecilik tehlikesi minimuma indirilmiştir (Alfuraih ve diğ., 2002).

2.2.5 ATM Dolandırıcılığı Yöntemi

Bu yöntem de günümüzde oldukça yaygın kullanılan sahtecilik yöntemlerindendir. Bankanın ATM cihazının kart giriş haznesine bir cisim yerleştirildikten sonra gelen müşterinin kartının cihaza sıkışması sağlanmaktadır. Daha sonra dolandırıcı yardım etmek bahanesiyle kart hamilinin şifresini öğrenip müşterinin bankadan ayrılmasından sonra kartı ATM cihazından çıkarıp kullanması ile gerçekleşen sahtecilik yöntemidir (Ahi, 2006).

Bu sahtecilik yönteminin bir başka şeklinde ise dolandırıcı gelen müşterinin kartının ATM cihazına sıkışmasını sağlamaktadır. Cihazın görünmeyen bir tarafına yerleştirilen kamera sayesinde kart hamilinin bilgilerini öğrendikten sonra aynı sahtecilik işlemi gerçekleştirilmektedir. Bankalar bu yönteme karşı ATM cihazlarında önlemler almakta, kart giriş haznelerinde özel aparatlar kullanarak bu sahtecilik yöntemini önlemeye çalışmaktadırlar.

2.3 Kredi Kartı Bilgisi Elde Etmede Sahtecilik Yöntemleri

Buraya kadar kredi kartı dolandırıcılarının kullandığı kredi kartı sahtecilik yöntemleri ele alınmıştır. Bunun dışında önemli bir konuda kredi kartı bilgilerinin nasıl ele geçirildiğidir. Kart bilgisi elde etmede e-posta yöntemi, veri transfer yöntemi, kartın kopyalanması ve gizli kamera yöntemi sayılmaktadır (Ahi, 2006). Bir diğer araştırmaya göre kartın klonlanması, kartın çalınması, bilgilerin bir şekilde duyulması, kredi kartının kayıt bilgilerinin ele geçirilmesi ve kart hamilinden kaynaklanan nedenler bu yöntemler arasında sayılırlar (Hsu ve diğ., 2007). Aşağıda kredi kartı bilgilerinin nasıl ele geçirildiği hakkında en çok kullanılan yöntemleri inceleyeceğiz.

2.3.1 Sahte E-Posta Yöntemi

En çok kullanılan kredi kartı bilgisi elde etme yöntemidir. Kullanıcıya bankadan veya kredi kartı kuruluşundan elektronik posta yoluyla gelen bazı duyurularda müşteri bilgilerinin güncellenmesi gerektiği aksi takdirde kartının kullanıma kapanacağı yazılmaktadır. Bu şekilde kredi kartı ile ilgili bütün gizli bilgiler üçüncü kişiler tarafından ele geçirilmektedir. Bu yöntemle elde edilen bilgiler ise sadece kredi kartı sahteciliğinde değil aynı zamanda internet bankacılığı dolandırıcılığında da kullanılmaktadır (Ahi, 2006).

Finans kurumları bu yönteme karşılık her fırsatta müşterilerini uyarmakta ve özellikle müşterilerin bu gibi mailleri dikkate almamaları gerekmektedir. Çünkü hiçbir finans kurumu müşterilerinin bilgilerini mail yoluyla güncellemelerini talep etmemektedir.

2.3.2 Kart Bilgilerini Kopyalama Yöntemi

Kredi kartının okuyucu denilen ve suç jargonunda papağan diye anılan kibrit kutusu büyüklüğünde bir cihazdan geçirilerek, kartın içindeki tüm verilerin kopyalanmak suretiyle sahte kredi kartı üreticilerinin eline geçmesidir. Özellikle market, restoran ve cafelerde hesap ödenirken kredi kartı ödemeyi alacak kişiye teslim edilir. Böylelikle bu tip bir sahteciliğin gerçekleşmesine davetiye çıkarmış olmakla beraber kartın güvenlik durumu diye bir şey ortadan kalkmaktadır. Bu esnada gerçekleşen bir kart bilgisi elde etme yöntemidir (Ahi, 2006).

Bu yöntem özellikle teknoloji destekli yapılan sahtecilik tiplerinden olup bir anlamda kartın klonlanmasıdır. Klonlama, sahte kart bilgisi elde etme yöntemleri içerisinde en çok işlenen üçüncü yöntem olması, sahtecilik açısından önemli bir yere sahip olmakla beraber günümüzde de bu konuda büyük sıkıntılar yaşanmaktadır. Bu sahtecilik yönteminin önüne geçebilmek yapılmış olan çalışmalarda kart sahibini koruyan ekstra güvenlik uygulamaları geliştirilmektedir.

2.3.3 Uzak Bilgisayardan Haklama Yöntemi

Bu yöntem genellikle uzaktaki bir sisteme saldırarak meydana gelen bir kredi kartı bilgisi elde etme yöntemidir. Günümüzde bilgisayarlara saldırma olayı kredi kartı bilgisi elde etmenin dışından çok popüler bir hal almıştır. Bu yöntemin meydana gelmesi özellikle GSM operatörlerinin, mağazaların ve büyük süpermarketlerin tüketicinin alışveriş alışkanlıklarını belirlemek ve buna göre hizmet üretmek amacıyla sakladıkları kredi kartı bilgilerinin, kötü niyetli üçüncü kişilerin eline geçmesiyle ortaya çıkan sakıncalı bir yöntemdir (Ahi, 2006).

Bu gibi kart bilgisi elde etme yöntemlerinin önüne geçebilmek için kartı çıkaran kuruluşların sistemin güvenliğini sağlamaları gerekmektedir. Bu yöntem kapsamında bilgisayara karşı yapılan saldırılar önemli bir konu olarak karşımıza çıkmaktadır. Saldırıların önlenmesi aşamasında karşımıza yapay sinir ağları, model tabanlı nedenleme, veri madenciliği, durum geçiş analizleri ve genetik algoritma teknikleri çıkmaktadır (Kou ve diğ., 2004). Bunun yanında telekomünikasyon ağları alanında gerçekleştirilen sahteciliklerde bu yöntemin kapsamına girmektedir. Küresel

telekomünikasyon sahteciliği ile her yıl milyarlarca dolar kaybedildiği tahmin edilmektedir. Telekomünikasyon ve bilgisayar saldırı alanlarındaki sahtecilik tespit ve önleme sistemleri için istatistik ve makine öğrenimi tekniklerini incelenmiştir.

2.3.4 Yerel Bilgisayardan Veri Transferi Yöntemi

Bu yöntemde kart bilgisi elde etme işlemi sistemin başında olan kişilerin kötü niyeti sonucunda meydana gelen güvenlik zaafından kaynaklanır. Kart bilgilerini saklayan mağazalarda çalışan kişilerin, kartı çıkaran kuruluşlarda çalışan personelin kendisine emanet edilen sistemi kötüye kullanarak kredi kart verilerini kopyalama yöntemidir (Ahi, 2006).

Bu yönteme karşı kurumlar kendi güvenlik mekanizmalarını geliştirerek etkili bir savunma sağlamak zorundadırlar. Zira bu yöntem için yapılacak veya kurgulanacak pek etkin bir koruma yolu bulunmamaktadır. Sonuçta sistemi kullanan sistemin içinde olan kişidir.

3. SAHTECİLİKTE KULLANILAN TEKNİKLER

Sahteciliğin hızla artmasından ötürü her yıl milyarlarca doların kaybedilmesi sonucunda bu artışı önlemek için birçok modern teknik geliştirilmekte ve birçok iş alanında uygulanmaktadır. Modern teknoloji ve küresel iletişimin hızlı gelişimi ile birlikte ortaya çıkan sahtecilik iş dünyasında da önemli bir kayba yol açmaktadır. Sahtecilik tespiti kullanıcı topluluklarının davranışlarını gözlemlemeyi ve tespit etmeyi gerektirken bu davranışlar uzun dönemler boyunca işlenen suçlar, sahtecilik, izinsiz saldırı ve hesap borçlarıdır. Sahtecilik aktiviteleri günlük hayatta kullandığımız birkaç alandan meydana gelmektedir. Bu alanları iletişim ağları, mobil iletişim, internet bankacılığı ve e-ticaret sistemleri olarak gösterebiliriz. Sayılan bu kanallar içerisinde kredi kartı kullanımının etkisi çok fazladır.

İstatistik ve makine öğrenimi sahtecilik tespiti için geçerli teknolojiler sunmakla beraber kara para aklama, e-ticaret, kredi kartı sahteciliği, telekomünikasyon sahteciliği ve bilgisayar saldırıları gibi konulara karşı başarıyla uygulanmaktadırlar. Birçok istatistik yöntemleri kredi kartı sahtecilik tespiti için çok geniş veri setlerine ihtiyaç duyarlar. Örneğin Barclaycard İngiltere'de tek başına 350 milyon kredi kartı kaydı üretmektedir. İstatistiksel analizler sonucunda çıkan sonuçlar sahtecilik tespiti için geçerli sonuçlardır (Bolton ve diğ., 2002). Bir başka kredi kartı sahtecilik tespit yöntemi ise hesaplama zekası kullanarak gerçek zamanlı sahtecilik tespiti yapan modeldir. Bu model, harcama desenlerini araştırarak gerçekleşebilecek sahtecilik durumlarına karşı yeni ve inovatif yaklaşımlar geliştirip, bu harcama desenlerini deşifre etmeyi amaçlamıştır. Sahteciliği tespit etmek için kendi kendini tertipleyen yöntemler (Self-Organizing Map) kullanılarak, müşteri davranışları yorumlanmış, süzülmüş ve analiz edilmiştir. Bu yöntemde kullanılan sinir ağları danışmasız öğrenme yoluyla oluşturulmuştur (Quah ve diğ., 2007).

Kompleks sahtecilik tespit görevlerin arkaplanlarını tanımlayan, geniş ve gürültülü gerçek dünya test setlerinin büyük oranda başarıyla çözümleyen uyarlanabilir durum tabanlı çıkarsama yönteminin geliştirilmesi farklı sahtecilik tespit yöntemlerinden biridir (Wheeler ve diğ., 2000).

Veri madenciliği, yapay zeka ve istatistiksel teknikler sahtecilik tespiti çalışmalarında kullanılan tekniklerdir (Kou ve diğ., 2004). Çalışmamızda sahtecilik tespiti için kullanılan tüm bu tekniklerden detaylı olarak bahsedeceğiz.

3.1. Yapay Zeka

İnsan beyni dünyanın en karmaşık makinesi olarak kabul edilebilir. İnsan beyni sayısal birkaç işlemi birkaç dakikada yapılabilmesine karşın, idrak etmeye yönelik olayları çok kısa bir sürede yapar. Örneğin yolda giden bir şoför yolun kayganlık derecesini, önündeki tehlikeden ne kadar uzak olduğunu, sayısal olarak değerlendiremezse dahi geçmişte kazanmış olduğu tecrübeler sayesinde aracın hızını azaltır. Çünkü o saniyelerle ölçülebilecek kadar kısa bir sürede tehlikeyi idrak etmiş ve ona karşı koyma gibi bir tepki vermiştir. Bu noktada akla gelen ilk soru şu olmaktadır: Acaba bir bilgisayar yardımı ile böyle bir zeka üretmek münkün olabilir mi? Bilgisayarlar çok karmaşık sayısal işlemleri anında çözümleyebilmelerine karşın, idrak etme ve deneyimlerle kazanılmış bilgileri kullanabilme noktasında çok yetersizdirler. Bu olayda insanı ya da insan beynini üstün kılan temel özellik, sinirsel algılayıcılar vasıtası ile kazanılmış ve göreli olarak sınıflandırılmış bilgileri kullanabilmesidir. **Uzman Sistemler** (US), **Bulanık Mantık** (BM), **Genetik Algoritma** (GA) ve **Yapay Sinir Ağları** (YSA) gibi yapay zeka alt dalları özellikle son yıllarda geniş bir araştırma ve uygulama alanı bulmaktadırlar (Elmas, 2007).

Uzman Sistemler, temelde insan düşüncelerini gerçekleştirmek amacıyla bilgisayarlar tarafından işlenen bir yazılımdır. Uzman Sistem geliştirilirken, uzmanların belli bir konudaki bilgi ve deneyimlerini bilgisayara aktarılması amaçlanmaktadır (Elmas, 2007).

Bulanık Mantık, bulanık küme teorisine dayanan bir matematiksel disiplindir. Bulanık mantık insan mantığında olduğu gibi, Uzun-Kısa, Sıcak-Soğuk, Hızlı-Yavaş, Siyah-Beyaz yerine Uzun-Ortadan Uzun-Orta-Ortadan Kısa-Kısa, Sıcak-Ilık-Az Soğuk-Soğuk-Çok Soğuk vb. gibi ara değerlere göre çalışmaktadır (Elmas, 2007).

Genetik algoritmalar yapay zekanın gittikçe genişleyen bir kolu olan evrimsel hesaplama tekniğinin bir parçasını oluşturmaktadır. Genetik algoritma Darwin'in

evrim kuramı doğada en iyinin yaşaması kuralından esinlenerek oluşturulan, bir veri öbeğinden özel bir veriyi bulmak için kullanılan bir arama yöntemidir. Genetik algoritma geleneksel yöntemlerle çözümü zor veya imkansız olan problemlerin çözümünde kullanılmaktadır (Elmas, 2007). Genetik algoritma kötü niyetli saldırıları normal durumlardan ayırmayı tespit etmekte kullanılır. Genetik algoritma mümkün olan herbir davranışsal modeli tasvir etmek için tasarlanmıştır. Bu yaklaşım yüksek tespit oranı ve düşük alarm oranı sağlar. Dokas ve Ertoz (Kou ve diğ., 2004) bilinen saldırıları tanımlamak için tahmin eden modeller inşaa edip sunmaktadırlar. Bu metod çarpık sınıf dağılımları ile çalışıldığı zaman standart veri madenciliği tekniklerinin yetersiz kaldığı durumlarda kullanılır (Kou ve diğ., 2004). Genetik algoritma, en iyilerin hayatta kalma prensibinden esinlenen genel bir araştırma ve optimizasyon yöntemidir.

Genel olarak yapay sinir ağları, insan beyninin sinir ağlarını taklit eden bilgisayar programlarıdır. Yapay sinir ağları bir anlamda paralel bilgi işleme sistemi olarak düşünülebilir. Yapay sinir ağlarına bu bilgiler ilgili olaya ait örnekler üzerinde eğitilerek verilir. Böylelikle, örnekler sayesinde açığa çıkarılmış özellikler üzerinde çeşitli genelleştirmeler yapılarak daha sonra ortaya çıkacak ya da o ana kadar hiç rastlanmamış olaylara da çözümler üretilmektedir (Elmas, 2007).

Yapay sinir ağları kullanılarak geliştirilmiş olan birçok sistem ve uygulama bulunmaktadır. Örneğin, Mellon bankasında kurulmuş olup halen kullanılmakta olan sistem sayesinde kaybolan kartlar, çalınmış kartlar, başvuru sahteciliği, sahte basılmış kartlar ve mail order sahteciliği ile elde edilmiş olan veriler, yapay sinir ağları kullanılarak analiz edilmiş ve toplam sahtecilik rakamlarında % 20 ile % 40 arasında azalma elde edilmiştir (Ghosh ve diğ., 1994). Yapılan bir başka araştırmaya göre sinirsel veri madenciliği kullanılarak kredi kartı sahtecilik tespiti yapılmaya çalışılmıştır. Amaç veri madenciliği teknikleri ile sinir ağları algoritmalarını birleştirerek daha düşük alarm oranı elde etmektir yani sahtecilik oranını düşürmektir. Sonuç olarak % 0.1 gibi çok düşük sahtecilik meydana oranı ortaya çıkmıştır (Brause ve diğ., 1999). Kredi kartı sahtecilik tespiti için kullanılan web servis tabanlı ortak projeler sayesinde bir araya getirilmiş birçok finansal kuruluşun sahtecilik şüphesi olan datalarının ortak bir platformda birleştirilerek yeni çözümler elde edilmesi sağlanmıştır. Bu şekilde kuruluşların sahtecilik tespit kabiliyetleri

artırılıp finansal kayıplarının azaltılması hedeflenmiştir. Geliştirilen bu sistem sayesinde, farklı kuruluşlarda ve endüstri dallarında bilginin paylaşım mekanizması daha aktif kullanılacaktır (Chiu ve diğ., 2004). Üç sınıflandırma metodunun kullanıldığı bir başka çalışmada (Yapay Sinir Ağları, Lojistik Regresyon ve Karar Ağaçları) amaç bankalar için anahtar görevlerden biri olan kredi kartı risk izleme sistemini inşaa etmek, çok daha doğru ve verimli analiz yapmayı sağlamaktır. Kullanılan üç yöntemden karar ağaçları, yapay sinir ağları ve lojistik regresyona göre daha iyi performans göstermiştir (Shen ve diğ., 2007). Kredi kartı sahtecilik tespiti için kullanılan yapay sinir ağı tabanlı veritabanı madenciliği sistemlerinden olan Cardwatch çok geniş, çeşitli kurumsal veritabanlarına kolaylıkla ve direkt olarak uygulanmıştır. Sonuç olarak sahtecilik tespit oranı % 85, legal kayıt tanımlama oranı ise % 100 olarak sonuçlandırılmıştır (Aleskerov ve diğ., 1997).

Yapay sinir ağları teknolojisi üzerinde bir diğer kredi kartı sahtecilik tespit sistemi, SOM algoritması kullanılarak tipik bir kart hamili davranışları modeli oluşturulmuş ve kayıtların sapmaları analiz edilerek şüpheli olan kayıtlar ortaya çıkarılmıştır. Modelin başarısı veri dağılımındaki istatistiksel dağılıma bağlı olmayıp gürültülü veri setlerinde de başarıyla uygulanmaktadır (Zaslavsky ve diğ., 2006). Veri madenciliği sınıflandırma tekniklerinin kullanıldığı yöntemde sahte finansal bilaçolarının tespit edilmesi hedeflenmiştir. Karar ağaçları, yapay sinir ağları ve bayes inanç ağlarının kullanıldığı çalışmada eğitim setinde % 100 ile en iyi sonucu yapay sinir ağları verirken, doğrulama aşamasında toplamda % 90.3 doğrulukla bayes inanç ağaçları en iyi performansı göstermiştir (Kirkos ve diğ., 2007).

3.1.1 Yapay Sinir Ağları

İnsanlığın doğayı araştırma ve taklit etme çabalarının en son ürünlerinden bir tanesi, yapay sinir ağları (YSA) teknolojisidir. Yapay sinir ağları basit biyolojik sinir sisteminin çalışma şekli simüle edilerek tasarlanan programlama yaklaşımıdır. Simüle edilen sinir hücreleri nöronlar içerirler ve bu nöronlar çeşitli şekillerde birbirlerine bağlanarak ağı oluştururlar. Bu ağlar öğrenme, hafızaya alma ve veriler arasındaki ilişkiyi ortaya çıkarma kapasitesine sahiptirler. Diğer bir ifadeyle, yapay sinir ağları, normalde bir insanın düşünme ve gözlemlemeye yönelik doğal yeteneklerini gerektiren problemlere çözüm üretmektedir. Bir insanın, düşünme ve

gözlemleme yeteneklerini gerektiren problemlere yönelik çözümler üretebilmesinin temel sebebi ise insan beyninin ve dolayısıyla insanın sahip olduğu yaşayarak veya deneyerek öğrenme yeteneğidir (Wikipedia, 2008).

1980'lerden sonra yaygınlaşan yapay sinir ağlarında amaç fonksiyon birbirine bağlı basit işlemci ünitelerinden oluşan bir ağ üzerine dağıtılmıştır. Yapay sinir ağlarında kullanılan öğrenme algoritmaları veriden üniteler arasındaki bağlantı ağırlıklarını hesaplar. YSA istatistiksel yöntemler gibi veri hakkında parametrik bir model varsaymaz yani uygulama alanı daha geniştir ve bellek tabanlı yöntemler kadar yüksek işlem ve bellek gerektirmez (Akpınar, 2000).

Yapay sinir ağları, her türlü bilgiyi işlemek ya da analiz etmek amacıyla kullanılırlar. İş hayatı, finans, endüstri, eğitim ve karışık problemli bilim alanlarında, bulanık veya var olan basit yöntemlerle çözülemeyen problemlerin çözümünde, doğrusal olmayan sistemlerde başarıyla uygulanmaktadırlar. Yapay sinir ağlarının başlıca uygulama alanları sınıflandırma, tahmin ve modelleme olarak ele alınabilir (Elmas, 2007). Başarılı uygulamalar incelendiğinde yapay sinir ağların çok boyutlu, gürültülü, karmaşık, kesin olmayan, eksik, kusurlu, hata olasılığı yüksek sensör verilerinin olması ve problemi çözmek için matematiksel modelin ve algoritmaların bulunmadığı, sadece örneklerin var olduğu durumlarda yaygın olarak kullanıldıkları görülmektedir.

3.1.1.1 Yapay Sinir Ağlarının Tarihçesi

Yapay sinir ağları ile ilgili çalışmalar 20.yy'ın ilk yarısında başlamış ve günümüze kadar büyük bir hızla devam etmiştir. Bu çalışmaları 1970 öncesi ve sonrası diye iki kısma ayırmak mümkündür. Zira 1970 yılları YSA için bir dönüm noktasını teşkil etmiş daha önce aşılması imkansız görünen pek çok problem bu dönemlerde aşılmıştır (Haykin, 1994).

İnsan beyni hakkındaki çalışmalar binlerce yıl öncesine dayanır. Modern elektroniğin gelişmesi ile birlikte bu düşünce işlemini kullanmaya çalışmak doğal bir hale gelmiştir. İlk yapay sinir ağı modeli 1943 yılında bir sinir hekimi olan Warren McCulloch ile bir matematikçi olan Walter Pitts tarafından gerçekleştirilmiştir.

McCulloch ve Pitts, insan beyninin hesaplama yeteneğinden esinlenerek elektrik devreleriyle basit bir sinir ağını modellemişlerdir (Altıntaş, 2008).

1948 yılında Wiener “Cybernetics” isimli kitabında sinirlerin çalışması ve davranış özelliklerine değinmiş, 1949 yılında ise Hebb “Organization of Behavior” isimli kitabında öğrenme ile ilgili temel teoriyi ele aldı. Hebb kitabında öğrenebilen ve uyum sağlayabilen sinir ağları modeli için temel oluşturacak "Hebb Kuralı" nı ortaya koymuştur. Hebb kuralı; sinir ağının bağlantı sayısı değiştirilebilirse, öğrenebileceğini ön görmekteydi. 1957 yılında Frank Rosenblatt'ın perseptron'u geliştirmesinden sonra, YSA'lar ile ilgili çalışmalar hız kazanmıştır. Perseptron; beyin işlevlerini modelleyebilmek amacıyla yapılan çalışmalar neticesinde ortaya çıkan tek katmanlı eğitilebilen ve tek çıkışa sahip bir ağ modelidir (Altıntaş, 2008).

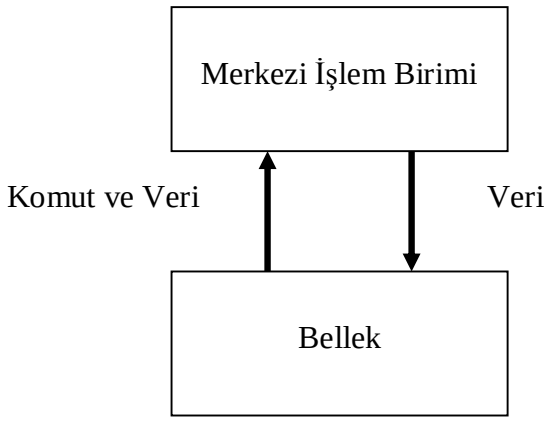
1959 yılında Bernand Widrow ve Marcian Hoff (Stanford Üniversitesi) ADALINE (Adaptive Linear Neuron) modelini geliştirmişler ve bu model YSA'ların mühendislik uygulamaları için başlangıç kabul edilmiştir. Bu model Rosenblatt'ın Perseptron'una benzemekle birlikte, öğrenme algoritması daha gelişmiştir. Bu model uzun mesafelerdeki telefon hatlarındaki yankıları ve gürültüleri yok eden bir adaptif filtre olarak kullanılmış ve gerçek dünya problemlerine uygulanan ilk YSA olma özelliğini kazanmıştır (Altıntaş, 2008). Bu yöntem günümüzde de aynı amaçla kullanılmaktadır. 1960’ların sonlarına doğru YSA çalışmaları durma noktasına gelmiştir. Buna en önemli etki; Minsky ve Pappert tarafından yazılan Perceptrons adlı kitaptır. Burada YSA’ların doğrusal olmayan problemleri çözemediği ispatlanmış, iki katmanlı ileri beslemeli ağların kullanılabileceğini ileri sürmüşler ve bunun tek katmanlı ağlardaki birçok sınırlamayı ortadan kaldırdığını göstermişlerdir. Böylece YSA çalışmaları bıçak gibi kesilmiştir (Yazici, 2008).

Tüm bunlara rağmen genç bilim adamları çalışmalarını sürdürmüşlerdir. 1980’li yıllar sinirsel hesaplama çalışmaları için bir atılım dönemi olmuştur (Elmas, 2007). 1982 – 1984 yıllarında Hopfield tarafından yayınlanan çalışmalar ile YSA’ların genelleştirilebileceği ve çözümü zor problemlere çözüm üretebileceğini göstermiş, ağların önemli sınıflarının matematik temellerini üretmiştir (Yazici, 2008).

1984'te Kohonen sinirlerin düzenli sıralanışına eşleme özelliği için danışmasız öğrenme ağlarını geliştirmiştir. Bu çalışmaların neticesi Hinton ve arkadaşlarının geliştirdiği Boltzman Makinası'nın doğmasına yol açmıştır. 1986 yılında Rumelhart ve McClelland karmaşık ve çok katmanlı ağlar için geriye yayılmalı öğrenme algoritması ortaya koymuştur. 1988 yılında radyal tabanlı fonksiyonlar modeli geliştirilmiş ve özellikle filtreleme konusunda başarılı sonuçlar elde edilmiştir. Daha sonra bu ağların daha gelişmiş şekli olan probabilistik ağlar ve genel regresyon ağları geliştirilmiştir. Bu tarihten günümüze kadar sayısız çalışma ve uygulama geliştirilmiştir (Elmas, 2007).

3.1.1.2 Yapay Sinir Ağlarına Giriş

Sayısal bilgisayarlar 1940'ların sonlarından günümüze değin hızla gelişmiştir. Önceleri matematik hesaplamalarda kullanılan sayısal bilgisayarlar daha sonra metin, sembol, resim ve ses işlemeyi de kapsayan geniş bir uygulama alanı bulmuştur. Şekil 3.1'de bir sayısal bilgisayarın işlevsel özelliğinin şekli görülmektedir. Sayısal bilgisayarlar Von-Neumann mimarı temeli üzerine kurulmuşlardır.



Şekil 3.1 Von-Neumann Makinesi (Elmas, 2007)

Yapay sinir ağları (YSA), insan beyninden esinlenerek geliştirilmiş, ağırlıklı bağlantılar aracılığıyla birbirine bağlanan ve her biri kendi belleğine sahip işlem elemanlarından oluşan paralel ve dağıtılmış bilgi işleme yapılarıdır. Yapay sinir ağları, bir başka değişle biyolojik sinir ağlarını taklit eden bilgisayar programlarıdır. Yapay sinir ağları zaman zaman bağlantıcılık, paralel dağıtılmış işlem, sinirsel-işlem, doğal zeka sistemleri ve makine öğrenme algoritmaları gibi isimlerle de

anılmaktadır. Yapay sinir ağları bir programcının geleneksel yeteneklerini gerektirmeyen, kendi kendine öğrenme düzenekleridir. Bu ağlar öğrenmenin yanı sıra ezberleme ve bilgiler arasında ilişkiler oluşturma yeteneğine de sahiptir.

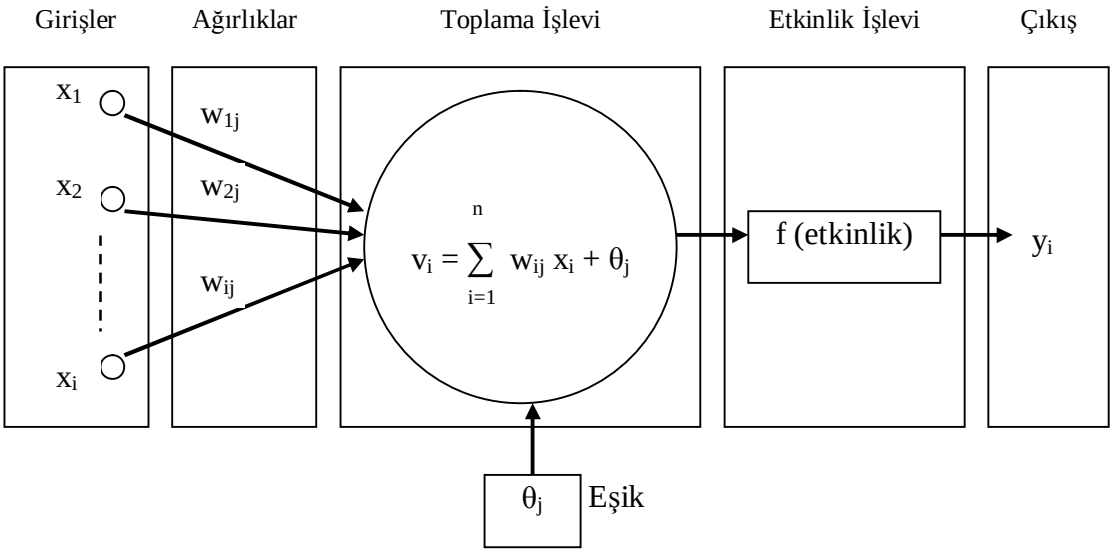
Yapay sinir ağları insan beyninin bazı organizasyon ilkelerine benzeyen özellikleri kullanmaktadırlar. Yapay sinir ağları bilgi işleme sistemlerinin yeni neslini temsil ederler. Genel olarak yapay sinir ağları model seçimi ve sınıflandırılması, işlev tahmini, en uygun değeri bulma ve veri sınıflandırılması gibi işlerde başarılıdır. Geleneksel bilgisayarlar ise özellikle model seçme işinde verimsizdir ve sadece algoritmaya dayalı hesaplama işlemleri ile kesin aritmetik işlemlerde hızlıdırlar (Elmas, 2007).

Birçok yapay sinir ağı tipi bulunmakla birlikte bazılarının kullanımı diğerlerinden daha yaygındır. En çok kullanılan yapay sinir ağı, Geri Yayılımlı Yapay Sinir Ağı olarak bilinendir. Bu tip yapay sinir ağı tahmin ve sınıflandırma işlemlerinde çok iyi sonuçlar vermektedir. Bu tip sinir ağları, karışık bilgi kümeleri arasında ilişki bulma konusunda başarılı sonuçlar vermektedir (Elmas, 2007).

Yapay sinir ağları, uygulanan ağ modeline göre değişik karakteristik özellikler göstermelerine karşın temel birkaç ortak özelliğe sahiptirler. Birinci özellik; yapay sinir ağlarında sistemin paralelliği toplamsal işlevin yapısal olarak dağılmışlığıdır. Yani karmaşık işlemler birçok nöronun eşzamanlı çalışması ile meydana getirilir. İkinci özellik; genelleme yeteneğidir. Yani eğitim sırasında kullanılmayan girdiler içinde anlamlı çıktılar üretebilmesidir. Üçüncü özellik; ağ fonksiyonları non-lineer olabilmektedir. Yani yapı üzerinde işlevin doğru biçimde yerine getirilmesini matematiksel olarak olası kılarlar. Dördüncü özellik; sayısal ortamda tasarlanan yapay sinir ağların donanımsal olarak gerçekleştirebilirlikleridir (Altıntaş, 2008).

Bütün bunlarla beraber, yapay sinir ağları pekçok sektörde değişik uygulama alanları bulmuştur. Bunlardan bazıları; Uzay, Otomotiv, Bankacılık, Savunma, Elektronik, Finans, Sigortacılık, Üretim, Sağlık, Telekomünikasyon, Güvenlik vs. alanlarında kullanılmaktadır (Altıntaş, 2008).

Yapay sinir ağlarının temel birimi işlem elemanı ya da düğüm olarak adlandırılan yapay bir sinirdir. Şekil 3.2'de yapay bir sinir (düğüm) gösterilmiştir.



Şekil 3.2 Yapay Bir Sinir (Düğüm) (Elmas, 2007)

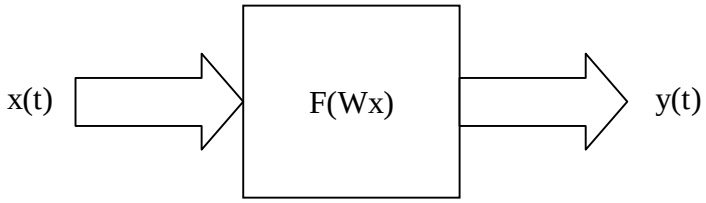
Girişler x_i sembolüyle gösterilmiştir. Bu girişlerin her biri ağırlık w ile çarpılır. Basitçe, bu ürünler eşik değeri θ_j ile toplanır ve sonucu oluşturmak için etkinlik işlevi ile işlem yapılır ve y_i çıkışı alınır. Tüm yapay sinir ağları bu temel yapıdan türetilmiştir. Bu yapıdaki farklılıklar yapay sinir ağlarının farklı sınıflandırılmalarını sağlar. Bir yapay sinirin öğrenme yeteneği, seçilen öğrenme algoritması içerisinde ağırlıkların uygun bir şekilde ayarlanmasına bağlıdır (Elmas, 2007).

3.1.1.3 Yapay Sinir Ağlarının Sınıflandırılması

Yapay sinir ağları, genel olarak birbirleri ile bağlantılı işlemci birimlerden veya diğer bir ifade ile işlemci elemanlardan oluşurlar. Her bir sinir hücresi arasındaki bağlantıların yapısı ağın yapısını belirler. İstenilen hedefe ulaşmak için bağlantıların nasıl değiştirileceği öğrenme algoritması tarafından belirlenir. Yapay sinir ağları yapılarına ve öğrenme algoritmalarına göre sınıflandırılırlar.

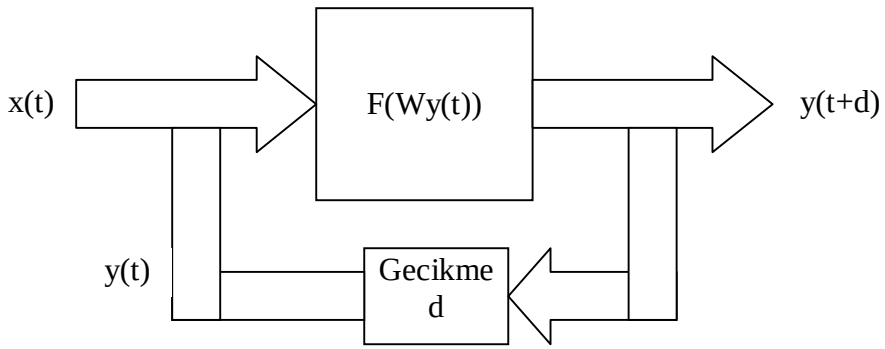
İleri beslemeli bir ağda işlemci elemanlar genellikle katmanlara ayrılmışlardır. İşaretler, giriş katmanından çıkış katmanına doğru tek yönlü

bağlantılarla iletilir. İşlemci elemanları bir katmandan diğer bir katmana bağlantı kurarlarken aynı katman içerisinde bağlantıları bulunmaz. Şekil 3.3'te ileri beslemeli ağ için blok diyagram gösterilmiştir. İleri beslemeli ağlara örnek olarak Çok Katmanlı Perseptron (Multilayer Perseptron-MLP) ve Vektör Kuantalama Öğrenme (Learning Vector Quantization - LVQ) ağları verilebilir (Altıntaş, 2008).



Şekil 3.3 İleri Beslemeli Ağ için Blok Diyagram (Şahin, 2008)

Geri beslemeli sinir ağı, çıkış ve ara katlardaki çıkışların giriş birimlerine veya önceki ara katmanlara geri beslendiği bir ağ yapısıdır. Böylece girişler hem ileri yönde hem de geri yönde aktarılmış olur. Şekil 3.4'te bir geri beslemeli ağ görülmektedir. Bu çeşit sinir ağlarının dinamik hafızaları vardır ve bir andaki çıkış hem o andaki hem de önceki girişleri yansıtır. Bundan dolayı, özellikle önceden tahmin uygulamaları için uygundurlar. Bu ağlar çeşitli tipteki zaman-serilerinin tahmininde oldukça başarı sağlamışlardır. Bu ağlara örnek olarak Hopfield, SOM (Self Organizing Map), Elman ve Jordan ağları verilebilir (Altıntaş, 2008).

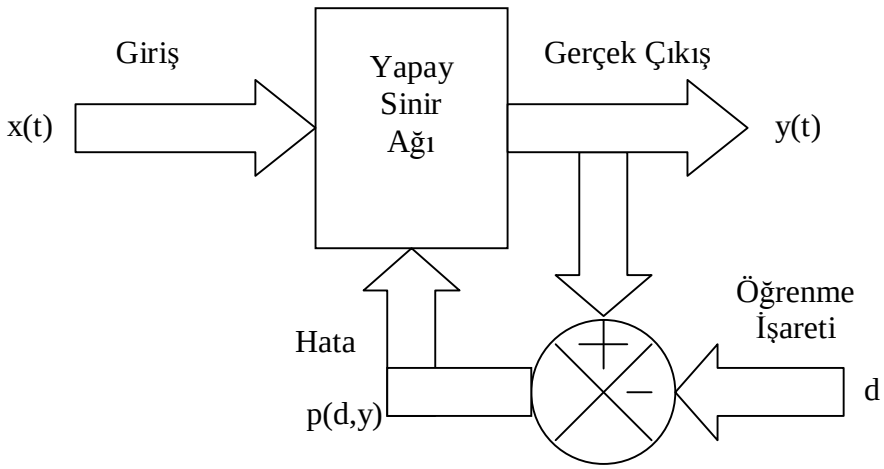


Şekil 3.4 Geri Beslemeli Ağ için Blok Diyagram (Şahin, 2008)

Öğrenme; gözlem, eğitim ve hareketin doğal yapıda meydana getirdiği davranış değişikliği olarak tanımlanmaktadır. O halde, bir takım metot ve kurallar, gözlem ve eğitime göre ağdaki ağırlıkların değiştirilmesi sağlanmalıdır. Bunun için

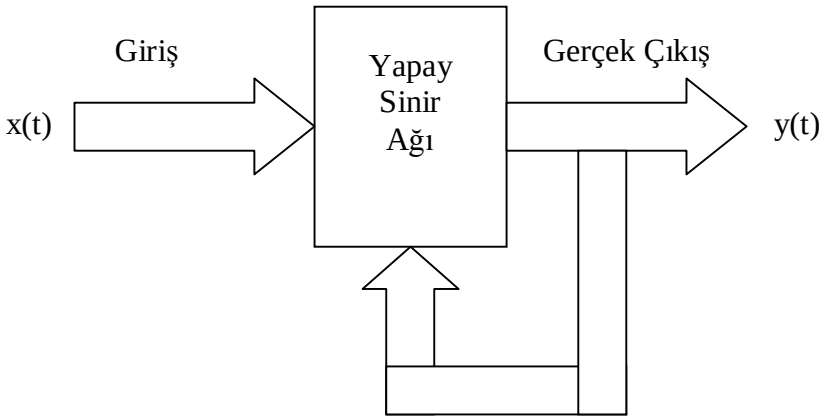
genel olarak üç öğrenme metodundan ve bunların uygulandığı değişik öğrenme kurallarından söz edilebilir.

Bu tip öğrenmede, yapay sinir ağlarına örnek olarak bir doğru çıkış verilir. İstenilen ve gerçek çıktı arasındaki farka (hataya) göre işlemci elemanlar arası bağlantıların ağırlığını en uygun çıkışı elde etmek için sonradan düzenlenebilir. Bu sebeple danışmanlı öğrenme algoritmasının bir "öğretmene" veya "danışmana" ihtiyacı vardır. Şekil 3.5'te danışmanlı öğrenme yapısı gösterilmiştir. Widrow-Hoff tarafından geliştirilen delta kuralı ve Rumelhart ve McClelland tarafından geliştirilen genelleştirilmiş delta kuralı veya geri yayılma algoritması danışmalı öğrenme algoritmalarına örnek olarak verilebilir (Altıntaş, 2008).



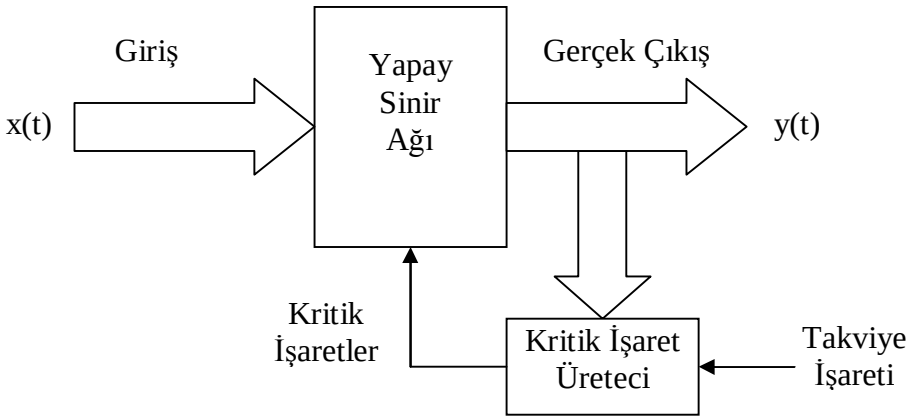
Şekil 3.5 Danışmalı Öğrenme Yapısı (Şahin, 2008)

Girişe verilen örnekten elde edilen çıkış bilgisine göre ağ sınıflandırma kurallarını kendi kendine geliştirmektedir. Bu öğrenme algoritmalarında, istenilen çıkış değerinin bilinmesine gerek yoktur. Öğrenme süresince sadece giriş bilgileri verilir. Ağ daha sonra bağlantı ağırlıklarını aynı özellikleri gösteren desenler oluşturmak üzere ayarlar. Şekil 3.6'da danışmasız öğrenme yapısı gösterilmiştir. Grossberg tarafından geliştirilen ART (Adaptive Resonance Theory) veya Kohonen tarafından geliştirilen SOM (Self Organizing Map) öğrenme kuralı danışmasız öğrenmeye örnek olarak verilebilir (Altıntaş, 2008).



Şekil 3.6 Danışmasız Öğrenme Yapısı (Şahin, 2008)

Bu öğrenme kuralı danışmanlı öğrenmeye yakın bir metoddur. Danışmasız öğrenme algoritması istenilen çıkışın bilinmesine gerek duymaz. Hedef çıktıyı vermek için bir "öğretmen" yerine, burada yapay sinir ağına bir çıkış verilmemekte fakat elde edilen çıkışın verilen girişe karşılık iyiliğini değerlendiren bir kriter kullanılmaktadır. Şekil 3.7'de takviyeli öğrenme yapısı gösterilmiştir. Optimizasyon problemlerini çözmek için Hinton ve Sejnowski'nin geliştirdiği Boltzmann kuralı veya Genetik Algoritma (GA) takviyeli öğrenmeye örnek olarak verilebilir (Altıntaş, 2008).

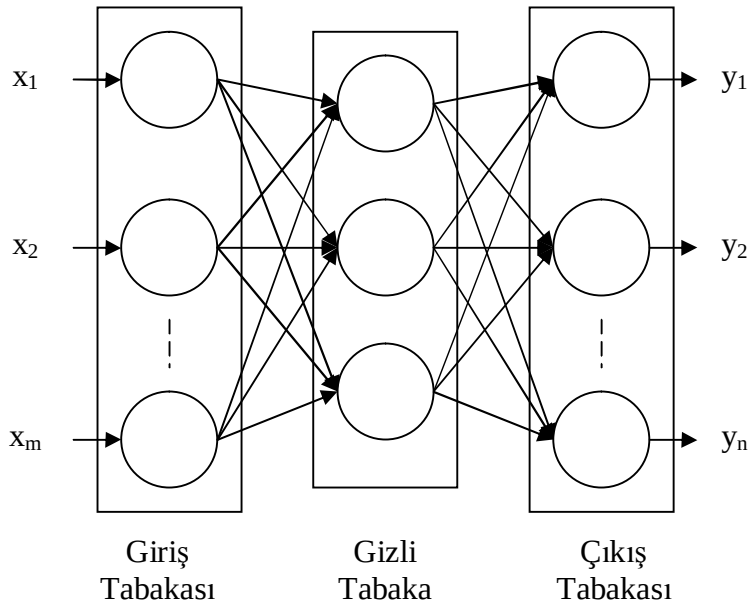


Şekil 3.7 Takviyeli Öğrenme Yapısı (Şahin, 2008)

3.1.1.3 Çok Katmanlı Perseptron Algoritması

Çok Katmanlı Perseptron sinir ağı modeli Şekil 3.8'de gösterilmiştir. Bu ağ modeli özellikle mühendislik uygulamalarında en çok kullanılan sinir ağı modeli olmuştur. Bir çok öğretme algoritmasının bu ağı eğitmede kullanılabilir olması, bu

modelin yaygın kullanılmasının sebebidir. Çok katmanlı perseptron modeli; bir giriş, bir veya daha fazla ara ve bir de çıkış katmanından oluşur. Bir katmandaki bütün işlem elemanları bir üst katmandaki bütün işlem elemanlarına bağlıdır. Bilgi akışı ileri doğru olup geri besleme yoktur. Bunun için ileri beslemeli sinir ağı modeli olarak adlandırılır. Giriş katmanında herhangi bir bilgi işleme yapılmaz. Buradaki işlem elemanı sayısı tamamen uygulanan problemlerin giriş sayısına bağlıdır. Ara katman sayısı ve ara katmanlardaki işlem elemanı sayısı ise deneme-yanılma yolu ile bulunur. Çıkış katmanındaki eleman sayısı ise yine uygulanan probleme dayanılarak belirlenir (Altıntaş, 2008).



Şekil 3.8 Çok Katmanlı Perseptron Yapısı (Şahin, 2008)

Çok katmanlı perseptron ağlarında, ağa bir örnek gösterilir ve örnek neticesinde nasıl bir sonuç üreteceği de bildirilir. Bu danışmanlı öğrenme yapısını göstermektedir. Örnekler giriş katmanına uygulanır, ara katmanlarda işlenir ve çıkış katmanından da çıkışlar elde edilir. Kullanılan eğitme algoritmasına göre ağın çıkışı ile arzu edilen çıkış arasındaki hata tekrar geriye doğru yayılarak hata minimuma düşünceye kadar ağın ağırlıkları değiştirilir (Şahin, 2008).

Çok katmanlı perseptron prensiplerinin kullanılarak oluşturulan kredi puanlama modeli yapay sinir ağları üzerine inşaa edilmiştir. Simülasyon sonuçlarına

göre yapay sinir ağları üzerinde kurulan kredi puanlama modelinin yüksek sınıflama doğruluk oranı ve güçlü yetenekleri olduğunu göstermiştir (Pang ve diğ., 2002).

Çok katmanlı perseptron ağlarında birçok öğrenme algoritması kullanılmaktadır. Geri yayılım algoritması, delta bar delta algoritması, genişletilmiş delta bar delta algoritması, hızlı yayılım algoritması ve genetik algoritma bu öğrenme algoritmalarındandır. Burada geri yayılım algoritmasını detaylı olarak inceleyeceğiz.

Geri yayılım algoritması, karmaşık verilerin sınıflandırılmasında kullanılan etkin yapay sinir ağı modellerinden birisidir. İlk olarak Werbos tarafından düzenlenen daha sonra Parker, Rummelhart ve McClelland tarafından geliştirilen bir geri yayılım ağıdır. İlk uygulamaları yazılı metinden söz sentezi, robot kollarının kontrolüdür. Geri yayılım ağları günümüzde en yaygın kullanılan öğrenimi kolay, sonuçları etkin bir yapay sinir ağı algoritmasıdır.

Yayılma ve uyum gösterme olmak üzere iki aşamada işlemleri gerçekleştiren geri yayılım ağı, katmanlar arasında tam bir bağlantının bulunduğu çok katmanlı, ileri beslemeli ve denetimli olarak eğitilen bir yapay sinir ağı modelidir. Geri yayılım algoritması bir çok uygulamalarda kullanılmış en yaygın öğrenme algoritması olup anlaşılması kolay ve tercih edilen en yaygın öğretme algoritmasıdır. Bu algoritma hataları geriye doğru çıkıştan girişe azaltmaya çalışmasından dolayı geri yayılım ismini almıştır. Geri yayılmalı öğrenme kuralı ağ çıkışındaki mevcut hata düzeyine göre herbir tabakadaki ağırlıkları yeniden hesaplamak için kullanılmaktadır. Bir geri yayılımlı ağ modelinde giriş, gizli ve çıkış olmak üzere 3 katman bulunmakla birlikte problemin özelliklerine göre gizli katman sayısını artırabilmek mümkündür.

Giriş katmanı, veri gruplarının ağa sunulduğu katmandır. Bu katmanda nöron sayısı veri giriş sayısı kadardır ve herbir nöron bir giriş verisi alır. Veri buradan bir sonraki katman olan gizli katmana geçer. Gizli katman temel işlevi gören katmandır. Bazı uygulamalarda ağda birden fazla gizli katman bulunur. Katman sayısı ve katmanlardaki nöron sayısı ağın tasarımcının tecrübesine ve problemin türüne göre değişir. Bu katman giriş katmanından aldığı ağırlıklandırılmış veriyi uygun bir fonksiyonla işleyerek bir sonraki katmana iletir. Çıkış katmanı ise gizli katmandan aldığı veriyi ağın kullandığı fonksiyonla işleyerek çıktısını verir. Bu katmandan elde

edilen değerler ağın bu problem için ürettiği çıkış değerleridir. Geri yayılım ağında bir katmandan başka bir katmana aradaki katmanı atlayarak geçmek mümkün değildir (Şahin, 2007).

$$e_i(n) = d_i(n) - y_i(n) \qquad (3.1)$$

3.1'deki ifade bize çok katmanlı perseptronun mimari planını göstermektedir. Ağın çıkışındaki hata (e) istenen değer (d) ile gerçek çıkış (y) arasındaki farktır ve 3.1'de ifade edilmiştir. Bu algoritma ile asıl ulaşılmak istenen hatanın kabul edilebilir seviyeye getirilmesidir.

Geri yayılım algoritmasının gerçek zamanlı uygulamasında ağırlıkların sıralı olarak güncellemeyi tercih ettiğini farzediyorduk. Sürecin bu şekli için algoritma döngüsü eğitim örnekleri vasıtası ile aşağıdaki gibi olmaktadır (Haykin, 1994).

- *Başlangıç* : Giriş veri değerleri ağımızın giriş katmanına uygulanarak öğrenme başlatılır.

- *Eğitim Örneklerinin Sunumu* : Eğitim örnekleri devreleriyle ağ sunulur. Eğitim seti içindeki her bir örnek için ileri ve geri hesaplamalar sırasıyla bir sonraki aşamalarda gerçekleştirilecektir.

- *İleri Hesaplama* : (x(n), d(n)) şeklinde yazdığımız eğitim örneğimizde, x(n) giriş vektörümüz giriş katmanına uygulanır ve çıkış katmanındaki hesaplama noktasında d(n) çıkış vektörü olarak sunulacaktır. *l* tabakasındaki *j* nöronu için lokal alanın indüklenmesi 3.2'deki ifadedeki gibi olmaktadır.

$$v_j^{(l)}(n) = \sum_{i=0}^{Mo} w_{ji}^{(l)}(n)\, y_i^{(l-1)}(n) \qquad (3.2)$$

Burada $y_i^{(l-1)}(n)$, *l*-1 katmanındaki *i* nöronunun çıkış sinyali ve $w_{ji}^{(l)}(n)$, *l*-1 katmanındaki *i* nöronundan beslenen *l* katmanındaki *j* nöronunun ağırlığıdır. *i*=0 için

$y_0^{(l-1)}(n) = +1$ ve $w_{j0}^{(l)}(n) = b_j^{(l)}(n)$ *l* katmanındaki *j* nöronuna uygulandı. Sigmoid fonksiyonu kullandığımızı varsayalım, *l* katmanındaki *j* nöronunun çıkış sinyali,

$$y_j^{(l)} = \varphi_j (v_j(n)) \tag{3.3}$$

eğer *j* nöronu ilk gizli tabakada olsaydı (yani $l = 1$),

$$y_j^{(0)} = x_j(n) \tag{3.4}$$

$x_j(n)$, $x(n)$ giriş vektörünün *j* nci elemanıdır. Eğer *j* nöronu çıkış katmanında ise (yani $l = L$)

$$y_j^{(L)} = o_j(n) \tag{3.5}$$

Çıkış sinyalini hesaplarsak

$$e_j(n) = d_j(n) - o_j(n) \tag{3.6}$$

$d_j(n)$, $d(n)$ cevap vektörünün *j* nci elemanıdır.

- *Geri Hesaplama* : Ağın δs ı hesaplanır ve aşağıdaki gibi tanımlanır.

$$\delta_j^{(l)}(n) = \begin{cases} e_j^{(L)}(n)\, \varphi_j'(v_j^{(L)}(n)) & \text{L çıkış katmanındaki } j \text{ nöronu için} \\ \varphi_j'(v_j^{(l)}) \sum_k \delta_k^{(l+1)}(n)\, w_{kj}^{(l+1)}(n) & l \text{ gizli katmanındaki } j \text{ nöronu için} \end{cases} \tag{3.7}$$

Genelleştirilmiş delta kuralına göre *l* katmanındaki ağın ağırlıklarını düzeltilmesi,

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \alpha[w_{ji}^{(l)}(n-1)] + \eta\delta_j^{(l)}(n)\, y_i^{(l-1)}(n) \tag{3.8}$$

η öğrenme oranı parametresi ve α momentum sabitidir.

- *Yineleme* : Yineleme, yukarıda anlatılan ileri ve geri hesaplamalarla durma kriteri sağlanana kadar ağdaki eğitim örneklerinin yeniden devretmesi şekliyle sağlanır.

3.1.2. ANFIS (Uyarlamalı Ağlara Dayanan Bulanık Çıkarım Sistemi)

ANFIS, bulanık çıkarım sistemleri ve çok katmanlı perseptronlar uyarlamalı ağların çok genel hesaplama çalışmalarının özel örnekleridir. Her iki örnekte uyarlamalı ağın geriye yayılma öğrenme yeteneğini almıştır. ANFIS uyarlamalı ağların, işlevsel olarak bulanık çıkarım sistemine eşdeğer olan bir sınıfıdır. ANFIS, açık olarak uyarlamalı ağlara dayanan bulanık çıkarım sistemi veya uyarlamalı sinirsel bulanık çıkarım sistemi anlamına gelmektedir (Elmas, 2007).

ANFIS'in öğrenme algoritması, en küçük kareler yöntemi ile geri yayılmalı öğrenme algoritmasının bir arada kullanılmasından oluşan melez öğrenme algoritmasıdır (Çavuş, 2006).

ANFIS'in yapısında hem yapay sinir ağları hemde bulanık mantık kullanılır. Bulanık modelleme veya bulanık tanımlama ilk olarak Takagi ve Sugeno tarafından sistematik olarak araştırıldı ve kontrol, tahmin ve çıkarım için elverişli uygulamalar bulundu (Jang, 1993).

Esasen ANFIS yapısı, Sugeno tipi bulanık sistemlerin, sinirsel öğrenme kabiliyetine sahip bir ağ yapısı olarak temsilinden ibarettir. Bu ağ, her biri belli bir fonksiyonu gerçekleştirmek üzere, katmanlar halinde yerleştirilmiş düğümlerin birleşiminden oluşmuştur (Özçalık ve diğ., 2003).

Aslında bulanık çıkarım sistemi çok katmanlı perseptrona göre daha kuvvetlidir. Örnek olarak ANFIS denetleyicilerin bazı eşsiz özellikleri tanımlanabilir.

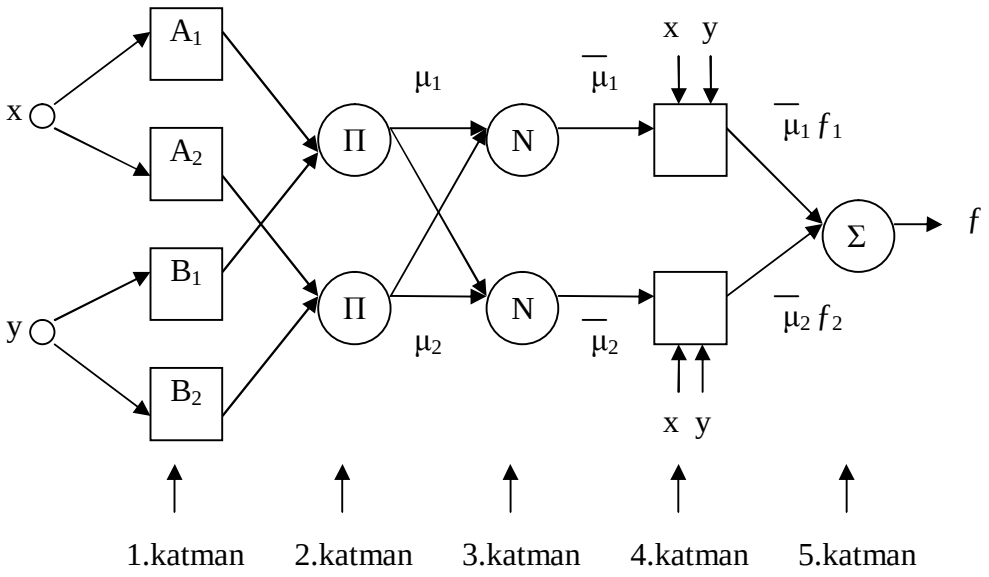
a) Öğrenme Yeteneği
b) Paralel İşlem
c) Yapılandırılmış Bilgi Temsili
d) Diğer Denetim Tasarım Yöntemleriyle Daha İyi Bütünleşme

çok katmanlı perseptron sadece a ve b özelliklerine sahiptirler fakat c ve d özelliklerine sahip değillerdir.

ANFIS'in yapısındaki bulanık çıkarım sisteminin mimarisinin kolaylıkla anlatılabilmesi için x ve y olmak üzere iki girişi ve f gibi bir çıkışı olduğu kabul edilirse, birinci derece Sugeno bulanık modeli için iki bulanık "EĞER O HALDE" kuralı aşağıda 3.9'daki gibi olur.

Kural 1 : EĞER $x=A_1$ ve $y=B_1$ ise O HALDE $f_1=p_1x+q_1y+r_1$ (3.9)
Kural 2 : EĞER $x=A_2$ ve $y=B_2$ ise O HALDE $f2=p_2x+q_2y+r_2$

Bu yapıya karşılık gelen eşdeğer ANFIS mimarisi Şekil 3.9'da görülmektedir. ANFIS mimarisi içerisindeki her katmana ait düğüm işlevleri ve dolayısıyla katmanların işlevleri sırasıyla aşağıda verilmiştir.



Şekil 3.9 İki Girişli ve İki Kurallı Sugeno Tip Bulanık Çıkarıma Eşdeğer ANFIS Mimarisi (Jang, 1993)

x ve y düğümleri giriş sinyallerinin diğer katmanlara aktarıldığı giriş düğümleridir. Birinci katman, her bir düğümün A_i ve B_i gibi bir bulanık kümeyi ifade eder. Bu katmandaki düğümlerin çıkışı giriş örneklerine ve kullanılan üyelik işlevine bağlı olan üyelik dereceleridir. İkinci katmandaki her düğüm Π ile etiketlenmiştir ve giren tüm işaretlerin çarpımını gösterir. Her bir düğümün çıkışı bir kuralın ateşleme seviyesini temsil eder. Üçüncü katmandaki her düğüm N ile etiketlenmiştir ve bir kuralın normalleştirilmiş ateşleme seviyesi hesaplanır. Dördüncü katmandaki her i düğümü, düğüm işlevi ile uyarlamalı bir düğümdür, kural katmanıdır. Her i düğümü

sonuç ağırlıkları değerlerini hesaplar. Beşinci katmanda sadece bir düğüm vardır ve Σ ile etiketlenmiştir. Burada, dördüncü katman çıkışından alınan sinyaller toplanır ve elde edilen sonuç sistemin gerçek çıkışı f değerini verir (Elmas, 2007).

Böylece Sugeno bulanık çıkarım modeline işlevsel olarak eşdeğer olan, örnek ANFIS yapısı tanımlanmıştır. Ağın yapısı tamamen sabit değildir. Ağın oluşturulması ve düğüm işlevlerinin görevlerine göre ayrılması, her katmandaki her bir düğümün sağladıklarına ve modüler işlevselliğine göre keyfi olarak seçilebilir. Sugeno tip ANFIS'ten Tsukamoto tip ANFIS'e kolaylıkla geçilebilir. Genellikle yaygın olarakta bu iki tip kullanılır. Mamdani tip bulanık çıkarıma karşılık gelen ANFIS için, Max-Min kompozisyonu ve sonuç çıkış için ağırlık merkezi durulama yöntemi ile elde edilebilir. Fakat bu Sugeno veya Tsukamoto tip ANFIS'e göre çok karmaşık ve zordur. Ayrıca öğrenme yeteneğine ve yaklaşım gücüne önemli bir katkıda sağlamamaktadır. Sonuç ve üyelik işlevlerine ait değişkenlerin ayarlanmasında geri yayılım öğrenme algoritması kullanılabilir (Elmas, 2007).

Bulanık mantık ve sinir ağları, akıllı sistemlerin geliştirilmesinde birlikte kullanılan tamamlayıcı araçlardır. Yapay sinir ağları, ham verilerle uğraşıldığında iyi sonuçlar veren düşük seviyeli yapılardır. Bulanık mantık ise, uzman görüşü sonucu elde edilen dilsel bilgileri kullanarak daha yüksek seviyeli sonuçlar çıkarmaktadır. Aslında bulanık sistemlerin öğrenme kabiliyeti yoktur ve kendilerini yeni çevreye adapte edemezler. Diğer yandan yapay sinir ağları öğrenme kabiliyetine sahiptir, fakat kullanıcı tarafından anlaşılmazlar. Sinirsel bulanık sistemler, yapay sinir ağlarının paralel hesaplayabilme ve öğrenme kabiliyeti ile bulanık mantığın uzman bilgisini kullanarak sonuçlar çıkarabilme özelliklerinin birleşiminden oluşur. Sonuç olarak sinirsel bulanık sistemler sayesinde yapay sinir ağları daha anlaşılır hale gelir (Çavuş, 2006).

Kredi kartı puanlama sistemi için birbirine zıt olan iki tip sınıflandırma sisteminden; bulanık sınıflandırma kuralları ve bulanık sinir ağları kullanılmış, gerçek dünya veri setleri üzerinde başarılı bir şekilde uygulanmıştır. Genetik bulanık sınıflandırma, sinirsel bulanık sınıflandırmaya göre çok daha iyi performans sergilemiştir (Hoffmann ve diğ., 2002).

3.1.3. Kural Tabanlı Öğrenme

Kural tabanlı öğrenme dediğimizde birçok metodla karşılaşmaktayız. Ancak tezimizi hazırlarken bu metodlardan; tembel öğrenme metodu (lazy learning) yada diğer bir adıyla bellek tabanlı metod ve karar ağaçlarını (decision tree) kullanacağız. Dolayısıyla sadece bu metodlardan bahsediyor olacağız.

Tembel öğrenme metodları veya bellek tabanlı metodlar saklı öğrenme örneklerinin sınıflandırmaları ile tanım kümesinin yapısını öğrenirler. Tanım kümesi modeli sonuçları önceden tanımlanmış uzaklık fonksiyonu kullanılarak tembel öğrenme ile genelleştirilebilir. Yeni bir bileşen için tanım kümesi modeli kullanılarak yeni bir sınıflandırmaya ihtiyaç duyulduğu zaman saklı olan en yakın görünmeyen örnek bulunarak sınıflandırma yapılır. Tembel öğrenmeye basit bir örnek vermek gerekirse aslan ve sivrisinek gibi iki hayvanın sınıflandırmasına bakalım. Aslan memeli olarak sınıflandırılırken sivrisinek böcek olarak sınıflandırılır. Öğrenme algoritmasına aslan için [büyük, saçlı, hayvan yer] bilgisi verilirken sivrisinek için [küçük, uçabilir, iğneli] bilgisi sunulur. Bu algoritma yeni hayvanların özelliklerini ve memeli veya böcek sınıfında olduğunu karşılaştıracaktır. Kaç tane ortak özelliği olduğunu arayacaktır. Örneğin yarasanın özellikleri [küçük, uçabilir, saçlı] olduğu göz önüne alınıp sınıflandırılırsa böcek sınıfına girecektir. Çünkü böcek sınıfında iki ortak özelliğe memeli sınıfında ise bir ortak özelliğe sahiptir (Erik ve diğ., 1998).

Karar ağaçları, veri madenciliğinde kuruluşlarının ucuz olması, yorumlanmalarının kolay olması, veri tabanı sistemleri ile kolayca entegre edilebilmeleri ve güvenilirliklerinin iyi olması nedenleri ile sınıflama modelleri içerisinde en yaygın kullanıma sahip tekniktir. Karar ağacı, adından da anlaşılacağı gibi bir ağaç görünümünde, tahmin edici bir tekniktir. Ağaç yapısı ile, kolay anlaşılabilen kurallar yaratabilen, bilgi teknolojileri işlemleri ile kolay entegre olabilen en popüler sınıflama tekniğidir.

Karar ağacı karar dügümleri, dallar ve yapraklardan oluşur. Karar dügümü, gerçekleştirilecek testi belirtir. Bu testin sonucu ağacın veri kaybetmeden dallara ayrılmasına neden olur. Her düğümde test ve dallara ayrılma işlemleri ardışık olarak gerçekleşir ve bu ayrılma işlemi üst seviyedeki ayrımlara bağımlıdır. Ağacın her bir

dalı sınıflama işlemini tamamlamaya adaydır. Eğer bir dalın ucunda sınıflama işlemi gerçekleşemiyorsa o dalın sonucunda bir karar düğümü oluşur. Ancak dalın sonunda belirli bir sınıf oluşuyorsa, o dalın sonunda yaprak vardır. Bu yaprak, veri üzerinde belirlenmek istenen sınıflardan biridir. Karar ağacı işlemi kök düğümünden başlar ve yukarıdan aşağıya doğru yaprağa ulaşana dek ardışık düğümleri takip ederek gerçekleşir.

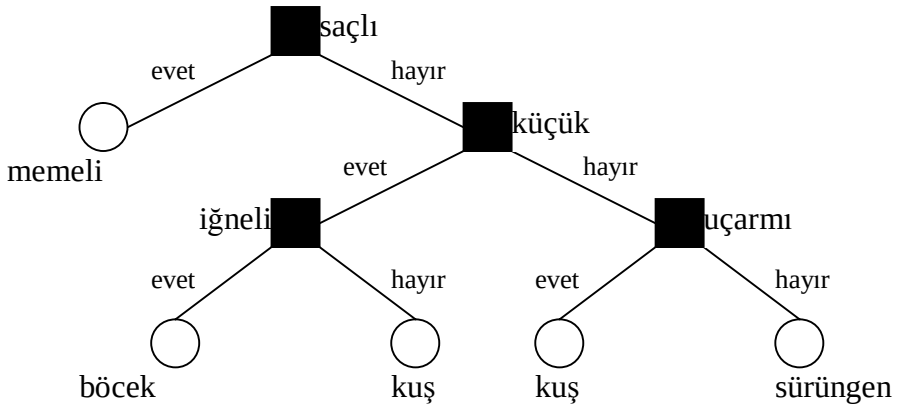
Karar ağacı tekniğini kullanarak verinin sınıflanması iki basamaklı bir işlemdir. İlk basamak öğrenme basamağıdır. Öğrenme basamağında önceden bilinen bir eğitim verisi, model oluşturmak amacıyla sınıflama algoritması tarafından analiz edilir. Öğrenilen model, sınıflama kuralları veya karar ağacı olarak gösterilir. İkinci basamak ise sınıflama basamağıdır. Sınıflama basamağında test verisi, sınıflama kurallarının veya karar ağacının doğruluğunu belirlemek amacıyla kullanılır. Eğer doğruluk kabul edilebilir oranda ise, kurallar yeni verilerin sınıflanması amacıyla kullanılır.

Test verisine uygulanan bir modelin doğruluğu, yaptığı doğru sınıflamanın test verisindeki tüm sınıflara oranıdır. Her test örneğinde bilinen sınıf, model tarafından tahmin edilen sınıf ile karşılaştırılır. Eğer modelin doğruluğu kabul edilebilir bir değer ise model sınıfı bilinmeyen yeni verileri sınıflama amacıyla kullanılabilir. Oluşturulan bu modelin doğruluğu, bir test verisi aracılığı ile onaylandıktan sonra model, sınıfı belli olmayan yeni bir veriye uygulanabilir ve sınıflama kuralı gereği yeni verinin sınıfı "mükemmel" olarak belirlenebilir.

Tekrarlamak gerekirse bir karar ağacı, bir alandaki testi belirten karar düğümlerinden, testteki değerleri belirten dallardan ve sınıfı belirten yapraklardan oluşan akış diyagramı şeklindeki ağaç yapısıdır. Ağaç yapısındaki en üstteki düğüm kök düğümüdür (Özekeş, 2003).

En sık kullanıma sahip karar ağacı modelleri ID3 ve daha gelişmiş modeli olan C4.5 tür. Bunun yanında IBM'in IntelligentMiner modeli de yaygın olarak kullanılmaktadır. ID3 ve C4.5 modellerinde bütün değişkenlerin ayrık olduğu varsayılır. Modelde bilgi kazanımı esas alınır ve herbir değişkenin entropisi hesaplanır. Entropi ile ayrık değişkenlerin bilgi kazancı ölçülür ve bilgi kazanımı en

fazla olan değişken kök düğüm olarak seçilir ve bu şekilde hesaplamalara devam edilerek ağaç oluşturulur. IBM'in modelinde bütün değişkenlerin sürekli olduğu varsayılır. Modelde herbir değişkenin ayrı ayrı gini index değeri hesaplanır. En düşük gini değerini veren ayrıma sahip değişken kök düğüm olarak seçilir ve bu şekilde hesaplamalara devam edilerek ağaç yapısı oluşturulur. Aşağıda örnek bir karar ağacı modeli gösterilmiştir.



Şekil 3.10 Hayvan Sınıflandırması için Karar Ağacı Örneği (Erik ve diğ., 1998)

Belirli bir sınıfın muhtemel üyesi olacak elemanların belirlenmesi, çeşitli durumların yüksek, orta, düşük risk grupları gibi çeşitli kategorilere ayrılması, gelecekteki olayların tahmin edilebilmesi için kurallar oluşturulması, sadece belirli alt gruplara özgü olan ilişkilerin tanımlanması, kategorilerin birleştirilmesi gibi alanlarda karar ağaçları kullanılmaktadır.

Karar ağaçları, hangi demografik grupların mektupla yapılan pazarlama uygulamalarında yüksek cevaplama oranına sahip olduğunun belirlenmesi (Direct Mail), bireylerin kredi geçmişlerini kullanarak kredi kararlarının verilmesi (Credit Scoring), geçmişte işletmeye en faydalı olan bireylerin özelliklerini kullanarak işe alma süreçlerinin belirlenmesi, tıbbi gözlem verilerinden yararlanarak en etkin kararların verilmesi, hangi degişkenlerin satışları etkilediğinin belirlenmesi, üretim verilerini inceleyerek ürün hatalarına yol açan değişkenlerin belirlenmesi gibi uygulamalarda kullanılmaktadır (Akpınar, 2000). Bunun yanında kredi kartı dataları ile sahtecilik tespiti için yapılan sınıflandırma modeli çalışmasında karar ağaçları etkin olarak kullanılmış ve olumlu sonuçlar ortaya koyduğu görülmüştür (Shen ve

diğ., 2007). Yapılan bir başka çalışmada, bankanın sahtecilik tespit sistemi karar ağacı teknikleri üzerinde geliştirilmiş olup sistem sonuçları oldukça başarılı olmuştur (Chiu ve diğ., 2004). Veri madenciliği tekniklerinden biri olan karar ağaçları şirketlerin bilançolarının incelenmesinde kullanılmış ve tahmin doğruluğu olarak tatmin edici sonuçlar ortaya koymuştur (Kirkos ve diğ., 2007). Bunun yanında ensemble, bagging ve boosting algoritmalarının uygulamalarında da karar ağaçları teknikleri kullanılmıştır.

3.1.4. Destek Vektör Makineleri

Destek vektör makineleri, sınıflandırma ve doğrusal olmayan fonksiyon yaklaşımı problemlerinin çözümü için 1960'ların sonunda Vapnik tarafından önerilen eğiticili bir öğrenme algoritmasıdır. Son yıllarda daha yaygın olarak kullanılmaya başlanan destek vektör makineleri; yazı tanıma, nesne tanıma, ses tanıma, yüz tanıma gibi örüntü tanıma uygulamalarında kullanılmıştır (Burges, 1998). Bu sınıflandırma yöntemi, birçok bilim ve mühendislik alanındaki konulara başarıyla uygulanmıştır. Ayrıca diğer geleneksel öğrenme yöntemleriyle karşılaştırıldığında bu sınıflandırıcının doğrusal olmayan problemleri çözmesindeki performansı ve yeteneği çok daha iyidir.

Günümüzde performansı sayesinde oldukça popüler olmuş bir metottur. Temelde lineer olarak ayrıştırılabilir iki sınıfın karar yüzeyinin destek vektörler olarak tanımlanan ve sınıf sınırlarını belirleyen örnekler arasında maksimum marjinin oluşturulması ilkesine dayanan bir algoritmadır. Marjın maksimizasyonu işlemi bir kuadratik sınırlamalı optimizasyon problemi şeklinde yazılarak, Lagrangian fonksiyonu şeklinde ifade edilerek dual forma dönüştürülür. Doğrusal problemler için gerçekleştirilen bu yaklaşım doğrusal olmayan ayrıştırma problemleri için kernel dönüşümleri kullanılarak genelleştirilebilir (Polat ve diğ., 2007) Destek vektör makineleri tekniği, sınıfları birbirinden ayıran marjini en büyük, dogrusal bir ayırt edici fonksiyon bulunmasını amaçlar. Doğrusal olarak ayrılamayan örnekler için, örnekler dogrusal olarak ayrılabildikleri daha yüksek boyutlu başka bir uzaya taşınır ve sınıflandırma o uzayda yapılır (Amasyalı ve diğ., 2006).

Daha öncede açıklandığı gibi destek vektör makineleri doğrusal olmayan örnek uzayını, örneklerin doğrusal olarak ayrılabileceği bir yüksek boyuta aktararak, farklı örnekler arasındaki maksimum sınırın bulunması esasına dayanır. Destek vektör makinelerinde karşılaşılabilecek iki durum bulunmaktadır. Bu durumları verilerin doğrusal olarak ayrılabilecekleri bir yapıda olması ve verilerin doğrusal olarak ayrılamayan bir yapıda olması şeklinde sıralayabiliriz. Verilerin doğrusal olarak ayrılabilen bir yapıda olması destek vektör makineleri tekniğinin en basit modelidir. Bu veriler arasında direkt olarak maksimum sınırın bulunması oldukça kolaydır. Burada klasik yöntemler kullanılarak bu analizler yapılabilmektedir. Ancak gerçek dünya problemlerinin büyük çoğunluğu birçok farklı bileşenden oluşan problemler olmakta ve doğrusal olarak ayrılmış bir yapı halinde karşımıza çıkmamaktadır. Bu durumda doğrusal olarak ayrılamayan veriler öncelikle doğrusal olarak ayrılabilecekleri farklı bir uzaya aktarılmalıdır. Böyle problemlerde de doğrusal olmayan sınıflandırma yöntemi kullanılmaktadır.

Makine öğrenme tekniklerinden biri olan destek vektör makineleri, kredi kartı puanlama ve kredi risk araştırmalarında kullanılmış olup diğer veri madenciliği teknikleri ile birlikte yüksek doğruluk sonuçları üretmiştir. Bunun yanında kredi kartı başvuru onayı üzerine geliştirilmiş olan sistemde karar ağaçları, geri-yayılım sinir ağları, destek vektör makineleri ve geliştirilen MOGP sınıflandırma metodu kullanılmış olup en iyi sonucu MOGP metodu vermiştir (Zhao, 2007). Çin'de 2000 yılında, destek vektör makineleri tekniği kullanılarak geliştirilen kredi puanlama sistemi ile listelenen 106 firma incelenmiş olup firmalar performanslarına göre iyi ve kötü olarak 2 grupta toplanmıştır. Simülasyon sonuçları destek vektör makineleri kullanılarak geliştirilen kredi puanlama sisteminin % 98.11 oranında doğru sınıflandırma yaptığını göstermiştir (Tian ve diğ., 2004). Oldukça yeni olan makine öğrenme tekniklerinden destek vektör makineleri, kurumsal kredi derecelendirme problemlerine ve daha doğru tahmin için yeni modeller geliştirmeyi sağlar (Lee, 2007). Destek vektör makinelerinin yanında yapay sinir ağları, kural tabanlı öğrenme ve istatistiksel yaklaşımlar gibi akıllı veri analizi yöntemleri görüntü işleme, ses tanıma, örüntü tanıma ve tıbbi alanda hastalık teşhisinde de kullanılmaya başlanmıştır (Özkaya ve diğ., 2005). Kısa dönem portföy yönetimi için çekirdek metodları üzerinde yapılmış çalışmada destek vektör makineleri algoritması kullanılmış ve destek vektör makineleri için % 68.35 doğru sınıflandırma oranı elde

edilmiştir. Bu çalışma ekonomi piyasaları üzerinde yapılan modellemelerde destek vektör makineleri algoritmasının kullanımının iyi bir seçim olacağını göstermiştir (Ince ve diğ., 2006) Döviz kuru tahmini üzerine yapılmış bir diğer çalışmada destek vektör makineleri ve yapay sinir ağları teknikleri kullanılarak iki safhalı tahmin modeli sunulmuştur. Bu modellerin karşılaştırması giriş seçimlerinin çok önemli olduğunu ve bulgular sonucunda destek vektör makinelerinin yapay sinir ağlarına göre daha iyi performans sergilediğini göstermiştir. (Ince ve diğ., 2006)

Destek vektör makineleri doğrusal (lineer) ve doğrusal olmayan (lineer olmayan) olmak üzere iki farklı alana ayrılmaktadır. Doğrusal destek vektör makineleri de kendi içinde doğrusal olarak ayrılabilen ve doğrusal olarak ayrılamayan verilerle işlem yapan destek vektör makineleri olarak ayrılmaktadırlar.

3.1.4.1 Doğrusal Destek Vektör Makineleri

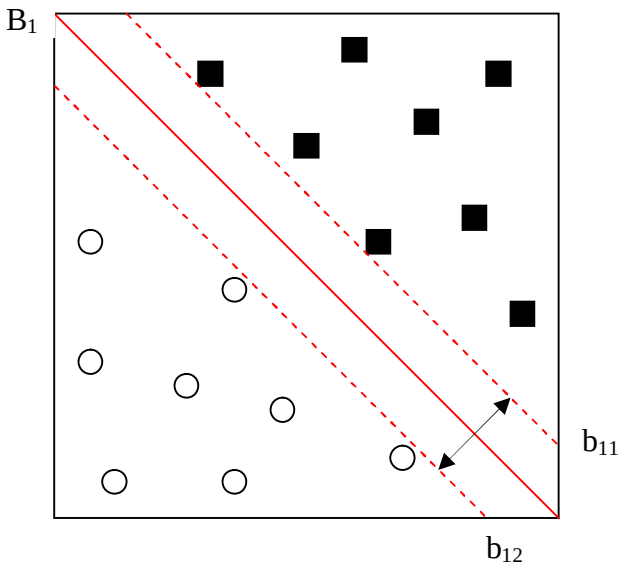
Eğitim için kullanılacak N elemandan oluşan verinin $x=\{x_i,y_i\}$, $i=1,2,\ldots,N$ olduğunu varsayalım. Burada $y_i=\{-1,1\}$ etiket değerleri ve $x_i \in R^d$ özellikler vektörüdür. Doğrusal olarak ayrılabilme durumunda, bu iki değerli veriler direkt olarak bir aşırıdüzlem ile ayrılabilecektir. Bu aşırıdüzleme ayırıcı aşırıdüzlem adı verilir. Destek vektör makinelerinin amacı bu aşırıdüzlemin iki örnek grubuna eş uzaklıkta olmasını sağlamaktır (Demirci, 2007).

Pozitif örnekleri negatif örneklerden ayıran bir hiper düzlemimiz olduğunu varsayarsak, iki sınıf probleminde doğrusal sınıflandırıcı hiper düzlemin normal vektörü w ve ofset değeri b ile tanımlanır. Karar sınırı $w^Tx + b = 0$ doğrusudur. Bu hiper düzlemin ayırdığı yarı uzaylardan her biri bir sınıf belirtir. Bu durumda aşağıdaki koşulların gerçeklenmesi gerekir.

$$w^Tx_i + b \geq 1 \text{ , } y_i = 1 \qquad (3.10)$$
$$w^Tx_i + b \leq -1 \text{ , } y_i = -1$$

Sınıf etiketi genelleştirilmiş olarak $y_i = \text{sgn}(w^Tx_i + b)$ eşitliği ile belirtilebilir. Doğrusal sınıflandırıcının sınırı hiper düzlem ile öğrenme verileri arasındaki en kısa uzaklık olarak tanımlanır. Ayırt etme yüzeyine en yakın veriler destek vektörleri

olarak adlandırılır ve karar sınırı sadece destek vektörleri ile belirlenir. Az sayıda veri noktası sınıflandırmada etkindir. Destek vektör algoritması en büyük margine sahip ayırıcı hiper düzlem ile sınıflandırma yaparak öğrenme hatasını en aza indirmeye çalışır (Özkaya ve diğ., 2005).



Şekil 3.11 Doğrusal Destek Vektör Makineleri

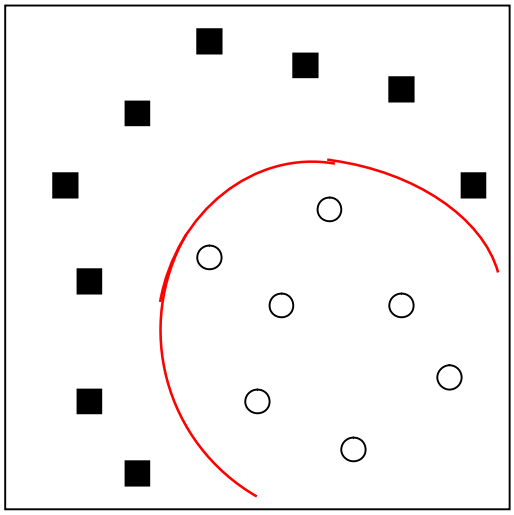
3.1.4.2 Doğrusal Olmayan Destek Vektör Makineleri

Doğrusal olmayan problemlerde çözümü bulmanın bir başka yolu da çekirdek fonksiyonları ile örneklerin öncelikle daha yüksek boyutlu ve lineer olarak ayrılabilecekleri bir uzaya taşınıp çözümün bu yeni uzayda aranmasıdır (Demirci, 2007).

Doğrusal olarak ayırt edilemeyen sınıflarda ise çekirdek fonksiyonları devreye girer ve n boyutlu bir veri kümesi m > n olmak üzere m boyutlu yeni bir veri kümesine dönüştürülerek, yüksek boyutta doğrusal sınıflandırma işlemi gerçekleştirilir. Çekirdek fonksiyonları destek vektör makineleri algoritmasında önemli rol oynar. Çeşitli çekirdek fonksiyonları tanımlanmıştır. Bu çekirdek fonksiyonlarından bazıları; doğrusal, polinom, gauss (radyal tabanlı fonksiyon), iki katmanlı perseptron olarak sayılabilir (Özkaya ve diğ., 2005).

Doğru çekirdek fonksiyonun seçimi sınıflandırma işleminin başarısında oldukça etkilidir. Destek vektör makineleri temelde ikili sınıflandırıcı olduğundan

çoklu sınıf problemleri basit olarak veri kümesinin değişik şekillerde ikiye bölünmesi, her bir ikili veri kümesi için destek vektör makinesi algoritmasının çalıştırılması ve sonuçların birleştirilmesi ile çözümlenir (Özkaya ve diğ., 2005).



Şekil 3.12 Doğrusal Olmayan Destek Vektör Makineleri

3.2. Klasik İstatistiksel Teknikler

Fen, sosyal ve tıp bilimleri araştırmalarından devlet yönetimine, üretim sektöründen mühendisliğe kadar pek çok alanda istatistiksel teknikler kullanılmaktadır. İstatistiksel teknikler, veri madenciliği'nin temelini oluşturmakla birlikte geniş bir kullanım alanına sahiptir. Buradan yola çıkarak Veri Madenciliği'nin istatistiksel tekniklerin evrim geçirmiş hali olarak düşünmemiz yanlış bir düşünce olmayacaktır.

Klasik istatistiksel teknikler oldukça katı olan bazı varsayımlar (verilerin simetrik olması, normal dağılış göstermesi, sapan gözlem olmaması vb.) geçerli olduğunda en iyi olacak şekilde tasarlanmışlardır. Bununla beraber, uygulamalarda karşılaşılan ve klasik analizlerin gerektirdiği varsayımların sağlanmadığı durumlarda bu tekniklerin uygulamaları oldukça yanıltıcı sonuçlar verebilir. Bu amaçla son yıllarda geliştirilen sağlam ve keşfedici yöntemler istatistiksel analizlerin etkinliğini arttırmıştır. Günümüzde istatistik tekniklerinin en yaygın olarak kullanıldığı sektörlerin başında otomotiv parçası üreten kuruluşlar gelmektedir (Özler, 2006).

Klasik istatistiksel teknikler olarak lojistik regresyon analizi, diskriminant analizi, kümeleme analizi, ki kare analizi, temel bileşenler analizi, varyans analizleri (ANOVA, MANOVA), ölçekleme analizi, bitişiklik analizi, doğal korelasyon analizi ve faktör analizi sayılabilir. Biz bu bölümde farklı gruplardan gelen bireyleri sınıflandırmak amacıyla kullanılabilen lojistik regresyon analizi ve diskriminant analizi üzerinde detaylı olarak duracağız. Her iki yöntemde teorik bir çerçevede incelenecek ve uygulamada izlenebilecek adımlar ile ilgili detaylar bölümlerin içinde verilecektir.

3.2.1. Lojistik Regresyon Analizi

Lojistik regresyon, bağımlı değişken üzerinde hangi bağımsız değişken/ değişkenlerin önemli risk faktörü olduğunu ve bu bağımsız değişken/değişkenlerin bağımlı değişkenin değerlerinin tahmininde ne düzeyde etkide bulunduklarını belirler. Kısaca bağımsız değişkenlerin etkilerine dayanarak verilerin sınıflandırılmasında kullanılan bir yöntemdir. Bu yöntem olasılık kurallarına uygun olarak verilerin belirli sınıflara atanmasını sağlar (Kurt ve diğ., 2005). Lojistik regresyon analizinin kullanım amacı, istatistikte kullanılan diğer model yapılandırma teknikleri ile aynıdır. En az değişkeni kullanarak en iyi uyuma sahip olacak şekilde bağımlı ile bağımsız değişkenler arasındaki ilişkiyi tanımlayabilen ve biyolojik olarak kabul edilebilir bir model kurmaktır (Bircan, 2004).

Lojistik regresyon son yıllarda popülerliği gittikçe artan bir regresyon yöntemidir. Temelde amacı diğer regresyon yöntemleri gibi bir ya da birden çok bağımsız değişken ile bağımlı değişken arasındaki ilişkiyi modellemektir. Özellikle tahmin edilmek istenen değişken iki durum alıyorsa yani başarılı-başarısız, hasta-sağlıklı, makine çalışır-çalışmaz gibi durumlarda lojistik regresyon modelleri kullanılmaktadır. Lojistik regresyon daha çok durum kontrolü çalışmalarında kullanılan bir yöntemdir.

İleri parametrik olmayan bir istatistiksel yöntem olan lojistik regresyon analizi, bağımlı değişken mutlaka ikili sonucu olan değişken olduğu durumlarda kullanılır. Ayrıca zorunlu olmamakla beraber bağımsız değişkenler genellikle sürekli olurlar. Bu avantajından dolayı gözlemlerin gruplara ayrılmasında ve yeni

gözlemlerin bu uygun gruplara atanmasında sıkça tercih edilen bir yöntem olmaktadır (Atan ve diğ., 2004).

Lojistik regresyon, bağımsız değişkenleri kullanarak ikili çıktısı olan bağımsız değişkenin istenilen durumunun gerçekleşme olasılığını hesaplar. Lojistik regresyon şu şekilde yazılır.

$$\text{Log} \frac{p}{(1-p)} = b_0 + b_1 * X_1 + \ldots + b_n * X_n \qquad (3.11)$$

p istenilen durumun gerçekleşme olasılığını, b_0 sabit değerinin, b_i (i=1,2,…n) ise herbir bağımsız değişkenin X_i (i=1,2,…n) katsayısını belirtir. Lojistik regresyon varsayıma ihtiyaç duymaz ve bununla beraber bir olasılık değeri vereceğinden yorumlanması daha kolaydır (Çinko, 2006).

Lojistik regresyon analizinin uygulamadaki adımları aşağıdaki gibidir: (Ünsal, 2005).

- Önsel grup üyelikleri belirlenir.

- Modele girecek değişkenler belirlenir. Bu amaçla önsel bilgiden ya da istatistiksel tekniklerden yararlanılabilir.

- Modelin parametreleri Newton-Raphson yöntemi ile tahmin edilir. Ardından modelin tümünün anlamlılığı olabilirlik oranı ile test edilir. Model anlamlı değilse analize son verilir. Eğer model anlamlı bulunursa diğer aşamaya geçilir.

- Tahmin edilen model parametrelerinin tek tek anlamalılığı incelenir. Bu amaçla olabilirlik oranı ya da Wald istatistiği kullanılabilir. Her katsayının anlamlılığı incelendikten sonra, teklik oranları incelenerek, açıklayıcı değişkenlerin bağımlı değişken üzerindeki etkileri yorumlanabilir.

- Tahmin edilen model parametreleri kullanılarak, her bir gözlemin hangi gruptan geldiği tahmin edilir.

- Modelin uyum iyiliğini incelemek amacıyla doğru sınıflandırma yüzdesi ve yapay R^2 ölçütleri kullanılır. Modelin uyum iyiliği kabul edilebilir düzeyde ise beşinci aşamadaki grup tahminleri kullanılabilir. Aksi halde ikinci aşamaya geçilerek modele girecek değişkenler yeniden gözden geçirilir ve işlemler tekrar edilir.

Lojistik regresyon bir regresyon tekniği olmakla beraber varsayımlarda diğer regresyon yöntemlerine göre farklılıklar gösterir. Bu farklılıklar bu yöntemin daha çok tercih edilmesini sağlar. Regresyon analizinde bağımsız değişkenlerin çoklu normal dağılım göstermesi ve özellikle bağımlı değişkenin sürekli olması koşulu aranırken, lojistik regresyonda bu koşullar aranmaz. Kesikli değişkenin hata terimi normal dağılım yerine binom dağılımı gösterir ve bu da regresyonda kullanılan tüm istatistiksel testlerin geçersiz olmasına sebebiyet verir. Dolayısıyla bu durumda lojistik regresyonun kullanılması araştırmacıya avantaj sağlar. Ayrıca lojistik regresyon analizinde diskriminant analizindeki varyans-kovaryans matrislerinin eşitliği şartı da aranmamaktadır. Lojistik regresyon analizi değişkenler arasında çoklu bağlantı olmadığını varsayar. Herhangi bir değişken diğer değişkenlerin doğrusal kombinasyonu cinsinden yazılmamalıdır. Böylece analizde bazı değişkenlerin toplamı ya da bazı değişkenlerin ortalamaları orijinalleriyle aynı yeni bir değişken olarak kullanılmamalıdır. Çoklu bağlantı regresyon analizinde katsayıların yanlış tahmin edilmesine, katsayıların standart hatalarının yüksek çıkmasına, yapılan *t* testlerinin geçersiz olmasına ve modelin tahmin gücünün azalmasına sebebiyet verebilir. Lojistik regresyon analizinde de benzer sorunlara yol açabilir. Bu yüzden eğer varsa, çoklu bağlantı durumunun tespit edilmesi ve gerekli düzeltmelerin yapılması gereklidir. Ancak araştırmacının amacı karar birimini uygun sınıfa yerleştirmekse, çoklu bağlantı problemi ihmal edilebilir. Bu sorunun tespiti için değişkenler arasındaki korelasyonlara bakmak gereklidir. Yüksek korelasyon çoklu bağlantıya işaret eder. Ayrıca doğrusal regresyonun varsayımları olan bütün ilişkili bağımsız değişkenlerin modele alınması buna karşın ilişkisiz bağımsız değişkenlerin modelden çıkarılması, hata terimlerinin birbirinden bağımsız olması gibi varsayımlar lojistik regresyon analizi için de geçerlidir (Atan ve diğ., 2004).

Doğrusal olmayan regresyon metodu ile bölmeleme metodunun kullanıldığı çalışmada test datası olarak kullanılan 944 kaydın içinde 255 adet kayıt şüpheli olarak seçilmiştir. Her kayıt 71 adet nitelik içermektedir. Analiz sonucunda her iki metodunda verdiği sonuçlar tatmin edici boyuttadır (Fischer, 2005). Sahtecilik tespiti için kullanılan sınıflandırma metodlarından lojistik regresyon, karar ağaçları ve sinir ağları teknikleri içinden en doğru oranı % 59 ile yapay zeka, % 58 ile lojistik regresyon sağlamıştır (Shen ve diğ., 2007). Genellikle, kredi puanlama modeli inşaa etmek için iki temel lineer istatistik metodu olan diskriminant analizi ve lojistic regresyon kullanılmıştır. Uygulama sonuçlarında birinci tip hataya lojistik regresyon % 23.81 ile diskriminant analizine göre % 25.43 ile daha doğru sınıflandırma gerçekleştirmiştir (Lee ve diğ., 2002).

3.2.2. Diskriminant Analizi

Diskriminant analizi, hatalı sınıflandırma olasılığını en aza indirgeyerek birimleri ait oldukları gruplara ayırmak amacına yönelik olan tamamıyla istatistiksel bir karar verme sürecidir (Tatlıdil, 1996). Diskriminant analizi, iki ya da daha fazla sayıdaki gruba ait birimler arasındaki farklılıkları maksimum yapan ve değişkenlerin doğrusal bileşiminden meydana gelen bir veya daha çok fonksiyonun belirlenmesi işlemidir. Bunun yanında sınıflandırma modellerinin geliştirilmesinde kullanılan ilk çok değişkenli istatistiksel sınıflandırma yöntemidir.

Çok degişkenli istatistiksel teknikler Fisher'in doğrusal diskriminant analizi üzerindeki çalışmalarına kadar uzanmaktadır. Diskriminant analizi, araştırmacılar ve uygulamacılar tarafından sınıflandırma modelleri geliştirmede kullanılan ilk çok değişkenli istatistiksel sınıflandırma yöntemi ve en yaygın olarak kullanılan metodoloji olmuştur. Fisher tarafından tanıtıldığında diskriminant analizinin amacı, iki grubu birbirinden ayırabilmekti. Fisher değişkenlerin doğrusal bir bileşimi olan diskriminant fonksiyonunu oluşturmuştur. Fisher'in çalışmasından yaklaşık olarak on yıl sonra Smith doğrusal diskriminant analizinden, karesel diskriminant analizini elde etmiştir. Daha sonra Welch diskriminant fonksiyonunun oluşturulmasında Neyman-Pearson olabilirlik oranı kriterini kullanmıştır (Sığırlı, 2006).

Bu analiz, araştırmacıya iki veya daha fazla grubun çeşitli değişkenlere bağlı olarak ortaya çıkan farklılıklarını ortaya koymasına imkan vermektedir. Diskriminant analizinde birimler en az hata ile ait oldukları kitlelere ayrılmaktadırlar. Bu analizin temelinde, incelenen bireyin kitlesinin belirlenmesini sağlayacak bir fonksiyon bulunmaktadır.

Diskriminant analizi, birbirleriyle yakından ilişkili istatistiksel birkaç yaklaşımı kapsayan geniş bir kavramdır. Araştırmacı bu yaklaşımların hepsini birarada kullanmayabilir. Bu yaklaşımlar iki ana kategoride ele alınabilir. Birinci kategori gruplar arası farklılıkların yorumlanmasından faydalanırken, ikinci kategori birimleri gruplara ayırmak amacıyla kullanılmaktadır. Genel olarak birimlerin gruplanmasında bazı matematiksel eşitliklerden faydalanılmaktadır. Diskriminant fonksiyonu olarak adlandırılan bu eşitlikler birbirine en çok benzeyen grupları belirlemeye imkan tanıyacak şekilde grupların karakteristiklerini ortaya koymak amacıyla kullanılmaktadır. Kısaca diskriminant analizi, iki veya daha fazla sayıdaki grubun farklılıklarının diskriminant değişkenleri vasıtasıyla ortaya konması işlemidir (Atan ve diğ., 2004).

Sınıflama tekniği olan diskriminant analizinde istenilen şey gruplar arası varyansın grup içi varyansa oranını maksimum kılmaktır. Diskriminant fonksiyonu şu şekildedir.

$$D = w_0 + w_1 * X_1 + \ldots + w_n * X_n \qquad (3.12)$$

D diskriminant değerini, w_0 sabit değerini, w_i (i=1,2,…n) ise bağımsız değişkenlerin X_i (i=1,2,…n) katsayı değerini göstermektedir. Model tahmin edildikten sonra veri seti için diskriminant değerleri hesaplanır ve her bir grubun ortalama değeri bulunur. Grupların ortalama diskriminant değerlerinden bir kritik değer elde edilir.Test verisi kullanılarak elde edilecek olan diskriminant değerleri kritik değer ile karşılaştırılır ve gözlemin hangi sınıfa ait olduğuna karar verilir (Çinko, 2006).

Diskriminant analizi tek faktör çok değişkenli varyans analizi MANOVA'nın uzantısı olan çok değişkenli bir analiz türüdür. Gruplar arası fark yoktur anlamını

taşıyan H_o hipotezi red edildikten sonra, gruplar arası farkın olduğu sonucuna varılır. Bu farklılığın ana nedenleri diskriminant analizi tekniğiyle ortaya çıkarılır. Diskriminant analizi aracılığıyla elde edilen diskriminant fonksiyonları, tahmin değişkenlerinin doğrusal bileşenlerinden oluşur. Diskriminant fonkiyonları gruplar arası farklılığa etki eden tahmin değişkenlerinin hangileri olduğunu ortaya çıkarır. Gruplar arası farklılığa etki eden bu değişkenlere de diskriminant değişkenler adı verilir. Diskriminant analizinin bir diğer işlevi ise, gruplardan herhangi birisine ait olan fakat hangi gruptan geldiği bilinmeyen bir birimin ait olduğu grubu en az hata ile saptamaktır.

O halde diskriminant analizinin amacını iki grupta toplamak olanaklıdır. Birincisi diskriminant fonksiyonları saptayıp ve bu fonksiyonlar aracılığıyla gruplar arası ayırıma en fazla etki eden diskriminant değişkenlerini belirlemek, ikincisi hangi gruptan geldiği bilinmeyen bir birimin hangi gruba dahil edileceğini belirlemektir. Birinci amaca yönelik diskriminant analizi betimsel amaçlı analiz, ikinci amaca yönelik olarak diskriminant analizi karar amaçlı analiz olarak adlandırılır (Ünsal, 2000).

Diskriminant analizinin uygulama adımları aşağıdaki gibidir: (Ünsal, 2005).

- Önsel grup üyelikleri belirlenir.

- Değişkenler için gruplar arasında fark olup olmadığı, Wilks'in Λ istatistiği ile belirlenir. Bu amaçla yapılacak MANOVA testi sonucunda gruplar arasında anlamlı bir fark varsa analize devam edilir. Eğer anlamlı bir fark bulunamazsa tüm grupların ortalamalarının eşit olduğu, dolayısıyla grup farkı olmadığı söylenebilir. Bu durumda diskriminant analizi yapılamaz.

- Kullanılacak değişkenler seçilir. Değişken seçiminde önsel bilgi ya da istatistiki yöntemler uygulanabilir.

- Değişkenler arasında çoklu bağlantının olup olmadığı incelenir. Bu amaçla birleştirilmiş grup içi korelasyon matrisi incelenir. Bu matristeki korelasyon

değerleri mutlak değerce % 75'ten büyük ise değişkenlerden bir kısmının atılması gerekir. Bu adımın sonunda değişken kümesi belirlenmiş olur.

- $\mathbf{W^{-1}B}$ matrisinin özdeğerleri ve bu özdeğerlere ilişkin özvektörler bulunur. Bu özvektörler, diskriminant fonksiyonları için gerekli ağırlıkları verir. Diskriminant fonksiyonlarının anlamlılık testi de bu özdeğerler kullanılarak yapılır. Eğer herhangi bir fonksiyon anlamlı ise yaptığı ayrımın başarılı olduğu söylenebilir.

- Standartlaştırılmamış diskriminant fonksiyonu kullanılarak her bir birey için diskriminant fonksiyonu değerleri elde edilir. Bu değerler sınıflandırma aşamasında kullanılacaktır.

- Grup üyelikleri için önsel olasılıklar belirlenir. Daha sonra bu olasılıklar ve diskriminant skorları kullanılarak sonsal olasılıklar elde edilir. Bireyin sahip olduğu en büyük sonsal olasılık tespit edilir. Bu olasılığı veren grubun o bireyin ait olduğu grup olduğu tahmin edilir ve birey sınıflandırılmış olur.

- Her bir birey sınıflandırıldıktan sonra, diskriminant fonksiyonunun başarısı, doğru sınıflandırma yüzdesi incelenerek tespit edilebilir.

Özetleyecek olursak; Diskriminant Analizi X veri setindeki değişkenlerin iki veya daha fazla gerçek gruba ayrılmasını sağlayabilmek amacıyla yararlanılan bir yöntemdir. Araştırmacının p tane özelliği bilinen birimleri doğal ortamdaki gerçek guruplarına, sınıflarına optimal düzeyde atamasını sağlayacak fonksiyonlar bulmasına yarayan bir yöntemdir. Elde edilebilecek somut özetleyici bilgiler açısından istatistiksel değerlendirmede önemli bir konudur. Çünkü hatalı sınıflandırma olasılığını en aza indirgeyerek birimleri ait oldukları guruplara ayırır, ait oldukları ana kütleleri belirler (Cangül, 2006).

Bu tekniğin yarar sağlayacağı alanlar arasında, personel yerleştirme testleri, çocukların psikolojik testleri, tıbbi tedavilerin etkileri, coğrafi bölgeler arasındaki ekonomik farklılıklar, seçim sonuçlarını tahmin etme, finansal başarısızlık tahmini çalışmaları ve benzerleri yer almaktadır. Gereken tek şey ise bazı değişkenlerde

farklılık gösteren ve bir aralık veya orana göre ölçülebilen iki veya daha fazla grubun var olmasıdır. Diskriminant Analizi bize bu gruplar arasındaki farklılıkları analiz etmekte ve herhangi bir yeni durumun en uygun olan gruba yerleşmesinde yardımcı olur (Cangül, 2006).

Finansal kuruluşların analistleri daima kredi vermek için kurallar ve prensipler kullanırlar. Ekonomik ve insangücü koşullarında kredi isteyen kişi sayısındaki artıştan dolayı bu işlemin manuel yürütülmesi mümkün değildi. Dolayısıyla kredi onay karar süreci gibi bir sistem otomasyonuna ihtiyaç duyuldu. Sonuçta kredi kararları istatistiksel metodlar, parametrik olmayan istatistiksel metodlar ve yapay zeka yaklaşımları ile desteklenerek sunulmuştur. Model inşaa edilirken doğrusal istatistiksel araçlardan olan diskriminant analizi uygulanmıştır. Diskriminant analizi, kredi verilerine göre sınıflandırma yaparak verinin iyi ve kötü kredi sınıfları olarak ayrılmasına ve bir sonraki aşamaya hazır hale getirilmesi sağlar (Lee ve diğ., 2002). Bunların yanında, diskriminant analizi kurumsal bankacılık sektöründe kredi notu belirlemede kullanılan modelin geçerliliğini denetleme adımında uygulanan yöntemlerden olup, gözlemleri iki farklı sınıfa ayırarak problem hakkında tahmin yapmaya olanak sağlar (Emel ve diğ., 2003). Bunların yanında kredi puanlama ve kredi risk hesaplama sistemlerinde geleneksel istatistik tekniklerinden olan diskriminant analizi kullanılmıştır.

4. KREDİ KARTI SAHTECİLİK TESPİT SİSTEMLERİ ÜZERİNE BİR UYGULAMA

4.1 Araştırmanın Amacı

Kredi kartı sahteciliği, son yıllarda gittikçe artan bir hızla büyümekte, şirketleri bu sahteciliğe karşı her geçen gün yeni araştırmalar yapmaya ve yeni sistemler geliştirmeye itmektedir. Yapılan bu çalışmalar ve araştırmalar süresince farklı istatistiksel teknikler kullanılmakta olup farklı sonuçlar elde edilmektedir. Elde edilen bu sonuçlar ile şirketler kendi savunma mekanizmalarını inşaa etmekte, kredi kartı sahteciliğine karşı etkin çözümler geliştirmektedirler.

Bu çalışmada kredi kartı sahtecilik tespitini en etkin şekilde gerçekleştiren metodu bulabilmek için veri madenciliği tekniklerinin yanı sıra klasik istatistiksel teknikler de kullanılmıştır. Veri madenciliği tekniklerinden sinir ağları kapsamında Çok Katmanlı Perseptron (Multilayer Perceptron) ve Destek Vektör Makinelerini, klasik istatistiksel teknikler olarakta Lojistik Regresyon ve Diskriminant Analizini kullanarak analizlerimiz gerçekleştirilmiştir.

Araştırma; kredi kartı sahtecilik tespiti için gerçek kredi kartı harcamaları üzerinden gidilerek yapay zeka ve istatistiksel teknikleri uygulamak suretiyle, kullanılan tekniklerden elde edilen doğru sınıflandırma oranı, birinci tip hata ve ikinci tip hata değerlerini kıyaslayarak en doğru yöntemi bulmak amacıyla yapılmıştır.

4.2 Araştırmanın Sınırları

Bu araştırma, bankacılık sektöründe önemli bir yere sahip olan kredi kartı ve kartın kullanımı sonucu meydana gelen harcama kayıtları üzerinde yapıldığından bankacılık sektörü içinde geçerli olmaktadır. Fakat benzer çalışma koşulları banka dışında çeşitli kurum ve kuruluşlarda da geçerli olabileceğinden çalışmamız bu tür yapılara da uygun bir model olabilir.

Araştırmada bulunan sonuçlar, bankacılık sektöründe kredi kartı sahteciliğine karşı alınması gereken önlemlerle ilgilidir. Bu önlemlerle birlikte kredi kartı sahtecilik tespiti için en doğru sonucu veren teknik belirlenerek, kurulan model açıklanmıştır.

4.3 Örneklem

Analizlerimizde kullandığımız veri setimiz bir finans kurumundan temin edilmiştir. Toplam 4637 kredi kartı harcama işleminin bulunduğu veri setimizde 4523 normal, 114 sahtecilik şüphesi olan kayıt bulunmaktadır. Veri setinin % 70'i modeli oluşturmak (eğitmek) geri kalan % 30'u ise oluşturulan modeli test etmek için kullanılmıştır. Modellerin tahmin güçlerini doğru bir şekilde karşılaştırabilmek için bütün analizlerde aynı veri seti kullanılmıştır.

Bu çalışmada kullandığımız veri setimizde bağımlı değişken fraud değişkeni, bağımsız değişkenler ise TLTutar, GünlükToplamTutar, GünlükToplamKayıt, pos_no, kart_no, Or_para_kodu ve kart_tipi değişkenleridir. Bunun yanında analizleri gerçekleştirirken Or_para_kodu, kart_tipi ve fraud değişkenlerini nominal tipinde, diğer değişkenlerimizi ise nümerik tipinde alarak analizlerde kullanılmıştır. Veri setimizdeki değişkenlerimiz Şekil 4.1'de görülmektedir.

No.		Name
1	☐	TLTutar
2	☐	GünlükToplamTutar
3	☐	GünlükToplamKayıt
4	☐	pos_no
5	☐	kart_no
6	☐	Or_para_kodu
7	☐	kart_tipi
8	☐	fraud

Şekil 4.1 Analizlerde Kullanılan Veri Setini Oluşturan Değişkenler

4.4 Yöntem

Analizler için en önemli kriter hata oranlarıdır. Bu hata oranları iki tip şeklinde sınıflandırılabilirler. Hatalardan biri olan kötü hata oranı, kötü müşterilerin doğru tahmin edilen gözlemleridir. Diğeri iyi hata oranı ise iyi müşterilerin hatalı

tahmin edilen gözlemleridir. Sonuçları değerlendirmek için tasvir edilen matris modeli iki sınıf hata oranını karşılaştırmaya karar verir (Li ve diğ., 2004). Yapılan analizlerde bu matrislerden faydalanıp, yorumlar bu çerçevede yapılmaktadır.

Analizlerimiz boyunca iyi bir kıyas yapabilmek için doğru sınıflandırma oranı, birinci tip hata ve ikinci tip hata değerleri kullanılmıştır. Doğru sınıflandırma oranı veri setimizdeki gerçek sınıflandırmanın analiz sonucunda öngörülen sınıflandırma ile aynı değeri veren kayıtların toplamının veri seti içindeki tüm toplama oranıdır. Yani gerçekte normal olan bir kaydın analiz sonucunda normal yada şüpheli olan bir kaydın analiz sonucunda şüpheli çıkan rakamlarının toplamının tüm kayıt sayısına oranıdır. Birinci tip hata kaydın kendisinin normal bir kayıt olduğu halde analiz sonucu bu kaydın sahtecilik şüphesi taşıdığı, ikinci tip hata ise kaydın kendisi sahtecilik şüphesi taşıdığı halde analiz sonucu normal bir kayıt olarak tanımlanmasıdır. Analiz sonucu, finansal kuruluş her iki tip hata durumda da kaynakları etkin kullanamadığından zarar görmektedir.

4.5 Veri Analizi

Veri seti üzerinde yaptığımız analizleri Weka (Witten ve diğ., 2005) veri madenciliği programının 3.5.6 versiyonu ve SPSS istatistik analiz programının 13 versiyonu kullanılarak gerçekleştirilmiştir.

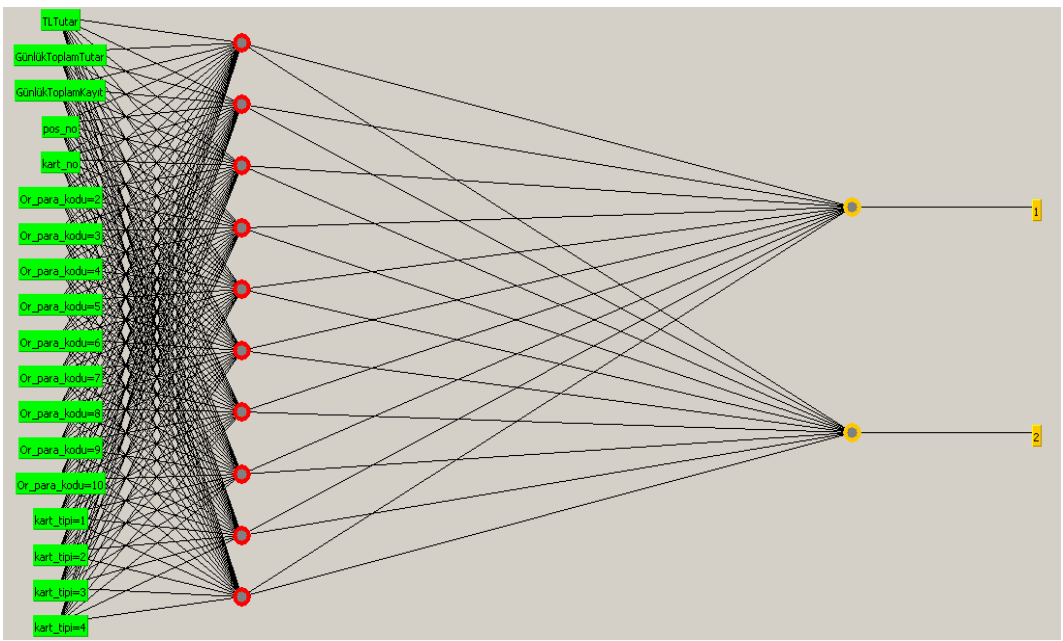
Weka, veri seti üzerinde kolaylıkla uygulanabilen öğrenme algoritmalarının gerçekleştirilmesini sağlar. Aynı zamanda veri setlerini dönüştürmek için çok çeşitli araçlar içermektedir. Weka analiz programı standart veri madenciliği problemleri için regresyon, sınıflandırma, kümeleme, birleştirilmiş kural madenciliği ve özellik seçme gibi metodlarını içermektedir (Witten ve diğ., 2005). Sinir ağları, destek vektör makineleri ve lojistik regresyon analizlerini Weka veri madenciliği programı, diskriminant analizini ise SPSS istatistik analiz programı kullanılarak gerçekleştirilmiştir. Kullanılan yöntemler ve elde edilen sonuçlar şu şekilde ifade edilmiştir.

4.6 Bulgular

İleriki bölümlerde yapay sinir ağları, destek vektör makineleri, lojistik regresyon ve diskriminant analizi sonucunda elde edilen bulgular yorumlarıyla birlikte detaylı olarak açıklanacaktır.

4.6.1 Yapay Sinir Ağları (Çok Katmanlı Perseptron Algoritması)

Yapay sinir ağları analizi için geri beslemeli algoritma kullanılmıştır. Yapay sinir ağları tekniğinin sahip olduğu bazı serbest parametreler bulunmaktadır. Bu parametreler algoritma eğitime başlamadan önce tanımlanması gereklidir. Yapay sinir ağlarında karşımıza çıkan en önemli zorluk, optimum parametre seçimidir. Optimum parametre seçimi için şu yöntem izlenmiştir. Algoritma eğitime başlamadan önce bazı parametreler belirtilmelidir. Bu parametrelerden öğrenme oranı en önemli serbest parametrelerdendir. Öğrenme oranı parametresi 0.25 ile 0.50 arasında 6 tane farklı değer alınmıştır. Bir diğer serbest parametre olan iterasyon sayısı parametresi ise 2500 iterasyon olarak belirtilmiştir. GUI parametresinin değeri de True olarak seçilmiştir. Analiz sırasında sistemin oluşturduğu sinir ağı haritası ve sonuçları aşağıdaki gibidir.



Şekil 4.2 Multilayer Perseptron Analizinde Programın Oluşturmuş Olduğu Yapay Sinir Ağı Haritası

Analiz sonucunda öğrenme oranı parametre değerlerine karşılık gelen doğru sınıflandırma oranları ise Tablo 4.1'de gösterilmektedir.

Tablo 4.1 Öğrenme Oranı Parametre Değerine Karşılık Doğru Sınıflandırma Oranları

Öğrenme Oranı	**Doğru Sınıflandırma Oranı**
0.25	98.78 %
0.30	98.71 %
0.35	98.56 %
0.40	98.85 %
0.45	**98.92 %**
0.50	98.56 %

Tablo 4.1'de görüldüğü gibi, öğrenme oranı parametresi 0.25 ile 0.50 arasındaki değerler ile denenerek test edilmiştir. En iyi sonucun öğrenme oranı parametresine 0.45 değerini set ederek yapmış olduğumuz analiz sırasında elde edildiği görülmüştür. Böylece doğru sınıflandırma oranı en yüksek olan değer elde edilmiştir. % 98.92 doğru sınıflandırma oranını veren analizin karmaşıklık matrisi Tablo 4.2'de gösterilmektedir.

Tablo 4.2 En Yüksek Doğru Sınıflandırma Oranını veren Multilayer Perseptron Analizinin Karmaşıklık Matrisi Değerleri

Karmaşıklık Matrisi		
Tahmin / Gerçek	**Normal**	**Şüpheli**
Normal	1361	2
Şüpheli	13	16

Tablo 4.2'de görünen matrisi yorumlamak gerekirse; modelimiz normal olan 1361 kaydı normal, normal olan 2 kaydı da şüpheli olarak göstermiştir. Buna karşın şüpheli olan 13 kaydı normal, şüphesi olan 16 kaydı ise şüpheli kayıt olarak göstermiştir. Analiz sonucunda doğru tahmin edilen gözlemlerin, bütün gözlemlere

oranı yani doğru sınıflandırma oranı değeri % 98.92 olarak elde edilmiştir. Normal gözlemlerin şüpheli tahmin edilenlerinin sayısının toplam normal gözlemlere oranı olan birinci tip hata değeri % 0.14, şüpheli gözlemlerin normal tahmin edilenlerinin sayısının şüpheli gözlemlere oranı olan ikinci tip hata değeri ise % 44.82 olarak bulunmuştur.

4.6.2 Destek Vektör Makineleri

Destek vektör makineleri analizi için SMO (Sequential Minimal Optimization) algoritması kullanılmıştır. Yapay sinir ağlarında olduğu gibi destek vektör makinelerinde de serbest parametreler bulunmaktadır. Bu parametrelerden Kernel ve c karmaşıklık parametresi en önemli serbest parametrelerdendir. c karmaşıklık parametresi için (1, 5, 10, 15, 20) değerleri ve kernel parametresi içindeki RBF Kernel metodunun gamma pararmetresi için ise (0.01, 0.05, 0.10, 0.20) değerleri seçilerek en doğru analizi yapan ikili bulunmaya çalışılmıştır. Analiz sırasında sistemin meydana getirdiği doğru sınıflandırma oranları Tablo 4.3'te gösterilmektedir.

Tablo 4.3 SVM için yapılan SMO Analizinin Sonucunda En Yüksek Doğru Sınıflandırma Oranını Veren İkilinin Analiz Sonuçları

c / **gamma**	**1**	**5**	**10**	**15**	**20**
0.01	97.91	98.56	98.56	98.56	98.56
0.05	98.56	98.56	98.56	98.42	98.42
0.10	98.56	98.56	98.42	98.42	98.49
0.20	98.56	98.56	**98.64**	**98.64**	**98.64**

Tablo 4.3'te görüldüğü gibi analiz sonucunda en iyi doğru sınıflandırma oranını veren RBF Kernel fonksiyonunun gamma parametresi değeri 0.20 ve karmaşıklık parametresi değerleri ise 10, 15, 20 olarak bulunmuştur. En yüksek

doğru sınıflandırma oranını veren analizin karmaşıklık matrisi ise Tablo 4.4'te gösterilmektedir.

Tablo 4.4 En Yüksek Doğru Sınıflandırma Oranını Veren SMO Analizinin Karmaşıklık Matris Değerleri

Karmaşıklık Matrisi		
Tahmin / Gerçek	**Normal**	**Şüpheli**
Normal	1362	1
Şüpheli	18	11

Tablo 4.4'te görüldüğü gibi modelimiz normal olan 1362 kaydı normal, normal olan 1 kaydı şüpheli olarak göstermiştir. Buna karşın şüpheli olan 18 kaydı normal, şüpheli olan 11 kaydı ise şüpheli kayıt olarak göstermiştir. Analiz sonucunda doğru tahmin edilen gözlemlerin, bütün gözlemlere oranı yani doğru sınıflandırma oranı değeri % 98.64 olarak elde edilmiştir. Normal gözlemlerin şüpheli tahmin edilenlerinin sayısının toplam normal gözlemlere oranı olan birinci tip hata değeri % 0.07, şüpheli gözlemlerin normal tahmin edilenlerinin sayısının şüpheli gözlemlere oranı olan ikinci tip hata değeri ise % 62.06 olarak bulunmuştur.

4.6.3 Lojistik Regresyon

Lojistik regresyon analizi için Weka (Witten ve diğ., 2005) veri madenciliği programı kullanılarak analiz yapılmıştır. Analizde değişken seçim yöntemi olan adım adım (stepwise) analiz uygulanmıştır. Adım adım analiz bize model oluşturmada en iyi sonucu verecek değişkenleri belirlemede yardımcı olmaktadır. Bunun sonucunda modelde gereksiz değişken yer almamakta ve modelin fonksiyonel ilişkiyi açıklama gücü yüksek olmaktadır. En iyi model az değişkenle açıklanabilen ve amaca uygun olan modeldir (Lee ve diğ., 2006).

Logistic fonksiyonunun varsayılan olarak gelen parametreleri üzerinde herhangi bir değişiklik yapmadan analizimizi gerçekleştirdiğimiz için sadece tek bir

doğru sınıflandırma oranı elde edilmiştir. Analiz sırasında sistemin oluşturduğu karmaşıklık matrisi Tablo 4.5'te gösterilmektedir.

Tablo 4.5 Logistic Fonksiyonu Kullanarak Elde Edilen Karmaşıklık Matrisi Değerleri

Karmaşıklık Matrisi		
Tahmin / Gerçek	**Normal**	**Şüpheli**
Normal	1360	3
Şüpheli	16	13

Tablo 4.5'te görüldüğü gibi modelimiz normal olan 1360 kaydı normal, normal olan 3 kaydı da şüpheli olarak göstermiştir. Buna karşın şüpheli olan 16 kaydı normal, şüpheli olan 13 kaydı ise şüpheli kayıt olarak göstermiştir. Analiz sonucunda doğru tahmin edilen gözlemlerin, bütün gözlemlere oranı yani doğru sınıflandırma oranı değeri % 98.63 olarak elde edilmiştir. Normal gözlemlerin şüpheli tahmin edilenlerinin sayısının toplam normal gözlemlere oranı olan birinci tip hata değeri % 0.22, şüpheli gözlemlerin normal tahmin edilenlerinin sayısının şüpheli gözlemlere oranı olan ikinci tip hata değeri ise % 55.17 olarak bulunmuştur.

4.6.4 Diskriminant Analizi

Diskriminant analizi için SPSS 13 istatistiksel paket programı kullanılmıştır. Diskriminant analizinde, kredi kartı sahtecilik tespit modelini inşaa etmede en iyi sonucu verecek değişken seçim prosedürü olan adım adım (stepwise) analiz uygulanmıştır (Lee ve diğ., 2006).

Adım adım analiz yaklaşımında SPSS'te Analyze – Classify yolu izlenerek Discriminant fonksiyonu seçilmiş ve gelen ekranda bağımlı değişken olan fraud değişkeni gruplanacak değişken, diğer değişkenler ise bağımsız değişken olarak seçildikten sonra analiz işlemi gerçekleştirilmiştir. Analiz sırasında sistemin oluşturduğu karmaşıklık matrisi Tablo 4.6'da gösterilmektedir.

Tablo 4.6 Diskriminant Fonksiyonu Kullanarak Elde Ettiğimiz Karmaşıklık Matris Değerleri

Karmaşıklık Matrisi		
Tahmin / Gerçek	**Normal**	**Şüpheli**
Normal	1279	23
Şüpheli	21	13

Tablo 4.6'da görüldüğü gibi model normal olan 1279 kaydı normal, normal olan 23 kaydı da şüpheli olarak göstermiştir. Buna karşın şüpheli olan 21 kaydı normal, şüpheli olan 13 kaydı ise şüpheli kayıt olarak göstermiştir. Analiz sonucunda doğru tahmin edilen gözlemlerin, bütün gözlemlere oranı yani doğru sınıflandırma oranı değeri % 96.48 olarak elde edilmiştir. Normal gözlemlerin şüpheli tahmin edilenlerinin sayısının toplam normal gözlemlere oranı olan birinci tip hata değeri % 1.76, şüpheli gözlemlerin normal tahmin edilenlerinin sayısının şüpheli gözlemlere oranı olan ikinci tip hata değeri ise % 61.76 olarak bulunmuştur.

4.6.5 Analizlerin Karşılaştırılması

Bu bölümde uygulamayı gerçekleştirdiğimiz yapay sinir ağları, destek vektör makineleri, lojistik regresyon ve diskriminant analizi tekniklerini kullanarak gerçekleştirmiş olduğumuz sahtecilik tespit modelleri karşılaştırılacaktır. Karşılaştırma için her bir modelin en iyi sonuçları alınmıştır. Tablo 4.7'de her bir model için en iyi analizler sonucu ortaya çıkan doğru sınıflandırma oranlarını ve karmaşıklık matrisi değerlerini karşılaştırmalı olarak görmekteyiz.

Tablo 4.7 Karşılaştırmalı Karmaşıklık Matris Değerleri ve Doğru Sınıflandırma Oranları

			Gerçek		**Doğru Sınıflandırma Oranı**
			1	2	
YSA	**Tahmin**	1	1361	13	98.92 %
		2	2	16	
SVM	**Tahmin**	1	1362	18	98.64 %
		2	1	11	
Lojistik Regresyon	**Tahmin**	1	1360	16	98,64 %
		2	3	13	
Diskriminant Analizi	**Tahmin**	1	1279	21	96,70 %
		2	23	13	

Tablo 4.7'de görüldüğü gibi doğru sınıflandırma oranına bakılarak analiz yapmak gerekirse destek vektör makineleri ve lojistik regresyon aynı doğruluk oranını vermektedir. Kullanılan tüm teknikleri bir arada analiz etmek gerekirse doğru sınıflandırma oranına bakıldığında sırasıyla yapay sinir ağları, destek vektör makineleri, lojistik regresyon ve diskriminant analizi olarak sıralandığı görülmektedir.

İkinci bir kriter olarak birinci tip hata (Type I) ve ikinci tip hata (Type II) kriter değerleri kullanılmaktadır. Tablo 4.8'de bu değerler gösterilmektedir.

Tablo 4.8 Karşılaştırmalı olarak Birinci Tip Hata ve İkinci Tip Hata Değerleri

	Type I	**Type II**
YSA	0.14 %	44.82 %
SVM	0.07 %	62.06 %
Lojistik Regresyon	0.22 %	55.17 %
Diskriminant Analizi	1.76 %	61.76 %

Tablo 4.8'de görüldüğü gibi birinci tip hataya bakıldığında ise en iyi tahmin modelinin sırasıyla destek vektör makineleri, yapay sinir ağları, lojistik regresyon ve

diskriminant analizi olduğu görülmektedir. İkinci tip hataya bakıldığında ise en iyi tahmin edilen modelin yapay sinir ağları olduğu devamında lojistik regresyon, diakriminant analizi ve destek vektör makineleri gelmektedir.

Buradan görüldüğü gibi birinci tip hatanın mı yoksa ikinci tip hatanın mı yüksek olmasını istiyoruz. İkinci tip hata, birinci tip hataya göre daha maliyetlidir. Bu sebepten herhangi bir hatayı minimum kılarken diğeri maksimum olabilir. Modeli kurarken asıl amaç iki hata oranı değerleri arasında denge sağlayıp, her iki hata değerini de optimum yapan modeli oluşturmaktır.

Tablo 4.9 Doğru Sınıflama, Birinci Tip Hata ve İkinci Tip Hata Gözlem Sayıları

	Doğru Sınıflama Sayısı	**Birinci Tip Hata Sayısı**	**İkinci Tip Hata Sayısı**
YSA	1377	2	13
SVM	1373	1	18
Lojistik Regresyon	1373	3	16
Diskriminant Analizi	1292	23	21

Tablo 4.9’da ise kullanmış olduğumuz tahmin modellerinin doğru sınıflama sayıları, birinci tip hata ve ikinci tip hata sayıları verilmiştir. Yukarıda yapmış olduğumuz yorumları bu tabloya da bakarak çıkarabiliriz.

Sonuç olarak, kullandığımız tüm bu teknikler neticesinde elde edilen sonuçlarla ilgili yorum yapmak gerekirse; Tablo 4.9’da görüldüğü gibi kredi kartı sahtecilik tespitinde diskriminant analizini kullanmak sakıncalı olabilir. Çünkü diskriminant analizinin çıktılarına baktığımız zaman rakamsal olarak her açıdan olumsuz sonuçlar görülmektedir. Bunun yanında sinir ağları, destek vektör makineleri ve lojistik regresyon kullanılarak yapılan analizler sonucunda birbirine yakın sonuçlar elde edilmiş olup, yeri geldiği zaman her üç tekniğinde kullanılmaya elverişli olduğu gösterilmiştir.

5. SONUÇ VE ÖNERİLER

Bu çalışmamızda, kredi kartlarında sahtecilik tespit sistemleri konusu ayrıntılı olarak ele alınmış, veri madenciliği ve istatistiksel teknikler kullanılarak inceleme yapılmış, çıkan sonuçlar ise doğru sınıflandırma oranı, birinci tip hata ve ikinci tip hata kriterlerine göre yorumlanarak hangi yöntemin en iyi sonuç verdiği ortaya konulmuştur.

Kredi kartlarında sahteciliğinin çarpıcı bir şekilde artması her yıl dünya çapında milyonlarca doların kaybedilmesine sebep olmakta, bununla beraber birçok iş alanında sahtecilik için modern teknikler geliştirilmekte ve bu alanlarda uygulanmaktadır. Çalışmamıza başlarken sahtecilik kavramı ile ilgili geniş bilgi verilmiştir. Daha sonra kredi kartları, sahtecilik çeşitleri ve sahtecilik yöntemleri ayrıntılı olarak açıklanmıştır. Sahteciliğin kapsamını belirttikten sonra işin teori ve analiz kısmına giriş yapılmıştır. Burada gerek veri madenciliği gerekse istatistiksel teknikleri detaylı olarak inceleme fırsatı bulup, hem teorik bilgiyi hemde kredi kartlarında sahtecilik tespiti ile ilgili makale ve kaynaklardan yararlanılmıştır. Bilimsel çalışmalarda yapay zeka sistemleri, veri madenciliği kısmının temelini teşkil etmekle beraber istatistiksel teknikler olarakta, lojistik regresyon ve diskriminant analizi kullanılmıştır. Yapay zeka kavramının içinde yapay sinir ağları, ANFIS, destek vektör makineleri ve kural tabanlı öğrenme yapıları hakkında da geniş bilgi verilmiştir.

Uygulama kısmında bir finans kurumundan elde ettiğimiz gerçek kredi kartı harcama kayıtları üzerinden, yapay zeka ve istatistiksel teknikler kullanılarak analizlerimizi gerçekleştirdik. Veri setimizde bulunan niteliklerden biri, kaydın normal veya şüpheli olduğunu belirtmektedir. Analizlerimizi yaparken yapay sinir ağları, destek vektör makineleri, lojistik regresyon ve diskriminant analizi teknikleri kullanılmış ve analiz sonucunda elde edilen doğru sınıflandırma oranı, birinci tip hata ve ikinci tip hata değerleri kullanılarak yorumlar gerçekleştirilmiştir. Doğru sınıflandırma oranına bakıldığında en doğru tahmin edilen modelin sırasıyla yapay sinir ağları, destek vektör makineleri, lojistik regresyon ve diskriminant analizi olduğu görülmüştür. Birinci tip hataya göre en doğru tahmin edilen modelin sırasıyla

destek vektör makineleri, yapay sinir ağları, lojistik regresyon ve diskriminant analizi olduğu, ikinci tip hataya bakıldığında ise en iyi tahmin edilen modelin yapay sinir ağları, lojistik regresyon, destek vektör makineleri ve diskriminant analizi olduğu görülmektedir.

Çalışmalarımızı yaparken veri setimizdeki bir nitelik ile kaydın normal yada şüpheli olduğunu bildiğimiz için danışmalı öğrenme algoritmaları kullanılmıştır. Aslında böyle bir nitelik kullanmadan yani danışmasız öğrenme algoritmalarını kullanarakta çok çarpıcı sonuçlar ortaya çıkabilir. Bunun yanında melez öğrenme algoritmaları kullanarak yani veri setimiz üzerinde önce gruplandırma sonra sınıflandırma yaparak farklı sonuçlar elde edilebilir. Melez öğrenme algoritmasına ANFIS güzel bir örnek olarak gösterebilir.

Kredi kartı sahtecilik tespiti için kullanabileceğimiz bir diğer yöntem ise ensemble algoritmalarıdır. Ensemble algoritmaları genellikle saldırı tespit sistemleri, anormal tespitlerinde ve kredi kartı uygulamalarında çok daha fazla kullanılan bir yöntemdir. Kredi kartı sahteciliğinde kullanılması doğruluk oranı daha büyük farklı sonuçların ortaya çıkmasına sebep olabilir. Bir başka önemli yöntem aykırı, sıradışı örnekleri ortaya çıkarmak için kullanılan öğrenme algoritmalarıdır. Örnek olarak aykırı örnek tespiti (outlier detection) verilebilir. Bu algoritma geniş veritabanları içinde farklı olan gözlemlerin ayrıştırılması esasına dayanır ve bu yöntemdeki asıl amaç, sıradışı olan gözlemleri bulmak olduğu için bu ismi almıştır. Sıradışı tespit sistemini hem danışmalı öğrenmede, hemde danışmasız öğrenmede kolaylıkla uygulanabilen bir yöntem olmakla birlikte kredi kartı sahteciliğinde de etkin bir şekilde kullanılabileceği düşünülmektedir.

Uygulama sonucuna göre, kullandığımız tekniklerden yapay sinir ağları, destek vektör makineleri ve lojistik regresyon tekniklerini kullanarak birbirine çok yakın sonuçlar elde edildiğinden dolayı her üç tekniğinde kredi kartı sahtecilik tespiti için kullanılmaya elverişli teknikler olduğu görülmüştür. Bir diğer teknik olan diskriminant analizinin sonuçlarından görüldüğü gibi kredi kartı sahtecilik tespiti için kullanılmasının sakıncalı olabileceği sonucuna varılmıştır.

KAYNAKÇA

Ahi M.G, (2008), "Kredi Kartı Sahteciliği", http://hukukcu.com/modules/smartsection/item.php?itemid=72, (03.06.2008)

Akpınar H, (2000), "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği" İstanbul Üniversitesi İşletme Fakültesi Dergisi, c.29, s.1-22

Aleskerov E, Fieisleben B, Rao B, (1997), "CARDWATCH: A Neural Network Based Database Mining System for Credit Card Fraud Detection" Proceedings of the IEEE/ IAFE, s.220-226

Alfuraih S.I, Sui N.T, Mcleod D, (2002), "Using Trusted Email to Prevent Credit Card Frauds in Multimedia Products", Internet and Web Information Systems, vol.5, s.245-256

Alliston M, (2002), "Modelling Credit Card Fraud" NY INFORMS Chapter, October

Altıntaş E, (2008), "Yapay Sinir Ağları (Artificial Neural Networks)" http://www.yapay-zeka.org/modules/wiwimod/index.php?page=ANN, (01.04.2008)

Amasyalı M.F, Diri B, Türkoğlu F, (2006), "Farklı Özellik Vektörleri ile Türkçe Dokümanların Yazarlarının Belirlenmesi" 15.Türkiye Yapay Zeka ve Sinir Ağları Sempozyumu - TAINN

Atan M, Çatalbaş E, (2004), "Çok Değişkenli İstatistiksel Analiz Yöntemleri ile Türk Bankacılık Sektöründe Çok Boyutlu Mali Başarısızlık Tahmin Modelleri Oluşturulması" 4. İstatistik Günleri Sempozyumu İzmir

Bankalararası Kart Merkezi, (2008), http://www.bkm.com.tr/bkm.html, (24.05.2008)

Bankalararası Kart Merkezi, (2008), http://www.bkm.com.tr/istatistik/raporlar1.html, (24.05.2008)

Bankalararası Kart Merkezi, (2008), Faaliyet Raporu 2006, http://www.bkm.com.tr/faaliyetraporu.html, (25.05.2008)

Bankalararası Kart Merkezi, (2008), Faaliyet Raporu 2007, http://www.bkm.com.tr/faaliyetraporu.html, (25.05.2008)

Bankalararası Kart Merkezi, (2008), http://www.bkm.com.tr/basin/bultenler/chip_basin_toplantisi_10052005.pdf, (24.05.2008)

Bhatla T.P, Prabhu V, Dua A, (2003), "Understanding Credit Card Frauds" Card Business Review

Bircan H, (2004), “Lojistik Regresyon Analizi: Tıp Verileri Üzerine Bir Uygulama” Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, vol.2, s.185-208

Brause R, Langsdorf T, Hepp M, (1999), “Neural Data Mining for Credit Card Fraud Detection” J.W.Goethe-University, Computer Science Department Report, Frankfurt-Germany

Bolton R.J, Hand D.J, (2002), “Statistical Fraud Detection: A Review” Statistical Science, vol.17, no.3, s.235-255

Burges C.J.C, (1998), “A Tutorial on Support Vector Machines for Pattern Recognition” Data Mining and Knowledge Discovery, vol.2, s.121-167

Cangül O, (2006), “Diskriminant Analizi ve Bir Uygulama Denemesi” Uludağ Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı İstatistik Bilim Dalı Yüksek Lisans Tezi, Bursa

Chiu C.C, Tsai C.Y, (2004), “A Web Services-Based Collaborative Scheme for Credit Card Fraud Detection” Proceedings of the IEEE International Conference on e-Technology, e-Commerce and e-Service

Çavuş M.F, (2006), “Bireysel Finansmanın Temininde Kredi Kartları: Türkiye’de Kredi Kartı Kullanımı Üzerine Bir Araştırma” Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, Konya, s.173-187

Çinko M, (2006), “Kredi Kartı Değerlendirme Tekniklerinin Karşılaştırılması” İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi, yıl:5, sayı:9, s.143-153

Demirci D.A, (2007), “Destek Vektör Makineleri ile Karakter Tanıma” Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi, İstanbul

Elmas Ç, (2007), “Yapay Zeka Uygulamaları” Seçkin Yayıncılık San. ve Tic. A.Ş, Kasım - 2007

Emel A.B, Oral M, Reisman A, Yolalan R, (2003), “A Credit Scoring for the Commercial Banking Sector” Socio-Economic Planning Sciences, vol.37, s.103-123

Erik F, Tjong K.S, (1998), “Machine Learning of Phonotactics” University of Groningen, The Netherlands, s.95-129

Fischer M, (2005), “Automatic Fraud Detection” MIKAN Solutions

Ghosh S, Reilly D.L, (1994), “Credit Card Fraud Detection with a Neural-Network” Proceedings of the Twenty-Seventh Hawaii International Conference, vol.3, p.621-630

Haykin S, (1999), “Neural Networks: A Comprehensive Foundation” Prentice Hall, s.173-175

Hoffmann F, Beasens B, Martens J, Put F, Vanthienen J, (2002), "Comparing a Genetic Fuzzy and a Neurofuzzy Classifier for Credit Scoring" International Journal of Intelligent Systems, vol.17, s.1067-1083

Hsu C.M, Chao H.M, (2007), "An Online Fraud-Resistant Technology for Credit Card E-Transactions", IEEE, s.1-4

Huang C.L, Chen M.C, Wang C.J, (2007), "Credit Scoring with a Data Mining Approach Based on Support Vector Machines" Expert Systems with Applications vol.33, s.847–856

Ince H, Trafalis T.B, (2006), "Kernel Methods for Short-Term Portfolio Management" Expert Systems with Applications vol.30, s.535–542

Ince H, Trafalis T.B, (2006), "A Hybrid Model for Exchange Rate Prediction" Decision Support Systems vol.42, s.1054–1062

"İnternette Kredi Kartı Sahtekarlığı", (2008), http://eticaret.garanti.com.tr/icerik/goster.asp?t=a&c=3&i=46, (02.06.2008)

Jang J.S.R, (1993), "ANFIS: Adaptive Network-Based Fuzzy Inference System" IEEE Transaction on Systems, Man. and Cybernetics, vol.23, no.3, s.665-685

Kaylan K, (2008), "Kamu Güvenine Karşı Suçlar" http://www.ceza-bb.adalet.gov.tr/makale/104.doc, (03.04.2008)

"Kredi kartı nedir? Kredi kartı tanımı.", (2008), http://www.tuketicifinansman.net/2007/09/kredi-karti-nedir-kredi-kartlarinin.html, (11.05.2008)

Kirkos E, Spathis C, Manolopoulos Y, (2007), "Data Mining Techniques for the Detection of Fraudulent Financial Statements" Expert Systems with Applications 32, s.995–1003

Kotsiantis S, Koumanakos E, Tzelepis D, Tampakas V, (2006), "Forecasting Fraudulent Financial Statements using Data Mining" International Journal of Computational Intelligence, vol.3, no.1, s.104-110

Kou Y, Lu C.T, Sinvongwattana S, Huang Y.P, (2004), "Survey of Fraud Detection Techniques" International Conference on Networking, Sensing & Control Taipei, Taiwan, s.749-754

Kurt İ, Türe M, (2005), "Yapay Sinir Ağları ile Lojistik Regresyon Analizi'nin Karşılaştırılması" Trakya Üniversitesi Tıp Fakültesi, cilt.22, vol.3, s.142-153

Lee T.S, Chiu C.C, Lu C.J, Chen I-F, (2002), "Credit Scoring Using the Hybrid Neural Discriminant Technique" Expert Systems with Applications 23, s.245-254

Lee T.S, Chiu C.C, Chou Y.C, Lu C.J, (2006), "Mining the customer credit using classification and regression tree and multivariate adaptive regression splines" Computational Statistics & Data Analysis vol.50, s.1113-1130

Lee Y.C, (2007), "Application of Support Vector Machines to Coprorate Credit Rating Prediction" Expert Systems with Applications 33, s.67-74

Li X, Ying W, Tuo J, Li B, Liu W, (2004), "Applications of Classification Trees to Consumer Credit Scoring Methods in Commercial Banks" IEEE International Conference on Systems, Man and Cybemetics, vol.5, s.4112-4117

Özçalık H.R, Uygur A.F, (2003), "Dinamik Sistemlerin Uyumlu Sinirsel-Bulanık Ağ Yapısına Dayalı Etkin Modellenmesi" KSÜ Fen ve Mühendislik Dergisi vol.6, s.36-46

Özekeş S, (2003), "Veri Madenciliği Modelleri ve Uygulama Alanları" İstanbul Ticaret Üniversitesi Dergisi, vol.3, s.65-82

Özkaya A.U, Kaya M.E, Gürgen F, (2005), "Destek Vektör Makineleri Kullanılarak Aritmi Sınıflandırılması" National Symposium on Biomedical Engineering, İstanbul, s.12-16

Özler C, (2006), "Keşfedici Veri Analizi Tekniklerinin Süreç Yetenek Analizi Çalışmalarında Uygulanması" Muğla Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, vol.16

Pang S.L, Wang Y.M, Bai Y.H, (2002), "Credit Scoring Model Based on Neural Network" Proceeding of the First International Conference on Machine Learning and Cybernetics, Beijing, vol.4, s.1742- 1746

Polat G, Altun H, (2007), "Ses Öznitelik Gruplarının Duygu Tespitinde Etkinliklerinin Belirlenmesi", IEEE 15. Sinyal İşleme ve İletişim Uygulamaları Kurultayı, Eskişehir

Quah J.T.S, Sriganesh M, (2007), "Real-Time Credit Card Fraud Detection Using Computational Intelligence" Expert Systems with Applications, s.863-868

Shen A, Tong R, Deng Y, (2007), "Application of Classification Models on Credit Card Fraud Detection" School of Management, Graduate University of the Chinese Academy of Sciences, China, s.1-4

Sığırlı D, (2006), "Sınıflandırma Probleminin Çözümlenmesinde Yapay Sinir Ağları ile Diskriminant Analizinin Karşılaştırılması ve Bir Uygulama" Uludağ Üniversitesi Yüksek Lisans Tezi, Bursa

SPSS, SPSS 13.0 for Windows, http://www.spss.com, (03.04.2008)

Srivastava A, Kundu A, Sural S, Majumdar A.K, (2008), "Credit Card Fraud Detection Using Hidden Markov Model" IEEE Transactions on Dependable and Secure Computing, vol.5, no.1, s.37-48

Şahin B, (2007),"Geri Yayılım Algoritması", http://www.kirbas.com/index.php?id=166, (05.03.2008)

Şahin B, (2008), "Yapay Sinir Ağları ve Uygulamaları" http://ysa.somee.com/download.aspx?id_no=1, (05.03.2008)

Tatlıdil H, (1996), "Uygulamalı Çok Değişkenli İstatistiksel Analiz" Engin Yayınları, sayı.1, s.35-39

Tian X, Deng F, (2004), "A Credit Scoring Model Using Support Vector Machine" 5th World Congress, China, vol.3, s.1945-1949

Ünsal A, (2000), "Diskriminant Analizi ve Uygulaması Üzerine Bir Örnek" Gazi Üniversitesi İİBF Dergisi, cilt.2, sayı.3, s.19-36

Ünsal A, Güler H, (2005), "Türk Bankacılık Sektörünün Lojistik Regresyon ve Diskriminant Analizi ile İncelenmesi" VII. Ulusal Ekonometri ve İst. Sempozyomu, İstanbul

Wheeler R, Aitken S, (2000), "Multiple Algorithms for Fraud Detection" Knowledge-Based Systems vol.13, s.93-99

Wikipedia, (2008), http://tr.wikipedia.org/wiki/Kredi_kartı, (26.05.2008)

Wikipedia, (2008), http://tr.wikipedia.org/wiki/Yapay_sinir_ağları, (26.05.2008)

Witten I.H, Frank E, (2005), "Data Mining Practical Machine Learning Tools and Techniques", Second Edition, Elsevier, s. 62-68

Yazici M, (2008), "Yapay Sinir Ağlarına Genel Bir Bakış" http://www.yapay-sinir-aglari.uzerine.com, (13.07.2008)

Yetgin C, (2008), "Kredi Kartı ve Bilgi Güvenliği" http://www.kpl.gov.tr/tr/kredi_bilgi.htm, (25.05.2008)

Zaslavsky V, Strizhak A, (2006), "Credit Card Fraud Detection Using Self-Organizing Map" Information & Security. An International Journal, vol.18, s.48-63

Zhao H, (2007), "A Multi-Objective Genetic Programming Approach to Developing Pareto Optimal Decision Trees" Decision Support Systems vol.43, s.809-826

ÖZGEÇMİŞ

1979 yılında Diyarbakır'da doğan Yavuz Selim Keresteci, ilk öğrenimi Diyarbakır 5 Nisan İlkokulu'nda, orta öğrenimini Diyarbakır Gazi Ortaokulu'nda ve lise öğrenimini Diyarbakır Ziya Gökalp Lisesi'nde 1996 yılında bitirdi. 1997 yılında İstanbul Üniversitesi Mühendislik Fakültesi Bilgisayar Bilimleri Mühendisliğini kazandı. 2001 yılında adı geçen üniversiteden mezun oldu. 2006 yılında Gebze Yüksek Teknoloji Enstitüsü Sosyal Bilimler Enstitüsü Strateji Bilimi Anabilim Dalı, Bilim ve Teknoloji Stratejileri yüksek lisans programına başladı. Özel bir finans kurumunda çalışma hayatına devam etmektedir.