T.C.
ERCİYES ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI

BİR İPLİK ÜRETİM TESİSİNDE NİTELİK SEÇİMİ VE SINIFLANDIRMA İLE İPLİK KALİTESİNİN BELİRLENMESİ

Hazırlayan
Tayfun ÖZMEN

Danışman
Yrd. Doç. Dr. Pınar Zarif TAPKAN

Yüksek Lisans Tezi

Mayıs 2017
KAYSERİ

T.C.
ERCİYES ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI

BİR İPLİK ÜRETİM TESİSİNDE NİTELİK SEÇİMİ VE SINIFLANDIRMA İLE İPLİK KALİTESİNİN BELİRLENMESİ

(Yüksek Lisans Tezi)

Hazırlayan
Tayfun ÖZMEN

Danışman
Yrd. Doç. Dr. Pınar Zarif TAPKAN

Mayıs 2017
KAYSERİ

BİLİMSEL ETİĞE UYGUNLUK

Bu çalışmadaki tüm bilgilerin, akademik ve etik kurallara uygun bir şekilde elde edildiğini beyan ederim. Aynı zamanda bu kural ve davranışların gerektirdiği gibi, bu çalışmanın özünde olmayan tüm materyal ve sonuçları tam olarak aktardığımı ve referans gösterdiğimi belirtirim.

Tayfun ÖZMEN

YÖNERGEYE UYGUNLUK

Bir İplik Üretim Tesisinde Nitelik Seçimi ve Sınıflandırma İle İplik Kalitesinin Belirlenmesi adlı Yüksek Lisans tezi, Erciyes Üniversitesi Lisansüstü Tez Önerisi ve Tez Yazma Yönergesi'ne uygun olarak hazırlanmıştır.

Tezi Hazırlayan

Tayfun ÖZMEN

Danışman

Yrd. Doç. Dr. Pınar Zarif TAPKAN

Endüstri Mühendisliği ABD Başkanı

Prof. Dr. Mithat ZEYDAN

Yrd. Doç. Dr. Pınar Zarif TAPKAN danışmanlığında **Tayfun ÖZMEN** tarafından hazırlanan "**Bir İplik Üretim Tesisinde Nitelik Seçimi ve Sınıflandırma İle İplik Kalitesinin Belirlenmesi**" adlı bu çalışma, jürimiz tarafından Erciyes Üniversitesi Fen Bilimleri Enstitüsü Endüstri Anabilim Dalında **Yüksek Lisans** tezi olarak kabul edilmiştir.

24 / 05 / 2017

JÜRİ:

Başkan : Yrd. Doç. Dr. Pınar Tapkan

Üye : Doç. Dr. İbrahim Akpon

Üye : Doç. Dr. Sinem Kulluk

ONAY:

Bu tezin kabulü Enstitü Yönetim Kurulunun 30/05/2017 tarih ve 2017/24-06 sayılı kararı ile onaylanmıştır.

31 / 05 / 2017

Prof. Dr. Mehmet AKKURT

Enstitü Müdürü

TEŞEKKÜR

Çalışmalarım boyunca, tez konumun belirlenmesi ve yürütülmesinde değerli görüş ve katkılarıyla beni yönlendiren, her konuda desteğini esirgemeyen, kıymetli tecrübelerinden faydalandığım tez danışmanım Sayın Yrd. Doç. Dr. Pınar Zarif TAPKAN'a teşekkürü bir borç bilirim.

Yüksek Lisans tez çalışmamın her aşamasında beni teşvik eden, anlayış ve desteği ile her zaman yanımda olan eşim Çiğdem ÖZMEN ve bugüne gelmemde büyük emeği geçen aileme saygı ve teşekkürlerimi sunarım.

Çalışmanın geliştirilmesinde yardımlarını esirgemeyen İplik Mühendisi Sayın Ersen GÖK'e, İplik laboratuvar çalışanları Sayın Tuncay SAVAÇ'a ve Sayın Ali Osman ÖZBEK'e yardımlarından dolayı ayrıca teşekkür ederim.

BİR İPLİK ÜRETİM TESİSİNDE NİTELİK SEÇİMİ VE SINIFLANDIRMA İLE İPLİK KALİTESİNİN BELİRLENMESİ

Tayfun ÖZMEN

Erciyes Üniversitesi, Fen Bilimleri Enstitüsü
Yüksek Lisans Tezi, Mayıs 2017
Danışman: Yrd. Doç. Dr. Pınar Zarif TAPKAN

ÖZET

Günümüz bilgisayar teknolojisi hızla ilerlemekte ve bilgisayarların hafıza kapasiteleri her geçen gün artmaktadır. Bilgisayarların hafıza kapasitelerinin artmasıyla birlikte bilgi kaydı yapılan alanların sayısı da artmaktadır ve bu sayede veriye ulaşmanın kolaylaştığı görülmektedir. Bilgisayar teknolojisi ile kaydı yapılan ve üretilen veriler tek başlarına bir anlam ifade etmemektedir. Bu veriler belli bir amaç doğrultusunda işlendiği zaman bir anlam ifade etmeye başlamaktadır. Bu ham veriyi bilgiye veya anlamlı hale dönüştürme işlemleri veri madenciliği ile yapılabilmektedir.

Bu çalışmada veri madenciliği tüm detayları ile incelenip veri madenciliği süreci, modelleri açıklanmış ve bir iplik üretim tesisinde iplik kalitesinin değişmesine sebep olabilecek faktörler belirlenmiştir. Faktörlerin önem sıraları Taguchi yöntemi ile belirlenmiştir. İplik kalitesini etkileyen faktör sayısının fazla olması sebebiyle nitelik seçimi ve sınıflandırma yöntemleri kullanılarak kural oluşturulmuştur. Sınıflandırma işlemi için Weka 3.8.1 ve MT-VeMa 1.0 paket programları kullanılmıştır. Kurallar, kaliteli iplik üretimi için işletmeye yol gösterici olmuştur. Bu çalışma ile veri madenciliği uygulamalarının, bir tekstil şirketinde gerçek verilerle nasıl sonuca ulaştığı gösterilmiş ve ilgili sürece katkıda bulunulmuştur.

Anahtar Sözcükler: Veri Madenciliği, Veri Madenciliği Süreci, Veri Madenciliği Modelleri, Taguchi Yöntemi, Sınıflandırma Kural Çıkarımı, Weka 3.8.1, MT-VeMa 1.0.

DETERMINING THE YARN QUALITY BY FEATURE SELECTION AND CLASSIFICATION IN A YARN PRODUCTION FACILITY

Tayfun ÖZMEN

Erciyes University, Graduate School Natural and Applied Science
M. Sc. Thesis, May 2017
Supervisor: Yrd. Doç. Dr. Pınar Zarif TAPKAN

ABSTRACT

Today's computer technology is advancing rapidly and computers memory capacity is increasing everyday. The number of the fields which are storing information is increasing with the increasing of computers information storage capacity and therefore the facilitator of attaining data. Data which are produced by computers are worthless alone because they are meaningless. These data become meaningful when they are processed for an aim. Changing this raw data to information and to significant state is possible with data mining.

In this study data mining's all of details are researched, data mining's process, models are explained and the factors of the yarn quality caused change are designated in yarn manufacturing plant. Factors order of importance is determined with Taguchi method. Hence the yarn quality is effected from number of factors is much, while qualification election and classification method are used, the rule is created. Weka 3.8.1 and MT-VeMa 1.0 package programs are used for classification method. The rules is a guided to the plant for produced quality yarn. With this study, how data mining's practises are reached the result, is showed with real datas in the textile company and concerning this process are contributed.

Keywords: Data Mining, Data Mining's Process, Data Mining's Models, Taguchi Method, Classification Rule Extraction, Weka 3.8.1, MT-VeMa 1.0.

İÇİNDEKİLER

BİR İPLİK ÜRETİM TESİSİNDE NİTELİK SEÇİMİ VE SINIFLANDIRMA İLE İPLİK KALİTESİNİN BELİRLENMESİ

KISALTMA VE SİMGELER

YSA : Yapay sinir ağları

TAKKO : Tur atan karınca koloni optimizasyonu

KKO : Karınca koloni optimizasyonu

TAKKOYSA-sınıflandırıcısı: Tur atan karınca koloni optimizasyonu ile yapay sinir ağları sınıflandırıcısı

ÇKA : Çok katmanlı algılayıcı

ÇDP : Çoklu denklem programlama

GP : Genetik programlama

DE : Diferansiyel gelişim

tp : Pozitif doğru

fp : Pozitif yanlış

tn : Negatif doğru

fn : Negatif yanlış

KS : Karınca sistemi

KKS : Karınca koloni sistemi

SKKO : Sürekli karınca koloni optimizasyonu

ITKKO : Izgara tabanlı karınca koloni optimizasyonu

DT : Decision Table

S/N : Sinyal gürültü oranı

TD : Test doğruluğu

KS : Kural sayısı

YSM : Yanlış sınıflandırma maliyeti

CS : Cevap verme süresi (saniye)

TABLOLAR LİSTESİ

ŞEKİLLER LİSTESİ

1. GİRİŞ

Bilgisayar sistemleri ile üretilen veriler tek başlarına değersizdir, çünkü bakıldığında bir anlam ifade etmezler. Bu veriler belli bir amaç doğrultusunda işlendiği zaman bir anlam ifade etmeye başlar [1]. Diğer taraftan geçmişte yaşanan kötü bir tecrübeden kaynaklanan kaybın engellenmesi mümkün değildir. Önemli olan geçmişe ait olaylara dair gizli bilgileri keşfetmek, ileriye yönelik durumsal öngörüler veren modeller ile önceden tedbir almamızı sağlayacak bir yönetim anlayışına geçmek ve olası kayıpları öngörebilmektir [2]. Bu yüzden büyük miktardaki verileri işleyebilen teknikleri kullanabilmek büyük önem kazanmaktadır. Bu ham veriyi bilgiye veya anlamlı hale dönüştürme işlemleri veri madenciliği ile yapılabilmektedir [1]. Veri madenciliği, bu gibi durumlarda kullanılan büyük miktardaki veri kümelerinde saklı durumda bulunan örüntü ve eğilimleri keşfetme işlemidir [3].

Bilgisayar ve iletişim teknolojilerindeki gelişmelere paralel olarak donanımın ucuzlaması verilerin uzun süre depolanmasına dolayısıyla büyük kapasiteli veritabanlarının oluşmasına neden olmuştur. Günümüzde veritabanları artık tera byte'larla ölçülmektedir. Bu ölçekte büyük veriler, stratejik öneme sahip bilgileri gizlemektedir. Veri madenciliği, büyük veri tabanlarındaki gizli bilgi ve yapıyı açığa çıkarmak için çok sayıda veri analizi aracını kullanan bir süreçtir [4].

Temel olarak dört alanda meydana gelen gelişmeler veri kaynaklarını bulma, onları sorgulama, sorgulanan verileri derleme ve verilerden bilgi üretmek için analiz yapma konusunda yeni fırsatlar ve araştırma konuları doğmasına neden olmaktadır. Bu alanlar aşağıda açıklandığı gibidir:

1. Veri işleme ve saklama alanındaki teknolojik gelişmeler gün geçtikçe artmakta ve daha fazla veriyi daha kısa sürede işlememize olanak sağlamaktadır.

2. Bilgisayar kullanımı üçüncü dünya ülkeleri de dahil, yıllar içerisinde artmakta ve gün geçtikçe daha fazla kişi, daha fazla dijital ortamda çalışmaya başlayarak, daha fazla dijital veri üretmektedir.
3. İletişim teknolojileri ve internet gibi altyapılar hızla tüm dünyayı sarmakta, yer ve zamandan bağımsız bir yaşam şekli gelişmektedir.
4. İnsanlar daha hızlı ve doğru karar almak için veriye dayalı bir araştırma, inceleme ve muhakeme kültürünü benimsemektedir.

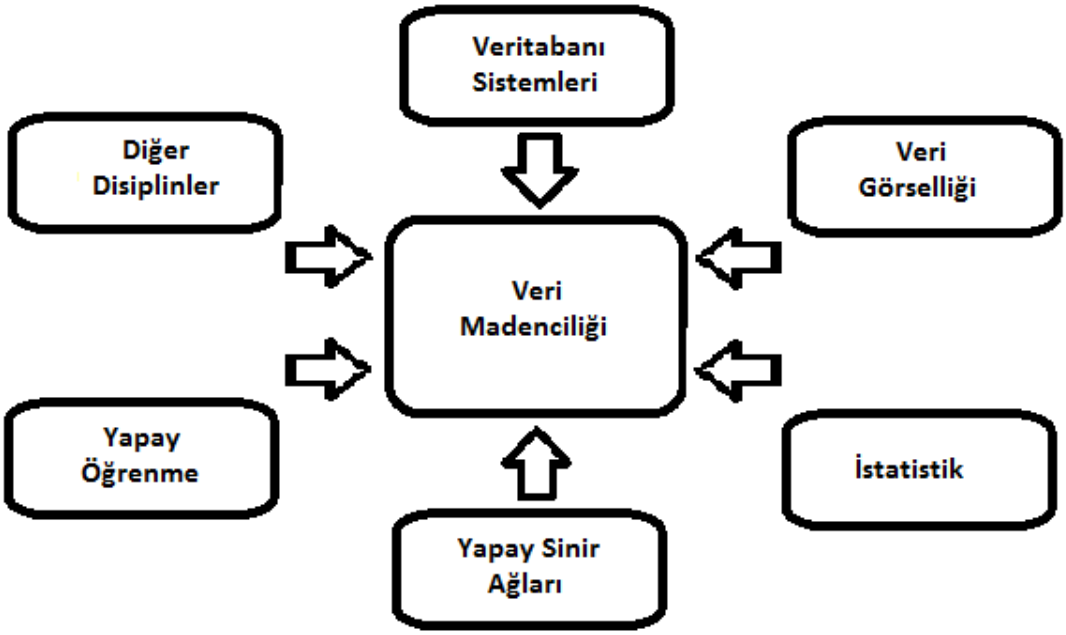
Veri madenciliği bu gelişmelere paralel doğmuş ve veri yığınları içindeki önceden keşfedilmemiş ilişkileri bulmaya odaklanmış sıcak bir araştırma alanıdır [5].

Bu çalışmada bir iplik üretim tesisinde üretilen ipliğin kalitesi, veri madenciliği sınıflandırma yöntemi kullanılarak belirlenecektir. Çalışmanın 2. bölümünde veri madenciliği genel kapsamda sunulacaktır. Bölüm 2.1'de veri madenciliğin tarihinden, Bölüm 2.2'de veri madenciliğin kullanıldığı alanlardan, Bölüm 2.3'de veri madenciliğini etkileyen etmenlerden, Bölüm 2.4'de veri madenciliğinde karşılaşılan problemlerden, Bölüm 2.5'de veri madenciliği sürecinden bahsedilecektir. Bölüm 3'de veri madenciliği modelleri detayları ile açıklanacaktır. Bölüm 4'de iplik kalitesini etkileyen faktörlerden bahsedilecektir. Ayrıca bu faktörlerin önem sıraları Taguchi yöntemi ile belirlenecektir. Bölüm 5'de çıkan sonuçlar veri madenciliği paket programlarında incelenip iplik kalitesini etkileyen faktörlerden bir sınıflandırma yapılarak kural oluşturulacaktır. Sonuç bölümünde ise çıkan veri sonuçları ile birlikte veri madenciliğinin genel değerlendirilmesi yapılacaktır.

2. VERİ MADENCİLİĞİ

Veri madenciliği, büyük miktarlardaki veri içinden, geleceğin tahmin edilmesinde yardımcı olacak anlamlı ve yararlı bağlantı ve kuralların bilgisayar programları aracılığıyla aranması ve analizidir. Ayrıca veri madenciliği, çok büyük miktardaki verilerin içindeki ilişkileri inceleyerek aralarındaki bağlantıyı bulmaya yardımcı olan ve veri tabanı sistemleri içerisinde gizli kalmış bilgilerin çekilmesini sağlayan veri analizi tekniğidir [1]. Veri madenciliği büyük miktarda veri inceleme amacı üzerine kurulmuş olduğu için veri tabanları ile yakından ilişkilidir. Günümüzde yaygın olarak kullanılmaya başlanan veri ambarları günlük kullanılan veri tabanlarının birleştirilmiş ve işlemeye daha uygun bir özetini saklamayı amaçlar. Veri madenciliği kendi başına bir çözüm değil çözüme ulaşmak için verilecek karar sürecini destekleyen, problemi çözmek için gerekli bilgileri sağlamaya yarayan bir araçtır. Veri madenciliği; analiste, iş yapma aşamasında oluşan veriler arasındaki şablonları ve ilişkileri bulması konusunda yardım etmektedir [6].

Veri madenciliğinin etkileşimde olduğu alanlar oldukça geniştir. Bu alanlar içerisinde Şekil 2.1.'de gösterildiği gibi, veri tabanı sistemleri, veri görselliği, yapay sinir ağları, istatistik, yapay öğrenme vb. disiplinler bulunmaktadır.



Şekil 2.1. Veri madenciliği ile ilişkili disiplinler.

Veri madenciliği araçları kullanılarak, işletmelerin daha etkin kararlar almasına yönelik karar destek sistemlerinde gerekli olan eğilimlerin ve davranış kalıplarının ortaya çıkarılması mümkün olmaktadır. Geçmişteki klasik karar destek sistemlerinin kullanıldığı araçlardan farklı olarak, veri madenciliğinde çok daha kapsamlı ve otomatik hale getirilmiş analizler yapmaya yönelik, birçok farklı özellik bulunmaktadır [2].

Günümüzün ekonomik koşulları ve yaşanan hızlı değişim ortamlarında, tecrübe ve önsezilere dayanarak alınan kararlarda yanlış karar alma riski çok yüksektir. Riski azaltmanın tek yolu bilgiye dayalı yönetimi öngören karar destek çözümleridir. Veri madenciliği teknikleri, gerçek anlamda bir karar destek sistemi oluşturmada olmazsa olmaz araçlardır. Bu noktada bilgi teknolojilerinden yararlanmak kaçınılmaz olmuştur [7].

Veri madenciliğinin özelliklerini sıralayacak olursak;

- Büyük ve karmaşık verilerle çalışır.
- Her türlü veriyi kullanarak çözümler üretebilir.
- Veritabanı sistemleri, veri görselliği, yapay sinir ağları, istatistik, yapay öğrenme vb. disiplinlerden faydalanır.

- Daha önceden bilinmeyen, doğrulanabilir, etkinleştirilebilir bilgiyi arar.
- Otomatik veya yarı otomatik olarak çalışan çözüm araçları kullanır.
- Birçok endüstride kullanılmaktadır.
- Sorunlara göre değişen çözüm araçları vardır.
- Hızla büyümekte olan bir alandır.

Bugüne kadar farklı kaynaklarda veri madenciliğinin pek çok tanımıyla karşılaşılmıştır. Bu kaynaklardan bazılarına göre veri madenciliğinin tanımı şöyledir:

Veri madenciliği, eldeki verilerden üstü kapalı, çok net olmayan, önceden bilinmeyen ancak potansiyel olarak kullanışlı bilginin çıkarılmasıdır. Bu da; kümeleme, veri özetleme, değişikliklerin analizi, sapmaların tespiti gibi belirli sayıda teknik yaklaşımları içerir [8].

Hand [9] veri madenciliğini istatistik, veritabanı teknolojisi, örüntü tanıma, makine öğrenme ile etkileşimli yeni bir disiplin ve geniş veritabanlarında önceden tahmin edilemeyen ilişkilerin ikincil analizi olarak tanımlamıştır. Kitler ve Wang [10] veri madenciliğini tahminci anahtar değişkenlerin binlerce potansiyel değişkenden izole edilmesini sağlama yeteneği olarak tanımlamışlardır. Jacobs [11] veri madenciliğini, ham verinin tek başına sunamadığı bilgiyi çıkaran, veri analizi süreci olarak tanımlamıştır.

Veri madenciliği, diğer bir adla veritabanında bilgi keşfi; çok büyük veri hacimleri arasında tutulan, anlamı daha önce keşfedilmemiş potansiyel olarak faydalı ve anlaşılır bilgilerin çıkarıldığı ve arka planda veritabanı yönetim sistemleri, istatistik, yapay zekâ, makine öğrenme, paralel ve dağıtık işlemlerin bulunduğu veri analiz teknikleridir [12].

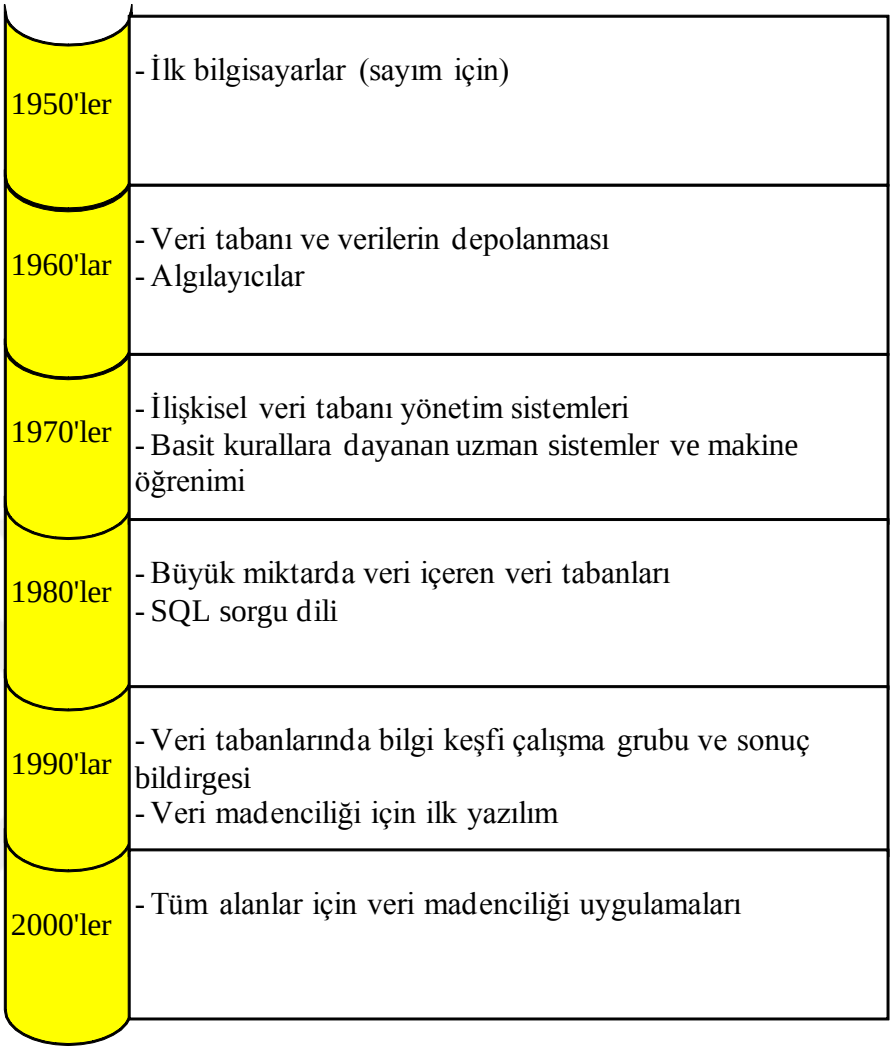
Veri madenciliği; veri ambarlarında tutulan çok çeşitli ve çok miktarda veriye dayanarak daha önce keşfedilmemiş bilgileri ortaya çıkarma, bunları karar verme ve eylem planını gerçekleştirmek için kullanma sürecidir [13].

Veri madenciliği, örüntü tanıma teknolojileri ile istatistiksel ve matematiksel teknikleri kullanarak havuzlarda saklanan yüklü miktardaki verilerin incelenmesi ile anlamlı yeni korelasyonlar, örüntüler ve eğilimler ortaya çıkarma sürecidir [14].

Veri madenciliği, büyük veri yığınları arasından gelecekle ilgili tahminde bulunabilmemizi sağlayabilecek bağlantıların, bilgisayar programı kullanarak aranması işidir [15].

2.1. Veri Madenciliğinin Tarihi

1950'li yıllarda ilk bilgisayarlar sayımlar için kullanılmaya başlanmıştır. 1960'larda ise veritabanı ve verilerin depolanması kavramı teknoloji dünyasında yerini almıştır. 1960'ların sonunda bilim adamları basit öğrenmeli bilgisayarlar geliştirebilmişlerdir. Minsky ve Papert, günümüzde sinir ağları olarak bilinen algılayıcıların sadece çok basit olan kuralları öğrenebileceğini göstermişlerdir. 1970'lerde İlişkisel Veri Tabanı Yönetim Sistemleri uygulamaları kullanılmaya başlanmıştır. Bununla beraber bilgisayar uzmanları basit kurallara dayanan uzman sistemler geliştirmişler ve basit anlamda makine öğrenimini sağlamışlardır. 1980'lerde veri tabanı yönetim sistemleri yaygınlaşmış ve bilimsel alanlarda uygulanmaya başlanmıştır. Bu yıllarda şirketler; müşterileri, rakipleri ve ürünleri ile ilgili verilerden oluşan veri tabanları oluşturmuşlardır. Bu veri tabanlarının içerisinde çok büyük miktarlarda veri bulunmakta ve bunlara SQL veri tabanı sorgulama dili ya da benzeri diller kullanarak ulaşılabilmektedir. 1990'larda artık içindeki veri miktarı katlanarak artan veri tabanlarından, faydalı bilgilerin nasıl bulunabileceği düşünülmeye başlanmış, bunun üzerine çalışmalara ve yayınlara başlanmıştır. KDD (IJCAI)-89 Veri Tabanlarında Bilgi Keşfi Çalışma Grubu toplantısı ile bilgi keşfi ve veri madenciliği ile ilgili temel tanım ve kavramların ortaya konması ile süreç daha da hızlanmış ve nihayet 1992 yılında veri madenciliği için ilk yazılım gerçekleştirilmiştir. 2000'li yıllarda veri madenciliği sürekli gelişmiş ve hemen hemen tüm alanlara uygulanmaya başlanmıştır. Alınan sonuçların faydaları görüldükçe, bu alana ilgi artmıştır [7]. Veri madenciliğinin tarihsel gelişim süreci, Şekil 2.2.'de gösterilmiştir.



Şekil 2.2. Veri madenciliğinin tarihsel süreci.

Tablo 2.1.'de ise veri madenciliğinin gelişim adımları yine tarihi süreç içerisinde detaylarla belirtilmiştir.

Tablo 2.1. Veri madenciliğinin gelişim adımları [16].

Gelişim Adımları	Cevaplanan Karar Ploblemi	Kullanılabilen Teknolojiler	Ürün Sağlayıcıları	Karakteristikler
Veri Toplama (1960'lar)	"Benim toplam karım geçen 5 yılda ne kadardı?"	Bilgisayarlar, Teypler, Diskler	IBM, CDC	Geriye dönük, statik veri dağıtımı
Veri Erişimi (1980'ler)	"İngiltere'de geçen mart ayında birim satışları ne kadardı?"	İlişkisel veritabanları, SQL, ODBC	Oracle, Sybase, İnformix, IBM, Microsoft	Kayıt düzeyinde geriye dönük, dinamik veri dağıtımı
Veri Ambarlama ve Karar Destek Sistemleri (1990'lar)	"Türkiye'de geçen nisan ayında birim satışları ne kadardı?"	OLAP, Çok boyutlu veritabanı sistemleri, Veri ambarları	Pilot, Comshare, Arbor, Cognos, Microstrategy	Çoklu düzeylerde, geriye dönük dinamik veri dağıtımı
Veri Madenciliği (Bugün)	"Gelecek ay Boston'daki birim satışlar muhtemelen ne olabilir, niçin?"	İleri düzeyde algoritmalar, Çok işlemcili bilgisayarlar, Büyük veritabanları	Pilot, Lockheed, IBM, SGI, SPSS, SAS, Microsoft vs.	Geleceğe dönük, proaktif enformasyon dağıtımı

2.2. Veri Madenciliğinin Kullanıldığı Alanlar

Büyük hacimde veri bulunan her yerde veri madenciliği kullanmak mümkündür. Günümüzde karar verme sürecine ihtiyaç duyulan birçok alanda veri madenciliği uygulamaları yaygın olarak kullanılmaktadır [2], [17], [18]. Veri madenciliği astronomi, biyoloji, finans, pazarlama, sigorta, tıp ve birçok başka dalda uygulanmaktadır. Bununla birlikte günümüzde veri madenciliği teknikleri özellikle işletmelerde çeşitli alanlarda başarı ile kullanılmaktadır. Bu uygulamaların başlıcaları ilgili alanlara göre şöyledir [19];

Pazarlama

- Müşterilerin satın alma örüntülerinin belirlenmesi
- Hedef pazarın belirlenmesi
- Müşteri memnuniyetinin belirlenmesi
- Çapraz satış
- Hilekârlık tespiti
- Kampanya yönetimi
- Dağıtım kanalı performans analizi

- Proses kontrol, kalite kontrol
- Envanter yönetimi
- Müşterilerin demografik özellikleri arasındaki bağlantıların bulunması
- Posta kampanyalarında cevap verme oranının artırılması
- Mevcut müşterilerin elde tutulması, yeni müşterilerin kazanılması
- Pazar sepeti analizi
- Müşteri ilişkileri yönetimi
- Müşteri değerlendirme
- Satış tahmini

Bankacılık

- Farklı finansal göstergeler arasında gizli korelasyonların bulunması
- Kredi kartı dolandırıcılıklarının tespiti
- Usulsüzlük tespiti
- Risk analizleri
- Risk yönetimi
- Kredi kartı harcamalarına göre müşteri gruplarının belirlenmesi
- Kredi taleplerinin değerlendirilmesi

Sigortacılık

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

Perakendecilik

- Satış noktası veri analizleri
- Alışveriş sepeti analizleri
- Tedarik ve mağaza yerleşim optimizasyonu

Borsa

- Hisse senedi fiyat tahmini
- Genel piyasa analizleri
- Alım-satım stratejilerinin optimizasyonu

Endüstri

- Kalite kontrol analizleri
- Lojistik
- Üretim süreçlerinin iyileştirilmesi

Bilim ve Mühendislik

- Ampirik veriler üzerinde modeller kurarak bilimsel ve teknik problemlerin çözümlenmesi

Telekomünikasyon

- İletişim desenlerinin belirlenmesi
- Kalite ve iyileştirme analizleri
- Hatların yoğunluk tahmini
- Kaynakların daha iyi kullanımı
- Servis kalitesinin artırılması

İnsanların Gen Haritalarının Oluşturulması

- Genetik veritabanı
- Gen sıralarının tespiti
- Genler ile hastalıkların ilişkilendirilmesi

Tıp

- Semptomlara göre hastalık tespiti
- Tedavi sürecinin belirlenmesi
- Test sonuçlarının tahmini
- Fonksiyonel magnetik rezonans verileri kullanılarak fiziksel hareketler ile etkin sinir sistemi bölgeleri ilişkilerinin belirlenmesi
- Klinik bilgi sistemleri
- Tıbbi karar sistemleri
- Tıbbi kayıt sistemleri
- Entegre akademik bilgi yönetim sistemleri
- Hastane yönetim sistemleri
- Bilgisayar destekli tıp eğitimi
- Tıbbi uzman sistemler
- Görüntü işleme ve analizi
- Sinyal analizi
- Hemşirelik bilgi sistemleri
- İdari karar sistemleri
- Tıbbi bilgi ağı
- Sağlık bilgi standartları
- Tıbbi bilişim eğitimi

2.3. Veri Madenciliğini Etkileyen Etmenler

Temel olarak veri madenciliğini beş ana harici eğilim etkiler [20]:

- **Veri:** Veri madenciliğinin bu kadar gelişmesindeki en önemli etkendir. Son yirmi yılda sayısal verinin hızla artması, veri madenciliğindeki gelişmeleri hızlandırmıştır. Bu kadar fazla veriye bilgisayar ağları üzerinden erişilmektedir. Diğer yandan bu verilerle uğraşan bilim adamları, mühendisler ve istatistikçilerin sayısı hala aynıdır. O yüzden, verileri analiz etme yöntemleri ve teknikleri geliştirilmektedir.
- **Donanım:** Veri madenciliği, sayısal ve istatistiksel olarak büyük veri kümeleri üzerinde yoğun işlemler yapmayı gerektirir. Gelişen bellek ve işlem hızı kapasitesi, birkaç yıl önce madencilik yapılamayan veriler üzerinde çalışmayı mümkün hale getirmiştir.
- **Bilgisayar Ağları:** Yeni nesil internet yaklaşık 100 Gbits/sn'lik, hatta belki daha da üzerinde hızları kullanmamızı sağlayacaktır. Bu da günümüzde kullanılan bilgisayar ağlarındaki hızın 100 katından daha fazla bir sürat ve taşıma kapasitesi demektir. Böyle bir bilgisayar ağı ortamı oluştuktan sonra, dağıtık verileri analiz etmek ve farklı algoritmaları kullanmak mümkün olacaktır. Buna bağlı olarak, veri madenciliğine uygun ağların tasarımı da yapılmaktadır.
- **Bilimsel Hesaplamalar:** Günümüz bilim adamları ve mühendisleri, simülasyonu bilimin üçüncü yolu olarak görmektedirler. Veri madenciliği ve bilgi keşfi, şu 3 metodu birbirine bağlamada önemli rol almaktadır: teori, deney ve simülasyon.
- **Ticari Eğilimler:** Günümüzde ticaret çok kârlı olmalı, daha hızlı ilerlemeli ve daha yüksek kalite hizmet verme yönünde olmalı, bütün bunları yaparken de minimum maliyeti ve en az insan gücünü göz önünde bulundurmalıdır. Bu tip hedef ve kısıtların yer aldığı iş dünyasında veri madenciliği, temel teknolojilerden biri haline gelmiştir. Çünkü veri madenciliği sayesinde müşterilerin ve müşteri faaliyetlerinin yarattığı fırsatlar daha kolay tespit edilebilmekte ve riskler daha açık görülebilmektedir.

2.4. Veri Madenciliğinde Karşılaşılan Problemler

Küçük veri kümelerinde hızlı ve doğru bir biçimde çalışan bir sistem, çok büyük veritabanlarına uygulandığında tamamen farklı davranabilir. Bir veri madenciliği sistemi tutarlı veri üzerinde mükemmel çalışırken, aynı veriye gürültü eklendiğinde kayda değer bir biçimde kötüleşebilir. Aşağıda veri madenciliğinin günümüz sistemlerinde karşı karşıya olduğu problemler incelenmiştir [21].

2.4.1. Veritabanı Boyutu

Veri tabanı boyutları inanılmaz bir hızla artmaktadır. Pek çok makine öğrenme algoritması birkaç yüz tutanaklık oldukça küçük örneklemleri ele alabilecek biçimde geliştirilmiştir. Aynı algoritmaların yüz binlerce kat büyük örneklemlerde kullanılabilmesi için azami dikkat gerekmektedir. Örneklemin büyük olması, örüntülerin gerçekten var olduğunu göstermesi açısından bir avantajdır ancak böyle bir örneklemden elde edilebilecek olası örüntü sayısı çok büyüktür. Bu yüzden veri madenciliği sistemlerinin karşı karşıya olduğu en önemli sorunlardan biri veritabanı boyutunun çok büyük olmasıdır. Dolayısıyla veri madenciliği yöntemleri ya sezgisel bir yaklaşımla arama uzayını taramalı ya da örneklemi yatay/dikey olarak indirgemelidir.

Yatay indirgeme çeşitli biçimlerde gerçekleştirilebilir. İlkinde, belirli bir niteliğin alan değerleri önceden sıradüzensel olarak sınıflandırılır (ya da kategorize edilir) ki buna genelleştirme işlemi de denilmektedir. Sonrasında, ise ilgili niteliğin değerleri önceden belirlenmiş genelleme sıra düzeninden aşağıdan yukarıya doğru seviye seviye güncellenir (yani, üst nitelik değeri ile değiştirilir) ve tekrarlı (mükerrer) çoklular çıkarılır [22]. İkincisinde, oldukça sağlam olan örnekleme kuramı kullanılarak çok büyük boyutlu veri öyle bir boyuta indirgenir ki hem kaynak veri belirli bir güven aralığında temsil edilebilir hem de indirgenen veri kümesi makine öğrenme algoritmalarınca işlenebilir [23]. Sonuncusunda ise, sürekli değerlerden oluşan bir alana sahip nitelik üzerine kesiklendirme tekniğinin uygulanmasıdır [24]. Sürekli değerlerin belirli aralık değerlerine dönüştürülmesi ile tekrarlılık arz eden çoklular ortadan kaldırılarak yatay indirgeme sağlanabilir. Aslında bu kesiklendirme tekniği, sürekli sayısal değerler için geçerli olmayan makine öğrenme algoritmaları için bir önkoşul ya da önişlemdir ki bu konu ayrı bir alt başlık olarak verilecektir. Dikey indirgeme, artık niteliklerin indirgenmesi işlemidir ki bu, artık veri alt başlığında tartışılacaktır.

2.4.2. Gürültülü Veri

Büyük veri tabanlarında pek çok niteliğin değeri yanlış olabilir. Bu hata, veri girişi sırasında yapılan insan hataları veya girilen değerin yanlış ölçülmesinden kaynaklanır. Veri girişi ya da veri toplanması sırasında oluşan sistem dışı hatalara gürültü adı verilir. Ancak günümüzde kullanılan ticari ilişkisel veritabanları veri girişi sırasında oluşan hataları otomatik biçimde gidermek konusunda düşük seviyede bir destek sağlamaktadır. Hatalı veri gerçek hayat veritabanlarında ciddi problem oluşturabilir. Bu durum, bir veri madenciliği yönteminin kullanılan veri kümesinde bulunan gürültülü verilere karşı daha az duyarlı olmasını gerektirir. Gürültülü verinin yol açtığı problemler, tümevarımsal karar ağaçlarında uygulanan metotlar bağlamında kapsamlı bir biçimde araştırılmıştır [25]. Eğer veri kümesi gürültülü ise sistem bozuk veriyi tanımalı ve ihmal etmelidir. Quinlan [25] gürültünün sınıflandırma üzerindeki etkisini araştırmak için bir dizi deney yapmıştır. Deneysel sonuçlar, etiketli öğrenmede etiket üzerindeki gürültü, öğrenme algoritmasının performansını doğrudan etkileyerek düşmesine sebep olmuştur. Buna karşın eğitim kümesindeki nesnelerin özellikleri/nitelikleri üzerindeki en çok %10'luk gürültü miktarı ayıklanabilmektedir. Chan ve Wong [26] gürültünün etkisini analiz etmek için istatiksel yöntemler kullanmışlardır.

2.4.3. Boş Değerler

Bir veri tabanında boş değer, birincil anahtarda yer almayan herhangi bir niteliğin değeri olabilir. Boş değer tanımı gereği kendisi de dahil olmak üzere hiç bir değere eşit olmayan değerdir. Bir girdinin eğer bir nitelik değeri boş ise o nitelik bilinmeyen ve uygulanamaz bir değere sahiptir. Bu durum ilişkisel veritabanlarında sıkça karşımıza çıkmaktadır. Bir ilişkide yer alan tüm çoklular aynı sayıda niteliğe (niteliğin değeri boş olsa bile) sahip olmalıdır. Örneğin kişisel bilgisayarların özelliklerini tutan bir ilişkide bazı model bilgisayarlar için ses kartı modeli niteliğinin değeri boş olabilir.

Lee [27] boş değeri, (1) bilinmeyen, (2) uygulanamaz, ve (3) bilinmeyen veya uygulanamaz olacak biçimde üçe ayıran bir yaklaşımı ilişkisel veritabanlarını genişletmek için öne sürmüştür. Mevcut boş değer taşıyan veri için herhangi bir çözüm sunmayan bu yaklaşımın dışında bu konuda sadece bilinmeyen değer üzerinde çalışmalar yapılmıştır [28], [29], [30], [31]. Boş değerli nitelikler veri kümesinde

bulunuyorsa, ya bunlar tamamıyla ihmal edilmeli ya da bunlara niteliğe olası en yakın değer atanmalıdır [32].

2.4.4. Eksik Veri

Veri kümesindeki eksiklik yatay ve dikey boyutta mevcuttur. Bu eksiklikler şu şekilde olabilir [33];

- Yatay boyutta: Yatay boyuttaki eksiklik, veri kümesinde olması gereken nitelik veya niteliklerin olmamasıdır.
- Dikey boyutta: Dikey boyuttaki eksiklik veri kümesindeki kayıtların eksik olmasıdır.

2.4.5. Artık Veri

Verilen veri kümesi, eldeki probleme uygun olmayan veya artık nitelikler içerebilir. Bu durum pek çok işlem sırasında karşımıza çıkabilir. Örneğin, eldeki problem ile ilgili veriyi elde etmek için iki ilişkiyi ortak nitelikler üzerinden birleştirirsek sonuç girdide kullanıcının farkında olmadığı artık nitelikler bulunur. Artık nitelikleri elemek için geliştirilmiş algoritmalar nitelik seçimi olarak adlandırılır [34].

Nitelik seçimi, tümevarıma dayalı öğrenmede budama öncesi yapılan bir işlemdir. Başka bir deyişle, nitelik seçimi, verilen bir ilişkinin içsel tanımını, dışsal tanımın taşıdığı (veya içerdiği) bilgiyi bozmadan onu eldeki niteliklerden daha az sayıdaki niteliklerle (yeterli ve gerekli) ifadeleyebilmektir. Nitelik seçimi yalnızca arama uzayını küçültmekle kalmayıp, sınıflama işleminin kalitesini de arttırmaktadır [35], [36], [37], [38], [39].

2.4.6. Dinamik Veri

Kurumsal çevrim-içi veri tabanları dinamiktir, yani içeriği sürekli olarak değişir. Bu durum, bilgi keşfi metotları için önemli sakıncalar doğurmaktadır. İlk olarak sadece okuma yapan ve uzun süre çalışan bilgi keşfi metodu, bir veri tabanı uygulaması olarak mevcut veri tabanı ile birlikte çalıştırıldığında mevcut uygulamanın da performansı ciddi ölçüde düşer. Diğer bir sakınca ise, veritabanında bulunan verilerin kalıcı olduğu varsayılıp, çevrimdışı veri üzerinde bilgi keşif metodu çalıştırıldığında, değişen verinin elde edilen örüntülere yansıması gerekmektedir. Bu işlem, bilgi keşfi metodunun

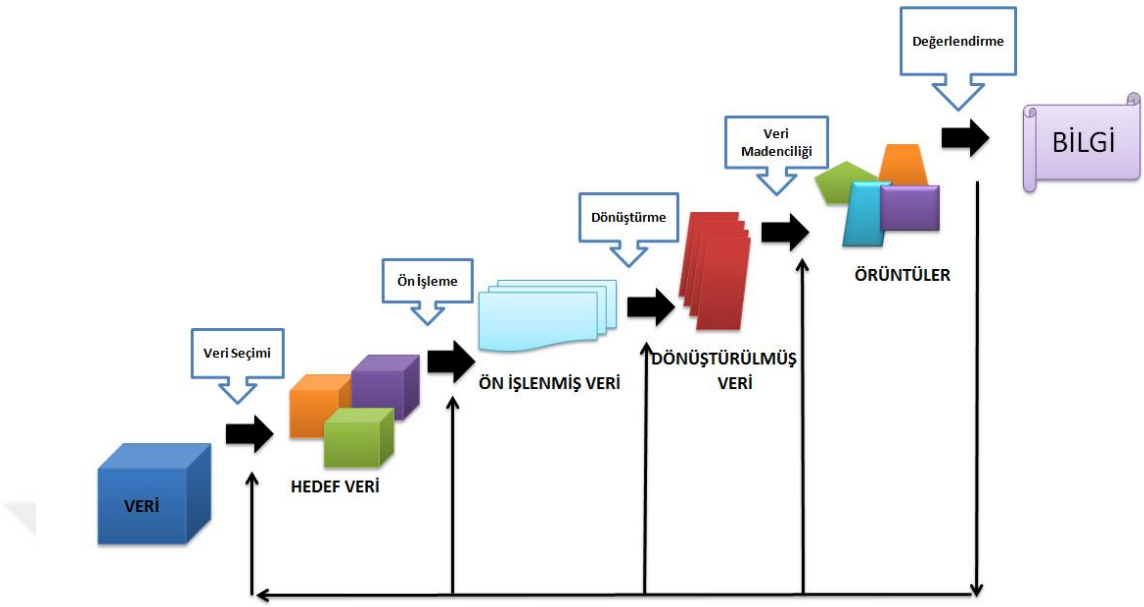
ürettiği örüntüleri zaman içinde değişen veriye göre sadece ilgili örüntüleri yığmalı olarak güncelleme yeteneğine sahip olmasını gerektirir [40]. Aktif veri tabanları, tetikleme mekanizmalarına sahiptir ve bu özellik bilgi keşif metotları ile birlikte kullanılabilir [41].

2.4.7. Farklı Tipteki Verileri Ele Alma

Gerçek hayattaki uygulamalar makine öğrenmede olduğu gibi yalnızca sembolik veya kategorik veri türleri değil; tamsayı, kesirli sayılar, çoklu ortam verisi, coğrafi bilgi içeren veri gibi farklı tipteki veriler üzerinde aynı anda işlem yapılmasını gerektirir. Kullanılan verinin saklandığı ortam, düz bir kütük veya ilişkisel veritabanında yer alan tablolar olacağı gibi, nesneye yönelik veritabanları, çoklu ortam veritabanları, coğrafî veritabanları vb. olabilir. Saklandığı ortama göre veri, basit tipte olabileceği gibi karmaşık veri tipleri (çoklu ortam verisi, zaman içeren veri, yardımlı metin, coğrafi vb.) de olabilir. Bununla birlikte veri tipi çeşitliliğinin fazla olması bir veri madenciliği algoritmasının tüm veri tiplerini ele alabilmesini olanaksızlaştırmaktadır. Bu yüzden veri tipine özgü veri madenciliği algoritmaları geliştirilmektedir [42].

2.5. Veri Madenciliği Süreci

Veri madenciliği, aynı zamanda bir süreçtir. Veri yığınları arasında, soyut kazılar yaparak veriyi ortaya çıkarmanın yanı sıra, bilgi keşfi sürecinde örüntüleri ayrıştırarak süzmek ve bir sonraki adıma hazır hale getirmek de bu sürecin bir parçasıdır. Bu süreç Şekil 2.3.'de gösterilmiştir. Üzerinde inceleme yapılan işin ve verilerin özelliklerinin bilinmemesi durumunda ne kadar etkin olursa olsun hiç bir veri madenciliği algoritmasının fayda sağlaması mümkün değildir. Bu sebeple, veri madenciliği sürecine girilmeden önce, başarının ilk şartı, iş ve veri özelliklerinin detaylı analiz edilmesidir [7].



Şekil 2.3. Bilgi keşfi sürecinde veri madenciliği [43].

Fayyad [43]'a göre veri tabanlarından bilgi keşfi sürecinde yer alan adımlar şöyledir;

- **Veri Seçimi**: Veri kümesinin birleştirilerek, sorguya uygun örneklem kümesinin elde edildiği adımdır.

- **Veri Temizleme ve Önişleme**: Seçilen örneklemde yer alan hatalı örneklerin çıkarıldığı ve eksik nitelik değerlerinin değiştirildiği aşamadır. Bu aşama keşfedilen bilginin kalitesini arttırır.

- **Veri İndirgeme**: Seçilen örneklemden ilgisiz niteliklerin atıldığı ve tekrarlı tutanakların ayıklandığı adımdır. Bu aşama seçilen veri madenciliği sorgusunun çalışma zamanını iyileştirir.

- **Veri Madenciliği**: Verilen bir veri madenciliği sorgusunun (sınıflandırma, kümeleme, birliktelik analizi vb.) işletilmesidir.

- **Değerlendirme:** Keşfedilen bilginin geçerlilik, yenilik, yararlılık ve basitlik gibi ölçütlere göre değerlendirilmesi aşamasıdır.

Akpınar [19]'a göre veri tabanlarında bilgi keşfi sürecinde izlenmesi gereken temel aşamalar şunlardır;

- Problemin tanımlanması,
- Verilerin hazırlanması,
- Modelin kurulması ve değerlendirilmesi,
- Modelin kullanılması,
- Modelin izlenmesi.

2.5.1. Problemin Tanımlanması

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

2.5.2. Verilerin Hazırlanması

Modelin kurulması aşamasında ortaya çıkacak sorunlar, bu aşamaya sık sık geri dönülmesine ve verilerin yeniden düzenlenmesine neden olacaktır. Bu durum verilerin hazırlanması ve modelin kurulması aşamaları için, bir analizcinin veri keşfi sürecinin toplamı içerisinde enerji ve zamanının % 50 - % 85'ini harcamasına neden olmaktadır. Verilerin hazırlanması aşaması kendi içerisinde toplama, değer biçme, birleştirme ve temizleme, seçme ve dönüştürme adımlarından meydana gelmektedir.

- **Toplama**

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

- **Değer Biçme**

Veri madenciliğinde kullanılacak verilerin farklı kaynaklardan toplanması, doğal olarak veri uyumsuzluklarına neden olacaktır. Bu uyumsuzlukların başlıcaları farklı zamanlara ait olmaları, kodlama farklılıkları (örneğin bir veri tabanında cinsiyet özelliğinin e/k, diğer bir veri tabanında 0/1 olarak kodlanması), farklı ölçü birimleridir. Ayrıca verilerin nasıl, nerede ve hangi koşullar altında toplandığı da önem taşımaktadır. Bu nedenlerle, iyi sonuç alınacak modeller ancak iyi verilerin üzerine kurulabileceği için, toplanan verilerin ne ölçüde uyumlu oldukları bu adımda incelenerek değerlendirilmelidir.

- **Birleştirme ve Temizleme**

Bu adımda farklı kaynaklardan toplanan verilerde bulunan ve bir önceki adımda belirlenen sorunlar mümkün olduğu ölçüde giderilerek veriler tek bir veri tabanında toplanır. Ancak basit yöntemlerle ve baştan savma olarak yapılacak sorun giderme işlemlerinin, ileriki aşamalarda daha büyük sorunların kaynağı olacağı unutulmamalıdır.

- **Seçim**

Bu adımda kurulacak modele bağlı olarak veri seçimi yapılır. Örneğin tahmin edici bir model için, bu adım bağımlı ve bağımsız değişkenlerin ve modelin eğitiminde kullanılacak veri kümesinin seçilmesi anlamını taşımaktadır. Sıra numarası, kimlik numarası gibi anlamlı olmayan ve diğer değişkenlerin modeldeki ağırlığının azalmasına da neden olabilecek değişkenlerin modele girmemesi gerekmektedir. Bazı veri madenciliği algoritmaları konu ile ilgisi olmayan bu tip değişkenleri otomatik olarak elese de, pratikte bu işlemin kullanılan yazılıma bırakılmaması daha akılcı olacaktır. Verilerin görselleştirilmesine olanak sağlayan grafik araçlar ve bunların sunduğu ilişkiler, bağımsız değişkenlerin seçilmesinde önemli yararlar sağlayabilir. Genellikle yanlış veri girişinden veya bir kereye özgü bir olayın gerçekleşmesinden kaynaklanan verilerin, önemli bir uyarıcı enformasyon içerip içermediği kontrol edildikten sonra veri kümesinden çıkarılması tercih edilir. Modelde kullanılan veri tabanının çok büyük olması durumunda tesadüfiliği bozmayacak şekilde örnekleme yapılması uygun olabilir. Günümüzde hesaplama olanakları ne kadar gelişmiş olursa olsun, çok büyük veri

tabanları üzerinde çok sayıda modelin denenmesi zaman kısıtı nedeni ile mümkün olamamaktadır. Bu nedenle tüm veri tabanını kullanarak birkaç model denemek yerine, tesadüfi olarak örneklenmiş bir veri tabanı parçası üzerinde birçok modelin denenmesi ve bunlar arasından en güvenilir ve güçlü modelin seçilmesi daha uygun olacaktır.

- **Dönüştürme**

Kredi riskinin tahmini için geliştirilen bir modelde, borç/gelir gibi önceden hesaplanmış bir oran yerine, ayrı ayrı borç ve gelir verilerinin kullanılması tercih edilebilir. Ayrıca modelde kullanılan algoritma, verilerin gösteriminde önemli rol oynayacaktır. Örneğin bir uygulamada bir yapay sinir ağı algoritmasının kullanılması durumunda kategorik değişken değerlerinin evet/hayır olması; bir karar ağacı algoritmasının kullanılması durumunda ise örneğin gelir değişken değerlerinin yüksek/orta/düşük olarak gruplanmış olması modelin etkinliğini artıracaktır.

2.5.3. Modelin Kurulması ve Değerlendirilmesi

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

Model kuruluş süreci danışmanlı ve danışmansız öğrenmenin kullanıldığı modellere göre farklılık göstermektedir. Örnekten öğrenme olarak da isimlendirilen danışmanlı öğrenmede, bir denetçi tarafından ilgili sınıflar önceden belirlenen bir kritere göre ayrılarak, her sınıf için çeşitli örnekler verilir. Sistemin amacı verilen örneklerden hareket ederek her bir sınıfa ilişkin özelliklerin bulunması ve bu özelliklerin kural cümleleri ile ifade edilmesidir. Öğrenme süreci tamamlandığında, tanımlanan kural cümleleri verilen yeni örneklere uygulanır ve yeni örneklerin hangi sınıfa ait olduğu kurulan model tarafından belirlenir. Danışmansız öğrenmede, kümeleme analizinde olduğu gibi ilgili örneklerin gözlenmesi ve bu örneklerin özellikleri arasındaki benzerliklerden hareket ederek sınıfların tanımlanması amaçlanmaktadır.

Danışmanlı öğrenmede seçilen algoritmaya uygun olarak ilgili veriler hazırlandıktan sonra, ilk aşamada verinin bir kısmı modelin öğrenimi, diğer kısmı ise modelin geçerliliğinin test edilmesi için ayrılır. Modelin öğrenimi eğitim kümesi kullanılarak gerçekleştirildikten sonra, test kümesi ile modelin doğruluk derecesi belirlenir. Bir modelin doğruluğunun test edilmesinde kullanılan en basit yöntem basit geçerlilik testidir. Bu yöntemde tipik olarak verilerin % 5 ile % 33 arasındaki bir kısmı test verileri olarak ayrılır ve kalan kısım üzerinde modelin öğrenimi gerçekleştirildikten sonra, bu veriler üzerinde test işlemi yapılır. Bir sınıflama modelinde yanlış olarak sınıflanan olay sayısının, tüm olay sayısına bölünmesi ile hata oranı, doğru olarak sınıflanan olay sayısının tüm olay sayısına bölünmesi ile ise doğruluk oranı hesaplanır (Doğruluk Oranı = 1 - Hata Oranı).

Sınırlı miktarda veriye sahip olunması durumunda, kullanılabilecek diğer bir yöntem çapraz geçerlilik testidir. Bu yöntemde veri kümesi tesadüfi olarak iki eşit parçaya ayrılır (*a* ve *b*). İlk aşamada *a* parçası üzerinde model eğitimi ve *b* parçası üzerinde test işlemi; ikinci aşamada ise *b* parçası üzerinde model eğitimi ve *a* parçası üzerinde test işlemi yapılarak elde edilen hata oranlarının ortalaması kullanılır. Birkaç bin veya daha az satırdan meydana gelen küçük veri tabanlarında, verilerin *n* gruba ayrıldığı *n*-katlı çapraz geçerlilik testi tercih edilebilir. Verilerin örneğin 10 gruba ayrıldığı bu yöntemde, ilk aşamada birinci grup test, diğer gruplar öğrenim için kullanılır. Bu süreç her defasında bir grubun test, diğer grupların öğrenim amaçlı kullanılması ile sürdürülür. Sonuçta elde edilen 10 hata oranının ortalaması, kurulan modelin tahmini hata oranı olacaktır.

Önyükleme, küçük veri kümeleri için modelin hata düzeyinin tahmininde kullanılan bir başka tekniktir. Çapraz geçerlilikte olduğu gibi model bütün veri kümesi üzerine kurulur. Daha sonra en az 200, bazen 1000'in üzerinde olmak üzere çok fazla sayıda öğrenim kümesi tekrarlı örneklemelerle veri kümesinden oluşturularak hata oranı hesaplanır.

Model kuruluşu çalışmalarının sonucuna bağlı olarak, aynı teknikle farklı parametrelerin kullanıldığı veya başka algoritma ve araçların denendiği değişik modeller kurulabilir. Model kuruluş çalışmalarına başlamadan önce, imkansız olmasa da hangi tekniğin en uygun olduğuna karar verebilmek güçtür. Bu nedenle farklı

modeller kurarak, doğruluk derecelerine göre en uygun modeli bulmak üzere denemeler yapılmasında yarar bulunmaktadır. Özellikle sınıflandırma problemleri için kurulan modellerin doğruluk derecelerinin değerlendirilmesinde basit ancak faydalı bir araç olan risk matrisi kullanılmaktadır.

Önemli diğer bir değerlendirme kriteri modelin anlaşılabilirliğidir. Bazı uygulamalarda doğruluk oranlarındaki küçük artışlar çok önemli olsa da, birçok işletme uygulamasında ilgili kararın niçin verildiğinin yorumlanabilmesi çok daha büyük önem taşıyabilir. Çok ender olarak yorumlanamayacak kadar karmaşıklaşsalar da, genel olarak karar ağacı ve kural temelli sistemler model tahmininin altında yatan nedenleri çok iyi ortaya koyabilmektedir. Kaldıraç oranı ve grafiği, bir modelin sağladığı faydanın değerlendirilmesinde kullanılan önemli bir yardımcıdır. Örneğin kredi kartını muhtemelen iade edecek müşterilerin belirlenmesi amacını taşıyan bir uygulamada, kullanılan modelin belirlediği 100 kişinin 35'i gerçekten bir süre sonra kredi kartını iade ediyorsa ve tesadüfi olarak seçilen 100 müşterinin aynı zaman diliminde sadece 5'i kredi kartını iade ediyorsa kaldıraç oranı 7 olarak bulunacaktır. Kurulan modelin değerinin belirlenmesinde kullanılan diğer bir ölçü, model tarafından önerilen uygulamadan elde edilecek kazancın bu uygulamanın gerçekleştirilmesi için katlanılacak maliyete bölünmesi ile elde edilecek olan yatırımın geri dönüş oranıdır.

Kurulan modelin doğruluk derecesi ne denli yüksek olursa olsun, gerçek hayatı tam anlamı ile modellediğini garanti edebilmek mümkün değildir. Yapılan testler sonucunda geçerli bir modelin doğru olmamasındaki başlıca nedenler, model kuruluşunda kabul edilen varsayımlar ve modelde kullanılan verilerin doğru olmamasıdır. Örneğin modelin kurulması sırasında varsayılan enflasyon oranının zaman içerisinde değişmesi, bireyin satın alma davranışını belirgin olarak etkileyecektir.

2.5.4. Modelin Kullanılması

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

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

2.5.5. Modelin İzlenmesi

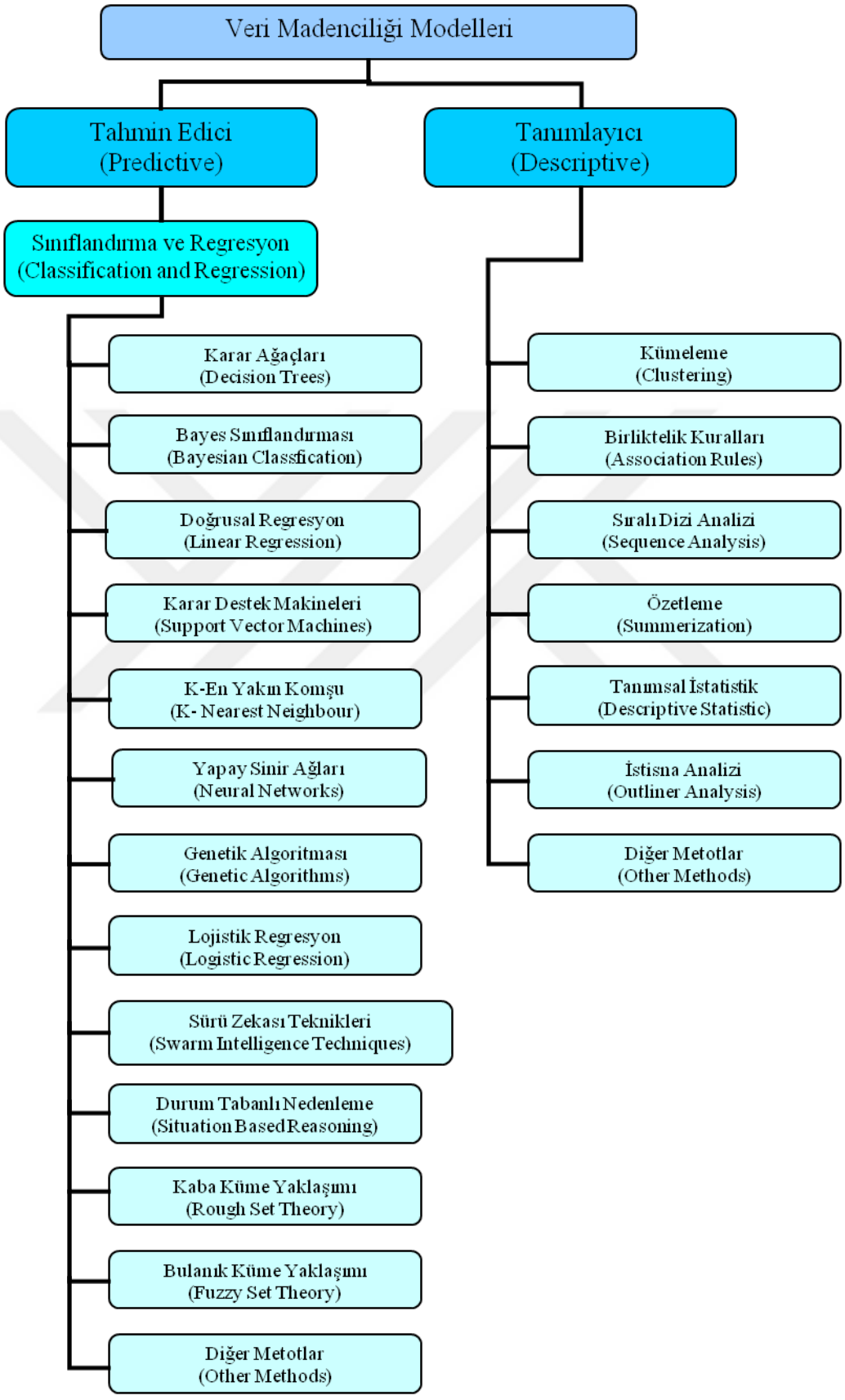
Zaman içerisinde bütün sistemlerin özelliklerinde ve dolayısıyla ürettikleri verilerde ortaya çıkan değişiklikler, kurulan modellerin sürekli olarak izlenmesini ve gerekiyorsa yeniden düzenlenmesini gerektirecektir. Tahmin edilen ve gözlenen değişkenler arasındaki farklılığı gösteren grafikler model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

3. VERİ MADENCİLİĞİ MODELLERİ

Veri madenciliği ile ilgili kullanılan pek çok yöntemin yanına hemen her geçen gün yeni yöntem ve algoritmalar eklenmektedir. Bunlardan bir kısmı onlarca yıldır kullanılan klasik teknikler diyebileceğimiz, ağırlıklı olarak istatistiksel yöntemlerdir. Diğer yöntemler de genellikle istatistiği temel alan ama daha çok makine öğrenme ve yapay zeka destekli yeni nesil yöntemlerdir. Veri madenciliği modelleri, gördükleri işlevlere göre temel olarak 3 grupta toplanır. Bunlar:

1. Sınıflandırma ve Regresyon,
2. Kümeleme,
3. Birliktelik Kurallarıdır.

Sınıflama ve regresyon modelleri tahmin edici, kümeleme ve birliktelik kuralları modelleri tanımlayıcı özelliktedir [44]. Şekil 3.1.'de veri madenciliği modelleri gösterilmektedir.



Şekil 3.1. Veri madenciliği modelleri [45].

Veri madenciliğinde kullanılan modeller, tahmin edici ve tanımlayıcı olmak üzere iki ana başlık altında incelenmektedir.

3.1. Tahmin Edici Modeller

Sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır. Tahmin edici modeller sınıflandırma ve regresyon yöntemleridir [19].

3.1.1. Sınıflandırma ve Regresyon Modelleri

Mevcut verilerden hareket ederek geleceğin tahmin edilmesinde faydalanılan ve veri madenciliği teknikleri içerisinde en yaygın kullanıma sahip olan sınıflandırma ve regresyon modelleri arasındaki temel fark, tahmin edilen bağımlı değişkenin kategorik veya süreklilik gösteren bir değere sahip olmasıdır. Ancak çok terimli lojistik regresyon gibi kategorik değerlerin de tahmin edilmesine olanak sağlayan tekniklerle, her iki model giderek birbirine yaklaşmakta ve bunun bir sonucu olarak aynı tekniklerden yararlanılması mümkün olmaktadır [19].

Sınıflandırma ve regresyon modellerinde kullanılan başlıca algoritmalar şöyle sıralanabilir [33];

- Karar ağaçları,
- Yapay sinir ağları
- Genetik algoritmalar
- K-en yakın komşu
- Bayes sınıflandırıcılar
- Doğrusal regresyon
- Doğrusal olmayan regresyon
- Lojistik regresyon
- Sürü zekası teknikleri
- Durum tabanlı nedenleme
- Kaba küme yaklaşımı
- Bulanık küme yaklaşımı

3.1.1.1. Karar Ağaçları

Karar ağaçları, sınıfları bilinen örnek veriden tümevarım yöntemiyle öğrenilen ağaç şekilli bir karar yapısı çeşididir. Bir karar ağacı, basit karar verme adımları uygulanarak, büyük miktarlardaki kayıtları, çok küçük kayıt gruplarına bölerek kullanılan bir yapıdır. Her başarılı bölme işlemiyle, sonuç gruplarının üyeleri bir diğeriyle çok daha benzer hale gelmektedir. Büyük veri tabanlarının kullanıldığı pek çok sınıflandırma probleminde ve karmaşık ya da hata içeren bilgilerde karar ağaçları yararlı bir çözüm olmaktadır [46].

Tahmin edici ve tanımlayıcı özelliklere sahip olan karar ağaçları, veri madenciliğinde [19];

- Kuruluşlarının ucuz olması,
- Yorumlanmalarının kolay olması,
- Veri tabanı sistemleri ile kolayca entegre edilebilmeleri,
- Güvenilirliklerinin daha iyi olması,

nedenleri ile sınıflandırma modelleri içerisinde en yaygın kullanıma sahiptir.

Karar ağacı temelli analizlerin yaygın olarak kullanıldığı sahalar [19];

- Belirli bir sınıfın muhtemel üyesi olacak elemanların belirlenmesi,
- Çeşitli vakaların yüksek, orta, düşük risk grupları gibi çeşitli kategorilere ayrılması,
- Gelecekteki olayların tahmin edilebilmesi için kurallar oluşturulması,
- Parametrik modellerin kurulmasında kullanılmak üzere çok miktardaki değişken ve veri kümesinden faydalı olacakların seçilmesi,
- Sadece belirli alt gruplara özgü olan ilişkilerin tanımlanması, kategorilerin birleştirilmesi ve sürekli değişkenlerin kesikliye dönüştürülmesidir.

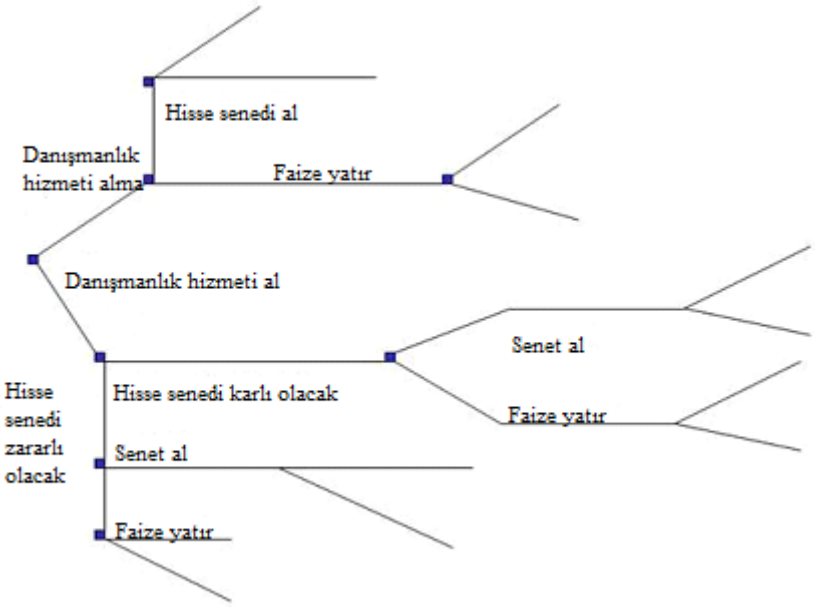
Karar ağacı temelli tipik uygulamalar ise [19];

- Hangi demografik grupların mektupla yapılan pazarlama uygulamalarında yüksek cevaplama oranına sahip olduğunun belirlenmesi,
- Bireylerin kredi geçmişlerini kullanarak kredi kararlarının verilmesi,
- Geçmişte işletmeye en faydalı olan bireylerin özelliklerini kullanarak işe alma süreçlerinin belirlenmesi,

- Tıbbi gözlem verilerinden yararlanarak en etkin kararların verilmesi,
- Hangi değişkenlerin satışları etkilediğinin belirlenmesi,
- Üretim verilerini inceleyerek ürün hatalarına yol açan değişkenlerin belirlenmesidir.

Gerçek dünyanın sosyal ve ekonomik olaylarını daha güvenilir bir şekilde gösterebilmek için standart istatistik tekniklerin dışında yeni analiz tekniklerinin geliştirilmesi ile ilgilenen Morgan ve Sonquist tarafından University of Michigan'da 1970'li yılların başlarında kullanıma alınan Automatic Interaction Detector (AID), karar ağacı temelli ilk algoritma ve yazılımdır. Automatic Interaction Detector (AID) tekniği en kuvvetli ve en iyi tahmini gerçekleştirebilmek için bağımlı ve bağımsız değişkenler arasındaki mümkün bütün ilişkilerin incelenmesine dayanmaktadır. En kuvvetli ilişkiye sahip bağımsız değişken bulunduğunda, veri kümesi bu bağımsız değişken değerlerine göre ikiye ayrılmakta ve süreç mümkün bölünmeler tamamlanıncaya kadar devam etmektedir [19].

Karar ağaçları ile ilgili bir örnek verecek olursak; Siz bir şirket yöneticisisiniz ve elinizde şirkete dair yüklü bir miktar para var. Bu parayı sizden en yüksek getiriyi sağlayacak şekilde faiz veya senet alarak değerlendirmeniz isteniyor. İsterseniz bir danışmandan yardım alabilir isterseniz kendiniz karar verebilirsiniz. Olasılıkları çıkartacak olursak; ilk olarak danışmana başvuralım. Danışman size senet al veya faize yatır seçeneklerini sunacaktır. Bu seçeneklerde kendi aralarında başarılı veya başarısız olarak ikiye ayrılacaktır. İlk etapta finans ile ilgili yeterli bilgiye sahip olmadığınız düşünülerek danışmana başvurmak mantıklı gelecektir. Ama bu seçenek sonucunda danışmana da bir miktar ödeme yapmamız gerekecektir. Diğer yandan danışmanlık hizmeti almazsınız ve kendiniz karar verirsiniz. Hisse senedi kârlı olacaktır (kârlı olmasına karşın faize yatır veya senet al) veya hisse senedi zararlı olacaktır. Bahsettiğimiz bu olayların karar ağacı Şekil 3.2.'deki gibidir [47].



Şekil 3.2. Örnek karar ağacı gösterimi [47].

3.1.1.2. Yapay Sinir Ağları

1980'lerden sonra yaygınlaşan yapay sinir ağlarında amaç fonksiyonu, birbirine bağlı basit işlemci ünitelerinden oluşan bir ağ üzerine dağıtılmıştır [48]. Yapay sinir ağlarında kullanılan öğrenme algoritmaları, veriden üniteler arasındaki bağlantı ağırlıklarını hesaplar. Yapay sinir ağları istatistiksel yöntemler gibi veri hakkında parametrik bir model varsaymaz yani uygulama alanı daha geniştir ve bellek tabanlı yöntemler kadar yüksek işlem ve bellek gerektirmez [49].

Yapay Sinir Ağları [50];

- İlk kez 1943'te ortaya çıkmış ancak bilgisayarlarda kullanımı 1980'lerde başlamıştır.
- Yapay sinir ağları beynin yapısından esinlenmiş bir bilgi işleme sistemidir. Nöronlara benzeştirilmiş işlem öğeleri arasındaki ilişkilerle yapılandırılmıştır.
- İnsan beyni gibi yapay sinir ağı da birbirine bağlı birçok işlem biriminden oluşmuştur. Birçok düğüm (işlem birimi) ve arkla (iç bağlantılar) yönetilen bir grafik olarak yapılandırılır.

- Bu işlem birimleri birbirlerinden bağımsız işlev görürler ve yalnızca yerel veriyi (düğüme gelen girdi ve düğümden çıkan çıktı) kullanırlar. Bu özellik, sinir ağlarının dağıtık ya da paralel ortamlarda kullanımını kolaylaştırır.
- Sinir ağları, kaynak (girdi), çıktı ve iç (gizli) düğümlerle yönetilen bir grafik olarak görülebilir. Girdi düğümü girdi katmanında, çıktı düğümü ise çıktı katmanında bulunur. Gizli düğümler, bir ya da daha çok gizli katmanda bulunur. Veri madenciliğinde, çıktı düğümü tahmini belirler.
- Tek bir girdi düğümünün olduğu (ağacın kökü) karar ağaçlarından farklı olarak sinir ağlarında, her öznitelik değeri için bir girdi düğümü vardır.
- Sinir ağları karmaşık sorunları çözebilir, ayrıca temel uygulamalardan öğrenebilir. Yani soruna kötü bir çözüm bulunduysa, ağ bu soruna bir dahaki sefer daha iyi bir çözüm bulacak biçimde değiştirilir.

Sinir ağları üç bölümden oluşur [50];

1. Sinir ağının veri yapısını tanımlayan sinir ağı grafiği.
2. Öğrenmenin nasıl gerçekleşeceğini belirten öğrenme algoritması.
3. Bilginin ağdan nasıl elde edileceğini belirleyen teknikler.

Yapay sinir ağları [50];

- Örüntü tanımada,
- Ses tanıma ve çözümlemede,
- Tıbbi uygulamalarda (tanı, ilaç),
- Hata algılamada,
- Sorun tanılamada,
- Robot denetiminde,
- Herhangi bir işlevi hesaplamada kullanılabilir.

3.1.1.3. Genetik Algoritmalar

Genetik algoritmalar; doğal seleksiyon prensibinden yola çıkılarak geliştirilmiş arama algoritmalarıdır. Algoritma, belirli bir uzunluğa sahip dizilerden oluşmuş bir popülasyona sahiptir. Popülasyon içindeki her bir kromozom, çözüm uzayında bir noktayı temsil eder. Bu diziler aynı zamanda üreme yolu ile varlığını sürdürmeye aday olan birer bireydir. Algoritmanın temel işleyişi; çözüme uygun olmayan bireyleri elemek, çözüme daha uygun bireyleri seçmek ve seçilen bireylerden yeni bireyler üretmek doğrultusundadır. Algoritmanın isleyişi aşamalı olarak düşünüldüğünde temel prensibinin eleme olduğu anlaşılmaktadır. İkinci prensip ise, elemeyi aşan bireylerden yararlanılarak olası yeni çözümler elde etmektir. Bu da bireyler arasındaki bilgi alışverişi ile sağlanır. Bireyler arası rasgele bilgi alışverişi, arama işleminin, çözüm uzayının daha uygun noktalarında devam etmesini sağlar [51].

John Holland [52] evrim sürecinin bir bilgisayar yardımıyla kullanılarak, bilgisayara anlayamadığı çözüm yöntemlerinin öğretilebileceğini düşünmüştür. Genetik algoritma John Holland tarafından bu düşüncenin bir sonucu olarak bulunmuştur. Genetik algoritma stokastik bir arama yöntemidir. “En uygun olan hayatta kalır” ilkesine dayanmaktadır. Sezgisel bir yöntem olan genetik algoritma, problem için optimum sonucu bulamayabilir, ancak bilinen metotlarla çözülemeyen veya çözüm zamanı çok büyük olan problemlerde optimuma çok yakın çözümler vermektedir. Holland’ın araştırması iki yönlüdür [51];

- Doğal sistemlerde görülen adaptasyon (ortama uyum yeteneği) kavramını açıklamak,
- Doğal sistemlerin temel işleyişini, yapay yazılım sistemleri aracılığı ile modellemek.

Doğal sistemler oldukça sağlam ve kararlı yapıdadırlar. Uyum yeteneğinin nasıl gelişip işlediğini öğrenmenin en iyi yolu, biyolojik sistemler üzerindeki çalışmalardır. Genetik algoritmaların karmaşık arama uzaylarına uygulandığında kararlı çözümlere ulaştıkları, kuramsal ve deneysel çalışmalar ile kanıtlanmıştır. Etkili ve verimli bir yöntem olduğu kanıtlandıktan sonra çeşitli bilim dallarının yanı sıra iş dünyası ve mühendislikte geniş bir uygulama alanı bulmasının en önemli nedeni, kuşkusuz hesaplamalarda sağladığı basitlik ve arama işlemlerindeki iyileştirme gücüdür [51].

3.1.1.4. K-en Yakın Komşu Algoritması

Kayıtlar, bir veri uzayındaki noktalar olarak düşünülürse, birbirine yakın olan kayıtlar, birbirinin civarında (yakın komşu) olur. *K*-en yakın komşuluğunda temel düşünce "komşunun yaptığı gibi yap" tır. Eğer belirli bir kişinin davranışı tahmin edilmek isteniyorsa, veri uzayında o kişiye yakın, örneğin on kişinin davranışlarına bakılır. Bu on kişinin davranışlarının ortalaması hesaplanır ve bu ortalama, belirlenen kişi için tahmin olur. *K*-en yakın komşuluğunda, *K* harfi araştırılan komşuların sayısıdır. 5-en yakın komşuluğunda, 5 kişiye ve 1-en yakın komşuluğunda 1 kişiye bakılır [45].

Bu modellerin olası dezavantajı özellikle çok fazla açıklayıcı değişken varsa çok fazla hesaplama yükü getirmesidir. Bu durumda komşuluklar ilgisiz noktalarda belirebilir, bu sebepten dolayı ortalamalarını almak anlamlı sonuçlar vermeyebilir [51].

3.1.1.5. Bayes Sınıflandırıcılar

Bayes sınıflandırıcılar istatistiksel sınıflandırıcılardır. Özel bir sınıfa ait olarak verilen bir olasılık gibi, sınıf üyeliklerine ait olasılıkları önceden söyleyebilirler [45]. Bunun yanı sıra büyük veri tabanları üzerinde işlem yapan bayes sınıflandırıcılar hız ve doğruluk yönünden oldukça yüksek performansa sahiptirler. Bayes sınıflandırıcılar, verinin olasılık dağılımı verildiğinde en düşük hata oranıyla etkin bir şekilde çalışabilirler [53].

Bayes, verilen bir sınıf üzerindeki bir özelliğe ait değerin etkisinin diğer özelliğe ait değerlerden bağımsız olduğunu farz eder. Bu kabul, sınıf koşullu bağımsızlık olarak adlandırılır. Bu ise gerekli hesaplamaları basitleştirir, anlaşılırlığı kolaylaştırır ve doğal olarak saf yani "naive" kelimesi ile ifade edilir [33].

Bayes sınıflandırıcılar verinin tek bir kez taranmasını gerektiren basit sınıflandırıcılardır. Bu nedenle büyük veri yığınlarında yüksek doğruluk ve hız elde edebilirler. Performansları karar ağaçları ve sinir ağları ile rekabet edebilecek ölçüdedir [54].

3.1.1.6. Doğrusal Regresyon

Regresyon analizi istatistiksel bir teknik olup, bir veya daha çok değişkenin başka değişkenler cinsinden tahmin edilmesini sağlayan ilişkilerin bulunmasına yardımcı olur. Regresyon analizinin değişik türleri vardır. Doğrusal regresyonda veri, düz bir doğru kullanılarak modellenir. Doğrusal regresyon, regresyonun en basit halidir. İki değişkenli doğrusal regresyon, rassal bir değişken olan *Y*'yi (yanıt değişkeni) diğer bir rassal değişken olan *X*'in (tahminleyici değişken) doğrusal bir fonksiyonu olarak modeller [45]. Fonksiyon şu şekildedir;

$$Y=\alpha+\beta x \tag{3.1.}$$

Burada yanıt değişkeni *Y*'nin varyansı sabit olarak kabul edilir. α ve β, sırasıyla *Y*-kesişim ve doğrunun eğimini sağlayan regresyon sabitleridir. Bu sabitler, gerçek veri ve tahmin edilen doğru arasındaki hatayı en küçükleyen, en küçük kareler yöntemi ile çözülebilir. Doğrusal regresyon, bağımlı ve bağımsız tüm değişkenler nümerik olduğunda seçilecek en doğal yöntemdir [55].

3.1.1.7. Doğrusal Olmayan Regresyon

Veri doğrusal bir bağımlılık göstermiyorsa, yanıt değişkeni ve tahminleyici değişkenler ancak çok terimli bir fonksiyonla modellenebiliyorsa bu durumda doğrusal olmayan regresyon kullanılır. Çok terimli regresyon, temel doğrusal modele çok terimli ifadeler ekleyerek modellenir. Doğrusal olmayan regresyonda değişkenlere dönüşümler uygulanarak, doğrusal olmayan model doğrusal bir model haline dönüştürülür ve bu doğrusal model daha sonra en küçük kareler yöntemi ile çözülür [33].

3.1.1.8. Lojistik Regresyon

Lojistik regresyon analizinin kullanım amacı, istatistikte kullanılan diğer model yapılandırma teknikleriyle aynı olup en az değişkeni kullanarak en iyi uyuma sahip olacak şekilde bağımlı (sonuç) değişken ile bağımsız değişkenler kümesi (açıklayıcı değişkenler) arasındaki ilişkiyi tanımlayabilen ve genel olarak kabul edilebilir bir model kurmaktır. Lojistik regresyonu, doğrusal regresyondan ayıran en belirgin özellik ise, lojistik regresyonda bağımlı değişkenin kategorik değişken olmasıdır. Lojistik

regresyon ve doğrusal regresyon arasındaki bu fark, hem parametrik model seçimine, hem de varsayımlara yansımaktadır. Lojistik regresyonda da, doğrusal regresyon analizinde olduğu gibi bazı değişken değerlerine dayanarak kestirim yapılmaya çalışılır, ancak iki yöntem arasında üç önemli fark vardır [56]:

1. Doğrusal regresyon analizinde tahmin edilecek olan bağımlı değişken sürekli iken, lojistik regresyonda bağımlı değişken kesikli bir değer olmalıdır.

2. Doğrusal regresyon analizinde bağımlı değişkenin değeri, lojistik regresyonda ise bağımlı değişkenin alabileceği değerlerden birinin gerçekleşme olasılığı tahmin edilir.

3. Doğrusal regresyon analizinde bağımsız değişkenlerin çoklu normal dağılım göstermesi koşulu aranırken, lojistik regresyonun uygulanabilmesi için bağımsız değişkenlerin dağılımına ilişkin hiç bir ön koşul yoktur.

3.1.1.9. Sürü Zekası

Sürü zekası, özerk yapıdaki basit bireyler grubunun kolektif bir zeka geliştirmesi olarak tanımlanır [57]. Sürü zekası tanımı ilk olarak 1989 yılında Gerardo Beni ve Jing Wang tarafından hücresel robotik sistemler kavramı içinde kullanılmıştır. Sürü zekası, merkezi olmayan, kendi kendini yöneten sistemlerde toplu davranış çalışmasına dayanmaktadır.

Sürü zekası sistemleri genel olarak, birbirleriyle ve çevreleriyle etkileşen basit ajanlar topluluğundan oluşur. Birey olarak ajanların nasıl davranacağını dikte eden merkezileşmiş bir kontrol yapısı yoktur. Bu ajanlar arasındaki yerel etkileşimler çoğunlukla küresel davranışın ortaya çıkışını gösterir. Bu tip sistemlere örnekler doğada mevcuttur. Bunlara örnek olarak karınca kolonileri, kuş sürüleri, hayvan sürüleri, bakteri küfleri, arı kolonileri, balık sürüsü vb. verilebilir.

Sürü zekası, etkileşim ve kendi kendine organizasyon olmak üzere iki mekanizma üzerine kuruludur. Etkileşim, sürünün iletişim kurmasına ve sürüdeki elamanların birbirlerinin hareketlerini düzenlemelerine yardımcı olur. Kendi kendine organizasyon

ise, sürünün herhangi bir plan olmadan sonuç üretebilmesini, esnek, sağlam ve merkezi bir yönetim birimi olmadan yapılanmasını sağlar.

Sürü zekası, karınca koloni optimizasyonu [58], parçacık sürü optimizasyonu [59], arı algoritması [60] ve stokastik difüzyon arama [61] olmak üzere farklı algoritmalar içermektedir. Bu algoritmalar optimizasyonda başarı ile kullanılmaktadır. Günümüzde sürü zekası teknikleri veri madenciliği alanında da kullanılmaya başlanmıştır ve yapılan uygulamalar bu tekniklerin sınıflandırmada iyi sonuçlar elde edebildiğini göstermektedir.

3.1.1.10. Durum Tabanlı Nedenleme

İlk olarak Schank [62] tarafından ortaya atılan durum tabanlı nedenleme sınıflandırıcıları, örnek tabanlı sınıflandırıcılardır [45]. Durum tabanlı nedenleme, deneyimle öğrenir ve verilen problemle daha önce karşılaşılan problemlerin benzerlik ve farklılıklarından yararlanır. Eğitim örneklerini öklit uzayında noktalar olarak saklayan en yakın komşu sınıflandırıcılarından farklı olarak, durum tabanlı nedenleme tarafından saklanan örnekler veya durumlar karmaşık sembolik tanımlamalardır. Yapay sinir ağlarından farklı olarak ise durum tabanlı nedenlemeye dayalı sınıflandırıcılar genelleme yapmazlar. Durum tabanlı nedenleme sınıflandırıcıları şunları içerir [63];

- Yeni istekleri karşılamak için eski çözümleri uyarlamak,
- Yeni durumları açıklamak veya yeni çözümleri ispatlamak için eski durumları kullanmak,
- Yeni durumları yorumlamak için önceki durumları nedenlemek.

Sınıflandırılacak yeni bir durum ele alındığında, bir durum tabanlı nedenleyici ilk olarak belirli bir eğitim durumunun mevcut olup olmadığını kontrol eder. Eğer bir eğitim durumu mevcutsa, ilişik bir çözüm o duruma geri gönderilir. Eğer hiç bir belirli durum bulunmazsa, bu durumda durum tabanlı nedenleyici yeni durumun bileşenleriyle aynı bileşenlere sahip eğitim durumlarını arar. Kavramsal olarak bu eğitim durumları yeni durumun komşuları olarak düşünülebilir. Durum tabanlı nedenleyici yeni duruma bir çözüm önermek için komşu eğitim durumlarının çözümlerini birleştirmeye çalışır. Çözümlerde uyuşmazlık ortaya çıkarsa bu durumda yeni çözümler için yeniden arama

yapmak gerekebilir. Durum tabanlı nedenleyici, uygun bileşik bir çözüm bulmak için geçmiş bilgileri ve problem çözme stratejilerini kullanabilir [33].

Durum tabanlı nedenlemeye yeni yaklaşımlar iyi bir benzerlik ölçüsünün bulunması, eğitim durumlarını indekslemek için etkin tekniklerin geliştirilmesi ve çözümlerin birleştirilmesini içerir. Mevcut bazı durum seçim metotları şunlardır [64]; özetlenmiş en yakın komşu, durum genişletme veya budama ile örnek tabanlı öğrenme (IB3 gibi), nitelik ağırlıklandırma ile örnek tabanlı öğrenme (IB4 gibi).

3.1.1.11. Kaba Küme Yaklaşımı

İlk olarak Pawlak [65] tarafından ortaya atılan kaba küme yaklaşımı sınıflandırmada kesin olmayan, gürültülü verideki yapısal ilişkileri bulmakta kullanılır ve kesikli değerli değişkenlere uygulanır. Kaba küme yaklaşımı, klasik kümedeki, kümenin yalnızca elemanları ile tanımlandığı ve kümenin elemanları hakkında ilave hiçbir bilginin bulunmadığı yaklaşımın aksine, bir kümenin tanımlanması için başlangıçta uzay hakkında bazı bilgilere gereksinim olduğu varsayımına dayanır. Kaba küme yaklaşımının temelini ayırt edilememe ilişkisi oluşturur. Bilginin temelini oluşturan ve aynı nesnelerin kümesi olan kümeye elemanter küme denir. Elemanter kümelerin herhangi bir birleşimi ise kesin küme olarak adlandırılır, aksi halde bir küme kaba olarak ifade edilir. Her kaba kümenin kesinlikle kümenin kendisinin veya tümleyen kümesinin elemanları olarak sınıflandırılamayan elemanları vardır, bunlara sınır hattı elemanları denir [66].

Kaba küme yaklaşımı, ele alınan bir eğitim verisinde denklik sınıflarının kurulumuna dayanır. Bir denklik sınıfını oluşturan tüm veri örnekleri ayırt edilemez niteliktedir. Verilen bir sınıf için kaba küme tanımı iki küme ile tahmin edilir, bunlar alt yaklaşım ve üst yaklaşımdır. Alt yaklaşım kesin olarak sınıfa ait olan tüm nesnelerden oluşur. Üst yaklaşım ise sınıfa ait olması olası bütün nesneleri içerir. Alt ve üst yaklaşımlar arasındaki fark sınır bölgesini oluşturur [67]. Karar kuralları her bir sınır için oluşturulur. Genel olarak kaba küme yaklaşımında, kuralları göstermekte karar tabloları kullanılır [45].

Kaba küme yaklaşımı kullanılarak çözülebilen ana problemler; özellik değerleri cinsinden nesnelerin kümesinin tanımı, özellikler arasındaki tam veya kısmi bağımlılıkların belirlenmesi, özelliklerin indirgenmesi, özelliklerin öneminin ortaya konulması ve karar kurallarının oluşturulmasıdır [65].

3.1.1.12. Bulanık Küme Yaklaşımı

Zadeh [68] tarafından ortaya atılan ve bulanık mantığa dayanan bulanık küme yaklaşımı, belirsizlik ile ilgilenir. Bulanık mantık, karmaşık, iyi tanımlanmamış ya da matematiksel olarak kolay analiz edilemeyen sistemlerin davranışlarını tanımlamak için hemen hemen doğru ve etkin yöntemler sunar. Diğer bir deyişle, bulanık mantık, belirsiz ve kesin olmayan bilginin kullanılması için bir platform oluşturur. Kategoriler arasında kesin bir ayrımdan ziyade, 0-1 arasında bir doğruluk değeri kullanırlar.

Bulanık küme yaklaşımları, veri madenciliği uygulamalarında özellikle sınıflandırma amacıyla kullanılmaktadır. Ancak sürekli değişkenler için keskin sınırların olması sınıflandırmada önemli bir dezavantajdır. Bununla birlikte yüksek düzeyde bir soyutlama sağlamaları sınıflandırmada bulanık küme yaklaşımına avantaj sağlamaktadır [45].

3.2. Tanımlayıcı Modeller

Tanımlayıcı modellerde ise karar vermeye rehberlik etmede kullanılabilecek mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır. Tanımlayıcı modeller kümeleme ve birliktelik kurallarıdır [19].

3.2.1. Kümeleme Yöntemi

Kümeleme, örnekler arasındaki yakınlık veya benzerliğin uygun niteliğine göre, örnek kümesinin belirli parça veya kümelere gruplanması süreci olarak tanımlanır [69]. Kümeleme ile birbirinden farklı olan ve birbirine benzeyen gruplar keşfedilmeye çalışılır. Kümelemede amaç, verileri alt kümelere ayırmaktır [70]. Kümelemede sınıflandırmadan farklı olarak, başlangıçta kümelerin ne olacağı veya hangi değişkenlerle verinin kümeleneceği bilinmez. Ele alınan konu ile ilgili uzman bir kişinin kümeleri yorumlaması gerekir. Kümeler belirlendikten sonra veri kümesi uygun bir

şekilde parçalanmalıdır. Böylece bu kümeler daha sonra yeni verileri sınıflandırmakta kullanılabilir [33].

Kümelemede, benzerlikleri bulmak için çoğunlukla Manhattan ve Öklit uzaklık fonksiyonları kullanılır. Uzaklık fonksiyonu sonucunun yüksek bir değer olması az benzerlik olduğunu, düşük bir değer olması ise çok benzerlik olduğunu ifade eder. Veri kümeleri için uygulanacak uzaklık fonksiyonlarının verimi farklı olabilir. Bundan dolayı Manhattan ve Öklit haricindeki uzaklık fonksiyonları bazı veri kümeleri için daha uygun olabilir [33].

Kümeleme modellerinin özellikleri kısaca şöyle sıralanabilir;

- Danışmansız öğrenmedir,
- Kümelerin yapıları doğrudan veriden bulunmalıdır,
- Önceden tanımlanan sınıf ve sınıf etiketli öğrenme örnekleriyle çalışmamaktadır,
- Bir veri madenciliği fonksiyonudur,
- Veri dağılımını anlamaya yardımcı olur,
- Her bir kümenin özelliklerini izler.

Tipik bir örnek kümeleme faaliyeti beş adımdan oluşmaktadır. Bunlar [71];

1. Uygun örnek gösterimi
2. Nesneler arasında uygun uzaklık veya benzerlik ölçütü ile yakınlığın tanımı
3. Kümeleme
4. Veri soyutlama
5. Çıktının değerlendirilmesi

Kümeleme modellerinde yer alan algoritmalar Bölümleme Kümeleme Algoritması ve Hiyerarşik Kümeleme Algoritması olmak üzere ikiye ayrılır [72]. Bölümleme kümelemede, kümeler arasında ilişki bulunmaz, hiyerarşik kümelemede ise, her kümede veri örneklerini içerecek bir bağlantı kurulur. Bu algoritmaların sınıflandırılması aşağıda verilmiştir;

- Bölümleme Kümeleme Algoritması

    - *K*-ortalamalar
    - *K*-medoid (CLARANS)
    - Beklenen maksimizasyon algoritması
    - Danışmansız bayes
    - Mod bulma
    - Yoğunluk tabanlı yaklaşım

- Hiyerarşik Kümeleme Algoritması
    - Toplayıcı hiyerarşik kümeleme algoritması
    - Bölücü hiyerarşik kümeleme algoritması

Veri madenciliği amacı ile kullanılan kümeleme algoritmalarının aşağıdaki özellikleri taşıması gerekir [54];

- Ölçeklenebilir olmalı,
- Farklı veri tipleri ile başa çıkabilmeli,
- Göreli şekildeki kümeleri keşfedebilmeli,
- Girdi parametrelerini belirlemekte alan bilgisi için minimum ihtiyaç göstermeli,
- Aykırı değerlerle ve gürültü ile başa çıkabilmeli,
- Girdi kayıtlarının sırasına duyarsız olmalı,
- Yüksek boyutta olmalı,
- Kullanışlı ve yorumlanabilir olmalı.

3.2.2. Birliktelik Kuralı

Bir alışveriş sırasında veya birbirini izleyen alışverişlerde müşterinin hangi mal veya hizmetleri satın almaya eğilimli olduğunun belirlenmesi, müşteriye daha fazla ürünün satılmasını sağlama yollarından biridir. Satın alma eğilimlerinin tanımlanmasını sağlayan birliktelik kuralları ve ardışık zamanlı örüntüler, pazarlama amaçlı olarak Pazar Sepeti Analizi (Market Basket Analysis) adı altında veri madenciliğinde yaygın olarak kullanılmaktadır. Bununla birlikte bu teknikler, tıp, finans ve farklı olayların birbirleri ile ilişkili olduğunun belirlenmesi sonucunda değerli bilgi kazanımının söz konusu olduğu ortamlarda da önem taşımaktadır. Birliktelik kuralları aşağıda sunulan

örneklerde görüldüğü gibi eş zamanlı olarak gerçekleşen ilişkilerin tanımlanmasında kullanılır [19].

- Müşteriler kola satın aldığında, %75 ihtimalle patates cipsi de alırlar,
- Düşük yağlı peynir ve yağsız yoğurt alan müşteriler, %85 ihtimalle diyet süt de satın alırlar.

Ardışık zamanlı örüntüler ise aşağıda sunulan örneklerde görüldüğü gibi birbirleri ile ilişkisi olan ancak birbirini izleyen dönemlerde gerçekleşen ilişkilerin tanımlanmasında kullanılır [19].

- X ameliyatı yapıldığında, 15 gün içinde %45 ihtimalle Y enfeksiyonu oluşacaktır,
- İMKB endeksi düşerken A hisse senedinin değeri %15'den daha fazla artacak olursa, 3 iş günü içerisinde B hisse senedinin değeri %60 ihtimalle artacaktır,
- Çekiç satın alan bir müşteri, ilk üç ay içerisinde %15, bu dönemi izleyen üç ay içerisinde %10 ihtimalle çivi satın alacaktır.

4. İPLİK KALİTESİ SEÇİMİNE ETKİ EDEN FAKTÖRLERİN TAGUCHİ DENEYSEL TASARIM YÖNTEMİ İLE BELİRLENMESİ

Bu bölümde; uygulama çalışmasının gerçekleştirileceği firmadan, makineden ve iplik kalitesi seçimine etki eden faktörlerin, deneysel tasarım metodu ile önem sıralarının belirlenmesi ve sayılarının azaltılmaya çalışılmasından bahsedilecektir.

Uygulama çalışmasının gerçekleştirileceği firma bir Karamancı Holding şirketi olan Orta Anadolu Tekstildir. 1953 yılında Kayseri'de entegre bir iplik ve dokuma fabrikası olarak kurulmuştur. 1986 yılındaki yeniden yapılanma sonunda faaliyet alanını %100 pamuklu jean giyim kumaşları üretimi olarak belirleyen şirket, bu alanda dünyanın ilerici, saygın ve tartışmasız lideri olmayı kendisine vizyon edinmiştir. Bugün itibari ile 50 milyon metrelik üretimi ile Türkiye'de birinci, Avrupa'da ikinci konumda bulunan Orta Anadolu, ürettiği "Ordenim" ve "Orcotton" markası ile dünya denim (jean kumaşı) sektöründe ürün ve hizmet kalitesinde haklı bir üne sahiptir.

Modern tekstil teknolojisini, geçmiş deneyimleri ve uzman kadrosuyla birleştiren, Orta Anadolu üretiminin %70'ini başta Avrupa Topluluğu ülkeleri olmak üzere çeşitli dünya ülkelerine ihraç etmektedir. Diğer taraftan Orta Anadolu dünya denim kumaş pazarının %1 ini elinde tutmaktadır.

Uygulama çalışmasının gerçekleştirileceği, Orta Anadolu fabrikasının iplik üretim bölümünde işleyiş prosesi Şekil 4.1.‘de olduğu gibidir.

Şekil 4.1. İşleyiş prosesi şematik çizimi.

İşletmeye gelen pamuk, pamuk ambarında stoklanır ve burada giriş testlerine tabi tutulur. Harman-hallaç makineleri olan Unifloc, Unimix, Erm ve Tarak makinelerinden sırasıyla geçirilerek temizlenip şerit haline getirilir. Kovalara yerleştirilen şeritler Cer-1 ve Cer-2 makinelerinde düzgünleştirilip düzenli hale getirilir. Bu şeritler kovalar halinde Fitil makinesine gelir, burada incelip büküm verilen şeritler fitil şeklinde makaralara sarılır. Makaralar daha sonra Ring İplik makinesine bağlanıp, eğrilerek inceltilip iplik elde edilir. Bu süreç içerisinde iplik kalitesine birçok faktör etki eder. Ring İplik makinesi, ipliği son haline getiren makine olması sebebiyle, iplik kalitesinde en çok etkiye sahiptir.

Bu çalışmada Ring İplik makinesinde iplik kalitesine etki eden 6 faktör belirlenmiştir. Bu 6 faktörün her birine ait 3 adet seviye vardır. Bu faktör ve seviyeler Tablo 4.1.'de görüldüğü gibidir.

Tablo 4.1. İplik kalitesine etki eden faktör ve seviyeleri.

Seviye/Faktör	**Manşon**	**Klips**	**Kopça**	**Kafes**	**Devir**	**Büküm**
1.Seviye	Yumuşak	3.5 mm	h2dr02	Dar	7500	581
2.Seviye	Orta	4 mm	h2dr04	Mevcut	8000	618
3.Seviye	Sert	4.5 mm	h2dr08	Geniş	8500	655

İplik kalitesine en çok etki eden Ring İplik makinesi ve iplik kalitesini etkileyen faktörlerden aşağıdaki kısımda detaylı olarak bahsedilecektir.

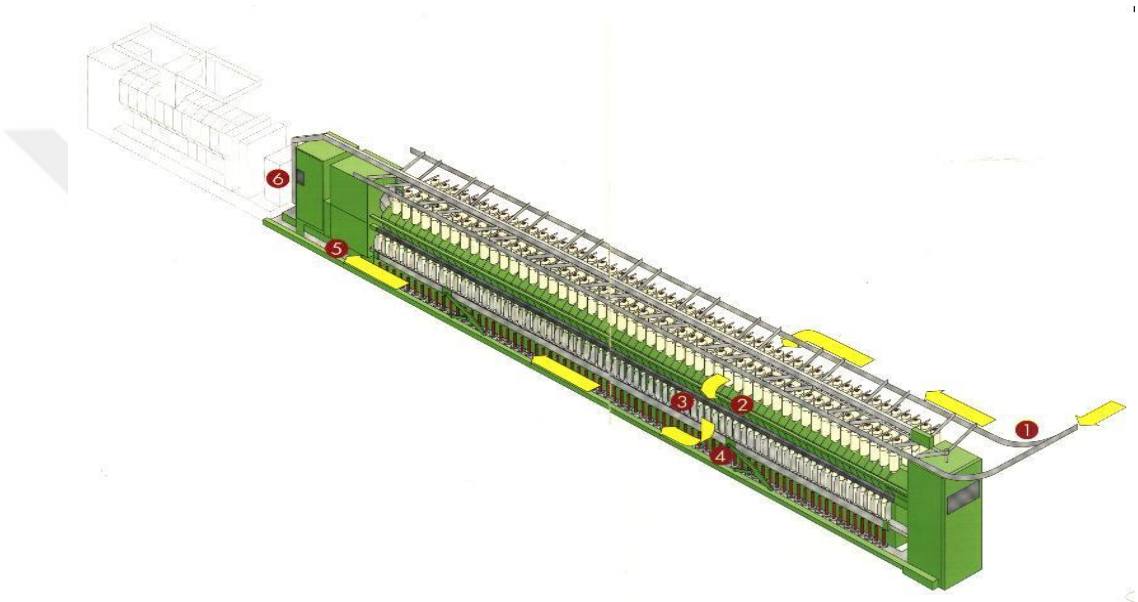
4.1. Ring İplik Makinesi

Bu makinede fitil, çekilerek tasarlanan derecede inceltilir, büküm verilerek istenilen mukavemete getirilir ve masura üzerine sarılır.

Ring iplik makinesi aşağıdaki 3 görevle mükelleftir:

1. Çalışılmakta olan fitili istenen iplik numarası nispetinde çekerek inceltmek,
2. İpliğe büküm vererek yeterli mukavemeti sağlamak,
3. İpliği masura üzerine kops teşkil edilecek şekilde sarmak.

Ring İplik makinesinin şematik gösterimi Şekil 4.2.'deki gibidir.



1. Fitil Cağlığı
2. Çekim sistemi
3. Büküm bölümü
4. Doffer
5. Masura transportu
6. Masura toplama bölgesi

Şekil 4.2. G5/11 Ring İplik makinesi şematik gösterimi.

Ring İplik makinesi çalışma elemanları aşağıdaki gibidir:

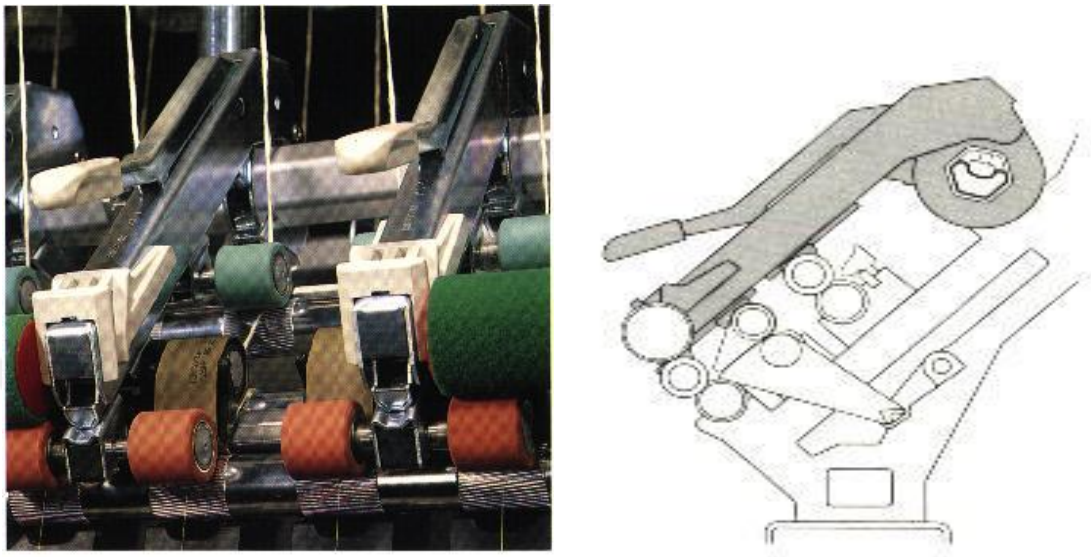
- Fitil bobinlerinin takılacağı bir cağlık,
- Fitilin çekilerek inceltileceği çekim sistemi,
- İplik üzerine bükümün dağıtılacağı kopça-bilezik tertibatı,
- İpliğin kops halinde masuraya sarılma düzeni,

- Otomatik takım değiştirme tertibatı.

Fitil cağlığı: Fitil bobinleri Ring İplik makinesinin cağlığına takılır. Bu cağlığa takılan fitil sayısı, iplik makinesinin iğ sayısına eşittir. Her fitil bir iğe hizmet vermektedir. Fitil bobinlerinin cağlığa takıldıkları cisme "fitil askısı" denir. Fitil askısının görevini emniyetli bir şekilde yerine getirebilmesi için; fitil bobinlerinin rahatlıkla takılabilir olması, düşey konumda tutulabilmesi, zor koşullarda bile aksaksız fitil akışının sağlanması, boş fitil masurasının rahatça çıkarılabilmesi gerekir.

Çekim sistemi: Çekim, bu kısımdaki silindirlerin birbirinden daha hızlı dönmesiyle oluşur. Çekim aparatında, arkadaki silindire az çekim verilir. Çünkü buraya giren fitilde bir miktar büküm vardır, çözülmesini rahatlıkla sağlayabilmek için az çekim verilmelidir. Çekimi muntazam yapabilmek için ve silindirler arasındaki mesafeyi kısaltıp elyaf yığılmasını önlemek maksadıyla 2. silindire apronlar geçirilmiştir. Apronlar sayesinde fitil alımı bozulmadan ön silindire yakın bir mesafeye taşınır. Çekimi değiştirmek için ön silindirden arka silindire hareket veren 2 dişliden herhangi birinin diş adedini değiştirmek gerekir. Dişlilerden birisi numara, diğeri arka dişlidir. Çekime etki eden faktörler; elyafın uzunluğu, kısa elyaf miktarı, fitilin kalınlığı, silindirlerin ölçü ve yapısı, çekim aparatının konstrüksiyonudur. Ring İplik makinelerinde 3 silindir ve 2 apronlu çekim sistemi kullanılmaktadır. Çekim sisteminin görevi; cağlığa takılı olan fitil bobininden sağılan fitilin, istenen iplik inceliği elde edilinceye kadar çekilerek inceltilmesini sağlamaktadır. 3 adet üst baskı silindirini ve üst apronu bünyesinde barındıran sisteme baskı kolu veya tabanca denilmektedir. Baskı kolunda giriş ve çıkış üst silindirleri olarak genellikle 28-35 mm çaplı, orta baskı silindir olarak da 25 mm çaplı silindirler kullanılır. Kullanılan hammaddeye bağlı olarak baskı silindirlerine uygulanabilecek baskı değerlerini değiştirebilmek mümkündür. Burada lif uzunluğuna ve lif yapısına bağlı olarak en iyi iplik neticelerini elde etmek üzere uygun olan baskı değeri ayarlanır. Çıkış silindiri üzerindeki baskının istenen değere ayarlanması özel bir anahtar yardımıyla yapılır. Baskının üzerindeki her farklı renk, farklı bir değeri ifade eder. Baskı yayları, konstrüksiyonları itibariyle özelliklerini uzun yıllar korur. Bu da ayarlanan baskı değerinin kendiliğinden değişmeyeceği anlamını taşır. Her üst silindirin kendine ait, bağımsız bir baskı elemanı mevcuttur. Dolayısıyla her çekim alanı için uygulanan baskı ayarı uygulandığı bölgeye ait kalır ve

komşu üst silindirler tarafından etkilenmezler. Şekil 4.3.'de çekim sistemi ve şematik gösterimi yer almıştır.



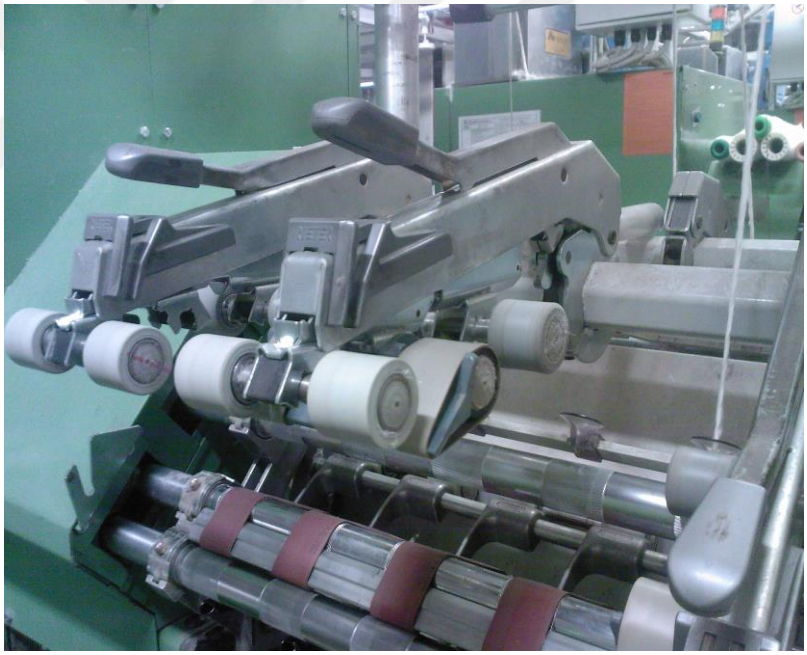
Şekil 4.3. Çekim sistemi ve şematik gösterimi.

Çekim sisteminde apronlar ve baskı manşonları: Çekim sistemi içinde liflerin transportunu gerçekleştiren apronlar üst ve alt apron olarak ikiye ayrılır. Kısa lif iplikhanesinde kullanılan apronlar aşağıdaki özelliklere sahip olmalıdır:

- Lif transportunu başarıyla gerçekleştirebilmeli,
- Aşınmaya karşı yüksek mukavemet göstermeli,
- Kopma mukavemeti yüksek olmalı,
- Sürtünme katsayısı düşük olmalı.

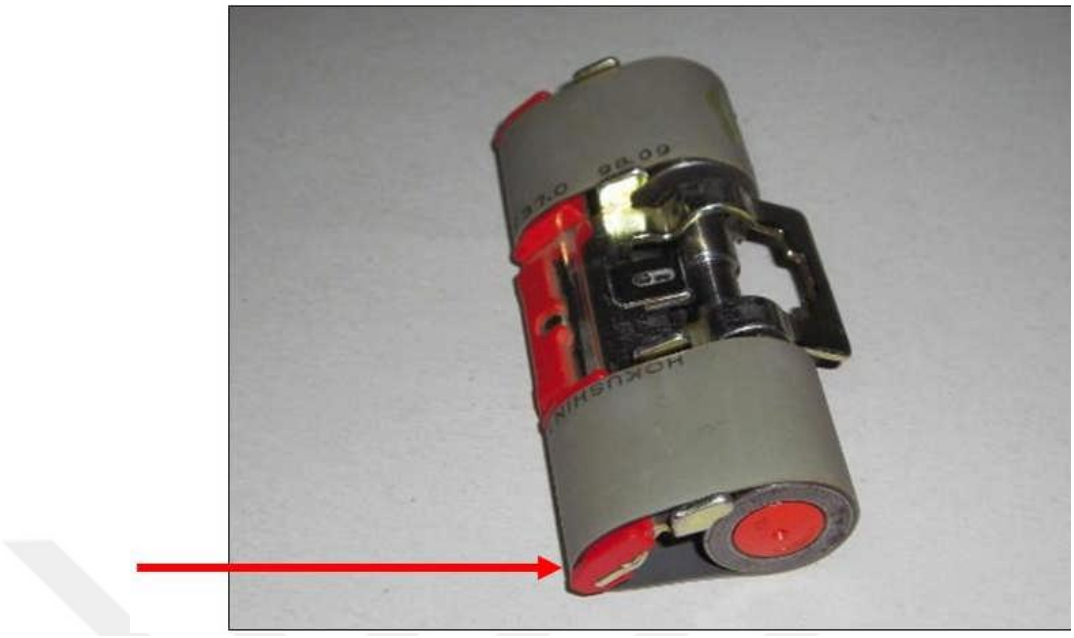
Apronlar açık ve kapalı apron olarak piyasada kullanılır. Kapalı apronlarda ek yeri olmayacağı için çalışma sırasında en ufak bir darbe olmayacaktır. Apronlar genellikle sentetik malzemeden imal edilir. Çekim sistemindeki lif kıstırma noktalarında liflerin zedelenmemesi için üst baskı silindirlerinin her iki tarafına kauçuk malzemeden imal edilmiş manşon geçirilmiştir. Ring İplik makinelerinde kullanılan manşonlar 60^0 shore ile 85^0 shore arasında imal edilir. Yumuşak manşonlarda liflerin kıstırılması ve sevki daha kontrollü olmakta, sert manşonlar ise daha fazla baskıya dayanıklı olmakta ve

elyaf sarığı teşkiline meyili daha az olmaktadır. Mesela, Ne 30/1 penye veya daha ince ipliklerde genellikle giriş silindirinde 82^0 shore, orta silindirde 73^0 shore ve çıkış silindirinde 65^0 shore sertliğinde manşon kullanılabilir. Belirli bir çalışma süresi sonunda (ortalama 1000-2000 çalışma saati) manşonların üzerinde aşınmalar, çapaklar ve ince yarıklar oluşur. Bu aşınmalar malzeme yorulması, yüksek baskı kolu basıncı veya manşon üzerinde oluşabilen elyaf sarıklarını temizlemek üzere işçinin, kullanımı yasak olan kesici madde kullanmasıyla meydana gelebilir. Aşınma durumunda bu manşonların taşlanması (rektifiye edilmesi) gerekir. Manşonların rektifiyesi, manşon asgari çapa düşünceye kadar tekrarlanır. Daha sonra manşon çekirdeğinden çıkarılarak atılır ve yenisi takılır. Şekil 4.4.'de manşon görülmektedir.



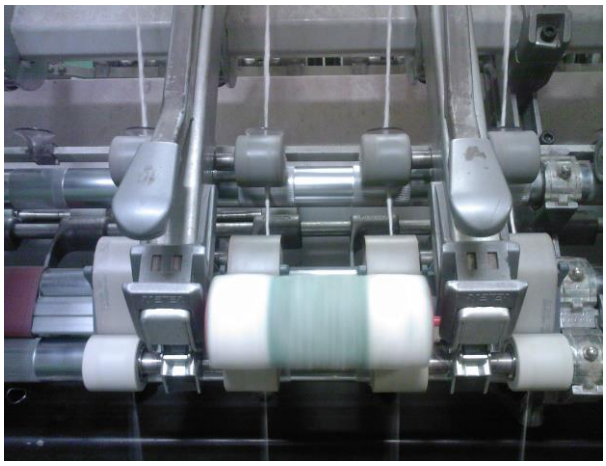
Şekil 4.4. Manşon görseli.

Üst apron kafesinde klipsler bulunur. Klipslerin görevi, alt apron ve üst apron arasındaki basıncı ayarlamaktır. Klips olmazsa apronlar arasındaki basınç bozulur, iplikte sürekli kopmalar meydana gelir. Aynı zamanda kalitesiz iplik üretilmiş olur. Çalışan ipliğin numarasına göre klips renkleri değişir. Şekil 4.5.'de klips görülmektedir.



Şekil 4.5. Klips görseli.

Çekim sistemi temizliği; hem gezer temizleyiciler hem de üzeri çuha kaplı silindirler vasıtasıyla sağlanır. Ayrıca Ring temizlik ekibi her gün belli sayıda Ring İplik makinesini temizliğe açarak özellikle çekim sistemini temizlik tabancalarıyla temizler. Şekil 4.6.'da temizleme silindiri görülmektedir.



Şekil 4.6. Temizleme silindiri görseli.

Büküm bölümü: Ring İplik makinesinin çekim sistemini terk eden lif bandını bir arada tutmak için büküm verilir. Fitil makinesine oranla çekim sistemini terk eden lif bandına

Ring İplik makinesinde kuvvetli miktarda büküm verilmesi gerekmektedir. Çünkü Ring İplik makinesinde çekim sistemini terk eden liflerin kesitteki sayısı, fitildekine göre çok azdır. Ring İplik makinesinde ipliğe verilen bükümü aşağıdaki faktörler etkiler:

- Kullanılan hammadde özellikleri (lif boyu vs.),
- Eğrilecek ipliğin inceliği (numarası),
- İpliğin kullanım yerinin dokuma (çözgü ve atkı ipliği) veya trikoya yönelik olacağı (kapalı büküm ya da açık büküm olacağı).

Ring İplik makinesinde iğ devri aşağıdaki faktörlere bağlı olarak değişiklik gösterir:

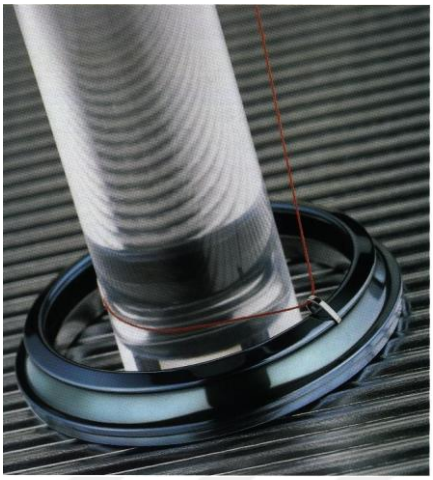
- Teknolojik faktörler (kopça hızı, bilezik çapı vs.),
- İpliğe bağlı faktörler (büküm).

Bükümün meydana gelmesinde son çekim silindirlerinden çıkan ipliğin hızının, iğ devir sayısının, bilezik ve kopçanın etkisi vardır.

Kopça ve bilezik: Bükümün iplik üzerine dağıtılması ve ipliğin kopsa sarılmasında etkili olan, iğ devir sayısını sınır koyan kopça ve bilezik, Ring İplik ve Ring Büküm proseslerinin iki elemanıdır. Makinenin verimini ve çalışma koşullarını büyük oranda etkilerler. Bilezik üzerinde büyük bir hızla dönen kopça şu iki fonksiyonu yerine getirir:

- Çekim sisteminden sevk edilen ipliğe bükümün dağıtılması,
- İpliğe uygun bir gerginlik kazandırarak masura üzerine kops formunda sarılması.

Düzgün bir sarım ve kops yapısı için bilezik kopçaya rehberlik eder. Kopça tarafından yerine getirilen iplik üzerine büküm dağıtılması ve ipliğin kops halinde sarılması işlemleri, aşırı derecede gerilmeden yapılmalıdır. Pamuk iplikhanesinde, C tipi, N tipi, eliptik veya yarı eliptik kopçalar önem arz etmektedir. Kopçalar yuvarlak veya basık kesitlidir. Yuvarlak kesitler pamukta tercih edilir. İplikhanelerde çelik kopçalar kullanılır. Çelik kopçalar belli oranda sertleştirilmiş ve yüksek derecede polisajlanmıştır. C tipi, N tipi, eliptik kopçalar gibi flanş bilezik kopçaları yağlanmazlar. Şekil 4.7.'de kopça ve bilezik görülmektedir.



Şekil 4.7. Kopça ve bilezik görseli.

Düzgün ve sarsıntısız kopça çalışması için aşağıdaki şartların geçekleşmesi gerekir:

- Bilezik bankosunun iğlere göre konumu kusursuz ve hareketi her zaman aynı olmalıdır.
- Bilezik ve iğlerde olduğu kadar balon kırıcı ve iplik kılavuzunun da aynı eksende ve iyi merkezlenmiş olması gerekmektedir.
- Bilezik tam yuvarlak ve temas yüzeyinin yatay pozisyonda olması gereklidir.
- İğin salgısız çalışması ve masuranın eksantrik bir çevre yüzeyine sahip olmaması gerekmektedir.
- Bileziğin kopça çalışma yüzeyi kusursuz olmalıdır. Bilezik üzerinde yanlış ve farklı kopça kullanımından kaynaklanan iz ve çapaklar bulunmamalıdır.
- Kopça temizleyicisi doğru ayarlanmalıdır (kopça ile temizleyici arası yaklaşık 0.5 mm olmalıdır).
- Bilezik çapı ile masura çapı oranı uygun olmalıdır. Burada bilezik çapının masura çapına oranının 2 veya daha küçük olması tavsiye edilir.
- Çalışılan iplik numarası, iplik kalitesi ve kopça formatı dikkate alınarak flanş bileziklerinde flanş eni doğru seçilmelidir.

Değişmeyen bir ısı ve nem oranı kopçanın çalışmasını olumlu yönde etkiler. Salon klima değerlerindeki oynamalar (örneğin artan rutubet) kopça aşınmasını arttırır. En uygun bilezik ve kopça formunun tayini en yüksek verimlere ulaşmanın başta gelen

koşuludur. Bilezik profili ile kopçanın formunun iyi uyumu sayesinde kopçanın bilezik üzerinde stabil konumu sağlanmış olur. Kopça bilezik üzerinde yeterince serbest hareket edebilmelidir.

Kopça numarası: Kopça ağırlığının (numara) çalışılan ipliğin numarasına uygun olması gerekmektedir. Burada aynı şekilde iğ devri, hammadde, balon büyüklüğü ve kops sertliği göz önüne alınmalıdır. Balon formunun iplik kopuşlarına olan etkisi büyüktür. İplik balonu normal koşullarda balon kırıcıya hafifçe temas etmelidir. Çok gevşek ve çok gergin iplik balonu oluşmasına (çok hafif ve çok ağır kopça nedeniyle) engel olunmalıdır. Bu tür balonlar iplik kopuşlarına, kopça aşınmalarına ve iplik kalitesinin bozulmasına neden olur. Kopçaya rehberlik yapan bileziğin hammaddesi birinci sınıf rulman çeliğidir. Bilezikler, beklenilen yüksek performansı karşılayabilmek için kopça sertliğine uygun olarak tamamen sertleştirilir. Flanş formu kadar bilezik çalışma yüzeyinin yuvarlaklığı da kopçanın çalışma özelliklerini etkileyen faktördür.

Kopça hızı: İğ devir ve bilezik çapına bağlı olarak değişir ve aşağıdaki formüle göre hesaplanır:

$$Kopça\ hızı = bilezik\ çapı(mm) \times iğ\ devri(devir/dakika) \times \Pi / 1000 \times 60 = m/sn \qquad (4.1.)$$

Bilezik alıştırması (Rodaj işlemi): Yeni bilezik değişimlerinde uygulanan rodaj işlemi, metalik çalışma yüzeyinin düzeltilmesi, kopçanın bilezik üzerinde kendisine bir kulvar hazırlaması ve ilk çalışma şartlarının optimum seviyeye en kısa zamanda getirilmesi için yapılır.

Bileziklerin temizlenmesi: Bilezikler müşteriye teslim edilmeden önce üzerleri pas önleyici solüsyon ile kaplanırlar. Satın alınan bilezikler yağ çözücü madde kullanılmadan silinmelidir. Flanş bilezikler kuru olarak çalışırlar. Bileziklerin yerine takılması; bileziklerin tutucularına sıkı olarak oturtulması ve oynamamalıdır. Bilezik ve iğ merkezlenmelidir.

Kopça seçimi: Kopçalar, rodaj kopçası ve çalışma kopçası olarak ikiye ayrılır. Rodaj sırasında kopça firmasının tavsiye ettiği rodaj kopçası kullanılır. Rodaj işleminin

bitirilmesi: Tavsiye edilen rodaj işleminin devamı esnasında eğer kopça aşınması tespit edilemiyorsa rodaj işlemi o safhada sona erdirilir.

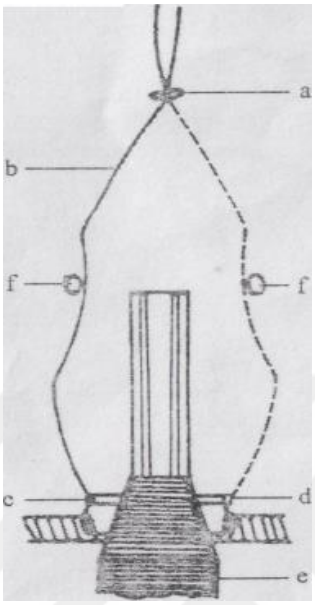
Büküm: Bükümün artması ile iplikteki mukavemet artarken, özellikle o iplikten örülmüş kumaşın tuşesi (tutumu) sertleşir ve Ring İplik makinesinde birim zamanda çıkan üretim azalır. Bir ipliğe verilecek büküm miktarı, o ipliğin daha sonraki kullanım yerine göre seçilir. Bükümün derecesi genellikle ikiye ayrılır. Açık büküm, triko(örme) ipliklerine verilir. Örme makinesinde ipliklerin karşılayacağı direnç nispeten düşük olduğu için triko ipliği üzerindeki büküm miktarı dokuma ipliğine göre düşüktür. Kapalı büküm, dokuma ipliklerine verilir, dokuma tezgahında özellikle çözgü ipliklerinin üzerine binen yük oldukça fazla olduğu için dokuma ipliklerine, örmeye nispeten yüksek büküm verilir. Dolayısıyla dokuma ipliğinin mukavemeti trikoya göre daha yüksektir.

İpliğin sarılması: İğin dönmesi sonucu çekim sistemiyle iğ arasındaki iplik parçası bir devir hareketi kadar kaydırılarak iğ etrafına dolanır. Bu ipliğin temas ettiği makine elemanı kopçadır. İplik ve kopça bilezik etrafında döner ve planganın o sırada bulunduğu konumu itibarıyla masuraya sarılır. Kopça, kendi ağırlığından dolayı hareketsiz kalmak isterken, iplikteki gerdirme kuvveti kopçayı bilezik üzerinde döndürür. Bilezik ile kopça arasında meydana gelen sürtünme kuvveti ilave bir frenleme olarak çalışır. İpliğin sevki ile beraber kopça üzerindeki gerdirme kuvveti azalır; fakat kopça iğe göre, sevk edilen iplik parçası kadar geriye kalır. Bunun neticesinde iplik üzerine sarılmış olur.

Sarım gerginliği: Masuranın konik yapısından dolayı iplik gerginliği devamlı değişmektedir. Küçük sarım çapında iplik gerginliği büyük sarım çapına göre daha fazladır. Kops sarım başlangıcındaki gerginlik oranı kops koniğinin tamamlanması sırasında da aynı şekildedir. İplik gerginliğinin arttığı bu konumlarda iplikte artması muhtemel iplik kopuşlarının önlenmesi için iğ devri, programlanabilen otomatik devir regülatörü ile değiştirilir.

Balon kırıcılar: İplik kopuş sayısının azalmasını sağlarlar. Kopuş sayısını azaltmak için makinenin çalışma hızını düşürmek üretim kaybıyla eş anlama geldiği için uygulanmaz.

Bunu hafif kopça ile sağlamak sadece teorik olarak mümkündür, çünkü bu durumda iplik balonu genişler. Şekil 4.8.'de balonlaşmanın oluşumu görülmektedir.



a) İplik kılavuzu
b) İplik
c) Kopça
d) Bilezik
e) İplik kopsu
f) Balonlaşma bileziği

Şekil 4.8. Ring iplik eğirmede balonlaşmanın oluşumu.

İğ devrini etkileyen faktörler: İplik makinesinde iğ devrinin yükselmesiyle üretim artar. İğ devrini etkileyen faktörler aşağıdaki gibidir:

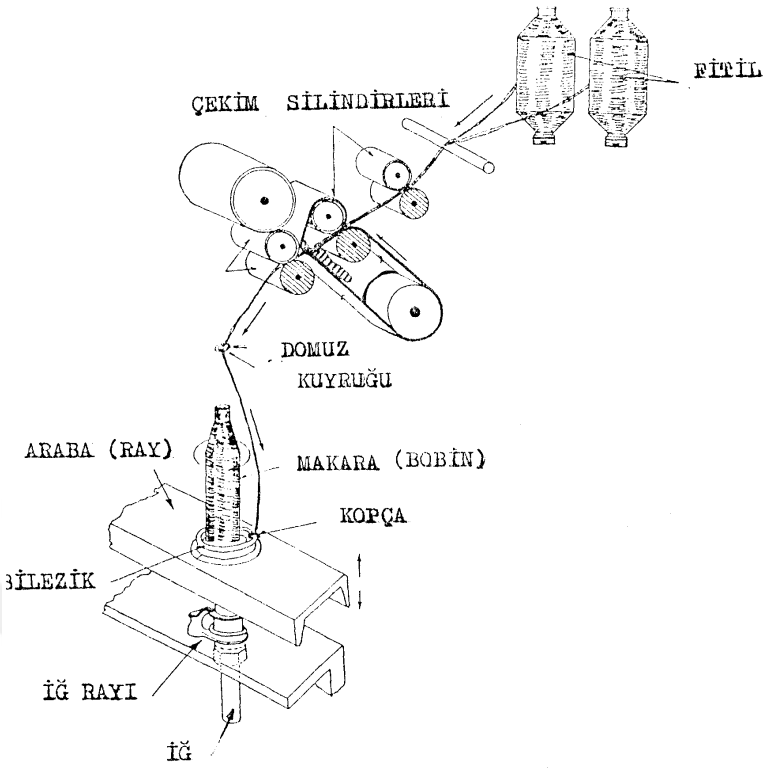
- Bilezik çapı küçüldükçe iğ devri artar,
- Kopça hızı yükseldikçe iğ devri artar,
- İplik numarası inceldikçe iğ devri artar,
- Bükümün ve dolayısıyla mukavemetin daha yüksek olmasından dolayı dokuma ipliklerinde, trikoya oranla daha yüksek iğ devirlerine çıkılır.

İplik masuraları: Ring İplik makinelerinde kullanılan ve ipliğin üzerinde kops formu teşkil edecek şekilde sarıldığı masuralar, büyük çoğunlukla sertleştirilmiş plastik maddeden imal edilmiştir. Takımın elle değiştirildiği makinelerde daha ucuz polipropilen malzemeden mamul iplik masuraları kullanılırken, sürtünmenin daha yüksek olduğu dofferli makinelerde ABS veya poliüretan malzeme kullanılmaktadır. Masuranın boyu genel olarak kullanılan bilezik çapının 5 katı olarak seçilmelidir.

Otomatik takım değiştirme tertibatı: Otomatik takım değiştirmenin faydaları aşağıdaki gibidir:

- Takım değiştirme süresi 2-2.5 dakika gibi kısa bir zaman süresinde yapıldığı için duruş süresi azalmakta ve üretim randımanı artmaktadır,
- Erken takım düşürmeye gerek kalmadığı için hem ring iplik hem de bobin makinesinde kops daha uzun süre çalışmakta, böylece makine randımanı artmaktadır,
- Takım değiştirme çok az bir süre aldığı için kalın numaralar için bile geçerli olmak üzere bilezik çapı küçük seçilmekte ve küçük bilezik çapından dolayı yüksek iğ hızlarına ulaşılabilmektedir,
- Takımcıya gerek kalmadığı için personel sayısı azalır,
- Boş masuranın iğe oturtulması robot kol vasıtasıyla ve bütün iğlerde aynı güç sarfıyla yapılacağından, iğ dibi sarığı ve masura kusuru olmadığı sürece bütün masuralar iğlere aynı derinlikte oturacak ve sarım, masura dibinden itibaren ayarlanan yükseklikten itibaren (örneğin 15 mm) başlayacaktır.

Şekil 4.9.'da ring iplik eğirmesinin şematik hali görülmektedir.



Şekil 4.9. Ring iplik eğirme şematik gösterimi.

İplik kalitesine etki eden faktörlerden ve Ring İplik makinesi çalışma prensibinden bahsedilmiştir. Bu faktörlerden iplik kalitesine en çok etki eden 6 faktör, işletme mühendisi ile tekstil makineleri literatürü baz alınarak belirlenmiştir. 6 faktörün her birinde 3'er seviye bulunmaktadır (Tablo 4.1.). Bu faktörlerin birbirini etkileme ihtimali baz alındığında 6 faktörün her biri için iplik kalitesini ne kadar etkilediği, hangi seviyede optimum çalıştığı, kaliteyi olumsuz anlamda en az düzeyde etkileyen faktör ve seviye hangisi gibi bir çok soru ortaya çıkmaktadır. Her makinenin 2.5 saatte 100 kg ve günlük 1 ton iplik ürettiği, iplik maliyetinin 5 dolar/kg olduğu düşünüldüğünde, 76 Ring İplik makinesi olan işletmede iplik kalitesinde oluşabilecek sorunlar ciddi maliyetler yaratmaktadır. İşletmenin bu soruları deneme yanılma ile çözmeye çalışması hem uzun sürmekte hem de maliyetli olmaktadır. Çünkü üretilen kalitesiz iplik hem para hem de zaman kaybına sebep olmaktadır. Bu çalışmada bu zaman ve para kaybını en az düzeye indirecek şekilde bir bilgi işleme sistematiği getirmek hedeflenmektedir.

İplik kalitesini ölçme testi, uluslararası bir standart olan Uster standartlarına göre değerlendirilmektedir. İşletme testlerini en son teknoloji Uster Tester 5 cihazı ile

yapmaktadır. İplik kalitesinde; ince yer, kalın yer, tüylülük ve neps değerlerine bakılmaktadır. Ayrıca standardın bir ifadesi olan ve bahsedilen 4 değerin çeşitli katsayılarla çarpılması ile oluşturulan bir Uster değeri de bulunmaktadır. İşletme 5 değeri de inceleyerek ipliğin kalitesine karar vermektedir. Bu değerlerden en az bir tanesinin sonucunun olumsuz çıkması o ipliğin kalitesinin bozuk olacağı anlamına gelmektedir.

Bu çalışmada işletmede sürekli üretilen ve diğer iplikler konusunda da fikir sahibi edecek KLZ140NA iplik tipi üzerinde uygulama çalışması gerçekleştirilmiştir.

- K:Karde
- L:Likralı
- Z: Z büküm
- 140: 14 Ne
- N: Normal İplik
- A: İşletme versiyonu anlamları taşımaktadır.

Testlerin sonucunun değişkenlik göstermesini engellemek için çalışma 75 nolu makinede ve 1000-1005 iğlerinde yapılmıştır. Makine ve iğler sabit tutulmuştur.

Yukarıdaki bilgiler ışığında 6 faktörün ve 3 seviyesinin birbiri ile etkileşimini anlamak için 3^6=729 adet test yapılması gerekmektedir. Günde 4 adet test yapılabildiği göze alındığında bu gözlem için 182 gün gerekmektedir. Bu zaman kaybı yaratmaktadır ve testlerin olumsuz çıkması sonucu; 1 ton/gün * 5 $/kg = 5000 $/gün para kaybı oluşacağı aşikardır.

Test sayısını daha aza indirmek için faktörleri azaltmamız ya da faktörlerin en etkili seviyelerini belirlememiz gerekecektir. Bu çalışmada bu işlem için Taguchi deneysel tasarım metodu kullanılmış olup, faktör önem sıralamaları ve en etkili seviyeler belirlenmiştir. Çıkan sonuçlardaki veriler, veri madenciliği paket programları ile işlenmiş olup sınıflandırma yöntemleri ile kural kümesi ortaya çıkarılmıştır.

4.2. Taguchi Deneysel Tasarım Metodu

Taguchi yöntemi Dr. Genichi Taguchi tarafından 1950'lerde süreç en iyileme tekniği olarak geliştirilmiştir. Taguchi'nin kalite alanına getirmiş olduğu en dikkat çekici katkı, kalite sistemini üretim öncesi ve üretim süreci olarak ikiye ayırarak bir ürünün kalitesinin ve müşteri memnuniyetinin, üretim öncesindeki aşamada tasarım ve geliştirmenin mükemmelliği ile yakından ilgili olduğunu gösteriyor olmasıdır [73].

Taguchi yöntemi farklı parametrelerin, farklı seviyeleri arasından en iyi kombinasyonu saptamak için oldukça kullanışlı bir yöntemdir. Her bir parametrenin, her bir seviyesini içeren tüm kombinasyonlar için oldukça fazla deneysel çalışma yapılması gereken durumlarda Taguchi yöntemi ile dikey dizi tablosu (Tablo 4.2.) kullanılarak çok daha az sayıda deneysel çalışmayla sonuca ulaşmak mümkündür [74].

Parametreler belirlendikten sonra Tablo 4.2.'deki Taguchi dikey dizisinden bir dizi seçmek gerekmektedir. Tablo 4.2.'ye göre; iplik kalitesine etki eden 6 parametre ve her bir parametrenin 3 seviyesi olduğuna göre L27 dizisi en uygun dizi olarak seçilir. Tablo 4.2.'nin dışında kalan parametre ve seviyeler için deney şartları daha zor olduğundan parametre ya da seviye küçültülerek uygun diziye getirilmesi gerekmektedir.

Tablo 4.2. Taguchi dikey dizi seçim tablosu.

PARAMETRE SAYISI	SEVİYE SAYISI													
	2			3			4			5		
P=2	S=2	L4	P=2	S=3	L9	P=2	S=4	L16	P=2	S=5	L25
P=3	S=2		P=3	S=3		P=3	S=4		P=3	S=5	
P=4	S=2	L8	P=4	S=3		P=4	S=4		P=4	S=5	
P=5	S=2		P=5	S=3	L27	P=5	S=4		P=5	S=5	
P=6	S=2		**P=6**	**S=3**					P=6	S=5	
P=7	S=2		P=7	S=3							
P=8	S=2	L12	P=8	S=3							
P=9	S=2		P=9	S=3							
P=10	S=2		P=10	S=3							
P=11	S=2		P=11	S=3							
P=12	S=2	L16	P=12	S=3							
P=13	S=2		P=13	S=3							
P=14	S=2										
P=15	S=2										
P=16	S=2	L32									
P=17	S=2										
P=18	S=2										
P=19	S=2										
P=20	S=2										
P=21	S=2										
P=22	S=2										
P=23	S=2										
P=24	S=2										
P=25	S=2										
P=26	S=2										
P=27	S=2										
P=28	S=2										
P=29	S=2										
P=30	S=2										
P=31	S=2										

Taguchi tasarımında bir ürünün kalite sağlama seviyesi hem ürün tasarımı hem de süreç tasarımını kapsayan 3 tasarım üzerine kurulmuştur. Bunlar:

- Sistem tasarımı: kavram oluşturma aşamasıdır.
- Parametre tasarımı: ürün ve süreç için hedef oluşturma aşamasıdır.
- Tolerans tasarımı: sonuç istenen hedefe ulaşılamadığında yapılan ilave çalışmalardır.

Sistem tasarımı: Bu adımda eldeki bütün materyaller değerlendirilirken mevcut teknolojik yenilikler araştırılır ve sistem içerisinde kullanabilirliği üzerine fizibilitesi yapılır. Bu adımda amaç, en az maliyetle en iyi ürün tasarımı ve maksimum müşteri memnuniyetidir [75].

Parametre tasarımı: Süreç iyileştirme ve geliştirmenin en önemli adımı parametre tasarımıdır. Bu adımda üretilecek ya da geliştirilecek olan ürünün özelliklerinin en iyi seviyeye getirilebilmesi için üretimde kullanılan parametrelerin iyileştirilmesi sağlanır. Parametrelere en iyi seviyeler seçilir. Üretim esnasında ürünün kalitesini olumsuz etkileyecek kontrol edilemeyen etkiler belirlenir ve bunlara kontrol edilemeyen parametre adı verilir. Ardından bu parametrelerin etkileri en küçüklenir. Bu adımda parametreler bloklanırken Taguchi'nin geliştirmiş olduğu dikey diziler kullanılır. Aynı zamanda sinyal gürültü oranı (S/N - Signal/Noise ratio) analizi ile de hesaplama yapılabilir. Aynı zamanda ortalamalar ve gürültü oranı (S/N means) değerleri de hesaplanarak kaydedilir [75].

Tolerans tasarımı: Tolerans tasarımı, parametre belirleme çalışmaları sonucu istenilen hedefe ulaşılamadığında yapılan ilave çalışmalardır. Bu aşamada gözlenen değerlerden faydalanılarak ürünün hedef değerden sapma göstermesinin getirdiği kayıplar bulunarak sapmalar azaltılır [75].

Taguchi kayıp fonksiyonu olarak bilinen ve aynı zamanda gürültü oranı (S/N-Sinyal/Noise ratio) fonksiyonu olarak da ifade edilen 3 farklı amaca uygun fonksiyon bulunmaktadır. Buna göre, amacın "en küçük en iyi", "en büyük en iyi" ve "nominal en iyi" olmasına göre aşağıdaki eşitlikler (Eş. 4.2.-4.6.) kullanılarak S/N oranları hesaplanır.

En düşük (küçük) en iyi olduğu durumda:

$$S/N = -10\, log\left(\frac{1}{n}\sum_{i=1}^{n} y_i^2\right) \tag{4.2.}$$

En yüksek (büyük) en iyi olduğu durumda:

$$S/N = -10\, log\left(\frac{1}{n}\sum_{i=1}^{n} \frac{1}{y_i^2}\right) \tag{4.3.}$$

Nominal en iyi olduğu durumda:

$$\hat{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (4.4.)$$

$$S/N = 10\ log\left(\frac{\hat{y}^2}{S^2}\right) \quad (4.5.)$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{y})^2 \quad (4.6.)$$

Eşitliklerde y_i : performans yanıtının *i.* gözlem değeri, *n:* bir denemedeki test sayısı, $\hat{y}$: gözlem değerlerinin ortalaması ve S^2: gözlem değerlerinin varyansını ifade etmektedir.

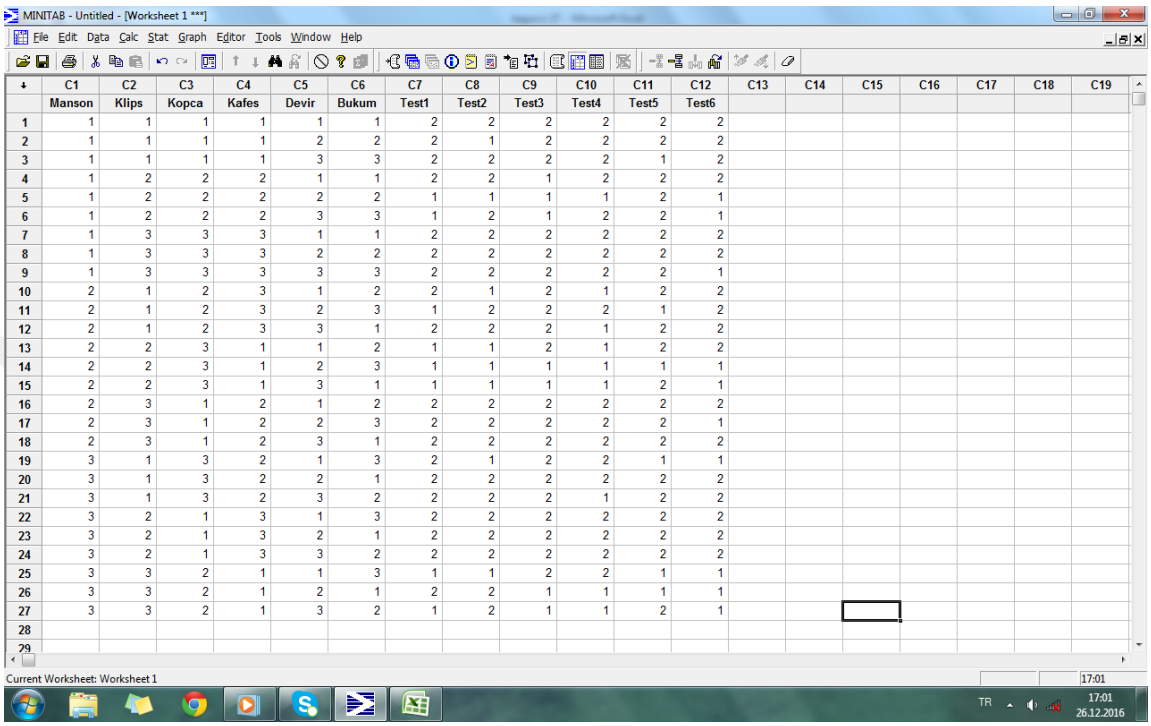
Taguchi yönteminin gerek geniş kullanım alanına sahip olması, gerekse daha az deney yaparak hem zaman kazancı, hem de daha az maliyetle sonuçların elde edilmesine imkân sağlaması gibi avantajlar sunması, tez kapsamında kullanılmasında etkili olmuştur. Eğer Taguchi deneysel tasarım metodu kullanılmasa idi, 6 faktörün ve 3 seviyesinin birbiri ile etkileşimini anlamak için 3^6=729 adet test yapılması gerekecek idi. Bu durum, işletme için zaman ve maliyet açısından imkansızdır.

4.3. Taguchi Deney Tasarım Metodu İle İplik Kalite Seçimine Etki Eden Faktörlerin İncelenmesi

Bu çalışmada Ring İplik makinesinde iplik kalitesine etki eden 6 faktör belirlenmiştir. Bu 6 faktörün her birine ait 3 adet seviyesi vardır. Bu faktör ve seviyeler Tablo 4.1.'de yer almaktadır.

Her biri üçer seviyeli 6 faktör için Minitab14 paket programından çıkan sonuç L27 Taguchi Tasarımı modelidir. L27 tasarımındaki faktör seviyeleri kullanılarak yapılan deneyler sonucunda elde edilen "en küçük en iyi" durumu için hesaplanan S/N oranları Tablo 4.3.'de yer almaktadır. Test işleminin sağlıklı olması açısından her test 6 kez tekrarlanmıştır.

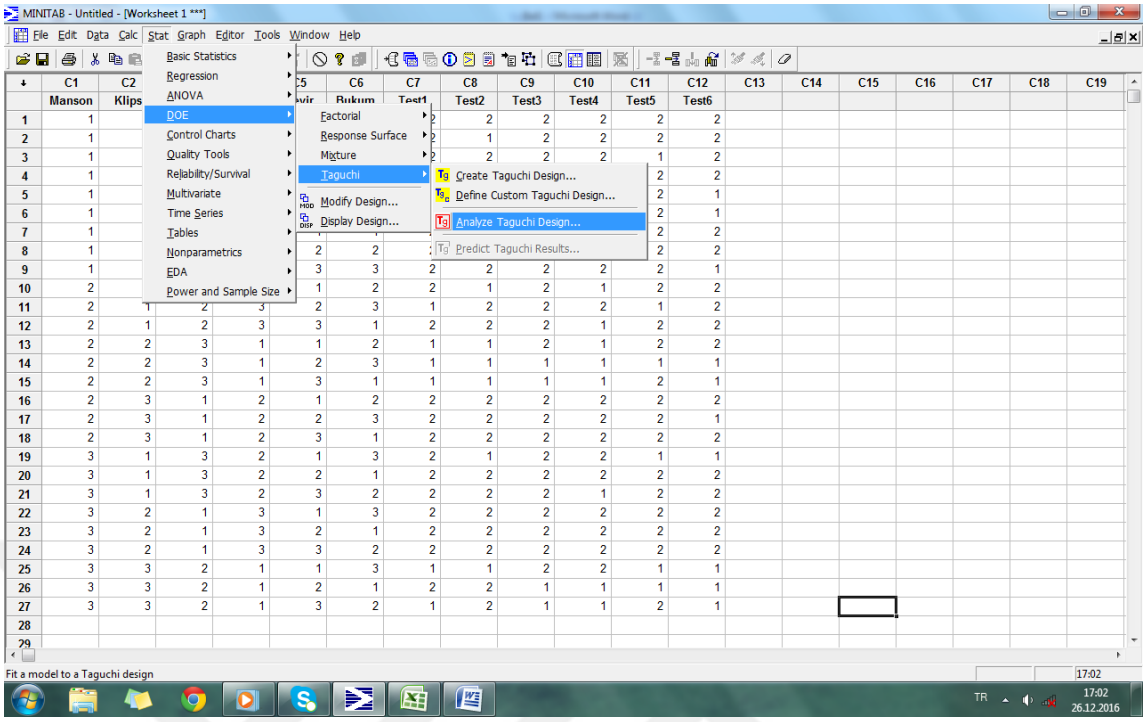
Şekil 4.10.-4.16.'da iplik test verilerinin Minitab14 paket programındaki Taguchi deneysel tasarım metodu uygulama ekranları görülmektedir.



	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19
	Manson	Klips	Kopca	Kafes	Devir	Bukum	Test1	Test2	Test3	Test4	Test5	Test6							
1	1	1	1	1	1	1	2	2	2	2	2	2							
2	1	1	1	1	2	2	2	1	2	2	2	2							
3	1	1	1	1	3	3	2	2	2	2	1	2							
4	1	2	2	2	1	1	2	2	1	2	2	2							
5	1	2	2	2	2	2	1	1	1	1	2	1							
6	1	2	2	2	3	3	1	2	1	2	2	1							
7	1	3	3	3	1	1	2	2	2	2	2	2							
8	1	3	3	3	2	2	2	2	2	2	2	2							
9	1	3	3	3	3	3	2	2	2	2	2	1							
10	2	1	2	3	1	2	2	1	2	1	2	2							
11	2	1	2	3	2	3	1	2	2	2	1	2							
12	2	1	2	3	3	1	2	2	2	1	2	2							
13	2	2	3	1	1	2	1	1	2	1	2	2							
14	2	2	3	1	2	3	1	1	1	1	1	1							
15	2	2	3	1	3	1	1	1	1	1	2	1							
16	2	3	1	2	1	2	2	2	2	2	2	2							
17	2	3	1	2	2	3	2	2	2	2	2	1							
18	2	3	1	2	3	1	2	2	2	2	2	2							
19	3	1	3	2	1	3	2	1	2	2	1	1							
20	3	1	3	2	2	1	2	2	2	2	2	2							
21	3	1	3	2	3	2	2	2	2	1	2	2							
22	3	2	1	3	1	3	2	2	2	2	2	2							
23	3	2	1	3	2	1	2	2	2	2	2	2							
24	3	2	1	3	3	2	2	2	2	2	2	2							
25	3	3	2	1	1	3	1	1	2	2	1	1							
26	3	3	2	1	2	1	2	2	1	1	1	1							
27	3	3	2	1	3	2	1	2	1	1	2	1							
28																			
29																			

Şekil 4.10. Test verilerinin Minitab14 paket programına girilmesi.

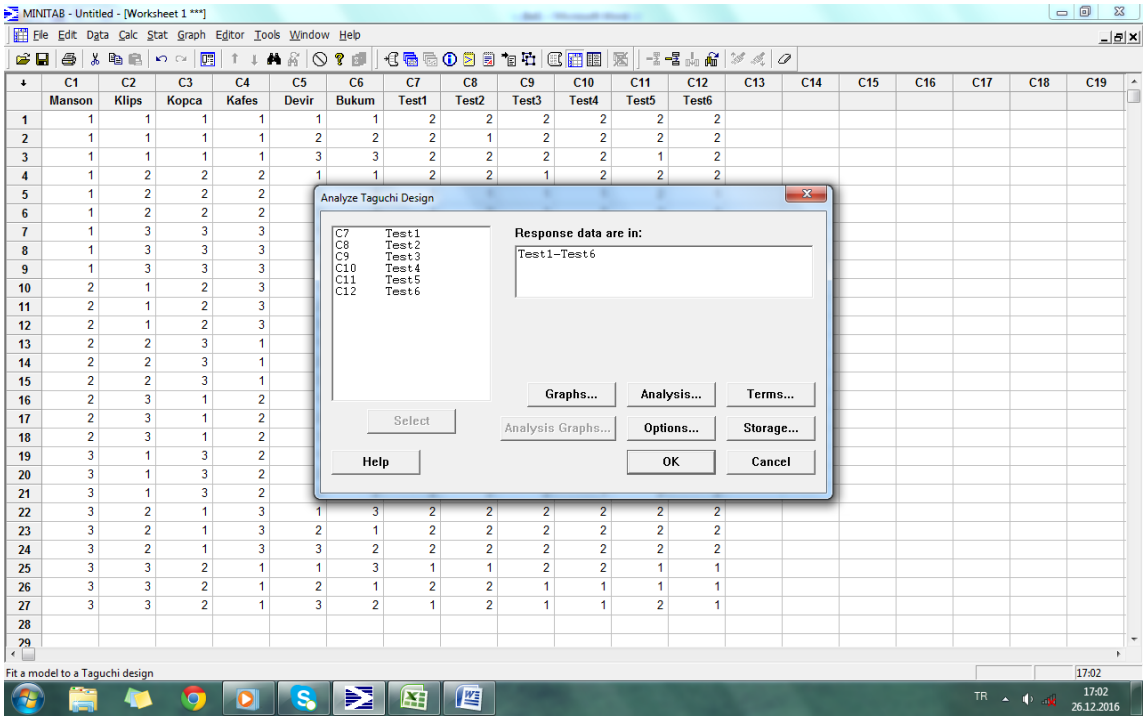
C1-C6 sütunları, iplik kalitesini etkileyen faktörleri göstermektedir. Faktörlerin seviyeleri 1, 2 ve 3 olarak ifade edilmiştir. C7-C12 sütunları, her bir veri kümesinin 6 test sonucunu belirtmektedir. Sonuç pozitif ise 1, negatif ise 2 ile ifade edilmiştir.



Şekil 4.11. Test verilerinin Minitab14 paket programında analiz ekranı.

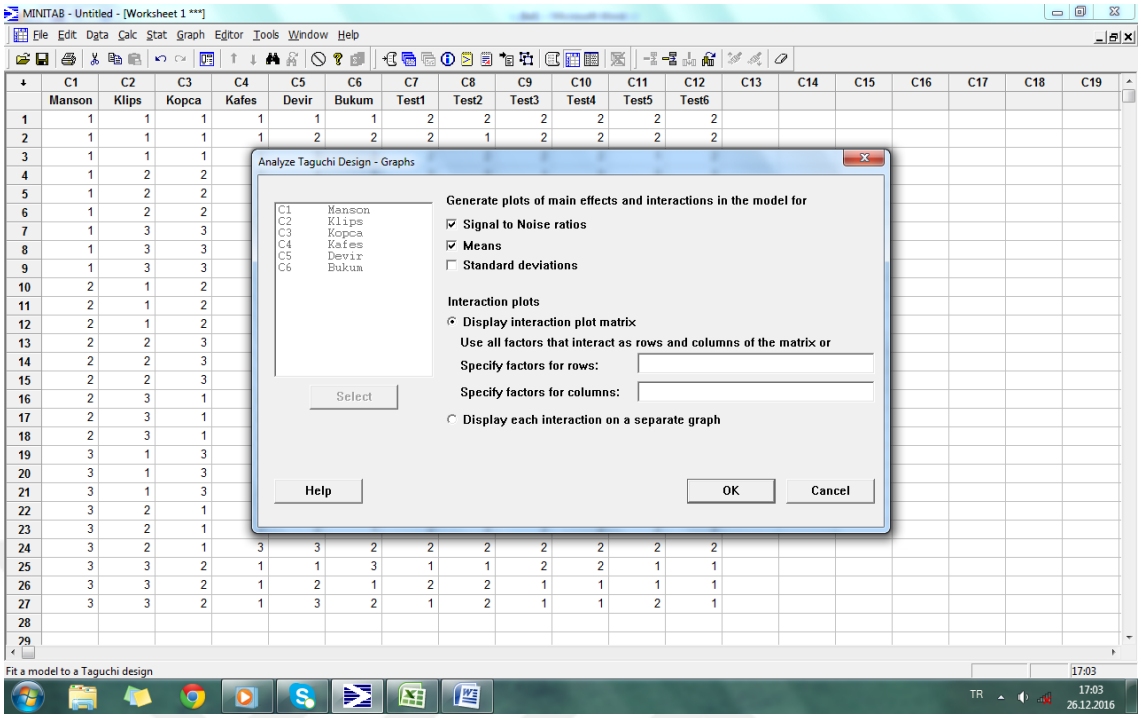
Analiz ekranı, Minitab14 de; Stat- DOE-Taguchi işlemleri takip edilerek açılır.

.



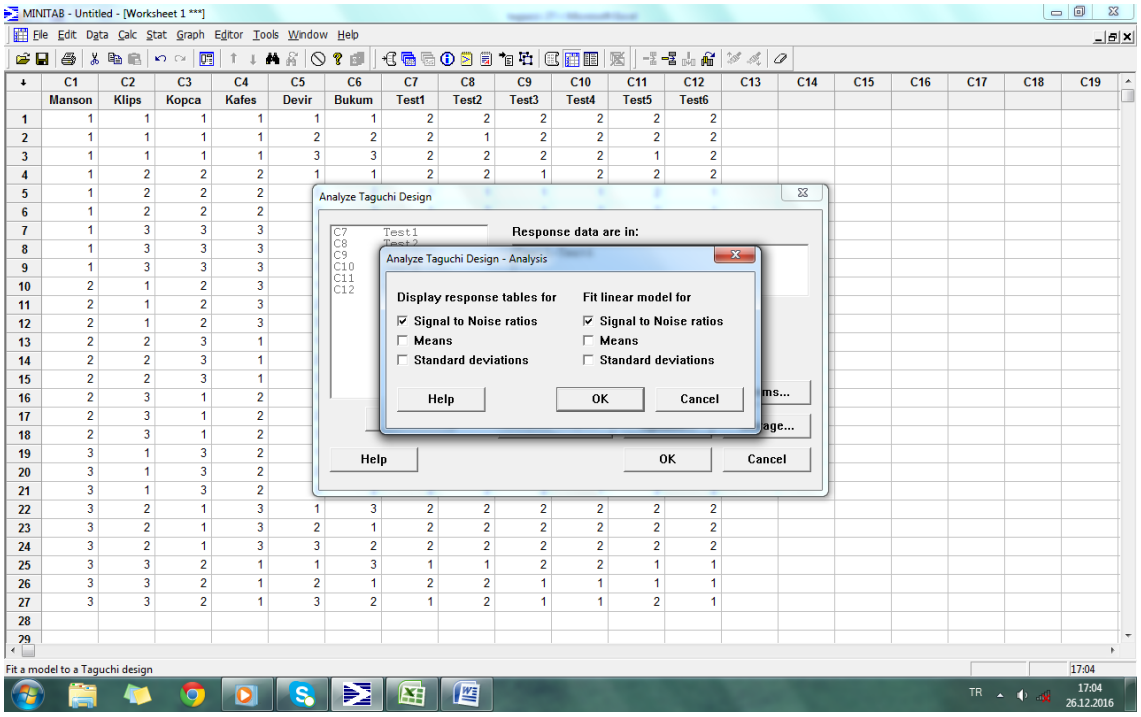
Şekil 4.12. Minitab14 paket programında girdi ve çıktı parametrelerinin belirlenmesi.

Girdi bilgileri sol kutucuğa, çıktı bilgileri sağ kutucuğa girilir.



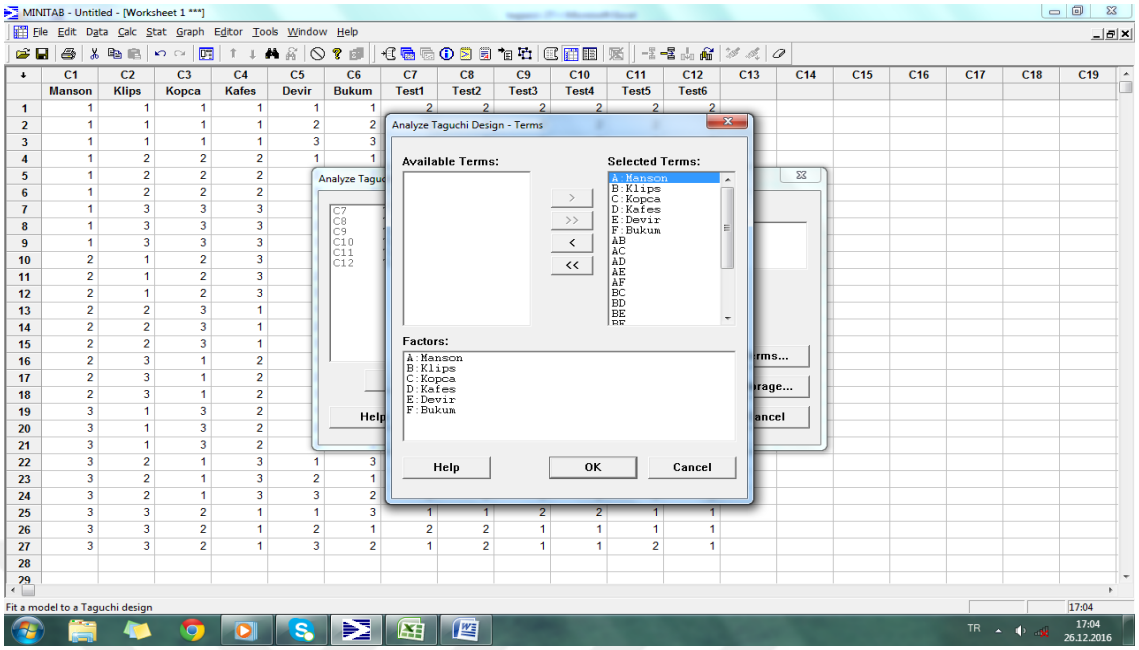
Şekil 4.13. Minitab14 paket programında sonuç verilerinin belirlenmesi.

Girdilerin, ana etkileri ve etkileşim modelleri belirlenir.



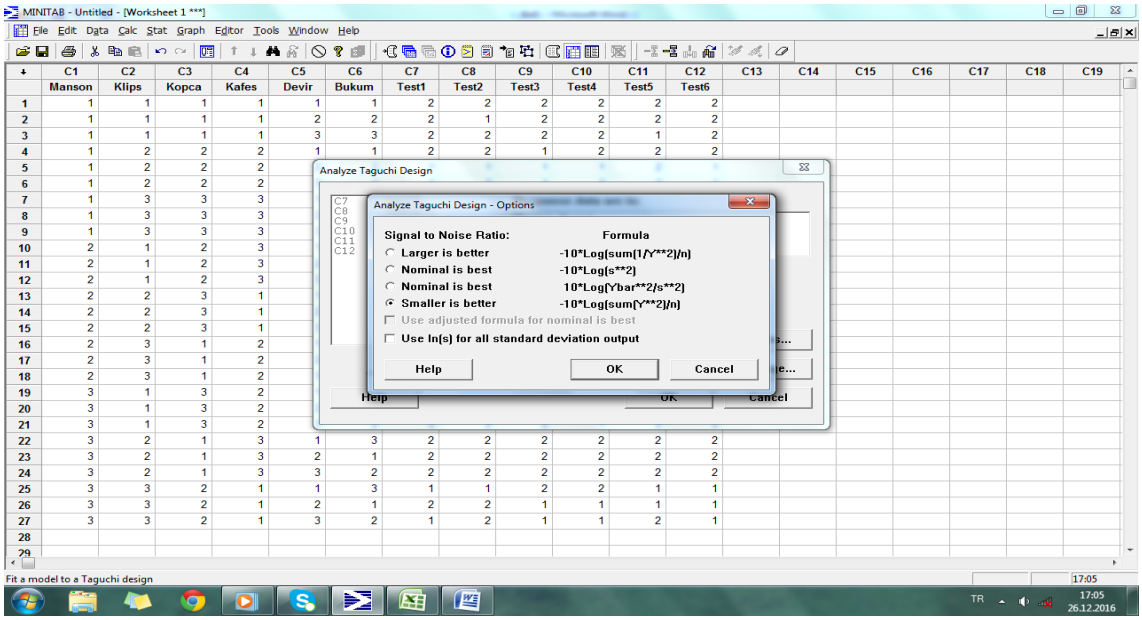
Şekil 4.14. Minitab14 paket programında sonuç tablolarının belirlenmesi.

Çıktıların, etki tabloları ve modelleri belirlenir.



Şekil 4.15. Minitab14 paket programında faktörlerin ikili etkilerinin belirlenmesi.

Faktörlerin birbiri ile etkileşim içinde olup olmadıkları belirlenir. Şekil 4.17.'deki sonuçlara göre, S/N oranları baz alındığında %95 güven düzeyinde Klips ile Devir arasında düşük ihtimali de olsa bir etkileşim vardır. Devir *p* değerinin çok yüksek çıkması ve önem düzeyinde son sırada yer alması sebebiyle bu etkileşim göz ardı edilmiştir.



Şekil 4.16. Minitab14 paket programında sinyal gürültü oranının belirlenmesi.

Gürültü oranı (S/N-Sinyal/Noise ratio) fonksiyonu seçimi yapılır.

Şekil 4.17.'de Taguchi tasarım Minitab14 sonuçları görülmektedir.

26.12.2016 16:58:07

Welcome to Minitab, press F1 for help.

Taguchi Design

Taguchi Orthogonal Array Design

L27(3**6)

Factors: 6
Runs: 27

Columns of L27(3**13) Array

1 2 3 4 5 6

Taguchi Analysis: Test1; Test2; Test3; ... versus Manson; Klips; Kopca; ...

The following terms cannot be estimated, and were removed.

Manson*Klips
Manson*Kopca
Manson*Kafes
Manson*Devir
Manson*Bukum
Klips*Kopca
Klips*Kafes
Klips*Bukum
Kopca*Kafes
Kopca*Devir
Kopca*Bukum
Kafes*Bukum
Devir*Bukum

Linear Model Analysis: SN ratios versus Manson; Klips; Kopca; Kafes; Devir; ...

Estimated Model Coefficients for SN ratios

Term	Coef	SE Coef	T	p
Constant	-4,67791	0,1195	-39,137	0,000
Manson 1	-0,38485	0,1690	-2,277	0,063
Manson 2	0,43287	0,1690	2,561	0,043

Klips 1	-0,58051	0,1690	-3,434	0,014
Klips 2	0,79090	0,1690	4,679	0,003
Kopca 1	-1,14939	0,1690	-6,800	0,000
Kopca 2	0,76735	0,1690	4,540	0,004
Kafes 1	1,15867	0,1690	6,854	0,000
Kafes 2	-0,22249	0,1690	-1,316	0,236
Devir 1	-0,35136	0,1690	-2,079	0,083
Devir 2	0,40173	0,1690	2,377	0,055
Bukum 1	-0,40604	0,1690	-2,402	0,053
Bukum 2	-0,04042	0,1690	-0,239	0,819
Klips*Devir 1 1	0,68604	0,2391	2,870	0,028
Klips*Devir 1 2	-0,55415	0,2391	-2,318	0,060
Klips*Devir 2 1	-0,90852	0,2391	-3,800	0,009
Klips*Devir 2 2	0,89144	0,2391	3,729	0,010
Kafes*Devir 1 1	-0,46617	0,2391	-1,950	0,099
Kafes*Devir 1 2	0,30052	0,2391	1,257	0,255
Kafes*Devir 2 1	0,10486	0,2391	0,439	0,676
Kafes*Devir 2 2	0,09127	0,2391	0,382	0,716

S = 0,6211 R^2 **= 96,6%** R^2**(adj) = 85,5%**

Analysis of Variance for SN ratios

Source	DF	Seq SS	Adj SS	Adj MS	F	p
Manson	2	3,040	3,040	1,5201	3,94	0,081
Klips	2	9,061	9,061	4,5305	11,74	**0,008**
Kopca	2	18,503	18,503	9,2514	23,98	**0,001**
Kafes	2	20,416	20,416	10,2080	26,46	**0,001**
Devir	2	2,586	2,586	1,2932	3,35	0,105
Bukum	2	3,293	3,293	1,6463	4,27	0,070
Klips*Devir	4	7,776	7,776	1,9440	5,04	0,040
Kafes*Devir	4	2,033	2,033	0,5084	1,32	0,362
Residual Error	6	2,314	2,314	0,3857		
Total	26	69,023				

Unusual Observations for SN ratios

Observation	SN ratios	Fit	SE Fit	Residual	St Resid
5	-1,761	-2,383	0,548	0,622	2,12 R

R denotes an observation with a large standardized residual.

Response Table for Signal to Noise Ratios
Smaller is better

Level	Manson	Klips	Kopca	Kafes	Devir	Bukum
1	-5,063	-5,258	-5,827	-3,519	-5,029	-5,084
2	-4,245	-3,887	-3,911	-4,900	-4,276	-4,718
3	-4,726	-4,888	-4,296	-5,614	-4,728	-4,231
Delta	0,818	1,371	1,917	2,095	0,753	0,853
Rank	**5**	**3**	**2**	**1**	**6**	**4**

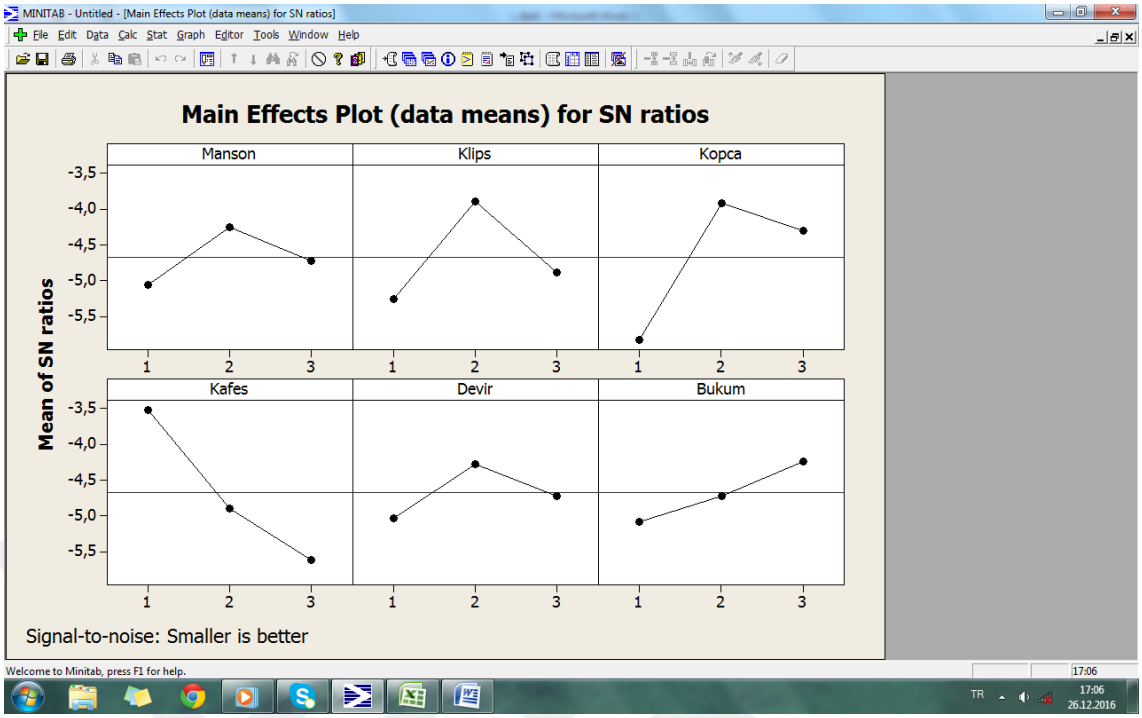
Şekil 4.17. Taguchi tasarım Minitab14 sonuçları.

İplik kalitesi için hesaplanan S/N oranı ve ortalamalar için varyans analizi sonuçları Şekil 4.17.'de sunulmaktadır. S/N oranları baz alındığında %95 güven düzeyinde Kafes, Kopça ve Klips faktörlerinin üçünün de sonuçlar üzerinde etkili faktörler olduğu (p<0.05) görülmektedir. Büküm, Manşon ve Devir değerleri ise daha az etkili faktörlerdir. Ayarlı R^2 değerinden modelin tahmin gücü %85.5 olarak bulunmuştur (Şekil 4.17.).

Tablo 4.3. Taguchi tasarım Minitab14 sonuçları özet tablosu.

Level	Manson	Klips	Kopca	Kafes	Devir	Bukum
1	**-5,063**	**-5,258**	**-5,827**	-3,519	**-5,029**	**-5,084**
2	-4,245	-3,887	-3,911	-4,900	-4,276	-4,718
3	-4,726	-4,888	-4,296	**-5,614**	-4,728	-4,231
Rank	**5**	**3**	**2**	**1**	**6**	**4**
p value	0,081	**0,008**	**0,001**	**0,001**	0,105	0,070

Tablo 4.3.'de Taguchi yöntemi uygulaması sonucu elde edilen faktör seviyeleri verilmektedir. S/N oranları için verilen yanıt ve ana etki grafiğinden (Tablo 4.3. ve Şekil 4.18.) de görüldüğü gibi en iyi faktör kombinasyonları Kafes-3, Kopça-1, Klips-1, Büküm-1, Manşon-1 ve Devir-1 olarak belirlenmiştir.



Şekil 4.18. S/N oranları için ana etkiler.

Taguchi deneysel tasarımını uygulamadaki amaç elimizde bulunan 6 faktörden hangilerinin daha etkili, hangilerinin az etkili olduğunu tespit etmek ve faktörlerin en iyi seviyelerini belirlemektir. Sonuç olarak 3 faktörün; Kafes, Kopça ve Klips etkisinin çok olduğu görülmüştür. Diğer 3 faktörün; Büküm, Manşon ve Devirin en etkili seviyeleri belirlenmiştir. Bu sayede seçtiğimiz 3 faktörün birbiri ile etkileşiminin tespit edilebilmesi için sadece $3^3 = 27$ test gereklidir. Böylece ana amaçlardan, zaman ve hatalı üründen kaynaklı para kaybı olmayacaktır. Etkili olan 3 faktörün tüm ihtimallerinin deneneceği, etkisiz olan faktörlerin en iyi seviyeleri ile test edileceği yeni veri kümesi Tablo 4.4.'de görülmektedir.

Tablo 4.4. Taguchi deney tasarımı sonucunda tekrar oluşturulan veri kümesi.

Kafes	Kopça	Klips	Büküm	Manşon	Devir	Test1	Test2	Test3	Test4	Test5	Test6
1	1	1	1	1	1	2	2	2	2	2	2
1	1	2	1	1	1	2	2	2	2	2	2
1	1	3	1	1	1	2	2	2	2	2	1
1	2	1	1	1	1	2	2	1	1	2	1
1	2	2	1	1	1	2	2	1	2	1	2
1	2	3	1	1	1	2	1	1	2	2	2
1	3	1	1	1	1	1	2	2	2	2	2
1	3	2	1	1	1	2	2	2	2	2	2
1	3	3	1	1	1	1	1	1	1	2	1
2	1	1	1	1	1	2	2	2	2	2	2
2	1	2	1	1	1	2	2	2	2	2	2
2	1	3	1	1	1	2	2	2	2	2	2
2	2	1	1	1	1	2	2	1	2	2	2
2	2	2	1	1	1	1	1	2	2	1	1
2	2	3	1	1	1	1	1	1	2	2	1
2	3	1	1	1	1	2	2	1	1	2	2
2	3	2	1	1	1	1	1	2	2	1	2
2	3	3	1	1	1	2	2	2	2	2	2
3	1	1	1	1	1	2	2	2	2	2	2
3	1	2	1	1	1	2	2	2	2	2	2
3	1	3	1	1	1	2	2	2	2	2	2
3	2	1	1	1	1	2	2	2	2	2	2
3	2	2	1	1	1	2	2	2	2	2	2
3	2	3	1	1	1	2	2	2	2	2	2
3	3	1	1	1	1	2	2	2	2	2	2
3	3	2	1	1	1	2	2	2	2	2	2
3	3	3	1	1	1	2	2	2	2	2	2

5. İPLİK KALİTESİ SEÇİMİNE ETKİ EDEN FAKTÖRLERDEN VERİ MADENCİLİĞİ SINIFLANDIRMA YÖNTEMİ İLE KURAL OLUŞTURULMASI

Sınıflandırma işlemine, sonuç olarak bir kural kümesine sahip olunabilmesi için gerek duyulmaktadır. Kural kümesi, işletme için esneklik açısından gereklidir. Elde olmayan ve iyi sonuç verdiği bilinen bir parçanın yerine, elde olan parça kullanılırsa makine nasıl tepki verecektir ya da diğer parametrelerden hangileri değiştirilirse sonuç beklenen değere ulaşmış olur? Örneğin işletmede 4.5 mm klips mevcut değilse yerine 3.5 mm klips takıldığında sonucun iyi olması için kafes genişliği ne olmalıdır? Bu bilgi işletme açısından çok önemlidir. Sonuç olarak bir kural kümesine sahip olunmasına, bu sebep dolayısı ile ihtiyaç vardır.

Taguchi deney tasarımı sonuçlarından oluşturulan yeni veri kümesinin, veri madenciliği paket programlarında işlenmesi ile nihai sonuç elde edilecektir. Paket program olarak; Weka 3.8.1 ve MT-VeMa 1.0 kullanılacaktır. Veri madenciliği paket programlarında sınıflandırma yaparken veriler, maliyete duyarlı ve maliyete duyarsız olarak işlenecektir.

Maliyete-duyarlı öğrenmenin, maliyete-duyarsız öğrenmeden en önemli farkı, yanlış sınıflandırma maliyetini dikkate alması ve beklenen yanlış sınıflandırma hataları sayısını en küçüklemek yerine, beklenen yanlış sınıflandırma maliyetini en küçüklemekle ilgilenmesidir. Maliyete-duyarsız durum için yanlış sınıflandırma maliyetleri arasında bir fark yokken, maliyete-duyarlı sınıflandırmada belli bir sınıfa ait bir örneği başka bir sınıfta tahmin etmenin maliyeti, tahmin edilen sınıfa göre değişkenlik göstermektedir. Diğer bir deyişle, maliyete-duyarsız sınıflandırma, maliyete-duyarlı sınıflandırmanın özel bir şeklidir [76].

Beklenen yanlış sınıflandırma hataları sayısını en küçüklemeyi amaçlayan maliyete-duyarsız sınıflandırıcılar, maliyete-duyarlı sınıflandırma problemlerini çözmekte yetersiz kalmaktadır. Bunun temel sebebi, basit bir 0/1 maliyet fonksiyonunu en küçükleyen geleneksel sınıflandırıcıların daha karmaşık bir maliyet matrisinin kayıp fonksiyonunun alanını değiştirmesi sebebiyle kayıp fonksiyonunun en küçükleneceğini garanti edememesidir. Maliyete-duyarlı öğrenme genellikle birçok gerçek hayat uygulamasında karşılaşılan dengesiz sınıf dağılımına sahip veri kümeleri üzerinde kullanılır. Bu tür bir veri kümesi maliyete-duyarsız sınıflandırıcılar kullanılarak sınıflandırıldığında, iki sınıfın söz konusu olduğu varsayımıyla, neredeyse bütün örnekler negatif olarak sınıflandırılacağından dolayı dengesizlik oranı yüksek veri kümelerinde büyük bir problem teşkil eder. İkili-sınıflandırma açısından ele alındığında daha az sayıda yer alan sınıf pozitif, diğer sınıf ise negatif sınıf olarak nitelendirilir [76].

Provost [77], tarafından da belirtildiği gibi, maliyete-duyarsız sınıflandırıcılarda iki temel varsayım yapılmaktadır: amacın doğruluğu en büyüklemek (ya da hata oranını en küçüklemek) olması ve eğitim ve test veri kümelerinin sınıf dağılımının aynı olması. Ancak bu iki varsayım altında yüksek derecede dengesiz bir veri kümesi için neredeyse bütün örnekler negatif olarak sınıflandırılacaktır. İşte maliyete-duyarlı sınıflandırma bu tür veri kümelerinin sınıflandırılmasında kullanılabilecek etkin bir yöntemdir.

Elkan [78] ve Zadrozny ve Elkan [79] çalışmalarında ikili-sınıflandırma için maliyete duyarlı öğrenmenin genel yapısını sunmuş ve yanlış sınıflandırma maliyetinin farklı maliyete duyarlı öğrenme algoritmaları üzerindeki rolünü detaylı olarak açıklamışlardır. Pozitif ve negatif olarak adlandırılan iki sınıf içeren ikili-sınıflandırmaya dayalı örnek bir maliyet matrisi Tablo 5.1.'de verilmiştir.

Tablo 5.1. İkili-sınıflandırma için maliyet matrisi örneği.

	Tahminlenen negatif	**Tahminlenen pozitif**
Gerçek negatif	C(0,0) TN	C(1,0) FP
Gerçek pozitif	C(0,1) FN	C(1,1) TP

Maliyete-duyarlı öğrenmede yanlış pozitif (FP - gerçekte negatif ancak pozitif olarak tahmin edilmiş), yanlış negatif (FN – gerçekte pozitif ancak negatif olarak tahmin edilmiş), doğru pozitif (TP – gerçekte pozitif ve pozitif olarak tahmin edilmiş) ve doğru negatif (TN – gerçekte negatif ve negatif olarak tahmin edilmiş) maliyetleri, maliyet matrisi içinde verilmiştir. Tabloda, *C(i,j)* aslında sınıfı *i* olan bir örneği *j* sınıfı olarak tahmin etmenin maliyetini; *C(i,i)* değerleri ise fayda olarak nitelendirilmekte olup doğru tahmin edilen örnekleri temsil etmektedir. Verilen örnekte 1 pozitif sınıfı, 0 ise negatif sınıfı göstermektedir. Maliyete-duyarlı öğrenmede bu tür bir maliyet matrisinin bilindiği varsayılır ve çoklu sınıflandırmaya ait bir maliyet matrisi için satır ve sütunlar kolaylıkla genişletilebilir. Verilen bu maliyet matrisine bağlı olarak eldeki örnekler minimum beklenen maliyet değerlerine göre sınıflandırılacaktır. *x* örneğini *j* sınıfı olarak sınıflandırmanın beklenen maliyeti aşağıdaki gibi tanımlanmaktadır.

$$R(^{j}/_{x}) = \sum_i P(^{i}/_{x}) C(i,j) \qquad (5.1.)$$

Verilen eşitlikte *P(i/x)*, *x* örneğini *i* sınıfı olarak sınıflandırmanın olasılık tahminini göstermektedir. *x* örneğinin pozitif olarak sınıflandırılması aşağıdaki şartın sağlanmasına bağlıdır.

$$P(0/x)C(1,0) + P(1/x)C(1,1) \leq P(0/x)C(0,0) + P(1/x)C(0,1) \qquad (5.2.)$$

veya,

$$P(0/x)(C(1,0) - C(0,0)) \leq P(1/x)(C(0,1) - C(1,1)) \qquad (5.3.)$$

Verilen maliyet matrisi üzerinde ilk sütundan *C(0,0)* değeri ve ikinci sütundan *C(1,1)* değeri çıkarıldığında daha basit bir maliyet matrisi elde edilecektir (Tablo 5.2.).

Tablo 5.2. İkili-sınıflandırma için basitleştirilmiş maliyet matrisi.

	Tahminlenen negatif	**Tahminlenen pozitif**
Gerçek negatif	0	C(1,0) – C(0,0)
Gerçek pozitif	C(0,1) – C(1,1)	0

Netice olarak verilen herhangi bir maliyet matrisi *C(0,0)=C(1,1)*=0 olan başka bir matrise dönüştürülebilir. *C(0,0)=C(1,1)*=0 olduğu varsayımı altında *x* örneğinin pozitif olarak sınıflandırılma şartı da Eşitlik 5.4.'deki hale dönüşecektir.

$$P(0/x)C(1,0) \leq P(1/x)C(0,1) \tag{5.4.}$$

Diğer taraftan *P(0/x)=1-P(1/x)* olduğundan *x* örneğinin pozitif olarak sınıflandırılabilmesi için ilgili olasılık tahmini değerinin belli bir eşik değerini geçmesi gerekmektedir (p^*).

$$P\,(1/x) \geq \frac{C(1,0)}{C(0,1)+C(1,0)} = \frac{FP}{FP+FN} = p^* \tag{5.5.}$$

Dolayısıyla maliyete-duyarsız bir sınıflandırıcı için *P(1/x)* değeri üretilebilirse bu sınıflandırıcı maliyete-duyarlı hale dönüştürülebilmekte ve literatürde geliştirilmiş birçok maliyete-duyarlı öğrenme algoritması da bu mantığa dayanmaktadır.

Maliyete-duyarlı sınıflandırmada en yaygın pratik yaklaşım, öğrenme algoritmasının kararların maliyetini en küçükleyen çıktıyı elde edebilmesi için eğitim verisinin sınıf dağılımını değiştirmektir. İki-sınıflı problemler için bunu yapmanın en basit ve en yaygın kullanılan yolu, eğitim kümesindeki örneklerin sınıf oranlarının maliyet değerlerine orantılı olarak değiştirilmesine dayanır. Tablo 5.3.'de gerçekleştirilen, çalışmaya ait maliyet matrisi örnek alınırsa,

Tablo 5.3. Yapılan çalışmaya ait maliyet matrisi.

	Tahminlenen negatif	**Tahminlenen pozitif**
Gerçek negatif	0	2200
Gerçek pozitif	500	0

pozitif sınıfa ait örnekler geri kalan örneklere göre 4.4 (=2200/500) kat daha fazla olacaktır. Bu yönteme tabakalaşma ya da tekrar dengeleme adı verilmekte olup genellikle daha az pahalı sınıftan seyrek örnekleme ya da daha pahalı sınıftan fazla örnekleme yaparken kullanılmaktadır [80].

Tabakalaşmanın avantajı herhangi bir iki-sınıflı probleme kolaylıkla uygulanabilir olmasıdır. Ancak bu avantaj iki soruna neden olmaktadır: (1) orjinal formülasyonda tabakalaşma sadece iki-sınıflı problemler üzerinde çalışabilmekte ve (2) tabakalaşma öğrenme algoritması tarafından hipotezi kurmak için kullanılan içsel mekanizmayı dikkate almadığından verinin orijinal dağılımına zarar vermektedir.

Tablo 5.3.'de ki veriler şu şekilde belirlenmiştir. 1 Ring İplik makinesi 1 takımda 100 kg iplik üretmektedir. Üretilen ipliğin satış fiyatı 5 $/kg'dır. Üretilen 100 kg iplik ile 666 metre kumaş üretilmektedir. Kumaşın satış fiyatı 3.30 $/metre'dir. Sonuç olarak eğer gerçek sonuç pozitif iken, negatif değerlendirilirse 100 kg iplik hurdaya ayrılmaktadır ve 500 $ maliyet kaybına sebebiyet vermektedir. Eğer gerçek sonuç negatif iken pozitif kabul edilirse, 666 metre kumaş hurdaya ayrılmaktadır. Bu da 2200 $ maliyet kaybına sebebiyet vermektedir.

5.1. Veri Madenciliği Paket Programları

Çalışma sonuçları, Weka 3.8.1 ve MT-VeMa 1.0 olmak üzere 2 farklı paket programda incelenecektir.

5.1.1. Weka 3.8.1 Paket Programı

WEKA (Waikato Environment for Knowledge Analyses), Waikato Üniversitesi tarafından geliştirilerek 1996'da ilk resmi sürümü yayınlanmış olan bir makine öğrenme ve veri madenciliği yazılımıdır. Akademik araştırmalar, eğitim ve endüstriyel uygulama alanlarında kullanım yeri olan WEKA, veri analizi ve tahminleyici modelleme için geliştirilmiş algoritma ve araçların görsel bir birleşimini içerir. Geliştirilen yazılımın temel avantajları geniş veri önişleme ve modelleme tekniklerine sahip olması, grafiksel kullanıcı ara yüzü sayesinde kullanımının kolay olması ve Java programlama dili ile uygulandığından herhangi bir platformda kullanılabilmesi yani taşınabilir olmasıdır [81].

5.1.2. MT-VeMa 1.0 Paket Programı

MT-VeMa 1.0 paket programı; BEE-miner, MEPAR-miner ve DIFACONN-miner algoritmalarının ortak bir platformda kullanımına imkan verir. Paket program "Veri Önişleme", "Sınıflandırma" ve "Raporlama" olmak üzere üç bölümden oluşmaktadır. Veri Önişleme bölümünde; kesiklendirme yöntemleri, veri kümesi ayrımı test-eğitim olarak ve kayıp değerleri doldurma yöntemlerinin tamamı yer almaktadır. Sınıflandırma bölümünde sınıflandırma yapabilmek amacıyla maliyete-duyarlı ve maliyete-duyarsız sınıflandırıcılar, Raporlama bölümünde ise elde edilen sonuçların kullanıcıya sunulduğu özet bilgiler yer almaktadır.

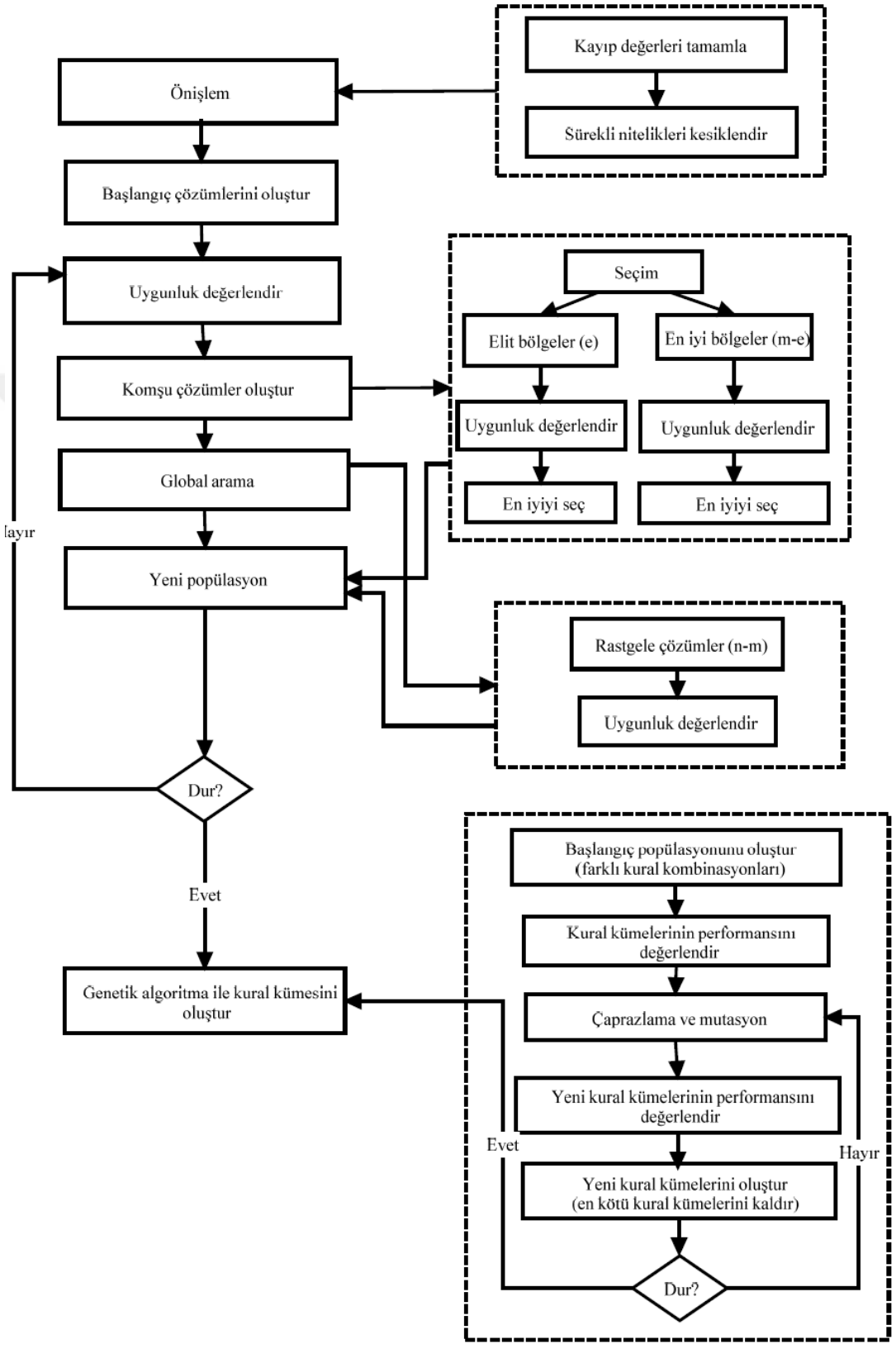
BEE-miner sınıflandırma algoritması

BEE-miner algoritması iki/çok sınıflı yapılara uygulanabilen ve Arı Algoritmasına [60] dayanan maliyete duyarsız/duyarlı bir sınıflandırıcı özelliği taşımaktadır. Geliştirilen algoritma doğrudan bir yöntem olma özelliği taşıdığından girdi ya da çıktı üzerinde yapılacak güncellemeler yerine, maliyet bileşeni algoritmanın çalışma prensibi içine dâhil edilmiştir. Maliyet bileşeni içeren bu aşamalar uygunluk fonksiyonunun ve komşuluk yapılarının belirlenmesidir [76].

BEE-miner algoritması, sürekli niteliklerin kesiklendirildiği ve eksik verilerin tamamlandığı önişlem aşamasıyla başlar. Sonrasında başlangıç çözümleri oluşturulur ve bu çözümlerin uygunluk fonksiyonu değerleri değerlendirilir.

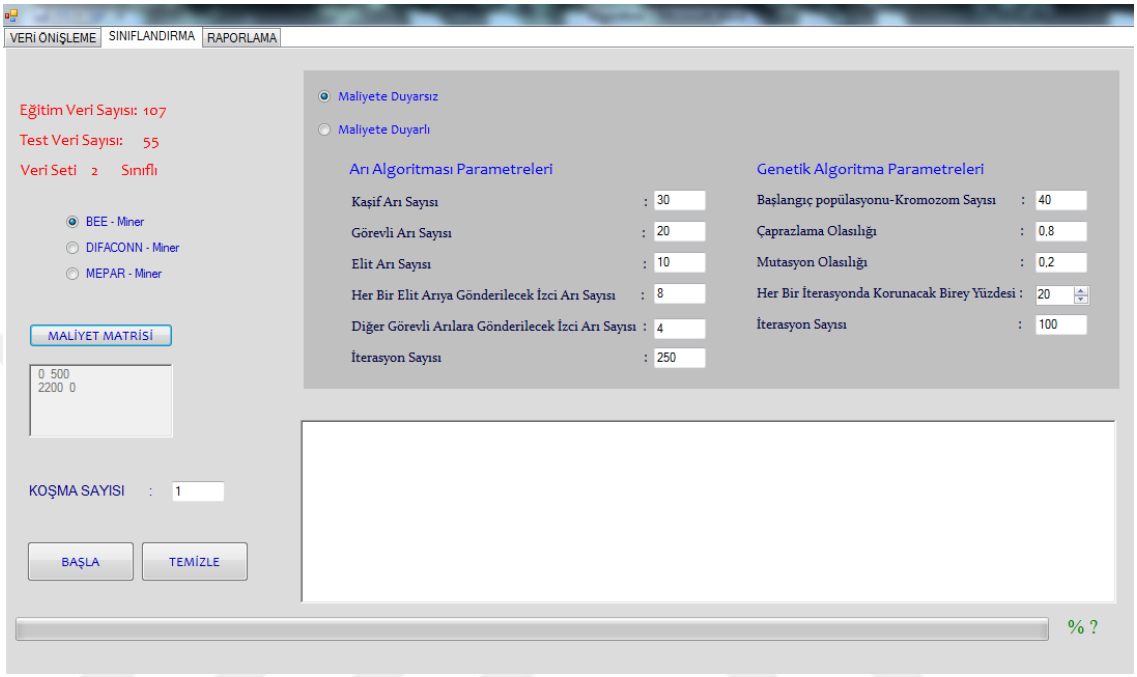
Arı Algoritması ile yeni popülasyon oluşturulur. Doğadaki arılarda, yiyecek arama işlemini önce kâşif arılar başlatır ve rastgele yiyecek kaynakları seçerler. Ayrıca yiyecek toplama süreci boyunca popülasyonun belli bir yüzdesi kâşif arılar olarak kalır. Benzer olarak Arı Algoritmalarında *n* adet kâşif arının araştırma uzayına rastgele yerleştirilmesi ile başlar. Kâşif arılarca ziyaret edilen noktaların uygunlukları değerlendirilir. Doğadaki arılara bakıldığında, bir arı kolonisi çok sayıda yiyecek kaynağını tüketmek için birçok yönde ve büyük uzaklıklar boyunca arama yapabilmektedir. Prensipte nektar miktarı fazla olan ve düşük bir çabayla tüketilebilen yiyecek kaynakları daha fazla arı tarafından, az miktarda nektara sahip yiyecek kaynakları ise daha az arı tarafından ziyaret edilecektir. Bu gerçekten hareketle en iyi uygunluk değerine sahip arılar (*m*) komşuluk araması için seçilir. Seçilen arıların komşuluğunda araştırma başlar ve daha umut verici çözümleri temsil eden en iyi *e* bölgeye, seçilen diğer bölgelere (*m*-*e*) göre daha fazla arı gönderilerek daha detaylı arama yapılır. Yeni popülasyonun oluşturulması için her bölgedeki en iyi uygunluk değerine sahip arı seçilir. Popülasyondaki diğer arılar (*n*-*m*) yeni potansiyel çözümler elde etmek için rastgele olarak araştırma uzayına atanırlar. Böylece her iterasyonun sonunda yeni popülasyon, seçilen her bölgenin temsilcileri ve rastgele arama yapan kâşif arılar olmak üzere iki parçadan oluşacaktır. Diğer taraftan Arı Algoritması kâşif arı sayısı ile temsil edilen popülasyon boyutu (*n*), ziyaret edilen *n* bölge içinden seçilen bölge sayısı (*m*), seçilen *m* bölge içindeki en iyi bölge sayısı (*e*), en iyi *e* bölgeye gönderilen izci arı sayısı (*nep*), seçilen diğer bölgelere (*m*-*e*) gönderilen izci arı sayısı (*nsp*), komşuluk arama boyutu (*ngh*) ve durdurma kriteri olmak üzere birçok parametre içermektedir.

Son olarak genetik algoritma kullanılarak kural kümesi belirlenir. BEE-miner algoritmasına ait akış diyagramı Şekil 5.1.'de sunulmuştur [76].



Şekil 5.1. BEE-miner algoritması akış diyagramı [76].

BEE-miner algoritmasında kullanılacak parametrelerin değerleri ve MT-VeMa programındaki görünümü Şekil 5.2.'deki gibidir.

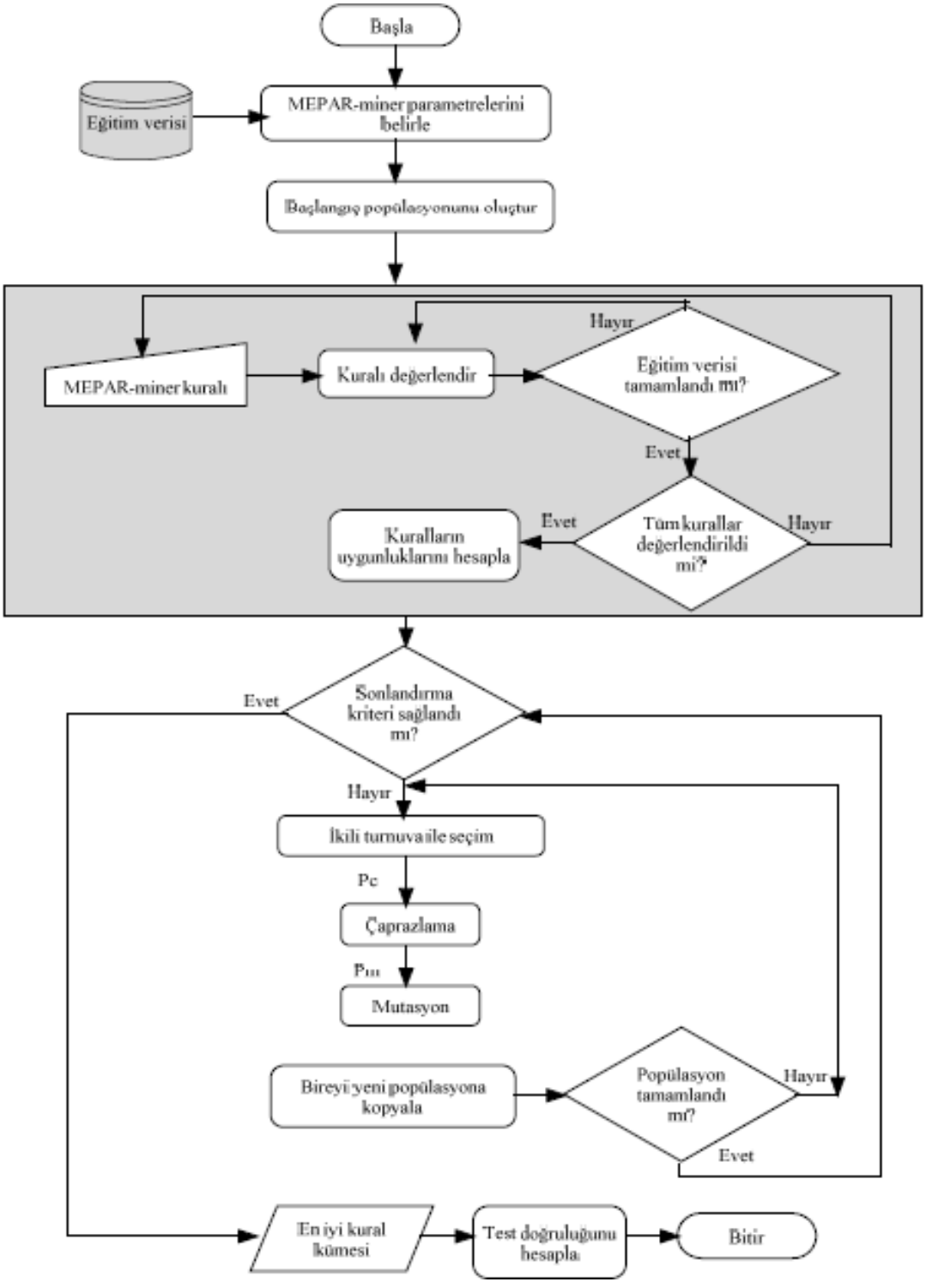


Şekil 5.2. BEE-miner algoritmasında kullanılacak parametrelerin MT-VeMa programındaki görünümü.

MEPAR-miner algoritması

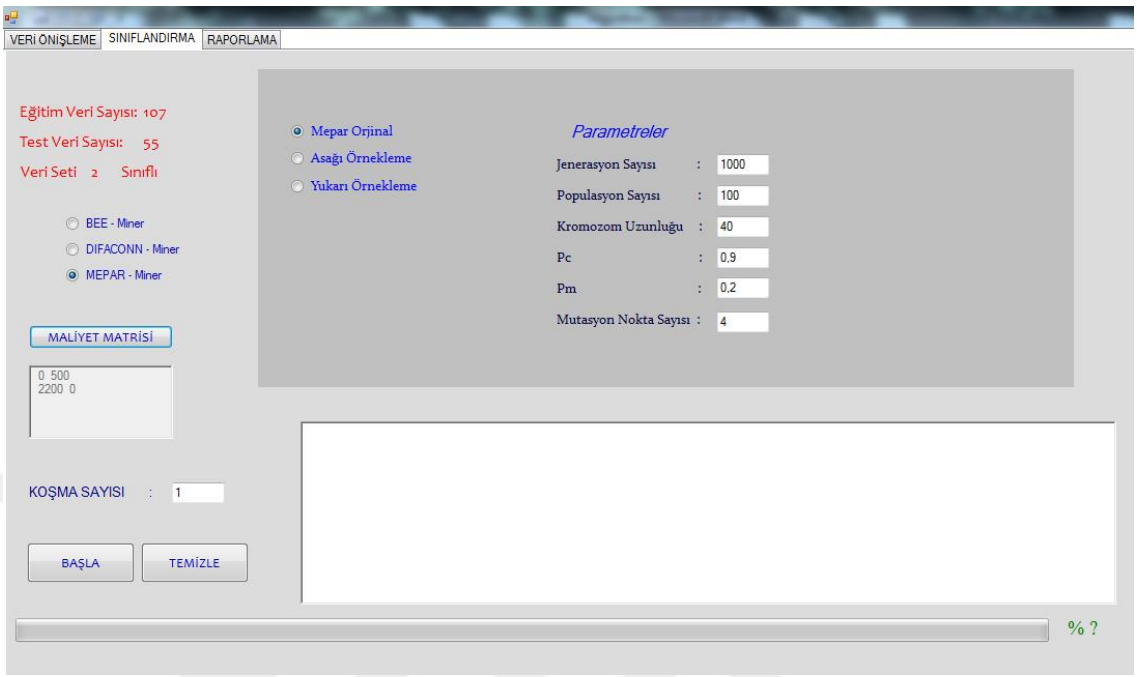
MEPAR-miner algoritması, sembolik regresyona yönelik geliştirilen çoklu denklem programlama (ÇDP) yaklaşımının, sınıflandırma kuralları türetmek üzere geliştirilmesi ile ortaya konmuştur. Çoklu denklem programlama, Genetik Programlamanın (GP) kodlanması, ağaç yapısından dolayı uygun olmayan bireylerin ortaya çıkması, ağaç derinliğinin genetik operatörler ile aşırı artması ve buna bağlı olarak çözüm süresinin yüksek olması gibi dezavantajlarının üstesinden gelmek için geliştirilmiş, doğrusal gösterime sahip değişken uzunlukta bireyler türeten bir algoritmadır. Etkin bir kromozom gösterimi ile genetik programlamanın avantajlarını içerisinde barındıran ve değişken uzunlukta sınıflandırma kuralları türeten MEPAR-miner yaklaşımı, genetik operatörlerle uygun olmayan bireylerin türetilmesi, kural boyutunun aşırı büyümesi gibi dezavantajları da ortadan kaldırmıştır [82].

MEPAR-miner algoritmasında, Oltean ve Dumitrescu [83] tarafından ortaya konulan çoklu denklem programlama (ÇDP) yaklaşımı, sınıflandırma kuralı türetmek üzere değiştirilmiş ve yazılımı gerçekleştirilmiştir. Çoklu denklem programlama algoritması, rastgele seçilmiş bireylerden oluşan bir popülasyonla başlar. Mevcut popülasyondaki her birey, probleme bağlı olarak belirlenen uygunluk fonksiyonuna göre değerlendirilir. Belirli sayıda seçilen en iyi bireyler, bir sonraki jenerasyona geçirilir. Eşleşme havuzu ikili turnuva seçimi ile doldurulur. Daha sonra bu havuzdaki bireyler rastgele eşleştirilerek çaprazlanır. İki ebeveynin çaprazlanması sonucu iki yeni birey elde edilir. Elde edilen bireyler mutasyona uğrar ve bir sonraki jenerasyona girerler. Algoritma belirli sayıda jenerasyon için tekrarlanır [82]. MEPAR-miner algoritmasının genel işleyişi Şekil 5.3.'de gösterilmektedir.



Şekil 5.3. MEPAR-miner algoritması genel akışı [82].

MEPAR-miner algoritmasında kullanılacak parametrelerin değerleri ve MT-VeMa programındaki görünümü Şekil 5.4.'deki gibidir.



Şekil 5.4. MEPAR-miner algoritmasında kullanılacak parametrelerin MT-VeMa programındaki görünümü.

DIFACONN-Miner algoritması

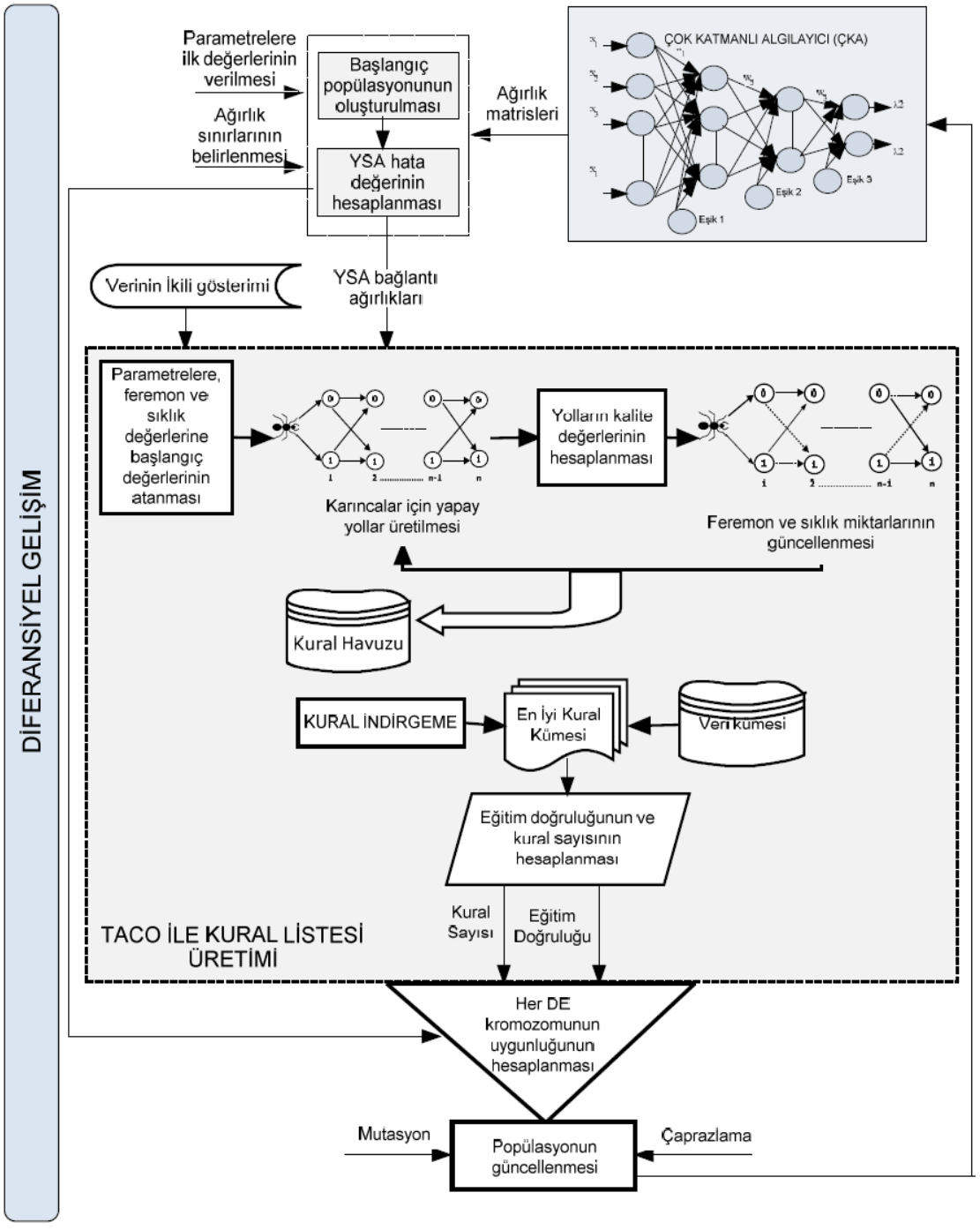
Yapay sinir ağları (YSA) genelleme, tahminleme, sınıflandırma gibi kabiliyetlerinden dolayı veri madenciliği alanında yaygın olarak kullanılmaktadır. Yapay sinir ağları oldukça yüksek sınıflandırma doğrulukları elde etmelerine rağmen elde ettikleri bilgileri kullanıcıların anlayabileceği kurallar şeklinde sunamamaktadırlar ve tanımlama kabiliyetleri oldukça sınırlıdır. Bu nedenle günümüzde veri madenciliğinde yapay sinir ağlarının kullanımında eğilim, yapay sinir ağlarından kural çıkarımına yöneliktir. DIFACONN-miner algoritması da bu eğilimden yola çıkarak, ileri beslemeli yapay sinir ağlarından sınıflandırma kuralları çıkarımına yönelik Özbakır vd. tarafından [84] yılında geliştirilmiş yeni bir sınıflandırma algoritmasıdır. Literatürde yapay sinir ağlarından kural çıkarımına yönelik pek çok çalışma mevcuttur ancak ele alınan DIFACONN-miner bu algoritmalardan oldukça farklıdır. Geçmiş çalışmalarda YSA eğitimi ve kural çıkarımı birbirinden bağımsız olarak gerçekleştirilmesine rağmen, sunulan algoritmada sınıflandırma kuralları çıkarımı için yapay sinir ağları eğitimi ve kural çıkarımı çoklu-amaç değerlendirme yapısı ile eşzamanlı olarak gerçekleştirilmektedir. DIFACONN-

miner yapay sinir ağlarının eğitimi için diferansiyel gelişim (DE) algoritmasını, kural çıkarımı için ise tur atan karınca koloni optimizasyonu (TACO) algoritmasını kullanmaktadır. DE algoritmasının her eğitim adımında YSA-ağırlıkları, TACO algoritmasına kural üretimi için gönderilmektedir. Daha sonra YSA yapısının uygunluğu YSA hatası, kural sayısı ve eğitim doğruluğu olmak üzere üç performans ölçütü içeren çoklu amaç fonksiyonuna göre değerlendirilmektedir [82].

DIFACONN-miner algoritması bir kural çıkarım algoritması olmaktan ziyade, aslında bir kural üretim algoritmasıdır ve birbirine bağlı üç bölümden oluşur. Bunlar;

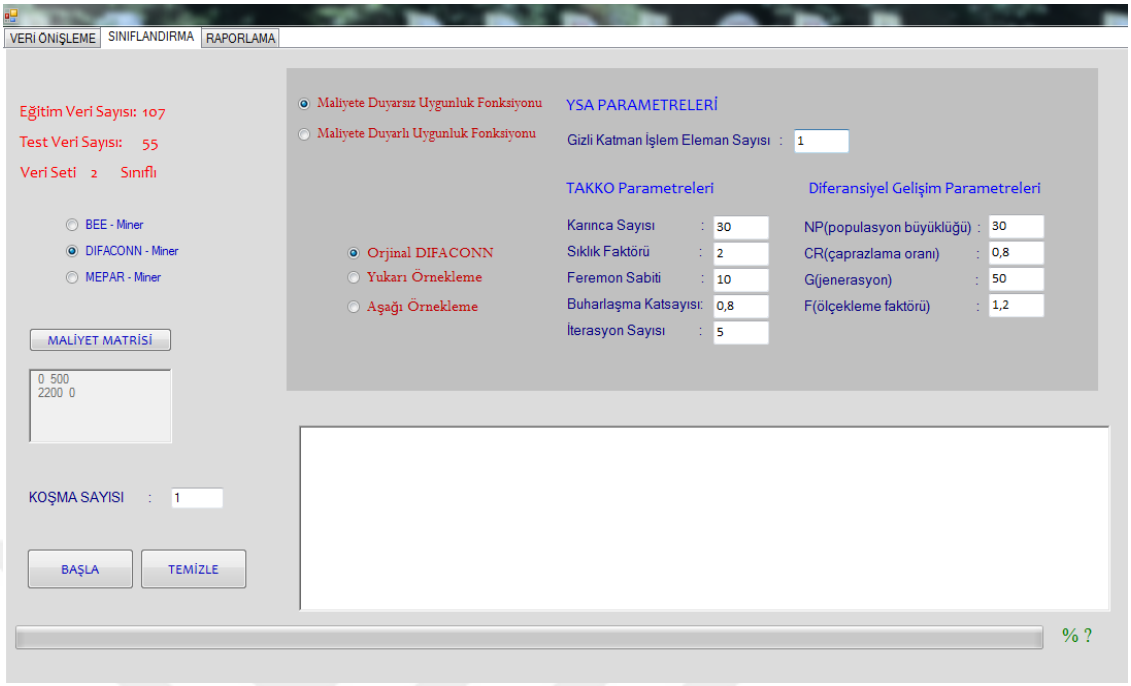
- DE algoritması ile ileri beslemeli YSA'ların eğitimi
- TACO algoritması ile kural çıkarımı
- Uygunluk değerlendirme

DIFACONN-miner algoritmasının ana çatısını, DE tabanlı eğitim algoritması oluşturmaktadır ve TACO algoritması, ilgili ağırlık grupları için kural kümeleri üretmek amacıyla DE algoritmasına dahil edilmektedir. İkili kodlanmış veri kullanıcı girdileri, ağırlık grupları ve kural kümeleri ise DIFACONN-miner algoritması tarafından hesaplanan değerlerdir [82]. DIFACONN-miner algoritmasının genel işleyişi Şekil 5.5.'de gösterilmektedir.



Şekil 5.5. DIFACONN-Miner algoritması genel akışı [82].

DIFACONN-Miner algoritmasında kullanılacak parametrelerin değerleri ve MT-VeMa programındaki görünümü Şekil 5.6.'daki gibidir.



Şekil 5.6. DIFACONN-Miner algoritmasında kullanılacak parametrelerin MT-VeMa programındaki görünümü.

5.2. Sonuçların Değerlendirilmesi

Maliyete duyarlı olmayan sınıflandırma için WEKA 3.8.1 veri madenciliği yazılımındaki, literatürde en çok kullanılan 6 maliyete-duyarsız sınıflandırıcı değerlendirmeye alınmıştır. Bu algoritmalar PART [85], C4.5 [86], SimpleCART [87], RIPPER [88], NBTree [89] ve DecisionTable'dır [90]. Bu 6 algoritma ile birlikte BEE-miner [76], MEPAR-miner [82], ve DIFACONN-miner [82]'ın maliyete duyarlı olmayan algoritmaları hep birlikte karşılaştırılacaktır ve 9 algoritma içinde en yüksek doğruluğa sahip olan algoritma bulunacaktır.

Maliyete duyarlı sınıflandırmada karşılaştırmaları adilane yapabilmek ve aynı eğitim-test kümeleri, aynı maliyet matrisleri ile çalışabilmek için BEE-miner, MEPAR-miner ve DIFACONN-miner algoritmaları WEKA 3.8.1 veri madenciliği yazılımındaki iki popüler maliyete-duyarlı meta-öğrenme algoritması olan "CostSensitiveClassifier" ve "MetaCost" ile karşılaştırılmıştır. CostSensitiveClassifier ve MetaCost algoritmaları, temel maliyete-duyarsız sınıflandırıcılarını maliyete-duyarlı hale getirirler. CostSensitiveClassifier'da maliyet duyarlılığını sunmak için iki yöntemden biri

kullanılabilir; her bir sınıfa atanan toplam maliyete göre eğitim örneklerini yeniden ağırlıklandırmak veya minimum beklenen yanlış sınıflandırma maliyetleriyle sınıfı tahminlemek. WEKA yazılımındaki MetaCost ise Domingos [91] tarafından ortaya atılan MetaCost algoritmasını kullanmaktadır.

Diğer taraftan eğitim kümesi, algoritmalara iyi öğrenme kabiliyeti kazandırmak için kullanılırken, test kümesi algoritmaların performansını değerlendirmek için kullanılan veri kümesidir. Veri kümelerinin eğitim ve test kümelerine ayrılmasında yüzde bölme (percentage split) kullanılmıştır. Veri kümesinin % 66'sı eğitim için kalanı ise test için seçilmiştir.

BEE-miner, MEPAR-miner ve DIFACONN-miner algoritmalarının parametrelerinde, algoritma makalelerindeki orijinal parametreler kullanılacaktır. Bu algoritmaların 30 koşma sonuçları değerlendirilecektir. 30 koşma değerlerinin ortalaması (ort), minumum değeri (min), maksimum değeri (mak) ve standart sapması (ss)'da Tablo 5.4. ve 5.5.'de sırasıyla maliyete duyarsız ve maliyete duyarlı sınıflandırma için belirtilmiştir.

Tablo 5.4. ve 5.5.'de karşılaştırmaya alınan algoritmaların, test doğrulukları (TD), kural sayıları (KS), yanlış sınıflandırma maliyetleri (YSM) ve hesapsal zamanı (HZ) ile gösterilmektedir.

Çalışmada paket programların çalıştırıldığı bilgisayar; İntel Core 2 Duo 800 Mhz. 2.0 Ghz. 2 mb Cache Bellek T6400 işlemciye ve Windows 7 Ultimate Service Pack1 32 bit işletim sistemine sahiptir.

Tablo 5.4. Maliyete-duyarlı olmayan sınıflandırma sonuçları.

ALGORİTMA		TD	KS	HZ(sn)
BEE-Miner 30 koşma	ort.	78.18	2.03	9.03
	s.s	0	0.18	0.49
	min.	78.18	2	8
	mak.	78.18	3	10
MEPAR-Miner 30 koşma	ort.	54	3	112.6
	s.s	15	0	4
	min.	36	3	109
	mak.	78	3	108
DIFACONN-Miner 30 koşma	ort.	72	6.87	637.3
	s.s.	5	1.53	13.47
	min.	56	8	649
	mak.	78	4	655
PART		**83.63**	8	2
Ripper		**83.63**	3	2
DecisionTable		**83.63**	**1**	3
C4.5 (J48)		**83.63**	**1**	**1**
SimpleCART		**83.63**	12	6
NBTree		**83.63**	**1**	7

Tablo 5.4.'de görüldüğü gibi, test doğruluğu açısından Weka algoritmaları %83.63 değeri ile en yüksek doğruluğa ulaşmıştır. Diğer algoritmaların maksimum değerlerine baktığımızda, Bee-miner %78.18, Mepar-miner ve Difaconn-miner %78 doğrulukla Weka algoritmalarına yakın değer elde etmişlerdir. Ortalama değer açısından, Bee-miner %78.18 ile aynı kalmasına karşın, Mepar-miner %54 ve Difaconn-miner %72 değerlerine düşmüştür. Kural sayısı açısından bakıldığından Weka algoritmalarından Decision Table, C4.5 ve NBTree bir kural ile hem en yüksek doğruluk hem de en az kural oluşturmuşlardır. Bee-miner iki, Mepar-miner ve Ripper üç, Difaconn-miner dört kuralda kalmışlardır. Part 8 kural ile ve SimpleCART 12 kural ile yüksek seviyede kalmıştır. Hesapsal zaman açısından en iyi değere bir sn ile tüm değerleri en iyi olan C4.5 algoritması ulaşmıştır. Weka algoritmaları süre açısından birbirine yakındır. Difaconn-miner ikili yapısından dolayı en iyi doğruluğu 655 sn'de bulabilmiştir. Mepar-miner 108 sn'de bulabilmiştir. Weka algoritmalarına en yakın değeri 10 sn ile Bee-miner algoritması vermiştir.

Tablo 5.5. Maliyete-duyarlı sınıflandırma sonuçları.

ALGORİTMA		YSM	TD	KS	HZ(sn)
BEE-miner 30 koşma	ort.	366.18	63.03	4.37	13.97
	s.s	88.28	13.37	0.67	1
	min.	256.36	67.27	4	14
	mak.	550.91	45.45	4	15
MEPAR-miner Aşağı Örnekleme 30 koşma	ort.	372.36	58	3	68.3
	s.s	102.19	9	0	2.44
	min.	256.36	67	3	66
	mak.	474.55	55	3	69
MEPAR-miner Yukarı Örnekleme 30 koşma	ort.	293.45	53	3	131.73
	s.s	32.99	13	0	4.86
	min.	256.36	67	3	124
	mak.	321.82	42	3	134
DIFACONN-miner Aşağı Örnekleme 30 koşma	ort.	377.76	48	5.13	88.2
	s.s	97.32	19	0.97	1.54
	min.	256.36	67	4	87
	mak.	620	25	5	89
DIFACONN-miner Yukarı Örnekleme 30 koşma	ort.	336.36	67	5.83	14179.87
	s.s	95.74	6	1.23	263.31
	min.	256,36	67	5	13627
	mak.	572.73	47	6	14311
MetaCost(PART)		180	**76.36**	7	14
MetaCost(Ripper)		220	74.54	5	15
MetaCost(DecisionTable)		321.81	41.81	**1**	11
MetaCost(C4.5)		189.09	74.54	9	8
MetaCost(SimpleCART)		229.09	72.72	5	42
MetaCost(NBTree)		267.27	52.72	**1**	49
CostSensitiveClassifier (PART)		**176.36**	70.90	6	3
CostSensitiveClassifier (Ripper)		207.27	70.90	3	3
CostSensitiveClassifier (DecisionTable)		207.27	70.90	4	3
CostSensitiveClassifier (C4.5)		185.45	69.09	8	**1**
Cost SensitiveClassifier (SimpleCART)		220	74.54	9	6
Cost SensitiveClassifier (NBTree)		207.27	70.90	**1**	7

Tablo 5.5.'de görüldüğü üzere tüm algoritmaların maliyete duyarlı kalabilmek adına test doğruluklarında bir düşüş olmuştur. Yanlış sınıflandırma maliyetleri dikkate alındığında en düşük yanlış sınıflandırma maliyetini 176.36 değeri ile CostSensitiveClassifier (PART) vermektedir. Bee-miner, Mepar-miner ve Difaconn-miner 256.36 değeri ile aynı sonucu vermektedir. Weka algoritmalarından MetaCost(DecisionTable) 321.81 değeri ve MetaCost(NBTree) 267.27 değeri ile bu algoritmalardan yüksek değer almıştır. Diğer algoritmalar daha düşük değer almışlardır. Test doğruluğu açısından en iyi değeri %76.36 ile MetaCost(PART) vermiştir. En iyi yanlış sınıflandırma maliyetine sahip olan CostSensitiveClassifier(PART) ise %70.9 doğrulukda kalmıştır. Bee-miner, Mepar-miner ve Difaconn-miner algoritmaları ise %67 doğrulukda kalmıştır. Bu sonuç ile Weka algoritmalarından yalnızca iki tanesinden daha iyi cevap vermişlerdir. Kural sayıları açısından bakıldığında 3 adet Weka algoritması bir kural sayısı ile en düşük yanlış sınıflandırma maliyetlerine ulaşmışlardır. Fakat en iyi yanlış sınıflandırma maliyetine sahip olan CostSensitiveClassifier(PART)'dan daha yüksek maliyette kalmışlardır. Bee-miner 4 kural, Mepar-miner 3 kural ve Difaconn-miner 4 kural ile en düşük yanlış sınıflandırma maliyetine sahip olmuşlardır. Hesapsal zaman açısından yine en iyi değere 1 sn ile CostSensitiveClassifier (C4.5) algoritması ulaşmıştır. Weka algoritmaları süre açısından 1 ile 49 sn arasında değerler almaktadır. Difaconn-miner ikili yapısından dolayı en düşük yanlış sınıflandırma maliyetine aşağı örneklemede 87 sn de ulaşmıştır. Mepar-miner ise aşağı örneklemede 66 sn de, yukarı örneklemede 124 sn de ulaşmıştır. Weka algoritmalarına en yakın değeri 14 sn ile Bee-miner algoritması vermiştir.

En düşük yanlış sınıflandırma maliyetine sahip CostSensitiveClassifier(PART) algoritması incelendiğinde test doğruluğu %70.9 çıkmıştır. Sonuç olarak 6 kural oluşmuştur. Kurallar şu şekildedir;

- Kopca > 1 AND Kafes <= 2 AND Kopca <= 2: 1
- Kopca <= 2 AND Kopca <= 1: 0
- Kafes > 2 AND Klips > 1: 0
- Kopca > 2 AND Klips > 2 AND Kafes <= 1: 1
- Klips <= 2 AND Kafes > 1: 1
- Diğer ihtimaller: 0

1: Sonuç olumludur. 0: Sonuç olumsuzdur.

Örneğin 4. kural incelendiğinde iplik kalitesinin iyi olması için klips 4.5 mm, kopça h2dr08 ve kafesin dar olması gereklidir. İşletmenin elinde 4.5 mm klips yoksa 4 veya 3.5 mm klips mevcutsa, 5. kurala bakılarak kafes mevcut veya geniş yapılarak yine kaliteli iplik üretilebilir. Bu durum işletmeye çok büyük esneklik sağlamaktadır. Bu çalışmada amaçlanan sınıflandırma yöntemleri ile kural kümesi oluşturmak ve işletmeye esneklik sağlamaktır ve istenilen amaca ulaşılmıştır.

6. SONUÇ

Veri madenciliğinin uygulanabilmesi için yığın halinde verilerin elimizde bulunması ön koşuldur. Veri madenciliği farklı formatlarda çok sayıda veritabanındaki yığın halindeki veriler arasında gizli bir şekilde bulunan örüntülerin çıkarılmasına yarayan bir araçtır. Özellikle zaman içinde verinin azlığının değil, çokluğunun bir sorun olması ve bilgisayarların veri saklama ve işleme hızlarındaki inanılmaz artışların sonucunda veri madenciliğinin güncelliği her geçen gün artmış ve artmaktadır.

Birçok veri madenciliği fonksiyonunun içinde, veri tabanlarında gizli olan bilgiyi kullanıcıların anlayabileceği kural listeleri veya karar ağaçları formunda sunan sınıflandırma dikkat çekmektedir. Literatürde pek çok araştırmacı, büyük boyutlu veri kümelerinde gizli olan bilgiyi anlaşılır ve doğru kurallar şeklinde sunan etkin sınıflandırma algoritmaları geliştirmek için çalışmalar yapmaktadırlar.

Bu çalışma da veri madenciliği detaylı olarak incelenmiştir. İplik üretim süreci detaylı olarak anlatılmıştır. İplik kalitesini etkileyen faktörlerden bahsedilmiştir. Bu faktörlerin önem sıraları Taguchi yöntemi ile belirlenmiştir. 6 faktörden, Kafes, Kopça ve Klips faktörlerinin üçünün de sonuçlar üzerinde etkili faktörler olduğu görülmüştür. Büküm, Manşon ve Devir ise daha az etkili faktörlerdir. Etkili 3 faktörün tüm seviyeleri ile denendiği yeni veri kümesi oluşturulmuştur. Her test 6 kez tekrarlanmıştır. Bu sayede test hatası engellenmeye çalışılmıştır. Çıkan sonuçlar veri madenciliği paket programlarında maliyete duyarsız ve maliyete duyarlı olarak incelenmiştir. Paket program olarak; Weka 3.8.1 ve MT-VeMa 1.0 kullanılmıştır. Çalışmada; PART, C4.5, SimpleCART, RIPPER, NBTree, DecisionTable, BEE-miner, MEPAR-miner ve DIFACONN-miner algoritmaları üzerinde sınıflandırma işlemi gerçekleştirilmiştir. İplik kalitesini etkileyen faktörlerden bir sınıflandırma yapılarak kural oluşturulmuştur. Sınıflandırma, işletmenin gerçek maliyet değerlerine dayalı olarak yapılmıştır. Bu

sayede elde edilen kural kümesinin etkin kullanımı ile işletmenin ilerde olası maliyet kayıplarını azaltacaktır. Çıkan sonuçlar işletmeyle paylaşılmıştır.

Bu çalışma ile veri madenciliği uygulamalarının, bir tekstil şirketinde gerçek verilerle nasıl sonuca ulaştığı gösterilmiş ve sürece katkıda bulunulmuştur.

KAYNAKLAR

[1] Kalıkov, A., Veri Madenciliği ve Bir E-Ticaret Uygulaması, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 2006.

[2] İnan, O., Veri Madenciliği, Yüksek Lisans Tezi, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Konya, 2003.

[3] Thuarisingham, B.M., Web Data Mining and Applications in Business Intelligence and Counter Terrorism, CRC Press LLC, Boca Raton, FL,USA, 2003.

[4] Zhou, Z., Three Perspectives of Data Mining, **Artificial Intelligence**, **No.143** (2003), pp.139-146, 2002.

[5] Güven, A., Bozkurt, Ö. ve Kalıpsız, O., Veri Madenciliğinin Geleceği, Yıldız Teknik Üniversitesi, Bilgisayar Mühendisliği Bölümü, İstanbul, 2007.

[6] Baykal, A., Veri Madenciliği Uygulama Alanları, **D.Ü.Ziya Gökalp Eğitim Fakültesi Dergisi**, **7**, s. 95-107, 2006.

[7] Savaş, S., Topaloğlu, N., ve Yılmaz, M., Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri, **İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi**, **S. 21**, s. 1-23, 2012.

[8] Frawley, William J., Piatetsky-Shapiro, Gregory; Matheus, Christopher J., Knowledge Discovery in Databases: an Overview, AAAI/MIT Press, 1992.

[9] Hand, D.J., Data Mining: Statistics and More?, **The American Statistician**, **Cilt 52**, s. 112-118, 1998.

[10] Kitler R., Wang W., The Emerging Role of Data Mining, **Solid State Technology**, **Cilt 42**, S.11, s. 45,1998.

[11] Jacobs, P., Data Mining: What General Managers Need to Know, **Harvard Management Update**, **Cilt 4**, S.10, s. 8, 1999.

[12] Berry M.J., Linoff G.S., Mastering Data Mining; John Wiley & Sons, s. 5-407, Inc New York, 2000.

[13] SWIFT, R.S., Accelerating Customer Relationship:Using CRM and Relationship Technologies, Prentice Hall PTR, Eylül 2000

[14] Larose, Daniel T., Discovering Knowledge in Data, An Introduction to Data Mining, John Wiley & Sons,Inc., s. 2, 2005.

[15] Doğan, Ş., ve Türkoğlu,İ., Hypothyroidi and Hyperthyroidi Detection from Thyroid Hormone Parameters by Using Decision Trees, **Doğu Anadolu Bölgesi Araştırmaları Dergisi**, **Cilt 5**, S.2, s.163-169, 2007.

[16] Pilot Software, An Introduction to Data Mining, Whitepaper, Pilot Software, 1998.

[17] Albayrak, M., EEG Sinyallerindeki Epileptiform Aktivitenin Veri Madenciliği Süreci ile Tespiti, Doktora Tezi, Sakarya Üniversitesi, Sakarya, 2008.

[18] Akgöbek, Ö., Çakır, F., Veri Madenciliğinde Bir Uzman Sistem Tasarımı, **Akademik Bilişim**, **9**, Harran Üniversitesi, Şanlıurfa, s. 801-806, 11-13 Şubat 2009.

[19] Akpınar, H., Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği, **İ.Ü. İşletme Fakültesi Dergisi**, **Cilt 29**, S.1, s.1-22, 2000.

[20] Vahaplar, A., İnceoğlu, M., Veri Madenciliği ve Elektronik Ticaret, Türkiye'de İnternet Konferansları, Harbiye İstanbul, 1-3 Kasım 2001.

[21] Sever, H. ve Oğuz, B., Veritabanlarında Bilgi Keşfine Formal Bir Yaklaşım, Kısım 1: Eşleştirme Sorguları ve Algoritmalar, **Bilgi Dünyası**, **3**, 2, s. 173-204, Ekim 2002.

[22] Han, J., Cai, Y., ve Cercone, N., Knowledge discovery in databases: An attribute-oriented approach, Bildiri Kitabı, 18. VLDB Konferansı, Vancouver, British Columbia, Canada, s. 547-559, 1992.

[23] Rastogi, R. ve Shim, K., Scalable algorithms for mining large databases, 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA: ACM Pres, s. 73-140, 1999.

[24] Fayyad, U. M. ve Irani, K. B., Multi interval discretization of continuous attributes for classification learning. R. Bajcsy, (ed.), 13th International Joint Conference on Artificial Intelligence, Morgan Kauffmann, New York, NY, s. 1022-1027, 1993.

[25] Quinlan, J. R., The effect of noise on concept learning, R. Michalski, J. Carbonell, ve T. Mitchell (eds.), Machine Learning: An Artificial Intelligence Approach, San Mateo, CA: Morgan Kauffmann Inc., Cilt 2, s. 149-166, 1986.

[26] Chan, K.C.C., Wong, A. K. C., A statistical technique for extracting classificatory knowledge from databases, G. Piatetsky-Shapiro, W. J. Frawley (eds.), **Knowledge Discovery in Databases** , s. 107-123, Cambridge, MA: AAAI/MIT, 1991.

[27] Lee, S. K., An extended relational database model for uncertain and imprecise information, 18th International Conference on Very Large Databases, VLDB'92 Konferans Bildiri Kitabı, Vancouver, British Columbia, Canada, s. 211-218, 1992.

[28] Grzymala-Busse, J. W., On the unknown attribute values in learning from examples, Z. W. Ras ve M. Zemankowa (eds.), Methodologies for Intelligent Systems, **Lecture Notes in AI**, **Cilt 542**, s. 368-377, New York: Springer-Verlag., 1991.

[29] Grzymala-Busse, D. M. ve Grzymala-Busse, J. W. ,Comparison of machine learning and knowledge acquisition methods of rule induction based on rough sets, The International Workshop on Rough Sets and Knowledge Discovery , s. 297-306, Banff, Alberta, Canada, 1993.

[30] Luba, T., Lasocki, R., On unknown attribute values in functional dependencies. T.Y. Lined, The International Workshop on Rough Sets and Soft Computing, San Jose, California, s. 490-497, 1994.

[31] Thiesson, B., Accelerated quantification of Bayesian networks with incomplete data, U. Fayyad ve R. Uthurusamy (eds.), The First International Conference on Knowledge Discovery and Data Mining, Montreal, Quebec, Canada, s. 306-311, 1995.

[32] Quinlan, J. R., Induction of decision trees, **Machine Learning**, **Cilt l**, s. 81-106, 1986.

[33] Kulluk, S., Karınca Koloni Optimizasyonu ile Yapay Sinir Ağlarından Kural Çıkarımı, Doktora Tezi, Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Kayseri, 2009.

[34] Choubey, S. K., Deogun, J. S., Raghavan, V. V. ve Sever, H., A Comparison of Feature Selection Algorithms in the Context of Rough Classifiers, **The 5th IEEE International Conference on Fuzzy Systems**, New Orleans, LA, **cilt. 2**, s. 1122-1128, 1996.

[35] Pawlak, Z., Slowinski, K., ve Slowinski, R., Rough classification of patients after highly selective vagotomy for duodenal ulcer. **International Journal of Man-Machine Studies**, **Cilt 24**, s. 413-433, 1986.

[36] Baim, P., A method for attribute selection in inductive learning systems, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, s. 888-896, 1988.

[37] Almuallim, H., Dietterich, T., Learning with many irrelevant features, 3rd Conference of American Association on Artificial Intelligence, s. 547-552, AAAI Press. Menlo Park, CA, 1991.

[38] Kira, K., Rendell, L., The feature selection problem: Tradational methods and a new algorithm, 4th Conference of American Association on Artificial Intelligence, s. 129-134, AAAI Pres,1992.

[39] Deogun, J. S., Raghavan, V. V. ve Sever, H., Exploiting upper approximations in the rough set methodology, U. Fayyad ve R. Uthurusamy (eds.), The First International Conference on Knowledge Discovery and Data Mining, s. 69-74, Montreal, Quebec, Canada,1995

[40] Hulten, G., Spencer, L. ve Domingos, P., Mining time-changing data streams, Bildiri Kitabı, 7th Acm Sıgkdd International Conference on Knowledge Discovery and Data Mining, San Fransisco, CA: ACM Pres, 2001.

[41] Paton, N.W., Diaz, O., Active database systems, **Computing Surveys**, **no. 31**(1), s. 63-103, 1999.

[42] Ching, J.Y., Wong, A.K.C., ve Chan, K.C.C.,Class-dependent discretization for inductive learning from continuous and mixed mode data, **IEEE Transactions on Knowledge and Data Engineering**, **no.17**(7), s. 641-651, 1995.

[43] Fayyad, U.M, Piatetsky-Shapiro, G., & Smyth, P., From data mining to knowledge discovery in databases , **AI Magazine**, **no.17**(3), pp. 37-54, 1996.

[44] Ozekes S., Veri Madenciliği Modelleri ve Uygulama Alanları, **İstanbul Ticaret Üniversitesi Dergisi**, **no. 3** (3), s. 65-82, 2003.

[45] Han, J., Kamber, M., Data Mining: Concept and Techniques, s. 24-703, Morgan Kaufmann Publishers, San Francisco, 2001.

[46] Albayrak, A.S., Yılmaz, Ş.K., Veri Madenciliği: Karar Ağacı Algoritmaları ve İMKB Verileri Üzerine Bir Uygulama, **S.D.Ü. İktisadi ve İdari Bilimler Fakültesi Dergisi**, **Cilt 14**, S.1, s. 31-52, 2009.

[47] İnal, E., Olap, Lisans Ödevi, Gebze Yüksek Teknoloji Enstitüsü Bilgisayar Mühendisliği Bölümü, Kocaeli, 2007.

[48] Bishop, C., Neural Networks for Pattern Recognition, Oxford Univ Pres, 1996.

[49] Alpaydın, E., Zeki Veri Madenciliği:Ham Veriden Altın Bilgiye Ulaşma Yöntemleri, Bilişim 2000 Eğitim Semineri, Boğaziçi Üniversitesi Bilgisayar Mühendisliği Bölümü, İstanbul, 2000.

[50] İstatistik Portalı, Veri Tabanı Ve Veri Madenciliği, Marmara Üniversitesi, İstanbul, 2011.

[51] Telcioglu M.B., Montaj Hattı Dengeleme Problemlerinin Genetik Algoritma Teknigi Kullanılarak Çözülmesi ve Bilgisayar Programı Uygulaması Pegasus Yazılımı, Yüksek Lisans Tezi, Erciyes Üniversitesi, Kayseri, 2002.

[52] Holland, J. H., Adaption in Natural and Artificial Systems, University of Michigan Pres, Ann Arbor, MI, 1975.

[53] Tou, J. T., Gonzalez, R. C., Pattern Recognition Principles, London: Addison-Wesley, 1974.

[54] Mitra, S., Acharya, T., Data Mining: Multimedia, Soft Computing and Bioinformatics, Wiley-Interscience, Inc. Hoboken, New Jersey, 2003.

[55] Witten, I. H., Frank, E., Data Mining: Practical Machine Learning Tools and Techniques, The Morgan Kaufmann Publishers, San Fransisco, CA, 2005.

[56] Coşkun, S., ve diğerleri, Lojistik Regresyon Analizinin incelenmesi ve Diş Hekimliğinde Bir Uygulaması, **Cumhuriyet Üniversitesi Diş Hekimliği Fakültesi Dergisi**, **Cilt 7**, S.1, s. 42-50, 2004.

[57] Bonabeau, E., Theraulaz, G., Swarm Smarts, Scientific American Inc., 72-79, 2000.

[58] Dorigo, M., Optimization, Learning and Natural Algorithms, PhD Thesis, Politecnico di Milano, Italy, 1992.

[59] Kennedy, J., Eberhart, R., Particle Swarm Optimization, In Proc. of The IEEE Int. Conf. on Neural Networks, Piscataway, NJ, 1942-1948, 1995.

[60] Pham, D.T., Ghanbarzadeh, A., Koç, E., Otri, S., Rahim, S., Zaidi, M. 2006a. "The Bees Algorithm – A Novel Tool for Complex Optimisation Problems", Proc. Innovative Production Machines and Systems Virtual Conf., 454-461.

[61] Bishop, J. M., Stochastic Searching Networks, Proc. 1st IEEE Conf. on Artificial Neural Networks, 329-331, 1989.

[62] Schank, R., Dynamic Memory: A Theory of Learning in Computers and People, New York: Cambridge University Press, 1982.

[63] Kolodner, J. L., Case-Based Reasoning, San Mateo: Morgan Kaufmann, 1993.

[64] Aha, D. W., Kibler, D., Albert, M. K., Instance-based Learning Algorithms, **Machine Learning**, **6**, 37-66, 1991.

[65] Pawlak, Z., Rough Sets Theoretical Aspects of Reasoning About Data, Kluwer Academic Publishers, 1991.

[66] Binay, H. S., Yatırım Kararlarında Kaba Küme Yaklaşımı, Doktora Tezi, 2002.

[67] Pawlak, Z., Skowron, A., Rough Set Rudiments, The International Workshop on Rough Sets and Soft Computing, San Jose, California, 72, 1994.

[68] Zadeh, L. A., Fuzzy Sets, **Information and Control**, **8**(3), 338-353, 1965.

[69] Ye, N., The Handbook of Data Mining, Lawrance Erlbaum, 1st edition, 2003.

[70] Michalski, R. S., Step, R. E., Learning from Observation: Conceptual Clustering. In R. S. Michalski, J. G. Oneli C., and Mite T. M., hell. Editors, Machine Learning: An Artificial Intelligence Approach, 1, Morgan Kaufmann, 331-363,1983.

[71] Jain, K. K., Dubes, R. C., Algorithms for Clustering Data, Englewood Cliffs, NJ: Prentice Hall, 1988.

[72] Aydoğan, F., E-ticarette Veri Madenciliği Yaklaşımlarıyla Müşteriye Hizmet Sunan Akıllı Modüllerin Tasarımı ve Gerçekleştirimi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, Yüksek Lisans Tezi, 2003.

[73] Ranjit, K.R., A primer on the Taguchi method: Van Nostrand Reinhold,1990.

[74] Güral,G., Gaz Kaynağında proses parametrelerinin optimizasyonu, Dokuz Eylül Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi Eylül, İzmir, 2003.

[75] Gökçe G, Taşgetiren S., Kalite için Deney Tasarımı, **Makine Teknolojileri Elektronik Dergisi**, **Cilt 6**, No:1 (71-83), 2009.

[76] Tapkan, P., Özbakır, L., Baykasoğlu, A. ve Kulluk, S., Acost sensitive classification algorithm:BEE-Miner, **Knowledge-Based Systems**, **95**, s. 99–113, 2016.

[77] Provost, F. 2000. Machine learning from imbalanced data sets, Proceedings of the AAAI'2000 Workshop on Imbalanced Data.

[78] Elkan, C. 2001. The foundations of cost-sensitive learning, Proceedings of the 17th International Joint Conference of Artificial Intelligence, Sattle, Washington, Morgan Kaufman, 973-978.

[79] Zadrozny, B., Elkan, C. 2001. "Learning and making decisions when costs and probabilitiesare both unknown", Proceedings of The 7th International Conference on KnowledgeDiscovery and Data Mining, ACM SIGKDD, 204-213.

[80] Japkowicz, N. 2000. The class imbalance problem:significance and strategies, In Proceedings of The 2000 International Conference on Artificial Intelligence (ICAI'2000).

[81] Tapkan, P., Özbakır, L., Baykasoğlu, A. 2011. "Weka ile veri madenciliği süreci ve örnek uygulama", Endüstri mühendisliği yazılımları ve uygulamaları kongresi, İzmir, Türkiye.

[82] Kulluk, S., Özbakır, L., Baykasoğlu, A. ve Tapkan, P., Cost-sensitive meta-learning classifiers: MEPAR-miner and DIFACONN-miner, **Knowledge-Based Systems**, **98**, s. 148–161, 2016.

[83] Oltean, M., Dumitrescu, D. 2002. Multi expression programming. Department of ComputeerScience, Technical Report, Romania.

[84] Özbakır, L., Baykasoğlu, A., Kulluk, S., Tapkan, P. Z.., Arı Sistemi ve Meta-öğrenme Yaklaşımları ile Maliyete Duyarlı Veri Madenciliği, Tübitak Program Kodu: 1001 Proje No: 111M528 , 2014.

[85] Frank, E., Witten, I.H. 1998. Generating Accurate Rule Sets without Global Optimization, Proc. 15th Int. Conf. On Machine Learning, 144-151.

[86] Quinlan, J.R. 1993. *C4.5:* Programs for Machine Learning. San Francisco, USA: MorganKaufmann.

[87] Breiman, L., Friedman, J., Stone, C. J., Olshen, R. A. 1984. Classification and regressiontrees (1. Basım) Chapman and Hall.

[88] Cohen, W. 1995. Fast Effective Rule Induction, Proc. 12th Int. Conf. On Machine Learning, 115-123.

[89] Kohavi,R.1996: Scaling Up the Accuracy of Naive-Bayes Classifiers: A Decision-Tree Hybrid. In: Second International Conference on Knoledge Discovery and Data Mining, 202-207.

[90] Kohavi, R. 1995. The Power of Decision Tables, Proc. 8th Eur. Conf. on Machine Learning, 174-189.

[91] Domingos, P. 1999. MetaCost: a general method for making classifiers cost-sensitive, In Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining, ACM Press, 155-164.

ÖZGEÇMİŞ

Tayfun ÖZMEN 1986 yılında Kayseri'de doğdu. İlk, orta ve lise öğrenimini Kayseri'de tamamladı. 2003 yılında Erciyes Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği bölümünü kazandı ve 2008 yılında mezun oldu. 2010 yılında Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Endüstri Mühendisliği Anabilim dalında yüksek lisans eğitimine başladı. 2010 yılı Nisan ayında Orta Anadolu Tekstil T.A.Ş'de Elektrik Bakım Mühendisi olarak göreve başladı ve halen aynı işyerinde görevine devam etmektedir.

İletişim Bilgileri:
Orta Anadolu Tekstil Fabrikası
Aydınlık Evler Mevkii
38060 KAYSERİ
Tel: (0 352) 207 87 00 / 12520
e-mail: ozmentayfun@hotmail.com