

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**KABA KÜMELER YARDIMIYLA ÖNEMSİZ VE KAYIP
TÜRDEKİ EKSİK VERİLERDEN BULANIK KURALLARIN
ÇIKARILMASI**

Gülnur AVŞAR

Tez Yöneticisi:
Yrd.Doç.Dr. Mehmet KAYA

YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

ELAZIĞ, 2007

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**KABA KÜMELER YARDIMIYLA ÖNEMSİZ VE KAYIP
TÜRDEKİ EKSİK VERİLERDEN BULANIK KURALLARIN
ÇIKARILMASI**

Gölnur AVŞAR

Yüksek Lisans Tezi
Bilgisayar Mühendisliği Anabilim Dalı

Bu tez, 22.06.2007 tarihinde aşğıda belirtilen jüri tarafından oybirliği /oyçokluğu ile başarılı / başarısız olarak değeriendirilmiştir.

Danışman: Yrd.Doç.Dr. Mehmet KAYA

Üye: Prof.Dr. Yakup DEMİR

Üye: Yrd.Doç.Dr. Burhan ERGEN

Bu tezin kabulü, Fen Bilimleri Enstitüsü Yönetim Kurulu'nun/...../..... tarih ve sayılı kararıyla onaylanmıştır.

TEŞEKKÜR

Bu tez çalışması boyunca, bilgi ve yardımlarını esirgemeyen danışmanım Sayın Yrd. Doç. Dr. Mehmet KAYA'ya teşekkürlerimi ve şükranlarımı sunarım.

İÇİNDEKİLER

Sayfa

TEŞEKKÜR	
İÇİNDEKİLER	I
ŞEKİLLER LİSTESİ	III
TABLolar LİSTESİ	IV
ÖZET	V
ABSTRACT	VI
1. GİRİŞ	1
1.1. Tezin Amacı	2
1.2. Tezde Geliştirilenler	2
1.3. Tez Taslağı	2
2. KABA KÜME VE BULANIK KÜME KAVRAMLARI	3
2.1. Kaba Küme Teorisi	3
2.2. Kaba Küme Teorisinin Temel Kavramları	3
2.2.1. Bilgi Sistemi	3
2.2.2. Ayırtedilemez İlişki	5
2.2.3. Düşük ve Yüksek Yaklaşımlar	6
2.2.4. Yaklaşımların Doğruluk Oranı	8
2.2.5. Niteliklerin Bağımsızlığı	9
2.2.6. Niteliklerin Özü ve İndirgemeleri	9
2.2.7. Nitelik Değerlerinin Özü ve İndirgemeleri	12
2.2.8. Sınıflandırma	13
2.3. Karar Tabloları	15
2.3.1. D-Lüzumsuz Nitelikler	16
2.3.2. Niteliklerin Bağlı Öz ve Bağlı İndirgemeleri	16
2.4. Karar Tablosu Analizinin Temel Adımları	17
2.4.1. Karar Kuralları	22
2.5. Bulanık Küme Teorisi	22
2.5.1. Giriş	22
2.5.2. Bulanık Küme	23
2.5.3. Üyelik Fonksiyonları	24
2.5.4. Küme İşlemleri	25
3. EKSİK NİTELİK DEĞERİ İÇEREN VERİ KÜMELERİ ÜZERİNDE KABA KÜME TEORİSİ	26
3.1. Giriş	26
3.2. Eksik Nitelik Değerleri ve Karakteristik İlişkiler	27
3.3. Yüksek ve Düşük Yaklaşım	31

4. KABA KÜME TEORİSİYLE EKSİK NİCEL VERİLERDEN BULANIK KURAL	
ÇIKARIMI.....	35
4.1.Giriş.....	35
4.2. Önerilen Algoritma	35
4.3. Algoritmanın Eksik Bir Veri Kümesine Uygulanışı	40
5. UYGULAMA YAZILIMI.....	47
5.1. Uygulama Ortamı	49
5.2. Geliştirilen Yazılım ile Elde Edilen Kurallar	50
6. SONUÇLAR.....	61
KAYNAKLAR.....	63
ÖZGEÇMİŞ	65

ŞEKİLLER LİSTESİ

	<u>Sayfa</u>
Şekil 2.1. X kümesinin yüksek ve düşük yaklaşımlarının şematik gösterimi	8
Şekil 2.2. Standart üyelik fonksiyonu çeşitleri	25
Şekil 4.1. A1 niteliği için üyelik fonksiyonu.....	40
Şekil 4.2. A2 niteliği için üyelik fonksiyonu.....	40
Şekil 5.1. Tiroid hastalığı verileri.....	47
Şekil 5.2. Uygulama yazılımında kullanılan üyelik fonksiyonları	48
Şekil 5.3. Uygulama yazılımı arayüzü	49
Şekil 5.4. Uygulama yazılımı ile kesin ve olası kuralların elde edilmesi	50
Şekil 5.5. Uygulama yazılımı ile ikinci durum için kesin ve olası kuralların elde edilmesi	53
Şekil 5.6. Kesin ve olası kurallar için başarı oranları	59
Şekil 5.7. Analiz sonucu kesin kural sayısı ve ortak kesin kural sayısı.....	59
Şekil 5.8. Analiz sonucu olası kural sayısı ve ortak olası kural sayısı	59

TABLÖLAR LİSTESİ

	<u>Sayfa</u>
Tablo 2.1. 10 x 3 boyutlu matris	4
Tablo 2.2. Örnek bir bilgi fonksiyonu.....	4
Tablo 2.3. Veri kümesindeki özdeş nesneler.....	5
Tablo 2.4. a_1 ve a_2 nitelikleri için birincil kümeler.....	5
Tablo 2.5. Niteliklerin birinin hariç tutulması durumunda birincil küme sayıları	9
Tablo 2.6. Örnek bir ayırtedilebilirlik matrisi	10
Tablo 2.7. İndirgeme sonrası elde edilen birincil küme	11
Tablo 2.8. Ayırtedilebilirlik fonksiyonlarının elde edileceği ayırtedilebilirlik matrisi	12
Tablo 2.9. İndirgemeler sonucunda elde edilen birincil küme	13
Tablo 2.10. Sınıflara ait düşük yaklaşım,yüksek yaklaşım ve doğruluk oranı değerleri.....	14
Tablo 2.11. Karar tablosu örneği.....	15
Tablo 2.12. İki karar niteliği içeren karar tablosu örneği.....	16
Tablo 2.13. Kümeye aitlikleri doğrultusunda elde edilen karar tablosu	17
Tablo 2.14. Sınıflandırmalara ait yaklaşım ve doğruluk oranı değerleri.....	18
Tablo 2.15 Ayırtedilebilirlik matrisi	18
Tablo 2.16. $\{a_1, a_2\}$ indirgemesi için Tablo 2.12’den elde edilen indirgenmiş karar tablosu ..	19
Tablo 2.17. $\{a_2, a_3\}$ indirgemesi için Tablo 2.12’den elde edilen indirgenmiş karar tablosu ..	20
Tablo 2.18. $\{a_1, a_2\}$ indirgemesi için D-ayırtedilebilirlik matrisi.....	20
Tablo 2.19. $\{a_1, a_2\}$ alt uzayında karar tablosunun son şekli.....	21
Tablo 3.1.Tam olarak tanımlanmış karar tablosuna bir örnek	27
Tablo 3.2. Tüm eksik niteliklerin ‘kayıp’ türde olduğu eksik bir karar tablosu örneği	28
Tablo 3.3. Tüm eksik niteliklerin “önemsiz”olarak tanımlandığı karar tablosu.	29
Tablo 3.4. Eksik niteliklerin bir kısmının kayıp, bir kısmının önemsiz olarak tanımlandığı karar tablosu	30
Tablo 4.1. Eksik veri kümesi.....	40
Tablo 4.2. Bulanık veri kümesi	41
Tablo 4.3. A1 niteliği için tahmin sonrası elde edilen veri kümesi.....	42
Tablo 4.4. A1 niteliği için tahmin sonrası elde edilen bulanık kümeler	42
Tablo 4.5. A2 niteliği için tahmin sonrası elde edilen veri kümesi.....	43
Tablo 4.6. A2 niteliği için tahmin sonrası elde edilen bulanık kümeler	43
Tablo 5.1. Uygulama yazılımı sonuçlarının karşılaştırılması	58

ÖZET

Yüksek Lisans Tezi

KABA KÜMELER YARDIMIYLA ÖNEMSİZ VE KAYIP TÜRDEKİ EKSİK VERİLERDEN BULANIK KURALLARIN ÇIKARILMASI

Gölnur AVŞAR

Fırat Üniversitesi

Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

2007, Sayfa: 65

Kaba küme teorisi, doğrulanmış mantığa, tutarsızlık gösteren verilere ve kesin olmayan gizli çıkarımların keşfine izin veren matematiksel bir araçtır. Kaba kümeler, bulanık kümelerde olduğu gibi kesin sınırlamaları kabul etmeyen ve eksik, yetersiz ve belirsiz bilgileri düzenleyerek, veri analizi için uygun hale getiren yapılardır. Bu tez çalışmasında, kaba küme teorisi kullanılarak, eksik veri kümelerinden bilgi çıkarımı yapılmıştır. Eksik veri kümeleri, günümüzde en çok hastalık verilerinde bulunmaktadır. Bu verilerden kurallar çıkarmak hastalık teşhisinde son derece önemlidir. Önemsiz ve kayıp olmak üzere, iki farklı tipte eksik veri içeren, bir veri kümesi ile çalışılmıştır. Nicel değerli eksik veri kümelerinden, kaba küme yaklaşımı ile bulanık kurallar çıkaran bir algoritma ile bilgi çıkarımı sağlanmıştır. Sadece önemsiz türde eksik nitelik içeren veri kümeleri için tanımlanmış olan algoritma, bu tez çalışması sırasında hem önemsiz, hem de kayıp türde eksik niteliği birlikte içeren veri kümelerinde uygulanabilecek biçimde geliştirilmiştir. Geliştirilen algoritma için tiroit hastalığı verileri üzerinde çalışan bir uygulama yazılımı oluşturulmuştur. Eksik niteliklerin sayıları ve türleri üzerinde yapılan değişikliklerle elde edilen altı farklı durum için yazılım test edilmiştir. Geliştirilen algoritmada, kesin kurallar da %100, olası kurallarda %91 başarı sağlanmıştır. En düşük başarı oranları tüm nitelik değerlerinin kayıp türde olduğu durum için alınmıştır. Tüm eksik niteliklerin kayıp türde olmasının elde edilen kural sayısını arttırdığı ve başarı oranında azalmaya neden olduğu sonucuna varılmıştır.

Anahtar Kelimeler: Kaba küme teorisi, Eksik veri kümesi, Kayıp veri, Önemsiz veri, Bulanık Kurallar

ABSTRACT

Masters Thesis

EXTRACTION OF FUZZY RULES FROM INCOMPLETE DATA WITH DO NOT CARE AND LOST VALUE BY ROUGH SETS

Gülnur AVŞAR

Firat University

Graduate School of Natural and Applied Sciences

Department of Computer Engineering

2007, Page: 65

Rough set theory is a mathematical approach to imprecision, vagueness and uncertainty in data analysis. This study deals with the problem of producing a set of certain and possible fuzzy rules from incomplete quantitative data. In this thesis, knowledge extraction has been done from incomplete data sets using rough set theory. In the present, incomplete data mainly exist in medical data sets. It is extremely important to extract rules from these data for medical diagnosis. Through this thesis, a data set including two different incomplete data which are called as “do not care” and “lost” data was used. A novel algorithm has been proposed to extract fuzzy rules by rough sets from incomplete data with quantitative values, as apart from the previous studies which handle the only “do not care” data or the only missing values. For this purpose, a software has been developed for evaluating the proposed algorithm on thyroid data set.

The algorithm has been tested for six different cases which use various type and number of incomplete data. The experimental results demonstrates that the proposed algorithm provides the accuracy rate of 100% in certain rules and 91% in possible rules. The lowest accuracy rate has been obtained for the case in which all of incomplete values are lost. In such a case, as the number of rules increases, the accuracy rate decreases.

Keywords: Rough set theory, incomplete data set, lost value, do not care value, fuzzy rules

1. GİRİŞ

Pawlak tarafından 1982 yılında önerilen kaba küme teorisi, uzman sistemler için muhakeme ve bilgi çıkarımında kullanılan matematiksel bir metottur [1,2]. Temel prensip olarak, eşitlik sınıfı kavramını kullanan kaba küme teorisinin, pek çok alanda uygulamaları mevcuttur. Orłowska'nın eksik bilgi ile muhakemesi [2], Germano ve Alexandre'nın bilgi tabanı indirgemesi [3], Lingras ve Yao'nun veri madenciliği [4], Zhong ve diğerlerinin kural keşfi [5] bu uygulamalara örnek verilebilir. Kaba küme teorisinin eğitim verileri üzerinde yararlı bilginin çıkarımındaki başarısı, veri madenciliği ve veritabanı üzerine yapılan çalışmalarda da sıklıkla kullanılmasına sebep olmuştur.

Çok sayıda öğrenme yaklaşımı, tam veri kümelerinden kural çıkarımı yapar. Ancak mevcut veri kümesi her zaman tam olmayabilir. Zaten gerçek dünya uygulamalarında, eksik verilerin ve tutarsızlıkların olması zaman zaman kaçınılmazdır. Şayet bir veri kümesinde bazı nitelik değerleri bilinmiyorsa, bu veri kümesine eksik veri kümesi denir. Eksik veri kümelerindeki bu problemi çözümlmek için bir çok metot önerilmiştir [6-8]. Eksik bilgi sistemlerinden, karar kurallarının elde edilmesi için geliştirilen bu teknikler, eksik verilerin yönetiminde farklı metotlar uygulamaktadır. Bu metotlardan en basiti, eksik nitelik içeren nesnelerin doğrudan silinmesidir [9]. Diğer basit bir metot da, eksik nitelik değerini istatistiksel analiz sonucu elde ettiği mümkün bir değerle ya da eksik veri içeren bu nesne ile aynı giriş verilerine sahip, bir başka nesnenin bu niteliğe ait değeri ile değiştirilmesidir [4]. Tüm bu teknikler verileri düzeltme yoluyla eksik bilgi sistemlerini tam bilgi sistemlerine dönüştürür. Ancak, bu metotlar yoluyla elde edilmiş kurallar, eksik veri ihtimalinden dolayı kesinsizlikler içerir. Yani, kesin kural sayısı net bir şekilde belirlenemez. Kaba küme teorisi, kural çıkarımı ve sınıflandırma konusunda temel problem olan tutarsız ve eksik bilgi ile uğraşan doğal bir metot vermektedir. Kaba küme teorisinde, diğer bir grup teknikte olduğu gibi veri kümesi boyutu değiştirilmez ya da eksik veriler tahmin edilmez [3, 10]. Bu teoride kural çıkarımı yaklaşımlar yoluyla gerçekleşir.

Gerçek dünya uygulamalarındaki eğitim verileri, çoğunlukla nicel değerlerden oluşur. Bu nedenle, nicel verilerle çalışan bir öğrenme algoritmasına olan ihtiyaç kaçınılmaz olmuştur. Bulanık küme kavramı, bu gibi nicel değerli veri kümelerinin, üyelik fonksiyonu ve sözel terimlerle gösterimindeki basitliği ve insanların muhakeme yeteneğine olan benzerliği nedeniyle sıklıkla kullanılmıştır. Bu sebeple, kaba küme teorisi ve bulanık kümeleri hatasız bilginin yüksek doğruluk miktarı ile elde edilmesi için birleştirilmesi fikri ortaya çıkmıştır [11]. Bu iki yaklaşımı birleştiren bu gibi uygulamalardan oldukça gerçekçi sonuçlar alınmıştır.

1.1. Tezin Amacı

Bu tez çalışması, kaba küme teorisini kullanarak, eksik veri kümelerinden bilgi çıkarmayı amaçlamıştır. Eksik veri kümeleri günümüzde en çok hastalık verilerinde bulunmaktadır. Bu verilerden kurallar çıkarmak son derece önemlidir. Bu nedenle bu tezde, hastalık teşhisi için eksik bir veri kümesi üzerinde, bulanık küme ve kaba küme mantığı birlikte kullanılmıştır. Tez çalışması sırasında, iki tip eksik veri ile çalışılmıştır. Bunlardan birincisi önemsiz (do not care) olarak ifade edilen eksik veridir. Bu verinin küme için önemsiz olduğu ve niteliğe ait herhangi bir değeri alabileceği düşünülür. İkinci tip eksik veri ise kayıp (lost) veridir. Bu verinin önemli olduğu ancak herhangi bir sebepten dolayı istenildiğinde ona erişilemediği ya da yanlışlıkla silindiği varsayılır.

1.2. Tezde Geliştirilenler

Bu tez çalışmasında, nicel değerli eksik veri kümeleri üzerinden bilgi çıkarımını sağlayan bir algoritma geliştirilmiştir. Başlangıçta, sadece önemsiz türde eksik veri içeren kümelere uygulanabilen bu algoritma, geliştirilerek hem önemsiz, hem de kayıp nitelik değerini birlikte içeren veri kümelerinde uygulanabilir hale getirilmiştir. Algoritma, öncelikle mevcut veri kümesini bulanıklaştırır. Ardından birincil kümelerin elde edilmesi, düşük ve yüksek yaklaşımların hesaplanması gibi kaba küme teorisine ait adımları gerçekleştirerek, önce eksik veriler hakkında tahminlerde bulunur, sonra yaklaşımlar yardımıyla kuralları elde eder. Düşük yaklaşımlardan kesin kurallar, yüksek yaklaşımlardan da mümkün (olası) kurallar elde edilir. Bu çalışmada, geliştirilen algoritma, tiroit hastalığına ait bir veri kümesi üzerinde denenmiş ve elde edilen sonuçlar, tezin ilerleyen bölümlerinde verilmiştir.

1.3. Tez Taslağı

İkinci bölümde, kaba küme teorisi ve teorinin temel kavramları ile geliştirilecek algoritmada kullanılan bulanık küme mantığı hakkında bilgi verilmektedir. Üçüncü bölümde, eksik nitelik değeri içeren veri kümeleri üzerinde kaba küme teorisinin nasıl uygulandığından bahsedilmektedir. Dördüncü bölümde, kaba küme teorisi ile eksik nicel verilerden bulanık kural çıkarımını sağlayan bir algoritmadan bahsedilmektedir. Beşinci bölümde, geliştirilen uygulama yazılımı ve bu yolla elde edilen kurallara, altıncı bölümde de tez çalışması ile elde edilen sonuçlara yer verilmektedir.

2. KABA KÜME VE BULANIK KÜME KAVRAMLARI

2.1. Kaba Küme Teorisi

Zdzislaw Pawlak [12] tarafından, 1980'li yılların başlarında tanıtılan kaba küme teorisi, belirsizliklerin ve şüphelerin üstesinden gelen matematiksel bir araçtır. Kaba küme teorisi doğrulanmış mantığa, tutarsızlık gösteren verilere ve kesinlik olmayan gizli çıkarımların keşfine izin verir. Bulanık kümelerde olduğu gibi, kaba küme teorisi de kesin sınırlamaları kabul etmeyen bir yapıdadır. Bulanık ve kaba kümeler gibi yapılar, eksik, yetersiz ve belirsiz bilgileri düzenleyerek veri analizi için uygun hale getirmektedirler. Bu gibi yeteneklerinden dolayı, kaba küme yaklaşımı, yapay zeka, makine öğrenmesi, karar analizi, veri tabanından bilgi keşfi, uzman sistemler, tümevarımsal mantık ve örüntü tanıma gibi alanların temeli olarak görülür.

Kaba küme teorisinde veriler, nitelik ve şart niteliklerinden oluşan bir tablo şeklinde saklanır. Kaba küme teorisi eğitim verilerini belli kriterler doğrultusunda bölümlere ayırmak için eşitlik sınıfı kavramını benimsemiştir. Öğrenme işleminde, düşük ve yüksek yaklaşım olmak üzere iki tür bölüm oluşturulur. Kaba küme teorisinin temelini oluşturan bu kavramlardan düşük yaklaşım yardımıyla kesin kurallar, yüksek yaklaşım yardımıyla da mümkün olabilecek, olası kurallar elde edilir.

2.2. Kaba Küme Teorisinin Temel Kavramları

2.2.1. Bilgi Sistemi

Bir bilgi sistemi IS (veya yaklaşım uzayı), bir sistem olarak görülebilir.

$$IS = (U, A) \quad (2.1)$$

Burada, U evren (nesnelerin sonlu bir kümesi $U = (\{x_1, x_2, \dots, x_m\})$), A'da nitelikler kümesidir (özellikler, değişkenler). $a \in A$ olmak üzere her nitelik (a, A nitelik kümesine ait bir nitelik olmak üzere) bir bilgi fonksiyonu tanımlar. $f_a: U \rightarrow V_a$, ile tanımlanan bilgi fonksiyonunda, V_a a niteliğinin tanım kümesi olarak isimlendirilir ve a'nın değer kümesidir.

Örnek 2.1.

Aşağıdaki veri kümesi 10 adet nesnenin üç ölçüm değerine göre sonuçlarını göstermektedir. Sonuçlar (10 x 3) yapısında bir matrisle gösterilmiştir.

Tablo 2.1. 10 x 3 boyutlu matris

2	1	3
3	2	1
2	1	3
2	2	3
1	1	4
1	1	2
3	2	1
1	1	4
2	1	3
3	2	1

Kaba küme teorisinin terminolojisi kullanılarak, bu veri kümesi bir bilgi sistemi $IS=(U,A)$ olarak düşünülebilir. U evreni ve A nitelikleri sırasıyla, nesneler kümesine ve değişkenler kümesine tekabül eder.

$$U = \{x_1, x_2, x_3, x_4, x_5, x_6, \dots, x_{10}\}$$

$$A = \{a_1, a_2, a_3\}.$$

Niteliklerin tanım kümesi;

$$V_1 = \{1,2,3\}$$

$$V_2 = \{1,2\}$$

$$V_3 = \{1,2,3,4\}$$

olur. Yani her niteliğin tanım kümesi, bu niteliğin aldığı değerlerin kümesidir. Bu sistem için bilgi fonksiyonu f_a , Tablo 2.2' de gösterilmiştir.

Tablo 2.2. Örnek bir bilgi fonksiyonu

U	a_1	a_2	a_3
x_1	2	1	3
x_2	3	2	1
x_3	2	1	3
x_4	2	2	3
x_5	1	1	4
x_6	1	1	2
x_7	3	2	1
x_8	1	1	4
x_9	2	1	3
x_{10}	3	2	1

2.2.2. Ayırtedilemez İlişki

$B \subset A$ olmak üzere niteliklerin her bir kümesi için ayırtedilemez ilişki $IND(B)$ (2.2)'deki gibi tanımlanır. X_i ve X_j gibi iki nesne, A'daki B nitelikler kümesi tarafından ayırtedilemezdir.

$$\text{Eğer } b(xi) = b(xj) \text{ ise } \forall b \subset B \text{ için} \quad (2.2)$$

$IND(B)$ eşitlik sınıfı, nesnelerin en küçük ayırtedilemez grubunu oluşturduğu için B'deki birincil küme olarak isimlendirilir. U 'nun her X_i elemanı için, $IND(B)$ ilişkisinde X_i 'nin eşitlik sınıfı $[X_i]_{IND(B)}$ biçiminde gösterilir. Birincil küme yapısı, kaba küme ile sınıflandırmada ilk adımdır.

Örnek 2.2.

Tablo 2.2'de gösterilen veri kümesinde bazı özdeş nesneler vardır. Örneğin x_1 ve x_3 nesnesi varolan verilerle birbirinden ayırtedilemez. Kullanılan üç niteliğe bağlı olarak tüm nesneler Tablo 2.3' deki gibi gruplanabilir.

Tablo 2.3. Veri kümesindeki özdeş nesneler

U/A	a_1	a_2	a_3
$\{x_1, x_3, x_9\}$	2	1	3
$\{x_2, x_7, x_{10}\}$	3	2	1
$\{x_4\}$	2	2	3
$\{x_5, x_8\}$	1	1	4
$\{x_6\}$	1	1	2

Tablo 2.3'deki her satır, bir birincil kümeyi tanımlar. U/A gösterimi, A uzayında U evreninin birincil kümelerinin dikkate alındığı anlamındadır. Örneğin, sadece a_1 ve a_2 gibi iki nitelik ile ilgileniliyorsa, ayırtedilemez ilişki $B = \{a_1, a_2\}$ olarak sınırlanır ve sonuçta elde edilen birincil kümeler Tablo 2.4'deki gibi olur.

Tablo 2.4. a_1 ve a_2 nitelikleri için birincil kümeler

U/B	a_1	a_2
$\{x_1, x_3, x_9\}$	2	1
$\{x_2, x_7, x_{10}\}$	3	2
$\{x_4\}$	2	2
$\{x_5, x_6, x_8\}$	1	1

2.2.3. Düşük ve Yüksek Yaklaşımlar

Kaba küme yaklaşımı, düşük ve yüksek yaklaşım adında iki kavrama dayanır. Düşük yaklaşım, kümeye kesin olarak ait olan elemanları, yüksek yaklaşımda kümeye ait olma olasılığı olan elemanları ifade eder.

X , U evrenindeki elemanların kümesini gösterebilir ($X \subset U$). $B \subseteq A$ olmak üzere, X 'in düşük yaklaşımı $\underline{BX}$ olarak gösterilir ve X 'de yer alan tüm bu birincil kümelerin birleşimi olarak tanımlanır.

Daha açık olarak;

$$\underline{BX} = \{x_i \in U \mid [x_i]_{Ind(B)} \subset X\} \quad (2.3)$$

şeklinde ifade edilir ve şu anlama gelir: X 'in düşük yaklaşımı, X 'de yer alan birincil kümelere ait X_i nesnelerinin kümesidir (B uzayında).

X 'in yüksek yaklaşımı, $\overline{BX}$ biçiminde gösterilir ve X ile kesişimi boş küme olmayan birincil kümelerin birleşimidir.

$$\overline{BX} = \{x_i \in U \mid [x_i]_{IND(B)} \cap X \neq \emptyset\} \quad (2.4)$$

X 'in düşük yaklaşımının, her X_i nesnesi (yani $x_i \in \underline{BX}$) için X 'e ait olduğu kesindir. X 'in yüksek yaklaşımının, her X_i nesnesi (yani $x_i \in \overline{BX}$) için ise, X 'e ait olabileceği söylenebilir. İkisinin farkı, U içinde X 'in sınırı olarak isimlendirilir.

$$BNX = \overline{BX} - \underline{BX} \quad (2.5)$$

Düşük ve yüksek yaklaşımlar özdeş (yani $\overline{BX} = \underline{BX}$) ise, X kümesi tanımlanabilir. Aksi halde X kümesi, U içerisinde tanımlanabilir değildir. U'da tanımlanamayan kümelerin dört ayrı türü vardır.

1. $\underline{BX} \neq \emptyset$ ve $\overline{BX} \neq U$ ise U'da kaba tanımlanabilir olarak isimlendirilir.
2. $\underline{BX} \neq \emptyset$ ve $\overline{BX} = U$ ise U'da dıştan tanımlanamaz olarak isimlendirilir.
3. $\underline{BX} = \emptyset$ ve $\overline{BX} \neq U$ ise U'da içten tanımlanamaz olarak isimlendirilir.
4. $\underline{BX} = \emptyset$ ve $\overline{BX} = U$ ise U'da tamamen tanımlanamaz olarak isimlendirilir.

Örnek 2.3.

X'in beş nesnesinden oluşan bir alt kümesi ile çalışıldığını varsayalım $\{X = x_1, x_3, x_4, x_5, x_9\}$. $B = \{a_1, a_2, a_3\}$ gibi üç niteliğin uzayındaki tüm verilerden bu kümenin ayırtedilmesi aşağıdaki gibi gerçekleşir. Tablo 2.3'de gösterilen sonuçlar doğrultusunda, bu kümenin düşük ve yüksek yaklaşımı aşağıdaki gibi hesaplanabilir.

Tablo 2.3' de gösterilen birincil kümeler;

$$\{x_1, x_3, x_9\}, \{x_4\}.$$

Buna göre düşük yaklaşım aşağıdaki gibi verilebilir.

$$\underline{BX} = \{x_1, x_3, x_4, x_9\}$$

X alt kümesinin yüksek yaklaşımını hesaplamak için, X alt kümesi ile en az bir tane elemanı ortak olan tüm birincil kümeler Tablo 2.3'den bulunur. Bunlar;

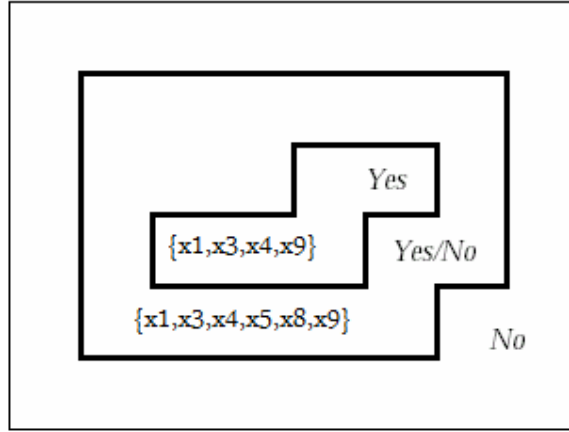
$$\{x_1, x_3, x_9\}, \{x_4\}, \{x_5, x_8\}$$

Bundan dolayı yüksek yaklaşım;

$$\overline{BX} = \{x_1, x_3, x_4, x_5, x_8, x_9\}.$$

U'da, X'in sınırı, düşük ve yüksek yaklaşımlar arasındaki fark olarak tanımlanır. Yüksek yaklaşımda olan ancak düşük yaklaşımda olmayan elemanlar olarak ifade edilebilir.

$$\begin{aligned} BNX &= \{x_1, x_3, x_4, x_5, x_8, x_9\} - \{x_1, x_3, x_4, x_9\} \\ &= \{x_5, x_8\}. \end{aligned}$$



Şekil 2.1. X kümesinin yüksek ve düşük yaklaşımlarının şematik gösterimi

2.2.4. Yaklaşımların Doğruluk Oranı

$B \subseteq A$ olmak üzere, X kümesinin doğruluk oranının ölçümü (2.6)'daki gibi tanımlanabilir.

$$\mu_B(X) = \text{card}(\underline{BX}) / \text{card}(\overline{BX}) \quad (2.6)$$

Bir kümenin önemi, X kümesinin düşük ve yüksek yaklaşımında yer alan nesnelerin sayısı ile ifade edilir. Doğruluk oranı, $0 \leq \mu_B(x) \leq 1$ aralığındadır.

X, U içerisinde tanımlanabilir ise $\mu_B(x) = 1$ 'dir. X, U içerisinde tanımlanamaz ise $\mu_B(x) < 1$ 'dir.

Örnek 2.4

Örnek 2.3'ün düşük yaklaşımında yer alan nesne sayısı 4'e, yüksek yaklaşımda yer alan nesne sayısı 6'ya eşittir. Buna göre, X kümesinin doğruluk oranı;

$$\mu_B(x) = 4/6 \text{ olur.}$$

Buradan çıkarılabilecek anlam, X kümesinin, U içerisinde kaba olarak tanımlanabilir olduğudur.

2.2.5. Niteliklerin Bağımsızlığı

Nitelik kümesinin bağımsız olup olmadığını kontrol etmek amacıyla, her nitelik için bu niteliğin silinmesinin bilgi sistemindeki birincil kümelerin sayısını arttırıp arttırmadığı kontrol edilir. Eğer $Ind(A) = Ind(A - a_i)$ ise a_i niteliği gereksiz olarak isimlendirilir. Aksi halde, a_i niteliği A kümesinde zorunlu bir niteliktir.

Örnek 2.5.

Tablo 2.2 ele alınsın. Eğer üç nitelik (a_1, a_2 ve a_3) hesaba katılırsa, beş birincil küme elde edilebilir (Tablo 2.3). Tablo 2.5. niteliklerin birinin hariç tutulması durumunda elde edilecek, birincil kümelerin sayısını verir. Örneğin, sadece a_2 ve a_3 kullanılmışsa, beş birincil küme elde edilebilir. Eğer a_1 ve a_3 kullanılmışsa, birincil küme sayısı 4 olur.

Eğer a_2 veya a_3 niteliği silinirse, birincil kümelerin sayısı daha küçük olur, ancak a_1 niteliğinin silinmesi birincil küme sayısını değiştirmez. a_2 ve a_3 nitelikleri zorunlu iken, a_1 niteliği gereksizdir.

Tablo 2.5. Niteliklerin birinin hariç tutulması durumunda birincil küme sayıları

	silinen nitelik			
	hiçbiri	a_1	a_2	a_3
birincil küme sayısı	5	5	4	4

2.2.6. Niteliklerin Özü ve İndirgemeleri

Öz ve indirgeme kavramları, kaba küme yaklaşımının iki temel kavramıdır. İndirgeme, orjinal bilgi sistemi yoluyla tüm ayırtedilebilir nesneleri ayırteden, bilgi sisteminin zorunlu bir parçasıdır. Öz, tüm indirgemelerin ortak parçalarıdır. İndirgeme ve özü hesaplamak için, ayırtedilebilirlik matrisi kullanılır. Ayırtedilebilirlik matrisi, n birincil küme sayısı olmak üzere nxn boyutlu bir matristir ve ayırtedilebilirlik matrisinin elemanları $[X]_i$ ve $[X]_j$ birincil kümeleri ayırteden, tüm niteliklerin kümesi olarak tanımlanır.

Örnek 2.6.

Ayırtedilebilirlik matrisi D, Tablo 2.3' de gösterilmiş beş eleman için Tablo 2.6'da oluşturulmuştur. d_{ij} elemanını hesaplamak için, ilk olarak i. ve j. birincil kümeleri ayırteden niteliklerin kümesi bulunmalıdır. Birincil küme 1 ve 2'yi ayırteden nitelik kümesi a_1, a_2 ve a_3 niteliklerini içerir yani $d_{21} = d_{12} = \{a_1, a_2, a_3\}$, $d_{31} = d_{13} = \{a_2\}$ yani birincil küme 3 ve 1'i ayırteden sadece a_2 niteliğidir. Ayırtedilebilirlik matrisi simetrik olduğundan (yani $d_{ij} = d_{ji}$) sadece alt diagonal parçayı almak yeterlidir. Şüphesiz ki her birincil küme, sonuçta elde edilen birincil kümeden en az bir nitelikle de olsa farklıdır, bundan dolayı ayırtedilebilirlik matrisinde boş hücre yoktur.

Ayırtedilebilirlik matrisi, A niteliklerinin tüm kümelerinde verilerin aynı parçalarını gösteren niteliklerin minimal kümesini (indirgemeler) bulmakta kullanılabilir. Bunu yapmak için, ayırtedilebilirlik fonksiyonu $f(A)$ adı verilen yapının elde edilmesi gerekir. Bu fonksiyon, aşağıdaki yolla inşa edilen bir boolean fonksiyondur. İki birincil kümeyi ayırteden nitelik kümelerindeki niteliklerin her biri için (örneğin a_1, a_2, a_3) a boolean değişkeni atandığında, sonuçtaki boolean fonksiyonu $(a_1 + a_2 + a_3)$ formunda elde edilir. $a_1 \vee a_2 \vee a_3$ şeklinde de gösterilebilir. Eğer nitelik kümesi boş ise, boolean sabit olarak 1 değeri atanır. Ayırtedilebilirlik matrisi, aşağıdaki Tablo 2.6'da gösterilmiştir.

Tablo 2.6. Örnek bir ayırtedilebilirlik matrisi

	Küme 1	Küme2	Küme3	Küme 4	Küme 5
Küme1					
Küme2	a_1, a_2, a_3				
Küme3	a_2	a_1, a_3			
Küme4	a_1, a_3	a_1, a_2, a_3	a_1, a_2, a_3		
Küme5	a_1, a_3	a_1, a_2, a_3	a_1, a_2, a_3	a_3	

Ayırtedilebilirlik fonksiyonu da aşağıdaki gibi elde edilir.

$$\begin{aligned} f(A) &= (a_1 + a_2 + a_3) a_2 (a_1 + a_3) (a_1 + a_3) \\ &\quad \times (a_1 + a_3) (a_1 + a_2 + a_3) (a_1 + a_2 + a_3) \\ &\quad \times (a_1 + a_2 + a_3) (a_1 + a_2 + a_3) a_3. \end{aligned}$$

$f(A)$ 'nın son şeklini elde etmek için eleme kurallarına başvurulur. Eleme kurallarına göre, birincil küme 1, a_1, a_2 ve a_3 niteliklerinden dolayı birincil küme 2'den ve a_2 niteliğinden dolayı birincil küme 3'den farklı ise 2 ve 3 kümelerinin her ikisini de ayırt ettiği için sadece a_2 niteliğini hesaba katmak yeterlidir. Yani;

$$(a_1 + a_2 + a_3)a_2 = a_2$$

Yine birincil küme 1'in, 2, 3, 4 ve 5 nolu birincil kümelerden ayırtedileceği varsayılın. Bu durumda, aşağıdaki nitelik kümelerini hesaba katmak gerekir. (Tablo 2.6'daki ilk sütun)

$$\{a_2\}, \{a_1, a_2, a_3\}, \{a_1, a_3\} \text{ ve } \{a_1, a_3\}$$

Sonrasında, ayırtedilebilirlik fonksiyonu aşağıdaki gibi ifade edilir.

$$\begin{aligned} a_2(a_1 + a_2 + a_3)(a_1 + a_3)(a_1 + a_3) &= a_2(a_1 + a_3) \\ &= a_2 a_1 + a_2 a_3 \end{aligned}$$

Pratikte, öz ve indirgemelerin hesaplanması çok basittir. Öz, ayırtedilebilirlik matrisinin tüm tek elemanlı girişlerinin kümesi iken indirgeme, ayırtedilebilirlik matrisinde boş olmayan her giriş ile en az bir ortak elemana sahip olan niteliklerin minimal alt kümesidir.

Örnek 2.6 için ayırtedilebilirlik matrisinden elde edilen sonuç

$$f(A) = a_2 a_3$$

olur. Bu yolla, niteliklerin dikkate alınan kümesinden bir indirgeme bulunur. Genellikle, bilgi sisteminin alternatif yollarla gösterimine izin verildiğinde birden fazla indirgeme vardır. Bir tane indirgeme bulunduğu için sistem Tablo 2.7' deki gibi gösterilebilir.

Tablo2.7. İndirgeme sonrası elde edilen birincil küme

U/R	a_2	a_3
$\{x_1, x_3, x_9\}$	1	3
$\{x_2, x_7, x_{10}\}$	2	1
$\{x_4\}$	2	3
$\{x_5, x_8\}$	1	4
$\{x_6\}$	1	2

Bu gösterim, Tablo 2.3'e eşdeğerdir. a_2 ve a_3 nitelikleri, a_1, a_2 ve a_3 nitelik kümesindeki gibi U evreni hakkında aynı bilgiyi sağlar.

Daha önce tanımlanan öz, ayırtedilebilirlik matrisinde meydana gelen tek niteliklerin tümünün kümesidir. Öz, a_2 ve a_3 niteliklerinin kümesidir. Bu durum, belirli durumlarda indirgemenin, öze eşit olabileceği anlamına gelir. Bu aşamada, aynı sonuçlar ortaya çıktığından, gereksiz nitelikleri eleme ve öz ve indirgemeleri belirlemede bazı zorluklar yaşanabilir.

2.2.7. Nitelik Değerlerinin Öz ve İndirgemeleri

Bilgi sisteminin basitleştirilmesi, sistem için gereksiz olan niteliklerin kesin değerlerinin iptal edilmesiyle elde edilir. Mesela, bu değerlerin bazılarının elenmesi ile sistemdeki tüm birincil kümeler ayırtedilebilir. Nitelik değerlerinin öz ve indirgemelerini bulma işlemi, niteliklerin öz ve indirgemelerini bulma işlemiyle benzerdir. Tüm hesaplamalar ayırtedilebilirlik matrisi temelli olarak gerçekleştirilir, ancak ayırtedilebilirlik fonksiyonunun tanımı burada biraz farklıdır. Ayırtedilebilirlik fonksiyonunun yerine, bilgi sisteminde birden fazla birincil küme olduğu için, çok sayıda ayırtedilebilirlik fonksiyonu yapılandırılır.

Örnek 2.7

Tablo 2.7'yi temel alarak, $\{a_2, a_3\}$ nitelikler evrenindeki beş birincil küme için beş ayırtedilebilirlik fonksiyonu $f_1(A), f_2(A), \dots, f_5(A)$ şeklinde yapılandırılır. $f_1(A)$ fonksiyonu 2, 3, 4, 5 kümelerinden, birincil küme 1'i ayırteden nitelikler kümesi ile ilgilidir. $f_2(A)$ fonksiyonu 1, 3, 4, 5 kümelerinden birincil küme 2'yi ayırteden nitelikler kümesi ile ilgilidir. Tablo 2.7'nin ayırtedilebilirlik matrisi, bahsedilen fonksiyonların hızlı bir biçimde yapılandırılmasına uygun bir formda, Tablo 2.8'de gösterilmiştir.

Tablo 2.8. Ayırtedilebilirlik fonksiyonlarının elde edileceği ayırtedilebilirlik matrisi

	Küme1	Küme2	Küme3	Küme4	Küme5
Küme1		a_2, a_3	a_2	a_3	a_3
Küme2	a_2, a_3		a_3	a_2, a_3	a_2, a_3
Küme3	a_2	a_3		a_2, a_3	a_2, a_3
Küme4	a_3	a_2, a_3	a_2, a_3		a_3
Küme5	a_3	a_2, a_3	a_2, a_3	a_3	

$f_1(A)$ 'yı oluşturmak için, Tablo 2.8.'in birinci sütunundaki tüm nitelik kümeleri hesaba katılır. Benzer şekilde, $f_2(A)$ 'yı oluşturmak için de, Tablo 2.8.'in ikinci sütunundaki tüm nitelik kümelerini hesaba katmak gerekir.

$$\begin{aligned} f_1(A) &= (a_2 + a_3)a_2a_3a_3 = a_2a_3 \\ f_2(A) &= (a_2 + a_3)a_3(a_2 + a_3)(a_2 + a_3) = a_3 \\ f_3(A) &= a_2a_3(a_2 + a_3)(a_2 + a_3) = a_2a_3 \\ f_4(A) &= a_3(a_2 + a_3)(a_2 + a_3)a_3 = a_3 \\ f_5(A) &= a_3(a_2 + a_3)(a_2 + a_3)(a_2 + a_3)a_3 = a_3 \end{aligned}$$

$f_1(A)$ nitelik değerlerinin indirgemesi olarak bir tane indirgeme ile sonuçlanmıştır(a_2a_3). Bu indirgeme a_2 ve a_3 niteliklerinin değerlerine dikkat edilmesini önerir. $f_2(A)$ ifadesinden de a_2 nitelik değerinin bilinmesinin gerekli olmadığı sonucuna varılabilir. Sonuç olarak Tablo 2.7 ,Tablo 2.9'da verildiği biçimde basitleştirilebilir. Bu tabloda, '*' gösterimi, önemsiz anlamında kullanılır.

Tablo 2.9. İndirgemeler sonucunda elde edilen birincil küme

U/R	a_2	a_3
$\{x_1, x_3, x_9\}$	1	3
$\{x_2, x_7, x_{10}\}$	*	1
$\{x_4\}$	2	3
$\{x_5, x_8\}$	*	4
$\{x_6\}$	*	2

2.2.8.Sınıflandırma

$X_i \subset U$ olmak üzere $F=\{x_1, x_2, \dots, x_n\}$, U evreninin altkümelerinin ailesi olsun. F 'deki altkümeler örtüşürse, yani $X_i \cap X_j = \emptyset$ ise ve bunların tümü birincil kümelerin hepsini içerirse, diğer bir deyişle $\cup X_i = U$ ise F , U 'nun bir sınıflandırması, dolayısı ile X_i 'ler de sınıflar olarak isimlendirilir.

F'in $B \subseteq A$ 'daki düşük ve yüksek yaklaşımları sırasıyla aşağıdaki gibi tanımlanır.

$$\underline{B}(F) = \{\underline{B}(X_1), \underline{B}(X_2), \dots, \underline{B}(X_n)\} \quad (2.7)$$

$$\overline{B}(F) = \{\overline{B}(X_1), \overline{B}(X_2), \dots, \overline{B}(X_n)\} \quad (2.8)$$

Sınıflandırma kalitesi de aşağıdaki gibi tanımlanır.

$$\eta_B F = \cup card \underline{B}(X_i) / card U \quad (2.9)$$

B'de, F sınıflandırmasının doğruluk değeri aşağıdaki formüle göre hesaplanabilir.

$$\beta_B F = \cup card \underline{B}(X_i) / \cup card \overline{B}(X_i) \quad (2.10)$$

Örnek 2. 8

Tablo 2.2.'deki veriler kullanılarak, aşağıdaki üç nesne kümesinin sınıflandırılacağı varsayalım.

$$X_1 = \{x_1, x_3, x_9\}$$

$$X_2 = \{x_4, x_5, x_6\}$$

$$X_3 = \{x_2, x_7, x_8, x_{10}\}$$

Her sınıfın düşük ve yüksek yaklaşımı ve sınıflandırmaların doğruluk oranları Tablo 2.10'da gösterilmiştir.

Tablo 2.10. Sınıflara ait düşük yaklaşım, yüksek yaklaşım ve doğruluk oranı değerleri.

sınıf no	nesne no	düşük yaklaşım	yüksek yaklaşım	doğruluk oranı
1	3	3	3	1.0
2	3	2	4	0.5
3	4	3	5	0.6

Tablo 2.10'daki son sütun ayrı ayrı her sınıf için hesaplanmış doğruluk oranlarını içerir. Örneğin, 2 numaralı sınıfın doğruluk oranı 0.5'dir. Çünkü düşük yaklaşımında 2, yüksek yaklaşımında 4 nesne bulunur ($2/4=0.5$).

1 nolu sınıf, sistem içerisinde tanımlanabilir, bununla beraber 2 ve 3 sınıfları sırasıyla 0.5 ve 0.6 doğruluk oranları ile kaba olarak tanımlanabilir.

Tüm sınıflandırmanın doğruluk oranı:

$$\eta_B F = \cup \text{card } \underline{B}(X_i) / \cup \text{card } \bar{B}(X_i) \\ = (3 + 2 + 3) / (3 + 4 + 5) = 0.67$$

Tüm sınıflandırmanın kalitesi:

$$\eta_B F = \cup \text{card } \underline{B}(X_i) / \text{card } U \\ = (3 + 2 + 3) / (3 + 3 + 4) = 0.8.$$

2.3. Karar Tabloları

Nitelik kümesi ve karar nitelikleri kümesi içeren bilgi gösterim sistemleri ‘karar tablosu’ olarak isimlendirilir. Karar tabloları sınıflandırma içinde yararlıdır.

Örnek 2. 9.

Tablo 2.11’de yer alan karar tablosu üç şart niteliği $\{a_1, a_2, a_3\}$ ve bir karar niteliği $\{d\}$ içerir. Karar niteliği $\{d\}$, aşağıdaki üç sınıf için 10 nesnenin aitliğini tanımlar.

sınıf 1: $\{x_1, x_3, x_9\}$,
sınıf 2: $\{x_2, x_4, x_7, x_{10}\}$,
sınıf 3: $\{x_5, x_6, x_8\}$.

Tablo 2.11. Karar tablosu örneği

U	a_1	a_2	a_3	d
x_1	2	1	3	1
x_2	3	2	1	2
x_3	2	1	3	1
x_4	2	2	3	2
x_5	1	1	4	3
x_6	1	1	2	3
x_7	3	2	1	2
x_8	1	1	4	3
x_9	2	1	3	1
x_{10}	3	2	1	2

Karar tablosu, çok sayıda karar niteliği içerebilir. Örneğin, Tablo 2.12’de üç şart niteliği $A=\{ a_1, a_2, a_3 \}$ ve iki karar niteliği $D=\{ d_1, d_2 \}$ vardır.

Tablo 2.12. İki karar niteliği içeren karar tablosu örneği

U	a_1	a_2	a_3	d_1	d_2
x_1	2	1	3	2	3
x_2	3	2	1	3	1
x_3	2	1	3	2	3
x_4	2	2	3	3	1
x_5	1	1	4	1	3
x_6	1	1	2	1	3
x_7	3	2	1	3	1
x_8	1	1	4	1	3
x_9	2	1	3	2	3
x_{10}	3	2	1	3	1

2.3.1. D-Lüzumsuz Nitelikler

B niteliklerinin şart kümesine ait a_i niteliği ($B \subseteq A$ olmak üzere), D’nin düşük yaklaşımı üzerinde etkili değilse ‘D-lüzumsuz’ nitelik olarak ifade edilir. Yani;

$$POS_B(D) = POS_{(B-a_i)}(D) \quad (2.11)$$

Aksi durumda, nitelik a_i , A da ‘D-zorunlu’ nitelik olarak ifade edilir.

2.3.2. Niteliklerin Bağlı Öz ve Bağlı İndirgemeleri

A’da tüm D-zorunlu niteliklerinin kümesi, A’nın D-öz değeri olarak isimlendirilirken, niteliklerin tüm kümeleri tarafından $IND(D)$ ayırtebilebilir ilişkinin tüm eşitlik sınıflarını ayırteden şart niteliklerinin minimal kümesi, D-indirgeme olarak isimlendirilir.

Bağlı indirgemeler, az miktarda değişmiş ayırtebilebilirlik matrisi kullanılarak hesaplanabilir. A’nın D-ayırtebilebilirlik matrisinin bir elemanı, $IND(D)$ ilişkisinin aynı eşitlik sınıfına ait olmayan X_i ve X_j nesnelerini ayırteden tüm niteliklerin kümesi olarak tanımlanır. D-öz, A’nın ayırtebilebilirlik matrisinin tüm tek elemanlarının kümesidir.

2.4. Karar Tablosu Analizinin Temel Adımları

- 1.D'deki birincil kümelerin inşası
- 2.D'deki birincil kümelerin yüksek ve düşük yaklaşımlarının hesaplanması
- 3.A niteliklerinin D-öz ve D-indirgemelerinin bulunması
- 4.A nitelik değerlerinin D-öz ve D-indirgemelerinin bulunması

Örnek 2.9.

Tablo 2.12 ve D'de ilk olarak aşağıdaki kümeler göz önüne alınsın.

sınıf 1: $\{x_1, x_3, x_9\}$,
sınıf 2: $\{x_2, x_4, x_7, x_{10}\}$,
sınıf 3: $\{x_5, x_6, x_8\}$.

İki şart niteliği d_1 ve d_2 , D'deki birincil kümelerde tanımlanmış yeni şart niteliği d ile geçici olarak yer değiştirebilir.

d karar niteliği, birincil kümeye ait tüm nesneler için 1 değerine, birincil küme 2'ye ait olan nesneler için 2 değerine sahiptir. Tablo 2.13 sınıflandırma problemini gösteren Tablo 2.12 ile aynıdır. Bunun anlamı, sonraki analizin tüm basamakları her iki karar tablosu için aynı olacaktır.

Tablo 2.13. Kümeye aitlikleri doğrultusunda elde edilen karar tablosu

U	a_1	a_2	a_3	d
x_1	2	1	3	1
x_2	3	2	1	2
x_3	2	1	3	1
x_4	2	2	3	2
x_5	1	1	4	3
x_6	1	1	2	3
x_7	3	2	1	2
x_8	1	1	4	3
x_9	2	1	3	1
x_{10}	3	2	1	2

Tablo 2.14'de sınıflandırmalara ait yaklaşım ve doğruluk oranı değerleri verilmiştir. Bu tabloya göre tüm sınıfların $A=\{a_1, a_2, a_3\}$ şart niteliklerinde tam anlamıyla tanımlanabilir.

Tüm sınıflandırmaların doğruluk oranı ve onların kaliteleri 1.0'e eşittir.

A niteliklerinin, D-indirgeme ve D-öz değerlerini bulmak için D-ayırte dılebilirlik matrisini ve D'deki farklı gruplardan nesneleri ayırteden elemanları oluşturmak gerekir.

Tablo 2.14. Sınıflandırmalara ait yaklaşım ve doğruluk oranı değerleri

sınıf no	nesne no	düşük yaklaşım	yüksek yaklaşım	doğruluk oranı
1	3	3	3	1.0
2	4	4	4	1.0
3	3	3	3	1.0

X_1 nesnesi, D uzayındaki X_3 ve X_9 nesneleri ile aynı sınıfta görüldüğünden, bu üç nesneyi ayırteden nitelik kümesi ile ilgilenilmez. Bu sebepten dolayı, Tablo 2.15'deki ilk sütunda 1, 3 ve 9 nolu satırlar dikkate alınmaz.

Tablo 2.15. Ayırte dılebilirlik matrisi

	1	2	3	4	5	6	7	8	9	10
1	–									
2	a_1, a_2, a_3	–								
3	–	a_1, a_2, a_3	–							
4	a_2	–	a_2	–						
5	a_1, a_3	a_1, a_2, a_3	a_1, a_3	a_1, a_2, a_3	–					
6	a_1, a_3	a_1, a_2, a_3	a_1, a_3	a_1, a_2, a_3	–	–				
7	a_1, a_2, a_3	–	a_1, a_2, a_3	–	a_1, a_2, a_3	a_1, a_2, a_3	–			
8	a_1, a_3	a_1, a_2, a_3	a_1, a_3	a_1, a_2, a_3	–	–	a_1, a_2, a_3	–		
9	–	a_1, a_2, a_3	–	a_2	a_1, a_3	a_1, a_3	a_1, a_2, a_3	a_1, a_3	–	
10	a_1, a_2, a_3	–	a_1, a_2, a_3	–	a_1, a_2, a_3	a_1, a_2, a_3	–	a_1, a_2, a_3	a_1, a_2, a_3	–

$f_A(D)$ ayırtedilebilirlik fonksiyonu aşağıdaki forma sahiptir.

$$\begin{aligned}
f_A(D) &= (a_1 + a_2 + a_3) a_2 (a_1 + a_3) (a_1 + a_3) \\
&\quad \times (a_1 + a_2 + a_3) (a_1 + a_3) (a_1 + a_2 + a_3) \\
&\quad \times (a_1 + a_2 + a_3) (a_1 + a_2 + a_3) \\
&\quad \times (a_1 + a_2 + a_3) (a_1 + a_2 + a_3) \\
&\quad \times (a_1 + a_2 + a_3) a_2 (a_1 + a_3) (a_1 + a_3) \\
&\quad \times (a_1 + a_2 + a_3) (a_1 + a_3) (a_1 + a_2 + a_3) \\
&\quad \times (a_1 + a_2 + a_3) (a_1 + a_2 + a_3) \\
&\quad \times (a_1 + a_2 + a_3) a_2 (a_1 + a_3) \\
&\quad \times (a_1 + a_2 + a_3) (a_1 + a_3) (a_1 + a_2 + a_3) \\
&\quad \times (a_1 + a_2 + a_3) (a_1 + a_3) (a_1 + a_2 + a_3) \\
&\quad \times (a_1 + a_2 + a_3) (a_1 + a_2 + a_3) (a_1 + a_3) \\
&\quad \times (a_1 + a_2 + a_3) (a_1 + a_2 + a_3) \\
&= a_2 (a_1 + a_3) = a_1 a_2 + a_2 a_3.
\end{aligned}$$

$\{a_1, a_2\}$ ve $\{a_2, a_3\}$ olmak üzere iki tane D-indirgeme vardır ve D-öz $\{a_2\}$ 'ye eşittir. Bunun anlamı, karar Tablosu 2.12'nin indirgenebilir ve Tablo 2.16 ve 2.17'deki gibi iki alternatif yolla gösterilir olduğudur.

Tablo 2.16. $\{a_1, a_2\}$ indirgemesi için Tablo 2.12'den elde edilen indirgenmiş karar tablosu

U	a_1	a_2	d
x_1	2	1	1
x_2	3	2	2
x_3	2	1	1
x_4	2	2	2
x_5	1	1	3
x_6	1	1	3
x_7	3	2	2
x_8	1	1	3
x_9	2	1	1
x_{10}	3	2	2

Tablo 2.17. $\{a_2, a_3\}$ indirgemesi için Tablo 2.12'den elde edilen indirgenmiş karar tablosu

U	a_2	a_3	d
x_1	1	3	1
x_2	2	1	2
x_3	1	3	1
x_4	2	3	2
x_5	1	4	3
x_6	1	2	3
x_7	2	1	2
x_8	1	4	3
x_9	1	3	1
x_{10}	2	1	2

Karar tablosunda şart niteliklerinin gereksiz değerlerinin elenmesi işlemiyle de ilgilenilir. Bunu yapabilmek $\{a_1, a_2\}$ veya $\{a_2, a_3\}$ alt uzayları için D-ayırtebilirlik matrisi temelli olarak, nitelik değerlerinin bağıl indirgeme ve bağıl öz değerlerinin hesaplanmasına bağlıdır. Tablo 2.18, $\{a_1, a_2\}$ indirgemeleri için D-ayırtebilirlik matrisini gösterir.

Tablo 2.18. $\{a_1, a_2\}$ indirgemesi için D-ayırtebilirlik matrisi

	1	2	3	4	5	6	7	8	9	10
1	—	a_1, a_2	—	a_2	a_1	a_1	a_1, a_2	a_1	—	a_1, a_2
2	a_1, a_2	—	a_1, a_2	—	a_1, a_2	a_1, a_2	—	a_1, a_2	a_1, a_2	—
3	—	a_1, a_2	—	a_2	a_1	a_1	a_1, a_2	a_1	—	a_1, a_2
4	a_2	—	a_2	—	a_1, a_2	a_1, a_2	—	a_1, a_2	a_2	—
5	a_1	a_1, a_2	a_1	a_1, a_2	—	—	a_1, a_2	—	a_1	a_1, a_2
6	a_1	a_1, a_2	a_1	a_1, a_2	—	—	a_1, a_2	—	a_1	a_1, a_2
7	a_1, a_2	—	a_1, a_2	—	a_1, a_2	a_1, a_2	—	a_1, a_2	a_1, a_2	—
8	a_1	a_1, a_2	a_1	a_1, a_2	—	—	a_1, a_2	—	a_1	a_1, a_2
9	—	a_1, a_2	—	a_2	a_1	a_1	a_1, a_2	a_1	—	a_1, a_2
10	a_1, a_2	—	a_1, a_2	—	a_1, a_2	a_1, a_2	—	a_1, a_2	a_1, a_2	—

Bağıl ayırtedilebilirlik fonksiyonları aşağıdaki gibi ifade edilir.

$$\begin{aligned}
f_1(D) &= (a_1 + a_2) a_2 a_1 a_1 (a_1 + a_2) a_1 (a_1 + a_2) \\
&= a_1 a_2 \\
f_2(D) &= (a_1 + a_2) (a_1 + a_2) (a_1 + a_2) (a_1 + a_2) \\
&\quad \times (a_1 + a_2) (a_1 + a_2) = a_1 + a_2 \\
f_3(D) &= (a_1 + a_2) a_2 a_1 a_1 (a_1 + a_2) a_1 (a_1 + a_2) \\
&= a_1 a_2 \\
f_4(D) &= a_2 a_2 (a_1 + a_2) (a_1 + a_2) (a_1 + a_2) a_2 = a_2 \\
f_5(D) &= a_1 (a_1 + a_2) a_1 (a_1 + a_2) (a_1 + a_2) \\
&\quad \times a_1 (a_1 + a_2) = a_1 \\
f_6(D) &= a_1 (a_1 + a_2) a_1 (a_1 + a_2) (a_1 + a_2) \\
&\quad \times a_1 (a_1 + a_2) = a_1 \\
f_7(D) &= (a_1 + a_2) (a_1 + a_2) (a_1 + a_2) (a_1 + a_2) \\
&\quad \times (a_1 + a_2) (a_1 + a_2) = a_1 + a_2 \\
f_8(D) &= a_1 (a_1 + a_2) a_1 (a_1 + a_2) (a_1 + a_2) \\
&\quad \times a_1 (a_1 + a_2) = a_1 \\
f_9(D) &= (a_1 + a_2) a_2 a_1 a_1 (a_1 + a_2) a_1 (a_1 + a_2) \\
&= a_1 a_2 \\
f_{10}(D) &= (a_1 + a_2) (a_1 + a_2) (a_1 + a_2) (a_1 + a_2) \\
&\quad \times (a_1 + a_2) (a_1 + a_2) = a_1 + a_2.
\end{aligned}$$

$\{a_1, a_2\}$ alt uzayında karar tablosunun son şekli Tablo 2.19'daki gibi gösterilebilir. Bu tabloda, ‘*’ gösterimi, “önemsiz” anlamındadır.

Tablo 2.19. $\{a_1, a_2\}$ alt uzayında karar tablosunun son şekli

U	a_1	a_2	d
x_1	2	1	1
x_2	*	2	2
x_3	2	1	1
x_4	*	2	2
x_5	1	*	3
x_6	1	*	3
x_7	*	2	2
x_8	1	*	3
x_9	2	1	1
x_{10}	*	2	2

2.4.1. Karar Kuralları

Yukarıda tanımlanan karar tablosu (2.12)' deki formda karar kuralları kümesi olarak ele alınabilir.

$$a_{ki} \Rightarrow d_j \quad (2.12)$$

Bu ifadede a_{ki} 'nin anlamı a_k niteliğinin i değerine sahiptir ve $\Rightarrow$ sembolü önermesel uzantıdır. Karar kuralında $\theta \Rightarrow \Phi$, θ ve Φ formülleri sırasıyla şart ve karar olarak isimlendirilir. Nitelik kümelerinin minimizasyonu ve başka nitelik kümeleri ile ilgili nitelik değerleri, basitçe verilerden karar kurallarının üretimi olarak bilinen, karar kurallarında gereksiz şartların indirgenmesi anlamındadır.

Örnek 2.10

Aşağıdaki kuralların kümesi Tablo 2.19'dan elde edilebilir.

$$a_{1_2} a_{2_1} \Rightarrow d_1$$

$$a_{2_2} \Rightarrow d_2$$

$$a_{1_1} \Rightarrow d_3$$

d, D içerisindeki birincil kümeleri tanımlar. Tablo 2.11 için, karar niteliğinin aldığı 1, 2 veya 3 değeri, sınıf 1, 2 veya 3'e olan aitliğini ifade eder. Tablo 2.12 için, 1, 2, 3 değerleri D'deki birincil kümeleri ifade eder ve aşağıdaki kurallar $\{d_1, d_2\}$ karar niteliği uzayında gösterilmiştir.

$$a_{1_2} a_{2_1} \Rightarrow d_{1_2} d_{2_3}$$

$$a_{2_2} \Rightarrow d_{1_3} d_{2_1}$$

$$a_{1_1} \Rightarrow d_{1_1} d_{2_3}$$

2.5. Bulanık Küme Teorisi

2.5.1. Giriş

Mühendislikte ve diğer bilim dallarında olaylar ve sistemler, kesin matematiksel modeller kullanılarak tanımlanır. Oluşturulan bu modellerin kullanılması ile olayın veya sistemin gelecekte alacağı durum veya göstereceği davranış biçimi tahmin edilmeye çalışılır. Günlük yaşantıda karşılaşılan problemlerin büyük bir çoğunluğu ya çeşitli nedenlerden dolayı tam olarak modellenemeyebilir ya da kesin bir durumu ifade edemeyebilirler. Bu tip problemlerin incelenmesinde ve çözümlenmesinde Bulanık Mantık yaklaşımı kullanılabilir.

Klasik mantığın aksine Bulanık Mantık, tamdan ziyade yaklaşık olan muhakeme modelleri ile uğraşan kavram ve tekniklerin esasını sağlamada yardımcı olmaktadır. Bulanık Mantık, doğada istatistiki olarak kesin olmayan ve belirsiz/şüpheli kaynaklar ile uğraşan bir tekniktir. Bulanık mantık'ın esası, bulanık küme teorisidir [13].

İlk defa Zadeh tarafından geliştirilen Bulanık Küme Teorisi, esas olarak insan düşünce ve algılarındaki belirsizlikle ilgilenir ve bu belirsizliği sayısallaştırmaya çalışır [14]. Diğer bir deyişle, bulanık mantığın amacı "insanların tam ve kesin olmayan bilgiler ışığında tutarlı ve doğru kararlar vermelerini sağlayan düşünme ve karar verme mekanizmalarının modellenmesi" olarak ifade edilebilir.

Bulanık Mantık düşüncesine uygun düşen modelleme problemleriyle karşılaşıldığında, genellikle bir uzman kişinin bilgi ve deneyimlerinden yararlanma yoluna gidilir. Uzman operatör; dilsel değişkenler/niteleyiciler olarak tanımlanabilen “uygun, çok uygun değil, yüksek, biraz yüksek, fazla, çok fazla” gibi günlük hayatımızda sıkça kullandığımız kelimeler doğrultusunda esnek bir denetim mekanizması geliştirir. İşte bulanık denetim bu tür mantıksal ilişkiler üzerine kurulmuştur [15].

Başlangıçta sadece teorik bir araştırma alanı olarak ortaya çıkmış olmasına rağmen, sonraki yıllarda pek çok farklı alanda uygulama imkanı bulmuştur. Bilgisayar bilimleri, kontrol, meteoroloji, tıp, yapay zeka ve uzman sistemler, sosyal bilimler, psikoloji ve yönetim bilimleri bu uygulama alanlarına örnek verilebilir.

2.5.2. Bulanık Küme

X, elemanları x biçiminde gösterilen ve belirli elemanlarla oluşturulmuş bir evrensel küme olsun. $X=\{x\}$ şeklinde ifade edilebilir. A, X kümesi ile ilgili olarak tanımlanmış bir kesin küme olsun. Bu kümeye ait karakteristik fonksiyon $\mu_A(x)$ ile gösterilsin. Bu fonksiyon, X kümesine ait x elemanın, A kümesinin üyesi olup olmadığının tespiti yoluyla kümeye olan aitliğini belirler. μ_A , 2.13'deki gibi tanımlanır.

$$\mu_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \quad (2.13)$$

Diğer bir deyişle, A kümesine ait karakteristik fonksiyon aşağıdaki gibi gösterilebilir.

$$\mu_A : X \longrightarrow \{0,1\} \quad (2.14)$$

Örneğin , “uzun boylu olarak nitelendirilen bir kimsenin boyu kaç olmalıdır ? ” sorusu için, tanım alanı $\{0,1\}$ olan bir üyelik fonksiyonunun ne kadar uygun olduğu sorgulansın. Uzun boylu olma durumu, kişinin çocuk ya da yetişkin olmasına bağlı olarak farklı anlamlar taşır. Elbette bazı genellemeler yapılabilir. Örneğin 1.80 m uzun boylu ya da 1.35 m uzun boylu değil gibi. Ancak bu durumda, 1.45 m boyundaki bir kişiyi sınıflandırmak oldukça zordur. Bundan dolayı, insanlar için kullanılan yaşlı, genç, kısa, şişman v.b. kavramların gösterimi bu metot da benzerdir ve ayırım yapmak oldukça zordur.

Bu sorunu aşmak için üyelik değerleri kullanılır. Üyelik değerlerini kullanan bu türlü bir metotta, kümeye kesin olarak ait olmama durumu 0 ile kümeye kesin olarak ait olma durumu 1 ile ifade edilir. $[0,1]$ aralığında üyelik değerlerine sahip bu metotta, üyelik değerlerinin elde edilmesi matematiksel olarak da oldukça kolaydır. Bu nedenle, evrensel küme gerçek sayılardan oluşmuş ise, bulanık küme biçiminde gösterimi, kesin küme ile gösterimine göre çok daha uygun olacaktır.

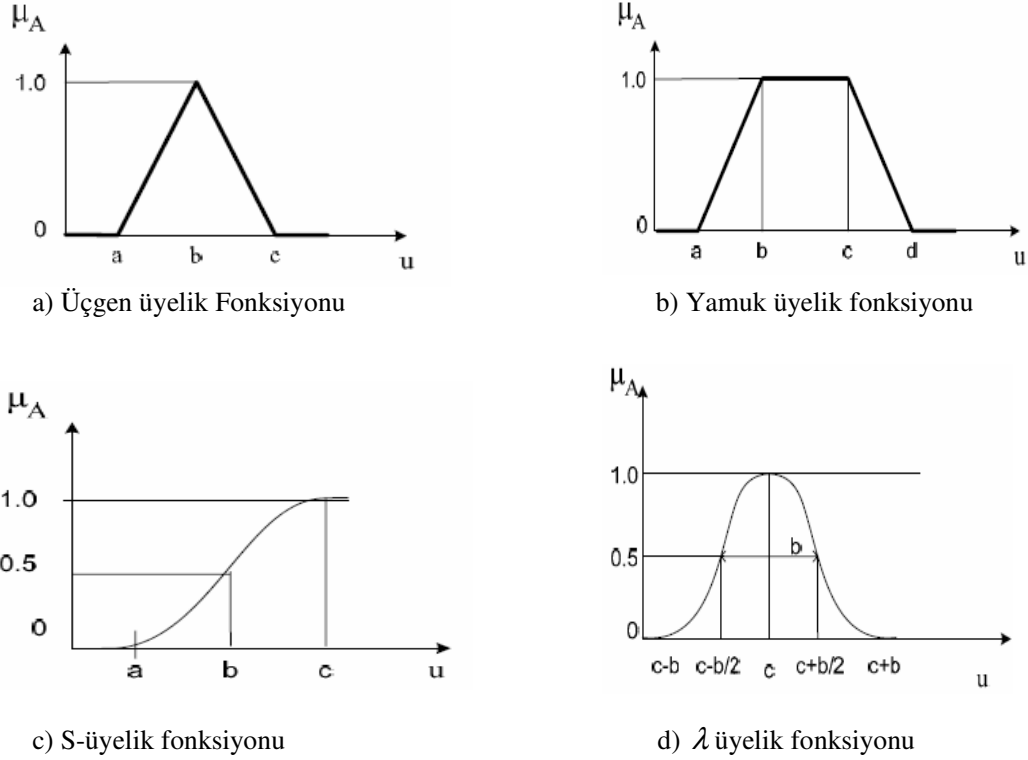
Karakteristik fonksiyonun aralığı $\{0,1\}$ çiftinden, $I=[0,1]$ aralığına genişletilmesi yoluyla bulanık küme teorisinin, kesin kümelerin bir genellemesi olduğu düşünülebilir. Bu işlem, üyelik fonksiyonu olarak isimlendirilen ve X 'e ait elemanların bu belirlenmiş aralıktaki değerini elde etmemizi sağlayan bir fonksiyon tarafından gerçekleştirilir. $\mu_A(x)$ ifadesi, $x \in X$ olmak üzere A kümesindeki x 'in üyelik derecesini ifade eder. Buna göre; X evrensel küme olmak üzere, X 'in A gibi bir bulanık kümesi, aşağıdaki gibi bir üyelik fonksiyonu ile ifade edilebilir:

$$\mu_A : X \longrightarrow [0,1] \quad (2.15)$$

2.5.3. Üyelik Fonksiyonları

X 'in her bir A bulanık kümesi, $x \in X$ olmak üzere her bir x için $\mu_A(x)$ üyelik derecesinin belirlenmesi yoluyla tanımlanabilir. Pratikte, bu sadece evrensel küme sonlu sayıda eleman içeriyorsa mümkündür. X kümesinin eleman sayısı oldukça büyük ise $\mu_A(x)$ en iyi üyelik fonksiyonu olarak bilinen bir yaklaşım parametre fonksiyonu ile gösterilir. Bu fonksiyon, kullanıcının verilen probleme göre parametreleri düzenlemesine izin verir.

Bulanık kümeleri göstermek için standart fonksiyonlar kullanılır. Sık kullanılan üyelik fonksiyonlarına Şekil 2.2’deki fonksiyonlar örnek verilebilir.



Şekil 2.2. Standart üyelik fonksiyonu çeşitleri

2.5.4. Küme İşlemleri

Belirsizliklerin sistematik işlenmesinde üyelik fonksiyonlarıyla gerçekleştirilen bulanık küme işlemleri büyük kolaylık sağlarlar. A ve B, X evrensel kümesinde sırasıyla μ_A ve μ_B üyelik fonksiyonuna sahip iki küme olsunlar.

- **Birleşim:** İki bulanık kümenin birleşimi aşağıdaki üyelik fonksiyonu ile gösterilir.

$$\forall x \in X, \mu_{A \cup B}(x) = \max\{\mu_A(x), \mu_B(x)\} \quad (2.16)$$

- **Kesişim:** İki bulanık kümenin kesişimi aşağıdaki üyelik fonksiyonu ile gösterilir.

$$\forall x \in X, \mu_{A \cap B}(x) = \min\{\mu_A(x), \mu_B(x)\} \quad (2.17)$$

- **Tümleyen:** Bulanık A kümesinin tümleyen bulanık kümesi aynı evrensel kümede aşağıdaki üyelik fonksiyonu ile tanımlanır.

$$\forall x \in X, \mu_{A^-}(x) = 1 - \mu_A(x) \quad (2.18)$$

3. EKSİK NİTELİK DEĞERİ İÇEREN VERİ KÜMELERİ ÜZERİNDE KABA KÜME TEORİSİ

3.1. Giriş

Kaba küme teorisinin tüm fikirleri, karar tablosu kullanılarak keşfedilmiştir. Karar tabloları, nitelik değerlerini ve kararları kullanarak durumları tanımlar. Karar bağımlı bir değişken iken nitelikler bağımsız değişkenlerdir. Kaba küme teorisi ile ilgili yayınların büyük çoğunluğunda bilgilerin tam olduğu varsayılır, yani tüm durumlar için tüm nitelik değerleri ve karar değerleri belirlenmiştir. Bu tür karar tablolarının tam olarak tanımlanmış olduğu söylenir. Ancak pratikte, karar tablosu olarak gösterilen giriş değerleri, eksik nitelik ve karar değerlerine sahip olabilir. Asıl amaç, örneklerden bilgi çıkarmak olduğu için, eksik karar değeri içeren bir örnek (yani sınıflandırılmamış) yararsızdır. Bu yüzden, sadece nitelik değerlerinin eksik olabileceği varsayılır.

Bir nitelik değerinin eksik olmasının iki temel sebebi vardır; bunlar değer kayıptır (örneğin silinmiştir) veya değer önemli değildir. İlk durumda, nitelik değeri yararlıdır ancak erişilemez. Sonraki durumda, değer önemli değildir, bu türlü değerler “önemsiz” olarak isimlendirilir.

Veri madenciliğinde, eksik nitelik değerleri ile ilgili iki temel strateji kullanılır. İlk strateji, eksik veri kümesini, tam veri kümesine dönüştürme temellidir. Bu stratejide, eksik nitelik kümelerinin, tam nitelik değerlerine dönüşümü, veri madenciliğinin temel süreci için bir önışlemedir. Diğer stratejide, bilgi bazı nitelik değerlerinin eksik olduğu veri kümelerinden elde edilir. Orjinal veri kümesi, tam veri kümesine dönüştürülmez. Bu çalışmada bu stratejiye uyan yaklaşım kavramları ile çalışılmıştır.

İlk strateji için tipik bazı uygulamalar aşağıda verilmiştir.

- Eksik nitelik değerinin niteliğın en genel nitelik değeri ile değıştirilmesi.
- Sayısal nitelikler için, eksik nitelik değerinin, niteliğın ortalama değeri ile değıştirilmesi.
- Niteliğın olası tüm değerlerinin atanması. Eksik nitelik değeri içeren bir veri kümesinin, eksik nitelik değerlerinin tüm olası nitelik değerleriyle değıştirilmesi yoluyla elde edilen yeni veri kümesi ile değıştirilmesi.
- Eksik nitelik değeri içeren durumların ihmal edilmesi. Eksik nitelik değeri içeren orjinal veri kümesi, varolan eksik nitelik değerlerinin silindiğı, yeni bir veri kümesi ile değıştirilir.
- Eksik nitelik değerleri özel değerler olarak muhafaza edilir.

3.2. Eksik Nitelik Değerleri ve Karakteristik İlişkiler

Bu bölümde, durumlar hakkındaki bilginin karar tablosu tarafından gösterildiği varsayılacaktır. Tablonun satırları durumlar tarafından ve sütunlar değişkenler tarafından etiketlenir. Tüm durumların kümesi U ile gösterilir. Bağımsız değişkenler ‘nitelik’ olarak isimlendirilir. Bağımlı nitelikler ise ‘karar’ olarak isimlendirilir ve d ile gösterilir. Tüm niteliklerin kümesi A ile ifade edilir. Karar tabloları ile ilgili bir örnek Tablo 3.1’de görülmektedir.

Tablo 3.1. Tam olarak tanımlanmış karar tablosuna bir örnek

	Nitelikler			Karar
	Konum	Bodrum	Şömine	Değer
1	iyi	evet	evet	yüksek
2	kötü	hayır	hayır	düşük
3	iyi	hayır	hayır	orta
4	kötü	evet	hayır	orta
5	iyi	hayır	evet	orta

Her karar tablosu ρ gibi bir fonksiyon olarak tanımlanır ki bu tüm nitelik kümesi içerisinde sıralanmış (durum, nitelik) çifti kümesini oluşturur. Örneğin, $\rho(1, \text{Konum}) = \text{iyi}$.

Kaba küme teorisi ayırteğilemez ilişki düşüncesini temel alır. B , tüm niteliklerin kümesi A ’nın boş olmayan bir alt kümesi olsun. U üzerinde, ayırteğilemez ilişki $\text{IND}(B)$, aşağıdaki gibi tanımlanır.

$$x, y \in U \text{ için } (x, y) \in \text{IND}(B) \text{ şayet } \rho(x, a) = \rho(y, a) \text{ tüm } a \in B \text{ için} \quad (3.1)$$

Tam olarak tanımlanmış karar tablosu için ayırteğilemez ilişki $\text{IND}(B)$ bir eşitlik ilişkisidir. $\text{IND}(B)$ ’nin eşitlik sınıfı, B ’nin birincil kümesi olarak isimlendirilir.

Örneğin, Tablo 3.1 için, $\text{IND}(\{\text{Konum}, \text{Bodrum}\})$ için birincil kümeler $\{1\}$, $\{2\}$, $\{3, 5\}$ ve $\{4\}$ biçimindedir.

Tablo 3.1.’de ρ fonksiyonu tam olarak tanımlanmıştır. Uygulamada, veri madenciliği için giriş verisi sıklıkla eksik nitelik değerleri içerebilir. Diğer bir deyişle, ρ fonksiyonu tam olarak tanımlanmamıştır.

Tezin devam eden bölümlerinde, tüm karar niteliklerinin tanımlanmış olduğu yani eksik olmadığı varsayılmış ve tüm eksik nitelik değerleri, kayıp veriler için “?”, önemsiz nitelik değerleri için “*” sembolleriyle gösterilmiştir.

Tam olarak tanımlanmamış tablolar, ayırtedilemezlik ilişkisi yerine karakteristik ilişkiler tarafından tanımlanır. Tam olarak tanımlanmamış tabloya bir örnek, Tablo 3.2’de gösterilmiştir. Buradaki eksik değerler kayıp nitelik türündedir.

Tablo 3.2. Tüm eksik niteliklerin ‘kayıp’ türde olduğu eksik bir karar tablosu örneği

	Nitelikler			Karar
	Konum	Bodrum	Şömine	Değer
1	iyi	evet	evet	yüksek
2	kötü	?	hayır	düşük
3	iyi	hayır	?	orta
4	kötü	evet	hayır	orta
5	?	?	evet	orta

Karar tabloları için, tüm eksik nitelik değerlerinin kayıp olduğu durumda, özel karakteristik bir ilişki Stefanowski ve Tsoukias tarafından tanımlanmıştır[16-18]. Bu bölümde tanımlanan, bu karakteristik ilişki $LV(B)$ olarak gösterilmiştir ve burada B kümesi, tüm niteliklerin kümesi A ’nın boş olmayan bir alt kümesini göstermektedir. $x, y \in U$ için karakteristik ilişki $LV(B)$ aşağıdaki gibi tanımlanır.

$$(x, y) \in LV(B) \text{ şayet } \rho(x, a) = \rho(y, a) \text{ ise tüm } a \in B \text{ için öyle ki } \rho(x, a) \neq ? \quad (3.2)$$

Her bir x durumu için, karakteristik ilişki $LV(B)$, $IB(x)$ karakteristik kümesi tarafından gösterilebilir.

$$IB(x) = \{y \mid (x, y) \in LV(B)\} \quad (3.3)$$

Tablo 3.2 için karakteristik küme $IA(x)$, $x \in U$ olmak üzere aşağıdaki gibidir.

$$IA(1) = \{1\}$$

$$IA(2) = \{2, 4\}$$

$$IA(3) = \{3\}$$

$$IA(4) = \{4\}$$

$$IA(5) = \{1, 5\}$$

Karar tablosuna başka bir örnek, tüm eksik nitelik değerlerinin önemsiz olduğu durumdur ve Tablo 3.3’ de gösterilmiştir.

Tablo 3.3. Tüm eksik niteliklerin “önemsiz ”olarak tanımlandığı karar tablosu.

	Nitelikler			Karar
	Konum	Bodrum	Şömine	Değer
1	iyi	evet	evet	yüksek
2	kötü	*	hayır	düşük
3	iyi	hayır	*	orta
4	kötü	evet	hayır	orta
5	*	*	evet	orta

Tüm eksik değerlerin önemsiz türde olduğu karar tabloları için, DCC(B) olarak gösterilen özel karakteristik ilişki M. Kryszkiewicz tarafından tanımlanmıştır [19,20]. $x, y \in U$ için karakteristik ilişki DCC(B) aşağıdaki gibi tanımlanır:

$$(x, y) \in DCC(B) \text{ ancak ve ancak } \rho(x, a) = \rho(y, a) \text{ ya da } \rho(x, a) = * \text{ ya da } \rho(y, a) = * \text{ tüm } a \in B \text{ için} \quad (3.4)$$

Benzer biçimde, bir x durumu için karakteristik ilişki DCC(B), karakteristik küme $JB(x)$ olarak gösterilebilir.

$$JB(x) = \{y \mid (x, y) \in DCC(B)\} \quad (3.5)$$

Tablo 3.3 için $x \in U$ olmak üzere, karakteristik küme $JA(x)$ aşağıdaki gibidir:

$$JA(1) = \{1, 5\}$$

$$JA(2) = \{2, 4\}$$

$$JA(3) = \{3, 5\}$$

$$JA(4) = \{2, 4\}$$

$$JA(5) = \{1, 3, 5\}$$

Tablo 3.4. Eksik niteliklerin bir kısmının kayıp, bir kısmının önemsiz olarak tanımlandığı karar tablosu

	Nitelikler			Karar
	Konum	Bodrum	Şömine	Değer
1	iyi	evet	evet	yüksek
2	kötü	?	hayır	düşük
3	iyi	hayır	?	orta
4	kötü	evet	hayır	orta
5	*	*	evet	orta

Tablo 3.4 daha genel bir durumu gösterir. Karar tablosunda, her iki türde eksik veri bulunur. Benzer şekilde, her iki türde eksik nitelik içeren karar tablosu için, U kümesi üzerinde $R(B)$ karakteristik ilişkisi aşağıdaki gibi tanımlanır.

$$(x, y) \in R(B) \text{ ancak ve ancak } \rho(x, a) = \rho(y, a) \text{ ya da } \rho(x, a) = * \text{ ya da } \rho(y, a) = * \text{ tüm } a \in B \text{ için öyle ki } \rho(x, a) \neq ? \quad (3.6)$$

$x, y \in U$ ve B 'de A tüm nitelikler kümesinin boş olmayan bir alt kümesi olmak üzere, $R(B)$ karakteristik ilişkisi $KB(x)$ karakteristik kümesi olarak da gösterilebilir.

$$KB(x) = \{y \mid (x, y) \in R(B)\} \quad (3.7)$$

Tablo 3.4 için $x \in U$ olmak üzere, $KA(x)$ karakteristik kümeleri aşağıdaki gibidir.

$$KA(1) = \{1, 5\}$$

$$KA(2) = \{2, 4\}$$

$$KA(3) = \{3, 5\}$$

$$KA(4) = \{4\}$$

$$KA(5) = \{1, 5\}$$

$LV(B)$ ve $DCC(B)$ karakteristik ilişkileri, $R(B)$ karakteristik ilişkisinin özel durumlarıdır. Tam olarak tanımlanmış bir karar tablosu için, $R(B)$ karakteristik ilişkisi $IND(B)$ 'ye indirgenir.

3.3. Yüksek ve Düşük Yaklaşım

Tam olarak tanımlanmış karar tabloları için düşük ve yüksek yaklaşımlar, ayırtedilemez ilişki temel alınarak tanımlanır. Birincil kümenin her sonlu birleşimi, B-tanımlanabilir küme olarak isimlendirilir. $IND(B)$ 'nin eşitlik sınıfı $[x]B$ olarak gösterilir.

X , tüm durumlar kümesi U 'nun herhangi bir alt kümesi olsun. X kümesi kavram olarak isimlendirilir ve genellikle kararın belirli bir değeri tarafından tanımlanan, tüm durumların kümesi olarak tanımlanır. X , B-tanımlanabilir küme değildir. Bundan dolayı, X kümesine, iki B-tanımlanabilir küme tarafından yaklaşılabılır. Bunlardan ilki, X kümesinin B-düşük yaklaşımı olarak isimlendirilir.

Düşük yaklaşım $\underline{B}X$ ile gösterilir ve aşağıdaki gibi tanımlanır.

$$\{x \in U \mid [x]B \subseteq X\} \quad (3.8)$$

İkinci küme, X 'in B-yüksek yaklaşımı olarak isimlendirilir. Yüksek yaklaşım $\overline{B}X$ olarak gösterilir ve aşağıdaki gibi tanımlanır.

$$\{x \in U \mid [x]B \cap X \neq \emptyset\} \quad (3.9)$$

Tam olarak tanımlanmamış karar tabloları için düşük ve yüksek yaklaşımlar, birkaç farklı yol ile tanımlanabilir. Yine, X kümesi U evrensel kümenin herhangi bir alt kümesi, B kümesi de tüm niteliklerin kümesi olan A 'nın alt kümesi ve $R(B)$ karakteristik küme $K(x)$ ile eksik tanımlanmış karar tablosunun karakteristik ilişkisi olsun. ($x \in U$)

İlk tanımlamaya göre düşük ve yüksek yaklaşımlar, U uzayında tatmin edici özellikteki taslakların (singleton) kümesidir. Bu tanım “taslak” olarak isimlendirilir. X 'in taslak B-düşük yaklaşımı

$$\underline{B}X = \{x \in U \mid K_B(x) \subseteq X\} \quad (3.10)$$

şeklinde tanımlanır. X 'in taslak B-yüksek yaklaşımı

$$\overline{B}X = \{x \in U \mid K_B(x) \cap X \neq \emptyset\} \quad (3.11)$$

olur. Tablo 3.2’ de gösterilen örnekte $B=A$ olarak kabul edilirse, $R(A) = LV(A)$ olur. Buna göre, A-yüksek ve A-düşük yaklaşımlar aşağıda gösterilmiştir.

$$\begin{aligned}\underline{A}\{1\} &= \{1\} \\ \underline{A}\{2\} &= \emptyset \\ \underline{A}\{3,4,5\} &= \{3,4\} \\ \overline{A}\{1\} &= \{1,5\} \\ \overline{A}\{2\} &= \{2\} \\ \overline{A}\{3,4,5\} &= \{2,3,4,5\}\end{aligned}$$

Tablo 3.3’de gösterilen örnekte $B=A$ olarak kabul edilirse, $R(A) = DCC(A)$ olur. Buna göre, A-yüksek ve A-düşük yaklaşımlar aşağıda gösterilmiştir.

$$\begin{aligned}\underline{A}\{1\} &= \emptyset \\ \underline{A}\{2\} &= \emptyset \\ \underline{A}\{3,4,5\} &= \{3\} \\ \overline{A}\{1\} &= \{1,5\} \\ \overline{A}\{2\} &= \{2,4\} \\ \overline{A}\{3,4,5\} &= \{1,2,3,4,5\} = U\end{aligned}$$

İkinci tanım farklı bir düşünce kullanır. Düşük ve yüksek yaklaşımlar, U ’ nun alt kümesi olan karakteristik kümelerin birleşimidir. Bu tanım “alt küme” olarak isimlendirilir. X ’in altküme B-düşük yaklaşımı 3.12’deki gibi tanımlanır.

$$\underline{B}X = \cup\{K_B(x) \mid x \in U, K_B(x) \subseteq X\} \quad (3.12)$$

X ’ in altküme B-yüksek yaklaşımı 3.13’deki gibi tanımlanır.

$$\overline{B}X = \cup\{K_B(x) \mid x \in U, K_B(x) \cap X \neq \emptyset\} \quad (3.13)$$

Tablo 3.2’de görülen karar tablosu örneğinde $R(A) = LV(A)$ ‘dır. Altküme A-düşük ve A-yüksek yaklaşımları, aşağıda gösterilmiştir.

$$\begin{aligned}\underline{A}\{1\} &= \{1\} \\ \underline{A}\{2\} &= \emptyset \\ \underline{A}\{3,4,5\} &= \{3,4\} \\ \overline{A}\{1\} &= \{1,5\} \\ \overline{A}\{2\} &= \{2,4\} \\ \overline{A}\{3,4,5\} &= \{1,2,3,4,5\} = U\end{aligned}$$

Tablo 3.3’de görülen karar tablosu örneğinde $R(A) = DCC(A)$ ‘dır. Altküme A-düşük ve A-yüksek yaklaşımları, aşağıda gösterilmiştir.

$$\begin{aligned}\underline{A}\{1\} &= \emptyset \\ \underline{A}\{2\} &= \emptyset \\ \underline{A}\{3,4,5\} &= \{3,5\} \\ \overline{A}\{1\} &= \{1,3,5\} \\ \overline{A}\{2\} &= \{2,4\} \\ \overline{A}\{3,4,5\} &= \{1,2,3,4,5\} = U\end{aligned}$$

Son yaklaşımda ise bir önceki tanımlamadaki U uzayı, kavram X ile yer değiştirmesiyle elde edilir. Kavram X ’ in, kavram B-düşük yaklaşımı aşağıdaki gibi tanımlanır.

$$\underline{B}X = \cup\{K_B(x) \mid x \in X, K_B(x) \subseteq X\} \quad (3.14)$$

Kavram X ’in, kavram B-yüksek yaklaşımı aşağıdaki gibi tanımlanır.

$$\overline{B}X = \cup\{K_B(x) \mid x \in X, K_B(x) \cap X \neq \emptyset\} \quad (3.15)$$

Tablo 3.2’de gösterilen karar örneğinde $R(A) = LV(A)$ ’dır. Kavram A-yüksek yaklaşımı aşağıda gösterilmiştir.

$$\begin{aligned}\overline{A}\{1\} &= \{1\} \\ \overline{A}\{2\} &= \{2,4\} \\ \overline{A}\{3,4,5\} &= \{1,3,4,5\}\end{aligned}$$

Tablo 3.3’de gösterilen karar örneğinde $R(A) = DCC(A)$ ’dır. Kavram A-yüksek yaklaşımı aşağıda gösterilmiştir.

$$\overline{A}\{1\} = \{1,5\}$$

$$\overline{A}\{2\} = \{2,4\}$$

$$\overline{A}\{3,4,5\} = \{1,2,3,4,5\} = U$$

Tam olarak tanımlanmış karar tabloları için bu üç tanım, aynı tanımda birleşir. Eksik tanımlanmış karar tabloları için bunun doğru olmadığı yukarıdaki örneklerde görülmüştür. Karakteristik ilişki $R(B)$ için, taslak düşük ve yüksek yaklaşımlar sırasıyla, altküme düşük ve yüksek yaklaşımlarının altkümesidir. Bundan başka, yüksek yaklaşımın kavram tanımı, yüksek yaklaşımın altküme tanımının altkümesidir. Böylece, yüksek yaklaşımın kavram tanımı, kural çıkarma için daha uygundur. Bunun nedeni, ‘yüksek yaklaşım kapsanan kavramların en küçük kümesi olmalıdır’ düşüncesidir. Bundan dolayı, taslak düşük ve yüksek yaklaşım için kabul edilen bazı özellikler genelde altküme ve kavram düşük ve yüksek yaklaşımları için uygun değildir.

4. KABA KÜME TEORİSİYLE EKSİK NİCEL VERİLERDEN BULANIK KURAL ÇIKARIMI

4.1. Giriş

Bu bölümde, nicel değerler içeren eksik bir veri kümesinden, kaba küme teorisi kullanılarak bulanık kuralların elde edilmesi işleminden bahsedilecektir. Burada, bulanık küme teorisi ve kaba küme teorisi bir arada kullanılmıştır. Bulanık eksik nicel değerli veri kümesinden kesin ve olası kuralların elde edilmesi ve eksik nitelik değerlerinin tahmini için bir algoritma önerilmiştir. Önerilen öğrenme algoritması ilk olarak üyelik fonksiyonu yardımıyla veri kümesini bulanık kümeye dönüştürür. Sonra, bulanık eksik düşük yaklaşımı hesaplar ve bu yaklaşım yardımıyla eksik değerleri tahmin etmeye çalışır. Kalan eksik değerlerde, bulanık eksik yüksek yaklaşımın hesaplanması yoluyla tahmin edilmeye çalışılır. Şu ana kadar yapılan çalışmalarda eksik veriler ya sadece önemsiz (*) ya da sadece kayıp (?) türdedir. Bu tez çalışması sırasında, hem önemsiz (*), hem de kayıp (?) türde eksik verileri birlikte içeren bir nicel değerli veri kümesi için yeni bir algoritma geliştirilmiş ve kesin ve olası kurallar çıkarılabilir hale getirilmiştir.

4.2. Önerilen Algoritma

Giriş: Her biri m tane nitelik değerli ve C sınıf kümesinden birine ait, n nesneli U eksik nicel değerli veri kümesi ve üyelik fonksiyonu.

Çıkış: Kesin ve olası bulanık kurallar kümesi

1.Adım: Sınıf etiketlerine göre nesneleri ayrık nesne alt kümelerine böler. Aynı C_l sınıfına ait olan nesnelerin her bir kümesi X_l olarak gösterilir.

2.Adım: $v^{(i)}$, her bir $Obj^{(i)}$ nesnesinin nicel değerini göstermek üzere her bir A_i nesnesi için $i=1$ 'den n 'e kadar, $j=1, \dots, m$ olmak üzere, f_j^i bulanık kümesi, verilen üyelik fonksiyonu kullanılarak,

$$\left(\frac{f_{j_1}^{(i)}}{R_{j_1}} + \frac{f_{j_2}^{(i)}}{R_{j_2}} + \dots + \frac{f_{j_l}^{(i)}}{R_{j_l}} \right) \quad (4.1)$$

şeklinde gösterilir. R_{jk} , A_j niteliğinin k'ncı bulanık alanı; f_j^i , R_{jk} bölgesinde $v_j^{(i)}$ 'nin bulanık üyelik değeri ve $l (=|A_j|)$, A_j için bulanık bölgenin sayısıdır. $Obj^{(i)}$, A_j için eksik ya da kayıp niteliğe sahipse nitelik değeri (*) ya da (?) olarak muhafaza edilir.

3.Adım: Niteliklerin bulanık eksik birincil kümeleri bulunur. Eğer $Obj^{(i)}$, A_j için kesin bulanık üyelik değerine $f_{jk}^{(i)}$ sahipse, $A_j = R_{jk}$ 'den bulanık eksik eşitlik sınıfı içine $(Obj^{(i)}, c)$ biçiminde; $Obj^{(i)}$, A_j için önemsiz değere (*) sahipse, her bir bulanık eksik eşitlik sınıfı içine $(Obj^{(i)}, u)$ biçiminde; $Obj^{(i)}$, A_j için kayıp değere (?) sahipse, aynı R_{jk} bulanık bölgede bulunan nesnelerin olduğu bulanık eksik eşitlik sınıfı içine $(Obj^{(i)}, l)$ olarak yerleştirilir.

$A_j = R_{jk}$, için bulanık eksik sınıfın $\mu_{A_{jk}}$ üyelik değeri şöyle hesaplanır:

$Obj^{(i)}$ kesin ve $f_{jk}^{(i)} \neq 0$ olmak üzere;

$$\mu_{A_{jk}} = \min f_{jk}^{(i)} \quad (4.2)$$

4.Adım: $q=1$ ile başlat. q , bulanık eksik düşük yaklaşımlar için o anki işletilen nitelik sayısını ifade etmek için kullanılır.

5.Adım: Her bir X_l sınıfı için q nitelikli, her bir B alt kümesinin bulanık eksik düşük yaklaşımı (4.3)'deki gibi hesaplanır.

$$B_*(x) = \{ (B_k(Obj^{(i)}), \mu_{B_k}(Obj^{(i)})) \mid 1 \leq i \leq n, Obj^{(i)} \in X_l, \\ B_k^c(Obj^{(i)}) \subseteq X_l, 1 \leq k \leq |B(Obj^{(i)})| \} \quad (4.3)$$

$B(Obj^{(i)})$, $Obj^{(i)}$ nesnesini içeren ve B nitelik altkümesinden türetilmiş bulanık eksik eşitlik sınıflarının kümesi, $B_k^c(Obj^{(i)})$, $B(Obj^{(i)})$ 'deki k'ncı bulanık eksik eşitlik sınıfının kesin bölümüdür.

6.Adım: Bulanık eksik düşük yaklaşımlardaki her bir kesin olmayan (önemsiz veya kayıp) $Obj^{(i)}$ örneği için aşağıdaki adımlar takip edilir:

(a) $Obj^{(i)}$ nesnesi, bulanık eksik düşük yaklaşımda B nitelik alt kümesinde k'ncı alan kombinasyonunun R_B^k , bulanık eksik eşitlik sınıflarından sadece birinde yer alıyorsa, $Obj^{(i)}$, nin kesin olmayan değeri (4.4)' deki gibi atanır:

$$\frac{\sum_{Obj^{(r)} \in B_k^c(Obj^{(r)})} v_j^{(r)} x f_{jk}^{(r)}}{\sum_{Obj^{(r)} \in B_k^c(Obj^{(r)})} f_{jk}^{(r)}} \quad (4.4)$$

$v_j^{(r)}$, A_j niteliği için $Obj^{(r)}$ 'nin nicel değeridir ve $f_{jk}^{(i)}$, R_B^k 'deki $v_j^{(r)}$ 'nin bulanık üyelik değeridir. Böylece, tahmin edilen $Obj^{(i)}$ değerleri, bulanık kümesinde değiştirilir. Bulanık eksik eşitlik sınıfında sıfır üyelik değerine sahip ($Obj^{(i)}$,u) ya da ($Obj^{(i)}$,l) silinir ve ($Obj^{(i)}$,c) ile değiştirilir. Minimum operasyonla, bunları içeren bulanık eksik eşitlik sınıflarının üyelik değerleri yeniden hesaplanır. Ayrıca, $Obj^{(i)}$ nesnesi üzerinde aynı işlemler yapılarak bulanık eksik düşük yaklaşımlar bulunur.

(b) $Obj^{(i)}$ nesnesi, bulanık eksik düşük yaklaşımda birden fazla bulanık eksik eşitlik sınıfında mevcutsa, birkaç nitelik onları belirleyinceye kadar eksik verinin tahmini ertelenir.

7.Adım: $q=q+1$ yapılır. $q > m$ oluncaya kadar Adım 5-7 tekrarlanır.

8.Adım: $Obj^{(i)}$ nesnesi, hala bulanık eksik düşük yaklaşımda birden fazla bulanık eksik eşitlik sınıfında mevcutsa, $Obj^{(i)}$ nesnesinin kesin olmayan değerinin tahmini için maksimum sayısal önemlilik ifadesini kullanır. Tahmin ve işleme Adım 6 (a)'da ki gibi yapılır.

9.Adım: Her bir altküme için bulanık eksik düşük yaklaşımlarından kesin bulanık kurallar türetilir ve gelecek veriler için etkililik ölçüsü, düşük yaklaşımdaki eşitlik sınıflarının üyelik değerlerinden belirlenir.

10.Adım: Kesin bulanık kurallardan, şart kısmı daha özelleşmiş olanlarını ve aynı kesin bulanık kurallardan etkililik ölçüsü eşit ve daha küçük olanlar silinir.

11.Adım: $q=1$ yapılır. q , bulanık eksik yüksek yaklaşımlar için o anki işletilen nitelik sayısını ifade etmek için kullanılır.

12.Adım: Her bir X_l sınıfı için q nitelikli, her bir B altkümesinin bulanık eksik yüksek yaklaşımı (4.5)'deki gibi hesaplanır:

$$B^*(x) = \{ (B_k(Obj^{(i)}), \mu_{B_k}(Obj^{(i)})) \mid 1 \leq i \leq n, B_k^c(Obj^{(i)}) \cap X_l \neq \emptyset, B_k^c(Obj^{(i)}) \not\subset X_l, 1 \leq k \leq |B(Obj^{(i)})| \} \quad (4.5)$$

$B(Obj^{(i)}), Obj^{(i)}$ nesnesini içeren ve B nitelik altkümesinden türetilmiş bulanık eksik eşitlik sınıflarının kümesi, $B_k^c(Obj^{(i)})$, $B(Obj^{(i)})$ deki k 'nci bulanık eksik eşitlik sınıfının kesin bölümüdür.

13.Adım: Bulanık eksik yüksek yaklaşımlardaki her bir kesin olmayan (önemsiz veya kayıp) $Obj^{(i)}$ örneği için aşağıdaki adımlar takip edilir.

(a) $Obj^{(i)}$ nesnesi, bulanık eksik yüksek yaklaşımda, B nitelik altkümesinde k 'nci alan kombinasyonunun R_B^k , bulanık eksik eşitlik sınıflarından sadece birinde yer alıyorsa, $Obj^{(i)}$, nin kesin olmayan değeri (4.6)'daki gibi atanır:

$$\frac{\sum_{Obj^{(r)} \in B_k^c(Obj^{(r)}) \& Obj^{(r)} \in X_l} v_j^{(r)} x f_{jk}^{(r)}}{\sum_{Obj^{(r)} \in B_k^c(Obj^{(r)}) \& Obj^{(r)} \in X_l} f_{jk}^{(r)}} \quad (4.6)$$

$v_j^{(r)}$, A_j niteliği için $Obj^{(r)}$ 'nin nicel değeridir ve $f_{jk}^{(i)}, R_B^k$ deki $v_j^{(r)}$ 'nin bulanık üyelik değeridir. Böylece, tahmin edilen $Obj^{(i)}$ değerleri bulanık kümesinde değiştirilir. Bulanık eksik eşitlik sınıfında sıfır üyelik değerine sahip $(Obj^{(i)}, u)$ ya da $(Obj^{(i)}, l)$ silinir ve $(Obj^{(i)}, c)$ ile değiştirilir. Minimum operasyonla, bunları içeren bulanık eksik eşitlik sınıflarının üyelik değerleri yeniden hesaplanır. Ayrıca, $Obj^{(i)}$ nesnesi üzerinde aynı işlemler yapılarak, bulanık eksik yüksek yaklaşımlar bulunur.

(b) $Obj^{(i)}$ nesnesi, bulanık eksik yüksek yaklaşımda birden fazla bulanık eksik eşitlik sınıfında mevcutsa, birkaç nitelik onları belirleyinceye kadar eksik verinin tahmini ertelenir.

14.Adım: $q=q+1$ yapılır. $q > m$ oluncaya kadar Adım 12-14 tekrarlanır.

15.Adım: Her bir X_l sınıfı için yüksek yaklaşımdaki her bir bulanık eksik eşitlik sınıfının uygunluk derecesi (4.7)'deki gibi hesaplanır.

$$P(B_k^c(Obj^{(i)})) = \frac{\sum_{Obj^{(r)} \in B_k^c(Obj^{(r)}) \& Obj^{(r)} \in X_l} f_{jk}^{(r)}}{\sum_{Obj^{(r)} \in B_k^c(Obj^{(r)})} f_{jk}^{(r)}} \quad (4.7)$$

16.Adım: $Obj^{(i)}$ nesnesi, hala bulanık eksik yüksek yaklaşımda, birden fazla bulanık eksik eşitlik sınıfında mevcutsa, $Obj^{(i)}$ nesnesinin kesin olmayan değerinin tahmini için maksimum uygunluk ölçüsü kullanır. Tahmin ve işleme Adım 13 (a)'daki gibi yapılır.

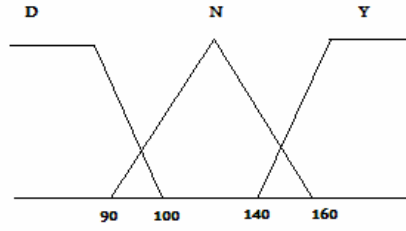
17.Adım: Tahmin edilmiş nesnelerin yeniden hesaplanmış uygunluk ölçüleri ile her bir B altkümesi için bulanık eksik yüksek yaklaşımlarından olası bulanık kurallar türetilir. Ayrıca, gelecek veriler için etkililik ölçüsü, yüksek yaklaşımdaki eşitlik sınıflarının üyelik değerlerinden belirlenir.

18.Adım: Olası bulanık kurallardan şart kısmı daha özelleşmiş olanlarını ve diğer kesin ve olası bulanık kurallarla aynı olan, etkililik ve uygunluk ölçüsü değerlerinin her ikisi de eşit veya daha küçük olanlarını silinir.

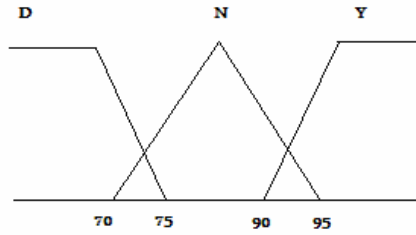
4.3. Algoritmanın Eksik Bir Veri Kümesine Uygulanışı

Tablo 4.1. Eksik veri kümesi

	A1	A2	K
1	80	?	N
2	155	70	Y
3	130	?	N
4	87	68	D
5	*	93	Y
6	140	100	Y
7	95	75	D
8	95	93	N
9	*	78	Y



Şekil 4.1. A1 niteliği için üyelik fonksiyonu



Şekil 4.2. A2 niteliği için üyelik fonksiyonu

Tablo 4.2. Bulanık veri kümesi

	A1	A2	K
1	1D	?	N
2	0,1N+0,75Y	1D	Y
3	0,85N	?	N
4	1D	1D	D
5	*	0,16 N+0,6Y	Y
6	0,57 N	1Y	Y
7	0,5D+0,14 N	0,4 N	D
8	0,5D+0,14 N	0,16 N+0,6Y	N
9	*	0,64 N	Y

Verilen veri kümesi için eşitlik sınıfları aşağıdaki gibi elde edilir:

$$XN=\{1, 3, 8\}$$

$$XD=\{4, 7\}$$

$$XY=\{2, 5, 6, 9\}$$

$$U/\{A1\} = \{ ((1,c)(4,c) (7,c)(8,c) (5,u)(9,u), 0.5) ((2,c)(3,c) (6,c)(7,c) (8,c)(5,u)(9,u), 0.1) ((2,c)(5,u)(9,u), 0.75) \}$$

A1 için düşük yaklaşım

$$A1_*(XN)=\emptyset$$

$$A1_*(XD)=\emptyset$$

$$A1_*(XY)=\{ ((2,c)(5,u)(9,u), 0.75) \}$$

5 ve 9 sadece XY için düşük yaklaşımda yer aldığından, Y sınıfına ait olduğu kabul edilir.

Değeri, $155 \cdot 0,75 / 0,75 = 155$ olarak bulunur.

Tablo 4.3. A1 niteliği için tahmin sonrası elde edilen veri kümesi

	A1	A2	K
1	80	?	N
2	155	70	Y
3	130	?	N
4	87	68	D
5	155	93	Y
6	140	100	Y
7	95	75	D
8	95	93	N
9	155	78	Y

Tablo 4.4. A1 niteliği için tahmin sonrası elde edilen bulanık kümeler

	A1	A2	K
1	1D	?	N
2	0,1N+0,75Y	1D	Y
3	0,85N	?	N
4	1D	1D	D
5	0,1N+0,75Y	0,16 N+0,6Y	Y
6	0,57N	1Y	Y
7	0,5D+0,14 N	0,4 N	D
8	0,5D+0,14 N	0,16 N+0,6Y	N
9	0,1N+0,75Y	0,64 N	Y

$$U/\{A1\} = \{ ((1,c)(4,c)(7,c)(8,c), 0.5) ((2,c)(3,c)(5,c)(6,c)(7,c)(8,c)(5,c)(9,c), 0.1) ((2,c)(5,c)(9,c), 0.75) \}$$

$$A1_*(XN)=\emptyset$$

$$A1_*(XD)=\emptyset$$

$$A1_*(XY)=\{ ((2,c)(5,c)(9,c), 0.75) \} \quad (\text{Kural çıkarılacak ifade})$$

$$U/\{A2\} = \{ ((2,c)(4,c), 1) ((5,c)(7,c)(8,c)(9,c), 0.16) ((5,c)(6,c)(8,c), 0.6) \}$$

$$A2_*(XN)=\emptyset$$

$$A2_*(XD)=\emptyset$$

$$A2_*(XY)=\emptyset$$

$U/\{A1,A2\}=\{((4,c)(1,l),1) ((2, c), 0.1) ((2,c), 0.75) ((6,c)(3,l), 0.57) ((4,c), 1) ((5,c) (7,c)(8,c) (9,c), 0.1) ((5,c) (6,c) (8,c), 0.1) ((5,c) (9,c), 0.16) ((5,c), 0.6) ((7,c) (8,c), 0.16) ((8,c), 0.5) \}$

$A1 A2_*(XN)= \{((8,c), 0.5) \}$

$A1 A2_*(XD)=\{ ((4,c)(1,l),1) ((4,c), 1) \}$

$A1 A2_*(XY)=\{ ((2c), 0.1) ((2,c),0.75) ((6,c)(3,l), 0.57) ((5,c) (9,c), 0.16) ((5,c) ,0.6) \}$

Birinci nesnedeki ? ifadesine ilişkin bilgi, sadece XD’ ye ait düşük yaklaşımda yer aldığı için D sınıfına ait olduğu kabul edilir. Nicel değeri $68*1/1 =68$ biçiminde elde edilir.
Üçüncü nesnedeki ? ifadesine ilişkin bilgi sadece XY’ ye ait düşük yaklaşımda yer aldığı için Y sınıfına ait olduğu kabul edilir. Nicel değeri $100*1/1 =100$ biçiminde elde edilir.

Tablo 4.5. A2 niteliği için tahmin sonrası elde edilen veri kümesi

	A1	A2	K
1	80	68	N
2	155	70	Y
3	130	100	N
4	87	68	D
5	155	93	Y
6	140	100	Y
7	95	75	D
8	95	93	N
9	155	78	Y

Tablo 4.6. A2 niteliği için tahmin sonrası elde edilen bulanık kümeler

	A1	A2	K
1	1D	1D	N
2	0,1N+0,75Y	1D	Y
3	0,85N	1Y	N
4	1D	1D	D
5	0,1N+0,75Y	0,16 N+0,6Y	Y
6	0,57 N	1Y	Y
7	0,5D+0,14 N	0,4 N	D
8	0,5D+0,14 N	0,16 N+0,6Y	N
9	0,1N+0,75Y	0,64 N	Y

$U/\{A1,A2\}=\{((4,c)(1,c), 1) ((7,c) (8,c), 0.16) ((8,c), 0.5) ((2,c), 0.1) ((5,c)(7,c)(8,c) (9,c), 0.1) ((3,c) (5,c) (6,c) (8,c), 0.1) ((2,c), 0.75) ((5,c) (9,c), 0.16) ((5,c), 0.6) \}$

$A1 \ A2 \ *(XN)= \{((8,c), 0.5) \}$

$A1 \ A2 \ *(XD)=\emptyset$

$A1 \ A2 \ *(XY)=\{ ((2,c), 0.1) ((2,c), 0.75) ((5,c) (9,c), 0.16) ((5,c), 0.6) \}$

Düşük yaklaşımlardan elde edilen kesin kurallar aşağıdaki gibidir:

1. $E\check{G}ER$ A1=Yüksek $\acute{I}SE$ K=Yüksek $f=0.75$
2. $E\check{G}ER$ A1=Düşük VE A2=Yüksek $\acute{I}SE$ K=Normal $f=0.5$
3. $E\check{G}ER$ A1=Normal VE A2=Düşük $\acute{I}SE$ K=Yüksek $f=0.1$
4. $E\check{G}ER$ A1=Yüksek VE A2=Düşük $\acute{I}SE$ K=Yüksek $f=0.75$
5. $E\check{G}ER$ A1=Yüksek VE A2=Normal $\acute{I}SE$ K=Yüksek $f=0.16$
6. $E\check{G}ER$ A1=Yüksek VE A2=Yüksek $\acute{I}SE$ K=Yüksek $f=0.6$

Sol tarafı daha özelleşmiş olan kurallar elenir. Buna göre, 4-5-6 numaralı kurallar, 1 numaralı kurala göre sol tarafı daha özelleşmiş olduğu için elenir.

Bu aşamadan sonra yüksek yaklaşımların hesaplanması işlemine geçilir:

$U/\{A1\} = \{ ((1,c)(4,c) (7,c)(8,c), 0.5) ((2,c)(3,c) (6,c)(7,c) (8,c)(5,c)(9,c), 0.1) ((2,c)(5,c)(9,c),0.75) \}$

$A1 \ *(XN)= \{ ((1,c)(4,c) (7,c)(8,c) ,0.5) ((2,c)(3,c) (6,c)(7,c) (8,c) (5,c)(9,c), 0.1) \}$

$A1 \ *(XD)= \{ ((1,c)(4,c) (7,c)(8,c) 0.5) ((2,c)(3,c) (6,c)(7,c) (8,c) (5,c)(9,c), 0.1) \}$

$A1 \ *(XY)=\{ ((2,c)(3,c) (6,c)(7,c) (8,c) (5,c)(9,c), 0.1) \}$

A1 yüksek yaklaşımdan elde edilen olası kurallar aşağıdaki gibidir.

1. $E\check{G}ER$ A1=Düşük $\acute{I}SE$ K=Normal $f=0.5$ $p=1,5/3=0.5$
2. $E\check{G}ER$ A1=Normal $\acute{I}SE$ K=Normal $f=0.1$ $p=(0.85+0.14)/2=0.495$
3. $E\check{G}ER$ A1=Düşük $\acute{I}SE$ K=Düşük $f=0.5$ $p=1,5/3=0.5$ (X)
4. $E\check{G}ER$ A1=Normal $\acute{I}SE$ K=Düşük $f=0.1$ $p=0.14/2=0,07$ (X)
5. $E\check{G}ER$ A1=Normal $\acute{I}SE$ K=Yüksek $f=0.1$ $p=0,87/2=0,435$ (X)

Elde edilen kurallardan sol tarafı eşit olanların uygunluk ve etkililik değerlerine bakılır. Bu değerlerin eşit veya daha küçük olduğu kurallar silinir. (Elenen kurallar (X) ile gösterilmiştir)

$$U/\{A2\} = \{ ((1,c) (2,c)(4,c), 1) ((5,c)(7,c)(8,c) (9,c), 0.16) ((3,c)(5,c)(6,c)(8,c), 0.6) \}$$

$$A2^*(XN) = \{ ((1,c) (2,c)(4,c), 1) ((5,c)(7,c)(8,c) (9,c), 0.16) ((3,c)(5,c)(6,c)(8,c), 0.6) \}$$

$$A2^*(XD) = \{ ((1,c) (2,c)(4,c), 1) ((5,c)(7,c)(8,c) (9,c), 0.16) \}$$

$$A2^*(XY) = \{ ((1,c) (2,c)(4,c), 1) ((5,c)(7,c)(8,c) (9,c), 0.16) ((3,c)(5,c)(6,c)(8,c), 0.6) \}$$

A2 niteliğinin yüksek yaklaşımından elde edilen olası kurallar aşağıdaki gibidir.

- | | | |
|--|--------|--------------------|
| 1. $E\check{G}ER$ A2=Düşük $\acute{I}SE$ K=Normal | f=1 | p=1/3 =0.34 |
| 2. $E\check{G}ER$ A2=Normal $\acute{I}SE$ K=Normal | f=0.16 | p=0.16/1,6 =0.1(X) |
| 3. $E\check{G}ER$ A2=Yüksek $\acute{I}SE$ K=Normal | f=0.6 | p=1.6/3.2 =0.5 |
| 4. $E\check{G}ER$ A2=Düşük $\acute{I}SE$ K=Düşük | f=1 | p=1/3 =0.34(X) |
| 5. $E\check{G}ER$ A2=Normal $\acute{I}SE$ K=Düşük | f=0.16 | p=0.4/1,6 =0.25(X) |
| 6. $E\check{G}ER$ A2=Düşük $\acute{I}SE$ K=Yüksek | f=1 | p=1/3 =0.34(X) |
| 7. $E\check{G}ER$ A2=Normal $\acute{I}SE$ K=Yüksek | f=0.16 | p=0.8/1,6 =0.5 |
| 8. $E\check{G}ER$ A2=Yüksek $\acute{I}SE$ K=Yüksek | f=0.6 | p=1.6/3.2 =0.5(X) |

$$U/\{A1,A2\} = \{ ((4,c)(1,c), 1) ((7,c) (8,c), 0.16) ((8,c), 0.5) ((2,c), 0.1) ((5,c)(7,c)(8,c) (9,c), 0.1) ((3,c) (5,c) (6,c) (8,c), 0.1) ((2,c), 0.75) ((5,c) (9,c), 0.16) ((5,c), 0.6) \}$$

$$A1 A2^*(XN) = \{ ((4,c)(1,c), 1) ((7,c) (8,c), 0.16) ((5,c)(7,c)(8,c) (9,c), 0.1) ((3,c) (5,c) (6,c) (8,c), 0.1) \}$$

$$A1 A2^*(XD) = \{ ((4,c)(1,c), 1) ((7,c) (8,c), 0.16) ((5,c)(7,c)(8,c) (9,c), 0.1) \}$$

$$A1 A2^*(XY) = \{ ((5,c)(7,c)(8,c)(9,c), 0.1) ((3,c) (5,c) (6,c) (8,c), 0.1) \}$$

A1 ve A2 nitelikleri için yüksek yaklaşımından elde edilen olası kurallar aşağıdaki gibidir.

- | | | |
|---|--------|---------------------|
| 1. $E\check{G}ER$ A1=Düşük VE A2=Düşük $\acute{I}SE$ K=Normal | f=1 | p=2/4=0.5 |
| 2. $E\check{G}ER$ A1=Düşük VE A2=Normal $\acute{I}SE$ K=Normal | f=0.16 | p=0.66/1.56=0.43(X) |
| 3. $E\check{G}ER$ A1=Normal VE A2=Normal $\acute{I}SE$ K=Normal | f=0.1 | p=0.3/1,84=0.16(X) |
| 4. $E\check{G}ER$ A1=Normal VE A2=Yüksek $\acute{I}SE$ K=Normal | f=0.1 | p=2.59/4.86=0.54 |

- | | | | |
|----|---|--------|---------------------|
| 5. | <i>EĞER</i> A1=Düşük <i>VE</i> A2=Düşük <i>İSE</i> K=Düşük | f=1 | p=2/4=0.5(X) |
| 6. | <i>EĞER</i> A1=Düşük <i>VE</i> A2=Normal <i>İSE</i> K=Düşük | f=0.16 | p=0.9/1.56=0.57 |
| 7. | <i>EĞER</i> A1=Normal <i>VE</i> A2=Normal <i>İSE</i> K=Düşük | f=0.1 | p=0.54/1.84=0.3(X) |
| 8. | <i>EĞER</i> A1=Normal <i>VE</i> A2=Normal <i>İSE</i> K=Yüksek | f=0.1 | p=1/1.84=0.54 |
| 9. | <i>EĞER</i> A1=Normal <i>VE</i> A2=Yüksek <i>İSE</i> K=Yüksek | f=0.1 | p=1.37/4.86=0.28(X) |

Sonuç olarak elde edilen kesin ve olası kurallar aşağıdaki gibidir.

Kesin kurallar

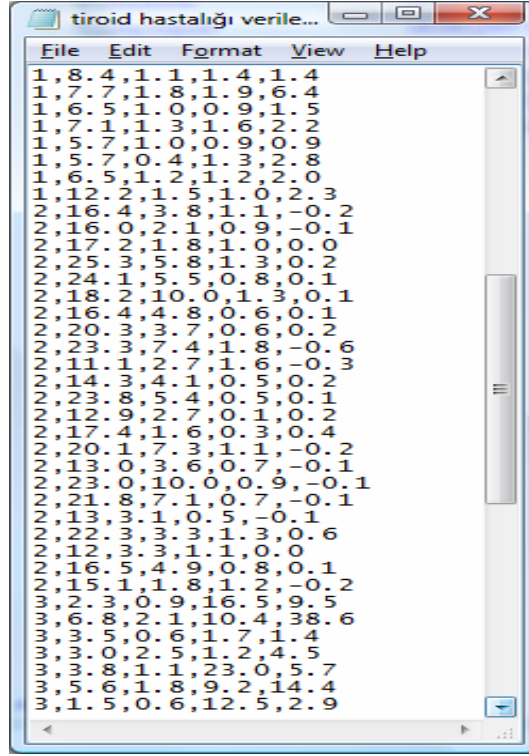
- | | | |
|----|--|--------|
| 1. | <i>EĞER</i> A1=Yüksek <i>İSE</i> K=Yüksek | f=0.75 |
| 2. | <i>EĞER</i> A1=Düşük <i>VE</i> A2=Yüksek <i>İSE</i> K=Normal | f=0.5 |
| 3. | <i>EĞER</i> A1=Normal <i>VE</i> A2=Düşük <i>İSE</i> K=Yüksek | f=0.1 |

Olası kurallar

- | | | | |
|----|---|--------|------------------|
| 1. | <i>EĞER</i> A1=Düşük <i>İSE</i> K=Normal | f=0.5 | p=1,5 /3 =0.5 |
| 2. | <i>EĞER</i> A1=Normal <i>İSE</i> K=Normal | f=0.1 | p=0.99 /2 =0.495 |
| 3. | <i>EĞER</i> A2=Düşük <i>İSE</i> K=Normal | f=1 | p=1/3 =0.34 |
| 4. | <i>EĞER</i> A2=Yüksek <i>İSE</i> K=Normal | f=0.6 | p=1.6/3.2 =0.5 |
| 5. | <i>EĞER</i> A2=Normal <i>İSE</i> K=Yüksek | f=0.16 | p=0.8/1,6 =0.5 |
| 6. | <i>EĞER</i> A1=Düşük <i>VE</i> A2=Düşük <i>İSE</i> K=Normal | f=1 | p=2/4=0.5 |
| 7. | <i>EĞER</i> A1=Normal <i>VE</i> A2=Yüksek <i>İSE</i> K=Normal | f=0.1 | p=2.59/4.86=0.54 |
| 8. | <i>EĞER</i> A1=Düşük <i>VE</i> A2=Normal <i>İSE</i> K=Düşük | f=0.16 | p=0.9/1.56=0.57 |
| 9. | <i>EĞER</i> A1=Normal <i>VE</i> A2=Normal <i>İSE</i> K=Yüksek | f=0.1 | p=1/1.84=0.54 |

5. UYGULAMA YAZILIMI

Bu tezde, önerilen algoritmanın uygulanabilirliği ve etkisini test etmek için bir yazılım geliştirilmiştir. Bu yazılım, <http://www.ics.uci.edu/~mlearn/MLRepository.html> sitesinden alınan tiroit hastalığı verilerini kullanmaktadır. Aşağıda uygulama yazılımının kullandığı tiroit verisi görülmektedir.



File	Edit	Format	View	Help
1	8.4	1.1	1.4	1.4
1	7.7	1.8	1.9	6.4
1	6.5	1.0	0.9	1.5
1	7.1	1.3	1.6	2.2
1	5.7	1.0	0.9	0.9
1	5.7	0.4	1.3	2.8
1	6.5	1.2	1.2	2.0
1	12.2	1.5	1.0	2.3
2	16.4	3.8	1.1	-0.2
2	16.0	2.1	0.9	-0.1
2	17.2	1.8	1.0	0.0
2	25.3	5.8	1.3	0.2
2	24.1	5.5	0.8	0.1
2	18.2	10.0	1.3	0.1
2	16.4	4.8	0.6	0.1
2	20.3	3.7	0.6	0.2
2	23.3	7.4	1.8	-0.6
2	11.1	2.7	1.6	-0.3
2	14.3	4.1	0.5	0.2
2	23.8	5.4	0.5	0.1
2	12.9	2.7	0.1	0.2
2	17.4	1.6	0.3	0.4
2	20.1	7.3	1.1	-0.2
2	13.0	3.6	0.7	-0.1
2	23.0	10.0	0.9	-0.1
2	21.8	7.1	0.7	-0.1
2	13.3	1.0	0.5	-0.1
2	22.3	3.3	1.3	0.6
2	12.3	3.3	1.1	0.0
2	16.5	4.9	0.8	0.1
2	15.1	1.8	1.2	-0.2
3	2.3	0.9	16.5	9.5
3	6.8	2.1	10.4	38.6
3	3.5	0.6	1.7	1.4
3	3.0	2.5	1.2	4.5
3	3.8	1.1	23.0	5.7
3	5.6	1.8	9.2	14.4
3	1.5	0.6	12.5	2.9

Şekil 5.1. Tiroit hastalığı verileri

Uygulama yazılımında 100 adet tiroit hastalığı verisiyle çalışılmıştır. Bu veri kümesi 1 karar niteliği ve 4 adet şart niteliğinden oluşmaktadır. Aşağıda her bir niteliğin ne anlama geldiği konusunda bilgi verilmiştir.

1: Sınıf niteliği (Karar niteliği)

1 = normal tiroit 2 = hiper tiroit 3 = hipo tiroit

2: İzotopik yerdeğiştirme metoduyla hesaplanan Toplam Serum Tiroksin miktarı (TST)

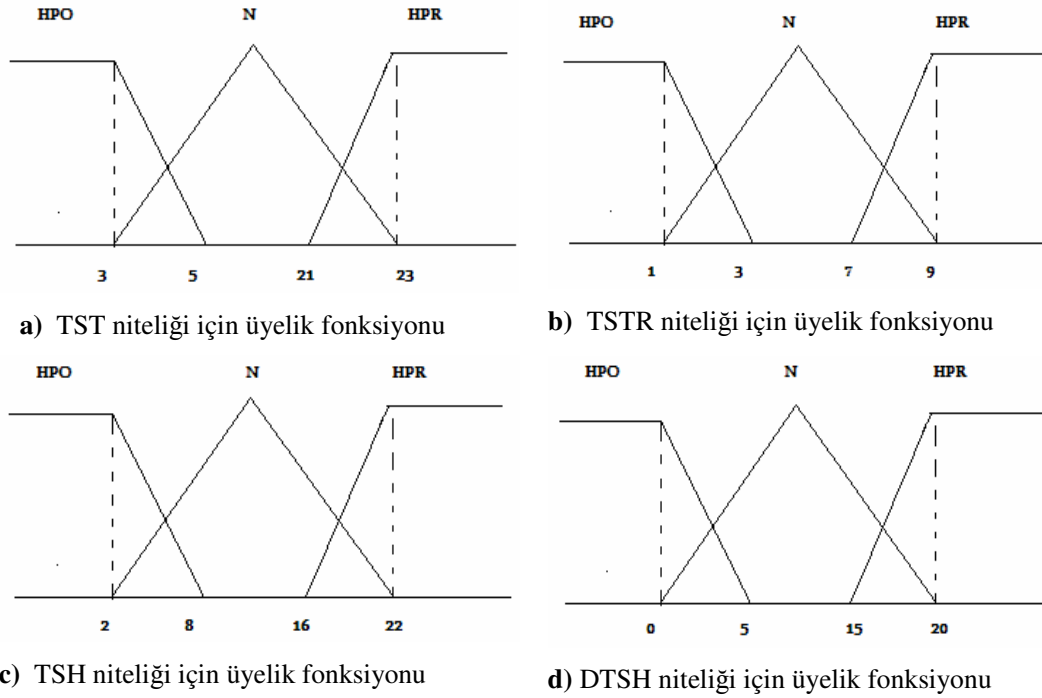
3: Radioimmün analiz ile ölçülen Toplam Serum Triodotironin miktarı (TSTR)

4: Radioimmün analiz ile ölçülen Bazal tiroit-uyarıcı hormon (TSH) miktarı

5: Temel değerle karşılaştırmak üzere 200 mikro gram tirotiropin salgılayıcı hormon enjekte edildikten sonra TSH'nın azami mutlak farklılık miktarı (DTSH)

Yukarıda da gösterildiği gibi uygulama yazılımında 2.nitelik TST, 3.nitelik TSTR, 4.nitelik TSH ve 5.nitelik DTSH sembolleriyle ifade edilmiştir.

4.2' deki algoritma eksik veriler üzerinde çalıştığı için öncelikle tiroit verileri üzerinde rasgele yerlere önemsiz ve kayıp nitelik değerleri yerleştirilmiştir. Sonra bu veriler aşağıdaki üyelik fonksiyonları doğrultusunda bulanıklaştırılmıştır. Bu üyelik fonksiyonları ifade edilirken sınıf nitelikleri olan Normal tiroit 'N', Hiper tiroit 'HPR' ve Hipo tiroit 'HPO' sembolleriyle ifade edilmiştir.



Şekil 5.2. Uygulama yazılımında kullanılan üyelik fonksiyonları

Yukarıdaki üyelik fonksiyonları doğrultusunda bulanıklaştırılan tiroit hastalığı verileri, 4.2' deki algoritma ile ele alınan nitelik sayısı her iterasyonda arttırılarak, önce birincil kümeler, ardından düşük ve yüksek yaklaşımlar elde edilmiştir. Elde edilen bu yaklaşımlar doğrultusunda, önemsiz ve kayıp niteliklerin tahmini yapılmıştır. Her iterasyonda elde edilen yaklaşımlar göz önüne alınarak, düşük yaklaşımlardan kesin kurallar, yüksek yaklaşımdan olası kurallar çıkarılmıştır. Tüm bu adımlar sonrasında, sol tarafı daha özelleşmiş olan kurallar ve sol tarafı aynı olan kuralların, etkililik ölçüsü ya da uygunluk derecesi eşit veya düşük olanları silinmiştir.

Bahsedilen tüm adımları gerçekleştiren uygulama yazılımının arayüzü Şekil 5.3’de gösterilmiştir.

Kaba Küme Yaklaşımı

Eksik Veri kümesi

	SINIF	TST	TSTR	TSH	DTSH
1	N	10,1	2,2	0,9	2,7
2	N	9,9	3,1	2	5,9
3	N	12,9	2,4	*	0,6
4	N	5,3	1,6	1,4	1,5
5	N	7,3	1,5	1,5	-0,1
6	N	6,1	2,1	1,4	7

Tahmin sonucu veri kümesi

	SINIF	TST	TSTR	TSH	DTSH
1	N	10,1	2,2	0,9	2,7
2	N	9,9	3,1	2	5,9
3	N	12,9	2,4	29	0,6
4	N	5,3	1,6	1,4	1,5
5	N	7,3	1,5	1,5	-0,1
6	N	6,1	2,1	1,4	7

Kesin Kurallar

EĞER	TST	=	HPR	İSE	SINIF	=	HPR	f	=	0,25
EĞER	TSTR	=	N	İSE	SINIF	=	HPR	f	=	0,0499
EĞER	TSH	=	N	İSE	SINIF	=	HPO	f	=	1
EĞER	DTSH	=	HPO	İSE	SINIF	=	HPO	f	=	0,8

Mümkün Kurallar

EĞER	TST	=	N <th>İSE</th> <th>SINIF</th> <th>=</th> <th>N</th> <th>f</th> <th>=</th> <th>0,04</th> <th>p</th>	İSE	SINIF	=	N	f	=	0,04	p
EĞER	TST	=	HPO	İSE	SINIF	=	N	f	=	0,15	p
EĞER	TST	=	N	İSE	SINIF	=	HPR	f	=	0,04	p
EĞER	TST	=	N	İSE	SINIF	=	HPO	f	=	0,04	p
EĞER	TST	=	HPO	İSE	SINIF	=	HPO	f	=	0,15	p

☒ Kesin Kural Analizi
 ☒ Mümkün Kural Analizi
 ☒ Analiz Sonucu

Şekil 5.3. Uygulama yazılımı arayüzü

5.1. Uygulama Ortamı

Uygulama yazılımı gerçekleştirilirken Delphi 7.0 programlama dili kullanılmıştır. Uygulama yazılımı, Pentium III 2.4 GHz işlemciye ve 1 GB RAM’e sahip bir bilgisayarda oluşturulmuştur.

5.2. Geliştirilen Yazılım ile Elde Edilen Kurallar

Bu bölümde altı farklı durum için elde edilen kurallara yer verilecektir. İlk durumda, veri kümesinde 5 adet kayıp (?), 11 adet önemsiz (*) nitelik değeri yer almaktadır. Uygulama yazılımının çalıştırılması sonucu, 70 adet kesin kural, 239 adet olası kural elde edilmiştir. Toplamda elde edilen 309 adet kuraldan, daha fazla özelleşmiş olanları, uygunluk ve etkililik değerleri eşit veya küçük olanları silinmiştir. Sonuçta 21 adet kural elde edilmiştir. Bu kurallar aşağıda listelenmiştir. Uygulama yazılımının bu duruma ait ekran görüntüsü Şekil 5.4'te gösterilmiştir.

The screenshot displays the 'Kaba Küme Yaklaşımı' software interface. It features two main data tables at the top: 'Eksik Veri kümesi' and 'Tahmin sonucu veri kümesi'. Below these are sections for 'Kesin Kurallar' (Certain Rules) and 'Mümkün Kurallar' (Possible Rules). At the bottom, there are three buttons: 'Kesin Kural Analizi' (green checkmark), 'Mümkün Kural Analizi' (green checkmark), and 'Analiz Sonucu' (red X). The 'Analiz Sonucu' button is active, showing a list of 21 rules with their corresponding f and p values.

	SINIF	TST	TSTR	TSH	DTSH
1	N	10,1	2,2	0,9	2,7
2	N	9,9	3,1	2	5,9
3	N	12,9	2,4	*	0,6
4	N	5,3	1,6	1,4	1,5
5	N	7,3	1,5	1,5	-0,1
6	N	6,1	2,1	1,4	7

	SINIF	TST	TSTR	TSH	DTSH
1	N	10,1	2,2	0,9	2,7
2	N	9,9	3,1	2	5,9
3	N	12,9	2,4	2,9	0,6
4	N	5,3	1,6	1,4	1,5
5	N	7,3	1,5	1,5	-0,1
6	N	6,1	2,1	1,4	7

Kesin Kurallar

EĞER	TST	=	HPR	İSE	SINIF	=	HPR	f	=	0,25
EĞER	TSTR	=	N	İSE	SINIF	=	HPR	f	=	0,0499
EĞER	TSH	=	N	İSE	SINIF	=	HPO	f	=	1
EĞER	DTSH	=	HPO	İSE	SINIF	=	HPO	f	=	0,8

Mümkün Kurallar

EĞER	TST	=	N <td>İSE</td> <td>SINIF</td> <td>=</td> <td>N</td> <th>f</th> <th>=</th> <th>0,04</th> <th>p</th>	İSE	SINIF	=	N	f	=	0,04	p
X	TST	=	HPO	İSE	SINIF	=	N	f	=	0,15	p
X	TST	=	N	İSE	SINIF	=	HPR	f	=	0,04	p
X	TST	=	N	İSE	SINIF	=	HPO	f	=	0,04	p
EĞER	TST	=	HPR	İSE	SINIF	=	HPR	f	=	0,15	p

Analiz Sonucu

18 EĞER TSH = HPR İSE SINIF = HPO	f = 1 p = 0,554625984251969
19 EĞER DTSH = N İSE SINIF = N	f = 0,01 p = 0,45097037793667
20 EĞER DTSH = HPO İSE SINIF = HPR	f = 0,06 p = 0,520319027725029
21 EĞER DTSH = HPR İSE SINIF = HPO	f = 0,8 p = 0,932065217391304

Şekil 5.4. Uygulama yazılımı ile kesin ve olası kuralların elde edilmesi

Kesin Kurallar

1.EĞER TST = HPR İSE SINIF = HPR	f = 0,25
2.EĞER TSTR = N İSE SINIF = HPR	f = 0,05
3.EĞER TSH = N İSE SINIF = HPO	f = 1
4.EĞER DTSH = HPO İSE SINIF = HPO	f = 0,8
5.EĞER TST = HPO VE TSH = HPR İSE SINIF = HPO	f = 1
6.EĞER TST = HPO VE DTSH = HPR İSE SINIF = HPO	f = 1
7.EĞER TSH = HPR VE DTSH = HPO İSE SINIF = N	f = 0,46
8.EĞER TSH = HPR VE DTSH = HPR İSE SINIF = HPO	f = 1
9.EĞER TSTR = N VE TSH = N VE DTSH = HPO İSE SINIF = N	f = 0,075
10.EĞER TST = N VE TSH = N VE DTSH = HPO İSE SINIF = N	f = 0,41

Olası Kurallar

1.EĞER TST = N İSE SINIF = N	f = 0,04	p = 0,550603299679882
2.EĞER TST = HPO İSE SINIF = HPO	f = 0,15	p = 0,91566265060241
3.EĞER TSTR = N İSE SINIF = HPR	f = 0,025	p = 0,603792415169661
4.EĞER TSTR = HPR İSE SINIF = HPR	f = 0,05	p = 0,646017699115044
5.EĞER TSTR = HPO İSE SINIF = HPO	f = 0,15	p = 0,478632478632479
6.EĞER TSH = HPO İSE SINIF = N	f = 0,084	p = 0,479525620944283
7.EĞER TSH = N İSE SINIF = HPO	f = 0,01	p = 0,90089358245329
8.EĞER TSH = HPR İSE SINIF = HPO	f = 1	p = 0,554625984251969
9.EĞER DTSH = N İSE SINIF = N	f = 0,01	p = 0,462788259958071
10.EĞER DTSH = HPO İSE SINIF = HPR	f = 0,06	p = 0,520319027725029
11.EĞER DTSH = HPR İSE SINIF = HPO	f = 0,4	p = 0,907407407407407

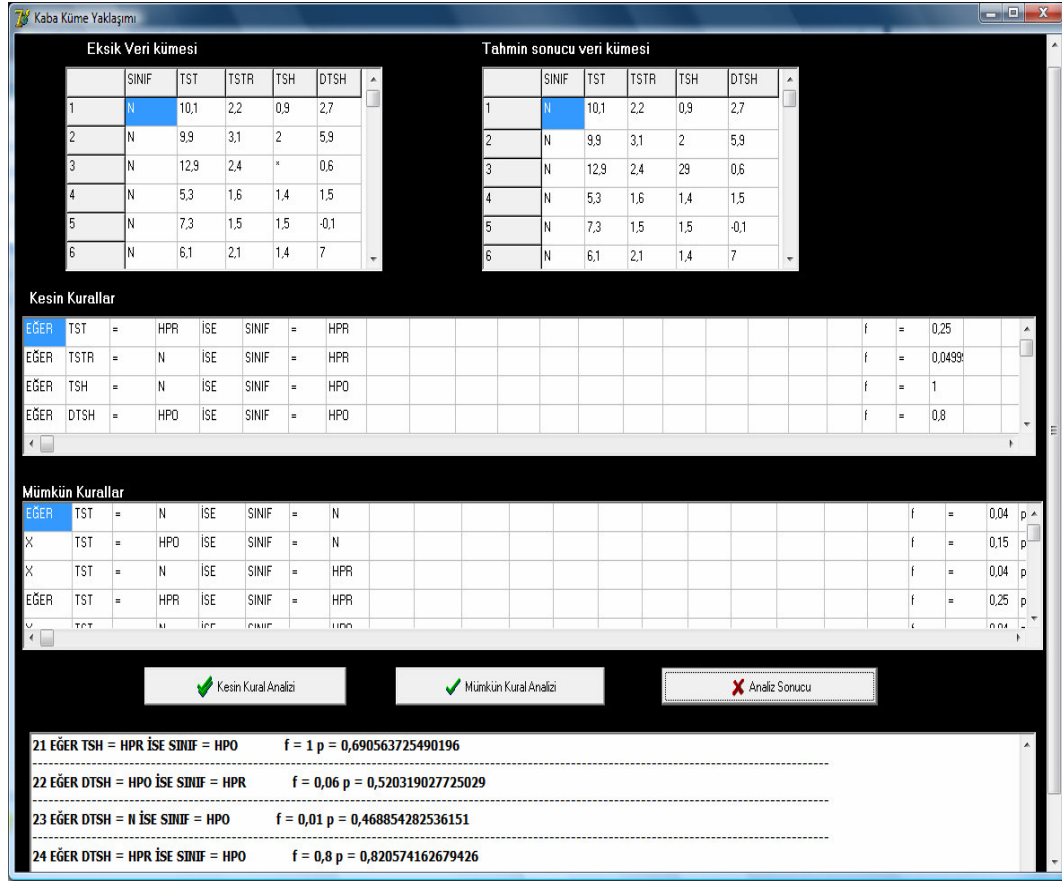
İkinci durumda ise kayıp niteliklere yer verilmemiştir. Veri kümesinin sadece önemsiz türde eksik nitelik içermesi durumu ele alınmıştır. 16 adet önemsiz türde eksik nitelik içeren veri kümesi üzerinde çalışılmıştır. Bu durumda 75 adet kesin kural, 255 adet olası kural elde edilmiştir. Toplamda elde edilen 330 adet kuraldan, daha fazla özelleşmiş olanları, uygunluk ve etkililik değerleri eşit veya küçük olanları silinmiştir. Sonuçta 24 adet kural elde edilmiştir. Bu kurallar aşağıda listelenmiştir. Uygulama yazılımının bu duruma ait ekran görüntüsü Şekil 5.5.'de gösterilmiştir.

Kesin Kurallar

1. <i>EĞER</i> TST = HPR <i>İSE</i> SINIF = HPR	f = 0,25
2. <i>EĞER</i> TSTR = N <i>İSE</i> SINIF = HPR	f = 0,05
3. <i>EĞER</i> TSH = N <i>İSE</i> SINIF = HPO	f = 1
4. <i>EĞER</i> DTSH = HPO <i>İSE</i> SINIF = HPO	f = 0,8
5. <i>EĞER</i> TST = HPR <i>VE</i> TSH = HPR <i>İSE</i> SINIF = HPO	f = 1
6. <i>EĞER</i> TST = HPO <i>VE</i> TSH = HPR <i>İSE</i> SINIF = HPO	f = 1
7. <i>EĞER</i> TST = HPR <i>VE</i> DTSH = HPR <i>İSE</i> SINIF = HPO	f = 1
8. <i>EĞER</i> TST = HPO <i>VE</i> DTSH = HPR <i>İSE</i> SINIF = HPO	f = 1
9. <i>EĞER</i> TSH = HPR <i>VE</i> DTSH = HPO <i>İSE</i> SINIF = N	f = 0,56
10. <i>EĞER</i> TSH = HPR <i>VE</i> DTSH = HPR <i>İSE</i> SINIF = HPO	f = 1
11. <i>EĞER</i> TSTR = N <i>VE</i> TSH = N <i>VE</i> DTSH = HPO <i>İSE</i> SINIF = N	f = 0,17
12. <i>EĞER</i> TST = N <i>VE</i> TSH = N <i>VE</i> DTSH = HPO <i>İSE</i> SINIF = N	f = 0,17

Olası Kurallar

1. <i>EĞER</i> TST = N <i>İSE</i> SINIF = N	f = 0,04	p = 0,563082347015865
2. <i>EĞER</i> TST = HPR <i>İSE</i> SINIF = HPR	f = 0,25	p = 0,902912621359223
3. <i>EĞER</i> TST = HPO <i>İSE</i> SINIF = HPO	f = 0,15	p = 0,91566265060241
4. <i>EĞER</i> TSTR = N <i>İSE</i> SINIF = HPR	f = 0,025	p = 0,603792415169661
5. <i>EĞER</i> TSTR = HPR <i>İSE</i> SINIF = HPR	f = 0,05	p = 0,646017699115044
6. <i>EĞER</i> TSTR = HPO <i>İSE</i> SINIF = HPO	f = 0,15	p = 0,478632478632479
7. <i>EĞER</i> TSH = HPO <i>İSE</i> SINIF = N	f = 0,084	p = 0,479525620944283
8. <i>EĞER</i> TSH = N <i>İSE</i> SINIF = HPO	f = 0,01	p = 0,939034716342083
9. <i>EĞER</i> TSH = HPR <i>İSE</i> SINIF = HPO	f = 1	p = 0,690563725490196
10. <i>EĞER</i> DTSH = HPO <i>İSE</i> SINIF = HPR	f = 0,06	p = 0,520319027725029
11. <i>EĞER</i> DTSH = N <i>İSE</i> SINIF = HPO	f = 0,01	p = 0,468854282536151
12. <i>EĞER</i> DTSH = HPR <i>İSE</i> SINIF = HPO	f = 0,8	p = 0,820574162679426



Şekil 5.5. Uygulama yazılımı ile ikinci durum için kesin ve olası kuralların elde edilmesi

Üçüncü durumda, önemsiz niteliklere yer verilmemiştir. Veri kümesinin sadece kayıp türde eksik nitelik içermesi durumu ele alınmıştır. 16 adet kayıp türde eksik nitelik içeren veri kümesi üzerinde çalışılmıştır. Bu durumda 80 adet kesin kural, 251 adet olası kural elde edilmiştir. Toplamda elde edilen 331 adet kuraldan, daha fazla özelleşmiş olanları, uygunluk ve etkililik değerleri eşit veya küçük olanları silinmiştir. Sonuçta 34 adet kural elde edilmiştir. Bu kurallar aşağıda listelenmiştir.

Kesin Kurallar

1. EĞER TST = HPR İSE SINIF = HPR f = 0,25
2. EĞER TSH = N İSE SINIF = HPO f = 1
3. EĞER DTSH = HPO İSE SINIF = HPO f = 0,8
4. EĞER TST = N VE TSTR = HPR İSE SINIF = HPR f = 0,05
5. EĞER TST = N VE TSH = HPR İSE SINIF = HPO f = 0,08

6. $E\check{G}ER\ TST = HPO\ VE\ TSH = HPR\ \acute{I}SE\ SINIF = HPO$	$f = 1$
7. $E\check{G}ER\ TST = N\ VE\ DTSH = HPR\ \acute{I}SE\ SINIF = HPO$	$f = 0,21$
8. $E\check{G}ER\ TST = HPO\ VE\ DTSH = HPR\ \acute{I}SE\ SINIF = HPO$	$f = 0,8$
9. $E\check{G}ER\ TSTR = HPR\ VE\ TSH = HPO\ \acute{I}SE\ SINIF = HPR$	$f = 0,4$
10. $E\check{G}ER\ TSTR = N\ VE\ TSH = HPR\ \acute{I}SE\ SINIF = HPO$	$f = 0,025$
11. $E\check{G}ER\ TSTR = HPO\ VE\ TSH = HPR\ \acute{I}SE\ SINIF = HPO$	$f = 0,3$
12. $E\check{G}ER\ TSTR = HPR\ VE\ DTSH = N\ \acute{I}SE\ SINIF = HPR$	$f = 0,01$
13. $E\check{G}ER\ TSTR = HPR\ VE\ DTSH = HPO\ \acute{I}SE\ SINIF = HPR$	$f = 0,4$
14. $E\check{G}ER\ TSTR = N\ VE\ DTSH = HPR\ \acute{I}SE\ SINIF = HPO$	$f = 0,025$
15. $E\check{G}ER\ TSTR = HPO\ VE\ DTSH = HPR\ \acute{I}SE\ SINIF = HPO$	$f = 0,3$
16. $E\check{G}ER\ TSH = HPR\ VE\ DTSH = N\ \acute{I}SE\ SINIF = HPO$	$f = 0,06$
17. $E\check{G}ER\ TSH = HPR\ VE\ DTSH = HPR\ \acute{I}SE\ SINIF = HPO$	$f = 0,8$
18. $E\check{G}ER\ TSH = HPO\ VE\ DTSH = HPR\ \acute{I}SE\ SINIF = HPO$	$f = 0,16$
19. $E\check{G}ER\ TSTR = N\ VE\ TSH = N\ VE\ DTSH = HPO\ \acute{I}SE\ SINIF = N$	$f = 0,17$
20. $E\check{G}ER\ TSTR = N\ VE\ TSH = HPR\ VE\ DTSH = HPO\ \acute{I}SE\ SINIF = N$	$f = 0,075$
21. $E\check{G}ER\ TSTR = HPO\ VE\ TSH = HPR\ VE\ DTSH = HPO\ \acute{I}SE\ SINIF = N$	$f = 0,56$
22. $E\check{G}ER\ TST = N\ VE\ TSH = N\ VE\ DTSH = HPO\ \acute{I}SE\ SINIF = N$	$f = 0,17$
23. $E\check{G}ER\ TST = N\ VE\ TSH = HPR\ VE\ DTSH = HPO\ \acute{I}SE\ SINIF = N$	$f = 0,41$

Olası Kurallar

1. $E\check{G}ER\ TST = N\ \acute{I}SE\ SINIF = N$	$f = 0,04$	$p = 0,51508868924211$
2. $E\check{G}ER\ TST = HPO\ \acute{I}SE\ SINIF = HPO$	$f = 0,15$	$p = 0,91566265060241$
3. $E\check{G}ER\ TSTR = N\ \acute{I}SE\ SINIF = HPR$	$f = 0,025$	$p = 0,61531007751938$
4. $E\check{G}ER\ TSTR = HPR\ \acute{I}SE\ SINIF = HPR$	$f = 0,05$	$p = 0,56989247311828$
5. $E\check{G}ER\ TSTR = HPO\ \acute{I}SE\ SINIF = HPO$	$f = 0,15$	$p = 0,478632478632479$
6. $E\check{G}ER\ TSH = HPO\ \acute{I}SE\ SINIF = N$	$f = 0,084$	$p = 0,479525620944283$
7. $E\check{G}ER\ TSH = N\ \acute{I}SE\ SINIF = HPO$	$f = 0,01$	$p = 0,939034716342083$
8. $E\check{G}ER\ TSH = HPR\ \acute{I}SE\ SINIF = HPO$	$f = 1$	$p = 0,690563725490196$
9. $E\check{G}ER\ DTSH = HPO\ \acute{I}SE\ SINIF = HPR$	$f = 0,06$	$p = 0,520319027725029$
10. $E\check{G}ER\ DTSH = N\ \acute{I}SE\ SINIF = HPO$	$f = 0,01$	$p = 0,480413492927095$
11. $E\check{G}ER\ DTSH = HPR\ \acute{I}SE\ SINIF = HPO$	$f = 0,6$	$p = 0,811557788944724$

Dördüncü durumda, önemsiz nitelik sayısı iki katına çıkarılmıştır. 22 adet önemsiz ve 5 adet kayıp türde eksik nitelik içeren veri kümesi üzerinde çalışılmıştır. Bu durumda 85 adet kesin kural, 263 adet olası kural elde edilmiştir. Toplamda elde edilen 348 adet kuraldan, daha fazla özelleşmiş olanları, uygunluk ve etkililik değerleri eşit veya küçük olanları silinmiştir. Sonuçta 25 adet kural elde edilmiştir. Bu kurallar aşağıda listelenmiştir.

Kesin Kurallar

1. <i>EĞER</i> TST = HPR <i>İSE</i> SINIF = HPR	f = 0,25
2. <i>EĞER</i> TSTR = N <i>İSE</i> SINIF = HPR	f = 0,05
3. <i>EĞER</i> TSH = N <i>İSE</i> SINIF = HPO	f = 1
4. <i>EĞER</i> DTSH = HPO <i>İSE</i> SINIF = HPO	f = 0,8
5. <i>EĞER</i> TST = HPO <i>VE</i> TSH = HPO <i>İSE</i> SINIF = HPO	f = 0,15
6. <i>EĞER</i> TST = HPR <i>VE</i> DTSH = HPR <i>İSE</i> SINIF = HPO	f = 1
7. <i>EĞER</i> TST = HPO <i>VE</i> DTSH = HPR <i>İSE</i> SINIF = HPO	f = 1
8. <i>EĞER</i> TSH = HPR <i>VE</i> DTSH = HPO <i>İSE</i> SINIF = N	f = 0,46
9. <i>EĞER</i> TSH = HPR <i>VE</i> DTSH = HPR <i>İSE</i> SINIF = HPO	f = 1
10. <i>EĞER</i> TST = HPR <i>VE</i> TSTR = HPO <i>VE</i> TSH = HPR <i>İSE</i> SINIF = HPO	f = 1
11. <i>EĞER</i> TST = HPR <i>VE</i> TSTR = HPO <i>VE</i> DTSH = N <i>İSE</i> SINIF = HPO	f = 0,29
12. <i>EĞER</i> TSTR = N <i>VE</i> TSH = N <i>VE</i> DTSH = HPO <i>İSE</i> SINIF = N	f = 0,075
13. <i>EĞER</i> TST = N <i>VE</i> TSH = N <i>VE</i> DTSH = HPO <i>İSE</i> SINIF = N	f = 0,41

Olası Kurallar

1. <i>EĞER</i> TST = N <i>İSE</i> SINIF = N	f = 0,05	p = 0,572745901639345
2. <i>EĞER</i> TST = HPR <i>İSE</i> SINIF = HPR	f = 0,25	p = 0,774436090225564
3. <i>EĞER</i> TST = HPO <i>İSE</i> SINIF = HPO	f = 0,15	p = 0,905405405405405
4. <i>EĞER</i> TSTR = N <i>İSE</i> SINIF = HPR	f = 0,025	p = 0,612348178137652
5. <i>EĞER</i> TSTR = HPR <i>İSE</i> SINIF = HPR	f = 0,05	p = 0,477124183006536
6. <i>EĞER</i> TSTR = HPO <i>İSE</i> SINIF = HPO	f = 0,15	p = 0,490749756572541
7. <i>EĞER</i> TSH = HPO <i>İSE</i> SINIF = N	f = 0,084	p = 0,465164405610485
8. <i>EĞER</i> TSH = N <i>İSE</i> SINIF = HPO	f = 0,01	p = 0,884514435695538
9. <i>EĞER</i> TSH = HPR <i>İSE</i> SINIF = HPO	f = 1	p = 0,683806146572104
10. <i>EĞER</i> DTSH = HPO <i>İSE</i> SINIF = HPR	f = 0,06	p = 0,523076923076923
11. <i>EĞER</i> DTSH = N <i>İSE</i> SINIF = HPO	f = 0,01	p = 0,460404150737302
12. <i>EĞER</i> DTSH = HPR <i>İSE</i> SINIF = HPO	f = 0,8	p = 0,820574162679426

Beşinci durumda, kayıp nitelik sayısı iki katına çıkarılmıştır. 11 adet önemsiz ve 10 adet kayıp türde eksik nitelik içeren veri kümesi üzerinde çalışılmıştır. Bu durumda 65 adet kesin kural, 257 adet olası kural elde edilmiştir. Toplamda elde edilen 322 adettan, daha fazla özelleşmiş olanları, uygunluk ve etkililik değerleri eşit veya küçük olanları silinmiştir. Sonuçta 21 adet kural elde edilmiştir. Bu kurallar aşağıda listelenmiştir.

Kesin Kurallar

1. <i>EĞER</i> TST = HPR <i>İSE</i> SINIF = HPR	f = 0,25
2. <i>EĞER</i> TSTR = N <i>İSE</i> SINIF = HPR	f = 0,05
3. <i>EĞER</i> TSH = N <i>İSE</i> SINIF = HPO	f = 1
4. <i>EĞER</i> DTSH = HPO <i>İSE</i> SINIF = HPO	f = 0,8
5. <i>EĞER</i> TST = HPO <i>VE</i> TSH = HPR <i>İSE</i> SINIF = HPO	f = 1
6. <i>EĞER</i> TST = HPO <i>VE</i> DTSH = HPR <i>İSE</i> SINIF = HPO	f = 1
7. <i>EĞER</i> TSH = HPR <i>VE</i> DTSH = HPO <i>İSE</i> SINIF = N	f = 0,46
8. <i>EĞER</i> TSH = HPR <i>VE</i> DTSH = HPR <i>İSE</i> SINIF = HPO	f = 1
9. <i>EĞER</i> TSTR = N <i>VE</i> TSH = N <i>VE</i> DTSH = HPO <i>İSE</i> SINIF = N	f = 0,5
10. <i>EĞER</i> TST = N <i>VE</i> TSH = N <i>VE</i> DTSH = HPO <i>İSE</i> SINIF = N	f = 0,41

Olası Kurallar

1. <i>EĞER</i> TST = N <i>İSE</i> SINIF = N	f = 0,04	p = 0,548986486486487
2. <i>EĞER</i> TST = HPO <i>İSE</i> SINIF = HPO	f = 0,15	p = 0,91566265060241
3. <i>EĞER</i> TSTR = N <i>İSE</i> SINIF = HPR	f = 0,025	p = 0,593719332679097
4. <i>EĞER</i> TSTR = HPR <i>İSE</i> SINIF = HPR	f = 0,05	p = 0,646017699115044
5. <i>EĞER</i> TSTR = HPO <i>İSE</i> SINIF = HPO	f = 0,15	p = 0,486486486486486
6. <i>EĞER</i> TSH = HPO <i>İSE</i> SINIF = N	f = 0,084	p = 0,479525620944283
7. <i>EĞER</i> TSH = N <i>İSE</i> SINIF = HPO	f = 0,01	p = 0,893634165995165
8. <i>EĞER</i> TSH = HPR <i>İSE</i> SINIF = HPO	f = 1	p = 0,703495630461923
9. <i>EĞER</i> DTSH = HPO <i>İSE</i> SINIF = HPR	f = 0,06	p = 0,521507422915874
10. <i>EĞER</i> DTSH = N <i>İSE</i> SINIF = HPO	f = 0,01	p = 0,451893551688843
11. <i>EĞER</i> DTSH = HPR <i>İSE</i> SINIF = HPO	f = 0,2	p = 0,915014164305949

Altıncı durumda ise, hem önemsiz hem de kayıp nitelik sayısı iki katına çıkarılmıştır. 22 adet önemsiz ve 10 adet kayıp türde eksik nitelik içeren veri kümesi üzerinde çalışılmıştır. Bu durumda 75 adet kesin kural, 287 adet olası kural elde edilmiştir. Toplamda elde edilen 362 adet kuraldan, daha fazla özelleşmiş olanları, uygunluk ve etkililik değerleri eşit veya küçük olanları silinmiştir. Sonuçta 25 adet kural elde edilmiştir. Bu kurallar aşağıda listelenmiştir.

Kesin Kurallar

1. $E\check{G}ER\ TST = HPR\ \acute{I}SE\ SINIF = HPR$	$f = 0,25$
2. $E\check{G}ER\ TSTR = N\ \acute{I}SE\ SINIF = HPR$	$f = 0,05$
3. $E\check{G}ER\ TSH = N\ \acute{I}SE\ SINIF = HPO$	$f = 1$
4. $E\check{G}ER\ DTSH = HPO\ \acute{I}SE\ SINIF = HPO$	$f = 0,8$
5. $E\check{G}ER\ TST = HPO\ VE\ TSH = HPO\ \acute{I}SE\ SINIF = HPO$	$f = 0,15$
6. $E\check{G}ER\ TST = HPR\ VE\ DTSH = HPR\ \acute{I}SE\ SINIF = HPO$	$f = 1$
7. $E\check{G}ER\ TST = HPO\ VE\ DTSH = HPR\ \acute{I}SE\ SINIF = HPO$	$f = 1$
8. $E\check{G}ER\ TSH = HPR\ VE\ DTSH = HPO\ \acute{I}SE\ SINIF = N$	$f = 0,46$
9. $E\check{G}ER\ TSH = HPR\ VE\ DTSH = HPR\ \acute{I}SE\ SINIF = HPO$	$f = 1$
10. $E\check{G}ER\ TST = HPR\ VE\ TSTR = HPO\ VE\ TSH = HPR\ \acute{I}SE\ SINIF = HPO$	$f = 1$
11. $E\check{G}ER\ TST = HPR\ VE\ TSTR = HPO\ VE\ DTSH = N\ \acute{I}SE\ SINIF = HPO$	$f = 0,29$
12. $E\check{G}ER\ TSTR = N\ VE\ TSH = N\ VE\ DTSH = HPO\ \acute{I}SE\ SINIF = N$	$f = 0,56$
13. $E\check{G}ER\ TST = N\ VE\ TSH = N\ VE\ DTSH = HPO\ \acute{I}SE\ SINIF = N$	$f = 0,41$

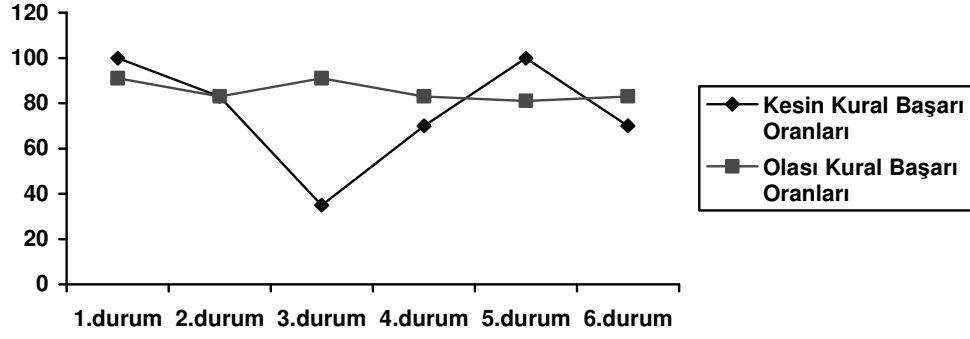
Olası Kurallar

1. $E\check{G}ER\ TST = N\ \acute{I}SE\ SINIF = N$	$f = 0,05$	$p = 0,570209464701319$
2. $E\check{G}ER\ TST = HPR\ \acute{I}SE\ SINIF = HPR$	$f = 0,25$	$p = 0,774436090225564$
3. $E\check{G}ER\ TST = HPO\ \acute{I}SE\ SINIF = HPO$	$f = 0,15$	$p = 0,848101265822785$
4. $E\check{G}ER\ TSTR = N\ \acute{I}SE\ SINIF = HPR$	$f = 0,025$	$p = 0,596059113300493$
5. $E\check{G}ER\ TSTR = HPR\ \acute{I}SE\ SINIF = HPR$	$f = 0,05$	$p = 0,477124183006536$
6. $E\check{G}ER\ TSTR = HPO\ \acute{I}SE\ SINIF = HPO$	$f = 0,15$	$p = 0,499009900990099$
7. $E\check{G}ER\ TSH = HPO\ \acute{I}SE\ SINIF = N$	$f = 0,084$	$p = 0,465164405610485$
8. $E\check{G}ER\ TSH = N\ \acute{I}SE\ SINIF = HPO$	$f = 0,01$	$p = 0,884514435695538$
9. $E\check{G}ER\ TSH = HPR\ \acute{I}SE\ SINIF = HPO$	$f = 1$	$p = 0,683806146572104$
10. $E\check{G}ER\ DTSH = HPO\ \acute{I}SE\ SINIF = HPR$	$f = 0,06$	$p = 0,5229793977813$
11. $E\check{G}ER\ DTSH = N\ \acute{I}SE\ SINIF = HPO$	$f = 0,01$	$p = 0,44663631765301$
12. $E\check{G}ER\ DTSH = HPR\ \acute{I}SE\ SINIF = HPO$	$f = 0,2$	$p = 0,801488833746898$

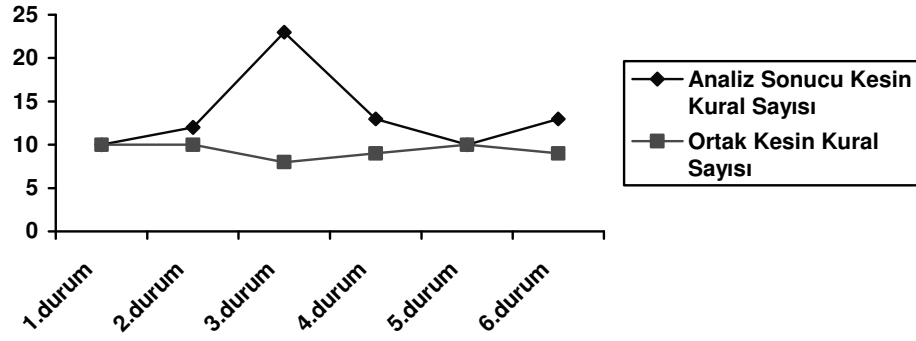
Bu altı özel durumu karşılaştırmadaki temel amaç, sadece önemsiz nitelik değeri içeren veri kümeleri için geliştirilmiş algoritma ile bu tez çalışması sırasında geliştirilen ve her iki tipte eksik nitelik içeren veri kümeleri üzerinde çalışabilen algoritmanın birbirine olan uygunluğunu test etmektir. Ayrıca eksik nitelik sayısının ve türünün değiştirilmesinin elde edilen kurallar üzerinde nasıl bir değişime yol açtığı tespit edilmek istenmiştir. Farklı altı durum için elde edilen sonuçlara aşağıdaki tabloda yer verilmiştir.

Tablo 5.1. Uygulama yazılımı sonuçlarının karşılaştırılması

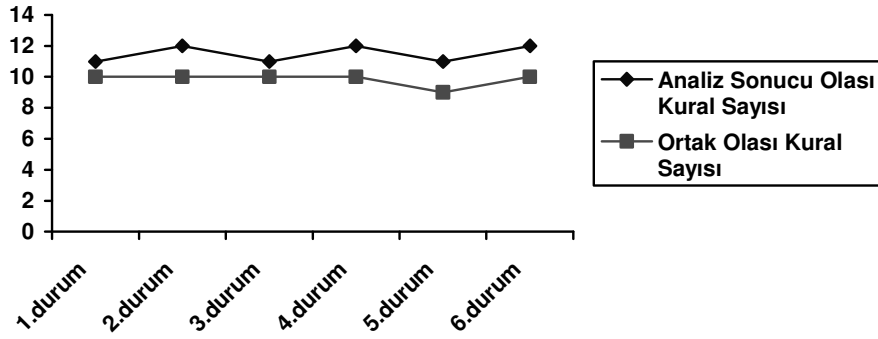
	Kesin kural sayısı	Olası kural sayısı	Analiz sonucu kesin kural sayısı	Analiz sonucu olası kural sayısı	Ortak kesin kural sayısı	Ortak olası kural sayısı	Kesin kurallar için başarı oranları	Olası kurallar için başarı oranları
1.durum (Önerilen algoritma)	70	239	10	11	10	10	%100	%91
2.durum (Hong'un[21] çalışması)	75	255	12	12	10	10	%83	%83
3.durum (Eksik niteliklerin tümü kayıp)	80	251	23	11	8	10	%35	%91
4.durum (iki kat önemsiz nitelik)	85	263	13	12	9	10	%70	%83
5.durum (iki kat kayıp nitelik)	65	257	10	11	10	9	%100	%81
6.durum (iki kat önemsiz ve kayıp nitelik)	75	287	13	12	9	10	%70	%83



Şekil 5.6. Kesin ve olası kurallar için başarı oranları



Şekil 5.7. Analiz sonucu kesin kural sayısı ve ortak kesin kural sayısı



Şekil 5.8. Analiz sonucu olası kural sayısı ve ortak olası kural sayısı

Tablo 5.1'e göre ařağıdaki sonuçlar çıkarılabilir.

- Başarının en yüksek olduğı durum, birinci durumdur (önerilen algoritma).
- Başarının en düşük olduğı durum, üçüncü durumdur.
- Tüm eksik niteliklerin kayıp türde oluşu, elde edilen kural sayısını arttırmaktadır.
- Analiz sonucu elde edilen kural sayısının artışı, başarı oranını düşürmektedir.
- Tüm eksik nitelik değerlerinin kayıp türde olduğı üçüncü durumda, kesin kural başarısında azalma görölmüştür. Ancak olası kural başarısı üzerinde bir değışikliğe neden olmamıştır.
- Önemsiz türdeki eksik nitelik değerlerinin iki katına çıkarıldığı dördüncü durumda, hem kesin hem de olası kural başarısında azalma olmuştur.
- Kayıp türdeki eksik nitelik değerlerinin iki katına çıkarıldığı beşinci durumda, kesin kural başarısı sabit kalırken, olası kural başarısında azalma olmuştur.
- Kayıp ve önemsiz türdeki eksik nitelik değerlerinin her ikisinin de iki katına çıkarılması, hem kesin hem de olası kural başarısında azalmaya neden olmuştur.
- Başarı oranında meydana gelen azalmaların sebebi, ortak kural kaybı değil, fazladan kural üretilmesidir. Dolayısıyla, eksik nitelik sayısı artışı büyük ölçekli bilgi kayıplarına neden olmamaktadır.

6. SONUÇLAR

Kaba küme teorisi doğrulanmış mantığa, tutarsızlık gösteren verilere ve kesinlik olmayan gizli çıkarımların keşfine izin veren matematiksel bir metottur. Bulanık kümelerde olduğu gibi, kaba küme teorisi de kesin sınırlamaları kabul etmeyen bir yapıdadır. Bu tez çalışmasında, kaba küme teorisi kullanılarak, eksik veri kümelerinden bilgi çıkarımı üzerinde durulmuştur. Kaba küme teorisinin temelini oluşturan yaklaşımların, tam ve eksik veri kümeleri üzerinde nasıl elde edildiği hakkında bilgi verilmiştir. Önemsiz ve kayıp olmak üzere, iki tip eksik nitelik değeri üzerinde çalışılmıştır. Önemsiz nitelik değeri, niteliğe ait herhangi bir değeri alabileceği düşünülen nitelik değeridir. Kayıp nitelik değeri ise veri kümesi için yararlı olan ancak herhangi bir sebeple şuan da erişemediğimiz nitelik değeridir.

Tez çalışmasında, ayrıca kaba küme teorisi ile nicel değerli eksik veri kümelerinden bulanık kural çıkarımını sağlayan bir algoritma üzerinde durulmuştur. Algoritma, öncelikle nitelikler için verilen üyelik fonksiyonları sayesinde eksik veri kümesini bulanıklaştırır. Ardından, üzerinde çalışılan nitelik sayısını her iterasyonda arttırarak, birincil kümeleri elde eder. Birincil kümelerden, kaba küme teorisinin temelini oluşturan kavramlar olan, düşük ve yüksek yaklaşımları hesaplar. Bu yaklaşımlar sayesinde kurallar çıkarılır. Düşük yaklaşımlardan kesin, yüksek yaklaşımlardan olası kurallar elde edilir.

Sadece önemsiz türdeki eksik nitelik değerleri için geliştirilen algoritma, bu tez çalışması sırasında hem önemsiz, hem de kayıp nitelik değerini birlikte içeren, eksik veri kümeleri üzerinde uygulanabilir biçimde geliştirilmiştir. Geliştirilen algoritma, uygulama yazılımı haline getirilmiş ve elde edilen kesin ve olası kurallara yer verilmiştir. Uygulama yazılımında, tiroit hastalığına ait 100 adet nesne ile çalışılmıştır. Bu çalışmada, hastalık verilerinin tercih edilmesindeki temel sebep, eksik nitelik değerlerinin bu tür veri kümelerinde daha sık görülmesi ve tahmin edilmesinin hastalık teşhisinde çok önemli bir yer tutmasıdır.

Altı farklı durum için, uygulama yazılımı test edilmiştir ve elde edilen kesin ve olası kurallara yer verilmiştir. Bu yolla, yeni geliştirilen algoritmanın uygulanabilirliği ölçülmüştür. Yapılan bu test işleminde, önerilen algoritmanın kesin kurallarda %100, olası kurallarda %91 ortak kural başarısına sahip olduğu saptanmıştır. Ayrıca bu algoritmanın sadece kayıp türde eksik nitelik değeri içeren veri kümelerinde uygulanmasının, elde edilen kural sayısını arttırdığı ve bu yolla başarı oranını düşürdüğü belirlenmiştir. Eksik nitelik sayılarındaki artışların

meydana getirdiđi başarı oranı azalmalarının, elde edilen kural sayısının artışından kaynaklandığı ve elde edilen ortak kuralda büyük ölçekli kayıplara neden olmadığı sonucuna varılmıştır.

Bundan sonra, önemsiz ve kayıp türdeki eksik nitelik değerlerinin yanı sıra, farklı bir eksik nitelik değeri olan, kısmi bir önemsiz durumu ifade eden, nitelik-kavram olarak isimlendirilen ve ‘-’ sembolü ile gösterilen, eksik veri değerini de içeren veri kümeleri için, bu algoritma geliştirilebilir.

KAYNAKLAR

1. J.W. Grzymala-Busse, 1988, "Knowledge acquisition under uncertainty: A rough set approach", *Journal of Intelligent Robotic systems*, Vol 1, pp. 3-16
2. E. Orlowska, 1994, "Reasoning with incomplete information: rough set based information logics" *Incompleteness and Uncertainty in Information Systems*, New York: Springer-Verlag, pp. 16-33
3. L.T. Germano, P. Alexandre, 1996, "Knowledge-base reduction based on rough set techniques" *The Canadian Conference on Electrical and Computer Engineering*, pp. 278-281
4. P.J. Lingras and Y.Y. Yao, 1998, "Data mining using extensions of the rough set model" *Journal of American Society for Information Science*, Vol. 49, No. 5, pp. 415-422
5. N. Zhong, J.Z. Dong, S. Ohsuga, T.Y. Lin, 1998, "An incremental probabilistic rough set approach to rule discovery" *The IEEE International Conference on Fuzzy Systems*, Vol. 2, pp. 933-938
6. M.R. Chmielewski, J.W. Grzymala-Busse, N.W. Peterson, S. Than, 1993, "The Rule Induction systems LERS-a version for personal computers", *Foundations of computing and Decision Sciences*, Vol. 18, pp. 181-212
7. R. Slowinski, J. Stefanowski, 1989, "Rough classification in incomplete information systems", *Mathematical and Computer Modelling*, Vol. 12, pp. 1347-1357
8. R. Slowinski, J. Stefanowski, 1994, "Handling various types of uncertainty in the rough set approach" in: W. Ziarko (Ed.), *Rough sets, fuzzy sets and Knowledge Discovery (RSKD'93)*, Springer, Berlin
9. Chmielewski, M.R. Grzymala-Busse, J.W., 1996, "Global discretization for continuous attributes as preprocessing for machine learning" *International journal of approximate reasoning*, 15, 319-331
10. I. Graham, P.L. Jones, 1988, "Experts Systems-Knowledge, Uncertainty and Decision" (Chapman and Computing) 117-158
11. D. Dubois, H. Prade, 1992, "Putting rough sets and fuzzy sets together" *Intelligent decision Support, Handbook of Applications and Advances of the Rough Sets Theory*, pp. 203-232
12. Z. Pawlak, 1982, "Rough set" *International Journal of Computer and Information Sciences*, pp. 341-356
13. Zhang, H.C. Huang, S.H., 1994, "A fuzzy approach to process plan selection", *I. J. of Prod. Res.*, 32 (6), 1265-1279
14. Zadeh, L.A. Fu, K.S. Tanaka, K. Shimura, 1975, "Fuzzy sets and their applications to cognitive and decision processes", Academic Press, N.Y.

15. Zimmermann, H.J. , 1991, "Fuzzy set theory and its applications", Kluwer Academic Publishers, Boston,
16. J. Stefanowski, 2001, "Algorithms of Decision Rule Induction in Data Mining", Poznan University of Technology Press, Poznan, Poland.
17. J. Stefanowski, A. Tsoukias, 1999, "On the extension of rough sets under incomplete information", Proceedings of the 7th International Workshop on New Directions in Rough Sets, Data Mining, Granular-Soft Computing, RSFDGrC', Yamaguchi, Japan, 73–81.
18. J. Stefanowski, A. Tsoukias, 2001, "Incomplete information tables and rough classification", Computational Intelligence 17, 545–566.
19. M. Kryszkiewicz, 1995, "Rough set approach to incomplete information systems", Proceedings of the Second Annual Joint Conference on Information Sciences, Wrightsville Beach, NC, 194–197.
20. M. Kryszkiewicz, 1999, "Rules in incomplete information systems", *Information Sciences* 113 271–292.
21. Tzung-Pei Hong, Li-Huei Tseng, Been-Chian Chien, 2002, "Learning fuzzy rules from incomplete quantitative data by rough sets", IEEE International Conference, 1438-1443

ÖZGEÇMİŞ

Gölnur AVŞAR, 1982 yılında Adana’da doğdu. İlk ve orta öğretimimi Adana’da tamamladı. 2000 yılında Elazığ Mehmet Akif Ersoy Lisesi’nden mezun oldu. Aynı yıl Fırat Üniversitesi Mühendislik Faköltesi Bilgisayar Mühendisliği bölümünde lisans eğitimi almaya hak kazandı. 2004 yılında ilgili bölümden mezun olup, 2004 yılı güz döneminde aynı bölümde yüksek lisans eğitimi almaya başladı.