

ANALYSIS OF INSURANCE DATA BY FUZZY DATA MINING

Mehmet Ali ALAN*, Fuat ÇAMLİBEL, Ali Rıza İNCE*****

* Assoc. Prof.Dr. , Cumhuriyet University

** Lecturer., Cumhuriyet University

***Assoc. Prof.Dr. , Cumhuriyet University

E-mail: alan@cumhuriyet.edu.tr, fcamlibel@cumhuriyet.edu.tr , arince@cumhuriyet.edu.tr

Copyright © 2018 Mehmet Ali ALAN, Fuat ÇAMLİBEL, Ali Rıza İNCE. This is an open access article distributed under the Eurasian Academy of Sciences License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

In this study, data mining was studied using insurance data by fuzzy data mining method. By using insurance data, the algorithm of the fuzzy class which classifies this data most successfully as well as the classes that this algorithm will generate are tried to be determined. As a result of the study, it was found that the VQNN algorithm is the most successful algorithm in classifying the insurance data and the vehicle characteristics are more determinative than the driver characteristics in determining the accident status.

Keywords: Data mining, Fuzzy Logic, Classification, VQNN, Insurance

SİGORTA VERİLERİNİN BULANIK VERİ MADENCİLİĞİ İLE ANALİZİ

ÖZET

Bu çalışmada, bulanık veri madenciliği yöntemiyle sigorta verileri kullanılarak veri madenciliği yapılmıştır. Sigorta verilerinden yararlanarak, hem bu verileri en başarılı sınıflandıran bulanık sınıfa ait algoritma, hem de bu algoritmanın üreteceği sınıflar tespit edilmeye çalışılmıştır. Çalışmanın sonucunda VQNN algoritmasının sigorta verilerini sınıflandırmada en başarılı algoritma olduğu ve araç özelliklerinin kaza durumunu belirlemede sürücü özelliklerine göre daha belirleyici olduğu anlaşılmıştır.

Anahtar Kelimeler: Veri Madenciliği, Bulanık Mantık, Sınıflandırma, VQNN, Sigortacılık

1. Giriş

Kurumların veri tabanlarındaki büyük miktarlardaki verileri, klasik yöntemlerle işlemeleri ve bunlardan yararlı sonuçlar çıkarmaları oldukça güçtür. Bu verilerden anlamlı çakarsamalar yapmak için bilgisayar destekli çözümler sunmak gerekir. Bu bağlamda yapay zekâ uygulamalarının, veri madenciliği ile birlikte kullanılması, işletmelerde karar süreçleri için ihtiyaç duyulan bilgi üretimini kolaylaştırabilir.



Veri tabanlarındaki veriler, standart kullanım amaçları dışında, kurumların işine yarayacak, rasyonel karar almayı sağlayacak gizli bilgiler ya da ilişkiler barındırabilir. Günümüzde kurumlar büyük miktarlarda veri üretmekte, ancak bu veriler içinde anlamlı ve yararlı bilgiyi ortaya çıkarmakta zorluklar yaşamaktadırlar. Geleneksel istatistik yöntemlerle büyük boyuttaki veriyi çözümlemek kolay değildir. Bu nedenle verileri işlemek ve çözümlemek için özel yöntemlere gereksinim duyulmuştur. Veri madenciliği yöntemleri bu gereksinimi karşılamak üzere ortaya çıkmıştır (Özkan, 2008:IV).

Veri madenciliğinin popüler yöntemleri; sınıflandırma, kümeleme ve birliktelik kuralları madenciliğidir. Sınıflandırma, veri madenciliğinin önemli ve çok kullanılan yöntemlerinden biridir. Sınıflandırma ve tahmin, gelecek veri trendlerinin tahmininde veya önemli veri sınıflarının açıklanmasında kullanılan iki önemli veri analiz tekniğidir. Bu analizler büyük miktarlardaki verilerin daha iyi anlaşılmasında kullanılabilmektedir (Han and Kamber, 2006:285).

Pek çok yapay zekâ uygulaması, veri madenciliği ile birlikte kullanılmaktadır. Bunlardan birisi de bulanık mantıktır. Günlük hayatta farkında olmadan çok sık kullanılan bulanık mantık, özellikle pek çok verinin gözlem yoluyla elde edildiği, bilginin sağlıklı alınmadığı veya yeterli olmaması nedeniyle yaşanan problemlere çözüm bulması açısından son derece önemli bir kavramdır. Bulanık mantık sayesinde gerçek hayattaki kavramlar konusunda yapılmış olan sözel değerlendirmeleri rakamlara dökme yani bu değerlendirmeler üzerinden işlem yapabilme imkânı ortaya çıkmıştır (Paksoy 2013:201).

Bu çalışmanın amacı, kasko sigorta verileri üzerinde bulanık veri madenciliği tekniğini kullanarak, sınıflandırma analizi ile karar süreçlerinde yardımcı olabilecek bilgileri ortaya çıkarmaktır.

Çalışma üç bölümden oluşmaktadır. Birinci bölümde veri madenciliği, sınıflandırma, bulanık mantık ve VQNN algoritmasıyla ilgili tanımlayıcı bilgilere yer verilmiştir. İkinci bölümde konuyla ilgili yapılmış çalışmalar gözden geçirilmiş; son bölümde ise sigorta verileri üzerinde veri madenciliği yapılmıştır.

2. Bulanık Veri Madenciliği ve VQNN Algoritması

2.1. Veri Madenciliği

Veri madenciliği, hem yararlı hem de anlaşılabilir verilerle, alışılmamış şekilde, verileri özetleyen ve gizli ilişkileri ortaya koyan bir analiz yöntemidir (Larose, 2006). Bu yöntem, öncelikle bilinmeyen örüntülerin ortaya konması amacıyla bilimsel ve teknik veri araştıran, veri tabanındaki bilgi keşfi süreçlerinden biridir (Rokach and Maimon, 2005:2).

Veri madenciliği, disiplinler arası bir alandır, veri tabanları sistemleri, istatistikler, makine öğrenimi, görselleştirme ve bilgi bilimini de içeren bir dizi disiplinin birleşmesidir. Ayrıca, kullanılan veri madenciliği yaklaşımına bağlı olarak, diğer disiplinlerden yapay sinir ağları, bulanık veya kaba küme teorisi, bilgi gösterimi, endüktif mantık programlama veya yüksek performanslı hesaplama gibi teknikler uygulanabilir. Veri madenciliği sistemi, madencilik yapılacak veri türlerine veya belirli veri madenciliği uygulamasına bağlı olarak, uzamsal veri analizi, bilgi alımı, model tanıma, görüntü analizi, sinyal işleme, bilgisayar grafikleri, web



teknolojisi, ekonomi, işletme, biyo enformatik veya psikoloji gibi alanlarla birlikte kullanılabilir. (Han and Kamber 2006:29-30).

Veri madenciliği, verinin bütünü kullanması bakımından diğer istatistiksel yöntemlerden ayrılmaktadır. Bu yöntemle, geleneksel yollarla elde edilmiş küçük verilerle çalışma yerine daha kolay değerlendirme yapabilecek, yeni bağımsız veriler tercih edilebilmektedir (Weiss and Zhang, 2003:426).

Gartner Group’a göre veri madenciliği, istatistiksel ve matematiksel yöntemler kadar örüntü tanıma teknolojilerini kullanmasıyla, muazzam miktarlardaki depolanmış verileri eleyerek yeni anlamlı birliktelikler, örüntüler ve trendler keşfetme sürecidir (Larose, 2005:2).

Veri madenciliği, reklamcılık, biyo informatik, veri tabanı pazarlama, dolandırıcılık tespiti, e-ticaret, sağlık, güvenlik gibi alanların içinde olduğu, farklı alanlarda uygulanabilen, değişik bakış açısı ve çalışması ile yeni bilgiler ortaya koyar (Jain et al., 2011).

En yaygın veri madenciliği algoritmaları ve modelleri içinde karar ağaçları (sınıflandırma ağaçları olarak ta adlandırılır (Bramer, 2007:6)); diskriminant analizi, yapay sinir ağları, genetik algoritmalar, birliktelik kuralları, kümeleme, sınıflandırma, çoklu lineer regresyon, sıralı örüntüler ve zaman serileri tahmini, örüntü tanıma ve özelliklerinin belirlenmesi sayılabilir. Bu algoritmalar sıradan bilgiyi bulup çıkarmak ve bir yöneticinin kararlarını yönlendirebilen özel bilgiye ulaşmak için çeşitli alanlardan elde edilen verileri işlemek amacıyla kullanılır (Wu and Li, 2003, (Tadesse, 2009)). Sınıflandırma, regresyon ve zaman serisi analizleri gizli örüntülerin ortaya çıkarılmasında ve şekillendirilmesinde uygun iken; birliktelik kuralları, kümeleme ve sırasal keşif yaklaşımları, hava tahmini ve şiddeti araştırmak ve tanımlamak için yararlı araçlar olabilir (Tadesse, 2009).

2.2. Sınıflandırma

Pek çok pratik karar verme işlemi, bir sınıflandırma problemi olarak formüle edildiğinde sınıflandırma, günlük yaşamda sıklıkla başvurulmuş bir işlemdir. Sınıflandırma ile nesneler bölünerek ayrıştırılır, yani karşılıklı olarak özel ya da genel kategorilerden her biri bir sınıf olarak atanabilir (Bramer, 2007:23).

Sınıflandırma, farklı sınıflardaki, değişik öğeleri ayırma sürecidir. Bu sınıflar, iş kuralları, sınıf sınırları veya bazı matematiksel fonksiyonlar olabilir. Sınıflandırma işlemi, sınıflandırılmış olan ögenin, bilinen bir sınıf değeri ile özellikleri arasındaki bir ilişki üzerine bina edilebilir. Bu sınıflandırma tipi, “denetimli öğrenme” olarak isimlendirilir. Eğer bir sınıfın bilinen örnekleri yoksa bu sınıflandırma denetimsizdir. En yaygın denetimsiz sınıflandırma yaklaşımı “kümeleme”dir. Kümeleme teknolojisinin en yaygın uygulamaları, perakende ürünlerde birliktelik analizi (market sepet analizi) ve dolandırıcılık tespittir (Nisbet, et al., 2009: 235, Dong-Peng, et al., 2008:36).

2.3. Bulanık mantık



Doğada ya da günlük yaşamda her şey, sınırları kesin olarak belirlenmiş, kati ve kesin hesaplamalar yapılabilecek durumda değildir. Bunun bir sonucu olarak esnek ya da dereceli hesaplama söz edilmektedir. Esnek hesaplama (Soft Computing), asıl karakteristiği yanlışlık ve belirsizliğe tolerans tanıyan ve ona değişen bir ortamda ve düşük maliyetle sorunların kapsamlı bir şekilde çözülmesini sağlayan bir metodolojidir. Esnek hesaplamanın temelinde, kesin veya yanlış olan bir problemi yaklaşık çözümler bulmak için düşük maliyetli kabul edilebilir bir çözüm sunan bilgisayar yöntemlerinin geliştirilmesinde tekniklerin melezleştirilmesi yatar. Bulanık Mantık, Sinir Ağları ve Evrimsel Hesaplama gibi esnek hesaplama metodolojiler buna dâhil edilebilir (Jesus et al.,2005)

Yapay zeka, veri madenciliğinin bütün yöntemlerinde kullanılmaya başlanmıştır. Yapay zekâ yöntemleri arasında; yapay sinir ağları (ANN), kendi kendini düzenleyen haritalar, Hopfield ağı, genetik algoritmalar, evrimsel algoritmalar, bulanık sistemler, kaba setler, kural tabanlı sistemler, destek vektör makineleri, karar ağaçları, Bayes ve olasılık modelleri sayılabilir (Lappas 2009: 81).

Veri Madenciliğinin hedeflerinden biri de tahmini doğrulukla, kullanıcının potansiyel sonuçlarını ortaya koyabilmesidir. Bu bağlamda, bulanık mantık, kullanıcı için basit ve kolayca yorumlanabilecek belirsizlik ve yanlışlıkların ayarlanmasını sağlayan bir bilgi gösterimi aracıdır (Jesus et al.,2005).

İlk olarak Zadeh tarafından 1965 yılında tanıtılan bulanık mantık, bilişsel belirsizlik ile ilgilenmekte ve klasik küme teorisinde bulunan birçok sorunun üstesinden gelmeye çalışmaktadır (Rokach and Maimon 2008:159; Jesus et al.,2005).

Bulanık kümelere dayalı olan bulanık mantık genelde, insan düşüncesine özdeş işlemlerin gerçekleşmesini sağlamakla, gerçek dünyada sık sık meydana gelen belirsiz ve kesin olmayan verilerin modellenmesinde kullanılmaktadır. Klasik mantıkta bir önerme “Doğru” ya da “Yanlış” tır. Fakat gerçek dünyada olayların ne derece iyi veya yanlış olmasının belirlenmesi gerekmektedir (Nabiyev 2010:622).

Bulanık küme teorisi, akılcı sistemlerde basitliği ve insanın mantığına benzerliği nedeniyle giderek daha fazla kullanılmaktadır. Bulanık sistemler, endüstriyel proses kontrol sistemleri ve bunun gibi uygulamalar yanında, tahmin etmede, karar vermede, iklimlendirmede, otomobil kontrol sistemleri gibi mekanik kontrol sistemlerinde ve akıllı yapılarda da kullanılır (Paksoy vd., 2013:17, Hong 2014).

Klasik matematiksel modellerin tatmin edici sonuçlar üretmediği, karmaşık, net olarak açıklanamayan ve belirsiz sistemlerle uğraşmak için bulanık sistem modellemesi üzerinde çalışılmaktadır. Belli başlı bulanık sistemler bulanık kümeler ve bulanık mantık üzerine dayalıdır ve özellikle kavramları muğlak olan sistemler için kullanışlıdır (Türkşen 2015:2).

Bulanık küme, kesin geçişleri elimine ederek belirsizlik kavramını yeniden tanımlar ve evrendeki bütün bireylere üyelik derecesi değerini atayarak matematiksel olarak ifade eder. Bu derece, bulanık küme tarafından verilen kavram ile uyumludur ve benzer bir bireyin derecesine uyar. Böylece bireyler, bulanık küme içerisinde üyelik dereceleri tarafından gösterilen daha büyük ve daha küçük değerlere ait olabilirler. Bu üyelik dereceleri $[0,1]$ aralığında gerçel değerler ile ifade edilir (Nabiyev, 2010:622-624; Cao et al.,2005; Rokach



and Maimon 2008:160, Kantardzic,2003 :247). Bulanık mantık değerlerini $[0,1]$ aralığında alan sonsuz değerli mantıktır (Nabiyev 2010:625).

Bulanık küme teorisi, klasik küme teorisi gibi, birleşim, kesişim, tamamlayıcı ve dahil olma gibi işlemleri de kapsar; aynı zamanda, değiştirici konsantrasyon ve dilatasyon ve bağlaşıklık bulanık kümeleme gibi klasik bir karşılık olmayan işlemleri içerir (Rokach and Maimon 2008:163).

Bulanık kümeler, bir değişkenin tam değerinin bilinmediği herhangi bir durumla uğraşırken yararlıdır. Bulanık mantık, değişkenin değerinin bir tahmin veya olası değerlerinin dağılımı yerine, bir dizi değer üzerinde değişkenin alınma olasılığı ile ilgilenir. Bu şekilde, daha az varsayım yapar ve açıkça hem neyin bilindiğini hem de bilinmediğini gösterir (Mazlack, 2005)

2.4. Bulanık veri madenciliği

Bazı gerçek uygulamalarda, birçok veri niceliksel ve dilsel niteliktedir. Bulanık veri madenciliği, bu tür veriden bulanık bilgiyi bulması için önerilmektedir (Wu at al, 2012). Veri Madenciliğinde bulanık mantık kullanımı mantıklıdır. Çünkü bulanık modeller değişkenler arasında anlamlı ve yararlı ilişkiler kuran niteliksel modeller kümesiyle kullanıcıya yönelik verilerin bir tanımını temsil eder. Bulanık kümeler, kenarlara en yakın öğeleri yok saymadan veya aşırı vurgulamadan farklı anlam seviyeleri arasında esnek sınırlar belirlenmesini sağlar. Herhangi bir bilgi çıkarma sürecinde insan etkileşiminin bir bileşeni vardır ve bulanık kümeler bilgiyi dilsel bir biçimde temsil edilmesine, önceki bilgiyi kolayca birleştirip yorumlanabilir çözümler sunulmasına izin verir (Jesus et al.,2005).

Bulanık kümeler, sürekli sınırlarda sözel enformasyon tanelerini yakalama kavramlarıdır. Veri madenciliğine katkı sağlayan teknolojiler arasında doğal olarak uyum sağlamaktadır. İnsanlar rakamlar yerine enformasyon taneleri üzerinde işlem yapmaktadır. Enformasyon tanelerinin hacmi verideki bakılması gereken kesin noktaları ima etmektedir ve enformasyonun belirlenmesi düzeyinde ilişki kurulmasına yardımcı olmaktadır.

Bulanık kümeler, sistem karmaşıklığını yönetilebilir yapmada belirsizlikten istifade etmektedir. Bulanık mantık sadece eksik, gürültülü veya kesin olmayan veriyle uğraşmada değil aynı zamanda, veriden, geleneksel sistemlere göre daha kabiliyetli ve düzgün performans sağlayan belirsiz modeller geliştirmekte de yardımcı olmaktadır. Bulanık sistemler belirsizliği tolere edebildiğinden ve verideki sözel belirsizlikleri düzeltmede yardımcı olduğundan kesin girdilerin geçerli olmadığı veya çok pahalı olduğu durumlarda sağlam, gürültü model veya tahmin fırsatı sunabilmektedir

Bulanık mantık, bilgi tanelenedirmesiyle ilişkili olarak, yüksek düzeyde özetleme üzerine çalışma avantajı da sunmaktadır. Bu nedenle, bulanık mantık sınıflandırma yapan veri madenciliği sistemleri için kullanışlıdır. Bulanık mantık sınıflandırma için kullanıldığında özellik değerleri bulanık değerlere dönüştürülür (Koyuncugil 2006:96-97)

Bulanık veri madenciliğinde bulanıklık iki düzeyde görülebilir; ya veriler bulanıktır ya da bilgi bulanıktır. Veriler bulanık olduğunda, öğrenme algoritması bulanık kümeleri ele



almalıdır. Bilgi, belirsiz olabilir veya verilere ilişkin ek bilgi getiren bir kanı dağılımı oluşturabilir. Bu durumda, her veri, bir olasılık, bir inanç veya herhangi bir belirsizlik derecesi ile ağırlıklandırılabilir (Marsala ve Meunier 2015).

Belirsiz En Yakın Komşular Sınıflandırması

VQNN (Panigrahi and Patra, 2016) her sınıfın benzerliklerinin toplamına dayanır. Dilsel olan “en çok” ve “Bazı” niceleyicileri kullanır. Verilen bulanık niceleyiciler (Q_U , Q_I) C 'nin alt ve üst yaklaşık değerini, sırayla “en çok” ve “bazı” yı gösterir. VQNN her sınıfı hedef bir uzaklığa y aşağıdaki gibi atar.

i. NN yi belirle, y nin en yakın komşuluğu K .

ii. y 'yi C sınıfına atayın $(R \downarrow^{Q_U})(y) + (R \uparrow^{Q_I} C)(y)$, maksimaldir. Belirsizce niceleştirilmiş kaba kümelerin al ve üst sınırları aşağıdaki gibidir.

$$((R \downarrow^{Q_U} C)(y)) = Q_U \left(\frac{\sum_{x \in X} \min(R_{x,y}, C(x))}{\sum_{x \in X} R_{x,y}} \right)$$

$$((R \uparrow^{Q_I} C)(y)) = Q_I \left(\frac{\sum_{x \in X} \min(R_{x,y}, C(x))}{\sum_{x \in X} R_{x,y}} \right)$$

Operatörler Q_U ve Q_I bulanık niceleyiciler olup “en çok ve “bazı” yı gösterir. Gösterildiği şekilde bu $[0, 1] \rightarrow [0, 1]$ aralıkta artarlar;

$$Q_U(1) = Q_I(1) = 1 \text{ and } Q_U(0) = Q_I(0) = 0$$

3. Literatür Özeti

Konuyla ilgili literatürde farklı veri setleri üzerinden yapılmış çalışmalar bulunmaktadır. Bunlardan Fang and Pecheux (2009) Sinyalize kavşaklardaki servis kalitesinin algılanışı ile ilgili toparladıkları verilerle bulanık veri madenciliği analizi yapmışlar ve bulanık üyeliğin öznel algıya hitap etmeye uygun olduğuna vurgu yapmışlardır. Maksimovic et al (2014), gerçek zamanlı olarak kaydedilmiş yangın sensör verilerini kullanarak, çeşitli sınıflandırma algoritmaları ile analizler yapılmışlar ve bulanık veri madenciliği algoritmalarının mevcut verileri daha iyi performansla sınıflandırdığını ortaya koymuşlardır. Aburrou et al (2010), elektronik bankacılıktaki kimlik hırsızlıklarının ortaya çıkarılmasında bulanık veri madenciliği yöntemini kullanarak, kimlik avı türlerini sınıflandırmışlardır. Simha and Iyengar (2006), bir telekom operatöründen aldıkları verileri bulanık veri madenciliği ile analiz etmişler ve müşteri ilişkileri yönetiminde müşteri sadakatinin artırılmasında bu yöntemin oldukça umut verici sonuçlar doğurduğuna vurgu yapmışlardır. Jing (2009), insan kaynakları verilerini kullanarak, bu veriler üzerinde bulanık veri madenciliği ile analizler yapmış ve bu yöntemin işletmelerin insan kaynaklarını değerlendirmede ve seçiminde hızlı ve etkili bir biçimde değerlendirmesine yardımcı olabileceğini ortaya koymuştur. Hüllermeier (2011), bulanık küme teorisinin, makine öğrenmesi, veri madenciliği ve ilgili alanlarda yapabileceği olası katkıları irdelemiş ve bu bağlamda, bulanık yöntemlerin veri içindeki belirsiz kalıpları



temsil etmedeki bazı avantajlarına vurgu yapmıştır. Mollazade et al (2009), hidrolik pompa sinyalleri veri seti üzerinde veri madenciliği yapmış ve bulanık veri madenciliğinin arıza teşhis potansiyeline sahip olduğunu vurgulamışlardır. Literatür taramalarında da görüldüğü gibi daha önce bu konuda yapılan bir çalışmanın olmaması bu çalışmanın özgünlüğünü ortaya koymaktadır.

4. Veri Madenciliği ve Analizi

4.1. Veri Seti ve Yöntem

Bu çalışmada Sivas il merkezinde faaliyet gösteren bir sigorta acentesine ait 2016 yılında yaptığı 500 adet kasko sigortası verileri kullanılmıştır. Analizde uygulamaya temel teşkil edecek, sınıflarla birlikte toplam 9 sütun başlığı tanımlanmıştır. Excel makroları kullanılarak veriler, bulanık mantık algoritmalarının formatına uygun biçime getirilerek veri ambarı hazırlanmıştır. Gerekli dönüşümler yapıldıktan sonra veriler “sigorta.arff” adlı metin dosyasına yazdırılmıştır. Kasko sigortası değişkenleri, aldığı değerler ve sınıflar tablo 1’de verilmiştir.

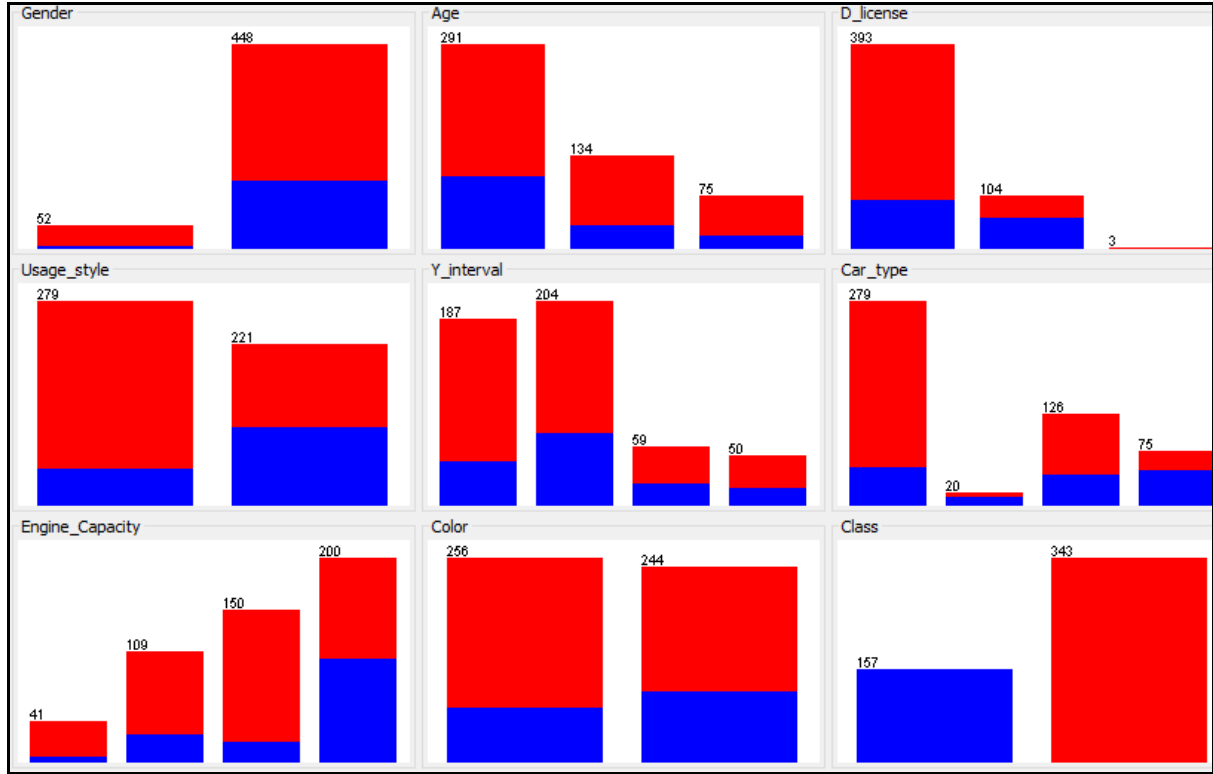
Tablo 1. 0,1,2 ve 3 Değerlerinin Karşılık Geldiği Değerler

DEĞİŞKENLER	DEĞERLER			
	0	1	2	3
Cinsiyet	Bayan	Erkek	-	-
Yaş	Genç	Orta yaşlı	Yaşlı	-
Ehliyet	B	C	F	-
Kullanım tarzı	Hususi	Ticari	-	-
Model yaşı	2014-2016	2011-2013	2008-2010	≤2007
Cinsi	Otomobil	Minibüs	Kamyonet	Kamyon
Motor hacmi	<1300	1300-1599	1600-2000	>2000
Renk	Açık renk	Koyu renk	-	-
SINIFLAR	Kaza var	Kaza yok	-	-

4.2. Uygulama

Yapılan çalışmada Waikato Üniversitesince geliştirilmiş olan WEKA Programının (Waikato Environment for Knowledge Analysis) 3.7.2 sürümü kullanılmıştır (Wekafull paket). WEKA Programı, açık kaynak kodlu bir yazılımdır. Bu program pek çok sınıflandırma, kümeleme ve birliktelik kurallarına ait algoritmayı desteklemektedir. WEKA, metin tabanlı pek çok dosya tipinin yanı sıra, veri tabanlarını ve verilerin olduğu URL adreslerini de desteklemektedir. WEKA programının bu sürümünde, bulanık mantık ve genetik algoritma desteği de bulunmaktadır.

Uygulamada kullanılan veri setindeki her bir sütun bir değişken olarak tanımlanmıştır. Bu değişkenlerin görselleri ve değerleri şekil 1’de sunulmuştur.



Şekil 1. Değişken Görselleri

Mevcut veri seti ile 10 kat çapraz doğrulama ve tam eğitilmiş set kullanılarak yapılan uygulama sonucunda izleyen tablodaki sonuçlar algoritmaların başarımlarına göre sıralanarak sunulmuştur:

Tablo 2. Bulanık Sınıflayıcılar ile Elde Edilen Analiz Sonuçları

Algoritmalar	Doğru sınıflandırılan Örnek	Kappa İstatistiği	Ortalama Mutlak Hata	Ortalama Hata Karekök	Görelî Mutlak Hata %	Görelî Hata Karekök %	TP Oranı	FP Oranı	F-Ölçütü
VQNN	380	0.3948	0.2939	0.4474	68.1785	96.3884	0.76	0.397	0.747
OWANN	376	0.3699	0.3583	0.4362	83.0953	93.9836	0.752	0.414	0.737
FURIA	373	0.3435	0.2575	0.4887	59.7273	105.283	0.746	0.437	0.727
FuzzyOwnershiN	361	0.3081	0.3262	0.449	75.6677	96.7364	0.722	0.435	0.71
VQQuickRules	243	0.3804	0.2477	0.4898	90.7216	133.973	0.752	0.384	0.748
DiscernibilityNN	322	0.2827	0.4625	0.4759	107.265	102.528	0.644	0.318	0.657
VQRules	320	0.1691	0.3534	0.5854	83.4887	127.311	0.652	0.48	0.647



								7	
FuzzyRoughNN	29 7	0.237 9	0.444 6	0.471 8	103.11 3	101.64 3	0.59 4	0.3	0.60 5

Yapılan uygulama çalışmasında WEKA programınca desteklenen pek çok bulanık algoritma denenmiş, başarımlar dereceleri Tablo 2’de verilmiştir.

VQNN algoritması, 380 doğru sınıflandırılmış örnek derecesiyle en başarılı algoritma olmuştur. Bu algoritmanın, sınıflar arası uyumu veren kappa istatistiği 0.3948, birinci sınıftaki doğru olarak sınıflandırılmış kayıtların sayısını veren TP (True Positive) oranı 0.76, birinci sınıfta sınıflandırılmış, ikinci sınıftaki kayıtların sayısını veren FP (False Positive) oranı 0.397 olarak elde edilmiştir. Kesinlik ve duyarlılığın harmonik ortalaması olan F-ölçütü ise 0.747 olarak bulunmuştur. VQNN algoritmasından sonra en başarılı algoritma ise 376 doğru örnek sınıflandırmasıyla OWANN algoritması olmuştur. Bu algoritmaların kappa istatistiği 0.3699, TP oranı 0.752, FP oranı, 0.414 ve F-Ölçütü 0.737 olarak bulunmuştur. Bu algoritmaları sırasıyla FURIA, FuzzyOwnershipNN, VQQuickRules, DiscernibilityNN, VQRules ve FuzzyRoughNN algoritmaları izlemektedir. F-ölçütü aşağıdaki formülle elde edilmektedir:

$$F - \text{Ölçütü} = \frac{2 \times \text{Duyarlılık} \times \text{Kesinlik}}{\text{Duyarlılık} + \text{Kesinlik}}$$

Formülde kullanılan kesinlik ve duyarlılık ise aşağıdaki şekilde hesaplanmaktadır:

$$\text{Kesinlik} = \frac{TP}{TP + FP}$$

$$\text{Duyarlılık} = \frac{TP}{TP + FN}$$

Tablo 3’de araç özellikleri (Model Yılı, Araç Tipi, Motor Hacmi, Renk) ile sürücü özelliklerine (Cinsiyet, Yaş, Sürücü belgesi türü, Kullanım tarzı) ait verilerin sınıflandırma performansları verilmiştir.

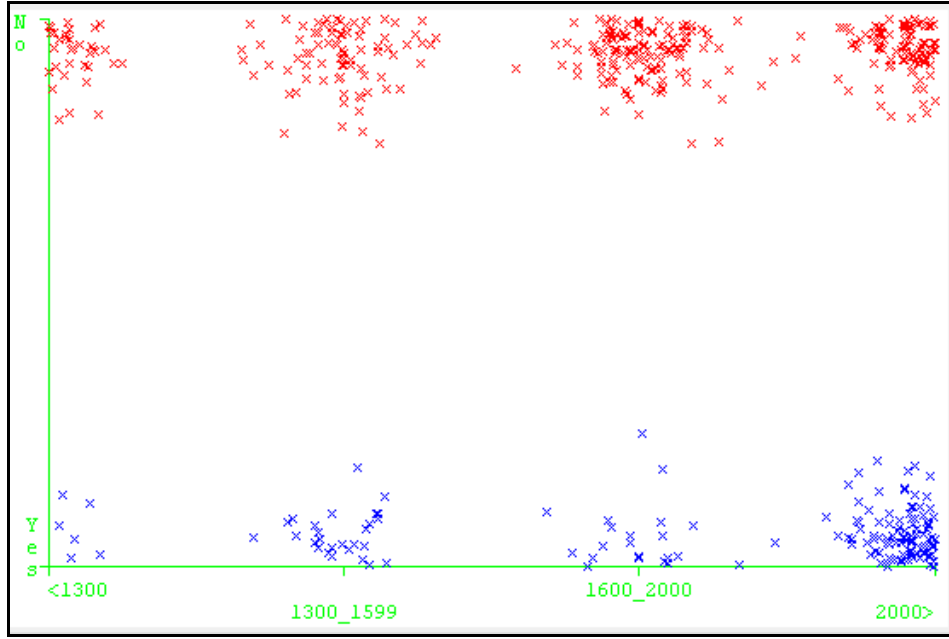
Tablo 3. Sürücü/Araç özellikleri sonuçları

Algoritmalar	Doğru sınıflandırılan Örnek	Kappa İstatistiği	Ortalama Mutlak Hata	Ortalama Hata Karekök	Görelî Mutlak Hata %	Görelî Hata Karekök %	TP Oranı	FP Oranı	F-Ölçütü
Araç Özellikleri	365	0.3153	0.311	0.4644	72.1355	100.061	0.73	0.441	0.714
Sürücü Özellikleri	357	0.3011	0.3164	0.473	73.396	101.907	0.714	0.428	0.705

Tablo 3’ten de anlaşılacağı gibi Kaza Var/Kaza Yok sınıflandırma temelinde yapılan analizlerde araç özelliklerinin sürücü özelliklerine göre daha belirleyici olduğu anlaşılmaktadır.

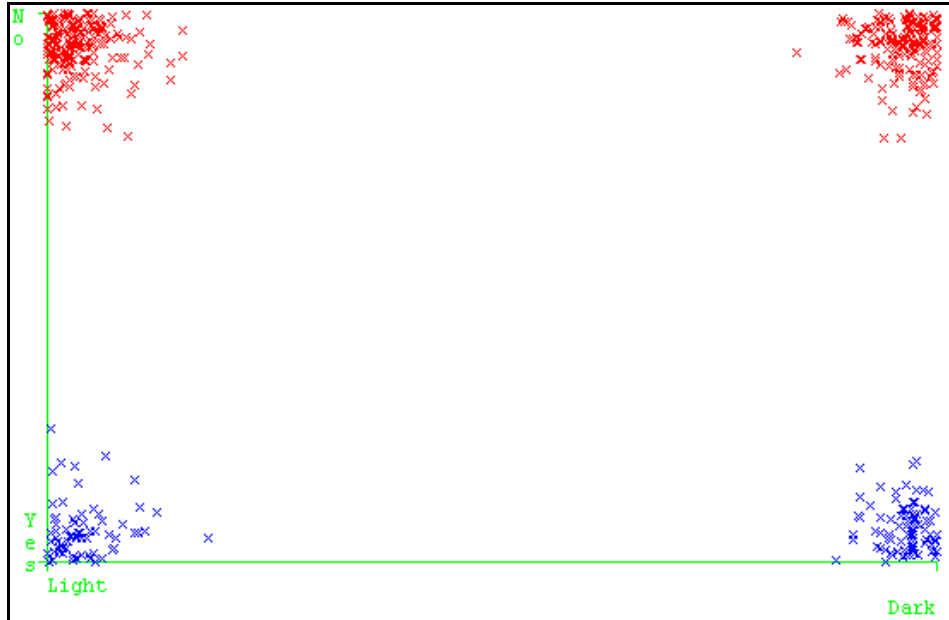


Veri madenciliği görsel destek bakımından da oldukça yararlı bilgiler sunabilir. İzleyen şekillerde sigorta verilerine ait çeşitli görseller sunulmuştur.



Şekil 2. Motor Kapasitesi Kaza İlişkisi

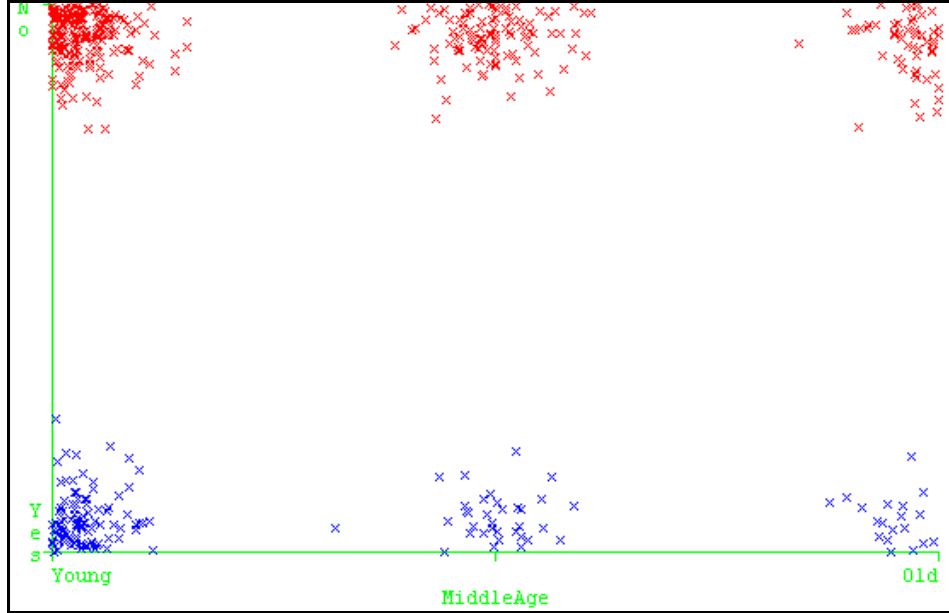
Şekil 2’de motor kapasitesi ile kaza yapma arasındaki ilişki görseli sunulmaktadır. Şekilden de anlaşılabileceği gibi motor gücü arttıkça kaza sayısı da artmaktadır.



Şekil 3. Renk/ Kaza Durumu

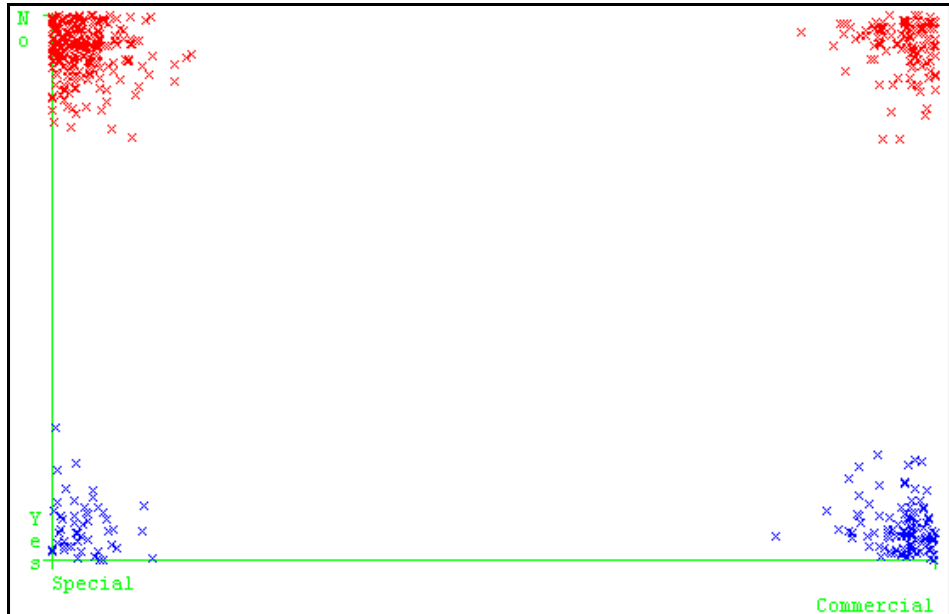


Şekil 3’de araçların renklerinin kaza yapma durumunu nasıl etkilediği görsel sunulmaktadır. Şekilden de anlaşılabacağı gibi arabanın renginin açık ya da koyu olması kaza yapma durumu belirgin şekilde etkilememektedir.



Şekil 4. Yaş/Kaza Durumu

Şekil 4’te sürücülerin yaşları ile kaza arasındaki ilişkiyi veren görsel sunulmaktadır. Şekilden de anlaşılabacağı gibi genç yaştaki sürücülerin kaza sayısı diğer yaş gruplarına göre daha fazladır.



Şekil 5. Kullanım Tarzı/Kaza Durumu



Şekil 5’de araçların kullanım tarzı ile kaza yapma durumu arasındaki durumu gösteren görsel sunulmaktadır. Şekilden anlaşılabacağı gibi ticari olarak kullanılan araçlar, hususi kullanıma göre daha fazla kazaya karışmaktadır.

Araç ve sürücü özelliklerindeki diğer değişkelerin görsellerinden somut şeyler çıkarmak zor olduğundan, onlara ait görseller verilmemiştir.

5.Sonuç

Bu çalışmada, bulanık veri madenciliği yöntemi kullanılarak sigorta şirketine ait kasko verileri analiz yapılmıştır. Veri madenciliği, gizli, önemli, önceden bilinmeyen, yararlı bilgileri ortaya çıkaran bir veri analiz tekniğidir. Bu yöntemle, alışlagelmiş analiz tekniklerinden farklı olarak, yalnızca sayısal verilerle değil, sayısal olmayan veriler ile de analizler yapılabilmekte ve gizli örüntüler ortaya çıkarılabilmektedir. Bulanık mantık ise kesin kurallar yerine esnek bilişim uygulamalarına izin veren bir yöntemdir. Bulanık veri madenciliği, standart veri madenciliğine göre verileri sınıflandırmada ve kümeleme daha başarılı sonuçlar üretebilmektedir. Mevcut verilerle bulanık veri madenciliğine ait WEKA programınca desteklenen pek çok algoritma denenmiş ve en başarılı algoritma olarak VQNN algoritması bulunmuştur. Ayrıca araç özelliklerinin kaza durumunu belirlemede, sürücü özelliklerine göre daha belirleyici olduğu anlaşılmıştır. Bu çalışmada başarılı olarak belirtilen algoritma ile sigorta şirketlerinin aktüerya bölümleri aracın motor kapasitesi, yaşı, rengi gibi araç özelliklerini de dikkate alarak risk primini hesaplayabilirler. Sonuç olarak sigorta şirketleri, hasar prim oranını daha doğru hesaplayacak ve şirketlerin teknik karlılıkları olumlu yönde etkilenecektir.

REFERENCES

- Aburrous, Maher, Hossain, M.A., Dahal, Keshav and Thabtah, Fadi (2010), “Intelligent phishing detection system for e-banking using fuzzy data mining”, **Expert Systems with Applications** 37, p. 7913–7921
- Bramer, Max (2007), **Principles of Data Mining**, Springer, London
- Cao, Yingjun, Wang, Paul P. and Alade Tokuta (2005), “Gene Regulating Network Discovery”, **Intelligent Data Mining**, Vol.5, p.49-78.
- Dong-Peng Yang, Li Jin-Lin, Lun Ran and Chao Zhou, (2008), Applications of Data Mining Methods in the Evaluation of Client Credibility, **Applications of Data Mining in E-Business and Finance** C. Soares et al. (Eds.), IOS Press, Amsterdam, p.35-43
- Fang, Clara Fang and Pecheux, Kelly Klaver (2009), “Fuzzy Data Mining Approach for Quantifying Signalized Intersection Level of Services Based on User Perceptions”, **Journal Of Transportation Engineering**, Jun 2009, p. 349-359
- Giudici, Paolo and Figini, Silvia, (2009), Applied Data Mining For Business and Industry, Second Edition, Wiley Publication, West Sussex
- Han, Jiawei and Kamber, Micheline, (2006), **Data Mining: Concepts and Techniques**, Second Edition, Morgan Kaufmann Publications, San Francisco
- Hong, Tzung-Pei, Lin, Chun-Wei, And Lin, Tsung-Ching (2012), “The Mffp-Tree Fuzzy Mining Algorithm To Discover Complete Linguistic Frequent Itemsets”, **Computational Intelligence**, Volume 30, Number 1, p. 145-166



- Hüllermeier, Eyke (2011), “Fuzzy sets in machine learning and data mining”, **Applied Soft Computing** 11 (2011), p. 1493–1505
- Jain, Y. K., Yadav, V. K. and Panday, G. S., (2011), “An Efficient Association Rule Hiding Algorithm for Privacy Preserving Data Mining”, **International Journal On Computer Science And Engineering**, Vol. 3 No. 7, p. 2792-2798.
- Jesus, M.J. del, Gonzalez, P., Herrera, F., and Mesonero, M.(2005), “Evolutionary Induction of Descriptive Rules in a Market Problem”, **Intelligent Data Mining**, Vol.5, p. 267-292.
- Jing, Han (2009), “Application of Fuzzy Data Mining Algorithm in Performance Evaluation of Human Resource”, **2009 International Forum on Computer Science-Technology and Applications**, p.343-346
- Kantardzic, Mehmed, (2003). Data Mining: Concepts, Models, Methods, and Algorithms, John Wiley & Sons J.B. Speed Scientific School, University of Louisville IEEE Computer Society.
- Koyuncugil, Ali Serhan (2006), Bulanık Veri Madenciliği Ve Sermaye Piyasalarına Uygulanması” yayınlanmamış doktora tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara,
- Lappas, Georgios (2009), “Machine Learning and Web Mining: Methods and Applications in Societal Benefit Areas”, **Data Mining Applications for Empowering Knowledge Societies**, Edited by Hakikur Rahman, Information Science Reference, New York, , p.76-95
- Larose, Daniel T., (2005), **Discovering Knowledge In Data**, Wiley Publication, New Jersey
- Maksimović ,Mirjana, Vujović, Vladimir and Milošević, Vladimir (2014), “Applying Fuzzy Logic and Data Mining Techniques in Wireless Sensor Network for Determination Residential Fire Confidence”, **Journal of Engineering Science and Technology**, Review 7 (4), p. 89- 96
- Marsala, Christophe and Bouchon-Meunier, Bernadette (2015), “Fuzzy data mining and management of interpretable and subjective information”, **Fuzzy Sets and Systems** 281, p. 252–259
- Mazlack, Lawrence J. (2005), “Granular Nested Causal Complexes”, **Intelligent Data Mining**, Vol.5, p. 23-48
- Mollazade, K, Ahmadi, H, Omid, M and Alimardani, R (2009), An Intelligent Model Based on Data Mining and Fuzzy Logic for fault diagnosis of external gear hydraulic pumps, **Intelligent Fault Diagnosis**, Insight Vol 51 No 11 November 2009, p.594-600
- Nabiyev, Vasif V. (2010), **Yapay Zekâ**, Seçkin Yayınları, Ankara
- Nisbet, R., Elder J., and Miner G., (2009), Handbook of Statistical and Data Applications, Elsevier Inc, Burlington.
- Özkan, Yalçın (2008), **Veri Madenciliği Yöntemleri**, Papatya Yayınları, İstanbul
- Paksoy, Turan, Yapıcı Pehlivan, Nimet ve Özceylen, Eren (2013), **Bulanık Küme Teorisi**, Nobel Yayınları, Ankara



- Panigrahi, Ashalata and Patra, Manas Ranjan (2016), “Fuzzy Rough Classification Models for Network Intrusion Detection”, **Transactions on Machine Learning and Artificial Intelligence**, Volume 4 No 2 April (2016); pp: 7-22
- Rokach, Lior and Maimon, Oded (2008), **Data Mining with Decision Trees**, World Scientific, New Jersey
- Simha, Jay B., Iyengar S.S (2006), “Fuzzy data mining for customer loyalty analysis”, **9th International Conference on Information Technology (ICIT'06)**, 2006 IEEE
- Tadesse, T., Wardlow, B. And Hayes, M.J. (2009), The Application of Data Mining for Drought Monitoring and Prediction, **Data Mining Applications for Empowering Knowledge Societies**, Edited by Hakikur Rahman, Information Science Reference, New York, p.280-291
- Türkşen, İ. Burhan (2015), **Dereceli(Bulanık) Sistem Modelleri**, Abaküs Yayınları, İstanbul
- Wu, Min-Thai, Hong, Tzung-Pei and Lee, Chung-Nan (2012), “A continuous ant colony system framework for fuzzy data mining”, **Soft Comput (2012)** 16, p. 2071–2082
- Wu, Tong and Li Xiangyang (2003), Data Storage and Management, **The Handbook of Data Mining**, Edited by. Nong Ye, Lawrence Erlbaum Associates Publishers. London, pp.393-407
- Weiss, Sholom M. And Zhang, Tong (2003), Performance Analysis and Evaluation, **The Handbook of Data Mining**, Edited by. Nong Ye, Lawrence Erlbaum Associates Publishers. London, p.436-439