

Belirsizlik Koşullarında Fuzzy Rough Algoritması: Kredi Skorlama'da Bir Uygulama

Fuzzy Rough Set Algorithm under Uncertainty: An Application in Credit Scoring

Nilgün GÜLER BAYAZIT⁽¹⁾, Yasemen UÇAN⁽²⁾

Öz: Günümüzün artan rekabetçi ortamı, bankaların tüketicilerin artan kredi taleplerine çabuk ve hızlı karar vermelerini gerektirmektedir. Bu amaçla bankalar müşterilere kredi verirken karar vermelerine yardımcı olan istatistik ya da makina öğrenmesi tabanlı kredi skorlama modelleri kullanmaktadırlar. Çalışmada kredi skorlama modellerindeki özellikle belirsizlik konusundaki eksikliği gidermek için bulanık-kaba küme tabanlı bir kredi skorlama modeli önerilmektedir. Bulanık ve kaba kümeler teoremine dayanan yöntem veri kümesindeki örneklerin bulanıklık benzerliklerini hesaplayarak tüketicinin kredi almaya olan uygunluğunu belirleyen kararlar vermektedir. Model sonuçları, yaygın olarak kullanılan diğer kredi skorlama yöntemleriyle karşılaştırılmış ve önerdiğimiz kredi skorlama modellerinden daha iyi olduğunu göstermiştir.

Anahtar Kelimeler: Kredi Skorlama, Bulanık Kümeler, Kaba Kümeler

Abstract: *The increasing competitive environment in today's world necessitates a prompt response from the banks to the increasing credit demands of consumers. To serve this purpose, the banks employ statistics or machine learning based credit scoring models that help them in their decision making to give credit to their clients. In this work, a fuzzy rough set based credit scoring model is proposed to remedy the deficiency due to the uncertainty in the credit scoring models. The method is based on fuzzy and rough set theory and makes decisions to determine the suitability of a consumer to receive credit by evaluating the fuzzy similarities of the samples in the data set. The results obtained with the model has been compared with other widely used credit scoring methods and has shown the superiority of our proposed credit scoring method.*

Key words: Creding Scoring, Fuzzy Set, Rough Set

JEL Classifications: C38,C63

1. Giriş

Bankacılıkta, risklerin etkin bir şekilde yönetilebilmesi için; risk tanımlamalarının yapılarak risklerin ölçülmesi için gerekli uygulamaların ve takip-değerlendirme sistemlerinin gerçekleştirilmesi gerekmektedir (Kavcıoğlu, 2014:11). 2004 yılında Basel bankacılık komitesinin yayınladığı Basel II uzlaşısında, finansal piyasalarda kalıcı istikrarın sağlanması amacıyla bankaların müşterilerinin kredi değerliliğini ve işlemlerinin risklilik düzeyini sınıflandırmaya tabi tuttıkları derecelendirme sistemlerine sahip olmalarını gerektirmektedir (Ayanoğlu ve Ertürk, 2007:76). Kredi riskinin ölçülmesinde, kredi skorlama (KS) modelleri, gelişmiş modeller ve uluslararası kredi riski ölçüm yazılımları kullanılmaktadır. Bankalar kredi risklerini ölçmek için çeşitli içsel kredi değerlendirme modellerine başvurmaktadır.

⁽¹⁾Yıldız Teknik Üniversitesi, Kimya-Metalurji Fakültesi, Matematik Mühendisliği Bölümü; guler@yildiz.edu.tr

⁽²⁾ Yıldız Teknik Üniversitesi, Kimya-Metalurji Fakültesi, Matematik Mühendisliği Bölümü; ucan@yildiz.edu.tr

Literatürde çeşitli yöntemler ile KS ölçümleri yapılmaktadır. Bu metotlar ile kredi borçlusunun özellikleri, temerrüt ihtimalinin ortaya çıkarılması veya borçluların çeşitli temerrüt gruplarına ayrılması için test edilmektedir. Bu modellerde öncelikle değerlendirme kriterleri belirlenir daha sonra istatistiki metotlar ile temerrüt olasılığı hesaplanır ya da temerrüt riski grupları belirlenir. Bu modeller genellikle üç başlık altında toplanabilir. Bunlar; lineer olasılık modeli, logit ve probit modeller ve lineer diskriminant modelleridir. Bu KS modelleri, kredi riskinin ölçümü bağlamında önemli avantajlar sağlamakta ise de bu modellerin uygulanmasında birtakım eksiklikler söz konusudur. Bu modeller, bazı önemli faktörleri analize dahil etmemekte diğer yandan bu kredi skorlama modellerinde, tahmin edilen ağırlıkların ve seçilen değişkenlerin çok kısa dönemde sabit kalacağı ve değişmeyeceği varsayılmaktadır (Kavcıoğlu, 2014:14). Üstelik bu modellerde belirsizlik yeterince analiz edilmemektedir.

Bu çalışma KS modellerinde ki özellikle belirsizlik konusunda ki eksikliği gidermek için bulanık-kaba küme tabanlı bir KS modeli önermektedir. Kaba küme teorisine dayanan *belirsiz nicelikli en yakın komşuluk modeli* (vaguely quantified nearest neighbour-VQNN), belirsizlikle baş edebilmek için mükemmel bir çerçeve sunmaktadır. Yöntem, küme teorisine dayandığı için optimizasyon tabanlı yöntemlerde olduğu gibi (yapay sinir ağları (YSA), destek vektör makinaları (DVM) vb.) iyi bir sonuç elde etmek için erken durdurma kriterine ve kullanıcı tarafından belirlenen parametrelere ihtiyaç duymamaktadır. VQNN tekniğinin diğer bir avantajı ise basitliği ve kolay anlaşılabilir olmasıdır. Sadece veri kümesindeki örneklerin bulanıklık (fuzzy) benzerliklerini hesaplayarak buna dayanan kararlar verir. Verideki saf hata terimi, tolere edebilir. Başlangıç tahminlerine ihtiyaç duymaz. Özellikle kesin olmayan belirsiz verilerle baş edebilir. Bu anlamda belirsizlik modellerde sorun olmaktan çıkmaktadır.

Makale şu şekilde düzenlenmiştir. KS literatürü ikinci bölümde verilmiştir. Üçüncü bölümde veri tanıtılmış ve bulanık-kaba en yakın komşuluk (BK-EYK) algoritması verilmiştir. Dördüncü bölümde sonuçlar tartışılmıştır. Son bölümde genel değerlendirme sonuçları verilmiştir.

2. Kredi Skorlama Literatürü

KS yalnızca tüketicinin kredi alma konusundaki yeterliliğini ölçümleyen bir sistem değildir. Aynı zamanda, finansal kuruluşların kaliteli kredi başvuruları elde etmesini, kredi kullanan tüketicilerin daha sonra da elde tutulmasını ve bir portföy oluşturarak tüketicilerin davranışlarının kontrol edilmesini de sağlamaktadır. KS modelinde bireysel nitelikli tüketicilerin kredi riskinin belirlenmesinde, finansal göstergeler (borçlunun toplam aktifleri, bürüt geliri vb), demografik göstergeler (yaş, cinsiyet, medeni durum vb.), iş durumu göstergeleri (borçlunun çalıştığı işin türü, aynı işyerinde kaç yıldır çalıştığı vb) ve finansal davranışıyla ilgili göstergeler kullanılmaktadır.

KS modelleri, bilgisayar teknolojisindeki gelişmelerin etkisinde (Mester, 1997) hızla gelişmiş olsa da literatür genellikle Beaver (1966) çalışmasına dayandırılır. Beaver (1966) derecelendirme konusunda öncü çalışmadır. Beaver finansal verileri oranlayarak iflas etmesi olası firmaları önceden belirlemeyi amaçlamış ve 30 farklı oranı 6 ana grupta toplamıştır. Bu öncü çalışmanın izinde Altman (1968) birden çok

boyuttan hareket ile derecelendirme yaptığı çalışmasında çoklu diskriminant analizinden hareket etmiştir. Altman “Z” modelinin geliştirildiği çalışmada, 33 iflas etmiş ve 33 iyi durumdaki firmanın verileri kullanılmış 5 farklı rasyo aracılığı ile; 1 yıl öncesine ait veriler ile %95 doğruluk oranıyla, 2 yıl öncesine ait veriler ile %72 doğruluk oranıyla, 3 yıl öncesine ait veriler ile %48 doğruluk oranıyla, 4 yıl öncesine ait veriler ile %29 doğruluk oranıyla ve 5 yıl öncesine ait veriler ile %36 doğruluk oranıyla sonuçlar elde edilmiştir. Meyer-Fiber (1970) ise ABD’de 1948 ile 1965 yılları arasında 39 iflas etmiş 39 karlı banka verilerinden hareket ile regresyon analizi uyguladıkları çalışmalarında iflasa ilişkin %79 oranında doğru öngöründe bulunmuşlardır. Bhatia (1988) ise Hindistan’da 1976 ile 1995 yılları arasında faaliyet göstermiş 18 kötü durumda ve 18 kârlı durumdaki firmayı analiz ettiği çalışmada diskriminant analizinden yararlanmıştır. 7 rasyonun kullanıldığı çalışmada %87.1 oranında doğru tahmin yapılmıştır. Desai (1996) YSA, doğrusal diskriminant analizi (DDA) ve lojistik regresyon (LR)’dan oluşan üç farklı algoritma kullanmış ve kötü sınıfa ait kredilerin sınıflandırılmasında YSA ile elde edilen modellerin daha iyi sonuçlar verdiğini belirlemiştir. Ancak, iyi ve kötü gruba ait kredilerin sınıflandırılmasında LR ve YSA yakın sonuçlar vermiştir.

Bu çalışmaların izinde farklı yöntemler kullanılarak daha etkin sonuçlar elde edilmeye çalışan birçok çalışma yapılmıştır. 2010 sonrası ise LR yöntemini, k-en yakın komşu (k-EYK), DVM ve YSA kullanarak kredi skoru elde etmeye yönelik çalışmalar izlemiştir. Nitekim, Ceren (2010) DVM ve LR yöntemlerinden hareket etmiş ve DVM algoritmasının %75 oranında doğruluk değeri ile %71,8 oranında doğruluk değeri veren LR metoduna üstün olduğunu belirlemiştir. Tabagari (2015), kredi talebinde bulunan 500 müşterinin bilgisinden oluşan veri setine LR yöntemini uygulamış ve %82,8 oranında doğruluk elde etmiştir. Sousa ve Reginaldo (2014) müşterilerin kredi profillerini iyi veya kötü olarak gruplamak için C4.5 karar ağacı ve YSA metodlarından yararlanmıştır. YSA %95,58’lik bir doğruluk oranı verir iken C4.5 karar ağacı %90.07 doğruluk vermiştir. Her iki oranda yüksek bir doğruluk oranı verse de YSA daha başarılı bulunmuştur. Demirbulut vd. (2017) en uygun KS modelinin tespiti için makine öğrenimi ve istatistiksel sınıflandırma algoritmaları kullanmışlardır. k-EYK, C4.5 karar ağacı, YSA, DVM, LR, probit regresyon, poisson regresyon ve genelleştirilmiş katkı modeli (GKM) yöntemlerinin karşılaştırıldığı çalışmada algoritmaların herbiri müşterinin kredi skorunun tespit edilmesine imkân sunmuştur.

3. Veri ve Matematiksel Altyapı: Bulanık-Kaba Küme Tabanlı En Yakın Komşuluk Sınıflandırıcısı

3.1. Veri

Çalışmada UCI Machine Learning Repository’deki açık kaynaklı German credit (kredi-g) ve Australian credit approval (kredi-a) kullanılmıştır. Veriler https://archive.ics.uci.edu/ml/machine-learning-databases/credit-screening/crx_data ve <https://archive.ics.uci.edu/ml/machine-learning-databases/statlog/german/> adreslerinden elde edilmiştir (Dua ve Karra Taniskidou, 2017).

Her iki veri kümesi, hem nümerik hem de sembolik (nominal) değerler içeren özniteliklerden oluşmaktadır. Kredi-a veri kümesi, 15 öznitelikten oluşmaktadır. Veri kümesi 307 (%44,5) kredi vermeye değer, 383 (%55,5) kredi verilmesi sakıncalı toplam 690 müşteri kayıtlarından oluşmaktadır. Veri güvenliği açısından öznitelik

isimleri anonim hale getirilmiştir. (A1, A2,..., A14). 20 öznelikten oluşan Kredi-g veri kümesi ise 700(%70) kredi vermeye değer, 300'ü (%30) kredi verilmesi sakıncalı 1000 müşteri kaydından oluşmaktadır.

3.2. Matematiksel Altyapı

Bu bölümde öncelikle kaba küme ve bulanık küme kavramları tanıtılıp daha sonra BK-EYK algoritması verilecektir.

3.2.1. Kaba Kümeler (Rough Sets)

Pavlak (1982) tarafından önerilen kaba kümelerde sınıflandırma yapılabilmesi için ilk olarak ayırtedilemezlik bağıntısı ile denklik sınıflarının oluşturulması gerekmektedir. Bunun için öncelikle aşağıdaki tanımların verilmesi gerekir.

$E = \{e_1, e_2, \dots, e_n\}$ sonlu bir küme, $P(E)$ kuvvet kümesi, X özneliklerin kümesi olmak üzere $BS = (E, X)$ bilgi sistemi olarak tanımlansın.

Tanım:

V_x, x özneliklerinin aldığı değerlerin kümesi olmak üzere f_x bilgi fonksiyonu; $\forall x \in X$ için $f_x : E \rightarrow V_x$ ile tanımlanır.

Tanım:

$\forall Y \subseteq X$ alt kümesinde ayırtedilemezlik bağıntısı β olmak üzere $\forall e_i, e_j \in E$ ve $\forall x \in Y$ için $x(e_i) = x(e_j)$ şeklinde tanımlanan β bağıntısı bir denklik bağıntısıdır ve E/Y bölüm kümesinin elamanları $[e_i]_\beta = \{f : e\beta f, \forall e \in E\}$ şeklinde tanımlanan denklik sınıflarından oluşur.

Tanım:

$A \subset E$ olsun. A kümesinin $Y \subset X$ alt kümesinde sırasıyla alt ve üst yaklaşımları $YA_* = \{e_i \in E : [e_i]_\beta \subset A\}$ ve $YA^* = \{e_i \in E : [e_i]_\beta \cap A \neq \emptyset\}$ şeklinde tanımlanır. (YA_*, YA^*) ikilisine kaba küme (rough set) denir.

E 'de A nın sınırı ise $YA_s = YA^* - YA_*$ şeklinde tanımlanır.

3.2.2. Bulanık Kümeler(Fuzzy Sets)

E 'de bir bulanık küme, $E \rightarrow [0,1]$ tanımlanan eşleme öyle ki $E \times E$ deki R bulanık bağıntısı ile tanımlanan bir kümedir. $\forall f \in E$ için R_f bulanık kümesi

$\forall e \in E$ için $R_f(e) = R(e, f)$ ile tanımlanır.

Eğer R yansıyan ve simetrik bir bağıntı ise R bulanık tolerans bağıntısı olarak isimlendirilir.

U sonlu bir küme ise A 'nın kardinalitesi $|A| = \sum_{e \in U} A(e)$ ile hesaplanır.

R tolerans bağıntısı ile A 'nın alt ve üst yaklaşımları bir çok yolla oluşturulabilir. Bunlar genelde t-norm ve implikatörler olup sırasıyla aşağıdaki gibi tanımlanırlar;

$T : [0,1]^2 \rightarrow [0,1]$ eşlemesi $\forall e \in [0,1]$ için $T(1, e) = e$

ve

$I : [0,1]^2 \rightarrow [0,1]$ eşlemesi $\forall e \in [0,1]$ için $I(0,0) = 1$, $I(1, e) = e$

Bu bulanık lojik bağlaçlar bulanık-kaba küme teorisinin gelişiminde önemli bir rol almışlardır (Dubois ve Prade, 1990).

3.2.3. Bulanık-Kaba En Yakın Komşuluk Algoritması (Fuzzy-Rough Nearest Neighbour-FRNN)

Bulanık-kaba kümeler, ayırdilemezlik kavramına dayanmaktadır. Bulanık kümeler teoremi, ayrık değerler içeren veri kümelerine uygulanmaktadır. Oysaki gerçek hayatta veriler sürekli değerler içermektedir. Bu yüzden, sürekli değerlerin bir ön işlem uygulanarak ayrıklaştırılması gerekmektedir. Bu ön işlem bilgi kaybına sebebiyet verir. Oysaki sezgisel ve esnek bir yaklaşım olan E 'deki R bulanık bağıntısı sayesinde, sürekli öznelik değerlerine sahip nesnel arasındaki yaklaşık eşitlik modellenebilir (Radzikowska ve Kerre, 2002).

E üzerindeki R bağıntısı $E \rightarrow [0,1]$ eşlemesi bulanık tolerans bağıntısıdır. Yani E 'de verilen R bağıntısına göre $R(e, e) = 1$ ve $R(e_1, e_2) = R(e_2, e_1)$ dir. Bu R bağıntısıyla $A \subset E$ 'de alt ve üst yaklaşımlar için bulanık kaba yaklaşım uzayı

$$(RA)_*(e) = \inf_{f \in E} I(R(e, f), A(f)) \quad (1)$$

$$(RA)^*(e) = \sup_{f \in E} T(R(e, f), A(f)) \quad (2)$$

ile verilir.

Eğer bulanık-kaba yaklaşım uzayında, bulanıklık nicelikleri yerine (Q_u, Q_l) nicelikleri alındığında, A kümesinin alt ve üst yaklaşımları

$$(RA)_*(f) = Q_u \left(\frac{\sum_{e \in E} \min(R(e, f), A(e))}{\sum_{e \in E} R(e, f)} \right) \quad (3)$$

$$(RA)^*(f) = Q_l \left(\frac{\sum_{e \in E} \min(R(e, f), A(e))}{\sum_{e \in E} R(e, f)} \right) \quad (4)$$

ile belirlenir.

Bu yaklaşımlar kullanılarak Jensen ve Cornelis (2011) tarafından geliştirilen BK-EYK algoritmasının sözde kodu Şekil 1'de verilmiştir. Komşuluk değerleri belirlenirken (1) ve (2) denklemlerindeki alt ve üst küme yaklaşımları kullanılırsa algoritma FRNN, (3) ve (4) denklemleri kullanılırsa algoritma VQNN olarak adlandırılmaktadır.

FRNN (X, K, C, y)
 Girdi - K : En yakın komşuluk sayısı
 X : Eğitim veri kümesi
 C : Karar sınıf kümesi
 y : Test kümesi
 Çıktı - S : Test kümesinin sınıfları
 Başla
 $N \leftarrow$ En yakın komşuluk(y, K)
 $\zeta \leftarrow 0, S \leftarrow \emptyset$
 Tekrarla $C \in C$
 Eğer $\frac{(R_*C)(y) + (R^*C)(y)}{2} \geq \tau$ ise
 $S \leftarrow C$
 $\zeta \leftarrow (R_*C)(y) + (R^*C)(y)/2$
 Bitir
 Bitir
 Çıktı: S

Şekil 1: BK-EYK algoritmasının sözde kodu (Jensen ve Cornelis, 2011).

4. Modellerin Sonuçları

Çalışmada bulanık-kaba küme teorisine dayanan VQNN ve FRNN yöntemleri kullanılmıştır. Önerdiğimiz modellerin etkinliğini gösterebilmek için sonuçlarımız bulanık en yakın komşuluk (FNN), LR, C4.5 karar ağacı ve DVM yöntemlerinin sonuçlarıyla karşılaştırılmıştır.

Gözetimli öğrenmede, iki sınıflı bir veri kümesi üzerinde kurulan modellerin başarımlarını ölçmek için kullanılan ölçütler 2x2'lik hata dizeyi aracılığıyla hesaplanır (Tablo 1).

Tablo 1: 2x2'lik hata dizeyi

	Model Kestirimi		
	Pozitif	Negatif	Toplam
Gerçek Pozitif	DP	YN	GP=DP+YN
Gerçek Negatif	YP	DN	GN=YP+DN
Gerçek Toplam	MP=DP+YP	MN=YN+DN	E=GP+GN

Doğru pozitif (DP): Gerçekte pozitif sınıfa ait gözlem olup model kestirimi tarafından pozitif olarak sınıflandırılan gözlem sayısı.

Yanlış negatif (YN): Gerçekte pozitif sınıfa ait gözlem olup model kestirimi tarafından negatif olarak sınıflandırılan gözlem sayısı.

Yanlış pozitif (YP): Gerçekte negatif sınıfa ait gözlem olup model kestirimi tarafından pozitif olarak sınıflandırılan gözlem sayısı.

Doğru negatif (DN): Gerçekte negatif sınıfa ait gözlem olup model kestirimi tarafından negatif olarak sınıflandırılan gözlem sayısı.

Literatürde bir modelin başarısı geçirme kümesindeki örnekler modele girdi olarak sunulup doğruluk oranı ($DO = \frac{DP+DN}{E}$), anma ($A = \frac{DP}{GP}$) ve kesinlik ($K = \frac{DP}{MP}$) başarımları ölçütlere ile ölçülmektedir. Anma, model tarafından pozitif olarak sınıflandırılan örneklerin kaçının gerçekten pozitif sınıfa ait olduğu, kesinlik ise modelin pozitif

sınıf dediği örneklerin gerçekte kaçının pozitif olduğu bilgisini verir. İki sınıflandırma modeli karşılaştırılırken anma ve kesinlik ölçütünün her ikisinin de yüksek olduğu model tercih edilir. Bu amaçla anma ve kesinlik ölçütünün harmonik ortalaması olan F ölçütü kullanılır.

$$F_{\text{ölçütü}} = \frac{2 * A * K}{(A + K)} \quad (5)$$

Model başarımını ölçmek için kullanılan bir başka yöntem ise doğru pozitif oranı ile yanlış pozitif oranları kullanılarak çizilen *alıcı işletim özellikleri eğrisidir*. İki sınıflandırma modeli karşılaştırılırken bir modele ait alıcı işletim özellikleri eğrisi, her zaman diğer modele ait alıcı işletim özellikleri eğrisinin üzerinde kalıyorsa, birinci model tercih edilir (Alpaydın, 2011:419).

Bu eğrileri karşılaştırmak oldukça zahmetlidir. Bu eğrinin altında kalan alan hesaplanarak modelin başarımını tek bir sayıyla ifade edilir. Alıcı işletim özellikleri eğrisinin altındaki alan (AİÖ-A), eğrinin altındaki ardışık noktaların oluşturduğu yamukların alanlarının toplamı olarak kestirilir (Fawcett, 2006). AİÖ-A, [0, 1] aralığında bir değer alır ve 1'e ne kadar yakın bir değer alırsa model başarım oranı o kadar yüksektir.

4.1. Model Sonuçları

Güvenilir bir KS modeli oluşturmak ve veri bağımlılığının etkisini azaltmak için K-kat çapraz geçeleme kullanılmıştır. Bu örnekleme yönteminde veri kümesi rastgele olarak eşit büyüklükte K parçaya ayrılır. Bu parçalardan bir tanesi geçeleme için kullanılırken, geriye kalan K-1 parça eğitim kümesi için kullanılarak model oluşturulur. Bu işlem K parçanın herbiri bir kez geçeleme kümesi olacak şekilde K kez tekrarlanır. Model başarımını olarak, oluşturulan modellerin başarı ortalamaları alınmaktadır. Çalışmada K sayısı 10 olarak alınmıştır. Daha güvenilir hata kestirimleri yapabilmek amacıyla 10 x10 kat çaprazlama yapılarak ortalamaların ortalaması alınıp daha güvenilir hata kestirimi yapılmaya çalışılmıştır.

Deneyler, veri madenciliği yazılımı olan açık kaynak kodlu Hall ve diğerleri (2009) tarafından geliştirilen Weka üzerinde R. Jensen¹ tarafından geliştirilen yazılım araç takımı kullanılarak yapılmıştır. Tablo 2 ve Tablo 3'de bulanık-kaba küme tabanlı algoritmaların; FNN, LR, DVM ve C4.5 algoritmalarının DO ve F-ölçütü metrikleriyle sonuçları gösterilmektedir. Bulanık-kaba küme tabanlı algoritmalar her iki veri kümesi içinde, diğer yaygın kullanılan makina öğrenmesi yöntemleriyle benzer sonuçlar vermiştir.

Tablo 2: Algoritmaların DO sonuç değerleri

Veri kümesi	VQNN %	FRNN %	FNN %	LR %	DVM %	C4.5 %
Kredit-a	86,75	81,04	86,38	85,25	84,37	85,28
Kredit-g	73,76	70,29	74,56	75,24	75,09	71,25

¹Fuzzy yazılım araç takımı http://users.aber.ac.uk/rkj/site/?page_id=139 (e.t. 10.07.2018) adresinden indirilebilir.

Tablo 3: Algoritmaların F-ölçütü sonuç değerleri

Veri kümesi	VQNN	FRNN	FNN	LR	DVM	C4.5
Kredit-a	0,85	0,78	0,85	0,84	0,84	0,83
Kredit-g	0,83	0,79	0,83	0,83	0,83	0,80

Ling, Huang ve Zhang (2003) sınıflandırma yöntemlerinin sonuçlarını karşılaştırmada DO oranı yerine AİÖ-A değerlerinin kullanılmasının daha doğru olduğunu göstermiştir. Son yıllarda sınıflandırma yöntemlerinin performansları karşılaştırılırken DO ve $F_{ölçütü}$ yerine AİÖ-A değerlerinin karşılaştırılması tercih edilmektedir.

Bulanık-kaba-küme tabanlı FRNN ve VQNN'nin AİÖ-A sonuçları düzeltilmiş eşlenik t-testi (paired t-test (corrected)) kullanılarak diğer algoritmaların sonuçlarıyla karşılaştırılmış sonuçlar Tablo 4'de verilmiştir. Tablo 4'teki "*" ilgili yöntemin VQNN'den istatistiksel olarak daha kötü, "v" ise yöntemin VQNN'den daha iyi bir başarı gösterdiğini %5 anlamlılık düzeyinde ifade etmektedir.

Tablo 4: Algoritmaların AİÖ-A değerleri için eşlenik t-testi sonuçları

Veri kümesi	VQNN	FRNN	FNN	LR	DVM	C4.5
Kredit-a	0,92	0,88*	0,86 *	0,91	0,86 *	0,88 *
Kredit-g	0,74	0,71*	0,63*	0,78v	0,67*	0,65*

Tablo 4'te de görüldüğü üzere VQNN, LR hariç diğer sınıflandırma algoritmalarına göre istatistiksel olarak daha iyi sonuç vermiştir. VQNN, verideki belirsizliğin yüksek olduğu durumlarda FRNN'ye göre daha iyi sonuç verebilmektedir (Jensen ve Cornelis, 2011).

5. Sonuç

Bu çalışmada, kredi skorlama için bulanık-kaba küme tabanlı model önerilmiştir. Önerilen model, açık kaynaklı German credit (kredi-g) ve Australian credit approval (kredi-a) verilere uygulanmıştır. Seçilmiş veri kümeleri, hem nümerik hem de sembolik değerler içeren özneliklerden oluşmaktadır. Kredi-a veri kümesi, 15 öznelikten oluşmaktadır. Veri kümesi 307 (%44,5) kredi vermeye değer, 383 (%55,5) kredi verilmesi sakıncalı toplam 690 müşteriyi kapsar iken 20 öznelikten oluşan Kredi-g veri kümesi ise 700(%70) kredi vermeye değer, 300'ü (%30) kredi verilmesi sakıncalı 1000 müşteri içermektedir. İki kredi veri kümesi, yaygın olarak kullanılan makina öğrenmesi yöntemleriyle karşılaştırılmıştır.

Kredi skorlama için kullanılan modellerde son yıllarda önemli gelişmeler olmuşsa da en etkin yöntem konusunda tartışmalar hala devam etmektedir. Çalışma bu konudaki eksikliği gidermeyi amaçlamıştır.

Küme teorisine dayanan VQNN,

- belirsizlikle başedebilmek için daha etkin bir çerçeve sunmaktadır.
- kullanıcı tarafından önceden belirlenen parametrelere ihtiyacı yoktur.
- basit ve kolay anlaşılabilir. Sadece veri kümesindeki örneklerin bulanıklık benzerliklerini hesaplayarak kararlar verir.

- iv. verideki net hatayı tolere ederek kesin olmayan belirsiz verilerle kolaylıkla baş edebilir.

Önerilen yöntem belirsizlik ortamında kredi skorlama yöntemi olarak literatürdeki yöntemlerden daha etkin sonuç verdiği için bu yöntem uygulayıcılar ve özellikle politika yapıcılar için önemlidir. Politika yapıcılara, daha etkin politikaların belirlenmesinde yol gösterici rol üstlenebilir.

6. Kaynakça

- Alpaydın, E. (2011). Yapay öğrenme. İstanbul: Boğaziçi Üniversitesi Yayınevi.
- Altman E. (1968). Financial Ratios, Discriminant Analysis and The Prediction of Corporate Bankruptcy. *Journal of Finance*, 23(4), 589-609.
- Ayanoglu, Y ve Ertürk, B. (2007). Modern kredi riski yönetiminde derecelendirmenin yeri ve IMKB'ye kayıtlı şirketler üzerinde bir uygulama. *Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 9 (2), 75-90.
- Beaver W., (1966), Financial ratios as predictors of failure, *Journal of Accounting Research*, 4, 71-111.
- Bhatia, B. S. ve Batra, G. S. (1996). *Management of Financial Service*. New Delhi : Deep & Deep Publications.
- Demirbulut, Y., Aktaş, M., Kalıpsız, O. ve Bayracı S. (2017). *İstatistiksel ve makine öğrenimi yöntemleriyle kredi skorlama*. Paper presented at UYMS'17. Erişim adresi http://ceur-ws.org/Vol-1980/UYMS17_paper_83.pdf
- Desai, Vijay S., Jonathan N. Crook, and George A. Overstreet. "A comparison of neural networks and linear scoring models in the credit union environment." *European Journal of Operational Research* 95.1 (1996): 24-37.
- Dubois, D. ve Prade, H. (1990). Rough fuzzy sets and fuzzy rough sets, *Internat. J. General Systems*, 17 (2-3), 191-209.
- Dua, D. ve Karra Taniskidou, E. (2017). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.
- Fawcett, T. (2006), An introduction to ROC analysis, *Pattern Recognition Letters*, 27, 861– 874.
- Hall, M., Frank, E., Holmes G., Pfahringer, B., Reutemann, P., Witten I.H. (2009). The WEKA data mining software: an update. *SIGKDD Explorations*, (11), Issue1.
- Jensen, R. ve Cornelis, C. (2011). Fuzzy-Rough nearest neighbour classification and prediction. *Theoretical Computer Science*, 412(42), 5871-5884.
- Kavcıoğlu, Ş. (2014). Ticari bankacılıkta kredi riskinin ve kredi riski ölçüm modellerinin değerlendirilmesi. *Finansal Araştırmalar ve Çalışmalar Dergisi*, 3 (5), 11-19.
- Ling C.X., Huang J. ve Zhang H. (2003) AUC: A Better Measure than Accuracy in Comparing Learning Algorithms. In: Xiang Y., Chaib-draa B. (eds) *Advances in*

Artificial Intelligence. AI 2003. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol 2671. Springer, Berlin, Heidelberg.

Mester, L. (1997). What's the point of credit scoring?, Federal Reserve Bank of Philadelphia, Business Review, September, 3-16. Erişim adresi https://fraser.stlouisfed.org/files/docs/historical/frbphi/businessreview/frbphil_rev_199709.pdf

Önder, C. (2010). *Bankruptcy prediction with support vector machines*. (Yüksek lisans tezi, Humboldt-Universität zu Berlin). Erişim adresi <http://dx.doi.org/10.18452/14130>

Pavlak Z. (1982). Rough set. *Internat.J.Comput.Inform.Sci.*, 11(5), 341-356.

Tabagari, Salome. Credit scoring by logistic regression. Diss. Tartu Ülikool, 2015.

Radzikowska, A. M., Kerre, E. E. (2002). Comparative study of fuzzy rough sets. *Fuzzy Sets and Systems*, 125, 137-155.

Sousa, M. M. ve Reginaldo S. F. (2014). Credit analysis using data mining: application in the case of a credit union. *JISTEM-Journal of Information Systems and Technology Management*, 11(2), 379-396.

Walczak, B., Massart, D.L. (1999). Rough sets theory. *Chemometrics and Intelligent Laboratory Systems*, 47, 1-16.

Yanpeng, Q., Shen, Q., Mac Parthaláin, N., Shang, C. ve Wua W. (2013). Fuzzy similarity-based nearest-neighbour classification as alternatives to their fuzzy-rough parallels. *International Journal of Approximate Reasoning*, 54, 184-195