





Mamdani ve Sugeno-tip Bulanık Çıkarım Sistemleri ile Sosyal Medya Haber Popülerliğinin Tahmini

Prediction of Social Media News Popularity with Mamdani and Sugeno Type Fuzzy Inference Systems

İsmail Atacak*¹ , Ömer Ayberk Şencan¹ 

¹Gazi Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, 06560, Ankara, Türkiye

Başvuru/Received: 01/09/2022

Kabul / Accepted: 14/12/2022

Çevrimiçi Basım / Published Online: 31/12/2022

Son Versiyon/Final Version: 31/12/2022

Öz

Haber popülerliği, internet ya da sosyal ağ sitelerinde yayınlanmış haberlere gösterilen ilgi düzeyinin bir göstergesidir. Haber sağlayıcıları, bu göstergenin kendi lehlerine değişmesini sağlamak adına, rekabetçi ve kullanıcılar açısından okunurluğu yüksek haberler yapmak zorundadır. Bu durum hem haber servislerinin sürekliliğine hem de haber kalitesinin artırılmasına önemli katkılar sağlar. Bu yüzden, haber popülerliğini otomatik olarak tespit eden sistemlerin varlığı günümüzde kaçınılmaz haline gelmiştir. Bu çalışmada, Kaliforniya Üniversitesi (KU)-Irvine Makine Öğrenmesi Deposu veri tabanından indirilen veriler bileştirilerek oluşturulmuş dengeli veri seti ve bu veri setinden Sentetik Azınlık Örnekleme Tekniği (Synthetic Minority Oversampling Technique-(SMOTE)) ile üretilen dengeli veri setine Mamdani ve Sugeno-tip bulanık çıkarım sistemi temelli modeller uygulanarak haber popülerliği tahmini yapılmıştır. Haber popülerliği tahmini için, farklı çıkarım ve durulaştırma yöntemleri kullanılarak oluşturulan 8 farklı bulanık mantık temelli model oluşturulmuştur. Oluşturulan modellerden 6 tanesi Mamdani-tip bulanık çıkarım sistemi ile yapılandırılırken, 2 tanesinde Sugeno-tip bulanık çıkarım sistemi kullanılmıştır. Karışıklık matrisi metrikleri ve R^2 eğrileri ile performansları değerlendirilen tahmin modellerinin deneysel sonuçları, dengeli ve dengeli veri setlerinin her ikisinde de tüm metrikler açısından en iyi performansı mak-min çıkarım yöntemi ve ağırlık merkezi durulaştırma yöntemini kullanan Mamdani-tip bulanık çıkarım sisteminin sağladığını göstermiştir. Ayrıca, burada geliştirilen modellerin deneysel sonuçları literatürde yer alan benzer çalışmalarınki ile kıyaslandığında, ağırlıklı ortalama yöntemini kullanan Sugeno-tip bulanık çıkarım sistemi dışındaki tüm bulanık mantık temelli modellerin benzer çalışmalardaki modellerin en iyileri kadar rekabetçi bir performans sergilemiş olduklarını da ispatlamıştır.

Anahtar Kelimeler:

“Sosyal Ağlar, Mamdani-tip Bulanık Çıkarım Sistemi, Sugeno-tip Bulanık Çıkarım Sistemi, Makine Öğrenmesi, Haber Popülerliği”

Abstract

News popularity is an indicator that measures the relevance of news published on the internet or social networking sites. News providers have to deliver competitive and user-readable news to ensure this indicator changes in their favour. This contributes to both the continuity of news services and the improvement of news quality. Therefore, it has become a necessity nowadays to have systems that automatically detect news popularity. In this study, news popularity prediction is made by applying Mamdani and Sugeno type fuzzy inference-based models to the unbalanced data set created by combining data downloaded from the University of California (UC)-Irvine Machine Learning Repository database and the balanced data set produced from this dataset by Synthetic Minority Oversampling Technique (SMOTE). For news popularity prediction, 8 different fuzzy logic-based configurations were created using different inference and defuzzification methods. Of these configurations, 6 are Mamdani-type fuzzy inference system and 2 are Sugeno-type fuzzy inference system. Experimental results of prediction models, whose performances are evaluated with confusion matrix metrics and R^2 curves, showed that the best performance in terms of all metrics in both unbalanced and balanced datasets was provided by the Mamdani type fuzzy inference system using the max-min inference method and the centroid defuzzification method. In addition, when the experimental results of the models developed here were compared with those of similar studies in the literature, it was proven that all fuzzy logic-based models, except for the Sugeno type fuzzy inference system, which uses the wtaver method, performed as competitive as the best models in similar studies.

Key Words:

“Social Networks, Mamdani Type Fuzzy Inference System, Sugeno Type Fuzzy Inference System, Machine Learning, News Popularity”

Terminoloji

OSM	Çevrimiçi Sosyal Medya (Online Social Media)	SMOTE	Sentetik Azınlık Örnekleme Tekniği (Synthetic Minority Over-Sampling Technique)
KU	Kaliforniya Üniversitesi (University of California)	FS	Fischer Score
NLP	Doğal Dil İşleme (Natural Language Processing)	TP	Doğru Pozitif (True Positive)
ANN	Yapay Sinir Ağları (Artificial Neural Networks)	TN	Doğru Negatif (True Negative)
DL	Derin Öğrenme (Deep Learning)	FP	Yanlış Pozitif (False Positive)
NB	Naïve Bayes	FN	Yanlış Negatif (False Negative)
SVM	Destek Vektör Makinesi (Support Vector Machine)	Acc	Doğruluk (Accuracy)
DT	Karar Ağacı (Decision Tree)	Rec	Duyarlılık (Recall)
MTL	Çok-Görevli Öğrenme (Multi-Task Learning)	Prec	Kesinlik (Precision)
AdaBoost	Adaptif Yükseltme (Adaptive Boosting)	Fscr	F-score
RF	Rasgele Orman (Random Forest)		

1. Giriş

Çevrimiçi Sosyal Medya platformlarının (Online Social Media (OSM)) günlük yaşantımızdaki etkisi her geçen gün katlanarak artmaktadır. Bu konu üzerinde yapılan araştırmalara göre (Statista, 2022), aylık aktif sosyal medya kullanıcı sayısının 2023 yılında 3.43 milyara erişmesi beklenmektedir. Bu değer yaklaşık olarak dünya nüfusunun üçte birine denk gelmektedir. Teknolojideki gelişmeler sonucunda, herkes için erişilebilirliği artan bu platformlar, Web 2.0 sayesinde kullanıcılara diğer kullanıcılar ile etkileşime girmek ve içerik oluşturmak gibi imkanlar sunmanın yanı sıra (Dhawan et al., 2022), haberlere ve bilgilere de ulaşma fırsatı yaratmıştır (Luo & Hancock, 2020). Sosyal medya platformlarının kullanım sıklığının bu şekilde artması, neredeyse bütün bilinir kişilerin ve kurumların sosyal medyayı düzenli şekilde kullanması ve hatta kendileri ile ilgili yeni bilgileri geleneksel medya araçlarından önce sosyal medya platformlarından duyurmasına yol açmıştır. Doğal olarak bu haberleri takip etmek isteyen kullanıcıların da sosyal medyaya katılımı ile, kullanıcılar tarafından oluşturulan içerik sayısının da katlanarak arttığı görülmektedir (Zaman et al., 2014).

Sosyal medyanın kullanım oranının bu denli artması, aynı zamanda haberlere erişim için kullanılan kaynakların da çoğalmasına neden olmuştur. Popüler haber ajansları, günlük bazda onlarca haberi sosyal medya platformları aracılığıyla okurlarına aktarmaktadır. Bu durumun bir sonucu olarak, editörler kendi kurumlarına ait haberlerin, popülerlik açısından diğer kurumların haberlerinin önüne geçebilmesi için bir rekabet durumuna girmişlerdir. Dolayısı ile bu noktada, paylaşılan haberlerin popülerliğinin analizi ve performans takibi ihtiyacı ortaya çıkmıştır (Saeed et al., 2022). Bu ihtiyaç sonucunda haber siteleri okurlarına günün popüler haberlerini önermek için sistemler ve mekanizmalar oluşturmuştur (B. Wu & Shen, 2015). Ancak bu sistemler incelendiğinde, oluşturulan mekanizmaların haberlerin anlık görüntülenme sayısını baz aldığı görülmektedir (Xiong et al., 2021). Bu durum, anlık olarak popüler olan haberin zaman çerçevesi içerisinde en çok görüntülenen haberden daha popüler haberler olarak kullanıcılara sunulmasına neden olabilmektedir. Haberlerin zaman bazlı olarak da popülerliğinin incelenmesi, ilgili haber kaynaklarının kullanıcılarına hem zaman açısından hem de görüntülenme sayısı olarak popüler olan haberlerin sunulabilmesine olanak sağlaması açısından daha verimli sonuçlar elde edilmesini sağlayacaktır.

Hem sayısal hem de içerik anlamında bu kadar çok verinin ise sadece insan gücü kullanılarak incelenmesinin imkânsız olduğu ortadadır. Bu noktada araştırmacılar sıklıkla bilgisayarların hesaplama gücünden faydalanarak ilgili verileri incelemekte ve çıkarımları da bilgisayar üzerinde çalışan modeller kullanarak yapmaktadır (Whelan et al., 2020). Makine öğrenmesini kullanan modellerin dışarıdan sürekli programlama ihtiyacı duymadan öğrenebilmeleri, onları çalışma zamanı ve verimliliği açısından insan gücünün kullanıldığı sistemlerin önüne taşımıştır (T.K. et al., 2021). Sosyal medya verilerinin analizi noktasında farklı tipte makine öğrenmesi temelli yöntemler kullanılmaktadır. Yapay Sinir Ağları (Artificial Neural Network-(ANN)) (Chew et al., 2021), Naïve Bayes (NB) (Liu et al., 2020), Destek Vektör Makinesi (Support Vector Machine-(SVM)) (Xia et al., 2017) ve Doğrusal Regresyon (Lineer Regresyon-(LR)) (H. Wu et al., 2016) sosyal medya analizi konusunda sıklıkla kullanılan makine öğrenmesi yöntemlerinden bazılarıdır. Yukarıda verilen metotların yanı sıra, Bulanık Mantık (Fuzzy Logic-(FL)) temelli modeller de günümüzde sosyal medya analizi konusunda kullanılmakta ve popülerliği her geçen gün artmaktadır (Vashishtha & Susan, 2019).

Bu çalışmada, Mamdani ve Sugeno-tip bulanık çıkarım sistemleri kullanılarak Facebook sosyal medya platformundaki haberlerin popülerlik tahmini yapılmıştır. Haber verisi olarak, KU-Irvine Makine Öğrenmesi Deposu veri tabanında bulunan Multi-Source Social Feedback of Online News Feeds isimli veri setinden (Moniz & Torgo, 2018) yararlanılmıştır. Bu veri setinde bulunan "Economy, Microsoft, Obama & Palestine" konu başlıklarından oluşan haberleri içeren 4 ayrı veri seti KU-Irvine Makine Öğrenmesi Deposundan indirilerek birleştirilmiş ve tek bir veri seti haline getirilmiştir. Haberlerin gelecekteki popülerliği, zamana bağlı olarak geçmişteki haberlerin görüntülenme sayıları göz önünde bulundurularak tahmin edilmiştir. Tahmin Süreci, farklı çıkarım ve durulaştırma metotlarına sahip bulanık çıkarım sistemlerini içeren 8 farklı model üzerinden test edilmiştir. Çalışmanın geriye kalanı şu şekilde organize edilmiştir: Literatür Taraması Bölüm 2'de verilmiştir. Bölüm 3'te kullanılan Materyal ve Metotlar ile seçilmiş veri setinin

özniteliklerinden ve bu veri setine uygulanan ön işleme aşamalarından bahsedilir. Oluşturulan modeller kullanılarak elde edilen deneysel sonuçlar Bölüm 5’te sunulur. Son bölümde ise çalışmanın genel sonuçları tartışılarak değerlendirilir.

2. Literatür Taraması

Daha önce yapılan çalışmalar sosyal medya platformlarının, popüler birer kullanıcılar arası etkileşim aracı olmalarının yanı sıra, haber kaynağı olarak da kullanılabilirliklerini göstermiştir (Kwak et al., 2010). Bu platformlarda popüler olan konuların çoğu toplumun o zaman diliminde odaklandığı başlıklar olduğundan, geçmişten günümüze haber popülerliğinin değerlendirilmesi üzerine çok sayıda bilimsel araştırma ortaya konmuştur (Beştaş, 2020). Bu bilimsel çalışmalar incelendiğinde, araştırmacıların haberlerin popülerliği konusunda farklı yaklaşımlar izledikleri ve farklı metotlardan yararlanarak sonuçları elde ettikleri görülmektedir.

Naseri & Zamani (2019) Çok-Görevli Öğrenme (Multi-task Learning (MTL)) yöntemini (Caruana et al., 1997) kullanarak Telegram platformundaki farklı haber ajanslarının haberlerinin popülerliğini analiz etmişler ve bu haber ajanslarının her biri için %5 ve %25’lik dilimdeki en popüler haberleri tespit etmişlerdir. Bu çalışmalarında maksimum %93,99 doğruluk ve %81,32 F-score değeri elde edilmiştir.

Deshpande (2018) KU-Irvine Makine Öğrenmesi Deposu veri tabanından elde edilen veri setini kullanarak sosyal medyada paylaşılan haberlerin gelecekteki popülerliğini tespit etmiştir. Bu bağlamda öncelikle Lineer Diskriminant Analiz yöntemi kullanılarak veri setinde öznitelik azaltma işlemi gerçekleştirilmiş, sonrasında ise Adaptif Yükseltme (Adaptive Boosting-(AdaBoost)) yöntemi ile geliştirilen model kullanılarak popüler haberlerin tespiti yapılmıştır. Araştırmacı bu çalışmada %69 doğruluk ve %73 F-score değerlerine ulaşmıştır.

Fernandes ve arkadaşları (2015) bir makale yayınlanmadan önce popüler olup olamayacağını tespitini yapmak amacıyla Akıllı Karar Destek Sistemi (Intelligent Decision Support System-(IDSS)) adı verilen bir modeli geliştirmiştir. Bu çalışmada bir online haber sitesi olan Mashable’dan elde edilen veriler kullanılmış ve beş çok bilindik makine öğrenmesi modelinin elde ettiği performanslar test edilmiştir. Test edilen algoritmalar içerisinde en başarılı sonuç %67 doğruluk oranı ve %69 F-score ile Rastgele Orman (Random Forest-(RF)) algoritması ile elde edilmiştir.

Ren & Yang (2015) ise gerçekleştirdikleri çalışma ile Fernandes ve arkadaşları (2015) tarafından gerçekleştirilen çalışmayı daha da geliştirerek 10 çok bilindik makine öğrenmesi algoritmasını aynı veri seti ile test etmişler ve bu algoritmalar içerisinde en başarılı sonucu veren modelin %69 doğruluk ve %71 duyarlılık ile RF algoritması olduğunu ortaya koymuşlardır.

Arapakis ve arkadaşları (2014) bir sisteme yeni bir nesnenin dahil edilme durumu anlamına gelen “Cold-start” haber popülerliğinin tespitini gerçekleştirmiştir. Sisteme yeni bir öğenin dahil olması ile kullanıcıların buna nasıl tepki verecekleri bilinmediği için tahmin süreci zorlaşır. Bu problemin üstesinden gelmek için Yahoo sosyal medya platformundan elde edilen veri setini kullanan araştırmacılar SVM yönteminden %79,7’lik doğruluk değerine ulaşmışlardır.

AL-Mutairi & Khan (2015) Wikipedia sitesindeki Arapça makalelerin popülerliğini tespit etmek amacıyla veri madenciliği yöntemlerinden faydalanarak bir yöntem geliştirmişlerdir. Bu yöntemde, Wikipedia makalesinin diğer kullanıcılar (ziyaretçiler) tarafından görüntülenme sayısına etki eden, daha önce gerçekleşmiş olan özel etkinlikler, sosyal medya platformlarındaki mesaj ve paylaşımlar gibi dış etkenler ile bu dış etkenlerin görüntülenme sayısına olan etkisi araştırılmıştır. Çalışmalarında Karar Ağacı (Decision Forest-(DF)), NB ve W-Jrip algoritmalarından yararlanan araştırmacılar en iyi performansı %62,62’lik doğruluk oranı ile DF algoritmasından elde etmişlerdir.

Ahmed ve arkadaşları (2013) Seçimli Perseptron (Voted Perceptron-(VP)) algoritmasının genişletilmiş bir versiyonu olan Benzerlik Öğretimi (Similarity Learning-(SiLA)) algoritmasını kullanarak, Twitter sosyal medya platformundaki paylaşımların popülerliğini araştırmışlardır. Bu çalışmada ilgili tweetlerin re-tweet sayılarından faydalanılmıştır. İkili sınıflandırma ve çoklu sınıflandırmanın kullanıldığı çalışmada %85’lik doğruluk değeri ile en yüksek başarımın ikili sınıflandırma ile elde edildiği görülmüştür. Ayrıca, SiLA ve K-En Yakın Komşu (K-Nearest Neighbors-(KNN)) metotları karşılaştırılmış ve SiLA metodunun KNN metoduna göre daha başarılı sonuçlar elde ettiği belirtilmiştir.

Szabo & Huberman (2008) YouTube ve Digg sosyal medya platformlarından elde edilen verileri kullanmış ve kullanıcıların erişimine sunulan içeriklerin elde ettikleri ilk görüntüleme değerlerini kullanarak, bu içeriklerin ne kadar popüler olma potansiyeli olduğunu hesaplamışlardır.

Phuvipadawat & Murata (2010) Hotstream adını verdikleri modeli geliştirerek, bu modeli Twitter’dan haberleri toplamak, gruplamak, sıralamak ve takip etmek için kullanmışlardır.

Sosyal medyada haber tespiti alanında kullanılan bir başka metot Bulanık Mantıktır. Bulanık mantığın sosyal medya üzerine uygulanması alanında farklı çalışmalar göze çarpar. Morente-Molinera ve arkadaşları (2019) kullanıcıların fikirlerinden bilgi elde

etmek için duygu analizi prosedürlerini kullanabilen yeni bir yöntem sunmuşlardır. Alınan bilgiler, bulanık ontoloji üzerinde organize bir şekilde saklanmaktadır. Bu sayede, diğer kullanıcıların oluşturulan bulanık ontolojilerden yararlanarak burada depolanan bilgileri alması ve bu bilgilerden yararlanması sağlanır.

Augusto ve arkadaşları (2019) havayolu şirketleri hakkındaki kötü haberlerin, sosyal medya kullanıcılarının şirket hakkındaki görüşlerini ne kadar etkilediğini bulanık mantığı kullanarak tespit etmeye çalışmıştır. Sosyal-kimlik üzerine inşa edilen bu çalışma ile müşteri-marka ilişkisi incelenerek kullanıcıların negatif haberlere olan tepkisinin ölçülmesi amaçlanmıştır.

Aghasian ve arkadaşları (2020) yapılandırılmamış bilgilerin gizliliğini sağlamak için otomatik bir model sunmuşlardır. Bu çalışmada, çevrimiçi sosyal ağlarda yapılandırılmamış veriler için gizlilik ölçümünün zorluklarını vurgulamanın yanı sıra, bu verilerdeki gizlilik riskini değerlendirmiştir. Parçalanmış verilerle ilgili gizlilik riskini ortaya çıkarmak için bir makine öğrenmesi modeli ve bulanık tabanlı sistemi bileştiren bir model ortaya koymuşlardır.

Howells & Ertugan (2017) sosyal medya verilerinin pazarlama anlamında incelenmesi amacı ile sosyal ağ verilerinin bulanık mantık yöntemleri kullanılarak duygu analizinin yapılmasını sağlayan bir model önermişlerdir. Geliştirilen model, bir mikroblogun içeriğini (Twitter platformundaki tweet gibi) ve müşteri geribildirimlerini veya algısını analiz edebilecek düzeyde bir yapı sunmaktadır. Elde edilen bu yapı bir kuruluşun bir ürün veya hizmet üzerinde müşterilerinin algılarını temsil etmek için geliştirilen bir bilgisayar uygulamasının temelini oluşturur.

Tavana ve arkadaşları (2013) sosyal medyadaki kullanıcıların, toplulukların ya da paylaşımların incelenmesi yerine pazarlama kârı düşünülerek sosyal medya platformu seçimi problemi üzerinde bir çalışma gerçekleştirmişlerdir.

Francisco & Castro (2020) özellikle Twitter gibi mikroblog siteleri ile benzer etkileşim mekanizmaları sosyal ağlarda, kullanıcı profillerini tamamlamak, değiştirmek ve zenginleştirmek için bulanık bir model sunmuşlardır. Araştırmacılar, kullanıcı profillerine karmaşık öznitelikler (belirli konulara yönelik ilgi ve fikirler) eklemek için sosyal ağ sitelerinde gerçekleştirilen eylemlerin gerçek anlamına dayanan bulanık bir çerçeve tasarlamışlardır.

Karyotis ve arkadaşları (2015) insan duygularını akıllı bilgisayar sistemlerine dahil etmek için yeni bir duygu modelleme metodolojisi sunulmuşlardır. Önerilen yaklaşım, kullanıcılardan duygu bilgilerini ortaya çıkarmak için bir yöntem, genetik olarak optimize edilmiş uyarlanabilir bulanık mantık tekniği kullanılarak modellenmiş yeni bir duygu temsili (AV-AT modeli) ve kullanıcının zaman içinde duygusal yörüngesini tahmin etmek ve izlemek için bir çerçeve içermektedir. Bu çalışmada, aynı zamanda bulanık mantık tekniği, mevcut diğer makine öğrenme yaklaşımlarına göre duygusal durumları modelleme yeteneği açısından karşılaştırılmıştır. Duygu kavramı ve yapısal unsurları doğal olarak bulanık yapıdadır ve belirsizlik içerirler. D. Wu (2012)'nin belirttiği gibi, duygular kişisel ve kişiler arası belirsizliğe tabidir. Kişiler arası belirsizlik, bireylerin aynı duygu hakkında sahip olduğu farklı algı ve ifadeleri ilgilendirirken, kişisel belirsizlik farklı zamanlarda veya bağlamlarda kendi duyguları hakkında sahip olduğu belirsizliktir. Bulanık mantık sistemleri bu belirsizlikleri, ilişkileri temsil etmek ve modellemek için bir araç olarak kullanır (Kazemzadeh et al., 2013). Bu çalışmada, duygu temsili yani AV-AT modeli geliştirilmiştir. Bu model, bir takım çevrimiçi ve çevrimdışı testlere tabi tutularak performans değerlendirmesi yapılmıştır. Yapılan performans değerlendirmesi sonucunda, geliştirilmiş olan bu modelin Popüler Uyarılma Değeri (AV) Temsili (Russell, 2003) veya Duygusal Yörüğe Hipotezi (Kirkland & Cunningham, 2012) ile karşılaştırıldığında daha etkili sonuçlar elde ettiği görülmüştür.

İncelenen bir diğer çalışmada Appel ve arkadaşları (2016) cümle düzeyinde duygu analizi sorununa hibrit bir yaklaşım sunarak çözüm aramışlardır. Bu yöntem, Doğal Dil İşleme (Natural Language Processing-(NLP)) temel tekniklerinden faydalanılarak oluşturulmuştur. SentiWordNet kullanılarak geliştirilmiş olan bir duygu sözlüğü, duygusal yönelim kutupluluğunu ve yoğunluğunu tahmin ederek, duygularla işlem yapabilmek adına bir temel oluşturmaktadır. Önerilen hibrit yöntem, üç farklı veri setine uygulanmış ve elde edilen sonuçlar, NB ve Maximum Entropy yöntemleri ile karşılaştırılmıştır. Miller (1956)'a göre, 7 ∓ 2 kategori bir alan için efektif kategori sayısı olarak belirlenmiştir. Bu çalışmada da bu durum göz önünde bulundurularak 5 (7-2) simetrik etki alanında dağıtılmış etiket bulunmuştur (0,1). Bu çalışmanın Twitter verileri ile yapılan performans değerlendirmesi sonucunda elde edilen değerler verilmiştir.

Her ne kadar bulanık mantığın sosyal medyaya uygulanmasına dair birçok çalışma incelenmiş olsa da karşılaştırmanın daha doğru ve efektif bir şekilde yapılabilmesi için sosyal medyada farklı metotlar ile haber popülerliğinin tespiti alanında yapılan çalışmalar da karşılaştırmaya dahil edilmiştir. Bu çalışmalarda kullanılan metotlar ve elde edilen performans değerleri Tablo 1'de verilmiştir. Bu çalışmada kullanılan yöntemin klasik makine öğrenmesi temelli yöntemlerden farklı olarak bulanık mantığı, sosyal medyadaki haberlerin popülerliğini ölçmek için kullanması en belirgin akademik katkısı olarak verilebilir.

Tablo 1. Sosyal Medyada haber popülerliği tespiti yapan çalışmalar ve performans metrikleri

Çalışma	Yıl	Platform	Metot	Performans
(Naseri & Zamani, 2019)	2019	Telegram	MTL	81,3% F-score
(Deshpande, 2018)	2018	UCI MLR	LDA & AdaBoost	73% F-score
(Fernandes et al., 2015)	2015	Mashable	RF	69% F-score
(Ren & Yang, 2015)	2015	Mashable	RF	69% Doğruluk
(Arapakis et al., 2014)	2014	Yahoo	SVM	70,7% Doğruluk
(AL-Mutairi & Khan, 2015)	2015	Wikipedia	DT	62,6% Doğruluk
(Ahmed et al., 2013)	2013	Twitter	SiLA & k-NN	85% Doğruluk

3. Materyal ve Metotlar

Bu bölümde, yapılmış olan çalışmada kullanılan yöntem ve metotlardan detaylı bir şekilde bahsedilmektedir. İlk olarak haber başlığı formundaki ham veri setine ön işlem basamakları uygulanarak çalışmada kullanılan dengeli ve dengesiz veri setinin nasıl oluşturulduğu açıklanmıştır. Sonrasında bulanık mantık temelli modellerin girişlerinin belirlenmesini sağlayan öznitelik seçim süreci anlatılmıştır. Daha sonra ise, tahmin sürecinde kullanılan bulanık mantık temelli modellerin tasarım süreci sunulmuştur.

3.1. Veri setinin oluşturulması

Bu çalışmada kullanılan veri setinin ham hali, KU-Irvine Makine Öğrenmesi Deposu veri tabanından elde edilen farklı haber başlıklarına ait veri setlerinin bir araya getirilmesi ile oluşturulmuş birleşik bir veri setidir. Konuyla ilgili Makine Öğrenmesi Deposunda bulunan veri setleri, Kasım 2015 ile Haziran 2016 tarihleri arasında 8 aylık süreçte Facebook, Google+ ve LinkedIn platformlarından toplanmış yorumları içeren bir takım haber başlığı verilerden oluşur. Ancak, bu çalışmada, sadece Facebook platformunda bulunan “Microsoft, Economy, Obama & Palestine” konu başlıklarını hakkındaki yorumları içeren veri setlerini birleştirilerek ham veri seti oluşturulmuştur. Birleştirme işlemi öncesi konu başlıklarına ait veri sayılarını içeren bilgiler Tablo 2’de verilmiştir.

Tablo 2. Veri birleştirme işlemi öncesi konu başlıklarına ait veri sayıları

Konu Başlığı	Veri Sayısı
Microsoft	29928
Economy	18531
Obama	27015
Palestine	7687
Toplam	83161

Birleştirme işlemi sonucunda oluşturulan veri seti toplamda 83161 veri içermektedir. Ayrıca, veri setinde haber etiketi ve IDLink indeksleri de dâhil olmak üzere 146 öznitelik bulunmaktadır. Bu özniteliklerden her biri, ilgili haberin 20 dakikalık periyotlar halinde aldığı görüntülenme sayısını belirtmektedir. Örneğin, TS1 isimli öznitelik paylaşılan haberin ilk 20 dakika içerisindeki görüntülenme sayısını tutarken, TS144 ise paylaşımından iki gün sonraki son yirmi dakikanın verisini tutar. Eğer ilgili haber hakkında Facebook üzerinde paylaşım yapılmamış ise veya ilgili zaman diliminde bu haberin popülerlik değeri hesaplanamıyorsa, bu özniteliklere otomatik olarak “-1” değeri atanmıştır. Birleştirme işlemi sonucunda oluşturulan veri setinin ham haline ait bir kesit Tablo 3’te verilmiştir.

Tablo 3. Oluşturulan ham veri setine ait bir kesit

Haber Konusu	IDLink	TS1	TS2	TS3	...	TS144
Microsoft	98809	105	200	279	...	13543
Economy	31272	-1	184	389	...	15250
Obama	90905	99	117	131	...	1164
Palestine	6922	-1	304	352	...	802

3.2. Veri ön işleme aşamaları

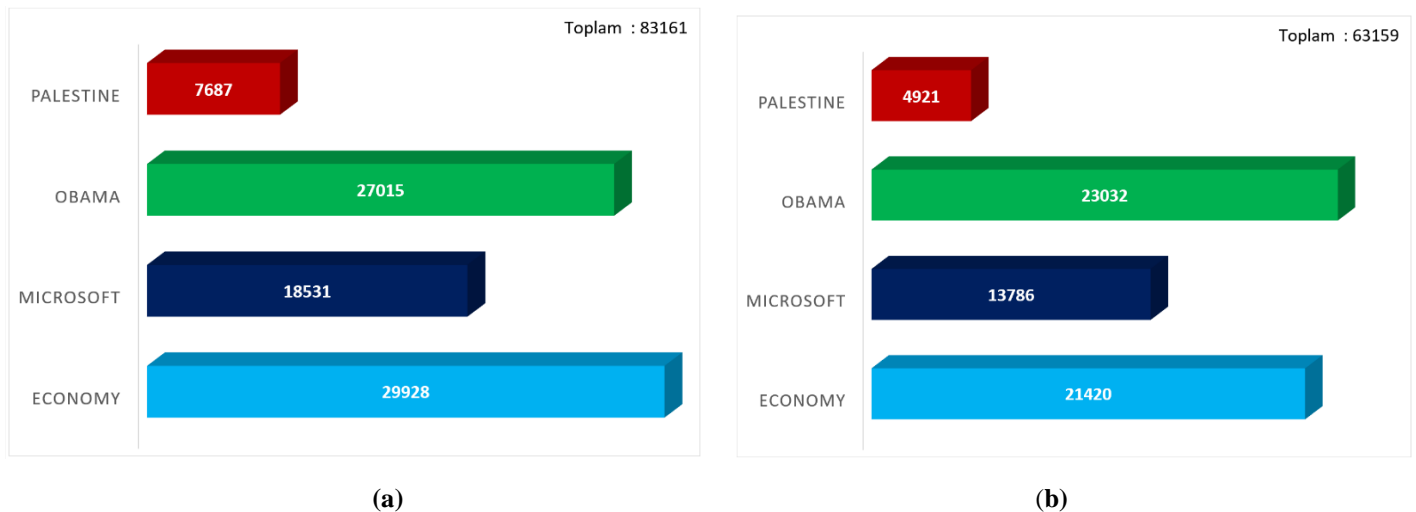
Deneysel çalışmalarda dengesiz ve dengeli olmak üzere iki tip veri seti kullanılmıştır. İlk ön işleme aşamasında ham veri setinden yararlanarak dengesiz veri seti oluşturulurken, ikinci ön işleme aşamasında dengesiz veri setine istatistiksel bir veri çoklama metodu olan SMOTE uygulanarak dengeli veri seti elde edilmiştir. Ham veri setindeki 20 dakikalık periyotlar halindeki görüntüleme sayısı çoğu zaman dilsel değerlendirme için yeterli bir örneklem sağlamaz. Bu yüzden, bu çalışmada dört saatlik periyotlar halindeki verilerden yararlanarak TS144 yani haberlerin son popülerlik düzeyinin hesaplanması amaçlanmıştır. Bu hedef doğrultusunda ön işleme aşamasının ilk basamağında veri seti dört saatlik periyotlar halinde yeniden düzenlenmiştir. Buna göre, TS144 etiketi ile belirlenmiş olan popülerlik değerleri ilgili haberin gerçek popülerlik değeri olarak kabul edilmiş ve TS144 sınıfı kullanılarak haberler ÖNEMLİ ve ÖNEMSİZ olmak üzere iki sınıfa ayrılmıştır. TS144 periyodu içerisindeki en yüksek değerlerin %10'u eşik değeri kabul edilerek,

- TS144 Max değerinin %10'undan daha küçük değere sahip olan haberler ÖNEMSİZ haber,
- TS144 Max değerinin %10'undan daha büyük değere sahip olan haberler ÖNEMLİ haber,

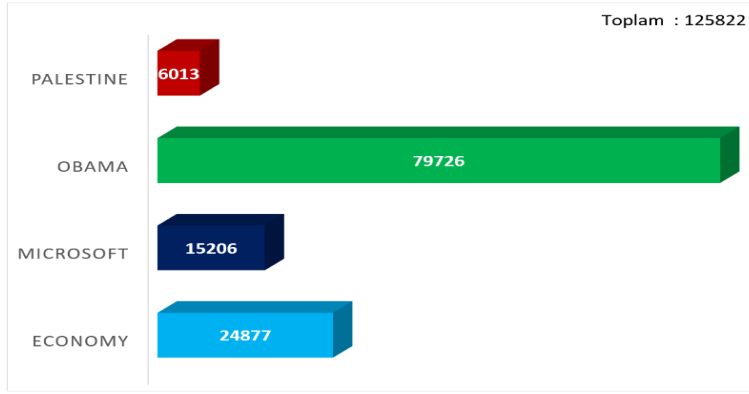
olarak sınıflandırılmıştır. ÖNEMSİZ haber sınıfı sayısal olarak "0" ile etiketlenirken, ÖNEMLİ haber sınıfı ise sayısal olarak "1" ile etiketlenmiştir. Bunun yanı sıra, veri setinde bulunan ancak yorumlanması mümkün olmayan ya da kendini tekrar eden veriler de mevcuttur. Bu kapsamda, TS1-TS144 arası tüm özniteliklerinin sayısal değeri "-1" olan veriler ve bu aralıkta sayısal değerleri aynı olan veriler veri setinden silinmiştir. Sonuçta 83161 verinin 20.002' ü silinerek 63159 satırlı ve Konu ve IDlink başlıkları ile birlikte 14 sütunlu (öznitelikli) bir veri seti elde edilmiştir. İlgili veri seti, içerisinde 247 ÖNEMLİ haber ve 62912 ÖNEMSİZ haber bulunduğundan dengesiz bir veri seti yapısı ihtiva eder.

Dengesiz veri setlerinin kullanıldığı tahmin modellerinde birçok performans metriği modelin performansını ölçmede yetersiz kalır. Bu yüzden yaptığımız çalışmada kullanılan bulanık mantık temelli modellerin performansının doğru bir şekilde ölçülebilmek için bu veri setine ek, yapay veri oluşturma yöntemlerinden faydalanılarak yeni bir veri setinin oluşturulma gereksinimi duyulmuştur. Yapay verilerin oluşturulması ve veri setinin dengeli hale getirilmesi için ise ikinci ön işlem aşaması olarak SMOTE (Chawla et al., 2002) uygulanmıştır. SMOTE veri dengesizliği problemlerinde en yaygın kullanılan çözüm yöntemidir. Bu yöntemde, azınlıkta olan veri sınıfından bir örneklem alt kümesi oluşturulur (Garcia, 2022). Sonrasında bu veriler göz önünde bulundurularak, yeni sentetik veriler üretilir. SMOTE doğrusal enterpolasyonu kullandığı için (Fernandez et al., 2018) yeni üretilen verilerin örneklem olarak alınan verilere yakın olduğu söylenebilir. İlgili yöntem sayesinde, rastgele aşırı örnekleme nedeniyle oluşan aşırı öğrenme ve rastgele yetersiz örnekleme nedeniyle oluşan bilgi kaybı gibi istenmeyen durumların (Zhang et al., 2022) çözülebilmesi için yapay veriler oluşturularak veri setinin dengelenmesi sağlanır. SMOTE'nin kullanılmasındaki neden onun rastgele örnekleme yöntemlerine kıyasla sahip olduğu avantajlarıdır. Öyle ki, SMOTE ile hem aşırı öğrenme problemine çözüm bulunabilir, hem de azınlık veri sınıfının öngörülebilirliği artışı için (Fernandez et al., 2018) modellerin daha verimli çalışması sağlanabilir.

SMOTE ile yapılan işlemin sonucunda, veri seti 62910 ÖNEMLİ ve 62912 ÖNEMSİZ veri olmak üzere, toplamda 125822 veri içeren dengeli bir veri seti elde edilmiştir. Ham veri setine, birinci ön işlem aşamasının uygulanması sonucunda elde edilen dengesiz veri seti ve ikinci ön işlem aşaması olan SMOTE'un uygulanması sonucunda elde edilen dengeli veri setlerinin konu başlıklarına göre haber dağılımları Şekil 1'de gösterilmiştir.



Şekil 1. Ham, dengesiz ve dengeli veri setlerinin konu başlıklarına göre haber dağılımları: (a) Ham veri seti; (b) İlk ön işlem aşaması sonrası elde edilen dengesiz veri seti; (c) İkinci ön işlem aşaması (SMOTE işlemi) sonrası elde edilen dengeli veri seti

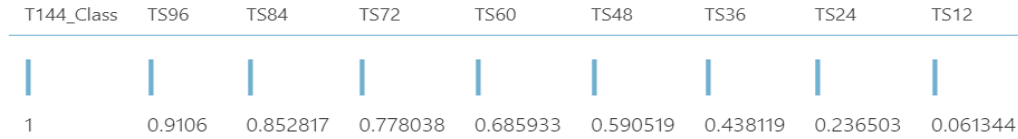


(c)

Şekil 1 (devam). Ham, dengesiz ve dengeli veri setlerinin konu başlıklarına göre haber dağılımları: (a) Ham veri seti; (b) İlk ön işlem aşaması sonrası elde edilen dengesiz veri seti; (c) İkinci ön işlem aşaması (SMOTE işlemi) sonrası elde edilen dengeli veri seti

3.3. Öznitelik seçimi

Bulanık mantık temelli sistemlerde giriş sayısı arttıkça giriş/çıkış arasındaki ilişkiyi temsil eden kuralların oluşması çok daha karmaşık hale gelmektedir. Bu durum göz önünde bulundurularak bu çalışmada giriş sayısı üç ile sınırlandırılmıştır. Bu noktada, mevcut veri setindeki özniteliklere öznitelik seçim süreci uygulanarak sonuca en çok etki eden üç öznitelik belirlenmiş ve bulanık mantık temelli modellerin giriş verisi olarak bu öznitelikler kullanılmıştır. Yapılan çalışmada öznitelikleri seçmek için, istatistiksel fonksiyonlar üzerinden elde edilen kazanç sıralamasına göre seçim sürecini yürüten filtre temelli bir öznitelik seçim metodu olan Fisher skor metodu (Fischer et al., 1999) kullanılmıştır. Bu metod özniteliklere ait sütunların ortalama ve standart sapma değerlerini kullanarak kazanç skorunu belirler ve seçim sürecini en yüksek değerli kazanç skoruna sahip öznitelikten başlayarak sıralı bir şekilde gerçekleştirir. Şekil 2’ de dengesiz veri setindeki TS12-TS96 aralığındaki 8 öznitelige sahip veriye Fisher skor metodu uygulanarak elde edilmiş kazanç değerlerine göre öznitelik derecelendirmesi gösterilmiştir.

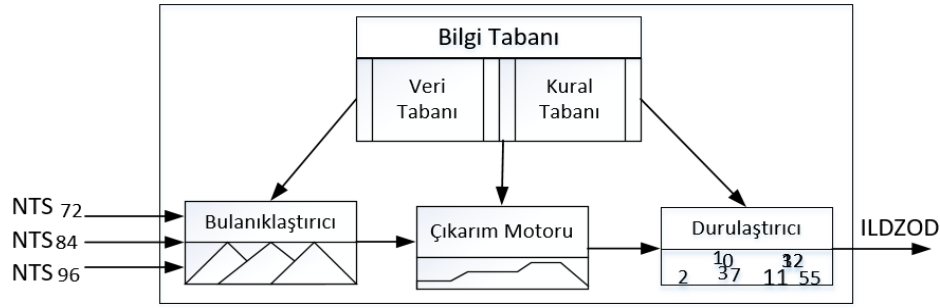


Şekil 2. Fisher skorlarına göre özniteliklerin derecelenmesi

Kazanç değerlerine göre ilk üç öznitelik sırası ile TS96, TS84 ve TS72 olup, bu öznitelikler oluşturulan tahmin modellerine giriş olarak uygulanmıştır.

3.4. Bulanık mantık temelli modellerin oluşturulması

Bu çalışmada kullanılan bulanık mantık temelli modeller, dengesiz ve dengeli veri setinden öznitelik seçim süreci ile belirlenmiş TS96, TS84 ve TS72 periyotlarındaki haberlerin görüntülenme sayılarına bağlı olarak, bu haberlerin gelecekteki (TS144 periyodundaki) ilgi düzeyinin önem derecelerini (ILDZOD) hesaplamak için kullanılır. Üç giriş ve tek çıkış bulanık mantık sistemi yapısında kurgulanmış olan bu modeller, Mamdani ve Sugeno-tip bulanık çıkarım sistemleri ile oluşturulmuştur. Veri setleri; Mamdani-tip bulanık çıkarım sisteminde çıkarım metodolojisi olarak maksimum-minimum (mak-min) ve maksimum-çarpım (mak-çarp) yöntemlerini, durulaştırma metodolojisi olarak ağırlık merkezi (centroid of area), alan açıortayı (bisector of area) ve en büyüklerin ortası (mean of maximum) yöntemlerini birleştiren 6 farklı model üzerinden test edilirken, Sugeno-tip bulanık çıkarım sisteminde metodolojik olarak ağırlıklı ortalama (wtaver) ve ağırlıklı toplam (wtsum) yöntemlerini birleştiren 2 farklı model ile test edilmiştir. Sugeno bulanık çıkarım sisteminde çıkış üyelik fonksiyonlarının sabit ya da doğrusal olması, bu sistemlerin çıkarım süreci sonunda sayısal sonucu doğrudan elde etmesine olanak sağlar. Bu durum, Mamdani çıkarım sisteminde olduğu gibi ek bir durulaştırma gereksinimini ortadan kaldırır. Bunun dışında her iki bulanık çıkarım sisteminde de işlem süreçleri birbiri ile aynıdır. Bu yüzden burada bulanık mantık temelli modellerin oluşturulması, Mamdani-tip bulanık çıkarım modeli üzerinden anlatılmıştır. Haber popülerliği tahmini için oluşturulan Mamdani-tip bulanık çıkarım sisteminin temel bileşenlerini gösteren blok diyagram Şekil 3’te verilmiştir.



Şekil 3. Haber popülerliği tahmini için oluşturulan Mamdani-tip bulanık çıkarım sistemi

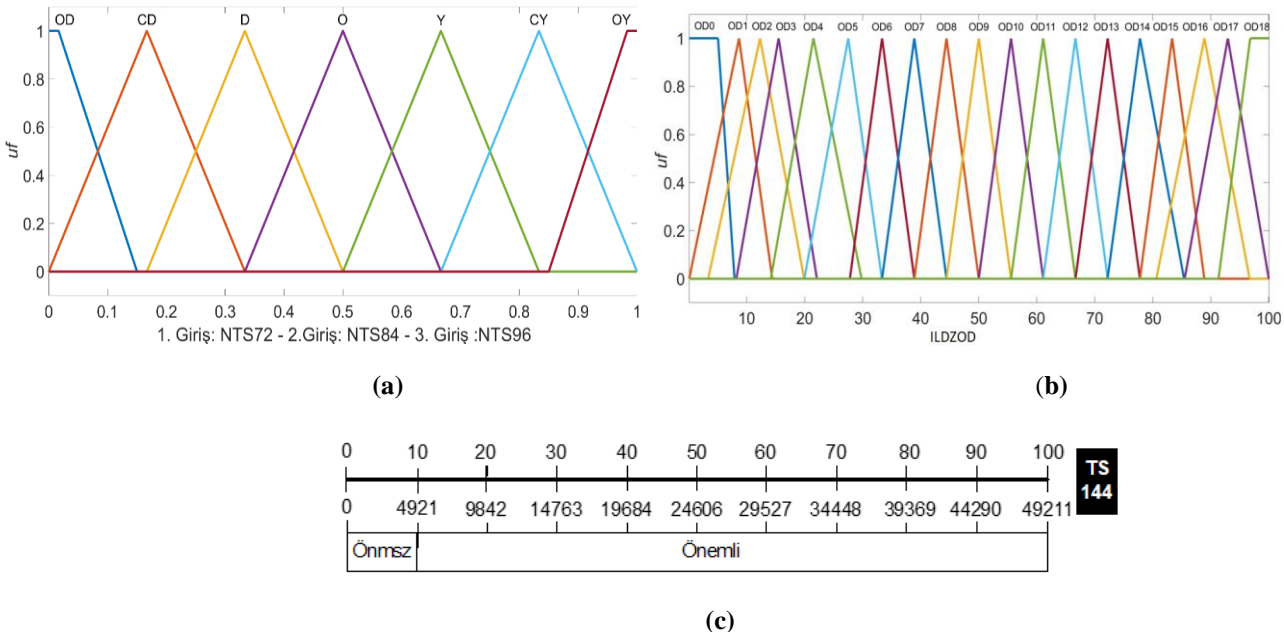
Önerilen bulanık mantık temelli modellerde, giriş olarak kullanılan TS72, TS84 ve TS96 periyotlarındaki haberlerin görüntülenme sayıları ait oldukları sütunlarındaki en yüksek değere bölünerek 0-1 aralığında normalize edilmiştir. Buna göre, bulanık mantık temelli modellerin normalize edilmiş formdaki girdileri Denklem 1’de verilen formül kullanılarak elde edilebilir.

$$NTS_X(i) = \left\{ \begin{array}{l} 1 \quad \text{Eğer } 1 \leq \frac{TS_X(i)}{TS_XMAK} \\ \frac{TS_X(i)}{TS_XMAK} \quad \text{Değilse} \end{array} \right\} \quad (1)$$

Burada $NTS_X(i)$ X’inci periyottaki i’inci haber verisinin 0-1 aralığında normalize edilmiş değerini, $TS_X(i)$ X’inci periyottaki i’inci haber verisinin görüntülenme sayısını ve TS_XMAK X’inci periyottaki haber verisinin en yüksek değerini gösterir. Her bir giriş 0-1 aralığında tanımlanan bir bulanık evrende üçgen tip üyelik fonksiyonları üzerinden isimleri,

- Oldukça Düşük: OD
- Düşük: D
- Yüksek: Y
- Oldukça Yüksek: OY
- Çok Düşük: CD
- Orta: O
- Çok Yüksek: CY

olan 7 bulanık küme ile tanımlanmıştır. Bulanık mantık temelli modellerin çıktıkları olarak, haberlere olan ILDZOD’i alınmıştır. ILDZOD günlük yaşantımızda yüzde cinsinden oransal değerler ile değerlendirildiği için bulanık mantık temelli modellerin çıkış evreni 0-100 aralığında ölçeklenmiş ve üçgen tipi üyelik fonksiyonları kullanılarak isimleri OD0, OD1...OD18 olan 19 bulanık küme ile tanımlanmıştır. Bu bilgiler doğrultusunda oluşturulan bulanık mantık temelli modellere ait giriş ve çıkış üyelik fonksiyonları ile sınıflandırma ölçeği sırası ile Şekil 4(a), Şekil 4(b) ve Şekil 4(c)’de gösterilmiştir. Şekil 4(c)’deki sınıflandırma ölçeğinden anlaşılacağı üzere, 10%’luk ilgi düzeyinin altında kalan haberler ÖNEMSİZ, bu oranın üzerinde ilgi düzeyine sahip olan haberler ise ÖNEMLİ haber sınıfı olarak etiketlenmiştir.



Şekil 4. Giriş-çıkış üyelik fonksiyonları ve sınıflandırma ölçeği: (a) Giriş üyelik fonksiyonları; (b) Çıkış üyelik fonksiyonları; (c) Sınıflandırma Ölçeği

Önerilen bulanık mantık temelli modeller için kural tabanı, her bir giriş değişkeni 7 farklı düzeyde popülerlik seviyesine (7 bulanık kümeye) sahip olduğu için toplamda 343 adet tanımlanmış kuraldan oluşur. Kuralların oluşturulması sırasında, her bir girişin, çıkışa eşit düzeyde etki ettiği göz önünde bulundurulmuştur. Bulanık mantık temelli modellerde haber popülerliği tespiti için oluşturulan kurallara ait bir kesit, Şekil 5'te verilmiştir.

1. Giriş: NTS72

	OD	CD	D	O	Y	CY	OY		
2.Giriş: NTS84	OD	OD0	OD1	OD2	OD3	OD4	OD5	OD6	OD
	CD	OD1	OD2	OD3	OD4	OD5	OD6	OD7	
	
	OY	OD6	OD7	OD8	OD9	OD10	OD11	OD12	CD
	OD	OD1	OD2	OD3	OD4	OD5	OD6	OD7	
	CD	OD2	OD3	OD4	OD5	OD6	OD7	OD8	
	
	OY	OD7	OD8	OD9	OD10	OD11	OD12	OD13	OY
	
	OD	OD6	OD7	OD8	OD9	OD10	OD11	OD12	
CD	OD7	OD8	OD9	OD10	OD11	OD12	OD13		
...		
OY	OD12	OD13	OD14	OD15	OD16	OD17	OD18		

3.Giriş: NTS96

Şekil 5. Bulanık mantık temelli modeller için oluşturulan kural tablolarının yapısı

3.5. Performans Metrikleri

Birçok sınıflandırma probleminde modellerin performanslarını test etmek için farklı ölçüm metrikleri kullanılır. Bu metriklerin bir kısmı model ile kullanılan veri seti arasındaki ilişkinin derecesini veren katsayılar şeklinde karşımıza çıkarken, diğerleri bir modelin tahmin ettiği değerler ile gerçek değerlerini karşılaştıran ve karışıklık (hata) matrisi olarak isimlendirilen bir matrise dayalı hesaplanmış değerler olarak kendini gösterir. Model ve veri seti arasındaki bağıntının gücünü ölçen katsayılar, bir değişkenin diğer bir değişkeni tanımlarken ne kadar öngörülebilir olduğunu gösterir. Bu tip bir performans ölçümünde katsayılar, değişkenlerin gerçek ve tahmini sınıflar üzerindeki örnek dağılımları üzerinden doğruluk değerlendirmesi yapılarak elde edilir. Bu çalışmada kullanılan R^2 performans metriği (Determinasyon Katsayısı) de yukarıda bahsedilen tanımlamaya dahildir. Buna göre, en başarılı durumda, modelin çıktısı olarak elde edilen değerler ile veri setindeki etiketlenmiş olan değerler tamamen birbirine eşit olur ve bu durum $R^2 = 1$ sonucunu ortaya koyar.

Ancak, (Goodman & Kruskal, 2012) önerilmiş olan modellerin değerlendirilebilmesi için sadece determinasyon katsayısının kullanılmasının yeterli olmadığını ortaya koymuştur. Çünkü, determinasyon katsayısı gibi ilişkisel ölçütlerin verdiği sonuçlar yalnızca, doğru şekilde sınıflandırılmış olan verilerin mevcut olması halinde henüz sınıflandırılmamış olan verilerin de doğru bir şekilde sınıflandırılabilceği anlamına gelir. Bununla birlikte, bazı durumlarda önerilen modeller performans açısından yüksek determinasyon katsayısı elde etseler de aynı performansını karışıklık matrisinin bazı metrikler açısından sağlamayabilirler. Bu yüzden bu çalışmada önerilen modellerin performansını tam olarak tespit edebilmek için R^2 metriğine ek olarak karışıklık matrisi temelli metrikler de kullanılmıştır.

Karışıklık matrisi temelli metriklerin hesaplanmasında, isimleri Doğru Pozitif (True Positive-(TP)), Doğru Negatif (True Negative-(TN)), Yanlış Pozitif (False Positive-(FP)) ve Yanlış Negatif (False Negative-(FN)) olan bu matrise ait dört temel parametreden yararlanır. Burada TP model tarafından doğru tahmin edilen pozitif örnek sayısını, TN model tarafından doğru tahmin edilen negatif örnek sayısını, FP model tarafından yanlış tahmin edilen negatif örnek sayısını ve FN model tarafından yanlış tahmin edilen pozitif örnek sayısını tanımlar. Bu çalışmada önerdiğimiz modellerin performansını ölçmek için karışıklık matrisi temelli metrikler olarak yaygın kullanılan doğruluk (accuracy), duyarlılık (recall) ve F-score (F-score) metrikleri kullanılmıştır.

Doğruluk (Acc) modelin genel performansını yansıtan bir metrik olup, elde ettiği değer modelin ürettiği doğru örnek sayısının toplam örnek sayısına oranlanması ile elde edilebilir. Bu metriğin hesaplanmasında kullanılan formül Eşitlik 2' de verilmiştir.

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Duyarlılık(Rec), modelin pozitif örnekleri sınıflandırmadaki başarısını ölçen bir metrik olup, onun değeri modelin doğru olarak sınıflandırdığı pozitif örnek sayısının tüm pozitif örnekler oranlanması ile bulunabilir. Eşitlik 3'te bu metriğe ait formül gösterilmiştir.

$$Rec = \frac{TP}{TP+FN} \quad (3)$$

Kesinlik (Prec), modelin pozitif olarak tahmin ettiği örneklerin gerçekten pozitif olma olasılığın ölçen bir metriktir. Bu metriğin değeri Eşitlik 4'te verilen formül kullanılarak hesaplanabilir.

$$Prec = \frac{TP}{TP+FP} \quad (4)$$

F-Score (Fscr) diğer adıyla F-ölçümü (F-Measure), (Caruana & Niculescu-Mizil, 2004) Kesinlik ve Duyarlılık metriklerinin harmonik ortalamasıdır. Eşitlik 5'te bu metriğin değerini hesaplayan genel formül verilmiştir.

$$Fscr = 2 \times \frac{Prec \times Rec}{Prec + Rec} \quad (5)$$

F-score metriği, (Labatut & Cherifi, 2011) tarafından etiketlenmiş gerçek veriler ile tahmin edilen sınıfların arasındaki benzerliğin bir ölçüsü olarak tanımlanır ve 0 (hiç örtüşme yok) ile 1 (tamamen örtüşme var) aralığında değerler alır. Elde edilen değer 1'e ne kadar yakınsa modelin de o kadar başarılı olduğu söylenebilir.

4. Deneysel Sonuçlar

Deneysel çalışmalar, "Materyal ve Metotlar" bölümünde detayları sunulan Mamdani ve Sugeno-tip bulanık çıkarım sistemlerinin çıkarım işlemleri ve durulaştırma işlemlerinde kullanılan yöntemlerin farklı şekilde yapılandırılmasından oluşan 8 farklı bulanık mantık temelli modelin dengesiz ve dengeli veri setlerine uygulanması ile gerçekleştirilmiştir. Mamdani-tip bulanık çıkarım sisteminin çıkarım işlemleri mak-min ve mak-çarp yöntemleri ve durulaştırma işlemleri ağırlık merkezi, alan açortayı ve en büyüklerin ortası yöntemleri kullanılarak gerçekleştirilirken, Sugeno-tip bulanık çıkarım sisteminde metodolojik olarak bu işlemler için ağırlıklı ortalama ve ağırlıklı toplam yöntemlerinden yararlanılmıştır. Bu yüzden yöntemlerin farklı şekilde yapılandırılmasına bağlı olarak Mamdani-tip bulanık çıkarım sistemi kullanılarak 6, Sugeno-tip bulanık çıkarım sistemi kullanılarak 2 model olmak üzere toplamda 8 model geliştirilmiş ve veri setlerine uygulanmıştır. Modellerin performansları R^2 metriği ile karışıklık matrisi temelli doğruluk, duyarlılık ve F-Score metrikleri üzerinden test edilmiştir. Dengesiz ve dengeli veri setlerinde her bir model için R^2 eğrileri ve bu metriğin değeri, 0-100 aralığında bulanık mantık modelinin ürettiği tahmin edilen ilgi düzeyinin önem derecesi (TAHMİN-ILDZOD) sonuçları ile TS144 periyodu için 0-100 aralığına ölçeklenmiş gerçek ilgi düzeyinin önem derecesi (GERÇEK-ILDZOD) sonuçları kullanılarak elde edilmiştir. Benzer şekilde karışıklık matrisi temelli metriklerin elde edilmesinde her bir model için bulanık çıkarım sisteminin ürettiği değerler ve TS144 periyodunun değerleri %10'dan küçükse ÖNEMSİZ (0) ve %10'dan büyükse ÖNEMLİ (1) sınıfına dahil olacak şekilde yeniden yapılandırılarak gerçek ve tahmin edilen sınıflar oluşturulmuştur. Bu sınıflara göre karışıklık matrisi parametre (TP, TN, FP ve FN) değerleri elde edilerek doğruluk, duyarlılık ve F-Score metrikleri hesaplanmıştır. Dengesiz ve dengeli veri setleri için deneysel çalışmalarda elde edilen performans sonuçları ve değerlendirmeleri aşağıda alt başlıklar halinde verilmiştir.

4.1. Dengesiz veri seti için deneysel sonuçlar

Bulanık mantık temelli modellerin dengesiz veri setine uygulanması ile elde edilen karışıklık matrisi parametrelerini ve bu matris parametrelerine göre hesaplanmış doğruluk, duyarlılık ve F-score metriklerine ait değerleri gösteren tablolar sırası ile Tablo 4 ve Tablo 5'te verilmiştir. Performans metrikleri dışında sadece karışıklık matrisi parametreleri üzerinden modelin performansı, onun doğru tahmin ettiği örnek sayısı (TP+TN) ve yanlış tahmin ettiği örnek sayısı (FN+FP) kullanılarak da doğrudan değerlendirilebilir. Bu kapsamda Tablo 4 incelendiğinde en iyi performansı, 63112 doğru tahmin edilen örnek sayısı ve 47 yanlış tahmin edilen örnek sayısı ile çıkarım sürecinde mak-min yöntemini ve durulaştırma sürecinde ağırlık merkezi yöntemini kullanan Mamdani-tip bulanık çıkarım sistemi göstermiştir. Bununla birlikte mak-çarp çıkarım yöntemi ve ağırlık merkezi durulaştırma yöntemini kullanan Mamdani-tip bulanık çıkarım sisteminin de 63111 doğru tahmin edilen örnek sayısı ve 48 yanlış tahmin edilen örnek sayısı ile bu performansa oldukça yakın bir sonuç ürettiği tablodaki sonuçlardan net bir şekilde anlaşılmaktadır. En kötü performans ise 62205 doğru tahmin edilen örnek sayısı ve 954 yanlış tahmin edilen örnek sayısı ile ağırlıklı toplam yöntemini kullanan Sugeno-tip bulanık çıkarım sistemi tarafından sergilenmiştir.

Tablo 5'te verilen sonuçlardan anlaşılacağı üzere doğruluk, duyarlılık ve F-Score metrikleri açısından en iyi performansı sırası ile 0.9993, 0.9028 ve 0.9047 değerleri ile mak-min çıkarım yöntemi ve ağırlık merkezi durulaştırma yöntemini kullanan Mamdani-tip bulanık çıkarım sistemi elde etmiştir. Sonuçlardan dengesiz veri setinde modellerin doğruluk değerleri birbirine oldukça yakın olduğu için, karışıklık matrisi temelli metriklerden duyarlılık ve F-Score metriklerinin doğruluk metriğinden daha ayırt edici olduğu sonucuna da ulaşılabilir.

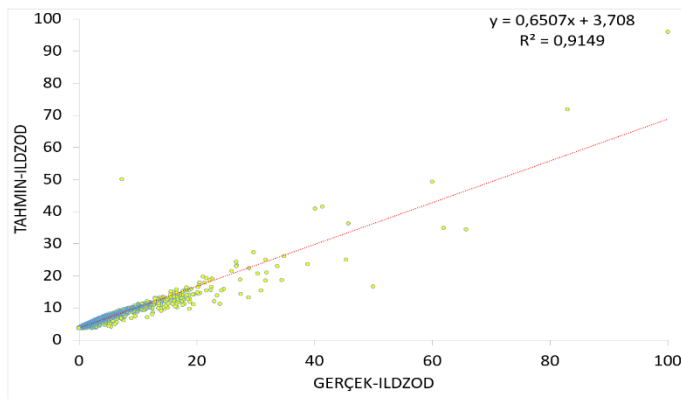
Tablo 4. Dengesiz veri seti için bulanık mantık temelli modellerin karışıklık matrisi parametreleri

Çıkarım Metodu	Çıkarım Süreci	Durulaştırma Süreci	TP	FN	FP	TN	TP+TN	FP+FN
Mamdani	Mak-Min	Ağırlık merkezi	223	24	23	62889	63112	47
		Alan Açıortayı	202	45	5	62907	63109	50
		En Büyüklerin Ortası	211	36	17	62895	63106	53
	Mak-Çarp	Ağırlık merkezi	203	44	4	62908	63111	48
		Alan Açıortayı	197	50	4	62908	63105	54
		En Büyüklerin Ortası	213	34	23	62889	63102	57
Sugeno	Sugeno Çıkarımı	Ağırlıklı ortalama	124	123	1	62911	63035	124
		Ağırlıklı toplam	218	29	925	61987	62205	954

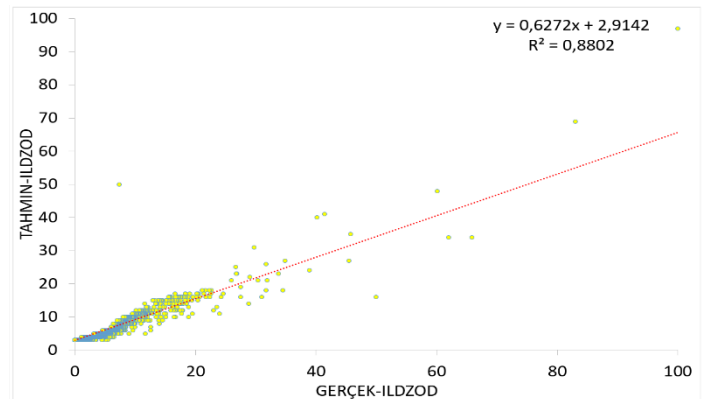
Tablo 5. Dengesiz veri seti için bulanık mantık temelli modellerin karışıklık matrisi performans metrikleri

Çıkarım Metodu	Çıkarım Süreci	Durulaştırma Süreci	Doğruluk	Duyarlılık	F-Score
Mamdani	Mak-Min	Ağırlık merkezi	0.9993	0.9028	0.9047
		Alan Açıortayı	0.9992	0.8178	0.8899
		En Büyüklerin Ortası	0.9992	0.8543	0.8884
	Mak-Çarp	Ağırlık merkezi	0.9992	0.8219	0.8943
		Alan Açıortayı	0.9991	0.7976	0.8795
		En Büyüklerin Ortası	0.9991	0.8623	0.8820
Sugeno	Sugeno Çıkarımı	Ağırlıklı ortalama	0.9980	0.5020	0.6667
		Ağırlıklı toplam	0.9849	0.8826	0.3137

Dengesiz veri seti için bulanık mantık temelli modellerin R^2 performans değerlerini içeren grafikler Mamdani-tip bulanık çıkarım sistemleri için Şekil 6'da ve Sugeno-tip bulanık çıkarım sistemleri için Şekil 7'de gösterilmiştir. R^2 eğrileri genellikle doğrusal regresyon modellerinde kullanılan bir uyum iyiliği ölçüsü (Colin Cameron & Windmeijer, 1997) olarak tanımlanır. Dolayısı ile bu değer yüksek olması, kullanılan modelin, eldeki veriye uygunluğunun bir göstergesidir. Şekil 6'da ve Şekil 7'de verilen sonuçlar incelendiğinde, R^2 metriği açısından en iyi performansın 0.9149 değeri ile mak-min çıkarım yöntemi ve ağırlık merkezi durulaştırma yöntemini kullanan Mamdani-tip bulanık çıkarım sisteminin sergilediği görülebilir. Bu metrik açısından en kötü performansı ise 0.6260 değeri ile mak-çarp çıkarım yöntemi ve en büyüklerin ortası durulaştırma yöntemlerini kullanan Mamdani-tip bulanık çıkarım sistemi göstermiştir.

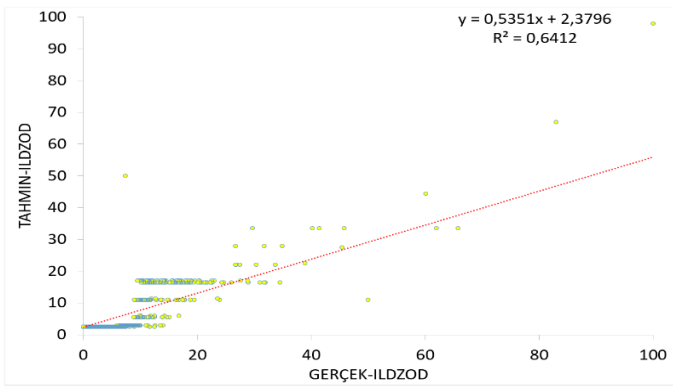


(a)

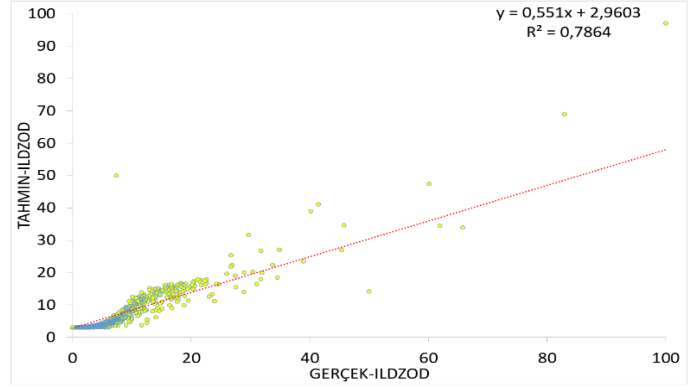


(b)

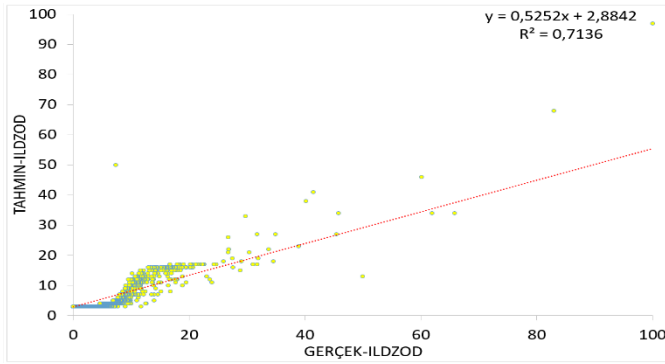
Şekil 6. Dengesiz veri seti için, Mamdani-tip bulanık çıkarım sistemlerine ait R^2 eğrileri (a) Mamdani – Mak-Min, Ağırlık merkezi; (b) Mamdani – Mak-Min, Alan Açıortayı; (c) Mamdani – Mak-Min, En Büyüklerin Ortası; (d) Mamdani – Mak-Çarp, Ağırlık merkezi; (e) Mamdani – Mak-Çarp, Alan Açıortayı; (f) Mamdani – Mak-Çarp, En Büyüklerin Ortası



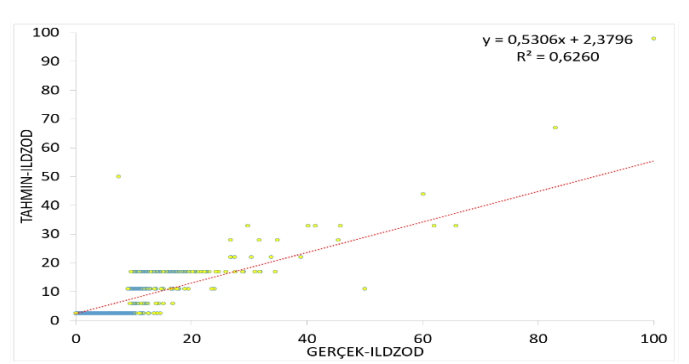
(c)



(d)

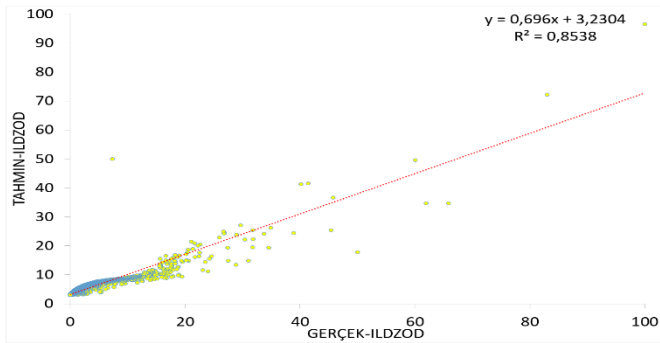


(e)

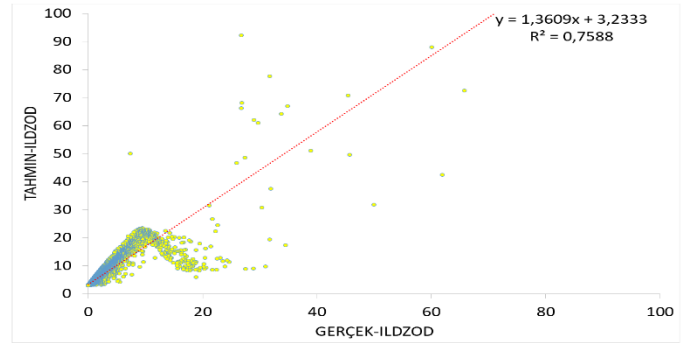


(f)

Şekil 6 (devam). Dengesiz veri seti için, Mamdani-tip bulanık çıkarım sistemlerine ait R^2 eğrileri (a) Mamdani – Mak-Min, Ağırlık merkezi; (b) Mamdani – Mak-Min, Alan Açığı; (c) Mamdani – Mak-Min, En Büyüklerin Ortası; (d) Mamdani – Mak-Çarp, Ağırlık merkezi; (e) Mamdani – Mak-Çarp, Alan Açığı; (f) Mamdani – Mak-Çarp, En Büyüklerin Ortası



(a)



(b)

Şekil 7. Dengesiz veri seti için, Sugeno-tip bulanık çıkarım sistemlerine ait R^2 eğrileri (a) Sugeno-Ağırlıklı ortalama; (b) Sugeno-Ağırlıklı toplam

Bulanık mantık temelli modellerin dengesiz veri setine uygulanması sonucu elde edilen sonuçlar genel olarak değerlendirilirse, tüm metrikler açısından en iyi performansın mak-min çıkarım yöntemi ve ağırlık merkezi durulaştırma yöntemini kullanan Mamdani-tip bulanık çıkarım sisteminin sergilediği sonucuna varılabilir.

4.2. Dengeli veri seti için deneysel sonuçlar

Bulanık mantık temelli modellerin SMOTE ile elde edilen dengeli veri setine uygulanması sonucu ulaşılan karışıklık matrisi parametrelerinin değerleri ve bu matrisin parametrelerine bağlı olarak hesaplanmış doğruluk, duyarlılık ve F-Score metriklerinin değerleri sırası ile Tablo 6 ve Tablo 7'de sunulmuştur. Tablo 6'daki sonuçlardan mak-min çıkarım yöntemi ve ağırlık merkezi durulaştırma yöntemini kullanan Mamdani-tip bulanık çıkarım sisteminin doğru tahmin ettiği 121689 örnek sayısı ve yanlış tahmin

ettiği 4133 örnek sayısı ile en iyi performansı sergilediği net bir şekilde görülmektedir. Bu sonuca en yakın performansı veren model 120240 doğru tahmin edilen örnek sayısı ve 5582 yanlış tahmin edilen örnek sayısı ile mak-çarp çıkarım yöntemi ve en büyüklerin ortası durulaştırma yöntemini kullanan Mamdani-tip bulanık çıkarım sistemidir. Karışıklık parametreleri açısından en kötü performansı ise 94410 doğru tahmin edilen örnek sayısı ve 31412 yanlış tahmin edilen örnek sayısı ile ağırlıklı ortalama yöntemini kullanan Sugeno-tip bulanık çıkarım sistemi vermiştir.

Tablo 7’de verilmiş olan dengeli veri seti için elde edilen performans sonuçları, bütün performans metrikleri açısından en iyi sonucun mak-min çıkarım yöntemi ve ağırlık merkezi durulaştırma yöntemini kullanan Mamdani-tip bulanık çıkarım sistemi tarafından elde edildiğini göstermektedir. Bu model, en iyi performans sonucuna 0.9672’lik doğruluk, 0.9347’lik duyarlılık ve 0.9660’lık F-Score değerleri ile ulaşmıştır. Dengeli veri seti için karışıklık matrisinin parametre temelli performans sonuçlarında olduğu gibi; doğruluk, duyarlılık ve F-Score metrikleri açısından bu sonuca en yakın değerleri sırası ile 0.9556, 0.9116 ve 0.9536 değerleri ile mak-çarp çıkarım yöntemi ve en büyüklerin ortası durulaştırma yöntemini kullanan Mamdani-tip bulanık çıkarım sistemi sağlamıştır. Benzer şekilde, karışıklık parametreleri temelli performans sonuçlarında olduğu gibi dengeli veri seti için doğruluk, duyarlılık ve F-Score metrikleri açısından en kötü performans sırası ile 0.7503, 0.5007 ve 0.6673 değerleri ile ağırlıklı ortalama yöntemini kullanan Sugeno-tip bulanık çıkarım sistemi tarafından elde edilmiştir.

Dengeli ve dengesiz veri setlerinde modellerin performansı karışıklık matrisi temelli metrikler açısından Tablo 5 ve Tablo 7’ de verilen değerler üzerinden karşılaştırılırsa; dengesiz veri setinde modellerin performansı açısından hemen hemen hiç belirleyiciliği olmayan doğruluk metriğinin, dengeli veri setinde duyarlılık ve F-Score metriği ile performans açısından oldukça belirleyici olduğu görülmektedir. Diğer taraftan dengeli veri setinde ağırlıklı ortalama yöntemine sahip Sugeno-tip bulanık çıkarım modeli dışındaki tüm modellerin duyarlılık ve F-Score metriklerinin dengesiz veri setindeki değerlerine göre sayısal olarak daha yüksek bir performans değerine sahip oldukları gözlemlenmiştir.

Tablo 6. Dengeli veri seti için bulanık mantık temelli modellerin karışıklık matrisi parametreleri

Çıkarım Metodu	Çıkarım Süreci	Durulaştırma Süreci	TP	FN	FP	TN	TP+TN	FP+FN
Mamdani	_____	Ağırlık merkezi	58801	4109	24	62888	121689	4133
		Mak-Min	53102	9808	5	62907	116009	9813
		En Büyüklerin Ortası	56242	6668	17	62895	119137	6685
		Ağırlık merkezi	53267	9643	5	62907	116174	9648
		Mak-Çarp	51967	10943	5	62907	114874	10948
		En Büyüklerin Ortası	57351	5559	23	62889	120240	5582
Sugeno	Sugeno Çıkarımı	Ağırlıklı ortalama	31499	31411	1	62911	94410	31412
		Ağırlıklı toplam	55153	7757	925	61987	117140	8682

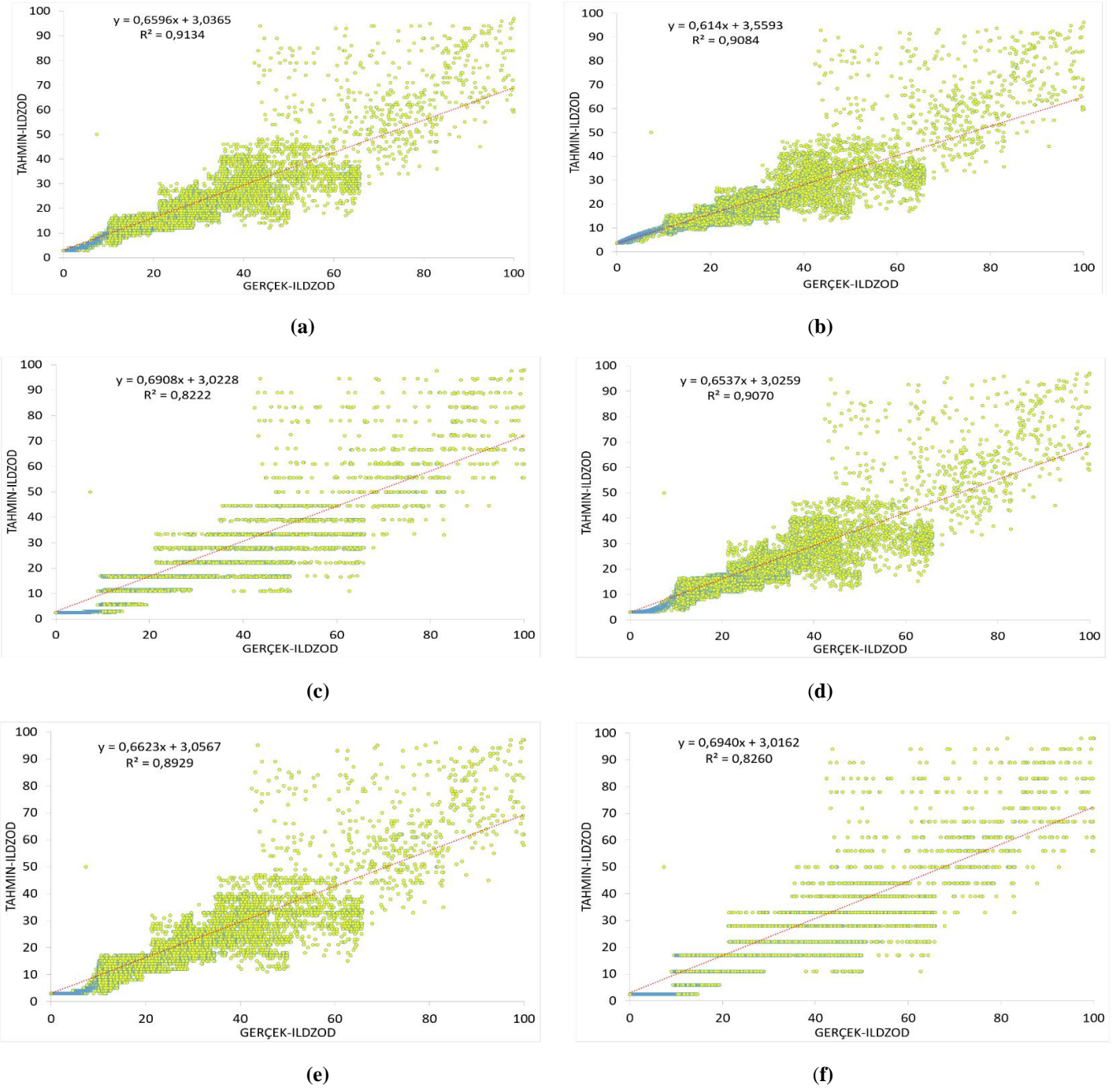
Tablo 7. Dengeli veri seti için bulanık mantık temelli modellerin karışıklık matrisi performans metrikleri

Çıkarım Metodu	Çıkarım Süreci	Durulaştırma Süreci	Doğruluk	Duyarlılık	F-score
Mamdani	_____	Ağırlık merkezi	0.9672	0.9347	0.9660
		Mak-Min	0.9220	0.8441	0.9154
		En Büyüklerin Ortası	0.9469	0.8940	0.9439
		Ağırlık merkezi	0.9233	0.8467	0.9170
		Mak-Çarp	0.9130	0.8261	0.9047
		En Büyüklerin Ortası	0.9556	0.9116	0.9536
Sugeno	Sugeno Çıkarımı	Ağırlıklı ortalama	0.7503	0.5007	0.6673
		Ağırlıklı toplam	0.9310	0.8767	0.9270

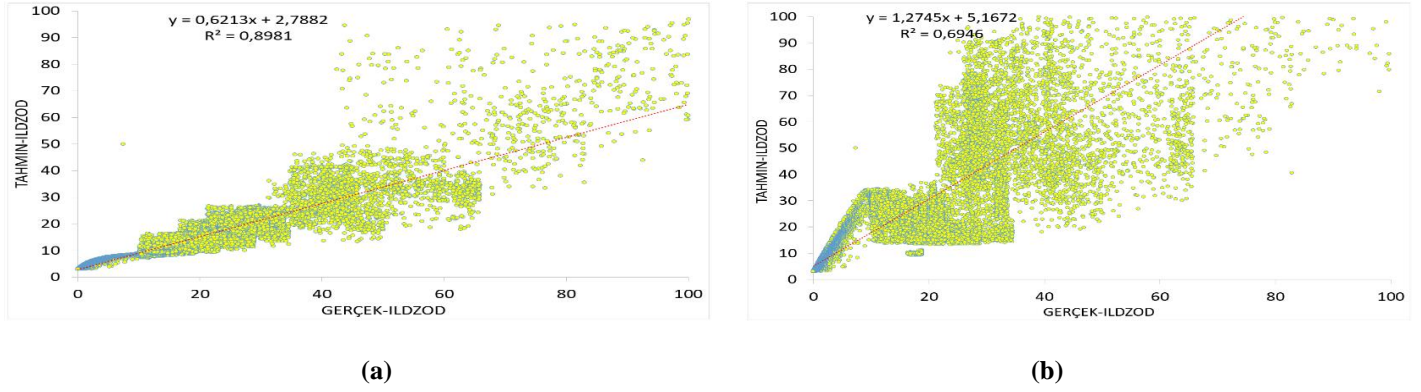
Dengeli veri seti için bulanık mantık temelli modellerin R^2 performans değerlerini içeren grafikler Mamdani-tip bulanık çıkarım sistemleri için Şekil 8’de ve Sugeno-tip bulanık çıkarım sistemleri için ise Şekil 9’da verilmiştir. Bu şekillerde resmedilen sonuçlar, R^2 metriği açısından en iyi performansın 0.9134 değeri ile mak-min çıkarım yöntemi ve ağırlık merkezi durulaştırma yöntemini kullanan Mamdani-tip bulanık çıkarım sistemi tarafından elde edildiğini göstermiştir. Bununla birlikte mak-min çıkarım yöntemi ve alan açıcı ortası durulaştırma yöntemini kullanan Mamdani-tip bulanık çıkarım sistemi de 0.9084’lük R^2 değeri ile en iyi performansa oldukça

yakın bir sonuç üretmiştir. R^2 metriği açısından en kötü performansı ise 0.6946 değeri ile ağırlıklı toplam yöntemini kullanan Sugeno-tip bulanık çıkarım sistemi sergilemiştir.

Modellerin dengesiz ve dengeli veri setlerinde elde ettikleri R^2 değerleri karşılaştırıldığında, mak-min çıkarım yöntemi ve ağırlık merkezi durulaştırma metodunu kullanan Mamdani-tip bulanık çıkarım yöntemi haricindeki tüm modellerin, dengeli veri seti için daha yüksek R^2 değerleri elde ettikleri gözlemlenmiştir. Sugeno-tip bulanık çıkarım sisteminde dengeli veri setinde ağırlıklı ortalamama yöntemini kullanan modelin R^2 değeri dengesiz veri setine göre artarken, ağırlıklı toplam yöntemini kullanan modelin R^2 değerinde bir düşüş olduğu görülmüştür.



Şekil 8. Dengeli veri seti için, Mamdani-tip bulanık çıkarım sistemlerine ait R^2 eğrileri (a) Mamdani – Mak-Min, Ağırlık merkezi; (b) Mamdani – Mak-Min, Alan Açıortayı; (c) Mamdani – Mak-Min, En Büyüklerin Ortası; (d) Mamdani – Mak-Çarp, Ağırlık Merkezi; (e) Mamdani – Mak-Çarp, Alan Açıortayı ; (f) Mamdani – Mak-Çarp, En Büyüklerin Ortası



Şekil 9. Dengeli veri seti için, Sugeno-tip bulanık çıkarım sistemlerine ait R^2 eğrileri (a) Sugeno-Ağırlıklı ortalama; (b) Sugeno-Ağırlıklı toplam

Elde edilen sonuçlar hem dengeli hem de dengesiz veri setlerinde R^2 , doğruluk, duyarlılık ve F-Score performans metrikleri açısından mak-min çıkarım yöntemi ve ağırlık merkezi durulaştırma yöntemini kullanan Mamdani-tip bulanık çıkarım modelinin en yüksek performansı sergilediğini göstermiştir. Ayrıca, bu sonuçlar dengesiz veri setinde ÖNEMLİ sınıfta bulunan örneklem sayısının az olması sebebi ile modellerin doğruluk metriği açısından ortaya çıkan ölçüm belirsizliğinin, SMOTE yönteminin kullanılması ile dengeli hale getirilen veri setinde ortadan kalktığı ve bu yöntemin belirleyicilik açısından duyarlılık ve F-Score metrikleri ile modellerin performansının ölçülmesinde önemli bir role sahip olduğunu da ortaya koymuştur.

Bu çalışmada kullanılan Mamdani ve Sugeno-tip bulanık çıkarımı temelli modeller literatürdeki çalışmalar ile karşılaştırıldığında, bu modellerin bir kısmının literatürdeki çalışmalara göre oldukça başarılı sonuçlar ürettiği görülmüştür. Oluşturulan modeller arasında en başarılı sonucu elde eden çıkarım yöntemi olarak mak-min ve durulaştırma yöntemi olarak ağırlık merkezi yöntemini içeren Mamdani-tip bulanık çıkarım sisteminin literatürdeki çalışmalar ile performans açısından karşılaştırılmasını gösteren bilgiler Tablo 8’de verilmiştir.

Tablo 8. Sosyal Medyada haber popülerliği tespiti için Mamdani-tip bulanık çıkarım sisteminin literatürdeki çalışmalar ile performans açısından karşılaştırılması

Çalışma	Yıl	Platform	Metot	Performans
(Naseri & Zamani, 2019)	2019	Telegram	MTL	81,3% F-score
(Deshpande, 2018)	2018	UCI MLR	LDA & AB	73% F-score
(Fernandes et al., 2015)	2015	Mashable	RF	69% F-score
(Ren & Yang, 2015)	2015	Mashable	RF	69% Doğruluk
(Arapakis et al., 2014)	2014	Yahoo	SVM	70,7% Doğruluk
(AL-Mutairi & Khan, 2015)	2015	Wikipedia	DT	62,6% Doğruluk
(Ahmed et al., 2013)	2013	Twitter	SiLA & k-NN	85% Doğruluk
Geliştirilen Mantık Temelli Model	Bulanık 2022	Facebook	Mamdani-tip bulanık çıkarım sistemi: Çıkarım yöntemi-->mak-min, Durulaştırma yöntemi-->ağırlıklı ortalama	96.72 % Doğruluk, 96.60 % F-score

Tablo 8’den bu çalışmada önerilen modeller arasında en iyi performansa sahip modelin, daha önce literatürde haber popülerliği alanında yapılan çalışmalara kıyasla hem doğruluk hem de F-Score metrikleri açısından daha başarılı sonuçlar elde ettiği açık bir şekilde görülmüştür. Bu bilgiler ışığında, sosyal medyada haber popülerliği analizinde makine öğrenmesi temelli metotların yanı sıra, bulanık mantık temelli metotların da efektif bir şekilde kullanılabileceği sonucuna ulaşılabilir.

5. Sonuçlar

Bu çalışmada Facebook platformunda paylaşılan “Economy, Microsoft, Obama & Palestine” konu başlıklarının bir araya getirilmesi ile oluşturulan veri setine; Mamdani ve Sugeno-tip bulanık çıkarım sistemlerinin çıkarım ve durulaştırma yöntemlerine göre yapılandırılmasından oluşan 8 tahmin modeli uygulanarak, bulanık mantık temelli tahmin modellerinin sosyal medya haber popülerliğinin tespitindeki kullanılabilirliği literatürdeki çalışmaların performansları ile karşılaştırılarak değerlendirilmiştir. Gerçekleştirilen çalışma veri setinin oluşturulması, model girdilerinin belirlenmesi ve modellerin yürürlüğe konması üzere üç aşamadan oluşur. İlk aşamada “Economy, Microsoft, Obama & Palestine” konu başlıklarından oluşan haberleri içeren 4 ayrı veri seti KU-Irvine Makine Öğrenmesi Deposundan indirilerek birleştirilmiş ve tek bir veri seti haline getirilmiştir. Bu veri setinde

görsütülenmeyen haberleri içeren veriler silinmiş ve 20 dakikalık görüntülenme periyodu şeklinde verilen öznitelikler 4 saatlik görüntülenme sayılarını yansıtabilecek şekilde yeniden düzenlenerek, veri setinin hem öznitelik sayısı azaltılmış hem de haberin görüntülenme sayısının periyodik olarak yorumlanması daha kolay hale getirilmiştir. Elde edilen veri setinde ÖNEMLİ haber sınıfında 247 örneklem ve ÖNEMSİZ haber sınıfında 62912 örneklem olduğu için dengesiz bir veri setidir. Dengesiz veri setlerinde sınıflardan birindeki örneklem sayısı az olduğu için bazı metrikler açısından performans ölçümleri doğru bir şekilde ölçülemez. Bu yüzden dengesiz veri setine SMOTE'nin uygulanması sonucunda ÖNEMLİ sınıfında 62910 ve ÖNEMSİZ sınıfında 62912 örneklem içeren dengeli bir veri seti oluşturulmuştur. Modellerin performansı her iki veri setinde de test edilmiştir. İkinci aşamada filtre temelli öznitelik seçim metodlarından Fisher skor metodu uygulanarak en yüksek kazançla sahip üç öznitelik model girdileri olarak belirlenmiştir. Son aşamada yapılandırılmış 8 farklı bulanık mantık modeli yürürlüğe konarak veri setinde haberlerin önem dereceleri belirlenmiştir. Modellerin performansının test edilmesinde R² metriği ile karışıklık matrisi temelli ölçümlerden doğruluk, duyarlılık ve F-Score metriklerinden yararlanılmıştır. Deneysel çalışmalardan elde edilen sonuçlar hem dengesiz veri setinde hem de dengeli veri setinde mak-min çıkarım yöntemi ve ağırlık merkezi durulaştırma yöntemini kullanan Mamdani-tip bulanık çıkarım sisteminin tüm performans metrikleri açısından en yüksek performansı sağladığını göstermiştir. Bu çalışmada önerilen modellerin dengeli veri setindeki doğruluk ve F-Score metrikleri açısından literatürdeki çalışmalar ile karşılaştırılması sonucunda, ağırlıklı ortalama yöntemini kullanan Sugeno-tip bulanık çıkarım yöntemi dışındaki bulanık mantık temelli modellerin, çoğunluğu makine öğrenmesi temelli modellerden oluşan, literatürdeki çalışmalardan daha yüksek bir performansa sahip olduğu görülmüştür. Bu durum aynı zamanda, önerilen bulanık mantık temelli modellerin, diğer birçok alanda olduğu gibi haber popülerliğinin tespiti için de efektif bir yöntem olarak kullanılabileceğini göstermiştir.

Bilgilendirme

Bu çalışma mühendislik bilimleri araştırma ve geliştirme uygulamaları özel sayısı çağırısı kapsamında değerlendirilmiştir.

Referanslar

- Aghasian, E., Garg, S., & Montgomery, J. (2020). An automated model to score the privacy of unstructured information—Social media case. *Computers & Security*, 92. <https://doi.org/10.1016/j.cose.2020.101778>
- Ahmed, H., Razzaq, M. A., & Qamar, A. M. (2013). Prediction of popular tweets using Similarity Learning. *ICET 2013 - 2013 IEEE 9th International Conference on Emerging Technologies*. <https://doi.org/10.1109/ICET.2013.6743524>
- AL-Mutairi, H. M., & Khan, M. B. (2015). Predicting the Popularity of Trending Arabic Wikipedia Articles Based on External Stimulants Using Data/Text Mining Techniques. *2015 International Conference on Cloud Computing, ICC3 2015*. <https://doi.org/10.1109/CLOUDCOMP.2015.7149651>
- Appel, O., Chiclana, F., Carter, J., & Fujita, H. (2016). A hybrid approach to the sentiment analysis problem at the sentence level. *Knowledge-Based Systems*, 108, 110–124. <https://doi.org/10.1016/j.knsys.2016.05.040>
- Arapakis, I., Barla Cambazoglu, B., & Lalmas, M. (2014). On the feasibility of predicting news popularity at cold start. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8851, 290–299. https://doi.org/10.1007/978-3-319-13734-6_21/COVER/
- Augusto, M., Godinho, P., & Torres, P. (2019). Building customers' resilience to negative information in the airline industry. *Journal of Retailing and Consumer Services*, 50, 235–248. <https://doi.org/10.1016/j.jretconser.2019.05.015>
- Beştaş, M. (2020). SOSYAL MEDYADA HABER POPÜLERLİĞİNİN TAHMİNİ: LİTERATÜR İNCELEMESİ. *International Journal of Social Humanities Sciences Research (JSHSR)*, 7(61), 3140–3155. <https://doi.org/10.26450/jshsr.2144>
- Caruana, R., & Niculescu-Mizil, A. (2004). *Data mining in metric space*. 69. <https://doi.org/10.1145/1014052.1014063>
- Caruana, R., Pratt, L., & Thrun, S. (1997). Multitask Learning. *Machine Learning 1997* 28:1, 28(1), 41–75. <https://doi.org/10.1023/A:1007379606734>
- Chawla, N. v., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*, 16(1), 321–357.
- Chew, A. W. Z., Pan, Y., Wang, Y., & Zhang, L. (2021). Hybrid deep learning of social media big data for predicting the evolution of COVID-19 transmission. *Knowledge-Based Systems*, 233. <https://doi.org/10.1016/j.knsys.2021.107417>
- Colin Cameron, A., & Windmeijer, F. A. G. (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, 77(2), 329–342. [https://doi.org/10.1016/S0304-4076\(96\)01818-0](https://doi.org/10.1016/S0304-4076(96)01818-0)

- Deshpande, D. (2018). Prediction Evaluation of Online News Popularity Using Machine Intelligence. *2017 International Conference on Computing, Communication, Control and Automation, ICCUBEA 2017*. <https://doi.org/10.1109/ICCUBEA.2017.8463790>
- Dhawan, A., Bhalla, M., Arora, D., Kaushal, R., & Kumaraguru, P. (2022). FakeNewsIndia: A benchmark dataset of fake news incidents in India, collection methodology and impact assessment in social media. *Computer Communications*, *185*, 130–141. <https://doi.org/10.1016/j.comcom.2022.01.003>
- Fernandes, K., Vinagre, P., & Cortez, P. (2015). A proactive intelligent decision support system for predicting the popularity of online news. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9273*, 535–546. https://doi.org/10.1007/978-3-319-23485-4_53
- Fernandez, A., Garcia, S., Herrera, F., & Chawla, N. v. (2018). SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. *Journal of Artificial Intelligence Research*, *61*, 863–905. <https://doi.org/10.1613/jair.1.11192>
- Fischer, U., Kopka, L., & Grabbe, E. (1999). Breast Carcinoma: Effect of Preoperative Contrast-enhanced MR Imaging on the Therapeutic Approach. *Radiology*, *213*(3), 881–888. <https://doi.org/10.1148/radiology.213.3.r99dc01881>
- Francisco, M., & Castro, J. L. (2020). A fuzzy model to enhance user profiles in microblogging sites using deep relations. *Fuzzy Sets and Systems*, *401*, 133–149. <https://doi.org/10.1016/j.fss.2020.05.006>
- Garcia, J. (2022). Bankruptcy prediction using synthetic sampling. *Machine Learning with Applications*, *9*, 100343. <https://doi.org/10.1016/J.MLWA.2022.100343>
- Goodman, L. A., & Kruskal, W. H. (2012). Measures of Association for Cross Classifications*. <https://doi.org/10.1080/01621459.1954.10501231>, *49*(268), 732–764. <https://doi.org/10.1080/01621459.1954.10501231>
- Howells, K., & Ertugan, A. (2017). Applying fuzzy logic for sentiment analysis of social media network data in marketing. *Procedia Computer Science*, *120*, 664–670. <https://doi.org/10.1016/j.procs.2017.11.293>
- Karyotis, C., Doctor, F., Iqbal, R., & James, A. (2015). An intelligent framework for monitoring students Affective Trajectories using adaptive fuzzy systems. *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1–8. <https://doi.org/10.1109/FUZZ-IEEE.2015.7338001>
- Kazemzadeh, A., Lee, S., & Narayanan, S. (2013). Fuzzy Logic Models for the Meaning of Emotion Words. *IEEE Computational Intelligence Magazine*, *8*(2), 34–49. <https://doi.org/10.1109/MCI.2013.2247824>
- Kirkland, T., & Cunningham, W. A. (2012). Mapping emotions through time: How affective trajectories inform the language of emotion. *Emotion*, *12*(2), 268–282. <https://doi.org/10.1037/a0024218>
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Proceedings of the 19th International Conference on World Wide Web - WWW '10*, 591–600. <https://doi.org/10.1145/1772690.1772751>
- Labatut, V., & Cherifi, H. (2011). Evaluation of Performance Measures for Classifiers Comparison. *Ubiquitous Computing and Communication Journal*, *6*, 21–34.
- Liu, S.-Y., Xiao, J., & Xu, X.-K. (2020). Sign prediction by motif naive Bayes model in social networks. *Information Sciences*, *541*, 316–331. <https://doi.org/10.1016/j.ins.2020.05.128>
- Luo, M., & Hancock, J. T. (2020). Self-disclosure and social media: motivations, mechanisms and psychological well-being. *Current Opinion in Psychology*, *31*, 110–115. <https://doi.org/10.1016/j.copsyc.2019.08.019>
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. <https://doi.org/10.1037/h0043158>
- Moniz, N., & Torgo, L. (2018). *Multi-Source Social Feedback of Online News Feeds*. <https://doi.org/10.48550/arxiv.1801.07055>
- Morente-Molinera, J. A., Kou, G., Pang, C., Cabrerizo, F. J., & Herrera-Viedma, E. (2019). An automatic procedure to create fuzzy ontologies from users' opinions using sentiment analysis procedures and multi-granular fuzzy linguistic modelling methods. *Information Sciences*, *476*, 222–238. <https://doi.org/10.1016/j.ins.2018.10.022>
- Naseri, M., & Zamani, H. (2019). Analyzing and Predicting News Popularity in an Instant Messaging Service. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1053–1056. <https://doi.org/10.1145/3331184.3331301>

- Phuvipadawat, S., & Murata, T. (2010). Breaking News Detection and Tracking in Twitter. *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 120–123. <https://doi.org/10.1109/WI-IAT.2010.205>
- Ren, H., & Yang, Q. (2015). *Predicting and Evaluating the Popularity of Online News*.
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. <https://doi.org/10.1037/0033-295X.110.1.145>
- Saeed, R., Abbas, H., Asif, S., Rubab, S., Khan, M. M., Iltaf, N., & Mussiraliyeva, S. (2022). A framework to predict early news popularity using deep temporal propagation patterns. *Expert Systems with Applications*, *195*. <https://doi.org/10.1016/j.eswa.2021.116496>
- Statista. (2022, June 15). *Social Meia - Statistics & Facts*. Statista.
- Szabo, G., & Huberman, B. A. (2008). Predicting the Popularity of Online Content. *SSRN Electronic Journal*, *53*(8). <https://doi.org/10.2139/ssrn.1295610>
- Tavana, M., Momeni, E., Rezaeiniya, N., Mirhedayatian, S. M., & Rezaeiniya, H. (2013). A novel hybrid social media platform selection model using fuzzy ANP and COPRAS-G. *Expert Systems with Applications*, *40*(14), 5694–5702. <https://doi.org/10.1016/j.eswa.2013.05.015>
- T.K., B., Annavarapu, C. S. R., & Bablani, A. (2021). Machine learning algorithms for social media analysis: A survey. *Computer Science Review*, *40*. <https://doi.org/10.1016/j.cosrev.2021.100395>
- Vashishtha, S., & Susan, S. (2019). Fuzzy rule based unsupervised sentiment analysis from social media posts. *Expert Systems with Applications*, *138*. <https://doi.org/10.1016/j.eswa.2019.112834>
- Whelan, E., Islam, A. K. M. N., & Brooks, S. (2020). Applying the SOBC paradigm to explain how social media overload affects academic performance. *Computers & Education*, *143*. <https://doi.org/10.1016/j.compedu.2019.103692>
- Wu, B., & Shen, H. (2015). Analyzing and predicting news popularity on Twitter. *International Journal of Information Management*, *35*(6), 702–711. <https://doi.org/10.1016/j.ijinfomgt.2015.07.003>
- Wu, D. (2012). Fuzzy sets and systems in building closed-loop affective computing systems for human-computer interaction: Advances and new research directions. *2012 IEEE International Conference on Fuzzy Systems*, 1–8. <https://doi.org/10.1109/FUZZ-IEEE.2012.6250779>
- Wu, H., Yue, K., Pei, Y., Li, B., Zhao, Y., & Dong, F. (2016). Collaborative Topic Regression with social trust ensemble for recommendation in social media systems. *Knowledge-Based Systems*, *97*, 111–122. <https://doi.org/10.1016/j.knosys.2016.01.011>
- Xia, B., Ni, Z., Li, T., Li, Q., & Zhou, Q. (2017). VRer: Context-Based V enue R ecommendation using e mbedded space r anking SVM in location-based social network. *Expert Systems with Applications*, *83*, 18–29. <https://doi.org/10.1016/j.eswa.2017.04.020>
- Xiong, J., Yu, L., Zhang, D., & Leng, Y. (2021). DNCP: An attention-based deep learning approach enhanced with attractiveness and timeliness of News for online news click prediction. *Information & Management*, *58*(2). <https://doi.org/10.1016/j.im.2021.103428>
- Zaman, T., Fox, E. B., & Bradlow, E. T. (2014). A Bayesian approach for predicting the popularity of tweets. *The Annals of Applied Statistics*, *8*(3), 1583–1611. <https://doi.org/10.1214/14-AOAS741>
- Zhang, Z., Liu, H., Chen, D., Zhang, J., Li, H., Shen, M., Pu, Y., Zhang, Z., Zhao, J., & Hu, J. (2022). SMOTE-based method for balanced spectral nondestructive detection of moldy apple core. *Food Control*, *141*. <https://doi.org/10.1016/j.foodcont.2022.109100>