DÜMF Mühendislik Dergisi

web: http://dergipark.gov.tr/dumf

Çevrimiçi sosyal medyada sahte haber tespiti

Feyza ALTUNBEY ÖZBAY
Fırat Üniversitesi, Mühendislik Fakültesi, Yazılım Mühendisliği Bölümü, Elazığ
faltunbey@firat.edu.tr ORCID: 0000-0003-0629-6888, Tel: (424) 237 00 00 (5575)

Bilal ALATAŞ*
Fırat Üniversitesi, Mühendislik Fakültesi, Yazılım Mühendisliği Bölümü, Elazığ
balatas@firat.edu.tr ORCID: 0000-0002-3513-0329, Tel: (424) 237 00 00 (5599)

Geliş: 04.10.2019, Revizyon: 16.01.2020, Kabul Tarihi: 27.01.2020

Öz

Son yıllarda, sosyal medyadan haber almak, bilginin hızlı bir şekilde yayılması, ucuz maliyet ve kolay erişim nedeniyle giderek daha popüler hale gelmiştir. Sosyal medyanın dünyadaki insanlar için temel bilgi kaynaklarından biri haline gelmesi toplum, kültür ve iş dünyası üzerinde olumlu ve olumsuz etkilere sahiptir. Sosyal medyadaki haberlerin kalitesi geleneksel haber kaynaklarından daha düşüktür ve sosyal medya sahte haber yaymak için çok uygundur. Sahte haberlerin insanlar ve toplum üzerindeki zararlı etkileri nedeniyle, sahte haberlerin tespiti dikkat çekmektedir. Bu çalışmada, sahte haberleri tespit etmek için iki aşamalı bir model önerilmiştir. İlk adımda, yapılandırılmamış verileri yapılandırılmış verilere dönüştürmek için sahte haberler içeren verilere bir dizi ön işlem uygulanmıştır. Bir sonraki adımda, yapılandırılmış sahte haber veri setine on denetimli yapay zekâ algoritması uygulanmıştır. Önerilen model dört farklı eğitim - test bölümlemesi ile incelenmiştir. Erişime açık veri seti üzerinde; Naive Bayes, JRip, J48, Rastgele Orman, Stokastik Gradyan İnişi, Yerel Ağırlıklı Öğrenme, Naive Bayes ile Karar Ağacı, Yerine Koyarak Öğrenme, Regresyon ile Sınıflandırma denetimli yapay zekâ algoritmaları test edilmiş ve bu algoritmalar üç değerlendirme ölçütüne bağlı olarak karşılaştırılmıştır.

Anahtar Kelimeler: *Yapay zekâ algoritmaları; Sahte haber tespit; Çevrimiçi sosyal medya*

* Yazışmaların yapılacağı yazar

DOI: 10.24012/dumf.629368

Giriş

Teknolojideki hızlı gelişim, insanların bilgiye erişim için kullandıkları kaynakları değiştirmiştir. İnternetin ortaya çıkışı ile birlikte sosyal medya, dünya genelinde insanların bilgiye erişim için kullandığı temel bir kaynak haline gelmiştir. Özellikle son yıllarda gazete, televizyon, radyo gibi geleneksel haber kaynakları yerine Twitter, Facebook gibi çevrimiçi sosyal medya platformları popüler duruma gelmiştir. İnsanların sosyal medyadaki haber kaynaklarını kullanmalarındaki temel neden; sosyal medya kaynaklarının düşük maliyetli ve kolay erişilebilir olması, ayrıca bilginin hızlı yayılmasını sağlamasıdır. Bu nedenle, her geçen gün sosyal medyadaki haberleri takip eden kullanıcı sayısı artmaktadır. Bu avantajlar, sosyal etkileşim ve bilgi paylaşımı için her yerde hazır bir platform oluşturmuştur. Sosyal medya, coğrafi sınırları olmayan milyonlarca üyesi olan sosyal gruplar oluşturulmasını kolaylaştırmıştır. Ayrıca, sosyal medya kullanıcıları, sosyal medyadaki "Paylaş" butonu ile haber makalelerini tüm grup üyeleri ile paylaşabilir. Böylelikle, sosyal medya tek bir tıklama ile milyonlarca insanın erişebileceği bir makaleyi yayma yeteneğine sahiptir. Sosyal medya birçok avantaj sağlamasına rağmen, sosyal medyadaki haberlerin kalitesi geleneksel haber platformlarına kıyasla daha düşüktür. Bazen sosyal medyadaki haberlerin içerikleri kötü niyetli kullanıcılar tarafından farklı amaçlar için değiştirilmektedir. Buna ek olarak, bu tip içerikler kontrol edilmeden iyi niyetli kişiler tarafından paylaşılarak yayılmaktadır.

Sosyal medyadaki haberler ve yorumlar kullanıcıların fikirlerini önemli ölçüde etkiler. Sahte haber olarak adlandırılan düşük kaliteli bu tip haberlerin yayılması bireyleri ve toplumları olumsuz yönde etkilemektedir. Sahte haberler sadece bireyler ve toplumlar için değil, iş dünyası ve hükümetler için de tehlike oluşturabilmektedir. Bu nedenle, çevrimiçi sosyal medyadaki sahte haberlerin belirlenip tespit edilmesi gerekmektedir.

Sahte haberlerin tespit edilmesi problemi sosyal medyada yeni bir araştırma alanı olmasına rağmen, son yıllarda önemli ölçüde dikkat çekmiştir. Problem, araştırmacılar tarafından farklı bakış açıları ile ele alınmıştır. Literatürde sahte haberlerin tespitine ve algılanmasına odaklanarak, sosyal medyadaki sahte haberlerin tespitinin geniş kapsamlı bir incelemesi yapılmıştır (Shu vd., 2017). Bir diğer çalışmada, Facebook haber gönderileri kullanılarak oluşturulan bir yazılım sistemi ile sahte haber tespiti yapılmıştır (Granik ve Mesyura, 2017). Yazarlar, oluşturdukları sistemde Naive Bayes algoritmasını kullanarak % 74 doğruluk oranı ile bir başarım elde etmişlerdir. Sahte haber olarak kabul edilen haberlerin yayılmasını engellemek için yapılan başka bir çalışmada sosyal ağ kullanıcılarının bayrakları bir araya getirilerek, uzmanlar tarafından sahte haber alt kümeleri belirlenir ve haberlerin yayılması durdurulur (Tschiatschek vd., 2018). Makine öğrenmesi algoritmaları temel alınarak önerilen bir yöntemde, Facebook verileri kullanılarak yöntemlerinin doğruluğunu test edilmiştir. Önerilen yöntemin doğruluk değeri % 82'dir (Vedova vd., 2018). Haber makaleleri üzerinde yapılan araştırmaları kolaylaştırmak için haber bilgi getiriminde son eğilimler konferansı (Corney vd., 2016) ile bağlantılı olan Sinyal Medya tarafından yayınlanan bir veri kümesi, makine öğrenmesi yöntemleri kullanılarak sahte haberleri belirlemek için kullanılmıştır (Gilda, 2017). Sahte haberleri, gerçek haberlerden ayırt etmek için hiciv ipuçları kullanılan bir diğer çalışmada, önerilen yöntemi test etmek için sivil toplum, bilim, ticaret ve yumuşak haber olmak üzere 4 ayrı alandaki sahte ve gerçek haberlerden, hiciv haberlerden yararlanılmıştır (Rubin vd., 2016). Sahte haberleri tespit etmek için 3 farklı özelliği birleştiren melez bir yöntem önerilmiştir. Özellikle, hem kullanıcıların ve makalelerin davranışları, hem de sahte haberleri yayınlayan kullanıcıların davranışları birleştirilmiştir. Bu üç özellikten hareketle, Yakalama, Puanlama ve Birleştirme adımlarını bir araya getiren melez bir yöntem geliştirilmiştir. Önerilen yöntem Twitter ve Weibo sosyal ağları üzerinde test edilmiştir (Ruchansky vd., 2017). Sahte haberleri tespit

etmek için, haber konuları ve haberi oluşturanlar arasındaki bağlantı bilgisini kullanarak, derin ayrıntılı ağ modeli önerilmiştir (Zhang vd., 2018). Bir diğer çalışmada, sosyal ağdaki kullanıcı hesap bilgileri de göz önüne alınarak sahte haberleri tespit etmek için yeni bir yöntem önerilmiştir (Long vd., 2017). Bu yöntemde, herhangi bir kullanıcının yaptığı haberlerin doğruluğunu tespit etmek için, aynı kullanıcının hesabındaki paylaşımlar değerlendirilir. Önerilen yöntem, sahte haber veri kümesi üzerinde test edilerek % 41.5 oranında bir doğruluk elde edilmiştir. Bir diğer çalışmada, güvenilmez metinlerin dil özelliklerini belirlemek için, gerçek haberleri, hiciv ve aldatmaca haberler ile kıyaslanmıştır (Rashkin vd., 2017). Deneyleri, metinlerin doğruluğunu tespit etmek için biçimsel ipuçları belirlemeye uyarlanmıştır. Metasezgisel optimizayon yöntemlerini kullanarak sosyal medyadaki sahte haberleri tespit etmek için 2 aşamadan oluşan bir yöntem önerilmiştir. Önerilen yöntemi test etmek için 3 farklı gerçek dünya verisi kullanarak farklı değerlendirme kriterlerine göre deney sonuçlarını vermişlerdir (Ozbay ve Alatas, 2019). Sahte haberleri tespit etmek için geometik derin öğrenme temelli bir model önerilmiştir (Monti vd., 2019). Önerilen yöntemi test etmek için Twitter tarafından doğrulanmış haber makalelerini kullanılmıştır. Bir diğer çalışmada, haberi yayınlayanın duygusu ve sosyal duygunun bir araya getirilerek duygu tabanlı sahte haber tespiti yöntemi önerilmiştir (Guo vd., 2019).

Bu çalışmada çevrimiçi sosyal medyadaki sahte haberleri tespit etmek için iki adımdan oluşan bir yöntem önerilmiştir. Model metin analizi ve denetimli yapay zekâ algoritmalarının birleşiminden oluşmuştur. Metin analizi yöntemleri ile yapısal olmayan haber metinleri yapısal biçime dönüştürülerek, modelin ikinci adımında kullanılan denetimli yapay zekâ algoritmalarının işleyebileceği duruma getirilmiştir. Çalışmada 10 adet denetimli yapay zekâ algoritması (Naive Bayes, JRip, J48, Rastgele Orman, Stokastik Gradyan İniş (SGD), Yerel Ağırlıklı Öğrenme (LWL), Karar Ağacı ve Naive Bayes (DTNB), Yerine Koyarak Öğrenme (Bagging), Regresyon ile Sınıflandırma (CvR)) ile yapısal haber veri kümeleri üzerinde test edilmiştir. Yapay zekâ yöntemleri gerçek dünya sahte haber verisi üzerinde 3 farklı metrik göz önüne alınarak test edilmiş ve sonuçlar karşılaştırmalı grafikler ve tablolar şeklinde sunulmuştur.

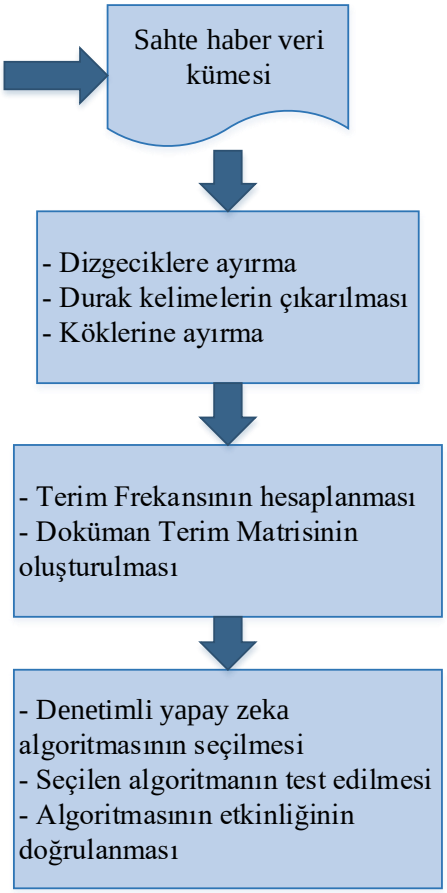
Sahte Haber Tespit Modeli

Bu bölümde, sahte haber tespiti için önerilen modelin ayrıntıları verilmektedir. Önerilen yöntem birbiri ile ilişkili adımlardan oluşmaktadır. Çevrimiçi sosyal medyadaki haberlerden oluşan yapısal olmayan veriden, sayılar ve durak kelimeler gibi gereksiz terim ve karakterler filtrelenerek ön işlem adımı gerçekleştirilir. Özellik uzayının boyutunu azaltmak için gereksiz terimler veriden çıkarılır ve daha sonra kelime sıklıklarına göre kök terimler seçilir. Elde edilen veri kümesinde her bir doküman içerisindeki terimler, Terim Frekans ağırlıklandırma yöntemi ile ağırlıklandırılmış ve her bir doküman terim ağırlıklarını içeren bir vektör olan Vektör Uzay Modeli ile temsil edilmiştir. Modelin son adımında, sahte haber tespit problemi bir sınıflandırma problemi olarak ele alınmış ve denetimli yapay zekâ algoritmaları sahte haber veri kümesine uygulanarak doğruluk, hassasiyet ve duyarlılık metriklerine göre karşılaştırılmıştır. Bu algoritmalar: Naive Bayes, JRip, J48, Random Forest, OneR, SGD, LWL, DTNB, Bagging, ClassificationViaRegression yöntemleridir. Önerilen sahte haber modelinin adımları Şekil 1'de verilmiştir.

Metin Madenciliği

Sosyal medya kullanıcıları, çevrimiçi sosyal medyada genellikle yapısal olmayan metin temelli veriler paylaşırlar. Bu verilerin, öğrenme sistemlerinde kullanılması için yapısal biçime dönüştürülmesi gerekir. Metin madenciliği yöntemleri bu aşamada kullanılarak, veriler yapısal biçime ve anlaşılabilir biçime dönüştürülerek yapay zekâ yöntemleri tarafından

işlenecek hale getirilir. Genel bir ifade olarak; metin madenciliği, büyük miktardaki metin temelli veriden kullanışlı bilginin aranması ve çıkarılmasıdır (Kumar ve Bhatia, 2013). Ön işlem adımları metin madenciliği tekniklerinde önemli rol oynar. Ön işlem adımları dizgeciklere ayırma, durak kelimelerin çıkarılması ve kök bulma olmak üzere 3 aşamada incelenmiştir.



Şekil 1. Sahte haber tespiti modelinin adımları

a) Dizgeciklere Ayırma

Dizgeciklere ayırma işlemi, metni dizge adı verilen küçük parçalara (sözcükler/ifadeler) ayırma görevidir. Aynı zamanda, tüm noktalama işaretleri, boşluk ve satır sonu karakterleri metin verisinden kaldırılır (Allahyari vd., 2017). Rakam içeren bütün terimler kaldırılır. Büyük/küçük harf ayrımı giderilir ve tek bir biçime dönüştürülür. Bu çalışmada metin içerisinde geçen terimlerin tamamı küçük harflere dönüştürülmüştür. En son adımda ise karakter sayısı *N*'den küçük olan kelimeler *N*-karakter filtresi ile silinmektedir. Bu çalışmada *N* = 3 olarak belirlenmiştir.

b) Durak Kelimelerin Çıkarılması

Metin içinde çok sık geçen fakat önemli olmayan kelimelerin çıkarılmasıdır. Durak kelimeler çıkarılarak terim uzayının boyutu azaltılır. Bağlaçlar, edatlar ve zarflar gibi kelimeler, durak kelimeler olarak kabul edilir.

c) Köklerine Ayırma

Bazı kelimeler aynı kökten oluşmalarına rağmen aldıkları eklere göre farklı anlamlar kazanmaktadır. Bu tip kelimelerin metin içerisinde geçme sıklığını belirlemek için köklerinin bulunması gerekir.

Yapılan çalışmada uygulanan metin ön işleme adımları Tablo 1'de listelenmiştir.

Tablo 1. Metin ön işlem adımları

Girdi: Metin verisi
Çıktı: İşlenmiş veri
1. Sayısal ifadelerin kaldırılması.
2. Noktalama işaretlerinin silinmesi.
3. *N* karakterden az kelimelerin silinmesi.
4. Büyük küçük harf ayrımının giderilmesi.
5. Durak kelimelerin çıkarılması.
6. Kelimelerin köklerine ayrılması.

Metin madenciliği çalışmalarındaki en büyük sorun yüksek boyutlu verilerdir. Bu nedenle, önerilen modelin başarımını artırmak için gereksiz özelliklerin metinden çıkarılması gerekir. Metin madenciliğinde özellik çıkarımı, metin özellik boyutunu azaltmak ve kullanışlı veri kümesi elde etmek için kullanılır. Veri kümesindeki her bir doküman için terimler ağırlıklandırılır ve her bir doküman terimlerin ağırlık vektörüne dönüştürülür. Bu temel gösterim Vektör Uzay Modeli (VUM) olarak adlandırılır. VUM'da, her bir kelime, kelimenin doküman içindeki ağırlığını gösteren bir değer ile ifade edilir. Bu çalışmada terimleri ağırlıklandırmak için Terim Frekansı (TF) yöntemi kullanılmıştır. TF bir kelimenin bir

dokümanda görülme sıklığını ifade eder ve Denklem 1 ile hesaplanır.

$$Terim\ Frekansı: TF = \frac{n_{ij}}{|d_i|} \quad (1)$$

Denklemde *di i*. dokümandaki tüm terimlerin toplam sayısıdır. *nij* ise *i*. dokümanda *j*. kelimenin sayısını temsil etmektedir.

Her bir dokümandaki kelimelerin TF değerleri hesaplandıktan sonra, terim ağırlıklarına göre Şekil 2'de gösterilen *m* x *n* boyutunda bir Doküman Terim Matrisi (DTM) oluşturulur.

$$\begin{matrix} D_1 \\ D_1 \\ . \\ . \\ . \\ D_m \end{matrix} \begin{bmatrix} A_{11} & A_{12} & A_{13} & \cdots & A_{1n} \\ A_{21} & A_{22} & A_{23} & \cdots & A_{2n} \\ . & & & & . \\ . & & & & . \\ . & & & & . \\ A_{m1} & A_{m2} & A_{m3} & \cdots & A_{mn} \end{bmatrix}$$

Şekil 2. Doküman Terim Matrisi

Matriste her bir satır dokümanları, sütunlar ise terimleri ifade eder. Her bir hücre ise, dokümandaki terimlerin ağırlıklarını gösterir.

Performans Değerlendirme Kriterleri

Sahte haberleri tespit etmek için veri kümesine uygulanan denetimli yapay zekâ algoritmalarının başarılarını değerlendirmek için birçok farklı performans değerlendirme kriteri bulunmaktadır. Algoritmaların başarılarını ölçmek için öncelikle bir karmaşıklık matrisinin oluşturulması gerekmektedir. Sahte haber tespiti problemi için oluşturulan karmaşıklık matrisi Tablo 2'de verilmiştir. Matriste verilen terimler (DP, YP, YN, DN) sahte haber problemi için aşağıdaki gibi ifade edilmiştir.

- Doğru Pozitif (DP): Tahmin edilen sahte haber aslında sahte bir haber ise, tahmin DP'dir.
- Yanlış Pozitif (YP): Tahmin edilen sahte haber aslında gerçek bir haber ise, tahmin YP'dir.
- Doğru Negatif (DN): Tahmin edilen gerçek haber aslında gerçek bir haber ise, tahmin DN'dir.
- Yanlış Negatif (YN): Tahmin edilen gerçek haber aslında sahte bir haber ise, tahmin YN'dir.

Tablo 2. Karmaşıklık matrisi

Karmaşıklık matrisi		**Gerçek sınıf**	
		Pozitif Sınıf	**Negatif Sınıf**
Tahmin edilen sınıf	**Pozitif Sınıf**	Doğru Pozitif (DP)	Yanlış Pozitif (YP)
	Negatif Sınıf	Yanlış Negatif (YN)	Doğru Negatif (DN)

Karmaşıklık matrisindeki ifadelere göre, uygulanan denetimli yapay zekâ algoritmalarının değerlendirilmesi için bu çalışmada Denklem (2-4) kullanılmıştır (Sokolova vd. 2006).

$$Doğruluk = \frac{|DN|+|DP|}{|YN|+|YP|+|DN|+|DP|} \quad (2)$$

$$Hassasiyet = \frac{|DP|}{|YP|+|DP|} \quad (3)$$

$$Duyarlılık = \frac{|DP|}{|YN|+|DP|} \quad (4)$$

Sahte haber tespiti problemi için doğruluk değeri, tüm veri kümesi içerisinden doğru tahmin edilen haberlerin oranıdır. Hassasiyet, sahte olarak tahmin edilen haberler içerisindeki, doğru tahmin edilen sahte haberlerin oranını gösterir. Duyarlılık ise, tahmin edilen sahte haberlerin, tüm sahte haberlere oranını ifade eder.

Yapay Zekâ Algoritmaları

Denetimli yapay zekâ algoritmaları, eğitim setine dayalı bir öğrenme şeklidir. Bu algoritmalar, eğitim verisindeki örneklerin etiketlerin bilindiğini varsayarlar. Denetimli yapay zekâ algoritmaları çalışmak için etiketli eğitim veri kümesi gerektirir ve denetlenecek verilere örnek olacak bir model döndürürler. Modelin görevi, giriş verisi için en az hata oranı ile bir çıktı

tahmin etmektir. Aşağıdaki bölümde, bu çalışmada kullanılan denetimli yapay zekâ algoritmaları hakkında bilgi verilmiştir.

a) Naive Bayes

Naive Bayes adını İngiliz matematikçi Thomas Bayes'ten alan veri madenciliği ve makine öğrenmesi için etkili ve verimli bir algoritmadır. Naive Bayes sınıflandırıcısı, tüm özelliklerin eşit derecede bağımsız olduğunu varsayan güçlü bağımsızlık varsayımlarıyla birlikte Bayes Teoremini uygulamaya dayanan basit bir olasılık sınıflandırıcısıdır. Bayes sınıflandırıcısı, bir sınıfın belirli bir özelliğinin varlığının (veya yokluğunun), başka bir özelliğin varlığı (veya yokluğu) ile ilgisiz olduğunu varsayar (Vaishali vd. 2014).

b) JRip

JRip algoritması, "Repeated Incremental Pruning to Produce Error Reduction (RIPPER) - Hata Üretmek İçin Tekrarlanan Artımsal Budama" olarak adlandırılan bir önermeli kural öğrenicisi kullanır. Bu yöntemde, sınıflar büyüyen boyutlarda incelenir. Sınıflar için ilk kural kümesi RIPPER kullanılarak üretilir. Algoritma, bir sınıfın tüm örnekleri kapsayan kural buluncaya kadar ilerler (Vaishali vd. 2014).

c) J48

J48 algoritması, genellikle sınıflandırma uygulamaları için tercih edilen algoritmadır. J48, ID3 ve C4.5 algoritmalarına dayanan istatistiksel bir karar ağacı algoritmasıdır. Her ağaçtan bir düğüm kullanma mantığı ile düğümler üzerinde çalışır. Bu nedenle, sınıflandırma algoritmaları arasında en hızlı ve ne yüksek hassasiyete sahip algoritmalardan biridir (Coşkun ve Baykal, 2011).

d) Rastgele Orman

Rastgele Orman algoritması, Tin Hou tarafından önerilmiştir (Ho, 1995). Rastgele orman, rastgele bir vektörü değerlerine bağlı olacak şekilde ağaç tahmin ediciler topluluğudur. Belirtilen vektör, ormandaki tüm ağaçlar için aynı dağılımdan bağımsız olarak örneklenir (Khan vd. 2010).

e) OneR

OneR, Holte tarafından önerilen basit ve hızlı bir algoritmadır (Holte, 1993). Bu yöntemde, tek bir niteliğe dayalı basit kurallar üretilir. Algoritma, tahmin için minimum hata oranına sahip bir kural seçer (Frank ve Witten, 1998). İki veya daha fazla kural aynı hata oranına sahipse, kural rastgele seçilir.

f) Stokastik Gradyan İniş (SGD – Stochastic Gradient Descent)

Stokastik Gradyan İniş, amaç fonksiyonunu optimize etmek için yinelemeli bir metot kullanan modern bir sınıflayıcıdır. Algoritma, gradyanları değerlendirmek için rastgele seçilen örnekleri kullanır, bu nedenle stokastik olarak isimlendirilir (Ruder, 2016).

g) Yerel Ağırlıklı Öğrenme (LWL - Locally Weighted Learning)

Yerel ağırlıklı öğrenme, sınıf tahmininde karar vermek için ağırlıklandırılmış en yakın komşulara benzer bir yerel ağırlıklı eğitim kullanır. Bu yöntem, yerel yakınlardaki bir veri grubu arasındaki mesafeleri ve olasılıkları dikkate alarak uygun bir çıktı tahmin eder (Fong vd. 2013).

h) Naive Bayes ile Karar Ağacı (DTNB – Decision Tree with Naive Bayes)

DTNB, karar ağaçları ve Naive Bayes algoritmasının birleştirilmesiyle oluşturulmuş melez bir algoritmadır. Aramanın her noktasında, algoritma, nitelikleri iki ayrı alt gruba ayırmanın yararını değerlendirir (bir grup Karar ağacı için, diğeri de Naive Bayes için). İleri seçim araması kullanılır. Her adımda seçilen özelliklerin modellenmesi için Naive Bayes, geri kalanlar

için karar ağaçlarının kullanıldığı ileri seçim araması kullanılır. Bütün özellikler ise karar tabloları tarafından modellenir (Mahajan ve Ganpati, 2014).

i) Yerine Koyarak Örnekleme (Bagging)

Bagging algoritması Breiman tarafından önerilen ve iyi bilinen topluluk algoritmalarından biridir (Ruder, 2016). Bagging, eğitim alt kümesini değiştirme ile ayarlanan tüm verilerden alır, çok sayıda temel öğrenici oluşturur ve son tahminleri yapmak için temel öğrenenlerin çıktılarını toplar (Hido vd., 1996).

j) Regresyon ile Sınıflandırma (CvR – Classification via Regression)

Regresyon, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi ampirik olarak belirlenmiş bir fonksiyonla değerlendirmek için kullanılan bir yöntemdir. Bu yöntem, geleneksel karar ağacını düğümlerde doğrusal regresyon olasılığı ile birleştiren M5P baz sınıflandırıcısı kullanır (Quinlan, 1992).

Deney Sonuçları

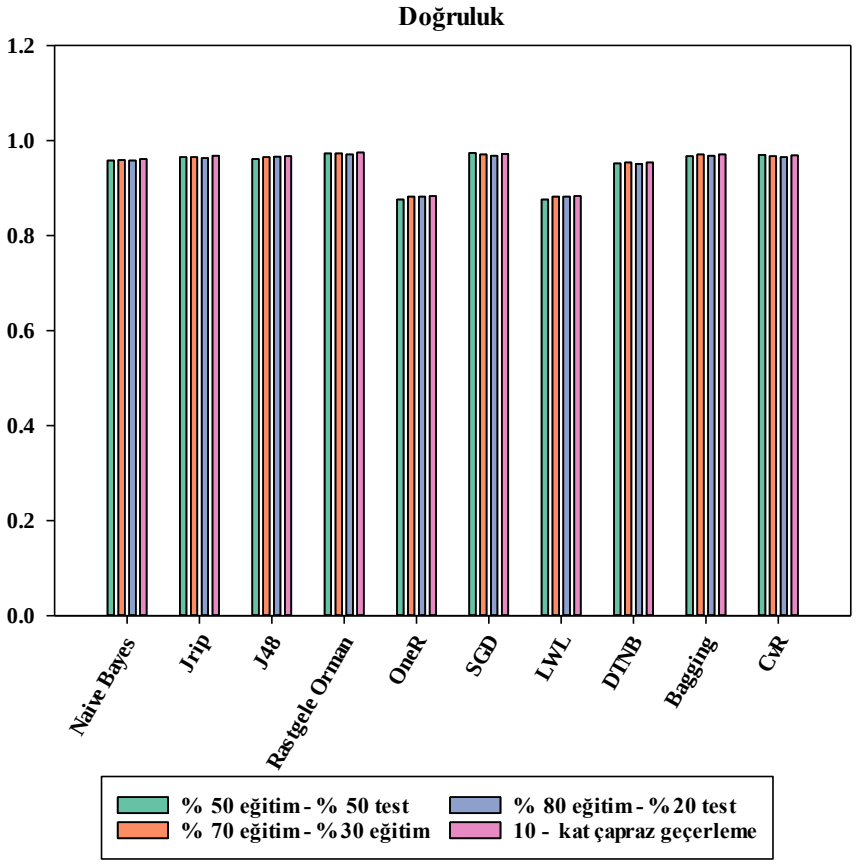
Önerilen yöntemi test etmek için ISOT sahte haber veri kümesi kullanılmıştır (Ahmed vd., 2017). Veri kümesinde sahte ve gerçek haber makalelerinden oluşmuştur. Doğru haberler Reuters.com haber sayfasından, sahte haberler ise farklı haber kaynaklarından derlenerek elde edilmiştir. ISOT veri kümesinde 21,417 gerçek etiketli 23,481 sahte etiketli haber makalesi yer almaktadır. Veri kümesinde farklı konularda haber makaleleri bulunmaktadır, özellikle politik ve dünya haberleri yer almaktadır.

Veri kümesi üzerinde, Naive Bayes, Jrip, J48, Random Forest, OneR, SGD, LWL, DTNB, Bagging ve CvR olmak üzere 10 farklı yapay zekâ algoritması ile sahte haberleri tespit etmek için deneyler yapılmıştır. Yapay zekâ algoritmaları, veri üzerinde test edilirken 4 farklı eğitim-test bölümlemesi yapılmıştır: % 50 eğitim - % 50 test, % 70 eğitim - % 30 test, % 80 eğitim - % 20 test ve 10 – kat çapraz geçerleme olarak belirlenmiştir. Belirlenen bölümlemeler ile ISOT sahte haber kümesi üzerinde önerilen yöntemin performansları doğruluk, hassasiyet ve duyarlılık istatistiksel değerlendirme ölçütlerine göre karşılaştırılmıştır.

ISOT veri kümesi sahte ve gerçek haberleri tahmin etmek için kullanılan 10 yapay zekâ algoritması ile elde edilen doğruluk değerleri Tablo 3'te verilmiştir. Elde edilen sonuçlara göre % 50 eğitim - % 50 test deney grubunda SGD 0.974 değeri ile en yüksek doğruluk değerini vermiştir. Rastgele Orman algoritması ise % 70 eğitim - % 30 test,% 80 eğitim - % 20 test ve 10 - kat çapraz geçerleme deney gruplarında en yüksek değeri vermiştir. 4 farklı deney için 10 farklı algoritma ile elde edilen doğruluk değerlerinin grafik ile gösterimi Şekil 3'te verilmiştir.

Tablo 3. Yapay zekâ algoritmaları ile elde edilen doğruluk değerleri

Yapay Zekâ Algoritmaları	Test Kriterleri			
	% 50 eğitim - % 50 test	**% 70 eğitim - % 30 test**	**% 80 eğitim - % 20 test**	**10 - kat çapraz geçerleme**
Naive Bayes	0.958	0.959	0.958	0.961
Jrip	0.965	0.965	0.963	0.968
J48	0.961	0.965	0.966	0.967
Rastgele Orman	0.973	**0.973**	**0.971**	**0.975**
OneR	0.876	0.882	0.882	0.883
SGD	**0.974**	0.971	0.968	0.972
LWL	0.876	0.882	0.882	0.883
DTNB	0.952	0.954	0.951	0.954
Bagging	0.967	0.971	0.968	0.971
CvR	0.970	0.967	0.965	0.969

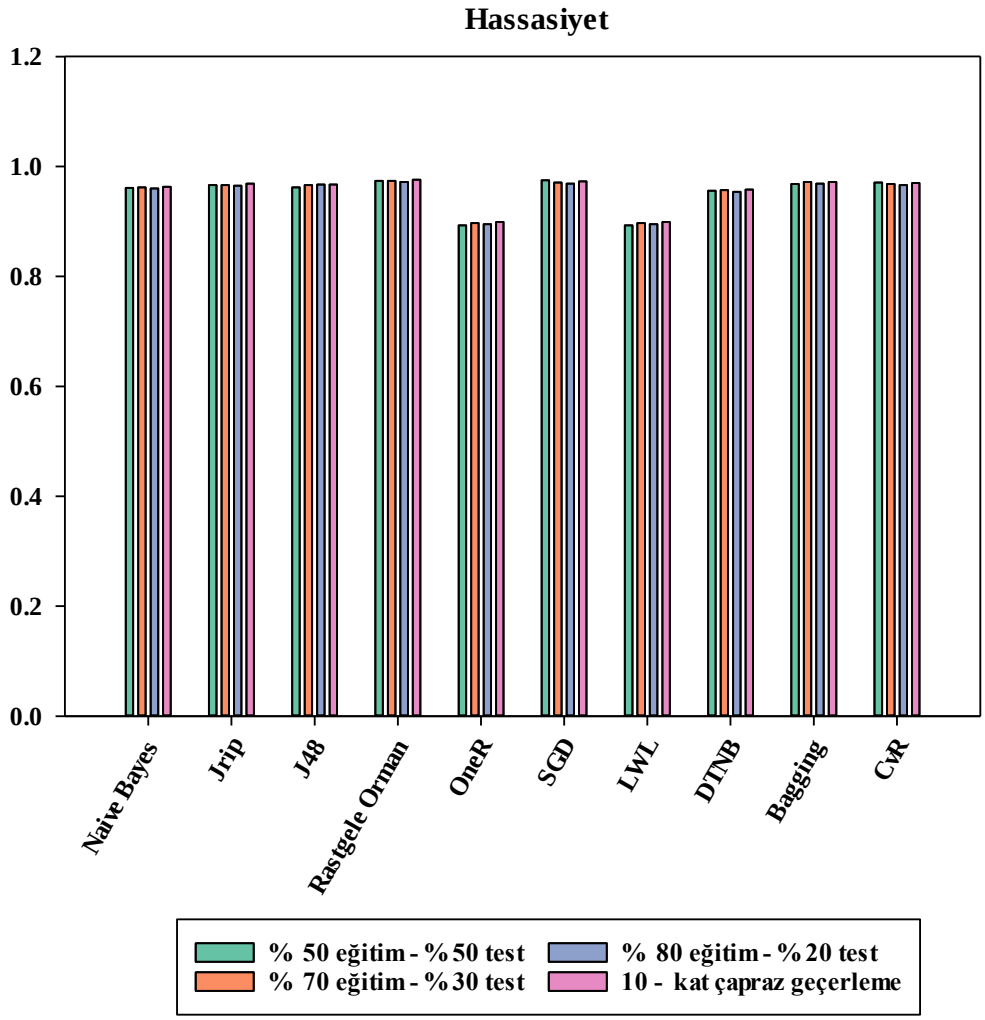


Şekil 3. Yapay zekâ algoritmalarının doğruluk değerleri

ISOT sahte haber veri kümesi üzerinde 10 farklı denetimli yapay zekâ algoritması ile elde edilen hassasiyet değerleri Tablo 4'te verilmiştir. Elde edilen değerlere göre, 0.975 değeri ile SGD algoritması % 50 eğitim - % 50 test deneyinde en yüksek hassasiyete sahiptir. Diğer 3 deney grubu için Rastgele Orman algoritması en iyi değerleri vermiştir. Elde edilen hassasiyet değerleri Şekil 4'te grafiksel olarak gösterilmiştir.

Tablo 4. Yapay zekâ algoritmaları ile elde edilen hassasiyet değerleri

Yapay Zekâ Algoritmaları	**Test Kritleri**			
	% 50 eğitim - % 50 test	**% 70 eğitim - % 30 test**	**% 80 eğitim - % 20 test**	**10 - kat çapraz geçerleme**
Naive Bayes	0.961	0.962	0.960	0.963
Jrip	0.966	0.966	0.965	0.969
J48	0.962	0.966	0.967	0.967
Rastgele Orman	0.974	**0.974**	**0.972**	**0.976**
OneR	0.893	0.897	0.895	0.899
SGD	**0.975**	0.971	0.969	0.973
LWL	0.893	0.897	0.895	0.899
DTNB	0.956	0.957	0.954	0.958
Bagging	0.968	0.972	0.969	0.972
CvR	0.971	0.968	0.966	0.970

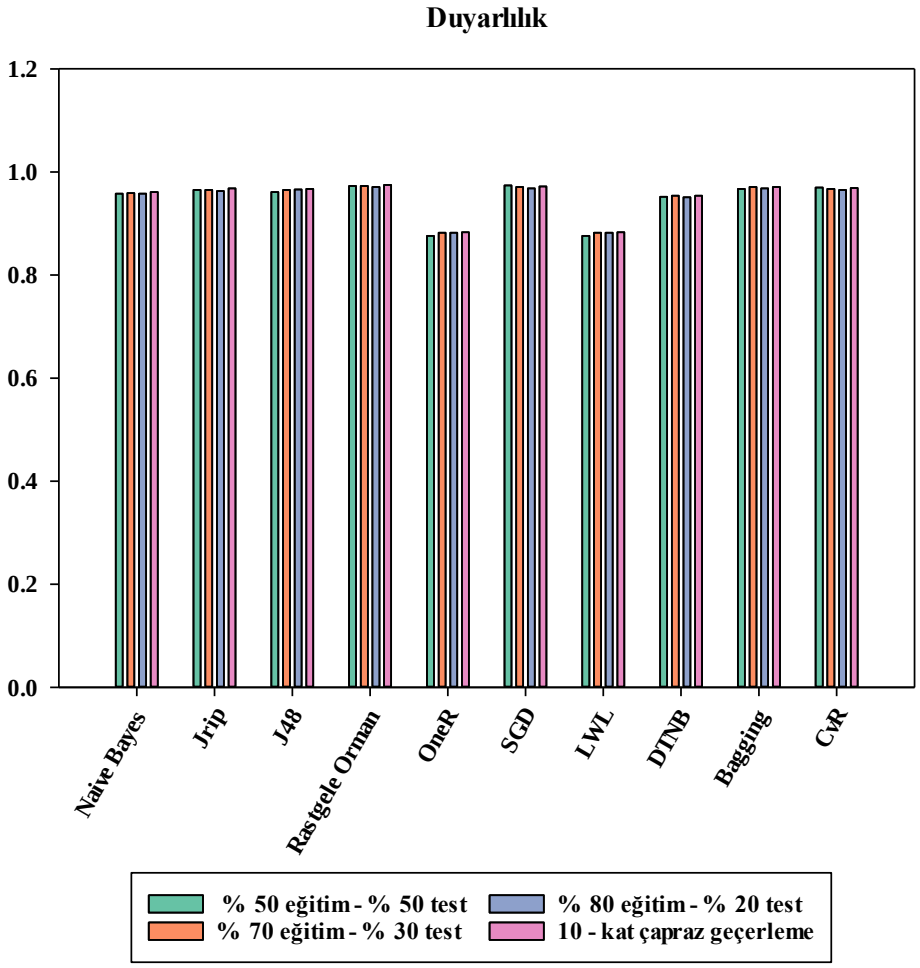


Şekil 4. Yapay zekâ algoritmalarının hassasiyet değerleri

Tablo 5. Yapay zekâ algoritmaları ile elde edilen duyarlılık değerleri

Yapay Zekâ Algoritmaları	Test Kritleri			
	% 50 eğitim - % 50 test	% 70 eğitim - % 30 test	% 80 eğitim - % 20 test	10 - kat çapraz geçerleme
Naive Bayes	0.959	0.960	0.959	0.962
Jrip	0.966	0.966	0.964	0.968
J48	0.962	0.966	0.966	0.967
Rastgele Orman	0.973	**0.974**	**0.971**	**0.976**
OneR	0.877	0.883	0.883	0.884
SGD	**0.975**	0.971	0.969	0.973
LWL	0.877	0.883	0.883	0.884
DTNB	0.953	0.954	0.951	0.955
Bagging	0.968	0.972	0.969	0.971
CvR	0.971	0.968	0.966	0.970

Kullanılan yapay zekâ algoritmaları ile 4 farklı deney grubu üzerinde elde edilen duyarlılık değerleri Tablo 5'de verilmiştir. Bu değerlere göre SGD algoritması 0.975 değeri ile % 50 eğitim - % 50 test deney grubunda en iyi değeri vermiştir. Diğer performans ölçütlerinde olduğu gibi bu ölçütte de Rastgele Orman algoritması kalan 3 deney grubu için en yüksek değeri vermiştir. Bu değerler, Şekil 5'de grafiksel olarak verilmiştir.



Şekil 5. Yapay zekâ algoritmalarının duyarlılık değerleri

Sonuçlar

Sosyal medya sitelerinin giderek artan popülerliği ile kullanıcılar tarafından oluşturulan mesajlar hızla geniş bir kitleye ulaşabilir. Böylece, sosyal medya sahte haberlerin yayılması için ideal bir ortam haline geldi. Sahte haberlerin hızlı yayılımı ve geniş kitleler tarafından erişilebilir olması toplumsal ve ekonomik olarak pek çok zarara neden olabilir ve bununla birlikte siyasi olayların sonuçlarını da manipüle edebilir. Bu nedenle, sosyal medyadaki sahte haberlerin tespit edilmesi önemli bir araştırma konusu haline gelmiştir.

Bu çalışmada çevrimiçi sahte haber tespit problemi sınıflandırma problemi olarak ele alınmıştır ve Naive Bayes, JRip, J48, Rastgele Orman, Stokastik Gradyan İniş, Yerel Ağırlıklı Öğrenme, Naive Bayes ile Karar Ağacı, Yerine Koyarak Öğrenme, Regresyon ile Sınıflandırma denetimli yapay zekâ algoritmalarının toplu olarak gerçek veri setindeki başarısı ilk kez incelenmiştir. Elde edilen sonuçlara göre, Stokastik Gradyan İniş algoritması % 50 eğitim - % 50 test deney grubunda doğruluk, hassasiyet ve duyarlılık performans değerlendirme ölçütlerine göre sırasıyla 0.974, 0.975 ve 0.975 değerleri ile en iyi sonucu veren algoritmadır. % 70 eğitim - % 30 test, % 80 eğitim - % 20 test ve 10 - kat çapraz geçerleme deney gruplarında doğruluk, hassasiyet ve duyarlılık ölçütlerine göre Rastgele Orman algoritması ISOT veri kümesi üzerinde en iyi performansı göstermiştir.

İleride sahte haber tespit problemi için farklı metrikler cinsinden daha performanslı sistemlerin geliştirilmesi planlanmaktadır. Bu amaçla denetimli makine öğrenmesi yöntemlerinin optimizasyonu ve parametre analizlerinin gerçekleştirilmesi, bu problemin

bir optimizasyon problemi olarak modellenmesi amaçlanmaktadır.

Kaynaklar

Ahmed, H., Traore, I., Saad, S., (2017). Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments. 127-138.

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., Kochut, K., (2017). A brief survey of text mining: Classification, clustering and extraction techniques. arXiv preprint arXiv:1707.02919.

Breiman, L., (1996). Bagging Predictors. *Machine Learning*, 24(2), 123-140.

Corney, D., Albakour, D., Martinez, M., and S. Moussa, S., (2016). What do a million news articles look like?. First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval, Italy, 42–47.

Coşkun, C., Baykal, A., (2011). An Application for Comparison of Data Mining Classification Algorithms. *Akademik Bilişim*, 1-8.

Fong, S., Luo, Z., Yap, B. W., (2013). Incremental learning algorithms for fast classification in data stream. 2013 International Symposium on Computational and Business Intelligence, India, 186-190.

Frank, E., I. Witten, I., (1998). Generating Accurate Rule Sets Without Global Optimization. Fifteenth International Conference on Machine Learning, San Francisco.

Gilda, S., (2017). Evaluating Machine Learning Algorithms for Fake News Detection. 15th Student Conference on Research and Development, Malaysia, 110-115.

Granik, M., Mesyura, V., (2017). Fake news detection using naive Bayes classifier. First Ukraine Conference on Electrical and Computer Engineering, Ukranie, 900-903.

Guo, C., Cao, J., Zhang, X., Shu, K., Yu, M., Exploiting emotions for fake news detection on social media, arXiv preprint arXiv:1903.01728, 2019.

Hido, S., Kashima, H., Takahashi, Y., (2009). Roughly balanced bagging for imbalanced data. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 2, 412-426.

Ho, T. K., (1995). Random decision forests. 3rd International Conference on Document Analysis and Recognition, Canada, 1, 278-282.

Holte, R. C., (1993). Very simple classification rules perform well on most commonly used data sets, *Machine Learning*, 11, 63-90.

Khan, R., Hanbury, A., Stoettinger, J., (2010). Skin detection: A random forest approach. IEEE International Conference on Image Processing, China, pp. 4613-4616.

Kumar, L., Bhatia, P. K., (2013). Text Mining: concepts, process and applications. International Journal of Global Research in Computer Science, 4(3), 36-39.

Long, Y., Lu, Q., Xiang, R., Li, M., Huang, C. R., (2017). Fake news detection through multi-perspective speaker profiles. Eighth International Joint Conference on Natural Language Processing, Taiwan, 2, 252-256.

Mahajan, A., Ganpati, A., (2014). Performance evaluation of rule based classification algorithms. *International Journal of Advanced Research in Computer Engineering & Technology*, 3(10), 3546-3550.

Monti, F., Frasca, F., Eynard, D., Mannion, D., Bronstein, M. M., (2019). Fake news detection on social media using geometric deep learning, arXiv preprint arXiv:1902.06673, 2019.

Ozbay, F. A., Alatas, B., (2019). A Novel Approach for Detection of Fake News on Social Media Using Metaheuristic Optimization Algorithms. *Elektronika ir Elektrotechnika*, 25(4), 62-67.

Quinlan JR., (1992). Learning with continuous classes. 5th Australian Joint Conference on Artificial Intelligence, 92, 343–348.

Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., Choi, Y., (2017). Truth of varying shades: Analyzing language in fake news and political fact-checking. 2017 Conference on Empirical Methods in Natural Language Processing, Denmark, 2931-2937.

Rubin, V., Conroy, N., Chen, Y., & Cornwell, S., (2016). Fake news or truth? using satirical cues to detect potentially misleading news. Second Workshop on Computational Approaches to Deception Detection, California, 7-17.

Ruchansky, N., Seo, S., Liu, Y., (2017). Csi: A hybrid deep model for fake news detection. 2017 ACM on Conference on Information and Knowledge Management, Singapore, 797-806.

Ruder, S., (2016). An overview of gradient descent optimization algorithms, arXiv preprint arXiv:1609.04747, 2016.

Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H., (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19, 22-36.

Sokolova, M., Japkowicz, N., Szpakowicz, S., (2006). Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. Australasian joint conference on artificial intelligence, Australia, 1015-1021.

Tschiatschek, S., Singla, A., Gomez Rodriguez, M., Merchant, A., Krause, A., (2018). Fake news detection in social networks via crowd signals. In Companion of the The Web Conference 2018 on The Web Conference 2018, 517-524.

Vaishali, V., Bhalodiya, N., N.N Jani, N. N., (2014). Applying Naïve bayes, BayesNet, PART, JRip and OneR Algorithms on Hypothyroid Database for Comparative Analysis, International Journal of Darshan Institute on Engineering Research & Emerging Technologies, 3(1).

Vedova, M. D., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M., de Alfaro, L., (2018). Automatic Online Fake News Detection Combining Content and Social Signals. 22st Conference of Open Innovations Association, 272-279.

Zhang, J., Cui, L., Fu, Y., Gouza, F. B., (2018). Fake news detection with deep diffusive network model. arXiv preprint arXiv:1805.08751.

Fake News Detection in Online Social Media

Extended abstract

In recent years, consuming news from social media has become increasingly popular due to its rapid dissemination of information, cheap cost, and easy access. The fact that social media becomes one of the main sources of information for people in the world has positive and negative effects on society, culture and business world. The quality of news on social media is lower than traditional news sources and social media are very suitable for spreading fake news. Owing to the detrimental effects of fake news on people and society, the detection of fake news is attracting attention. Online social media are attracted worldwide attention and increased its popularity by day by. The digital information age offers content creators the opportunity to publish which is known as "fake news" that is deliberately designed to mislead the reader. Major online social networking sites like Facebook, Twitter and Weibo make it easy to spread fake news to crowded readers. People can be affected negatively by fake news and accept false ideas. In addition, the propagation of fake news can reduce the reliability of the right news. Due to the large enormous of textual data generated by people and machines around the world, text mining applications in different areas have grown significantly. Text mining is a combination of algorithms and methods designed to extract hidden information and explore interesting patterns from unstructured textual data for different objectives.

Supervised artificial intelligence algorithms use the idea of learning from samples that try to find the relationship between input attributes and target attributes. The relationship found is represented by a structure expressed as a model. These algorithms provide a robust and fast model. They require a train set of labeled data. Their working principle is to learn from a set of labeled samples in the training set. As a result, unlabeled samples in the test set can be predicted.

In this study, a two-step model is proposed for detecting fake news. In the first step, a number of pre-processing is applied to the data set containing fake news to convert unstructured data into structured data. In the next step, ten supervised artificial intelligence algorithms are implemented on a structured fake news dataset. The proposed model is examined with four different train-test partitions. This experimental evaluation is performed with the Naive Bayes, JRip, J48, Random Forest, Stochastic Gradient Descent, Locally Weighted Learning, Decision Tree with Naive Bayes, Bagging, Classification via Regression supervised artificial intelligence algorithms on an existing public dataset and these algorithms are compared depending on accuracy, precision, and recall evaluation metrics.

Keywords: *Artificial intelligence algorithms, Fake news detection, Online social network*