Research Article / Araştırma Makalesi GU J Sci, Part A, 9(3): 323-333 (2022) https://doi.org/10.54287/gujsa.1170640

Gazi University
Journal of Science
PART A: ENGINEERING AND INNOVATION
http://dergipark.org.tr/gujsa

Detecting Turkish Fake News Via Text Mining to Protect Brand Integrity

Ozge DOGUC[1*]

[1]Istanbul Medipol University, Kavacik Guney Campus, Beykoz Istanbul

Keywords
Text Mining
Fake News
Brand Management
Data Management

Abstract
Fake news has been in our lives as part of the media for years. With the recent spread of digital news platforms, it affects not only traditional media but also online media as well. Therefore, while companies seek to increase their own brand awareness, they should also protect their brands against fake news spread on social networks and traditional media. This study discusses a solution that accurately classifies the Turkish news published online as real and fake. For this purpose, a machine learning model is trained with tagged news. Initially, the headlines were analyzed within the scope of this study that are collected from Turkish online sources. As a next step, in addition to the headlines of these news, news contexts are also used in the analysis. Analysis are done with unigrams and bigrams. The results show 95% success for the headlines and 80% for the texts for correctly classifying the fake Turkish news articles. This is the first study in the literature that introduces an ML model that can accurately identify fake news in Turkish language.

Cite
Doguc, O. (2022). Detecting Turkish Fake News Via Text Mining to Protect Brand Integrity. *GU J Sci, Part A*, *9*(3), 323-333.

Author ID (ORCID Number)
O. Doguc, 0000-0002-5971-9218

Article Process	
Submission Date	03.09.2022
Revision Date	20.09.2022
Accepted Date	22.09.2022
Published Date	30.09.2022

1. INTRODUCTION

Today, Internet users refer to online content for the decisions they make in their daily lives, the shopping they do, and the school they choose. In other words, many Internet users treat online sources as their main source of information. E-commerce customers use reviews and news about the products before making purchase decisions, and for this purpose companies use online content to impress customers (Ahmed et al., 2017). False or misleading news and comments are also used extensively to influence the decisions of internet users (Obadă, 2019). The biggest sources of spreading fake news or rumors are social media platforms such as Google, Facebook and Twitter (Becker, 2017). Given the lack of control people have over the spread of fake content along with misleading information, the issue of fake news has been around for years since the traditional media era (Lemann, 2016). Fake news has received more attention in the last few years, especially since the 2016 US elections.

In addition to advances in computer science, the dynamic nature of the web and social media, which can be updated with information from each user, helps people creating and spreading fake news. On the other hand, it is very difficult to identify institutions affected by fake news. According to the report of Trend Micro, a cyber security company, companies can easily purchase services from many fake news producers around the world. According to the same report, political parties also use these services to rig the election results by manipulating voters' views on certain issues (Levin, 2017). There are studies showing that detecting fake news is difficult and requires more sophisticated decision support systems than detecting fake products (Chu et al., 2021).

*Corresponding Author, e-mail: odoguc@medipol.edu.tr

Fake news targeting brands can negatively affect consumers' perceptions of them. It can even change consumers' behavior towards brands and cause them to leave bad reviews on the brand line. As the growing middle class in Turkey in recent years often conspicuously consumes premium brands as a form of self-identity and associates them with the class, the fake news about premium brands has begun to affect a wider audience (Bankole & Reyneke, 2020). Fake news can affect the Turkish named brands as well as the global ones (Mertoglu, 2020).

This paper presents an approach to detect Turkish fake news about e-commerce websites, using text analysis based on n-gram features and machine learning classification techniques. In this paper, using the Python programming language and machine learning tools in the literature, the data were created from Turkish news headlines and texts. This study introduces a trained ML model by using real and fake Turkish news. This is the first study in the literature that shows that Turkish fake news can be effectively and efficiently detected through analysis of headlines and body text.

1.1 Literature Survey

Machine learning, under the concept of AI, has emerged as the preferred method for applications such as computer vision, speech, recognition, natural language processing to process information quickly and learn correctly (Murphy, 2012). AI developers recognize that training can be done much easier by showing examples of desired input-output behavior, rather than programming (if-else) by predicting the expected response for all possible inputs. The impact of machine learning is evident in many data-intensive industrial areas such as computer science, consumer services, diagnostics of complex systems, and control of logistics chains (Mahoney, 2011). This section discusses studies that are in the literature that use ML techniques for detecting fake news.

1.1.1 Fake News

Conroy et al. (2015) lists the types of fake news as follows:

1. Image-based: Fake news posts use more graphics as content.

2. User-based: Most fake news is produced by fake accounts and is aimed at an audience that represents certain age groups, gender, and political affiliation.

3. Knowledge-based: Some fake news produces pseudo-scientific solutions to insoluble or difficult problems and make users believe it is real.

4. Genre-based: Mostly written by pseudo-authors who imitate and copy the style of some journalists and brands.

5. Postural-based: Presented in a way that changes the meaning and purpose of correct statements.

In the literature, many automatic detection techniques of fake news and spoofing posts have been discussed (Altunbey Özbay & Alataş, 2020; Toğaçar et al., 2021). These studies use semantic and frequency analysis of the Turkish language to detect fake news patterns. Fake news detection has many facets, ranging from the use of chatbots to spread false information to the use of clickbait to spread rumors. Many social media networks, including Facebook, have many click tricks that increase sharing and likes for posts and then spread fake information (Chen & Cheng, 2019). Significant amount of work has been done to date to detect fake information (Stahl, 2018). The following detection techniques were introduced by Parikh and Atrey (2018):

1. Modeling of deception

2. Clustering

3. Predictive modeling

4. Methods based on content markup

5. Methods not based on text sign

While certain level of success has been achieved in detecting fake news and posts using Machine Learning (ML) techniques, the ever-changing characteristics of fake news on social media networks pose challenges in accurate classification.

1.1.2 Detecting Fake News with Machine Learning

In recent years, detecting fraudulent online reviews and fake news has played an important role in business, and politics because of the potential impact of fake reviews on consumer behavior and purchasing decisions. Researchers used machine learning techniques with large datasets to enhance learning in intelligent mechanisms that can detect fake news so they can distinguish between words syntactically and semantically. It includes tasks such as detecting fake news, reviewing tags, keywords, text content and media used (image, video, etc.). There have been few studies in the literature that have done textual analysis of topics such as fake news, rumors or spam (Aytaç et al., 2020). There are two aspects to detecting fake news. First, it is based on linguistic and psychological reasons, in other words, features that can capture the meaning of the text. However, these features had trouble understanding text properly and did not perform well on long texts. The second approach, based on the neural network model, provides document-level representation to make sense of fake news text (Englmeier, 2021). In these studies, artificial neural network models were used to learn semantic representations in the field of Natural Language Processing (NLP) and achieved successful results (Spicer, 2018). In addition, Albahar (2021) and Zhao et al. (2020) use hybrid methods that combine text mining and statistical analysis. These studies examine propagation of the fake news over time and investigate how news articles are distributed over time through radial basis function (RBF). Similarly, de Souza et al. (2021) use iterative versions of Gaussian fields and harmonic functions for labeling the fake news. These studies are done for English and Chinese languages and focus on how fake news spread on Internet over time.

2. MATERIAL AND METHOD

In this paper, a machine learning model was created that can train the news and classify it as "real" or "not real" (fake) with the data obtained. Data was collected from online Turkish news sources. The sources are Turkish online media websites and articles that are published on social media. Web crawlers were used to find and collect article that had been published between January 2021 and December 2021. More than 1000 news articles, containing more than 50000 words were collected during this process. News articles about popular brand names (e.g. Nike, Prada, Gucci, Beymen, Apple, Vakko, etc.) were taken into consideration, and they were preprocessed to be used in ML algorithms. Table 1 shows an example fake news article collected through Turkish news sources. The words like *sızdırıldı (leaked), söylentiler (rumors)* are big hints about authenticity of the article.

Table 1. *Example Turkish fake news article*

	Turkish Original	**English Translation**
Title	***Apple iPhone 13 kamera özellikleri sızdırıldı***	***Apple iPhone 13 camera features leaked***
Text	*Sony, Apple iPhone ailesi için kamera sensörü tedarik etmeye devam edecek. Şirketin lens konusunda da Genius ve Largan ile çalışacağı belirtiliyor. Bu yılın başlarında yayımlanan raporlarda, iPhone 13 özellikleriyle ilgili bazı iddialar yer alıyordu. Analist Ming-Chi Kuo, iPhone 13 kamera özellikleri ile ilgili söylentileri doğruladı.*	*Sony will continue to supply camera sensors for the Apple iPhone family. It is stated that the company will also work with Genius and Largan on lenses. In the reports released earlier this year, there were some claims about iPhone 13 features. Analyst Ming-Chi Kuo has confirmed rumors about iPhone 13 camera features.*

An ML model was developed that can decide whether the news is real or fake by using the transformed data and various functions written, using Naïve Bayes statistical calculation. Collected data was split into two sets; majority of the data was used to train the model and the second set was used to validate the classification results.

In the light of previous academic studies on the subject, (Doguc et al., 2020) AI models for pattern-detection and sorting were designed for this research. Secondly, for efficiency the code needed for operations was written in Python language. After this stage, suitable news titles were searched for the dataset of the study. Data analyzed in the application are the headlines of the news published between 2020 and 2022. In addition, the texts of the news were also analyzed.

Naïve Bayes probability calculation was used for text data among artificial intelligence models. In Naive Bayes, the data was kept in separate lists and dictionaries according to whether they were real or fake, and how much of all the words passed as real and fake was analyzed.

2.1 Naïve Bayes

Naive Bayes (NB) is one of the best-known data mining algorithms for classification (Wu et al., 2008). It is a supervised learning algorithm, based on the Bayes' theorem. It is called as a 'Naive' classifier, as it assumes that all features in a class are independent (unrelated) of each other. For example, given that a fruit is apple, its three properties (being round, red and small) can be considered as dependent on each other, they are actually independent features of the fruit. In practice, most combinations of attribute values are either not available in the training data or are not available in sufficient numbers. NB circumvents this impasse by assuming conditional independence. Despite this independence assumption, Bayes is a competent classifier in many real-world applications (Zhang et al., 2016).

2.2 N-Gram

N-Gram is a n-character fragment taken from a string (Violos et al., 2018). N-Gram is used in the model building process by dividing a sentence into parts of words. In N-Gram, 'N' indicates the number of words to be grouped in a single section. There are three types of N-Gram (Drus & Khalid, 2019):

- Unigram: A token consisting of only one word.
- Bigram: A token consisting of two words.
- Trigram: A token consisting of three words.

2.3 Bag of Words

The bag of words, or BoW for short, is a way to extract features from text for use in modeling machine learning algorithms. A BoW is a representation of text that models the location, relationships, and interactions of words in a document. The reason why the model is called the word bag is that it provides summary information about the order or structure of the words in the document and can be used easily when necessary. The model checks if known words are found in the document. In this approach, the histogram of the words in the text is examined and each word is counted as a feature (Goldberg, 2017).

2.4 Text Classification

The text classification problem has been extensively studied in practice for the last decade. Recent developments in NLP and text mining techniques allowed researchers work in applications that use text classification methods efficiently. Most text classification and document classification systems are divided into four stages: feature extraction, size reduction, classifier selection, and evaluation (Jiang et al., 2018).

Feature extraction and preprocessing are crucial steps for text classification applications. Most text and document datasets have problems with words to block, misspellings, slang, etc. Contains many unnecessary words. Some text cleaning techniques are tokenization, stop words, capitalization, slang and abbreviation, noise removal, spelling correction, stemming and lemmatization.

3. RESULTS AND DISCUSSION

In this paper, detection of fake news is analyzed in 3 stages. In the first stage, it is discussed if a news article can be correctly identified as fake just by checking its headline, as headlines in the news usually contain the important summary. In addition, the words that most affect whether news is real, or fake is determined. Figure 1 shows the flowchart of the fake news analysis method.

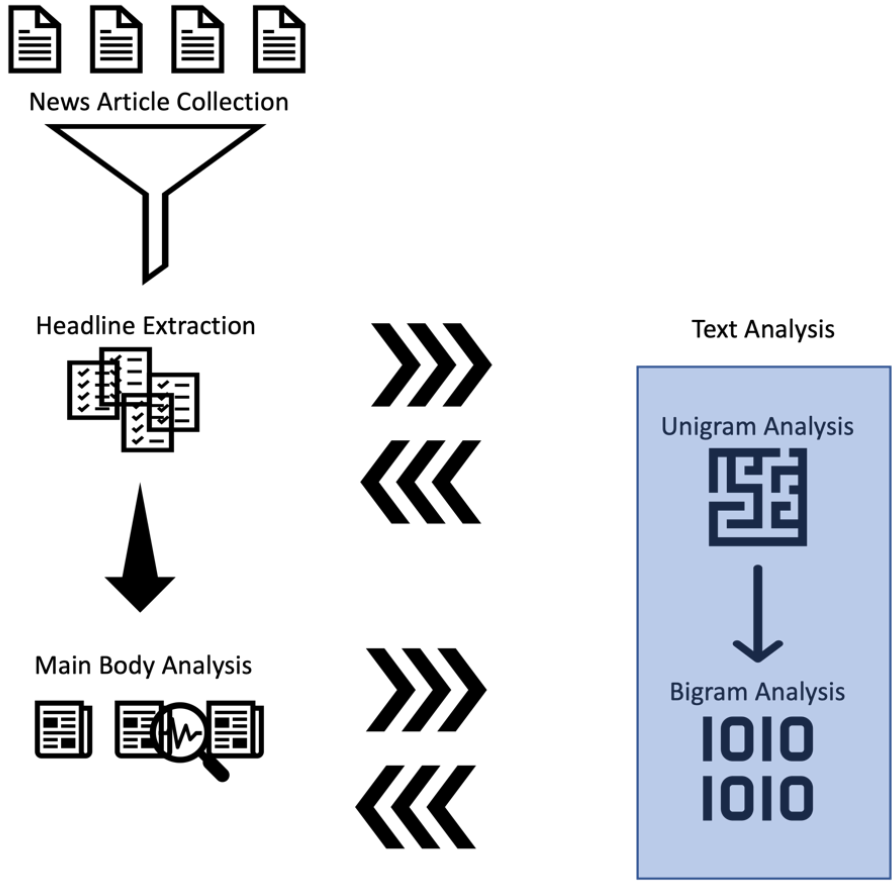

Figure 1. *Flowchart of the fake news detection method*

As shown in Figure 1, news articles are collected from Turkish online media, including articles from newspapers and tweets. Next, headlines are extracted from the articles, to be analyzed separately. The text analysis engine performs two-step analysis of the given text: unigram and bigram analysis. It determines if text is from a fake news or not, and a confidence score for the evaluation.

After eliminating the empty news headers, the detection application uses a tokenizer. The tokenizer serves to divide the text into tokens or words, and stopwords like *şey*. The stopwords, which are used in Turkish but have no direct meaning are excluded from analysis to prevent them from affecting classification. Then, using the *count vectorizer*, the given text is converted into a vector containing the frequencies of the words. In this way, words are converted into integer values that can processed easily. Figure 2 shows the pseudocode.

```
dic = {}
y = 0
for x in vectorizer3.get_feature_names():
    dic[x] = y
    y = y + 1
```

Figure 2. *Creating dictionary of words and their frequencies*

In order to use the Naïve Bayes probabilities using the formula in Figure 3, the news articles are labeled as 'Positive' and 'Negative'. The application counts the words in each article and prepares a learning bed based on the number of words and article's sentiment. Next, the same is done with the bigrams, as shown in Figure 4. The purpose of doing this is to calculate probabilities among their own classes according to the Naive Bayes formula. All labels and probabilities are stored in dictionaries to improve efficient storage and fast lookup.

Likelihood

Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

Posterior Probability

Predictor Prior Probability

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Figure 3. *Naive Bayes Probability Calculation Formula*

```
bigram_vectorizer3 = CountVectorizer(ngram_range=(2, 2),token_pattern=r'\b\w+\b', min_df=1)
bXY = bigram_vectorizer3.fit_transform(fixedTarget)
buniqCount = len(bigram_vectorizer3.get_feature_names())
bidic = {}
y = 0
for x in bigram_vectorizer3.get_feature_names():
    bidic[x] = y
    y += 1
biallPos = 0
biallNeg = 0
for x,y in zip(bXY,slabel):
    if(y == 0):
        biallPos += np.sum(x)
    else:
        biallNeg += np.sum(x)
```

Figure 4. *Using Count Vectorizer for Double Words and Finding the Number of Positive and Negative Data*

Next, the news data is split into 2 sets: training and test. During the training phase, the application calculates scoring for each news article; based on the formula below:

$$\frac{\text{number of repetitions of the word (negative, positive)} + 1}{\text{total number of words in positive or negative case} + \text{number of unique words in the file}}$$

The code snippet for the training algorithm is given in Figure 5.

```
testTitle = target
testLabel = label

posnoncount = 0
negnoncount = 0

unicount = 0
bicount = 0
stringcount = 0
for i,a in zip(testTitle,testLabel):
    if(type(i) != str):
        continue
    stringcount += 1
    i = re.sub(r'[^\w\s]','',i)
    words = i.split()
    for ax in range(len(words)):
        words[ax] = words[ax].lower()
    x = positiveNaive(XY, words, uniqCount, vectorizer3, allPos, label, dic, posDic)
    x = x + (np.log(posCount/(posCount + negCount)))
    y = negativeNaive(XY, words, uniqCount, vectorizer3, allNeg, label, dic, negDic)
    y = y + (np.log(negCount/(posCount + negCount)))

    uniflag = 0
    if(x > y):
        uniflag = 0
    else:
        uniflag = 1
```

Figure 5. *Code snippet for the training algorithm*

The generated ML model has a success rate of 95% for single words and 99% for bigrams. The reason for high the success rates is that presence of some words or word groups in the headlines provide too much information about that classification of that news article. List in Figure 6 contains information about frequencies of the words which words have more impact on the classification results.

```
Real News Unigram
Istanbul     |     moda          |     yaz
5362         |     5083          |     4936

Fake News Unigram
dolar        |     Rusya         |     Ukrayna
3884         |     3368          |     3349

Real News Bigram
Yaz modası   |     babalar günü  |     indirim kuponu
4030         |     3589          |     3309

Fake News Bigram
Rusya Ukrayna      |     dolar kuru    |     hediye garantili
3608               |     2915          |     2318
```

Figure 6. *Most frequent words without stopwords*

Looking at the list above, it is seen that the content of some news, both real and fake, can be guessed from the title. In the second stage, the same methodology is used with news text context, not just the headlines. The functions mentioned above are reused for similar operations. In addition, the BoW method is used. Likewise, separate analyzes are made for unigrams and bigrams. Dictionaries are prepared for single words, the data is separated as positive / negative, and the calculation of the total number of words is carried out. Same is done for bigrams as well. The accuracy results are given below:

```
Accuracy with Unigrams: 81.44%

Accuracy with Bigrams: 62.63%
```

In news headlines, the result obtained as a result of the operation performed for bigrams was higher, while a higher result was obtained in the operation performed with single words using the entire news text. The reason for this may be that better words should be chosen in news headlines in order to express more with fewer words, and the words in succession should be in harmony.

Finally, this paper examines the impact of the strong words on the classification of fake news. For that purpose, 10 words whose presence strongly indicates that the news is real, 10 words whose absence strongly indicates that the news is real, 10 words whose presence strongly indicates that the news is fake, and 10 words whose absence strongly indicates that the news is fake were picked. In all cases, effect of the stopwords is also tested.

The data for 10 words whose presence strongly indicates that the news is real are as follows (Table 2):

Table 2. *10 Words that appear in the real news articles*

Term	***Weight***
Ve	0.303
Bir	0.142
Şey	0.130
Hangi	0.118
Ne	0.106
İle	0.098
Bu	0.095
Kadar	0.091
De/Da	0.088
Hep	0.078

The list consists of mostly the stopwords, Table 3 shows the results without the stopwords:

Table 3. *10 Words that appear in the real news articles (without stopwords)*

Term	***Weight***
İstanbul	0.102
Moda	0.091
Hafta	0.085
Indirim	0.083
Baba	0.081
Yaz	0.081
Sipariş	0.069
Sezon	0.069
Garanti	0.069
Hediye	0.067

Next, this paper analyzes if absence of words can be used to decide if a news article is real or fake. Table 4 shows the top the 10 words whose absence strongly separates a real news article from a fake one (appear in the fake news but not in the real ones):

Table 4. *10 words the absence of which strongly indicates the truth of the news*

Term	***Weight***
Iddia	0.037
Söylendi	0.037
Çarpıcı	0.036
Edildi	0.035
Bedava	0.030
Anında	0.028
iPhone	0.028
Hediye	0.028
Kazandınız	0.025
Hemen	0.025

In other words, real news articles usually do not include words such as *iddia* and *söylendi*, while fake ones do. Similarly, Table 5 includes the top 10 words that appear in the fake news most frequently.

Table 5. *10 words whose existence strongly indicates that the news is fake*

Term	***Weight***
Dolar	0.072
Rusya	0.061
Ukrayna	0.055
İndirim	0.023
Faiz	0.021
Elektronik	0.021
Yaz	0.019
Telefon	0.019
Kupon	0.019
Ürün	0.017

Trained with hundreds of news articles and their analysis results, this study creates an ML model for Turkish that can accurately detect fake news. Unlike the studies that are available in the literature (Obadă, 2019; Yalcin & Simsek, 2020; Zhao et al., 2020; Albahar, 2021; de Souza et al., 2021), this study doesn't rely on statistical analysis or temporal distribution of news articles. It combines NLP and ML techniques to analyze Turkish news articles and their headlines; and devise rules that can distinguish fake news through syntactic and semantic features.

4. CONCLUSION

This paper focused on how the news can be accurately classified as *real* or *fake* by using the news articles available in the online media. Fake news impact brand integrity and equity greatly (Chen & Cheng, 2019), and more than 80% of the online users come across with fake news every week (Belin, 2020). Turkish brands are not immune to the fake news, and this paper uses Turkish news articles to create an intelligent model that can successfully detect fake news.

Data used in the research are the news (texts) published in the past, these texts were first cleansed and preprocessed to be used in the AI models. Among the AI models, Naïve Bayes probability calculation was used for text data. While there are studies in the literature that achieve 91% success with natural language processing (NLP) (Toğaçar et al., 2021), in the established methodology, a success rate of over 95% was achieved for headlines and around 80% for texts. As a result, it is shown that the news titles provide more important information that the article itself in the form of a summary. In addition to these, how the absence of words affects the reality of the news were examined. Naive Bayes gives more weight to the more frequent words and a classification can be done based on the frequencies. This is the first study that provides machine learning model that can accurately identify Turkish fake news.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Ahmed, H., Traore, I., & Saad, S. (2017, October 26-28). *Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques*. In: I. Traore, I. Woungang & A. Awad (Eds.), Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environment, First International Conference, ISDDC 2017, Vancouver, BC, Canada, (pp. 127–138). doi:10.1007/978-3-319-69155-8_9

Albahar, M. (2021). A hybrid model for fake news detection: Leveraging news content and user comments in fake news. *IET Information Security*, *15*(2), 169–177. doi:10.1049/ise2.12021

Altunbey Özbay, F., & Alataş, B. (2020). Çevrimiçi sosyal medyada sahte haber tespiti. *DÜMF Mühendislik Dergisi*, *11*(1), 91–103. doi:10.24012/dumf.629368

Aytaç, Ö. B., Silahtaroğlu, G., & Doğuç, Ö. (2020). Analysis of Digital Marketing Strategies of Deposit Banks in Turkey via Text Mining Twitter Posts. In: H. Dincer & S. Yüksel (Eds.) *Strategic Outlook for Innovative Work Behaviours* (pp. 361–376). Springer. doi:10.1007/978-3-030-50131-0_20

Bankole, O., & Reyneke, M. (2020). The Effect of Fake News on the Relationship between Brand Equity and Consumer Responses to Premium Brands: An Abstract. In: S. Wu, F. Pantoja & N. Krey (EdS.), *Marketing Opportunities and Challenges in a Changing Global Marketplace* (pp. 461–462). Springer International Publishing. doi:10.1007/978-3-030-39165-2_189

Becker, R. (2017, June 26). Your short attention span could help fake news spread. URL

Belin, A. (2020, June 25). How to Protect and Defend your Brand from Fake News. URL

Chen, Z. F., & Cheng, Y. (2019). Consumer response to fake news about brands on social media: the effects of self-efficacy, media trust, and persuasion knowledge on brand trust. *Journal of Product & Brand Management*, *29*(2), 188–198. doi:10.1108/JPBM-12-2018-2145

Chu, S. K. W., Xie, R., & Wang, Y. (2021). Cross-Language Fake News Detection. *Data and Information Management*, *5*(1), 100–109. doi:10.2478/dim-2020-0025

Conroy, N. K., Rubin, V. L., & Chen, Y. (2015, November 6-10). Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology (ASIS&T)*, *52*(1), 1-4. doi:10.1002/pra2.2015.145052010082

de Souza, M. C., Nogueira, B. M., Rossi, R. G., Marcacini, R. M., dos Santos, B. N., & Rezende, S. O. (2021). A network-based positive and unlabeled learning approach for fake news detection. *Machine Learning*. doi:10.1007/s10994-021-06111-6

Doguc, O., Aytac, O. B., & Silahtaroglu, G. (2020). Lemmatizer: Akıllı Türkçe kök bulma yöntemi. *Turkish Studies - Information Technologies and Applied Sciences*, *15*(3), 289-299. doi:10.47844/TurkishStudies.44220

Drus, Z., & Khalid, H. (2019). Sentiment Analysis in Social Media and Its Application: Systematic Literature Review. *Procedia Computer Science*, *161*, 707-714. doi:10.1016/j.procs.2019.11.174

Englmeier, K. (2021). The Role of Text Mining in Mitigating the Threats from Fake News and Misinformation in Times of Corona. *Procedia Computer Science*, *181*, 149-156. doi:10.1016/j.procs.2021.01.115

Goldberg, Y. (2017). *Neural Network Methods in Natural Language Processing*. Morgan & Claypool Publishers. doi:10.1007/978-3-031-02165-7

Jiang, M., Liang, Y., Feng, X., Fan, X., Pei, Z., Xue, Y., & Guan, R. (2018). Text classification based on deep belief network and softmax regression. *Neural Computing and Applications*, *29*(1), 61–70. doi:10.1007/s00521-016-2401-x

Lemann, N. (2016, November 30). Solving the Problem of Fake News. URL

Levin, S. (2017, June 13). Pay to sway: report reveals how easy it is to manipulate elections with fake news. URL

Mahoney, M. W. (2011). Randomized Algorithms for Matrices and Data. *Foundations and Trends in Machine Learning*, *3*(2), 123–224. doi:10.1561/2200000035

Mertoglu, U. (2020). A Fake News Detection Model for Turkish Language (*Türkçe için Sahte Haber Tespit Modelinin Oluşturulması*). PhD Thesis. Hacettepe University.

Murphy, K. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.

Obadă, D.-R. (2019). Sharing Fake News about Brands on Social Media: a New Conceptual Model Based on Flow Theory. *Argumentum. Journal of the Seminar of Discursive Logic, Argumentation Theory and Rhetoric*, *17*(2), 144-166.

Parikh, S. B., & Atrey, P. K. (2018, April 10-12). *Media-Rich Fake News Detection: A Survey*. In: 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), (pp. 436–441). doi:10.1109/MIPR.2018.00093

Spicer, R. N. (2018). Lies, Damn Lies, Alternative Facts, Fake News, Propaganda, Pinocchios, Pants on Fire, Disinformation, Misinformation, Post-Truth, Data, and Statistics. In: *Free Speech and False Speech*, (pp 1-31). Springer International Publishing. doi:10.1007/978-3-319-69820-5_1

Stahl, K. (2018). Fake news detection in social media. California State University Stanislaus, 6.

Toğaçar, M., Eşidir, K. A., & Ergen, B. (2021). Yapay Zekâ Tabanlı Doğal Dil İşleme Yaklaşımını Kullanarak İnternet Ortamında Yayınlanmış Sahte Haberlerin Tespiti. Journal of Intelligent Systems: Theory and Applications, *5*(1), 1–8. doi:10.38016/jista.950713

Violos, J., Tserpes, K., Varlamis, I., & Varvarigou, T. (2018). Text Classification Using the N-Gram Graph Representation Model Over High Frequency Data Streams. *Frontiers in Applied Mathematics and Statistics*, *4*, 41. doi:10.3389/fams.2018.00041

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z.-H., Steinbach, M., Hand, D. J., & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, *14*(1), 1–37. doi:10.1007/s10115-007-0114-2

Yalcin, F., & Simsek, Y. (2020). A New Class of Symmetric Beta Type Distributions Constructed by Means of Symmetric Bernstein Type Basis Functions. *Symmetry*, *12*(5), 779. doi:10.3390/sym12050779

Zhang, L., Jiang, L., Li, C., & Kong, G. (2016). Two feature weighting approaches for naive bayes text classifiers. *Knowledge-Based Systems*, *100*, 137–144.

Zhao, Z., Zhao, J., Sano, Y., Levy, O., Takayasu, H., Takayasu, M., Li, D., Wu, J., & Havlin, S. (2020). Fake news propagates differently from real news even at early stages of spreading. *EPJ Data Science*, *9*(1), 7. doi:10.1140/epjds/s13688-020-00224-z