Int. J. Pure Appl. Sci. 8(1);31-42 (2022)

IJPAS
ijpas@munzur.edu.tr
ISSN: 2149-0910

Research article/Araştırma makalesi
DOI: 10.29132/ijpas.994840

Machine Learning Based Deception Detection System in Online Social Networks

Harun Bingol[1]*, Bilal Alatas[2]
[1]Malatya Turgut Ozal University, Faculty of Engineering and Natural Sciences, Departmant of Software, Elazig, Turkey
[2]Firat University, Faculty of Engineering, Departmant of Software, Elazig, Turkey
**harun_bingol@hotmail.com, balatas@firat.edu.tr*
Received date:13.09.2021, Accepted date:04.02.2022

Abstract

The rapid dissemination of Internet technologies makes it easier for people to live in terms of access to information. However, in addition to these positive aspects of the internet, negative effects cannot be ignored. The most important of these is to deceive people who have access to information whose reliability is controversial through social media. Deception, in general, aims to direct the thoughts of the people on a particular subject and create a social perception for a specific purpose. The detection of this phenomenon is becoming more and more important due to the enormous increase in the number of people using social networks. Although some researchers have recently proposed techniques for solving the problem of deception detection, there is a need to design and use high-performance systems in terms of different evaluation metrics. In this study, the problem of deception detection in online social networks is modeled as a classification problem and a methodology that detects misleading contents in social networks using text mining and machine learning algorithms is proposed. In this method, since the content is text-based, text mining processes are performed and unstructured data sets are converted to structured data sets. Then supervised machine learning algorithms are adapted and applied to the structured data sets. In this paper, real public data sets are used and Support Vector Machine, k-Nearest Neighbor (k-NN), Naive Bayes (NB), Random Forest, Decision Trees, Gradient Boosted Trees (GBT), and Logistic Regression algorithms are compared in terms of many different metrics. In Dataset 1, the highest average accuracy value was obtained with 74.4% GBT algorithm, while in Dataset 2, the highest average accuracy value was obtained from the NB algorithm with 71.2%.

Keywords: Classification, deception detection, machine learning, social networks

Çevrimiçi Sosyal Ağlarda Makine Öğrenmesi Tabanlı Aldatma Tespit Sistemi

Öz

İnternet teknolojilerinin hızla yaygınlaşması, insanların bilgiye erişim açısından yaşamlarını kolaylaştırmaktadır. Ancak internetin bu olumlu yönlerine ilaveten olumsuz etkileride göz ardı edilemez. Bunların en önemlisi ise sosyal medya üzerinden güvenilirliği tartışmalı olan bilgiye erişmek isteyen insanların aldatılmasıdır. Aldatma, genel olarak insanların belirli bir konuda düşüncelerini yönlendirmeyi ve belirli bir amaca yönelik toplumsal bir algı oluşturmayı amaçlar. Bu fenomenin tespiti, sosyal ağları kullanan insan sayısındaki muazzam artış nedeniyle giderek daha önemli hale geliyor. Bazı araştırmacılar son zamanlarda aldatma tespiti problemini çözmek için teknikler önermiş olsa da, farklı değerlendirme ölçütleri açısından yüksek performanslı sistemler tasarlamaya ve kullanmaya ihtiyaç vardır. Bu çalışmada, çevrimiçi sosyal ağlarda aldatma tespiti problemi bir sınıflandırma problemi olarak modellenmiş ve metin madenciliği ve makine öğrenmesi algoritmaları kullanılarak sosyal ağlardaki yanıltıcı içerikleri tespit eden bir metodoloji önerilmiştir. Bu yöntemde içerik metin tabanlı olduğu için metin madenciliği işlemleri yapılmakta ve yapılandırılmamış veri kümeleri yapılandırılmış veri kümelerine dönüştürülmektedir. Ardından denetimli makine öğrenmesi algoritmaları uyarlanmata ve yapılandırılmış veri kümelerine uygulanmaktadır. Bu çalışmada, gerçek halka açık veri setleri kullanılmış ve Destek Vektör Makinesi, k-Nearest Neighbor (k-NN), Naive Bayes (NB), Random Forest, Decision Trees, Gradient Boosted Trees (GBT) ve Logistic Regresyon algoritmaları birçok farklı metrik açısından karşılaştırılmıştır. Veri seti 1'de en yüksek ortalama doğruluk değerini %74.4 GBT algoritmasında elde edilirken, Veri seti 2'de en yüksek ortalama doğruluk değeri %71.2 ile NB algoritmasından elde edilmiştir.

IJPAS
ijpas@munzur.edu.tr
ISSN: 2149-0910

Research article/Araştırma makalesi
DOI: 10.29132/ijpas.994840

Anahtar Kelimeler: Sınıflandırma, aldatma tespiti, makine öğrenmesi, sosyal ağlar

INTRODUCTION

The development of Internet brought the concepts of social media and social networking. In social networks, people produce and share content that expresses their feelings and thoughts. For example, they share information about a hotel they have gone to, or they comment on a restaurant where they eat. In other words, social life is being carried into the digital world.

The rapid expansion of Internet technologies has redefined the concept of electronic commerce. Because reviews and sharing of ideas by the users of the products that are intended to be sold have become an integral part of online shopping today. Consumers provide fast, easy, and inexpensive access to information through social networks. Traditional newspapers and magazines are too slow and expensive to be compared to social networks. However, the reliability of a system with such advantages can be controversial. This situation, which cannot be ignored, brings some risks. Malicious people try to influence the idea of buying negatively or positively by making deceptive comments that do not reflect the truth in order to mislead the opinions and thoughts of the public about a product (Dematis et al., 2018).

People who use the well-known social media tools such as Twitter, Facebook, and Instagram get an idea and opinion about the product by reading user reviews under an advertised product. They make positive or negative decisions in line with these ideas and thoughts. These decisions, not only affect the decision-makers but can also become public opinion. Because nowadays, access to information is extremely fast. Thus, it is not possible to foresee the extent, severity, and harm of the danger. One of the two competing yogurt brands may try to reduce sales rates with fake comments by putting a few untrue deceptive contents on a web page about the other yogurt company it competes with. In addition, consumers express opinions about the services they receive and the money they pay for this service. These ideas may also include manipulation.

One recent study shows that while 90% of consumers make a decision to buy a service or product, they read and evaluate consumer comments available on the Internet (Rudolph, 2015). This is a very important rate. Moreover, the study shows that 88% of consumers rely on personal recommendations and online consumer comments (Rudolph, 2015). User reviews of such particular services and products pave the way for manipulating the thoughts of people who will buy this product (Dematis et al., 2018). There are many websites around the world whose purpose is to produce only deceptive content and manage people's ideas.

Even if some researchers have recently proposed techniques for solving the problem deception detection in online social network, there is a need to design and use high-performance systems in terms of different evaluation criteria. In this study, a method determining the content that deceives people in online social media by using machine learning algorithms is proposed. In this method, deception detection is modeled as a classification problem. Since the content is text-based, unstructured data sets are converted to structured data sets with text mining stages. Then machine learning algorithms are applied to the data in the structured data set. Performances of adapted algorithms are evaluated in two real public data sets in terms of many evaluation criteria such as precision, accuracy, f-measure, and recall. In Dataset 1, the highest average accuracy value was obtained with 74.4% GBT algorithm, while in Dataset 2, the highest average accuracy value was obtained from the NB algorithm with 71.2%.

The remainder of this paper is structured as follows. Section 2 reviews the related works on the deception detection problem. In Section 3, the details of the proposed model is presented. How to obtain structured data sets from unstructured data sets by text mining is also described in this section. In Section 4, used machine learning algorithms and the data sets that are used to test the deception detection method are introduced. The performance comparison of machine learning algorithms with respect to different metrics are presented. In Section 5, the results of the study are examined and the article is finalized.

RELATED WORKS

The problem of deception detection has an important place in social network analysis. There are original articles in the literature that contain various approaches to solve this important and complex

IJPAS
ijpas@munzur.edu.tr
ISSN: 2149-0910

Research article/Araştırma makalesi
DOI: 10.29132/ijpas.994840

problem. The scientific world is making serious efforts to solve this problem and its popularity is increasing day by day. In this section, the articles and reports related to the detection of deception are introduced. Ott et al. stated that web sites containing consumer reviews have become targets of deceptive content (Ott et al., 2011). In their study, three approaches to detect deceptive content were developed and their performances were compared. As a result, it was stated that a classifier with an accuracy rate of approximately 90% was developed.

Delgado et al. used the machine learning techniques such as Decision Tree (DT), Logistic Regression (LR), Naive Bayesian (NB), Support Vector Machine (SVM), Neural Networks (NN), Random Forest (RF) methods. Articles related with news and e-mails were used as data sources to detect deception. They carried out the classification process using Bag of Words and Part of Speech tag features (Ceballos Delgado et al., 2021).

Krishnaveni and Radha stated in their study that text classification algorithms can be used together with clustering algorithms to get better results. In the experiments, NB, SVM, and DT classification algorithms and K-means, One-Pass, and DBScan clustering algorithms were used together. They observed that it was the most successful case when K-means and SVM algorithms were used together (Krishnaveni et al, 2021).

Kesarvani et al. used data obtained from Facebook to detect deception in their study. Machine learning algorithms such as Logistic Regression (LR), Random Forest, and SVM were used in the experiments. LR classification algorithm showed the highest performance with 98.25% accuracy (Kesarwani et al., 2021).

In his Ph.D. thesis, Merritts aimed to realize the deception detection system automatically and autonomously by using the BDI (Belief, Desire, and Intention) agents. With the prototype developed as a result of the study, it was stated that the data could be classified as “deceptive” or “non-deceptive” with an accuracy rate of 85% (Merritts, 2013).

Wani et al. used the Covid-19 dataset to detect deceptive content. They used deep learning techniques such as Long Short Term Memory (LSTM) and Convolutional Neural Network (CNN) in their experiments. They reached 98.41% as the highest accuracy rate in their study (Wani et al., 2021).

Feng and Hirst tried to distinguish whether the products were original using op_spam_v1.3 (Ott et al., 2011), a data set for hotel reviews. Profile compatibility was used during the experiments. In this way, it was stated that the classification performance was significantly improved (Feng et al, 2013). Sternglanz et al. examined the methods of detecting deception by law enforcement. In addition, they provided information about meta-analytic studies (Sternglanz et al., 2019).

Rill-Garcia et al. tried to detect deception through video data. They tried to improve the results obtained by using the multimodal fusion method. They also used it during the Spanish dataset experiments (Rill-García et al., 2019). Li et al. investigated a general approach to identify online deceptive ideas using new data sets from three different areas (hotel, restaurant, and medicine) (Li et al, 2014a).

Huayi Li et al. reported that in the previous deception detection studies, the texts mentioned generally in English were used and deception detection was proposed for English. In their study, they tried to identify the deceptions in the Chinese texts (Li et al., 2014b). The data sets from the Chinese review site, Dianping, were used to perform these operations.

Jaume Masip stated that the detection of deception was a lively, dynamic field of psychology that has experienced significant developments recently (Masip, 2017). Conroy et al. proposed an innovative hybrid approach by combining networked behavioral data, linguistics, and machine learning approaches (Conroy et al., 2015). Rubin et al. proposed a system that identifies potential types of deceptive news for users and aims to assist users in filtering (Rubin et al., 2015). Three types of fake news were discussed in their work. For predictive modeling, pros, and cons were analyzed.

Litvinova et al. used the data set prepared in Russian language in their study. They reached 68.3% accuracy with the classifier proposed according to some selected parameters (Litvinova et al., 2017). Kumari and Srivastava discussed binary classifiers commonly used in text mining (Kumari et al., 2017). Ding et al. stated that it was very difficult to detect cheating from real life videos. They recommended evaluating the human body and face together to detect deception. (Ding et al., 2019).

IJPAS
ijpas@munzur.edu.tr
ISSN: 2149-0910

Research article/Araştırma makalesi
DOI: 10.29132/ijpas.994840

Kleinberg et al. suggested that cross validation technique should be used to detect deception in their study (Kleinberg et al., 2019). Van der Walt et al. identified the deceptive event by using profile image features in their study. They performed a classification using machine learning methods (Van der Walt et al., 2018). Psychological, linguistic, and computational processes consistently were presented as difficulties in detecting deception (Rosso et al., 2017).

Krishnamurthy et al., developed a multi-modal neural model for the detection of deception. In the experimental results, it was stated that it gave better results than all known methods with an accuracy rate of 96.14% (Krishnamurthy et al., 2018).

Bessi performed a study on the statistical properties of unproven claims and hoaxes in social networks (Bessi, 2017). Zu et al. proposed an efficient model, which focus on the relationship between updated information and false information to reduce the impact of fabricated fake information (Zhu et al., 2018). Supervised machine learning methods for detecting deception as false and fabricated news in social media were examined by Altunbey Ozbay and Alatas (Altunbey Ozbay et al., 2019).

Van Der Zee et al. stated in their study that they found significant differences in correct and incorrect tweets. They stated that they developed a quantitative model. It was stated that this system achieved an accuracy value of 73% (Van Der Zee et al., 2021).

In their study, Levine et al. stated that the accuracy rate of the system they proposed as the deception detection model was 55%, and the error rate was 10%. (Levine et al., 2021).

DECEPTION DETECTION MODEL

The representation of data in intelligent systems significantly affects the performance of results. Particularly, text-based problems need to be converted into a suitable representation. Text mining is a data mining study that accepts the text as a data source (Aggarwal et al., 2012). Text mining is defined as the extraction of previously unknown, useful, and meaningful information from textual data. In this study, text mining was used as pre-processing methods in order to construct a complete deception detection model. According to this model, the basic operations of text mining were applied to the data set. The deception detection model proposed and used in this study is shown in Figure 1.

IJPAS
ijpas@munzur.edu.tr
ISSN: 2149-0910

Research article/Araştırma makalesi
DOI: 10.29132/ijpas.994840

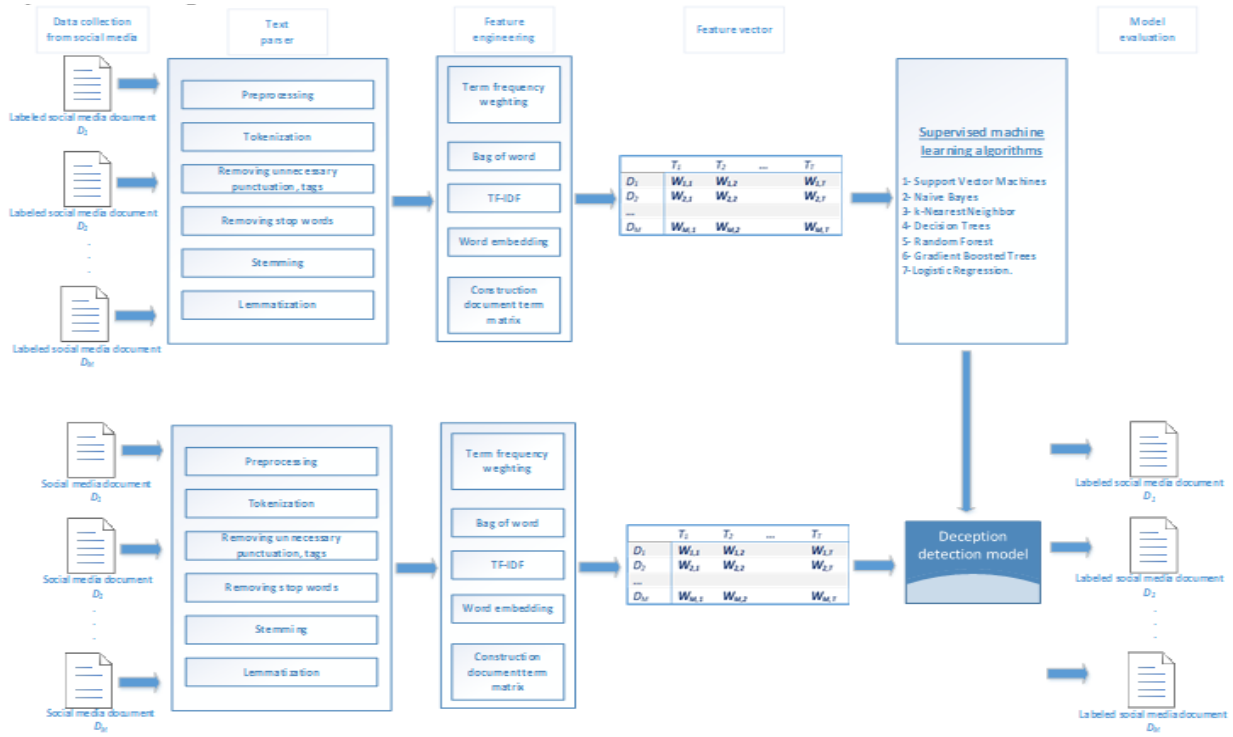

Figure 1. Deception detection model (Baloglu et al., 2019)

Data Preprocessing Steps

Text mining studies are included in the field of natural language processing. Natural language processing studies, mostly include studies based on linguistics under artificial intelligence. On the other hand, text mining aims to reach statistical results through text (Aggarwal et al., 2012; Can et al., 2017).

Text mining applications generally require processing on unstructured data. In order to make sense of unstructured data, we need to make the data workable. In other words, it is necessary to obtain structured data from unstructured data.

Tokenization

This pre-processing step slices the textual data into smaller pieces that are called as tokens. All of the punctuations were removed from the text data in this process (Mullen et al., 2018). Number filter was also applied to delete numbers. In this step, words consisting of less than N characters were deleted from the text (N = 3).

Removing stop words

Stop words in the text do not carry information, but are found in the unique structure of each language. Pronouns, prepositions, and conjunctions are stop-words. Some stop words in English are as follows: a, an, after, about, by, but, when, that, too, on, above, once, until, am, and so on.

Stemming

In the step of stemming, the root states of the words which have the same meaning but different word forms is tried to be found. For instance, the words, tire- tired- tiring, interested- interesting, bored- boring, surprised- surprising, and so on.

Feature Extraction and Selection

In the step of stemming, the root states of the words which have the same meaning but different word forms is tried to be found. For instance, the words, tire- tired- tiring, interested- interesting, bored- boring, surprised- surprising, and so on.

The feature extraction and selection process is the determination of the features that determine which class the data belongs to among the many features of the data (Göker et al., 2017). High dimensional data are one of the biggest problems encountered in text mining. Therefore, in order to develop the model, the higher dimension must be reduced. This is performed by removing unnecessary features from the data. Thus, the search space is reduced that can be more easily studied. The terms in the textual data for each document were weighted and each document was

IJPAS
ijpas@munzur.edu.tr
ISSN: 2149-0910

Research article/Araştırma makalesi
DOI: 10.29132/ijpas.994840

converted into a term weights vector. In vector space model, each word is represented by a numerical value that shows the weight of the word in the text data (Altunbey Ozbay et al., 2019).

Inverse Document Frequency (IDF), Term Frequency (TF), Term Frequency-Inverse Document Frequency (TF-IDF), or binary representation are proposed to indicate the weights (Altunbey Ozbay et al., 2019). Generally, TF and TF-IDF are used among these approaches. In this paper, TF was used for the weights (Azam et al., 2012).

Binary Vectors: The text-containing data in the data set is represented as 0's and 1's.

TF: It refers to the number of repetitions of the word roots in the data as shown in Eq. (1).

TF-IDF: It gives a measure of the number of repetitions (TF) of the word root in the data and of the infrequently repeated words (IDF) in the entire data set. It is computed as in Eq. (2).

$$TF_{ij} = \frac{n_{ij}}{|d_i|} \quad (1)$$

$$IDF_{ij} = log_2^{(\frac{n}{n_j})} \quad (2)$$

d_i is the sum of all terms in the i-th document. n_{ij} is the number of j-th word in the i-th document. When calculating the IDF, n represents the total number of documents, while n_j represents the number of documents in which the j-th term appears.

After the TF value is computed for each word of document, Document Term Matrix (DTM) is constructed using weights of the words. In DTM, each row represents the documents, column indicates the term and cell indicates the term weights (Göker et al., 2017).

During the experiments, a document matrix which was reduced in size according to TF was created. A part of the document matrix was determined to be training data and the rest as test data. Machine learning algorithms were applied to the document matrix and the results were observed.

Machine Learning Algorithms Adapted for Deception Detection

Seven machine learning algorithms were used during the experiments. The reason for this is that there is no machine learning algorithm that works perfectly for each data set. The performance of these algorithms was evaluated in terms of different metrics such as precision, recall, F-measure, and accuracy. The adapted machine learning algorithms as deception detection are SVM, Naive Bayes, K-Nearest Neighbor (k-NN), Decision Trees, Random Forest, Gradient Boosted Trees, and Logistic Regression.

Naive Bayes method greatly simplifies classification task by assuming that attributes are independent given class. Naive Bayes often competes well with more complex supervised machine learning methods although independence is generally a poor assumption (Altay et al., 2019).

SVM is a discriminative machine learning technique formally defined by a separating hyperplane (Osuna et al., 1997). Given labeled training data sets, it outputs an optimal hyperplane which categorizes new samples.

Fix and Hodges proposed a non-parametric algorithm for pattern classification that has since become known as the k-nearest neighbor (k-NN) rule (Fix et al., 1951). k-NN is one of the most simple and fundamental supervised machine learning algorithms and is generally the first choice for a classification task when there is little or no prior knowledge about the data distribution (Peterson, 2009).

The decision tree builds a classification model in the form of a tree structure. It breaks down the data into smaller pieces while at the same time an associated decision tree is incrementally constructed. The final result is a tree with leaf nodes and decision nodes (Friedl et al., 1997).

Random forest method consists of a combination of tree predictors where each tree depends on the values of a random vector sampled independently from the input vector, and each tree casts a unit vote for the most popular class to classify an input vector (Breiman, 2001).

Gradient boosting combines many weak supervised machine learning methods to construct a strong predictive model. Generally decision trees are used when performing gradient boosting (Friedman, 2002).

The logit-the natural logarithm of an odds ratio is the central mathematical concept that underlies logistic regression. Generally, logistic regression is well suited for relationships between one or more continuous or categorical predictor variables and a categorical output variable (Peng et al., 2002).

EXPERIMENTAL RESULTS

In this study, the problem of deception detection was handled as a classification problem. This section

IJPAS
ijpas@munzur.edu.tr
ISSN: 2149-0910

Research article/Araştırma makalesi
DOI: 10.29132/ijpas.994840

compares the classification capabilities of adapted machine learning methods on different sets of experiments. The performances of the algorithms in different metrics are shown in comparative tables. Real-world data sets (TripAdvisor, Hotels.com) were used to evaluate the proposed deception detection models. Details of two real data sets and experimental results obtained from these data sets are presented in the following subsections.

Data Set 1 (Trip Advisor)

The TripAdvisor data set contains an equal number of real and deceptive data on customer satisfaction collected from 20 hotels in Chicago. It is a data set containing 400 correct comments from TripAdvisor and 400 deceiving comments from Mechanical Turk (Ott et al., 2011).

Seven different machine learning algorithms were applied to data set 1 to detect deception. 70% of the data set was used for training and remaining was used for testing the algorithms. The standard parameter values selected in the literature were used for machine learning algorithms. In addition, no parameter analysis and optimization were performed. The results are listed in Table 1.

Table 1. *Performance of machine learning algorithms in data set 1 (70% training, 30% testing)*

	Machine Learning Algorithms						
	Naive Bayes	**Decision Tree**	***k*-NN (*k*=3)**	**Random Forest**	**SVM**	**Logistic Regression**	**Gradient Boosted Trees**
Accuracy	0.746	0.633	0.650	0.754	0.750	0.742	**0.771**
F-Measure	0.729	0.607	0.689	0.755	0.747	0.746	**0.766**
Precision	0.781	0.654	0.620	0.752	0.730	0.734	**0.783**
Recall	0.683	0.567	**0.775**	0.758	0.748	0.758	0.750

Table 2. *Performance of machine learning methods in data set 1 (80% training, 20% testing)*

	Machine Learning Algorithms						
	Naive Bayes	**Decision Tree**	***k*-NN (*k*=3)**	**Random Forest**	**SVM**	**Logistic Regression**	**Gradient Boosted Trees**
Accuracy	0.700	0.694	0.656	0.719	0.715	**0.756**	0.725
F-Measure	0.692	0.703	0.696	0.717	0.709	**0.766**	0.718
Precision	0.711	0.682	0.624	0.722	0.702	**0.736**	0.735
Recall	0.675	0.725	0.787	0.713	0.703	**0.800**	0.700

When Table 1 is examined, Gradient Boosted Trees seems the best machine learning algorithm in terms of accuracy, f-measure, and precision metrics. Its accuracy is 0.771 for data set 1. Decision Tree is the worst method in terms of accuracy, f-measure, and recall values for this dataset.

When Table 2 is examined, Logistic Regression for data set 1 outperformed all machine learning

IJPAS
ijpas@munzur.edu.tr
ISSN: 2149-0910

Research article/Araştırma makalesi
DOI: 10.29132/ijpas.994840

algorithms with respect to all metrics. The accuracy of Logistic Regression in this dataset is 0.756. Naive Bayes seems the worst method in terms of f-measure and recall values obtained in this dataset. k-NN is the worst method in terms of precision and accuracy metrics.

The results of another experiment in which the cross-validation test was performed are shown in Table 3. 5-fold cross-validation was performed during this experiment. The standard parameters in the literature were used for machine learning methods. No parameter analysis and optimization were performed. When Table 3 is checked, it is seen that Naive Bayes outperformed other machine learning methods in term of the f-measure, accuracy, and precision values. Its accuracy value is 0.752. Decision Tree is the worst method in terms of all metrics.

Table 3. *Performance of machine learning algorithms in data set 1 (5-Fold Cross Validation)*

	Machine Learning Algorithms						
	Naive Bayes	**Decision Tree**	***k*-NN (*k*=3)**	**Random Forest**	**SVM**	**Logistic Regression**	**Gradient Boosted Trees**
Accuracy	**0.752**	0.598	0.640	0.733	0.751	0.731	0.736
F-Measure	**0.755**	0.662	0.686	0.737	0.753	0.727	0.742
Precision	**0.748**	0.572	0.609	0.729	**0.748**	0.739	0.726
Recall	0.753	0.599	**0.785**	0.745	0.758	0.715	0.760

Table 4. *Mean performances of machine learning algorithms in data set 1*

	Machine Learning Algorithms						
	Naive Bayes	**Decision Tree**	***k*-NN (*k*=3)**	**Random Forest**	**SVM**	**Logistic Regression**	**Gradient Boosted Trees**
Accuracy	0.733	0.642	0.649	0.735	0.739	0.743	**0.744**
F-Measure	0.726	0.657	0.690	0.736	0.736	**0.746**	0.742
Precision	0.747	0.636	0.618	0.734	0.727	0.736	**0.748**
Recall	0.704	0.630	**0.782**	0.739	0.736	0.758	0.737

When the values obtained from these three experiments are averaged for the algorithms, the mean metric values of the algorithms are demonstrated in Table 4. As seen in this table, SVM has the highest mean accuracy value according to the averaged results of three experiments.

Data Set 2 (Hotels.com)

The Hotels.com data set contains an equal number of real and deceptive data on customer satisfaction collected from 20 hotels in Chicago. It is a data set containing 400 correct comments from (TripAdvisor, Expedia, Orbitz, Hotels.com, Priceline and Yelp) and 400 deceiving comments from Mechanical Turk (Sternglanz et al.,2019).

Seven different machine learning methods were applied to data set 2 to detect deception. 70% of the data set was selected for training and remaining of the total data set was selected for testing the methods. The standard parameter values were used for machine learning algorithms. Parameter analysis and optimization were not performed. The results are shown in Table 5.

When Table 5 is examined, Naive Bayes and Logistic Regression for data set 2 are the best machine learning algorithms with an accuracy of 0.704. The worst-case machine learning algorithm for Data set 2 was the Decision Tree with an accuracy of 0.625. Logistic Regression outperformed all methods in terms of f-measure, accuracy, and recall values

IJPAS
ijpas@munzur.edu.tr
ISSN: 2149-0910

Research article/Araştırma makalesi
DOI: 10.29132/ijpas.994840

obtained from this data set for deception detection problem. Decision Tree seems the worst method according to the accuracy, f-measure, and precision metrics for this data set.

80% of the data set was used for training and remaining was used for testing the algorithms as another experiment. The standard parameter values used in the literature were used for machine learning algorithms. In addition, no parameter analysis and optimization were performed. The results obtained are demonstrated in Table 6.

Random Forest outperformed all methods in terms of f-measure, accuracy, and recall values obtained from this data set for deception detection problem according to the results of this experiment. Naive Bayes and Random Forest for data set 2 are the best machine learning algorithms with an accuracy of 0.706. Decision Tree is the worst method in terms of all metrics.

The results of another experiment in which the cross-validation test was performed are shown in Table 7. 5-fold cross-validation was used during the experiment. The standard parameter values in the literature were selected for machine learning algorithms. No parameter analysis and optimization were performed. When Table 7 is examined, Naive Bayes seems the best machine learning algorithm in terms of accuracy, recall, and f-score for data set 2. The worst machine learning algorithm for data set 2 was the *k*-NN with an accuracy of 0.651.

Table 5. *Performance of machine learning methods in data set 2 (70% training, 30% testing)*

	Machine Learning Algorithms						
	Naive Bayes	**Decision Tree**	***k*-NN (*k*=3)**	**Random Forest**	**SVM**	**Logistic Regression**	**Gradient Boosted Trees**
Accuracy	**0.704**	0.625	0.692	0.692	0.700	**0.704**	0.667
F-Measure	0.682	0.648	0.686	0.704	0.691	**0.715**	0.688
Precision	**0.738**	0.610	0.698	0.677	0.712	0.690	0.647
Recall	0.633	0.692	0.675	0.733	0.652	**0.742**	0.733

Table 6. *Performance of machine learning methods in data set 2 (80% training, 20% testing)*

	Machine Learning Algorithms						
	Naive Bayes	**Decision Tree**	***k*-NN (*k*=3)**	**Random Forest**	**SVM**	**Logistic Regression**	**Gradient Boosted Trees**
Accuracy	**0.706**	0.606	0.700	**0.706**	0.702	0.681	0.681
F-Measure	0.693	0.623	0.696	**0.728**	0.718	0.691	0.695
Precision	**0.726**	0.598	0.705	0.677	0.670	0.671	0.667
Recall	0.662	0.650	0.688	**0.787**	0.760	0.713	0.725

Table 7. *Performance of machine learning algorithms in data set 2 (5-Fold Cross Validation)*

	Machine Learning Algorithms

IJPAS
ijpas@munzur.edu.tr
ISSN: 2149-0910

Research article/Araştırma makalesi
DOI: 10.29132/ijpas.994840

	Naive Bayes	**Decision Tree**	***k*-NN (*k*=3)**	**Random Forest**	**SVM**	**Logistic Regression**	**Gradient Boosted Trees**
Accuracy	**0.725**	0.696	0.651	0.720	0.721	0.701	0.708
F-Measure	**0.728**	0.682	0.659	0.717	0.724	0.702	0.695
Precision	0.721	0.680	0.644	0.724	0.717	0.700	**0.730**
Recall	**0.735**	0.653	0.675	0.710	0.730	0.698	0.663

Table 8. *Mean performance of machine learning algorithms in data set 2*

	Machine Learning Algorithms						
	Naive Bayes	**Decision Tree**	***k*-NN (*k*=3)**	**Random Forest**	**SVM**	**Logistic Regression**	**Gradient Boosted Trees**
Accuracy	**0.712**	0.642	0.681	0.706	0.708	0.695	0.685
F-Measure	0.701	0.651	0.680	**0.716**	0.711	0.703	0.693
Precision	**0.728**	0.629	0.682	0.693	0.700	0.687	0.681
Recall	0.677	0.665	0.679	**0.743**	0.714	0.718	0.707

When the values obtained from these three experiments are averaged for the algorithms in data set 2, the mean metric values of the algorithms are demonstrated in Table 8. As seen in this table, Random forest has the best mean f-measure and recall values while Naïve Bayes has the highest mean accuracy and precision values according to the averaged results of three experiments performed in data set 2.

CONCLUSION

Today, the accuracy and reliability of information have gained more importance with the widespread use of social media. In this article, a methodology for the problem of deception detection in online social networks was proposed by using text mining techniques and machine learning algorithms. The problem of deception detection was modeled as a classification problem in this study. The proposed model was tested on two different real data sets. Performances of the machine learning algorithms were evaluated with respect to accuracy, precision, f-measure, and recall metrics using three experiments.

When all the experimental results were evaluated as a whole, GTB performed better in terms of the obtained mean f-measure and accuracy values for dataset1. NB and RF algorithms showed equal performance for data set 2. NB algorithm reached the highest accuracy value in data set 2. Decision Trees was the worst algorithm in both data set 1 and data set 2 with respect to all evaluation metrics for the deception detection problem that was modeled as a classification problem in this study. There is no single algorithm that solves all problems.

In Dataset 1, the highest average accuracy value was obtained with 74.4% GBT algorithm, while in Dataset 2, the highest average accuracy value was obtained from the NB algorithm with 71.2%.

In future studies, the model proposed and used in this article may be improved by exploring new algorithms, integrating metaheuristic search and optimization methods, and hybridizing the current algorithms for more efficient results. Different feature extraction techniques and ensemble methods can also be integrated for enhancing the performance of the deception detection system in terms of many important metrics.

CONFLICT OF INTEREST

The Authors report no conflict of interest relevant to this article

RESEARCH AND PUBLICATION ETHICS STATEMENT

The authors declare that this study complies with research and publication ethics.

REFERENCES

Aggarwal, C. C., Zhai, C. (Eds.). (2012). Mining text data. Springer Science & Business Media.

Altay, O., Ulas, M., Mahmut, O. Z. E. R., Ece, G. E. N. C. (2019). An expert system to predict warfarin dosage

IJPAS
ijpas@munzur.edu.tr
ISSN: 2149-0910

Research article/Araştırma makalesi
DOI: 10.29132/ijpas.994840

in Turkish patients depending on genetic and non-genetic factors. In IEEE 7th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-6).

Altunbey Ozbay, F., Alatas, B. (2019). Fake news detection within online social media using supervised artificial intelligence algorithms, Physica A, https://doi.org/10.1016/j.physa.2019.123174.

Azam, N., Yao, J. (2012). Comparison of term frequency and document frequency based feature selection metrics in text categorization. Expert Systems with Applications, 39(5), 4760-4768.

Baloglu, U. B., Alatas, B., Bingol, H. (2019). Assessment of Supervised Learning Algorithms for Irony Detection in Online Social Media. In 2019 1st International Informatics and Software Engineering Conference (UBMYK) (pp. 1-5). IEEE.

Baydogan, C., Alatas, B. (2021). Metaheuristic Ant Lion and Moth Flame Optimization-Based Novel Approach for Automatic Detection of Hate Speech in Online Social Networks. IEEE Access, 9, 110047-110062.

Bessi, A. (2017) On the statistical properties of viral misinformation in online social media, Physica A 469, 459–470

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Can, Ü., Alataş, B. (2017). Review of Sentiment Analysis and Opinion Mining Algorithms. International Journal of Pure and Applied Sciences, 3(1), 75-111.

Ceballos Delgado, A. A., Glisson, W., Shashidhar, N., Mcdonald, J., Grispos, G., Benton, R. (2021). Deception Detection Using Machine Learning. In Proceedings of the 54th Hawaii International Conference on System Sciences (p. 7122).

Conroy, N. J., Rubin, V. L., Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. In Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community (p. 82). American Society for Information Science.

Dematis, I., Karapistoli, E., Vakali, A. (2018). Fake Review Detection via Exploitation of Spam Indicators and Reviewer Behavior Characteristics. In International Conference on Current Trends in Theory and Practice of Informatics (pp. 581-595). Edizioni Della Normale, Cham.

Ding, M., Zhao, A., Lu, Z., Xiang, T., & Wen, J. R. (2019). Face-focused cross-stream network for deception detection in videos. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7802-7811).

Feng, V. W., Hirst, G. (2013). Detecting deceptive opinions with profile compatibility. In Proceedings of the Sixth International Joint Conference on Natural Language Processing (pp. 338-346).

Fix, E., Hodges Jr, J. L. (1951). Discriminatory analysis-nonparametric discrimination: consistency properties. California Univ Berkeley.

Friedl, M. A., Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. Remote sensing of environment, 61(3), 399-409.

Friedman, J. H. (2002). Stochastic gradient boosting. Computational statistics & data analysis, 38(4), 367-378.

Göker, H., Tekedere, H. (2017). FATIH Projesine Yönelik Görüşlerin Metin Madenciliği Yöntemleri İle Otomatik Değerlendirilmesi. Bilişim Teknolojileri Dergisi, 10(3), 291-299.

Kesarwani, A., Chauhan, S. S., Nair, A. R., & Verma, G. (2021). Supervised Machine Learning Algorithms for Fake News Detection. In Advances in Communication and Computational Technology (pp. 767-778). Springer, Singapore.

Kleinberg, B., Arntz, A., & Verschuere, B. (2019). Being accurate about accuracy in verbal deception detection. PloS one, 14(8), e0220228.

Krishnamurthy, G., Majumder, N., Poria, S., & Cambria, E. (2018). A deep learning approach for multimodal deception detection. arXiv preprint arXiv:1803.00344.

Krishnaveni, N., & Radha, V. (2021). Performance Evaluation of Clustering-Based Classification Algorithms for Detection of Online Spam Reviews. In Data Intelligence and Cognitive Informatics (pp. 255-266). Springer, Singapore.

Kumari, R., Srivastava, S. K. (2017). Machine learning: A review on binary classification. International Journal of Computer Applications, 160(7).

Levine, T. R., Daiku, Y., & Masip, J. (2021). The Number of Senders and Total Judgments Matter More Than Sample Size in Deception-Detection Experiments. Perspectives on Psychological Science, 1745691621990369.

Li, H., Liu, B., Mukherjee, A., Shao, J. (2014). Spotting fake reviews using positive-unlabeled learning. Computación y Sistemas, 18(3), 467-475.

Li, J., Ott, M., Cardie, C., Hovy, E. (2014). Towards a general rule for identifying deceptive opinion spam. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (Vol. 1, pp. 1566-1576).

Litvinova, O., Seredin, P., Litvinova, T., & Lyell, J. (2017). Deception detection in russian texts. In Proceedings of the Student Research Workshop at the 15th Conference of the European Chapter of the Association for Computational Linguistics (pp. 43-52).

Masip, J. (2017). Deception detection: State of the art and future prospects. Psicothema, 29(2), 149-159.

IJPAS
ijpas@munzur.edu.tr
ISSN: 2149-0910

Research article/Araştırma makalesi
DOI: 10.29132/ijpas.994840

Merritts, R. A. (2013). Online Deception Detection Using BDI Agents.

Mullen, L. A., Benoit, K., Keyes, O., Selivanov, D., & Arnold, J. (2018). Fast, Consistent Tokenization of Natural Language Text. Journal of Open Source Software, 3(23), 655.

Osuna, E., Freund, R., Girosit, F. (1997). Training support vector machines: an application to face detection. In Proceedings of IEEE computer society conference on computer vision and pattern recognition (pp. 130-136). IEEE.

Ott, M., Choi, Y., Cardie, C., Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. In Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1 (pp. 309-319).

Peng, C. Y. J., Lee, K. L., Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. The journal of educational research, 96(1), 3-14.

Peterson, L. E. (2009). K-nearest neighbor. Scholarpedia, 4(2), 1883.

Rill-García, R., Jair Escalante, H., Villasenor-Pineda, L., & Reyes-Meza, V. (2019). High-level features for multimodal deception detection in videos. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 0-0).

Rosso, P., Cagnina, L. C., (2017). Deception Detection and Opinion Spam, A practical guide to sentiment analysis, 155-171, Springer.

Rubin, V. L., Chen, Y., Conroy, N. J. (2015). Deception detection for news: three types of fakes. In Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community (p. 83). American Society for Information Science.

Rudolph, S. (2015). The impact of online reviews on customers' buying decisions. Business 2 Community.

Sternglanz, R. W., Morris, W. L., Morrow, M., & Braverman, J. (2019). A review of meta-analyses about deception detection. The Palgrave handbook of deceptive communication, 303-326.

Van der Walt, E., Eloff, J. H., & Grobler, J. (2018). Cyber-security: Identity deception detection on social media platforms. Computers & Security, 78, 76-89.

Van Der Zee, S., Poppe, R., Havrileck, A., & Baillon, A. (2021). A personal model of Trumpery: linguistic deception detection in a real-world high-stakes setting. Psychological science, 09567976211015941.

Wani, A., Joshi, I., Khandve, S., Wagh, V., & Joshi, R. (2021). Evaluating Deep Learning Approaches for Covid19 Fake News Detection. arXiv preprint arXiv:2101.04012.

Zhu, H., Wu, H., Cao, J., Fu, G., Li, H. (2018). Information dissemination model for social media with constant updates, Physica A 502, 469–482