Available online at www.dergipark.org.tr/en

INTERNATIONAL ADVANCED RESEARCHES
and
ENGINEERING JOURNAL

Journal homepage: www.dergipark.org.tr/en/pub/iarej

e-ISSN: 2618-575X

International Open Access

Volume 08
Issue 03

December, 2024

Research Article

Predicting medical drug usage intentions via SGD-based text classification model

Duygu Bağcı Daş [a,*]

[a] *Department of Computer Programming, Ege University, Izmir, 35100, Turkiye*

ARTICLE INFO

Article history:
Received 03 Juney 2024
Accepted 14 December 2024
Accepted 20 December 2024

Keywords:
Drug usage intention
Machine learning
Medical drug usage
Text classification
Text mining

ABSTRACT

The effects of medical drugs and their usage purposes vary among individuals due to the chemical composition of drugs, side effects, genetics, etc. Even if those effects are to be discovered pharmacologically, they cannot be fully understood. Hence, it becomes essential to analyze the individuals' reviews and experiences to unearth such effects and find out which other purposes drugs are used for, in addition to the target disease they are developed to cure. Text classification methods present various solutions to analyze those reviews effectively. Generally, these effects are investigated in terms of emotional analysis of medical drug usage experience as positive or negative. However, some drugs can be used for more than one specific treatment. For example, an antipsychotic drug can be used for both depression and anxiety or ADHD. Therefore, the effects of medical drug users and drug names to be associated with the review of the studies should be covered comprehensively. Based on this motivation, this study proposed a lightweight model for the prediction of medical drug usage intentions using text-based patient reviews. For this purpose, TF-IDF and bigram methods are used for text classification in the feature extraction step, then the Stochastic Gradient Descent (SGD) classifier is used for prediction and compared to other popular machine learning algorithms. Classification results indicate that the SGD and TF-IDF-Bigram approach effectively predicts drug usage intentions for medical purposes with an accuracy of 98.42%. Based on the outcomes, it is concluded that the findings of this study may be beneficial in pharmaceutics or medicine considering drug design, reducing side effects, health management, treatment adherence and process design, and personalized medicine.

1. Introduction

Machine learning techniques have been increasingly used in recent decades in the medical field [1-3]. As a subdomain of Machine Learning, text classification is a methodology for structuring, organizing, and tagging a topic based on texts. This methodology is utilized in the medical field, including diagnosis, prognosis, personalized medicine, etc. For instance, a disease could be diagnosed by using former patients' complaints, or by analyzing their sentiments on social media. All of these abilities make text classification more popular among researchers. Some of the studies presented by the researchers are briefly described below.

Liu [4] presented a high-quality classifier for supporting healthcare information support in Chinese disease-related information. The classifier was designed to help identify the cause, symptom, curing, side effects, and prevention of cancer. Billyan et al. [5] performed sentiment analysis with machine learning approaches for restaurant businesses. They used Fuzzy K-Nearest Neighbor as the classification method to predict whether the customer reviews are positive or negative. Pratama and Sarno [6] predicted the personality of individuals based on the text that they were written on Twitter. They classified English and Indonesian languages by using Naive Bayes (NB), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). Sulea et al. [7] investigated the text classification methods that can support law professionals. They applied several machine learning methods to predict the ruling and the law area. Olsson et al. [8] aimed to classify Czech documents using cross-language text classification with English training data.

In addition to those studies, researchers have investigated the implementation and combination of machine learning algorithms to improve the effectiveness of text classification. Some of the studies related to this

* Corresponding author. Tel.: +0-232-311-1475; Fax: +0-232-388-7599.
E-mail addresses: duygu.bagci.das@ege.edu.tr
ORCID: 0000-0003-4519-3531
DOI: 10.35860/iarej.1495330

topic are as follows. Li and Park [9] presented a novel method that helps to classify documents effectively. They utilized the learning phase evaluation back propagation neural network and singular value decomposition that improved the document classification efficiently. Tang et al. [10] proposed a Bayesian classification approach for automatic text classification. They used class-specific features by following Baggenstoss's PDF Projection Theorem to rebuild the PDF within the raw data space. Kumar et al. [11] investigated the text classification performance on machine learning by using the Back Propagating Artificial Neural Network (ANN) model. Shen et al. [12] presented an improved parallel Bayesian algorithm theoretically and experimentally to improve the speed of the classification on the Hadoop cloud platform. Dalal and Zaveri [13] proposed an enhanced hybrid classification method, which is the combination of NB and SVM for text document preprocessing.

Shokrpour et al. [14] investigated the best algorithm that classifies drug interactions. They used NB, Multinominal NB, Zero R, Jrip, LibSVM, j48, Random Forest, and Bagging algorithms to evaluate the classes. Chai et al. [15] used 25K tweets that describe drug use to obtain drug effectiveness by performing relation extraction. Aramaki et al. [16] investigated the extraction of adverse drug events and effects by using clinical records. They divided their study into determining the amount of information that existed in the files and automatic extracting of the Natural Language Processing (NLP) system. Graber et al. [17] used online reviews of drugs related to five disorders to perform aspect-based sentiment analysis, applying cross-domain, and cross-data learning. They used two datasets, namely, Drugscom and Druglib. Lavecchia [18] investigated machine-learning approaches within the scope of drug discovery. Five different machine learning algorithms are compared with each other in terms of classification error, computational cost, memory requirements, complexity for implementation, on/offline application, interpretation, advantages, and disadvantages.

Abada et al. [19] investigated the relationships between substance use and mental health conditions (alcohol, opioids, tobacco, anxiety, depression, insomnia) in addiction. They identified patterns and correlations across these categories using five Machine Learning algorithms and Principal Component Analysis (PCA). Elahi et al. [20] developed an attention-based graph network, a recommendation framework that combines both relational and contextual information to improve recommendation systems. Their model addresses challenges in large, sparse datasets, specifically in health informatics, by using a user-specific attention mechanism to personalize recommendations.

This research proposes a predictive model for drug usage intentions, which explores how drugs may be used for conditions beyond their intended purpose whereas most of the studies in this field generally focus on drug-based adverse effects and sentiment analysis. This study investigated the prediction of drug usage intentions using machine learning-based text classification from an online review dataset by using the Python 3.7 NLTK tool. The Stochastic Gradient Descent (SGD) classifier with TF-IDF (Term Frequency-Inverse Document Frequency) and bigram feature extraction techniques is used for the first time for predicting drug usage intentions.

The rest of the paper is arranged as follows: Section 2 explains the necessary preparations, preprocessing, and feature extraction methods to perform machine learning-based text classification. Section 3 presents the effects of the preprocessing techniques on the model performance and the classification results. Finally, the discussion about the outcomes is given in Section 4.

2. Material and Method

2.1 The Dataset and Preprocessing

Drugs.com dataset, which contains user reviews on specific medical drugs, is used to evaluate an effective method for the prediction of drug usage intentions. The dataset contains 215063 reviews of 6345 drugs [17]. The top fifteen usage based on the reviews are selected because the dataset is unbalanced (e.g., there are 2 reviews for only one drug condition, whereas there are 28788 for another drug condition). Table 1 shows the dataset summary and Table 2 contains some example reviews in the pharmaceutical field. For all experiments, the dataset is split into training and testing sets at a ratio of 75% - 25%.

Table 1. Dataset summary [17]

Abb.	Condition Name	Number of Reviews
C1	Birth Control	28788
C2	Pain	6145
C3	Acne	5588
C4	Insomnia	3673
C5	Obesity	3568
C6	Diabetes Type 2	2554
C7	High Blood Pressure	2321
C8	Abnormal Uterine Bleeding	2096
C9	Depression	9069
C10	Anxiety	5904
C11	Bipolar Disorder	4224
C12	Weight Loss	3609
C13	ADHD	3383
C14	Emergency Contraception	2463
C15	Vaginal Yeast Infection	2274

Table 2. Example reviews [17]

This drug has turned down the volume on a lot of my voices. Which is great. I sleep better and take more naps during the day. Also notice my appetite increased which is where I think people mean they gain weight on this med. I just exercise an hour a day. My mood is kinda flat lined.
I used them and within 6 hours my eyes swelled almost shut and cheeks puffed up 2 weeks later. I still have red watery eyes but thankfully the swelling is finally gone. I will never use this product again obviously but it is also incredibly expensive. I have never had an allergic reaction to any medication before so this was very unexpected.

Figure 1 depicts the steps of a machine learning-based text classification procedure. In the first step, the corpus is purified from unnecessary data by using various pre-processing techniques on the corpus data as illustrated in Figure 2. This is a critical process since it has a direct impact on the model performance. Then, feature extraction is performed and a feature vector is created. The created feature vectors are associated with labels and the classifier model is trained with this data. The trained model is tested with new unseen data.

Many stop words in English are used in Natural Language Processing. Table 3 shows a list of the used stop words existing in NLTK library.

Figure 1. Machine learning-based text classification

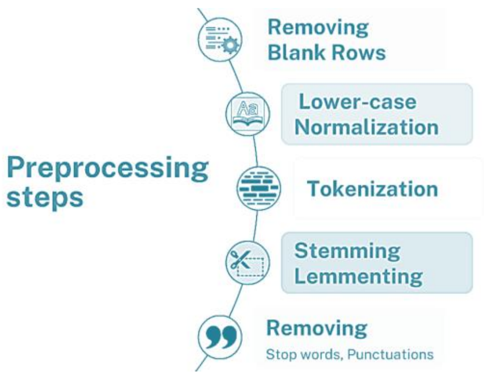

Figure 2. Preprocessing steps for the proposed text classification model

Table 3. Stop words list

i	they	a	in	own	doesn
me	them	an	out	same	doesnt
my	their	the	on	so	hadn
myself	theirs	and	off	than	hadnt
we	themselves	but	over	too	hasn
our	what	if	under	very	hasnt
ours	which	or	again	s	haven
ourselves	who	because	further	t	havent
you	whom	as	then	can	isn
youre	this	until	once	will	isnt
youve	that	while	here	just	ma
youll	thatll	of	there	don	mightn
youd	these	at	when	dont	mightnt
your	those	by	where	should	mustn
yours	am	for	why	shouldve	mustnt
yourself	is	with	how	now	needn
yourselves	are	about	all	d	neednt
he	was	against	any	ll	shan
him	were	between	both	m	shant
his	be	into	each	o	shouldn
himself	been	through	few	re	shouldnt
she	being	during	more	ve	wasn
shes	have	before	most	y	wasnt
her	has	after	other	ain	weren
hers	had	above	some	aren	werent
herself	having	below	such	arent	won
it	do	to	no	couldn	wont
its	does	from	nor	couldnt	wouldn
its	did	up	not	didn	wouldnt
itself	doing	down	only	didnt	

2.2 TF-IDF and Bigram methods

In this study, feature extraction procedure is performed by using the TF-IDF, and the bigrams. The SGD is used to classify drug usage intentions.

The TF-IDF and the bigram calculations are frequently used techniques in text classification processes. TF-IDF is derived from Term Frequency (TF) and Inverse Document Frequency (IDF). TF refers to the number of times a term occurs in a document. Generally, the frequency of a term in a document indicates the importance of such a term in that document. IDF is a metric indicating how common or rare a term is. If a term is very common, its IDF value will be low; if it is rare, the IDF value will be high. This shows the importance of a term existing in the corpus of text [21].

TF and IDF are utilized to evaluate the importance of a specific term in a text.

The multiplication of TF and IDF values of a term gives the importance of that term in both the document and the overall collection. The mathematical expressions of TF, IDF, and TF-IDF are briefly given as follows.

$$TF(i,d) = \frac{n_{i,d}}{\sum_{i' \in d} n_{i',d}} \tag{1}$$

$$IDF(i,C) = log \frac{N_c}{|\{d \in C: i \in d|} \tag{2}$$

$$TF\text{-}IDF(i,d)= TF(i,d)\ x\ IDF(i,C) \tag{3}$$

where $n_{i,d}$ is number of the word i shows up in the document d, $\sum_{i' \in d} n_{i',d}$ is the total word count in document d, N_c is the number of documents in a Corpus, and $|\{d \in C: i \in d|$ represents the number of documents where the word i exists. N-grams can take different forms, such as single words (unigrams), pairs of words (bigrams), trios of words (trigrams), or sequences of any length. Here n indicates the number of consecutive words. N-grams are often used to extract unique patterns in a text, obtaining meaningful data from textual data.

In this study, the bigrams are denoted in texts as follows. For example, “side effects” or “allergic symptoms” are some bigrams due to the frequent use of these two words together, and such bigrams hold representative characteristics for understanding pharmaceutical text data regarding the use of drugs.

2.3. Stochastic Gradient Descent (SGD) Classifier

SGD Classifier is a linear classifier a classification model (such as SVM or logistic regression) optimized by the SGD. As a type of SVM, SGD is defined as a classification approach whose goal is finding the optimal decision boundary in order to distinguish instances of different classes [22]. SGD is an effective technique to be implemented in large datasets since it iteratively updates the model parameters by calculating gradients on randomly chosen mini-batches of the train data instead of considering the entire dataset. Owing to this positive aspect, its training procedure is fast and it is scalable for large datasets. Mathematically, SGD can be expressed as follows.

$$O(w) = 1\frac{1}{n}\sum_{i=1}^{n} L(y_i \hat{y}_i) + \mu\|w\|^2 \tag{4}$$

where $O(w)$ represents the objective function while n is the number of training samples. $L(y_i \hat{y}_i)$ is the loss function and μ is the regularization parameter depending on w representing the weight vector. In this paper, the modified Huber loss function and L2 regularizer are considered.

The success of the considered approach is determined by interpreting the accuracy results of the classification.

2.4 One-vs-Rest (OvR) approach

One-vs-Rest (OvR) is a method in multi-class classification that converts the problem into binary classification problems. A separate binary classifier is trained for each class, and each classifier tries to distinguish that class (positive instances) from all other classes (negative instances). The test data is then applied to all classifiers, and each classifier makes a prediction. The ultimate decision is based on the classifier with the highest score. The evaluation of the SGD model is performed based on the OvR approach within the study.

3. Results

The analysis results of the different text classification processes by using machine learning algorithms are evaluated for the bigram and the TF-IDF methods. The assessment of the classifier is made considering accuracy, precision, recall, and F1 scores.

Figure 3 shows the performance metrics for the bigram considering NB (Clf-1), SVM (Clf-2), KNN (Clf-3, and RF (Clf-4) classifiers as basic classifiers. As seen in Figure 3, Clf-2 that used the bigrams had the highest performance owing to its accuracy (81%), precision (0.85), recall (0.71), and f1 score (0.81) values whereas Clf-1 classifier had the lowest performance when used Bigrams.

Figure 4 presents the performance metrics for the same classifiers where they are utilized with the Bigrams and the TF-IDF. According to the results given in Figure 4, it is observed that Clf-2 classifier shows the best performance due to its accuracy (89%), precision (0.90), recall (0.85), and f1 score (0.87). On the other hand, Clf-1 classifier has the worst performance when considering the same text processing combination.

Regarding both Figures 3 and 4, Clf-3 is placed second and Clf-4 is the third best approach in classifying other purposes of drug use. Comparing the text processing techniques, it is observed that using the TF-IDF with the bigrams as a combination increased the accuracy by 8%, meaning that it is beneficial to use such text processing methods together. On the other hand, the performance metrics did not reach the desired level using Clf-1, Clf-2, Clf-3, and Clf-4. Due to its promising performance in handling large datasets as expressed in Section 2.3 the SGD classifier, as an improved version of Clf-2 (SVM), is employed to leverage the classification performance to a higher level.

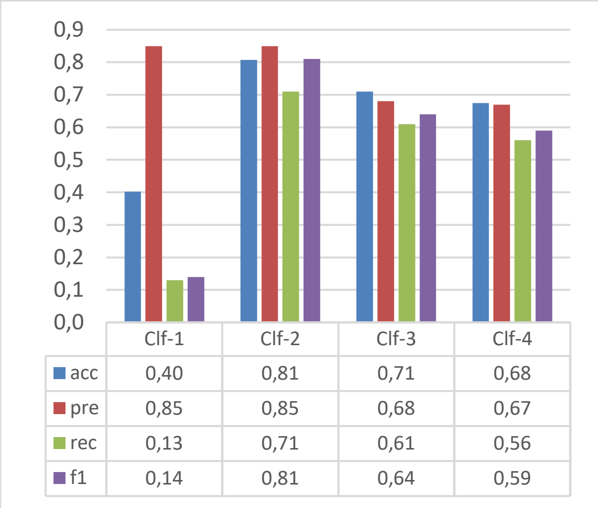

Figure 3. The performance metrics of different base classifiers for the Bigram technique

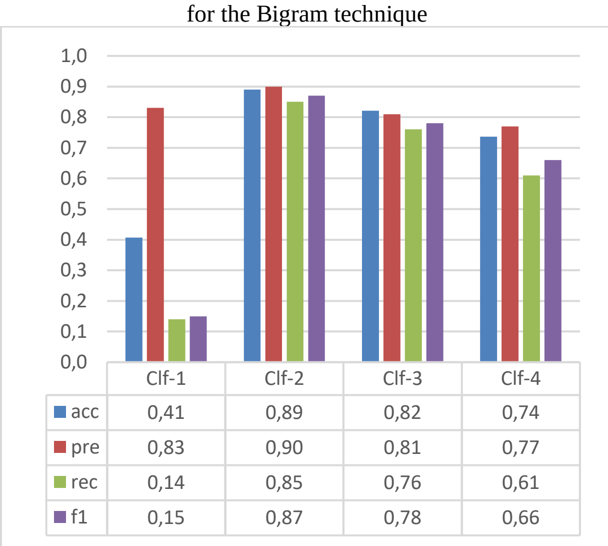

Figure 4. The performance metrics of different base classifiers for the Bigram and TF-IDF techniques

The classification results using the SGD (Clf-5) classifier and the bigrams – TF-IDF text processing combination are presented in Table 4. As seen in Table 4, Clf-5 significantly increased the accuracy from 89% to 98.42%. Owing to its high precision (0.98) and recall (0.98) values given in Table 5 indicate the robustness and responsiveness of the model meaning that it can deal with distorted and different data effectively.

Class-based results given in Table 5 show that classes C9 (Depression) and C14 (Emergency Contraception) are perfectly associated with the corresponding drug. The use of drugs for Diabetes Type 2 and High Blood Pressure are classified as less accurate than others, with accuracy values of 96% and 97%, respectively. All other medical drug usage conditions are successfully classified with accuracy values between 97% and 100%, meaning that Clf-5 with the bigrams and TF-IDF text processing can be utilized as an effective tool for understanding the effect of the drugs in treating diseases or conditions different from that they are developed for.

Table 4. Accuracy Metrics of the Bigram-TF-IDF/SGD

Class No	Accuracy	Class No	Accuracy
C1	98.26%	***C9***	99.84%
C2	99.01%	***C10***	99.30%
C3	99.47%	***C11***	98.97%
C4	96.62%	***C12***	97.79%
C5	98.23%	***C13***	99.07%
C6	97.00%	***C14***	99.87%
C7	95.71%	***C15***	97.90%
C8	99.31%	***AVG***	**98.42%**

Table 5. Weighted average performance metrics of the Bigram-TF-IDF/SGD Method

Class No.	Precision	Recall	F1 Score
C1	0.98	0.98	0.98
C2	0.99	0.99	0.99
C3	0.99	0.99	0.99
C4	0.96	0.97	0.96
C5	0.98	0.98	0.98
C6	0.97	0.97	0.97
C7	0.96	0.96	0.96
C8	0.99	0.99	0.99
C9	1.00	1.00	1.00
C10	0.99	0.99	0.99
C11	0.99	0.99	0.99
C12	0.98	0.98	0.98
C13	0.99	0.99	0.99
C14	1.00	1.00	1.00
C15	0.98	0.98	0.98
AVG	***0.98***	***0.98***	***0.98***

Some deep learning-based approaches are proposed in the related field. For example, Al-Hadhrami et al. [23] evaluated the bidirectional long short-term memory and convolutional neural network (BiLSTM-CNN) models for sentiment analysis of patient drug reviews. They considered two different datasets where the BiLSTM-CNN model achieved an accuracy of 96%. Dandala et al. [24] studied adverse drug events based on Deep learning techniques. They employed a combined BiLSTM and conditional random fields (CRF) model for medical entity detection and a BiLSTM-attention network to understand the relationship between drugs and symptoms. The F-measure values of the proposed combined approach are 0.83 for entity detection and 0.87 for relation detection. Colon-Ruiz and Segura-Bedmar [25] compared different Deep learning architectures for sentiment analysis of drug reviews. They considered CNN, LSTM, and the combinations of these models. Additionally, they investigated the impact of pre-trained word embeddings, including Bidirectional Encoder Representations from Transformers (BERT) combined with Bi-LSTM. They found that BERT achieved the best results with an accuracy of 90%, although it required a significant amount of training time. On the other hand, CNN achieved acceptable results with a shorter training time.

This study offers an alternative model for predicting medical drug usage conditions, and intentions are successfully classified with accuracy values between 97% and 100%.

4. Conclusions

This study investigated the prediction of the use of drugs for different purposes than their main target condition using machine learning-based text classification from an online review dataset. The conclusions of the study are presented as follows:

- The proposed technique, SGD, TFIDF-Bigrams is effective for the prediction of the use of medical drugs for different conditions than they are developed to treat.
- Preprocessing (stemming, lemmatizing, tokenizing, removing stop words and punctuations, methods affected the performance positively and therefore, should be considered meticulously.
- It is seen from the results that using the bigram and the TF-IDF in combination improved the classification regardless of the classifier.
- The best result is obtained for the SGD – TF-IDF & Bigrams method by an accuracy value of 98.42%. In addition, owing to its high precision (0.98), recall (0.98), and F1 score (0.98), it can be concluded that it is a robust and responsive technique.
- Similar conditions are not recognized during classification. For instance, Acne and Birth Control, which are related to hormone treatment, are poorly classified. A similar situation is observed for Anxiety and Depression. Taking this conclusion into account may improve the model's accuracy in future works.
- The findings of this study may be beneficial in pharmaceutics or medicine considering drug design, reducing side effects, health management, treatment adherence and process design, and personalized medicine.

Declaration

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The author(s) also declared that this article is original, was prepared in accordance with international publication and research ethics, and ethical committee permission or any special permission is not required.

Author Contributions

Duygu Bagci Das developed the methodology, performed the analysis, supervised and improved the study, wrote the manuscript, and proofread the manuscript.

References

1. Şen Ö., S.Bozkurt Keser, and K. Keskin, *Early stage diabetes prediction using decision tree-based Ensemble Learning Model.* International Advanced Researches and Engineering Journal, 2023. **7**(1): p. 62-71.
2. Jahan S., *Major depressive disorder diagnosis from electroencephalogram data and potential treatment with dimethyltryptamine.* International Advanced Researches and Engineering Journal, 2023. **7**(2):p. 90–96.
3. Tuncer T., E. Aydemir, F. Özyurt, S. Dogan, S. B. Belhaouarı, and E. Akbal, *An automated COVID-19 respiratory sound classification method based on novel local symmetric Euclidean distance pattern and Relieff Iterative MRMR feature selector.* International Advanced Researches and Engineering Journal, 2021. **5**(3): p.334–343.
4. Liu R.-L., *Text classification for healthcare information support,* in 20th International Conference on Industrial, Engineering, and Other Applications of Applied Intelligent Systems. 2007. Kyoto, Japan: p. 44–53.
5. Billyan B., R. Sarno, K.R. Sungkono, and I.R. Tangkawarow, *Fuzzy k-nearest neighbor for restaurants business sentiment analysis on TripAdvisor*, in 2019 International Conference on Information and Communications Technology. 2019. Kuala Lumpur, Malaysia: p. 543-548.
6. Pratama B. Y. and R. Sarno, *Personality classification based on Twitter text using naive Bayes, KNN and SVM,* in 2015 International Conference on Data and Software Engineering (ICoDSE), 2015. Yogyakarta, Indonesia: p. 170-174.
7. Suela O-M., M. Zampieri, S. Malmasi, M. Vela, L.P. Dinu, and J. van Genabith [cited 2024 1 Jun]; Available from: https://arxiv.org/abs/1710.09306
8. Olsson J. S., D. W. Oard, and J. Hajič, *Cross-language text classification*, in Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, 2005. Salvador Brazil: p.645-646.
9. Li C.H. and S.C. Park, *An efficient document classification model using an improved back propagation neural network and singular value decomposition*. Expert Systems with Applications, 2009. **36**(2): p.3208–3215.
10. Tang B., H. He, P. M. Baggenstoss, and S. Kay, *A bayesian classification approach using class-specific features for text categorization.* IEEE Transactions on Knowledge and Data Engineering, 2016, **28**(6):p. 1602–1606.
11. Kumar R. R., M. B. Reddy, and P. Praveen, *Text Classification Performance Analysis on Machine Learning.* International Journal of Advanced Science and Technology, 2019. **28**(20): p. 691–697.
12. Shen P., H. Wang, Z. Meng, Z. Yang, Z. Zhi, R. Jin, and A. Yang, *An Improved Parallel Bayesian Text Classification Algorithm.* Review of Computer Engineering Studies, 2016. **3**(1): p. 6–10.
13. Dalal M. K. and M.A. Zaveri, *Automatic text classification: A technical review.* International Journal of Computer

Applications, 2011. **28**(2): p. 37–40.
14. Shokrpour N., R. Rezaee, R. Akbari, M. Nasiri, and F. Foroughinia, *An evaluation of classification algorithms for prediction of drug interactions: Identification of the best algorithm.* International Journal of Pharmaceutical Investigation, 2018. **8**(2): p. 92-99.
15. Chai Z., X. Wan, Z. Zhang, and M. Li, *Harvesting drug effectiveness from social media,* in Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019. Paris, France: p.55-64.
16. Aramaki E., Y. Miura, M. Tonoike, T. Ohkuma, H. Masuichi, K. Waki, and K. Ohe, *Extraction of adverse drug effects from clinical records.* Studies in health technology and informatics, 2010. **160**(1): p.739–743.
17. Gräßer F., S. Kallumadi, H. Malberg, and S. Zaunseder, *Aspect-based sentiment analysis of drug reviews applying cross-domain and cross-data learning,* in Proceedings of the 2018 International Conference on Digital Health, 2018. Lyon France: p.121-125.
18. Lavecchia A., *Machine-learning approaches in drug discovery: Methods and applications.* Drug Discovery Today, 2015. **20**(3): p.318–331.
19. Abada W., A. Bouramoul, and M. S. Kahil, *How machine learning algorithms can examine patterns in multiple substance addictions: Predicting the effects and interactions of psychoactive substances*. International Journal of Computers and Applications, 2024. **46**(11): p.1045–1055.
20. Elahi E., E. Elahi, S. Anwar, B. Shah, Z. Halim, A. Ullah, I. Rida, M. Waqas, *Knowledge graph enhanced contextualized attention-based network for responsible user-specific recommendation.* ACM Transactions on Intelligent Systems and Technology, 2024. **15**(4): p. 1–24.
21. Korkmaz T., A. Çetinkaya, H. Aydin, and M. A. Barişkan, *Analysis of whether news on the internet is real or fake by using deep learning methods and the TF-IDF algorithm.* International Advanced Researches and Engineering Journal, 2021. **5**(1): p. 31–41.
22. Zhang T., *Solving large scale linear prediction problems using stochastic gradient descent algorithms,* in Twenty-first international conference on Machine learning- ICML '04, 2004.
Banff, Alberta, Canada: p.116-124.
23. Al-Hadhrami S., T. Vinko , T. Al-Hadhrami, F. Saeed, and S.N. Qasem, *Deep learning-based method for sentiment analysis for patients' drug reviews*. PeerJ Computer Science, 2024. **10**: p. e1976.
24. Dandala B., V. Joopudi, and M. Devarakonda, Adverse *Drug Events Detection in clinical notes by jointly modeling entities and relations using Neural Networks.* Drug Safety, 2019. **42**(1): p.135–146.
25. Colón-Ruiz C. and I. Segura-Bedmar, *Comparing deep learning architectures for sentiment analysis on drug reviews.* Journal of Biomedical Informatics, 2020. **110**: p.103539.