Drug Prediction

by Techpower Solutions

Submission date: 27-Feb-2024 07:47AM (UTC-0500)
Submission ID: 2305804066
File name: PREDICTION_THE_DRUG_-_Paper_Work.docx (947.2K)
Word count: 4696
Character count: 28249

Prediction The Drug Rating Based on Reviews by Using Natural Language Processing

***Abstract*— The ability of Natural Language Processing (NLP) to analyze and comprehend human language has drawn a lot of interest in recent years. This research proposes an NLP-based drug evaluation system based on reviews. The idea is to develop a system that can automatically read and evaluate medication reviews, pull out pertinent data, and assign scores according to user sentiment. There are multiple phases to the suggested strategy. Initially, a collection of medicine reviews is gathered from multiple platforms, including social media, internet forums, and health websites. Reviews undergo pre-processing in order to manage linguistic difficulties, tokenize, and eliminate noise. The polarity of every review is ascertained by employing sentiment analysis tools It consists of assigning a favourable, negative, or neutral classification to the sentiment that is expressed in the passage. Numerous techniques, including Recurrent Neural Networks (RNN), Support Vector Machines (SVM), and Naive Bayes, can be utilized for this purpose.**

Key Words – Sentiment Analysis, Sentiment Score, Rating System, Natural Language Processing, Drug Evaluation, Reviews, and Health.

INTRODUCTION

The advent of user-generated content and the rapid expansion of internet platforms in recent years have made it possible to analyze and glean important insights from vast amounts of textual data in a way never before possible. Drug review analysis is an especially fascinating application, where users discuss different pharmaceuticals in internet forums and share their thoughts and experiences with them. By combining natural language processing (NLP) methods with these user-generated assessments, new avenues for comprehending the sentiments, emotions, and complex variables influencing drug ratings become accessible. The public and medical experts are largely informed on the effectiveness, negative effects, and general user satisfaction of different drugs by means of drug testing and evaluation. The public and medical experts are largely informed on the effectiveness, negative effects, and general user satisfaction of different drugs by means of drug testing and evaluation.

The real-world experiences of people using these treatments are not fully captured by traditional drug evaluation procedures, which frequently rely on expert judgments and clinical trials. On the other hand, the abundance of user-generated content on digital platforms presents a singular chance to leverage the combined knowledge of a wide range of users, offering a comprehensive and dynamic viewpoint on medication effectiveness.

The goal of this research is to predict medicine evaluations based on rich textual data found in user reviews by analyzing and using advanced natural language processing algorithms. From these reviews, we extracted important patterns, attitudes, and contextual data in an effort to create a prediction model that would offer insights into the variables affecting drug ratings. This research's interdisciplinary approach combines machine learning, computational linguistics, and health analytics to produce a potent tool for

gauging and forecasting public opinion toward medications.

By analysing and forecasting drug evaluations using rich textual data from user reviews, this study will employ sophisticated natural language processing techniques. We tried to create a prediction model that could offer insights into the variables impacting drug ratings by taking significant patterns, attitudes, and contextual data out of these reviews. Because of its interdisciplinary nature, this research offers a potent tool for analysing and forecasting public opinion toward medications by combining computational linguistics, machine learning, and health analytics.

Through the demonstration of how natural language understanding may be used to extract valuable insights from the collective voice of patients, this research marks a significant advancement in the utilization of NLP in the healthcare domain. The objective is to enhance the quality of healthcare decision-making by promoting a more knowledgeable and data-driven approach to drug evaluation by delving into the subtleties of drug reviews.

I. LITERATURE SURVEY

2.1 Sairamvinay Vijayaraghavan Debraj Basu .et.al One crucial algorithm in natural language processing called sentiment analysis is used to identify the sentiment present in a given text. For our experiment, we decided to examine text-based reviews of different medications that also included a rating on a range of 1 to 10. This data set was acquired from the UCI machine learning repository, which included the train and test (split 75–25%) data sets. In general, we had divided the drug's numerical rating into three classes: neutral (4–7), negative (1-4), and positive (7–10). We looked into how different conditions' reviews, which use different language, affect the drugs' ratings because there are several reviews for the same medication. Our primary goal was to apply supervised machine learning classification algorithms that use the textual review to predict the rating class. The main embeddings that we had used were Count Vectors (CV) and Term Frequency Inverse Document Frequency (TFIDF). Our models were trained on the most common conditions found in the data set, such as "birth control," "depression," and "pain," and we were able to predict the test data sets with good accuracy.

2.2 Julia Chan .et.al Drug safety and risk-benefit characteristics must be understood in order to identify adverse drug reactions (ADRs). ADRs are a significant contributor to patient morbidity, accounting for 5–10% of acute care hospitalizations globally and maintaining a constant incidence over the previous 30 years.Adverse drug reactions (ADRs) are typically reported voluntarily. However, the significant underreporting rates linked to this approach provide a challenge to pharmacovigilance initiatives. Automated adverse reaction reporting may be less helpful if other drug-related adverse events are overreported, yet this nevertheless gives rise to another way to boost reporting rates. We created a deep learning natural language processing system to identify ADRs in discharge summary reports for a university hospital center. We created our model in two stages: first, we used 150,000 training data points to further train a pre-trained model (DeBERTa).We used this model to fit 861 annotated discharge summaries and a corpus of unlabeled discharge summaries in order to identify AMR remarks. To make sure that our algorithm can differentiate ADRs from other drug-related adverse events, the annotated corpus is enriched for both validated ADR reports and confounding drug-related adverse events. With a ROC-AUC of 0.934 (95% CI: 0.931 - 0.955) for the task of detecting discharge summaries incorporating ADR signals, the final model demonstrated strong performance. [2].

2.3 Sharon Fong Mei Toh .et.al A rising alternative for self-paced rehabilitation is the use of wearable technology at home. Its application as a home stroke rehabilitation treatment has not been thoroughly studied. The objectives of this review were to: 1) delineate wearable technology-based interventions in stroke patients' at-home

physical rehabilitation; and 2) demonstrate the efficacy of wearable technology as a therapeutic option. We conducted a thorough search of the Cochrane Library, Medline, SinaHL, and Web of Science electronic databases to find articles published between the beginning and February of 2022. The Arksey and O'Malley paradigm was practically used in this scoping research. Studies were evaluated and chosen by two impartial reviewers. For this review, twenty-seven were chosen. The degree of evidence was evaluated and a descriptive analysis was performed on these studies [3].

2.4 Sharon Fong Mei Toh .et.al The majority of studies, according to this review, have concentrated on enhancing hemiparetic upper extremity (UL) function, with relatively few investigating the application of wearable technology in lower extremity (LL) rehabilitation that takes place at home. Wearable technology interventions include robotic therapy, virtual reality (VR), stimulus-based training, and activity monitors. We found "moderate" evidence for activity monitors, "limited" evidence for virtual reality, "inconsistent evidence" for robotic training, and "strong" evidence for stimulus-based training among UL treatments. With so few research conducted, the effects of wearable LL technologies are "very limited." The field of wearable robotics is one where new technology will lead to exponential growth in research. Subsequent investigations could concentrate on pinpointing elements of LL rehabilitation that can be efficiently tackled using wearable technologies [4].

2.5 Kona Sravya.et.al In making decisions, an advisory framework can offer a multitude of well-informed solutions with unclear facts and assist in determining the client's needs. Since customer-generated material is considered to contain human language in multiple ways, sentiment research is suggested to be an unreliable test by all accounts. Exams frequently concentrate on uninteresting topics like movies, coffee shops, and electrical problems, but not enough on health and wellness-related topics. Understanding healthcare in general and the drugs individuals take in particular can provide in-depth understandings of how to make the best choice possible while keeping the public interest in mind. We also want to apply the Framework program for medicine prescribers in this endeavor to conduct experimental advancements including medication audits. Creating a flexible support system to assist patients in making better decisions about prescription medicine prescriptions was the aim of this experiment. Initially, we provide an unusual method of assessment to deal with drug research and create drug evaluations. Furthermore, we are aware that medication audits can be helpful for clients, patient circumstances, and medication opinion survey participation. At that stage, we include those elements into a framework for a proposal that lists suitable drugs. In order to estimate the age and the proposed hybrid model in light of a given open data set, experiments have been carried out utilizing decision trees, K nearest neighbors, and linear support vector classifiers. To achieve the most crucial processing and modify the restrictions, every calculation is examined. For a very long period, the linear support vector classifier was the method of choice for age estimation since it provided a reasonable balance between model diversity, model accuracy, and model efficiency.[5].

2.6 Md. Ferdous Bin Hafiz .et.al Sentiment analysis, or the analysis of meaning regarding a certain topic from plain text, is one of the most recently developed subdimensions of natural language processing. Sentiment analysis of medication is very important these days since it can assist future consumers make better judgments by providing information about a particular medication and helping to rank medications based on effectiveness through the study of user reviews. The aim of this proposed study is to quantify a specific drug's level of effectiveness. The majority of text mining research currently uses unsupervised machine learning techniques to group data. The primary goal when using supervised learning techniques for text mining is often to divide the data into two groups.

Classification is made considerably more difficult by the absence of technical words in similar data sets. To improve the accuracy of categorizing medications according to their efficacy using various algorithms, the suggested research focuses on keyword recognition utilizing tokenization and lemmatization. Improving health and well-being and curing illnesses can both benefit from this classification. The UCI machine learning repository's medication review dataset was subjected to binary classification using four different machine learning techniques. Support vector classifier (SVC), naive Bayes classifier, random forest, and multilayer perceptron are some of the machine learning methods used for binary classification; linear SVC was one of these algorithms used for multi-class classification.[6].

II. METHODOLOGY

A. Problem Statement

All functional flaws must be removed in order to produce mobile apps that are successful. User evaluations for mobile apps are crucial for developers to discover hidden issues, since limited time and resources may cause usability testing to overlook some functional faults. Regretfully, manually handling each revision takes a lot of time and is prone to mistakes. Because the majority of existing research on mobile app reviews has been on review classification, requirements engineering, feelings, and abstraction, it has not been possible for developers to efficiently detect complex code based on reviews. They searched the program code but could not identify the functional faults mentioned in customer evaluations. Furthermore, current research on mapping revisions to complicated source files searches for matches between words in the revisions and words in the source code, error reports, messages, and stack traces—words that are ignored and can result in false positives and false negatives. Take Account. Every word has a part-of-speech label and a semantic meaning. We present a novel method in this research for localizing operational problems in mobile applications by leveraging context data from user reviews and establishing semantic connections between reviews and bytecodes.

Drawbacks of System

- The process of implementation is not facilitated by them.
- They just compare algorithms to determine which one has the highest accuracy.
- They were unable to determine the degree of confusion and effectiveness.
- They don't produce confusing measurements, charts, or performance graphs.
- There is a low level of accuracy and efficiency.

B. Proposed System

Natural language processing (NLP) is the basis of the suggested method for drug evaluation based on reviews, which offers a number of upgrades and improvements over the current setup. These are the key components of the suggested system. Better Data Collection: A greater variety of sources, including patient surveys, clinical trial data, and electronic medical records, are included in the proposed system's expanded data collection process. The system attempts to acquire comprehensive and representative drug reviews by merging data from several sources. Advanced preprocessing: To address issues unique to medication reviews, including typos, medical jargon, and abbreviations, the suggested solution employs advanced preprocessing techniques. To ensure meaningful representation of drug-related information and to increase the accuracy of text pre-processing, domain-specific dictionaries or ontologies can be employed.

Micro-Sentiments Analysis: This proposed framework incorporates the analysis of micro-sentiments in an effort to surpass a basic binary classification of positive/negative sentiments. Take note of subtle remarks made in reviews on the degree of side effects, the medication's effectiveness for a certain ailment, or the satisfaction level with drug interactions. This makes it possible to use a more complex and insightful approach of evaluating quality.

Benefits of the system:

- To improve accuracy, we compare more than two algorithms.
- We have accurately determined the significance of measures related to confusion and efficiency.
- We have produced performance and confusion metrics graphs and charts.
- We raise the bar for efficiency and precision.
- We appropriately apply machine learning.
- We use the data to systematically train the machine. 70% or 80% of the train data

III.ARCHITECTURE

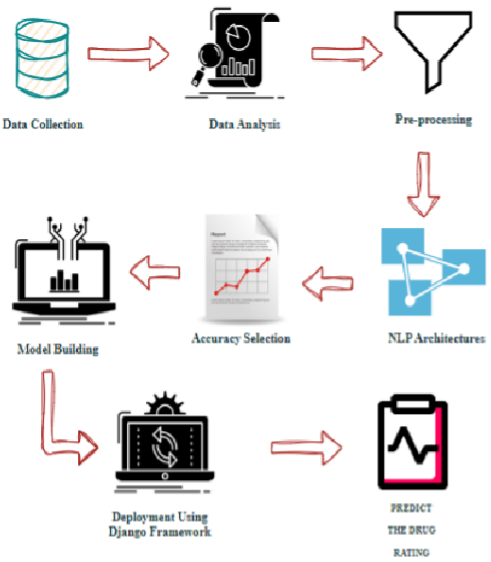


Figure 1: Machine learning techniques are used in a natural language processing (NLP) framework to predict drug ratings by preprocessing text input, extracting pertinent features, and modeling connections between reviews and ratings. By employing sophisticated natural language processing (NLP) methods, the system offers a reliable and effective way to forecast drug ratings based on textual content analysis and user feelings.[1]

IV.DISCUSSION

It has important functions like predicting medicine ratings from reviews, evaluating comments, and employing natural language processing (NLP) to reliably predict model and user ratings from text data. Personalized healthcare decision-making is aided by the application of natural language processing (NLP), which enables a detailed comprehension of user feelings and offers significant insights into aspects influencing drug ratings. The advancement of natural language processing (NLP) methods can enhance the efficacy of assessments in predictive medicine and lead to better clinical judgments for patients and healthcare professionals.

System Modules:

- Module 1 : Information Gathering
- Module 2 : Pre-processing of Data
- Module 3 : Information projection
  - Random Forest
  - Decision Tree
  - Gradient Boosting

  Algorithm
- Module 4: Deployment Using DJANGO

Module 1: Information Gathering:

Data gathering is the first step in the procedure. In this step, pertinent data must be gathered from a variety of sources, including databases, external data sets, APIs, and sensors. Data annotation, such as categorizing text or tagging photos, may also be a part of data collecting.

Module 2: Pre-processing of Data :

Validation approaches are used in machine learning to extract an error rate from an ML model, which is thought to be almost exactly the same as the actual error rate of the data set. When data sizes are sufficiently large to accurately represent the population, validation techniques are

not required. In real-world scenarios, it is crucial to work with data samples that do not fairly represent the population in a particular data collection. To locate the missing value, whether it be an integer or a float variable, duplicate the value and describe the data type. While modifying the model's hyperparameters, a data model is utilized to offer an objective assessment of a model's fit to a training data set.The flexibility of the estimation increases when the validation data set's performance is incorporated into the model architecture. A regularly updated model validation set is used for evaluation. This data is used by machine learning developers to modify the upper boundaries of the model.Managing the content, quality, and structure of data as well as gathering and evaluating data can become into laborious to-do lists. We assist you in comprehending your data and its properties during the data identification phase; this understanding will help you decide which algorithm to utilize to construct your model.

The Python Pandas library is used for many data cleaning tasks, particularly since it can quickly clean the data and is focused on huge data cleaning tasks without values. More time should be dedicated to data cleansing, analysis, and modeling.

Random errors account for part of these sources. In some cases, missing data might have a deeper cause. It is crucial to comprehend these various kinds of missing data from a statistical standpoint. The kind of missing data, how to handle filling in the gaps and detecting missing values, how to do some simple imputation, and a thorough statistical method for handling missing data are all covered.

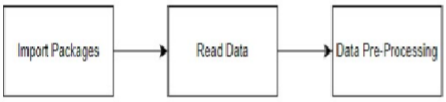


Figure 2: In the data preprocessing phase, which uses natural language processing to predict drug ratings based on reviews, textual data is subjected to tasks such as tokenization and deblocking to convert it into a format suitable for analysis. Furthermore, normalization and word removal techniques are used to improve the quality and relevance of the features extracted from the reviews. [2]

Module 3: Information Projection:

The visualization of data is a key component of machine learning and applied statistics. Actually, quantitative analysis and data evaluations are the primary areas of study for statistics. Data visualization offers valuable tools for obtaining qualitative insights. When examining a data set, it can assist you in identifying trends, skewed data, outliers, and more. Plots and diagrams that are visceral and compelling, as opposed to those that are connected or significant to a limited subject knowledge base, might benefit from the usage of data visualizations to emphasize and illustrate essential linkages. I recommend devouring the books I mentioned at the conclusion to learn more about exploratory data analysis and data visualization.

Without being displayed in a visual format, such graphs and charts, data can occasionally lack meaning. In applied statistics and applied machine learning, the ability to rapidly visualize data and other patterns is crucial. You will learn about the many kinds of layers that you need to be aware of when using Python to visualize data and how to use them to have a deeper understanding of your own data.

How to use line and bar graphs to interpret time series data with categorical measurements.

How to use boxplots and histograms to summarize data distributions.

Figure 3: Using natural language processing, an information scheme is developed to predict drug ratings from reviews. It maps important characteristics from textual data into a numerical representation, which makes it

possible to identify patterns and comprehend the underlying factors that affect drug ratings. This procedure makes sure that important information from reviews is transferred into a format that machine learning algorithms can comprehend, which increases the predictive model's accuracy[3].

Random Forest Algorithm:

Use natural language processing (NLP) and random forest techniques to predict medicine ratings based on reviews by pre-processing text input, extracting pertinent features, and building a random forest model. In order to precisely forecast medication evaluations, the system examines attitudes and trends seen in reviews. This strategy makes use of NLP techniques and ensemble learning to create a strong model that can handle the intricacy of textual input, offering an efficient way to forecast drug ratings.

Decision Tree Algorithm:

Our goal in this research is to use Natural Language Processing (NLP) and Decision Tree Algorithm to predict medicine ratings based on reviews. By removing pertinent elements from the text, we preprocess and examine the textual reviews. We create a predictive model that discovers patterns and connections between extracted data and drug ratings using a decision tree algorithm. This method explains how particular words or phrases affect drug ratings by using the natural structure of decision trees to create predictions. Our goal is to create a practical and comprehensible model for drug rating prediction from textual reviews by fusing natural language processing and decision trees.

GBC:

Utilizing a gradient boosting technique in natural language processing (NLP), medicine evaluations are predicted from reviews. This method makes use of NLP to extract information from textual material, and the gradient boosting strategy combines several weak learners to increase prediction accuracy. The goal of this approach is to tie pertinent drug evaluations to language patterns seen in reviews.

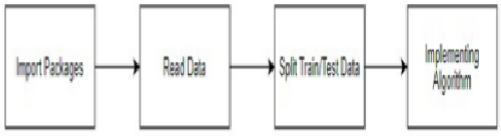


Figure 4: Algorithmic processing to predict drug ratings from reviews using natural language processing uses machine learning techniques such as sentiment analysis and feature extraction to train a model on pre-processed text data. By using algorithms such as Decision Tree and Random Forest, the system efficiently learns patterns in reviews, enabling accurate predictions of drug ratings based on identified sentiments and related features. [4]

Module 4: Establishing Using DJANGO

In order to improve user interface and forecast whether an image is CKD or not, the deep learning model learned in this module is transformed into a Hierarchical Data Format file (.h5 file) and utilized in our Django framework.

Django

A safe and readily managed website may be created quickly with the help of the high-level Python web framework Django. The majority of web development is handled by Django, which was created by seasoned developers, freeing you up to concentrate on refining your application rather than beginning from scratch. It features a vibrant and active community, great documentation, a ton of free and paid support options, and is open source and free.

You can develop software with Django that: I'll complete it.

Django implements the "batteries included" philosophy and provides all the functionalities developers require "out of the box." All you need is contained in a single "product," which means that it all works together as a cohesive unit, satisfies industry standards for design, and comes with up-to-date, thorough documentation.

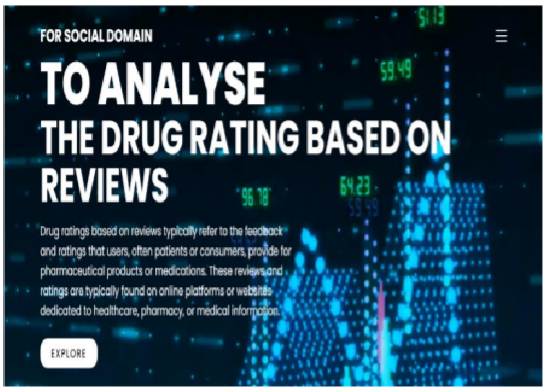
Flexible

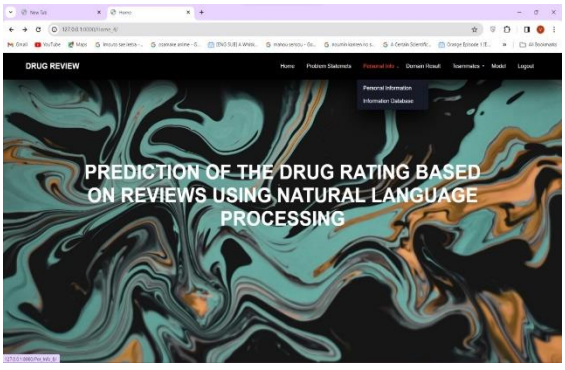
Any type of website, including social networks, news sites, wikis, and content management systems, may be built using Django. It can provide material in any format (HTML, RSS feeds, JSON, XML, etc.) and interact with any client-side framework. You are currently on a Django-built website!

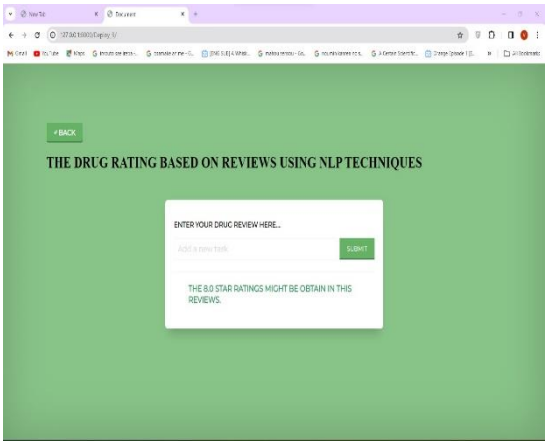
It can be expanded to incorporate more components if necessary, and it offers options for the internal functionality you choose (such as several well-known databases, templating engines, etc.).

V.RESULTS

The efficacy of natural language processing (NLP) models in extracting pertinent information and interpreting sentiment is demonstrated by their ability to predict medicine evaluations from user reviews with accuracy. Through the examination of textual data, the model pinpoints crucial attributes that impact drug assessments, thereby advancing a refined comprehension of drug users' encounters. By using user feedback to forecast drug ratings more accurately, this strategy helps healthcare providers make well-informed decisions.







VI. CONCLUSION

Finally, a revolutionary method for drug rating prediction through review leveraging natural language processing (NLP) is presented. Through the process of identifying important emotions, themes, and background data from user comments, the approach offers a sophisticated comprehension of medication efficacy and user happiness. By utilizing cutting-edge natural language processing (NLP) techniques, a solid model that can precisely forecast drug valuations can be created, enhancing pharmaceutical and healthcare industry decision-making processes. This novel method has the potential to completely change how we assess and comprehend the efficacy of medications based on actual user experiences.

VII. FUTURE ENRICHMENT

Our goal for the future is to create a reliable natural language processing (NLP) model that can be used to forecast medicine ratings according to user feedback. By utilizing sophisticated machine learning methods, we construct a predictive model by examining sentiment, important terms, and contextual details in reviews. Our methodology leverages advanced natural language processing (NLP)

techniques to enhance the precision of drug rating forecasts by obtaining valuable insights from unstructured textual data. Furthermore, our intention is to investigate deep learning techniques in order to enhance the model's functionality and flexibility with respect to various data sets.

VIII. REFERENCES

[1] World Health Organization. 2020. Pharmacovigilance. [online] Available at: https://www.who.int/teams/regulationprequalification/regulation-and-safety/pharmacovigilance [Accessed 12 April 2022].

[2] Pharmacologyeducation, 2022. [Online]. Available: https://www.pharmacologyeducation.org/clinical pharmacology/adverse-drug-reaction. [Accessed: 31- Aug- 2022].

[3] A. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media", Business Horizons, vol. 53, no. 1, pp. 59-68, 2010. Available: 10.1016/j.bushor.2009.09.003.

[4] R. Alt & M .Wittwer,. "Towards an ontology-based approach for social media analysis", In Proceedings of 22 European Conference on Information Systems (ECIS 2014), 2014.

[5] PatientsLikeMe, 2022. [Online]. Available: https://www.patientslikeme.com/. [Accessed: 30- Aug- 2022].

[6] Drugs.com,2022. [Online]. Available: https://www.drugs.com/. [Accessed: 30- Aug- 2022].

[7] blog.23andme.com, 2022. [Online]. Available: https://blog.23andme.com/. [Accessed: 30- Aug- 2022].

[8] M. Wiley, C. Jin, V. Hristidis and K. Esterling, "Pharmaceutical drugs chatter on Online Social Networks", Journal of Biomedical Informatics, vol. 49, pp. 245-254, 2014. Available: 10.1016/j.jbi.2014.03.006.

[9] Z.Min," Drugs Reviews Sentiment Analysis using Weakly Supervised Model". In 2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA),pp. 332-336, 2019.

[10] F.Gräßer, S.Kallumadi, H.Malberg, & S.Zaunseder," Aspect-based sentiment analysis of drug reviews applying cross-domain and crossdata learning". In Proceedings of the 2018 International Conference on Digital Health , pp. 121-125, 2018.

[11] UCI Machine Learning Drug Review dataset. Kaggle.com, from https://www.kaggle.com/jessicali9530/kuc-hackathon-winter-2018. [Accessed 12 April 2022].

[12] UCI Machine Learning Repository: Drug Review Dataset (Drugs.lib) Data Set. Archive.ics.uci.edu. from https://archive.ics.uci.edu/ml/datasets/Drug+Review+Dataset+%28Dr uglib.com%29, [Accessed 12 April 2022].

[13] Y. Han, M. Liu and W. Jing, "Aspect-Level Drug Reviews Sentiment Analysis Based on Double BiGRU and Knowledge Transfer", IEEE Access, vol. 8, pp. 21314-21325, 2020. Available: 10.1109/access.2020.2969473.

[14] S.Garg, " Drug recommendation system based on sentiment analysis of drug reviews using machine learning". In the proceedings of 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) ,pp. 175-181. IEEE, 2021.

[15] J. Ramos et al., "Using TF-IDF to determine word relevance in document queries," in Proceedings of the first instructional conference on machinelearning, vol. 242, pp. 133–142, Piscataway, NJ, 2003.

[16] Yoav Goldberg and Omer Levy. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method, 2014; arXiv:1402.3722.

[17] M.N.Uddin, M.F.B. Hafiz,S.Hossain& S.M.M. Islam. "Drug Sentiment Analysis using Machine Learning Classifiers". International

Journal of Advanced Computer Science and Applications, vol 13, no.1, 2022.

[18] Sztandera, "Self-augmenting Knowledge Base for Informed Decision Making With Biomedical Applications in Cancer Diagnosis", WSEAS TRANSACTIONS ON CIRCUITS AND SYSTEMS, vol. 19, pp. 69-74, 2020. Available: 10.37394/23201.2020.19.8.

[19] D. Koppad, "Arrhythmia Classification Using Deep Learning: A Review", WSEAS TRANSACTIONS ON BIOLOGY AND BIOMEDICINE, vol. 18, pp. 96-105, 2021. Available: 10.37394/23208.2021.18.11.

[20] S.Loria, P.Keen, M. Honniba, R.Yankovsky, D. Karesh, , E.Dempsey, "Textblob: simplified text processing.", Secondary TextBlob: simplified text processing, 3, 2014.

[21] WordNet | A Lexical Database for English. Wordnet.princeton.edu., from https://wordnet.princeton.edu/.[Accessed 12 April 2022] [22] J. Friedl, Mastering Regular Expressions, 3rd Edition. [S.l.]: O'Reilly Media, Inc., 2006.

[23] pandas - Python Data Analysis Library. Pandas.pydata.org. (2022). Retrieved 29 July 2022, from https://pandas.pydata.org/.

[24] S.Tongman & N.Wattanakitrungroj , "Classifying Positive or Negative Text Using Features Based on Opinion Words and Term Frequency-Inverse Document Frequency". In 2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA), pp. 159-164, IEEE,2018.

[25] F.Pedregosa, G.Varoquaux, A.Gramfort, V.Michel, B.Thirion, , O.Grisel, M.Blondel, P. Prettenhofer, R.Weiss, , V.Dubourg, J.Vanderplas, A.Passos, & D.Cournapeau, "Scikit-learn: Machine learning in Python", the Journal of machine Learning research, vol 12, pp.2825-2830, 2011

[26] A. M Kibriya, E.Frank, B. Pfahringer & G. Holmes,. 2004,"Multinomial naive bayes for text categorization revisited". In Australasian Joint Conference on Artificial Intelligence, pp. 488-499, Springer, Berlin, Heidelberg,2004.

[27] J. Brownlee," Master Machine Learning Algorithms: discover how they work and implement them from scratch", Machine Learning Mastery. Chapter 13: Logistic Regression, pp.51-55, 2016a.

[28] T. Hastie, T.Robert, and F. Jerome," The elements of statistical learning: data mining, inference, and prediction". Springer Science & Business Media, 2009.

[29] J. Brownlee,"Master Machine Learning Algorithms: discover how they work and implement them from scratch", Machine Learning Mastery. Chapter 28: Bagging and Random Forest, pp.126-129, 2016b

[30] G.Biau, 2012. Analysis of a random forests model. The Journal of Machine Learning Research, vol 13, no.1, pp.1063-1095,2012.

Drug Prediction

ORIGINALITY REPORT

7%	5%	6%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.researchgate.net Internet Source	2%
2	Sharon Fong Mei Toh, Kenneth N. K. Fong, Pablo Cruz Gonzalez, Yuk Ming Tang. "Application of Home-Based Wearable Technologies in Physical Rehabilitation for Stroke: A Scoping Review", IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023 Publication	1%
3	Mariam Gawich, Marco Alfonse. "A Proposed Model for Drugs' Review Analysis and Adverse Drug Reaction Discovery", 2022 7th International Conference on Mathematics and Computers in Sciences and Industry (MCSI), 2022 Publication	1%
4	thesai.org Internet Source	<1%
5	www.icicel.org Internet Source	<1%

6	Submitted to South Bank University Student Paper	<1%
7	odr.chalmers.se Internet Source	<1%
8	www.diva-portal.org Internet Source	<1%
9	www.scitepress.org Internet Source	<1%
10	Hasan Faraz Khan, Noor Fatima. "Analyzing Customer Sentiment in Drug Reviews Using Natural Language Processing", 2023 7th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech), 2023 Publication	<1%
11	K. O. Vedasree Anusha, Akanksha Dhar, Shashwat Dixit, Amay Saraf, Ishaan Lonial, Anusha N. "Automated Personality-Based Candidate Shortlisting using Machine Learning and Natural Language Processing", 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), 2023 Publication	<1%
12	M. Dhilsath Fathima, R. Hariharan, S. P. Raja. " Multiple Imputation by Chained Equations– - Nearest Neighbors and Deep Neural Network	<1%

	Architecture for Kidney Disease Prediction ", International Journal of Image and Graphics, 2021 Publication	
13	Rizwan Qureshi, Muhammad Irfan, Taimoor Muzaffar Gondal, Sheheryar Khan et al. "AI in Drug Discovery and its Clinical Relevance", Heliyon, 2023 Publication	<1%
14	aclweb.org Internet Source	<1%
15	hdl.handle.net Internet Source	<1%
16	libweb.kpfu.ru Internet Source	<1%
17	machine-learning-course.readthedocs.io Internet Source	<1%
18	Kerstin Denecke. "Sentiment Analysis in the Medical Domain", Springer Science and Business Media LLC, 2023 Publication	<1%
19	Christopher McMaster, Julia Chan, David FL Liew, Elizabeth Su, Albert G Frauman, Wendy W Chapman, Douglas EV Pires. "Developing A Deep Learning Natural Language Processing Algorithm For Automated Reporting Of	<1%

Adverse Drug Reactions", Cold Spring Harbor Laboratory, 2021

Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		