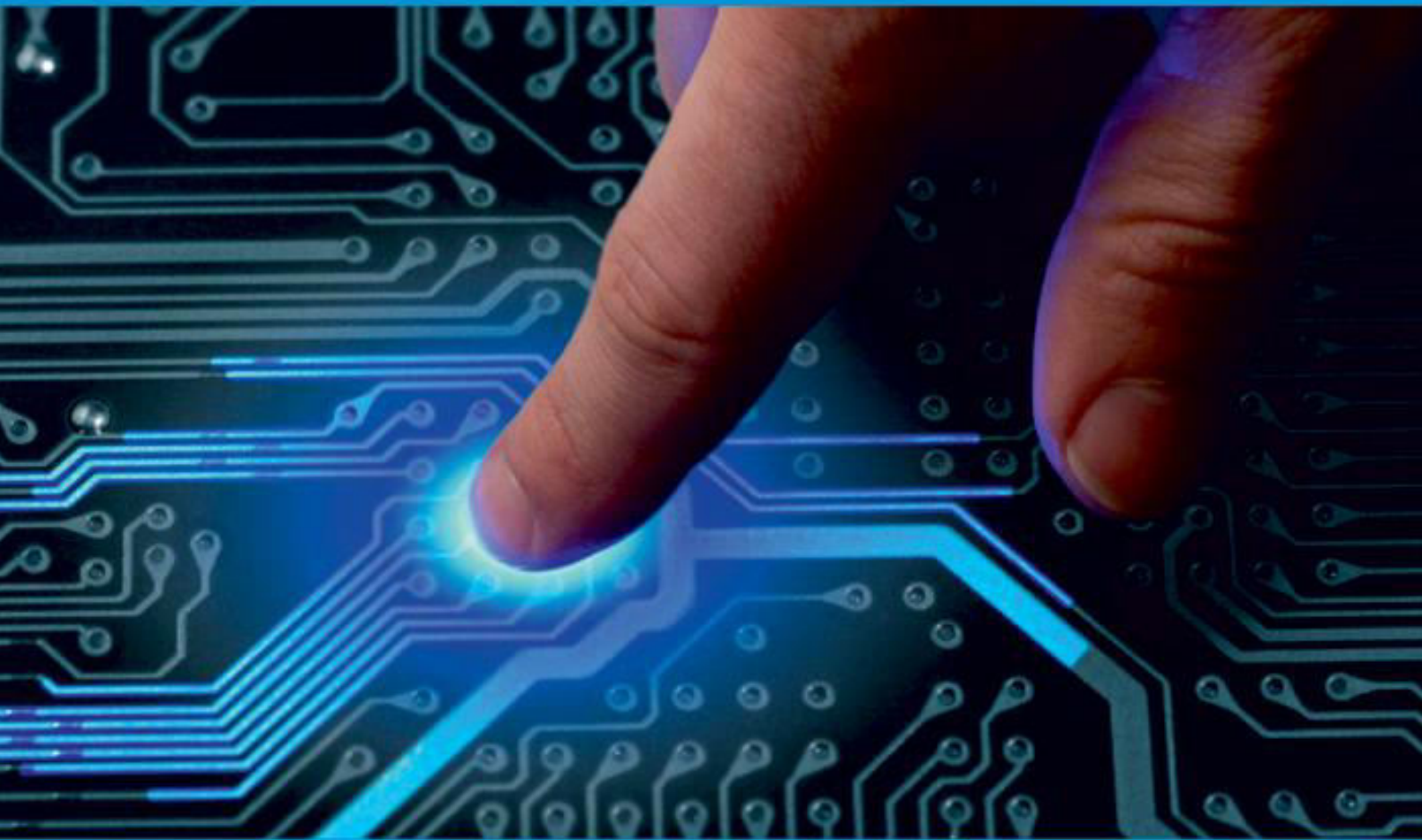




**IJIRCCCE**

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 12, December 2022

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 8.165**

 9940 572 462

 6381 907 438

 [ijircce@gmail.com](mailto:ijircce@gmail.com)

 [www.ijircce.com](http://www.ijircce.com)

# Pharmaceutical Guidance System Based on Sentimental Analysis of Chemist Reviews Using Machine Learning

G.Dharani, Dr.G.Singaravel M.E.,Ph.D, Mr.S.Gowtham M.E.,

Department of Information Technology, KSR College of Engineering, Tiruchengode, India

Professor & Head, Department of Information Technology, KSR College of Engineering, Tiruchengode, India

Assistant Professor, Department of Information Technology, KSR College of Engineering, Tiruchengode, India

**ABSTRACT:** Since coronavirus has shown up, inaccessibility of legitimate clinical resources is at its peak, like the shortage of specialists and healthcare workers, lack of proper equipment and medicines etc. The entire medical fraternity is in distress, which results in numerous individual's demise. Due to unavailability, individuals started taking medication independently without appropriate consultation, making the health condition worse than usual. As of late, machine learning has been valuable in numerous applications, and there is an increase in innovative work for automation. This paper intends to present a drug recommender system that can drastically reduce specialists heap. In this research, we build a medicine recommendation system that uses patient reviews to predict the sentiment using various vectorization processes like Bow, TF-IDF, Word2Vec, and Manual Feature Analysis, which can help recommend the top drug for a given disease by different classification algorithms. The predicted sentiments were evaluated by precision, recall, f1score, accuracy, and AUC score. The results show that classifier Linear SVC using TF-IDF vectorization outperforms all other models with 93% accuracy.

## I. INTRODUCTION

With the number of coronavirus cases growing exponentially, the nations are facing a shortage of doctors, particularly in rural areas where the quantity of specialists is less compared to urban areas. A doctor takes roughly 6 to 12 years to procure the necessary qualifications. Thus, the number of doctors can't be expanded quickly in a short time frame. A Telemedicine framework ought to be energized as far as possible in this difficult time. Clinical blunders are very regular nowadays. Over 200 thousand individuals in China and 100 thousand in the USA are affected every year because of prescription mistakes. Over 40% medicine, specialists make mistakes while prescribing since specialists compose the solution as referenced by their knowledge, which is very restricted. Choosing the top level medication is significant for patients who need specialists that know wide-based information about microscopic organisms, antibacterial medications, and patients. Every day a new study comes up with accompanying more drugs, tests, accessible for clinical staff every day. Accordingly, it turns out to be progressively challenging for doctors to choose which treatment or medications to give to a patient based on indications, past clinical history. With the exponential development of the web and the web-based business industry, item reviews have become an imperative and integral factor for acquiring items worldwide.

Individuals worldwide become adjusted to analyze reviews and websites first before settling on a choice to buy a thing. While most of past exploration zeroed in on rating expectation and proposals on the E-Commerce field, the territory of medical care or clinical therapies has been infrequently taken care of. There has been an expansion in the number of individuals worried about their well-being and finding a diagnosis online. As demonstrated in a Pew American Research center survey directed in 2013, roughly 60% of grown-ups searched online for health-related subjects, and around 35% of users looked for diagnosing health conditions on the web. A medication recommender framework is truly vital with the goal that it can assist specialists and help patients to build their knowledge of drugs on specific health conditions. A recommender framework is a customary system that proposes an item to the user, dependent on their advantage and necessity.

These frameworks employ the customers' surveys to break down their sentiment and suggest a recommendation for their exact need. In the drug recommender system, medicine is offered on a specific condition dependent on patient reviews using sentiment analysis and feature engineering. Sentiment analysis is a progression of strategies, methods, and tools for distinguishing and extracting emotional data, such as opinion and attitudes, from language. On the other hand, Featur engineering is the process of making more features from the existing ones; it improves the performance

of models. This examination work separated into five segments: Introduction area which provides a short insight concerning the need of this research, Related works segment gives a concise insight regarding the previous examinations on this area of study, Methodology part includes the methods adopted in this research, The Result segment evaluates applied model results using various metrics, the Discussion section contains limitations of the framework, and lastly, the conclusion section.

## II. LITERATURE SURVEY

### 1. Drug Recommendation System based on Sentiment Analysis of Drug Reviews using Machine Learning

**Author: SatvikGarg**

With a sharp increment in AI advancement, there has been an exertion in applying machine learning and deep learning arXiv:2104.01113v2 [cs.LG] 5 Apr 2021 strategies to recommender frameworks. These days, recommender frameworks are very regular in the travel industry, e-commerce, restaurant, and so forth. Unfortunately, there are a limited number of studies available in the field of drug proposal framework utilizing sentiment analysis on the grounds that the medication reviews are substantially more intricate to analyze as it incorporates clinical wordings like infection names, reactions, a synthetic names that used in the production of the drug [8]. The study [9] presents GalenOWL, a semantic-empowered online framework, to help specialists discover details on the medications. The paper depicts a framework that suggests drugs for a patient based on the patient's infection, sensitivities, and drug interactions. For empowering GalenOWL, clinical data and terminology first converted to ontological terms utilizing worldwide standards, such as ICD-10 and UNII, and then correctly combined with the clinical information.

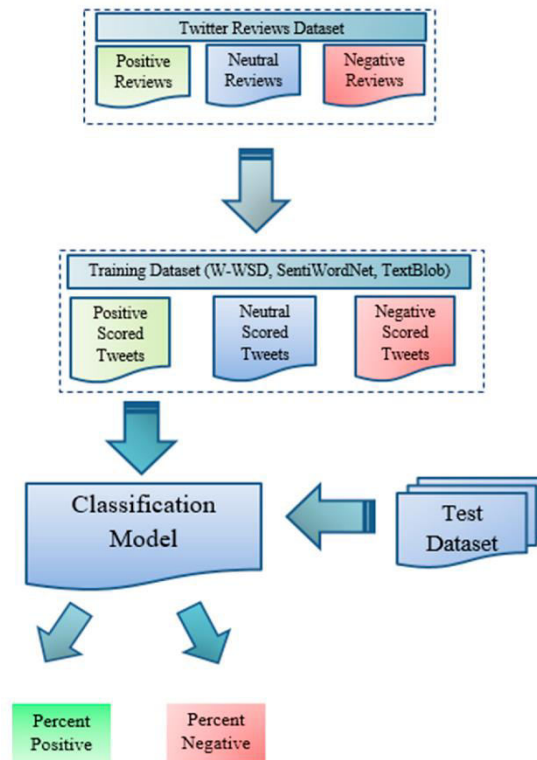
### 2. Examined Large Scale Treatment Records To Locate The Best Treatment Prescription For Patients.

**Author: Leilei Sun**

The idea was to use an efficient semantic clustering algorithm estimating the similarities between treatment records. Likewise, the author created a framework to assess the adequacy of the suggested treatment. This structure can prescribe the best treatment regimens to new patients as per their demographic locations and medical complications. An Electronic Medical Record (EMR) of patients gathered from numerous clinics for testing. The result shows that this framework improves the cure rate. In this research, multilingual sentiment analysis was performed using Naive Bayes and Recurrent Neural Network (RNN). Google translator API was used to convert multilingual tweets into the English language. The results exhibit that RNN with 95.34% outperformed Naive Bayes, 77.21%. The study is based on the fact that the recommended drug should depend upon the patient's capacity. For example, if the patient's immunity is low, at that point, reliable medicines ought to be recommended. Proposed a risk level classification method to identify the patient's immunity. For example, in excess of 60 risk factors, hypertension, liquor addiction, and so forth have been adopted, which decide the patient's capacity to shield himself from infection. A web-based prototype system was also created, which uses a decision support system that helps doctors select first-line drugs. Xiaohong Jiang et al. examined three distinct algorithms, decision tree algorithm, support vector machine (SVM), and backpropagation neural network on treatment data.

## III. METHODOLOGIES

The dataset used in this research is Drug Review Dataset (Drugs.com) taken from the UCI ML repository. This dataset contains six attributes, name of drug used (text), review (text) of a patient, condition (text) of a patient, useful count (numerical) which suggest the number of individuals who found the review helpful, date (date) of review entry, and a 10-star patient rating (numerical) determining overall patient contentment. It contains a total of 215063 instances shows the proposed model used to build a medicine recommender system. It contains four stages, specifically, Data preparation, classification, evaluation, and Recommendation.



(b)

#### IV. EXISTING SYSTEM

Recommender frameworks point to supply clients with personalized stock and repair to alter the expanding online information over-burden drawback. Various recommender framework methods are anticipated since the mid1990s, and numerous shapes of recommender framework code were created as of late for a spread of applications. The health-related substance shared through on-line feedbacks or surveys contains covered up assumption designs that emerges through totally distinctive sources from medical world which offer benefits to the pharmaceutical industry. Amid this, the on-line component is fantastically standard of late for online looking, diverse stock through distinctive websites like on-line buying of drugs at entryway step. Numerous websites and blogs offers clients to rate their stock with their fulfillment and quality of stock, logistics, administrations and criticism etc., which the clients examines for a particular medicine or on quality of administration.

#### DISADVANTAGE

- In the existing work, the system did not implement an exact sentiment analysis for large data sets.
- This system is less performance due to lack Data Classification and Data Fragmentation technique.

#### PROPOSED SYSTEM

A recommender framework is a customary system that proposes an item to the user, dependent on their advantage and necessity. These frameworks employ the customers' surveys to break down their sentiment and suggest a recommendation for their exact need. In the drug recommender system, medicine is offered on a specific condition dependent on patient reviews using sentiment analysis and feature engineering. Sentiment analysis is a progression of strategies, methods, and tools for distinguishing and extracting emotional data, such as opinion and attitudes, fro [7]. On the other hand, Featuring engineering is the process of making more features from the existing ones; it improves the performance of models.

#### ADVANTAGE

- The system is more effective since it presents the proposed algorithm used in natural language processing responsible for counting the number of times of all the tokens in review or document.
- The system has exact sentiment analysis prediction techniques for Data Cleaning and Visualization.

## V. RESULT

In this work, each review was classified as positive or negative, depending on the user's star rating. Ratings above five are classified as positive, while negative ratings are from one to five-star ratings. Initially, the number of positive ratings and negative ratings in training data were 111583 and 47522, respectively. After applying smote, we increased the minority class to have 70 percent of the majority class examples to curb the imbalances. The updated training data contains 111583 positive classes and 78108 negative classes. Four different text representation methods, namely Bow, TF-IDF, Word2Vec, Manual feature and ten different ML algorithms were applied for binary classification. Results belonging to 5 different metrics given Table III shows the results using evaluation metrics on a bag of words vectorization technique. We can easily see that perceptron outperforms all other classification algorithms. All algorithms showed similar types of results ranging from 89% to 91% accuracy. Logistic regression and LinearSVC accomplished a 90% AUC score. Even after achieving accuracy more prominent than logistic and LinearSVC, Perceptron achieved only 89.8% AUC score.

## VI. CONCLUSION

Reviews are becoming an integral part of our daily lives; whether go for shopping, purchase something online or go to some restaurant, we first check the reviews to make the right decisions. Motivated by this, in this research sentiment analysis of drug reviews was studied to build a recommender system using different types of machine learning classifiers, such as Logistic Regression, Perceptron, Multinomial Naive Bayes, Ridge classifier, Stochastic gradient descent, LinearSVC, applied on Bow, TF-IDF, and classifiers such as Decision Tree, Random Forest, Lgbm, and Catboost were applied on Word2Vec and Manual features method. We evaluated them using five different metrics, precision, recall, f1score, accuracy, and AUC score, which reveal that the Linear SVC on TF-IDF outperforms all other models with 93% accuracy. On the other hand, the Decision tree classifier on Word2Vec showed the worst performance by achieving only 78% accuracy. We added best-predicted emotion values from each method, Perceptron on Bow (91%), LinearSVC on TF-IDF (93%), LGBM on Word2Vec (91%), Random Forest on manual features (88%), and multiply them by the normalized usefulCount to get the overall score of the drug by condition to build a recommender system. Future work involves comparison of different oversampling techniques, using different values of n-grams, and optimization of algorithms to improve the performance of the recommender system.

## REFERENCES

- [1] Telemedicine, <https://www.mohfw.gov.in/pdf/Telemedicine.pdf>
- [2] Wittich CM, Burkle CM, Lanier WL. Medication errors: an overview for clinicians. *Mayo Clin Proc.* 2014 Aug;89(8):1116-25.
- [3] CHEN, M. R., & WANG, H. F. (2013). The reason and prevention of hospital medication errors. *Practical Journal of Clinical Medicine*, 4.
- [4] DrugReview Dataset, <https://archive.ics.uci.edu/ml/datasets/Drug%2BReview%2BDataset%2B%2528Drugs.com%2529#>
- [5] Fox, Susannah, and Maeva Duggan. "Health online 2013. 2013." URL: <http://pewinternet.org/Reports/2013/Health-online.aspx>
- [6] Bartlett JG, Dowell SF, Mandell LA, File TM Jr, Musher DM, Fine MJ. Practice guidelines for the management of community-acquired pneumonia in adults. *Infectious Diseases Society of America. Clin Infect Dis.* 2000 Aug;31(2):347-82. doi: 10.1086/313954. Epub 2000 Sep 7. PMID: 10987697; PMCID: PMC7109923.
- [7] Fox, Susannah & Duggan, Maeva. (2012). Health Online 2013. Pew Research Internet Project Report.
- [8] T. N. Tekade and M. Emmanuel, "Probabilistic aspect mining approach for interpretation and evaluation of drug reviews," 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPEs), Paralakhemundi, 2016, pp. 1471-1476, doi: 10.1109/SCOPEs.2016.7955684.
- [9] Doulaverakis, C., Nikolaidis, G., Kleontas, A. et al. GalenOWL: Ontology-based drug recommendations discovery. *J Biomed Semant* 3, 14 (2012). <https://doi.org/10.1186/2041-1480-3-14>
- [10] Leilei Sun, Chuanren Liu, ChonghuiGuo, HuiXiong, and YanmingXie. 2016. Data-driven Automatic Treatment Regimen Development and Recommendation. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 1865–1874. DOI:<https://doi.org/10.1145/2939672.2939866>
- [11] V. Goel, A. K. Gupta and N. Kumar, "Sentiment Analysis of Multilingual Twitter Data using Natural Language Processing," 2018 8th International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 2018, pp. 208-212, doi: 10.1109/CSNT.2018.8820254.



- [12] Shimada K, Takada H, Mitsuyama S, et al. Drug-recommendation system for patients with infectious diseases. AMIA AnnuSymp Proc. 2005;2005:1112.
- [13] Y. Bao and X. Jiang, "An intelligent medicine recommender system framework," 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), Hefei, 2016, pp. 1383-1388, doi: 10.1109/ICIEA.2016.7603801.
- [14] Zhang, Yin & Zhang, Dafang& Hassan, Mohammad & Alamri, Atif&Peng, Limei. (2014). CADRE: Cloud-Assisted Drug REcommendation Service for Online Pharmacies. Mobile Networks and Applications. 20. 348-355. 10.1007/s11036-014-0537-4.
- [15] J. Li, H. Xu, X. He, J. Deng and X. Sun, "Tweet modeling with LSTM recurrent neural networks for hashtag recommendation," 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, 2016, pp. 1570-1577, doi: 10.1109/IJCNN.2016.7727385.
- [16] Zhang, Yin & Jin, Rong& Zhou, Zhi-Hua. (2010). Understanding bag-of-words model: A statistical framework. International Journal of Machine Learning and Cybernetics. 1. 43-52. 10.1007/s13042-010-0001-0.
- [17] 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.
- [18] Yoav Goldberg and Omer Levy. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method, 2014; arXiv:1402.3722.
- [19] DanushkaBollegala, TakanoriMaehara and Kenichi Kawarabayashi. Unsupervised Cross-Domain Word Representation Learning, 2015; arXiv:1505.07184.
- [20] Textblob, <https://textblob.readthedocs.io/en/dev/>.
- [21] van der Maaten, Laurens & Hinton, Geoffrey. (2008). Viualizing data using t-SNE. Journal of Machine Learning Research. 9. 2579-2605.
- [22] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer. SMOTE: Synthetic Minority Over-sampling Technique, 2011, Journal Of Artificial Intelligence Research, Volume 16, pages 321-357, 2002; arXiv:1106.1813. DOI: 10.1613/jair.953.
- [23] Powers, David & Ailab,. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness& correlation. J. Mach. Learn. Technol. 2. 2229-3981. 10.9735/2229-3981



**INNO**  **SPACE**  
SJIF Scientific Journal Impact Factor

**Impact Factor: 8.165**

**doi**<sup>®</sup>  
**cross** **ref**

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



[www.ijircce.com](http://www.ijircce.com)

Scan to save the contact details