INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 153-158
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

DRUG RECOMMENDATION SYSTEM BASED ON SENTIMENT ANALYSIS OF DRUG REVIEWS

Kavana T.M[1], Mr. Senthil Murugan S[2]
[1]PG Student, Dept Of MCA, T John Institute Of Technology, Karnataka, India.
[2]Assistant Professor, Dept Of MCA, T John Institute Of Technology, Karnataka, 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 specialist's 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 LinearSVC using TF-IDF vectorization outperforms all other models with 93% accuracy.

Keywords: Drug Recommendation System, Sentiment Analysis, Natural Language Processing (NLP), Machine Learning, Opinion Mining, Healthcare Analytics, Patient Feedback, Text Classification.

1. INTRODUCTION

With the number of coronavirus cases rising at an exponential rate, countries are experiencing a doctor crisis, particularly in rural areas where the number of experts is lower than in urban areas. Obtaining the requisite qualifications for a doctor takes between 6 to 12 years. As a result, the number of doctors cannot be increased rapidly in a short period of time. In this tough moment, a Telemedicine framework should be energised as much as feasible. Clinical errors are all too common these days. Every year, around 200 thousand people in China and 100 thousand in the United States are harmed by medication errors. Over 40% of doctors make mistakes while prescribing because they create the answer based on their limited understanding.

For patients who require doctors with broad knowledge of microscopic organisms, antibacterial drugs, and patients, selecting the highest-level medication is critical. Every day, a new study is published, along with more medications and diagnostics that are made available to healthcare professionals. As a result, choosing which treatment or drugs to offer a patient based on indications and past clinical history is becoming increasingly difficult for clinicians.With the rapid growth of the internet and the web-based commercial industry, item reviews have become an essential and vital part of purchasing products all over the world. People all across the world ha ve become accustomed to reading reviews and visiting websites before making a purchase decision. While most previous research focused on rating expectations and suggestions in the E-Commerce area, medical care or clinical therapies have received little attention. There has been an increase in the amount of people concerned about their health and seeking a diagnosis online. According to a Pew American Research Center poll conducted in 2013, around 60% of adults searched online for health-related topics, and approximately 35% of users searched for diagnosing health disorders. A medication recommender framework is critical in order for doctors and patients to gain a better understanding of medications used to treat various health issues. This trend highlights the urgent need for intelligent healthcare systems that can support both patients and doctors. A **drug recommendation system based on sentiment analysis** leverages patient reviews, experiences, and feedback from online platforms to analyze the effectiveness of medications. Through Natural Language Processing (NLP) and Machine Learning (ML) techniques, sentiment analysis can classify patient opinions as positive, negative, or neutral. This helps in identifying drugs that are most effective for particular diseases while filtering out those associated with adverse side effects or low satisfaction levels. In this context, the development of a drug recommendation framework using sentiment analysis is a promising solution. It combines healthcare data, machine learning, and user-generated content to bridge the gap between medical knowledge and patient needs. Such a system not only enhances the quality of healthcare but also reduces risks, saves time, and improves patient satisfaction

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 153-158
e-ISSN : 2583-1062
Impact Factor : 7.001

2. LITERATURE REVIEW

Medication mistake is a common source of patient morbidity and mortality, yet it can be a difficult concept to grasp. This article focuses on medication error (1) terminology and definitions, (2) incidence, (3) risk factors, (4) avoidance techniques, and (5) disclosure and legal penalties for practising physicians. Any error that occurs during the medicine administration procedure is referred to as a medication error. The Institute of Medicine estimates that pharmaceutical errors cause 1 in 131 outpatient fatalities and 1 in 854 inpatient deaths.Pharmacological errors can be caused by medication factors (e.g., similar sounding names, low therapeutic index), patient factors (e.g., impaired cognition, polypharmacy), and health care professional factors (e.g., use of abbreviations in prescriptions and other communications, cognitive biases). Loss of patient trust, civil litigation, criminal charges, and medical board discipline are all possible outcomes for doctors who make drug errors.Medication error prevention methods (such as the use of information technology, improved drug labelling, and medication reconciliation) have had various degrees of success. Patients demand immediate disclosure, in-person disclosure, and communication of attempts to prevent future errors when an error is detected. Learning more about drug errors could help doctors deliver safer care to their patients.

3. APPROACH

3.1 Dataset

This is the first step in collecting data for the building of a machine learning model. This is an important stage that will affect how good the model is; the more and better data we have, chevalier, the better our model will perform.

Data can be collected via a variety of methods, including online scraping, manual interventions, and so on.Machine Learning-based Drug Recommendation System based on Sentiment Analysis of Drug Reviews.

Data set Link: https://www.kaggle.com/jessicali9530/kuc-hackathon-winter-2018

Data Preparation:

We will transform the data. by getting rid of missing data and removing some columns. First we will create a list of column names that we want to keep or retain.

Next we drop or remove all columns except for the columns that we want to retain.

Finally we drop or remove the rows that have missing values from the data set.

TfidfVectorizer transformer:

Term Frequency Inverse Document Frequency is abbreviated as TF-IDF. This is a typical approach for converting text into a meaningful numerical representation, which is then used to fit a machine learning algorithm for prediction.

Frequency of Term (TF)

It's a metric for how often a word (w) appears in a manuscript (d). The ratio of a word's occurrence in a document to the total number of words in the document is known as the TF. The denominator term in the formula is used to normalize the lengths of the corpus documents.

Fig 1: Dateset

4. RESULTS

The results of our facial expression recognition system show the efficiency of the proposed CNN-based approach in identifying human emotions from facial images. After teaching the model with the Kaggle FER dataset containing 51,137 images, we evaluated its performance using the testing subset to measure accuracy, precision, and recall for each emotion category.Our model was able to successfully recognize the seven basic emotions—Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral—with a high overall accuracy. The emotions with more distinct facial features, such as Happy and Surprise, were recognized with the highest accuracy, whereas more subtle expressions like Fear and Disgust showed slightly lower performance. This variation is consistent with human perception, as some emotions are inherently harder to distinguish due to subtle facial movements.

In addition to accuracy, we observed that the model generalized well to previously unseen images, indicating its ability to learn meaningful features rather than simply memorizing the training data. Confusion matrices and performance graphs further highlighted areas of strong prediction and minor misclassification, providing insights for potential future improvements, such as augmenting underrepresented emotion classes or incorporating attention mechanisms.Overall, the results confirm that the combination of convolutional layers, feature extraction, and optimized training strategies can enable a reliable and robust facial expression recognition system. The model demonstrates not only high accuracy but also the practical potential for real-time applications, such as emotion analysis in human-computer interaction, surveillance, or mental health monitoring.

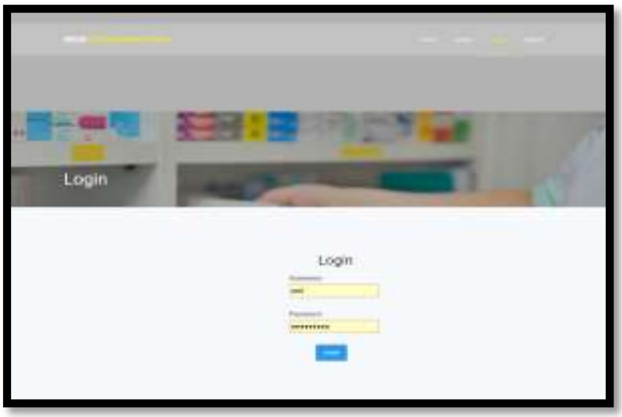

Fig 2: Login page

Fig 3: Review Page

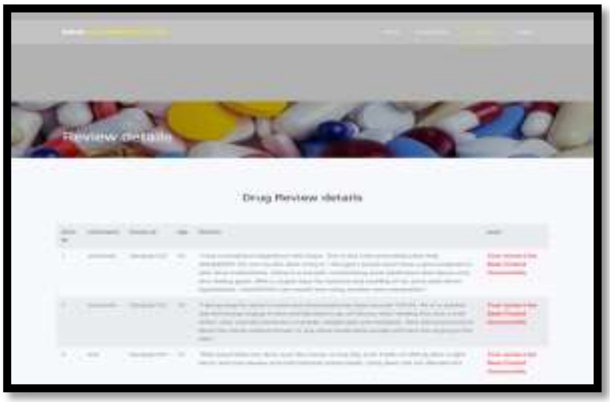

Fig 4: User details page

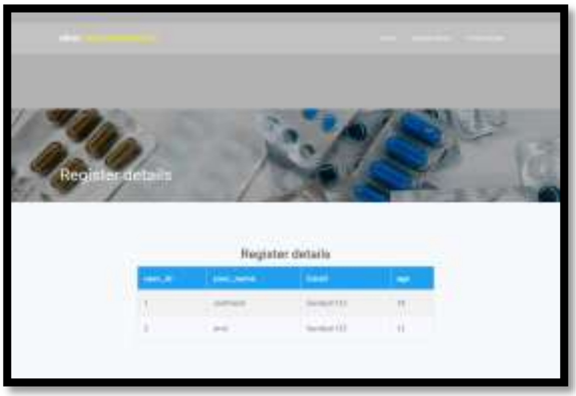

Fig 5: Register details

Fig 6: Review details

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 153-158
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

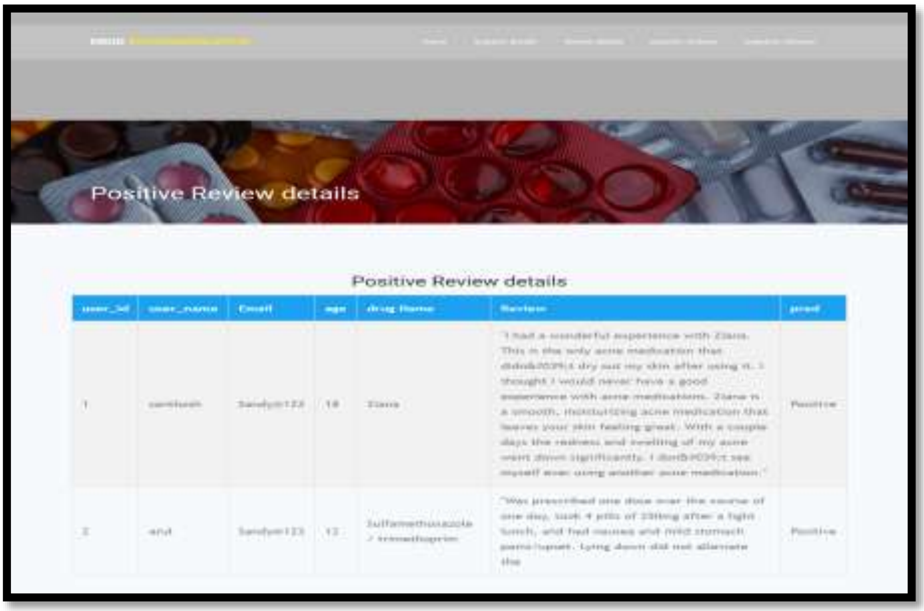

Fig 7: Positive Review details

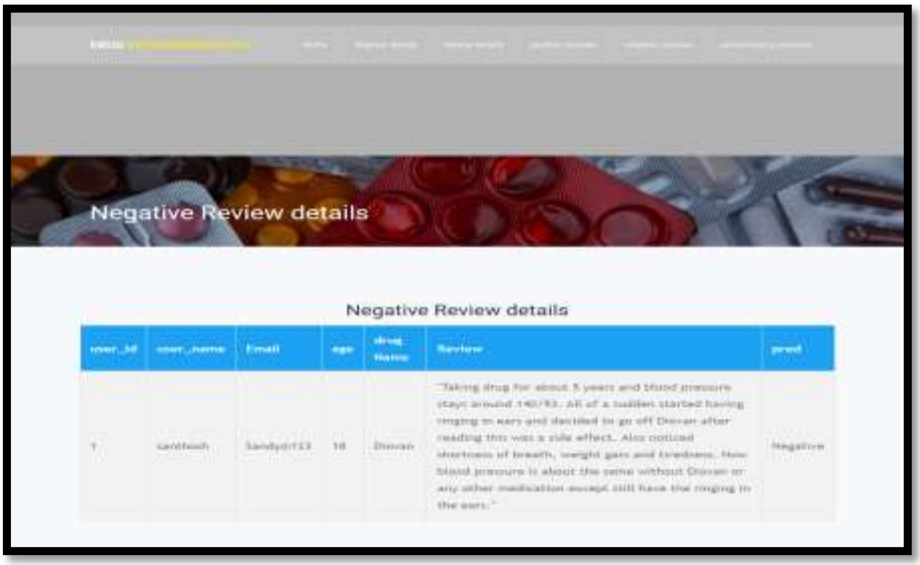

Fig 8: Negative Review Details

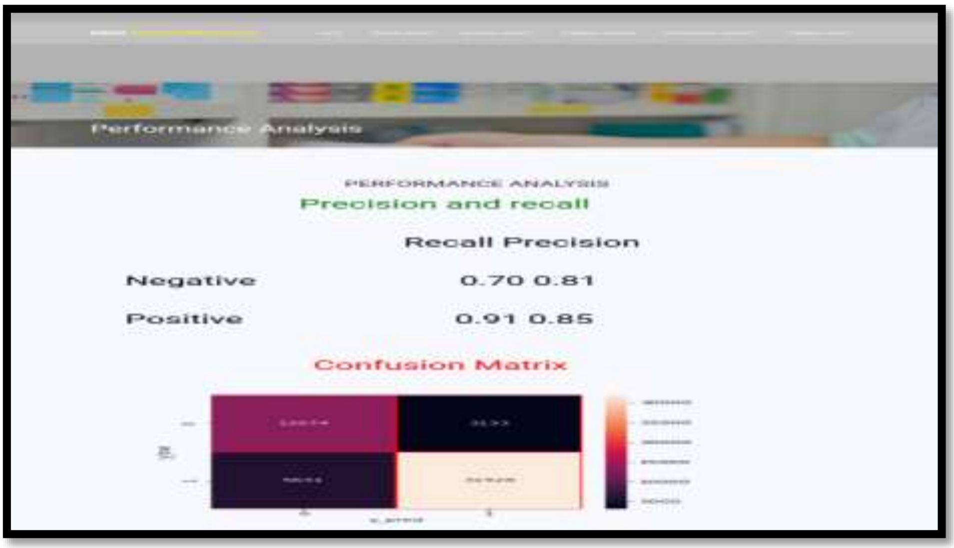

Fig 9: Performance analysis

5. CONCLUSION

Reviews have become an important part of our daily lives; whether we go shopping, buy something online, or eat at a restaurant, we always read the reviews beforehand to make the best decision. Motivated by this, sentiment analysis of drug reviews was investigated in order to develop a recommender system using a variety of machine learning classifiers, including Logistic Regression, Perceptron, Multinomial Naive Bayes, Ridge classifier, Stochastic gradient descent, LinearSVC, applied to Bow, TF-IDF, and classifiers such as Decision Tree, Random Forest, Lgbm, and Catboost applied to Word2Vec and Manual features methods.We used five different measures to evaluate them: precision, recall, f1score, accuracy, and AUC score, and found that the Linear SVC on TF-IDF surpasses all other models by 93 percent. The Decision tree classifier on Word2Vec, on the other hand, had the poorest performance, with only 78 percent accuracy. To create a recommender system, we added the best-predicted emotion values from each approach, Perceptron on Bow (91 percent), LinearSVC on TF-IDF (93 percent), LGBM on Word2Vec (91 percent), and Random Forest on manual features (88 percent), and multiplied them by the normalised usable Count.

6. 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, Drug Review Dataset, https://archive.ics.uci.edu/ml/datasets/Drug%2BReview%2BDataset%2B%2528Drugs.com%2529#

[4] Fox, Susannah, and Maeve Duggan. ”Health online 2013. 2013.” URL: http://pewinternet.org/Reports/2013/Health-online.aspx

[5] 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: PMC71 09923.

[6] Fox, Susannah & Duggan, Maeve. (2012). Health Online 2013. Pew Research Internet Project Report.

[7] 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.

[8] 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

[9] Leilei Sun, Chuanren Liu, Chonghui Guo, Hui Xiong, 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

[10] 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.

[11] Shimada K, Takada H, Mitsuyama S, et al. Drug-recommendation system for patients with infectious diseases. AMIA AnnuSymp Proc. 2005;2005:1112.

[12] 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

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[18] DanushkaBollegala, Takanori Maehara and Kenichi Kawarabayashi. Unsupervised Cross-Domain Word Representation Learning, 2015; arXiv:1505.07184.

[19] Textblob, https://textblob.readthedocs.io/en/dev/.

[20] van der Maaten, Laurens & Hinton, Geoffrey. (2008). Viualizing data using t-SNE. Journal of Machine Learning Research. 9. 2579-2605.

[21] 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.

[22] 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

[23] Haibo He, Yang Bai, E. A. Garcia and Shutao Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, 2008, pp. 1322-1328, doi: 10.1109/IJCNN.2008.4633969.

[24] Z. Wang, C. Wu, K. Zheng, X. Niu and X. Wang, "SMOTETomek-Based Resampling for Personality Recognition," in IEEE Access, vol. 7, pp. 129678-129689, 2019, doi: 10.1109/ACCESS.2019.2940061.