International Journal of Research in Engineering, Science and Management
Volume 7, Issue 10, October 2024
https://www.ijresm.com | ISSN (Online): 2581-5792

Drug Recommendation System Using TF-IDF Vectorization and Cosine Similarity

B. R. Poorna[1*], Edwin Charles Mathew[2]
[1]Assistant Professor, Department of Computer Science and Engineering, Mar Baselios College of Engineering and Technology, Trivandrum, India
[2]Student, Department of Computer Science and Engineering, Mar Baselios College of Engineering and Technology, Trivandrum, India

***Abstract*: This paper presents a machine learning-based drug recommendation system implemented as a Python web application using Flask. The system enables users to input symptoms and receive a list of recommended drugs based on historical patient data. The project leverages natural language processing (NLP) techniques, specifically TF-IDF vectorization and cosine similarity, to match user-provided symptoms with reviews of drugs used for similar conditions. The system was developed to assist in the efficient selection of drugs, improve patient care, and demonstrate the application of machine learning in a real-world healthcare scenario. This paper details the project's development, including its background, methodology, implementation, results, and potential future improvements.**

***Keywords*: cosine similarity, drug recommendation, Flask framework, machine learning, natural language processing (NLP), patient reviews, Python web application, sentiment analysis, symptom-based recommendation, TF-IDF vectorization.**

1. Introduction

The integration of machine learning (ML) in healthcare has become increasingly important, offering new opportunities for improved patient outcomes and operational efficiencies. One such application is the development of drug recommendation systems, which can provide valuable support to medical professionals and patients by suggesting drugs based on input symptoms. This paper presents a machine learning-based drug recommendation system designed as a basic web application using Flask, a Python web framework. The primary objective of this project is to demonstrate the feasibility of using machine learning techniques, particularly NLP, to analyze patient data and recommend suitable drugs efficiently.

2. Literature Review

A. Integrated Opinion-Based Drug Recommendation with Transformers (Job et al., 2023) [1]

This study explores the use of Transformers for analyzing patient feedback to recommend drugs. The authors fine-tuned models like BERT to enhance the context-sensitivity and accuracy of drug recommendations. By leveraging pre-trained language models, the system can better interpret nuanced medical texts and patient reviews, leading to more precise drug suggestions. The research underscores the potential of transformer-based models to significantly improve the effectiveness of drug recommendation systems in clinical settings.

Key Contributions:

1) Utilized Transformers for sentiment analysis of drug reviews.
2) Improved recommendation accuracy by focusing on contextual understanding.
3) Demonstrated the feasibility of integrating advanced NLP models in healthcare applications.

B. Deep Learning-Based Sentiment Analysis on Drug Reviews (Ganesh et al., 2023) [2]

Ganesh et al. applied deep learning models, specifically Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM), to perform sentiment analysis on drug reviews. The models achieved high accuracy rates (85.12% for LSTM and 89.3% for BiLSTM), demonstrating the strength of deep learning in capturing complex patterns in text data. This study highlights the potential of deep learning to outperform traditional machine learning approaches in sentiment analysis tasks.

Key Contributions:

1) Applied LSTM and BiLSTM for drug review sentiment analysis.
2) Achieved high accuracy in classifying sentiments.
3) Emphasized the advantages of deep learning models in handling sequential data.

C. Sentiment Analysis of Drug Reviews Using Transfer Learning (Punith NS & Krishna Raketla, 2021) [3]

This paper compares various transformer-based models (BERT, XLNET, Bio_Clinical BERT, and ConvBERT) for sentiment analysis of drug reviews. The authors found that XLNET and ConvBERT outperformed other models, with ConvBERT achieving the highest accuracy (94.3%). The study underscores the importance of pre-trained models and their adaptation to domain-specific tasks, showing how advanced transformers can enhance the performance of sentiment analysis systems in healthcare.

Key Contributions:

1) Compared multiple transformer-based models for

*Corresponding author: poorna.br@mbcet.ac.in

sentiment analysis.
2) Demonstrated superior performance of XLNET and ConvBERT.
3) Highlighted the effectiveness of domain-specific pre-training in improving model accuracy.

D. Determining the Efficiency of Drugs Under Special Conditions from Users' Reviews on Healthcare Web Forums (Saad et al., 2021) [4]

This paper investigates a hybrid sentiment analysis approach, combining lexicon-based and learning-based techniques, to evaluate the efficiency of drugs under specific conditions. The study applied sentiment lexicons like AFFIN, TextBlob, and VADER alongside machine learning models, demonstrating that this hybrid approach improves classification accuracy. The findings emphasize the importance of post-market drug surveillance and the role of user reviews in assessing drug efficacy and safety.

Key Contributions:
1) Combined lexicon and learning-based sentiment analysis methods.
2) Enhanced classification accuracy of drug reviews.
3) Highlighted the utility of user reviews for drug efficacy assessments.

E. Drug Recommendation System Based on Sentiment Analysis of Drug Reviews Using Machine Learning (Garg, 2021) [5]

This research presents a drug recommendation system that uses sentiment analysis of drug reviews through traditional machine learning algorithms. The study employed TF-IDF vectorization and classifiers such as logistic regression and random forest to categorize reviews into sentiment classes. The work illustrates how sentiment analysis can effectively inform drug recommendations by analyzing user experiences.

Key Contributions:
1) Implemented traditional machine learning algorithms for sentiment analysis.
2) Demonstrated the effectiveness of feature engineering techniques.
3) Provided insights into how sentiment analysis can guide drug recommendations.

3. Methodology

A. Data collection and preprocessing

The dataset used in this project was obtained from the UC Irvine Machine Learning Repository. It consists of patient reviews of drugs from the "Drugs.com" website, including information such as the drug name, the condition for which it was prescribed, and patient reviews. The dataset was loaded into a Pandas DataFrame for processing.

Data preprocessing steps included:
1) Converting text to lowercase to standardize the data.
2) Handling missing values in the `drugName` column by filling them with the most frequent drug name.
3) Dropping rows with missing values in the `condition` column to ensure data integrity.

B. Natural Language Processing

To recommend drugs based on symptoms, the project employed NLP techniques:
1) TF-IDF Vectorization: The `TfidfVectorizer` from the Scikit-Learn library was used to convert the patient reviews into a matrix of TF-IDF features. This step helps in quantifying the importance of words in each review.
2) Cosine Similarity: The cosine similarity metric was used to compute the similarity between input symptoms and the available drug reviews. This metric helps identify drugs whose reviews closely match the symptom description provided by the user.

C. Machine Learning Model

The core functionality of the recommendation system is built around a similarity-based approach:
1) The input symptom is matched against the `condition` column in the dataset.
2) If a match is found, the reviews for the corresponding drugs are processed using TF-IDF vectorization.
3) Cosine similarity is computed to identify the most relevant drug reviews, and the top-ranked drugs are recommended.

D. Web Application Development

The project used the Flask web framework to create a basic Python web application. The application consists of two main routes:
1) Home Route (`/`): Displays a message confirming the server is running.
2) Predict Route (`/predict`): Accepts user input via a POST request containing the symptom, processes the input, and returns drug recommendations in JSON format.

Cross-Origin Resource Sharing (CORS) is enabled to allow the application to handle requests from different origins, enhancing its usability and accessibility.

4. Results and Discussions

The drug recommendation system successfully recommends drugs based on input symptoms by utilizing NLP techniques to analyze patient reviews. The use of TF-IDF and cosine similarity allows the system to identify and rank relevant drug reviews, providing users with a list of potentially suitable medications.

A. Limitations

The current implementation has certain limitations:
1) The system relies on the quality and comprehensiveness of the dataset. Any biases or gaps in the data can affect the recommendations.
2) The model does not account for patient-specific factors, such as age, gender, allergies, or medical history, which may be important in drug prescription.
3) The recommendation approach is primarily based on the similarity of text data, without considering the

pharmacological properties or potential drug interactions.

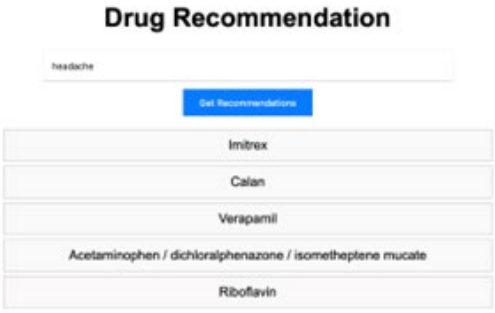

Fig. 1. Drug recommendations for 'Headache'

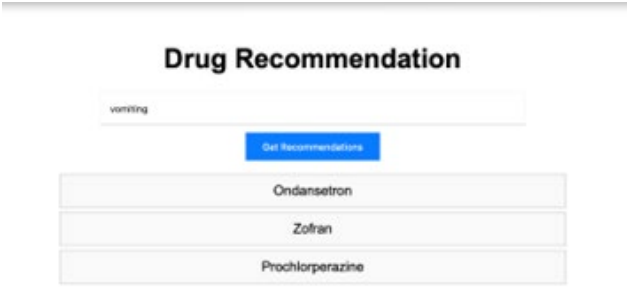

Fig. 2. Drug recommendations for 'Vomiting'

5. Conclusion

This project successfully developed a basic machine learning-based drug recommendation system using Flask, demonstrating the application of NLP techniques in healthcare. The system effectively matches input symptoms with drug reviews to suggest relevant medications, offering a valuable tool for medical professionals and patients. However, there is room for improvement, particularly in terms of data quality, personalized recommendations, and the integration of additional medical data.

A. Future Work

1) Integrating more comprehensive datasets to improve the diversity and accuracy of recommendations.
2) Incorporating patient-specific information such as demographics, allergies, and medical history for more personalized recommendations.
3) Exploring advanced machine learning techniques, including deep learning models, to handle more complex patterns and interactions.
4) Developing a full-scale web or mobile application to enhance usability and accessibility.

Acknowledgment

We would like to express our sincere gratitude to CERD (Centre for Engineering Research and Development) for their generous support that made this research possible.

References

[1] S. Job, X. Tao, Y. Li, L. Li and J. Yong. (2023, March). Topic Integrated Opinion-Based Drug Recommendation with Transformers. *IEEE Transactions on Emerging Topics in Computational Intelligence.* [Online]. 7(6), pp. 1676-1686.

[2] G. C, A. R, K. V, K. D and S. M, "Deep Learning Based Sentiment Analysis on Drug Reviews," in ICIDeA, Imphal, Manipur, 2023, pp. 401-406.

[3] P. NS and K. Raketla, "Sentiment Analysis of Drug Reviews using Transfer Learning," in ICIRCA, Coimbatore, Tamil Nadu, 2021, pp. 1794-1799.

[4] E. Saad et al., (2021). Determining the Efficiency of Drugs Under Special Conditions from Users' Reviews on Healthcare Web Forums. *IEEE Access.* [Online]. 9, pp. 85721-85737.

[5] S. Garg (2021), "Drug Recommendation System based on Sentiment Analysis of Drug Reviews using Machine Learning," in International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, Uttar Pradesh, 2021, pp. 175-181.