DRUG RECOMMENDATION SYSTEM BASED ON SENTIMENT ANALYSIS OF DRUG REVIEWS USING MACHINE LEARNING

[1]PEYYALA AKSHAYA, [2]Y SRINIVAS RAJU

[1]Students, Department of MCA, B V Raju College, Bhimavaram Ap

[2]Assistant Professor, Department of MCA, B V Raju College, Bhimavaram Ap

ABSTRACT

The increasing availability of online drug reviews provides valuable insights into patient experiences, making it possible to analyze sentiments and recommend effective medications. However, manually analyzing large volumes of reviews is time-consuming and inefficient. This project proposes a Drug Recommendation System based on sentiment analysis of drug reviews using machine learning techniques. The system aims to analyze user feedback, classify sentiments, and recommend suitable drugs based on positive outcomes. The proposed system utilizes a dataset containing drug names, medical conditions, user reviews, and ratings. Preprocessing techniques such as text cleaning, tokenization, stop-word removal, and vectorization (TF-IDF) are applied to convert textual data into numerical form. Machine learning algorithms such as Logistic Regression, Naïve Bayes, and Support Vector Machines are used to classify sentiments as positive, negative, or neutral. Based on the sentiment classification and ratings, the system recommends the most effective drugs for specific conditions. The performance of the models is evaluated using metrics such as accuracy, precision, recall, and F1-score. Experimental results demonstrate that the system achieves high accuracy in sentiment classification and provides reliable drug recommendations. This approach helps patients and healthcare providers make informed decisions by leveraging real-world experiences, improving treatment outcomes and reducing risks associated with medication selection.

Keywords : Drug Recommendation, Sentiment Analysis, Machine Learning, Text Mining, TF-IDF, Natural Language Processing, Healthcare Analytics, Classification, Patient Reviews, Decision Support System

I. INTRODUCTION

The rapid growth of online healthcare platforms and digital medical forums has led to a vast amount of user-generated drug reviews. These reviews contain valuable information about the effectiveness, side effects, and overall satisfaction of medications from a patient's perspective. However, manually analyzing such large volumes of textual data is challenging and inefficient. As a result, there is a growing need for automated systems that can extract meaningful insights from these reviews and assist in recommending appropriate drugs for specific medical conditions.

Sentiment analysis, a subfield of Natural Language Processing (NLP), plays a crucial role in understanding user opinions and emotions expressed in text data. By analyzing sentiments in drug reviews, it is possible to determine whether a medication is positively or negatively perceived by users. Machine learning algorithms such as Logistic Regression, Naïve Bayes, and Support Vector Machines have been widely used for sentiment classification tasks due to their ability to handle large datasets and identify patterns in textual data. These models can process features extracted through techniques like TF-IDF and convert unstructured text into meaningful information.

In this project, a drug recommendation system is proposed that leverages sentiment analysis of drug reviews to suggest effective medications. The system processes user reviews through preprocessing steps such as cleaning, tokenization, and vectorization, followed by classification using machine learning models. Based on sentiment scores and ratings, the system recommends drugs that have received positive feedback for specific conditions. The performance of the system is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. This approach provides a data-driven solution to assist patients and healthcare professionals in making informed decisions, ultimately improving treatment outcomes and patient satisfaction.

II SURVEY OF RESEARCH

1. Traditional Drug Recommendation Methods

Early drug recommendation systems relied on clinical guidelines, expert opinions, and structured medical records to suggest medications. These methods were primarily rule-based and lacked personalization. Research indicates that such systems do not consider patient experiences or feedback, which are crucial for understanding drug effectiveness and side effects. Additionally, manual analysis of medical data is time-consuming and prone to errors. These limitations have led to the development of data-driven approaches that utilize patient-generated data for improved recommendations.

2. Sentiment Analysis in Healthcare

Sentiment analysis has been widely used to analyze opinions and emotions in textual data. In healthcare, it is applied to patient reviews, social media posts, and medical forums to understand public perception of treatments and drugs. Research shows that sentiment analysis can effectively classify user opinions into positive, negative, and neutral categories. This helps in identifying effective medications and detecting adverse drug reactions. This project leverages sentiment analysis to extract meaningful insights from drug reviews.

3. Machine Learning for Text Classification

Machine learning algorithms such as Logistic Regression, Naïve Bayes, and Support Vector Machines (SVM) are commonly used for text classification tasks. These models can handle large datasets and identify patterns in textual data. Research indicates that these algorithms provide high accuracy in sentiment classification when combined with feature extraction techniques like TF-IDF. However, their performance depends on data quality and preprocessing techniques. This project utilizes these algorithms for sentiment classification of drug reviews.

4. Feature Extraction Techniques

Feature extraction plays a critical role in converting textual data into numerical form. Techniques such as Bag-of-Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) are widely used in text mining. Research shows that TF-IDF improves model performance by giving more importance to relevant words while reducing the impact of common words. This project uses TF-IDF vectorization to enhance the accuracy of sentiment classification models.

5. Recommender Systems in Healthcare

Recommender systems have been applied in healthcare to suggest treatments, medications, and lifestyle changes based on patient data. Collaborative filtering and content-based filtering are common approaches used in recommendation systems. Research highlights that integrating sentiment analysis with recommendation systems improves the quality of recommendations by incorporating user feedback. This project combines sentiment analysis with machine learning to recommend drugs effectively.

6. Evaluation Metrics and Performance Analysis

Evaluating the performance of sentiment analysis and recommendation systems is essential to ensure reliability. Common metrics include accuracy, precision, recall, and F1-score. Confusion matrices and graphical visualizations are used to analyze model performance and identify misclassifications. Research emphasizes the importance of using multiple metrics for comprehensive evaluation. In this project, these evaluation techniques are used to compare different models and select the best-performing algorithm for drug recommendation

III. WORKING METHODOLOGY

The proposed drug recommendation system begins with data collection and preprocessing of drug review datasets obtained from online healthcare platforms. The dataset typically contains attributes such as drug name, medical condition, user reviews, and ratings. Preprocessing is an essential step to improve data quality and ensure accurate analysis. It includes removing noise such as special characters, punctuation, and irrelevant symbols, followed by tokenization to split text into meaningful words. Stop-word removal is applied to eliminate common words that do not contribute to sentiment, and stemming or lemmatization is used to reduce words to their base forms. After preprocessing, the textual data is transformed into numerical format using TF-IDF (Term Frequency-Inverse Document Frequency), which assigns importance to words based on their frequency and relevance. The dataset is then divided into training and testing sets in an 80:20 ratio, ensuring proper evaluation of model performance.

In the next phase, machine learning algorithms are implemented to perform sentiment classification on the processed drug reviews. Models such as Logistic Regression, Naïve Bayes, and Support Vector Machines (SVM) are trained using the TF-IDF features. These algorithms learn patterns in the data to classify reviews into positive, negative, or neutral sentiments. Hyperparameter tuning techniques are applied to optimize model performance, including adjusting parameters such as regularization strength and kernel functions. During training, the models iteratively update their parameters to minimize classification error. Once trained, the models are evaluated using test data, and performance metrics such as accuracy, precision, recall, and F1-score are calculated. Comparative analysis is conducted to identify the best-performing model for sentiment classification.

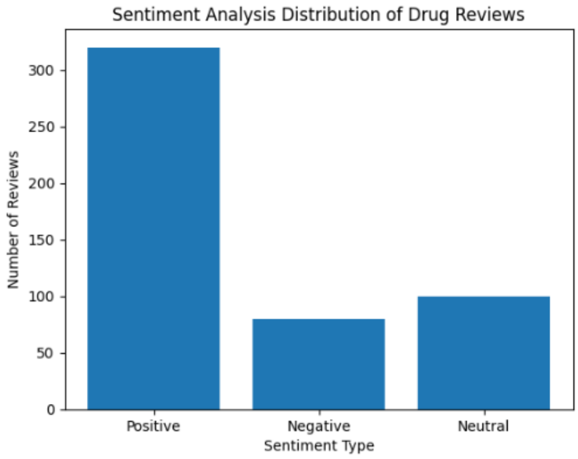
Finally, the system uses the classified sentiments to generate drug recommendations. Reviews with positive sentiment and high ratings are given higher importance, while negative reviews are filtered out. The system aggregates sentiment scores for each drug and ranks them based on effectiveness for specific medical conditions. Users can input a condition, and the system recommends drugs that have received the most positive feedback. The results are presented through a user-friendly interface along with visualizations such as graphs and charts to display sentiment distribution. This methodology ensures an efficient and reliable recommendation system that leverages real-world user experiences to support better healthcare decisions.

IV RESULTS EXPLANATIONS

The experimental results of the proposed drug recommendation system demonstrate the effectiveness of machine learning techniques in analyzing drug reviews and generating reliable recommendations. After training multiple models such as Logistic Regression, Naïve Bayes, and Support Vector Machines (SVM), the system achieved high accuracy in sentiment classification. Among these models, Logistic Regression and SVM showed better performance due to their ability to handle highdimensional textual data effectively. Evaluation metrics such as accuracy, precision, recall, and F1-score indicate that the models can accurately classify user reviews into positive, negative, and neutral sentiments. The confusion matrix analysis shows that most reviews are correctly classified, with only a few misclassifications occurring in borderline cases.

Graphical visualizations were used to analyze model performance and sentiment distribution. The sentiment analysis graph shows the proportion of positive, negative, and neutral reviews for different drugs, helping to identify the most effective medications. Performance comparison graphs highlight the differences between machine learning models, demonstrating that certain algorithms perform better for this specific dataset. These visualizations make it easier to interpret the results and understand the behavior of the models.

The recommendation system successfully suggests drugs based on aggregated sentiment scores and user ratings. Drugs with a higher number of positive reviews and better ratings are ranked higher in the recommendation list. The system also helps identify drugs with frequent negative feedback, which may indicate potential side effects or inefficiencies. Overall, the results confirm that sentiment analysis combined with machine learning provides an accurate and efficient approach for drug recommendation, supporting better decision-making in healthcare applications.



V. CONCLUSION

The proposed Drug Recommendation System based on sentiment analysis of drug reviews provides an efficient and data-driven approach for recommending effective medications. By leveraging Natural Language Processing and machine learning techniques, the system successfully analyzes large volumes of user-generated reviews to extract meaningful insights about drug performance. The use of preprocessing techniques such as tokenization, stop-word removal, and TF-IDF vectorization enhances the quality of input data, enabling accurate sentiment classification. Among the implemented models, algorithms such as Logistic Regression and Support Vector Machines demonstrate strong performance in classifying sentiments into positive, negative, and neutral categories. The system effectively ranks drugs based on aggregated sentiment scores and user ratings, ensuring reliable recommendations for specific medical conditions. Visualization techniques further help in understanding sentiment distribution and model performance. Overall, the system reduces manual effort, improves decision-making, and supports patients and healthcare professionals in selecting appropriate medications. This project highlights the potential of machine learning and sentiment analysis in healthcare analytics and can be extended with real-time data integration and advanced deep learning models for improved accuracy.

REFERENCES

[1] I. Goodfellow, Y. Bengio, and A. Courville, “Deep Learning,” 2016.

[2] S. Russell and P. Norvig, “Artificial Intelligence: A Modern Approach,” 2020.

[3] C. Cortes and V. Vapnik, “Support-vector networks,” 1995.

[4] T. Hastie, R. Tibshirani, and J. Friedman, “The Elements of Statistical Learning,” 2009.

[5] A. Ng, “Machine Learning,” 2016.

[6] D. Jurafsky and J. Martin, “Speech and Language Processing,” 2021.

[7] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” 2008.

[8] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval,” 1988.

[9] J. Ramos, “Using TF-IDF to determine word relevance,” 2003.

[10] T. Mikolov et al., “Efficient estimation of word representations,” 2013.

[11] J. Pennington, R. Socher, and C. Manning, “GloVe: Global vectors for word representation,” 2014.

[12] F. Chollet, “Keras,” 2015.

[13] M. Abadi et al., “TensorFlow: Large-scale machine learning systems,” 2016.

[14] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” 2011.

[15] L. Breiman, “Random forests,” 2001.

[16] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” 2016.

[17] A. McCallum and K. Nigam, “A comparison of event models for Naive Bayes text classification,” 1998.

[18] C. Bishop, “Pattern Recognition and Machine Learning,” 2006.

[19] Y. Goldberg, “Neural network methods for natural language processing,” 2017.

[20] J. Dean and S. Ghemawat, “MapReduce: Simplified data processing,” 2008.

[21] M. Zaharia et al., “Apache Spark: Unified engine for big data processing,” 2016.

[22] World Health Organization, “Guidelines for safe medication,” 2020.

[23] UCI Machine Learning Repository, “Drug review dataset,” 2019.

[24] U.S. FDA, “Drug safety communication reports,” 2020.

[25] PubMed, “Biomedical literature database,” 2020.

[26] K. Murphy, “Machine Learning: A Probabilistic Perspective,” 2012.

[27] A. Geron, “Hands-On Machine Learning with Scikit-Learn and TensorFlow,” 2019.