WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Machine Learning Based sentiment Analysis System for Personalized Drug Recommendation System

Pugazhvendhan M[1], Srinivasan Sujith[2], Kishore Kumar V[3], Prasanth S[4], Mrs. A.Manju[5]

[1,2,3,4] Department of Computer Science, Kingston Engineering College, Vellore

[5] AP/CSE, Kingston Engineering College, Vellore

Email: [1]**pugazhvendhan68@gmail.com**, [2]**srinivasansujith7@gmail.com**, [3]**kishorev9268@gmail.com**, [4]**prasanth07012@gmail.com**

[5]**amanjudurai@gmail.com**

Abstract

This paper proposes a personalized drug recommendation system leveraging machine learning and sentiment analysis on patient reviews. The system integrates vectorization techniques and classification algorithms to predict review sentiments, combining model outputs with user ratings for effective drug ranking. The best performing configuration—LinearSVC with TF-IDF vectorization—achieves 93% accuracy. Results demonstrate the framework's potential in assisting healthcare professionals with informed prescribing and supporting patients with better medication choices.

Keywords : Drug Recommendation, Sentiment Analysis, Machine Learning, TF-IDF, LinearSVC, Ensemble Learning, Patient Reviews, Feature Extraction, SMOTE, Personalized Healthcare

1. Introduction

Drug recommendation systems are gaining prominence due to their ability to assist healthcare providers by leveraging AI to improve prescription accuracy. With a rise in publicly shared medication experiences, analyzing sentiment in patient reviews has become vital. However, challenges exist due to the complexity of medical terminology and variability in patient feedback. This work presents a sentiment-aware drug recommendation framework that classifies sentiments from drug reviews and uses these insights to rank medications. It addresses gaps in existing systems by combining multiple machine learning models and integrating a weighted scoring system to enhance recommendation reliability.

2. Related Work

Several studies have explored sentiment analysis and drug recommendation frameworks:

- GalenOWL (2012): Introduced ontology-based recommendations using clinical standards (ICD-10, UNII) for semantic integration.
- Sun et al. (2016): Applied semantic clustering on treatment records to recommend regimens based on patient demographics.
- Hassan et al. (Cloud-based CADRE, 2016): Proposed collaborative filtering with tensor decomposition for enhanced drug suggestions.
- Jiugang Li (2018): Developed a CNN-LSTM framework for hashtag recommendations that inspired semantic feature processing in drug reviews.

Despite these advancements, real-time and personalized drug suggestion systems based on patient sentiment remain underexplored.

3. Methodology

3.1 Data Collection & Preparation

- Dataset: Drug review dataset from UCI Machine Learning Repository.
- Preprocessing: Tokenization, stopword removal, lemmatization.

3.2 Feature Engineering

- Techniques: BoW, TF-IDF, Word2Vec, manual extraction.
- Sentiment Scoring: TextBlob and label encoding for sentiment classification.

3.3 Classification Models

- Algorithms: Logistic Regression, Perceptron, Naive Bayes, Decision Trees, Random Forest, LGBM, LinearSVC.
- Best performance: LinearSVC with TF-IDF achieved 93% accuracy.

3.4 Balancing and Ensemble

- SMOTE applied to handle class imbalance.

- Weighted scoring combines sentiment score and review usefulness to rank drugs.

4. System Architecture

The system architecture includes four stages:

- Input Module: Patient review collection and preprocessing.
- Sentiment Classification: Multiple models evaluate sentiment polarity.
- Scoring System: Weighted combination of sentiment score and usefulness.
- Recommendation Engine: Top-N drug suggestions for specific conditions.

5. Results and Discussion

Model Performance:

- TF-IDF + LinearSVC: 93%
- Word2Vec + LGBM: 91%
- BoW + Perceptron: 91%
- Manual + Random Forest: 88%

Key Observations:

- TF-IDF with LinearSVC showed the highest sentiment prediction accuracy.
- Ensemble model ensured balanced recommendations, reducing model bias.
- SMOTE improved classification fairness between positive and negative sentiments.

Stakeholder Benefits:

- Doctors can mitigate prescription errors.
- Patients gain insights from similar user experiences.

6. Conclusion and Future Work

Conclusion:

The proposed sentiment-driven drug recommendation system effectively combines multiple ML models and vectorization methods to produce accurate and personalized drug suggestions. It reduces the burden on healthcare professionals and improves patient awareness.

Future Enhancements:

- Incorporate additional oversampling techniques (ADASYN, SmoteTomek).
- Fine-tune hyperparameters.
- Extend to multilingual reviews with NLP transformers.
- Develop a mobile-friendly, cloud-based deployment model.

7. Limitations and Recommendations

Limitations:

- Dependence on review availability and quality.
- Cold-start problem for lesser-reviewed drugs.

Recommendations:

- Integrate clinical data for hybrid suggestions.
- Use advanced models like BERT for improved language understanding.
- Enhance interpretability through model explainability tools.