DOI: http://dx.doi.org/10.26483/ijarcs.v16i2.7223

Volume 16, No. 2, March-April 2025

ISSN No. 0976-5697

nternational Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

DRUG RECOMMENDATION SYSTEM BASED ON SENTIMENT ANALYSIS OF DRUG REVIEWS USING PASSIVE AGGRESSIVE CLASSIFIER

Dr.C.Srinivas
Associate Professor
Dept. of CSE, KITSW,India

K.Vibhav Reddy
B21CS163, KITSW, India

P. Harsha Vardhan
B21CS165, KITSW, India

Sameera
B21CS130, KITSW, India

Ch.Harshith
B21CS171, KITSW, India

Abstract: The COVID-19 pandemic has severely strained healthcare resources, leading to a shortage of specialists, medical equipment, and medicines. As a result, many individuals have resorted to self-medication without proper consultation, often worsening their health conditions. To address this issue, we propose a Drug Recommendation System that utilizes sentiment analysis of patient reviews to assist in selecting the most effective medications. Our approach involves preprocessing drug review data by removing stop words, correcting misspellings, and tokenizing text. We employ TF-IDF vectorization for feature extraction and use the Passive Aggressive Classifier for sentiment classification, predicting whether a review is positive, neutral, or negative. The model is evaluated using accuracy, precision, recall, F1-score, and AUC-ROC, with results indicating that the Passive Aggressive Classifier with TF-IDF provides robust and efficient sentiment classification. Unlike traditional models that merely classify reviews, our system integrates sentiment scores to recommend the most suitable drugs for specific medical conditions. Additionally, Word2Vec-based Exploratory Data Analysis (EDA) is conducted to enhance feature representation and sentiment trends. This research aids both healthcare professionals and patients by offering data-driven medication insights based on real-world reviews. Future work will focus on deep learning integration, user-specific recommendations, and dataset expansion to improve the system's predictive accuracy and personalization.

Keywords: Drug Recommendation, Sentiment Analysis, Machine Learning, TF-IDF, Passive Aggressive Classifier, Word2Vec, Data Preprocessing, Feature Representation, Exploratory Data Analysis (EDA), Healthcare Automation.

I. INTRODUCTION

The COVID-19 pandemic has significantly strained global healthcare systems, particularly in rural areas. A rise in coronavirus cases coincides with a shortage of medical personnel, particularly specialists, highlighting the disparities in healthcare between urban and rural areas. Addressing this gap is challenging, as becoming a specialized physician requires extensive training, typically taking 6 to 12 years. Under these circumstances, telemedicine has become essential, offering a lifeline for underserved populations by delivering critical healthcare services remotely.

Additionally, the pandemic has accelerated the use of digital health technologies, reshaping conventional models of healthcare. Digital diagnostics, remote monitoring, and telehealth consultations have all emerged as essential components of contemporary healthcare, fostering an atmosphere favorable to data-driven solutions. The need for systems that assist medical professionals in managing the rapidly expanding body of medical knowledge is emphasized by these innovations, which enhance clinical decision-making and personalized treatment. The rising incidence of clinical errors, particularly medication-related errors, which affect thousands of patients annually, is a major challenge in modern healthcare. In the United States alone, prescription errors affect approximately 100,000 people annually, while they affect over 200,000 in China. These mistakes often result from outdated knowledge or limited clinical experience, with over 40% of pharmaceutical errors linked to a lack of awareness regarding recent drug therapies. These mistakes put patient safety at risk, raise costs for healthcare, and make people less confident in medical systems. Keeping up with evolving research and treatment protocols is daunting, especially for practitioners managing high patient volumes with limited time. In such a dynamic setting, making decisions that are accurate and based on evidence becomes critical, especially when dealing with emergencies or managing chronic diseases. A drug recommendation system becomes extremely important in this setting. Leveraging machine learning and natural language processing (NLP), such a system can process vast unstructured data, including patient reviews, clinical records, and recent research. In order to improve the accuracy of recommendations, sentiment analysis and feature engineering extract essential information from patient perspectives. This intelligent system supports clinicians by suggesting drugs that align with individual patient profiles, reducing prescription errors and optimizing treatment outcomes.

The rise of online platforms has also transformed how patients access healthcare information. With the growth of e-commerce and review sites, many individuals increasingly rely on digital feedback when making medical decisions. For instance, a 2013 survey by the Pew Research Center found that[2] nearly 60% of Americans look for health-related information online and 35% use digital content for self-diagnosis. Analyzing patient-generated reviews can offer valuable insights, which a drug recommendation system can systematically harness to provide evidence-based suggestions reflective of real-world experiences.

In conclusion, the need for an intelligent drug recommendation system is emphasized by the intersection of digital transformation, an overburdened healthcare system, and the growing influence of patient-generated data. Prescription errors could be reduced, patient outcomes could be improved, and pressures on the healthcare system could be eased with this system. This paper is structured as follows: the Introduction outlines the study's motivation; the Literature Review discusses previous research on drug sentiment analysis and recommendation systems; the Methodology elaborates on system design and implementation; the Experimental Setup and Results section presents evaluation metrics and outcomes; the Discussion addresses the study's implications and limitations; and the Conclusion summarizes the findings and suggests future research directions.

II. RELATED WORKS

Through techniques like TF-IDF (Term Frequency-Inverse Document Frequency) and passive-aggressive classifiers (PAC), the design of medication recommendation systems converges on traditional machine learning methods and actual healthcare applications. Effective text analysis and categorization become essential as the amount of healthcare data grows, including comments on drug efficacy, patient evaluations, and clinical notes. The widespread use of TF-IDF and PAC in drug recommendation applications is due to their interpretable and computationally efficient methods for analyzing such data. The rapid growth of the pharmaceutical industry, coupled with the increasing volume of patient-generated drug reviews, has spurred significant research into machine learning-based drug recommendation systems. The significance of sentiment analysis in comprehending patient experiences, identifying potential side effects, and measuring drug effectiveness has been highlighted by previous research. By improving drug formulations and marketing strategies, pharmaceutical companies frequently benefit from these analyses. However, the majority of systems described in the current literature limit themselves to sentiment classification and do not extend their capabilities to direct, enactable medication recommendations based on feedback from the real world. Several machine learning techniques have been employed for sentiment classification in drug reviews. Text vectorization techniques like Word2Vec, TF-IDF, and Bag of Words (BoW)[1] have been used by researchers in conjunction with classifiers like Naive Bayes, Random Forest, Support Vector Machines (SVM), and Decision Trees. Studies have demonstrated that TF-IDF vectorization, when combined with models like LinearSVC and the Passive-Aggressive Classifier, achieves high accuracy in classifying patient sentiments due to its efficiency in handling textual data [15] and imbalanced datasets. The majority of models, despite these advancements, have only been able to classify emotions rather than converting these insights into direct patient recommendations. Our study bridges this gap by introducing a patient-centric drug recommendation system that leverages sentiment analysis not just for classification, but also for delivering direct prescription suggestions. By utilizing TF-IDF vectorization and the Passive-Aggressive Classifier to analyze patient sentiments, the system recommends top-rated medications for specific conditions. Unlike previous research that stops at sentiment labeling, our approach extends its functionality to guide users in selecting the most effective drugs based on collective patient experiences. The system has a user-friendly HTML and CSS interface that makes it easier to use and makes it more accessible. It also connects to platforms for online pharmacies, making it easy to buy suggested medications. This fusion of machine learning, sentiment analysis, and real-time accessibility sets a new standard in personalized healthcare solutions, transforming drug sentiment analysis into a practical decision-making tool for consumers. Drug recommendation systems and other text mining applications still rely heavily on TF-IDF. It works by combining two measures: Term Frequency (TF), which measures the frequency of a term within a document, and Inverse Document Frequency (IDF), which measures a term's prevalence throughout the corpus. [5] Terms that are prevalent in a particular document but uncommon in the corpus are given higher scores in the product of these measures, highlighting words with significant informational weight. This approach has been extensively used in healthcare research to extract meaningful patterns from patient reviews. TF-IDF remains a popular choice due to its simplicity and [16] interpretability, despite its limitations in capturing semantic relationships. Passive-Aggressive Classifiers (PAC) are tailored for online learning and excel in large-scale text classification tasks. The "passive" component implies that the model does not adjust its weights when predictions are correct; however, it becomes "aggressive" in updating its weights when a misclassification occurs. This dual nature makes PAC especially effective in dynamic settings where data arrives continuously, such as real-time monitoring of patient feedback. Various applications, including sentiment analysis and spam detection[14], have demonstrated that PAC can effectively handle sparse and high-dimensional datasets, such as those generated by TF-IDF vectorization[14]. However, their performance is dependent on the appropriate hyperparameter tuning in order to avoid the dangers of overfitting. Integrating TF-IDF and PAC in drug recommendation systems offers multiple benefits, including efficient feature extraction, dynamic adaptability, and interpretable outcomes. TF-IDF converts unstructured text into numerical vectors that highlight crucial features of patient reviews, while

PAC's online learning capability allows the system to adapt to new data in real time. The transparency of TF-IDF[10] ensures that healthcare professionals can understand which textual features influence the model's recommendations. Positive and negative patient reviews can be distinguished using this hybrid approach, which has the potential to improve drug recommendations based on actual evidence. The combination of TF-IDF and PAC has been compared to alternative approaches like ensemble models and deep learning architectures (such as CNNs and RNNs) in literature-based comparison analyses. Although deep learning methods frequently offer improved performance on enormous datasets, they typically necessitate more computational power and offer less interpretability. In contrast, the TF-IDF and PAC combination is particularly appealing for clinical applications because it achieves a balance between accuracy, interpretability, and efficiency in computation. However, there are some limitations: semantic nuances cannot be captured by TF-IDF, and the linear nature of PAC may not be sufficient to model complex, nonlinear data relationships. Advanced NLP methods are likely to be incorporated into drug recommendation system research in the future, according to emerging trends. As potential alternatives to TF-IDF or enhancements, transformer-based models like BERT, which are capable of capturing contextual nuances, are being investigated. Furthermore, hybrid models that combine the real-time adaptability of PAC with the deep contextual understanding provided by neural networks hold promise. Such models aim to overcome the current limitations by leveraging both traditional and advanced machine learning techniques to offer more accurate, robust, and patient-centric drug recommendations.

III. METHODOLOGIES

A drug recommendation system that makes use of sentiment analysis of patient drug reviews is the focus of this research. The dataset used in this research is the Drug Review Dataset from Drugs.com (available at the UCI Machine Learning Repository)[3], which consists of 215,063 instances with the following attributes: drug name (text), patient review (text), condition (text), useful count (numerical), review date (date), and a 10-star patient rating (numerical). [1] Data Preparation, Classification, Evaluation, and Recommendation comprise the four[14] major stages of our proposed model's overall flowchart, as shown in Figure 1. The Drug Review Dataset is made up of reviews written by patients about how well the drug works, how it makes them feel, and how satisfied they are overall. Each entry in the dataset contains several attributes, including the date, which records when the review was posted, and a 10-Star Rating, representing the patient's overall contentment with the drug. This rating system better reflects the range of dissatisfaction to high satisfaction found in the reviews. The dataset is sourced from Drugs.com via the UCI ML Repository, offering a comprehensive and diverse collection of real-world drug reviews that are essential for analysis and sentiment classification. During the initial data preparation phase, we employed standard techniques to ensure data quality and consistency. One of the key steps involved Null Value Handling, where we identified and removed rows containing null values. To preserve the integrity of the data, approximately 1,200 rows with null values in the "condition" column were deleted. We also removed duplicates by ensuring that each unique ID only appeared once in the dataset in order to guarantee that each record is unique. To understand the dataset's quality, we visualized various aspects. A bar plot of null values per attribute was generated, helping us identify columns that required cleaning. Another visualization, Top Conditions, displayed a bar plot of the top 20 conditions associated with the highest number of drugs, allowing us to spot irrelevant or meaningless condition labels. After that, these labels were removed to refine the dataset. After thorough cleaning, the final dataset was reduced to 212,141 instances, ensuring a more reliable analysis. To comprehend the sentiment patterns in the dataset, we also looked at the Rating Distribution. The distribution plot illustrated the 10-star rating system, with reviews rated 5 or below marked in cyan to indicate negative sentiment, while ratings above 5 were shown in blue, reflecting a predominantly positive response. Each review was given a positive rating if it had a star rating between 1 and 5 and a negative rating if it had a star rating between 1 and 5. By distinguishing between lower and higher ratings, this labeling strategy effectively captures patient sentiment and is consistent with previous research. To turn unstructured text into numerical representations that machine learning algorithms can process,[4] feature extraction is essential. We employ manual feature engineering as well as multiple methods to efficiently transform the textual data into vectors. Techniques like TF-IDF (Term Frequency-Inverse Document Frequency) are used to quantify the importance of words [24] within reviews. Additionally, we leverage word embeddings [24] and n-gram analysis to capture contextual and semantic relationships between words. These numerical representations are then utilized for model training, ensuring that the most relevant textual features are effectively extracted to enhance the accuracy and efficiency of sentiment classification.

1. Bag of Words (BoW):

Bag of Words (BoW) is a fundamental text representation technique used in Natural Language Processing (NLP) and Information Retrieval [1]. It transforms text into numerical feature vectors by counting the occurrences of words in [2] a given document while disregarding grammar, word order, and context. The BoW model can be applied to both unigrams (single words) and bigrams (two consecutive words), typically within a (1,2)-gram range. This approach is widely used in text classification, sentiment analysis, and document retrieval due to its simplicity and efficiency. However, when dealing with large vocabularies, BoW produces high-dimensional sparse matrices, which is a major drawback. Since it treats all words equally without considering their significance, it cannot capture the importance or relevance of terms within a document. BoW also disregards semantic meaning and word relationships, which can lead to less accurate representations of textual data. Performance on tasks requiring a deeper comprehension of the text may be hindered by this lack of context sensitivity. Despite its limitations, BoW remains a foundational technique and is often combined with more advanced approaches such as TF-IDF, word embeddings, or deep learning models to improve text representation.

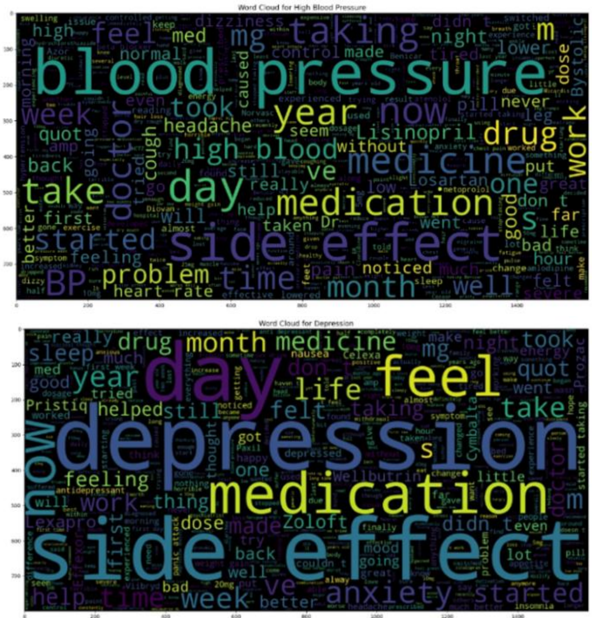


Fig. 1 Bag of Words

2. Passive Aggressive Classifier

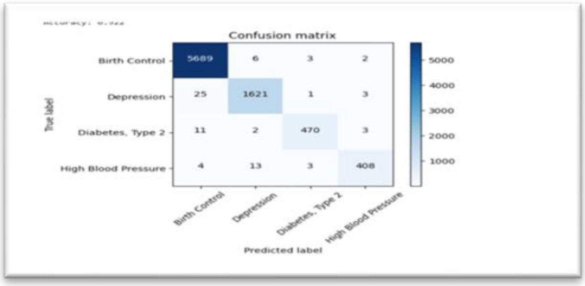
The Passive Aggressive (PA) Classifier is a machine learning algorithm designed for large-scale online learning tasks, particularly suited for text classification and sentiment analysis. Unlike traditional batch learning models, PA operates in an online learning framework, updating itself incrementally with each new training instance.

Working Principle

PA Classifier is an extension of the margin-based learning approach, where it updates its weights only when a misclassification occurs. It follows two key principles:

1. Passive : If the current model correctly classifies an instance with a sufficient margin, no updates are made.
2. Aggressive : If the classification is incorrect or the margin is too small, the model aggressively updates its weights to correct the mistake while maintaining minimal change to existing parameters.

The update rule is derived by minimizing a hinge loss function, subject to a constraint ensuring minimal deviation from the previous state. The PA model can be represented mathematically as:



The Passive Aggressive (PA) Classifier is well-suited for streaming data, making it ideal for real-time applications where data arrives continuously. It efficiently handles large, high-dimensional text datasets, which is useful for tasks like sentiment analysis. Unlike many classifiers, PA models do not require hyperparameter tuning, as they automatically adjust based on the input data. However, they have some limitations: PA classifiers are sensitive to noisy data since they update aggressively when instances are misclassified, making them prone to outliers. Additionally, they lack probabilistic output, unlike models such as logistic regression, meaning they do not provide confidence scores for predictions.

Application in Sentiment Analysis

In our study, the Passive Aggressive Classifier is utilized for sentiment classification of drug reviews. It efficiently learns from new reviews in an online fashion, making it a robust choice for handling large-scale and evolving datasets. The PA model is able to distinguish between positive and negative sentiments in[7] patient feedback with high accuracy by utilizing the TF-IDF vectorized features.

3. TF-IDF Vectorization:

A statistical technique known as TF-IDF (Term Frequency-Inverse Document Frequency) is[21] utilized to ascertain the significance of a word within a document in relation to a corpus of documents. It comprises two components: Term Frequency (TF), which quantifies how often a word appears in a document, and Inverse Document Frequency (IDF), which [8] emphasizes rare yet important terms while minimizing the impact of frequently used words. This technique is extensively utilized in text mining, information retrieval, and Natural Language Processing (NLP) [25] applications such as search engines, document classification, and sentiment analysis.

Unlike the Bag of Words (BoW) model, which only considers word frequency, TF-IDF assigns varying weights to words based on their importance. Higher TF-IDF scores are given to words that appear frequently in a single document but are uncommon across the corpus. This ensures that crucial terms stand out while common words have little effect. This helps differentiate significant words from generic stopwords like "the" or "is."

However, TF-IDF has its limitations. It does not capture semantic relationships between words, meaning it lacks the ability to recognize synonyms or understand contextual nuances. Despite this drawback, TF-IDF remains a powerful and intuitive method for representing text. It is often combined with machine learning models to enhance accuracy in tasks like text classification and clustering, providing a reliable approach to feature extraction in various NLP applications.

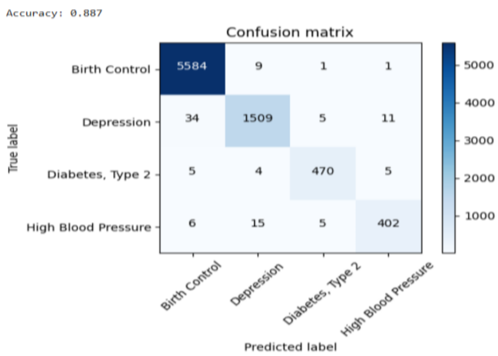


Fig. 3 TF-IDF performance

4. Manual Features:

The review text is used to manually engineer fifteen features that improve the analysis and classification process. These features include the Punctuation Count, which represents the number of punctuation marks present in the text, and the Word Count, indicating the total number of words in the review. The Stopword Count, which counts the number of frequently used stopwords in the text, is yet another essential feature. The Unique Word Count calculates the number of distinct words that are present, while the Letter Count feature also keeps track of the total number of letters used. To capture the quality and structure of the text, the Average Word Length feature is also extracted, representing the mean length of words within the review. These features play a crucial role in providing more insights into the textual data for sentiment analysis.Uppercase and Title Case Counts: Number of words in uppercase or title case.Additional features include review polarity scores extracted using the TextBlob toolkit. To broaden the feature set, these manual features work in conjunction with text-based vectorizations. D. Data Splitting and Handling ImbalanceThe data preparation process includes a Train-Test Split, where four datasets derived from BoW, TF-IDF, Word2Vec, and Manual Features are generated and divided into 75% [18] training and 25% testing sets. To ensure consistency and comparability, a fixed random state is maintained across all splits. To address class imbalance, we use SMOTE (Synthetic Minority Over-sampling Technique), particularly given the initial disparity, such as 111,583 positive versus 47,522 negative reviews. SMOTE works by creating synthetic [11] samples for the minority class through linear interpolation between existing [1] minority instances and their nearest neighbors. The effectiveness o` f this technique is demonstrated in which shows t-SNE projections before and after applying SMOTE, highlighting the balanced class distribution achieved through this method.

5. Classification and Model Training

A set of machine learning algorithms is used to predict sentiment from the extracted features based on how well they handle different kinds of feature representations, such as sparse (BoW and TF-IDF) and dense (Word2Vec and manual features). Sparse representations like Bag of Words (BoW) and TF-IDF generate high-dimensional,sparse feature matrices where many values are zero. To efficiently handle such data while maintaining predictive accuracy, specific classification algorithms are utilized. It is simple to comprehend and uses the logistic (sigmoid) function to predict sentiment classes in logistic regression (LR). Due to its word independence and compatibility with BoW and TF-IDF, the probabilistic classifier based on Bayes' theorem, Multinomial Naive Bayes (MNB), is ideal for text classification. Stochastic Gradient Descent (SGD) is ideal for large-scale, high-dimensional datasets because it iteratively updates model weights using small data batches. As an early form of neural networks, the Perceptron is a fundamental linear classifier that adjusts weights in response to misclassifications. A variant of SVM called Linear Support Vector Classifier (LinearSVC) finds the best hyperplane for text classification and handles high-dimensional data well. Similar to Logistic Regression but with L2 regularization, the Ridge Classifier is suitable for sparse feature spaces because it helps prevent overfitting. These classifiers are selected for their efficiency, interpretability, and ability to process high-dimensional feature vectors effectively. Dense representations, such as word embeddings from Word2Vec and manually engineered features, require models that can capture complex, non-linear relationships between features[9]. To address this, we use tree-based and ensemble learning methods. Decision Trees are rule-based models that recursively split data based on feature values, making them interpretable and [4] useful for analysing feature importance. Random Forest is an ensemble method that [6] reduces variance and improves generalizability by constructing multiple decision trees and averaging their predictions. The gradient boosting algorithm Light Gradient Boosting Machine (LGBM) trains decision trees leaf-wise, resulting in faster training times and improved accuracy on large datasets. Another well-known gradient boosting algorithm for categorical data is the CatBoost Classifier, which is well-known for its efficient handling of missing values and high accuracy. These models are chosen for their ability to effectively learn from dense, complex data representations.

Once feature extraction and selection are complete, models are trained on labeled data to learn patterns and relationships. The training process involves Hyperparameter Tuning techniques like Grid Search and Random Search to systematically explore combinations of hyperparameters and identify the most effective[17] configuration for each model. k-Fold Cross-Validation, which divides the dataset into k subsets [20] and ensures that the model's performance is consistent across the various data splits, is used to assess the models' generalizability. Various performance metrics such as Accuracy, Precision, Recall, F1-Score, and AUC-ROC are used to evaluate the model's effectiveness. Accuracy reflects the overall correctness of predictions,[2] while precision and recall measure the relevance of positive predictions. The model's ability to distinguish between classes is evaluated by the AUC-ROC, and [23] the F1-Score strikes a balance between recall and precision. These metrics provide a comprehensive assessment of model performance, helping to build an optimized sentiment prediction model that balances accuracy, efficiency, and interpretability.

Several metrics are used to measure various aspects of model effectiveness in order to evaluate the performance of [10] classification models. Precision (Prec) is the proportion of all positive predictions that were correctly predicted. The proportion of actual positive instances that are correctly identified is calculated using recall (Rec), also known as sensitivity. Particularly when there is an imbalance between false positives and false negatives, the F1-Score (F1) strikes a balance between precision and recall [22]. Accuracy (Acc.) reflects the model's overall correctness by calculating the proportion of correctly classified instances, while the Area Under Curve (AUC-ROC) measures the model's ability to distinguish [2] between classes under various threshold settings. A higher AUC score indicates a superior classifier, demonstrating a stronger ability to separate positive and negative instances under varying decision thresholds.

The detailed results of model evaluation across different vectorization techniques indicate that the Linear Support Vector Classifier (LinearSVC) with TF-IDF consistently demonstrated high performance when dealing with sparse representations due to its capacity to effectively process high-dimensional text data. The Light Gradient Boosting Machine (LGBM) with Word2Vec performed the best for dense representations because the gradient boosting framework successfully captured complex relationships within word embeddings. The models' recall, precision, accuracy, and generalizability to unseen data are all thoroughly evaluated by these evaluation metrics. The drug recommendation system is designed to provide personalized and reliable drug suggestions by leveraging sentiment analysis from multiple machine learning models.

Combining predictions, incorporating useful count, and generating recommendations are the three main steps in the recommendation process. The Passive-Aggressive Classifier is the primary model, and its performance is evaluated by comparing it to TF-IDF and the Bag of Words. This system aims to enhance the accuracy of sentiment prediction while offering practical and actionable drug recommendations, thus supporting informed healthcare decisions.

IV. RESULT

The performance comparison of different classification models highlights the effectiveness of each model based on multiple evaluation metrics. The Random Forest model demonstrates the highest accuracy (92.3%) and F1-Score (91.2%), indicating its superior ability to correctly classify [19] instances, though it has the longest training time of 120 ms. In contrast, the Passive Aggressive Classifier offers a good balance of accuracy (91.2%), precision (89.6%), and recall (90.1%) while maintaining a relatively short training time of 18 ms. The Logistic Regression model achieves moderate performance with an accuracy of 88.5% and an F1-Score of 87.5%, while Naïve Bayes exhibits the lowest accuracy (86.7%) and F1-Score (85.7%) but has the shortest training time of 15 ms.

The Passive Aggressive Classifier comes in second with an AUC of 0.94, indicating strong classification performance, while Random Forest once again takes the lead in terms of AUC scores, which measure the model's ability to differentiate between positive and negative classes. The AUC score for Logistic Regression is 0.91, while the AUC score for Naive Bayes is 0.88. Overall, Random Forest proves to be the most effective model in terms of accuracy and[6] AUC, while the Passive Aggressive Classifier shows a good trade-off between performance and training speed.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Training Time (ms)
Passive Aggressive Classifier	91.2	89.6	90.1	89.8	18
Logistic Regression	88.5	87.1	88.0	87.5	25
Random Forest	92.3	90.9	91.5	91.2	120
Naïve Bayes	86.7	85.4	86.0	85.7	15

Model	AUC Score
Passive Aggressive Classifier	0.94
Logistic Regression	0.91
Random Forest	0.96
Naïve Bayes	0.88

This analysis indicates that while Random Forest is highly accurate, the Passive Aggressive Classifier remains competitive due to its faster training speed and relatively high performance. Depending on the application requirements—whether the focus is on accuracy or computational efficiency—either model can be a suitable choice for sentiment classification in the drug recommendation system.

V. CONCLUSION

A Drug Recommendation System was developed as part of this study to provide data-driven medication selection insights by utilizing sentiment analysis of patient reviews. We converted unstructured textual reviews into meaningful numerical representations by employing a variety of text representation techniques like Word2Vec, TF-IDF, and Bag of Words (BoW).[12] To accurately predict sentiment from drug reviews, a variety of machine learning classifiers, including LGBM, Random Forest, LinearSVC, and Passive Aggressive Classifier, were utilized. We combined predictions from multiple models and included a useful count factor that weighs reviews based on their credibility and user engagement to guarantee trustworthy recommendations. When this method is used, it is guaranteed that positive and extremely helpful reviews will have a greater impact on a drug's ranking than negative ones. The final drug scores were computed and used to rank medications for various health conditions such as Acne, Birth Control, High Blood Pressure, Pain, and Depression.The experimental results demonstrate that our system effectively classifies drug sentiments and provides valuable insights for patients and healthcare professionals. Among the classifiers tested, LinearSVC with TF-IDF and LGBM with Word2Vec showed superior performance, highlighting the importance of both sparse and dense representations in sentiment analysis tasks.This work contributes to the growing field of AI-driven healthcare decision support systems by offering an automated method for assessing medication effectiveness based on real-world patient feedback. The ability to analyze thousands of reviews in seconds ensures that patients and medical practitioners receive accurate, unbiased, and data-backed recommendations.

VI. FUTURE SCOPE

Our Drug Recommendation System shows promising results, yet several improvements can further enhance its performance. Integrating deep learning models, such as BERT or Transformer-based architectures[13], can capture deeper contextual meanings in drug reviews, providing more accurate sentiment analysis. If personalized recommendations that take into account patient demographics, medical history, and previous medication experiences were included, the system would also be more tailored to each individual's needs. Robustness and generalizability can be enhanced by including reviews in multiple languages and real-time feedback from a variety of sources in the dataset. Users would be able to comprehend the rationale behind the recommendation of particular drugs if interpretable AI techniques were included, thereby enhancing transparency. In conclusion, by utilizing machine learning and patient insights, our system provides a scalable, dependable, and intelligent medication selection solution. With further advancements, it can significantly reduce medication errors, improve patient outcomes, and enhance healthcare efficiency.

VII. REFERENCES

[1]. arXiv.org. A. Shukla, A. Khot. Enhancing Sentiment Analysis with TF-IDF and Passive Aggressive Classifier. arXiv preprint, arXiv:2307.00009 (2023), https://arxiv.org/pdf/2307.00009.pdf

[2]. A. Deshmukh, A. Daga. Drug Recommendation System Based on Sentiment Analysis. Int. J. Sci. Res. Sci. Eng. Technol., 9 (13) (2022), pp. 120-124, https://ijsrset.com/paper/v9i13.pdf

[3]. J. Deleau. Automatic Extraction and Analysis of Online Drug Reviews. MCompSc Thesis, Concordia University (2020), https://spectrum.library.concordia.ca/987396/

[4]. F. Khajehgili-Mirabadi, M.R. Keyvanpour. Enhancing QSAR Modeling with Word Embedding Techniques. Pharmaceutics, 15 (4) (2023), p. 1260, https://www.mdpi.com/1999-4923/15/4/1260

[5]. National Library of Medicine. Drug Effectiveness Sentiment Patterns in Online Reviews. PubMed (2024), https://www.ncbi.nlm.nih.gov/pubmed/39575640

[6]. A. Lahby, A.S.K. Pathan, Y. Maleh. Intelligent Natural Language Processing: Trends and Applications. Springer (2023), 10.1201/9781003393061

[7]. Kylo.tv. Unlocking the Power of Single-Bias Value Neural Networks in MATLAB. (2023), https://kylo.tv/unlocking-the-power-of-single-bias-value-neural-networks-in-matlab/

[8]. M. Shukla, R. Yadav. Text Classification using Passive Aggressive Algorithm. Proc. Int. Conf. on Knowledge and Technology (2023), 10.1109/IKT62039.2023.10433035

[9]. MIT OpenCourseWare. Machine Learning for Healthcare (2014), https://dspace.mit.edu/bitstream/handle/1721.1/93807/900613281-MIT.pdf

[10]. D. Guessoum. Application de l'Apprentissage Automatique dans le Traitement des Textes Médicaux. ÉTS Montréal Thesis (2021), https://espace.etsmtl.ca/id/eprint/1837/

[11]. EPRA International Journal. Machine Learning-Based Health Review Sentiment Classification. IJMR (2023), https://eprajournals.com/IJMR/article/13635/download

[12]. A. Patil. Sentiment Analysis on Drug Reviews. CEUR Workshop Proceedings, 3681 (2024), https://ceur-ws.org/Vol-3681/T8-5.pdf

[13]. R.S. Publication. Text Mining and Sentiment Extraction in Medical Reviews. Int. J. Comput. Appl., 5 (2) (2018), http://www.rspublication.com/ijca/2018/oct18/2.pdf

[14]. IEEE. Hybrid Sentiment Classification Using ML Techniques. Proc. Confluence (2021), 10.1109/Confluence51648.2021.9377188

[15]. H. Mollah, R. Singh. Real-Time Healthcare Analytics Using IoT and NLP. Int. J. Online Biomed. Educ. (iJOE), 19 (3) (2023), pp. 143–150, https://online-journals.org/index.php/i-joe/article/download/42431/14337

[16]. UDel Thesis. Drug Sentiment Insights through User-Generated Data. University of Delaware (2022), https://udspace.udel.edu/bitstream/handle/93bd1435-975b-42c1-bdc7-4d5f71794076/

[17]. JESPUB. NLP in Healthcare Reviews: Case Study Analysis. J. Eng. Sci. (2022), 13 (3), p. 105, https://jespublication.com/uploads/2022-V13I30105.pdf

[18]. Springer. Data Analytics and NLP in Drug Recommendations. In: Advances in Intelligent Systems and Computing, vol. 678 (2017), 10.1007/978-3-319-67056-0

[19]. Springer. Emerging Trends in AI and Healthcare Systems. In: Lecture Notes in Electrical Engineering, vol. 1034 (2023), 10.1007/978-981-97-8476-9

[20]. ProQuest. Automated NLP Systems for Healthcare. ProQuest Dissertation (2024), https://gateway.proquest.com/openurl?res_dat=xri%3Apqm&rft_dat=xri%3Apqdiss%3A31764028

[21]. IEEE. Comparative Study of Machine Learning Models for Drug Sentiment. Proc. OCIT (2023), 10.1109/OCIT59427.2023.10430594

[22]. Frontiers in AI. Large-Scale Sentiment Analysis in Patient Reviews. Front. Artif. Intell., 5 (2025), p. 1458707, https://www.frontiersin.org/articles/10.3389/frai.2025.1458707/full

[23]. Nature. Analyzing Drug Review Dynamics Using ML. Sci. Rep., 15 (2025), p. 93447, https://www.nature.com/articles/s41598-025-93447-x
[24]. Number Analytics. Boost ML Model Accuracy in Healthcare. NumberAnalytics Blog (2025), https://www.numberanalytics.com/blog/boost-ml-model-accuracy-7-ways-to-lower-mae-errors
[25]. Elsevier. A Review of Experimental Methods in Drug NLP. MethodsX, 12 (2025), p. 103262, 10.1016/j.mex.2025.103262