Journal of Artificial Intelligence and Emerging Technologies
Volume 2, Issue 7, pp 6-10, July 2025
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2025.207002
Available Online at https://www.jaiet.com

# Development of a Machine Learning and NLP-Based Framework for Automated Drug Recommendations

[1]Gopinath M, [2]Manikandan D

[1,2]Department of Computer Science and Engineering (Data Science), Gates Institute of Technology, Gooty, Andhra Pradesh, India

***Abstract:*** The increasing complexity of healthcare decision-making and the growing volume of medical literature present significant challenges in identifying optimal drug therapies for patients. This study proposes the development of an integrated Machine Learning (ML) and Natural Language Processing (NLP) framework for automated drug recommendation, designed to assist healthcare professionals in clinical decision-making. The framework leverages NLP techniques to extract relevant information from unstructured medical texts, such as electronic health records, clinical notes, and research articles, while machine learning algorithms analyze patient-specific data, including demographics, medical history, and symptom profiles. By combining predictive analytics with semantic understanding, the system recommends personalized drug therapies that are clinically appropriate and contextually relevant. The framework also incorporates a feedback loop for continuous learning, enabling it to improve accuracy over time as new patient data and medical knowledge are incorporated. Experimental evaluation demonstrates that the proposed system significantly enhances the efficiency and precision of drug recommendation, reduces potential medication errors, and provides actionable insights to healthcare providers. This research highlights the potential of AI-driven solutions in transforming clinical decision support systems, promoting evidence-based medicine, and improving patient outcomes through intelligent, automated drug guidance.

***Keywords:*** Drug, Recommender System, Machine Learning, NLP, Smote, Bow, TF-IDF, Word2Vec, Sentiment analysis.

## I. INTRODUCTION

With the number of corona virus cases growing exponentially, the nations are facing a shortage of doctors, particularly in rural areas where the quantity of specialists is less compared to urban areas. A doctor takes roughly 6 to 12 years to procure the necessary qualifications. Thus, the number of doctors can't be expanded quickly in a short time frame. A Telemedicine framework ought to be energized as far as possible in this difficult time.

[1]. Clinical blunders are very regular nowadays. Over 200 thousand individuals in China and 100 thousand in the USA are affected every year because of prescription mistakes. Over 40% medicine, specialists make mistakes while prescribing since specialists compose the solution as referenced by their knowledge, which is very restricted [2][3]. Choosing the top-level medication is significant for patients who need specialists that know wide-based information about microscopic organisms, antibacterial medications, and patients [6].

Every day a new study comes up with accompanying more drugs, tests, accessible for clinical staff every day. Accordingly, it turns out to be progressively challenging for doctors to choose which treatment or medications to give to a patient based on indications, past clinical history. With the exponential development of the web and the web- based business industry, item reviews have become an imperative and integral factor for acquiring items worldwide. Individuals worldwide become adjusted to analyse reviews and websites first before settling on a choice to buy a thing.

While most of past exploration zeroed in on rating expectation and proposals on the E-Commerce field, the territory of medical care or clinical therapies has been infrequently taken care of. There has been an expansion in the number of individuals worried about their well-being and finding a diagnosis online.

## II. RELATED WORK

From the existing literature, different health recommender systems (HRS) are available.

Collaborative filtering utilizes past user behavior to

Copyright © 2025 JAIET All Rights Reserved

Journal of Artificial Intelligence and Emerging Technologies
Volume 2, Issue 7, pp 6-10, July 2025
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2025.207002
Available Online at https://www.jaiet.com

examine similar profiles and determine preferences to make clear recommendations. A hospital recommendation system was proposed by Fedelucio et al. [14] based on the treatments, consulted physicians, hospitals, and patient health indicators of a patient. An alternative hybrid recommender system based on available information on family doctors and available patients was suggested [15].

Various HRS help to support medical treatment and prognosis [16]. Recommendations made on content- based filtering are dependent on specific features only. Different features selected using rough set feature reduction can predict diabetes [17]. Content-driven models are used to evaluate radial doses and weights for elements in cancer treatments [18]. It is reported that content-based models achieve a better performance than traditional models in predicting the risk of heart attack [19].

A model called iCARE uses collaborative filtering and hybrid learning to predict disease risk based on a patient's previous illnesses [20]. The risks of delivery for pregnant women can be predicted using a collaborative filtering algorithm that includes Mahalanob is distance and fuzzy membership [21]. Ontologies and methods of problem-solving are fundamental components of knowledge-based systems [22]. Based on the knowledge of users and products, knowledge-based filtering selects products that are suitable for users [23]. Meanwhile, hybrid systems combine different filtering approaches [24].

Demographic filtering offers recommendations based on demographic data such a sage, gender, nationality, and residency [25]. In medical emergencies such as the COVID-19pandemic, older people have a higher risk of complications and contracting serious illnesses if they are untreated. Through information filtering, the HRS can handle such emergencies by collecting patient messages and recommending treatment [26]. With the help of the patient's demographic information, these messages for smoking cessation users used hybrid filtering to assess similarity.

A semantic web is a fast-evolving technology that utilizes a content-based recommendation system with machine-readable annotations [27]. The social-based filtering algorithm considers information about an individual's neighborhood, along with similar tastes [28].To prescribe the most appropriate treatment to patients, semantic clustering assesses the similarities between records, taking into account the patient's demographics, location, and medical complications [29].

The DRS offers medicine based on patient reviews using sentiment analysis and feature engineering. The risk level classification identifies a patient's immune system and recommends medicines if the patient has a low immune system [30]. Doulaverakis et al. [31] proposed GalenOWL, a semantic-driven online framework with the help of a specialist to manage drug recommendations based on the past profile of the patient. By considering worldwide standards such as ICD-10 and UNII, this framework converts clinical data and drug interactions to ontological terms. Cloud-assisted drug recommendation (CADRE)also considers the patient's side effects and shifts to the cloud to advance the quality of the patient's experience [32]. No particular DRS system was developed for the COVID-19 emergency. Therefore, we aimed to develop a DRS modeling framework by incorporating a stacked artificial neural network (ANN) for the fair and safe usage of drugs in pandemics.

## III. PROPOSED SYSTEM

The proposed Drug Recommendation System represents a state-of-the-art integration of machine learning and natural language processing, addressing the complexities of personalized drug recommendations. This intelligent system amalgamates diverse datasets from reputable sources, employing advanced natural language processing techniques for nuanced analysis of unstructured patient reviews. A hybrid machine learning model, blending collaborative and content-based filtering, is implemented to provide accurate and personalized drug suggestions based on individual medical conditions and preferences. The user interface is designed for simplicity, allowing users to input medical information and visualize detailed drug information for informed decision-making. Ethical considerations are paramount, with strict privacy measures, bias mitigation, and compliance with regulations. The system operates on a scalable infrastructure, deploying robust security measures and continuous monitoring for optimal performance. A feedback loop ensures adaptability and iterative improvement, while comprehensive documentation and user support mechanisms facilitate effective system utilization.

Here's a structured outline for a Proposed System section on an "Integrated ML with NLP Framework for Drug Recommendation":

Proposed System: Integrated ML with NLP Frame work for Drug Recommendation

The proposed system integrates Machine Learning (ML) and Natural Language Processing (NLP) techniques to create an intelligent, automated framework capable of recommending

Copyright © 2025 JAIET All Rights Reserved

Journal of Artificial Intelligence and Emerging Technologies
Volume 2, Issue 7, pp 6-10, July 2025
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2025.207002
Available Online at https://www.jaiet.com

appropriate drugs based on patient data, symptoms, medical history, and clinical notes.

**1. System Architecture Overview**

The framework consists of the following key components:

Data Collection Module: Aggregates structured and unstructured medical data including EHRs (Electronic Health Records), doctor notes, prescriptions, and lab results.

Preprocessing Unit:

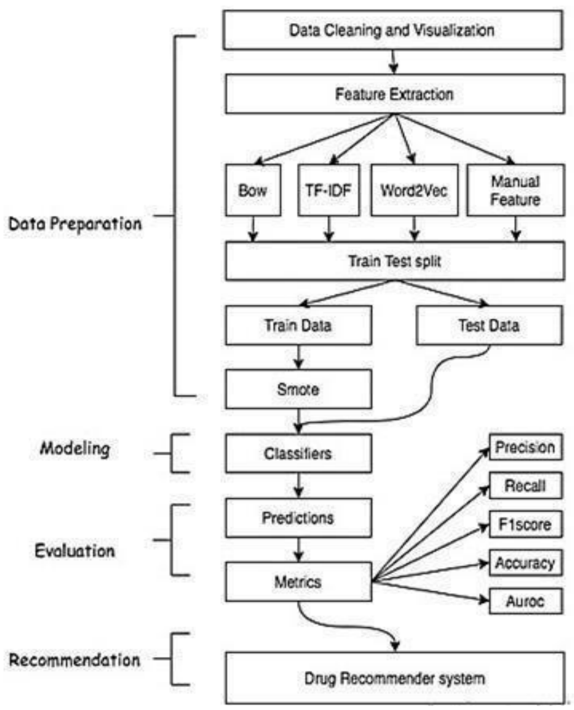
Structured Data: Normalization and missing value imputation.

Unstructured Data: NLP techniques (e.g., tokenization, POS tagging, named entity recognition) are applied to extract medical entities and relationships.

Feature Extraction and Representation:

Utilizes word embeddings (e.g.,Word2Vec, BioBERT) for textual data.

Encodes clinical data into feature vectors suitable for ML models.




**2. Machine Learning Layer**

Model Selection: Supervised learning models such as Random Forest, XGBoost, or deep learning models like LSTM or BERT-based architectures for sequence modeling.

Training: Uses annotated datasets to train models to recognize patterns and predict suitable drugs.

Evaluation: Employs metrics such as accuracy, precision, recall, F1-score, and AUC-ROC.

**3. Drug Recommendation Engine**

Predicts top-N most effective drugs based on:

Patient symptoms

Diagnoses and comorbidities Historical treatment responses

The system supports real-time querying and updates recommendations based on new inputs.

**4. User Interface and API**

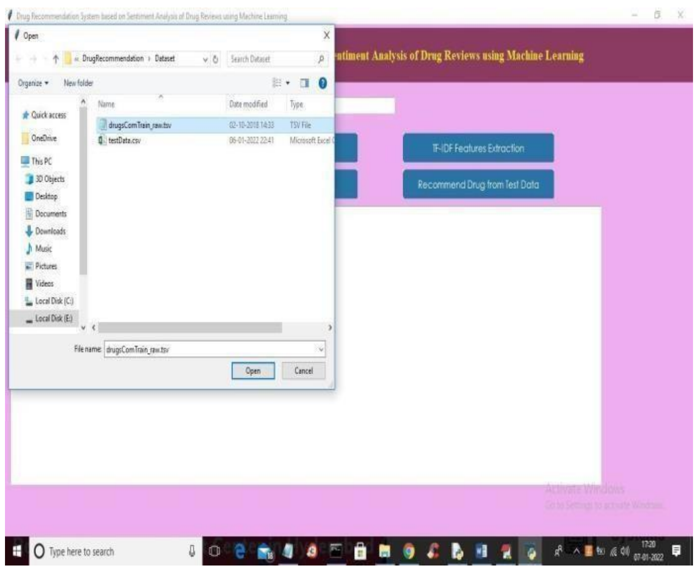
Provides a user-friendly interface for healthcare professionals.API endpoints enable integration with hospital management systems or third-party healthcare apps.

**5. Security and Compliance**

Ensures data privacy through encryption and access control.

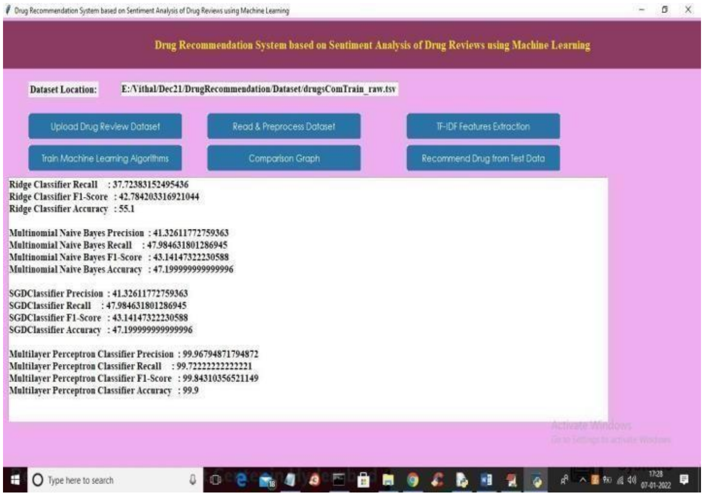
Compliant with healthcare standards such as HIPAA and HL7.

## IV. RESULT




Copyright © 2025 JAIET All Rights Reserved

Journal of Artificial Intelligence and Emerging Technologies
Volume 2, Issue 7, pp 6-10, July 2025
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2025.207002
Available Online at https://www.jaiet.com




## V. CONCLUSION

Reviews are becoming an integral part of our daily lives; whether go for shopping, purchase something online or go to some restaurant, we first check the reviews to make the right decisions. Motivated by this, in this research sentiment analysis of drug reviews was studied to build a recommender system using different types of machine learning classifiers, such as Logistic Regression, Perceptron, Multinomial Naive Bayes, Ridge classifier, Stochastic gradient descent, Linear SVC, applied on Bow, TF-IDF, and classifiers such as Decision Tree, Random Forest, Lgbm, and Cat boost were applied on Word2Vec and Manual features method. We evaluated them using five different metrics, precision, recall, f1score, accuracy, and AUC score, which reveal that the Linear SVC on TF-IDF outperforms all other models with 93% accuracy. On the other hand, the Decision tree classifier on Word2Vec showed the worst performance by achieving only 78% accuracy. We added best-predicted emotion values from each method, Perceptron on Bow (91%), Linear SVC on TF-IDF (93%), LGBM.

## REFERENCES

[1] Battineni, G.; Sagaro, G.G.; Chintalapudi, N.; Di Canio, M.; Amenta, F. Assessment of Awareness and Knowledge on Novel Coronavirus (COVID-19) Pandemic among Seafarers. Healthcare 2021, 9, 120.

[2] Goh, J.M.; Gao, G.; Agarwal, R. The creation of social value: Can an online health community reduce rural–urban health disparities? MIS Q. 2016, 40, 247–263.

[3] Cook, S.F.; Bies, R.R. Disease Progression Modeling: Key Concepts and Recent Developments. Curr. Pharmacol. Rep. 2016, 2,221–230.

[4] Koren, Y. Factorization Meets the Neighborhood: A Multifaceted Collaborative Filtering Model. In Proceedings of the 14thACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, NV, USA, 24–27 August 2008;pp. 426– 434.

[5] Ye, Q.; Hsieh, C.Y.; Yang, Z.; Kang, Y.; Chen, J.; Cao, D.; He, S.; Hou, T. A unified drug-target interaction prediction framework based on knowledge graph and recommendation system. Nat. Commun. 2021, 12, 6775.

[6] Fox, S.; Duggan, M. Health Online 2013; Pew Research Internet Project Report: Washington, DC, USA, 2013. pp. 191–194.

[7] Subramaniyaswamy, V.; Manogaran, G.; Logesh, R.; Vijayakumar, V.; Chilamkurti, N.; Malathi, D.; Senthil selvan, N. An ontology driven personalized food recommendation in IoT-based healthcare system. J. Super comput. 2019, 75, 3184–3216.

[8] Liang, T.P. Recommender systems for decision support. Expert Syst. Appl. 2008, 45, 385–38.

[9] Chintalapudi, N.; Angeloni, U.; Battineni, G.; di Canio, M.; Marotta, C.; Rezza, G.; Sagaro, G.G.; Silenzi, A.; Amenta, F. LASSO Regression Modeling on Prediction of Medical Terms among Seafarers' Health Documents Using Tidy Text Mining. Bioengineering2022, 9, 124.

[10] Lu, J.; Wu, D.; Mao, M.; Wang, W.; Zhang, G. Recommender system application developments: A survey. Decis. Support Syst.2015, 74, 12–32.

[11] Huang, F.; Wang, S.; Chan, C.-C. Predicting disease by using data mining based on healthcare information system. In Proceedings of the 2012 IEEE International Conference on granular.

Copyright © 2025 JAIET All Rights Reserved

Journal of Artificial Intelligence and Emerging Technologies
Volume 2, Issue 7, pp 6-10, July 2025
ISSN (online): 3049-1967
DOI: https://doi.org/10.47001/JAIET/2025.207002
Available Online at https://www.jaiet.com

**Citation of this Article:**

Gopinath M, & Manikandan D. (2025). Development of a Machine Learning and NLP-Based Framework for Automated Drug Recommendations. *Journal of Artificial Intelligence and Emerging Technologies.* 2(7), 6-10. Article DOI: https://doi.org/10.47001/JAIET/2025.207002

***** End of the Article *****

Copyright © 2025 JAIET All Rights Reserved 