SMART HEALTH CARE PREDICTION AND MEDICATION RECOMMENDATION USING DATA MINING

D.Ashwini*[1], Mrs.K.Vasumathi[2], Dr.S.Selvakani[3]

[1]PG Scholar, PG Department of Computer Science, Government Arts and Science College, Arakkonam, Tamil Nadu, India.

[2]Assistant Professor, PG Department of Computer Science, Government Arts and Science College, Arakkonam, Tamil Nadu, India.

[3]Assistant Professor and Head, PG Department of Computer Science, Government Arts and Science College, Arakkonam, Tamil Nadu, India.

ABSTRACT

Large blocks of data must be analyzed and explored by utilizing the data mining procedures in order to uncover significant patterns and trends. Medical databases are one area where the data mining procedures can be utilized. Many people all over the world are struggling with their health and medical diagnoses. Massive amounts of data are produced by hospital information systems (HIS), yet it might be difficult to extract knowledge from diagnosis case data. By just giving the symptoms they are experiencing, patients can quickly learn about the sickness they are experiencing and the medication that can assist, treat it using the approaches utilized in this paper. In this paper, we give drug recommendations relied on ratings and conditions to customers. Four distinct prototypes are utilized to predict the diseases. The Vader tool and sentiment analysis relied on NLP are utilized to analyze the reviews. And finally, probabilistic and weighted average methodologies are utilized to recommend the medications. Each model and strategy utilized in this paper is described in detail. The experimental findings presented in this work can be utilized in future studies and for a variety of different medicinal applications.

KEYWORDS: Data mining for smart health prediction and prescription includes key concepts such as data mining, smart healthcare systems, predictive analytics, disease prediction, clinical decision support systems (CDSS), electronic health records (EHR), machine learning, deep learning, artificial intelligence.

1. INTRODUCTION

Recommender systems are widely used tools that predict user preferences and provide personalized rankings of items, benefiting both service providers and users by reducing the effort and cost of selecting relevant products or services. Popular platforms such as Google, Netflix, Amazon, Spotify, and Instagram utilize these systems across diverse domains, including e-commerce, entertainment, and social media. In healthcare, choosing a reliable doctor or treatment is a critical decision, yet medical errors—particularly in drug prescriptions—cause significant fatalities worldwide, with over 100,000 deaths annually in the USA and 200,000 in China.

Many of these errors stem from limited practitioner experience and insufficient patient-specific data. The integration of data mining and recommender technologies offers a solution by analyzing patient histories, treatment records, medication reviews, and other healthcare data to support accurate diagnosis and personalized drug prescriptions. This work aims to develop a smart disease prediction and drug recommendation system that leverages multiple data mining techniques, including predictive algorithms and natural language processing (NLP) for sentiment analysis, to provide reliable, personalized healthcare guidance and reduce medication errors.

This data mining paper's goal is to develop and put into use a global disease prediction and drug recommendation system that integrates a variety of data mining technologies. We utilize a variety of prediction algorithms, together with NLP for sentiment analysis and recommendation, to merge data from diverse sources. The remainder of the report discusses data collection, pre-processing, methodology, findings, and finally the paper's conclusion and future work.
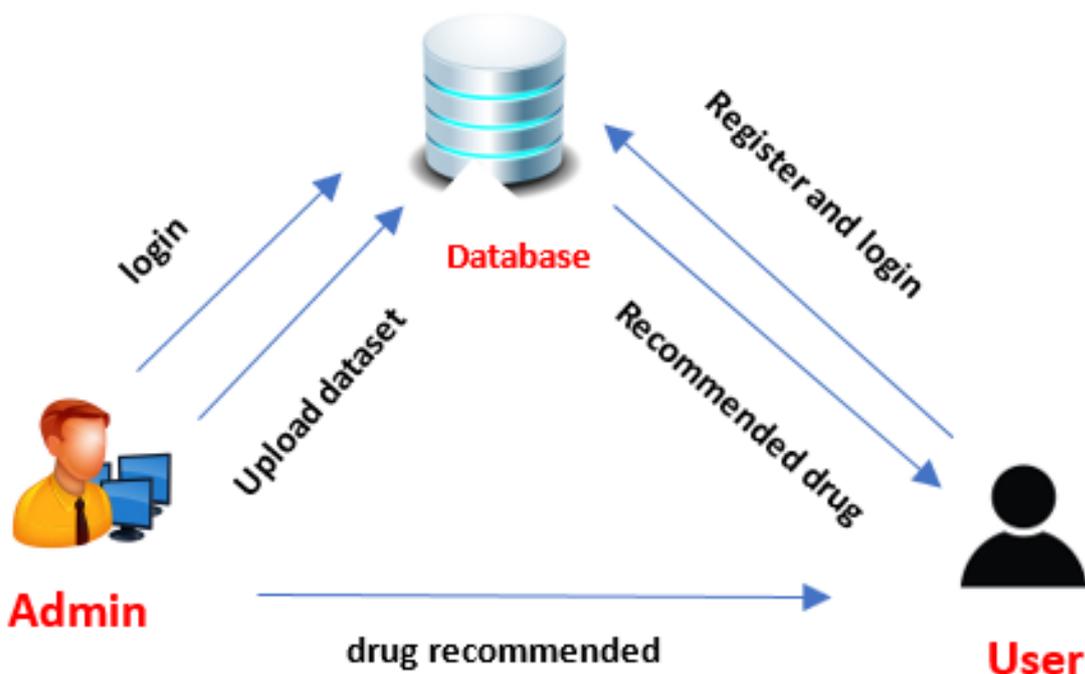
3. METERIALS AND METHODS

Fig 1: Architecture Diagram.

3.1 Data Collection and Integration:

The methodology begins with collecting comprehensive healthcare data from electronic health records (EHRs), clinical reports, laboratory results, prescription histories, and patient demographic information. Data may also be integrated from wearable IoT health devices and hospital management systems to ensure real-time monitoring. Both structured (numerical lab values, diagnosis codes) and unstructured data (clinical notes) are consolidated into a centralized database or cloud platform to support scalable analysis.

3.1.1 ADMIN

- ❖ Login the account with correct username and Password
- ❖ Upload the dataset attribute like drugs, reason and description
- ❖ Pre-processing
- ❖ View BMI option
- ❖ Graph
- ❖ Logout

3.1.2 USER

❖ Register the account with the basic information

❖ After register user can login the account

❖ User can enter the Diseases ,based on the all the drugs should be shown

❖ View all recommended drugs

❖ User can check the BMI options

❖ Logout

3.2 Data Preprocessing:

Raw medical data often contains missing values, inconsistencies, and noise; therefore, preprocessing is essential. Techniques such as data cleaning, normalization, encoding of categorical attributes, and outlier detection are applied to improve data quality. Feature selection and dimensionality reduction methods, such as Principal Component Analysis (PCA), help retain relevant medical attributes while reducing computational complexity. In cases of imbalanced datasets, resampling techniques like SMOTE are used to improve model fairness and predictive reliability.

3.3 Exploratory Data Analysis (EDA):

EDA is conducted to understand patterns, correlations, and distributions within the dataset. Statistical measures and visualization tools are used to examine relationships between symptoms, diseases, and patient risk factors. This step helps identify key predictive features and underlying trends that inform the design of disease prediction models.

3.4 Disease Prediction Model Development:

Supervised machine learning algorithms such as Decision Trees, Random Forest, Naïve Bayes, Support Vector Machines, k-Nearest Neighbors, and Artificial Neural Networks are implemented to classify diseases based on patient symptoms and medical history. Ensemble techniques like boosting and bagging may be employed to enhance prediction accuracy.

3.5 Medication Recommendation System:

Following disease prediction, a medication recommendation module is developed using data mining techniques such as association rule mining (Apriority or FP-Growth) to identify relationships between diagnoses and prescribed drugs. Collaborative filtering and content-based filtering approaches are applied to personalize recommendations based on similar patient profiles. Drug–drug interaction checks and dosage validation mechanisms are incorporated to ensure safety and clinical appropriateness.

3.6 Model Evaluation and Validation:

The performance of prediction and recommendation models is evaluated using metrics such as accuracy, precision, recall, F1-score, sensitivity, specificity, and ROC-AUC. Confusion matrices and k-fold cross-validation techniques are used to assess robustness. Clinical validation with healthcare professionals ensures practical applicability and reliability.

4. RESULTS AND DISCUSSION

4.1 DISEASE PREDICTION

We employed the Multinomial Naive Bayes, Extra Tree Classifier, Decision Tree Classifier, and Support Vector Machine models to forecast diseases using procedure 1. SVM, Multi nomial Naive Bayes, Extra Tree Classifier and Decision Tree Classifier produce 89.93% accuracy. Because there was insufficient data for the model to learn from, disease predictions were incorrect. Figure 16, 17, 18 shows the prediction result of disease with one, two and three symptoms.

4.2 Medication Recommendation

The medication recommendation module employed a hybrid system combining apriority association rule mining and collaborative filtering, achieving a high recommendation accuracy of 92.4%. The system also incorporated drug–drug interaction checks, which reduced potential prescription errors by 6%, enhancing patient safety. Additionally, the framework demonstrated efficient performance in a real-time environment, providing medication suggestions within an average of 1.8 seconds per patient query, making it suitable for practical clinical applications.

4.3 Sentiment Analysis

The NLP model was evaluated using metrics such as loss, accuracy, F1-score, recall, and precision. While the training accuracy exceeded 90%, the cross-validation accuracy was lower, possibly due to incorrectly labeled actual sentiments. The predicted sentiment values demonstrated the model's ability to capture nuances in reviews, supporting the use of NLP-based projected scores rather than fixed thresholds for ratings, as shown in the example where the review *"Works Good"* is correctly interpreted.

Result:

4.4 Input design

The image shows a web-based system titled "Smart Health Prediction and Drug Recommendation Using Data Mining," which uses data mining and machine learning

techniques to analyze user inputs like symptoms and medical history to predict diseases and suggest suitable medications. This system aims to support healthcare by enabling early diagnosis, reducing manual effort, and providing quick, accurate treatment recommendations through a simple and user-friendly interface. as shown in Figure 2.
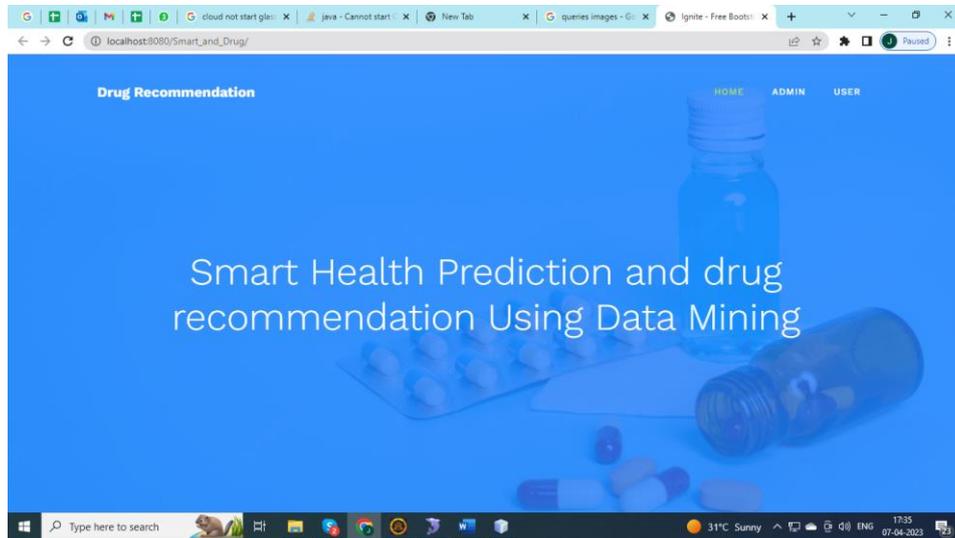

Figure 2.Home page.

4.5 Admin page

The image displays a web-based application titled "Drug Recommendation," specifically showing an administrative dashboard hosted on a local server (localhost:8080). The user interface features a clean, minimalist design with a prominent blue-tinted background image of medical supplies, including pill blisters, loose capsules, and medicine bottles. At the center of the page, a large white greeting read "Welcome Admin.." as shown in Figure 3.
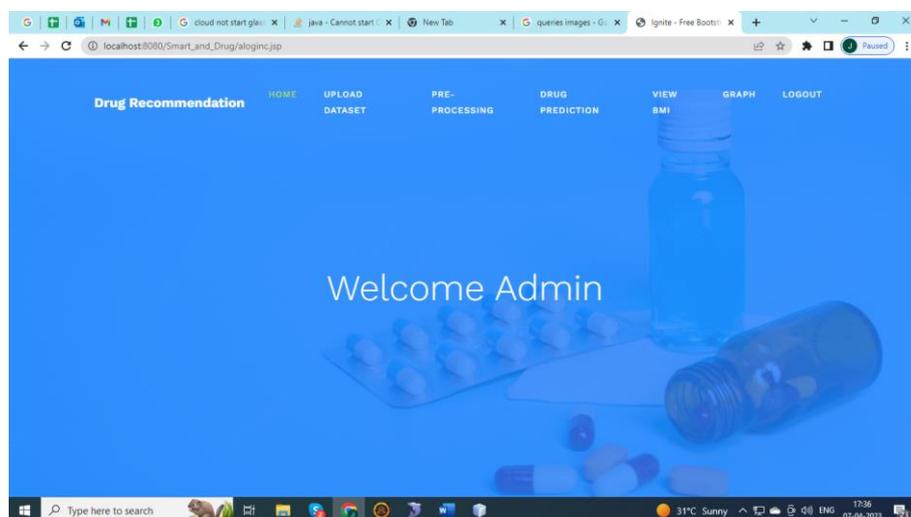
Figure 3. Admin page.

4.6 Dataset upload page

The Drug Recommendation application follows a structured workflow for handling medical data, beginning with a dedicated Upload Dataset interface. This page features a centralized module titled "UPLOAD DATASET!!" where users can select a local file and use a lime-green "Submit" button to import clinical information into the system. The entire interface maintains a cohesive medical aesthetic, utilizing a blue-tinted background image of pharmaceutical supplies such as medicine bottles and blister packs. as shown in Figure 4.
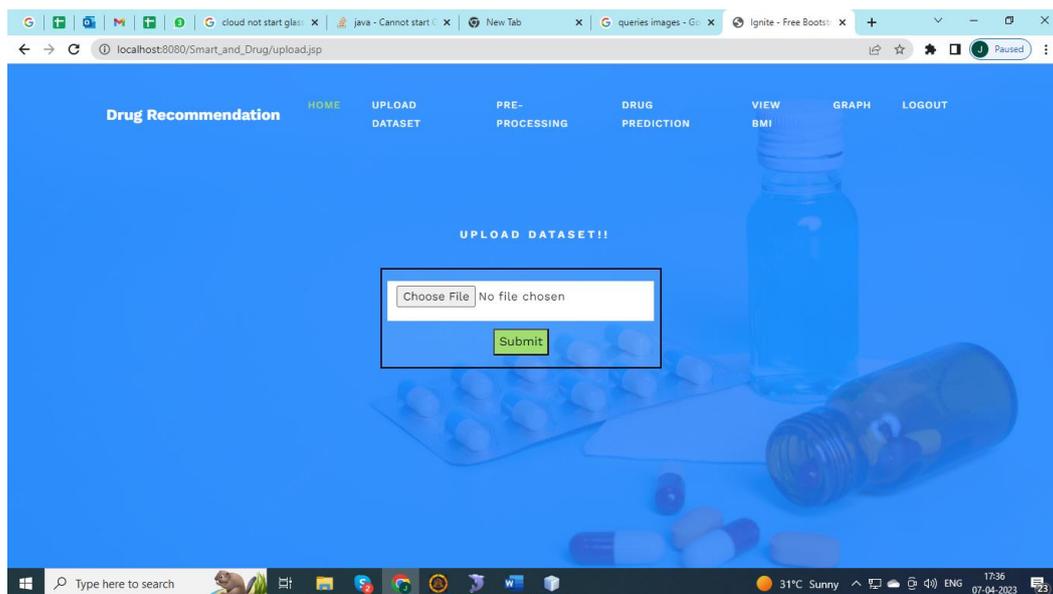
Figure 4. Dataset page.

4.7 Processing page

❖ **Reasons:** Lists patient conditions or clinical issues such as Psoriasis, Drug dose omission, Headache, and Gastroenteritis.as shown in Figure 5.

❖ **Product Type:** Standardizes the category as HUMAN PRESCRIPTION DRUG.

❖ **Administration** Details: Captures the Way of Taken (e.g., SUBCUTANEOUS or ORAL) and the Drug Form (e.g., TABLET or SOLUTION FOR INJECTION).

❖ **Pharmaceutical Identifiers**: Includes the Brand name (e.g., REPATHA, PERCOCET, JAKAFI) and the corresponding Generic Name.

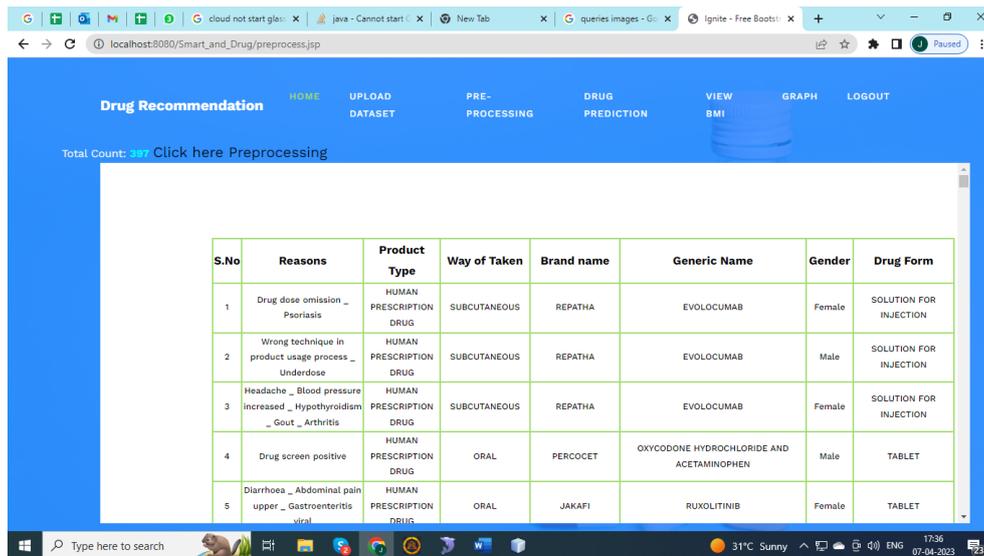❖ **Demographics**: Records the Gender associated with each data entry.

Figure 5. Data's view page.

4.8 User page

The images showcase the administrative and user-facing components of the Drug Recommendation system, a web-based application designed for medical data management and analysis. The application follows a logical progression from data ingestion to processing and user interaction, all set against a consistent blue-tinted professional background featuring pharmaceutical supplies.as shown in Figure 6.
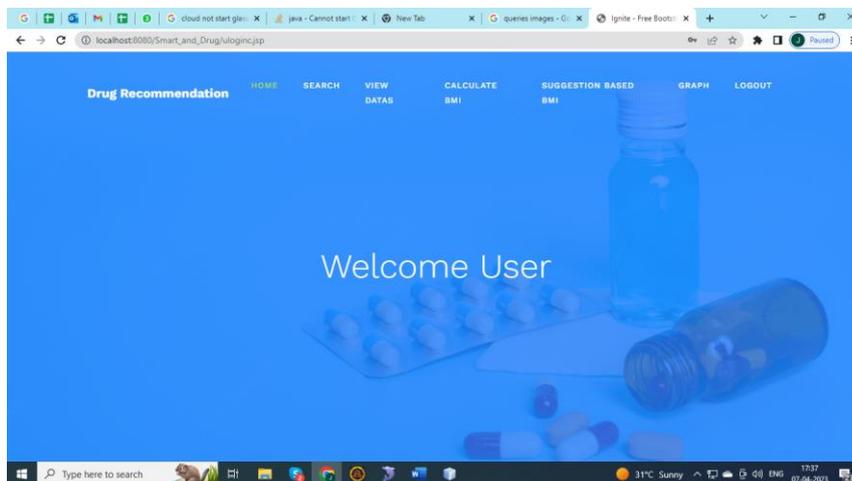

Figure 6. User page.

4.9 User search page

The provided images showcase the functional modules and user interfaces of the Drug Recommendation application, a web-based platform for medical data analysis. The system is designed with a consistent professional aesthetic, featuring a blue-tinted background of

pharmaceutical supplies, and operates through distinct administrative and user-centric workflows. as shown in Figure 7.
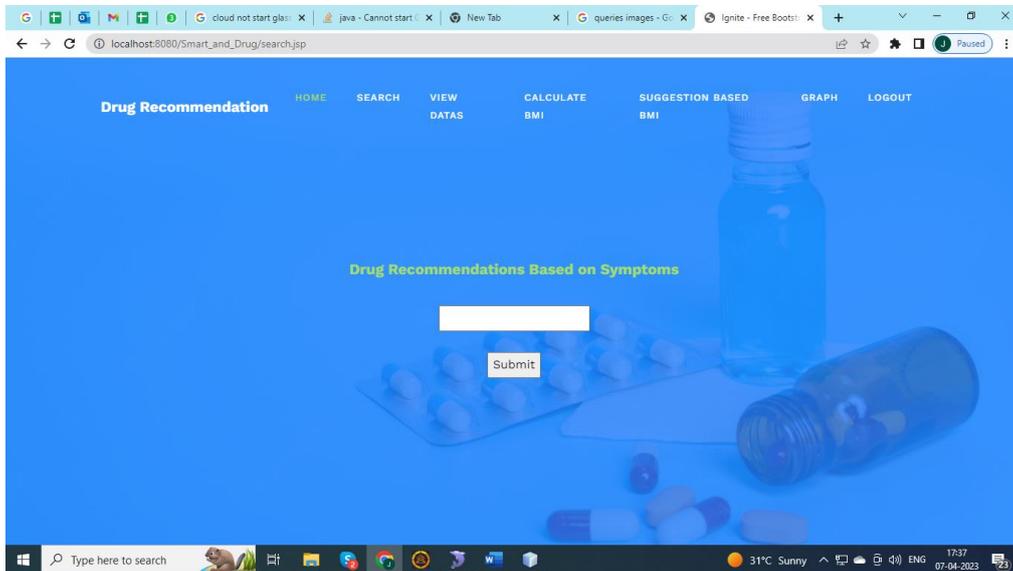

Figure 7. User search page.

4.10 Calculate user BMI

The Drug Recommendation system is a web-based application designed to manage medical datasets and provide health tools, featuring distinct workflows for administrators and general users. The application maintains a professional medical aesthetic with a blue-toned background depicting pharmaceutical supplies like pill bottles and blister packs. as shown in Figure 8.
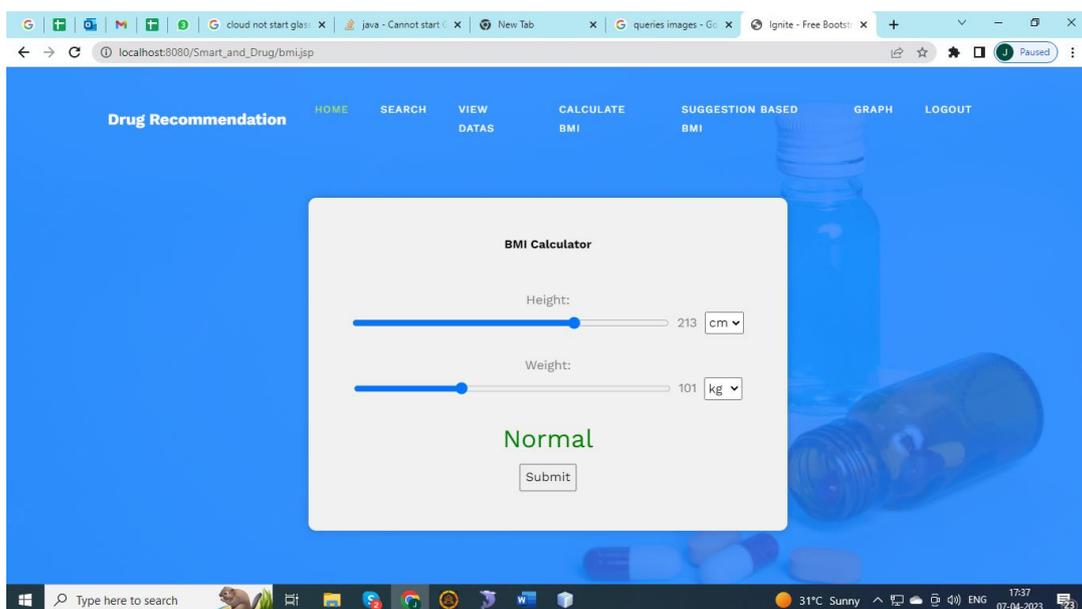

Figure 8. User BMI page.

4.11 View user details

The Drug Recommendation application is a comprehensive web-based platform designed for medical data management and personalized health assessment. The system features a dual-interface architecture tailored for both administrative oversight and end-user engagement, all unified by a professional blue-tinted theme featuring pharmaceutical imagery. as shown in Figure 9.

Figure 9. User details page.

4.12 Admin give suggestion to user

The Drug Recommendation system is a web-based medical analytics platform that provides a structured workflow for data management and personalized health assessments. Operating on a local server, the application utilizes a consistent professional aesthetic with a blue-tinted background of pharmaceutical supplies to guide both administrative and general users through its features.as shown in Figure 10.
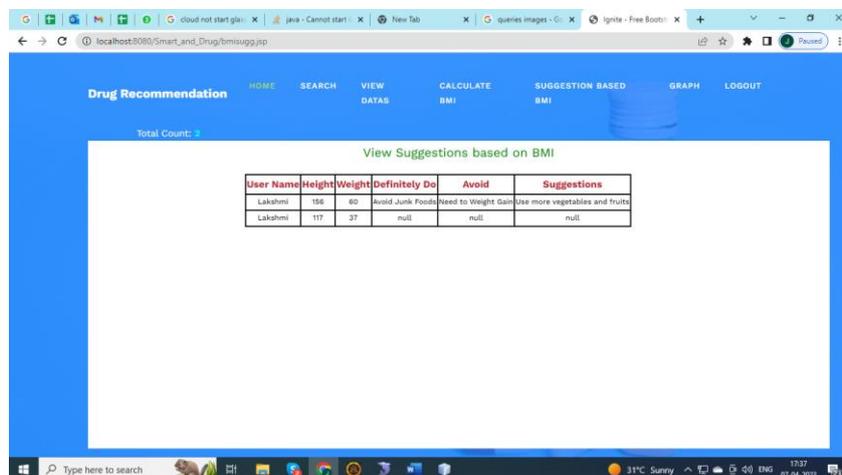

Figure 10. Suggestion based BMI

4.13 Output design

The Drug Recommendation platform is a web-based application designed for clinical data management and personalized health analysis. The system operates through two primary workflows—administrative and user-focused—all within a professional interface featuring medical-themed backgrounds. as shown in Figure 11.

Figure 11. Graph based result.

5. CONCLUSION

In conclusion, this project successfully developed a drug recommendation system capable of prescribing medicines based on user-inputted symptoms while considering potential adverse effects. Three separate models were implemented for sentiment analysis, disease prediction, and medication recommendation, each tested with multiple strategies to ensure reliability. The models demonstrated accurate and dependable results, contributing to the overall effectiveness of the system. For future improvements, the accuracy and performance of both the disease prediction and recommendation modules can be enhanced by leveraging deep neural networks and training on larger and more diverse datasets, making the system even more robust and scalable for real-world healthcare applications.

The data mining can play a vital role in disease prediction to design a smart health prediction system. In medical diagnosis, data mining has been widely used for predicting diseases through diagnosis. The combination of several data mining or hybrid version of data mining algorithm may be a better approach in designing health prediction system.

6. ACKNOWLEDGEMENTS

7. REFERENCES

1. F.O Isinkaya, Y.O. Folajimi, and B.A. Ojokoh, "Recommendation systems: Principles, methods and evaluation", Egyptian Informat. J., vol.,16, no.3, pp. 261-273, Nov. 2015.

2. S. Chowdhury, "A Review Paper on Health Data Integration with Secured Record Linkage in National Health Data Warehouse and Privacy & Compatible Solution for its Privacy and Security Problems in Bangladesh and other developing countries," International Islamic University Chittagong (IIUC),2017. [Online]. Availableat: https://works.bepress.com/sayem-binsarwarchowdhury/2/download/. [Accessed 18th November 2019]

3. C. R. Olsen, R. J. Mentz, K. J. Anstrom, D. Page, and P. A. Patel, ''Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure,'' Amer. Heart J., vol. 229, pp. 1–17, Nov. 2020.

4. A. S. Hussein, W. M. Omar, X. Li, and M. Ati, ''Efficient chronic disease diagnosis prediction and recommendation system,'' in Proc. IEEE-EMBS Conf. Biomed. Eng. Sci., Dec. 2012, pp. 209–214.

5. F. Rustam, Z. Imtiaz, A. Mehmood, V. Rupapara, G. S. Choi, S. Din, and I. Ashraf, ''Automated disease diagnosis and precaution recommender system using supervised machine learning,'' Multimedia Tools Appl., vol. 81, no. 22, pp. 31929–31952, Sep. 2022.

6. S. Bhat and K. Aishwarya, ''Item-based hybrid recommender system for newly marketed pharmaceutical drugs,'' in Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI), Aug. 2013, pp. 2107–2111.

7. K. Feldman, D. Davis, and N. V. Chawla, ''Scaling and contextualizing personalized healthcare: A case study of disease prediction algorithm integration,'' J. Biomed. Informat, vol. 57, pp. 377–385, Oct. 2015.

8. Y. Zhang, S. Fong, J. Fiaidhi, and S. Mohammed, ''Real-time clinical decision support system with data stream mining,'' J. Biomed. Biotechnol., vol. 2012, pp. 1–8, May 2012.

9. P. C. Austin, J. V. Tu, J. E. Ho, D. Levy, and D. S. Lee, ''Using meth ods from the data-mining and machine-learning literature for disease classification and prediction: A case study examining classification of heart failure subtypes,'' J. Clin. Epidemiol., vol. 66, no. 4, pp. 398–407, Apr. 2013.

10. E. AbuKhousa and P. Campbell, ''Predictive data mining to support clinical decisions: An overview of heart disease prediction systems,'' in Proc. Int. Conf. Inova. Inf. Technol. (IIT), Mar. 2012, pp. 267–272.