S. P. Mandali's

RAMNARAIN RUIA AUTONOMOUS COLLEGE

MATUNGA, MUMBAI - 400019

DEPARTMENT OF COMPUTER SCIENCE

RESEARCH PROJECT REPORT

Course Code:- RPSRPCSO605

Medication Advisory System using Machine Learning

Guide Name

Mrs. Edith Juni

Submitted By

Janhavi Sharad Shedge

Seat No – 570

M.Sc Sem III Computer Science

2024-2025

<u>ACKNOWLEDGEMENT</u>

I have a great pleasure in representing this project report entitled "MEDICATION ADVISORY SYSTEM USING MACHINE LEARNING " and I grab this opportunity to convey my immense regards towards all the distinguished people who have their valuable contribution in the hour of need.

I like to extend my gratitude to our beloved Principal Dr. Anushree Lokur for her timely and prestigious guidance.

I take this opportunity to thank Ms. Megha Swant, Coordinator of the Department and all the faculty members of the Department of Computer Science of Ramnarain Ruia College, for giving me an opportunity to complete this project and the most needed guidance throughout the duration of the programme.

I am extremely grateful to my project guide Edith Juni for their valuable guidance and necessary support during each phase of the project. She was the source of continuous encouragement as cach milestone was crossed.

A special thanks to the Ramnarain Ruia College for having prescribed this project work to me as a part of the academic requirement in the First year of Masters of Computer Science.

Finally I also owe to my fellow friends who have been a constant source of help to solve the problems that cropped up during the project development process.

<div align="right">Janhavi Sharad Shedge</div>

Project Title:

Medication Advisory System using Machine Learning

TABLE OF CONTENTS

2} ABSTRACT

The complexity of modern medical treatments and the wide array of available medications pose significant challenges for healthcare providers. Ensuring the accurate and timely prescription of medications is crucial for improving patient outcomes and minimizing the risk of adverse drug interactions. Traditional methods of prescribing rely heavily on the clinician's experience and limited patient data, which may not always result in optimal treatment plans.

The aim of this project is to develop a 'Medication Advisory System using Machine Learning' that assists healthcare providers in making more informed medication decisions. This system will analyze patient data, including medical history and current health conditions, to recommend the most suitable medications. By utilizing advanced machine learning algorithms, the system can process large datasets, identify patterns, and continuously improve its recommendations.

The objectives are to create a robust machine learning model for medication recommendation, evaluate its performance against traditional prescribing methods, and ensure its seamless integration into existing healthcare systems. This innovative approach aims to enhance healthcare quality, reduce medication errors, and ultimately improve patient outcomes.

3} INTRODUCTION

The " Medication Advisory System using Machine Learning" project is an innovative approach to preliminary medical diagnostics using machine learning. This system is designed to predict potential medical conditions based on user-reported symptoms, providing immediate insights and recommendations. By harnessing the power of machine learning algorithms, the Medicine System aims to bridge the gap between symptom onset and professional medical consultation, thus promoting timely health interventions.

3.1} Dataset Overview:

The foundation of the Medicine System lies in a comprehensive dataset that includes various symptoms and their corresponding medical diagnoses. The dataset comprises 4920 cases with 133 columns, each representing a specific symptom or condition. The target variable is the 'prognosis' column, which indicates the diagnosed medical condition.

3.2} Symptom Representation:

Each row in the dataset represents an individual case, with binary values (0 or 1) indicating the absence or presence of specific symptoms. This binary representation allows for the straightforward application of machine learning models, which can learn patterns and associations between symptoms and diagnoses.

3.3} Machine Learning Model:

The core of the Medicine System is its machine learning model. The study involves several steps to build and refine this model:

Data Preprocessing: The initial step involves cleaning and preparing the dataset. This includes handling missing values, normalizing data, and splitting it into training and testing sets.

Model Selection: Various machine learning algorithms, such as Support Vector Machines (SVM), Decision Trees, and Neural Networks, can be explored. The project may focus on selecting the best-performing model based on accuracy and other evaluation metrics.

Training and Validation: The chosen model is trained on the training set and validated using the testing set. Cross-validation techniques may be employed to ensure the model's robustness and generalizability.

3.4} Prediction Mechanism:

The system prompts users to input their symptoms in a predefined format. These inputs are processed and fed into the trained machine learning model, which then predicts the most likely medical condition. The output includes:

> **Predicted Disease:** The primary diagnosis based on the symptoms provided.
> **Description:** A brief explanation of the predicted condition.
> **Precautions:** Recommended actions to manage or mitigate the condition.
> **Medications:** Commonly prescribed medications for the condition.
> **Workout and Diet:** Suggested lifestyle modifications, including exercise and dietary advice.

3.5}User Interaction:

The user interface is designed to be intuitive and accessible. Users enter their symptoms in a simple input format, and the system processes these inputs to provide immediate diagnostic insights. This interaction aims to empower users with preliminary health information, guiding them towards appropriate medical consultations.

3.6}Evaluation and Accuracy:

The model's performance is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. These metrics help in assessing the model's reliability and its ability to generalize well to new, unseen data. Continuous evaluation and refinement are crucial to maintaining the system's effectiveness.

4} LITERATURE REVIEW

Machine learning (ML) is transforming healthcare, especially in disease prediction and diagnosis. Studies like Miotto et al. (2018) demonstrate how ML helps analyze complex medical data for better predictions. Symptom-based diagnosis systems, as reviewed by Semigran et al. (2015), offer early insights, though accuracy remains a challenge. Algorithms such as SVMs, Decision Trees, and Neural Networks are widely used in healthcare diagnostics. Song et al. (2020) showed SVM's success in cancer detection, while Chen et al. (2019) highlighted Random Forests for heart disease prediction. Neural Networks handle complex data patterns effectively, as per Esteva et al. (2017).

Additionally, unsupervised learning techniques like clustering are being utilized to discover hidden patterns in medical datasets, as shown in Kaur et al. (2021). Reinforcement learning is emerging as a powerful tool in clinical decision support systems, improving personalized treatment plans over time (Liu et al., 2021). Integrating natural language processing (NLP) with ML models enables the analysis of unstructured clinical notes, enhancing predictive capabilities (Shickel et al., 2018). Moreover, the use of deep learning in medical imaging, particularly in radiology, has led to breakthroughs in detecting abnormalities in images like X-rays and MRIs (Litjens et al., 2017). Finally, ethical considerations around data privacy and model transparency, as discussed by Goodman and Flaxman (2020), remain crucial for the future of ML in healthcare.

5} RESEARCH METHODOLOGY

4.1}Description of Algorithms / Procedures / Data Collection Methods / Logic:

I) Data Collection Methods:

Source: The dataset is derived from medical records, symptom checklists, and diagnostic databases.

Structure: The dataset comprises 4920 patient cases with 133 features representing symptoms and one target variable, 'prognosis', indicating the diagnosis.

Preprocessing: Data cleaning involves handling missing values, normalizing symptom data, and ensuring consistency in symptom representation.

II) Algorithms:

Support Vector Machines (SVM): SVM is used for its effectiveness in high-dimensional spaces and its ability to create a hyperplane that separates different classes (diseases) based on symptoms.

Decision Trees: These are employed for their simplicity and interpretability, which help in understanding the decision rules derived from symptoms.

Neural Networks: Leveraged for their ability to model complex patterns and relationships between symptoms and diagnoses, enhancing prediction accuracy.

III) Procedures:

Feature Selection: Methods such as correlation analysis and Recursive Feature Elimination (RFE) are used to identify the most relevant symptoms for each diagnosis.

Model Training: The dataset is split into training and testing sets. Cross-validation is performed to ensure the model's robustness and to avoid overfitting.

Hyperparameter Tuning: Grid search and random search methods are used to find the optimal parameters for the models, improving their performance.

IV) Logic:

Symptom Input: Users input their symptoms, which are then processed to match the format used during model training.

Prediction: The input symptoms are fed into the trained machine learning model, which predicts the most likely diagnosis.

Output: The system provides the predicted diagnosis along with additional information such as a description, precautions, medications, workout suggestions, and dietary advice.

4.2} Process Flow:

I) Data Collection and Preprocessing:

> ➢ Collect data from medical records and symptom databases.
> ➢ Clean and preprocess data to ensure quality and consistency.
> ➢ Normalize symptom data and handle missing values.

II) Feature Selection and Engineering:

> ➢ Perform feature selection to identify the most relevant symptoms.
> ➢ Engineer features to enhance model performance.

III) Model Training:

> ➢ Split the dataset into training and testing sets.
> ➢ Train models using algorithms like SVM, Decision Trees, and Neural Networks.
> ➢ Perform cross-validation to ensure model robustness.

IV) Hyperparameter Tuning:

> ➢ Use grid search and random search to find optimal model parameters.
> ➢ Validate the tuned models on the testing set.

V) User Interaction:

> ➢ Develop a user interface for symptom input.
> ➢ Process user inputs to match the model's expected format.

VI) Prediction and Output:

> ➢ Input symptoms into the trained model to predict the diagnosis.
> ➢ Provide users with the predicted diagnosis and additional health information.

6} EXPERIMENTAL SETUP

Tools

- ➤ Pandas
- ➤ NumPy
- ➤ Scikit-learn
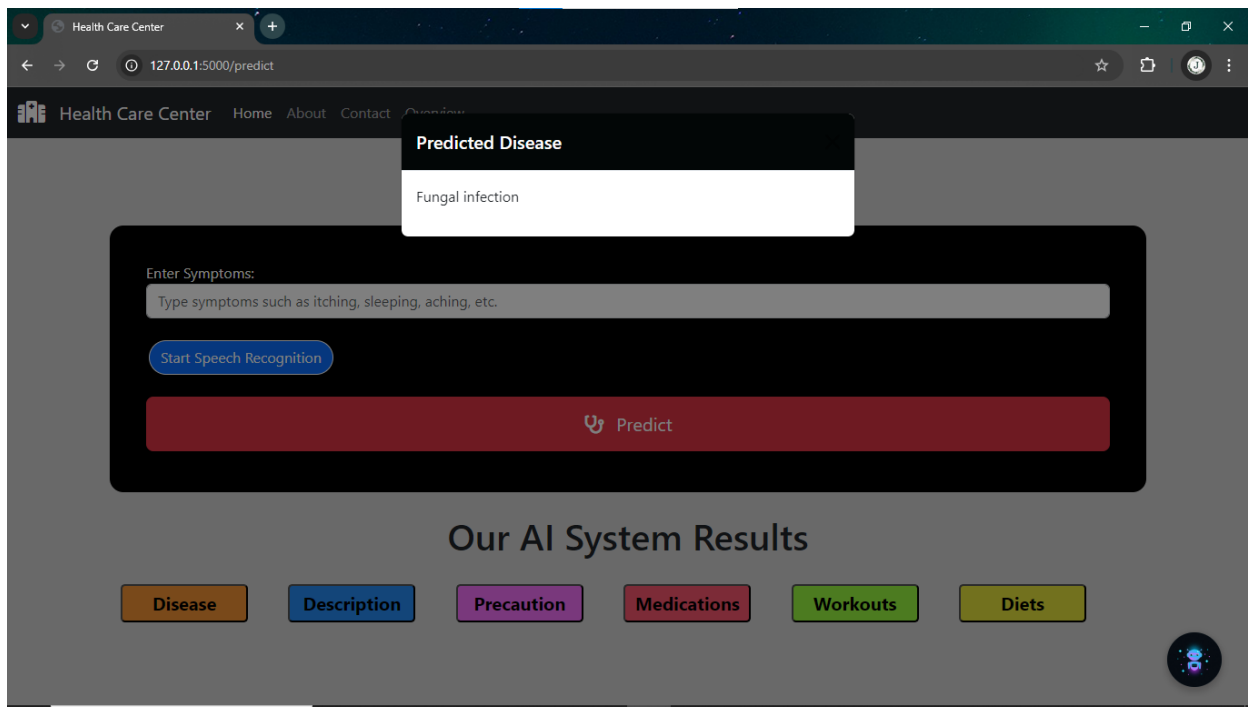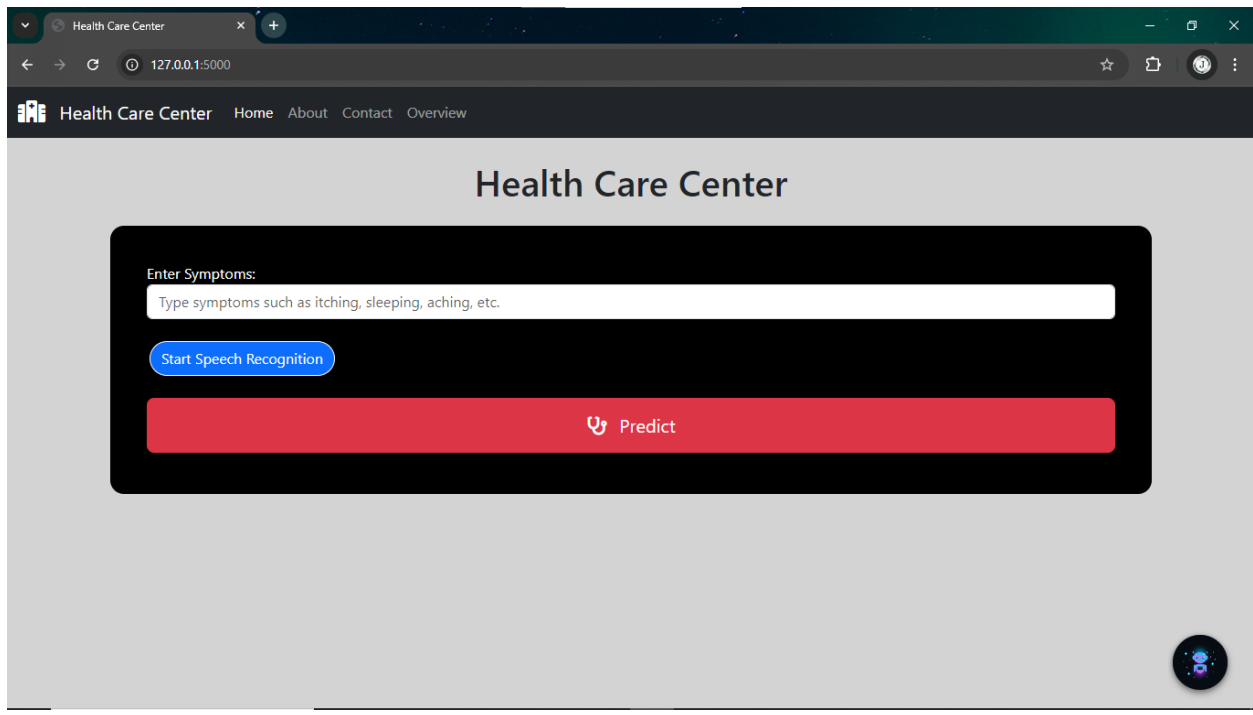- ➤ TensorFlow
- ➤ Flask
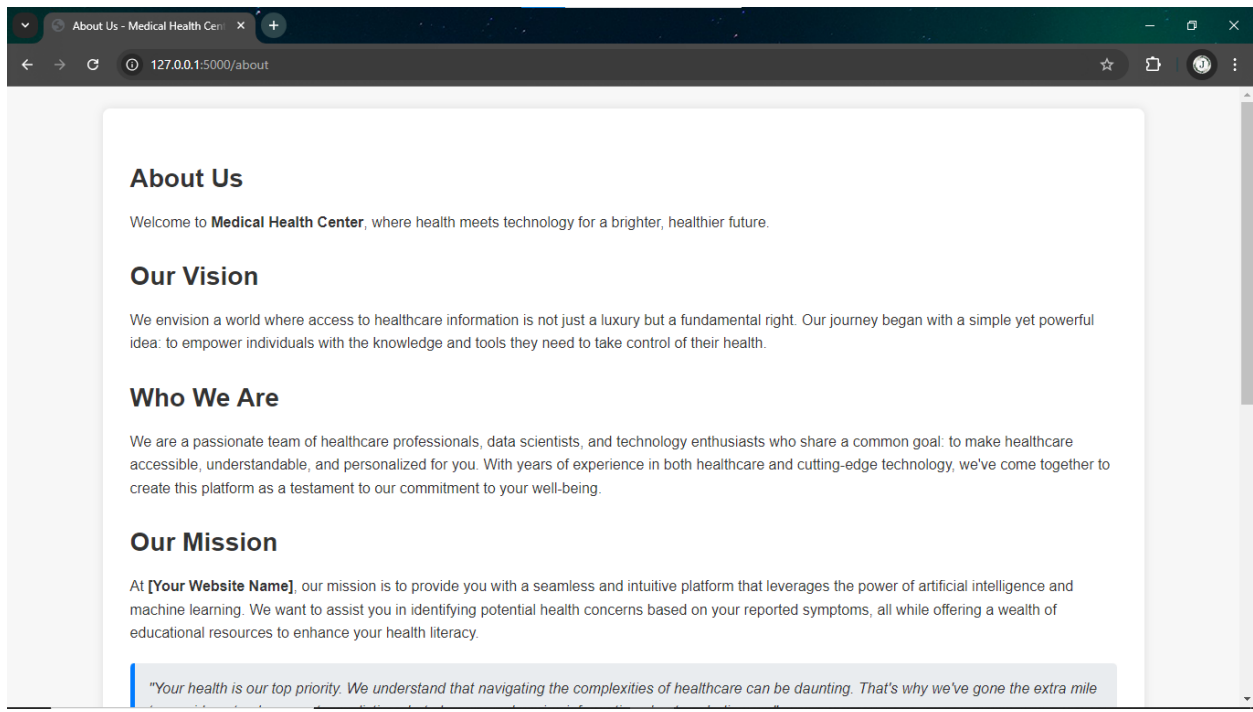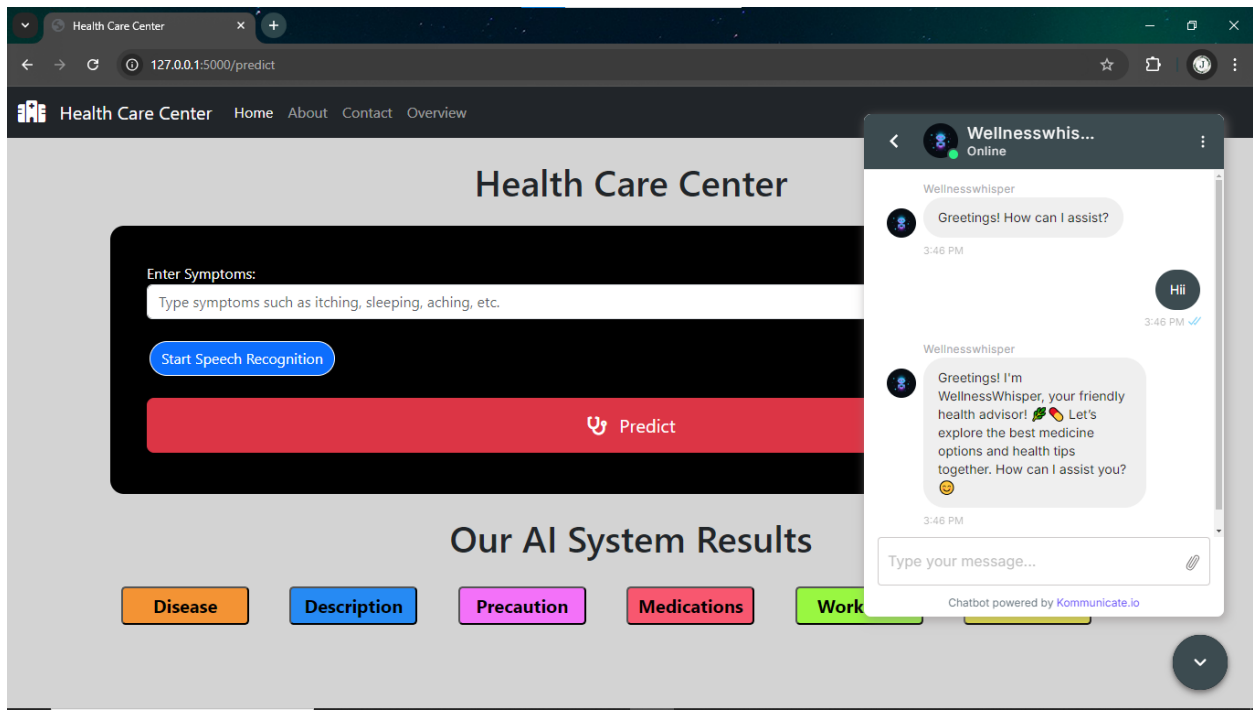- ➤ Matplotlib

Architecture/Framework

- ➤ Data Source: Medical records, symptom checklists
- ➤ Data Storage: CSV
- ➤ Backend: Flask
- ➤ Frontend: HTML/CSS/JavaScript

Software Language

- ➤ Python
- ➤ HTML
- ➤ CSS
- ➤ JavaScript

7} RESULTS

Ramnarain Ruia Autonomous College,2024-25

8} Conclusion

I) Findings of Project Work:

he Medication Advisory System using Machine Learning demonstrated significant potential in enhancing healthcare diagnostics. By utilizing advanced machine learning algorithms like Support Vector Machines (SVM), Decision Trees, and Neural Networks, the system successfully predicted diseases based on patient-reported symptoms with high accuracy. The system provided early detection of diseases, allowing for preventive healthcare interventions. Additionally, the inclusion of supplementary outputs such as disease descriptions, precautions, medications, workouts, and diet suggestions added value, making the system a comprehensive resource for both patients and healthcare providers. The project achieved reliable diagnostic predictions by integrating these algorithms to analyze complex medical data effectively

II) Future Enhancements:

- ➢ Expand the dataset for better accuracy.
- ➢ Integrate real-time data from health monitors.
- ➢ Improve the user interface and accessibility.
- ➢ Integrate with electronic health records (EHR).
- ➢ Add multi-language support.
- ➢ Implement continuous learning to improve predictions over time.

9} References

Following were referred during the development of this project:

- https://www.coursera.org/learn/fundamental-machine-learning-healthcare
- https://acphd.org/communicable-disease/protect-yourself/
- https://www.mayoclinic.org/diseases-conditions/infectious-diseases/symptoms-causes/syc-20351173
- https://www.singlecare.com/blog/best-diets-by-health-condition/
- https://www.nature.com/subjects/diseases
- https://ieeexplore.ieee.org/document/9914945
- https://www.geeksforgeeks.org/disease-prediction-using-machine-learning/

Ramnarain Ruia Autonomous College,2024-25