A Healthcare System using Machine Learning Techniques for Disease Prediction with Chatbot Assistance

ABSTRACT:

This project aims to develop a healthcare system that can predict multiple diseases, including liver disease, diabetes and heart disease using machine learning techniques. These are the most common types of Diseases that plaque the Indians today. This system will use XGBoost (eXtreme Gradient Boosting) for these diseases while maintaining accuracy, speed, resource utilization, stability and reducing complexity. The system will also include a healthcare chatbot that can provide personalized health recommendations to users.

The system will be developed using Python and its many Machine Learning Libraries & Streamlit Framework. Machine learning models will be trained on publicly available datasets to predict diseases accurately. The developed models will be evaluated using performance metric like accuracy of training and testing data. The healthcare chatbot will be developed with NLP and/or OpenAI API Integration Techniques to provide personalized health recommendations to users. The system will be deployed on a cloud-based platform to ensure scalability and availability.

The Healthcare System has the potential to improve the accuracy of disease diagnosis, care and early intervention, thereby reducing healthcare costs and improving patient outcomes. The healthcare chatbot can provide accessible and personalized health recommendations, leading to a healthier and happier life.

Batch: A8

Details:

S.No.	Hall Ticket No.	Name	Contact	Email Id
1.	19WJ1A0519	Anagondi Shiva Prasad	9347025164	shiva.prasad3345@gmail.com
2.	19WJ1A0531	Badempet Pavan	7989256101	pavan9b@gmail.com
3.	19WJ1A0561	Cheerala Prashanth	9347604784	cheeralaprashanth10@gmail.com

EXISTING SYSTEM	PROPOSED SYSTEM
• In case of Existing System there are almost none which have the capability of turning into a complete All in One Healthcare System for Disease Prediction. Even if there are they suffer from having the requirement of Significant amount of Computing Resources and yet suffer from low accuracy. The existing system uses Decision Trees for Liver and Heart Disease and KNN for Diabetes Prediction. Each of these falters in different ways. • On top of that they lack a proper UI and falter with scalability, ease of use and cannot provide personalized recommendations.	• In case of Proposed System XGBoost (eXtreme Gradient Boosting) an Ensemble Boosting Technique is used for Liver Disease Prediction, Prediction, Heart Disease prediction and Diabetes Prediction. • Proposed System is a powerful and efficient for disease prediction, and it can provide accurate and personalized health recommendations to users through the healthcare chatbot. It aims to overcomes the limitations of existing systems by using robust algorithm that can handle complex and noisy medical datasets and provide accurate and interpretable results. A UI for accessing all the tools in one place.
EXISTING ALGORITHM • Decision Trees, KNN	**PROPOSED ALGORITHM** • XGBoost (eXtreme Gradient Boosting)
ALGORITHM DEFINITIONS • A decision tree is a machine learning algorithm that uses a hierarchical structure of branching decisions to model the relationship between the input features and the output label. Each internal node of the tree represents a decision on a particular feature, and each leaf node represents a class label or a numerical value. The goal of the decision tree is to recursively split the data into subsets that are as homogeneous as possible with respect to the target variable. • K-Nearest Neighbors (KNN) works by finding the K nearest data points to a new input data point in the training dataset, and	**ALGORITHM DEFINITIONS** • XGBoost is a powerful ensemble learning algorithm that is used for both classification and regression tasks. It uses a gradient boosting framework to iteratively train multiple weak models (decision trees) on the data, where each subsequent model tries to correct the errors of the previous models. • XGBoost is known for its speed and scalability, and it has become a popular algorithm for winning machine learning competitions on Kaggle. It achieves this speed by implementing parallel processing and using hardware optimization techniques such as cache-aware computing and out-of-core computing. These

then predicting the output value based on the majority class or average value of the K nearest neighbors.	optimizations allow XGBoost to handle large datasets with millions of features and millions of instances in a relatively short amount of time.
DRAWBACKS • KNN can be computationally expensive and slow when dealing with large datasets, as it requires calculating the distance between the new input point and every point in the training dataset. • Decision Trees are prone to overfitting, especially when dealing with noisy data or complex datasets with many features. This can lead to poor generalization and low accuracy on new data. • Both KNN and Decision Trees can suffer from the curse of dimensionality, where the performance of the model decreases as the number of features increases. This can lead to increased computation time and poor accuracy.	**ADVANTAGES** • XGBoost uses optimized algorithms and parallel processing techniques to handle large datasets more efficiently. • XGBoost addresses overfitting issue by using regularization techniques, such as L1 and L2 regularization, to prevent overfitting and improve the generalization of the model. • XGBoost uses feature selection and feature engineering techniques to reduce the number of features and improve the performance of the model.

SYSTEM ARCHITECTURE

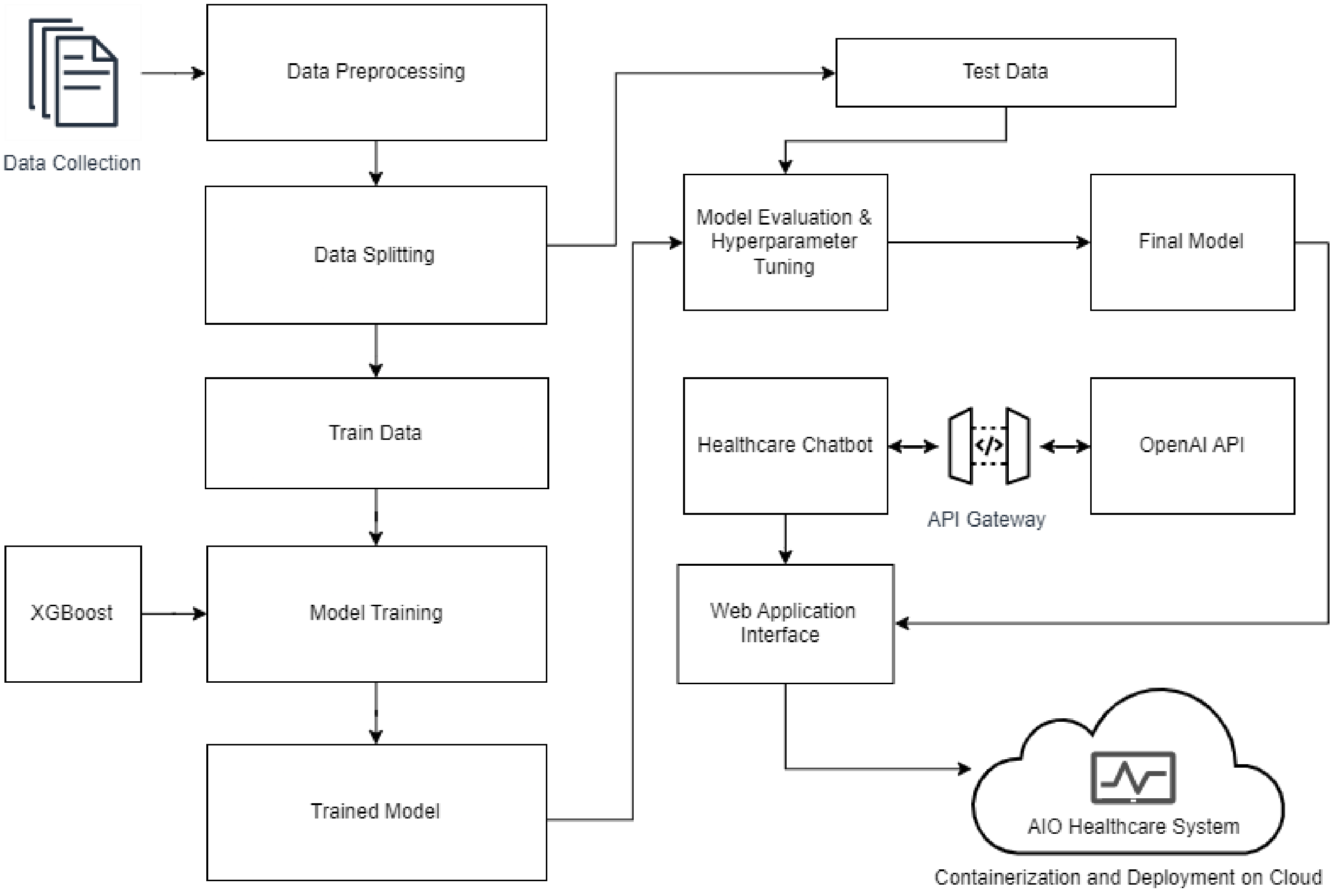

Fig: Proposed System Architecture

MINIMUM SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS

• Processor	:	Core i3 Processor
• RAM	:	4GB DDR3 RAM
• Hard Disk	:	500 GB

SOFTWARE REQUIREMENTS

• Operating System	:	Linux 3.0, WINDOWS 7
• Back End	:	Heroku, Python
• Version Control	:	GitHub
• Front End	:	Streamlit, Python