Group A5: DISEASE PROGNOSIS USING RANDOM FOREST

Presented by

D. Rahul 245321748022
G. Chandu 245321748025
N. Vishal 245321748064

Under the supervision of

Mr. Nageswara Rao

Project Introduction

- **What exactly is Disease Prognosis?**

  Disease prognosis refers to the prediction of the likely course and outcome of a medical condition.

  Importance: Importance: Early prognosis can lead to timely interventions and personalized treatment plans.

- **Aim of our project**

  In our project we predict health outcomes with advanced data analysis based on user symptoms.

- **Benefits of Using ML for Disease Prognosis**

  - ML considers individual health data for customized disease prognosis, optimizing treatment strategies.
  - ML algorithms analyze subtle patterns, enabling early disease identification before symptoms manifest.
  - Enhanced prognostic accuracy leads to more effective and targeted interventions, elevating the overall quality of patient care.
  - ML-based prognosis contributes valuable insights for ongoing medical research, fostering advancements in understanding and treating diseases.

Technologies Used

Html, Css, Js - HTML structures content, CSS styles presentation, and JS adds interactivity to web pages.

Node.JS – A runtime environment for executing javascript code outside of a web browser

Express.JS – Express.js is a web application framework for Node.js, simplifying server-side JavaScript development.

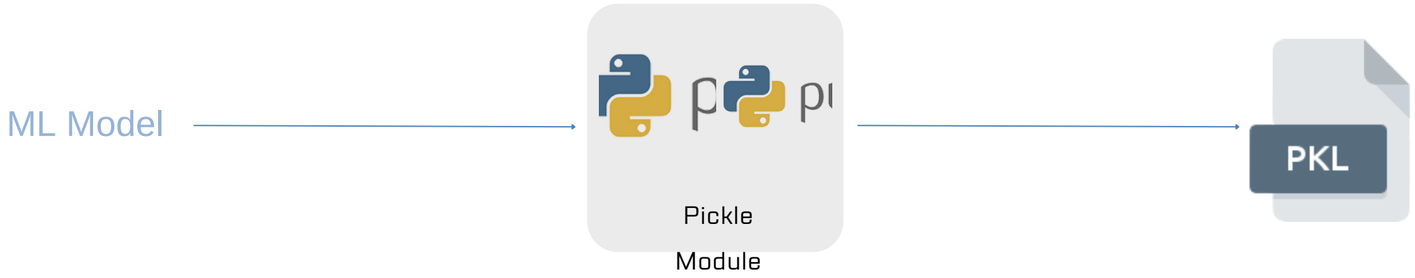
Python – A high level, multi-purpose programming language.
We used python specifically for implementing machine learning and extracting trained features from external windows pe files

The Machine Learning Model

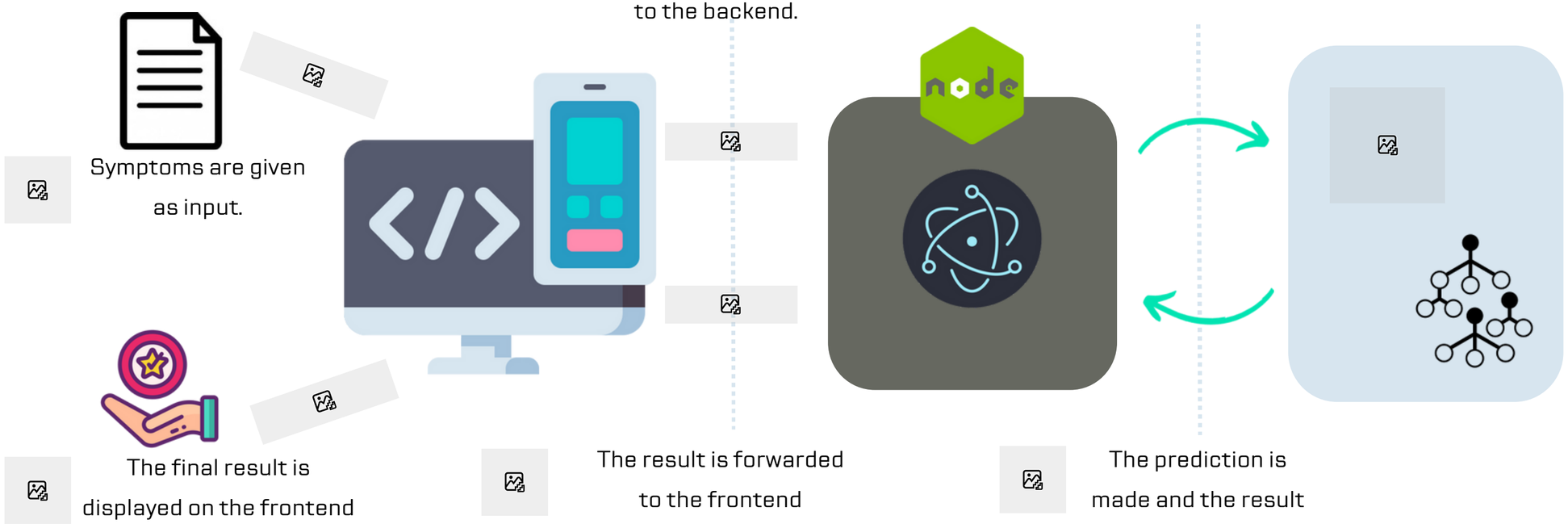
Dataset source: Dataset

- We trained a machine learning model with the dataset, which contains 132 parameters on which 42 different types of diseases can be predicted.
- The model was trained using the Random Forest Classifier algorithm, with an observed training accuracy of 100%.

- In order to leverage the model later using a python script, we saved it with the help of a python module called ‘pickle’.



Project Workflow



RANDOM FOREST

KEY POINTS :

- Ensemble Learning: Random Forest is an ensemble learning method, meaning it combines the predictions of multiple individual models to improve overall performance.
- Decision Trees: The base models used in Random Forest are decision trees. Each tree is constructed independently, and the final prediction is based on the aggregate of predictions from all trees.
- Bagging: Random Forest employs a technique called bagging (Bootstrap Aggregating) to train each decision tree on a randomly selected subset of the training data with replacement. This helps reduce overfitting and variance in the model.
- Random Feature Selection: At each split in the decision tree, only a random subset of features is considered. This introduces further randomness into the model and helps decorrelate the trees, leading to more diverse and robust predictions.
- Out-of-Bag (OOB) Error Estimation: Random Forest uses the out-of-bag samples (samples not used in the training of a particular tree) to estimate the model's performance without the need for cross-validation.
- Feature Importance: Random Forest provides a measure of feature importance, indicating the contribution of each feature to the model's predictive performance. This can be useful for feature selection and understanding the underlying data patterns.


Thank
you