



# Multi Disease Prediction

BY

RAMGOPAL DEVARUPPALA

# Overview

- ▶ Problem Statement
- ▶ Introduction
- ▶ Data Description
- ▶ Data Analysis
- ▶ Results
- ▶ Conclusion

# Problem Statement

Multi-disease prediction remain a challenge in the domain of machine learning mainly because of the existence of different symptoms for the same disease and also the accuracies have direct impact on the mental health of the patient. Suppose, by giving the symptoms as an input for the designed model and it has a low accuracy and predicts that the patient has a deadly disease such as cancer, the patient may be depressed mentally without actually having the disease. Here we will check for different model. So that We can prefer unique model for the multiple diseases.

# Introduction

The project mainly discuss about how the different types of model performs on the same datasets. The dataset is taken out from the Kaggle. In order to get clarity about their health status, this multi disease prediction models are used, where the past records were trained and tested and get the results. We used Random forest, K-NN, SVM and Logistic regression models. Matplotlib is a library used for the visualization.

# Data Description

- ▶ The Datasets are taken out from the Kaggle.
- ▶ Three different types of datasets are used in the project. That is Cancer dataset (569,27) , Liver dataset (583,11) and Diabetics dataset(768,9). All values are significant digits.
- ▶ The dataset is divided into training and testing dataset
- ▶ Dataset split into 70% of training dataset and remaining 30% for testing data.

# Data Analysis

- ▶ Data cleaning:  
Data contains numerous useless and missing values. For that, we have to drop the rows or columns or it should be replaced by the other values.
- ▶ Correlation : Explains how the variables are strongly related to each other.
- ▶ Performed Random forest and other models like SVM, KNN, Logistic Regression.
- ▶ Applied Hyperparameter tuning for the random forest to achieve good results.

# Data Analysis ( Contd.)

- ▶ Hyper parameter tuning is nothing but trying some different combination to get the more accurate results.
- ▶ The analysis was done to compare the model performances over the same datasets.

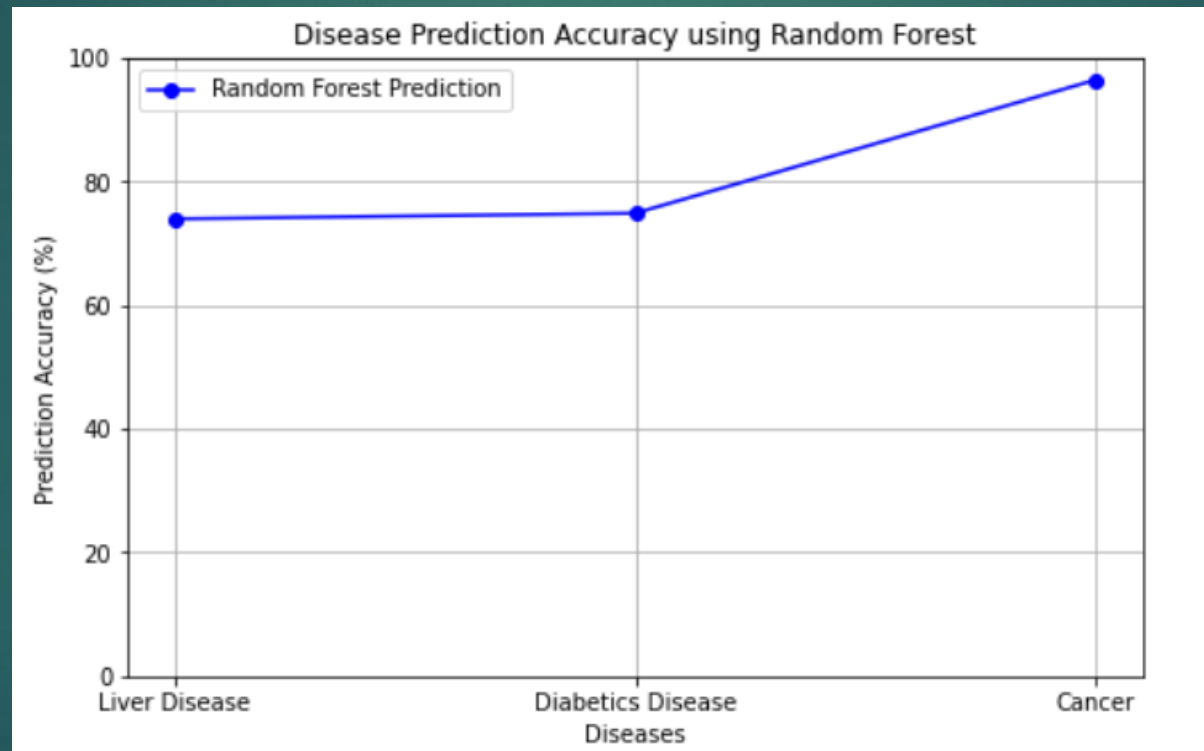
# Results

- ▶ The accuracy for the Liver disease prediction while using Random Forest is 73.97, SVM- 74.65, KNN – 70.54, Logistic regression- 70.54
- ▶ The accuracy for the Diabetics disease prediction while using Random Forest is 74.89, SVM – 74.025, KNN- 69.26, Naïve bayes- 72.29, Logistic regression – 74.025
- ▶ The accuracy for the cancer prediction while using Random Forest is 96.5, SVM – 96.50, KNN – 96.50, Logistic regression – 96.503496



# Results (contd.)

Comparison graph for different diseases on Random forest model :



# Conclusion

- Therefore, This Project concluded with different models with different datasets.

Out of these, below models performed well on different disease with different datasets

- > Liver Disease prediction using SVM
- > Diabetics Disease prediction using Random Forest
- > Cancer prediction using Random Forest

So, using Multi disease prediction system patients can save their time and improves mental health.

Future Scope : This project can be expanded further into web-app implemented using HTML, CSS, Java script and python framework