MAIN PROJECT

SYNOPSIS ON

Statistical Machine Learning Approaches to Liver Disease Prediction

Submitted By,

Robin Biju

MCA S4

Roll no: MCK20MCA-2037

Guided By,

Deepa Thomas

Assistant Professor

Department of CSE

Statistical Machine Learning Approaches to Liver Disease Prediction

INTRODUCTION

Patients with Liver disease have been continuously increasing because of excessive consumption of alcohol, inhale of harmful gases, intake of contaminated food, pickles and drugs. This dataset was used to evaluate prediction algorithms in an effort to reduce burden on doctors. The data set include Age of the patient, Gender of the patient, Total Bilirubin. Direct Bilirubin, Alkaline Phosphatase, Alamine Aminotransferase, Aspartate Aminotransferase, Total Proteins, Albumin, Albumin and Globulin Ratio. Dataset: field used to split the data into two sets (patient with liver disease, or no disease). Given a dataset containing biological and diagnostic data of 583 Indian patients, this project aims to identify a suitable machine learning algorithm which is capable of identifying whether a person has liver disease or not.

ABSTRACT

Medical diagnoses have important implications for improving patient care, research, and policy. For a medical diagnosis, health professionals use different kinds of pathological methods to make decisions on medical reports in terms of the patients' medical conditions. The use of artificial intelligence and machine learning in combination with clinical findings has further improved disease detection. Machine learning (ML), data-driven algorithms can be utilized to validate existing methods and help researchers to make potential new decisions. The purpose of this study was to extract significant predictors for liver disease from the medical analysis of 615 humans using ML algorithms. The study compared binary classifier machine learning algorithms (i.e., artificial neural network, random forest (RF), and support vector machine), which were utilized on a published liver disease data set to classify individuals with liver diseases, which will allow health professionals to make a better diagnosis. ML methods predict liver disease by incorporating the risk factors, which may improve the inference-based diagnosis of patients.

HARDWARE SPECIFICATION

Machine	I3 8th or above
Memory	8 GB
SSD	128 GB
Monitor	18.5" LED Monitor
Keyboard	USB/3
Mouse	USB/3

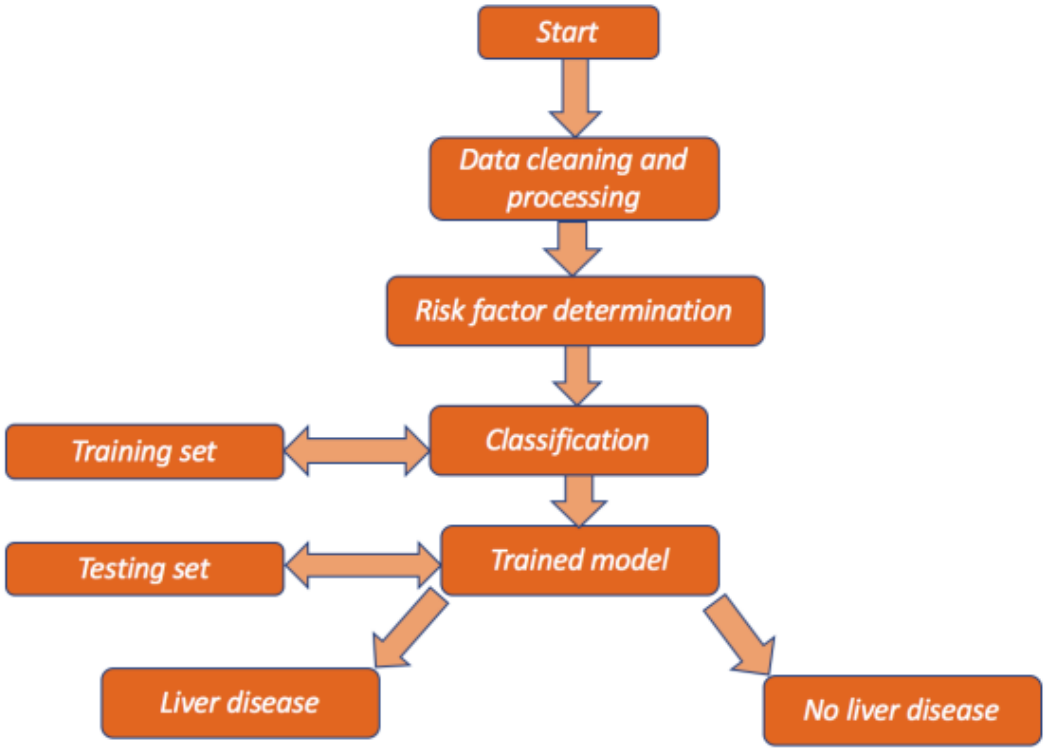
SOFTWARE SPECIFICATION

Operating System	Windows
Web Technologies	Python, Django, HTML, CSS
Database	MySQL
Web Browser	Google Chrome/Mozilla Firefox

OBJECTIVE

This project helps to diagnose liver disease using Random Forest Algorithm.

METHODOLOGY



In this project we collect data from data set, and the health specialist can enter the data for test using our web application in our application we are going through Data Cleaning and Data Pre-processing, Expletory Data Analysis, Data Visualization, Machine Learning - Supervised Learning Algorithms, Decision Trees, K Nearest Neighbours, Logistic Regression ,Support Vector Machines.

LIMITATIONS

This is website and most of the users does not able to access the website. If it is an application then everyone can access the site. Here the user can only register and login into the site. The user does not have any other options to interact with the site. The admin can only give the inputs to the system.

CONCLUSION

Initially, the dataset was explored and made ready to be fed into the classifiers. This was achieved by removing some rows containing null values, transforming some columns which were showing skewness and using appropriate methods (Label Encoding) to convert the labels so that they can be useful for classification purposes. Performance metrics on which the models would be evaluated were decided. The dataset was then split into a training and testing set. Firstly, a naive predictor and a benchmark model ('Logistic Regression') were run on the dataset to determine the benchmark value of accuracy. The greatest difficulty in the execution of this project was faced in two areas- determining the algorithms for training and choosing proper parameters for fine-tuning. Initially, I found it very vexing to decide upon 3 or 4 techniques out of the numerous options available in sklearn. This exercise made me realize that parameter tuning is not only a very interesting but also a very important part of machine learning. I think this area can warrant further improvement, if we are willing to invest a greater amount of time as well as computing power.