

CSE 523 Machine Learning

Progress Report - 1

Section 1

Date of Submission: 9th February 2022

Group Name: **Bug Smashers**

Group Details

| Roll No | Name |
| --- | --- |
| AU1940028 | Moksh Doshi |
| AU1940120 | Jaimik Patel |
| AU1940130 | Nandish Patel |
| AU1940164 | Jenil Bagadiya |

2022 Winter Semester

# Tasks Performed in the week

* + Literature Survey

# Outcomes of the tasks performed

* <https://link.springer.com/article/10.1007/s42979-020-00365-y>

**Our Understanding:** This article details various different approaches to predict heart diseases using machine learning techniques. It explores supervised and unsupervised learning for the given dataset. An extensive description of outputs of various different techniques such as K-Nearest Neighbors, Random Forest, Decision Trees, Naive Bayes were compared. Furthermore, some modifications were applied to the above mentioned algorithms to achieve even better results. This paper gave us a holistic view of various algorithms which could be used to successfully predict whether or not a person will be diagnosed with heart failure in the future.

* <https://www.hindawi.com/journals/cin/2021/8387680/>

**Our Understanding:** Correct prediction of heart diseases helps in saving lives, however, incorrect prediction can lead to irrecoverable conditions as well. Using classical machine learning algorithms with help of analysis and results of Heart Disease dataset 94.2% accuracy can be achieved. In today’s time 17.9 million people die annually just because of cardiovascular disease which is one of leading causes of death globally. Using SVM (Supervised Machine Learning) algorithm for the patients who already have diabetes and after taking some features like blood sugar level, age of patient and the blood pressure data of the patient we can predict the future outcome with a high degree of accuracy.

* <https://www.researchgate.net/publication/326733163_Prediction_of_Heart_Disease_Using_Machine_Learning_Algorithms>

**Our Understanding:** The paper discusses different techniques which could be used on datasets for heart illness prediction. Post selection of attributes and pre-processing, Naive Bayes algorithm which uses Bayes theorem was discussed for making independent assumptions. Then the ID3 algorithm is used to build decision trees. Further k-means clustering technique is used to cluster datasets based on nearest-neighbor. No one way was sufficient and for accurate prediction in small dataset the former is suited and the latter for large dataset. Thus, it provided a way on how to go about the project.

# Tasks to be performed in the upcoming week

* + Google Collab Setup
  + Dataset Collection from UCI ML Repo
  + Data Cleanup & Analysis
  + Compile Final Features to be used in the future