Cardiovascular diseases (CVDs) are a group of disorders of the heart and blood vessels, and they include:

Coronary heart disease – a disease of the blood vessels supplying the heart muscle

Cerebrovascular disease – a disease of the blood vessels supplying the brain

Peripheral arterial disease – a disease of blood vessels supplying the arms and legs

Rheumatic heart disease – damage to the heart muscle and heart valves from rheumatic fever, caused by a bacteria

Congenital heart disease – non-function of heart structure existing at birth

Deep vein thrombosis and pulmonary embolism – blood clots in the leg veins, which can cause blockage and move to the heart, lungs, and various other parts of the body

Many people die annually from CVDs than from any other cause.

An estimated 17.9 million people died from CVDs in 2021, representing 31% of all global deaths. Of these deaths, 85% are due to heart attack and stroke

Our Project goal is to predict the possibility of a person having cardiovascular disease or not based on various parameters specified in the dataset provided by Svetlana Ulianova on Kaggle.

The dataset had impurities which had to be rectified using various preprocessing/PCA methods

We ran Logistic Regression, Random Forest, and a Simple neural network on the preprocessed data set.

The conclusion states Simple neural network had the highest accuracy of 0.73 followed by Random Forest at 0.72. The precision and recall were also at 0.72 for RF. We also went ahead and applied NLTK to the output of models’ It is a powerful tool to preprocess text data for further analysis like with ML models for instance. It helps convert text into numbers, which the model can then easily work with.

The output of the NLTK program would pop a user input where we need to give all the basic inputs like age, systolic blood pressure, diastolic blood pressure, cholesterol level, and pulse rate, separated by commas. Output would either be 1 or 0