**Introduction**

Diabetes is one of the most common human diseases and has become a significant public health concern worldwide. There were approximately 450 million people diagnosed with diabetes that resulted in around 1.37 million deaths globally in 2017. More than 100 million US adults live with diabetes, and diabetes was the seventh leading cause of death in the US in 2020. One in ten US adults have diabetes now, and if the current trend continues, it is projected that as many as one in three US adults could have diabetes by 2050. . Diabetes patients are at elevated risk of developing health complications such as kidney failure, vision loss, heart disease, stroke, premature death, and amputation of feet or legs, which can lead to dysfunction and chronic damage of tissue.

This document is a short and brief explanation of results obtained after building the diabetes prediction model using Logistic Regression Classifier. It also includes the Receiver Operating Characteristic.

**Brief clarification of the method used**

**Logistic Regression**

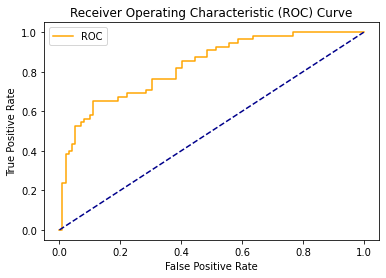
Logistic regressions model a relationship between the categorical response variable and covariates. There is a linear combination of independent variables with log-odds of the probability of an event in a logistic model. Binary logistic regressions estimate the likelihood that a characteristic of a binary variable is present, given the values of the covariates.

**Obtained results**

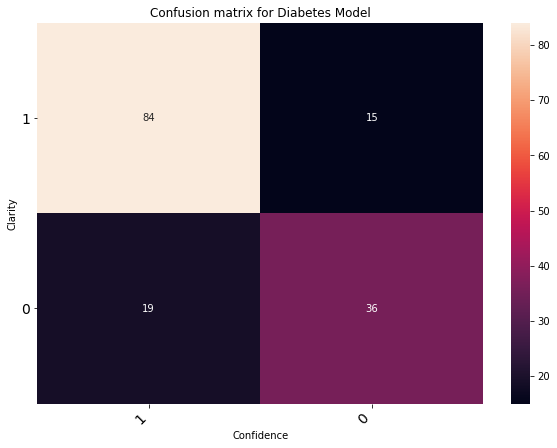
**Prediction Accuracy**

Our variable outcome, called “Outcome”, is a binary numeric variable indexed by 1 and 0. Accordingly, we recoded variable labels so that “1” indicates the diabetes patient and “0” otherwise. We conclude that our model has a prediction accuracy of 78%.

**Receiver Operating Characteristic curve**



**Confusion matrix**



**Conclusion**

Identifying individuals at high risk of developing diabetes is a critical component of disease prevention and healthcare. This study presents a predictive equation of diabetes to provide a better understanding of risk factors that could assist in classifying high-risk individuals, make the diagnosis, and prevent and manage diabetes. Five critical variables identified in predicting type 2 diabetes are age, BMI, pedigree, glucose and frequency of pregnancies. We conclude that our proposed model has a prediction accuracy of 78%.