**ABSTRACT**

Diabetes is one of the deadliest and chronic diseases. It can be characterized by high blood sugar levels over a prolonged period. It prevents our body from properly using the energy from the food we eat. Many complications arise if diabetes remains untreated and unidentified. The tedious identifying process results in consulting a doctor. But the rise in machine learning approaches solves this critical problem. The motive of this project is to prognosticate the likelihood of diabetes in patients with maximum accuracy. Machine learning classification algorithms such as Logistic regression, K-nearest neighbor, Decision tree, Support vector machines and Random Forest are used for predictions at initial stage. Predictions was done on the Pima Indians Diabetes Dataset (PIDD) which is sourced from UCI Machine Learning Repository. The dataset includes attributes such as pregnancies, BMI, blood pressure, age, skin thickness, diabetic pedigree function, glucose level and insulin level which are crucial factors for predicting the presence of diabetes in patients. We randomly choose 700 healthy people and ‘diabetic patients’ data as training set. The performance of the algorithm are evaluated using Accuracy score. Results obtained show logistic regression algorithm outperforms with highest accuracy of 80.51% comparatively with other algorithms. The results are verified using ROC Receiver Operating Characteristic curves in a proper and systematic manner. The Machine learning model is deployed at Watson Studio in IBM Cloud. Node Red is a flow based programming tool which is been used to create an user interface for Machine Learning model.