**PREDICTING THE READMISSION FOR DIABETIC PATIENTS**

ABSTRACT:

Hospital readmission is considered an effective measurement of care provided within healthcare. This study acknowledges that while HbA1c is important, it may not be critical in predicting readmissions. It also investigates the hypothesis that using machine learning on a wide feature, making use of model diversity, and blending prediction will improve the accuracy of readmission risk predictions compared with existing techniques. A dataset originally containing 1010 admissions and 8 features was used to evaluate the hypothesis. The results from the study are encouraging and can help healthcare providers improve inpatient diabetic care.

INTRODUCTION:

Python is a popular programming language. It was created by GUIDO VAN ROSSUM and released in 1991.It is used for

* Web development (server-side),
* Software development,
* Mathematics,
* System scripting.

Python works on different platforms (Windows, Mac, Linux, Raspberry Pi, etc).Python runs on an interpreter system, meaning that code can be executed as soon as it is written. This means that prototyping can be very quick.

Machine learning is a subset of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of Computer Programs that can change when exposed to new data. The term Machine Learning was coined by Arthur Samuel in 1959, an American pioneer in the field of computer gaming and artificial intelligence and stated that “it gives computers the ability to learn without being explicitly programmed”. Machine Learning is classified into three types .They are 1.Supervised Learning 2.UnSupervised Learning 3.Reinforcement Learning. Again the supervised learning is classified as 1.Regression and 2.Classification.

Now-a-days we can observe that most of the people are suffering from Diabetes. They use to face problems in controlling or managing blood sugar levels. The most frequent problems are hyperglycaemia or hypoglycaemia. When patients face any of these conditions, they must be admitted immediately to hospital to receive proper care. In general, hospitalization consumes about half of health care expenditure. Hospital readmissions have considerable potential as an indicator of the quality of care. Several studies have documented the relationship between hospital readmission and quality of care. On the other hand, no studies have addressed the relationship between quality of diabetes care and early readmission. Therefore by using Machine Learning we can determine the readmission rate for diabetic patients at a hospital to access compliance of health care providers and to identify factors predicting readmission. In this particularly for the prediction of readmission of patients with diabetes in hospital we select the **classification** model. The classification model is classified as

1. Logistic Regression

2. K-NN

3. SVM

4. Decision Tree

5. Random Forest

Here we perform the prediction by using all the models. Basing on the accuracy of AUC curve we choose the model which has the highest accuracy. In this we selected **Decision Tree** model.

OBJECTIVE OF RESEARCH:

Hospital readmission risk prediction facilitates the identification of patients potentially at high risk so that resources can be used more efficiently in terms of cost-benefit. In this context, several models for readmission risk prediction have been proposed in recent years. The goal of this review is to give an overview of prediction models for hospital readmission, describe the data analysis methods and algorithms used for building the models, and synthesize their results.



PROBLEM STATEMENT:

Predicting hospital readmission for patients with diabetes using machine learning. By using DECISION TREE model which comes under classification helps in predicting the readmission of diabetic patients within particular span of time.

INDUSTRY PROFILE:

The readmission of diabetes patients comes under the Healthcare sector. Healthcare systems are organizations established to meet the health needs of targeted populations. According to the World Health Organizations(WHO), a well functioning healthcare system requires a financing mechanism, a well trained and adequately paid workforce, reliable information on which to base decisions and policies, and well maintained health facilities to deliver quality medicines. So to meet the needs of targeted population we go with a machine learning model where we can train a model which will predict the readmissions of people in hospital with diabetes. In this model it will get trained by its self to predict the readmissions before 3 months. From the dataset the required parameters are considered, basing on the parameters the model predicts whether the readmission of a particular patient takes place or not.

REVIEW OF LITERATURE:

Data.gov, Kaggle, Google public-Datasets are some of the sources where we collected the dataset for readmission of diabetes patient in a hospital. In this dataset we can know the details of diabetic patients according to the survey. Based on the parameters like pregnancies, glucose, blood pressure, skin thickness, insulin, BMI (body mass index), diabetes pedigree function, age we can get the outcome. For the prediction of this model using decision tree we do consider glucose, blood pressure, insulin level, BMI, diabetes pedigree function and age to get the accurate output whether a person readmits in a hospital or not.

DATA COLLECTION:

In this the parameters are pregnancies , glucose, blood pressure, skin thickness, insulin, BMI (body mass index), diabetes pedigree function, age.

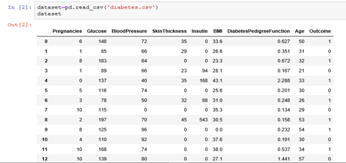
METHODOLOGY:

For predicting a particular model in machine learning there are 6 major steps to be followed:

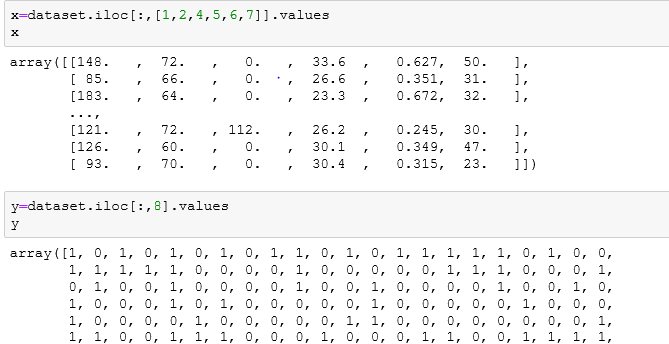


1. Collecting data: Import the data science libraries like pandas, numpy and matplotlib then import the dataset from data source that we have downloaded from data.gov.





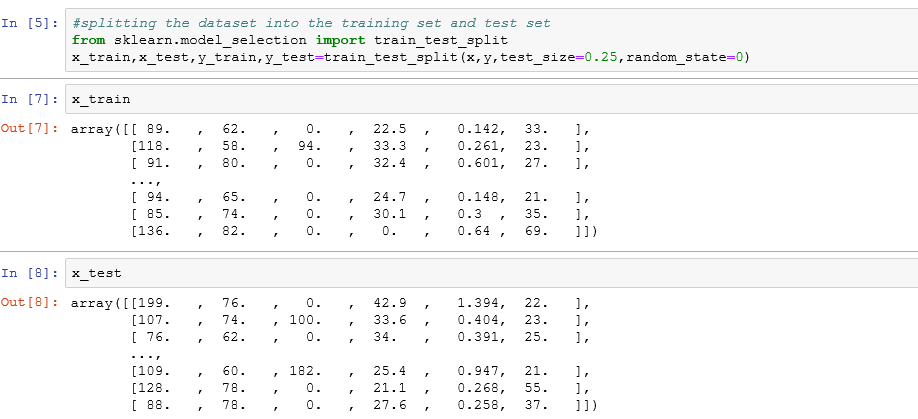
2.Data Wrangling: It is the process of converting and mapping data from its raw form to another format with the purpose of making it more valuable and appropriate for further data analysis.



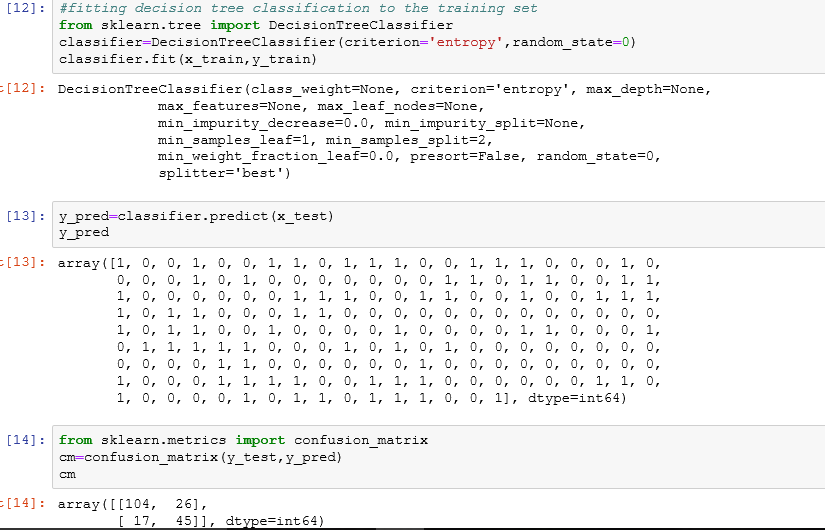
3.Analyse data: The dataset is formatted in a way that best fits in the decision tree model. In the same way, standardizing values in a column will ensure that our data is aggregated correctly.

4.Train algorithm: In this step ,we will use our data to incrementally improve the model’s ability to predict whether readmit or not.

5. Test algorithm: Once training is complete, it’s time to test the model .The test the algorithm allows us to test the model against data that has never been used for training. This is meant to be representative of how the model might perform in the real world.

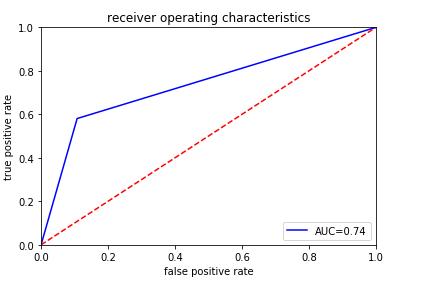


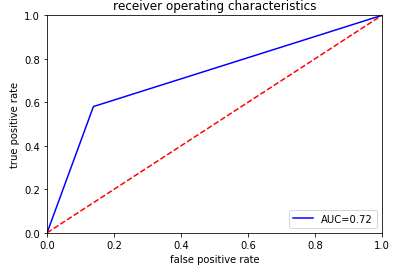
6.Development: After the test is over ,the further improvement in our training can be possible by tuning the parameters. There are few parameters that are implicitly assumed when the training is done.

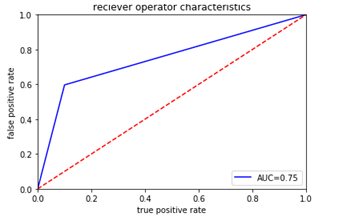
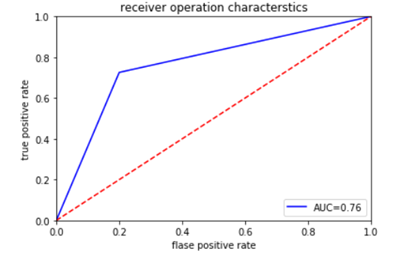


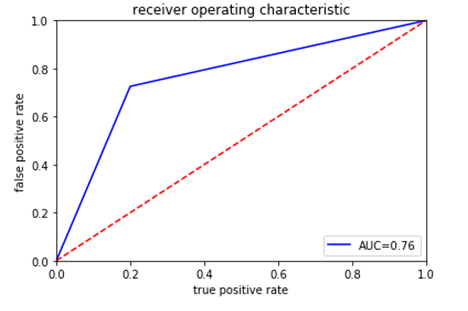
DATA MODELING AND VISUALIZATION:

Data visualization is an integral part of any data science project. Here we use different graphs and plots to visualize complex data to ease the discovery of data patterns. Through visualization one can easily understand the model .This visualization also helps in identifying the plots. In this project, we will utilize the Area under the ROC curve (AUC) to evaluate the best model. This is a good data science performance metric for picking the best model since it captures the tradeoff between the true positive and false positive and does not require selecting a threshold.

**LOGISTIC REGGRESSION K-NEAREST NEIGHBOURS** 



** SIMPLE VECTOR MACHINE RANDOM FOREST**

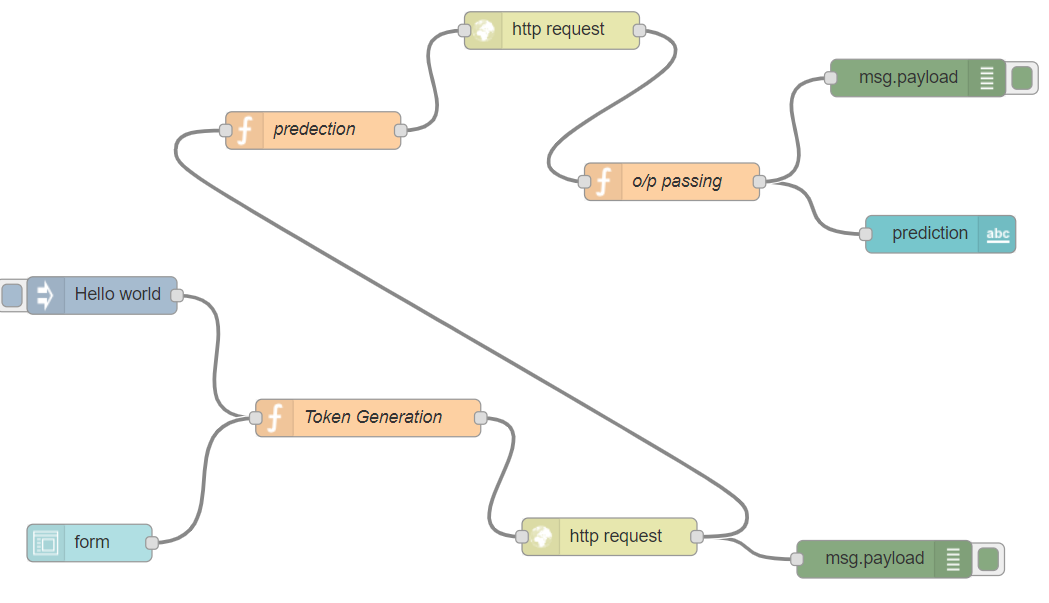
 **DECISION TREE**

From the above graphs we get the highest AUC (Area under curve) value for decision tree so we prefer decision tree classification for this prediction.

DEPLOYMENT

After developing the model we need to deploy the model. Here we use node red for deployment the model. In node red we have different kinds of nodes for function of model.

Node red is a programming tool for wiring together hardware devices, APIs and online services in new and interesting ways. Node-RED provides a browser-based flow editor that makes it easy to wire together flows using the wide range of nodes in the palette. Flows can be then deployed to the runtime in a single-click. The flows created in Node-RED are stored using JSON which can be easily imported and exported for sharing with others. An online flow library allows you to share your best flows with the world. The given below is the flow of the model whether a person with diabetes will readmits in hospital or not,.



FINDINGS AND SUGGESTIONS:

The data analyzed in this study was from a single hospital. This study can be conducted with a larger hospital sample size in rural and urban community settings in multiple states across India. This would allow to determine the readmission factors based on the patient geographical location or if similar traits are observed nationwide. In addition, this would strengthen both urban and rural models while assessing the importance of age categorization.

CONCLUSION:

Diabetes, like other chronic medical conditions, is associated with increased risk of hospital readmission. However, diabetes is not always the main priority for providers in the inpatient setting, and opportunities for optimization of care may be missed without a systematic approach. Preliminary studies suggest that acute and chronic glycemic control may be of importance when diabetes is the primary diagnosis or whether it is a comorbidity. Prospective interventional studies targeting interventions for improving glycemic control in the hospital and after discharge are needed to determine whether glycemic control impacts the frequency of readmissions. In the meantime, efforts to reduce readmissions should be multi factorial and encompass both general and diabetes-specific measures.

By Kingpins Team