SMARTBRIDGE
Let's Bridge the Gap

INTERNSHIP REPORT APPROVAL FORM

July 1, 2019

With immense pleasure, this is to approve that the students of Gudlavalleru Engineering College i.,e

Swetha Verma(1516102143),

Rahul Kurmi(1516102099),

Shankar C(1516102118) and

Pradeep R(1516102090)

successfully completed their Project and Project Report on **"Prediction of Readmission of Diabetic Patients"** under our guidance.

We are highly impressed with the work that they have done and commend them on their quick grasping skills. They have shown good intent to learn and have put the knowledge gained into application in the from of this project. We appreciate the hard work and commitment shown by them.

We, hereby approve that this document is completely checked and accepted by SmartBridge Technical Team. Its been an absolute pleasure to educate and mentor these students. We hope that this document will also serve as a Letter of Recommendation, to whomsover applied.

We wish them success in all future endeavors and a great career ahead.

GD Abhishek

AI Developer

ABSTRACT

Diabetes is linked to an increased likelihood for a number of serious, sometimes life-threatening complications. Proper diabetes management can reduce the risk of complications and improve quality of life. It is widely known that diabetes patients who are educated on disease management techniques are more effective in maintaining proper diet, blood pressure, and blood glucose. They are also more likely to make healthy lifestyle adjustments that result in positive changes, such as ideal body weight; they suffer fewer complications and do not experience frequent acute episodes that require emergency medical care or hospitalizations. However, little is known about which aspects of disease management and patient care factors or what combination of the factors may predict an acute situation that leads to unnecessary hospitalizations.

The objective of Project is to determine whether many independent variables or combinations of variables help to predict a diabetes-related hospital readmission for patients with diabetes. The variables are categorized into different main groups of the patient record in hospital. A convenience sample of 11 historical medical records of patients who were admitted to a rural hospital with a diagnosis of, or relating to, diabetes has studied. After comparing various predictive algorithms Support Vector Classifier performed better and was chosen to analyze a convenience sample of patients admitted to the hospital for a diabetes-related diagnosis. In an effort to simplify the prediction process, all the independent variables were examined. The results of the analysis returned prediction of variable combinations that correctly classified cases for readmission.

While each predictor model for hospital readmission of diabetes patients demonstrated both statistical and practical significance, the overall hit rates were impressive. Most Accurate Model Based on the findings of this study, it can be concluded that, for this sample, the most accurate prediction of readmissions to the hospital for conditions caused by or related to diabetes within a period of time for patients with diabetes.

1. PREDICTION OF READMISSION OF DIABETIC PATIENTS

1.1-INTRODUCTION

Artificial Intelligence is the simulation of human intelligence processes by machine, especially computer systems. These processes include learning, reasoning and self-correction. Particular application of AI includes expert system, speech recognition and machine vision. Popular AI cloud offering includes Amazon AI services, IBM Watson Assistant, Microsoft cognitive services and Google AI services.

Examples of AI technology:

1. Automation
2. Machine Learning
3. Natural language processing (NLP)
4. Robotic
5. Self-driving cars

Machine Learning

Machine is the part of Artificial intelligence that allows computer system to learn from examples, data and experience. Through enabling computers to perform specific tasks intelligently, machine learning systems can carry out complex processes by learning from data, rather than following pre-programmed rules. It includes algorithmic interpretability, robustness, privacy, fairness, inference of causality, human-machine interaction, and security

Python for Machine Learning

Machine learning focuses on the development of Computer Programs that can change when exposed to new data. In this Project we'll implement a simple machine learning algorithm using python. Python is an interpreted, object-oriented, high level programming language with dynamic semantics. Python is widely used for data analytics. Almost all cloud platforms offer support and often new features becomes available in Python first.

Anaconda

Directly from the platform and without involving DevOps, data scientists can develop and deploy AI and machine learning models rapidly into production. Anaconda provides the tools needed to easily:

- Collect data from files, databases, and data lakes
- Manage environments with Conda
- Share, collaborate on, and reproduce projects
- Deploy projects into production with the single click of a button

1.2-OBJECTIVES OF RESEARCH

Diabetes is historically categorized into two categories, Type 1 and Type 2. Until recently, Type 1 was considered a childhood disease and was usually diagnosed before age 40. Nearly 50% of all people diagnosed with Type 1 diabetes are younger than 20 years of age. Type 2 is the most common form, with half of all cases diagnosed in persons over 55 years of age. As with other chronic diseases, persons with diabetes can take measures to reduce the likelihood of developing the disease.

Promoting healthy lifestyles and reducing healthcare costs through effective management of chronic conditions has been a challenge among a variety of diseases, including diabetes. Diabetes, with its complex interaction with other health threats, is particularly challenging for a considerable portion 3 of the population. Attention to biomarkers such as blood glucose and blood pressure can help patients manage and control the disease in order to reduce acute episodes, chronic side effects and unnecessary hospital admissions. Management of diabetes is ideally a team effort. Main players include the physician, diabetes educator, and the patient, while friends and family members play an important supporting role. Much of the disease management responsibility falls to the patient; therefore, effective patient education is pivotal.

The objective of Project is to determine whether many independent variables or combinations of variables help to predict a diabetes-related hospital readmission for patients with diabetes. The variables are categorized into different main groups of the patient record in hospital. A convenience sample of 11 historical medical records of patients who were admitted to a rural hospital with a diagnosis of, or relating to, diabetes has studied. After comparing various predictive algorithms Support Vector Classifier performed better and was chosen to analyze a convenience sample of patients admitted to the hospital for a diabetes-related diagnosis. In an effort to simplify the prediction process, all the independent variables were examined. The results of the analysis returned prediction of variable combinations that correctly classified cases for readmission.

1.3- PROBLEM STATEMENT

Diabetes is linked to an increased likelihood for a number of serious, sometimes life-threatening complications. Proper diabetes management can reduce the risk of complications and improve quality of life. It is widely known that diabetes patients who are educated on disease management techniques are more effective in maintaining proper diet, blood pressure, and blood glucose. They are also more likely to make healthy lifestyle adjustments that result in positive changes, such as ideal body weight; they suffer fewer complications and do not experience frequent acute episodes that require emergency medical care or hospitalizations. However, little is known about which aspects of disease management and patient care factors or what combination of the factors may predict an acute situation that leads to unnecessary hospitalizations.

2. REVIEW OF LITERATURE

This study investigated the relationship between unnecessary hospital readmissions of diabetes patients and the key aspects related to diabetes: patient demographics, lifestyle components, biomarkers, and disease management. The concept of the study is based in the interconnected dynamic relationship of diabetes care aspects with the predictability of emergency treatment or unplanned admissions into a hospital and the economic impact of suboptimal disease management of diabetes. The purpose of this study was to evaluate the contribution of patient demographics, lifestyle components, biomarkers and disease management aspects to predicting hospitalization readmissions for patients with diabetes. This review of the literature will first provide background and context, including definitions of the two types of diabetes, an overview of prevalence of the disease, and an explanation of the economic impact of diabetes. It will present the demographic groups of patients with diabetes and their lifestyle components. The review of patient characteristics and lifestyle will consider age, gender, ethnicity, and patient controlled components that are known to directly affect diabetes such as smoking status and social support.

Essential biomarkers will be explained. Biomarkers such as blood glucose level and blood pressure level are indicators of the level of success of disease management efforts. The types of diabetes education available to assist in patient self-management are described, and the relationship between the physician specialty and avoidable readmission is also examined. Examination of these factors may help predict future acute episodes of diabetes that lead to recidivistic hospital admissions. The economic impact of suboptimal hospital care and self-care practices resulting in recidivistic readmissions will be detailed.

3. DATA COLLECTION

Currently in the U.S., diabetes affects 23.6 million children and adults or 8.0% of the population and is the nation's fifth leading cause of death among diseases. The reported death rate is considered to be underestimated, however, due to the fact that diabetes is often not listed as the cause of death. The American Diabetes Association (2008) reports that of the 280,000 deaths related to diabetes in 2007, only 77,000 noted diabetes as the primary cause. While over 23 million people are diagnosed with the chronic disease, it is believed that 5.7 million people remain undiagnosed; and 57 million people are considered to have pre-diabetes (ADA, 2008).

Many researchers posit that prevalence of diabetes will increase considerably over the next 3 to 5 decades with the largest increase occurring in those over 60 years of age (Narayan Venkat et al., 2006).

It is estimated that 5 to 10% 10 of Americans who are diagnosed with diabetes have Type 1 diabetes. The ADA (2008) describes Type 2 diabetes, the most common form of diabetes, as the body's failure to produce enough insulin or failure to use insulin to convert sugar to energy. Previously known as juvenile diabetes, Type 1 diabetes is usually diagnosed in children and young adults. Type 2 diabetes was formerly known as adult onset diabetes. While obesity is strongly related to Type 2 diabetes, there are other causes that remain unidentified. Although genetics is considered to play a significant role in both types, the main cause still remains a mystery (Christensen et al., 2004). The quality of the patient's self-care and attention to medical biomarkers may predict the extent of recurring acute episodes resulting in emergency room visits or hospitalizations (Smith et al., 2000).

Economic Impact of Diabetes is a complex and costly lifelong disease with no cure in sight. Brandle et al. (2003) state the worldwide prevalence of diabetes is increasing, as is the demand for and cost of medical care. In 2007 the estimated economic cost of diabetes in the U.S. was $174 billion. Twenty percent of these healthcare dollars is spent on diabetes care, and one in ten dollars is used on health care directly attributable to complications of diabetes. But the cost does not stop here; the loss to the nation in economic productivity is nearly $60 billion dollars. Interestingly these astounding totals are incomplete because the cost associated with undiagnosed diabetes, nearly 25%, and the uncompensated care and loss of productivity of family members of patients with diabetes are not included (Fradkin & Rodgers, 2008).

Many persons with diabetes can live healthy lives with proper personal and medical disease management; however, a percentage of persons with diabetes will 11 experience episodes of acute illness that require emergency treatment and often hospitalization. Unnecessary readmissions to the hospital are unfortunately common and costly in terms of quality of life for the patient and in financial terms for both the patient and the healthcare system (Ashton, Kuykendall, Johnson, Wray, & Wu, 1995). The ADA (2008) and Fradkin and Rodgers (2008) suggest that medical costs are 11% higher in patients with diabetes compared to those without.

4. METHODOLOGY

4.1-EXPLORATIVE DATA ANALYSIS

4.1.1-FIGURES AND TABLES

```
In [40]:
data.shape
```

executed in 5ms, finished 12:13:31 2019-06-29

```
(101766, 50)
```

```
In [43]:
data.head()
```

executed in 30ms, finished 12:15:02 2019-06-29

	encounter_id	patient_nbr	race	gender	age	weight	admission_type_id	discharge_disposition_id	admission_source_id	time_in_hospital	...	citc
0	2278392	8222157	Caucasian	Female	[0-10)	?	6	25	1	1	...	No
1	149190	55629189	Caucasian	Female	[10-20)	?	1	1	7	3	...	No
2	64410	86047875	AfricanAmerican	Female	[20-30)	?	1	1	7	2	...	No
3	500364	82442376	Caucasian	Male	[30-40)	?	1	1	7	2	...	No
4	16680	42519267	Caucasian	Male	[40-50)	?	1	1	7	1	...	No

5 rows × 50 columns

Fig4.1 Table definition

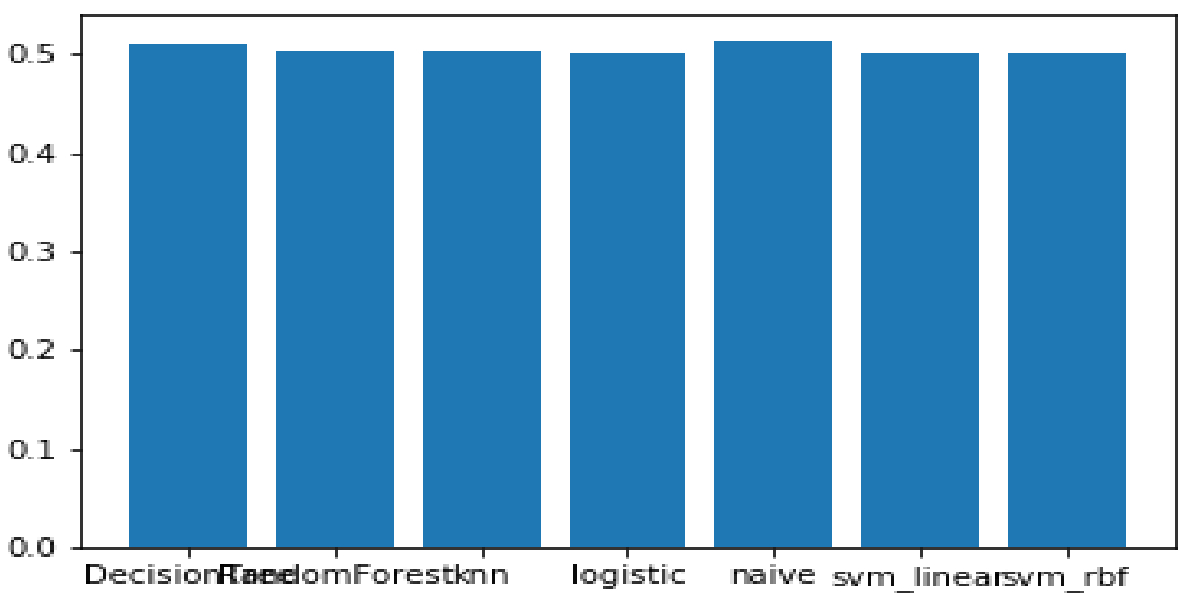


Fig4.2 Accuracy of different Algorithms for prediction

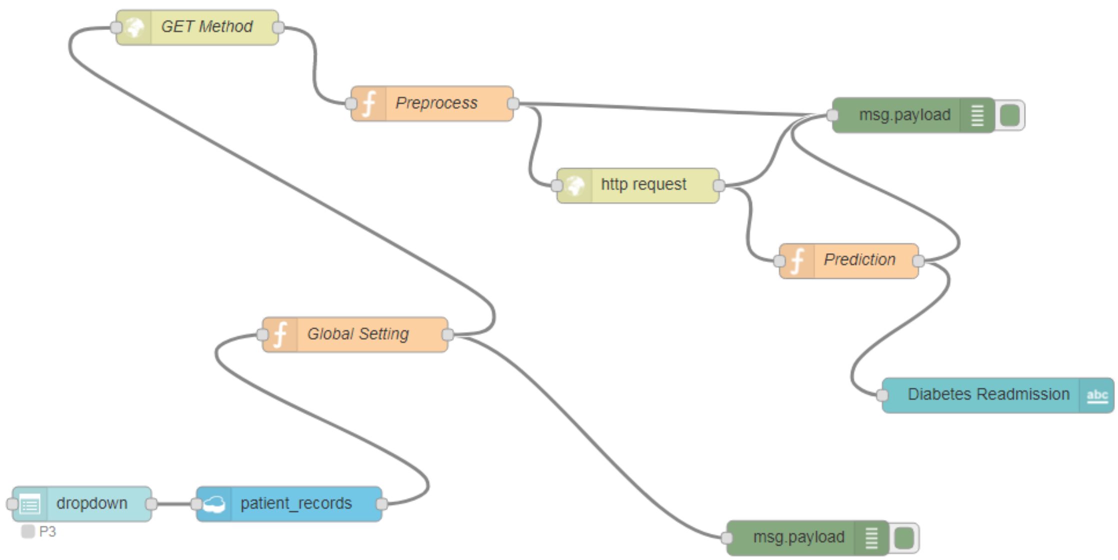


Fig4.3 Web application creation using NodeRed

4.2-DATA MODELLING

We have our final datset. If you won't have the true values how would that the prediction is correct. Now you will realize that, how important the training data phase is. We train the model in a way that it can predict(almost) correct results. In this dataset, we filled missing values using mean, median mode methods. We have split the data into train and test. We will train the model on training data and predict the results on the test data. For this project we will use all the algorithms and find best algorithm i.e.) Support Vector Machines RBF kernel. Before jumping into the code, let's get a little background about the SVM RBF kernel. Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well.

Implementation

Importing data set

```
import types
import pandas as pd
from botocore.client import Config
import ibm_boto3

def __iter__(self): return 0

# @hidden_cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove those credentials before you share your notebook.
client_fb79abb3e90b4d559452e958b889a0e1 = ibm_boto3.client(service_name='s3',
    ibm_api_key_id='up0ENSdaqBPsAqbQB3uKFIkKFL24yCp4_CByktL9K8q8',
    ibm_auth_endpoint="https://iam.bluemix.net/oidc/token",
    config=Config(signature_version='oauth'),
    endpoint_url='https://s3.eu-geo.objectstorage.service.networklayer.com')

body = client_fb79abb3e90b4d559452e958b889a0e1.get_object(Bucket='diabetes-donotdelete-pr-svs2om1aramqgp',Key='diabetic_data.csv')['Body']
# add missing __iter__ method, so pandas accepts body as file-like object
if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType( __iter__, body )
```

```
import numpy as np

import matplotlib.pyplot as plt


data = pd.read_csv(body)

data.describe()
```

Slicing of data

```
y=data.iloc[:,-2:-1].values

data.drop('readmit',axis=1,inplace=True)

x=data.values
```

Encoding of data

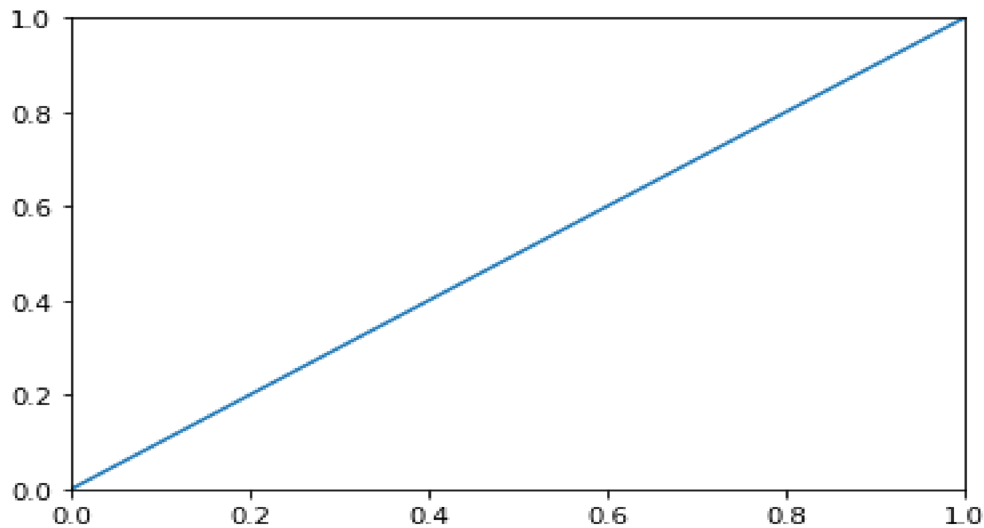
```
from sklearn.preprocessing import LabelEncoder

lb=LabelEncoder()

x[:,0]=lb.fit_transform(x[:,0])

x[:,7]=lb.fit_transform(x[:,7])

x[:,8]=lb.fit_transform(x[:,8])

x[:,10]=lb.fit_transform(x[:,10])

x[0:5,:]

from sklearn.preprocessing import OneHotEncoder

oh1=OneHotEncoder(categorical_features=[7])

oh2=OneHotEncoder(categorical_features=[-1])


from sklearn.model_selection import train_test_split

x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2)
```

Feature Scaling

```
from sklearn.preprocessing import MinMaxScaler

sc = MinMaxScaler()
```

Support Vector Classifier

```
from sklearn.svm import SVC

svc = SVC(kernel='linear')

pipeline = Pipeline([('scaler', sc), ('pip_svc', svc)])

pip_svc = pipeline.fit(x_train , y_train)

y_pred3 = pip_svc.predict(x_test)

accuracy_score(y_pred3,y_test)

0.88812460079595146
import sklearn.metrics as metrics
fpr, tpr, threshold = metrics.roc_curve(y_test, y_pred3)
roc_auc_svm = metrics.auc(fpr, tpr)
roc_auc_svm

0.5
plt.plot(fpr,tpr)

plt.xlim([0,1])

plt.ylim([0,1])

(0, 1)
```




```
from watson_machine_learning_client import WatsonMachineLearningAPIClient
wml_credentials = {
   "url" : "https://eu-gb.ml.cloud.ibm.com",
   "access_key" : "VQbhZuK6LzUUo1vhzPJSfeWebY64FzVLlXmmhgWk-cH-",
   "username" : "c2aa491b-31f4-4ea8-8915-abf87efd9972",
   "password" : "52a130bf-6db5-4818-a69c-4523177f0db9",
   "instance_id" : "fd1c0cea-986c-40ea-82a3-b3c8a908aff4"
}
client=WatsonMachineLearningAPIClient(wml_credentials)
model_props = {client.repository.ModelMetaNames.AUTHOR_NAME:"veeppiaar",
          client.repository.ModelMetaNames.AUTHOR_EMAIL:"veeppiaar.cse@outlook.com",
          client.repository.ModelMetaNames.NAME:"Diabetes Readmission"}
model_s=client.repository.store_model(pip_svc1,meta_props=model_props)
client.repository.list()
pubilshed_model_uid=client.repository.get_model_uid(model_s)
```

```
pubilshed_model_uid
'084379bd-9e2b-48b9-bcc7-0e130b35130d'

d=client.deployments.create(pubilshed_model_uid,name='Diabetes Readmission')

pip_svc1.predict([[0,1,1,0,0,0,1,41,0,1,0,1,0,1]])
array([0])

scoring_endpoint=client.deployments.get_scoring_url(d)

scoring_endpoint
'https://eu-gb.ml.cloud.ibm.com/v3/wml_instances/fd1c0cea-986c-40ea-82a3-b3c8a908aff4/depl
oyments/69308ca3-99ef-4de5-8d1c-0e8fd47d197a/online'
```

5.REFERENCES

[1] https://www.kaggle.com/pavan2029/diabetic-data/downloads/diabetic_data.csv/1

[2] https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html

[3] https://www.geeksforgeeks.org/introduction-machine-learning-using-python/

[4] https://medium.com/datadriveninvestor/classification-algorithms-in-machine-learning-85c0ab65ff4

[5] https://towardsdatascience.com/supervised-machine-learning-classification-5e685fe18a6d

[6] https://towardsdatascience.com/predicting-hospital-readmission-for-patients-with-diabetes-using-scikit-learn-a2e359b15f0

6.CONCLUSION

While each predictor model for hospital readmission of diabetes patients demonstrated both statistical and practical significance, the overall hit rates were impressive. Most Accurate Model Based on the findings of this study, it can be concluded that, for this sample, the most accurate prediction of readmissions to the hospital for conditions caused by or related to diabetes within a period of time for patients with diabetes.