**Predicting hospital readmission for patients with Diabetes**

**PROJECT REPORT**

**Submitted by**

**Kaggle geeks**

**S. Mahammad Ali**

**C. Niveda Reddy**

**L. Manjula**

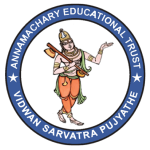
**P. Reddy Rani**

**Harish kumar Reddy**

***In partial fulfilment for the award of the Certificate***

**of**

**SUMMER INTERNSHIP PROGRAM**

**Department of Computer Science and Engineering**

**Annamacharya Institute of Technology and Sciences**

**Venkatapuram Village , Renigunta Mandal , Tirupati , Andhra Pradesh 517520**

**July 2019.**

### BONAFIDE CERTIFICATE

This is to certify that the project entitled ” **Predicting hospital readmission for patients with Diabetes**” submitted by **S. Mahammad Ali , C.Niveda Reddy , L.Manjula , P. Reddy Rani , Harish kumar Reddy** in partial fulfilment for the requirements for the award of internship certification in technologies of Machine learning and Deep learning is an authentic work carried out by them under my supervision and guidance.

To the best of my knowledge, the matter embodied in the project report has not been submitted to any other University/Institute for the award of any Degree or Diploma.

### Signature of Supervisor                                       Signature of Head of the Department

 Akshay Kumar Kothuri, M.Tech       Mrs.B.Rupa Devi M.Tech,Ph.D.,

AI & IoT Developer Assistant Professor & HOD,

SmartBridge Educational Services PVT LTD Department of CSE,

Hyderabad – 500096 AITS,Tirupati.

# Predicting hospital readmission for patients with Diabetes

* 1. **Introduction:**

Predicting hospital readmission to provide the better services to the patients

**Machine learning :**

The term Machine Learning was coined by Arthur Samuel in 1959, “it gives computers the ability to learn without being explicitly programmed”.By using mathematical funtions to it by using variance , standarad devition an soon concepts those are already defined in python packages “A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E . Machine learning has 3 categories

* Supervised learning : guided data used for modelling
* Unsupervised learning : unguided data used for modelling
* Reinforcement learning :learning from mistakes by taking 0 or 1

**Supervised learning :**

* **Classification** : A classification problem is when the output variable is a category, such as “Red” or “blue” or “disease” and “no disease”.
* **Regression** : A regression problem is when the output variable is a real value, such as “dollars” or “weight”.

In this project we have used supervised with in that classification model is used .we have considered KNN (K Nearest neighbor) algorithm to train the machine

**Python :**

Python is a widely used, general purpose, text-based programming language. Python is commonly used as a first text-based language because it’s very readable and you can get things done with less code than many other languages. Python has support for turtle graphics and other modules that make coding more fun and colour ful .Python is used for real projects in industry so it’s a useful language to learn. Our projects make use of the latest version - Python 3.

# Packages used in this project

1. Pandas
2. Sklearn (sikit learn)
   1. **Objective of research :**

* To solve the waiting for treatment problem of a diabetes patients.
* To improve the hospital management services to patients not to be waited.
  1. **Problem Statement :**

As the healthcare system moves toward value-based care, hospitals has created many programs to improve the quality of care of patients. One of these programs is called the Hospital Readmission Program , due to more number of patients, unable to provide hospitality to them . one solution is to create interventions to provide additional assistance to patients with increased risk of readmission. But how do we identify these patients? We can use predictive modeling from data science to help prioritize patients .This project takes Machine learning concept to solve the problem

* 1. **Industry profile :** This project is helpful for the hospital managements to provide better and more services to the patients and to reduce the risk of hospital management.

**2. Literature Review :**

**2.1 Project profile :**

Due to problems in a hospital management system to serve the patients without delay of treatment not to be waited in the wating hall.hospiatal mangers don’t have a future mechanism to know feature how many patients will come . our project will provide these mechanism to know feature statistics of patients

**2.2 project intiation**

As few days back I have suffered a problem in hospital that is waiting for the treatment even I paid payment . because having less services in the hospital for more no.of patients

we considered as this problem and we made a solution using machine learning to improve the hospital services. There after hospital managements will know future how many patients are coming regarding that statistics management people will maintain those required services for the patients.

**2.3 Blogs used**

To get data about the project we gone through [www.kaggle.com](http://www.kaggle.com) to get the data and to know about the project I gone to [www.machinelearningmastery.com](http://www.machinelearningmastery.com) to know the brief explanation about the project and to know the best algorithm suitable for our project.

**3.Data collection :**

**3.1 Data fetched from :**

We have fetched data from kaggle database ([www.kaggale.com](http://www.kaggale.com))

In this project we have used attribures

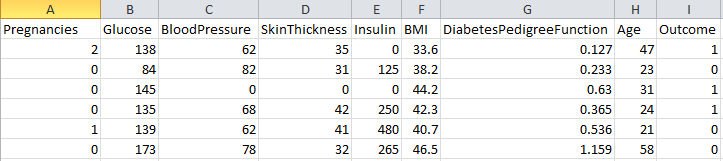
* Pregnancies
* Glucose
* Blood Pressure
* Skin Thickness
* Insulin
* BMI
* Diabetes Pedigree Function
* Age

To predict whether patients will come or not

* outcome

0 – admit not required

1 – admit required



**4 .Methodology :**

I have divided whole data as x as independent values and y as dependent value .

**X values :**

* Pregnancies
* Glucose
* Blood Pressure
* Skin Thickness
* Insulin
* BMI
* Diabetes Pedigree Function
* Age

**Y Value :**

* Outcome

We have choosen classification model and KNN Algorithm to train the data with out standard scalling the data . why because for KNN model accuracy\_score is very good and also auc value is very good in svm, decision tree, random decision forest iam getting accuracy\_score=0.5 ~ 0.6 but for KNN accuracy is 0.8 so I concluded it is best for my project .

**4.1 Project code :**

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\_51bf8485c7f7413f91b3642079f342d5 = ibm\_boto3.client(service\_name='s3',

ibm\_api\_key\_id='ovBjexrAcoSWkzt6DmN3dO24X3iQvUlSDRs4DDU3Hn0W',

ibm\_auth\_endpoint="https://iam.eu-gb.bluemix.net/oidc/token",

config=Config(signature\_version='oauth'),

endpoint\_url='https://s3.eu-geo.objectstorage.service.networklayer.com')

body = client\_51bf8485c7f7413f91b3642079f342d5.get\_object(Bucket='multiplelinearregression-donotdelete-pr-37blgboxafmsao',Key='diabetes.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 )

data = pd.read\_csv(body)

data .head()

**#seperating coloumns for train data & test data**

x=data.iloc[:,:8].values

y=data.iloc[:,8].values

#spliting train , test data from x ,y

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,random\_state=0)

**#we have used KNN algorithm to train the data**

from sklearn.neighbors import KNeighborsClassifier

classifier=KNeighborsClassifier(n\_neighbors=5,metric='minkowski',p=2,weights='uniform').fit(x\_train,y\_train)

**# creating watson API**

from watson\_machine\_learning\_client import WatsonMachineLearningAPIClient

**#proving machine learning credentials to watson studio or a model**

wml\_credentials={

"url": "https://eu-gb.ml.cloud.ibm.com",

'access\_key':"aPtRn2Zid8m180WwA3ExVG-l0C-JJ\_zNVkOlxVNOMUER",

'username':"4cad6cbd-e69f-426f-941c-df61ffa87029",

'password':"7661d008-beaf-4936-a03a-f582d64c3404",

'instance\_id':"904116f2-0752-4e8d-8da4-d344ad2b4456"

}

client=WatsonMachineLearningAPIClient(wml\_credentials)

**#Assigning details to the model**

model\_props={

client.repository.ModelMetaNames.AUTHOR\_NAME:'shaikmahammadali',

client.repository.ModelMetaNames.AUTHOR\_EMAIL:'shaikmahammadali400@gmail.com',

client.repository.ModelMetaNames.NAME:'hospital',

}

**#To store the model in to the machine learning service**

model\_artifact=client.repository.store\_model(classifier,meta\_props=model\_props)

**# Getting model UID**

published\_model\_uid=client.repository.get\_model\_uid(model\_artifact)

**# creating a deployment**

created\_deployment=client.deployments.create(published\_model\_uid,name='hospital')

**# getting End scoring url**

scoring\_endpoint=client.deployments.get\_scoring\_url(created\_deployment)

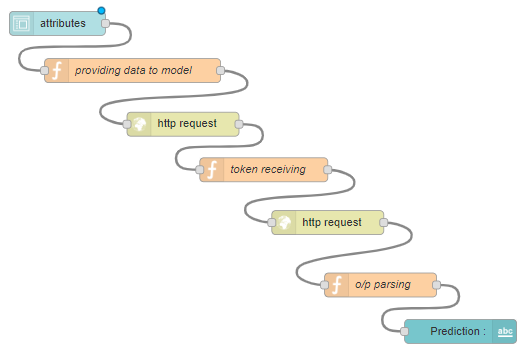
**Deployment :**

we have deployed this model in Watson studio successfully to provide user interface we have used NODE RED

**End\_Scoring\_url** :

<https://eu-gb.ml.cloud.ibm.com/v3/wml_instances/904116f2-0752-4e8d-8da4-d344ad2b4456/deployments/b41619c5-42ce-4638-80e9-d196e40e96c3/online>

**4.2 NODE RED flow :**



**4.2.1 Function 1 code (providing data to model) :**

Var Base64={\_keyStr:"ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz0123456789+/=",encode:function(e){var t="";var n,r,i,s,o,u,a;var f=0;e=Base64.\_utf8\_encode(e);while(f<e.length){n=e.charCodeAt(f++);r=e.charCodeAt(f++);i=e.charCodeAt(f++);s=n>>2;o=(n&3)<<4|r>>4;u=(r&15)<<2|i>>6;a=i&63;if(isNaN(r)){u=a=64}else if(isNaN(i)){a=64}t=t+this.\_keyStr.charAt(s)+this.\_keyStr.charAt(o)+this.\_keyStr.charAt(u)+this.\_keyStr.charAt(a)}return t},decode:function(e){var t="";var n,r,i;var s,o,u,a;var f=0;e=e.replace(/++[++^A-Za-z0-9+/=]/g,"");while(f<e.length){s=this.\_keyStr.indexOf(e.charAt(f++));o=this.\_keyStr.indexOf(e.charAt(f++));u=this.\_keyStr.indexOf(e.charAt(f++));a=this.\_keyStr.indexOf(e.charAt(f++));n=s<<2|o>>4;r=(o&15)<<4|u>>2;i=(u&3)<<6|a;t=t+String.fromCharCode(n);if(u!=64){t=t+String.fromCharCode(r)}if(a!=64){t=t+String.fromCharCode(i)}}t=Base64.\_utf8\_decode(t);return t},\_utf8\_encode:function(e){e=e.replace(/\r\n/g,"n");var t="";for(var n=0;n<e.length;n++){var r=e.charCodeAt(n);if(r<128){t+=String.fromCharCode(r)}else if(r>127&&r<2048){t+=String.fromCharCode(r>>6|192);t+=String.fromCharCode(r&63|128)}else{t+=String.fromCharCode(r>>12|224);t+=String.fromCharCode(r>>6&63|128);t+=String.fromCharCode(r&63|128)}}return t},\_utf8\_decode:function(e){var t="";var n=0;var r=c1=c2=0;while(n<e.length){r=e.charCodeAt(n);if(r<128){t+=String.fromCharCode(r);n++}else if(r>191&&r<224){c2=e.charCodeAt(n+1);t+=String.fromCharCode((r&31)<<6|c2&63);n+=2}else{c2=e.charCodeAt(n+1);c3=e.charCodeAt(n+2);t+=String.fromCharCode((r&15)<<12|(c2&63)<<6|c3&63);n+=3}}return t}};

var username="4cad6cbd-e69f-426f-941c-df61ffa87029"

var password="7661d008-beaf-4936-a03a-f582d64c3404";

var encodedString = Base64.encode(username+':'+password);

msg.headers = {"Authorization":"Basic "+ encodedString};

global.set("preg",msg.payload.preg)

global.set("glu",msg.payload.glu)

global.set("bp",msg.payload.bp)

global.set("skn",msg.payload.skn)

global.set("ins",msg.payload.ins)

global.set("bmi",msg.payload.bmi)

global.set("dpf",msg.payload.dpf)

global.set("age",msg.payload.age)

return msg;

**Function 2 token receiving :**

var token=msg.payload.token;

var ad=global.get("ad");

var rd=global.get("rd");

var ma=global.get("ma");

var st1=global.get("st1");

var st2=global.get("st2");

msg.payload={"fields":["Admimistration","R&D","Marketing","state code","state code1"],"values":[[ad,rd,ma,st1,st2]]};

msg.headers={"Authorization":"Bearer "+token}

return msg;

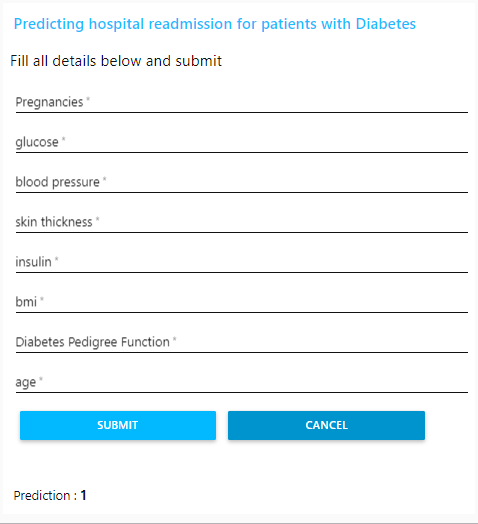
**4.2.2Function 3 o/p parsing :**

msg.payload=msg.payload.values[0][0]

return msg;

**GET and POST method URL in NODE RED** : <https://eu-gb.ml.cloud.ibm.com/v3/identity/token>

**4.3 Our project URL:** <https://mulregmodel.eu-gb.mybluemix.net/ui>



**USER INTERFACE SCREENSHOT**

**5. Findings and Suggestions :**

* It is only helpful for the hospital persons only not that much help for the health insurance companies.
* To increase the income of a company like a health insurance we need to provide outcome as 3 ouputs and also need to groupby() the categories those who are joining in with in 30 days , more than 30 days , those who will not join and display the count for the providing hospital reimbrussments .
* We will provide this feature as next update .
* Due to lack of time we have not included this feature in this version
* Less user attractiveness . provided in future

**6. conclusion :**

**6.1 Pros :**

* It is helpful for the hospital management people to provide better services to the patients
* Improves hospital services
* Get to know about investment where to keep
* Optimize the cost of expenditure
* Know to feature investment of health insurance compan**y**

**6.2 Cons :**

* Will not predict 100% accuractly but predit 80% accuractly
* Predicting outcomes rarely not expected outcomes

**6.3 Feature inventions :**

* To increase the income of a company like a health insurance we need to provide outcome as 3 ouputs and also need to groupby() the categories those who are joining in with in 30 days , more than 30 days , those who will not join and display the count for the providing hospital reimbrussments .
* We have not implemented date wise patient visit to the hospital
* We will provide this feature as next update .

**6.4 Good thing in project :**

* Reducing hospital management risk to maintain the patients in waiting hall.
* Suggesting them to provide services to the patients when patients are more that day