The Estimation of Smart Glucose Predictor with The Sense of Machine Learning Techniques

Mr. T . RAMALINGAM , M.E. Assistant Professor,

Department of Computer Science and Engineering

Mr. R. JAGADESH, B.E, Student of Computer Science Engineering

Mr. S. GOWSHIK, B.E, Student of Computer science and Engineering

St. Joseph College of Engineering, Sriperumbudur, Chennai.

Abstract

Diabetes Is A Chronic Disease With The Potential To Cause A Worldwide Health Care Crisis. According To International Diabetes Federation 382 Million People Are Living with Diabetes Across The Whole World. By 2035, This Will Be Doubled As 592 Million. Diabetes Is A Disease Caused Due To The Increase Level Of Blood Glucose. This High Blood Glucose Produces The Symptoms Of Frequent Urination, Increased Thirst, And Increased Hunger. Diabetes Is A One Of The Leading Cause Of Blindness, Kidney Failure, Amputations, Heart Failure And Stroke. When We Eat, Our Body Turns Food Into Sugars, Or Glucose. At That Point, Our Pancreas Is Supposed To Release Insulin. Insulin Serves As A Key To Open Our Cells, To Allow The Glucose To Enter And Allow Us To Use The Glucose For Energy. But With Diabetes, This System Does Not Work. Type 1 And Type 2 Diabetes Are The Most Common Forms Of The Disease, But There Are Also Other Kinds, Such As Gestational Diabetes, Which Occurs During Pregnancy, As Well As Other Forms. Machine Learning Is An Emerging Scientific Field In Data Science Dealing With The Ways In Which Machines Learn From Experience. The Aim Of This Project Is To Develop A System Which Can Perform Early Prediction Of Diabetes For A Patient With A Higher Accuracy By Combining The Results Of Different Machine Learning Techniques. The Algorithms Like K Nearest Neighbour, Logistic Regression, Random Forest, Support Vector Machine And Decision Tree Are Used. The Accuracy Of The Model Using Each Of The Algorithms Is Calculated. Then The One With A Good Accuracy Is Taken As The Model For Predicting The Diabetes

Key Terms: Machine Learning, Diabetes, Decision Tree, K Nearest Neighbour, Logistic Regression, Support Vector Machine, Accuracy.

Introduction

Diabetes Is The Fast Growing Disease Among The People Even Among The Youngsters. In Understanding Diabetes And How It Develops, We Need To Understand What Happens In The Body Without Diabetes. Sugar (Glucose) Comes From The Foods That We Eat, Specifically Carbohydrate Foods. Carbohydrate Foods Provide Our Body With Its Main Energy Source Everybody, Even Those People With Diabetes, Needs Carbohydrate. Carbohydrate Foods Include Bread, Cereal, Pasta, Rice, Fruit, Dairy Products And Vegetables (Especially Starchy Vegetables).

The Glucose Moves Around The Body In The Bloodstream. Some Of The Glucose Is Taken To Our Brain To Help Us Think Clearly And Function. The Remainder Of The Glucose Is Taken To The Cells Of Our Body For Energy And Also To Our Liver, Where It Is Stored As Energy That Is Used Later By The Body. In Order For The Body To Use Glucose For Energy, Insulin Is Required. Insulin Is A Hormone That Is Produced By The Beta Cells In The Pancreas. Insulin Works Like A Key To A Door. Insulin Attaches Itself To Doors On The Cell, Opening The Door To Allow Glucose To Move From The Blood Stream, Through The Door, And Into The Cell.

Types Of Diabetes Type 1 Diabetes Means That The Immune System Is Compromised And The Cells Fail To Produce Insulin In Sufficient Amounts. There Are No Eloquent Studies That Prove The Causes Of Type 1 Diabetes And There Are Currently No Known Methods Of Prevention. Type 2 Diabetes Means That The Cells Produce A Low Quantity Of Insulin Or The Body Can't Use The Insulin Correctly. This Is The Most Common Type Of Diabetes, Thus Affecting 90% Of Persons Diagnosed With Diabetes. It Is Caused By Both Genetic Factors And The Manner Of Living. Gestational Diabetes Appears In Pregnant Women Who Suddenly Develop High Blood Sugar. In Two Thirds Of The Cases, It Will Reappear During Subsequent Pregnancies. There Is A Great Chance That Type 1 Or Type 2 Diabetes Will Occur After A Pregnancy Affected By Gestational Diabetes.

Literature Survey

Yasodhaet Al.[1] Uses The Classification On Diverse Types Of Datasets That Can Be Accomplished To Decide If A Person Is Diabetic Or Not. The Diabetic Patient's Data Set Is Established By Gathering Data From Hospital Warehouse Which Contains Two Hundred Instances With Nine Attributes. These Instances Of This Dataset Are Referring To Two Groups I.E. Blood Tests And Urine Tests. In This Study The Implementation Can Be Done By Using WEKA To Classify The Data And The Data Is Assessed By Means Of 10-Fold Cross Validation Approach, As It Performs Very Well On Small Datasets, And The Outcomes Are Compared.

The Naïve Bayes, J48, REP Tree And Random Tree Are Used. It Was Concluded That J48 Works Best Showing An Accuracy Of 60.2% Among Others

Aiswaryaet Al. [2] Aims To Discover Solutions To Detect The Diabetes By Investigating And Examining The Patterns Originate In The Data Via Classification Analysis By Using Decision Tree And Naïve Bayes Algorithms. The Research Hopes To Propose A Faster And More Efficient Method Of Identifying The Disease That Will Help In Well-Timed Cure Of The Patients. Using PIMA Dataset And Cross Validation Approach The Study Concluded That J48 Algorithm Gives An Accuracy Rate Of 74.8% While The Naïve Bayes Gives An Accuracy Of 79.5% By Using 70:30 Split.

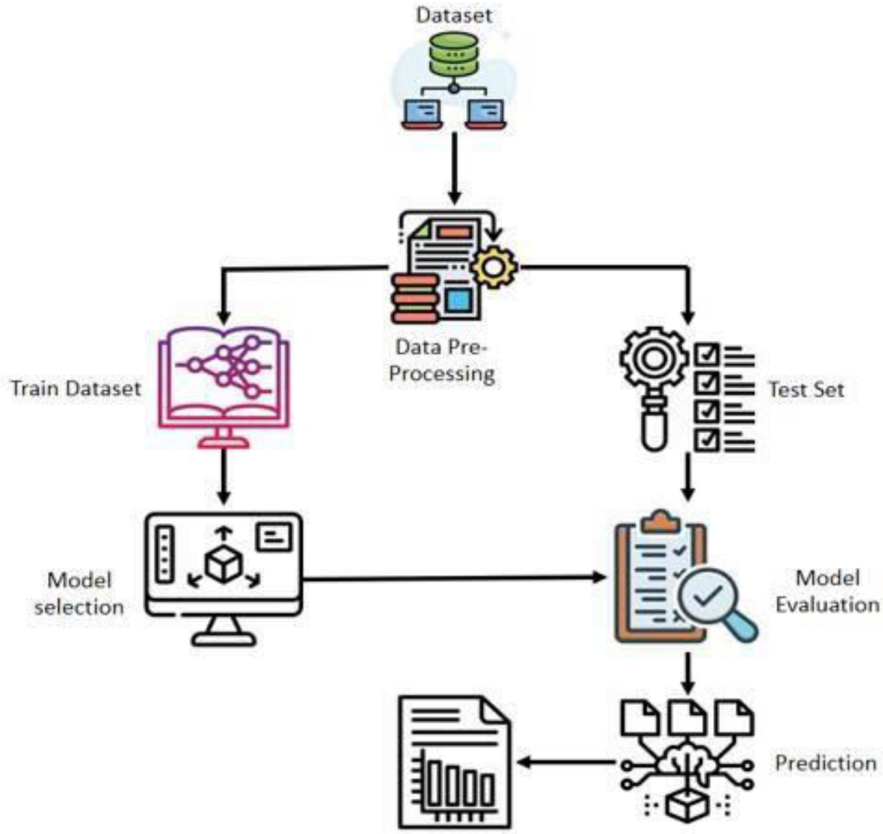
System Design

The Dataset Is The Collection Of Data That Is Used To Train And Evaluate The Machine Learning Model. It Typically Includes Features Such As Age, Gender, Medical History, Lifestyle Factors, And Other Relevant Information That May Contribute To The Risk Of Stroke.

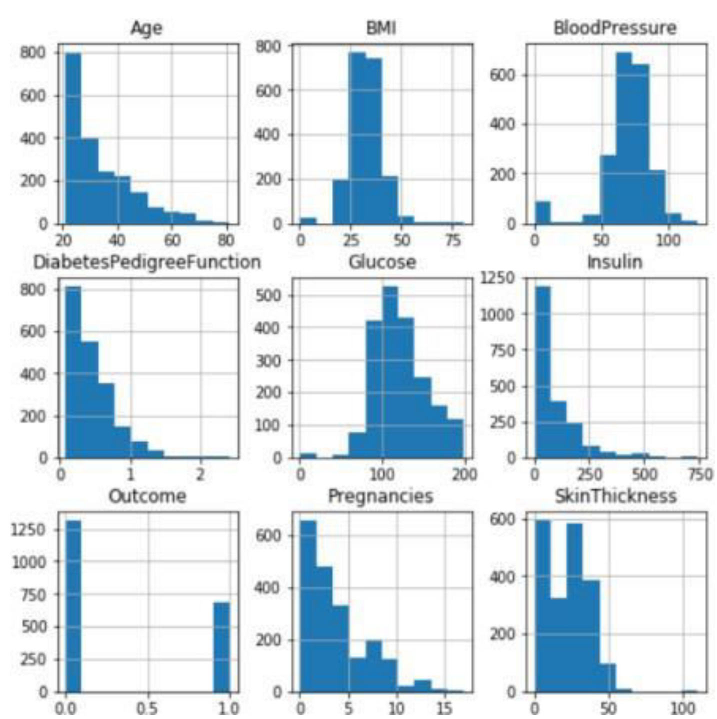
Data Preprocessing Involves Preparing The Dataset For Use In The Machine Learning Model. This May Include Cleaning The Data, Removing Outliers And Errors, Filling In Missing Values, And Scaling Or Normalizing The Data.

The Training Dataset Is A Subset Of The Overall Dataset That Is Used To Train The Machine Learning Model. The Model Learns From The Patterns In The Training Dataset To Make Accurate Predictions.

The Test Dataset Is A Separate Subset Of The Overall Dataset That Is Used To Evaluate The Performance Of The Machine Learning Model.



Histogram:



IMPLEMENTATION

Importing Library Files:

```
import numpy as np
import pandas as pd
import statsmodels.api as sm
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.preprocessing import scale, StandardScaler
from sklearn.model_selection import train_test_split, GridSearchCV, cross_val_score
from sklearn.metrics import confusion_matrix, accuracy_score, mean_squared_error, r2_score, roc_auc_score, roc_curve, classification_report
from sklearn.linear_model import LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
from sklearn.neural_network import MLPClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import GradientBoostingClassifier
from lightgbm import LGBMClassifier
from sklearn.model_selection import KFold
import warnings
warnings.simplefilter(action = "ignore")
```

Reading the dataset:

```
df = pd.read_csv("../input/pima-indians-diabetes-database/diabetes.csv")
df.info()
```

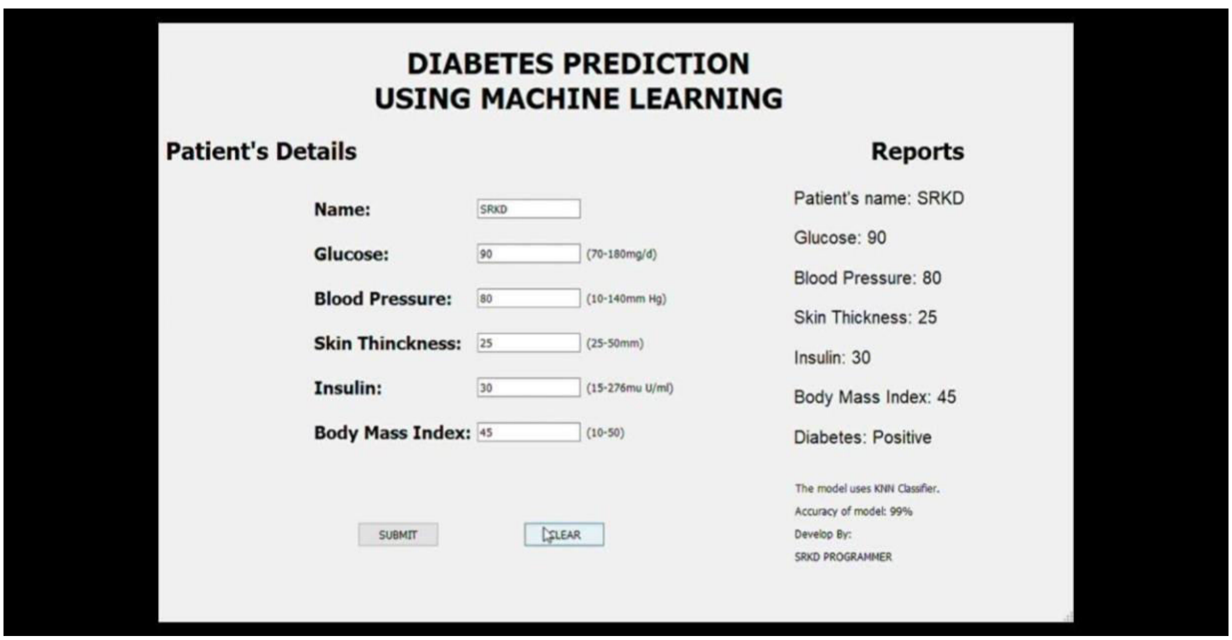
```
df.describe([0.10,0.25,0.50,0.75,0.90,0.95,0.99]).T

df["Outcome"].value_counts()*100/len(df)

print("Max Age: " + str(df["Age"].max()) + " Min Age: " + str(df["Age"].min()))
```

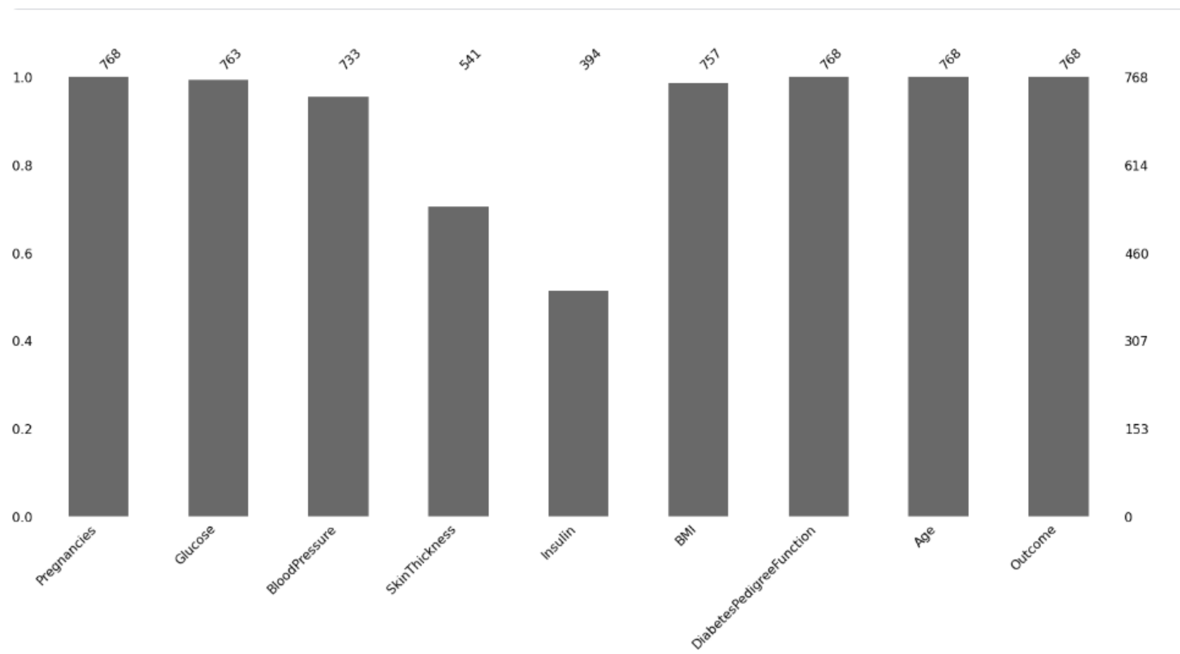
Histogram and density graphs of all variables were accessed:

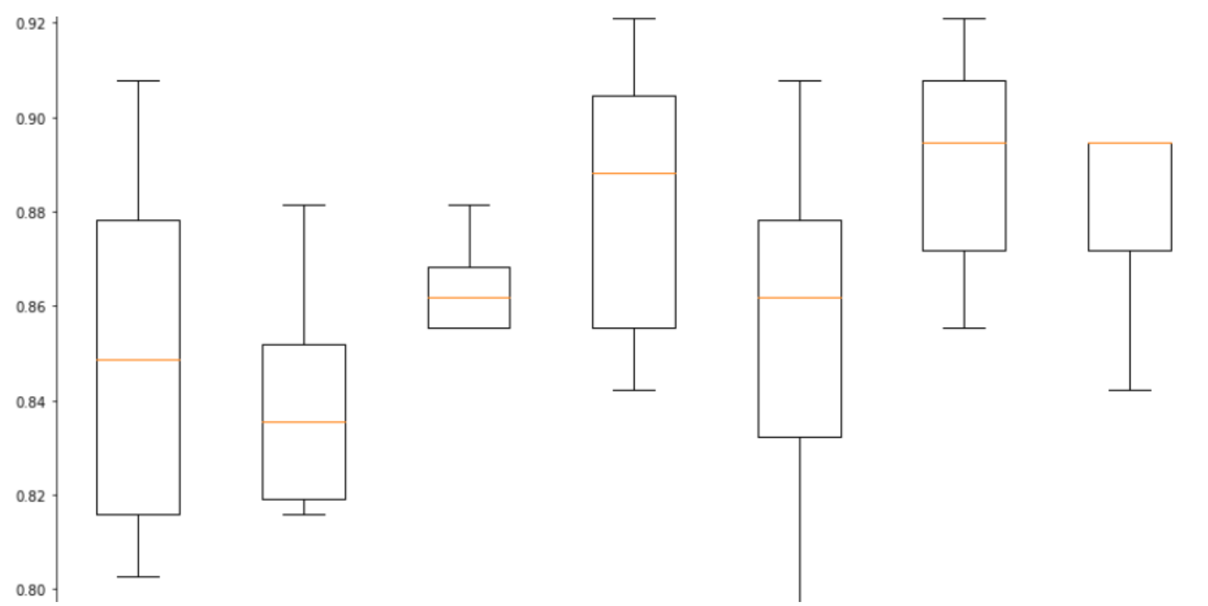
```
fig, ax = plt.subplots(4,2, figsize=(16,16))

sns.distplot(df.Age, bins = 20, ax=ax[0,0])

sns.distplot(df.Pregnancies, bins = 20, ax=ax[0,1])

sns.distplot(df.Glucose, bins = 20, ax=ax[1,0])

sns.distplot(df.BloodPressure, bins = 20, ax=ax[1,1])

sns.distplot(df.SkinThickness, bins = 20, ax=ax[2,0])

sns.distplot(df.Insulin, bins = 20, ax=ax[2,1])

sns.distplot(df.DiabetesPedigreeFunction, bins = 20, ax=ax[3,0])

sns.distplot(df.BMI, bins = 20, ax=ax[3,1])

NewBMI = pd.Series(["Underweight", "Normal", "Overweight", "Obesity 1", "Obesity 2", "Obesity 3"], dtype = "category")

df["NewBMI"] = NewBMI

df.loc[df["BMI"] < 18.5, "NewBMI"] = NewBMI[0]

df.loc[(df["BMI"] > 18.5) & (df["BMI"] <= 24.9), "NewBMI"] = NewBMI[1]

df.loc[(df["BMI"] > 24.9) & (df["BMI"] <= 29.9), "NewBMI"] = NewBMI[2]

df.loc[(df["BMI"] > 29.9) & (df["BMI"] <= 34.9), "NewBMI"] = NewBMI[3]

df.loc[(df["BMI"] > 34.9) & (df["BMI"] <= 39.9), "NewBMI"] = NewBMI[4]

df.loc[df["BMI"] > 39.9 ,"NewBMI"] = NewBMI[5]

def set_insulin(row):

    if row["Insulin"] >= 16 and row["Insulin"] <= 166:
```

```
return "Normal"

else:

return "Abnormal"

NewGlucose = pd.Series(["Low", "Normal", "Overweight", "Secret", "High"], dtype = "category")

df["NewGlucose"] = NewGlucose

df.loc[df["Glucose"] <= 70, "NewGlucose"] = NewGlucose[0]

df.loc[(df["Glucose"] > 70) & (df["Glucose"] <= 99), "NewGlucose"] = NewGlucose[1]

df.loc[(df["Glucose"] > 99) & (df["Glucose"] <= 126), "NewGlucose"] = NewGlucose[2]

df.loc[df["Glucose"] > 126 ,"NewGlucose"] = NewGlucose[3]
```

SNAPSHOTS




768
763
733
541
394
757
768
768
768
1.0
0.8
0.6
0.4
0.2
0.0
768
614
460
307
153
0
Pregnancies
Glucose
BloodPressure
SkinThickness
Insulin
BMI
DiabetesPedigreeFunction
Age
Outcome


0.92
0.90
0.88
0.86
0.84
0.82
0.80

CONCLUSION

One Of The Important Real-World Medical Problems Is The Detection Of Diabetes At Its Early Stage. In This Study, Systematic Efforts Are Made In Designing A System Which Results In The Prediction Of Diabetes. During This Work, Five Machine Learning Classification Algorithms Are Studied And Evaluated On Various Measures. Experiments Are Performed On John Diabetes Database. Experimental Results Determine The Adequacy Of The Designed System With An Achieved Accuracy Of 99% Using Decision Tree Algorithm.

FUTURE ENHANCEMENTS

In Future, The Designed System With The Used Machine Learning Classification Algorithms Can Be Used To Predict Or Diagnose Other Diseases. The Work Can Be Extended And Improved For The Automation Of Diabetes Analysis Including Some Other Machine Learning Algorithms.

REFERENCES:

[1]. Aljumah, A.A., Ahamad, M.G., Siddiqui, M.K., 2013. Application of data mining: Diabetes health care in young and old patients. Journal of King Saud University - Computer and Information Sciences 25, 127–136. doi:10.1016/j.jksuci.2012.10.003.

[2]. Arora, R., Suman, 2012. Comparative Analysis of Classification Algorithms on Different Datasets using WEKA. International Journal of Computer Applications 54, 21–25. doi:10.5120/8626-2492.

[3]. Bamnote, M.P., G.R., 2014. Design of Classifier for Detection of Diabetes Mellitus Using Genetic Programming. Advances in Intelligent Systems and Computing 1, 763–770. doi:10.1007/978-3- 319-11933-5.

[4]. Choubey, D.K., Paul, S., Kumar, S., Kumar, S., 2017. Classification of Pima indian diabetes dataset using naive bayes with genetic algorithm as an attribute selection, in: Communication and Computing Systems: Proceedings of the International Conference on Communication and Computing System (ICCCS 2016), pp. 451– 455.

[5]. Dhomse Kanchan B., M.K.M., 2016. Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis, in: 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication, IEEE. pp. 5–10.

[6]. Sharief, A.A., Sheta, A., 2014. Developing a Mathematical Model to Detect Diabetes Using Multigene Genetic Programming. International Journal of Advanced Research in Artificial Intelligence (IJARAI) 3, 54–59. doi:doi:10.14569/IJARAI.2014.031007.

[7]. Sisodia, D., Shrivastava, S.K., Jain, R.C., 2010. ISVM for face recognition. Proceedings - 2010 International Conference on Computational Intelligence and Communication Networks, CICN 2010 , 554– 559doi:10.1109/CICN.2010.109. [

8]. Sisodia, D., Singh, L., Sisodia, S., 2014. Fast and Accurate Face Recognition Using SVM and DCT, in: Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012, Springer. pp. 1027–1038.

AUTHOR 1

Mr.T.RAMALINGAM, M.E., is a Assistant Professor in the Department of Computer Science and Engineering at St. Joseph College of Engineering, Sriperumbudur, Chennai, Tamil Nadu. He has completed his Ph.D, in Cyber Security - Computer Science and Engineering in 2017 from Manonmaniam Sundaranar University (MSU) Tirunelveli, Tamilnadu. He has done his M.E, CSE in Anna University Chennai in the year 2015.

AUTHOR 2

Mr. R. JAGADESH,B.E., Student of Computer Science and Engineering at St. Joseph College of Engineering, Sriperumbudur, Chennai, Tamil Nadu. I had attended many Workshops, Seminars in Python, Machine Learning.

AUTHOR 3

Mr. S. GOWSHIK ,B.E., Student of Computer Science and Engineering at St. Joseph College of Engineering, Sriperumbudur, Chennai, Tamil Nadu. I had attended many Workshops and Seminars in the area of Python and Machine Learning.