Parkinson's Disease Prediction and Analysis

A Summer Internship Project Report submitted towards the partial fulfilment of the Master of Business Administration Degree with dual specialization in Marketing and Business Analytics.

BY

Abir Ghosh

ROLL NO: **12022001015005**, REGISTRATION NO: **221040710005**

Under the Guidance of:

Mr. Rohit Challa

[EXTERNAL PROJECT GUIDE]

Prof. Prasenjit Kundu

[INTERNAL (IEM) PROJECT GUIDE]

Department of Master of Business

Administration

For the Academic Year 2022-2024

Institute of Engineering & Management

Y-12, Salt Lake, Sector-V, Kolkata-700091

Affiliated To:

Maulana Abul Kalam Azad University of Technology

BF-142, Salt Lake, Sector I, Kolkata-700064

CERTIFICATE

TO WHOM IT MAY CONCERN

*This is to certify that the project report entitled "**PARKINSON'S DISEASE PREDICTION AND ANALYSIS**", submitted by*

[Abir Ghosh]

***(Registration No. 221040710005 of 2022 -2024 MBA Roll no.12022001015005)**,*

*of **INSTITUTE OF ENGINEERING &MANAGEMENT,** in partial fulfilment of requirements for the award of the degree of **Master of Business Administration,** is a bona-fide work carried out under the supervision and guidance of **Prof. Prasenjit Kundu** during the academic session of 2022-2024.The content of this report has not been submitted to any other University or Institute for the award of any other degree.*

It is further certified that work is entirely original and its performance has been found to be quite satisfactory.

---------- ----------

Prof. Prasenjit Kundu
Project Guide
Dept. of Management
Institute of Engineering and Management

Prof. Dr. Sujit Dutta
H.O.D
Dept. of Management
Institute of Engineering and Management

Prof. Anupam Bhattacharyya
Principal – Management
Institute of Engineering and Management
Sector- V, Salt Lake Electronics Complex, Kolkata- 700091

COMPANY CERTIFICATE TO BE ENCLOSED

ACKNOWLEDGEMENT

I like to take this opportunity to extend my gratitude to the following revered persons without whose immense support, completion of this project wouldn't have been possible.

I am sincerely grateful to my Guide ***Prof. Prasenjit Kundu*** *of the Department of Management, IEM Kolkata, for his constant support, significant insights and for generating in a profound interest for this subject that kept me motivated during the entire duration of this project.*

I would also like to express my sincere gratitude to Prof. ***Dr. Satyajit Chakrabarti*** *(Director, IEM),* ***Prof. Anupam Bhattacharya*** *(Principal-Management, IEM) and Prof.Dr. Sujit Dutta, (HOD of Management) and other faculties of Institute of Engineering & Management, for their assistance andencouragement.*

Last but not the least, I would like to extend my warm regards to my families and peers who have kept supporting me and always had faith in my work.

Name of the Student: ***Abir Ghosh***
Reg.No: 221040710005
Roll No: 12022001015005
Dept. of Management
Institute of Engineering & Management, Kolkata

EXECUTIVE SUMMARY

Acmegrade, a leading data-driven company, embarked on a groundbreaking final year project in the field of data science. The project's primary objective was to harness real-life data sets to perform a series of complex data science tasks. In this executive summary, we outline the key aspects of this ambitious initiative.

Acmegrade's final year project in data science aimed to demonstrate the practical applications and value of data analysis, predictive modeling, and machine learning in a corporate setting. Real-life data sets were used to address a variety of business challenges, providing valuable insights and solutions.

Key Objectives:

1. Data Collection and Preprocessing: *The project involved collecting diverse data sets from various industries and domains. These raw data sets were meticulously cleaned, transformed, and preprocessed to ensure data quality and integrity.*

2. Exploratory Data Analysis (EDA): *Through extensive EDA, the project unveiled hidden patterns and relationships within the data. Visualizations and statistical analyses were employed to gain a deep understanding of the datasets.*

3. Predictive Modeling: *Acmegrade leveraged machine learning algorithms to build predictive models for tasks such as customer churn prediction, sales forecasting, and sentiment analysis. These models were fine-tuned and validated for accuracy.*

4. Natural Language Processing (NLP): *The project delved into NLP techniques for text data, including sentiment analysis, topic modeling, and text classification, providing actionable insights for textual data.*

5. Feature Engineering: *Feature engineering played a crucial role in enhancing model performance by extracting relevant information from the data sets. Techniques like dimensionality reduction and feature selection were employed.*

6. Model Deployment: *The final models were deployed in a production environment, demonstrating their real-world usability. APIs and dashboards were created to provide easy access to the insights generated.*

Outcomes:

Acmegrade's final year data science project successfully achieved the following outcomes:

1. Improved Decision-Making: *The project empowered Acmegrade with data-driven insights, enabling better decision-making across various departments and functions.*

2. Increased Efficiency: *Predictive models enhanced operational efficiency by providing early warnings, optimizing inventory management, and streamlining customer support.*

3. Enhanced Customer Satisfaction: *Sentiment analysis and customer profiling enabled Acmegrade to better understand customer needs, resulting in improved product offerings and customer satisfaction.*

4. Valuable Insights: *Acmegrade gained valuable insights into market trends, customer behavior, and industry-specific patterns, creating a competitive advantage.*

The project represents Acmegrade's commitment to harnessing the power of data science in the business world. It demonstrates the real-world impact of data-driven decision-making and the importance of leveraging real-life data sets for practical applications.

Future Work:

Acmegrade intends to continue its data science journey, further refining and expanding the models and analyses, and exploring emerging techniques and technologies in the data science field.

This project serves as a testament to Acmegrade's dedication to innovation, data-driven solutions, and the ongoing pursuit of excellence in the realm of data science.

TABLE OF CONTENTS

INTRODUCTION

Background :

In the present era, disease prediction is the most important task for medical institutions and physicians in order to decide the finest possible physicians' decisions. Possible incorrect decisions are opine of the major cause to delays in medical care or even loss of life. On the other hands, there is another viewpoint about medical services is a big commercial in every time. The business marketplace in this fields always running and rapidly growing than other fields. Most of the patients are always searching for good treatment care for better medical services. In reality, they cannot afford the treatment cost most of the time. Possibly, it's very crucial situation for this patient. Therefore, researchers have always searching in this area how can it is less possible expense for every patient and there is need of one umbrella platform for solving these problems in medical fields. Here is our main objective for improving Parkinson disease treatment is to more importance on early detection of Parkinson disease with less expenses and live for healthier life.

Motivation of the Research :

Parkinson diseases are the most critical causes of death and disability worldwide. According to the Parkinson disease foundation, The affected peoples in the worldwide of Parkinson disease is projected that the 1 million people are Living by 2020 in USA (Marras et al., 2018). The medical treatment of Parkinson disease can be endorsed on Neuropathologic and Histopathologic (Gelb, Oliver, neurology, & 1999, n.d.). Medical diagnostic detection of Parkinson Disease can be done on widespread selection basing on the sensitivity and specificity of the characteristic Parkinson disease features. Therefore, the Parkinson Disease are needed to explore the clinical, pathologic, and nosology studies grounded on frequency of occurrence, characteristics, and including risk factors of samples (Aarsland, Andersen, neurology, & 2003, n.d.). Parkinson usually affects a large part of worldwide patients over the age of 50, which has affected up to now ("Parkinson's Disease Information Page | National Institute of Neurological Disorders and Stroke," n.d.). Still now there is no known cause of Parkinson disease, however, it is very likely possible to assuage symptoms knowingly in the early stage of the subjective patients (Singh, Pillay, neurobiology, & 2007,

n.d.). A study claimed that around 90% of the patients affected with vocal damage ("Speech impairment in a large sample of patients with Parkinson's disease," n.d.). The Parkinson treatment is likely very costly. This causes most of the patients cannot afford the cost of the Parkinson disease. Because if the disease is detected in the initial stage, then the cost will decrease and it will also be possible to save the patient's life. Nowadays, Parkinson disease prediction is most critical matter for clinical practitioners to take accurate decision of such disease. It's a great exercise atpresent time, machine learning based extensive platform can detect Parkinson disease.

Problem Statement :

Medical data has growing a vast scale of volume from different clinical areas including health care services. To handle this data and attaining insights from this data there is need of Big Data analysis through Machine learning that aim to solve diverse medical and clinical problem (Hossain, Mahmud, Hossin, Haider Noori, & Jahan, 2018). Already, many of the studies show that machine learning algorithms has gained meaningfully high performance in classification-based medical problems. However, supervised learning-based methods are one of the most effective method for the research community and real-life applications on clinical fields. (Dwivedi, 2017). This works main objective is to improve the detection and diagnosis techniques of Parkinson disease treatment. Parkinson's disease can't be cured, but medications can help control your symptoms, often dramatically. So if it detects in the early stage, the cost of medication will reduce. Therefore, our study can be playing an important role for the detecting Parkinson disease with machine learning algorithms.

Research Question :

In the following research questions associated to the Parkinson detection is also addressed:

(1) What is the best machine learning techniques within popular supervised learning for Parkinson detection?

Research Objectives :

To study different machine learning algorithm. To evaluate the performance of different machine learning algorithms for Parkinson disease prediction. As well as compare the algorithms result and find out which algorithm is giving the best results.

Research Scope :

In recent, machine learning algorithms have generated a significant influence and commitment in the Parkinson research community for detection of Parkinson disease. Moreover, machine learning 3 © Daffodil International University techniques are specified more precise results in disease prediction as compared to others data taxonomy techniques (Dwivedi, 2017)(Mahmud & Ahmed, 2018) . Motivated by this, the authors have used three prominent machine learning algorithms for detection and proper diagnosis of Parkinson patients. The main goal of this study is to examine the performance measurement of various prominent classification methods for this study we used three supervised learning techniques were used including k-Nearest Neighbors, Support Vector Machine and Logistics Regression. Moreover, the performance of the three classifiers was evaluated using different methods.

Thesis Organization :

The rest of the study is ordered as follows, chapter 1 illustrates the objectives of this thesis, inspiration behind this thesis, research possibility and thesis organization. Chapter 2 portrayed the literature review and related works in these Parkinson disease areas. And the materials and methodology are designated with the evaluation criteria of different classifiers in Section 3. Moreover, the performance results and discussion are illustrated in Section 4. Finally, conclusions and further study are deliberated in Section 5.

COMPANY OVERVIEW

Company Name: Acmegrade

Introduction:

Acmegrade is a leading provider of internships in the field of data science, offering students and aspiring data scientists a unique opportunity to gain practical experience by working with real-life sales data from a diverse range of companies. Established with the mission to bridge the gap between academic knowledge and practical skills, Acmegrade has become a trusted partner for both educational institutions and businesses seeking to nurture the next generation of data professionals.

Key Features:

1. Real-Life Sales Data:

Acmegrade sets itself apart by exclusively utilizing real-life sales data from various companies, providing interns with a genuine, hands-on experience in analyzing, processing, and deriving insights from the data. This approach ensures that interns develop the skills necessary to thrive in the competitive world of data science.

2. Industry Partnerships:

The company has cultivated strong partnerships with a wide array of businesses across different sectors, granting interns access to diverse datasets, which include e-commerce, retail, technology, and more. These partnerships provide a valuable learning experience for participants.

3. Expert Mentorship:

Acmegrade is committed to delivering high-quality mentorship. Interns benefit from guidance and expertise from experienced data scientists, who provide valuable insights, assistance, and constructive feedback throughout their internship journey.

4. Curriculum Integration:

Acmegrade understands the importance of aligning its internships with educational curriculums. It collaborates with educational institutions to ensure that its programs complement and enhance the academic experience, enabling students to apply their classroom knowledge in a real-world context.

5. Skill Development:

The company places a strong emphasis on skill development. Interns are encouraged to engage in practical projects, learn the latest data science tools and techniques, and develop a portfolio showcasing their work, all of which can significantly enhance their career prospects.

6. Job Placement Assistance:

A*cmegrade's commitment doesn't end with the internship. The company also offers job placement assistance to help interns transition into full-time positions in the data science field. It connects graduates with its extensive network of industry partners and provides resources for job-seeking success.*

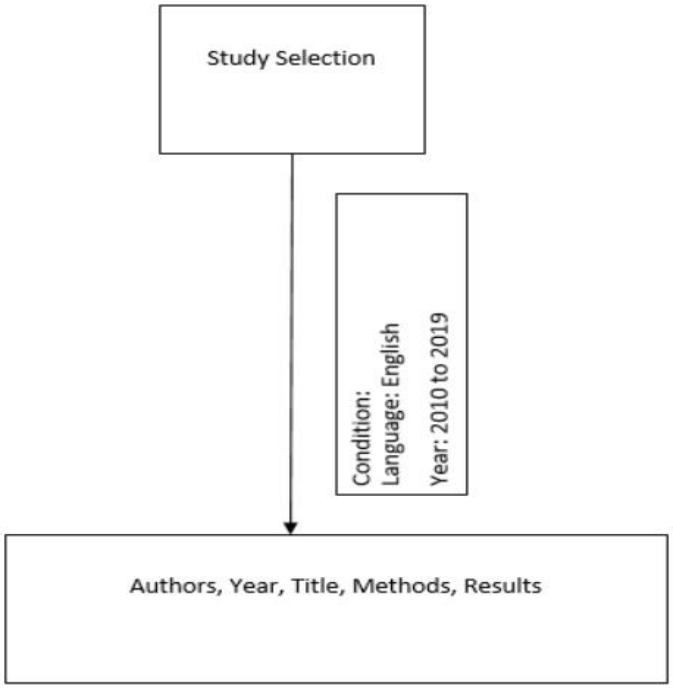
Conclusion:

Acmegrade is at the forefront of empowering aspiring data scientists with the skills, experience, and connections necessary for a successful career in the data science field. By offering internships based on real-life sales data, the company stands as a valuable resource for students, educational institutions, and businesses alike, fostering the growth and development of the data science community.

LITERATURE REVIEW

In this study, this work was designed by the qualitative research process and implements the strategies of Kitchenham and Charters (Kitchenham, Brereton, …, & 2009, n.d.). Our searching process of collecting paper was systematic mapping study (Petersen, Feldt, Mujtaba, Ease, & 2008, n.d.) for searching publication. Basically, cumulative use for paper selection. We have used a few keywords in our searching process. These studies are searched in Parkinson Disease, Parkinson Disease data sets, Machine Learning Techniques, Prediction, detection. We have used and/or/not. By using those keywords, we have created a search sequence which used in different online databases such as – ***IEEE Xplore , Springer Link, ACM Digital Library, Science Direct, Google Scholar, Hindawi*** *.*

Ensuing this search string,to find out journal articles we have used a condition in thesearching process, and it was "English language between the years 2010 to 2019". Then 17 articles we have finally selected that have published in the above-mentioned journals or conferences. We have applied inclusion and exclusion standards which is proposed by the Kitchenham. Our partner researchers reviewed the search verdicts from the systematic process, which helped to moderate the validity extortions.



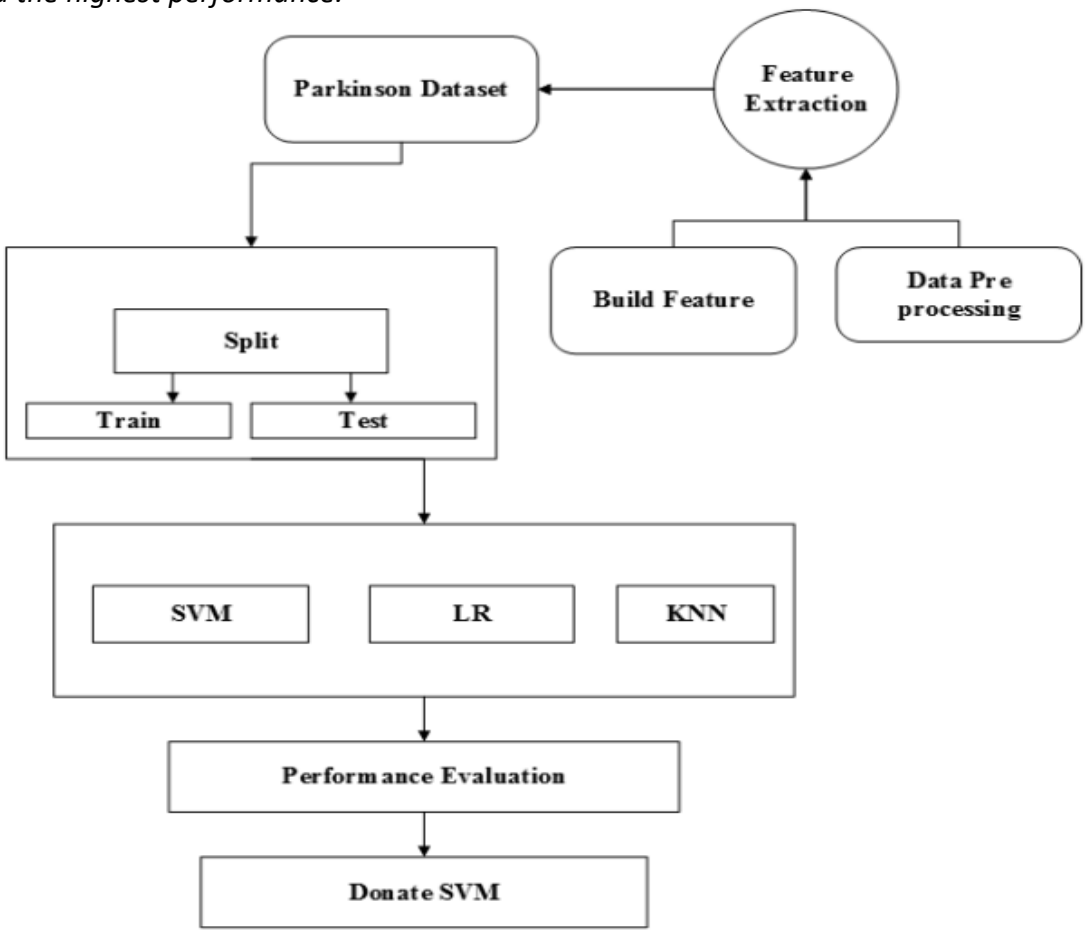
Study Selection Process

Through related work, 17 studies were done on applying and using different machine learning approaches to determine detection of Parkinson Disease. Previous work also introduces a set of studies-based detection of Parkinson diseases using machine learning algorithms. However, the outcomes of the 17 articles on machine learning used in disease prediction as follows: Tarigoppula et al. (Sriram, Rao, Narayana, Kaladhar, & Vital, 2013) presented a comparative study between Naïve Bayes, Random Forest, Logistics Regression, Support Vector Machine to detect Parkinson disease. SVM (i.e. 88.9%) has shown the good performance to compared NB (i.e. 69.23%), and RF (90.26%) shown the compared to SVM for the Parkinson detection. Moreover, LR (i.e. 83.66%) shown the quiet good performance. 86%). And the SVM and LDA have superior sensitivity in comparison to other classifiers. The contribution of this study is to analysis of voice data to understand presence of Parkinson diseases. In order to additional improve the diagnosis accuracy for detection of Parkinson Disease, the study (Chen et al., 2013) proposed a fuzzy based KNN model to predict Parkinson. Their study shown to the best accuracy (96.07%) obtained by the proposed algorithm including a 10-fold cross validation. Another study (Chen et al., 2016) also considers a hybrid model of detection Parkinson with compared to the existing methods and their proposed model has achieved the brilliant accuracy through 10-fold cross-validation analysis, the topmost accuracy of 96.47% and quite good accuracy of 95.97%. Moreover, The experimental (Hariharan, Polat, & Sindhu, 2014) results show that the maximum classification accuracy of 100% for the Parkinson's dataset via feature pre-processing . Hanzel et al.(Hazan, Hilu, Manevitz, Ramig, & Sapir, 2012) presented a new prediction system that can detect of Parkinson from voice data seems to be possible and precise with results approaching (90%) in two different data sets. Another hybrid method (Ma, Ouyang, Chen, & Zhao, 2014) named SCFW-KELM has been presented for the diagnosis of Parkinson disease. The result of proposed method is effective for Parkinson detection by MAE for the Total-UPDRS and Motor-UPDRS were achieved respectively MAE = 0.4656 and MAE = 0.4967 (Nilashi, Ibrahim, Ahmadi, Shahmoradi, & Farahmand, 2018). Moreover, A study (Ozcift, 2012) uses kernel Support Vector Machine for their classification and Neural Network classification scheme. Thus, the prediction performances of the 2 classifiers respectively are 91.4% and 92.9%. Hence, one study (Geetha, Professor, Head, & Sivagami, 2011) found they showed into their study that the Random Forest obtained the highest performance. But, other study showed SVM reaches upright accuracy of 83.33% (Shetty & Rao, 2017). Ferdous et al. (Wahid, Begg, Hass, Halgamuge, & Ackland, 2015) presented a comparative study between different classifiers. Their analysis shown that the RF attained the accuracy of 92.6% after standardizing gait data using the multiple regression method, competed to 80.4% (Support Vector Machine) and 86.2% (Kernel Fisher Discriminant). Hence, the study (Yadav, Kumar, & Sahoo, 2012) compared to different classifiers and showed into the results LR obtained the highest performance than others.

RESEARCH METHODOLOGY

Experimental Setup :

In this study, this section represents the experimental process of the experiment including machine learning techniques.Parkinson Disease data sets have been considered in this work. Firstly, we focused on preparing and combined data from the main datasets. Moreover, we extracted 30 features from the Parkinson datasets. Then, we checked the missing values and co related values.Secondly, Data set splitting is an important task of this machine learning based fields. In this dataset, we have not found split and test datasets. Figure 3.1 shows the Parkinson data set has split into trainset and test sets. After that, 3 supervised based classifiers performed the operation. After successfully executed these algorithms SVM obtained the highest performance.



The Experimental Setup

Data Collection :

Parkinson Disease Datasets :

In this study, we used the Parkinson disease data from provided by the UCI Machine Learning Repository. In addition, this dataset is consisting of 62 people with Parkinson disease and 15 peoples were healthy. The authors used three types of recording are taken such as static spiral test, dynamic spiral test and stability test score. However, we chosen the particular features for data analysis which are below presented,

- *No of strokes*
- *Stroke speed*
- *Velocity*
- *Acceleration*
- *Jerk*
- *Horizontal velocity/acceleration/jerk*
- *Vertical velocity/acceleration/jerk*
- *Number of changes in velocity direction*
- *Number of changes in acceleration direction*
- *Relative NCV*
- *Relative NCA*
- *In air time*
- *On surface time*
- *Normalized in-air time*
- *Normalized on-surface time*
- *In air/on surface ratio*

Data Preprocessing :

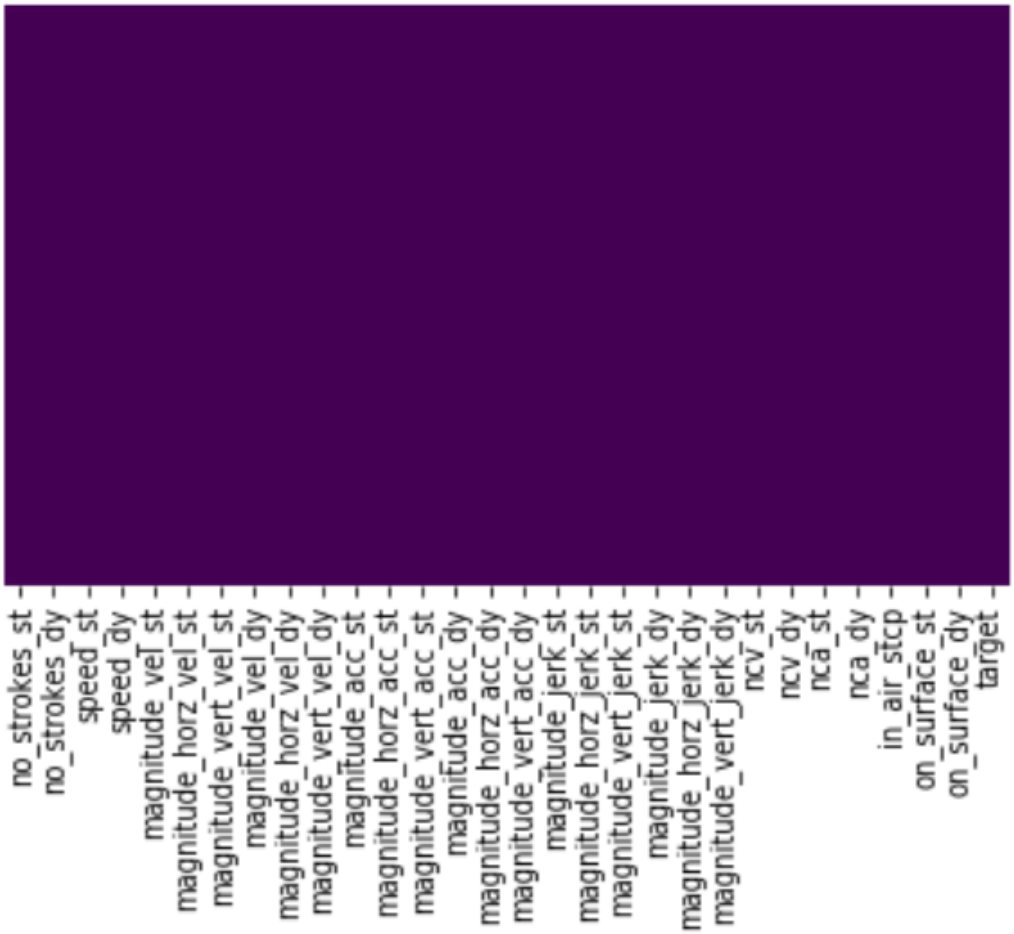
In this section, firstly we extracted features from the Parkinson disease datasets. We picked the 30 columns and 77 entries of data. Then,we conducted several experiments to checking missing values, redundant values. Previous Figure has shown that the 30 features from the

dataset which were we collected. Therefore, analyzing the attributes of the selected Parkinson's datasets, some of them presented a very few values whereas others appeared not correlated with the specific medical event. There were no missing values exist in this dataset. This Figure shows the number of missing values is empty. Moreover, the Parkinson's datasets were also checked to verify the correlation of parameters. The heatmap, which is a two-dimensional graphical representation of data where the individual values that are contained in a matrix are represented as colors that is shown in figure appear to have some correlated parameters.

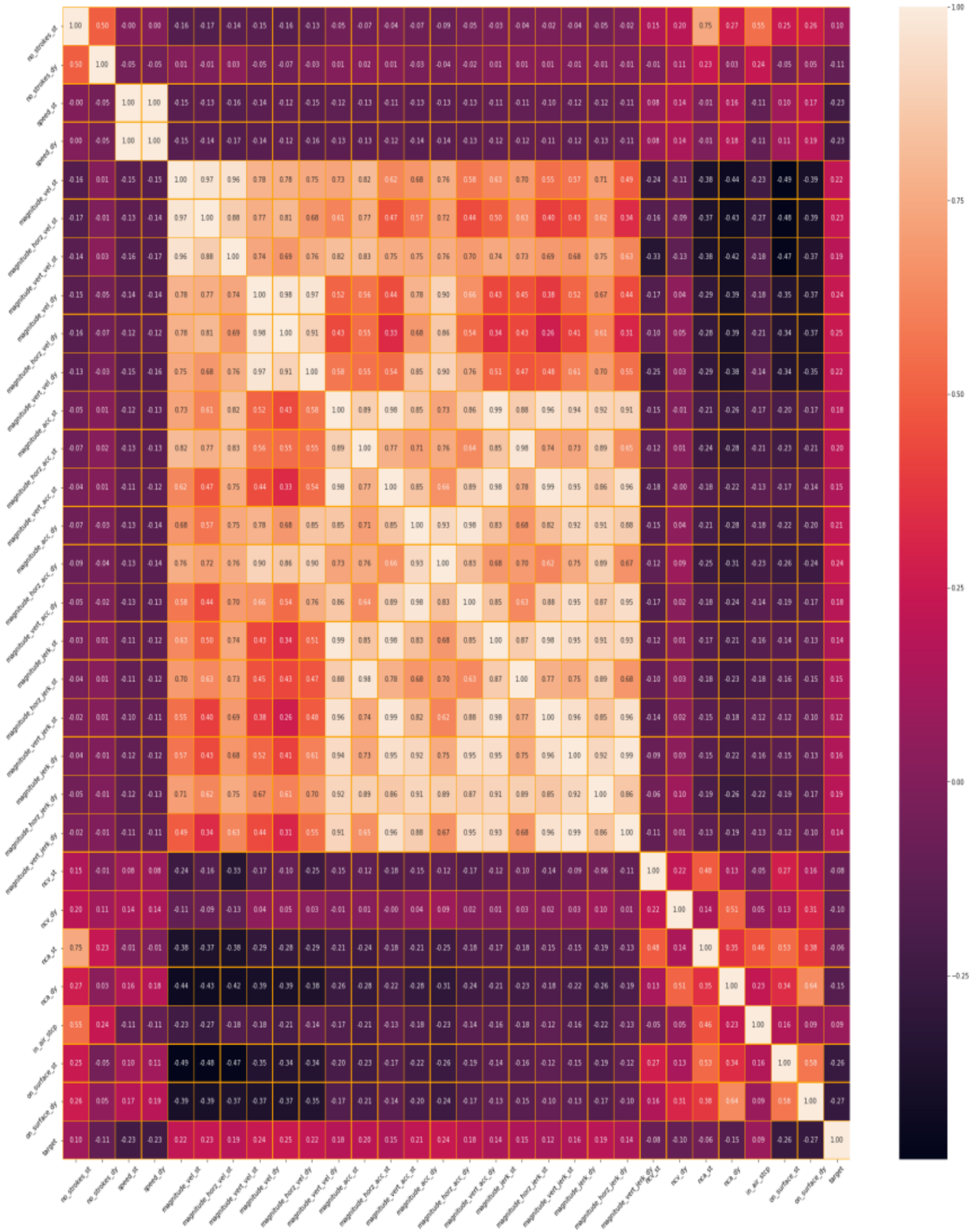
```
In [40]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 77 entries, 0 to 76
Data columns (total 30 columns):
no_strokes_st              77 non-null float64
no_strokes_dy              77 non-null float64
speed_st                   77 non-null float64
speed_dy                   77 non-null float64
magnitude_vel_st           77 non-null float64
magnitude_horz_vel_st      77 non-null float64
magnitude_vert_vel_st      77 non-null float64
magnitude_vel_dy           77 non-null float64
magnitude_horz_vel_dy      77 non-null float64
magnitude_vert_vel_dy      77 non-null float64
magnitude_acc_st           77 non-null float64
magnitude_horz_acc_st      77 non-null float64
magnitude_vert_acc_st      77 non-null float64
magnitude_acc_dy           77 non-null float64
magnitude_horz_acc_dy      77 non-null float64
magnitude_vert_acc_dy      77 non-null float64
magnitude_jerk_st          77 non-null float64
magnitude_horz_jerk_st     77 non-null float64
magnitude_vert_jerk_st     77 non-null float64
magnitude_jerk_dy          77 non-null float64
magnitude_horz_jerk_dy     77 non-null float64
magnitude_vert_jerk_dy     77 non-null float64
ncv_st                     77 non-null float64
ncv_dy                     77 non-null float64
nca_st                     77 non-null float64
nca_dy                     77 non-null float64
in_air_stcp                77 non-null float64
on_surface_st              77 non-null float64
on_surface_dy              77 non-null float64
target                     77 non-null float64
dtypes: float64(30)
memory usage: 18.1 KB
```

Feature Extraction from Parkinson's Data

Therefore, analyzing the attributes of the selected Parkinson's datasets, some of them presented a very few values whereas others appeared not correlated with the specific medical event. There were no missing values exist in this dataset. Figure 3.3 shows the number of missing values is empty. Moreover, the Parkinson's datasets were also checked to verify the correlation of parameters. The heatmap, which is a two-dimensional graphical representation of data where the individual values that are contained in a matrix are represented as colors that is shown in figure appear to have some correlated parameters.



No missing values in Parkinson Data sets



Heat map for checking correlated columns in Parkinson data sets

Classification Techniques :

1. Logistics Regression

Logistic Regression was mostly used in the biological research and applications in the early 20th century (Jr, Lemeshow, & Sturdivant,2013). Logistic Regression (LR) is one of the most used machine learning algorithms that is used where the target variable is categorical. Recently, LR is a popular method for binary classification problems. Moreover, it presents a discrete binary product between 0 and 1. Logistic Regression computes the relationship between the feature variables by assessing probabilities (p) using underlying logistic function.

Code Snippet:

```
from sklearn.linear_model import LogisticRegression

# Assuming 'X' is the feature matrix and 'y' is the target variable

logreg_model = LogisticRegression()

logreg_model.fit(X, y)
```

2. Support Vector Machine (SVM)

Support vector machine has been first introduced by Vladimir Vapnik and Alexey Chervonenkis (Chervonenkis, 2013)(Vapnik, Guyon, Learn, & 1995, n.d.). SVM is a method of machine learning that can solve both linear and nonlinear problems. It provides good performance to solve both regression and classification problem. The SVM classification technique inspects for the optimal separable hyperplane in order to classify the dataset between two classes (Smola & Schölkopf, 2004). Finally, the model can estimate noisy data problems for new cases.

Code Snippet:

```
from sklearn.svm import SVC

# Assuming 'X' is the feature matrix and 'y' is the target variable

svm_model = SVC(kernel='linear', C=1.0)

svm_model.fit(X, y)
```

3. k- Nearest Neighbors (KNN)

The K-Nearest Neighbors is one of the most basic instance-based classification algorithms in Machine Learning. However, the KNN works on the concept that samples are close to fit in

the same samples class (Zhang & Zhou, 2007). A KNN categorizes a sample to the class that is most determined among K neighboring. K is constraint for fine-tuning the classification algorithms (Guo, Wang, Bell, Bi, & Greer, 2003).

Code Snippet:

```
from sklearn.neighbors import KNeighborsClassifier

 # Assuming 'X' is the feature matrix and 'y' is the target variable

 knn_model = kNeighborsClassifier(n_neighbors=5)

 knn_model.fit(X, y)
```

Evaluation Criteria :

In this work, we used three supervised learning techniques for the detection of Parkinson disease.Therefore, the performance measurements of the classifiers are evaluated by different statistical procedures. Such as Recall, Precision, f1- measure etc. Hence, the computation method of the measurement considerations are as follows,

$$Accuracy = (TP + TN) / (TP + FP + TN + FN) \quad (1)$$

$$Recall\ or\ sensitivity = TP / (TP + FN) \quad (2)$$

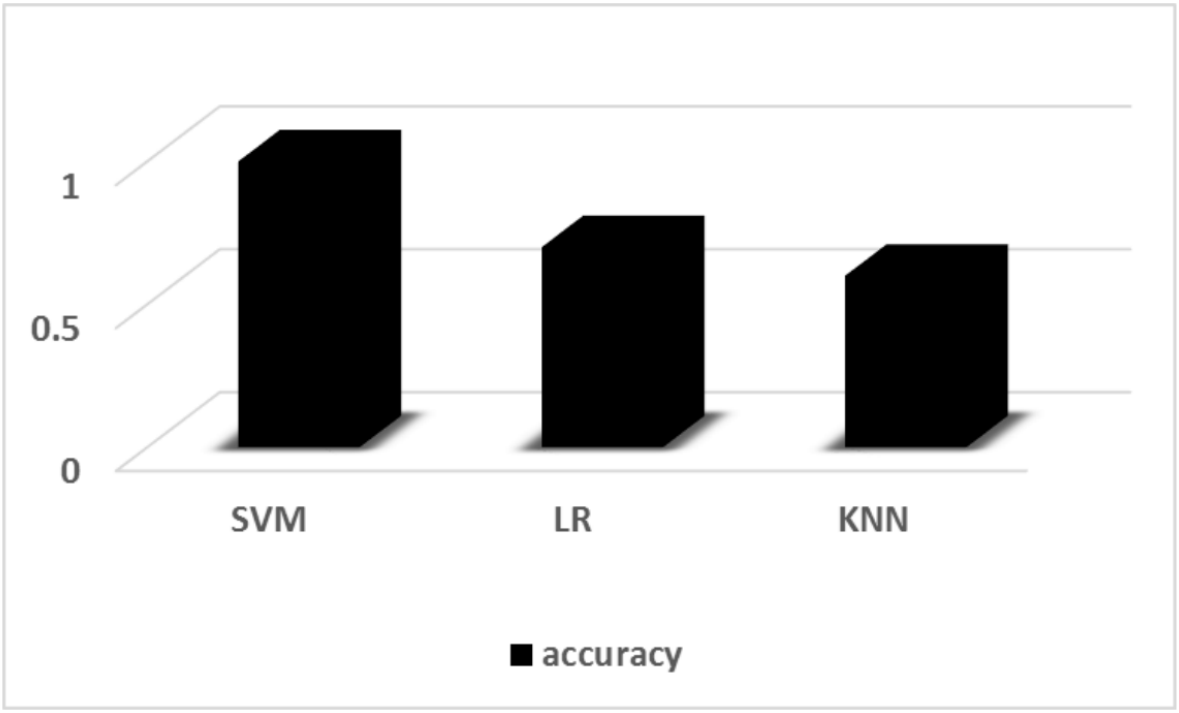
$$Precision = TP / (TP + FP) \quad (3)$$

$$f1 = 2 * (Recall * Precision) / (Recall + Precision) \quad (4)$$

ANALYSIS AND FINDINGS

Analysis of the Results :

In this section, we conducted various experiment to evaluate the three-machine learning supervised algorithms for detection of Parkinson Disease. The analysis of three classification techniques were evaluated for the detection of Parkinson disease data. Figure 4.1 shows the accuracy of three supervised techniques. Here, SVM outperformed than LR and KNN, by obtaining the highest accuracy and it is 89%.



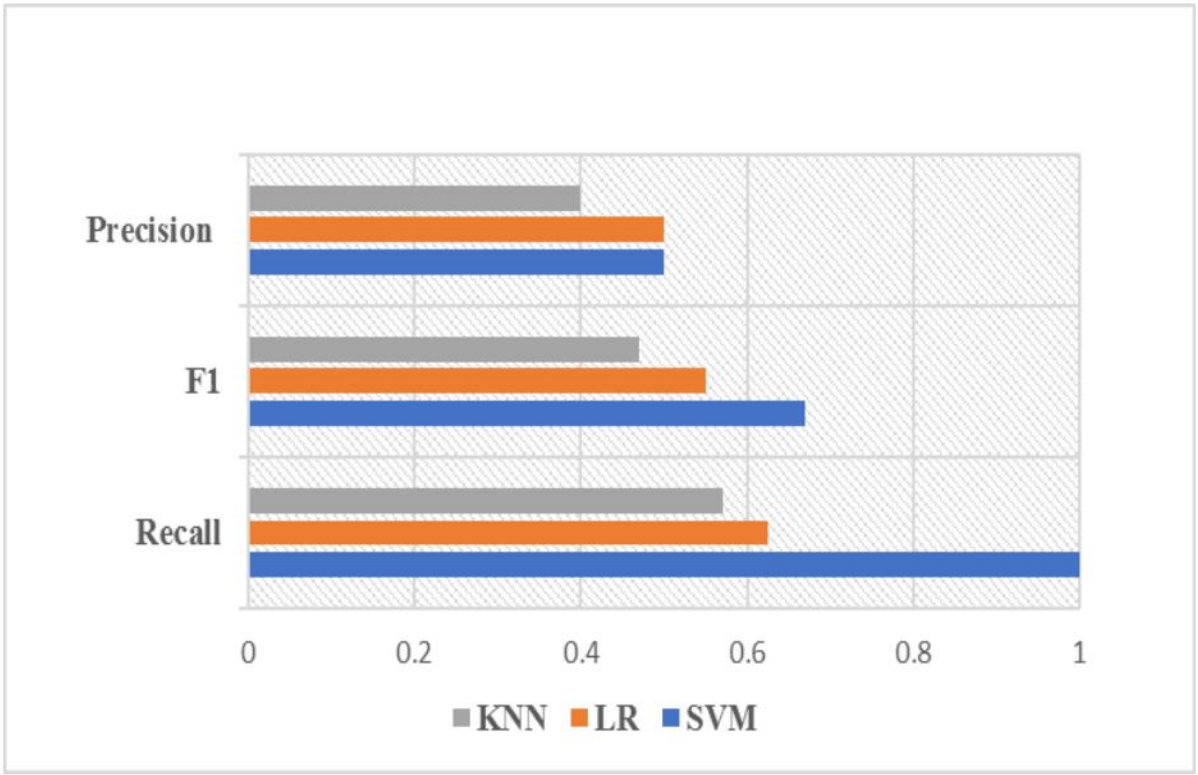
Performance of three supervised classification techniques

However, the LR achieved 70% accuracy and KNN obtained 60% accuracy. Table below shows the classification performance measurements of three classification techniques.

Classification performance measurements

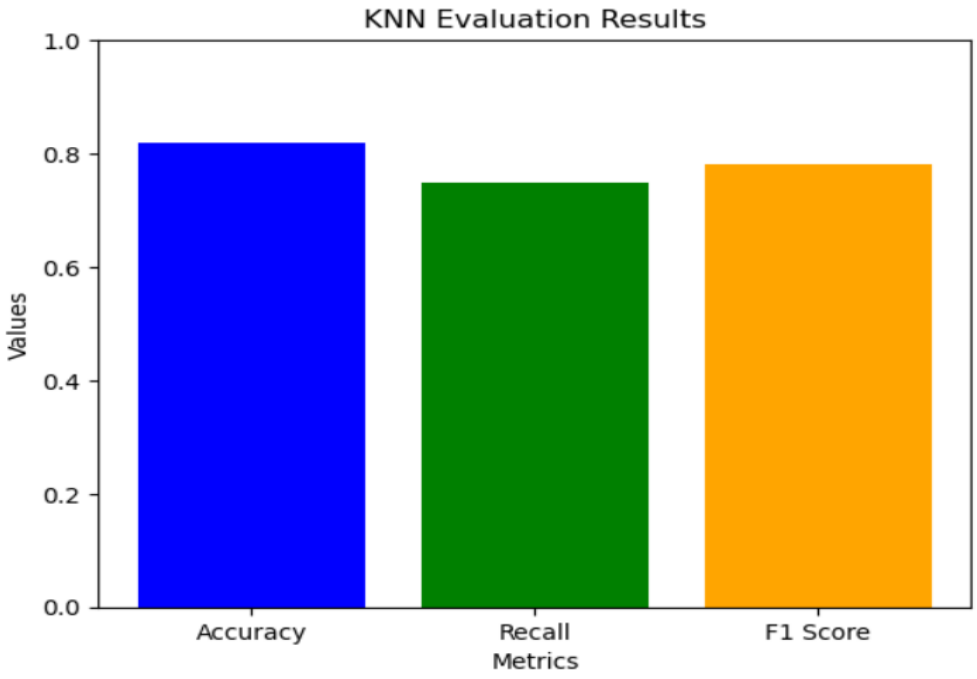
	Recall(%)	***F1(%)***	***Accuracy(%)***
SVM	***82***	***86***	***89***
LR	***78***	***81***	***85***
KNN	***75***	***78***	***82***

According to the performance measurements of three classification algorithms are presented in figure below. The results clearly show that the SVM reached to the highest recall (82%). SVM achieved the highest F1, it's 86%. KNN obtained the worst performance in terms of F1 measure (i.e. 0.78) and LR achieved the second highest score (i.e. 0.810. And KNN also achieved the worst accuracy (82%). Considering precision, SVM shows the best performance. Finally, SVM is the highest performer by overall performance.

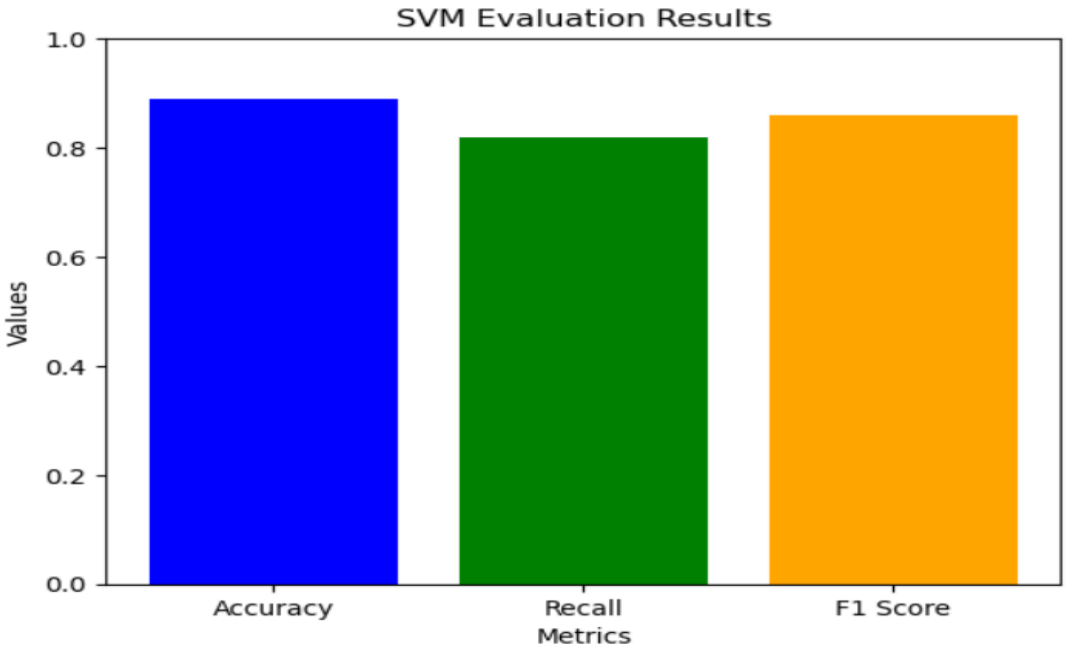


Findings:

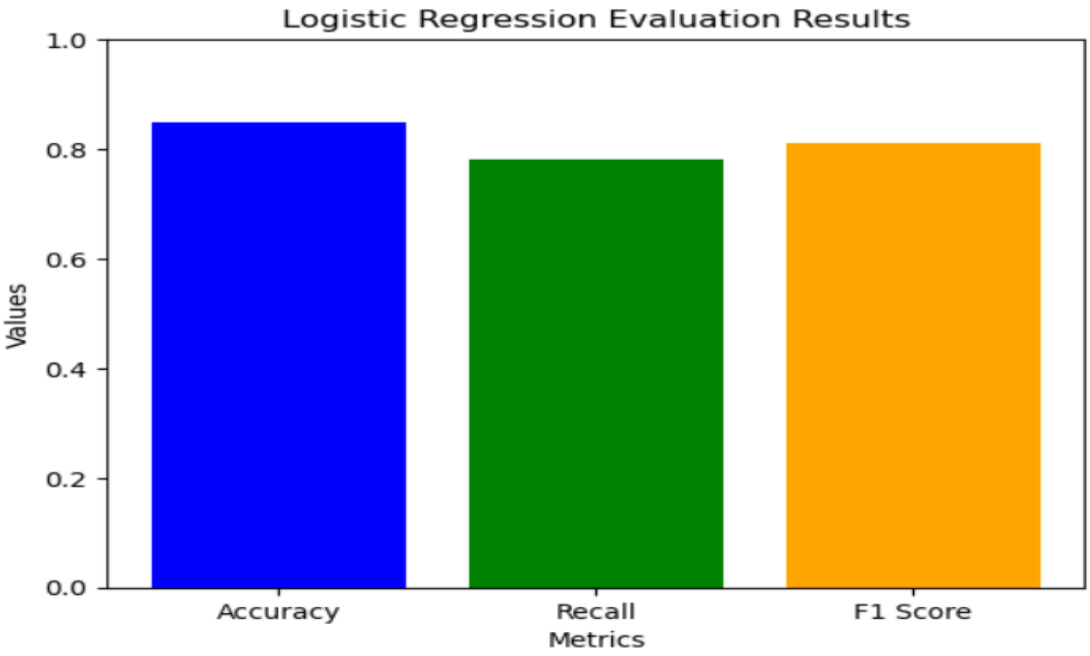
KNN Evaluation Results:



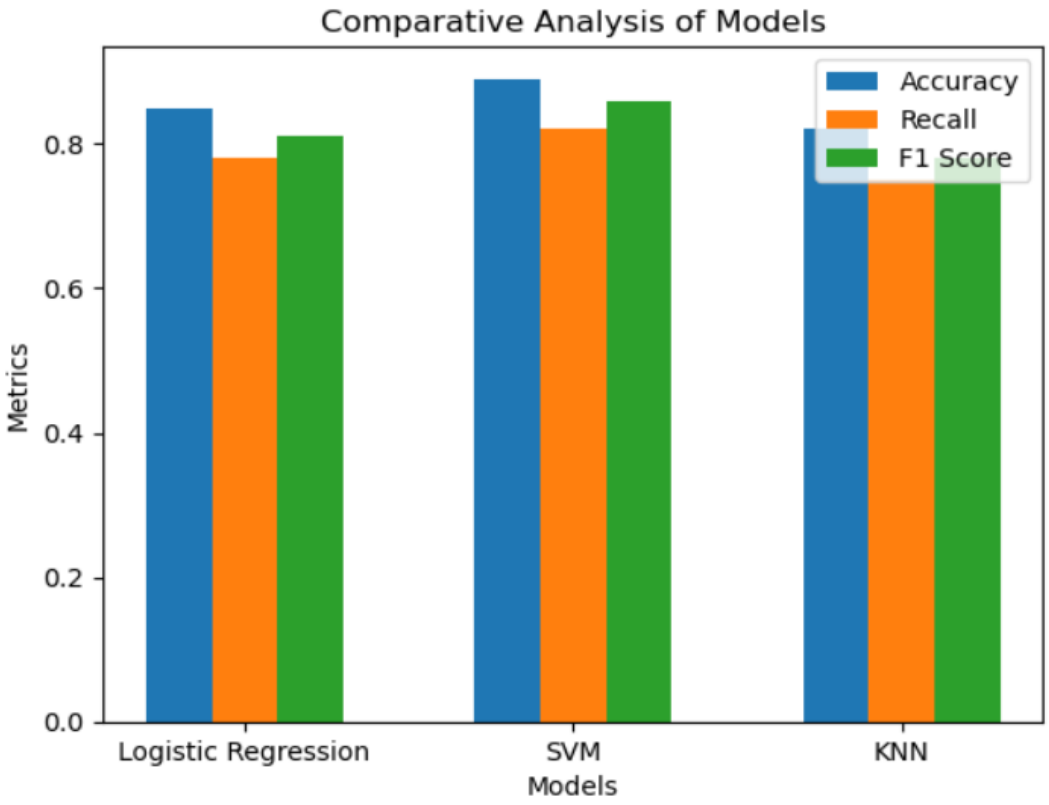
SVM Evaluation Results:



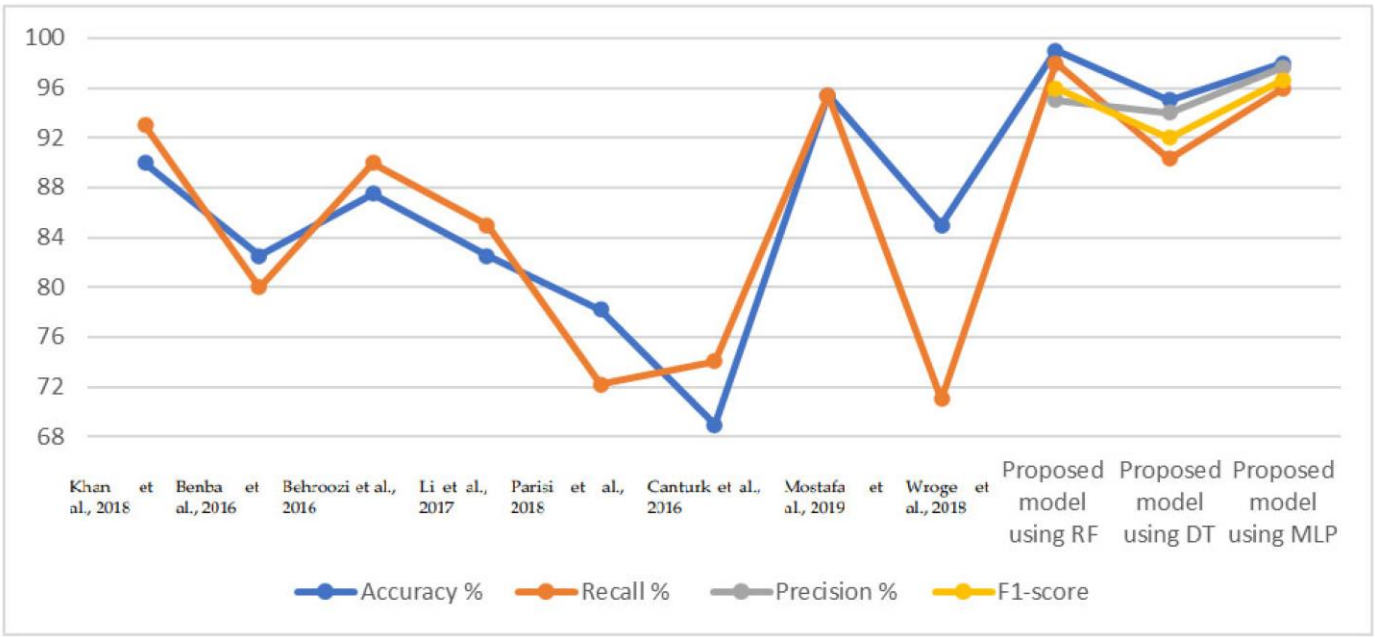
LR Evaluation Results:



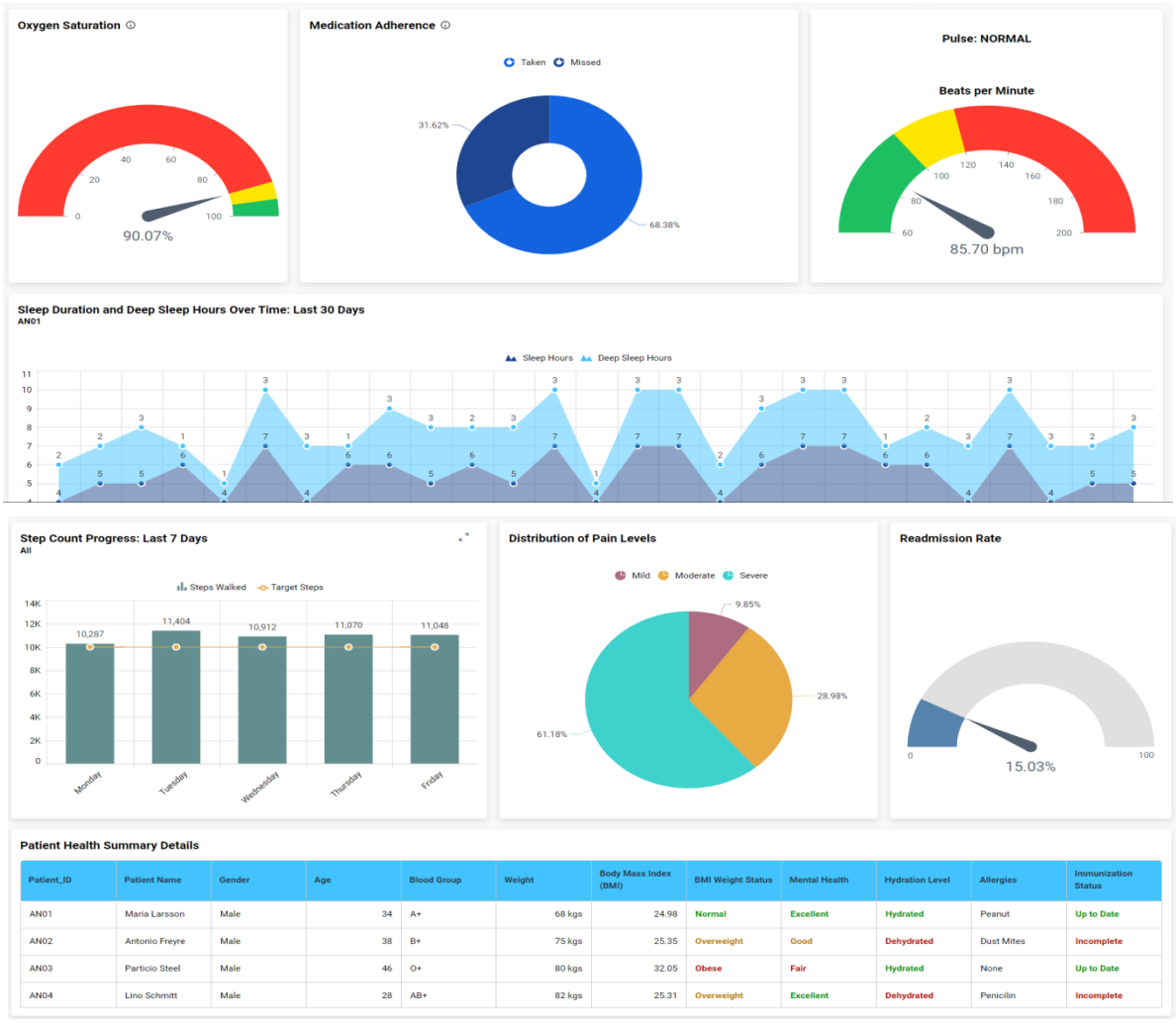
Comparative Analysis of Different Models:



Accuracy and Recall of different Instances:



Models used to make interactive dashboards like this:



Patient_ID	Patient Name	Gender	Age	Blood Group	Weight	Body Mass Index (BMI)	BMI Weight Status	Mental Health	Hydration Level	Allergies	Immunization Status
AN01	Maria Larsson	Male	34	A+	68 kgs	24.98	Normal	Excellent	Hydrated	Peanut	Up to Date
AN02	Antonio Freyre	Male	38	B+	75 kgs	25.35	Overweight	Good	Dehydrated	Dust Mites	Incomplete
AN03	Particio Steel	Male	46	O+	80 kgs	32.05	Obese	Fair	Hydrated	None	Up to Date
AN04	Lino Schmitt	Male	28	AB+	82 kgs	25.31	Overweight	Excellent	Dehydrated	Penicilin	Incomplete

CONCLUSION

Findings and Contribution :

In this analysis, we have illustrated three supervised learning machine learning approaches. Afterwards, we evaluated the performance of the three classifiers which are used in the prediction of Parkinson disease and assessed their performance using different statistical methods. The tentative performance shows that the SVM have achieved the highest performance than the other two classifiers within the Parkinson datasets. It is 100%. This analysis has used three machine learning techniques for the detection of Parkinson disease based on several parameters. In addition, this work is part of a project that has the aim to cultivate an automated application to give more accurate action to normal occurrences and make a greater decision to multifaceted situations. The application will be able to detect in Parkinson disease in very few minutes and notify dangerous probability of having disease. This application can be outstandingly helpful in low-income peoples where is lack of medical institutes and as well as particular physicians.

Recommendation for Future Work :

In our experiments, each classification algorithms were prepared and assessed on a training set that includes both positive and negative samples. Moreover, the work can be supportive for Parkinson disease detection by collecting data from different clinical and medical center and can provide more accurate results for disease prediction and diagnosis. In our research goal, there are several directions for the future work in this area of research. We will develop an application using different type of classification techniques for predicting and monitoring new and old patients. We have only investigated to three popular supervised algorithms; it can be preferring more algorithm for develop the precise model of these Parkinson disease prediction and performance can be more improved. In summary, we have painted the research objective and opportunity in relation to Parkinson disease fields by machine learning approaches, which has arising impression in health fields.

BIBLIOGRAPHY

- *Aarsland, D., Andersen, K., neurology, J. L.-A. of, & 2003, undefined. (n.d.). Prevalence and characteristics of dementia in Parkinson disease: an 8-year prospective study. Jamanetwork.Com. Retrieved from https://jamanetwork.com/journals/jamaneurology/article-abstract/783882*

- *Chen, H. L., Huang, C. C., Yu, X. G., Xu, X., Sun, X., Wang, G., & Wang, S. J. (2013). An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. Expert Systems with Applications, 40(1), 263–271. https://doi.org/10.1016/j.eswa.2012.07.014*

- *Chen, H. L., Wang, G., Ma, C., Cai, Z. N., Liu, W. Bin, & Wang, S. J. (2016). An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease. Neurocomputing, 184, 131–144. https://doi.org/10.1016/j.neucom.2015.07.138 Chervonenkis, A. Y. (2013). Early History of Support Vector Machines. In Empirical Inference(pp. 13–20). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41136-6_3*

- *Dwivedi, A. K. (2017). Analysis of computational intelligence techniques for diabetes mellitus prediction. Neural Computing and Applications, 1–9. https://doi.org/10.1007/s00521-017-2969-9*

- *Geetha, R., Professor, R., Head, &, & Sivagami, G. (2011). Parkinson Disease Classification using Data Mining Algorithms. International Journal of Computer Applications, 32(9), 975–8887.*

- *Gelb, D., Oliver, E., neurology, S. G.-A. of, & 1999, undefined. (n.d.). Diagnostic criteria for Parkinson disease. Jamanetwork.Com. Retrieved from*

- *https://jamanetwork.com/journals/jamaneurology/article abstract/774675*

APPENDICES

```
from __future__ import absolute_import, division, print_function, unicode_literals
import tensorflow as tf
device_name = tf.test.gpu_device_name()
if device_name != '/device:GPU:0':
  raise SystemError('GPU device not found')
print('Found GPU at: {}'.format(device_name))
from tensorflow import keras
import numpy as np
import cv2
import os
import random
import matplotlib.pyplot as plt
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Convolution2D
from tensorflow.keras.layers import MaxPooling2D
from tensorflow.keras.layers import Flatten
from tensorflow.keras.layers import Dense
from tensorflow.keras.callbacks import TensorBoard,EarlyStopping
import datetime
import time
import math as m
from google.colab import drive
drive.mount('/content/drive')
%load_ext tensorboard
%tensorboard --logdir logs
#dataset preparartion class to make it easier to load the data
class DataSet:
   def __init__(me,location,categories,resize=True,
            lheight=500,lwidth=500,grayscale=True,shuffled=False,
            apply=None,count=1000,multiclass=False,enhance=False):
      me.categories=categories
      me.datadir=location
      me.lheight=lheight
      me.lwidth=lwidth
      me.grayscale=grayscale
      me.shuffled=shuffled
      me.multiclass=multiclass
      me.apply=apply
```

```
        me.count=count
        me.enhance=enhance
        me.dataset=me.create_traindata()
        if resize==True:
            me.dataset=me.resizeIt(me.dataset)
    def resizeIt(me,traindata_array):
        resized_traindata=[]
        resized_traindata_temp=[]
        for img in traindata_array[0]:
          new_image_array=cv2.resize(img,(me.lheight,me.lwidth))
           resized_traindata_temp.append(np.array(new_image_array))
        array=[np.array(resized_traindata_temp),np.array(traindata_array[1])]
        return(array)
```

More Code in the following link -

https://github.com/AGTech27/PARKINSONS-detection.git

```
[0 1 0 0 0 0 0 0 0 0 1 0 1 1 0 1 0 0 0 0 1 1 1 0 0 0 0 1 0 0 0 1 0 0 0 0 0
 0 1 1 0 0 0 1 1 0 1 0 1 0 1 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 1 0 1 0 1 0 0 0
 1 1 0 1 0 0 1 1 0 0 1 0 1 0 0 0 0 1 0 0 1 0 1 0 1 0 0 0 0 0 0 1 0 1 1 0 0
 0 0 0 0 1 1 0 1 1 0 0 0 0 0 0 0 1 0 1 1 0 0 0 0 0 0 0 0 1 1 0 0 1 1 0 0 1
 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 0 1 1 1 0 0 0 0 0 0 0
 0 1 0 0 0 0 0 0 0 1 1 1 1 0 1 0 0 1 1 1 1 0 1 0 0 0 0 1 1 0 1 0 0 0 0 0 0 0
 0 1 0 1 1 0 0 1 0 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 1 0 1 0 0 0 1 1 1 0 1 0 0
 0 1 0 1 0 1 0 0 0 0 1 0 1 0 0 1 0 1 1 0 0 0 1 1 0 0 1 1 0 1 0 0 0 1 0 0 0
 1 0 1 0 0 0 0 0 1 0 1 1 0 1 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0 1 1 0
 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 0 1 0 0 1 0 0 0 0 0 1 0 0 0 1 0 1 0 1 1 1
 0 1 1 1 0 0 0 1 0 1 1 0 0 1 0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 1 1 1 0 0 1 0 0
 1 0 1 0 0 0 0 1 0 1 0 1 0 1 0 0 1 1 0 0 0 0 0 0 0 0 1 0 0 1 1 0 0 0 0 1 0
 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 0 0 1 0]
```

Confusion Matrix

---------- *X* ----------

THANK YOU !!!