iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** III **Month of publication:** March 2019

DOI: http://doi.org/10.22214/ijraset.2019.3284

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

Prediction of Parkinson Disease by Best Accuracy using Supervised Classification Machine Learning Approach

Gopinath. K[1], B. Ajay Kishore[2], K. Arul Selvan[3], D. Mohan Raj[4]
[1, 2, 3, 4]*Dept. Of Computer Science (Anna University) Panimalar Institute of Technology, Chennai, India*

Abstract: Parkinson's disease is the most prevalent neurodegenerative disorder affecting more than 10 million people worldwide. There is no single test which can be administered for diagnosing Parkinson's disease.
Because of these difficulties, to investigate a machine learning approach to accurately diagnose Parkinson's, using a given dataset. To prevent this problem in medical sectors have to predict the disease affected or not by finding accuracy calculation using machine learning techniques.
The aim is to investigate machine learning based techniques for Parkinson disease by prediction results in best accuracy with finding classification report.
The analysis of dataset by supervised machine learning technique(SMLT) to capture several information's like, variable identification, univariate analysis, bivariate and multivariate analysis, missing value treatments and analyze the data validation, data cleaning/preparing and data visualization will be done on the entire given dataset
Keywords: Dataset, Machine learning-Classification method, python, Prediction of Accuracy result.

I. INTRODUCTION

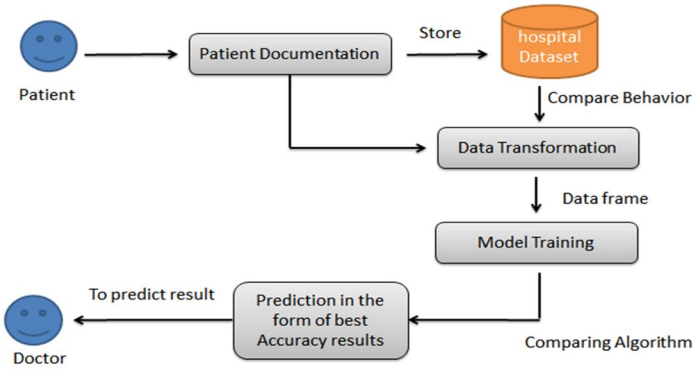
Parkinson's disease [1] (PD) is a neurodegenerative disorder, and millions of people suffer from it all over the world. The Occurrences of PD increases with the age growth, about 6.32 million people are suffering from this disease. Notably, in a developed country, the number of patients with PD has increased significantly in recent years. However, there are no methods which can measure the PD progression efficiently and accurately in its early stages.
The last known drug for Parkinson's disease was found in 1967.Machine learning is to predict the future from past data. Machine learning (ML)[2] is a type 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 and the basics of Machine Learning, implementation of a simple machine learning algorithm using python. Movement of training and prediction involves use of specialized algorithms. It feed the training data to an algorithm, and the algorithm uses this training data to give predictions on a new test data.
Machine learning can be categorize in to three categories. There are supervised learning, unsupervised learning and reinforcement learning. Supervised learning program is both given the input data and the corresponding labeling to learn data has to be labeled by a human being beforehand.
Unsupervised learning is no labels. It provided to the learning algorithm. This algorithm has to figure out the clustering of the input data. Finally, Reinforcement learning dynamically interacts with its environment and it receives positive or negative feedback to improve its performance.

II. DESIGN

A. General

Design is meaningful engineering representation of something that is to be built. Software design is a process design is the perfect way to accurately translate requirements in to a finished software product. Design creates a representation or model, provides detail about software data structure, architecture, interfaces and components that are necessary to implement a system.There are two datasets are discussed here speech [3] and tremor.

B. System Architecture



Design Architecture

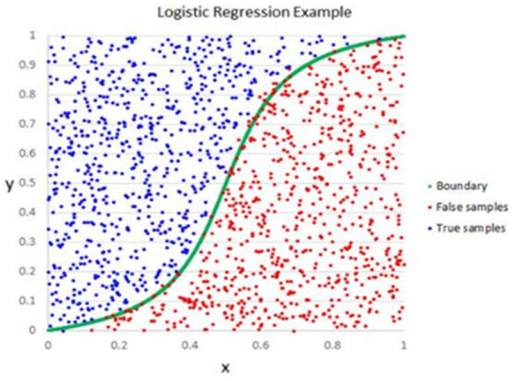
III. ALGORITHM

A. Algorithm Explanation

In machine learning, classification is a supervised learning access in which the computer programs Studies from the input given to the system and then uses this learning to analyze new Knowledge, speech detection, handwriting detection, bio metric identification, document distribution etc. In Supervised Learning, algorithms learn from labeled data. After understanding the data, the algorithm determines which classification should be given to new data based on pattern and associating the patterns to the unclassified new data.

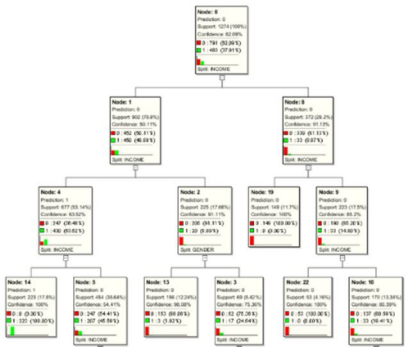
B. Logistic Regression

It is a statistical Procedure for analysing a data set in which there are one or more independent variables that verify an outcome. The outcome is measured with a divided variable (in that there are solely 2 potential outcomes). The objective of this logistic regression [4] is to find the best convenient model to describe the relationship between the divided characteristic of interest (dependent variable = response or outcome variable) and a collection of independent (predictor or explanatory) variables. Logistic regression is a Machine Learning classification algorithm that is used to coclude the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as one (yes, success, etc.) or zero (no, failure, etc.).



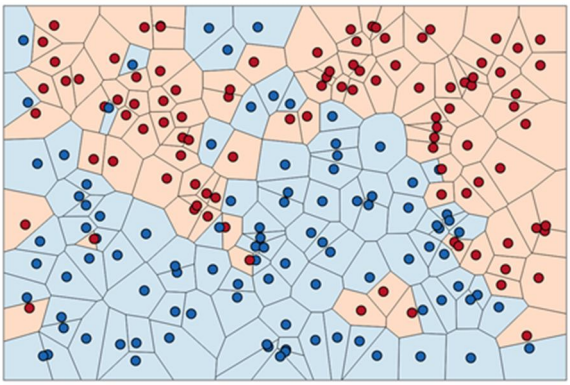
C. Decision Tree

It is one of the most powerful and popular algorithm. Decision-tree algorithm[5] falls under the category of supervised learning algorithms. It works for each continuous in addition as categorical output variables. Decision tree builds classification or regression models with in the type of tree structure. It breaks down a data set into smaller and smaller subsets while at the constant time an associated decision tree is incrementally Formed. A call node has two or more than two branches and a leaf node symbolize a classification or decision. The upmost decision node in a tree which corresponds to the best predictor refered to as root node. Decision trees handles categorical as well as numerical data. Decision tree builds classification or regression models in the form of a tree structure. The rules are learned one by one using the training data one at a time. Each time a rule is learned, the covered rules are removed by tuples.



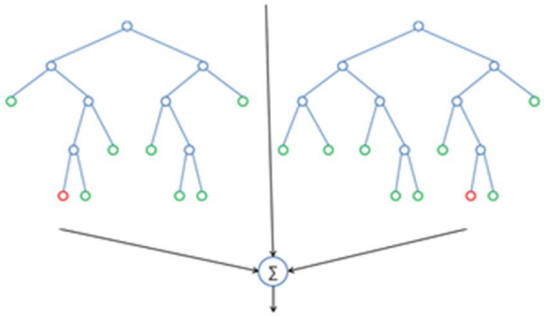
D. K-Nearest Neighbor (KNN)

K-Nearest Neighbor [6] is a supervised machine learning algorithm which stores all instances correspond to training data points in n-dimensional space. When an unknown disorganize info or data is received, it analyzes the closest k number of instances saved (nearest neighbors) and returns the most common class as the prediction and for actual-value data it returns the mean of k nearest neighbors. In the distance-weighted nearest neighbor algorithm, it weights the contribution of each of the k neighbors according to their distance using the following query giving greater weight to the closest neighbors.



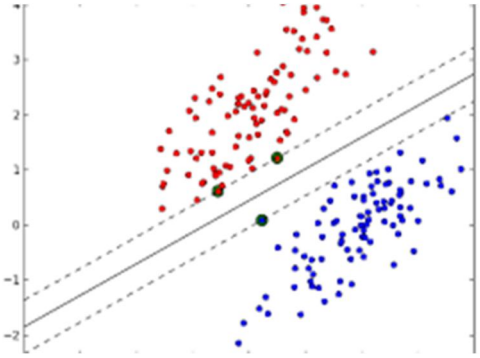
E. Random Forest

Random forests or random decision forests [7] are an altogether learning method for distribution, regression and alternative tasks, that operate by constructing a large number of decision trees at training time and outputting the class that is the mode of the classes (distribution) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set. Random forest is a supervised machine learning algorithm based on altogether learning. Ensemble learning also known as altogether learning is a type of learning where you join different types of algorithms or same algorithm multiple times to form a more powerful prediction model. The random forest algorithm combines numerous algorithm of the same type, hence the name "Random Forest". The random forest algorithm can be used for regression as well as distribution tasks.



F. Support Vector Machines

A classifier that categorizes the data set by setting an optimal hyper plane between data. I chose this classifier as it is incredibly versatile in the number of different kernelling functions that can be applied and this model can yield a high predictability rate. Support Vector Machines [8] are perhaps one of the most popular and talked about machine learning algorithms. They were extraordinarily common round the time they were developed within the Nineties and still be the go-to technique for a high-performing algorithmic rule with very little standardization.



IV. COMPARISONS OF RESULT

A. Speech Result

Algorithm	Precision	Recall	F1-Score	Accuracy (100%)
LR	0.86	0.85	0.85	84.74
DT	0.84	0.83	0.83	83.05
SVC	0.56	0.75	0.64	74.57
RF	0.93	0.93	0.93	93.22
KNN	0.95	0.95	0.95	94.91

B. Tremor Result

Algorithm	Precision	Recall	F1-Score	Accuracy (100%)
LR	0.75	0.79	0.75	79.16
DT	0.92	0.92	0.92	91.66
SVC	0.63	0.79	0.70	79.16
RF	0.96	0.96	0.96	95.83
KNN	0.78	0.79	0.78	79.16

V. FUTURE WORKS

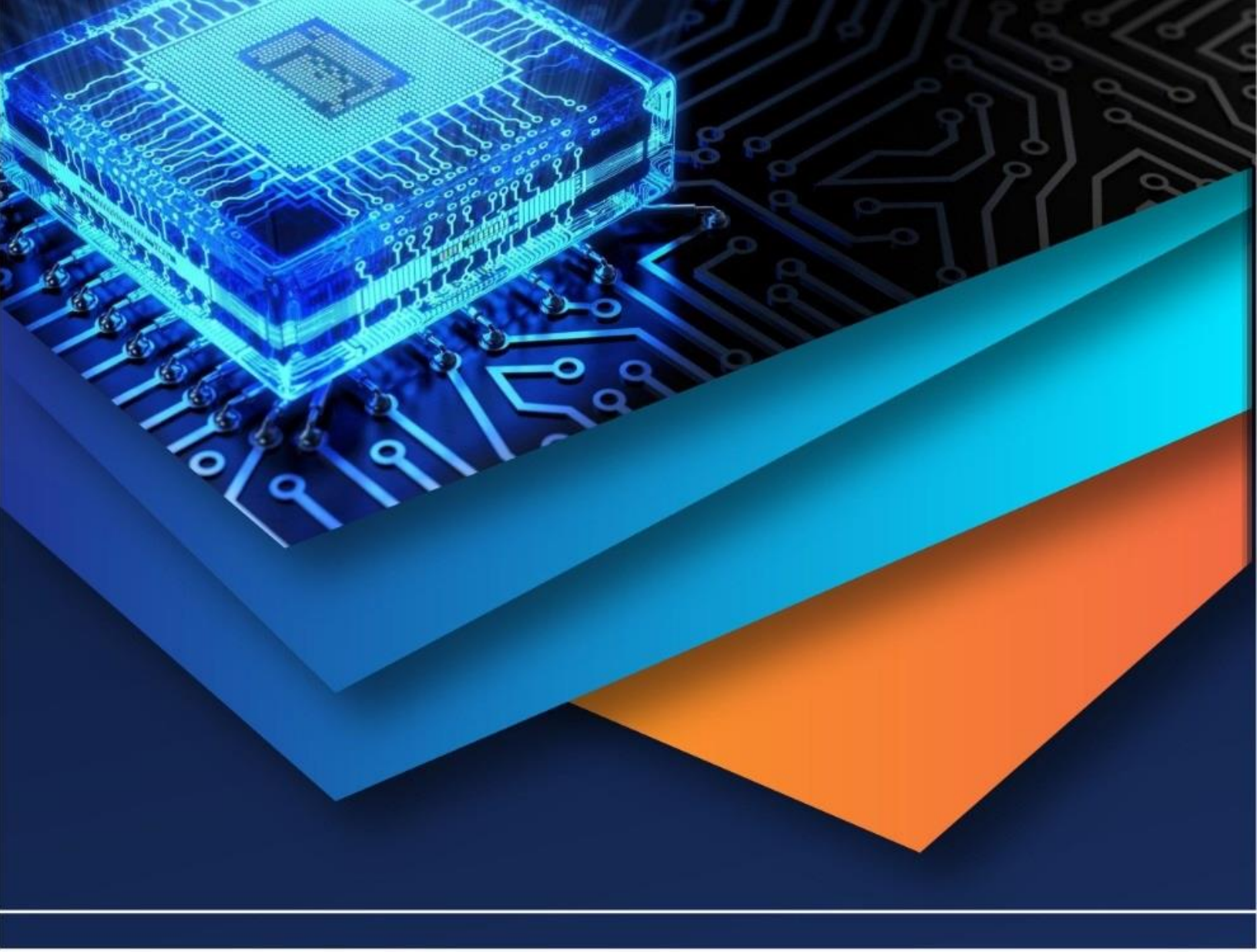
A. Hospitals want to automate the detecting the disease persons from eligibility process (real time) based on the account detail.
B. To automate this process by show the prediction result in web application or desktop application.
C. To optimize the work to implement in Artificial Intelligence environment.

VI. CONCLUSION

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on public test set is higher accuracy score of Speech for KNN and Tremor for Random Forest algorithm. This brings some of the following insights about diagnose the Parkinson disease. Early diagnosis of Parkinson's is most important for the patient to reduce its impact. To presented a prediction model with the aid of artificial intelligence to improve over human accuracy and provide with the scope of early detection. With our proposed prediction model we aim to make it easier for doctors to do precise diagnosis and prediction of PD, both of which have human limitations due to the method of detection of PD that is used now. It can be inferred from this model that, area analysis and use of machine learning technique is useful in developing prediction models that can help a doctor reduce the long process of diagnosis and eradicate any human error. To separate the work of detection and prediction methods to detect and measure the area of brain that is affected due to PD and use that data in machine learning to create the prediction model with accuracy is higher comparing other models.

REFERENCES

[1] J . Parkinson, Essay on the Shaking Palsy. London: Sherwood, 1817.
[2] Aravind Kumar Tiwari, Machine Learning Based Approaches for prediction of Parkinsons Disease, June 2016
[3] Meysam Asgari and Izhak Shafran, Predicting Severity Of ParkinsonsDisease from Speech,September 2010
[4] S.Sharanyaa, Classification Of Parkinson's Disease Using Logistic Regression
[5] Themis P.Exarchos , Using partial decision trees to predict Parkinson's symptoms: A new approach for diagnosis and therapy in patients suffering from Parkinson's disease
[6] Murat Gok, An ensemble of k-nearest neighbours algorithm for detection of Parkinson's disease.
[7] Alexander Zlotnik andJuan M. Montero, Random Forest-Based Prediction of Parkinson's Disease Progression UsingAcoustic, ASR and Intelligibility Features.
[8] Surangsrirat D and Thanawattano C, Support vector machine classification of Parkinson's disease and essential tremor subjects based on temporal fluctuati



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY