PARKINSON'S DISEASE DETECTION USING MACHINE LEARNING

RESEARCH PAPER

LOVELY PROFESSIONAL UNIVERSITY

PHAGWARA, PUNJAB

NAME:HEMANT VG

SUBJECT:INT 247

REGNO:11910284

SECTION:KM055

ROLL NO:A-16

ABSTRACT:

The Parkinson's disease is a neurological disorder. It Predominantly produces neurons which affect the specific area a brain. And the symptoms of Parkinson's disease will slowly develop over years and this disease will make our brain dead. but the symptoms are different from other people because diversity. The symptoms are Tremor, slowness of movements, shivering, Bradykinesia, Limb rigidity, Vocal Impairment, and balance problems. There is no permanent cure for this disease. Some affected people can feel better by taking medicines. And for some patients will undergo surgery to improve from the common symptoms.

The Parkinson's Disease is an most common disease in the world. There are two types of disease Communicable and non-communicable diseases. Parkinson disease is a communicable disease which can be spread from person to person.
And the symptoms will slowly increases over years. And finally it may lead to dead. It is not cure able still now. To overcome from the symptoms surgery has to be taken.

In this project I have used the several machine learning models to predict the different accuracy i.e. Support vector machines (SVM), Random Forest classifier, K-nearest neighbors and XGBoost. To predict the accuracy of the models.

Keywords: Parkinson's disease, SVM, KNN, Random Forest, XGBoost

INTRODUCTION:

Parkinson's Disease is a brain neurological disorder. This leads to an reduction called dopamine in the brain. And many different symptoms are associated with Parkinson's disease and the most common symptoms that the affected patients will have slowness in body movements, muscle hardness. That the Parkinson's disease cannot be cured at the advanced stage as this disease will slowly increase day by day. But with the help of medications and surgery may be used to control the symptoms.

Parkinson's disease occurs when neurons in an area of the brain called the substantia nigra become impaired or die. These cells normally produce dopamine, a chemical called neurotransmitter that helps the cells of the brain

communicate, transmits signals and messages between areas in the brain. When these nerve cells become impaired or die, they produce less dopamine. Dopamine is especially important for the operation of another area of the brain called the basal ganglia. This area of the brain is responsible for organizing the brain's commands for body movement. The loss of dopamine causes the movement symptoms seen in people with Parkinson's disease.

Parkinson's disease is a nervous system disease that affects your ability to control movement. The disease usually starts out slowly and worsens over time. If you have Parkinson's disease, you may shake, have muscle stiffness, and have trouble walking and maintaining your balance and coordination. As the disease worsens, you may have trouble talking, sleeping, have mental and memory problems, experience behavioral changes and have other symptoms. About 50% more men than women get Parkinson's disease. It is most commonly seen in persons 60 years of age and older. However, up to 10% of patients are diagnosed before age 50. About 60,000 new cases of Parkinson's disease are diagnosed in the United States each year. Parkinson's disease occurs when nerve cells (neurons) in an area of the brain called the substantia nigra become impaired or die. These cells normally produce dopamine, a chemical (neurotransmitter) that helps the cells of the brain communicate (transmits signals, "messages," between areas in the brain).

When these nerve cells become impaired or die, they produce less dopamine. Dopamine is especially important for the operation of another area of the brain called the basal ganglia. This area of the brain is responsible for organizing the brain's commands for body movement. The loss of dopamine causes the movement symptoms seen in people with Parkinson's disease. People with Parkinson's disease also lose another neurotransmitter called norepinephrine. This chemical is needed for proper functioning of the sympathetic nervous system. This system controls some of the body's autonomic functions such as digestion, heart rate, blood pressure and breathing. Loss of norepinephrine causes some of the non-movement-related symptoms of Parkinson's disease. Symptoms of PD include Tremor, bradykinesia, rigid muscles, speech changes, etc.

Problem Statement:

Parkinson's disease is a progressive nervous system disorder that affects movement leading to shaking, stiffness, and difficulty with walking, balance, and coordination. Parkinson's symptoms usually being gradually and get worse over time.

The productive screening process does not demand a medical visit and can be more helpful. People who are having Parkinson's Disease show distinctive voice characteristics, therefore voice recordings are considered to be a beneficial tool for the diagnosis. Implementation of machine learning algorithms on the speech dataset for accurate diagnosis of the disease would be a productive screening step before visiting the doctor.

OBJECTIVE:

The main aim of this project is to predict weather the person is affected with the Parkinson's Disease are not using machine learning. In This project the different machine learning models has been used to predict accurately.

The following objective of the model are.

- Developing the model to predict the patient affected with the Parkinson's disease or not.
- Predicting the accuracy of each machine learning models
- Analysing the dataset and splitting the dataset into training and testing for the prediction of the Parkinson's disease
- Different Classification model is used for the Prediction of Parkinson's Disease.

Literature Review:

As nowadays We all know that each person has very less nutrition in our body. Some People doesn't prefer a good food or lack of water content in our body may leads to a disease.

Parkinson's Disease is one of the most common diseases in the world. There are two types of diseases they are communicable and non-communicable diseases.
The Parkinson's Disease is a communicable disease as it spread from people to people. And it affects the main region of the brain and cell does not work properly and it may lead to brain dead also.

The Parkinson's disease symptoms will slowly start and affect our body.

Still now this disease is not cure able at the advanced stage.

As I have said before it will mostly affect the people age above 60.

And the symptoms can be cured with some medications and surgery.

Disease Society stated that one individual in every 500 British people has this disease, and it is expected that this number will increase 3-fold in the next 50 years. Normally this illness becomes worse over time and mostly affects people between 50-70 years old. PD was first described by James Parkinson, a British physician, in 1817 and there is still no treatment for PD (Lones et al., 2014; Pereira et al., 2015).

Shahbakhi et alia (2014) presented a Genetic Algorithm (GA) and SVC algorithm to classify between people having Parkinson's disease and people who aren't suffering (healthy people). Fourteen features of Voice signals are based on pitch, jitter, shimmer, and noise to harmonic ratio, which are vital characteristics of the voice signals. This provides classification accuracy of 94.50, 93.66, and 94.22 per 4, 7, and 9 optimize features respectively.

Yahia A. et alia (2014) presented a comparison between naive Bayes Classifier and KNN algorithm using Parkinson's voice dataset with sound recordings of people having Parkinson's disease and healthy people. The accuracy achieved by the KNN classifier and Naive Bayes algorithm is 80% and 93.3% respectively.

Dataset Information:

This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds one of 195 voice recording from these individuals ("name" column). The main aim of the data is to discriminate healthy people from those with PD, according to "status" column which is set to 0 for healthy and 1 for PD. The data is in ASCII CSV format. The rows of the CSV file contain an instance corresponding to one voice recording. There are around six recordings per patient, the name of the patient is identified in the first column.

MDVP: Fo(Hz) - Average vocal fundamental frequency
MDVP: Fhi(Hz) - Maximum vocal fundamental frequency
MDVP: Flo(Hz) - Minimum vocal fundamental frequency
MDVP: Jitter(%),

MDVP: Jitter(Abs),MDVP:RAP,MDVP:PPQ,Jitter:DDP - Several

measures of variation in fundamental frequency
MDVP:Shimmer,MDVP:Shimmer(dB),Shimmer:APQ3,Shimmer:APQ5,

MDVP: APQ,Shimmer:DDA - Several measures of variation in amplitude
NHR,HNR - Two measures of ratio of noise to tonal components in the voice
status - Health status of the subject (one) - Parkinson's, (zero) - healthy
RPDE,D2 - Two nonlinear dynamical complexity measures
DFA - Signal fractal scaling exponent
spread1,spread2,PPE - Three nonlinear measures of fundamental frequency variation

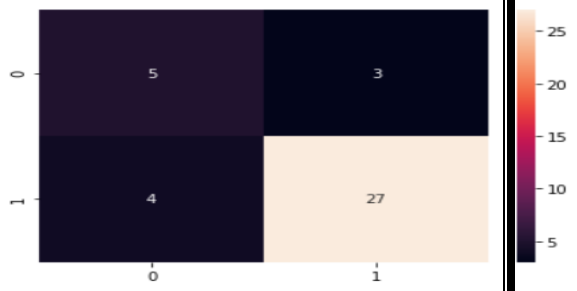
formatted way. So for this, we use data preprocessing task. A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data preprocessing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

In this project I have used Min MaxScaler for feature scaling , it helps in normalization of data. Normalization scales each input variable separately to the range 0-1, which is the range for floating-point values where we have the most precision

Data-Pre-processing:

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a

```
In [24]: from sklearn.metrics import confusion_matrix
         y_pred=model.predict(x_test)
         cm = confusion_matrix(y_test, y_pred)
         sns.heatmap(cm ,annot=True)
Out[24]: <AxesSubplot:>
```



Machine Learning Classifiers:

Random Forest:

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

```
In [23]: from sklearn.ensemble import RandomForestClassifier
         model= RandomForestClassifier(random_state=2)
         model.fit(x_train,y_train)
         y_pred=model.predict(x_test)
         model.score(x_test, y_test)
Out[23]: 0.8205128205128205
```

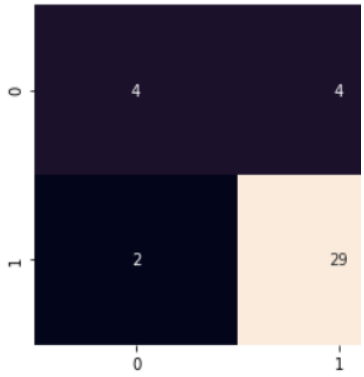
Support Vector machines:

Support Vector Machine (SVM) is a relatively simple Supervised Machine Learning Algorithm used for classification and/or regression. It is more preferred for classification but is sometimes very useful for regression as well. Basically, SVM finds a hyper-plane that creates a boundary between the types of data. In 2-dimensional space, this hyper-plane is nothing but a line. In SVM, we plot each data item in the dataset in an N-dimensional space, where N is the number of features/attributes in the data. Next, find the optimal hyperplane to separate the data. So by this, you must have understood that inherently, SVM can only perform binary classification (i.e., choose between two classes). However, there are various techniques to use for multi-class problems. Support Vector Machine for

Multi-Class Problems To perform SVM on multi-class problems, we can create a binary classifier for each class of the data. The two results of each classifier will be: The data point belongs to that class OR The data point does not belong to that class.

```
In [21]: from sklearn.svm import SVC
         model = SVC(kernel='linear')
         model.fit(x_train, y_train)
         y_pred=model.predict(x_test)
         print(accuracy_score(y_test, y_pred)*100)

         84.61538461538461
```

```
In [22]: from sklearn.metrics import con
         y_pred=model.predict(x_test)
         cm = confusion_matrix(y_test, y
         sns.heatmap(cm ,annot=True)

Out[22]: <AxesSubplot:>
```

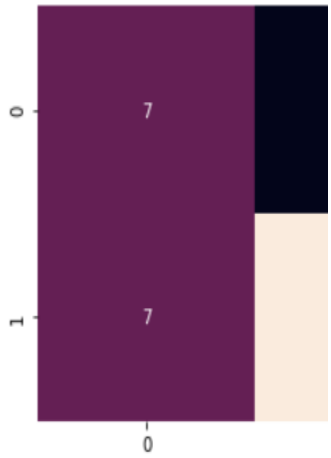


K-Nearest Neighbors is one of the most basic yet essential classification algorithms in Machine Learning. It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining and intrusion detection. It is widely disposable in real-life scenarios since it is nonparametric, meaning, it does not make any underlying assumptions about the distribution of data (as opposed to other algorithms such as GMM, which assume a Gaussian distribution the given data). We are given some prior data (also called train- ing data), which classifies coordinates into groups identified by an attribute

```
In [25]: from sklearn.neighbors import KNeighborsClassifier
         model = KNeighborsClassifier(n_neighbors=8)
         model.fit(x_train, y_train)

Out[25]: KNeighborsClassifier(n_neighbors=8)

In [26]: model.score(x_test, y_test)

Out[26]: 0.7948717948717948
```

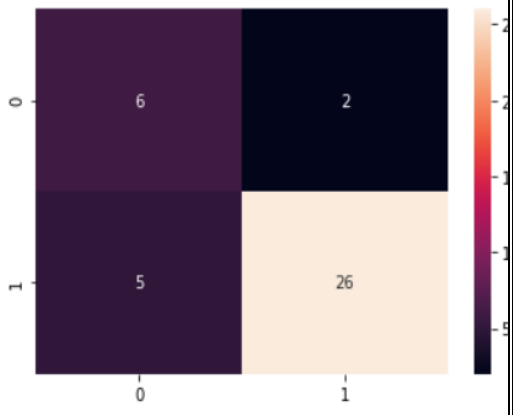
K-Nearest Neighbors:

```
In [27]: from sklearn.metrics import
         y_pred=model.predict(x_test
         cm = confusion_matrix(y_tes
         sns.heatmap(cm ,annot=True)

Out[27]: <AxesSubplot:>
```




```
In [18]: model=XGBClassifier(eval_metric='mlogloss')
         model.fit(x_train,y_train)

         C:\Anaconda\lib\site-packages\xgboost\sklearn.py:1224: UserWarning: The use
         will be removed in a future release. To remove this warning, do th following
         tructing XGBClassifier object; and 2) Encode your labels (y) as integers st
           warnings.warn(label_encoder_deprecation_msg, UserWarning)

Out[18]: XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
                       colsample_bynode=1, colsample_bytree=1, enable_categorical=Fa
                       eval_metric='mlogloss', gamma=0, gpu_id=-1, importance_type=N
                       interaction_constraints='', learning_rate=0.30000001,
                       max_delta_step=0, max_depth=6, min_child_weight=1, missing=na
                       monotone_constraints='()', n_estimators=100, n_jobs=8,
                       num_parallel_tree=1, predictor='auto', random_state=0,
                       reg_alpha=0, reg_lambda=1, scale_pos_weight=1, subsample=1,
                       tree_method='exact', validate_parameters=1, verbosity=None)

In [19]: y_pred=model.predict(x_test)
         print(accuracy_score(y_test, y_pred)*100)

         82.05128205128204
```

```
In [20]: from sklearn.metrics import confusion_matri
         y_pred=model.predict(x_test)
         cm = confusion_matrix(y_test, y_pred)
         sns.heatmap(cm ,annot=True)

Out[20]: <AxesSubplot:>
```



XGBOOST:

XGBoost is an algorithm. That has recently been dominating applied gadget learning. XGBoost set of rules is an implementation of gradient boosted choice timber. That changed into the design for pace and overall performance.

CONCLUSION:

Parkinson's disease affects the CNS of the brain and has yet no treatment unless it's detected early. Late detection leads to no treatment and

loss of life. Thus, its early detection is significant. For early detection of the disease, we utilized various machine learning algorithms to detect Parkinson's disease. We checked our Parkinson disease data and found out that Support Vector Machines is the best Algorithm with accuracy of 84% to predict the onset of the disease which will enable early treatment and save a life

REFERENCES:

1. Shahbakhi, M., Far, D. T., & Tahami, E. (2014). Speech Analysis for Diagnosis of Parkinson's Disease Using Genetic Algorithm and Support Vector Machine. Journal of Biomedical Science and Engineering, 2014

2. Yahia A, Laiali A. (2014). Detection of Parkinson Disease through Voice Signal Features. Journal of American Science 2014; 10 (10), 44-47.

3. Khan Academy. (2015). Movement signs and symptoms of Parkinson's disease | NCLEX-RN | Khan Academy. [online] Available at: https://www.youtube.com/watch?v=4qdD4Ny34cc [Accessed 11 Apr. 2018].

4. Pereira, C.R., Weber, S.A., Hook, C., Rosa, G.H. and Papa, J.P., 2016, October. Deep LearningAided Parkinson's Disease Diagnosis from Handwritten Dynamics. In Graphics, Patterns and Images (SIBGRAPI), 2016 29th SIBGRAPI Conference on(pp. 340-346). IEEE.

5.Prechelt, L., 1998. Early stopping-but when?. In Neural Networks: Tricks of the trade (pp. 55-69). Springer, Berlin, Heidelberg.

6. Samii A., J.G. Nutt, B.R. Ransom.Parkinson's disease. Lancet, 363 (2004), pp. 1783-1793

7.Geeks for Geeks- Data Preprocessing

8. Chen, A. H., Lin, C. H., & Cheng, C. H. New approaches to improve the performance of disease classification using nested–random forest and nested– support vector machine classifiers. Cancer, 2 (10509), 102.

9. Das, R. (2010). A comparison of multiple classification methods for diagnosis of Parkinson disease. Expert Systems with Applications, 37 (2), 1568-1572

10: Souillard-Mandar, W., Davis, R., Rudin, C., Au, R., Libon, D.J., Swenson, R., Price, C.C., Lamar, M. and Penney, D.L., 2016. Learning classification models of cognitive conditions from subtle behaviors in the digital Clock Drawing Test. Machine learning, 102(3), pp.393-441.

11. C. Lu, B. Deng, Y. Zhu, J. Wang and C. Liu, "Effect of DBS Targeting Striatum on beta Oscillations in Parkinson's Disease," 2019 Chinese Control Conference (CCC), Guangzhou, China, 2019, pp. 502-507. doi: 10.23919/ChiCC.2019.8866389

12. Shalev-Shwartz, S. and Ben-David, S., 2014. Understanding machine learning: From theory to algorithms. Cambridge university press.