Parkinson's Disease Detection using Machine Learning

Nithin Chandra Gupta Samudrala
Metallurgical Engineering and Materials Science
200110076@iitb.ac.in

Abstract: This report contains what is Parkinson's disease and to detect the early onset of the disease. We will use here XGBoost and Support Vector Machines (SVMs) and utilize the data-set available for training models.

Introduction:

Parkinson Disease is a brain neurological disorder. It leads to shaking of the body, hands and provides stiffness to the body. No proper cure or treatment is available yet at the advanced stage. Treatment is possible only when done at the early or onset of the disease. More than 145,000 people have been found alone suffering in the U.K and in India, almost one million population suffers from this disease and it's spreading fast in the entire world.

What causes Parkinson's disease is still unclear, but researchers have research that several factors are responsible for triggering the disease. It includes Genes, Environment etc. We will make use of XGBoost and SVMs to check which is the best algorithm for detection of the onset of disease.

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.

Support Vector Machine(SVM):

Algorithm for the analysis of classification and regression is the support vector machine. It is a supervised machine algorithm used. Image classification and hand-written recognition are where the support vector machine comes to hand used. It sorts the data in one out of two categories and displays the output with the margin between the two as far as possible.

Features considered for Detection:
MDVP:Fo(Hz), MDVP:Fhi(Hz), MDVP:Flo(Hz), MDVP:Jitter(%),
MDVP:Jitter(Abs), MDVP:RAP, MDVP:PPQ, Jitter:DDP, MDVP:Shimmer,
MDVP:Shimmer(dB), Shimmer:APQ3, Shimmer:APQ5, MDVP:APQ,
Shimmer:DDA, NHR, HNR, RPDE, DFA, spread1, spread2, D2, PPE

Pre-processing of the data:
In pre-processing the data, it contains data of patients who are suffering
from Parkinson's Disease, so the result can be biased to the patient status
to be active. So, to make the data unbiased, I used SMOTE in oversampling
the minority data.
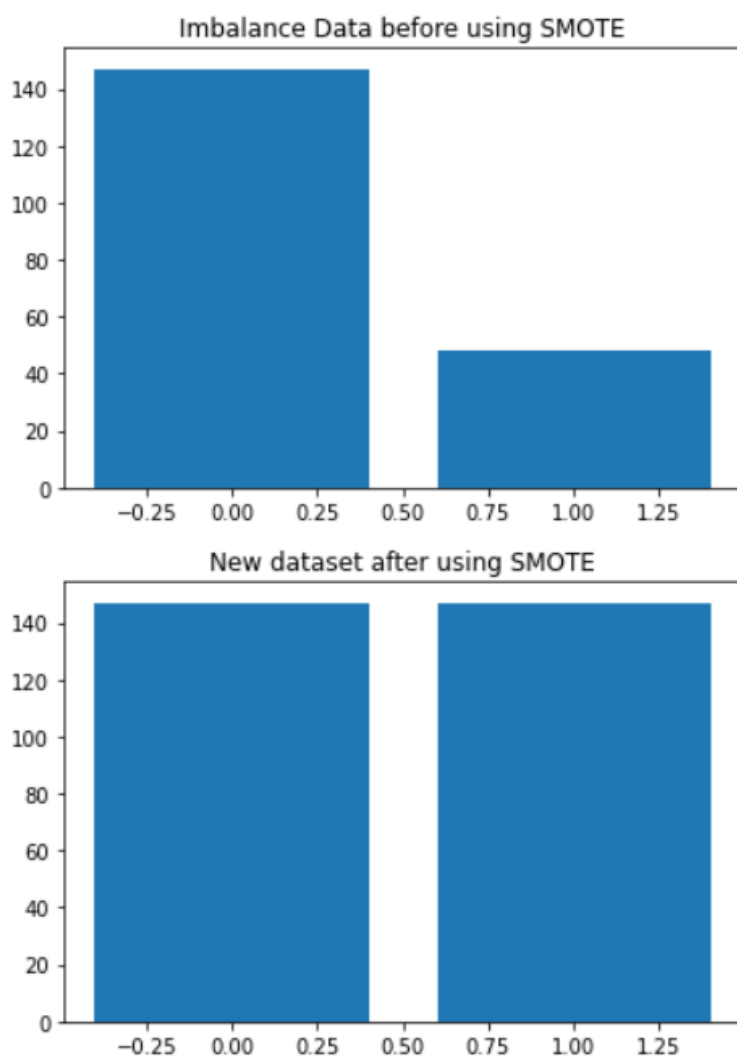
Feature Selection:
All features cannot play a role in identifying the person suffering from
Parkinson's Disease. And using many features may cause overfitting of the
model. So, I used Lasso Regression to select features which can have
higher weight over predicting the person suffering from parkinson's Disease.
And it turned out only *spread1 and spread2* has greater influence over the
prediction.

Implementation:

After pre-processing of the data, we load the data to tain the XGBoost and SVM models. We train these models using the above features. And after training the XGBoost model, for prediction, XGBoost gives the disease status of the person between 0 and 1. I took 0.5 value as threshold, so that, symptomatic person disease status and accuracy score of the model can be retrieved.

Results:

After experimenting on XGBoost and SVM, SVM model gave better accuracy. Below is the Classification report for the SVM.

	precision	recall	f1-score	support
0	0.86	0.97	0.91	32
1	0.96	0.81	0.88	27
accuracy			0.90	59
macro avg	0.91	0.89	0.90	59
weighted avg	0.90	0.90	0.90	59

Below is the confusion matrix for SVM model

```
[[31  1]
 [ 5 22]]
```

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 machine learning algorithms such as XGBoost and SVM. We checked our Parkinson disease data and find out Support Vector Machine(SVM) is the best Algorithm to predict the onset of the disease which will enable early treatment and save a life with 89.83% accuracy.