Speech Recognition and Diagnosis of Parkinson’s disease using voice impairments

Abstract

In our project we build a speech recognition system which recognises what a person is talking and one application of it namely diagnosis of Parkinson’s using different machine learning techniques. We used UCLA data. We performed feature engineering and **reduced 27 to 16 features**. Overall, we found the **logistic regression** was our best **model with 90% accuracy and f1 score of 90%** with **recall** (diseased) of **96 %** for our test subjects. The **extra trees with 93.6% overall score** seems to be better but due to explain ability, we came to conclusion that Logistic regression is the best overall model.

Introduction and Motivation

The Speech recognition is one of the main applications of speech processing and we try to extract it for different local languages. And we followed it up with one of the applications which is diagnosis of Parkinson’s disease which will follow our speech recognition model. There are several diseases which can be identified by speech like depression, mental disorders. In our study we have used UCLA dataset repository which has data about 240 patients with 17 features in our final dataset. Initial feature engineering, data cleaning was performed using excel. This integration of both is yet to be done and if done it could be a major deal breaker as we can cater diagnosis to many patients in a large country like India where many people don’t speak English.

Dataset Description

The original dataset had a lot of features namely 28 features and there were also sensitive information like patient ID, age, duration of disease, gender and a lot of **Personally identifiable information**, We removed all of PII (Personally identifiable information) so that we are in compliance of various regulation like GDPR. GDPR asks us to use less of PII data while building any machine learning models. PII can also introduce bias into our model unconsciously. The features which we selected were Jitter and shimmer. Jitter and shimmer are acoustic characteristics of voice signals, and they are caused by irregular vocal fold vibration. They are perceived as roughness, breathiness, or hoarseness in a speaker’s voice. All natural speech contains some level of jitter and shimmer, but measuring them is a common way to detect voice pathologies. The speech production system is not a rigid, mechanical machine, but composed of an assortment of soft-tissue components. Therefore, although parts of a speech signal might seem stationary, there are always small fluctuations in it, as vocal fold oscillation is not exactly periodic. Variations in signal frequency and amplitude are called jitter and shimmer, respectively. There are several different ways to measure jitter and shimmer. For instance, when detecting voice disorders, they are measured as percentages of the average period, where values above certain thresholds are potentially related to pathological voices. Jitter and shimmer are most clearly detected from long, sustained vowels.

A commonly used jitter value is the absolute jitter. This measure expresses the average absolute difference between consecutive periods.

Let us see the formula for absolute and relative Jitter

$$Jitter(relative) = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\|T_i - T_{i+1}\|}{\frac{1}{N}\sum_{i=1}^{N} T_i}$$

$$Jitter(absolute) = \frac{1}{N-1}\sum_{i=1}^{N-1}\|T_i - T_{i+1}\|$$

Similarly Shimmer also has two types of calculations namely

$$Shimmer(dB) = \frac{1}{N-1}\sum_{i=1}^{N-1}\|20\log(A_{i+1}/A_i)\|$$

$$Shimmer(relative) = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\|A_i - A_{i+1}\|}{\frac{1}{N}\sum_{i=1}^{N} A_i}$$

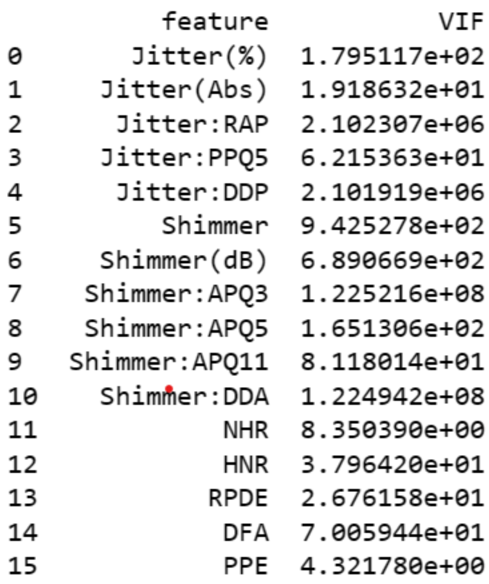
The Noise to harmonic ratio, Harmonics-to-noise ratio were also used along with Recurrence Period Density Entropy (RPDE) is a feature that is used in speech processing to determine the periodicity of the time series by measuring the repetitiveness in the phase space of the system. DFA and PPE features were also used.

Implementation

Variation Inflation factor

The main features were selected using feature importance from initial classifiers and research papers. The **Variation Inflation factor** was done to check multicollinearity among **independent input variables.**

VIF Value	Description
1-5	No multicollinearity
5-20	Medium multicollinearity
Above 20	High multicollinearity



```
          feature           VIF
0       Jitter(%)  1.795117e+02
1     Jitter(Abs)  1.918632e+01
2      Jitter:RAP  2.102307e+06
3     Jitter:PPQ5  6.215363e+01
4      Jitter:DDP  2.101919e+06
5         Shimmer  9.425278e+02
6     Shimmer(dB)  6.890669e+02
7    Shimmer:APQ3  1.225216e+08
8    Shimmer:APQ5  1.651306e+02
9   Shimmer:APQ11  8.118014e+01
10    Shimmer:DDA  1.224942e+08
11            NHR  8.350390e+00
12            HNR  3.796420e+01
13           RPDE  2.676158e+01
14            DFA  7.005944e+01
15            PPE  4.321780e+00
```

Since VIF Is less than 5, there is no significant cross correlation or multi collinearity between input independent variables

Test Train Split

The test train split was done with 80% of data being used for training and 20% for testing at random.

The practise which is commonly used is given below

Purpose	Data Percentage
Training	60
Validation	20
Testing	20

Standard Scaler

The standard scaler is applied to all the input independent variables

$$x_{scaled} = \frac{x - mean}{sd}$$

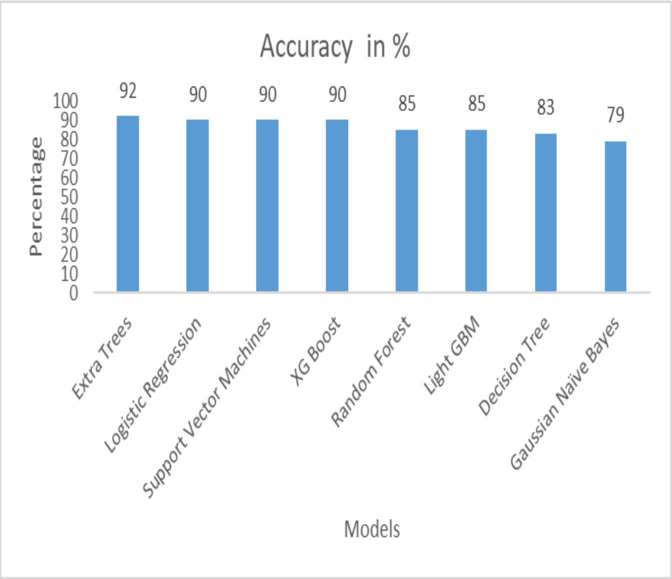
Models

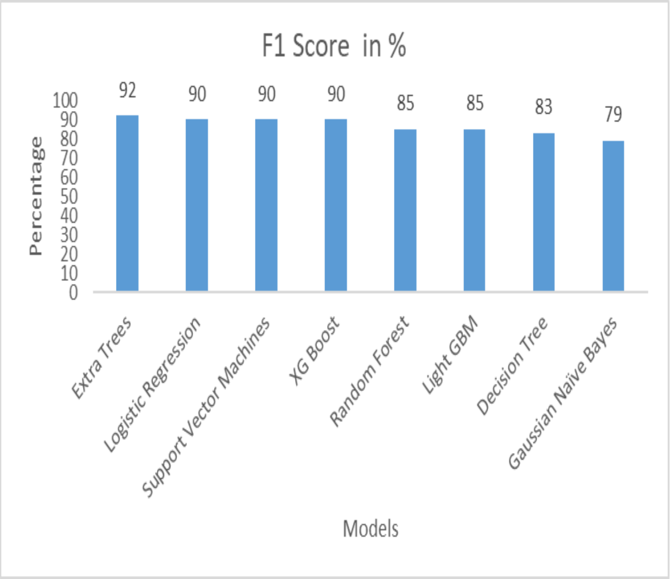
A variety of models were tried out and summary of models tried out is given in the below table

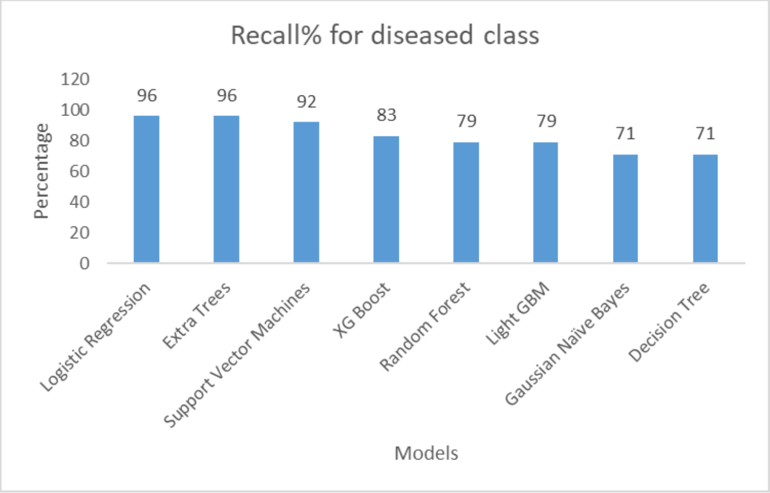
Initial Metrics

S.No	Model	Accuracy %	Recall% for diseased class	F1 Score %
1	Logistic Regression	90	96	90
2	Gaussian Naïve Bayes	79	71	79
3	Support Vector Machines	90	92	90
4	Decision Tree	83	71	83
5	Random Forest	85	79	85
6	Extra Trees	92	96	92
7	Light GBM	85	79	85
8	XG Boost	90	83	90

To determine the best models, we have used three main metrics which is accuracy, recall for 1 class which is diseased class and F1 score. The reason to choose for recall for diseased class is if a person is not diagnosed with disease but later turns out to have disease is more dangerous, hence we decided to give importance to recall for 1 class.







Hyperparameter Tuning

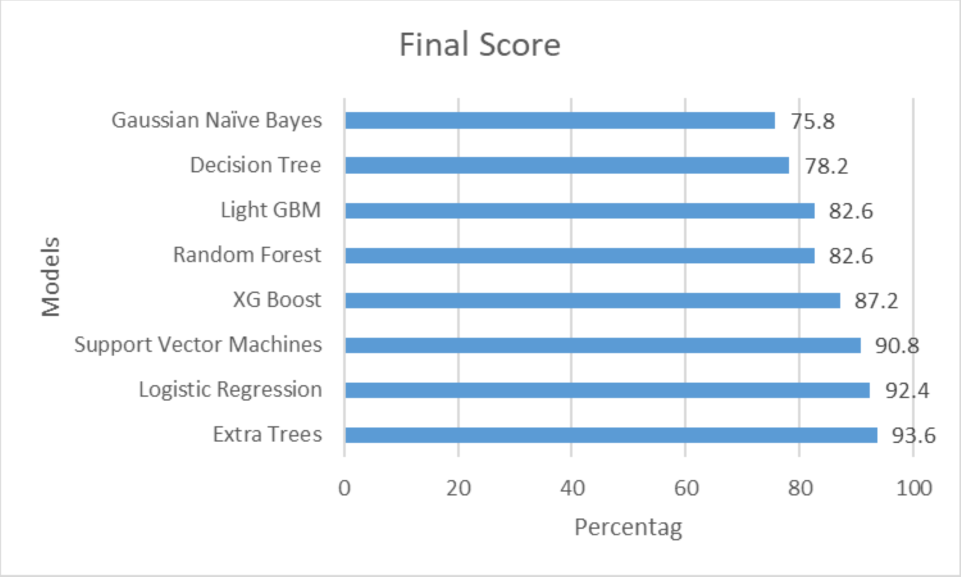
The hyperparameter tuning was done to Support vector machine using **grid search** and best model turned out to be SVC(C=10, gamma=0.1, kernel='poly'). The poly kernel with slack variable of 10 and gamma value of o.1 turned out to be the best fit.

The Overall weighted scores are given below with more weightage given to recall for diseased class

Metric	Weightage
Accuracy	0.4
Recall for diseased class	0.3
F1 Score	0.3

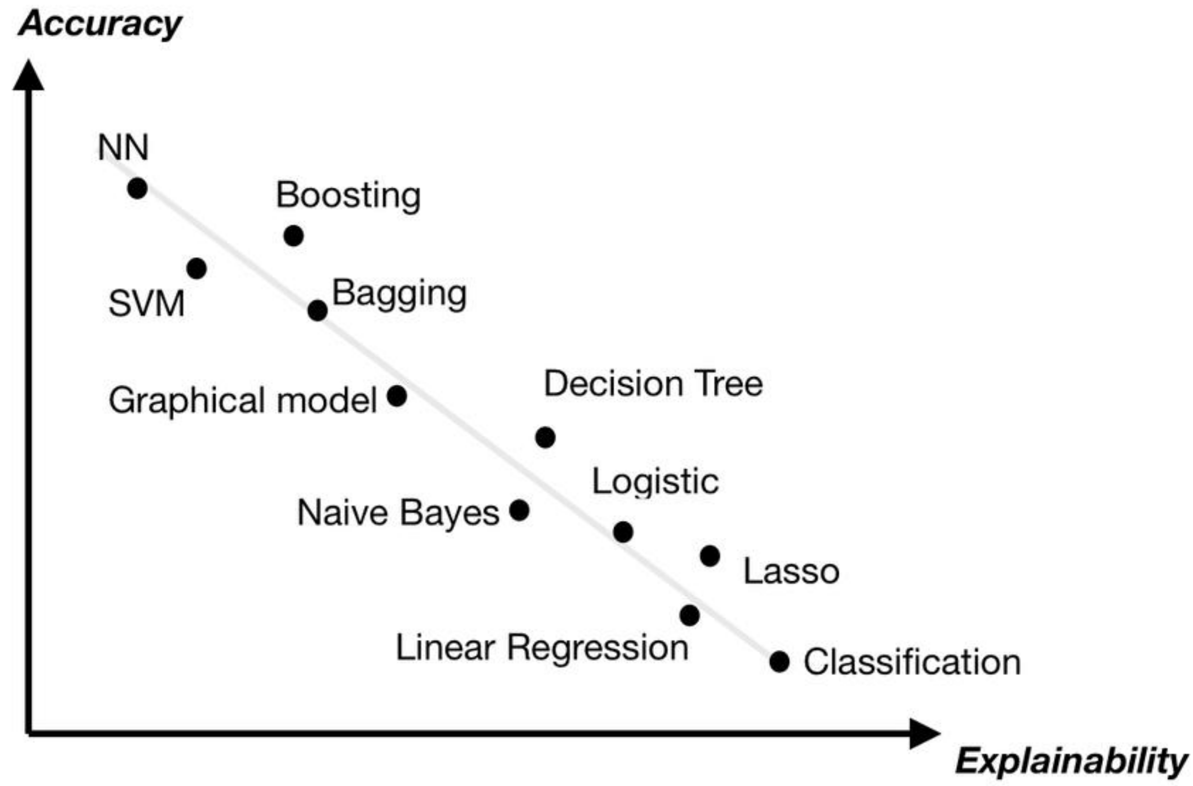
The final consolidated scores are given below

S.No	Model	Overall Score
1	Extra Trees	93.6
2	Logistic Regression	92.4
3	Support Vector Machines	90.8
4	XG Boost	87.2
5	Random Forest	82.6
6	Light GBM	82.6
7	Decision Tree	78.2
8	Gaussian Naïve Bayes	75.8



From this, we can clearly see **extra trees is the best classifier but we still choose logistic regression because of explainability.**

In **Logistic Regression** we can easily say why it is predicted as normal or diseased using **regression coefficients** and it is also provide faster results, thus suitable for larger datasets. In Extra trees we have use **LIME** or **SHAP** further additionally to interpret model results.



Hence, we come to conclusion that Logistic regression is the best model among considering all factors

Conclusion

Thus in our project we have built a speech recognition system which recognises what a person is talking and one application of it namely diagnosis of Parkinson's using different machine learning techniques. The Logistic regression with overall score of **92.4%** with **90% accuracy, F1 score and 96% recall for diseased class is the best model considering all factors like speed, explainability.**