Parkinsons Data Set Predictions

By Himanee T.



Citation

The following paper: 'Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection', Little MA, McSharry PE, Roberts SJ, Costello DAE, Moroz IM. BioMedical Engineering OnLine 2007, 6:23 (26 June 2007)

Information about Data

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 195 entries, 0 to 194
Data columns (total 24 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   name              195 non-null    object 
 1   MDVP:Fo(Hz)       195 non-null    float64
 2   MDVP:Fhi(Hz)      195 non-null    float64
 3   MDVP:Flo(Hz)      195 non-null    float64
 4   MDVP:Jitter(%)    195 non-null    float64
 5   MDVP:Jitter(Abs)  195 non-null    float64
 6   MDVP:RAP          195 non-null    float64
 7   MDVP:PPQ          195 non-null    float64
 8   Jitter:DDP        195 non-null    float64
 9   MDVP:Shimmer      195 non-null    float64
 10  MDVP:Shimmer(dB)  195 non-null    float64
 11  Shimmer:APQ3      195 non-null    float64
 12  Shimmer:APQ5      195 non-null    float64
 13  MDVP:APQ          195 non-null    float64
 14  Shimmer:DDA       195 non-null    float64
 15  NHR               195 non-null    float64
 16  HNR               195 non-null    float64
 17  status            195 non-null    int64  
 18  RPDE              195 non-null    float64
 19  DFA               195 non-null    float64
 20  spread1           195 non-null    float64
 21  spread2           195 non-null    float64
 22  D2                195 non-null    float64
 23  PPE               195 non-null    float64
dtypes: float64(22), int64(1), object(1)
memory usage: 36.7+ KB
```

More Information about Data

Grouping the 3 columns Status, D2, PPE

```
status  D2
0       1.423287    0.044539
        1.512275    0.096320
        1.544609    0.056141
        1.743867    0.085569
        1.840198    0.103561
                       ...
1       3.184027    0.454721
        3.274865    0.418646
        3.317586    0.301952
        3.413649    0.457533
        3.671155    0.332086
Name: PPE, Length: 195, dtype: float64
```

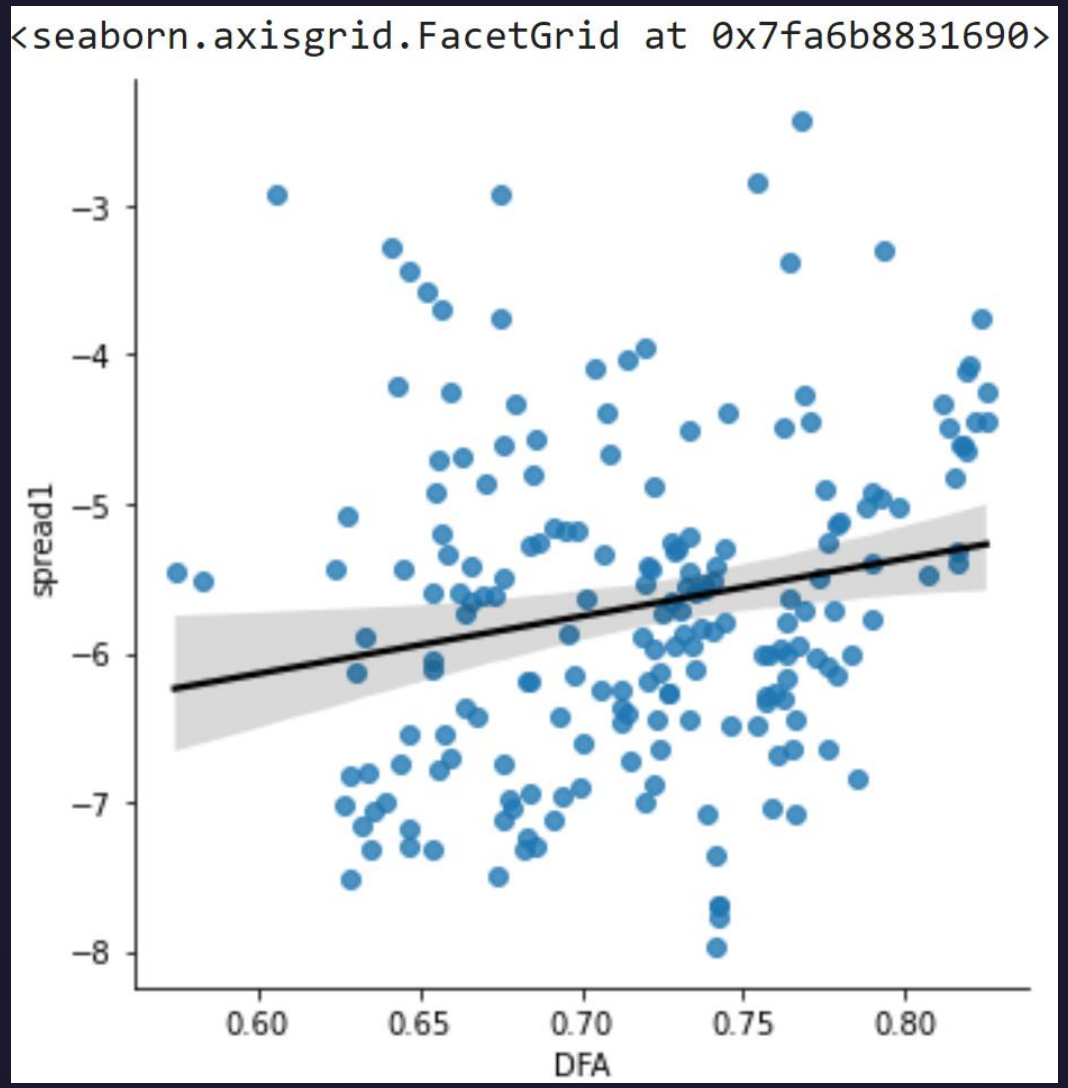
More Information about Data

Grouping the 3 columns Status, D2, PPE, MDVP, MDVP(Abs)

```
status  MDVP:Jitter(%)
0       0.00178           0.000009
        0.00180           0.000009
        0.00185           0.000007
        0.00198           0.000010
        0.00205           0.000009
                            ...
1       0.01936           0.000150
        0.02714           0.000140
        0.03011           0.000220
        0.03107           0.000160
        0.03316           0.000260
Name: MDVP:Jitter(Abs), Length: 181, dtype: float64
```

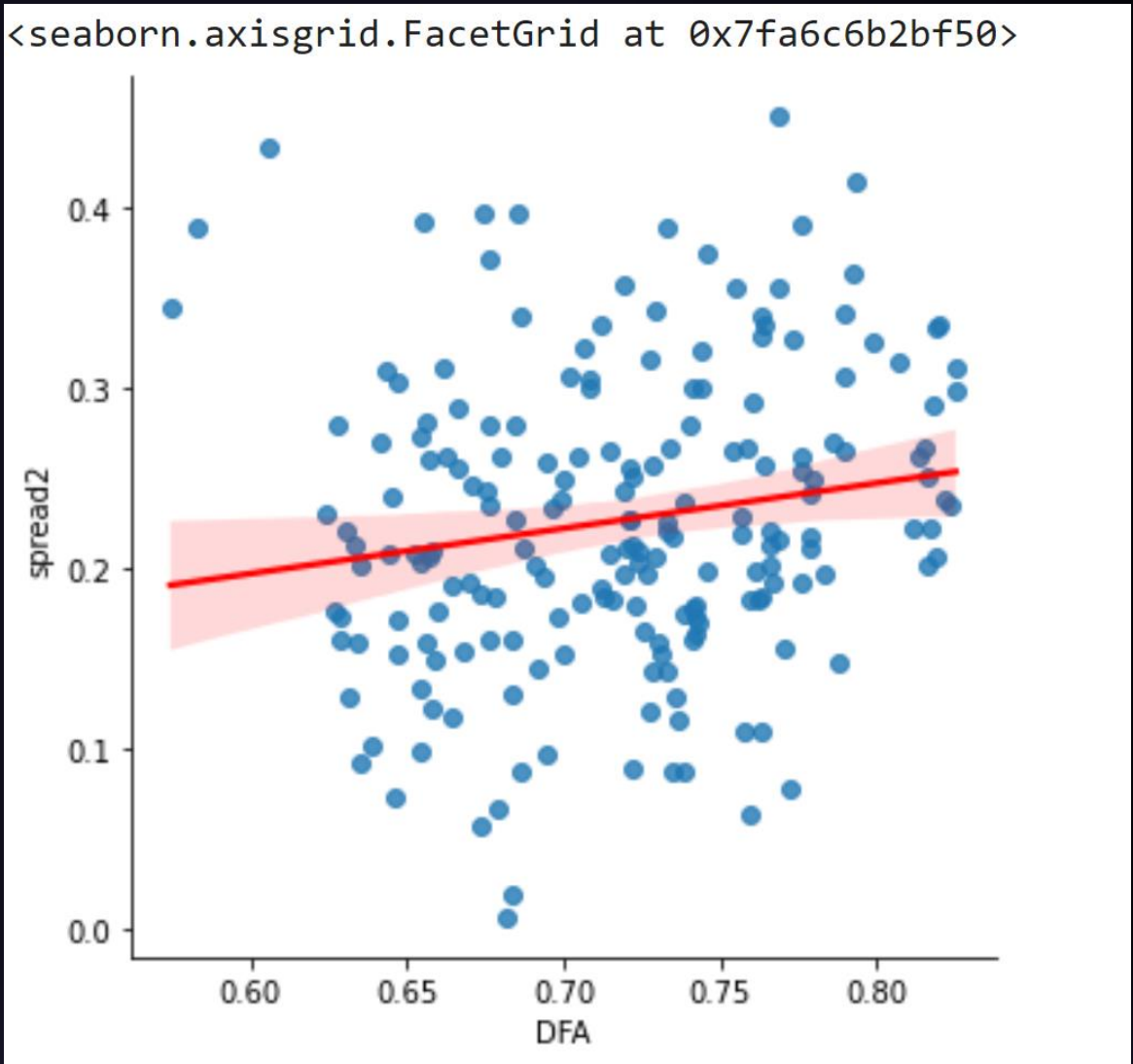
Visualization of Data

Graph Showing Positive Corelation between Spread1 and DFA



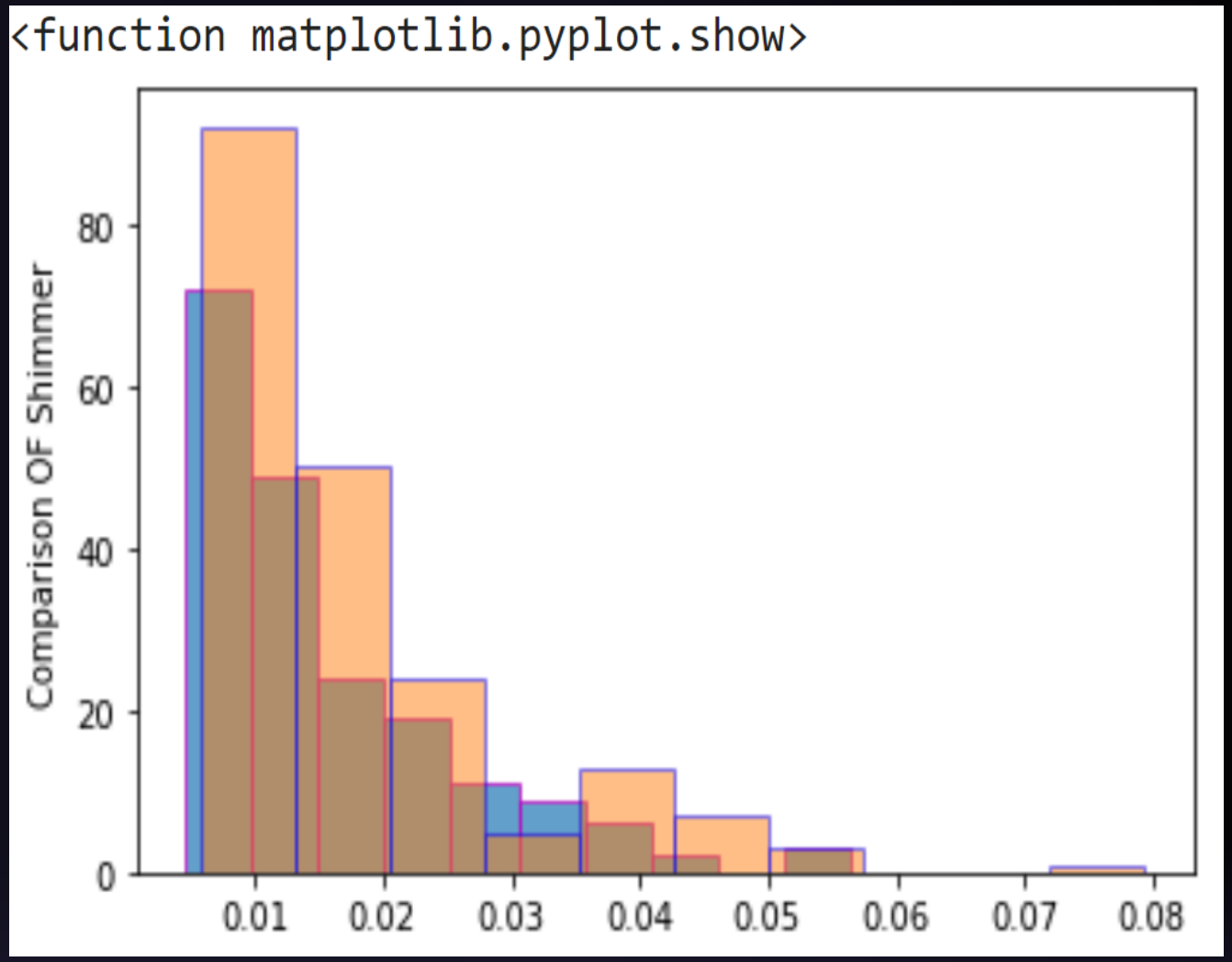
Visualization of Data

Graph Showing Positive Corelation between Spread2 and DFA

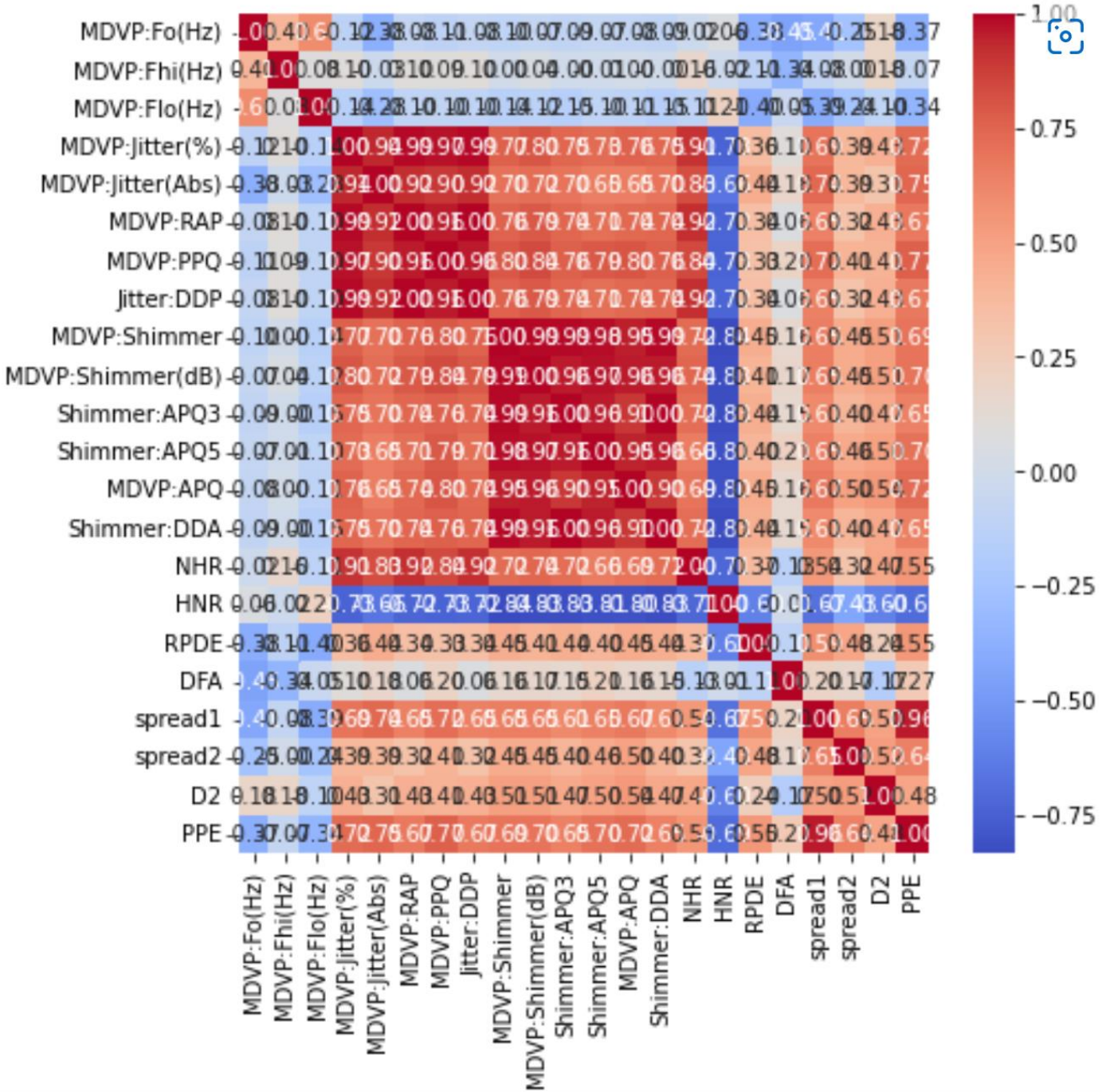


Visualization of Data

Graph Showing Comparison of Shimmer APQ3 and APQ5



<matplotlib.axes._subplots.AxesSubplot at 0x7fa6afa229d0>



Visualization of Data

Heatmap showing relation between all columns

KNeighbours Classifier

Classification Report showing
Precision, recall and F1-score

	precision	recall	f1-score	support
0	0.88	0.64	0.74	11
1	0.90	0.97	0.94	38
accuracy			0.90	49
macro avg	0.89	0.81	0.84	49
weighted avg	0.90	0.90	0.89	49

Decision Tree Classifier

Classification Report showing Precision, recall and F1-score

	precision	recall	f1-score	support
0	0.78	0.64	0.70	11
1	0.90	0.95	0.92	38
accuracy			0.88	49
macro avg	0.84	0.79	0.81	49
weighted avg	0.87	0.88	0.87	49

Dummy Classifier

Classification Report showing
Precision, recall and F1-score

	precision	recall	f1-score	support
0	0.00	0.00	0.00	11
1	0.78	1.00	0.87	38
accuracy			0.78	49
macro avg	0.39	0.50	0.44	49
weighted avg	0.60	0.78	0.68	49

Principal Component Analysis (PCA)

PCA speeds up training time and, sometimes improve predictions.

```
Training accuracy: 0.9452054794520548
Testing accuracy: 0.8979591836734694
```