

Parkinson's Disease Prediction

IDS-506

Health Information Management and Analytics

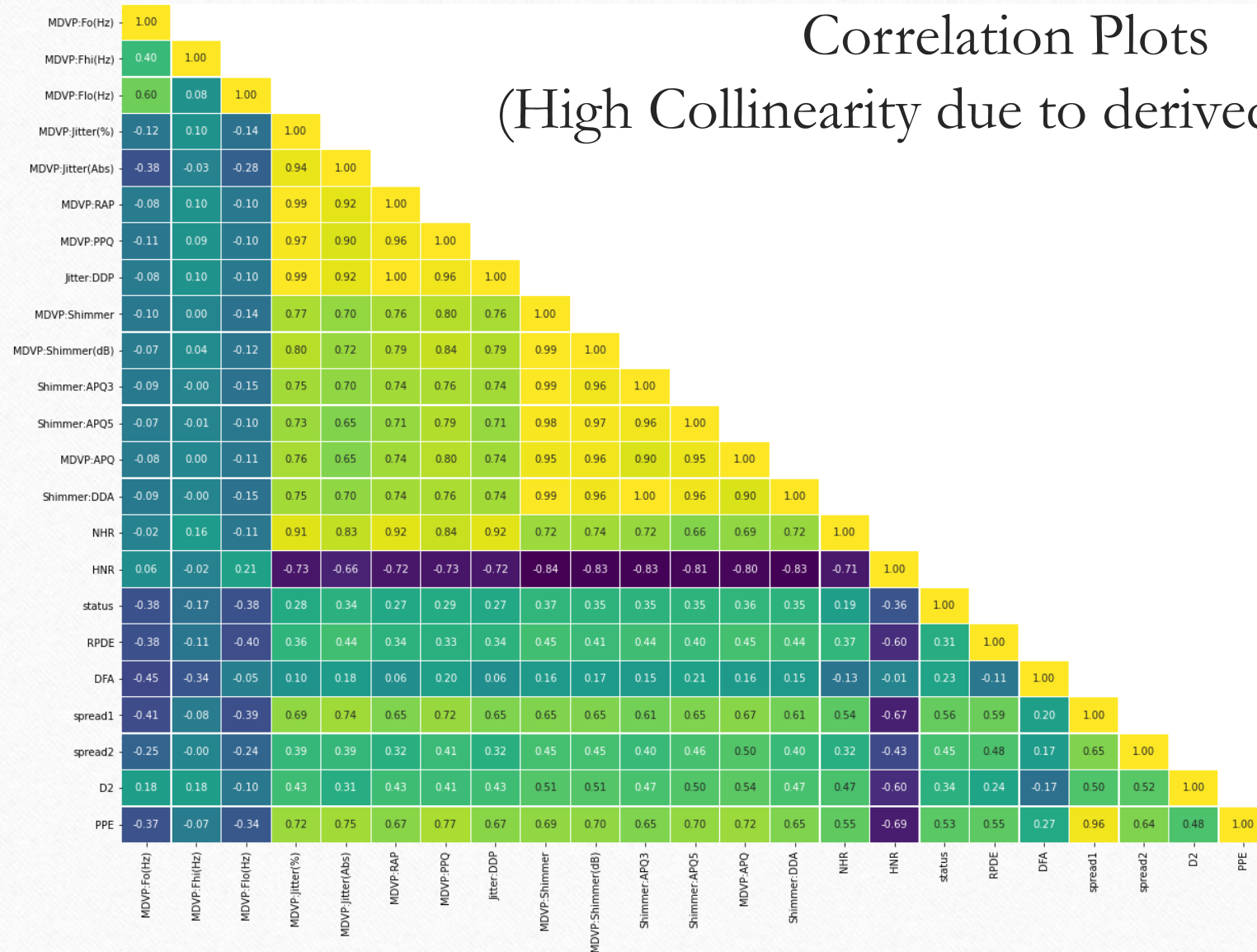
Vivek Kumar

The Case & Data

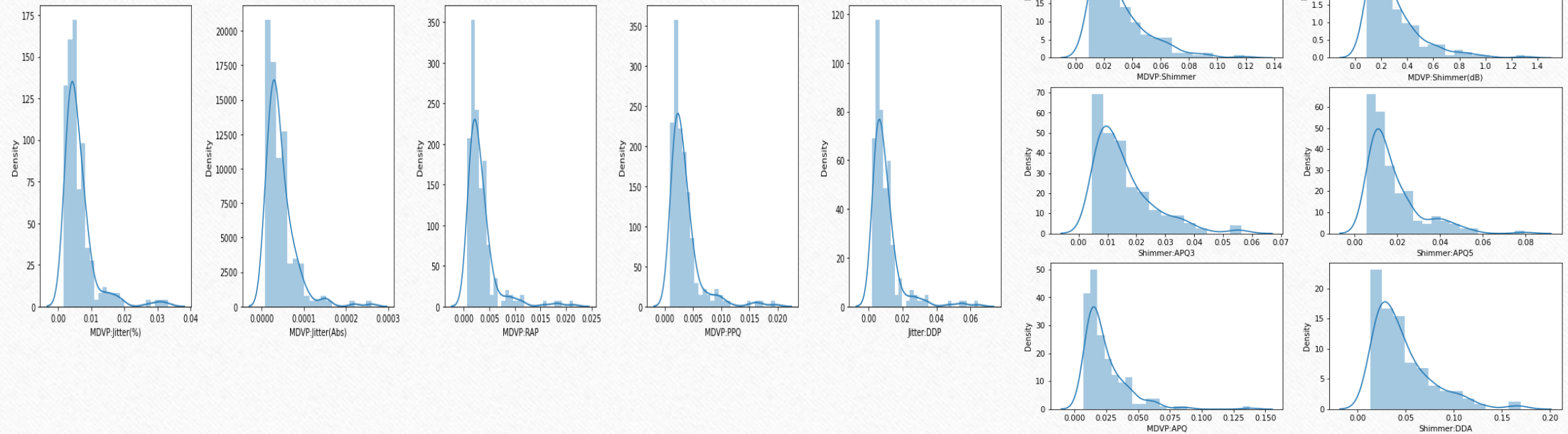
- The dataset provided has biomedical voice measurements of around 6 observation each from 32 people, out of these 24 have Parkinson's disease
- The objective of this assignment is to differentiate healthy people from those having Parkinson's disease
- Dataset has 195 records and 24 attributes, other than name & status rest 22 attributes are voice related characteristics derived from voice recordings from 32 people (3/4th having Parkinson's Disease)
- The dataset has zero missing values across all attributes

Correlation Plots

(High Collinearity due to derived variables)



Distribution Plots for Variation in Frequency and Amplitudes



Data Transformation using Linear Discriminant Analysis

- The highly correlated vocal parameters are combined by using the linear discriminant analysis method
- Created MDVP_LDA for `MDVP:Jitter(%)'`, `'MDVP:Jitter(Abs)'`, `'MDVP:RAP'`, `'MDVP:PPQ'`, `'Jitter:DDP'`
- Created Shimmer_LDA for `'MDVP:Shimmer'`, `'MDVP:Shimmer(dB)'`, `'Shimmer:APQ3'`, `'Shimmer:APQ5'`, `'MDVP:APQ'`, `'Shimmer:DDA'`

Logistic Regression (Baseline Model)

Classification Report

	precision	recall	f1-score	support
0	1.00	0.43	0.60	7
1	0.89	1.00	0.94	32
accuracy			0.90	39
macro avg	0.94	0.71	0.77	39
weighted avg	0.91	0.90	0.88	39

Odds-Ratio

MDVP:F0(Hz)	0.992036
MDVP:Fhi(Hz)	0.995126
MDVP:Flo(Hz)	0.998248
NHR	1.003644
HNR	0.978807
RPDE	0.670464
DFA	1.131930
spread1	4.888483
spread2	1.347137
D2	5.548352
PPE	1.195863
MDVP_LDA	0.896149
Shimmer_LDA	0.896149

Logistic Regression on reduced variables based on ICPR (Interclass Probability Risk) and MKLD (Modified Kullback-Leibler Divergence) values

(ICPR < 0.6; MKLD > 1)

Features	ICPR	MKLD
MDVP:F0	0.59	1.02
MDVP:Fhi	0.71	0.26
MDVP:Flo	0.67	1.13
HNR	0.66	0.73
RPDE	0.71	0.25
D2	0.76	0.22
Spread1	0.56	0.89
Spread2	0.69	0.46
MDVP-LDA	0.59	1.96
Shimmer-LDA	0.46	5.72

Classification Report

	precision	recall	f1-score	support
0	1.00	0.43	0.60	7
1	0.89	1.00	0.94	32
accuracy			0.90	39
macro avg	0.94	0.71	0.77	39
weighted avg	0.91	0.90	0.88	39

Odds-Ratio

MDVP:F0(Hz)	0.991919
MDVP:Fhi(Hz)	0.995120
MDVP:Flo(Hz)	0.998271
HNR	0.977413
RPDE	0.654947
spread1	4.903031
spread2	1.317114
D2	5.487451
MDVP_LDA	0.897023
Shimmer_LDA	0.897023

NCBI Article Referred: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5434464/#sec3>

Decision Tree

Feature Importance

	Imp
MDVP:F0(Hz)	0.000000
MDVP:F1i(Hz)	0.138056
MDVP:F1o(Hz)	0.000000
NHR	0.000000
HNR	0.034905
RPDE	0.000000
DFA	0.118203
spread1	0.526898
spread2	0.121141
D2	0.049802
PPE	0.000000
MDVP_LDA	0.010995
Shimmer_LDA	0.000000

	precision	recall	f1-score	support
0	0.50	0.33	0.40	9
1	0.82	0.90	0.86	30
accuracy			0.77	39
macro avg	0.66	0.62	0.63	39
weighted avg	0.74	0.77	0.75	39

