

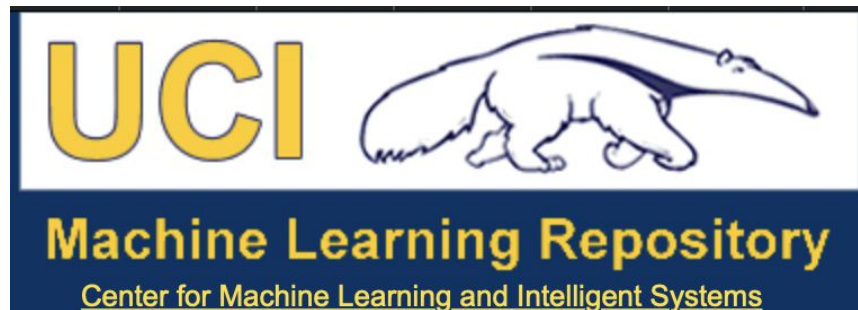
# Predicting a Parkinson's Diagnosis

---

Data Analyst: Rebecca Lorey

# Data Set - General Information

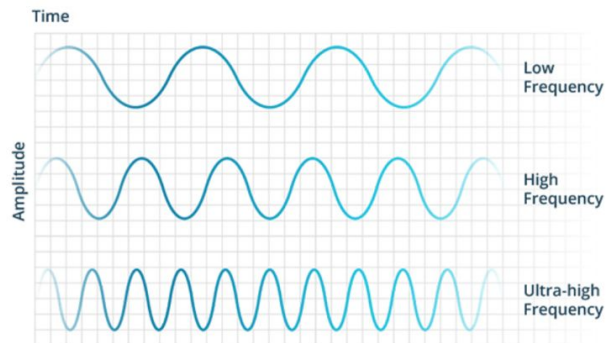
- UCI Machine Learning Repository
- Collected by Max Little at the University of Oxford
- MDVP - Multidimensional Voice Program
- 31 people
- 23 diagnosed with Parkinson's
- 195 voice recordings



# Data Set - Features

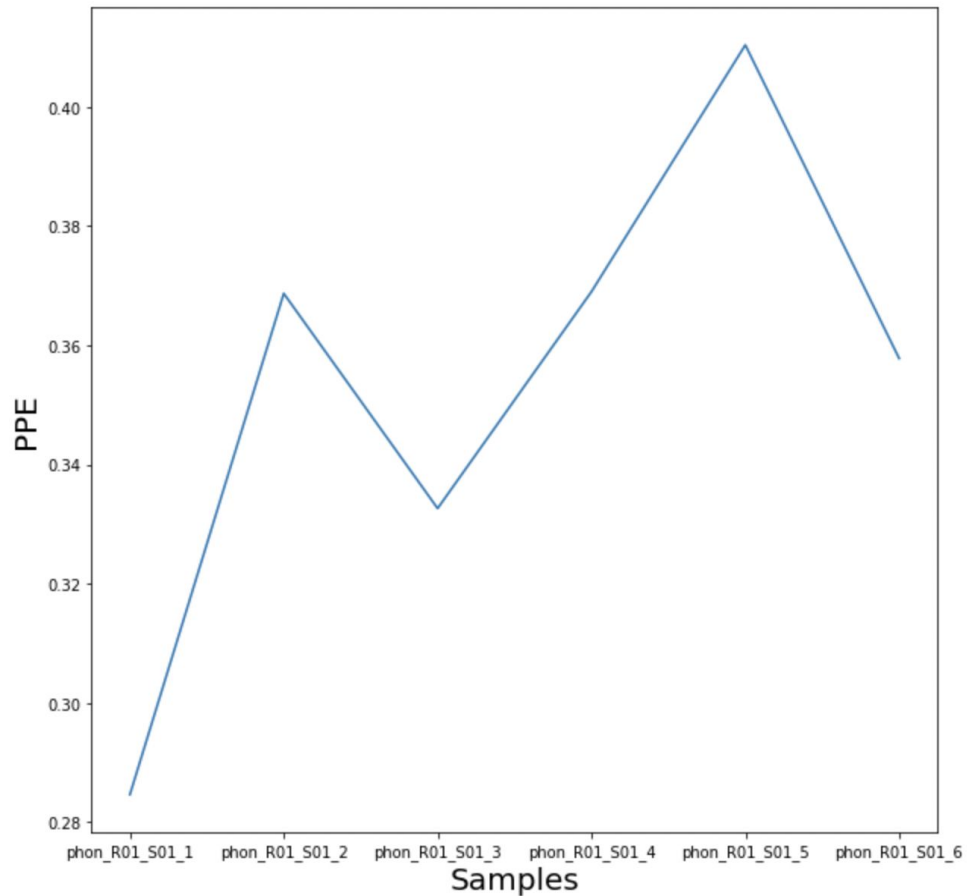
## Positive Diagnosis of Parkinson's

### Biomedical Voice Measurements

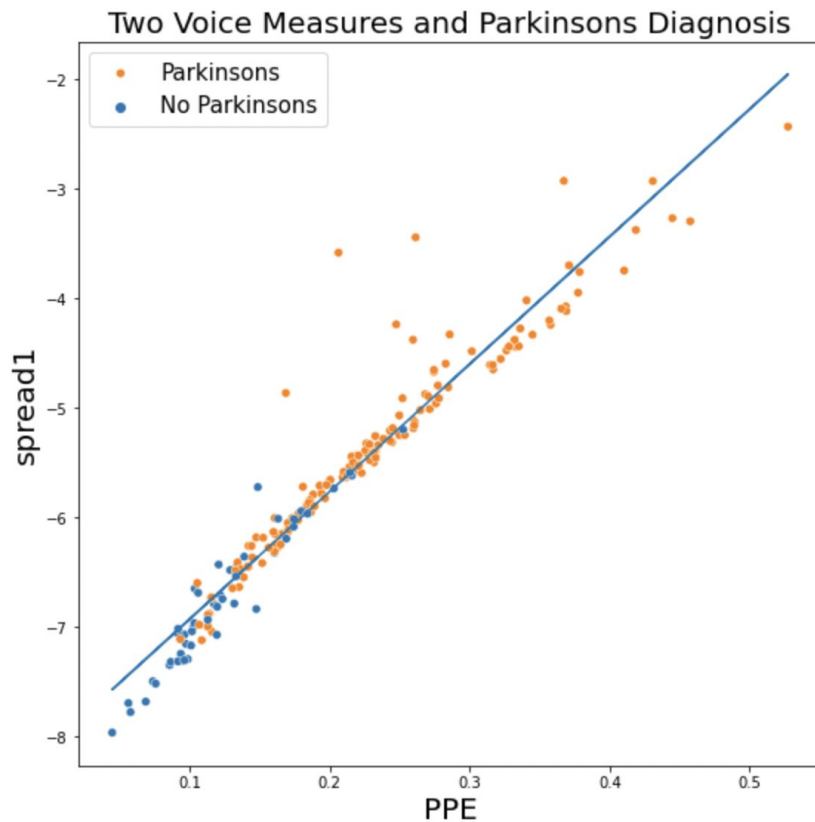


- 23 Different Features
- Various Vocal Frequencies
- Measures of Variation in Fundamental Frequency
- Measures of Variation in Amplitude
- Ratio of Noise to Tonal Components in Voice

# Multiple Samples per Person



# Voice Measure Relationships and Parkinson's Diagnosis



# Chosen Model

Light Gradient Boosted Machine - Optimized



LightGBM

# Model Evaluation

- 92% Accuracy on ability to correctly predict Parkinson's Disease
- 83% Precision at predicting No Parkinson's Disease
- 95% Precision at predicting Parkinson's Disease
- Lower percentage of false negatives
- Limitations on size of data set used to create model and variable of time not given
- Test run - 85% Accuracy at predicting Parkinson's over 5,800 samples



# Final Recommendations:

- Improve model with more data including more data from non-PD diagnosed individuals
- Determine sample timestamps from given data set if possible
- Explore possible relationship between early MDVP parameters and future diagnosis of Parkinson's
- Model MDVP parameters along with clinical symptoms to predict PD

