Predicting Parkinson's Disease Severity & Progression with Machine Learning Models

Annika Dinulos

Milestone Project 3

TABLE OF CONTENTS

01

Key Takeaways

Background and Main Findings

01

1 Million People in the US Live with Parkinson's Disease

PARKINSON'S DISEASE (PD)

Disease is progressive, worsening over time

Tremors

In hands, arms, legs, jaw, head

Muscle Stiffness

Where muscle is contracted for a long time

Slowness of Movement

Including impaired balance and coordination, could lead to falls

National Institute on Aging. (n.d.). Parkinson's Disease: Causes, Symptoms, and Treatments. National Institute on Aging. Retrieved May 1, 2023, from https://www.nia.nih.gov/health/parkinsons-disease

01

Important Topic for Healthcare Advancement

- This model and presentation is intended for anyone in the healthcare industry, specifically neurological departments or academia.

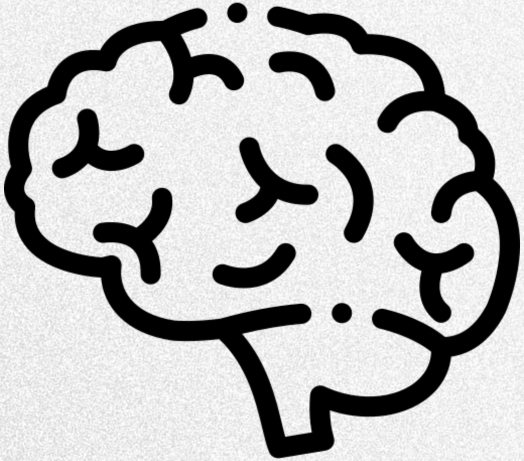
- This information may also be interesting to people who have family history of PD since the Unified Parkinson’s Disease Rating Scale (UPDRS) may be an accessible way to track symptoms.

01

Key Takeaways

- Disease progression can be predicted with high degree accuracy with an XGBoost model.

- Could be an important advancement in PD care plans and in research!



href="https://www.flaticon.com/free-icons/brain" title="brain icons">Brain icons created by Freepik - Flaticon

02

Research question:
Can we predict disease severity & progression in early PD patients using their motor and total UPDRS scores?

03

Data

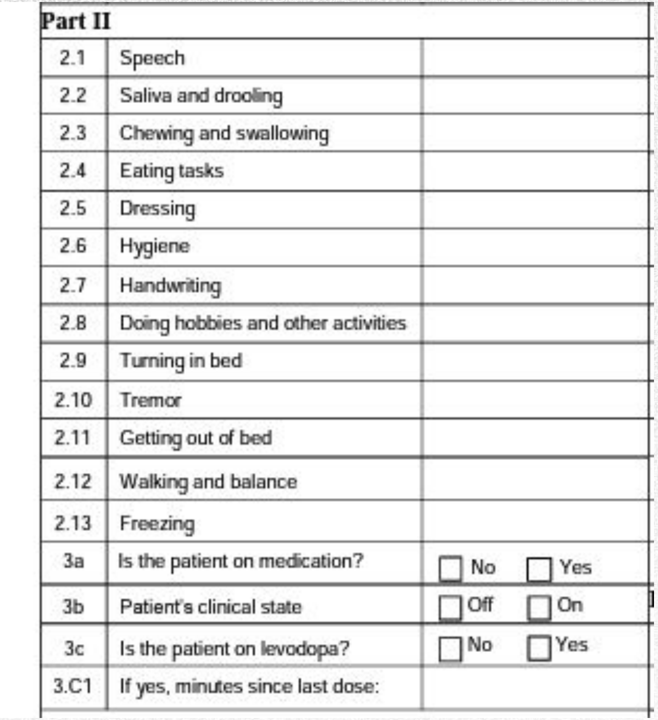
Parkinson’s Telemonitoring Dataset

03 Parkinson’s Telemonitoring Dataset

- Data: Parkinson's Telemonitoring Dataset from the UCI Machine Learning Repository
- Biomedical voice measurements from 42 case studies
- The Unified Parkinson’s Disease Rating Scale (UPDRS) is a metric used to track disease progression. Higher UPDRS scores indicate more severe PD symptoms

Two target variables were chosen: **motor UPDRS** and **total UPDRS**

Vocal symptoms fall under the motor section of the UPDRS, and total UPDRS scores include all symptoms.



Part II		
2.1	Speech	
2.2	Saliva and drooling	
2.3	Chewing and swallowing	
2.4	Eating tasks	
2.5	Dressing	
2.6	Hygiene	
2.7	Handwriting	
2.8	Doing hobbies and other activities	
2.9	Turning in bed	
2.10	Tremor	
2.11	Getting out of bed	
2.12	Walking and balance	
2.13	Freezing	
3a	Is the patient on medication?	☐ No ☐ Yes
3b	Patient's clinical state	☐ Off ☐ On
3c	Is the patient on levodopa?	☐ No ☐ Yes
3.C1	If yes, minutes since last dose:	

Snapshot of UPDRS Part II

Data Citation: Tsanas,Athanasios & Little,Max. (2009). Parkinsons Telemonitoring. UCI Machine Learning Repository. https://doi.org/10.24432/C5ZS3N.

03 Parkinson’s Telemonitoring Dataset

Attribute Information

subject# - Integer that uniquely identifies each subject
age - Subject age
sex - Subject gender '0' - male, '1' - female
test_time - Time since recruitment into the trial. The integer part is the number of days since recruitment.
motor_UPDRS - Clinician's motor UPDRS score, linearly interpolated
total_UPDRS - Clinician's total UPDRS score, linearly interpolated
Jitter(%),Jitter(Abs),Jitter:RAP,Jitter:PPQ5,Jitter:DDP - Several measures of variation in fundamental frequency
Shimmer,Shimmer(dB),Shimmer:APQ3,Shimmer:APQ5,Shimmer:APQ11,Shimmer:DDA - Several measures of variation in amplitude
NHR,HNR - Two measures of ratio of noise to tonal components in the voice
RPDE - A nonlinear dynamical complexity measure
DFA - Signal fractal scaling exponent
PPE - A nonlinear measure of fundamental frequency variation

Data Citation: Tsanas,Athanasios & Little,Max. (2009). Parkinsons Telemonitoring. UCI Machine Learning Repository. https://doi.org/10.24432/C5ZS3N.

04

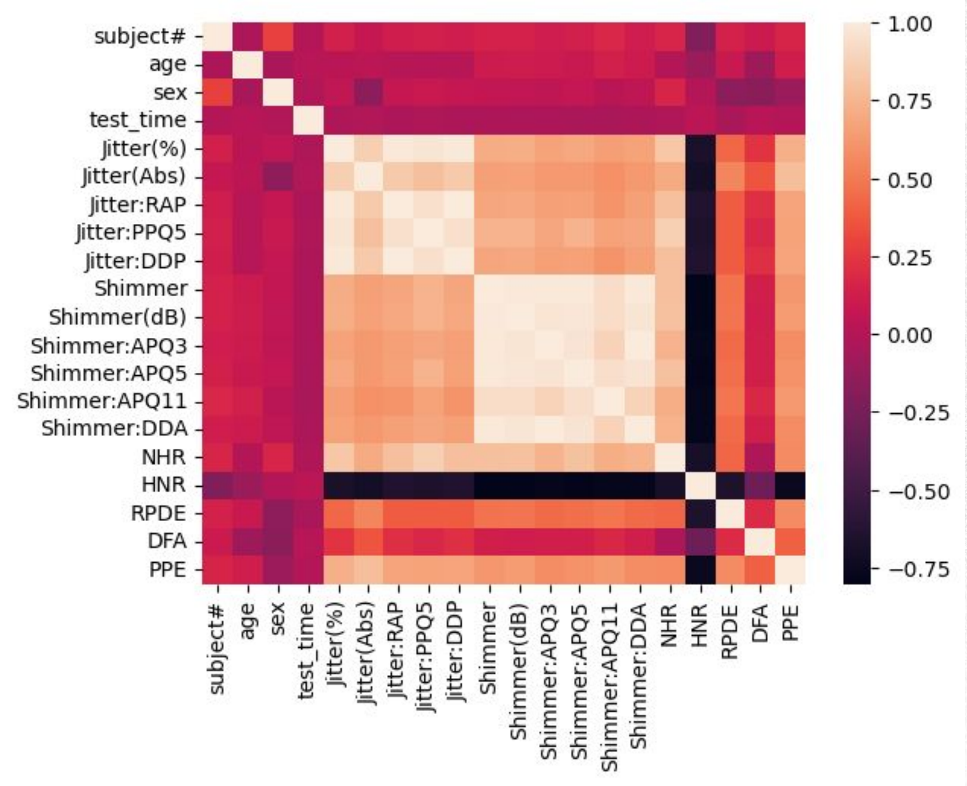
Exploratory Data Analysis

Parkinson’s Telemonitoring Dataset

04 Exploratory Data Analysis

- 5875 instances and 26 features
- No missing values, no categorical data
- Data will be standardized
- Many redundant features
- No text features
- Target is a continuous variable: requires regression methods

- Jitter (frequency) and shimmer (amplitude) features are highly correlated with each other - Jitter (%) and Shimmer were chosen as the main features



05 Modeling

06 Evaluation

Features and Models, Evaluation Metric

05 Model Features

- All models used the same features
  - Subject #, age, sex, test_time, Jitter (%), Shimmer, NHR, HNR, RPDE, DFA, PPE
- Data was split 75% training, 25% test
- Random state 42
- XGBoost default parameters

06 XGBoost Performed Best on Training Dataset

1. Linear Regression
   R2 score = 14%
2. Ridge Regression
   R2 score = 9.6%
3. Random Forest
   R2 score = 96.4%
4. **XGBoost**
   R2 score = 99.5%

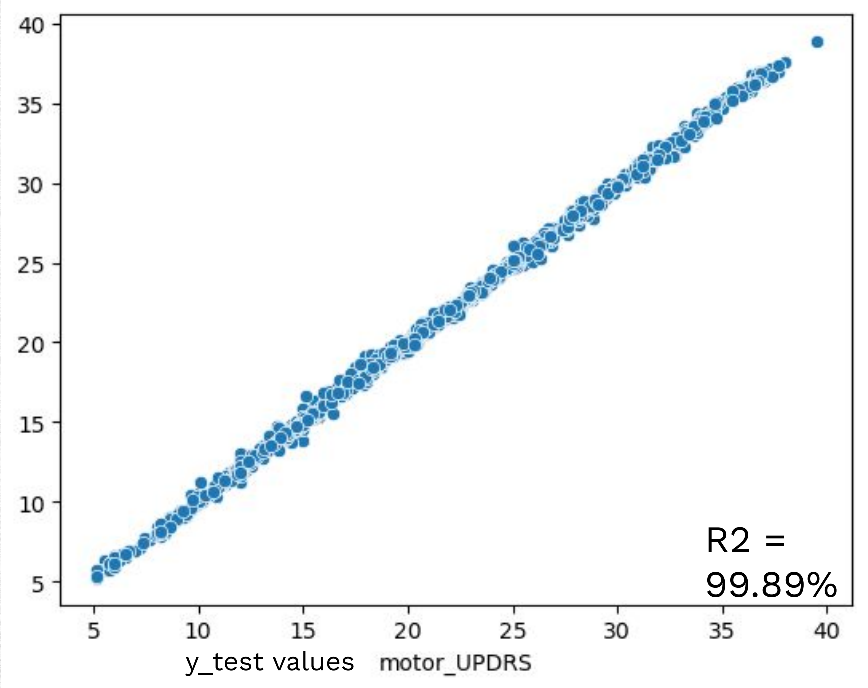
Evaluation Metric: R2 scores were used across all models since they are regression problems

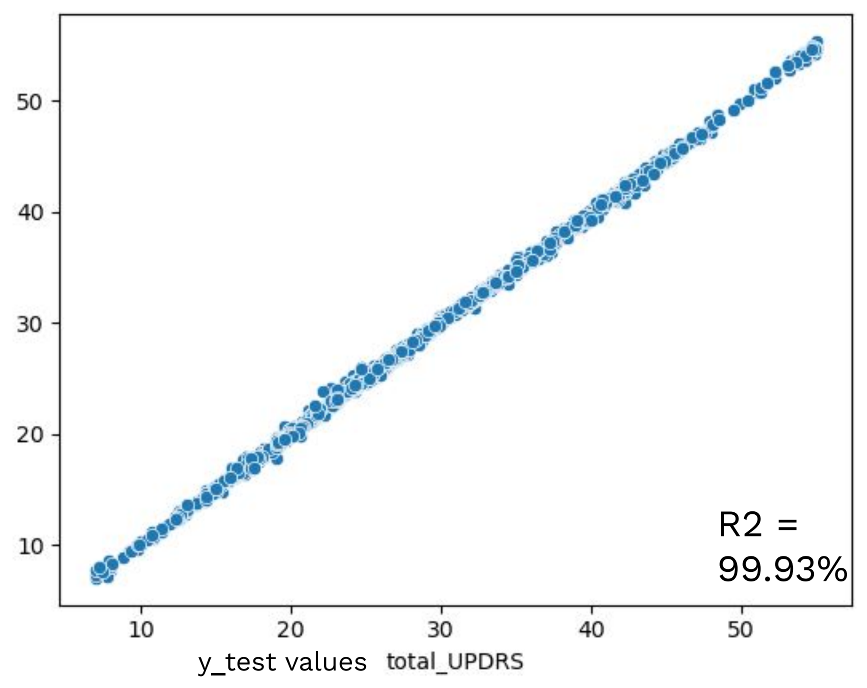
07

Results

XGBoost on Test Data

07 XGBoost Performed Well on Test Data for Both Target Variables





Mean R2 score of 10 fold cross validation = 89.33%

Standard Deviation = 0.02

Mean R2 score of 10 fold cross validation = 91.27%

Standard Deviation = 0.27

08

Conclusion

Prediction and Next Steps

08 We Can Predict PD Severity & Progression Using XGBoost

- Based on our models, we are able to predict disease severity & progression with 91% accuracy in early PD patients using their total UPDRS scores and with fewer attributes than called for in the study.
- Since jitter and shimmer variables are highly dependent on each other, one variable can be assigned to each and still predict progression accurately.
- The total UPDRS score model is more accurate compared to the motor UPDRS scores, indicating that vocal data may be more impactful to more than just the motor section of the UPDRS questionnaire.

08 Next Steps..

- Moving forward, I would like to assess other metrics to evaluate PD progression, and test the other versions of the Jitter and Shimmer parameters.
- These results are exciting news, since documentation of disease progression is important for patient care, and any improvement upon PD symptoms can be helpful.