



Predicting Parkinson's Disease Progression

By: Samuel Flusche, Ruben Valdovinos, Eric Vandament

Table of Contents

Background Information & Kaggle

About our Data


Submissions on Kaggle

Findings

Moving Forward



Background Information



What is Parkinson's Disease?

AMP-PD

"Parkinson's disease (PD) is a chronic and progressive neurological disease that is marked by tremors in the resting muscles, rigidity, slowness of movement, impaired balance and a shuffling gait. In addition, many people with PD develop non-motor symptoms such as behavioral changes and cognitive impairment."

Currently there is no cure, however there are treatments to relieve symptoms.



What is a Kaggle Competition?

- Competitions are hosted on Kaggle.com
- Anyone can host a competition, in our case AMP-PD is hosting ours
- The goal is to make predictions using models
- Predictions must be formatted in specific ways
- Scores are graded by host
- Kaggle updates a live leaderboard with the the entries while the competition is ongoing

About our Data

A photograph showing an elderly person's hands holding a small yellow and blue ball. A younger person's hands are visible, one resting on the elderly person's wrist and the other near the ball, suggesting assistance or a physical therapy exercise. The elderly person is wearing a beige cable-knit sweater and dark blue jeans. The younger person is wearing a grey long-sleeved shirt. The background is a plain, light-colored wall.

About our Competition

**Hosted by Accelerating Medicines
Partnership® Parkinson's Disease
(AMP®PD)**

- Public-Private partnership made to help study Parkinson's Disease
- Goal is predict the severity of one Parkinson's throughout its progression
- Uses real-world data from thousands of patients



About our Competition

Hosted by AMP-PD

- Teams: 1,225
- Competitors: 1,347
- Entries: 17,745
- Max team size is 5
- Limited to 5 submissions per day

Variables

- **Updrs** - Rating scale for Parkinson's Disease symptoms, range 1-4, 4 being worst
- **Peptide** - Sequence of amino acids within the peptide
- **PeptideAbundance** - Frequency of the amino acid in the sample
- **UniProt** - The UniProt ID code for the associated protein
- **NPX** - Normalized protein expression
- **upd23b_clinica_state_on_medication** - Was the patient taking medication
- **visit_month** - month patient came on visit

A close-up photograph showing a young person's hand gently holding an elderly person's hand. The young hand is on the left, with smooth skin and short, clean nails. The elderly hand is on the right, with wrinkled skin and visible age spots. The background is blurred, showing hints of a hospital or care facility setting with blue and white colors. The text "Submissions on Kaggle" is overlaid in white, bold font across the center of the image.

Submissions on Kaggle

Submitting findings

- Each competition has a different way of submitting findings
- need to estimate both their UPDRS scores for that visit and predict their scores for any potential visits 6, 12, and 24 months later

A prediction...	# rating	# group_key
3342_0_updrs_1_plus_0_months	0	0
3342_0_updrs_1_plus_6_months	0	0
3342_0_updrs_1_plus_12_months	0	0
3342_0_updrs_1_plus_24_months	0	0
3342_0_updrs_2_plus_0_months	0	0
3342_0_updrs_2_plus_6_months	0	0
3342_0_updrs_2_plus_12_months	0	0
3342_0_updrs_2_plus_24_months	0	0
3342_0_updrs_3_plus_0_months	0	0
3342_0_updrs_3_plus_6_months	0	0

```
count = 0
for (test, test_peptides, test_proteins, sample_submission) in iter_test:

    targets = sample_submission.prediction_id.str.split('_').str[2:4].apply(lambda x: '_'.join(x))
    if count == 0:
        print(targets)
    print(10* '---sample submission ---')
    print(sample_submission.head())
    print(10* '-----')

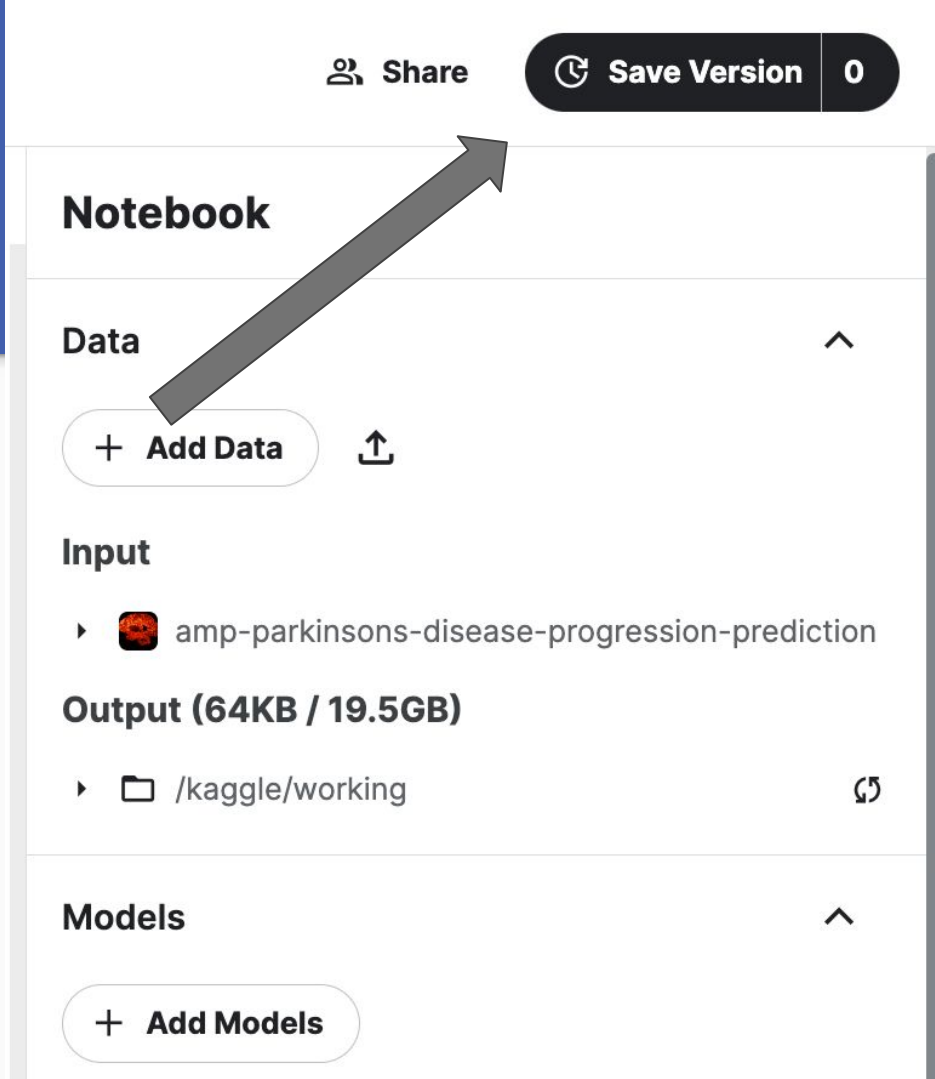
    sample_submission['rating'] = targets.map(mean_dict)
    env.predict(sample_submission)

    count += 1
```

How that data frame is formatted

Submission

- First step after code is done
- Click save version




The screenshot displays the top right corner of a Kaggle Notebook interface. At the top, there are two buttons: 'Share' with a user icon and 'Save Version' with a clock icon and a counter '0'. Below these, the 'Notebook' section is visible. Under the 'Data' section, there is a '+ Add Data' button and an upload icon. Below the 'Input' section, there is a list item with a red icon and the text 'amp-parkinsons-disease-progression-prediction'. Below the 'Output (64KB / 19.5GB)' section, there is a list item with a folder icon and the text '/kaggle/working'. At the bottom, the 'Models' section is visible with a '+ Add Models' button. A large grey arrow points from the 'Add Data' button towards the 'Save Version' button.


Share Save Version 0

Notebook



Data

+ Add Data 

Input

▸  amp-parkinsons-disease-progression-prediction

Output (64KB / 19.5GB)

▸  /kaggle/working 

Models

+ Add Models

Submission

- Change name
- Save or save and commit
- Advance settings allow you to pick how you want to run the code
- Saving output

×

Save version

VERSION NAME

Version 1

9 / 50

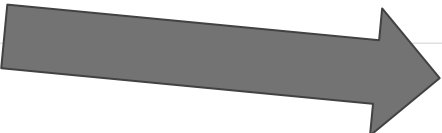
VERSION TYPE

✓ Save & Run All (Commit)

▼

Run a fresh copy of your notebook and save the output


▼ **Advanced Settings**



Save

Submission


- Submission will run on bottom left
- Wait until it finishes running



Submission trend
Version #1
Running: just now

just now


...



Submission trend
Interactive Session
Running: 3 minutes

3 minutes


...



Linear model RV
Interactive Session
Cancelled

41 minutes

...



2 Active Events

▼

+ Create

Home

Competitions

Datasets

Models

Code

Discussions

Learn

More

Your Work

RECENTLY VIEWED

Submission trend

PDP Almost Only Tre...

AMP-Parkinson's Dis...

Parkinson's Disease Pr...

Using Feature selectio...

RECENTLY EDITED

Using Feature selectio...

View Active Events

Featured Code Competition

AMP®-Parkinson's Disease Progression Prediction

Use protein and peptide data measurements from Parkinson's Disease patients to predict progression of the disease.

\$60,000

Prize Money

AMP AMP-PD · 1,546 teams · 16 days to go (9 days to go until merger deadline)

Overview Data Code Discussion Leaderboard Rules Team

Submissions

Submit Predictions

Your Team

Everyone that competes in a Competition does so as a team - even if you're competing by yourself. [Learn more.](#)

General

TEAM NAME

St. Edward's University Data

This name will appear on your team's leaderboard position.

Team Members

Your team can have a maximum of 5 members.

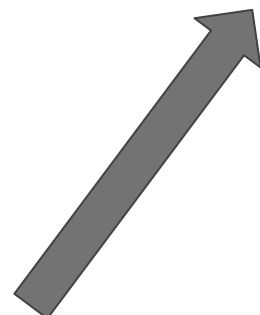


Eric Vandament

Team Leader



Samuel Flusche



Come back to here

Submission

- Choose which notebook
- Pick correct version
- Pick correct output file

✕ Submit to Competition

Notebook File Upload



AMP®-Parkinson's Disease Progression Prediction

You have 3 submissions remaining today. This resets in 19 hours.

SELECT NOTEBOOK

Submission trend

NOTEBOOK VERSION

Version 1

OUTPUT FILE

submission.csv

In this competition, we will privately re-run your selected Notebook Version with a hidden test set substituted into the competition dataset. We then extract your chosen Output File from the re-run and use that to determine your score.

DESCRIPTION

Enter a description

0 / 500



Submit



Submissions

Select up to 2 submissions that will count towards your final leaderboard score. If less than 2 are selected, Kaggle will automatically select from your best scoring submissions. [Learn More](#)

0/2

☒ Auto-selection candidates ?

All

Successful

Selected

Errors

Recent ▾

Submission and Description

Public Score ⓘ

Select



Submission trend - Version 1

Succeeded · Ruben Valdovinos · 2m ago



56.1



Linear model RV - Version 1

Succeeded · Ruben Valdovinos · 5d ago

56.2



submission.csv

Succeeded · Ruben Valdovinos · 7d ago

60.3



[LB: 57.2] Ensemble: RandomForest, SVR, Linear - Version 1

Succeeded · Samuel Flusche · 16d ago · Test Submission

57.2



Basic Regressions (Linear, Bayesian ...) - Version 1

Succeeded · Ruben Valdovinos · 16d ago · first try

56.5



This is what the final submission will look like

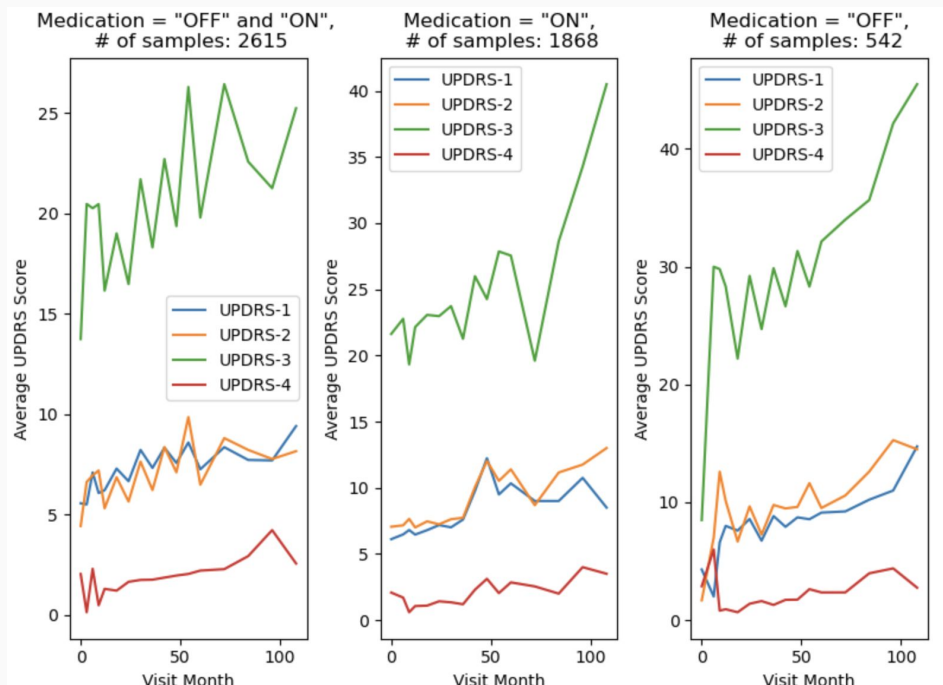
A close-up photograph of an elderly person's arm and hand. The skin is wrinkled and aged. The person is wearing a maroon short-sleeved shirt and dark brown trousers. They are holding a light-colored wooden cane with their right hand. A gold ring is visible on the ring finger. The background is a blurred green, suggesting an outdoor setting.

Findings

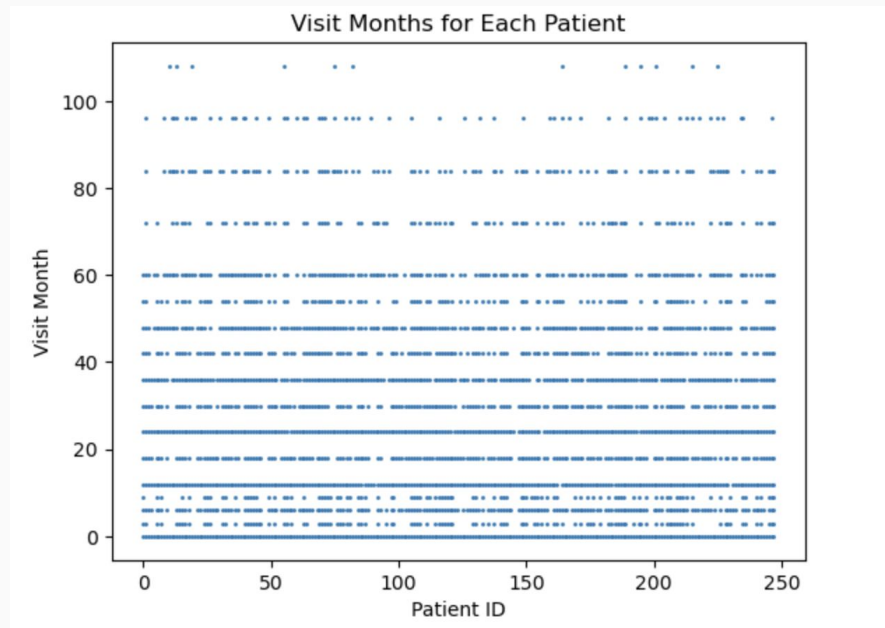
Analysis of Others

- Top scorers have not made their scores public
- Successful models we found almost 'cheat'
- Many of the public models don't even touch the protein and peptide data
- This means that a lot of the data focus exclusively on the typical trends of parkinson's progression and not the corresponding medical factors of the patient.

Our Analysis



Our Analysis



Our Analysis

- Wanted to try to utilize the protein and peptide data in models
- A few ways to do this, always involves combining all dataframes
- We combined protein and peptide data into the dataframe and ran the same regression,
 - Gave a score of 56.9
- Experimented with ensembling, only lowered the score

Analysis Challenges and Aids

- Failed submissions still count as submissions
- Complicated models can be hard to submit correctly
- Data prep is easy, most of what you need is already out there

Our Results

Our Best Score: 56.1 (from SGD regressor code)

Place: 244

Overall Best Score: 53.6

- Top 22% of entries
- Tied with 143 other people
- Top 25 only ones below score of 56.0

244

St. Edward's University Data



56.1

6

1h



Your Best Entry!

Your submission scored , which is not an improvement of your previous score. Keep trying!

A close-up photograph of an elderly person's hands, which are wrinkled and aged, clasped together in a prayer-like position. The hands are resting on a light blue fabric surface. The person is wearing a maroon long-sleeved shirt and a silver bracelet on their left wrist. The background is slightly blurred, showing a wicker chair and a yellow object.

Moving Forward

Improving Models

- Successfully utilizing protein/peptide will require more complicated models
- Some people have started creating deep learning models or models better utilizing the time series
- The best models usually have a combination of different techniques

Thank you