Group 4

We propose to develop a machine learning model that uses protein and peptide measurements from patients to predict Parkinson's disease progression. Considering Parkinson's disease is a progressive neurodegenerative disorder that affects millions of people worldwide, this is an important issue. Developing better treatment options and improving patient quality of life can be improved by predicting disease progression.

This project uses the Parkinson's Progression Markers Initiative (PPMI) dataset, which contains measurements of proteins and peptides from several hundred patients. The dataset includes mass spectrometry readings from cerebrospinal fluid (CSF) samples collected from Parkinson's disease patients and control subjects of similar age. A machine learning model can be trained on the dataset.

To develop the machine learning model, we will use a variety of algorithms, including neural networks, decision trees, and random forests. We will use Python as the software to implement the machine learning algorithms because of its popularity and the availability of several powerful libraries for machine learning, such as Scikit-learn, Pytorch, and TensorFlow.

To obtain sufficient background on applying the chosen algorithm to the specific problem, we will review research papers and books on machine learning, particularly those that discuss the application of machine learning to medical data. We will also refer to online resources, such as blogs and forums, to seek advice from experts in the field. e.g.: [Towards Data Science blog](https://towardsdatascience.com/choosing-the-correct-error-metric-mape-vs-smape-5328dec53fac)**,** [Machine learning for medical diagnosis: history, state of the art and perspective by Guojie Song, Xuelong Li, and Xiaohua Tong,](https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-1023-5) [Parkinson’s Disease: Causes, Symptoms, and Treatments](https://www.mayoclinic.org/diseases-conditions/parkinsons-disease/symptoms-causes/syc-20376055)

We will use a number of metrics, such as accuracy, RMSE, and [SMAPE](https://en.wikipedia.org/wiki/Symmetric_mean_absolute_percentage_error) to assess the performance of the machine learning model. To make sure that the model is not overfitting the training data, we will also employ strategies like cross-validation.

The project timeline will be as follows:

 Week 1: Data cleaning and pre-processing

Week 2: Feature selection and engineering

Week 3: Model selection and training

Week 4: Model evaluation and optimization &

               Finalize the model and write the report.

In summary, this project aims to develop a machine learning model to predict the progression of Parkinson's disease using protein and peptide data measurements from patients. The proposed approach includes using various machine learning algorithms, including neural networks, decision trees, and random forests, and Python as the implementation software. The project will be evaluated using several metrics, and the timeline for completing the project is four weeks.