

# Project Proposal

Javier Echavarren, Victor Boveda, Carlos Castillo

28 Feb, 2023

## Project title

The title of our project is: “Parkinson’s Disease Progression Prediction”. It is an active competition from Kaggle.

## Team Members

We are three students from the CS584 course:

- Javier Echavarren Suarez A20521368
- Victor Boveda Prado A20521372
- Carlos Castillo Martinez A20520708

## Description of the problem

Parkinson’s disease (PD) is a disabling brain disorder that affects movements, cognition, sleep, and other normal functions. Unfortunately, there is no current cure—and the disease worsens over time. It’s estimated that by 2037, 1.6 million people in the U.S. will have Parkinson’s disease, at an economic cost approaching \$80 billion.

## State of the art

Research indicates that protein or peptide abnormalities play a key role in the onset and worsening of this disease. Gaining a better understanding of this—with the help of data science—could provide important clues for the development of new pharmacotherapies to slow the progression or cure Parkinson’s disease.

Current efforts have resulted in complex clinical and neurobiological data on over 10,000 subjects for broad sharing with the research community. A number of important findings have been published using this data, but clear biomarkers or cures are still lacking.

## Preliminary plan

Competition host, the Accelerating Medicines Partnership® Parkinson’s Disease (AMP®PD), is a public-private partnership between government, industry, and nonprofits that is managed through the Foundation of the National Institutes of Health (FNIH). The Partnership created the AMP PD Knowledge Platform, which includes a deep molecular characterization and longitudinal clinical profiling of Parkinson’s disease patients, with the goal of identifying and validating diagnostic, prognostic, and/or disease progression biomarkers for Parkinson’s disease.

The goal of this competition is to predict the course of Parkinson’s disease (PD) using protein abundance data. The complete set of proteins involved in PD remains an open research question and any proteins that have

predictive value are likely worth investigating further. The core of the dataset consists of protein abundance values derived from mass spectrometry readings of cerebrospinal fluid (CSF) samples gathered from several hundred patients. Each patient contributed several samples over the course of multiple years while they also took assessments of PD severity.