CS109 Group 10 - Scope of Work

Guilherme Braz, James May, Shristi Pandey, Zach Werkhoven

Group Number: 10

Have you met/communicated with your fellow teammates?

We have met twice as a group via video conference and have setup a slack team to handle our day-to-day communication.

Have you met/communicated with your assigned TF?

We have been in contact with our TF (Justin Lee) by email and held a video conference with him over the weekend.

Has your team formulated a well-defined question to address in your project, based on the project description and references?

We have defined an intermediate goal of conducting exploratory data analysis to identify features from that ADNI database that will allow us to develop a model for predicting patient prognosis. Our goal will become more well-defined once we have conducted additional research and have a better grasp on the structure and availability of the many possible features in the data set. An important step toward this goal will be defining our response variable. Potential response variables include categorical features such as diagnosis (DXSUM or DXCHANGE) or continuous features such as the Alzheimer's Disease Assessment Score (ADAS). We will focus on these response variable candidates in greater detail during preliminary EDA.

Briefly describe your team's plans for work to be completed by Nov 28 (milestone three). Please assign specific tasks to team members and deadlines for when these tasks are to be completed.

Please see the below (Table 1) for a detailed description of our tentative work plan.

Task	Description	Deliverable	Due Date	Assigned
ADNI database research	Initial study of types of data and formats available, get human readable names/descriptions for various metrics. Identify response variable.		10/26	Zach, Guilherme
ADNI study background info	Research the project itself to figure out how study was conducted, what measures were collected, when how etc.		10/26	James, Shristi
Aggregate data into single file	Get all the data into a single tabular format (eg. one row = one patient) if possible. We may to consider different data types that can't be combined in this way. Preferrably we will start with a large feature space.	ADNI_unformatted_all_data.csv + code	11/2	All members
Format data	Standardized format for all missing values, ensure proper data types for metrics	code to format dataframe	11/2	Zach, Guilherme
Conduct preliminary EDA	Visualize data and fit/validate models and compare model types. Maybe use some early models to perform feature selection to narrow the scope of our study to most interesting predictor. Identify colinear predictors. Try regularization/dimensionality for dealing with colinearity. Settle on specific goal for project.	Individual Jupyter notebooks and visualizations	11/9	All members
Post EDA group meeting	Discuss/compare results. Narrow down set of features to the final subset.		11/9	All members
Format categorical data	Format categorical data with appropriate/interpretable names.	code to rename categories	11/11	Shristi
Individual EDA	Visualize data, fit/validate models.	Individual Jupyter notebooks and visualizations	11/19	All members
Group meeting for revised project statement	Meet to discuss compare models. Select a model for the final report. Identify key visualizations for revised project statment.		11/19	All members
Data summary for project statement	Write summary of what the data is, how it was formatted/processed, and intial EDA	~1 page text on the data	11/26	Zach
Data/EDA summary visualizations	Create visualizations for data summary portion of report	visualizations of important EDA findings	11/26	Guilherme
Model summary	Write summary of model including type of model, how it was trained/validated/tested, and key findings (eg. best predictors)	~1 page text summarizing how model was chosen, key findings of model, and performance	11/26	James
Model visualizations	Create visualizations for model summary portion of report	visualizations model: comparisons to other model types, coefficients, performance etc	11/26	Shristi
Combine materials into written report	Combine the above materials and submit to Canvas	PDF combining visualizations and text	11/28	Shristi
Format visualization code into single notebook for submission	Format code and submit to Canvas	*.ipynb with code for visualizations and markdown comments explaining process*	11/28	James
Create website	Create website to summarize results of the project (will further subdivide this category in the future)	a website	12/12	All members

Table 1 - Tentative description of tasks, due dates, and division of labor for Group 10 final project