The Elections Challenge

Congratulations! Upon a successful completion of this exercise, you will be a valid candidate to work as an analyst for your favorite pollster, predict the outcome of the elections, and participate in the great democratic celebration in ways you did not even imagine.

Your mission is to create a solution that will help predict the elections results and to offer the various parties data-driven services that will assist their campaigning.

The polls data available for analysis (i.e. the "training set") is a dataset of 10K voter. There are 36 characteristics per voter, which represent demographics, various polls results and more. The dataset is labeled with the party each person intend to vote for. This data has been collected from various locations, all around the nation, and it can be considered as an unbiased representative of the entire population. Please note that some of the voters didn't respond to all the questions. Hence, there are missing values throughout the data.

The assignments that will be handed along the course will walk you, step by step, through tasks such as data preparation, analysis, modeling, predictions, conclusions and suggestions. The following are examples for such tasks:

- Per party: Identify the key factors influencing most of its voters
- Predict which party would win the majority of votes
- Predict to the division of votes between the various parties
- On the Election Day, each party would like to suggest transportation services for its voters. Provide each party with a list of its most probable voters
- Build a steady coalition (one that its voters are most similar) – show why this coalition will be more stable than other coalitions.
  - Terms for a "steady coalition"
    - Over 51% of the votes, relatively homogeneous with respect to the participating parties, and very much different from the opposition
- Identify the factor (voters' characteristic) which by manipulating you are most likely to change which party will win the election
- What would be a group of factors that by being able to manipulate, it will allow you to create a strong coalition per your choice?
  - see the definition of a steady coalition

Whenever applicable, the value of your work will be evaluated on a new, unlabeled, testing data, for which you will need to provide your predictions and the results of your analysis. At other times, we will use privileged information (but one that can be validated … ☺)

Dataset details:

- A labeled data set
  - 10K rows
  - 38 columns
    - 1st column is the label ("vote"), columns 2 to 38 are the features
      - Label (the "Vote" field)
        - 10 parties (classes): Yellows, Blues, Reds, Greens, Browns, Purples, Pinks, Greys, Whites, Oranges
      - There are both nominal and numeric features
  - Data set includes missing values

The assignments will be provided throughout the course … stay tuned!

… and don't forget to vote yourself and make a difference