CA675 – Assignment 2

Mid-Way Report

# Project Idea:

Analyzing US primary election results of 2016 to develop an election strategy that can either help Democrat or Republican candidate for the general election.

# Project Description:

2016 primary election dataset was collected from Kaggle (https://www.kaggle.com/benhamner/2016-us-election) .

This dataset contains fraction of votes for each candidates of the democrats and the republican in each county. This helps to understand the win margins of each candidate in that region.

There is another dataset which has the information of the detailed demographic distribution of the United States. This will help us in understanding the alignment towards a candidate by a particular section of the society based on gender, age, race, ethnicity etc.

By merging these two datasets we will be able to determine the various factors that helped a candidate to win the primary elections. This can help to develop an election strategy for the general elections.

# Technologies Planned to Use:

* Hadoop HDFS:

For distributed file storage.

* PySpark + Mlib:

For analytics.

* Tableau:

For visualization.

# Task Distribution:

|  |  |  |
| --- | --- | --- |
| Team Member | Role | Task |
| Akash Bhargava(18210613) | Developer | Dataset processing, Data Analysis, Data Loading |
| Geetha Rapalli(18210710) | Developer | Data Analysis, Env. Setup, Visualization |
| Harshali Patil(18212797) | Developer | Dataset processing, Data Analysis, Visualization |
| Ramneesh Vaidyambath(18210732) | Developer | Data Analysis, Env. Setup, Data Loading |
| Tajinderpal Singh Gill(16213078) | Developer | Dataset processing, Data Analysis, Report Creation |