

CS253 A3

Yash Pratap Singh

April 2024

1 Introduction

This report presents an analysis of election data using machine learning techniques. The naive-Bayes model was used to predict the education level of candidates. The report outlines data preprocessing steps, the models used, and their hyperparameters.

* [GitHub](#) (Link to Github Repo)

2 Data Preprocessing

1. Eliminating Unnecessary Features: The 'ID' and 'Candidate' columns were removed. The 'Constituency' column was also excluded as it didn't significantly enhance the model's performance.
2. Normalizing Features: All features were normalized using MinMaxScaler to ensure uniform contribution to the model.

3 Models Used

1. **Naive-Bayes:**
 - No hyperparameters to tune.
 - Performance: Precision, Recall and F1-score were calculated for each education level.

3.1 Classification Results

3.1.1 Naive-Bayes

Table 1: Classification Report

Class	Precision	Recall	F1-Score	Support
10th Pass	0.04	0.02	0.03	54
12th Pass	0.17	0.11	0.14	79
5th Pass	0.00	0.00	0.00	1
8th Pass	0.00	0.00	0.00	27
Doctorate	0.00	0.00	0.00	9
Graduate	0.23	0.35	0.28	130
Graduate Professional	0.18	0.10	0.13	90
Literate	0.00	0.00	0.00	1
Others	0.00	0.00	0.00	8
Post Graduate	0.26	0.40	0.31	116
Accuracy	0.22			
Macro Avg	0.09	0.10	0.09	515
Weighted Avg	0.18	0.22	0.19	515

4 Hyper parameters Selection

1. Training-Validation Split: A split of 0.8 for training and 0.2 for validation was chosen along with a random state of 42. This split ratio was found to yield the best results through experimentation.
2. Naive-Bayes: No hyperparameters to tune. It performed better than Random Forest, likely due to its simplicity and assumption of feature independence.

5 LeaderBoard Rankings

- F1 score : 0.25568
- Public LeadeardBoard Rank : 44
- Private LeadeardBoard Rank : 58

6 Plots

6.1 Percentage Distribution of Parties with Candidates Having the Most Criminal Records

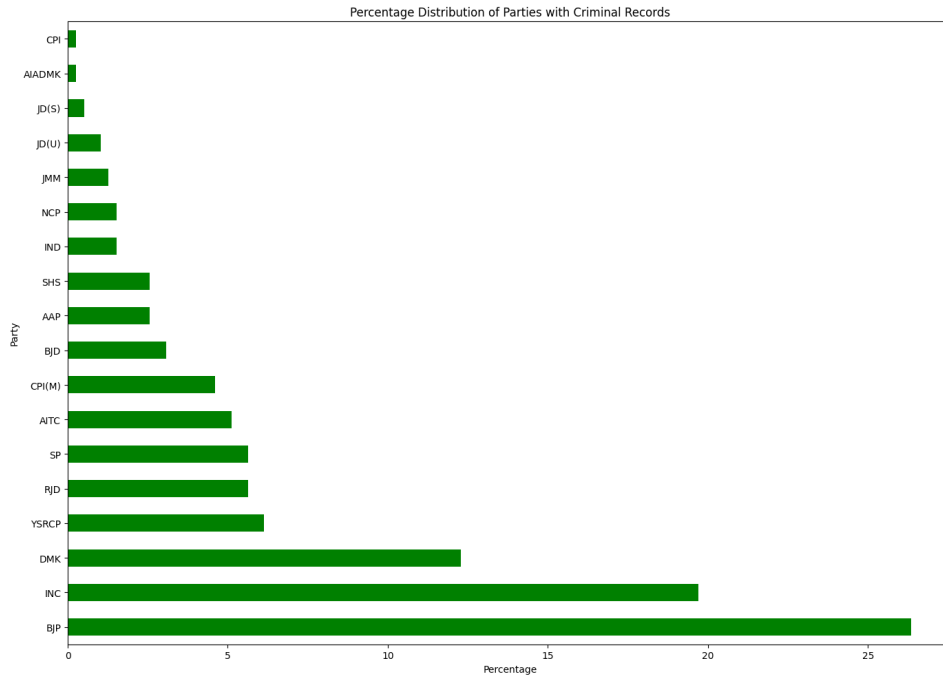


Figure 1: Percentage Distribution of Parties with Criminal Records

6.2 Percentage Distribution of Parties with the Most Wealthy Candidates

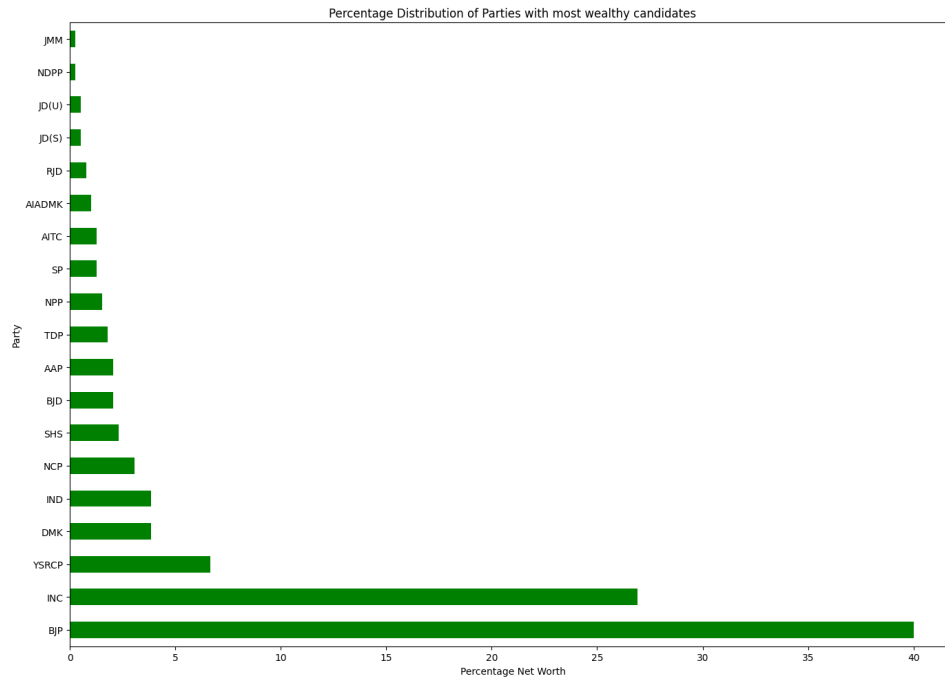


Figure 2: Percentage Distribution of Parties with Wealthy Candidates