Machine Learning Model Report

Mohammed Anas

April 14, 2024

1 Methodology

1. **Data Preprocessing Steps:**

   - Loaded the dataset from a CSV file.
   - Dropped unnecessary columns ('ID', 'Candidate', 'Constituency').
   - Handled missing values by dropping rows with missing values.
   - Encoded categorical variables ('Party', 'state') using LabelEncoder.
   - Converted 'Total Assets' and 'Liabilities' to numerical values.
   - Mapped 'Education' to numerical values using a predefined mapping.

2. **Feature Engineering:** The model performs feature engineering on the 'Party' and 'state' columns by encoding these categorical variables into numerical values using LabelEncoder from sklearn.preprocessing. This is necessary because machine learning algorithms work better with numerical data.

3. **Transformation Used:** The model transforms the 'Total Assets' and 'Liabilities' columns from string values to numerical values. If the value contains 'Crore', it multiplies the numerical part by 100. If the value contains 'Lac', it keeps the numerical part as is. If the value does not contain either 'Crore' or 'Lac', it assigns a value of 0.

2 Experiment Details

This is the case of ordinal targets.

Model	Hyperparameters
HistGradientBoostingClassifier[1]	Default hyperparameters

3   Results

- **Final F1 Score:** Public=0.22091, Private=0.23410.

- **Public and Private Leaderboard Rank:** Public=175, Private=105.

4   References

[1]Histogram-Based Gradient Boosting
[2]Intro to Machine Learning
[3]Machine Learning for Everybody – Full Course