

Predicting the election results using the Historical IEC Election results and an API for real-time data.

Using the machine learning algorithm, Random Forest Classifier to predict the South African electoral results.

INTRO

- In an age of ever-changing political landscapes, the need for accurate and predictive electoral modelling has become increasingly important. The primary objective of the project is to address this by developing an initial predictive model framework using historical IEC electoral data and API technologies.
- This model aims to predict election results, analyse voter behaviour, identify potential trends, and understand the drivers of voter preferences and party performance.
- This initial framework may be used to aid political analysts, campaign managers, and policymakers in making informed decisions. Furthermore, it is designed to serve as a foundation for future developers and analysts, to facilitate the development and enhancement of subsequent iterations.

METHODS

- Given the nature of the historical IEC electoral data, we identified three classifier models: Decision Tree, Random Forest and Gradient Boosting Machines (GBM).
- We selected the best performing classifier model with respect to their predictive abilities.
- Model Performance comparison:

Model	Model Accuracy
Decision Tree Classifier	13%
Random Forest Classifier	29%
Gradient Boosting Machines	19%

Figure 1: Predictive accuracy of the three classifier models considered- **Random Forest Selected**

- Develop an Application Programming Interface (API) to cater for real-time data input.

RESULTS

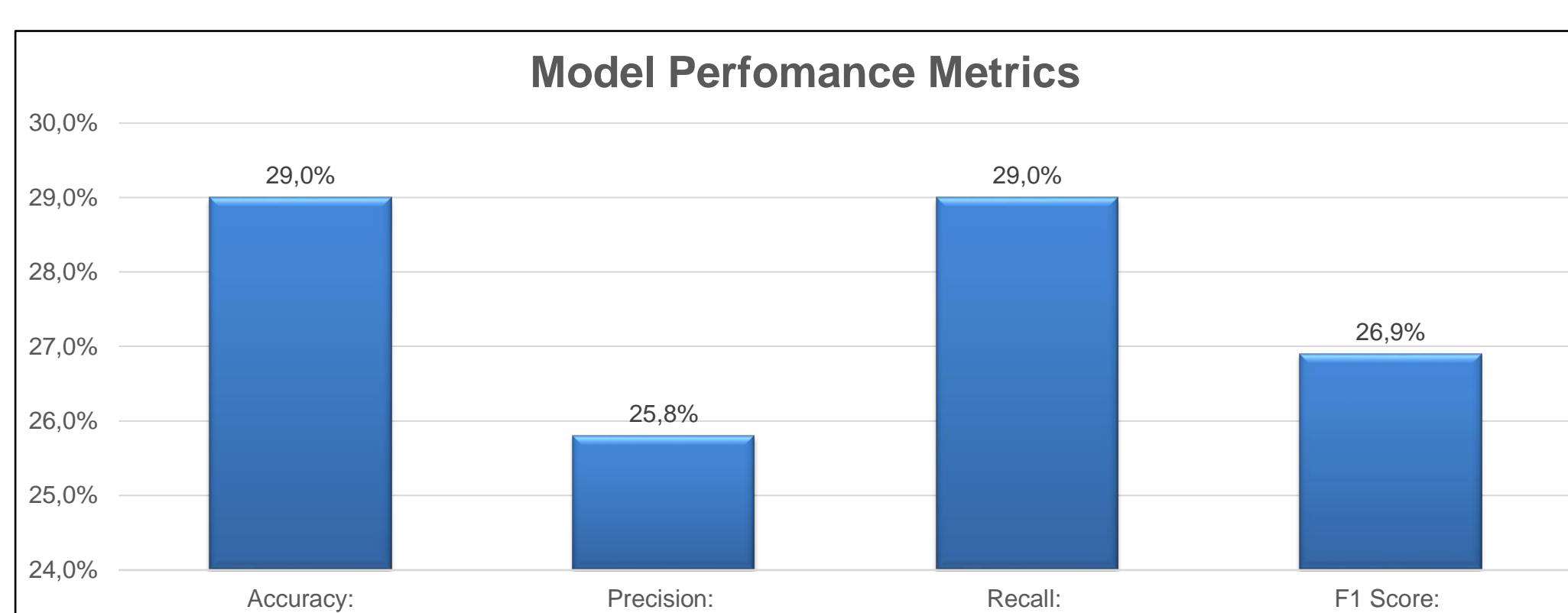


Figure 2: The Random Forest Model Performance Metrics

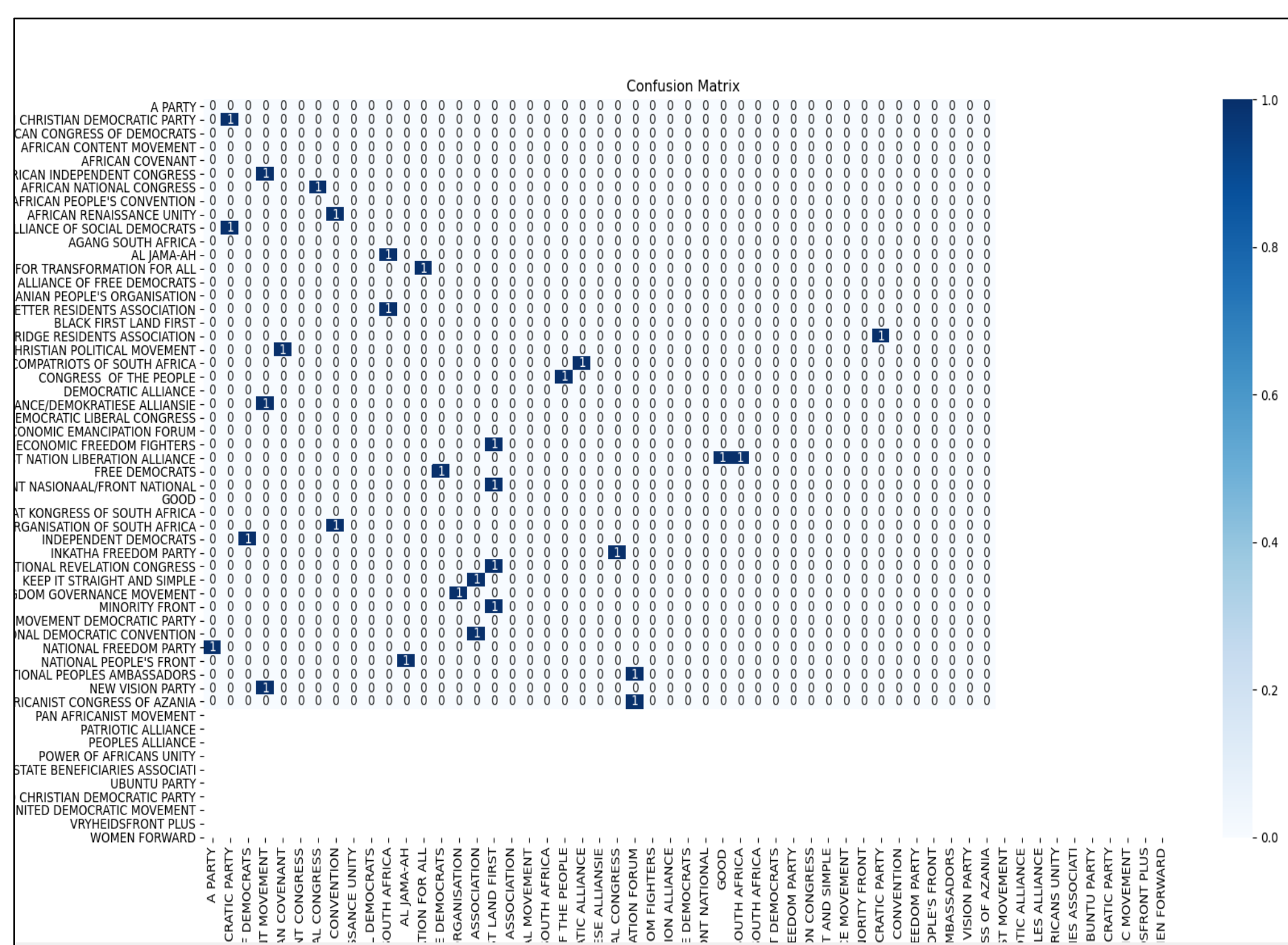


Figure 3: The Random Forest Model Performance - Confusion Matrix Graph.

DISCUSSION

- Model accuracy:** 29% for predicting the winning party in the next election.
- Key constraint:** Limitations of our dataset.
- Performance impact:** Insufficient historical data and incomplete voter information/features (such as voter demographics)
- Other factors:** External influences outside the dataset these include IEC data access limitations and incomplete data due to POPI regulations
- Conclusion:** Although low, predictive value does exist, but accuracy is limited by the data quality and scope.

Additional Visualisations

```
Number of appearances per party:
2019    48
2014    29
2009    26
Name: Election Year, dtype: int64
```

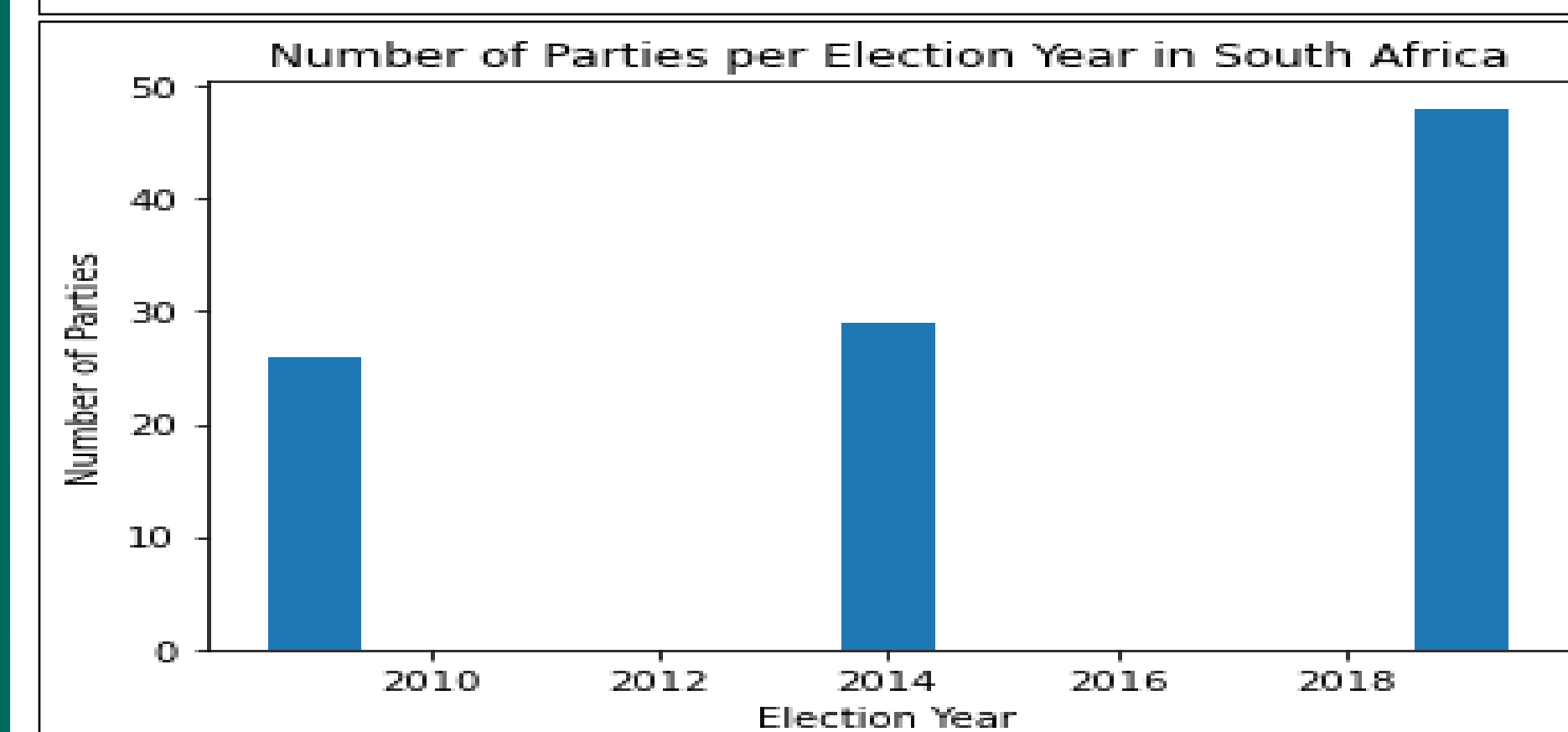


Figure 4: The number of registered parties from 2009 to 2019

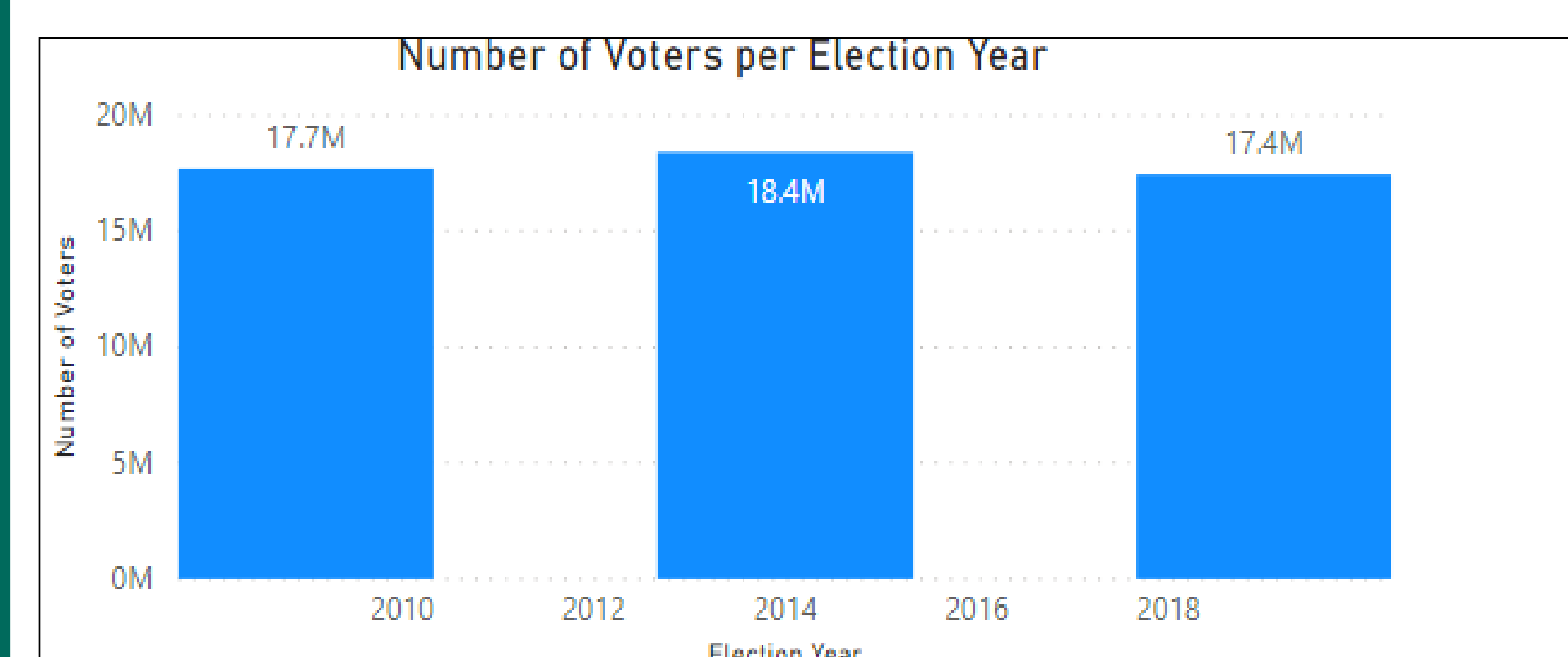


Figure 5: The number of registered voters from 2009 to 2019

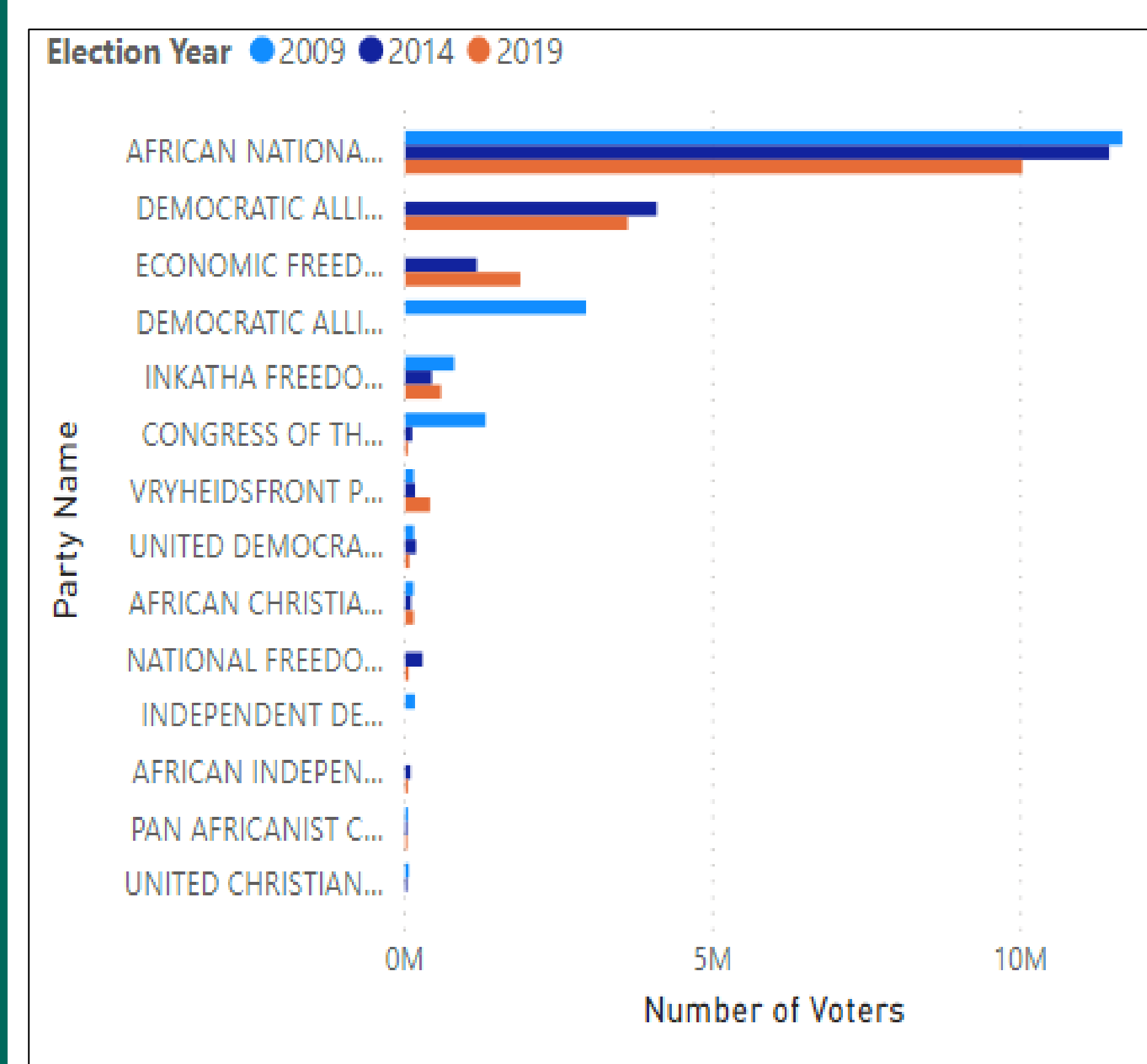


Figure 6: The number of voters VS the party voted for over the period 2009 to 2019

