# Understanding Election Dynamics: A Graph Analysis Approach

PROBLEM

The results of National Assembly elections in Pakistan over the years have been nothing short of a mystery that most do not understand. This is why it is important for analysis to be carried out in a manner that is understandable for a wider audience. Analyzing the election results leveraging graphs make newer trends and analytics come to the surface.

DATASET & DATA MODEL

The dataset for this project was collected from Election Commission of Pakistan (ECP) website which was then divided into four entities and three edges. The data includes all the candidates that competed for a constituency in National Assembly elections from 1993 to 2013. The four entities in focus are candidates, constituencies, political parties, and provinces.
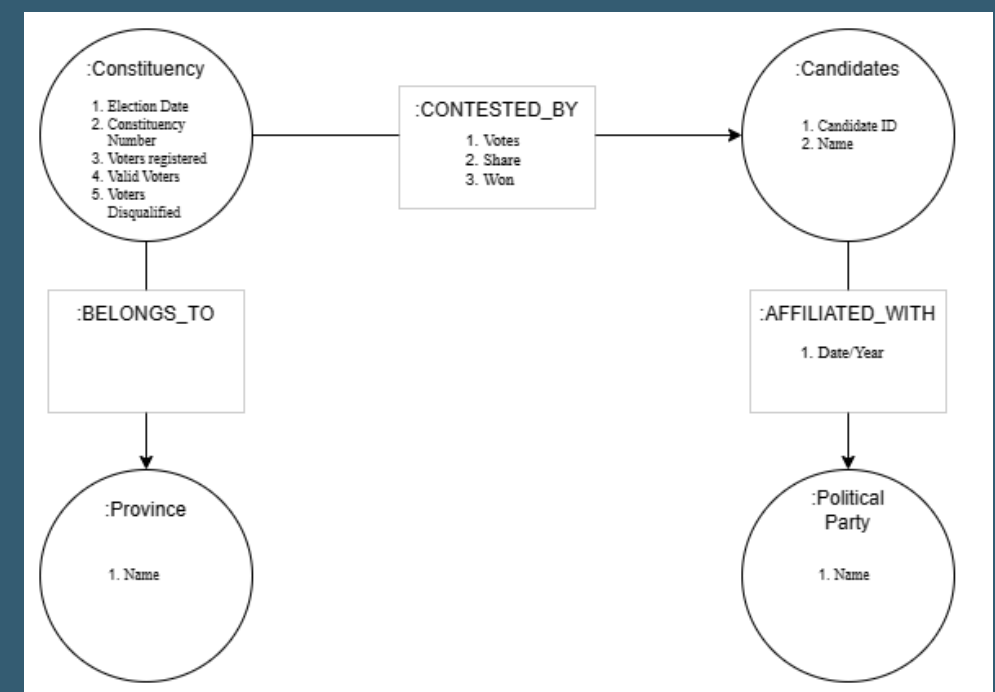

GRAPH ANALYTICS

Analysis for understanding the graph database was carried out. Other than calculating and using the graph properties such as PageRank and centrality metrics, the analysis first took place by trying to answer the following questions:

1. Most dominant party in each election from 1993-2013: Most dominant political party was PMLN winning majority of the seats in two elections out of the five we focused on.
2. Candidates who won in multiple unique election years in the same constituencies: Sardar Ayaz Sadiq from NA-122 has won his designated constituency for the past three national assembly elections.
3. Average voter registration per province over the election years: Sindh has had the highest voter registration.
4. Candidates with the highest votes share each year, highlighting the clearest winners: In elections from 1993-2013, the clearest win was of Sufyan Yousuf from Na-246 accumulating 96.4% of the vote-share.
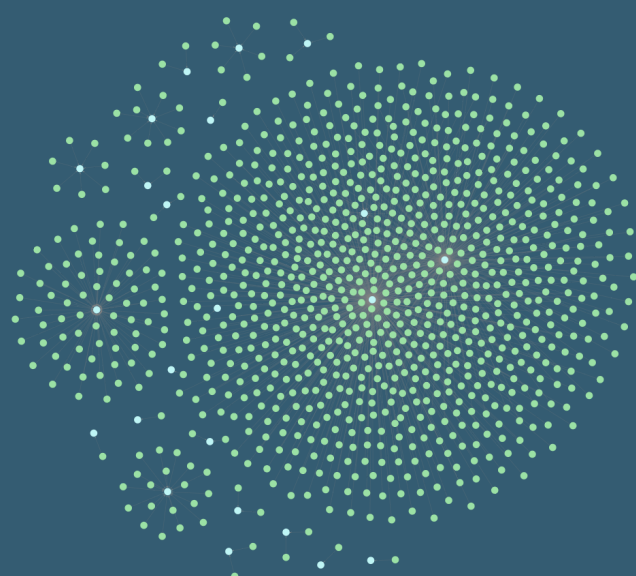
Viz of candidates affiliated with political parties
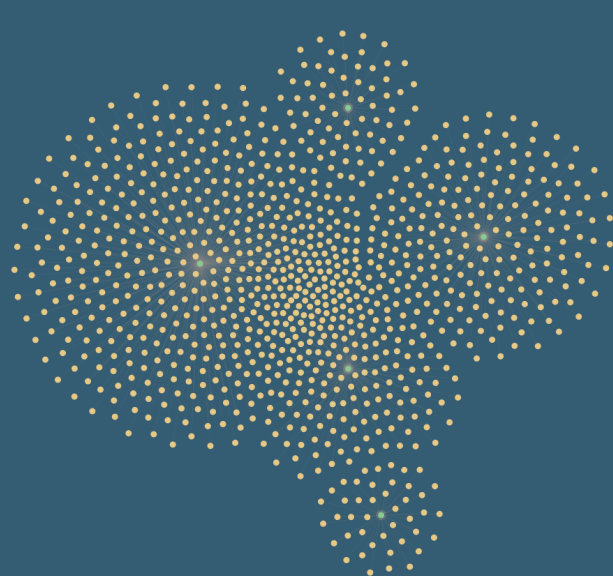
Viz of constituencies belonging to provinces

MACHINE LEARNING

A classification algorithm has been implemented to predict whether a candidate will win or lose from a constituency. To test our model, we trained our model without graph properties such as PageRank and then trained the model with graph properties. The model was trained on constituency, political party, province, and outcome. The following graph and confusion matrix show the result of the model without graph properties.
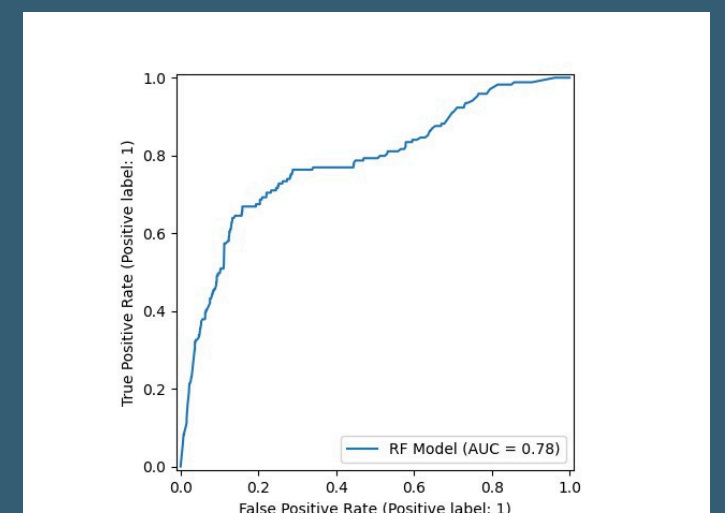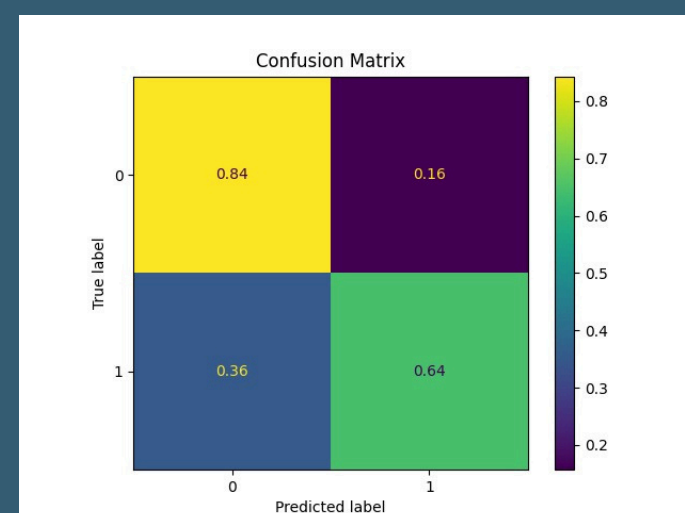
The model was then trained using the same four properties with the addition of the graph property PageRank and an increase in the accuracy was noticed. The increase accuracy hints that graph properties play an important role in achieving better graph machine learning models.
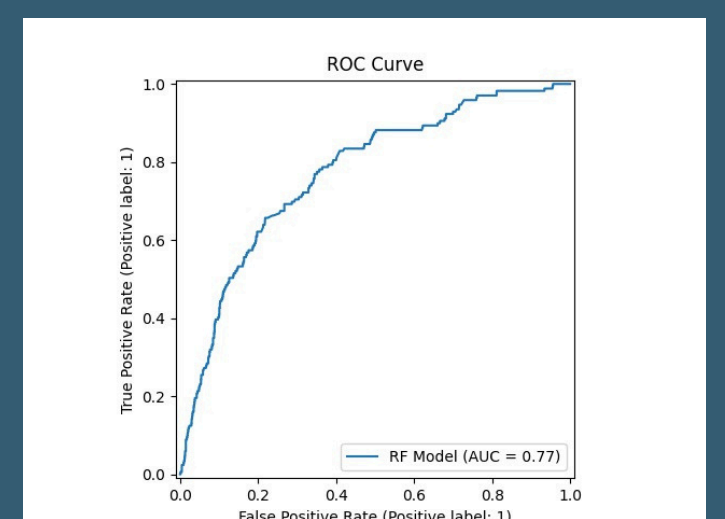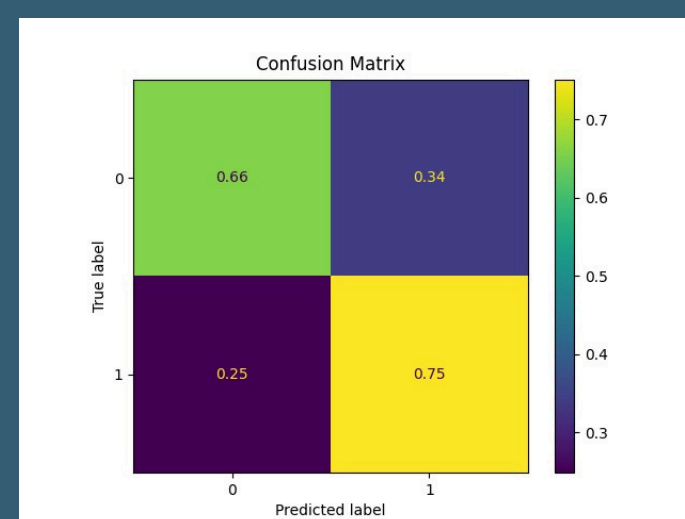

APPLICATION

Application created for the project leverages the Tkinter library of Python to create the frontend of our application. This application is then connected to the database locally through the neo4j library of the graph database.

CONCLUSION & FUTURE WORK

National Assembly elections in Pakistan have a certain mystique around them. Most citizens consume the result based off of news channels rather than official streams such as the Election Commission of Pakistan. What it tends to do is that while most people know the results, they miss out on finer details such as what was the turnout and who has been continuously winning from the same National Assembly seat. Graph Databases also help the users in identifying how close two candidates are in a graph network of the National Assembly which can be used to analyze and predict if there is direct communication between the users. For future work, this project can leverage from more data including 2018, 2023, and election years prior to 1993. We can create classification problems on this graph database as well where the machine learning model would predict if a candidate might win or lose based on its characteristics.

Supervisor: Dr. Qasim Pasta   Project by: Faraz Ali | M. Azeem Haider | Ronit Kumar