Tamil Nadu Assembly Election 2021 Analysis

Saptharishee M[#], Chenchu Aravind[*],Karthick Roshan[#]

saptharishee.m2020@vitstudent.ac.in
karthickroshan.m2020@vitstudent.ac.in
chenchuaravind.v2020@vitstudent.ac.in
Notebook

***Abstract*— With 234 constituencies in the Indian state of Tamil Nadu, the 2021 Legislative Assembly election poll was Tamil Nadu's first assembly election after the demise of the two most prominent Chief Ministers in the state's modern history, J. Jayalalithaa and M. Karunanidhi. In order to improve the economy after the COVID-19 pandemic, both the AIADMK and DMK promised jobs in their manifestos. Industries, especially MSMEs, have been hit hard by the slowdown in the economy. These factors helped and made us take up this topic and analyze it.**

The problem uses "TCPD Indian Elections Data v2.0" of Tamil Nadu General Legislative Election 2021 and analyses the result with the list of political parties participating, performance of political parties,performance of women candidates, and other related data,percentage of votes cast, also a logistic regression model with 98.72 % accuracy and a bring out a visual idea. With the loss of 2 major leaders from both the major parties, Tamil Nadu faced its most anticipated 16th legislative election on 6-April-2021 at a single phase. The results of this election were released on 2 May-2021 by Election Commission Of India(ECI) .
***Keywords*— Tamil Nadu,Election,Parties,ECI (Election Commission Of India).**

I. Introduction

India, the largest democracy of the world, has a parliamentary system defined by constitution . The constitution states two types of elections held for Lok Sabha(Union Government) and Vidhan Sabha(State Government).where elected members of both work together on several aspects. Both elections happen in a 5 year interval.The Lok Shabha candidates are chosen by General election and the Vidhan Sabha by Assembly/Legislative assembly election.The independent organisation Election Commission of India(ECI) is responsible for holding,managing and releasing the results of election .State elections are conducted by State ECI and General Elections are managed by ECI .

The state Tamil Nadu(India) has seen 16 Assembly elections with several major parties and ideologies over the time.The state has 234 constituencies(38 Districts). The polling held on 6th April 2021 had voter turnover percentage of 72.81%.DMK(Dravida Munnetra Kazhagam) and allies(SPA-Secular Progressive Alliance) won the election with 159 seats (DMK-133 ,INC-18 ,VCK-4 ,CPI-2 ,CPI(M)-2).The opponent party AIADMK(All India Anna Dravida Munnetra Kazhagam) and allies(NDA-National Democratic Alliance) won 75 seats (AIADMK-66,PMK-5,BJP-4).

II. Literature

Is the Anti-Incumbency Sentiment Changing in Tamil Nadu? A Case Study of 2021 Tamil Nadu Assembly Elections - Anurag Piyamrao Wasnik , Divya Saroja Chayanam

The author of the paper analyzed 5 elections of Tamil Nadu and made an analysis to check if there is a shift in the strong sentiment in favor of voting out politicians who already hold power. At the intro it speaks briefly about the politics in India and the voting behavior where the authors mention about electoral 'swing' and how it happens .The author speaks about Tamil-Nadu politics and categorizes it as "anti-incumbency" because of the political shift that happens every 5 years. Author also mentions that "data shows that a swing of 5% votes is enough to make a winner out of a loser and vice versa." Which is an interesting and healthy point for a democracy.

3.1 speaks about consistent and deviant constituency which is basically whether they want the same party/person to continue or change them .This analysis shows us that except for 2011 every year the consistent constituency has been more than 48%.It can also be inferred that consistent constituency percentage has been increasing for past 10 years.

In 3.2 The author calculates the switching frequency where the author takes the data of past 3 elections 2011,2016,2021 and grouped it into incumbent and anti-incumbent. Where the author categorizes constituencies into "swing constituencies" and consistent constituency .Author gives an inference that constituencies with high switching frequency are deviant ,experimental and more volatile whereas the constituency with zero deviant or zero switch frequency it is a consistent constancy .There are 25 parties with 0 frequency switches out of which 17 remained incumbent(same party won that constituency for 3 years in a row) . Author infers that the number of constituencies that stayed with their previous voting option is about 3.5 times the number of constituencies

that shifted their vote, indicating a strong preference for the incumbent.

In 3.3 Author speaks and raises concern about the raise in caste based voting among constituencies of Tamil Nadu .Author also highlighted various factors such as Ideological blur and distribution percentage that which has lead in this way .It also categorizes TN into 4 parts such as Chola Naadu, Thondai Naadu, Nadu Naadu, Kongu Naadu and Pandya Naadu and lists out Present day dominant caste- based parties and the dominant cast in that region Where it is clearly evident that caste plays a major role . The author also mentions the vote share of a caste based political party called PMK has grown about 1476% since 1991-2016.

To conclude the author states that this trend of voting based on caste-line is not a good trend for democracy, author also cites an example of Lebanon that ensures diversity .Author also says that Dravidian parties must reconsider whether they are exacerbating or lessening the sectoral gap in the Tamil Nadu political arena. Aside from that, such a sectoral difference adds to the administrative cost of maintaining communal harmony and peace inside the state.

How Tamil Nadu voted in 28 charts: DMK won a clean victory but AIADMK remain competitive adversary- Gilles Verniers, Vignesh Karthik KR, Mohit Kumar & Neelesh Agrawal

The article starts with an account on what results were expected and what came out after elections , which it states as " not a surprise " .It also states that the opponent AIADMK was also not bad , They only had 7% of difference in vote than 2016 legislative election .It also states that DMK gained 6% of increase in overall vote percentage .There is just a 4% difference in vote share between the two parties, but that disparity yielded significantly different outcomes in terms of seats: the DMK won 70.7 percent of its contests, while the AIADMK won 34.5 percent. As a result, the DMK received twice as many seats as its opponent.

The seat distribution of election shows a clear majority of DMK(alone) with 57%(133) of seats and its alliance added up 13% (26)of seats whereas on the other hand AIADMK took the 2nd place in terms of separate vote share with 28%(66) of seats.BJP and PMK both the alliance of AIADMK won 4 and 5 seats each which is 2% each.

The next visualization of assembly election(2016), General election(2019) ,Legislative election (2021) showed rise in support for DMK .Even though all these elections are for different frameworks it still shows gradual rise .Other small-party combinations fared poorly in this election. The AMMK-led alliance received only 2.8 percent of the vote, while the MNM-led alliance received only 2.7 percent. They were unable to get any seats. However, the NTK ran on its own and received 6.6 percent of the vote share, up to 1.1 percent in 2016. In 2019, it received 3.9 percent of the vote .After several other analysis on previous election data the article concludes that ,there was a clear win by DMK over AIADMK .Also it says in 2019 and 2021, the election was won by a party that was successful in incarnating a strong regional identity and forming a big alliance not just against a local enemy but also against Hindu majoritarianism. The DMK fought on the notion of regional identity, as manifested via cultural and even civilizational purity. It campaigned just as hard against the BJP, which has no presence in the state, as it did against the AIADMK. Attacking the BJP on subjects of broad interest was also an effective technique of making the AIADMK appear bad by association.

PRS Legislative Research - Manish Kanadje

This specific analysis visualizes the previous election(2016) and 2021 election and draws few conclusions. First it compares the number of seats won by each party in both elections and concludes that DMK won the 2021 election and AIADMK won the 2016 election with 133 and 134 constituencies respectively. On visualizing Women's representation it is inferred that women's representation reduced nearly by half; visualization of the age of MLA's shows a sharp increase in MLAs above 70 years .Visualization of educational qualification of MLA's ,62% of the MLAs have at least a bachelor's degree in 2021 ,which is up from 59% in 2016.

III. Outline

We discovered 50 variables which include variables with high frequency null values in the dataset. We chose a feature subset from them and reduced it to 18 variables. We now have 18 independent variables and the candidate's position. We changed the position of the candidate by replacing every 1 with True and the rest with False. We also deleted NOTA values from rows due to an overabundance of null values. Now that we have the preprocessed data, we need to convert it to int and float types so that it can be used in the model.

Later, we divided the data into train data and test data, using train data to train the model and test data to create predictions and assess the data's correctness.Then we used 2 methods to predict the data .We used Scikit-learn ,Numpy , Pandas and matplotlib libraries in python.For visualization we used Tableau.

IV. Methods Used

We have Used k-NN(k-nearest neighbors algorithm) and logistic regression to build a predictive model for the given data

A. k-nearest neighbors algorithm (k-NN)

The KNN algorithm presumes that comparable objects exist nearby. KNN encapsulates the concept of similarity (distance) with mathematics - calculating the distance between points on a graph.

The KNN Algorithm:

1.Load the data

2.Set K to the number of neighbours you choose.

3. For each sample in the data

3.1 Determine the distance between the query example and the current example in the data.

3.2 To an ordered collection, add the distance and the example's index.

4. Using the distances, sort the ordered collection of distances and indices from smallest to greatest (in ascending order).

5. Select the first K elements from the sorted collection.

6. Obtain the labels of the chosen K entries.

7. If regression, return the mean of the K labels.

8. If classification, return the mode of the K labels.

B. Logistic Regression

Logistic Regression is used to divide a set's items into two categories (binary classification) by estimating the likelihood of each member in the set.

Logistic regression is a linear approach, but the logistic function transforms the predictions. As a result, we can no longer comprehend the predictions as a linear combination of the inputs, as we can with linear regression. To continue from above, the model may be described as:

$$P(y(i)=1)=1/1+\exp(-(\beta 0+\beta 1x(i)1+\ldots+\beta pxp(i)))$$

Here bi(the coefficients) are estimated using Maximum likelihood estimation and substitute necessary values in the equation

The coeffs that we got are
[[-2.16433924e-10,
3.42310734e-08,
2.93638991e-04,
-1.83139373e-04,
4.21978808e-05,
3.92527358e-08,
-1.24441208e-09,
1.37838008e-09,
-3.47612659e-09,
1.21797583e-09,
1.86723372e-08,
6.81547236e-11,
-8.85143776e-10,
-7.24403946e-10,
-2.42701385e-08,
-1.56184025e-08,
9.91338264e-09,
-1.88270744e-09]]

By substituting the value
Lose if p(x) < 0.5
Win if p(X) >= 0.5.

TABLE I
COMPARING THE OUTCOMES

	ML Model	
	k-NN	**Logistic Regression**
Accuracy	97.62 %	98.72 %
Standard deviation	0.84 %	0.62 %
Confusion Matrix		

	Predicted No	Predicted Yes
Actual No	752	9
Actual Yes	5	32

	Predicted No	Predicted Yes
Actual No	754	7
Actual Yes	1	38

V. VISUALIZATION

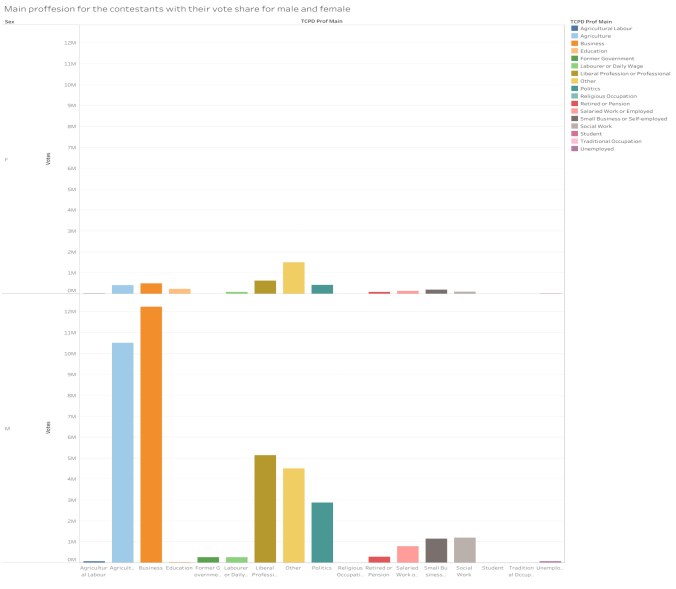

Fig 1 :Main profession for the contestants with their vote share for male and female

Candidates with Business as their profession have got more vote share compared to other professions in Males,whereas in females other professions got the upper hand.

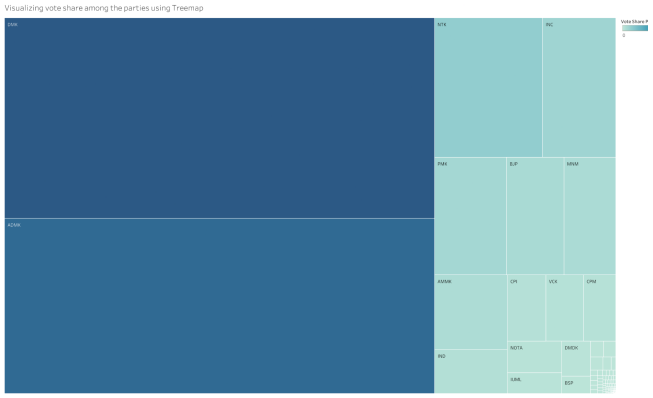

Fig 2 :Visualizing vote share among the parties using Treemap

In fig.2 DMK has got more vote share which has darker color in treemap following DMK, ADMK has got second highest votes.

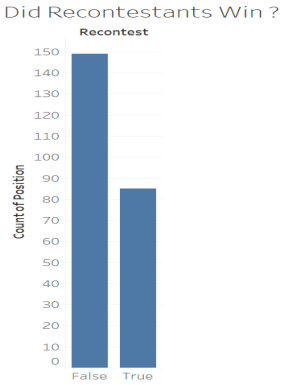

Fig 3: Recontestents winning

There will be a big question: did contestants win the election this will be answered in fig.3 that contestants have not won the election with big margin between won and lost.

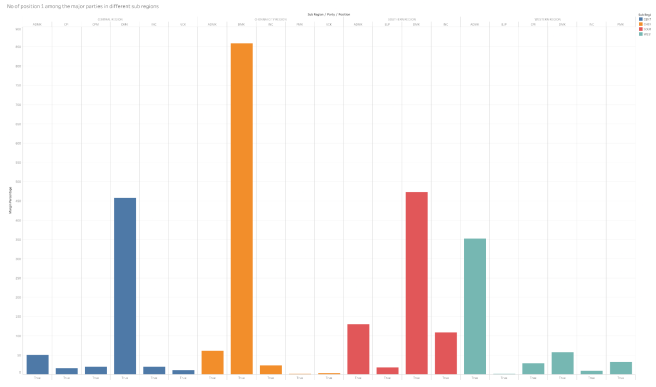

Fig 4: No of winning candidate in different subregions with respect to parties.

DMK has flourished in chennai region then in any other region following in southern region. Next to DMK, ADMK has flourished in the southern region with the lowest percentage than DMK.

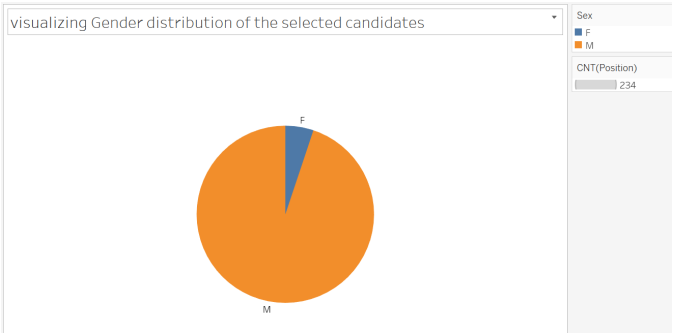

Fig 5: Gender distribution of winning candidates

The number of male contestants are higher so automatically the number of winners i.e in number one position but the number of female contestants are considerably low compared to male, so that concludes females contestants has won less

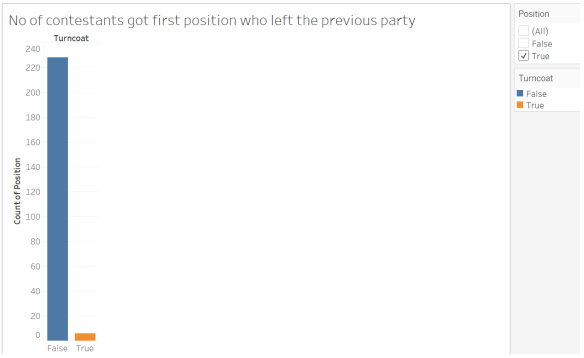

Fig 6:Turncoat (joined another party) winning percentage

Winners/Contestants leaving the previous parties is more common in Indian politics. So in that case Number of contestants who left their previous party has won very less.

VI. Conclusions

First, we clean the dataset and convert all the values into a numerical format. Then we made feature subset selection in the dataset to land on a final dataset to perform logistic and k-NN predictions.

The k-NN classifier model gives us the accuracy of 97.62 % with standard deviation 0.84 %. The k-NN model gave 14 wrong predictions out of a data that consisted of 800 values.

The Logistic regression model gives us the accuracy of 98.72 % with standard deviation 0.62 %which is satisfactory and

better than the k-NN model . The logistic model gave 8 wrong predictions out of a data that consisted of 800 values.

VII. References

[1] DATA:"TCPD Indian Elections Data v2.0″, Trivedi Centre for Political Data, Ashoka University.
Ananay Agarwal, Neelesh Agrawal, Saloni Bhogale, Sudheendra Hangal, Francesca Refsum Jensenius, Mohit Kumar, Chinmay Narayan, Basim U Nissa, Priyamvada Trivedi, and Gilles Verniers. 2021.

[2] Wasnik, Anurag Piyamrao and Chayanam, Divya Saroja, Is the Anti-Incumbency Sentiment Changing in Tamil Nadu? A Case Study of 2021 Tamil Nadu Assembly Elections (November 8, 2021). Journal of Social and Political Sciences, Vol.4 No.4 (2021), Available at SSRN: https://ssrn.com/abstract=3958686

[3] How Tamil Nadu voted in 28 charts: DMK won a clean victory but AIADMK remain competitive adversary- Gilles Verniers, Vignesh Karthik KR, Mohit Kumar & Neelesh Agrawal .Scroll Article on May 09, 2021

[4] PRS Legislative Research - Manish Kanadje on May 3, 2021.https://prsindia.org/files/policy/policy_vital_state/Tamil_Nadu_Profile_2021.pdf

[5] Feature Engineering for Election Result Prediction -Thamindu Dilshan Jayawickrama on Apr 6, 2020 in towardsdatascience article in Medium.

[6] Machine Learning Basics with the K-Nearest Neighbors Algorithm-Onnel Harrison. on Sep 11, 2018 in towardsdatascience article in Medium.