**About the data**

The dataset for this report comes from different data sources. Our primary data acquisition model comes under Social networks, where Tweets were pulled from Twitter using web scrapping methodology. The appropriate permission was reserved from Twitter. The Tweets dataset includes information about the user name, Location or active Location of the user, and the posted tweets. Twitter dataset metadata are shown in table 1.

**Table 1** Details of twitter dataset

|  |  |  |
| --- | --- | --- |
| **S.no** | **Criteria** | **Details** |
| 1 | **Name** | Twitter Dataset |
| 2 | Type | Text data |
| 3 | No. of rows | ~3000 |
| 4 | No. of columns | 3 |
| 5 | Missing values | No |
| 6 | Target Type | Data Mining |
| 7 | Applicable Technique | NLP |
| 8 | Data repository | Social Media (Public domain) |

This data source model is considered public data, where anyone can access the trending tweets using the internet. The geographical survey data acquisition is considered our secondary data source. The survey conducted 'Survey concerning Tamil Nadu Legislative Assembly election, 2021' by approaching local people and circulated the link through WhatsApp, Instagram, and other social media. Survey dataset metadata are shown in table 2.

**Table 2** Details of survey dataset

|  |  |  |
| --- | --- | --- |
| **S.no** | **Criteria** | **Details** |
| 1 | Name | Survey Dataset |
| 2 | Type | Multivariate Dataset |
| 3 | No. of rows | ~407 |
| 4 | No. of columns | 24 |
| 5 | Missing values | No |
| 6 | Target Type | Ordinal/Nominal/Categorical |
| 7 | Applicable Technique | Classification |
| 8 | Data repository | Survey dataset (Owned) |

This data source model is considered private data. Since chosen topic requires historical data for statistical and other geographical analysis, we're taking past and current election details from the governmental repository. Which is accessible by every citizen with the help of the internet. Historical dataset metadata were shown in table 3.

**Table 3** Details of Historical dataset

|  |  |  |
| --- | --- | --- |
| **S.no** | **Criteria** | **Details** |
| 1 | Name | Election Dataset |
| 2 | Type | Multivariate Dataset |
| 3 | No. of rows | 1638 |
| 4 | No. of columns | 9 |
| 5 | Missing values | No |
| 6 | Target Type | Categorical |
| 7 | Applicable Technique | Statistical Analysis |
| 8 | Data repository | Governmental repository |

Those datasets include information about the election, Public opinions, Assembly constituency details, Parties & Alliances details, Party details, Candidate profile, Past result, Public twitter tweets, and another necessary questionnaire with enough pollster’s opinions. Those features descriptions for all the dataset were detailly discussed in following table 4,5,6.

**Table 4** features descriptions of twitter data source

|  |  |  |
| --- | --- | --- |
| **S.no** | **Feature name** | **Description** |
| **1** | User | Registered User name of the twitter user. |
| **2** | Location | Registered Location, or active location of the twitter user, while posting the tweets. |
| **3** | Tweets | Posted tweets in form of Text, Emojis or an Image. |

**Table 5** features descriptions of survey data source

|  |  |  |
| --- | --- | --- |
| **S.no** | **Feature name** | **Description** |
| 1 | Which of the following best describes your age? | This fields help to identify the age grouping of pollsters.  In generally, ‘18 – 24’, ‘25 – 40’, ‘41 – 50’, and ‘Above 50’ are the provided options. |
| 2 | What gender do you most identify with? | To indicate the gender of the pollsters. ‘Male’, ‘Female’ and ‘Other’ are fixed options. |
| 3 | Are you a registered to vote? | To identify whether pollsters are registered to vote or not?  It’s an Yes or No Mandatory field. |
| 4 | Are you registered to vote at the current address you reside at? | To identify whether pollsters are registered to vote at the current address?  It’s an ‘Yes’ or ‘No’ mandatory field. |
| 5 | Do you feel that you fully understand the election process? | This fields help to identify how pollsters understand the election process.  ‘Little’, ‘Lot’ and ‘Moderate’ are some of the options provided for pollsters. |
| 6 | In the last 5 years did you vote in a local election? This includes voting for Councillor & mayors | This fields help to identify whether pollsters has voted in any local election in the last 5 years. It’s an ‘Yes’ or ‘No’ mandatory field. |
| 7 | Did you voted in the 2021 elections? | This fields help to identify whether pollsters voted in 2021 election and It’s an ‘Yes’ or ‘No’ mandatory field. |
| 8 | If not voted, why ? | If Question 7, is answered No. This field capture the reasons why pollster did not voted in 2021 election. FAQ was provided as options. |
| 9 | On what basis do you assess a political candidate? | To identify how assessable are the political candidate over the election period. In general-poster ad, Digital ad, Newspaper, TV ad are some common options provided for pollsters. |
| 10 | Which of the following best describes your decision to vote in the 2021 election? | To identify how assessable are the political candidate during/over the election period. |
| 11 | On what basis you select your political candidate. | This fields help to identify how pollster selecting their political candidate.  Example here some of the provided options.  - Die-hard fan of the Party/ Candidate  - On past experience/Leadership  - On their promises  - Influenced by others |
| 12 | Contributed or collected money | To identify whether pollsters has Contributed or collected money? and It’s an ‘Yes’ or ‘No’ mandatory field |
| 13 | Attended election meetings/rallies | To identify whether pollsters has Attended election meetings/rallies? and It’s an ‘Yes’ or ‘No’ mandatory field |
| 14 | Participated in door to door canvassing | To identify whether pollsters has Participated in door to door canvassing and It’s an ‘Yes’ or ‘No’ mandatory field |
| 15 | Distributed election leaflets or put up posters | To identify whether pollsters has Distributed election leaflets or put up posters and It’s an ‘Yes’ or ‘No’ mandatory field |
| 16 | Do you think the existing government is going in the right direction to benefit Tamil Nadu's people? | To identify whether the existing government is going in the right direction to benefit people and It’s an ‘Yes’ or ‘No’ mandatory field |
| 17 | How would you rate Edappadi Palanisamy’s performance as the Chief Minister of Tamil Nadu | To analysis the performance rating for our Chief Minister of Tamil Nadu.  In general, Very good, Good, Neutral, Bad, and Very Bad are the provided option to choose. |
| 18 | What is your assessment of the performance of the AIADMK government in Tamil Nadu in the last five years? Would you say that you have been satisfied or dissatisfied with it? | To analysis the performance rating for ruling party of Tamil Nadu.  In general, ‘Fully satisfied’, ‘somewhat satisfied’, ’somewhat dissatisfied’, ‘Fully dissatisfied’ and ‘I’m not sure’ are the provided option to choose. |
| 19 | During the last two- three years have you or any of your family members benefited from any Government scheme ? | To identify whether the pollster or any of the family members received any governmental scheme during the last two- three years. It’s an ‘Yes’ or ‘No’ mandatory field. |
| 20 | On the day of voting will you vote for the same party which you voted now or your decision may change? | This filed help to analysis the decision of the pollsters on their day of voting. |
| 21 | Which party do you support in your ward? | This fields help to identify the party which is having majority in their constituency. |
| 22 | What is your opinion about the candidate? | Text descriptive field to process the pollsters opinion about their candidate. |
| 23 | Which party will rule Tamil Nadu for the next five years? | To identify Public concerns to determine which party rule Tamil Nadu for the next five years. Options provided with 5 majority part in Tamil Nadu. |
| 24 | Why do you think your candidate will win? | To identify Public concerns to determine why pollster wanted his/her candidate to win.  Reason were collected in text format and analysed for NLP |

**Table 5** features descriptions of Historical data source

|  |  |  |
| --- | --- | --- |
| **S.no** | **Feature name** | **Description** |
| 1 | S.No | Unique identity number for each record, ranging from 1 to 1638. |
| 2 | Election year | Historical year, only for last two successful election and current 2021 election. (2011, 2016, 2021) |
| 3 | District | Name of the 38 District inside Tamil Nadu |
| 4 | AC No. | Unique assembly constituency number for each assembly constituency. Ranging from 1 to 234. |
| 5 | Assembly Constituency | Given name of the Assembly Constituency under each district. Each of 234 AC named uniquely. |
| 6 | Parties & Alliances | Details of Parties & Alliances group |
| 7 | Party | Details of Party |
| 8 | Candidate | Party-wise name of the candidate for each Assembly Constituency. |
| 9 | Result | Assembly Constituency-wise election result for past record and ‘pending’ status for ongoing election, 2021. |