

Modeling U.S. Election Engagement on Twitter: Classification, NER, and Stance Detection

1 Project Overview

The purpose of this project is to categorize tweets on the US elections into important themes such as campaign updates, policy issues, polling data, and controversy. Additionally, the research tries to extract relevant named entities, such as candidates, political parties, events, and locations. The model classifies tweets as supporting, opposing, or neutral towards certain individuals, policies, or political parties.

2 Objective

This method tracks the current state of political talk on social media in real time, providing journalists, analysts, and campaign staff with valuable insights. Beyond basic categorization, stance classification provides a more contextual view of how various entities and issues are addressed, emphasizing both trends and public mood.

3 Data

Primary Dataset: Kaggle - US Election 2020 Tweets Dataset

Description: Contains tweets related to the 2020 US elections, with tweet content, timestamps, user information, and engagement metrics (likes, retweets), class tagging using GPT4

4 Models

4.1 Task 1: Election Tweet Classification

Classify tweets into the following categories: Campaign Updates, Polling Data, Policy Issues, Controversies or Misinformation, Non-election-related (other)

Baseline Models: Logistic Regression with TF-IDF, Multinomial/Bernoulli Naive Bayes, XGBoost/LightGBM
Advanced Model: Fine-tune LLM for classification

4.2 Task 2: Named Entity Recognition (NER)

Extract key named entities such as Politicians' Names, Political Parties Locations

NER Models: Fine-tune BERT-NER, SpaCy with custom dictionaries

5 Tools and Libraries

- **Libraries:** Hugging Face Transformers, SpaCy, Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn

6 Related Work

- Devlin et al. (2019), "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding": Introduces BERT, which we leverage for tweet classification.
- Mohammad et al. (2016), "Stance and Sentiment in Tweets": Focuses on stance and sentiment detection, crucial for analyzing political tweets.
- Derczynski et al., "Broad Twitter Corpus: A Diverse Named Entity Recognition Resource": Provides a Twitter NER corpus applicable to election-related tweets.
- Garg Caragea, "Stanceformer: Target-Aware Transformer for Stance Detection": Proposes a transformer-based model for stance detection relevant to political tweets.
- Keraghel et al., "A Survey on Recent Advances in Named Entity Recognition": Reviews the latest NER advances, guiding our use of modern techniques.
- Murthy (2015), "Twitter and Elections: Are Tweets Predictive, Reactive, or a Form of Buzz?": Explores Twitter's role in elections, informing our analysis of election tweets.

7 Timeline and Working Plan

Task	Deadline	Description
Data Exploration and EDA	11/1	Load and explore datasets, analyze tweet content
Preprocessing and Cleaning	11/5	Tokenization, lemmatization, and noise removal
Baseline Model Implementation	11/10	Logistic Regression with TF-IDF, Naive Bayes
Advanced Model Implementation and NER Implementation	11/18	Fine-tune BERT for tweet classification, Extract political entities with BERT-NER and SpaCy
Evaluation and Visualization and Report Writing	12/2	Analyze metrics and produce visualizations , refine the final project report

8 Expected Results and Visualizations

Visuals

- **Bar Chart:** Sentiment per Category
- **Confusion Matrix:** Model Performance
- **Word Clouds:** Frequent Words per Category
- **NER Examples:** Sample Tweets with Highlighted Entities
- **Time-Series Plot:** Sentiment Trends Over Time
- **Precision-Recall Curves:** Model Evaluation

Evaluation Metrics

- **Confusion Matrix Statistics:**
 - True Positives (TP), True Negatives (TN)
 - False Positives (FP), False Negatives (FN)
- **Accuracy, Precision, Recall, and F1-Score**

9 Team Roles

- **Shreyas Bhaskar:** Preprocessing and EDA, Classical Algorithms
- **Chaitanya Agarwal:** BERT Model Fine-tuning and Classification
- **Marasanige Samarth Mahendra:** NER Implementation and Evaluation