A Project outline submission

On

Web claim versification/fact-checking for political claims

in

Information Retrieval and Extraction

By team - 12

Matrix

Parvatam Pavan Kumar	2023900008
Kagolanu Anantha Lakshmi Tanuja	2023900010

Introduction:

In an age where information is readily accessible and disseminated through the vast expanse of the internet, political claims and statements hold a significant sway over public opinion, policy decisions, and even the course of elections. However, this unprecedented access to information also comes with a proliferation of misinformation, misleading statements, and outright falsehoods in the realm of politics.

Web claim verification and fact-checking for political claims have emerged as essential tools in the pursuit of truth, transparency, and informed decision-making. The rapid spread of political claims across social media, news outlets, and online forums has made it imperative to distinguish between fact and fiction, to ensure that citizens are equipped with accurate information as they engage in the democratic process.

Problem statement:

The problem revolves around addressing the proliferation of political misinformation, misleading statements, and false claims that are often spread across various online platforms. These claims can come from politicians, news articles, social media posts, and other online sources.

The primary goal is to determine the accuracy of these political claims. It involves assessing whether the claims are supported by credible evidence and whether they accurately represent the facts or context related to a political issue.

It is crucial for promoting transparency, accountability, and informed decision-making in politics. Misinformation can influence public opinion, elections, and policymaking, making it essential to separate fact from fiction.

This process typically involves consulting credible and non-partisan fact-checking organizations, analyzing multiple sources, evaluating context, examining evidence, considering potential bias, and assigning ratings based on the accuracy of the claim.

An important aspect of addressing this problem is educating individuals on how to critically evaluate political claims and recognize reliable fact-checking sources. Developing critical thinking skills is essential in combating misinformation. Reduce the spread of false information, and promote a healthier political discourse.

Summary of the papers (or methods) and datasets that will be implemented.

Here we will be primarily leveraging methods and datasets related to natural language processing (NLP), information retrieval, and data extraction.

Methods:

1. Named Entity Recognition (NER): NER methods identify and classify entities (e.g., names of people, organizations, and locations) in text data. Papers on state-of-the-art NER models like BERT-based models or domain-specific NER techniques can be implemented to extract relevant entities from political claims

2. Information Retrieval Techniques: Implementation of information retrieval methods such as TF-IDF (Term Frequency-Inverse Document Frequency) and BM25 to retrieve relevant documents or articles from a corpus based on the content of the political claim. can adapt and fine-tune these techniques for the task.

3. Keyword Extraction: Utilize keyword extraction techniques to identify important keywords or phrases within political claims. This can help in retrieving relevant documents and understanding the context of claims.

4. Contextualized Embeddings: Implementations of methods that use pre-trained contextualized word embeddings (e.g., BERT) to understand the context of political claims better. These embeddings can be used to find related articles or fact-checks.

5. Claim Stance Classification: models for claim stance classification to determine whether a political claim supports, opposes, or is neutral on a particular issue.

6. Fact-checking Models: Explore fact-checking models that use machine learning, deep learning, or rule-based approaches to verify the accuracy of political claims. Incorporate state-of-the-art models or techniques from research papers in this area

7. Sentiment Analysis: Implementation sentiment analysis techniques to determine the sentiment expressed in political claims and related content. This can provide additional context for claim verification.

8. Performance Metrics: To evaluate the performance of algorithms using Accuracy, recall, precision or F1-score.

Datasets:

1. Political Claim Datasets: Collecting or curating datasets containing political claims from various sources, including social media, news articles, and political speeches.
2. News Article Corpus: A corpus of news articles, blogs, and reports related to politics and current events will be gathered. This corpus will be essential for information retrieval and contextual understanding.
3. Knowledge Graph Data: Collecting or creating a knowledge graph dataset with structured information about political topics, entities, and their relationships.
4. Stance Detection Datasets: Explore stance detection datasets to train models for determining the stance of political claims. Datasets like the Fake News Challenge dataset can be adapted for this purpose.
5. Stance Detection Datasets: Explorin stance detection datasets to train models for determining the stance of political claims. Datasets like the Fake News Challenge dataset can be adapted for this purpose
6. Entity Datasets: Will consider using entity recognition datasets such as CoNLL-03 or custom datasets relevant to our domain.
7. Sentiment Analysis Datasets: For sentiment analysis, we will use sentiment-labeled datasets in the political domain, or may create own sentiment-labeled dataset from political texts.
8. Fact-checking Datasets: will use fact-checking datasets such as those from PolitiFact, FactCheck.org, and Snopes. These datasets contain labeled claims, fact-check ratings, and related information for training and evaluation.

Working plan for expected timeline.

Project Setup and Planning (1-2 weeks)
Data Preprocessing and Exploration (1-2 weeks)
Model Development (6-7 weeks)
Evaluation and Refinement (4-5 weeks)
Deployment and Reporting (1-2 weeks)
Maintenance and Monitoring(1-2 weeks)

The interim and final deliverables.

1. Project Proposal and Scope Document: A detailed project proposal outlining the objectives, scope, and methodology of the fact-checking project. Which will be delivered early in the project to align all stakeholders on the project's goals.
2. Data Collection Report: A report detailing the sources of data, methods used for data collection, and any challenges or limitations encountered during data acquisition. This helps ensure transparency in data gathering.

3. Data Preprocessing Report: Documentation of the data cleaning and preprocessing steps performed on the collected datasets will be provided highlighting any decisions made regarding handling missing data or outliers.
4. Model Development Progress Report: An update on the progress of model development will be provided. This includes details about the selected machine learning or deep learning models, feature engineering, and initial model training results.
5. Model Evaluation Report: Will share the results of model evaluation, including performance metrics, confusion matrices, and any areas of improvement identified during the evaluation phase.
6. Prototype or Demo: If applicable, will deliver a prototype or demo of the fact-checking system. This can be an early version of the web-based platform or application, showcasing the core functionality.
7. Final Project Report: A comprehensive report summarizing the entire project, including objectives, methodology, findings, and conclusions. This report provides an overview of the project's impact and lessons learned.
8. Claim Verification Model: The final, trained claim verification model(s) ready for deployment. Which include any code, scripts, or instructions necessary for using the model.
9. Documentation: A detailed documentation for users and maintainers of the system, including user guides, technical documentation, and instructions for updating data and models will be provided.
10. Performance Metrics and Evaluation Summary: Will share the final model's performance metrics, along with a summary of how well the system has performed in practice, including any insights into the system's effectiveness in verifying political claims.