Truth Guard

"Truth Guard: A Holistic Fake News Detection Application"

Phase 3: Development Part 1

DATA LOADING

Fact-Checking Organizations:

- Data Source: Obtain a dataset from fact-checking organizations, such as Snopes, PolitiFact, and FactCheck.org.
- Data Format: The dataset is typically in structured formats like CSV or JSON.
- Data Loading: Utilize data loading libraries (e.g., pandas) to load the dataset into the project.

Trusted News Sources:

- Data Source: Collect news articles from reputable sources, including major newspapers, international news agencies, and respected news websites.
- Data Format: Text data in various formats, potentially with metadata.
- Data Loading: Implement a data loading module to retrieve and store news articles in a structured format.

DATA PREPROCESSING

Text Data Preprocessing:

- Text Cleaning: Remove special characters, HTML tags, and unwanted characters.
- Tokenization: Split text into words or phrases.
- Stopword Removal: Eliminate common words that don't carry significant meaning.
- Stemming or Lemmatization: Reduce words to their root form.

Multimedia Data Preprocessing:

- Image and Video Analysis: Implement tools for image and video preprocessing, including resizing, filtering, and feature extraction.

INITIAL DATA ANALYSIS

Text Data Analysis:

- Word Frequency Analysis: Determine the most common words in both true and fake news articles.
- Sentiment Analysis: Investigate the sentiment of the news articles.
- Language Pattern Recognition: Identify patterns that can distinguish between true and fake news.

Multimedia Data Analysis:

- Image and Video Features: Extract and analyze features from images and videos, such as image dimensions, color histograms, and video frame rates.

CONCLUSION

In this document, Outline of the initial steps in building the "Truth Guard" project, specifically focusing on data loading, preprocessing, and initial data analysis. These steps are fundamental in setting the foundation for the development of our fake news detection system.