Software Requirements Specification for EchoTruth

Sarron Tadesse, Kalif Byrd, and Dev Raiyani

February 7, 2024

Contents

1 Introduction

1.1 Purpose

This document is a detailed Software Requirements Specification (SRS) for EchoTruth, a machine learning-based solution for detecting fake news. It serves as a formal agreement between stakeholders and the development team.

1.2 Scope

EchoTruth aims to classify news content as 'true' or 'fake' to combat the spread of misinformation.

1.3 Definitions, Acronyms, and Abbreviations

SRS: Software Requirements Specification
ML: Machine Learning
NLP: Natural Language Processing

1.4 References

To be filled with relevant references and documents.

1.5 Overview

The subsequent sections of this document will provide detailed descriptions of the requirements and constraints for the EchoTruth system.

2 Overall Description

2.1 Product Perspective

EchoTruth will function as a part of a larger ecosystem of news validation tools, providing a layer of initial screening for authenticity.

2.2 Product Functions

EchoTruth is a machine learning model that will process news articles, classify them as true or fake based on the machine learning algorithms.

2.3 User Classes and Characteristics

The primary users will be individuals seeking to verify the accuracy of news articles.

2.4 Operating Environment

EchoTruth will operate within Jupyter Notebook or Google Colab environments, leveraging the computational resources provided by these platforms.

3 System Features and Requirements

3.1 Functional Requirements

3.1.1 Data Collection and Preprocessing

FR1: The system shall collect a comprehensive dataset of news articles.
FR2: The system shall preprocess the dataset to facilitate machine learning.

3.1.2 Machine Learning Model

FR3: The system shall implement a machine learning model for classification.

3.1.3 User Interface

FR4: The system shall provide an interactive notebook-based interface for submitting articles and displaying results.

3.2 Performance Requirements

3.2.1 Classification Performance

PR1: The system shall classify news articles with a minimum accuracy of 95

3.3 Environment Requirements

3.3.1 Development Environment

The system shall be developed using Python with support from libraries such as NLTK and Scikit-learn.

3.4 External Interface Requirements

3.5 User Interfaces

The system shall provide an interactive notebook-based interface for users to interact with the EchoTruth service. This interface will be implemented using platforms such as Jupyter Notebooks or Google Colab, allowing for a dynamic and user-friendly environment. Users will be able to input text or upload files directly into the notebook for analysis. The system will then process the input through the fake news detection model and display the results within the same notebook.

This approach is designed to facilitate hands-on interaction with EchoTruth, enabling users to easily test and visualize the model's performance on their inputs. The notebook will include instructions for users on how to use the interface, provide inputs, and interpret the system's output. This setup is particularly suitable for academic, research, and educational purposes, allowing the functionality of the EchoTruth model to be showcased with minimal web development effort.

3.6 Software Interfaces

The EchoTruth project will be implemented within Jupyter Notebooks or Google Colab environments, which will serve as the primary interface for interaction with the machine learning model. These platforms provide a comprehensive ecosystem for data analysis and machine learning tasks, integrating directly with Python libraries such as `Pandas` for data manipulation, `scikit-learn` for machine learning tasks, and `matplotlib` for data visualization. The model itself will be built, trained, and executed within these notebooks, leveraging the extensive support these platforms offer for interactive development and exploration.

4 Other Nonfunctional Requirements

4.1 Security Requirements

Given the interactive notebook environment, the EchoTruth project will prioritize the security of the data and code within the notebook. While traditional web security measures like HTTPS are inherent to the platforms used (Jupyter, Google Colab), attention will be given to:

- Ensuring that any sensitive data used in the notebooks is anonymized or securely managed, especially when notebooks are shared publicly.
- Avoiding the inclusion of sensitive credentials or personal information within the notebook. Any access to external resources will be conducted in a manner that secures personal and sensitive information, using environment variables or secure token storage solutions where applicable.

4.2 Software Quality Attributes

For the EchoTruth project within the notebook environment, the focus on software quality attributes will adapt to emphasize:

- **Availability:** While the notebooks themselves are hosted on platforms that ensure high availability (e.g., GitHub for Jupyter Notebooks, Google Drive for Colab Notebooks), the project will ensure that all necessary dependencies and data (where permissible) are readily accessible for users to run the notebooks with minimal setup.

- **Maintainability:** The code within the notebooks will adhere to best practices in software development, including clear documentation, modular design, and the use of version control to manage changes. This ensures that the project can be easily updated, extended, or modified by others.
- **Accuracy and Timeliness:** The machine learning model's design and implementation will aim for high accuracy in fake news detection, with considerations for how the model can be updated as new data becomes available or as the landscape of fake news evolves.

5 Detailed Requirements

5.1 Functional Requirements Introduction

EchoTruth will offer features critical to the verification of news authenticity. The system will allow users to submit news articles and provide a classification based on its authenticity as 'true' or 'fake'. This functionality is crucial in today's digital age where information can be rapidly disseminated and consumed without proper validation. The sections below detail each functional requirement necessary for EchoTruth to perform its intended operations.

5.2 Functional Requirements

5.2.1 Article Interaction and Analysis

FR7: Users will be able to input news articles directly into the notebook or upload text files for classification. This process will be facilitated through code cells designed to accept user inputs or file uploads within the Jupyter Notebook or Google Colab environment.
FR8: Upon processing the input, the system will immediately display the classification results, including the credibility assessment and any relevant analysis, directly within the notebook. Users will be able to interactively explore the results and optionally modify their query or input new data for analysis in real-time.

5.2.2 News Classification

FR9: The notebook will classify input news articles using a pre-trained machine learning model, incorporating interactive code cells where users can initiate the classification process by providing articles as input.
FR10: For each article classified, the notebook will generate and display a reliability score, alongside other relevant metrics or insights that contribute to the understanding of the article's credibility. This output will be presented directly within the notebook, allowing for immediate review and analysis by the user.

5.3 Performance Requirements

5.3.1 Response Time

PR2: Classification results will be provided within the notebook environment in an acceptable time frame, aiming for immediate feedback upon execution of the analysis code. While exact times may vary based on the complexity of the model and the length of the input article, the goal is to maintain response times that keep the interactive experience engaging, typically within a few seconds to a minute for processing.