Software Requirements Specification (SRS) for News Identifier Project

By: Hardi Dave, Dhairya Patel, and Avi Patel

1. **Introduction:** The News Identifier Project aims to address the challenge of identifying fake news on social media using machine learning. The project seeks to enhance the accuracy of news shared online by analyzing the authenticity of information through predictive models. It will focus on both technical aspects and human behavior, creating a comprehensive solution for misinformation.

2. **Scope:** The system will be designed to classify news as either real or fake. The primary users will be individuals or platforms needing to verify the authenticity of news content. The system will leverage a Kaggle dataset to train machine learning models, which will then classify news articles.

3. **Functional Requirements:**
    - **Data Input:** Accepts news data from social media or other sources in the form of text.
    - **Data Preprocessing:** Cleans and processes the news data for training and testing.
    - **Model Training:** Uses machine learning models (Decision Tree, XGBoost, KNN) to train on the dataset.
    - **Fake News Detection:** Analyzes and classifies news as true or false using the trained models.
    - **Result Output:** Provides results indicating the authenticity of the news with a confidence score.

4. **Non-Functional Requirements:**
    - **Performance:** The system should process and classify news articles within seconds.
    - **Scalability:** The system must handle large datasets efficiently.
    - **Accuracy:** Target an accuracy rate of over 85% for news classification.
    - **Usability:** The system will be user-friendly and accessible via a simple interface.

5. **Tools and Technologies:**
    - **Programming Language:** Python
    - **Tools:** TextBlob, Jupyter Notebook
    - **Machine Learning Models:** Decision Tree Classifier, XGBoost, KNN
    - **Libraries:** TensorFlow
    - **Dataset:** Kaggle Fake News Detection Dataset

6. **Stakeholders:**
    - Social media platforms
    - News consumers
    - Fact-checkers
    - Media companies

7. **Conclusion:** This project provides a solution to the growing problem of misinformation by integrating machine learning with media literacy efforts. It will help mitigate the spread of fake news, protect public trust, and ensure the dissemination of accurate information.