Dhirubhai Ambani Institute of Information and Communication Technology • DA-IICT

Software Engineering

IT314

Fake News Faux Real

Group Members:

Student Id	Name
202001412	ZEEL BHANDERI
202001238	AAYUSH BRAHMBHATT

Table of Contents:

1. Introduction

1.1. Purpose

Fake news has existed since the beginning of time. However, since the advent of the Internet, the rate at which fake news has spread has increased manifold. Given the ease of access to information, it has become very easy for people to spread false information and manipulate public opinion. This has led to significant consequences, including spreading misinformation, manipulating public opinion, and undermining trust in journalism and other sources of credible information. As a result, such technological solutions become extremely important to help people identify and flag false news and distinguish between credible and unreliable sources of information. In conclusion, developing a fake news classifier such as FakeRanks Faux Real is essential for addressing the challenges posed by the spread of false information online and ensuring the credibility and integrity of information in the digital age.

1.2. Intended Audience and Reading Suggestions

This document is intended for software developers, project managers, and stakeholders who will be involved in the development, testing, and implementation of the software. It is recommended to read this document thoroughly to gain an understanding of the requirements before beginning development.

1.3. Product Scope

- TheFakeNewsDetectorisasoftwareapplicationthatisdesigned to detect and classify news articles as either real or fake based on certain criteria. The application will analyze the content of the article and compare it to known sources of fake news. It will also check the reputation of the source, the language used, and the tone of the article to determine its veracity.
- Theapplicationwillbedesignedtobeuser-friendlyandeasyto use. It will have a simple and intuitive interface that allows users to input a news article or url of the article and receive a classification of whether it is real or fake.
- TheFakeNewsDetectorwillbedesignedtobehighlyaccurate, with a low false positive rate.
- Itwillberegularlyupdatedtoensurethatitremainsup-to-datewith the latest sources of fake news and continues to provide accurate classifications.
- Theapplicationwillbedesignedtobeusedbybothnoviceand expert users, with user-friendly features and advanced functionality for more experienced users.
- Overall,theFakeNewsDetectorapplicationwillprovidean important tool for combating the spread of fake news and ensuring that users have access to accurate and reliable information.
- The"FakeNewsDetector"usesacombinationofmachinelearning and natural language processing techniques (NLP) to analyze news articles and assess their credibility. It may also use a database of known fake news sources to cross-reference the article's source and content.

2. Overall Description

2.1. Product Perspective

The Fake News Detector will be a standalone software application that will run on various operating systems, including Windows and Mac. It will interact with external data sources, such as news articles and user input.

2.2. Product Functions

The software application will perform the following functions:

- Analyzenewsarticlestodeterminetheprobabilityofthearticle being fake
- Provideascoreindicatingthelikelihoodofthearticlebeingfake
- Displaytheresultsinauser-friendlyinterface
- Allowuserstoinputnewsarticlesforanalysis
- Allowuserstoadjustcertainsettingsfortheanalysis

2.3. User Classes and Characteristics

The Fake News Detector will be used by various individuals, including journalists, researchers, and the general public. Users will require a basic understanding of how to access and use the software application.

2.4. Operating Environment

The software application will be able to run on various operating systems, including Windows, Mac, and Linux. Users will require an internet connection to access the application.

2.5. Design and Implementation Constraints

The software application will be designed using advanced algorithms and techniques to analyze news articles. The design will be scalable to accommodate a growing user base.

2.6. User Documentation

The software application will provide user documentation in the form of online help and tutorials. The documentation will include instructions on how to use the software and interpret the results.

The software will provide user documentation in FAQ form to provide instructions how to use the software and interpret the results.

2.7. Assumptions and Dependencies

The software application assumes that the news articles provided for analysis are in English. It also depends on external data sources for the analysis, including news articles and user input.

3. External Interface Requirements

3.1. User Interfaces

The software application will have a user-friendly interface that allows users to input news articles for analysis and adjust certain settings for the analysis. The interface will also display the results of the analysis in a clear and concise manner.

The interface shall have the following features:

- InputfieldtoenternewsarticleURLortext
- Submitbuttontotriggertheanalysisprocess
- Outputsectiontodisplaytheanalysisreport

- Clearbuttontoresettheinputandoutputsections.

3.2. Hardware Interfaces

The software application will require a computer or mobile device with an internet connection to access the application.

3.3. Software Interfaces

The following software dependencies shall be required to run the Fake News Detector:
Python 3.5 or higher Natural Language Toolkit (NLTK) library
Scikit-learn library
Flask web application framework

3.4. Communication Interfaces

The Fake News Detector shall communicate with external systems through HTTP requests and responses. It shall use the following communication protocols: HTTP for web-based communication JSON for data exchange

4. System Features

- Advancedmachinelearningalgorithms:Theapplicationusesadvanced machine learning algorithms such as Decision Tree, Naïve Bayes, Logistic Regression, Support Vector Machines to analyze and classify news articles and other forms of digital content based on various features, such as the style of writing and the use of language.

- News article classification: The application provides an automatic classification of news articles into credible or unreliable categories, helping users quickly and easily identify false information.

- Userinterface:FakeRanksFauxRealprovidesauser-friendlyinterfacethat allows users to submit articles for classification and view the results.

- Regular updates: FakeRanks Faux Real is regularly updated to ensure that it remains effective and accurate in identifying false information online.

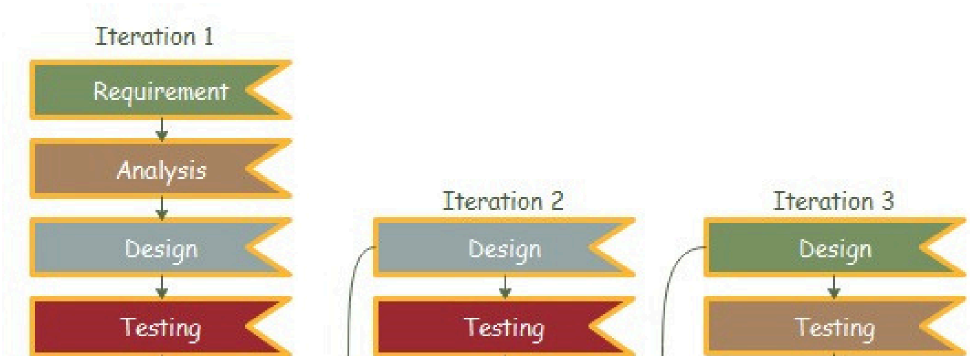
- SignUp:TousetheFakeRanksFauxRealsystem,ausermustfirstcreate an account. The sign-up process typically involves entering personal information such as a username, password, and email address. The user will then receive a confirmation email to activate their account.

- Login:Aftercreatinganaccount,theusercanlogintotheFakeRanksFaux Real system by entering their username and password. The system will verify the credentials and grant access to the user if they are correct.

- SessionManagement:Onceloggedin,thesystemwillkeeptrackofthe user's session to ensure that they remain logged in until they log out or their session expires.

5. SDLC Model

We have used the Iterative Modelfor FakeRanks Faux Real.



An iterative model can be used in the development of a fake news classifier like FakeRanks Faux Real for several reasons:

- ImprovingAccuracy:Aniterativemodelallowsforcontinuousimprovement of the classifier's accuracy. The model can be trained on new data and updated based on feedback from users, allowing it to adapt and improve over time.

- Addressing Complexity: Fake news classification is a complex task, and an iterative model can handle this complexity by continually refining the classification process.

- Incorporating Feedback: The iterative model allows for user feedback to be incorporated into the training process, which can improve the accuracy of the classifier.

- BetterModelSelection:Aniterativemodelcanhelpdeterminethebest algorithms and techniques for the classification task, as well as the best parameters for those algorithms.

- HandlingEvolvingData:Thenatureoffakenewsisconstantlyevolving,and an iterative model can handle this by incorporating new data and evolving to address new types of fake news.

- Overall, using an iterative model in the development of a fake news classifier like FakeRanks Faux Real can provide greater accuracy, flexibility, and robustness, allowing the classifier to continually improve over time.

6. Other Non-Functional Requirements

Performance:

The system should be able to process a large volume of news articles and social media posts in real-time, and provide accurate results within seconds. This is important to ensure that fake news articles are detected as quickly as possible, before they can be shared and potentially cause harm.

Usability:

The system should be easy to use and navigate, with clear instructions and feedback for users. This is important to ensure that users can effectively utilize the system without requiring extensive technical knowledge, and to encourage widespread adoption of the system.

Compatibility:

The system should be compatible with different devices and platforms. It should be accessible to all users, regardless of the device they use.

Scalability:

To ensure scalability, the system utilizes machine learning and natural language processing techniques (NLP). These techniques help in automating the detection of fake news, which are trained on large datasets of fake and real news. By doing so, the system can quickly and accurately analyze vast amounts of news articles, making it scalable to analyze a large volume of data.

Security:

The system ensures the security of user data by implementing encryption methods to protect sensitive information such as passwords and personal information. It also implements user access controls to ensure that only authorized personnel can access and modify data which is achieved by implementing authentication protocols, such as two-factor authentication and role-based access control. Users are validated utilizing numerous security stages so solid security is given.

Flexibility:

It is achieved by delivering the system through multiple platforms, updating content, collaborating with others, and gathering continuous feedback. These measures can help ensure that the system remains relevant, adaptable, and effective in promoting media literacy and combating fake news.

Speed:

Given how rapidly false news may propagate, the fake news detection system must be able to analyze massive volumes of data quickly. The detection and halting of the spread of bogus news may be delayed by a sluggish system.

Accuracy:

The false news detection system's accuracy is crucial since it is essential to make sure the system is dependable and capable of doing so. Inaccurate detection might result in false positives or negatives, which can make users doubt the system and perhaps propagate misleading information.

Customizability:

The system must be customizable to meet the needs of different users. This is important because different users may have different requirements for the system.

Adaptability:

The system must be adaptable to changing trends and patterns of fake news. This is important because fake news can change rapidly and the system must be able to keep up with these changes.

Reliability:

The project takes a comprehensive approach to ensuring the reliability of its

tool, combining reliable sources, advanced AI algorithms, human review, and a transparent methodology to provide users with accurate and trustworthy information.
Prediction of whether the news is real or fake ought to be quick.

Security: Validation of users is done by using a database. All the user information is stored in a database and users are not allowed to move to the main portal if their login fails.

7. Functional Requirements

- User authentication: System allows users to create accounts and log in securely.
- News article submission: Users are able to submit news articles for classification.
- Text analysis: System analyzes the text of the submitted news articles and identifies fake news.
- News classification: System classifies news articles as either "real" or "fake" based on the results of the text analysis.
- Results display: System displays the results of the news classification to the user, along with an explanation of the decision.
- Performance evaluation: System evaluates its performance based on user feedback and continuously improves its accuracy.
- Fetchingarticles: System fetches the data upon entering the url of the news source.

Appendix A: Glossary

NLP :Natural language processing is a subfield of computer science, information.

NLTK:Natural Language Toolkit [3] (NLTK) is a platform used for building Python programs that work with human language data for applying in statistical natural language processing (NLP)

Python: An interpreted high-level programming language for generalpurpose programming.

Appendix B: Analysis Models

Domain Analysis Model:

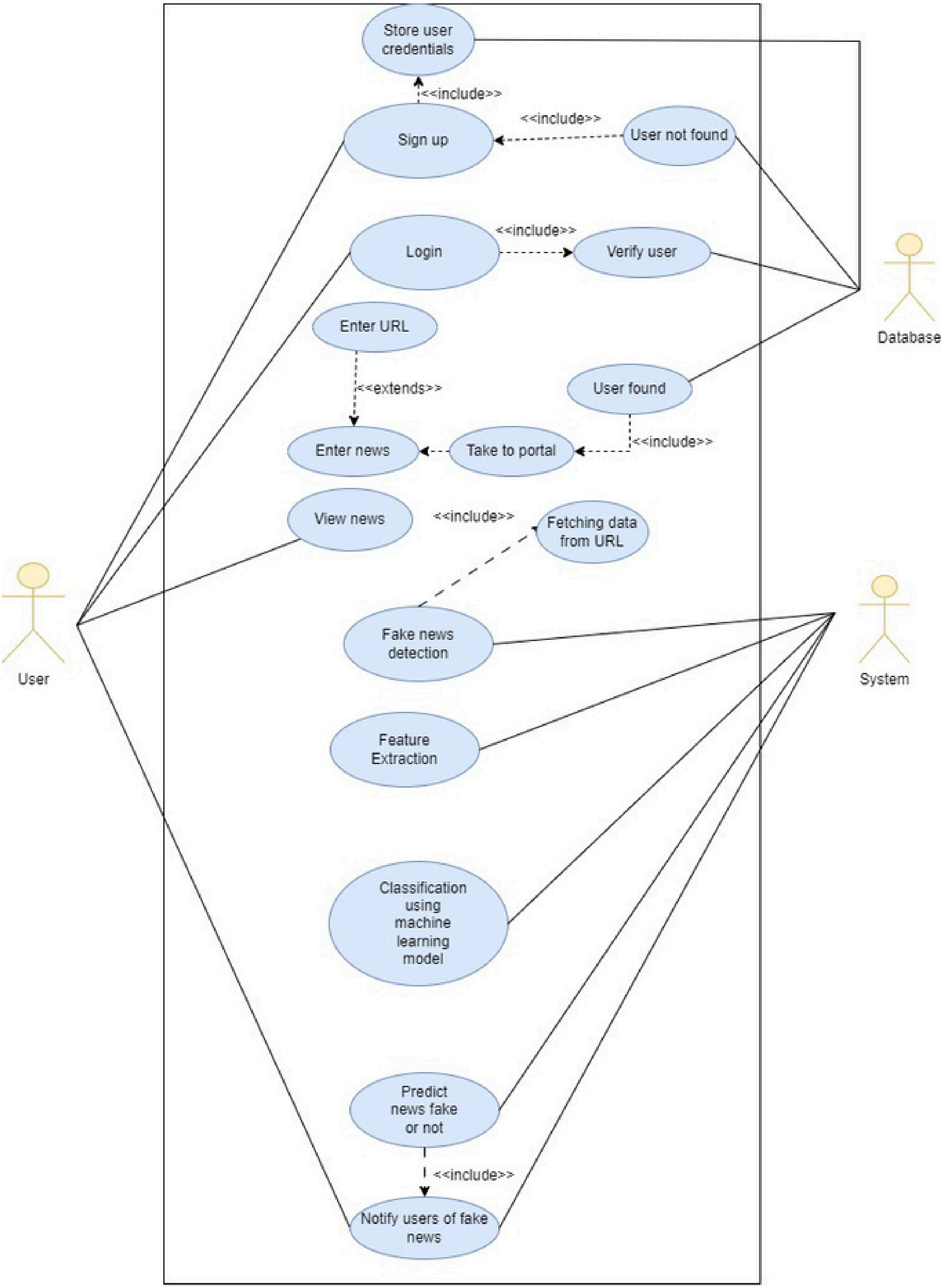
The "Fake News Detector" is a tool designed to identify and flag potentially fake news stories. To create a domain analysis model for this tool, we can consider the following factors:

1. Domain: The domain of the "Fake News Detector" is online news media and journalism.
2. Purpose: The purpose of the "Fake News Detector" is to identify and flag potentially fake news stories that are being circulated on the internet. The tool aims to assist users in identifying news articles that may not be credible or reliable.
3. Functionality: The "Fake News Detector" uses a combination of machine learning and natural language processing techniques (NLP) to analyze news articles and assess their credibility. It may also use a database of known fake news sources to cross-reference the article's source and content.
4. User interface: The user interface of the "Fake News Detector" may be a website or mobile app that allows users to input a news article's URL or text. The tool will then analyze the article and provide a credibility score or a warning if the

article is deemed to be potentially fake.

5. Data sources: The "Fake News Detector" may use a variety of data sources to assess the credibility of news articles, including news databases, social media, and online news sources. It may also use user feedback to improve its accuracy over time.
6. Fact-checking: The third area of focus is fact-checking. The model should be able to compare the information in the news with known facts and information from reliable sources. This could involve cross-referencing news articles with external sources of information or using data from fact-checking websites.
7. Accuracy: The accuracy of the "Fake News Detector" will depend on the quality of its machine learning algorithms and the amount and quality of data it has access to. The tool may need to be updated regularly to keep up with changes in the online news media landscape.
8. Legal and ethical considerations: The "Fake News Detector" must be designed in compliance with legal and ethical standards. It must protect users' privacy and not be used to suppress or censor legitimate news stories. It should also be transparent in how it assesses the credibility of news articles and how it makes decisions about what articles to flag as potentially fake.

Use Case Diagrams and Description:



Use Case	Sign up
Description	To allow the user to either sign in to an existing account or create a new account to access the system.
Primary Actor	User
Secondary Actor	Database
Pre condition	1. Theuserhasaccesstoadevice with an internet connection and a web browser. 2. Theuserhasnotalready registered an account on the FakeRanks Faux Real platform. 3. Theuserhasaninterestinusinga fake news classifier tool.
Post condition	The user is either signed in to the system or has created a new account and is automatically signed in.
Flow	1. Theuseropensthesystem'slogin page. 2. Iftheuseralreadyhasanaccount, they select the "Sign in" option and enter their username/email and password. 3. Thesystemverifiestheuser's credentials and logs them in.

	4. Iftheuserdoesnothavean account, they select the "Sign up" option and provide the required information, such as their name, email, password, and any other relevant details. 5. Thesystemverifiestheuser's information and creates a new account for them. 6. Thesystemautomaticallylogsthe user into the new account. 7. Theuserisredirectedtothe system's main page or the page they were trying to access before being prompted to sign in/up.
Alternate Flow	3.a If the user enters incorrect login credentials, the system prompts them to try again. 5.a If the user provides incomplete or incorrect information during the sign-up process, the system prompts them to correct the errors or provide additional information.
Extension	If the system is offline or not functioning properly, the user cannot sign in or sign up

Use Case	Feeding article text for fake news classification
Description	To input article text into the FakeRanks Faux Real platform for fake news classification
Primary Actor	User
Secondary Actor	Database
Pre condition	1. Theuserisloggedintotheir FakeRanks Faux Real account 2. Theuserhasaccesstoanarticle that they want to classify as either real or fake news 3. Theuserhascopiedthearticletext to their device's clipboard
Post condition	The user has received a classification result for the article text they provided.
Flow	1. Theusernavigatestothe FakeRanks Faux Real platform and logs in to their account. 2. Onceloggedin,theuseris directed to the platform's dashboard. 3. Onthedashboard,theuserselects the "Classify News" option. 4. Theuserispromptedtopastethe article text into a text box provided by the platform.

	5. Theuserpastesthearticletext from their device's clipboard into the text box. 6. Theplatformanalyzesthearticle text and provides a classification result, indicating whether the article is real or fake news. 7. Theuserispresentedwiththe classification result.
Alternate Flow	
Extension	1. Iftheuserencountersanyerrors while pasting the article text or submitting it to the platform, they are prompted with an appropriate error message and given instructions on how to rectify the issue. 2. Ifthearticletextprovidedbythe user is in a format that is not compatible with the platform's analysis algorithm, the user is prompted with an error message and given instructions on how to reformat the text.

Use Case	Feeding article text for fake news classification by providing a link to the article

Description	To input article text into the FakeRanks Faux Real platform for fake news classification by providing a link to the article
Primary Actor	User
Secondary Actor	System
Pre condition	1. Theuserisloggedintotheir FakeRanks Faux Real account 2. Theuserhasaccesstoanarticle that they want to classify as either real or fake news 3. TheuserhastheURLlinkforthe article 4. Thearticletextshouldbein between 2015 to 2020 and it is also related to Donald Trump.
Post condition	The user has received a classification result for the article text retrieved from the provided URL link.
Flow	1. Theusernavigatestothe FakeRanks Faux Real platform and logs in to their account. 2. Onceloggedin,theuseris directed to the platform's dashboard. 3. Onthedashboard,theuserselects the "Classify News" option.

	4. Theuserispromptedtoprovide the URL link to the article they want to classify. 5. TheuserpastesortypestheURL link into the text box provided by the platform. 6. Theplatformretrievesthearticle text from the provided URL link and analyzes it. 7. Theplatformprovidesa classification result, indicating whether the article is real or fake news. 8. Theuserispresentedwiththe classification result.
Alternate Flow	
Extension	1. Iftheuserencountersanyerrors while providing the URL link or submitting it to the platform, they are prompted with an appropriate error message and given instructions on how to rectify the issue. 2. Ifthearticletextretrievedfromthe provided URL link is in a format that is not compatible with the platform's analysis algorithm, the user is prompted with an error

	message and given instructions on how to reformat the text.

Use Case	Notifying the user of the classification result for a given article
Description	To receive a notification of the classification result (real or fake news) for a given article
Primary Actor	System
Secondary Actor	User
Pre condition	1. Theuserisloggedintotheir FakeRanks Faux Real account 2. Theuserhasprovidedarticletext or a link to an article for classification by the platform
Post condition	The user has received a notification of the classification result for the provided article.
Flow	1. Afterprovidingarticletextoralink to an article, the user waits for the classification result. 2. Oncetheplatformhasanalyzed the provided article text, it provides a classification result.

	3. Theplatformsendsanotificationto the user indicating the classification result, either "real news" or "fake news". 4. Theuserreceivesthenotification and views the classification result. 5. Theusercanchoosetoviewmore information about the classification result, such as the reasons behind the classification or the platform's confidence level in the classification.
Alternate Flow	
Extension	If the platform encounters any errors during the classification process, the user is prompted with an appropriate error message and given instructions on how to rectify the issue

Use Case	Feature Extraction for Fake News Classification
Description	The goal of this use case is to extract relevant features from the raw text data to be used in the classification of fake news articles.
Primary Actor	FakeRanks Faux Real System (preprocessing of text data)

Secondary Actor	
Pre condition	1. Theinputtothisusecaseisraw text data obtained from various sources such as news articles, social media, and other online content. 2. Thesystemhasaccesstothe necessary libraries and resources for processing text data.
Post condition	1. Theoutputofthisusecaseisaset of relevant features extracted from the raw text data, which can be used as input to the fake news classification algorithm. 2. Theextractedfeaturesareina suitable format for the classifier, and the system is ready to proceed with the classification process.
Flow	1. Thesystemreadsintherawtext data. 2. Thesystemappliesvarious preprocessing techniques such as tokenization, stop word removal, and stemming to clean the data. 3. Thesystemappliesfeature extraction technique such as tf-idf vectorization to extract relevant features from the cleaned text.

	4. Thesystemoutputstheextracted features in a suitable format for use in the fake news classification algorithm.
Alternate Flow	1a. If the raw text data is not available, the system prompts the user to provide the input. 2a. If the preprocessing techniques fail to clean the data effectively, the system may resort to manual inspection or other advanced techniques. 3a. If the feature extraction techniques fail to extract relevant features, the system may experiment with other techniques or adjust the parameters of the existing techniques. 4a. If the dimensionality reduction techniques fail to reduce the number of features effectively, the system may experiment with other techniques or adjust the parameters of the existing techniques.
Extension	

Use Case	Classification using Machine Learning for FakeRanks Faux Real with Preprocessed Text Input by the System

Description	The goal of this use case is to use Machine Learning to classify preprocessed text data fed into the FakeRanks Faux Real system as either real or fake news.
Primary Actor	FakeRanks Faux Real system
Secondary Actor	
Pre condition	1. TheFakeRanksFauxRealsystem has access to a preprocessed dataset of labeled news articles to use for training the machine learning model. 2. TheFakeRanksFauxRealsystem has access to a preprocessed dataset of unlabeled news articles to classify using the machine learning model.
Post condition	1. TheFakeRanksFauxRealsystem has a trained machine learning model that can classify new, preprocessed news articles as either real or fake. 2. Theperformanceofthemachine learning model has been evaluated using metrics such as accuracy, precision, recall, and F1 score.

Flow	1. TheFakeRanksFauxRealsystem receives preprocessed news articles to be classified. 2. Thepreprocessedtextdataissplit into training and testing sets. 3. Themachinelearningalgorithmis applied to train a model using the preprocessed text data and their corresponding labels (real or fake). 4. Theperformanceofthemachine learning model is evaluated using metrics such as accuracy, precision, recall, and F1 score. 5. Thetrainedmachinelearning model is used by the system to classify new, preprocessed news articles as real or fake.
Alternate Flow	
Extension	