**Department of Information Technology**

**Faculty of Computing**

**The Islamia University of Bahawalpur**



**Software Design Document**

**FAKE NEWS DETECTION USING NLP**

**Submitted by**

***ZAINAB BIBI***

***F22BINFT1M01069***

**Submitted to**

**Sir Faisal Shahzad**

**Summery**

The rapid growth of online media platforms has significantly increased the spread of fake news, posing serious threats to public opinion, social harmony, and decision-making processes. This project focuses on developing an automated fake news detection system using Natural Language Processing (NLP) and machine learning techniques. The system preprocesses and analyzes textual news data to extract meaningful linguistic features, such as word frequency and semantic patterns. These features are then used to train supervised machine learning models that classify news articles as either real or fake. By leveraging NLP-based text analysis and classification algorithms, the proposed system aims to achieve high accuracy in identifying misleading content, thereby helping to combat misinformation and promote reliable information dissemination.

**Table of Content**

[**1.** **Introduction** 3](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.b0zucao0arfi)

[**1.1.** **Purpose of SDD** 4](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.3q29tjw07goc)

[**1.2.** **Scope** 4](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.toch3nd8wbxp)

[**2.** **Application Architecture**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.p95ikicl4lj7) 4

[**2.1.** **Architectural Overview**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.vda44ve2w9t9) 4

[**2.2.** **High-Level Architecture Components**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.9r016r8eup50) 5

[**3.** **Data Model Schema**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.3uemzja4z1sn) 5

**3**[**.1.** **Dataset Structure**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.vda44ve2w9t9) 5

**3**[**.2.** **Data Entities**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.9r016r8eup50) 6

**3**[**.3.** **Model Data**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.9r016r8eup50) 7

**4**[**.** **User Interface Design**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.m9bk2fnzgago) 7

**4**[**.1.** **UI Overview**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.vda44ve2w9t9) 7

**4**[**.2.** **Main Screens**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.9r016r8eup50) 7

**4**[**.3.** **UI Flow**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.9r016r8eup50) 8

**5**[**.** **Non-Functional Design Considerations**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.m9bk2fnzgago) 8

**5**[**.1.** **Performance**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.vda44ve2w9t9) 8

**5**[**.2.** **Scalability**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.9r016r8eup50) 8

**5**[**.3.** **Usability**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.9r016r8eup50) 8

**5**[**.4.** **Maintainability**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.9r016r8eup50) 8

**5**[**.5.** **Security**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.9r016r8eup50) 8

**6**[**.** **Design Diagrama**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.m9bk2fnzgago) 9

**6**[**.1.** **Architecture Diagram**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.vda44ve2w9t9) 9

**6**[**.2.** **Sequence Diagram**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.9r016r8eup50) 10

**6**[**.3.** **Component Diagram**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.9r016r8eup50) 10

**7**[**.** **Conclusion**](https://docs.google.com/document/d/1232Lieab35d8PVQl1jv1VyrsNPnRLbMcFTFat1L8reQ/edit#heading=h.m9bk2fnzgago) 11

1. **Introduction**
   1. **Purpose of SDD**

The purpose of this project is to develop an automated fake news detection system using Natural Language Processing and machine learning techniques. The system aims to analyze and classify news articles as real or fake with high accuracy, helping to reduce the spread of misinformation and support users in accessing reliable and trustworthy information online.

* 1. **Scope**

The scope of this project is to design and develop a fake news detection system using Natural Language Processing and machine learning techniques. The system will process textual news articles, perform data preprocessing such as tokenization, stop-word removal, and feature extraction, and apply supervised machine learning algorithms to classify news content as real or fake. The project focuses on text-based analysis and does not include image, video, or social media metadata analysis. The system aims to provide accurate classification results and can be used as a decision-support tool to help identify misleading information in online news platforms.

1. **Application Architecture**
   1. **Architectural Overview**

The Fake News Detection System follows a modular, layered architecture to ensure scalability and maintainability. The system consists of the following key components:

1. **User Interface Layer**  
   Provides an interface for users to input news articles or upload text files and view classification results (real or fake).
2. **Data Preprocessing Layer**  
   Handles text cleaning and normalization processes such as tokenization, stop-word removal, stemming/lemmatization, and removal of irrelevant characters to prepare data for analysis.
3. **Feature Extraction Layer**  
   Converts processed text into numerical representations using techniques such as TF-IDF or word embeddings for model training and prediction.
4. **Machine Learning Layer**  
   Includes trained classification models (e.g., Naïve Bayes, Logistic Regression, or Support Vector Machine) that analyze extracted features and classify news articles as real or fake.
5. **Prediction & Result Layer**  
   Generates classification results along with confidence scores and sends them back to the user interface.
6. **Data Storage Layer**  
   Stores datasets, trained models, and prediction logs for future reference and system improvement.

This layered architecture ensures separation of concerns, efficient data processing, and accurate classification of news articles.

1. **Data Model Schema**
   1. **Dataset Structure**

The Data Module is responsible for managing all data-related operations required for the fake news detection system. It includes the following components:

1. **Dataset Collection Module**  
   Stores labeled news articles (real and fake) used for training and testing the machine learning models. Each record contains the news text, label, and optional metadata such as source and publication date.
2. **Data Preprocessing Module**  
   Handles text cleaning operations including tokenization, stop-word removal, lowercasing, stemming or lemmatization, and removal of punctuation and special characters.
3. **Feature Storage Module**  
   Stores extracted features such as TF-IDF vectors or embeddings generated from the preprocessed text for efficient model training and prediction.
4. **Model Data Module**  
   Maintains trained machine learning models and related parameters for reuse during classification.
5. **Prediction Data Module**  
   Logs input news articles and their classification results (real or fake) for evaluation and system performance analysis.

The Data Module ensures secure, organized, and efficient handling of textual data throughout the system lifecycle.

* 1. **Data Entities**
     1. **Image Data Entity**

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| image\_id | Unique identifier |
| image\_path | URL of image |
| Image name | Name of image file |
| Image size | Size of the image |

* + 1. **Prediction Result Entity**

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| Result\_ID | Unique identifier for each prediction result |
| Article\_ID | Reference ID of the analyzed news article |
| Predicted\_Label | |  | | --- | |  |  |  | | --- | | Classification result (Real / Fake) | |
| Confidence\_Score | Probability or confidence level of the prediction |

* 1. **Model Data**

|  |  |
| --- | --- |
| **Component** | **Description** |
| Model\_ID | Unique identifier for the trained model |
| Model\_Name | Name of the machine learning model |
| Model\_Type | |  | | --- | |  |  |  | | --- | | Algorithm used (Naïve Bayes, SVM, Logistic Regression, etc.) | |

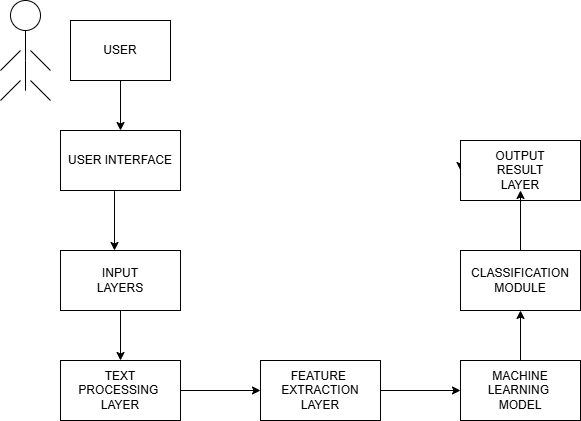
1. **User Interface Design**
   1. **UI Overview**
2. The User Interface (UI) of the Fake News Detection System is designed to be simple, user-friendly, and interactive, allowing users to easily analyze news content and receive instant classification results.
3. Main UI Components:
4. 1. Home Page
5. Provides a short introduction to the system.
6. Explains the purpose of detecting fake news using NLP and machine learning.
7. Contains navigation buttons to access other features.

2. News Input Interface

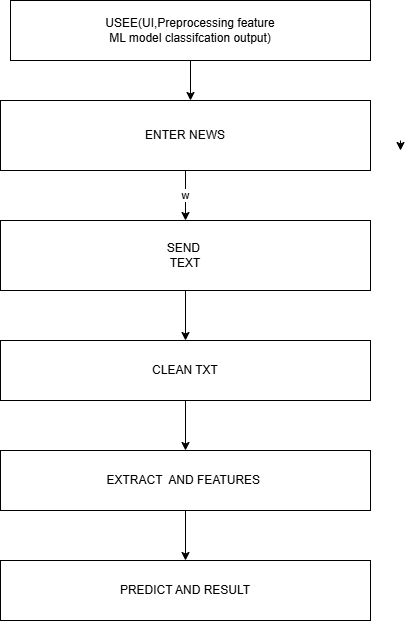
1. A text area where users can paste or type news articles.
2. Option to upload a text file (optional).
3. Ensures easy input of news content for analysis.

3. Analyze Button

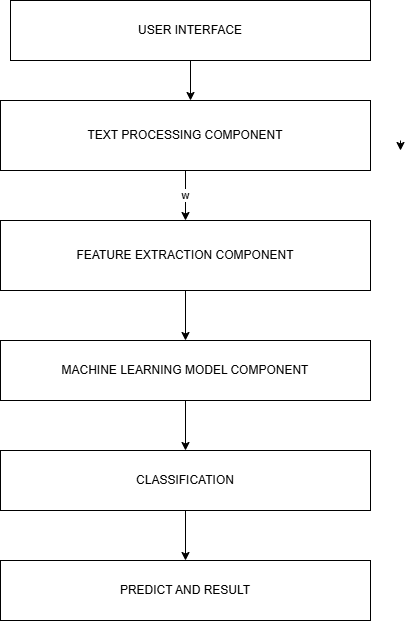
1. Allows users to submit the entered news article.
2. Triggers preprocessing, feature extraction, and classification processes.
3. 4. Result Display Section
4. Displays the classification result as “Real News” or “Fake News.”
5. May include confidence score or probability for better understanding.
6. Uses clear labels or color indicators for quick interpretation.
7. 5. Model Information Page (Optional)
8. Provides brief details about the NLP techniques and machine learning models used.
9. Enhances transparency and user trust.
10. 6. Error Handling & Validation
11. Shows alerts for empty or invalid inputs.
12. Ensures smooth interaction and reliable output.
13. UI Design Goals
14. Clean and minimal layout
15. Easy navigation
16. Fast response time
17. Accessible for non-technical users
18. **Design Diagrams**
    1. **Architecture Diagram**

****

* 1. **Sequence Diagram**

****

* 1. **Component Diagram**

****

1. **Conclusion**

The Fake News Detection System successfully demonstrates the application of Natural Language Processing and machine learning techniques to identify and classify news articles as real or fake. By performing effective text preprocessing and feature extraction, the system is able to analyze news content accurately. Trained machine learning models enable reliable classification with high accuracy, helping to reduce the spread of misinformation. The system is user-friendly, scalable, and can be further enhanced with advanced deep learning models and real-time data sources in the future. Overall, this project provides an efficient and practical solution to combat fake news in online media.