

Phase 1: Problem Definition and Design Thinking

In this part you will need to understand the problem statement and create a document on what have you understood and how will you proceed ahead with solving the problem. Please think on a design and present in form of a document.

Problem Definition: The problem is to develop a fake news detection model using a Kaggle dataset. The goal is to distinguish between genuine and fake news articles based on their titles and text. This project involves using natural language processing (NLP) techniques to preprocess the text data, building a machine learning model for classification, and evaluating the model's performance.

Design Thinking:

1. **Data Source:** Choose the fake news dataset available on Kaggle, containing articles titles and text, along with their labels (genuine or fake).
2. **Data Preprocessing:** Clean and preprocess the textual data to prepare it for analysis.
3. **Feature Extraction:** Utilize techniques like TF-IDF (Term Frequency-Inverse Document Frequency) or word embeddings to convert text into numerical features.
4. **Model Selection:** Select a suitable classification algorithm (e.g., Logistic Regression, Random Forest, or Neural Networks) for the fake news detection task.
5. **Model Training:** Train the selected model using the preprocessed data.
6. **Evaluation:** Evaluate the model's performance using metrics like accuracy, precision, recall, F1-score, and ROC-AUC.