Model Development Phase Template

Date	15 March 2024
Team ID	SWTID1728136330
Project Title	Fake News Analysis in Social Media using NPL
Maximum Marks	5 Marks

Model Selection Report

Fake news detection involves classifying news articles as either fake or real based on their textual content. In this task, we compare two models—**LSTM** (Long Short-Term Memory) and **Logistic Regression**—to determine the most appropriate model for fake news detection, considering factors such as performance, complexity, and computational requirements.

Model Selection Report:

Model	Description
Model 1: LSTM (Long Short-Term Memory)	LSTM is a type of recurrent neural network (RNN) specifically designed to overcome the vanishing gradient problem, which is prevalent in traditional RNNs. LSTMs are well-suited for sequential data, such as text, because they can capture long-range dependencies and contextual information. In fake news detection, LSTM can learn the underlying patterns in news articles by analysing the sequence of words and their contextual relationships. • **Architecture:** LSTM networks consist of layers of LSTM cells, which use gates to regulate the flow of information. The network is typically trained using backpropagation through time (BPTT), and often includes an embedding layer to transform words into dense vectors.

	• **Strengths:** **Contextual understanding:** LSTM can effectively capture long-term dependencies and the context in text, making it suitable for understanding the nuanced structure of fake news. **Handling sequential data:** LSTM is ideal for text data where the order of words matters for classification (e.g., detecting sarcasm, tone, etc.).
Model 2: Logistic Regression	Logistic Regression is a linear model used for binary classification tasks. In the context of fake news detection, it works by learning the relationship between the features (e.g., word frequencies, term presence) and the class label (fake or real news). It is a simple yet effective model for text classification when combined with feature extraction methods like TF-IDF or word embeddings. • **Architecture:** Logistic Regression computes a weighted sum of the input features and applies a sigmoid function to predict the probability of the input belonging to one of two classes. • **Strengths:** **Simplicity and speed:** Logistic Regression is computationally efficient and easy to implement. **Interpretable results:** The weights of the model can be directly interpreted, which is valuable for understanding the significance of different features. **Good performance with feature engineering:** When combined with strong feature engineering techniques like TF-IDF or word embeddings, it can perform well on tasks like fake news detection.