Model Development Phase Template

Date	27 October 2024
Team ID	739894
Project Title	Toxic Comment Clssification for Social Media using NLP
Maximum Marks	5 Marks

Model Selection Report

In the model selection report for future deep learning and computer vision projects, various architectures, such as CNNs or RNNs, will be evaluated. Factors such as performance, complexity, and computational requirements will be considered to determine the most suitable model for the task at hand. The objective of this report is to evaluate various neural network architectures for future deep learning and computer vision projects. Specifically, we will assess models for toxic comment classification on social media using Natural Language Processing (NLP). The evaluation will consider factors such as performance, complexity, and computational requirements to determine the most suitable model for the task at hand.

Model Selection Report:

Model	Description
Model 1	Logistic Regression is a popular and straightforward machine learning algorithm used for binary classification tasks. In the context of toxic comment classification, logistic regression helps in predicting whether a comment is toxic or not based on the features extracted from the text. Key Components

Feature Extraction: The first step involves converting the text data into numerical features. Common techniques include:

Bag of Words (BoW): Represents text as a collection of word counts.

Term Frequency-Inverse Document Frequency (TF-IDF): Weighs the importance of words based on their frequency across documents.

Word Embeddings: Techniques like Word2Vec or GloVe map words into high-dimensional vectors capturing semantic meaning.

Model Training:

Input Features: The extracted features from the comments.

Binary Classification: The model is trained on labeled data where each comment is marked as toxic (1) or non-toxic (0).

Sigmoid Function: Logistic regression uses the sigmoid function to map the input features to a probability between 0 and 1. This probability indicates the likelihood of a comment being toxic.

Prediction: During prediction, the model calculates the probability of a comment being toxic. If the probability exceeds a certain threshold (commonly 0.5), the comment is classified as toxic.

Evaluation Metrics

Precision: Measures the proportion of true positive toxic comments among all comments classified as toxic.

	Recall: Measures the proportion of true positive toxic comments detected among all actual toxic comments. F1 Score: The harmonic mean of precision and recall, providing a balance between the two