Model Optimization and Tuning Phase Template

Date	10 April 2025
Team ID	259453
Project Title	SMS Spam Detection using NLP
Maximum Marks	10 Marks

Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase focuses on refining machine learning models for better performance. We evaluated multiple classifiers using techniques like GridSearchCV and adjusted hyperparameters such as smoothing, regularization strength, and vectorizer settings to maximize accuracy and minimize false positives.

Hyperparameter Tuning Documentation (8 Marks):

Model	Tuned Hyperparameters
Model 1	Multinomial Naive Bayes - `alpha`: 0.5 (Laplace smoothing parameter to prevent zero probability) - Vectorizer: `TfidfVectorizer` with `ngram_range=(1,2)`, `max_df=0.95`, `min_df=2` These parameters improved generalization by reducing overfitting and capturing useful word combinations (bigrams).
Model 2	Logistic Regression - `C`: 10 (inverse of regularization strength) - `penalty`: `'l2'` - `solver`: `'liblinear'`

	- Vectorizer: `TfidfVectorizer` with tuned parameters same as above Logistic Regression was tuned to balance model complexity and accuracy, helping avoid overfitting.

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Model 1	Multinomial Naive Bayes This model was selected as the final optimized model because of its **fast training time**, **high accuracy**, and **simplicity**. It performed best in terms of **precision and recall**, especially minimizing false positives (ham predicted as spam), which is crucial for SMS classification tasks.