Model Development Phase Template

Date	15 March 2024
Team ID	SWTID1728285970
Project Title	SMS- Spam Detection Using NLP
Maximum Marks	5 Marks

Feature Selection Report Template

In the forthcoming update, each feature will be accompanied by a brief description. Users will indicate whether it's selected or not, providing reasoning for their decision. This process will streamline decision-making and enhance transparency in feature selection.

Feature	Description	Selected (Yes/No)	Reasoning
Message Length	The total number of characters in the SMS.	Yes	Spam messages are often longer/shorter than typical non-spam messages.
Word Frequency	Frequency of specific words often found in spam messages (e.g., "win," "free").	Yes	Helps identify common spam keywords to differentiate spam from non-spam messages.

Special Characters	Count of special characters (e.g., $, %, @).	yes	Spam messages often use symbols to grab attention or bypass filters.

N-grams	Frequency of word sequences (bigrams, trigrams).	Yes	Spam messages may have specific word patterns (e.g., "win now," "free gift").
Sender Information	Presence of specific sender patterns (e.g., shortcodes or unknown senders).	Yes	Spam messages often come from specific types of senders.
Message Time	Timestamp of the SMS (e.g., late-night messages).	No	Spam messages may be sent at unusual times to avoid detection.