

## Model Development Phase Template

<b>Date</b>	11 March 2025
<b>Team ID</b>	740052
<b>Project Title</b>	AI-Based Intelligent Insight Extractor
<b>Maximum Marks</b>	6 Marks

### Model Selection Report:

In the forthcoming Model Selection Report, various models will be outlined, detailing their descriptions, hyperparameters, and performance metrics, including Accuracy or F1 Score. This comprehensive report will provide insights into the chosen models and their effectiveness.

<b>Model</b>	<b>Description</b>	<b>Hyperparameters</b>	<b>Performance Metric (e.g., Accuracy, F1 Score)</b>
Spacy Pipeline	spaCy is a powerful natural language processing library that can assist in extractive text summarization by providing robust linguistic features. Although it doesn't offer built-in summarization models, it is commonly used for preprocessing tasks like sentence segmentation, tokenization, stop word removal, and lemmatization.	-	Accuracy score = 75%
Word Tokenisation	Word tokenization is the process of splitting text into individual words or tokens. It is a key step in natural language processing, helping computers understand and analyze language. For example, the sentence."	-	Accuracy score = 67%
Sentence Tokenisation	Sentence tokenization is the process of dividing a block of text into individual sentences. It helps in understanding the	-	Accuracy score = 66%

	structure of the text and is often the first step in many NLP tasks. For example, the paragraph "I love coding. It is fun and creative." would be split into two sentences: ["I love coding.", "It is fun and creative."].		
HeapQ	The heapq module in Python provides an efficient way to implement heaps, which are special tree-based data structures commonly used for priority queues. It supports a min-heap by default, where the smallest element is always at the root.	-	Accuracy score = 70%