**Model Development Phase Template**

|  |  |
| --- | --- |
| **Date** | 11 March 2025 |
| **Team ID** | 740052 |
| **Project Title** | AI-Based Intelligent Insight Extractor |
| **Maximum Marks** | 5 Marks |

**Feature Selection Report Template:**

Here's a Feature Selection Table specifically tailored for an AI-Based Intelligent Insight Extractor, following the format you provided. I've assumed that the system processes structured and possibly unstructured input data (text, logs, metadata, etc.) to derive actionable insights.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | **Feature** | **Description** | **Selected (Yes/No)** | **Reasoning** | | --- | --- | --- | --- | | Document\_ID | Unique identifier for each document or data record | No | Serves only as an identifier, not useful for predictive modelling. | | Source\_Type | Type of input source (e.g., email, PDF, chat, database) | Yes | May influence the format, structure, and reliability of insights. | | Content\_Text | Raw unstructured text content | Yes | Core feature for NLP-based insight extraction; provides semantic meaning. | | Timestamp | Time when the data was generated | Yes | Useful for temporal analysis, trends, and identifying insight relevance. | | Author\_Role | Role of the person generating the content (e.g., manager, analyst) | Yes | Helps prioritize or interpret insights based on domain authority. | | Sentiment\_Score | Polarity score derived from text analysis | Yes | Indicates emotional tone, often relevant for insights on user feedback or mood. | | Keyword\_Count | Number of relevant keywords detected in the text | Yes | Higher keyword density often correlates with relevance of insight. | | Named\_Entities | Count or type of named entities extracted (ORG, LOC, PERSON, etc.) | Yes | Identifies key subjects involved in the content and context. | | Topic\_Label | Topic derived from classification or topic modeling | Yes | Allows clustering and thematic understanding of the data. | | Content\_Length | Length of the text/document | Yes | Influences processing time and may correlate with information richness. | | Language | Language of the content (e.g., EN, FR, DE) | Yes | Needed for applying correct NLP models and understanding context. | | Confidence\_Score | Model-generated score for insight reliability | Yes | Useful for filtering low-confidence predictions or uncertain extractions. | | Insight\_Category | Predefined category (e.g., Risk, Opportunity, Feedback) | Yes | Helps in grouping and routing extracted insights. | | User\_ID | ID of the user associated with the data source | No | Acts as metadata; not predictive unless tied to behavioral features. | |