**Phase-2 Submission Template**

**Student Name:** [Enter Your Name]

**Register Number:** [Enter Your Register Number]

**Institution:** [Insert College Name]

**Department:** [Enter Your Department Name]

**Date of Submission:** [Insert Date]

**Github Repository Link:** [Update the project source code to your Github Repository]

### **1. Problem Statement**

### *[Clearly articulate the real-world problem being solved, refined from Phase-1*

* *Revisit and refine the problem based on additional understanding of the dataset.*
* *Clearly define the type of problem (classification, regression, clustering, etc.).*
* *Explain why solving this problem matters (impact, relevance, or application area).]*

### **2. Project Objectives**

### *[Update the project goals now that you're entering practical implementation.*

* *Define the key technical objectives*
* *Specify what the model aims to achieve (e.g., accuracy, interpretability, real-world applicability).*
* *Mention if the goal has changed or evolved after data exploration.]*

### **3. Flowchart of the Project Workflow**

### *[Visually represent the entire workflow from start to finish.]*

### 

### **4. Data Description**

### *[Provide a short recap of the dataset and its source.*

* *Dataset name and origin (e.g., Kaggle, UCI, open APIs).*
* *Type of data: structured, unstructured, image, text, time-series, etc.*
* *Number of records and features.*
* *Static or dynamic dataset.*
* *Target variable (if supervised learning).]*

### **5. Data Preprocessing**

*[Perform and document data cleaning and preparation.*

* *Handle missing values (removal, imputation, etc.).*
* *Remove or justify duplicate records.*
* *Detect and treat outliers.*
* *Convert data types and ensure consistency.*
* *Encode categorical variables (label encoding, one-hot encoding).*
* *Normalize or standardize features where required.*
* *Document and explain each transformation step clearly in code and markdown.]*

### 

### **6. Exploratory Data Analysis (EDA)**

*[Perform detailed statistical and visual exploration of the data.*

* *Univariate Analysis:*
  + *Distribution of features using histograms, boxplots, countplots, etc.*
* *Bivariate/Multivariate Analysis:*
  + *Correlation matrix, pairplots, scatterplots, grouped bar plots, etc.*
  + *Analysis of relationship between features and the target variable.*
* *Insights Summary:*
  + *Highlight patterns, trends, and interesting observations.*
  + *Mention which features may influence the model and why.]*

### 

### **7. Feature Engineering**

*[Enhance or transform data to improve model performance.*

* *Create new features based on domain knowledge or EDA insights.*
* *Combine or split columns (e.g., extracting date parts).*
* *Use techniques like binning, polynomial features, ratios, etc.*
* *Apply dimensionality reduction (optional, e.g., PCA).*
* *Justify each feature added or removed.]*

### 

### **8. Model Building**

### *[Build and compare multiple models to solve the defined problem.*

* *Select and implement at least 2 machine learning models.*
  + *E.g., Logistic Regression, Decision Tree, Random Forest, KNN, etc.*
* *Justify why these models were selected (based on problem type and data).*
* *Split data into training and testing sets (with stratification if needed).*
* *Train models and evaluate initial performance using appropriate metrics.*
  + *For classification: accuracy, precision, recall, F1-score.*
  + *For regression: MAE, RMSE, R² score].*

### 

### 

### **9. Visualization of Results & Model Insights**

*[Use plots and charts to explain model behavior.*

* *Confusion matrix, ROC curve, feature importance plot, residual plots, etc.*
* *Include visual comparisons of model performance.*
* *Interpret top features influencing the outcome.*
* *Clearly explain what each plot shows and how it supports conclusions.]*

### 

### **10. Tools and Technologies Used**

*[Mention all tools used in this phase of the project.*

* *Programming Language: Python or R.*
* *IDE/Notebook: Google Colab, Jupyter Notebook, VS Code, etc.*
* *Libraries: pandas, numpy, seaborn, matplotlib, scikit-learn, XGBoost, etc.*
* *Visualization Tools: Plotly, Tableau, Power BI.]*

### 

### **11. Team Members and Contributions**

***[****List names and responsibilities.*

* *Clearly mention who worked on:*
  + *Data cleaning*
  + *EDA*
  + *Feature engineering*
  + *Model development*
  + *Documentation and reporting]*