



Data Science and AI Capstone Project Instructions

Objective: Select a data science problem of your interest and apply the skills you've learned to solve it. Use a dataset from Kaggle with a rating of 7 or above.

Steps:

1. **Choose a Topic:** Pick a problem that aligns with your interests, such as healthcare analytics, climate change predictions, sports statistics, business intelligence, or other domains.
2. **Select a Dataset:** Browse Kaggle for datasets rated 7 and above. Ensure the dataset is suitable for your analysis and aligns with your chosen topic.
3. **Define the Problem Statement:** Describe the problem, why it's important, and the expected outcome.
4. **Collect and Explore Data:** Examine the chosen dataset, clean it if necessary, and explore its features and potential.
5. **Data preprocessing:**
 - A. handle missing values: identify and address any missing values in the dataset using techniques such as imputation, deletion, or data interpolation. Ensuring a clean and complete dataset is crucial for accurate model training and performance.
 - B. , convert data types: Transform categorical variables into numerical formats, as many machine learning algorithms require numerical inputs. This may involve one-hot encoding, label encoding, or other appropriate techniques based on the specific dataset and problem at hand.
 - C. normalize and scale data,
6. **Perform Exploratory Data Analysis :** conduct visualization techniques such as histograms, scatter plots and heatmaps to explore the distribution of features and identify relationships between variables, statistical analysis and anomaly detection.
7. **Feature Engineering:**
8. **Model selection:** logistic regression, decision trees and random forests, support vector machines, neural networks etc
9. **Model training and evaluation:** accuracy, precision, recall and F1-score
10. **Model optimization and reporting:** hyperparameter tuning, cross-validation, visualizing results
11. **Summarize in a Report:** Write a clear and concise report with your approach, methods, results, and conclusion. Include code snippets and visualizations where necessary.

12. Prepare for Presentation: Summarize your project in a 5-10 minute presentation for the class.

Deadline: Provide the final submission link to your Github and/or Google Colab by **12th September 2025 at 11.59 p.m.**, and be ready to present on **19th September 2025**.

Evaluation Criteria: Projects will be assessed based on creativity, technical execution, and clarity of communication.

Capstone Project Grading Rubric (Total: 100%)

Criteria	Description	Weight (%)
1. Topic Selection & Dataset Quality	Relevance of the chosen problem and dataset (Kaggle rating ≥ 7), alignment with domain interest	10%
2. Problem Statement	Clarity, significance, and articulation of expected outcomes	10%
3. Data Collection & Exploration	Depth of data understanding, cleaning, and initial insights	10%
4. Data Preprocessing	Handling missing values, data type conversion, normalization/scaling	10%
5. Exploratory Data Analysis (EDA)	Use of visualizations, statistical summaries, and anomaly detection	10%
6. Feature Engineering	Creation or selection of meaningful features to improve model performance	10%
7. Model Selection & Justification	Appropriateness of chosen algorithms for the problem	10%

8. Model Training & Evaluation	Use of metrics (accuracy, precision, recall, F1-score), validation strategy	10%
9. Optimization & Reporting	Hyperparameter tuning, cross-validation, result visualization	10%
10. Final Report & Presentation	Clarity, structure, inclusion of code/visuals, and effectiveness of oral presentation	10%