

Progress Report 1, MATH 6388

Due Date: October 20, 2023

Congratulations on selecting your data set for the course project! This report is your opportunity to showcase your understanding of the data, your exploratory data analysis (EDA) skills, and to present the overall machine learning pipeline (i.e., preprocessing techniques, learning algorithms, evaluation metrics, and refinement procedures).

1. Introduction:

Begin your report with a brief introduction to your chosen data set. Include details such as the source and the nature of the data (e.g., numerical, categorical, sample size, number of features). Clearly state the problem or question you aim to address using this data set.

2. Exploratory Data Analysis (EDA):

Describe your EDA process in detail. Discuss the steps you took to clean and preprocess the data. Highlight any challenges you faced and how you overcame them. Use visualizations such as histograms, scatter plots, or heatmaps to convey key insights about the data set. Explain the patterns or anomalies you discovered and their relevance to your project.

3. Snapshot of Machine Learning Pipeline:

Outline the machine learning pipeline you plan to implement for your course project. Provide a high-level overview of each step, from data preprocessing to model evaluation. Include any algorithms or techniques you intend to use and justify your choices. Discuss potential challenges you anticipate during the implementation phase.

4. Page Limit and Grading Guidelines:

Your report should not exceed two pages. Conciseness and clarity are essential. Ensure that each section provides valuable insights into your understanding of the data set and your project's direction. Grading will consider the depth of your EDA, the soundness of your machine learning pipeline, and the clarity of your communication. Creativity in visualization and problem-solving will also be appreciated.