

Expectation Project Update

Please ensure that your presentation includes some (if not most) of the following guidelines. We expect a single PowerPoint presentation that incorporates the following:

❖ Machine Learning Presentation Checklist

Across all these requirements, make sure to use high-quality visualizations

1. Problem Outlook

- Clearly define the problem
- Explain the importance and real-world applications
- Describe the dataset (source, size, and type of data)

2. Technical Terms & Concepts

- Define and explain relevant ML terms
- How are you approaching the problem
- If the model has assumptions, make sure to mention them
- Include the goal and requests from the mentors

3. Feature Engineering

- List new features created from existing data
- Explain why these features were generated and how they improve the model
- Show examples of transformed or combined features

4. Feature Imputation (Handling Missing Data)

- Identify missing values and explain why they occur
- Describe the imputation method used (e.g., mean, median, mode, predictive imputation)
- Show a before-and-after comparison (the effect on your metrics)

5. Feature Correlation & Selection

- Show correlation heatmaps or other correlation metrics
- Explain which features are highly correlated and why that matters
- Identify and justify which features were selected or removed
- If you use a feature selection approach explain its pitfalls and how did you overcome this

6. Handling Outliers

- Define and detect outliers (e.g., Z-score, IQR method, boxplots) • Explain their impact on the dataset
- Describe how they were handled (removal, transformation, etc.)
- Are they outliers or valuable information Beware of the Simpson paradox (data belonging to different groups)

7. Deleted Features

- List features that were removed and explain why
- Show impact analysis (before and after feature removal)

8. Exploratory Data Analysis (EDA)

- Provide key summary statistics (mean, median, standard deviation, etc.)
- Include important visualizations (histograms, scatterplots, pair plots)
- Identify patterns, trends, or biases in the data
- Show maps of relevant features

9. Data Preprocessing & Scaling

- Explain any normalization or standardization techniques applied
- Justify why scaling was necessary (if applicable)

10. Model building

- Show models tested and how they compare
- Explain the assumptions of the model used and why it is used
- Discuss how the data splitting affects your results
- Show how hyperparameters affect model predictions

10. Challenges & Next Steps

- Identify any issues encountered during data preprocessing
- Suggest possible solutions or alternative approaches
- Discuss plans for model fine-tuning.