

guidance

February 7, 2025

1 CAP 5768: Introduction to Data Science - Project Guidelines

1.1 Project Overview

- **Type:** Group Project (Up to 3 students per group)
- **Group Formation:** Self-formed
- **Total Points:** 50
- **Deliverables:**
 - **Project Proposal:** 10 points (**Suggest Using ACM conference template**), due to [Feb 28]
 - **Midway Presentation:** 5 points (5-10 minutes), [Mar 18, 20]
 - **Final Presentation:** 10 points (20 minutes) [Apr. 8-19]
 - **Final Report:** 25 points (**Suggest Using ACM conference template**) [Apr. 23]

1.1.1 Project Scope

Students will work on a real-world data science problem, demonstrating their ability to: - Collect and preprocess data - Perform Exploratory Data Analysis (EDA) - Develop and evaluate machine learning models - Interpret and visualize results - Discuss challenges and limitations

1.2 1. Project Proposal [Feb 28]– 10 Points

1.2.1 Deliverable: A 2-3 page written proposal

Purpose: Define a clear problem statement, dataset, and methodology before proceeding.

1.2.2 Proposal Structure

1. **Title and Team Members**
2. **Problem Statement**
 - Define the research question or problem.
3. **Dataset Selection**
 - Describe the dataset (source, size, format).
4. **Methodology**
 - Data cleaning and feature engineering plan.
 - Intended statistical/ML methods to apply.
5. **Evaluation Metrics**
 - Define success criteria (e.g., accuracy, RMSE, AUC).

6. Expected Challenges

- Any foreseen issues in data or modeling.

1.2.3 Grading Criteria (10 Points)

Criteria	Points
Clarity of problem definition	3
Feasibility of data and methods	3
Appropriateness of evaluation metrics	2
Writing quality and organization	2

1.3 2. Midway Presentation – 5 Points

1.3.1 Deliverable: 5-10 minute presentation

Purpose: Ensure students make meaningful progress and receive feedback before final submission.

1.3.2 Presentation Structure

1. Introduction & Problem Recap
2. Exploratory Data Analysis (EDA) Results
3. Feature Engineering & Model Selection
4. Initial Model Performance & Findings
5. Challenges Faced & Next Steps

1.3.3 Grading Criteria (5 Points)

Criteria	Points
Depth of EDA and insights	2
Logical methodology and model selection	2
Clarity of challenges and next steps	1

1.4 3. Final Report & Presentation (Finals Week) – 35 Points

1.4.1 Deliverable:

- **Final Report:** 25 points (7-9 page paper)
- **Final Presentation:** 10 points (10-12 slides)

1.4.2 Final Report Structure (25 Points)

1. Title and Abstract
2. Introduction & Problem Definition
3. Dataset Description & Preprocessing

4. **EDA and Feature Engineering**
5. **Model Development & Evaluation**
6. **Results & Discussion**
7. **Limitations & Future Work**
8. **Conclusion & References**

1.4.3 Grading Criteria for Final Report (25 Points)

Criteria	Points
Depth and clarity of analysis	10
Logical structure and methodology	6
Interpretation of results and discussion	5
Writing quality and citations	4

1.4.4 Final Presentation Structure (10 Points)

1. **Introduction and Research Question**
2. **Data Pipeline Overview**
3. **Key Insights from EDA**
4. **Model Selection & Performance**
5. **Challenges & Lessons Learned**
6. **Future Work**

1.4.5 Grading Criteria for Final Presentation (10 Points)

Criteria	Points
Clarity and engagement	5
Justification of approach and results	5

1.5 Project Timeline

Week	Deliverable	Points
6	Project Proposal	10
10	Midway Presentation (5-10 min)	5
Finals Week	Final Report & Presentation	35