



Electrical and Computer Engineering Department
Machine Learning and Data Science - ENCS5341

Assignment #3
Submission deadline: 6.01.2026

In this assignment, you will use the travel-destination dataset collected in Phase 1 to build and evaluate machine learning models. You will also communicate your findings through a comprehensive written report.

You need to complete the following tasks:

1. Choose a Learning Task [10 Points]

Your first step is to select a machine learning task based on the dataset. You may choose **any task**, as long as it uses **only the collected data**. Try to be creative—more challenging or insightful tasks will be rewarded, while overly simple tasks may limit your potential.

Example tasks include:

- Predicting the **country** from an image
- Predicting the **mood/emotion** from the description
- Predicting the **weather** from the image
- Predicting the **time of day**
- Predicting the **activity** shown

You may define your own task or even combine multiple tasks.

2. Data Preprocessing and Exploratory Data Analysis (EDA) [15 Points]

Before building any models:

- Clean, preprocess, and organize your dataset.
- Perform exploratory data analysis to understand trends, class distributions, correlations, and potential issues.
- Use both **quantitative summaries** (e.g., statistics) and **visualizations** (e.g., histograms, scatter plots).

3. Baseline Model [10 Points]

As a baseline model, evaluate a nearest neighbor baseline using a distance of your choice. Report the performance of this baseline using **both k=1 and k=3**.

4. Proposed Machine Learning Models [30 points]

Develop and evaluate **two additional ML models** on your task.

For each model:

- Explain why you selected it.
- Report its performance.
- Tune at least **one hyperparameter**, testing **at least four different values**.
- Discuss whether and why performance improved (or did not improve) compared to the baseline.

You are encouraged to explore different model families (e.g., tree-based models, linear models, SVMs, CNNs, etc., depending on your task).

5. Performance analysis [20 Points]

Analyze the behavior of your **best-performing model**:

- Examine test examples with **errors** (misclassifications or high errors for regression).
- Look for **patterns** in these mistakes:
 - Are errors associated with certain labels?
 - Are some categories or conditions particularly difficult?
 - Are there potential issues with the data itself?

Provide thoughtful, data-driven insights.

6. Final Report (max 8 pages) [15 Points]

Along with your **code**, you must submit the **cleaned dataset**, and a **report**. The quality and clarity of this report will form a major part of your grade. Your report should include the following sections:

- **Introduction**

- Describe the task you selected.
- Summarize the models you explored.
- Define the evaluation metrics used.

- **Exploratory Data Analysis**

- Present key steps from your EDA.
- Include descriptive statistics and visualizations.
- Highlight interesting trends or challenges discovered in the data.

- **Experiments and Results**

- Describe your baseline model, proposed models, and hyperparameter tuning.
- Report all results clearly (e.g., tables, plots, metrics).
- Interpret and compare the performance of each model.

- **Analysis**

- Discuss how you analyzed model errors as required in Task 5.
- Present meaningful observations and insights.

- **Conclusions and Discussion**

- Summarize your findings and main takeaways.
- Discuss limitations of your models, your dataset, or your evaluation metrics.
- Briefly suggest possible improvements or future directions.

Remark:

All components will be assessed based on both the reported materials and the accuracy of the implementation. The 15 points allocated for the report are specifically for the quality and structure of the report itself. Therefore, it is essential to prepare a detailed, well-organized, and clearly structured report.