Project Overview

# Tasks

## Compare Performance of Algorithms

Classification: Apply and compare Logistic Regression and K-Nearest Neighbors (K-NN) on a reasonable-sized classification dataset.

## Regression

Measure the performance of a Linear Regression model on a reasonable-sized regression dataset.

## Data Analysis

Summarize the characteristics of the chosen dataset(s) using descriptive statistics.

## Reporting

Write a report covering your data collection and pre-processing methods, dataset statistics, methodology, evaluation strategy, and results.

Identify a clear topic or problem relevant to machine learning or artificial intelligence (examples include housing price prediction, disease detection, spam email detection, and anomalous transaction identification).

## Develop a Solution

Identify and collect data (using existing, publicly available datasets if possible) to perform experiments.

Specify data collection strategies for data-driven applications using machine learning.

## Compare Models

Apply and evaluate Logistic Regression and K-NN on the classification dataset.

Apply Linear Regression to the regression dataset.

Use appropriate evaluation metrics to measure model performance (e.g., RMSE, RSS, Sensitivity, Specificity).

## Error Analysis

Perform a comprehensive manual error analysis on model outputs.

# Project Deliverables

## Final Report:

Submit a PDF report, formatted in the IEEE conference style (4-5 double-column pages).

## Source Code and Datasets:

Provide all code and datasets used in a compressed ZIP file.

# Key Requirements

## Data Requirements:

Use a dataset for predictive analytics tasks with a clearly identifiable response variable.

The dataset should have at least 10,000 rows.

## Number of Methods:

Use and evaluate two classification methods (Logistic Regression and K-NN) and one regression method (Linear Regression).

Generate descriptive statistics for each dataset.

## Performance Metrics:

Compare models using multiple performance measures (e.g., RMSE, RSS, Sensitivity, Specificity).

# Project Report Structure

The report should follow the IEEE format and be 4-5 double-column pages. Here’s a suggested structure:

## Abstract:

100-150 words providing a high-level description of the project, key findings, and dataset domains.

## Introduction:

Motivate the work, describe project objectives, and provide a concise overview of subsequent sections.

## Methodology:

Describe your approach to solving the problem, including data collection (if applicable) and data pre-processing.

Discuss key preliminary aspects, such as dataset preparation and pre-processing methods used.

## Results and Discussion:

Explain your evaluation methodology, including chosen performance measures and their relevance.

Discuss results in detail, including implications, limitations, and insights.

Discuss any sampling methods used.

## Conclusions and Future Work:

Summarize findings, note limitations, and suggest potential extensions or improvements.

## References:

Include all references in IEEE style.

# Potential Data Sources

You can use datasets from these sources, among others:

Statista: https://www.statista.com

European Data Portal: https://data.europa.eu/

UK Open Government Data: https://data.gov.uk

Central Statistics Office, Ireland: https://www.cso.ie

Ireland’s Open Government Data: https://data.gov.ie

Run My Code: https://www.runmycode.org

Amazon Public Datasets: https://aws.amazon.com/datasets

Google Public Data Directory: https://www.google.com/publicdata/directory

UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/

Zenodo: https://zenodo.org

Dublinked: https://data.smartdublin.ie

Data.gov: https://www.data.gov

Quandl: <https://www.quandl.com>

| **Criteria** | **High H1 (80-100)** | **H1 (70-79)** | **H2.1 (60-69)** | **H2.2 (50-59)** | **Pass (40-49)** | **Fail (<40)** |
| --- | --- | --- | --- | --- | --- | --- |
| **Objectives and Motivation** (15%) | Very challenging project objectives are well-presented, met, and thoroughly motivated. It is hard to find a fault in the approach. | Challenging project objectives are well-presented, met, and thoroughly motivated. All steps are rigorously studied and implemented. | Appropriate project objectives are well-presented, met, and motivated. All steps are rigorously studied and implemented. Some minor shortcuts or errors may be present. | Appropriate project objectives are presented, mostly met, and motivated, but some key decisions lack depth. | Clear objectives are partially met. All steps are appropriately applied but lack depth, with potential significant mistakes in approach. | Project objectives are not discernible and/or objectives are not met. |
| **Methodology** (30%) | All key decisions are appropriately justified. The project extends well beyond applying models to datasets, thoroughly investigating a range of situations for a rich understanding of performance. | All key decisions are appropriately justified. The project extends beyond applying models to datasets, investigating a diverse range of situations for a rich understanding of performance. | Most key decisions are appropriately justified. The project extends beyond applying models to datasets, making a good attempt to investigate a range of situations for better understanding. | Some key decisions are justified, but more depth is needed. The project only extends beyond applying models to datasets to some extent, and needs more differentiated evaluation for better understanding. | Some key decisions are justified but lack depth. The project only marginally extends beyond applying models to datasets, with limited differentiated evaluation. | Key decisions are unjustified, and the project lacks depth in several key aspects. |
| **Results and Discussion** (30%) | Insightful conclusions appreciate the project’s limitations and implications, grounded in relevant literature. Future work is well-conceived and thoroughly discussed. | Insightful conclusions appreciate the project’s limitations and implications, grounded in relevant literature. Future work is well-conceived. | Implications and limitations are well understood, with key takeaways highlighted. Future work is appropriate but could benefit from more depth and creativity. | Implications and limitations are understood but lack depth. Future work is appropriate but lacks creativity. | Implications and limitations are not well understood. Future work is appropriate but lacks depth and creativity. | Implications and limitations are poorly understood or missing. Future work seems arbitrary or inconsistent with project findings. |
| **Conclusion and Future Work** (10%) | Exceptionally well-written and presented, with no mistakes in formatting or referencing. | Well-written with no major language errors. All figures are clear and readable. IEEE template is followed, length limits adhered to, and references are correct. | Appropriate future work is discussed. Document has few language or style errors. Figures are well-presented, IEEE template is followed, and references are mostly complete. | Future work is appropriate but lacks depth and creativity. Document is readable with some language or style errors. Some figures are well-presented. IEEE template is largely followed, and references are mostly complete. | Document is readable but has several language or style errors. Some figures are suboptimal. IEEE template is mostly followed, references are mostly complete. | Document has typos or poor English usage, IEEE template may be broken, figures are hard to read, and references are likely incomplete. |
| **Quality** (15%) | Thorough understanding of project findings, exceptional writing quality, no formatting or referencing errors. | High writing quality with no significant language errors, well-presented figures, adherence to IEEE template and length limits, and correct referencing. | Some language or style errors, but overall well-written. Figures are mostly well-presented, IEEE template is followed, and references are correctly used. | Readable with some language or style errors, with mostly well-presented figures. IEEE template is followed, and references are mostly correct. | Document is readable but has language or style errors, and figures may be suboptimal. IEEE template is mostly followed, and references are mostly correct. | Document has many typos or poor English usage, IEEE template may be broken, figures may be hard to read, and references are likely incomplete. |