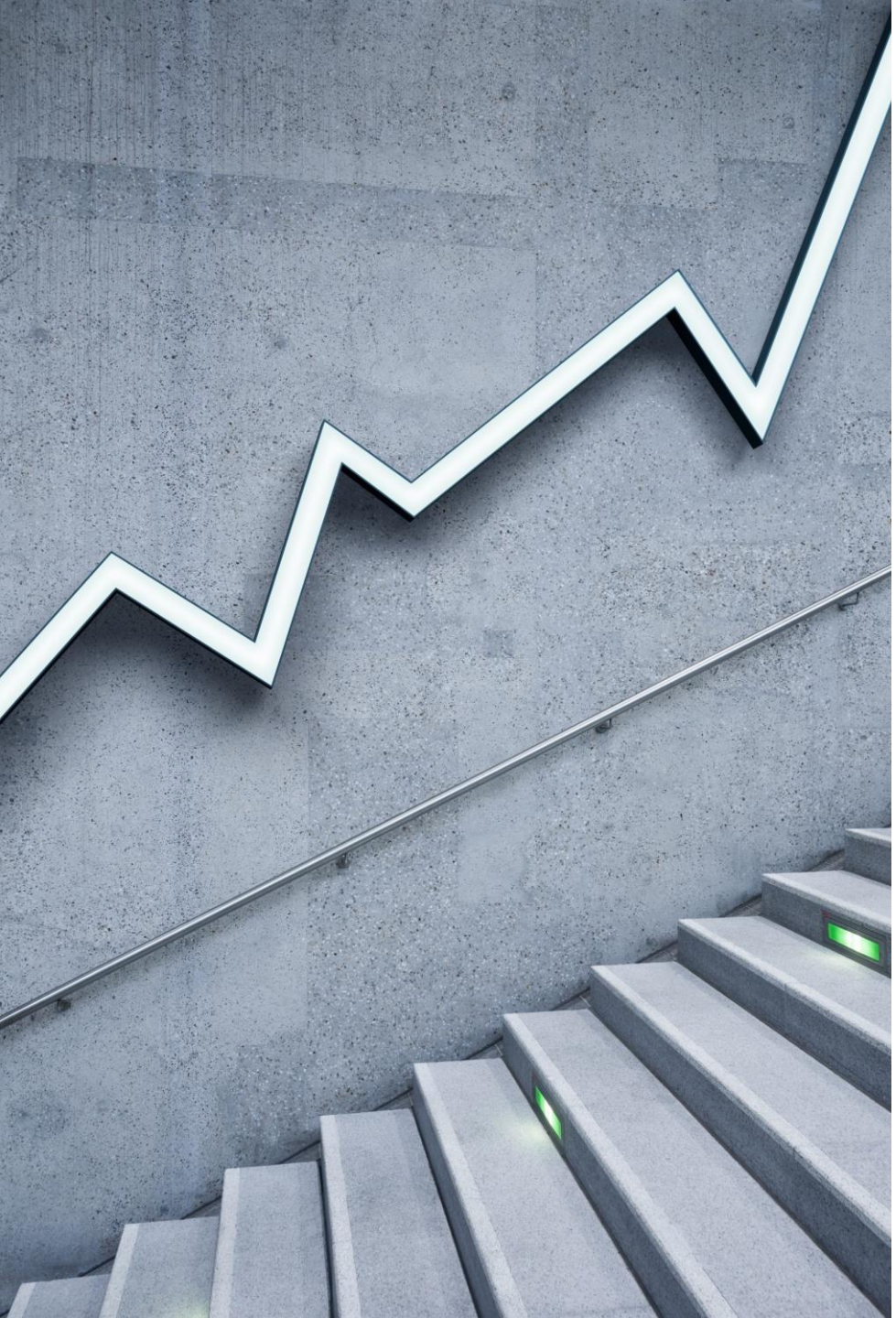

BDAT1007 Final Project

FINAL PROJECT: 40% - GROUP WORK (3 STUDENTS PER GROUP)

PART A: Dataset Selection

Each group must find a unique dataset that meets the following criteria:

- **Minimum Size**: The dataset should contain at least 1,000 records/rows.
- **Originality**: The dataset should not have been used in any prior assignments, projects, or classwork.
- **Source**: Provide the URL of the dataset source with your submission.
- **Relevance**: Ensure the dataset includes fields suitable for the analysis and techniques you plan to use.

Recommended Sources for Datasets:

- Kaggle Datasets
- Google Dataset Search

PART B: Data Cleaning

After selecting the dataset, each group should perform the following steps:

- **Data Consistency**: Ensure there are no null values in key fields.
- **Redundancy**: Remove redundant data to support accurate analysis.
- **Transformation**: Clean and transform the dataset as needed for analysis.
- **Dimensionality Reduction**: Consider applying dimensionality reduction techniques to simplify the dataset while retaining essential features, making subsequent analysis more efficient.

PART C: Implementation of Techniques

Choose and apply any three techniques covered in the course for data analysis and mining. Examples of techniques include:

- K-Nearest Neighbors (KNN)
- Logistic Regression
- Support Vector Machines (SVM)
- Decision Trees
- Naive Bayes
- Neural Networks
- Clustering
- Regression Analysis

PART C:

For each technique you select: Justify your answers.

- **Justification**: Explain why you chose this technique.
- **Target/Label**: Identify the target or label in your dataset.
- **Model Purpose**: Describe what the model does.
- **Results Explanation**: Explain the results obtained from each technique.

PART D: Visualization

Create three visualization charts to support your analysis and justify the outcomes from Part C. The visualizations should:

- Clearly represent the results and insights from your analysis.
- Be relevant and effectively convey the impact of each technique.

PART E: Results Comparison

Compare the results of the three techniques:

- **Differing Results**: Explain why results may vary between techniques.
- **Consistent Results**: If results are similar, explain the reasons.

If the results differ – why?

If the results are the same – why?

Submission Requirements

- **Deadline : Week 13**
- **Dataset URL**: Submit the source URL of the dataset.
- **PowerPoint Presentation**: Submit a PowerPoint presentation that includes:
  - **Slide 1**: Team Members – List all group members.
  - **Slide 2**: Dataset Overview – Display the dataset with all attributes.
  - Summary of Work – Explain the work done by each team member.
- **Code and Outputs**: Submit your code file along with the outputs and visualizations generated from your analysis.

Final Presentation

Final presentations will take place in **Weeks 14–15.**

Time limit : 15 minutes

Each student will get 5 minutes (total 15 minutes) for presentations.

Each student should complete one technique fully A-E.

Thank you