NAMAL UNIVERSITY
Department of Computer Science
BS Computer Science Program
6th Semester, Session (2025)
Semester Project

CS-352   **Data Analysis and Visualization**   Max Marks: 10

Project Objective

The goal of this project is to apply various data analysis, preprocessing, and visualization techniques to a dataset of your choice. You will explore, clean, visualize, and model the data using techniques learned in class. This project will demonstrate your ability to handle real-world data and extract meaningful insights, culminating in a research report formatted according to IEEE standards.

Project Phases

Phase 1: Project Proposal

- **Choose a Dataset**: Select a dataset related to your field of interest. Ensure the dataset is substantial and publicly available.
- **Submit Proposal**: Write a brief proposal including:
  - Dataset source and description
  - Research question(s) you aim to answer
  - Preliminary thoughts on potential challenges and how you plan to address them

Phase 2: Exploratory Data Analysis (EDA)

- **Data Description**: Provide a detailed description of your data including variables, types, and summary statistics.
- **Visualizations**: Create different types of plots (e.g., histograms, scatter plots, box plots, etc.) to visualize your data.
- **Initial Findings**: Discuss initial observations and insights based on your EDA.

Phase 3: Data Preprocessing

- **Handle Missing Values**: Identify and address missing values.
- **Remove Duplicates**: Identify and remove any duplicate records.
- **Outliers**: Detect and handle outliers appropriately.
- **Data Transformation**: Apply necessary transformations (e.g., scaling, encoding categorical variables).
- **Document Process**: Provide a detailed report on the steps taken and the rationale behind your decisions.

Phase 4: Correlation Analysis

- **Correlation Matrix**: Calculate and visualize the correlation matrix.
- **Significant Correlations**: Identify and discuss significant correlations in your data.

Phase 5: Time Series Analysis (if applicable)

- **Time Series Data**: If your dataset includes time-based data, perform a time series analysis.
- **Trends and Seasonality**: Identify trends and seasonal patterns.
- **Forecasting**: Apply a forecasting model if relevant.

Phase 6: Modelling

- **Regression Analysis**: Apply at least one regression model. Evaluate the model using appropriate metrics.
- **Classification Analysis**: If applicable, apply and compare different classification algorithms. Evaluate the model using appropriate metrics.
- **Clustering Analysis**: Apply different clustering algorithms. Visualize and interpret the clusters.

Phase 7: Final Report and Presentation

- **Comprehensive Report**: Compile a final report detailing each phase of your project. Include code snippets, visualizations, and findings. Format the report according to IEEE standards. Your report should include the following sections:
  - **Abstract**
  - **Introduction**
  - **Related Work**
  - **Methodology**
  - **Data Analysis and Results**
  - **Conclusion**
  - **References**
- **Presentation**: Prepare a 10-minute presentation summarizing your project, methods, and key insights.

Deliverables

- **Proposal Document**
- **EDA Report**
- **Preprocessing Report**
- **Correlation Analysis Report**
- **Time Series Analysis Report** (if applicable)
- **Modelling Report**
- **Final Comprehensive Report (IEEE Format)**
- **Presentation Slides**

Evaluation Criteria

- **Data Selection and Proposal**: Appropriateness and significance of the chosen dataset and research questions.
- **Exploratory Data Analysis**: Depth and clarity of the initial data exploration and visualization.
- **Data Preprocessing**: Thoroughness and justification of preprocessing steps.
- **Correlation and Time Series Analysis**: Accuracy and insights from correlation and time series analysis.
- **Modelling**: Effectiveness and evaluation of applied models.
- **Reporting and Presentation**: Clarity, completeness, and professionalism of the final report and presentation.

Tools and Technologies

- **Programming Language**: Python
- **Libraries**: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, etc.
- **Documentation**: Jupyter Notebooks or Google Colab

Important Notes

- Ensure all code is well-documented.
- Interpret and discuss the results of your analysis and models.
- Be prepared to explain your methods and decisions during the presentation.
- Follow the IEEE formatting guidelines for your final report. A sample IEEE template can be found [here](#).