**Data Mining Project Report**  
**Course:** IT160IU - Data Mining  
**Semester:** Spring 2025

**1. Introduction**

Provide an overview of the project, its objectives, and the methodologies employed. Include a brief explanation of data mining concepts, machine learning algorithms, and their significance in the context of this assignment.

* **Objective:** To build a data mining framework incorporating a classification/prediction model and a sequence mining algorithm.
* **Dataset Used:** Specify the dataset selected from the provided options (e.g., Production Quality Dataset from Kaggle).

**2. Data Pre-Processing**

**Objective:** Clean and prepare the raw dataset for analysis and modeling.

* **2.1 Raw Data Overview**  
  Describe the dataset, including the number of attributes, instances, and key characteristics. Present a summary table if necessary.
* **2.2 Data Cleaning Process**  
  Outline the steps taken to clean the data:
  + Handling missing values.
  + Removing duplicates.
  + Addressing outliers.
* **2.3 Data Transformation**  
  Discuss any transformations applied, such as normalization, encoding categorical variables, or feature selection.
* **Output:** Present the final cleaned dataset.

**3. Classification/Prediction Algorithm**

**Objective:** Implement a classification or prediction model using the Weka library.

* **3.1 Model Selection**  
  Explain the algorithm chosen (e.g., Decision Tree, Random Forest) and justify the choice.
* **3.2 Implementation Process**  
  Detail the steps to convert data to ARFF format and integrate Weka into the program. Mention any challenges faced during implementation.
* **3.3 Results**  
  Share initial results, including accuracy, precision, recall, and runtime.

**4. Improvement of Results**

**Objective:** Enhance the model's performance using clustering, different algorithms, or advanced data analysis techniques.

* **4.1 Methodology**  
  Explain the additional algorithm or improvement method used (e.g., K-Means Clustering, PCA for dimensionality reduction).
* **4.2 Comparison of Results**  
  Use tables or charts to compare the performance of the initial and improved models.

**5. Model Evaluation**

**Objective:** Evaluate the final models using 10-fold cross-validation.

* **5.1 Performance Metrics**  
  Present metrics such as accuracy, F1-score, and runtime for all models.
* **5.2 Analysis of Results**  
  Interpret the outcomes, discuss any trade-offs, and provide insights into the quality of the models.

**6. Conclusions**

Summarize the key findings, lessons learned, and potential future improvements. Reflect on the project objectives and whether they were achieved.

**7. References**

List all references, including:

* Dataset sources.
* Weka documentation and tutorials.
* Any additional literature or tools used.

**Appendix (Optional)**

Include supplementary materials such as:

* Code snippets.
* Detailed charts or graphs.
* Instructions for running the program.

**Submission Checklist**

Ensure the following items are submitted:

1. **Report**: report.pdf
2. **Code**: A folder containing all scripts, datasets, and executable files.
3. **Structure**: All files organized as specified in the project assignment.
4. **Testing**: Verify that your program handles both relative and absolute file paths.