CCS2213: Machine Learning
Academic Session: Semester 2, 2024-2025
SCHOOL OF COMPUTING & INFORMATICS
AlBukhary International University (AIU)

ASSIGNMENT

Task Type

Assignment 1 is an **individual assignment**.

The student is required to choose one of the listed problems/data sets in **Appendix A** and propose the solution to the problem.

You are required to **perform** experiments on the chosen algorithms.

Assignment Description

Topic Selection: See *Appendix A*.

Based on the title chosen:

1. **Data set background and characteristics**. Study the data set chosen carefully. Based on the data set:
   a. The background study of the data set. Perform the literature that covers (not limited to):
      (i) The usage of the dataset and the trend, including the proposed approaches
      (ii) Search for the literature for the past 5 years (2018 and above) that has used the chosen dataset or similar ones. The required information includes the title, the authors, year, journal/conferences, and page number.
   b. Report on the class distribution of the given dataset (dataset in Appendix A).
   c. Determine whether the data set is balanced or unbalanced. Give your justification and explain how this condition will affect performance.
2. **Pre-processing** options. Discuss and select the suitable pre-processing options.
3. **Model Evaluation Technique**. Based on the size and characteristics of the data set, choose the suitable *Model Evaluation* techniques to be used in your machine learning evaluation. Explain the reason for using that option (hold-out, cross validation etc.).
4. **Choice of classifier**. Choose **two** classifiers that you have learned in the first half of the semester.
   a. State the reasons for your choice based on the data set characteristics (and any others) and how these choices will affect performance. You may justify your choice based on the literature done in 1(a).

CCS2213: Machine Learning
Academic Session: Semester 2, 2024-2025
SCHOOL OF COMPUTING & INFORMATICS
AlBukhary International University (AIU)

ASSIGNMENT

b. Based on data set, select, and justify a suitable metric to evaluate the performance of your classification model (Confusion matrix, F1 Score etc.).

Report Requirement and Format

- A report must be prepared using font type Arial, size 12 in single line (minimum 5 pages, maximum of 10 pages, PDF document file). Every chapter should start with a **new** page (Chapter 1.0 to 5.0, and references).
- A **cover page** should contain course name (including semester and year), assignment title, name, matrix no and dataset title.
- Table of Contents

  1.0 Dataset Background
  2.0 Pre-processing options
  3.0 Model Evaluation Technique
  4.0 Choice of the classifiers
  5.0 Conclusion

  References

  Note: You may create additional subsections as deemed necessary.

Report Submission Instruction

- Submit soft copy (**pdf**) and **ipynb** to Moodle@AIU). You are required to **zip** the files.
- The Zip file must be named according to the following notation:
  CCS2213_ Name_MatrixNo_TitleNo_Lab#

Assignment Evaluation

This assignment will be graded (A to F scale).

IMPORTANT: Students who copied or plagiarized other's work or let their work be copied or plagiarized will be given an F grade. The student may be barred from sitting for final exam and reported to the university's disciplinary board.

Assignment Due Date: Refer to the course planner.

CCS2213: Machine Learning
Academic Session: Semester 2, 2024-2025
SCHOOL OF COMPUTING & INFORMATICS
AlBukhary International University (AIU)

ASSIGNMENT

Grading Rubric – Assignment

Course Learning Outcome (CLO):

- CLO2: Compare various machine learning techniques and algorithms (C4, PLO2).

Rubric (Report) - (10%)

Component	2-1 (Poor)	5-3 (Average)	8-6 (Good)	10-9 (Excellent)	Weight
Dataset set background and characteristics	The dataset description is **absent**.	The dataset description is **minimal**.	The dataset description is **adequately** complete.	The dataset description is **complete** and **comprehensive**.	20%
Pre-processing options	Pre-processing options are **minimally** discussed and justified.	Pre-processing options are **fairly** discussed and justified.	Pre-processing options are **adequately** discussed and justified.	Pre-processing options are **clearly** discussed and justified.	20%
Model Evaluation Technique	The model is **poorly** presented, and discussion of the model is **absent**. Insights and contributions are **poorly** discussed or absent.	The best-suited model is **minimally** discussed and justified. Insights from the analysis are **vague.**	The best-suited model is **fairly** discussed and justified. Insights from the analysis are **less evident,** and contribution is **fairly** discussed.	The best-suited model is **clearly** discussed and justified. Insights from the analysis evident and contribution are discussed and well-explained.	15%
Choice of the classifiers	The choice of classifiers is **minimally** discussed based on the LR and justified.	The choice of classifiers is **fairly** discussed based on the LR and justified.	The choice of classifiers is **adequately** discussed based on the LR and justified.	The choice of classifiers is **clearly** discussed based on the LR and justified.	25%

CCS2213: Machine Learning
Academic Session: Semester 2, 2024-2025
SCHOOL OF COMPUTING & INFORMATICS
AlBukhary International University (AIU)

ASSIGNMENT

Conclusion & References	The conclusion is **absent,** and no references provided.	The conclusion is of **simplistic** summary and **few** references are provided.	The conclusion is a **partially** complete summary and **adequate** references are given.	The conclusion contains a **comprehensive** summary and good references are provided.	10%
Report Formatting	Some writings are inaccurate and unclear. Follow the format given and somewhat organized.	Some writings are inaccurate and unclear. Follow the format given and somewhat organized.	Most writings are accurate, clear and concise. Somewhat follow the format and organized.	Most writings are accurate, clear and concise language used throughput. The report follows the format given and is properly arranged and well-organized.	10%

Grading Rubric – Presentation (5%)

Component	2-1 (Poor)	5-3 (Average)	8-6 (Good)	10-9 (Excellent)	Weight
Dataset set background and characteristics	The dataset description is **absent**.	The dataset description is **minimal**.	The dataset description is **adequately** complete.	The dataset description is **complete** and **comprehensive**.	10%
Pre-processing options	Pre-processing options are **minimally** discussed and justified.	Pre-processing options are **fairly** discussed and justified.	Pre-processing options are **adequately** discussed and justified.	Pre-processing options are **clearly** discussed and justified.	15%
Model Evaluation Technique	The model is **poorly** presented, and discussion of the model is **absent**. Insights and	The best-suited model is **minimally** discussed and justified. Insights from the	The best-suited model is **fairly** discussed and justified. Insights from the analysis are	The best-suited model is **clearly** discussed and justified. Insights from the analysis evident	15%

ASSIGNMENT

	contributions are **poorly** discussed or absent.	analysis are **vague.**	**less evident,** and contribution is **fairly** discussed.	and contribution are discussed and well-explained.	
Choice of the classifiers	The choice of classifiers is **minimally** discussed based on the LR and justified.	The choice of classifiers is **fairly** discussed based on the LR and justified.	The choice of classifiers is **adequately** discussed based on the LR and justified.	The choice of classifiers is **clearly** discussed based on the LR and justified.	15%
Scripting Code Walk Through	The walk through the scripting code is poor.	The walk through the scripting code is fair.	The walk through the scripting code is good.	The walk through the scripting code is excellent.	25%
Presentation Delivery	Students(s) lacks confidence and it is hard to understand what was spoken.	Students' confidence is fair. Speaks clearly and holds the attention of the listener fairly.	Student(s) confidence is good. Speaks clearly and eloquently. Emphasize important ideas and hold the listener's attention.	Student(s) confidence is noteworthy. Speaks very clearly and eloquently. Emphasize important ideas and hold the listener's attention greatly.	10%
Q & A		.	Student(s) can answer the question with confidence and clarity.	Student(s) can answer the question with confidence and clarity.	10%