TDS3851 MACHINE LEARNING

COURSEWORK ASSIGNMENT

[LAB EXERCISES (20%) + PROJECT (40%)]

TRIMESTER 3, 2024/25

© 2024 Ya Ping Wong <yp@ypwong.net>

INTRODUCTION

Welcome to the pinnacle of your Machine Learning journey this trimester! This coursework assignment aims to gauge your comprehension of concepts covered throughout the trimester and assess your ability to apply them to real-world data analysis, modeling, and prediction.

This Course Assignment comprises two integral components:

1. **Lab Exercises Enhancements (20%)**
2. **Project (40%)**

These components include reviewing and extending the lab materials in the course, as well as selecting a real-world dataset, applying diverse machine learning models, and critically evaluating their performance and accuracy.

OBJECTIVES

The objectives of the **Lab Exercises Enhancements** are:

- To thoroughly understand the lab materials and to create enhancements to the lab exercises.
- To demonstrate your understanding of the lab materials and enhance the lab materials to bring your understanding to the deeper levels.

The objectives of the **Project** are:

- To demonstrate your understanding of the course materials.
- To select suitable real-world datasets and to evaluate suitable machine learning models (including those not covered in the lab materials) to aid in analysis of the data.
- To create and train suitable machine learning models to predict patterns and trends in the data.
- To evaluate and assess the quality of the models and predictions.

DEADLINE

The submission **deadline** for all components is **12:00 noon on Thursday, 06 February 2025** (Week #14). Deadline extension is unlikely.

GROUPING

This coursework assignment is a collaborative effort for groups of AT LEAST three BUT NOT MORE THAN four students, with marks assigned based on individual contributions.

MINIMUM REQUIREMENTS

The minimum requirements of the **Lab Exercises Enhancements** are:

- Select at least 4 labs (but NOT the lab on Python) and add further materials to it so that future students of those lab materials can use them to have deeper understanding of the materials.
- Possible enhancements are:
    - To clarify certain concepts already presented.
    - To explore more advanced concept of the lab materials.
    - To provide visualizations to aid deeper understanding of the lab materials.
    - To apply the materials to real-world datasets to further demonstrate how those materials can be used.
    - To compare different between the various methods, approaches, or models.
    - To demonstrate the advantages and disadvantages of various methods, approaches, or models.
- Do not include the original lab materials unless you need to reuse some of the original codes and materials.
- Put each of the lab exercises enhancements in one Jupyter Notebook (.ipynb extension), meaning you should have at least 4 Jupyter Notebooks.
- Your Jupyter Notebooks must be executable in **Google Colab** AND **Visual Studio Code with Jupyter Extension** without the need of other codes.

The minimum requirements of the **Project** are:

- Select an **interesting real-world problem or issue in Malaysia** to be explored and analyzed to gain further understanding of the patterns or trends.
- Select open datasets freely available **about Malaysia** to achieve your goals. You may choose more than one dataset. The datasets must contain data from any year but must be up to year 2019 or later.
- Analyze, explore and visualize possible patterns and trends in the data.
- Make appropriate use of Principle Component Analysis in your analysis.
- Analyze and create suitable machine learning models to assist you to do predictions of the patterns or trends.
- Explain and justify why the predictions are useful and important.
- Evaluate and assess the quality of the models and predictions.
- Make recommendations on how to your analysis can help in solving the selected problems.

© 2024 Yap Ping Wong <yp@ypwong.net>

DELIVERABLES

Your submissions must include:

- **Jupyter Notebooks**
  - In your submission, you should have folders for your project submission with the naming convention as shown in the example below:

    1_LabEnhancements_PCA
    2_LabEnhancements_LinearRegression
    3_LabEnhancements_LogisticRegression
    4_LabEnhancements_NeuralNetworks
    5_Project

  - The numbering should reflect the sequence that reflects the complexity of the topics.
  - Make sure you do one clean run for each of the Jupyter Notebooks so that the output of the codes are embedded into the notebooks.

- **Documentations**
  - Your root folder should also contain a README.md in markdown format to describe the overview of your project submissions and the folder structures.
  - In each of the folders, you must also include a README.md in markdown format to give an overview of the particular lab enhancement or project.
  - State the full official student names (as registered at MMU) and student IDs, and state each member's contribution clearly.
  - Cite all references and data sources etc. within the Jupyter Notebooks.
  - PDF version of your Jupyter Notebooks.
  - For Project, please include an "Executive Summary" in not more than 300 words.

- **Video Presentation**
  - Prepare a video recording of around 20 minutes explaining all your submissions, longer video does not mean better.
  - You are free to choose whatever presentation format, each member of the group must participate in the presentation.
  - Upload your presentation as a private video on YouTube, and share the link in the documentations.

SUBMISSION INSTRUCTIONS

Put all your files in GitHub and share with me the link in a online Form to be made available later.

Only one member of each group needs to submit the GitHub link in the online Form.

Make sure you put your GitHub repository as **private** and add my GitHub ID (ypwong@mmu.edu.my) as your collaborator. Apply for a GitHub Education student account immediately if you have not done so.

LIVE PRESENTATION

You are required to do a short live online presentation to summarize all your submissions and for Q&A session. The presentation shall be 10 minutes only followed by 5 minutes of Q&A. The presentation shall be conducted after the submission deadline. Presentation slots in Week #14, Study Week or Exam Weeks would be given for you to book later.

© 2024 Ping Wong <yp@ypwong.net>

EVALUATION CRITERIA

Overall Evaluation Criteria:

Criteria	Lab Exercises Enhancements	Project
Understanding (30%)	*• Demonstrated comprehension of lab materials.* *• Clear explanations and insights in additional exercises.*	*• Clear understanding of chosen problem and datasets.* *• Adequate application of course materials in the analysis.*
Application & Analysis (30%)	*• Effective application of enhancements to deepen understanding.* *• Appropriate visualizations and comparisons in added exercises.*	*• Skillful use of Principal Component Analysis (PCA).* *• Thorough exploration and analysis of datasets.*
Modelling & Prediction (20%)	*• Effective application of modeling concepts in additional exercises.*	*• Appropriate selection and application of machine learning models.* *• Sound justification for predictions.*
Evaluation & Reflection (10%)	*• Thoughtful reflections on enhancements and contributions.*	*• Thorough evaluation of model performance.* *• Clear reflections on the significance and limitations of predictions.*
Presentation & Documentation (10%)	*• Clear and well-documented Jupyter Notebooks.*	*• Structured and well-documented analysis in Jupyter Notebooks.* *• README files providing clear overviews.*

Criteria	Poor (0-49%)	Fair (50-69%)	Good (70-89%)	Excellent (90-100%)
Understanding (30%)	Limited understanding and minimal contributions.	Adequate understanding with satisfactory contributions.	Clear understanding with effective contributions.	Comprehensive understanding demonstrated with insightful contributions.
Application & Analysis (30%)	Ineffective application with unclear visualizations.	Basic application of enhancements with some visualizations.	Appropriate application of enhancements with clear visualizations.	Effective application of enhancements, demonstrating a deepened understanding.
Modelling & Prediction (20%)	Ineffective application of modeling concepts.	Basic application of modeling concepts.	Appropriate selection and application of machine learning models in the project.	Successful application of modeling concepts in additional exercises.
Evaluation & Reflection (10%)	Limited evaluation and reflections.	Adequate evaluation with some reflections.	Sound evaluation and clear reflections on predictions.	Thorough evaluation of model performance and insightful reflections.
Presentation & Documentation (10%)	Poorly documented analysis with unclear README files.	Adequately documented analysis with some clarity in README files.	Structured and well-documented analysis with clear README files.	Clear, well-documented Jupyter Notebooks with insightful reflections.

© 2024 Ya Ping Wong <ypwong.mmg.net>

Pay careful attention to the evaluation criteria to maximize your marks. Good luck, and enjoy the learning process!

END OF DOCUMENT.