MONASH University

Department of Econometrics and Business Statistics
ETW3482 DATA MINING AND PREDICTIVE MODELLING

Semester 2, 2024
Predictive Modelling Project: Project Technical Report
Due date: Tuesday, 24 September 2024, 11:55 p.m.

This project is 40% of your assessment for this unit. The total number of marks for this project is 100.

Objective:
The objectives of this assignment are to apply predictive modelling knowledge to a real-world business scenario based on previously selected datasets with binary targets and prepare an industrial-format technical report for a data analytics manager to evaluate its relevancy, significance, reliability and data integrity.

Background:
Students have previously identified a dataset suitable for a predictive modelling task with a binary target and have prepared a proposal outlining their intended approach. Building on this proposal, students will develop a predictive modelling project using SAS Viya Model Studio and write a technical report detailing the predictive modelling process.

Project Scope:
- **Technical Reporting:** Students will document their predictive modelling process based on their proposal.
- **Focus:** The report will concentrate on the steps involved in preparing the dataset and developing the predictive model, excluding the evaluation and implementation phases.
- **Predictive Modelling Project:**
    - Conduct the necessary data preprocessing for the selected data.
    - The project must ideally include all types of models learned in ETW3482.
    - You could also choose to only use some of the models learned in ETW3482 (e.g. you choose to create different variants of a particular model, so you could only use a total of 3 of the 4 models that you have learned in ETW3482).
    - You could also choose to include one or two models that are available in SAS Viya but was not covered in ETW3482, although this is not encouraged.
    - The project aims to identify a champion model with a reasonable and acceptable low misclassification rate/accuracy/error.

Deliverables:

A Predictive Modeling Technical Report that includes the following components:

Report Structure:

Your technical report should include the following sections:

1. **Title Page**
   - Include the unit code and title, report title, your name, student ID, course title, and submission date.
2. **Executive Summary**
   - Summarise the chosen dataset, business problem, objective, predictive modelling approach, and potential business applications (not exceeding 10% of the maximum word count).
3. **Table of Contents**
   - List sections and subsections with page numbers.
4. **Introduction**
   - Reiterate the business problem and the objective of the predictive model.
   - Outline the report structure, indicating how it builds upon the previously submitted proposal.
5. **Data Description**
   - Provide a detailed description and an overview of input variables and the binary target of the dataset identified in the previous assignment, including the inherited data issues due to the collection process or operation nature.
6. **Methodology**
   - Explain the data preprocessing conducted, including data separation, data cleaning, imputation, feature engineering, or any necessary transformation(s) that were done in processing the data.
   - Describe the methods used for feature selection and the rationale behind these choices.
   - Document the predictive modelling techniques explored, detailing the selection process for the final model.
7. **Model Development**
   - Outline the model development process, including algorithm selection, training, and parameter tuning.
   - Discuss how the binary nature of the target variable influenced the modelling approach.
8. **Discussion**
   - Analyse how the model addresses the business problem.
   - Reflect on the model's alignment with business goals and potential integration into business processes.
9. **Conclusion**
   - Summarise the predictive modelling process and its business implications.
10. **References (if applicable)**
    - List any sources or literature referenced in the latest APA format.
    - This section can be omitted if you have not cited any references from any sources throughout your report.
11. **Appendices (if applicable)**
    - Include any supporting material.

- This section can be omitted if you have no materials to be included as an appendix(ces).

Report Formatting:

- The report should be professional, clean, and adhere to the assessment criteria.
- Please use the following formatting options throughout thre body of your report (i.e. all sections of the report excluding the front cover and appendix):
  - Line spacing: 1.5
  - Alignment: Justify
  - Font: Calibri (Body)
  - Font size: 12
  - Page numbering: Bottom of page (centre) & style: Page xx of xx
- Visual aids such as charts and graphs should be included to enhance the report. Lack of the appropriate visual aids (e.g. charts and graphs) will be penalized accordingly.
- The report should be written clearly and concisely, and the level of technical language must be suitable and appropriate for business technical personnel.
- The word count for this report **should not exceed 2000 words (excluding the title page, table of contents, references, and appendices). Please state the total word count at the end of the technical report. Anything beyond 2000 words will NOT be read and marked.**

Assessment Criteria:

- **Clarity and Depth of Technical Reporting:** The report should clearly and thoroughly describe the predictive modelling process and how it relates to the business problem.
- **Application of Predictive Modeling Techniques:** Demonstrate an understanding of the appropriate use of predictive modelling techniques, given the characteristics of the dataset.
- **Professional Presentation:** The report should be well-structured, emphasising readability and presentation quality. Be sure to use a formal writing style and appropriate technical terms as this is a formal business technical report.
- **Feasibility and Business Alignment:** The proposed model should be feasible for implementation and aligned with business objectives.

Note: Always refer to the marking rubric for this assignment that has been uploaded on Moodle for guidance on each evaluation item stated in the rubric.

Submission:

Submit the report as a **PDF file** using the following naming convention: **ETW3482_PredMod Tech Report_Student Name**

Deadline:

All submissions must be made by **11.55 p.m. on Tuesday, 24 September 2024** through the **Moodle submission link**.

MONASH University

Additional remarks:

1) While evaluation and implementation are not the focus of this report, please design your model with these aspects in mind for the next assignment (Assignment 3: Model Assessment Assignment). **The focus should be on developing a robust model that is ready for the subsequent stages of deployment and evaluation.**

2) Assignment 3: Model Assessment (40% weightage) will be released the week after this assignment is due, sometime between 30 September and 2 October 2024. You will be given 10 days to complete Assignment 3. Assignment 3 will include several short-answer questions that you will need to answer based on the pipeline you constructed in Assignment 2 and the results obtained.

3) Before sending any questions via email or requesting appointments, please watch the detailed briefing video for Assignment 2, which has been posted on Moodle. The video addresses many common and frequently asked questions.

4) For further clarifications related to the assignment, please contact your tutors: Ms. Jaya Jothi (Tutorials 02, 03, & 04) or Prof. Ganeshsree (Tutorial 01).

5) Alternatively, you can contact Prof. Ganeshsree (the CE for this unit) via email for quick questions or request a Zoom or face-to-face appointment (subject to availability). If you wish to meet for an in-person or Zoom session, it is recommended to come in groups so that everyone can ask their questions, and others can benefit from the responses. This approach saves time for everyone.

GOOD LUCK!

~~~THE END~~~