

Introduction

This document contains a description of the work you must complete for your CA 2 Predictive modelling work. You will analyse your data variables, build the predictive model, and then analyse and evaluate the model before forecasting output from it. Finally you will critique the results from the model.

Use the following section headings in your CA 2 document.

Abstract

Complete your abstract when you have all other sections finished. Describe in a few paragraphs what problem you are examining. and describe what your research question will answer. Then briefly describe the predictive model you created. Finish the abstract by describing your model validation, as well as your model forecasting and appraisal. And describe in detail what the results of the predictive model indicate. The abstract should be approx. 250 – 300 words.

Research question

In this section you must define a research question you will attempt to answer using the variables within your dataset. And you must include a discussion on the research question and how it can be answered by the dataset variables you selected. The research question must be suited to the content of the data variables in your dataset. And you must use multiple variables to answer your research question.

Building the predictive model

Describe in detail all the steps you completed when validating the suitability of your variables to fit with the model assumptions. Include numerical values of outputs from the processes you used. And include relevant charts to strengthen your discussions when appropriate. Discuss each of the model assumptions in detail, and clearly show whether each of your variables meet with the model assumptions.

Model validation

Once your model is constructed, you must evaluate and describe its accuracy. The validation method should use training and testing data to evaluate model performance. Describe in detail what the test results show by referencing and describing the test result outputs.

Model forecasting

Devise several tests that contain relevant variations in dependent variable inputs for your model that will help you to evaluate if it is accurate or not. And then provide predictions of these tests from the model. Critically appraise each of these test predictions in detail. Discuss what these predictions indicate.

Evaluate in detail whether your model proves / disproves your research question that you set out to investigate using the model.

Conclusion

Conclude your predictive modelling by discussing in detail your work including the predictive model selection, the validation you performed, the forecasting outputs it produced, and any relevant points of interest you found in your analysis.

Breakdown of assessment marks

This CA assessment is worth 40% of the total marks for the Data Science module.

The maximum available mark for each section is shown below.

Section	% mark
Abstract	5
Research question	5
Building the predictive model	20
Model validation	20
Model forecasting	20
Conclusion	20
Neatness, readability, document flow, references	10
Total	100

Important Information

Plagiarism will not be accepted and will result in an automatic mark of zero.

If you use references, the Harvard referencing must be adopted. Please use the following link which might help you create the references required:

<http://www.neilstoolbox.com/bibliography-creator/>.

Late submissions will not be accepted without a valid medical certificate.

Any deviation from the above project specification must be approved by myself before submission.

Due Date: 23:59 pm on Sunday 30th May.

Work submitted after this date and time will incur a late penalty. Refer to the cover sheet for late penalty information.

You must submit your work through Blackboard. Submit your work as a pdf document. A cover sheet must be first page of your final submitted document. Use the cover sheet on Blackboard.

Include a link to the GitHub repo containing your R code.