MXB344 WIL Project

# Task Overview

This is a Work Integrated Learning (WIL) project where you work in teams in collaboration with analysts and managers from the Bank of Queensland (BoQ). This project will assess your ability to work as a team to apply generalised linear modelling to create a reproducible piece of analysis of real-world customer loan data. The analysis will culminate in the proposal of a credit risk model and an evaluation of its performance using modern industry standard measures and techniques.

The way you will work will closely mirror that of real analysts working in the financial sector. You will be expected to summarise your findings for executives, document your analysis in a reproducible way, and review your performance against your measures of success.

## Task Summary

Task: group modelling task

Due date:

1. Project plan due 11:59pm Friday 6th September 2019 (Week 7)
2. Group oral presentations in week 12 and 13 in the lecture time (9-11am, Friday 18th and 25th of October)
3. Technical report, executive summary and project plan review due 11:59pm Friday 1st November 2019 (Week 14)

Weighting: 80%

Individual or group: groups of 3-4

## Unit Learning Outcomes Addressed:

1. Expertly and critically carry out statistical analysis using statistical models in the analysis of various data sets and examples.

2. Use R to carry out statistical analyses.

3. Communicate statistical conclusions clearly and concisely both in written form and orally.

# Deliverables

1. A project plan in pdf or .docx format (10 marks) – due 11:59pm Friday 6th September 2019 (Week 7)

2. A group oral presentation of maximum 15min duration with 5min for questions (25 marks) – (9-11am, Friday 18th and 25th of October)

3. An executive summary document of maximum 2 pages length in pdf or .docx format (10 marks) – due 11:59pm Friday 1st November 2019 (Week 14)

4. Technical report describing the analysis, methodology and conclusions in pdf format (50 marks) – due 11:59pm Friday 1st November 2019 (Week 14)

5. Project plan review in pdf or .docx format (5 marks) – due 11:59pm Friday 1st November 2019 (Week 14)

6. Zip file containing the RMarkdown code files, the data, and any other supplementary information. (**Code must be provided or you will receive 0 marks for assignment 1**)

Each item must be uploaded to its corresponding Blackboard TurnItIn link by the due date.

# Background and Objectives

You are a team of analysts working for peer-to-peer personal lending startup that has been recently acquired by a regional Australian Bank (BoQ). After much back-slapping and champagne cork popping it has come to light that the risk models used by your company previously were too ad-hoc to be suitable for use in a bank that is subject to strict regulatory requirements. Your new management has ordered your team to do a ‘ground-up’ rebuild of your credit risk (loan default) models using Generalised Linear Models (GLMs) and the tools, methods and performance measures that you have learned.

You have access to your historical lending data, which you will need to utilise to build your new credit risk model.

With regards to your new credit risk model your management has a number of primary concerns:

1. What are appropriate approaches to modelling credit risk and what is the current state-of-the-art in this arena?

2. How does this new model perform compared to the one you used previously? How can it be expected to perform on new loans? There are some performance benchmarks available in the project folder on Blackboard.

3. What are the important variables in this model and how do they compare to variables the bank has found to be traditionally important in its own modelling?

4. What assurances and justifications can you make about the statistical rigor of your model and modelling methodology?

# Groups

You are free to choose your own group of 3-4 students and must register that on Blackboard by 11:59pm Friday 9th August 2019 (Week 3). After that you will be automatically assigned to a group; if your group has less than 4 students than one of the remaining students may be assigned to your group. Please contact the lecturer if you have troubles forming a group.

# Data

Your IT team has provided a database export of your lending activity between 2007 and 2011, see Blackboard. This is the data set on which your old model was benchmarked. There is also a data dictionary explaining the covariates in an accompanying Excel file.

# Tasks and Requirements

## Task 1 Project Plan (due Week 7)

Produce a maximum 5-page project plan for your manager outlining your proposed work schedule. This should include:

1. research objectives and requirements,
2. project milestones and deliverables,
3. work tasks, including who in the team is leading or contributing to each task,
4. scheduling, critical path analysis and constraints

You could use RMarkdown for this document; however, Microsoft Word .docx is also acceptable.

## Task 2 Oral Presentation (Due Week 12/13)

This boardroom-style presentation will give you the opportunity to experience presenting your results to a panel of senior analysts and project stakeholders. The panel will be made up of a representative from BoQ and QUT faculty member(s).

Unlike many presentations you have given during your time at university, this one will be highly interactive, and you can expect questions and interjections. Rather than exclusively present, it will be your role to facilitate a meeting where you aim to summarise your work and argue for a particular credit risk model to be adopted. You will have to keep track of time to make sure the work you want to showcase is seen by the panel.

Your time allotted with the panel is 20 minutes (strict limit), with a suggested maximum presentation time of 15 minutes and minimum 5 minutes of discussion.

The panel is interested in evaluating your proposed model to decide if it could be adopted. To do this, they will need understand the data you used and the process you went through to arrive at it. They will also need to see measures of the model’s performance and be satisfied they reflect its performance on future loan application data. Your analysis will not be concluded at this stage but you will need to come to the meeting with a proposed model, and make a presentation that addresses the manager’s four questions (see Background and Objectives).

Your presentation should be concise and flow logically. You may share the presenting role as much or as little as you deem appropriate, but all team members must be present and respond to questions directed at them. You will present a united front as a team of analysts and work together to respond to questions. You may wish to assign duties to each team member: timekeeper, note-taker, presenter… etc, to ensure the best possible outcome.

During the course of the presentation the panel may offer suggestions to improve your model or make your analysis more robust. You will make note of these suggestions and address them in your final report.

You may use any software to create your presentation but the slides must be submitted as part of the project zip file.

## Task 3 Executive Summary (Due Week 14)

Produce a maximum 2 page summary for your manager that is understandable by a highly trained individual who is not necessarily a statistician. Marks will be deducted if you go over the limit. Use plots to make your point clearly and succinctly where appropriate. This document must:

1. summarise the problem and context,
2. summarise the data collected,
3. summarise the reviewed literature and justify the chosen method,
4. summarise the results, clearly communicating uncertainty (e.g. state confidence intervals for covariate coefficients and show confidence bounds on plots),
5. summarise the validity/fit of the model, and
6. relate the results to the manager’s four questions above (see Background and Objectives) and justify your recommendations with evidence from your analysis.

This is not the technical report, it is an executive summary that captures clearly and concisely the findings of the study and their impact in relation to the objectives/manager’s questions. In practice, busy managers and executives rarely read the technical report in their entirety so it is crucial to be compelling, concise and convincing in the executive summary.

You should use RMarkdown for this document; however, Microsoft Word .docx is also acceptable.

## Task 4 Technical Report (Due Week 14)

The overarching task is the creation of a statistical model for loan default. You will use binomial regression within the generalised linear modelling framework to build this model. Models like the one you will build are used by money lending organisations to assign probability of default to loan applications. This information allows effective management of a lending (credit risk) portfolio.

You will need to:

1. understand and clean the data, documenting decisions on how you handled missing and/or unusual data,
2. perform exploratory data analysis to guide modelling and analysis,
3. review the literature to understand how credit risk is currently modelled and what is the state-of-the-art,
4. apply variable selection algorithms or heuristics to narrow down the large number of covariates for modelling,
5. fit binomial GLMs selecting among three link functions: logit, probit and cloglog,
6. select a model from a space of candidate models using goodness of fit and cross validation measures (including AIC, Gini/AUC),
7. test and validate the assumptions of the model,
8. evaluate the performance of the model against the existing benchmark and use cross-validation,
9. draw conclusions about the effects of important variables and make statements of the uncertainty relating to the effects,
10. combining all of the above in a reproducible report written in RMarkdown.

The structure of the technical report should be as follows:

* Introduction - describe the problem and context,
* Literature review – describe what other people do, what are the pros and cons of their approach,
* Data – summaries, exploratory data analysis and the like,
* Methods – modelling method, variable selection, validation
* Results – model coefficients, assumptions and validation under uncertainty,
* Discussion – model interpretation and use evidence from models to answer the four manager’s questions (see Background and Objectives)
* References – you can use EndNote or BibTeX or the like to help you manage references.

You **must** use RMarkdown for this document. Maximum page length for the entire report including references is 20 pages – marks will be deducted if you go over the limit. However, you may include supporting information in appendices and these are not included in the page count, i.e. no page limit on appendices. Typically, supporting information that is not central to your argument goes into appendices. No marks are allocated to appendices. You should also include a cover page with your title, date, author list (include student number) and organisation.

## Task 5 Project Plan Review (Due Week 14)

This is a short document of up to a maximum of 5 pages in length reviewing how your project went and which deliverables were achieved. Marks will be deducted if you go over the limit. Reflect on your project plan from week 7 and discuss:

1. what was achieved – include a table listing your original deliverables and milestones and add a column showing what was achieved and when,
2. who contributed to which items of work,
3. reflect on any changes to the project plan, why it happened and how it went, and
4. discuss how the project could be better managed in the future.

# Useful Resources

1. “Story telling with data”, presentation at Google: <https://www.youtube.com/watch?v=8EMW7io4rSI>

2. Making presentations with Rmarkdown: <http://rmarkdown.rstudio.com/ioslides_presentation_format.html>

3. R data wrangling cheatsheet: <https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf>

4. Lecture notes from week 7