Actuaries Institute Australia

DATA SCIENCE PRINCIPLES

SYLLABUS 2023

Education Program
ACTUARIES INSTITUTE

Education that Counts

1. Overview and Aim

This subject forms part of an actuary's toolkit, to quantify uncertainty and solve problems in a range of business environments.

The aim of this subject is to extend students' knowledge of modern data science tools and techniques beyond those introduced in the Foundation Program subjects and to teach students how to apply this knowledge in real-life business settings, preparing them for more complex and practice specific applications which will be taught in the Fellowship Program.

2. Student Outcomes

After successfully completing this subject, students will be able to contribute to all stages of a data science project across multiple industries or domains, including:

- explaining where and how their data science work can add value to the business environment and strategy;
- sourcing, interpreting, evaluating and preparing data for modelling;
- using judgement to select appropriate predictive analytic techniques for a given business problem;
- applying predictive analytic techniques to solve regression and classification problems;
- evaluating and comparing performance of different models; and
- communicating findings to a range of audiences.

3. Prerequisites

Students will have attempted (but not necessarily passed) Foundation subjects Actuarial Mathematics (CM1), Actuarial Statistics (CS1) and Risk Modelling and Survival Analysis (CS2), or their equivalents, before attempting this subject. A good understanding of the following topics is particularly important:

- CM1 topic 1 – data and basics of modelling
- CS1 topic 2.1 – exploratory data analysis
- CS1 topic 3.1 – estimation and estimators
- CS2 topic 4 – regression theory and applications
- CS2 topic 5 – machine learning

Students will also need a basic understanding of one programming language and each university will choose their preferred programming language (and ensure students have a basic understanding of that programming language prior to undertaking this subject).

4. Assessment Skill Level

Assessment of this subject will be split across the following skill levels:

- Simple Application (20%): demonstration of a detailed knowledge and understanding of the topic
- Application (50%): demonstration of an ability to apply the principles underlying the topic within a given context; and
- Higher Order (30%): demonstration of an ability to perform deeper analysis and assessment of situations, including forming judgements, taking into account different points of view, comparing and contrasting situations, suggesting possible solutions and actions and making recommendations.

5. Assessment Method

Each university decides on the method of assessment

6. Learning Objectives

1. Business Environment (5%)

1.1. Justify the importance of having a good understanding of the business, including its objectives, constraints and operating environment, when designing and implementing a data science project.

1.2. Explain the key iterative steps involved in building a model (business understanding, data understanding and preparation, modelling, evaluation, communication and deployment).

1.3. Explain how the model building process is a specific application of the Actuarial Control Cycle.

2. Data understanding and preparation (30%)

2.1. Data exploration

2.1.1. Explain what is meant by domain knowledge and discuss how it is used to interpret data and make feature selections.

2.1.2. Explore features of data using data visualisation.

2.1.3. Explain the purpose of dimension reduction and contrast the advantages and disadvantages of a number of dimension reduction techniques.

2.1.4. Explain the difference between correlation and causation and how this influences model building and communication of results.

2.2. Assessing data quality

2.2.1. Explain common features of data, such as categorical variables, missing values, unreliable/non-validated data, outliers and high cardinality features (eg post codes, medical condition coding or similar), that may lead to problems.

2.2.2. Apply a range of techniques to assess data quality such as data visualisation, outlier detection, frequency counting (for categorical variables) and using summary statistics (for numerical variables).

2.2.3. Validate data against other sources and the same data from previous runs.

2.2.4. Explain the importance of talking to people who input and use the data to assess data quality.

2.3. Data manipulation and cleansing

2.3.1. Use a programming language to manipulate data (eg filter, merge, sort, group by and summarise).

2.3.2. Apply appropriate methods to deal with common data problems such as those listed in 2.2.1.

2.3.3. Explain how data with an imbalance between control and treatment observations can cause modelling problems and list different methods used to deal with such data.

2.3.4. Construct a balanced training set to improve modelling outcomes for imbalanced data.

3. Modelling (45%)

3.1. Explain the difference between regression and classification problems.

3.2. Explain the iterative process of defining an appropriate target response variable when building a predictive model.

3.3. Perform feature selection and transformation, taking into consideration domain knowledge, the impact of nuisance variables, information leakage, variable interactions and the treatment of highly correlated variables.

3.4. Explain and perform common techniques to split data into training, validation and testing sub-sets.

3.5. Explain and compare the advantages and disadvantages of the following predictive analytics techniques; GLM, shrinkage techniques (eg LASSO and ridge regression), and tree based models (eg random forests and GBM).

3.6. Explain what is meant by estimator bias and variance and how model selection can impact these.

3.7. Select an appropriate predictive analytics technique for a given problem, taking into account the type of problem to be solved and the relative importance of accuracy, efficiency, interpretability and scoring speed for the given situation.

3.8. Perform predictive analytics modelling using the techniques outlined in 3.5 and compare their outcomes.

3.9.	Apply a range of cross validation techniques to control for overfitting (eg k-fold, leave-one-out and nested cross validation).
3.10.	Explain potential limitations of ex-ante analysis (eg black swan events) and what they mean for the model fit process.
3.11.	Describe the main characteristics of neural networks and the circumstances in which they should be considered as an alternative to the techniques considered in 3.5.
4.	**Evaluation(15%)**
4.1.	Explain common model errors such as information leakage, collinearity and discretising error.
4.2.	Use a range of statistical measures and techniques to assess a model's performance and distinguish between those which can be used for regression and for classification.
4.3.	Assess the appropriateness and performance of a model, taking into account business context and objectives.
5.	**Communication (5%)**
5.1.	Communicate modelling results to a range of business decision making audiences, taking into account the audience's needs and relating findings back to the original business objective(s).
5.2.	Explain the benefits of reproducible analysis, including the adoption of good coding habits, version control and backup.

About the Actuaries Institute

The Actuaries Institute is the sole professional body for actuaries in Australia. The Institute provides expert comment on public policy issues where there is uncertainty of future financial outcomes. Actuaries have a reputation for a high level of technical financial skills and integrity. They apply their risk management expertise to allocate capital efficiently, identify and mitigate emerging risks and to help maintain system integrity across multiple segments of the financial and other sectors. This expertise enables the profession to comment on a wide range of issues including life insurance, health insurance, general insurance, climate change, retirement income policy, enterprise risk and prudential regulation, finance and investment and health financing.

Published October 2022

Institute of Actuaries of Australia

ABN 69 000 423 656
Level 2, 50 Carrington Street,
Sydney NSW 2000, Australia
t +61 (0) 2 9239 6100
f +61 (0) 2 9239 6170
actuaries@actuaries.asn.au
www.actuaries.asn.au