

# Advanced Topics in Predictive Analytics – Draft Syllabus

## October 2021 Draft

The following is a **draft** syllabus for this exam. It is possible that there will be small changes to this syllabus when the final version is released prior to the opening of the e-Learning modules in January 2022.

### LEARNING OUTCOMES

<b>1. Topic: Working with Data (20-30%)</b>
<b>Learning Objectives</b>
The Candidate will be able to understand basic database calculations and manipulations and communicate with those who manage databases.
<b>Learning Outcomes</b>
The Candidate will be able to: <ul style="list-style-type: none"><li>a) Explain the basics concepts of database management, in particular, extract, transform, and load (ETL) operations.</li><li>b) Describe how data collection practices and assumptions affect data quality.</li><li>c) Evaluate the quality of appropriate data sources for a problem.</li><li>d) Explain the terminology and structure of relational databases.</li><li>e) Clean and organize data by performing each of the following:<ul style="list-style-type: none"><li>• Check for outliers, both univariate and multivariate</li><li>• Validate the data with regard to internal consistency</li><li>• Handle missing data (including understanding the types of missing data) by selecting the appropriate action from deletion of the record, imputation, and adding a missing value flag.</li></ul></li><li>f) Understand how different data structures can be used in different analytical tasks, including the difference between a database and a data lake.</li></ul>

## 2. Topic: Data and Model Ethics (10-20%)

### Learning Objectives

The Candidate will be able to work with data and models in an ethical and responsible manner.

### Learning Outcomes

The Candidate will be able to:

- a) Apply a general ethical framework for working with data and models.
- b) Discuss and comply with relevant standards of practice.
- c) Discuss and comply with relevant regulations that apply to working with data and models.
- d) Explain, evaluate, and correct for analytical bias, such as proxy bias and target variable bias.
- e) Explain why being blind to sensitive or prohibited variables is not sufficient to ensure lack of bias.
- f) Adjust a model when a variable must be excluded.
- g) Describe the lifecycle of a predictive model and the data which supports it.
- h) Perform data and model governance and develop model documentation in an ethical context.

### 3. Topic: Advanced Predictive Analytics Models (30-40%)

#### Learning Objectives

The Candidate will fit a variety of models and select one appropriate to the circumstances and intended use.

#### Learning Outcomes

The Candidate will be able to:

- a) Explain, fit, evaluate, and make predictions with each of the following models:
  - Additive models
  - Linear mixed models
  - Neural networks
- b) Apply Bayesian techniques to predictive models
- c) Compare model results with those from linear and tree-based methods.
- d) Explain the benefits of and demonstrate the combination of multiple models via stacking and blending.
- e) Select and justify a modeling approach based on accuracy, explainability, stability, analytical effort, computational efficiency, and table importability, taking into account the business context of the problem.
- f) Recognize and mitigate the effect of:
  - Starting with too many variables
  - Repeated use of train/test/validate sets

#### 4. Topic: Model Explainability and Communication (20-30%)

##### Learning Objectives

The Candidate will be able to effectively communicate the results of applying predictive analytics, including the relationship between model input and output, to solve a business problem.

##### Learning Outcomes

The Candidate will be able to:

- a) Understand aspects of explainability, in particular:
  - Suitability, decomposability, algorithmic transparency, and post-hoc interpretability
  - The difference between explainability and interpretability
  - When a lack of explainability may be acceptable
- b) Communicate and justify a recommended analytics solution, including use as appropriate of:
  - Variable importance plots
  - Partial dependence plots
  - Individual conditional expectation plots
  - Shapley values
  - Lift and gain charts
- c) Explain why a model is predicting certain values for certain records
- d) Communicate in a clear and straightforward manner using common language that is appropriate for the intended audience.
- e) Structure a report in an effective manner while following standards of practice for actuarial communication.