

Advanced Topics in Predictive Analytics Syllabus

The Advanced Topics in Predictive Analytics course provides candidates with the ability to employ selected analytics techniques to solve business problems and effectively communicate the solution. After learning this material from e-Learning modules, candidates are given a take-home assessment requiring analysis of a data set in the context of a business problem and submission of a report. The report/assessment will be [graded on a pre-set schedule](#).

A thorough knowledge of probability (as covered in Exam P), mathematical statistics (as covered in VEE Mathematical Statistics), and selected models and methods for analyzing data (as covered in Exams SRM and PA) is assumed.

The learning objectives and outcomes provided on the following pages follow the four modules of the e-Learning support provided. The ranges of weights shown in the Learning Objectives below are intended to apply to the large majority of exams administered. On occasion, the weights of topics on an individual exam may fall outside the published range. Candidates should also recognize that tasks often cover multiple learning objectives, including some weight for communication and ethics in many tasks.

LEARNING OUTCOMES

1. Topic: Ethical Foundations (5-10%)
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: <ul style="list-style-type: none">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.

2. Topic: Working with Data (20-30%)

Learning Objectives

The Candidate will be able to understand basic database calculations and manipulations and prepare data for predictive modeling applications.

Learning Outcomes

The Candidate will be able to:

- a) Understand the basic structure of a data pipeline, including being able to perform the following tasks:
 - Evaluate the quality of appropriate data sources for a problem
 - Explain the difference between a database, data lake, and data warehouse
 - Describe how different data structures can be used in different analytical tasks.
- b) Explain the basics concepts of database management, in particular, extract, transform, and load (ETL) operations. Demonstrate these skills by performing the following:
 - Extract data from various file structures
 - Subset, aggregate, summarize, and otherwise modify data for specific exploratory or modeling purposes
 - Create data sets as a final product of extracting and transforming data that can be used in a predictive model.
- c) Assess the accuracy and quality of data.
 - Describe how data collection practices and assumptions affect data quality.
 - Validate the data for internal consistency.
- d) Explain the terminology and structure of relational databases and be able to use common keys between collections of data to merge information from multiple sources.
- e) Clean and organize data by performing each of the following:
 - Check for outliers, both univariate and multivariate
 - 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.
- f) Detect possible biases introduced when preparing data for a predictive model.

3. Topic: Advanced Predictive Analytics Models (30-45%)

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 the importance of model accuracy.
- b) Explain, fit, evaluate, and make predictions with each of the following models:
 - Additive models
 - Linear mixed models
 - Neural networks.
- c) Apply Bayesian techniques to predictive models.
- d) Compare model results with those from linear and tree-based methods.
- e) Explain the benefits of and demonstrate the combination of multiple models via stacking and blending.
- f) 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.
- g) Recognize and mitigate the effect of:
 - Starting with too many variables
 - Repeated use of train/test/validate sets
 - Model bias, including fairness concepts and proxy discrimination.

4. Topic: Model Explainability and Communication (25-35%)

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:
 - The connection between ethics and explainability
 - 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) Perform data and model governance and develop model documentation in an ethical context.
- e) Communicate in a clear and straightforward manner using common language that is appropriate for the intended audience.
- f) Structure a report in an effective manner while following standards of practice for actuarial communication.

REQUIRED RESOURCES

All required resources are embedded in the e-Learning modules.

The modules contain both R and Python code for completing the analyses. Module 2 (the first module with code) will indicate the versions under which the code has been tested.

For your assessment you are free to use any programming language or statistical software. It is your responsibility to have access to a sufficiently robust platform. The only required deliverable will be a Word document with your responses. The assessment team will ensure

that the tasks can be completed using R or Python as described in the modules. SOA staff will not be available to provide coding help, regardless of the language or software used.