Note: The exam is currently in beta review and will soon be available for testing by customers (to participate, email DRU@datarobot.com).

Citizen Data Scientist Professional

Exam Description

Welcome, we are glad that you are interested in participating in our certification program. This guide is designed to help you evaluate your readiness to pass the Citizen Data Scientist Professional exam. This includes information about the target audience, the exam objectives, and recommended exam preparation. We strongly recommend you combine exam preparation study with hands-on on-the-job experience to maximize your likelihood of passing the exam.

This guide should be used in conjunction with the ***DataRobot Professional Certification Program Guide*** which outlines the DataRobot certification and credential program administration, guidelines, and rules.

About the Citizen Data Scientist Professional

This credential is for individuals experienced in solving business problems using automated AI with DataRobot.

Audience

This exam is aimed at Citizen Data Scientists who have at least 1 year of experience performing data analysis and 6 months using DataRobot. The candidate for this exam is proficient in:

- Defining suitable machine learning projects and sourcing data that will support it
- Preparing foundational data
- Performing Exploratory Data Analysis (EDA)
- Using DataRobot to set-up and build models
- Evaluating and interpreting models
- Recognizing and eliminating bias
- Understanding model deployment approaches
- Communicating results to stakeholders
- Handling common use cases and edge cases

A candidate for this exam will likely need assistance with:

- Modeling strategies and approach
- Putting a model into production
- Models the involve time series

A candidate for this exam is **not** expected to know how to:
- Code in Python or R

Typical job roles of a Citizen Data Scientist Professional candidate may include:

- Business Analyst
- Data Analyst
- Data Engineer
- Associate Data Scientist

About the Exam

Format: 40 multiple choice questions; 75 minutes to complete via proctored exam
Prerequisite exams: None
Registration Fee: 100 USD (per attempt)
Languages: English.
For more information, see the DataRobot Professional Certification Program Guide.

Exam Objectives

Project Identification and Problem Framing (10%)	• Identify machine learning uses cases to solve and frame a business problem • Given a problem scenario, determine an appropriate prediction target • Define the input (features) and predicted output format
Effectively Work with Data (15%)	• Identify suitable dataset sources; evaluate relevance and quality • Enumerate common data preparation issues • Define the unit of analysis
EDA (10%)	• Handle common data quality issues • Select features for building models • Perform automated feature engineering
Modeling Setups (25%)	• Describe how DataRobot uses blueprints and how those differ from models • Address various use cases for partitioning data • Detect and mitigate target leakage • Employ model efficacy techniques such as overfitting, underfitting, partitioning, and cross-validation • Describe causes of drift and the impact to model efficacy • Identify and mitigate model bias • Select an appropriate prediction target
Evaluate, Interpret, Understand, and Communicate (25%)	• Use the appropriate tool to evaluate a model • Interpret regression visuals • Use prediction explanations to understand the logic behind DataRobot predictions • Leverage non-leaderboard visualizations • Evaluate a model's value and options for improvement • Describe key areas of documentation that are critical to communicate
Deploy Models (10%)	• Describe the tasks to implement and maintain models including selecting models for deployment • Differentiate between the various model deployment options • Deploy a model

Exam Knowledge, Experience, and Preparation

The best preparation combines hands-on practical experience, course completion, and self-study in the areas listed in the Exam Objectives. The following courses will help you prepare for the exam:

- AutoML I (instructor led) or equivalent self-paced quests:
  - Citizen Data Scientist Starter
  - Citizen Data Scientist Foundation
- MLOps Starter (Self-Paced or Instructor-Led)
- Time Series Modeling (Self-Paced)
- Data Prep Starter Quest (Self-Paced)
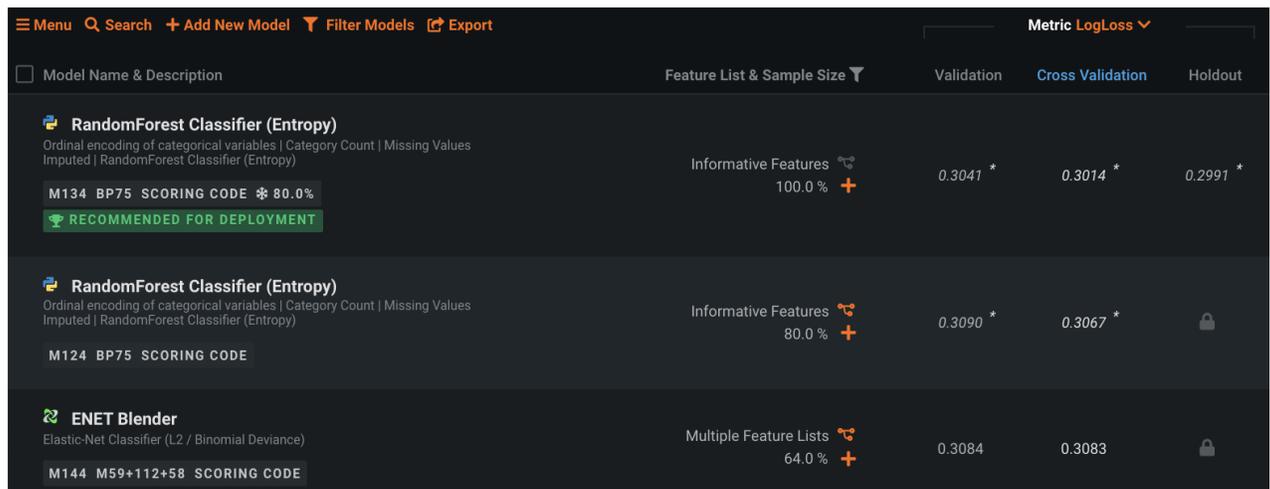- Working with Text (Self-Paced)

Additionally, there are hands-on labs related to many of the exam topics listed. Go to http://university.datarobot.com, in the catalog you can filter on *labs*.

If at any point, you have questions about the topics covered, you should leverage the DataRobot documentation and community as resources.

Sample Questions

The following are examples of questions types that would be typical in the exam.

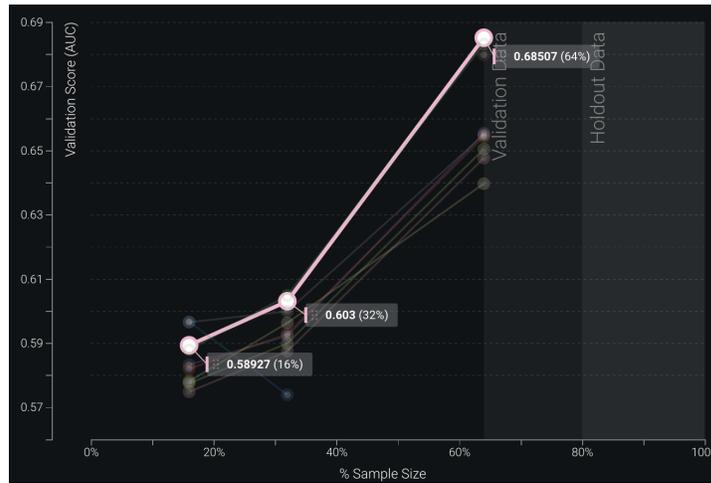1. In this example, which column of scores should be most similar to performance on future data?

    a. Validation because it uses out-of-sample data.
    b. Cross validation because it reduces sample bias.
    c. Holdout because it is the last step in model selection.
    d. It depends on the use case.

2. What does the highlighted learning curve show about the use of the blueprint?

    a.   Performance would not likely improve by adding more rows of data
    b.   Increasing the sample size used for training improved performance
    c.   Adding features to the training set resulted in better scoring models
    d.   Prediction times became longer as more data was used for training

3. A university wants you to predict the demand for tutors for the coming semester. They can supply data about tutors, courses, students, and enrollments (course-student combinations).

What level of detail (unit of analysis) would best predict the demand?

    a.   A row for each tutor to account for a given tutor's prior utilization rate
    b.   A row for each course to account for course difficulty and need for tutor support
    c.   A row for each student to account for the student's history of tutor utilization
    d.   A row for each enrollment to account for both course difficulty and student history

Answers: 1. B   2. B   3. D

Maintenance Requirements

Credentials are valid for 1 year from the date of successful completion. To maintain credentials, you must complete 6 hours of continuing education which you can fulfill through DataRobot education programs. For more information, see the *DataRobot Professional Certification Program Guide*.

Note: DataRobot reserves the right to change certification requirements or program rules at any time. For more information, see the most recent exam guide and DataRobot Professional Certification Program Guide.