AI Data Science Specialist v1.0

Exam 1D0-184

Domain 1: Data Science Overview

1.1: Fundamentals

1.1.1: Define machine learning

1.1.2: Explain data science applications for business

1.1.3: Distinguish the difference between AI and data science

1.1.4: List applications of data science

1.1.5: Describe what is the purpose of data science?

1.1.6: Explain what a correlation coefficient is and how it is calculated

1.2: Legal, Ethics and Privacy Considerations

1.2.1: Explain societal impact of AI

1.2.2: Explain the implications of biased predictions by data models

1.2.3: Apply ethical reasoning in decision making scenarios

1.2.4: Identify ethical guidelines to be applied in data science

1.2.5: Discuss web security standards

1.2.6: Explain data protection security methodologies

1.2.7: Demonstrate risks associated with data privacy and integrity

1.2.8: Demonstrate data collection security principles

1.3: Career

1.3.1: Apply data evaluation and data modeling for business solutions

1.3.2: Describe industries in need of data science

1.3.3: Read scientific articles, conference papers, etc. to identify emerging analytic trends and technologies

1.3.4: Learn about the latest developments in your professional field

Domain 2: Analysis

2.1: Exploratory Data Analysis

2.1.1: Use data mining techniques

2.1.2: Explain clustering techniques and their use cases

2.1.3: Conduct exploratory data analysis

2.1.4: Explain how to capture properties of distributions (mean, variance, skewness, kurtosis)

2.1.5: Analyze sets of data using descriptive statistical methods

2.1.6: Construct frequency distributions

2.2: Modeling and Visualization Techniques

2.2.1: Create a visualization of one or two variables in order to understand the data better

2.2.2: Perform feature selection for supervised and unsupervised analysis

2.2.3: Explain curse of dimensionality

2.2.4: Explain the difference between model underfitting and overfitting

2.2.5: Explain the different types of errors made by a predictive model

2.2.6: Apply dimensionality reduction techniques (e.g., PCA) for data visualization

2.2.7: Explain the difference between classification and regression

2.2.8: Identify different performance metrics for classification (accuracy, ROC curve, AUC, F1)

2.2.9: Analyze data using correlation and linear regression methods

2.2.10: Describe data analyzing techniques

2.3: Statistics

2.3.1: Provide statistical and mathematical solutions

2.3.2: Explain linear models and generalized linear models

2.3.3: Explain bias-variance trade off

2.3.4: Compare and contrast different model evaluation techniques and their pros and cons

2.3.5: Define causal inference and with which kind of data it can be performed

2.3.6: Explain importance of checking model assumptions before deciding on final model

2.3.7: Explain how to detect bias in a model

2.3.8: Explain how to evaluate success of model fitting

2.3.9: Describe statistical power and why it is important

2.3.10: Explain difference between parametric and non-parametric models

2.3.11: Explain how to decide which performance metrics to use given a prediction problem

2.3.12: Explain how to create confidence intervals around estimations

2.3.13: Explain the difference between the frequentist and Bayesian approaches to probability

2.3.14: Explain the concept of hypothesis testing

Domain 3: Managing Data

3.1: General Data Management

3.1.1: Develop data structures and data warehousing solutions

3.1.2: Explain how to analyze big datasets through distributed systems (e.g., Hadoop, MapReduce)

3.1.3: Write SQL queries to fetch the data

3.1.4: List the different stages in the data cycle

3.1.5: Explain how to maintain a dataset through integration and scrubbing

3.1.6: Demonstrate data source attributes, benefits and collection strategies

3.1.7: Explain data selection criteria and procedures

3.1.8: Describe methods for acquiring data

3.2: Querying Databases

3.2.1: Types of databases and query languages

3.2.2: Query languages strengths and weaknesses

3.2.3: Indexes and Query efficiency

3.3: Data Preparation

3.3.1: Handle categorical variables

3.3.2: Explain missing value problem and handling strategies

3.3.3: Explain what outlier is and how an outlier detection process works

3.3.4: Demonstrate data preprocessing and normalization

Domain 4: Professional Skills

4.1: Programming

4.1.1: Explain basic concepts about algorithm design such as computational complexity

4.1.2: Program in R

4.1.3: Use matplotlib and/or seaborn to visualize data

4.1.4: Use Pandas to represent data

4.1.5: Use common machine learning packages

4.1.6: Write syntax for an analysis package (e.g., SPSS, SAS, R)

4.1.7: Program in Python

4.1.8: Solve statistical problems using programming languages

4.2: Conduct Research

4.2.1: Design and conduct surveys, opinion polls, or other instruments to collect data

4.2.2: Perform an A/B test to decide of treatment effect

4.2.3: Describe training and testing datasets and their role in analysis and modeling

4.3: Consulting

4.3.1: Provide technical support for existing reports, software, databases, dashboards, or other tools.

4.3.2: Advise others on analytical techniques

4.4: Communicating Results

4.4.1: Deliver oral or written presentations of the results of modeling and data analysis

4.4.2: Compile reports, charts, papers, presentations or white papers that describe and interpret findings of analyses

4.4.3: Prepare data visualizations to communicate complex results to non-statisticians

4.4.4: Describe how to interpret and report data analysis results

4.5: Deploy Models

4.5.1: Maintain and update existing models using fresh data or to make new predictions.

4.5.2: Choose a methodology for deploying machine learning models for applications.

4.5.3: Develop scalable frameworks

4.5.4: Describe how to scale a data science solution

4.6: Problem Identification

4.6.1: Identify problems that can be solved using machine learning models or data analyses.

4.6.2: Identify business problems or management objectives that can be addressed through data analysis

4.6.3: Identify solutions to problems (staffing, marketing, etc.) using the results of data analysis