CDSP(Associate):Certified Data Science Professional - Associate

Modality: On Demand

Duration: 4 Hours

Overview of CDSP(Associate)

Your role as a data scientist must aim to introduce companies with enhanced strategies of doing business by leveraging the power and limitless opportunities big data offers. The **"CDSP(Associate) Certified Data Science Professional - Assosicate"** certification upskills you to advance and successfully transition from data analytics to data science career. There is a huge skills gap in the field of data science, and talented, knowledgeable data scientists will get countless opportunities, but not every data scientist can fit the roles that require a high level of expertise, reflected with passion to immerse in this science of big data. The course curriculum takes candidates deep into the topics of machine learning, model interpretation, machine learning optimization and compute optimization. Get certified and accelerate your success into leading the big data league.

Who Is This Cert Suitable For?

This certification exam is suitable for incumbents who possess a strong background in data science with a firm understanding of methodologies and their application in relevant disciplines. Having worked as a data analyst for a couple years makes you a suitable candidate to attempt the certification exam, because the topics this exam will test you require pre-established conceptual and practical expertise in some areas of big data, while a deep understanding of other areas. If you have worked as a data analyst, have built an understanding of data analysis and data science challenges and also hold a track record of using various data analysis methodologies to one or more companies' advantage, this certification will further magnify your aptitude and acumen with in-depth data science learning and exposure to technologically updated big data theories and practices.

Pre-requisites for the Cert Exam

If you aim to start preparing for this ultimate Data Science certification, you must come with a solid base of managing databases and spreadsheets. You must be knowledgeable with predictive analysis techniques and instruments while also possess evident exposure around analytical platforms, including SPSS and SAS. You should also have some experience with basic R programming, other quantitative methods, object-oriented programming and RDBMS fundamentals. A data science professional who is friendly with operating on widely used Big Data programming and analytics platforms, would be able to grasp the complicated and technical knowledge delivered while preparing for the certification exam.

Course Outline:

Professional Data Science Skills

- The right metric for business impact

- Understand business need and select correct machine learning error metric

- Explain why you would not use a specific metric for a problem

- Understanding the Business Need

- Analyse a business problem and determine if it is a machine learning problem

- Select which business objective is to be optimized by Machine Learning

- Communicating Data Science to a Non-Technical Audience

- Describe a machine learning problem in simple terms

- List simple explanations of a highly technical problem

- Data Drift and Model Retraining

- Describe what data drift is

- Select when data drift is a problem and what to do about it

- Data Science Change Management in Business

ML Optimization

- Understanding the Error Surface

- Define what an error surface is in terms of modeling

- Explain how to sample from the error surface and optimize models

- Hyperparameter Tuning

- Explain what a hyperparameter is

- Describe hyperparameter tuning in terms of the error surface

- Gradient Descent Optimization

- Explain what Gradient Descent is doing

- Explain Gradient Descent in terms of the error surface

- Genetic Optimization

- Explain what Genetic Optimization is doing

- Explain Genetic Optimization in terms of the error surface

- Bayesian Optimization

- Explain what Bayesian Optimization is doing

- Explain Bayesian Optimization in terms of the error surface

Compute Optimization

- Runtime Optimization

- List factors that affect runtime in Machine Learning

- Select the most optimal combination of factors for runtime

- Vectorization

- Explain what vectorization is

- Explain how vectorization speeds un modeling

- RAM Considerations

- Describe what RAM is

- Describe methods to limit impact on RAM

- Using fast Data Structures

- Describe Fast Data Structures in R and Python

- Analyse when to use these data structures

- GPUs and Parallelization

- Understand why and when GPUs are used

- Understand why and when parrelization is used

Machine Learning Algorithms

- Linear Regression

- Describe linear regression in simple terms

- Explain when to use linear regression

- Logistic Regression

- Describe logistic regression in simple terms

- Explain when to use logistic regression

- K-Nearest Neighbors

- Describe k-nearest neighbors in simple terms

- Explain when to use k-nearest neighbors

- K-Means Clustering

- Describe k-means clustering in simple terms

- Explain when to use k-means clustering

- Regularization

- Describe regularization in simple terms

- Explain when to use regularization

- Neural Networks and transfer learning

- Describe neural networks and transfer learning in simple terms

- Explain when to use neural networks and transfer learning

- DBSCAN

- Describe DBSCAN in simple terms

- Explain when to use DBSCAN

- Random Forest

- Describe random forest in simple terms

- Explain when to use random forest

- Gradient Boosting

- Describe gradient boosting in simple terms

- Explain when to use gradient boosting

- Common Sources of Error

- Describe common problems with modeling (overfitting, underfitting, leakage)

- Explain how to identify common problems in modeling

Model Interpretability

- Partial Dependency

- Describe partial dependency in simple terms

- Explain when to use partial dependency

- Linear Model Interpretation

- Describe linear model interpretation in simple terms

- Explain when to use linear model interpretation

- Tree based variable importance

- Describe tree based variable importance in simple terms

- Explain when to use tree based variable importance

- SHAP

- Describe SHAP values in simple terms

- Explain when to use SHAP values

- Target Leakage

- Describe target leakage in simple terms

- Explain how to identify target leakage