CERTNEXUS®

Certified Data Science Practitioner™(DSP-110) Exam Blueprint

CDSP

Introduction to CertNexus

CertNexus is a vendor-neutral certification body, providing emerging technology certifications and micro-credentials for business, data, developer, IT, and security professionals. CertNexus' mission is to assist closing the emerging tech global skills gap while providing individuals with a path towards rewarding careers in Cybersecurity, Data Science, Data Ethics, Internet of Things, and Artificial Intelligence (AI)/ Machine Learning (ML).

We rely on our Subject Matters Experts (SMEs) to provide their industry expertise and help us develop these credentials by participating in a Job Task Analysis, Exam Item Development, and determining the Cut Score. We also depend upon practitioners in the field to participate in a survey of the Job Task Analysis and beta testing to ensure that our certifications validate knowledge and skills relevant to the industry.

Acknowledgements

CertNexus was honored to have the following Subject Matter Experts contribute to the development of this exam blueprint.

Indrani Gorti	Canadian Imperial Bank of Commerce (CIBC)	https://www.cibc.com/	in
Sahar Nasiri	Delta Airlines	https://www.delta.com	in
Nishtha J	ADAPT Centre	https://www.adaptcentre.ie	in
Sarah Haq	upday	https://corporate.upday.com/	in
Mahsa Seifikar	Pardazeshgaran Saman (SPR)		in
Semih Kumluk	PwC	https://www.pwc.com	in
David Lengacher	Reynolds American Inc.	https://www.reynoldsamerican.com	in

CertNexus Certified Data Science Practitioner™ (CDSP) Exam DSP-110

Exam Information

The *Certified Data Science Practitioner™ (CDSP)* is an industry-validated certification which helps professionals differentiate themselves from other job candidates by demonstrating their ability to put data science concepts into practice. Data can reveal insights and inform—by guiding decisions and influencing day-to-day operations. This calls for a robust workforce of professionals who can analyze, understand, manipulate, and present data within an effective and repeatable process framework. This certification validates candidates' ability to use data science principles to address business issues, use multiple techniques to prepare and analyze data, evaluate datasets to extract valuable insights, and design a machine learning approach. In addition, it will validate skills to design, finalize, present, implement and monitor a model to address issues regardless of business sector.

Candidate Eligibility

The *Certified Data Science Practitioner™ (CDSP)* exam requires no application fee, supporting documentation, or other eligibility verification measures for you to be eligible to take the exam. An exam voucher will come bundled with your training program or can be purchased separately here. Once purchased, you will receive more information about how to register for and schedule your exam through Pearson Vue. You can also purchase a voucher directly through Pearson Vue. Once you have obtained your voucher information, you can register for an exam time here. By registering, you agree to our Candidate Agreement included here.

Exam Prerequisites

There are no formal prerequisites to register for and schedule an exam. Successful candidates will possess the knowledge, skills, and abilities as identified in the domain objectives in this blueprint. It is also strongly recommended that candidates possess the following knowledge, skills, and abilities:
- A working level knowledge of programming languages such as Python® and R
- Proficiency with a querying language
- Strong communication skills
- Proficiency with statistics and linear algebra
- Demonstrate responsibility based upon ethical implications when sharing data sources
- Familiarity with data visualization

You can obtain this level of skill and knowledge by taking the following courseware, which is available through training providers located around the world, or by attending an equivalent third-party training program:
- Introduction to Programming with Python®
- Python® Programming: Advanced
- Using Data Science Tools in Python®
- Data Wrangling with Python®
- Applied Data Science with Python® and Jupyter®
- Big Data Analysis with Python®
- Certified Ethical Emerging Technologist™ (CEET)

Exam Specifications

Number of Items: 100, of which 75 count toward your score

Passing Score: 70%

Duration: 120 minutes (**Note:** Published exam times include the 10 minutes you are allotted for reading and signing the Candidate Agreement and reviewing exam instructions.)

Exam Options: Online through Pearson OnVUE or in person at Pearson VUE test centers.

Item Formats: Multiple Choice / Single Response

The exam should comprise multiple-choice, single-response items as a default. Other item types may be used if content calls for it. For example, manipulating snippets of code, including SQL; reading data visualizations.

Exam Description

Target Candidate:
The *Certified Data Science Practitioner™ (CDSP)* exam is designed for professionals across different industries seeking to demonstrate the ability to gain insights and build predictive models from data.

Exam Objective Statement:
The exam will certify that the successful candidate has the knowledge, skills, and abilities required to answer questions by collecting, wrangling, and exploring datasets, applying statistical models and artificial-intelligence algorithms, to extract and communicate knowledge and insights.

To ensure exam candidates possess the aforementioned skills, the *Certified Data Science Practitioner™ (CDSP)* exam will test them on the following domains with the following distribution:

Domain	# of Items
1.0 Defining the question to be addressed through the application of data science	6
2.0 Extracting, Transforming, and Loading Data	16
3.0 Performing exploratory data analysis	23
4.0 Building models	18
5.0 Testing models	6
6.0 Communicating findings	6
Total	**75**

The information that follows is meant to help you prepare for your certification exam. This information does not represent an exhaustive list of all the concepts and skills that you may be tested on during your exam. The exam domains, identified previously and included in the objectives listing, represent the large content areas covered in the exam. The objectives within those domains represent the specific tasks associated with the job role(s) being tested. The information beyond the domains and objectives is meant to provide examples of the types of concepts, tools, skills, and abilities that relate to the corresponding domains and objectives. All of this information represents the industry-expert analysis of the job role(s) related to the certification and does not necessarily correlate one-to-one with the content covered in your training program or on your exam. We strongly recommend that you independently study to familiarize yourself with any concept identified here that was not explicitly covered in your training program or products.

Objectives

Domain 1.0   Defining the question to be addressed through the application of data science

Objective 1.1   Identify the project scope

- Identify project specifications, including objectives (metrics/KPIs) and stakeholder requirements
- Identify mandatory deliverables, optional deliverables
- Identify project limitations (time, technical, resource, data, risks)

Objective 1.2   Understand stakeholder challenges

- Understand stakeholder terminology
- Become aware of data privacy, security, and governance policies
- Obtain permission/access to data

Objective 1.3   Classify a question into a known data science problem

- Access references
- Identify data sources and type
- Select modeling type

Domain 2.0   Extracting, Transforming, and Loading Data

Objective 2.1   Gather relevant datasets

- Read data
- Research third-party data availability
- Collect open-source data

Objective 2.2   Clean datasets

- Identify and eliminate irregularities in data
- Parse the data
- Check for corrupted data
- Correct the data format for storing/querying purposes
- Deduplicate data

Objective 2.3   Merge datasets

- Join data from different sources

Objective 2.4	**Apply problem-specific transformations to datasets**

- Apply word embeddings
- Generate latent representations for image data

Objective 2.5	**Load data**

- Load into DB
- Load into DataFrame
- Export to CSV files
- Load into visualization tool
- Make an endpoint

Domain 3.0 Performing exploratory data analysis

Objective 3.1	**Examine data**

- Generate summary statistics
- Examine feature types
- Visualize distributions
- Identify outliers
- Find correlations
- Identify target feature(s)

Objective 3.2	**Preprocess data**

- Identify missing values
- Make decisions about missing values (e.g., imputing method, record removal)
- Normalize, standardize, or scale data

Objective 3.3	**Carry out feature engineering**

- Apply encoding to categorical data
- Assign feature values to bins or groups
- Split features
- Convert dates to useful features
- Apply feature reduction methods

Domain 4.0 Building models

Objective 4.1	**Prepare datasets for modeling**

- Decide proportion of dataset to use for training, testing, and (if applicable) validation
- Split data to train, test, and (if applicable) validation sets

Objective 4.2	**Build training models**

- Define algorithms to try
- Train model
- Tune hyperparameters, if applicable

Objective 4.3	**Evaluate models**

- Define evaluation metric
- Compare model outputs
- Select best performing model
- Store model for operational use

Domain 5.0   Testing models

Objective 5.1   **Test hypotheses**

- Design A/B tests
- Define success criteria for test
- Evaluate test results

Objective 5.2   **Test pipelines**

- Put model into production
- Ensure model works operationally
- Monitor pipeline for performance of model over time

Domain 6.0   Communicating findings

Objective 6.1   **Report findings**

- Implement model in a basic web application for demonstration (POC implementation)
- Derive insights from findings
- Identify features that drive outcomes (e.g., explainability, variable importance plot)
- Show model results
- Generate lift or gain chart

Recertification Requirements

The *Certified Data Science Practitioner™ (CDSP)* certification is valid for 3 years from the date that it is initially granted. In order to maintain a continuously valid certification, candidates can recertify by retaking the most recent version of the exam before their certification expires.

Certified Data Science Practitioner (CDSP) Acronyms

Acronym	Expanded Form
AI	artificial intelligence
ANN	artificial neural network
AUC	Area Under the Curve
CNN	convolutional neural network
CSV	Comma-Separated Values
ETL	exact, transform, and load
GAN	generative adversarial network
GDPR	General Data Protection Regulation
HIPPA	Health Insurance Portability and Accountability Act
KNN	K-nearest neighbors
KPI	key performance indicator

LSTM	Long Short-Term Memory
ML	machine learning
NLP	natural language processing
NLTK	Natural Language Toolkit
NLU	natural language understanding
PCA	principal component analysis
RBF	Radial Basis Function Kernel
REST API	Representational State Transfer Application Programming Interface
ROC	Receiver Operating Characteristic
RNN	Recurrent Neural Network
SVM	Support Vector Machine
SQL	Structured Query Language
TF-IDF	Term Frequency–Inverse Document Frequency

CertNexus offers personnel certifications and micro credentials in a variety of emerging technology skills including Cybersecurity, Cyber Secure Coding, the Internet of Things (IoT), IoT Security, Data Science, Artificial Intelligence, and Data Ethics. For a complete list of our credentials visit **https://certnexus.com/certification/**.