CDS | CHARTERED DATA SCIENTIST

www.adasci.org
info@adasci.org

Chartered Data Scientists Curriculum 2023

Section 1: Probability Theory, Statistics and Linear Algebra (weightage 12%)

Counting, Random variables, distributions, quantiles, mean-variance, p-Value, Confidence Interval, Hypothesis testing, t-test, z-test, Chi-Square test, Analysis of Variance (ANOVA), Conditional probability, base rate fallacy, Joint distributions, covariance, correlation, independence, Central limit theorem, Frequentist significance tests and confidence intervals, Maximum Likelihood Estimation, Bayes' theorem and Bayesian statistics, Scalars, Vectors, Matrices, and Tensors. Multiplying Matrices and Vectors, Eigen decomposition, Singular Value Decomposition.

Section 2: Data Engineering and Databases (weightage 8%)

Relational databases, Non-relational databases, Graphical databases, batch processing, stream processing, in-memory processing, data management, SQL Databases, NoSQL Databases, data warehouses, data lakes, ETL and Data pipelines, data access, governance and integration, operations and security.

Section 3: Exploratory Data Analysis (weightage 8%)

Data understanding, Data visualization, Visualization techniques, Univariate analysis, Multivariate analysis, EDA and visualization tools.

Section 4: Supervised and Unsupervised Learning (weightage 15%)

Linear and Non-linear Models, Classification, Regression, K-Nearest Neighbours, Naïve Bayes, Clustering, K-Means Clustering, Hierarchical Clustering, Various learning errors, regularization, estimator bias-variance trade-off, active learning, Support vector machine (SVM) and kernels, Model selection and model selection criteria, Ensemble learning - bagging and boosting, Expectation-Maximization (EM) algorithm, Hidden

Markov models, Bayesian networks, Probabilistic inference, Association Rule Learning, Reinforcement Learning, Time-Series Analysis, Cross-Validation.

Section 5: Neural Networks and Deep Learning (weightage 11%)

Feedforward Networks, Backpropagation Learning, Gradient Descent, Optimization techniques for neural networks, Regularization techniques, Neural net training algorithms, Convolutional Networks, Recurrent and Recursive Neural Networks, Representation Learning, Autoencoders, Deep Generative Models, Factor Analysis, t-Distributed Stochastic Neighbour Embedding (t-SNE), Transfer learning frameworks.

Section 6: Natural Language Processing (weightage 8%)

Text extraction and preprocessing, Text Classification, Sentiment analysis, Information retrieval, Parsing, Tokenization, Vectorization, Part of Speech (POS) Tagging, Sequence modelling, Word embedding, Word2Vec, Language Modelling, NLP Transformers, Generative models for NLP.

Section 7: Computer Vision (weightage 8%)

Image processing, Image transformations, Image segmentation, Image classification, Object recognition, Image reconstruction, Image augmentation, Image tagging, Video analytics, Transformers in computer vision, Generative models for computer vision.

Section 8: Deployment and Model management (weightage 8%)

Deployment of machine learning model, CI/CD and deployment pipelines, tracking model quality, reporting and visualization mechanisms for model performance, MLOps, Data drift and concept drift.

Section 9: Programming Frameworks for Data Science (weightage 10%)

Python Lists, Dictionaries, Arrays, Conditional statements, Loops, Data Frame, Functions, Object-oriented programming, Files, Exceptions, Sci-Kit Learn, Keras, TensorFlow, PyTorch, Python tools for deep learning.

Section 10: Business and Data Science (weightage 12%)

Identifying stakeholders, Handling data privacy concerns, Determining problem-data science fit, defining problem statements for multiple stakeholders, understanding constraints and scope of data science projects, Defining and communicating business benefits, identifying data sources and creating initial reports, Decision Modelling.

Reference Textbooks

The CDS exam has the most likelihood to come from the below mentioned books, though candidates are advised to refer to the textbooks that they are most comfortable with and get the maximum learning from.

Section 1:

A Course in Probability Theory, Kai Lai Chung, Academic Press.
An Introduction to Statistical Learning: With Applications in R, Daniela Witten, Gareth James, Robert Tibshirani, and Trevor Hastie, Springer Publication.
Introduction to Probability Models, 9th Edition, Sheldon M. Ross, Academic Press.

Section 2:

Database System Concepts Textbook by Avi Silberschatz, Henry F. Korth, and S. Sudarshan, McGraw Hill Publication.
Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems, Martin Kleppmann, O'Reilly Publication.

Section 3:

Practical Statistics for Data Scientists: 50 Essential Concepts, Peter Bruce and Andrew Bruce, O'Reilly Publication.

Section 4:

Pattern Recognition and Machine Learning, Christopher Bishop, Springer Publication.
Machine Learning, Tom M. Mitchell, McGraw Hill Publication.
The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2nd Edition, T Hastie, R Tibshirani and J Friedman, Springer Series in Statistics, Springer Publications.

Section 5:

Deep Learning Book by Aaron C. Courville, Ian Goodfellow, and Yoshua Bengio, MIT Press.
Machine Learning A Probabilistic Perspective, Kevin P. Murphy, MIT Press.
Neural Networks and Learning Machines, 3rd Edition, Simon Haykin, Pearson Publication.

Section 6:

Foundations of Statistical Natural Language Processing, Christopher D. Manning and Hinrich Schutze, The MIT Press.
Natural Language Processing with Python, Steven Bird, Ewan Klein and Edward Loper, O'Reilly Publication.
Generative Deep Learning, 2nd Edition, David Foster, O'Reilly Media

Section 7:

Computer Vision: Algorithms and Applications, Richard Szeliski, Springer Publication.
Generative Deep Learning, 2nd Edition, David Foster, O'Reilly Media

Section 8:

Evaluating Machine Learning Models, Alice Zheng, O'Reilly Publication.
Building Machine Learning Powered Applications, Emmanuel Ameisen, O'Reilly Publication.

Section 9:

Python Cookbook: Recipes for Mastering Python 3, 3rd Edition, David Beazley & Brian K. Jones, O'Reilly Publication.
Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow : Concepts, Tools and Techniques to Build Intelligent Systems, 2nd Edition, Aurelien Geron, O'Reilly Publication.
R for Data Science: Import, TIDY, Transform, Visualize, and Model Data, Hadley Wickham and Garrett Grolemund, O'Reilly Publication.

Section 10:

1. Laursen GHN, Thorlund J (2016) Business Analytics for Managers: Taking Business Intelligence Beyond Reporting, 2nd ed.(John Wiley & Sons, Hoboken, NJ).
2. Business Analytics, 2nd Edition, James Evans, Pearson Publication.