Data Science Curriculum

A comprehensive guide to learn data science from scratch

# Introduction

Data science is the interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data. Data science combines various skills such as mathematics, statistics, computer science, domain knowledge, and communication to solve real-world problems.

This curriculum is designed to help you learn data science from the basics to the advanced level. It covers the most important topics, concepts, tools, and techniques that you need to know as a data scientist. You will also learn how to apply your data science skills to different domains and industries, such as business, healthcare, social media, and more.

The curriculum is divided into four module levels: beginner, intermediate, advanced, and expert. Each module level consists of several topics that are further divided into subtopics. The table below shows the overview of the curriculum, with the topics and subtopics for each module level.

|  |  |  |
| --- | --- | --- |
| Module Level | Topics | Subtopics |
| Beginner | Data Science Fundamentals | * What is data science and why it matters * Data types and sources * Data collection and storage * Data cleaning and preprocessing * Data exploration and visualization * Data ethics and privacy |
| Beginner | Programming for Data Science | * Python basics * Python libraries for data science: NumPy, pandas, matplotlib, seaborn * Data structures and algorithms * Object-oriented programming * Debugging and testing * Web scraping and APIs |
| Intermediate | Statistics for Data Science | * Descriptive statistics * Inferential statistics * Hypothesis testing and confidence intervals * Correlation and regression * ANOVA and chi-square tests * Bayesian statistics |
| Intermediate | Machine Learning | * What is machine learning and how it works * Supervised learning: classification and regression * Unsupervised learning: clustering and dimensionality reduction * Reinforcement learning: Q-learning and deep Q-networks * Machine learning algorithms: linear regression, logistic regression, k-nearest neighbors, decision trees, random forests, support vector machines, k-means, principal component analysis, etc. * Machine learning evaluation: accuracy, precision, recall, F1-score, confusion matrix, ROC curve, AUC, etc. * Machine learning optimization: gradient descent, regularization, hyperparameter tuning, etc. * Machine learning frameworks: scikit-learn, TensorFlow, Keras, PyTorch, etc. |
| Advanced | Deep Learning | * What is deep learning and how it differs from machine learning * Artificial neural networks: perceptron, multilayer perceptron, activation functions, backpropagation, etc. * Convolutional neural networks: convolution, pooling, dropout, batch normalization, etc. * Recurrent neural networks: long short-term memory, gated recurrent unit, etc. * Generative adversarial networks: generator, discriminator, etc. * Deep learning applications: computer vision, natural language processing, speech recognition, etc. * Deep learning frameworks: TensorFlow, Keras, PyTorch, etc. |
| Advanced | Natural Language Processing | * What is natural language processing and why it is important * Text preprocessing: tokenization, stemming, lemmatization, stop words, etc. * Text representation: bag of words, term frequency-inverse document frequency, word embeddings, etc. * Text analysis: sentiment analysis, topic modeling, text summarization, etc. * Text generation: language modeling, text completion, text paraphrasing, etc. * Text classification: spam detection, sentiment analysis, document categorization, etc. * Natural language processing frameworks: NLTK, spaCy, gensim, etc. |
| Expert | Data Engineering | * What is data engineering and how it supports data science * Data pipelines: ETL, ELT, batch processing, stream processing, etc. * Data warehousing: data models, schemas, dimensions, facts, etc. * Data lakes: concepts, architecture, benefits, challenges, etc. * Big data: characteristics, sources, challenges, solutions, etc. * Big data frameworks: Hadoop, Spark, Kafka, etc. * Cloud computing: concepts, services, providers, etc. * Cloud platforms: AWS, Google Cloud, Azure, etc. |
| Expert | Data Science Projects | * Data science project lifecycle: CRISP-DM, OSEMN, etc. * Data science project management: agile, scrum, kanban, etc. * Data science project examples: recommendation systems, fraud detection, image recognition, etc. * Data science project tools: Jupyter Notebook, GitHub, Colab, etc. * Data science project documentation: README, report, presentation, etc. * Data science project deployment: web app, API, dashboard, etc. |

# Conclusion

This data science curriculum is a comprehensive and structured guide to help you learn data science from scratch. It covers the most essential topics, concepts, tools, and techniques that you need to know as a data scientist. It also provides you with practical examples and projects to apply your data science skills to different domains and industries. By following this curriculum, you will be able to master data science and become a successful data scientist.