Gigaversity

Data Science Curriculum- Full Breakdown

Module 1: Python Basics and Advanced Concepts

Week 1: Python Basics

Session 1: Introduction to Python

- Overview of Python's history, importance, and features
- Comparison with other programming languages
- Installation and setup
- Basic syntax and writing Python scripts

Session 2: Variables and Data Types

- Keywords and variable rules
- Operators and data types
- Type conversion (casting)
- Introduction to control flow statements

Session 3: Control Flow and Loops

- if, else, elif
- Loops: for, while, break, continue

Session 4: Functions and Lambdas

- Defining functions, parameters, return values
- Lambda expressions
- Functional programming basics (map, reduce, filter)

Session 5: Lists and Dictionaries

- List operations and methods
- Dictionary key-value pairs and manipulation

Session 6: Sets and Tuples

- Set operations and applications
- Tuple characteristics and methods

Week 2: Modules, Exceptions, and Git

Session 1: Modules and Exception Handling

- Modules: importing and using
- Packages: creating and managing
- Exception handling: try, except, finally, raise
- Version Control: Git basics

Week 3: File Handling and Iterators

Session 1: File Handling

- Reading/writing files
- Using with statement for context management

Session 2: Iterators and Generators

- Custom iterator protocols
- Generators and yield

Week 4: OOP in Python

Session 1: Classes and Objects

- Creating classes, attributes, and methods

Session 2: Inheritance and Polymorphism

- Inheritance structures
- Method overriding

Session 3: Encapsulation and Magic Methods

- Abstraction and encapsulation
- Magic methods (__init__, __str__, etc.)

Week 5: NumPy and Pandas

Session 1: Introduction to NumPy

- Arrays, operations
- Indexing, slicing, reshaping

Session 2: Mastering Pandas

- Series and DataFrames
- Missing values, merging, aggregation
- ufuncs and broadcasting

Week 6: Data Visualization and API Frameworks

Session 1: Matplotlib + Advanced NumPy/Pandas

- Line, scatter, histogram plots
- Titles, legends, subplots
- Categorical data plots
- Advanced indexing, grouping, time series

Session 2: Getting Started with FastAPI

Session 3: MLflow + Flask

Session 4: Streamlit

Gigaversity

Module 2: SQL Fundamentals and Advanced Concepts

Week 7: SQL Basics

Session 1: Introduction to SQL

- Importance in data science
- SQL command categories (DQL, DDL, DML, DCL)
- Version control for scripts

Session 2: Basic SELECT Queries

- SELECT syntax and WHERE clause

Session 3: Data Types and Functions

- String, numeric, date/time functions
- NULL handling

Session 4: Aggregation and Grouping

- COUNT, SUM, AVG, MIN, MAX
- GROUP BY and HAVING

Week 8: Joins, Subqueries, and CTEs

Session 1: SQL Joins

- INNER, LEFT, RIGHT, Self Joins

Session 2: Subqueries and CTEs

- Subqueries and correlated subqueries
- Common Table Expressions

Week 9: Advanced SQL Techniques

Session 1: Advanced Querying

- UNION, INTERSECT, CASE, COALESCE

Session 2: Window Functions

- RANK, ROW_NUMBER
- Partitioning and sliding windows

Week 10: Data Manipulation

Session 1: CRUD Operations

- INSERT, UPDATE, DELETE, MERGE

Week 11: Time Series and Transformation

Session 1: Time Series in SQL

- Date functions, time-based aggregations

Session 2: Data Transformation

- PIVOT, UNPIVOT, string splitting/concatenation

Week 12: Views and Optimization

Session 1: Views and Temp Tables

- Creating views and materialized views

Session 2: Performance Optimization

- Indexes, query plans

Session 3: Python + SQL Integration

- SQL with pandas, SQLAlchemy
- Execute SQL queries from Python

Session 4: Final Review and Capstone

- Mini-project
- Best practices
- EDA project using SQL + Python

Gigaversity

Module 3: Mathematics for Data Science

Session 1: Introduction to Mathematics in Data Science

- Overview of mathematical concepts used in data science
- Importance of mathematics in data analysis and modeling
- Basic set theory and logic

Session 2: Descriptive Statistics

- Measures of central tendency: Mean, Median, Mode
- Measures of dispersion: Range, Variance, Standard Deviation
- Understanding data distributions

Session 3: Probability Basics

- Fundamentals of probability
- Probability distributions: Normal, Binomial, Poisson
- Bayes' Theorem and Conditional Probability

Session 4: Calculus Basics

- Differentiation rules and techniques
- Partial derivatives for multivariate functions

Session 5: Optimization Techniques

- Optimization problems in data science
- Gradient Descent and its variants (Stochastic, Mini-batch)
- Constrained optimization and Lagrange multipliers

Session 6: Linear Algebra Foundations

- Vector and Matrix operations
- Singular Value Decomposition (SVD)
- Eigenvalues and Eigenvectors

Session 7: Advanced Linear Algebra

- Linear transformations and their geometric interpretations
- Principal Component Analysis (PCA) introduction

Session 8: Review and Integration

- Recap of concepts from Sessions 1–7
- Case Study: Apply all math concepts to solve a data science problem

---

Module 4: Exploratory Data Analysis (EDA)

Session 1: Introduction to EDA and Data Collection

- Definition and importance of EDA
- Data sources and importing data
- Initial inspection of data

Session 2: Data Cleaning Basics

- Techniques for dealing with missing data
- Identifying and removing duplicates
- Handling datetime data

Session 3: Data Transformation

- Data type conversions
- Normalization and standardization
- Handling skewed data

Session 4: Summary Statistics and Data Distributions

- Measures of central tendency and dispersion
- Creating and interpreting histograms, box plots, and scatter plots
- Understanding data distributions

Session 5: Data Visualization Fundamentals

- Introduction to visualization libraries (Matplotlib, Seaborn)
- Creating basic plots: bar charts, line graphs, scatter plots
- Customizing plots: labels, titles, legends

Session 6: Advanced Visualization and Feature Engineering

- Creating advanced plots: heatmaps, pair plots, violin plots
- Introduction to feature engineering

Session 7: Outlier Detection and Handling

- Techniques for identifying outliers (Z-score, IQR method)
- Visualizing outliers
- Strategies for handling outliers

Session 8: Hypothesis Testing and Presenting Findings

- Conducting simple statistical tests (t-tests, chi-square tests)
-

- Interpreting p-values and significance levels
- Best practices in data visualization

Module 5: Machine Learning Curriculum

Module 1: Introduction to Machine Learning

Session 1: Introduction to Machine Learning

- Definition and Importance of Machine Learning
- Types of Machine Learning: Supervised, Unsupervised, Reinforcement
- Applications and real-world examples

Session 2: Supervised Learning Basics

- Introduction to Classification and Regression Problems
- Differences between Classification and Regression
- Evaluation Metrics Overview

Module 2: Supervised Learning Algorithms

Session 3: Decision Trees and Random Forests

- Understanding Decision Trees
- Building and Visualizing Decision Trees
- Implementing Random Forests
- Applications and real-world examples

Session 4: K-Nearest Neighbors (KNN) and Logistic Regression

- KNN Algorithm Explanation
- Choosing the Right K Value
- Logistic Regression Theory

- Implementing Logistic Regression

Module 3: Advanced Supervised Learning

Session 5: Support Vector Machines (SVM)

- SVM Theory and Kernel Trick
- Implementing SVM for Classification
- Support Vector Regression (SVR)

Session 6: Linear and Polynomial Regression

- Simple and Multiple Linear Regression
- Polynomial Regression
- Implementing Linear Regression
- Ridge Regression (L2 Regularization)

Module 4: Regularization and Boosting

Session 7: Regularization Techniques

- Lasso Regression (L1 Regularization)
- Elastic Net Regression
- Overfitting and Underfitting

Session 8: Gradient Boosting Machines

- Introduction to Boosting
- Implementing and Tuning Gradient Boosting Models
- XGBoost and LightGBM

Module 5: Neural Networks and Deep Learning

Session 9: Neural Networks

- Introduction to Neural Networks
- Feedforward Neural Networks
- Backpropagation Algorithm

Session 10: Unsupervised Learning

- Introduction to Clustering
- K-Means Clustering Algorithm
- Hierarchical Clustering
- DBSCAN (Density-Based Spatial Clustering)

Module 6: Advanced ML Topics

Session 11: Dimensionality Reduction

- Principal Component Analysis (PCA)
- t-Distributed Stochastic Neighbor Embedding (t-SNE)
- Feature Selection vs. Dimensionality Reduction

Session 12: Association Rule Learning

- Apriori Algorithm
- Market Basket Analysis
- Eclat Algorithm

Session 13: Advanced Unsupervised Learning

- Gaussian Mixture Models (GMM)
- Self-Supervised Learning Basics

- Autoencoders
- Introduction to Anomaly Detection Techniques

Module 7: Model Evaluation and Reinforcement Learning

Session 14: Model Evaluation and Ensemble Methods

- Cross-Validation Techniques
- Bagging and Boosting
- Stacking Models

Session 15: Introduction to Reinforcement Learning

- Basics of Reinforcement Learning
- Markov Decision Processes (MDPs)
- Q-Learning Algorithm

Module 8: Advanced Topics and Future Trends

Session 16: Advanced Topics and Future Trends

- Transfer Learning Basics
- Ethical Considerations in AI
- Future Trends in Machine Learning

Module 6: Deep Learning + NLP Curriculum

Module 1: Deep Learning Foundations

Session 1: Introduction to Deep Learning and Neural Networks

- Basics of Artificial Neural Networks (ANNs)
- Structure and Components of Neural Networks
- Importance and Applications of Deep Learning

Session 2: Training Neural Networks

- Backpropagation
- Gradient Descent
- Activation Functions (ReLU, Sigmoid, Tanh)

Session 3: Loss Functions and Optimization Algorithms

- Loss Functions (MSE, Cross-Entropy)
- Optimization Algorithms (SGD, Adam, RMSProp)
- Introduction to Learning Rate Scheduling

Module 2: Convolutional Neural Networks (CNNs)

Session 4: Introduction to Convolutional Neural Networks (CNNs)

- Overview of CNNs
- Convolutional Layers
- Pooling Layers (Max, Average)
- Flattening Layers

Session 5: Applications of CNNs in Computer Vision

- Image Classification
- Transfer Learning in CNNs
- Data Augmentation

Module 3: Recurrent Neural Networks (RNNs)

Session 6: Introduction to RNNs

- RNN Fundamentals
- Applications in Time-Series and Sequential Data
- Challenges with Vanilla RNNs (Vanishing Gradient Problem)

Session 7: Long Short-Term Memory (LSTM) Networks

- LSTM Concepts
- Applications of LSTM
- Sequence-to-Sequence Models (basic intro to Attention)

Module 4: Natural Language Processing (NLP)

Session 8: Introduction to NLP

- Basics of NLP Tasks (Text Classification, Translation, Summarization)
- Introduction and Applications

Session 9: Feature Extraction for NLP

- Text Preprocessing: Cleaning, Tokenization
- Word Embeddings (Word2Vec, GloVe)
- TF-IDF Vectorization

Session 10: Introduction to Transformers

- Transformer Architecture Basics (Encoder-Decoder)
- Attention Mechanisms (Self-Attention, Multi-Head Attention)

Session 11: BERT and Its Variants

- Overview of BERT
- Applications of BERT in NLP (Text Classification, QA)
- Fine-tuning Pre-trained BERT Models

Session 12: Advanced NLP Techniques

- Sequence Labeling Tasks (NER, POS Tagging)
- Text Classification with CNNs and RNNs

- Introduction to Anomaly Detection in Text Data

Module 5: Self-Supervised Learning and Advanced Topics

Session 13: Self-Supervised Learning in NLP

- Introduction to Self-Supervised Learning
- Techniques like Masked Language Modeling (MLM)
- Practical Implementations of Self-Supervised Models

Module 7: Generative AI (GenAI) Curriculum

Module 1: Introduction to Generative Models

Session 1: What are Generative Models?

- Definition and purpose of generative models
- Types: GANs, VAEs, LLMs
- Key applications: Text generation, image generation, data synthesis

---

Session 2: Overview of GANs

- Components of GANs: Generator and Discriminator
- How GANs work: Adversarial process
- Types of GANs: StyleGAN, CycleGAN

Session 3: Applications of GANs

- Applications: Image generation, super-resolution, text generation
- GANs for data augmentation and art creation
- GANs for text generation

Module 2: Autoencoders and VAEs

Session 4: Introduction to Autoencoders and VAEs

- Basic structure: Encoder, latent space, decoder
- Applications: Data compression, anomaly detection, unsupervised learning
- Variational Autoencoders for generating images and text

Session 5: Advanced Applications of Autoencoders and VAEs

- Use of VAEs in generating images/text
- Feature extraction with VAEs
- Comparison of VAEs and GANs

Module 3: Introduction to LLMs (Large Language Models)

Session 6: Introduction to LLMs

- What are LLMs?
- History: GPT, BERT, T5 evolution
- Pre-training and fine-tuning in LLMs

Session 7: Transformer Architecture in LLMs

- Introduction to Transformer model
- Encoder-decoder structure
- Self-attention, Multi-head Attention, Positional Encoding

Session 8: Pre-trained Language Models

- Overview of GPT Models (GPT-2, GPT-3, GPT-4)
- Introduction to BERT: Bidirectional Understanding
- Differences between GPT and BERT

Module 4: Applications of LLMs in Generative AI

Session 9: Applications of LLMs

- Use cases: Chatbots, text generation, summarization, translation
- Content creation using LLMs: Articles, poetry, storytelling
- Text-to-Text Transformation tasks

Session 10: Feature Learning and Evaluation Techniques

- Feature Learning in Generative AI
- How LLMs learn latent semantic features
- Evaluation Metrics for Generative Models:
  - Inception Score (IS), Fréchet Inception Distance (FID) for images
  - Perplexity, BLEU, ROUGE for text