

Data Science Course

A Data Science syllabus typically includes a combination of foundational concepts, tools, and techniques in statistics, programming, data manipulation, machine learning, and data visualization. Here is a comprehensive overview of the typical content covered in a Data Science course.

1. Introduction to Data Science

- What is Data Science?
- The Data Science Process: From data collection to analysis and visualization
- Overview of Data Science tools and technologies
- Real-world applications of Data Science in various industries

2. Data Collection and Acquisition

- Understanding Data Types: Structured, Semi-structured, and Unstructured data
- Data Sources: Databases, APIs, Web Scraping, and IoT
- Introduction to Big Data (Hadoop, Spark, etc.)
- Data Storage: Databases (SQL, NoSQL)

3. Data Preprocessing and Cleaning

- Importance of Data Cleaning
- Handling Missing Data: Imputation, Deletion, etc.
- Data Transformation: Scaling, Normalization, and Encoding
- Handling Categorical Data: One-Hot Encoding, Label Encoding
- Removing Duplicates, Outliers, and Inconsistent Data

4. Exploratory Data Analysis (EDA)

- Descriptive Statistics: Mean, Median, Mode, Standard Deviation, etc.
- Visualizing Data: Histograms, Box Plots, Scatter Plots
- Correlation and Covariance
- Identifying Patterns and Trends in Data
- Using libraries like Pandas, Matplotlib, Seaborn for EDA

5. Introduction to Programming (Python or R)

- Basic Programming Concepts: Variables, Data Types, Loops, Functions
- Libraries for Data Science: NumPy, Pandas, Matplotlib, Seaborn
- Working with Data Frames and Series
- Introduction to Jupiter Notebooks and IDEs for Data Science

6. Probability and Statistics for Data Science

- Probability Theory: Events, Conditional Probability, Bayes Theorem
- Probability Distributions: Normal, Binomial, Poisson, etc.
- Statistical Inference: Hypothesis Testing, p-values, Confidence Intervals
- Sampling Techniques and Central Limit Theorem
- Regression Analysis: Linear and Logistic Regression

7. Data Visualization

- Principles of Data Visualization
- Creating Charts and Graphs: Bar Charts, Line Graphs, Pie Charts, Histograms
- Advanced Visualization: Heatmaps, Pair Plots, Word Clouds
- Using Visualization Tools: Matplotlib, Seaborn, Plotly, Tableau, Power BI
- Interactive Visualizations for Data Exploration

8. Machine Learning Basics

- Introduction to Machine Learning: Supervised vs. Unsupervised Learning
- Types of Machine Learning Algorithms
- Supervised Learning: Linear Regression, Logistic Regression, Decision Trees, Random Forests, Support Vector Machines (SVM)
- Unsupervised Learning: K-Means Clustering, Hierarchical Clustering, PCA (Principal Component Analysis)
- Model Evaluation: Cross-validation, Confusion Matrix, Precision, Recall, F1-Score
- Model Overfitting and Underfitting, Bias-Variance Tradeoff
- Training and Testing Data Split

9. Advanced Machine Learning Techniques

- Ensemble Learning: Random Forest, Gradient Boosting, AdaBoost, XGBoost
- Neural Networks and Deep Learning: Basics of Artificial Neural Networks (ANN)
- Introduction to Deep Learning Libraries: TensorFlow, Keras, PyTorch
- Natural Language Processing (NLP): Text Preprocessing, Tokenization, Sentiment Analysis, Word

Address: 9, FCI Housing Society, Above Medplus Medical, Manish Nagar, Nagpur -440015

Call: +91 9657959184 Email: ssitnagpur285@gmail.com Website: - www.ssinfotech.org

Embeddings (Word2Vec, GloVe)

- Time Series Analysis: ARIMA, Exponential Smoothing
- Recommender Systems: Collaborative Filtering, Content-Based Filtering

10. Big Data and Data Engineering

- Big Data Technologies: Hadoop, Spark, Hive, Pig
- Data Processing: Batch vs. Stream Processing
- Data Pipelines: ETL (Extract, Transform, Load) Concepts
- Cloud Computing in Data Science: AWS, Google Cloud, Microsoft Azure
- Real-time Data Processing with Apache Kafka

11. Model Deployment and Productionalization

- Introduction to Model Deployment
- Deploying Models with Flask, Fast API, or Django
- Introduction to Docker and Containers for Deployment
- Continuous Integration/Continuous Deployment (CI/CD)
- Model Monitoring and Maintenance
- Serving Models via REST APIs

12. Ethics and Privacy in Data Science

- Ethical Considerations in Data Collection and Usage
- Bias and Fairness in Machine Learning Models
- Data Privacy and Security: GDPR, HIPAA, and other regulations
- Transparency and Explainability in AI and ML Models
- Responsible AI Practices

13. Capstone Project

- Building a Data Science Project from Scratch
- End-to-End Workflow: Data Collection, Cleaning, Analysis, Modeling, Visualization
- Presenting Results and Insights to Stakeholders
- Documentation and Reporting

14. Tools and Technologies

- Programming Languages: Python, R
- Libraries: Pandas, NumPy, SciPy, Scikit-learn, TensorFlow, Keras, PyTorch, Matplotlib, Seaborn, Plotly
- Big Data Tools: Hadoop, Apache Spark, Hive

- Databases: SQL, NoSQL, MongoDB
- Visualization Tools: Tableau, Power BI, Plotly, D3.js
- Cloud Platforms: AWS, Google Cloud, Microsoft Azure