

Data Analytics Course Syllabus

Module 1: Foundations of Data Analytics

1.1 Introduction to Data Analytics

- Definition, scope, and significance of data analytics
- Historical perspective and evolution of the field
- Key components of the data analytics ecosystem

1.2 Data Collection and Preprocessing

- Techniques for data acquisition (APIs, web scraping, databases)
- Data cleaning and preprocessing using Pandas
- Exploratory Data Analysis (EDA) and feature engineering

1.3 Descriptive Statistics and Visualization

- Measures of central tendency and dispersion
- Data visualization using Matplotlib, Seaborn, and Plotly
- Dashboard creation and design principles

1.4 Inferential Statistics

- Hypothesis testing, confidence intervals, and p-values
- Regression analysis and correlation
- Advanced statistical methods and experimental design

Module 2: Programming for Data Analytics

2.1 Python Programming Fundamentals

- Variables, data types, and control structures
- Functions, modules, and error handling
- Object-oriented programming concepts

2.2 Data Manipulation and Visualization with Python

- Pandas for data manipulation
- Data visualization libraries (Matplotlib, Seaborn, Plotly)
- Interactive visualizations and geospatial data

2.3 Advanced Topics in Python

- Web scraping using BeautifulSoup and Selenium
- Machine learning libraries (Scikit-Learn, TensorFlow, PyTorch)
- API integration and working with JSON data

2.4 Database Management and SQL

- Relational databases and normalization
- SQL basics and advanced querying
- Connecting Python to databases for seamless integration

Module 3: Machine Learning and Predictive Analytics

3.1 Introduction to Machine Learning

- Supervised learning, unsupervised learning, and reinforcement learning
- Model evaluation, validation, and overfitting
- Ensemble methods and model deployment considerations

3.2 Regression and Classification Models

- Linear regression, logistic regression
- Decision trees, random forests, and support vector machines
- Neural networks and deep learning fundamentals

3.3 Clustering and Dimensionality Reduction

- K-means clustering, hierarchical clustering
- Principal Component Analysis (PCA) and t-SNE
- Anomaly detection and outlier analysis

3.4 Time Series Analysis and Forecasting

- Time series data exploration and preprocessing
- ARIMA models and forecasting techniques
- Advanced time series models and applications

Module 4: Big Data and Advanced Technologies

4.1 Introduction to Big Data

- Characteristics of big data and its challenges
- Distributed computing frameworks (Hadoop, Spark)
- Big data storage solutions (HDFS, S3, Google Cloud Storage)

4.2 Cloud Computing for Data Analytics

- Cloud platforms (AWS, Google Cloud, Azure)
- Setting up cloud-based data analytics environments
- Scalability, security, and cost considerations

4.3 Advanced Analytics with Spark

- Spark architecture and RDDs
- Spark SQL and DataFrame API
- Machine learning with Spark MLlib

Module 5: Capstone Project and Industry Applications

5.1 Capstone Project

- Real-world data analytics project in collaboration with industry partners
- Project planning, execution, and presentation
- Peer review and feedback sessions

5.2 Industry Applications and Guest Lectures

- Case studies from various industries (finance, healthcare, marketing, etc.)
- Guest lectures from industry experts and data scientists
- Emerging trends in data analytics and future directions