

Data Science Course Content – 9Globes Technologies

Module 1: Introduction to Python:

- Concepts of Python programming
- Configuration of Development Environment
- Using the Python Interpreter
- Numbers and Strings

Module 2: More on Python:

- Tuples and Lists
- Functions
- Control Flow and Loops
- Dictionaries

Module 3: Data science Fundamentals:

- Introduction to Data science
- Real world use-cases of Data science
- Walkthrough of data types
- Data science project lifecycle

Module 4: Introduction to NumPy:

- Basics of NumPy Arrays
- Mathematical operations in NumPy
- NumPy Array manipulation
- NumPy Array broadcasting

Module 5: Data Manipulation with Pandas:

- Data Structures in Pandas-Series and Data Frames
- Data cleaning in Pandas
- Data manipulation in Pandas
- Handling missing values in datasets
- Hands-on: Implement NumPy arrays and Pandas Data Frames

Module 6: Data Visualization in Python:

- Plotting basic charts in Python
- Data visualization with Matplotlib
- Statistical data visualization with Seaborn
- Hands-on: Coding sessions using Matplotlib, Seaborn packages

Module 7: Exploratory Data Analysis:

- Introduction to Exploratory Data Analysis (EDA) steps
- Plots to explore relationship between two variables
- Histograms, Box plots to explore a single variable
- Heat maps, Pair plots to explore correlations
- Perform EDA to explore survival using titanic dataset

Module 8: Introduction to Machine Learning:

- What is Machine Learning?
- Use Cases of Machine Learning
- Types of Machine Learning - Supervised to Unsupervised methods
- Machine Learning workflow

Module 9: Linear Regression:

- Introduction to Linear Regression
- Use cases of Linear Regression
- How to fit a Linear Regression model?
- Evaluating and interpreting results from Linear Regression models
- Predict Bike sharing demand

Module 10: Logistic Regression:

- Introduction to Logistic Regression
- Logistic Regression use cases
- Understand use of odds & Logit function to perform logistic regression
- Predicting credit card default cases

Module 11: Decision Trees & Random Forest:

- Introduction to Decision Trees & Random Forest
- Understanding criterion (Entropy & Information Gain) used in Decision Trees
- Using Ensemble methods in Decision Trees
- Applications of Random Forest
- Predict passenger survival using Titanic Data set

Module 12: Model Evaluation Techniques:

- Introduction to evaluation metrics and model selection in Machine Learning
- Importance of Confusion matrix for predictions
- Measures of model evaluation - Sensitivity, specificity, precision, recall & f-score
- Use AUC-ROC curve to decide best model
- Applying model evaluation techniques to Titanic dataset

Module 13: Dimensionality Reduction using PCA:

- Unsupervised Learning: Introduction to Curse of Dimensionality
- What is dimensionality reduction?
- Technique used in PCA to reduce dimensions
- Applications of Principle component Analysis (PCA)
- Optimize model performance using PCA on SPECTF heart data

Module 14: K Nearest Neighbours:

- Introduction to KNN
- Calculate neighbours using distance measures
- Find optimal value of K in KNN method
- Advantage & disadvantages of KNN
- Classify phishing site data using close neighbour technique

Module 15: Naive Bayes Classifier:

- Introduction to Naive Bayes Classification
- Refresher on Probability theory
- Applications of Naive Bayes Algorithm in Machine Learning
- Classify spam emails based on probability

Module 16: K-means Clustering:

- Introduction to K-means clustering
- Decide clusters by adjusting centroids
- Find optimal 'k value' in K-means
- Understand applications of clustering in Machine Learning
- Segment hands in Poker data and segment flower species in Iris flower data

Module 17: Support Vector Machines:

- Introduction to SVM
- Figure decision boundaries using support vectors
- Identify hyperplane in SVM
- Applications of SVM in Machine Learning
- Predicting wine quality using SVM