



DATA SCIENCE

Data Science Overview

Learning Objective

- ❖ Understanding Data Science and Data Science Project Life cycle.

Topics Covered

- Introduction To Data Science
- Life Cycle of Data Science
- Skills required for Data Science
- Careers Path in Data Science
- Applications of Data Science
- Future Of Data Science
- Relation between Statistics, Math's, Programming, BI tools and Data Science.

Statistics

Learning Objective

- ❖ Learning Statistics required for data science

Topics Covered:

- Introduction to Data:

- ❑ Data types
- ❑ Data Collection Techniques
- ❑ **Descriptive Statistics:**
 - ❑ Measures of Central Tendency
 - ❑ Measures of Dispersion
 - ❑ Measures of Skewness and Kurtosis
 - ❑ Visualization
- ❑ **Inferential Statistics:**
 - ❑ Sampling variability and Central Limit Theorem
 - ❑ Confidence Interval for Mean
 - ❑ Hypothesis ,t- Test, F-Test, Chi-square Test
 - ❑ ANOVA
- ❑ **Correlation coefficient(R-value),r squared, adjusted r squared.P value**
- ❑ **MAE,MSE,RMSE**
- ❑ **Confusion Matrix,TP,TN,FP,FN,Precision,Recall,Classification Report**
- ❑ **Random Sampling and Probability Distribution:**
 - ❑ Probability and Limitations, Discrete Probability, Continuous Probability
 - ❑ Binomial, Poisson Distributions, Normal Distribution

Python for Data Science

Learning Objective

- ❖ **Learning and Building a foundation for the most in-demand programming language of the 21st century**

Topics Covered

- ❑ Python programming:
- ❑ Environment Setup
- ❑ Jupyter Notebook Overview

- ❑ Data types:Numbers,Strings,Printing,Lists,Dictionaries,Booleans,Tuples ,Sets
- ❑ Comparison Operators
- ❑ if, elif, else Statements
- ❑ Loops: for Loops, while Loops
- ❑ range()
- ❑ list comprehension
- ❑ functions
- ❑ lambda expressions
- ❑ map and filter
- ❑ methods
- ❑ Programming Exercises.
- ❑ Object Oriented Programming
- ❑ Modules and packages
- ❑ Errors and Exception Handling
- ❑ Python Decorators
- ❑ Python generators
- ❑ Collections
- ❑ Regular Expression

- ❑ Python for Exploratory Data Analysis:Numpy
 - ❑ Installing numpy
 - ❑ Using numpy
 - ❑ NumPy arrays
 - ❑ Creating numpy arrays from python list
 - ❑ Creating arrays using built in methods(arrange(),zeros(),ones(),linspace(),eye(),rand(),etc.
 - ❑ Array attributes : shape, type
 - ❑ Array methods: Reshape(),min(),max(),argmax(),argmin(),etc.

- ❑ Python for Exploratory Data Analysis:Pandas
 - ❑ Introduction to Pandas
 - ❑ Series

- Data Frames
- Missing Data
- GroupBy
- Merging, Joining and Concatenating
- Operations
- Data Input and Output

- Python for Data Visualization (matplotlib)
 - Installing Matplotlib, Basic Matplotlib commands
 - Creating Multiplot on same canvas
 - Object Oriented Method:figure(),plot(),add_axes(),subplots(),etc.
 - Matplotlib Exercise

- Python for Data Visualization (seaborn)
 - Categorical plot
 - Distribution plot
 - Regression plot
 - Seaborn Exercise

- Python for Data Visualization (using pandas)
 - Scatter plot
 - Histograms
 - Box plot

- Data Analysis Case Study
- Web Scraping

Note :All topics are delivered as Hands-On sessions.

SQL

Learning Objective

- ❖ Learn the fundamental of database and extract information from RDBMS using the structured query language.

Topics Covered

- Introduction to RDBMS
- Installing mysql
- Retrieving
- Updating
- Inserting
- Deleting
- Sorting AND Filtering
- Summarizing AND Grouping
- Using Subqueries
- Joining Tables
- Views
- Stored Procedure
- Python Database Connection API Subnets

Machine Learning

Learning Objective

- ❖ Learning Machine Learning Fundamentals
- ❖ Using Machine Learning to solve problems
- ❖ Mastering supervised and unsupervised learning

Topics Covered

- Relationship between Data Science and Machine Learning
- Supervised Learning
- Unsupervised Learning
- Supervised Learning
 - Linear Regression
 - Ridge Regression
 - Lasso Regression
 - Polynomial Regression
 - Support vector regression
 - Decision Tree Regression
 - Random Forest Regression
 - Logistic Regression
 - Support Vector Machines
 - Kernel SVM
 - Decision Trees and Random Forest
 - Ensemble Of Decision Trees
 - Model Evaluation and Improvement
- Unsupervised Learning
 - Challenges in Unsupervised Learning
 - Preprocessing AND Scaling
 - Dimensionality Reduction, Feature Extraction
 - Principle Component Analysis (PCA)
 - Clustering
 - KMEANS
 - Hierarchical clustering
 - Model evaluation and improvement
 - Cross validation, Grid search, Evaluation metrics and scoring

Case Study on Machine Learning: Apply the machine learning concept learnt to solve a real life predictive analytics problem.

Note :All topics are delivered as Hands-On sessions.

Artificial Neural Network and Deep Learning

Learning Objective

- ❖ **Learning Artificial Neural Network and Deep Learning**

Topics Covered

- ❑ **Introduction to Deep Learning** : Deep Learning Applications, Artificial Neural Network, TensorFlow Demo, Deep Learning Frameworks
- ❑ **Up and Running with TensorFlow** : Installation, Creating Your First Graph and Running It in a Session, Managing Graphs, Lifecycle of a Node Value, Linear Regression with TensorFlow, Implementing Gradient Descent, Feeding Data to the Training Algorithm, Saving and Restoring Models, Visualizing the Graph and Training Curves Using TensorBoard, Name Scopes, Modularity, Sharing Variables
- ❑ **Introduction to Artificial Neural Networks** : From Biological to Artificial Neurons, Training an MLP with TensorFlow's High-Level API, Training a DNN Using Plain TensorFlow, Fine-Tuning Neural Network Hyperparameters
- ❑ **Training Deep Neural Nets** : Vanishing / Exploding Gradients Problems, Reusing Pretrained Layers, Faster Optimizers, Avoiding Overfitting Through Regularization, Practical Guidelines
- ❑ **Convolutional Neural Networks** : The Architecture of the Visual Cortex, Convolutional Layer, Pooling Layer, CNN Architectures

- ❑ **Recurrent Neural Networks** : Recurrent Neurons, Basic RNNs in TensorFlow, Training RNNs, Deep RNNs, LSTM Cell, GRU Cell, Natural Language Processing
- ❑ **Autoencoders** : Efficient Data Representations, Performing PCA with an Undercomplete Linear Autoencoder, Stacked Autoencoders, Unsupervised Pretraining Using Stacked Autoencoders, Denoising Autoencoders, Sparse Autoencoders, Variational Autoencoders

Note: All topics are delivered as Hands-On sessions.

Case Study on Deep Learning: Apply the Deep Learning concept learnt to solve a real life predictive problem.

NLP and Recommender Systems

Learning Objective

- ❖ Understanding natural language processing.

Topics Covered

- ❑ Corpus
- ❑ Text preprocessing using Bag of words technique
- ❑ TF(Term Frequency)
- ❑ IDF(Inverse Document Frequency)
- ❑ Normalization
- ❑ Vectorization
- ❑ NLP with Python

Tableau

Learning Objective

- ❖ Data Visualization using Tableau

Topics Covered

- Tableau basics
- Architecture of Tableau
- Meta Data and Data Blending
- Working with sets,filter,mapping
- Using Calculation,Expression,Parameters
- Charts, Graphs
- Dashboards, Storytelling.
- Integrating Tableau with MySQL database, Hadoop

Note: All topics are delivered as Hands-On sessions.

Case Study using Tableau: Solving a real life problem using Tableau.