



Data Science Foundation Program

- ✓ 250+ Hiring Partners
- ✓ Hybrid Model for Project Sessions
- ✓ 97% Average Salary Hike

In Collaboration with



TABLE OF CONTENTS

Program Details	1
Why Choose Us	2
IBM Certification	3
Job Referrals	4
Success Stories	5
Transition Process	6
Program Fee & Financing	7
Program Outline	8
Real-Time Projects & Detailed Syllabus	9

PROGRAM DETAILS

COURSE PREREQUISITE

There are no such hard prerequisite criteria. Just the urge to learn programming and basic ideas about advanced math is enough.

WHO IS THIS PROGRAM FOR?

- ✓ Working professional having more than 6 months of experience in Non-Technical Domain
- ✓ Qualification: BE/B.Tech (from any branch), BBA/MBA, MCA/M.Tech, B.Com, B.Sc (in any branch)

INDUSTRIAL EXPERTS

Our trainers are working professionals having more than 8+ years of experience as Sr. Data Scientist, Machine Learning Engineer, AI Engineer, BI Developer, Big Data Architect, Sr. Data Analyst etc.



Course Duration

Weekday Batch : 5 Months
Monday to Friday - 2 Hours/Day

Weekend Batch : 7 Months
Saturday & Sunday - 3.5 Hours/Day

200+ Hours
of Learning



WHY CHOOSE US?

We focus on working professionals and help them achieve the peak of success without losing their designation or wasting their existing experience.

SUPPORT FOR NON-PROGRAMMERS

- We provide special assistance and support for non-technical professionals
- All tools & modules are covered from scratch



PROJECT INNOVATION LAB

- Experts from MNCs and MAANG assist in **online and offline** project sessions
- Attain classroom session in 7+ cities (Pune, Mumbai, Delhi, Kolkata, Hyderabad, Chennai, Bangalore)

*Project Sessions are also available in Online Mode

WHY CHOOSE US?

1-ON-1 DEDICATION

- Live interactive session with expert for every individual
- Each session is guided by industrial expert
- 24*7 seamless technical support from our dedicated team



1.5 YEARS SUBSCRIPTION

- Limitless access for all the learning materials, live batches, and project sessions
- Professionals get to switch between weekdays and weekends
- Make your learning calendar as per your convenience

WHY CHOOSE US?

Learnbay



- ✓ **100%** Live Interactive Sessions from Expert
- ✓ **1.5 Years** of Subscription to **Live Classes**
- ✓ **Hybrid Model**
Live + Classroom Project Sessions in 7+ Cities
- ✓ **5 Guaranteed** Interview calls
- ✓ **1:1 Doubt Clearing** Session with Expert

Others



- ✗ **Boring** Recorded Sessions
- ✗ Access to only **Recorded Videos**
- ✗ **No Model**
No Classroom sessions for projects
- ✗ **No Guaranteed** Interview calls
- ✗ **No Doubt Clearing** Session with Expert

Course Completion Certificate



Internationally
Recognized Certificate



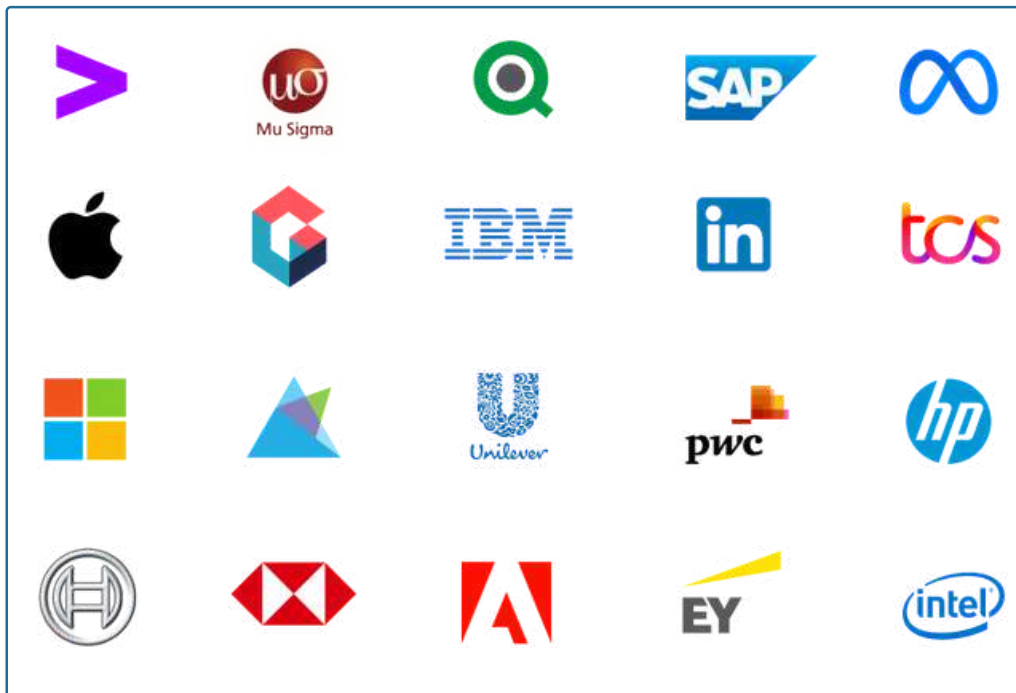
Validate your Data
Science skills



Get acknowledged in
IT sector

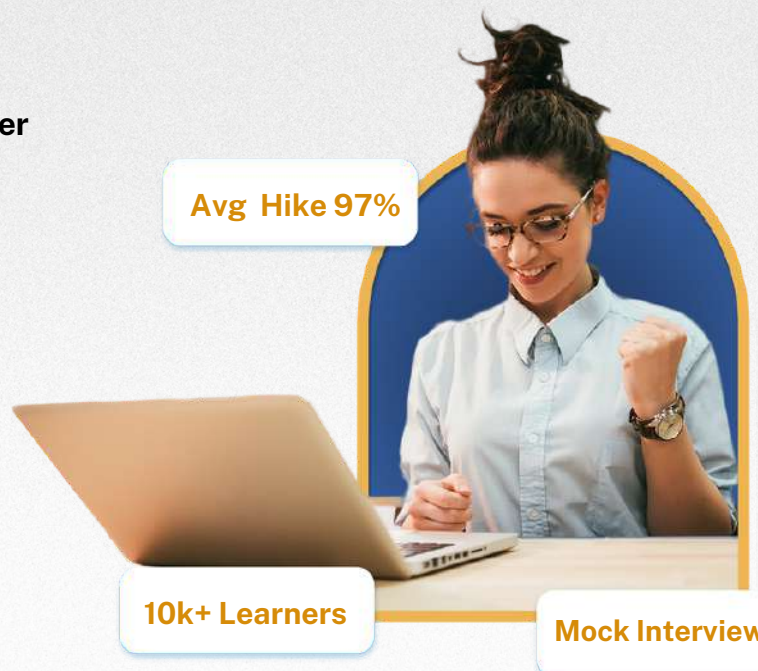
5 GUARANTEED INTERVIEW CALLS

We have partnered with **250+ Top MNC'S & FinTech Startups** across the globe to offer genuine job leads. Most of our learners were hired for their **dream jobs** one month before the course completion.



Dedicated Placement Cell for Working Professionals to ensure a smooth Career Transition

- Prioritize growth and salary hike with in-demand skillset
- Make a transition without losing the designation or wasting your existing experience





Success Stories



Preksha Mishra

Designation

Lead Data Scientist

Got placed in:

HCL

With a hike of

140%



The course structure is excellent with emphasis on concept building and tools & software at the same time. The support team is excellent and supportive and quite agile to respond to doubts.

Domain: Telecom



Saurabh Kumar

Designation

Data Scientist & Statistician

Got placed in:

Teleperformance

With a hike of

135%



When I joined Learnbay I did not have any knowledge apart from the very basics. I gradually build my concept via various trainers and get trained in data science with strong knowledge/concepts.

Domain: Math Professor



Ritesh Kumar

Designation

Associate Consultant

Got placed in:

Capgemini

With a hike of

150%



I knew nothing about data science before I joined Learnbay. But through a variety of instructors, I steadily developed my notion and received solid knowledge and conceptual training in data science with hike of 119%.

Domain: Mechanical



[Read More Reviews](#)





Success Stories



Shrvanathi A

Designation

Data Scientist

Got placed in:



With a hike of

230%

Learnbay has helped me a lot to learn data science applications in the e-commerce industry. The live class concept was really helpful in receiving proper DS training. Thanks to all my mentors and the placement team.

Domain: **Mechanical**



Mohd. Israr

Designation

Data Scientist

Got placed in:



With a hike of

210%

Thanks to the Learnbay data science course and excellent guidance, I was able to ace the TCS interview and secure a job with a 400% pay raise. All of the real-world time projects helped me develop my concepts as a data scientist.

Domain: **Mechanical**



Ankit Biswas

Designation

Data Scientist

Got placed in:



With a hike of

180%

I come from a nontechnical background. However, with Learnbay's well-structured course, amazing mentorship, and consistent support, I was able to not only enhance my skills but also land my dream career.

Domain: **Software Engineer**



[Read More Reviews](#)



TRANSITION PROCESS

LEARNING PHASE

Learn updated tools and modules from basic to advance by industry expert



ASSESSMENT

Evaluate your skillset with real-time case studies and assignments



8 LIVE PROJECTS

Work on domain specific industrial projects and make your experience relevant



IBM CERTIFICATION

Earn Certification from IBM and get globally recognized



PROFILE GROOMING

Get Interview Ready with experts. Attain Resume Build-Up, 1:1 Mock Interview



5 INTERVIEW CALLS

Guaranteed Interviews call from FinTech Startups and top MNCs



PROGRAM FEE & FINANCING



We provide a choice of financing alternatives to make it more cost-effective, and make our programs accessible for all learners

No Cost EMI

We have partnered with the following financing companies to provide competitive finance options at 0% interest rate with no hidden costs.



To know more about course fees & scholarship, [click here](#)

Financing as low as

Rs. 6,391/month

For one-time payment



Internet
Banking



Credit/Debit
Card

Program Fee

Rs. 65000/- +18% GST

Rs. 76,700/-



PROGRAM OUTLINE

MODULE
00

Cohort Orientation + Special Programming Classes

8 Hrs

TERM
01

Python Programming (Basic + Advance)

Python, Anaconda, GitHub, Pandas

40 Hrs

TERM
02

Statistics and Machine Learning

Matplotlib, Scikit-Learn, Seaborn

70 Hrs

TERM
03

Data Science Tools

SQL, MongoDB, Tableau, PowerBI,
Big Data & Spark Analytics, Time Series

86 Hrs

REAL-TIME INDUSTRIAL PROJECTS

Domain: HR



1. Career progression planning of employees with workforce defections & efficiency

IBM intends to boost its HR department by identifying employees' masked inconsistency. They need models to identify the graphical variations in their 14000+ employees' performances. Help them build models with your regressions and other ML abilities.

Machine Learning

Python

SQL

PySpark

Domain: Marketing



2. Descriptive study of trends and irregularities with prediction analysis for conversion.

Swiggy seeks a broad marketing campaign. But they need automated keyword generation tools. They also require proper message preparation and delivery of the same to the right audience at the right time. You can help them with text analytics and NLP-based keyword research.

Exploratory Data Analysis

Big Data

NLP

Domain: Sales



3. Forecasting future sales with trends and price maximization

BMW customers can sell old vehicles, but rivals provide superior resale prices. BMW's data science-powered software will deliver the greatest market value for used vehicles based on Km travelled, daily price changes, production dates, etc. Such tasks build analytical abilities.

Scikit-learn

XG Boost

Customer Segmentation

Domain: Healthcare



4. Understanding covid-19 cases and fatality rate by time series forecasting

Samsung will launch a new healthcare app soon. The key goal of this app is an accurate human activity tracking and providing relevant health-related recommendations. Continuous analysis of a massive amount of mobile data is required for such an app.

Supervised Machine Learning

Python (Pandas Library)

REAL-TIME INDUSTRIAL PROJECTS

Domain: BFSI



5. Learn and develop classification techniques for the digital transformation of banking

JPMorgan offers tax-friendly insurance choices. You can help them forecast insurance premiums. Targeted marketing using your random forest algorithm skills can help obtain better premium values.

Data Analytics

Matplotlib

Logical Regression

Domain: Media



6. Building a content recommendation model on the basis of regional viewer categorization

Netflix is a global entertainment video streaming site. They offer content in various regional languages. Build a local recommendation engine for Netflix customers residing in south Bangalore on their weekend and weekdays activities, utilizing NLP.

ML Customer Segmentation

Python (Data-Preprocessing)

Domain: Transportation



7. Reduction of waiting time via a highly precise forecasting model

Make a demand forecasting model based on specific time period rider demands. Such a model will help both riders and cab drivers to ensure the least possible waiting time. You can include measures like latitude and longitude identification.

Machine Learning

Hadoop

Time Series Analysis

Domain: Oil, Gas and energy



8. Understanding in-depth about logging while drilling (LWD) technique

Saudi Aramco company is working on the development of high-efficiency drilling models. Use the bright sides of big data analytics to identify the most cost-effective and highly productive drilling sites.

Matplotlib in Python

Big Data

8 hours

1. Cohort Orientation

A brief introduction to tools related to data

Learn about particular real-time projects and Capstone projects

Data and its impact on career opportunities

Fundamental relevance of projects using data

Role of data in businesses

2. Cohort Orientation

Significance of data in decision-making

Scope of data in research and development

Utilizing data, to enhance industrial operations and management

Data in performance evaluation

Data in customer segmentation

3. Fundamentals of programming

Types of code editors in python

Introduction to Anaconda & Jupyter notebook

Flavors of python

Introduction to Git, GitHub

Python Fundamentals

Source code vs Byte code vs

Machine code

Compiler & Interpreter

Memory Management in Python

4. Fundamentals of Statistics

Mean, Median, Mode

Standard Deviation, Average.

Probability, permutations, and combinations

Introduction to Linear Algebra

Tools Covered**git****github**

Note: This module 0 is for those who are from a non-technical background like Mechanical, BBA, MBA, B.Com, M.Com, etc. Or for those who work in Non-IT sectors, who are new to programming & statistics (basic mathematics)

40 hours

1. Programming Basics & Environment Setup

Installing Anaconda, Anaconda Basics and Introduction
Get familiar with *version control, Git and GitHub*.
Basic Github Commands.
Introduction to Jupyter Notebook environment. Basics Jupyter notebook Commands.
Programming language basics.

2. Python Programming Overview

Python Overview
Python 2.7 vs Python 3
Writing your First Python Program
Lines and Indentation, Python Identifiers
Various Operators and Operators Precedence
Getting input from User, Comments, Multi line Comments.

3. Strings, Decisions & Loop Control

Working With Numbers, Booleans and Strings, String types and formatting, String operations
Simple if Statement, if-else Statement
if-elif Statement.
Introduction to while Loops, for Loops, Using continue and break.

Class hands-on :

6 programs/coding exercise on string, loop and conditions in classroom

4. Python Data Types

List, Tuples, Dictionaries
Python Lists, Tuples, Dictionaries
Accessing Values, Basic Operations
Indexing, Slicing, and Matrixes
Built-in Functions & Methods
Exercises on List, Tuples And Dictionary

5. Functions And Modules

Introduction To Functions
Defining & Calling Functions
Functions With Multiple Arguments.
Anonymous Functions - Lambda
Using Built-In Modules, User-Defined Modules, Module Namespaces, Iterators And Generators

Class hands-on :

8+ Programs to be covered in class of functions, Lambda, modules, Generators and Packages.

Class hands-on (Python Data Types):

- Program to convert tuple to dictionary*
- Remove Duplicate from Lists*
- Python program to reverse a tuple*
- Program to add all elements in list.*
- + 3 more programs to be covered in class*

40 hours

6. File I/O And Exceptional Handling and Regular Expression

Opening and Closing Files
open Function, file Object Attributes
close() Method ,Read,write,seek.
Exception Handling, try-finally Clause
Raising an Exceptions, User-Defined Exceptions
Regular Expression- Search and Replace

Regular Expression Modifiers
Regular Expression Patterns

Class hands-on :

10+ Programs to be covered in class from File IO, Reg-ex and exception handling.

7. Data Analysis Using Numpy

Introduction to **Numpy**. Array Creation, Printing Arrays, Basic Operation - Indexing, Slicing and Iterating, Shape Manipulation - Changing shape, stacking and splitting of array
Vector stacking, Broadcasting with Numpy, *Numpy for Statistical Operation.*

8. Data Analysis Using Pandas

Pandas : Introduction to Pandas
Importing data into Python
Pandas Data Frames, Indexing Data Frames ,Basic Operations With Data frame, Renaming Columns, Subsetting and filtering a data frame.

Assignment 1 (Week 2):

10 Coding exercises on Python Basics - Variables, Operators, Strings, Loops, Control Statement

Assignment 2 (Week 3):

10 Python programs and practice set on List, Tuples, Dictionaries & Matrices operations

Assignment 3 (Week 4):

10 Coding exercises on Functions, Lambda, Input-Output, File and Regular Expression

40 hours

9. Data Visualization using Matplotlib

Matplotlib: Introduction, plot(), Controlling Line Properties, Subplot with Functional Method, Multiple Plot, Working with Multiple Figures, Histograms

10. Data Visualization using Seaborn

Seaborn :
Intro to Seaborn And Visualizing statistical relationships , Import and Prepare data. Plotting with categorical data and Visualizing linear relationships.
Seaborn Exercise

3 Case Study on Numpy, Pandas, Matplotlib

1 Case Study on Pandas And Seaborn

Assessment Test in Python :

2 hour of Assesment Test in Python (Coding & Objective Questions)

REAL TIME USE CASES IN PYTHON TO BE COVERED IN CLASS

WITH 5 ASSIGNMENTS



1. Fundamentals of Math and Probability

Probability distributed function & cumulative distribution function. Conditional Probability, Baye's Theorem

Problem solving for probability assignments

Random Experiments, Mutually Exclusive Events, Joint Events, Dependent & Independent Events

2. All about Population & Sample

Population vs Sample, Sample Size Simple Random Sampling, Systematic Sampling, Cluster Sampling, Stratified Sampling, Convenience Sampling, Quota Sampling, Snowball Sampling and Judgement Sampling

3. Introduction to Statistics, Statistical Thinking

Variable and its types Quantitative, Categorical, Discrete, Continuous, *all with examples

Five Point Summary and Box Plot

Outliers, Causes of Outliers, How to treat Outliers, I-QR Method and Z-Score Method

4. Descriptive Statistics

Measures of Central Tendency – Mean, Median and Mode Measures of Dispersion – Standard Deviation, Variance, Range, IQR (Inter-Quartile Range)

Measure of Symmetry/ Shape – Skewness and Kurtosis

5. Inferential Statistics

Central Limit Theorem Point estimate and Interval estimate Creating confidence interval for population parameter Characteristics of Z-distribution and T-Distribution. Type of test and rejection region. Type of errors in Hypothesis Testing

6. Hypothesis Testing

Type of test and Rejection Region Type o Errors-Type 1 Errors, Type 2 Errors. P value method, Z score Method. The Chi-Square Test of Independence.

Regression. Factorial Analysis of Variance. Pearson Correlation Coefficients in Depth. Statistical Significance

Null and Alternative Hypothesis
One-tailed and Two-tailed Tests,
Critical Value, Rejection region,
 Inference based on Critical Value

Binomial Distribution

Assumptions of Binomial Distribution,
 Normal Distribution, Properties of
 Normal Distribution, Z table,
 Empirical Rule of Normal Distribution
 & Central Limit Theorem and its
 Applications

7. Linear Algebra

Dot Product, Projecting Point on
 Axis.

Matrices in Python, Element
 Indexing, Square Matrix, Triangular
 Matrix, Diagonal Matrix, Identity
 Matrix, Addition of Matrices, Scalar
 Multiplication, Matrix Multiplication,
 Matrix Transpose, Determinant,
 Trace

**T-Test, Analysis of variance (ANOVA),
 and Analysis of Covariance (ANCOVA)**

Regression analysis in ANOVA

Class Hands-on:

Problem solving for C.L.T
 Problem solving Hypothesis Testing
 Problem solving for T-test, Z-score test

**Case study and model run for
 ANOVA, ANCOVA**

8. Data Processing & Exploratory Data Analysis

What is Data Wrangling
 Data Pre-processing and cleaning?
 How to Restructure the data?
 What is Data Integration and
 Transformation

9. EDA

Finding and Dealing with Missing
 Values.

What are Outliers?

Using Z-scores to Find Outliers.

**Bivariate Analysis, Scatter Plots and
 Heatmaps.**

Introduction to Multivariate
 Analysis

*Note: Problem-Solving Techniques
 and Case Studies using Statistics will
 be covered in class from week 2.*

**Statistics Assignments : Total 4
 practice set and Assignments from
 Statistics**

40 hours

1. Machine Learning Introduction

Definition, Examples, Importance of Machine Learning
Definition of ML Elements: Algorithm, Model, Predictor Variable, Response Variable, Training - Test Split, Steps in Machine Learning,
ML Models Type: Supervised Learning, Unsupervised Learning and Reinforcement Learning.

2. Regression and Classification Models

Definition of regression, OLS Algorithm, Sum of Squares of residuals, Gradient Descent Algorithm, Cost Function
Evaluation Metrics for Regression Model: MAE, MSE, RMSE, R Square, Adjusted R Square

3. Linear Regression Model

Comparing MAE, MSE, and RMSE. Significance of Adjusted R square. Overfitting and Underfitting. Bias and Variance.
Regularization methods:
Ridge and Lasso
Multicollinearity, VIF. Using Python library **Sklearn** to create the Linear Regression Model and evaluate the model created.

4. Data Preprocessing

Types of Missing values (MCAR, MAR, MNAR), Methods to handle missing values
Outliers, Methods to handle outliers: IQR Method, Z Method
Feature Scaling: Definition, Methods: Absolute Maximum Scaling, Min-Max Scaler, **Normalization, Standardization, Robust Scaling**

5. Data Preprocessing

Encoding the data: Definition, Methods: OneHot Encoding, Mean Encoding, Label Encoding, Target Guided Ordinal Encoding

6. Logistic Regression Model

Definition. Why is it called the "Regression model"?
Sigmoid Function, Transformation & Graph of Sigmoid Function

7. Evaluation Metrics for Classification model

Confusion Matrix, Accuracy, Misclassification, **TPR, FPR, TNR, Precision, Recall, F1 Score, ROC Curve**, and AUC. Using Python library Sklearn to create the Logistic Regression Model and evaluate the model created

40 hours

8. K Nearest Neighbours Model

Definition, Steps in KNN Model, Types of Distance: Manhattan Distance, Euclidean Distance, 'Lazy Learner Model'.

Confusion Matrix of Multi Class Classification

Using Python library Sklearn to create the K Nearest Neighbours Model and evaluate the model

9. Decision Tree Model

Definition, Basic Terminologies, Tree Splitting Constraints, Splitting Algorithms:

CART, C4.5, ID3, CHAID

Splitting Methods:

GINI, Entropy, Chi-Square, and Reduction in Variance

Using Python library Sklearn to create the Decision Tree Model and evaluate the model created

10. Random Forest Model

Ensemble Techniques: Bagging/bootstrapping & Boosting. Definition of Random Forest, OOB Score

K-Fold Cross-Validation**11. Hyperparameter Tuning**

GridSearchCV, Variable Importance. Using Python library Sklearn to create the **Random Forest Model** and evaluate the model created.

Use cases

12. Naive Baye's Model

Definition, Advantages, **Baye's Theorem** Applicability, Disadvantages of **Naive Baye's Model**, Laplace's Correction, Types of Classifiers: Gaussian, Multinomial and Bernoulli Using Python library Sklearn to create the Naive Baye's Model and evaluate the model created

Case Study

- Business Case Study for Kart Model
- Business Case Study for Random Forest
- Business Case Study for SVM
- To classify an email as spam or not spam using logistic Regression.
- Application of Linear Regression for Housing Price Prediction

40 hours

13. K Means and Hierarchical Clustering

Definition of Clustering, Use cases of Clustering
K Means Clustering Algorithm,
Assumptions of **K Means Clustering**
Sum of Squares Curve or Elbow Curve

14. Hierarchical Clustering

Dendrogram, Agglomerative Clustering, Divisive Clustering,
Comparison of K Means Clustering and Hierarchical Clustering
Using Python library Sklearn to create and evaluate the clustering model

15. Principal Component Analysis(PCA):

Definition, Curse of Dimensionality, Dimensionality Reduction
Technique, When to use PCA,
Use Cases
Steps in **PCA**, EigenValues and EigenVectors, Scree Plot.
Using Python library Sklearn to create Principal Components

16. Support Vector Machine(SVM)

Model: Definition, Use Cases,
Kernel Function, Aim of Support Vectors, Hyperplane, Gamma Value, **Regularization Parameter**
Using Python library Sklearn to create and evaluate the SVM Model

Summary of all Machine Learning Models and Discussion about the Capstone Project

Note :

All Machine Learning Algorithms are covered in depth with real time case studies for each algorithm.

Once 60% of ML is completed, Capstone Project will be released for the batch.

Case Study

- Recommendation Engine for e-commerce/retail chain
- Twitter data analysis using NLP



14 hours

1. SQL and RDBMS

RDBMS And SQL Operations.
Single Table Queries - SELECT, WHERE, ORDER BY, Distinct, And, OR
Multiple Table Queries: INNER, SELF, CROSS, and OUTER, Join, Left Join, Right Join, Full Join, Union

2. Advance SQL

Advance SQL Operations
Data Aggregations and summarizing the data
Ranking Functions: Top-N Analysis
Advanced SQL Queries for Analytics

3. NoSQL, HBase & MongoDB

NoSQL Databases
Introduction to HBase
HBase Architecture, HBase Components, Storage Model of HBase.
HBase vs RDBMS
Introduction to **Mongo DB**, CRUD
Advantages of MongoDB over RDBMS

4. JSON Data & CRUD

Basics and CRUD Operation
Databases, Collection & Documents
Shell & MongoDB drivers
What is JSON Data
Create, Read, Update, Delete
Finding, Deleting, Updating, Inserting Elements
Working with Arrays
Understanding Schemas and Relations

5. Programming with SQL

Mathematical Functions
Variables
Conditional Logic
Loops
Custom Functions
Grouping and Ordering
Partitioning
Filtering Data
Subqueries

Assignments

- Working with multiple tables
- Practice Joins, Grouping and Subqueries
- Using GROUP BY and HAVING Clauses
- Practice Aggregation Queries

14 hours

1. Introduction to MongoDB

What is MongoDB
Characteristics and Features
MongoDB Ecosystem
Installation process
Connecting to MongoDB database
Introduction to NoSQL
Introduction of MongoDB module
What are Object Ids in MongoDB

2. MongoDB (Advance)

MongoDB Use cases
MongoDB Structures
MongoDB Shell vs MongoDB Server
Data Formats in MongoDB
MongoDB Aggregation Framework
Aggregating Documents

2. MongoDB (Advance)

Working with MongoDB Compass
& exploring data visually
Understanding Create, Read, Update, Delete
Schemas & Relations
Document Structure
Working with Numeric Data
Working on Scheme Designing

Tool Covered



Assignment

Obtain the data in the format you want by formulating queries that are both effective and high-performing.

14 hours

1. Introduction to Tableau

Connecting to data source
Creating dashboard pages
How to create calculated columns
Different **charts**

2. Visual Analytics

Getting Started With Visual Analytics
Sorting and grouping
Working with sets, set action
Filters: Ways to filter, Interactive Filters
Forecasting and Clustering

3. Dashboard and Stories

Working in Views with Dashboards and Stories
Working with Sheets
Fitting Sheets
Legends and Quick Filters
Tiled and Floating Layouts, Floating Objects

4. Tableau (Advance)

Mapping
Coordinate points
Plotting Latitude and Longitude
Custom Geocoding
Polygon Maps
WMS and Background Image

Hands-on Assignments

- Connecting data source and data cleansing
- Working with various charts
- Deployment of Predictive model in visualization

Tool Covered

14 hours

1. Getting Started With Power BI

Installing Power BI Desktop and
Connecting to Data

Overview of the Workflow in Power
BI Desktop

Introducing the Different Views of
the Data Mode

Query Editor Interface

Working on Data Model

2. Programming with Power BI

Working with Time Series

Understanding aggregation and
granularity

Filters and Slicers in Power BI Maps

Scatterplots and BI Reports

Connecting Dataset with Power BI

Creating a Customer

Segmentation Dashboard

Analyzing the Customer
Segmentation Dashboard

Assignments

Create Bar charts

Create Pie charts

Create Tree maps

Create Donut Charts

Create Waterfall Diagrams

Creating Table Calculations for
Gender

Tool Covered

Note: All the assignments will be
covered in-depth with real-time
examples

16 hours

1. Introduction To Hadoop & Big Data

Distributed Architecture - A Brief Overview. Understanding **Big Data**
Introduction To Hadoop, Hadoop Architecture
HDFS, Overview of MapReduce Framework
Hadoop Master: Slave Architecture
MapReduce Architecture
Use cases of **MapReduce**

2. What is Spark

Introduction to Spark RDD
Introduction to **Spark SQL** and Data frames
Using **R-Spark** for machine learning
Hands-on:
Installation and configuration of Spark
Using **R-Spark** for machine learning programming

3. Getting to know PySpark

Pyspark Introduction
Pyspark Environment Setup
pySpark - Spark context
RDD , Broadcast and Accumulator
Sparkconf and Sparkfiles
Spark MLlib Overview Algorithms and utilities in Spark Mlib

Hands-on

Map reduce Use Case 1: Youtube data analysis

Map reduce Use Case 2: Uber data analytics

Spark RDD programming

Spark SQL and Data frame programming

Tools Covered



14 hours

1. Introduction to Time Series Forecasting

Basics of Time Series Analysis and Forecasting
Method Selection in Forecasting
Moving Average (MA) Forecast Example
Different Components of Time Series Data
Log Based Differencing, Linear Regression for Detrending

2. Introduction to ARIMA Models

ARIMA Model Calculations, Manual ARIMA Parameter Selection
ARIMA with Explanatory Variables
Understanding Multivariate Time Series and their Structure
Checking for Stationarity and Differencing the MTS

Case Study

- Time series classification of smartphone data to predict user behavior

Note: All the assignments and case studies will be covered in-depth with real-time examples

Case Study

- Performing Time Series Analysis on Stock Prices
- Time series forecasting of sales data

Contact Us



“

*Learn Here,
Lead Anywhere*

Click on the icon to follow us!



Address

#1090

1st floor, 18th Cross Rd,
above Sangam Sweets,
Sector 3, HSR Layout,
Bengaluru, Karnataka
560102

Book a counselling session
with expert!

[Book a session](#)