



COURSE BROCHURE

DATA SCIENCE

WHAT IS DATA SCIENCE ?

Data Science is all about “using various techniques, algorithms to analyze large amounts of datasets (both structured & Unstructured), to extract useful data insights, thus applying them in various business domains.”

WHY THERE'S A DEMAND FOR DATA SCIENTISTS ?

Data is being generated day by day at a massive rate and in order to process such massive data sets, Big Firms, Companies are hunting for good data scientists to extract valuable data insights from these data sets and using them for various business strategies, models, plans

HOW DO THEY DO IT ?



HOW GMAIL FILTERS YOUR EMAILS IN THE SPAM AND NON-SPAM CATEGORIES?



HAVE YOU EVER WONDERED HOW AMAZON; EBAY SUGGEST ITEMS FOR YOU TO BUY?



HOW NETFLIX PREDICTS THE SHOWS OF YOUR LIKING?

These are the few questions we ponder from time to time. In reality, doing such tasks are impossible without the availability of data.



DATA SCIENCE IS ALL ABOUT USING DATA TO SOLVE PROBLEMS.

- 1 The problem could be decision making such as identifying which email is spam and which is not
- 2 Or a product recommendation such as which movie to watch?
- 3 Or predicting the outcome such as who will be the next President of the USA?.



So, the core job of a data scientist is to understand the data, extract useful information out of it and apply this in solving the problems.

MODULES COVERED

- What is Data science?
- What is Machine Learning?

- Implementation of Data science in real time industries
- What is the difference between Machine learning, Deep Learning and AI?

- Role of Data Scientist
- What are the skills required to become a Data Scientist
- Introduction to Mathematics and Statistics



LEARN STATISTICS

If Data Science is a language, then statistics is basically the grammar. Statistics is basically the method of analyzing, interpretation of large data sets. When it comes to data analysis and gathering insights, statistics is as note worthy as air to us. Statistics help us understand the hidden details from large datasets.

MODULES COVERED

DATA TYPES

- Continuous Variables
- Ordinal Variables
- Categorical Variables

CENTRAL TENDENCIES

- Mean, Median, Mode
- Standard Deviation
- Probabilities
- Variance
- Plotting

BASIC AND BUSINESS STATISTICS

- Distance Formula
- Euclidean Distance
- Cosine Similarity
- Mean Squared Error
- Root Mean Square Error

DESCRIPTIVE STATISTICS

- Sampling
- Random Sampling
- Systematics Sampling
- Stratified Sampling



LEARN PYTHON

The First and Foremost Step Towards Data Science should learning be a programming language (i.e., Python). Python is the most common coding language, used by the Majority of Data Scientist, because of its simplicity, versatility and being pre-equipped with powerful libraries (like NumPy, SciPy, and Pandas) useful in data analysis and other aspects in Data Science. Python is an open-source language and supports various libraries.

MODULES COVERED

INTRODUCTION TO PYTHON

BASICS OF PYTHON

- Number system
- Operators
- Loops
- Functions
- List
- Tuple
- Set
- Dictionaries

DIFFERENT
PLOTTING
TECHNIQUES

INDEXING AND
SLICING

EXPLORATORY
DATA ANALYSIS

DEPLOYMENT
METRICS

VALIDATION
TECHNIQUES

MODEL
ACCURACY

DATA
VISUALIZATION

BASICS OF PYTHON

- Introduction to Pandas
- Introduction to Numpy
- Scipy
- Matplotlib
- Seaborn



ACQUAINTANCE WITH EDA (EXPLORATORY DATA ANALYSIS)

EDA (Exploratory data analysis) is the most important aspect in the vast field of data science. It includes analysing various data, variables, various data patterns, trends and extracting useful insights from them with help of various graphical and statistic methods. EDA identifies various pattern which Machine learning algorithm might fail to identify. It includes all Data Manipulation, Analysis, and Visualization.

MACHINE LEARNING & DEEP LEARNING

Machine learning is the core skill required to be a Data Scientist.

Machine learning is used to build various predictive models, classification models, etc., and is being used by big firms, Companies to Optimize their planning as per the predictions. For example, Car Price prediction

Deep Learning on the other hand is and an advanced version of Machine Learning which deploys the use of Neural Network, a framework that combines various machine learning algorithms for solving various tasks, for training data. Various Neural networks are recurrent neural network (RNN) or a convolutional neural network (CNN) etc

MODULES COVERED IN MACHINE LEARNING

INTRODUCTION TO MACHINE LEARNING

- Supervised Learning
- Unsupervised Learning
- Statistical Modelling

MACHINE LEARNING ALGORITHM

- Linear Regression
- Stochastic Gradient
- Descent Model
- Cost Function
- Multiple Linear
- Regression
- Assumptions of
- Linear Regression

LOGISTIC REGRESSION

- Sigmoid Function
- Solving Classification Problems
- The Logistic Model
- Confusion Matrix

NAÏVE BAYES

- Naïve Bayes' Probability

SUPPORT VECTOR MACHINE

- Maximum Margin Classifier
- What is Hyperplane
- SVM Linear, Non-Linear Model
- Kernel and Power Kernel Method

K-NEAREST NEIGHBOR

- How to find an Appropriate K value?
- Advantages and Disadvantages of KNN
- KNN for regression and Classification
- How to find the Nearest Neighbor

MODULES COVERED IN MACHINE LEARNING

RANDOM FOREST

- What is Random Forest?
- Classification with Random Forest
- Implementation of Random Forest
- Introduction to Adaptive Boosting

DECISION TREES

- How to build Decision Tree
- Information Gain
- Entropy
- Gini index and Entropy methods

ENSEMBLE TECHNIQUE

- Bagging
- Boosting
- XGBoost
- Gradient Boosting Methods

FORECASTING ANALYSIS

- Why Forecasting?
- Types of Forecasting
- Techniques for Trend
- Techniques for Seasonality
- Predictive Analysis

TIME SERIES ANALYSIS

- ARIMA
- Trend and Seasonality

INTRODUCTION TO UNSUPERVISED MACHINE LEARNING

- K-means Clustering
- Apriori Algorithm

MODULE COVERED IN DEEP LEARNING

Introduction to Deep Learning

Artificial Neural Network (ANN)

Convolution Neural Network (CNN)

Recurrent Neural Network (RNN)

Long short-term Memory (LSTM)

Activation Function

Sigmoid

1. RELU 2. ELU 3. PRELU 4. Softmax



Testing and Validation of the Machine Learning Model after Deployment Should Be done in order to check its effectiveness and accuracy. Testing is an Important Step in Data Science for keeping the efficiency and effectiveness of the ML model in check.

MODULES COVERED

How to define problem/ prepare problem

Load the libraries

Load the respective dataset

How to summarize the data

Descriptive Statistics

Data Visualization

Exploratory Data Analysis

How to prepare Data

Data cleaning

Feature selection

How to Evaluate Algorithm

Validation test

Evaluation Metrics

How to improve

Accuracy/Results

Hyper tuning Parameter

Ensemble techniques

Prediction on Validation

Dataset

ANALYTICAL VISUALIZATION IN TABLEAU

1

As the market-leading choice for modern business intelligence, Tableau's analytics platform makes it easier for people to explore and manage data, and faster to discover and share insights that can change businesses and the world.

2

Tableau helps people see and understand data, whether they're an analyst, data scientist, student, teacher, executive, or business user. From connection through collaboration, Tableau is the most powerful, secure, and flexible end-to-end analytics platform.

MODULE COVERED

Introduction to Tableau

Features of Tableau

Connecting Tableau to a Data File

Navigating Tableau –
Dimension and Measure

Creating a Calculating Field

Adding Colors and Labels

Creating groups and bins

Creating groups and bins

How to do sorting

MODULE COVERED

Doing Analytics in Tableau

How to Forecast

How to custom Distribution
band

How to custom reference line

How to model Trend line

How to summarize Totals,
Constant line and Average lines

Calculations in Tableau

Different Charts

Tableau Dashboard

ADDITIONAL SUPPORT

End to End Project Explanation

2 Real time Project Explanation

Assignments and Assessments

All classes with real time examples

Real time interview questions

Resume Building

Interview preparation

Mock interviews

Contact Us



+91-40-48538611

+91-40-48538711



www.ihaconsulting.in



trainings@ihaconsulting.in



(+91) 8375985217