SAAG Technologies

by Pralhad. P

# Course - Data science and Machine Learning Using Python

## About the course

This is a complete Data Science specialization training course provides you detailed learning in data science, data analytics, project life cycle, data acquisition, analysis, statistical methods, and machine learning.

## Duration of the Course – 50Hrs with including 5 projects

## Key Learning Objectives

- Gain fundamental knowledge of what is Data Science and Machine Learning
- Learn about Data Science in a business context and what is the future of Data Science
- Understand Data Science applications and discover some use cases for Data Science

## Course Contents

| 1 | Introduction to Data Science |
|---|---|
| | • What is data Science? - Introduction<br>• Importance of Data Science<br>• Demand for Data Science Professional<br>• Lifecycle of data science<br>• Tools and Technologies used in data Science<br>• Business Intelligence vs Data Science<br>• Role of a data scientist |

PART A - INTRODUCTION TO STATISTICS

| 2 | Fundamentals of Math and Probability |
|---|---|
| **Theory** | • Basic understanding of linear algebra, Matrices, vectors<br>• Addition and Multiplication of matrices<br>• Fundamentals of Probability<br>• Probability distributed function and cumulative distributed function |
| **Hands on** | • Problem solving using Python for vector manipulation<br>• Problem solving for probability assignments |

| 3 | Descriptive Statistics |
|---|---|
| Theory | • Describe or summaries a set of data Measure of central tendency and measure of dispersion.<br>• The mean, median, mode, curtosis and skewness Computing Standard deviation and Variance. Types of distribution |
| Hands on | • 5 Point summary Box Plot Histogram and Bar Chart Exploratory analytics Python Methods |

| 4 | Inferential Statistics |
|---|---|
| Theory | • What is inferential statistics Different types of Sampling techniques Central Limit Theorem<br>• Point estimate and Interval estimate<br>• Creating confidence interval for population parameter Characteristics of Z-distribution and T-Distribution Basics of<br>**Hypothesis Testing**<br>• Type of test and rejection region<br>• Type of errors in Hypothesis resting, Type-l error and Type-ll errors<br>• P-Value and Z-Score Method<br>• T-Test, Analysis of variance (ANOVA) and Analysis of Co variance (ANCOVA)<br>• Regression analysis in ANOVA |
| Hands on | • Problem solving for C.L.T Problem solving Hypothesis Testing Problem solving for T-test, Z-score test<br>• Case study and model run for ANOVA, ANCOVA |

| 5 | Inferential Statistics |
|---|---|
| Theory | • What is inferential statistics Different types of Sampling techniques Central Limit Theorem<br>• Point estimate and Interval estimate<br>• Creating confidence interval for population parameter Characteristics of Z-distribution and T-Distribution Basics of<br>**Hypothesis Testing**<br>• Type of test and rejection region<br>• Type of errors in Hypothesis resting, Type-l error and Type-ll errors<br>• P-Value and Z-Score Method<br>• T-Test, Analysis of variance (ANOVA) and Analysis of Co variance (ANCOVA)<br>• Regression analysis in ANOVA |
| Hands on | • Problem solving for C.L.T Problem solving Hypothesis Testing Problem solving for T-test, Z-score test<br>• Case study and model run for ANOVA, ANCOVA |

| 6 | Introduction to Machine Learning |
|---|---|
| **Theory** | • What is Machine Learning? What is the Challenge?<br>• Introduction to Supervised Learning, Unsupervised Learning<br>• What is Reinforcement Learning?<br><br>**1.Linear Regression**<br>• Introduction to Linear Regression with Multiple Variables Disadvantage of Linear Models Interpretation of Model Outputs Understanding Covariance and Co linearity Understanding Heteroscedasticity<br><br>**2.Logistic Regression**<br>• Introduction to Logistic Regression. – Why Logistic Regression.<br>• Introduce the notion of classification Cost function for logistic regression Application of logistic regression to multi-class classification.<br>• Confusion Matrix, Odd's Ratio and ROC Curve Advantages and Disadvantages of Logistic Regression |

PART B – UNDERSTANDING AND IMPLEMENTING MACHINE LEARNING

| | |
|---|---|
| | **3. Decision Trees and Supervised Learning**<br>• Decision Tree – data set<br>• How to build decision tree? Understanding CART Model Classification Rules-Over fitting Problem Stopping Criteria and Pruning<br>• How to find final size of Trees? Model A decision Tree<br><br>**4. Random Forests**<br>• How random forest works?<br>• Bagging and Bootstrapping<br>• Out of Bag Error<br>• Proximity |
| **Hands on** | • Application of Linear Regression for Housing Price Prediction<br>• Employee performance Prediction<br>• Business Case Study for CART Model<br>• Purchase Prediction |

| 7 | Unsupervised Learning |
|---|---|
| **Theory** | **Hierarchical Clustering**<br>• K-Means algorithm for clustering – groupings of unlabeled data points.<br>• Principal Component Analysis(PCA)- Data Independent components analysis(ICA)<br><br>**Anomaly Detection**<br>• Association rules |
| **Hands on** | • Churn Prediction |

PART C – PYTHON PROGRAMMING

| 8 | PYTHON PROGRAMMING |
|---|---|
| **Theory** | • Introduction<br>• Installing python<br>• Python Basics: Expressions, Data Types & Variables<br>• Control Structures: if-else, loops break, continue etc...<br>• Scoping Rules<br>• Functions<br>• Working with Lists<br>• Working NumPy arrays<br>• Reading and Writing Files<br>• Organizing Files<br>• Working with documents – PDF, Word<br>• Working with CSV and JSON Data<br>• Basic plotting with Matplotlib<br>• Scatter plot, Line plot, Bar charts, Box plot, Plotting with Pandas<br>• Working with Data Frames<br>• Aggregate functions using Pandas<br>• Data Processing, cleaning using Pandas<br>• Working with NumPy Library<br>• Dictionaries and Structuring Data<br>• Manipulating Strings |

- ❖ Resume Preparation
- ❖ Naukri or LinkedIn updation Process
- ❖ Important Interview Questions
- ❖ Job Assistance

**3rd Floor, No: 92/5, ACR Greens, Opp. Salarpuria, Outer Ring Road, Beside Biryani Zone, Marathahalli, Bangalore - 560037; info@saagtech.com / www.saagtech.com**
**Phone: +91-94 916 28222 / +91-94 916 29222**