Course Outline

DATA SCIENCE WITH PYTHON (BASIC)

Enhance your skills on Python Programming

Introduction

This course will introduce you to what data science is and what data scientists do. You'll discover the applicability of data science across fields, and learn how data analysis can help you make data driven decisions, as well as build machine learning models and deploy data science solution to a web page.

Training Duration

3 Full Days

Who Should Attend?

- Software Engineer
- Computer Science Engineer
- Data Analyst

No Pre-requisite

- Python Programming

Objectives

- Use principles of statistics and probability to design and execute A/B tests and recommendation engines to assist businesses in making data-automated decisions
- Build Machine Learning Models & make predictions
- Deploy a data science solution to a web app
- Manipulate and analyse distributed datasets using Apache Spark
- Communicate results effectively to stakeholders

Level	Objectives	Duration
Basic	Data Scientist (Basic) - *Basic Python Programming* • Overview of data science workflow, data gathering, role of data scientist • Understand Data Engineering, data base programming • Data wrangling and preparing for Machine Learning • Data Visualisation	3 days

Course Outline

Module 1: Introduction to Data Science

- The Data Science Process
- Communicating to Stakeholders

Module 2: Linear Algebra

- Vectors
- Linear Combination
- Linear Transformation and Matrices

Module 3: Practical Statistics

- Data Types
- Measures of centre (mean, median, mode)
- Standard Deviation, Variance, Outliers
- Probability
- Binomial Distribution
- Conditional Probability
- Bayes Rule
- Normal Distribution theory
- Sampling distributions and the Central Limit
- Theorem
- Confidence Intervals
- Hypothesis Testing
- Type I and type II errors
- P-values
- Null Hypothesis, Alternate Hypothesis

Module 4: Data Engineering

- ETL pipelines
- Extract - CSV, JSON, XML, SQL databases
- Transform - combing, cleaning, encoding, missing data, duplicate data
- Dummy data, Outlier data, scaling data
- Feature Engineering
- Load

Module 5: Database Programming

Module 6: Data Preparation, Data Wrangling

Module 7: Data Visualisation

- Data Visualisation in Data Analysis
- Design of Visualisations
- Univariate Exploration of Data
- Bar Charts
- Pie Charts
- Histograms
- Bivariate Exploration of Data
- Scatterplots and Correlation
- Overplotting, Transparency, and Jitter
- Heat Maps
- Violin Plots
- Box Plots
- Clustered Bar Charts
- Faceting
- Line Plots
- Swarm Plots
- Multivariate Exploration of Data
- Feature Engineering
- Explanatory Visualisations
- A Data Visualisation in Data Analysis - Case study

Testimonials

"This session has exposed new tools for us to use in our day-to-day task. It was really an interesting course. Will recommend this to others."

- Intel employee -

"This sessions has exposed us the method on data modelling and visualisation. The trainer also applied real life application of data in the content which is very useful. Totally recommended."

- Intel employee -

"The instructor was passionate, knowledgable and patient as he was helpful in answering our questions and he explained the examples really well."

- Intel Employee -

"The instructor prepared the explanations of the data with clear visuals which actually helped us understand better and clearer. Great job!"

- Intel employee -

About Elite Indigo

Elite Indigo Consulting provides corporate training to the semiconductor and manufacturing industries. With a humble beginning of one founding member with passion and desire to share his 20 years of experiences in Smart Manufacturing for global manufacturing facilities, now, we have a strong and competent team of 20 members, all aligned with company mission, vision and core values.

Our Mission

"Transform Data into Insights - Leap Forward"

Our Vision

Be a Global Trusted Advisor in the Areas of Skills Development, Consultancy & Software Solutions specialising in Semiconductor & Manufacturing industries.

Our Core Values