ADVANCED COURSE IN
DATA SCIENCE

BEST JOB IN IT INDUSTRY :
DATA SCIENTIST!

AI, ML and Deep Learning Constitute 60% of the Demand

Experience is losing its premium. With traditional IT job roles vanishing,
Techies are turning to short-term Deep-learning courses to stay in the race.

Data Scientist

$110k median base salary

Data Engineer

$106k median base salary
4.3/5.0 level of job satisfaction

Analytics Manager

2000 openings
$112k median base salary
4.1/5.0 level of job satisfaction

WHY ARE THESE JOBS IN DEMAND ?

Influx of data to be captured, cleaned and analyzed
Technology is struggling to keep up
Academic institutions scrambling to prepare students
Candidates in short supply
Limited technological capacity
Overwhelming demand

Top Majors for Data Scientists

Statistics
Math
Quantitative IT
Operations Research
Highly technical social
Science concentrations
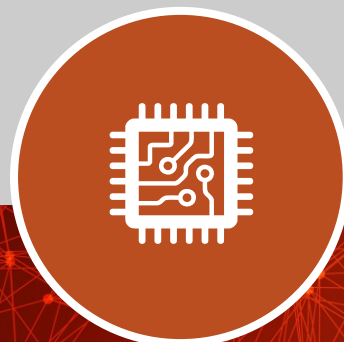
In-demand soft skills include:

In-demand soft skills include:
Intellectual curiosity
Openness to learning new things
Ability to solve problems in
unique ways
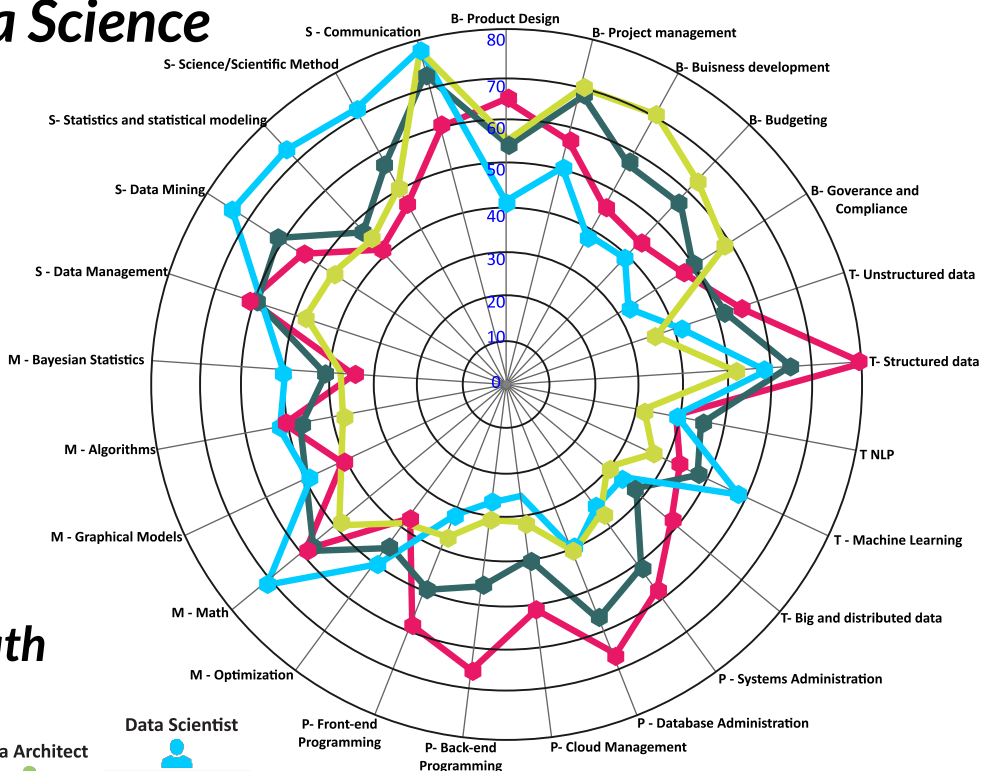Passion for innovation

Must-have skills and expertise:

Must-have skills and expertise:
Classical statistics
Bayesian statistics
Linear algebra for machine
learning applications
In-depth knowledge of R
and Secondarily SAS
Coding in Python, SQL,
NoSQL and Hadoop
Knowledge of the Hadoop platform
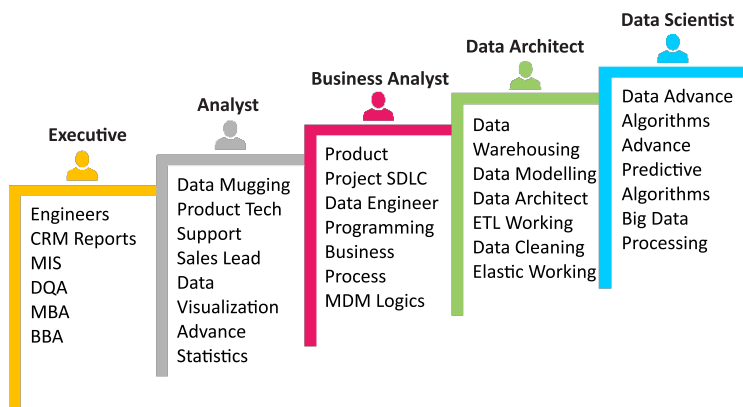Experience with spark,
Hive or Pig

" Data are becoming the new raw material of business.

Proficiency in Data Science Skills by Role

Legend:
- Business Management People
- Developer - Technology / Programming
- Creative – Jack of all Trades
- Research – Scientist, Statistician and Researcher

Radar chart axes (clockwise from top):
- B- Product Design
- B- Project management
- B- Buisness development
- B- Budgeting
- B- Goverance and Compliance
- T- Unstructured data
- T- Structured data
- T NLP
- T - Machine Learning
- T- Big and distributed data
- P - Systems Administration
- P - Database Administration
- P - Cloud Management
- P- Back-end Programming
- P- Front-end Programming
- M - Optimization
- M - Math
- M - Graphical Models
- M - Algorithms
- M - Bayesian Statistics
- S - Data Management
- S- Data Mining
- S- Statistics and statistical modeling
- S- Science/Scientific Method
- S - Communication

Scale: 0, 10, 20, 30, 40, 50, 60, 70, 80

Data Scientist Career Path

Executive
Engineers
CRM Reports
MIS
DQA
MBA
BBA

Analyst
Data Mugging
Product Tech Support
Sales Lead
Data Visualization
Advance Statistics

Business Analyst
Product
Project SDLC
Data Engineer
Programming
Business Process
MDM Logics

Data Architect
Data Warehousing
Data Modelling
Data Architect
ETL Working
Data Cleaning
Elastic Working

Data Scientist
Data Advance Algorithms
Advance Predictive Algorithms
Big Data Processing

Top 10 Use Cases for Data Science & Machine Learning

Healthcare: Patient Diagnosis

Finance: Fraud Detection

Manufacturing: Anomaly Detection

Retail: Inventory Optimization

Government: Smarter Services

Transportation: Demand Forecasting

Network: Intrusion Detection

E-Commerce: Recommender Systems

Media: Interaction & Speed

Education: Research Insight

> Without data you're just another person with an opinion
> ~ W. Edwards Deming

COURSE CURRICULUM
ADVANCED COURSE IN DATA SCIENCE
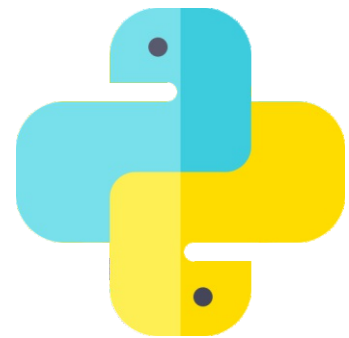
COURSE DURATION 3 MONTHS	SESSION HOURS 120 HRS	CASE STUDIES & PROJECTS

A: DATA SCIENCE ADVANCED TOPICS

Module I: INTRODUCTION

1. Introduction to Jupyter Notebook
2. Getting Started with Data Science
3. Unix Introduction

Module II: PYTHON

4. Python Basics
5. Python Introduction
6. Python Data Structure: Lists and Arrays
7. Python : Conditions and Branching
8. Python : Functions and Methods
9. Python : Objects and Classes
10. Practice Questions in Python
11. Introduction to NumPy
12. Linear Algebra in NumPy
13. Seaborn, Matplotlib
14. **Project 1 : Satellite Image Data Analysis using NumPy**
15. Introduction to Pandas

Module III: PROBABILITY AND MATHEMATICS

16. Introduction to Probability Theory
17. Operation in Sets
18. Permutations and Combinations
19. Probability : Programming in Python
20. Conditional Probability
21. Continuous Probability Distributions (Gaussian, Exponential)
22. Central Limit Theorem (Introduction)
23. Estimating the Sample Mean
24. 2D data Plots and Analysis
25. Plotting Probability Distributions (Histogram, BoxPlot)
26. T Test/Z Test/Anova
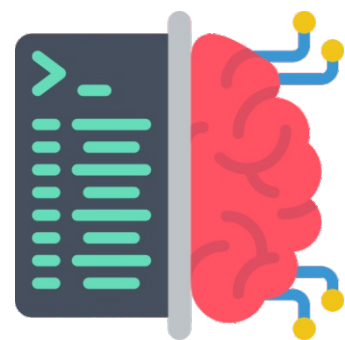
Module IV : DATA SCIENCE AND METHODOLOGY

27. Data Acquisition
28. Data Wrangling
29. Data Statistical Analysis, Grouping and Correlation
30. Model Development
31. Model Evaluation and Refinement
32. Getting started in scikit-learn with the famous iris dataset
33. Training a Machine Learning Model with scikit-learn
34. Comparing Machine Learning Models in scikit-learn
35. Data Science Pipeline: Pandas, Seaborn, and scikit-learn
36. Cross-Validation for Parameter Tuning, Model Selection, and Feature Selection
37. Efficiently Searching for Optimal Tuning Parameters
38. Evaluating a Classification Model : Confusion Matrix and ROC

Module V : DATA VISUALIZATION

39. Basic Plotting for Data Visualisation
40. Data Manipulation for Visualisation
41. 1D Data Analysis: Histograms, Boxplots, and Violin Plots
42. **Project 2 : Visualization of world GDP and carbon dioxide emission**
43. **Project 3 : Using Folium Library for Geographic Overlays**

Module VI : MACHINE LEARNING

44. Simple Linear Regression
45. Multiple Linear Regression
46. Non-Linear Regression
47. Regression Methods
48. Ridge Regression and Lasso Regression
49. Linear Regression and Decision Tree Regression
50. Random Forest Regression
51. Logistic Regression
52. **Project 4 : Sentiment Analysis using Logistic Regression**
53. Decision Tree Classification
54. **Project 5 : Daily Weather Data Analysis using Decision Tree Classification**
55. Random Forest Classification
56. Boosting Algorithms
57. Bagging
58. K- Nearest Neighbours Classification

59. **Project 6 : Nearest Neighbour for Handwritten Digit Recognition**
60. Naive Bayes Classification
61. K-Means Clustering
62. **Project 7 : Minute Weather Data Clustering using K-Means Clustering**
63. Hierarchical Clustering
64. K-Means and Hierarchical Clustering on the same dataset
65. Density-Based Spatial Clustering of Applications with Noise (DB-SCAN)
66. Support Vector Machines & Regression
67. **Project 8 : Sentiment Analysis with Support Vector Machines**
68. Principal Component Analysis (PCA)
69. Applying Principal Component Analysis on Handwritten Digits Dataset
70. Market Basket Analysis

Module VII: NATURAL LANGUAGE PROCESSING

71. Stemming, Phrase identification, word sense disambiguation
72. POS tagging
73. TF and IDF
74. N-gram models of language
75. Word to Vector, Doc to Vector
76. Applying ML Algorithms
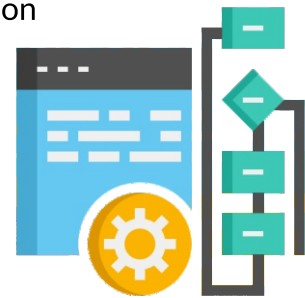
Module VII: ALGORITHM DESIGN AND ANALYSIS

77. Evaluate the speed, runtime and memory dependencies of algorithmic models
78. Parallel computing systems such as SISD (Single Instruction SingleData Stream),
    SIMD (Single Instruction Multiple Data Streams),
    MISD (MultipleInstructions Single Data Stream),
    MIMD (Multiple Instructions Multiple DataStreams)
79. How to use coding tools
80. Create, review and execute unit test cases
81. Corrective and Preventive actions for problems and defects can improve future designs
82. Measure and Optimize performance of algorithm
83. Deployment of the Models
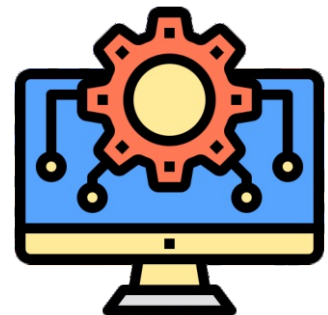
B: ADD-ON MODULES

Module A: SQL: DATABASE QUERY PROCESSING

01. RDBMS Principals
02. Install a DB Engine
03. SQL syntax and Data types
04. Operators, Expressions, Comments
05. Data Definition Language (DDL)
06. Data Manipulation Language (DML)
07. Grant and Revoke
08. SQL Functions (Sum , Count, Avg etc)
09. Joins (self, left, right, full outer)
10. Queries and Sub Queries
11. SQL Clauses
12. SQL Window functions
13. SQL Real time examples
14. SQL live Practice Session

Module B: TABLEAU

01. Tableau Desktop
02. Tableau Products
03. Tableau Terminology
04. Data Connection
05. Working with Data
06. Visualizing Data
07. Statistical Models
08. Dashboards
09. Sharing the Visuals