

ELEVATE YOUR CAREER WITH APPLIED DATA SCIENCE PROGRAMME

Powered by
SEED Infotech & Hitachi Vantara

Gain in demand skills and
become industry ready professional



About SEED Infotech Ltd.

Incorporated in 1994, SEED Infotech has relentlessly endeavored to provide end to end solutions to the Information Technology Industry. With our expertise developed through profound experience, we provide Training in Information Technology, Staffing, Consulting, Software Development and Testing Services to clients globally. With our headquarters in Pune, the IT hub of India, our activities are spread over many locations and cities in Maharashtra and other states in India.

We can rightfully be proud of a formidable manpower of over 300+ qualified professionals including 150+ highly qualified and certified trainers, as also 30+ training locations in India. We have strong associations with Global Technology Leaders such as PMI, Red Hat, HP, Pearson, Salesforce, Prometric, Gasq, Kryterion, EC-Council & SAP. These are undoubtedly our core strengths. Since inception, we have trained more than 8, 00,000+ students and professionals. Capitalizing on our strengths and expertise, we are looking forward to having an outreach in IT Cities, Metros and Colleges all over India through the Hub and Spoke Model.

About Hitachi Vantara

Hitachi Vantara, a wholly owned subsidiary of Hitachi, Ltd., helps data-driven leaders find and use the value in their data to innovate intelligently and reach outcomes that matter for business and society – what we call a double bottom line. We help enterprises store, enrich, activate and monetize their data to improve their customers' experiences, develop new revenue streams and lower their business costs



COURSE GOAL

There is much written about the shortage of Data Scientist and their value in Data driven economy. This concern had been recognized by SEED Infotech and we decided to join hands with Hitachi Vantara to address this .

The aim of this course is to increase the number of “citizens of data science”, who will understand these imperatives and drive the necessary change from within. It will also be a launch-pad for those that do want to continue towards a career in data science. Organizations must stop the “school science projects”, focus only on those which will deliver business results and effectively align their existing data scientists to these. In parallel, they must build a pipeline of new data scientists to support future innovation projects.

The Applied Data Science course from SEED Infotech is centered on the business-led application of data science. It includes insider learnings from successful customer projects across a diverse range of industries. It provides access to methods and technology which is powering data-driven projects today. It offers a fast-track for students to higher employment market value and can augment university courses in science, technology, engineering, mathematics and business.

Required Knowledge and Skills

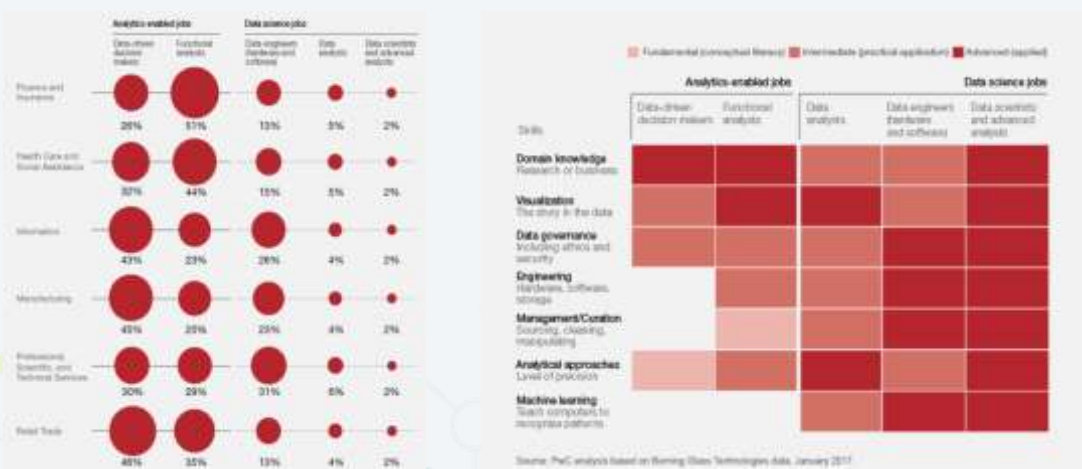
No pre-requisites

Target Audience

Fresher Graduate / Experience Professional aspiring career in Data Science (knowledge of Mathematics and Statistics)

A Growing Need

The 2020 estimate calls for 2.7 million job postings for data science and analytics roles

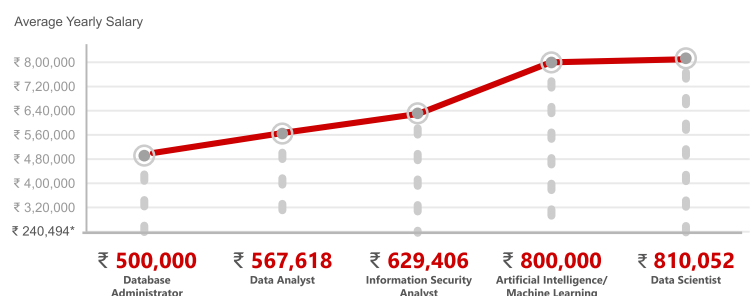


<https://www.pwc.com/us/en/library/data-science-and-analytics.html>

Increasing demand for Data Science professionals

Analysts predict that the country will have more than 11 million job openings by 2026. In fact, since 2019, hiring in the data science industry has increased by 46%. Yet, around 93,000 jobs in Data Science were vacant at the end of August 2020 in India

Future Job Outlook (India)



<https://www.payscale.com/research/IN/Job/Information-Technology>
* Average Indian salary in 2019

Data Science Fundamental

C Programming

- C Language Basics
- Selection Control Statements
- Iterative Statements
- Functions
- Pointers
- Storage Classes
- 1D, 2D Arrays
- Sorting Technique
- String Handling
- Structures and Union

SQL

- Introduction
- Retrieving Data
- Single Row Function
- Summerizing Data
- Joins
- Subqueries
- DML
- DDL
- Database Design
- Mongo Db

Responsive Web Designing

- HTML, HTML5
- CSS, CSS3
- JavaScript
- Bootstrap



Exploring Data Science

Introduction to data science and statistics techniques - ILT or VILT

- 5 Hours

- The process of problem solving
- Descriptive statistics – mean, median, mode, standard deviation, variance, inferential statistics
- Sample, population, regression, hypothesis testing, correlation, co-variance
- Data collection and visualization

Data analysis methods - 1 ILT or VILT - 5 Hours

- Overview of data analysis methods, Performing Exploratory Data Analysis, Feature Exploration, Types of Variables, Univariate Analysis, Multivariate Analysis
- The Normal Distribution, Handling Missing Data Applying QC method, Outlier Treatment Applying multivariate analysis
- Application of Probability Theory and Statistical Hypothesis Testing, Business application

Introduction to Artificial Intelligence and Machine Learning - ILT or VILT - 5 Hours

- Introduction to AI, Evolution & Revolution of AI, Introduction of Applications in various Domains (Scientific including Health Science, Engineering, Financial services and other industries)

Ethical AI – WBT 5 Hours

- Ethical AI and why it matters
- Exercise – research and write article

Industrial IoT Primer – WBT 5 Hours

- Introduction to IoT
- Two worlds – IT and OT
- OT terminology
- Industrial process control
- Spotlight on manufacturing
- Vendor landscape
- Digital Transformation of OT

Introduction to Python (can skip if prior experience on Python) – ILT or VILT - 5 Hours

- Introduction to Python, The Python Interpreter, Working with Command Line/IDLE Python Data Types, Built in operators, Functions and Methods, Block and Indentation
- Conditional Statements, Control Flow Statements, Data Structures
- The Zip Function , Range, Collection framework, Membership and Identity operators, Functions , Default Arguments, Lambda Expressions, Map Filter and Reduce Functions, Scope of Variables
- Object Oriented Programming, Instantiating & Inheritance Classes • Method and Operator Over Loading

Python for Machine Learning –ILT or VILT

- 5 Hours

- Handling the Exceptions, File I/O Handling, Regular Expressions, Serializing Python Objects, Django: Web Development Programming, Flask: Web Development Programming
- Data Science Programming Tools, Pandas Introduction, Constructing Series Objects, Constructing Data Frame Objects, Reading CSV Files into Data Frame Object Reading Excel Files into Data Frame Object, Selecting columns and Slicing, Row and Column Filtering, Finding Unique values in DataFrame, Delete Duplicates in Data Frame, Merging and Concatenating DataFrames, Pivoting, Group by Operations, Working with Dates
- Introduction to Numpy, Comparison on Memory and run-time with native list, Function to Create Numpy Arrays, Attributes of Arrays, Indexing of Arrays, Vectorized operations
- Introduction to Matplotlib, Plotting Line Plot, Plotting with Categorical Data, Plotting Scatter Plots
- Using Seaborn to Visualize Data, Density plots, Histograms, Heat Maps, Violin Plots

Introduction to Machine Learning –ILT or VILT

- 5 Hours

- Overview of machine learning, What is machine learning?, Types of machine learning, Process of machine learning, Examples of applying machine learning, achine learning and data mining, Machine learning and deep learning
- Verifying machine learning with analysis tool, Scikit Learn, H2O Framework
- Sample Selection, Training Data, Testing Data, Validation Data Feature Scaling, Stand-ardization
- Linear Regression, Multiple Linear Regression Logistic Regression, Multi Logistic Re-gression Gradient Boosting Algorithm
- Model Validation, Confusion Matrix, ROC Curve, Cross Validation AUC, R2 Value, Lift, Gain, K-fold Validation • Bootstrapping & Bagging, Over Fitting vs Under-fitting Diagnosis
- SMOTE, Random Over Sampling, Random Under Sampling
- Probabilistic Classifier - Naive Bayes Classifier , Non-probabilistic Classifier - K-nearest neighbor(KNN), Decision Tree, Random Forest Algorithm, Support Vector Machines(SVM)
- Need For Dimensionality Reduction, Principal Component Analysis(PCA) Hierarchical and K-means Clustering
- Case 1: Optimization, Case 2: Anomaly detection, Case 3: Numerical prediction
- Implementing machine learning, Exercise - four lab exercises

WBT - 5 Hours

- Right data, right place, right time
- The changing landscape
- Agile, DataOps and DevOps
- Containerization
- GIGO
- Governance
- Data catalogues

Analytics Data Pipeline - ILT or VILT - 10Hours

- Traditional data pipelines – OLTP, OLAP, ETL
- Big Data challenges and the need for a new way
- Tools, vendor ecosystem
- Modern Data Lakes
- Analytics maturity levels - DEPPA
- IOT and Edge Computing
- Data sources
- Exercise – data pipeline vendor map

Big Data Ecosystem - ILT or VILT - 10 Hours

- Hadoop ecosystem (HDFS, MapReduce, Spark etc)
- NoSQL
- Streaming data
- Public Cloud

Data Modelling and Visualization - ILT or VILT - 5 Hours

- Introduction to data modelling and visualization
- Common visualization techniques
- Exercise – view some visualization examples using Pentaho and Excel

Project example using Lumada – WBT - 5 Hours

- Use case outline
- Business initiative and KPI's
- Features / Deliverables in waves
- Focus on value
- Solution overview
- Key learnings
- Exercise - research an industry domain

Project write-up WBT - 10 Hours

- Builds on activities throughout the course
- Pulls concepts together and applies them to a scenario

Final workshop WBT - 10 Hours

- Sharing of ideas and next steps
- Final wrap-up of the course

*ILT = Instructor Led Training , these sections will be delivered by face to face or

virtually by an instructor remotely.

*WBT = Web Based Training, this can be accessed via internet.

Awards



Awarded as 'Best Instructor-Led Training Center' – Tier II Cities by SAP in 2015



Awarded as **BEST CENTER for CUSTOMER EXPERIENCE** by SAP India.



MACCIA Award for Excellence in Entrepreneurship, 2013 in appreciation of exemplary achievement in IT and ITES



Winner of 'Microsoft Learning Solutions Partner of the Year 2012' Award

Associations



As a R.E.P., SEED Infotech has been approved by Project Management Institute (PMI) to issue Professional Development Units (PDUs) for its training courses. The PMI Registered Education Provider logo is a registered mark of the Project Management Institute, Inc.

Our Branches

- Pune : Chinchwad, Hadapsar, Satara Road, Sinhgad Road, Viman Nagar, Wagholi
- Maharashtra : Mumbai, Ahmednagar, Aurangabad, Jalgaon, Nashik



SEED Infotech Ltd. - Head Office
 'Nalanda', Opp. Gandhi Lawns, Near ICICI Bank,
 Beside Dr. Kalmadi Shamarao High School,
 Gulawani Maharaj Road, Erandawana,
 Pune-411004, Maharashtra (INDIA) Call : 92255 20000
 job@seedinfotech.com | www.seedinfotech.com

