What is Data Science

Data science is the study of data to extract meaningful insights for business.
It has a multidisciplinary approach that combines principles and practices from the fields of mathematics, statistics, artificial intelligence, and computer engineering to analyze large amounts of data.

Tools and SkillSet of Data Scientist

DataScientist require skill sets that are centered on Computer Science, Mathematics, and Statistics.
Data Scientists use several unique techniques to analyze data such as machine learning, trends, linear regressions, and predictive modeling.
The tools Data Scientists use to apply these techniques include

Statistical Analysis System(SAS)	Kinme	Google Analytics	Microsoft HDInsight
Apache Hadoop	RapidMiner	Python	Jupyter
Tableau	Excel	R(RSTUDIO)	Matplotlib
TensorFlow	Apache Flink	DataRobot	MATLAB
BigML	PowerBI	D3.js	QlikView

RoadMap Of Learning Data Science

Python

- Setting Up Environment
- Jupyter Overview
- Basics of Python
- For Data Analysis
  - NumPy
  - Pandas
  - Matplotlib
  - Seaborn
- For Data Visualization
  - Pandas Built-in
  - Plotly and Cufflinks
  - Geographical Plotting
- Basics Of Machine Learning
- Linear Regression
- K Nearest Neighbors
- Decision Trees and Random Forests
- Support Vector Machines
- K Means Clustering
- Principal Component Analysis
- Natural Language Processing
- Neutral Nets and Deep Learning
- Big Data and Spark

Statistics and Probability

- Analyzing Categorical Data
  - Analyzing One Categorical Variable
  - Distributions in Two-Way Tables
- Comparing, Displaying and Summarizing Quantitative Data
  - Display of Qualitative Data with Graphs
  - Comparing and Describing Distributions
  - Measuring Center in Quantitative Data
  - Mean, Median and Interquartile Range
  - Variance and Standard Deviation of Data
  - Box and Whisker Plot
- Modeling Data Distributions
  - Percentiles
  - Z-scores
  - Effects of linear transformations
  - Density curves
  - Normal distributions and the empirical rule
  - Normal distribution calculations
  - More on normal distributions
- Exploring bivariate numerical data
  - Introduction to scatterplots
  - Correlation coefficients
  - Introduction to trend lines
  - Least-squares regression equations
  - Assessing the fit in least-squares regression
  - More on regression
- Study Design
  - Statistical questions
  - Sampling and observational studies
  - Sampling methods
  - Types of studies (experimental vs. observational)
  - Experiments
- Probability
  - Basic theoretical probability
  - Probability using sample spaces
  - Basic set operations
  - Experimental probability
  - Randomness, probability, and simulation
  - Addition rule
  - Multiplication rule for independent events
  - Multiplication rule for dependent events
  - Conditional probability and independence
- Counting, Permutations and Combinations
  - Counting principle and factorial
  - Permutations
  - Combinations
  - Combinatorics and probability
- Random Variables
  - Discrete random variables
  - Continuous random variables
  - Transforming random variables
  - Combining random variables
  - Binomial random variables
  - Binomial mean and standard deviation formulas
  - Geometric random variables
  - More on expected value
  - Poisson distribution
- Sampling Distributions
  - What is a sampling distribution?
  - Sampling distribution of a sample proportion
  - Sampling distribution of a sample mean
- Confidence Intervals
  - Introduction to confidence intervals
  - Estimating a population proportion
  - Estimating a population mean
  - More confidence interval videos
- Hypothesis Testing
  - The idea of significance tests
  - Error probabilities and power
  - Tests about a population proportion
  - Tests about a population mean
  - More significance testing videos
- Interference for Categorical Data
  - Chi-square goodness-of-fit tests
  - Chi-square tests for relationships
- Advanced Regression
  - Inference about slope
  - Nonlinear regression
- Analysis of variance (ANOVA)

Linear Algebra

- Basics of Matrices
- Matrices and Systems of Linear Equation
- Matrix Algebra and Operations
- Determinant and Inverse of Matrix
- Basics of Vector and Vector Spaces
- Subspace Null Space
- Span and Spanning Sets
- Linear Dependence and Independence
- Eigenvalues and Eigenvectors

Machine Learning

- Algorithm 1
  - Introduction to Algorithms and Machine Learning
  - Tools to Analyze Algorithms
  - Algorithmic Technique: Divide and Conquer
  - Divide and Conquer Example: Investing
  - Randomization in Algorithms
  - Application Area Scheduling
- Algorithm 2
  - Graphs
  - Some Ideas Behind Map Searches
  - Dictionaries and Hashing
  - Search Trees
  - Dynamic Programming
- Algorithms 3 and Application to Personal Genomics
  - Linear Programming
  - NP-completeness
  - Introduction to Personal Genomics
  - Massive Raw Data In Genomics
  - Data Science On Personal Genomes
  - Interconnectedness Of Personal Genomes
  - Personal Genomics Case Studies
- Machine Learning
  - Algorithms in Machine Learning
  - What Is Machine Learning
  - Classification
  - Linear Classifiers
  - Ensemble Classifiers
  - Model Selection
  - Cross Validation
- Machine Learning Applications
  - Introduction to Probabilistic Topic Models
  - Probabilistic Modeling
  - Topic Modeling
  - Probabilistic Inference
  - Prediction of Preterm Birth
  - Data Description and Preparation
  - Methods for Prediction of Preterm Birth
  - Relation Between Machine Learning and Statistics

Deep Learning

- Introduction to Deep Learning
  - What is Deep Learning?
  - Use cases of Deep Learning
  - What is Perceptron?
  - Learning Rate
  - Epoch
  - Batch Size
  - Activation Function
  - Single Layer Perceptron
- Getting Started with TensorFlow
  - Introduction and Installing TensorFlow
  - Defining Sequence model layers
  - Layer Types
  - Model Compilation, Optimizer, Loss Function and Training
  - Digit Classification using Simple Neural Network in TensorFlow
  - Adding Hidden Layer and Dropout
  - Using Adam Optimizer
- Convolution Neural Network
  - Image Classification Example
  - What is Convolution and Convolutional Layer Network
  - Filtering
  - ReLU Layer
  - Pooling
  - Data Flattening
  - Fully Connected Layer
  - Saving and Loading a Model
  - Face Detection using OpenCV
- Regional CNN
  - Regional-CNN
  - Selective Search Algorithm
  - Bounding Box Regression
  - SVM in RCNN
  - Pre-trained Model
  - Model Accuracy, Inference Time and Size Comparison
  - Transfer Learning
  - Object Detection – Evaluation
  - mAP
  - IoU
  - Fast RCNN – Speed Bottleneck
  - Feature Pyramid Network (FPN)
  - Regional Proposal Network (RPN)
  - Mask R-CNN
- Boltzmann Machine & Autoencoder
  - What is Boltzmann Machine (BM)?
  - Identify the issues with BM
  - Why did RBM come into picture?
  - Step by step implementation of RBM
  - Distribution of Boltzmann Machine
  - Understanding Autoencoders Architecture of Autoencoders
  - Brief on types of Autoencoders
  - Applications of Autoencoders
- Generative Adversarial Network(GAN)
  - Understanding GAN
  - What is Generative Adversarial Network?
  - Step by step Generative Adversarial Network implementation
  - Types of GAN
  - Recent Advances: GAN
- Introduction RNN and GRU
  - Issues with Feed Forward Network
  - Recurrent Neural Network (RNN)
  - Architecture and Calculation in RNN
  - Backpropagation and Loss calculation
  - Applications of RNN
  - Vanishing andExploding Gradient
  - What is GRU and Components of GRU
  - Update gate and Reset gate
- LSTM and Auto Image Captioning Using CNN LSTM
  - What is LSTM and Structure of LSTM
  - Forget Gate, Input Gate and Output Gate
  - Types of Sequence-Based Model
  - Sequence Prediction, Sequence Classification and Sequence Generation
  - Types of LSTM
  - Stacked LSTM
  - CNN LSTM and Bidirectional LSTM
  - How to increase the efficiency of the model?
  - Backpropagation through time and Workflow of BPTT
  - Auto Image Captioning
  - COCO dataset
  - Pre-trained model
  - Inception V3 model
  - Architecture of Inception V3
  - Modify last layer of pre-trained model and Freeze model
  - CNN for image processing and LSTM or text processing

BigData + PySpark + Tableau

- Setup and Installations
  - Python Installation
  - Installing Apache Spark
  - Installing Java
  - Installing MongoDB
  - Installing NoSQL
- Data Processing with PySpark and MongoDB
  - Integrating PySpark with Jupyter Notebook
  - Data Extraction
  - Data Transformation
  - Loading Data into MongoDB
- Machine Learning with PySpark and MLib
  - Data Pre-processing
  - Building the Predictive Model
  - Creating the Prediction Dataset
- Creating Data Pipeline Scripts
  - Installing Visual Studio Code
  - Creating the PySpark ETL Script
  - Creating the Machine Learning Script
- Tableau Data Visualization
  - Installing Tableau
  - Installing MongoDB ODBC Drivers
  - Creating a System DSN for MongoDB
  - Loading the Data Sources
  - Creating a Geo Map
  - Creating a Bar Chart
  - Creating a Magnitude Chart
  - Creating a Table Plot
  - Creating a Dashboard

Courses to Learn Data Science From

Coursera Free Course :- https://in.coursera.org/courses?query=free%20courses%20data%20science

- Machine Learning :- https://in.coursera.org/learn/uol-machine-learning-for-all
- Python and Statistics :- https://in.coursera.org/learn/python-statistics-financial-analysis
- Math Skills for Data Science :- https://in.coursera.org/learn/datasciencemathskills
- Data Processing :- https://in.coursera.org/learn/python-data-processing
- K-Means Clustering for Python :- https://in.coursera.org/learn/data-science-k-means-clustering-python

DataCamp Free Course :- https://www.datacamp.com/courses-all

- Free Courses of All Sub Points of Data Science

IntelliPaat IIT Madras Data Science Course :-
https://intellipaat.com/advanced-certification-data-science-artificial-intelligence-iit-madras/?utm_source=google&utm_medium=search&utm_term=data%20science%20course&utm_campaign=s_datascience_in_state&utm_source=google&utm_medium=cpc&campaignid=11476410440&adgroupid=111616916629&utm_term=data%20science%20course&gclid=CjOKCQiAn4SeBhCwARIsANeF9DIPmC0tWAdbhMGvT35rD8S8ikZBJmHzkKR9jyMktuiLblbJ_UnzHFgaAoXuEALw_wcB

SimpliLearn IITK Data Science Course :-
https://www.simplilearn.com/iitk-professional-certificate-course-data-science?utm_source=google&utm_medium=cpc&utm_term=data%20scientist%20training&utm_content=19485826713-143598343423-643998210188&utm_device=c&utm_campaign=Search-DataCluster-PG-DSBI-CDS-IITK-Main-IN-AllDevice-adgroup-STAG-IITK-DS-Theme-Training&gclid=CjOKCQiAn4SeBhCwARIsANeF9DIaxJhIQljqFqEeVHxfCrK4vxEaOSdR_GmZZuiV9MSPr0uYwKC1-mYaApW8EALw_wcB

FreeCodeCamp Course :- https://www.classcentral.com/provider/freecodecamp?subject=data-science

- Free Courses of All Sub Points of Data Science

Free Udemy Data Science Courses :- https://www.udemy.com/topic/data-science/free/

- Free Courses of All Sub Points of Data Science

Free Udacity Data Science Courses :- https://www.udacity.com/courses/all?field=school-of-data-science price=Free

Projects

Beginner Level : -

- Fake News Detection using R Language
- Detecting Forest Fire
- Detection of Road Lane Lines
- Sentiment Analysis Backed by R Dataset
- Influences of Climatic Pattern on the food chain supply globally

Intermediate Level

- Recognition of Speech with Libraries
- Prediction of Age & Gender through Deep Learning
- Creating ChatBot with Python
- Implementing Driver Fatigue Detection System

Advanced Level

- Detecting Credit Cards with Python
- Project on the recognition of traffic signals
- Movie Recommendation Platform with R Packages
- Segmentation of Customers Group with ML