Learnbay

DATA SCIENCE PROGRAM

CO-DEVELOPED WITH IBM

Duration

250+ hours of Live Interactive Training from Industry Experts

Job Support

Guaranteed Job Referral for Working Professionals with Resume Prep and Mock Interview

Projects

12+ real-time projects and Capstone projects.

Specially crafted for working professionals.

Live Online Interactive Mode

+91 77 956 87 988

Program Highlights

Live Sessions By Expert

- Live Online Interactive Training
- Get trained from FAANG / Industrial Experts
- 250+ hours of training
- 7.5 Months Weekday
- 9 Moths Weekend

Project Based Learning

- 12+ Real Time Projects
- 2 Capstone Projects
- Mentorship & Guidance by Expert

3 Year Flexible Subscription

- Flexibility to attend multiple batches from different trainers.
- Life time access to Recordings.
- Access to change batch between weekday and weekend session

Special Support to Non Programmers

- Learn Python from scratch
- Special classes for Non programming background students
- Real time Use Cases.

Global Certification in Data Science

- Certified Data science program.
- Industry Accredited Global Certification Course.
- In Collaboration with IBM.

Job Assistance Program for Working Professionals

- Resume support from expert
- Interview prep session and Mock interview
- **Guaranteed job referrals** for working professionals

Top Rated Training Institute in India For Data Science And AI Certification

4.8 ★★★★½

300+ user Review

Quora

Top Rated

4.9 ★★★★½

PROGRAM DETAILS

Eligibility

Work Experience :
Working Professionals With **1+ Years of experience in any domain** (tech or non technical)
Academics :
BE/B.Tech (from any branch) , BBA/MBA, MCA/M.Tech, B.Com, Graduation in Mathematics, Statistics, IT

Who should apply

- Software developers/ Programmers, Project Managers, Manual And Automation Test Engineer, Java and .NET Developer, Business Analyst.
- SAP domain expert, Python , Embedded developer , Android/ IoS developer.
- Professionals planning for Masters or higher education in Data Science.

Pre - requisite

There is no Prerequisite for this course as we cover programming and statistics from basics. We provide special classes & support for professionals from non-programming/ non-technical background.

Why enroll for this program?

Learnbay offers **instructor-led interactive** program with live doubt solving session as learning from recorded videos can be boring.

Get 1:1 doubt clearing session with expert after your live class. Flexibility to batches, get back up classes and attain session from multiple instructors.

As per the industrial requirement 2 or 3 modules is not sufficient, hence **we offer a Full Stack program** specially crafted for working professionals.

We over our program with complete **flexibility to attain Live Session/Classroom Session for 3 years** and Life - Time Access of LMS. With access to change batches, instructors, etc.

Program Modules & Tools

TERM 1 & 2

Python

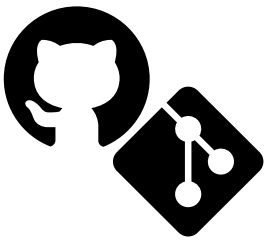
Statistics

Machine Learning

Deep Learning (Tensorflow)

Time Series Analysis & Forecasting

Natural Language Processing



Git & GitHub

R Programming

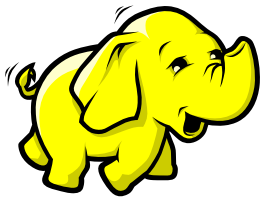
TERM 3 & 4

SQL for Data Science

Tableau

Power BI

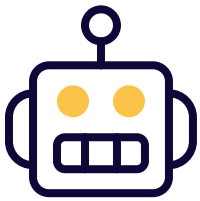
Mongo DB



Hadoop

Apache Spark

Google Cloud



Advance AI

Global Certification & Fees

Program Fee

Rs. 65,000 +18% GST

Weekday Batches : 7.5 Months	Weekend Batches : 9 Months
Monday - Friday	Saturday & Sunday
2 hours everyday	3.5 hours everyday

Payment modes

INTEREST FREE INSTANT LOAN
WITHOUT CREDIT CARD

Aadhaar Card, Pan Card & 3 months salary slip required

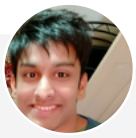
NO COST EMI UPTO 9 MONTHS ON MAJOR CREDIT CARDS

ICICI, HDFC, RBL, Standard Chartered, Axis bank,Kotak credit cards

UPI, Net Banking, Bank Transfer, No Cost EMI (Credit Cards), Interest Free Loan

GENERATE DISCOUNT COUPON

Success Stories



Shezan Baig

Working at Accenture AI

Learnbay is one of the most remarkable data science institutes I've come across. In comparison to other institutes in India, it offers a data science course at a low cost. Excellent value for money. I would strongly advise everyone to attend this institute. All of the trainers are excellent in their own subject, but the Stats & ML trainer in particular is outstanding.

View LinkedIn profile

Viraj Ghodke

Working at Affine Analytics

For me, learning using the Learnbay platform has been a great experience. The teaching and management team are very helpful. They are constantly willing to clarify each individual's doubts and meet the needs of working professionals. I owe my gratitude to the trainers and the whole Learnbay team for assisting me in getting placed.

View LinkedIn profile

Pooja Sharma

Working at Learnvista

I have done Data Science and AI certification and got placed within 8 months. Journey was really tough for me because i was from mechanical domain. But the mentors were really helpful and they have good industrial knowledge. Facility of recording classes is very useful.

View LinkedIn profile

Shubhangi J. Waghmare

Working at Infrasoft Technologies

The offering here is best in the industry I would say both cost and curriculum wise. One advantage joining here is you can access their resources for lifetime unlike others where you have accessibility only for a year or so. Most importantly, there is continuous assistance for recruitment. Well,one enrolls for any course and ends up getting a handsomely paying job.

View LinkedIn profile

Pawan Yadav

Working at Oracle

I have done Data Science certification and i placed in Oracle. Journey was really tough for me because i was from core electronics domain. Mentors are really helpful and they have good knowledge. Personally i liked teaching style of Trainer Nishant. Facility of recording classes is very useful.

View LinkedIn profile

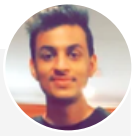
Suman Karmakar

Working at IBM

It was a good and effective course with dedicated faculties for modules.You get flexibility to attend classes from multiple instructors.Very Supportive environment for learning.

View LinkedIn profile

Success Stories



Neelesh Dugar

Working at Act21 Softwares

Very well designed and structured. I really appreciate him and would want to put some light on Utkarsh Kulshrestha. Cheers to you guys! I had an amazing experience at Learnbay, which got me where I am today. Thank you to each one of you and also Abhishek who is handling very well. All the best guys!!

Deevraj

Working at Mindtree

The quality of content is very nice mainly the instructor concentrating on the practical part, live project sessions make you feel confident to attend interviews. Multiple batch options, access for any instructor class videos or materials. Totally positive environment around. One can join here with no second thought.

Srikanth Saurav

Working at Mediamarksaturn

Machine Learning concepts & Statistics are very well explained by Utkarsh. Best thing was completing the syllabus on-time as they have promised. Trainers are clearing the doubts . Got multiple joining offers from different MNCs for Data Science and AI developer

View LinkedIn profile

Read more reviews

Watch Transition Videos

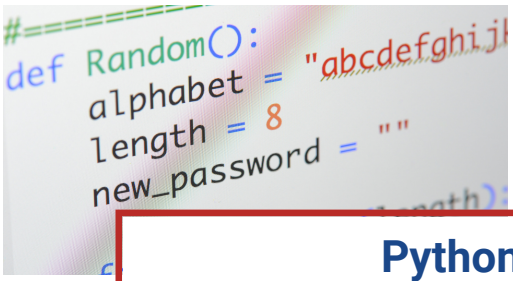
STILL CONFUSED?

Apply for FREE Career Counselling Session with our Expert

+91 77 956 87 988

BOOK NOW

Demo Recordings



Python

Statistics

Machine Learning

Deep Learning

NLP

R Programming

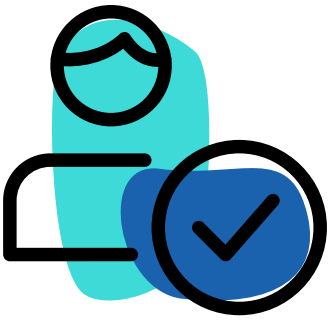
Real Time Projects

SUBSCRIBE

SUBSCRIBE US TO WATCH MORE DATA SCIENCE AND AI VIDEOS

Application Process

Talk to Our Admission Executive



Qualification: BE, B. Tech, ME, M.Tech. BCA, MCA (Any Branch), MBA, Etc. All technical or managerial degree. Professionals having 1+ years of experience in any domain. To know more about Eligibility Whatsapp Us

Whatsapp Now

Talk to our admission executive & get your profile reviewed

Attend Personalised Career Counselling and profile review session with expert. This session will help you to understand whether your profile is suitable for Data Science and AI certification course.

Apply For Profile Review

Pay and Enroll for this program

Contact our Admission Officer for discount coupon. Apply the discount coupon and enrol for IBM certified Program.

Payment Mode: Debit Card, Credit Card, UPI, Bank Transfer, Interest Free Loan, No Cost EMI (Credit Card)

Pay and Enroll for the program

Program Syllabus

MODULE 0

Basic programming fundamentals for Non-programming background aspirants

4 days (8 hours)

TERM 1

Core Python + Numpy + Pandas + Matplotlib + Seaborn

(40 hours) : : 1 Month Weekday :: 1.25 Months Weekend

TERM 2

Statistics + Machine Learning + Capstone Project

(70 hours) : : 1.75 Months Weekday :: 2.25 Months Weekend

TERM 3

SQL + MongoDB + Tableau + PowerBI + Hadoop basics & Apache Spark + R Programming

(70 hours) : : 1.75 Months Weekday :: 2.25 Months Weekend

TERM 4

Deep Learning using Tensor-flow + Natural Language Processing & Text Analytics + Cloud Deployment of ML Model using GCP + Capstone Project

(48 hours) : : 1.25 Months Weekday :: 1.5 Months Weekend

Important Note :

After successful completion of Term 1, 2 and 3, candidates become eligible for Job Assistance Program (2- 3 weeks) which includes :

- Resume Session and Assistance.
- Interview Prep Session & Mock Interview with Expert.
- Participating in Live Kaggle Competitions.
- List of Important Interview Questions from each modules.
- **Guaranteed Job Referrals for Data Science/ ML Engineer roles.**
- You can start attending interviews after Term 3 and keep learning other modules from Term 4 simultaneously.

Module 0 Programming Fundamentals

Chapter 1: Introduction to Programming (3 hrs)

What is a programming language ?
Source code Vs bytecode Vs machine code
Compiler Vs Interpreter
C/C++, Java Vs Python

Chapter 2: Jupyter notebook basics (1 hrs)

Different type of code editors in python. Introduction to Anaconda and jupyter notebook.
Flavours of python.

Chapter 3: Python Programming Basics (2 hrs)

Variable Vs identifiers Vs strings
Operators Vs operand Procedure oriented Vs modular programming

Chapter 4: Statistics basics (2 hrs)

Introduction to statistics Mean, median, mode, Standard deviation, Average. Introduction to probability, permutations and combinations. Introduction to linear Algebra

Chapter 5: Git and GitHub (2 hrs)

Learn the key concepts of the Git source control system
Step through the entire basic Git workflow
Configure SSH for authentication
Create and use a remote repository on GitHub
Git Overview
Set up & configuration
Working with git locally

[NOTE]

This module 0 is for those who are from non-technical background like Mechanical, BBA, MBA, B.Com, M.Com, etc.
Or for those who work in Non-IT sectors to get in-depth knowledge of programming and how to use it in Data Science.

Module 1 Python : 40 hours

1. Programming Basics & Environment Setup

Installing Anaconda, Anaconda Basics and Introduction
Get familiar with *version control, Git and GitHub.*
Basic Github Commands.
Introduction to Jupyter Notebook environment. Basics Jupyter notebook Commands.
Programming language basics.

2. Python Programming Overview

Python Overview
Python 2.7 vs Python 3
Writing your First Python Program
Lines and Indentation, Python Identifiers
Various Operators and Operators Precedence
Getting input from User, Comments, Multi line Comments.

3. Strings, Decisions And Loop Control

Working With Numbers, Booleans and Strings,String types and formatting, String operations
Simple if Statement, if-else Statement if-elif Statement.
Introduction to while Loops.
Introduction to for Loops,Using continue and break.

Class hands-on :

6 programs/coding exercise on string, loop and conditions in classroom

4. Python Data Types

List,Tuples,Dictionaries
Python Lists,Tuples,Dictionaries
Accessing Values,Basic Operations
Indexing, Slicing, and Matrixes
Built-in Functions & Methods
Exercises on List,Tuples And Dictionary

Class hands-on :

- *Program to convert tuple to dictionary*
- *Remove Duplicate from Lists*
- *Python program to reverse a tuple*
- *Program to add all elements in list.*
- *+ 3 more programs to be covered in class*

5. Functions And Modules

Introduction To Functions – Why
Defining Functions
Calling Functions
Functions With Multiple Arguments.
Anonymous Functions - Lambda
Using Built-In Modules,User-Defined Modules,Module Namespaces, Iterators And Generators

Class hands-on :

8+ Programs to be covered in class from functions, Lambda, modules, Generators and Packages.

6. File I/O And Exceptional Handling and Regular Expression

Opening and Closing Files
open Function,file Object Attributes
close() Method ,Read,write,seek.
Exception Handling, try-finally Clause
Raising an Exceptions,User-Defined Exceptions
Regular Expression- Search and Replace
Regular Expression Modifiers
Regular Expression Patterns,re module

Class hands-on :

10+ Programs to be covered in class from File IO,Reg-ex and exception handling.

Module 1 Python : 40 hours

7. Data Analysis Using Numpy And Pandas

Introduction to **Numpy**. Array Creation, Printing Arrays, Basic Operation - Indexing, Slicing and Iterating, Shape Manipulation - Changing shape, stacking and splitting of array
Vector stacking, Broadcasting with Numpy, *Numpy for Statistical Operation.*

Pandas : Introduction to Pandas
Importing data into Python
Pandas Data Frames, Indexing Data Frames ,Basic Operations With Data frame, Renaming Columns, Subletting and filtering a data frame.

8. Data Visualization using Python: Matplotlib and Seaborn

Matplotlib: Introduction, plot(),Controlling Line Properties, Subplot with Functional Method, Multiple Plot, Working with Multiple Figures, Histograms

Seaborn :
Intro to Seaborn And Visualizing statistical relationships , Import and Prepare data. Plotting with categorical data and Visualizing linear relationships.
Seaborn Exercise

3 Case Study on Numpy, Pandas , Matplotlib

1 Case Study on Pandas And Seaborn

Assessment Test in Python :

2 hour of Assesment Test in Python
(Coding & Objective Questions)

REAL TIME USE CASES IN PYTHON TO BE COVERED IN CLASS

WITH 5 ASSIGNMENTS

Module 2 Statistics : 24 hours

1. Fundamentals of Math and Probability

Basic understanding of linear algebra, Matrics, vectors
Addition and Multimplication of matrics
Fundamentals of Probability
Probability distributed function and cumulative distributed function.

Class Hand-on

Problem solving using R for vector manupulation
Problem solving for probability assignments

2. Descriptive Statistics

Creating Headers, Footers, and Page Numbers
Adjusting Page Margins and Orientation
Adding Print Titles and Gridlines, rows to repeat at top of each page
Formatting Fonts & Values
Adjusting Row Height and Column Width
Changing Cell Alignment
Adding Borders
Applying Colors and Patterns
Using the Format Painter
Merging Cells, Rotating Text
Using Auto Fill

3. Inferential Statistics

What is inferential statistics
Different types of Sampling techniques
Central Limit Theorem
Point estimate and Interval estimate
Creating confidence interval for population parameter
Characteristics of Z-distribution and T-Distribution
Basics of Hypothesis Testing
Type of test and rejection region
Type of errors in Hypothesis resting, Type-I error and Type-II errors
P-Value and Z-Score Method
T-Test, Analysis of variance(ANOVA) and Analysis of Co variance(ANCOVA)
Regression analysis in ANOVA

Class Hands-on:

Problem solving for C.L.T
Problem solving Hypothesis Testing
Problem solving for T-test, Z-score test
Case study and model run for ANOVA, ANCOVA

4. Hypothesis Testing

Hypothesis Testing
Basics of Hypothesis Testing. Type of test and Rejection Region. Type o errors-Type 1 Errors, Type 2 Errors. P value method, Z score Method. **The Chi-Square** Test of Independence. Regression. Factorial Analysis of Variance. Pearson Correlation Coefficients in Depth. Statistical Significance, Effect Size, and Confidence Intervals

5. Data Processing & Exploratory Data Analysis

Introduction to Data Cleaning & Data Pre-processing. What is Data Wrangling?
How to Restructure the data?
What is Data Integration, Data Transformation
EDA : Finding and Dealing with Missing Values. What are Outliers? Using Z-scores to Find *Outliers*. Introduction to Bivariate Analysis, Scatter Plots and Heatmaps. Introduction to Multivariate Analysis

Module 3 Machine Learning : 46 hours

Introduction To Machine Learning

What is Machine Learning?
Introduction to Supervised and Unsupervised Learning
Introduction to SKLEARN (Classification, Regression, Clustering, Dimensionality reduction, Model selection, Preprocessing)
What is Reinforcement Learning?
Machine Learning applications
Difference between Machine Learning and Deep Learning

1. Supervised Learning

Support Vector Machines
Linear regression
Logistic regression
Naive Bayes
Linear discriminant analysis
Decision tree
k-nearest neighbor algorithm
Neural Networks (Multilayer perceptron)
Similarity learning

2. Linear Regression

Introduction to Linear Regression
Linear Regression with Multiple Variables
Disadvantage of Linear Models
Interpretation of Model Outputs
Understanding Covariance and Colinearity
Understanding Heteroscedasticity

Case Study – Application of Linear Regression for Housing Price Prediction

3. Logistic Regression

Introduction to Logistic Regression.– Why Logistic Regression .
Introduce the notion of classification
Cost function for logistic regression
Application of logistic regression to multi-class classification.
Confusion Matrix, Odd's Ratio And ROC Curve
Advantages And Disadvantages of Logistic Regression.
Case Study: To classify an *email as spam or not spam using logistic Regression.*

4. Decision Trees

Decision Tree – data set
How to build decision tree?
Understanding Kart Model
Classification Rules- Overfitting Problem
Stopping Criteria And Pruning. How to Find final size of Trees. Model A decision Tree. Naive Bayes. Random Forests and Support Vector Machines. Interpretation of Model Outputs

Case Study:
1 Business Case Study for Kart Model
2 Business Case Study for Random Forest
3 Business Case Study for SVM

Module 3 Machine Learning : 46 hours

5. Unsupervised Learning
Hierarchical Clustering
k-Means algorithm for clustering – groupings of unlabeled data points.
Principal Component Analysis(PCA)-Data
Independent components analysis(ICA)
Anomaly Detection
Recommender System-collaborative filtering algorithm
Case Study– Recommendation Engine for e-commerce/retail chain

6. Natural language Processing
Introduction to natural Language Processing(NLP).
Word Frequency Algorithms for NLP
Sentiment Analysis
Case Study :
Twitter data analysis using NLP

7. Introduction to Time Series Forecasting
Basics of Time Series Analysis and Forecasting, Method Selection in Forecasting
Moving Average (MA) Forecast Example, Different Components of Time Series Data, Log Based Differencing, Linear Regression for Detrending

8. ARIMA and Multivariate Time Series Analysis
Introduction to ARIMA Models, ARIMA Model Calculations, Manual ARIMA Parameter Selection, ARIMA with Explanatory Variables
Understanding Multivariate Time Series and Their Structure, Checking for Stationarity and Differencing the MTS

Case Study : Performing Time Series Analysis on Stock Prices

Important Note :
All Machine Learning Algorithms are covered in depth with Real time case studies for each Algorithm
Once *60% of ML is completed,* **Capstone Project will be released for the batch.**

Assignments:
Statistics Assignments : Total 4 practice set and Assignments from Statistics
Machine Learning Assignments : Total 3 Practice Set And 2 Real time use case as Assignments

Assessment Test For Term2:
Duration : 3 hours
Question Type : Objective & ML Case Studies

Module 4 SQL & MongoDB : 24 hours

1. RDBMS And SQL Operations :

Introduction To RDBMS
Single Table Queries - SELECT, WHERE, ORDER BY, Distinct, And, OR

Multiple Table Queries: INNER, SELF, CROSS, and OUTER, Join, Left Join, Right Join, Full Join, Union
Advance SQL Operations:
Data Aggregations and summarizing the data
Ranking Functions: Top-N Analysis
Advanced SQL Queries for Analytics

2. NoSQL Databases :

Topics - What is HBase?
HBase Architecture, HBase Components,
Storage Model of HBase,
HBase vs RDBMS
Introduction to Mongo DB, CRUD
Advantages of MongoDB over RDBMS
Use cases

3. Programming with SQL :

Mathematical Functions
Variables
Conditional Logic
Loops
Custom Functions
Grouping and Ordering
Partitioning
Filtering Data
Subqueries

4. MongoDB Overview :

Where MongoDB is used?
MongoDB Structures
MongoDB Shell vs MongoDB Server
Data Formats in MongoDB
MongoDB Aggregation Framework
Aggregating Documents
What are MongoDB Drivers?

5. Basics and CRUD Operation :

Databases, Collection & Documents
Shell & MongoDB drivers
What is JSON Data
Create, Read, Update, Delete
Finding, Deleting, Updating, Inserting Elements
Working with Arrays
Understanding Schemas and Relations

6. Introduction to MongoDB :

What is MongoDB?
Characteristics and Features
MongoDB Ecosystem
Installation process
Connecting to MongoDB database
Introduction to NoSQL
Introduction of MongoDB module
What are ObjectIds in MongoDb

Module 5 Tableau & PowerBI : 22 hours

1. Introduction to Tableau :
Connecting to data source
Creating dashboard pages
How to create calculated columns
Different charts
Hands-on :
Hands on on connecting data source and data cleansing
Hands on various charts

2. Visual Analytics :
Getting Started With Visual Analytics
Sorting and grouping
Working with sets, set action
Filters: Ways to filter, Interactive Filters
Forecasting and Clustering
Hands-on :
Hands on deployment of Predictive model in visualization

3. Dashboard and Stories :
Working in Views with Dashboards and Stories
Working with Sheets
Fitting Sheets
Legends and Quick Filters
Tiled and Floating Layout
Floating Objects

4. Mapping :
Coordinate points
Plotting Latitude and Longitude
Custom Geocoding
Polygon Maps
WMS and Background Image

5. Getting Started With Power BI :
Installing *Power BI Desktop* and Connecting to Data
Overview of the Workflow in Power BI Desktop
Introducing the Different Views of the Data Mode
Query Editor Interface
Working on Data Model

6. Programming with Power BI :
Working with Timeseries
Understanding aggregation and granularity
Filters and *Slicers in Power BI*
Maps, Scatterplots and BI Reports
Connecting Dataset with Power BI
Creating a Customer Segmentation Dashboard
Analyzing the Customer Segmentation Dashboard

Module 6 Big Data & Spark Analytics : 14 hours

1. Introduction To Hadoop :
Distributed Architecture - A Brief Overview
Understanding Big Data
Introduction To Hadoop, Hadoop Architecture
HDFS, Overview of MapReduce Framework
Hadoop Master – Slave Architecture
MapReduce Architecture
Use cases of MapReduce

2. Apache Spark Analytics :
What is Spark
Introduction to Spark RDD
Introduction to Spark SQL and Dataframes
Using R-Spark for machine learning
Hands-on:
installation and configuration of Spark

Using R-Spark for machine learning programming

3. Apache Spark Analytics :
Getting to know PySpark
Pyspark Introduction
Pyspark Environment Setup
pySpark - Spark context
RDD , Broadcast and Accumulator
Sparkconf and Sparkfiles
Spark MLlib Overview Algorithms and utilities in Spark Mlib

Hands-on:
Map reduce Use Case 1 : Youtube data analysis
Map reduce Use Case 2: Uber Data Analytics

Hands-on:
Spark RDD programming

Hands-on:
Spark SQL and Dataframe programming

Module 7 R Programming : 10 hours

1. Introduction To R :

Installation Setup
Quick guide to RStudio User Interface
RStudio's GUI3
Changing the appearance in RStudio
Installing packages in R and using the library
Development Environment Overview
Introduction to R basics
Building blocks of R
Core programming principles
Fundamentals of R

2. Programming with R :

Creating an object
Data types in R
Coercion rules in R
Functions and arguments
Matrices
Data Frame
Data Inputs and Outputs with R
Vectors and Vector operation
Advanced Visualization
Using the script vs. using the console

3. Manipulating Data :

Data transformation with R - the Dplyr package - Part
Data transformation with R - the Dplyr package - Part
Sampling data with the Dplyr package
Using the pipe operator in R
Tidying data in R - gather() and separate()
Tidying data in R - unite() and spread()

4. Visualizing Data :

Intro to data visualization
Introduction to ggplot2
Building a histogram with ggplot2
Building a bar chart with ggplot2
Building a box and whiskers plot with ggplot2
Building a scatterplot with ggplot2

Module 8 Tensor Flow & Deep Learning : 20 hours

1. Introduction to Deep Learning And Tensor Flow

Neural Network
Understanding Neural Network Model
Installing TensorFlow
Simple Computation, Constants And Variables
Types of file formats in TensorFlow
Creatting A Graph – Graph Visualization
Creating a Model – Logistic Regression
Model Building using tensor flow
TensorFlow Classification Examples

2. Introduction to Tensor Flow

Installing TensorFlow
Simple Computation ,Contants And Variables
Types of file formats in TensorFlow
Creatting A Graph - Graph Visualization
Creating a Model - Logistic Regression
Model Building
TensorFlow Classification Examples

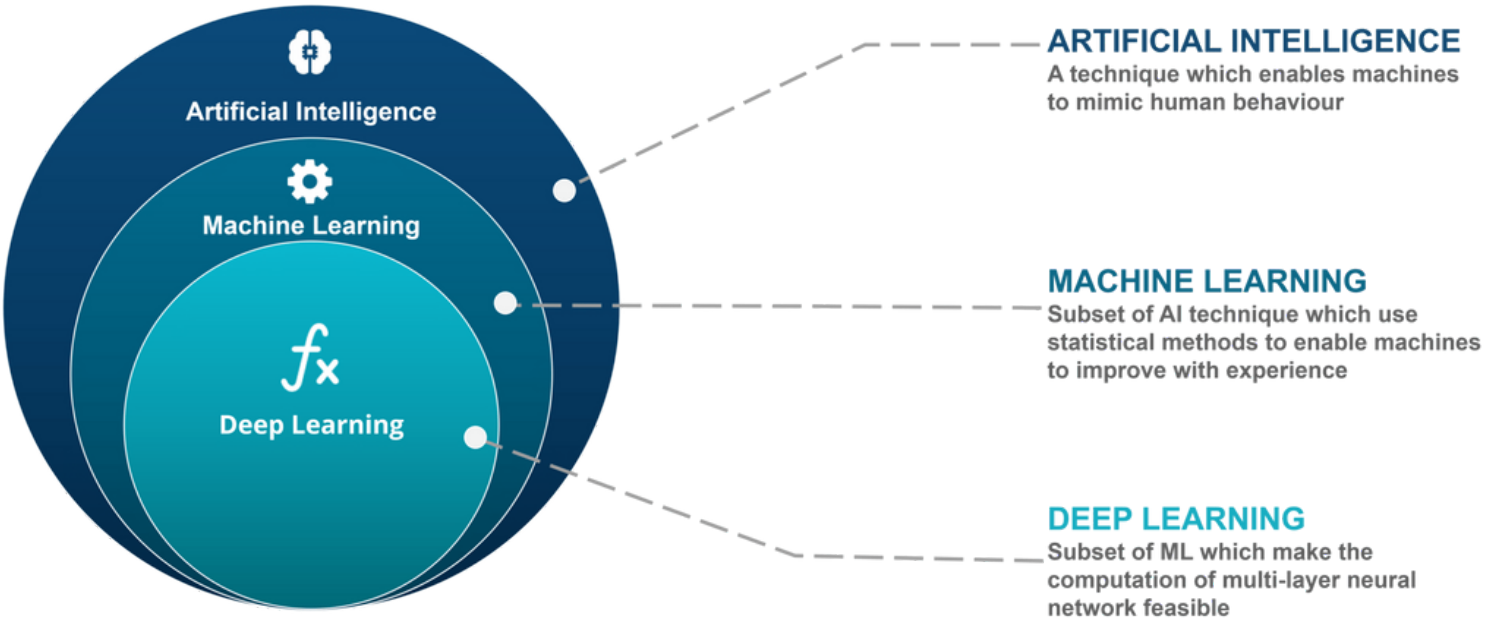
3.Understanding Neural Networks With Tensor Flow

Basic Neural Network
Single Hidden Layer Model
Multiple Hidden Layer Model
Backpropagation – Learning Algorithm and visual representation
Understand Backpropagation – Using Neural Network Example
TensorBoard
Project on backpropagation

4. Convolutional Neural Network (CNN)

Convolutional Layer Motivation
Convolutional Layer Application
Architecture of a CNN
Pooling Layer Application
Deep CNN
Understanding and Visualizing a CNN

Project : Building a CNN for Image Classification



Module 9 Natural Language Processing : 28 hours

1. Introduction to NLP & Text Analytics

Introduction to Text Analytics
Introduction to NLP
What is Natural Language Processing?
What Can Developers Use NLP Algorithms For?
NLP Libraries
Need of Textual Analytics
Applications of Natural Language Procession
Word Frequency Algorithms for NLP
Sentiment Analysis

2. Text Pre Processing Techniques

Need of Pre-Processing
Various methods to Process the Text data
Tokenization ,Challenges in Tokenization
Stopping ,Stop Word Removal
Stemming - Errors in Stemming
Types of Stemming Algorithms - Table lookup Approach ,N-Gram Stemmers

3. Distance Algorithms used in Text Analytics

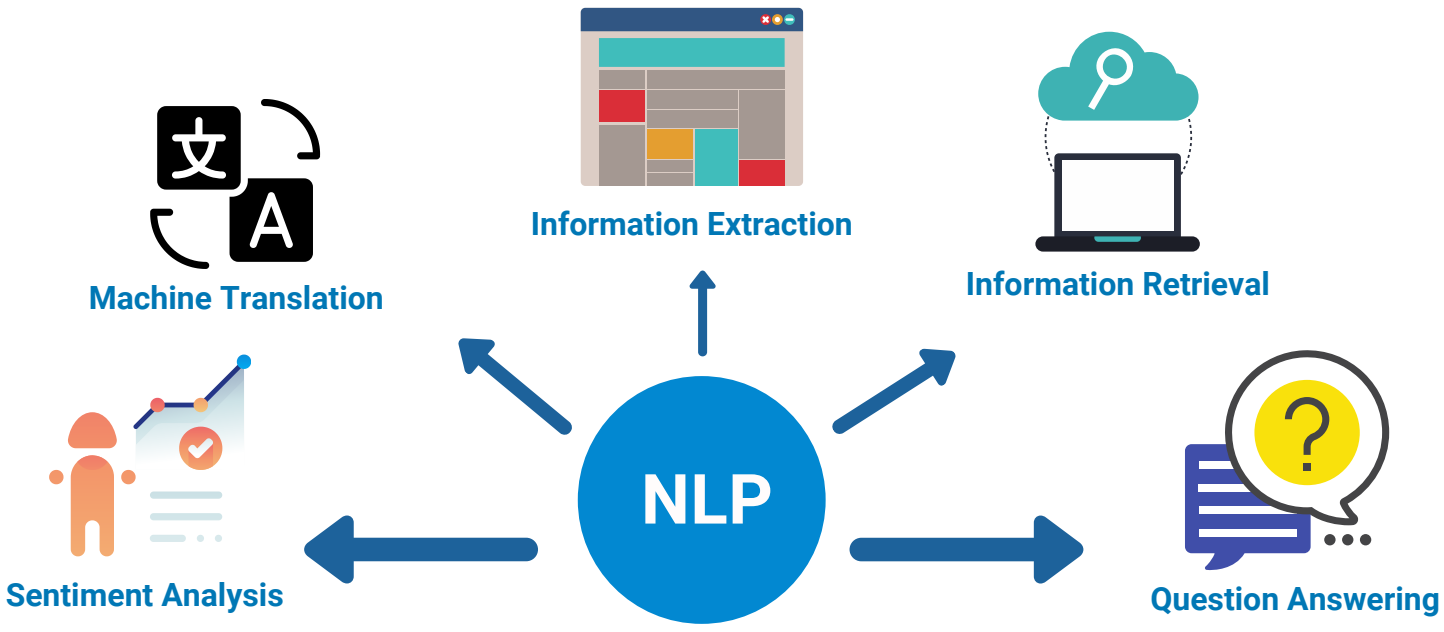
String Similarity
Cosine Similarity Mechanishm - Similarity between Two text documents
Levenshtein distance - measuring the difference between two sequences.
Applications of Levenshtein distance
LCS(Longest Common Sequence)
Problems and solutions , LCS Algorithms.

4. Information Retrieval Systems

Information Retrieval - Precision,Recall,F- score
TF-IDF
KNN for document retrieval
K-Means for document retrieval
Clustering for document retrieval

5. Projects And Case Studies

a. Sentiment analysis for twitter, web articles
b. Movie Review Prediction
c. Summarization of Restaurant Reviews



Module 10 TRAINING AND DEPLOYING ML MODEL USING GCP : 6 hours

1. Introduction To GCP Cloud ML Engine :

Introduction to Google CloudML Engine
CloudML Engine in Machine Learning WorkFlow
Components of Cloud ML Engine - Google Cloud Platform Console.
gcloud command-line tool and Rest API

2. Training Machine Learning Model :

Developing a training application
Packaging a training application
Running and monitoring a training job
Using hyperparameter tuning
Using GPUs for training models in the cloud

Real - Time Projects

Project No.1 : Loan Default Prediction

DataSet : Banking Data

Domain - Banking & Finance

The bank wants to improve their services by finding interesting groups of clients. Fortunately, the bank stores data about their clients, the accounts (transactions within several months), the loans already granted, the credit cards issued. This process of loan default prediction can be done with machine learning algorithms.

Project No. 2 : Clustering Customers

DataSet : BigBazar/Future Group

Domain - Retail industry

Big Bazaar has retail outlets across major metropolitan cities in India. With the help of machine learning algorithms we can better understand customer behaviour and understand their buying needs better. BigBazaar runs various loyalty programs, festive offers which provide their customer more opportunities to avail discounts.

Project No. 3 : IBM HR Analytics

DataSet : IBM

Domain - Demand/Supply

Applying analytic processes to the human resource department of an organization in the hope of improving employee performance and therefore getting a better return on investment. This is especially concerning if your business is customer facing, as customers often prefer to interact with familiar people.

Project No.4 : Forecasting Uber Demand

DataSet : Uber & Rapido

Domain - Demand/Supply

The goal is to create an interactive dashboard using Tableau. This Tableau Dashboard can be used to get historical insights into a neighborhood.

For example, see its upcoming forecasted demand, increase the accuracy, decrease surge pricing events.

Project No. 5 : Analyzing Health Data and tracking human activity

DataSet : Samsung

Domain - Healthcare

The goal is to breakdown all the data that the Samsung Health app has collected and see what useful insights we can gain by analyzing it.

Project No.6 : Identify fraudulent credit card transactions.

DataSet : Banking Dataset

Domain - Banking & Finance

To recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase. It involves various processes like Data Cleaning, Data Visualization, Insights generation, Model generation, Feature Engineering and so on.

Real - Time Projects

Project No. 7 : Consumer Reviews of Amazon Products

DataSet : Amazon Data

Domain - E-Commerce

The goal is to analyze Amazon's most successful consumer electronics product launches; discover insights into consumer reviews and assist with machine learning models.
What are the most reviewed Amazon products?
How do the reviews in the first 90 days after a product launch?

Project No. 8 : Airbnb New User Bookings

DataSet : Airbnb

Domain - Travel & Hospitality

The goal is to predict which country a new user's first booking destination will be. By accurately predicting where a new user will book their first travel experience, Airbnb can share more personalized content with their community, decrease the average time to first booking, and better forecast demand.

Project No. 9 : Netflix Movies and TV Shows

DataSet : Netflix

Domain - Media and Entertainment

Explore what all other insights can be obtained from the list of tv shows and movies available on Netflix as of 2019. Understanding what content is available in different countries
Identifying similar content by matching text-based features
Network analysis of Actors / Directors and find interesting insights.

Project No. 10 : Walmart Sales Forecasting

DataSet : Walmart

Domain - Retail

This dataset contains the sales for each department from the Walmart dataset containing data of 45 Walmart stores, selected holiday markdown events are also included. These markdowns are known to affect sales, but it is challenging to predict which departments are affected and the extent of the impact.

Project No. 11 : BMW Pricing Challenge

DataSet : BMW dataset

Domain - Automation

To find a good statistical model to describe the value of a used car depending on the basic description
How does the estimated value of a car change over time? Can you detect any patterns?
How big is the influence of the factors not represented in the data on the price?

Project No. 12 : Bosch Production Line Performance

DataSet : Bosch

Domain - Manufacturing

To predict internal failures using thousands of measurements and tests made for each component along the assembly line.
This would enable Bosch to bring quality products at lower costs to the end user. The goal is to predict which parts will fail quality control

Real - Time Projects

Project No. 13 : Trending YouTube Video Statistics

DataSet : youtube

Domain - Social Media

The dataset of this project are daily record of the top trending YouTube videos, to generate insights like : Sentiment analysis in a variety of forms. Categorizing YouTube videos based on their comments and statistics. Training ML algorithms like RNNs to generate their own YouTube comments.

Project No. 14 : Identify And Predict Customer churn

DataSet : Telecom

Domain - Telecom

The goal is to develop a churn prediction model which assists telecom operators to predict customers who are most likely subject to churn.
Also to understand the customer behavior and reasons for churn.
Multiple classification models to predict the customer churn in telecom industry.

Project No. 15 : Smart Supply Chain for Big Data Analysis

DataSet : Dataco

Domain - Supply Chain

A DataSet of Supply Chains used by the company DataCo Global is used for the analysis. Dataset of Supply Chain , which allows the use of Machine Learning Algorithms and R Software. It also allows the correlation of Structured Data with Unstructured Data for knowledge generation.

Get in touch!

Call Us
+91 77 956 87 988

Mail Us
contacts@learnbay.co

Visit Us
www.learnbay.co

Whatsapp Us
7349-2222-63

click on icon to

Follow Us On

HSR Office :
Learnbay, 147, 5th Main Rd,
Rajiv Gandhi Nagar, HSR Sector
7, Near Salarpuria Serenity,
Bengaluru, Karnataka 560102
INDIA

Marathahalli Office :
Learnbay, 19/1, 2nd Floor,
Classic Aura (Beside Aricent),
Marathahalli - Outer Ring Road,
Kadubeesanahalli, Bengaluru,
Karnataka 560103 INDIA