upGrad

iiit-b
ज्ञानमुक्तमम्

International
Institute of Information
Technology Bangalore

The
PLAN BOOK

• • •

Executive Programme in
Machine Learning and Artificial Intelligence

DATA SCIENCE TOOLKIT (PREPARATORY COURSE)

WHY THIS MATTERS

Learn the basics of Python in order to lay a strong foundation for your data journey.

Use Python as a data science tool rather than a general-purpose programming language. Focus on Python based libraries specially used for data science such as NumPy and Pandas.

Data visualisation is essential before you start modelling and drawing inferences. Data professionals use different visualisation techniques to analyse and investigate data sets.

KEY OUTCOMES

● ● ●

80% of a data scientist's time is spent on cleaning data. Learn how to clean, manipulate and visualise data using Python libraries such as NumPy, Pandas, Seaborn, and more.

Learn different visualisation tools, such as heatmap, bar plots and scatter plots, under the Seaborn and Plotly libraries.

Learn how to find and analyse patterns in the data to draw actionable insights.

ACTIVITIES PLANNED FOR YOU

● ● ●

Live sessions to help you with the applications of the concepts learned in real time cases.

Personalised Industry (PI) Sessions Connect with a dedicated industry coach in a group of 8-10 peers; all matched based on similar work experience and expected outcomes.

Engage in peer-to-peer learning.

COURSE REQUIREMENTS

Three weeks are given to complete the course as optional content.

STATS AND EDA

WHY THIS MATTERS

The industry is gradually shifting to cloud. Most applications are serviced from cloud platforms. Learn about popular cloud services such as Git and AWS.

Statistical methods are needed to understand the data used to train ML models and interpret their results.

Learn how to formulate and validate different hypotheses to solve real-life business problems.

KEY OUTCOMES

● ● ●

Version control using Git, creating a portfolio on GitHub, essential services on the AWS cloud.

Descriptive statistics, fundamentals of probability and probability distributions, sampling and inferring results about a population.

You will learn the details of hypothesis testing and the different types of techniques such as the p-value method, A/B testing, z-test, and chi-squared test.

ACTIVITIES PLANNED FOR YOU

● ● ●

Live sessions to help you with the applications of the concepts learned in real time cases.

Personalised Industry (PI) Sessions Connect with a dedicated industry coach in a group of 8-10 peers; all matched based on similar work experience and expected outcomes.

Engage in peer-to-peer learning.

Contribute to open-source libraries such as NumPy.

Host code on GitHub.

COURSE REQUIREMENTS

Basic knowledge of Python and Excel.

Five weeks with 12-15 hours a week.

MACHINE LEARNING — I
WHY THIS MATTERS

Using the concepts learnt earlier, you can help an asset management company make different investments by identifying the best sectors in other countries.

The core of data science comprises understanding data and training it to make predictions on unseen data points. While several advanced algorithms have emerged, a

fundamental understanding of the basic concepts and algorithms is necessary.

The naive Bayes classification model performs better and can train on less data than logistic regression.

Given the options available for modelling a solution, it is imperative that you learn how to select the best model to solve any problem.

KEY OUTCOMES

Use the concepts you learned previously to help an asset management firm in finding the best sectors in other countries to invest in.

The fundamentals of linear regression.

Assume the role of an analyst at an investment bank and determine where the bank should invest. Also, explain the recommendations instead of the analysis conducted.

Logistic regression.

Naive Bayes classification: Understand the different evaluation metrics and learn how you can decide which algorithms yield the best value.

Learn about the generalised and regularised regression models and apply them to solve problems that cannot be solved using linear regression.

Model selection will help you apply the learning to build a better model, which performs well on the training as well as the test data set. Also, evaluate different models using different evaluation metrics.

ACTIVITIES PLANNED FOR YOU

Live sessions to help you with the applications of the concepts learned in real time cases.

Personalised Industry (PI) Sessions
Connect with a dedicated industry coach in a group of 8-10 peers; all matched based on similar work experience and expected outcomes.

Linear regression assignment
Implement the linear regression concepts to build a linear model to predict the demand for bike sharing.

COURSE REQUIREMENTS

Basic knowledge of Python and Excel.

Six weeks with 12-15 hours a week.

MACHINE LEARNING — II

WHY THIS MATTERS

Although you have learnt about linear regression in the previous modules, the industry presents many challenges that require techniques in extension to linear regression.

Expand your understanding of machine learning with industry benchmarked methods — random forests.

Expand your understanding of machine learning with industry benchmarked methods — boosting. These techniques can drastically increase the performance of any classification/regression task.

Clustering introduces you to the concept of unsupervised learning techniques, wherein you try to find patterns based on similarities in the data since you do not have a target or outcome variable.

Real-life data sets can have many correlated features.

Principal component analysis (PCA) can help you reduce the dimensions of your original data set for resolving such issues.

KEY OUTCOMES

• • •

Learn about the generalised and regularised regression models and apply them to solve problems that cannot be solved using linear regression.

Learn about tree models and boosting and implement them to build cutting-edge machine learning solutions for classification problems.

You will also learn about decision trees, random forests, Adaboost, gradient boosting, and XGBoost.

Unsupervised ML using different clustering algorithms such as k-means clustering, hierarchical clustering, and k-mode clustering.

Learn essential concepts related to dimensionality reduction, the learning algorithm of PCA and its practical applications in supervised and unsupervised problems.

ACTIVITIES PLANNED FOR YOU

• • •

Live sessions to help you with the applications of the concepts learned in real time cases.

Personalised Industry (PI) Sessions
Connect with a dedicated industry coach in a group of 8-10 peers; all matched based on similar work experience and expected outcomes.

Case Study
Apply your learnings from the course on Machine Learning — II and solve a real-world case study.

COURSE REQUIREMENTS

Basic knowledge of Python and Excel.

Eight weeks with 12-15 hours a week.

INTRODUCTION TO NEURAL NETWORKS

WHY THIS MATTERS

Introduction to neural networks constitutes the fundamental block.

Convolutional neural networks (CNNs) are used widely to solve many computer vision problems.

Understand how CNNs are used in real-time industrial applications and recurrent neural networks (RNNs) help solve problems involving sequences of data, such as text classification, language modelling, and time series analysis.

KEY OUTCOMES

• • •

Understand the working of an artificial neural network to perform a simple image classification task.

Leverage a CNN to extract complex features in images or videos and build models that can recognise different classes of images.

Apply and utilise transfer learning in different scenarios for image classification or object detection. You will also be introduced to the concept of OpenCV.

Learn about the basic architecture of an RNN, its different architectural variants, such as bidirectional RNNs, and some popular gated variants such as long short term memory (LSTM) and gated recurrent units (GRU).

ACTIVITIES PLANNED FOR YOU

• • •

Live sessions

Personalised Industry (PI) Sessions

Get hands-on experience of building an end-to-end pipeline for training CNNs in different classification tasks, such as flower and chest X-ray images.

In the RNN module, apply RNNs to tag words in an English corpus with their part-of-speech (POS) tags. Additionally, you will learn how to implement 1D CNN-RNN architecture.

Apply the learnings on gesture recognition to build a smart TV system.

COURSE REQUIREMENTS

Basic knowledge of Python and Excel.

Eight weeks with 12-15 hours a week.

NATURAL LANGUAGE PROCESSING (NLP)
WHY THIS MATTERS

Lexical processing introduces you to the intricacies involved in text cleaning and pre-processing.
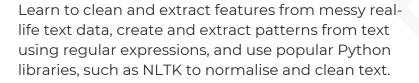
Syntactic processing helps understand the grammatical structure and, hence, the meaning of text.

Semantic processing methods are the cornerstone of modern NLP and it is important to learn how the contextual meaning of a corpus of text can be understood and modelled using ML methods.

KEY OUTCOMES
● ● ●

Learn to clean and extract features from messy real-life text data, create and extract patterns from text using regular expressions, and use popular Python libraries, such as NLTK to normalise and clean text.

Learn how to form grammatical rules to extract meaningful information from free-form text. You will also be able to train a custom model to recognise common word entities.

Learn how to create word vectors and model the contextual meaning of words.

Extract latent topics from a text corpus and understand what are the broad themes that are mentioned in the text.

ACTIVITIES PLANNED FOR YOU
● ● ●

Live sessions to help you with the applications of the concepts learned in real time cases.

Personalised Industry (PI) Sessions
Connect with a dedicated industry coach in a group of 8-10 peers; all matched based on similar work experience and expected outcomes.

Case Study
Apply your learnings from the NLP course and solve a real-world case study.

COURSE REQUIREMENTS

Basic knowledge of Python and Excel.

Seven weeks with 12-15 hours a week.

CAPSTONE

WHY THIS MATTERS

Choose from a variety of capstones based on your industry to complete the course.

KEY
OUTCOMES

Multiple end-to-end industry projects from the most relevant domains in the present-day landscape where data is being used to drive a majority of the decisions. Gain expertise in your desired domain and add a solid proof of work in your portfolio.

COURSE REQUIREMENTS

Four weeks with 12-15 hours a week.

READY TO GO THE DISTANCE?

● ● ●

Kickstart your journey on the road of learning!