

Data Science 1 Month Hands-on Workshop Session Coverage

Main Modules

1	Introduction to AI Big Data Analytics and Foundational Statistics
2	Exploratory Data Analysis (EDA)
3	Data Cleaning and Preprocessing
4	Python Programming
5	Data visualization
6	Machine Learning Basics
7	Database Management
8	Data Warehousing
9	Probability Concepts
10	Time Series Analysis
11	Employability Skills
12	Soft Skills

Modules in Detail

Module 1	Introduction to AI Big Data Analytics and Foundational Statistics	Duration: 24 Hrs
Key Learning Outcomes	<ul style="list-style-type: none"> • Explain the relevance of AI & Big Data Analytics for the society • Explain the various use-cases of AI & Big Data in the industry • Define “general” and “narrow” AI • Describe the fields of AI such as image processing, computer vision, robotics, NLP, etc. • Outline a career map for roles in AI & Big Data Analytics • Analyse the differences between key terms such as Supervised Learning, Unsupervised Learning and Deep Learning • Distinguish between different probability distributions such as Normal, Poisson, Exponential, Bernoulli, etc. • Identify correlation between variables using scatterplots and other graphical techniques • Apply basics of descriptive statistics including measures of central tendency such as mean, median and mode • Apply different correlation techniques such as Pearson’s Correlation Coefficient, Methods of Least Squares etc. • Apply different techniques for regression analysis including linear, logistic, ridge, lasso, etc. • Use hypothesis testing to draw inferences and measure statistical significance • Distinguish between prior and posterior measures of probability • Comprehend the difference between supervised and unsupervised learning • Describe the suitable statistical models for structured and unstructured data • Use maximum likelihood estimation to estimate the parameters of a statistical model • Apply the concepts of conditional probability including Bayes theorem • Discover the basics of using statistical software packages and IDEs such as RStudio, Jupyter Notebooks • Apply basic functions and libraries present in statistical software packages and IDEs • Make use of statistical packages, frameworks and libraries such as NumPy and Pandas in developing applications 	

Module 2 & 5	Exploratory Data Analysis (EDA) & Data Visualization	Duration: 18 Hrs
Key Learning Outcomes	<ul style="list-style-type: none"> • Identify the type of data, volume of data, and variables required for the analysis • Distinguish between different types of data such as numerical, categorical, etc. • Identify common open and paid data sources • Discuss the uses and characteristics of different open source and paid data sources • Describe the purpose of metadata • Describe various Data validation tools and processes • Demonstrate the process of capturing various types of data such as enterprise data, consumer data etc. from various data sources • Conduct the process of importing data from both public and private databases or data stores and store it in datasets or data frames • Organize and map metadata as per the needs of the analysis • Perform data profiling for data quality assessment and validation • Uses statistical graphics and other methods to summarize the main characteristics of a dataset, • Uncover patterns and relationships • Use statistical techniques like descriptive statistics (mean, median, standard deviation) and data visualization methods (charts, graphs, etc.) 	

Module 3	Data Cleaning and Preprocessing	Duration: 18 Hrs
Key Learning Outcomes	<ul style="list-style-type: none"> • Differentiate the unprocessed and processed data • Explain the impact of unprocessed data on subsequent analytical operations • Describe the various anomalies that may be found in unprocessed data (e.g. missing values, incorrect data types, and redundant data) • Explain the Data Normalization techniques and concepts • Describe the properties of different tools that can be used to validate the pre-processed data • Analyze unprocessed data to discover anomalies such as missing values, incorrect data types, etc. • Apply different techniques and functions to clean unprocessed data including removing missing values, transforming incorrect data types, etc. • Apply different approaches to normalize datasets such as feature scaling etc. • Apply appropriate tools and techniques to perform pre-processed data validation 	

Module 4 & 6	Python Programming & Basics of Machine Learning	Duration: 24 Hrs
Key Learning Outcomes	<ul style="list-style-type: none"> • Distinguish between the limitations of different programming, command line or scripting languages to develop machine learning algorithms • Select the most suitable programming languages to develop or optimize the statistical machine learning algorithm • Use object-oriented programming concepts such as abstraction, encapsulation, modularity, etc. to write user defined functions and classes • Apply dynamic programming concepts to solve complex optimization problems • Use the streaming model to compute real-time or large amounts of data that cannot be stored in the memory • Identify the objective of the analysis and define the suitable predesigned algorithms, libraries, packages, frameworks, applications to address the objective • Categorize the different use cases for algorithms such as recommender engines, chatbots, image annotation • Distinguish between the pros and cons of different open-source libraries and packages to satisfy specific use cases • Apply algorithms to specific use cases and scenarios such as vision, text recognition, image recognition, natural language processing based on suitable open-source libraries selected • Validate the models implemented using appropriate tools and processes • Optimize and iterate algorithm models to achieve quality of outputs 	

Module 7 & 8	Database Management & Data Warehousing	Duration: 12 Hrs
Key Learning Outcomes	<ul style="list-style-type: none"> • Introduction to Structured and non-structured database management system. • Data Modelling/Relationship • Database and Objects • Handling Structured data using MySQL/Oracle/ etc. • Handling NO-SQL data handling using MongoDB 	

Module 9 & 10	Probability Concepts & Time Series Analysis	Duration: 18 Hrs
Key Learning Outcomes	<p><u>Intro (2 Hrs)</u></p> <ul style="list-style-type: none"> • Collect data points at regular intervals over time • Build a model that represents the data • Validate the model • Use the model to predict future values or impute missing values <p><u>Graph Algorithms (6 Hrs)</u></p> <ul style="list-style-type: none"> • Evaluate different algorithms for decomposing graphs into parts • Evaluate different algorithms for finding shortest paths in graphs such as breadth-first search, shortest-path-tree, Dijkstra's algorithm and Bellman-Ford algorithm • Evaluate greedy algorithms such as Kruskal's algorithm and Prim's algorithm to solve minimum spanning tree problems • Apply the basics behind undirected graphs such as representing and exploring graphs, previsit and postvisit orderings etc. • Apply the basics behind directed graphs such as acyclic graphs, topological sorting and computing strongly connected components <p><u>String Algorithms (4 Hrs)</u></p> <ul style="list-style-type: none"> • Summarize the concepts behind algorithms such as suffix trees that are used for pattern matching • Use brute force approaches for pattern matching • Use algorithms such as suffix arrays and Burrows-Wheeler Transform for approximate pattern matching • Use algorithms such as Knutt-Morris-Pratt for exact pattern matching • Apply different techniques to construct suffix trees and arrays <p><u>Neural Networks (6 Hrs)</u></p> <ul style="list-style-type: none"> • Distinguish between different types of recurrent neural networks and commonly used variants such as GRUs and LSTMs • Summarize the concepts behind convolutional neural networks and recurrent neural networks • Build shallow and deep neural networks using techniques such as forward propagation and back propagation • Apply the foundational layers of convolutional neural networks such as pooling and convolutions and stack them properly in a deep network to solve multi-class image classification problems • Build convolutional neural networks and apply it to object detection problems • Use word vector representations and embedding layers to train recurrent neural networks • Apply attention model intuition and trigger word detection to speech recognition problems 	

Module 11	Employability Skills	Duration: 18 Hrs
Key Learning Outcomes	<ul style="list-style-type: none"> • Career Development and Goal Setting (2 Hrs) • Environment, health, and safety (EHS) (4 Hrs) • Getting Ready for Apprenticeship and Jobs (4 Hrs) • Inclusive and environmentally sustainable workplaces (8 Hrs) 	

Module 12	Soft Skills	Duration: 18 Hrs
Key Learning Outcomes	<ul style="list-style-type: none"> • Practical Communication Skills (6 Hrs) • Essential Digital Skills with Office Etiquettes (4 Hrs) • Getting Ready for Apprenticeship and Jobs (4 Hrs) • Introduction to Entrepreneurship (4 Hrs) 	

Total Duration: 150 Hrs - 25 Days hands on workshop with 6 hrs per day sessions.