CORPNCE

CORPNCE TECHNOLOGIES PVT LTD

TRAINING | PRODUCT | CONSULTING

ADVANCED DATA SCIENCE & MACHINE LEARNING

NOV 2024

🕒 3 Hrs Introduction to Data Science

1. What is Data Science
2. How it is different from Big Data and Data Analytics
3. Data Driven decision making
4. Purpose and Business problems
5. How Data Scientist work
6. Intro to basic AI terminology
7. Different sectors using Data science
8. Overview of course roadmap
9. Course ethics and student expectations
10. Intro to training schedule
11. Student queries

MODULE 1 🕒 36 Hrs Introduction to Python Programming

Day 1 - 3hrs - Intro | installation | Operators

1. A Brief History of Python
2. Python Versions
3. Installing Python
4. Environment Variables
5. Executing Python From The Command Line
6. IDLE – JupyterLab, VSCode
7. Python Documentation
8. Intro to Syntax
9. Operators - Arithmetic, Logical, Comparison etc.
10. Intro to Dunder Methods
11. Intro to ChatGPT and StackOverflow for help

Day 2 - 3hrs - Containers | Tuple | Loops

1. Comments - How to make comments in Python
2. Intro to containers - Lists, Tuple, Set, Dict etc.
3. Tuple - Intro
4. Dunder Methods of Tuple
5. Basic properties
6. Heterogeneous
7. Indexing and slicing
8. Immutability
9. Iteration - Intro to For & While loops
10. Examples of For loops with Tuple
11. Intro to IF conditions
12. Intro to ELIF and ELSE conditions
13. Break and Continue
14. Examples of Loops with Conditions

Day 3 - 3hrs - Lists | Loops | Examples

1. List Syntax
2. List Dunder Methods
3. Mutability
4. Pre-Defined Methods - Append, Pop etc.
5. Slicing and Loops in List
6. Zip and Enumeration
7. Examples of List with Loops
8. List Comprehension
9. Application of List in Real World

Day 4 - 3hrs - Set | Strings

1. Intro to Set - Collection of Unique Numbers
2. Syntax and Dunder Methods
3. Union, Intersection and Update Method
4. Application of Set in Real World
5. Examples of Set
6. Intro to Strings
7. String Dunder Methods
8. Predefined Methods of String
9. String Slicing and Immutability
10. Example of String with Loops

Day 5 - 3hrs - Regular Expression

1. Importance of Regular Expression
2. Intro to Re Module in Python
3. Regular Expression Syntax
4. Flags in Re
5. Difference between FindIter, Match and Search
6. Phone Number, Email, Website Example
7. Practice session on Strings and RegEx

Day 6 - 3hrs - Error Handling | Dictionary

1. Errors
2. Run-Time Errors
3. The Exception Model
4. Exception Hierarchy
5. Handling Multiple Exceptions
6. Try and Except

7. Raise and Assert
8. Writing Your Own Exceptions using PyDantic (optional)
9. Intro to Dict
10. Dict Dunder Methods
11. Dict Pre-Defined Methods
12. Mutability
13. Keys, Values & Items
14. Looping in Dict
15. Dictionary Comprehension
16. Similarity between Dict and JSON
17. Examples

Day 7 - 3hrs - Functions | Examples

1. Why Functions
2. Function Syntax
3. 4 Types of Arguments in Python
4. Pass and Return Example in Functions
5. Doc Strings in Function
6. Type hint in Function
7. Intro to Typing Module
8. Examples of Functions with Data Types

Day 8 - 3hrs - String Formatting | File Handling

1. String Formatting
2. Difference between - % , .format and f-string
3. Examples
4. Intro to File Handling
5. Access Modes

6. Reading & Writing Data from a File
7. Additional File Methods
8. Handling IO Exceptions
9. Difference between Module, Package and Library
10. Importing Files from Directories
11. Intro to Sys and Os Module
12. Sys Path setup
13. Byte Modes
14. Examples of Copying Text and Images between Files

Day 9 - 3hrs - Intro to Oops I VSCode

1. VSCode installation
2. Classes in Python
3. Principles of Object Orientation
4. Creating Classes
5. Attributes and Methods
6. Dunder Methods In Python Class
7. Inheritance
8. Polymorphism
9. Encapsulation
10. Decorators
11. ClassMethods and StaticMethods
12. Examples

Day 10 - 3hrs - Intro to Design Patterns | Git | Github

13. Intro to Design Pattern
14. Strategy Design pattern
15. Intro to Interface and Abstraction
16. Intro to Resources

17. Intro to Design Pattern Book
18. GeekTrust Challenge Example
19. Intro to Git
20. Git installation
21. Git Commands
22. Intro to Github
23. Pushing the first code to Github

Python Test | Doubt Clearing Class - 6hrs

MODULE 2 24 Hrs on Stats and Probability

Day 13 - 3hrs - Introduction | Descriptive Statistics

1. Statistical Analysis - How it is different from Assumptions
2. Major categories of statistics – Frequency and Bayesian
3. Intro to Data collection
4. OLTP vs OLAP
5. Data Warehouse Vs Data Lake etc.
6. AWS services for Data Pipeline
7. Central tendency measures – Mean, Median, Mode
8. Measures of dispersion – Variance, Standard Deviation
9. Interquartile Range and box plots with outliers
10. Different type of Graphs - When to Use What!
11. Frequency Distribution – Histograms
12. Bar graph, line plot, Pie chart, Scatter plots, Hexagonal Bins, Heat Map
13. Examples with Python Code
14. Python code explanation of Box plots

Day 14 - 3hrs - Introduction to Probabilities

1. Difference between probabilities and Statistics
2. Common Terms & Fundamental Definition
3. Mutually exclusive events
4. Marginal Probabilities
5. Conditional Probabilities
6. Joint Probability
7. Example of Each type
8. Rules of Independence
9. AND or Intersection in probability
10. Rule of Addition
11. Baye's Theorem
12. Examples of Bayes
13. Intro to Baysian Network
14. Examples

Day 15 - 3hrs - Introduction to Discrete Distributions

1. Discrete Probability Distributions
2. Random Variables
3. Expected Value in Probability
4. Probability Mass Functions
5. Bernoulli Distribution
6. Binomial & Multinomial Distribution
7. Poisson distribution
8. Examples

Day 16 - 3hrs - Introduction to Continuous Distributions

1. Continuous Probability Distributions
2. Difference between PMF and PDF (probability Density Function)

3. Intro to Integration and Area under the Curve
4. Uniform Distributions
5. Intro Univariate Gaussian distribution
6. Empirical Rules with Z- Score and Python codes
7. Multivariate Normal Distribution
8. Covariance matrix for multivariate gaussian
9. Marginalisation and Conditioning Intro
10. Examples & Practice

Day 17 - 3hrs - Inferential Statistics

1. What is Sampling
2. Central Limit Theorem
3. Sampling Distribution of Sample Mean
4. Standard Error
5. Example of Sampling Distribution of Sample Mean
6. Confidence Intervals for Z-Distributions
7. Margin of Error
8. Limitation of Z-Distribution
9. Degrees of freedom
10. T- Distributions and T- Statistics
11. PDF of T-Distribution (optional)
12. Intro to T- table
13.Examples

Day 18 - 3hrs - Hypothesis Testing

1. What is Hypothesis Testing
2. Conducting Experiments
3. Examples of Hypothesis Testing
4. What is the p-Value
5. Critical Value
6. Type -1 and Type - 2 Errors
7. Examples
8. Similarity – Cohen Kappa
9. Correlation Vs Covariance

Stats Test | Doubt Clearing Class - 6hrs

MODULE 3 🕒 3 Hrs SQL Databases | APIs

Day 21 - 3hrs - Intro to Databases | APIs

1. Basics of data categorisation and different formats of data
2. Difference between SQL and NOSQL database
3. Importance of storing data
4. AWS data pipeline Revision- OLTP vs OLAP
5. Introduction to PostgreSQL
6. Common SQL commands
7. Introduction to FastAPI
8. Resource Assignment for FastAPI
9. Introduction to Postman for testing
10. CRUD operation Example on real-time data

MODULE 5 🕒 3 Hrs NumPy

Day 22 - 3hrs - Intro to NumPy | Arrays | Random Data Generation

1. Download Resources for NumPy from Corpnce Link
2. What is NdArray and Why to use it instead of List
3. PyObject Head - How NumPy Arrays are Faster
4. Computation on arrays – Universal Functions
5. Computation on arrays – Aggregate Function
6. Computation on arrays – Array Broadcasting
7. Boolean Arrays and Masks
8. Advanced Indexing
9. Numpy Tricks and Application

10.Random Process Intro with Numpy

11. Numpy as the backbone of Pandas

12. Examples

MODULE 6 🕒 6 Hrs on Pandas

Day 23 - 3hrs - Pandas | DataFrames | Merging

1. Why Pandas, What is Data Wrangling?
2. Different ways of creating pandas Series
3. Intro to DataFrame
4. Hierarchal indexing
5. Missing Values
6. Concatenating
7. Merging DataFrame
8. Grouping and Aggregation
9. Pandas Melt
10. Crosstab
11. Working with Strings - Regex with pandas

Day 24 - 3hrs - Data Loading | Real Time Analysis

1. Importing data with Pandas
2. Read and write Excel, CSV files and JSON
3. Dealing Missing Values on real time data
4. Merge, Apply and GroupBy examples
5. Netflix Data Loading and preprocessing
6. Time series Analysis (optional)

MODULE 7 🕒 6 Hrs Data Visualisation

Day 25 - 3hrs - Matplotlib | Deciding Graphs

1. Importance of data visualisation
2. Different verities of plots necessary for data science toolbox
3. Which plot to use where and how to derive maximum information
4. Storytelling with data
5. Matplotlib – Normal vs. Object Oriented
6. Line Plots and Scatter Plots
7. Errorbars – Uniform and Gaussian
8. Hex-bin plots as a replacement
9. Density and Contour plots
10. Histograms and Binning
11. Customising Legends and Colorbars
12. Text and Annotation
13. Customising Ticks

Day 26 - 3hrs - Seaborn | Plotly

1. Intro to Seaborn
2. DistPlot, JointPlot, KDEPlot
3. Boxplot, Violinplot, CatPlot
4. Intro to Plotly Express - Dynamic Plotting
5. Examples of different plotting with Plotly
6. Figure Object in Plotly
7. Geo plot with Plotly
8. Examples

MODULE 8 🕒 6 Hrs EDA

Day 27 - 3hrs - EDA | Examples

1. Data Analytics Process
2. Exploratory Data Analysis (EDA)
3. How to start with Data Analytics Project
4. Data Analytics on Netflix Data
5. Intro to Kaggle

Day 28 - 3hrs - EDA | Mini Project

1. Continue EDA
2. First Mini-project Assignment

MODULE 9 🕒 35 Hrs Machine Learning

Day 31 - 5hrs - Intro to Machine Learning | Regression Analysis

1. Intro to Machine Learning
2. Supervised and unsupervised learning
3. Example of Train and Test Data
4. Classification vs Regression
5. What is a ML Model?
6. What is Training?
7. Examples
8. Intro to Regression Analysis
9. Definition of Variance
10. Mathematical Equation
11. What is Hypothesis
12. Simple Linear Regression
13. Definition of Regression as Collection of Sample Means
14. Why Confidence Interval around prediction?
15. Dependent Variable - The Gaussian Explanation
16. MSE - The Gaussian Error

Day 32 - 5hrs - OLS | Metrics

1. What is Model Optimisation?
2. What is OLS (Ordinary Least Square)?
3. What is Gradient Descent?
4. Derivation of OLS (optional)
5. Regression Model Metrics

6. Difference between MSE, MAE, RMSE
7. Intro to variance
8. Total Variance and Captured Variance
9. R2 Score
10. Adjusted R2 Score
11. Implementation of Regression with SkLearn
12. Intro to Hyperparameters
13. Train Test Split
14. Cross Validation
15. Intro to Non Linear data
16. Polynomial Regression
17. Over-Fitting Vs Under Fitting
18. Validation Curve
19. Selecting the right Degree
20. Grid Search CV

Day 33 - 5hrs - Lasso | Ridge | Feature Engineering

1. Impact of Outliers on MSE
2. Intro to Regularisation
3. L1 and L2 Regularisation
4. Impact of Regularisation Error on MSE
5. L1 & L2 Error Optimisation -The Visual Explanation
6. Lasso Regression with SkLearn
7. Ridge Regression with SkLearn
8. Finding Feature importance with Lasso
9. When to use Lasso over Ridge and Vice Versa
10. Feature Engineering
11. Processing Categorical data
12. Intro to OHE
13. TF-IDF

14. Housing Data Analysis
15. Project Assignment

Day 34 - 5hrs - Logistic | Cross Entropy | Gradient Descent

1. Intro to Classification
2. Intro to Logistic Regression
3. Why Linear Regression can not be used for classification
4. Why MSE can not be the error for Classification
5. Maximum Likelihood Estimation (MLE)
6. Cross Entropy Loss
7. Sigmoid Function
8. Gradient Descent
9. Stochastic, Batch & Mini Batch Gradient Descent
10. How to decide the Threshold for LR
11. Metrics - Accuracy Score
12. Classification challenges - Imbalanced Data
13. Confusion Matrix
14. Specificity and Sensitivity
15. Threshold for Specificity
16. F1 - Score
17. ROC and AUC
18. Example of Classification with SkLearn
19. Assignment and Github upload

Day 35 - 5hrs - Naive Bayes | Decision Trees | RF | XGBoost

1. Intro to Naive Bayes
2. Generative Models
3. Gaussian NB
4. Multinomial NB
5. Advantages and Limitation
6. Spam Vs Ham classification with Text Processing and NB

7. Probability in Decision Trees
8. Entropy as Loss Function
9. Overfitting in Decision Trees
10. Examples
11. Random Forest
12. Bagging
13. Feature selection with RF
14. Bagging Vs Boosting
15. Implementation with Sklearn and XGBoost

Day 36 - 5hrs - SVM | Statistical Tests | Projects

1. Intro to SVM
2. Why Margins?
3. Support Vectors
4. Linear Kernels
5. RBF
6. Kernel Trick
7. Intro to statistical tests
8. Parametric Vs Non Parametric
9. When to use which test
10. First Major Project for ML

Day 37 - 5hrs - Unsupervised Learning

1. Intro to unsupervised Learning
2. Clustering with K-Means
3. Expectation and Maximisation
4. Deciding the no of clusters
5. Silhouette Analysis
6. Example
7. Intro to Dimensionality Reduction
8. Intro to PCA

9. Maths behind PCA
10. Deciding the Captured Variance with right Degree
11. Advantages of PCA
12. Limitations
13. Implementation with Sklearn
14. Target Encoding
15. Global Information Loss Vs Local Information Loss
16. Intro to TSNE
17. TSNE Vs PCA
18. Hyper-parameter selection for TSNE
19. Example

MODULE 10 🕒 60Hrs Deep Learning & NLP

Day 39 - 5hrs | ANN | FCNN

1. Introduction to Deep Learning
2. Difference with ML
3. Different types of ANN
4. Relation with Logistic Regression
5. FCNN Architecture
6. Forward Propagation
7. SoftMax Probability
8. Calculating Loss
9. BackProp
10. Gradient Descent in FCNN
11. SGD, RMS prop and ADAM
12. Intro to PyTorch
13. What is parallel computing?
14. MNIST implementation with PyTorch
15. Limitation of FCNN

Day 40 - 5hrs - CNNs

1. Why FCNN fails on noisy data
2. Spatial Information importance
3. Intro to CNN Architecture
4. How Convolution works?
5. MaxPooling and Edge Detection
6. Image classification with ResNet
7. Data Preparation with TorchVision
8. Implementation with PyTorch
9. Intro to Huggingface Models
10. Other Image processing Models

Day 41 - 5hrs - Transfer Learning | Example

1. What is fine tuning?
2. What is transfer learning?
3. Transfer Learning in CNNs
4. PyTorch Example of transfer learning
5. Limitation of Transfer Learning in Modern Models
6. Feature augmentation and Batch Normalisation
7. Project Assignment
8. Discussion and Doubt Clearing

Day 42 - 5hrs - NLP | RNN | LSTM

1. NLP Vs NLG Vs NLU
2. Intro to Word Embedding and word2vec
3. Context capturing
4. Sequential data - How Language is different from Image
5. What is a Language Model?
6. Why N-Gram LM failed?
7. RNN fundamentals

8. Vanishing and Exploding Gradients
9. Gradient Clipping
10. Intro to LSTM
11. Advantages and Disadvantages of LSTM
12. Stacked LSTM
13. Context capture in Bi-directional LSTMs

Day 43 - 5hrs - Transformers | Stepping GenAI

1. What is an Encoder and Decoder Model
2. SMT VS NMT
3. Neural Machine Translation with seq2seq
4. Bottle neck problem
5. Intro to Attentions
6. Different type of Attentions
7. Attention is all you Need
8. Intro to Transformer architecture
9. Self Attention - Key, Value & Query
10. Multi Head Attention

Day 44 - 5hrs - LLM | LangChain

1. Examples of LLM
2. Latest Trends
3. Intro to LangChain
4. Interacting LLMs with LangChain
5. Zero Shot Learning Vs Few Shot Learning
6. Prompt Engineering Basics
7. Hardware requirements to run LLMs
8. What is RAG?
9. What is Vector Database?

Day 45 - 5hrs - LLM | LangGraph

1. Intro to LanGraph
2. RAG with LanGraph
3. Example Code
4. Difference between LLM and Agent
5. Building Agent with LangGraph
6. Example
7. Road Ahead for LLMs

Day 46 - 5hrs - Coding

1. Instructor led final project
2. Deployment Instructions
3. Final Project Assignment (student led)
4. Doubt clearing session

Day 47 - 5hrs - Deployment and Packaging

1. Model packaging
2. Backend instructions with FastAPI
3. Intro to AWS and Databricks
4. Examples

PROJECTS

- There will be 15 minor projects, 2 Major and one portfolio project.
- Candidates are eligible for 6 months internship upon test
- Major and portfolio projects will have grades. Mandatory submission for certificate.
- Coding and Practice sessions will be extra.
- Faculty might slowdown on certain days according to progress.

Contact – Ranjan Mishra

Mobile - +91-9739604796

Email – info@corpnce.com

Address – #191, WOC service road, Rajajinagar, Near Mahalaxmi Metro Station