EdYoda

Data Scientist Program

Program Curriculum

Learning outcomes:

- Strong fundamental concepts of Frontend Development
- Strong fundamental concepts of Server-side Development using REST APIs
- Work with database seamlessly
- Create beautiful end to end websites
- Gain Industry standard knowledge

Data Wrangling

1. Introduction to Data Wrangling

- What is Data Wrangling?
- Why do all Python Developers need to learn Data Wrangling?
- Common Libraries for Data Wrangling - NumPy, Pandas & Seaborn
- Comparison with Python List

2. Essential NumPy

- Introduction to NumPy
- Creation & Access
- Numpy Statistics Function
- Stacking and Splitting
- Reshaping, Slicing and Flatten
- Image Processing

3. Data Manipulation, Cleaning & Transformation using Pandas

- Introduction to Pandas
- Why is Pandas used?
- Importing Data from Various Sources
- High and Low Level Data Understanding
- Univariate and Bivariate Analysis
- Data Transformation
- Grouping & Statistical Functions
- Merge & Join
- Feature Engineering
- Sorting and Indexing Dataframe

- Filtering Dataframe
- Usage of loc and iloc functions
- Pivot Table

4. Data Visualization using Seaborn:

- Why Visualization?
- Introduction to Seaborn
- Cat Plot
- Dist Plot
- Pair Plot
- Joint Plot
- Bar Plot
- Density Plot
- Joint Distribution Plot

Fundamental Maths for Data Scientist

1. Essential Maths & Statistics

- Essential Linear Algebra
- Matrix Operations
- Understanding distributions
- Probability Concepts
- Calculus
- Understanding distributions
- Mean, Median, Mode, Quantile
- Other statistics Concepts
- Sampling Techniques

Machine Learning

1. Nearest Neighbours & Basics of Clustering

- Introduction to Distance Vectors

- Classification using NN
- Understand, Train & Test Datasets
- Implement Nearest Neighbours using NumPy
- Importance of Metrices
- Classification Metrices
- Regression using NN
- Understanding Scikit Datasets
- Understanding Kaggle & UCI Datasets
- Understanding Unsupervised ML
- What is Clustering and where is it used?
- Introduction to KMeans
- Implement KMeans using NumPy

2. Data Preprocessing & Feature Selection

- Why Preprocessing is Required
- Preprocessing Numbers, Ordinal, and Nominal Values
- Linear and Non-Linear Models
- Linear Regression using Scikit-learn
- The Loss Function
- Gradient Descent Algorithm for Weight Optimization
- Linear Regression on House Price Data
- Implement Linear Regression using NumPy/Python
- Multiple and Polynomial Linear Regression
- Case Study for Linear Regression

3. Linear Models

- Linear Model for Classification
- Understanding Decision Boundary
- Challenges in Building Machine Learning Products
- Concept of Pipelining
- Connecting Preprocessing with Estimators
- Concept of Dealing with Heterogeneous Data

- Combining Multiple Pipeline with Column Transformer
- Case Studies

4. Pipelining, ColumnTransformer, Hyper-parameter Tuning

- Hyper Parameter Tuning
- Course Tuning vs Fine Tuning
- Cross Validation
- How to Choose the Best Model & Best Parameters
- Fundamentals of Decision Tree
- Decision Boundary of Decision Tree/Logistic Regression/Nearest Neighbour
- Generate Decision Tree using Information Gain & Gini Impurity
- When to use Entropy and Gini Index
- Variants of Decision Tree
- Regression Decision Tree
- Implement Decision Tree using Scikit

5. NLTK, Text Processing, Naive Bayes

- Introduction to Text Processing
- CountVectorizer
- Tf-IDF Vectorizer
- Hashing Vectorizer
- Effects of Preprocessing on Text Data
- Introduction to NLTK
- Stemming & Lemmatization
- Tokenization
- Handling Punctuation
- Using NLTK Preprocessor with Scikit
- Introduction to Conditional Probability
- Bayes Theorem Fundamentals
- Maths Behind Naive Bayes Theorem
- Assumptions of Bayes Theorem
- Types of Naive Bayes - Gaussian, Multinomial & Binomial
- Implement Naive Bayes using NumPy

6. Anomaly Detection, Imbalanced Classes & Ensemble Methods

- Anomaly Detection
- Techniques for Anomaly Detection
- What are Imbalanced Classes?
- Sampling Techniques - Over, Under, SMOTE
- Concept of Ensemble Models
- Bagging vs Boosting
- RandomForest
- AdaBoost
- VotingClassifier
- Deployment

Deep Learning Fundamentals

1. Introduction to Deep Learning

- What is Deep Learning?
- Applications of Deep Learning
- Differences between Machine Learning & Deep Learning

2. Neural Networks

- Introduction to Neural Networks
- Deep Dive into Neural Networks
- Single & Deep NN
- Back Propagation
- Activation Functions
- Linear Regression vs Neural Network
- Understanding Keras & Tensorflow
- DNN on Structured Data
- DNN on MNIST Data

- Applications of Deep Learning on Images
- Introduction to CNN
- Applications of CNN

Mindset for Problem Solving

1. Mathematical Aptitude
- Percentages
- Profit and Loss
- Simple Interest and Compound Interest
- Work And Time
- Probability
- Permutation and Combination
- Profit and Loss
- Time & Speed
- Ratios and Proportions
- Data Interpretation

2. Art of Learning Anything
- What is Intelligence
- Relation of success with intelligence
- Illusion of Learning
- Focussed Mode vs Diffused Mode
- Procrastination
- Improving Recall
- Creating Brain Links
- Visual memory & Data Memory
- Slow Thinking

3. Computational Thinking
- Thinking before Doing/Coding
- Problem Identification
- Decomposition
- Pattern Recognition

- Abstraction
- Algorithm Design
- Computational Thinking Use Case 1
- Computational Thinking Use Case 2

4. Technical Puzzles
- Why are Puzzles part of interviews?
- The Art of solving puzzles
- Approach more important than the solution
- Puzzles for Vertical Thinking
- Puzzles for Horizontal Thinking

Productivity and Decision Making

1. Art of being Super Productive
- Start with Why to make objectives clear
- Thinking Limitless
- The magic of computing returns
- Deciding what to work on
- Time Management Skills
- Measuring what matters
- Choosing wisely habits to inculcate

2. Effective Decision Making
- Why is decision making a key skill?
- Components of Decision Making
- Understanding common biases
- Letting emotions not clutter decision making
- Difference between quick decision making & slow decision making

Professional Communication

1. Reading comprehension & Short writing
- Building vocabulary
- Extracting insights from the textual information
- Drawing inferences from multiple stories
- Writing you inferences for others to understand
-

2. Book Reading & Writing Reviews
- Reading 10 books during the entire course & writing book reviews
- 2 Biographies
- 2 Fictions
- 6 Non-Fictions

3. Effective Understanding & Articulation
- Watching 20 movies from our suggested list
- Writing 1000 words essay on those movies
- Writing a summary of the movies

4. Group Discussion for decision making
- Understanding why GD is so important in personal & professional life
- The objective of GD - Collectively making the right decision
- 5 GD on various topics

5. Writing Professional chat/E-mail
- Writing as the most common method of professional communication
- Factors to keep in mind before starting to write
- Points to consider while writing
- Activities after writing
- Difference between chat writing & email writing

6. Making Impressive Presentation

- Why making a presentation is a professional job
- The objective of the presentation
- Attributes of good presentation
- Why research is key to the presentation
- Making a presentation interactive
- Doing 10 video/live presentation

Computer Fundamentals

1. Operating System Concepts
- Operating System Architecture
- Processes and Process Management
- Threads and Concurrency control
- Scheduling
- Memory Management
- Inter-Process Communication
- Synchronization Constructs
- I/O Management
- Resource Virtualization
- Remote Services
- Distributed Systems
- Introduction to Data Center Technologies

2. Linux Administration
- Introduction to Linux Operating Systems
- Basic Linux Commands
- File Management and Security
- The directory structure of Unix
- User Management

- Groups
- Shell types and basic commands
- Permissions
- sudo
- Systemd Services Start and Stop
- Resource Mgmt with systemctl
- Process Management (top, ps)
- Package Management(yum, apt, rpm)
- Managing disks (lsblk, df, mount, umount,du)
- File systems

3. Data Structures and Algorithms

- Built-in Data Type
    - o Integers
    - o Boolean
    - o Floating
    - o Character and Strings
- Derived Data Type
    - o List
    - o Array
    - o Stack
    - o Queue
- Linked List
    - o Singly Linked List
    - o Doubly Linked List
    - o Circular Linked List
- Array
- Stack
- Queue
- Tree
- Basic Operations
    - o Traversing
    - o Searching
    - o Sorting
    - o Hashing
    - o Insertion
    - o Deletion
    - o Merging

- Searching techniques
  - Binary search
  - Linear search
- Recursion
- Fibonacci series
- Sorting Algorithm
  - Bubble sort
  - Insertion sort
  - Selection sort
  - Quick sort
  - Merge sort
  - Bucket sort

4. Database concepts

- Introduction to Databases
- Entity Relationship Model
- Relational Model
- Relational Algebra
- Normalization
- Transactions and Concurrency Control
- DBMS Architecture 2-level 3-level
- Data Abstraction and Data Independence
- Database Objects
- Entity-Relationship Model
- Generalization
- Specialization
- Aggregation
- Entity Relationship Diagrams
- Keys in Relational Model
- Candidate key,
- Super key
- Primary key
- Alternate key
- Foreign key
- Strategies for Schema design
- Schema Integration

- Data modelling
- Star Schema in Data Warehouse modelling
- Data Warehouse Modeling

5. Basic SQL - Syntax

- Data Types
- Operators
- Expressions
- Create Database
- Drop Database
- Select Queries
- Create Table
- Drop Table
- Other Table Operations
- Insert Query
- Where Clause
- AND & OR Clauses
- Update operations
- Delete operations
- Order By clause
- Group By Clause
- Sorting operations
- SQL Constraints
- Type of Joins
- Unions Clause
- NULL Values
- Indexing
- Views

6. Software Engineering

- Software Engineering Overview
- Features of Good Software:
  - o Operational Features
  - o Transitional Features
  - **o** Maintenance Features

- Software Development:
  - Requirement Gathering
  - Software Design
  - Programming
- Software Design
  - Design
  - Maintenance
  - Programming
- Programming:
  - Coding
  - Testing
  - Integration
- Software Development Life Cycle
  - Requirement Gathering
  - System Analysis
  - Software Design
  - Coding
  - Testing
  - Integration
  - Deployment
  - Operation and Maintenance
- Types of SDLC
  - Waterfall model
  - Iterative Model
  - Spiral model
  - V Model
- Agile Concepts
- DevOps Concepts
- Microservices Architecture
- Features of Microservices Architecture
- Software Requirements
- Software Design Basics
- Analysis & Design Tools
  - Data Flow Diagram
  - Flow Chart
- Design Strategies
  - Function-Oriented Design
  - Object-Oriented Design
- User Interface Design
  - Command Line Interface(CLI)

- o   Graphical User Interface (GUI)
- Design Complexity

- Software Testing Overview

  - o   Manual Vs Automated Testing
  - o   Testing Approaches
  - o   Black-box testing
  - o   White-box testing
  - o   Unit Testing
  - o   Integration Testing
  - o   Functionality testing
  - o   Acceptance Testing
  - o   Regression Testing
- Quality Control

- Deployment Methods

  - o   Blue-Green Deployment
  - o   Rolling Deployment
- Software Monitoring

- Software Maintenance

7. Tools

- Git

  - o   What is Git?
  - o   Installing Git
  - o   First-Time Git Setup
  - o   Git Basics
  - o   Getting a Git Repository
  - o   Recording Changes to the Repository
  - o   Viewing the Commit History
  - o   Undoing Things
  - o   Working with Remotes
  - o   Tagging
  - o   Git Branching
  - o   Basic Branching and Merging
  - o   Branch Management
  - o   Branching Workflows
  - o   Remote Branches
  - o   Rebasing

- Putty

  - o   Installation
  - o   Types of connections

- o   Connecting to a remote server
- o   Using Auth keys
- o   Customizing putty

- ● Vim
  - o   Vim Basics
  - o   Insert Mode
  - o   Visual Mode
  - o   Command Mode
  - o   Create and Edit a file
  - o   Search and replace in Vim
  - o   Vim diff
  - o   Copy operations
  - o   .vimrc file
  - o   Vim Commands