Certificate Program in Data Science

The following is the detailed syllabus for the certificate program offered at our institute:

1. PYTHON FOR DATA SCIENCE

Module 1: Python Intro and Basics

- Python Intro and Get Started
- Python Syntax and Comments
- Variables and Data types
- Operators and types

Module 2: Python Data Structures

- String and String methods
- List: Object, methods, comprehensions
- Tuple: Object, methods, comprehensions
- Sets: Object, methods, comprehensions
- Dictionary: Object, methods, comprehensions

Module 3: Conditional Operators

- If Else statement
- Nested If
- For Loops
- While Loops
- Break and Continue statements

Module 4: Python Functions and Classes

- Python Functions Intro
- Creating and Calling Functions
- Types of Function arguments
- Lambda Functions
- Pass statement

Module 5: Numpy Library

- Numpy Intro
- Arrays – 1D, 2D and 3D
- Array Creation
- Array Manipulation
- Matrix Operations
- Numpy Random Module
- Arrays – Stacking, Splitting & Slicing

- Broadcasting and Boolean Mask
- Other Functions

Module 6: Pandas Library

- Pandas Intro
- Pandas Series and DataFrame
- Reading the Data
- DataFrame methods

Module 7: Python Visualization Library

- About Matplotlib, Seaborn libraries
- Plot, subplot creation
- Basic Plots: line, bar, scatter, pie charts
- multi-plot, pair plots
- Plotting with Seaborn

2. STATISTICS FOR DATA SCIENCE

- Descriptive And Inferential statistics
- Types Of Data, Sampling types
- Measures of Central Tendencies
- Data Variability: Standard Deviation
- Z-Score, Outliers, Normal Distribution
- Central Limit Theorem
- Histogram, Normality Tests
- Skewness & Kurtosis
- Understanding Hypothesis Testing
- P-Value Method, Types Of Errors
- T Distribution, One Sample T-Test
- Independent And Relational T Tests
- Direct And Indirect Correlation
- Regression Theory

3. MACHINE LEARNING

Module 1: Machine Learning Introduction

- Machine Learning Introduction
- ML core concepts
- Supervised and Unsupervised Learning
- Reinforcement Learning Intro
- Regression and Classification Algorithms`

Module 2: Linear Regression (Hands-on)

- About Linear Regression
- Mathematical Intuition
- Cost Function Minimization Overview
- Model Evaluation Metrics: MSE, MAE, RMSE etc.

Module 3: Logistic Regression (Hands-on)

- Logistic Regression – Introduction
- Sigmoid Function
- Cost Function Minimization Overview
- Model Evaluation Metrics: Accuracy, Precision, Recall etc.

Module 4: K Nearest Neighbour Classifier (KNN)

- KNN Intro and working
- Nearest Neighbour concepts
- Model Evaluation Metrics

Module 5: K Means Clustering

- Clustering Introduction
- Algorithm Working
- Cost Function Minimization Overview
- Model Evaluation Metrics

Module 6: Decision Trees

- Regression and Classification Trees
- Ensemble Techniques: Bagging and Boost
- Random Forest and XG Boost Algorithms
- Cost Function Minimization Overview
- Model Evaluation Metrics

Module 7: Naive Bayes Classifier

- Introduction to Naive Bayes
- How it works: Bayes' Theorem
- Naive Bayes For Text Classification
- Modelling and Evaluation in Python

Module 8: Support Vector Machine (SVM)

- About SVM Classifier
- Support Vectors and Hyperplanes
- Linear and Non-linear SVMs
- Kernel methods in SVM

Module 9: Artificial Neural Network

- Introduction to ANN
- Modelling and Evaluation of ANN in Python
- Image Classification using CNN

Module 10: Time Series Forecasting

- Time Series Intro and Data
- Trend, Seasonality, cyclical and random
- Autoregressive Model (AR)
- Moving Average Model (MA)
- Stationarity of Time Series
- ARIMA Model
- Autocorrelation and AIC

Module 11: Principal Component Analysis (PCA)

- About PCA
- Working and Methodology of PCA
- Modelling PCA in Python

Module 12: Other ML Concepts

- Adv. Metrics (Roc_Auc, R2, Precision, Recall)
- K-Fold Cross-validation
- Hyper Parameter Tuning
- Grid And Randomized Search CV
- Balancing Imbalanced Data Set
- Feature Selection Techniques

Module 13: Feature Engineering

- Introduction to Feature Engineering
- Transforming Predictors
- Feature Selection methods
- Backward elimination technique
- Feature importance from ML modelling

Module 14: ML Model Deployment

- Introduction to Flask, Streamlit
- URL and App routing
- Flask, Streamlit application – ML Model Deployment

4. GIT FOR DATA SCIENTISTS

Module 1: GitHub Repo

- About GitHub and Repo
- Creating GitHub Account
- Create a new Git Repo
- Copy existing repo
- Various Functionalities

Module 2: Git commands

- Git pull
- Code commits
- Push to remote repo
- Checkout branch
- Merge branches, Pull requests
- Other commands

Module 3: GIT

- Purpose of Version Control
- GIT Introduction
- Terminologies
- Git workflow

5. PROJECT WORK

The project work will be on real data set. The learners can choose their project theme from the available options.