DATA SCIENCE PROFESSIONAL

SYLLABUS

www.gteceducation.com

G-TEC VIRTUAL UNIVERSITY

www.gteccollege.com

www.gensmartacademy.com

G-TEC EDUCATION

Making Professionals Globally

Course Contents:

1. Python and Statistics for Data Science

Basics of Python
Fundamentals of Statistics
Probability
Linear Algebra
Calculus

2. Data Analysis and Visualization with Python

Introduction to NumPy

NumPy Arrays
Mathematical operations in NumPy
NumPy Array manipulation
NumPy Array broadcasting

Data Manipulation with Pandas

Data Structures in Pandas - Series and Data Frames
Data cleaning in Pandas
Data manipulation in Pandas
Handling missing values in datasets

Data Visualization

Visualization with Python
Plotting basic charts in Python
Data visualization with Matplotlib
Statistical data visualization with Seaborn

3. Machine Learning

Introduction to Machine Learning (ML)

What is Machine Learning ?
Use Cases of Machine Learning
Types of Machine Learning - Supervised, Unsupervised, Reinforcement
Machine Learning workflow

Supervised Learning

Regression

Multi Linear Regression

Introduction to Linear Regression
Use cases of Linear Regression
Fitting a Linear Regression model
Evaluating and interpreting results from Linear Regression models

Classification

Logistic Regression

Introduction to Logistic Regression
Logistic Regression use cases

Understand use of Sigmoid function to perform logistic regression.

Model Evaluation Techniques

Introduction to evaluation metrics and model selection in Machine Learning
Importance of Confusion matrix for predictions
Measures of model evaluation - Sensitivity, specificity, precision, recall & f-score
Use ROC curve to decide best model

Decision trees & Random Forests

Introduction to Decision Trees & Random Forest
Understanding criterion (Entropy & Information Gain) used in Decision Trees
Using Ensemble methods in Decision Trees
Applications of Random Forest.

Support vector machines (SVM)

Introduction to SVM
Figure decision boundaries using support vectors
Identify hyperplane in SVM
Applications of SVM in Machine Learning

Unsupervised Learning

Clustering

K-Means

Introduction to K-means clustering
Decide clusters by adjusting centroids
Find optimal 'k value' in kmeans
Applications of clustering in Machine Learning

Recommendation Systems

KNN (K- Nearest neighbors)

Introduction to KNN
Calculate neighbors using distance measures
Find optimal value of K in KNN method
Advantage & disadvantages of KNN

Dimensionality Reduction

Introduction to Curse of Dimensionality
What is dimensionality reduction?
PCA to reduce dimensions
Applications of Principle component Analysis (PCA)

4. Deep Learning Foundation

5. Introduction to Computer Vision

6. Introduction to Natural Language Processing