PG Program in Data Science

Course Information

Duration
12 Months

Mode of Delivery
Online

General Information

Description

The curriculum, learning content, and projects are designed and created by Careerera faculties in coordination with leading Industry experts. This specialization program helps you to build extra skill sets in Data Science and Advanced Machine Learning thereby making you a future-ready technology professional.

Learning Goals

- Students will develop relevant programming abilities.
- Students will demonstrate proficiency with statistical analysis of data.
- Students will develop the ability to build and assess data-based models.
- Students will execute statistical analyses with professional statistical software.
- Students will demonstrate skills in data management.

MODULE 01 : PRE-PROGRAM CONTENT

1. INTRODUCTION TO ANACONDA ENVIRONMENT AND INSTALLATION

WEEK 0
1 Hour

---

2. WORKING ENVIRONMENTS

WEEK 0
2 Hours

- Jupyter Notebook
- Kaggle
- Google Colab
- MySQL Workbench

---

3. INTRODUCTION TO:

WEEK 0
2 Hours

- Data Science
- Machine Learning

- Artificial Intelligence
- Deep Learning

---

4. THE CRISP-DM FRAMEWORK

WEEK 0
2 Hours

- Business and Data Understanding
- Data Preparation
- Modelling
- Evaluation
- Deployment

MODULE 02 : PYTHON & ITS LIBRARIES FOR DATA SCIENCE

1. GETTING STARTED WITH PYTHON

WEEK 1-4
24 Hours

- Syntax, variables and Data types
- Data structures in python – List, Tuple, Set, Dictionary
- Conditional statements – if, Else, Elif
- Loops – While & For Loops
- Functions – lambda, recursions, map, filter& reduce functions
- OOP – Class, Objects, Inheritance, Polymorphism, Encapsulation, Abstraction, Generators, Iterators
- Programming in Python – Questions Practice
- Doubts Clearing/Assessment

PYTHON PROGRAMMING ASSIGNMENT

- Problem Statement
- Final Submission
- Solution

---

2. PYTHON FOR DATA SCIENCE

WEEK 5-7
18 Hours

- Introduction to NumPy
- Introduction to Pandas
- Introduction to Matplotlib
- Introduction to Seaborn
- Data Visualization using Seaborn & Matplotlib
- Doubts Clearing/Assessment

---

3. EXPLORATORY DATA ANALYSIS

WEEK 8-9

- Data Sourcing
- Data Cleaning – null and infinite values, outliers, capping, sanity checks, data formatting, etc.
- Types of Variables – Categorical & Continuous
- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
- Doubts Clearing/Assessment

12 Hours

EDA CASE STUDY ASSIGNMENT

- Problem Statement
- Final Submission
- Solution

MODULE 03 : DATA ANALYSIS USING SQL

4. DATA ANALYSIS USING SQL

WEEK 10-12

18 Hours

- Introduction to DBMS & RDBMS
- OLAP vs OLTP
- Database Design
- Database creation in MYSQL Workbench
- DDL and DML statements
- MySQL Data Types &Clauses
- Retrieve Data from Single Tables
- Retrieve and Transform Data From Multiple Tables Using JOINS and SET Operations
- Grouping and Summarizing Records
- Creating Views with Aggregate Functions
- Writing Subqueries
- Doubt Clearing/Assessment

SQL ASSIGNMENT

- Problem Statement
- Final Submission
- Solution

MODULE 04 : QUANTITATIVE RESEARCH & STATISTICS FOR DATA SCIENCE

5. STATISTICS

WEEK 13-14
12 Hours

- Introduction to Statistics
- Basics of Probability
- Discrete Probability Distribution
- Continuous Probability Distribution
- Normal Distribution
- Poisson's Distribution
- Bayes' Theorem
- Central Limit Theorem
- Doubt Clearing/Assessment

STATISTICS ASSIGNMENT

- Problem Statement
- Final Submission
- Solution

---

6. HYPOTHESIS TESTING

WEEK 15-16
12 Hours

- Concepts of Hypothesis Testing: Null and Alternate Hypothesis
- Making a Decision and Critical Value Method
- p-Value Method and Types of Errors
- One-Sample T-Test, Two Sample T-Test
- Z-Test, ANOVA, Chi-Square, A/B Testing
- Pearson Co-Relation, Co-Variance, Chebyshev-Inequality Formula
- Doubt Clearing/Assessment

HYPOTHESIS TESTING ASSIGNMENT

- Problem Statement
- Final Submission
- Solution

MODULE 05 : MACHINE LEARNING(ML) FOR DATA SCIENCE & ANALYTICS

1. SUPERVISED LEARNING - REGRESSION

WEEK 17-18
15 Hours

- Simple Linear Regression
- Multiple Linear Regression
- Performance Metrics: Accuracy Scores, R^2 And Adjusted R^2 To Compare the Model with Different Numbers of Independent Variables.
- Approaches to Feature Selection: Univariate Selection, Feature Importance, RFE
- Parameter Tuning and Model Evaluation
- Data Transformation and Normalization
- Ridge & Lasso Regression (L1 & L2)
- Doubts Clearing/Assessment

LINEAR REGRESSION ASSIGNMENT

- Problem Statement
- Final Submission
- Solution

---

2. LOGISTIC REGRESSION

WEEK 19
6 Hours

- Concept of Logistic Regression
- Univariate Logistic Regression
- Multivariate Logistic Regression: Model Building and Evaluation
- Dealing with Categorical Independent Variable – One Hot Encoding Vs Dummy Variable
- Doubt Clearing/Assignment

LOGISTIC REGRESSION ASSIGNMENT

- Problem Statement
- Final Submission
- Solution

---

3. DECISION TREES

WEEK 20
6 Hours

- Concept of Decision Trees
- Importance and Usage of Ginny And Entropy
- Information Gain
- Visualizing Decision Trees Nodes and Splits

- Working of The Decision Tree Algorithm
- Evaluating Decision Trees Models: Accuracy, Precision, Recall, Confusion Matrix
- Doubt Clearing/Assignment

DECISION TREES ASSIGNMENT

- Problem Statement
- Final Submission
- Solution

---

4. CLASSIFICATION TECHNIQUES – CART

WEEK 21
6 Hours

- Extending Decision Trees to Regressing Problems
- Advantages of Using CART
- The Bayes Theorem
- KNN Classifier
- The Gaussian Naïve's Bayes Classifier – Assumptions of The Naïve Bayes Classifier, Functioning of The Naïve's Bayes Algorithm
- Doubts Clearing/Assessment

---

5. SUPPORT VECTOR MACHINE(SVM)

WEEK 22
6 Hours

- What is Support Vector Machine?
- How does SVM Work?
- Different Types of SVM
- SVM Kernels
- SVM Use Cases
- How to Implement SVM?
- Doubts Clearing/Assessment

SVM ASSIGNMENT

- Problem Statement
- Final Submission
- Solution

---

6. UNSUPERVISED LEARNING: CLUSTERING

WEEK 23
6 Hours

- Introduction to Clustering
- K-Means Clustering: Linkage, Use of Elbow Curve & Silhouette Score
- Hierarchical Clustering – Agglomerative &

Divisive, Distance Matrix, Dendrogram
- Doubts Clearing/Assessment

CLUSTERING ASSIGNMENT

- Problem Statement
- Final Submission
- Solution

---

7. PRINCIPAL COMPONENT ANALYSIS(PCA)

WEEK 24
6 Hours

- Noise in The Data and Dimensionality Reduction
- Capturing Variance – The Concept of Principal Components
- Assumptions in Using PCA
- Eigenvectors and Orthogonality of Principal Components
- What Is the Complexity Curve?
- Advantages of Using PCA
- The Working of The PCA Algorithm
- Singular Value Decomposition
- Doubts Clearing/Assessment

PCA ASSIGNMENT

- Problem Statement
- Final Submission
- Solution

---

8. ENSEMBLE MODELLING

WEEK 25-26
12 Hours

- Popular Ensembles
- Bagging
- Bootstrap
- Introduction to Random Forests
- Feature Importance in Random Forests
- Boosting
- How Boosting Algorithm Works?
- Adaboost
- Doubts Clearing/Assessments

ENSEMBLING MODELLING ASSIGNMENT

- Problem Statement

- Final Submission
- Solution

ML CASE STUDY ASSIGNMENT

- Overview of The Domain and Associated Concepts
- Problem Statement
- Evaluation Rubric
- Final Submission

MODULE 06 : BUSINESS ANALYSIS WITH TABLEAU

TABLEAU BASICS

WEEK 27
6 Hours

- Introduction to Data Visualization and Tableau
- Tableau Products Suite
- Data Types
- Tableau File Types
- Connecting to Data Sources
- Handling R Data
- Connecting to MS Access Database
- Loading & Reshaping Data Aggregation
- Working with Continuous and Discrete Data Using Filters
- Working with Dates
- Using Calculated Fields & Parameters
- Creating Tables and Charts
- Building Dash Boards and Storyboards
- Sharing Your Work and Publishing for A Wider Audience

TABLEAU ASSIGNMENT

- Problem Statement
- Final Submission
- Solution

CAPSTONE PROJECT PHASE

- Overview of The Domain and Associated Concepts
- Problem Statement
- Evaluation Rubric
- Final Submission

MODULE 07 : AI & DEEP LEARNING

1. TIME SERIES ANALYSIS

WEEK 28-30
18 Hours

- What is Time Series?
- Regression vs Time Series
- Examples of Time Series data
- Trend, Seasonality, Noise and Stationarity
- Time Series Operations
- Detrending
- Successive Differences
- Moving Average and Smoothing
- Exponentially weighted forecasting model
- Lagging
- Correlation and Auto-correlation
- Holt Winters Methods
- Single Exponential smoothing
- Holt's linear trend method
- Holt's Winter seasonal method
- ARIMA and SARIMA
- Doubts Clearing/Assessment

TIME SERIES ANALYSIS ASSIGNMENT

- Problem Statement
- Final Submission
- Solution

---

2. NATURAL LANGUAGE PROCESSING AND SPEECH RECOGNITION

WEEK31-32
12 Hours

- REGEX and Introduction To NLP
- Feature Engineering on Text Data
- Natural Language Understanding Techniques
- Natural Language Generation
- Natural Language Processing Libraries
- Natural Language Processing with Machine Learning and Deep Learning
- Basic Lexical Processing
- Advanced Lexical Processing
- Introduction to Speech Recognition
- Signal Processing and Speech Recognition Models
- Speech to Text
- Text to Speech
- Voice Assistant Devices

- Doubts Clearing/Assessment

NLP ASSIGNMENT

- Problem Statement
- Final Submission
- Solution

---

3. TEXT MINING AND SENTIMENTAL ANALYSIS

WEEK 33-34
12 Hours

- Text Cleaning, Regular Expressions, Stemming, Lemmatization
- Word Cloud, Principal Component Analysis, Bigrams & Trigrams
- Web Scrapping, Text Summarization, Lex Rank Algorithm
- Latent Dirichlet Allocation (LDA) Technique
- Word2vec Architecture (Skip Grams Vs CBOW)
- Text Classification, Document Vectors, Text Classification Using Doc2Vec
- Doubts Clearing/Assessment

TEXT MINING ASSIGNMENT

- Problem Statement
- Final Submission
- Solution

---

4. REINFORCEMENT LEARNING

WEEK35
6 Hours

- Introduction to Reinforcement Learning
- Reinforcement Learning Framework and Elements
- Multi-Arm Bandit
- Markov Decision Process
- Solution Methods
- Q-value and Advantage Based Algorithms

REINFORCEMENT LEARNING ASSIGNMENT

- Problem Statement
- Final Submission
- Solution

5. INTRODUCTION TO AI & DEEP LEARNING

WEEK 36-37
12 Hours

- Artificial Neural Network
- Deep Neural Network and Tools
- Deep Neural Net Optimization, Tuning & Interpretability
- Convolutional Neural Net(CNN)
- Famous CNN Architectures
- Recurrent Neural Networks
- Autoencoders
- Doubts Clearing/Assessment

6. ADVANCED DEEP LEARNING & COMPUTER VISION

WEEK38-40
12 Hours

- Course Introduction
- RBM and DBNS
- Variational Auto Encoder
- Working with Deep Generative Models
- Applications: Neural Style Transfer and Object Detection
- Distributed & Parallel Computing for Deep Learning Models
- Transfer Learning
- Object Detection
- Semantic Segmentation
- Instance Segmentation
- Other Variants of Convolution
- Metric Learning
- Siamese Networks
- Triplet Loss
- Doubts Clearing/Assessments

AI & DEEP LEARNING ASSIGNMENT

- Problem Statement
- Final Submission
- Solution

MODULE 08 : BUSINESS ANALYSIS WITH MS POWER BI & GOOGLE DATA STUDIO

1. BI TOOLS: GOOGLE DATA STUDIO

WEEK 41
6 Hours

- Introduction to Google Data Studio
- Data Types, Sources, Connections
- Loading, Reshaping, Data Aggregation
- Working with Continuous and Discrete Data
- Report Edit Mode in Data Studio
- Using Filters in Data Studio
- Using Calculated Fields and Parameters
- Creating Tables and Charts
- Building Dash Boards and Story Boards in Data Studio
- Doubts Clearing/Assessment

---

2. BI TOOLS: POWER BI

WEEK 42
6 Hours

- Introduction to Power BI
- The Key Features of Power BI Workflow
- Desktop Application BI Service
- Sourcing Data from The Web (OData And Azure)
- Building A Dashboard
- Data Visualization
- Publishing to The Cloud
- DAX Data Computation
- Row Context, Filter Context
- Analytics Pane, Creating Columns and Measures
- Data Drill Down and Drill Up
- Creating Tables, Binned Tables
- Data Modeling and Relationships
- Power BI Components: Power View, Map, Query and Pivot
- Doubt Clearing/Assessment

MODULE 09 : R PROGRAMMING FOR DATA SCIENCE

GETTING STARTED WITH R

WEEK 43
6 Hours

- Introduction to R Programming
- How to Install R
- Packages In R
- Data Types In R
- Data Structures
- Conditional Statements
- Loops
- Functions
- Lists
- Arrays

- Data Frames
- R Plot
- R Line
- R Scatterplot
- R Pie Charts
- R Bars
- Doubts Clearing/Assessments

LIVE INDUSTRY GRADED PROJECT

- Overview of The Domain and Associated Concepts
- Problem Statement
- Evaluation Rubric
- Final Submission

EMPLOYABILITY ENHANCEMENT PROGRAM

WEEK44-45
12 Hours

- Training Need Analysis
- Communication Skills
- Placement Skills
- Behavioral Science
- Resume Building
- Corporate Etiquette
- Personality Enhancement, Etc.

DISCLAIMER:

Modules may change in order to keep up with the demand and developing issues in the relevant discipline or to reflect the research interests of teaching staff in current affairs.