INSAID

GLOBAL CERTIFICATE IN DATA SCIENCE

Online | 6 months | 6 Terms

PROGRAM SNAPSHOT

PRE-TERM PREP
- Python and Stats for Data Science

TERM 01
- Data Analysis with Python

TERM 02
- Machine Learning - I

TERM 03
- Data Visualization with Tableau

PROJECT
- Capstone Project - I

TERM 04
- Machine Learning - II

TERM 05 [ELECTIVE]
- Machine Learning - III [E-1]
- Data Analytics with R [E-2]
- Deep Learning Foundation [E-3]

TERM 06
- Capstone Project - II & Industry Immersion

INSAID

TERM WISE SYLLABUS

Term 1:

DATA ANALYSIS WITH PYTHON

Module 1 : Data Science Fundamentals

- Thought Experiment : Data science @ Amazon and Target
- Introduction to Data Science
- Real world use-cases of Data Science
- Walkthrough of data types
- Data Science project lifecycle

Module 2 : Introduction to Numpy

- Basics of Numpy Arrays
- Mathematical operations in Numpy
- Numpy Array manipulation
- Numpy Array broadcasting

Module 3 : Data manipulation with Pandas

- Data Structures in Pandas-Series and DataFrames
- Data cleaning in Pandas
- Data manipulation in Pandas
- Handling missing values in datasets
- Hands-on: Implement Numpy arrays and Pandas Dataframes

Module 4 : Data Visualization in Python

- Plotting basic charts in Python
- Data visualization with Matplotlib
- Statistical data visualization with Seaborn
- Interactive data visualization with Bokeh
- Hands-on: Coding sessions using Matplotlib, Seaborn, Bokeh packages

Module 5 : Exploratory Data Analysis - 1

- Introduction to Exploratory Data Analysis (EDA) steps
- Plots to explore relationship between two variables
- Histograms, Box plots to explore a single variable
- Heat maps, Pair plots to explore correlations
- Case study: Perform EDA to explore survival using titanic dataset

Module 6 : Exploratory Data Analysis - 2

- Case study: Identify stock price patterns of Google & Apple
- Case study: Perform EDA to analyze spread of tuberculosis

Term Projects

NYC Flight data

Analyse flight delays from airports in NYC

The dataset contains 300k+ observations of flights from NYC airports in 2013

European Premier league data

Analyse key sucess factors of top football teams at European Premier League

The dataset contains 17 variables and over 400 observations on Premier league football from years 2012-16

Wine quality data

Analyse factors that determine best wine quality from France

The data set contains 12 variables and over 4000 observations on varieties of red wine.

Automobile pricing data

Analyse factors affecting automobile pricing

The dataset contains 25 variables and over 200 observations on performance & pricing of automobiles in USA

Facebook activity data

Analyse Facebook data to understand hidden patterns in social activity

The dataset contains 15 variables and 100K observations on pseudo Facebook activity data

Term 2:

MACHINE LEARNING - I

Module 1 : Introduction to Machine Learning (ML)

- What is Machine Learning ?
- Use Cases of Machine Learning
- Types of Machine Learning - Supervised to Unsupervised methods
- Machine Learning workflow

Module 2 : Linear Regression

- Introduction to Linear Regression
- Use cases of Linear Regression
- How to fit a Linear Regression model?
- Evaluating and interpreting results from Linear Regression models
- Case study:Predict Bike sharing demand

Module 3 : Logistic Regression

- Introduction to Logistic Regression
- Logistic Regression use cases
- Understand use of odds & Logit function to perform logistic regression
- Case study:Predicting credit card default cases

Module 4 : 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
- Case study:Predict passenger survival using Titanic Data set

Module 5 : 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 AUC-ROC curve to decide best model
- Case study: Applying model evaluation techniques to Titanic dataset

Term Projects

Air Quality Analysis

Predict Relative Humidity

The dataset contains 15 variables and over 5000 observations on weather conditions

Loan Payment data

Predict customer loan default class

The dataset contains 11 variables and over 300 observations on bank customers

Mercedes-Benz Greener Manufacturing

Predict the time it takes to pass testing

The dataset contains 386 variables & 4210 observations

Wine Quality data

Predict wine quality

The dataset contains 12 variables and over 1500 observations on wine quality

Predict whether or not a person is diabetic

Predict positive & negative case for diabetes

The dataset contains 9 variables and 769 observations on suspected paitents

Term 3:

DATA VISUALIZATION WITH TABLEAU

Module 1 : Introduction to Visual Analytics

- Introduction to data visualization
- Understanding Tableau ecosystem in industry
- Loading data files in Tableau
- Creating first visualizations
- Case Study: Sales performance Analysis

Module 2 : Data Visualization using Tableau

- Introduction to graphs - bar graph and line graph
- Working with continuous measures & discrete variables
- Heat maps and Geographical data visualizations
- Creating map Views
- Case Study: Analyse Earthquake data from 1900 till 2014

Module 3 : Data joining & blending in Tableau

- Introduction to SQL joins
- Performing data blending in Tableau
- Creating dual axis charts in Tableau
- Introduction to descriptive statistics and Visual analytics
- Case Study : Analyse revenue trends in Retail businesses

Module 4 : Predicitve Analytics using Tableau and R

- Introduction to R programming tool & R studio
- Installing R and R studio
- Applications of linear regression in prediction
- Data crunching: Creating groups, sets & parameters
- Case Study: Forecast revenues in Retail Scenario

Module 5 : Interactive Dashboard Design

- Introduction to principles of dashboard design
- Custom geocoding in Tableau
- Developing dashboard products using Tableau
- Introduction to writing storyline in Tableau
- Case Study: Build banking customer segmentation dashboard

Module 6 : Advanced Calculations using Tableau	• Introduction to calculations: Date calculations • Using LOD calculations: INCLUDE, EXCLUDE & FIXED functions • Working with Table calculations • Exporting data from Tableau • Case Study: Analyse Retail sales across geographies, products & customers
Module 7 : Applications of advanced Calculations using Tableau	• Introduction to customer churn analysis • Estimating customer life time value • Applications of context filtering • Applications of logical functions in Tableau • Case Study: Analyse retail sales data to predict customer behaviour
Module 8 : Revision of concepts and Project discussion	• Revision of key concepts: data blending, writing calculations, LOD calcuations etc. • Review of Tableau project portfolio • Communicating data insights using reporting tools • Tableau Interview prep • Discussing EDA objectives of final project

Term Projects

Hubway data visualization challenge

Produce visualizations that reveal interesting user patterns about how people in Boston gets around on Hubway

The dataset contains 1 million observations on bike usage by residents of Boston

Term 4:

MACHINE LEARNING - II

Module 1 : Dimensionality Reduction using PCA	• Unsupervised Learning: Introduction to Curse of Dimensionality • What is dimensionality reduction? • Technique used in PCA to reduce dimensions • Applications of Principle component Analysis (PCA) • Case study: Optimize model performance using PCA on SPECTF heart data
Module 2 : KNN (K- Nearest neighbours)	• Introduction to KNN • Calculate neighbours using distance measures • Find optimal value of K in KNN method • Advantage & disadvantages of KNN • Case Study:Classify phishing site data using close neighbour technique
Module 3 : Naïve Bayes classifier	• Introduction to Naïve Bayes classification • Refresher on Probability theory • Applications of Naive Bayes Algorithm in Machine Learning • Case study : Classify spam emails based on probability
Module 4 : K-means clustering technique	• Introduction to K-means clustering • Decide clusters by adjusting centroids • Find optimal 'k value' in kmeans • Understand applications of clustering in Machine Learning • Case study : Segment hands in Poker data and segment flower species in Iris flower data
Module 5 : Support vector machines (SVM)	• Introduction to SVM • Figure decision boundaries using support vectors • Identify hyperplane in SVM • Applications of SVM in Machine Learning • Case Study : Predicting wine quality using SVM

Module 6 : Time series forecasting

- Introduction to Time Series analysis
- Stationary vs non stationary data
- Components of time series data
- Interpreting autocorrelation & partial autocorrelation functions
- Stationarize data and implement ARIMA model
- Case Study: Forecast demand for Air passengers

Term Projects

Predict burned area of forest

The dataset contains 518 observations & 13 variables regarding forest fire in the north-east region of Portugal

Children Killed & injured in Gun Violence in the USA

The data set contains 6 attributes with 501 instances of Children killed & injured across all the states of the USA.

Predict whether or not a person is diabetic

The dataset contains 9 variables and 769 observations on suspected paitents

Predict Adults' income whether or not <=50k or >50k

The dataset has 15 attributes & 32562 instances

Supermarket Purchase Analysis

The dataset contains 6 variables identifying shopping behaviour of a segment of customers.

Term 5 [Elective]:

MACHINE LEARNING - III

Module 1 : Introduction to Apriori Algorithm

- Applications of Apriori algorithm
- Understand Association rule
- Developing product recommendations using association rules
- Case study:Analyse online retail data using association rules

Module 2 : Recommender Systems

- Introduction to Recommender systems
- Types of Recommender systems - collaborative, content based & Hybrid
- Types of similarity matrix (cosine , Jaccard, Pearson correlation)
- Case Study:Build Recommender systems on Movie data using KNN basics

Module 3 : Linear Discriminant Analysis (LDA)

- Recap of dimensionality reduction concepts
- Types of dimensionality reduction
- Dimensionality reduction using LDA
- Case Study : Apply LDA to determine Wine Quality

Module 4 : Anomaly Detection

- Introduction to Anomaly detection
- How Anomaly detection works?
- Types of Anomaly detection: Density based, Clustering etc.
- Case Study:Detect anomalies on electrocardiogram data

Module 5 : Ensemble learning

- Introduction to Ensemble Learning
- What are Bagging and Boosting techniques?
- What is Bias variance trade off?
- Case study : Predict wage (annual income) classes from adult census data

Module 6 : Stacking

- Introduction to stacking
- Use Cases of stacking
- How stacking improves machine learning models?
- Case Study:Predict survivors in Titanic case

Module 7 : Optimization

- Introduction to optimization in ML
- Applications of optimization methods
- Optimization techniques: Linear Programming using Excel solver
- How Stochastic Gradient Descent(SGD) Works?
- Case study: Apply SGD on Regression data (sklearn dataset)

Module 8 : Neural Networks

- Introduction to Neural networks
- What are Perceptrons & Types of Perceptrons?
- Workflow of a Neural network & analogy with biological neurons
- Case Study : Apply computer vision for digit recognition on MNIST data

Term Projects

Recommend groceries on e-commerce platforms

This data contains 10 variables and over 1000 observations on online grocery purchases

Predict sales prices in real estate market

The train dataset contains 80 variables & over 1000 obervatons& test data contains 79 variables & over 1000 observations on Iowa city real estate

Predict customer satisfaction of Santander Bank

The high dimensional train dataset contains 76021 obervations & test data contains 75819 observations(Variables given are anonymous)

Predict level of risk associated with insurance

The dataset contains 124 attributes & 59382 instances

Detect unusual Credit Card transaction

The dataset contains 31 attributes & 284808 instances

Term 5 [Elective]:

DATA ANALYTICS WITH R

Module 1 : Data Science Fundamentals

- Thought Experiment: Data science @ Google
- Introduction to Data Science
- Real world use-cases of Data Science
- Walkthrough of data types
- Data Science project lifecycle

Module 2 : Introduction to programming in R

- Installing R and R Studio
- Basic Commands in R
- Installing packages
- Setting working directory
- Exercises: Basic exercises in R Programming

Module 3 : Playing around with Data objects in R

- Data structures
- Basic Data management
- Loops and Functions
- Saving output
- Exercises: Loops and functions in R

Module 4 : Descriptive statistics - 1

- Introduction to Statistics
- Descriptive Statistics
- Measures of central tendency
- Measures of Dispersion and shape
- Case Study: Investigation of Crime statistics in Beaufort

Module 5 : Descriptive statistics - 2

- Introduction to Probability
- Probability Distributions used in Data Science
- Quantiles, percentiles, and standard score
- Case Study : Analyse student's performance at school

Module 6 : Inferential Statistics - 1	• Introduction to Inferential Statistics • Population and Samples • Central Limit Theorem • Case Study: Sampling data for Business analysis
Module 7 : Inferential Statistics - 2	• Introduction to Hypothesis Testing • Confidence Intervals • Tests of significance: p-value • Case Study: Apply Inferential statistics & Central limit theorem using Python
Module 8 : Intermediate R: Importing data	• Loading data from R libraries • Importing data from Excel and CSV files • Connecting SQL databases • Webscraping using R • Case study: Webscraping websites using scrapy package
Module 9 : Intermediate R : Data Manipulation using Tidyverse	• Identifying NULL values in datasets • Introduction to data imputation methods • Creating new variables and recoding variables • Type conversions • Case Study: Using Tidyverse in Data Manipulation
Module 10 : Intermediate R: Restructuring Data	• Managing Date values • Numerical and Character functions • Aggregating & Restructuring data • Sorting, merging datasets • Exercises : Subsetting datasets for use in Predictive analytics

Module 11 : Intermediate R : Exploratory data analytics using ggplot2

- Introduction to basic graphs: Barplots, Scatterplots & line graphs
- Using Boxplots in univariate analysis
- Applications of Histograms
- Using ggplot2 for advanced visualizations

Term Projects

Kickstarter funding data

Derive insights from successful and failed projects on Kickstarter platform

The dataset contains 15 variables and around 400,000 observations

Air Quality Analysis

Derive insights on air quality in metropolitan cities

The dataset contains 6 variables and over 100 observations

Retail banking data

Derive insights on direct marketing campaigns of a Portuguese Bank

The dataset contains 17 variables and over 45000 observations

Characters in superhero comics

Derive insights on the nature of characters in Marvel comics

The dataset contains 11 variables and over 20000 observations

International crisis behaviour

Derive insights on factors which caused major international crisis events in the last 100 years

The dataset contains 96 Variables and over 1000 observations

Term 5 [Elective]:

DEEP LEARNING FOUNDATION

Module 1 : Artificial Intelligence

- Introduction to Artificial Intelligence
- Breakthroughs in the field of AI
- Overview of advanced Machine Learning algorithms
- Weights and Bias estimation using gradient descent optimization

Module 2 : Getting started with Tensorflow

- Installing Tensorflow in Python
- Introduction to data flow graphs in Tensorflow
- Functions, operations and execution pipeline in Tensorflow
- Regression technique in Tensorflow
- Case Study: Predict Boston Housing Prices using Tensorflow

Module 3 : Tensorflow programming in Python

- Classification in Tensorflow
- Introduction to Tensorboard visualization
- Activation functions in Tensorflow

Module 4 : Introduction to Deep Learning

- An overview on Deep Neural Networks
- Real world applications of Deep Neural Networks
- Neural Networks using Tensorflow
- Optimization techniques employed in Neural Networks
- Case Study: Classify handwritten digits (MNIST) using logistic regression

Module 5 : Optimization of Deep Neural Networks

- Hyperparameters in deep neural networks
- Filters in Convolutional Neural Networks
- Max pooling and padding
- Dropout and Regularization in Deep learning

Module 6 : Introduction to Convolutional Neural Networks	• Introduction to Convolutional Neural Networks(CNN's) • Evaluate, Improve and tune Convolutional Neural Networks • Object Classification, localization and segmentation • Reusing models with Transfer learning • Case Study: Classify handwritten digits (MNIST) using Deep Neural networks with RELU on Keras
Module 7 : Introduction to Natural Language Processing	• Introduction to NLP • Introdution to Word embeddings • Simple Word Vector representations: word2vec, GloVe • Implementation of word2vec model in Keras
Module 8 : Recurrent Neural networks	• Intoduction to Recurrent neural networks based language models • Introduction to Gated Recurrent units • LSTMs for machine translation • Case Study: Perform Sentiment Analysis using word embedding Seq2Seq LSTM Model translation
Module 9 : Recursive neural networks for Sentiment Analysis	• Introduction to Sequence to sequence learning • Convolutional Neural networks for Sentence classification • Train Recursive Neural Networks for Sentiment analysis • Introduction to Dynamic memory networks • Case Study: Explore Pride and Prejudice book to perform Char-RNN
Module 10 : Revision of concepts and Term Project	• Project: Bulid a Chatbot using Slack Class • Integrate Chatbot with Bot server

Term Projects

Build Chatbot using slack Class

Develop a chatbot to simulate a human as a conversational partner

This chatbot allows us translate user submitted conversations from English to Hindi

Term 6:

CAPSTONE PROJECT - I

James Telco Bond

In this capstone project, students will be provided with data collected by a major Telecom operator on the demographic behaviour of users using different handsets.

Students are required to do the initial bit of data cleansing, pre-processing and then upload this data to SQL server via a web hosting platform that will be provided to them.

This data from SQL server will be used to create a dashboard for the company using D3.js scripts. D3,js scripts will be provided to students upfront. These dashboards are reflective of how interactive visualizations can help companies make strategies such as what demographies to cater to, how men and women customers behave differently, which geographies are popular and ones that need more investment from the company in terms of finance and marketing?

Term 6:

CAPSTONE PROJECT - II

Demand Planners

This capstone project will focus more on applying machine learning concepts rather than data gathering and storing aspects. Students will be provided with data collected by a major Taxi Aggregator of taxi bookings done in a leading city. As budding data science consultants, students are required to do exploratory data analysis & present an initial report.

After that the students are required to create an UI that displays the observations regarding taxi usage across the city from the analysis and the website should also have a provision for the company to forecast demand for taxis at a specific time in the day.

The taxi bookings data provided will be in csv format and dashboards for the company need to be created using D3.js scripts. The D3.js scripts will be provided to the students beforehand.

Need to know

Program Start

August 2018

Duration

- 06 months (Incl. Capstone Projects)

Prerequisite

- Background in Programming (Not Mandatory)
- Laptop with 4 GB RAM

Program Fee

INR 2 Lakh + GST

Scholarships

50 scholarships (each with 70% tuition waiver) for professionals passionate about making a career in Data Science & furthering INSAID's mission of putting India on the global AI map.

Talk to our Admissions team today or **attend the next Data Science MasterClass** to know more

For Further details, write to us info@insaid.co