INSAID

6 MONTH CERTIFICATE IN

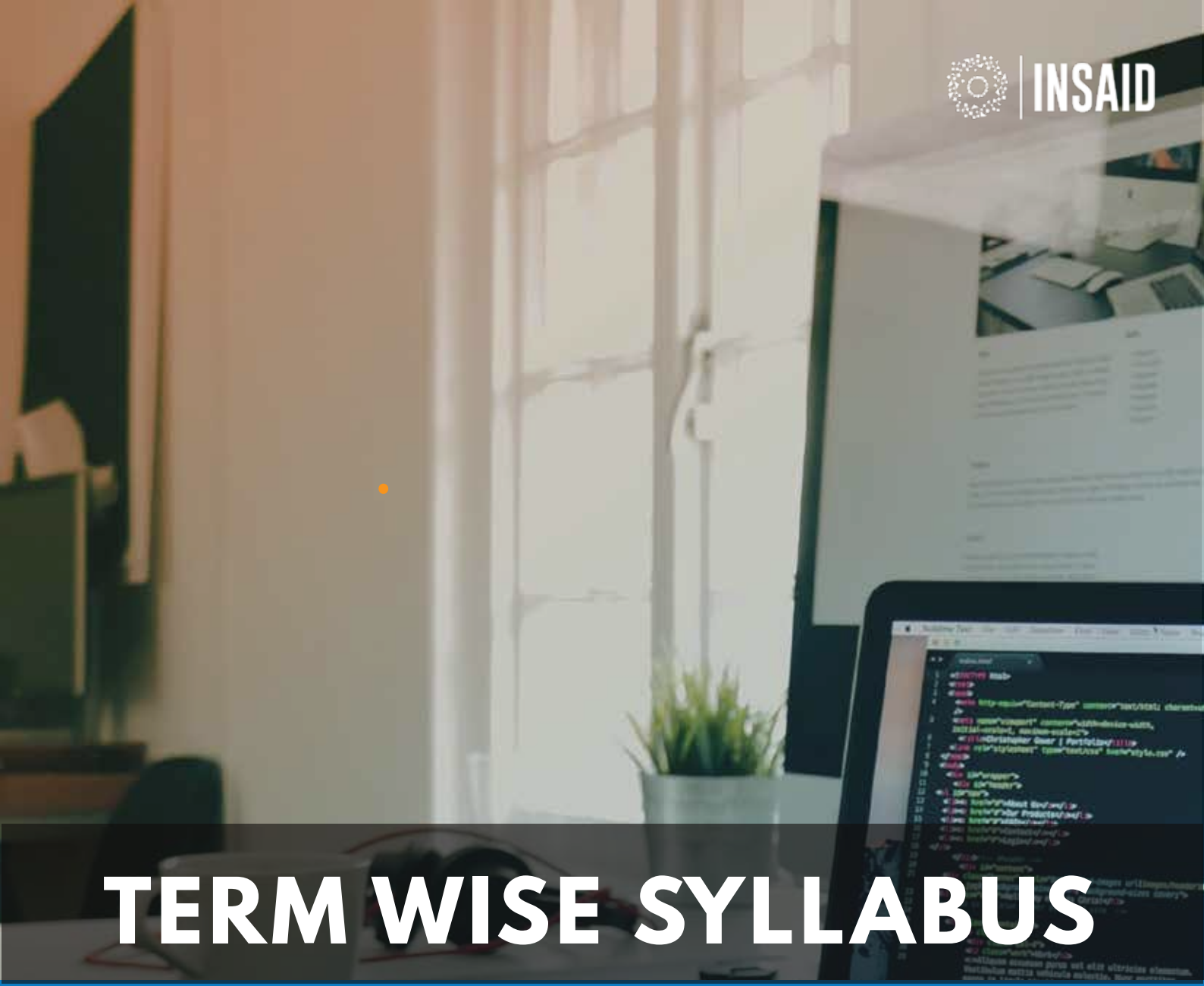
DATA SCIENCE & MACHINE LEARNING

Online | 6 months | 6 Terms

PROGRAM SNAPSHOT

6 month Certificate in Data Science and Machine Learning

- PRE - TERM PREP — Python & 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 — Machine Learning - III [Elective-1]
Data Analytics with R [Elective-2]
Deep Learning Foundation [Elective-3]
- TERM 06 — Capstone Project - II & Industry Immersion



INSAID

TERM WISE SYLLABUS

Term 1:

DATA ANALYSIS WITH PYTHON

Module 1: Data Science Fundamentals

Module Topics	Material
– 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	– Pre reads: Fifty Years of Data Science by David Donoho – Assignment: Practice questions on python data types – Post reads: Getting your first Data Science job

Module 2: Introduction to Numpy

Module Topics	Material
– Basics of Numpy Arrays – Mathematical operations in Numpy – Numpy Array manipulation – Numpy Array broadcasting	– Pre reads: Python for Data Analysis(ebook) – Assignment: Practice questions on Numpy and Pandas – Post reads: Python for Data Analysis(ebook)

Module 3: Data manipulation with Pandas

Module Topics	Material
– 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	– Pre reads: Python for Data Analysis(ebook) – Assignment: Practice questions on Numpy and Pandas – Post reads: Python for Data Analysis(ebook)

Module 4: Data Visualization in Python

Module Topics	Material
– 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	– Pre reads: Data wrangling in Python – Assignment: Case Study on IMDB Movie Reviews to identify gender movie preference – Assignment: Practice questions on Matplotlib and Seaborn – Post reads: Plotting and Visualization in Python

Module 5: Exploratory Data Analysis - 1

Module Topics	Material
– 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	– Pre reads: Data wrangling in Python – Assignment: Case Study on IMDB Movie Reviews to identify gender movie preference – Assignment: Practice questions on Matplotlib and Seaborn – Post reads: Plotting and Visualization in Python

Module 6: Exploratory Data Analysis - 2

Module Topics	Material
– Case study: Identify stock price patterns of Google & Apple – Case study: Perform EDA to analyze spread of tuberculosis	– Pre reads: Data wrangling in Python – Assignment: Case Study on IMDB Movie Reviews to identify gender movie preference – Assignment: Practice questions on Matplotlib and Seaborn – Post reads: Plotting and Visualization in Python

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)

Module Topics	Material
– What is Machine Learning ? – Use Cases of Machine Learning – Types of Machine Learning - Supervised to Unsupervised methods – Machine Learning workflow	– Pre read: What is Machine Learning & how is it transforming the world? – Post reads:Top 10 Machine Learning algorithms

Module 2: Linear Regression

Module Topics	Material
– 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	– Pre read: Examples of Linear Regression – Assignment: Predicting sepal length in Iris flower dataset – Post reads: Python package Scikit learn on Linear Regression

Module 3: Logistic Regression

Module Topics	Material
– 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	– Pre read: Examples of Logistic Regression – Assignment: Abalone gender classification – Postread: Python package Scikit learn on Logistic Regression

Module 4: Decision trees & Random Forests

Module Topics	Material
– 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	– Pre read: How do Decision Trees & Random forests work? – Assignment: Classify Iris flower species – Post read: Articles on Decision Tree classifiers

Module 5: Model evaluation techniques

Module Topics

- 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

Material

- Assignment: Apply various evaluation metrics to judge a credit card approval model

Term Projects

Air Quality Analysis	Predict Relative Humidity	The dataset contains 15 variables and over 5000 observations on weather conditions
Supermarket Purchase Analysis	Segment retail customers	The dataset contains 6 variables identifying shopping behaviour of a segment of customers.
Loan Payment data	Predict customer loan default class	The dataset contains 11 variables and over 300 observations on bank customers
NYC Taxi trips	Predict commute duration in urban metropolis	The dataset contains 1 million observations on trip durations in New York city
Wine Quality data	Predict wine quality	The dataset contains 12 variables and over 1500 observations on wine quality

Term 3:
DATA VISUALIZATION USING TABLEAU

Module 1: Introduction to Visual Analytics

Module Topics	Material
– Introduction to data visualization – Understanding Tableau ecosystem in industry – Loading data files in Tableau – Creating first visualizations – Case Study: Sales performance Analysis	– Prereads: Tableau Usecases (Ferrari, Ernst & Young, Cornell University) – Assignment: EDA of Hollywoods most profitable movies – Postreads: Business Statistics (by Amir Aczel)

Module 2: Data Visualization using Tableau

Module Topics	Material
– 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	– Pre reads: Tufte Principles of Data Visualization – Assignment : EDA of H1B Visa data – Postreads: Data cleaning using Excel

Module 3: Data joining & blending in Tableau

Module Topics	Material
– 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	– Prereads: SQL Joins fundamentals – Assignment: Mini Project on Startup investment Analysis – Postreads: Level of detail calculations

Module 4: Predictive Analytics using Tableau and R

Module Topics	Material
– 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	– Pre reads: Applications of Linear regression in business forecasting – Reference book: Business Statistics by Aczel – Assignment: EDA Miniprojects on USA Housing data, Cruchbase funding – Post Reads: Datacamp - R practice exercises

Module 5: Interactive Dashboard Design

Module Topics	Material
– 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	– Prereads: Fundamentals of Dashboard Design – Postreads: Developing Dashboards guide (e-book)

Module 6: Advanced Calculations using Tableau

Module Topics	Material
– 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	– Prereads: Applications of LOD Calulations – Postreads: More examples of LOD calculations

Module 7: Applications of advanced Calculations using Tableau

Module Topics	Material
– 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	– Prereads: Customer analytics methods and applications – Postreads: Tableau student reference guide

Module 8: Revision of concepts and Project discussion

Module Topics	Material
– 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	– Prereads: Tableau student reference guide – Project: Passenger demand analysis using Hubway Transportation data

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

Module Topics	Material
– 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	– Pre read : Dimensionality reduction method – Assignment : Implementing PCA on Students grades data – Post read : Applications of PCA

Module 2: KNN (K- Nearest neighbours)

Module Topics	Material
– 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	– Pre read : Introduction to KNN classification algorithm – Assignment : Classification of prostate cancer data – Post read : Implementing KNN using python from Scratch

Module 3: Naïve Bayes classifier

Module Topics	Material
– 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	– Pre read : Introduction to Bayes theorem – Assignment : Movie sentiment analysis – Post read : Bayes theorem examples

Module 4: K-means clustering technique

Module Topics	Material
– Introduction to K-means clustering – Decide clusters by adjusting centroids – Find optimal 'k value' in k means – Understand applications of clustering in Machine Learning – Case study : Segment hands in Poker data and segment flower species in Iris flower data	– Pre read : Introduction to K Means clustering – Assignment : Clustering crime data to identify similar patterns – Post reads : K Means clustering using Python

Module 5: Support vector machines (SVM)

Module Topics	Material
– 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	– Pre read : Introduction to SVM – Assignment : Gender classification on Abalone dataset – Post reads : Learning SVM in more detail (MIT opencourseware course)

Module 6: Time series forecasting

Module Topics	Material
– 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	– Pre read : Introduction to Time series analysis – Assignment : Forecast spot exchange rates or Euro vs US dollar – Post reads: Six types of decomposition in Time series method

Term Projects

Predict Relative Humidity

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

Segment retail customers

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

Predict customer loan default class

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

Predict commute duration in urban metropolis

The dataset contains 1 million observations on trip durations in New York city

Predict wine quality

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

Term 5 [Elective]:

MACHINE LEARNING - III

Module 1: Introduction to Apriori Algorithm

Module Topics	Material
– Applications of Apriori algorithm – Understand Association rule – Developing product recommendations using association rules – Case study:Analyse online retail data using association rules	– Pre read: Introduction to Apriori algorithm – Assignment : Build product associations using Online Retail data – Post read: Detailed tutorial of Apriori algorithm

Module 2: Recommender Systems

Module Topics	Material
– 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	– Pre read:The wonderful world of Recommendation systems – Assignment: Build a recommender system using Jokes ratings data – Post read: Recommender systems by Ritchi Ng (e-book)

Module 3: Linear Discriminant Analysis (LDA)

Module Topics	Material
– Recap of dimensionality reduction concepts – Types of dimensionality reduction – Dimensionality reduction using LDA – Case Study : Apply LDA to determine Wine Quality	– Pre reads: Is LDA a dimensionality reduction technique or classifier? – Assignment: Classify Prostate Cancer using LDA – Post reads: LDA by Sebastian Raschka

Module 4: Anomaly Detection

Module Topics	Material
– Introduction to Anomaly detection – How Anomaly detection works? – Types of Anomaly detection: Density based, Clustering etc. – Case Study:Detect anomalies on electrocardiogram data	– Pre Reads: Introduction to Anomaly detection – Assignment: Detect anomalies on server log data – Post Reads: Application guide for Anomaly detection

Module 5: Ensemble learning

Module Topics	Material
– 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	– Pre read : Basic introduction to ensemble learning – Assignment : Predict wine quality using XGboost – Post read : Why does Xgboost win every ML competition?

Module 6: Stacking

Module Topics	Material
– Introduction to stacking – Use Cases of stacking – How stacking improves machine learning models? – Case Study:Predict survivors in Titanic case	– Pre read: What is Stacking in Machine Learning? – Assignment: Classify Diabetes cases using stacking method – Post reads:How to implement stacked generalization from scratch?

Module 7: Optimization

Module Topics	Material
– 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)	– Pre read: Introduction to Optimization – Assignment: Stochasstic Gradient Descent on Linear Regression with given coefs & input data – Post read: Evolution of Gradient Descent

Module 8: Neural Networks

Module Topics	Material
– 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	– Pre read:Introduction to Neural Networks – Assignment: Predict house pricing in Boston – Post read:Google's Deep mind Explained

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 clty 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

Module Topics	Material
– Thought Experiment: Data science @ Google – Introduction to Data Science – Real world use-cases of Data Science – Walkthrough of data types – Data Science project lifecycle	– Pre reads: Fifty Years of Data Science by David Donoho – Pre reads: R in Action ebook (chapter 1) – Post reads: Getting your first Data Science job

Module 2: Introduction to programming in R

Module Topics	Material
– Installing R and R Studio – Basic Commands in R – Installing packages – Setting working directory – Exercises: Basic exercises in R Programming	– Prereads: R in Action ebook (chapters 2 and 3) – Assignment: More practice questions on R programming – Post reads: Basic Statistics

Module 3: Playing around with Data objects in R

Module Topics	Material
– Data structures – Basic Data management – Loops and Functions – Saving output – Exercises: Loops and functions in R	– Prereads: R in Action ebook (chapters 2 and 3) – Assignment: More practice questions on R programming – Post reads: Basic Statistics

Module 4: Descriptive statistics - 1

Module Topics	Material
– Introduction to Statistics – Descriptive Statistics – Measures of central tendency – Measures of Dispersion and shape – Case Study: Investigation of Crime statistics in Beaufort	– Pre reads:Basic Statistics by open intro ebook – Assignment:Basic Statistics and Probability Assessment – Post reads: Tutorial on Skewness and Kurtosis

Module 5: Descriptive statistics - 2

Module Topics	Material
– Introduction to Probability – Probability Distributions used in Data Science – Quantiles, percentiles, and standard score – Case Study: Analyse student's performance at school	– Pre reads:Basic Statistics by open intro ebook – Assignment:Basic Statistics and Probability Assessment – Post reads: Tutorial on Skewness and Kurtosis

Module 6: Inferential Statistics - 1

Module Topics	Material
– Introduction to Inferential Statistics – Population and Samples – Central Limit Theorem – Case Study: Sampling data for Business analysis	– Pre reads: Hypothesis driven thinking in data science – Assignment: Advanced Statistics Assessment – Post reads: Statistics and Data Analysis(ebook)

Module 7: Inferential Statistics - 2

Module Topics	Material
– Introduction to Hypothesis Testing – Confidence Intervals – Tests of significance: p-value – Case Study: Apply Inferential statistics & Central limit theorem using Python	– Pre reads: Hypothesis driven thinking in data science – Assignment: Advanced Statistics Assessment – Post reads: Statistics and Data Analysis(ebook)

Module 8: Intermediate R: Importing data

Module Topics	Material
– 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	– Prereads: R in Action ebook (chapters 4 and 5) – Assignment: More practice questions on R programming – Post reads: Applications of Tidyverse package

Module 9: Intermediate R: Data Manipulation using Tidyverse

Module Topics	Material
– 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	– Prereads: R in Action ebook (chapters 6 and 7) – Assignment: More practice questions on R programming – Post reads: Applications of ggplot2 package

Module 10: Intermediate R: Restructuring Data

Module Topics	Material
– Managing Date values – Numerical and Character functions – Aggregating & Restructuring data – Sorting, merging datasets – Exercises: Subsetting datasets for use in Predictive analytics	– Prereads: R in Action ebook (chapters 6 and 7) – Assignment: More practice questions on R programming – Post reads: Applications of ggplot2 package

Module 11: Intermediate R: Exploratory data analytics using ggplot2

Module Topics	Material
– Introduction to basic graphs: Barplots, Scatterplots & line graphs – Using Boxplots in univariate analysis – Applications of Histograms – Using ggplot2 for advanced visualizations	– Prereads: R in Action ebook (chapters 6 and 7) – Assignment: More practice questions on R programming – Post reads: Applications of ggplot2 package

Term Projects

Exploratory data analysis on Kickstarter funding data

Derive insights from successful and failed projects on Kickstarter platform

The dataset contains 15 variables and around 400,000 observations

Term 5 [Elective]:

DEEP LEARNING FOUNDAION

Module 1: Artificial Intelligence

Module Topics	Material
– Introduction to Artificial Intelligence – Breakthroughs in the field of AI – Overview of advanced Machine Learning algorithms – Weights and Bias estimation using gradient descent optimization	– Pre read: The Last Five years in Deep learning – Assignment: Basic Artificial Intelligence quiz – Post read: Large-scale Machine Learning using Tensor Flow

Module 2: Getting started with Tensorflow

Module Topics	Material
– 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	– Pre read: The Last Five years in Deep learning – Assignment: Basic Artificial Intelligence quiz – Post read: Large-scale Machine Learning using Tensor Flow

Module 3: Tensorflow programming in Python

Module Topics	Material
– Classification in Tensorflow – Introduction to Tensorboard visualization – Activation functions in Tensorflow	– Pre read: Black-box optimization – Assignment: Classify of hand written digits (MNIST) using Deep Neural Networks – Post read: Using neural nets to recognize handwritten digits

Module 4: Introduction to Deep Learning

Module Topics	Material
– 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	– Pre read: Black-box optimization – Assignment: Classify of hand written digits (MNIST) using Deep Neural Networks – Post read: Using neural nets to recognize handwritten digits

Module 5: Optimization of Deep Neural Networks

Module Topics	Material
– Hyperparameters in deep neural networks – Filters in Convolutional Neural Networks – Max pooling and padding – Dropout and Regularization in Deep learning	– Pre reads: Deep Learning and Convolutional Neural Networks – Assignment: Classify of hand written digits (MNIST) using Convolutional Neural Networks – Post reads: A Beginner's Guide To Understanding Convolutional Neural Networks

Module 6: Introduction to Convolutional Neural Networks

Module Topics	Material
– 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 –	– Pre reads: Deep Learning and Convolutional Neural Networks – Assignment: Classify of hand written digits (MNIST) using Convolutional Neural Networks – Post reads: A Beginner's Guide To Understanding Convolutional Neural Networks

Module 7: Introduction to Natural Language Processing

Module Topics	Material
– Introduction to NLP – Introdution to Word embeddings – Simple Word Vector representations: word2vec, GloVe – Implementation of word2vec model in Keras	– Pre read: Natural Language Processing and Machine Learning – Assignment: Build a recurrent neural network to classify SMS as either spam or not spam. – Post reads: An Intuitive Understanding of Word Embeddings

Module 8: Recurrent Neural networks

Module Topics	Material
– 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	– Pre read: Natural Language Processing and Machine Learning – Assignment: Build a recurrent neural network to classify SMS as either spam or not spam. – Post reads: An Intuitive Understanding of Word Embeddings

Module 9: Support vector machines (SVM)

Module Topics	Material
– 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	Pre reads: How to make an intelligent chatbot? – Assignment: Train Char-RNN on Colab GPU, – Assignment: Build a Word-RNN in place of Char-RNN Post reads: Creating an Ordering chatbot with simple NLP

Module 10: Revision of concepts of Deep Learning with NLP and Term Project

Module Topics	Material
– 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	– Pre reads: How to make an intelligent chatbot? – Assignment: Train Char-RNN on Colab GPU, – Assignment: Build a Word-RNN in place of Char-RNN – Post reads: Creating an Ordering chatbot with simple NLP

Term Projects

Build Chatbot using slack Class	This chatbot allows us translate user submitted conversations from English to Hindi

CAPSTONE PROJECT

James Telco Band

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?

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.

INSAID

PROGRAM STARTS

August 2018

DURATION

06 Months (Incl. Capstone Project)

PROGRAM FLOW

3 Months - Data Analysis → Data Visualization → Machine Learning - I → Capstone Project

3 Months - Machine LEarning - II → Elective → Capstone Project - II & Industry Immersion

WEEKLY COMMITMENT

10 Hours per week

> 4-5 Live virtual Instructor Class,
> 5-6 hours assignments & projects

PROGRAM FEE

INR 79,000/ + GST

Interest Free EMI Starting at INR 14,000/ month

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