

DATA SCIENCE COURSE AND IT'S DOMAIN SPECIFIC TRAINING (HANDS-ON USING R)

By Dinesh Babu, B.E, MBA, Ph.D.

THE COURSE APPROACH IS A CASE-STUDY- ORIENTED APPROACH, WHEREIN EVERY SESSION WILL HAVE A CASE STUDY ASSIGNMENT THAT THE PARTICIPANT HAS TO IMPLEMENT. BEYOND THEORETICAL KNOWLEDGE, LEARNING METHODOLOGY HEAVILY FOCUSES ON USE OF EXAMPLES, CASE STUDIES AND PROJECT WORK LEADS TO ACQUIRE INDEPTH EXPOSURE IN EVERY FACET OF DATA SCIENCE.



Real Time Case Studies and Capstone Projects to be covered



DATA SCIENCE COURSE CONTENT

Course Objective

This course gives an introduction to the field of business intelligence along with business analytics, which extensively use data, statistical and quantitative analysis, exploratory and predictive models, and fact-based management to take decisions and actions. Knowledge of R, SPSS and Excel is all the more important for interacting with Consultants should their help be needed for the organization. And, To facilitate statistical analysis of data to support the research.

1. Data Science Introduction

- 1.1 Data Science Process
- 1.2 What Makes Data Science different from Business Intelligence?
- 1.3 Data Statistics - Descriptive and Inferential
- 1.4 Data Visualization
- 1.5 Machine Learning Algorithm in Detail
 - Supervised Learning Algorithm
 - Unsupervised Learning Algorithm
 - Reinforcement Learning Algorithm
- 1.6 How Machine Learning different from Traditional Statistics
- 1.7 Data Science Importance and Key challenges
- 1.8 Life of Data Scientist
- 1.9 Data Science Application

2. Measurement and Scaling

- 2.1 Measurement and Scaling Introduction
- 2.2 Primary Scales of Measurement
 - 2.2.1 Nominal Scale and Ordinal Scale
 - 2.2.2 Interval Scale and Ratio Scale
- 2.3 Comparative Scaling Techniques
 - 2.3.1 Paired Comparison Scaling

- 2.3.2 Rank Order Scaling
- 2.3.3 Constant Sum Scaling
- 2.3.4 Q-Sort and Other Procedures
- 2.4 Non Comparative Scaling Techniques
 - 2.4.1 Continuous Rating Scale
 - 2.4.2 Itemized Rating Scale
 - 2.4.3 Likert Scale
 - 2.4.4 Semantic Differential Scale
 - 2.4.5 Stapel Scale



3. R Introduction

- 3.1 R – Overview and R Environmental Setup
- 3.2 Why R? And what is R?
- 3.3 Different flavors of R and Understanding R Studio
- 3.4 Packages and Libraries in R studio
- 3.5 Operators in R and R - Built in functions
- 3.6 R Studio Windows – R Script, R Console, R Plot, R Environment

4. R – Data Structures

- 4.1 R – Vectors and Lists
- 4.2 R – Strings and Matrices
- 4.3 R – Arrays and Factors
- 4.4 R – Data Frames
- 4.5 Top 12 – R Packages

5. R – Data Interfaces

- 5.1 R- CSV files Read and Write and analyses the data
- 5.2 R- Excel files Read and Write and analyses the data
- 5.3 Getting and Setting the Working Directory

5.4 Connecting Databases with R Studio

6. Data Visualization Using SPSS, R and Excel

6.1 Bar Graph and Line Graph

6.2 Area Chart and Pie Chart

6.3 Scatter Diagram and Histogram

6.4 High-Low Graph

6.5 Box Plot and Dual Axis Graph



7. Questionnaire Design

7.1 Questionnaire Definition

7.2 Objectives of a Questionnaire

7.3 Questionnaire Design Process

7.3.1 Specify the Information Needed

7.3.2 Type of Interviewing Method

7.3.3 Individual Question Content

7.3.4 Overcoming Inability to Answer

7.3.5 Overcoming Unwillingness to Answer

7.3.6 Choosing Question Structure

7.3.7 Choosing Question Wording

7.3.8 Determining the Order of Question

7.3.9 Form and Layout

7.3.10 Reproduction of the Questionnaire

7.3.11 Pretesting

8. Descriptive Statistics using R, SPSS and Excel

8.1 Central Tendency

8.1.1 Mean and Weighted Mean and Geometric Mean

8.1.2 Median, Mode, Percentiles and Quartiles

8.2 Dispersion

- 8.2.1 Variance, Standard Deviation and Range
- 8.2.2 Interquartile Range and Coefficient of Variation
- 8.3 Numerical Measures: Z-Scores, Chebyshev's Theorem, Empirical Rule and Detecting Outliers
- 8.4 Exploratory Data Analysis – Five – Number Summary, Box Plot
- 8.5 Measures of Association: Covariance and Correlation Coefficient



9. Inferential Statistics: Introduction to Probability

- 9.1 Probability and Statistical Experiment
- 9.2 Decision Tree Diagram
- 9.3 Counting Rule – Permutation and Combination
- 9.4 Assigning Probabilities – Classical, Frequency and Subjective method
- 9.5 Events and Their Probabilities
- 9.6 Relationships of Probability – Union, Intersection, Compliments and Mutually Exclusive events
- 9.7 Conditional Probability and Bayes' Theorem

10. Discrete Probability Distribution

- 10.1 Discrete Probability Distribution
- 10.2 Random Variable – Discrete and Continuous
- 10.3 Binomial Probability Distribution
 - 10.3.1 Evans Electronics Real time example using Binomial Probability
- 10.4 Poisson Probability Distribution
 - 10.4.1 Mercy Hospital Real time example using Binomial Probability distribution
- 10.5 Hyper geometric Probability Distribution
 - 10.5.1 Neveready's Hospital Real time example using Binomial Probability

11. Continuous Probability Distribution

11.1 Continuous Probability distribution

11.2 Uniform Probability Distribution

11.2.1 Slater's Buffet Real time example using Uniform Probability Distribution

11.3 Normal Probability Distribution

11.3.1 Pep Zone Real time example using Normal Probability distribution

11.4 Exponential Probability Distribution

11.4.1 Real time example using Exponential Probability distribution

12. Data Preparation

12.1 Data Preparation Process

12.2 Coding and Transcribing

12.3 Data Cleaning

12.4 Statistically Adjusting the Data

12.5 Selecting a Data Analysis Strategy

12.6 Classification of Statistical Technique – Uni variant and Multi Variant

13 Primary and Secondary Data

13.1 Primary Data Collection

13.2 Secondary Data Collection

13.3 Comparison of Primary and Secondary Data

13.4 Classification of Secondary Data

14 Experimental Design

14.1 Experimental Design Introduction

14.2 Pre Experimental Design

14.2.1 One-Shot Case Study

14.2.2 One Group Pre-test - Post-test Design

14.2.3 Static Group Design

14.3 True Experimental Designs

14.3.1 Pre-test - Post-test Control Group Design

14.3.2 Post-test Only Control Group Design

14.4 Quasi Experimental Designs

14.4.1 Time Series Design and Multiple Time Series Design

14.3 Statistical Design

14.4 Randomized Block Design

14.5 Latin Square Design

14.6 Factorial Design

15 Hypothesis Test

15.1 Introduction of Hypothesis Testing

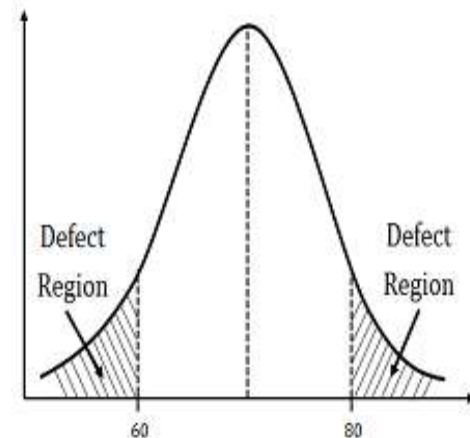
15.2 Types of Hypothesis Test

15.3 Formulation of Hypothesis Testing

15.4 Type I and Type II Error

15.5 Calculation of Test Statistic

15.6 Mapping Hypothesis Test with Real time Example



16 Correlation using R, SPSS and Excel

16.1 Correlation Analysis

16.2 Formulation of Correlation Matrix

16.2.1 Product Moment Correlation

16.2.2 Partial Correlation

16.2.3 Non metric Correlation

16.3 Mapping Correlation concept with Real Time Example

17 Regression using R, SPSS and Excel

- 17.1** Regression Analysis
- 17.2** Formulation of Regression Model
- 17.3** Bivariate Regression
- 17.4** Statistics Associated with Bivariate Regression Analysis
- 17.5** Conducting Bivariate Regression Analysis
- 17.6** Multiple Regressions
- 17.7** Conducting Multiple Regression
- 17.8** Mapping Bivariate Regression with Real Time Example

18 ANOVA – Analysis of Variance using R, SPSS and Excel

- 18.1** One way ANOVA
 - 18.1.1 Statistics associated with ANOVA
 - 18.1.2 Conducting One-Way Analysis of Variance
 - 18.1.3 Identification of Dependent & Independent Variables
 - 18.1.4 Decomposition of the Total Variation
 - 18.1.5 Measurement of Effects
 - 18.1.6 Significance Testing
 - 18.1.7 Interpretation of Results
 - 18.1.8 Mapping One-Way with real time example
- 18.2** Two - Way ANOVA
- 18.3** N – Way ANOVA

19 ANCOVA – Analysis of Covariance using R, SPSS and Excel

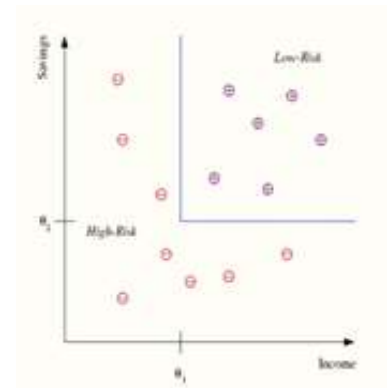
- 19.1** ANCOVA Introduction
- 19.2** Conducting ANCOVA
- 19.3** Mapping ANCOVA with Real time example

20 Dimension Reduction - Factor Analysis

- 20.1 Factor Analysis Introduction
- 20.2 Factor Analysis Model
- 20.3 Statistics associated with Factor Analysis
- 20.4 Conducting Factor Analysis
- 20.5 Construction of Factor Analysis
- 20.6 Factor Analysis Method
 - 20.6.1 Principal Component Analysis
 - 20.6.2 Rotation Method
- 20.7 Mapping Factor Analysis with Real Time Example

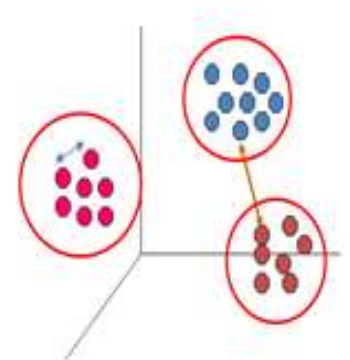
21 Discriminant Analysis

- 21.1 Relationship between ANOVA, Regression and Discriminant Analysis
- 21.2 Discriminant Analysis Model
- 21.3 Statistics associated with Discriminant Analysis
- 21.4 Conducting Discriminant Analysis
- 21.5 Multiple Discriminant Analysis
- 21.6 Mapping Discriminant Analysis with Real Time Example



22 Cluster Analysis using R, SPSS and Excel

- 22.1 Cluster Analysis Introduction
- 22.2 Statistics associated with Cluster Analysis
- 22.3 Conducting Cluster Analysis
- 22.4 Classification of Clustering Procedure
 - 22.4.1 Hierarchical Clustering
 - 22.4.2 Non Hierarchical Clustering

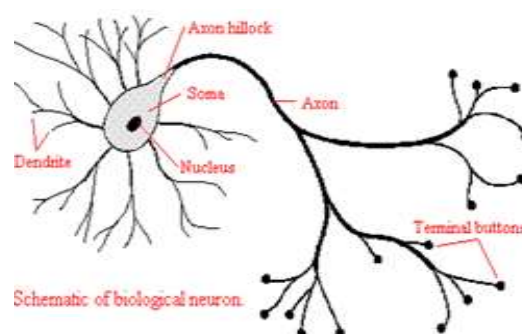


23 Logistic Regression using R, SPSS and Excel

- 23.1 Logistic Function
- 23.2 Single Predictor Model
- 23.3 Determine Logistic Cut off
- 23.4 Estimated Equation for Logistic Regression

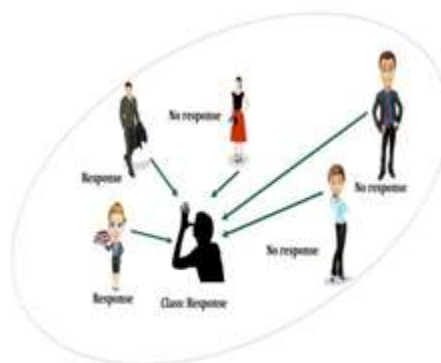
24 Artificial Neural Network (ANN)

- 24.1 Artificial Neural Network Introduction
- 24.2 ANN - Architecture and Schematic Diagram
- 24.3 ANN – Architectural Types:
 - 24.3.1 Single Layer Feed Forward
 - 24.3.2 Multiple Layer Feed Forward
- 24.4 Pre-processing steps of ANN
- 24.5 Back propagation Algorithm
- 24.6 Real Time case study using ANN
- 24.7 Advantage and Disadvantage of ANN
- 24.8 What ANN can do and What Not?
- 24.9 Comparison of ANN and Digital Computers
- 24.10 Application of Artificial Neural Network



25 K – Nearest Neighbour (KNN)

- 25.1 K – Nearest Neighbour Introduction
- 25.2 K – Nearest Neighbour Algorithm
- 25.3 Pre-Processing your dataset for KNN
- 25.4 How to measure “Nearby”
- 25.5 Choosing “K” and High “K” vs. Low “K”
- 25.6 Real Time case study using KNN



25.7 Advantage and Disadvantage of KNN

26 Naïve Bayes Classification

26.1 Naïve Bayes Introduction

26.2 Probabilistic Basics and Probabilistic Classification

26.3 Characteristics of Naïve Bayes

26.4 Real Time Case study using Naïve Bayes

26.5 Advantage and Shortcoming of Naïve Bayes

27 Association Rule

27.1 Association Rule Introduction

27.2 Apriori Algorithm

27.3 Multiple Association Rules

27.4 Market Basket Analysis (MBA)

27.5 Application of Apriori Algorithm and Market Basket Analysis



28 Decision Trees

28.1 What is a Decision Tree?

28.2 How to create Decision Tree

28.3 Choosing and Identifying attributes for Decision Tree

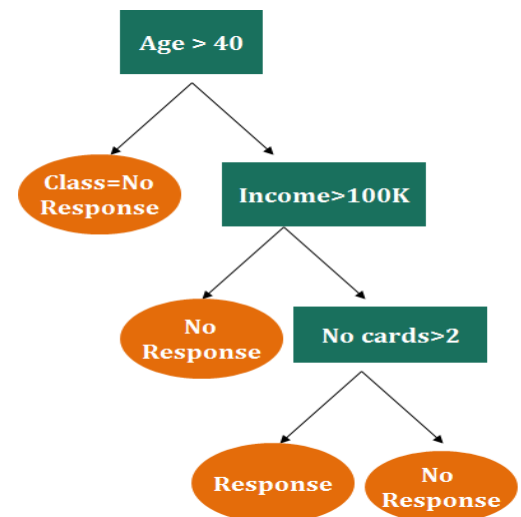
28.4 Entropy and Information Gain with Intuitions

28.5 Pruning Trees and its types

28.5.1 Forward Pruning and Backward Pruning

28.6 Sub tree Replacement and Raising

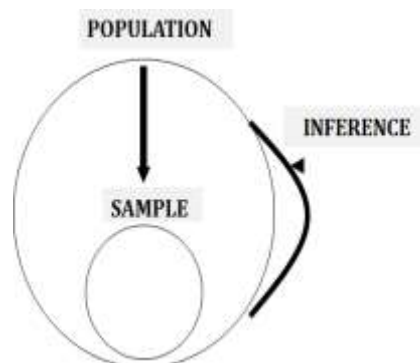
28.7 Real time case study with Decision Tree



29 Inferential Statistics using R, SPSS and Excel

29.1 Non – Parametric Test

- 29.1.1 Wilcoxon Sign Test
- 29.1.2 Friedman Test
- 29.1.3 Mann – Whitney Test
- 29.1.4 Kruskal – Wallis Test
- 29.1.5 Chi-Square Test



29.2 Parametric Test

- 29.2.1 T –test (One and Two Sample)
- 29.2.2 Z - test (One and Two Sample)
- 29.2.3 F – Test (One and Two Sample)

30 Data Warehousing

- 30.1 Data Warehouse Introduction
- 30.2 Online Transaction Processing (OLTP)
- 30.3 Online Analytical Processing (OLAP)
- 30.4 Data Warehousing Modelling
 - 30.4.1 Star Schema
 - 30.4.2 Snowflake Schema
 - 30.4.3 Fact Constellations

31 Support vector Machine

- 31.1 SVM Introduction
- 31.2 Classification as hyper plane location problem
- 31.3 Motivation for linear support vectors in classification problem
- 31.4 SVM as a quadratic optimization problem
- 31.5 Non Linear Support Vector Machine
- 31.6 Introduction to Kernel Methods

31.7 SVM Advantage and Disadvantage

32 Random Forest and its Application

32.1 Random Forest Introduction

32.2 Bagging Process

32.3 Real Time Case Study using Random Forest

32.4 Random Forest Application

33 Time Series

33.1 Time series Basics

33.2 Time series Component

33.2.1 Trend Component

33.2.2 Seasonal Component

33.2.3 Cyclical Component

33.2.4 Irregular Component

33.3 Smoothing Methods

33.3.1 Moving Average Method

33.3.2 Exponential Smoothing method

33.4 Trend Based Forecasting

33.4.1 Linear trend forecasting

33.4.2 Non Linear trend forecasting

33.4.3 Exponential trend forecasting

33.5 Auto Regressive (AR) Model and Moving Average (MA) Model

33.6 Autoregressive Moving Average (ARIMA) model using R

33.7 Plotting Time Series using R

CAPSTONE PROJECTS – REAL TIME PROJECT EXECUTIONS

All candidates would be pursuing an application oriented capstone project in the field of Data Analytics. The project shall be mentored and evaluated by Brainswig. The project will be presented to the Brainswig board as part of the requirement for successful completion of the program.

Project 1: Predictive Analytics with Model Simulation - Ames Housing Authority



- Formulate the Predictive model with respect to home Sale prices. Predict Home Sale Price for Test Dataset with lowest possible error.
- Compare efficiency and error rates for all Predictive models. And, Identification of best Predictive model with minimum prediction error rate.

33 Prediction of English Premier League (EPL) Championship – Gaming Analytics

- Predicting a standing Table for EPL championship and trying to figure out/predict the champions of EPL.
- Extract WOW factors and important trends from the given data that might help the stakeholders i.e., Advertisers, Bidders and clubs owners to take decisions.
- Calculation of Attack & Defense Power for each club



Project 3: Principal Component Analysis – Dimension Reduction – LKP Shares & Securities

- Identification of most influencing variable of investing in Prepaid Scheme
 - Restructuring the model for acquiring more investors, Competitor Analysis and predicting the best model



Project 4: Unsupervised Machine Learning – Merger and Acquisitions Analytics

- Credit Risk Analytics using Logistic Regression and Company's Short Term and Long Term Performance Analysis
- Financial Ratio Analysis, Sector wise and Industry wise Analysis Using Financial Analytics



Project 5: Predicting house prices for using Supervised



Machine Learning

- Build two Machine Learning model for predicting house

- Model 1: price vs. square feet. Referred as “square feet model”
- Model 2: price vs. Number of bedrooms. Referred as “bedroom model.
- Compare two different models for predicting house prices



Project 6: Machine Learning Classification - Portuguese banking institution

- Predict the term deposit subscription for the Bank additional dataset and conclude if the telemarketing campaign was a success or not.
- Classification goal is to predict if the client will subscribe a term deposit
- To predict which clients will secure a term deposit based on a set of information on client and purchase of term deposit.



Project 7: Automation of loan eligibility process - Dream Housing Finance Company

- Identification of target customers segments, those are eligible for loan amount
- Dream Housing Finance Company deals in all home loans and validates the customer eligibility for loan.
- Predicting bank’s revenue pattern based on saving account and loans they offer.

Short Brief about Myself:

I have Completed B.Tech, MBA from ICFAI Business School (IBS), Hyderabad and completed Ph.D. in the field of Data Analytics in BITS, Pilani. Currently, I am working for Capgemini as Senior Data Scientist, also worked for iGATE Global solution as Consultant in Analytics division. I am CBAP certified professional. And, National Level winner in Data Science Competition out of 4 Lakh participants and they selected best 30 Data Scientist Participants across India; I am also one among 30 Data Scientist which I attached at the end of this document for your reference. As a part-time training, Overall, I have 12 years of experience in the training profession to both Indian and overseas students. I can deliver below mentioned subject’s i.e.

List of Trainings offered:

- **Data Science Using Python, R Software, IBM SPSS and MS Excel**
- **Machine Learning using Python, R Software and IBM SPSS**
- **Advanced Statistics using R, IBM SPSS and MS Excel**
- **Business Analytics using R and Python**
- **Business Analysis as per CBAP Certification**

- **Data Analytics using Python and R Software**

With respect to training, past 2 years I received **Excellence award Winner** for Best Data Science and Business Analysis Teacher from **Urbanpro** (For your reference please [Click Here](#))



As teaching pedagogy, approach is a case – study oriented, wherein every session will have a case study assignment that the participant has to implement practically. Beyond theoretical knowledge, learning methodology heavily focus on use of examples, case studies and project work leads to acquire in-depth exposure in every facet of syllabus. Also, I will share the live project along with the training session. So, that participant won't have any skill gap after attended the training program.

List of Additional Trainings conducted:

- Conducted Data Analysis using SPSS and R training program for **Jadavpur University Students, Kolkata**. (India's 5th Ranked University)
- Conducted Data Analysis using SPSS training program for **Bangalore University Ph.D. Research Scholars**.
- Conducted Data Analysis using SPSS training program for **Vishweshwaraiah Technological University (VTU), Belgaum** – Ph.D. Research Scholars.
- Conducted Data Analysis using R program training for Research Scholars of **University of Madras, Chennai**.
- Conducted Data Analysis using R program training for MBA Students of **Velammal Engineering College, Chennai**.
- Conducted Advanced Statistics training program to **TATA Power Strategic Engg Division (SED)**, Bangalore.
- Conducted Data Analysis using SPSS training program for **Bangalore Institute Technology (BIT), Bangalore** – MBA Students.
- Conducted Data Analysis using SPSS training Program for **Wipro, Sholinganallur – Chennai** to Data Analyst Professional.
- Conducted Data Analysis using Python training Program for **Infosys, Mahindra City** – Chennai to Data Analyst Professional.
- Conducted Data Analysis Using SPSS training program for **Trichy – BHEL (Bharat Heavy Electrical Limited)** Top Management Committee.
- Conducted Data Analysis as Business Analysis training program for MBA Students of **Indus Business Academy, Bangalore**.
- Conducted Data Analysis using SPSS training program for **Shri Sakthikailash Women's College, Salem**.
- Conducted Data Analysis using SPSS training program for Research Scholars of **Sathyabama University, Chennai**.

Contact Details:

R. DINESH BABU, Cell: +91 9626201705 Email: rrrdinesh88@gmail.com

National Level Data Analyst Winner out of 2 lakh Top B-School Contestant like IIM's & IIT's

LatentView Winners out of 2 Lakh Participants

Dpl LV Timeline Recent Status Photo Place Life Event

Dpl LV shared a link, 2 seconds ago Only Me

iPad Air + PPI Winners

Ruchik Gandhi IIM Indore Aravindan Ingersol IMI Delhi Amitabh Vajpayee XLRI Jamshedpur

Like Comment Share 412 people like this.

Dpl LV shared a link, 2 seconds ago Only Me

iPod Shuffle + PPI Winners

1	Manveen Kaur	XLRI Jamshedpur
2	S Nitish Sivaramakrishnan	IIM Indore
3	Deepak Nesarikar	SDMIMD
4	Abhijeet Singh	IIT Madras
5	Manjunatha.C	IIFT Delhi
6	Manoj Hartharan	IIM Raipur
7	Yogendra Sarfare	TAPMI
8	Dharani Dharan	IIM Lucknow
9	Krishna Kumar	SDMIMD
10	S Abhishek	SCMHRD
11	Amit Kumar Sahoo	IIM Calcutta
12	Mayuresh Sanap	JBIMS
13	Bharat Bansal	IIM Kashipur
14	Kishor Kumar	IIM Kozhikode
15	Ashesh Pandey	IIT Bombay
16	Anirudh Singh Rana	IMT Ghaziabad
17	Sumit Guri	IIM Kozhikode
18	Vijaykumar Kurdagi	SDMIMD
19	Pallavi Agrawal	MDI Gurgaon
20	Avinash Devulapalli	IIM Kashipur
21	Sandeep Kumar V G	IIM Calcutta

Like Comment Share 170 1

Dpl LV shared a link, 2 seconds ago Only Me

PPI Winners

Ankit Dutt	IMT-G	Abhinav Bhatia	IMI Delhi
Chintan Gala	JBIMS	Abhishek Anand	IIM Rohtak
Dinesh Babu	IBS-H	Abhishek Gupta	IIM Kozhikode
Maharshi Nitish	IIM-C	Janak Lotwala	NMIMS Mumbai
Md Shadab Hussain	IIM-C	Kashish Chopra	SCMHRD
Pankaj Parasar	TAPMI	Kattamuri Hemanth	IIM Shillong
Paulson Mathew	IIM-L	Monika Sahu	IIT Kharagpur
Rony K roy	IIM-B	Ramaditya Goparaju	SCMHRD
Sameer Kumar	MDI-G	Randeep Singh Ibrar	IIM Ranchi
Venkatraman C	IMT-G	Rohit Garg	IIM Raipur
Vivek Kale	IBS-H	Subhankar Barik	XLRI-Jamshedpur

Like Comment Share 170 1

About Create Ad Create Page Developers Careers Privacy Cookies Terms Help

Training and Workshops conducted in Corporate as well as Universities

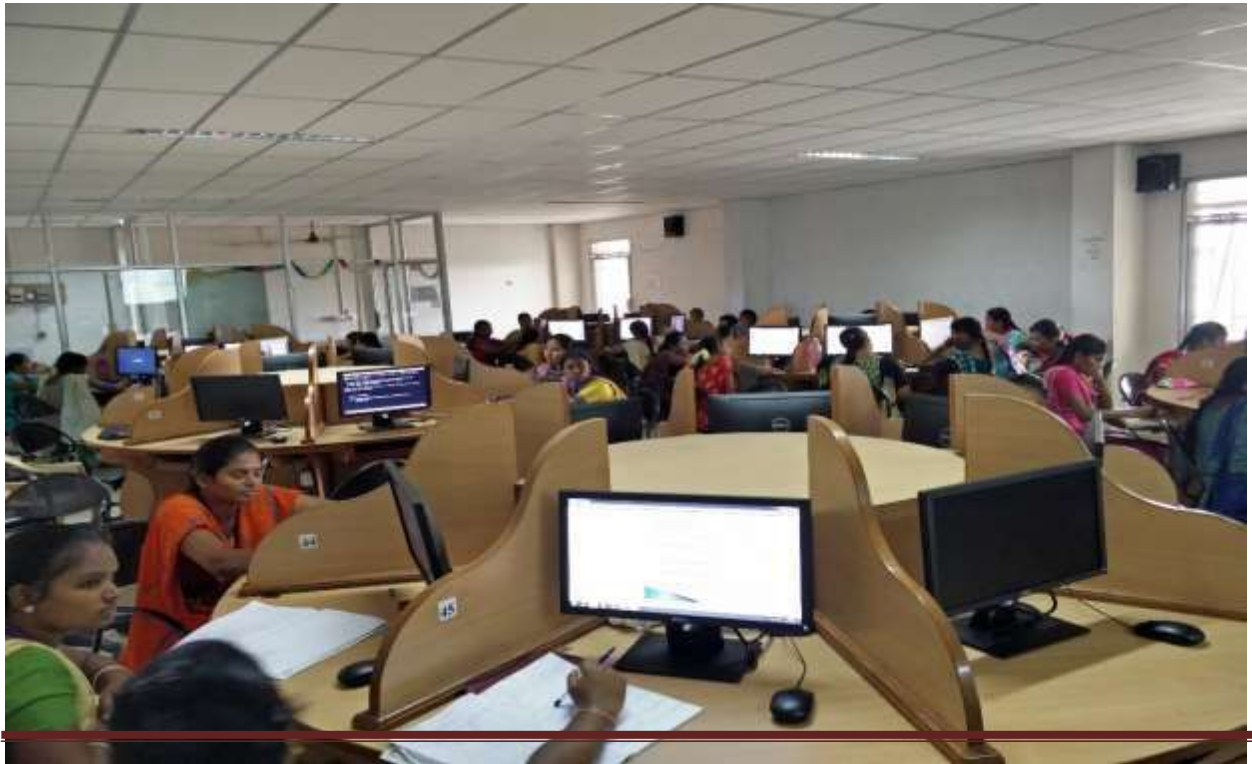
















PLACEMENT ASSISTANCE

The Career Assistance Service (CAS) team works hand in hand with you to further your career aspirations. From the first placement session during the program launch to final mock interviews, we thoroughly prepare to be interview – ready.



RESUME BUILDING

Refining and polishing the candidate's resume with insider tips to help them land their dream job



INTERVIEW PREP

Preparing candidate to ace HR and Technical interview rounds with model interview questions and answer



MOCK

Preparing candidates to face interview scenarios through 1:1 and panel mock interviews with industry veterans



INTERVIEW OPPORTUNITIES

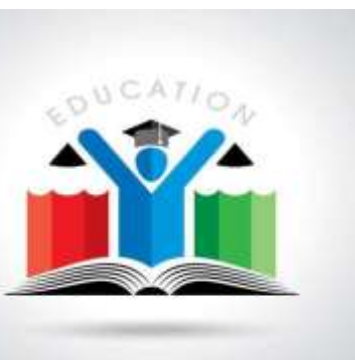
Assured 3-5 interview opportunities at leading firms to ensure you get placed



PLACEMENT PORTAL

Access to all available leads and references from open and private networks on our placement portal

COURSE OUTCOME:



- Ability to their Projects on their own.
- Ability to get into Machine Learning Jobs and Turn their professional Careers into Machine Learning Field.
- Also can Earn Money by doing machine Learning Project as a freelancer during their college Days itself.