

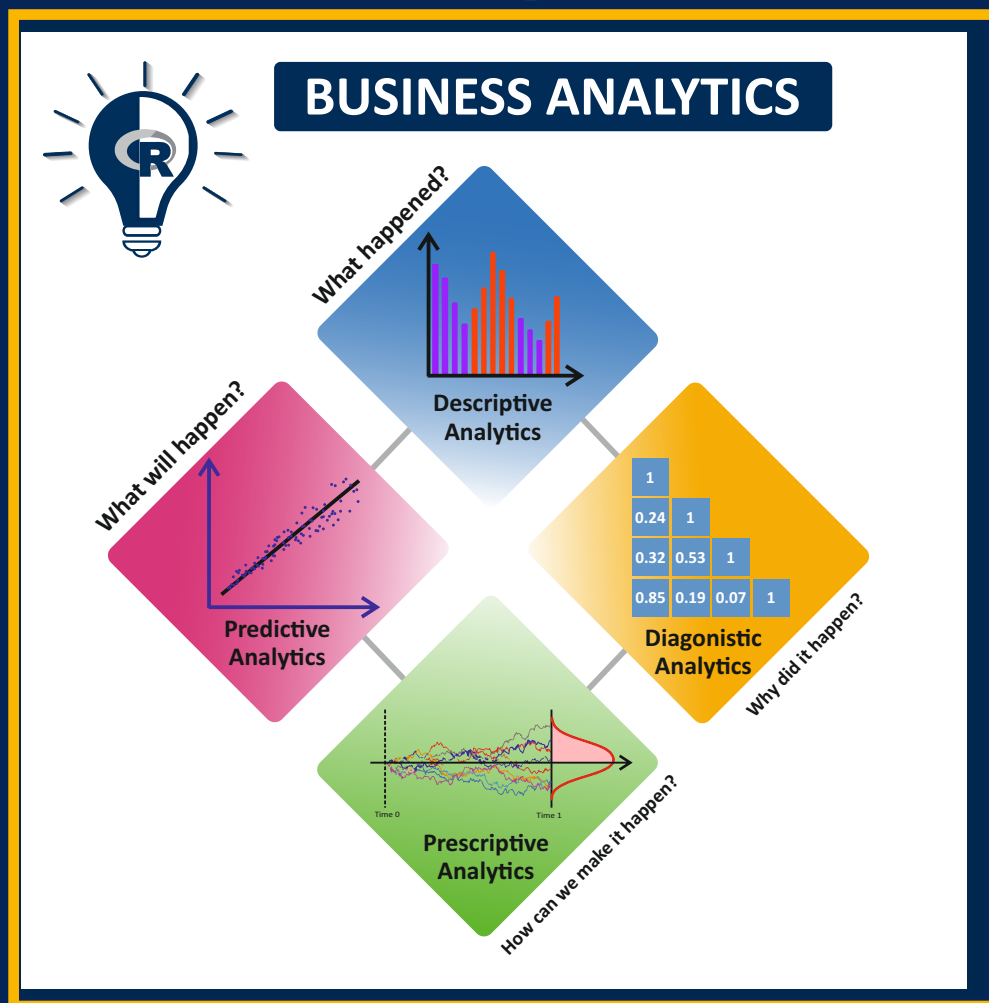
BUSINESS ANALYTICS

using R

50+ hours

Case Study and Project- driven Methodology

Blended Learning Methodology



PEAKS²TAILS



DETAILED CURRICULUM

MODULE 1 -FEATURE ENGINEERING

STEPS OF DATA CLEANING & PROCESSING

- Identifying Data type
- Exploratory data analysis - Data Summarisation and visualisation
- Missing value treatment for Categorical Variables
- Missing value treatment for Continuous Variables
- Outliers treatment for Categorical Variables
- Outliers treatment for Countinuous Variables
- Balancing of data
- Covariaties Creation (Enrichment)
- Dimensions Reductions
- Dummy Coding for Categorical Variables
- Scaling for continuous Variables
- Discretisation or Weight of Evidence
- Data Partitioning

MODULE 2 -MODEL BUILDING

STEPWISE REGRESSION

- Removing problems of Multicollinearity,
- Selecting Important Variables
- Correcting Problems of autocorrelation
- Checking problems of Non-Linearity
- Then correcting the Problems of Heteroskedasticity



DETAILED CURRICULUM

TYPES OF REGRESSION

- Multiple Regression
- Dummy Independent Variable - Dummy Regression
- Dummy Dependant Variable - Logistic Regression
- Penalised Regression - Ridge & Lasso Regression
- Forward Selection, Forward stagewise and least angle Regression
- Bayesian regression with Spike & Slab Selection
- Support Vector Regression
- Principal Component Regression

LOGISTIC REGRESSION

- General Linear Modelling in Excel using 4 main link functions Normal, Poisson, Binomial, Gamma.
- Multinomial Logit, Ordered Logit Model

DECISION TREE

- Classification and Decision Tree
- Classification & Regression Tree
- CHAID
- Boosting, Bagging & Random Forest

SUPPORT VECTOR MACHINE

- Primal & Dual Formulation
- Linear SVM, Non Linear SVM using Slack Variables
- Kernel Trick and Radial Basis Function

LINEAR DISCRIMINANT ANALYSIS

- Maximum Likelihood
- Fisher's Discriminant
- Bayesian Discrimination

K -NEAREST NEIGHBOUR

- 1.K-Nearest neighbour
- 2.K-means prototype



DETAILED CURRICULUM

NAIVE BAYES CLASSIFIER

- Bayes Theorem
- Naïve Bayes Classifier

CLUSTERING

- Bottom up Clustering a.k.a K means clustering
- Bottom up Clustering a.k.a Hierarchical clustering
- Top down Clustering a.k.a Minimal Spanning Tree
- Clustering using Expectation Maximization
- Mixed Variables Clustering - K means Prototype

MODULE 3 - MODELLING VALIDATION

VALIDATION METRICS

- Gini/AR
- AUROC/ CAP
- Unconditional Entropy
- Conditional Entropy
- Kullback-Leibler Divergence
- Kolmogorov-Smirnov (KS)
- Information Value

MODULE 4 - PROJECTS

PROJECTS

- Application to Credit data
- Application to HR data
- Application to Fraud detection
- Application to Marketing data

BACKGROUND

OBJECTIVE

ATTENDEES

PEDADOGY

BACKGROUND

The amount of data that is being generated on a daily basis has increased multiple times, the technological advancement in terms of storing this data has improved considerably. Therefore, in turn the ability to analyse the data and generate insights for data driven decision making becomes primarily importance However the availability of manpower with data science and machine learning skills is limited. The objective of the course is to introduce the concepts of data science and machine learning to the participants using R.



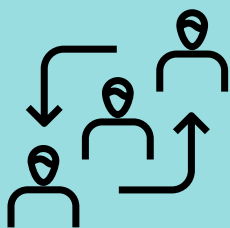
OBJECTIVE

Develop predictive models using various statistical and machine learning techniques ,Interpret and evaluate various models and its generalization , Hands on experience on the usage of open source packages like R and R Studio.



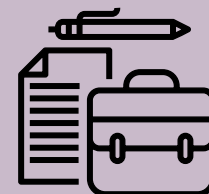
WHO CAN ATTEND

Beginner candidates from various quantitative backgrounds, like Engineering, Finance, Maths, Business Management who are looking for Business Analytics training to start their career in the field of Analytics and Data Science.



PEDADOGY

Professionals working in analytics field or students interested to make a career in analytics



DEMO MODELS

**DURING THE PROGRAM YOU WILL LEARN TO
CREATE EXCEL MODELS LIKE SHOWN BELOW**

```
30 }
31 }
32 }
33 }
34 }
35 }
36 }
37 }
38 creditminmax<-as.data.frame(lapply(credit_new,normalize))
39 creditzscore<-scale(credit_new)
40 summary(creditzscore)
41
42 set.seed(100)
43 train<-sample(1:999,700,replace = FALSE)
44 trainingdata<-creditminmax[train,]
45 testingdata<-creditminmax[-train,]
46
47 #using zscore values
48 trainingdata1<-creditzscore[train,]
49 testingdata1<-creditzscore[-train,]
50
51
52 library(class)
53 model_mm_27<-knn(trainingdata,testingdata,credit$Default[train],27)
54 table(model_mm_27,credit$Default[-train])
55 195/21+12+71
56 216/299
57
58 model_z_27<-knn(trainingdata1,testingdata1,credit$Default[train],27)
59 table(model_z_27,credit$Default[-train])
60 ##improving the model
61 #changing k
62 model_mm_25<-knn(trainingdata,testingdata,credit$Default[train],25)
63 table(model_mm_25,credit$Default[-train])
64 217/299
65
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
stepis on movieR
53 summary(fit.train)
54
55 #linearity
56 fit.train<- lm(Collection ~Marketing_expense + Production_expense + Budget + Producer_rating + Critic_rating + Trailer_views + X3
57 fitted2 <-fit.train$fitted.values^2
58 fit.train<-lm(Collection ~Marketing_expense + Production_expense + Budget + Producer_rating + critic_rating + Trailer_views + X3D
59 summary(fit.train)
60
61 Marketing_expense2 <-Marketing_expense^2
62 Production_expense2 <- Production_expense^2
63 Budget2 <- Budget^2
64 Producer_rating2 <- Producer_rating^2
65 Critic_rating2 <- Critic_rating^2
66 Trailer_views2 <-Trailer_views^2
67 X3D_available2 <- X3D_available^2
68 Time_taken2 <- Time_taken^2
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115 #?
116 # Outlier Treatment for Categorical Variables
117 # categories that occur less than 5% of the times in a column is considered as an outlier
118 table(cat_data$BusinessTravel)
119 36/length(cat_data$BusinessTravel) #if the least value is appearing more than 5% times,then others would definatey
120
121 table(cat_data$Department)
122 19/length(cat_data$Department) ##Appearing less than 5% of the times so to be replaced
123 cat_data$Department[cat_data$Department==1] <- cat_data$Department[cat_data$Department==3]
124
125 table(cat_data$Education)
126 60/length(cat_data$Education) # no outlier
127
128 table(cat_data$EducationField)
129 12/length(cat_data$EducationField) #should be replaced
130 18/length(cat_data$EducationField) #should be replaced
131 45/length(cat_data$EducationField) # need not be replaced
132 cat_data$EducationField[cat_data$EducationField==1] <- cat_data$EducationField[cat_data$EducationField==6]
133 cat_data$EducationField[cat_data$EducationField==5] <- cat_data$EducationField[cat_data$EducationField==6]
134
135 table(cat_data$EnvironmentSatisfaction)
136 96/length(cat_data$EnvironmentSatisfaction)
137
```

FREQUENTLY ASKED QUESTIONS

PREREQUISITE



Knowledge of Basic Excel

CERTIFICATE



Silver Certificate on successful completion of projects .
Gold Certification on passing a 2 hours MCQ based exam.

FEES



Rs.15000

DURATION



50+ hours

ABOUT THE TRAINER



Karan Aggarwal is one of India's leading trainers in Financial Modelling, Risk Modelling, Data Analytics and academic programs like Financial Risk Manager (FRM) & Actuarial Science. He has spearheaded several solution accelerators and spreadsheet-based prototypes in Risk and Analytics space. Karan has also authored a number of books on Advanced Excel, Statistical Modelling, Risk Modelling & Machine Learning. He is widely regarded for his problem-solving, thought leadership and intrapreneurship skills. His analytical mindset, solid fundamentals & the thirst to keep learning set him apart as a true authority in this field. Karan has also been awarded the Young Indian Entrepreneur Award by the Confederation Of Indian Industries in the year 2017.



OUR TRAINEES WORK IN



OUR SERVICES

1



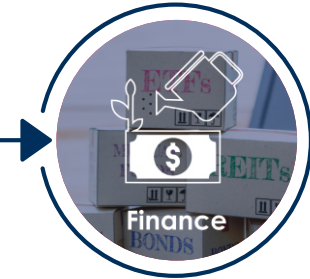
2



3



4



98 74 98 74 98

