Online

Certification Program on Business Analytics and Data Mining

During April 2022 – June 2022
[18 days]

Conducted by:

SQC & OR Unit, Mumbai

Indian Statistical Institute,

Room No 320, 3rd Floor Old C G O Building
101 Maharshi Karve Road, Mumbai 400 020
Tel 22014588 / 22004574
Email: sqcbombay@gmail.com
www.isimumbai.co.in

Introduction:

Business analytics refers to the skills, technologies and practices for continuous, iterative exploration and investigation of past business performance to gain insight and drive business planning. Data mining is the set of methods and techniques for exploring and analyzing large data sets, in an automatic or semi-automatic way, in order to find among these data certain unknown or hidden rules, association or tendencies that could be used to enhance business performance. Both the topics require analysis of data using statistical, machine learning and other quantitative techniques using various software. This program is planned to impart problem formulation and data analysis skills to the participants using appropriate statistical and machine learning techniques. The program focuses on practical applications to enable participants to solve real problems.

About the Institute:

The Indian Statistical Institute is a quasi-central organization under the Ministry of Statistics & Programme Implementation. It is declared by an Act of Parliament as an Institution of National Importance. Over the years the Institute has grown as a multi-disciplinary organization. It functions as a University empowered to award degrees up to Ph.D.; as a Corporation in undertaking large scale projects for public and private sectors as well as central and state governments; as a Firm of Consultants to industries to improve Quality, Reliability and Efficiency; and as a meeting place for academics and professionals.

Programme Objectives:

This program is designed to guide business analytics and data mining professionals in extracting implicit, previously unknown and potentially useful knowledge from large data sets. Practical data analysis including insight development, data cleaning, and building predictive and explanatory models in R using statistical and machine learning techniques will be covered in detail.

Program Benefits:

The participants will acquire knowledge on various statistical and analytical techniques required to carry out business analytics and data mining tasks effectively. All the topics will be covered using R-studio and MS-Excel.

Course Coverage: The course is divided in 6 modules as follows:

Module – 1: Introduction – Basic Concepts, Descriptive Statistics and Exploratory Data Analysis (EDA)

Introduction to Data Mining and Business Analytics; Understanding Data and Summary Measures; Sampling Concept and Methods; Introduction to R Programming; Basic Concepts of Probability, conditional probability and Bayes theorem; Introduction to Probability Distributions and their usage; Introduction to Exploratory Data Analysis (EDA) for effective visualisation and enhanced understanding of business problems

Module – 2: Basics of Explanatory and Predictive Analytics, Estimation and Hypothesis Testing, and Concepts of Machine Learning (ML)

Introduction to Explanatory and Predictive analytics; Estimation and hypothesis Testing in the context of Business Analytics; Methods of estimation – concepts of likelihood function and its usages; Estimation and Hypotheses Tests of means, variances and proportions; Estimation and hypotheses testing for categorical data; Missing Data Handling using MCAR, MAR and MNAR; Introduction to statistical models and problem formulation; Testing equality of several means using one and two-way ANOVA. Analysis of Covariance; Basic Concepts of Prediction Modelling and Machine Learning and their usage

Module – 3: Prediction Modeling – I (Regression and Classification)

Introduction to Multiple Linear Regression (MLR); Practical Model Building – Variable Selection and Pre-processing for MLR; Model Validation; Variants of MLR – Ridge and Principal Component Regression; Classification using Linear and Non Linear Models: Introduction to Discriminant analysis and Naive Bayes classifier.

Module – 4: Time Series Analysis, Survival Analysis and Tree Based Methods

Survival Analysis - Estimation of survival function using Kaplan – Meier method, introduction to Log Rank Tests, Cox Regression; **Univariate Forecasting Models –** ARIMA and Seasonal ARIMA (SARIMA), ARIMA with exogenous variable; **Time Series Regression** – Dynamic regression and Cochrane-Orcutt procedure; **Classification and Regression Trees** – Basics of tree building, concepts of pruning, adjustments to reduce classification costs, bagging and its usage, random forests and its usage to improve predictive accuracy, variable importance measures and their usages, tuning hyper-parameters.

Module – 5: Prediction Modeling – II: Logistic Regression and ML Techniques

Boosting Methods – L2Boosting, Ada Boost, Stochastic Gradient Boosting, XG Boost, Tuning boosting hyper-parameters, understanding why boosting usually works well; **Logistic regression** – logistic regression for risk assessment, parameter estimation and interpretation & Model diagnostics; **Support Vector Machines (SVM)** – Introduction, concept of kernels, kernelized regression Hyper-parameter tuning**: K-Nearest Neighbour (KNN)** – Introduction, Advantages and disadvantages, Pitfalls and requirements of pre-processing**: Introduction to Artificial Neural Networks (ANN):** Understanding ANN from the perspective of non-linear regression, feed forward networks, fitting neural networks.

Module – 6: Association Rule Mining, Clustering, Neural Networks (Advanced) and Introduction to Deep Learning:

Segmentation and Grouping – Cluster Analysis, Hierarchical clustering, Bayesian Hierarchical Clustering (BHC), K-means and K-medoids clustering, usage of clustering in business analytics; **Market Basket Analysis –** Introduction**,** Mining Frequent Patterns, Associations and Correlation; **Artificial Neural Networks** – Feed forward neural network, multilayer perceptron, problems of overfitting, hyperparameter tuning and fitting ANN with multiple layers; **Deep Learning** – Meaning and applications, where do we apply deep learning techniques, advantages and disadvantages, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), brief introduction to applications like Natural Language Processing (NLP), Sentiment Analysis and Text Mining.

After each module except the 6th module, test will be conducted in MCQ format. Assignment will be given after alternate modules. Participants are expected to carry out the assignments and present their finding in the next module to enhance their grasp on the subject.

Target Participants:
- Managers / Executives/ Professionals/ Students associated with data analysis or planning to have a career in Analytics.
- Should be graduate in any discipline. Preliminary knowledge of Excel is desirable.

Schedule:
April	08-10:	Module – 1 (Friday, Saturday and Sunday)
April	22-24:	Module – 2 (Friday, Saturday and Sunday)
May	13-15:	Module – 3 (Friday, Saturday and Sunday)
May	27-29:	Module – 4 (Friday, Saturday and Sunday)
June	10-12:	Module – 5 (Friday, Saturday and Sunday)
June	24-26:	Module – 6 (Friday, Saturday and Sunday)

Session Timings: 0930 to 1800 hrs. (0930-1100; 1130-1300; 1400-1530; 1600-1730)

Faculty:
Experienced faculties of SQC & OR Division having in-depth experience in data analysis and its application in various industries.

Course Fee:
- INR 60000 per participant + 18% GST (INR 70800/-) **for students and individual participants**
- INR 80000/- per participant + 18% GST (INR 94400/-) **for corporate/sponsored persons**
- US$ 1200 for overseas participants (inclusive of course material).

Registration:
- Please send your nomination along with appropriate course fee in the attached registration format given in the last page of this brochure. Before making online payment, please check eligibility and availability of seats. Fees once paid will not be refunded under any circumstances
- Registrations are done on 'first come first served' basis.

Important Dates:
Last date of submission of nominations**: *April 1, 2022***
Online Test and wrap-up: ***July 3,2022***

Venue: Online
The program will be conducted by using **Zoom** as on-line platform. Participants must sign up in zoom with the same e-mail id which they will provide in their nomination form.

Sponsored/Individual/Overseas (Please tick the appropriate one and cross off other options)

<div style="border:1px solid black;">

Training Programme on
Certification program for Business Analytics and Data Mining
April – June 2022

Organization:

Mailing Address:

Contact Person:

E-mail:
Phone: Fax :
Details of the participant(s) attending the program:

1. Name :
Position :

2. Name :
Position :

*Details of DD/Cheque attached (*Rs *) per participant inclusive of GST @18%) in favour of*
"Indian Statistical Institute" payable at Mumbai
Amount : _____
DD/Cheque No. :_____ Date :_____
Bank :_____Branch :_____

Signature :   *Name :*   *Date :*

Bank Details for online payments [please inform details of online payment to office]
Bank Name: STATE BANK OF INDIA
Bank Account No: 10996682279
Account Name;   INDIAN STATISTICAL INSTITUTE, Type: Current
Branch: MUMBAI MAIN BRANCH
Bank Address: MUMBAI SAMACHAR MARG, HORNIMAN CIRCLE, FORT
                    MUMBAI 400023
IFSC code: SBIN0000300

Signature with Date :

Name :

Please use copy of this form in case there are more than two participants. Please e-mail the form
to: sqcbombay@gmail.com or send to :
Program Director, SQC & OR Unit, Indian Statistical Institute, Room No,320, 3rd
Floor, Old CGO Bldg,101, Maharshi Karve Road, Mumbai 400 020
Tel. 022 - 22014588 / 22004574 *www.isimumbai.co.in*

</div>