

FORE School of Management

CERTIFICATION COURSE IN BBBB DATA ANALYTCS FOR BUSINESS & MANAGEMENT







PROGRAM COVERAGE

- Data Mining and Data Analytics
 - o Machine Learning algorithms using R and Python
 - o Hadoop and Kafka eco-systems
 - o NoSQL & Graph Databases
 - o Deep Learning, NLP and AI
- Business Analytics (Capstone, Python oriented)
- Web Analytics



Be able to clean, transform and visualize the dataset to gain deeper insights and make it ready for analysis

Be able to select a subset of appropriate machine learning algorithms that could be applied to get the desired predictive results

Gain sufficient proficiency in tools necessary to implement algorithms

Apply the knowledge of Deep Learning to a wide array of disciplines such as health, process control, navigation and others



Should be able to himself install, setup and configure and experiment with a complete hadoop and Kafka ecosystem

Should be able to install, configure and be sufficiently familiar with the variety of NoSQL databases and decide for himself which one to use, when and how



Put to use relevant tools and techniques to get a reasonable predictive accuracy



Be able to Install & Configure important Analytics and Storage Systems such as Hadoop-ecosystem, NoSQL databases & others

WHO SHOULD ATTEND

Specifically, the course will be useful to:



EXECUTIVES

Ambitious Executives (from Private/Public sectors) looking forward to sharpening their skills in making sense of data in order to innovate and add more value to their organization and to society



ACADEMICIANS Lecturers and Professors for extending the horizon of their knowledge through deepening their research skills



DATA SCIENTISTS/ DEVELOPERS

Techniques taught to them will have applications in a broad array of disciplines



STUDENTS / RESEARCH SCHOLARS

II nd year students currently enrolled in Engg. /PGDM/ MBA or any graduate or Postgraduate program who have had an introductory course in statistics. These students can look forward to better placement opportunities with added skill set

ELIGIBILITY

Graduate in any discipline

CERTIFICATE ISSUED BY

University of California, Riverside, Extension USA and FORE School of Management, New Delhi

PROGRAM SCHEDULE

| Duration 150 hours | Frequency Twice a week | | Class Schedule Saturday-Sunday | | Timings 10:30 am - 1:30 pm | |
|-----------------------|---------------------------|-----------------------------|-----------------------------------|-----------------------------|-------------------------------|-----------------------------|
| PROGRAM FEE | | Registration Fees | | | on Fee | 1 st Installment |
| ₹55,000/- +GST | Date | At the time of registration | | 20 th March 2018 | | 25 th May 2018 |
| | Amount | ₹10,000/- | * | ₹25,000/- | | ₹20,000/- |

Books and Material Fee: ₹5,500 +GST per participant, payable to 'FORE School of Management, New Delhi'

* Any request for refund of registration fees on account of valid reason prior to the closure of registrations or 10 working days before the date of course commencement whichever is earlier, the amount paid shall be refunded with a deduction of ₹5,000 + applicable taxes.



PEDAGOGY

- Projects
- together in a lab.

The Data Analytics program is Project based not Pure Theory-based

Learning with question/answers are in real-time: Live Virtual Interactive Learning Algorithms are first explained conceptually, avoiding mathematics and then these are implemented with real data from Industry as

Datasets for implementation are made available in advance and so also a copy of code to be executed. The code is numbered and copiously commented so that long after the lecture has finished, students can go back through the code/comments and refresh their knowledge During During the lecture, code is explained and

executed line-by-line. At his end the student also executes it. Consequently, results are available at our end as also on Students Laptop. The whole experience is as if everyone is sitting

PROGRAM CONTENT

INTRODUCTORY BUSINESS STATISTICS

- Measures of Central Tendency and Dispersion
- Probability Theory (Different Approaches, Rules of Probability, Baye's Theorem)
- Random Variables and Probability Distributions Discrete Probability Distributions - Binomial and Poisson Distribution
- Continuous Probability Distributions Normal Distribution
- Correlation and Regression Analysis: Simple & Multiple Regression
- Concept of Hypotheses Testing, Type I & Type II Errors, Power of The Test, Hypothesis Testing of Mean and Proportion, Two Sample Tests, Tests for Difference in Means and Proportions.
- Chi-Square Goodness-of-Fit Test, Test of Independence

DATA MINING AND DATA ANALYTICS

Module 2.1: Machine Learning Algorithms (using R and Python*)

- Developing familiarity with R; Data structures; Summarizing data; Data Exploration and transformation; integrating datasets; data & dates wrangling
- Data Visualization and story-telling. Developing relationships between various features and plotting distributions
- Data Mining: Measures of Proximity; Cluster Analysis: Curse of Dimensionality;
- K-means clustering and Model based clustering
- Text clustering and Agglomeration clustering;
- Evaluation of clusters; Cluster Validation; Clustering tendency
- Classification Analysis: Decision tree Induction; Cross-validation, parameter tuning & grid search
- Techniques of Dimensionality Reduction: PCA and SVD (Singular Value Decomposition)
- Neural Network
- Random Forest and Regression Trees; Determining feature importance with Boruta
- Gradient Boosting Technique for Machine Learning & grid search of its parameters
- Evaluating Classification: ROC, AUC, Precision, Recall, Specificity, Sensitivity; kappa metric; Overfitting;

Bias-variance trade-off; L1 & L2 regularization

- Ensemble modeling: A review of variety of techniques; Balancing datasets
- eXtreme Gradient Boosting (XGBoost)
- LightGBM: Light Gradient Boosting Machine

Module 2.2: Hadoop and Kafka Eco System; Processing streaming data and analysis

- Introduction to Hadoop and its ecosystem; Hadoop file storage formats
- Linux and Hadoop shell commands
- Hadoop streaming
- Hive on Tez and hadoop
- Pig on Tez and hadoop
- Pyspark and SparkSQL: Data storage and Extraction with SQL; Executing ML algorithms (including grid-search)) using MLlib and ML libraries
- Recommender Engine using Mahout on hadoop
- Installation of Hadoop ecosystem
- Apache Kafka: Stream data processing

Module 2.3: NoSQL and Graph Databases

- Introduction to NoSQL Databases and CAP theorem; Comparison with RDBMS
- Redis in-memory data structure store
- MongoDB Document Database
- Hbase column family database on hadoop
- Neo4j Graph Database

Module 2.4: Deep learning, NLP & AI

- Autoencoders and anomaly detection
- Deep Learning with Convolution Neural Network
- Using very Deep Convolution networks and Data Augmentation
- Transfer Learning
- Generative-Adversarial Networks (GAN)
- Recurrent Neural Networks & LSTM
- Natural Language Word2Vec Processing & transformation

BUSINESS ANALYTICS CAPSTONE (PYTHON ORIENTED)

Introduction to python; Using iPython; Basic data types

WEB ANALYTICS

STUDENTS EXERCISES/PROJECTS

- Random Forest
- Feature plotting

and data structures in python and pandas; Loops a Conditionals in python; Exploring data with pandas—Quick Start Numpy: Arrays; Basic arrays operations; Comparison operators and value testing for arrays; Array item selection and manipulation; Data Visualization in python; Data Visualization using t-distributed stochastic neighbor embedding (t-sne) k-means clustering with scikit-learn Decision trees classifier Ensemble Modeling ■ Logistic Regression (along with Dimensionality Reduction, PCA) Support Vector Machines Introduction to Keras on Tensorflow

Basics of Web analytics Analytic techniques and Tools: Google trends, Google Website optimizer, Google Analytics, Google Tag manager Data Analysis and Data Visualization

Data Visualization and story-telling. K-means clustering Model based clustering Dimensionality reduction and t-sne visualization Decision trees Induction K-Nearest Neighbour Neural Network Naïve Bayes Modeling eXtreme Gradient Boosting (XGBoost) Support Vector Machines Regression trees Apache Pig Exercises Analyse data on Spark/PySpark mongoDB Exercises Deep-Learning: Autoencoder Deep Learning

FACULTY PROFILE

Prof. Ashok Kumar Harnal, Professor in IT Area at FORE SChool of Management: Graduated from IIT Delhi; M.Phil, MA (Economics): Expert in Big Data and Data Analytics both on the technology side as also on Analytics side. Extensively taught faculty and students on the subject of big data technology and analytics. Participated in various machine learning problems in areas of business and marketing.



Prof. Kemal Oflus, Professor at UCR: Capstone Project Faculty covering Python module, Ex-rocket scientist. Highly motivated and versatile data scientist with fifteen plus years of proven analytics performance. Skilled at building effective and productive working relationships with customers, team members, executive management. Excellent time management, negotiation, interpersonal and presentation skills. A talent for analyzing problems, developing simplified procedures, and finding innovative solutions those improve operating efficiency and lower costs for customer. Successful in bringing methods long have been used in engineering and scientific communities to business customers and decision makers.

Prof. Hitesh Arora is a Professor in the area of Quantitative Techniques/ Operations Management at FORE School of Management, New Delhi. A graduate in Mathematics and a post graduate in Operational Research from University of Delhi, he has earned his Doctorate in Mathematical Programming from Department of Operational Research, University of Delhi. He has qualified National Eligibility Test (NET) conducted jointly by CSIR & UGC for Lectureship with Junior Research Fellowship (JRF) in Mathematical Sciences. He started his teaching career from University of Delhi and taught subjects like Optimization, Queuing Theory, Inventory Management and Statistics besides guiding students in their project work. As an actuarial consultant, his work involved Data Modeling and Reserving for Personal and Commercial Lines of different UK-based insurance companies. He has over seventeen years of experience in academics and industry.



Prof. Rakhi Tripathi, Associate Professor at FORE: PhD (IIT, Delhi) and MS (Computer Science) from Bowie State University (University of Maryland System). She has an experience of more than 10 years in research. She has worked on prestigious projects on Computer Networks and E-government at IIT Delhi. She has also presented and published several research papers in national as well as international reputed journals, conferences and books. Her current areas of interest include Computer Networks, E-government, Cloud computing, Mobile computing, Digital strategy and Social Media.

Prof. Dhanya Jothimani is PhD (Financial Analytics) – Thesis Submitted, IIT Delhi; M.Tech (Industrial Engineering and Management), IIT Kharagpur; B.Tech (Production Engineering), NIT Trichy. During her doctoral programme at IIT Delhi, Dhanya has presented her research work in well-reputed conferences including INFORMS Annual Meet, Annual meeting of Decision Sciences Institute (DSI) and Annual conference of Midwest Association for Information Systems (MWAIS). She was sponsored by Department of Science and Technology (DST) to present her research work at INFORMS Annual Meet 2016 at Nashville, Tennessee, USA. She has deliverd few lectures on R language and multi-criteria decision-making tools to postgraduate and doctoral students at IIT Delhi.







ABOUT EDUCATION LANES

Education Lanes is Tech Mahindra Growth Factories' initiative that offers certificate programs from premier institutes on a virtual platform. Education Lanes offers a comprehensive direct-to-device education suite with real-time interactive and participative virtual classroom sessions.

EDUCATION LANES

A Mahindra Group Initiative

For queries, call us at: 9975806184, 9811243210 Email: info@educationlanes.com | www.educationlanes.com

CLICK HERE TO REGISTER