Data Analytics for Business

Data analysis is the need of the hour. Today, different organizations are generating huge amounts of data without knowing how to make use of it for their benefit. To change this, machine learning and statistical techniques are now being to develop predictive models from existing data to forecast future outcomes.

Objective

Expecting to build a solid foundation of business analytics, this course has been designed to impart knowledge of machine learning and statistical methods for data analysis. The course shall also provide sufficient knowledge of python programming language to use for machine learning algorithm and python/R programming for statistical methods. A brief introduction of neural networks and deep learning will also be covered.

Target Audience

1. Students who've passed 10+2 examinations with Mathematics.
2. Professionals having knowledge of Mathematics.

Course Duration: 125 Hours

Fee: Rs. 40000/- plus taxes if any.

Probable Resource Persons:

1. Prof. Sanjeet Singh, IIM, Kolkata
2. Prof. R. K. Agrawal, JNU, Delhi
3. Prof. Aparna Mehra, IIT Delhi
4. Mr. Premnath Dalai, Co-founder Stepup Analytics (Practitioner)
5. Experts from industries and institutions of repute like the University of Delhi, etc.

Course Contents

Module 1: **Introduction to Business Analytics** **(35 hours)**

Descriptive Analytics: Describing and summarizing data sets, measures of central tendency, dispersion, skewness, kurtosis, Correlation.

Probability: Measures of probability, conditional probability, independent event, Bayes' theorem, random variable, discrete (binomial, Poisson, geometric, hypergeometric, negative binomial) and continuous (uniform, exponential, normal, gamma). Expectation and variance, markov inequality, chebyshev's inequality, central limit theorem.

Inferential Statistics: Sampling & Confidence Interval, Inference & Significance. Estimation and Hypothesis Testing, Goodness of fit, Test of Independence, Permutations and Randomization Test, t-test/z-test (one sample, independent, paired), ANOVA, chi-square.

Module 2: Data Manipulation using Python **(25 hours)**

Introduction to Python Editors & IDE's (Jupyter, Spyder, pycharm, etc.), custom environment settings, basic data types (numeric, string, float) and their operations, control flow (if-elif-else), loops (for, while), inbuilt functions for data conversion, writing user defined functions.

Concepts of packages/libraries – important packages like NumPy, SciPy, scikit-learn, Pandas, Matplotlib, seaborn, etc., installing and loading packages, reading and writing data from/to different formats, tuples, sets, dictionaries, simple plotting, functions, list comprehensions, database connectivity.

Module 3: Data Analysis **(15 hours)**

Relevance in industry, Statistical learning vs machine learning, types and phases of analytics.

Data pre-processing and cleaning: data manipulation steps (sorting, filtering, duplicates, merging, appending, subsetting, derived variables, data type conversions, renaming, formatting, etc.), normalizing data, sampling, missing value treatment, outliers.

Exploratory data analysis: Data visualization using matplotlib, seaborn libraries, creating graphs (bar/line/pie/boxplot/histogram, etc.), summarizing data, descriptive statistics,

univariate analysis (distribution of data), bivariate analysis (cross tabs, distributions and relationships, graphical analysis).

Module 4: Machine learning – Part 1 **(20 hours)**

Introduction, Applications of Machine Learning, Key elements of Machine Learning, Supervised vs. Unsupervised Learning.

Supervised Machine Learning: Linear Regression, Multiple Linear Regression Polynomial Regression.

Classification: Using Logistic Regression, Logistic Regression vs. Linear Regression, Logistic Regression with one variable and with multiple variables, Application to multi-class classification. The problem of Overfitting, Application of Regularization in Linear and Logistic Regression. Regularization and Bias/Variance. Classification using K-NN, Naive Bayes classifier, Decision Trees (CHAID Analytics), Random Forest, Support Vector Machines.

Model Evaluation: Cross validation types (train & test, bootstrapping, k-fold validation), parameter tuning, confusion matrices, basic evaluation metrics, precision-recall, ROC curves.

Case study

Module 5: Machine learning – Part 2 **(20 hours)**

Neural Networks: Introduction, Model Representation, Gradient Descent vs. Perceptron Training, Stochastic Gradient Descent, Multiclass Representation, Multilayer Perceptrons, Backpropagation Algorithm for Learning, Introduction to Deep Learning.

Association Rule Mining: Mining frequent itemsets, Apriori algorithm, market basket analysis.

Case study

Unsupervised Machine Learning: Introduction, Clustering, K-Means algorithm, Affinity Propagation, Agglomerative Hierarchical, DBSCAN, Dimensionality Reduction using Principal Component Analysis.

Case study: Application of PCA

Time Series Forecasting: Trends and seasonality in time series data, identifying trends, seasonal patterns, first order differencing, periodicity and autocorrelation, rolling window estimations, stationarity vs. non-stationarity, ARIMA and ARIMAX Modeling

Case Study

Module 6: Optimization in Analytics (10 hours)

Introduction to Operations Research (OR), Linear Programming Problems (LPP), Geometry of linear programming, Sensitivity and Post-optimal analysis, Duality and its economic interpretation.

Network models and project planning, Non-linear Programming – KKT conditions, Introduction to Stochastic models, Markov models, Classification of states, Steady-state probability, Dynamic Programming.

References:

1. Kumar, U.D. :Business Analytics – The Science of Data – Driven Decision Making, Wiley.
2. Gert, H.N., Thorlund, L. and Thorlund, J. :Business Analytics for Managers – Taking Business Intelligence Beyond Reporting, Wiley.
3. Johnson, R.A., Miller, I. and Freund, J. :Probability and Statistics for Engineers, Pearson.
4. Jose, J. and Lal, S.P. :Introduction to Computing & problem solving with Python, Khanna Publishers.
5. Bowles, M. :Machine Learning in Python – Essential Techniques for Predictive Analysis, Wiley.
6. Larose, D.T. and Larose, C.T.: Data Mining and Predictive Analytics, Wiley.
7. Bishop, C.M. :Pattern recognition & Machine Learning, Springer New York.
8. Falch, P. :Machine Learning, Wiley.
9. Deepa, S.N. and Sivanandam, S.N. :Principles of Soft Computing, Wiley.
10. Taha, A.H. :Operations Research – An Introduction, Prentice Hall.
11. Raschka, S. :Python Machine Learning

Course Co-coordinators:

1. **Dr. Sameer Anand**
2. **Dr. Ajay Jaiswal**