



Data Science

Data Visualization and Summarization

Data Visualization and Interpretation

Different types of data

Data summarization and visualization methods

Tables, Graphs, Charts, Histograms

Frequency distributions

Relative frequency

Measures of central tendency and dispersion

Box Plot

Chebychev's Inequality

Data visualization and Storytelling with Data

Basic probability concepts

Conditional probability

Bayes Theorem

Monty Hall Problem

Probability distributions

Continuous and discrete distributions

Sequential decision-making

Decision Tree

Sampling and estimation: Estimation problems, Point and interval estimates

Case Studies: 1. Central Parking Solutions Private Limited (IIMB Case);

2. A Dean's Dilemma: To Admit or Not to Admit (IIMB Case)

Descriptive Statistics

Introduction to Advanced Data Analytics

Statistical inferences for various Business problems

Types of Variables

Measures of central tendency

Dispersion

Variable Distributions

Probability

Distributions

Normal Distribution and Properties

Histograms

Exploratory Data Analysis

Histograms

Probability Theory

Bayes Theorem

Random Variables

Cumulative Distribution Function

Continuous Distributions



Skewness
Anova
Central limit Theorem
Monte Carlo Method
Estimate
Kernel Density Estimate
Regression
Covariance
Correlation
Causation
Euclidean Distance

Data quality and outlier treatment

Outlier treatment with robust measurements (median)
Outlier treatment with central tendency Mean
Replacing with series means or median values
Z score Calculation
Data Normalization
Sampling and estimation

Test of Hypothesis

Null/Alternative Hypothesis formulation
Type I and Type II errors
One Sample TTEST
Paired TTEST
Independent Sample TTEST
Analysis of Variance (ANOVA),
MANOVA
Chi Square Test (Non Parametric Tests)
Kruskal-Wallis,
Mann-Whitney,
Wilcoxon, McNemar test

Data Preparation and Quality Check

Data Validation and Data Imputation
Proc Univariate techniques analysis (SAS)
Q-Q probability plots
Cumulative frequency (P P) plots
Explorer analysis (SPSS)
Steam and leaf analysis
Kolmogorov Smirnov test
Shapiro Wilks test



Data Transformation

Log transformation (s)
Arcsine transformation
Box- Cox transformation
Square root transformation
Inverse transformation

Predictive Analytics

Simple linear regression
Coefficient of determination
Significance tests
Residual analysis
Confidence and Prediction intervals
Multiple linear Regression
Coefficient of determination
Interpretation of regression coefficients
Categorical variables
Heteroscedasticity
Multicollinearity
Outliers
Auto regression and Transformation of variables
Regression Model Building
Logistic and Multinomial Regression
Logistic function
Estimation of probability using logistic regression
Deviance
Wald Test
Hosmer Lemshow Test
Classification table
Gini co-efficient
Forecasting
Moving average
Exponential smoothing
Casual Models
"Application of predictive analytics in retail, direct marketing, health care, financial services, insurance, supply chain, etc."
Case Studies: Pricing of players in the Indian Premier League (IIMB Case)
Colonial Broadcasting Company (HBS Case)
Pedigree vs Grit: Predicting Mutual Fund Manager Performance (Kellogg Case)
Breaking Barriers – Micro-Mortgage Analytics (IIMB Case)
A Game of Two Halves: In-Play betting in Football (IIMB Case)

Optimization Analytics (Prescriptive Analytics)



Introduction to Operations Research (OR)
Linear programming (LP)
Formulating decision problems using linear programming
Interpreting the results and sensitivity analysis
Concepts of shadow price and reduced cost
Multi-period LP models
"Applications of linear programming in product mix, blending, cutting stock, transportation, transshipment, assignment, scheduling, planning and revenue management problems."
Network models and project planning.
Integer Programming (IP) problems
Mixed-integer and Zero-one programming
Applications of IP in capital budgeting, location decisions, contracts
Multi-criteria decision making (MCDM) techniques
Goal Programming (GP) and analytic hierarchy process (AHP)
Applications of GP and AHP in solving problems with multiple objectives
Non-linear programming, portfolio theory
Case Studies: 1. Merton Truck Company (HBS Case),
2. Supply Chain Optimization at Madurai Aavin Milk Dairy (IIMB Case)
3. Red Brand Canners (Stanford Case)

Stochastic Models (Marketing and Retail Analytics)

Introduction to stochastic models
Markov models
Classification of states
Steady-state probability estimation
Brand switching and loyalty modeling
Market share estimation in the short and long run
Poisson process
Cumulative Poisson process
Applications of Poisson and cumulative Poisson in operations, marketing and insurance
Measuring effectiveness of retail promotions, warranty analytics
Renewal theory, Applications of renewal theory in operations and supply chain management
Markov decision process, Applications of Markov decision process in sequential decision making
Case Studies: 1. Browser Wars: Microsoft Vs Netscape (Darden Case)
2. Consumer choices between house brands and national brands in detergent purchase at Reliance Retail (IIMB Case)"
3. MNB ONE Credit card Portfolio (Darden Case).

Market Research and Operations Analytics

Principal component analysis



Factor analysis
Conjoint analysis
Discriminant analysis
"ARCH (autoregressive conditional heteroscedasticity)
ARCH (Generalized autoregressive conditional heteroscedasticity)"
Monte Carlo simulation
Supply chain analytics
Classification and regression trees (CART)
Chi-squared automatic interaction detector (CHAID)
Statistical process control, Value stream mapping, TRIZ
Case Studies: 1. Apollo Hospitals: Differentiation through Hospitality (IIMB Case)
2. Dean's Dilemma: To Admit or Not to Admit (IIMB Case)
3. Dosa King – A Standardized Masala Dosa for Every Indian (IIMB Case)
4. Delivering Doors in a Window – Supply Chain Management at Hindustan Aeronautics Limited (IIMB Case)

Analytics in Finance and Insurance

"Dynamic pricing and revenue management, high dimensional data analysis, financial data analysis and prediction"
Survival analysis and its applications:
Life tables
KapMeier estimates
Proportional hazards
Predictive hazard modeling using customer history data
Analytics in finance, discounted cash flows (DCF), Profitability analysis
Asset performance:
Sharpe ratio, Calmar ratio, Value at risk (VaR), Brownian motion process, Pricing options and Black-Scholes formula, Game theory
Insurance loss models: Aggregate loss models, discrete time ruin models, Continuous time ruin models

Predictive Modeling & Diagnostics

Correlation - Pearson, Kendall
SLR Regression
MLR Regression
Residual analysis
Auto Correlation
VIF Analysis
Indexing Eigen Value interpretation
Homoscedasticity
Homogeneity
Stepwise regression
Transformation of variables
Logistic Regression Analysis



Discriminant and Logit Analysis
Multiple Discriminant Analysis
Stepwise Discriminant Analysis Binary
Logit Regression
Estimation of probability using logistic regression, Wald Test
Hosmer Lemshow

Advanced Analysis

Factor Analysis
Introduction to Factor Analysis – PCA
Reliability Test
KMO MSA tests
Eigen Value Interpretation
Rotation and Extraction
Varimax Models
Principle component analysis
Conformity Factor Analysis
Exploitory Factor Analysis

Cluster Analysis

Introduction to Cluster Techniques
Distance Methodologies
Hierarchical and Non-Hierarchical Procedures
K Means clustering
Wards Method

Conjoint Analysis

Statistics and terms Association with Conjoint Analysis
Assumption and limitation of conjoint analysis
Hybrid Conjoint Analysis

Time Series Forecasting

Introduction to Time Series Data
Visualizing Time Series data
Data Exploration
Intro to AR, MA, ARMA, ARIMA Models
Smoothing and annual Time series
Time series forecasting for seasonal data
Multiplicative Models
Additive Models
Case studies with ARIMA Models (Using SAS & R)



Loss Forecasting & Portfolio Monitoring

Introduction to Survival Analysis
Using Survival Models to build Credit Baselines
Loss forecasting
Portfolio and monitoring & strategy

Introduction to Risk Management Practices by Bank

Overview of Scorecard's
Stages of Scorecard Development
Types of Scorecards

Scorecard Development Project

Scorecard Development Project planning
Roles of Responsibilities
Data Review and Project Parameters
Project plan

Scorecard Development Preliminaries

Data Exclusion
Variable selection
Target Variable definition
Data Description

Scorecard Development

Missing Value Check
Exploratory Data Analysis
Correlation Analysis
Assigning predictive power
Sampling
Model development
Model Selection
Model finalizations
Model validation
Variable Selection using Weight of Evidence & Information Value
Scorecard Development using grouped variables
Concept of variable grouping
Computing Weight of Evidence (WOE) and Information Values
Model development with WOE Variables

Reject Inference



Introduction to Reject Inference
Reject inference techniques
Strategies developed with Reject Inference
Performing reject inference and remodeling on SAS

Model Validations, Finalization and Uses

Introduction to Scaling
Finalize models/Model selection based on performance stats
Model Validations
Scorecard Management reports and documentation

Data Mining

Data Mining
Data partition (Training, Validating Testing)
Data Explore
Data Testing
Data Transform
Linear Model
SVM Model
Tree Analysis
RandomForest Analysis
Model Evaluation
ROC
Lift Curve
Sensitivity
Error/ Confusion matrices

Text Mining

Vocabulary Mapping
Classify Text
Using NLTK
Feature Extraction
Market Basket Analysis
Association Rules
Support Vector Machines
Term Frequency and Weight
Term Document Matrix
UIMA
Text Analysis
Named Entity Recognition
Corpus

Business Analytics



Introduction and Data Analytics
Linear Regression
Logistic Regression
Decision Tree and Clustering
Time Series Modeling
Logistic Regression
Market Basket Analysis
Cross Sell Model
Market Mix Modeling
Churn Analytics
Buy till You Die Model
Customer Lifetime Value Analysis
Telecom Model to Estimate Bill

Machine Learning

Numerical and Categorical var
Supervised and Unsupervised Learning
Concepts, Inputs and Attributes
Classifier
Prediction
Bias and Variance
Trees and Classification
Decision trees
Boosting
Naive Bayes Classifiers
K-Nearest Neighbor
Logistic and Linear Regression
Ranking and Preceptor
Hierarchical and K-Means Clustering
Neural networks
Sentimental Analysis
Collaborative Filtering
Tagging
Vocabulary mapping

Linear Regression with One Variable

Model and Cost Function
Parameter Learning

Linear Regression with Multiple Variables



Environment Setup Instructions
Multivariate Linear Regression
Computing Parameters Analytically

Logistic Regression

Classification and Representation
Logistic Regression Model
Multiclass Classification

Regularization

Solving the Problem of over fitting

Neural Networks: Representation

Motivations
Neural Networks
Applications

Neural Networks: Learning

Cost Function and Back propagation
Back propagation in Practice
Application of Neural Networks

Advice for Applying Machine Learning

Evaluating a Learning Algorithm
Bias vs. Variance

Machine Learning System Design

Building a Spam Classifier
Handling Skewed Data
Using Large Data Sets



Support Vector Machines

Large Margin Classification
Kernels
SVMs in Practice

Unsupervised Learning

Clustering

Dimensionality Reduction

Motivation
Principal Component Analysis
Applying PCA

Anomaly Detection

Density Estimation
Building an Anomaly Detection System
Multivariate Gaussian distribution (Optional)

Recommender Systems

Predicting Movie Ratings
Collaborative Filtering
Low Rank Matrix Factorization

Natural language Processing (NLP)

Language modeling
Hidden Markov models
Basic Text Processing
Minimum Edit Distance
Language Modeling
Spelling Correction
NLP Tasks and Text Similarity



Syntax and Parsing
Language Modeling and Word Sense Disambiguation
Part of Speech Tagging and Information Extraction
Text Summarization
Collocations and Information Retrieval
Sentiment Analysis and Semantics
Discourse, Machine Translation, and Generation
Tagging problems
Probabilistic context-free grammars
Parsing problem
Statistical approaches to machine translation
Log-linear models and their application to NLP problems
Unsupervised and semi-supervised learning in NLP

Deep Learning

Intro to Deep Learning
Simple Word Vector representations: word2vec, GloVe
Advanced word vector representations: language models, softmax, single layer networks
Neural Networks and back propagation -- for named entity recognition
Project Advice, Neural Networks and Back-Prop (in full gory detail)
Practical tips: gradient checks, over fitting, regularization, activation functions, details
Introduction to Tensor flow
Recurrent neural networks -- for language modeling and other tasks
GRUs and LSTMs -- for machine translation
Recursive neural networks -- for parsing
Recursive neural networks -- for different tasks (e.g. sentiment analysis)
Convolutional neural networks -- for sentence classification
The future of Deep Learning for NLP: Dynamic Memory Networks
Machine learning and artificial neural networks
Generative models
Restricted Boltzmann Machine
Contrastive Divergence algorithm
Deep Belief Network
Back propagation algorithm
Logistic regression and Soft max regression
Unsupervised pertaining in deep neural networks



Regularization in neural networks
Dropout

Convolutional Neural Networks

Invariance, stability
Variability models (deformation model, stochastic model).
Scattering networks
Group Formalism
Supervised Learning: classification
Properties of CNN representations: invertibility, stability, invariance
Covariance/invariance: capsules and related models.
Connections with other models: dictionary learning, LISTA.
Other tasks: localization, regression.
Embeddings (DrLim), inverse problems
Extensions to non-euclidean domains
Dynamical systems: RNNs.

Deep Unsupervised Learning

Auto encoders (standard, denoising, contractive, etc)
Variational Autoencoders
Adversarial Generative Networks
Maximum Entropy Distributions
Non-convex optimization for deep networks
Stochastic Optimization
Attention and Memory Models

Artificial Intelligence

Introduction
Overview
Agents: Perception, Decisions, and Actuation
Search and Planning
Uninformed Search (Depth-First, Breadth-First, Uniform-Cost)
Informed Search (A*, Greedy Search)
Heuristics and Optimality
Constraint Satisfaction Problems
Backtracking Search
Constraint Propagation (Arc Consistency)



Exploiting Graph Structure
Game Trees and Tree-Structured Computation
Minimax, Expectimax, Combinations
Evaluation Functions and Approximations
Alpha-Beta Pruning
Decision Theory
Preferences, Rationality, and Utilities
Maximum Expected Utility
Markov Decision Processes
Policies, Rewards, and Values
Value Iteration
Policy Iteration
Reinforcement Learning
TD/Q Learning
Exploration
Approximation

Geospatial Analysis

GIS Principles and Technology
Mapping Science
Principles of Spatial Analysis
Representation, Structures and Algorithms
Airborne Data Acquisition
Climate Modeling
Geo demographics and Population Geography
GIS Design
Network and Locational Analysis
Spatial Decision Support Systems
Spatio-temporal Analysis and Data Mining
Surface Water Modeling
Terrestrial Carbon: Modeling and Monitoring
Web and Mobile GIS

Big Data analysis

Hadoop
Introduction to big data
Sources of big data
Hadoop distributed file system
Statistical Analysis of Big Data

Tools: R, SPSS, SAS, SAS Enterprise Miner, WEKA, Excel

**** Hadoop and Tableau Training without any Additional charges****