# SAVITRIBAI PHULE PUNE UNIVERSITY (Formerly University of Pune)



# **POST GRADUATE PROGRAM IN BUSINESS ANALYTICS**

# SYLLABUS

**Department of Technology** 

w.e.f. from Academic Year 2018-19

# Syllabus 2018\_PGPBA

| Term 1                      | Term 2               | Term 3                  |
|-----------------------------|----------------------|-------------------------|
| Business Analytics Approach | Big Data Spark       | Analytics in Finance    |
| Introduction to Python      | Customer Analytics   | Bayesian Networks       |
| Machine Learning 1          | Data Visualization   | Deep Learning           |
| Managerial Communication    | Econometrics         | Introduction to SAS     |
| Market Research             | Introduction to R    | NoSQL                   |
| RDBMS and Data Warehousing  | Machine Learning 2   | Optimisation Techniques |
| Statistics 1                | Model Interpretation | Retail Analytics        |
|                             | Statistics 2         | Text Analytics          |
|                             |                      | Capstone Project        |

## Term 1

# **Business Analytics Approach**

| Session | Topics  |  |
|---------|---|--|
| 1 - 2   | <ul> <li>Analytics landscape</li> <li>What it's all about - Definitions and some facts</li> <li>Evolution - human/machine/ai , reactive/proactive/simulation</li> <li>Why now? - Raw materials available now - data, tech and stat maturity</li> <li>Applications - Analytics across value chain, connect back to ROI, harnessing machine data</li> </ul> |  |
| 3       | <ul> <li>Data extraction and data cleaning - Process of data extraction and ensuring<br/>validity, accuracy, completeness</li> </ul>  |  |
| 4 - 5   | <ul> <li>Analytical Thinking</li> <li>Problem Solving Approach to Analytics - MECE, Trees, SCQA, etc.</li> </ul>  |  |
| 6       | <ul> <li>Raw materials</li> <li>All about data - Types of data</li> </ul>   |  |
| 7 - 8   | <ul> <li>All about models - Overview of statistical and data mining models</li> <li>All about technology - Technology stacks available</li> </ul>   |  |
| 9       | <ul> <li>Analytics framework</li> <li>CRISP-DM - Understanding the framework with an end to end project example</li> </ul>  |  |
| 10      | <ul> <li>Workshop on problem identification using CRISP-DM - Rosemann dataset EDA using Tableau</li> </ul>  |  |

# Introduction to Python

|   | Торіс  |
|---|--|
| 1 | Getting Started with Python  |
| 2 | LPHW - 1 - 17  |
| 3 | LPHW - 18 - 26   |
| 4 | LPHW - 27 - 39   |
| 5 | Handling Sequences         Additional info         1.       Data structures         2.       More data structure |
| 6 | Essential Python Libraries   |
| 7 | same as above  |
| 8 | Python Exam  |
| x | Data Science with Python   |
| x | Two more examples  |
|   |  |

# Machine Learning 1

### **Session Plan**

| Session 1:               | Introduction to machine learning  |
|--------------------------|---|
| Session 2:               | Understanding the data  |
| Session 3:               | Data Manipulations (Part 1)   |
| Session 4:<br>Tutorial 1 | Data Manipulations (Part 2)<br>Tutorial Session on Data manipulations   |
| Session 5:               | Dealing with dimensions (PCA)   |
| Session 6:               | Dealing with dimensions (SVD)   |
| Session 7:               | Kernel function and Kernel PCA  |
| Session 8:               | Dealing with missing values in data (part 1)                            |
| Session 9:<br>Tutorial 2 | Dealing with missing values in data (part 2)<br>Tutorial Session on EDA |

| Session 10:                | Numerical Methods (using Python)   |  |
|----------------------------|--|--|
| Session 11:                | Class Test   |  |
| Session 12:                | Understanding data clustering concepts and techniques (K-means and Hierarchical<br>clustering) |  |
| Session 13:                | Running clustering in Python with Silhouette distance measure for cluster validity             |  |
| Session 14:                | Hands on exercise  |  |
| Session 15:                | Understanding data classification concepts and techniques using decision trees (C4.5 and CART) |  |
| Session 16:                | Model building concepts and techniques (Split and Cross Validation); model validation          |  |
| Session 17:<br>Tutorial 3: | Hands on exercise<br>Tutorial on Clustering and Classification                                 |  |
| Session 18:                | Concepts of Association Rule Mining  |  |
| Session 19:                | Building association rules and interpretation  |  |
| Session 20:                | Class Test   |  |

# **Managerial Communication**

| Session 1 -<br>2 | <ul> <li>Topic: Communication - a theoretical framework</li> <li>Definition, Significance, Function, Process</li> <li>Communication flow and network structures</li> <li>Organizational Communication</li> </ul>    |
|------------------|---|
| Session 3        | <ul> <li>Topic: Listening Skills</li> <li>Barriers to Communication</li> <li>The importance of Listening Skills</li> <li>Types of Listening Modes – active listening</li> <li>Improving Listening Skills</li> </ul> |
| Session 4        | <ul> <li>Topic: Non Verbal communication</li> <li>Reading Non verbal Communication Cues</li> <li>Role of non-verbal cues</li> <li>Kinesics</li> </ul>   |
| Session 5 -<br>6 | Activity: Present a BA topic to the class   |
| Session 7 -<br>8 | <ul> <li>Topic: Presentation Skills (Content and form)</li> <li>Defining purpose</li> <li>Audience Analysis</li> <li>Organizing Contents &amp; Outlining</li> <li>Design &amp; Timing</li> </ul>                    |

| Session 9 -<br>12 | <ul> <li>Activity: Presentation Skills Workshop - learning by doing</li> <li>Preparing audience-centric presentations</li> <li>Agenda</li> <li>Outline</li> <li>Content</li> <li>Delivery</li> </ul>   |
|-------------------|--|
| Session 13        | <ul> <li>Topic: Selling and Negotiation Skills;</li> <li>Selling in a one-to-one and one-to-many situations</li> <li>Characteristics of a good salesperson</li> <li>Negotiating effectively</li> <li>Characteristics of a good negotiator</li> </ul>             |
| Session 14        | <ul> <li>Topic: Business oral and written communication</li> <li>Talking on the telephone</li> <li>Writing emails</li> <li>Conferencing</li> <li>Business Etiquette</li> <li>Social media Etiquette</li> <li>Conducting and participating in meetings</li> </ul> |

### **Market Research**

### **Course Content**

After **two initial sessions** on **"Introduction to Marketing Research"** the contents of the course can be broken up into 2 sections:

**1.** How to Design Market Research Studies (5 sessions): This would include inputs on "Problem Identification" in real-life Marketing situations with specific emphasis on how to Design a Research to find solutions to such Problems. Technicalities of research design and application of Statistics related areas (like Sampling and Sample Size Determination) will be covered extensively.

**2. Measurement Scales, Questionnaires, Data Cleaning and Data Preparation (3 sessions)**: This would cover technicalities of collecting information and preparing the data.

| Case<br>Asses | Study<br>sment | / | Readings | / | Group | Contents | Торіс | Ses<br>No. |
|---------------|----------------|---|----------|---|-------|----------|-------|------------|
|               |                |   |          |   |       |          |       |            |

| 2 Cases-lets ('client briefs') on areas<br>related to<br>(i) FMCG Product Test<br>(ii) Media Research   | Intro to Marketing Research?<br>Steps in the MR Process<br>QuantivsQuali Research /<br>Primary vs Secondary Research<br>Real life 'client briefs' will be<br>shared. Discussions on research<br>approach to the 'client briefs' | MR – An<br>Introduction,<br>Overview from a<br>Practitioner | 1 –<br>2 |
|---|---|---|----------|
| Case-let (on soap packaging) to<br>demonstrate management problem<br>and MR problem<br>Recommended Reading: Chapt 1 and 2<br>from text-book                                     | Management Decision Problem<br>and Marketing Research<br>Problem.<br>Formulating Research<br>Objectives.  | Problem<br>Identification                                   | 3        |
| <u>Case for Class Discussion</u> : Cases to<br>understand brand-switching patterns<br>in Longitudinal Studies   | Exploratory, Descriptive and<br>Causal design<br>Cross-Sectional vs Longitudinal<br>Study   | Research Design   | 4        |
| Outputs from Indian Pop Census and<br>IRS<br><u>Group Assignment 1</u> : Market selection<br>for an FMCG company using<br>Population Census data                                | Secondary data and working<br>with important Syndicated<br>researches in India (e.g. Indian<br>Population Census, IRS, TAM)   | Secondary Research  | 5        |
| <u>Case for class discussion</u> : Sampling<br>process of Indian Readership Study<br>(IRS)<br><u>Recommended Reading</u> : Chapt 11 of<br>text book                             | Census and Sample<br>Sampling Design Process<br>Probability and Non-Probability<br>sampling. Examples and<br>applications   | Sampling: Design,<br>Procedures and use<br>in MR            | 6 –<br>7 |
| <u>Group Assignment 2</u> : Sample size<br>determination for an ad hoc study of<br>HUL  | Practical considerations for<br>selecting sample size<br>Statistical approach of sample<br>size determination   | Sample Size<br>determination                                | 8        |
| <u>Case for Class Discussion</u> : Q're<br>preparation for a short section of<br>brand health study including Brand<br>Awareness, Usage, Intention to try or<br>repeat, Imagery | Information Collection as Flow<br>Charts<br>Routing, Wording of qns,<br>General Dos and Don'ts<br>Measurement Scales  | Questionnaire<br>Design, Data Entry,<br>Data Cleaning       | 9–<br>10 |

# **RDBMS and Data Warehousing**

| #   | Торіс  |  |
|-----|--|--|
| 1   | Introduction to DBMS / RDBMS   |  |
| 2   | Data Modelling - Entity Relationships  |  |
| 3   | Data Modelling - Normalization   |  |
| 4   | Physical Data Model  |  |
| 4A  | Accessing SQL  |  |
| 4B  | DDL Creating Tables, Loading Data, Insert, Delete, Update  |  |
| 5   | DML Basic Select   |  |
| 6   | DML Advanced Select  |  |
|     | SQL for Smarties   |  |
| 6A  | <ul> <li>Assignment : Pick out an unusual example of SQL usage. Demonstrate it by creating one or more tables with sample data. Create a Google Slide deck to explain the example. Embed the Slide Deck in a blog post and submit the URL of the blogpost in the form given here.</li> <li>Your deck should contain 4 slides <ol> <li>Brief overview of what is interesting about the SQL</li> <li>DDL required to create the schema and populate the tables</li> <li>The actual SQL to be executed</li> <li>The output generated</li> </ol> </li> </ul> |  |
| 7   | Practice   |  |
| 8   | Some Advanced Topics   |  |
| 9   | Introduction to Data Warehousing   |  |
| 10  | Dimensional Modelling :  |  |
| 11  | Dimensional Modelling (contd)  |  |
| 11A | DW with MySQL  |  |
| 12  | Olap Cubes   |  |
| 13  | QUIZ   |  |
| 14  | Practical BI   |  |
| 9A  | Business Analytics Approach  |  |

Extra stuff

| 14  | Practice                      |
|-----|-------------------------------|
| 14a |                               |
| 5   | Getting Started with SQL Lite |
| 9   | Even More SQL!                |
| 14  | Olap Cubes Practicals         |
| 14  | Even More SQL!                |
| 7B  | Review of Normalisation       |
| 14  | Practical BI                  |

### Statistics 1

### Session 1: Introduction to data

Data Basics (Observations, Variables and types, Relationship between variables) Overview of data collection principles (Populations and samples, Sampling, Explanatory and response variables, Observational studies and sampling strategies, Sampling methods) Experiments - Principles of experiment design

### Session 2 - 5: Introduction to data (continued)

Examining Numerical and Categorical data (Dot plots, Frequency Distribution and Histograms, Ogives, Stem and Leaf plot, Bar Charts, Pie Chart, Scatter Plot, Contingency Table, Segmented bar and mosaic plots) Outliers, Mean, Median, Mode, Percentile, Variance and Standard Deviation, Explanation of Standard Deviation in the context of Empirical Rule for variability of data, Coefficient of Variation, Box plots – quartiles and median, Box plots – outliers, Box plots – skewness, Robust statistics, Transformation of data (very briefly), Mapping data) Comparing numerical data across groups

### Session 6 – 8 : Introduction to Probability

Introduction (Defining probability, Mutually exclusive outcomes, Event, Sample space, Probability distribution, Probabilities when events are not disjoint, Addition rule, Complement, Independence)

Conditional probability (Definition, Contingency table, Marginal and joint probabilities, General multiplication rule, Tree diagrams)

### Bayes' Rule

Random variables (Expectation, Variability, Linear combinations, Variability in linear combinations) Continuous distribution (with reference to histogram) and probabilities

### Session 9 - 10: Continuous Distributions

Normal distribution (The distribution model, Standardizing with z scores, Normal probability table and the model with examples, Empirical rule of standard deviation for variability of data) Evaluating the normal probability plot (comparing with histogram, constructing a plot)

Session 11 - 13: Introduction to linear regression

Correlation, Line fitting, Fitted values, Residuals

Fitting a line by least squares regression, Standard Error of Estimate (Residual Standard Error), Using R-quare to describe the strength of a fit

Types of outliers in linear regression, Residual Analysis to test the assumptions of the Regression Model

Basic introduction to multiple regression

Session 14 – 15: Foundations for inference and estimation

Variability in estimates (Point estimates and their basic properties, Standard error of the mean) Sampling distribution of mean and proportion and the central limit theorem

Confidence intervals of population mean and population proportion (significance, interpretation, estimating the population parameter, different confidence levels, changing the confidence levels) Margin of error and ascertaining a sample size

Session 16 - 18: Foundations for inference and hypothesis testing Nearly normal population with known SD Hypothesis testing framework Two Tailed and One Tailed tests Testing hypothesis using confidence intervals, and Critical Z values One-sample means with the *t* distribution with unknown population SD (Normality condition, *t* distribution and its use as a solution the standard error problem, one-sample *t* confidence intervals, one-sample *t* tests) Inference for a single proportion (Normality condition, hypothesis testing, and sample size) Decision errors (Type 1 and 2) Hypothesis testing using p-values Choosing a significance level Power and the Type 2 error rate Statistical significance versus practical significance

Session 19 – 20:<u>Linear Regression and Multiple Regression</u> Hypothesis Tests for the Slope of the Regression Model and Testing the Overall Model Confidence Interval and Prediction Interval for the response variable Model building – a case with data

### Term 2

**Big Data Spark** 

|   | Торіс              |
|---|--------------------|
| 1 | "Big Data" concept |
|   | Intro to Colab     |
| 2 | Handling Sequences |
| 3 | Spark Intro        |

| 4 | Spark Tutorial "DailyShow" |
|---|----------------------------|
| 5 | Working with RDD           |
| 6 | RDD vs DataFrame           |
| 7 | Machine Learning           |
| 8 | Machine Learning           |

## **Customer Analytics**

| Case Study / Readings / Group<br>Assessment   | Contents  | Торіс  | Ses<br>No. |
|---|---|--|------------|
| Charts and tables (MR Agency<br>Outputs) for discussion   | Cross-tabulations and how<br>meaningful marketing implications<br>can be drawn using them<br>Need of higher-order statistical<br>tools            | Basics of data<br>analysis and<br>marketing reco<br>using analyzed<br>data | 1          |
| Tables from multiple-panel blind product tests  | Testing of Statistically Significant<br>Differences – for proportions and<br>means  | Test of significance<br>and ANOVA  | 2          |
| Demonstrations using raw data of a blind product test   | Multiple regression to find "Drivers<br>of brand preference"<br>SPSS navigations for familiarization  | Multiple<br>Regression   | 3          |
| Demonstrations using raw data of a blind product test   | Concept of dimension reduction<br>Principal Component Analysis<br>(PCA)   | Dimension<br>Reduction<br>Techniques                                       | 4          |
| Demonstrations using outputs of<br>psychographic traits   | Need for Factor Analysis, Basic principles, Reading Factor Outputs  | Factor Analysis  | 5          |
| Psychographic segmentation of<br>consumers of different financial<br>instruments – Raw data of a large<br>primary research will be used to<br>demonstrate Clustering  | Need for Cluster Analysis, Basic<br>principles<br>Psychographic segmentation<br>studies   | Cluster Analysis   | 6          |
| Cases:<br>1. Planning a positioning platform<br>for a face wash, using product test<br>data<br>2. Psychographic segmentation of<br>independent home builders in India | Basic principles of other tools like<br>Discriminant and Conjoint Analysis<br>Marketing problems / situations<br>and use of multivariate analysis | Application of<br>multivariate tools<br>in marketing<br>decisions          | 7 –<br>8   |

| The Analytics Process with special<br>emphasis to the data crunching<br>part will be demonstrated for<br>areas like<br>• Marketing-mix Modelling<br>• Campaign Management<br>• Market Basket Analysis | Applications in<br>Marketing<br>Analytics | 9 -<br>10 |
|---|---|-----------|
|---|---|-----------|

## **Data Visualization**

| #  | Торіс                                 |
|----|---------------------------------------|
| 1  | Overview                              |
| 2  | Introduction to Tableau               |
| 3  | Tableau - Continued                   |
| 4  | QlikView                              |
| 5  | QlikView                              |
| 6  | Getting Started : Basic Google Charts |
| 7  | Dynamic Charts with Google Docs       |
| 8  | Data Visualisation with R             |
| 8A | Google Visualisation with R           |
| 9  | Geomapping                            |
| 10 | GGPlot2 - visualisation with R        |
| x  | Dashboard with data from Quandl       |
| 11 | Presentation                          |

# Econometrics

| Session<br>No | Proposed Topic To be Covered   |
|---------------|--|
| 1             | Multiple and Multivariate Regression - theory and illustrative problem solving |

| 2 | Forward Insertion, Backward Removal and Stepwise Regression illustration using R, concepts of AIC and BIC   |
|---|---|
| 3 | Detection and Handling Multicollinearity in Regression, Tests of Multicollinearity,<br>Significance VIF   |
| 4 | Time Series Analysis, Creation of Time Series Object in R, Decomposition of Time Series,<br>Studying Trend, Seasonality, Randomness and Cyclical Pattern in a Time Series |
| 5 | Use of Holtwinters Function in R for Time Series Forecasting Problems, Illustrations using R, Study of the Residuals  |
| 6 | Concept of Stationarity in Time Series, Stationarity Test - Augmented Dickey Fuller Test  |
| 7 | Autoregressive, Moving Average and ARMA models, ARIMA models with different orders  |

## Introduction to R

| Session 1:  | Working with RStudio and simple R operations |
|-------------|--|
| Session 2:  | Understanding data structures in R           |
| Session 3:  | Conditional operations in R                  |
| Session 4:  | Working with loops in R                      |
| Session 5:  | Apply family functions and their uses        |
| Session 6:  | Working with strings                         |
| Session 7:  | Working with dates                           |
| Session 8:  | Basic R plots                                |
| Session 9:  | Working with ggplot package                  |
| Session 10: | Class test                                   |

# Machine Learning 2

| Session 1: | Clustering concepts (DBSCAN) |
|------------|------------------------------|
|------------|------------------------------|

| Session 2:  | Working with DBSCAN                               |
|-------------|---|
| Session 3:  | Clustering concepts (EM Clustering)               |
| Session 4:  | EM clustering with Python and interpretation      |
| Tutorial 1  | Tutorial on clustering                            |
| Session 5:  | Classification Technique (Naive Bayesian)         |
| Session 6:  | Using Naive Bayesian Techniques with Python       |
| Session 7:  | Classification Technique (KNN)                    |
| Session 8:  | Working with KNN and model comparison             |
| Session 9:  | Mini project 1                                    |
| Tutorial 2  | Tutorial on classification                        |
| Session 10: | Classification Technique (SVM)                    |
| Session 11: | Using SVM in Python for classification            |
| Session 12: | Ensemble techniques (Boosting and Bagging)        |
| Session 13: | Working with ensembles in Python                  |
| Session 14: | Classification and regression using Random Forest |
| Session 15: | Working with random forest in Python              |
| Tutorial 3  | Tutorial on advanced classification techniques    |
| Session 16: | Recommendation Systems (RS)                       |
| Session 17: | Building RS in Python                             |
| Session 18: | Mini project 2                                    |
| Session 19: | Kaggle inclass competition                        |
| Session 20: | Class Test  |

# **Model Interpretation**

| Session # | Topics Covered                  |
|-----------|---------------------------------|
| 1         | Concept of Predictive Modelling |

| 2 - 5   | Building Regression Models - Linear Models, Robust and Resistant Regression Models,<br>Subset Selection and Shrinkage Methods, Ridge Regression, Lasso and Least Angel<br>Regression, Partial Least Square Regression, Generalized Linear Models, Bagging,<br>Boosting, Random Forest. |
|---------|--|
| 6 - 9   | Building Classification Models - Logistic Regression, Linear Discriminant Analysis,<br>Classification Tree, Random Forests.  |
| 10 - 11 | Building Clustering Models - k- Means Clustering, Hierarchical Clustering  |
| 12      | Principal Component Analysis in model building   |
| 13      | Multi dimensional scaling  |

### **Statistics 2**

Session 1: Hypothesis Testing for Single Population (Revision): Solving for Type 2 error, Operating Characteristic, and Power Curve

Session 2 - 4: Hypothesis for Two Populations

Hypothesis Testing and Confidence Intervals about the differences in two means (using the *z*-Statistic), Hypothesis Testing and Confidence Intervals about the differences in two means: Independent Samples and Population Variances unknown, Statistical Inferences for two related populations, Statistical Inferences about two population proportions *Testing of Hypotheses about two population variances in detail* 

Session 5 – 7: Analysis of Variance

Concept of Design of Experiments, Completely Randomized Design (One-Way ANOVA), Multiple Comparison Tests, Randomized Block Design, Factorial Design (Two-Way ANOVA) - with two treatments

Session 8 - 10: Multiple Regression Analysis

The model with two explanatory variables or more, multiple regression equation, significance test of the overall model

Residuals, Standard error of estimate and coefficient of determination (R-square), Adjusted R-square etc.

Interpreting multiple regression software output

Session 11 - 12Multiple Regression Analysis (Advanced) Nonlinear methods (Polynomial regression, Tukey's ladder of transformations, Regression models with interaction, Model transformation), *Indicator (dummy) variables* 

Session 13 - 14: Other Distributions Binomial, Poisson, Hyper-geometric, Uniform distributionsetc Session 15 -16: Logistic Regression Concept, Logistic Regression Model, Testing overall model,

Session 17 – 18:<u>Chi-square Analysis</u> Chi-Square Goodness-of-Fit Test, Contingency Analysis: Chi-Square Test of Independence

Session 19 -20: Non-parametric tests

### Term 3

### Analytics in Finance

| Session 1-2 | Concept of Income & Expense, Asset & Liability:  |
|-------------|--|
|             | Measurement of profit<br>Different profits (EBITDA, EBIT, EBT, EAT, NOPAT, Gross Profit)<br>Calculation of cost of goods sold & gross profit |
|             | Types of assets and liabilities (Current & Noncurrent)   |
|             | Measure of equity  |
|             | Risk of equity   |
| Session 3-4 | Types of business decisions:<br>Operating, investing & financing decisions   |
|             | Profit & cash flow:<br>Linkage between profit & cash flow<br>Meaning of operating cash flow  |
| Session 5-6 | Profitability analysis:  |
|             | DuPont analysis and its break up<br>Relation between profit and asset utilisation efficiency, capital structure, margin                      |
|             | Liquidity & Solvency study:  |
|             | Use of financial ratios to measure liquidity & solvency<br>Meaning & interpretation  |

### **Bayesian Networks**

| Session No | Topics to be covered   |
|------------|--|
| 1 - 2      | Revision of Probability Theory   |
| 3          | Introduction to Bayesian Inference Models, Bayesian view of uncertainty.                                 |
| 4          | Learning in Bayesian Inference System - why Bayesian inference in machine learning?                      |
| 5 - 6      | Model overfitting and bias-variance tradeoff, Selecting models of optimum complexity, Bayesian Averaging |
| 7          | Bayesian Regression Models - Use of the "arm" package in R, simulation of posteriori distributions       |
| 8 - 9      | Bayesian Classification Models - Naive Bayes classifier, Bayesian logistic regression model              |
| 10 -11     | Bayesian Moles for Unsupervised Learning - Bayesian Mixture Models                                       |
| 12 - 13    | Bayesian Neural Networks - brnn package in R, Deep belief networks - concepts and illustrative examples  |

# **Deep Learning**

| Session | Topics to be covered  |
|---------|---|
| 1 - 2   | Fundamentals of neural networks: neurons and layers, activation function, different types of activation function, the backpropagation algorithm, an illustrative hand computed example. |
| 3       | Multiple Regression and logistic regression using neural networks   |
| 4 - 5   | Deep Learning fundamentals - feature learning, deep learning algorithms and various applications - speech recognition, object recognition and classification                            |
| 6       | Popular Open source libraries in Deep Learning - Theano, TensorFlow and Keras, A sample deep learning example code using Keras  |
| 7 - 8   | Unsupervised Feature Learning - Autoencoders, Regularization techniques for autoencoders, denoising autoencoders  |

| 9  | Boltzman Machine, Restricted Boltzman Machine, Implementation in TensorFlow - an example illustration |
|----|---|
| 10 | Recurrent Neural Networks and Language models - LSTM - example illustration                           |

# Introduction to SAS

| # | Торіс                            | Remarks   |
|---|----------------------------------|---|
| 1 | A SAS Table or Dataset           | <ul> <li>The SAS Program Data Vector (PDV)</li> <li>Variables &amp; Observations</li> <li>Libraries</li> </ul>  |
| 2 | The SAS Data Step                | <ul> <li>SAS Statements, Functions, and Procedures</li> <li>A SAS program</li> <li>Fundamentals of data processing by SAS</li> </ul>  |
| 3 | Reading/Extracting data into SAS | <ul> <li>FILENAME, INFILE statement for external files</li> <li>SET statement for reading SAS Files</li> <li>Basics of SAS/CONNECT to connect to external databases such as SQL Server, DB2, others.</li> </ul>                                 |
| 4 | SAS Statements                   | Core set of Statements and their options used in<br>manipulating data, e.g.;<br>- DATA<br>- LIBNAME<br>- INPUT, VAR<br>- ARRAY, DO Loops<br>- Assigning, manipulating variable values<br>- BY statement<br>- MERGE (joining 2 or more datasets) |
| 5 | SAS Functions                    | Core set of SAS Functions for transforming data:<br>- SUM<br>- SUBSTR<br>- PUT, INPUT<br>- INDEX<br>- SCAN<br>- RANUNI (For random sampling)  |

| 6  | SAS Dates & Formats                                     | <ul> <li>Date Functions</li> <li>Numeric &amp; Character Formats</li> <li>Creating a Format or Lookup Table.</li> <li>Using a Format for Lookup.</li> </ul>  |
|----|---|--|
| 7  | Defining SAS Macros                                     | <ul> <li>The Macro facility in SAS</li> <li>Global and Local Macros</li> <li>Creating a SAS Macro</li> <li>Invoking a Macro</li> </ul>   |
| 8  | PROC SQL<br>A short Test                                | <ul> <li>The SQL facility in SAS</li> <li>SQL Joins</li> <li>SAS Dictionary Tables</li> </ul>  |
| 9  | SAS Procedures: Exploratory                             | <ul> <li>The PROC SORT procedure with options</li> <li>PROC MEANS</li> <li>The PROC SUMMARY procedure with options</li> <li>Generating statistics such as MEAN, STD, VAR, etc.</li> <li>PROC GPLOT</li> <li>PROC DATASETS with options.</li> </ul> |
| 10 | SAS Procedures: Sampling,<br>Hypothesis Testing & ANOVA | PROC SURVEYSELECT<br>PROC STANDARD to standardize vars.<br>PROC ANOVA<br>PROC GLM for hypothesis testing. ANOVA, others  |
| 11 | SAS PROCEDURES: Regression                              | PROC CORR for Correlation.<br>PROC REG with options.   |
| 12 | Wrap Up & Test  | Review all material<br>A Final Test  |

### NoSQL

Lecture 1 & 2 NoSQL Introduction to NoSQL and History Where to use NoSQL Benefits of NoSQL List of NoSQL Databases Storage Types for NoSQL databases CAP and BASE Theorem Exercise MongoDB Introduction to MongoDB RDBMS vs. MongoDB NoSQL Operations in MongoDB Examples of MongoDB with various methods Exercise

### Lecture 3 & 4

Indexes in MongoDB with various techniques Indexing Strategies Atomic Operations Exercise Data Modelling with MongoDB References vs. Embedded Documents Tree structures Sharding fundamentals with illustration

### Lecture 5 & 6

Aggregation in MongoDB Aggregation Pipeline and MapReduce with examples Full Text Search in MongoDB with example Text Analytics Information Extraction Terminology and Approaches in Text Analytics MongoDB as Analytics Platform with examples Big Data Traditional vs. Big Data Approach Applications of Big Data Solutions

### **Optimisation Techniques**

| Session 1 | Linear Programming Problems (Graphical Method and Simplex)                                |
|-----------|---|
| Session 2 | Solving Linear Programming Problems using python  |
| Session 3 | Shortest route problem  |
| Session 4 | Dynamic Programming - Theory and illustrative problem solving using "Scperf" package in R |

| Session 5 | Queuing Theory - Theory and Problem Solving using "queuecomputer" package in R -<br>simulation of arrival pattern, service pattern and estimating various performance<br>metrics like average queue length, average throughput etc. |
|-----------|---|
| Session 6 | Project Management Using PERT and CPM - illustrative problem solving using the<br>"Concept Draw" tool.  |

### **Retail Analytics**

The following topics are to be covered in the course:

- Introduction to Retail, with focus on
  - Traditional retail business model
  - Franchising
  - Modern Format Retail
  - Private Labels
  - Online Retail
  - Developments in the Indian Retail market
- Supply Chain Analytics for Retail
- Assortment Optimization
- Forecasting
- Segmentation and Clustering
- Customer Loyalty Analytics
- Prediction of Customer Lifetime Value (CLV)

The above concepts will be taught through class lectures. Plus a large retail-transaction-dataset will be used, where concepts like Clustering, CLV, etc will be demonstrated using R

### **Text Analytics**

| Session 1: | Basics of text analysis processes       |
|------------|---|
| Session 2: | Web crawling                            |
| Session 3: | Web Scraping from downloaded html files |
| Session 4: | Text classification                     |
| Session 5: | Singular Value decomposition concept    |
| Session 6: | Latent Semantic Analysis                |
| Session 7: | Document clustering                     |
| Session 8: | Topic Modeling                          |
| Session 9: | Class Assignments                       |

Session 10: Presentation/Test

### **Capstone Project**

### **Session Plan**

### Defining analytics problem, EDA and data engineering

- Arriving at a analytical problem from a business problem
- Defining analytics objectives
- Exploratory data analysis to understand the business and customers
- Data engineering joins, outlier treatment, and smoothening

### Modelling

- Determining the statistical technique
- Build model
- Evaluate result (model accuracy)

### Presenting insights via visualization

- Deduce insights or story behind the numbers
- Structure insights
- Prepare a dashboard to storyboard the client

### Project

- 1. Analyzing customers of a supermarket
- 2. Prediction of diseases
- 3. Prediction of stock prices based on 5-min interval stock data
- 4. Prediction of various sectoral data of Indian economy and identifying their association and causal relationships.
- 5. Classification of potential customers of a bank