|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| EB-301 | **ESSENTIAL OF BUSINESS ANALYTICS** | **100** | **4** | **0** | **0** | **3** |

**COURSE OBJECTIVES** This course helps the students to understand and analyze basic essentials of business analytics business framework. They shall be exposed to fundamental statistical techniques to solve real life problems and enable them to take better decisions.

## Unit I:

Introduction to Business Analytics, Types of data, Integrating Analytics with business, Business Analytics for Competitive Advantage, Descriptive, Predictive, and Prescriptive Analytics, Dashboards History; Subdivisions within Statistics; Data collection, Editing, Classification, Tabulation, Diagrammatic and Graphical representation of data.

## Unit II:

Measures of Central tendency and Dispersion: Arithmetic Mean, Geometric Mean, Harmonic Mean, Median, Mode, Mean Deviation, Quartile Deviation, Standard Deviation, Skewness, Kurtosis and Moments.

## Unit III:

Probability and Probability Distributions: Introduction to Probability, Probability Rules, Probabilities under Conditions of Statistical Independence, Probabilities under Conditions of Statistical Dependence, Revising Prior Estimates of Probabilities, Bayes‘ Theorem, Random Variables, Use of Expected Value in Decision Making, Binomial Distribution, Poisson Distribution, Normal Distribution.

## Unit IV:

Sampling and Estimation: Random Sampling, Introduction to Sampling Distributions, Relationship Between Sample Size and Standard Error, Point Estimates, Interval Estimates, Confidence Intervals, Calculating Interval Estimates of the Mean from Large Samples.

## Unit V:

Testing of Hypotheses: Hypothesis, Steps in Hypothesis Testing, Measuring the Power of a Hypothesis Test, Hypothesis Testing of Means and Proportions, Hypothesis Testing for Differences between Means and Proportions, Analysis of Variance, One way ANOVA and Two way ANOVA, Non-parametric tests: Chi-Square Test, The Sign Test for Paired Data, The MannWhitney U Test, Kruskal-Wallis Test, The Kolmogorov-Smirnov test. Unit 6: Correlation, Regression and Time Series: Correlation, Product moment correlation, Rank correlation, Bi- variate correlation, Regression, Simple linear Regression, Line of best fit, Time Series, Trend Analysis, Cyclical Variation, Seasonal Variation, Irregular Variation, Time Series Analysis in Forecasting.

## References:

1. Richard I. Levin & David S.Rubin, Statistics for Management, PHI.1999, New Delhi.
2. Kishor S. Trivedi, Probability and Statistics with Reliability, Queuing and Computer Science Applications, John Wiley & Sons, Singapore, 2002.
3. John E.Freund & Ronald E. Walpole, Mathematical statistics, PH, New Jersey, 1980.
4. E.L.Lehmann, Testing Statistical Hypotheses, John Wiley & Sons, New York, 1986.
5. S.P. Gupta, Statistical Methods, Sultan Chand &Sons, New Delhi 1998.
6. Sundar Rao P.S.S, Richard J, Introduction to biostatistics – A manual for students in Health Sciences, PHI Learning Pvt. Ltd. 1996, New Delhi.
7. Susan Miltan , Statistical methods in the Biological and Health Sciences,1999, McGraw-Hill

8.B. Burt Gerstman, Basic Biostatistics: Statistics for Public Health Practice, Jones & Bartlett Learning, 2008.

9.Wayne W. Danial, John Wiley, Biostatistics: A Foundation for analysis in the Health Sciences

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **EB-302** | **Text, Social Media& Web Analytics** | **100** | **4** | **0** | **0** | **3** |

## UNIT I:

Text Analytics: Text Analytical Approach and Tools to Analyze Data: Analytical Approaches-- History of Analytical Tools-- Introducing Popular Analytical Tools-- Comparing Various Analytical Tools. Text mining –unstructured text, episode rule discovery for texts, hierarchy of categories, text clustering.

## UNIT II:

Social Media Analytics: Introduction to Semantic Web: Limitations of current Web-- Development of Semantic Web-- Emergence of the Social Web. Social Network analysis: Development of Social Network Analysis -Key concepts and measures in network analysis. Electronic sources for network analysis: Electronic discussion networks- Blogs and online communities - Web-based networks.

## UNIT III

Knowledge representation on the Semantic web: Ontology and their role in the Semantic Web: Ontology-based knowledge Representation – Ontology languages for the Semantic Web: Resource Description Framework - Web Ontology Language. Modelling and aggregating social network data: State-of-the-art in network data representation - Ontological representation of social individuals - Ontological representation of social relationships - Aggregating and reasoning with social network data. Social-semantic applications: Generic Architecture- Sesame- Elmo – Graph util, Flink-Open academia. Social network extraction: Survey method-electronic data extraction- Data collection Optimiztionprediction- Evaluation.

## UNIT IV:

Understanding web analytics: The foundations of Web analytics: Techniques and Technologies – Present and Future of Web analytics.---Data Collection: Importance and Options –Web server log files: Click stream data – User submitted information – Web server performance data – Page tags –First and third party tracking - Web Analytics Strategy: Key performance indicators – Web analytics process – Heuristics evaluations – Site visits – Surveys – Measuring reach – Measuring acquisition – Measuring conversion – Measuring retention – Security and privacy implications of Web analytics.

## UNIT V:

Web Analytics Tools: Content organization tools – Process measurement tools – Visitor segmentation tools – Campaign analysis tools – Commerce measurement tools – Google analytics – Omniture – Web trends – Yahoo! Web analytics. Google Analytics: Key features and capabilities – Quantitative and qualitative data - Working of Google analytics – Privacy - Tracking visitor clicks, Outbound links and Non HTML files.

## Reference Books:

1. Bernard J. Jansen, ―Understanding User-Web Interactions via Web analytics‖, Morgan and Claypool, 2009.
2. Avinash Kaushik, ―Web Analytics2.0‖, John Wiley and Sons, 2010.
3. Brian Clifton, ―Advanced web metrics with Google analytics‖, John Wiley and Sons, 2012.
4. Justin Cutroni, ―Google Analytics‖, O‟Reilly, 2015.
5. Jerri L. Ledford, Joe Teixeira and Mary E. Tyler, ―Google Analytics‖, John Wiley and Sons, 2013.
6. Charu C. Aggarwal and ChengXiang Zhai, Mining Text Data. Springer, 2012.
7. Dan Jurafsky and James H Martin,Speech & Language Processing. Pearson Education India, 2000.
8. Guandong Xu, Yanchun Zhang and Lin Li, Web Mining and Social Networking – Techniques and applications, First Edition, Springer, 2011.
9. Dion Goh and Schubert Foo - Social information Retrieval Systems: Emerging Technologies and Applications for Searching the Web Effectively, IGI Global Snippet, 2008.
10. Max Chevalier, Christine Julien and Chantal Soulé-Dupuy, Collaborative and Social Information Retrieval and Access: Techniques for Improved user Modelling, IGI Global Snippet, 2009.
11. John G. Breslin, Alexander Passant and Stefan Decker, -The Social Semantic Web, Springer, 2009.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **EB-303** | **PREDICTIVE ANALYTICS** | **100** | **4** | **0** | **0** | **3** |

Course Objective: This course will enable students to apply specific statistical and regression analysis methods applicable to predictive analytics to identify new trends and patterns, uncover relationships, create forecasts and to develop and use various quantitative and classification predictive models based on various regression and models.

## Unit I

Linear Regression: Coefficient of determination-- Significance test, Residual analysis - Standard Error - Ratio of variance- Galton Graph – Ratio of Regression – Interpretation of Galton‘s Graph

* Confidence and Prediction intervals.

## Unit II

Multiple Linear Regression: Coefficient of determination--Interpretation of regression coefficients-- Categorical variables— heteroscedasticity - Multi-co linearity outliers-- Auto regression and Transformation of variables—Regression--Model Building.

## Unit III

Logistic And Multinomial Regression: Logistic function-- Estimation of probability using Logistic regression, Variance-- Wald Test-- Hosmer Lemshow Test-- Classification Table-- Gini Co-efficient.

## Unit IV

Forecasting: Moving average-- Exponential Smoothing-- Casual Models. Time Series Analysis-- Moving Average Models-- ARIMA models-- Multivariate Models.

## Unit V

Index numbers: construction of Index numbers – selection of items- selection of base – selection of average and system of weighting – construction of various types of index numbers. Theory of probability ad sampling: statistical probability – the Laws of probability – permutations and combinations.

## Reference Books:

* 1. Anderson, Sweeney and Williams ―Statistics for business and economics‖, Cengage Learning, 2011.
  2. Richard I. Levin. David S. Rubin, ―Statistics for Management‖, Pearson Education, 2012.
  3. Richard A. Johnson, Irwin Miller and John Freund, ―Probability and Statistics for Engineers‖, Pearson Education, 2014.
  4. Ronald E. Walpole, Raymond H. Meyers, Sharon L. Meyers, ―Probability and Statistics for Engineers and Scientists‖, Pearson Education.
  5. Asthana B.N., ―Elements of Statistics‖ Chaitanya publishing house, Allahabad.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **EB-304** | **BIG DATA ANALYTICS** | **100** | **4** | **0** | **0** | **3** |

**COURSE OBJECTIVES** : • Understand the Big Data Platform and its Use cases • Provide an overview of Apache Hadoop • Provide HDFS Concepts and Interfacing with HDFS • Understand Map Reduce Jobs • Provide hands on Hodoop Eco System • Apply analytics on Structured, Unstructured Data. • Exposure to Data Analytics with R.

## Unit I

**Introduction to Big Data:** Big Data-definition, Characteristics of Big Data (Volume, Variety,Velocity, Veracity, Validity), Importance of Big Data , Patterns for Big Data Development, Data in the Warehouse and Data in Hadoop [Zikopoulos] - Introduction to Hadoop**:** Hadoop- definition, Understanding distributed systems and Hadoop,Comparing SQL databases and Hadoop, Understanding MapReduce, Counting words with Hadoop—running your first program, History of Hadoop, Starting Hadoop - The building blocks of Hadoop, NameNode, DataNode, Secondary NameNode, JobTracker and Task Tracker.

## Unit II

**HDFS:** Components of Hadoop -Working with files in HDFS, Anatomy of a MapReduceprogram, Reading and writing the Hadoop Distributed File system -The Design of HDFS, HDFS Concepts, The Command-Line Interface, Hadoop Filesystem, The Java Interface, Data Flow, Parallel Copying with distcp, Hadoop Archives. Hadoop I/O: Compression— Serialization-- Avro and File-Based Data structures.

## Unit III

**MapReduce Programming:** Writing basic Map Reduce programs - Getting the patent data set,constructing the basic template of a Map Reduce program, Counting things, Adapting for Hadoop‘s API changes, Streaming in Hadoop. MapReduce Advanced Programming: Advanced MapReduce - Chaining Map Reduce jobs,joining data from different sources.

## Unit IV

Hadoop Eco System --User Defined Functions-- Data Processing operators. Hive : Hive Shell-- Hive Services-- Hive Metastore-- Comparison with Traditional Databases—HiveQL-- Tables, Querying Data and User Defined Functions. Hbase : HBasics—Concepts—Clients—Example-- Hbase Versus RDBMS. Big SQL : Introduction

## Unit V

**Graph Representation in MapReduce:** Modeling data and solving problems with graphs,Shortest Path Algorithm, Friends-of-Friends Algorithm, PageRank Algorithm, BloomFilters. Data Analytics with R Machine Learning : Introduction, Supervised Learning, Unsupervised Learning, Collaborative Filtering. Big Data Analytics with BigR.

## References

1. Tom White ― Hadoop: The Definitive Guide‖ Third Edit on, O‘reily Media, 2012.
2. Seema Acharya, Subhasini Chellappan, "Big Data Analytics" Wiley 2015.
3. Michael Berthold, David J. Hand, "Intelligent Data Analysis‖, Springer, 2007.
4. Jay Liebowitz, ―Big Data and Business Analytics‖ Auerbach Publications, CRC press (2013)
5. Tom Plunkett, Mark Hornick, ―Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle R Enterprise and Oracle R Connector for Hadoop‖, McGraw- Hill/Osborne Media (2013), Oracle press.
6. Anand Rajaraman and Jef rey David Ulman, ―Mining of Massive Datasets‖, Cambridge University Press, 2012.
7. Bill Franks, ―Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics‖, John Wiley & sons, 2012.
8. Glen J. Myat, ―Making Sense of Data‖, John Wiley & Sons, 2007
9. Pete Warden, ―Big Data Glossary‖, O‘Reily, 2011.
10. Michael Mineli, Michele Chambers, Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley Publications, 2013.
11. ArvindSathi, ―BigDataAnalytics: Disruptive Technologies for Changing the Game‖, MC Press, 2012
12. Paul Zikopoulos ,Dirk DeRoos , Krishnan Parasuraman , Thomas Deutsch , James Giles , David Corigan , "Harness the Power of Big Data The IBM Big Data Platform ", Tata McGraw Hill Publications, 2012.