Shiromani Gurudwara Parbandhak Committee's

Guru Nanak Khalsa College of

Arts, Science and Commerce (Autonomous)

Matunga, Mumbai – 400 019, Maharashtra, India

Syllabus for M.Sc. Semester I and II

Program: MSc in Big Data Analytics

Course: Computer Science

(M.Sc. Computer Science –Part-1 Syllabus as per NEP-2020)

(To be implemented from 2023-24)

About the Course

Data Science is the extraction of actionable knowledge directly from data through a process of discovery, hypothesis, and analytical hypothesis analysis.

A Data Scientist is a practitioner who has sufficient knowledge of the overlapping regimes of expertise in business needs, domain knowledge, analytical skills and programming expertise to manage the end-to-end scientific method process through each stage in the big data lifecycle.

Recently Economic Times has reported that India's demand for data scientists grows over 400%. Based on our market research and conversation with the industry, we have identified Data Science and Big Data Analytics as one of the sectors with critical supply demand imbalance.

The Course will be divided into 4 semesters spread across 2 years. The emphasis will be on solving business problems across various domains using necessary Statistical and ML Tools and Techniques. Course will focus on most widely used tools like Python, R, Tableau, Hadoop, Spark, SPSS etc.

Course Objectives

• Understand and use Big Data Technologies to solve business problems.

• Start your career as Data Scientist, Data Analyst, Business Intelligence Developer, or Data Engineer.

• To produce Data Scientists in diverse domains who are at the forefront of the coming AI revolution.

Course Highlights

• One-on-One Mentoring by expertise from Academics / Industry veterans

• 100 % Placements

• Project based Learning

• Domain Specialization

Who should take this course Data Analyst Course?

Aspiring professionals of any educational background with an analytical frame of mind are best suited to pursue the Data Analyst course, including:

• IT professionals

• Banking and finance professionals

• Marketing managers

• Sales professionals

• Supply chain network managers

• Beginners in the data analytics domain

• Students in UG/ PG programs

Eligibility criterion

• Candidates who pass B.Sc. (I.T./C.S./Mathematics/Statistics/Physics/Biotech) /B.E. (I.T./C.S.)/ B.Tech (I.T. /C.S.) /MCA are eligible to seek admission in M.Sc. in Big Data Analytics. OR

• The candidate must have pursued a Bachelor's degree in relevant specialization with minimum 50% aggregate score.

• Management students are also eligible to take admission.

Course Outcome

After the successful completion of the degree the student can be placed in various multinational companies belonging to various sectors like Banking & Finance, Healthcare, Technology, IT etc.

Some of the job titles include –

• Data Scientist

• Data Analyst

• Data Engineer

• Data Architect

• Business Intelligence Manager

• Business Intelligence Developer

• Business Intelligence Analyst

• Business Analyst

Batch Size: 20 students

Pre-requisites for joining the course

1. Microsoft Excel for Data Analysis
    a. Excel Tables, Filters, Sorting
    b. Pivot Tables and Charts
    c. Formats, Formulas, Dates
    d. Functions – Mathematical, Statistical, Text, Date
2. Basic UNIX Programming
    a. Basic UNIX Commands
    b. Handling files and folders
    c. Concatenation, find and replace, modify file & texts
    d. Basic summary commands
3. Basics of R Programming language and R studio.
4. Professionals wishing to succeed in this Data Analyst course should have basic knowledge of mathematics.

Semester –I (Total credit: 20)

The following table gives the details of the theory courses in Semester –I

Sr. no	CourseType	Course Title	Theory credits	Practical credits
1	MandatoryPaper-I	Statistical methods, Linear algebra and Computing for Data Science	04 credit	02 credit
2	MandatoryPaper-II	Python for Data Analytics	04 credit	02 credit
3	MandatoryPaper-III	Sports Analytics	02 credit	-
4	ElectivePaper	Database Management and Data Mining	03 credit	01 credit
		Data Analysis	03 credit	01 credit
5		Research Methodology	04 credit	-

Semester –II (Total credit: 20)

The following table gives the details of the theory courses in Semester –II

Sr. no	Course Type	Course Title	Theory credits	Practical credits
1	Mandatory Paper-I	Advance statistical methods with time series	04 credit	02 credit
2	Mandatory Paper-II	Machine Learning	04 credit	02 credit
3	Mandatory Paper-III	Human Resource Analytics	02 credit	-
4	Elective Paper	Foundations of data science (programming for big data)	03 credit	01 credit
	Elective Paper	Enabling Technologies for data science-I	03 credit	01 credit
5		On job training		04 credit

Semester –I

Course Code	Course Title	Course credit
GNKPSBDA1501	**Statistical methods, Linear algebra and Computing for Data Science**	**04**
Unit 1	**Data Collection & Visualization** Concepts of measurement, scales of measurement, design of data collection formats with illustration, data quality and issues with date collection systems with examples from business, cleaning and treatment of missing data. **Linear Algebra** Linear equations and matrices, matrix operations, solving system of linear equations, Concept & Computation of determinant and inverse of matrix, Eigen values and Eigen vectors	**15L**
Unit 2	**Basic Statistics** Frequency table, histogram, measures of location, measures of spread, skewness, kurtosis, percentiles, box plot, correlation and simple linear regression. **Contingency Tables:** Two way contingency tables, measures of association, testing for dependence.	**15L**
Unit 3	**Core Java Concepts** Introduction to Java programming, Object-oriented programming concepts, Interface, Exception Handling, Packages, Threads **Data Structure & Concepts of Computation Using Java** Algorithms, Convergence, Complexity with illustrations, some sorting & searching algorithms, some numerical methods	**15L**
Unit 4	**Computing Methodologies Using R** Monte-Carlo simulations of random numbers and various statistical methods, memory handling strategies for big data. **Basic Probability** Concepts of experiments, Outcomes, Sample space, Events, Principle of inclusion & exclusion, Conditional probability, Independence, Bayes Theorem. **Probability Distribution** Random Variables: discrete and continuous probability models, Binomial distribution	**15L**

Suggested Books:

1. Statistics: David Freedman, Pobert Pisani & Roger Purves, WW.Norten & Co. 4th Edition 2007.
2. The visual display of Quantitative Information: Edward Tufte, Graphics Press, 2001.
3. Best Practices in Data Cleaning: Jason W. Osborne, Sage Publications 2012.

4. Introduction to Data Science (Data Analysis and Prediction Algorithms with R), Rafael A. Irizarry, https://rafalab.github.io/dsbook/
5. Hands-On Programming with R - Write Your Own Functions and Simulations, Grolemund Garrett, O'Reilly
6. Linear Algebra and Its Application: Gilbert Strang, 4th Edition, Academic Press.

Course Code	Course Title	Course credit
GNKPSBDA1P501	**Statistical methods, Linear algebra and Computing for Data Science**	**02**
Note : *Practical's to be conducted using Java and R, Python*		
1. Basic of statistics: Frequency table, skewness, kurtosis etc. 2. Implement Data visualization 3. Implement Object-oriented programming in Java 4. Implement Sorting and searching algorithms in Java 5. Implement Exception handling in Java 6. Implement Multithreading in Java 7. Implement Monte Carlo Simulation 8. Implement Eigen Values and Eigen vectors. 9. Computing probability 10. Implement Binomial distribution		

Course Code	Course Title	Course credit
GNKPSBDA2501	**Python for Data Analytics**	**04**
Unit 1	**Basic Concepts** Introduction to Python interpreter, Control statements, Data Types **Writing Functions** Defining a function, calling a function, passing by value or reference, anonymous function	**15L**
Unit 2	**File Handling** Opening and Closing Files, Reading and Writing Files, Directories in Python **Packages** What are Packages? Import package	**15L**
Unit 3	**Exception Handling** Python errors and Built-in exceptions, user defined exceptions, exception handling	**15L**
Unit 4	**OO Programming Concepts** OOP, class, Inheritance, overloading **Python libraries for Big Data** NumPy, Pandas, SciPy etc	**15L**
Suggested Books: 1. Core Python Programming: Dr. R. Nageswara Rao, Dream Tech, Second Edition		

2. Python for Everybody: Exploring Data in Python 3: Charles Severance
3. Python Cookbook: Recipes for Mastering Python 3: David Beazley, 3rd Edition

Course Code	Course Title	Course credit
GNKPSBDA2P501	**Python for Data Analytics**	**02**

Note : *Practical's to be conducted using Python*

1. Programs based on Data Types.
2. Programs based on Functions.
3. Programs based on File Handling.
4. Programs based on Packages.
5. Programs based on Control Structures.
6. Programs based on exception handling.
7. Programs based on Classes and objects.
8. Programs based on Inheritance.
9. Programs based on Overloading.
10. Working on Big Data libraries: NumPy, Pandas, Matplotlib etc.

Course: MSc-I Big Data Analytics
Semester-I Paper
Course Title: Sports Analytics
Course Code: GNKPSBDA00000
Credits: 3
No of lectures (Hours): 45
Marks: 100 (75:25)

Course Objectives:

Sr. No.	Course objectives
The course aims at:	
1	To understand the importance of data analytics in sports and ethical considerations and challenges in sports data analytics

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed
CO 1	Develop skills in collecting, cleaning, and managing sports data	PO1	-	U,Ap,An
CO 2	Gain proficiency in using statistical analysis techniques to analyze sports data	PO1,PO2	PS02	U,Ap
CO 3	Apply data visualization methods to present sports data effectively	PO2,PO3	PSO2	U,Ap,An
CO 4	Learn how to apply predictive modeling techniques to sports data	PO1,PO4	PS04	U,Ap,An

Unit		Title	No. of lectures	CO Mapping

Unit 1			**15**	
	1.1	Introduction to Sports Data Analytics Overview and Evolution of analytics in sports, Importance and applications of sports data analytics		**CO 1**
	1.2	Data Collection and Preprocessing Sources of sports data, Database management for sports analytics. Creating effective visualizations for sports data, Interactive dashboards for sports analytics		**CO 2**
	1.3	**Data Visualization for Sports Analytics:** Statistical Analysis in Sports,Descriptive statistics for sports data, Hypothesis testing in sports analytics, Regression analysis in sports, Analysis of variance (ANOVA) in sports		**CO 3**
Unit 2			**15**	
	2.1	**Advanced Techniques in Sports Data Analytics** Predictive Modeling, Feature selection and engineering, Linear regression models and Classification models for sports outcomes		**CO 4**
	2.2	Machine Learning in Sports Analytics: Decision trees and random forests, Support vector machines and Neural networks and deep learning in sports analytics		**CO 5**
	2.3	Advanced Topics in Sports Data Analytics: Sports performance analysis, Player tracking and motion analytics, Sports marketing, Sports injury prediction and prevention Sports Business Analytics: Revenue generation and marketing in sports, Fan engagement and customer analytics		**CO 6**

References:

1. Sports Analytics: A Guide for Coaches, Managers, and Other Decision Makers by

Benjamin C. Alamar, Columbia university press, 2013
2. Sports Analytics and Data Science: Winning the Game with Methods and Models by

Thomas Miller, 1st edition, Pearson FT Press, 2015
3. Sports Analytics: Analysis, Visualisation and Decision Making in Sports Performanceby

Daniel Memmert, Tim McGarry, and Tony Reilly, 2018
4. Cricket Analytics: Analytics and Data Science in Cricket by Tapan Bagchi and S.

Raghunathan
5. Machine Learning using Python by Manaranjan Pradhan and U. Dinesh Kumar, Weily ,

2020

URL for Online Study Material –

1. www.coursera.org
2. www.kaggle.com
3. www.datacamp.com

4. www.sportsanalyticsinstitute.com

Course Code	Course Title	Course credit
GNKPSBDA3A501	**Database Management and Data Mining**	**03**
Unit 1	**Basic Concepts** Different data models, ER and EER diagram, schema, table, Big Data Concepts and Hadoop Ecosystem	**15L**
Unit 2	**Relational and Non-Relational Databases** Structure, various operations, normalization, SQL, No- SQL, Graph Database, Parallel and distributed database, Map-Reduce. Lab using SQL/Oracle/MySql for Relational databases; Hadoop(any), MangoDB, GraphDB for Big Data	**15L**
Unit 3	**Implementation** ORACLE SQL/MS SQL/MySQL, Hadoop Ecosystem, Concept of database security **Introduction to data mining** Knowledge discovery from databases, Data Mining Functionalities	**15L**
Suggested Books: 1. Database system concepts: Abraham Silberschartz, Henry F. Korth and S. Surarshan, McGraw Hill, 2011. 2. Hadoop 2 Quick-Start Guide: Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem, Douglas Eadline, Addison-Wesley, Pearson Education India; First edition (1 March 2016) 3. Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, EMC Education Services, 2015		

4. Data Mining: Concepts and Techniques: Jain Pei, Jiawei Han, Micheline Kamber , 3rd Edition,2012

Course Code	Course Title	Course credit
GNKPSBDA3AP501	**Database Management and Data Mining**	**01**

Note : *Practical's to be conducted using weka, R, SQL Server and upcoming technologies*

1. For given scenario:
Draw E-R diagram and convert entities and relationships to table.
2. Perform the following:
i. Creating a Database
ii. Viewing all databases
iii. Viewing all Tables in a Database
iv. Creating Tables (With and Without Constraints)
v. Inserting/Updating/Deleting Records in a Table
3. Demonstrate the Accessing and Storing and performing CRUD operations in
i. MongoDB
ii. Redis
4. Demonstrate the Accessing and Storing and performing CRUD operations in
i. HBase
ii. Apache Cassandra
5. Demonstrate the indexing and ordering operations in
i. MongoDB
ii. CouchDB
iii. Apache Cassandra
6. Knowledge discovery from databases part-1:
i. Create tables using different applications.
ii. Develop an application to design a warehouse by importing various tables from external sources.
7. Develop an application to create dimension tables in a cube and form
i. Star schema
ii. Snowflake schema
iii. Parent-Child schema
8. Develop an application to
i. Demonstrate operations like roll-up, drill-down, slice, and dice
ii. Pre-process data imported from external sources.

Course Code	Course Title	Course credit
GNKPSBDA3B501	**Data Analysis**	**03**
Unit 1	**Predictive Analysis** What and Why Analytics, Introduction to Tools and Environment, Application of Modeling in Business Types of data and variables, Data Modeling Techniques, Missing imputations etc. Regression analysis	**15L**

	Sale Analytics Sales data and common challenge, Effective Sales analysis, types of Social analysis, tools to analysis sale data	
Unit 2	**Social Media Analytics** Introduction to Social Media Data, Visualizing and Modeling Patterns in Social Media Data, Social Media Networks, Monitoring Customer Engagement in Social Media	**15L**
Unit 3	**Sport Analytics** Introduction to Sport Analytics, Need and use of Sport Analytics, Use of Analytics in different games, Gambling and Betting Market, EPL forecasting	**15L**
Suggested Books: 1. Textbook: Szabo, G., G. Polatkan, O. Boykin & A. Chalkiopoulus (2019), *Social Media Data Mining and Analytics*, Wiley, ISBN 978-1-118-82485-6 2. Predictive & Advanced Analytics (IBM ICE Publication)		

Course Code	**Course Title**	**Course credit**
GNKPSBDA3BP501	**Data Analysis**	**01**
Note : *Practical's can be conducted using Python, R*		
1. Sale analysis with Power BI or Tableau 2. Implement analysis of Social Networking site: Twitter 3. Forecasting EPL result 4. Gambling and betting odd analysis 5. Case study of health analysis 6. Case study of Netflix's Content Recommendation System 7. Case study of Airbnb's Dynamic Pricing Strategy 8. Case study of Fraud Detection in Financial Transactions		

Course Code	**Course Title**	**Course credit**
GNKPSBDA4501	**Research methodology**	**04**
Unit 1	**Research Fundamentals and Terminology:** Meaning and Objective of research, features of a good research study, types of Research (qualitative and quantitative research) Study designs and variations: basic, applied, historical, exploratory, experimental, ex-post-facto, case study, diagnostic research, crossover design, case control design, cohort study design, multifactorial design. **Literature Survey Methods**	**15L**

	Journal and abbreviation, current tites and review, monographs, textbooks, introduction to abstract, Beilstein, subject and author index Digital: Web sources, E-journals, Journal access, TOC alerts, Hot articles, Citation Index, Impact factor, H-index, E- consortium, UGC Infonet, E-books, Internet discussion groups and communities, Blogs, preprint servers, Search engines, Sciurus, Google Scholar, Wiki-databases, Science Direct, SciFinder, Scopus	
Unit 2	**Research writing** Scientific writing- Reporting practical and project work, writing literature surveys and reviews, organizing a poster display, giving an oral presentation. Writing Scientific Papers: Justification for scientific contributions, bibliography, description of methods, conclusions, the need for illustration, style, publications of scientific work. Project Proposal and research funding agencies, Research grants, scholarships and funding (CSIR, DBT, DST, DST-INSPIRE Fellowship, ICMR, INSA, BRNS, MoEFCC, UGC- RFSMS, Fulbright Fellowships for Indian students, Lady Tata Memorial Trust, EPA, Bill and Melinda Gates Foundation, Welcome Trust, Erasmus Mundus) **Publication ethics and Bibliography** Publication ethics : definition, introduction Best practices/ Standards settings initiative and guidelines COPE, WAME Conflict of interest Publication Misconduct: definition, concept, problems that lead to unethical behavior Violation of publication ethics, authorship and contributor ship, Identification of publication misconduct, Predatory publisher and journals Use of reference management software (MS Word / Zotero / Mendeley)	**15L**
Unit 3	**Research Design** Choosing appropriate research methods, Experimental, observational, case study, survey, etc., Sampling techniques and sample size determination, Validity, and reliability in research **Data Collection Methods** Surveys and questionnaires, Interviews: structured, semi-structured, unstructured, Observations and participant observations, Ethnographic studies and fieldwork, Data Collection Tools and Technologies	**15L**

	Online survey platforms Data collection software, Sensor-based data collection, Wearable devices and Internet of Things (IoT) for data collection, Data Analysis Techniques **Quantitative data analysis using statistical tools (e.g., SPSS, R)** Qualitative data analysis: coding, thematic analysis, content analysis, Mixed-methods data analysis	
Unit 4	**Research Presentation and Communication** Writing research papers and technical reports, creating effective research presentations, Peer review process and responding to feedback, presenting research findings at conferences, Research Presentation and Communication **Writing research papers and technical reports** Creating effective research presentations, Peer review process and responding to feedback Presenting research findings at conferences	**15L**

Semester –II

Course Code	Course Title	Course credit
GNKPSBDA1502	**Advance Statistical Methods with time series**	**04**
Unit 1	**Estimation** Unbiasedness, Consistency, UMVUE, Maximum likelihood estimates. **Test of Hypotheses** Two types of errors, test statistic, parametric tests for equality of means & variances.	**15L**
Unit 2	**Linear Model** Gauss Markov Model, least square estimators, Analysis of variance. **Regression** Multiple linear regression, forward, backward & stepwise regression (practical's only), Logistic Regression.	**15L**
Unit 3	**Review of Linear Programming** Non-Linear Programming. Assignment Models. Transportation Models.	**15L**
Unit 4	**Queuing Models** Characteristics of Queuing Process, Poisson Process, Birth-Death Process, Single-Server Queues, Multi-Server Queues, Queues with Truncation, Finite- Source Queues, Numerical Techniques & Simulation. **Introduction to Time Series:** Components of time series, Smoothing auto correlation, stationarity, concepts of AR, MA, ARMA & ARIMA models with illustrations	**15L**

Suggested Books:

1. Statistical Inference: P. J. Bickel and K. A. Docksum, 2nd Edition, Prentice Hall.
2. Introduction to Linear Regression Analysis: Douglas C. Montgomery
3. Operations Research: Prem Kumar Gupta & D. S. Hira
4. Fundamentals of Queuing Theory: Donald Gross, John F. Shortle, James M. Thompson & Carl M. Harris, Fourth Edition, Wiley

Course Code	Course Title	Course credit
GNKPSBDA1P502	**Advance Statistical Methods with time series**	**02**

1. Computing Summary Statistics using real time data
2. Testing of hypothesis for large sample tests for real-time problems.
3. Testing of hypothesis for small sample tests for One and Two Sample mean and paired comparison

4. Testing of hypothesis for Small Sample tests for F-test
5. Implementation of Gauss Markov Model/ least square estimation model
6. Performing ANOVA (one-way and two-way) for real dataset.
7. Applying stepwise linear regression models to real dataset and interpreting the coefficient of determination for scale data.
8. Implementation of Logistic Regression.
9. Understanding of Queuing models with Single-Server and multi-Server.
10. Implement time series analysis with ARMA/ARIMA model.

Course Code	**Course Title**	**Course credit**
GNKPSBDA2502	**Machine Learning**	**04**
Unit 1	**Machine Language Overview**: Applications of Machine Learning Algorithms, Steps involved in Machine Learning, Types of machine learning. **Linear Regression**: simple linear regression, multiple linear regression	**15L**
Unit 2	**Preparing data for classification:** removing outliers, handling missing data, normalizing the data, dimensionality reduction, handling skewed data, using large datasets **Resampling Methods:** cross-validation, the Bootstrap, percentage split	**15L**
Unit 3	**Classification using Nearest Neighbors:** k-NN algorithm **Probabilistic Classifiers:** generative (Naïve Bayes) and conditional(Logistic) classifiers **Evaluating Model Performance:** different performance evaluation metrics	**15L**
Unit 4	**Neural Networks** Representation Learning, Different Models like single and multi-layer perceptron, back propagation, Application. **Support Vector Machines** Model, Large Margin Classification, Kernels, SVMs in practice.	**15L**

Suggested Books:

1. Machine Learning: Tom Mitchell
2. Pattern Recognition and Machine Learning: Christopher Bishop, Springer,2006
3. An Introduction to Statistical Learning: Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer,2015
4. Python Machine Learning: Sebastian Raschka,2015

Course Code	Course Title	Course credit
GNKPSBDA2P502	**Machine Learning**	**02**

1. Perform Data Cleaning.
2. Fit a classification model using K Nearest Neighbour (KNN) Algorithm on a given data set. [One may use data sets like Caravan, Smarket, Weekly, Auto and Boston].
3. Classification using Naïve Bayes algorithms.
4. Implement simple and multiple linear regression model on a standard data set and plot the least square regression fit. Comment on the result. [One may use inbuilt data sets like Boston, Auto etc]
5. Perform cross-validation types.
6. Implement logistic regression on standard dataset and evaluate using accuracy, precision, recall and F1 score metrics.
7. Fit a support vector classifier for a given data set.
8. Study of different performance evaluation metrics.
9. Evaluate the performance of a model:
(i) Boosting
(ii) Bagging
(iii) Random Forest
10. Implement Neural Network.

Course: MSc-II Big Data Analytics
Semester-I Paper
Course Title: Human Resource Analytics
Course Code: GNKPSBDA00000
Credits: 3
No of lectures (Hours): 45
Marks: 100 (75:25)

Course Objectives:

Sr. No.	Course objectives
The course aims at:	
1	To analyze problems and issues in HR and the relevance of HR analytics.

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed
CO 1	Logically synthesize the tools, methods and techniques of HR analytics to understand real world corporate scenario.	PO1	-	U,Ap,An
CO 2	Identify the application and uses of HR analytics in various HR sub-systems.	PO1,PO2	PS02	U,Ap

Unit		Title	No. of lectures	CO Mapping

Unit 1			**15**	
	1.1	**HR Measurement:** Need for HR Measurement, Significance and concept of HR Analytics, HR Analytics and business linkages, Prerequisites of HR Analytics		**CO 1**
	1.2	**Models and frameworks of HR Analytics;** Measuring intellectual capital, need and rationale for HR Accounting & Audit, Approaches and methods of HR Accounting & Audit		**CO 2**
	1.3	**HRIS for HR Analytics:** What is Human Resource Information System; Role of HRIS in analytics; HRIS development and Implementation, the development process- need analysis, systems design, structure and culture; HRIS Applicationsmaking HRIS work.		**CO 3**
Unit 2			**15**	
	2.1	HR Analytics for Staffing, Training & Development, Performance Management Systems, Career Planning Systems, Rewards and Compensation Management, Employee Relations Systems.		**CO 4**
	2.2	Analytics for HR system: HR performance frameworks and measurement systems; Measuring HR Climate and People Management Capabilities; Competency Management Frameworks & Competency Mapping, Integration of competency-based HR System. Measuring HR Effectiveness		**CO 5**
	2.3	The HR Scorecard Trends and Future Challenges:Technology and changes in HR Analytics, Role of social media, Big Data and Predictive Analytics in HR, Assessing the effectiveness of HR Analytics, Post analysis steps, Review and monitoring, Issues in HR valuation and measurement; Emerging challenges: Global and Indian Experience		**CO 6**

References:

1. Ulrich, D. & Brockbank, W., The HR Value Proposition. Harvard Business School Press 2016
2. How to measure HRM by Jac Fitz-enz 2002
3. Predictive Analytics for Human Resources by Jac Fitz-enz, John Mattox II, Wiley 2014
4. Making Human Capital Analytics Work: Measuring the ROI of Human Capital Processes and Outcomes. By by Jack Phillips,Patricia Pulliam Phillips- 2014

Course Code	Course Title	Course credit
GNKPSBDA3A502	**Foundations Of Data Science (Programming For Big Data)**	**03**
Unit 1	**Graph Theory** Basic Concepts, Algorithms for connectedness, shortest path, Minimum Sampling Tree, Random Graphs Large graphs, G(n,p) model, Giant Component, Connectivity, Cycles, Non-Uniform models, Applications.	**15L**
Unit 2	**High Dimensional Space** Properties, Law of large numbers, Sphere and cube in high dimension, Generating points on the surface of a sphere, Gaussians in High dimension, Random projection, Applications. Singular Value Decomposition (SVD) Best rank k approximation, Power method for computing the SVD, Applications.	**15L**
Unit 3	**Random Walks** Reflection Principle, Long leads, Changes of Sign, Illustrations. Algorithm for Massive Data Problems Frequency Moments of data streams, matrix algorithms	**15L**

Suggested Book: Foundations of Data Science: John Hopcroft & Ravindran Kannan.

Course Code	Course Title	Course credit
GNKPSBDA3AP502	**Foundations Of Data Science (Programming for Big Data)**	**01**
1. Implement Minimum sampling tree. 2. Implement random graphs showing large graphs and cycles. 3. Implement non-uniform models. 4. Implement random projections in high dimensional space. 5. Show implementation of Sphere and cube in high dimensional space 6. Implement best rank k approximation. 7. Implement and discuss working of SVD. 8. Implement SVD on any suitable dataset and show power method for computing SVD on the dataset.		

Course Code	Course Title	Course credit
GNKPSBDA3B502	**Enabling Technologies for Data Science-I**	**03**
Unit 1	**Big data and Hadoop:** Hadoop architecture, Single node & Multi-node Hadoop, Hadoop commands, Hadoop daemon, Task instance, Hadoop Ecosystem and its installation, Illustrations.	**15L**
Unit 2	**Map-Reduce:** Framework, Developing Map-Reduce program, Life cycle method, Serialization, Running Map-Reduce in local and pseudo- distributed mode, Illustrations. **HIVE:** Data types and commands, Illustrations. **SQOOP:** Importing data, exporting data, Running, Illustrations	**15L**
Unit 3	**PIG:** Schema, Commands, Illustrations **NoSQL database:** Features, Types, NoSQL vs. SQL, Advantages and Disadvantages **Oozie:** What is Oozie? Workflow, packaging and deploying an Oozie workflow application, Features.	**15L**
Suggested Book: 1. Hadoop The Definitive Guide : Tom White , 4th Edition, 2017 2. Hadoop in Action : Chuck Lam, 2010 3. Data-intensive Text Processing with Map Reduce : Jimmy Lin and Chris Dyer, Morgan & Claypool Publishers, 2010		

Course Code	Course Title	Course credit
GNKPSBDA3BP502	**Enabling Technologies for Data Science-I**	**01**

1. Multi-node Hadoop eco-system configuration with HDFS
2. Improving Map-Reduce performance using combiners
3. Creating map reduce jobs
4. Create Bloom filter
5. Creating Map-Reduce Programs in local and pseudo-distributed mode
6. Working of HADOOP with HIVE
7. Working of HADOOP with SQOOP
8. Working of HADOOP with PIGs

Course Code	Course Title	Course credit
GNKPSBDA4502	**On Job Training**	**04**

A learner has to show an on job training of 60 hours with proper documentation.
-Joining letter
-Job training letter