

Progressive Education Society's
**Modern College of Arts, Science and
Commerce,**
Shivajinagar, Pune 5

(An Autonomous College Affiliated
to Savitribai Phule Pune University)

Detailed Syllabus
For Post Graduate Diploma in
Data Science
(2020-21 Course)
(with effect from 2020-21)

Abbreviations:

CIA:Continuous Internal Assessment **TH:** Theory **TUT:** Tutorial **PR:**Practical
AEC:Ability Enhancement Courses **SECT:** Skill Enhancement Courses Theory
ESE: End Sem Exam **DSET:**Discipline Specific Elective Theory
DSEP:Discipline Specific Elective Practical

Program Structure:

Year	Number of subjects per Semester	Number of lectures per Semester	Number of credits per Subject	Total Credits
Sem -I	3 Theory Courses	48 (4 Lectures/Week)	4	12
	2 Lab Courses	(4 Lab Hours/Week)	4	8
Sem -II	3 Theory Courses	48 (4 Lectures/Week)	4	12
	2 Lab Courses (including 1Project)	(4 Lab Hours/Week)	4	8
Total Credits:				40

Eligibility: Any Bachelor's degree in Science / Technology / Engineering faculty with minimum 50% marks for Open Category and minimum 45% for Reserved Category.

Duration: The entire Programme is a One Year and Two Semester Full Time Programme.

Objective:The objective of the Programme is to produce trained Software professionals with hands on experience and specialization in Data Science.

Course pre-requisites:

- 1) Basic Computer Knowledge
- 2) Programming Knowledge

NOTE: Bridge courses will be offered at the beginning of a Semester for students who have not completed the desired pre-requisites for courses listed in the current Semester. This includes Introductory lectures followed by self learning.

Examination:

The conduct of evaluation will be done by the College in the following manner:

- 1) Paper setting will be done by the respective teacher(s) and be handed over to CEO of the College
- 2) The Examination will be conducted strictly as per the Examination Time Table
- 3) The assessment will be done by the respective teacher(s) within 30 days time under the directives of the CEO/Principal

Evaluation of Students:

- 1) The In-semester and End-Semester examinations will be of 50 marks each
- 2) There shall be reevaluation of answer script of End semester examination, but not of Internal assessment papers as the Internal assessment comprises of various activities

Continuous Internal Assessment (CIA/CA):

Internal assessment for each course would be continuous and dates for each tutorials/practical tests will be pre-notified in the time table for teaching or placed separately as a part of time table. A teacher may select a variety of the procedures for Continuous Internal Assessment(CIA) suggested as follows:

- a) Mid-term test
- b) On-line test
- c) Open book test (concerned teacher will decide the books allowed)
- d) Tutorial
- e) Surprise test
- f) Oral
- g) Theory Assignments

- h) Review of Research paper
- i) Seminar presentation
- j) Group Discussion
- k) Programming Assignments

It is the Student's responsibility to preserve the documentation of the Internal assessment except midterm test answer script.

Completion of Degree Program:

- 1) As soon as a student earns 40 credits, the student will be deemed to have completed the requirements of the 'Post Graduate Diploma in Data Science' Programme
- 2) If a student has failed in a subject, then the said subject will not be taken into account for calculating GPA and overall grade. In fact, all the subjects in which a student has passed will be taken into account for calculating the GPA and overall grade
- 3) The policies and procedures decided will be followed for the conduct of examinations and declaration of the result of a candidate

Semester I

Year / Semester	Subject	Paper code	Title of Paper	Hrs per week	Credits	IA	ES	Total
Year – I Semester – I	CCT-1	20CsCdsP101	Python for Data Science	4	4	50	50	100
	CCT-2	20CsCdsP102	Mathematics and Statistics for Data Science	4	4	50	50	100
	CCT-3	20CsCdsP103	Big Data Analytics	4	4	50	50	100
	CCP-1	20CsCdsP104	Lab on Python	4	4	50	50	100
	CCP-2	20CsCdsP105	Lab on Big Data Analytics	4	4	50	50	100

Semester II

Year / Semester	Subject	Paper code	Title of Paper	Hrs per week	Credits	IA	ES	Total
Year – I Semester – II	CCT-1	20CsCdsP201	Data Analytics	4	4	50	50	100
	CCT-2	20CsCdsP202	Business Intelligence	4	4	50	50	100
	CCT-3	20CsCdsP203	Machine Learning	4	4	50	50	100
	CCP-1	20CsCdsP204	Lab on Machine Learning	4	4	50	50	100
	CCP-2	20CsCdsP205	Project (on Data Science)	4	4	50	50	100

Total Credits for Post Graduate Diploma in Data Science :40

Semester –I

Progressive Education Society's
Modern College of Arts, Science and Commerce (Autonomous)
Shivajinagar, Pune - 5

Post Graduate Diploma in Data Science (2020 Course)

Course Code :20CsCdsP101

Course Name :Python for Data Science

Teaching Scheme: TH:4Hours/Week

Credits :04

Examination Scheme: CIA : 50 Marks

End-Sem : 50 Marks

Prerequisites:

- An understanding of programming in an imperative language (e.g., C/C++, Java)
- Knowledge of basic algorithms and data structures (e.g., sorting, searching, lists, stacks, and trees)

Course Objectives:

- An understanding of programming language paradigm
- Learning functional programming language Python

Course Outcomes:

On completion of the course, student will be able to–

- Students can solve problems by using Python language
- Students can implement projects by using Python

Course Contents

Unit No.	Title	No. of Lectures
Unit 1	Introduction to programming languages	2
	1.1 The Art of Language Design 1.2 The Programming Language Spectrum 1.3 Why Study Programming Languages? 1.4 Programming Environments 1.5 Declarative style of programming,	
Unit 2	Introduction To Python	2
	2.1 Installation and 2.2 Working with Python	

	2.3 Understanding Python variables 2.4 Python basic Operators 2.5 Understanding python blocks	
Unit 3	Python Data Types	2
	3.1 Declaring and using Numeric data types: int, float, complex 3.2 Using string data type and string operations 3.3 Defining list and list slicing 3.4 Use of Tuple data type	
Unit 4	Python Program Flow Control	5
	4.1 Conditional blocks using if, else and elif 4.2 Simple for loops in python 4.3 For loop using ranges, string, list and dictionaries 4.4 Use of while loops in python Loop manipulation using pass, continue, break and else 4.5 Programming using Python conditional and loops Block	
Unit 5	Python Functions	3
	5.1 Modules And Packages 5.2 Organizing python codes using functions 5.3 Organizing python projects into modules Importing own module as well as external modules 5.4 Understanding Packages	
Unit 6	Python String, List And Dictionary Manipulations	5
	6.1 Building blocks of python programs 6.2 Understanding string in build methods 6.3 List manipulation using in build methods 6.4 Dictionary manipulation Programming using string, list and dictionary in build functions	
Unit 7	Python File Operation	5
	7.1 Reading config files in python 7.2 Writing log files in python 7.3 Read functions, read(), readline() and readlines() 7.4 Write functions, write() and writelines() 7.5 Manipulating file pointer using seek Programming using file operations	
Unit8	Python Regular Expression	6
	8.1 Powerful pattern matching and searching Power of 8.2 Pattern searching using regex in python 8.3 Real time parsing of networking or system data using regex Password, email, url validation using regular expression 8.4 Pattern finding programs using regular expression	
Unit 9	Python Database Interaction	8
	9.1 SQL Database (SQLite) connectivity 9.2 Creating and searching tables, Reading and storing config information in database 9.3 Reading and storing config information in database 9.4 Basic SQL queries	

Unit 10	Python Libraries	8
	10.1 Numpy 10.2 Pandas 10.3 Matplotlib 10.4 Scipy(Only Introduction) 10.5 Plotly 10.6 Seaborn	
Unit 11	Experiential Learning	2
	11.1 Analysis of all Functional programming with respect to Python 11.2 Analysis and study of Libraries provided by Python to support Data science	

References:

1. Functional Programming: Practice and Theory by Bruce J. MacLennan • ISBN-10: 0201137445 • ISBN-13: 978-0201137446
2. An Introduction to Functional Programming Through Lambda Calculus (Dover Books on Mathematics) Paperback by Greg Michaelson • ISBN-10: 0486478831 • ISBN-13: 978-0486478838
3. Computational Semantics with Functional Programming by Jan van Eijck (Author), Christina Unger (Author) • ISBN-10: 0521757606 • ISBN-13: 978-0521757607
4. Introduction to Computer Science Using Python: A Computational Problem-Solving Focus by Charles Dierbach
5. Programming Languages: Principles and Practice By Kenneth C. Louden • ISBN-10: 1575864967 • ISBN-13: 978-1575864969
6. LEARNING TO PROGRAM WITH PYTHON by Richard L. Halterman
Python 3 Object-oriented Programming Second Edition by Dusty Phillips

Progressive Education Society's
Modern College of Arts, Science and Commerce (Autonomous)
Shivajinagar, Pune - 5

Post Graduate Diploma in Data Science (2020 Course)

Course Code :20CsCdsP102

Course Name :Mathematics and Statistics for Data Science

Teaching Scheme: TH:4Hours/Week

Credits :04

Examination Scheme: CIA : 50 Marks

End-Sem : 50 Marks

Prerequisites:

- Knowledge of basic Mathematics and Statistics

Course Objectives:

- To understand the underlying Mathematical and Statistical Concepts in Data Science

Course Outcomes:

On completion of the course, student will be able to–

- Learning to describe basic features of the data in a study
- Provide brief summary about the sample using different quantitative measures
- To fit predictive models for the sample data
- Develop analytical thinking by using the ability to see a problem or solution from different points of view.
- To find chance of an event based on prior knowledge of conditions that might be related to the event.
- To apply different forms of probability distribution when the values of observed data are discrete and continuous.

Course Contents:

Unit No.	Title	No. of Lectures
Unit 1	Significance of Mathematics and Statistics in Data Science	1
Unit 2	Vectors	6
	2.1 Vector: Vector addition, Scalar Vector multiplication, Inner Product 2.2 Complexity of Vector Computations 2.3 Linear Functions: Linear Functions, Taylor Approximation 2.4 Regression Model Norms and Distance: Norm distance, Standard deviation, Angle, Complexity 2.5 Clustering: Clustering, a clustering Objective, The K means algorithm, Examples and Applications 2.6 Linear Independence: Linear Dependence, Basis, Orthonormal 2.7 Vectors	
Unit 3	Matrices	6
	3.1 Introduction to matrices 3.2 Zero and identity Matrices, Transpose, addition and norm 3.3 Matrix Vector Multiplication, Complexity 3.4 Matrix Examples: Geometric Transformation, Selectors, Incidence Matrix and Convolution 3.5 Linear Equations: Linear and affine functions, Linear function models, System of Linear Equations 3.6 Matrix Multiplication: Matrix Multiplication, Composition of Linear Functions, Matrix Power and QR Factorization 3.7 Matrix Inverses: Left and right inverses, Inverse, Solving Linear Equations, Examples, Pseudo Inverse	
Unit 4	Least Squares	3
	4.1 Least Squares :Least Squares Problem, Solution, Solving Least Squares Problems, Examples 4.2 Least squares data fitting: Least Squares data fitting, Validation, Feature Engineering. Least Squares 4.3 Classification: Classification, Least Squares Classifier, Multi-classifiers 4.4 Multi Objective Least Squares: Multi Objective Least Squares, Control, Estimation and Inversion, Regularized data fitting	
	Statistics :	

Unit 5	Descriptive Statistics:	6
	<p>5.1 Measures of Central Tendency: Mean, Median, Mode</p> <p>5.2 Partition Values: Quartiles, Percentiles, Box Plot</p> <p>5.3 Measures of Dispersion: Variance, Standard Deviation, Coefficient of variation</p> <p>5.4 Skewness: Concept of skewness, measures of skewness</p> <p>5.5 Kurtosis: Concept of Kurtosis, Measures of Kurtosis (All topics to be covered for raw data using R software)</p> <p>5.6 Bias, Variance</p>	
Unit 6	Introduction to Probability:	8
	<p>6.1 Probability - classical definition, probability models, axioms of probability, probability of an event</p> <p>6.2 Concepts and definitions of conditional probability, multiplication theorem $P(A \cap B) = P(A) \cdot P(B A)$</p> <p>6.3 Bayes' theorem (without proof)</p> <p>6.4 Naïve Bayes</p> <p>6.5 Concept of Posterior probability, problems on posterior probability</p> <p>6.6 Definition of sensitivity of a procedure, specificity of a procedure. Application of Bayes' theorem to design a procedure for false positive and false negative</p> <p>6.7 Concept and definition of independence of two Events</p> <p>6.8 Random Forest</p>	
Unit 7	Statistical Testing	5
	<p>7.1 Basic Terminology</p> <p>7.2 Confidence Intervals</p> <p>7.3 Hypothesis Testing</p> <p>7.4 Chi-Square Tests, P-test, Z-test,</p> <p>7.5 P value, T-Test</p>	
Unit 8	Types of Data	6
	<p>8.1 Quantitative Discrete data, Continuous data (Interval scale, Ration scale)</p> <p>8.2 Qualitative Nominal data, Ordinal data</p>	
Unit 9	Special Distributions	6
	<p>9.1 Normal Distribution</p> <p>9.2 Standard Normal Distribution</p> <p>9.3 Pareto Distribution (using R software)</p>	
Unit 10	Experiential learning	1

Reference Books:

Mathematics

1. Introduction to Applied Linear Algebra Vectors, Matrices and Least Squares by Stephen Boyd (Stanford University) and Lieven Vandenberghe (University of California, Los Angeles) Cambridge University Press

Statistics:

1. Fundamentals of Applied Statistics (3rd Edition), Gupta and Kapoor, S.Chand and Sons, New Delhi, 1987.
2. An Introductory Statistics, Kennedy and Gentle.
3. Statistical Methods, G.W. Snedecor, W.G. Cochran, John Wiley & sons, 1989.
4. Introduction to Linear Regression Analysis, Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, Wiley
5. Modern Elementary Statistics, Freund J.E., Pearson Publication, 2005.
6. Probability, Statistics, Design of Experiments and Queuing theory with applications Computer Science, Trivedi K.S., Prentice Hall of India, New Delhi, 2001.
7. A First course in Probability 6th Edition, Ross, Pearson Publication, 2006.
8. Introduction to Discrete Probability and Probability Distributions, Kulkarni M.B., Ghatpande S.B., SIPF Academy, 2007.
9. A Beginners Guide to R, Alain Zuur, Elena Leno, Erik Meesters, Springer, 2009
10. Statistics Using R, Sudha Purohit, S.D.Gore, Shailaja Deshmukh, Narosa, Publishing Company

Progressive Education Society's
Modern College of Arts, Science and Commerce (Autonomous)
Shivajinagar, Pune - 5

Post Graduate Diploma in Data Science (2020 Course)

Course Code :20CsCdsP103

Course Name :Big Data Analytics

Teaching Scheme: 4Hours/Week

Credits :04

Examination Scheme: CIA : 50 Marks

End-Sem : 50 Marks

Prerequisites:

- Basic Concepts of Database Management System and Structured Query Language

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 Hadoop Eco System
- Apply analytics on Structured, Unstructured Data.
- Exposure to Data Analytics with R.

Course Outcomes:

On completion of the course, student will be able to–

- Identify Big Data and its Business Implications.
- List the components of Hadoop and Hadoop Eco-System
- Access and Process Data on Distributed File System
- Manage Job Execution in Hadoop Environment
- Develop Big Data Solutions using Hadoop Eco System
- Analyze InfosphereBigInsights Big Data Recommendations.
- Apply Machine Learning Techniques using R.

Course Contents:

Unit No.	Title	No. of Lectures
Unit 1	INTRODUCTION TO BIG DATA AND HADOOP	12
	1.1 Types of Digital Data 1.2 Introduction to Big Data, 1.3 Big Data Analytics 1.4 History of Hadoop, Apache Hadoop 1.5 Analysing Data with Unix tools, 1.6 Analysing Data with Hadoop, 1.7 Hadoop Streaming, 1.8 Hadoop Echo System, 1.9 IBM Big Data Strategy, 1.10 Introduction to Infosphere Big Insights and Big Sheets	
Unit 2	HDFS(Hadoop Distributed File System)	10
	2.1 The Design of HDFS 2.2 HDFS Concepts 2.3 Command Line Interface 2.4 Hadoop file system interfaces 2.5 Data flow 2.6 Data Ingest with Flume and Scoop and Hadoop archives 2.7 Hadoop I/O: Compression, Serialization, Avro and File-Based Data structures	
Unit 3	Map Reduce	10
	3.1 Anatomy of a Map Reduce Job Run 3.2 Failures, Job Scheduling 3.3 Shuffle and Sort, Task Execution, Map Reduce Types and Formats, Map Reduce Features	
Unit 4	Hadoop Eco System	15
	4.1 Pig :Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined Functions, Data Processing operators. 4.2 Hive :Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, Querying Data and User Defined Functions. 4.3 Hbase :HBasics, Concepts, Clients, Example, Hbase Versus RDBMS 4.4 Big SQL : Introduction	
Unit 5	Experiential learning	1
	Use of Bit search CV	

Reference Books:

1. Tom White “ Hadoop: The Definitive Guide” Third Edit on, O’reily Media, 2012.
2. SeemaAcharya, SubhasiniChellappan, "Big Data Analytics" Wiley 2015.
3. Michael Berthold, David J. Hand, "Intelligent Data Analysis”, Springer, 2007.
4. Jay Liebowitz, “Big Data and Business Analytics” AuerbachPublications, 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. AnandRajaraman and Jeffrey 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. MichaelMineli, Michele Chambers, AmbigaDhiraj, "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 ,DirkDeRoos , Krishnan Parasuraman , Thomas Deutsch , James Giles , David Corigan , "Harness the Power of Big Data The IBM Big Data Platform ", Tata McGraw Hill Publications, 2012.

Progressive Education Society's
Modern College of Arts, Science and Commerce (Autonomous)
Shivajinagar, Pune - 5

Post Graduate Diploma in Data Science (2020 Course)

Course Code :20CsCdsP104

Course Name :Lab on Python

Teaching Scheme: TH:4Hours/Week

Credits :04

Examination Scheme: CIA : 50 Marks

End-Sem : 50 Marks

Assignment 1	Introduction To Python
	Installation of Python on different OS Working with Python as a calculator
Assignment 2	Programs on Flow Control
	Basic programs for understanding of different control flow in Python
Assignment 3	Functions
	Writing Programs using functions Use of Modules Use of packages
Assignment 4	Python programs for String, List
	Python programs for String, List
Assignment 5	Dictionary Manipulations
	Dictionary manipulation Programming using string, list and dictionary in build functions
Assignment 6	Python File Operation
	Reading config files in python Writing log files in python read functions, read(), readline() and readlines() write functions, write() and writelines()
Assignment 7	Python Object Oriented Programming
	Object Oriented programs

Progressive Education Society's
Modern College of Arts, Science and Commerce (Autonomous)
Shivajinagar, Pune - 5

Post Graduate Diploma in Data Science (2020 Course)

Course Code :20CsCdsP105

Course Name :Lab on Big Data Analytics

Teaching Scheme: TH:4Hours/Week

Credits :04

Examination Scheme: CIA : 50 Marks

End-Sem : 50 Marks

Assignment 1	Installation
	Installation of tools required for bigdata (i) Perform setting up and Installing Hadoop in its two operating modes: <ul style="list-style-type: none"> • Pseudo distributed, • Fully distributed. (ii) Use web based tools to monitor your Hadoop setup.
Assignment 2	File Management
	(i) Implement the following file management tasks in Hadoop: <ul style="list-style-type: none"> • Adding files and directories • Retrieving files • Deleting files ii) Benchmark and stress test an Apache Hadoop cluster
Assignment 3	Map Reduce
	Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm. <ul style="list-style-type: none"> • Find the number of occurrence of each word appearing in the input file(s) • Performing a MapReduce Job for word search count (look for specific keywords in a file)
Assignment 4	Stop word elimination problem
	(i) Implement the following file management tasks in Hadoop: <ul style="list-style-type: none"> • Adding files and directories • Retrieving files • Deleting files ii) Benchmark and stress test an Apache Hadoop cluster
Assignment 5	Map Reduce with a given dataset
	Write a Map Reduce program that mines weather data. Weather sensors collecting data every hour at many locations across the globe gather large volume of log data, which is a good candidate for analysis with MapReduce, since it is

	<p>semi structured and record-oriented.</p> <ul style="list-style-type: none"> • Find average, max and min temperature for each year in NCDC data set? • Filter the readings of a set based on value of the measurement, Output the line • of input files associated with a temperature value greater than 30.0 and store it in a separate file
Assignment 6	Case studies
	<p>On a given Dataset: Instead of breaking the sales down by store, give us a sales breakdown by product category across all of our stores What is the value of total sales for the following categories? Toys, Consumer Electronics</p>
Assignment 7	Pig Apache Spark
	<p>Install and Run Pig then write Pig Latin scripts to sort, group, join, project, and filter your data. Write a Pig Latin scripts for finding TF-IDF value for book dataset Install and Run Hive then use Hive to create, alter, and drop databases, tables, views, functions, and indexes. Install, Deploy & configure Apache Spark Cluster. Run apache spark applications using Python Data analytics using Apache Spark on Amazon food dataset, find all the pairs of items frequently reviewed together</p>

Semester –II

Progressive Education Society's
Modern College of Arts, Science and Commerce (Autonomous)
Shivajinagar, Pune - 5

Post Graduate Diploma in Data Science (2020 Course)

Course Code :20CsCdsP201

Course Name :Data Analytics

Teaching Scheme: TH:4Hours/Week

Credits :04

Examination Scheme: CIA : 50 Marks

End-Sem : 50 Marks

Prerequisites:

- Knowledge of probability theory, statistics and programming

Course Objectives:

- Able to apply fundamental algorithmic ideas to process data
- Learn to apply hypotheses and data into actionable Predictions
- To understand Data Analytics Life Cycle and Business Challenges

Course Outcomes:

On completion of the course, student will be able to–

- Deploying the Data Analytics Lifecycle in data analytics projects
- Selecting appropriate data visualizations to clearly communicate analytic insights to business sponsors and analytic audiences

Course Contents:

Unit No.	Title	Number of Lectures
Unit 1	Introduction to Data Science	12
	1.1 Basics of Data What is Data Science? Data science process Stages in data science project 1.2 Basics of Data Analytics Types of Analytics – Descriptive, Predictive,	

	<p>Prescriptive</p> <p>1.3 Data Preprocessing, Normalization and Transformation, Use of BeautifulSoup Tool</p> <p>1.4 Exploratory data Analysis</p> <p>Feature engineering: missing value treatment, outliers identification, duplicate rows/columns, data types - continuous, finding important columns/variables/dimensions, scaling (Min- MAX), Standard scaler, OHE, Dummification</p> <p>1.5 Statistical Inference</p> <p>Populations and samples, Statistical modeling, Probability, Distribution, Correlation, Regression</p>	
Unit 2	Data Optimization	12
	<p>2.1 Hyper parameter tuning using GridsearchCV</p> <p>2.2 RandomSearchCV</p> <p>2.3 Stacking, boosting, Ensembling</p> <p>2.4 Feature Engineering</p> <p>2.5 Concept of Over fitting, Under fitting, Balance/Imbalanced data, Cross Validation, Normalization</p> <p>2.6 Web scrapping in 'Data Analytics' add tool 'Beautiful SOAP'</p>	
Unit 3	Data Analytics with Python Programming	12
	<p>3.1 Numpy</p> <p>Arrays, Array indexing</p> <p>Data types</p> <p>Numpy operations</p> <p>Array math</p>	

	<p>Broadcasting</p> <p>3.2 SciPy</p> <p>Image operations</p> <p>Distance between points</p> <p>3.3 Data analysis and manipulation using Pandas</p> <p>Package</p> <p>Importing Data, Creating A DataFrame, DataFrame Methods,</p> <p>Data Operations, Pandas SQL operations,</p> <p>Indexing DataFrames, Boolean Indexing,</p> <p>Indexing using Labels, Multi-Indexing,</p> <p>Merge, Joining, ConcatenatingDataFrames,</p> <p>Sorting DataFrames,</p> <p>Date and time operations in DataFrames,</p> <p>Input-output in Pandas,</p> <p>Pandas settings,</p> <p>Apply Function,</p> <p>Pivot Table, Crosstab, Group By</p> <p>Iterating over rows of a Dataframe</p> <p>Plotting with Pandas</p> <p>3.4Matplotlib</p> <p>3.5 Seaborn</p>	
Unit 4	Data Visualization	10
	<p>4.1 Basic Principles</p> <p>Ideas and tools for data visualization, Graph visualization, Data Summaries, Model Checking and Comparison</p> <p>Purpose of Visualization</p> <p>Multidimensional Visualization</p> <p>Tree Visualization</p> <p>Graph Visualization</p> <p>Visualization techniques</p> <p>Understanding analytics output and their usage</p> <p>4.2 Scikit package</p>	

	4.3 MatplotLib Library Plotting Sub plots Images 4.4Data Visualization tools: Tableau, PowerBI 4.5 Extracting or interpreting graphs	
Unit5	Experiential Learning	2
	Apply Data Analytics techniques in a Case study and also use Visualization techniques	

Reference Books:

1. Developing Analytic Talent: Becoming a Data Scientist, Vincent Granville, wiley, 2014
ISBN-10: 1118810082, **ISBN-13:** 978-1118810088
2. Introduction to Data Science, Jeffrey Stanton & Robert De Graaf, Version 2.0, 2013.

Progressive Education Society's
Modern College of Arts, Science and Commerce (Autonomous)
Shivajinagar, Pune - 5

Post Graduate Diploma in Data Science (2020 Course)

Course Code :20CsCdsP202

Course Name :Business Intelligence

Teaching Scheme: TH:4Hours/Week

Credits :04

Examination Scheme: CIA : 50 Marks

End-Sem : 50 Marks

Prerequisites:

- Relational database concepts, database design and entity-relationship (E-R) modeling, data normalization, and Structured Query Language (SQL)
- Data Mining techniques

Course Objectives:

- Understand the role of BI in enterprise performance management and decision support
- Understand the applications of data mining and intelligent systems in managerial work
- Understand data warehousing and online analytical processing (OLAP) concepts, including dimensional modeling, star and snowflake schemas, attribute hierarchies, metrics, and cubes
- Learn data analysis and reporting using an available BI software

Course Outcomes:

On completion of the course, student will be able to–

- Apply Data mining techniques in Managerial work
- Use BI Software

Course Contents:

Unit No.	Title	Number of Lectures
Unit 1	Introduction to Business Intelligence	6
	1.1.Definition and History of BI 1.2 Transaction processing versus analytical processing 1.3 BI implementation 1.4 Major tools and techniques of BI	
Unit 2	Data Warehousing	8
	2.1 Definition and concepts 2.2 Data warehouse architecture, ETL process, data warehouse development, Top down vs. Bottom up, Data Mart vs. EDW 2.3 Implementation issues 2.4 Real-time data warehousing	
Unit 3	Business performance management	10
	3.1 Key performance indicators and operational metrics 3.2 Balanced scorecard, Six Sigma, Dashboards and scorecards	
Unit 4	Data Mining for Business Intelligence	8
	4.1 Data mining process 4.2 Data mining methods 4.3 ANN for Data Mining	
Unit 5	Text, and Web mining for Business Intelligence	6
	5.1 Text mining Applications 5.2 Process and Tools 5.3 Web content, structure and usage mining	
Unit 6	BI implementation, Integration and Emerging Trends	8
	6.1 Implementing BI 6.2 BI Application Life Cycle 6.3 Connecting BI to Enterprise systems 6.4 On demand BI, Issues of legality, privacy and Ethics 6.5 Emerging topics in BI 6.6 Social Networking and BI, RFID and BI	
Unit 7	Experiential Learning	2
	Apply Data mining techniques in Managerial Case study	

Reference Books:

1. Business Intelligence: A Managerial Approach, 2nd Edition, PEARSON 2012
Authors: Efraim Turban, Ramesh Sharda, Dursun Delen, and David King,
ISBN-10: 0-13-610066-X, ISBN-13: 978-0-13-610066-9
2. Oracle Business Intelligence Applications, McGraw Hill Education 2013
Authors : Simon Miller, William Hutchinson ISBN-10: 93-5134-153-4
ISBN-13: 978-93-5134-153-6

Progressive Education Society's
Modern College of Arts, Science and Commerce (Autonomous)
Shivajinagar, Pune - 5

Post Graduate Diploma in Data Science (2020 Course)

Course Code :20CsCdsP203

Course Name :Machine Learning

Teaching Scheme: TH:4Hours/Week

Credits :04

Examination Scheme: CIA : 50 Marks

End-Sem : 50 Marks

Prerequisites:

- Knowledge of DBMS and Data mining
- Basic knowledge of Python programming language
- Basic knowledge of Probability, Linear Algebra and Statistics

Course Objectives:

- Understand human learning aspects
- To understand machine learning concepts and techniques
- To apply machine learning techniques for problem solving
- To use tools and techniques for understanding concept of machine learning

Course Outcomes:

On completion of the course, student will be able to–

- Understand a wide variety of learning algorithms

- Ability to analyze the data using any machine learning algorithm
- Ability to design and implement a learning model
- Understand how to perform evaluation of learning algorithms and model selection
- Ability to tackle real world problems of machine learning

Course Contents:

Unit No.	Title	Number of Lectures
Unit 1	Introduction to Machine Learning	12
	1.1 What is Machine learning? 1.2 When to use ML? 1.3 Why to study ML? 1.4 Applications of machine learning 1.5 Types of learning - Supervised learning, Unsupervised learning, Semi-supervised learning 1.6 Dimensionality reduction- Introduction to dimensionality reduction, Subset selection, Principal Component Analysis(PCA), tSNE algorithm, Factor analysis 1.7 Designing learning system- Steps in developing learning system, Data preprocessing, Missing data, categorical data, feature scaling, training set vs. test set, Data Modeling phases 1.8 Introduction to Auto Machine Learning, Cloud Machine Learning 1.9 Case studies	
Unit 2	Supervised Learning	12
	2.1 Supervised learning 2.2 Learning with Classification - Decision Tree, Support Vector Machine, Naive Bayes, K-Nearest Neighbors, Logistic regression 2.3 Learning with Regression - Simple Linear regression, Multiple linear Regression 2.4 Evaluating model performance - False positives, False negatives Confusion matrix, Accuracy, Precision, Recall, Cross Validation and Comparison 2.5 Regression Matrices R- Square, Adjusted R Square 2.6 Classification and Evaluation Matrix	

	2.7 Case studies	
Unit 3	Unsupervised Learning	8
	3.1 Un-supervised learning 3.2 Clustering algorithms - K-means clustering, Hierarchical clustering 3.3 Rule based learning - Association rule mining, Apriori, FP growth algorithms, Support and Confidence parameters and Comparison 3.4 Case studies	
Unit 4	Learning with Neural Network	7
	4.1 Limitations of Machine Learning 4.2 What is Deep learning? 4.3 Artificial Neural network - Neurons, Activation function, A Multilayer Feed-Forward Neural Network, Defining network topology, Back propagation algorithm 4.4 Introduction of Recurrent Neural Networks(RNN) 4.5 Convolutional Neural Networks 4.6 Reinforcement learning, Q Learning Algorithm 4.7 Introduction to Long Short Term Memory Networks 4.7 Case studies	
Unit 5	Trends in Machine Learning	7
	5.1 Introduction to Natural Language Processing(NLP) 5.2 Reinforcement learning - Elements of Reinforcement learning, Exploration, Exploitation, Rewards, Penalties 5.3 Ensemble methods - Bagging, Boosting 5.4 Cognitive Computing 5.5 Introduction to IBM Watson 5.6 Gradient boosting and XGBoost 5.7 Learning for Time series Forecasting	
Unit 6	Experiential Learning	2
	Case study of Weather Forecasting with the above Learning Algorithms	

Reference Books:

1) T. Mitchell, Machine Learning, McGraw-Hill, 1997, ISBN-13: 978-0070428072
ISBN-10: 0070428077

2) P. Langley. "Elements of Machine Learning" Morgan Kaufmann Publishers, Inc. 24

1996

- 3) EthemAlpaydin "Introduction to machine learning", MIT Press, **ISBN-10:** 026201243X
ISBN-13: 978-0262012430
- 4) William W. Hsieh, "Machine Learning Mehods in the Environmental Sciences", Cambridge,
ISBN-13: 978-0521791922, **ISBN-10:** 0521791928
- 5) Han Kamber, "Data Mining Concepts and Techniques", Morgann Kaufmann Publishers,
ISBN 978-0123814791
- 6) Margaret. H. Dunham, "Data Mining Introductory and Advanced Topics", Pearson Education,
ISBN-13: 978-0130888921, **ISBN-10:** 0130888923

Text Books:

- 1) Peter Harrington "Machine Learning In Action", DreamTech Press,
ISBN-10: 9781617290183, **ISBN-13:** 978-1617290183
- 2) Stephen Marsland, "Machine Learning An Algorithmic Perspective" CRC Press,
ISBN-10: 1466583282, **ISBN-13:** 978-1466583283

Progressive Education Society's
Modern College of Arts, Science and Commerce (Autonomous)
Shivajinagar, Pune - 5

Post Graduate Diploma in Data Science (2020 Course)

Course Code :20CsCdsP204

Course Name :Lab on Machine Learning

Teaching Scheme: 4Hours/Week

Credits :04

Examination Scheme: CIA : 50 Marks

End-Sem : 50 Marks

Assignments on Machine Learning with a given dataset

Assignment No.	Title
1.	Classification problem with Supervised learning and Unsupervised learning
2.	Clustering with Unsupervised learning
3.	Prediction with Supervised learning and Unsupervised learning
4.	API creation for the developed model

5.	Azure and Auto Machine Learning
----	---------------------------------

Progressive Education Society's
Modern College of Arts, Science and Commerce (Autonomous)
Shivajinagar, Pune - 5

Post Graduate Diploma in Data Science (2020 Course)

Course Code :20CsCdsP205

Course Name :Project (on Data Science)

Teaching Scheme:4Hours/Week

Credits :04

Examination Scheme: CIA : 50 Marks

End-Sem : 50 Marks

The Project can be Platform and Language independent but should be specific to 'Data Science'. CIA of the Project will be done by the Project Guide. Assessment will be done weekly in the respective batch. Evaluation will be on the basis of weekly progress of project work, progress report, orals, results, demonstration and documentation. Students should fill the status of the Project work on the Progress report and get the Signature of Project Guide regularly. Progress report should also include how much time you have spent on specific task/module.

Format of Project Progress Report

Name of the Student:	
Roll No:	Seat No:
Title of the Project:	
Name of Organization/Institute:	
Project Guide's Name:	

Sr. No.	From Date	To Date	Details of Project work	Project Guide's sign (withdate)

Head,
Dept. of Computer Science

Sample Index of the Project Report

1. College Certificate
2. Problem Definition
3. Existing System
4. Need for Computerization
5. Scope of the Proposed System
6. Objectives of the Proposed System
7. Requirement gathering and anticipation
8. Platform (H/W, S/W) with version details
9. Analysis Specification (Object Oriented Approach)
10. Design Specification (Object Oriented Approach)
11. Data Dictionary , Flow diagram(Whichever is applicable)
12. Implementation Strategies
13. Input /Output Screens
14. Decision Tools(If any)
15. Testing Strategies
16. Limitations and Drawbacks
17. Conclusion
18. Future Enhancements
19. User Manual
20. References & Bibliography

NOTE:Any 5 UML Diagrams (Use-case, ClassDiagrams, ObjectDiagrams,

SequenceDiagrams, CollaborationDiagram, ActivityDiagram, State Transition Diagram,

ComponentDiagram, Deployment Diagram)