SCHEME OF EXAMINATION

And

SYLLABUS

For

POST GRADUATE DIPLOMA

In

DATA SCIENCE & ANALYTICS

Offered by

Community College of Skill Development

J C Bose University of Science & Technology, YMCA

Sector-6, Mathura Road, Faridabad,

Haryana, India

2024-25

ABOUT THE COMMUNITY COLLEGE OF SKILL DEVELOPMENT

Community College of Skill Development has been running Diploma in Electrical Electrician since 2013 and also got approval from UGC for PG Diploma in Data Science & Analytics in 2019 with a mission to impart quality education along with extensive hands-on training on the equipment/systems in electrical laboratories and industries. At present CCSD offers one year skill programs in Data Science & Analytics. The training is based on the Dual Education System, which lays great emphasis on practical training. The curriculum also provides an excellent "feeder" degree for those students uncertain about choosing a specific career. The presence of highly skilled and qualified trainers helps the students to enhance their professional and skill levels.

ABOUT THE PROGRAM

Post Graduate Program in Data science is an interdisciplinary field that strives to extract knowledge or insights from data in various forms and employs techniques and theories drawn from many fields within the broad areas of mathematics, statistics, operations research, information science, and computer science. The objective of the Post Graduate Program in Data Science is to introduce participants to the tools and techniques used for handling, managing, analyzing and interpreting data. The PG Diploma in Data Science and Analytics is the course that introduces students to decision making driven by big data and analytics. The PG Diploma in Data Science is a 12- months, 2-semester course that takes students through the intricacies of Data Analytics. The program offers a right blend of statistics, technical and business knowledge. The curriculum has been designed with multiple industry leaders to ensure that the students learn exactly what the employer needs. The program builds a solid foundation in Data Science & Analytics by covering industry standard tools and techniques through a practical, industry-oriented curriculum. The program assumes no prior knowledge of coding in Python, R or SQL and begins from fundamentals. By the end of the program the students will get a deep understanding of statistical techniques critical to Data Analysis and they are able to create Analytical models using real life data to drive business impact.

PROGRAM EDUCATIONAL OBJECTIVES (PEOs)

PEO-1: To enhance the competence level for tackling real world problems in industry, academia, and research organizations

PEO-2: Ability to identify new sources of data and methods to improve data collection, analysis, and reporting.

PEO- 3: Ability to plan, implement, and assess high-level statistical models and strategies for application in the business’s most complex issues.

PEO- 4: Ability to develops econometric and statistical models for various problems including projections, classification, clustering, pattern analysis, sampling and simulations

PEO- 5: Ability to perform a vital role in the advancement of innovative strategies to understand the business's consumer trends and management as well as ways to

solve difficult business problems, for instance, the optimization of product fulfillment and entire profit.

PROGRAM OUTCOMES

After completing the program, students will be able to:

1. Analyze data using statistical techniques and providing reports.
2. Develop and implement databases and data collection systems.
3. Acquire data from primary and secondary sources and maintain data systems.
4. Use effective technology like soft Computing and machine learning to improve the prevalent solutions.

PROGRAM SPECIFIC OUTCOMES (PSOs)

To impart State-of-Art knowledge in the field of Data Science & Analytics and hand on application based practical training with regular Academic and Industry interaction.

SCHEME OF EXAMINATION

FIRST SEMESTER

Subject Code	Subject Name	L-T-P	Credits	Marks Weightage		Course Type
				Internal	External	
PGDSA-101	Introduction to Data Science	3-0-0	3	25	75	PCC
PGDSA-102	Descriptive Statistics	3-0-0	3	25	75	PCC
PGDSA-103	Advanced Database Management System	3-0-0	3	25	75	PCC
PGDSA-104	Data Structure and Algorithm	3-0-0	3	25	75	PCC
PGDSA-107	Python Lab	0-0-4	4	30	70	SDP
PGDSA-108	SQL Lab	0-0-4	4	30	70	SDP

Total	**12-0-8**	**20**	**160**	**440**	

SECOND SEMESTER

Subject Code	Subject Name	L-T-P	Credits	Marks Weightage		Course Type
				Internal	External	
PGDSA-201	Introduction to Big Data and Cloud Computing	3-0-0	3	25	75	PCC
PGDSA-202	Inferential Statistics	3-0-0	3	25	75	PCC
PGDSA-203	Machine Learning	3-0-0	3	25	75	PCC
PGDSA-204	Deep Learning	3-0-0	3	25	75	PCC
PGDSA-205	Mathematics	3-0-0	3	25	75	PCC
PGDSA-206L	Machine Learning Lab	0-0-1.5	1.5	30	70	SDP
PGDSA-207L	Python Lab	0-0-1.5	1.5	30	70	SDP
PGDSA-208L	Project	0-0-2	2	30	70	SDP
Total		**15-0-5**	**20**	**215**	**585**	

DETAILED SCHEME AND SYLLABUS

FIRST SEMESTER

Subject Code	Subject Name	L-T-P	Credits	Marks Weightage		Course Type
				Interna l	External	
PGDSA-101	Introduction to Data Science	3-0-0	3	25	75	PCC
PGDSA-102	Descriptive Statistics	3-0-0	3	25	75	PCC
PGDSA-103	Advanced Database Management System	3-0-0	3	25	75	PCC
PGDSA-104	Data Structure and Algorithm	3-0-0	3	25	75	PCC

PGDSA-107	Python Lab	0-0-4	4	30	70	SDP
PGDSA-108	SQL Lab	0-0-4	4	30	70	SDP
Total		**12-0-8**	**20**	**160**	**440**	

PGDSA-101: INTRODUCTION TO DATA SCIENCE

PG Diploma (**Data Science & Analytics*****) I Semester***

No. of Credits:			3	Sessional:	25 Marks
L	T	P	Total	Theory:	75 Marks
3	0	0	3	Total:	100 Marks
				Duration of Exam:	3 Hours

Pre- Requisite: Computer skills

Successive: Basics of Data Science and Machine Learning

Course Objectives: The objective of studying this course is

❖ Key concepts in data science, including tools, approaches, and application scenarios.

❖ Topics in data collection, sampling, quality assessment and repair. ❖ Topics in statistical analysis and machine learning.

Course Outcomes: At the end of the course, the student shall be able to:

CO1 Manipulate large data sets and use them to identify trends and reach meaningful conclusions to inform strategic business decisions.

CO2 Clean, aggregate, and organize data from disparate sources and transfer it to data warehouses.

CO3 Visualize data using the python module.

Course Contents:

Unit 1: Overview of Data Science

Data Science incorporates various Discipline, Data Science Importance, Data Science Process, Data Science Life Cycle, Data Science Applications and Use Cases, Challenges in Data Science, Data Science Team, Data Science tools and Platforms.

Unit 2: Mathematical Computing

Knowledge on Packages like The numpy Library- numpy, ndarray, dtype, Intinsic creation of array, Difference between list and numpy array , Indexing, Slicing, and Iterating, numpy functions for linear algebra operation , statistical operation , string operation.

Unit 3: Pandas

Introduction to pandas Data Structures, Data series, Data frame, Index object, Other Functionalities on Indexes, Function Application and Mapping, Data Preparation, Concatenating, Data Transformation, Discretization and Binning, String Manipulation, Data Aggregation, Group Iteration.

Unit 4: Data Visualization

The matplotlib and seaborn library , Plot , Scatter plot, Bar Graph , Histogram , Pie Chart, Factorplot, Boxplot, VoilinPlot, Stripplot, Swarmplot, barplot, Countplot, Distplot, JointPlot, PairPlot, RugPlot, Kdeplot, PairGrid, Pairplot, FaceGrid, Heatmap.

Unit 5: Scipy

File input/output, special function, Linear algebra operations, Interpolation, Optimization and fit, Statistics and random numbers, Numerical integration, Fast Fourier transformations.

Text Books/ Reference Books:

1. Introduction to Data Science, Rafael A. Irizarry
2. Doing Data Science Straight Talk from the Frontline "by Cathy O'Neil and Racheal Schutt

Note: It is recommended that some part of the syllabus is to be covered in online mode.

PGDSA-102: DESCRIPTIVE STATISTICS

PG Diploma (Data* Science & Analytics*) I Semester

No. of Credits:		3		Sessional:	25 Marks
L	T	P	Total	Theory:	75 Marks
3	0	0	3	Total:	100 Marks
				Duration of Exam:	3 Hours

Course Objectives: The main objective of this course is to provide students with the foundations of probabilistic and statistical analysis mostly used in varied applications in engineering and science like disease modeling, climate prediction and computer networks etc.

Course Outcomes: At the end of the course, the student shall be able to:

CO1 How to calculate and apply measures of location and measures of dispersion grouped and ungrouped data cases.

CO2 How to apply discrete and continuous probability distributions to various business problems.

CO3 Perform Test of Hypothesis as well as calculate confidence interval for a population parameter for single sample and two sample cases. Understand the concept of p-values.

CO4 Learn non-parametric test such as the Chi-Square test for Independence as well as Goodness of Fit.

Course Contents:

Unit 1: Trial, random experiment, sample point and sample space

Definition of an event. Operation of events, mutually exclusive and exhaustive events. Classical (Mathematical) and Empirical definitions of Probability and their properties. Theorems on Addition and Multiplication of probabilities. Independence of events, pairwise and mutual independence for three event Conditional probability, Bayes theorem and its applications.

Unit 2: Concept of Discrete random variable and properties of its probability distribution

Random variable. Definition and properties of probability distribution and cumulative distribution function of discrete random variable. Raw and Central moments (definition only) and their relationship. (upto order four). Concepts of Skewness and Kurtosis and their uses. Expectation of a random variable. Theorems on Expectation & Variance. Joint probability mass function of two discrete random variables. Marginal and conditional distributions.

Theorems on Expectation & Variance, Covariance and Coefficient of Correlation. Independence of two random variables

Unit 3: Some Standard Discrete Distributions

Discrete Uniform, Binomial and Poisson distributions and derivation of their mean and variance. Recurrence relation for probabilities of Binomial and Poisson distributions. Poisson approximation to Binomial distribution. Hypergeometric distribution, Binomial approximation to hypergeometric distribution.

Unit 4: Continuous random variable

Concept of Continuous random variable and properties of its probability distribution Probability density function and cumulative distribution function. Their graphical representation. Expectation of a random variable and its properties. Measures of location, dispersion, skewness and kurtosis. Raw and central moments (simple illustrations).

Unit 5: Some Standard Continuous Distributions

Uniform, Exponential (location scale parameter), memory less property of exponential distribution and Normal distribution. Derivations of mean, median and variance for Uniform and Exponential distributions. Properties of Normal distribution (without proof). Normal approximation to Binomial and Poisson distribution (statement only). Properties of Normal curve. Use of normal tables.

Text Books/ Reference Books:

1. Introduction to probability , Dimitri Bertsekas
2. Probability and Statisics Morris H.DeGoot

Note: It is recommended that some part of the syllabus is to be covered in online mode.

PGDSA-103: ADVANCED DATABASE MANAGEMENT SYSTEM

PG Diploma (Data Science & Analytics) I Semester

No. of Credits:			3	Sessional:	25 Marks
L	T	P	Total	Theory:	75 Marks
3	0	0	3	Total:	100 Marks
				Duration of Exam:	3 Hours

Pre- Requisite: Basics of Database Management System

Successive: Database Management System

Course Objectives: The objective of studying this course is to provide a strong foundation for database application development, and introduce key aspects of emerging database technology.

Course Outcomes: At the end of the course, the student shall be able to:

CO1 Create Stored Database Procedures for writing consistent, well-tuned backend code.

CO2 Develop database applications using XML data model.

CO3 Understand developments in database technologies.

Course Contents:

Unit 1: Introductory concepts of DBMS

Introduction and applications of DBMS, Purpose of database, Data Independence, Database System architecture- levels, Mappings, Database, users and DBA. Three Level Architecture of DBMS, The External Level or Subschema, The Conceptual Level or Conceptual Schema, The Internal Level or Physical Schema, Data Definition Language, Data Manipulation Language; Database Management System Structure, Database Manager, Database Administrator, Data Dictionary, Client/Server Architecture.EntityRelationship: Basic concepts, Design process, constraints, Keys, Design issues, E-R diagrams, weak entity sets, extended E-R features – generalization, specialization, aggregation, reduction to E-R database schema.

Unit 2: Relational Algebra model

Relational Algebraic Operations, Basic Operations, Union, Difference, Cartesian Product, Intersection, projection, selection, join, division.

Unit 3: Functional Dependency

Definition, trivial and non-trivial FD, closure of FD set, closure of attributes, irreducible set of FD, Normalization – 1Nf, 2NF, 3NF, Decomposition using FD- dependency preservation, BCNF, Multi- valued dependency, 4NF, Join dependency and 5NF

Unit 4: Concurrency Control and Transaction processing

Serializability: Serializability by Locks, Locking Systems With Several, Lock Modes, Architecture for a Locking Scheduler Managing Hierarchies of Database Elements, Concurrency Control by Timestamps, Concurrency Control by Validation, Database recovery management. Introduction of transaction processing, advantages and disadvantages of transaction processing system, online transaction processing system, serializability and recoverability, view serializability, resolving deadlock, distributed locking. Transaction management in multi-database systems, long duration transaction, high-performance transaction system.

Unit 5: Parallel and Distributed Databases

Database Architectures for parallel databases, Distributed Databases and Object Oriented Databases. Distributed Database Introduction of DDB, DDBMS architectures, Homogeneous and Heterogeneous databases, Distributed data storage, Overview of object: oriented paradigm, OODBMS architectural approaches, Object identity, procedures and encapsulation, Object oriented data model: relationship, identifiers, Basic OODBMS terminology, Inheritance.

Unit 6: Data warehouse and data mining:

Introduction to Data Warehousing – Concepts, Benefits and Problems, DW Architecture – Operational Data, load manager, meta data, DW Data flows – inflow, upflow, meta flow, DW tools and technologies – Extraction, cleansing and transformation tools, On-line Analytical Processing, Data mining techniques.

Text Books/ Reference Books:

1. Database Management Systems by Raghu Ramakrishnan
2. Database Systems: The complete book by Jeffrey Ullman

Note: It is recommended that some part of the syllabus is to be covered in online mode.

PGDSA-104: DATA STRUCTURES AND ALGORITHM

PG Diploma (Data Science & Analytics) I Semester

			3	Sessional:	25 Marks
			Total	Theory:	75 Marks
3	0	0	3	Total:	100 Marks
				Duration of Exam:	3 Hours

Pre- Requisite: Nil

Course Objectives: The objective of studying this course is to:

- To impart the basic concepts of data structures and algorithms.
- To understand concepts about searching and sorting techniques
- To understand basic concepts about stacks, queues, lists trees and graphs.
- To enable them to write algorithms for solving problems with the help of fundamental data structures

Course Outcomes: At the end of the course, the student shall be able to:

CO1 Understanding the core terms, concepts, and tools of relational database management systems.

CO2 Understanding database design and logic development for database programming.

Course Contents:

Unit 1: Basic Terminologies

Elementary Data Organizations, Data Structure Operations: insertion, deletion, traversal etc.; Analysis of an Algorithm, Asymptotic Notations, Time-Space trade off. **Searching:** Linear Search and Binary Search techniques and their complexity analysis.

Unit 2: Stacks and Queues

Stack and its operations: Algorithms, Applications of Stacks: Expression Conversion and evaluation – corresponding algorithms and complexity analysis. Queue, Types of Queue: Simple Queue, Circular Queue, Priority Queue; Operations on each type of Queues: Algorithms and their analysis.

Unit 3: Linked Lists

Singly linked lists: Representation in memory, Algorithms of several operations: Traversing, Searching, Insertion into, Deletion from linked list; Linked representation of Stack and Queue, Doubly linked list: operations on it and algorithmic analysis; Circular Linked Lists: all operations their algorithms and the complexity analysis.

No. of Credits:

L T P

Unit 4: Trees

Basic Tree Terminologies, Different types of Trees: Binary Tree, Threaded Binary Tree, Binary Search Tree, AVL Tree; Tree operations on each of the trees and their algorithms with complexity analysis. Applications of Binary Trees, B Tree, B+ Tree: definitions, algorithms and analysis.

Unit 5: Sorting and Hashing

Objective and properties of different sorting algorithms: Selection Sort, Bubble Sort, Insertion Sort, Quick Sort, Merge Sort, Heap Sort; Performance and Comparison among all the methods. Hashing and collision resolution.

Unit 6: Graph

Basic terminologies and representations, Graph search and traversal algorithms and complexity analysis.

Text Books/ Reference Books:

1. Data Structure with C by Seymour Lipschutz
2. Data Structure with C by Aaron M.Tanenbaum, Yedidyah Langsam

Note: It is recommended that some part of the syllabus is to be covered in online mode.

PGDSA-107: PYTHON LAB

PG Diploma (Data Science & Analytics) I Semester

			4	Internal:	30 Marks
			Total	External:	70 Marks
0	0	4	4	Total:	100 Marks
				Duration of Exam:	3 Hours

Course Objectives: The objective of studying this course is to implement Python programs with conditionals and loops, using functions for structuring Python programs and Represent compound data using Python lists, tuples, dictionaries.

Course Outcomes: At the end of the course, the student shall be able to:

CO1 Write, test, and debug simple Python programs.

CO2 Implement Python programs with conditionals and loops.

CO3 Develop Python programs step-wise by defining functions.

CO4 Use Python lists, tuples, dictionaries for representing compound data.

Course Contents:

1. Introduction to Python, Running Python Programs, Writing Python Code, Data Types and Variables, Numeric Variables, String Variables, Standard Data Types Printing with Parameters, Getting Input from a User, String Formatting, Multiple Variable Assignment, Type Conversion.
2. Arithmetic, Assignment, Comparison, logical, Membership, Identity operators, operator Precedence, Evaluating Expressions.
3. Logical Expressions, the “if” Statement, Logical Operators, More Complex Expressions, while loop, for loop, Pattern using for loop.
4. Accessing Values in Strings, slicing of the string, Various String Operators, Predefined Function for string, Reverse of the string.
5. Define a list, List indices, traversing a list, List operations, slices and methods, Map, filter and reduce, Deleting elements of list, Nested list.
6. Advantages of Tuple over List, Packing and Unpacking, comparing tuples, creating nested Tuple, using tuples as keys in dictionaries, Deleting Tuples, Slicing of Tuple, Tuple Membership Built-in functions with Tuple.
7. Advantages of Tuple over List, Packing and Unpacking, comparing tuples, creating nested Tuple, Using tuples as keys in dictionaries, Deleting Tuples, Slicing of Tuple, Tuple Membership Built-in functions with Tuple.
8. create a dictionary, python hashing, Python Dictionary Methods, Copying dictionary, Updating Dictionary, Delete Keys from the dictionary, Dictionary items Method, Sorting the Dictionary , Dictionary in-built Functions, len() Method.
9. What is a function, How to define and call a function in Python, Types of Functions, Significance of Indentation (Space) in Python, Function Return Value, Types of Arguments in Functions, Default Arguments and Non-Default Arguments, Keyword Argument and Non-keyword Arguments, Arbitrary Arguments, Rules to define a function in Python, Various Forms of Function Arguments, Nested Functions, Call By Value, Call by Reference, Anonymous Functions/Lambda functions, Passing functions to function, map(), filter(), reduce() functions, Docstring, Iterators, Generators, Closures, Decorators.

No. of Credits:

L T P

PGDSA-108: SQL LAB

PG Diploma (Data Analysis & Analytics) I Semester

			4	Internal:	30 Marks
			Total	External:	70 Marks
0	0	4	4	Total:	100 Marks
				Duration of Exam:	3 Hours

Course Objectives: The objective of this course is to:

❖ To learn the concepts of Relational Database Management System

❖ To have the hands of experience on SQL using Microsoft Server Management Studio ❖ To learn and practice various SQL queries.

Course Outcomes: After the completion of this course, the students will be able to:

CO1 Install, configure, and interact with a relational database management system.

CO2 Describe, define and apply the major components of the relational database model to database design Learn and apply the Structured Query Language (SQL) for database definition and manipulation

CO3 Define, develop and process single entity, 1:1, 1:M, and M:M database tables **Course contents:**

2. Details about SQL Fundamentals
3. Introduction to Microsoft SQL server and Management studio.
4. To Create a Database & SQL DataTypes
5. To use Alias in SQL
6. To use SELECT in SQL
7. To query data using Where clause in SQL
8. To use Insert Into in SQL
9. To Delete & Update data in SQL
10. To Create A Database & SQL DataTypes
11. To use SELECT in SQL
12. To query data using Where clause in SQL
13. To use Insert Into in SQL
14. To Delete & Update data in SQL
15. To use Group By, Wildcards and Regular Expressions in SQL.
16. Use of Null value & Keyword in SQL
17. How to use Alter, Drop & Rename function in SQL 18. How to use Limit keyword, joins, union an index in SQL.

DETAILED SCHEME AND SYLLABUS

SECOND SEMESTER

Subject Code	Subject Name	L-T-P	Credits	Marks Weightage		Course Type
				Internal	External	
PGDSA-201	Introduction to Big Data and Cloud Computing	3-0-0	3	25	75	PCC
PGDSA-202	Inferential Statistics	3-0-0	3	25	75	PCC
PGDSA-203	Machine Learning	3-0-0	3	25	75	PCC
PGDSA-204	Deep Learning	3-0-0	3	25	75	PCC
PGDSA-205	Mathematics	3-0-0	3	25	75	PCC
PGDSA-206L	Machine Learning Lab	0-0-1.5	1.5	30	70	SDP
PGDSA-207L	Python Lab	0-0-1.5	1.5	30	70	SDP

No. of Credits:

L T P

PGDSA-208L	Project	0-0-2	2	30	70	SDP
Total		**15-0-5**	**20**	**215**	**585**	

PGDSA-201: INTRODUCTION TO BIG DATA AND CLOUD COMPUTING

PG Diploma (Data Science & Analytics) II Semester

No. of Credits:			3	Sessional:	25 Marks
L	T	P	Total	Theory:	75 Marks
3	0	0	3	Total:	100 Marks
				Duration of Exam:	3 Hours

Course Objectives:

- ❖ Describe the Data Science Process and how its components interact.
- ❖ Use APIs and other tools to scrap the Web and collect data.
- ❖ Apply EDA and the Data Science process in a case study.
- ❖ Describe what Data Science is and the skill sets needed to be a data scientist

Course Outcomes: At the end of the course, the student shall be able to:

CO1 Identify and explain fundamental mathematical and algorithmic ingredients that constitute a Recommendation Engine (dimensionality reduction, singular value decomposition, principal component analysis). Build their own recommendation system using existing components.

CO2 Create effective visualization of given data (to communicate or persuade).

CO3 Work effectively (and synergically) in teams on data science projects.

Course Contents:

Unit 1: Big Data

Introduction to Big Data, Fields of Big Data, Benefits of Big Data, Big Data Challenges, Local vs distributed System, Hadoop, Map Reduce, Spark.

Unit 2: Pyspark

Introduction to pyspark, uses of pyspark, Features of pyspark, Advantage of pyspak, Architecture, Pyspark modules and packages, Download Spark and Dependencies: Java SetUp, Python SetUp, Spark SetUp.

Unit 3: Pyspark Resilient Distributed Dataset (RDD)

Introduction to RDD, Advantage of RDD, RDD limitations, RDD Creation, RDD Parallelize, RDD map(lambda and simple function), flatmap, filters, distinct, groupByKey, reduceByKey, count, countByValue, saveastextfile, partition.

Unit 4: Pyspark DataFrame

Introduction to DataFrame, DataFrame creation from RDD, Select DF columns, Column rename and Alias, show, collect, withColumn, where, drop, dropFDuplicates, union, unionAll, Filter rows, count, Distinct, Duplicate, sort, orderBy, groupBy, joins, sample, sampleBy, fill, fillna, pivot, partitionBy.

Unit 5: Pyspark SQL

Introduction to Pyspark SQL, Pyspark streaming, String functions, Date & time functions, Math functions, Aggregate functions, Window functions, Sort functions.

Unit 6: Cloud Computing

Introduction to Cloud Computing, Advantage, Disadvantage, Architecture, Technologies, Applications, Types of cloud, Cloud Service Models, Virtualization, Cloud Service Providers, Local Virtual Box SetUp, AWS EC2 pyspark setup, AWS EMR Cluster SetUp.

Text Books/ Reference Books:

1. Hands-On Big Data Analytics with PySpark By James Cross , Rudy Lai , Bartłomiej Potaczek
2. Analytics with Spark Using Python (Addison-Wesley Data & Analytics Series) 1st Edition by Jeffrey Aven (Author)

Note: It is recommended that some part of the syllabus is to be covered in online mode.

PGDSA-202: INFERENTIAL STATISTICS

PG Diploma (Data Science & Analytics) II Semester

No. of Credits:			3	Sessional:	25 Marks
L	T	P	Total	Theory:	75 Marks
3	0	0	3	Total:	100 Marks
				Duration of Exam:	3 Hours

Course Objectives: The objective of studying this course is:

- ❖ To calculate and apply measures of location and measures of dispersion grouped and ungrouped data cases.
- ❖ To apply discrete and continuous probability distributions to various business problems.
- ❖ Perform Test of Hypothesis as well as calculate confidence interval for a population parameter for single sample and two sample cases. Understand the concept of p-values.
- ❖ Learn non-parametric test such as the Chi-Square test for Independence as well as Goodness of Fit.

Course Outcomes: At the end of the course, the student shall be able to:

CO1 Student can calculate and apply all measures of location and measures of dispersion for grouped and ungrouped data.

CO2 Student can apply discrete and continuous probability distributions to all of business problems.

CO3 Student can perform all test of Hypothesis.

CO4 Student can compute and interpret all of the results of Bivariate Regression.

Course Contents:

Unit 1: Elementary topics on Estimation and Testing of hypothesis

Sample from a distribution: Concept of a statistic, estimate and its sampling distribution.

Parameter and it’s estimator. Concept of bias and standard error of an estimator. Central Limit theorem (statement only). Sampling distribution of sample mean and sample proportion. (For large sample only) Standard errors of sample mean and sample proportion. Point and Interval estimate of single mean, single proportion from sample of large size. Statistical tests: Concept of hypothesis Null and alternate hypothesis, Types of errors, Critical region, Level of significance. Large sample tests (using central limit theorem, if necessary) For testing specified value of population mean for testing specified value in difference of two means For testing specified value of population proportion For testing specified value of difference of population proportion (Development of critical region is not expected.) Use of central limit theorem.

Unit 2: Idea of Inference

Point & Interval Estimations and Testing of Hypothesis (2L) Point estimation: Requirements of a good estimator – notions of Mean Square Error, Unbiasedness: Minimum Variance Unbiasedness and Best Linear Unbiasedness, Sufficiency, Factorization Theorem (Discrete case only), Properties of minimum variance unbiased estimators, consistent estimators and asymptotic efficiency, Cramer-Rao lower bound, Rao-Blackwell Theorem. (17L) 9 Methods of Estimation – Moment, Least-square, Maximum Likelihood & Minimum $\chi 2$ methods and their properties (excluding proofs of large sample properties).

Unit 3: Elements of Hypothesis Testing

Null and Alternative hypotheses, Simple and Composite hypotheses, Critical Region, Type I and Type II Errors, Level of Significance and Size, p-value, Power (4L) Tests of Significance related to a single Binomial proportion and Poisson parameter; two Binomial proportions and Poisson parameters; the mean(s) and variance(s) of a single univariate normal distribution, two independent normal distributions and a single bivariate normal distribution; regression and correlation coefficients of a single bivariate normal distribution, Combination of Probabilities in tests of significance

Unit 4: Introduction to heterogeneity and Analysis of Variance and Covariance

Heterogeneity and Analysis of Variance and Covariance, Linear Hypothesis, Orthogonal splitting of total variation, Selection of Valid Error. (3L) Applications of the ANOVA technique to: one-way classified data, two-way classified data with equal number of observations per cell, testing simple regression coefficients, tests for parallelism and identity, correlation ratio, linearity of simple regression, multiple correlation and partial correlation coefficients.

Unit 5: Bayesian Paradigm

Introduction, Bayesian and Minimax decision rules, selection of a prior, Bayesian point estimation, Bayesian sufficiency, and Classical approximation methods.

Text Books/ Reference Books:

1.Introduction to Probability by Charles M.Grinstead, J. Laurie Snell
2. Probability and Statistics by Dacunha- Castelle Didier
3.. Mukhopadhyay P. (1999): Applied Statistics
4. Johnston J. & Dinardo J. (1997): Econometric Methods, McGraw Hill
5. Nagar A.L. & Das R.K. (1976): Basic Statistics

PGDSA-203: MACHINE LEARNING

PG Diploma (Data Science & Analytics) II Semester

No. of Credits:			3	Sessional:	25 Marks
L	T	P	Total	Theory:	75 Marks
3	0	0	3	Total:	100 Marks
				Duration of Exam:	3 Hours

Course Objectives: The objective of studying this course is to:

❖ To introduce students to the basic concepts and techniques of Machine Learning.

❖ To become familiar with regression methods, classification methods, clustering methods.

❖ To become familiar with Dimensionality reduction Techniques.

Course Outcomes: At the end of the course, the student shall be able to: CO1 Gain knowledge about basic concepts of Machine Learning

CO2 Identify machine learning techniques suitable for a given problem

CO3 Solve the problems using various machine learning techniques

CO4 Apply Dimensionality reduction techniques.

Course Contents:

Unit 1: Machine Learning (ML) Fundamentals

ML Modelling Flow, Parametric and Non-Parametric ML, Types of ML, Performance Measures, Bias-Variance Trade-Off, Overfitting and Underfitting, Optimization.
Classification, Regression, Linear Regression with OLS, Linear Regression with SGD, Evaluating Model Parameters, L1 and L2 Regularization, Logistic Regression MLE, Logistic Regression with SGD, Evaluating Model Performance, Measuring Performance Metrics: Precision, Recall, AUC, ROC

Unit 2: Decision Trees

Intro to Decision Tree,| Entropy and Information Gain, Standard Deviation Reduction, Gini, Index, CART and CHAID, Performance Metrics, Bootstrap Sampling, Bagging (Bootstrap Aggregation), Random Forest, Performance Metrics.

Unit 3: Support Vector Machines (SVM)

Understanding Vectors, Decision Boundary, Support Vectors, Understanding Hyperplane, Support Vector Machine, Working of SVM, Kernels and Types of Kernels, Strengths and Challenges of SVM

Unit 4: Model Selection Technique

Cross Validation, Types of Cross Validation, Hold-out, K-fold, Grid and random search for Parameter tuning. Ensemble Techniques: Boosting, AdaBoost, Gradient Boosting, XGBoost.

Unit 5: Principal Component Analysis

Introduction to Dimensionality Reduction, Computing Components in PCA, Dimensionality Reduction using PCA.

Unit 6: K-Means Clustering

Intro to Clustering, K-Means Clustering Algorithm, Choosing the Optimum K value (Elbow Method), Various Distance Measures**.**

Text Books/ Reference Books:

1. Machine Learning for Hackers Drew Coway,Jobn Myles White.
2. Python Machine Learning by example Liu Yuxi(Hayden)
   PGDSA-204: DEEP LEARNING

PG Diploma (Data Science & Analytics) II Semester

No. of Credits:			3	Sessional:	25 Marks
L	T	P	Total	Theory:	75 Marks
3	0	0	3	Total:	100 Marks
				Duration of Exam:	3 Hours

Course Objectives:

- ❖ To introduce neural networks concepts and associated techniques
- ❖ To design appropriate neural network-based technique for a given scenario.
- ❖ To reduce the dimension of an image and classification of images.
- ❖ To introduce the recurrent neural networks to overcome sequence learning problems.

Course Outcomes: At the end of the course, the student shall be able to:

CO1 To provide an understanding of different types of Deep Architectures, including Convolutional Networks and Recurrent Networks.

CO2 To introduce neural networks concepts and associated techniques **Course Contents:**

Unit 1: Introduction to Tensorflow

Computational Graph, Key highlights, Creating a graph, Regression example, Gradient Descent, TensorBoard, Modularity, Sharing variables, Keras, Perceptron: XOR Gate

Unit 2: Introduction to Neural Networks

Activation function: linear and non-linear, Sigmoid, RELU, Hyperbolic Fns, Softmax Artificial Neural Networks: Introduction to Perceptron Training Rule, Gradient Descent Rule.

Unit 3: Introduction to Gradient Descent

Gradient Descent and Backpropogation: Gradient Descent, Stochastic Gradient Descent, Backpropagation, Some problems in ANN Optimization and Regulization: Overfitting and Capacity, crosss Validation, Feature Selection, Regularization, Hyperparameters.

Unit 4: Deep Neural Networks

Introduction to Convolution Neural Networks: Introduction to CNNs, Kernel filter, Principles behind CNNs, Multiple Filters, CNN applications

Introduction to Recurrent Neural Networks: Unfolded RNNs, Seq2Seq RNNs, LSTM, RNN applications.

Unit 5: Deep learning applications

Image Processing, Natural Language Processing, Speech Recognition, Video Analytics.

Text Books/ Reference Books:

1. Deep Learning by Ian Goodfelow, Yoshua Bengio and Aaron Courville.
2. Hands on Deep Learning for images with Tensorflow, Ballard Will.
3. Goodfellow, I.,Bengio,Y.,and Courville, A.,Deep Learning, MIT Press,2016.
4. Bishop, C., M., Pattern Recognition and Machine Learning, Springer, 2006.
5. Satish Kumar, Neural Networks:A Classroom Approach, Tata McGraw-Hill Education
6. Golub, G.,H., and Van Loan, C., F., Matrix Computations. JHU Press, 2013
7. Yegnanarayana, B.,Artificial Neural Networks PHI Learning Pvt.Ltd, 2009

PGDSA-205: MATHEMATICS

PG Diploma (Data Science & Analytics) II Semester

No. of Credits:			3	Sessional:	25 Marks
L	T	P	Total	Theory:	75 Marks
3	0	0	3	Total:	100 Marks
				Duration of Exam:	3 Hours

Pre- Requisite: Nil

Course Objectives: The objective of studying this course is to analyze the solution set of a system of linear equations, express some algebraic concepts (such as binary operation, group, field), express a system of linear equations in a matrix form, describe the concepts of eigen value, eigenvector and characteristic polynomial and definite a vector space and subspace of a vector.

Course Outcomes: At the end of the course, the student shall be able to:

CO1 Practice residential and light commercial wiring in accordance with the codes and authorities for installation.

CO2 Identify various wiring methodologies and their testing procedures.

CO3 List the tools used in wiring.

C04 Recognize the different electrical accessories used in residential and light commercial wiring.

Course Contents:

Unit 1: Scalars and Vectors

Scalars, Vectors, Matrix, Tensors, Vector space, linearly independent and linearly dependent set of vectors, Basis and dimension of a vector space.

Unit 2: Norm of Matrix

Linear Transformations and its matrix representation, Elementary transformations, Rank of a transformation, Rank- nullity theorem, Matrix decomposition, Quadratic forms, Geometry of positive definite quadratic form,

Unit 3: Determinants

Determinant, Partitioning of matrices, Eigenvalue, Eigenvector, Cayley-Hamilton theorem, Similarity of matrices, Diagonalization of matrices.

Unit 4: Solution of Matrices

Inner product spaces, Isometry, Orthonormal bases, Gram-Schmidt process. Solution of the system of linear equations. Matrix differential operators, Jacobian of matrix transformation, and function of matrix arguments, Principal Component Analysis.

Text Books/ Reference Books:

1. Linear Algebra by Kenneth Hoffman, Ray Kunze

2. Linear Algebra by Gilbert Strang

Note: It is recommended that some part of the syllabus is to be covered in online mode.

PGDSA-206L: MACHINE LEARNING LAB

PG Diploma (Data Science & Analytics) II Semester

No. of Credits:			1.5	Internal:	30 Marks
L	T	P	Total	External:	70 Marks
0	0	1.5	1.5	Total:	100 Marks
				Duration of Exam:	3 Hours

Course Objectives: The objective of studying this course is to:

- To familiarize students with structure, configuration and working of a power derived vehicle.
- Understand the concept of Electric vehicle and Hybrid electric vehicle.
- To get familiarize with the various components and their application inside an electric vehicle.

Course Outcomes: At the end of the course, the student shall be able to:

CO1 Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a .CSV file.

CO2 Build an Artificial Neural Network by implementing the Back propagation algorithm and test the same using appropriate data sets.

Course Contents:

1. For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples
2. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.
3. Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your data set.
4. Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set. You can use Java/Python ML library classes/API.
5. Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment on the quality of clustering. You can add Java/Python ML library classes/API in the program.
6. Write a program to implement k-Nearest Neighbor algorithm to classify the iris data set.

Print both correct and wrong predictions. Java/Python ML library classes can be used for this problem.

7. Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.

Text Books/ Reference Books:

1. Python Machine learning by example, LiuYuxi(Hayden)

Note: It is recommended that some part of the syllabus is to be covered in online mode.

PGDSA-207L: PYTHON LAB

PG Diploma (Data Science & Analytics) II Semester

No. of Credits:		1.5		Internal:	30 Marks
L	T	P	Total	External:	70 Marks
0	0	1.5	1.5	Total:	100 Marks
				Duration of Exam:	3 Hours

Course Objectives: The objective of studying this course is to provide Basic knowledge of Python. Python programming is intended for software engineers, system analysts, program managers and user support personnel who wish to learn the Python programming language

Course Contents:

1. **Python Module:** Introduction to module, Types of Module, import Statement, from…import Statement, Import * Statement, Underscores in Python, The dir( ) Function , Creating User defined Modules, Command line Arguments , Python Module Search Path
2. Introduction to Package, .py file, Importing module from a package, Creating a Package, Creating Sub Package, Importing from Sub-Packages, Popular Python Packages
3. **Python predefined module:** Date & DateTime Class, Format Time Output, Timedelta Objects, Calendar Module,os module.
4. **File Handling:** Introduction to file handling, File Objects , File Different Modes and Object Attributes, a Text File and Append Data to a File and Read a File, Closing a file, Read, read line , read lines,write,write lines, Renaming and Deleting Files.
5. **Exception Handling:** importance Of Exception, Introduction to Exception Handling, Try … Except, Try .. Except .. else Try … finally, Argument of an Exception, Python Custom Exceptions, Ignore Errors, Assertions, Using Assertions Effectively.
6. **Object Oriented Programming System:** Define oops concept, Difference between class variable and instance variable, Difference between function and method, Define class, object and instance of a class. features of oops, encapsulation , Polymorphism, Inheritance, Differentiate among instance method , class method , static method.
7. **Regular Expression:** Introduction to regular expression, Regular Expression Syntax, Understanding Regular Expressions, Regular Expression Patterns, Literal characters, Repetition Cases, Example of w+ and ^ Expression, Example of \s expression in re.split function, Using regular expression methods, Using re.match(),Finding Pattern in Text (re.search()),re.findall for text, Flags, Methods of RE.
8. **Database Connectivity:** Creating Database connection, understanding cursor, Executing Queries, Parameterized Queries.
9. Introduction to Multithreading, Threading Module, Define a thread, Thread Synchronization.
10. **GUI Programming-tkinter:** Introduction, Components and events, Adding control, Label, Button, Entry, Text , Radio, Check widget, ListBox, Menu, Combobox.

Text Books/ Reference Books:

1. Head- First Python, 2nd edition by Paul Barry (o'Reilly,2016)
2. Python programming by zohn zelle

Note: It is recommended that some part of the syllabus is to be covered in online mode.