SCHEME OF EXAMINATION

And

SYLLABUS

For

POST GRADUATE DIPLOMA

In

DATA SCIENCE & ANALYTICS

Offered by

J C Bose University of Science & Technology YMCA

Sector-6, Mathura Road, Faridabad, Haryana, India

J C Bose University of Science & Technology YMCA

VISION

J C Bose University of Science & Technology YMCA aspires to be a nationally and internationally acclaimed leader in technical and higher education in all spheres which transforms the life of students through integration of teaching, research and character building.

MISSION

- To **contribute** to the development of science and technology by synthesizing teaching, research and creative activities.
- To **provide** an enviable research environment and state-of-the art technological exposure to its scholars.
- To **develop** human potential to its fullest extent and make them emerge as world class leaders in their professions and enthuse them towards their social responsibilities.

Community College of Skill Development

VISION

"**Community College of Skill Development** aspires to be a national leader in skill development in all spheres which **transforms** the life of students through the integration of skill, training and character building."

MISSION

- To **provide** an academically conducive environment, to **generate** future leaders and professionals of 21st Century.
- To **nurture** talent of underprivileged section of the society and **transforms** the lives of millions through academic excellence.
- To **incorporate** specific need-based education and training opportunities for skill up gradation.

About Program

- PG DIPLOMA IN DATA SCIENCE & ANALYTICS is a one year post graduate program which is designed with an aim to provide the students in depth knowledge of various advanced concepts of data science.
- This comprehensive curriculum covers the concepts of Data Analytics and Statistics foundation, analyzing data using Python and R programming languages, interacting with databases using SQL, visualizing the data using Tableau and Power BI and Machine Learning and deep learning using python.
- This Post Graduate Program in Data Analytics has been developed by industry experts to help you learn the applications of data science from scratch and build powerful models to generate useful business insights and predictions. It has been designed for fresher and budding professionals looking to build their career in Data Science & Analytics. This program comes with a job placement guarantee and provides complete placement support to prepare learners for future job opportunities.

PROGRAM EDUCATION OBJECTIVE

PEO1	To enhance the competence level for tackling real world problems in industry, academia and research organizations
PEO2	To sharpen problem solving ability using in depth analysis based upon stateof-the-art concepts and technology
PEO3	To create awareness about professional ethics, multidisciplinary approach, entrepreneurial thinking and effective communication.

PROGRAM OUTCOMES

PO1	Ability to describe the current state of reality for organizations by translating data into information accessible to the business.
PO2	Ability to identifying new sources of data and methods to improve data collection, analysis, and reporting.
PO3	Ability to plans, implements, and assesses high-level statistical models and strategies for application in the business's most complex issues.
PO4	Ability to develops econometric and statistical models for various problems including projections, classification, clustering, pattern analysis, sampling and simulations
PO5	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 SPECIFIC OUTCOMES

PSO1	Analyzing data using statistical techniques and providing reports
PSO2	Developing and implementing databases and data collection systems
PSO3	Acquiring data from primary and secondary sources and maintain data systems
PSO4	Identifying, analyzing, and interpreting trends or patterns in complex data sets . Filtering and cleaning data.
PSO5	Effectively technology like soft Computing and machine learning to improve the prevalent solutions

J.C BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD
COMMUNITY COLLEGE OF SKILL DEVELOPMENT

SCHEME OF POST GRADUATE DIPLOMA IN DATA SCIENCE & ANALYTICS

FIRST SEMESTER

SUBJECT CODE	SUBJECT NAME	CREDITS
PG-DSA-101	INTRODUCTION TO DATA SCIENCE & MACHINE LEARNING	4
PG-DSA-102	STATISTICS & PROBABILITY	4
PG-DSA-103	ADVANCED DATABASE MANAGEMENT SYSTEM	4
PG-DSA-104	DATA STRUCTURES & ALGORITHMS	4
PG-DSA-105	PYTHON LAB	8
PG-DSA-107	SQL LAB	3
TOTAL		30

SECOND SEMESTER

SUBJECT CODE	SUBJECT NAME	CREDITS
PG-DSA-201	FUNDAMENTALS OF DATA SCIENCE	4
PG-DSA-202	ELEMENTS OF STATISTICAL LEARNING	4
PG-DSA-203	MACHINE LEARNING	4
PG-DSA-205	MATHEMATICS	4
PG-DSA-206	MACHINE LEARNING LAB	3
PG-DSA-207	PYTHON LAB	3
PG-DSA-208	MINOR PROJECT	4
	ELECTIVE SUBJECT: Any one of the following	
PG-DSA-204 PG-DSA-209 PG-DSA-210	DEEP LEARNING IMAGE ANALYSIS AND COMPUTER VISION DIGITAL SPEECH PROCESSING AND RECOGNITION	4
TOTAL		30

Detailed Curriculum

J.C BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD

SUBJECT NAME: INTRODUCTION TO DATA SCIENCE & MACHINE LEARNING

PAPER CODE: PG-DSA-101

Semester -1st	Sessional – 25 Marks
Stream - Data Science & Analytics	Theory – 75 Marks
L T P Total Credits	Total - 100 Marks
4 0 0 4	Duration of Exam: 3 Hours

Course Objectives:

1. Key concepts in data science, including tools, approaches, and application scenarios
2. Topics in data collection, sampling, quality assessment and repair
3. Topics in statistical analysis and machine learning

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,

Course Outcomes:

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

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

3. Visualize data using python module.

J.C BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD

SUBJECT NAME: STATISTICS & PROBABILITY

PAPER CODE: PG-DSA-102

Semester -1st	Sessional – 25 Marks
Stream–Data Science & Analytics	Theory– 75 Marks
L T P Total Credits	Total: 100
Marks	
4 0 0 4	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 Content:

Unit-1

Introduction to data, descriptive statistics, basic concepts of statistics, sampling, Define probability, conditional probability, Baye's theorem, random variables.

Unit-2

Randomization, case studies of randomization, Hypothesis testing, simulation case studies, central limit theorem, Normal distribution

Unit-3

Inference for a single proportion, difference of two proportions, Testing for goodness of fit using chi-square, testing for independence in two-way tables, T- distribution, difference of two means, compairing many means with ANOVA.

Unit-4

Line fitting, residuals and correlation, fitting a line by least squares regression, types of outliers in linear regression, multiple regression, logistic regression.

Course Outcomes:

1. How to calculate and apply measures of location and measures of dispersion -- grouped and ungrouped data cases.
2. How to apply discrete and continuous probability distributions to various business problems.
3. 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.
4. Learn non-parametric test such as the Chi-Square test for Independence as well as Goodness of Fit.

J.C BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD

SUBJECT NAME: ADVANCED DATABASE MANAGEMENT SYSTEM

PAPER CODE: PG-DSA-103

Semester -1st	Sessional – 25 Marks
Stream–Data Science & Analytics	Theory– 75 Marks
L T P Total Credits	Total: 100
Marks	
4 0 0 4	Duration of Exam: 3 Hours

Course Objectives:

To provide strong foundation for database application development, introduce key aspects of emerging database technology.

Course Contents:

Unit-1

Introductory concepts of DBMS:

Introduction and applications of DBMS, Purpose of data base, Data Independence, Database System architecture- levels, Mappings, Database, users and DBA.

Unit-2

Database System Architecture:

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

Entity-Relationship: 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-3

Relational Algebra model:

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

Unit-4

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-5

Concurrency Control

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

Unit-6

Transaction processing:

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 system, long duration transaction, high-performance transaction system.

Unit-7

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-8

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.

Course Outcome:

1. Create Stored Database Procedures for writing consistent, well-tuned backend code.
2. Develop database application using XML data model.
3. Understand developments in database technologies.

J.C BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD

SUBJECT NAME: DATA STRUCTURES & ALGORITHMS

PAPER CODE: PG-DSA-104

Semester -1st	Sessional – 25 Marks
Stream–Data Science & Analytics	Theory– 75 Marks
L T P Total Credits	Total: 100
Marks	
4 0 0 4	Duration of Exam: 3 Hours

Course Objectives:

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

Course Content:

Unit-1

INTRODUCTION

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.

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.

Course Outcomes:

1. Understanding the core terms, concepts, and tools of relational database management systems.
2. Understanding database design and logic development for database programming.

J.C BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD

SUBJECT NAME: PYTHON LAB

PAPER CODE: PCC-WD-105

Semester 1st	Internal – 30 Marks
Stream– Data Science & Analytics	External – 20 Marks
L T P Total Credits	Total - 50 Marks
0 0 8 8	Duration of Exam: 3 Hours

Course Objectives:

1. To write, test, and debug simple Python programs.
2. To implement Python programs with conditionals and loops.
3. Use functions for structuring Python programs.
4. Represent compound data using Python lists, tuples, dictionaries.
5. Read and write data from/to files in Python.

Course Contents:

Unit-1

Fundamental of python:

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.

Unit-2

Python Operators and Operands:

Arithmetic, Assignment, Comparision , logical ,Membership ,Identity operators , operator Precedence, Evaluating Expressions.

Unit-3

Making Decisions and loop statement:

Logical Expressions , The "if" Statement , Logical Operators , More Complex Expressions, while loop, for loop , Pattern using for loop.

Unit-4

Python Strings:

Accessing Values in Strings, Slicing of the string, Various String Operators, Predefined Function for string, Reverse of the string.

Unit-5

Python List:

Define a list, List indices, Traversing a list, List operations, slices and methods, Map, filter and reduce, Deleting elements of list , Nested list.

Unit-6

Python Tuple:

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

J.C BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD

Unit-7

Python Set:

create a set, Iteration Over Sets, Python Set Methods, Python Set Operations, Union of sets, Built-in Functions with Set, Python Frozenset.

Unit-8:

Python Dictionary:

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

Unit-9

Python Functions:

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.

Course Outcomes:

1. Write, test, and debug simple Python programs.
2. Implement Python programs with conditionals and loops.
3. Develop Python programs step-wise by defining functions and calling them.
4. Use Python lists, tuples, dictionaries for representing compound data.
5. Read and write data from/to files in Python.

J.C BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD

SUBJECT NAME: SQL LAB

PAPER CODE: PCC-WD-107

Semester 1st	Internal – 30 Marks
Stream– Data Science & Analytics	External – 20 Marks
L T P Total Credits	Total - 50 Marks
0 0 3 3	Duration of Exam: 3 Hours

Course objective:

1. To learn the concepts of Relational Database Management System
2. To have the hands of experience on SQL using Microsoft Server Management Studio
3. To learn and practice various SQL queries.

Course contents:

1. Details about SQL Fundamentals
2. What is a database? What is SQL?
3. Introduction to Microsoft SQL server and Management studio.
4. How To Create A Database & SQL DataTypes
5. How to use Alias in SQL
6. How to use SELECT in SQL
7. How to query data using Where clause in SQL
8. How to use Insert Into in SQL
9. How to Delete & Update data in SQL
10. How To Create A Database & SQL DataTypes
11. How to use SELECT in SQL
12. How to query data using Where clause in SQL
13. How to use Insert Into in SQL
14. How to Delete & Update data in SQL
15. How sorting is done in SQL using ORDER BY, DESC and ASC
16. How to use Group By in SQL
17. How to use Wildcards in SQL
18. Using Regular Expressions & Wild Cards in SQL
19. Use of Null value & Keyword in SQL
20. How to use Alter, Drop & Rename function in SQL
21. How to use Limit keyword in SQL
22. How to use Joins in SQL
23. How to use Unions in SQL
24. How to use Index in SQL

Course Outcomes:

1. Install, configure, and interact with a relational database management system.
2. 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
3. Define, develop and process single entity, 1:1, 1:M, and M:M database tables

J.C BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD

SUBJECT NAME: FUNDAMENTALS OF DATA SCIENCE

PAPER CODE: PG-DSA-201

Semester -2nd	Sessional – 25 Marks
Stream–Data Science & Analytics	Theory– 75 Marks
L T P Total Credits	Total: 100
Marks	
4 0 0 4	Duration of Exam: 3 Hours

Course Objectives:

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

Course Contents:

UNIT-1

Text Analysis: Understanding nlp , install nltk , tokenize words , tokenizing sentences , stop word customization , stemming and lemmatization , Feature Extraction , Sentiment Analysis , Count Vectorizer, Tfidf Vectorizer.

UNIT-2

Image Processing: Install opencv module, Reading images , Understanding gray scaleimage , Resizing image , Understanding haar classifiers , Face and eyes Detection , Building image dataset , capturing video, Face classification in video.

UNIT-3

TIME SERIES:

Understanding Time Series Data , Visualizing and Understanding Time Series, Components, Autocovariance , ACF and PACF , Autoregressive models: AR,MA, ARMA, ARIMA , Exponential Smoothing , Holt-Winter's Model.

Unit-4

Recommendation System:

Content based technique, Collaborative filtering technique, Evaluating similarity based on correlation

Classification-based recommendations.

Unit-5

Data Visualization with Tableau :

Intro to Tableau Interface, Connnecting to Data, Visual Analytics, Mapping, Calculations, Dashboard and Stories.

Unit-6

Data Visualization with PowerBI:

Introduction to PowerBI, Visualisation with BI, Data Analysis Expressions

Course Outcomes:

1. 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.
2. Create effective visualization of given data (to communicate or persuade).
3. Work effectively (and synergically) in teams on data science projects.

J.C BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD

SUBJECT NAME: ELEMENTS OF STATISTICAL LEARNING

PAPER CODE: PG-DSA-202

Semester -2nd	Sessional – 25 Marks
Stream–Data Science & Analytics	Theory– 75 Marks
L T P Total Credits	Total: 100 Marks
4 0 0 4	Duration of Exam: 3 Hours

Course Objectives:

1. How to calculate and apply measures of location and measures of dispersion -- grouped and ungrouped data cases.
2. How to apply discrete and continuous probability distributions to various business problems.
3. 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.
4. Learn non-parametric test such as the Chi-Square test for Independence as well as Goodness of Fit.

Course Contents:

1. Introduction to Random Variable, Random Sample, Statistic and Parameter, Sampling Distributions , Expectation ,Variance ,Covariance, Correlation.
2. Common Probability Distribution , Marginal Probability , Conditional Probability ,
3. Chain Rule of Conditional Probability, Independence and Conditional Independence.
4. Theory of Estimation: Sufficient statistics, completeness, unbiased estimation, moment estimation, maximum likelihood estimation, notion of admissibility of estimators.
5. Testing of Statistical Hypothesis: Generalized NP lemma, Unbiased critical regions, unbiased tests and similar regions, invariant test, testing and confidence regions.
6. Model Building: Linear Regression(Least square methods and Gradient
7. Descent method) . Logistic Regression(logistic function , logits , odds ,odds ratio)
8. Bayesian Paradigm: Introduction, Bayesian and Minimax decision rules, selection of a prior, Bayesian point estimation, Bayesian sufficiency, and Classical approximation methods.

Course Outcomes:

1. Student can calculate and apply all measures of location and measures of dispersion for grouped and ungrouped data
2. Student can apply discrete and continuous probability distributions to all of business problems.
3. Student can perform all test of Hypothesis
4. Student can compute and interpret all of the results of Bivariate Regression

J.C BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD

SUBJECT NAME: MACHINE LEARNING

PAPER CODE: PG-DSA-203

Semester -2nd	Sessional – 25 Marks
Stream–Data Science & Analytics	Theory– 75 Marks
L T P Total Credits	Total: 100
Marks	
4 0 0 4	Duration of Exam: 3 Hours

Course Objectives:

1. To introduce students to the basic concepts and techniques of Machine Learning.
2. To become familiar with regression methods, classification methods, clustering methods.
3. To become familiar with Dimensionality reduction Techniques.

Course Contents:

Unit-1

ML Fundamentals:

ML Modelling Flow, Parametric and Non-Parametric ML, Algorithm, Types of ML, Performance Measures, Bias-Variance Trade-Off, Overfitting and Underfitting, Optimization

Unit-2

Linear Regression:

Linear Regression with OLS, Linear Regression with SGD, Evaluating Model Parameters , L1 and L2 Regularization, Mesuring Performance Metrics

Unit-3:

Logistic Regression:

Logistic Regression MLE , Logistic Regression with SGD , Evaluating Model Performance, Measuring Performance Metrics: Precision, Recall, AUC ROC, etc

Unit-4

Decision Trees:

Intro to Decision Tree ,| Entropy and Information Gain , Standard Deviation Reduction, Gini Index , CART and CHAID , Performance Metrics

Unit-5

Random Forests:

Bootstrap Sampling , Bagging (Bootstrap Aggregation) , Intro to Random Forest , Why Random Forest ,Performance Metrics

Unit-6

Model Selection Technique:

Cross Validation , Types of Cross Validation , Hold-out , K-fold , Grid and random search for Parameter tuning.

Unit-7

Support Vector Machines (SVM):

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

Unit-8

Ensemble Techniques:

Boosting , AdaBoost , Gradient Boosting , XGBoost

J.C BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD

Unit-9

Principal Component Analysis:

Intro to Dimensionality Reduction, Computing Components in PCA, Dimensionality Reduction using PCA

Unit-10

K-Means Clustering:

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

Unit-11

Hierarchical clustering | Intro to Hierarchical Clustering | Dendrogram | Types of Hierarchical Clustering: Agglomerative and Divisive , Cluster Linkage

Course Outcomes:

1. Gain knowledge about basic concepts of Machine Learning
2. Identify machine learning techniques suitable for a given problem
3. Solve the problems using various machine learning techniques
4. Apply Dimensionality reduction techniques.

J.C BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD

SUBJECT NAME: DEEP LEARNING

PAPER CODE: PG-DSA-204

Semester -2nd	Sessional – 25 Marks
Stream–Data Science & Analytics	Theory– 75 Marks
L T P Total Credits	Total: 100 Marks
4 0 0 4	Duration of Exam: 3 Hours

Course Objectives:

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

Course Contents:

MODULE-1: INTRODUCTION TO NEURAL NETWORKS

Artificial neurons, Neural networks and architectures, Activation function-linear and non-linear, Learning types-supervised, unsupervised and reinforced, McCulloch Pitts neuron model, Learning rules-Delta learning rule, Competitive learning rule, Hebbian learning rule and Perceptron learning rule, etc.

MODULE-2: CLASSIFIERS

Single Perceptron as classifier, Perceptron networks for linearly separable classification, Multi-layer perceptron model, Adaline, Madaline.

MODULE-3: DEEP NEURAL NETWORKS

Back Propagation Algorithms, Sequence learning problems, Introduction to Recurrent neural networks, Vanishing gradient problem, LSTM network, Basic structure of convolution neural networks.

MODULE-4: DIMENSION REDUCTION TECHNIQUES

Introduction to PCA, Dimensionality reduction Using PCA, Singular Value Decomposition.

Course Outcomes:

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

J.C BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD

SUBJECT NAME: MATHEMATICS

PAPER CODE: PG-DSA-205

Semester -2nd | Sessional – 25 Marks

Stream–Data Science & Analytics | Theory– 75 Marks

L T P Total Credits | Total: 100 Marks

4 0 0 4 | Duration of Exam: 3 Hours

Course Contents:

1. Scalars , Vectors , Matrix ,Tensors
2. Vector space ,Linearly independent and linearly dependent set of vectors, Basis and dimension of a vector space , Norm of Matrix
3. Linear Transformations and its matrix representation, Elementary transformations, Rank of a transformation, Rank- nullity theorem
4. Matrix decomposition, Quadratic forms, Geometry of positive definite quadratic form, Determinant, Partitioning of matrices
5. Eigen value, Eigen vector, Cayley-Hamilton theorem, Similarity of matrices, Diagonalization of matrices
6. 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
7. Principal Component Analysis

Course Outcomes:

1. analyze the solution set of a system of linear equations.
2. express some algebraic concepts (such as binary operation, group, field).
3. express a system of linear equations in a matrix form.
4. definite a vector space and subspace of a vector space.
5. describe the concepts of eigenvalue, eigenvector and characteristc polynomial.

J.C BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD

SUBJECT NAME: MACHINE LEARNING LAB

PAPER CODE: PCC-WD-206

Semester: 2nd	Internal – 30 Marks
Stream– Data Science & Analytics	External – 20 Marks
L T P Total Credits	Total - 50 Marks
0 0 3 3	Duration of Exam: 3 Hours

1. 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.
2. 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.
3. Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.
4. Build an Artificial Neural Network by implementing the Back propagation algorithm and test the same using appropriate data sets.
5. 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.
6. 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.
7. 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.
8. 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.
9. Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions. Java/Python ML library classes can be used for this problem.
10. Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.

Course Outcomes:

Real World Problem solving and programming capability

J.C BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD

SUBJECT NAME: PYTHON LAB

PAPER CODE: PCC-WD-207

Semester: 2nd | Internal – 30 Marks

Stream– Data Science & Analytics | External – 20 Marks

L T P Total Credits | Total - 50 Marks

0 0 3 3 | Duration of Exam: 3 Hours

Course Objectives:

The course is designed 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:

Unit-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

Unit-2

- **Packages in python:**

Introduction to Package, .py file, Importing module from a package, Creating a Package, Creating Sub Package, Importing from Sub-Packages, Popular Python Packages

Unit-3

Python predefined module

Date & DateTime Class, Format Time Output, Timedelta Objects, Calendar Module,os module.

Unit-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.

Unit-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.

Unit-6

Object Oriented Programming System:

Define oops concept, Difference between class variable and instance variable, Difference between function and method, Define class, What are object and instance of a class?, Name all features of oops, encapsulation , Polymorphism, Inheritance, Differentiate among instance method , class method , static method.

Unit-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.

Unit-8

Database Connectivity:

Creating Database connection, Understanding cursor, Executing Queries, Parameterised Queries

Unit-9

Multithreading:

Intrduction to Multithreading, Threading Module, Define a thread, Thread Synchronization.

Unit-10

GUI Programming-tkinter:

Introduction, Components and events, Adding control, Label, Button, Entry, Text , Radio, Check widget, ListBox, Menu, Combobox

Course Outcomes:

Problem solving and programming capability.

J.C BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD

SUBJECT NAME:IMAGE ANALYSIS AND COMPUTER VISION
PAPER CODE: PG-DSA-209

Semester -2nd	Sessional – 25 Marks
Stream–Data Science & Analytics	Theory– 75 Marks
L T P Total Credits	Total: 100 Marks
4 0 0 4	Duration of Exam: 3 Hours

Course Objectives:

The course is designed to provide knowledge of image analysis to scaling , transforming , object detection , classification of image.

Unit-1

Morphological Image Processing: Basic concept of set theory, logic operation involving binary images,dilation and erosion, opening and closing, and hit-or-miss transformation. Some basic morphological algorithms – Boundary extraction, region filling, extraction of connected components, convex hull, thinning, thickening, skeletons, and pruning. Extensions to gray-scale images – Dilation, Erosion, Opening and closing, and application of gray scale morphology.

Unit-2

Image segmentation: Detection of discontinuities – Point detection, line detection, edge detection – gradient operators, compass operators, laplace operators and zero crossing, stochastic gradients, performance of edge detector operators. Amplitude thresholding or window slicing, component labeling, boundary based approaches, region-based approaches and clustering, template matching, and texture segmentation.

Unit-3

Boundary Extraction: Connectivity, Contour following, Edge linking and heuristic graph searching, dynamic programming, and Hough transform

Unit-4

Region Representation: Run-length codes, Quad-trees, topological descriptor, texture and projections.

Unit-5

Moment representation: Moment representation theorem, Moment matching, Orthogonal moments, Moment invariants. Applications of moment invariants.

Unit-6

Shape feature: Geometry features, Moment-based features.

Unit-7

Texture: Statistical approaches, structural approaches, and other approaches.

Unit-8

Scene matching: Image subtraction, template matching and area correlation, and matched filtering.

Unit-9

Object recognition and image understanding: Patterns and pattern classes, decision theoretic and structural methods.

Course Outcomes:

Object detection, Image classification

SUBJECT NAME: DIGITAL SPEECH PROCESSING AND RECOGNITION

PAPER CODE: PG-DSA-210

Semester -2nd	Sessional – 25 Marks
Stream–Data Science & Analytics	Theory– 75 Marks
L T P Total Credits	Total: 100 Marks
4 0 0 4	Duration of Exam: 3 Hours

Course Objectives:

The course is designed to provide knowledge of speech analysis ,classification of speech etc.

Unit-1

Introduction to Digital Speech Processing: Review of DSP Fundamentals, Acoustic Theory of Speech Production, Speech Perception--Auditory Models, Sound Perception Models, MOS Methods, Sound Propagation in the Vocal Tract..

Unit-2

Speech Coding and Synthesis: Time Domain Methods in Speech Processing, Methods of Pitch Period Estimation, Speech Representations Based on STFT Analysis-Synthesis Methods, Homomorphic Speech Processing, Linear Predictive Coding (LPC) Methods, Speech Coding Methods--Model-Based Approaches. Cepstral Analysis.

Unit-3

Speech Enhancement: Introduction, Classification of Speech Enhancement Methods, Short-Term Spectral Amplitude Techniques, Speech Modeling and Wiener Filtering, Speech Enhancement and AllPol Modeling, Sequential Estimation via EM Theory, Constrained Iterative.

Unit-4

Speech Quality Assessment: The Need for Quality Assessment, Quality Versus Intelligibility. Subjective Quality Measures - Intelligibility Tests , Quality Tests. Objective Quality Measures - Articulation Index , Signal-to-Noise Ratio, Itakura Measure, Other Measures Based on LP Analysis, Weighted-Spectral Slope Measures, Global Objective Measures, Example Applications. Objective Versus Subjective Measures.

Unit-5

The Speech Recognition Problem: Introduction , Speaker-Dependent Versus Speaker-Independent, Recognition , Vocabulary Size, Isolated-Word Versus Continuous-Speech, Recognition, Linguistic Constraints, Acoustic Ambiguity and Confusability, Environmental Noise, Speaker Recognition and Verification. Dynamic Time Warping, Hidden Markov Model(HMM) based speech modeling, N-Gram Statistical Models, and Other Formal Grammars, standard Databases for Speech-Recognition Research.

Course Outcomes:

Speech Recognition, Speech classification

J.C. Bose University of Science & Technology, YMCA, Faridabad

(A Haryana State Government University)

(Established by Haryana State Legislative Act No. 21 of 2009 & Recognized by UGC Act 1956 u/s 22 to Confer Degrees)

Accredited 'A' Grade by NAAC

GOLDEN JUBILEE YEAR (1969-2019)

COMMUNITY COLLEGE OF SKILL DEVELOPMENT (CCSD)

Program PG Diploma in Data Science & Analytics (program Code: 271)
Revised Scheme Course Index of the year 2020-2021(BOS Dated 18/05/2020)
Mapping of the course with the employability/Entrepreneurship/Skill Development

S.NO	Course	Code	Skill development	Entrepreneurship	Employability
1	INTRODUCTION TO DATA SCIENCE & MACHINE LEARNING	PG-DSA-101	✓		✓
2	STATISTICS & PROBABILITY	PG-DSA- 102	✓	✓	
3	ADVANCED DATABASE MANAGEMENT SYSTEM	PG-DSA-103	✓		✓
4	DATA STRUCTURES & ALGORITHMS	PG-DSA-104	✓		✓
5	PYTHON LAB	PG-DSA-105	✓	✓	✓
6	SQL LAB	PG-DSA-107	✓	✓	✓
7	FUNDAMENTALSOF DATA SCIENCE	PG-DSA-201	✓		✓
8	ELEMENTS OFSTATISTICAL LEARNING	PG-DSA-202	✓	✓	
9	MACHINE LEARNING	PG-DSA-203	✓		✓
10	DEEP LEARNING	PG-DSA-204	✓		✓
11	MATHEMATICS	PG-DSA-205	✓	✓	
12	MACHINE LEARNING LAB	PG-DSA-206	✓	✓	✓
13	PYTHON LAB	PG-DSA-207	✓	✓	✓
14	MINOR PROJECT	PG-DSA-208	✓	✓	✓
15	IMAGE ANALYSIS AND COMPUTER VISION	PG-DSA-209	✓		✓
16	DIGITAL SPEECH PROCESSING AND RECOGNITION	PG-DSA-210	✓		✓

Principal CCSD