

**Savitribai Phule Pune University  
(Formerly University of Pune)**



**Department of Technology**

**STRUCTURE OF ONE YEAR FULL TIME POST GRADUATE DIPLOMA IN  
Data Science and Artificial Intelligence (PGD-DS&AI)**

**Semester 1**

Sr. No.	Course Code	Course Name	Teaching Scheme			Credits
			L	T	P	
1.	PGDS101	Introduction to Programming	2	1	0	2
2.	PGDS102	Python Programming	1	1	1	2
3.	PGDS103	R Programming for Data Science	1	1	1	2
4.	PGDS104	Statistics Essential for Data Science-1	2	1	0	2
5.	PGDS105	Introduction to Database Management System	2	1	0	2
6.	PGDS106	Programming Lab-1	0	2	2	4
7.	PGDS107	Seminar-1	0	1	0	2
		<b>Total Credits</b>				<b>16</b>

**Semester 2**

Sr. No.	Course Code	Course Name	Teaching Scheme			Credits
			L	T	P	
1.	PGDS201	Statistics Essential for Data Science-2	2	1	0	2
2.	PGDS202	Machine Learning & Artificial Intelligence	2	0	1	2
3.	PGDS203	Data Visualization	0	1	2	2
4.	PGDS204	Natural Language Processing	1	1	1	2
5.	PGDS205	Introduction to Deep Learning	1	1	1	2
6.	PGDS206	Programming Lab-2	0	2	2	4
7.	PGDS207	Seminar-2	0	1	0	2
		<b>Total Credits</b>				<b>16</b>

**Semester 3**

Sr. No.	Course Code	Course Name	Teaching Scheme			Credits
			L	T	P	
1	PGDS301	Data Science Capstone Project	0	3	3	10
2	PGDS302	Massive Open Online Courses (MOOCs)	0	4	0	2
		<b>Total Credits</b>				<b>12</b>
		<b>Course Total Credits</b>				<b>44</b>

**Subject Code: PGDS101**

**Subject Name: Introduction to Programming**

**Teaching Scheme:**

Lectures: 3 Hrs / week

**Credit**

2

**Examination Scheme:**

Mid Semester: 25

End Semester: 50

**Total: 100**

**Unit-1: Fundamentals of Programming Language**

**05 Hrs**

Importance of Studying Programming Languages, History of Programming Languages, Impact of Programming Paradigms, Role of Programming Languages, Programming Environments. Impact of Machine Architectures: The operation of a computer, Virtual Computers and Binding Times. Programming paradigms- Introduction to programming paradigms, Introduction to four main Programming paradigms- procedural, object oriented, functional, and logic & rule based.

**Unit-2: Introduction to C language**

**3 Hrs**

Introduction to c, Features of data types placeholders, Compilation Process control statements, Loop control structure, Functions, Pre-processing and header files, Arrays and strings, Structure and unions, File handling

**Unit-3: Introduction to OOP**

**5 Hrs**

Introduction to procedural, modular, generic, and object-oriented programming techniques, limitations of procedural programming, need of object-oriented programming, OOP Paradigms, Fundamentals of object-oriented programming: Namespaces, objects, classes, data members, methods, messages, data encapsulation, data abstraction and information hiding, inheritance, polymorphism. Benefits of OOP, C++ as object-oriented programming language.

**Unit-4: Introduction to Python**

**3 Hrs**

Introduction to Python and Computer Programming, Data Types, Variables, Basic Input-Output Operations, Basic Operators, Boolean Values, Conditional Execution, Loops, Lists and List Processing, Logical and Bitwise Operations, Functions, Tuples, Dictionaries, and Data Processing.

**Unit-5: Introduction to R**

**2 Hrs**

Overview of R, R data types and objects, reading and writing data, Control structures, functions, scoping rules, dates and times, Loop functions, debugging tools, Simulation, code profiling.

**Unit-6: Introduction to Natural Language Processing**

**2Hrs**

Introduction, what is NLP, what is Data Pre-processing, what are Stop words, what is Tokenization, Stemming, Lemmatization, Modelling Techniques in NLP.

**References:**

1. Sebesta R., "Concepts of Programming Languages", 4th Edition, Pearson Education, ISBN81- 7808-161-X
2. Introduction to Object Oriented Programming (English, Paperback, S. Sureshkumar)
3. A Textbook of Basics of C-Language Programming (English, Paperback, K Bikshalu)
4. Python Programming Fundamentals - A Beginner's Handbook (Nischay Kumar Hegde)

**Subject Code: PGDS102**

**Subject Name: Python Programming**

**Teaching Scheme:**

Lectures: 3 Hrs / week

**Credit**

2

**Examination Scheme:**

Mid Semester: 25

End Semester: 50

**Total: 100**

**Unit-1: Problem Solving, Programming and Python Programming 07 Hrs**

General Problem-Solving Concepts- Problem solving in everyday life, types of problems, problem solving with computers, difficulties with problem-solving, problem-solving aspects, top-down design. Problem Solving Strategies, Program Design Tools: Algorithms, Flowcharts and Pseudo-codes, implementation of algorithms. Basics of Python Programming: Features of Python, History and Future of Python, Writing and executing Python program, Literal constants, variables and identifiers, Data Types, Input operation, Comments, Reserved words, Indentation, Operators and expressions, Expressions in Python.

**Unit-2: Decision Control Statements 08 Hrs**

Decision Control Statements: Decision control statements, Selection/conditional branching Statements: if, if-else, nested if, if-elif-else statements. Basic loop Structures/Iterative statements: while loop, for loop, selecting appropriate loop. Nested loops, the break, continue, pass, else statement used with loops. Other data types- Tuples, Lists and Dictionary.

**Unit-3 Functions and Modules 08 Hrs**

Need for functions, Function: definition, call, variable scope and lifetime, the return statement. Defining functions, Lambda or anonymous function, documentation string, good programming practices. Introduction to modules, Introduction to packages in Python, Introduction to standard library modules.

**Unit-4: Strings 07 Hrs**

Strings and Operations- concatenation, appending, multiplication and slicing. Strings are immutable, strings formatting operator, built in string methods and functions. Slice operation, ord() and chr() functions, in and not in operators, comparing strings, Iterating strings, the string module.

**Unit-5: Object Oriented Programming 08 Hrs**

Programming Paradigms-monolithic, procedural, structured and object oriented, Features of Object-oriented programming-classes, objects, methods, and message passing, inheritance, polymorphism, containership, reusability, delegation, data abstraction and encapsulation. Classes and Objects: classes and objects, class method and self-object, class variables and object variables, public and private members, class methods.

**Unit-6: Working with Data in Python 07 Hrs**

Introduction• Working with NumPy Arrays, Pandas in Python, Useful Ways to View Data Frame objects in Python.

**References:**

1. Python Data Science Handbook - Essential Tools for Working with Data (Jake VanderPlas)
2. DATA SCIENCE AND ANALYTICS WITH PYTHON (JESUS ROGEL - SALAZAR)
3. Mastering Python for Data Science (Madhavan Samir)
4. R. G. Dromey, "How to Solve it by Computer", Pearson Education India; 1st edition, ISBN10: 8131705625, ISBN-13: 978-8131705629 Maureen Spankle, "Problem Solving and Programming Concepts", Pearson; 9th edition, ISBN-10: 9780132492645, ISBN-13: 978- 0132492645
5. Romano Fabrizio, "Learning Python," Packt Publishing Limited, ISBN: 9781783551712, 1783551712
6. Paul Barry, "Head First Python- A Brain Friendly Guide", SPD O'Reilly, 2nd Edition, ISBN:978-93-5213-482-3

7. Martin C. Brown, "Python: The Complete Reference", McGraw Hill Education, ISBN-10: 9789387572942, ISBN-13: 978-9387572942, ASIN: 9387572943
8. Jeeva Jose, P. Sojan Lal, "Introduction to Computing & Problem Solving with Python", Khanna Computer Book Store; First edition, ISBN-10: 9789382609810, ISBN-13: 978-9382609810

**Text Books:**

1. Reema Thareja, "Python Programming Using Problem Solving Approach", Oxford University Press, ISBN 13: 978-0-19-948017-6
2. R. Nageswara Rao, "Core Python Programming", Dreamtech Press; Second edition ISBN10: 938605230X, ISBN-13: 978-9386052308 ASIN: B07BFSR3LL

**Subject Code: PGDS103**

**Subject Name: R Programming for Data Science**

**Teaching Scheme:**

**Lectures: 3 Hrs / week**

**Credit**

**2**

**Examination Scheme:**

**Mid Semester: 25**

**End Semester: 50**

**Total: 100**

**08 Hrs**

**Unit 1: - Basics of R-Programming**

Evolution of R, Features of R, Local Environment support, R Command prompt, R Script File, Comment, R Data types, R Variables, R Operators-function.

**Unit 2: - R Fundamentals**

**08 Hrs**

R-Decision Making: - R-If statement, R-If... else statement, R- The if... else if...else statement-Switch Statement, R- Loop: - Repeat loop, While loop, for loop, Loop, Control statement: - Break, Next.

**Unit 3: -Data Structures in R**

**10 Hrs**

Vectors: -Vector Creation, Accessing Vector Elements, Vector Manipulation, Lists: - Creating Lists, Naming List Elements, Accessing List Elements, Manipulating List Elements, Merging Lists, Covering Lists to Vectors. Matrices: - Accessing Elements of Matrix, Matrix Computation. Arrays: - Naming columns & rows, Accessing Array Elements, Manipulating Array Elements, and Calculations across Array Elements. Factors: - Factors in Data Frames, changing order of levels, Generating factor levels. Data frames: - Extract data from data frame, Expand data frame.

**Unit 4: - Working with Data in R**

**10 Hrs**

R-Reshaping: - Joining rows and columns, merging data frames, melting and casting. R- CSV Files: - Getting and starting with directory, Input as a CSV file, Reading CSV file, Analyzing CSV file, writing to CSV file. R- EXCEL File: - Install xlsx Packages, Verify & Load "xlsx" packages, Input as a xlsx file, Reading excel file. R- Binary File: - writing binary file, reading binary file. R- XML File: - Input data, Reading XML file, details of the first node, xml to data node.

**Unit 5: - Statistics in R**

**08 Hrs**

R- Statistics Introduction: Mean, median and mode, Minimum and maximum value, Percentiles, Variance and Standard Deviation, Covariance and Correlation, Probability distributions. R Data Set: - Print Variable Values, Sort Variable Values, Analyzing the Data. R Max and Min, Max and Min, Outliers. R Mean: - Mean, Median, and Mode, R Percentiles.

**Unit 6: - Application of R- programming**

**08 Hrs**

R- Pie charts: - Pie chart title and colour, 3-D Pie Chart. R- Bar Chart: - Bar Chart Labels, Title and colour, Group Bar chart and stacked bar chart. R- Box Plot: - Creating a box plot, Box plot with notch. R- Histogram: - Range of x and y values.

**Reference Books: -**

1. Mark Gardener, Beginning R: The Statistical Programming Language, Wiley India, ISBN:9788126541201 16.
2. Mark Gardener, The Essential R Reference, Wiley India, ISBN: 9788126546015 17.
3. Judith Hurwitz, Alan Nugent, Big Data for Dummies, Wiley India, ISBN: 9788126543281.
4. "Practical Data Science with R", Second Edition, Nina Zumel and John Mount  
*Foreword by Jeremy Howard and Rachel Thomas*, November 2019, ISBN 9781617295874.
5. "Hands-On Programming with R: Write Your Own Functions and Simulations," by Garrett Golemund.

**Subject Code: PGDS104****Subject Name: Statistics Essential for Data Science-1****Teaching Scheme:****Credit****Examination Scheme:****Lectures: 3 Hrs / week****2****Mid Semester: 25****End Semester: 50****Total: 100****Unit-1: Maths and Statistics for Data Science****3 Hrs**

Introduction: Sample or Population Data, Understanding the Data, various data types, Various variable types, uses of variable types, Population and Sample, Sampling techniques, Data representation, Numerical parameters to represent data, Mean, Mode, Median, Sensitivity, Information Gain, Entropy, Statistical parameters to represent data.

**Unit-2: The Fundamentals of Descriptive Statistics****3 Hrs**

**Probability and its uses, Uses of probability, Need of probability, Bayesian Inference, Density Concepts, Normal Distribution Curve, Measures of Central Tendency, Asymmetry, and Variability, Practical Example: Descriptive Statistics.**

**Unit-3: Data Clustering****4 Hrs**

**Association and Dependence, Causation and Correlation, Covariance, Simpson's Paradox, Clustering Techniques, B testing.**

**Unit 4: Statistical Inference****5 Hrs**

Point Estimation, Confidence Margin, Hypothesis Testing, Levels of Hypothesis Testing, Estimators and Estimates, Confidence Intervals: Advanced Topics, Practical Example: Inferential Statistics, Hypothesis Testing: Introduction

**Unit-5: Testing the Data****3 Hrs**

Parametric Test, Parametric Test Types, Non- Parametric Test, Experimental Designing, Hypothesis Testing: Let's Start Testing, Practical Example: Hypothesis Testing, The Fundamentals of Regression Analysis, Subtleties of Regression Analysis, Assumptions for Linear Regression Analysis, Dealing with Categorical Data, Practical Example: Regression Analysis.

**Unit-6: Regression Modelling****2 Hrs**

Logistic and Regression Techniques, Problem of Collinearity, WOE and IV, Residual Analysis, Heteroscedasticity, Homoscedasticity.

**Reference Books: -**

1. Statistics for Data Science (Miller James D.)
2. Practical Statistics for Data Scientists (Bruce Peter)
3. Statistical Data Book 2nd Edition (R. S. Nagarajan, C. Muralidharan)

**Subject Code: PGDS105**

**Subject Name: Introduction to Database Management System**

**Teaching Scheme:**

**Credit**

**Examination Scheme:**

Lectures: 3 Hrs / week

2

Mid Semester: 25

End Semester: 50

**Unit-1: Data Wrangling with SQL**

**10 Hrs**

SQL Basics – Fundamentals of Structured Query Language, SQL Tables, Joins, Variables

Advanced SQL – SQL Functions, Subqueries, Rules, Views, Nested Queries, string functions, pattern matching, Mathematical functions, Date-time functions, etc.

Deep Dive into User Defined Functions -Types of UDFs, Inline table value, multi-statement table, Stored procedures, rank function, triggers, etc. SQL Optimization and Performance - Record grouping, searching, sorting, etc., Clustered indexes, common table expressions.

Record grouping, searching, sorting, etc.

**Unit-2: Introduction to Big Data and Spark**

**04 Hrs**

Apache spark framework, RDDs, Stopgaps in existing computing methodologies

**RDDs** – RDD persistence, caching, General operations: Transformation, Actions, and Functions, Concept of Key-Value pair in RDDs, Other pair, two pair RDDs, RDD Lineage, RDD Persistence, Word Count Program Using RDD Concepts, RDD Partitioning & How it Helps Achieve Parallelization.

**Unit-3: Advanced Concepts & Spark-Hive**

**04 Hrs**

Passing Functions to Spark, Spark SQL Architecture, SQL Context in Spark SQL, User-Defined Functions, Data Frames, Interoperating with RDDs, Loading Data through Different Sources, Performance Tuning, Spark-Hive Integration.

**Unit-4: Extract Transform Load**

**04 Hrs**

Web Scraping, Interacting with APIs, **Data Handling with NumPy** - NumPy Arrays, CRUD Operations, etc., Linear Algebra – Matrix multiplication, CRUD operations, Inverse, Transpose, Rank, Determinant of a matrix, Scalars, Vectors, Matrices.

**Data Manipulation Using Pandas** -Loading the data, data frames, series, CRUD operations, splitting the data, etc. **Data Pre-processing** -Exploratory Data Analysis, Feature engineering, Feature scaling, Normalization, standardization, etc. Null Value Imputations, Outliers Analysis and Handling, VIF, Bias-variance trade-off, cross validation techniques, train-test split, etc.

**Data Visualization**-Bar charts, scatter plots, count plots, line plots, pie charts, donut charts, etc, with **Python Matplotlib**, Regression plots, categorical plots, area plots, etc, with **Python seaborn**.

**Unit-5: Deploying Machine Learning Models with Cloud**

**04 Hrs**

Introduction to MLOps-MLOps lifecycle, MLOps pipeline, MLOps Components, Processes, etc. Deploying Machine Learning Models- Introduction to Azure Machine Learning, Deploying Machine Learning Models using Azure.

**References:**

1. Visual Data Storytelling with Tableau, (4 Colour), 1st Edition by Pearson  
Lindy Ryan
2. Mastering Tableau 2019.1 Meier Marleen
3. Tableau Your Data! - Fast and Easy Visual Analysis With Tableau Software  
Murray Dan

**Subject Code: PGDS106**

**Subject Name: Programming Lab-I**

**Teaching Scheme:**

Practical: 4 Hrs / week

**Credit**

4

**Examination Scheme:**

End Semester: 100

Assignments related to Python, R and SQL

**Python--**

**Program: 1-COMPUTE THE RESULT GRADEWISE**-To accept student is five courses' marks and compute his/her result. Student is passing if he/she scores marks equal to and above 40 in each course. If student scores aggregate greater than 75%, then the grade is distinction. If aggregate is  $60 \geq$  and  $< 75$  then the grade is first division. If aggregate is  $50 \geq$  and  $< 60$ , then the grade is second division. If aggregate is  $40 \geq$  and  $< 50$ , then the grade is third division.

**Program: 2-CHECK IF THE NUMBER IS ARMSTRONG NUMBER**-To check whether input number is Armstrong number or not. An Armstrong number is an integer with three digits such that the sum of the cubes of its digits is equal to the number itself.

**Program: 3-CALCULATOR PROGRAM**-To simulate simple calculator that performs basic tasks such as addition, subtraction, multiplication and division with special operations like computing  $xy$  and  $x!$

**Program: 4** - To accept the number and Compute a) square root of number, b) Square of number, c) Cube of number d) check for prime, d) factorial of number e) prime factors.

**Program: 5-REVERSING THE NUMBER** -To accept a number from user and print digits of number in a reverse order.

**Program: 6-CONVERSION OF BINARY TO DECIMAL-** To input binary number from user and convert it into decimal number.

**Program: 7 -PSEUDO RANDOM NUMBERS-** To generate pseudo random numbers.

**Program: 8-PARTITION IN EVEN AND ODD-** To accept list of N integers and partition list into two sub lists even and odd numbers.

**Program: 9-FIBONACCI SERIES-** To accept from user the number of Fibonacci numbers to be generated and print the Fibonacci series.

**Program: 11 -FILE OPERATIONS -AIM:** To copy contents of one file to other. While copying a) all full stops are to be replaced with commas b) lower case are to be replaced with upper case c) upper case are to be replaced with lower case.

**Program: 12 -FILE COUNT -To** count total characters in file, total words in file, total lines in file and frequency of given word in file.

**Program: 13 -EMPLOYEE CLASS -** Create class EMPLOYEE for storing details (Name, Designation, gender, Date of Joining and Salary). Define function members to compute a) total number of employees in an organization b) count of male and female employee c) Employee with salary more than 10,000 d) Employee with designation "Asst Manager"

## SQL—

1)Bank Database Consider the following database of Bank. A bank maintains the customer details, account details and loan details. It has the Branch information also. Following are the tables:

ACCOUNT(ACC\_NO INT, ACC\_TYPE CHAR(10), BALANCE FLOAT(8,2))

LOAN(LOAN\_NO INT, LOAN\_AMT DOUBLE(9,2) , NO\_OF\_YEARS INT)

BRANCH(BRANCH\_NO INT, BRANCH\_NAME CHAR(20), BRANCH\_CITY

VARCHAR(20)) CUSTOMER(CUST\_NO INT , CUST\_NAME CHAR(20), CUST\_STREET CHAR(15), CUST\_CITY VARCHAR(20))

The relationships are as follows. :-

CUSTOMER-ACCOUNT: 1-M

CUSTOMER- LOAN: 1-M

BRANCH-LOAN: 1-M

BRANCH-ACCOUNT: 1:M

Constraints:

1) use auto\_increment data type attribute for cust\_no

2) branch\_name should be not null.

**2) Bus transport System Consider the following database of Bus transport system . Many buses run on one route. Drivers are allotted to the buses shiftwise. Following are the tables:**

BUS (BUS\_NO INT , CAPACITY INT , DEPOT\_NAME VARCHAR(20))

ROUTE (ROUTE\_NO INT, SOURCE CHAR(20), DESTINATION CHAR(20),

NO\_OF\_STATIONS INT) DRIVER (DRIVER\_NO INT , DRIVER\_NAME CHAR(20),

LICENSE\_NO INT, ADDRESS CHAR(20), D\_AGE INT , SALARY FLOAT)

The relationships are as follows: BUS\_ROUTE : M-1 BUS\_DRIVER : M-M with descriptive attributes Date of duty allotted and Shift — it can be 1 (Morning) Or 2 ( Evening ).

Constraints:

1. License\_no is unique.

2. Bus capacity is not nul

**3)Client-Policy Database Consider an insurance company which has agents. Clients select a particular policy ang go for the policy through the agents.** Company manintains information about

the clients and agents . Whenever client takes a policy , agent validates the information of client such as age of the 11 client should be in the range of the selected policy( i.e. Age should be between minimum\_age\_limit and maximum\_age\_limit.) , sum\_assured also should be between the min\_sum\_assured and max\_sum\_assured. The client gets a unique policy number , decides the premium amount , type\_of\_premium , nominee name etc. The policy term is calculated as the

maturity age of the selected policy — age of the client.

POLICY(POLICY\_NAME VARCHAR(20), MIN\_AGE\_LIMIT INTEGER, MAX\_AGE\_LIMIT INTEGER, MATURITY\_AGE INTEGER, MIN\_SUM\_ASSURED INTEGER, MAX\_SUM\_ASSURED INTEGER); CLIENT (CLIENT\_ID INTEGER, NAME VARCHAR (25), BIRTH\_DATE DATE, NOMINEE\_NAME VARCHAR (25), RELATION\_WITH\_CLIENT VARCHAR (20)); AGENT (AGENT\_ID INTEGER, NAME VARCHAR (25), LICENSE\_NO INTEGER, BRANCH\_OFFICE VARCHAR (20)); Relationship between: POLICY, CLIENT and AGENT are ternary with described attributes POLICY\_NO, PREMIUM AMOUNT, POLICY\_DATE, TYPE\_OF\_PREMIUM, SUM\_ASSURED and POLICY TERM. The relationship table is:

AGENT\_CLIENT\_POLICY (AGENT\_ID INTEGER, CLIENT\_ID INTEGER, POLICY\_NAME VARCHAR (20), POLICY\_NO INTEGER, PREMIUM DECIMAL(7,2), POLICY\_DATE DATE, TYPE VARCHAR(20), SUM\_ASSURED DECIMAL(7,2), TERM INTEGER).

type: is the type of premium which can be 'q' (quarterly), 'h' (half yearly), 'y' (Yearly). While inserting records in relationship table enter the type value as one of the 'q', 'h', 'y'. Constraints: 1. use auto\_increment data type attribute for client\_id. 2. policy\_no is unique.

**4) Real Estate Database Consider the Real Estate Agency Database where estate agents are sold by many agents. Estates are purchased by Customers from Agents and agent will get a commission.**

Following are the tables:

1. AGENT (AID INT, ANAME VARCHAR (20), ADDRESS VARCHAR (20), CONTACTNO VARCHAR (10));
2. ESTATE (ENO INT, TYPE VARCHAR (20), LOCATION VARCHAR (20), PRICE INT);
3. CUSTOMER (CUSTID INT, CNAME VARCHAR (20), CONTACTNO VARCHAR (10), ADDRESS VARCHAR (30));

Type: Estate type can be 1bhk flat, land, 2bhk flat etc.

The Relationship Between: 12 Agent, Estate and Customer is a ternary relationship. A relationship table Transaction will store the transaction about customer purchased estate from agent. A customer can purchase many estates from one or more agents. A transaction table is:

TRANSACTION (AID, ENO, CUSTID, PURCHASEDATE DATE, COMMISSION DECIMAL (5,2)) Constraints:

1. Not null on price
2. Unique constraint on Eno on transaction table (one estate sold to only one customer)

**5) Mobile Billing Database Consider a database of Gigabyte Mobile Services which provide postpaid services to the customers. Different service plans are available from which a customer can select any one. The monthly customer call information is recorded (in custcallinfo table) and a bill is generated at the end. Following are the tables:**

1. PLAN (PLANNO INT, PNAME VARCHAR (20), NOOFFREECALLS INT, RATEPERCALLPERMIN DOUBLE, FREECALLTIME INT, FIXEDAMT DOUBLE);
2. CUSTOMER (CUSTNO INT, NAME VARCHAR (20), MOBILENO VARCHAR (10));
3. CUSTCALLINFO (RECNO INT, CUSTNO INT, NOOFCALLS INT, TOTALTAKLTIME INT, CYCLEPERIOD VARCHAR (20))

**6) BILL (RECNO INT, BILLNO INT, FINALBILLAMT DOUBLE, CYCLEDATE DATE, BILLDUEDATE DATE, STATUS VARCHAR (10), BILLPAYDATE DATE)**

**CYCLEPERIOD: shows the period between two particular months. e.g 'jan-feb', 'oct-nov'.  
CYCLEDATE: is in between the respective cycle period.**

Following are the relationships:

1. PLAN-CUSTOMER: 1-M
2. CUSTOMER-CUSTCALLINFO: 1-M
3. CUSTOMER-BILL :1-1 Constraint: 1. FIXEDAMT in plan is by default 0.

6) . Railway Reservation System Consider a railway reservation system of passengers. Passengers reserve berths of a bogie of trains. The bogie capacity of all the bogies of a train is same.

1. TRAIN (TRAIN\_NO INT, TRAIN\_NAME VARCHAR (20), DEPART\_TIME TIME, ARRIVAL\_TIME TIME, SOURCE\_STN VARCHAR (20), DEST\_STN VARCHAR (20), NO\_OF\_RES\_BOGIES INT, BOGIE\_CAPACITY INT) 13

2. PASSENGER (PASSENGER\_ID INT, PASSENGER\_NAME VARCHAR (20), ADDRESS VARCHAR (30), AGE INT, GENDER CHAR)

Relationship is as follows:

TRAIN\_PASSENGER: M-M with descriptive attributes as follows: TICKET (TRAIN\_NO INT, PASSENGER\_ID INT, TICKET\_NO INT COMPOSITE KEY, BOGIE\_NO INT, NO\_OF\_BERTHS INT, DATE DATE, TICKET\_AMT DECIMAL (7,2), STATUS CHAR) The status of a particular berth can be 'W' (waiting) or 'C' (confirmed).

**Subject Code: PGDS107**

**Subject Name: Seminar-I**

**Teaching Scheme:**

**Credit**  
2

**Examination Scheme:**

Presentation & Report Submission

Seminar related to any of their current subjects

**Subject Code: PGDS201**

**Subject Name: Statistics Essential for Data Science-2**

**Teaching Scheme:**

**Credit**  
2

**Examination Scheme:**

**Mid Semester: 25**

**End Semester: 50**

**Total: 100**

**Unit-1: Data Gathering and Data Discovery**

**3 Hrs**

Identifying potential data sources, Gathering data, Data discovery- understanding the data, assessing data, data formats, Parsing, Selecting features, Transformation, Scalability and real-time issues

**Unit-2: Cleaning and Conditioning Data**

**3 Hrs**

Data Preparation Basic Models: Data Integration, Data Cleaning, Data Normalization, Min-Max Normalization, Z-score Normalization, Decimal Scaling Normalization, Consistency checking, Heterogeneous and missing data, Dealing with missing values, Duplicate values, Noise, Inconsistent data, Outliers

**Unit-3: ETLT**

**4 Hrs**

Transform and enrich data: Data Transformation, Linear Transformations, Quadratic Transformations, Non-polynomial Approximations of Transformations, Polynomial Approximations of Transformations, Rank Transformations, Box-Cox Transformations, Spreading the Histogram, Nominal to Binary Transformation, Transformations via Data Reduction, ETL tools

**Unit 4: Exploratory Analysis**

**5 Hrs**

Formulating Hypothesis, Data Terminology, Data Exploration, Data Exploration through Summary Statistics, Data Exploration through Plots, Feature Engineering, Feature selection, Feature transformation, Dimensionality reduction

**Unit-5: Testing the Data**

**3 Hrs**

Parametric Test, Parametric Test Types, Non- Parametric Test, Experimental Designing, Hypothesis Testing: Let's Start Testing, Practical Example: Hypothesis Testing, The Fundamentals of Regression Analysis, Subtleties of Regression Analysis, Assumptions for Linear Regression Analysis, Dealing with Categorical Data, Practical Example: Regression Analysis.

**Unit-6: Advanced Tools for Data Preparation**

**2 Hrs**

Web scraping, Data from social networks, Open-source tools for data preparation: Open Refine, R/Python libraries for data preparation and visualization

**Textbooks:**

1. Glenn J. Myatt, “Making Sense of Data I: A Practical Guide to Exploratory Data Analysis and Data Mining” 2. Salvador García, JuliánLuengo, Francisco Herrera, “Data Preprocessing in Data Mining”

**Reference Books: -**

1. Statistics for Data Science (Miller James D.)
2. Practical Statistics for Data Scientists (Bruce Peter)
3. Statistical Data Book 2nd Edition (R. S. Nagarajan, C. Muralidharan)
4. ‘Python Crash Course’ by Eric Matthews
5. Big Data Analytics with Hadoop 3(Alla Sridhar)
6. Scikit-Learn: Machine Learning Simplified
7. Mark Gardner, “Beginning R: The Statistical Programming Language”, Wrox Publication, ISBN: 978-1-118-16430-3
8. David Dietrich, Barry Hiller, “Data Science and Big Data Analytics”, EMC education services, Wiley publications, 2012, ISBN0-07-120413-X
9. Ruben Verborgh; Max De Wilde, “Using OpenRefine : the essential OpenRefine guide that takes you from data analysis and error fixing to linking your dataset to the Web”

**Subject Code: PGDS202**

**Subject Name: Machine Learning & Artificial Intelligence**

**Teaching Scheme:**

Lectures: 3 Hrs / week

**Credit**

2

**Examination Scheme:**

Mid Semester: 25

End Semester: 50

**Unit-1: Decision Trees**

**04 Hrs**

Decision trees, a type of data mining algorithm that can select from among a large number of variables those and their interactions that are most important in predicting the target or response variable to be explained. Decision trees create segmentations or subgroups in the data, by applying a series of simple rules or criteria over and over again, which choose variable constellations that best predict the target variable.

**Unit-2: Random Forests**

**04 Hrs**

Random forests, a type of data mining algorithm that can select from among a large number of variables those that are most important in determining the target or response variable to be explained. Unlike decision trees, the results of random forests generalize well to new data.

### **Unit-3: Lasso Regression**

**04 Hrs**

Lasso regression analysis is a shrinkage and variable selection method for linear regression models. The goal of lasso regression is to obtain the subset of predictors that minimizes prediction error for a quantitative response variable. The lasso does this by imposing a constraint on the model parameters that causes regression coefficients for some variables to shrink toward zero. Variables with a regression coefficient equal to zero after the shrinkage process are excluded from the model. Variables with non-zero regression coefficients are most strongly associated with the response variable. Explanatory variables can be either quantitative, categorical or both. In this session, you will apply and interpret a lasso regression analysis. You will also develop experience using k-fold cross validation to select the best fitting model and obtain a more accurate estimate of your model's test error rate.

### **Unit-4: K-Means Cluster Analysis**

**04 Hrs**

Cluster analysis is an unsupervised machine learning method that partitions the observations in a data set into a smaller set of clusters where each observation belongs to only one cluster. The goal of cluster analysis is to group, or cluster, observations into subsets based on their similarity of responses on multiple variables. Clustering variables should be primarily quantitative variables, but binary variables may also be included. In this session, we will show you how to use k-means cluster analysis to identify clusters of observations in your data set. You will gain experience in interpreting cluster analysis results by using graphing methods to help you determine the number of clusters to interpret, and examining clustering variable means to evaluate the cluster profiles. Finally, you will get the opportunity to validate your cluster solution by examining differences between clusters on a variable not included in your cluster analysis.

### **Unit-5: GIT**

**04 Hrs**

Version Control -What is version control, types, SVN, Git Lifecycle, Common Git commands, working with branches in Git, GitHub collaboration (pull request), GitHub Authentication (SSH and Http), Merging branches, Resolving merge conflicts, Git workflow.

### **Unit-6: Deep Learning Using Tensor Flow**

**04 Hrs**

Artificial Intelligence Basics-Introduction to Keras API and TensorFlow, Neural Networks-Neural networks, Multi-layered Neural Networks, Artificial Neural Networks, Deep Learning-Deep neural networks, Convolutional Neural Networks, Recurrent Neural Networks, GPU in deep learning, Autoencoders, restricted Boltzmann machine.

### **References:**

1. Visual Data Storytelling with Tableau, (4 Colour), 1st Edition by Pearson  
Lindy Ryan
2. Mastering Tableau 2019.1 Meier Marleen
3. Tableau Your Data! - Fast and Easy Visual Analysis With Tableau Software  
Murray Dan

**Subject Name: Data Visualization****Teaching Scheme:**

Lectures: 3 Hrs / week

**Credit**

2

**Examination Scheme:**

Mid Semester: 25

End Semester: 50

Getting started with Importing Data Options	Unit 1	<p>Introduction to Data Visualization, BI Lifecycle, What is Analysis, importance of data visualization to the industry</p> <p>Why Data Visualization become so popular and where we used it. what is Data and what is Visualization Techniques, Data Wrangling Getting Started</p> <p>Importing Data : Excel As a source, SQL Server as a Source, Web as a Source</p> <p>Direct Query, Limitations,</p> <p>Live Connection, Limitation, which should I choose.</p>
Data Transformation Strategies	Unit 2	<p>Data Transformation Strategies</p> <p>The power query editor, transform basics</p> <p>Use first row as header, Remove Column, Change type</p> <p>Add Column from examples, Advanced Data Transformation options, Conditional Columns,</p> <p>Fill Down, unpivot, merging queries, appending queries.</p> <p>Leveraging R, Installation and configuration, The R Script transform, M Formula language, #shared</p>
Building the Data Model	Unit 3	<p>Building the Data Model</p> <p>building relationships, Editing relationships</p> <p>creating a new relationship</p> <p>Working with complex relationship</p> <p>many to many relationships, cross filtering directions</p> <p>Enabling filtering from the many side of a relationship</p> <p>role playing tables, Importing the date table</p> <p>Usability enhancements, Hiding tables and columns, Renaming tables and columns, Default summarization</p> <p>How to display one column but sort by another, Data Categorization, Creating hierarchies</p> <p>summary</p>
Leveraging DAX	Unit 4	<p>Leveraging DAX</p> <p>Building calculated columns</p> <p>String functions ; Month, Year</p> <p>Format function ; Month , Year</p> <p>Age Calculation, Switch () ; the basics calculated measures ; basic aggregation, Total Sales, Total Cost, Profit, Profit Margin, Optional Parameters</p> <p>Filter Context, Percentage of total calculation,</p> <p>Time Intelligence,</p> <p>year to date sales, YTD sales(Fiscal Calendar), Prior Year Sales</p> <p>Data Visualization Basics</p>

Visualizing Data	Unit 5	<p>Visuals for Filtering, Interactive Filtering, The Slicer Visual</p> <p>Visualizing tabular data : the table visual, the matrix visual</p> <p>Visualizing categorical data: Bar and Column charts, Pie and Donut Charts, The Treemap Visual, The scatter Chart</p> <p>Visualizing Trend Data: Line and Area Charts, The Waterfall Charts, The Funnel Chart</p> <p>Visualizing KPI Data: The Gauge Visual, The KPI Visual,</p> <p>Visualizing Geographical Data: The map visual, The Filled Map Visual, The ArcGIS map visual</p>
Digital Storytelling with Power BI	Unit 6	<p>Configuring drill through filters</p> <p>Storytelling with the selection pane and bookmarks</p> <p>Bookmarks pane</p> <p>Selection pane</p> <p>summary</p>

**Text Book:** Microsoft Power BI Quick Start Guide, By Devin Knight, Brain Knight, Mitchell Pearson and Manual quintana, Published by Packt Publishing Ltd, ISBN 978-1-78913-822-1

**Reference Book:** Microsoft Power BI Data Analyst Certification Guide by Orrin Edenfield-Edward

**Subject Code: PGDS204**

**Subject Name: Natural Language Processing**

**Teaching Scheme:**

Lectures: 3 Hrs / week

**Credit**

2

**Examination Scheme:**

Mid Semester: 25

End Semester: 50

**Total: 100**

**Unit-1: Introduction to NLP**

**03 Hrs**

What is NLP? Why NLP is Difficult? History of NLP, Advantages of NLP, Disadvantages of NLP, Components of NLP, Applications of NLP, How to build an NLP pipeline? Phases of NLP, NLP APIs, NLP Libraries

**Unit-2: Text Mining, Cleaning, and Pre-processing**

**03 Hrs**

Various Tokenizers, Tokenization, Frequency Distribution, Stemming, Types of Stemming, Stop Words, Normalization, POS Tagging, Lemmatization, Named Entity Recognition (NER), NLP Models- Bigrams, Trigrams & Ngram Model, Binary Weight.

**Unit-3: Text classification, NLTK, sentiment analysis, etc.**

**04 Hrs**

Overview of Machine Learning, Bag-of-Words, Term Frequency, Count vectorizer, Inverse Document Frequency, Text conversion, Confusion Matrix, Classification Metrics, Naive Bayes Classifier, Support vector Machine

**Unit-4: Sentence Structure, Sequence Tagging, Sequence Tasks, and Language Modelling**

**04 Hrs**

Language Modeling, Sequence Tagging, Sequence Tasks, Predicting Sequence of Tags, Syntax Trees, Context-Free Grammars, Chunking, Automatic Paraphrasing of Texts, Chinking.

**Unit 5: Machine Translation:**

Need of MT, Problems of Machine Translation, MT Approaches, Direct Machine Translations, Rule-Based Machine Translation, Knowledge Based MT System, Statistical Machine Translation (SMT), Parameter learning in SMT (IBM models) using EM, Encoder-decoder architecture, Neural Machine Translation

**Unit-6: AI Chabot's and Recommendations**

**Engine**

**03 Hrs**

Using the NLP concepts, build a recommendation engine and an AI chatbot assistant using AI.

**References:**

1. Natural Language Processing and Information Retrieval First Edition ( TIWARY, U.S, SIDDIQUI ,TANVEER)
2. Speech and Language Processing - Written by Daniel Jurafsky and James Martin.
3. Natural Language Processing with Python - Written by Steven Bird, Ewan Klein and Edward Loper.

**Subject Code: PGDS205**

**Subject Name: Introduction to Deep Learning**

**Teaching Scheme:**

Practical: 4 Hrs / week

**Credit**

4

**Examination Scheme:**

End Semester: 100

**Unit-1: Fuzzy Systems and Genetic Algorithms**

**03 Hrs**

Membership Functions, Fuzzification and Methods, Defuzzification and Methods, Fuzzy Logic, Fuzzy Rules and Fuzzy Reasoning, Fuzzy Inference Systems, Fuzzy Expert Systems, Fuzzy Decision Making. Fuzzy Control Systems, Fuzzy Classification. Genetic Algorithms: Introduction to Genetic Algorithms (GA), Search space, Working Principle, Simple GA, Operators, Fitness function, Multi-level Optimization.

**Unit-2: Introduction to Neural Networks**

**03 Hrs**

Perceptron's, Perceptron Learning Algorithm, Sigmoid Neuron, Shallow neural networks, Deep neural networks, Feedforward Neural networks, Gradient descent and the backpropagation algorithm

**Unit-3: Deep Learning**

**04 Hrs**

Learning Parameters of a feedforward neural network, the vanishing gradient problem, and ways to mitigate it, ReLU Heuristics for avoiding bad local minima, Heuristics for faster training, Nesterov Accelerated GD, Stochastic GD, AdaGrad, RMSProp, Adam, Momentum. Adagrad, Principal Component Analysis and its interpretations, Singular Value Decomposition.

**Unit-4: Neural Networks and its variants**

**04 Hrs**

Convolutional Neural Networks, LeNet, AlexNet, ZF-Net, VGGNet, GoogLeNet, ResNet, Visualizing Convolutional Neural Networks, Guided Backpropagation, Deep Dream, Deep Art, Fooling Convolutional Neural Networks

**Unit 5: Introduction to Soft Computing and Fuzzy logic**

**4 hrs**

Introduction to soft computing: , Paradigms soft computing, Features, Components, Techniques, Applications, Neural Networks, Fuzzy logic, Genetic Algorithms, Hybrid systems, Introduction to Fuzzy logic: Classical and Fuzzy sets, operations, properties, Fuzzy Relations.

### Unit-6: Sequence Models:

03 Hrs

RNN, LSTM, GRU models, Application to NLP, language models, machine translation, image captioning, video processing, visual question answering, video processing, learning from descriptions, Attention Mechanism, Attention over images

#### References:

Text Books:

1. S. N. Sivanandam & S.N.Deepa “Principles of Soft computing”, John Wiley & Sons,
2. S. Rajasekaran, G. A. Vijayalakshami, Neural Networks, Fuzzy Logic and Genetic Algorithms: Synthesis & Applications, PHI.
3. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.
4. David E. Goldberg., Genetic Algorithms: in Search and Optimization, PHI
5. Jyh: Shing Roger Jang, Chuen:Tsai Sun, Eiji Mizutani, Neuro:Fuzzy and Soft Computing, Prentice:Hall of India, 2003

#### Reference Books:

1. Timothy J. Ross, Fuzzy Logic with Engineering Applications (Wiley)
2. George J. Klir and Bo Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice Hall,
3. An Introduction to Genetic Algorithm Melanic Mitchell (MIT Press)
4. Evolutionary Algorithm for Solving Multi-objective, Optimization Problems (2nd Edition), Collelo, Lament, Veldhnizer ( Springer)
5. Neural Networks and Learning Machines Simon Haykin (PHI).
6. Neural Networks, Fuzzy logic, and Genetic Algorithms, S. Rajasekaran& G. A. V. Pai, PHI.

**Subject Code: PGDS206**

**Subject Name: Programming Lab-II**

**Teaching Scheme:**

Practical: 4 Hrs / week

**Credit**

4

**Examination Scheme:**

End Semester: 100

Assignments related to Machine Learning, Deep Learning and Visualization

#### Machine Learning---

1. Basics of Machine Learning with Python
2. Basic Package introduction: Numpy, Pandas, Seaborn, Matplotlib
4. Machine Learning using Simple Linear regression
4. Classification problem solving using Logistic Regression
5. Classification analysis using Decision tree Algorithm
6. Classification and regression analysis using Random forest algorithm
7. Classification and regression problem solving using Support Vector Machine algorithm 3
8. Classification problem solving using K-Nearest Neighbor algorithm
9. Unsupervised Learning using K-means Clustering
10. End to End Machine Learning Classification Project 1
11. End to End Machine Learning Regression Project 2

#### Natural Language Processing (Lab)

Objective: The main objective of this laboratory is to write programs that manipulate and analyze language data using Python.

Python Packages Students are expected to know/ learn the following Python NLP packages

- NLTK ( [www.nltk.org/](http://www.nltk.org/) (<http://www.nltk.org/>))
- Spacy ( <https://spacy.io/> )

- TextBlob ( <http://textblob.readthedocs.io/en/dev/>)
- Gensim(<https://pypi.python.org/pypi/gensim>)
- Pattern (<https://pypi.python.org/pypi/Pattern>)

Datasets: 1. NLTK includes a small selection of texts from the Project Gutenberg electronic text archive, which contains some 25,000 free electronic books, hosted at <http://www.gutenberg.org/>.

2. The Brown Corpus contains text from 500 sources, and the sources have been categorized by genre, such as news, editorial, and so on (<http://icame.uib.no/brown/bcmlos.html>).

3. Wikipedia Articles Or any other dataset of your choice Reference: Jacob Perkins. Python 3 Text Processing with NLTK 3 Cookbook. Packt Publishing. 2014

Exercises:

1. Text segmentation: Segment a text into linguistically meaningful units, such as paragraphs, sentences, or words. Write programs to segment text (in different formats) into tokens (words and word-like units) using regular expressions. Compare an automatic tokenization with a gold standard

2. Part-of-speech tagging: Label words (tokens) with parts of speech such as noun, adjective, and verb using a variety of tagging methods, e.g., default tagger, regular expression tagger, unigram tagger, and n-gram taggers.

3. Text classification: Categorize text documents into predefined classes using Naïve Bayes Classifier and the Perceptron model

4. Chunk extraction, or partial parsing: Extract short phrases from a part-of-speech tagged sentence. This is different from full parsing in that we're interested in standalone chunks, or phrases, instead of full parse trees

5. Parsing: parsing specific kinds of data, focusing primarily on dates, times, and HTML. Make use of the following preprocessing libraries: □ dateutil which provides datetime parsing and timezone conversion □ lxml and BeautifulSoup which can parse, clean, and convert HTML □ charade and UnicodeDammit which can detect and convert text character encoding

6. Sentiment Analysis: Using Libraries TextBlob and nltk, give the sentiment of a document

## **Domain : Social Media**

"Daily" Twitter Data Analysis for a Product

As more and more people are expressing their views and opinions on various microblogging websites about various products and services. There has been a surge of data generated by the users, these websites have people sharing their thoughts daily.

Sentiment Analysis with the help of Natural Language Processing technique for identifying the sentiments of a product or service

## **Domain : E commerce**

Natural Language Processing

Customers are looking for more information before buying a product on E-commerce websites. Amazon introduced a new feature 'question and answer' search field for products.

The project is to build information retrieval system from Amazon products data based on NLP techniques. Top 5 relevant answers to be retrieved based on input question

## **Domain : Banking**

Predicting Loan defaulters

Reducing the risk of fraudulent loans by carefully evaluating the risk & at the same time increasing profits by rejecting only those loans, which have the potential of defaulting

## **Domain : FMCG**

## Warranty Cost prediction

The objective of the analysis to predict an item when sold, what is the probability that customer would file for warranty and to understand important factors associated with them

### **Domain : Aviation**

Predict flight delays

Predict which flights would be delayed and by how long?

Flight delays costs the industry an estimated \$25 billion every year More than 60 percent of frequent flyers cite delays among the things about air travel that they find most dismaying. And the costs are spread around — an extra \$25 in parking here, a missed business meeting there. Carriers, meanwhile, pay an estimated \$62 per minute in crew, fuel, maintenance and other costs. It adds up.

### **Lab: Data Analytics and Visualization**

#### **Lab 01: Overview**

Exercise 1: Getting Started

Task 1: Record Your Account Details

Task 2: Sign in to the Power Bi Service

Task 3: Create Work space

Task 4: Open Power BI Desktop

Task 5: Update the lab database

#### **Lab02: Preparing data** in Power BI Desktop

Exercise1: Prepare Data

Task 1: Save the Power BI Desktop file from a source

Task 2: Sey Power BI Desktop options from a source

Task 3: Get the Data from SQL Server from a source

Task 4: Preview SQL Server queries from a source

Task 5: Get data from CSV file from a source

Task 6: Get data from Web as a source

#### **LAB 03: Loading** Data in Power BI desktop

Overview

Exercise 03: Load Data

Task 1: Configure the Salesperson Query

Task 2: Configure SalespersonRegion query

Task 3: Configure the Product query

Task 4: Configure the Reseller query

Task 5: Configure the Region query

Task 6: Congigure the Sales query

Task 7: Configure the Target query

Finish Up.

#### **Lab04: Data Modeling** in Power BI Desktop

Overview

Exercise: Create Model Relationships, Configure Tables, Review Model Interface, Create Quick Measures

Task 1: Create Model Relationships

Task 2: Configure Tables

2.1 Configure the Product Table

2.2 Configure the Region Table

2.3 Configure the Reseller Table

2.4 Configure the Sales table

## 2.5 Bulk update properties

Task 3: Review the Model interface

Task 4: Create Quick Measures

Finish up

### **Lab05: Advanced Data Modelling** in Power BI Desktop

Overview

Exercise: Create a Many to Many Relationship

Task 1: Create Many to Many Relationship

Task 2: Relate the Targets table

Task 3: Enforce Row Level Security

Finish up.

### **Lab06: Using DAX** in Power BI Desktop

Overview

Exercise 06: Create Calculated Tables

Task 1: Create the Salesperson table

Task 2: Create the Date Table

Task 3: Create calculated Columns

Task 4: Complete the Date table

Task 5: Mark the Date table

### **Lab 07: Measures**

Exercise 06.2: Create Measures

Task1: Create Simple Measures

Task2: Create Additional Measures

Finish up.

### **Lab 08: Publish** the Power BI Desktop

Overview

Exercise 1: Publish the File.

Finish up.

### **Lab 09 : Designing a Report** in Power BI Desktop

Overview

Exercise 7.1: Create Report

Task 1: Create a new File

Task 2: Create a live Connection

Task 3: Design page 1

Task 4: Design page 2

Task 5: Design page 3

Task 6: Publish the report.

Finish up.

### **Lab10: Creating a Power BI Dashboard with R Scripting**

Overview

Exercise: Create a Dashboard, Refresh Data Set, Review the Dashboard.

Task 1: Create a Dashboard

Task 2: Edit tile details

Task 3: Configure an alert

Task 4: Update lab database, Refresh Power BI Desktop file.

### **Case Studies:**

1. Image classifier for identifying cat vs dogs using CNN
2. Image classifier for identifying cat vs dogs using CNN

**Subject Code: PGDS207**

**Subject Name: Seminar-II**

**Teaching Scheme:**

**Credit**

2

**Examination Scheme:**

Presentation & Report Submission

Seminar related to any of their current subjects

**Subject Code: PGDS301**

**Subject Name: Capstone Project**

**Teaching Scheme:**

**Credit**

10

**Examination Scheme:  
Viva & Report Submission**

The Data Science capstone project focuses on establishing a strong hold of analyzing a problem and coming up with solutions based on insights from the data analysis perspective. The capstone project will help you master the following verticals:

**Project Steps:**

- Data Processing - In this step, you will apply various data processing techniques to make raw data meaningful.
- Model Building - You will leverage techniques such as regression and decision trees to build Machine Learning models that enable accurate and intelligent predictions. You may explore Python, R, or SAS to develop your model. You will follow the complete model-building exercise from data split to test and validate data using the k-fold cross-validation process.
- Model Fine-tuning - You will apply various techniques to improve the accuracy of your model and select the champion model that provides the best accuracy.
- Dash boarding and Representing Results - As the final step, you will be required to export your results into a dashboard with meaningful insights using Tableau.
- Assessment and monitoring of the model created using the machine learning models.

**Subject Code: PGDS302**

**Subject Name: Massive Open Online Courses (MOOCs)**

**Teaching Scheme:**

**Credit**

4

**Examination Scheme:  
Certificate Submission**

1. Introduction to Machine Learning: <https://nptel.ac.in/courses/106/106/106106139/>
2. Machine Learning: <https://nptel.ac.in/courses/106/106/106106202/>
3. Machine Learning for Science and Engineering applications: <https://nptel.ac.in/courses/106/106/106106198/>
4. Introduction to Machine Learning: <https://nptel.ac.in/courses/106/105/106105152/>
5. Deep Learning (Part-I): <https://nptel.ac.in/courses/106/106/106106184/>
6. Deep Learning: [https://onlinecourses.nptel.ac.in/noc19\\_cs54/preview](https://onlinecourses.nptel.ac.in/noc19_cs54/preview)
7. Naive Bayes from Scratch: <https://courses.analyticsvidhya.com/courses/naive-bayes>
8. Getting Started with Neural Networks: <https://courses.analyticsvidhya.com/courses/getting-started-with-neural-networks>
9. Machine Learning – Offered by Stanford Online - <https://www.coursera.org/learn/machine-learning>
10. Microsoft Exam DA-100: Analyzing Data with Microsoft Power BI
11. Microsoft Exam PL-300: Microsoft Power BI Data Analyst.