Diploma Engineering

Laboratory Manual

(Fundamental of Machine Learning) (DI04016031)

Information Technology Semester 4th

Enrolment No	
Name	
Branch	Information Technology
Academic Term	22/12/2025 to 29/04/2026
Institute	Government Polytechnic, Gandhinagar

Directorate of Technical Education
Gandhinagar - Gujarat

Course Coordinator: Ms. Hiral R. Patel

DTE's Vision:

To facilitate quality technical and professional education having relevance for both industry and society, with moral and ethical values, giving equal opportunity and access, aiming to prepare globally competent technocrats.

DTE's Mission:

1. Quality technical and professional education with continuous improvement of all the resources and personnel

2. To promote conducive ecosystem for Academic, Industry, Research, Innovations and Startups

3. To provide affordable quality professional education with moral values, equal opportunities, accessibility and accountability

4. To allocate competent and dedicated human resources and infrastructure to the institutions for providing world-class professional education to become a Global Leader ("Vishwa Guru")

Institute's Vision:

Develop technically proficient and ethically sound diploma engineers contributing to industry and society needs.

Institute's Mission:

- To impart quality technical education
- To shape students towards sensitizing in ethical values and contributing in nature
- To familiarize students with world of work

Department's Vision:

The department will strive to develop technically proficient and ethically sound diploma IT engineers to serve the industry and society.

Department's Mission:

M1: To impart IT engineering education through effective teaching learning process.
M2: Making our students contributory and believing in ethical values.
M3: Making students aware and practice current technology in industry.

<u>Certificate</u>

This is to certify that Mr./Ms ………………………………………………………………….
Enrolment No. ………….……………. of 4th Semester of *Diploma in Information Technology* of Government Polytechnic, Gandhinagar (623) has satisfactorily completed the term work in course Fundamental of Machine Learning (DI04016031) for the Academic Year: 2025-26 Term: 252 prescribed in the GTU curriculum.

Place: Gandhinagar

Date: /04/2026

Signature of Course Faculty   **Head of Department**

Course Coordinator: Ms. Hiral R. Patel

Programme Outcomes (POs):

1. **Basic and Discipline specific knowledge:** Apply knowledge of basic mathematics, science and engineering fundamentals and engineering specialization to solve the *engineering* problems.

2. **Problem analysis**: Identify and analyse well-defined *engineering* problems using codified standard methods.

3. **Design/ development of solutions:** Design solutions for *engineering* well-defined technical problems and assist with the design of systems components or processes to meet specified needs.

4. **Engineering Tools, Experimentation and Testing:** Apply modern *engineering* tools and appropriate technique to conduct standard tests and measurements.

5. **Engineering practices for society, sustainability and environment:** Apply appropriate technology in context of society, sustainability, environment and ethical practices.

6. **Project Management:** Use engineering management principles individually, as a team member or a leader to manage projects and effectively communicate about well-defined engineering activities.

7. **Life-long learning:** Ability to analyse individual needs and engage in updating in the context of technological changes *in field of engineering.*

Safety and necessary Precautions followed

- ✓ Read the experiment thoroughly before starting and ensure that you understand all the steps and concepts involved from underpinning theory.
- ✓ Keep the workspace clean and organized, free from clutter and unnecessary materials.
- ✓ Use the software according to its intended purpose and instructions.
- ✓ Ensure that all the necessary equipment and software are in good working condition.
- ✓ Never eat or drink in the lab, as it can cause contamination and create safety hazards.
- ✓ If any accidents or injuries occur, immediately notify the instructor and seek medical attention if necessary.

Course Coordinator: Ms. Hiral R. Patel

Practical Outcome - Course Outcome matrix

Course Outcomes (COs):
a) CO1: -To understand the need of machine learning for various problem solving.
b) CO2: - Prepare machine learning models by performing appropriate data pre-processing, and apply evaluation methods to assess model performance.
c) CO3: - Evaluate various supervised learning algorithms using appropriate dataset.
d) CO4: -Evaluate various unsupervised learning algorithms using appropriate dataset.
e) CO5:- Understand and apply various existing Python libraries for data preprocessing, visualization, and machine learning tasks.

S. No.	Practical Outcome/Title of experiment	CO1	CO2	CO3	CO4	CO5
1.	Numerical Computing with Python (NumPy, Matplotlib). Perform various operation mentioned in unit 6 corresponding to Numpy and Matplotlib					✓
2.	Introduction to Pandas for data import and export (Excel, CSV etc.) Perform various operation mentioned in unit 6 corresponding to Pandas					✓
3.	Basic Introduction to Scikit learn. Perform various operation mentioned in unit 6 corresponding to Scikitlearn)					✓
4.	Introduction to Machine Learning – Predicting Favourite Fruit. Understand how machines "learn" from examples, similar to human learning by Exploring a simple dataset and predict outcomes based on input features. Write Python code that takes age group and Gender as input and provide favourite fruit as output. Generate rule through python code.	✓				
5.	To understand different types of Machine Learning activities and learn how to handle and preprocess datasets for ML applications.		✓			
6.	To simulate a Machine Learning prediction scenario in an interactive game without training model explicitly. Build a confusion matrix from students' decisions, and calculate performance metrics. Students play a "Celebrity Guessing Game". A model predicts whether a person in a photo is a Famous Celebrity (Yes) or Not a Celebrity (No).		✓			

Course Coordinator: Ms. Hiral R. Patel

7.	Write a program to use a K-nearest neighbour it to predict class labels of test data. Training and test data must be provided explicitly			✓		
8.	Use the following dataset as music.csv			✓		

age	gender	genre
20	1	HipHop
23	1	HipHop
25	1	HipHop
26	1	Jazz
29	1	Jazz
30	1	Jazz
31	1	Classical
33	1	Classical
37	1	Classical
20	0	Dance
21	0	Dance
25	0	Dance
26	0	Acoustic
27	0	Acoustic
30	0	Acoustic
31	0	Classical
34	0	Classical
35	0	Classical

a. Store file as music.csv and import it to python using pandas
b. Prepare the data by splitting data in input(age ,gender) and output(genre) data set
c. Use SVM model from sklearn to predict the genre of various age group people. (Ex A male of age 21 likes hiphop whereas female of age 22 likes dance)
d. Calculate the accuracy of the model prepared.
e. Generate synthetic dataset and verify accuracy over larger dataset.

9.	Import vgsales.csv from kaggle platform. a. Find rows and columns in dataset b. Find basic information regarding dataset using describe command. c. Find values using values command. d. Apply logistic regression for predicting house price.			✓		
10.	Project on regression a. Import home_data.csv on kaggle using pandas			✓		

Course Coordinator: Ms. Hiral R. Patel

	b. Understand data by running head ,info and describe command c. Plot the price of house with respect to area using matplotlib library d. Apply linear regression model to predict the price of house e. Highlight the correctness of Plot using suitable Metrics					
11.	Write a program to cluster a set of points using K-means. Training and test data must be provided explicitly.				✓	

Industry Relevant Skills

The following industry relevant skills are expected to be developed in the students by performance of experiments of this course.

a) *Student will learn to automate variety of task making system more efficient and cost effective*
b) *Student will learn efficient handling of data that will cater to better data analytics*
c) *Student will lean to implement machine learning approaches to varied field of applications from healthcare to e-commerce.*

Instructions for Students

1. Organize the work in the group and make record of all observations.
2. Students shall develop maintenance skill as expected by industries.
3. Student shall attempt to develop related hand-on skills and build confidence.
4. Student shall develop the habits of evolving more ideas, innovations, skills etc.
5. Student shall refer technical magazines and data books.
6. Student should develop habit to submit the practical on date and time.
7. Student should well prepare while submitting write-up of exercise.

Course Coordinator: Ms. Hiral R. Patel

Continuous Assessment Sheet

Name: **Enrolment No:**   **Term:** 22/12/2025 to 29/04/2026

Sr No.	Practical Outcome/Title of experiment	Page	Date	Marks (20)	Sign
1	Numerical Computing with Python (NumPy, Matplotlib). Perform various operation mentioned in unit 6 corresponding to Numpy and Matplotlib				
2	Introduction to Pandas for data import and export (Excel, CSV etc.) Perform various operation mentioned in unit 6 corresponding to Pandas				
3	Basic Introduction to Scikit learn. Perform various operation mentioned in unit 6 corresponding to Scikitlearn)				
4	Introduction to Machine Learning – Predicting Favorite Fruit. Understand how machines "learn" from examples, similar to human learning by Exploring a simple dataset and predict outcomes based on input features. Write Python code that takes age group and Gender as input and provide favorite fruit as output. Generate rule through python code. <table><tr><td>Age Group</td><td>Gender</td><td>favorite Fruit</td></tr><tr><td>10-15</td><td>Male</td><td>Apple</td></tr><tr><td>10-15</td><td>Female</td><td>Banana</td></tr><tr><td>16-20</td><td>Male</td><td>Orange</td></tr><tr><td>16-20</td><td>Female</td><td>Mango</td></tr><tr><td>21-25</td><td>Male</td><td>Banana</td></tr><tr><td>21-25</td><td>Female</td><td>Mango</td></tr></table>				
5	To understand different types of Machine Learning activities and learn how to handle and preprocess datasets for ML applications. Consider following dataset: Student.csv				

Course Coordinator: Ms. Hiral R. Patel

Name	Age	Gender	Hours Studied	Marks	Passed
Alice	20	F	5	80	Yes
Bob	21	M	3	50	No
Charlie	22	M	4	60	Yes
Diana	20	F	2	40	No
Eve	23	F	6	90	Yes

1. Load data set using read_CSV
2. Explore data (data type and structure) check data type, missing value and display basic statistics
3. Handle missing values using fillna
4. Encode categorical data. Covert gender to numeric
5. Perform Feature Subset selection

| 6 | To simulate a Machine Learning prediction scenario in an interactive game without training model explicitly. Build a confusion matrix from students' decisions, and calculate performance metrics. Students play a "Celebrity Guessing Game". A model predicts whether a person in a photo is a Famous Celebrity (Yes) or Not a Celebrity (No).

Students are given 10–15 celebrities' photos with actual labels hidden. Students must Simulate the model by making predictions (Yes/No). After revealing the true labels, students assign each prediction to TP, FP, TN, and FN.

Input confusion matrix values in a python code and evaluate Accuracy, Kappa Value, Sensitivity, Specificity, Error Rate, Precision, Recall, and F-measure. | | | | |

| 7 | Write a program to use a K-nearest neighbor it to predict class labels of test data. Training and test data must be provided explicitly | | | | |

8	ML Project Use the following dataset as music.csv 	age	gender	genre
20	1	HipHop
23	1	HipHop
25	1	HipHop
26	1	Jazz
29	1	Jazz
30	1	Jazz
31	1	Classical
33	1	Classical
37	1	Classical
20	0	Dance
21	0	Dance
25	0	Dance
26	0	Acoustic
27	0	Acoustic
30	0	Acoustic
31	0	Classical
34	0	Classical
35	0	Classical | a. Store file as music.csv and import it to python using pandas b. Prepare the data by splitting data in input (age ,gender) and output(genre) data set c. Use SVM model from sklearn to predict the genre of various age group people.(Ex A male of age 21 likes hiphop whereas female of age 22 likes dance) d. Calculate the accuracy of the model prepared. e. Generate synthetic dataset and verify accuracy over larger dataset.				
9	Import vgsales.csv from kaggle platform. a. Find rows and columns in dataset b. Find basic information regarding dataset using describe command. c. Find values using values command. d. Apply logistic regression for predicting house price.				
10	Project on regression a. Import home_data.csv on kaggle using pandas b. Understand data by running head, info and describe command c. Plot the price of house with respect to area using matplotlib library d. Apply linear regression model to predict the price of house e. Highlight the correctness of Plot using suitable Metrics				
11	Write a program to cluster a set of points using K-means. Training and test data must be provided explicitly.				

Course Coordinator: Ms. Hiral R. Patel

Assessment-Rubrics

Criteria	Total Marks	Exceptional (4 Marks)	Satisfactory (3 Marks)	Developing (2 Marks)	Limited (1 Mark)
Engagement	4	Performed practical him/herself	Performed practical with others help	Watched other students performing practical but not tried him/herself	Present in practical session but not attentively participated in performance
Accuracy	4	Accurately done	1-2 errors/mistakes found	3-5 errors/mistakes identified	More than 5 errors/mistakes found
Documentation	4	No errors, Program is well Executed and Documented Properly	Complete write-up and output tables but presentation is poor	Some of the commands missing with missing outputs	Poor write-up and diagram or missing content
Understanding & Explanation	4	Fully understood the performance & can explain perfectly	Understood the performance but can't explain	Partially understood the performance & can give little explanation	Partially understood and can't give explanation
Time	4	Completed the work within 1st week	Work is submitted later than 1st week but by the end of 2nd week	Work done after 2nd week but before the end of 3rd week	Work submitted after 3rd week time

Course Coordinator: Ms. Hiral R. Patel

Practical No. 1: Numerical Computing with Python (NumPy, Matplotlib). Perform various operation mentioned in unit 6 corresponding to Numpy and Matplotlib

A.   **Objective:** Getting familiarized with python libraries related to visualization and computation.

B.   **Expected Program Outcomes (POs):-** PO1, PO2, PO3, PO4, PO7.

C.   **Expected Skills to be developed based on competency:**

    I.   To understand the use of well know python libraries.
    II.   To visualize data and implement logics based on data..

D.   **Expected Course Outcomes(Cos)**

    CO -5

E.   **Practical Outcome(PRo)**

    Store and represent data using python libraries.

F.   **Expected Affective domain Outcome(ADos)**

    Handle tools /components/equipment carefully with safety and necessary precaution.

    In software ethics Environment sustainability and environment consciousness whenever suitable.

G.   **Prerequisite Theory:**

    Refer Unit 1 of course curriculum. Also explore the link following link

    https://numpy.org/doc/stable/

    https://matplotlib.org/stable/tutorials/index

H.   **Resources/Equipment Required**

Sr.No.	Instrument/Equipment /Components/Trainer kit	Specification	Quantity
1	*System supporting Jupyter Notebook*	*Python 3.x*	*1*

I.   **Procedure to be followed/Source code (CE & IT software subjects):**
Student must use the space for writing source code. Understand and re-implement different methods for handling data. (Exhaustive use of functions must be done)

                Course Coordinator: Ms. Hiral R. Patel

J.   Practical related Quiz.

1. Which of the function is a function to create a numpy array?
a) empty()
b) array()
c) ones()
d) All the above

2. What is the output of the below code?
import numpy as np
arr = np.array([1, 2, 3, 4, 5, 6, 7])
result = arr + 1
print(result)
a) array([2, 3, 4, 5, 6, 7])
b) array([3, 4, 5, 6, 7])
c) array([2, 3, 4, 5, 6, 7, 8])
d) array([3, 4, 5, 6, 7, 8])

3. Find the output of the below code
a = np.array([[[1,2,3],[4,5,6]]])
a) 1
b) (1,3)
c) 3
d) (3,1)

4. By default, Plot() function plots a?
a) Bar chart
b) Line chart
c) Pie chart
d) Horizontal bar chart

5. Which of the following type of chart is not supported by pyplot?
a) Pie
b) Boxplot
c) Histogram
d) All of the above

6. To create histogram pyplot provides?
a) hist()
b) histo()
c) histg()
d) histogram()

Course Coordinator: Ms. Hiral R. Patel

K.   References / Suggestions

Numpy

https://www.youtube.com/watch?v=Rbh1rieb3zc

Matplotlib

https://www.youtube.com/watch?v=yZTBMMdPOww

L.   Assessment Rubrics

Criteria	Engagement (4)	Accuracy (4)	Documentation (4)	Understanding &Explanation (4)	Time (4)	Total Marks (20)
Marks						

Faculty Signature

Course Coordinator: Ms. Hiral R. Patel

Practical No. 2: Introduction to Pandas for data import and export (Excel, CSV etc.)
Perform various operation mentioned in unit 6 corresponding to Pandas

Objective: Getting familiarized with python machine learning libraries.

A.   Expected Program Outcomes (POs):-PO1, PO2, PO3, PO4, PO7.

B.   Expected Skills to be developed based on competency:

   I.   To understand the use of well know python machine learning libraries.
   II.   Using machine learning methods in python libraries.

C.   Expected Course Outcomes(Cos)

CO-5

D.   Practical Outcome(PRo)

 Using machine learning methods implemented in Pandas library.

E.   Expected Affective domain Outcome(ADos)

Handle tools /components/equipment carefully with safety and necessary precaution.

In software ethics Environment sustainability and environment consciousness whenever suitable.

F.   Prerequisite Theory:

Refer Unit 6 of course curriculum. Also explore the link following link

 https://pandas.pydata.org/docs/user_guide/index.html

G.   Resources/Equipment Required

Sr.No.	Instrument/Equipment /Components/Trainer kit	Specification	Quantity
1	System supporting Jupyter Notebook	Python 3.x	1

H.   Procedure to be followed/Source code (CE & IT software subjects):

Student must use the space for writing source code. Understand and re-implement different methods for handling data.(Exhaustive use of functions must be done)

I. **Practical related Quiz.**

1. Which of the following feature is not provided by the Pandas module?

   a) Merge and join the data sets

   b) Filter data using the condition

   c) Plot and visualize the data

   d) None of the above

2. From which of the following files, pandas can read data?

   a) JSON

   b) Excel

   c) HTML

   d) All the above

3. Given a dataset named 'data' containing the 5 columns and 10 rows, find the output of the below code?

   print(len(data.columns))

   a) 5

   b) 10

   c) 15

   d) 50

4. What does the attribute shape return?

   a) It returns the number of rows and columns respectively in the form of a tuple

   b) It returns the number of columns and rows respectively in the form of a list

   c) It returns the number of rows and columns respectively in the form of a list

   d) It returns the number of columns and rows respectively in the form of a tuple

5. Which of the following commands return the data type of the values in each column in the data frame

   a) print(df.dtype)

   b) print(dtypes(df))

c) print(df.dtypes)

d) None of the above

J. **References / Suggestions (lab manual designer should give)**

https://www.youtube.com/watch?v=RhEjmHeDNoA

K. **Assessment Rubrics**

Criteria	Engagement (4)	Accuracy (4)	Documentation (4)	Understanding &Explanation (4)	Time (4)	Total Marks (20)
Marks						

Faculty Signature

Course Coordinator: Ms. Hiral R. Patel

Practical No. 3: Basic Introduction to Scikit learn. Perform various operation mentioned in unit 6 corresponding to Scikitlearn)

 Objective: Getting familiarized with python machine learning libraries.

A. **Expected Program Outcomes (POs):-**PO1, PO2, PO3, PO4, PO7.

B. **Expected Skills to be developed based on competency:**

    I.   To understand the use of well know python machine learning libraries.
    II.   Using machine learning methods in python libraries.

C. **Expected Course Outcomes(Cos)**

CO-5

D. **Practical Outcome(PRo)**

 Using machine learning methods implemented in Scikit library.

E. **Expected Affective domain Outcome(ADos)**

Handle tools /components/equipment carefully with safety and necessary precaution.

In software ethics Environment sustainability and environment consciousness whenever suitable.

F. **Prerequisite Theory:**

Refer Unit 6 of course curriculum. Also explore the link following link

https://scikit-learn.org/stable/user_guide.html

G. **Resources/Equipment Required**

Sr.No.	Instrument/Equipment /Components/Trainer kit	Specification	Quantity
1	*System supporting Jupyter Notebook*	*Python 3.x*	*1*

H. **Procedure to be followed/Source code (CE & IT software subjects):**

Student must use the space for writing source code. Understand and re-implement any 2 examples provided in scikit learn portal

I.   Practical related Quiz.

1. Why do we need two sets: a train set and a test set?

   a) To train the model faster

   b) To validate the model on unseen data

   c) To improve the accuracy of the model

2. Cross-validation allows us to:

   a) train the model faster

   b) measure the generalization performance of the model

   c) reach better generalization performance

   d) estimate the variability of the generalization score

3. How is a tabular dataset organized?

   a) a column represents a sample and a row represents a feature

   b) a column represents a feature and a row represents a sample

   c) the target variable is represented by a row

   d) the target variable is represented by a column

4. A categorical variable is:

   a) a variable with only two different possible values

   b) a variable with continuous numerical values

   c) a variable with a finite set of possible values

J.   References / Suggestions

https://www.youtube.com/watch?v=OobqWEUrVKw

K.   Assessment Rubrics

Criteria	Engagement (4)	Accuracy (4)	Documentation (4)	Understanding &Explanation (4)	Time (4)	Total Marks (20)
Marks						

Faculty Signature

Course Coordinator: Ms. Hiral R. Patel

Practical No. 4: Introduction to Machine Learning – Predicting Favourite Fruit. Understand how machines "learn" from examples, similar to human learning by exploring a simple dataset and predict outcomes based on input features. Write Python code that takes age group and Gender as input and provide favourite fruit as output. Generate rule through python code.

A.   **Objective:** To understand the concept of learning

B.   **Expected Program Outcomes (POs):-**PO1, PO2, PO3, PO6, PO7.

C.   **Expected Skills to be developed based on competency:**

     I.   Understand and generate inference rules.
    II.   Implementing a procedure.
    III.   Testing use cases over implementing procedures.

D.   **Expected Course Outcomes(Cos)**

CO-1

E.   **Practical Outcome(PRo)**

Generating Inference rule for prediction.

F.   **Expected Affective domain Outcome(ADos)**

Handle tools /components/equipment carefully with safety and necessary precaution.

In software ethics Environment sustainability and environment consciousness whenever suitable.

G.   **Prerequisite Theory:**

Refer Unit 1 of course curriculum. Students are suggested to read chapter 1 of Machine Learning authored by TOM. M. Mitchell

Dataset

Age Group	Gender	favorite Fruit
10-15	Male	Apple
10-15	Female	Banana
16-20	Male	Orange
16-20	Female	Mango
21-25	Male	Banana
21-25	Female	Mango

H.   **Resources/Equipment Required**

          Course Coordinator: Ms. Hiral R. Patel

Sr.No.	Instrument/Equipment /Components/Trainer kit	Specification	Quantity
1	*System supporting Jupyter Notebook*	*Python 3.x*	*1*

I. **Procedure to be followed/Source code (CE & IT software subjects):**

Course Coordinator: Ms. Hiral R. Patel

J. **Observations and Calculations/Input-Output (CE & IT software subjects):**

Write inference rules here.

K.   Practical related Quiz.
Multiple choice questions

1. What is the input feature in the favourite fruit prediction problem?

   a) Fruit

   b) Age group and Gender

   c) Prediction accuracy

   d) Output label

2. What does the training data represent?

   a) Random guesses

   b) Past examples

   c) Test cases

   d) Errors

3. If the input is Age Group = 21–25 and Gender = Female, what is the predicted fruit?

   a) Apple

   b) Banana

   c) Orange

   d) Mango

L.   References / Suggestions

Tom M. Mitchell,

Machine Learning, McGraw-Hill Education, 1997.

→ Classic reference defining machine learning as learning from experience.

M.   Assessment Rubrics

Criteria	Engagement (4)	Accuracy (4)	Documentation (4)	Understanding &Explanation (4)	Time (4)	Total Marks (20)
Marks						

Faculty Signature

Course Coordinator: Ms. Hiral R. Patel

Practical No. 5: To understand different types of Machine Learning activities and learn how to handle and pre-process datasets for ML applications. Using the student.csv perform following operations

1. Load data set using read_CSV
2. Explore data (data type and structure)check data type ,missing value and display basic statistics
3. Handle missing values using fillna
4. Encode categorical data .Covert gender to numeric
5. Perform Feature Subset selection

A.   **Objective:** The primary objective of the practical is to understand data pre-processing along with identifying various types of data.

B.   **Expected Program Outcomes (POs):-** PO1, PO2, PO3,PO6, PO7

C.   **Expected Skills to be developed based on competency:**

      I.   Importing existing datasets from data repositories.
     II.   Applying feature selection algorithms on imported data
    III.   Selecting features based on the evaluation parameters

D.   **Expected Course Outcomes(Cos)**

     CO-2

E.   **Practical Outcome(PRo)**

     Performing feature Subset selection.

F.   **Expected Affective domain Outcome(ADos)**

     Handle tools /components/equipment carefully with safety and necessary precaution. In software ethics Environment sustainability and environment consciousness whenever suitable.

G.   **Prerequisite Theory:**

     Refer Unit 2 of course curriculum. Students are suggested to read chapter 2 of Machine Learning authored by Dutt, Chandramouli and das

Name	Age	Gender	Hours Studied	Marks	Passed
Alice	20	F	5	80	Yes
Bob	21	M	3	50	No
Charlie	22	M	4	60	Yes
Diana	20	F	2	40	No
Eve	23	F	6	90	Yes

H. **Resources/Equipment Required**

Sr.No.	Instrument/Equipment /Components/Trainer kit	Specification	Quantity
1	*System supporting Jupyter Notebook*	*Python 3.x*	*1*

I. **Procedure to be followed/Source code (CE & IT software subjects):**

Course Coordinator: Ms. Hiral R. Patel

J. **Observations and Calculations/Input-Output (CE & IT software subjects):**

Observation Table: Prepare a accuracy table by varying training and test data.

Course Coordinator: Ms. Hiral R. Patel

K.   Practical related Quiz.

1. Which pandas function shows the data types of all columns?

a) info()
b) describe()
c) shape()
d) head()

2. Which method is used to check missing values in a dataset?
a) isnull()
b) missing()
c) empty()
d) dropna()

3. What does describe() function return?
a) Column names
b) Missing values
c) Summary statistics
d) Data types

4. Which function is used to fill missing values?
a) replace()
b) fillna()
c) update()
d) dropna()

L.   References / Suggestions (lab manual designer should give)

Diabetes Prediction using Machine Learning from Kaggle-
https://www.youtube.com/watch?v=HTN6rccMu1k

https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

M.   Assessment Rubrics

Criteria	Engagement (4)	Accuracy (4)	Documentation (4)	Understanding &Explanation (4)	Time (4)	Total Marks (20)
Marks						

Faculty Signature

Course Coordinator: Ms. Hiral R. Patel

Practical No. 6: To simulate a Machine Learning prediction scenario in an interactive game without training model explicitly. Build a confusion matrix from students' decisions, and calculate performance metrics. Students play a "Celebrity Guessing Game". A model predicts whether a person in a photo is a Famous Celebrity (Yes) or Not a Celebrity (No). Students are given 10–15 celebrities' photos with actual labels hidden. Students must simulate the model by making predictions (Yes/No). After revealing the true labels, students assign each prediction to TP, FP, TN, and FN. Input confusion matrix values in a python code and evaluate Accuracy, Kappa Value, Sensitivity, Specificity, Error Rate, Precision, Recall, and F-measure.

Objective: Learning to design confusion matrix and calculate metrices.

A.   **Expected Program Outcomes (POs):-**PO1, PO2, PO3, PO6, PO7

B.   **Expected Skills to be developed based on competency:**

    I.   Draw the confusion matrix.
   II.   Identifying True Positive ,True Negative, False Positive, False Negative
  III.   Compute metrices like Accuray, Precision, Recall etc.

C.   **Expected Course Outcomes(Cos)**

CO -3

D.   **Practical Outcome(PRo)**

Generate Confusion Matrix and calculate metrices.

E.   **Expected Affective domain Outcome(ADos)**

Handle tools /components/equipment carefully with safety and necessary precaution.

In software ethics Environment sustainability and environment consciousness whenever suitable.

F.   **Prerequisite Theory:**

Refer Unit 3 of course curriculum. Students are suggested to read chapter 3 of Machine Learning authored by Dutt, Chandramouli and das

G.   **Resources/Equipment Required**

Sr.No.	Instrument/Equipment /Components/Trainer kit	Specification	Quantity
1	*System supporting Jupyter Notebook*	*Python 3.x*	*1*

H. **Procedure to be followed/Source code (CE & IT software subjects):**

I. **Observations and Calculations/Input-Output (CE & IT software subjects):**

Observation Table: Prepare a table of predicted class labels.

Course Coordinator: Ms. Hiral R. Patel

Calculations:

Calculate confusion matrix

J.   Practical related Quiz.
1. A confusion matrix is used to evaluate:
    a) Training time
    b) Model architecture
    c) Classification performance
    d) Feature selection

2. In a binary confusion matrix, True Positive (TP) means:
    a) Correctly predicted negative
    b) Incorrectly predicted positive
    c) Correctly predicted positive
    d) Incorrectly predicted negative

3. Which metric measures the overall correctness of a model?
    a) Precision
    b) Recall
    c) Accuracy
    d) Specificity
4. Precision is defined as:

a) TP / (TP + FN)

b) TP / (TP + FP)

c) TN / (TN + FP)

d) (TP + TN) / Total

5. Recall is also known as:

a) Accuracy

b) Specificity

c) Sensitivity

d) Kappa

K.   References / Suggestions

https://cse.iitkgp.ac.in/~pabitra/course/ml/ml.html

L.   Assessment Rubrics

Criteria	Engagement (4)	Accuracy (4)	Documentation (4)	Understanding &Explanation (4)	Time (4)	Total Marks (20)
Marks						

Faculty Signature

Practical No. 7: Write a program to use a K-nearest neighbor it to predict class labels of test data. Training and test data must be provided explicitly

A. **Objective:** To study and implement the K-Nearest Neighbour (KNN) classification algorithm and use it to predict class labels of given test instances based on explicitly defined training data.

B. **Expected Program Outcomes (POs):-**PO1, PO2, PO3, PO6, PO7

C. **Expected Skills to be developed based on competency:**

    I.   Ability to understand and explain the working of the K-Nearest Neighbor (KNN) algorithm.
    II.   Ability to compute distances and apply majority voting for classification.
    III.   Ability to implement KNN using given training and test datasets in Python.
    IV.   Ability to analyze and interpret classification results for test data.

D. **Expected Course Outcomes(Cos)**

    CO -3

E. **Practical Outcome(PRo)**

    The student will be able to implement and apply the K-Nearest Neighbor (KNN) algorithm to classify test data using explicitly provided training samples, distance measures, and majority voting.

F. **Expected Affective domain Outcome(ADos)**

    Handle tools /components/equipment carefully with safety and necessary precaution.In software ethics Environment sustainability and environment consciousness whenever suitable.

G. **Prerequisite Theory:**

    Refer Unit 3 of course curriculum. Students are suggested to read chapter 3 of Machine Learning authored by Dutt, Chandramouli and das

H. **Resources/Equipment Required**

Sr.No.	Instrument/Equipment /Components/Trainer kit	Specification	Quantity
1	System supporting Jupyter Notebook	Python 3.x	1

I.   Procedure to be followed/Source code (CE & IT software subjects):

J.   Observations and Calculations/Input-Output (CE & IT software subjects):

Observation Table: Prepare a table of predicted class labels.

Calculations:

Calculate confusion matrix

K.   Practical related Quiz.

1. K-Nearest Neighbor (KNN) is a:
    - a) Unsupervised learning algorithm
    - b) Supervised learning algorithm
    - c) Reinforcement learning algorithm
    - d) Deep learning algorithm

2. KNN is also known as:
    - a) Model-based learning
    - b) Rule-based learning
    - c) Instance-based learning
    - d) Probabilistic learning

3. What does the parameter K represent in KNN?
    - a) Number of classes
    - b) Number of features
    - c) Number of nearest neighbors
    - d) Number of test samples

4. Which distance measure is most commonly used in KNN?
    - a) Manhattan distance
    - b) Cosine similarity
    - c) Euclidean distance
    - d) Hamming distance

5. What happens if the value of K is very small?
    - a) Underfitting
    - b) Overfitting
    - c) High bias
    - d) No effect

L.   References / Suggestions

https://scikit-learn.org/stable/modules/neighbors.html

M.   Assessment Rubrics

Criteria	Engagement (4)	Accuracy (4)	Documentation (4)	Understanding &Explanation (4)	Time (4)	Total Marks (20)
Marks						

Faculty Signature

Course Coordinator: Ms. Hiral R. Patel

Practical No. 8: ML Project: Use the following dataset as music.csv.

age	gender	genre
20	1	HipHop
23	1	HipHop
25	1	HipHop
26	1	Jazz
29	1	Jazz
30	1	Jazz
31	1	Classical
33	1	Classical
37	1	Classical
20	0	Dance
21	0	Dance
25	0	Dance
26	0	Acoustic
27	0	Acoustic
30	0	Acoustic
31	0	Classical
34	0	Classical
35	0	Classical

a. Store file as music.csv and import it to python using pandas

b. Prepare the data by splitting data in input(age ,gender) and output(genre) data set

c. Use SVM model from sklearn to predict the genre of various age group people.(Ex A male of age 21 likes hiphop whereas female of age 22 likes dance)

d. Calculate the accuracy of the model prepared.

e. Generate synthetic dataset and verify accuracy over larger dataset.

A.   **Objective:** Effectively use sklearn library to make predictions using decision tree.

B.   **Expected Program Outcomes (POs):-**PO1, PO2, PO3,PO6,PO7.

C.   **Expected Skills to be developed based on competency:**

   I.   Importing existing datasets from data repositories.
   II.   Doing Prediction of class labels
   III.   Splitting of attribute based on criteria
   IV.   Learning on how to work on machine learning project.

D.   **Expected Course Outcomes(Cos)**

   CO-2

E.   **Practical Outcome(PRo)**

   Determine accuracy of the classification model.

F.   **Expected Affective domain Outcome(ADos)**

   Handle tools /components/equipment carefully with safety and necessary precaution.

   In software ethics Environment sustainability and environment consciousness whenever suitable.

G.   **Prerequisite Theory:**

   Refer: https://scikit-learn.org/stable/

   Refer Unit 3 of course curriculum. Students are suggested to read chapter 3 of Machine Learning authored by Dutt, Chandramouli and das

H. **Resources/Equipment Required**

Sr.No.	Instrument/Equipment /Components/Trainer kit	Specification	Quantity
1	System supporting Jupyter Notebook	Python 3.x	1

I. **Procedure to be followed/Source code (CE & IT software subjects):**

Course Coordinator: Ms. Hiral R. Patel

J.   Observations and Calculations/Input-Output (CE & IT software subjects):

Observation Table: Prepare table genre suggestion accuracy by varying test and train size.

Calculations:

Calculate confusion matrix

K.   Practical related Quiz.

  1. What is the main objective of a Support Vector Machine?

   a) Minimize classification error only
   b) Maximize the margin between classes
   c) Minimize training time
   d) Reduce dimensionality

  2. In SVM, the data points closest to the decision boundary are called:

   a) Centroids
   b) Neurons
   c) Support vectors
   d) Clusters

  3. Which of the following best describes the margin in SVM?

   a) Distance between any two data points
   b) Distance between class means
   c) Distance between the decision boundary and nearest data points
   d) Distance between training and testing data

  4. Which kernel is used when data is linearly separable?

a) Polynomial

b) RBF

c) Sigmoid

d) Linear

5. What does the kernel function do in SVM?

a) Reduces noise

b) Converts labels to numbers

c) Maps data to higher-dimensional space

d) Removes outliers

L.   References / Suggestions

Python Machine Learning Tutorial (Data Science):-

https://www.youtube.com/watch?v=7eh4d6sabA0

M.   Assessment Rubrics

Criteria	Engagement (4)	Accuracy (4)	Documentation (4)	Understanding &Explanation (4)	Time (4)	Total Marks (20)
Marks						

Faculty Signature

Course Coordinator: Ms. Hiral R. Patel

Practical No. 9: Import vgsales.csv from kaggle platform.
      a. Find rows and columns in dataset
      b. Find basic information regarding dataset using describe command.
      C. Find values using values command.
      d. Apply logistic regression for predicting house price.

A.   **Objective:** understand the imported data from known repositories.

B.   **Expected Program Outcomes (POs):-**PO1, PO2, PO3,PO6, PO7.

C.   **Expected Skills to be developed based on competency:**

     I.   Importing existing datasets from data repositories.
    II.   Understanding the data imported.
    III.   Using pandas library
    IV.   Understanding Logistic regression

D.   **Expected Course Outcomes(Cos)**

     CO-3

E.   **Practical Outcome(PRo)**

     Identifying data attributes.

F.   **Expected Affective domain Outcome(ADos)**

     Handle tools /components/equipment carefully with safety and necessary precaution.

     In software ethics Environment sustainability and environment consciousness whenever suitable.

G.   **Prerequisite Theory:**

     https://pandas.pydata.org/docs/user_guide/index.html

     Refer unit 4 of course curriculum. Students are suggested to read chapter 7 of Machine Learning authored by Dutt, Chandramouli and das

H.   **Resources/Equipment Required**

Sr.No.	Instrument/Equipment /Components/Trainer kit	Specification	Quantity
1	*System supporting Jupyter Notebook*	*Python 3.x*	*1*

     Course Coordinator: Ms. Hiral R. Patel

I.   **Procedure to be followed/Source code (CE & IT software subjects):**

Course Coordinator: Ms. Hiral R. Patel

J.   Practical related Quiz.

1. What is Pandas used for?
   a) Data analysis and manipulation
   b) Web development
   c) Machine learning
   d) Image processing

2. What are the two main data structures in Pandas?
   a) Series and Data Frames
   b) Arrays and lists
   c) Dictionaries and tuples
   d) Matrices and vectors

3. How do you read a CSV file into a Pandas DataFrame?
   a) pd.read_csv('filename.csv')
   b) pd.read_excel('filename.csv')
   c) pd.read_table('filename.csv')
   d) pd.read_json('filename.csv')

4. How do you select a subset of rows and columns from a Pandas DataFrame?
   a) df.loc[row_index, column_index]
   b) df.iloc[row_index, column_index]
   c) df[row_index, column_index]
   d) df.select(row_index, column_index)

5. How do you group data in a Pandas DataFrame?
   a) df.groupby(column_name)
   b) df.group_by(column_name)
   c) df.sort_by(column_name)
   d) df.filter_by(column_name)

K.   References / Suggestions

https://www.youtube.com/watch?v=7eh4d6sabA0

L.   Assessment Rubrics

Criteria	Engagement (4)	Accuracy (4)	Documentation (4)	Understanding &Explanation (4)	Time (4)	Total Marks (20)
Marks						

Faculty Signature

Course Coordinator: Ms. Hiral R. Patel

Practical No. 10: Project on regression

      a. Import home_data.csv on kaggle using pandas
      b. Understand data by running head ,info and describe command
      c. Plot the price of house with respect to area using matplotlib library
      d. Apply linear regression model to predict the price of house
      e. Highlight the correctness of Plot using suitable Metrics

A.   **Objective:** Understand the linear model.

B.   **Expected Program Outcomes (POs):-**PO1, PO2, PO3, PO6, PO7.

C.   **Expected Skills to be developed based on competency:**

      I.   Importing existing datasets from data repositories.
      II.   Understanding the data imported.
     III.   Using sklearn library to implement linear model.

D.   **Expected Course Outcomes(Cos)**

    CO-3

E.   **Practical Outcome(PRo)**

    Predicting values using linear regresssion.

F.   **Expected Affective domain Outcome(ADos)**

    Handle tools /components/equipment carefully with safety and necessary precaution.

    In software ethics Environment sustainability and environment consciousness whenever suitable.

G.   **Prerequisite Theory:**

    https://scikit-learn.org/stable/

    Refer unit 4 of course curriculum. Students are suggested to read chapter 8 of Machine Learning authored by Dutt, Chandramouli and das

H.   **Resources/Equipment Required**

Sr.No.	Instrument/Equipment /Components/Trainer kit	Specification	Quantity
1	*System supporting Jupyter Notebook*	*Python 3.x*	*1*

I. **Procedure to be followed/Source code (CE & IT software subjects):**

J.   Practical related Quiz.

    1. What is linear regression used for?

    a) Data visualization

    b) Clustering

    c) Predictive modeling

    d) Dimensionality reduction

    2. In linear regression, what is the objective?

    a) To minimize the mean squared error between the predicted and actual values

    b) To maximize the correlation coefficient between the features and target variable

    c) To maximize the R-squared value between the features and target variable

    d) To minimize the sum of absolute errors between the predicted and actual values

    3. How is linear regression implemented in Scikit-Learn?

    a) By instantiating a LinearRegression object and calling its fit method

    b) By instantiating a Regression object and calling its fit method

    c) By instantiating a LinearModel object and calling its fit method

    d) By instantiating a LinearSolver object and calling its fit method

    4. What is the R-squared value in linear regression?

    a) A measure of how well the model fits the data

    b) A measure of the correlation between the features and target variable

    c) A measure of the variance in the target variable that can be explained by the features

    d) A measure of the error between the predicted and actual values

K.   References / Suggestions

https://www.youtube.com/watch?v=8jazNUpO3lQ

L.   Assessment Rubrics

Criteria	Engagement (4)	Accuracy (4)	Documentation (4)	Understanding &Explanation (4)	Time (4)	Total Marks (20)
Marks						

Faculty Signature

Course Coordinator: Ms. Hiral R. Patel

Practical No. 11: Write a program to cluster a set of points using K-means. Training and test data must be provided explicitly.

A. **Objective:** Determining the correct number of clusters.

B. **Expected Program Outcomes (POs):-**PO1, PO2, PO3, PO6, PO7.

C. **Expected Skills to be developed based on competency:**

    I.   Learning data pre-processing task.
    II.   Determining optimal number of clusters.
   III. Understanding feature selection.

D. **Expected Course Outcomes(Cos)**

CO-4

E. **Practical Outcome(PRo)**

Predicting values using linear regresssion.

F. **Expected Affective domain Outcome(ADos)**

Handle tools /components/equipment carefully with safety and necessary precaution.

In software ethics Environment sustainability and environment consciousness whenever suitable.

G. **Prerequisite Theory:**

Refer unit 5 of course curriculum. Students are suggested to read chapter 9 of Machine Learning authored by Dutt, Chandramouli and das

H. **Resources/Equipment Required**

Sr.No.	Instrument/Equipment /Components/Trainer kit	Specification	Quantity
1	System supporting Jupyter Notebook	Python 3.x	1

        Course Coordinator: Ms. Hiral R. Patel

I.   Procedure to be followed/Source code (CE & IT software subjects):

J.   Practical related Quiz.

1. What is K-means clustering used for?

a) Dimensionality reduction

b) Data cleaning

c) Data clustering

d) Model selection

2. What is the objective of K-means clustering?

a) To minimize the sum of squared distances between data points and their centroids

b) To maximize the variance between data points and their centroids

c) To minimize the sum of absolute distances between data points and their centroids

d) To maximize the correlation between data points and their centroids

3. What is the value of K in K-means clustering?

a) The number of clusters

b) The number of data points

c) The number of features

d) The number of centroids

4. How is the initial centroid for K-means clustering selected?

a) Randomly

b) Based on the mean of the data points

c) Based on the median of the data points

d) Based on the mode of the data points

5. How do you evaluate the quality of the clustering in K-means clustering?

a) By calculating the sum of squared distances between data points and their centroids

b) By calculating the silhouette score

c) By calculating the F1 score

d) By calculating the Pearson correlation coefficient

K.   References / Suggestions (lab manual designer should give)

https://www.youtube.com/watch?v=EItlUEPCIzM

L.   Assessment Rubrics

Criteria	Engagement (4)	Accuracy (4)	Documentation (4)	Understanding &Explanation (4)	Time (4)	Total Marks (20)
Marks						

Faculty Signature

Course Coordinator: Ms. Hiral R. Patel