Diploma Engineering

Laboratory Manual

(Fundamentals of Machine Learning)
(4341603)

Information Technology Semester 4th

Enrolment No	
Name	
Branch	
Academic Term	
Institute	

Directorate Of Technical Education
Gandhinagar - Gujarat

DTE's Vision:

DTE's Mission:

Institute's Vision:

Institute's Mission:

Department's Vision:

Department's Mission:

Certificate

This is to certify that Mr./Ms …….. Enrolment No. ……. of ……. Semester of *Diploma in*……..of ……. (GTU Code) has satisfactorily completed the term work in course ……for the academic year: …… Term: Odd/Even prescribed in the GTU curriculum.

Place:……..

Date: ……..

Signature of Course Faculty **Head of the Department**

Preface

The primary aim of any laboratory/Practical/field work is enhancement of required skills as well as creative ability amongst students to solve real time problems by developing relevant competencies in psychomotor domain. Keeping in view, GTU has designed competency focused outcome-based curriculum -2021 (COGC-2021) for Diploma engineering programmes. In this more time is allotted to practical work than theory. It shows importance of enhancement of skills amongst students and it pays attention to utilize every second of time allotted for practical amongst Students, Instructors and Lecturers to achieve relevant outcomes by performing rather than writing practice in study type. It is essential for effective implementation of competency focused outcome- based Green curriculum-2021. Every practical has been keenly designed to serve as a tool to develop & enhance relevant industry needed competency in each and every student. These psychomotor skills are very difficult to develop through traditional chalk and board content delivery method in the classroom. Accordingly, this lab manual has been designed to focus on the industry defined relevant outcomes, rather than old practice of conducting practical to prove concept and theory.

By using this lab manual, students can read procedure one day in advance to actual performance day of practical experiment which generates interest and also, they can have idea of judgement of magnitude prior to performance. This in turn enhances predetermined outcomes amongst students. Each and every Experiment /Practical in this manual begins by competency, industry relevant skills, course outcomes as well as practical outcomes which serve as a key role for doing the practical. The students will also have a clear idea of safety and necessary precautions to be taken while performing experiment.

This manual also provides guidelines to lecturers to facilitate student-centred lab activities for each practical/experiment by arranging and managing necessary resources in order that the students follow the procedures with required safety and necessary precautions to achieve outcomes. It also gives an idea that how students will be assessed by providing Rubrics.

Fundamentals of machine learning course will help students to build up core competencies in understanding machine learning approaches and students will be able to design and train machine learning models for various use cases. The lab work of the course is designed to develop crisp understanding of the underpinning theory.

Although we try our level best to design this lab manual, but always there are chances of improvement. We welcome any suggestions for improvement.

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 analyze individual needs and engage in updating in the context of technological changes *in field of engineering.*

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 leaning model and learning the evaluation methods.*
c) *CO3: - Evaluate various supervised learning algorithms using appropriate dataset.*
d) *CO4: -Evaluate various unsupervised learning algorithms using appropriate dataset.*
e) *CO5:-To understand the use of various existing machine learning libraries.*

S. No.	Practical Outcome/Title of experiment	CO1	CO2	CO3	CO4	CO5
1.	Numerical Computing with Python (NumPy, Matplotlib)					✓
2.	Introduction to Pandas for data import and export (Excel, CSV etc.)					✓
3.	Basic Introduction to Scikit learn					✓
4.	Implement the Find-S concept learning algorithm that finds the most specific hypothesis that is consistent with the given training data. Conditions: Hypothesis can only be conjunction (AND) of literals. Literals are either attributes or their negations.		✓			
5.	Import Pima indian diabetes data Apply select K best and chi2 for feature selection Identify the best features		✓			
6.	Write a program to learn a decision tree and use it to predict class labels of test data Training and test data will be explicitly provided by instructor. Tree pruning should not be performed.			✓		
7.	ML Project Use the following dataset as music.csv			✓		

	<table><tr><th>age</th><th>gender</th><th>genre</th></tr><tr><td>20</td><td>1</td><td>HipHop</td></tr><tr><td>23</td><td>1</td><td>HipHop</td></tr><tr><td>25</td><td>1</td><td>HipHop</td></tr><tr><td>26</td><td>1</td><td>Jazz</td></tr><tr><td>29</td><td>1</td><td>Jazz</td></tr><tr><td>30</td><td>1</td><td>Jazz</td></tr><tr><td>31</td><td>1</td><td>Classical</td></tr><tr><td>33</td><td>1</td><td>Classical</td></tr><tr><td>37</td><td>1</td><td>Classical</td></tr><tr><td>20</td><td>0</td><td>Dance</td></tr><tr><td>21</td><td>0</td><td>Dance</td></tr><tr><td>25</td><td>0</td><td>Dance</td></tr><tr><td>26</td><td>0</td><td>Acoustic</td></tr><tr><td>27</td><td>0</td><td>Acoustic</td></tr><tr><td>30</td><td>0</td><td>Acoustic</td></tr><tr><td>31</td><td>0</td><td>Classical</td></tr><tr><td>34</td><td>0</td><td>Classical</td></tr><tr><td>35</td><td>0</td><td>Classical</td></tr></table> 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 decision tree 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. e. vary training and test size to check different accuracy values model achieves.					
8.	Write a program to use a K-nearest neighbour to predict class labels of test data.Training and test data must be provided explicitly.			✓		
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.			✓		
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					
11.	Write a program to cluster a set of points using K-means.Training and test data must be provided explicitly.				✓	
12.	Import Iris dataset a. Find rows and columns using shape command b. Print first 30 instances using head command c. Find out the data instances in each class. (use group by and size) e. Plot the univariate graphs (box plot and histograms) f. Plot the multivariate plot (scatter matrix) g. Split data to train model by 80% data values h. Apply K-NN and k means clustering to check accuracy and decide which is better.				✓	

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.

Guidelines to Course Faculty

1. Couse faculty should demonstrate experiment with all necessary implementation strategies described in curriculum.
2. Couse faculty should explain industrial relevance before starting of each experiment.
3. Course faculty should involve & give opportunity to all students for hands on experience.
4. Course faculty should ensure mentioned skills are developed in the students by asking.

5. Utilise 2 hours of lab hours effectively and ensure completion of write up with quiz also.
6. Encourage peer to peer learning by doing same experiment through fast learners.

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.

Continuous Assessment Sheet

Name: **Enrolment No:** **Term:**

Sr no	Practical Outcome/Title of experiment	Page	Date	Marks (25)	Sign
1	Numerical Computing with Python (NumPy, Matplotlib)				
2	Introduction to Pandas for data import and export (Excel, CSV etc.)				
3	Basic Introduction to Scikit learn				
4	Implement the Find-S concept learning algorithm that finds the most specific hypothesis that is consistent with the given training data. Conditions: Hypothesis can only be conjunction (AND) of literals. Literals are either attributes or their negations.				
5	Import Pima indian diabetes data Apply select K best and chi2 for feature selection Identify the best features				
6	Write a program to learn a decision tree and use it to predict class labels of test data Training and test data will be explicitly provided by instructor.				

<table>
<tr><td></td><td>Tree pruning should not be performed.</td><td></td><td></td><td></td><td></td></tr>
<tr><td>7</td><td>ML Project
Use the following dataset as music.csv
<table>
<tr><th>age</th><th>gender</th><th>genre</th></tr>
<tr><td>20</td><td>1</td><td>HipHop</td></tr>
<tr><td>23</td><td>1</td><td>HipHop</td></tr>
<tr><td>25</td><td>1</td><td>HipHop</td></tr>
<tr><td>26</td><td>1</td><td>Jazz</td></tr>
<tr><td>29</td><td>1</td><td>Jazz</td></tr>
<tr><td>30</td><td>1</td><td>Jazz</td></tr>
<tr><td>31</td><td>1</td><td>Classical</td></tr>
<tr><td>33</td><td>1</td><td>Classical</td></tr>
<tr><td>37</td><td>1</td><td>Classical</td></tr>
<tr><td>20</td><td>0</td><td>Dance</td></tr>
<tr><td>21</td><td>0</td><td>Dance</td></tr>
<tr><td>25</td><td>0</td><td>Dance</td></tr>
<tr><td>26</td><td>0</td><td>Acoustic</td></tr>
<tr><td>27</td><td>0</td><td>Acoustic</td></tr>
<tr><td>30</td><td>0</td><td>Acoustic</td></tr>
<tr><td>31</td><td>0</td><td>Classical</td></tr>
<tr><td>34</td><td>0</td><td>Classical</td></tr>
<tr><td>35</td><td>0</td><td>Classical</td></tr>
</table>
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 decision tree 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 like dance)
d. Calculate the accuracy of the model.
e. vary training and test size to check different accuracy values model achieves.</td><td></td><td></td><td></td><td></td></tr>
<tr><td>8</td><td>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.</td><td></td><td></td><td></td><td></td></tr>
<tr><td>9</td><td>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.</td><td></td><td></td><td></td><td></td></tr>
</table>

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				
11	Write a program to cluster a set of points using K-means. Training and test data must be provided explicitly.				
12	Import Iris dataset a. Find rows and columns using shape command b. Print first 30 instances using head command c. Find out the data instances in each class. (use group by and size) i. Plot the univariate graphs (box plot and histograms) j. Plot the multivariate plot (scatter matrix) k. Split data to train model by 80% data values l. Apply K-NN and k means clustering to check accuracy and decide which is better.				

Date: ……

Practical No.1: Numerical Computing with Python (NumPy, 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 -1

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. 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.

J. 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)

K. 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?
   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()

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

Numpy

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

Matplotlib

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

M. Assessment-Rubrics

Criteria	Total Marks	Exceptional (5 - Marks)	Satisfactory (4 to 3 - Marks)	Developing(2 - Marks)	Limited (1 -Mark)
Engagement	5	Performed practical him/her self	Performed practical with others help	Watched other students performing practical but not tried him/her self	Present in practical session but not attentively participated in performance
Accuracy	5	Accurately done	1-2 errors/mistakes found	3-5 errors/mistakes identified	More than 5 errors/mistakes committed
Documentation	5	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	5	Fully understood the performance & can explain perfectly	Understood the performance but cannot explain	Partially understood the performance &can give little explanation	Partially understood and cannot give explanation

Time	5	Completed thework within 1 week	Work is submitted later than 1 week but by the end of 2ndweek	Work done after 2nd week but before the end of3rd week	Work submittedafter 3 week time

Max 25 marks

Sign with Date

Practical No.2: Introduction to Pandas for data import and export (Excel, CSV etc.)

A. Objective: Getting familiarized with python machine learning libraries.

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 machine learning libraries.
II. Using machine learning methods in python libraries.

D. Expected Course Outcomes(Cos)

CO-5

E. Practical Outcome(PRo)

Using machine learning methods implemented in Pandas library.

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 6 of course curriculum. Also explore the link following link

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

H. Resources/Equipment Required

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

I. 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.

J. 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)

K. 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

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

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

M. Assessment-Rubrics

Criteria	Total Marks	Exceptional (5 - Marks)	Satisfactory (4 to 3 - Marks)	Developing(2 - Marks)	Limited (1 -Mark)
Engagement	5	Performed practical him/her self	Performed practical with others help	Watched other students performing practical but not tried him/her self	Present in practical session but not attentively participated in performance
Accuracy	5	Accurately done	1-2 errors/mistakes found	3-5 errors/mistakes identified	More than 5 errors/mistakes committed
Documentation	5	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	5	Fully understood the performance & can explain perfectly	Understood the performance but cannot explain	Partially understood the performance &can give little explanation	Partially understood and cannot give explanation
Time	5	Completed thework within 1 week	Work is submitted later than 1 week but by the end of 2ndweek	Work done after 2nd week but before the end of3rd week	Work submittedafter 3 week time

Max 25 marks

Sign with Date

Practical No.3: Basic Introduction to Scikit learn

A. Objective: Getting familiarized with python machine learning libraries.

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 machine learning libraries.
II. Using machine learning methods in python libraries.

D. Expected Course Outcomes(Cos)

CO-5

E. Practical Outcome(PRo)

Using machine learning methods implemented in Scikit library.

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 6 of course curriculum. Also explore the link following link

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

H. Resources/Equipment Required

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

I. 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.

J. 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

K. Practical related Quiz.

State True or False

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

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

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

M. Assessment-Rubrics

Criteria	Total Marks	Exceptional (5 - Marks)	Satisfactory (4 to 3 - Marks)	Developing(2 - Marks)	Limited (1 -Mark)
Engagement	5	Performed practical	Performed practical with others help	Watched other students performing practical but not tried him/her self	Present in practical session but not attentively

		him/her self			participated in performance
Accuracy	5	Accurately done	1-2 errors/mistakes found	3-5 errors/mistakes identified	More than 5 errors/mistakes committed
Documentation	5	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	5	Fully understood the performance & can explain perfectly	Understood the performance but cannot explain	Partially understood the performance &can give little explanation	Partially understood and cannot give explanation
Time	5	Completed thework within 1 week	Work is submitted later than 1 week but by the end of 2ndweek	Work done after 2nd week but before the end of3rd week	Work submittedafter 3 week time

Max 25 marks

Sign with Date

Practical No.4: Implement the Find-S concept learning algorithm that finds the most specific hypothesis that is consistent with the given training data.

Conditions:

Hypothesis can only be conjunction (AND) of literals. Literals are either attributes or their negations.

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. Organising search for an acceptable hypothesis.
II. Implementing a procedure.
III. Testing use cases over implementing procedures.

D. Expected Course Outcomes(Cos)

CO-2

E. Practical Outcome(PRo)

Identifying the best features using K best and chi2 algorithms on Pima indian diabetes data.

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 TOM. M. Mithell

Dataset

1	1	1	1	1	1	0	1	1
1	1	1	1	1	1	0	0	1
1	1	1	1	1	1	1	1	0
1	1	1	1	1	0	0	1	1
1	1	1	1	1	0	0	0	1
1	1	1	0	1	1	0	1	1
1	1	0	1	1	1	0	1	0

1	1	1	0	1	1	0	0	1
1	1	1	0	1	0	0	1	1
1	1	1	0	1	0	0	0	1
0	1	1	1	1	1	0	1	1
0	1	1	1	1	1	0	0	1
1	0	1	1	1	1	0	1	0
0	1	1	1	1	0	0	1	1
1	1	0	1	0	1	0	1	0
1	0	0	1	1	1	0	1	0
1	0	0	1	0	1	1	1	0
0	1	1	1	1	0	0	0	1
1	0	1	1	1	1	1	1	0
0	1	1	0	1	1	0	1	1

H. Resources/Equipment Required

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

I. 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.

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

Student must use the space for writing source code.

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

Observation Table: Prepare chart of Input use cases. : Boolean input attributes (x1, x2, … , x8) in first 8 columns. The last (9th) column represents the Boolean class label (y). Each row is a training instance. There are 20 training instances as mentioned in prerequisite theory.

L. Practical related Quiz.

State True or False

1. Find S algorithm only considers positive training examples and neglect negative training examples.
2. In Find-S algorithm we move bottom to top i.e. general hypothesis to specific hypothesis.
3. A maximally specific hypothesis covers none of the negative training examples.

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

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

N. Assessment-Rubrics

Criteria	Total Marks	Exceptional (5 - Marks)	Satisfactory (4 to 3 - Marks)	Developing(2 - Marks)	Limited (1 -Mark)
Engagement	5	Performed practical him/her self	Performed practical with others help	Watched other students performing practical but not tried him/her self	Present in practical session but not attentively participated in performance
Accuracy	5	Accurately done	1-2 errors/mistakes found	3-5 errors/mistakes identified	More than 5 errors/mistakes committed
Documentation	5	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	5	Fully understood the performance & can explain perfectly	Understood the performance but cannot explain	Partially understood the performance &can give little explanation	Partially understood and cannot give explanation
Time	5	Completed thework within 1 week	Work is submitted later than 1 week but by the end of 2ndweek	Work done after 2nd week but before the end of3rd week	Work submittedafter 3 week time

Max 25 marks

Sign with Date

Practical No.5: Import Pima indian diabetes data Apply select K best and chi2 for feature selection Identify the best features.

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)

Identifying the best features using K best and chi2 algorithms on Pima indian diabetes data.

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

H. Resources/Equipment Required

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

I. 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.

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

Student must use the space for writing source code.

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

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

L. Practical related Quiz.

1. What is the main advantage of using feature selection?
   a) speeding-up the training of an algorithm
   b) fine tuning the model's performance
   c) remove noisy features

2. When selecting feature, the decision should be made using:
   a) the entire dataset
   b) the training set
   c) the testing set

3. Given 20 potential features, How many models do you have to evaluate in all the subsets algorithm

   a) 20
   b) 40
   c) 1048576
   d) 1048596

4. The best fit model of size 5(i.e., with 5 features) always contains the set of features from best fit model of size 4.

   a) True
   b) False

M. 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

N. Assessment-Rubrics

Criteria	Total Marks	Exceptional (5 - Marks)	Satisfactory (4 to 3 - Marks)	Developing(2 - Marks)	Limited (1 -Mark)
Engagement	5	Performed practical him/her self	Performed practical with others help	Watched other students performing practical but not tried him/her self	Present in practical session but not attentively participated in performance

Accuracy	5	Accurately done	1-2 errors/mistakes found	3-5 errors/mistakes identified	More than 5 errors/mistakes committed
Documentation	5	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	5	Fully understood the performance & can explain perfectly	Understood the performance but cannot explain	Partially understood the performance &can give little explanation	Partially understood and cannot give explanation
Time	5	Completed thework within 1 week	Work is submitted later than 1 week but by the end of 2ndweek	Work done after 2nd week but before the end of3rd week	Work submittedafter 3 week time

Max 25 marks

Sign with Date

Practical No.6: Write a program to learn a decision tree and use it to predict class labels of test data.

- ✓ Training and test data will be explicitly provided by instructor.
- ✓ Tree pruning should not be performed.

A. Objective: Learning Decision tree and predicting class labels.

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

D. Expected Course Outcomes(Cos)

CO -3

E. Practical Outcome(PRo)

Identifying the class labels using decision tree.

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:

The learned tree should be tested on test instances with unknown class labels, and the predicted class labels for the test instances should be printed as output. Predicted class labels (0/1) for the test data must be exactly in the order in which the test instances are present in the test file.

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. 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.

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

Student must use the space for writing source code.

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

Observation Table: Prepare a table of predicted class labels.

Calculations:

Calculate confusion matrix

L. Practical related Quiz.

1. What is a decision tree?

a) A visual representation of decision-making using nodes and branches

b) A mathematical formula for predicting outcomes

c) A statistical model for regression analysis

2. What is the purpose of a decision tree?

a) To predict outcomes based on input variables

b) To summarize data using a graphical representation

c) To perform hypothesis testing on a dataset

3. What is a split in a decision tree?

a) A branch that represents a decision based on a feature or attribute

b) A point where the tree branches into different paths

c) A method for reducing the complexity of a decision tree

4. What is pruning in a decision tree?

a) A technique for simplifying the tree by removing branches that don't contribute to accuracy

b) A method for reducing the number of input variables

c) A way to increase the complexity of a decision tree

5. What is over fitting in a decision tree?

a) When the tree is too simple and doesn't capture all the relevant information

b) When the tree is too complex and fits the training data too closely

c) When the input variables are not correlated with the outcome variable

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

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

N. Assessment-Rubrics

Criteria	Total Marks	Exceptional (5 - Marks)	Satisfactory (4 to 3 - Marks)	Developing(2 - Marks)	Limited (1 -Mark)
Engagement	5	Performed practical him/her self	Performed practical with others help	Watched other students performing practical but not tried him/her self	Present in practical session but not attentively participated in performance
Accuracy	5	Accurately done	1-2 errors/mistakes found	3-5 errors/mistakes identified	More than 5 errors/mistakes committed
Documentation	5	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	5	Fully understood the performance & can explain perfectly	Understood the performance but cannot explain	Partially understood the performance &can give little explanation	Partially understood and cannot give explanation
Time	5	Completed thework within 1 week	Work is submitted later than 1 week but by the end of 2ndweek	Work done after 2nd week but before the end of3rd week	Work submittedafter 3 week time

Max 25 marks

Sign with Date

Practical No.7: 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

i. Store file as music.csv and import it to python using pandas
ii. Prepare the data by splitting data in input(age ,gender) and output(genre) data set
iii. Use decision tree 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)
iv. Calculate the accuracy of the model.
v. Vary training and test size to check different accuracy values model achieves.

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. 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.

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

Student must use the space for writing source code.

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

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

Calculations:

Calculate confusion matrix

L. Practical related Quiz.

1. What is the root node in a decision tree?

a) The topmost node that represents the output or decision

b) The node that has no parent

c) The node that has the maximum number of child nodes

d) The node that is located at the centre of the tree

2. What are some advantages of using decision trees for machine learning?

a) They are easy to interpret and visualize.

b) They can handle both categorical and numerical data.

c) They can handle missing values and noisy data.

d) All of the above.

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

Python Machine Learning Tutorial (Data Science):-

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

N. Assessment-Rubrics

Criteria	Total Marks	Exceptional (5 - Marks)	Satisfactory (4 to 3 - Marks)	Developing(2 - Marks)	Limited (1 -Mark)
Engagement	5	Performed practical him/her self	Performed practical with others help	Watched other students performing practical but not tried him/her self	Present in practical session but not attentively participated in performance
Accuracy	5	Accurately done	1-2 errors/mistakes found	3-5 errors/mistakes identified	More than 5 errors/mistakes committed

Documentation	5	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	5	Fully understood the performance & can explain perfectly	Understood the performance but cannot explain	Partially understood the performance &can give little explanation	Partially understood and cannot give explanation
Time	5	Completed thework within 1 week	Work is submitted later than 1 week but by the end of 2ndweek	Work done after 2nd week but before the end of3rd week	Work submittedafter 3 week time

Max 25 marks

Sign with Date

Practical No.8: 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: Learn simplest supervised machine learning algorithm used for classification.

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

D. Expected Course Outcomes(Cos)

CO-3

E. Practical Outcome(PRo)

Classifying data points based on how its neighbour are classified.

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:

Training data: data.csv

1	1	1	1	1	1	0	1	1
1	1	1	1	1	1	0	0	1
1	1	1	1	1	1	1	1	0
1	1	1	1	1	0	0	1	1
1	1	1	1	1	0	0	0	1
1	1	1	0	1	1	0	1	1
1	1	0	1	1	1	0	1	0
1	1	1	0	1	1	0	0	1
1	1	1	0	1	0	0	1	1
1	1	1	0	1	0	0	0	1
0	1	1	1	1	1	0	1	1
0	1	1	1	1	1	0	0	1
1	0	1	1	1	1	0	1	0
0	1	1	1	1	0	0	1	1

1	1	0	1	0	1	0	1	0
1	0	0	1	1	1	0	1	0
1	0	0	1	0	1	1	1	0
0	1	1	1	1	0	0	0	1
1	0	1	1	1	1	1	1	0
0	1	1	0	1	1	0	1	1

Test Data: test.csv

0	1	1	1	1	1	1	1
1	0	0	0	0	0	0	0
0	1	1	0	1	0	0	0
0	1	1	1	1	0	0	0

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*

I. 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.

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

Student must use the space for writing source code.

K. Practical related Quiz.

1. What does KNN stand for?

   a) K-Nearest Neighbors

   b) Kernel Nonlinear Network

   c) K-Means Nearest Neighbors

   d) None of the above

2. In KNN, how is the distance between a new data point and its neighbors typically measured?

   a) Euclidean distance

   b) Manhattan distance

   c) Cosine similarity

   d) All of the above

3. In what type of machine learning problems is KNN generally used?

   a) Regression problems

   b) Classification problems

   c) Clustering problems

   d) Dimensionality reduction problems

4. What are some advantages of using KNN for machine learning?

   a) It is a simple and easy-to-implement algorithm.

   b) It can handle both continuous and categorical data.

   c) It can adapt to complex decision boundaries.

   d) All of the above.

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

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

https://www.youtube.com/watch?v=4HKqjENq9OU

M. Assessment-Rubrics

Criteria	Total Marks	Exceptional (5 - Marks)	Satisfactory (4 to 3 - Marks)	Developing(2 - Marks)	Limited (1 -Mark)
Engagement	5	Performed practical him/her self	Performed practical with others help	Watched other students performing practical but not tried him/her self	Present in practical session but not attentively participated in performance
Accuracy	5	Accurately done	1-2 errors/mistakes found	3-5 errors/mistakes identified	More than 5 errors/mistakes committed
Documentation	5	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	5	Fully understood the performance & can explain perfectly	Understood the performance but cannot explain	Partially understood the performance &can give little explanation	Partially understood and cannot give explanation
Time	5	Completed thework within 1 week	Work is submitted later than 1 week but by the end of 2ndweek	Work done after 2nd week but before the end of3rd week	Work submittedafter 3 week time

Max 25 marks

Sign with Date

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.

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

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*

I. 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.

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

Student must use the space for writing source code.

K. 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 DataFrames

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)

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

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

M. Assessment-Rubrics

Criteria	Total Marks	Exceptional (5 - Marks)	Satisfactory (4 to 3 - Marks)	Developing(2 - Marks)	Limited (1 -Mark)
Engagement	5	Performed practical him/her self	Performed practical with others help	Watched other students performing practical but not tried him/her self	Present in practical session but not attentively participated in performance
Accuracy	5	Accurately done	1-2 errors/mistakes found	3-5 errors/mistakes identified	More than 5 errors/mistakes committed
Documentation	5	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	5	Fully understood the performance & can explain perfectly	Understood the performance but cannot explain	Partially understood the performance &can give little explanation	Partially understood and cannot give explanation
Time	5	Completed thework within 1 week	Work is submitted later than 1 week but by the end of 2ndweek	Work done after 2nd week but before the end of3rd week	Work submitted after 3 week time

Max 25 marks

Sign with Date

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

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. 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.

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

Student must use the space for writing source code.

K. 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

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

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

M. Assessment-Rubrics

Criteria	Total Marks	Exceptional (5 - Marks)	Satisfactory (4 to 3 - Marks)	Developing(2 - Marks)	Limited (1 -Mark)
Engagement	5	Performed practical him/her self	Performed practical with others help	Watched other students performing practical but not tried him/her self	Present in practical session but not attentively participated in performance
Accuracy	5	Accurately done	1-2 errors/mistakes found	3-5 errors/mistakes identified	More than 5 errors/mistakes committed
Documentation	5	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	5	Fully understood the performance & can explain perfectly	Understood the performance but cannot explain	Partially understood the performance &can give little explanation	Partially understood and cannot give explanation
Time	5	Completed thework within 1 week	Work is submitted later than 1 week but by the end of 2ndweek	Work done after 2nd week but before the end of3rd week	Work submittedafter 3 week time

Max 25 marks

Sign with Date

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*

I. 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.

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

Student must use the space for writing source code.

K. 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

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

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

M. Assessment-Rubrics

Criteria	Total Marks	Exceptional (5 - Marks)	Satisfactory (4 to 3 - Marks)	Developing(2 - Marks)	Limited (1 -Mark)
Engagement	5	Performed practical him/her self	Performed practical with others help	Watched other students performing practical but not tried him/her self	Present in practical session but not attentively participated in performance
Accuracy	5	Accurately done	1-2 errors/mistakes found	3-5 errors/mistakes identified	More than 5 errors/mistakes committed
Documentation	5	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	5	Fully understood the performance & can explain perfectly	Understood the performance but cannot explain	Partially understood the performance &can give little explanation	Partially understood and cannot give explanation
Time	5	Completed thework within 1 week	Work is submitted later than 1 week but by the end of 2ndweek	Work done after 2nd week but before the end of3rd week	Work submittedafter 3 week time

Max 25 marks

Sign with Date

Practical No.12: Import Iris dataset.

a. Find rows and columns using shape command
b. Print first 30 instances using head command
c. Find out the data instances in each class.(use groupby and size)
m. Plot the univariate graphs(box plot and histograms)
n. Plot the multivariate plot(scatter matrix)
o. Split data to train model by 80% data values
p. Apply K-NN and k means clustering to check accuracy and decide which is better.

A. Objective: Differentiate between supervised v/s unsupervised learning approaches

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

C. Expected Skills to be developed based on competency:

I. Learn to handle data efficiently.
II. Identifying the similarity between data sets.
III. Finding Neighbours and generating responses.
IV. Computing Accuracy of the method used.

D. Expected Course Outcomes(Cos)

CO-4

E. Practical Outcome(PRo)

Differentiate between supervised and unsupervised learning.

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*

I. 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.

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

Student must use the space for writing source code.

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

Observation Table: Draw accuracy table for both KNN and Kmeans.

L. Practical related Quiz.

1. Which algorithm is supervised and which one is unsupervised?

a) K-means clustering is supervised, KNN algorithm is unsupervised

b) K-means clustering is unsupervised, KNN algorithm is supervised

c) Both K-means clustering and KNN algorithm are supervised

d) Both K-means clustering and KNN algorithm are unsupervised

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

a) A classification of the data points into different classes

b) A prediction of the target variable for a given data point

c) A grouping of similar data points into K clusters

d) The K nearest neighbors for a given data point

3. What is the output of KNN algorithm?

a) A classification of the data points into different classes

b) A prediction of the target variable for a given data point

c) A grouping of similar data points into K clusters

d) The K nearest neighbors for a given data point

4. What is the primary objective of K-means clustering?

a) To classify data points into different classes

b) To find the K nearest neighbors for a given data point

c) To group similar data points into K clusters

d) To predict the target variable for a given data point

5. What is the primary objective of KNN algorithm?

a) To classify data points into different classes

b) To find the K nearest neighbors for a given data point

c) To group similar data points into K clusters

d) To predict the target variable for a given data point

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

https://youtu.be/6kZ-OPLNcgE

N. Assessment-Rubrics

Criteria	Total Marks	Exceptional (5 - Marks)	Satisfactory (4 to 3 - Marks)	Developing(2 - Marks)	Limited (1 -Mark)
Engagement	5	Performed practical him/her self	Performed practical with others help	Watched other students performing practical but not tried him/her self	Present in practical session but not attentively participated in performance
Accuracy	5	Accurately done	1-2 errors/mistakes found	3-5 errors/mistakes identified	More than 5 errors/mistakes committed
Documentation	5	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	5	Fully understood the performance & can explain perfectly	Understood the performance but cannot explain	Partially understood the performance &can give little explanation	Partially understood and cannot give explanation
Time	5	Completed thework within 1 week	Work is submitted later than 1 week but by the end of 2ndweek	Work done after 2nd week but before the end of3rd week	Work submittedafter 3 week time

Max 25 marks

Sign with Date