GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Diploma Engineering

Level: Diploma

Branch: Information Technology

Subject Code: DI04016031

Subject Name: Fundamental of Machine Learning

w. e. f. Academic Year:	2025-26
Semester:	4th
Category of the Course:	PCC

Prerequisite:	Students should have basic knowledge of mathematics including algebra, statistics, and probability. They should be familiar with Python programming and data structures. A foundational understanding of computer operations, file handling, and software installation is required.
Rationale:	Machine learning focuses on the use of data and algorithms to perform learning similar to the way human learns. To solve recent problems in IT domain it is important to understand the need of machine learning and apply machine learning methods in efficient ways. Every student of Information Technology must therefore understand the blue prints of machine learning approaches and must be able to apply learning methods on available datasets. This 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 modes for various use cases.

Course Outcome:

After Completion of the Course, Student will be able to:

No	Course Outcomes	RBT Level
01	To understand the need of machine learning for various problem solving.	Understand
02	Prepare machine learning models by performing appropriate data preprocessing, and apply evaluation methods to assess model performance.	Apply
03	Evaluate various supervised learning algorithms using appropriate datasets.	Apply
04	Evaluate various unsupervised learning algorithms using appropriate dataset.	Apply
05	Understand and apply various existing Python libraries for data preprocessing, visualization, and machine learning tasks.	Apply

**Revised Bloom's Taxonomy (RBT)*

GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Diploma Engineering

Level: Diploma

Branch: Information Technology

Subject Code: DI04016031

Subject Name: Fundamental of Machine Learning

Teaching and Examination Scheme:

Teaching Scheme (in Hours)			Total Credits L+T+ (PR/2)	Assessment Pattern and Marks				Total Marks
				Theory		Tutorial / Practical		
L	T	PR	C	ESE (E)	PA(M)	PA(I)	ESE(V)	
3	0	2	4	70	30	20	30	150

Course Content:

Unit No.	Content	No. of Hours	% of Weightage
1.	**Introduction to machine learning** 1.1 Describe basic concept of machine learning and its applications 1.1.1 Overview of Human Learning and Machine Learning 1.1.2 Types of Machine Learning 1.1.3 Applications of Machine Learning 1.1.4 Tools and Technology for Machine Learning	04	09
2.	**Preparing to Model** 2.1 Describe different types of Machine learning Activities. 2.1.1 Machine Learning activities 2.2 Explain types of data and data preprocessing 2.2.1 Types of data in Machine Learning 2.2.2 Structures of data 2.2.3 Data quality and remediation 2.2.4 Data Pre-Processing - Dimensionality reduction, Feature subset selection	07	15
3.	**Modeling and Evaluation** 3.1 Selecting a machine learning model 3.1.1 Selecting a Model - Predictive/Descriptive 3.2 Train the model for supervised learning 3.2.1 Training a Model for supervised learning: Holdout method, K-fold Cross-validation method 3.3 Evaluate the prepared model 3.3.1 Model representation and interpretability 3.3.2 Evaluating performance of a model Confusion Matrix 3.3.3 Improving Performance of a model	08	18

4.	**Supervised Learning - Classification and Regression** 4.1 Describe supervised learning 4.1.1 Introduction to supervised learning. 4.1.2 Classification Model 4.1.3 Learning steps 4.2 Explain classification Algorithms. 4.2.1 Classification Algorithms - k-Nearest Neighbor (KNN), Support Vector Machines 4.3 Explain Regression 4.3.1 Regression - Simple linear regression, Multiple linear regression, Logistic regression	10	22
5.	**Unsupervised Learning** 5.1 Explain unsupervised learning 5.1.1 Supervised vs. Unsupervised Learning 5.1.2 Applications of unsupervised learning 5.2 Describe Clustering 5.2.1 Clustering K-means Clustering algorithm 5.3 Describe pattern finding using association rule 5.3.1 Finding pattern using Association Rule, Apriori algorithm	08	18
6	**Python libraries for Machine learning** 6.1 Demonstrate the application of existing Python libraries for machine learning tasks, with emphasis on Pandas. 6.1.1 Data structures: Series, DataFrame 6.1.2 Data loading and storage: read_csv(), read_excel(), to_csv() 6.1.3 Data selection and filtering: loc[], iloc[] 6.1.4 Data cleaning: isnull(), dropna(), fillna() 6.1.5 Data aggregation: groupby(), pivot_table() 6.2 Demonstrate the application of existing Python libraries for machine learning tasks, with emphasis on Numpy 6.2.1 Array creation: array(), arange(), linspace() 6.2.2 Array operations: reshape(), concatenate(), transpose() 6.2.3 Mathematical functions: mean(), std(), dot(), sum() 6.2.4 Linear algebra: linalg.inv(), linalg.eig() 6.3 Demonstrate the application of existing Python libraries for	08	18

	machine learning tasks, with emphasis on Matplotlib 6.3.1 Basic plotting: plot(), scatter(), bar() 6.3.2 Histogram and pie charts: hist(), pie() 6.3.3 Customization: title(), xlabel(), ylabel(), legend() 6.3.4 Subplots: subplot(), subplots() 6.4 Demonstrate the application of existing Python libraries for machine learning tasks, with emphasis on Scikit-learn 6.4.1 Data preprocessing: train_test_split(), StandardScaler(), LabelEncoder() 6.4.2 Classification algorithms: LogisticRegression(), KNeighborsClassifier(), SVC() 6.4.3 Regression algorithms: LinearRegression(), 6.4.4 Model evaluation: accuracy_score(), confusion_matrix(), classification_report(), Clustering: KMeans()		
	Total	**45**	**100**

Suggested Specification Table with Marks (Theory):

Distribution of Theory Marks (in %)					
R Level	**U Level**	**A Level**	**N Level**	**E Level**	**C Level**
20	**40**	**40**	**-**	**-**	**-**

Where R: Remember; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create (as per Revised Bloom's Taxonomy)

References/Suggested Learning Resources:
(a) Books:

Sr. No.	Title of Book	Author	Publication with place, year and ISBN
1	Machine Learning: Step-by- Step Guide To Implement Machine Learning Algorithms with Python.	Rudolph Russell	Rudolph Russell Publications
2	Machine Learning	Saikat Dutt, S. Chandramouli	Das, Pearson
3	Machine Learning with Python Cookbook Practical Solutions from Preprocessing to Deep Learning.	Chris Albon	O'Reilly Media, Inc.

(b) Open source software and website:

1. https://www.geeksforgeeks.org/machine-learning/
2. https://www.tutorialspoint.com/machine_learning_with_python/index.htm
3. https://nptel.ac.in/
4. https://www.coursera.org/
5. https://scikit-learn.org/

Suggested Course Practical List:

The following practical outcomes (PrOs) are the subcomponents of the COs. These PrOs need to be attained to achieve the COs.

<table>
<tr><th>Sr. No.</th><th>Practical Outcomes (PrOs)</th><th>Unit No.</th><th>Approx. Hrs. required</th></tr>
<tr><td>1</td><td>Numerical Computing with Python (NumPy, Matplotlib). Perform various operation mentioned in unit 6 corresponding to Numpy and Matplotlib</td><td>VI</td><td>04</td></tr>
<tr><td>2</td><td>Introduction to Pandas for data import and export (Excel, CSV etc.) Perform various operation mentioned in unit 6 corresponding to Pandas</td><td>VI</td><td>04</td></tr>
<tr><td>3</td><td>Basic Introduction to Scikit learn. Perform various operation mentioned in unit 6 corresponding to Scikitlearn)</td><td>VI</td><td>04</td></tr>
<tr><td>4</td><td>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><th>Age Group</th><th>Gender</th><th>favorite Fruit</th></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></td><td>I</td><td>02</td></tr>
</table>

<table>
<tr>
<td>5</td>
<td>To understand different types of Machine Learning activities and learn how to handle and preprocess datasets for ML applications.
Consider following dataset: Student.csv
<table>
<tr><th>Name</th><th>Age</th><th>Gender</th><th>Hours Studied</th><th>Marks</th><th>Passed</th></tr>
<tr><td>Alice</td><td>20</td><td>F</td><td>5</td><td>80</td><td>Yes</td></tr>
<tr><td>Bob</td><td>21</td><td>M</td><td>3</td><td>50</td><td>No</td></tr>
<tr><td>Charlie</td><td>22</td><td>M</td><td>4</td><td>60</td><td>Yes</td></tr>
<tr><td>Diana</td><td>20</td><td>F</td><td>2</td><td>40</td><td>No</td></tr>
<tr><td>Eve</td><td>23</td><td>F</td><td>6</td><td>90</td><td>Yes</td></tr>
</table>
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</td>
<td>II</td>
<td>02</td>
</tr>
<tr>
<td>6</td>
<td>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.</td>
<td>III</td>
<td>02</td>
</tr>
<tr>
<td>7</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>IV</td>
<td>02</td>
</tr>
</table>

<table>
<tr>
<td>8</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 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.</td>
<td>IV</td>
<td>04</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.
d. Apply logistic regression for predicting house price.</td>
<td>IV</td>
<td>02</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 e. Highlight the correctness of Plot using suitable Metrics	**IV**	**02**
11	Write a program to cluster a set of points using K-means. Training and test data must be provided explicitly.	**V**	**02**
	Total		**30**

Note:-
More Practical Exercises can be designed and offered by the respective course teacher to develop the industry relevant skills/outcomes to match the COs. The above table is only a suggestive list.

List of Laboratory/Learning Resources Required:

Sr. No.	Laboratory/Learning Resources/Equipment Name with Broad Specifications	PrO. No.
1	Computer system with operating system: Windows 7 or higher version, macOS, and Linux, with 4GB or higher RAM, Python versions: 2.7.X, 3.6.X, or higher version.	**All**
2	Python IDEs and Code Editors Open Source : IDLE, Jupyter, Visual Studio Code (VS Code), Spyder	

Suggested Project List:

Only one project is planned to be undertaken by a student that needs to be assigned to him/her in the beginning of the semester. The project could be industry application based, internet-based, workshop-based, laboratory-based or field-based. Each project should encompass two or more COs which are in fact, an integration of PrOs, UOs and ADOs. Each student will have to maintain a dated work diary consisting of individual contributions in the project work and give a seminar presentation of it before submission. The duration of the project should be about 14-16 (fourteen to sixteen) student engagement hours during the course. The students ought to submit projects by the end of the semester to develop the industry-oriented COs.

A suggestive list of projects is given here. This has to match the COs. Similar projects could be added by the concerned course teacher:

i) BigMart Sales Prediction: BigMart sales dataset consists of 2013 sales data for 1559 products across 10 different outlets in different cities. The goal of the BigMart sales prediction ML project is to build a regression model to predict the sales of each of 1559 products for the following year in each of the 10 different BigMart outlets.

ii) Stock Price Prediction using machine learning is the process of predicting the future value of a stock traded on a stock exchange for reaping profits. With multiple factors involved in predicting stock prices, it is challenging to predict stock prices with high accuracy, and this is where machine learning plays a vital role.

Suggested Activities for Students:

Other than the classroom and laboratory learning, following are the suggested student related co-curricular activities which can be undertaken to accelerate the attainment of the various outcomes in this course: Students should conduct following activities in group and prepare reports of about 5 pages for each activity, also collect/record physical evidences for their (student's) portfolio which will be useful for their placement interviews:

a) Explore different data repositories and register for ML based competitions on platforms like kaggle.

b) Undertake micro-projects in teams

* * * * * * *