

**GUJARAT TECHNOLOGICAL UNIVERSITY (GTU)**

**Competency-focused Outcome-based Green Curriculum-2021 (COGC-2021)**

Semester -IV

**Course Title: Fundamentals of Machine Learning**

(Course Code: 4341603)

Diploma programme in which this course is offered	Semester in which offered
Information Technology	4 <sup>th</sup> semester

**1. 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.

**2. COMPETENCY**

The purpose of this course is to help the student to attain the following industry identified competency through various teaching-learning experiences:

- **Develop appropriate machine learning algorithms for problem solving**

**3. COURSE OUTCOMES (COs)**

The practical exercises, the underpinning knowledge, and the relevant soft skills associated with this competency are to be developed in the student to display the following COs:

The student will develop underpinning knowledge, adequate programming skills of competency for implementing various applications using python programming language to attain the following course outcomes.

- To understand the need of machine learning for various problem solving.
- Prepare machine learning model and learning the evaluation methods.
- Evaluate various supervised learning algorithms using appropriate dataset.
- Evaluate various unsupervised learning algorithms using appropriate dataset.
- To understand the use of various existing machine learning libraries.

**4. TEACHING AND EXAMINATION SCHEME**

Teaching Scheme (In Hours)			Total Credits (L+T/2+P/2)	Examination Scheme				Total Marks
L	T	P		Theory Marks		Practical Marks		
			C	CA	ESE	CA	ESE	
3	-	4	5	30	70	25	25	150

(\*): Out of 30 marks under the theory CA, 10 marks are for assessment of the micro-project to facilitate integration of COs and the remaining 20 marks is the average of 2 tests to be taken during the semester for the assessing the attainment of the cognitive domain UOs required for the attainment of the COs.

**Legends:** **L**-Lecture; **T** – Tutorial/Teacher Guided Theory Practice; **P** -Practical; **C** – Credit, **CA** - Continuous Assessment; **ESE** -End Semester Examination.

## 5. SUGGESTED PRACTICAL EXERCISES

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

S. No.	Practical Outcomes (PrOs)	Unit No.	Approx. Hrs. required
1	Numerical Computing with Python (NumPy, Matplotlib)	I	04
2	Introduction to Pandas for data import and export (Excel, CSV etc.)	VI	04
3	Basic Introduction to Scikit learn	VI	04
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.	II	06
5	Import Pima Indian diabetes data Apply select K best and chi2 for feature selection Identify the best features	II	04
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.	III	04
7	ML Project Use the following dataset as music.csv	III	06

	<b>age</b>	<b>gender</b>	<b>genre</b>		
	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 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 neighbor it to predict class labels of test data. Training and test data must be provided explicitly.			IV	04
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.			IV	04
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			IV	06
11	Write a program to cluster a set of points using K-means. Training and test data must be provided explicitly.			V	04
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)			V	06

	<ul style="list-style-type: none"> <li>e. Plot the univariate graphs(box plot and histograms)</li> <li>f. Plot the multivariate plot(scatter matrix)</li> <li>g. Split data to train model by 80% data values</li> <li>h. Apply K-NN and k means clustering to check accuracy and decide which is better.</li> </ul>		
	<b>Total</b>		<b>56</b>

**Note**

- i. 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.
- ii. The following are some **sample** 'Process' and 'Product' related skills (more may be added/deleted depending on the course) that occur in the above listed **Practical Exercises** of this course required which are embedded in the COs and ultimately the competency..

S. No.	Sample Performance Indicators for the PrOs	Weightage in %
1	Using the existing python libraries through Python Jupyter notebook.	25
2	Use python to read dataset and modify as per requirement.	20
3	Selecting appropriate machine learning method.	25
4	Train and test the model by importing existing data set.	10
5	Making predictions and improve learning parameters as well as improve accuracy.	20
	<b>Total</b>	<b>100</b>

**6. MAJOR EQUIPMENT/ INSTRUMENTS REQUIRED**

This major equipment with broad specifications for the PrOs is a guide to procure them by the administrators to usher in uniformity of practical in all institutions across the state.

S. No.	Equipment Name with Broad Specifications	PrO. No.
1	Computer system with operating system: Windows 7 or higher Ver., macOS, and Linux, with 4GB or higher RAM, Python versions: 2.7.X, 3.6.X	All
2	Python IDEs and Code Editors Open Source : Anaconda Navigator	

**7. AFFECTIVE DOMAIN OUTCOMES**

The following **sample** Affective Domain Outcomes (ADOs) are embedded in many of the above-mentioned COs and PrOs. More could be added to fulfill the development of this competency.

- a) Work as a Data scientist.
- b) Follow ethical practices.

The ADOs are best developed through the laboratory/field based exercises. Moreover, the level of achievement of the ADOs according to Krathwohl's 'Affective Domain Taxonomy' should gradually increase as planned below:

- i. 'Valuing Level' in 1<sup>st</sup> year
- ii. 'Organization Level' in 2<sup>nd</sup> year.
- iii. 'Characterization Level' in 3<sup>rd</sup> year.

## 9. UNDERPINNING THEORY

Only the major Underpinning Theory is formulated as higher-level UOs of *Revised Bloom's taxonomy* in order development of the COs and competency is not missed out by the students and teachers. If required, more such higher-level UOs could be included by the course teacher to focus on the attainment of COs and competency.

Unit	Unit Outcomes (UOs) (4 to 6 UOs at Application and above level)	Topics and Sub-topics
<b>Unit – I Introduction to machine learning</b>	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
<b>Unit – II Preparing to Model</b>	2.1 Describe different types of Machine learning Activities. 2.2 Explain types of data and data preprocessing.	2.1.1 Machine Learning activities 2.1.2 Types of data in Machine Learning 2.1.3 Structures of data 2.1.4 Data quality and remediation 2.1.5 Data Pre-Processing <ul style="list-style-type: none"> <li>• Dimensionality reduction</li> <li>• Feature subset selection</li> </ul>
<b>Unit– III Modeling and Evaluation</b>	3.1 selecting a machine learning model 3.2 Train the model for supervised learning 3.3 Evaluate the prepared model.	3.1.1 Selecting a Model <ul style="list-style-type: none"> <li>• Predictive/Descriptive</li> </ul> 3.2.1 Training a Model for supervised learning: <ul style="list-style-type: none"> <li>• Holdout method</li> <li>• K-fold Cross-validation method</li> </ul> 3.3.1 Model representation and interpretability 3.3.2 Evaluating performance of a model <ul style="list-style-type: none"> <li>• Confusion Matrix</li> </ul> 3.3.3 Improving Performance of a model
<b>Unit– IV Supervised Learning - Classification</b>	4.1 Describe supervised learning 4.2 Explain classification Algorithms. 4.3 Explain Regression.	4.1.1 Introduction to supervised learning. 4.1.2 Classification Model 4.1.3 Learning steps

<b>and Regression</b>		4.2.1 Classification Algorithms <ul style="list-style-type: none"> <li>• k-Nearest Neighbor (kNN)</li> <li>• Support Vector Machines</li> </ul> 4.3.1 Regression <ul style="list-style-type: none"> <li>• Simple linear regression</li> <li>• Multiple linear regression</li> <li>• Logistic regression</li> </ul>
<b>Unit– V Unsupervised Learning</b>	5.1 Explain unsupervised learning 5.2 Describe Clustering 5.3 Describe pattern finding using association rule	5.1.1 Supervised vs. Unsupervised Learning 5.1.2 Applications of unsupervised learning 5.2.1 Clustering <ul style="list-style-type: none"> <li>• K-means Clustering algorithm</li> </ul> 5.3.1 Finding pattern using Association Rule <ul style="list-style-type: none"> <li>• Apriori algorithm</li> </ul>
<b>Unit– VI Python libraries for Machine learning</b>	6.1 Explain the use of existing machine learning python libraries.	6.1.1 Pandas 6.1.2 Numpy 6.1.3 Matplotlib 6.1.4 Scikit-learn

**Note:** The UOs need to be formulated at the 'Application Level' and above of Revised Bloom's Taxonomy' to accelerate the attainment of the COs and the competency.

## 10. SUGGESTED SPECIFICATION TABLE FOR QUESTION PAPER DESIGN

Unit No.	Unit Title	Teaching Hours	Distribution of Theory Marks			
			R Level	U Level	A	Total Marks
I	Introduction to machine learning	04	02	04	02	8
II	Preparing to Model	06	02	06	04	12
III	Modeling and Evaluation	08	02	04	06	12
IV	Supervised Learning - Classification and Regression	10	02	04	08	14
V	Unsupervised Learning	08	02	06	04	12
VI	Python libraries for Machine learning	06	04	04	04	12
<b>Total</b>		<b>42</b>	<b>14</b>	<b>28</b>	<b>28</b>	<b>70</b>

**Legends:** R=Remember, U=Understand, A=Apply and above (Revised Bloom's taxonomy)

**Note:** This specification table provides general guidelines to assist students for their learning and to teachers to teach and question paper designers/setters to formulate test

items/questions assess the attainment of the UOs. The actual distribution of marks at different taxonomy levels (of R, U and A) in the question paper may vary slightly from the above table.

### 11. SUGGESTED STUDENT ACTIVITIES

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
- c) Give a seminar on any relevant topics.
- d) Collect various sensor data from smart phones and apply machine learning approach.

### 12. SUGGESTED SPECIAL INSTRUCTIONAL STRATEGIES (if any)

These are sample strategies, which the teacher can use to accelerate the attainment of the various outcomes in this course:

- a) Massive open online courses (**MOOCs**) may be used to teach various topics/subtopics.
- b) Guide student(s) in undertaking micro-projects.
- c) '**L**' in **section No. 4** means different types of teaching methods that are to be employed by teachers to develop the outcomes.
- d) About **20% of the topics/sub-topics** which are relatively simpler or descriptive in nature is to be given to the students for **self-learning**, but to be assessed using different assessment methods.
- e) With respect to **section No.11**, teachers need to ensure to create opportunities and provisions for **co-curricular activities**.
- f) Guide students for open source python editors.

### 13. SUGGESTED MICRO-PROJECTS

**Only one micro-project** is planned to be undertaken by a student that needs to be assigned to him/her in the beginning of the semester. In the first four semesters, the micro-project are group-based. However, in the fifth and sixth semesters, it should be preferably be **individually** undertaken to build up the skill and confidence in every student to become problem solver so that s/he contributes to the projects of the industry. In special situations where groups have to be formed for micro-projects, the number of students in the group should **not exceed three**.

The micro-project could be industry application based, internet-based, workshop-based, laboratory-based or field-based. Each micro-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 total duration of the micro-project should not be less than **16 (sixteen) student engagement hours** during the course. The student ought to submit a micro-project by the end of the semester to develop the industry oriented COs.

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

- **Project idea 1: 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.
- **Project idea 2: 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.
- **Project idea 3:** Data from leading music service can be taken to build a better music recommendation system.

#### 14. SUGGESTED LEARNING RESOURCES

S. 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	SaikatDull,S.Chjandr amouli	Das, Pearson
3	Machine Learning with Python Cookbook_ Practical Solutions from Preprocessing to Deep Learning.	Chris Albon	O'Reilly Media, Inc.

#### 15. SOFTWARE/LEARNING WEBSITES

- <https://www.geeksforgeeks.org/machine-learning/>
- [https://www.tutorialspoint.com/machine\\_learning\\_with\\_python/index.htm](https://www.tutorialspoint.com/machine_learning_with_python/index.htm)
- <https://nptel.ac.in/>
- <https://www.coursera.org/>
- <https://scikit-learn.org/>

#### 16. PO-COMPETENCY-CO MAPPING

Semester II	Fundamentals of Machine Learning(Course Code: 4341603)						
	POs and PSOs						
Competency & Course Outcomes	PO 1 Basic & Discipline specific knowledge	PO 2 Problem Analysis	PO 3 Design/ development of solutions	PO 4 Engineering Tools, Experimentation & Testing	PO 5 Engineering practices for society, sustainability & environment	PO 6 Project Management	PO 7 Life-long learning
<b>Competency</b> Develop a machine learning model to solve real world problems.							
<b>Course Outcomes</b> CO a)To understand the need of machine learning for various problem solving.	3	2	3	2	-	-	3



CO b) Prepare machine learning model and learning the evaluation methods.	3	2	3	2	-	2	3
CO c) Evaluate various supervised learning algorithms using appropriate dataset	3	3	3	3	-	3	3
CO d) Evaluate various unsupervised learning algorithms using appropriate dataset	3	3	3	3	-	3	3
CO e) To understand the use of various existing machine learning libraries	3	2	3	3	-	-	3

Legend: '3' for high, '2' for medium, '1' for low or '-' for the relevant correlation of each competency, CO, with PO/ PSO

## 17. COURSE CURRICULUM DEVELOPMENT COMMITTEE

### GTU Resource Persons

Sr. No.	Name and Designation	Institute	Email
1	Mr. Sunil K. Paryani- Head(IT)	Government Polytechnic Ahmedabad	mailtosunil9@gmail.com
2	Ms. Hiral R. Patel - Lect.(IT)	Government Polytechnic Gandhinagar	hiralit@gmail.com
3	Mr. Pramod K. Tripathi - Lect.(IT)	Government Polytechnic Gandhinagar	csharp.pramod@gmail.com