GUJARAT TECHNOLOGICAL UNIVERSITY (GTU)

Competency-focused Outcome-based Green Curriculum-2021 (COGC-2021)
Semester -V

Course Title: Introduction to Machine Learning
(Course Code: 4350702)

Diploma programme in which this course is offered	Semester in which offered
Computer Engineering	5th 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 Computer domain it is important to understand the need of machine learning and apply machine learning methods in efficient ways. Every student of Computer Engineering 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

Students acquire the ability to assess and analyze outcomes produced by machine learning algorithms and models, while enhancing their capacity for critical thinking in addressing practical challenges.

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.

a) Describe basic concept of machine learning and its applications

b) Practice Numpy, Pandas, Matplotlib, sklearn library's inbuilt function required to solve machine learning problems

c) Use Pandas library for data preprocessing

d) Apply supervised learning algorithms based on dataset characteristics

e) Apply unsupervised learning algorithms based on dataset characteristics

4.   TEACHING AND EXAMINATION SCHEME

Teaching Scheme (In Hours)			Total Credits (L+T/2+P/2)	Examination Scheme				
				Theory Marks		Practical Marks		Total Marks
L	T	P	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	Explore any one machine learning tool. (like Weka, Tensorflow, Scikit-learn, Colab, etc.)	I	4
2	Write a NumPy program to implement following operation • to convert a list of numeric values into a one-dimensional NumPy array • to create a 3x3 matrix with values ranging from 2 to 10 • to append values at the end of an array • to create another shape from an array without changing its data(3*2 to 2*3)	II	4
3	Write a NumPy program to implement following operation • to split an array of 14 elements into 3 arrays, each with 2, 4, and 8 elements in the original order • to stack arrays horizontally (column wise)	II	4
4	Write a NumPy program to implement following operation • to add, subtract, multiply, divide arguments element-wise • to round elements of the array to the nearest integer • to calculate mean across dimension, in a 2D numpy array • to calculate the difference between neighboring elements, element-wise of a given array	II	4
5	Write a NumPy program to implement following operation • to find the maximum and minimum value of a given flattened array • to compute the mean, standard deviation, and variance of a given array along the second axis	II	4

6	Write a Pandas program to implement following operation • to convert a NumPy array to a Pandas series • to convert the first column of a DataFrame as a Series • to create the mean and standard deviation of the data of a given Series • to sort a given Series	II	4
7	Write a Pandas program to implement following operation • to create a dataframe from a dictionary and display it • to sort the DataFrame first by 'name' in ascending order • to delete the one specific column from the DataFrame • to write a DataFrame to CSV file using tab separator	II	4
8	Write a Pandas program to create a line plot of the opening, closing stock prices of given company between two specific dates.	II	4
9	Write a Pandas program to create a plot of Open, High, Low, Close, Adjusted Closing prices and Volume of given company between two specific dates.	II	4
10	Write a Pandas program to implement following operation • to find and drop the missing values from the given dataset • to remove the duplicates from the given dataset	III	4
11	Write a Pandas program to filter all columns where all entries present, check which rows and columns has a NaN and finally drop rows with any NaNs from the given dataset.	III	4
12	Write a Python program using Scikit-learn to print the keys, number of rows-columns, feature names and the description of the given data.	III	4
13	Write a Python program to implement K-Nearest Neighbour supervised machine learning algorithm for given dataset.	IV	4
14	Write a Python program to implement a machine learning algorithm for given dataset. (It is recommended to assign different machine learning algorithms group wise – micro project)	V	4
	Total		**56**

<u>*Note*</u>

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.	30
2	Use python to read dataset and modify as per requirement.	25
3	Selecting appropriate machine learning method.	20
4	Train and test the model by importing existing data set.	15
5	Making predictions and improve learning parameters as well as improve accuracy.	10
Total		**100**

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 (jupyter, spyder, google colab) 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 1st year
   ii.   'Organization Level' in 2nd year.
   iii. 'Characterization Level' in 3rd year.

8.   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)	Topics and Sub-topics
Unit – I **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 • Supervised Machine Learning • Unsupervised Machine Learning • Reinforcement Learning. 1.1.3 Applications of Machine Learning 1.1.4 Tools and Technology for Machine Learning
Unit – II **Python libraries suitable for Machine Learning**	2.1. Develop Program using Python Libraries	2.1 Numpy • Creating Array: array() • Accessing Array: by referring to its index number • Stacking & Splitting: stack(), array_split() • Maths Functions: add(), subtract(), multiply(), divide(), power(), mod() • Statistics Functions: amin(), amax(), mean(), median(), std(), var(), average(), ptp() 2.2 Pandas • Series: Series() • Dataframes: DataFrames() • Read CSV File: read_csv() • Cleaning Empty Cells: dropna() • Cleaning Wrong Data: drop() • Removing Duplicates: duplicated() • Pandas Plotting: plot() 2.3 Matplotlib • Pyplot.plot: plot() • Show: show() • Labels: xlabel(), ylabel() • Grid: grid() • Bars: bar() • Histogram: hist() • Subplot: subplot() • pie chart: pie() • Save the plotted images into pdf: savefig() 2.4 sklearn • Key concepts and features • Steps to Build a Model in Sklearn: Loading a Dataset- `read_csv()`, train_test_split- train_test_split()

Unit – III **Preparing to Model and Preprocessing**	3.1 Describe different types of Machine learning Activities 3.2 Explain Data preprocessing	3.1.1 Machine Learning activities • Preparing to Model • Learning: Data Partition-*k*-fold cross validation, Model Selection • Performance Evaluation: confusion matrix • Performance Improvement: Ensemble 3.2.1 Types of Data • Qualitative/Categorical Data: Nominal, Ordinal • Quantitative/Numeric Data: Interval, Ratio 3.2.2 Data quality and remediation • Handling outliers • Handling missing values 3.2.3 Data Pre-Processing • Dimensionality reduction • Feature subset selection: Filter, Wrapper, Hybrid, Embedded
Unit– IV **Supervised Machine Learning Models**	4.1 Define Supervised Learning 4.2 List types of Supervised Learning, Describe K-Nearest Neighbour and Simple linear regression 4.3 Advantage and disadvantage of supervised machine learning	4.1.1 Introduction of Supervised Learning • Brief explanation of Supervised Machine Learning • Working of Supervised Machine learning • Real world Applications/Examples of Supervised Machine learning • Steps in Supervised Machine learning 4.2.1 Types of Supervised Learning ● Classification: Define Classification, list types of classification, list types of Machine learning classification algorithms (list linear models, nonlinear models), list use cases of classification algorithms. **K-Nearest Neighbour (K-NN) :** Working of K-NN, Need of KNN algorithm, steps of working of K-NN, Select value of K, advantage and disadvantage of K-NN algorithm ● Regression: Define Regression analysis, list types of regression analysis, list real world examples of regression analysis **Linear regression:** List types of linear regression, mathematical equation of linear regression, diagram of

Unit– V Unsupervised Machine Learning Models	5.1 Define Unsupervised Learning 5.2 List types of Unsupervised Learning 5.3 Differentiate Supervised and Unsupervised Learning	linear regression line (positive, negative) **Simple linear regression** : (Description, objective, demonstrate example of salary prediction using python) (Steps: Prepare dataset, split data set into training and testing set, visualize training data set and testing data set, i.e. plot it, initialize the training set and fitting it using training set, Predict) list applications of linear regression 4.3.1 Advantage and disadvantage of supervised machine learning
		5.1.1 Introduction of Unsupervised Learning • Brief explanation of unsupervised Machine Learning • Need of unsupervised learning • Working of unsupervised learning • Real world examples of unsupervised Learning • List unsupervised learning algorithms 5.2.1 Types of Unsupervised Learning • Clustering: Definition, list clustering methods, list real world applications/examples (fruits, vegetables, computer devices (input and output etc)), ● Association: Definition, list association methods, list real world applications/examples ● Advantage and Disadvantage of unsupervised learning algorithm 5.3.1 Differentiate Supervised and Unsupervised Learning

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.*

9. SUGGESTED SPECIFICATION TABLE FOR QUESTION PAPER DESIGN

Unit No.	Unit Title	Teaching Hours	Distribution of Theory Marks			
			R Level	U Level	A Level	Total Marks
I	Introduction to machine learning	06	4	4	4	12

II	Python libraries suitable for Machine Learning	09	3	4	7	14
III	Preparing to Model and Preprocessing	09	6	6	4	16
IV	Supervised Machine Learning Models	10	5	7	4	16
V	Unsupervised Machine Learning Models	08	5	4	3	12
Total		**42**	**23**	**25**	**22**	**70**

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.*

10.   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) Enroll in an online Course related to ML based
c) Undertake micro-projects in teams
d) Give a seminar on any relevant topics
e) Collect various sensor data from smart phones and apply machine learning approach

11.   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.

12.   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:** Breast Cancer Prediction: This machine learning project uses a dataset that can help determine the likelihood that a breast tumor is malignant or benign. You can build a classification model for this project.
- **Project idea 2:** Loan Prediction: The idea behind this ML project is to build a model that will classify how much loan the user can take. It is based on the user's marital status, education, number of dependents, and employments. You can build a linear model for this project.
- **Project idea 3:** Stock Price Prediction: This machine learning beginner's project aims to predict the future price of the stock market based on the previous year's data.
- **Project idea 4:** Phishing: Create a Python program that can predict if a URL is legitimate. Teach a computer to recognize phishing web links.
- **Project idea 5:** Titanic survived Prediction: This will be a fun project to build as you will be predicting whether someone would have survived if they were in the titanic ship or not.
- **Project idea 6:** 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 7:** Wine Quality Test: It will use the chemical information of the wine and based on the machine learning model, it will give you the result of wine quality.
- **Project idea 8:** Data from leading music service can be taken to build a better music recommendation system.

13.   SUGGESTED LEARNING RESOURCES

S. No.	Title of Book	Author	Publication with place, year and ISBN
1	Machine Learning	Saikat Dull, S. Chjandramouli	Das, Pearson
2	Machine Learning Using Python	Pradhan Manaranjan, U Dinesh Kumar	Wiley India Pvt. Ltd
3	Introduction to Machine Learning	Jeeva Jose	Khanna Publishers

4	Machine Learning in Action	Peter Harrington	Manning , dreamtech press
5	Machine Learning_ Step-by-Step Guide To Implement Machine Learning Algorithms with Python.	Rudolph Russell	Rudolph Russell Publications
6	Machine Learning with Python Cookbook_ Practical Solutions from Preprocessing to Deep Learning.	Chris Albon	O'Reilly Media, Inc.

14.   SOFTWARE/LEARNING WEBSITES
   a.   https://www.geeksforgeeks.org/machine-learning/
   b.   https://www.tutorialspoint.com/machine_learning_with_python/index.htm
   c.   https://www.javatpoint.com/machine-learning
   d.   https://nptel.ac.in/
   e.   https://www.coursera.org/
   f.   https://scikit-learn.org/
   g.   https://www.w3resource.com/python-exercises/pandas/index.php
   h.   https://machinelearningforkids.co.uk/
   i.   https://monkeylearn.com/machine-learning/
   j.   http://appinventor.mit.edu/explore/ai-with-mit-app-inventor

15.   PO-COMPETENCY-CO MAPPING

Semester II	Introduction to Machine Learning (Course Code: 4350702)						
	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 Manage ment	PO 7 Life-long learning
Competency Students acquire the ability to assess and analyze outcomes produced by machine learning algorithms and models, while enhancing their capacity for critical thinking in addressing practical challenges.							
Course Outcomes CO a) Describe basic concept of machine learning and its applications	2	-	-	2	-	1	1
CO b) Practice Numpy, Pandas, Matplotlib, sklearn library's inbuilt function required to solve machine learning problems	2	1	2	2	-	1	1
CO c) Use Pandas library for data preprocessing	1	2	2	2	-	1	1
CO d) Apply supervised learning algorithms based on dataset characteristics	2	2	2	2	1	1	1

CO e) Apply unsupervised learning algorithms based on dataset characteristics	2	2	2	2	1	1	1

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

16.   COURSE CURRICULUM DEVELOPMENT COMMITTEE

GTU Resource Persons

Sr. No.	Name and Designation	Institute	Contact No.	Email
1	Ms. Manisha P. Mehta HOD	Government Polytechnic Himatnagar	9879578273	manishamehtain@gmail.com
2	Ms. Jasmine J. Karagthala Lecturer	Government Polytechnic for Girls, Ahmedabad	9824799620	jdaftary@gmail.com
3	Miss. Kumundrini B. Prajapati Lecturer	Government Polytechnic Gandhinagar	9974543026	kumundrini13187@gmail.com