



GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Master of Computer Applications

Level: PG

Course / Subject Code: MC03094121

Course / Subject Name : Machine Learning

w. e. f. Academic Year:	2025-26
Semester:	3
Category of the Course:	Elective Group-2

Prerequisite:	Basic math (algebra, probability, and statistics), programming (preferably Python), and an understanding of data structures. Some knowledge of data analysis is helpful but not required.
Rationale:	<ul style="list-style-type: none"> To develop a strong foundation in machine learning concepts, algorithms, and applications. To gain hands-on experience with Python and machine learning tools for practical implementation. To understand and apply supervised and unsupervised learning techniques, feature engineering, and neural networks. To enhance problem-solving and analytical thinking skills for real-world machine learning applications. To acquire the necessary knowledge and expertise for careers in data science, artificial intelligence, and research.

Course Outcome:

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

No	Course Outcomes	RBT Level
01	Differentiate various learning approaches, and to interpret the concepts of supervised learning and unsupervised learning.	UN
02	Apply various techniques for data preprocessing & evaluating and improving performance of model	AP
03	Apply concepts of feature engineering, bayes' theorem and naive bayes' classifier	AP
04	Use different supervised learning algorithms on the given dataset	AP
05	Apply different clustering & association analysis algorithms on the dataset	AP

**Revised Bloom's Taxonomy (RBT): RM: Remember, UN: Understand, AP: Apply, AN: Analyze, EL: Evaluate, CR: Create*

Teaching and Examination Scheme:

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



GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Master of Computer Applications

Level: PG

Course / Subject Code: MC03094121

Course / Subject Name : Machine Learning

Course Content:

Unit No.	Content	No. of Hours	% of Weightage
1.	Introduction to Machine Learning: Human learning versus machine learning, types of machine learning, Applications of machine learning, tools for machine learning, Machine Learning Activities, Data structures for machine learning	4	16%
2.	Preparing to Model, Feature Engineering: Data Pre-processing Selecting a model, training a model, model Representation and interpretability Evaluating performance of a model, improving performance of a model.	6	21%
3.	Modelling and Evaluation: Introduction to feature engineering Feature transformation, feature subset selection, Concept learning through Bayes' theorem	6	21%
4.	Supervised Learning: Classification Model, k -Nearest Neighbour (kNN) algorithm, Decision tree algorithm Simple and multiple Regression algorithms, Assumptions in regression, Problems in regression, Accuracy improvement of linear regression.	7	21%
5.	Unsupervised Learning: Unsupervised Learning versus Supervised Learning, Applications of Unsupervised Learning, Clustering, Finding Pattern using Association Rules	7	21%
Total		30	100

Suggested Specification Table with Marks (Theory):

Distribution of Theory Marks (in %)					
R Level	U Level	A Level	N Level	E Level	C Level
10	30	60	0	0	0

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:

1. **Machine Learning** – Saikat Dutt, Subramanian Chandramouli, Amit Kumar Das, *Pearson Education. (Main Book as per Syllabus Contents)*
2. **Machine Learning** – Tom M. Mitchell, *McGraw Hill*
3. **Machine Learning** – Anuradha Srinivasaraghavan, Vincy Joseph, *Wiley India*
4. **Machine Learning in Action** – Peter Harrington, *DreamTech*
5. **Real-World Machine Learning** – Henrik Brink, Joseph Richards, Mark Fetherolf, *DreamTech*

(c) Open-source software and website:

1. Development Environment & IDEs
- Anaconda Distribution – <https://www.anaconda.com/>



GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Master of Computer Applications

Level: PG

Course / Subject Code: MC03094121

Course / Subject Name : Machine Learning

- Jupyter Notebook – <https://jupyter.org/>
- VS Code (with Python Extension) – <https://code.visualstudio.com/>
- 2. Programming Languages
 - Python – <https://www.python.org/>
- 3. Open-Source Machine Learning Libraries
 - Scikit-Learn – <https://scikit-learn.org/>
 - Keras – <https://keras.io/>
 - PyTorch – <https://pytorch.org/>
- 4. Data Manipulation & Visualization Tools
 - NumPy – <https://numpy.org/>
 - Pandas – <https://pandas.pydata.org/>
 - Matplotlib – <https://matplotlib.org/>
 - Seaborn – <https://seaborn.pydata.org/>
- 5. Datasets & Data Sources
 - Kaggle Datasets – <https://www.kaggle.com/datasets>
 - UCI Machine Learning Repository – <https://archive.ics.uci.edu/ml/index.php>
 - Google Dataset Search – <https://datasetsearch.research.google.com/>

Suggested Project List:

The practical exercises should be performed in Python, preferably using a Python notebook.

1. Write a Python program to perform feature engineering on a Titanic Dataset (<https://www.openml.org/d/42438>).
Perform the following tasks:
 - a. Handling missing values
 - b. Encoding categorical variables
 - c. Scaling numerical features
 - d. Creating new features
2. Write a Python program to perform Principal Component Analysis (PCA) on a given dataset. The objective is to reduce the dimensionality of the data while retaining the maximum variance. Use the dataset provided below and visualize the transformed features in a lower-dimensional space. Use wine dataset from UCI Machine Learning Repository (<https://archive.ics.uci.edu/dataset/109/wine>)
3. Write a Python program to implement a Decision Tree Classifier to predict whether a passenger survived the Titanic disaster based on passenger details. Evaluate the model using precision, recall, and F1-score. (Download dataset from <https://www.openml.org/d/42438>)
4. Implement a Python program that predicts house prices based on square footage using Linear Regression. Train the model on the given dataset and evaluate it using Mean Squared Error (MSE) and R^2 score. (Download dataset <https://www.kaggle.com/datasets/yasserh/housing-prices-dataset/data>)
5. Write a Python program to predict a car's fuel efficiency (MPG - miles per gallon) based on multiple factors such as horsepower, weight, and engine size using Multiple Linear Regression. Evaluate the model using R^2 score and RMSE. (Download dataset from <https://archive.ics.uci.edu/dataset/9/auto+mpg>)
6. Write a Python program to implement the k-Nearest Neighbors (kNN) algorithm to classify different species of iris flowers based on their petal and sepal measurements. Use different values



GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Master of Computer Applications

Level: PG

Course / Subject Code: MC03094121

Course / Subject Name : Machine Learning

of k and evaluate the model using accuracy and a confusion matrix. (Download dataset from <https://archive.ics.uci.edu/dataset/53/iris>)

7. Write a Python program to implement the K-Means clustering algorithm to segment customers based on just two features: Annual Income and Spending Score. Use the K-Means algorithm to group customers into clusters and visualize the results with a simple scatter plot.

Perform the following tasks:

- Load the dataset and explore the data.
- Standardize the features to ensure that both features contribute equally.
- Apply K-Means Clustering to group the customers into 2 clusters.
- Visualize the results using a scatter plot. Highlight the clusters with different colors.

Customer	Annual Income (k\$)	Spending Score (1-100)
1	15	39
2	25	81
3	45	62
4	55	23
5	65	95
6	75	44
7	85	35
8	95	75
9	105	15
10	120	50

8. Write a Python program to implement Hierarchical Clustering (Agglomerative Clustering) to group social media users based on their interaction patterns. Use the following dataset consists of two features:

- Number of Likes per Week
- Number of Comments per Week

Use Agglomerative Clustering to find natural groupings in the data and visualize the dendrogram to determine the optimal number of clusters.

User ID	Likes per Week	Comments per Week
1	5	2
2	20	8
3	30	12
4	40	20
5	60	30
6	70	35
7	80	40
8	90	50
9	100	55
10	110	65



GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Master of Computer Applications

Level: PG

Course / Subject Code: MC03094121

Course / Subject Name : Machine Learning

9. Write a Python program to implement the Apriori Algorithm for market basket analysis. The objective is to identify frequent itemsets and generate association rules from the following dataset of customer transactions in a retail store.

Transaction ID	Items Purchased
1	Milk, Bread
2	Milk, Bread, Butter
3	Bread, Butter
4	Milk, Butter
5	Bread, Butter, Eggs
6	Milk, Bread, Eggs
7	Milk, Bread, Butter, Eggs
8	Bread, Eggs
9	Milk, Eggs
10	Milk, Butter, Eggs

Suggested Activities for Students (Optional):

1. Train multiple algorithms (e.g., Decision Tree, Random Forest, SVM) on the same dataset and compare accuracy.
2. Perform cross-validation and hyperparameter tuning to improve performance.
3. **Mini Project:** Fake News Detection - Apply Logistic Regression or SVM to classify real vs. fake news.
4. Deploy the model on Google Colab, Hugging Face Spaces, or Heroku.

CO- PO Mapping:

Semester 3	Machine Learning							
	POs							
Course Outcomes	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	-	-	-	-	-	-	-
CO2	3	3	3	3	-	-	-	-
CO3	3	3	3	3	-	-	-	-
CO4	3	3	3	3	-	-	-	-
CO5	3	3	3	3	-	-	-	-

Legend: '3' for high, '2' for medium, '1' for low and '-' for no correlation of each CO with PO.

Note: The CO-PO mapping is indicative; the institute/faculty member can change as required.
