Level	Correlation
-	Nil
1	Slightly / Low
2	Moderate / Medium
3	Substantial / High

Assessment Rubrics:

- Quiz / Assignment/ Quiz/ Discussion / Seminar
- Midterm Exam
- Programming Assignments
- Final Exam

Mapping of COs to Assessment Rubrics :

	Internal Exam	Assignment	Project Evaluation	End Semester Examinations
CO 1	✓		✓	✓
CO 2	✓	✓	✓	✓
CO 3	✓		✓	✓
CO 4	✓	✓	✓	✓

12. INTRODUCTION TO MACHINE LEARNING

Discipline	COMPUTER SCIENCE				
Course Code	UK3DSCCSC211				
Course Title	INTRODUCTION TO MACHINE LEARNING				
Type of Course	DSC				
Semester	III				
Academic Level	2				
Course Details	Credit	Lecture per week	Tutorial per week	Practical per week	Total Hours/Week
	4	3 hours	-	2 hours	5 hours
Pre-requisites	PYTHON PROGRAMMING				

Course Summary	This course offers a comprehensive overview of machine learning fundamentals, spanning supervised, unsupervised, and reinforcement learning techniques. Students will gain practical skills in data preprocessing, visualization, and analysis using Python libraries like NumPy, Pandas, and Scikit-learn. Delving into regression and classification algorithms, including linear regression, logistic regression, and decision trees, learners will acquire the ability to interpret and predict data patterns effectively. Advanced topics explore unsupervised learning methods such as clustering and dimensionality reduction, providing students with essential tools for data analysis. Additionally, the flexi module introduces ensemble learning, neural networks, and autoencoders, paving the way for further exploration into artificial intelligence and machine learning applications.

Detailed Syllabus:

Module	Unit	Content	Hrs (L+P)
I	**Introduction to Machine Learning**		**15**
	1	Definition and Importance of Machine Learning:	
	2	Applications across Various Domains	
	3	Supervised Learning:-Definition and Examples, Regression vs. Classification	
	4	Unsupervised Learning:- Definition and Examples,Clustering vs. Dimensionality Reduction	
	5	Reinforcement Learning:- Definition and Examples, Agent-Environment Interaction, Exploration vs. Exploitation Tradeoff	
	6	Understanding data:- numeric variables – mean, median, mode, Measuring spread.	
	7	Introduction to NumPy, Pandas, and Scikit-learn:- Overview of their Features and Capabilities	
II	**Data Preprocessing and Visualization**		**15**
	7	Introduction to Data Preprocessing, Handling Missing Data: Imputation Techniques, Removal Strategies	
	8	Outlier Detection and Treatment: Z-score, IQR, Winsorization	

	9	Feature Scaling and Normalization: Min-Max Scaling, Z-score Normalization, Encoding Categorical Variables: One-Hot Encoding, Label Encoding	
	10	Introduction to Data Visualization:-Overview of Matplotlib and Seaborn Libraries	
	11	Basic Plot Types: Line Plot, Scatter Plot, Bar Plot, Histogram	
	12	Advanced Plot Types: Box Plot, Violin Plot, Heatmap,Multiple Subplots and Figures	
III	**Supervised Learning**		**15**
	12	Regression - Introduction, Types of Regression, Linear Regression, Multiple Linear Regression, Non-Linear Regression (Polynomial Regression)	
	13	Classification –Introduction, Logistic Regression, Decision Trees, Naïve Bayes Classification, Support Vector Machines:-Intuition and Optimization, K-Nearest Neighbours, Random Forest.	
IV	**Unsupervised Learning**		**15**
	19	Categorization of Major Clustering Methods - Partitioning Methods - K-means, K-medoids. Hierarchical Methods - Agglomerative Clustering, Density-based Methods – DBSCAN.	
	20	Principal Component Analysis (PCA):Understanding the PCA algorithm, Calculating principal components and eigenvalues, Reducing dimensionality using PCA, Interpretation of principal components, PCA implementation and applications	
	21	t-Distributed Stochastic Neighbour Embedding (t-SNE):Introduction to t-SNE algorithm, Similarities and differences between PCA and t-SNE	
V		**Flexi Module: Not included for end semester exams**	**15**
	26	Ensemble Learning: Understanding ensemble methods like bagging and boosting.	
	27	Introduction to Neural Networks: Basics of artificial neural networks (ANN), deep learning frameworks (e.g., TensorFlow).	
	28	Introduction to autoencoders, Encoding and decoding processes in autoencoders, Training autoencoders with backpropagation	

		Denoising autoencoders and variational autoencoders, Applications of autoencoders in unsupervised learning and feature learning	

References

- Introduction to Machine Learning with Python" by Andreas C. Müller & Sarah Guido
- Python Machine Learning" by Sebastian Raschka and Vahid Mirjalili
- Pattern Recognition and Machine Learning" by Christopher M. Bishop
- Machine Learning: A Probabilistic Perspective" by Kevin P. Murphy

Lab Exercises

1. Prepare a dataset of customer having the features date, price, product_id, quantity_purchased, serial_no, user_id,user_type, user_class, purchase_week and visualise the data with
    a. Plot diagram for Price Trends for Particular User, Price Trends for Particular User Over Time
    b. Create box plot Quantity and Week value distribution having parameters of quantity_purchased','purchase_week'
2. **Task:** Conduct exploratory data analysis (EDA) on a designated dataset utilizing NumPy and Pandas.

    Description: Select a dataset of choice (e.g., Iris dataset, Titanic dataset, etc.), and load it into a Pandas DataFrame. Leverage NumPy for numerical computations. Compute the mean, median, and mode of numeric variables within the dataset. Assess the data's spread through techniques such as standard deviation, variance, and range calculations. Employ histograms and box plots to visually represent the distribution of numeric variables. Provide insights and interpretations based on the outcomes of the EDA.

3. Task: Utilize Python programming to preprocess the "Titanic" dataset.

    Description:Implement data preprocessing steps to handle missing data by employing imputation techniques or removal strategies. Detects and treats outliers using Z-score, IQR, or Winsorization methods.

4. Task: Utilize Python programming feature scaling and normalization on the "Titanic" dataset.

    Description:Perform feature scaling and normalization on relevant features, and encode categorical variables using one-hot encoding or label encoding schemes. Utilize Matplotlib and Seaborn libraries to visualize the preprocessed dataset, creating basic plots such as Line Plot, Scatter Plot, Bar Plot, and Histogram, as well as advanced plots like Box Plot, Violin Plot, and Heatmap

5. Task: Utilize Python programming visualize on the "Titanic" dataset.

   Description: Utilize Matplotlib and Seaborn libraries to visualize the preprocessed dataset, creating basic plots such as Line Plot, Scatter Plot, Bar Plot, and Histogram, as well as advanced plots like Box Plot, Violin Plot, and Heatmap

6. Task: Train regression models on the "Boston Housing" dataset to predict house prices based on various features.

   Description: Utilize the "Boston Housing" dataset available in the scikit-learn library. Train a linear regression model to predict house prices using features such as area, number of bedrooms, and location. Additionally, implement multiple linear regression to predict sales revenue based on advertising spending across different channels. Explore the application of non-linear regression techniques like polynomial regression to capture more complex data patterns in the dataset. Visualize the regression results to understand the relationships between predictors and the target variable.

7. Task: Employ classification techniques on the "Titanic" dataset to predict survival outcomes based on passenger features.

   Description: Use the Titanic dataset to train a logistic regression model to predict survival outcomes based on passenger features.

8. Task: Employ classification techniques on the "MNIST dataset"

   Description: Implement a support vector machine classifier to classify handwritten digits using the MNIST dataset.

9. Task: Employ classification techniques on the "iris dataset"

   Description: Experiment with k-nearest neighbors and random forest classifiers on iris dataset and MNIST dataset and compare their performance.

10. Task: Apply K-means clustering on the "Online Retail" dataset to segment customers based on their purchasing behavior.

    Description: Utilize the "Online Retail" dataset, which contains information about customer transactions, including items purchased and their quantities. Implement K-means clustering to segment customers into distinct groups based on their purchasing patterns. Analyze the characteristics of each cluster to understand the preferences and behaviors of different customer segments. Identify potential marketing strategies tailored to each segment to enhance customer engagement and satisfaction.

    Dataset: The "Online Retail" dataset is available from the UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets/Online+Retail).

11. Task: Employ principal component analysis (PCA) on the "Labeled Faces in the Wild" dataset to reduce the dimensionality of facial images.

Description: Utilize the "Labeled Faces in the Wild" dataset, which contains a collection of facial images belonging to various individuals. Implement PCA to reduce the high-dimensional feature space of facial images while preserving essential information. Visualize the principal components to gain insights into the underlying structure of the data. Reconstruct the facial images using a reduced number of dimensions to observe the effectiveness of dimensionality reduction. Analyze the reconstructed images to understand the impact of dimensionality reduction on facial image quality and interpretability.

Dataset: The "Labeled Faces in the Wild" dataset is available from the scikit-learn library (https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_lfw_people.html).

Course Outcomes

No.	Upon completion of the course the graduate will be able to	Cognitive Level	PSO addressed
CO1	Summarise the definition and significance of machine learning	U	PSO – 1, 3
CO2	Cite the principles underlying supervised and unsupervised learning methods.	U	PSO – 1, 2, 3
CO3	Apply data preprocessing procedures using Python libraries to cleanse and organise datasets efficiently,	Ap	PSO – 1, 2, 3
CO4	Illustrate the effectiveness of machine learning models.	Ap	PSO – 1, 2, 3

R-Remember, U-Understand, Ap-Apply, An-Analyse, E-Evaluate, C-Create

Note: 1 or 2 COs/module

Name of the Course: INTRODUCTION TO MACHINE LEARNING

Credits: 3:0:1 (Lecture: Tutorial: Practical)

CO No.	CO	PO/PSO	Cognitive Level	Knowledge Category	Lecture(L)/ Tutorial(T)	Practical (P)
1	Summarise the definition and significance of machine learning	PO- 1, 2, 3, 4, 6 PSO – 1, 3	U	F, C, P	L	P
2	Cite the principles underlying supervised and unsupervised learning methods.	PO- 1, 2, 3, 4, 6 PSO – 1, 2, 3	U	F, C, P	L	P
3	Apply data preprocessing	PO- 1, 2,	Ap	F, C, P	L	P

	procedures using Python libraries to cleanse and organise datasets efficiently,	3, 4, 8 PSO – 1, 2, 3,3				
4	Illustrate the effectiveness of machine learning models.	PO- 1, 2, 3, 4, 6 PSO – 1, 2, 3	Ap	F, C,P	L	P

F-Factual, C- Conceptual, P-Procedural, M-Metacognitive

Mapping of COs with PSOs and POs:

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PSO1**	**PSO2**	**PSO3**	**PSO4**
CO1	**2**	**2**	**1**	**1**	**-**	**2**	**3-**	**-**	**2**	**2**	**2**	**-**
CO2	**3**	**3**	**2**	**1**	**-**	**3**	**3-**	**-**	**3**	**3**	**2**	**-**
CO3	**3**	**3**	**2**	**2**	**-**	**3**	**3**	**3**	**3**	**3**	**2**	**-**
CO4	**3**	**3**	**2**	**2**	**-**	**2**	**3**	**-**	**3**	**3**	**2**	**-**

Correlation Levels:

Level	Correlation
-	Nil
1	Slightly / Low
2	Moderate / Medium
3	Substantial / High

Assessment Rubrics:

- Quiz / Assignment/ Quiz/ Discussion / Seminar
- Midterm Exam
- Programming Assignments
- Final Exam

Mapping of COs to Assessment Rubrics:

	Internal Exam	Assignment	Lab Assessment	End Semester Examinations
CO1	✓		✓	✓
CO2	✓	✓	✓	✓
CO3	✓		✓	✓
CO4	✓	✓	✓	✓

13. MEDICAL IMAGING AND ANALYSIS

Discipline	Computer Science
Course Code	UK3DSCCSC212
Course Title	MEDICAL IMAGING AND ANALYSIS