FACULTY OF COMPUTER APPLICATIONS
B.Sc. (IT)

- **Sem.** :6
- **Subject Code** :05BS0604
- **Subject** : Machine Learning
- **Course Objectives** :

1. To get the idea about concepts of machine learning and Python libraries
2. To learn the concepts of feature engineering
3. Understanding various algorithms of Classification for supervised machine learning
4. Understanding various algorithms of Regression for supervised machine learning
5. Understanding various algorithms of clustering and association rule mining for unsupervised machine learning

- **Prerequisites** :

Unit No	Topics Covered	No of lectures required
1	Introduction What is machine learning? key terminology, Key tasks of machine learning, How to choose the right algorithm? Steps in developing machine learning applications Why python? Scikit-learn, Essential libraries and tools, Python 2 versus Python 3, A first application – classifying iris species Preprocessing and Scaling Different Kinds of Preprocessing, Applying Data Transformations, Scaling Training and Test Data the Same Way, The Effect of Preprocessing on Supervised Learning Dimensionality Reduction, Feature Extraction, and Manifold Learning	**9**

	Principal Component Analysis (PCA), Non-Negative Matrix Factorization (NMF), Manifold Learning with t-SNE	
2	Representing Data and Engineering Features Categorical Variables,One-Hot-Encoding, Binning, Discretization, Linear Models, and Trees, Interactions and Polynomials ,Univariate Nonlinear Transformations , Automatic Feature Selection Cross-Validation Cross-Validation in scikit-learn, Benefits of Cross-Validation, Stratified k-Fold Cross-Validation and Other Strategies, Evaluation Metrics and Scoring Keep the End Goal in Mind, Metrics for Binary Classification, Metrics for Multiclass Classification, Regression Metrics, Using Evaluation Metrics in Model Selection	**12**
3	Classifying with K- nearest neighbors, Splitting datasets one feature at a time – decision tree, Classifying with probability theory – Naïve Bayes, Logistic regression, Support vector machines	**10**
4	Predicting numerical values – Regression, Tree based regression	**6**
5	Clustering k-means clustering, Agglomerative Clustering ,DBSCAN ,Comparing and Evaluating Clustering Algorithms Association analysis with Apriori algorithm	**8**

Course Outcomes :(Students will be able)

1. To understand applications of ML in real life and usage of various Python libraries for ML
2. To implement preprocessing steps on data to make it ready for analysis
3. To apply Classification algorithms for supervised learning
4. To apply Regression algorithms for supervised learning
5. To apply clustering algorithms and association rule mining algorithms for real life problems

FACULTY OF COMPUTER APPLICATIONS
B.Sc. (IT)

Course Outcomes – Program Outcomes Mapping Table :

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PSO 1	PSO2	PSO 3
CO1	H		L			M			L		L
CO2	H	L	M	L		M	L	L	M	L	M
CO3	H	M	H	L	L	H	M	H	M	L	H
CO4	H	M	H	L	L	H	M	H	M	L	H
CO5	H	M	H	L	L	H	M	H	M	L	H

Text Book :

1. Introduction to Machine Learning with Python – A guide for data scientists, Andreas C. Müller & Sarah Guido, O'Reilly Media,1st Edition
2. Machine Learning in Action, Peter Harrington, Manning, 1st Edition

Reference Books :

1. Machine Learning, SaikatDuttet al. , Pearson Education, 2019, 3rd Edition
2. Practical Machine Learning with Python, Dipanjan Sarkar et al., Apress, 2018, 1st Edition
3. Machine Learning,Tom M Mitchell, McGraw Hill, 1st Edition

Web References :

1. **www.edx.com**
2. **www.coursera.com**
3. **www.kaggle.com**

App References :

1. Learn Machine Learning - ML Tutorials & Programs
2. Data Science 101 - Machine Learning Tutorials

FACULTY OF COMPUTER APPLICATIONS

B.Sc. (IT)

Syllabus Coverage from text /reference book & web/app reference:

Unit #	Chapter Numbers
1	Book 1 – Chapter 1,3 Book 2 – Chapter 1
2	Book 1 –Chapter 4,5
3	Book 2 – Chapter 2,3,4,5,6
4	Book 2 – Chapter 8,9
5	Book 1 – Chapter 3 Book 2 – Chapter 11

FACULTY OF COMPUTER APPLICATIONS
B.Sc. (IT)

PRACTICALS

<table>
<tr><th>Unit No</th><th>List of Practical</th></tr>
<tr><td>1&2</td><td>Data Identification and Feature Engineering :
Practical – 1 :
Download Auto-MPG data set (https://www.kaggle.com/uciml/autompg-dataset) and perform following operations
i. Write program to read dataset (Text, CSV,JSON,XML)
ii. Which of the attributes are numeric and which are categorical?
iii. Performing Data Cleaning
● Handling Missing Data ● Removing Null data
iv. Rescaling Data
v. Encoding Data
vi. Feature Selection and Dimensionality Reduction
● Implement Principle Component Analysis,
Pracitcal – 2 : Download Airline data set (https://www.kaggle.com/open-flights/airline-database) and perform all the above operations.</td></tr>
<tr><td>3</td><td>Supervised Learning–Classification:
Practical–1 :Write a python code to apply NaiveBayesian algorithm to classify that whether a person can buy computer or not based on given test data :
<table>
<tr><th>Age</th><th>Income</th><th>Student</th><th>Creditrating</th><th>Buyscomputer</th></tr>
<tr><td>Youth</td><td>High</td><td>No</td><td>Fair</td><td>No</td></tr>
<tr><td>Youth</td><td>High</td><td>No</td><td>Excellent</td><td>No</td></tr>
<tr><td>Middle</td><td>High</td><td>No</td><td>Fair</td><td>Yes</td></tr>
<tr><td>Senior</td><td>Medium</td><td>No</td><td>Fair</td><td>Yes</td></tr>
<tr><td>Senior</td><td>Low</td><td>Yes</td><td>Fair</td><td>Yes</td></tr>
<tr><td>Middle</td><td>Low</td><td>Yes</td><td>Excellent</td><td>No</td></tr>
<tr><td>Senior</td><td>Low</td><td>Yes</td><td>Excellent</td><td>Yes</td></tr>
<tr><td>Youth</td><td>Medium</td><td>No</td><td>Fair</td><td>No</td></tr>
<tr><td>Youth</td><td>Low</td><td>Yes</td><td>Fair</td><td>Yes</td></tr>
<tr><td>Senior</td><td>Medium</td><td>Yes</td><td>Fair</td><td>Yes</td></tr>
<tr><td>Youth</td><td>Medium</td><td>Yes</td><td>Excellent</td><td>Yes</td></tr>
<tr><td>Middle</td><td>Medium</td><td>No</td><td>Excellent</td><td>Yes</td></tr>
</table>
</td></tr>
</table>

Middle	High	Yes	Fair	Yes
Senior	Medium	No	Excellent	No

Practical–2 :Write a python code to implement **decisiontree**
For below given dataset. Identify Job offered or not.

CGPA	Communication	Aptitude	Programming_skill	Job_offered
High	Good	High	Good	Yes
Medium	Good	High	Good	Yes
Low	Bad	Low	Good	No
Low	Good	Low	Bad	No
High	Good	High	Bad	Yes
High	Good	High	Good	Yes
Medium	Bad	Low	Bad	No
Medium	Bad	Low	Good	No
High	Bad	High	Good	Yes
Medium	Good	High	Good	Yes
Low	Bad	High	Bad	No
Low	Bad	High	Bad	No
Medium	Good	High	Bad	Yes
Low	Good	Low	Good	No
High	Bad	Low	Bad	No
Medium	Bad	High	Good	No
High	Bad	Low	Bad	No
Medium	Good	High	Bad	Yes

Practical–3:Write a python code to implement **K-nearest neighbourhood** program for the given dataset. (for above both the data sets)

Practical – 4 :Implement supervised machine learning algorithm(Classification–K Nearest Neighbour) in python to classify breast tumor data into malignant breast tumor or benign breast tumor (use breast tumor dataset) and obtain its accuracy level.

Practical – 5 :Implement supervised machine learning algorithm(Classification–K Nearest Neighbour) in python to classify iris data into setosa,

<table>
<tr><td></td><td>virginica, versicolor using iris dataset and obtain its accuracy level.

Practical – 6 :Build a classification model in python that classifies if a student gets admission in a course or not given his last two examination scores for the dataset available at https://docs.google.com/spreadsheets/d/1g0mjTUZ9Ado5prXA 1UnAvNjmdzTrV0TzkFkIoU-Lpbk/edit?usp=sharing

Practical – 7 :Implement supervised machine learning algorithm(Classification–Support Vector Machine) in python to classify breast tumor data into malignant breast tumor or benign breast tumor (use breast tumor dataset) and obtain its accuracy level.

Practical–8:Write a python program to build an email spam classifier using support vector machines for the Spam base dataset from UCI machine learning repository.

Practical – 9 : Implement supervised machine learning algorithm(Classification-Naïve Bayes algorithm)in python on Pima Indians Diabetes dataset and obtain its accuracy level.</td></tr>
<tr><td>4</td><td>Supervised Learning–Regression:
Pracitcal – 1 :Predict the CPU time if the DiskIO = 40 on the basis of following data :
<table>
<tr><td>DiskIO</td><td>CPUTime</td></tr>
<tr><td>14</td><td>2</td></tr>
<tr><td>16</td><td>5</td></tr>
<tr><td>27</td><td>7</td></tr>
<tr><td>42</td><td>9</td></tr>
<tr><td>39</td><td>10</td></tr>
<tr><td>50</td><td>13</td></tr>
<tr><td>83</td><td>20</td></tr>
</table>

Practical – 2 :Write a python code to predict profit of hotel chain given the population of the area (city) using the data at https://docs.google.com/spreadsheets/d/1Ks20skBgEef</td></tr>
</table>

	HFU36s FqVzozoFtz2EZE2rxBIgXOrUg/edit?usp=sharing. **Practical–3**:Writeapythoncodetopredictsalaryonthe basis of experience in years using the data at https://github.com/tarunlnmiit/machine_learning/blob/master/SimpleLinearRegression.csv **Practical – 4** : Write a python code to predict the price of house given square feet and number of bed rooms in the house for the dataset available at https://docs.google.com/spreadsheets/d/1DHVK7gKo4TSyj7m FLwofHamj1Sl4SOZma2q51w1ZvyE/edit?usp=sharing **Practical–5**:Build a logistic regression model to classify flower type based on the dataset of iris flower. **Practical–6**:Buildamultivariatelogisticregressionmodelto classify glass type of glass givendifferent glass mixture features using the Glass Identification Dataset from UCIMachine Learning Repository.
5	**Unsupervised Learning–Clustering and Association Rule Mining:** **Practical – 1** :Implement unsupervised machine learning algorithm(Clustering–KMeans) in python on Titanic dataset to cluster data (use Titanic dataset) by removing the class label. **Practical – 2 :**Implement unsupervised machine learning algorithm(Clustering–KMeans)inpythononBreastTumourdatasettocluster data(useBreastTumourdataset)by removing the class label. **Practical–3 :** Implement unsupervised machine learning algorithm (Clustering – Hierarchical) in python on Titanic dataset to cluster data (use Titanic dataset). **Practical – 4 :**Implement unsupervised machine

learning algorithm (Clustering – Hierarchical) in python on Breast Tumor dataset to cluster data(use Breast Tumor dataset)by removing the class label.

Practical – 5 : Implement Apriori algorithm in python to find rules which explain association between different products for given transactions at a retail store. (The data is available at https://drive.google.com/file/d/1NUXoptUlHY8z4KcFKpFA6sQN 5KnWzk3p/view?usp=sharing)

Practical–6:GeneratingAssociationruleminingfor following data set.

1	Red,White,Green
2	White,Orange
3	White,Blue
4	Red,White,Orange
5	Red,Blue
6	White,Blue
7	White,Orange
8	Red,White,Blue,Green
9	Red,White,Blue
10	Yellow