

FACULTY OF COMPUTER APPLICATIONS
B.Sc. (DS)

- **Sem.** : 4
- **Subject Code** : 05DS0402
- **Subject** : Machine Learning I
- **Course Objectives** :
 1. To get the idea about concepts of machine learning and Python libraries
 2. To learn the concepts of feature engineering and data representation
 3. Understanding various model improvement and modeling techniques
 4. Understanding various algorithms of classification for supervised machine learning
 5. Understanding various algorithms of regression for supervised machine learning
- **Prerequisites** : Basics of python and statistics

Unit No	Topics Covered	No of lectures required
1	Introduction to Machine Learning --Why Machine Learning? --Problems Machine Learning Can Solve --Knowing Your Task and Knowing Your Data --Why Python? --scikit-learn --Installing scikit-learn --Essential Libraries and Tools --Jupyter Notebook --NumPy --SciPy --matplotlib --pandas --mlearn --Python 2 Versus Python 3 --Getting started with NumPy library	8
2	Representing Data and Engineering Features --Dealing with missing data -- Eliminating samples or features with missing values	10

FACULTY OF COMPUTER APPLICATIONS
B.Sc. (DS)

	<ul style="list-style-type: none"> -- Imputing missing values -- Understanding the scikit-learn estimator API -- Handling categorical data -- Mapping ordinal features -- Encoding class labels -- Performing one-hot encoding on nominal features -- Partitioning a dataset in training and test sets -- Bringing features onto the same scale -- Selecting meaningful features -- Sparse solutions with L1 regularization -- Sequential feature selection algorithms 	
3	<p>Model Evaluation and Improvement</p> <ul style="list-style-type: none"> -- Cross-Validation -- Cross-Validation in scikit-learn -- Benefits of Cross-Validation -- Stratified k-Fold Cross-Validation and Other Strategies -- Grid Search -- Simple Grid Search -- The Danger of Overfitting the Parameters and the Validation Set -- Grid Search with Cross-Validation -- 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 	10
4	<p>Classification</p> <ul style="list-style-type: none"> -- Classification model, Classification learning steps -- Common Algorithms <ul style="list-style-type: none"> - Baye's Theorem / Naïve Baye's - k-Nearest Neighbour (KNN) - Decision Tree - Random Forest Model - Support Vector Machine(SVM) - Ensemble Technique 	10
5	<p>Regression</p> <ul style="list-style-type: none"> -- Simple Linear Regression -- Multiple Linear Regression -- Polynomial Regression Model -- Lasso Regression -- Ridge Regression -- Logistic Regression 	12

FACULTY OF COMPUTER APPLICATIONS
B.Sc. (DS)

	--Using CART for regression --Maximum Likelihood Estimation (Least Square)	
--	---	--

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 model improvement and evaluation techniques
4. To apply classification algorithms for supervised learning
5. To apply regression algorithms

Course Outcomes – Program Outcomes Mapping Table:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PSO 1	PSO 2	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
3. Python machine learning - 3rd Edition by Sebastian Raschka & Vahid Mirjalili, Packt
4. “Machine Learning”, SaikatDutt et al. , Pearson Education, 2019, Third Edition

Reference Books :

1. Machine Learning, SaikatDutt et 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

FACULTY OF COMPUTER APPLICATIONS
B.Sc. (DS)**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

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

Unit #	Chapter Numbers
1	Book 1 - Chapter 1 Book 3 - Chapter 1
2	Book 2 - Chapter 4
3	Book 3 - Chapter 2 Book 2 - Chapter 4
4	Book 1 - Chapter 2 Book 3 - Chapter 2 Book 2 - Chapter 7
5	Book 1 - Chapter 2 Book 4 - Chapter 8

FACULTY OF COMPUTER APPLICATIONS
B.Sc. (DS)

PRACTICALS

Unit No	List of Practical																																																																						
<p align="center">2&3</p>	<p>Data Identification and Feature Engineering :</p> <p>Practical – 1 :</p> <p>Download Auto-MPG data set (https://www.kaggle.com/uciml/autompg-dataset) and perform following operations</p> <p>i. Write program to read dataset (Text,CSV,JSON,XML)</p> <p>ii. Which of the attributes are numeric and which are categorical?</p> <p>iii. Performing Data Cleaning</p> <ul style="list-style-type: none"> • Handling Missing Data • Removing Null data <p>iv. Rescaling Data</p> <p>v. Encoding Data</p> <p>vi. Feature Selection and Dimensionality Reduction</p> <ul style="list-style-type: none"> • Implement Principle Component Analysis, <p>Practical – 2 : Download Airline data set (https://www.kaggle.com/open-flights/airline-database) and perform all the above operations.</p>																																																																						
<p align="center">4</p>	<p>Classification:</p> <p>Practical-1</p> <p>:Write a python code to apply Naive Bayesian algorithm to classify that whether a person can buy a computer or not based on given test data :</p> <table border="1" data-bbox="467 1339 1274 1885"> <thead> <tr> <th>Age</th> <th>Income</th> <th>Student</th> <th>Credit rating</th> <th>Buy computer</th> </tr> </thead> <tbody> <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> <tr><td>Middle</td><td>High</td><td>Yes</td><td>Fair</td><td>Yes</td></tr> </tbody> </table>	Age	Income	Student	Credit rating	Buy computer	Youth	High	No	Fair	No	Youth	High	No	Excellent	No	Middle	High	No	Fair	Yes	Senior	Medium	No	Fair	Yes	Senior	Low	Yes	Fair	Yes	Middle	Low	Yes	Excellent	No	Senior	Low	Yes	Excellent	Yes	Youth	Medium	No	Fair	No	Youth	Low	Yes	Fair	Yes	Senior	Medium	Yes	Fair	Yes	Youth	Medium	Yes	Excellent	Yes	Middle	Medium	No	Excellent	Yes	Middle	High	Yes	Fair	Yes
Age	Income	Student	Credit rating	Buy computer																																																																			
Youth	High	No	Fair	No																																																																			
Youth	High	No	Excellent	No																																																																			
Middle	High	No	Fair	Yes																																																																			
Senior	Medium	No	Fair	Yes																																																																			
Senior	Low	Yes	Fair	Yes																																																																			
Middle	Low	Yes	Excellent	No																																																																			
Senior	Low	Yes	Excellent	Yes																																																																			
Youth	Medium	No	Fair	No																																																																			
Youth	Low	Yes	Fair	Yes																																																																			
Senior	Medium	Yes	Fair	Yes																																																																			
Youth	Medium	Yes	Excellent	Yes																																																																			
Middle	Medium	No	Excellent	Yes																																																																			
Middle	High	Yes	Fair	Yes																																																																			

FACULTY OF COMPUTER APPLICATIONS
B.Sc. (DS)

Senior	Medium	No	Excellent	No
--------	--------	----	-----------	----

Practical-2
:Writepythoncodetoimplement**decisiontree** forbelowgiven dataset.IdentifyJob offeredor not.

CGPA	Communica- tion	Aptitude	Program ming_ski ll	Job_offe red
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:Writepythoncodetoimplement**K-nearest neighbourhood**program for the given dataset. (for above both the data sets)

Practical - 4 :Implement supervised machine learning algorithm(Classification-KNearestNeighbourhood)inpython to classify breast tumour data into malignant breast tumour or benign breast tumour (use breast tumour dataset) and obtain its accuracy level.

Practical - 5 :Implement supervised machine learning algorithm(Classification-KNearestNeighbourhood)inpython to classify iris data into setosa, virginica, versicolor using iris dataset and

FACULTY OF COMPUTER APPLICATIONS
B.Sc. (DS)

	<p>obtain its accuracy level.</p> <p>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/1g0mjTUZ9A-do5prXA-1UnAvNjmdzTrV0TzkFkIoU-Lpbk/edit?usp=sharing</p> <p>Practical – 7 :Implement supervised machine learning algorithm(Classification-SupportVectorMachine)inpythonto classify breast tumour data into malignant breast tumourorbenign breast tumour (use breast tumour dataset) and obtain its accuracy level.</p> <p>Practical-8:Writeapythonprogramtobuildanemailspam classifier using support vector machines for the Spam base dataset from UCI machine learning repository.</p> <p>Practical – 9 : Implement supervised machine learning algorithm(Classification-NaïveBayesalgorithm)inpythonon Pima Indians Diabetes dataset and obtain its accuracy level.</p>																
<p align="center">5</p>	<p>Regression:</p> <p>Pracitcal – 1 :Predict the CPU time if the DiskIO = 40 on thebasis of following data :</p> <table border="1" data-bbox="467 1354 925 1654"> <thead> <tr> <th>DiskIO</th> <th>CPUTime</th> </tr> </thead> <tbody> <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> </tbody> </table> <p>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/1Ks20skBgEefHFU36s</p>	DiskIO	CPUTime	14	2	16	5	27	7	42	9	39	10	50	13	83	20
DiskIO	CPUTime																
14	2																
16	5																
27	7																
42	9																
39	10																
50	13																
83	20																

FACULTY OF COMPUTER APPLICATIONS
B.Sc. (DS)

[FqVzozoFtz2EZE2rxBIgXOrUg/edit?usp=sharing](https://github.com/tarunlnmiit/machine_learning/blob/master/SimpleLinearRegression.csv).

Practical-3: Write a python code to predict salary on the 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/1DHVK7gKo4TSyj7mFLwofHamj1SI4SOZma2q51w1ZvyE/edit?usp=sharing>

Practical-5: Build a logistic regression model to classify flower type based on the dataset of iris flower.

Practical-

6: Build a multivariate logistic regression model to classify glass type of glass given different glass mixture features using the Glass Identification Dataset from UCIMachine Learning Repository.