AML311	PYTHON AND MACHINE LEARNING LAB	CATEGORY	L	T	P	Credit	Year of Introduction
		PCC	0	0	3	2	2022

Preamble: This course enables the learners to get hands-on experience in most popular supervised learning algorithms (such as linear regression, logistic regression, decision trees, Bayesian learning and Naive Bayes algorithm) and unsupervised learning algorithms (such as basic clustering algorithms). This helps the learners to understand the process of knowledge inference from raw data through dataset preprocessing and analysis.

Prerequisite: Fundamentals of Programming, Python programming fundamentals, Machine learning.

Course Outcomes: After the completion of the course the student will be able to

CO#	Course Outcomes
CO1	Develop applications in Python programming. **(Cognitive Knowledge Level: Apply)**
CO2	Implement machine learning algorithms using packages and libraries in Python for various applications.**(Cognitive Knowledge Level: Apply)**
CO3	Implement python programs for supervised learning methods through Neural network, Regression and classification.**(Cognitive Knowledge Level: Apply)**
CO4	Implement clustering algorithms.**(Cognitive Knowledge Level: Apply)**
CO5	Apply dimensionality reduction as a dataset preprocessing step. **(Cognitive Knowledge Level: Apply)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓	✓				✓				✓
CO2	✓	✓	✓	✓	✓			✓				✓
CO3	✓	✓	✓	✓	✓	✓		✓				✓
CO4	✓	✓	✓	✓	✓	✓		✓				✓
CO5	✓	✓	✓	✓	✓	✓		✓				✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and teamwork
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Lifelong learning

Assessment Pattern

Bloom's Category	Continuous Assessment Test (Internal Exam) Marks in percentage	End Semester Examination Marks in percentage
Remember	**20**	**20**
Understand	**20**	**20**
Apply	**60**	**60**
Analyze		
Evaluate		
Create		

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	75	75	3 hours

Continuous Internal Evaluation Pattern:

Attendance : **15 marks**
Continuous Evaluation in Lab : **30 marks**
Continuous Assessment Test : **15 marks**
Viva voce : **15 marks**

Internal Examination Pattern:

The Internal examination shall be conducted for 100 marks, which will be converted to out of 15 while calculating Internal Evaluation marks. The marks will be distributed as, Algorithm - 30 marks, Program - 20 marks, Output - 20 marks and Viva - 30 marks.

End Semester Examination Pattern:

The End Semester Examination shall be conducted for 100 marks, which will be converted to out of 75 while calculating External Evaluation marks. The marks will be distributed as, Algorithm - 30 marks, Program - 20 marks, Output - 20 marks and Viva- 30 marks.

Operating System to use in lab : Ubuntu , Windows

Fair Lab Record:

All the students attending the Python and machine learning lab should have a Fair Record. Every experiment conducted in the lab should be noted in the fair record. For every experiment, in the fair record, the right-hand page should contain experiment heading, experiment number, date of experiment, aim of the experiment, procedure/algorithm followed, other such details of the experiment and final result. The left-hand page should contain a print out of the respective code with sample input and corresponding output obtained. All the experiments noted in the fair record should be verified by the faculty regularly. The fair record, properly certified by the faculty, should be produced during the time of End Semester Examination for the verification by the examiners.

SYLLABUS

***Mandatory**

1. Introduction to Python Programming.
2. Familiarization of basic Python Libraries such as Sklearn, Numpy, Pandas and Matplotlib.*
3. Write a Python program to find the union and intersection of two lists.
4. Design a Python program to count the occurrences of each word in a given sentence.
5. Write a Python program to multiply two matrices.*
6. Write a Python program to find the most frequent words in a text file.*
7. Implement and demonstrate Single, Multi variable and Polynomial Regression for a given set of training data stored in a .CSV file and evaluate the accuracy.*
8. Implement a Python program to perform logistic regression on a dataset.
9. Write a Python program to implement Naive Bayes classifier and calculate the accuracy, precision, and recall for your data set.*
10. Write a Python program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.*
11. Assuming a set of data that need to be classified, use a Support Vector Machine classifier to perform this task and evaluate the accuracy.*
12. Implement K-Nearest Neighbor algorithm to classify any dataset.
13. Implement K-Means Clustering using any given dataset.*
14. Build an Artificial Neural Network using Backpropagation algorithm and test the same with appropriate dataset.*
15. Implement dimensionality reduction using PCA.

(Use socially relevant dataset as far as possible)

Python and Machine Learning Lab-Practice Questions

1. Review of Python programming: Programs using matplotlib / plotly / bokeh / seaborn.
2. Write a program to find words which are greater than a given length k.
3. Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using the standard Heart Disease Dataset.
4. Implementation of Random Forest Classification in Python.
5. Write a program to implement k-Nearest Neighbor algorithm to classify the iris data set. Print both correct and wrong predictions.
6. Program to implement text classification using Support vector machine.
7. The probability that it is Friday and that a student is absent is 3%. Since there are 5 school days in a week, the probability that it is Friday is 20%. What is the probability that a student is absent given that today is Friday? Apply Bayes' rule in python to get the result. (Ans: 15%)

8. Implement Naïve Bayes theorem to classify the English text.
9. Program to implement Mean-Shift algorithm in python.
10. Implement Agglomerative Hierarchical Clustering.
11. Apply K-Means clustering to evaluate Student's performance. The results expected show the profile of a student with criteria for excellent performance, standard performance, and underperformance.

Reference Books:

1. Aurelien Geron, "Hands–On Machine Learning with Scikit–Learn and TensorFlow", O'Relily.
2. Christopher Bishop, "Pattern Recognition and Machine Learning", Springer, 2007.
3. Ian Good fellow and Yoshua Bengio and Aaron Courville, "Deep Learning", MIT Press, 2016.