SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited with "A" Grade by NAAC
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai - 600 119.
Phone: 044 - 2450 3150 / 51 / 52 / 54 / 55 Fax: 044 - 2450 2344
www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NEW COURSE INTRODUCED-2018-2019

2018-EVEN

SCS4305	PYTHON PROGRAMMING AND MACHINE LEARNING LAB	L	T	P	Credits	Total Marks
		0	0	4	2	100

Course Outcomes:

CO1 - Design and develop Python Program for simple Machine Learning tasks.

CO2 - Explain complexity, implementation and limitations of Machine Learning.

CO3 - Formulate the model for various Use cases by employing Supervised, Semi-Supervised and Unsupervised Machine learning algorithms.

CO4 - Interpretation of data by analyzing the Data set.

CO5 - Performance study on the Trained Model with Test and Validation data.

CO6 - Evaluate and assess the performance of Machine Learning systems.

List of Experiments:

1. **Analysis and implementation using Python /Jupiter Notebook/Colab**   **(CO1)**

    i. Compute the distance travelled by the robot from current position after a sequence of movement and original point.
    ii. Creation of scatter plot using sepal length and petal width to separate the Species classes

2. **Computation of Statistical details and Complexity**   **(CO2)**

    i. Calculate the Five Number Summary(Quartiles, IQR) for the attribute(age) of each employee at a Tea Factory.
    ii. Analyze the complexity of Heap sort, applied over different sized random lists.

3. **Preprocessing and construction of a quality dataset**   **(CO3)**

    i. Preprocess the given data to build good training sets (80%) and test sets (20%) by removing the missing values and imputing them with the mean value.
    ii. Examine the interrelations among the set of variables using Principal Component Analysis, display the PCA Components and generate Heatmap.

4. **Analysis and Interpretation of data** (CO4)

    i. Manipulate the Twitter Data Set by removing the Punctuation, Numbers, Special Characters and word length<=3. Tokenize the Words and Stem.

    ii. Generate a word cloud for the Twitter dataset and retrieve the top 15 positive and negative tags.

5. **Classification and Clustering (Performance Assessment)** (CO5)

    i. Find core samples of high density and expand clusters from them using DBSCAN Clustering.

    ii. Split the iris dataset into train and test data(80%-20%) and train or fit the data into the model and using the K Nearest Neighbor Algorithm calculate the performance for different values of k.

6. **Evaluate the Performance of Machine Learning algorithms** (CO6)

    i. Evaluate the performance of Machine Learning algorithms using Confusion Matrix, Accuracy, Sensitivity, Specificity, Precision and Recall.

    ii. Employ Linear Regression to check the linearity between the a)stock price and interest rate, b)stock price and unemployment rate.