MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
(A constituent unit of MAHE, Manipal)

M.C.A.

COURSE PLAN: Machine Learning Lab

Department:	DATA SCIENCE & COMPUTER APPLICATIONS				
Course Name & code:	Machine Learning Lab			DSE 2242	
Semester & branch:	4th Semester		DSE		
Name of the faculty:					
No of contact hours/week:	**L**	**T**	**P**	**C**	
	0	**0**	**3**	**1**	

Course Outcomes (COs)

At the end of this course, the student should be able to:		No. of Contact Hours	Marks
CO1	To understand the basics of Machine learning using Python.	6	10
CO2	To apply Supervised Machine Learning Algorithms	12	40
CO3	To apply Unsupervised Machine Learning Algorithms.	9	30
CO4	To apply Ensemble based Classifier methods .	6	20
	Total	**33**	**100**

Lesson Plan

L No	Topics	Course Outcome Addressed
Exp - 01	1) Create a program that reads the length and width of a farmer's field from the user in feet. Display the area of the field in acres. **Hint:** There are 43,560 square feet in an acre. 2) Write a program that asks the user to enter the width and length of a room. Once the values have been read, your program should compute and display the area of the room. The length and the width will be entered as floating point numbers. Include units in your prompt and output message; either feet or meters, depending on which unit you are more comfortable working with. 3) Python includes a library of functions for working with time, including a function called **asctime** in the time module. It reads the current time from the computer's internal clock and returns it in a human-readable format. Write a program that displays the current time and date. Your program will not require	

any input from the user.

4) Create a program that reads the following10 data values (integers) and displays them in sorted order (from smallest to largest). Use the min and max functions to find the smallest and largest values. Also compute the 5 number summary for the data. Find the IQR value. Also find the outliers if any.

Data: 16, 09, 14, 11, 13, 06, 18, 15, 10, 12.

5) Create a program that reads a letter of the alphabet from the user. If the user enters **a, e, i, o or u** then your program should display a message indicating that the entered letter is a vowel. If the user enters y then your program should display a message indicating that sometimes y is a vowel, and sometimes y is a consonant. Otherwise your program should display a message indicating that the letter is a consonant.

6) The following table lists the sound level in decibels for several common noises.

Noise	Decibel level (dB)
Jackhammer	130
Gas lawnmower	106
Alarm clock	70
Quiet room	40

Write a program that reads a sound level in decibels from the user. If the user enters a decibel level that matches one of the noises in the table then your program should display a message containing only that noise. If the user enters a number of decibels between the noises listed then your program should display a message indicating which noises the level is between. Ensure that your program also generates reasonable output for a value smaller than the quietest noise in the table, and for a value larger than the loudest noise in the table.

7) The marks obtained by 10 students in a class test were as follows:

38, 41, 36, 31, 45, 38, 27, 32, 29, 39

Find:

(a) The mean of their marks
(b) The mean of their marks when the marks of each student are increased by 2.
(c) The mean of their marks when one mark is deducted from marks of each Student.
(d) The mean of their marks when the marks of each student halved

8) When analysing data collected as part of a science experiment it may be desirable to remove the most extreme values before performing other calculations. Write a function that takes a list of values and an non-negative integer, n, as its parameters. The function should create a new copy of the list with the n largest elements and the n smallest elements removed. Then it should return the new copy of the list as the function's only result. The order of the elements in the returned list does not have to match the order of the elements in the original list.

	Your function should read a list of numbers from the user and remove the two largest and two smallest values from it. Display the list with the outliers removed, followed by the original list. Your program should generate an appropriate error message if the user enters less than 4 values.	
Exp-02	1) Generate numerical data using Range and Arange functions. 2) Understand and implement linspace function. 3) Using lists, numpy arrays, range and arrange functions implement the following plots using matplotlib library functions: a) Simple Scatter Plot b) Sine curve plot c) Data generation using random function d) Histograms, bar plots, box plot 4) Using the math library of Python implement the following Matrix Operations: a) Create an M X N matrix using numpy arrays. The values of M and N should be taken as input from the user. b) Perform the operations of Matrix addition and Multiplication using python lists and arrays. c) Using Python coding and also built in Math functions write a program to perform matrix inversion. 5) Using any random data write a program in Python illustrating the conversion of Lists to numpy Arrays and numpy Arrays to Lists. 6) Prepare and excel sheet for Student's data with the following content: (a) data rows of minimum 10 students data. (b) Columns Data: SNO, REGNO, NAME, M1, M2, M3, M4, M5. (c) Computer the Mean, Median, Mode, Average and 5 number summary of all Marks Columns (d) Draw a box plot for all numeric columns. (e) Draw various possible plots for student marks in each subject. (f) Display the outlier values if any in each of the students marks columns. 7) Generate random data using all available built in random functions in Python. 8) Write a program in Python to create a new csv file and fill in the data used in Q.4 and display the .csv file.	
Exp-03	1) With given data and using matplotlib library do the following plots: a) Histograms, Bar, Pie charts, Scatter and Box. 2) Write a program in Python to perform the following operations on matrices: a) Reading and Writing matrices using numpy arrays. b) Matrix Addition and Multiplication of given dimensions. c) Matrix Transpose and Matrix Inversion. 3) With the given sample data write a program to perform Simple Linear Regression and also plot the line of best fit for the given data. The required	

	linear regression equations along with the final parameter values should be displayed in the output. 4) Perform Simple Linear Regression using the data sets: mtcars.csv and abalone.csv and perform visualization of the Regression Line. The outputs should be as per the templates. **Outputs:** (i) The derived Regression equation with the appropriate parameter values. (ii) Correlation using heat maps. (iii) Using the Cost function derive the appropriate regression parameter values. The iterations should be tabulated in an excel sheet indicating the parameters chosen in each iteration and the final set of parameters that result in error reduction of the cost function.	
Exp-04	**Note: (1) Use the train_test_split function for each of the following algorithms with the test sizes = {0.3, 0.4, 0.5}.** **(2) Use the concept of heat maps to visualize the degree of correlation between the variables.** **(3) Data Sets: insurance.csv & salaries.csv** 1) With given data sets apply Multiple Linear Regression and determine the appropriate regression parameters using the concept of matrices and vectors representation of the regression variables and parameters. Also draw the scatter and regression plots for the given set of dependent and independent variables. 2) With given data sets apply Logistic Regression and determine the appropriate regression parameters. Also draw the scatter and regression plots for the given set of dependent and independent variables. 3)Apply Gradient Descent on the following three regression algorithms: (a) Simple Linear Regression; (b) Multiple Linear Regression; and Logistic Regression. And display the appropriate values of parameters which result in reduction of the respective cost functions. **Outputs:** (i) The derived Regression equation with the appropriate parameter values. (ii) Correlation using heat maps. (iii) Using the Cost function derive the appropriate regression parameter values. (iv) The iterations should be tabulated in an excel sheet indicating the parameters chosen in each iteration and the final set of parameters that	

	result in error reduction of the cost function.	
Exp-05	**Q.1)** With given data sets apply the KNN Classifier for the following data sets: (i) Iris.csv, Wine.csv, abalone.csv, mtcars.csv (ii) Tabulate the following information regarding each data set: (a) Purpose of the data sets. (b) Dimensions of the data sets. (c) Column Names and data types of the data sets. (iii) Use the number of nearest neighbors i.e. the k-value = {1, 2, 3}. (iv) Use the train_test-split function of python to split the given data sets using the following test_size values: {0.3, 0.4, 0.5}. (v) Keep the maximum epochs value = 10. (vi) Tabulate the following in an excel sheet the outputs according the above combinations. (vii) Display the classification report and confusion matrices of each combination of the above parameters.	
Exp-06	**Data Sets: diabetes.csv, cancer.csv** **Q.1)** With given data sets apply the concept of handling missing and null values using the built-in functions of Python library. **Q.2)** With given data sets apply the Naïve Bayes Classifier for the given data sets. (i) Tabulate the following information regarding each data set: a) Purpose of the data sets. b) Dimensions of the data sets. c) Column Names and data types of the data sets. (ii) Use the train_test-split function of python to split the given data sets using the following test_size values: {0.3, 0.4, 0.5}. (iv) Keep the maximum epochs value = 10. (v) Tabulate the following in an excel sheet the outputs according the above combinations. (vi) Display the classification report and confusion matrices of each combination of the above parameters.	
Exp-07	Implement Principal Component Analysis (PCA & LDA) in Python. Perform the following on the given datasets. a) Import all the libraries. b) Load Data c) Apply PCA and LDA: i. Standardize the dataset prior to PCA. ii. Import PCA, LDA from sklearn.decomposition. iii. Choose the number of principal components. d) Check Components e) Plot the components (Visualization) f) Calculate variance ratio	

Exp-08	**Implement Decision Tree Algorithm in Python** 1) Import the dataset. (Play Tennis dataset) 2) Initiate the implementation by performing data pre-processing steps. Detail the steps involved in cleaning, handling missing values, and encoding categorical variables. 3) Develop a Python function or script to fit a Decision Tree algorithm to the training set. Specify the key hyperparameters and configurations involved in training the Decision Tree model. 4) Implement code to predict the test results using the trained Decision Tree model. **Outputs:** i. Evaluate the accuracy of the decision tree model on the test set. Generate a confusion matrix to analyze the model's performance. ii. Create visualizations to represent the results of the Decision Tree model on the test set. Create visualizations to represent the results of the Decision Tree model on the test set.	
Exp-09	**Implement Support Vector Machine algorithm.** 1. Import libraries 2. Import datasets. 3. Explore the data to gain insights about the data (View dimension, preview the dataset, view the column names of the data frame, remove leading spaces from column names, rename column names, check the distribution of the target_class column) 4. View the percentage distribution of the target_class column, View the summary of the dataset. 5. Explore missing values in variables, view summary statistics in numerical variables, draw boxplots to visualize outliers) 6. Check the distribution of variables 7. Split data into separate training and test set 8. Run SVM with default hyperparameters Run SVM with rbf kernel and C=1.0 and gamma=auto Run SVM with rbf kernel and C=100.0 Run SVM with rbf kernel and C=1000.0 9. Run SVM with linear kernel i. Run SVM with linear kernel and C=1.0 ii. Run SVM with linear kernel and C=100.0 iii. Run SVM with linear kernel and C=1000.0 10. Run SVM with polynomial kernel i. Run SVM with polynomial kernel and C=1.0 ii. Run SVM with polynomial kernel and C=100.0 iii. Run SVM with polynomial kernel and =1000.0 11. Run SVM with sigmoid kernel i. Run SVM with sigmoid kernel and C=1.0 ii. Run SVM with sigmoid kernel and C = 100.0 iii. Run SVM with sigmoid kernel and C=1000.0	

	12. Compare the train-set and test-set accuracy 13. Visualize confusion matrix with seaborn heatmap 14. Generate classification report 15. Evaluate classification accuracy and classification error	
Exp-10	**K-Means Clustering** 1. Import libraries. 2. Import dataset (Iris dataset). 3. Explore the data to gain insights about the data. 4. Check the distribution of the target_class column. 5. View the percentage distribution of the target_class column. 6. View the summary of the dataset. 7. Explore missing values in variables.	
Exp-11	Considering various regression and classification models, explore the following ensemble approaches. a) Averaging method b) Max voting c) Stacking d) Blending e) Bagging f) Boosting **Import Necessary Libraries** **Creating Dataset:** Generate or import a suitable dataset for classification or clustering tasks. Explain the characteristics of the dataset chosen and its relevance to the ensemble methods. **Initializing the Models:** Initialize different classification and clustering models, considering a diverse set of algorithms for ensemble techniques. **Fitting Training Data:** Train each model on the training dataset. Specify the hyperparameters and configurations used for each model during the training phase. **Testing the Model:** Evaluate the performance of each ensemble method on the test dataset. Discuss the metrics used for assessment and compare the results obtained from the different approaches.	
Exp-12	**End Semester Examination**	

References:

1. Hans Peter Langtangen, Python Scripting for Computational Science, (3e), Springer Publishers, 2014
2. Naomi R. Ceder, The Quick Python Book, (2e), Manning Publications Co., 2010
3. Wesley J. Chun, Core Python Applications Programming, (3e), Prentice Hall Publishers, 2012
4. G. James, D. Witten, T Hastie, R Tibshirani, An introduction to statistical learning with applications in R, Springer, 2013.

Submitted by:
Mr. Nirmal Kumar Nigam, Mrs. Shavantrevva S B & Mrs. Rashmi M

(Signature of the Faculty)

Date: 22/07/2024

Approved by:

(Signature of HOD)

Date:

Faculty members teaching the course (if multiple sections exist):

Faculty	Section	Signature	Faculty	Section	Signature
Mr. Nirmal Kumar Nigam	**B**		Mrs. Shavantrevva S B	**C**	
Mrs. Rashmi M	**A**				