



INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal - 500 043, Hyderabad, Telangana

COURSE CONTENT

DATA MINING AND KNOWLEDGE DISCOVERY LABORATORY								
VI Semester: CSE (AI & ML)								
Course Code	Category	Hours / Week			Credits	Maximum Marks		
ACSD43	Core	L	T	P	C	CIA	SEE	Total
		0	0	2	1	40	60	100
Contact Classes: NIL	Tutorial Classes: NIL	Practical Classes: 45			Total Classes: 45			
Prerequisite: No Prerequisites required to this course								

I. COURSE OVERVIEW:

This course helps the students to practically understand a data warehouse, techniques and methods for data gathering and data pre-processing using different tools. The different data mining models and techniques will be discussed in this course. The main objective of this lab is to impart the knowledge on how to implement classical models and algorithms in data warehousing and data mining and to characterize the kinds of patterns that can be discovered by association rule mining, classification and clustering.

II. COURSE OBJECTIVES:

The students will try to learn:

- I The Data Object Exploration and visualization
- II The pre-processing on new and existing datasets.
- III Frequent item set generation and association rules on transactional data.
- IV The data model creation by using various classification and clustering algorithms.
- V The data models accuracy analysis by varying the sample size.

III. COURSE OUTCOMES:

At the end of the course students should be able to:

- CO1 Analyze the knowledge generated from data objects, matrix operations using Numpy.
- CO2 Demonstrate Numpy module methods to categorize and correlate the raw data.
- CO3 Select appropriate pre-processing techniques to manage the missing values of data.
- CO4 Apply Apriori Algorithm and logistic regression for classification of data mining.
- CO5 Identify Classification technique from Decision Tree, Bayesian Network and Support Vector Machines to mine knowledge from pre-processed data
- CO6 Examine Clustering algorithms to build predication model for solving real world problem.

IV. COURSE CONTENT:

WEEK 1: Data Object, Matrix Operations Using Numpy

- a) Create multi-dimensional arrays and find its shape and dimension
- b) Create a matrix full of zeros and ones ,do reshape and flatten of data
- d) Append data vertically and horizontally & Indexing and slicing

WEEK 2: Data Object, Matrix Operations Using Numpy

- a) Dot and matrix product of two arrays
- b) Compute the Eigen values of a matrix
- c) Solve a linear matrix equation such as $3 * x_0 + x_1 = 9$, $x_0 + 2 * x_1 = 8$
- d) Compute the multiplicative inverse of a matrix

WEEK 3: Exploration and visualization of Data objective

- a) Loading data from CSV file
- b) Compute the basic statistics of given data - shape, no. of columns, mean
- c) Splitting a data frame on values of categorical variables

WEEK 4: Exploration of Data, Correlation

- a) Load data, describe the given data and identify missing, outlier data items
- b) Find correlation among all attributes
- c) Visualize correlation matrix

WEEK 5: Data Preprocessing – Handling Missing Values

Write a python program to impute missing values with various techniques on given dataset.

- a) Remove rows/ attributes
- b) Replace with mean or mode

WEEK 6: Association Rule Mining – Apriori

Write a python program to find rules that describe associations by using Apriori algorithm between different products given as 7500 transactions at a French retail store.

- a) Display top 5 rows of data
- b) Find the rules with min_confidence : .2, min_support= 0.0045, min_lift=3, min_length=2

WEEK 7: Decision Tree, Bayesian Network, Support Vector Machines (SVM)

Write a python program to

- a) Explore data and visualize each attribute
- b) Predict the test set results and find the accuracy of the model

WEEK 8: Classification – Decision Tree

Write a python program to

- a) Calculate Euclidean Distance.
- b) Get Nearest Neighbors
- c) Make Predictions.

WEEK 9: Decision Tree, Bayesian Network, Support Vector Machines (SVM)

Write a python program to

- a) Explore data and visualize each attribute
- b) Predict the test set results and find the accuracy of the model

WEEK 10: Decision Tree, Bayesian Network, Support Vector Machines (SVM)

Write a python program to

-
- a) Explore data and visualize each attribute
 - b) Predict the test set results and find the accuracy of the model

WEEK 11: Decision Tree, Bayesian Network, Support Vector Machines (SVM)

Write a python program to

- a) Explore data and visualize each attribute

WEEK 12: Clustering – K- Means

Write a python program

- a. to perform preprocessing.
- b. to perform clustering using k-means algorithm to cluster the records into two i.e., the ones who survived and the ones who did not.

V. REFERENCE BOOKS:

1. Robert Layton, “Learning Data Mining with Python”, Packt Publishing, 2015.

VI. WEB REFERENCES:

1. <https://www.dataquest.io/blog/sci-kit-learn-tutorial/>
2. https://www.ibm.com/support/knowledgecenter/en/SS3RA7_sub/modeler_tutorial_ddita/modeler_tutorial_ddita_gentopic1.html
3. <https://archive.ics.uci.edu/ml/datasets.php/>
4. <https://www.datacamp.com/community/tutorials/svm-classification-scikit-learn-python>
5. <https://pythonprogramming.net/k-mean-titanic-dataset-machine-learning-tutorial/>
6. <https://newoutlook.it/download/python/learning-data-mining-with-python.pdf>

VII. MATERIALS ONLINE:

1. Course Template
2. Lab Manual