Laboratory Manual

for

Artificial Intelligence and Machine Learning Lab.

(MCA-263)

MCA - III Semester

Compiled by:

Dr. Rakhee

(Assistant Professor, BVICAM, New Delhi)

Bharati Vidyapeeth's
Institute of Computer Applications and Management (BVICAM)

A-4, Paschim Vihar, Rohtak Road, New Delhi-63

Visit us at: www.bvicam.in

Index

List of Abbreviations

BTL	Bloom's Taxonomy Level
CE	Communication Efficacy
CICP	Conduct Investigations of Complex Computing Problems
CK	Computational Knowledge
CO	Course Outcome
DAC	Departmental Advisory Committee
DDS	Design and Development of Solutions
I&E	Innovation and Entrepreneurship
I&T	Individual & Team Work
IQAC	Internal Quality Assurance Cell
LLL	Life-Long Learning
MTU	Modern Tool Usage
PA	Problem Analysis
PE	Professional Ethics
PEO	Programme Educational Objective
PMF	Project Management and Finance
PO	Programme Outcome
SEC	Societal and Envoirnmental Concern
SED	Stream Editor

Declaration

Department	:	Department of Computer Science and Applications
Course, Year and the Semester to which Lab is offered	:	MCA - II Year, III Semester
Name of the Lab Course	:	Artificial Intelligence and Machine Learning Lab.
Course Code	:	MCA-263
Version No.	:	
Name of Course/Lab Teacher	:	Dr. Rakhee
Laboratory Manual Committee	:	1. Dr. Ritika Wason 2. Prof. P. S. Grover 3. Mr. Amit Sharma, Alumni & Industry Expert 4. Dr. Sunil Pratap Singh
Approved by	:	DAC
Approved by	:	IQAC

Signature (Course Teacher)	**Signature** (Head of Department)	**Signature** (IQAC Coordinator)

1. Vision of the Department

To become a centre of excellence in the field of Computer Science and Applications to produce quality professionals in software development.

2. Mission of the Department

M1 To produce quality software professionals as per global industry standards.

M2 To foster innovation, entrepreneurial skills, research capabilities and bring all-round development amongst budding professionals.

M3 To promote analytical and collaborative life-long learning skills, among students and faculty members.

M4 To inculcate strong ethical values and professional behaviour while giving equal emphasis to social commitment and nation building.

3. Programme Educational Objectives (PEOs)

The PEO's for the MCA programme are as follows:

PEO1 Exhibit professional competencies and knowledge for being a successful technocrat.

PEO2 Adopt creative and innovative practices to solve real-life complex problems.

PEO3 Be a lifelong learner and contribute effectively to the betterment of the society.

PEO4 Be effective and inspiring leader for fellow professionals and face the challenges of the rapidly changing multi-dimensional, contemporary world.

4. Programme Objectives (POs)

PO1: Computational Knowledge (CK)

Demonstrate competencies in fundamentals of computing, computing specialisation, mathematics, and domain knowledge suitable for the computing specialisation to the abstraction and conceptualisation of computing models from defined problems and requirements.

PO2: Problem Analysis (PA)

Identify, formulate, and analyze complex real-life problems in order to arrive at computationally viable conclusions using fundamentals of mathematics, computer sciences, management and relevant domain disciplines.

PO3: Design and Development of Solutions (DDS)

Design efficient solutions for complex, real-world problems to design systems, components or processes that meet the specifications with suitable consideration to public health, and safety, cultural, societal, and environmental considerations.

PO4: Conduct Investigations of Complex Computing Problems (CICP)

Ability to research, analyze and investigate complex computing problems through design of experiments, analysis and interpretation of data, and synthesis of the information to arrive at valid conclusions.

PO5: Modern Tool Usage (MTU)

Create, select, adapt and apply appropriate technologies and tools to a wide range of computational activities while understanding their limitations.

PO6: Professional Ethics (PE)

Ability to perform professional practices in an ethical way, keeping in mind cyber regulations & laws, responsibilities, and norms of professional computing practices.

PO7: Life-Long Learning (LLL)

Ability to engage in independent learning for continuous self-development as a computing professional.

PO8: Project Management and Finance (PMF)

Ability to apply knowledge and understanding of the computing and management principles and apply these to one's own work, as a member and leader in a team, to manage projects in multidisciplinary environments.

PO9: Communication Efficacy (CE)

Ability to effectively communicate with the technical community, and with society at large, about complex computing activities by being able to understand and write effective reports, design documentation, make effective presentations, with the capability of giving and taking clear instructions.

PO10: Societal and Envoirnmental Concern (SEC)

Ability to recognize and assess societal, environmental, health, safety, legal, and cultural issues within local and global contexts, and the consequential responsibilities applicable to professional computing practices.

PO11: Individual & Team Work (I&T)

Ability to work in multi-disciplinary team collaboration both as a member and leader as per need.

PO12: Innovation and Entrepreneurship (I&E)

Ability to apply innovation to track a suitable opportunity to create value and wealth for the betterment of the individual and society at large.

5. Institutional Policy for Students' Conduct

The following guidelines shall be followed:-

5.1 All the students in their introductory Lab. shall be assigned a system, which shall be their workplace for the complete semester. Students can store records of all their Lab. assignments on their individual workstations.

5.2 Introductory Lab. shall include an introduction to the appropriate software/tool, followed by a basic Introductory Assignment having Practice Questions. All the students are expected to complete this assignment within a week time, as the same shall be assessed through a lab. test.

5.3 Each week, the instructor, in parallel to respective topics covered in the theory lecture, shall assign a set of practical problems to the students. The problems in these assignments shall be divided into two parts. The first set of problems shall be compulsory for all the students and its record need to be maintained in the Prcatical File, having prescribed format, as given in Appendix - A. All the students should get the Practical File checked and signed, weekly, by the respective teacher. The second set of problems is Advanced Problems and shall be optional. Student may solve these advanced problems for their further practice.

5.4 Cellular phones, pagers, CD players, radios and similar devices are prohibited in the classrooms, laboratories and examination halls.

5.5 Laptops, Tablets may be used in lectures/labs for the purpose of taking notes or working on team-projects.

5.6 The internal practical exam shall be conducted towards the end of the semester and shall include the complete set of lab exercises conducted as per syllabus. However, students shall be assessed on continuos basis through overall performances in regular lab. tests, both announced and surprise and viva-voce.

5.7 The respective faculty shall prepare and submit sufficient number of practical sets of computing problems to the Dean (Examinations), atleast two

weeks prior to the actual exam. It is the responsibility of the faculty to ensure that a set should not be repeated for more than 5 students in a given batch.

5.8 The internal practical exam shall be of 3 hours duration where the student shall be expected to implement solutions to his/her assigned set of problems on appropriate software tools in the lab.

5.9 Once implemented, student shall also appropriately document code implemented in the assigned answer sheets, which shall be submitted at the end of the examination. All the students shall also appear for viva-voce examination during the exam.

5.10 Co-operate, collaborate and explore for the best individual learning outcomes but copying or entering into the act of plagiarism is strictly prohibited.

6. Learning Outcomes of Laboratory Work

The student shall demonstrate the ability to:

- Verify and Implement the concepts and theory learnt in class.
- Code and use Software Tools to solve problems and present their optimal solutions.
- Apply numerical/statistical formulas for solving problems/questions.
- Develop and apply critical thinking skills.
- Design and present Lab as well as project reports.
- Apply appropriate methods for the analysis of raw data.
- Perform logical troubleshooting as and when required.
- Work effectively as a member of a team in varying roles as need be.
- Communicate effectively, both oral and written.
- Cultivate ethics, social empathy, creativity and entrepreneurial mindset.

7. Course/Lab Outcomes (COs)

CO1	Apply heuristic search-based algorithms to solve different puzzles. [BTL3]
CO2	Identify reduction techniques on large datasets and reduce their dimensionality. [BTL3]
CO3	Analyze the datasets for bias and apply appropriate regression techniques. [BTL4]
CO4	Evaluate the learning techniques for classification. [BTL5]
CO5	Implement the knowledge of inferences rules to design the knowledge base. [BTL6]
CO6	Create a domain specific intelligent application. [BTL6]

8. Mapping of CO's with PO's

Table 1: Mapping of CO's with PO's

PO/CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓		✓							
CO2	✓	✓	✓	✓	✓							
CO3	✓	✓	✓	✓	✓	✓				✓		
CO4	✓	✓	✓	✓	✓	✓				✓		
CO5	✓	✓	✓	✓	✓	✓				✓		
CO6	✓	✓	✓	✓	✓	✓	✓		✓		✓	✓

9. Course/Lab Description

Course (Lab) Title : Artificial Intelligence and Machine Learning Lab.

Course (Lab) Code : MCA-263

Credits : 01

Pre-requisites : Basic Understanding of AI & ML algorithms, Python

and Graph Theory

Academic Session : July to December

Contact Hours/Week : 02 (01 Labs of 02 Hours/Week)

Internal Assessment : 40 Marks

External Assessment : 60 Marks

10. Grading Policy

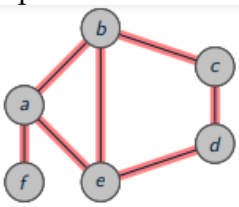
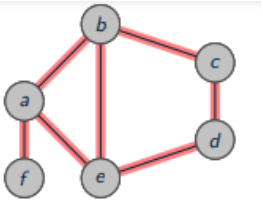
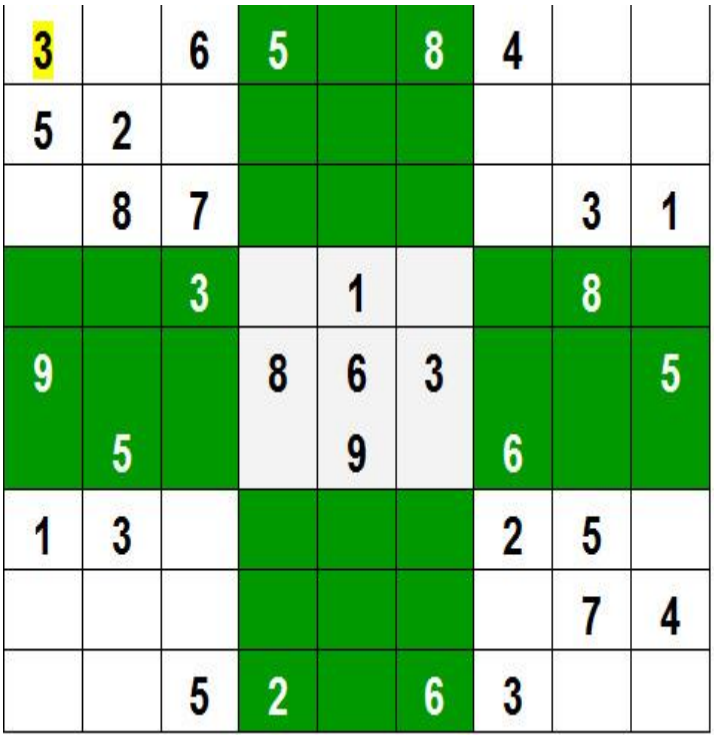
Item	Points	Marks	Remarks
Weekly Lab Exercises including Practical Files	10	10	Closed Book/Open Book
Internal End-Term Practical Examination	20	10	Closed Book
Viva-Voce	20	20	Closed Book
External End-Term Examinations	60	60	Closed Book (conducted and evaluated by the University)
	Total	**100**	

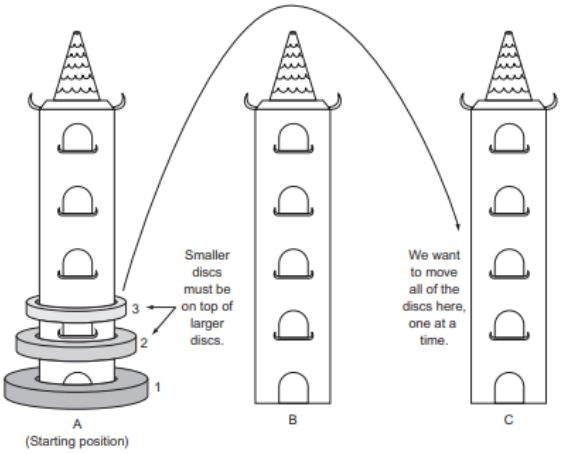
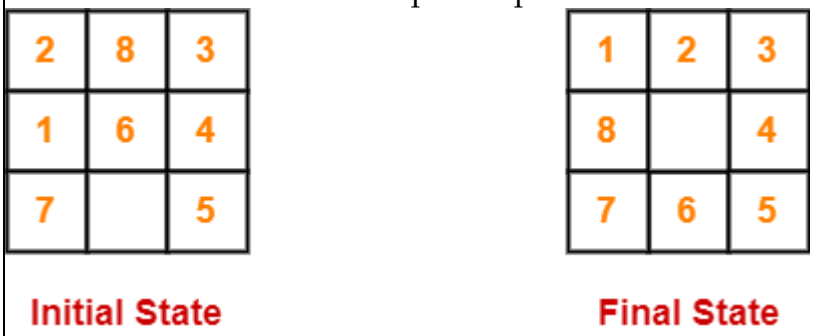
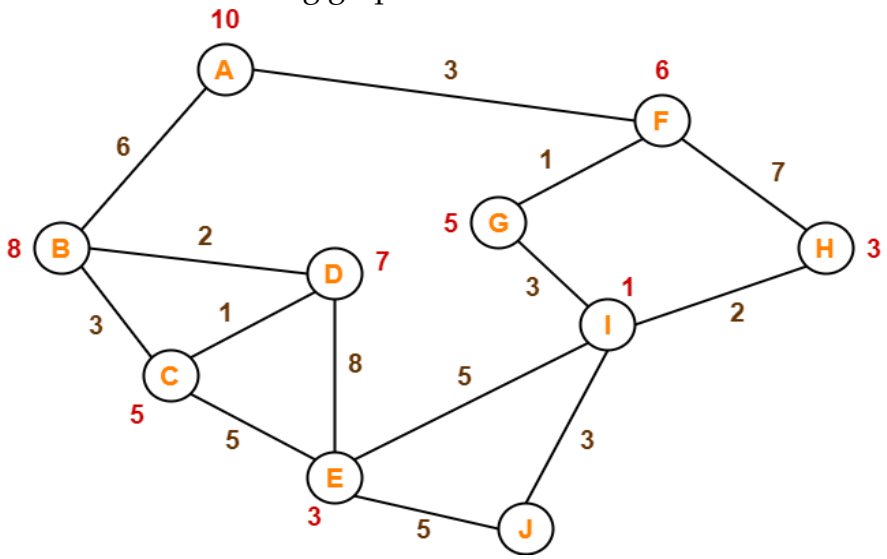
11. Lesson Plan

Week No.	Lab No.	Topics/Concepts to be Covered
1.	1.	Create a solution to solve the Graph Traversal using BFS and DFS.
2.	2.	Create a solution to solve problems of informed/Heuristic search
3.	3.	Create a solution to solve problems of informed/Heuristic search (Contd.)
4.	4.	Create a solution for data processing in google colab using python programs and to clean the dataset and evaluate on classification algorithms.
5.	5.	Create a solution to detect bias and variance in dataset. Learn to detect overfitting and underfitting.
6.	6.	Create a solution to predict the relationship between two variables using linear regression.

Week No.	Lab No.	Topics/Concepts to be Covered
7.	7.	Create a solution to predict the relationship between two variables using logistic regression.
8.	8.	Create a solution for dimensionality reduction implementing unsupervised learning.
9.	9.	Create a solution to define clusters using K-means, ensemble learning.
10.	10.	Create a solution to predict variable using neural networks.
11.	11.	Create a solution to build a recommender system.
12.	12.	Buffer for revision

12. Lab Exercises/Problems

P1	Create a solution to solve the Graph Traversal using BFS [CO1/BTL3] 
P2	Given a snake and ladder board, find the minimum number of dice throws to reach the destination cell starting from the source using BFS [CO1/BTL4]
P3	Create a solution to solve the graph traversal using DFS [CO1/BTL3] 
P4	Create a solution to solve the following Sudoku using DFS [CO1/BTL5] 
P5	The Towers of Hanoi Three vertical pegs (henceforth "towers") stand tall. We will label them A, B, and C. Doughnut-shaped discs are around tower A. The widest disc is at the bottom, and we will call it disc 1. The rest of the discs above disc 1 are labeled with increasing numerals and get progressively narrower. For instance, if we were to work with three discs, the widest disc, the one on the bottom, would be 1. The next widest disc, disc 2, would sit on top of disc 1. And finally, the narrowest disc, disc 3, would sit on top of disc 2.

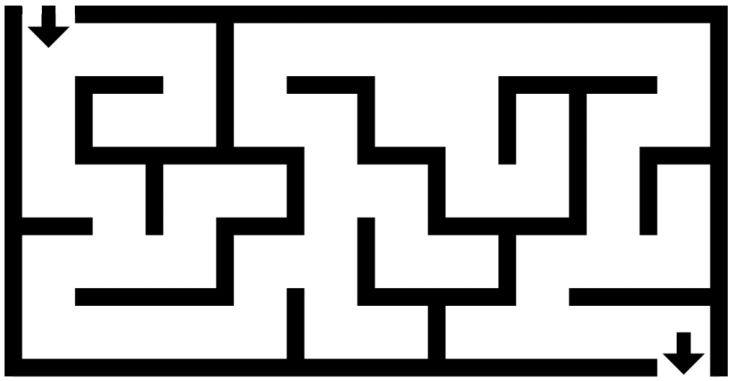
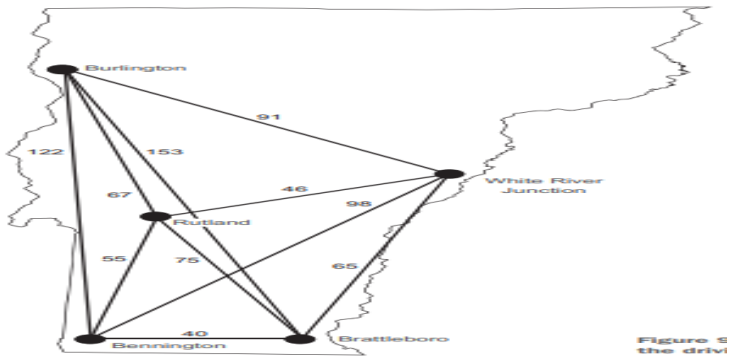
	Our goal is to move all of the discs from tower A to tower C given the following constraints: Only one disc can be moved at a time. The topmost disc of any tower is the only one available for moving. A wider disc can never be atop a narrower disc. Solve this problem [CO1/BTL6] 
P6	Given an initial state of an 8-puzzle problem and final state to be reached.  Find the most cost-effective path to reach the final state from initial state using A* Algorithm. F(n)=g(n)+h(n) Consider g(n) = Depth of node and h(n) = Number of misplaced tiles. [CO1/BTL6]
P7	Consider the following graph-  The numbers written on edges represent the distance between the nodes.

	The numbers written on nodes represent the heuristic value. Find the most cost-effective path to reach from start state A to final state J using A* Algorithm. [CO1/BTL6]
P8	Create a solution to load the IRIS dataset from the following URL: "https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data". Prepare the data, evaluate the algorithms and present the results through suitable visualizations? [CO2/BTL3]
P9	Clean the Iris Dataset of Question P8 of duplicate values and repeat your analysis. Which algorithm performs better now? [CO2/BTL3]
P10	Clean the Oil Spill dataset from the following URL: https://github.com/jbrownlee/Datasets/blob/master/oil-spill.csv. Clean the data of duplicate data, single value columns and low variance columns. Once the data is prepared evaluate it on the classification algorithms in CP1 and present the result through suitable visualizations? [CO2/BTL5]
P11	Clean the Heart Disease Database to create a classifier that can help diagnose patients[CO2/BTL4]
P12	Load the Boston housing dataset directly via URL and split it into train and test sets, then estimates the mean squared error (MSE) for a linear regression as well as the bias and variance for the model error over 200 bootstrap samples. Estimate the bias and variance for the regression model? [CO2/BTL5]
P13	Using linear regression predict the relationship between the experience of an individual and his salary. Predict the variance and bias for the same? [CO3/BTL5]
P14	Predict the CO2 emission of a car based on the size of the engine, but use multiple regression so we can throw in more variables, like the weight of the car? [CO3/BTL5]
P15	Plot the CO2 emission values wrt engine size using multiple linear regression? [CO3/BTL5]
P16	You have the following client list and some additional sales information stored in a CSV file (where the file name is 'Clients '): Person Name — Country — Product — Purchase Price Jon — Japan — Computer — $800 Bill — US — Tablet — $450 Maria — Canada — Printer — $150 Rita — Brazil — Laptop — $1,200 Jack — UK — Monitor — $300 Ron — Spain — Laptop — $1,200 Jeff — China — Laptop — $1,200 Carrie — Italy — Computer — $800 Marry — Peru — Computer — $800

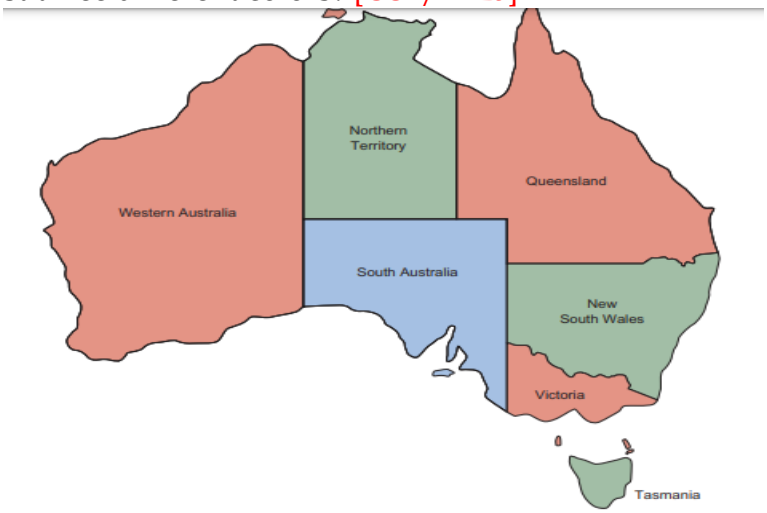
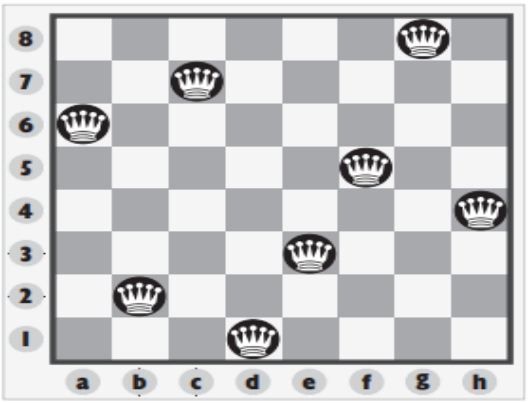
	Ben Russia Printer $150 Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a .CSV file. FIND-S Algorithm 1. Initialize h to the most specific hypothesis in H 1. For each positive training instance x 2. For each attribute constraint ai in h If the constraint ai is satisfied by x Then do nothing 3. Else replace ai in h by the next more general constraint that is satisfied by x Output hypothesis h[CO3/BTL4]
P17	Apply Linear Regression and build a model that studies the relationship between the head size and the brain weight of an individual? Evaluate by using least square regression method where RMSE (root mean squared error) and R-squared/R2 will be the model evaluation parameters. [CO3/BTL4]
P18	Apply logistic regression on userdata.csv dataset to predict the users who may be potential customers to purchase a SUV car? Also generate the confusion matrix to evaluate your model? [CO4/BTL5]
P19	Apply logistic regression on handwritten digits dataset to classify the digits. Evaluate your model too? [CO4/BTL5]
P20	Create a mall_customer_dataset.csv dataset and apply the K-means on the same after deciding the number of clusters using the elbow method to uncover the patterns? [CO4/BTL5]
P21	Use the pima indian diabetes database to perform ensemble predictions using the following bagging classifiers: Bagged Decision Trees, Random Forest Classifier and Extra trees? [CO5/BTL5]
P22	Implement a simple neural network with: - 2 inputs - a hidden layer with 2 neurons (h1, h2) - an output layer with 1 neuron (o1) [CO5/BTL6]
P23	Build a simplified clone of IMDB Top 250 Movies using metadata collected from IMDB. The following are the steps involved: 1. Decide on the metric or score to rate movies on. 2. Calculate the score for every movie. 3. Sort the movies based on the score and output the top results. 4. Use the Full Movie Lens dataset. [CO6/BTL6]
P24	Build a system that recommends movies that are similar to a particular movie. Compute pairwise cosine similarity scores for all movies based on their plot descriptions and recommend movies based on that similarity score threshold. The plot description is available to you as the overview feature in your metadata dataset. [CO6/BTL6]
P25	Build a recommender system based on the following metadata: the 3 top actors, the director, related genres, and the movie plot keywords.

	Note-The keywords, cast, and crew data are not available in your current dataset, so the first step would be to load and merge them into your main DataFrame metadata. [CO3, CO6/BTL5]

13. Advanced Lab Exercises/Problems

AQ1	Create a solution for indexing Web Pages by applying BFS? [CO1/BTL6]
AQ2	Create a solution to solve a maze using DFS? [CO1/BTL6] 
AQ3	The salesman is interested in visiting five of the major cities of Vermont. We will not specify a starting (and therefore ending) city. [CO1/BTL5] 

	Rutland	Burlington	White River Junction	Bennington	Brattleboro
Rutland	0	67	46	55	75
Burlington	67	0	91	122	153
White River Junction	46	91	0	98	65
Bennington	55	122	98	0	40
Brattleboro	75	153	65	40	0

AQ4	Imagine you have a map of Australia that you want to color by state/territory (which we will collectively call "regions"). No two adjacent regions should share a color. Can you color the regions with just three different colors? [CO1/BTL5] 
AQ5	A chessboard is an eight-by-eight grid of squares. A queen is a chess piece that can move on the chessboard any number of squares along any row, column, or diagonal. A queen is attacking another piece if in a single move, it can move to the square the piece is on without jumping over any other piece. (In other words, if the other piece is in the line of sight of the queen, then it is attacked by it.) The eight queens' problem poses the question of how eight queens can be placed on a chessboard without any queen attacking another queen. [CO1/BTL5] 
AQ6	Try to use a multiple linear regression with the same dataset as in Q EP3, but this time use FUEL CONSUMPTION in CITY and FUEL CONSUMPTION in HWY instead of FUELCONSUMPTION_COMB. Does it result in better accuracy? [CO1, CO2/BTL6]
AQ7	Apply PCA to EQ1 to speed up a machine learning algorithm (logistic regression) on the MNIST dataset. Also report the improvement in score? [CO3/BTL6]

AQ8	Apply logistic regression on handwritten digits dataset to classify the digits. Evaluate your model too. [CO3/BTL6]
AQ9	Create a random dataset using the make_blobs () function from sklearn and apply K-means on the same after deciding the number of clusters using the elbow method? [CO4/BTL6]
AQ10	Build an actual neural network on the MNIST dataset and train it using the back propagation algorithm. [CO5/BTL6]

Appendix – A: Index of Lab File

Week No.	Lab. Ex. No.	Detailed Description of the Lab Exercise	Outcome Mapping		Page No./Link of Online Document	Signature of Teacher with Date
			CO	BTL		
1.						
2.						
3.						
4.						
5.						
6.						
7.						
8.						
9.						
10.						
11.						
12.						

Note: *The students should use header and footer, mentioning their roll number & name in header and page number in footer.*