COURSEPACK (Fall 2023-24)

Scheme

The scheme is an overview of work-integrated learning opportunities and gets students out into the real world. This will give what a course entails.

Course Title	Machine Learning			Course Type		Comprehensive	
Course Code	E2UC508C			Class		B. Tech. V Sem.	

Instruction delivery	Activity	Credits	Weekly Hours	Total Number of Classes per Semester				Assessment in Weightage	
				Theory	Tutorial	Practical	Self-study	CIE	SEE
	Lecture	3	3						
	Tutorial	0	0						
	Practical	0	0						
	Self-study	1	9						
	Total	4	12	45	0	0	135	50%	50%

Course Lead: Dr. Santosh Kumar

Names Course Instructors

Theory	Practical
1 Santosh Kumar	
2 MD ARQUAM	
3 Akhilesh Kumar Singh (32563)	
4 Akhilesh Kumar Tripathi	
5 C. Ramesh Kumar	
6 Vartika Mishra	
7 Ashish Ranjan	
8 Sugan Patel	
9 Kaleemur Rehman	
10 Rakesh Bharati	
11 Pooja Singh	
12 Vijayant Pawar	
13 Albert Mundu	
14 Abdul Mazid (Course Coordinator)	
15 Anupam Kumar Sharma	
16 Krishna Kant Agrawal	
17 Mithilesh Kumar Yadav	
18 Mili Dhar	
19 Rakesh Bharati	
20 Pooja Singh	
21 G. Sakthi	
22 Rakesh Bharti	

Machine Learning (ML) gives computers the ability to learn without being explicitly programmed. Machine learning is a subfield of artificial intelligence (AI). The goal of machine learning generally is to understand the structure of data and fit that data into models that can be understood and utilized by people. Machine learning allow for computers to train on data inputs and use statistical analysis in order to output values that fall within a specific range. Machine learning facilitates computers in building models from sample data in order to automate decision-making processes based on data inputs.

PREREQUISITE COURSE

PREREQUISITE COURSE REQUIRED	YES	NO (√)

If, yes please fill in the details

Prerequisite:

Prerequisitecourse code	Prerequisitecourse name

COURSE OBJECTIVE

The objective of this course is to deliver concepts of machine learning including supervised, unsupervised, and reinforcement learning with practical implementation of these on datasets by building models that will be used for data analytics and predictions.

COURSE OUTCOMES(COs)

After the completion of the course, the student will be able to:

CO No.	Course Outcomes
E2UC508C.1	Apply supervised, unsupervised, and reinforcement learning algorithms to address a wide range of tasks, including classification, regression, clustering, ranking, prediction, pattern identification, and recommendation problems.
E2UC508C.2	Implement machine learning models encompassing classification, regression, clustering, decision trees, and reinforcement learning.
E2UC508C.3	Compare and evaluate the performance of ML models for specific tasks, considering accuracy, precision, recall, confusion matrices, log loss, F-measures, and AUC-ROC curves.
E2UC508C.4	Apply Genetic Algorithm to improve machine learning classifier processes, such as decision tree model creation and optimization.

CO numbering method: **CO**   **2**   **01**   **1**
(Semester no/ course no/ co sequence)

BLOOM'S LEVEL OF THE COURSE OUTCOMES

Comprehensive

CO No.	Remember BTL1	Understand BTL2	Apply BTL3	Analyse BTL4	Evaluate BTL5	Create BTL6
E2UC508C.1			∫			
E2UC508C.2			∫			
E2UC508C.3				∫	∫	
E2UC508C.4			∫			

PROGRAM OUTCOMES (POs):

PO1 Computing Science knowledge: Apply the knowledge of mathematics, statistics, computing science and information science fundamentals to the solution of complex computer application problems.

PO2 Problem analysis: Identify, formulate, review research literature, and analyze complex computing science problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and computer sciences.

PO3 Design/development of solutions: Design solutions for complex computing problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO4 Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO5 Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern computing science and IT tools including prediction and modeling to complex computing activities with an understanding of the limitations.

PO6 IT specialist and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional computing science and information science practice.

PO7 Environment and sustainability: Understand the impact of the professional computing science solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO8 Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the computing science practice.

PO9 Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO10 Communication: Communicate effectively on complex engineering activities with the IT analyst community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO11 Project management and finance: Demonstrate knowledge and understanding of the computing science and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO12 Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PROGRAMME SPECIFIC OUTCOME (PSO):

The students of Computer Science and Engineering shall:

PSO1: Have the ability to work with emerging technologies in computing requisite to Industry 4.0.

PSO2: Demonstrate Engineering Practice learned through industry internship and research project to solve live problems in various domains.

COURSE ARTICULATIONMATRIX

COs#/ POs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
E2UC508C.1	3	1	2		2	1			1	1	1	1		
E2UC508C.2	2		2		2				1	1	1	1	2	
E2UC508C.3	2	2	1	1	2				1	1	1	1		
E2UC508C.4	2	1	2		2				1	1	1	1	2	

Note: 1-Low, 2-Medium, 3-High

COURSE ASSESSMENT

Assessment Pattern for Comprehensive Course:

Type of Course (C)	CIE			Total Marks		Final Marks CIE*0.5+SEE*0.5
	LAB@ (Work+ Record)	MTE	Course-based Project^	CIE	SEE	
COMPREHENSIVE	25	50	25	100	100	100

@Lab Work-15 marks + Lab Record-10 marks

COURSE CONTENT

THEORY

Content
Basic concepts: Definition of learning systems, Goals and applications of machine learning. Aspects of developing a learning system: training data, concept representation, function approximation. History of ML, Introduction of Machine Learning Approaches – (Artificial Neural Network, Clustering, Reinforcement Learning, Decision Tree Learning, Bayesian networks, Support Vector Machine, Genetic Algorithm), Issues in Machine Learning and Data Science Vs Machine Learning. **Types of Learning:** Supervised learning and unsupervised learning. Overview of classification: setup, training, test, validation dataset, over fitting. **Classification:** Decision Trees – Attribute Selection Measures and Tree Pruning; Bayesian and Rule-based Classification; Model Evaluation and Selection; Cross-Validation; Classification Accuracy; Bayesian Belief Networks; Classification by Backpropagation; and Support Vector Machine
Regression: Linear Regression and Logistic Regression, Perceptron, Exponential family, Generative learning algorithms, Gaussian discriminant analysis, Naive Bayes, Support vector machines: Optimal hyper plane, Kernels. Model selection and feature selection. Combining classifiers: Bagging, boosting (The Ada boost algorithm), Evaluating and debugging learning algorithms, Classification errors. Introduction to Weka: a machine learning tool for Classifiers Introduction to Weka: a machine learning tool for Clustering
Decision Tree Learning - Decision tree learning algorithm, Inductive bias, Inductive inference with decision trees, Entropy and information theory, Information gain, ID-3 Algorithm, Issues in Decision tree learning. INSTANCE-BASED LEARNING – k-Nearest Neighbour Learning, Locally Weighted Regression, Radial basis function networks, Case-based learning
Clustering, K-means, EM Algorithm, Mixture of Gaussians, Factor analysis, PCA (Principal components analysis), ICA (Independent components analysis), latent semantic indexing, Spectral clustering, Markov models Hidden Markov models (HMMs).
Reinforcement Learning–Introduction to Reinforcement Learning , Learning Task,Example of Reinforcement Learning in Practice, Learning Models for Reinforcement – (Markov Decision process , Q Learning - Q Learning function, Q Learning Algorithm), Application of Reinforcement Learning, Introduction to Deep Q Learning.
Genetic Algorithms: Introduction, Components, GA cycle of reproduction, Crossover, Mutation, Genetic Programming, Models of Evolution and Learning, Applications.

LESSON PLAN FOR COMPREHENSIVE COURSES

FOR THEORY   15 weeks * 3 Hours = 45 Classes) (1credit = 1Lecture Hour)
FOR PRACTICAL 15 weeks * 2Hours = 30 Hours lab sessions (1 credit = 2 lab hours)

LNo	Topic for Delivery	Skill	Competency
1	Basic concepts: Definition of learning systems, Goals and applications of machine learning	Explain basic concepts of ML	E2UC508C.1
2	Aspects of developing a learning system: training data	Explain the role of training Data Set	
3	concept representation, function approximation	Explain concept representation	
4	History of ML, Introduction of Machine Learning Approaches – (Artificial Neural Network, Clustering)	Explain history of ML	
5	Introduction of Machine Learning Approaches: (Reinforcement Learning, Decision Tree Learning)		
6	Introduction of Machine Learning Approaches : (Bayesian networks, Support Vector Machine, Genetic Algorithm)		
7	Issues in Machine Learning and Data Science Vs Machine Learning	Explain difference between DS and ML	E2UC508C.1 E2UC508C.2
8	Types of Learning: Supervised learning and unsupervised learning	Explain types of learning	
9	Overview of classification: setup, training, test	Explain classification of objects	
10	Overview of classification: validation dataset, over fitting		
11	Classification Families: linear discriminative, non-linear discriminative, decision trees		
12	Classification: Decision Trees – Attribute Selection Measures and Tree Pruning;		
13	Bayesian and Rule-based Classification; Model Evaluation and Selection; Cross-Validation;	Explain and apply classification models	

14	Classification Accuracy; Bayesian Belief Networks; Classification by Backpropagation; and Support Vector Machine		
15	Regression: Linear Regression and Logistic Regression, Perceptron, Exponential family		
16	Generative learning algorithms, Gaussian discriminant analysis, Naive Bayes,	Explain and implement classifiers and regression models	
17	Support vector machines: Optimal hyper plane, Kernels. Model selection and feature selection		
18	Combining classifiers: Bagging, boosting (The Ada boost algorithm)		
19	Combining classifiers: Evaluating and debugging learning algorithms, Classification errors		
20	Introduction to Weka: a machine learning tool for Classifiers	Use weka ML tool	
21	Introduction to Weka: a machine learning tool for Clustering		
22	Decision Tree Learning - Decision tree learning algorithm	Explain use of decision tree	E2UC508C.2 E2UC508C.3
23	Inductive bias, Inductive inference with decision trees		
24	Entropy and information theory, Information gain	Explain, apply and build models based on entropy	
25	ID-3 Algorithm		
26	Issues in Decision tree learning		
27	INSTANCE-BASED LEARNING – k-Nearest Neighbour Learning	Explain, apply instance and case based learning	
28	Locally Weighted Regression, Radial basis function networks		
29	Case-based learning		
30	Clustering, K-means	Explain, apply and build models based clustering	E2UC508C.1 E2UC508C.2
31	EM Algorithm	Explain and apply PCA, ICA and HMM	
32	Mixture of Gaussians, Factor analysis		
33	PCA (Principal components analysis)		
34	ICA (Independent components analysis)		
35	latent semantic indexing, Spectral		

	clustering		
36	Markov models Hidden Markov models (HMMs).		
37	Reinforcement Learning–Introduction to Reinforcement Learning	Explain, apply and build models based on Reinforcement Learning	
38	Learning Task,Example of Reinforcement Learning in Practice		
39	Learning Models for Reinforcement – (Markov Decision process, Q Learning - Q Learning function, Q Learning Algorithm)		
40	Application of Reinforcement Learning		
41	Introduction to Deep Learning	Apply deep learning and GA	E2UC508C.4
42	Genetic Algorithms: Introduction, Components		
43	GA cycle of reproduction, Crossover, Mutation		
44	Genetic Programming		
45	Models of Evolution and Learning, Applications		

BIBLIOGRAPHY

- **Text Book**

1. Tom M. Mitchell, ―Machine Learning, McGraw-Hill Education (India) Private Limited, 2013.

2. Andreas C. Müller & Sarah Guido, "Introduction to Machine Learning", with Python, A GUIDE FOR DATA SCIENTISTS, O'Reilly.

3. M. Gopal, "Applied Machine Learning", McGraw Hill Education

4. Richard S. Sutton and Andrew G. Barto, "Reinforcement Learning: An Introduction", A Bradford Book The MIT Press, Cambridge, Massachusetts, London, England

- **Reference Books**

1. Chris Albon, Machine Learning with Python Cookbook: Practical Solutions from Preprocessing to Deep Learning, O'Reilly, 2018

2. Probability for Statistics and Machine Learning: Anirban DasGupta - 2011

3. Rudolph Russell, "Machine Learning: Step-by-Step Guide To Implement Machine Learning Algorithms with Python"

4. Peter Harrington, Machine Learning in Action , April 2012 ISBN 9781617290183

5. David Barber, "Bayesian Reasoning and Machine Learning", 2012

6. Stephen Marsland, ―Machine Learning: An Algorithmic Perspective, CRC Press, 2009.

7. Nils J. Nilsson, "INTRODUCTION TO MACHINE LEARNING", TEXTBOOK, Robotics Laboratory Stanford University, Stanford, CA 94305, 1998

▪ **Journals/Magazines/Govt. Reports/Gazatte/Industry Trends**

▪ **Webliography**

1. https://ifacet.iitk.ac.in/professional-certificate-course-in-ai-and-machine-learning/?utm_source=google&utm_medium=cpc&utm_term=machine%20learning%20program&utm_content=20161035334-154966829171-661014543750&utm_device=c&utm_campaign=Search-DataCluster-PG-DSAI-AIML-CAIML-IITK-Main-IITKDomain-IN-AllDevice-Core-adgroup-CAIML-IITK-Core-Keywords&gclid=CjwKCAjwv8qkBhAnEiwAkY-ahtf1DaJ-Y1LwSBrj-txj8SC4XbNNvivTroiHn11CXXzGox61khUlNhoCtXAQAvD_BwE

2. https://www.udemy.com/topic/machine-learning/

3. https://mrcet.com/downloads/digital_notes/CSE/IV%20Year/MACHINE%20LEARNING(R17A0534).pdf

4. https://www.cs.cmu.edu/~hn1/documents/machine-learning/notes.pdf

5. https://news.vidyaacademy.ac.in/wp-content/uploads/2018/10/NotesOnMachineLearningForBTech-1.pdf

▪ **SWAYAM/NPTEL/MOOCs Certification** (Two-Three numbers)

1. https://www.udemy.com/topic/machine-learning/

2. https://www.coursera.org/learn/machine-learning

3. https://developers.google.com/machine-learning/crash-course

4. https://www.edx.org/learn/machine-learning

5. Company Name : AMAZON AWS , Title of Course : Academy Machine Learning for Natural Language Processing

PRACTICE PROBLEMS

SNo	Problem
1	Explain how Support Vector Machine can be used for classification of linearly separable data.
2	Define Hidden Markov Model. What is meant by evaluation problem and how is this solved?
3	Use K Means clustering to cluster the following data into two groups. Assume cluster centroid are m1=2 and m2=4. The distance function used is Euclidean distance. { 2, 4, 10, 12, 3, 20, 30, 11, 25 }
4	Describe the concept on density based clustering and write the steps involved in DBSCAN algorithm.
5	Differentiate between supervised and unsupervised training. Explain with suitable examples.
6	Explain feature selection and feature extraction method for dimensionality reduction.
7	Illustrate the two approaches used in subset selection.
8	Define the terms Hypothesis space and Version space. Illustrate with an

	example.
9	Describe the random forest algorithm to improve classifier accuracy.
10	Compare Cross validation with Bootstrapping Techniques.
11	Calculate the output y of a three input neuron with bias. The input feature vector is $(x1, x2, x3) = (0.8, 0.6, 0.4)$ and weight values are $[w1, w2, w3, b] = [0.2, 0.1, -0.3, 0.35]$. Use binary Sigmoid function as activation function.
12	Describe the significance of Kernal functions in SVM. List any two kernel functions.
13	Explain the basic elements of a Hidden Markov Model (HMM). List any two applications of HMM.
14	Suppose 10000 patients get tested for flu; out of them, 9000 are actually healthy and 1000 are actually sick. For the sick people, a test was positive for 620 and negative for 380. For the healthy people, the same test was positive for 180 and negative for 8820. Construct a confusion matrix for the data and compute the precision and recall for the data.
15	Consider the training data in the following table where Play is a class attribute. In the table, the Humidity attribute has values "L" (for low) or "H" (for high), Sunny has values "Y" (for yes) or "N" (for no), Wind has values "S" (for strong) or "W" (for weak), and Play has values "Yes" or "No". What is class label for the following day (Humidity=L, Sunny=N, Wind=W), according to naïve Bayesian classification?
16	What are the benefits of pruning in decision tree induction? Explain different approaches to tree pruning?
17	Given the set of values $X = (3, 9, 11, 5, 2)^T$ and $Y = (1, 8, 11, 4, 3)^T$. Evaluate the regression coefficients.
18	Explain DBSCAN algorithm for density based clustering. List out its advantages compared to K-means.
19	Find the covariance and correlation coefficient of data X={1,2,3,4,5} and Y={1,4,9,16,25)
20	Design single layer perceptron with two iteration. Consider the perceptron having with the initial weights w1=0.5, w2 = 0, learning rate α=0.2 and bias θ =0.4 for AND Boolean function. The activation function is the Step function f(x) which gives the output either 0 or 1. If value of f(x) is greater than or equal to 0, it outputs 1 or else it outputs 0.
21	Consider a perception to represent the Boolean function And with the initial weights w1=0.3, w2 = -0.2, learning rate α=0.2 and bias θ =0.4 as shown in Figure. The activation function used here is the Step function f(x) which gives the output value as binary. i.e., 0 or 1. If value of f(x) is greater than or equal to 0, it outputs 1 or else it outputs 0. Design a perceptron that performs the Boolean function AND & update the weights until the Boolean function gives the desired output
22	How Agglomerative Hierarchical clustering works? Explain the procedure to measure distance between two clusters.
23	Develop a Classifier for Email spam detection model.
24	What is purpose of The Inductive Learning Hypothesis ?
25	Giving an example, Explain the Concept Learning Task in ML.

26	Explain Adeline Network, highlighting its advantages over Perception.
27	What is the assumption in Naïve Bayesian Algorithm that makes is different from Bayesian Theorem.
28	Differentiate between Conventional Learning and Instance based Learning.
29	What are the important elements of a Reinforcement Learning Model ?
30	How does Back Propagation Rule help in Learning in a Neural Network.
31	Illustrate Principal Component Analysis (PCA) method of dimensionality reduction technique with suitable example.
32	Discuss the Classification and regression trees (CART) algorithm with an example.
33	What is Entropy in Information Theory? How it is used in decision trees for classification.
34	"Find out Precision and recall for the following data set. Consider a dataset with a 1:100 minority to majority ratio, with 100 minority examples and 10,000 majority class examples.
35	Describe about Probabilistic models used for categorical data
36	A model makes predictions and predicts 120 examples as belonging to the minority class, 90 of which are correct, and 30 of which are incorrect."
37	Consider the training data in the following table where Play is a class attribute. In the table, the Humidity attribute has values "L" (for low) or "H" (for high), Sunny has values "Y" (for yes) or "N" (for no), Wind has values "S" (for strong) or "W" (for weak), and Play has values "Yes" or "No". What is class label for the following day (Humidity=L, Sunny=N, Wind=W), according to naïve Bayesian classification?
38	Develop a step-by-step procedure to cluster the given dataset using the k-means algorithm, initializing the objects 2 and 5 with the coordinate values (4,6) and (12,4) as initial seeds.
39	Develop a solution using the Weighted k-NN algorithm to determine the class of a test instance (7.6, 60, 8) based on a provided training dataset in a table. Set the value of K to 3 and apply the algorithm to calculate the weighted class by considering the distances and weights of the nearest neighbors.
40	Develop a decision tree using the CART algorithm based on the provided training dataset shown in the Table. Apply the CART algorithm to recursively partition the data based on the selected attributes and splitting criteria, resulting in a decision tree that can be used for classification or regression tasks. Also analyze the performance of the developed model.
41	How can reinforcement learning algorithms be used to train autonomous agents in complex environments, such as playing video games or controlling robots?
42	How can unsupervised learning algorithms, such as clustering and dimensionality reduction, be used to extract meaningful insights from large datasets?
43	Explain the concept of bias and fairness in machine learning models, and discuss approaches to mitigate bias and ensure fairness.
44	How can Bayesian inference be used to estimate model parameters and make predictions in machine learning?
45	Explain the concept of ensemble learning and discuss different techniques, such as bagging and boosting, for building ensemble models.
46	Discuss the challenges and potential solutions for handling imbalanced datasets in

	machine learning tasks.
47	Discuss the challenges and potential solutions for training machine learning models on resource-constrained devices, such as smartphones or IoT devices.
48	Discuss the challenges and potential solutions for handling missing data in machine learning tasks.
49	Explain the concept of self-supervised learning and discuss its applications in pretraining deep neural networks.
50	Discuss the challenges and potential solutions for training machine learning models that are robust to adversarial attacks.
51	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 Baye's rule in python to get the result. (Ans: 15%)
52	Extract the data from database using python
53	Implement k-nearest neighbours classification using python
54	Given the following data, which specify classifications for nine Combinations of VAR1 and VAR2 predict a classification for a case where VAR1=0.906 and VAR2=0.606, using the result of k-means clustering with 3 means (i.e., 3 centroids)
55	The following training examples map descriptions of individuals onto high, medium and low credit-worthiness. Input attributes are (from left to right) income, recreation, job, status, age-group, home-owner. Find the unconditional probability of 'golf' and the conditional probability of 'single' given 'medRisk' in the dataset
56	Implement linear regression using python
57	Implement naive Bayes theorem to classify the English text
58	Implement an algorithm to demonstrate the significance of genetic algorithm
59	Implement the finite words classification system using Back-propagation algorithm
60	Given the following vectors: A = [1, 2, 3, 4, 5, 6, 7, 8, 9 10] B = [4, 8, 12, 16, 20, 24, 28, 32, 36, 40] C = [10, 9, 8, 7, 6, 5, 4, 3, 2, 1] Ex. 1: Find the arithmetic mean of vector A, B and C Ex. 2: Find the variance of the vector A, B and C Ex. 3: Find the Euclidean distance between vector A and B Ex. 4: Find the correlation between vectors A & B and A & C
61	Load breast cancer dataset and perform classification using Euclidean distance. Use 70% data as training and 30% for testing.
62	To install the pypi libraries for Machine Learning.
63	Review of python datatypes for Artificial Intelligence and Machine Learning
64	Develop a machine learning model for standard database using Support Vector Machines
65	Develop a machine learning model for standard database using Decision Trees.
66	Develop a machine learning model for standard database using Random Forest.
67	Deploy standardization and normalization on some standard dataset.
68	Deploy Principal Component Analysis to extract relevant features on some standard database.
69	Understanding of github and conda environments.

70	To use the github packages and libraries to frame a standard project and commit back to github.
71	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 csv file.
72	For a given set of training data example stored in .csv file, implement and demonstrate the Candidate-Elimination Algorithm to output and describes the set of all hypotheses consistent with training example.
73	Write a 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.
74	Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets
75	Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets
76	Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your data set
77	Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set. You can use Java/Python ML library classes/API.
78	Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment on the quality of clustering. Add Python ML library classes/API in the program
79	Write a program to implement K-nearest neighbour algorithm to classify iris dataset. Print both correct and wrong predication using python machine learning
80	Implement the non-parametric Locally Weighted Regression algorithm to fit data points. Select appropriate data set for your experiment and draw graphs.

STUDENT-CENTEREDLEARNING (SELF-LEARNING TOWARDS LIFE-LONG-LEARNING)

Self-Learning (it's a typical course-based project to be carried out by a whole class in groups of four students each; they should exhibit higher level BTLs)

The students, in a group, are expected to conceive an idea based on the content(objectives/outcomes) and apply the suitable knowledge to demonstrate their learning.

A list of 30-40 project statements can be offered to the students to choose or students can conceive their own ideas(teamwork),design and develop the product/process/service and implement the same.

A) COURSE-BASED PROJECT (Psychomotor skills)

To enhance their skill set in the integrated course, the students are advised to execute course-based

design projects. Some sample projects are given below:

1.   Amazon Alexa Reviews Sentiment Analysis

Amazon Alexa is a cloud-based voice service developed by Amazon that allows

customers to interact with technology. There are currently over 40 million Alexa users around the world, so analyzing user sentiments about Alexa will be a good data science project. So, if you want to learn how to analyze the sentiments of users using Amazon Alexa, this article is for you. The machine learning project of Amazon Alexa Reviews Sentiment Analysis Using Python can be a good option.

2.   Amazon Product Reviews Sentiment Analysis

Amazon is an American multinational corporation that focuses on e-commerce, cloud computing, digital streaming, and artificial intelligence products. But it is mainly known for its e-commerce platform which is one of the biggest online shopping platforms today. There are so many customers buying products from Amazon that today Amazon earns an average of $ 638.1 million per day. So, for having such a large customer base, it will turn out to be an amazing data science project if we can analyze the sentiments of Amazon product reviews. So, the Amazon Product Reviews Sentiment Analysis project with Python can be the best option for you.

3.   Amazon Recommendation System

Recommendation Systems are one of the widely used applications of Data Science in most companies based on products and online services. Amazon is a great example of such companies. Being an online shopping website, Amazon needs to generate personalized recommendations to provide a better user experience. The Recommendation System of Amazon follows the principle of generating product-based recommendations which means you get to measure the similarities between two products and recommend the most similar products to the user. The methods of measuring similarities between two products have always been a major focus of researchers.

4.   Autocorrect Keyboard with Python and Machine Learning

Almost every smartphone brand irrespective of its price provides an autocorrect feature on their keyboards. In the context of machine learning, autocorrect is based on natural language processing. As the name suggests it is programmed to correct spellings and errors while typing. The Autocorrect model is programmed to correct spellings and errors while inputting text and locating the most comparable and related words. It is completely based on NLP that compares the words in the vocabulary dictionary and the typed words on the keyboard. If the typed word is found in the dictionary, the autocorrect feature assumes you typed the correct term. If the word does not exist, the tool identifies the most comparable words from the smartphone's history.

5.   Automatic License Number Plate Recognition System

This project aims to recognize license number plates. In order to detect license number plates, you will be using OpenCV to identify number plates and python pytesseract to extract characters and digits from the number plates. OpenCV is an open-source machine learning library and provides a common infrastructure for computer vision. Pytesseract is a Tesseract-OCR Engine to read image types and extract the information

present in the image.

6. Automatic Time Series Forecasting

Automatic Time Series Forecasting is a forecast of future values generated over time from past data. Think of how the price changes every day for your favorite stock. Time-series forecasting is can predict the price of that stock over multiple time periods. For example, forecasting what Tesla's stock price will be for the next 60 days or across other time periods. Other examples of time-series data are the weekly numbers of account signups, daily revenue, hourly transactions, and so on.

7. Build a Collaborative Filtering Recommender System in Python

Collaborative Filtering is the most common technique used when it comes to building intelligent recommender systems that can learn to give better recommendations as more user information is collected. Collaborative filtering is a technique that can filter out items that a user might like on the basis of reactions by similar users. It searches among a large group of people and finds a smaller set of users with tastes similar to a particular user. It looks at the items they like and combines them to create a ranked list of suggestions. There are many ways to decide which users are similar and combine their choices to create a list of recommendations.

8. Build Classification Algorithms for Digital Transformation [Banking]

Bank XYZ has a growing customer base where the majority of them are liability customers (depositors) vs. borrowers (asset customers). The bank is interested in expanding the borrower's base rapidly to bring in more business via loan interests. A campaign that the bank ran in the last quarter showed an average single-digit conversion rate. Digital transformation is the core strength of the business strategy – devising effective campaigns with better target marketing to increase the conversion ratio to double-digit with the same budget as per the last campaign. As a data scientist, you are asked to develop a machine learning model to identify potential borrowers to support focused marketing. Build a machine learning model to perform focused digital marketing by predicting the potential customers who will convert from liability customers to asset customers.

9. Build CNN for Image Colorization using Deep Transfer Learning

Image colorization is the process of taking an input grayscale (black and white) image and then producing an output colorized image that represents the semantic colors and tones of the input. In image colorization, a color is assigned to each pixel of a target grayscale image. Image colorization technique is helpful for developing many applications such as Medical Microscope, Medical Imagery, Denoising and recreating old Images, Night Vision Camera, etc. For example, you can use a fully automated data-driven technique called autoencoders for image colorization. Autoencoders are a specific type of feedforward neural network where the input is the same as the output. The VGG16 model will be used as a feature extractor. VGG16 is a classic neural network used as a backbone for many computer vision tasks. This project aims to build a

convolutional neural network that will best convert the grayscale images to RGB images.

10. Catching Illegal Fishing Project

Overfishing and illegal fishing are becoming big problems around the world. For example, there are many records of intensive and often illegal fishing in West African waters by Asian and European fleets that reduce the regular catch for the local populations, increasing their poverty levels. "Being able to see which vessels are fishing where would be a tremendous help in reducing illegal fishing," says Josephus Mamie, head of Sierra Leone's Fisheries Research Unit. In this project, you will collaborate with Global Fishing Watch to detect fishing activity in the ocean using data from the satellite Automatic Identification System (AIS) collected from different vessels around the world. The AIS data contains the latitude, longitude, speed, and course of the vessels at different times.

11. Census Income Dataset Project

A census is the process of collecting, compiling, and publishing demographic, economic, and social data pertaining to a specific time to all persons in a country or delimited part of a country. As part of a census count, most countries also include a census of housing. It is the process of collecting, compiling, and publishing information on buildings, living quarters, and building-related facilities such as sewage systems, bathrooms, and electricity, to name a few. In this project, we will use a standard imbalanced machine learning dataset referred to as the "Adult Income" or simply the "adult" dataset. For census salary prediction, we have to classify salaries that fall within a specified range. This mainly helps to understand the real estate demands and also, the demands for basic amenities based on one's salary range.

12. Classification with Neural Networks

Autoencoders are the simplest of the deep learning architectures. They are a specific type of feedforward neural network where the input is first compressed into a lower-dimensional code. Then, the output is reconstructed from the compact code representation or summary. Therefore, autoencoders have three components built inside them – encoder, code, and decoder. To begin the development process, you will need an encoding method, a decoding method, and a loss function. Binary cross-entropy and mean squared error are the two top choices for the loss function. And to train the autoencoders, you can follow the same procedure as artificial neural networks via back-propagation. Now, let us discuss the applications of these networks.

13. Count Objects in Image

Counting objects in an image is a task of computer vision. There are many computer vision libraries that you can use for this task, such as OpenCV, TensorFlow, PyTorch, Scikit-image, and cvlib. You must have not heard much about the cvlib library in Python. Well, this is a very simple, high-level, and easy-to-use computer vision library in Python. By using the features of this library, we can count the number of objects in an image using Python. To use this library, make sure you have OpenCV and TensorFlow installed

in your systems. You can easily install it by using the pip command; pip installs cvlib.

14.   Coupon Purchase Prediction

Recruit Ponpare is Japan's leading joint coupon site, offering huge discounts on everything from hot yoga to gourmet sushi, to a summer concert bonanza. Ponpare's coupons open doors for customers they've only dreamed of stepping through. They can learn difficult to acquire skills, go on unheard-of adventures, and dine like (and with) the stars.

Using past purchases and browsing behavior, this competition asks you to predict which coupons a customer will buy in a given period of time. The resulting models will be used to improve Ponpare's recommendation system, so they can make sure their customers don't miss out on their next favorite thing.

15.   Covid-19 Detection

The current COVID-19 pandemic threatens human life, health, and productivity. AI plays an essential role in COVID-19 case classification as we can apply machine learning models on COVID-19 case data to predict infectious cases and recovery rates using chest x-ray. Accessing patients' private data violates patient privacy and the traditional machine learning model requires accessing or transferring whole data to train the model. In recent years, there has been increasing interest in federated machine learning, as it provides an effective solution for data privacy, centralized computation, and high computation power.

16.   Currency Exchange Rate Prediction

Predicting the currency exchange rates is the regression problem in machine learning. There are changes in exchange rates every day that affect the income of a person, or a business and can even affect the economy of a country. Thus, predicting the currency exchange rates can help an individual as well as a country in many ways. There are so many machine learning algorithms that we can use to predict future currency exchange rates. You can also use artificial neural networks for this task.

17.   Customer Churn Prediction Analysis

A well-known bank has been observing a lot of customers closing their accounts or switching to competitor banks over the past couple of quarters. This has caused a huge dent in their quarterly revenues and might drastically affect annual revenues for the ongoing financial year, causing stocks to plunge and market cap to reduce significantly. The idea is to be able to predict which customers are going to churn so that necessary actions/interventions can be taken by the bank to retain such customers.

In this machine learning churn prediction project, we are provided with customer data pertaining to his past transactions with the bank and some demographic information. We use this to establish relations/associations between data features and customers' propensity to churn and build a classification model to predict whether the customer will leave the bank or not. We also go about explaining model predictions through

multiple visualizations and give insight into which factor(s) are responsible for the churn of the customers. This project walks you through a complete end-to-end cycle of a data science project in the banking industry, right from the deliberations during the formation of the problem statement to making the model deployment-ready.

18. Deep Learning Project for Text Detection in Images using Python

Text detection is the process of detecting the text present in the image. Several applications include solving the captcha, identifying vehicles by reading their license plates, etc. Convolutional neural networks are deep learning algorithms that are very powerful for the analysis of images. On the other hand, Recurrent Neural Networks (RNNs) are used for sequential data such as text. RNNs are ideal for solving problems where the sequence is more important than the individual items themselves. This model is best used for images with a single line of text in them so we will build this model on images with single-line texts. This project aims to build a convolutional recurrent neural network that can detect the text from a given image.

19. Deepfake Detection with Machine Learning

Machine learning algorithms such as a generative adversarial network (GAN) can be used to create deepfakes. Discriminative models can be used as a method of detecting deepfake videos. Generative adversarial networks (GANs) are an approach to training generative models, in which two neural nets work together to generate fake images that look real and have never been seen before. The first network is called the "generator" and it creates new fakes. The second network is called the "discriminator," which tries to detect whether the images are real or fake. The discriminative model can be used as a method of detecting deepfake videos by using adversarial learning techniques for example, where an attacker's system trains itself on examples of deepfake videos in order to fool a detector.

20. Driver Demand Prediction

Food delivery supported through advanced applications has emerged as one of the fastest-growing developments in the e-commerce space. We all love to order online, one thing that we don't like to experience is variable pricing for delivery charges. Delivery charges highly depend on the availability of riders in your area, the demand of orders in your area, and the distance covered. Due to driver unavailability, there is a surge in delivery pricing and many customers drop off resulting in loss to the company. To tackle such issues if we track the number of hours a particular delivery executive is active, we can efficiently allocate certain drivers to a particular area depending on demand.

21. Emojify – Create your own emoji with Python

Emojis or avatars are ways to indicate nonverbal cues. These cues have become an essential part of online chatting, product review, brand emotion, and many more. It also led to increasing data science research dedicated to emoji-driven storytelling.

With advancements in computer vision and Machine learning, it is now possible to

detect human emotions from images. In this deep learning project, we will classify human facial expressions to filter and map corresponding emojis or avatars.

22. End-to-end Fake News Detection System

The end-to-end fake news detection system is one of the top machine learning projects for students or aspiring tech professionals to work on. This machine learning project uses the applications of natural language processing approaches to detect fake news efficiently and effectively. The model can be developed on the basis of the count vectorizer as well as a TFIDF matrix (Term Frequency Inverse Document Frequency). Thus, developers can assemble a dataset consisting of real news and fake news for the model to successfully use the end-to-end fake news detection system.

23. Text Classification with Transformers-RoBERTa and XLNet Model

The text classification with transformers like RoBERTa and XLNet model machine learning project helps to have a deep understanding of how to load, fine-tune, as well as evaluate multiple transformer models for different text classification tasks. The data here needs to be put in tsv format with four columns including guide, lable, alpha, and text, and without any header.

24. Fake Currency Detection with Machine Learning

Fake currency is roaming around in the financial sector more often than before. Thus, one needs to know how to use machine learning for fake currency detection. It is known as a task of binary classification in machine learning to differentiate between fake and real currency all the time without any failure. The dataset usually consists of four input characteristics such as the variance of the image transformed into wavelets, asymmetry of the image transformed into wavelets, Kurtosis of the image transformed into wavelets, as well as image entropy.

25. Fraud Detection with Machine Learning

Fraud detection with machine learning models and algorithms is useful for tech professionals to work with for a better hands-on experience. It also includes the use of supervised learning, unsupervised learning, semi-supervised learning, as well as reinforcement learning. This project of effective fraud detection with machine learning can become a key tool to prevent cybercrimes.

26. Gold Price Prediction

For the gold price prediction machine learning project, one needs to use machine learning regression techniques for predicting the accurate price of one of the most valuable and flourishing metals across the world, gold. The model is known for gaining relevant information from the past Gold ETF prices and providing accurate gold price predictions for the very next day. The data may consist of the daily Gold ETF price for the last 12 years to predict the Gold ETF close price.

27. Google Play Store Sentiment Analysis

Google Play Store sentiment analysis project uses both machine learning and Python. The sentiment analysis task is to analyze any customer's reviews and comments by

downloading relevant datasets from Kaggle. It is needed to add three new columns in the dataset for a better understanding of the sentiments of each customer review categorized as positive, negative, as well as neutral.

28.   Handwriting Recognition

Handwriting recognition is one of the top machine learning projects that are very useful for different purposes in the future. The machine learning algorithm is known for performing handwriting recognition and can recognize characteristics from different media such as images, and touch-screen devices, and convert them to a machine-readable format. There are three categories of the character recognition algorithms such as image pre-processing, feature extraction, as well as classification.

29.   Hate Speech Detection

Developers must build one hate speech detection machine learning project with the integration of Python-based NLP machine learning techniques. The NLP technique is known as Tf-Idf vectorization for extracting relevant keywords that are popular for conveying the importance of hate speech. Logistic regression helps to train computers to classify hate speech with the data extracted from any library or repository.

30.   Heart Disease Prediction

This machine learning project is highly useful for the healthcare sector across the world for predicting heart disease to save the life efficiently. The project on predicting a heart disease consists of multiple machine learning algorithms such as neighbors classifier, decision tree classifier, support vector classifier, as well as random forest classifier. One can use multiple libraries for a better understanding of data and different numbers of algorithms to vary their multiple parameters to compare the final model efficiently and effectively.

31.   Hotel Reviews Sentiment Analysis

Developers can use sentiment analysis and machine learning algorithms to classify hotel reviews provided by customers from top leading travel sites. There are multiple techniques such as Naïve Bayes, support vector machine, logistic regression, and many more with an ensemble learning model to combine five classifiers and the result will be appropriate. One can collect necessary hotel reviews sentiment analysis from Kaggle and many other places for gaining data such as hotel services during vacations, and business trips.

32.   Human Activity Recognition using Smartphone Dataset

Human activity recognition(HAR) has wide applications in medical research and human survey system. In this project, we design a robust activity recognition system based on a smartphone. The system uses a 3-D smartphone accelerometer as the only sensor to collect time-series signals. This work focuses on the recognition of human activity using smartphone sensors using different machine learning classification approaches. Data retrieved from smartphones accelerometer and gyroscope sensors are classified to recognize human activity. Results of the approaches used are compared in terms of

efficiency and precision.

33. Inventory Demand Forecasting

Inventory forecasting is also known as demand planning, is the practice of using past data and trends. Inventory demand forecasting is the process of predicting customer demand for an inventory item over a defined period. Accurate inventory demand forecasting enables a company to hold the right amount of stock, without over or under-stocking, for optimum inventory control. This project works by helping companies strike a balance between having too much cash tied up in inventory and having enough stock to meet demand.

34. Language Classification Model

Language classification is the grouping of associated languages in the same category. There are two main kinds of language classification: genealogical and typological classification. Languages are grouped diachronically into language families. language classification is solved here is the classification of text into three possible languages like English, Dutch, and Afrikaans. The goal is to devise a machine-learning algorithm to analyze short conversations extracted from your Project corpora, and automatically classify them according to the language of the conversation.

35. Loan Eligibility Prediction

A loan eligibility prediction using several ML algorithms. The dataset with features, namely, gender, marital status, education, number of dependents, employment status, income, co-applicants income, loan amount, loan tenure, credit history, existing loan status, and property area, are used for determining the loan eligibility regarding the loan sanctioning process. This project uses SQL and Python to build a predictive model on GCP to determine whether an application requesting a loan is eligible or not. This application is working properly and meeting all Banker requirements.

36. MLOps Project for a Mask R-CNN on GCP using uWSGI Flask

MLOps advocates automation and monitoring at all steps of the ML system. It refers to Machine learning operations that represent different methodologies, techniques, and procedures used to automate the deployment and handling of machine learning algorithms. This project is to provide hands-on experience in MLOps by using cloud computing. Google cloud platform is used as a cloud provider. We would advise you to have a basic understanding of Image Segmentation using Mask R-CNN with Tensorflow before jumping into this project.

37. Movie Recommendation System using Machine Learning

Recommendation systems are pretty common these days. Netflix, Prime Video, YouTube, and other streaming platforms use these recommendation systems to suggest a movie that you might like to watch according to your previous watch history. A movie recommendation system is an ML-based approach to filtering or predicting the users' film preferences based on their past choices and behavior. It's an advanced filtration mechanism that predicts the possible movie choices of the concerned user and their

preferences towards a domain-specific item.

38. Music Recommendation System ML Project

This project will build a music recommendation system using real datasets. It will utilize a dataset sourced from outside called Million Songs Dataset which contains two files: triplet_file and metadata_file. The triplet_file has within it information regarding the singer or band's name, the title of the song, and how long the song lasts. It can actually understand the musical patterns of a listener with their playlist as the source and what factors are really useful in determining the taste and interest of the listener.

39. Netflix Stock Price Prediction

Netflix is one of the most popular OTT streaming platforms. It offers a vast collection of television series and films and owns its productions known as Netflix Originals. To predict the stock prices of Netflix with machine learning, using the LSTM neural network as it is one of the best approaches for regression analysis and time series forecasting. People who are highly active in stock market investments always keep an eye on companies like Netflix because of its popularity. Machine learning has significant applications in stock price prediction.

40. Network Security Analysis

It is an independent evaluation of the informational security of network infrastructure and the preparation of recommendations on raising the security level of the network infrastructure for the best international practices of providing informational security. Network security is the general practice of protecting computer networks and devices accessible to the network against malicious intent, misuse, and denial. This project is an improvement over traditional network intrusion detection. This dataset is widely used by security data science professionals to classify problems of Network Security.

41. Next Word Prediction Model

It is the task of predicting what word comes next. It is one of the fundamental tasks of NLP and has many applications. Google also uses the next word prediction model based on our browsing history. Google uses our browsing history to make next-word predictions, smartphones, and all the keyboards that are trained to predict the next word are trained using some data. Machine Learning model for next word prediction using Python.

42. NLP-based ChatBot

NLP enables the computer to acquire meaning from inputs given by users. It is a branch of informatics, mathematical linguistics, ML, and AI. This project is a computer program or artificial intelligence that communicates with a customer via textual or sound methods. Such programs are often designed to support clients on websites or via phone. The chatbots are generally used in messaging applications like Slack, Facebook Messenger, or Telegram. NLP-based chatbot project allows users to make chatbots by themself. This is a popular solution for those who do not require complex and sophisticated technical solutions.

43. Object Detection

This project takes an image as input and produces one or more bounding boxes with the class label attached to each bounding box. These algorithms are capable enough to deal with multi-class classification and localization as well as to deal with the objects with multiple occurrences. In object detection, the bounding boxes are always rectangular. Object detection is a fascinating field in machine learning, it is used for research purposes. Some applications of object detection are facial recognition, this can also be used to count people for crowd statistics, also used to identify products, or check the quality of a product.

44. Ola Bike Ride Request Demand Forecast

The Ola service industry is growing for the last couple of years and it is expected to grow in near future. This project is about Ola drivers' need to choose where to hang tight for passengers as they can get somebody quickly. Passengers also prefer a quick taxi service whenever needed. So many times people faced problems with taxi booking requests, which sometimes cannot be fulfilled or the wait time for ride arrival is very long due to the unavailability of a nearby Ola.

45. OpenAI Gym in Machine Learning

It is a toolkit that provides a wide variety of simulated environments. OpenAI is an artificial intelligence research company, funded in part by Elon Musk. OpenAI gym for developing and comparing Reinforcement Learning algorithms. Reinforcement learning is an area of machine learning that allows an intelligent agent to learn the best behaviors in an environment by trial and error. The user can choose their robot, environment, action, and rewards for testing their reinforcement learning algorithms in OpenAI Gym.

46. OpenCV Project to Master Advanced Computer Vision Concepts

OpenCV is a huge open-source library for computer vision, machine learning, and image processing. It can process images and videos to identify objects, faces, or even the handwriting of a human. When integrated with various libraries, such as "NumPy," a highly optimized library for numerical operations, the number of weapons increases in your Arsenal, i.e., whatever operations one can do in NumPy can be combined with OpenCV.

47. Plant Species Identification

Image-based methods are considered a promising approach for species identification. A user can take a picture of a plant in the field with the built-in camera of a mobile device and analyze it with an installed recognition application to identify the species or at least receive a list of possible species if a single match is impossible. By using a computer-aided plant identification system also non-professionals can take part in this process. Therefore, it is not surprising that large numbers of research studies are devoted to automating the plant species identification process.

48. Pneumonia Detection

With the growing technological advancements we have in this day and age, it is possible to use tools based on deep learning frameworks to detect pneumonia based on chest x-ray images. The challenge here would be to aid the diagnosis process which allows for expedited treatment and better clinical outcomes. The dataset that will be used for this project will be the Chest X-Ray Images (Pneumonia) from Kaggle. The dataset consists of training data, validation data, and testing data. The training data consists of 5,216 chest x-ray images with 3,875 images shown to have pneumonia and 1,341 images shown to be normal.

49. Predict Credit Default -Credit Risk Prediction Project

Credit default risk is simply known as the possibility of a loss for a lender due to a borrower's failure to repay a loan. Credit analysts are typically responsible for assessing this risk by thoroughly analyzing a borrower's capability to repay a loan — but long gone are the days of credit analysts, it's the machine learning age! Machine learning algorithms have a lot to offer to the world of credit risk assessment due to their unparalleled predictive power and speed. In this article, we will be utilizing machine learning's power to predict whether a borrower will default on a loan or not and to predict their probability of default.

50. Predict Tinder Matches with Machine Learning

The simple Tinder algorithm can swipe left and right based on the recommendations of a pre-trained deep neural network (Machine Learning). Convolutional Neural Networks are used in this process as they recognize objects, places, and people in your photos, signs, people, and lights in self-driving cars, crops, forests, and traffic in aerial imagery, various anomalies in medical images, and all kinds of other useful things. But once in a while, these powerful visual recognition models can also be warped for distraction, fun, and amusement.

51. Predicting Interest Levels of Rental Listings

Finding the perfect place to call your new home should be more than browsing through endless listings. RentHop makes apartment search smarter by using data to sort rental listings by quality. But while looking for the perfect apartment is difficult enough, structuring and making sense of all available real estate data programmatically is even harder. Two Sigma invites you to apply your talents in this recruiting competition featuring rental listing data from RentHop. We will predict the number of inquiries a new listing receives based on the listing's creation date and other features. Doing so will help RentHop better handle fraud control, identify potential listing quality issues, and allow owners and agents to better understand renters' needs and preferences.

52. Real-Time Face Mask Detection

Computer vision and image processing have an extraordinary impact on the detection of the face mask. Face detection has a range of case applications, from face recognition to facial movements, where the latter is required to show the face with extremely high accuracy. As machine learning algorithms progress rapidly, the threats posed by face

mask detection technology still seem effectively handled. This innovation is becoming increasingly important as it is used to recognize faces in images and real-time video feeds. However, for the currently proposed models of face mask detection, face detection alone is a very tough task. In creating more improved facial detectors, following the remarkable results of current face detectors, the analysis of events and video surveillance is always challenging.

53. Real-time Gender Detection System

Gender detection is one of the popular computer vision applications. When you use a camera to detect a person's gender instead of detecting it on a picture, it can be said to be a real-time gender detection system. There are many libraries and frameworks in Python that can be used to create a real-time gender detection system. Some of these libraries include Yolo, Tensorflow, OpenCV, and Cvlib.

54. Resume Parser NLP Spacy Python

This project uses Python's library, SpaCy to implement various NLP (natural language processing) techniques like tokenization, lemmatization, parts of speech tagging, etc., for building a resume parser in Python. And, considering all the resumes are submitted in PDF format, you will learn how to implement optical character recognition (OCR) for extracting textual data from the documents. The resulting application will require minimum human intervention to extract crucial information from a resume, such as an applicant's work experience, name, geographical location, etc.

55. Social Media Ads Classification

Classifying social media ads means analyzing their social media ads to find the most profitable customers for the products who are most likely to buy the product. Since not all products suit the target audience, using this project can help the enterprise leaders to determine whether a person will buy their product or not by classifying their social media ads. The dataset that will be used for the social media ads classification would contain information about a product's media advertising campaign that would enable the analysts to predict whether the target audience has purchased the product or not.

56. Social Media Followers Prediction

The activity of social media followers' prediction is quite crucial for content creators or business leaders. To predict the increase or decrease in the number of social media followers, professionals need to use a dataset that will yield the type of activities that the people on your social media perform. It is very difficult to find such a dataset on social media platforms like Facebook and Instagram since these platforms do not provide much data related to the followers. So, professionals need to find different sources from which they can gather this distinct information.

57. Speech Emotion Recognition

Emotion recognition is quite a challenging task, but emotions are not constant. A speech emotion recognition system is a collection of methodologies that process and classify speech signals to detect emotions embedded in them. These SER systems can

be used in a wide variety of application areas like an interactive voice-based assistant or caller-agent conversation analysis. The problem of speech emotion recognition can be solved by analyzing one or more of these features. With the help of this project, professionals can leverage machine learning to obtain the underlying emotion from speech audio data and some insights into the human expression of emotion through voice.

58. Survival Prediction on the Titanic Ship

Survival prediction of the Titanic ship is a whole process of building a machine learning model on the Titanic dataset that is used by several people all over the world. It provides information on the fate of passengers on the Titanic, and the collection of data is summarized according to economic status, gender, age, and survival. This machine learning project has revealed that almost 60% of the passenger in the first class survived. Less than 30% of the passengers in the third class survived, which indicates that almost half of the passengers that were originally on the Titanic would have survived.

59. Topic Modeling with Python

Topic modeling lets developers implement helpful features like detecting breaking news on social media, recommending personalized messages, detecting fake users, and characterizing information flow. Text mining or topic modeling generally indicates the process of automatically identifying topics present in a text object and deriving hidden patterns exhibited by a text corpus. Hence, it better assists better decision-making. It is basically an unsupervised approach used for finding and observing a bunch of words in large clusters of texts.

60. Uber Data Analysis Project

Uber is responsible for moving data for people and making hassle-free deliveries. Uber collects large volumes of data and its fantastic team handles Uber data analysis using machine learning tools and frameworks. Creating this project would help Uber customers and riders to get positive experiences. But before starting with any machine learning project, one should know that it is essential to realize how prevalent the exercise of exploratory data analysis is in these projects.

61. Ultrasound Nerve Segmentation

Even the bravest people can get squeamish at the very name of surgeries. It can cause discomfort and surgical pain. But using this machine learning, cum data science project one can identify nerve structures in a data set of ultrasound images. This will help enhance catheter placement and contribute to a more pain-free future for treatments. The creators of this project are working to improve pain management through the use of indwelling catheters that block or mitigate pain at the source.

62. Visualize a Solar System with Python

To complete this project, the participants can use the great selection of graphics libraries in Python which will allow them to plot all major types of graphics so that they

can visualize all the data. Participants can use a combination of Plotly Scatter3D and Plotly Surface plots to create an entire solar system inside a graph. By creating more code lines using Python, this project can be recreated according to the desired results.

63.   Whatsapp Group Chat Analysis

Whatsapp group chat analysis is a machine learning and data science project. The participants will receive a chat at the end of the project will not require any cleaning and preparation and can be used directly for the task. But before starting with the project, they will have to make sure that the data which is ready to use should have the format of the data and time of the messages changed, and it can be done extremely easily.

64.   Zillow Home Value Prediction ML Project

Zillow is a popular online real estate and rental marketplace dedicated to providing customers with data to make the best possible housing decision. The Zestimate was initially created to give consumers as much information as possible about homes and the housing market, marking the first time consumers had access to the type of home value information at no cost. This project would essentially demonstrate to improve the Zestimate residual error which is also called 'log error' that would in turn provide the actual sales price.

B) SELF-LEARNING THROUGH MOOCs (Cognitive Skills): Certification

1. https://www.udemy.com/topic/machine-learning/
2. https://www.coursera.org/learn/machine-learning
3. https://developers.google.com/machine-learning/crash-course
4. https://www.edx.org/learn/machine-learning

Course Lead   **Program Chair**   **Dean-SCSE**