Internal Assesment Test - I

Sub:	Machine Learning					Code:	22MBABA403
Date:	**08.08.2024**	Duration:	**90 minutes**	Max Marks: **50**	Sem: **IV**	Branch:	**MBA**

SET- II

			OBE	
		Marks	CO	RBT
Part A - Answer Any Two Full Questions (2* 20 = 40 marks)				
1 (a) Differentiate compiler and interpreter.		[03]	CO1	L2

Compiler
The Compiler is a translator which takes input i.e., High-Level Language, and produces an output of low-level language i.e. machine or assembly language. The work of a Compiler is to transform the codes written in the programming language into machine code (format of 0s and 1s) so that computers can understand.

    A compiler is more intelligent than an assembler it checks all kinds of limits, ranges, errors, etc.

    But its program run time is more and occupies a larger part of memory. It has a slow speed because a compiler goes through the entire program and then translates the entire program into machine codes.

Interpreter
An Interpreter is a program that translates a programming language into a comprehensible language. The interpreter converts high-level language to an intermediate language. It contains pre-compiled code, source code, etc.

    It translates only one statement of the program at a time.

    Interpreters, more often than not are smaller than compilers.

			CO	RBT
(b) Explain briefly any seven library packages / modules of Python with the purpose		[07]	CO1	L2

- Numpy: For Advanced array operations (e.g. add, multiply, slice, reshape, index), Comprehensive mathematical functions, Random number generation, linear algebra routines, Fourier transforms, etc.

- Pandas: If you work with tabular, time series, or matrix data, pandas is your go-to Python package. It is known as a fast, efficient, and easy-to-use tool for data analysis and manipulation. It works with data frame objects; a data frame is a dedicated structure for two-dimensional data.

- Matplotlib: Matplotlib is the most common data exploration and visualization library. You can use it to create basic graphs like line plots, histograms, scatter plots, bar charts, and pie charts. You can also create animated and interactive visualizations with this library.

- Seaborn: Seaborn is a high-level interface for drawing attractive statistical

graphics with just a few lines of code.

- Scikit-learn: Identify which category an object is likely to belong to (used in fraud detection, image recognition, cancer detection, etc.).Predict a continuous variable based on available features (used in predicting house prices and inflation).Group similar objects into clusters (used in customer segmentation, social network analysis, etc.).

- NLTK: one of the leading Python platforms for processing language data. It is a set of language processing libraries and programs that provide a toolkit for: Classification, Tokenization, Stemming, Tagging, Parsing, Semantic reasoning,

- Keras: Keras is a high-level, user-friendly API used for building and training neural networks. It is designed to be user-friendly, modular, and easy to extend. Keras allows you to build, train, and deploy deep learning models with minimal code.

- Tensorflow: TensorFlow: This library was developed by Google in collaboration with the Brain Team. It is an open-source library used for high-level computations. It is also used in machine learning and deep learning algorithms. It contains a large number of tensor operations. Researchers also use this Python library to solve complex computations in Mathematics and Physics.
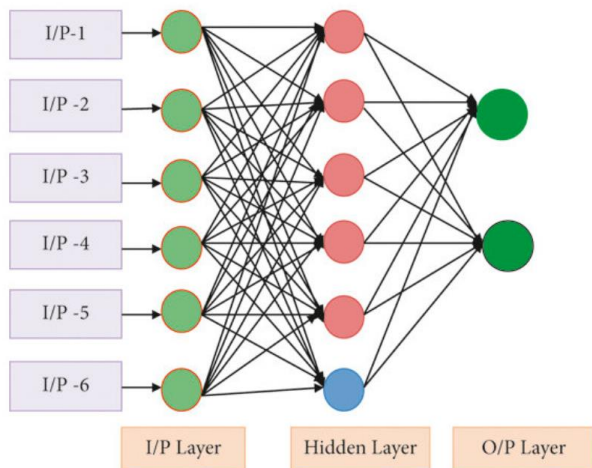
(c) Examine the functioning of a computer system with a block diagram.	[10] CO1	L3

Block Diagram of a Computer
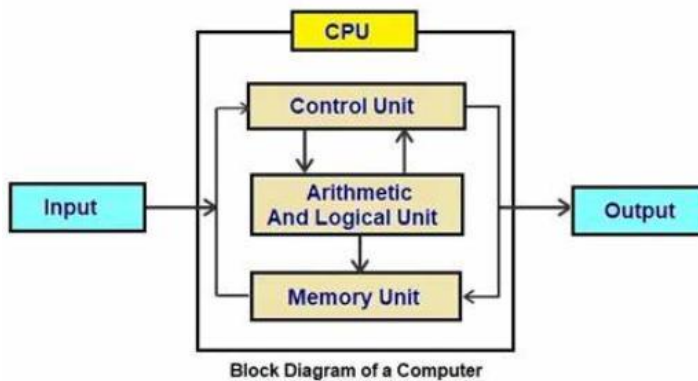
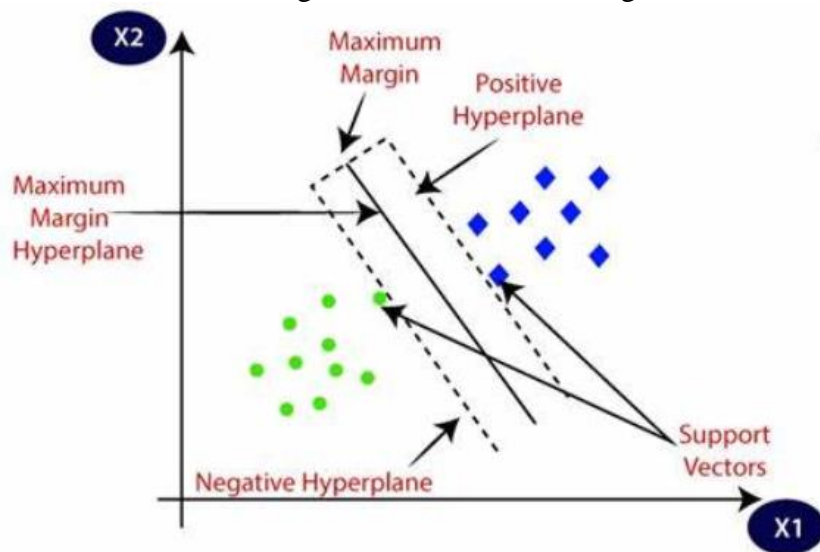|		
2 (a) What is the justification for comparing neural networks with biological systems?	CO1	L1

[03]

(b) Illustrate the functioning of SVM with a neat diagram. [07] CO1 L3

(c) Critically analyze the implementation of K- means clustering vis-a-vis KNN. [10] CO2 L4

- The key distinction between kNN and k−means clustering is that whereas KNN is a supervised learning technique used for classification and regression problems, k−means clustering is an unsupervised learning approach.
- K-Nearest Neighbors is a supervised machine learning algorithm used for classification and regression tasks.
- In KNN, data points are categorized based on the majority class of their k-nearest neighbors. The "k" represents the number of neighboring data points considered when making a decision.
- KNN relies on the similarity of data points in a feature space to classify or predict new data points.
- K-Means Clustering is an unsupervised machine learning algorithm used for grouping data points into clusters.
- K-Means identifies clusters by minimizing the sum of squared distances between data points and their respective cluster centroids.
- It iteratively assigns data points to clusters and updates cluster centroids until convergence is achieved.
- Use-Cases:
  KNN: KNN is commonly used for classification tasks such as text classification, image recognition, and recommendation systems. It's

particularly effective when you have labelled data and want to classify new data points based on their similarity to existing examples.

K-Means Clustering: K-Means clustering is used for tasks like customer segmentation, image compression, and anomaly detection. It's ideal for situations where you want to discover underlying patterns or group data points based on their intrinsic similarities.

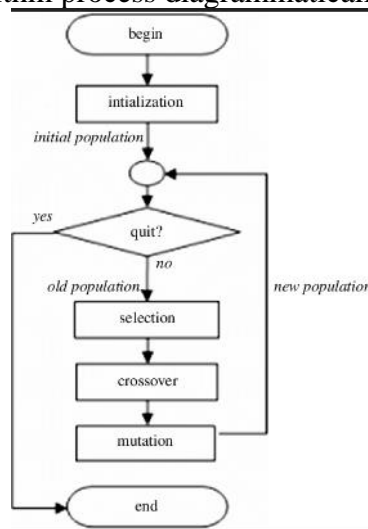3 (a) State the significance of Information gain in DT. [03] CO1 L1

- Information Gain, or IG for short, measures the reduction in entropy or surprise by splitting a dataset according to a given value of a random variable.
- A larger information gain suggests a lower entropy group or groups of samples.
- Information Gain helps us understand how much a particular feature contributes to making accurate predictions in a decision tree.

(b) Outline the Genetic algorithm process diagrammatically. [07] CO1 L4

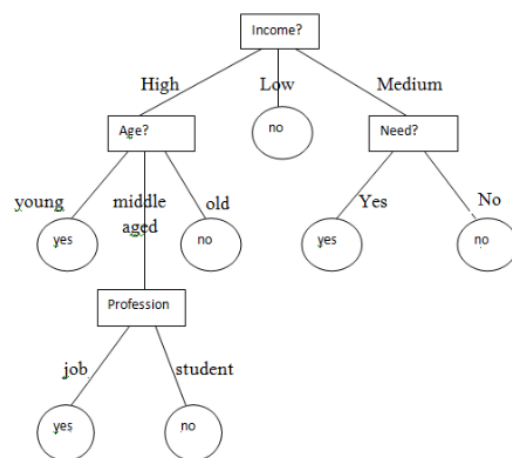(c) Develop a decision tree for identifying the target market size for computer games [10] CO1 L6
in Bangalore.

Part B - Compulsory (01*10=10 marks) – CASE STUDY
Given below is the data of weather conditions in which play/no play decisions were taken.

4.

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

(a) Evaluate the suitability of Naïve Bayes algorithm for the above dataset, if it is to be predicted whether playing happens or not. [5] CO2 L5

During the training phase of the Naïve Bayes algorithm, probabilities for all possible combinations of feature values and classes are calculated and stored in a hashing format. This involves estimating the prior probabilities of classes and conditional probabilities of feature values given each class. In the testing phase, the algorithm retrieves the corresponding probabilities based on the observed feature values, multiplies them together, and provides the final output, indicating the predicted class.

(b) Given the condition of sunny, cool, high humidity and wind false condition, predict the likelihood of playing tennis. [5] CO2 L3

$P(X|C_i) P(C_i)$ for i=1,2
$P(class = P) = 10 / 14 = 0.714$
$P(class = N) = 4 / 14 = 0.285$

To compute $P(X|C_i)$ for i=1,2 we compute the following conditional probabilities.
$P(outlook=Sunny/class = P) = 2/10 = 0.2$
$P(outlook=Sunny/class = N) = 3/4 = 0.75$

$P(temperature = cool/ class = P) = 3 /10 = 0.3$
$P(temperature = cool/ class = N) = 1 /4 = 0.25$

$P(humidity = high / class = P) = 4 /10 = 0.4$
$P(humidity = high / class = N) = 3 /4 = 0.75$

$P(windy = false / class = P) = 6 /10 = 0.6$
$P(windy = false / class = N) = 2 /4 = 0.5$

$P(X| class = P)$
$= P(outlook=Sunny/class = P) * P(temperature = cool/ class = P) *$
$P(humidity = high / class= P) * P(windy = false / class = P)$
$= 0.2 * 0.3 * 0.4 * 0.6$
$= 0.0144$

$P(X| class = N)$
$= P(outlook=Sunny/class = N) * P(temperature = cool/ class = N) *$
$P(humidity = high / class = N) * P(windy = false / class = N)$
$=0.75 * 0.25 * 0.75 * 0.5$
$=0.0703$

To find the class C_i that maximizes $P(X|C_i) P(C_i)$ we compute
$P(X | class = P) P(class = P) = 0.0144 * 0.714 = 0.01028$
$P(X | class = N) P(class = N) = 0.0703 * 0.285 = 0.0200$

The Naive Bayes classifier algorithm predicts class N for tuple X

Course Outcomes (COs)		PO1	PO2	PO3	PO4	PO5	PSO1	PSO2	PSO3	PSO4
CO1:	Understand the concepts of Machine learning	1a,2a,3a	1b,1c,2b,3b				1b,1c,2b,3b			
CO2:	Apply the knowledge of Data visualisation and accurate decision making		2c,3c				2c,3c	4a,4b		
CO3:	Analyse the Big data and pattern using machine learning algorithms									
CO4:	Evaluate the Data Structure and provide immersive experience to users									
CO5:										
CO6:										

Cognitive level	KEYWORDS
L1 - Remember	list, define, tell, describe, recite, recall, identify, show, label, tabulate, quote, name, who, when, where, etc.
L2 - Understand	describe, explain, paraphrase, restate, associate, contrast, summarize, differentiate interpret, discuss
L3 - Apply	calculate, predict, apply, solve, illustrate, use, demonstrate, determine, model, experiment, show, examine, modify
L4 - Analyze	classify, outline, break down, categorize, analyze, diagram, illustrate, infer, select
L5 - Evaluate	asses, decide, choose, rank, grade, test, measure, defend, recommend, convince, select, judge, support, conclude, argue, justify, compare, summarize, evaluate
L6 - Create	design, formulate, build, invent, create, compose, generate, derive, modify, develop, integrate

PO1–*Theoretical Knowledge;* PO2–*Foster Analytical and Critical Thinking Abilities for data based decision making;*
PO3– *Develop Value Based Leadership;* PO4 –*Ability to Understand and communicate various business aspects to global;*
PO5 – *Ability to lead themselves and others in the achievement of organizational goals contributing effectively to a team environment;*
PSO1- *Comprehend Contemporary features of Business Management Science and its administration*
PSO2- *Analyze and interpret the dynamic situations for making Business Management strategies*
PSO3- *Handle responsibility with the ethical values for all actions undertaken by them*
PSO4- *Adapt and focus on achieving the organizational goal and objectives with complete zeal and commitment.*

CI   CCI   HOD