USN										

Internal Assessment Test 1 - Mar 2025

Sub:	**Machine Learning**					Sub Code:	**BCS602**	Branch:	**ISE**
Date:	**24/03/2025**	Duration:	**90 min**	Max Marks:	**50**	Sem/Sec:	**VI / A, B & C**		**OBE**

Answer any FIVE FULL Questions

		MARKS	CO	RBT
1a	Define Machine Learning, Data, Information, Knowledge, and Intelligence with an example. Explain and Sketch the Knowledge pyramid representation	[06]	CO1	L1
1b	Consider the set: V = { 42, 80, 82, 98 }, Apply Min-Max procedure and map the marks to a new range 0-1.	[04]	CO1	L3
2a	Explain in detail about different types of Machine Learning with examples for each type.	[8]	CO1	L2
2b	Do the stem and Leaf plot for the following English Marks {46, 53, 72, 73, 83, 85,92 }	[2]	CO1	L3
3a	List out the challenges of Machine Learning in detail	[05]	CO1	L2
3b	For the student's age list { 15,23,25,34,36,47,58,59,68 }, Find the IQR	[05]	CO2	L3
4a	Explain Big Data Analytics and its Types with examples	[05]	CO1	L2
4b	Explain Flat files and list the popular spreadsheet formats with the relevant examples.	[05]	CO2	L3
5	Let the data points be $\begin{pmatrix}2\\6\end{pmatrix}$ and $\begin{pmatrix}1\\7\end{pmatrix}$. Apply PCA and find the transformed data. Apply reverse and prove that PCA Works.	[10]	CO2	L3
6a	Apply Find-S algorithm for the below training dataset consists of 5 instances.	[07]	CO2	L3
6b	Find the Covariance of the data X={ 4,5,6,7,9}, and Y = { 16,25,36,49,56}	[03]	CO2	L3

Eyes	Nose	Head	Fcolor	Hair	Smile
Round	Triangle	Round	Purple	Yes	Yes
Square	Square	Square	Green	Yes	No
Square	Triangle	Round	Yellow	Yes	Yes
Round	Triangle	Round	Green	No	No
Square	Square	Round	Yellow	Yes	Yes

Faculty Signature CCI Signature HOD Signature

USN										

Internal Assessment Test 1 Scheme & Solution - Mar 2025

Sub:	**Machine Learning**					Sub Code:	**BCS602**	Branch:	**ISE**
Date:	**24/03/2025**	Duration:	**90 min**	Max Marks:	**50**	Sem/Sec:	**VI / A, B & C**		**OBE**

Answer any FIVE FULL Questions

		MARKS	CO	RBT
1	Define Machine Learning, Data, Information, Knowledge, and Intelligence with an example. Explain and Sketch the Knowledge pyramid representation	[06]	CO1	L1

- Processed data is called information. This includes patterns, associations, or relationships among data.
- For example, sales data can be analyzed to extract information like which is the fast selling product.
- Condensed information is called knowledge.
- For example, the historical patterns and future trends obtained in the above sales data can be called knowledge.
- Unless knowledge is extracted, data is of no use. Similarly, knowledge is not useful unless it is put into action.
- Intelligence is the applied knowledge for actions. An actionable form of knowledge is called intelligence.
  Computer systems have been successful till this stage. — 4m
- The ultimate objective of knowledge pyramid is wisdom that represents the maturity of mind that is, so far, exhibited only by humans.
- Here comes the need for machine learning. The objective of machine learning is to process these archival data for organizations to take better decisions to design new products, improve the business processes, and to develop effective decision support systems.

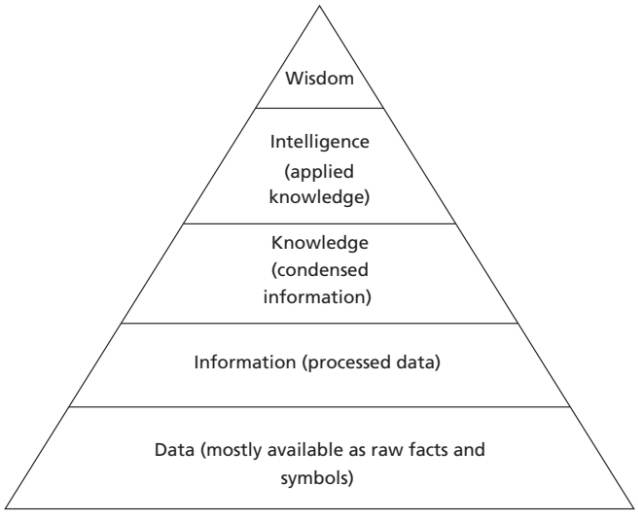
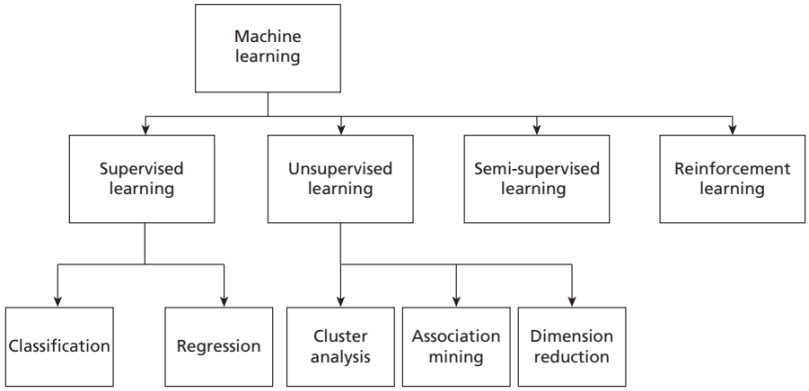
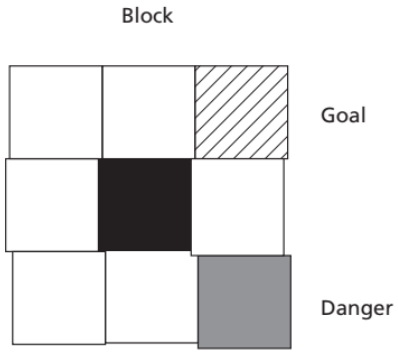


Figure 1.1: The Knowledge Pyramid — 2m

The objective of machine learning is to process these data for organizations to take better decisions to **design new products, improve the business processes,** and to **develop effective decision** support systems.

	Consider the set: V = { 42, 80, 82, 98 }, Apply Min-Max procedure and map the marks to a new range 0-1. For 42 : 0 For 80 : 0.678 For 82 : 0.714 For 98 : 1	[04] 1m 1m 1m 1m	CO1	L3
2	Explain in detail about different types of Machine Learning with examples for each type.  **Figure 1.5:** Types of Machine Learning Supervised Learning • Supervised algorithms use labelled dataset. As the name suggests, there is a supervisor or teacher component in supervised learning. • A supervisor provides labelled data so that the model is constructed and generates test data. • In supervised learning algorithms, learning takes place in two stages. • In layman terms, during the first stage, the teacher communicates the information to the student that the student is supposed to master. • The student receives the information and understands it. • During this stage, the teacher has no knowledge of whether the information is grasped by the student. Supervised learning has two methods: 1. Classification 2. Regression Unsupervised Learning • The second kind of learning is by self-instruction. • There are no supervisor or teacher components. • In the absence of a supervisor or teacher, self-instruction is the most common kind of learning process. • This process of self-instruction is based on the concept of trial and error.	[8] 1m 2m 2m	CO1	L2

	• Here, the program is **supplied with objects, but no labels are defined**. • The algorithm itself observes the examples and recognizes patterns based on the principles of grouping. • Grouping is done in ways that similar objects form the same group. • Cluster analysis and Dimensional reduction algorithms are examples of unsupervised algorithms. **Semi-supervised Learning** • There are circumstances where the dataset has a huge collection of unlabelled data and some labelled data. • Labelling is a costly process and difficult to perform by the humans. • Semi-supervised algorithms use **unlabelled data by assigning a pseudo-label.** • Then, the labelled and pseudo-labelled dataset can be combined. Reinforcement Learning • Reinforcement learning mimics human beings. • Like human beings use ears and eyes to perceive the world and take actions, reinforcement learning allows the agent to interact with the environment to get rewards. • The agent can be a human, animal, robot, or any independent program. • The rewards enable the agent to gain experience. • The agent aims to maximize the reward. The reward can be positive or negative (Punishment). When the rewards are more, the behavior gets reinforced and learning becomes possible. • Consider the following example of a Grid game • In this grid game, the gray tile indicates the danger, black is a block, and the tile with diagonal lines is the goal. • The aim is to start, say from bottom-left grid, using the actions left, right, top and bottom to reach the goal state.  	1m 2m		
	Do the stem and Leaf plot for the following English Marks {46, 53, 72, 73, 83, 85,92 }	[2]	CO1	L3

3	List out the challenges of Machine Learning in detail • 1. Problems – Machine learning can deal with the 'well-posed' problems where specifications are complete and available. • Computers cannot solve 'ill-posed' problems. | Input (x_1, x_2)	Output (y) | |---|---| | 1, 1 | 1 | | 2, 1 | 2 | | 3, 1 | 3 | | 4, 1 | 4 | | 5, 1 | 5 | • Huge data – This is a primary requirement of machine learning. Availability of a quality data is a challenge. • A quality data means it should be large and should not have data problems such as missing data or incorrect data. • 3. High computation power – With the availability of Big Data, the computational resource requirement has also increased. • Systems with Graphics Processing Unit (GPU) or even Tensor Processing Unit (TPU) are required to execute machine learning algorithms. • 4.Complexity of the algorithms – The selection of algorithms, describing the algorithms, application of algorithms to solve machine learning task, and comparison of algorithms have become necessary for machine learning or data scientists now. • Algorithms have become a big topic of discussion, and it is a challenge for machine learning professionals to design, select, and evaluate optimal algorithms. • 5. Bias/Variance – Variance is the error of the model. This leads to a problem called bias/ variance tradeoff. • A model that fits the training data correctly but fails for test data, in general lacks generalization, is called overfitting. • The reverse problem is called underfitting where the model fails for training data but has good generalization.	[05]	CO1	L2
	For the student's age list { 15,23,25,34,36,47,58,59,68 }, Find the IQR $Q_1 = 24$, $Q_3=58$, IQR = $Q_3-Q_1 = 34.5$	[05]	CO2	L3
4 (a)	Explain Big Data Analytics and Types of Analytics with examples • The primary aim of data analysis is to assist business organizations to take decisions. • For example, a business organization may want to know which is the fastest selling product, in order for them to market activities. • Data analysis is an activity that takes the data and generates useful information and insights for assisting the organizations. • Data analysis and data analytics are terms that are used interchangeably to refer to the same concept. However, there is a subtle difference. • Data analytics is a general term and data analysis is a part of it. • Data analytics refers to the process of data collection, preprocessing and analysis. It deals with the complete cycle of data management.	[05]	CO1	L2

	• Data analysis is just analysis and is a part of data analytics. It takes historical data and does the analysis. There are four types of data analytics: • 1. Descriptive analytics • 2. Diagnostic analytics • 3. Predictive analytics • 4. Prescriptive analytics Descriptive Analytics • It is about describing the main features of the data. • After data collection is done, descriptive analytics deals with the collected data and quantifies it. • It is often stated that analytics is essentially statistics. • There are two aspects of statistics – Descriptive and Inference. Descriptive analytics only focuses on the description part of the data and not the inference part. • What was our overall productivity? Diagnostic Analytics • It deals with the question – 'Why?'. This is also known as causal analysis, as it aims to find out the cause and effect of the events. • For example, if a product is not selling, diagnostic analytics aims to find out the reason. There may be multiple reasons and associated effects are analyzed as part of it. • Why did our company sales decrease in the previous quarter? Predictive Analytics • It deals with the future. It deals with the question – 'What will happen in future given this data?'. • This involves the application of algorithms to identify the patterns to predict the future. • The entire course of machine learning is mostly about predictive analytics and forms the core of this book. • Predicting maintenance issues, Predicting article popularity Prescriptive Analytics • It is about the finding the best course of action for the business organizations. • Prescriptive analytics goes beyond prediction and helps in decision making by giving a set of actions. • It helps the organizations to plan better for the future and to mitigate the risks that are involved. • Automatic adjustment of product pricing based on customer demand and external factors.			
5	Explain Flat files and list the popular spreadsheet formats with the relevant examples. These are the simplest and most commonly available data source. It is also the cheapest way of organizing the data. These flat files are the files where data is stored in plain ASCII or EBCDIC format. (Extended binary coded decimal interchange code) Minor changes of data in flat files affect the results of the data mining algorithms. Hence, flat file is suitable only for storing small dataset and not desirable if the dataset becomes larger. Some of the popular spreadsheet formats are listed below:	[05]	CO2	L3

	• CSV files – CSV stands for comma-separated value files where the values are separated by commas. These are used by spreadsheet and database applications. The first row may have attributes and the rest of the rows represent the data. • TSV files – TSV stands for Tab separated values files where values are separated by Tab. Both CSV and TSV files are generic in nature and can be shared. There are many tools like Google Sheets and Microsoft Excel to process these files.			
	Let the data points be $\begin{pmatrix}2\\6\end{pmatrix}$ and $\begin{pmatrix}1\\7\end{pmatrix}$ Apply PCA and find the transformed data. Apply reverse and prove that PCA Works. **Solution**: One can combine two vectors into a matrix as follows: The mean vector can be computed as Eq. (2.53) as follows: $$\mu = \begin{pmatrix}\frac{2+1}{2}\\ \frac{6+7}{2}\end{pmatrix} = \begin{pmatrix}1.5\\6.5\end{pmatrix}$$ As part of PCA, the mean must be subtracted from the data to get the adjusted data: $$x_1 = \begin{pmatrix}2-1.5\\6-6.5\end{pmatrix} = \begin{pmatrix}0.5\\-0.5\end{pmatrix}$$ $$x_2 = \begin{pmatrix}1-1.5\\7-6.5\end{pmatrix} = \begin{pmatrix}-0.5\\0.5\end{pmatrix}$$ One can find the covariance for these data vectors. The covariance can be obtained using Eq. (2.54): $$m_1 = \begin{pmatrix}0.5\\-0.5\end{pmatrix}\begin{pmatrix}0.5 & -0.5\end{pmatrix} = \begin{pmatrix}0.25 & -0.25\\-0.25 & 0.25\end{pmatrix}$$ $$m_2 = \begin{pmatrix}-0.5\\0.5\end{pmatrix}\begin{pmatrix}-0.5 & 0.5\end{pmatrix} = \begin{pmatrix}0.25 & -0.25\\-0.25 & 0.25\end{pmatrix}$$ The final covariance matrix is obtained by adding these two matrices as: $$C = \begin{pmatrix}0.5 & -0.5\\-0.5 & 0.5\end{pmatrix}$$ The eigen values and eigen vectors of matrix C can be obtained (left as an exercise) as $\lambda_1 = 1$, $\lambda_1 = 0$. The eigen vectors are $\begin{pmatrix}-1\\1\end{pmatrix}$ and $\begin{pmatrix}1\\1\end{pmatrix}$. The matrix A can be obtained by packing the eigen vector of these eigen values (after sorting it) of matrix C. For this problem, $A = \begin{pmatrix}-1 & 1\\1 & 1\end{pmatrix}$.	[10]	CO2	L3

$$A = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

One can check that the PCA matrix A is orthogonal. A matrix is orthogonal is $A^{-1} = A$ and $AA^{-1} = I$.

$$AA^T = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The transformed matrix y using Eq. (2.55) is given as:

$$y = A \times (x - m)$$

Recollect that $(x–m)$ is the adjusted matrix.

$$y = A(x - m) = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{pmatrix}$$

$$= \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \left(for\ convenience\ 0.5 = \frac{1}{2} \right)$$

$$= \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 \end{pmatrix}$$

One can check the original matrix can be retrieved from this matrix as:

$$\{(A)^T \times y\} + m$$

$$x = A^T y + m = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 1.5 \\ 6.5 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} + \begin{pmatrix} 1.5 \\ 6.5 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 6 & 7 \end{pmatrix}$$

Therefore, one can infer the original is obtained without any loss of information.

6 (a) Apply Find-S algorithm for the below training dataset consists of 5 instances. [07] CO2 L3

Eyes	Nose	Head	Fcolor	Hair	Smile
Round	Triangle	Round	Purple	Yes	Yes
Square	Square	Square	Green	Yes	No
Square	Triangle	Round	Yellow	Yes	Yes
Round	Triangle	Round	Green	No	No
Square	Square	Round	Yellow	Yes	Yes

G = { ‘?’, ‘?’, ‘?’, …..’?’}
S = {‘Φ’,’Φ’,’Φ’, ……,’Φ’}

	Result: Hypothesis h = < ? ? round ? yes >			
	Find the Covariance of the data X={ 4,5,6,7,9}, and Y = { 16,25,36,49,56} E(X) = 6.2, E(Y)=36.4, Cov(X,Y) = 24.588	[03]	CO2	L3

Faculty Signature CCI Signature HOD Signature