CMR INSTIT UTE OF TECHNOLOGY	USN	CMRIT

First Internal Test

Sub:	**Big Data Analytics**					Sub Code:	17CS82	Branch:	**ISE**
Date:		Duration:	90 min's	Max Marks:	50	Sem / Sec:	VIII / A ,B		OBE

	Answer any FIVE FULL Questions	MARKS	CO	RBT
1 (a)	**What are association rules? How do they help?**	[4]	CO5	L1

Solution:

In business environments a pattern or knowledge can be used for many purposes. In sales and marketing, it is used for cross-marketing and cross-selling, catalog design, e-commerce site design, online advertising optimization, product pricing, and sales/promotion configurations. This analysis can suggest not to put one item on sale at a time, and instead to create a bundle of products promoted as a package to sell other non-selling items. In retail environments, it can be used for store design. Strongly associated items can be kept close tougher for customer convenience. Or they could be placed far from each other so that the customer has to walk the aisles and by doing so is potentially exposed to other items. In medicine, this technique can be used for relationships between symptoms and illnesses; diagnosis and patient characteristics/treatments; genes and their functions; etc.

Representing Association Rules

A generic Association Rule is represented between a set X and Y: **XÞ Y [S%,C%]**

X, Y: products and/or services

X: Left-hand-side (LHS)

Y: Right-hand-side (RHS)

S: Support: how often **X** and **Y** go together in the dataset – i.e. P (**X** U **Y**)

C: Confidence: how often **Y** is found, given **X** – i.e. P (**Y** | **X**)

Example: {Hotel booking, Flight booking}Þ {Rental Car} [30%, 60%]

[Note: P (X) is the mathematical representation of the probability or chance of X occurring in the data set.}

Computation example:

Suppose there are 1000 transactions in a data set. There are 300 occurrences of X, and 150 occurrences of (X,Y) in the data set.

Support S for **XÞ Y** will be P(X U Y) = 150/1000 = 15%.

Confidence for **XÞ Y** will be P (**Y** | **X**); or P (**X** U **Y**) / P (**X**) = 150/300 = 50%

1 (b)	**Describe three business applications where cluster analysis will be useful. Write a pseudo code for K-Means algorithm.**	[06]	CO5	L2

Solution:

Applications of Cluster Analysis

Cluster analysis is used in almost every field where there is a large variety of transactions. It helps provide characterization, definition, and labels for populations. It can help identify natural groupings of customers, products, patients, and so on. It can also help identify outliers in a specific domain and thus decrease the size and complexity of problems. A prominent business application of cluster analysis is in market research. Customers are segmented into clusters based on their characteristics—want and needs, geography, price sensitivity, and so on.

Here are some examples of clustering:

1.*Market Segmentation*: Categorizing customers according to their similarities, for instance by their common wants and needs, and propensity to pay, can help with targeted marketing.
2. *Product portfolio*: People of similar sizes can be grouped together to make small, medium and large sizes for clothing items.
3. *Text Mining*: Clustering can help organize a given collection of text documents according to their content similarities into clusters of related topics.

Here is the pseudo code for implementing a K-means algorithm.

Algorithm K-Means (K number of clusters, D list of data points)

1. Choose K number of random data points as initial centroids (cluster-centers)
2. Repeat till cluster-centers stabilize
    a. *{ Allocate each point in D to the nearest of K centroids;*
    b. *Compute centroid for the cluster using all points in*

2	**List the advantages and disadvantages of Regression Models**	[10]	CO5	L1

Solution:
Regression Models are very popular because they offer many advantages.

1. Regression models are easy to understand as they are built upon basic statistical principles such as correlation and least square error.
2. Regression models provide simple algebraic equations that are easy to understand and use.

3. The strength (or the goodness of fit) of the regression model is measured in terms of the correlation coefficients, and other related statistical parameters that are well understood.
4. Regression models can match and beat the predictive power of other modeling techniques.
5. Regression models can include all the variables that one wants to include in the model.
6. Regression modeling tools are pervasive. They are found in statistical packages as well as data mining packages. MS Excel spreadsheets can provide simple regression modeling capabilities.

Regression models can however prove inadequate under many circumstances.

1. Regression models can not cover for poor data quality issues. If the data is not prepared well to remove missing values or is not well-behaved in terms of a normal distribution, the validity of the model suffers.
2. Regression models suffer from collinearity problems (meaning strong linear correlations among some independent variables). If the independent variables have strong correlations among themselves, then they will eat into each other’s predictive power and the regression coefficients will lose their ruggedness. Regression models

will not automatically choose between highly collinear variables, although some packages attempt to do that.

3. Regression models can be unwieldy and unreliable if a large number of variables are included in the model. All variables entered into the model will be reflected in the regression equation, irrespective of their contribution to the predictive power of the model. There is no concept of automatic pruning of the regression model.

4. Regression models do not automatically take care of non-linearity. The user needs to imagine the kind of additional terms that might be needed to be added to the regression model to improve its fit.

5. Regression models work only with numeric data and not with categorical variables. There are ways to deal with categorical variables though by creating multiple new variables with a yes/no value.

Q	Question	Marks	CO	Level
3	**Explain representation, design principles and business applications of Artificial Neural Networks.**	[10]	CO4	L2

Solution:

Business Applications of ANN

Neural networks are used most often when the objective function is complex, and where there exists plenty of data, and the model is expected to improve over a period of time. A few sample applications are:

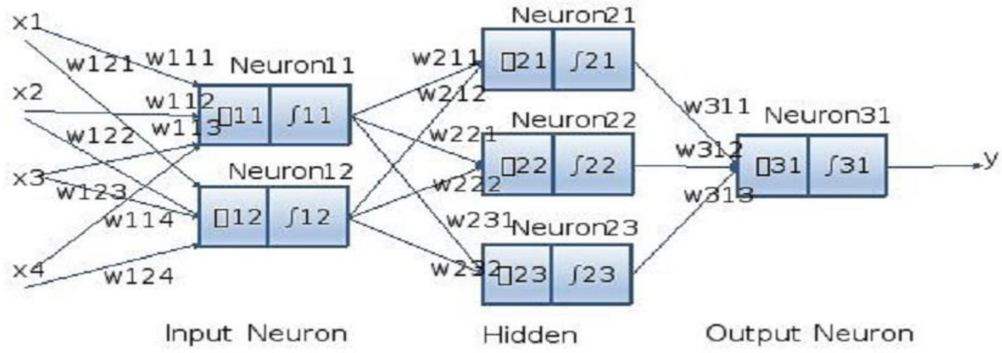
1. They are used in stock price prediction where the rules of the game are extremely complicated, and a lot of data needs to be processed very quickly.
2. They are used for character recognition, as in recognizing hand-written text, or damaged or mangled text. They are used in recognizing fingerprints. These are complicated patterns and are unique for each person. Layers of neurons can progressively clarify the pattern leading to are markably accurate result.
3. They are also used in traditional classification problems, like approving a financial loan application.

Design Principles of an Artificial Neural Network

1. A neuron is the basic processing unit of the network. The neuron (or processing element)receives inputs from its preceding neurons (or PEs),does some nonlinear weighted computation on the basis of those inputs, transforms the result into its output value, and then passes on the output to the next neuron in the network. X’s are the inputs, w’s are the weights for each input, and y is the output.

2. A Neural network is a multi-layered model. There is at least one input neuron, one output neuron, and at least one processing neuron. An ANN with just this basic structure would be a simple, single-stage computational unit. A simple task may be processed by just that one neuron and the result may be communicated soon. ANNs however, may have multiple layers of processing elements in sequence. There could

be many neurons involved in a sequence depending upon the complexity of the predictive action. The layers of PEs could work in sequence, or they could work in parallel.



3. The processing logic of each neuron may assign different weights to the various incoming input streams. The processing logic may also use nonlinear transformation, such as a sigmoid function, from the processed values to the output value. This processing logic and the intermediate weight and processing functions are just what works for the system as a whole, in its objective of solving a problem collectively. Thus, neural net works are considered to be an opaque and a black-box system.
4. The neural network can be trained by making similar decisions over and over again with many training cases. It will continue to learn by adjusting its internal computation and communication based on feedback about its previous decisions. Thus, the neural networks become better at making a decision as they handle more and more decisions. Depending upon the nature of the problem and the availability of good training data, at some point the neural network will learn enough and begin to match the predictive accuracy of a human expert. In many practical situations, the predictions of ANN, trained over a long period of time with a large number of training data, have begun to decisively become more accurate than human experts. At that point ANN can begin to be seriously considered for deployment in real situations in real time.

Representation of a Neural Network

A neural network is a series of neurons that receive inputs from other neurons. They do a weighted summation function of all the inputs, using different weights (or importance) for each input. The weighted sum is then transformed into an output value using a transfer function.
Learning in ANN occurs when the various processing elements in the neural network adjust the underlying relationship (weights, transfer function, etc)between input and outputs, in response to the feedback on their predictions. If the prediction made was correct, then the weights would remain the same, but if the prediction was incorrect, then the parameter values would change. The Transformation (Transfer) Function is any function suitable for the task at hand.

The transfer function for ANNs is usually a non-linear sigmoid function. Thus, if the normalized computed value is less than some value (say0.5) then the output value will be zero. If the computed value is at the cut-off threshold, then the output value will be a 1. It could be a nonlinear hyperbolic function in which the output is either a -1 or a 1. Many other functions could be designed for any or all of the processing elements.

Thus, in a neural network, every processing element can potentially have a different number of input values, a different set of weights for those inputs, and a different transformation function. Those values support and compensate for one another until the neural network as a whole learns to provide the correct output, as desired by the user.

4 **Create a decision tree for the following data set.** CO5 L3

Weekend (Example)	**Weather**	**Parents**	**Money**	**Decision (Category)**
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W5	Rainy	No	Rich	Stay in
W6	Rainy	Yes	Poor	Cinema
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W9	Windy	Yes	Rich	Cinema
W10	Sunny	No	Rich	Tennis

Solution:

The first thing we need to do is work out which attribute will be put into the node at the top of our tree: either weather, parents or money. To do this, we need to calculate:

Entropy(S) = $-p_{cinema} \log_2(p_{cinema}) - p_{tennis} \log_2(p_{tennis}) - p_{shopping} \log_2(p_{shopping}) - p_{stay\_in} \log_2(p_{stay\_in})$
$= -(6/10) * \log_2(6/10) - (2/10) * \log_2(2/10) - (1/10) * \log_2(1/10) - (1/10) * \log_2(1/10)$
$= -(6/10) * -0.737 - (2/10) * -2.322 - (1/10) * -3.322 - (1/10) * -3.322$
$= 0.4422 + 0.4644 + 0.3322 + 0.3322 = 1.571$

and we need to determine the best of:

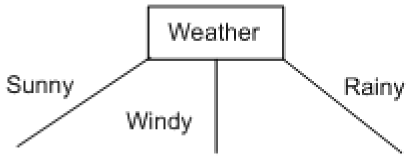
Gain(S, weather) = $1.571 - (|S_{sun}|/10)*\text{Entropy}(S_{sun}) - (|S_{wind}|/10)*\text{Entropy}(S_{wind}) - (|S_{rain}|/10)*\text{Entropy}(S_{rain})$
$= 1.571 - (0.3)*\text{Entropy}(S_{sun}) - (0.4)*\text{Entropy}(S_{wind}) - (0.3)*\text{Entropy}(S_{rain})$
$= 1.571 - (0.3)*(0.918) - (0.4)*(0.81125) - (0.3)*(0.918) = 0.70$

Gain(S, parents) = $1.571 - (|S_{yes}|/10)*\text{Entropy}(S_{yes}) - (|S_{no}|/10)*\text{Entropy}(S_{no})$
$= 1.571 - (0.5) * 0 - (0.5) * 1.922 = 1.571 - 0.961 = 0.61$

Gain(S, money) = $1.571 - (|S_{rich}|/10)*\text{Entropy}(S_{rich}) - (|S_{poor}|/10)*\text{Entropy}(S_{poor})$
$= 1.571 - (0.7) * (1.842) - (0.3) * 0 = 1.571 - 1.2894 = 0.2816$

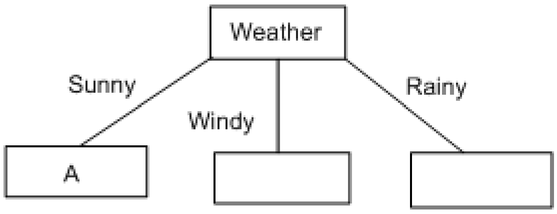
This means that the first node in the decision tree will be the weather attribute. As an exercise, convince yourself why this scored (slightly) higher than the parents attribute - remember what entropy means and look at the way information gain is calculated.

From the weather node, we draw a branch for the values that weather can take: sunny, windy and rainy:



Now we look at the first branch. S_{sunny} = {W1, W2, W10}. This is not empty, so we do not put a default categorization leaf node here. The categorizations of W1, W2 and W10 are Cinema, Tennis and Tennis respectively. As these are not all the same, we cannot put a categorization leaf node here. Hence we put an attribute node here, which we will leave blank for the time being.

Looking at the second branch, S_{windy} = {W3, W7, W8, W9}. Again, this is not empty, and they do not all belong to the same class, so we put an attribute node here, left blank for now. The same situation happens with the third branch, hence our amended tree looks like this:



Now we have to fill in the choice of attribute A, which we know cannot be weather, because we've already removed that from the list of attributes to use. So, we need to calculate the values for Gain(S_{sunny}, parents) and Gain(S_{sunny}, money). Firstly, Entropy(S_{sunny}) = 0.918. Next, we set S to be S_{sunny} = {W1,W2,W10} (and, for this part of the branch, we will ignore all the other examples). In effect, we are interested only in this part of the table:

Weekend (Example)	**Weather**	**Parents**	**Money**	**Decision (Category)**
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W10	Sunny	No	Rich	Tennis

Hence we can calculate:

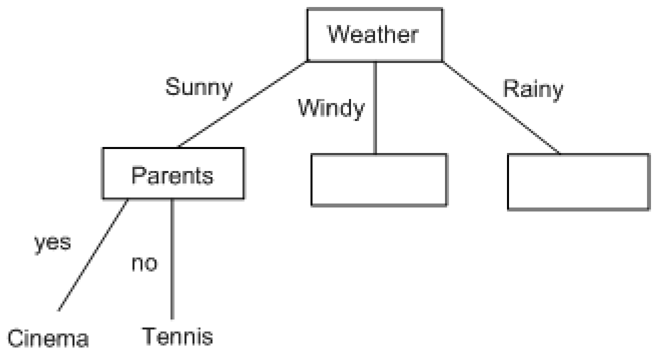
$$\text{Gain}(S_{sunny}, \text{parents}) = 0.918 - (|S_{yes}|/|S|)*\text{Entropy}(S_{yes}) - (|S_{no}|/|S|)*\text{Entropy}(S_{no})$$
$$= 0.918 - (1/3)*0 - (2/3)*0 = 0.918$$

$$\text{Gain}(S_{sunny}, \text{money}) = 0.918 - (|S_{rich}|/|S|)*\text{Entropy}(S_{rich}) - (|S_{poor}|/|S|)*\text{Entropy}(S_{poor})$$
$$= 0.918 - (3/3)*0.918 - (0/3)*0 = 0.918 - 0.918 = 0$$

Notice that Entropy(S_{yes}) and Entropy(S_{no}) were both zero, because S_{yes} contains examples which are all in the same category (cinema), and S_{no} similarly contains examples which are all in the same category (tennis). This should make it more obvious why we use information gain to choose attributes to put in nodes.

Given our calculations, attribute A should be taken as parents. The two values from parents are yes and no, and we will draw a branch from the node for each of these. Remembering that we replaced the set S by the set S_{Sunny}, looking at S_{yes}, we see that the only example of this is W1. Hence, the branch for yes stops at a categorisation leaf, with the category being

Cinema. Also, S_{no} contains W2 and W10, but these are in the same category (Tennis). Hence the branch for no ends here at a categorisation leaf. Hence our upgraded tree looks like this:



5	**Describe the Business Intelligence and Data Mining cycle. What are the different data mining techniques?**	[10]	CO3 CO4	L1

Business is the act of doing something productive to serve someone's needs, and thus earn a living and make the world a better place. business activities are recorded on paper or using electronic media, and then these records become data. There is more data from customers' responses and on the industry as a whole. All this data can be analysed and mined using special tools and techniques to generate patterns and intelligence, which reflect how the business is functioning. These ideas can then be fed back into the business so that it can evolve to become more effective and efficient in serving customer needs. And the cycle continues (Figure 1.1).

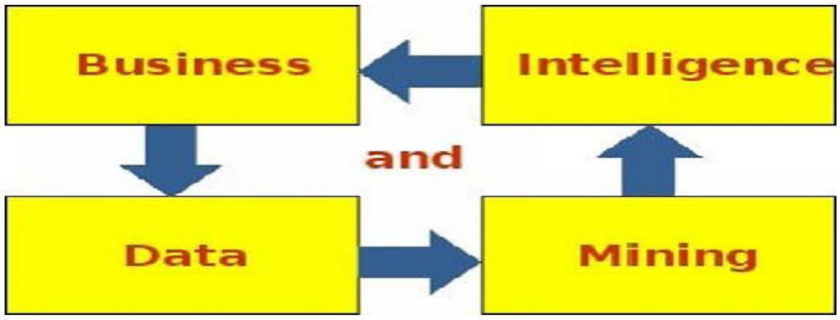


Figure 1.1: Business Intelligence and Data Mining Cycle

Business Intelligence

Any business organization needs to continually monitor its business environment and its own performance, and then rapidly adjust its future plans. This includes monitoring the industry, the competitors, the suppliers, and the customers. The organization needs to also develop a balanced scorecard to track its own health and vitality. Executives typically determine what they want to track based on their key performance Indexes (KPIs) or key result areas (KRAs). Customized reports need to be designed to deliver the require information to every executive. These reports can be converted into customized dashboards that deliver the information rapidly and in easy-to grasp formats.

Different data mining techniques

Here are brief descriptions of some of the most important data mining techniques used to generate insights from data.

<table>
<tr>
<td></td>
<td>1. Decision Trees: They help classify populations into classes. It is said that 70%of all data mining work is about classification solutions; and that 70% of all classification work uses decision trees. Thus, decision trees are the most popular and important data mining technique. There are many popular algorithms to make decision trees. They differ in terms of their mechanisms and each technique work well for different situations. It is possible to try multiple decision-tree algorithms on a data set and compare the predictive accuracy of each tree.

2. Regression: This is a well-understood technique from the field of statistics. The goal is to find a best fitting curve through the many data points. The best fitting curve is that which minimizes the (error) distance between the actual data points and the values predicted by the curve. Regression models can be projected into the future for prediction and forecasting purposes.

3. Artificial Neural Networks: Originating in the field of artificial intelligence and machine learning, ANNs are multi-layer non-linear information processing models that learn from past data and predict future values. These models predict well, leading to their popularity. The model's parameters may not be very intuitive. Thus, neural networks are opaque like a black-box. These systems also require a large amount of past data to adequate train the system.

4. Cluster analysis: This is an important data mining technique for dividing and conquering large data sets. The data set is divided into a certain number of clusters, by discerning similarities and dissimilarities within the data. There is no one right answer for the number of clusters in the data. The user needs to make a decision by looking at how well the number of clusters chosen fit the data. This is most commonly used for market segmentation. Unlike decision trees and regression, there is no one right answer for cluster analysis.

5. Association Rule Mining: Also called Market Basket Analysis when used in retail industry, these techniques look for associations between data values. An analysis of items frequently found together in a market basket can help cross-sell products, and also create product bundles.</td>
<td></td>
<td></td>
<td></td>
</tr>
</table>

6	**Consider the following dataset.**	[10]	CO5	L3

Student	Test_Marks	Grade
1	95	85
2	85	95
3	80	70
4	70	65
5	60	70

a) What linear regression equation best predicts statistics performance, based on math aptitude scores?
b) If a student made an 80 on the aptitude test, what grade would we expect her to make in statistics?
c) How well does the regression equation fit the data?

Solution:

In the table below, the x_i column shows scores on the aptitude test. Similarly, the y_i column shows statistics grades. The last two columns show deviations scores - the difference between the student's score and the average score on each test. The last two rows show sums and mean scores that we will use to conduct the regression analysis.

Student	x_i	y_i	(x_i-x)	(y_i-y)
1	95	85	17	8
2	85	95	7	18
3	80	70	2	-7
4	70	65	-8	-12
5	60	70	-18	-7
Sum	390	385		
Mean	78	77		

And for each student, we also need to compute the squares of the deviation scores (the last two columns in the table below).

Student	x_i	y_i	$(x_i-x)^2$	$(y_i-y)^2$
1	95	85	289	64
2	85	95	49	324
3	80	70	4	49
4	70	65	64	144
5	60	70	324	49
Sum	390	385	730	630

And finally, for each student, we need to compute the product of the deviation scores.

Student	x_i	y_i	$(x_i-x)(y_i-y)$
1	95	85	136
2	85	95	126
3	80	70	-14
4	70	65	96
5	60	70	126
Sum	390	385	470
Mean	78	77	

The regression equation is a linear equation of the form: $\hat{y} = b_0 + b_1x$. To conduct a regression analysis, we need to solve for b_0 and b_1. Computations are shown below. Notice that all of our inputs for the regression analysis come from the above three tables.

First, we solve for the regression coefficient (b_1):

$$b_1 = \Sigma [(x_i - x)(y_i - y)] / \Sigma [(x_i - x)^2]$$

$$b_1 = 470/730$$

$$b_1 = 0.644$$

Once we know the value of the regression coefficient (b_1), we can solve for the regression slope (b_0):

$$b_0 = y - b_1 * x$$

$$b_0 = 77 - (0.644)(78)$$

$$b_0 = 26.768$$

Therefore, the regression equation is: $\hat{y} = 26.768 + 0.644x$.

b) In our example, the independent variable is the student's score on the aptitude test. The dependent variable is the student's statistics grade. If a student made an 80 on the aptitude test, the estimated statistics grade (ŷ) would be:

$$\hat{y} = b_0 + b_1x$$

$$\hat{y} = 26.768 + 0.644x = 26.768 + 0.644 * 80$$

$$\hat{y} = 26.768 + 51.52 = 78.288$$

c) Whenever you use a regression equation, you should ask how well the equation fits the data. One way to assess fit is to check the coefficient of determination, which can be computed from the following formula.

$$R^2 = \{ (1 / N) * \Sigma [(x_i - x) * (y_i - y)] / (\sigma_x * \sigma_y) \}^2$$

where N is the number of observations used to fit the model, Σ is the summation symbol, x_i is the x value for observation i, x is the mean x value, y_i is the y value for observation i, y is the mean y value, σ_x is the standard deviation of x, and σ_y is the standard deviation of y.

Computations for the sample problem of this lesson are shown below. We begin by computing the standard deviation of x (σ_x):

$$\sigma_x = \text{sqrt} [\Sigma (x_i - x)^2 / N]$$

$$\sigma_x = \text{sqrt}(730/5) = \text{sqrt}(146) = 12.083$$

Next, we find the standard deviation of y, (σ_y):

$$\sigma_y = \text{sqrt} [\Sigma (y_i - y)^2 / N]$$

$$\sigma_y = \text{sqrt}(630/5) = \text{sqrt}(126) = 11.225$$

And finally, we compute the coefficient of determination (R^2):

$$R^2 = \{ (1 / N) * \Sigma [(x_i - x) * (y_i - y)] / (\sigma_x * \sigma_y) \}^2$$

$$R^2 = [(1/5) * 470 / (12.083 * 11.225)]^2$$

	$R^2 = (94 / 135.632)^2 = (0.693)^2 = 0.48$ A coefficient of determination equal to 0.48 indicates that about 48% of the variation in statistics grades (the dependent variable) can be explained by the relationship to math aptitude scores (the independent variable). This would be considered a good fit to the data, in the sense that it would substantially improve an educator's ability to predict student performance in statistics class.			
7	**List and explain different types of data. Compare database systems with data warehousing systems.** **Solution:** Data can be of different types. 1. Data could be an unordered collection of values. For example, a retailer sells shirts of red, blue, and green colour. There is no intrinsic ordering among these color values. One can hardly argue that any one colour is higher or lower than the other. This is called nominal (means names)data. 2. Data could be ordered values like small, medium and large. For example, the sizes of shirts could be extra-small, small, medium, and large. There is clarity that medium is bigger than small, and large is bigger than medium. But the differences may not be equal. This is called ordinal(ordered) data. 3. Another type of data has discrete numeric values defined in a certain range, with the assumption of equal distance between the values. Customer satisfaction score may be ranked on a 10-point scale with 1being lowest and 10 being highest. This requires the respondent to carefully calibrate the entire range as objectively as possible and place his own measurement in that scale. This is called interval (equal intervals) data. 4. The highest level of numeric data is ratio data which can take on any numeric value. The weights and heights of all employees would be exact numeric values. The price of a shirt will also take any numeric value. It is called ratio (any fraction) data. 5. There is another kind of data that does not lend itself to much mathematical analysis, at least not directly. Such data needs to be first structured and then analyzed. This includes data like audio, video, and graphs files, often called BLOBs (Binary Large Objects). These kinds of data lend themselves to different forms of analysis and mining. Songs can be described as happy or sad, fast-paced or slow, and so on. They may contain sentiment and intention, but these are not quantitatively precise.	[10]	CO3	L1

Compare database systems with data warehousing systems.

Function	Database	Data Warehouse
Purpose	Data stored in databases can be used for many purposes including day-to-day operations	Data stored in DW is cleansed data useful for reporting and analysis
Granularity	Highly granular data including all activity and transaction details	Lower granularity data; rolled up to certain key dimensions of interest
Complexity	Highly complex with dozens or hundreds of data files, linked through common data fields	Typically organized around a large fact tables, and many lookup tables
Size	Database grows with growing volumes of activity and transactions. Old completed transactions are deleted to reduce size.	Grows as data from operational databases i rolled-up and appende every day. Data is retai for long-term trend ana
Architectural choices	Relational, and object-oriented, databases	Star schema, or Snowf schema
Data Access mechanisms	Primarily through high level languages such as SQL. Traditional programming access DB through Open DataBase Connectivity (ODBC) interfaces	Accessed through SQ SQL output is forward reporting tools and dat visualization tools

DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING

Date:

CO's to PO's & PSO's mapping

Name of the course : Big Data Analytics
Name of the Faculty/s : Mrs. Vaishali M Deshmukh
Sub Code : 15CS82
Sem& Sec : 8th A,B

Course Outcomes		Modules covered	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3	PSO4
CO1	Master the concepts of HDFS and MapReduce framework	**1**	1	3	2	2	3	2	-	-	-	-	-	-	2	-	2	-
CO2	Investigate Hadoop related tools for Big Data Analytics and perform basic Hadoop Administration	**1,2**	1	3	2	2	3	2	-	-	-	-	-	-	2	-	2	-
CO3	Recognize the role of Business Intelligence, Data warehousing and Visualization in decision making	**3**	1	2	2	1	1	2	-	-	-	-	-	-	2	-	1	-
CO4	Infer the importance of core data mining techniques for data analytics	**3,4**	2	3	3	2	1	1	-	-	-	-	-	-	2	-	1	-
CO5	Analyze Data Mining Techniques	**3,5**	2	2	3	2	1	1	-	-	-	-	-	-	2	-	2	-

COGNITIVE LEVEL	REVISED BLOOMS TAXONOMY KEYWORDS
L1	List, define, tell, describe, identify, show, label, collect, examine, tabulate, quote, name, who, when, where, etc.
L2	summarize, describe, interpret, contrast, predict, associate, distinguish, estimate, differentiate, discuss, extend
L3	Apply, demonstrate, calculate, complete, illustrate, show, solve, examine, modify, relate, change, classify, experiment, discover.
L4	Analyze, separate, order, explain, connect, classify, arrange, divide, compare, select, explain, infer.
L5	Assess, decide, rank, grade, test, measure, recommend, convince, select, judge, explain, discriminate, support, conclude, compare, summarize.

PROGRAM OUTCOMES(PO), PROGRAM SPECIFIC OUTCOMES(PSO)				CORRELATION LEVELS	
PO1	Engineering knowledge	PO7	Environment and sustainability	0	No Correlation
PO2	Problem analysis	PO8	Ethics	1	Slight/Low
PO3	Design/development of solutions	PO9	Individual and team work	2	Moderate/ Medium
PO4	Conduct investigations of complex problems	PO10	Communication	3	Substantial/ High
PO5	Modern tool usage	PO11	Project management and finance		
PO6	The Engineer and society	PO12	Life-long learning		
PSO1	Implement and maintain enterprise solutions using latest technologies.				
PSO2	Develop and simulate wired and wireless NW protocols for various network applications using modern tools.				
PSO3	Apply the knowledge of Information technology and software testing to maintain legacy systems.				
PSO4	Apply knowledge of web programming and design to develop web based applications using database and other technologies				