CMR INSTIT UTE OF TECHNLOGY	USN										

Internal Assessment Test - 2

Sub:	Big Data Analytics					Sub Code:	17CS82	Branch:	ISE	
Date:	16/04/2019	Duration:	90 min's	Max Marks:	50	Sem / Sec:	VIII / A ,B			OBE

Solution for BDA

		MARKS	CO	RBT
1 (a)	**What is Web Mining? Explain its characteristics and three types of web mining.** **Solution:** Web mining is the art and science of discovering patterns and insights from the World-wide web so as to improve it. The world-wide web is at the heart of the digital revolution. More data is posted on the web every day than was there on the whole web just 20 years ago. Billions of users are using it every day for a variety of purposes. The web is used for electronic commerce, business communication, and many other applications. Web mining analyzes data from the web and helps find insights that could optimize the web content and improve the user experience. Data for web mining is collected via Web crawlers, web logs, and other means. Here are some characteristics of optimized websites: **1. Appearance:** Aesthetic design. Well-formatted content, easy to scan and navigate. Good color contrasts. **2. Content:** Well-planned information architecture with useful content. Fresh content. Search engine optimized. Links to other good sites. **3. Functionality:** Accessible to all authorized users. Fast loading times. Usable forms. Mobile enabled. This type of content and its structure is of interest to ensure the web is easy to use. The analysis of web usage provides feedback on the web content, and also the consumer's browsing habits. This data can be of immense use for commercial advertising, and even for social engineering. The web could be analyzed for its structure as well as content. The usage pattern of web pages could also be analyzed. Depending upon objectives, web mining can be divided into three different types: **1. Web usage mining** As a user clicks anywhere on a webpage or application, the action is recorded by many entities in many locations. The browser at the client machine will record the click, and the web server providing the content would also make a record of the pages served and the user activity on those pages. The entities between the client and the server, such as the router, proxy server, or ad server, too would record that click **2. Web content mining** A website is designed in the form of pages with a distinct URL (universal resource locator). A large website may contain thousands of pages. These pages and their content is managed using specialized software systems called Content Management Systems. Every page can have text, graphics, audio, video, forms, applications, and more kinds of content including user generated content.	[06]	CO5	L2

3. Web structure mining

The Web works through a system of hyperlinks using the hypertext protocol (http). Any page can create a hyperlink to any other page, it can be linked to by another page. The intertwined or self-referral nature of web lends itself to some unique network analytical algorithms. The structure of Web pages could also be analyzed to examine the pattern of hyperlinks among pages. There are two basic strategic models for successful websites: Hubs and Authorities.

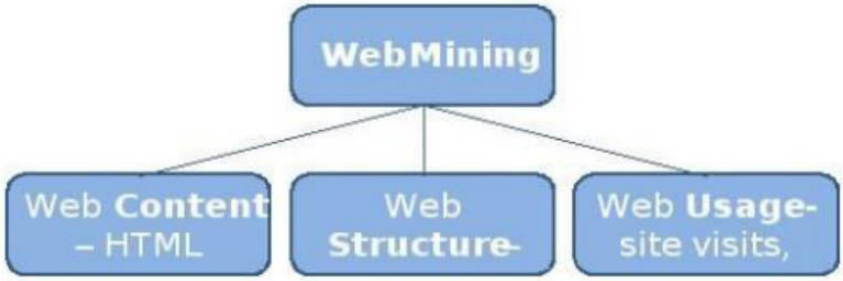


Figure: 1 Web Mining structure

1 (b) **Explain Text Mining Process.** [04] CO5 L2

Solution:

Text Mining is a rapidly evolving area of research. As the amount of social media and other text data grows, there is need for efficient abstraction and categorization of meaningful information from the text.

The first level of analysis is identifying frequent words. This creates a bag of important words. Texts – documents or smaller messages – can then be ranked on how they match to a particular bag-of-words. However, there are challenges with this approach. For example, the words may be spelled a little differently. Or there may be different words with similar meanings.

The next level is at the level of identifying meaningful phrases from words. Thus 'ice' and 'cream' will be two different key words that often come together. However, there is a more meaningful phrase by combining the two words into 'ice cream'. There might be similarly meaningful phrases like 'Apple Pie'.

The next higher level is that of Topics. Multiple phrases could be combined into Topic area.

Thus the two phrases above could be put into a common basket, and this bucket could be called 'Desserts'. Text mining is a semi-automated process. Text data needs to be gathered, structured, and then mined, in a 3-step process (Figure 1)

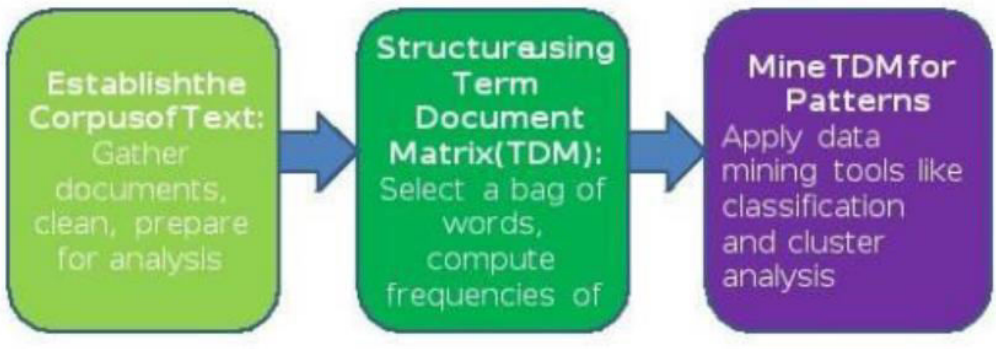
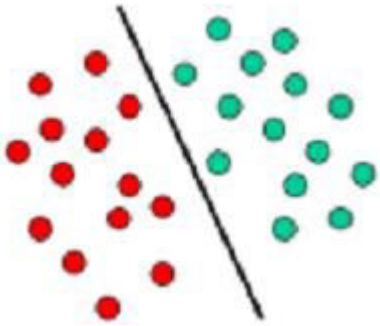
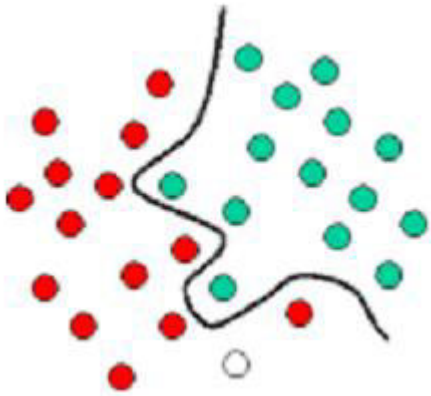
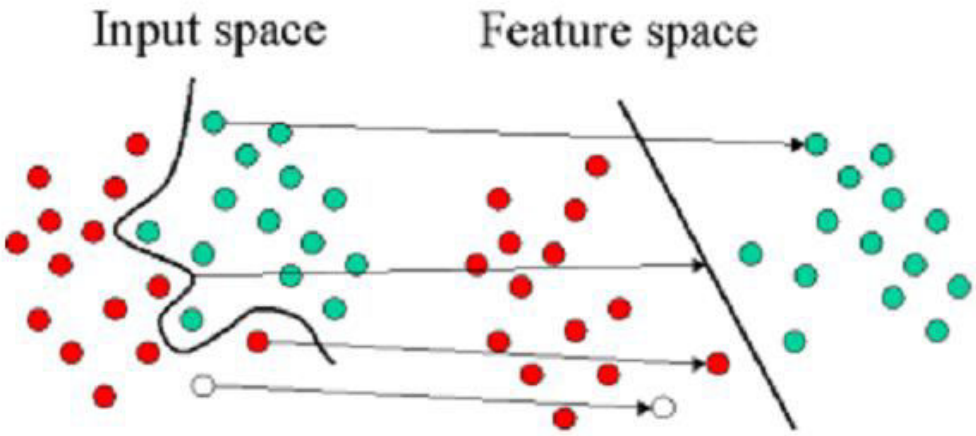


Figure 1 Text Mining Architecture

1. The text and documents are first gathered into a corpus, and organized.

	2. The corpus is then analysed for structure. The result is a matrix mapping important terms to source documents. 3. The structured data is then analysed for word structures, sequences, and frequency.			
2	**Explain the Support Vector Machine with its kernel types.** **Solution:** Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. A schematic example is shown in the illustration below. In this example, the objects belong either to class GREEN or RED. The separating line defines a boundary on the right side of which all objects are GREEN and to the left of which all objects are RED. Any new object (white circle) falling to the right is labeled, i.e., classified, as GREEN (or classified as RED should it fall to the left of the separating line).  The above is a classic example of a linear classifier, i.e., a classifier that separates a set of objects into their respective groups (GREEN and RED in this case) with a line. Most classification tasks, however, are not that simple, and often more complex structures are needed in order to make an optimal separation, i.e., correctly classify new objects (test cases) on the basis of the examples that are available (train cases). This situation is depicted in the illustration below. Compared to the previous schematic, it is clear that a full separation of the GREEN and RED objects would require a curve (which is more complex than a line). Classification tasks based on drawing separating lines to distinguish between objects of different class memberships are known as hyperplane classifiers. Support Vector Machines are particularly suited to handle such tasks.  The illustration below shows the basic idea behind Support Vector Machines. Here we see the original objects (left side of the schematic) mapped, i.e., rearranged, using a set of mathematical functions, known as kernels. The process of rearranging the objects is known as mapping (transformation). Note that in this new setting, the mapped objects (right side of the schematic) is linearly separable and, thus, instead	[10]	CO4	L2

of constructing the complex curve (left schematic), all we have to do is to find an optimal line that can separate the GREEN and the RED objects.



Support Vector Machine (SVM) is primarily a classier method that performs classification tasks by constructing hyperplanes in a multidimensional space that separates cases of different class labels. SVM supports both regression and classification tasks and can handle multiple continuous and categorical variables. For categorical variables a dummy variable is created with case values as either 0 or 1.

Thus, a categorical dependent variable consisting of three levels, say (A, B, C), is represented by a set of three dummy variables: A: {1 0 0}, B: {0 1 0}, C: {0 0 1} To construct an optimal hyperplane, SVM employs an iterative training algorithm, which is used to minimize an error function. According to the form of the error function, SVM models can be classified into four distinct groups:

- Classification SVM Type 1 (also known as C-SVM classification)
- Classification SVM Type 2 (also known as nu-SVM classification)
- Regression SVM Type 1 (also known as epsilon-SVM regression)
- Regression SVM Type 2 (also known as nu-SVM regression)

Following is a brief summary of each model.
Classification SVM

CLASSIFICATION SVM TYPE 1

For this type of SVM, training involves the minimization of the error function: subject to the constraints: where C is the capacity constant, w is the vector of coefficients, b is a constant, and represents parameters for handling non separable data (inputs). The index i label the N training cases. Note that represents the class labels and xi represents the independent variables. The kernel is used to transform data from the input (independent) to the feature space. It should be noted that the larger the C, the more the error is penalized. Thus, C should be chosen with care to avoid over fitting.

CLASSIFICATION SVM TYPE 2

In contrast to Classification SVM Type 1, the Classification SVM Type 2 model minimizes the error function: subject to the constraints: In a regression SVM, you have to estimate the functional dependence of the dependent variable y on a set of independent variables x. It assumes, like other regression problems, that the relationship between the independent and dependent variables is given by a deterministic function f plus the addition of some additive noise:

Regression SVM

y = f(x) + noise

The task is then to find a functional form for f that can correctly predict new cases that the SVM has not been presented with before. This can be achieved by training the SVM model on a sample set, i.e., training set, a process that involves, like classification (see above), the sequential optimization of an error function. Depending on the definition of this error function, two types of SVM models can be recognized:

REGRESSION SVM TYPE 1

For this type of SVM the error function is:

$$\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i + C\sum_{i=1}^{N}\xi^*_i$$

which we minimize subject to:

$$w^T\phi(x_i) + b - y_i \le \varepsilon + \xi_i^*$$
$$y_i - w^T\phi(x_i) - b_i \le \varepsilon + \xi_i$$
$$\xi_i, \xi_i^* \ge 0, i = 1,\ldots,N$$

REGRESSION SVM TYPE 2

For this SVM model, the error function is given by:

$$\frac{1}{2}w^T w - C\left(v\varepsilon + \frac{1}{N}\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right)\right)$$

which we minimize subject to:

$$\left(w^T\phi(x_i) + b\right) - y_i \le \varepsilon + \xi_i$$
$$y_i - \left(w^T\phi(x_i) + b_i\right) \le \varepsilon + \xi^*_i$$
$$\xi_i, \xi^*_i \ge 0, i = 1,\ldots,N, \varepsilon \ge 0$$

There are number of kernels that can be used in Support Vector Machines models. These include linear, polynomial, radial basis function (RBF) and sigmoid:

Kernel Functions

$$K(\mathbf{X_i}, \mathbf{X_j}) = \begin{cases} \mathbf{X_i} \cdot \mathbf{X_j} & \text{Linear} \\ (\gamma \mathbf{X_i} \cdot \mathbf{X_j} + C)^d & \text{Polynomial} \\ \exp(-\gamma |\mathbf{X_i} - \mathbf{X_j}|^2) & \text{RBF} \\ \tanh(\gamma \mathbf{X_i} \cdot \mathbf{X_j} + C) & \text{Sigmoid} \end{cases}$$

where $K(\mathbf{X_i}, \mathbf{X_j}) = \phi(\mathbf{X_i}) \cdot \phi(\mathbf{X_j})$

that is, the kernel function, represents a dot product of input data points mapped into the higher
dimensional feature space by transformation
***Gamma* is an adjustable parameter of certain kernel functions.**
The RBF is by far the most popular choice of kernel types used in Support Vector Machines. This is mainly because of their localized and finite responses across the entire range of the real x-axis.

3 (a)	**What are the 3 Vs of Big Data and its sources?** **Solution:** **1. Variety:** There are many types of data, including structured and unstructured data. Structured data consists of numeric and text fields. Unstructured data includes images, video, audio, and many other types. There are also many sources of data. The traditional sources of structured data include data from ERPs systems and other operational systems. Sources for unstructured data include social media, Web, RFID, machine data, and others. Unstructured data comes in a variety of sizes, resolutions, and are subject to different kinds of analysis. For example, video files can be tagged with labels, and they can be played, but video data is typically not computed, which is the same with audio data. Graphic data can be analyzed for network distances. Facebook texts and tweets can be analyzed for sentiments but cannot be directly compared. ***2. Velocity:*** The Internet greatly increases the speed of movement of data, from e-mails to social media to video files, data can move quickly. Cloud-based storage makes sharing instantaneous, and easily accessible from anywhere. Social media applications enable people to share their data with each other instantly. Mobile access to these applications also speeds up the generation and access to data. ***3. Volume:*** Websites have become great sourced and repositories for many kinds of data. User click streams are recorded and stored for future use. Social media applications such as Facebook, Twitter, Pinterest, and other applications have enabled users to become consumers of data (producers and consumers). There is an increase in the number of data shares, and also the size of each data element. High-definition videos can increase the total shared data. There are autonomous data streams of video, audio, text, data, and so on coming from social media sites, websites, RFID applications, and so on.	[06]	CO1	L1

	***Sources of Data*:** There are several sources of data, including some new ones. Data from outside the organization may be incomplete, and of a different quality and accuracy. **1. *Social Media*:** All activities on the web and social media are considered stores and are accessible. Email was the first major source of new data. Google searches, Facebook posts, Tweets, Youtube videos, and blogs enable people to generate data for one another. **2. *Organizations*:** Business organizations and government are a major source of data.ERP systems, e-Commerce systems, user-generated content, web-access logs, and many other sources of data generate valuable data for organizations. **3. *Machines*:** The Internet of things is evolving. Many machines are connected to the web and autonomously generate data that is untouched by humans. RFID tags and telematics are two major applications that generate enormous amounts of data. Connected devices such as phones and refrigerators generate data about their location and status. **4. *Metadata*:** There is enormous data about data itself. Web crawlers and web-bots scan the web to capture new web pages, their html structure, and their metadata. This data is used by many applications, including web search engines. The data also includes varied quality of data. It depends upon the purpose of collecting the data, and how carefully it has been collected and curated. Data from within the organization is likely to be of a higher quality. Publicly available data would include some trustworthy data such as from the government.			
3(b)	**Explain Management of big data in details.** **Solution:** Many organizations have started initiatives around the use of Big Data. However, most organizations do not necessarily have a grip on it. Here are some emerging insights into making better use of big data. 1. Across all industries, the business case for big data is strongly focused on addressing *customer-centric objectives*. The first focus on deploying big data initiatives is to protect and enhance customer relationships and customer experience. 2. *Solve a real pain-point*. Big data should be deployed for specific business objectives in order to avoid being overwhelmed by the sheer size of it all. 3. Organizations are beginning their *pilot* implementations by using existing and newly accessible internal sources of data. It is better to begin with data under one's control and where one has a superior understanding of the data. 4. Put *humans and data together* to get the most insight. Combining database analysis with human intuition and perspectives is better than going just one way. 5. Advanced *analytical capabilities* are required, yet lacking, for organizations to get the most value from big data. There is a growing awareness of building or hiring those skills and capabilities. 6. Use more *diverse data*, not just more data. This would provide a broader perspective into reality and better-quality insights.	[04]	CO1	L1

	7. The *faster* you analyze the data, the more its predictive value. The value of data depreciates with time. If the data is not processed in five minutes, then the immediate advantage is lost. 8. *Don't throw away data* if no immediate use can be seen for it. Data has value beyond what you initially anticipate. Data can add perspective to other data later in a multiplicative manner. 9. *Maintain one copy* of your data, not multiple. This would help avoid confusion and increase efficiency. 10. Plan for *exponential growth*. Data is expected to continue to grow at exponential rates. Storage costs continue to fall, data generation continues to grow, data-based applications continue to grow in capability and functionality. 11. A *scalable and extensible* information management foundation is a prerequisite for big data advancement. Big data builds upon resilient, secure, efficient, flexible, and real-time information processing environment. 12. Big data is transforming business, just like IT did. Big data is a new phase representing a *digital world*. Business and society are not immune to its strong impacts.			
4	**Differentiate between** **a) Text Mining and Data Mining**	[10]	CO5	L2

Dimension	Text Mining	Data Mining
Nature of data	Unstructured data: Words, phrases, sentences	Numbers; alphabetical and logical values
Language used	Many languages and dialects used in the world; many languages are extinct, new documents are discovered	Similar numerical systems across the world
Clarity and precision	Sentences can be ambiguous; sentiment may contradict the words	Numbers are precise.
Consistency	Different parts of the text can contradict each other	Different parts of data can be inconsistent, thus, requiring statistical significance analysis
Sentiment	Text may present a clear and consistent or mixed sentiment, across a continuum. Spoken words adds further sentiment	Not applicable
Quality	Spelling errors. Differing values of proper nouns such as names. Varying quality of language translation	Issues with missing values, outliers, etc
Nature of analysis	Keyword based search; co-existence of themes; Sentiment mining	A full wide range of statistical and machine learning analysis for Relationships and differences

b) Social Network Analysis and Traditional Data Mining

Dimensions	Social Network Analysis	Traditional Data Mining
Nature of learning	Unsupervised Learning	Supervised & Unsupervised Learning
Analysis of goals	Hub nodes, important nodes, sub networks	Key decision rules, cluster centroids
Dataset structures	A graph of nodes and directed links	Dataset with columns instances
Analysis Techniques	Visualization with statistics, iterative graphical computation	Machine learning Statistics
Quality measurements	Usefulness is key criteria	Predictive accuracy for classification techniques

5 (a)	**What are pre-processing steps involved in Test Data Mining (TDM)?** **Solution:** Here are some considerations in creating a TDM. 1. A large collection of documents mapped to a large bag of words will likely lead to a very sparse matrix if they have few common words. Reducing dimensionality of data will help	[05]	CO4	L2

	improve the speed of analysis and meaningfulness of the results. Synonyms, or terms will similar meaning, should be combined and should be counted together, as a common term. This would help reduce the number of distinct terms of words or 'tokens'. 2. Data should be cleaned for spelling errors. Common spelling errors should be ignored, and the terms should be combined. Uppercase lowercase terms should also be combined. 3. When many variants of the same term are used, just the stem of the word would be used to reduce the number of terms. For instance, terms like customer order, ordering, order data, should be combined into a single token word, called 'Order'. 4. On the other side, homonyms (terms with the same spelling but different meanings) should be counted separately. This would enhance the quality of analysis. For example, the term order can mean a customer order, or the ranking of certain choices. These two should be treated separately. "The boss ordered that the customer orders data analysis be presented in chronological order'. This statement shows three different meanings for the word 'order'. Thus, there will be a need for a manual review of the TD matrix. 5. Terms with very few occurrences in very few documents should be eliminated from the matrix. This would help increase the density of the matrix and the quality of analysis. 6. The measures in each cell of the matrix could be one of several possibilities. It could be a simple count of the number of occurrences of each term in a document. It could also be the log of that number. It could be the fraction number computed by dividing the frequency count by the total number of words in the document. Or there may be binary values in the matrix to represent whether a term is mentioned or not. The choice of value in the cells will depend upon the purpose of the text analysis. At the end of this analysis and cleansing, a well-formed, densely populated, rectangular, TDM will be ready for analysis. The TDM could be mined using all the available data mining techniques.			
5 (b)	**List the key requirements for good data visualization?** **Solution:** To help the client in understanding the situation, the following considerations are important: **1. Fetch appropriate and correct data for analysis.** This requires some understanding of the domain of the client and what is important for the client. E.g. in a business setting, one may need to understand the many measure of profitability and productivity. **2. Sort the data in the most appropriate manner.** It could be sorted by numerical variables, or alphabetically by name. **3. Choose appropriate method to present the data.** The data could be presented as a table, or it could be presented as any of the graph types. **4. The data set could be pruned to include only the more significant elements**. More data is not necessarily better, unless it makes the most significant impact on the situation. 5. The visualization could show additional dimension for reference such as the expectations or targets with which to compare the results. 6. The numerical data may need to be binned into a few categories. E.g. the orders per person were plotted as actual values, while the order sizes were binned into 4 categorical choices. 7. High-level visualization could be backed by more detailed analysis. For the most significant results, a drill-down may be required. 8. There may be needed to present additional textual information to tell the whole story.	[05]	CO3	L1

	For example, one may require notes to explain some extraordinary results			
6	**Write a Naïve Bayes algorithm for leaning and classifying text and apply it *on following* data to predict the category of the given text: *"A very close game".***	[10]	CO5	L3

Text	**Category**
"A great game"	Sports
"The election was over"	Not sports
"Very clean match"	Sports
"A clean but forgettable game"	Sports
"It was a close election"	Not sports

Solution:

LEARN_NAIVE_BAYES_TEXT($Examples$, V)

Examples is a set of text documents along with their target values. V is the set of all possible target values. This function learns the probability terms $P(w_k|v_j)$, describing the probability that a randomly drawn word from a document in class v_j will be the English word w_k. It also learns the class prior probabilities $P(v_j)$.

1. collect all words, punctuation, and other tokens that occur in Examples

- *Vocabulary* ← the set of all distinct words and other tokens occurring in any text document from *Examples*

2. calculate the required $P(v_j)$ and $P(w_k|v_j)$ probability terms

- For each target value v_j in V do
  - $docs_j$ ← the subset of documents from *Examples* for which the target value is v_j
  - $P(v_j) \leftarrow \frac{|docs_j|}{|Examples|}$
  - $Text_j$ ← a single document created by concatenating all members of $docs_j$
  - n ← total number of distinct word positions in $Text_j$
  - for each word w_k in *Vocabulary*
    - n_k ← number of times word w_k occurs in $Text_j$
    - $P(w_k|v_j) \leftarrow \frac{n_k+1}{n+|Vocabulary|}$

CLASSIFY_NAIVE_BAYES_TEXT(Doc)

Return the estimated target value for the document Doc. a_i denotes the word found in the ith position within Doc.

- *positions* ← all word positions in *Doc* that contain tokens found in *Vocabulary*
- Return v_{NB}, where

$$v_{NB} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j) \prod_{i \in positions} P(a_i|v_j)$$

Word	P(word | Sports)	P(word | Not Sports)
a	(2+1) / (11+14)	(1+1) / (9+14)
very	(1+1) / (11+14)	(0+1) / (9+14)
close	(0+1) / (11+14)	(1+1) / (9+14)
game	(2+1) / (11+14)	(0+1) / (9+14)

Now just multiply all the probabilities, and check the probabilities

$$P(a|Sports) \times P(very|Sports) \times P(close|Sports) \times P(game|Sports) \times P(Sports)$$
$$= 2.76 \times 10^{-5}$$
$$= 0.0000276$$

$$P(a|Not\ Sports) \times P(very|Not\ Sports) \times P(close|Not\ Sports) \times P(game|Not\ Sports) \times P(Not\ Sports)$$
$$= 0.572 \times 10^{-5}$$
$$= 0.00000572$$

The classifier gives "A very close game" the **Sports** tag as it is higher than other.

7 **Discuss the different types of data visualization techniques.** [10] CO3 L2

Solution:

1. Line graph. This is a basic and most popular type of displaying information. It shows data as a series of points connected by straight line segments. If mining with time-series data, time is usually shown on the x-axis. Multiple variables can be represented on the same scale on y-axis to compare of the line graphs of all the variables.

2. Scatter plot: This is another very basic and useful graphic form. It helps several the relationship between two variables. In the above case let, it shows two dimensions: Life Expectancy and Fertility Rate. Unlike in a line graph, there are no line segments connecting the points.

3. Bar graph: A bar graph shows thin colorful rectangular bars with their lengths being proportional to the values represented. The bars can be plotted vertically or horizontally. The bar graphs use a lot of more ink than the line graph and should be used when line graphs are inadequate.

4. Stacked Bar graphs: These are a particular method of doing bar graphs. Values of multiple variables are stacked one on top of the other to tell an interesting story. Bars can also be normalized such as the total height of every bar is equal, so it can show the relative composition of each bar.

5. Histograms: These are like bar graphs, except that they are useful in showing data frequencies or data values on classes (or ranges) of a numerical variable.

6. Pie charts: These are very popular to show the distribution of a variable, such as sales by
region. The size of a slice is representative of the relative strengths of each value.

7. Box charts: These are special form of charts to show the distribution of variables. The box shows the middle half of the values, while whiskers on both sides extend to the extreme values in either direction.

8. Bubble Graph: This is an interesting way of displaying multiple dimensions in one chart.
It is a variant of a scatter plot with many data points marked on two dimensions. Now imagine that each data point on the graph is a bubble (or a circle) ... the size of the circle and the color fill in the circle could represent two additional dimensions.

9. Dials: These are charts like the speed dial in the car, that shows whether the variable value (such as sales number) is in the low range, medium range, or high range. These ranges could be colored red, yellow and green to give an instant view of the data.

10. Geographical Data maps are particularly useful maps to denote statistic

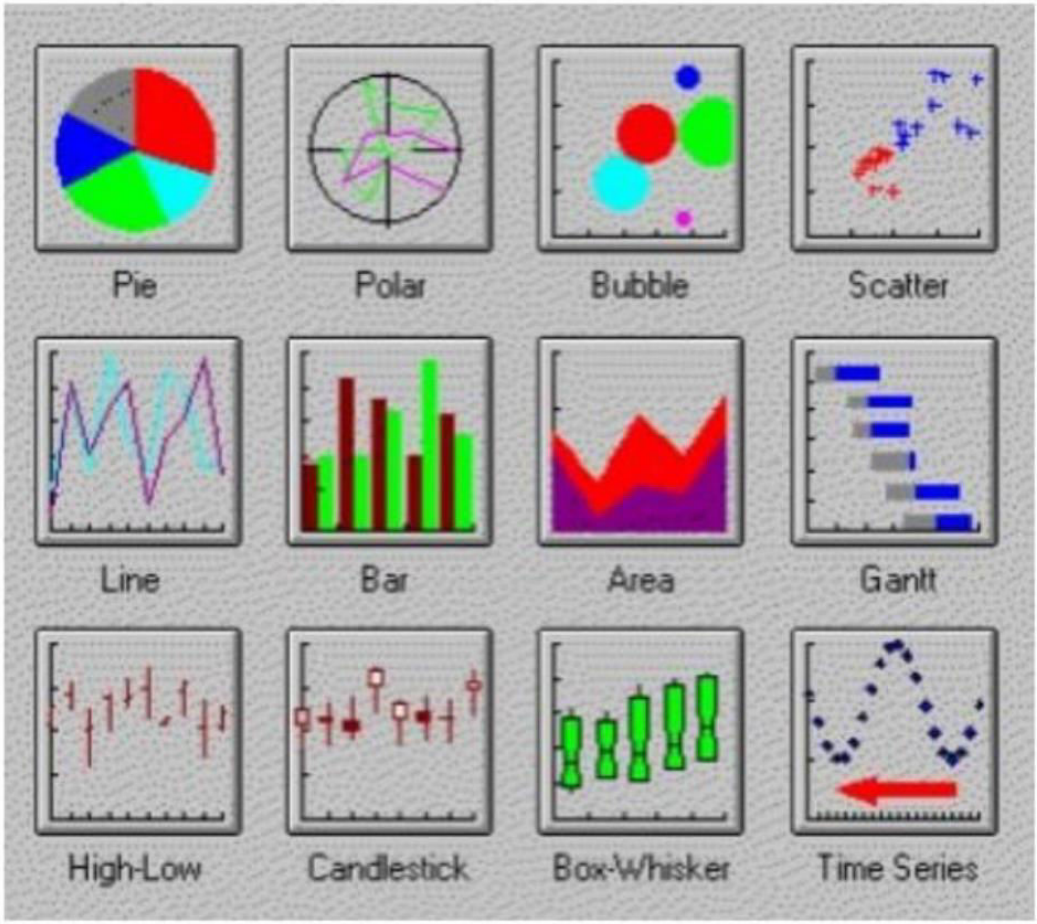


Fig.1 Different types of graphs

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				