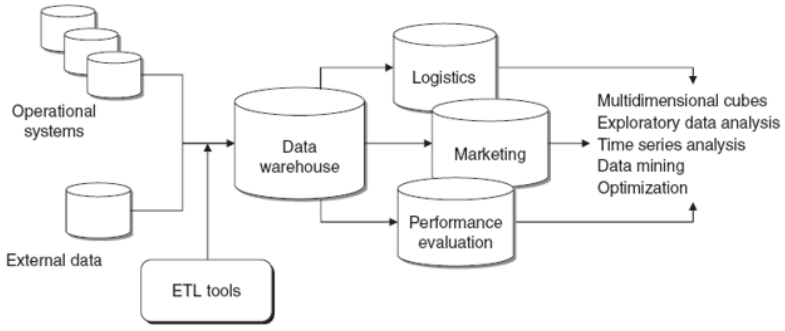
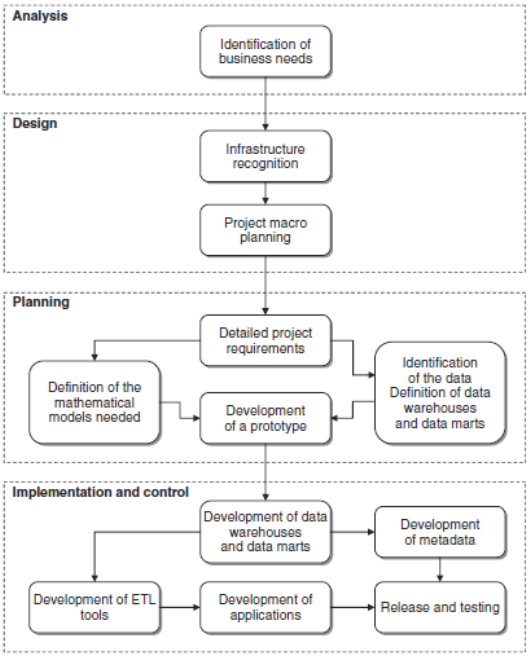
Q. P. Code:

(Time: 2½ hours)

Total Marks: 75

N. B.: (1) **All** questions are **compulsory**.
(2) Make **suitable assumptions** wherever necessary and **state the assumptions** made.
(3) Answers to the **same question** must be **written together**.
(4) Numbers to the **right** indicate **marks**.
(5) Draw **neat labeled diagrams** wherever **necessary**.
(6) Use of **Non-programmable** calculators is **allowed**.

1.	**Attempt *any three* of the following:**	**15**
a.	What is business intelligence? Explain architecture of the business intelligence. **Answer:** Business intelligence: The technology, methods, and applications used for the integration and presentation of business information, is known as Business Intelligence System. The business decisions making made by an enterprise is supported by the Business Intelligence System. Architecture:  Business Intelligence Architectures	
b.	Explain different phases in development of business intelligence system. **Answer:** **Phases in the development of Business Intelligence System:**	



Phases in the development of a business intelligence system

Analysis:

During the first phase, the needs of the organization relative to the development of a business intelligence system should be carefully identified.
This preliminary phase is generally conducted through a series of interviews of knowledge workers performing different roles and activities within the organization. It is necessary to clearly describe the general objectives and priorities of the project, as well as to set out the costs and benefits deriving from the development of the business intelligence system.

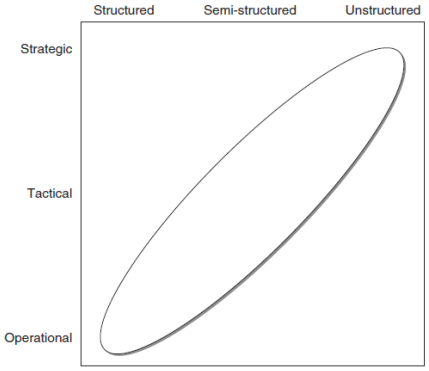
Design:

The second phase includes two sub-phases and is aimed at deriving a provisional plan of the overall architecture, taking into account any development in the near future and the evolution of the system in the mid-term. First, it is necessary to make an assessment of the existing information infrastructures. Moreover, the main decision-making processes that are to be supported by the business intelligence system should be examined, in order to adequately determine the information requirements. Later on, using classical project management methodologies, the project plan will be laid down, identifying development phases, priorities, expected execution times and costs, together with the required roles and resources.

Planning:

The planning stage includes a sub-phase where the functions of the business intelligence system are defined and described in greater detail. Subsequently, existing data as well as other data that might be retrieved externally are assessed. This allows the information structures of the business intelligence architecture, which consist of a central data warehouse and possibly some satellite data marts, to be designed. Simultaneously with the recognition of the available data, the mathematical models to be adopted should be defined, ensuring the availability of the data required to feed each model and verifying that the efficiency of the

	algorithms to be utilized will be adequate for the magnitude of the resulting problems. Finally, it is appropriate to create a system prototype, at low cost and with limited capabilities, in order to uncover beforehand any discrepancy between actual needs and project specifications. **Implementation and control:** The last phase consists of five main sub-phases. First, the data warehouse and each specific data mart are developed. These represent the information infrastructures that will feed the business intelligence system. In order to explain the meaning of the data contained in the data warehouse and the transformations applied in advance to the primary data, a *metadata* archive should be created, as described in Chapter 3. Moreover, ETL procedures are set `out to extract and transform the data existing in the primary sources, loading them into the data warehouse and the data marts. The next step is aimed at developing the core business intelligence applications that allow the planned analyses to be carried out. Finally, the system is released for test and usage.	
c.	What is decision support system(DSS)? What are the factors that affect the degree of success of a DSS? **Answer:** **Decision support system:** **A *decision support system* (DSS) is an interactive computer-based application that** combines **data and mathematical models to help decision makers solve complex problems faced in managing the public and private enterprises and organizations.** **<u>Factors that may affect the degree of success of DSS:</u>** • **Integration:** The design and development of a DSS require a significant number of methodologies, tools, models, individuals and organizational processes to work in harmony. • **Involvement:** The exclusion or feeling of isolation from the project team of knowledge workers who will actually use the system once it is implemented is a mistake that is sometimes made during the design and development of DSS. • **Uncertainty:** Implementation cost is less but driving more effective decisions may cost more.	
d.	Explain classification of decisions according to their nature and scope. **Answer:** **According to their nature, decisions can be classified as** • **Structured** • **Unstructured and** • **Semi-structured** **According to their scope, decisions can be classified as** • **Operational** • **Tactical and** • **Strategic**	



A taxonomy of decisions

e. Define system. Explain closed cycle and open cycle system with suitable example.Answer:

System definition:

The term *system* is often used in everyday language: for instance, we refer to the solar system, the nervous system or the justice system. The entities that we intuitively denominate *systems* share a common characteristic, which we will adopt as an abstract definition of the notion of system: each of them is made up of a set of components that are in some way connected to each other so as to provide a single collective result and a common purpose.

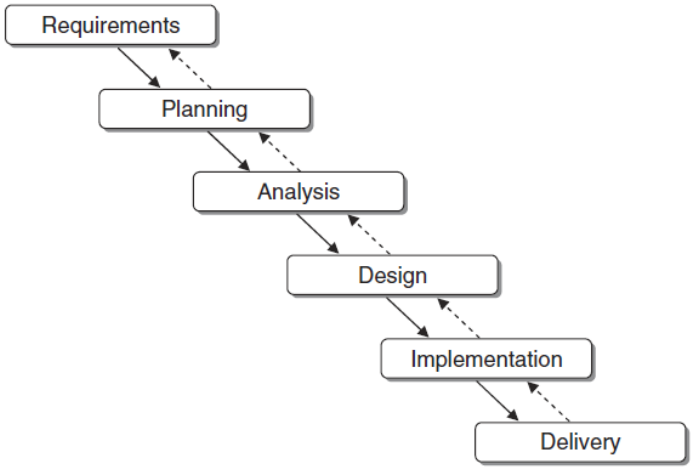
Closed cycle system:

Open cycle system:

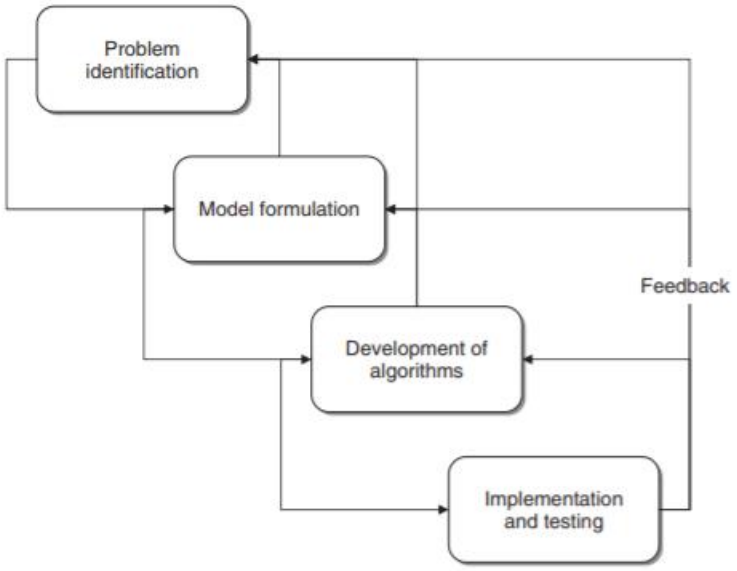
f. Describe the phases in the development of a decision support systems(DSS).

Answer:

- **Planning**: The main purpose of the planning phase is to understand the needs and opportunities, translate them into project & later into DSS.
- **Analysis**: Define detailed functions of DSS to be developed responses to the questions like *What should the DSS accomplish, and who will use it, when and how?*
- **Design**: How will the DSS work. Hardware + network + Software tools
- **Implementation:** Implementation + installation + testing



Phases in the developement of a decision support system

2.	**Attempt *any three* of the following:**	**15**
a.	What are the phases in the development of mathematical models for decision making? **Answer:**  *Phases in the development of mathematical models for decision making*	
b.	Explain the divisions of mathematical models according to their characteristics, probabilistic nature, temporal dimension. **Answer:** **According to their characteristics, models can be divided into** Iconic, analogical and symbolic **According to their probabilistic nature of models, which can be either** Stochastic or deterministic. **According to their temporal dimension in a model can be either** static or dynamic.	
c.	What is data mining? List the real life applications of data mining. **Answer:** **Data mining:** The term data mining refers therefore to the overall process consisting of data gathering and analysis, development of inductive learning models and adoption of practical decisions and consequent actions based on the knowledge acquired. The term mathematical learning theory is reserved for the variety of mathematical models and methods that can be found at the core of each data mining analysis and that are used to generate new knowledge. The term data mining refers therefore to the overall process consisting of data gathering and analysis, development of inductive learning models and adoption of practical decisions and consequent actions based on the knowledge acquired. The term mathematical learning theory is reserved for the variety of mathematical models and methods that can be found at the core of each data mining analysis and that are used to generate new knowledge.	

	Applications od data mining: • Relational marketing. • Fraud detection. • Risk evaluation. • Text mining. • Image recognition. • Web mining. • Medical diagnosis.	
d.	Explain categorical and numerical attributes with proper examples. **Answer:** The attributes contained in a dataset can be categorized as categorical or numerical, depending on the type of values they take on. **Categorical:** Categorical attributes assume a finite number of distinct values, in most cases limited to less than a hundred, representing a qualitative property of an entity to which they refer. **Examples** of categorical attributes are the province of residence of an individual (which takes as values a series of names, which in turn may be represented by integers) or whether a customer has abandoned her service provider (expressed by the value 1) or remained loyal to it (expressed by the value 0). Arithmetic operations cannot be applied to categorical attributes even when the coding of their values is expressed by integer numbers. **Numerical:** Numerical attributes assume a finite or infinite number of values and lend themselves to subtraction or division operations. **For example,** the amount of outgoing phone calls during a month for a generic customer represents a numerical variable. Regarding two customers A and B making phone calls in a week for $27 and $36 respectively, it makes sense to claim that the difference between the amounts spent by the two customers is equal to $9 and that A has spent three fourths of the amount spent by B.	
e.	Differentiate between supervised and unsupervised learning. **Answer:**	

Unsupervised Learning	Supervised Learning
The machine is given huge sets of data that are not labelled as inputs to analyse.	The input is in the form of raw data that is labelled.
The machine needs to figure out the output on its own by identifying patterns in the raw data provided to it.	The machine is already fed with the required feature set to classify between inputs (hence the term 'supervised').
Divided into two types of problems – Association (where we want to find a set of rules that describe our data) and Clustering (where we want to find groups in our data).	Divided into two types of problems – Regression (outputs are real values) and Classification (outputs are categories).
K-means for clustering problems and Apriori algorithm for association rule learning problems.	Linear regression for regression problems, Random Forest for classification and regression problems, Support Vector Machines for classification problems.

f. Explain the following normalization techniques:

(i) Decimal scaling

(ii) Min-max

Answer:

Decimal scaling:

Decimal scaling is based on the transformation

$x_{ij} = x_{ij} / 10^h$,

where h is a given parameter which determines the scaling intensity. In practice, decimal scaling corresponds to shifting the decimal point by h positions toward the left. In general, h is fixed at a value that gives transformed values in the range [−1, 1].

Min-max:

Min-max standardization is achieved through the transformation

$$x'_{ij} = \frac{x_{ij} - x_{\min,j}}{x_{\max,j} - x_{\min,j}}(x'_{\max,j} - x'_{\min,j}) + x'_{\min,j},$$

where

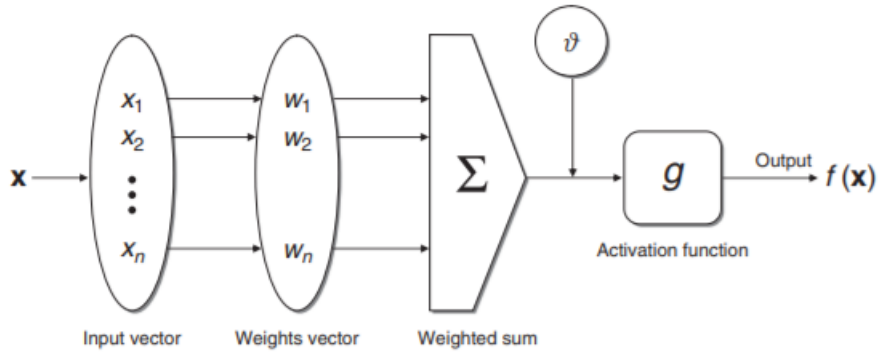
$$x_{\min,j} = \min_i x_{ij}, \quad x_{\max,j} = \max_i x_{ij},$$

are the minimum and maximum values of the attribute $\mathbf{a}_j$ before transformation, while $x'_{\min,j}$ and $x'_{\max,j}$ are the minimum and maximum values that we wish to obtain after transformation. In general, the extreme values of the range are defined so that $x'_{\min,j} = -1$ and $x'_{\max,j} = 1$ or $x'_{\min,j} = 0$ and $x'_{\max,j} = 1$.

3.	**Attempt *any three* of the following:**	**15**
a.	What are the criteria used to evaluate classification methods? **Answer:**	

	Following are the evaluation criteria of classification methods • Accuracy • Speed • Robustness • Scalability • Interpretability	
b.	Explain top-down induction of decision tree. Examine the components of the top-down induction of decision trees procedure. **Answer:** Top-down induction of decision trees 1. In the initialization phase, each observation is placed in the root node of the tree. The root is included in the list L of active nodes. 2. If the list L is empty the procedure is stopped, otherwise a node J belonging to the list L is selected, is removed from the list and is used as the node for analysis. 3. The optimal rule to split the observations contained in J is then determined, based on an appropriate preset criterion. The splitting rule generated in this way is then applied, and descendant nodes are constructed by subdividing the observations contained in J . For each descendant node the conditions for stopping the subdivision are verified. If these are met, node J becomes a leaf, to which the target class is assigned according to the majority of the observations contained in J . Otherwise, the descendant nodes are added to the list L. Finally, step 2 is repeated. In the following sections we will examine the following components of the top-down induction of decision trees procedure. **Splitting rules.** For each node of the tree it is necessary to specify the criteria used to identify the optimal rule for splitting the observations and for creating the descendant nodes. As shown in the next section, there are several alternative criteria, which differ in the number of descendants, the number of attributes and the evaluation metrics. **Stopping criteria.** At each node of the tree different stopping criteria are applied to establish whether the development should be continued recursively or the node should be considered as a leaf. In this case too, various criteria have been proposed, which result in quite different topologies of the generated trees, all other elements being equal. **Pruning criteria.**	

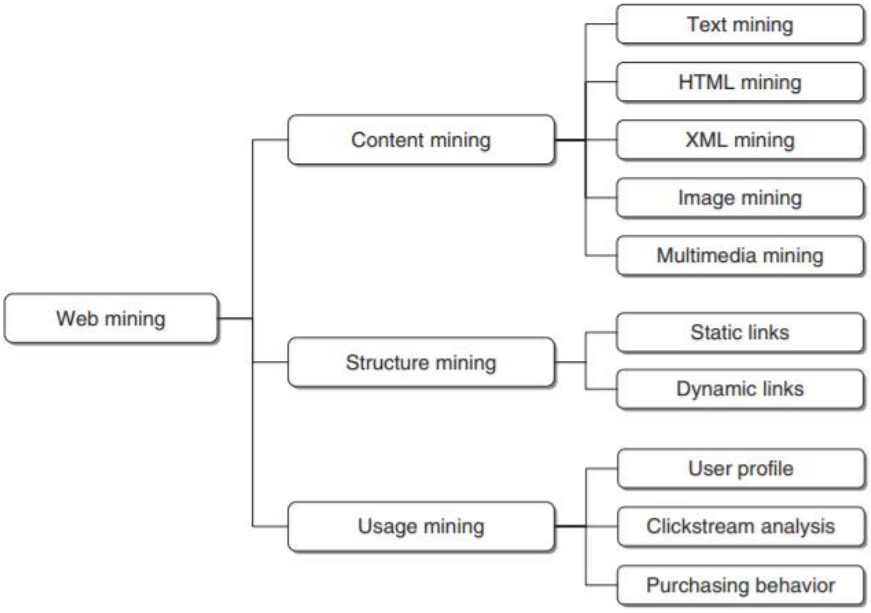
	Finally, it is appropriate to apply a few pruning criteria, first to avoid excessive growth of the tree during the development phase (pre-pruning), and then to reduce the number of nodes after the tree has been generated (post-pruning).	
c.	Short note on Naive Bayesian classifiers. **Answer:** Refer page No. 252 of reference book.	
d.	Write k-means algorithm for clustering. **Answer:** The K-means algorithm receives as input a dataset D, a number K of clusters to be generated and a function dist(x_i, x_k) that expresses the inhomogeneity between each pair of observations, or equivalently the matrix D of distances between observations. Given a cluster Ch, h = 1, 2,...K, the centroid of the cluster is defined as the point z_h having coordinates equal to the mean value of each attribute for the observations belonging to that cluster, that is, $$z_{hj} = \frac{\sum_{\mathbf{x}_i \in C_h} x_{ij}}{\text{card}\{C_h\}}$$ **K-means algorithm** **Step 1.** During the initialization phase, K observations are arbitrarily chosen in D as the centroids of the clusters. **Step 2.** Each observation is iteratively assigned to the cluster whose centroid is the most similar to the observation, in the sense that it minimizes the distance from the record. **Step 3.** If no observation is assigned to a different cluster with respect to the previous iteration, the algorithm stops. **Step 4.** For each cluster, the new centroid is computed as the mean of the values of the observations belonging to the cluster, and then the algorithm returns to step 2.	
e.	Explain the 'Rosenblatt perceptron' form of neural network with diagram. **Answer:** The perceptron, shown in the following figure, is the simplest form of neural network and corresponds to a single neuron that receives as input the values (x_1, x_2,...,x_n) along the incoming connections, and returns an output value f (x). The input values coincide with the values of the explanatory attributes, while the output value determines the prediction of the response variable y. Each of the n input connections is associated with a weight w_j . An activation function g and a constant ϑ, called the distortion, are also assigned. Suppose that the values of the weights and the distortion have already been determined during the training phase. The prediction for a new observation x is then derived by performing the following steps.	

	 *Operation of a single unit in a neural network*	
f.	Write a short note on confusion matrix. **Answer:** Refer page No. 231 of reference book.	
4.	**Attempt *any three* of the following:**	**15**
a.	Write a short note on market basket analysis. **Answer:** The purpose of market basket analysis is to gain insight from the purchases made by customers in order to extract useful knowledge to plan marketing actions. It is mostly used to analyze purchases in the retail industry and in e-commerce activities, and is generally amenable to unsupervised learning problems. It may also be applied in other domains to analyze the purchases made using credit cards, the complementary services activated by mobile or fixed telephone customers, the policies or the checking accounts acquired by a same household. The data used for this purpose mostly refer to purchase transactions, and can be associated with the time dimension if the purchaser can be tracked through a loyalty card or the issue of an invoice. Each transaction consists of a list of purchased items. This list is called a basket, just like the baskets available at retail points of sale. If transactions cannot be connected to one another, say because the purchaser is unknown, one may then apply association rules, to extract interesting correlations between the purchases of groups of items. The rules extracted in this way can then be used to support different decision-making processes, such as assigning the location of the items on the shelves, determining the layout of a point of sale, identifying which items should be included in promotional flyers, advertisements or coupons distributed to customers. Clustering models, are also useful in determining homogeneous groups of items, once an incidence matrix X has been created for the representation of the dataset, where the rows correspond to the transactions and the columns to the items. If customers are individually identified and traced, besides the above techniques it is also possible to develop further analyses that take into account the time dimension of the purchases. For instance, one may generate sequential association rules, or apply time series analysis.	
b.	What is use of web mining methods? What are the different purposes of web mining?	

Answer:

Web mining:

The web is a critical channel for the communication and promotion of a company's image. Moreover, e-commerce sites are important sales channels. Hence, it is natural to use web mining methods in order to analyze data on the activities carried out by the visitors to a website.



Taxonomy of web mining analyses

c. Explain "tactical planning" optimization model for logistics planning.

Answer:

The logistic system includes I products, which will be denoted by the index $i \in I = \{1, 2,...,I\}$. The planning horizon is subdivided into T time intervals $t \in T = \{1, 2,...,T\}$, generally of equal length and usually corresponding to weeks or months. The manufacturing process has at its disposal a set of critical resources shared among the different products and available in limited quantities. These resources may consist of production and assembly lines, to manpower, to specific fixtures and tools required by manufacturing. The R critical resources considered in the logistic production system will be denoted by the index $r \in R = \{1, 2,...,R\}$. Whenever a single resource is relevant to the manufacturing process, the index r will be omitted for sake of simplicity.

Tactical planning

In its simplest form, the aim of tactical planning is to determine the production volumes for each product over the T periods included in the medium-term planning horizon in such a way as to satisfy the given demand and capacity limits for a single resource, and also to minimize the total cost, defined as the sum of manufacturing production costs and inventory costs.

We therefore consider the decision variables

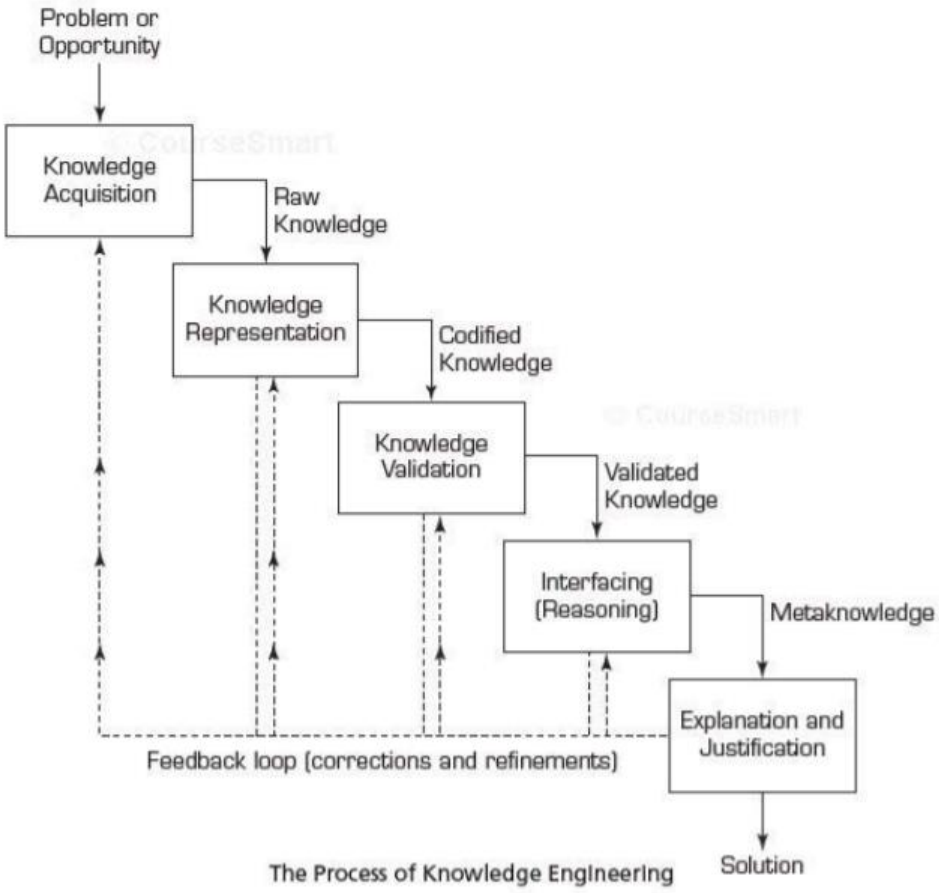
P_{it} = units of product i to be manufactured in period t,

I_{it} = units of product i in inventory at the end of period t, and the parameters

	d_{it} = demand for product i in period t, c_{it} = unit manufacturing cost for product i in period t, h_{it} = unit inventory cost for product i in period t, e_i = capacity absorption to manufacture a unit of product i, b_t = capacity available in period t. The resulting optimization problem is formulated as follows: $\min \quad \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}}(c_{it}P_{it}+h_{it}I_{it})$ $\text{s.to} \quad P_{it}+I_{i,t-1}-I_{it}=d_{it}, \quad i\in\mathcal{I}, t\in\mathcal{T},$ $\sum_{i\in\mathcal{I}} e_i P_{it} \le b_t, \quad t\in\mathcal{T},$ $P_{it}, I_{it} \ge 0, \quad i\in\mathcal{I}, t\in\mathcal{T}.$	
d.	Explain the CCR model **Answer:** Refer page No. 390 of reference book.	
e.	Write short note on efficient frontier. **Answer:** Refer page No. 386 of reference book.	
f.	What is relational marketing? What are the data mining in the field of relational marketing? **Answer:** Relational marketing: Data mining applications in the field of relational marketing, have significantly contributed to the increase in the popularity of these methodologies. Some relevant Applications within relational marketing are: • identification of customer segments that are most likely to respond to targeted marketing campaigns, such as cross-selling and up-selling; • identification of target customer segments for retention campaigns; • prediction of the rate of positive responses to marketing campaigns; • interpretation and understanding of the buying behavior of the customers; • analysis of the products jointly purchased by customers, known as market basket analysis.	

5.	**Attempt *any three* of the following:**	**15**
a.	*Define knowledge management. What are data, information and knowledge?* **Answer:** **Knowledge management** *Knowledge management (KM*) is the systematic and active management of ideas, information, and knowledge residing in an organization's employees. Its purposes include effective and efficient problem solving, dynamic learning, strategic planning, and decision making. KM initiatives focus on identifying knowledge, explicating it in such a way that it can be shared in a formal manner, and leveraging its value through reuse **Data** are facts, measurements, and statistics. **Information** is organized or processed data that is timely (i.e., inferences from the data are drawn within the time frame of applicability) and accurate (i.e., with regard to the original data). **Knowledge** is information that is contextual, relevant, and actionable. and describe its purpose.	
b.	Describe the knowledge management system (KMS) cycle. **Answer:** **The KMS cycle has the steps listed below.** **1. Create knowledge.** **2. Capture knowledge.** **3. Refine knowledge.** **4. Store knowledge.** **5. Manage knowledge.** **6. Disseminate knowledge.**	
c.	Describe how AI and intelligent agents support knowledge management. Relate XML to knowledge management and knowledge portals. **Answer:** Refer Page No. 497 of reference book.	
d.	List and explain characteristics of artificial intelligence. **Answer:** AI techniques usually have features describe below: **1. Symbolic processing:** AI is a branch of science that deals with symbolic, non- algorithmic methods of problem solving. This definition focuses on two characteristics a. **Numeric versus symbolic** Symbolic processing being core of AI still that doesn't mean AI cannot use math. b. **Algorithmic versus heuristic**	

	An algorithm is a step by step process and is intended to find same solution for a specific problem. Human processes are usually non-algorithmic rather human thinking relies more on rules, opinions and gut feelings, learned from previous experiences **2. Heuristics** Heuristics are intuitive knowledge learned from previous experience. By using heuristics in AI we don't have to rethink completely what we have if we encounter a similar problem. Many AI methods uses heuristics to reduce complexity of problem solving a. **Inferencing** AI also includes reasoning capabilities that can build higher level knowledge using existing knowledge represented as heuristics in the form of rules. Inference is the process of deriving logical outcome using set of facts and rules. b. **Machine learning** Learning is an important capability of human being it separates human from other creatures. AI have simplest learning capabilities called machine learning Machine learning allow computer systems to monitor and sense environmental factors and adjust their behavior to react to changes.	
e.	What is knowledge engineering? Explain the process of knowledge engineering. **Answer:** **<u>KNOWLEDGE ENGINEERING</u>** The collection of intensive activities encompassing acquisition of knowledge from human experts and conversion of this knowledge into repository is known as knowledge acquisition. Knowledge engineering requires cooperation and close communication between human experts and knowledge engineer to successfully codify and represent rules that human expert uses to solve problems within specific application domain. Knowledge possessed by human is often unstructured and not explicitly expressedThe major goal of knowledge engineering is to help experts articulate "*how they do what they do "and* to document this knowledge in a reusable form Knowledge engineering deals with steps necessary to build expert system: • Knowledge acquisition: The process of getting knowledge from experts • Knowledge representation: Selecting the most appropriate structures to represent the knowledge. • Knowledge validation: Testing that the knowledge of ES is correct and complete. • Inference: Ability to reason. • Explanation: Ability to explain its advice.	



The Process of Knowledge Engineering

f. What are the areas for expert system applications?

Answer:

Areas for ES Applications

- ***Finance.*** Finance ES include insurance evaluation, credit analysis, tax planning, fraud prevention, financial report analysis, financial planning, and performance evaluation.
- ***Data processing.*** Data processing ES include system planning, equipment selection, equipment maintenance, vendor evaluation, and network management.
- ***Marketing.*** Marketing ES include customer relationship management, market analysis, product planning, and market planning.
- ***Human resources.*** Examples of human resources ES are human resources planning, performance evaluation, staff scheduling, pension management, and legal advising.
- ***Manufacturing.*** Manufacturing ES include production planning, quality management, product design, plant site selection, and equipment maintenance and repair.
- ***Homeland security.*** Homeland security ES include terrorist threat assessment and terrorist finance detection.
- ***Business process automation.*** ES have been developed for help desk automation, call center management, and regulation enforcement.
- ***Health care management.*** ES have been developed for bioinformatics and other health care management issues.

and many more…

****