MEETING 14

DATA MINING

A. LEARNING OBJECTIVES

Students are able to do inversion subject and verb in a sentence, to speak about data mining process, able to understand reading text data mining and able to understand listening text about language translator.

B. INVERT THE SUBJECT AND VERB WITH QUESTION WORDS

The words such as what, when , where, why, and how have to different functions in the sentence, first they can introduce a question and in this case the subject and verb that follow are inverted.

Example:

What is data mining?

Where can I learn about data mining process ?

Why are computer science students studying data mining process?

Those words an join together two clauses, and in this case the subject and the verb that follow are not inverted (Deborah Phillips:2001)

Example:

I don't know what data mining process is

Where I can learn data mining process, I will take computer course.

Do you know why computer science students are studying data mining process.

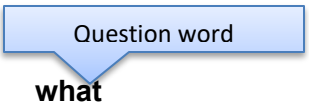
Each of these examples are two clauses joined by a question word. Notice that the subjects and verbs that follow the question words what, where and why are not inverted in this case. The following list of the question words and their sentence patterns:

Inverted subjects and verbs with question words are WHO, WHAT, WHEN, WHERE, WHY and HOW.

When the question world introduces a question, the subject and the verb are inverted. Example:

Question word	V	S	?
What	**are**	**they?**	

When the question word connects two clause, the subject and the verbs that follow are not inverted.



S	V		S	V
I	know	what	they	are

C. EXERCISES

1. Structure

1.1 Each of the following sentences contains question word then underline the subjects and the verbs.

1) The computer programmer is not sure when will the new application be ready.
   O True
   O False
2) The lecturer does not understand why many student didn't understand the data mining process.
   O True
   O False
3) How the students can understand the data mining process deeply?
   O True
   O False
4) How is the subject understood by students?
   O True
   O False
5) computer science students can explain how the computer process the data inside the motherboard.
   O True
   O False
6) He is explaining how does a system works inside the motherboard.
   O True
   O False
7) Not even the computer programmers know when the new application will be released.
   O True

O False

8) How long it has been since you studied in this university?
   O True
   O False
9) The computer programmers doubt that the new application will be ready soon.
   O True
   O False
10) Do you know why he wants to take computer course?
   O True
   O False
11) In the store were many new released computer devices.
   O True
   O False
12) Inside the motherboard is the computer processor.
   O True
   O False
13) There a processor is inside the motherboard.
   O True
   O False
14) At the electronic fair was the computer too expensive for me to buy.
   O True
   O False
15) On my table the devices are that need to be fixed.
   O True
   O False
16) on the corner of the room are the database books that you are looking for.
   O True
   O False

2. Speaking

2.1 work in pair to practice the conversation below!

Sam	Do you know what data mining is?
Nass	I know, data mining is process on finding pattern in big data.
Sam	What are the application of data mining?
Nass	Data mining can be found in financial banking, customer segmentation, future healthcare and many more.
Sam	How is data mining use in business?
Nass	A company can use data mining to process raw data to get clear information.
Sam	Can you give me example on how data mining can be helpful to business?
Nass	By using data mining businesses can learn about their customers to increase sell or make marketing strategies.
Sam	Why data mining is important business tool?
Nass	Because data mining can help to see sales trend, can develop smarter business marketing campaign.
Sam	Do you know how I can start learning data mining?
Nass	I know you start from popular languages for dating mining, they are R, Python and SQL
Sam	Thank you very much for the information, that is really helpful
Nass	Any time. Good luck on your language programming lesson.

3. Reading

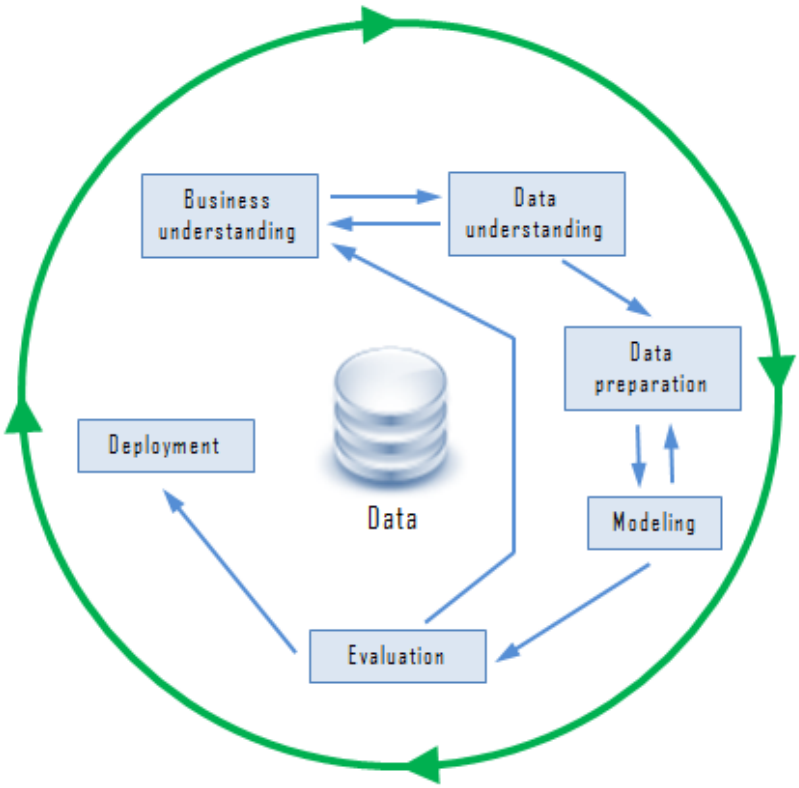
3.1 Read and record loud the reading text in your phone!

Data Mining Processes

Data mining is a promising and relatively new technology. Data mining is defined as a process of discovering hidden valuable knowledge by analyzing large amounts of data, which is stored in databases or data warehouse, using various data mining techniques such as machine learning, artificial intelligence(AI) and statistical. Many organizations in various industries are taking advantages of data mining including manufacturing, marketing, chemical, aerospace… etc, to increase their business efficiency. Therefore, the needs for a standard data mining process increased dramatically. A data mining process must be reliable and it must be repeatable by business people with little or no knowledge of data mining background. As the result, in 1990, a cross-industry standard process for data mining (CRISP-DM) first published after going through a lot of workshops, and contributions from over 300 organizations.

Cross-Industry Standard Process for Data Mining (CRISP-DM) consists of six phases intended as a cyclical process as the following figure:

Data Mining Processes - CRISP-DM



Cross-Industry Standard Process for Data Mining (CRISP-DM)

Business understanding

In the business understanding phase: First, it is required to understand business objectives clearly and find out what are the business's needs. Next, we have to assess the current situation by finding the resources, assumptions, constraints and other important factors which should be considered. Then, from the business objectives and current situations, we need to create data mining goals to achieve the business objectives within the current situation. Finally, a good data mining plan has to be established to achieve both business and data mining goals. The plan should be as detailed as possible.

Data understanding

First, the data understanding phase starts with initial data collection, which we collect from available data sources, to help us get familiar with the data. Some important activities must be performed including data load and data integration in order to make the data collection successfully. Next, the "gross" or "surface" properties of acquired data need to be examined carefully and reported. Then, the data needs to be explored by tackling the data mining questions, which can be addressed using querying, reporting, and

visualization. Finally, the data quality must be examined by answering some important questions such as “Is the acquired data complete?”, “Is there any missing values in the acquired data?”

Data preparation

The data preparation typically consumes about 90% of the time of the project. The outcome of the data preparation phase is the final data set. Once available data sources are identified, they need to be selected, cleaned, constructed and formatted into the desired form. The data exploration task at a greater depth may be carried during this phase to notice the patterns based on business understanding.

Modeling

First, modeling techniques have to be selected to be used for the prepared dataset. Next, the test scenario must be generated to validate the quality and validity of the model. Then, one or more models are created by running the modeling tool on the prepared dataset. Finally, models need to be assessed carefully involving stakeholders to make sure that created models are met business initiatives.

Evaluation

In the evaluation phase, the model results must be evaluated in the context of business objectives in the first phase. In this phase, new business requirements may be raised due to the new patterns that have been discovered in the model results or from other factors. Gaining business understanding is an iterative process in data mining. The go or no-go decision must be made in this step to move to the deployment phase.

Deployment

The knowledge or information, which we gain through data mining process, needs to be presented in such a way that stakeholders can use it when they want it. Based on the business requirements, the deployment phase could be as simple as creating a report or as complex as a repeatable data mining process across the organization. In the deployment phase, the plans for deployment, maintenance, and monitoring have to be created for implementation and also future supports. From the project point of view, the final report of the project needs to summary the project experiences and review the project to see what need to improved created learned lessons. The CRISP-DM offers a uniform framework for experience documentation and guidelines. In addition, the CRISP-DM can apply in various industries with different types of data. **adopted from http://www.zentut.com/data-mining/data-mining-processes/**

3.2 Answer the question below:

1) What is Data mining?
2) What are the three most common Data mining techniques?
3) What industries are using data mining?
4) What are the six phases of Cross-Industry Standard Process for Data Mining?
5) What is the last step of Business understanding?
6) What is the first phase of data understanding ?

3.3 Match the term (a-h) and the statement (1-8)

a. Data mining
b. deployment
c. evaluation
d. data preparation
e. data understanding
f. data mining techniques
g. CRISP-DM
h. modeling

1) A process of discovering hidden valuable knowledge by analizing large amounts of data, which of a stored in database.
2) It Consumes 90% of the time .
3) It needs to be assessed carefully involving stakeholders to make sure that the models are met business initiatives.
4) The model result must be evaluated in the context of business objectives in the first place.
5) Cross-industry Standard Process for Data Mining.
6) Machine learning, Artificial Intelligence and statistic.
7) The plans for deployment, maintenance, and monitoring have to be created for implementation and also future support.
8) First phase starts with data collection from avalaible data sources.

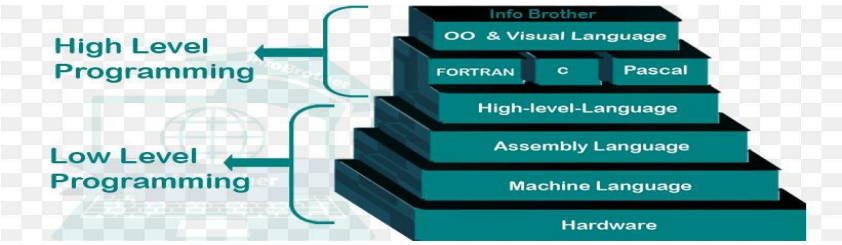
3.4 Write T for true statement and write F for false statement

1. Data mining is able to find the unseen data in database.
   O True
   O False
2. Data mining doesn't have relation to artificial intelligence.
   O True
   O False
3. Data mining is very useful for companies or organization.
   O True
   O False
4. The data mining used techniques such as machine leaning in finding data.
   O True
   O False
5. Data mining was first published without workshop in organization.
   O True
   O False
6. The first phase in business understanding is clear objective of the business necessities.
   O True
   O False
7. In data understanding phase, the last step is examining data carefully and reporting it. F
   O True
   O False
8. The data preparation takes the longest time.
   O True
   O False
9. The final models in Modeling techniques don't need to be assessed by the stakeholders.
   O True
   O False
10. Deployment phase should be complicated.
    O True O False

4. Listening

4.1 Listen to Sultan Issa Hampton talking about translator language and write T for true statement and F for false statements.

1. A translator is another name of programming language processor.
   O True
   O False
2. Function of programming language processor is to translate different language.
   O True
   O False
3. programming language processor will lose the functional or logical structure of the original program when translating different language.
   O True
   O False
4. C ++ and Java are low-level and human-unreadable computer languages.
   O True
   O False
5. Java bytecode is high -level languages.
   O True
   O False
6. assembly language and machine code are intermediate-level languages.
   O True
   O False



https://favpng.com/png_view/

5. writing

5.1 translate the sentences below

1) Data mining process must be reliable ad it must be reliable by business people with little or no knowledge of data mining background.

_____

_____

2) Many organizations in various industries are taking advantages of data mining including manufacturing, marketing, chemical. Aerospace..etc.

_____

_____

3) A cross-industry standard process for data mining first published after going through a lot of workshops, and contributions from over 300 organization.

_____

_____

4) The model result must be evaluated in the context of business objective.

_____

5) Data mining needs to be presented in such a way that stakeholders can use it.

_____

6) It is required to understand business objectives clearly and find out what are the business' needs.

_____

_____

7) Many companies are taking advantages of data mining including manufacturing, marketing, chemical etc.

_____

_____

8) Data mining process must be reliable and it must be repeatable by business people.

_____

_____

6. .Key word

English	Synonym	Indonesia
Affairs against, ambiguous, appreciation, approach, approaches, containing, conversations, corpus, determining, equivalent, estimate, Gained, handheld, include, intent, lexicons, perform. preserving, relies, representative, Represented, segment, sufficiently, usage,		

D. BIBLIOGRAPHY

Phillips Deborah, Complete Course for the Toefl Test , Preparation for computer and Paper test, Longman a Pearson Education company, N.Y 2001.

https://favpng.com/png_view/types-of-programming-languages-low-level-programming-language-high-level-programming-language-high-and-low-level-computer-programming-png/Y9XvYRpA

http://www.zentut.com/data-mining/data-mining-processes/