c. extracts useful information from data already present m databases

d. discovers algorithms previously unknown on existing data

B. Look at the contact lens data in this table and complete the structural description that follows.

| Age | Spectacle prescription | Astigmatism | Tear production rate | Recommended lenses |
|---|---|---|---|---|
| young | myope | no | reduced | none |
| young | myope | no | normal | soft |
| young | myope | yes | reduced | none |
| young | myope | yes | normal | hard |
| young | hypermetrope | no | reduced | none |
| young | hypermetrope | no | normal | soft |
| young | hypermetrope | yes | reduced | none |
| young | hypermetrope | yes | normal | hard |
| pre-presbyopic | myope | no | reduced | none |
| pre-presbyopic | myope | no | normal | soft |
| pre-presbyopic | myope | yes | reduced | none |
| pre-presbyopic | myope | yes | normal | hard |
| pre-presbyopic | hypermetrope | no | reduced | none |
| pre-presbyopic | hypermetrope | no | normal | soft |
| pre-presbyopic | hypermetrope | yes | reduced | none |
| pre-presbyopic | hypermetrope | yes | normal | none |
| presbyopic | myope | no | reduced | none |
| presbyopic | myope | no | normal | none |
| presbyopic | myope | yes | reduced | none |
| presbyopic | myope | yes | normal | hard |
| presbyopic | hypermetrope | no | reduced | none |
| presbyopic | hypermetrope | no | normal | soft |
| presbyopic | hypermetrope | yes | reduced | none |
| presbyopic | hypermetrope | yes | normal | none |

myope: a short-sighted person

If tear production rate = reduced then recommendation = ..... .

Otherwise, if age = young and astigmatic = ..... .

Then recommendation = ..... .

C. Work with a partner to answer the questions below.

1. The author suggests data mining for solving the problem of masses of data already present in databases and not used explicitly. Do you agree? Why or why not?

2. Can you think of any disadvantages of data mining? What do you think of data mining and ethics?

values

## Reading Strategy

### Understanding the Difference between Topic and Main Idea

A topic is the subject of a piece of writing.

A main idea is the writer's message about the topic.

Typically, writers organize their writing around one or two main ideas.

For example: The topic and main idea of Data Mining on pages 15-16:

Topic: Data mining

Main Idea: Data mining is the process of discovering meaningful patterns that are already in the data but were previously unseen.

A. **Topic and main idea of a paragraph. Read the paragraph below and find the topic and main idea. Discuss your answers with a partner.**

Security is a broad topic and covers a multitude of sins. In its simplest form, it is concerned with making sure that nosy people cannot read, or worse yet, secretly

large quantity violation of law excessively curious

modify messages intended (planned) for other recipients. It is concerned with people trying to access remote services that are not authorized (permitted) to use. It also deals with ways to tell whether that message purportedly (as it appears) from the IRS (Internet Revenue Service) "Pay by Friday, or else" is really from the IRS and not from the Mafia. Security also deals with the problems of legitimate (legal; valid) messages being captured and replayed, and with people later trying to deny that they sent certain messages.

(Tanenbaum & Wethe.rall: p. 763)

1. The topic of this paragraph is ..... .

a. Legitimacy (conforming to rules)

b. Unauthorized (illegal) services

c. Security

2. The main idea of this paragraph is ..... .

a. malicious (mean; ill-natured) people trying to get some benefit

b. security from different angles

c. numerous pitfalls that security deals with

B. Topic and main idea of a longer selection. Skim the reading on pages 27-31 and find the topic and main idea. Discuss your answers with a partner.

1. The topic of the reading is ..... .

a. Communication satellites

b. Computer networks

c. Information processing

2. The main idea of the reading is .....

a. technical issues involved in network design

b. development of the computer in the 21st century

c. classification of networks

## Building Vocabulary

### Compound Words

Compound words are created by combining two shorter words. Some of these words have hyphens that connect their component parts. You can usually figure out the meaning a compound word by breaking it down into its simpler parts.

*marketplace* -the place used as a market

The component parts of many compound words are separated by a hyphen:

customer-centered - the customer is the most important feature

well-worn -old- used very often

A. Underline the compound words in the following sentences. Then explain what each one means or provide a synonym. (Note: some sentences. have more than one compound word.).

1. There is a new section on Bayesian networks with a description of how to learn classifiers based on these networks and how to implement them efficiently using all-dimensions trees.

2. If costs are known, they can be incorporated into a financial analysis of the decision-making process.

3. Attribute-selected classifier selects attributes, reducing the data's dimensionality before passing it to the classifier.

4. CV parameter selection optimizes performance by using cross-validation to select parameters.

5. Ever since I started fumbling with joysticks and game controllers, I feel I have been falling slow-motion into a place I didn't really understand or appreciate.

6. A conceptually general way to address multiclt1ssproblems is known as pairwise classification.

B. The chart below includes several examples of compound word groups. Try adding an example of your own to each group. Then, give a simple phrase with you r word on the right side of the chart.

| | |
|---|---|
| well-worn<br>well-developed<br>well-documented | |
| database<br>databank | |
| service-oriented<br>instance-based | Instance-based learning |

B. Pair Work. Work with a partner on the word families and complete the table. Use a dictionary if necessary.

| Noun | Verb | Adjective | Adverb |
|---|---|---|---|
| | Learn | | |
| | | | Structurally |
| | | Intelligent | |
| | Mine | | |
| Definition | | | |
| Pattern | | | |
| | Generate | | |

D. Use the information in the table above to 1complete the sentences.

1. We would all test the growing gap between ..... of data and our understand mg of it.

2. People have been seeking ..... in data since human life began.

3. We are interested, in techniques for finding and describing ..... patterns in data as a tool for helping with explaining that data and making predictions from it.

4. Machine ..... provides the technical basis for data mining.

5. Data complexity calls for new techniques and tools that can ..... turn low-level data into high-level and useful knowledge.

6. Data mining is often ..... as process of extracting valid, previously unknown, comprehensible information from large databases in order to improve and optimize business decisions.

7. The patterns that are ..... can be examined and used to inform future decisions.

## Word Magnifier

**Study the word complexity used in these sentences.**

1. As the world grows in complexity, overwhelming us with the data it generates, data mining becomes our only hope for elucidating the patterns that underlie it.

2. I was astonished by the size and complexity of the problem.

In the first sentence, complexity means "the state of being formed by many parts", whereas in the second sentence, it means "the state of being difficult to understand".

**Use the information in the box to decide which idea the word 'complexity' presents in each sentence.**

1. He became quite nervous when he noticed the complexity of the math problems in his final exam.

2. To reduce the design complexity, most networks are organized as a stack of layers.

3. Deliberate complexity in his sentences confuses people.

4. The growth of databases incomplexity brings data mining to the forefront of new business technologies.

| Time Clauses: When; Once | |
|---|---|
| We can use *when* to show that one action happens immediately after another action.<br>1. When the process is complete, the mining software generates a report.<br><br>We can use *once* in place of *when* to emphasize the completion of the first action.<br><br>2. Once the distinguishing characteristics are found, they can be put to work. | If the subjects are the same in both actions, the time clause can be shortened. To do so, once is followed by the past participle of the verb required.<br><br>Once found the distinguishing characteristics can be put to work. |

A. Link each pair of actions using the time clauses above.

1.

a. The mining software generates a report.

b. An analyst goes over the report.

2.

a. Work on a data mining problem begins.

b. It is necessary to bring all the data together into a set of instances.

3.

a. A decision-tree induction method prunes away a subtree.

b. It applies a statistical test that decides whether that subtree is justified by the data.

4.

a. You use a search engine.

b. It provides a set of links related to your search.

5.

a. Data is identified.

b. It is mined.

6.

a. Data is cleaned.

b. It is freed from duplicate information.

7.

a. The discriminant function is constructed.

b. It is used to predict the class of a given data.

B. Complete the passage using the time clauses from the box.

| When identified/ Once you have discovered the patterns<br>When you have a proper domain understanding/ When integrated |
|---|

The main stages of data mining process are: domain understanding; data selection; cleaning and preprocessing; discovering patterns; interpretation; and reporting and using discovered knowledge. .....1..... you can identify useful data. Data is never clean and, in the form, suitable for data mining. Hence, .....2....., data must be cleaned and preprocessed. Preprocessing involves integrating data from different sources. .....3....., data is represented as inputs to· data discovery stage. The data-pattern-discovery stage is the heart of the entire data mining process. In the academic literature, it is the only stage that is referred to as data mining. .....4....., you report and put to use the discovered knowledge to generate new actions or products and services or marketing strategies.