Data mining summarizing(unit-1)

ملاحظه// هذا الكتيب فقط للمراجعه السريعه ولا يعد مرجعاً مكتملاً(فقط شرح للمفاهيم الاساسيه للماده)

Data mining (1):

is a process of using computing power to apply methodologies including new techniques of knowledge discovery to data.

هي عمليه استخدام القدره او الطاقه الحسابيه لعمل طرق علميه من خلال تقنيات جديده من المعرفه لاكتشاف البيانات

Data mining (2) :

is a process to analyzing data from different perspectives and summarizing it to be useful information

هي عمليه تحليل البيانات من تنبئات و مصادر مختلفه و تلخيصها لتصبح ذات فائده

Data:

any fact or number or text that processed by computer

هي أي حقائق او ارقام او نصوص يتم معالجتها بالكمبيوتر

Types of Data:

- ✓ Operational data- transaction data
- ✓ Non-operational data
- ✓ Meta data

Knowledge:

it's an information can be converted into knowledge about historical patterns and future.

المعرفه : هي معلومات يتم اخفائها داخل المعرفه حول انماط تاريخيه او مستقبليه

Warehouse:

It's a process of centralized data management retrieval

التخزين هو عمليه تخزين و اداره و استرجاع البيانات المركزه

Data mining Tasks:

- ✓ Classification
- ✓ Clustering
- ✓ Association Rule Discovery
- ✓ Sequential Pattern Discovery
- ✓ Regression
- ✓ Deviation Detection

Classification:

Arranges the data into predefined groups.

التصنيفات: هي عمليه ترتيب البيانات داخل مجموعات معرفه

Clustering:

Is like classification but the groups are not predefined, so the algorithm will try to group similar items together.

هي عمليه تشبه التصنيفات لكن المجموعات لا تكون معرفه لذلك الخوارزميات سوف تجرب المجموعات المتشابهه

Association Rule:

Given a set of records, each of which contain some number of items from a given collection..Or Searches for relationships between variables.

القاعده الاجتماعيه : تعطي مجموعه من السجلات التي تحتوي على بعض الارقام من العناصر المجتمعه

Regression :

Attempts to find a function which models the data with the least error.
A common method is to use Genetic Programming.

الارتداد: هو محاوله ايجاد داله لتشكيل البيانات مع اقل عدد ممكن من الاخطاء باستخدام البرمجه الوراثيه

Visualization:

complement to other DM techniques like Segmentation, … etc.
process of extracting previously unknown , valid , and actionable (understandable) information from large databases

<div dir="rtl">هو عمليه استخلاص البيانات المفقوده و غير المعروفه من قواعد بيانات سابقه ضخمه</div>

Data Mining Elements:

Data mining consists of five major elements:
- ✓ Extract, transform, and load transaction data onto the data warehouse system.
- ✓ Store and manage the data in a multidimensional database system.
- ✓ Provide data access to business analysts and information technology professionals.
- ✓ Analyze the data by application software.

- ✓ Present the data in a useful format, such as a graph or table.

Analysis Levels:

- ✓ Artificial neural networks:

Non-linear predictive models that learn through training and resemble (imitate) biological neural networks in structure.

<div dir="rtl">الشبكه العصبيه الاصطناعيه: هو عباره عن نموذج غير خطي يتعلم بالتدريب و التقليد و هو يشبه الجهاز العصبي الطبيعي</div>

- ✓ Genetic algorithms:

Optimization techniques that use processes such as genetic combination, mutation (change), and natural selection in a design based on the concepts of natural evolution.

<div dir="rtl">الخوارزميه الجينيه: هي عباره عن تقنيات اختياريه تستخدم مجموعات وراثيه لاختيار تصميم اساسي للمفاهيم الوراثيه الطبيعيه</div>

Decision trees:

Tree-shaped structures that represent sets of decisions.
These decisions generate rules for the classification of a data set.

<div dir="rtl">شجره القرارات هي مجموعه من التراكيب التي تشبه الشجره و هي مجموعه من القرارات التي تولد قواعد لتصنيف البيانات</div>

Rule induction:

The extraction of useful if-then rules from data based on statistical significance.

القاعده الحثيه: هي عمليه استخلاص القواعد المفيده (اذا كان فإن) من البيانات الاساسيه الموجوده في المستندات الاحصائيه

Technological infrastructure:

There are two critical technological drivers:

- ✓ Size of the database: The more data being processed and maintained, the more powerful the system required.

- ✓ Query complexity: The more complex the queries and the greater the number of queries being processed, the more powerful the system required.

- ✓ Relational database storage and management technology is adequate for many data mining applications less than 50 gigabytes.

- ✓ However, this infrastructure needs to be significantly enhanced to support larger applications.

- ✓ Some vendors have added extensive indexing capabilities to improve query performance

  - ✓ Data mining identifies trends within data that go beyond simple analysis.

  - ✓ Through the use of sophisticated algorithms, non-statistician users have the opportunity to identify key attributes of processes and target opportunities.

  - ✓ The term data mining is often used to apply to the two separate processes of knowledge discovery and prediction.

Data Mining Issues:

key issues raised by data mining technology are:

- ✓ 1.Business issues : analyzing routine business transactions and classifications.

- ✓ 2.social issues: Data mining makes it possible to analyze routine business transactions and glean a significant amount of information about individuals buying habits and preferences.

- ✓ 3. Mining Methodology Issues: Pertain to data mining approaches applied and their limitations.

- ✓ 4. Cost: While system hardware costs have dropped dramatically within the past few years, data mining and data warehousing tend to be self-reinforcing.

- ✓ 5. User Interface Issues: The knowledge discovered by data mining tools is useful as long as it is interesting, and above all understandable by the user.

- ✓ 6.Data Source issue:
  -An excess of data appear when we have more data than we can handle.
   - different types of data are stored in a variety of repositories.

Data Mining software:

- ✓ Data mining software is one of a number of analytical tools for analyzing data.

- ✓ It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified.[1]

Analysts separate data mining software into two groups: data mining tools and data mining applications.

- ✓ **Data mining tools**: provide a number of techniques that can be applied to any business problem. (SPSS is a general tool)

- ✓ **Data mining applications**, on the other hand, embed techniques inside an application customized to address a specific business problem.

Data Mining Tools
Data mining tools provide both developers and business users with an interface for discovering, manipulating, and analysing corporate data

---