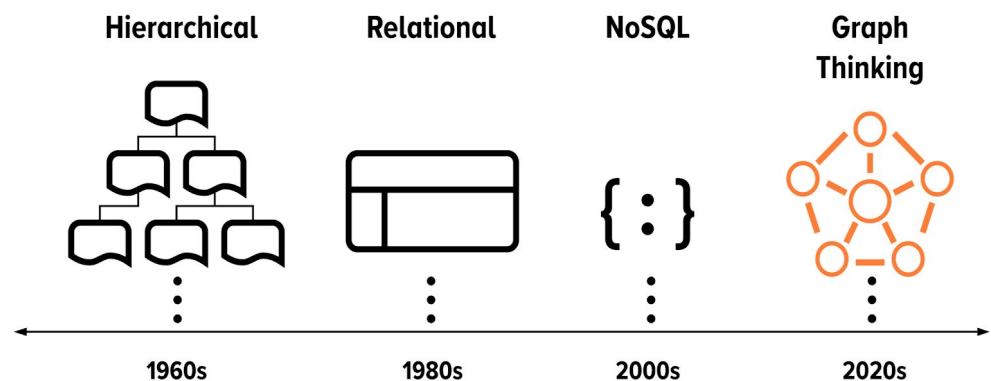


- 1) Articulate what data mining is within the context of its evolution, as an interdisciplinary subfield, as an essential step of knowledge and pattern discovery, and as a type of business analytics;

→As a context of its evolution

- ◆ Let me first tell you a brief history, way back in the 1960s up to 1970s was the first generation of modern computing were technologies about processing data collection and database management systems are still in research and development. Afterward, data analysis appears because that time the amount of data collection continued to increase significantly thanks to computer hardware technology for helping database performance to quickly acquire relevant information like relational, transactional data. Now technology is still a trend and continually emerging. There is a strategy called adopting cloud computing, it's because of the four fundamental blocks and concepts that define the cloud. These are highly available, fault-tolerant, scalability, and elasticity. Huge volumes of data have been stored beyond cloud databases but the question is, what would you do to this data? So data mining is often used to refer to the entire knowledge discovery process turning unused data into golden mines. The image below is a timeline graph referring to the evolution of database technology, they have different functionalities, capabilities, and techniques to utilize.

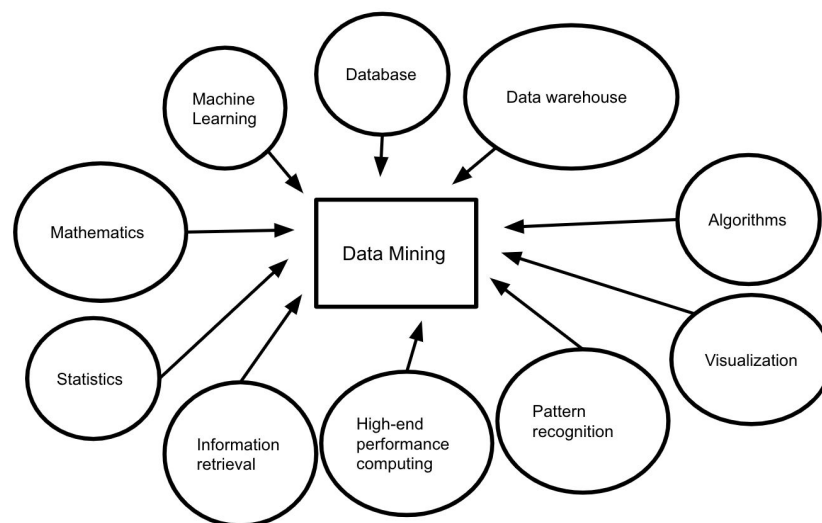


→As an interdisciplinary subfield

◆ Data mining has associated techniques (**computer science and statistics**) from other domains, the following are the most common discipline available for creating intelligent methods:

- Database
 - Also called DBMS (Database Management System), it consists of a collection of interrelated data and software programs that manage to access needed information. An example is an object-oriented database like PostgreSQL, NoSQL like DynamoDB.
- Statistics
 - The most common basic terms in statistics you'll ever come cross are mean, median, and mode. In regards to data mining, it is a mathematical function that describes model data behavior.
- Machine Learning
 - Is a process where computers can learn automatically to recognize the complex pattern to make an intelligent decision based on the data. For example, is a web application that provides powered predictions for tracking the metrics and progress of KPI(Key Performance Indicator).
- Algorithms
 - The role of algorithms takes the input to process output and a tool to solve computational-problems. On a side note, the algorithm is efficient when time and space are not huge. An example of an algorithm is the Insertion sort because this algorithm is efficient when sorting smaller inputs. Worst-case running of $O(n^2)$ complexity.
- Visualization
 - It has an inherent connection with data mining, is a process of analyzing data sets. It is referred to understandable if high dimensional data was reduced because data integrity is maintained. For example, nominal attributes are fit to use a pie chart because the order is arbitrary.

- Mathematics
 - It is an important aspect of data mining because it is dealing with **numerical** and **nominal** attributes.
- High-Performance computing
 - It is an ability to process data and perform complex calculations at high speed. In general, the computing method is optimized. An example is the usage of **indexes, normalization, query languages, architecture** remember speed is important.
- Data warehouse
 - Is a centralized repository of information collected from the different data source using a unified schema and modeled multiple dimensional data structures. In general query and analysis tools are used in this phase. An example is **archived-data, pre-processed data.**
- Pattern recognition
 - A data sets may contain **noise** and **outliers**, it is to be said interesting pattern if easily understood by **humans, valid data, useful,** and **interestingly new.**
- Information retrieval
 - Simply a process of searching for information to documents. An example is **document data** where the goal is how many times a word or text repeat in the documents.



→As an essential step of knowledge and pattern discovery

- ◆ Pattern discovery is the most important aspect of data mining. A pattern a set of items, subsequences, or substructures that occur frequently together or strongly correlated with datasets. It represents the intrinsic and important properties of datasets. Pattern discovery is a process of uncovering patterns from massive data sets. An example is a transactional data what products were often purchased together? Here is another interesting example in software engineering what code segments likely to contain copy-and-paste bugs? Another meaning of pattern discovery is a foundation for data mining tasks. The image below is an example of pattern discovery that pertains to transactional data. It refers to an item sets and count of an item repeated in a transaction.

<i>TID</i>	<i>Items</i>
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

→As a type of business analytics

- ◆ Business analytics is a subset of business intelligence it is an exploration and investigation of past business performance to gain insight and drive business planning. Data mining techniques cover strategic business decision making. The major concerns across the business are customer management, revenue growth, bringing operational efficiency, and so on. In summary without data mining, businesses may not be able to perform the best possible applied through market analysis, customer satisfaction, business competitors. I would say data mining is the core of the business if only doing it right.

2) Synthesize the benefits of data mining in extracting knowledge from data.

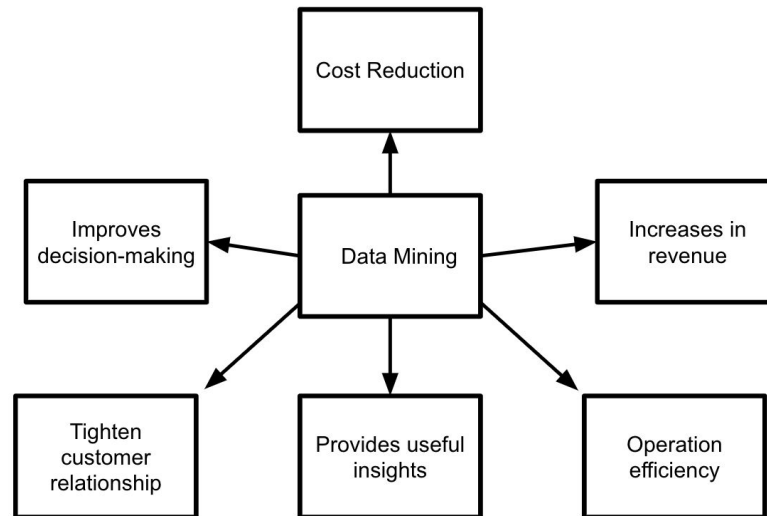
→Benefits of data mining

- ◆ Data mining offers pros and cons to individuals or businesses. But in my opinion, here are the practical advantages.
 - Cost reduction
 - This is an important matter in data mining because of one disadvantage is expensive. What do I mean by this? Well, if a company anticipates a higher loss of income versus the normal profit then it's best for investment in a solution that causes a decrease in income. Like data specialists, maintenance, system provisions, and so on.
 - Improved-decision making
 - Business intelligence provides historical, current, and future predictions as to where the company leading to. So businesses must have a solid understanding of the customer, service, resources, competitors, and market supply. Data mining allows decision-makers to decide what is the best possible decision from the knowledgeable discovery of the datasets.
 - Increase in revenue
 - Predictive analysis and data mining can boost revenues and profit margins by having a fantastic service from the customer. There are certain metrics to consider that may increase leads and sales and those are dependent from the knowledge discovered in the datasets.
 - Tighten customer relationship
 - One advantage of data mining is customer relationship. It offers the business to choose or what type of customers to approach with different kinds of products. Another to consider is loyalty it allows you to reward consumer behavior through coupons, discounts, incentives.
 - Provides useful insights
 - Data mining allows you to analyze historical data about overall business performance like customer

transactions, products with these as a flavor it provides a useful hint for hidden insights.

- Operation efficiency
 - Data mining helps the company to have quality service to the customer, remember better service more customers means more profit.

The image below pertains to the benefits of data mining.



3) Relate the challenges of data mining to a specific discipline.

→Data mining challenges to a specific discipline

- ◆ There are many constraints in data mining and these are the most common for big data.
 - Data quality
 - Data quality is the precondition for analyzing data and it refers to qualitative and quantitative pieces of information. Data quality problems like noise and outliers, missing values, duplicate data are a tedious task to accomplish especially if the origin of the source is coming to production. Remember it's very important when it comes to decision is having a data quality by asking where the problem is coming from?

- Data sources
 - The diversity of data sources brings abundant **data types** and **complex data structures** which increases the difficulty of **data integration**.
- Huge volume and speed
 - Data volume is **tremendous**, and it is difficult to judge quality within reasonable amount of time. To me, this is a great challenge to the existing techniques of **data preprocessing** quality. Efficiency and speed are really important to obtain data quality. It avoids the misleading conclusion, eventually leading to decision-making.
- Unified and standardization
 - No **unified** and approved **data quality standards** have been formed in order to guarantee the **product quality** and **better decision** valid standard should be imposed.

Links and References:

[*The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011,*](#)

<https://www.itexico.com/blog/what-is-cloud-adoption#:~:text=Cloud%20adoption%20is%20a%20strategy,manage%2C%20and%20process%20critical%20data.,>

[https://www.oreilly.com/library/view/the-practitioners-guide/9781492044062/ch01.html,](https://www.oreilly.com/library/view/the-practitioners-guide/9781492044062/ch01.html)

https://en.wikipedia.org/wiki/Data_mining#:~:text=Data%20mining%20is%20an%20interdisciplinary,comprehensible%20structure%20for%20further%20use.,

[https://cs.ccsu.edu/~markov/ccsu_courses/DataMining-1.html,](https://cs.ccsu.edu/~markov/ccsu_courses/DataMining-1.html)

[https://www.statisticshowto.com/statistics-basics/,](https://www.statisticshowto.com/statistics-basics/)

<https://www.slideshare.net/HoangNguyen446/business-analytics-and-data-mining>

.

[https://www.youtube.com/watch?v=GehnEqjj4X8,](https://www.youtube.com/watch?v=GehnEqjj4X8)

[https://www.youtube.com/watch?v=nrMeLHSO-pw,](https://www.youtube.com/watch?v=nrMeLHSO-pw)

[https://wisdomplexus.com/blogs/data-mining-pros-cons/,](https://wisdomplexus.com/blogs/data-mining-pros-cons/)

[https://www.smallbusinessrainmaker.com/small-business-marketing-blog/how-to-grow-sales-with-the-data-mining-information-already-at-your-fingertips,](https://www.smallbusinessrainmaker.com/small-business-marketing-blog/how-to-grow-sales-with-the-data-mining-information-already-at-your-fingertips)

[https://www.educba.com/advantages-of-data-mining/,](https://www.educba.com/advantages-of-data-mining/)

<https://referralrock.com/blog/benefits-of-data-mining-in-crm/>

<https://datascience.codata.org/articles/10.5334/dsj-2015-002/>