Data Analytics

Lesson 10.

Introduction to Data Mining

Dr. Hai Tran

hai.tran@sbsuni.edu.vn

Scholar: https://scholar.google.com/citations?user=kHZvlTkAAAAJ&hl=en&oi=ao

Co-Founder: XAI - https://xai.foo/

Saigon Business School

In partnership with

University Canada West

Macquarie University Sydney · Australia

University of Northampton

Amity Global Institute

Saigon Business School

Learning materials

- Textbook
  - Evans, J. (2016) Business Analytics. 2nd edn. Pearson.
  - Runkler, T. (2016) Data Analytics: Models and Algorithms for Intelligent Data Analysis. 2nd edn. Vieweg+Teubner Verlag.

- Online reference materials
  - archive.ics.uci.edu/ml/
  - powerbi.microsoft.com
  - https://github.com/topics/data-analysis-python
  - https://media.pearsoncmg.com/ph/esm/esm_evans_eba3e_20/tools/eba3e_analytic_solver.html
  - https://data.imf.org/

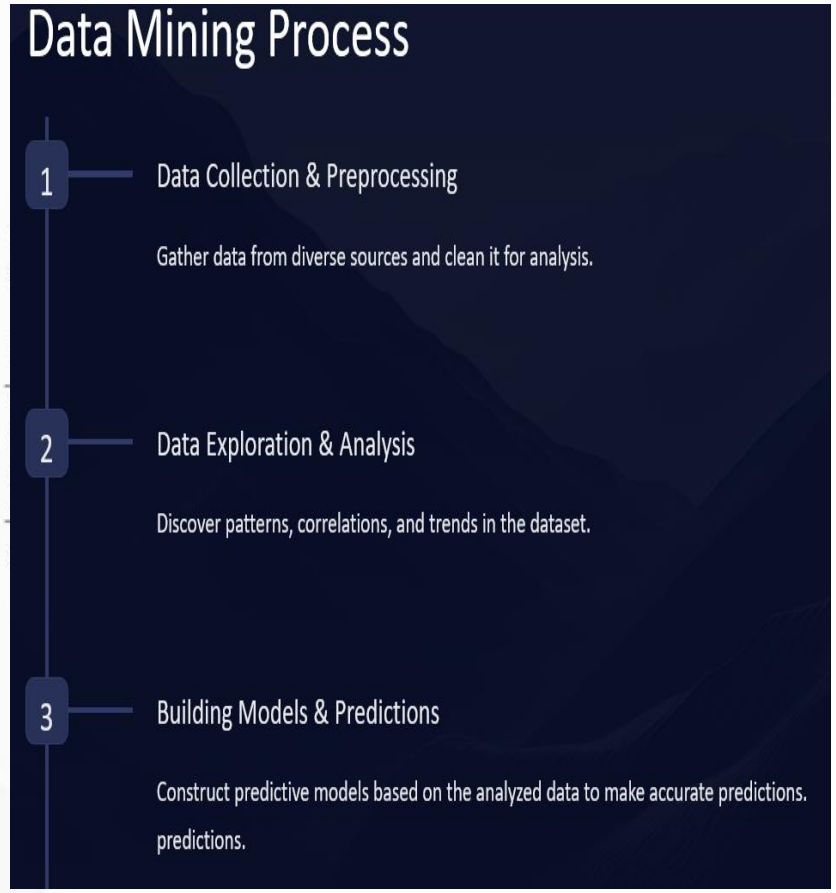
Saigon Business School

Agenda

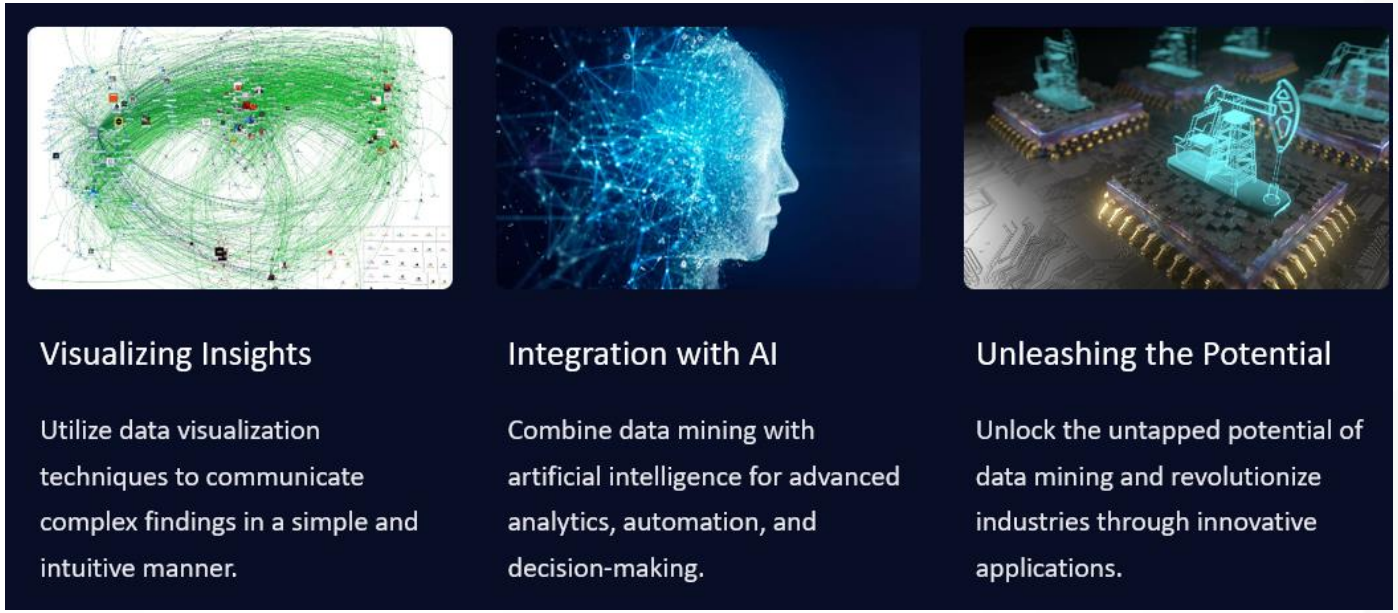
- Lesson 1: Understanding Data Analytics Terminologies.
- Lesson 2: Foundation of Business Analytics
- Lesson 3: Visualizing and Exploring data
- Lesson 4: Applying Descriptive Analytic Techniques
- Lesson 5: Data Modeling
- Lesson 6: Predictive Analytics
- Lesson 7: Regression, Classification and Clustering
- Lesson 8: Forecasting Techniques
- Lesson 9: Investigating Predictive Analytic Techniques
- Lesson 10: Introduction to Data Mining
- Lesson 11: Demonstrating Prescriptive Analytic Methods
- Lesson 12: Recap and advanced topics

Saigon Business School

Introduction to Data Mining



Explore the fascinating world of data mining. Discover its definition, benefits, and process. Uncover the challenges and applications that make it essential in various industries.



Saigon Business School

Introduction to Data Mining

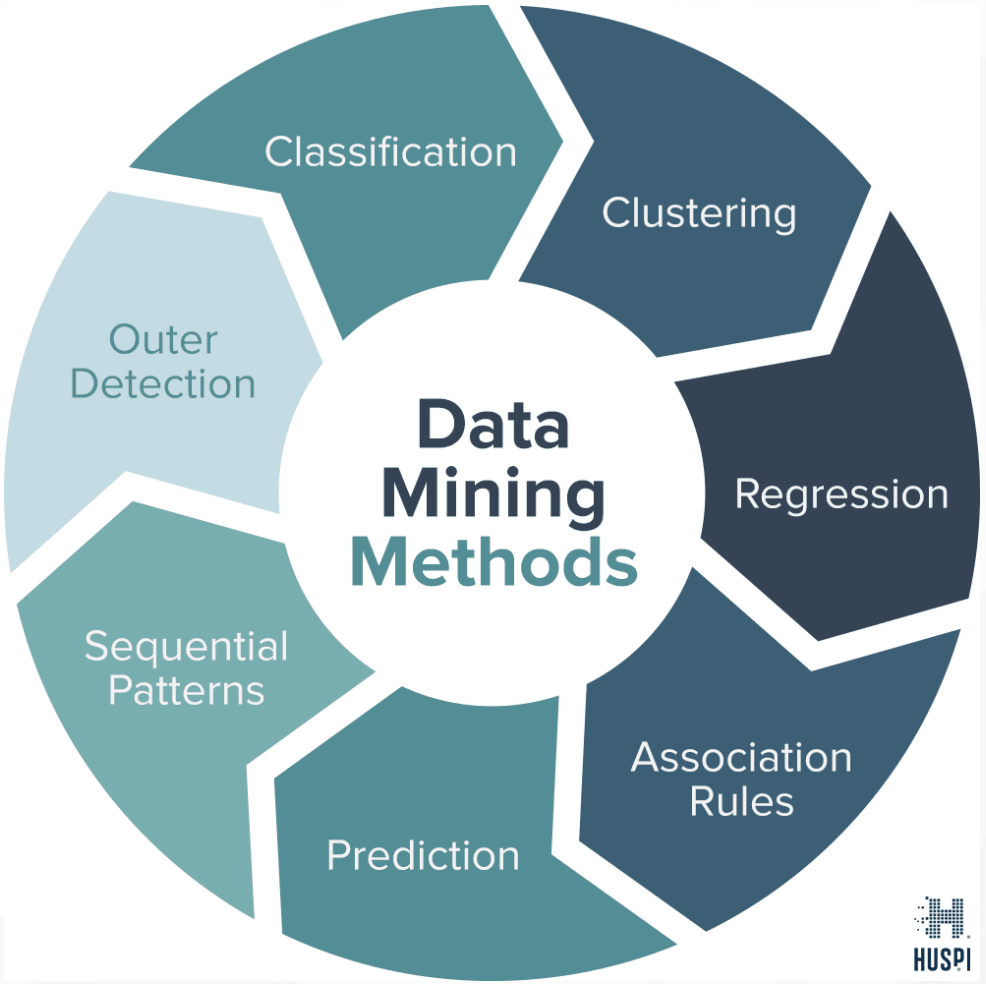
Data mining encounters challenges related to the vast volume of data generated in today's digital age, making it difficult to efficiently process and analyze large datasets. Another difficulty lies in ensuring the quality and reliability of the data, as incomplete or inaccurate information can lead to erroneous conclusions. Additionally, the ethical concerns surrounding privacy and the responsible use of sensitive data pose challenges, requiring careful consideration and adherence to ethical standards in the data mining process.

Interpreting results in data mining presents a challenge as complex algorithms may generate intricate patterns, requiring careful analysis to extract meaningful insights and avoid misinterpretation that could impact decision-making in a business context.

Saigon Business School

Introduction to Data Mining

1. **Classification.** This data mining method goes through the dataset and identifies information about this data that creates a foundation for further analysis. For example, “This is A,” “This is B”, and “This is C.”
2. **Clustering.** It is a bit similar to the classification, but besides simply identifying the information about the data (pardon the tautology), it looks for similar items and groups them together into clusters. For example, “This is A, it goes with the other As,” “This looks like B, so it goes with the Bs,” etc.
3. **Regression.** This data mining method helps to identify relationships between various data points. For example, the probability of A, given the amount of B and C.
4. **Outer Detection.** This method, also known as Outlier Mining or Outlier Analysis, is an important tool used for fraud, fault, anomaly detection, intrusions, etc. It analyses the data and focuses on the data points that don’t fit the norm.
5. **Association Rules.** This method is used for the discovery of hidden patterns and associations between two or more items.
6. **Sequential Patterns.** The method deals with the identification or discovery of trends and patterns in transaction data over a certain period. In a way, it is similar to Association Rules, but this one is more global.
7. **Prediction.** Last, but not least, the data mining method is a combination of other techniques. It uses the information gleaned from trend analysis, sequential patterns, clustering, etc. about the events in the past to make predictions about how things would work in the future.

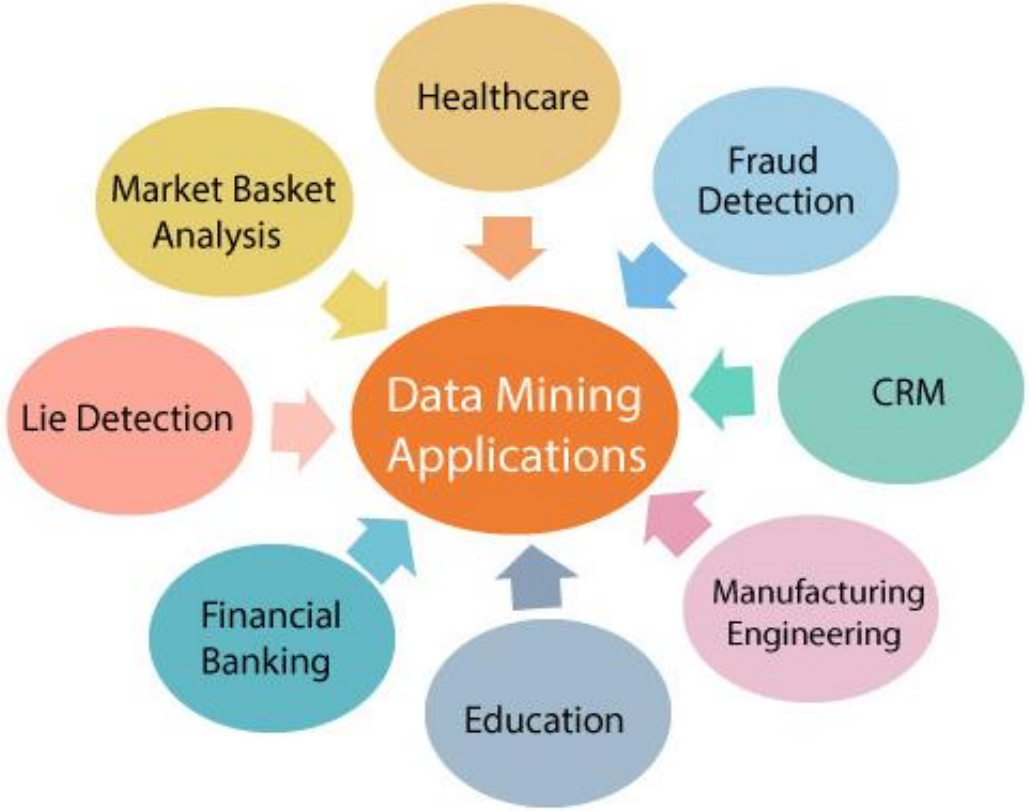


Saigon Business School

Introduction to Data Mining

Data mining Applications (Ordering based on your interest and sharing the reason)

1.
2.
3.
4.
5.
6.
7.
8.



Saigon Business School

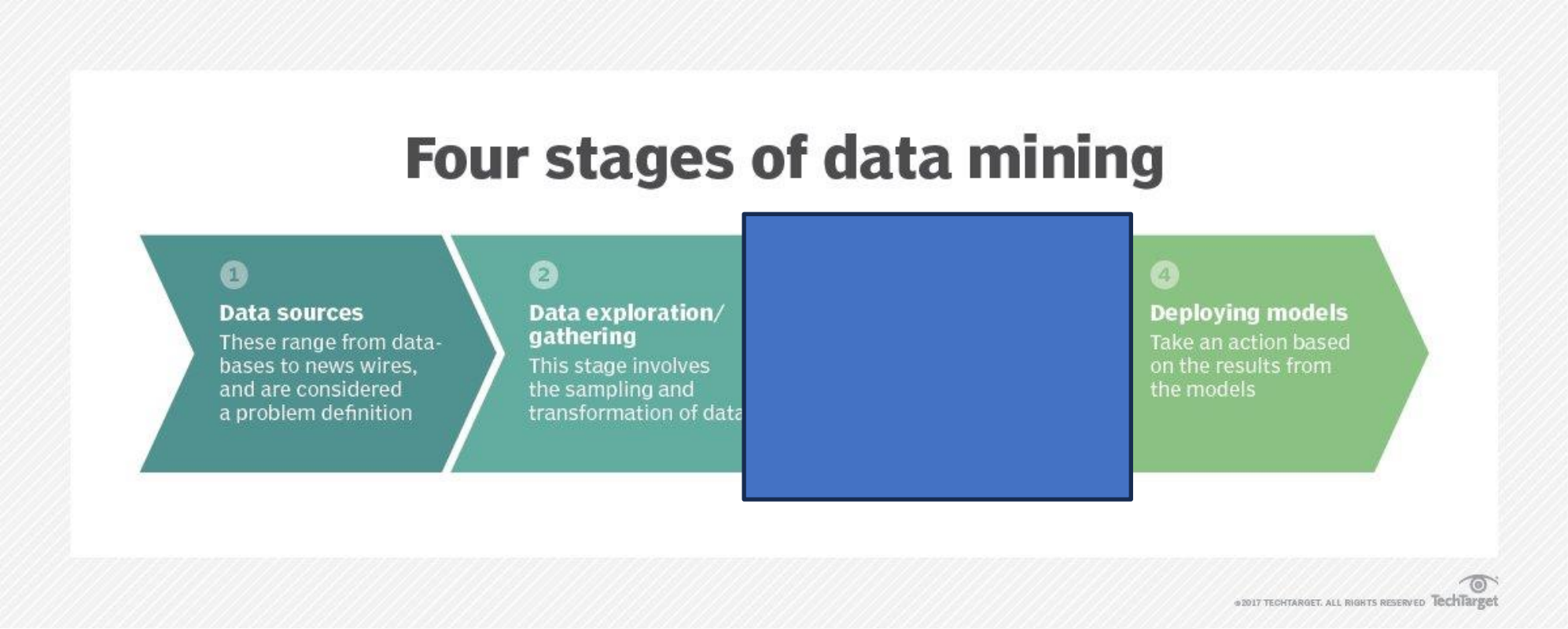
Introduction to Data Mining

1. What is text data mining?
2. What is game data mining?
3. What is financial data mining?
4. What is email data mining?
5. What is clinical data mining?
6. What is multimedia data mining?
7. What is sales data mining?

Saigon Business School

Investigating Predictive Analytic Techniques

Guess the stage 3 in the below diagram:



Saigon Business School

Learning Mission

30 mins

Reading Business Analytics textbook: Chapter 10, page 301 – 340.

Discussion and answer:

Classification matrix
Cluster analysis
Confidence of the (association) rule Data mining
Discriminant analysis Discriminant function
Divisive clustering methods Euclidean distance

Saigon Business School

Conclusion and Questions

Data mining is a rapidly growing field of business analytics that is focused on better understanding characteristics and patterns among variables in large databases using a variety of statistical and analytical tools. Many of the tools that we have studied in previous chapters, such as data visualization, data summarization.

Saigon Business School

Thank you