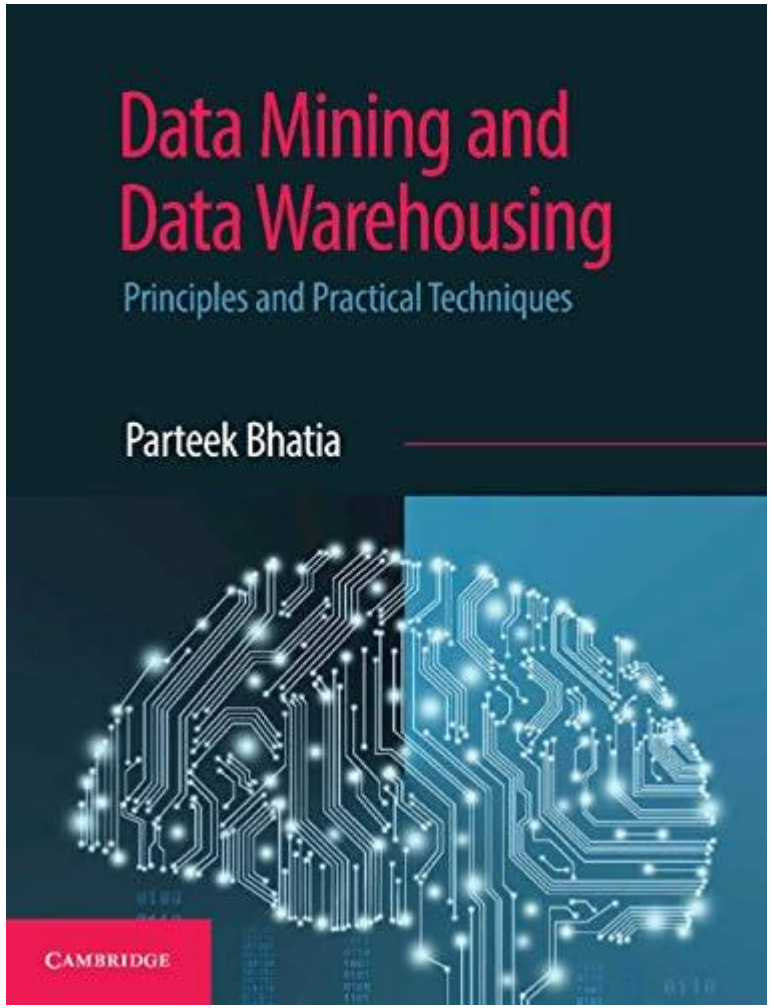
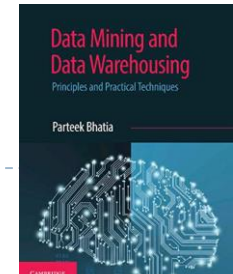
Chapter 2
Introduction to Data Mining



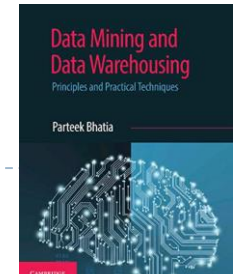
Chapter Objectives

1. **To learn about the concepts of data mining.**
2. **To understand the need for, and the applications of data mining**
3. **To differentiate between data mining and machine learning**
4. **To understand the process of data mining.**
5. **To understand the difference between data mining and machine learning.**



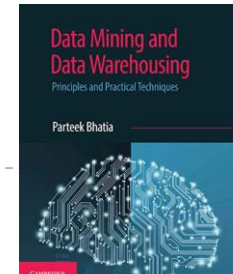
Defining Data Mining

- Data mining is a collection of techniques for efficient automated discovery of previously unknown, valid, novel, useful and understandable patterns in large databases. The patterns must be actionable so they may be used in an enterprise's decision making.'



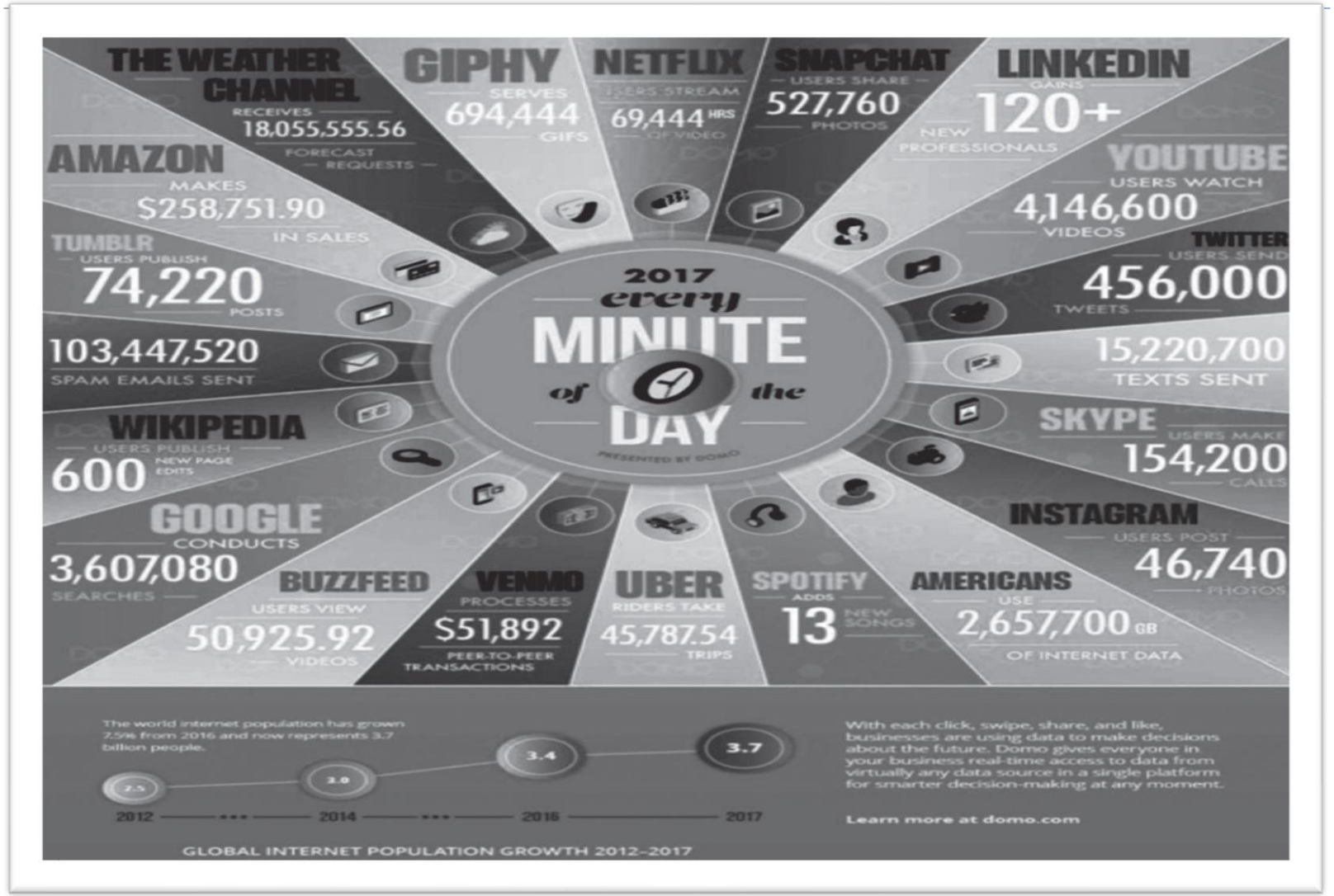
Introduction to Data Mining

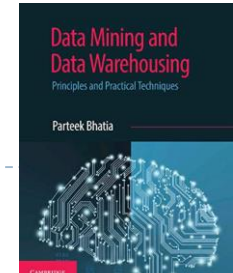
- From this definition, the important takeaways are:
- Data mining is a process of automated discovery of previously unknown patterns in large volumes of data.
- This large volume of data is usually the historical data of an organization known as the data warehouse.
- Data mining deals with large volumes of data, in Gigabytes or Terabytes of data and sometimes as much as Zetabytes of data (in case of big data).
- Data mining allows businesses to determine historical patterns to predict future behavior.



Need of Data Mining

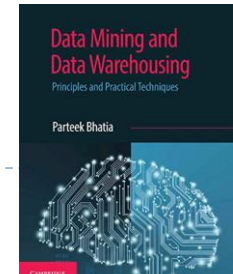
Per Minute Generation of Data over Internet





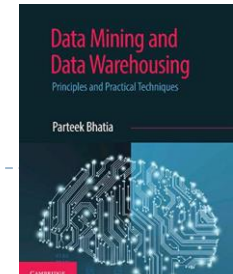
Need of Data Mining

- Growth in generation and storage of corporate data
- Need for sophisticated decision making
- Evolution of technology
- Availability of much cheaper storage, easier data collection and better database management for data analysis and understanding
- Decline in the costs of hard drives
- Growth in worldwide disk capacities

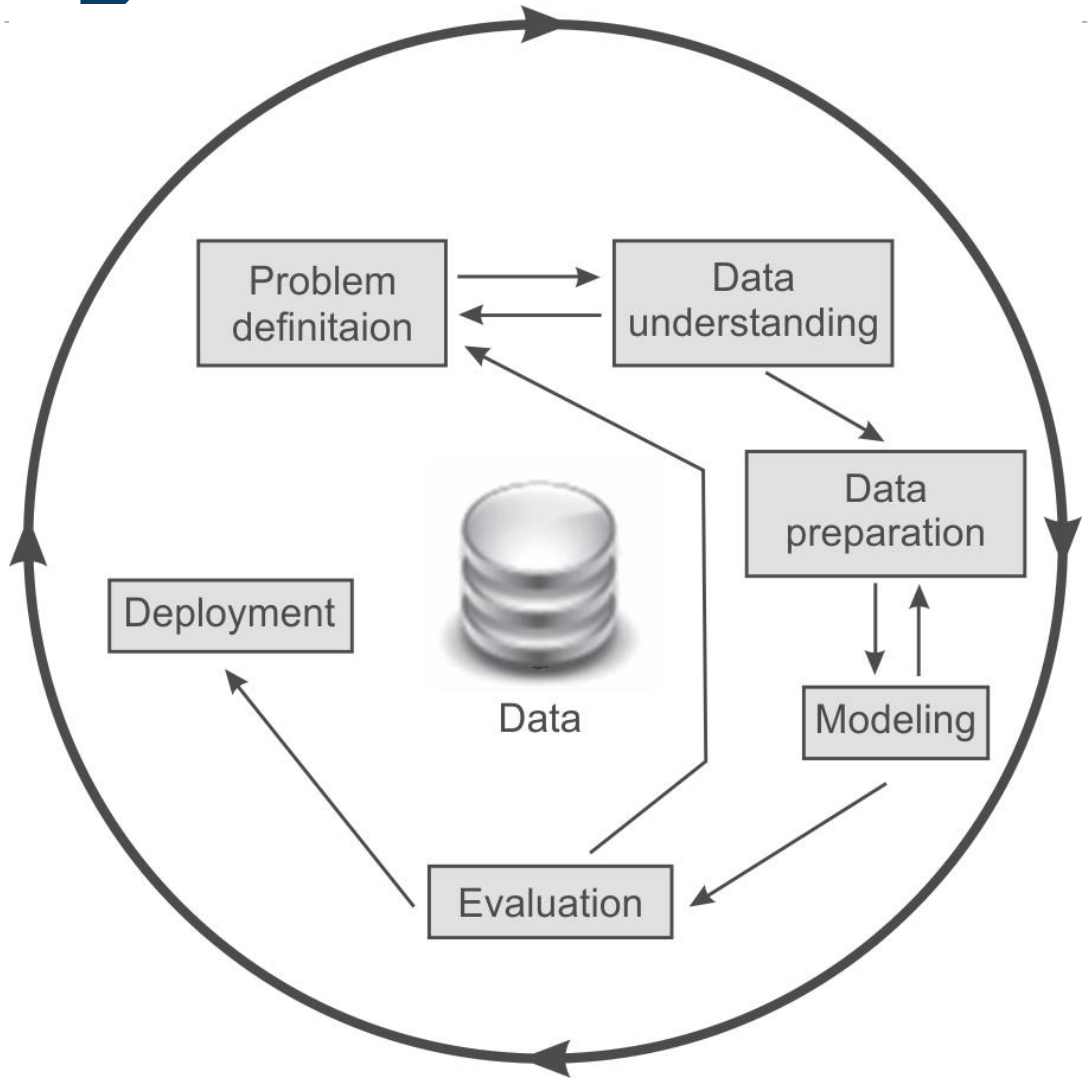


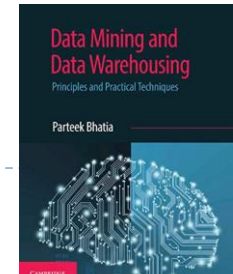
Applications of Data Mining

- **Loan/Credit card approvals**
- **Market segmentation**
- **Fraud detection**
- **Better marketing**
- **Trend analysis**
- **Market basket analysis**
- **Customer churn**
- **Website design**
- **Corporate analysis and risk management**

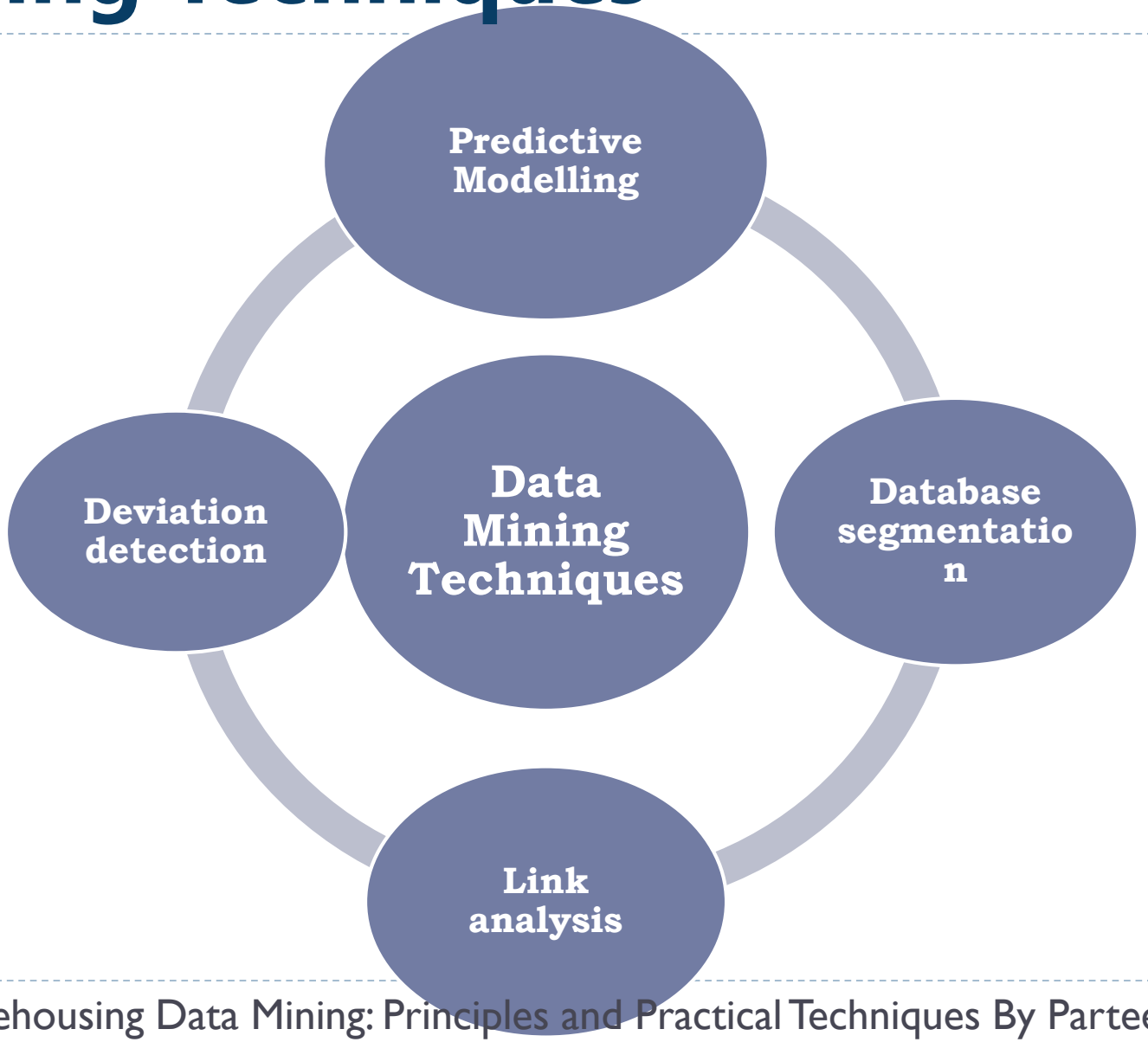


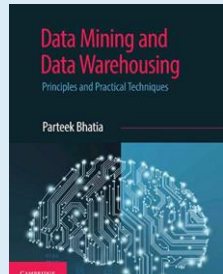
Data Mining Process





Data Mining Techniques

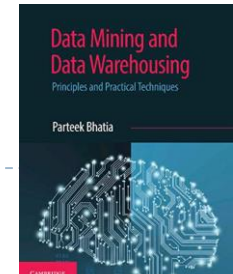




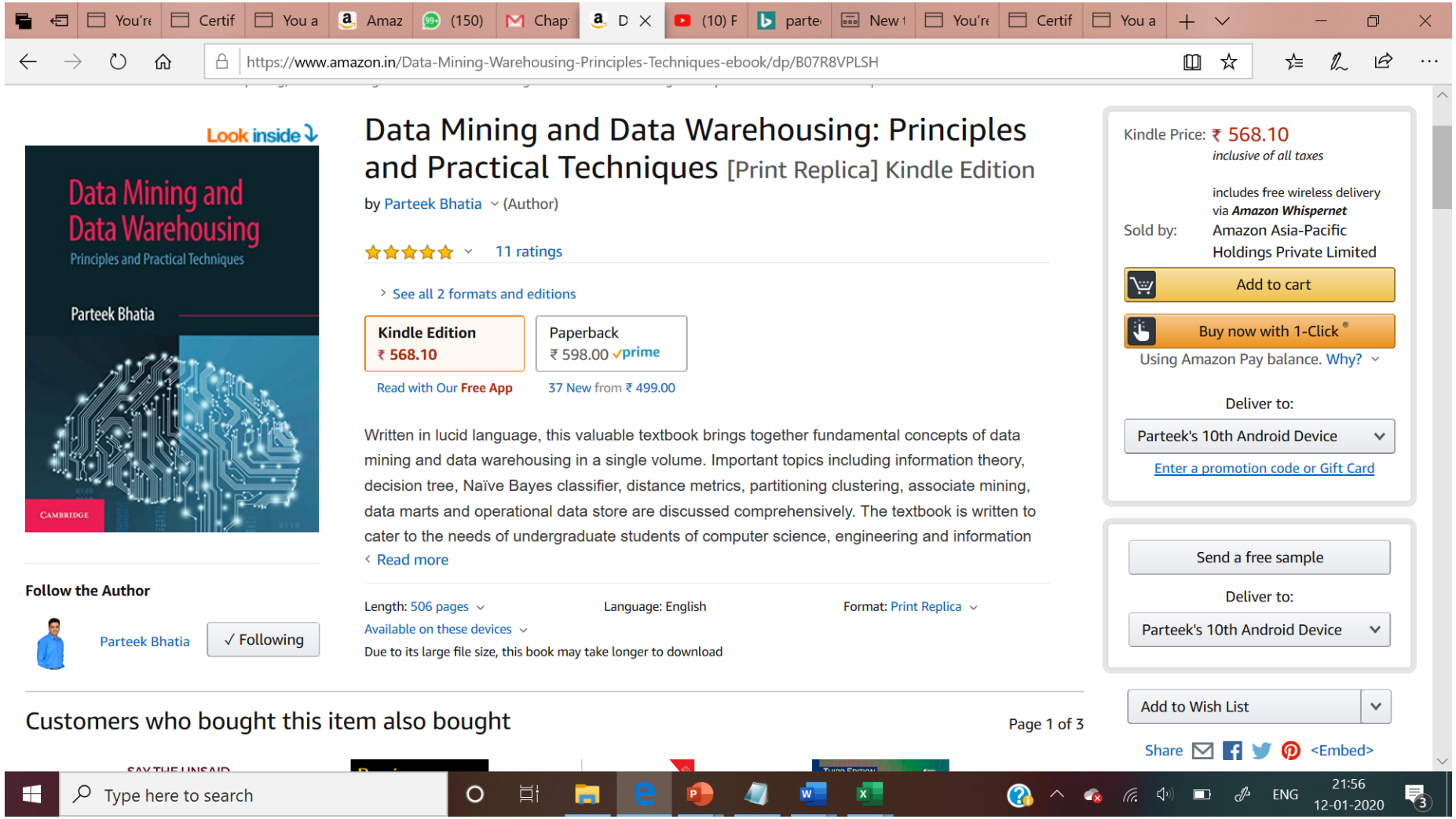
Difference between Data mining and ML

Table 2.1 Tabular comparison of data mining and machine learning

Basic for comparison	*Data mining*	*Machine learning*
Meaning	It involves extracting useful knowledge from a large amount of data.	It introduces new algorithm from data as well as past experience.
History	Introduced in 1930 it was initially called knowledge discovery in databases.	It was introduced in 1959.
Responsibility	Data mining is used to examine patterns in existing data. This can then be used to set rules.	Machine learning teaches the computer to learn and understand the given rules.
Nature	It involves human involvement and intervention.	It is automated, once designed it is self-implementing and no or very little human effort is required.

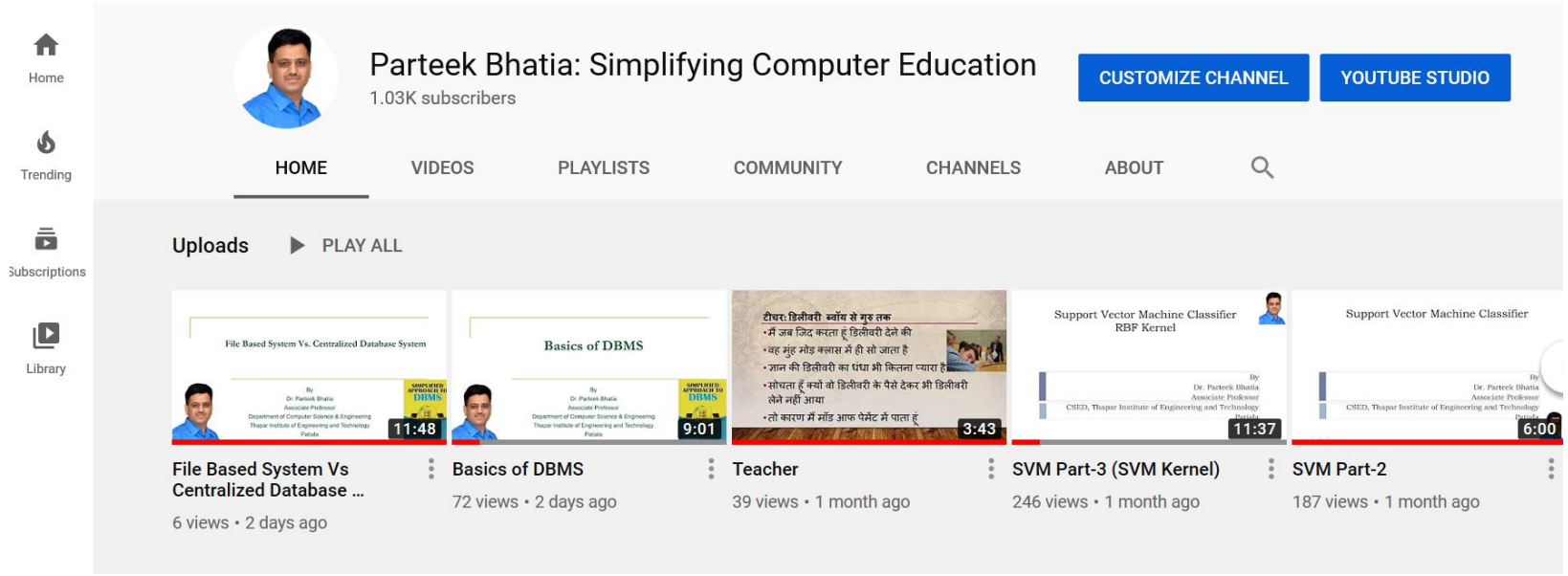


Reference


https://www.amazon.in/Data-Mining-Warehousing-Principles-Techniques-ebook/dp/B07R8VPLSH
Look inside
Data Mining and Data Warehousing
Principles and Practical Techniques
Parteek Bhatia
CAMBRIDGE
Data Mining and Data Warehousing: Principles and Practical Techniques [Print Replica] Kindle Edition
by Parteek Bhatia (Author)
11 ratings
See all 2 formats and editions
Kindle Edition ₹ 568.10
Paperback ₹ 598.00 prime
Read with Our Free App
37 New from ₹ 499.00
Written in lucid language, this valuable textbook brings together fundamental concepts of data mining and data warehousing in a single volume. Important topics including information theory, decision tree, Naïve Bayes classifier, distance metrics, partitioning clustering, associate mining, data marts and operational data store are discussed comprehensively. The textbook is written to cater to the needs of undergraduate students of computer science, engineering and information
Read more
Length: 506 pages
Language: English
Format: Print Replica
Available on these devices
Due to its large file size, this book may take longer to download
Follow the Author
Parteek Bhatia
✓ Following
Kindle Price: ₹ 568.10
inclusive of all taxes
includes free wireless delivery via Amazon Whispernet
Sold by: Amazon Asia-Pacific Holdings Private Limited
Add to cart
Buy now with 1-Click
Using Amazon Pay balance. Why?
Deliver to:
Parteek's 10th Android Device
Enter a promotion code or Gift Card
Send a free sample
Deliver to:
Parteek's 10th Android Device
Add to Wish List
Share
<Embed>
Customers who bought this item also bought
Page 1 of 3
Type here to search
ENG
21:56
12-01-2020

For more information

- Subscribe to YouTube Channel from the Author
  - To receive latest video tutorials on Data Mining, Machine Learning, DBMS, Big Data, NoSQL and many more.
- https://www.youtube.com/user/parteekbhatia



Free Online on SQL at Udemy

Udemy

ROYAL ACADEMY OF ENGINEERING Funded

THAPAR INSTITUTE OF ENGINEERING & TECHNOLOGY (Deemed to be University)

Simplified Approach to SQL

QUICK START
Ace Interviews and College Exams in 1 month

BI - LINGUAL
Explanation of each of the topic in English and Hindi

INTERACTIVE
Discussion forums and direct messaging to instructor

RELAXING
Study the course at your own pace

STUDY ANYWHERE
Study on your smartphone with udemy app

COURSE CONTENT

Week 1 : *Introduction to SQL and performing basic operations with SQL.*
Week 2 : *Creation of Tables with Integrity constraints.*
Week 3 : *Table Alterations and Joins.*
Week 4 : *Grouping of Data*

ABOUT THE INSTRUCTOR

Dr. Parteek Bhatia is Associate Professor in the Department of Computer Science and Engineering at Thapar Institute of Engineering and Technology, Patiala. He has more than 18 years of academic experience.He has authored several books in various areas of computer science. His book - Simplified approach to DBMS is one of the bestseller. Currently, he is working on plethora of Projects which are funded by Department of Science and Technology, CSIR and other funding agencies of India.

QR CODE

COURSE LINK: https://www.udemy.com/learn-sql-in-a-simplified-manner/

ABOUT THE AUTHOR

Dr. Bhatia is an Associate Professor in the Department of Computer Science and Engineering at Thapar Institute of Engineering and Technology, Patiala. He has more than twenty years of teaching experience and has published papers in journals. His current research includes natural language processing, machine learning and human-computer interface. He has taught courses including data mining and data warehousing, big data analysis and database management system at undergraduate and graduate levels. He also runs online courses on the Udemy portal.

www.parteekbhatia.com

For book related queries : parteek.bhatia@gmail.com