Data Mining and Business Intelligence

Data-driven strategy for business transformation

Dr. Jyotiranjan Hota

www.bpbonline.com

First Edition 2025

ISBN: 978-93-65892-239

LIMITS OF LIABILITY AND DISCLAIMER OF WARRANTY

The information contained in this book is true and correct to the best of author's and publisher's knowledge. The author has made every effort to ensure the accuracy of these publications, but the publisher cannot be held responsible for any loss or damage arising from any information in this book.

All trademarks referred to in the book are acknowledged as properties of their respective owners but BPB Publications cannot guarantee the accuracy of this information.

Dedicated to

My late parents

My wife ***Mandakini***

and

Daughters ***Mugdha*** *and* ***Snigdha***

About the Author

Dr. Jyotiranjan Hota is a distinguished academic with nearly 20 years of teaching, research and software consulting experience. He holds a B.E. in computer science and engineering from NIT Rourkela, a PGDBM from Xavier Institute of Management, Bhubaneswar, and a Ph.D. in Management Studies from Aligarh Muslim University, which was a joint program with AIMA, New Delhi. His areas of expertise includes business analytics, artificial intelligence, machine learning, data mining, text mining, visual analytics, and functional modules of SAP S4/HANA (SD, MM, PP, and FI-CO). He is also proficient in programming languages like R and Python and conversant with tools such as KNIME and Power BI. Dr. Hota's Ph.D. research delved into the adoption of Multivendor ATM technology in India. He analyzed challenges and opportunities from the perspectives of customers, suppliers, and bankers. Through this work, he developed and validated several qualitative and quantitative models addressing the drivers and barriers to technology adoption.

His broader interests focus on integrating information technology across various functional areas of management. An AIMA-accredited management teacher in the IT domain, Dr. Hota has an extensive record of publications in prominent top journals, including the International Journal of Bank Marketing, Asia Pacific Journal of Information Systems, International Journal of Management in Education, and The IUP Journal of Applied Economics, among others. He has also contributed to works by leading publishers such as Sage, Springer and Palgrave Macmillan. Dr. Hota has actively participated in international conferences in India and abroad, serving in key roles such as program committee member, advisory board member, technical committee member, track chair, and session chair. In recognition of his academic contributions, he was awarded with the ICBM-AMP Academic Excellence Award 2018 in the Best Professor in IT and Operations category in Hyderabad, India.

About the Reviewers

- **Tong Zhi** is a seasoned data scientist and data engineer with expertise in the private equity sector. His professional experience encompasses the development and deployment of advanced analytical and predictive models, such as Markov Chain Monte Carlo simulations, sophisticated time series forecasting, and classification algorithms.

  In addition to his role in private equity, he is also the founder of a startup venture, where he successfully architected and implemented a comprehensive business intelligence platform from scratch.

  Tong holds a master of science degree in business analytics and a bachelor of science degree in finance. Recognized for his expertise, he has frequently been invited to speak at prestigious international data science conferences and events. Currently, he serves at RoundShield Partners, a firm specializing in private equity and private credit, while concurrently managing HireHarbour, an innovative executive assistant outsourcing agency.

- **Anup Sahoo** is a Cloud Technical Lead at Insight India with over 14+ years of rich experience in the field of Quality Engineering, Test Automation, and DevOps. As a seasoned professional and a lifelong learner, Anup is passionate about solving real-world problems by merging deep technical expertise with cutting-edge technologies.

  He is a **Generative AI enthusiast and researcher**, exploring how large language models (LLMs) and AI-driven automation can transform the future of software testing and quality engineering. As a **Technical Author**, Anup has shared his insights through technical blogs, research-backed frameworks, and a growing portfolio of practical tools that aim to make QA smarter and more adaptive.

  When he's not immersed in designing intelligent test frameworks or experimenting with AI-infused pipelines, Anup channels his energy into mentoring aspiring professionals, creating impactful DevOps and automation content, and exploring nature through trekking. His curiosity-driven approach and commitment to innovation make him a driving force in both the tech and learning communities.

Acknowledgement

I would like to extend my heartfelt gratitude to everyone who has supported me on the challenging journey of writing this book as a sole author.

My wife, Mandakini, and my daughters, Mugdha and Snigdha, have been a constant source of love, motivation, and strength. Their encouragement and sentimental support have been a constant source of motivation.

At KSOM fraternity, I received consistent writing support from the early drafts to the final manuscript. The team's encouragement was instrumental in shaping this scholarly work into its final published version.

My sincere appreciation goes to BPB Publications for their invaluable guidance and expertise in bringing this book to life. I would also like to acknowledge the technical reviewers and editors who contributed their valuable feedback to this manuscript. Their insights and suggestions have greatly enhanced the quality of the book.

Last but not least, I want to express my gratitude to the readers who have shown interest in the book. Your support and encouragement have been deeply appreciated.

Thank you to everyone who has played a part in making this book a reality.

Preface

In today's era of digital transformation, information has become the cornerstone of progress and innovation. The ability to uncover actionable insights from data and effectively use business intelligence tools is a skill that spans industries, domains, and geographies. This book is carefully crafted to provide readers with the knowledge, techniques, case studies, and practical applications required to thrive in this transformative era.

The book is divided into eight thoughtfully designed chapters which offers a balanced blend of theoretical understanding and practical exercises. It provides a gradual progression that takes readers from foundational concepts to advanced analytics to prepare them to navigate the complexities of real-world data challenges.

The journey begins with Chapter 1, which introduces the basics of data mining and business intelligence. This chapter highlights the significance of these fields, explains core principles and emphasizes the importance of leveraging data effectively. Readers are also introduced to key differences between **online analytical processing** (**OLAP**) and **online transactional processing** (**OLTP**) systems. Chapter 2 focuses on pre-processing techniques, regression, and classification methods. It equips readers with essential tools to improve data quality and build reliable predictive models, laying a strong foundation for tackling real-world challenges.

Chapter 3 presents association rule mining, which is key to discovering patterns and relationships in data. This chapter explains metrics such as support, confidence and lift while introducing algorithms like A priori for identifying valuable insights. Chapter 4 discusses clustering techniques and their applications across various domains. It provides practical examples to illustrate foundational methods like k-means clustering, as well as advanced algorithms for grouping and analyzing data effectively.

The middle chapters explore the domain of business intelligence. Chapter 5 introduces its fundamentals, examining the driving forces, market dynamics, and tactical applications that define this transformative field. Chapter 6 puts these ideas into practice by exploring business intelligence architecture, concepts such as slicing and dicing and utilizing Power BI to model data and create impactful dashboards.

As the book progresses, Chapter 7 introduces innovative methodologies such as text mining, cognitive analytics, and big data analytics. These approaches equip readers with techniques to handle structured, semi-structured, and unstructured data effectively. The

final chapter, Chapter 8, discusses the ethical dimensions of data mining and business intelligence. It reflects on issues such as data governance, transparency, and responsible data practices, emphasizing the importance of trust and accountability in handling data in today's digital environment.

This book is thoughtfully written to be clear and useful for readers from various backgrounds, including students, professionals, and lifelong learners. With its clear explanations, practical examples, and comprehensive coverage, data mining and business intelligence provides a valuable resource for mastering the concepts and applications of these fields. It is hoped that this book inspires curiosity, fosters critical thinking, and empowers readers to unlock the potential of data to create meaningful and impactful solutions in their respective areas of study and work. Through practical examples, comprehensive explanations, sand a structured approach, this book aims to equip readers with a solid understanding of digital systems and technology. Whether you are a beginner or an experienced learner, I hope this book will serve as a valuable resource in your journey of exploring the foundations of data-driven insights.

Chapter 1: Introduction to Data Mining and Business Intelligence - This chapter briefly narrates the fundamental principles of data mining and business intelligence tools and techniques. Readers will gain an overall idea of how to use various techniques to extract meaningful information from large datasets. It also covers the scope, issues, and future trends.

Chapter 2: Regression and Classification Techniques with Applications - This chapter initially describes various pre-processing techniques with examples explained using R. Application of important supervised learning algorithms and applications of these algorithms on datasets from multiple functional areas are explained. Finally, ensemble algorithms are explained to improve the accuracy of prediction.

Chapter 3: Concept and Application of Association Rule Mining Algorithm - Association rule mining is an unsupervised learning algorithm that is used to decipher the best rules from frequent item sets to derive business insights. The most popular application of association rule mining is market basket analysis, which predicts the buying behaviors of customers in market place. Key metrics used to evaluate these algorithms are support, confidence and lift. These metrics evaluate the reliability, significance and strength of the generated rules to derive business decisions.

Chapter 4: Clustering - This chapter discusses clustering as an unsupervised learning algorithm which is used to group similar items based on their attributes. Here, different types of clustering algorithm like k-means clustering, hierarchical and density-based

clustering are discussed with examples using R. Few advanced clustering algorithms are discussed to deal with very large datasets.

Chapter 5: Introduction to Business Intelligence - This chapter lays a solid foundation of elementary concepts of business intelligence. It covers basic defection, markets, various key vendors, scopes, benefits, and future trends to prepare the readers with groundworks to further delve into business intelligence applications.

Chapter 6: Business Intelligence Architecture, Query and Reporting Practices - This chapter discusses the architecture of business intelligence, data reconciliation process, concept of data marts, OLAP cubes, and various data modelling, dashboards, and visualization techniques. Various applications are discussed using Power BI exercises for the readers.

Chapter 7: Advanced Data Mining and Business Intelligence Techniques - This chapter narrates advanced data mining techniques to deal with large volume of data of various functional domains to extract meaningful insights. This chapter covers topics like text mining, big data, edge analytics, cognitive analytics and real-time analytics to integrate with data mining and business intelligence tools for making informed decisions in organizations. Organizations gain strategic intent by uncovering diamonds from large datasets.

Chapter 8: Data Mining and Business Intelligence Ethical Framework - The lifeblood of data mining and business intelligence is data. Primary responsibility and challenges of data governance revolve around data collection, access, preservation, security, privacy, and democratization. Fairness, transparency, and trust can be achieved through a proper ethical framework based on good data governance practices. This chapter discusses inhibiting and facilitating forces of ethical directions of data mining and business intelligence to ensure trust, transparency, lowering costs incurred in organizations, and social implications for society as a whole.

Code Bundle and Coloured Images

Please follow the link to download the
Code Bundle and the ***Coloured Images*** of the book:

https://rebrand.ly/30gtva2

The code bundle for the book is also hosted on GitHub at
https://github.com/bpbpublications/Data-Mining-and-Business-Intelligence.
In case there's an update to the code, it will be updated on the existing GitHub repository.

We have code bundles from our rich catalogue of books and videos available at **https://github.com/bpbpublications**. Check them out!

Errata

We take immense pride in our work at BPB Publications and follow best practices to ensure the accuracy of our content to provide with an indulging reading experience to our subscribers. Our readers are our mirrors, and we use their inputs to reflect and improve upon human errors, if any, that may have occurred during the publishing processes involved. To let us maintain the quality and help us reach out to any readers who might be having difficulties due to any unforeseen errors, please write to us at :

errata@bpbonline.com

Your support, suggestions and feedbacks are highly appreciated by the BPB Publications' Family.

Piracy

If you come across any illegal copies of our works in any form on the internet, we would be grateful if you would provide us with the location address or website name. Please contact us at **business@bpbonline.com** with a link to the material.

If you are interested in becoming an author

If there is a topic that you have expertise in, and you are interested in either writing or contributing to a book, please visit **www.bpbonline.com**. We have worked with thousands of developers and tech professionals, just like you, to help them share their insights with the global tech community. You can make a general application, apply for a specific hot topic that we are recruiting an author for, or submit your own idea.

Reviews

Please leave a review. Once you have read and used this book, why not leave a review on the site that you purchased it from? Potential readers can then see and use your unbiased opinion to make purchase decisions. We at BPB can understand what you think about our products, and our authors can see your feedback on their book. Thank you!

For more information about BPB, please visit **www.bpbonline.com**.

Join our book's Discord space

Join the book's Discord Workspace for Latest updates, Offers, Tech happenings around the world, New Release and Sessions with the Authors:

https://discord.bpbonline.com

Table of Contents

1. Introduction to Data Mining and Business Intelligence 1

Introduction 1

Structure 1

Objectives 2

Reasons for studying data mining 2

Evolution 3

Introduction to OLTP, OLAP and data mining 4

Associated fields of data mining 5

Data mining techniques 8

Business intelligence techniques 11

Processes of data mining 14

Leading vendors of data mining 16

Introduction to business intelligence 17

Motivations 18

Evolution of business intelligence 18

Leading vendors of business intelligence 19

Privacy issues 21

Future trends 22

Conclusion 23

Multiple choice questions 23

Answers 24

Practice exercises 25

2. Regression and Classification Techniques with Applications 27

Introduction 27

Structure 27

Objectives 28

Data cleaning, transformation and normalization 28

Bias and variance 32

Regression ... 35
Metrics and coefficients of regression ... *36*
Assumption of linear regression ... *37*
Case study ... *39*
Holdout and cross-validation in classification tasks ... *42*
Performance metrices of classification techniques ... *42*
Decision tree ... 44
Terminologies of decision tree structure ... *44*
Tree construction and metrics of decision tree ... *45*
Gini value ... *46*
Logistics regression ... 51
Other classification techniques ... 55
Conclusion ... 60
Multiple choice questions ... 60
Answers ... *62*
Practice exercises ... 62

3. Concept and Application of Association Rule Mining Algorithm ... 65
Introduction ... 65
Structure ... 66
Objectives ... 66
Importance and evolution of association rule mining ... 66
Fundamental concepts ... 67
Popular algorithms and applications in business domains ... 69
Apriori algorithm and business applications ... *69*
FP-Growth and Eclat algorithm in business applications ... *75*
Data visualization and modeling using R ... 75
Case study ... 82
Aims of the study ... *82*
Dataset description ... *82*
Methodology of the case study ... *83*
Data exploration ... *83*
Algorithm implementation parameters ... *83*

Performance metrics ... *83*
Experimental setup ... *84*
Results and analysis ... *84*
Summary ... *87*
Key learnings ... *87*
Practical application ... *87*
Challenges and future prospects ... 87
Conclusion ... 88
Multiple choice questions ... 88
Answers ... *90*
Practice exercises ... 90
References ... 92

4. Clustering ... 93
Introduction ... 93
Structure ... 93
Objectives ... 94
Distance metrics ... 94
Euclidean distance measure ... *94*
Manhattan distance measure ... *95*
Cosine similarity measure ... *95*
Jaccard coefficient measure ... *96*
K-means clustering, applications and challenges ... 96
K-medoids clustering for data partitioning, applications and challenges ... *98*
Hierarchical clustering, applications and challenges ... *100*
Advanced clustering algorithms ... 102
Data visualization on clustering using R ... 102
K-means clustering application ... *103*
K-means clustering using R ... *105*
K-medoid application ... *106*
Hierarchical clustering application ... *109*
Conclusion ... 111
Multiple choice questions ... 112

Answers *113*
Practice exercises 114

5. Introduction to Business Intelligence 117
Introduction 117
Structure 117
Objectives 118
Business intelligence 118
From data to decisions in a transforming market 118
Revolutionizing business intelligence for strategic advantage *118*
Empowering innovation through business intelligence *119*
Global perspectives *120*
Key drivers in the global market *120*
Indian landscape *121*
Charting the landscape of business intelligence 122
Embracing opportunities and confronting challenges *122*
Scope and benefits *122*
Challenges in implementing business intelligence *123*
Understanding descriptive, diagnostic, predictive and prescriptive analytics 124
Descriptive analytics *124*
Uncovering business performance with diagnostic analytics *125*
Harnessing predictive analytics for future business insights *126*
Empowering decision making with prescriptive analytics *127*
Case let *130*
Business intelligence applications across industries 131
Business intelligence applications in the retail sector *132*
Analysis of business intelligence's impact on the airline industry *132*
Business intelligence in agriculture *133*
Transforming pharmaceutical practices through business intelligence *134*
Transforming media and entertainment practices through BI *134*
Transforming higher education with business intelligence 135
Case study 136
Conclusion 139

Multiple choice questions 140
Answers *142*
Practice exercises 142

6. Business Intelligence Architecture, Query and Reporting Practices 143
Introduction 143
Structure 143
Objectives 144
Introduction to business intelligence architecture 144
ETL process and data warehouse design 145
Data warehouse design *146*
OLAP and cube designs 147
Key components of OLAP *147*
Levels in OLAP hierarchies *148*
OLAP operations *149*
Application of OLAP operations *150*
Data modelling applications using Power BI 153
DAX operators *155*
DAX calculation types *156*
Sales table *157*
Calculating total revenue in Power BI example *157*
Dimensional modelling using Power BI *158*
Importing data into Power BI *159*
Dashboard design and data visualization using Power BI 162
Importing data into Power BI *163*
Designing an interactive dashboard *167*
Case study 168
Conclusion 173
Multiple choice questions 173
Answers *175*
Practice exercises 175
References 177

7. Advanced Data Mining and Business Intelligence Techniques 179
Introduction 179
Structure 179
Objectives 180
Text mining as advanced data mining technique 180
Preprocessing text data 180
Bag of words modelling 181
Importance of the three models 186
Visualizing with word clouds 186
Sentiment analysis 189
Topic modelling 189
Text classification and clustering 189
Question-answering systems and chatbots 189
Challenges and considerations 190
Scalability and efficiency 190
Future trends 190
Big data analytics in business 191
Characteristics of big data 191
Challenges to big data analytics 194
Emerging trends in big data analytics 194
Future ahead 195
Edge analytics 196
Importance and advantages of edge analytics 196
Challenges of edge analytics 197
Applications 197
Future of edge analytics 198
Cognitive analytics 198
Importance for data mining and BI professionals 199
Advantages of cognitive analytics 199
Challenges of cognitive analytics 199
Applications 200
Real-time analytics and data streaming 202

Concept and importance 202
Advantages of real-time analytics and data streaming 202
Challenges of real-time analytics and data streaming 203
Applications 203
Future of real-time analytics and data streaming 205
New era of creativity through generative AI 207
Core concepts of generative AI 207
Ethical considerations and challenges of generative AI 208
Agentic AI as autonomous intelligence's future 209
Applications of agentic AI 210
Future of continuous innovation and ethical integration 210
Key insights and lessons learned 212
Summary 213
Conclusion 213
Multiple choice questions 213
Answers 215
Practice exercises 216

8. Data Mining and Business Intelligence Ethical Framework 217
Introduction 217
Structure 217
Objectives 218
Ethical guidelines and frameworks 218
Data protection laws and their ethical imperatives 218
General Data Protection Regulation 218
California Consumer Privacy Act 219
Digital Personal Data Protection Act of India 219
Brazil's General Data Protection Law 220
Adverse impact of algorithmic bias in industry 220
Historical data bias 220
Facial recognition bias in Indian law enforcement 221
Sampling bias 221
Mitigation of algorithmic bias 222

Merging ethical insights, data mining and BI 224
Transforming data into decisive consumer insights *224*
Strengthening finance with risk management and fraud prevention *224*
Maximizing operational excellence for efficiency and productivity *225*
Enhancing hiring and workplace satisfaction *225*
Strategies for decision making and competitive advantage *226*
Turning ethical responsibilities into action 229
Establishing a strong ethical foundation *229*
Ensuring data privacy and security *230*
Promoting transparency and accountability *231*
Mitigating bias and ensuring fairness *231*
Fostering an ethical culture *232*
Outcome *235*
Expanding on the steps for ethical responsibility *235*
Conclusion 237
Multiple choice questions 237
Answers *239*
Practice exercises 239
Role-playing scenario questions 240

Index 241-247

Chapter 1
Introduction to Data Mining and Business Intelligence

Introduction

In this chapter, we will discuss the fundamentals, techniques, applications, and challenges of data mining and business intelligence. We will initially focus on its importance and motivation to grasp the subject foundationally. The difference between **Online Analytical Processing** (**OLAP**) and **Online Transactional Processing systems** (**OLTP**) is explained through real-life examples. Emerging trends of data mining and business intelligence, top vendors, tools, and markets for data mining and business intelligence, are explained to build a foundation for subsequent chapters.

Structure

This chapter covers the following topics:

- Reasons for studying data mining
- Evolution
- Introduction to OLTP, OLAP and data mining
- Associated fields of data mining
- Data mining techniques
- Business intelligence techniques

- Leading vendors of data mining
- Introduction to business intelligence
- Motivations
- Leading vendors of business intelligence
- Privacy issues
- Future trends

Objectives

After going through this chapter, you will understand the fundamental principles of data mining and business intelligence tools, techniques, markets, and privacy concerns. Readers will also gain an overall idea of how to deal with various techniques to extract meaningful information from large datasets.

Reasons for studying data mining

Although organizations today are sinking in data, they are starving for knowledge. Massive data is generated from social media, e-commerce transactions, online forums, **Internet of Things** (**IoT**) devices, and several streaming platforms. So, for competitive advantages, firms need to derive hidden patterns and insights from data. There is a curiosity to know the following:

- How to make use of data assets?
- How can organizations make the best use of generated data?
- How to decipher the gap from stored data to knowledge?
- What are the limitations of database queries in fetching the following results:
    - Generate a list of customers likely to purchase products or services.
    - Who are the prospects likely to respond to our advertising campaign?

In the correct context, data is abundant, and multiple data mining tools are readily available. It is not an uphill task to gather data in a warehouse. Similarly, computing power is cheap, and there is tremendous pressure on companies to build their strategic intent. Hence, it becomes indispensable to apply these tools and techniques to derive diamonds out of data for survival and excellence in the current business world.

"Data is like a faint light when you are lost in a dark room. Follow it, try to make sense of it, and you might actually know where you are and what is around you."

- David Sides

Evolution

Data mining is a blend of statistics, mathematics, computer science, machine learning, and big data. Though there is no fixed timeline, its inception was in the late eighteenth century with the development of Bayes' theorem based on conditional probability. Currently, the Bayes theorem is applied extensively in data mining. Regression as a basic prediction technique in data mining was developed during the early nineteenth century. These rich applications strengthened the field of statistics, followed by the application of computing by the *Alan Turing* model and neural networks by the mid-twentieth century. *Charlee Babbage* surprised the entire human civilization with the power of computing by a machine. However, *Turing* took a step forward with the Turing machine model, which stated that a machine can also think like a human. Similarly, the introduction of neural networks laid the foundation for data mining by developing a model in 1943.

A drastic development in data mining happened after the mid-twentieth century due to databases, genetic algorithms, and further evolutionary computation in the era of computing. Real business applications gained momentum during the last decade of the twentieth century due to data warehousing and data mining as a prediction technique with the introduction of data science. Due to the rapid development of social media, big data, cloud usage, and IoT applications, there was an increased usage of data mining in all functional areas of business during the twenty-first century. The stages of evolution of data mining are specified in *Figure 1.1*:

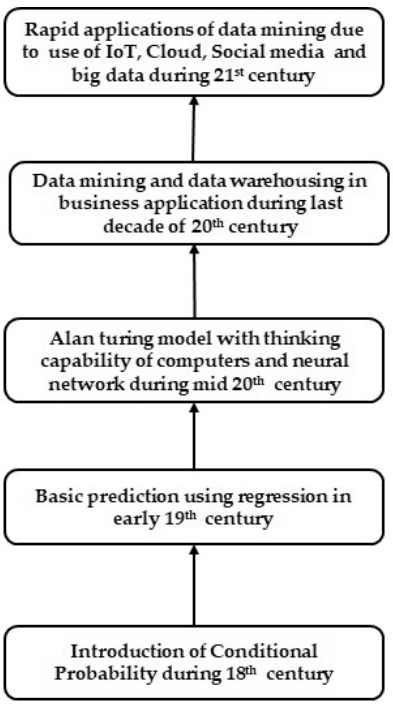

***Figure 1.1**: Evolution of data mining*

Introduction to OLTP, OLAP and data mining

Daily business transactions are captured in databases as online transaction processing in real-time. These OLTP transactions are captured online or offline. These transactions are common in all business activities like manufacturing, retail, sales, and finance. OLTP transactions are user-friendly and can be operated and managed by end users. Response time to these queries is quite fast. For example, withdrawal of money from ATMs, purchasing products online, and booking tickets are a few examples of OLTP systems. The volume of OLTP transactions is quite high, but transactions are small. Multiple users can access the OLTP system. So, a concurrency control mechanism is required to avoid data access by many users simultaneously. These transactions are basically database transactions. As past data is updated with new data, historical data cannot be retained for decisional analysis. However, OLAP systems are built offline, based on aggregate data from OLTP systems. These are offline systems that require huge resources, and the response time is high compared to OLTP systems.

The extracted data from OLTP systems is cleaned, transformed to a single format, and loaded into the data warehouse. Then OLAP queries are applied to these summary data. The queries touch a very large amount of data. Updates to OLAP systems are usually periodic and infrequent. As queries are complex, an individual query requires lots of resources. OLAP technology is based on a multidimensional data model. Ad-hoc queries can be applied linking functional areas like sales, marketing, operations, finance, and accounting for decisional analysis. OLAP uses cubes to store multiple categories of data. Usually, cubes have three dimensions. Say, for example, we take an example of a retail multidimensional model. The categories of dimensions can be product, customer and time dimensions. Actual transactions are stored as facts centrally. Further, analysis can be done using OLAP tools through multiple operations smoothly to derive insights from data. OLTP transactions are real-time transactions and OLAP transactions are offline transactions. So, OLTP and OLAP transactions are analyzed separately. There are three reasons for which we cannot mix both OLTP and OLAP queries. Firstly, performance requirements of OLTP and OLAP queries are not same. Since, OLTP queries require very less response time, and data should be always consistent. OLAP queries can saturate CPU time and consume lots of resources. Secondly, data modeling is different for both OLTP and OLAP. OLTP contains many tables with complex entity relationship models. However, OLAP contains very few dimension tables and a single fact table, which comprises transactions or facts. So fewer joins among tables are required as the architecture is different from OLTP. Thirdly, OLTP targets only a single source, whereas OLAP integrates data from multiple sources. So, OLTP and OLAP queries cannot be mixed. However, these queries are quite useful based on the requirements of different kinds.

Data mining techniques are applied to datasets to find hidden patterns and forecast future outcomes. The data mining technique uncovers hidden insights from data to

facilitate decision-making. There are multiple applications of data mining in business. Telecommunication companies that use data mining techniques to predict when their customer will leave their company to join other competitors. In retail, the data mining technique helps decipher customers to whom we can go for product bundling promotional offers based on past transaction history and preferences. Data mining techniques also predict the life insurance policies the customers are likely to purchase in the future.

You may refer to *Figure 1.2* to go through the difference between OLTP, OLAP, and data mining:

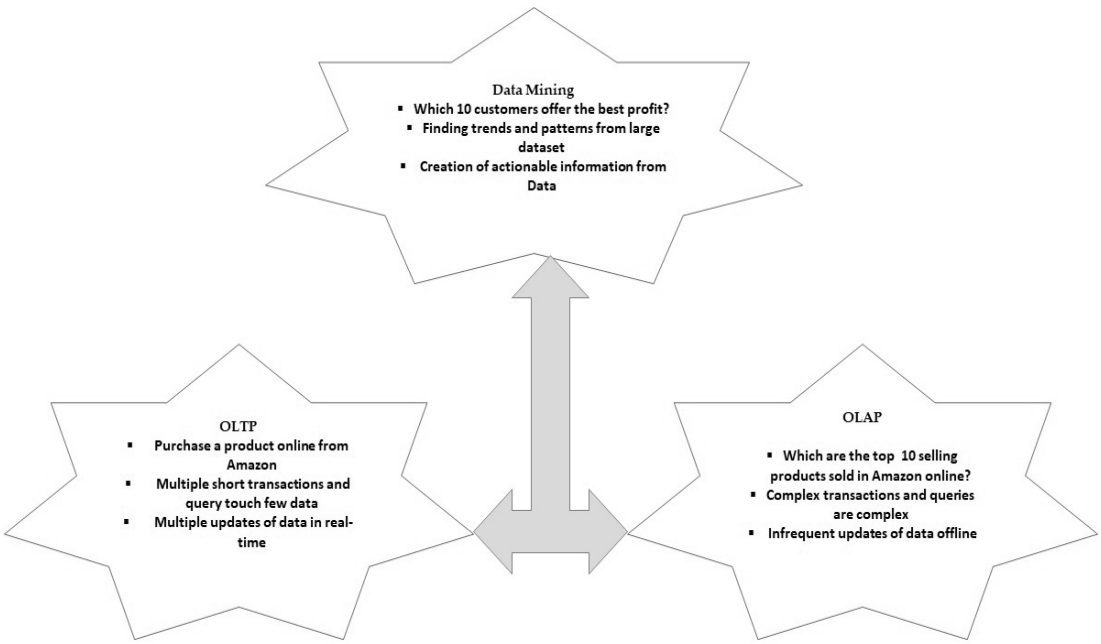

***Figure 1.2**: Difference between OLTP, OLAP and data mining*

From the purchase history, data mining techniques can identify the pattern and association among purchased products. Subsequently, marketers can plan for techniques like cross-selling, upselling, and product bundling.

Associated fields of data mining

Some associated fields of data mining are:

- **Statistics**: As a branch of applied mathematics, statistics deals with data collection, organization, interpretation, and presentation of data. As a related field of data mining, statistics is used to understand, described, and check the distribution pattern of data and also predicts using many statistical tools, like regression and clustering. Data visualization is an important contribution to statistics. Scatter plots, heat maps, histograms, and boxplots help to understand data pattern. To