DEMOCRATIC AND POPULAR REPUBLIC OF ALGERIA MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH UNIVERSITY OF MOHAMED BOUDIAF - M'SILA

FACULTY OF MATHEMATICS AND INFORMATICS

DEPARTMENT OF COMPUTER SCIENCE

A Dissertation in Fulfillment

For the Requirements of the Degree of Master in Computer Science

DOMAINE: Mathematics and Informatics

FILIERE: Informatics

OPTION : Information Systems and Software Engineering

By : SLIMANE KADI CHEYMA

Entitled

Analyzing Data Using Apriori Algorithm

Presented publicly to the jury:

Dr. HAMMANI Said	**University of M'sila**	**President**
Dr. LOUNNAS Bilal	**University of M'sila**	**Supervised**
Dr. BOUDIA Malika	**University of M'sila**	**Examiner**

Academic Year: 2020/2021

I dedicate this work:
To My Father And Mother for their continued support and I wish them good health and long life
SLIMANE KADI CHEYMA .

Acknowledgements

In The Name of **ALLAH**, The Most Beneficent, The Most Merciful.

All praise belongs to **ALLAH** alone, and blessings and peace be upon the final Prophet.

At first I would Like To thank my **Father and mother** Who were by my side and Still always by my side For helping me solve the Obstacles of this Dissertation and all my academic path.

Thanks to my **Sisters and Brother** Romaissa , safa , and Ahmed and **My aunt** Karima For Their Undless Love and for theirs encouragements and supports during my academic path.

Many thanks to my supervisors **Dr. LOUNANAS Bilal** for His supports, and advises From First Year Of University.

Special thanks To my TeamMates at MFormatik Company also My friends and All One Who Helped Me and Support Me.

Contents

List of Figures

List of Tables

1

GENERAL INTRODUCTION

1. General Introduction
    - 1.1. Overview
    - 1.2. Problem Statement
    - 1.3. Objectives
        - 1.3.1. Main Objectives
        - 1.3.2. Specific Objectives
2. Dissertation Organization

1.1 Overview

“We are drowning in information and starving for knowledge.” Rutherford D. Roger

How the world has changed in the This Past Years! large quantities of data are collected and mined in nearly every area of science, entertainment, business, and industry. Online movie and book stores study customer ratings to recommend or sell them new movies or books. Social networks mine information about members and their friends to try to enhance their online experience , There is a crucial need to sort through this mass of information, and pare it down to its bare essentials. For this process to be successful, we need Business intelligence.

In a 1958 article, IBM researcher Hans Peter Luhn used the term business intelligence. He defined intelligence as: "the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal." And In 1989 Howard Dresner (later a Gartner Group analyst) proposed "business intelligence" as an umbrella term to describe "concepts and methods to improve business decision making by using fact-based support systems." It was not until the late 1990s that this usage was widespread.

Business intelligence as its Understood today ;is the art of turning data (Big Data) into actions. This is accomplished through the creation of data products, which provide actionable information without exposing decision makers to the underlying data (e.g., buy/sell strategies for financial instruments, a set of actions to improve product yield, or steps to improve product marketing). Performing business intelligence requires the extraction of timely, actionable information from diverse data warehouses to drive data products. Examples of data products include answers to questions such as: “Which of my products should I advertise more heavily to increase profit? How can I improve my compliance program, while reducing costs? What manufacturing process change will allow me to build a better product?” The key to answering these questions is: understand the data you have and what the data inductively tells you.

As we move into the data economy, Business intelligence is the competitive advantage for organizations interested in winning, and wining came through improving decision-making witch leads us to the stark reality that whether or not information is available, decisions must be made. The process of extracting such valuable knowledge from a big Data is known as Data Mining.

Data mining is the process of automatically discovering useful information in large data Warehouse or Data Mart. Data mining techniques are deployed to scour large

database in order to find novel and useful patterns that might otherwise remain unknown. They also provide capabilities to predict the outcome of a future observation ,Data mining is a method of extracting what is useable within a database and separating it out from what is unusable. Such methods are necessary because, as human being, we lack the capacity to sort and organize such large volumes of data. The use of data mining techniques provides, in all areas, the ability to explore, focus on the most important information in databases, and data mining techniques also focus on building future predictions and exploring behavior and trends. This allows making the right decisions, and taking them at the right time

One of the main and important topic of data mining is Association Rule Mining. Association rule mining, one of the most important and well researched techniques ;finds interesting association or correlation relationships among a large set of data items. With massive amount of data continuously being collected and stored in databases, many industries are becoming interested in mining association rules from their databases. For example the discovery of interesting association relationships among huge amounts of business transaction records can help catalog design, crossmarketing, loss leader analysis, and other business decision making process[9].

There are various algorithms have been proposed to discover the frequent item sets. The Apriori algorithm is one of the most popular algorithms in the mining of association rules in a centralized database, which will explained broadly later.

1.1.1 Problem Statement

Nowadays, information and knowledge represent the fundamental wealth of an organization. Enterprises try to utilize this wealth to gain competitive advantage when making important decisions. Enterprise software and systems include Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), and Supply Chain Management (SCM) systems.

The decision-makers in the Organizations are facing important challenge in today’s competitive environment. And According to the definition of data mining which refers to extracting information from large amount of data. This information is hidden by nature and can not be extracted without special intelligent tools and domain of experts who can analyze this knowledge and introduce it to decision makers.

With its goal to build one-on-one relationships with customers by developing an intimate understanding of their needs and wants, data mining can come in very useful. With all the data that is generated from various events (product inquiries,

sales, product reviews), there are many different ways data mining can provide more insight Like Identify most likely buyers / responders of new products and services also Discover time-variant associations between products and services to maximize sales and customer value.

1.1.2 Objectives

Main Objectives

The main objective of this Dissertation is to develop an efficient Apriori approach Application to generate and discover important and interestiong frequent itemsets and use these frequent itemsets for generating association rules from the generated knowledge by applying data mining techniques and tools.

Specific Objectives

- 1. Designing the suitable Apriori model for Our Data.
- 2. Implementing the algorithm based on the designed model.
- 3. Implementation and Executing , testing the algorithm using the collected Data.
- 4. Evaluating the performance of the algorithm based on time, speedup.

1.2 Dissertation Organization

Our dissertation is divided into four parts. The first contains the general introduction and we talk about overview, Problem Statement and issues and objectives of Dissertation. The second part contains the Litterateur topics to help the reader to understand the contribution of this project. This part contains Four chapters: Business intelligence, Data Warehouse and Data Mining also Group Melouki. The Third Part about Association Rules and their application on Data . the Fourth part is the application part and contains one chapter in which the design and development of Our Application Hisba-BI. The last parts is the general conclusion and future work.

2

LITERATURE PART

1. Business Intelligence
    - 1.1. Why is Business Intelligence Important ?
    - 1.2. Business Intelligence Architecture
    - 1.3. Data , Information and Knowledge
    - 1.4. Business Intelligence System
2. Data Warehouse
    - 2.1. What is Data Warehouse ?
    - 2.2. Types Of Databases
    - 2.3. Database vs. Data Warehouse vs.Data Mart
    - 2.4. Axtract , Load , Transform
3. Business Intelligence and Data Mining
4. Data Mining
    - 4.1. Data Mining Definition
    - 4.2. Tasks Of Data Mining
    - 4.3. Process Of Data Mining
    - 4.4. Why do we need Data Mining?
    - 4.5. Data Mining Application
    - 4.6. Top 5 Data Mining Algorithms
5. Melouki Group
    - 5.1. Tasks Of Melouki Group
    - 5.2. Process Of Data Mining
    - 5.3. Why do we need Data Mining?
    - 5.4. Problem
    - 5.5. Solution

Business intelligence may be defined as a set of mathematical models and analysis methodologies that exploit the available data to generate information and knowledge useful for complex decision-making processes. This opening chapter will describe in general terms the problems entailed in business intelligence, highlighting the interconnections with other disciplines and identifying the primary components typical of a business intelligence environment[1].

2.1 Business Intelligence

Def 1:-

The term Business Intelligence (BI) refers to technologies, applications and practices for the collection, integration, analysis, and presentation of business information. The purpose of Business Intelligence is to support better business decision making. Essentially, Business Intelligence systems are data-driven Decision Support Systems (DSS). Business Intelligence is sometimes used interchangeably with briefing books, report and query tools and executiveinformation systems[12].

Def 2:-

Business intelligence (BI) is a technology-driven process for analyzing data and presenting actionable information to help corporate executives, business managers and other end users make more informed business decisions. BI encompasses a variety of tools, applications and methodologies that enable organizations to collect data from internal systems and external sources, prepare it for analysis, develop and run queries against the data, an create reports, dashboards and data visualizations to make the analytical results available to corporate decision makers as well as operational workers[1].

The potential benefits of business intelligence programs include accelerating and improving decision making; optimizing internal business processes; increasing operational efficiency; driving new revenues; and gaining competitive advantages over business rivals. BI systems can also help companies identify market trends and spot business problems that need to be addressed.

To illustrate BI in action, here are a few departmental specific examples of insights and benefits that can come from its adoption and application [12]:

1. **Human Resources:**

   HR can tremendously benefit from the implementation of Business Intelligence utilizing employee productivity analysis, compensation and payroll tracking, and insights into employee satisfaction.

2. **Finance:**

   Business Intelligence can help finance departments by providing invaluable and in-depth insights into financial data. The application of BI can also help to track quarterly and annual budgets, identify potential problem areas before they cause any negative impacts, and improve the overall organizational business health and financial stability.

3. **Sales:**

   Business Intelligence can assist your company's sales force by providing visualizations of the sales cycle, in-depth conversion rates analytics, as well as total revenue analysis. BI can help your sales team to identify what's working as well as points of failure which can result in dramatically improved sales performance.

4. **Marketing:**

   BI provides the marketing department with a convenient way to view all current and past campaigns, the performance and trends of those campaigns, a breakdown of the cost per lead and the return on investment, site traffic analytics, as well as a multitude of other actionable pieces of information.

5. **Executive Leadership:**

   Plain and simple, Business Intelligence allows organizations to reduce costs by improving efficiency and productivity, improving sales, and revealing opportunities for continuous improvement. Business Intelligence allows members of Executive Leadership to more easily measure the organization's pulse by removing gray areas and eliminating the need to play the guessing game on how the company is doing.

2.1.1 Why Is Business Intelligence Important?

Business intelligence can help companies make better decisions by showing present and historical data within their business context. Analysts can leverage BI to provide performance and competitor benchmarks to make the organization run smoother and more efficiently. Analysts can also more easily spot market trends to increase sales or revenue. Used effectively, the right data can help with anything from compliance to hiring efforts [6]. The main reasons to invest in a solid BI strategy and system are:

1. **Gain New Customer Insights:** One of the primary reasons companies are investing their time, money, and efforts into Business Intelligence is because it gives them a greater ability to observe and analyze current customer buying trends. Once you utilize BI to understand what your consumers are buying and the buying motive, you can use this information to create products and product improvements to meet their expectations and needs and, as a result, improve your organization's bottom-line.

2. **Actionable Information:** An effective Business Intelligence system serves as a means to identify key organizational patterns and trends. A BI system also allows you to understand the implications of various organizational processes and changes, allowing you to make informed decisions and act accordingly.

3. **Efficiency Improvements:** BI Systems help improve organizational efficiency which consequently increases productivity and can potentially increase revenue. Business Intelligence systems allow businesses to share vital information across departments with ease, saving time on reporting, data extraction, and data interpretation. Making the sharing of information easier and more efficient permits organizations to eliminate redundant roles and duties, allowing the employees to focus on their work instead of focusing on processing data.

4. **Sales Insight:** Sales and marketing teams alike want to keep track of their customers, and most utilize Customer Relationship Management (CRM) application to do so. CRMs are designed to handle all interactions with customers. Because they house all customer communications and interactions, there is a wealth of data and information that can be interpreted and used to strategic initiatives. BI systems help organizations with everything from identifying new customers, tracking and retaining existing ones, and providing post-sale services.

5. **Real-Time Data:** When executives and decision-makers have to wait for re-

ports to be compiled by various departments, the data is prone to human error and is at risk of being outdated before it's even submitted for review. BI systems provide users with access to data in real-time through various means including spreadsheets, visual dashboards, and scheduled emails. Large amounts can be assimilated, interpreted, and distributed quickly and accurately when leveraging Business Intelligence tools.

In summary, BI makes it possible to combine data from multiple sources, analyze the information into a digested format, and then disseminate the information to relevant stakeholders. This allows companies to see the big picture and make smart business decisions. There are always inherent risks when it comes to making any business decision, but those risks aren't as prominent or worrisome when implementing an effective and reliable BI solution. Business Intelligent organizations can move forward in an increasingly data-driven climate with confidence knowing they are prepared for any challenge that[6].

2.1.2 Business Intelligence Architecture

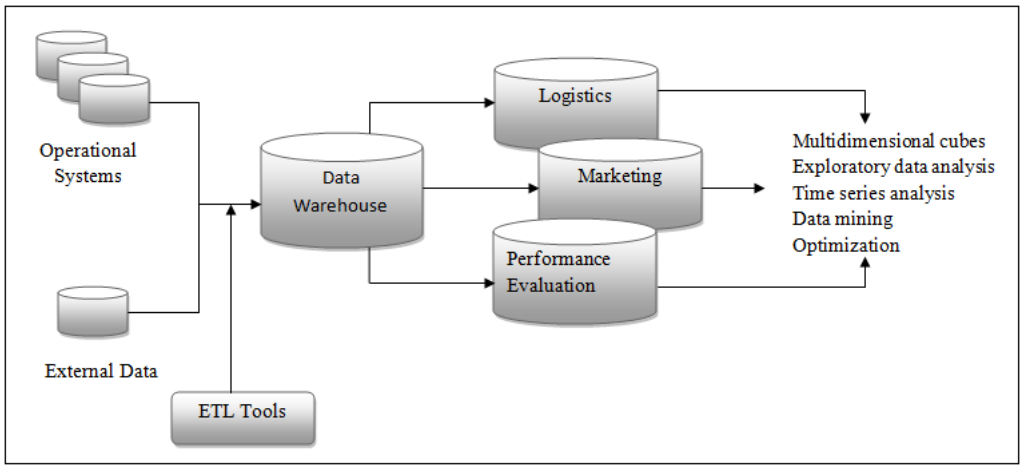


Figure 2.1: A typical business intelligence architecture

The architecture of a business intelligence system, depicted in Figure Figure 2.1, includes three major components. Data sources. In a first stage, it is necessary to gather and integrate the data stored in the various primary and secondary sources, which are heterogeneous in origin and type. The sources consist for the most part of data belonging to operational systems, but may also include unstructured documents, such as emails and data received from external providers. Generally speaking, a major effort is required to unify and integrate the different data sources(Data warehouses and data marts). Using extraction and transformation tools known as extract, transform, load (ETL), the data originating from the different sources

are stored in databases intended to support business intelligence analyses. These databases are usually referred to as data warehouses and data marts Data are finally extracted and used to feed mathematical models and analysis methodologies intended to support decision makers. Also Carlo (2009) uses the following pyramid Figure 2.2 to describe how business intelligence system is constructed[1].

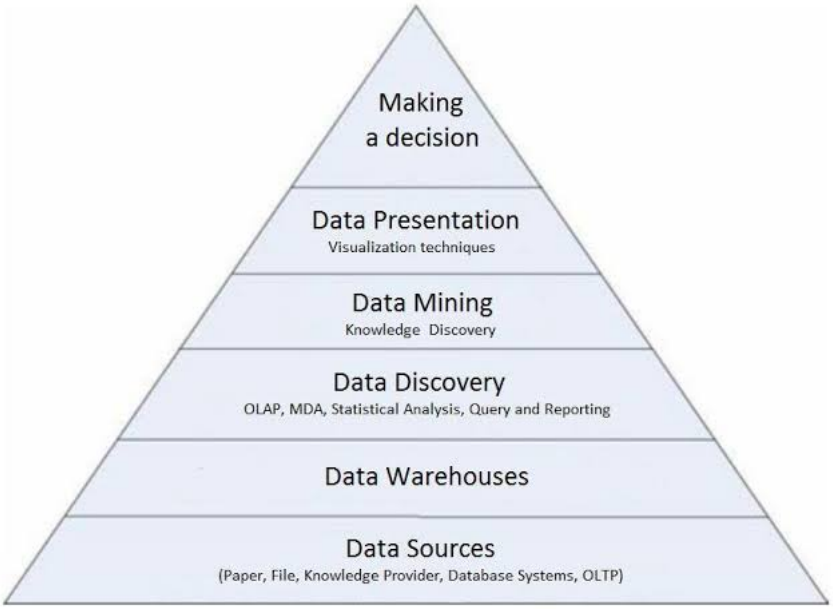


Figure 2.2: Business Intelligence System Carlo 2009

Data sources: The sources mostly consist of data belonging to operationalize systems, but may also include unstructured data, such as emails, and data received from external providers.

Data warehouse/Data mart: Data warehouses are used to consolidate different kinds of data into a central location using a process known as extract, transform and load (ETL) and standardize these results across systems that are allowed to be queried. Data marts are generally small warehouses that focus on information on a single department, instead of collecting data across a company. They limit the complexity of databases and are cheaper to implement than full warehouses[1].

Data exploration: Data exploration is a passive BI analysis consisting of query and reporting systems, as well as statistical method.

Data mining: Data mining is active BI methodologies with the purpose of information and knowledge extraction from data.

Optimization: Optimization model allows us to determine the best solution out of a set of alternative actions, which is usually fairly extensive and sometimes

even infinite.

Decisions: When business intelligence methodologies are available and successfully adopted, the choice of a decision pertains to the decision makers, who may also take advantage of informal and unstructured information available to adapt and modify the recommendations and the conclusions achieved through the use of mathematical models.

2.1.3 Data , Information and Knowledge

In BI context, we always see the word data, information, and knowledge which could lead us getting confused on its use and implication. Carlo (2009) distinguishes their definition.

Data: It refers to a structured codification of single primary entities and as well as of transactions involving two or more primary entities Carlo (2009). BI is popular among companies mainly because of analysis of data that is of any form and formulate a strategy accordingly. Generally data is classified into three types—structured data, semi-structured data, and unstructured data. Structured data are information that is fixed form, the data may be a collection of forms of websites, and detailed address that can be easily read by the computers since the data is already standardized. Unstructured data are information that cannot be easily read by computers, which may be text, documents, video tapes, websites, and pictures (Jermol et al. 2003), or any other type of information that cannot be clearly sorted or organized into rows and columns. Information is used many times to Company data are found across different locations and places in the form of Customer Relation Management (CRM) programs, marketing automation systems and social media platforms.

Information: It refers to the result of extraction and processing activities carried out on data, and it appears meaningful for those who receive it in a specific domain.

Knowledge: It is formed from information which is used to make decisions and develop the corresponding actions. Hence, we could say that knowledge consists of information that puts to work into a specific domain, and it is enhanced by the experience and competence of decision makers in tackling and solving complex problems[1].

2.1.4 Business intelligence System

The development of a business intelligence system can be assimilated to a project, with a specific final objective, expected development times and costs, and the usage and coordination of the resources needed to perform planned Project [1].

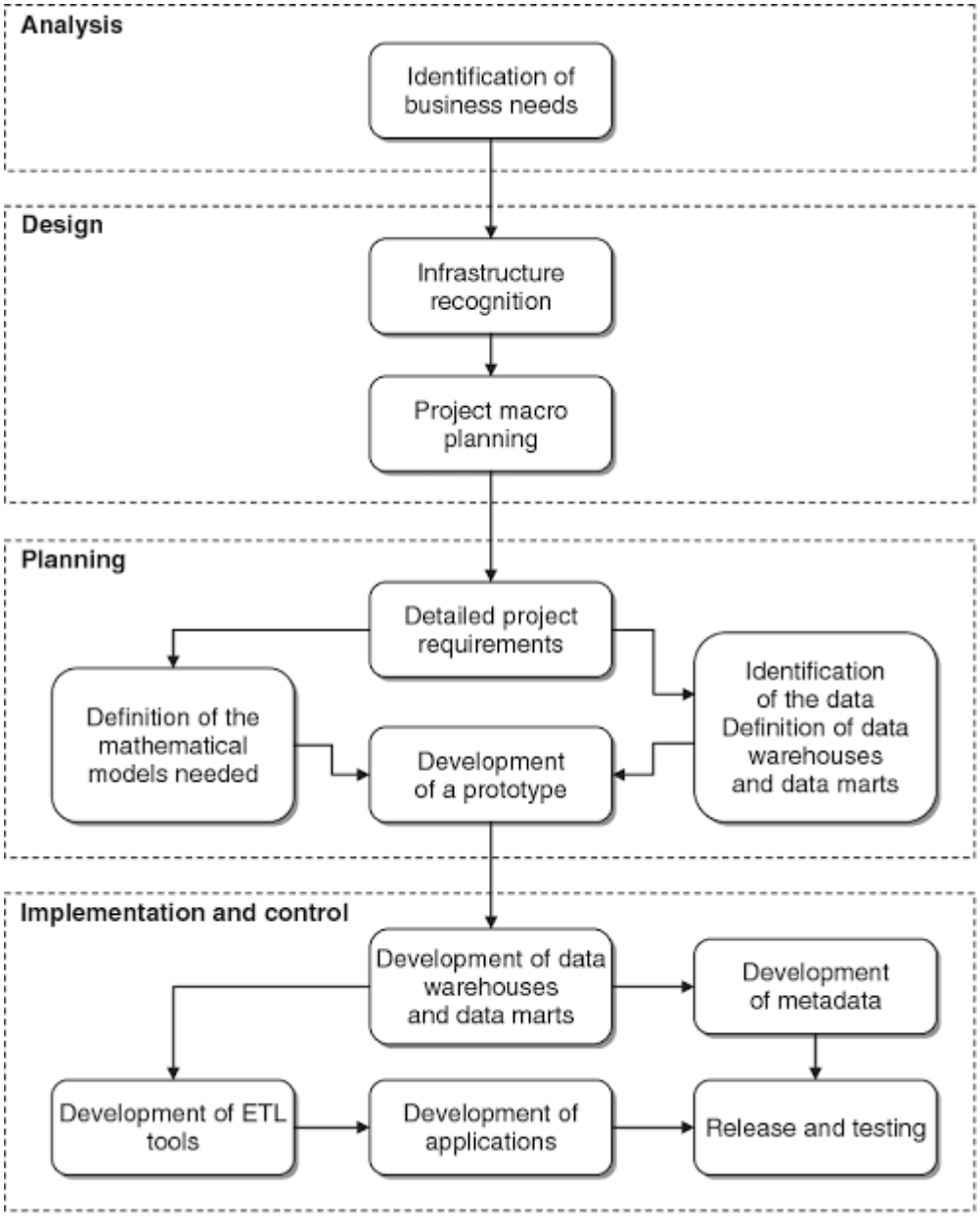


Figure 2.3: Phases in the development of a business intelligence system

Analysis. During the first phase, the needs of the organization relative to the development of a business intelligence system should be carefully identified. This

preliminary phase is generally conducted through a series of interviews of knowledge workers performing different roles and activities within the organization. It is necessary to clearly describe the general objectives and priorities of the project, as well as to set out the costs and benefits deriving from the development of the business intelligence system[1].

Design. The second phase includes two sub-phases and is aimed at deriving a provisional plan of the overall architecture, taking into account any development in the near future and the evolution of the system in the mid term. First, it is necessary to make an assessment of the existing information infrastructures. Moreover, the main decision-making processes that are to be supported by the business intelligence system should be examined, in order to adequately determine the information requirements. Later on, using classical project management methodologies, the project plan will be laid down, identifying development phases, priorities, expected execution times and costs, together with the required roles and resources.

Planning. The planning stage includes a sub-phase where the functions of the business intelligence system are defined and described in greater detail. Subsequently, existing data as well as other data that might be retrieved externally are assessed. This allows the information structures of the business intelligence architecture, which consist of a central data warehouse and possibly some satellite data marts, to be designed. Simultaneously with the recognition of the available data, the mathematical models to be adopted should be defined, ensuring the availability of the data required to feed each model and verifying that the efficiency of the algorithms to be utilized will be adequate for the magnitude of the resulting problems. Finally, it is appropriate to create a system prototype, at low cost and with limited capabilities, in order to uncover beforehand any discrepancy between actual needs and project specifications.

Implementation and control. The last phase consists of five main sub-phases. First, the data warehouse and each specific data mart are developed. These represent the information infrastructures that will feed the business intelligence system. In order to explain the meaning of the data contained in the data warehouse and the transformations applied in advance to the primary data, a metadata archive should be created, as described in Chapter 3. Moreover, ETL procedures are set out to extract and transform the data existing in the primary sources, loading them into the data warehouse and the data marts. The next step is aimed at developing the core business intelligence applications that allow the planned analyses to be carried out. Finally, the system is released for test and usage.

2.2 Data Warhouse

Big DataThough most companies have plenty of data, a lot of work needs to be done to get it ready. The data must be collected, cleaned and formatted properly, and stored in one place for analysis. This process is known as data warehousing[7].

2.2.1 What is a Data Warehouse?

As the name implies, a data warehouse takes a lot of different data from various sources and stores it in one place. That data can come from transactional systems, such as marketing, sales, CRM and ERP systems, as well as external sources such as the web. The warehouse then holds the data so it can be used for analytical processing and generating reports. Storing data this way is critical for running business intelligence. While the data held in a data warehouse is typically available in other locations, it's not really usable for analysis until warehousing is done. The data sources are usually transactional systems, meaning they're designed and built to use data when performing specific functions. Running reports and analysis from data held in those transactional systems would likely interrupt their normal operations, making it impossible to perform business intelligence and actually run the business at the same time. Data warehouses, on the other hand, are designed specifically for analysis, making it possible to use all records from all sources at the same time to answer questions [7].

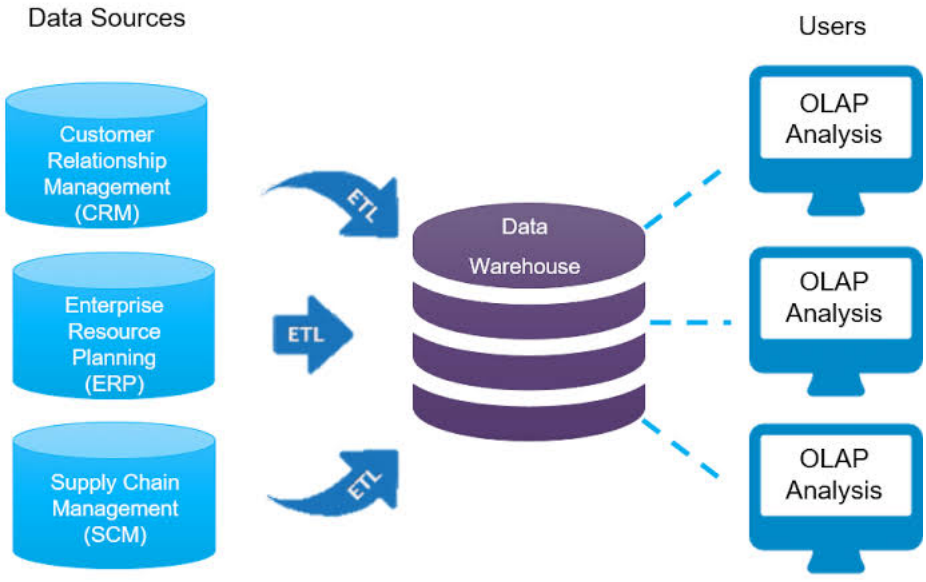


Figure 2.4: Data Warehouse Sources

2.2.2 Types of Databases

Another reason databases are necessary is that data must often be stored in different ways for analytics than for processing transactions. One key distinction in the ways data can be stored is the difference between a relational and multidimensional database:

Relational database can be thought of as using a two-dimensional structure. Imagine a simple database such as an Excel spreadsheet. The data is organized using rows and columns. One key is that the data must be normalized so each attribute can be put in the proper place and the entries can be sorted.

Multidimensional database, as the name implies, stores data based on more than two dimensions. Rather than an entry being defined by its row and column like in a spreadsheet, each entity in a multidimensional database can contain a number of different attributes and can exist independently of the other entries. One example of a relational database could be a table listing different products and the number of units sold in each state in a given calendar year. In this table, the rows could be labeled with the products, with the columns assigned to the states and the numerical values filled in accordingly. The table could then be sorted for each state based on the number of sales.

In this example, if you want to look at any other attributes – for example, the number of sales in each month – that would require a separate relational database. That could be a table for each product, with the columns representing each state and the rows representing each month. In comparison, a multidimensional database could hold all of that information in the same database. Each product could be thought of as its own entity in the database, with values stored for sales in each state, sales per month, etc [7].

2.2.3 Database vs. Data Warehouse vs. Data Mart

Using the concepts laid out by Inmon, IBM and others, a data warehouse can be defined as a top-down of the entire collection of data that can be used for analysis, whereas a database is a smaller set of data used for a specific purpose, often for transaction processing.
Another concept relevant to business intelligence is the data mart. Essentially, a data mart is a smaller, more focused version of a data warehouse. Whereas the warehouse holds all of the data from the entire organization necessary to perform business intelligence, a data mart holds all of the data about one particular area. For example, data marts may be broken up according to different operational ar-

eas, in which case the company would have a dedicated sales data mart, a dedicated finance data mart, etc. As with a data warehouse, the information would still come from multiple sources and be transformed into a common format, but the data mart would strictly contain sales or finance or another group of data. This concept comes in part from the work of Ralph Kimball, one of the "fathers of data warehousing," along with Inmon. Rather than Inmon's top-down approach, Kimball advocated for building individual data marts for each section of an enterprise and then integrating them [7].

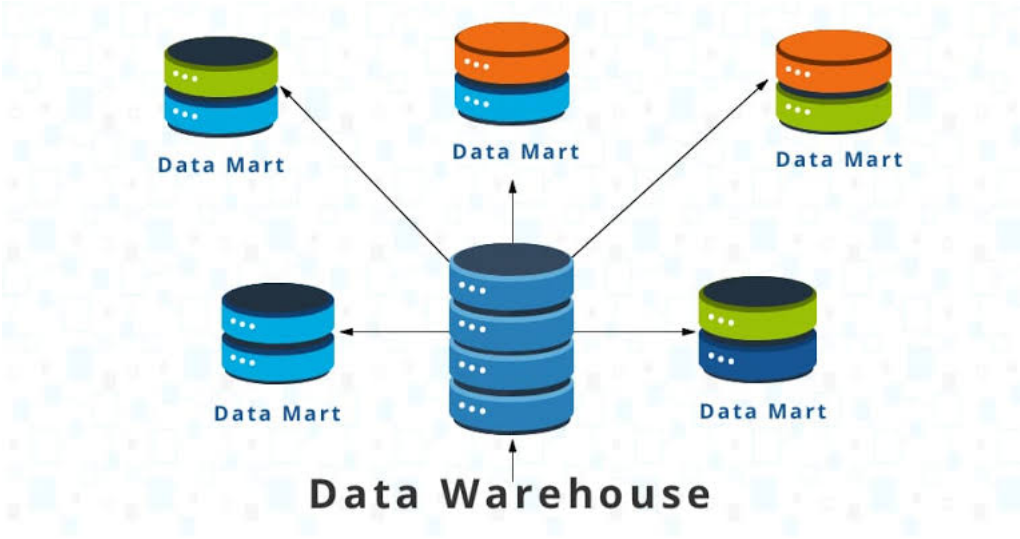


Figure 2.5: Data Warehouse and Data Marts

2.2.4 Extract, Transform and Load

Populating a data warehouse isn't as simple as just copying out of other systems. Data normally isn't ready for analysis when it's taken straight out of those transactional systems. Different applications may create data in different formats, or the data may be organized based on different criteria. In order to get a full picture of the entire business, it's necessary to do some work to the data so different sets can be analyzed in relation to each other. The quality of the data must also be checked before performing any business intelligence. Errors and other problems can lead to inaccurate data that ultimately impairs decision making. Data warehousing allows records to be verified before they're used for analysis, making sure the organization has one "single version of the truth" to use for business intelligence[7].

Getting data ready for analysis Gathering, preparing and storing data in the data warehouse is done through a process called Extract, Transform and Load (ETL). ETL tools:

1. Extract data from internal and external sources

2. Transform it into a standard format – for example, converting dates to the same format, splitting customer names into first and last, etc.

3. Load data into the data warehouse.

The transform stage is especially important when it comes to business intelligence. In order to get complete, accurate information about the organization as a whole, businesses must look at data from a variety of sources at the same time. However, those sources usually include systems that come from different vendors, running on different types of hardware, and managed by different employees. That means a lot of work likely must be done in order for all of the data to make sense together.

2.2.5 Extract, Load, Transform

In addition to ETL, **ELT** (Extract, Load and Transform) is also becoming a more viable option for businesses. In this process, data is pulled from the sources and then transferred to a Staging Database, where integrity and business rule checks are performed. The data is then moved into the warehouse, where it's transformed into the necessary formats. Previously, ETL was the only option because earlier data warehouses didn't have the capacity to perform the transformation. Therefore, other tools were needed in between the data sources and the warehouse. However, as technology advances and data volumes grow, ELT is becoming more attractive due to how long ETL can take. If sets of data don't make it to the warehouse until after they are transformed, that means there could be a significant delay before it becomes usable for the business. Transforming the data within the warehouse provides faster access to that data. In addition, ETL is less flexible. If data is transformed before it gets to the warehouse, that usually means only the transformed version of the data is available for analysis. On the other hand, if data is loaded and then transformed, the raw data will still be available. That means if the company's needs change while analysis is performed, it will be more likely to still have access to the data it needs[7].

2.3 Business Intelligence and Data Mining

Business is the act of doing something productive to serve someone's needs, and thus earn a living and make the world a better place. Business activities are recorded on paper or using electronic media, and then these records become data. There is more data from customers' responses and on the industry as a whole. All this data can be analyzed and mined using special tools and techniques to generate patterns and intelligence, which reflect how the business is functioning. These ideas can then be fed back into the business so that it can evolve to become more effective and efficient in serving customer needs. And the cycle continues [2].

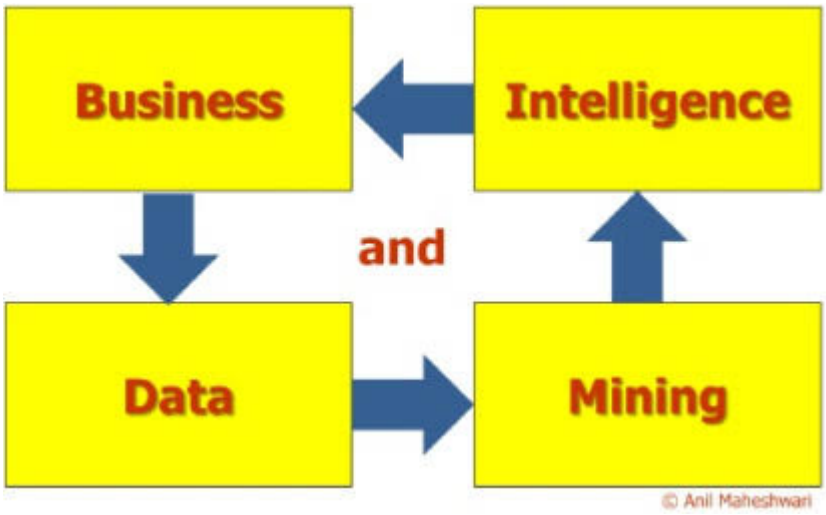


Figure 2.6: Business intelligence and Data Mining Cycle

2.4 Data Mining

As observed in previous chapters, the evolving technologies of information gathering and storage have made available huge amounts of data within most application domains, such as the business world, the scientific and medical community, and public administration. The set of activities involved in the analysis of these large databases, usually with the purpose of extracting useful knowledge to support decision making, has been referred to in different ways, such as data mining, knowledge discovery, pattern recognition and machine learning. In particular, the term data mining indicates the process of exploration and analysis of a dataset, usually of large size, in order to find regular patterns, to extract relevant knowledge and to obtain meaningful recurring rules. Data mining plays an ever-growing role in both theoretical studies and applications. In this chapter we wish to describe and

characterize data mining activities with respect to investigation purposes and analysis methodologies. The relevant properties of input data will also be discussed. Finally, we will describe the data mining process and its articulation in distinct phases[7].

2.4.1 Data Mining Definition

Data mining is the process of extracting meaningful information from large quantities of data. It involves uncovering patterns in the data and is often tied to data warehousing because it makes such large amounts of data usable. Data elements are grouped into distinct categories so that predictions can be made about other pieces of data. For example, a bank may wish to ascertain the characteristics that typify customers who pay back loans. Although this could be done with database queries, the bank would first have to know what customer attributes to query for. Data mining can be used to identify what those attributes are and then make predictions about futurecustomer behavior .

Data Mining also known as Knowledge Discovery in Databases, refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data stored in databases [7].

Steps Involved in KDD Process [12]:
Data Cleaning: Data cleaning is defined as removal of noisy and irrelevant data from collection. Cleaning in case of Missing values. Cleaning noisy data, where noise is a random or variance error. Cleaning with Data discrepancy detection and Data transformation tools.

Data Integration: Data integration is defined as heterogeneous data from multiple sources combined in a common source(DataWarehouse). Data integration using Data Migration tools. Data integration using Data Synchronization tools. Data integration using ETL(Extract-Load-Transformation) process.

Data Selection: Data selection is defined as the process where data relevant to the analysis is decided and retrieved from the data collection. Data selection using Neural network. Data selection using Decision Trees. Data selection using Naive bayes. Data selection using Clustering, Regression, etc.

Data Transformation: Data Transformation is defined as the process of transforming data into appropriate form required by mining procedure.

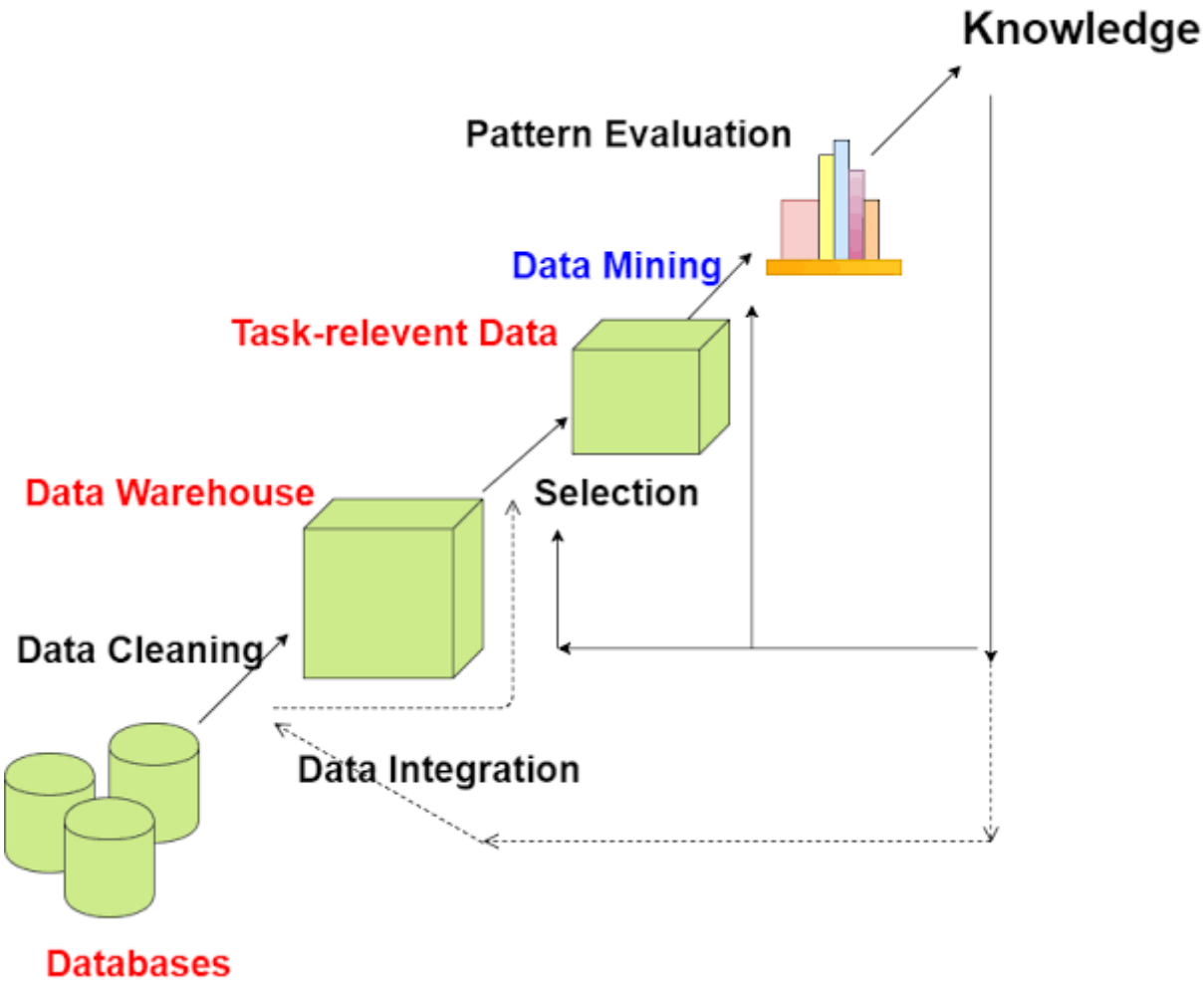


Figure 2.7: Business intelligence and Data Mining Cycle

Data Transformation is a two step process:

1. **Data Mapping:** Assigning elements from source base to destination to capture transformations.

2. **Code generation:**Creation of the actual transformation program..

Data Mining: Data mining is defined as clever techniques that are applied to extract patterns potentially useful. Transforms task relevant data into patterns. Decides purpose of model using classification or characterization.

Pattern Evaluation: Pattern Evaluation is defined as as identifying strictly increasing patterns representing knowledge based on given measures. Find interestingness score of each pattern. Uses summarization and Visualization to make data understandable by user. Knowledge representation: Knowledge representation is defined as technique which utilizes visualization tools to represent data mining results. Generate reports. Generate tables. Generate discriminant rules, classification rules, characterization rules, etc.

2.4.2 Tasks Of Data Mining

Many problems of intellectual, economic, and business interest can be phrased in terms of the following six tasks:

1. Classification

2. Estimation

3. Prediction

4. Affinity grouping

5. Clustering

6. Description and profiling

The first three are all examples of directed data mining, where the goal is tofind the value of a particular target variable. Affinity grouping and clustering are undirected tasks where the goal is to uncover structure in data without respect to a particular target variable. Profiling is a descriptive task that may be either

directed or undirected.

CLASSIFICATION Classification, one of the most common data mining tasks, seems to be a human imperative. In order to understand and communicate about the world, we are constantly classifying, categorizing, and grading. We divide living things into phyla, species, and general; matter into elements; dogs into breeds; people into races; steaks and maple syrup into USDA grades.

Classification consists of examining the features of a newly presented object and assigning it to one of a predefined set of classes. The objects to be classified are generally represented by records in a database table or a file, and the act of classification consists of adding a new column with a class code of some kind. The classification task is characterized by a well-defined definition of the classes, and a training set consisting of preclassified examples. The task is to build a model of some kind that can be applied to unclassified data in order to classify it. Examples of classification tasks that have been addressed using the techniques described in this book include:

1. Classifying credit applicants as low, medium, or high risk
2. Choosing content to be displayed on a Web page
3. Determining which phone numbers correspond to fax machines
4. Spotting fraudulent insurance claims
5. Assigning industry codes and job designations on the basis of free-text job descriptions

ESTIMATION Classification deals with discrete outcomes: yes or no; measles, rubella, or chicken pox. Estimation deals with continuously valued outcomes. Given some input data, estimation comes up with a value for some unknown continuous variable such as income, height, or credit card balance. In practice, estimation is often used to perform a classification task. A credit card company wishing to sell advertising space in its billing envelopes to a ski boot manufacturer might build a classification model that put all of its cardholders into one of two classes, skier or nonskier. Another approach is to build a model that assigns each cardholder a “propensity to ski score.” This might be a value from 0 to 1 indicating the estimated probability that the cardholder is a skier. The classification task now comes down to establishing a threshold score. Anyone with a score greater than or equal to the threshold is classed as a skier, and anyone with a lower score is

considered not to be a skier. The estimation approach has the great advantage that the individual records can be rank ordered according to the estimate[**?**].

PREDICTION Prediction is the same as classification or estimation, except that the records are classified according to some predicted future behavior or estimated future value. In a prediction task, the only way to check the accuracy of the classification is to wait and see. The primary reason for treating prediction as a separate task from classification and estimation is that in predictive modeling there are additional issues regarding the temporal relationship of the input variables or predictors to the target variable. Any of the techniques used for classification and estimation can be adapted for use in prediction by using training examples where the value of the variable to be predicted is already known, along with historical data for those examples. The historical data is used to build a model that explains the current observed behavior. When this model is applied to current inputs, the result is a prediction of future behavior.

AFFINITY GROUPING OR ASSOCIATION RULES

The task of affinity grouping is to determine which things go together. The prototypical example is determining what things go together in a shopping cart at the supermarket, the task at the heart of market basket analysis. Retail chains can use affinity grouping to plan the arrangement of items on store shelves or in a catalog so that items often purchased together will be seen together. Affinity grouping can also be used to identify cross-selling opportunities and to design attractive packages or groupings of product and services. Affinity grouping is one simple approach to generating rules from data. If two items, say cat food and kitty litter, occur together frequently enough, we can generate two association rules:

1. People who buy cat food also buy kitty litter with probability P1.

2. People who buy kitty litter also buy cat food with probability P2.

Clustering Clustering is the task of segmenting a heterogeneous population into a number of more homogeneous subgroups or clusters. What distinguishes clustering from classification is that clustering does not rely on predefined classes. In classification, each record is assigned a predefined class on the basis of a model developed through training on preclassified examples. In clustering, there are no predefined classes and no examples. The records are grouped together on the basis of self-similarity. It is up to the user to determine what meaning, if any, to attach to the resulting clusters. Clusters of symptoms might indicate different diseases. Clusters of customer attributes might indicate different market segments. Clustering is often done as a prelude to some other form of data mining or modeling.

For example, clustering might be the first step in a market segmentation effort: Instead of trying to come up with a one-size-fits-all rule for “what kind of promotion do customers respond to best,” first divide the customer base into clusters or people with similar buying habits, and then ask what kind of promotion works best for each cluster

2.4.3 Data mining and its process

Data mining is an interactive process. Take a look at the following steps.

1. **Requirement gathering**
   Data mining projects start with requirement gathering and understanding. Data mining analysts or users define the requirement scope with the vendor business perspective. Once the scope is defined, we move to the next phase.
2. **Data exploration**
   In this step, Data mining experts gather, evaluate, and explore the requirement or project. Experts understand the problems, challenges, and convert them to metadata. In this step, data mining statistics are used to identify and convert data patterns.
3. **Data preparations**
   Data mining experts convert the data into meaningful information for the modeling step. They use the ETL process – extract, transform, and load. They are also responsible for creating new data attributes. Here various tools are used to present data in a structural format without changing the meaning of data sets.
4. **Modeling**
   Data experts put their best tools in place for this step as this plays a vital role in the complete processing of data. All modeling methods are applied to filter the data in an appropriate manner. Modeling and evaluation are correlated steps and are followed at the same time to check the parameters. Once the final modeling is done the final outcome is quality proven.
5. **Evaluation**
   This is the filtering process after successful modeling. If the outcome is not satisfactory, then it is transferred to the model again. Upon final outcome, the requirement is checked again with the vendor so no point is missed. Data mining experts judge the complete result at the end.
6. **Deployment**
   This is the final stage of the complete process. Experts present the data to vendors in the form of spreadsheets or graphs. Have a look at the below diagram for CRISP DM- Cross Industry-standard process for Data mining.

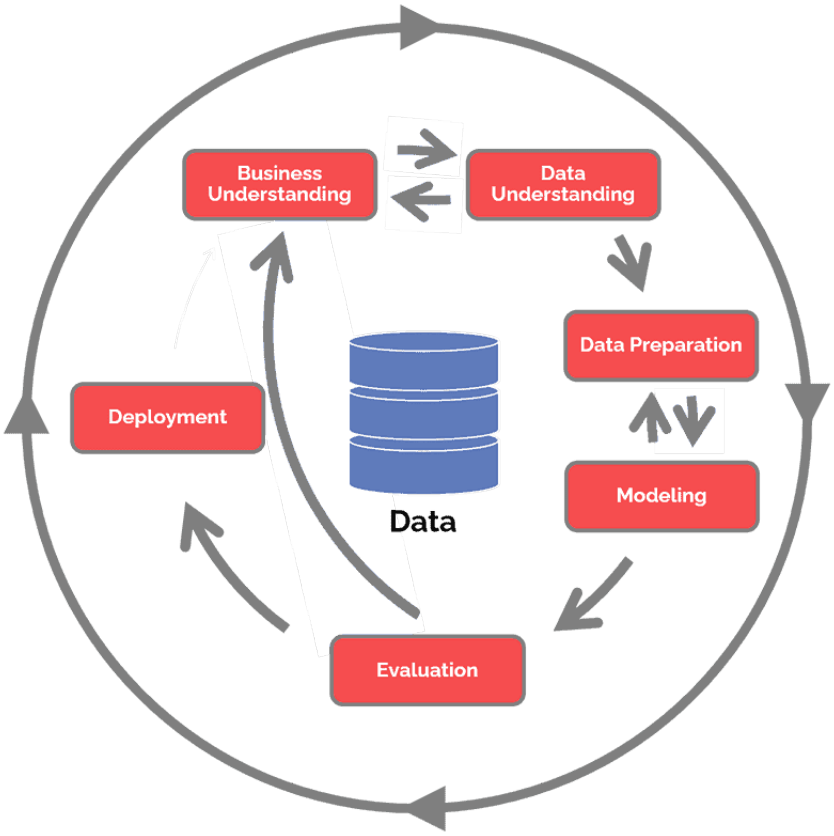


Figure 2.8: Cross-Industry Standard Process for Data Mining

2.4.4 Why do we need Data Mining?

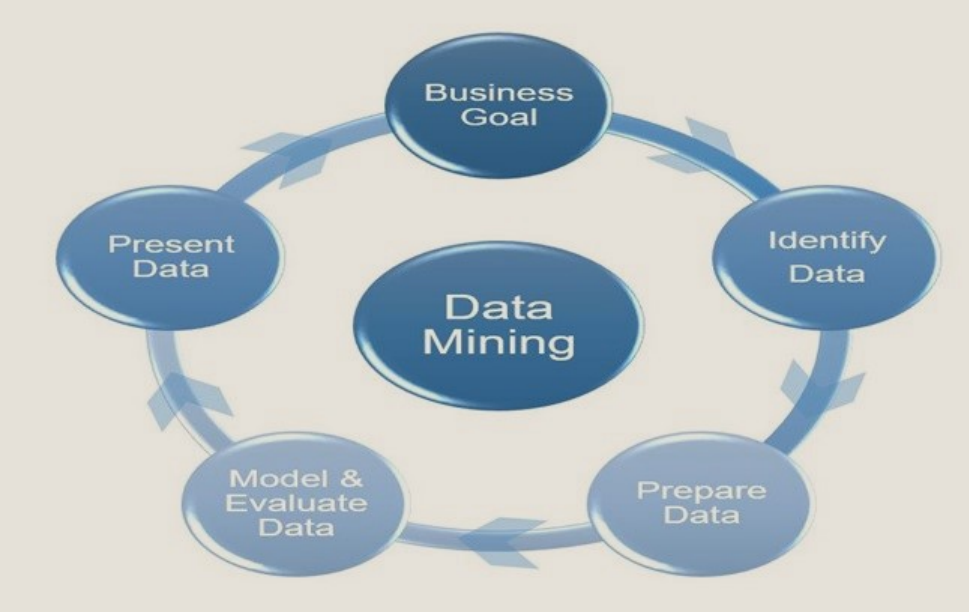


Figure 2.9: Goals Of Data Mining

1. Data mining is the procedure of capturing large sets of data in order to identify the insights and visions of that data. Nowadays, the demand of data industry is rapidly growing which has also increased the demands for data analysts and data scientists.
2. With this technique, we analyze the data and then convert that data into meaningful information. This helps the business to take accurate and better decisions in an organization.
3. Data mining helps to develop smart market decision, run accurate campaigns, make predictions, and more.
4. With the help of Data mining, we can analyze customer behaviors and their insights. This leads to great success and data-driven business.

2.4.5 Data Mining Application

Figure 2.10: Business intelligence and Data Mining Cycle

1. **Research and surveys.** Data mining can be used for product research, surveys, market research, and analysis. Information can be gathered that is quite useful in driving new marketing campaigns and promotions.

2. **Information collection.** Through the web scraping process, it is possible to collect information regarding investors, investments, and funds by scraping through related websites and databases.

3. Customer opinions. Customer views and suggestions play an important role in the way a company operates. The information can be readily be found on forums, blogs, and other resources where customers freely provide their views.

4. **Data scanning.** Data collected and stored will not be important unless scanned. Scanning is important to identify the patterns and similarities contained in data entries.

5. **Extraction of information.** This is the processing of identifying useful patterns in data that can be used in the decision-making process. This is so because decision making must be based on sound information and facts.

6. **Pre-processing of data.** Usually, the data collected is stored in the data warehouse. This data needs to be pre-processed. By pre-processing, we mean some data that may be deemed unimportant may be removed manually by

data mining experts.

7. **Web data.** Web data usually poses many challenges in mining. This is so because of its nature. For instance, web data can be deemed as dynamic, meaning it keeps changing from time to time. Therefore, it means the process of data mining should be repeated at regular intervals.

8. **Competitor analysis.** There is a need to understand how your competitors are fairing on in the business market. You need to know both their weaknesses and strengths. Their methods of marketing and distribution can be mined. How they reduce their overall costs is also quite important.

9. **Online research.** The internet is highly regarded for the vast amounts of information it contains. It is clear that it is the largest source of information. It is possible to gather a lot of information regarding different companies, customers, and business clients. It is possible to detect fraud through online means.

10. **News.** Nowadays, with almost all major newspapers and news sources posting their news online, it is possible to gather information regarding trends and other critical areas. In this way, it is possible to be in a better position of competing in the market.

11. **Updating data.** This is quite important. Data collected will be useless unless it is updated. This is to ensure that the information is relevant so as to make decisions from it.

2.4.6 Top 5 Data Mining Algorithms

1. **Algorithm C4.5**
   is one of the top data mining algorithms and was developed by Ross Quinlan. C4.5 is used to generate a classifier in the form of a decision tree from a set of data that has already been classified. Classifier here refers to a data mining tool that takes data that we need to classify and tries to predict the class of new data. Every data point will have its own attributes. The decision tree created by C4.5 poses a question about the value of an attribute and depending on those values, the new data gets classified. The training dataset is labelled with lasses making C4.5 a supervised learning algorithm. Decision trees are always easy to interpret and explain making C4.5 fast and popular compared to other data mining algorithms[16].

2. **K-mean Algorithm**
   One of the most common clustering algorithms, k-means works by creating a k number of groups from a set of objects based on the similarity between objects. It may not be guaranteed that group members will be exactly similar, but group members will be more similar as compared to non-group members. As per standard implementations, k-means is an unsupervised learning algorithm as it learns the cluster on its own without any external information [16].

3. **Support Vector Machines**
   In terms of tasks, Support vector machine (SVM) works similar to C4.5 algorithm except that SVM doesn't use any decision trees at all. SVM learns the datasets and defines a hyperplane to classify data into two classes. A hyperplane is an equation for a line that looks something like "y = mx + b". SVM exaggerates to project your data to higher dimensions. Once projected, SVM defined the best hyperplane to separate the data into the two classes [16].

4. **Apriori Algorithm**

   Apriori algorithm works by learning association rules. Association rules are a data mining technique that is used for learning correlations between variables in a database. Once the association rules are learned, it is applied to a database containing a large number of transactions. Apriori algorithm is used for discovering interesting patterns and mutual relationships and hence is treated as an unsupervised learning approach. Thought the algorithm is highly efficient, it consumes a lot of memory, utilizes a lot of disk space and takes a lot of time [16].

5. **Expectation-Maximization Algorithm**
   Expectation-Maximization (EM) is used as a clustering algorithm, just like the k-means algorithm for knowledge discovery. EM algorithm work in iterations to optimize the chances of seeing observed data. Next, it estimates the parameters of the statistical model with unobserved variables, thereby generating some observed data. Expectation-Maximization (EM) algorithm is again unsupervised learning since we are using it without providing any labelled class information [16]

2.5 Melouki Group

Group Melouki is A Group Orgnized into 9 Society in Algéria , M'sila , They Provide Services In : Software , Pharmacy , Analysis , Logistics ,nutrition and beauty. Group Melouki Was started in 2006 with A " Melouki Pharmacy " Then They Build a Laboratory after 4 Years Fom Its Start , and Since the Parapharmaceutical products market relatively linked with the parapharm markdet , in 2015 they started importing parapharmaceuticals to further meet the needs of their Customers . Group Melouki become morde demanding , espacially in IT Solutions , M-Formatik was the perfect Solution o attends their Needs [17].

2.5.1 Tasks Of Melouki Group

In Group Melouki [17] Each Company Has its Own Tasks :
M-FORMATIK : is a company providing IT Services , with the aim of Producing It Products and Services to bring Productivity and efficiency to the realization of Customer activities.

PARAPHARM STREET : is a wholesale company of parapharmaceuticals and medical equipement that guarantee new trends in products demanded by all categories .

TECHNICLAB ALGERIA : SARL TECHNICLAB is a Laboratoy equipement distribution company . Offre a wide and Varied range Of Quality

MELOUKISANTE: is a company speciliazes i, import of medical material and laboratory equiemet as a main activity .

There is also LogistiX , Melouki Laboratory and Other companies.

2.5.2 Problem

Due To the Big Data That Group Melouki Have From 9 Companies (Data Sources) Their Data are always in a big Development and Also Their relations With Customers and Their Needs Witch They Should Deal with it with A big Caution and fear and must maintain it Correctly by using a nice marketing strategy to increase Their Product Sales and helps the Customers buy their items with ease and enhance the sales performance .

2.5.3 Solution

The Solution That we suggest is Association Rules , exactly Apriori Algorithm .

3

ASSOCIATION RULES

1. Association Rules
    1.1. Algorithms For Association Rules
2. Apriori Algorithm
    2.1. Basic Of Apriori Algorithm
3. Market Basket Analysis

3.1 Association Rules

Association rule mining, one of the most important and well researched techniques. It aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items indata repositories. Association rules are widely used in various areas such as telecommunication networks, market and risk management, inventory control [13].

In a database of transactions D with a set of n binary attributes (items) I, a rule is defined as an implication of the form :

X ==>Y where X, Y $\subseteq$ **and X** $\bigcap Y = \varnothing$

There are two important basic measures for association rules, support (s) and confidence (c). Since the database is large and users concern about only those frequently items, usually thresholds of support and confidence are pre-defined by users to drop rules that are not interesting or useful. The two thresholds are called minimum support and minimum confidence [13].

Agrawal. [9] defined association rules as the implication rules that inform the user about items most likely to occur in some transactions of a database. They are advantageous to use because they are simple, intuitive and do not make assumptions of any models. Their mining requires satisfying a user-specified minimum support and a user-specified minimum confidence from a given database at the same time.

Support of an association rule is defined as the percentage of records to the total number of records in the database. The count for each item is increased by one every time the item is encountered in different transaction T in database D during the scanning process. Support(s) is calculated by the following [13].

$$\text{Support (XY)} = \frac{Support\ count\ of\ XY}{Total\ number\ of\ transaction\ in\ D}$$

Figure 3.1: Association Rule Support

Confidence of an association rule is defined as the percentage of the number of transactions to the total number of records that contain X, where if the percentage exceeds the threshold of confidence an interesting association rule X ==>Y can be generated [13], Confidence is a measure of strength of the association rules.

$$\text{Confidence (X|Y)} = \frac{Support\ (XY)}{Support(X)}$$

Figure 3.2: Association Rule Confidence

3.1.1 Apriori Algorithm

Apriori is an algorithm that has been proposed in [14]. The discovery of frequent itemsets is accomplished in several iterations. In each scan, a full scan of training data is required to count new candidate itemsets from frequent itemsets already found in the previous step. Apriori uses the Apriori property to improve the efficiency of the search process by reducing the size of the candidate itemsets list for each iteration. The Apriori property saysthat every sub (k-1)-itemsets of the frequent k-itemsets must be frequent.

Input: *database D, Mini Support* ϵ , *Mini Confidence* ε

Output: *Rt All association rules*

Method:

1- L1 = large 1-itemsets;

2- for(k=2; $L_{k-1} \neq \emptyset$*; k++) do begin*

3- Ck =apriori-gen(L_{k-1}*); //generate new candidates from* L_{k-1}

4- for all transactions T $\in D$ *do begin*

5- Ct=subset(C_k*,T); //candidates contained in T.*

6- for all candidates C $\in C_t$ *do*

7- Count(C)=Count(C)+1; // increase support count of C by 1

8- end

9- $L_k = \{C \in C_t \mid Count(C) \geq \epsilon \times |D|\}$

10- end

11- $L_f = \bigcup_k L_k$

12 Rt=GenerateRules(L_f*,* ε *)*

Figure 3.3: Pseudo Code for Apriori Algorithm

The Apriori algorithm for finding frequent itemsets is shown in Figure 3.3 , is used to produce Ck, and discarding all itemsets in Cn that do not pass the support threshold. Once these candidate itemsets are identified from Ck, then their supports are incremented. The rule generated that satisfy minimum confidence.

3.2 Algorithm flow in detail

Figure 3.4 depicts the flow diagram of the Apriori algorithm. As shown, the algorithm triggers with scanning all the transactions to get a list of items with support S. As long as the computed support is greater than the user defined minimum threshold support, the item set is taken forward. If it does not meet the minimum requirements, it gets dropped out of contention. Once the item set passes the threshold checks, it gets added to the frequent single item sets. Subsequently a join operation is performed to generate a (k+1) element item set. The threshold checks are performed again on the (k) element item sets and passing it would enable generating a (k) element frequent item set.

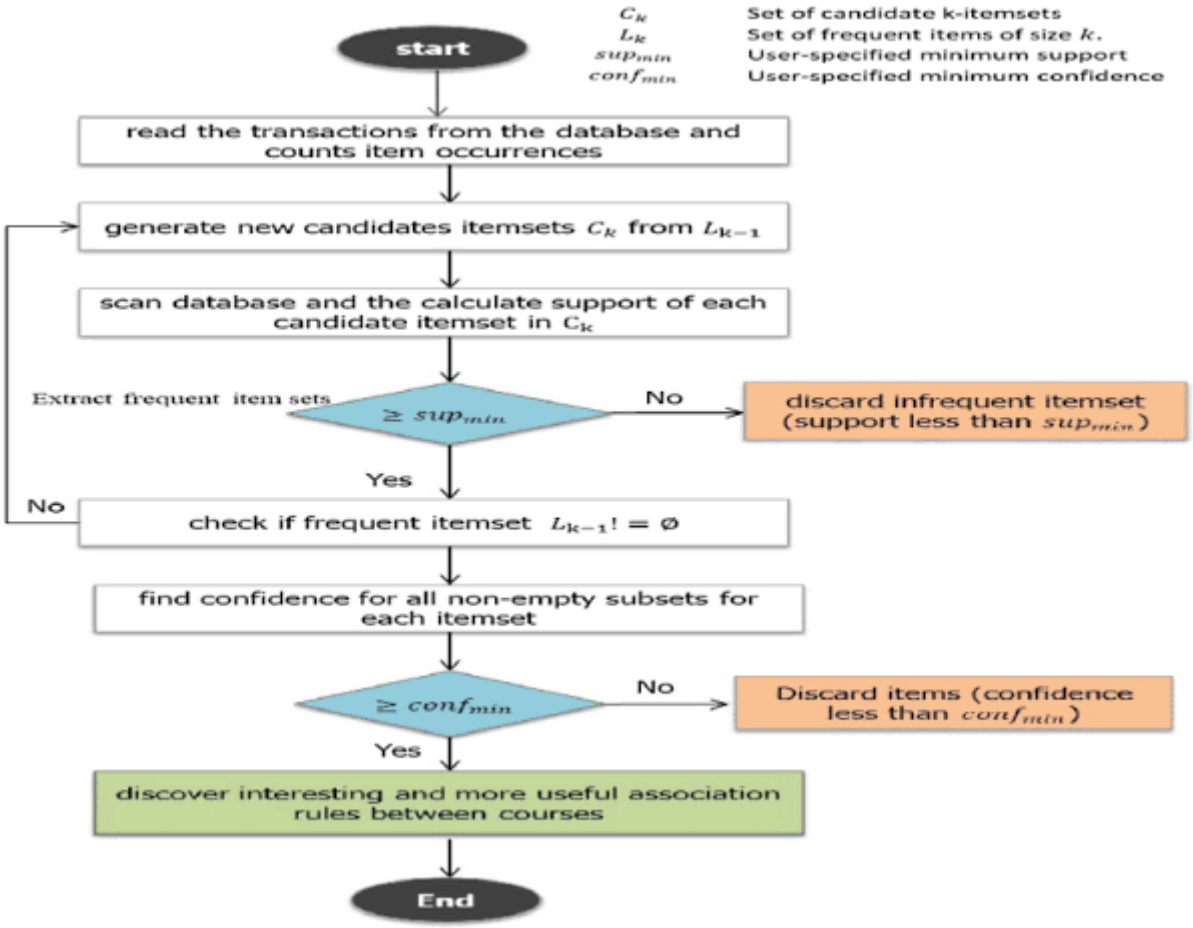


Figure 3.4: Apriori Algorithm Flow Chart

Post this step, if we can generate a (k+1) item set, we do a recursion on the above-mentioned process. However, if we reach a saturation on the item sets, for each frequent item sets gathered so far a complete list of non-empty subsets is gener-

ated. Foreach of these subsets, the confidence C is calculated and ran through a minimum threshold check. Passing this minimum requirement makes it a strong rule, else is discarded. The complete flow is pictorially shown in Figure 3.4 In the next subsections, we will take up an illustration of the algorithm, which also finds place in the developed courseware website for enhanced student learning.

3.3 Market Basket Analysis

Association rule mining searches for interesting relationships among items in a given data set. This paper presents a typical example of association rule technique and highlights the importance of knowledge discovery process in large databases. Considering the example of a store that sells DVDs, Videos, CDs, Books and Games, the store owner might want to discover which of these items customers are likely to buy together. With the information above, the store could strive for more optimum placement of DVDs and Games as the sale of one of them may improve the chances of the sale of the other frequently associated item. The mailing campaigns may be fine tuned to reflect the fact that offering discount coupons on Videos may even negatively impact the sales of DVDs offered in the same campaign. Abetter decision could be not to offer both DVDs and Videos in a campaign. In this case, it is appropriate to use association rule mining to generate the optimum combination of products to increase sales. Apriori algorithm data mining discovers items that are frequently associated together. The Apriori data mining analysis of the 9 transactions above is known as Market Based Analysis, as it is designed to discover which items in a series of transactions are frequently associated together. In this example I considered a database consisting of 9 transactions (table 1)[10].

1. Suppose minimum support count required is 2

2. We have to first find out the frequent itemset using Apriori algorithm;

3. Association rules will be generated using the two parameters minimum support and minimum confidence.

Assume that the dataset of 9 transactions below is selected randomly from a universe of 100000 transactions:

The Apriori algorithm would analyze all the transactions in a dataset for each items support count. Any item that has a support count less than the minimum support count required is removed from the pool of candidate items. First step of this

Transactions	Itemsets
Customer1	BOOKS , CD ,VIDEO
Customer2	CD, GAMES
Customer3	CD ,DVD
Customer4	BOOKS ,CD ,GAMES
Customer5	BOOKS ,DVD
Customer6	CD ,DVD
Customer7	BOOKS ,DVD
Customer8	BOOKS ,CD ,DVD ,VIDEO
Customer9	BOOKS ,CD ,DVD

Table 3.1: An example of database with transactions

algorithm is generating 1-itemset Frequent Pattern. The result is presented in the table 2

{BOOKS}	6
{CD}	7
{VIDEO}	2
{GAMES}	2
{DVD}	6

Table 3.2: Itemsets and their support count

At the beginning of association rule generation process each of the items is a member of a set of first candidate itemsets. The support count of each candidate item in the itemset is calculated (table 3) and items with a support count less than the minimum required support count are removed as candidates. The remaining candidate items in the itemset are joined to create second candidate itemsets each comprise of two items or members.

{BOOKS}	6
{CD}	7
{VIDEO}	2
{GAMES}	2
{DVD}	6

Table 3.3: Itemsets and their support count after condition verified

Step 2 is represented by the generation of 2-itemset Frequent Pattern. After a

combination process the matrix L1 provides the items for C2 matrix. The result is presented in the table 4.

Itemsets
{BOOKS, CD}
{CD , VIDEO}
{VIDEO , GAMES}
{BOOKS , DVD}
{CD , VIDEO}
{CD , GAMES}
{CD , DVD}
{VIDEO ,GAMES}
{VIDEO , DVD }
{GAMES , DVD}

Table 3.4: Itemsets and their support count after condition verified

Each item counts in the database to generate the next matrix, L2 as follow:

Itemsets	Support Count
{BOOKS, CD}	4
{BOOKS , VIDEO}	2
{BOOKS , DVD}	4
{CD , VIDEO}	2
{CD , GAMES}	2
{CD , DVD}	4

Table 3.5: Itemsets and their support count after L2xL2

The next step is to discover the set of frequent 2-itemsets, L2 and the algorithm uses the L1 Join L1 procedure to generate a candidate set of 2-itemsets, C2. The transactions in D are scanned and the support count for each candidate itemset in C2 is accumulated (table 6). The set of frequent 2-itemsets, L2, is then determined, consisting of those candidate 2-itemsets in C2 having minimum support. The following step consists in calculating the support count of each two member itemset from the database of transactions and 2 member itemsets that occur with a support count greater than or equal to the minimum support count are used to generate third candidate itemsets. The steps 1 and 2 are repeated to generate fourth and fifth candidate itemsets, the criteria used to stop this process being the value of support count of all the itemsets.

The next step consists of generation of 3-itemset frequent pattern as is shown in

the table 7. The generation of the set of candidate 3-itemsets, C3, involves use of the Apriori Property . In order to find C3, a L2 Join L2 procedure is used. Join step is complete and prune step will be used to reduce the size of C3. Prune step helps to avoid heavy computation due to large Ck[10].

Itemsets
{BOOKS, CD , VIDEO}
{BOOKS , CD , DVD}
{BOOKS , CD , GAMES}
{CD , VIDEO , GAMES}
{CD , GAMES , DVD}
{CD , DVD , VIDEO}

Table 3.6: Itemsets and their support count after condition verified

Based on the Apriori property [8] that all subsets of a frequent itemset must also be frequent, we can determine that four latter candidates cannot possibly be frequent. Considering this example, lets take BOOKS, CD, VIDEO. The 2-item subsets of it are BOOKS, CD, BOOKS, VIDEO and CD, VIDEO. Since all 2-item subsets of BOOKS, CD, VIDEO are members of L2, we will keep BOOKS, CD, VIDEO in C3. Therefore, C3= BOOKS, CD, VIDEO, BOOKS, CD, DVD after checking for all members of result of Join operation for Pruning. Now, the transactions in D are scanned in order to determine L3, consisting of those candidates 3-itemsets in C3 having minimum support [10].

Itemsets	Support Count
{BOOKS, CD , VIDEO}	2
{BOOKS , CD , DVD}	2

Table 3.7: 3-itemset Frequent Pattern

Step 4 is represented by the generation of 4-itemset frequent pattern.

Itemsets
{BOOKS, CD , VIDEO , DVD}

Table 3.8: 4-itemset Frequent Pattern

The algorithm uses L3 JoinL3 procedure to generate a candidate set of 4-itemsets, C4. Although the join results in BOOKS, CD, VIDEO, DVD, this itemset is pruned since its subset CD, VIDEO, DVD is not frequent in L2.

Thus, C4= 0, and algorithm finishes, having found all of the frequent items. This completes the Apriori Algorithm. All the candidate itemsets generated with a support count greater than the minimum support count form a set of frequent itemsets. These frequent itemsets will be used to generate strong association rules (where strong association rules satisfy both minimum support and minimum confidence).
The final step is to provide the association rules from frequent itemsets. The procedure consists in [6]: For each frequent itemset “l”, generate all nonempty subsets of l; For every nonempty subset s of l, output the rule “s (l-s)” if supportcount(l) / supportcount(s) >= minconf where minconf is minimum confidence threshold. For the example given in this paper, L = BOOKS, CD, VIDEO, GAMES, DVD, BOOKS, CD, BOOKS, VIDEO, BOOKS, DVD, CD, VIDEO, CD, GAMES, CD, DVD, BOOKS, CD, VIDEO, BOOKS, CD, DVD.
For example L = BOOKS, CD, VIDEO. Its all nonempty subsets are BOOKS, VIDEO, BOOKS, CD, CD, VIDEO, BOOKS, VIDEO, CD.

Let minimum confidence threshold is 60%. The resulting association rules are shown below, each listed with its confidence [10].

R1: BOOKS and VIDEO CD
Confidence = scBOOKS,CD,VIDEO/scBOOKS,VIDEO = 2/2 = 100% and R1 is selected.
R2: VIDEO and CD BOOKS
Confidence = scBOOKS,VIDEO,DVD/scVIDEO,CD = 2/2 = 100% and R2 is selected.
R3: BOOKS and CD VIDEO
Confidence = scBOOKS,VIDEO,DVD/scBOOKS,CD = 2/4 = 50% and R3 is rejected.
R4: BOOKS VIDEO and CD
Confidence = scBOOKS,VIDEO,DVD/scBOOKS = 2/6 = 33% and R4 is rejected.
R5: VIDEO BOOKS and CD
Confidence = scBOOKS,VIDEO,DVD/VIDEO = 2/2 = 100% and R5 is selected.
R6: CD BOOKS and VIDEO
Confidence = scBOOKS,VIDEO,DVD/ CD = 2/7 = 28% and R6 is rejected. In this way, we have found three strong association rules

4

DESIGN AND IMPLEMENTATION

1. Global Architecture Of Hisba-BI
    1.1. General Strycture Of The Environment
    1.2. Design
    1.3. Implementation
2. Experimental Results
    2.1. Basic Of Apriori Algorithm

4.1 Global Architecture Of Hisba-BI

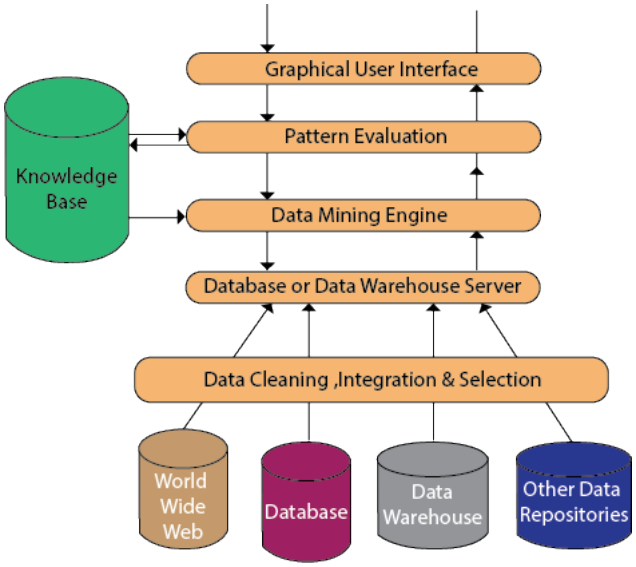


Figure 4.1: Phases in the development of a business intelligence system

Hisba-BI is an extention application of an ERP(Itegrated Management System) Application Called "HISBA" . it Aims to integrate all data Sources Of Melouki Group (data of Hisba ERP) , analyze it and present business data according to Apriori Algorithm Pattern to finds interseting Relations between Products with the intention of making data-driven business decisions and improve Product Sales and Purchases Of The Group .

4.1.1 General Structure Of The Eviroment

We Will talk in this section about the environment in which we develop our program with the mention of programming language used in development and some of the libraries And Packages Imported.

Programing Language

In The Development Of Hisba-BI, We Used The Microsoft C .
C SHARP
C (C SHARP IN BRITISH ENGLISH) Is an Object-Oriented Programming Language Marketed By Microsoft Since 2002 and Intended To Develop On The Microsoft .Net Platform . It is Derived From C ++ and very close to java which it takes the

general syntax and concepts , adding concepts such as operator overload , indexes and delegates.

Figure 4.2: C SHARP Logo

Develompent Environment

We Will Talk about Microsoft Environment And Data Storage Using Microsoft Sql Server.

MICROSOFT SQL SERVER

MICROSOFT SQL SERVER Is a Relational Database Management System (RDBMS) That Supports a wide variety of transaction Processing , Business Intelligence and Analytics in Corporate it Environments .Microsoft Sql Server in ne Of The Three Market-Leading Database Technologies , Along with ORACLE Database And IBM'S .

Figure 4.3: MICROSOFT SQL SERVER Logo

DEVELOPER EXPRESS

This Is A software Development Company Founded in 1998 and Headquartered In Glendale , Calefornia . Devexpress initially started Producing Interface Controls .

Currently DEVEXPRESS Offers Products For Developers Using C , DELPHI / C ++ BUILDER, VISUAL STUDIO AND HTML5 / JAVASCRIPT TECHNOLOGIES.

Figure 4.4: DEVELOPER EXPRESS Logo

ENTITY FRAMEWORK

ENTITY FRAMEWORK Is The FrameworkObject-Relational Mapping (ORM)That Microsoft Makes Available as Part Of The .NET Development . Its Purpose is to Abstract The Ties To a Relational Database , In Such a way the Developer Can Relate To The Database Entity As to a set of Objects and then to classes in addition to their properties.

4.1.2 Design

In This Section We Will Present The Conceptual Aspect Of Our Work Using The Unified Modeling Language (UML) and The Apriori Algorithm That I Used .

Class diagram :
In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among objects [11].

NOTE : In Our Application We Are Just Consuming Database Of Hisba ERP (We Use Just A part Of its Model) .

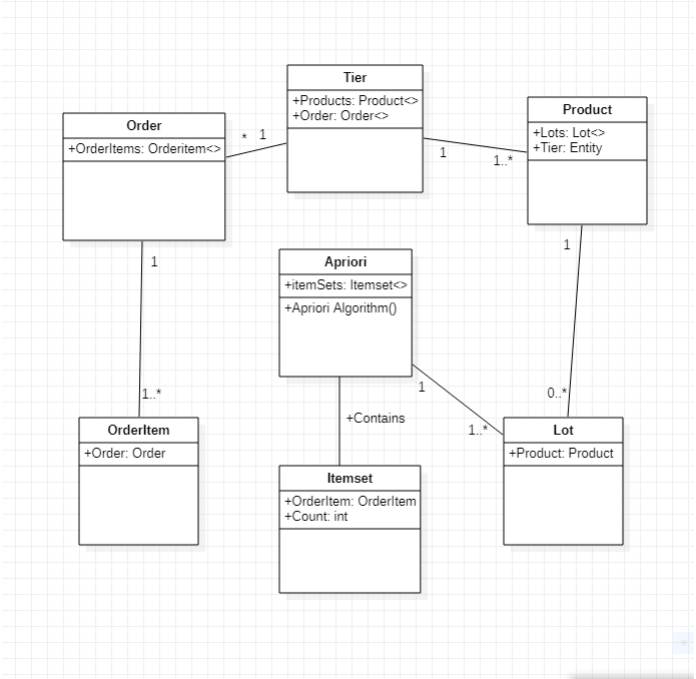


Figure 4.5: Class Diagram

Our Classes Are :

Tier : Tier is an Entity That Conatins Clients and Providers , Concurrent (every Kind Of Persons Job In Application).

```
public class Tiers
{
    [Key]
    public int Id { get; set; }

    public string Code => " " + Id;

    public string Reference { get; set; }

    public string Address { get; set; }

    public string TradeName { get; set; }

    // Cal
    public double Balance { get; set; }

    public int Coef { get; set; }

    public DateTime DateCreation { get; set; }

    public DateTime DateModification { get; set; }

    public string Commune { get; set; }

    public int? CreatorId { get; set; }

    public virtual User Creator { get; set; }

    public int? ModifierId { get; set; }

    public virtual User Modifier { get; set; }

    public virtual ICollection<Order> Orders { get; set; } = new ObservableCollection<Order>();

    public virtual ICollection<Product> Products { get; set; } = new ObservableCollection<Product>();

    public string FullName => TradeName.ToUpper();

    public double OldBalance { get; set; }

    public string TierType { get; set; }
}
```

Figure 4.6: shows Tier Class

Product : Product entity is the Main Entity That Have The Principle Content Of Apriori Algorithm We will Use all Of It Attribute For Constructing a Dashboard That Visualize all Thing Related To it . in Hisba Model each Product Have its Own Lots ; Lot is Our Product That we gone Sale and Apply Test On it.

```
[Key]
public int Id { get; set; }

public string Code => "P-" + Id;

public string Reference { get; set; }

public string Label { get; set; }

public string Label2 { get; set; }

public string Note { get; set; }

public string Description { get; set; }

public DateTime DateCreation { get; set; }

public DateTime DateModification { get; set; }

public int MinQty { get; set; }

public int MaxQty { get; set; }

public int MinQty2 { get; set; }

public int Q { get; set; }

public int Q1 { get; set; }

public int MaxQty2 { get; set; }

public int MaxSaleQty { get; set; }

// Cal
public double PMP { get; set; }

// Cal
public double LastPurchasePrice { get; set; }

public double MinPurchasePrice { get; set; }

public double MaxPurchasePrice { get; set; }

// Cal
public int ReserveQty { get; set; }
```

Figure 4.7: shows Product Class

Lot : Lot is The Sale Entity That we Apply Apriori On it.

```
[Key]
public int Id { get; set; }

public string Code => "P-" + Id;

public string Reference { get; set; }

public string Label { get; set; }

public string Label2 { get; set; }

public string Note { get; set; }

public string Description { get; set; }

public DateTime DateCreation { get; set; }

public DateTime DateModification { get; set; }

public int MinQty { get; set; }

public int MaxQty { get; set; }

public int MinQty2 { get; set; }

public int Q { get; set; }

public int Q1 { get; set; }

public int MaxQty2 { get; set; }

public int MaxSaleQty { get; set; }

// Cal
public double PMP { get; set; }

// Cal
public double LastPurchasePrice { get; set; }

public double MinPurchasePrice { get; set; }

public double MaxPurchasePrice { get; set; }

// Cal
public int ReserveQty { get; set; }
```

Figure 4.8: shows Lot Class

Order : Order Entity is The Transaction (Delivery Note) That we sale it , its Variables are all about saling . it have a List Of Order Items; OrderItem is The Entity That Gone be Sale in The Order(Lot).

```
public  class Order
 {
     [Key]
     public int Id { get; set; }

     public string Code { get; set; }

     public string Reference { get; set; }

     public string AssociatedRef { get; set; }

     public string OtherRef { get; set; }

     public DateTime Date { get; set; }

     public DateTime DateCreation { get; set; }

     public DateTime DateModification { get; set; }

     public string Note { get; set; }

     public double Weight { get; set; }

     public double Volume { get; set; }

     public bool IsLocked { get; set; }

     // Cal
     public double Margin { get; set; }

     // Cal
     public double Charge { get; set; }

     public double AmountHT { get; set; }

     public double NetAmountHT => AmountHT - Discount;

     public double NetAmountTTC { get; set; }

     public double NetToPay { get; set; }

     public double TotalTVA { get; set; }

     public double DiscountPercentage { get; set; }
```

Figure 4.9: shows Order Class

OrderItem : OrderItem is The Lot That Gone be Saled .

```
public class OrderItem
{
    [Key]
    public int Id { get; set; }

    public string Code => " " + Id;

    public string Reference { get; set; }

    public string Label { get; set; }

    public double Charge { get; set; }

    public double Weight { get; set; }

    public double Volume { get; set; }

    public double Discount { get; set; }

    public double DiscountPercentage { get; set; }

    public double UPriceHT { get; set; }

    public double UPriceTTC { get; set; }

    public double UNetPriceHT => UPriceHT * (1 - DiscountPercentage / 100);

    public double UNetPriceTTC => UPriceTTC * (1 - DiscountPercentage / 100);

    public double AmountHT => UPriceHT * Qty;

    public double NetAmountHT => UNetPriceHT * Qty;

    public double AmountTTC => UPriceTTC * Qty;

    public double NetAmountTTC => UNetPriceTTC * Qty;

    public double TotalTVA => NetAmountHT * (TVAPercentage / 100);

    public double TVAPercentage { get; set; }

    // Cal
    public string Margin { get; set; }

    public int Qty { get; set; }
```

Figure 4.10: shows OrderItem Class

Sequence diagram : Sequence Diagram shows object interactions arranged in time sequence . it depicts the objects involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario . sequence diagrams are typically associated with use case realizations in the logical view of the system under development [11].

Data Preparation Sequence Diagram

in The Data Preparation Of Sequence Diagram Our sequence of events is , the Data Preparation begin by Hisba-Bi System Request For Customer Informations in The Interface Of Apriori Algorithm (Apriori System) Then it Connect (Hisba) Database Then Ouur Filtering System Witch Is A group Of ethods and Queries That Retreive Information From Database , Eliminate Nulls , Delete UnUsed Odres , All Operations of data Cleaning Then After Data Cleaning We Rebuild Our Model and insert it in our Database and Rereive a useful information and pass it to apriori System.

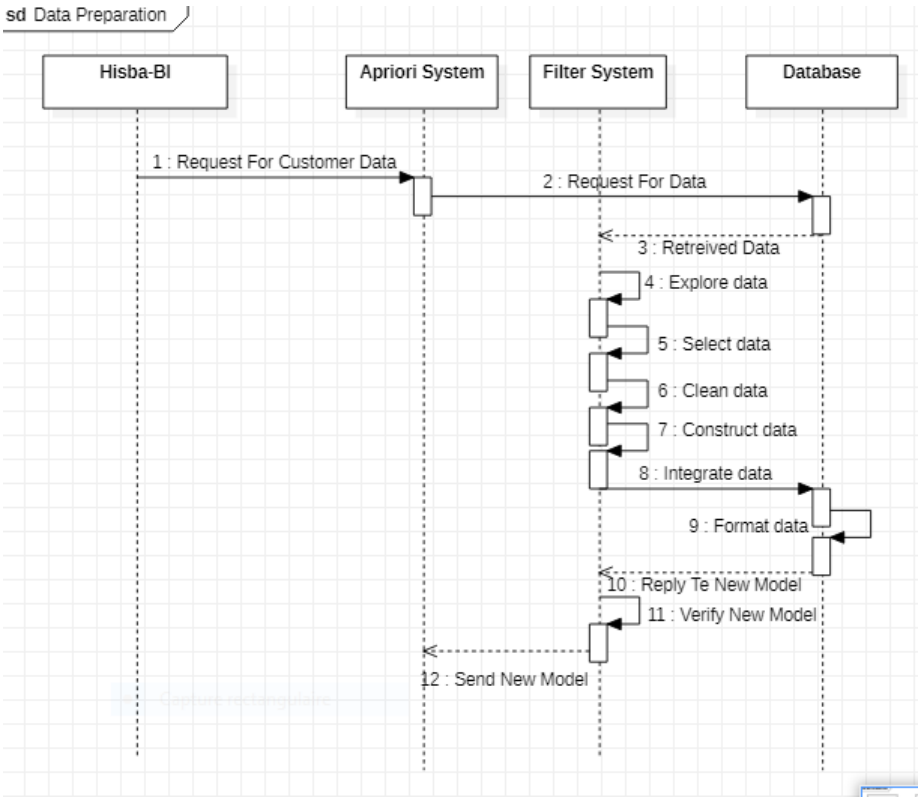


Figure 4.11: Data Preparation Sequence Diagram

Data Modeling Sequence Diagram

in The Data Modeling Of Sequence Diagram Our sequence of events , the Data That Have been Prepared and Cleaned (The New Model That We got) We pplicate Apriori Algorithm On it The Generate Test Design To Build Our Last Model (Association Rules) And Evaluate Our results .

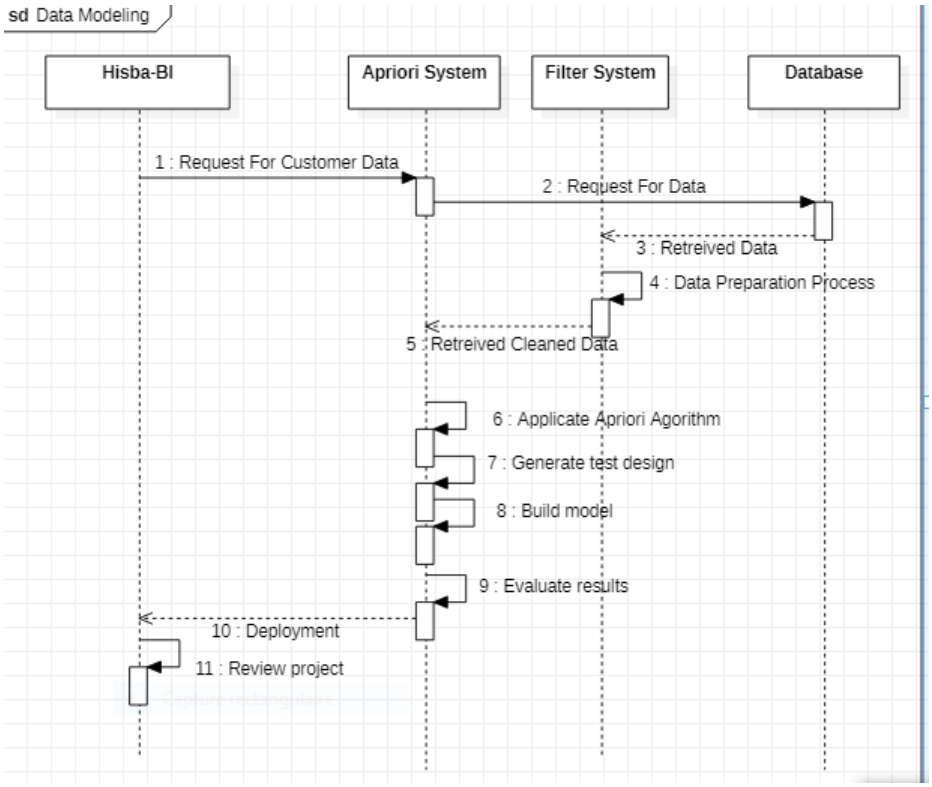


Figure 4.12: Data Modeling Sequence Diagram

4.1.3 Implementation

In The Implementation of Hisba-BI we will Pass by a Steps Process in order to get a Propre Data That We can use in Apriori Algorithm and Then generate A Correct Rules .

1. **Data Collection :** we collect largest Database From Different Companies with multiple domains ParaPharmaceuticals , Medicale Pieces ...
   In our Application we Consume Two Databases ParaharmStreet Company Dataset and Techniclab Company Data Set.
2. **Data Preprocessing:** some preprocessing in Datasets is performed. It includes applying Nulls removal, and all Methods Of Data Cleaning.
3. **Design the Apriori Algorithm :** we build the Apriori algorithm .
4. **Implementation of the algorithm :** we implement the proposed Algorithm using C Sharp programming language .

Data Collection : Data collection is a systematic process of gathering Data From Databases . In Our Application We Start By Collecting Access Files That Contains Archives Of The ERP Applicaion "Hisba" Thats Mean Collecting initial data and load it into your analysis tool(Sql Server).

In Our Data Collection We use two Methods For The Collect :

Microsoft SQL Server Migration Assistant (SSMA) :

is a tool designed to automate database migration to SQL Server from Microsoft Access, DB2, MySQL, Oracle, and SAP ASE [15].

Steps Microsoft SQL Server Migration Assistant (SSMA) :

1. Create a new SSMA project.

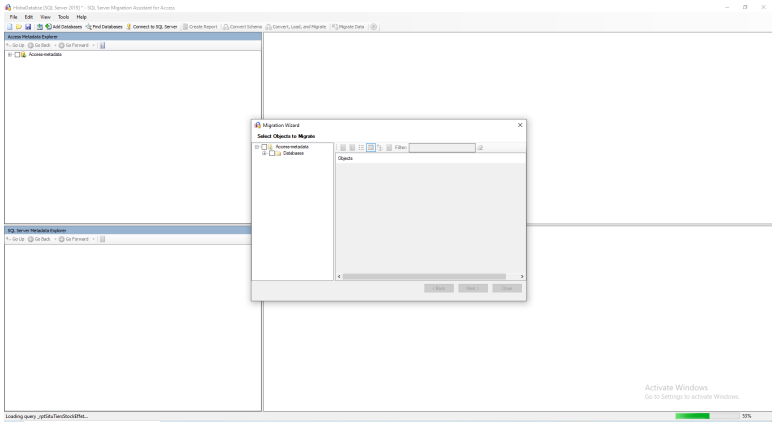


Figure 4.13: Create a new SSMA project

2. Connect to an instance of SQL Server.
3. Map Access Database schemas to SQL Server database schemas.
4. Convert Access database into SQL Server schemas.

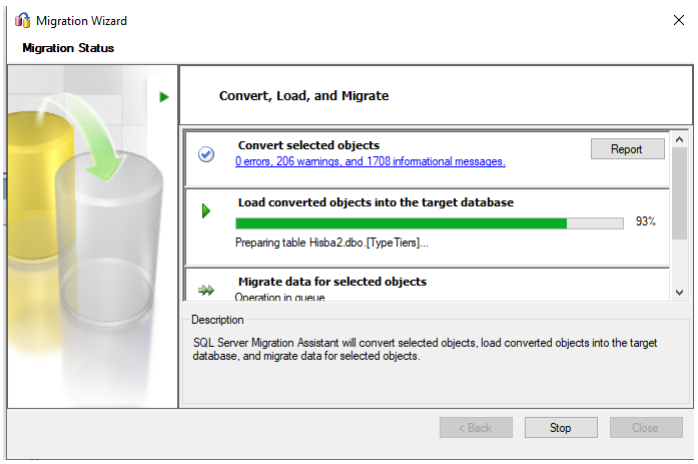


Figure 4.14: Convert Access database into SQL Server schemas

5. Load the converted database objects into SQL Server.

6. Migrate data to SQL Server.

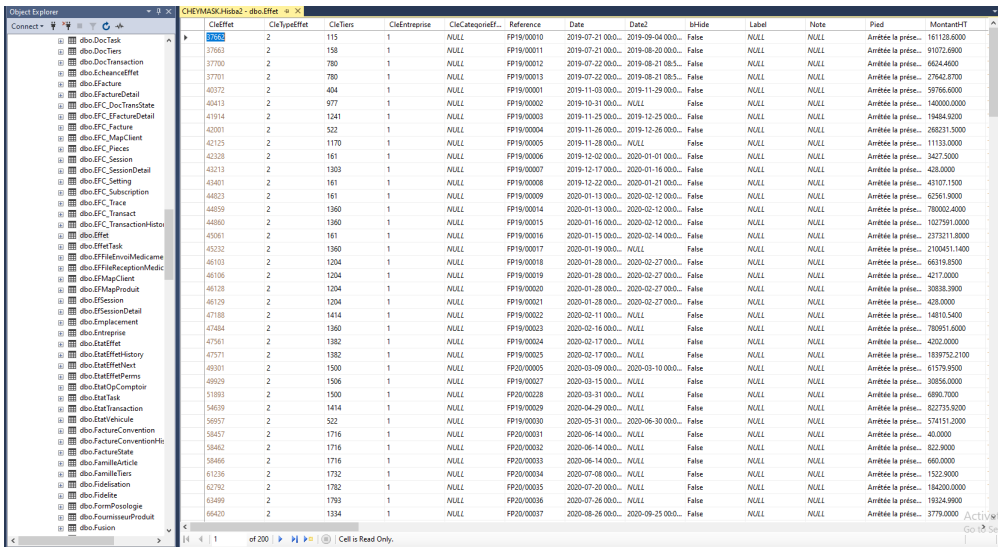


Figure 4.15: Migrate data to SQL Server

7. Link Converted Tables in Sql Server.

Implement an interface That Collect Data in Hisba-BI :

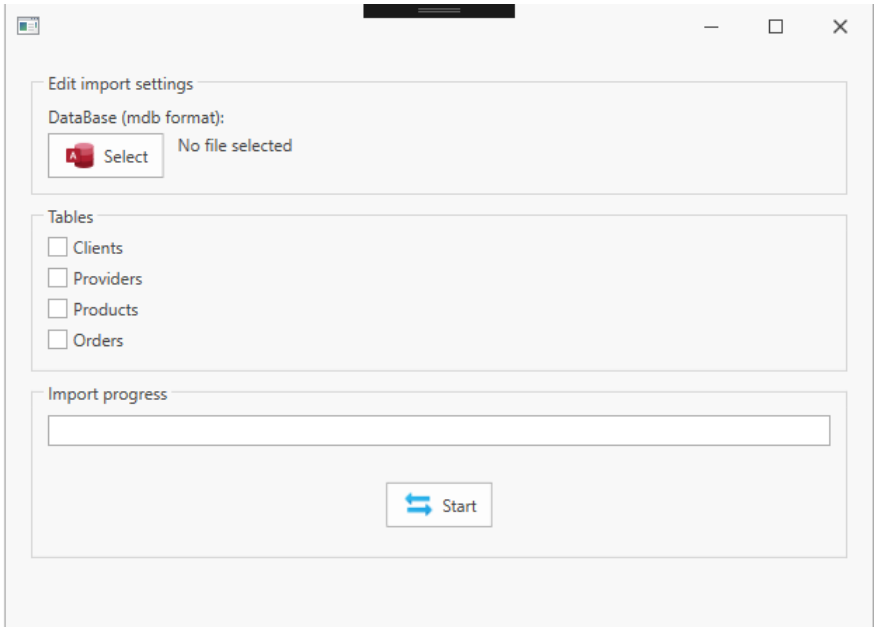


Figure 4.16: Show Access Database Import

This interface Shows The Access Databse and migration to sql server Process at First We Select Our Database Then we select witch tables we want to import Then we Start , at this process we Start The Second Step Of Our Implementation witch its Data Preprocessing.

Data Preprocessing: at This Phase We start our data Cleaning ; Often this is the lengthiest task. Without it, we will likely fall victim to garbage-in, garbage-out. A common practice during this task is to correct, impute, or remove erroneous values By A methods That we Incude In Our Queries . [12]. Then We Construct Our new Data ; Derive new attributes that will be helpful Then we Integrate our Data .

Design the Apriori Algorithm : at its shows page 41 we use This PseudoCode To Design The Apriori Algorithm and Also According To its page 42 chart Flow We Will Implements is Code .

Implementation of the algorithm :

In figure 3 an Uml schema over the final implementation is illustrated **??** . The first thing we need to do is to load the data from the database and represent it in some manner in our application Hisba-BI. like we have done in this :

Tiers Interfaces : Tiers Interface Their is Two Interfaces That Show Us The List Of Clients and Providers. Note : all Names of Cilents and Prociders in our Application are not real.

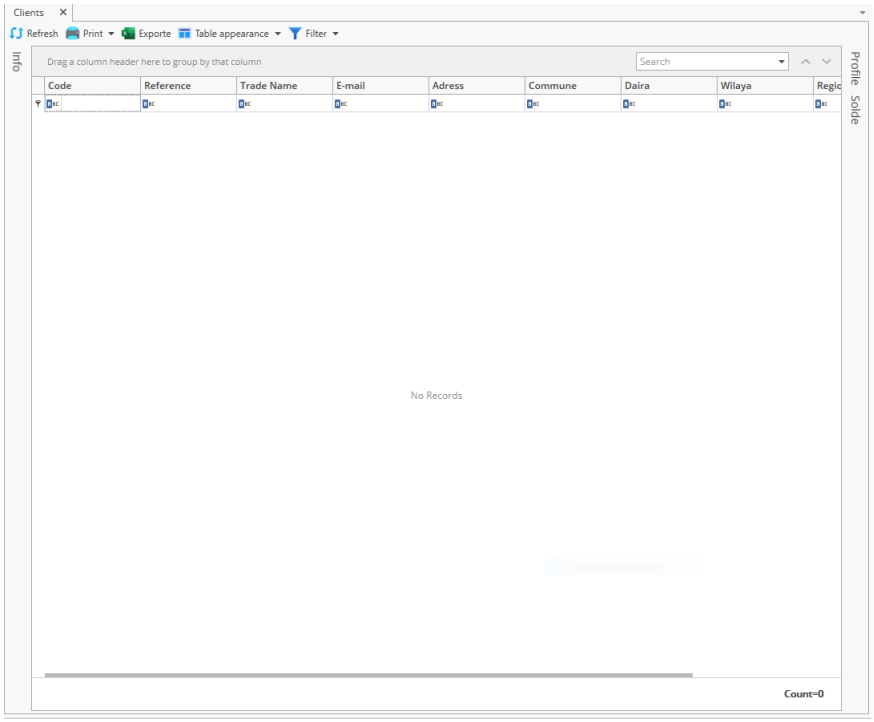


Figure 4.17: Show Client Interface

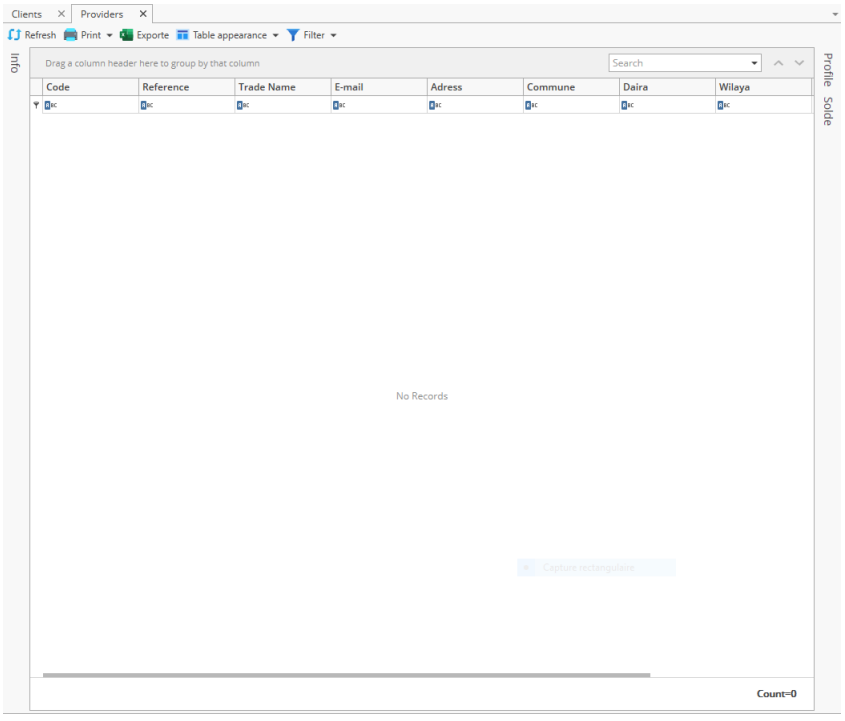


Figure 4.18: Show Provider Interface

Product Interface :

This Interface Shows The Product List Used In Hisba-BI after Data Cleaning.

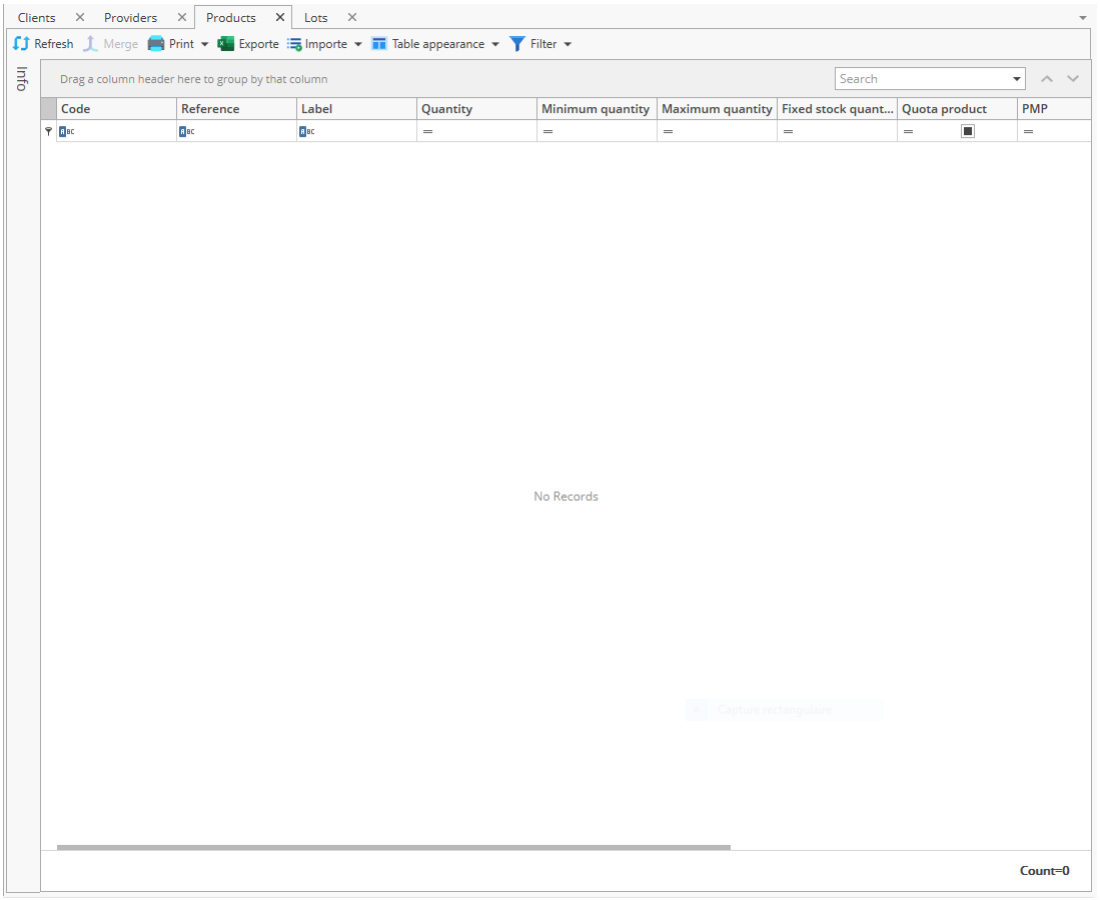


Figure 4.19: Show Prodcut Interface

Client Dashboard :

This interface Shows Client Dashboard That Helps Decision Making Person In The Company To Take Their Decision Easly By Showing Thm The Client State .

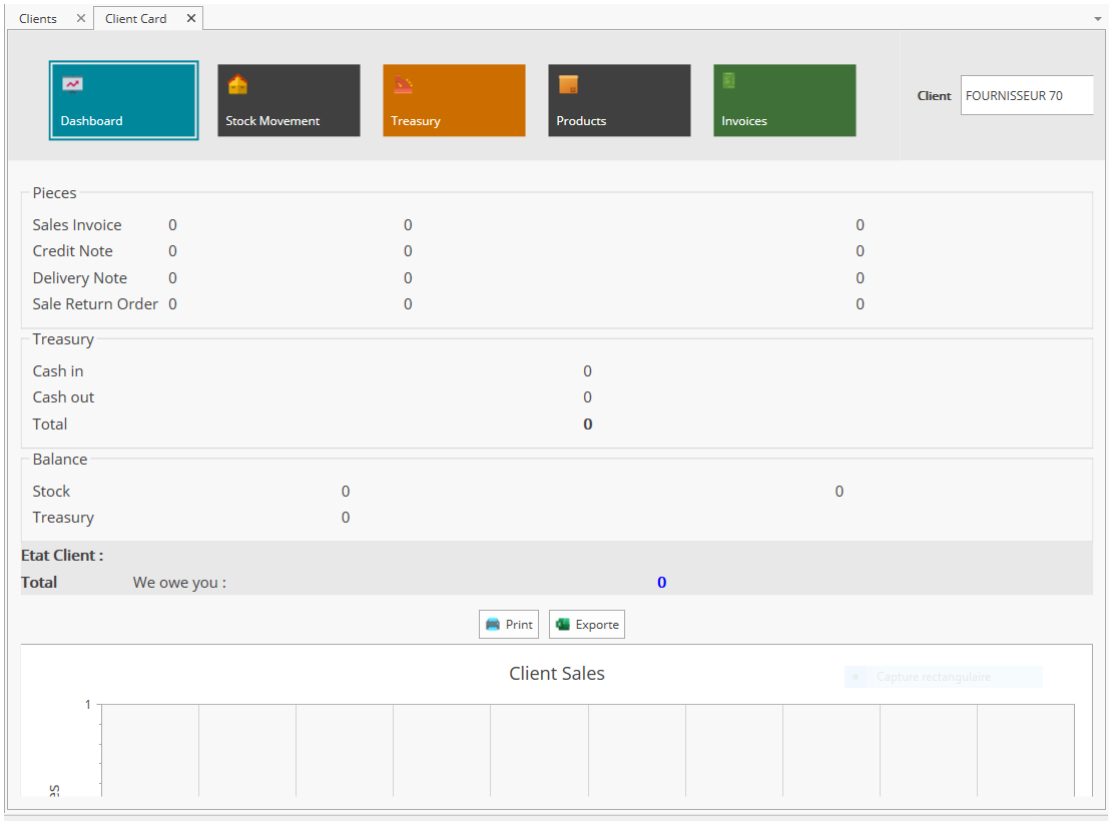


Figure 4.20: Show Client Dashboard

To Connecte To Database We Need To enter Our Server IPAddress we Implement a Splash Screen That Can Connecte Database :

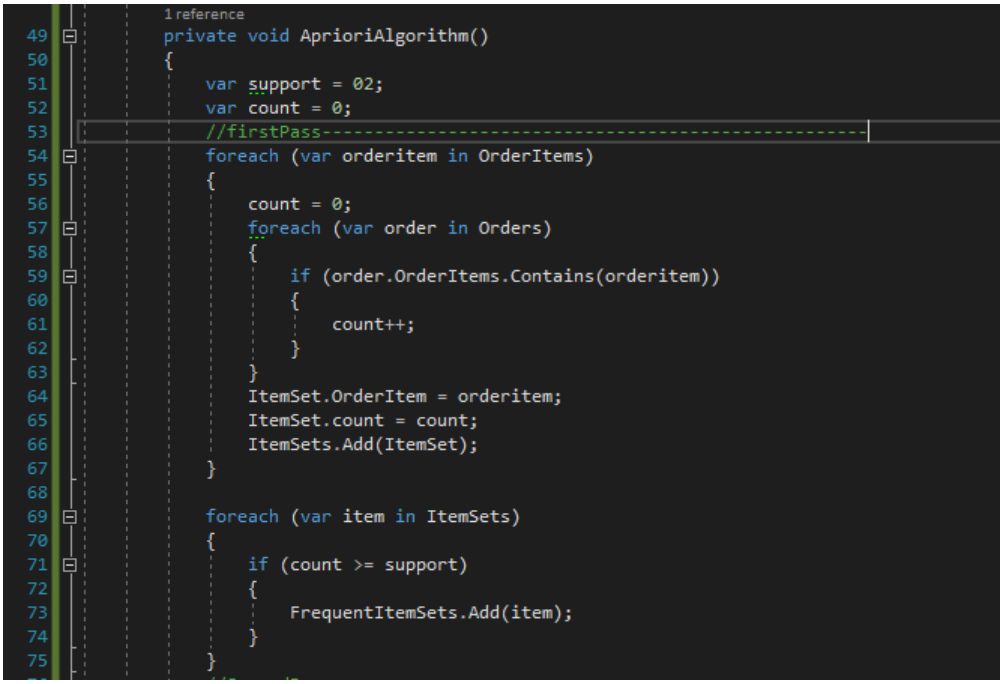
Figure 4.21: Show Hisba-BI Splash

Our Splash Screen Tried To Connect To Hisba Database if There is No Previous Connection String it will Pass To Figure 4.22 at This Figure We enter The Type Of

Figure 4.22: Show Hisba-BI Splash Login Via Ip Adress

Connection , Our Server Name , Username and Pasword if its Sql Server Authentication.

Then We need To iterate through the whole transaction table and group the Order Id by corresponding Order Item Id, Then we count each individual term in Order since only OrderItems that meets the support threshold are needed for consecutive iterations of the algorithm :



```
1 reference
private void AprioriAlgorithm()
{
    var support = 02;
    var count = 0;
    //firstPass-----------------------------------------------------
    foreach (var orderitem in OrderItems)
    {
        count = 0;
        foreach (var order in Orders)
        {
            if (order.OrderItems.Contains(orderitem))
            {
                count++;
            }
        }
        ItemSet.OrderItem = orderitem;
        ItemSet.count = count;
        ItemSets.Add(ItemSet);
    }

    foreach (var item in ItemSets)
    {
        if (count >= support)
        {
            FrequentItemSets.Add(item);
        }
    }
```

Figure 4.23: Apriori Algorithm First Pass

In this second step of the algorithm it takes the large item sets That Have been generated and generates all possible 2-sets that can be combined. In later steps it will take all large 2-set item sets and generate all possible 3-sets, all large 3-set item sets to generate all possible 4-sets and so on. and it Continue To iterate and in each iterate we count the support and generate new item set .

```
//SecondPass-------------------------------------------------------
var Pair_Count = 0;

foreach (var orderitem01 in FrequentItemSets)
{
    foreach (var orderitem02 in FrequentItemSets)
    {
        if (PairItemset.Contains(orderitem01) && PairItemset.Contains(orderitem02)) { }
        else
        {
            PairItemset.Add(orderitem01);
            PairItemset.Add(orderitem02);
            Pair_Count = PairItemset.Sum(item => item.count);
            PairItemsets.Add(new InternalOrderitems
            {
                PairOrderItems = PairItemset,
                PairCount = Pair_Count
            });
        }
    }
}

foreach (var pair in PairItemsets)
{
    if (pair.PairCount >= support)
    {
        ThirdItemsets.Add(pair);
    }
}

//ThirdPass-------------------------------------------------------
```

Figure 4.24: Apriori Algorithm Second Pass

4.1.4 Experiments and Results

we present and analyze the experimental results to show and test if Apriori algorithm can enhance execution time, speedup, and generate strong association rules in terms of support .

Experimental Setup

We implement Apriori Algorithm using C Sharp programming Language language and tested on a computer with Core I5 processor and 8 GB of memory.

Experimental Results and Evaluation

We use the collected Data Contains of 6335 Product and 10341 Lot , 41200 Order and 1650 Tiers . We evaluate the performance of the Apriori with respect to the execution time and Support.
Execution Time To measure the execution time, we have executed the Apriori algorithm with Support 4 on a Our system . We also have used different number of testing datasets to observe the effects of different problem on the performance. 2 sets where used with the number of 6335 Product and 10341 Lot , 41200 Order and 1650 Tiers and second set with 600 Product , 820 Lot , 1000 Order and 500 Tiers. Our specific execution time values for various runs are shown in Table 4.1

Apriori Accuracy	Runtime (ms)			
	Order Number	Product Number	Tiers Number	Lot Number
	41200	6635	1650	10341
ParaPharm Dataset	1000			
	1000	600	500	820
Techniclab Data set	400			

Table 4.1: The Execution Times (sec.) of Apriori ALgorithm

Support Support is an important measure (Section 3) because a rule that has very low support may occur simply by chance. The smaller the minimum support threshold is, the more frequent itemsets there will be, so the execution time will increase along with the decrease of the minimum support threshold. The number of frequent items increases quickly along with decrease of minimum support threshold. Table 4.1.4.2 shows the execution time for different support values.

Support	Excution Time (sec)
6	1200
10	800

5

GENERAL CONCLUSION

5.1 General Conclusion

The Work On Business Intelligence an its Tools is One Of The Motivational an improvement areas , In Our Memory we have introduced Hisba-BI an Application That Use In general Buusiness Intelligence , DataMining and apply Apriori Algorithm To Geerate An Associations Rules That Will Help in Decision Making. In Our Memory We Have Shown General Introduction About Business Intelligence and Data Mining , Data warehouse Then Association Rules And Apriori Algorithm That We Apply To get to Our Last Goal Witch Is Enhance Sales Performance at group Melouki Witch Man A Better Decion Making.

Making Our Application Full Of all Business Intelligence Tools is one Of The Thing we will be working in the Future witch is Considered as Perspective , wr try to add all Tasks and Algorithms of Data Mining Like Classification(To determine witch Client we Buy and not) , Estimation , Clusternig Also Other Advanced Tools Of Business Intelligence.

One of the key business intelligence future trends many experts are predicting is a growth of the digital business intelligence world into a space where tools and platforms will become more broad-spectrum and eventually, more collaborative.

Bibliography

[1] https://catalogimages.wiley.com/images/db/pdf/0470511397.C01.pdf.

[2] Business Intelligence and Data Mining ,Anil K. Maheshwari, PhD 2015.First published by Business Expert Press, LLC.

[3] Introduction to Business Intelligence : evolution of BI.

[4] Introduction to Business Intelligence : evolution of BI,Literature Review of business intelligence rasemy heang and raghul mohan school of business and engenineering halmstad university , sweden.

[6] Introduction to Business Intelligence : evolution of BI.

[7] The Definitive Guide to Business Intelligence Better Guys.

[8] data and business Analytics mark ferguson , Editor Mark Ferguson, Editor anil K.Maheshwari , Ph.D.

[9] R. Rakesh Agrawal, "Fast Algorithms for Mining Association Rules in Large Databases," Computer Science and Technology, vol. 15, pp. 487-499, 1994.

[10] Universitatea Petrol-Gaze din Ploieşti, Bd. Bucureşti 39, Ploieşti, Catedra de Informatică

[11] https://fr.wikipedia.org/wikia

[12] https://olap.com/learn-bi-olap/olap-bi-definitions/business-intelligence

[12] https://www.datascience-pm.com/crisp-dm-2/

[13] S.Q. Zaho, "Association Rule Mining: A Survey," anyang Technological University , 2003.

[14] T. L. Rakesh Arrawal, "Mining Association Rules Between Sets of Items in Large Databases," in SIGMOD, 1993.

[15] https://docs.microsoft.com/en-us/sql/ssma/sql-server-migration-assistant?view=sql-server-ver15

[16] https://www.upgrad.com/blog/common-data-mining-algorithms/

[17] https://www.melouki.com/

Abstract :

The Work On Business Intelligence an its Tools is One Of The Motivational an improvement that exploit the available data to generate information and useful knowledge for complex decision-making processes.

In Our Dissertation we have introduced Hisba-BI an Extension Application That Use an ERP And CRM Management Datastes To Extract From it knowledge That Help in Decision Making , Hisba-BI Use In general Buusiness Intelligence , DataMining and apply Apriori Algorithm To Generate An Associations Rules That Will Help in Decision Making also Hisba-Bi Have a Statistical Parts (Dashboard) That Visualize Data in a better way to Responsible Of The Company.

In end of this dissertation, we have developed a Desktop Application , that can Consume Data and Visualize it by Using Charts , Diagram and Generate Association Rules Using The Apriori Algorithm.

Key Word: Business Intelligence System, Data Mining Application , Apriori Algorithm Application , Data Science.

ملخص :

يعد العمل على ذكاء الأعمال وأدواته أحد التحسينات التحفيزية التي تستغل البيانات المتاحة لتوليد المعلومات والمعرفة المفيدة لعمليات صنع القرار المعقدة.

في أطروحتنا ، قدمنا تطبيقًا ملحقًا يستخدم تخطيط موارد المؤسسة لاستخراج المعرفة التي تساعد في اتخاذ القرار ، واستخدام ذكاء الأعمال بشكل عام ، واستخراج البيانات ، وتطبيق الخوارزمية لإنشاء قواعد الجمعيات التي ستساعد في اتخاذ القرار. الأجزاء (لوحة القيادة) التي تصور البيانات بطريقة أفضل لتكون مسؤولاً عن الشركة.

في نهاية هذه الرسالة ، قمنا بتطوير تطبيق سطح مكتب يمكنه استهلاك البيانات وتصورها باستخدام المخططات والرسم التخطيطي وإنشاء قواعد الارتباط باستخدام خوارزمية Apriori.

الكلمة الرئيسية: نظام ذكاء الأعمال ، تطبيق استخراج البيانات ، تطبيق خوارزم ابريوري

Abstrait :

Le travail sur la Business Intelligence et ses outils est l'une des améliorations motivantes qui exploitent les données disponibles pour générer des informations et des connaissances utiles pour les processus de prise de décision complexes.

Dans notre thèse, nous avons présenté Hisba-BI une application d'extension qui utilise un ERP et des données de gestion CRM pour en extraire des connaissances qui aident à la prise de décision, Hisba-BI utilise en général l'intelligence économique, l'exploration de données et applique l'algorithme Apriori pour générer des règles d'association. Cela aidera à la prise de décision et Hisba-Bi disposera d'une partie statistique (tableau de bord) qui visualise les données d'une meilleure manière pour le responsable de l'entreprise.

À la fin de cette thèse, nous avons développé une application de bureau, qui peut consommer des données et les visualiser en utilisant des graphiques, des diagrammes et générer des règles d'association à l'aide de l'algorithme Apriori.

Mot clé : Système de Business Intelligence, Application d'exploration de données, Application d'algorithme Apriori, Data Science.

.