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

FACULTE: Mathematics and Informatics
DEPARTEMENT: computer science
N°: ……..……..…..

DOMAINE: Mathematics and Informatics
FILIERE: Computer Science
OPTION: Information Systems and Software Engineering

A Dissertation in Fulfillment
for the Requirements of the Degree of Master

By: Abdoune Souad

Entitled

A Data Mining System for Customer Relationship Management

Case study: Géant Electronics Company (BBA)

Presented publicly to the jury:

……	University of M'sila	President
Brahimi Mahmoud	University of M'sila	Supervised
……	University of M'sila	Examiner

Academic Year: 2019 /2020

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

FACULTE: Mathematics and Informatics
DEPARTEMENT: computer science
N°: ……..……..

DOMAINE: Mathematics and Informatics
FILIERE: Computer Science
OPTION: Information Systems and Software Engineering

A Dissertation in Fulfillment
for the Requirements of the Degree of Master

By: Abdoune Souad

Entitled

A Data Mining System for Customer Relationship Management

Case study: Géant Electronics Company (BBA)

Presented publicly to the jury:

……	University of M'sila	President
Brahimi Mahmoud	University of M'sila	Supervised
……	University of M'sila	Examiner

Academic Year: 2019 /2020

Dedication

I dedicate this work:

To the candles of my life; my mother and my father for their support, and advice and I wish them good health and long life.

To my sisters and my brothers.

To my family members.

To all my teachers.

To all my friends.

Acknowledgements

First of all, we would like to express our gratitude to almighty Allah for giving us strength and ability to understand, learn and write this modest research.

I would like to express my sincere gratitude to my supervisor Dr. Brahimi Mahmoud for the continuous support, his patience, motivation, and immense knowledge.

Big thanks to my parents for their continued support during my studies.

Many thanks Mr. Saidani Tahar and all the employees of Géant Electronics Company for their help and their trust.

I would like to thank all the professors of the Computer science department for their valuable advice and encouragement.

I would like to thanks all the people who have helped me directly or indirectly to the realization of this work.

Thank you very much.

Content

Chapter 2: Presentation of the host organization

Chapter 3: Analysis and design

Chapter 4: Implementation and operation

List of figures:

GENERAL INTRODUCTION

General introduction

Nowadays, the large number of companies and the abundance of products generate a great competition in the market; also the dominance of offer to request has part in increasing the competition between firms, which oblige them to go to several methods for controlling the market and acquiring customers.

Customer attention is the more interested point of all the companies which search for success and continuity, and this is done by fulfill customers' demands, listen to their concerns, exploit great potentials and provides competitive services (in the price, payment modes, approximation of merchandise, duration of grantee and more) to gain their satisfaction. Thanks to the scientific and technological development that has touch in all fields, such problems have been researched to find solutions capable of treating them. Perhaps the most prominent of them is CRM.

CRM stands for Customer Relationship Management. This expression was used for the first time in the early of 1990s, which is a famous system and business strategy that uses technology to manage the communication between companies and its clients to optimize long term benefits and help them to achieve their goals.

Alongside good customer management; data mining techniques have been shown to be effective in increasing the understanding of the market and in ensuring a good vision on the decision-making level. The smooth running of the CRM is supported by the good quality and the high accuracy of information provided by the data mining tasks.

In this project we take a vivid example of Géant Electronics Company, which is a leading economic institution in the electronics industry in Algeria. This latter is characterized by a variety of high-quality products and to maintain its position in the market, especially with its many competitors, this company always seeks to develop its strategies and modernize its management by relying on technological methods and modern technologies. Our goal behind this study is to develop a CRM application and make it work as parallel as the Enterprise Resource Planning (ERP) of Géant Electronics Company.

The present dissertation is divided into four chapters which are organized as follows:

The first chapter presents the basic concepts of the business intelligence and data mining in order to show the complementary work that can be established between them to better understand and manage the market.

The second chapter is subdivided into two parts. The first is devoted to the presentation of the host organization and the second presents the basic concepts related to CRM which interests us in this project.

Then, in the third chapter, we approach the conceptual work of our application using the UML modeling language.

The last chapter is devoted to the implementation part of the work where we presented the different used tools and the technical environment of the development with of course the carried out application.

Finally, we will end this dissertation with a conclusion and some suggestions for future improvements.

CHAPTER 1

BUSINESS INTELLIGENCE AND DATA MINING

1. Introduction

It is not enough for a company to store a multitude of data in a specialized database, data Warehouse or big data, it is also necessary to exploit them in an efficient way. This is essentially the role of Data Mining and Business Intelligence fields. The first one is used for extracting insights contained in this mass of data which is far too large to be satisfied with classical statistical tools. However, Business Intelligence refers to all of the technologies that allow companies to analyze these voluminous data for the benefit of their decision-making.

In this chapter, we will present the key concepts of these two fields which generally work together for the satisfaction of business leaders.

2. What is Business Intelligence?

Business Intelligence or BI is a set of concepts, methods and technologies which are used to treat and present data for analysis operation and decisions- making. In other meaning Business Intelligence allows the collection, the organization, and data storage also extraction and restitution of pertinent information [1].

3. Reasons for business Intelligence

BI enables companies to make informed business decisions by improving the timelines and quality of information. BI reveals [**2**]:

- The position of the company in comparison to its concurrent.
- Changes in customer behavior and spending patterns.
- The capabilities of the organization.
- Market condition, future, trends, demographic and economic information.
- The social, regulatory, and political environment.
- What the other firms in the market are doing.

4. Components of Business Intelligence

Business Intelligence is founded around many components including [**3**]:

4.1. On-Line Analytical Processing (OLAP)

This technology stands behind the majority of BI applications. OLAP analyzes business information in a multi-dimensional manner. The difference between OLAP software and the other analysis tools is the ability to store data in multidimensional database structure.

4.2. Corporate Performance Management (CPM)

It includes methods, metrics, activities, and systems which are used to monitor and manage the business performance of the company. CPM is an important component of BI for companies that are searching for such changes like: budget remodeling, cost cutting, upgrading organization strategy and improving the process of financial planning.

4.3. Real Time BI (RTBI)

It is used for sorting and analyzing business data. RTBI allows the company to get a vision about the business process as quickly as possible to take strategic action.

4.4. Data Warehouse

Data warehouse allows the business aimer to go through different data subset and examine components that could help making the best business decisions. Data warehouse is different from a simple database; it is created to help analyze a large volume of data in a short time. Its goal is to receive and process data in a short time.

4.5. Data source

To integrate the business intelligence process and methodology, it is important to have the right understanding of data source.

Companies trend to store huge amounts of operational data, by collect them from many source like:

- Enterprise resource planning: that stores a large amount of transaction data.
- Customer Relationship Management (CRM): that store and analyze customer behavior and its important information.
- E-Commerce applications: that offer real-time seals activity.
- There are other sources like: company database, flat file, web services…etc.

5. Business Intelligence architecture [4]

Business Intelligence architecture is composed of the following principal components:

5.1. Data warehouse

It is a general database of company that is used to store all the historical data from all the departments of enterprise. The data warehouse presents itself as the heart of the system in view of its importance. Data warehouse is a database designed especially to facilitate data analysis and reporting operations.

5.2. Extract Transform Load (ETL)

There are many sources of data for business intelligence. So the consolidation operation which consists in extracting data from their sources, transforming them and then pouring them in the data warehouse. This operation is called Extract Transform Load (ETL).

5.3. Data model

Data warehouse is designed to be the source of analysis and reports, but data a warehouse isn't fast to be able to cover all needs. The requirement for faster processing and lower response time in part and aggregated information on another part leads to creation of another layer in business intelligence systems. This layer is called data model and it is divided into two technologies which are:

5.3.1. OLAP cube

OLAP is a file-based data storage that loads data from a data warehouse in a cube model, which contains descriptive information.

5.3.2. Table Memory Model (Tabular Model)

Table Memory Model based on a new engine in memory for tables. The memory engine loads all the data stored in tables and answers for questions directly from memory. It's very fast in terms of response time.

The model must provide the following features.

5.4. Data visualization

Data visualization is a part of the business intelligence system that users can see.

5.5. Master Data Management (MDM)

MDM is the process of maintaining a single, consistent version for all entities of database across multiple systems.

5.6. Data Quality Service (DQS)

DQS works according to knowledge base domains, which means a knowledge base can be created for different domains, and the knowledge base will be maintained and improved by a data steward according to its evolution over time. There are corresponding policies that can be used to enforce data normalization (figure 1.1).

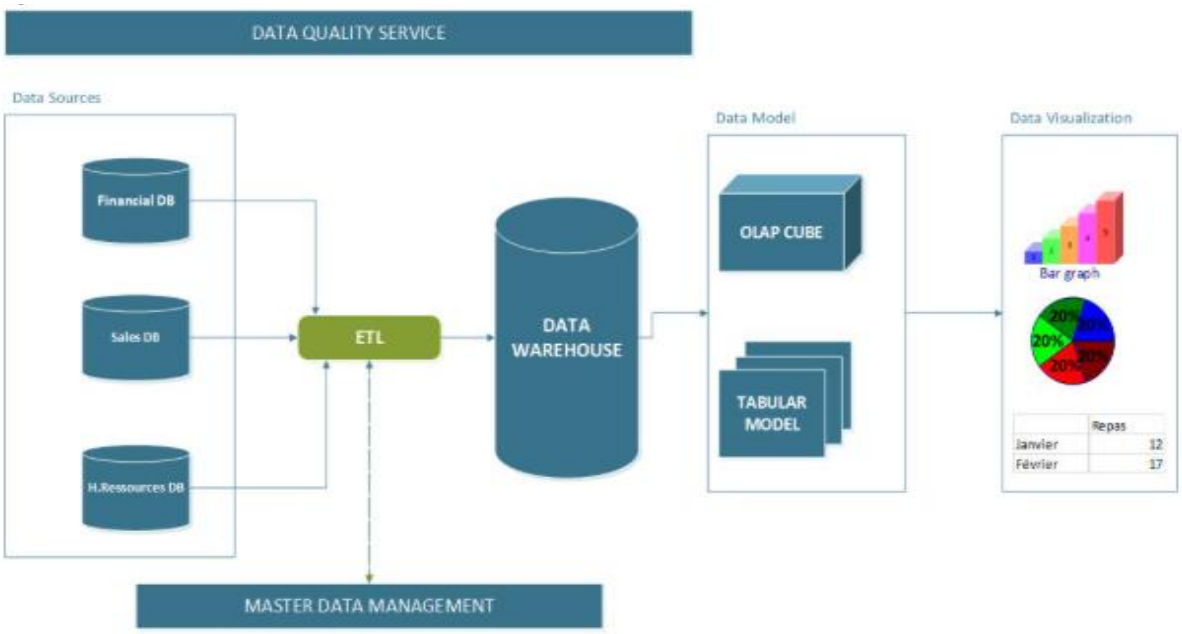


Figure 1.1: Business intelligence architecture

6. The steps of decision-making project [5]

A decision making project consists of 4 steps which are collection phase, integration phase, organizational phase, and restitution phase.

Figure 1.2: Decision making steps

6.1. Collection phase

This phase uses data that can be presented in different format:

- It can be flat files like: XML files, ASCII files…etc.
- Database systems like: MYSQL, ORACLE…etc.

6.2. Integration phase

This step is necessary to manipulate data sources using ETL (Extract, Transform, and Load) and ODS layers.

6.3. Organizational phase

This step allows the storage of data in a warehouse called a data warehouse.

6.4. Restitution phase

In this phase we use several tools for the restitution of the result like: reporting and querying tools, Analysis tools, and data mining phase.

7. Business Intelligence tools [6]

There are many tools for business intelligence but the best of them are: Microsoft Power BI, Tableau Desktop, Dun das, Sisense, and Zoho analytics.

8. Business Intelligence applications [7]

Business Intelligence can be applied in several fields such as:

8.1. Transporting Industry

Generally BI offers a lot of tools and platforms which enable the delivery of information to decision makers such as: On Line Analytical Processing (OLAP), reporting tools, data visualization tools, and data mining tools.

8.2. Banking Industry

There are many BI platforms to rely them with bunking industry to make more effective decisions in a few areas such as: customer analysis, operation and financial analysis, sales and marketing analysis, promotion analysis, fraud and risk analysis.

8.3. Health Care Industry

There are many BI platforms designed to provide users with simple and relevant data like: number of patients, treatment needed and their hospitalization period.

8.4. Retail Industry

BI is implemented for demand forecasting in the retail industry by generating reliable estimates for both short term and long term demand based on the available customer data.BI also used to monitor customer loyalty by evaluating which customers are loyal and predict which are want to leave.

8.5. Manufacturing Industry

BI systems allow Manufacturers to track their inventory usage by using alerts that send notification of low inventory levels. BI systems allow Manufacturers to analyze data from many sources to determine performance goals and sophisticated profitability and financial models.

8.6. Pharmaceuticals Industry

BI systems help pharmaceuticals to monitor customer behavior and identify which products are most profitable.

9. Business Intelligence benefits [8]

Business Intelligence guarantees many potential benefits such as:

9.1. Productivity

BI facilitates creation of reports so saving a lot of time and resources.

9.2. Visibility

BI helps in improving the visibility of such business processes which makes it possible to analyze the important area.

9.3. Accountability

BI takes care of the accountability factors to perform as per the business goals.

9.4. Feature

By means of BI, decision makers get full control of the business through dashboards and scorecards.

9.5. Analytics

BI eliminates all the complex processes and offers predictive analysis.

Therefore, the benefits listed below can generate other benefits which are shown in the following figure:

Figure 1.3: Benefits of business intelligence

10.What is Data mining?

Data mining is to analyze a huge amount of data in a data warehouse to find new relations that are impossible to identify without this mean [**9**]. In other meaning data mining is the extraction of useful information from a huge amount of data [**10**].

11.Knowledge Data Discovery(KDD) process [11]

KDD is the automatic extraction of patterns from large volumes of data; data mining is a step of knowledge discovery. In general, a knowledge discovery process consists of the following steps:

11.1. Data cleaning

Data cleaning allows removing noisy, erroneous, missing, or irrelevant data.

11.2. Data integration

Data integration allows integrating many data sources in a common source.

11.3. Data selection

Data selection allows choosing the data relevant to analysis.

11.4. Data transformation

Data transformation allows transforming the selected data into forms appropriate for mining procedure.

11.5. Data mining

Apply data mining algorithms to extract data patterns.

11.6. Pattern evaluation

In this step, strictly interesting patterns representing knowledge are identified based on given measures.

11.7. Knowledge presentation

This step uses visualization techniques to help users to understand the data mining result.

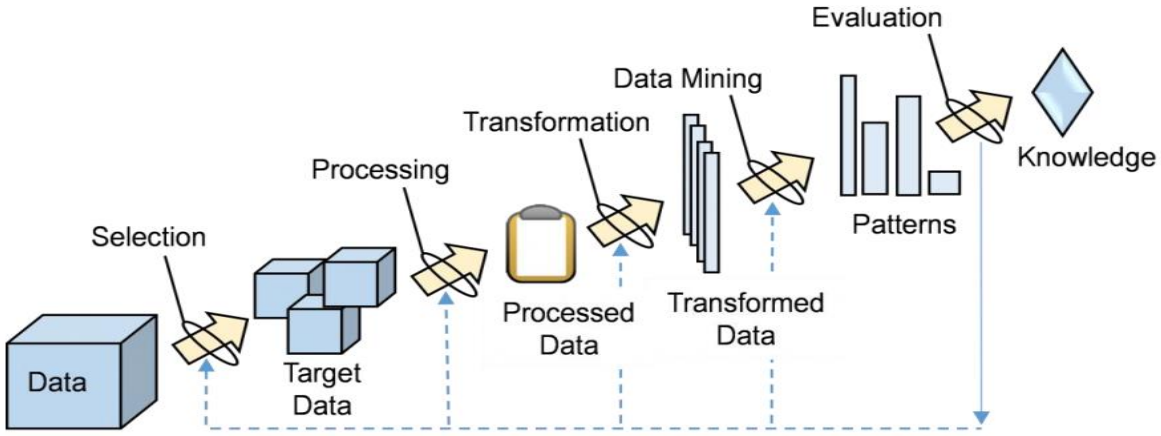


Figure 1.4: KDD process

12.Major tasks of data mining [12]

Generally, data mining tasks are classified into two categories: predictive task and descriptive task. The main goal of the first one is to predict the value of a particular attribute based on the other values; however the second one is to identify patterns that summarize relationships of data.

Figure 1.5: Data mining tasks

12.1. Classification

Classification is one of the most important tasks in data mining. This technique aims to organize data in a given classes [**13**]. In other meaning classification is a task that maps a data item into one set of predetermined classes [**14**].

12.1.1. Classification process

Classification process is divided into two steps; the first one is model construction, in this step the model is built from some training data set [**15**]. Model can be presented as:

classification rule, decision tree or mathematical formula1 [**14**]. The second step is testing the model, in this step the model is used to classify unknown tuples into class labels [**15**].

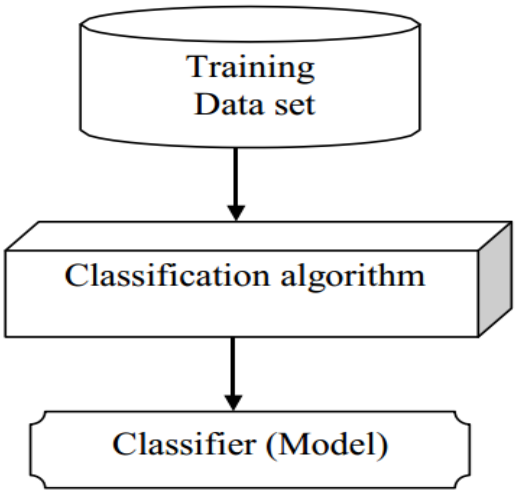


Figure 1.6: Model construction phase

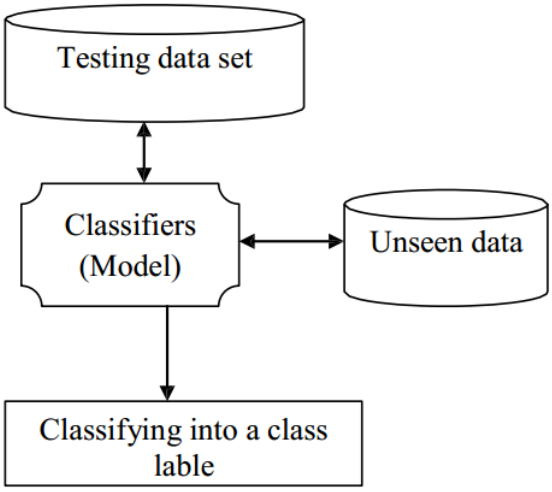


Figure 1.7: Testing model phase

12.1.2. Classification algorithms

There are several classification algorithms such as:

12.1.2.1. Decision Tree

Decision tree is one of the most effective algorithms in data mining; it is used in several domains because of its simplicity and robustness. The decision tree consists of nodes that form a rooted tree, meaning it is a direct tree with a node called root that has no incoming edges. Nodes with outgoing edges are called terminals or test nodes [**16**].

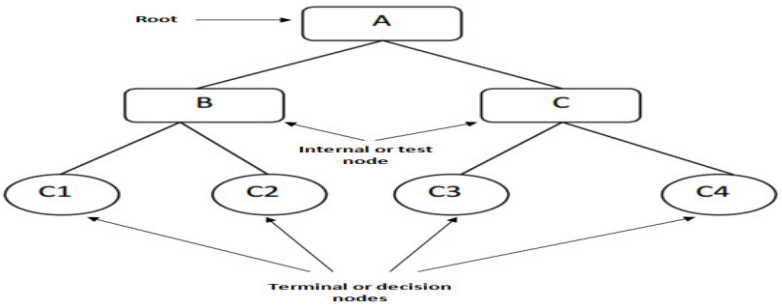


Figure1.8: Decision tree structure

Basically, the generation of tree decision consists of two phases: tree construction and tree pruning.

Tree construction [**17**]

- At the beginning all the training examples are at the root.
- If samples are all the same class then node becomes a leaf labeled with class label.
- Else select the attribute that has the best separates samples into individual classes.
- Repeat the two last steps.

- Stop if there are no samples with attribute value or samples in nodes belonging to the same class.
- There are different algorithms to choose the attribute to split the dataset such as: information gain (ID3/C4.5) and gini index.

Tree pruning [**17**]

- The constructed decision tree may have many branches or leaf nodes, this caused by noisy information and over-fitting trees.
- To avoid the problem, remove tree branches that may reflect noise in training data and leads to error when classifying test data.

12.1.2.2. Random Forest (RF) [**18**]

To avoid the problem, remove tree branches that may reflect noise in training data and leads to error when classifying test data. Random forest is a classifier that evolves from decision trees. It actually consists of many decision trees. To classify a new instance, each decision tree provides a classification for input data; random forest collects the classifications and chooses the most voted prediction as the result. The input of each tree is sampled data from the original dataset. In addition, a subset of features is randomly selected from the optional features to grow the tree at each node. Each tree is grown without pruning. Essentially, random forest enables a large number of weak or weakly-correlated classifiers to form a strong classifier.

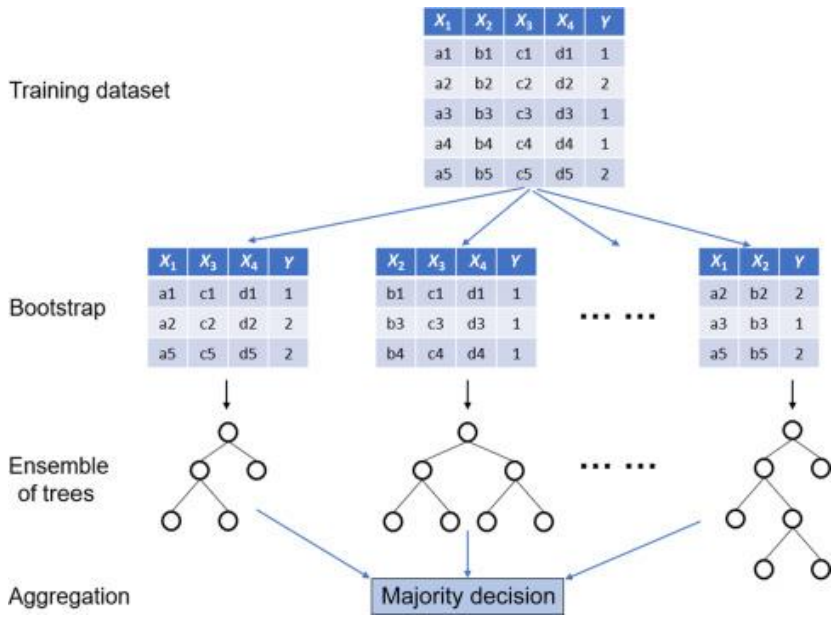


Figure 1.9: Implementation of RF classifier

12.1.2.3. Naïve Bayes [**19**]

Naïve Bayes is a statistical classification technique based on Bayes theorem. Naïve Bayes classifier assumes that the effect of a particular feature in a class is independent of other features. Naïve Bayes model is easy to build and useful for very large datasets. This is the Bayes theorem:

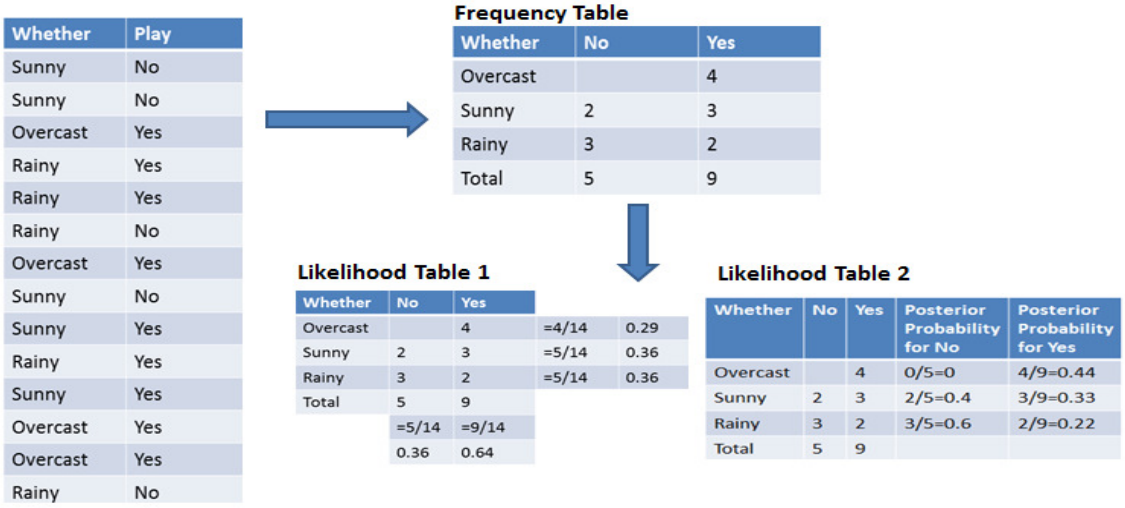
$$p(h|D) = \frac{p(D|h).p(h)}{p(D)} \qquad (1)$$

Where:

- P (h): the probability of hypothesis h being true (regardless of the data). This is known as the prior probability of h.
- P (D): the probability of the data (regardless of the hypothesis). This is known as the prior probability.
- P (h|D): the probability of hypothesis h given the data D. This is known as posterior probability.
- P (D|h): the probability of data D given that the hypothesis h was true. This is known as posterior probability.

The Naïve Bayes algorithm can be summarized in the following steps:

- Calculate the prior probability for given class labels.
- Find likelihood probability with each attribute for each class.
- Put these values in Bayes formula and calculate posterior probability.
- See which class has a higher probability, given the input belongs to higher probability class.



Whether	Play
Sunny	No
Sunny	No
Overcast	Yes
Rainy	Yes
Rainy	Yes
Rainy	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rainy	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table

Whether	No	Yes
Overcast		4
Sunny	2	3
Rainy	3	2
Total	5	9

Likelihood Table 1

Whether	No	Yes		
Overcast		4	=4/14	0.29
Sunny	2	3	=5/14	0.36
Rainy	3	2	=5/14	0.36
Total	5	9		
	=5/14	=9/14		
	0.36	0.64		

Likelihood Table 2

Whether	No	Yes	Posterior Probability for No	Posterior Probability for Yes
Overcast		4	0/5=0	4/9=0.44
Sunny	2	3	2/5=0.4	3/9=0.33
Rainy	3	2	3/5=0.6	2/9=0.22
Total	5	9		

Figure 1.10: Naïve Bayes example

12.1.2.4. Neural networks (NN) [**15**]

Neural networks or Artificial Neural Networks (ANN) is a mathematical or computational model based on biological nervous systems. It has a multiple interconnected group of artificial nerves. A neural network is used for classification and pattern recognition. An ANN changes its structure and adjusts its weight to minimize error. The operation of adjusting of weight is based on external and internal information that flows through the network during the learning phase.

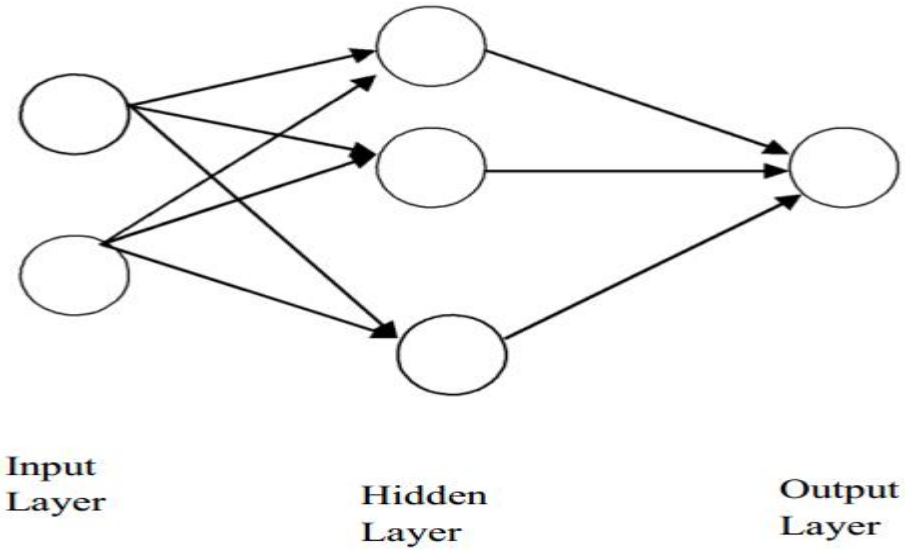


Figure 1.11: Neural Network structures

12.1.2.5. K-Nearest Neighbor (KNN)

KNN is one of the simplest classification algorithms; its purpose is to use a database in which the data points are separated into several classes to predict the classification of new sample point. KNN algorithm is based on similarity measure [**20**].K in KNN is a parameter that refers to the number of nearest neighbors to include in the majority of the voting process. Similarity is defined according to distance metric between points [**21**]. There are different distance functions such as:

Euclidean $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$

Manhattan $\sum_{i=1}^{k}|x_i - y_i|$

Minkowski $\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q}$

Figure 1.12: Some distance functions

The KNN algorithm can be summarized in the following steps [**22**]:

- Specify the value of K.
- Calculate the distance between test data and each row of training data.
- Sort the calculated distance in ascending order.
- Get the most frequent class of these rows.
- Return the predicted class.
- Repeat the last four steps from one to total number of training data points.

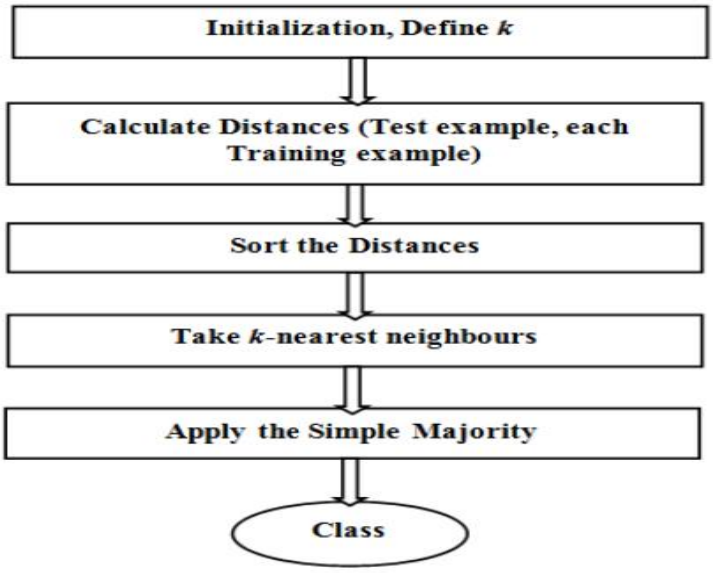


Figure 1.13: KNN algorithm steps

12.2. Regression [23]

Regression is a statistical method used in a lot of disciplines and domains that attempt to determine strength and character between one dependent variable and series of independent variables.

12.2.1. Regression methods

There are various kinds of regression methods available to make predictions. The most commonly used of them are:

12.2.1.1. Linear Regression

Linear regression is the simplest regression technique. Basically linear regression method use the formula of straight line (**y = m x +b**) to predict the relationship between two factors. The factor that is being predicted is called dependent variable; the factors that are used to predict the value of the dependent variable are called independent variables [**24**]. There are two types of linear regression:

- Simple linear regression (SLR): in this type there are only one dependent variable and one independent variable [**25**]. The equation of SLR is:

$$y = \beta_0 + \beta_1 x + \varepsilon \qquad (2)$$

Where y is response variable, x is the regressor variable and is not a random variable, ε is an error introduced to account for the randomness in data.

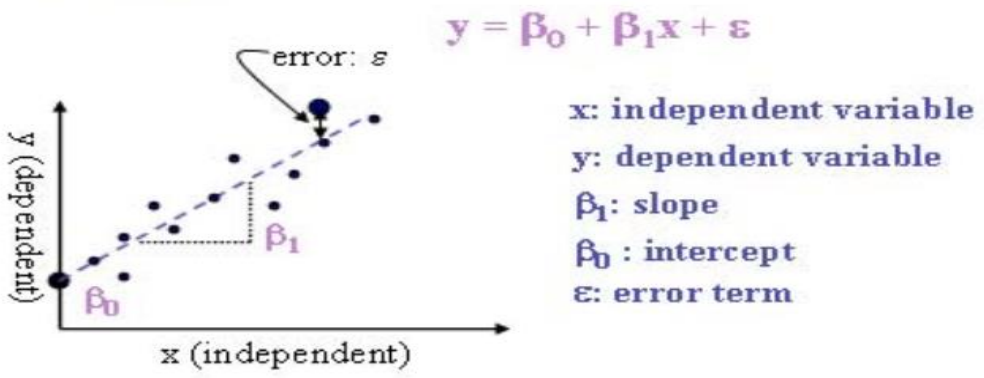


Figure 1.14: Simple linear regression representation

- Multiple linear regressions (MLR) [**25**]: in this method the dependent variable is one and there are multiple independent variables. This is the equation that represents this method:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \qquad (3)$$

12.2.1.2. Logistic Regression [**25**]

Logistic regression is the second popular form of regression after linear regression; it is preferred when the dependent variable is binary. This method is used in the classification of objects, it has one output that can be true or false and there can be a single or multiple inputs.

12.2.1.3. Polynomial Regression [**26**]

Polynomial is a technique to fit a nonlinear equation by taking the independent variable. The equation of this kind is:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_n x^n + \varepsilon \qquad (4)$$

12.2.1.4. Ridge Regression

Ridge regression is a method that helps to understand the regularization concept to solve fitting problems [**26**]. Regularization is appropriate for models that fail on testing data and pass on training data [**25**].

12.2.1.5. Multivariate Regression [**25**]

Multivariate regression is different from multiple regressions; it has multiple dependent variables and multiple independent variables. The dependent variables are in different formulae and it has more than one independent variable to predict the value of **y** s.

12.3. Clustering [27]

Clustering is a technique that divides the given data items into classes called clusters where any class must have similar data items.

12.3.1. Clustering applications

Clustering is mainly used to segment or classify a database (for example to sort customer data such as age, profession, place of residence, etc., to optimize customer relationship management) or to extract knowledge in order to identify data subsets that are difficult to identify with the classical methods. Therefore, Clustering can be used in various fields such as:

- Bioinformatics to identify similar genes.
- Marketing to segment customers in order to establish market research.
- Image processing for object recognition.
- Automatic Language Processing (LAT) for document classification.

12.3.2. Clustering algorithms

Generally, the clustering algorithms can be divided into five categories which are:

12.3.2.1. Partitioning clustering algorithms [**27**]

The partitioning algorithms divide the dataset into various partitions where each partition represents a cluster. The resulting clusters have the following characteristics:

- Each cluster must have at least one object.
- Each object must be a part of exactly one cluster.

The commonly used partitioning clustering algorithms are:

- K-means: is one of the simplest algorithms that solve the well-known clustering problem. It divides data items into k-clusters which are similar between them and dissimilar to the objects belonging to another cluster [**28**]. The K-means algorithm is implemented in four steps which are [**29**]:

1) Given K, Partitioning object into K non empty subset.

2) Compute seed point as the center (mean point) of the cluster of the current partition.

3) Assign each object to the cluster with the nearest seed point.

4) Go pack to step 2, stop when no more new assignment.

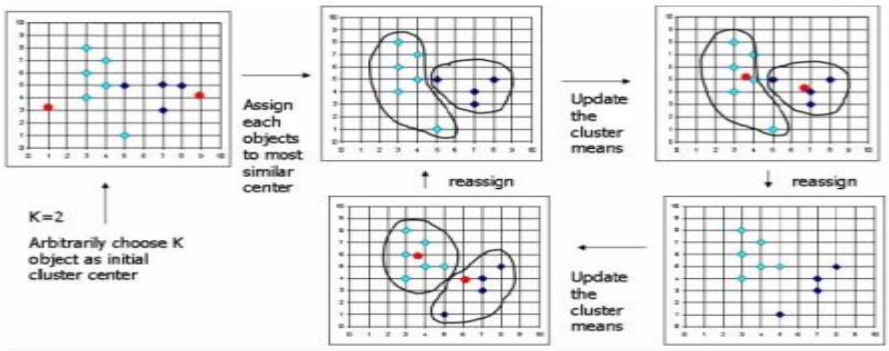


Figure 1.15: K-means example

- K-medoids: it is one of the important methods of clustering for partitioning a dataset into k-clusters [**30**]. The process of k-medoids contain this steps: firstly, randomly select k of the n data points as the medoids, after this associate each data points to the closet medoid, then for each medoid m and for each non medoid data point o swap m and o and compute the total cost of the configuration, then select the configuration with the lowest cost. Finally, repeat the three last steps until there is no change in medoids [**28**].

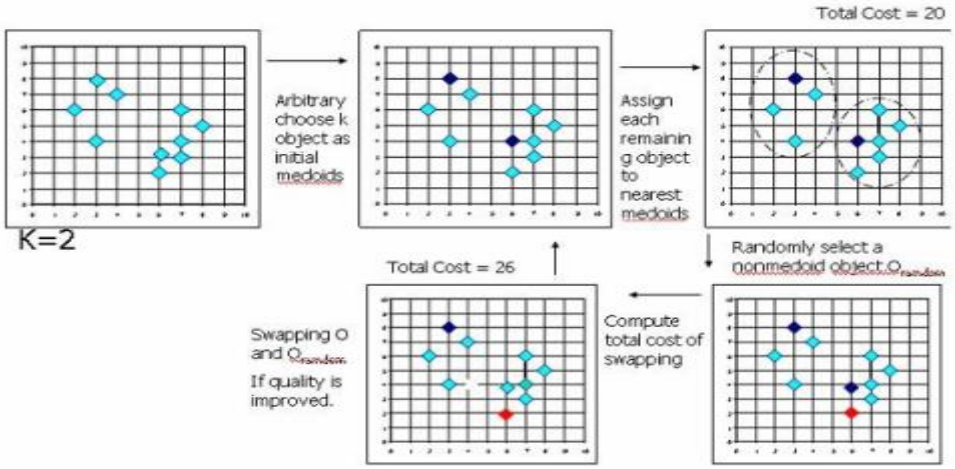


Figure 1.16: K-medoid example

12.3.2.2. Hierarchical clustering algorithms [**31**]:

These methods construct the clusters by recursively partitioning the instances in either a top-down or bottom-up fashion. This method is divided to two types are:

- Agglomerative (bottom-up): at the beginning each object represents a cluster, and then merges clusters until the desired cluster structure is obtained.
- Divisive (top-down): at the beginning all the points in one cluster and in each step split a cluster until each cluster consists of only one data instance.

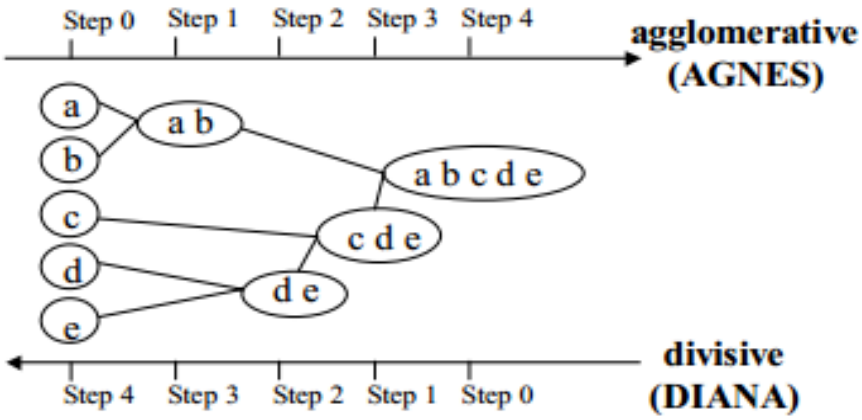


Figure 1.17: Hierarchical algorithm

12.3.2.3. Density-based algorithms [**31**]

These methods assume that the points that appertain to each cluster are drawn from a specific probability distribution. The goal of this method is to discover clusters of arbitrary shape which are not necessarily convex. Some of common density-based clustering techniques are [**32**]:

- Density-Based Spatial Clustering of Application with Noise (DBSCAN): that discovers high density regions in DBSCAN and creates clusters out of them.
- Varied Density Based Spatial Clustering of Application with Noise (VDBSCAN) [**32**]: this algorithm can detect clusters with varied density.
- Density-Based Algorithm for Discovering Density Varied Clusters in large spatial database (DVBSCAN): this algorithm handles local density variation within the cluster.
- A Distribution-Based Clustering Algorithm for Mining Large Spatial Database (DBCLASD): It is an algorithm that detects clusters with arbitrary shape.
- Spatial-Temporal Density-Based Clustering (ST-DBSCAN): It is a suitable algorithm for applications involving spatial temporal data like: weather forecasting, medical imaging, and geographical information systems.
- Ordering Points to Identify the Clustering Structure (OPTICS): it is a method that is proposed to overcome the difficulty in using one set of global parameters in clustering analysis.
- DEnsity-Based CLUstering (DECLUE): is based on a set of density distribution functions. It is used to estimate the Kernel density.

The density based algorithms can be summarized with their different characteristics in the following table:

Algorithm	Varied Density	Primary Input Requirement	Time Complexity	Cluster Type	Type of Data
DBSCAN	No	Cluster Radius, Minimum number of Objects	$O(n^2)$; O(nLogn) for spatial indexed data	Arbitrary Shaped	Spatial Data with Noise
VDBSCAN	Yes	Automatically Generated	$O(n^2)$	Arbitrary Shaped	Spatial Data with Varied Density
DVBSCAN	Yes	Two Input Parameters to be given by User	$O(n^2)$	Arbitrary Shaped	Spatial Data with Varied Density
DBCLASD	Yes	Automatically Generated	$O(n^2)$	Arbitrary Shaped	Spatial Data with Uniformly Distributed Points
ST-DBSCAN	No	Three Input Parameters to be given by User	$O(n^2)$	Arbitrary Shaped	Spatio-Temporal
OPTICS	Yes	Density Threshold	$O(n^2)$; O(nLogn) for spatial indexed data	Arbitrary Shaped	Spatial Data with Varied Density
DENCLUE	Yes	Radius	$O(n^2)$	Arbitrary Shaped	Spatial Data with Varied Density

Table 1.1 Summary of density based algorithms

12.3.2.4. Model-based clustering algorithms [**31**]

These methods use some mathematical models for clusters and attempt to optimize the fit between these models and data. These Algorithms also find characteristic descriptions for each cluster. The most frequently used methods are: decision tree and neural networks.

12.3.2.5. Grid-based clustering algorithms [**27**]

These methods divide the space into a finite number of cells, and then perform all the operations of clustering in a grid structure.

12.4. Association Rule

Association rule is a data mining technique that is used to find the relationships between objects that are frequently used together [**33**].

The formal definition of association rules is as follows: Let I= $\{I_1, I_2, I_n\}$ a set of m distinct attributes. Let D a set of database transactions where each transaction T represents a set of items such that $T \subseteq I$. An association rule is an implication of the form $X \Rightarrow Y$, where $X, Y \subseteq I$ and $X \cap Y \neq 0$ [**34**]. There are two basic criteria that association rule uses: support (S) and confidence (α).

The support (S) of an association rule is the percentage of the transactions that contain $X \cup Y$ to the total transactions in the database.

The confidence (α) of an association rule is the percent of the transactions that contain $X \cup Y$ to the number of transactions that contain X [**35**].

12.4.1. Kinds of association rule mining [**36**]

There are five dimension types of association rules:

12.4.1.1. Single-dimensional association rule

Single-dimensional association rule is an association rule that has items or attributes that reference only one dimension, for example: x is an item set ⇒ the single dimensional rule is buys(x, milk).

12.4.1.2. Multidimensional association rule

If a rule references more than one dimension.

12.4.1.3. Boolean association rule

If a rule involves associations between the presence or the absence of attributes, for example: buys(x, computer) ⇒ buys (x, scanner).

12.4.1.4. Quantitative association rule:

If it describes associations between quantitative items or attributes. In these rules, quantitative values for items or attributes are partitioned into intervals. For example: age (X,"30-34") ∧ income(X,"24K - 48K")⇒ buys(X,"high resolution TV").

12.4.1.5. Correlation association rule

In general, such mining generates a lot of rules, many of which are redundant and don't indicate a correlation relationship among item sets. As a consequence the discovered associations can be further analyzed to uncover statistical correlations, leading to correlation rules.

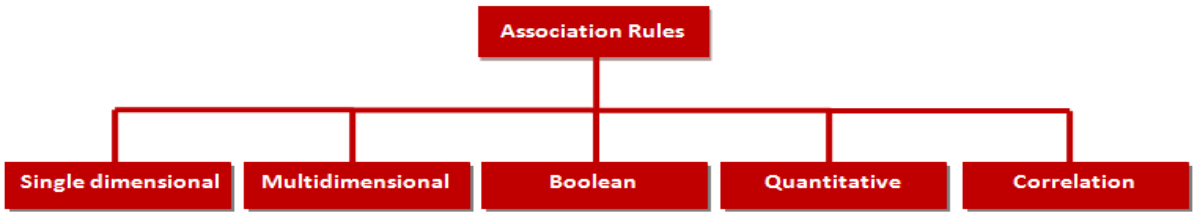


Figure 1.18: Kinds of association rule

12.4.2. Association rule algorithms:

12.4.2.1. AIS algorithm

The AIS algorithm is the first published algorithm for mining association rules. It focused on the enhancement of database with necessary functionality to process decision support queries. In this algorithm the consequent of association rule contains only one item, for example: x ∩ y ⇒ z [**35**] [**33**]. AIS algorithm proceeds according to the following steps [**37**]:

- Generate candidate itemsets and count on the fly as the database is scanned.
- For each transaction, it is determined which of the large itemsets of the previous pass are contained in this transaction.
- New candidate itemsets are generated by extending these large itemsets with other items in this transaction.

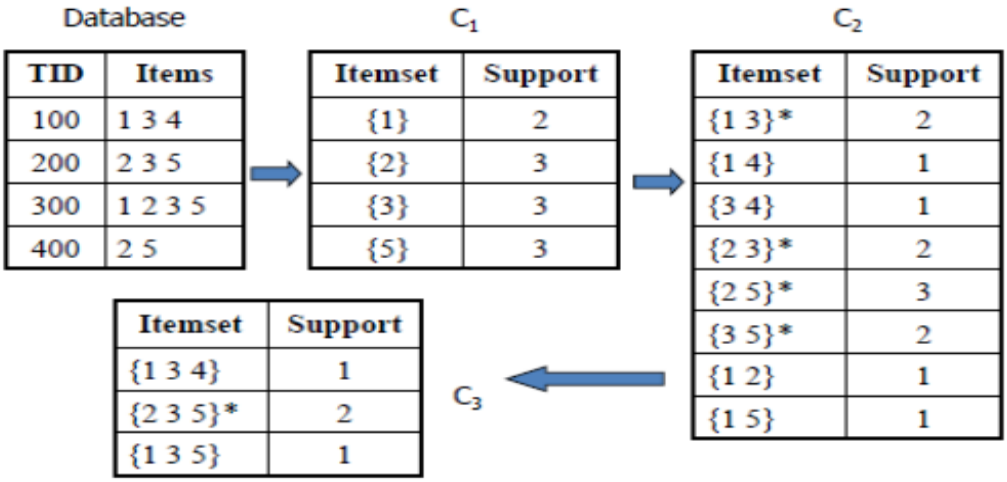


Figure 1.19: Example of AIS algorithm

12.4.2.2. SETM algorithm [**33**]:

The steps of this algorithm are the same as the AIS algorithm with small changes.

- Candidate itemsets are generated on the fly as the database is scanned but counted at the end of the pass.
- New candidate itemsets are generated in the same way as the AIS algorithm but the TID of generating transactions is saved with the candidate itemset in sequential structure.
- At the end the support count of candidate itemsets is determined by aggregating this sequential structure.

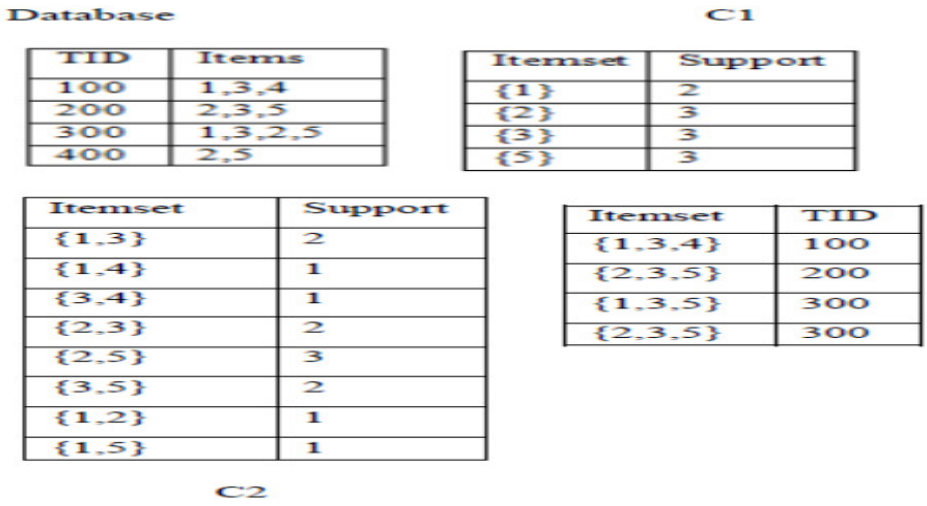


Figure 1.20: Example of SETM algorithm

12.4.2.3. Apriori algorithm [**33**]:

Apriori algorithm is used for frequent item set mining and association rule learning. This algorithm proceeds according to the following steps:

- Candidate itemsets are generated using only the large itemsets of the previous pass without considering the transactions in the database.
- The large item set of the previous pass is joined with itself to generate all item sets whose size is higher by 1.
- Each generated item set that has a subset which is not large is deleted, the remaining item sets are candidate ones.

Items	Count Number
I1	7
I2	8
I3	6
I4	2
I5	3
I6	1

a) C1

Large 1 Items
I1
I2
I3
I5

b) L1

Items	Count Number
I1,I2	5
I1,I3	4
I1,I5	3
I2,I3	4
I2,I5	3
I3,I5	1

c) C2

Large 2 Items
I1,I2
I1,I5
I2,I5
I2,I3
I1,I3

d) L2

Items	Count Number
I1,I2,I5	3
I1,I2,I3	2

e) C3

Figure1.21: Example of Apriori algorithm

13.Data mining architecture [38]

Data mining is a very important process where potentially useful and previously unknown information is extracted from large volumes of data. There are a number of components involved in the data mining process which constitute their architecture. The major of them are:

13.1. Data source

There are many sources of data but the more important are: database, data warehouse, World Wide Web (WWW), text files and other documents.

13.2. Database or data warehouse server

The database or warehouse server contains the data that is ready to be processed.

13.3. Data mining engine

It consists of a number of modules for data mining tasks.

13.4. Pattern evaluation modules

It is responsible for the evaluation of the pattern by using a threshold value.

13.5. Graphical User Interface

Graphical User Interface is a module that helps users to communicate with the data mining system easily without knowing the real complexity behind the process.

13.6. Knowledge base:

Knowledge base is useful for guiding the searching or evaluating the result patterns.

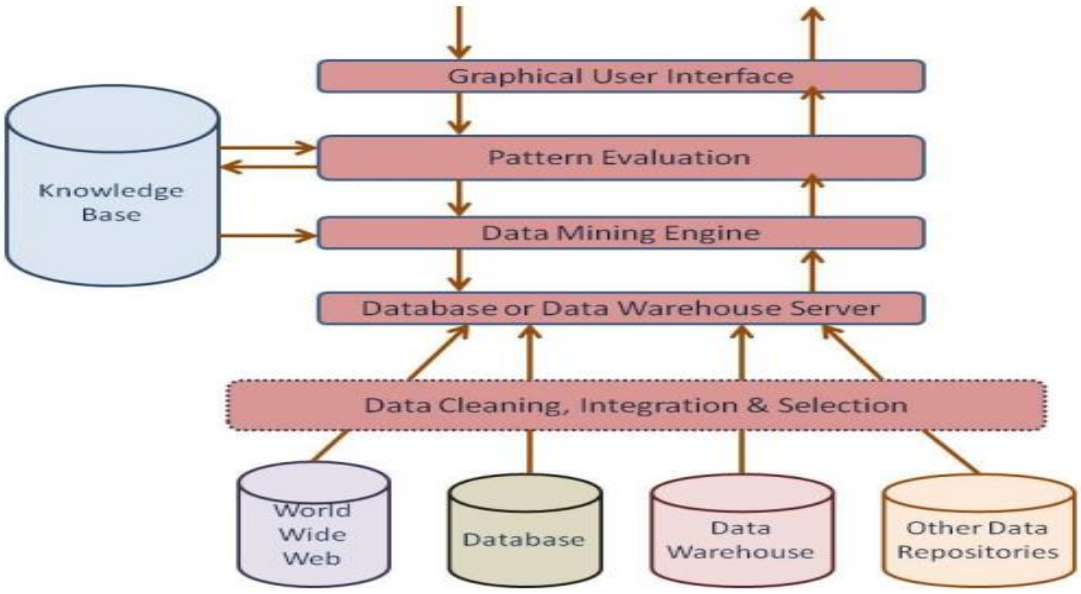


Figure 1.22: Data mining architecture

14.Data mining tools [39]:

There are a lot of tools for data mining; some of them work in classification, some work in regression, clustering, association, and other work with all the algorithms. The best of them are:

14.1. Orange

Orange is open source data visualization and analysis, data mining is done through python or visual programming.

14.2. R-Software Environment

R is a free software environment for statistical computing and graphics in all the platforms (windows, UNIX, MacOs).

14.3. RapidMiner (YALE)

RapidMiner was written in java programming language, this tool holds top position on the list of data mining tools. In addition to data mining RapidMiner also provides functionality like: data processing and visualization, predictive analysis and statistical modeling, evaluation, and deployment.

14.4. KNIME

KNIME stands for the Konstanz Information Miner. It is an open source data analytics, reporting and integration platform. It gives a graphical user interface for the assembly of nodes for data processing.

14.5. NLTK

NLTK stands for Natural Information Toolkit, written in python programming language. It provides a pool of language processing tools including data mining, machine learning, data scraping, sentiment analysis and other various tasks.

14.6. WEKA

WEKA stands for Waikato Environment for Knowledge Analysis. It contains tools for data preprocessing, classification, clustering, association rules and visualization. It supports different data files such as: CSV (Comma Separated Value), ARFF (Attribute Relation File Format)…etc.

15.Data mining process [40]

Cross-Industry Standard Process for Data Mining (CRISP-DM) consists of six phases intended as a cyclical process as the following figure:

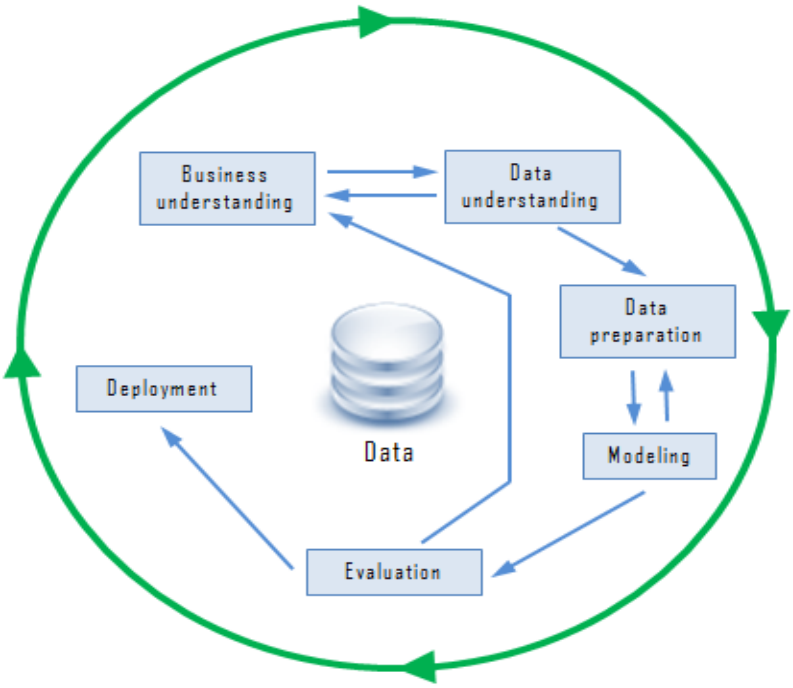


Figure 1.23: Cross-Industry Standard Process for Data Mining (CRISP-DM)

15.1. Business understanding

Determine the goal of data mining and produce of the project plan.

15.2. Data understanding

Collect data, explore this data and evaluate it.

15.3. Data preparation

Select, clean, construct and format the data.

15.4. Modeling

Select the modeling technique, generate test design and implement the model.

15.5. Evaluation

Evaluate the result and determine the next step.

15.6. Deployment

Maintenance, produce the final report and review the project.

16.Data mining application [41]

There are many application domains of data mining but it is highly useful in this following domains:

- Future healthcare.
- Market Basket Analysis.
- Education.

- Manufacturing Engineering.
- Fraud Detection.
- Intrusion Detection.
- Financial Bunking.
- Research Analysis.
- Bioinformatics.
- Customer relationship management (CRM).

17.Conclusion

In this chapter we have presented the basic concepts of business intelligence and data mining. These two domains contain methods and techniques that are very important to present and analyze data to extract useful information that help us to take a new decision.

In summary, business intelligence includes a set of concepts, methods, and technologies that are used to present data for analyzing them and making decisions. In addition, Data mining offers the possibility to analyze and explore a huge amount of data from storage support to extract useful information.

Business intelligence and data mining are two important concepts that can be used together in a lot of economic, social and health domains.

In our study, we focus on the Customer Relationship Management domain (CRM).To do this, we will present in the following our work concerning the implementation of a data mining based system for the “Géant” company specialized in the production of household appliances.

CHAPTER 2

PRESENTATION OF THE HOST ORGANIZATION

1. Introduction

In this chapter we present our host organization, which concerns the Customer Relationship Management (CRM) in Géant Electronics Company of Bordj Bou Arreridj. This last is among the leaders in electronics and household appliances production in Algeria.

In order to properly approach things, we focus in this chapter on two components. The first one concerns the presentation of the company and more precisely the CRM part which constitutes the heart of our work. The second one concerns the basic concepts of CRM management and its strategies in order to clearly show the link between this and the data mining ideas that we have developed.

2. Géant Company presentation

Géant or "SARL LOTFI ELECTRONICS" is a specialized company in the fabrication, transformation, assembly, montage and the marketing of electronic products, households and computer components. It's equipped with an infrastructure adopting a modern installation.

3. Géant strategies

The strategies of Géant electronics are characterized by the following axes:

- The position of the company in comparison to its concurrent.
- The diversification of products while ensuring technological differentiation to increase the market share.
- Invest in resources while valuing the skills acquired, by enhancing the implication and training of staff.
- Register the process in logic of durable development.
- Satisfy the requirement to which the company is subscribed.

4. Géant quality politic

Géant quality policy is brought to the attention of all Géant staff, customers, and interested parties and will have to be reviewed, to always remain consistent with Géant strategic directions.

For this, the general management of SARL LOTFI ELECTRONICS undertakes to provide the necessary resources and therefore, at all hierarchical levels, create a dynamic of

continuous improvement for the development of the company and the well-being of its employees.

5. Géant coordinates

Name: SARL LOTFI ELECTRONICS.

Work: Fabrication and montage of electronics, households, and informatics apparatus.

Address: Activity area N°94 B.B.Arreridj 3400 – Algeria.

Phone: +213 (0) 35 876 171/181.

Fax: +213 (0) 35 876 160.

Bank AGB: N°032 0010 2344 7300 1208 72.

Capital: 3.640.000.000 DA.

RCN°: 05 B 463338.

NIF: 000534046333811.

Site: www.geant-dz.com | info@geant-dz.com.

Facebook: geantelectronics.

6. Géant organizational chart

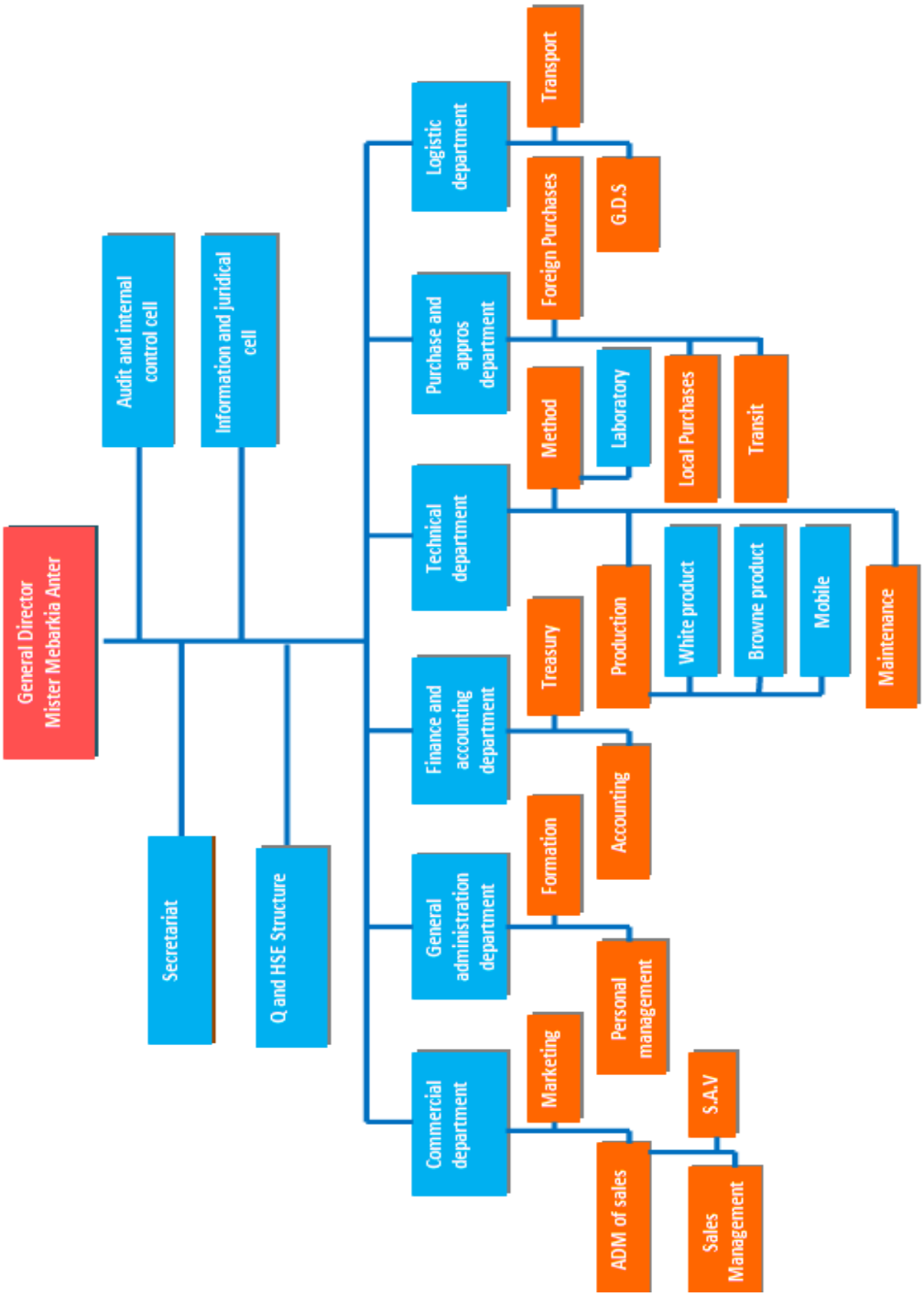
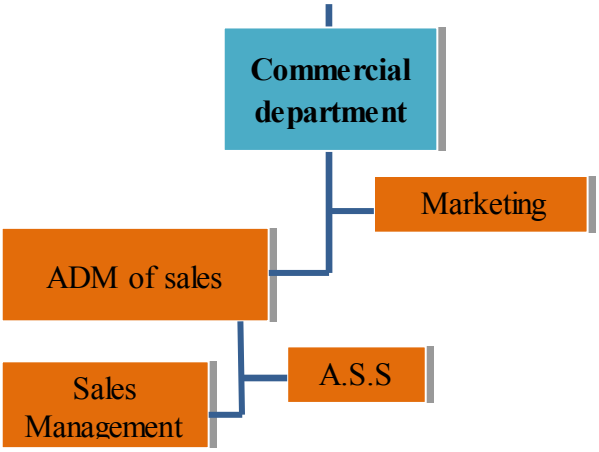


Figure 2.1 Géant organizational chart



Our case study focuses on the commercial department. This department is tasked to research customer needs in order to meet them in the best possible way, to sell the products and services of the company on which it depends, and also to manage its customer while taking care of its prospects.

This service consists of the following components:

Figure 2.2: Commercial department components

6.1. ADM of sales [42]

The ADV service (sales administration) regulates the administrative processes related to customer-company relationships. For example, it ensures that product delivery times are respected, the quality of the services provided or the compliance of the prices applied. The ADV is also in charge of the company's invoicing and customer management. The statistical monitoring of these actions allows the ADV to design dashboards, analyzes and reports, which help decision-makers to improve the functioning of the company.

6.2. Marketing

It is a process of identifying and qualifying profitable business opportunities (needs, desires of consumers, companies, etc.), then defining and implementing the actions necessary to exploit them sustainably [**43**]. There are three primary purposes of marketing [**44**]:

- Capturing the attention of your target market.
- Persuading a consumer to purchase your product.
- Providing the customer with a specific, low risk action that is easy to take.

6.3. Sales Management [45]

Sales management is a part of the company's marketing mix. It deals with the formation of sales strategies; product merchandising and pricing; sales promotion activities; distribution function; and planning, staffing, supervising, motivating and controlling of sales personnel to attain the desired sales objectives.

6.4. After-sales service (A.S.S) [46]

After-sales service refers to the set of services provided by the company to their customers after the sale, more or less varied, depending on the activities and core business of the company:

- Follow up after purchase of a product or service.
- Troubleshooting.
- Delivery, installation merchandise.

In short, the most relevant missions of the commercial department are linked to the missions of the CRM. In order to understand this idea in depth, we present in the following sections the specificities of the CRM paradigm.

7. What is CRM (Customer Relationship Management)?

CRM is a comprehension and marketing strategy, CRM also is a famous system for managing the communication between organization and its clients, customer and sales prospects, this system uses technology for organizing, synchronizing and automating business and sales activities [**47**]. In other words, CRM is a business strategy that selects and manages clients to optimize long term benefits [**48**].

8. What are CRM objectives [47]?

CRM objectives are numerous, but the most important of them are:

- Simplifying marketing and sales systems.
- Providing the best customer services.
- Discovering new customers and increasing customer profits.
- Increasing more effectively cross-selling products.

9. Benefits of CRM strategy implementing [49]

The most important benefits are global satisfaction of clients that leads to gain customer fidelity. CRM benefits are divided into two types:

9.1. Tangible benefits:

- Increase in profitability.
- Speed in time investment.
- Evolution of productivity.
- Reduce marketing costs.
- Increase rates of customer relation.
- Achieve higher return in marketing investments.

9.2. Intangible benefits:

- Achieve customer satisfaction.
- Increase the quality of services provided to customers.
- High quality of communication channels.
- Understand customer needs.
- Reduce the gap between the organization services and customer expectations.

10.Aspects of CRM [50]

Generally, the CRM has four aspects which are: customer identification, customer attraction, and customer retention and customer development.

10.1. Customer identification

Customer identification is the first step in Customer Relationship Management. This stage includes targeting individuals that will become company customers. It gives more important to analyze the lost customers and how to attract them again.

10.2. Customer attraction

Customer attraction step follows customer identification. After identifying customer's potential segments, companies can put their energy into attracting target customer segments; one of the factors for attracting clients is direct marketing. Direct marketing is an advertising process that encourages clients to command products through many different channels.

10.3. Customer retention

Customer retention is a paramount concern in Customer Relationship Management. Companies must maintain their customers by working to satisfy them and meet their needs. Customer retention includes simultaneous marketing programs for creating customer loyalty and complaints management.

10.4. Customer development

Customer development requires regularly increasing the number of interactions, the value of interactions and the personal profitability of the client. The components of customer development are: customer life cycle time analysis, average sales growth, and analysis of product basket.

11.Type of CRM

We distinguish three types of CRM which are:

11.1. Operational CRM [51]

Operational CRM is defined as the automation and the constant improvement of sales, marketing, and customer service. It includes these automations:

- **Sales Force Automation (SFA):** is the application of technology for managing selling activities.
- **Marketing Automation:** includes event-based marketing, promotion, campaigns management and market segmentation. This strategy enables marketing force to access to customer data for design, execute, and evaluate targeted offers.
- **Service Automation:** includes resolving issues or cases, service, level management, addressing inbound communication and diagnosing and solving the issues about product.

11.2. Analytical CRM [52]

Analytical CRM composed of collections of actions, tools, methods and applications that allow analyzing data of customer relationship. This kind of CRM is used to discover patterns of customer behavior, to determine target, additional sales potential. Among the methods used in this type of CRM are: data mining, scoring, OLAP, modeling…etc.

11.3. Collaborative CRM [53]

Collaborative CRM type allows various services departments to share customer information. The main goal of this type is to promote company-wide harmony. Collaborative CRM can be used to help departments of a company to understand how they fit in when fulfilling a customer's wants and needs. The sharing operation is done through these two ways:

- **Interaction management:** interaction management helps companies to track all the interactions between them and their customers. The collected information from this tool allows keeping mails, social media, phone calls…etc.
- **Channel management:** channel management allows adapting the company strategy to the needs of their customers.

12.The function of CRM [54]

The main functions of CRM are:

12.1. Sales Forces Automation (SFA)

Sales Forces Automation is the automation of the sales cycle. This functionality integrates the different stages of the sales: seizure objectives, elaborate forecasts, managing customer portfolios…etc.

12.2. Enterprise Marketing Automation

The automation of marketing campaigns allows collecting, analyzing, and segmenting customer information in order to optimize the actions to be undertaken.

12.3. Call Center

Customer center the automation of the various processes of taking calls from numbering, receiving calls until analyzing and elaborating reports.

12.4. Contact Center

In addition to the classical CRM call center solution, contact center solution allows us to manage multimedia relations such as: telephone, email, fax, mail…etc.

12.5. Marketing Database

with the different functionality offered by CRM solutions, the implementation of such solution requires a real infrastructure of information management which is designated by marketing databases, that store different information relating to each customer (last name, first name, age, sex, family situation, contact…etc), all this data are exploited through data mining tools which allow the realization of market segmentation.

13.CRM process [55]

The CRM process passes by five stages which are:

13.1. Identification

Identification consists of collecting information about each customer, or at least those how have been targeted in the action plan. There are a lot of information sources such as: questionnaire, sales force, accounting, loyalty cards, after sales services, call center, internet, etc. Companies often have partial and imperfect databases but they should not be neglected. The build of customer database pass by some big steps which are:

- Audit of the current customer information sources, internal or external to the company.
- Evaluation of the potential information sources in terms of interest, coast, and feasibility.

- Definition of the policy that systematically collects information.
- Evaluation of the feasibility to integrate existing databases.

13.2. Segmentation

Customer identification is the first step. We have to analyze data and group customers. Database segmentation consists of grouping customers according to common characteristics that can be affecting their behavior. Thus, we can segment customers according to different criteria:

- Importance of their purchases (small |medium |large customers).
- The type of product purchased.
- Their expressed interest.
- Their calculated probability of defection.
- And the more important one is the segmentation by profitability.

13.3. Adapting

Identification and segmentation of customers must lead to adapt the service and the communication towards customers. This adaptation focused on offers, on contact channels or on the content of the communication.

13.4. Exchanging

The interaction results either from campaigns organized by the company or from a response to customer requests. In the first case, the company posts a letter, makes a phone call, and sends emails to make an offer to the customer or maintain a relationship. In the second case, it is the customer who solicits the company by calling a standard phone, sending a letter or email, or by going to the purchasing point, in this case companies must take advantage of contact opportunities by feeding the database with new information about customers.

13.5. Evaluation

The customer relationship is built in the time with each interaction therefore learning is an essential dimension of this process. The objectives must be quantified in order to be evaluated. Indicators can take several forms: satisfaction index, attrition rate, profitability per customer, etc.

Within the company, partial and demonstrable results are the best argument in progressive approach to advance the file of customer relation.

14.What is Enterprise resource planning (ERP)?

Enterprise Resource Planning is a process used by companies to manage and integrate the important parts of the business such as: purchase, sales, manufacturing, human resource, marketing, finance and more [**56**]. ERP is business management software used to improve performance, profitability, and productivity of the company [**57**].

Figure 2.3: ERP modules

15.Function of ERP [58]

ERP allow companies to:

- Enter data and find it instantly wherever it counts in their system.
- Optimize its department and their organization.
- Link each department to automate the flow of activities.
- Manage the company with a single database.
- Have a global portrait about the situation in order to make better decisions available for a long term.

16.The difference between CRM and ERP [57]

- CRM is defined as a system that lets the organization to trace all the transactions of customers. However, ERP refers to a software program that helps a company to manage its business process.
- CRM focuses on increasing sales, ERP gives emphasis on reducing costs.
- CRM is oriented towards the management of customer relationship with the enterprise while ERP is mainly concerned with planning the resources of the company to insure its best possible uses.
- CRM is one of the modules of ERP.

17.Applying data mining in CRM [59]

It is a new technology used to take advantage of historical data. Data mining techniques in CRM help to assist a company in finding and selecting the relevant information that can be used to get a global view of the customer life-cycle, among these techniques there are: association rule, clustering, classification and regression.

18.The process of data mining in CRM [60]

18.1. Define business problem

Each CRM application has at least one business objective for which there is a need to build the appropriate model.

18.2. Build a marketing database

Building a marketing database is the second step in the process that aims to create a marketing database based on an operational database and corporate data warehouse.

18.3. Explore the data

This step understands data because it is very important before building the model.

18.4. Prepare data for modeling

Prepare data for modeling is the last data preparing stage before building a model. This step is composed into four parts which are:

- Select the variable on which we build a model.
- Construct new predictors derived from the raw data.
- Decide to select a subset or sample of our data on which to build models.

- Transform variables in accordance with the requirement of the algorithm that we choose it for building a model.

18.5. Data mining model building

Data mining model building is an iterative stage that aims to find and build the best model in solving business problems.

18.6. Result evaluation

Evaluate the building model using data which is not involved in the modeling stage.

18.7. Deploy model and result

Data mining is a small in building CRM application but a critical part of the final product. The way of integrating data mining in the application is determined by the nature of customer interaction.

19.The benefits of applying data mining in CRM [61]

- **Provide useful insights:** the use of data mining allows us to analyze historical customer data and provides useful hidden insights.
- **Helps to understand customer behavior:** data mining process helps to identify the shopping pattern of customers and gives more importance for customer's behavior, which helps to understand customer's opinions.
- **Helps to design a results-driven marketing campaign:** with the help of accurate data models companies can implement successful marketing campaigns.
- **Helps to measure profitability:** applying data mining in CRM helps to measure the profitability of the company. In addition, with the analysis operation of customer data, we can identify which products, services, and coasts keep our customers come back for more.
- **Enhance customer and brand loyalty:** data mining helps to enhance customer loyalty, brand loyalty and increase the number of product purchases.
- **Helps to interpret market trends:** data mining can help to predict future market trends and identify the new target markets.

Figure 2.4: Benefits of data mining in CRM

20.Some applications of data mining in CRM [59] [61] [62]

Data mining is involved in several applications in the company, namely:

- Fraud detection.
- Product customization.
- Market segmentation.
- Sales forecasting.
- Basket analysis.
- Database marketing.
- Customer analysis.
- Customer feedback.

21.Conclusion

In this chapter, we have provided a global overview about Géant Electronics which constitutes our host organization. So, we gave a global overview about this company, its strategy, its quality policy, its coordinates and its organizational chart.

In addition, and in order to understand our mission through this project we have presented the CRM paradigm. We have learned a lot of things about it, beginning with its definitions, its types, its implementation steps and its roles by applying data mining within its tasks.

In the next chapter, we will approach the conceptual work by detailing the CRM process in our host organization by using the appropriate tools.

CHAPTER 3

ANALYSIS AND DESIGN

1. Introduction:

Before starting the implementation process of our application, we must first tackle on its analysis and design through the use of graphical modeling languages that help us to understand it more. At the present time there are a lot of designing languages, but we have chosen the Unified Modeling Language (UML).

In this chapter, we are going to give a global overview of UML and use some of its diagrams to present the different spots of the CRM application.

2. What is UML [63]?

UML is an acronym that stands for Unified Modeling Language. It was created by the Object Management Group (OMG) in January 1997; it is the fusion of three older languages which are: Booche, Object Modeling Technique (OMT), and Object Oriented Software Engineering (OOSE).

UML is a graphical modeling language based on diagrams, designed to provide a standardized method for visualizing the design of software systems and particularly the software systems that are built using the Object Oriented (OO) style.

3. What is UML diagram [64]?

UML diagram is a graphical representation that divides the software systems into components and subcomponents.

4. Types of UML diagrams

There are 14 types of UML diagrams and each one of them serves a different purpose. These diagrams are organized in two main categories: structural diagrams and behavioral diagrams.

4.1. Structural diagrams [64]

Structural diagrams show the structure of the software system (class, components, objects, packages…) and the relationships between those items. The main diagrams in this type are:

- Class diagram.
- Object diagram.
- Components diagram.
- Deployment diagram.

- Package diagram.
- Composite structure diagram.
- Profile diagram

4.2. Behavioral diagrams [64]

Behavioral diagrams show the interaction of the system with itself, users, others systems and other entities. The main diagrams in this type are:

- Use case diagram.
- Sequence diagram.
- Activity diagram.
- State diagram.
- Communication diagram.
- Timing diagram.
- Interaction overview diagram.

5. The presentation of starUML [65]

Star UML is open source UML modeling software, developed in Delphi. It works in the most popular operating system; it allows designing different types of diagrams such as: diagrams of classes, components, use cases, sequences...etc.

6. Project tasks modeling

For the modeling of our application, we used three UML diagrams which are:

6.1. Use case diagram

- The use case diagram allows identifying the possibilities of interaction between the system and its actors [**66**].
- The actor: is a person or material or another system that interacts with our application in order to achieve the added value.
- In this diagram (figure3.1), we show four actors: Storekeeper, Sales manager, Customer and Analyst dealing with our system.

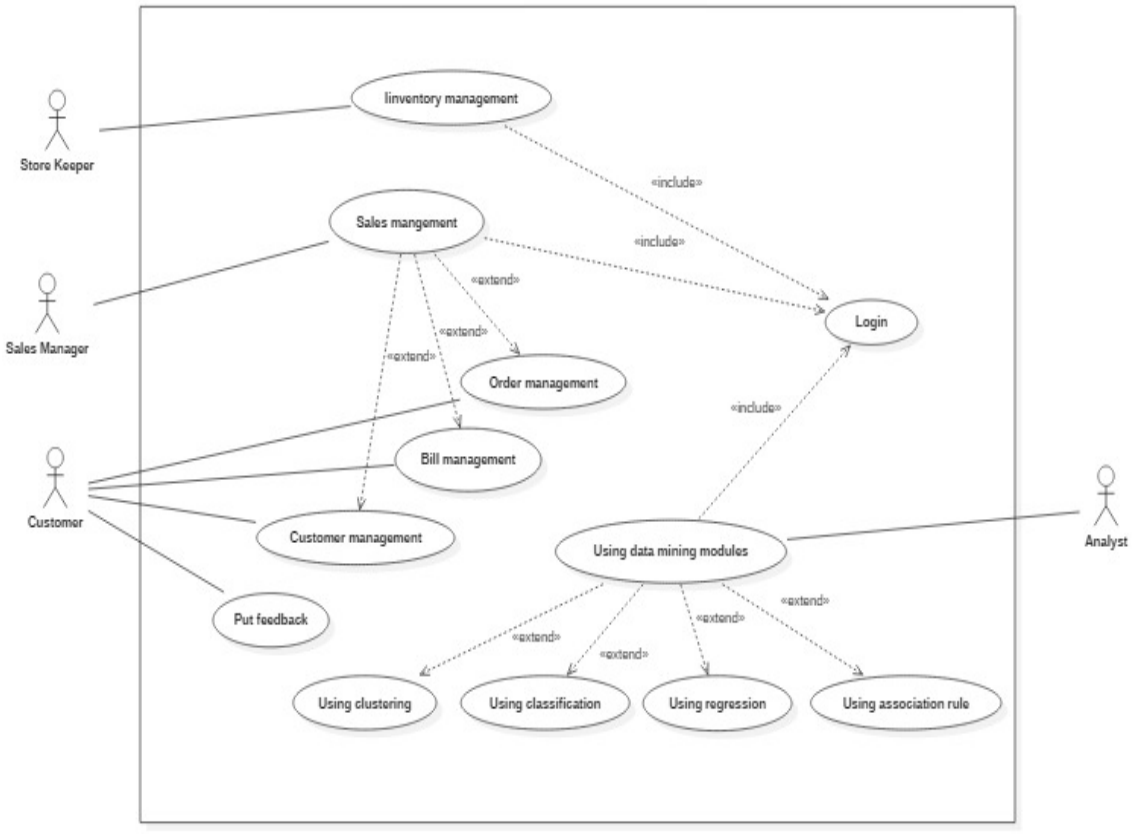


Figure 3.1: Use case diagram of CRM system

6.2. Sequence diagram

It is a sequential representation of the processing sequence and the interactions between the system elements and its actors [**66**].

In this section we will present the sequence diagram of the different spots of the system: login, customer management, bill management, order management, inventory management and using data mining modules.

6.2.1. Login

In the login diagram represented in figure 3.2 , we show the sequence of steps of the process, beginning by entering user information, then testing this information and in the end determining whether he can enter the system or not.

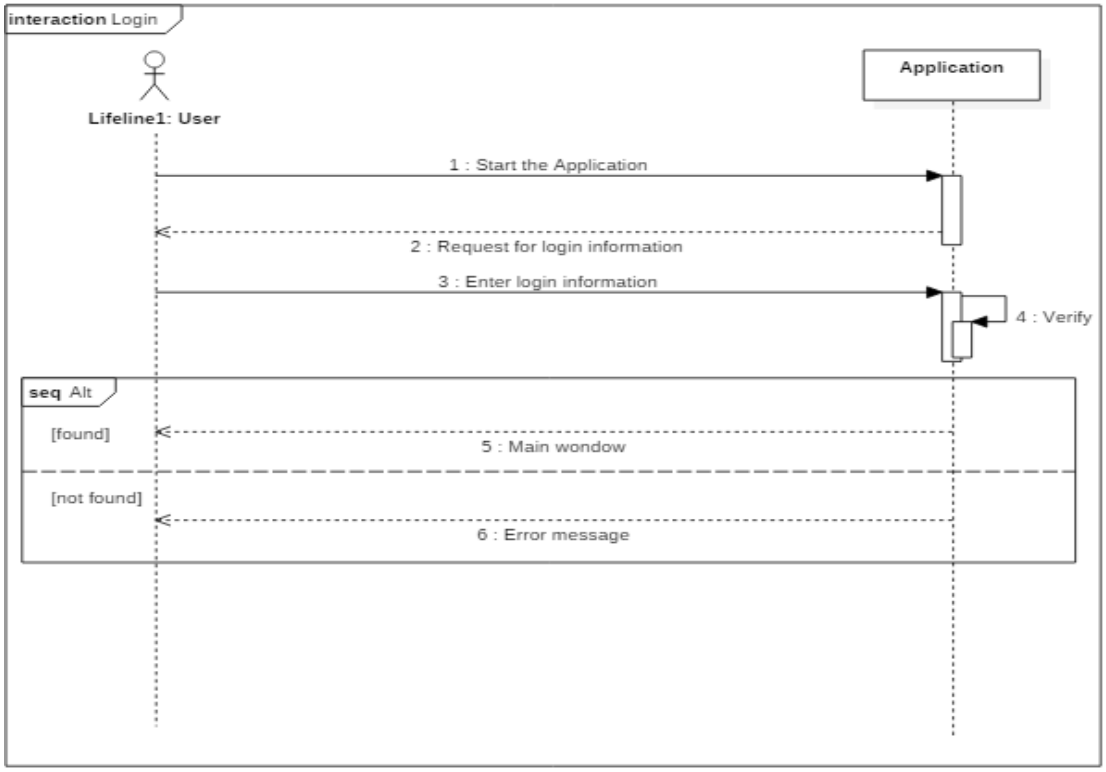


Figure 3.2: Sequence diagram of user login

6.2.2. Customer management

In this section we show the sequence diagram of adding, updating and deleting customers, mentioning the followed steps.

Add Customer Diagram represented in the figure below:

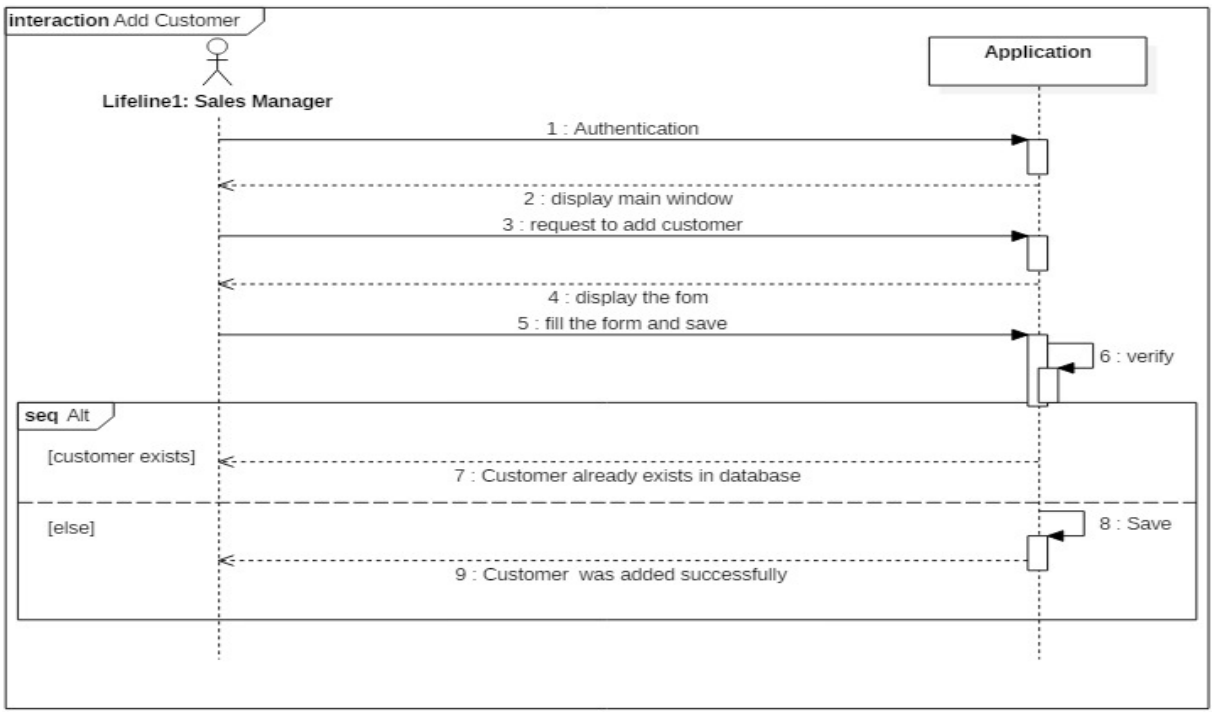


Figure 3.3: Sequence diagram of add customer

Update Customer Diagram represented in the figure below:

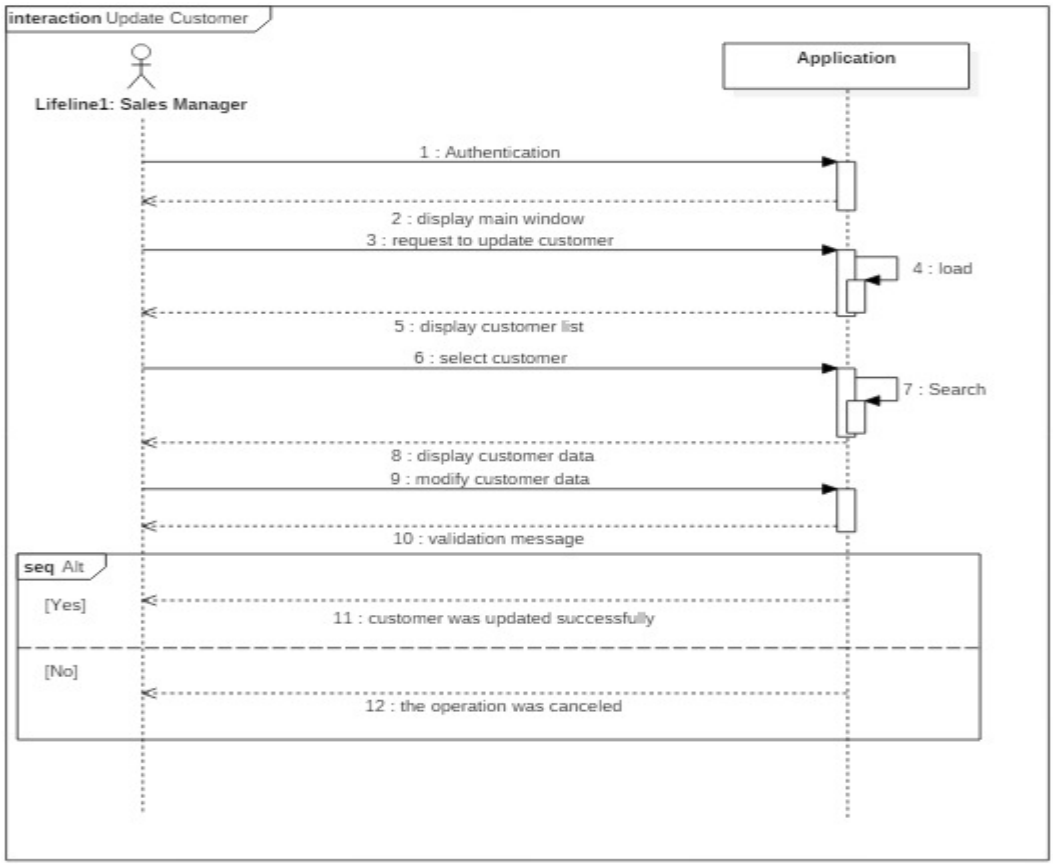


Figure 3.4: Sequence diagram of update customer

Delete Customer diagram represented in the figure below:

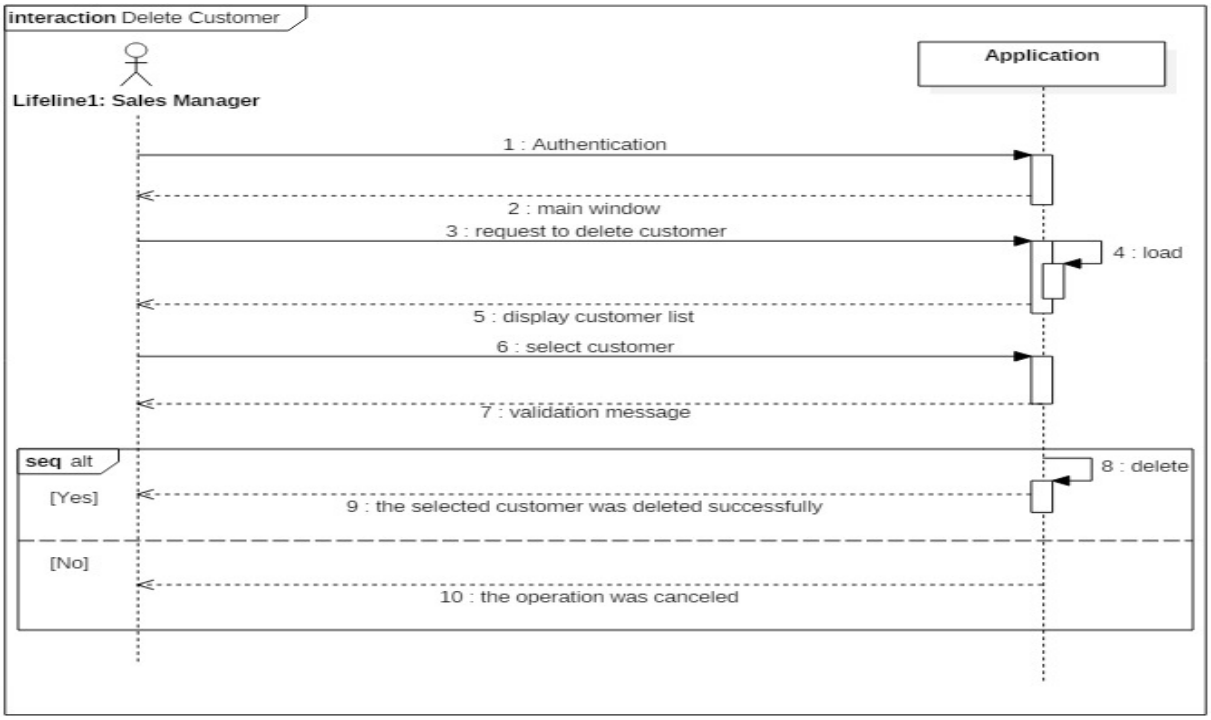


Figure 3.5: Sequence diagram of delete customer

6.2.3. Order management

In this section we show the sequence diagram of adding and printing order, mentioning the followed steps.

Add Order diagram represented in the figure below:

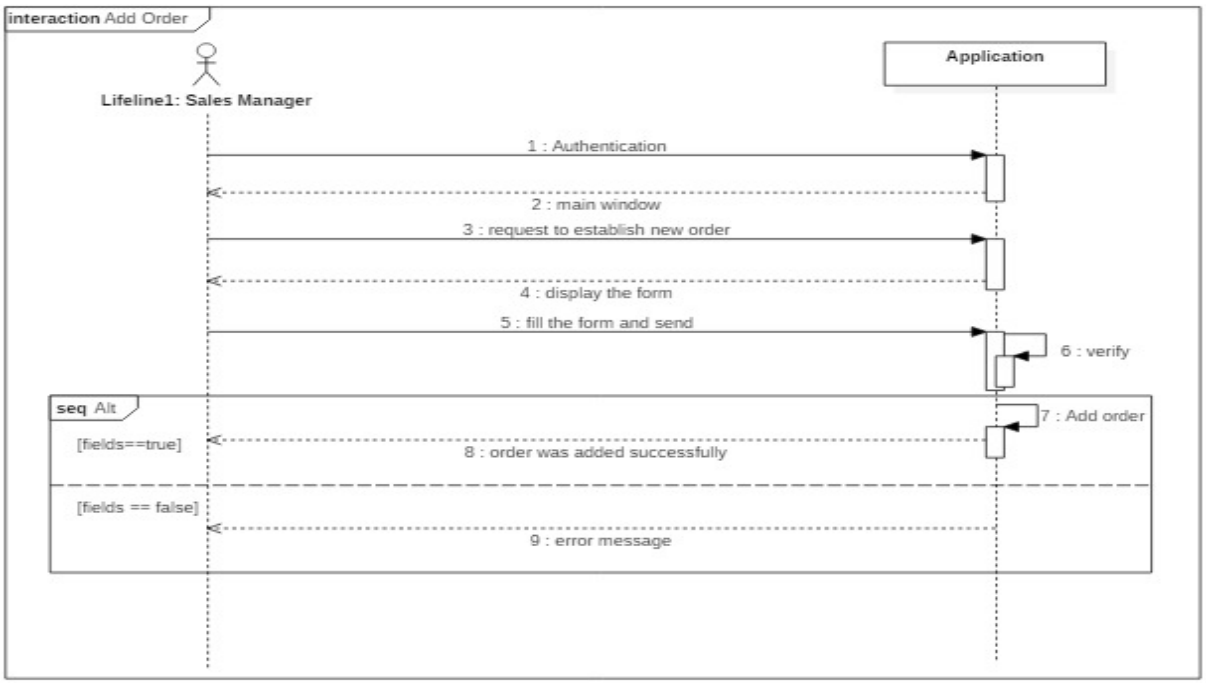


Figure 3.6: Sequence diagram of add order

Print Order diagram represented in the figure below:

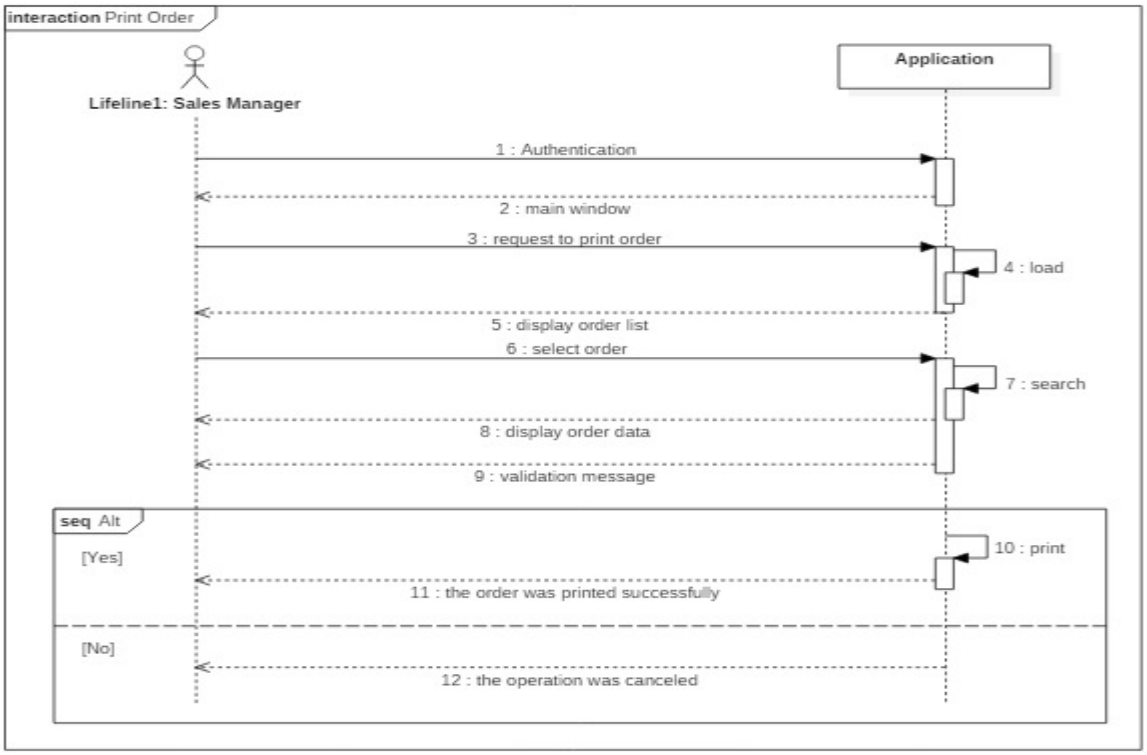


Figure 3.7: Sequence diagram of print order

6.2.4. Bill management

In this section we show the sequence diagram of adding, updating, and printing bill, mentioning the followed steps.

Add bill diagram represented in the figure below:

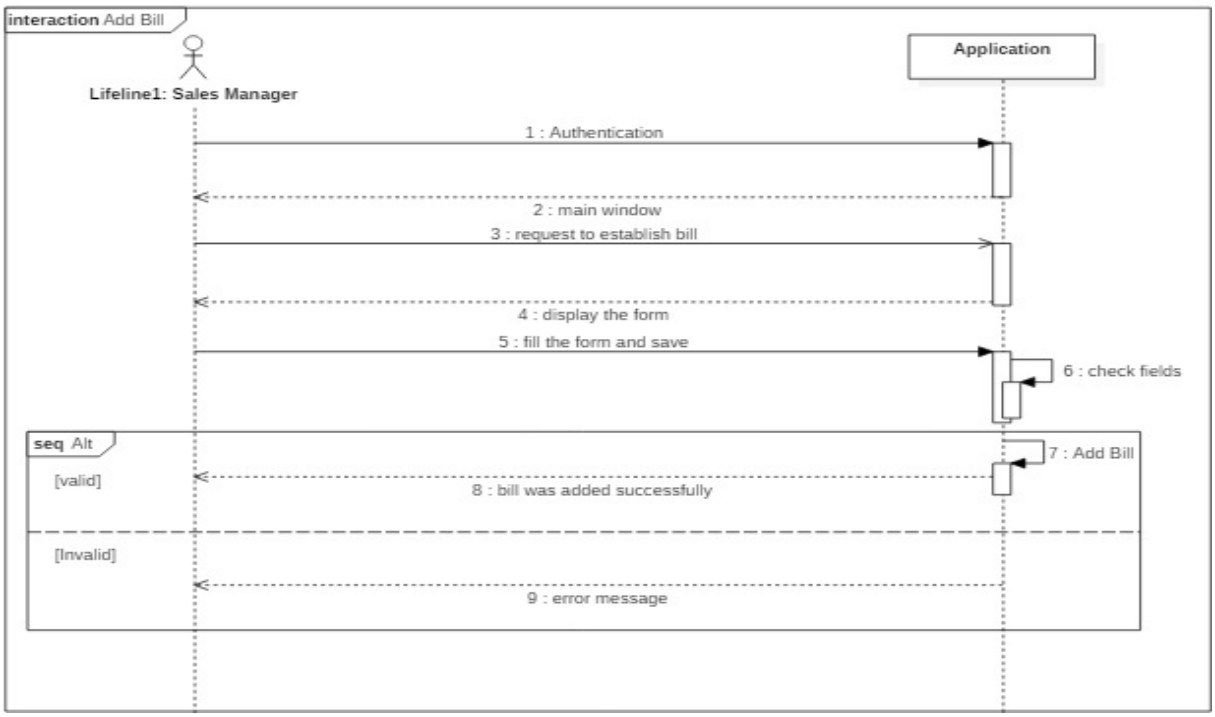


Figure 3.8: Sequence diagram of add bill

Update bill diagram represented in the figure below:

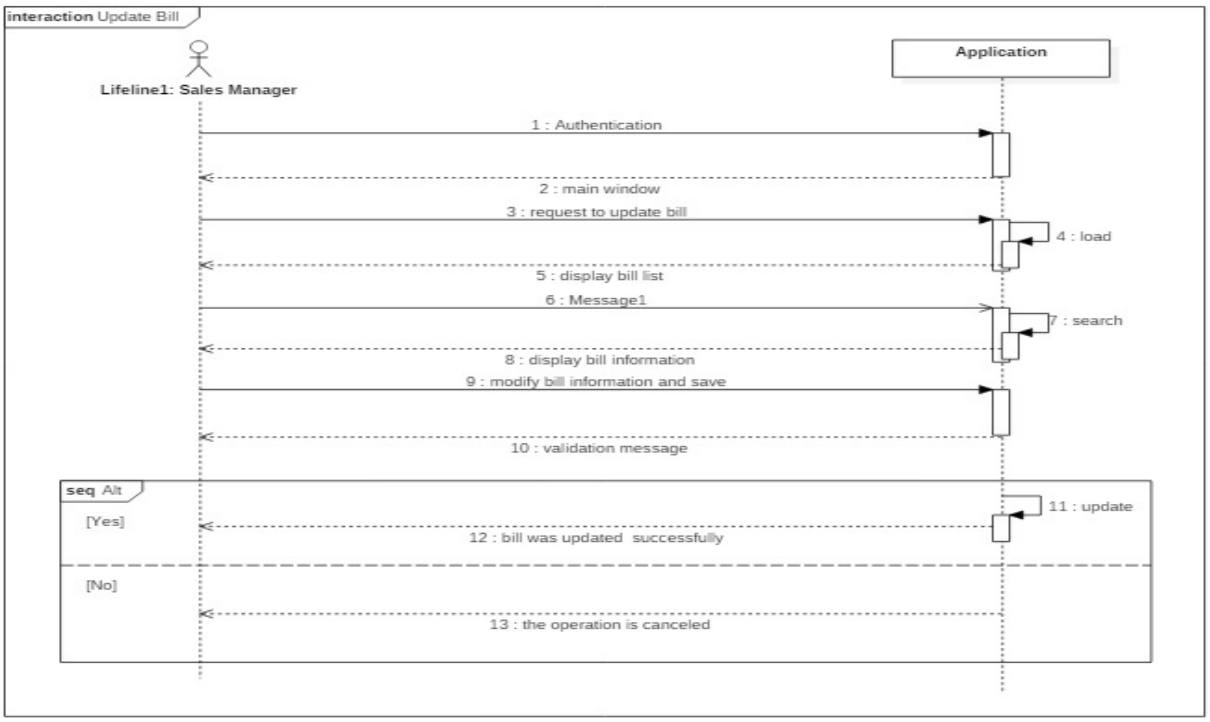


Figure 3.9: Sequence diagram of update bill

Print bill diagram represented in the figure below:

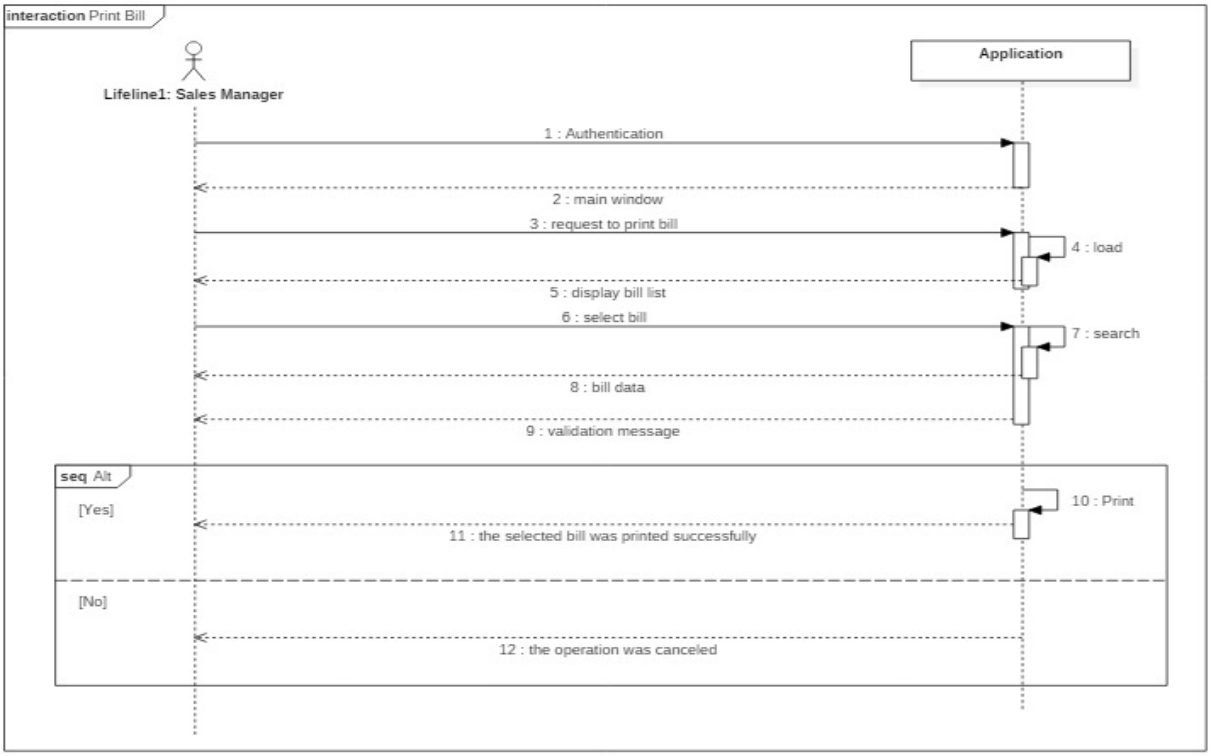


Figure 3.10: Sequence diagram of print bill

6.2.5. Inventory Management

In this section we show the sequence diagram of checking inventory level, adding and printing input voucher, adding and printing output voucher, mentioning the followed steps. Consult Inventory level diagram represented in the figure below:

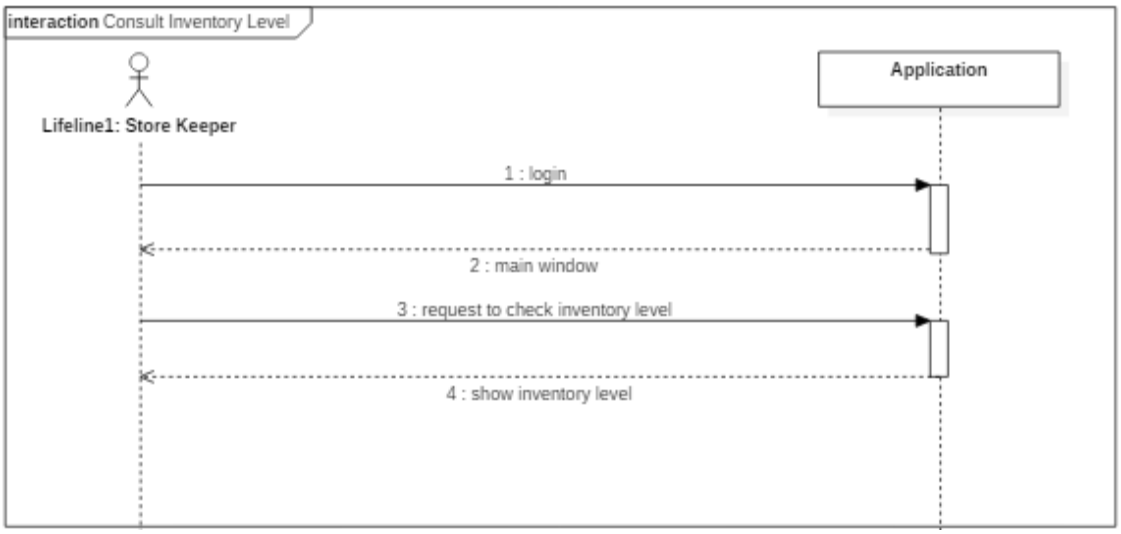


Figure 3.11: Sequence diagram of consult inventory level

Add Input Voucher diagram represented in the figure below:

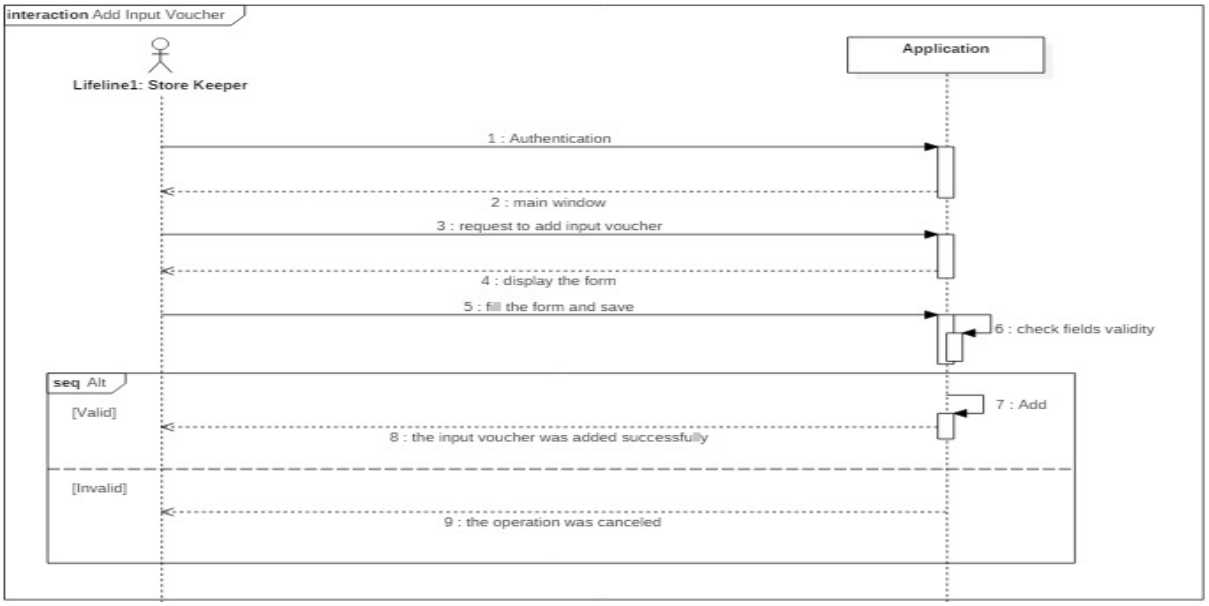


Figure 3.12: Sequence diagram of add input voucher

Print Input Voucher diagram represented in the figure below:

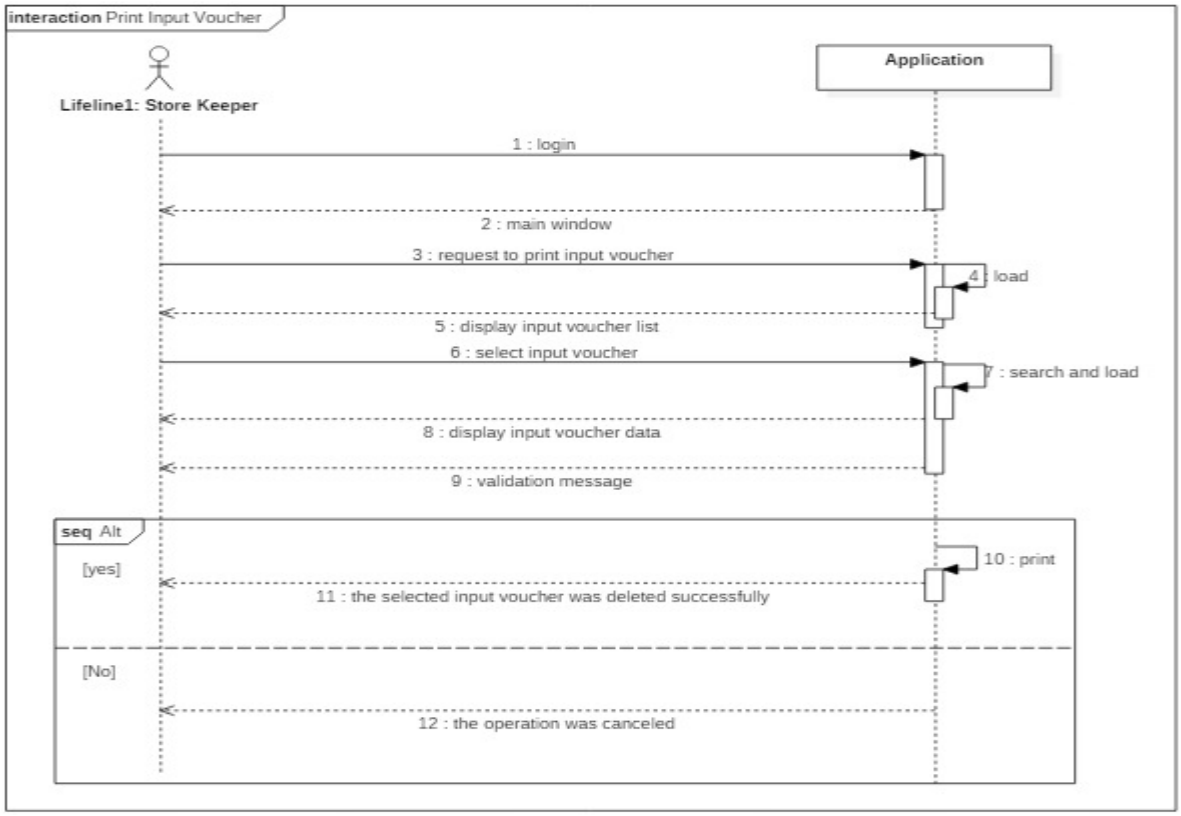


Figure 3.13: Sequence diagram of print input voucher

Add Output Voucher diagram represented in the figure below:

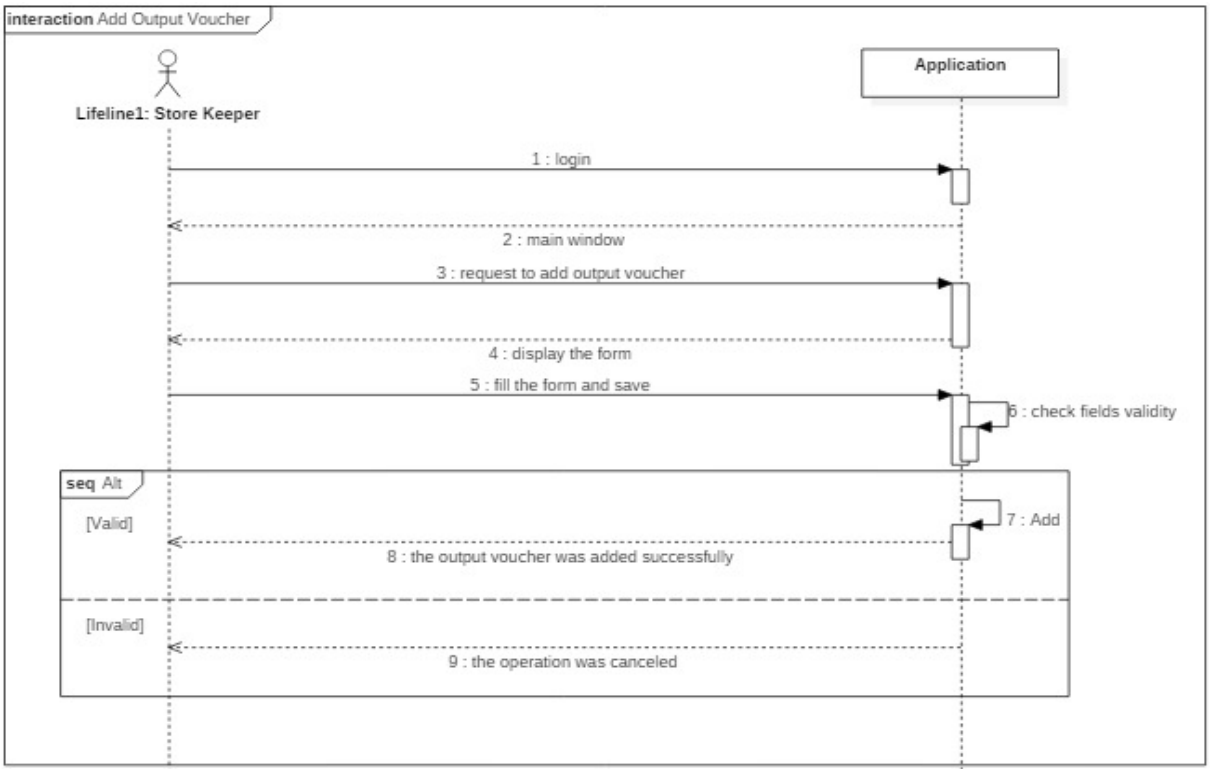


Figure 3.14: Sequence diagram of add output voucher

Print Output Voucher diagram represented in the figure below:

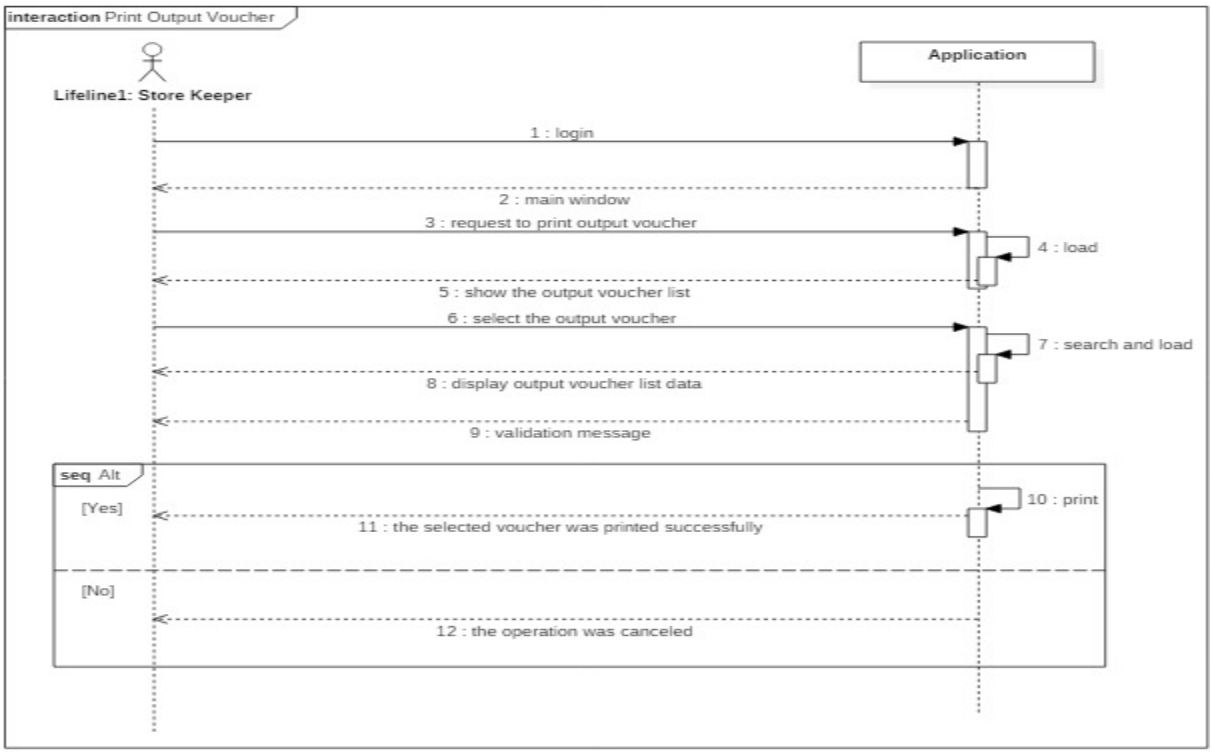


Figure 3.15: Sequence diagram of print output voucher

6.2.6. Using data mining modules

In this section we present sequence diagrams of the different data mining tasks used in the system, mentioning the followed steps.

Use Clustering algorithm diagram represented in the figure below:

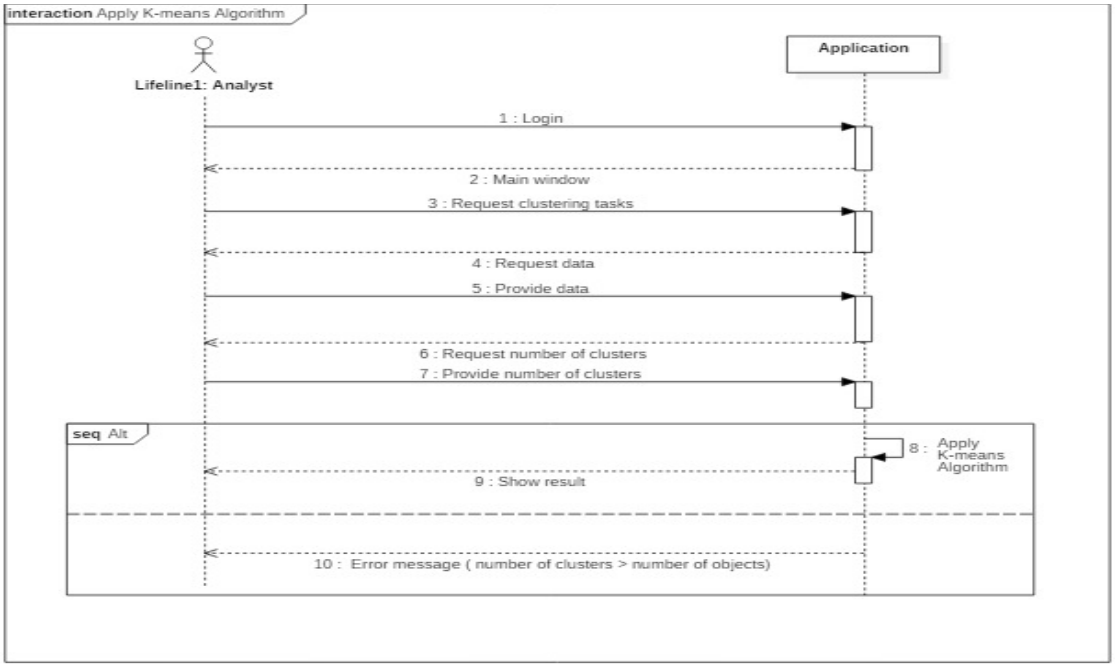


Figure 3.16: Sequence diagram of apply clustering algorithm

Use Classification algorithm diagram represented in the figure below:

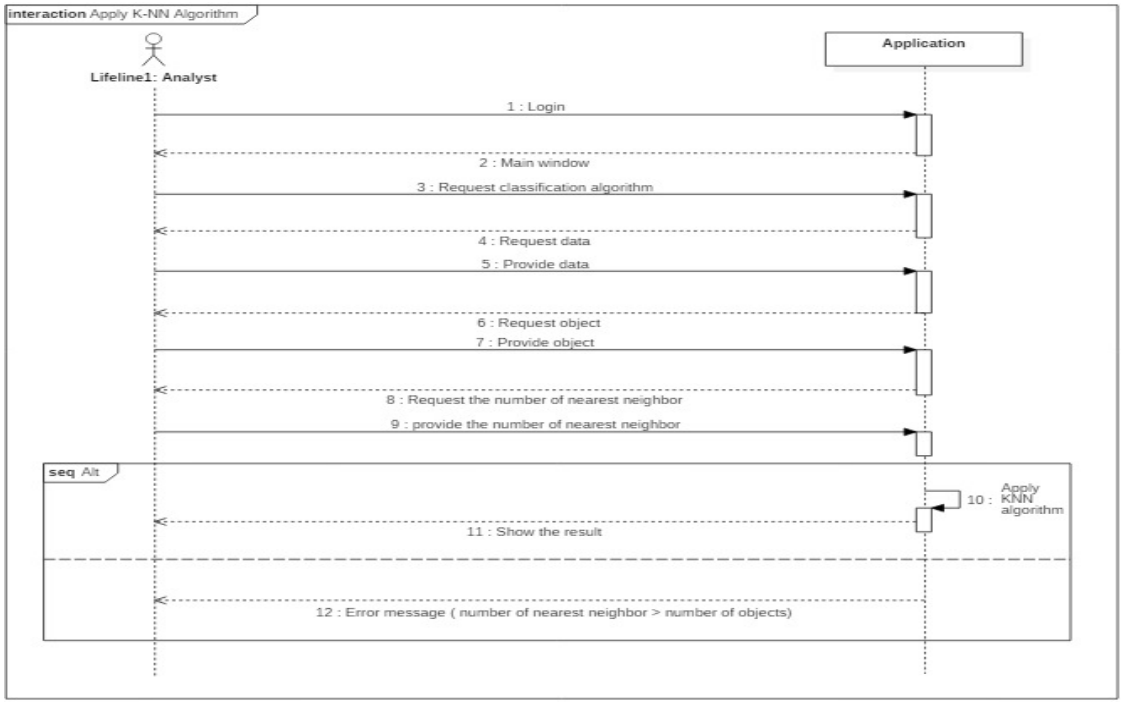


Figure 3.17: Sequence diagram for apply classification algorithm

Use Association Rule algorithm diagram represented in the figure below:

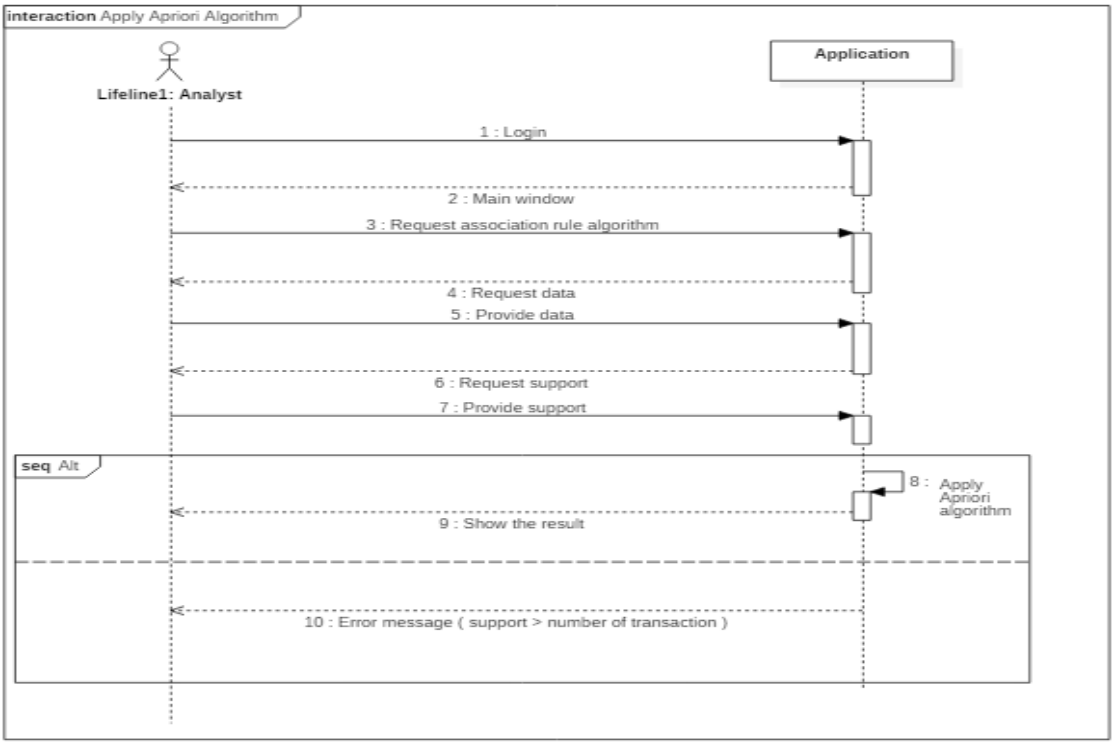


Figure 3.18: Sequence diagram of apply association rule algorithm

Use Regression algorithm diagram represented in the figure below:

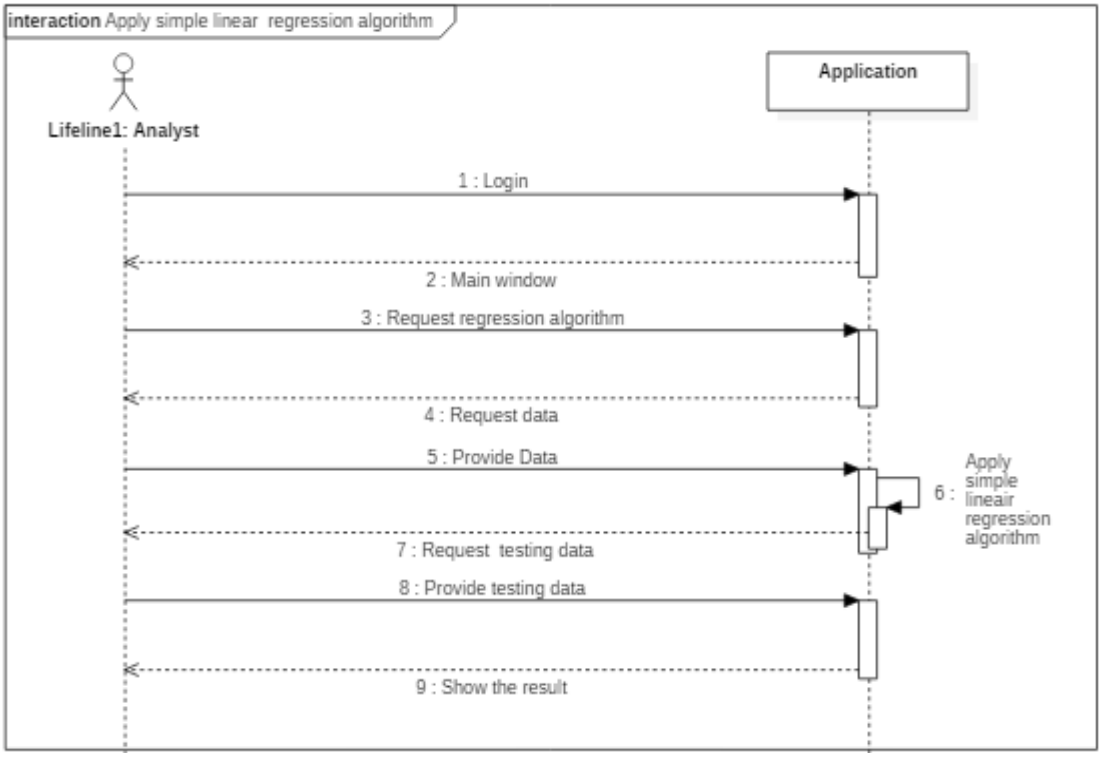


Figure 3.19: Sequence diagram of apply regression algorithm

6.3. Class diagram

It is the most commonly used type of UML diagrams. It shows the static structure of the software system, including classes, attributes, methods, and the relationships between classes [**64**].

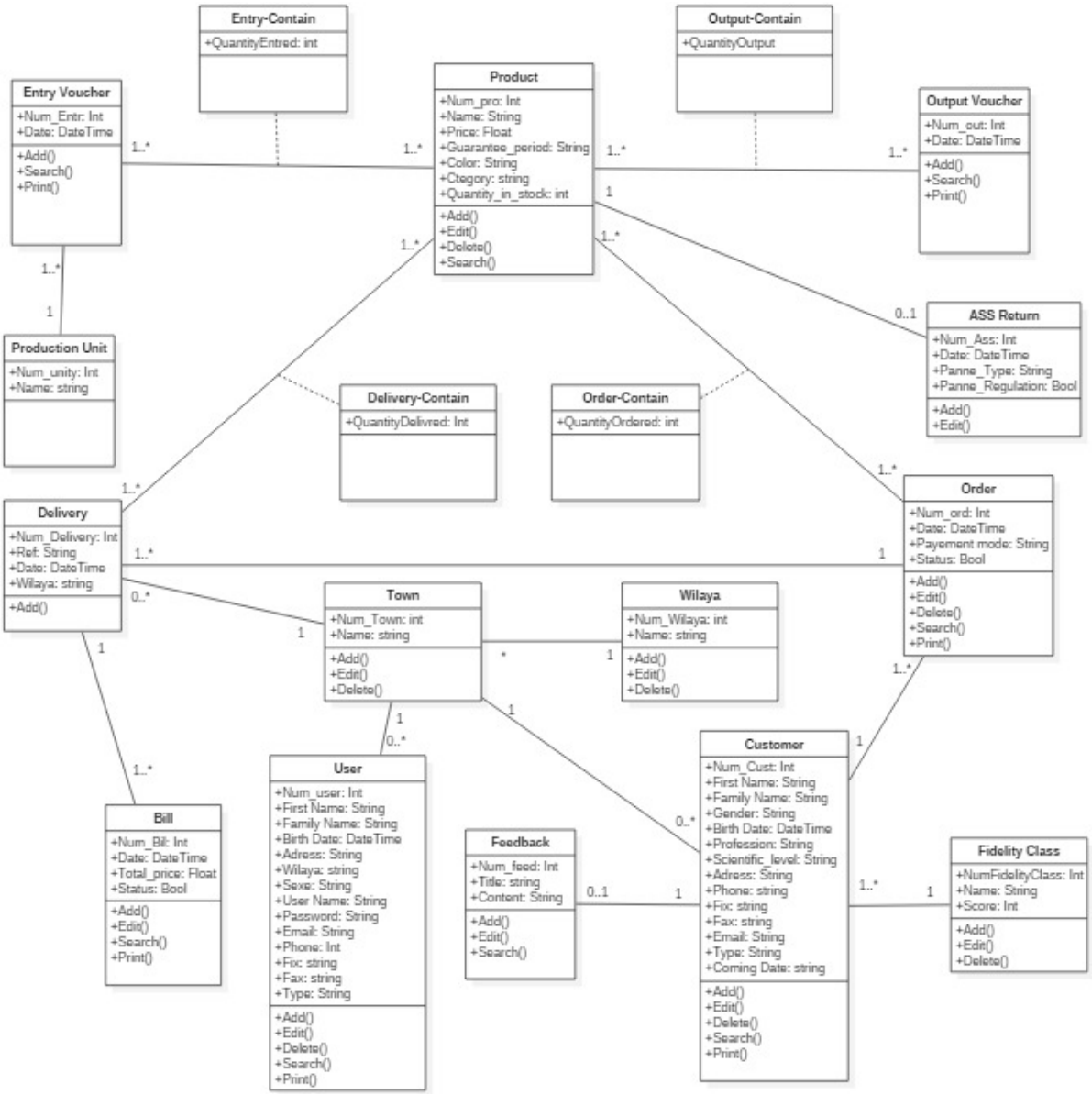


Figure 3.20: Class diagram of CRM system

7. Conclusion

In this chapter we have approached the analysis and design phase of our development. We have also presented the UML language which is a graphical modeling language used to present the software in graphical way to simplify the understanding of the system components. We also presented the different tasks of our CRM application by the use of three diagrams which are: Use case diagram, sequence diagrams and class diagram. In our case

these three diagrams are sufficient to clearly express our needs and to properly present the components and functionalities of our application.

In the next chapter, we will present the implementation phase and we will present the environment and the technical tools used as well as the application carried out.

CHPTER 4

IMPLEMENTATION AND OPERATION

1. Introduction

The content of this chapter is devoted to the technical aspect. Let's start with the rules governing the class diagram used in our design phase to the relational model. By the following, we will present the different tools and software solutions used in the development of our application. Finally, we will present the great functionalities provided by our developed system.

2. What is Relational Model (RM) [67]

The relational Model (RM) was introduced by Codd in 1970 at IBM's San Jose research laboratories, in order to achieve these following objectives:

- Ensure the independence of application and internal representation of data.
- Manage data consistency and redundancy problems.
- Use data languages based on solid theories.

3. Transformation Rules for passing from class diagram to relational model

There are four principal rules which are [**68**]:

3.1. Class Transformation

Each class of the UML diagram becomes an entity; you have to choose an attribute of the class that can play the role of key.

3.2. Association transformation with multiplicity (1..)

A foreign key attribute must be added in the child relationship of the association. The attribute bears the name of the parent of the association.

3.3. Association transformation with multiplicity (*..*)

The association becomes relation. The primary key of this relation is the concatenation of the identifiers of the classes connected to the association.

3.4. Association transformation with multiplicity (1..1)

Foreign key attributes must be added to the relation derived from the class having the minimum multiplicity equal to the attributes bearing the name of the primary key of the relation derived from the class connected to the association. If the two minimum multiplicities are one, it is better to merge the two classes into one.

4. Results of transition to the relational model

By applying the different rules mentioned above, we will obtain the following relational models:

- EntryVoucher (Num_EntryVoucher, Date, Num_ProductionUnit)
- OutputVoucher (Num_OutputVoucher, Date)
- ProductionUnit (Num_ProductionUnit, Name)
- Product(Num_Product, Name, Price, Guarantee_Period, Color, Category, Quantity_in_Stock).
- Output-Contain (Num_Product, Num_OutputVoucher, QuantityOutput)
- Entry-Contain (Num_Product, Num_EntryVoucher, QuantityEntred)
- ASSReturn (Num_ASSReturn, Date, PanneType, PanneRegulation, Num_Product)
- Delivery (Num_Delivery, Ref, Wilaya, Date, NumTown, Num_Order, Num_Town)
- Order (Num_Order, Date, PayementMode, Status, Num_Customer)
- Delivery-Contain (Num_Delivery, Num_Product, QuantityDelivred)
- Order-Contain (Num_Order, Num_Product, QuantityOrdred)
- Town (Num_Town, Name, Num_Wilaya)
- Wilaya (Num_Wilaya, Name)
- Bill (Num_Bill, Date, TotalPrice, Status, Num_Delivery)
- User (Num_User, FirstName, FamilyName, BirthDate, Gender, Type, Address, Wilaya, Phone, Fix, Fax, Email, UserName, Password, Num_Town).
- Feedback (Num_Feedback, Title, Content, Num_Customer)
- Customer (Num_Customer, FirstName, FamilyName, BirthDate, Gender, Type, Address, Wilaya, Phone, Fix, Fax, Email, ComingDate, Num_FidelityClass)
- FidelityClass (Num_FidelityClass, Name, Score)

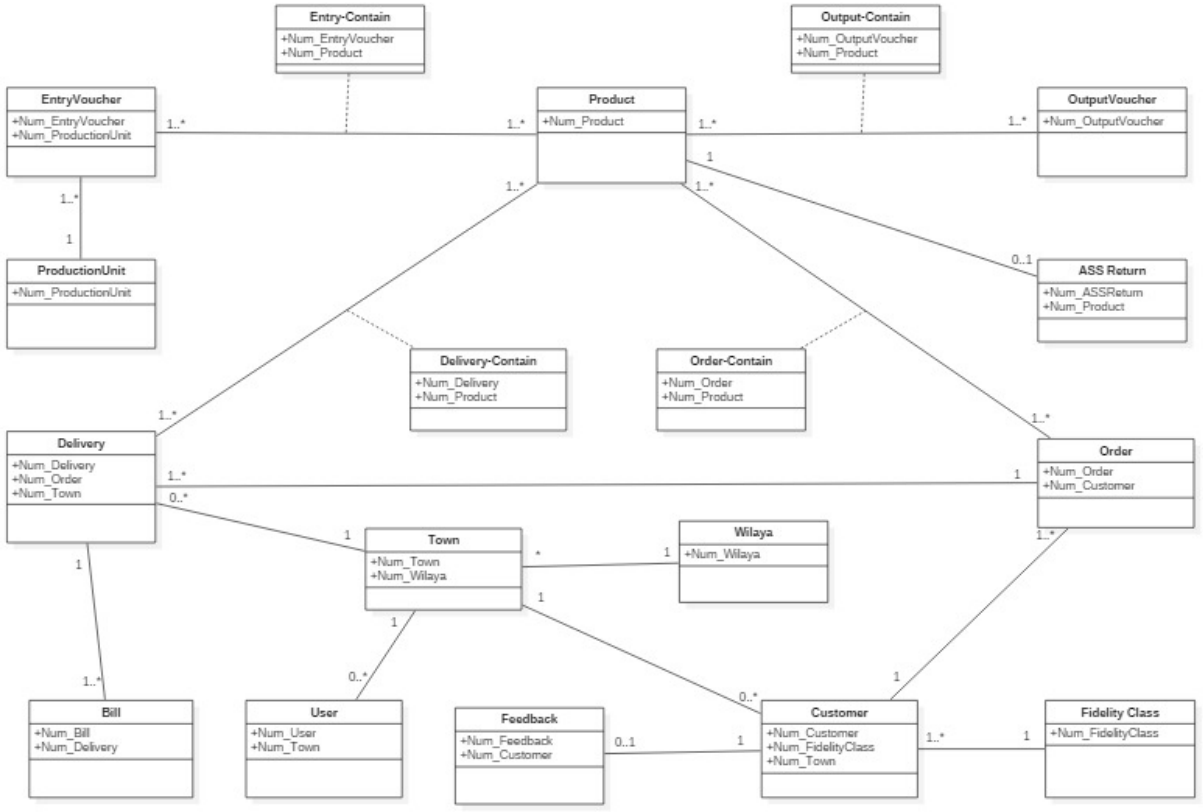


Figure 4.1: Relational Model of CRM application

5. Development tools

We used several tools in our development work including:

5.1. Microsoft SQL Server [69]

Microsoft SQL Server is a Relational Database Management System (RDBMS) developed by Microsoft in 1989. Available under commercial license, a free license is also available but is limited in functionality. It has different editions such as: Compact Edition (CE for smartphone), Express (free version), developer…etc.

5.2. Microsoft Visual Studio [70]

Microsoft Visual Studio is a suite of development software for windows designed by Microsoft. Visual studio is a complete set of development tools for generating ASP.NET web applications, XML web services, desktop applications and mobile applications. It is an integrated development environment (IDE) used by visual basic, visual C#, visual C++, visual J#, this IDE allows them to share tools and facilitate the reaction of solution that uses multiple languages.

5.3. R studio [71]

RStudio is an integrated development environment (IDE) for R programming language; it is available in open source and commercial edition. It runs on Windows, Linux, and Mac or in a browser connected to the R studio server or R studio server Pro.

5.4. DevExpress [72]

DevExpress is a society that works closely with Microsoft. It develops extensions for visual studio to provide new features also to create more ergonomic and pleasant design.

6. Programming languages

In our development the used programming languages are.

6.1. C# programming language [73]

C# (pronounced C-Sharp) is Object Oriented Programming language that was invented by Microsoft in 2000 to build secure and robust applications on their .NET platforms. It is derived from C++ but more close to Java.

6.2. R programming language [74]

R is a language and environment for statistical computing and graphics. It is similar to S programming language which was developed at Bell laboratories. R provides a wide variety of statistical (linear and non linear modeling, classical, statistical tests, clustering, classification…) and graphical techniques. It is available as free software under the terms of the free software foundation. R runs in Windows, Unix, and Mac OS.

7. Frameworks

In the development of our CRM application we used these following frameworks:

7.1. Entity Framework [75]

Entity framework was released the first time in 2008; it is the primary Microsoft means of interaction between .NET applications and relational databases. It is an open source ORM framework for ADO.NET, which is a part of .NET framework. Entity framework has three main approaches: Code First, Model First, and Database First. We are interested in Code First.

7.2. Code First Approach [76]

Code first approach was introduced with entity framework for 4.1 , it allows us to define our model classes which will become the domain model of the application. Then design our database first and then create the classes which match the database design (figure 4.2).

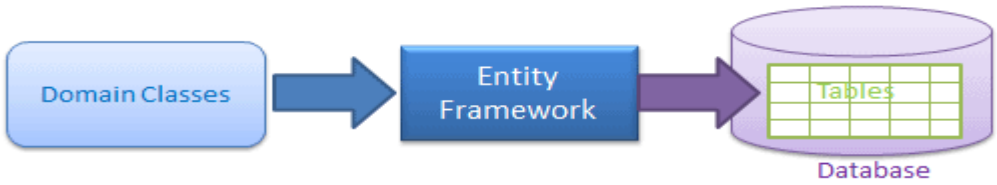


Figure 4.2: Code First approach

The development workflow in code first would be: first create or modify domain classes, then configure these domain classes using data annotation attributes, after this create or update the database schema using automated migration or code based migration as shown in figure below:

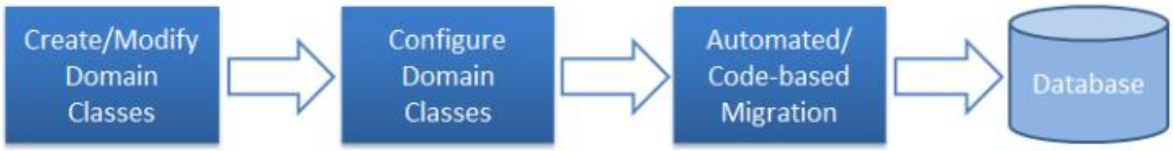


Figure 4.3: Code First workflow

7.3. WPF Application [77]

Windows Presentation Foundation (WPF) is a User Interface (UI) framework that creates desktop applications. It supports many applications development features such as: application model, resources controls, graphics, layout, data binding, documents, and security. WPF uses the Extensible Application Markup Language (XMAL) to provide a declarative model for application programming.

7.4. R.NET.Community [78]

R.NET.Community is the bridge between R programming language and .Net applications. It allows .NET framework to interact with R programming language in the same process. R.NET requires the .NET framework 4 and native R shared libraries installed with R environment.

8. Implementation:

After the presentation of our development environment, we will present in what follows the result of our work through the presentation of the features provided by it.

8.1. Database Creation:

For the creation of the database we used Entity Framework and exactly code first approach. Firstly we defined each table presented in figure 4.1 as a class as shown in the figure 4.4, then we have created two classes which are DbContext (figure 4.5) and DbInitializer (figure 4.6) that allow to Create database and initialize it if there are changes .

```
public class Product
{
    [Key]
    public int NumProduct { get; set; }
    public string Name { get; set; }
    public double Price { get; set; }
    public string Color { get; set; }
    public int GranteePeriod { get; set; }
    public string Type { get; set; }
    public int QuantityInStock { get; set; }
    public int ComStockQuantity { get; set; }
}
```

Figure 4.4: Product class source code

```
public class MarketingEnforcementsDbContext: DbContext
{
    public MarketingEnforcementsDbContext(string connectionString)
        : base(connectionString)
    {
        Database.SetInitializer(new MarketingEnforcementsDbInitializer());
        Database.CreateIfNotExists();
    }
    public virtual DbSet<ASSReturn> ASSReturns { get; set; }
    public virtual DbSet<Bill> Bills { get; set; }
    public virtual DbSet<Customer> Customers { get; set; }
    public virtual DbSet<Delivery> Deliveries { get; set; }
    public virtual DbSet<Feedback> Feedbacks { get; set; }
    public virtual DbSet<FidelityClass> FidelityClasses { get; set; }
    public virtual DbSet<InputVoucher> InputVouchers { get; set; }
    public virtual DbSet<Order> Orders { get; set; }
    public virtual DbSet<OutputVoucher> OutputVouchers { get; set; }
    public virtual DbSet<Product> Products { get; set; }
    public virtual DbSet<ProductionUnit> ProductionUnits { get; set; }
    public virtual DbSet<User> Users { get; set; }
    public virtual DbSet<Wilaya> Wilayas { get; set; }
    public virtual DbSet<Town> Towns { get; set; }
    public virtual DbSet<EntryContain> EntryContains { get; set; }
    public virtual DbSet<OutputContain> OutputContains { get; set; }
    public virtual DbSet<OrderContain> OrderContains { get; set; }
    public virtual DbSet<DeliveryContain> DeliveryContains { get; set; }
    protected override void OnModelCreating(DbModelBuilder modelBuilder)
    {
        modelBuilder.Entity<EntryContain>().HasKey(ev => new { ev.NumProduct,ev.NumInputVoucher });
        modelBuilder.Entity<OutputContain>().HasKey(ev => new { ev.NumProduct, ev.NumOutputVoucher });
        modelBuilder.Entity<OrderContain>().HasKey(ev => new { ev.NumProduct, ev.NumOrder });
        modelBuilder.Entity<DeliveryContain>().HasKey(ev => new { ev.NumProduct, ev.NumDelivery });
        modelBuilder.Entity<ASSReturn>().HasOptional(ev => ev.Product);
        Configuration.ProxyCreationEnabled = false;
        base.OnModelCreating(modelBuilder);
        modelBuilder.Conventions.Remove<System.Data.Entity.ModelConfiguration.Conventions.PluralizingTableNameConvention>();
    }
```

Figure 4.5: DbContext source code

```
public class MarketingEnforcementsDbInitializer : DropCreateDatabaseIfModelChanges<MarketingEnforcementsDbContext>
{
    protected override void Seed(MarketingEnforcementsDbContext context)
    {
        context.SaveChanges();
    }
}
```

Figure 4.6: DbInitializer source code

8.2. Splash Screen and loading data

Splash screen is a window that loads data from a database.

```
private async void SplashScreen_OnLoaded(object sender, RoutedEventArgs e)
{
    var connectionStringMainValue = GeneralSettings.GetSettingValue("ConnectionString");
    try
    {
        SharedBll.Db = new MarketingEnforcementsDbContext(connectionStringMainValue);
    }
    catch (Exception sqlException)
    {
        DXMessageBox.Show(sqlException.ToString());
        Application.Current.Shutdown();
        return;
    }
    var x = SharedBll.Db.ASSReturns.LoadAsync();  await x;
    x = SharedBll.Db.Bills.LoadAsync();  await x;
    x = SharedBll.Db.Customers.LoadAsync();  await x;
    x = SharedBll.Db.Deliveries.LoadAsync();  await x;
    x = SharedBll.Db.Feedbacks.LoadAsync();  await x;
    x = SharedBll.Db.FidelityClasses.LoadAsync(); await x;
    x = SharedBll.Db.InputVouchers.LoadAsync(); await x;
    x = SharedBll.Db.EntryContains.LoadAsync();
    await x;
    x = SharedBll.Db.OutputContains.LoadAsync();  await x;
    x = SharedBll.Db.Orders.LoadAsync();  await x;
    x = SharedBll.Db.OutputVouchers.LoadAsync();  await x;
    x = SharedBll.Db.ProductionUnits.LoadAsync();  await x;
    x = SharedBll.Db.Products.LoadAsync();  await x;
    x = SharedBll.Db.Users.LoadAsync(); await x;
    x = SharedBll.Db.Wilayas.LoadAsync();  await x;
    x = SharedBll.Db.Towns.LoadAsync(); await x;
    x = SharedBll.Db.OrderContains.LoadAsync(); await x;
    x = SharedBll.Db.DeliveryContains.LoadAsync(); await x;
    var login = new LoginWindow();
    login.Show();
    Hide();
}
```

Figure 4.7: Loading data source code

Figure 4.8: Splash Screen

8.3. Login Window

Login window tests the entered user data, if this data is true the main window is shown, else the error message is displayed.

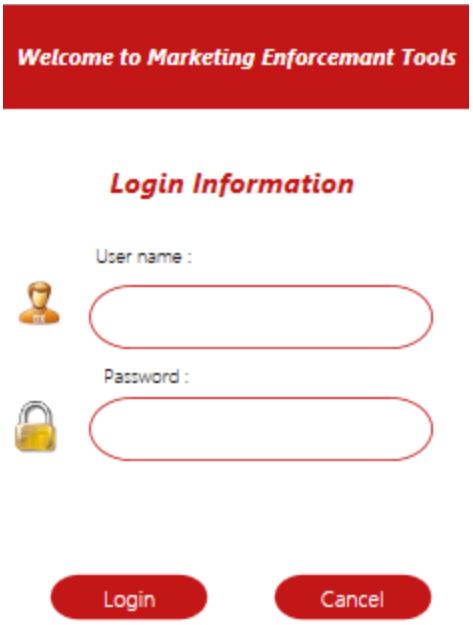


Figure 4.9: Login window

```
private void LoginButton_Click(object sender, RoutedEventArgs e)
{
    var found = false;
    IQueryable<User> UserList = UserBll.GetAll();
    foreach(User user in UserList)
    {
        if(user.UserName == username.Text && user.Password == password.Password)
        {
            found = true;
            SharedBll.User = user;
            break;
        }
    }
    if(found == true)
    {
        var mainwindow = new MainWindow();
        mainwindow.Show();
        Hide();
    }
    else
    {
        DXMessageBox.Show("Authentification Error");
        username.Text = "";
        password.Password = "";
        x++;
        if (x == 3)
        {
            Application.Current.Shutdown();
        }
    }
}
```

Figure 4.10: Login source code

8.4. Main Window

Main Window shows the different spots of the application.

8.4.1. Customer Management

In this section we can Add, Edit, and Delete customers from the database, Add Customer feedback, and add fidelity class.

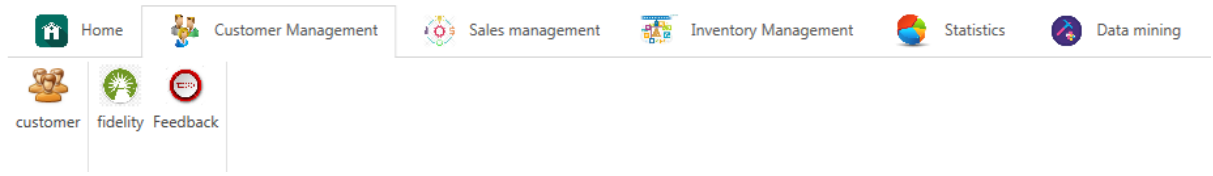


Figure 4.11: Customer management spots

Add, edit, delete customer interface are shown in the figure below:

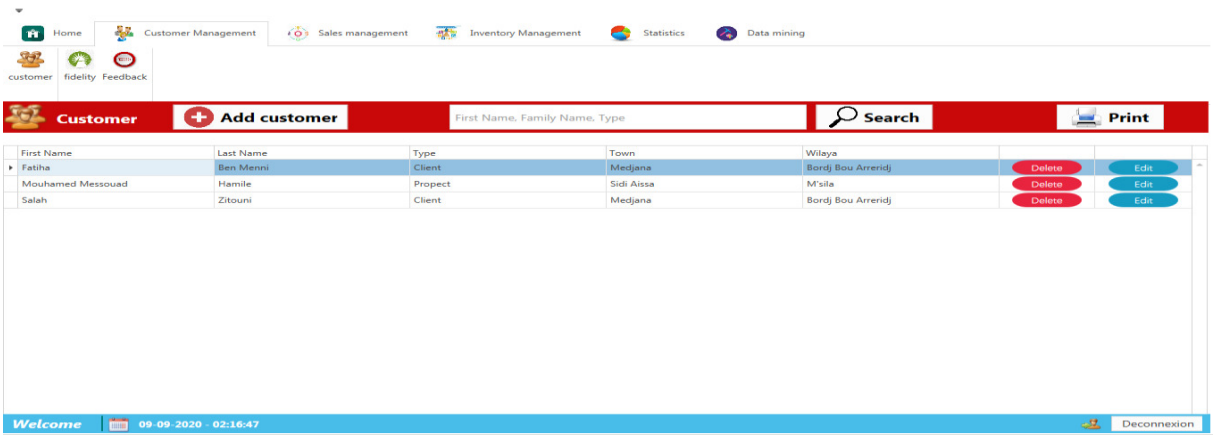


Figure 4.12: Customer management interface

Add customer feedback interface is shown in the figure below:

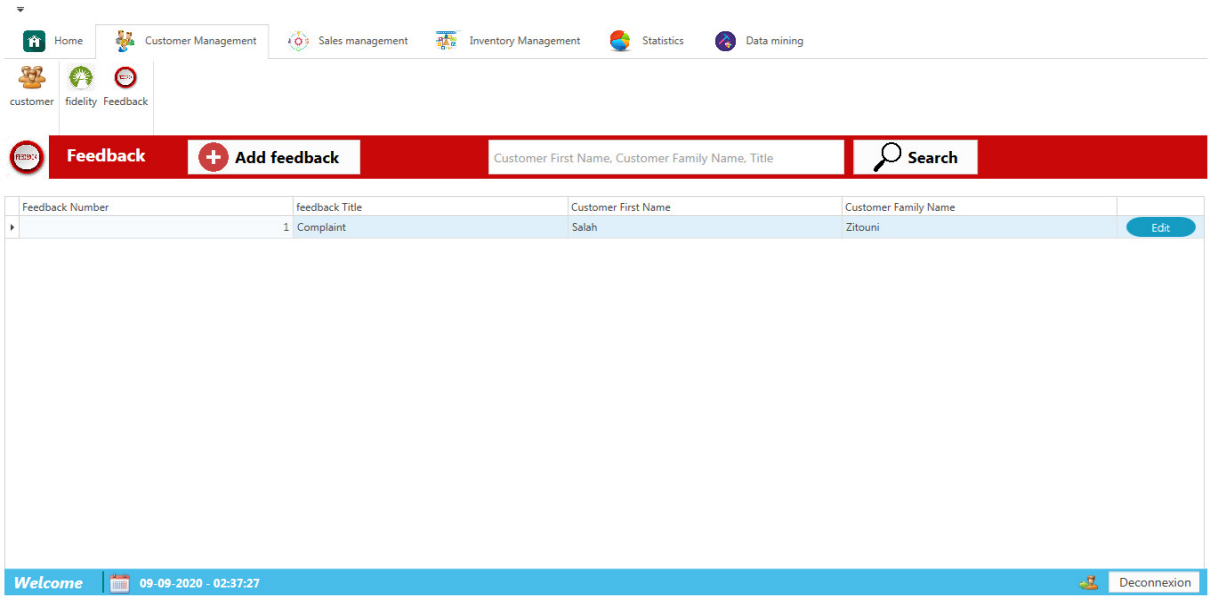


Figure 4.13: Feedback management interface

Add Fidelity Class interface is shown in the figure below:

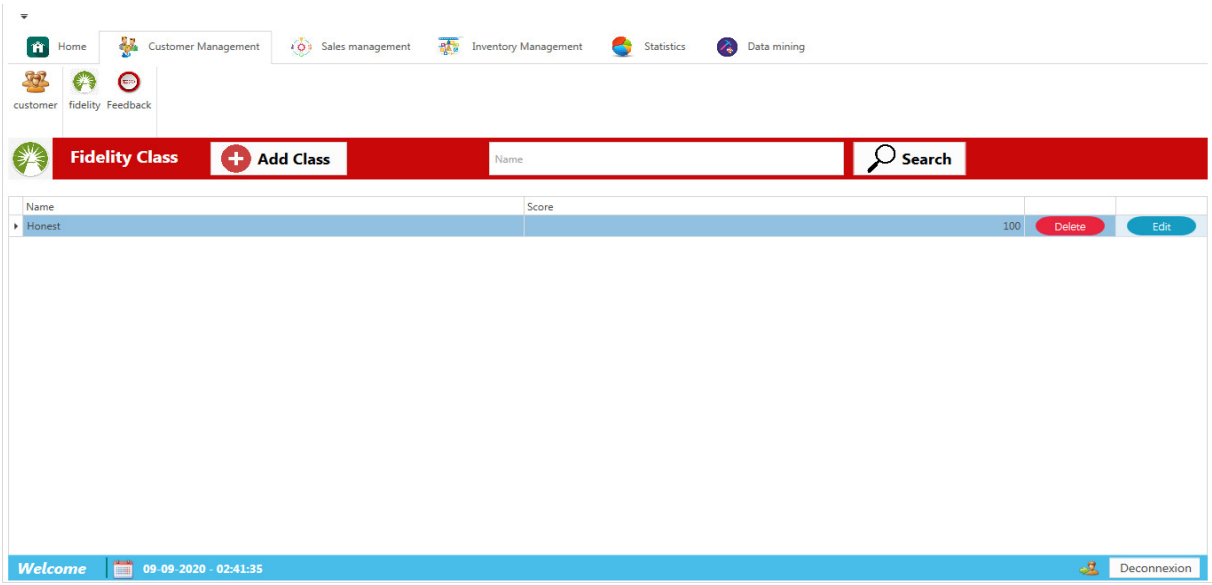


Figure 4.14: Fidelity class management interface

8.4.2. Inventory Management

In this section we can check inventory level, add a new type of product, and create a new input and output voucher.

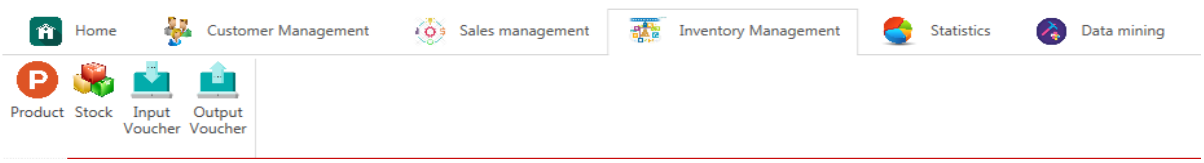


Figure 4.15: Inventory management spots

Add New Product is shown in the figure below:

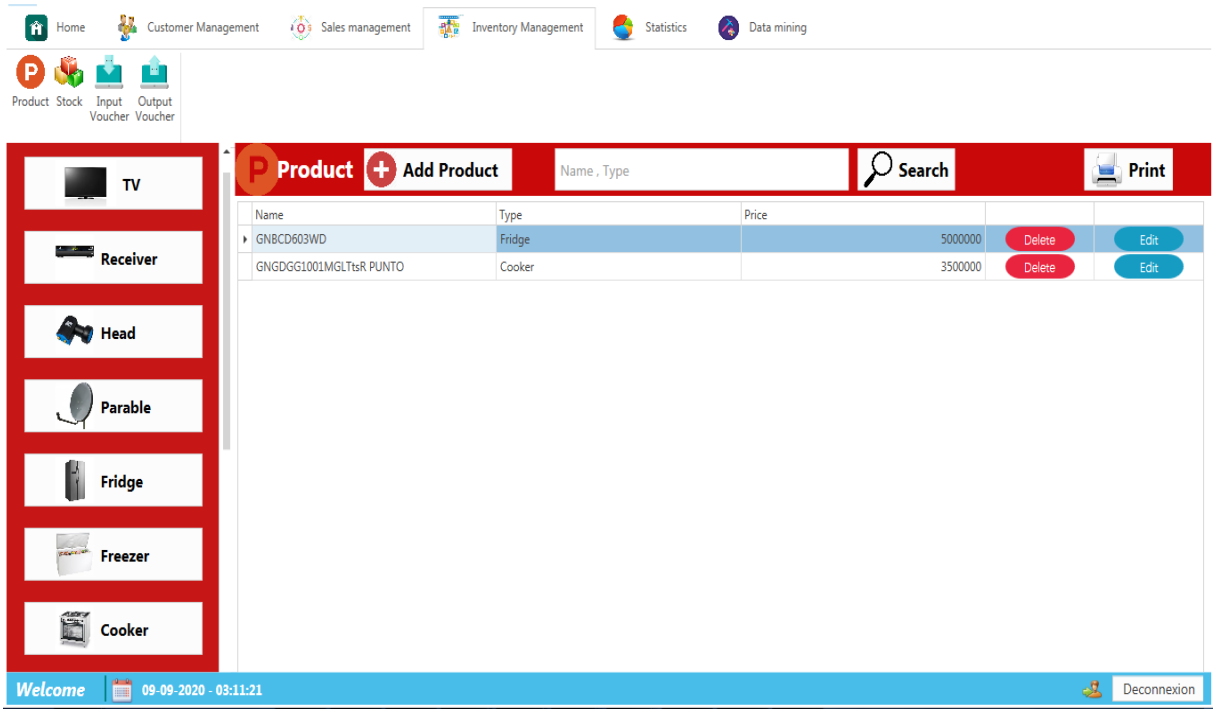


Figure 4.16: Product management interface

Check Inventory Level interface is show in the figure below:

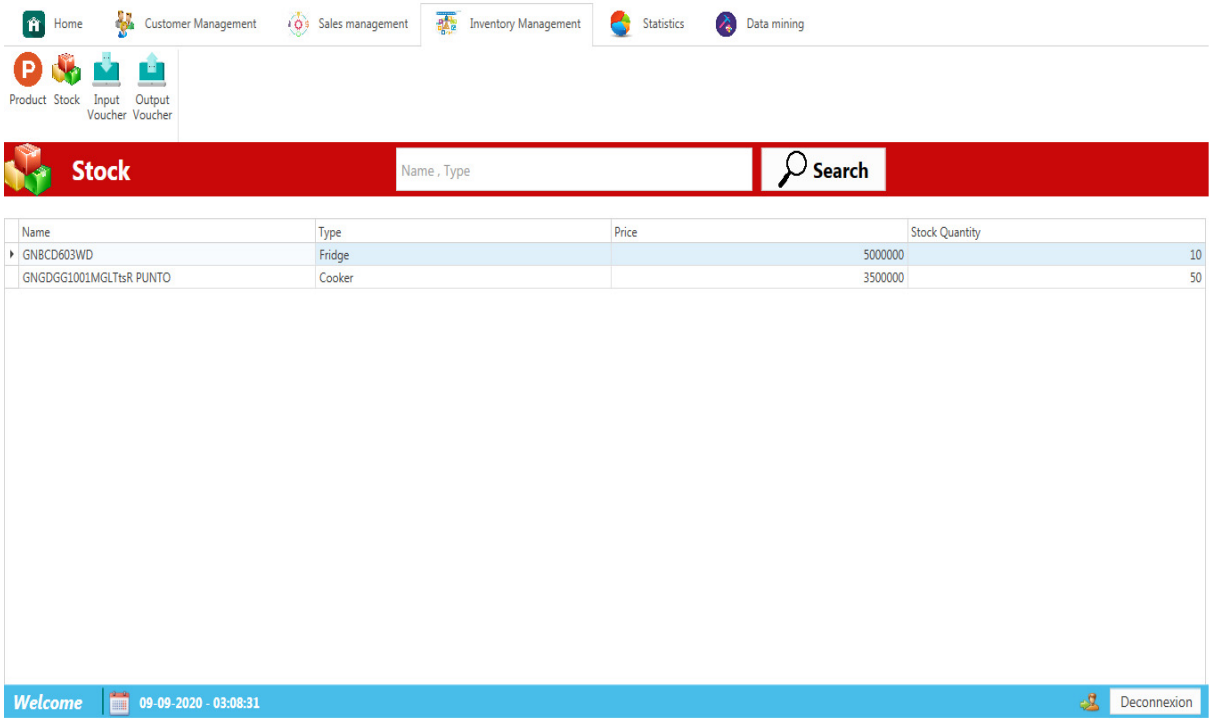


Figure 4.17: Display inventory level interface

Input Voucher management interface is shown in the figure below:

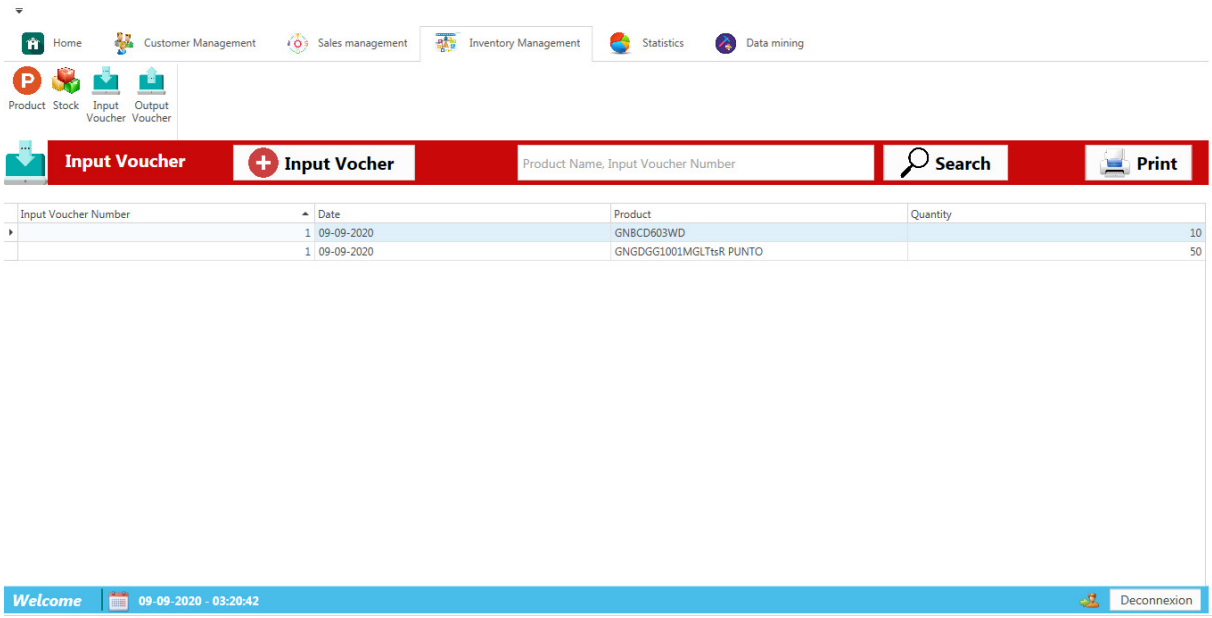


Figure 4.18: Input voucher management interface

Output Voucher management interface is shown in the figure below:

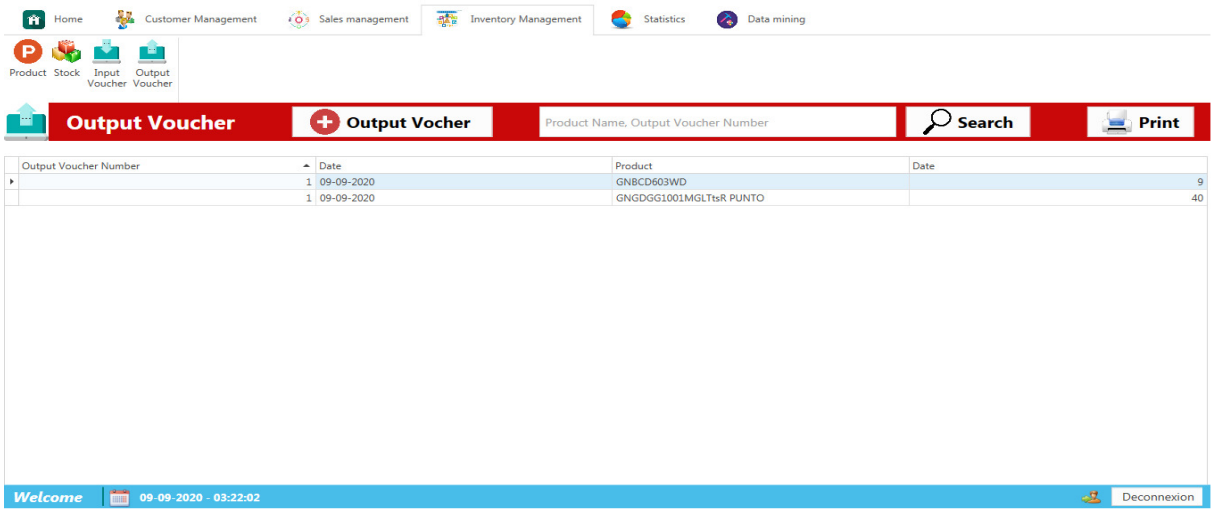


Figure 4.19: Output voucher management interface

8.4.3. Classical statistics

In this section we can divide our customer data by different criteria, classify products, and Count sales revenue.

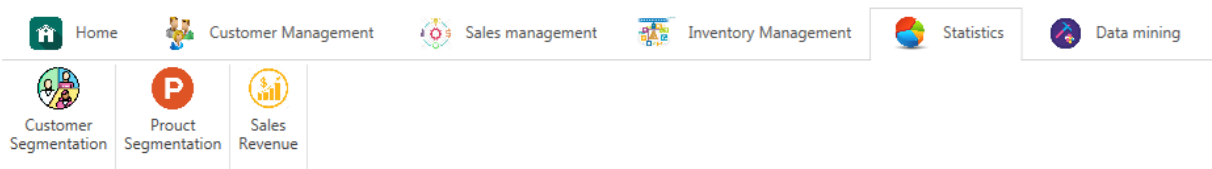


Figure 4.20: Classical statistics

8.4.3.1. Customer Segmentation

Classify Customer by Age interface is shown in the figure below:

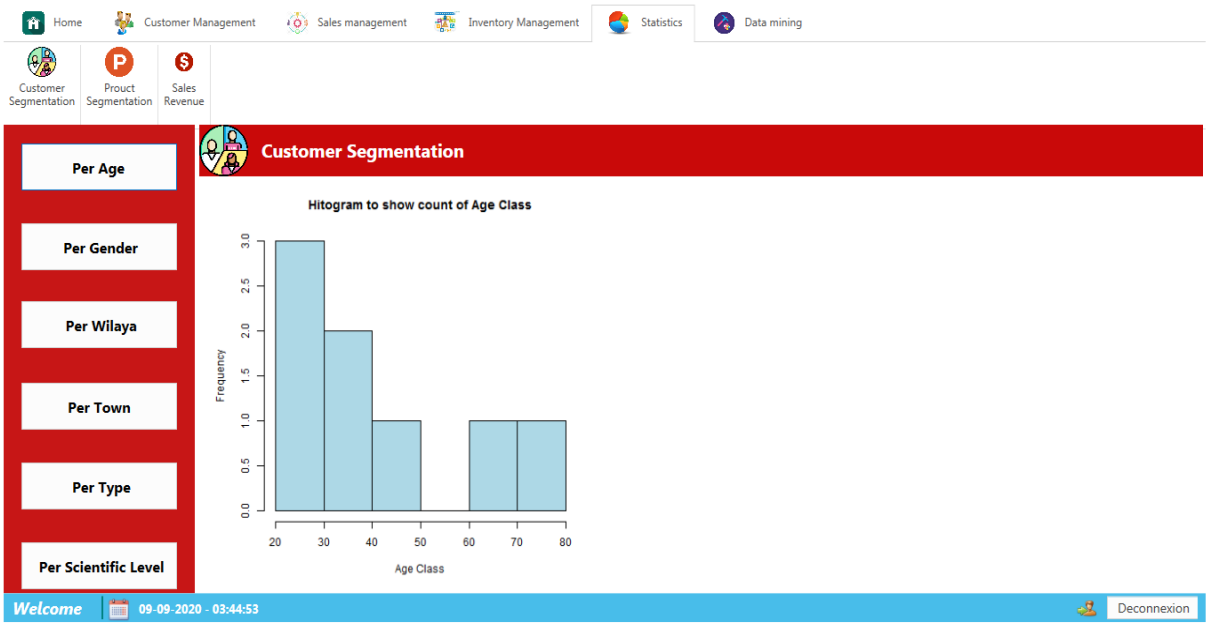


Figure 4.21: Classifying customer by age

Classify Customer by Gender interface is shown in the figure below:

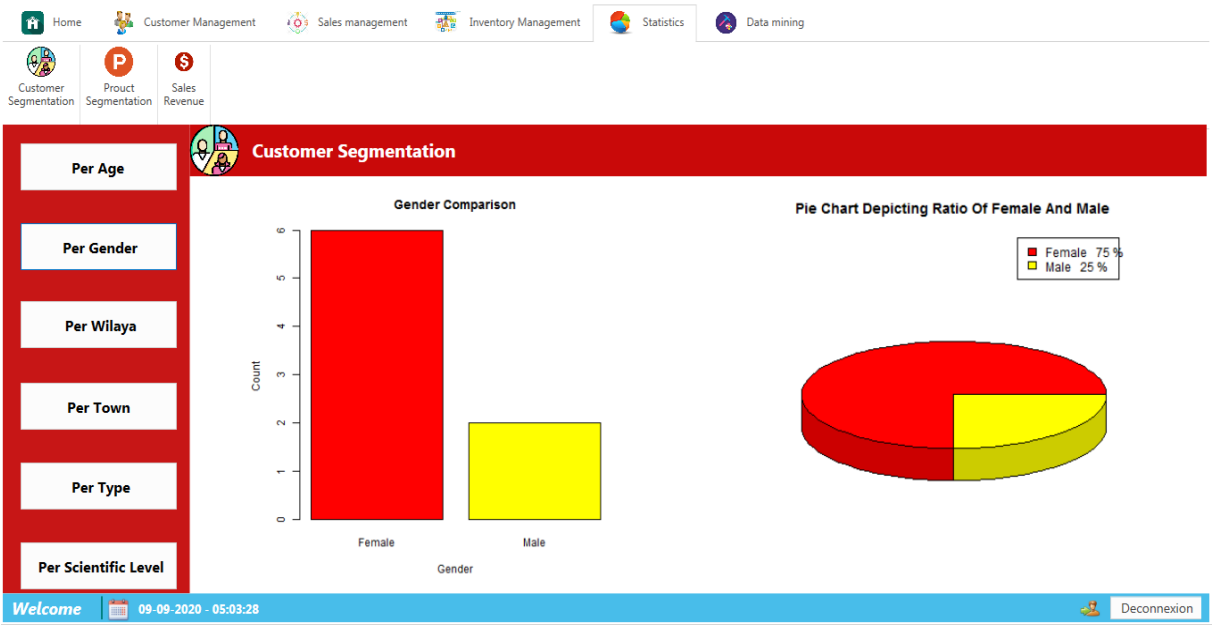


Figure 4.22: Classify customer by gender

Classify Customer by Type interface is shown in the figure below:

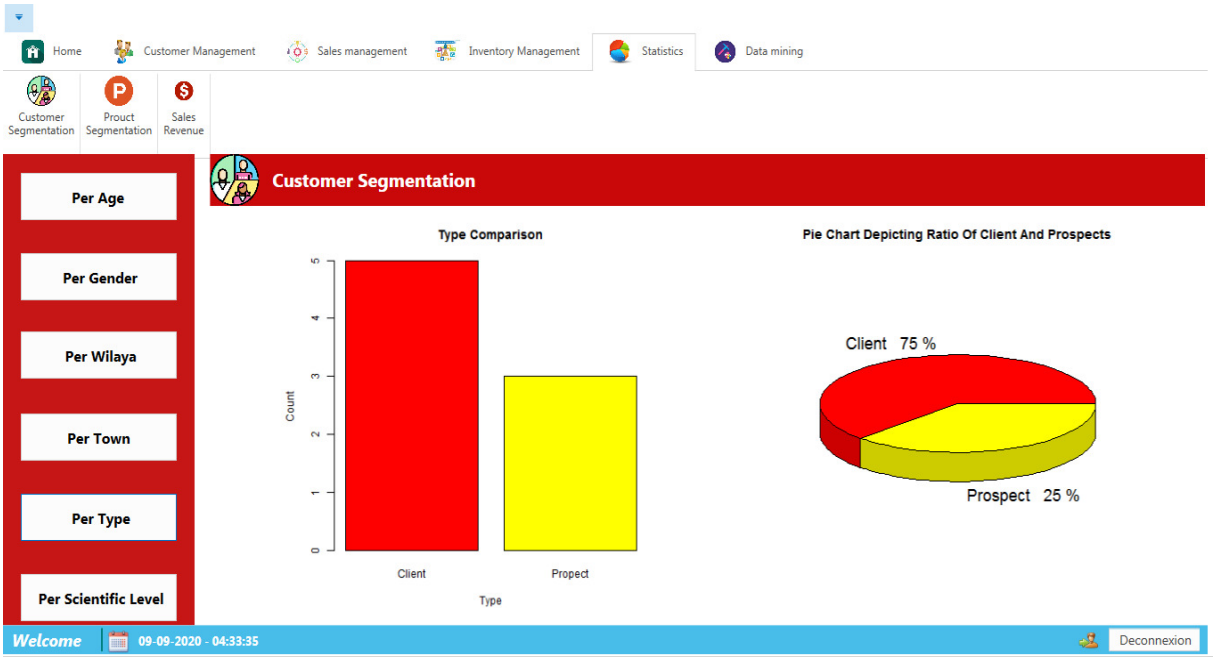


Figure 4.23: Classify customer by Type

Classify Customer by wilaya interface is show in the figure below:

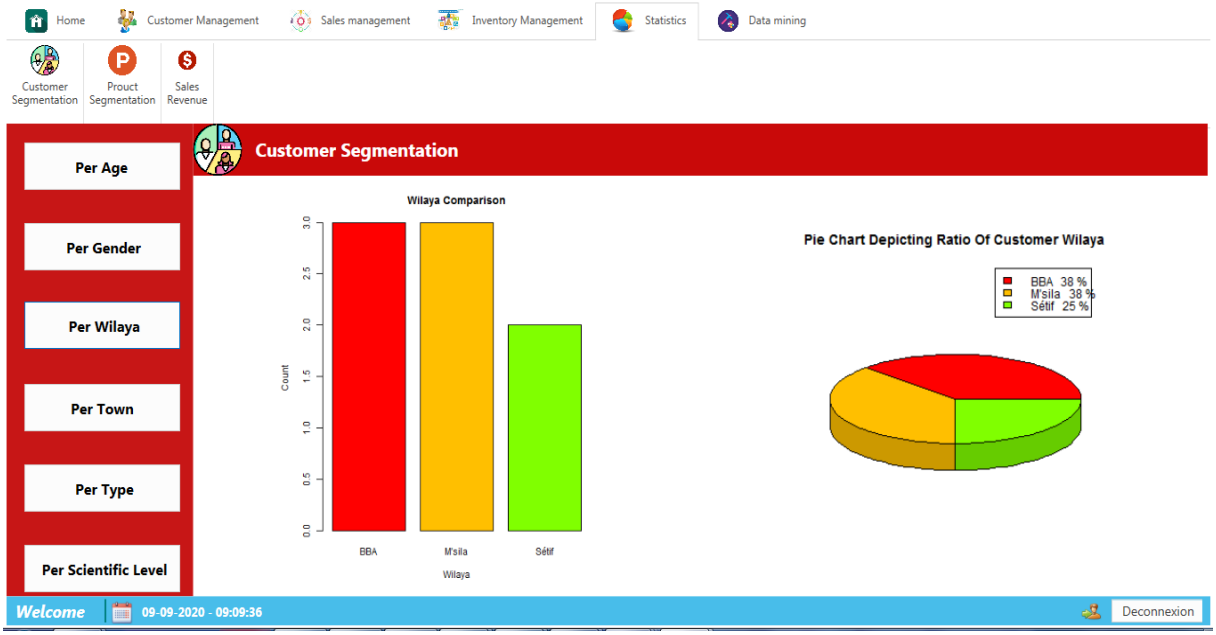


Figure 4.24: Classify customer by wilaya

The interface of classifying Customer by Town is shown in the figure below:

Figure 4.25: Classify customer by town

The interface of classifying customer by scientific level is shown in the figure below:

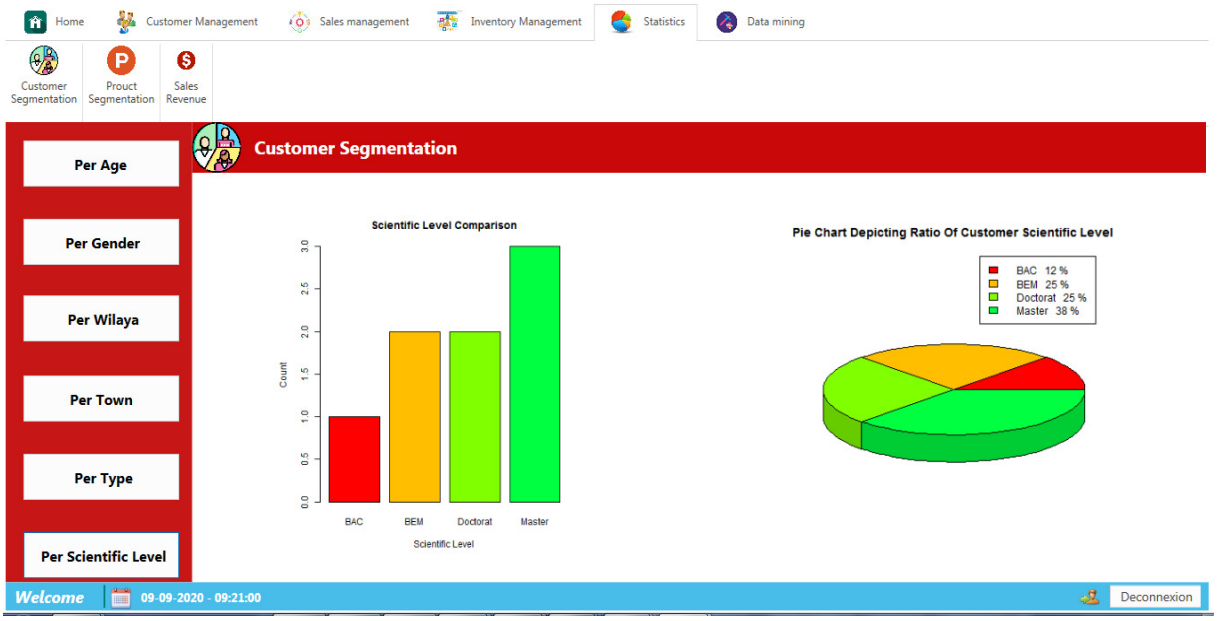


Figure 4.26: Classify customer by scientific level

8.4.3.2. Product Classification

Classify the selling products by type in a specific month is shown in the figure below:

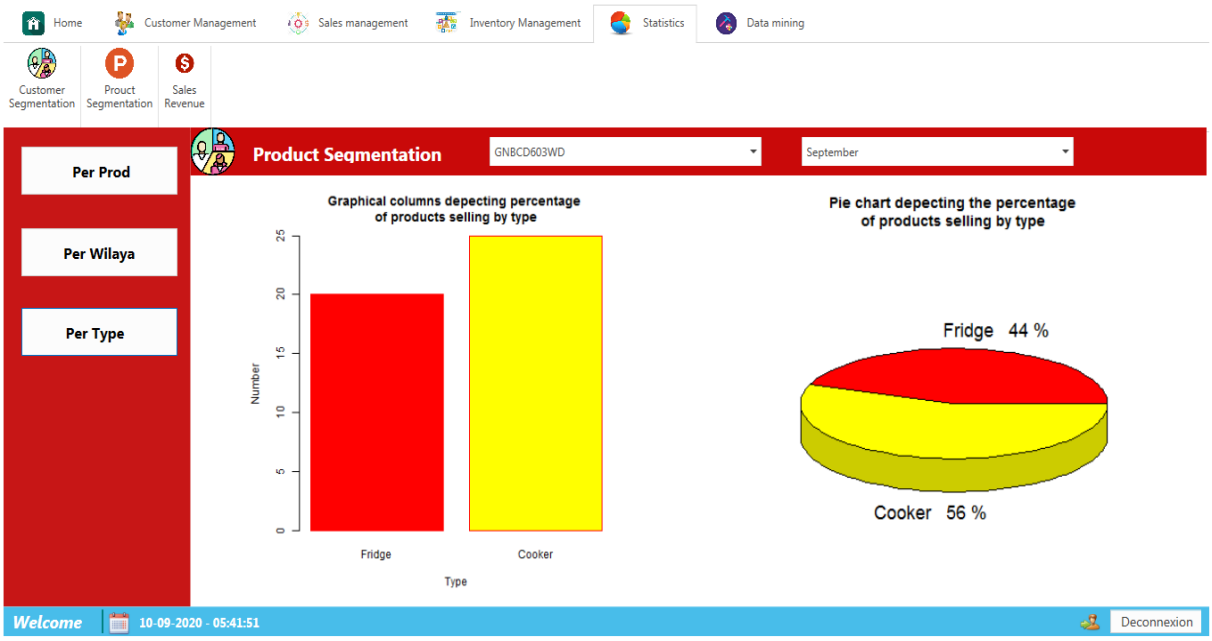


Figure 4.27: Classify the products sold by type

Classify the number of products sold in each wilaya is shown in the figure below:

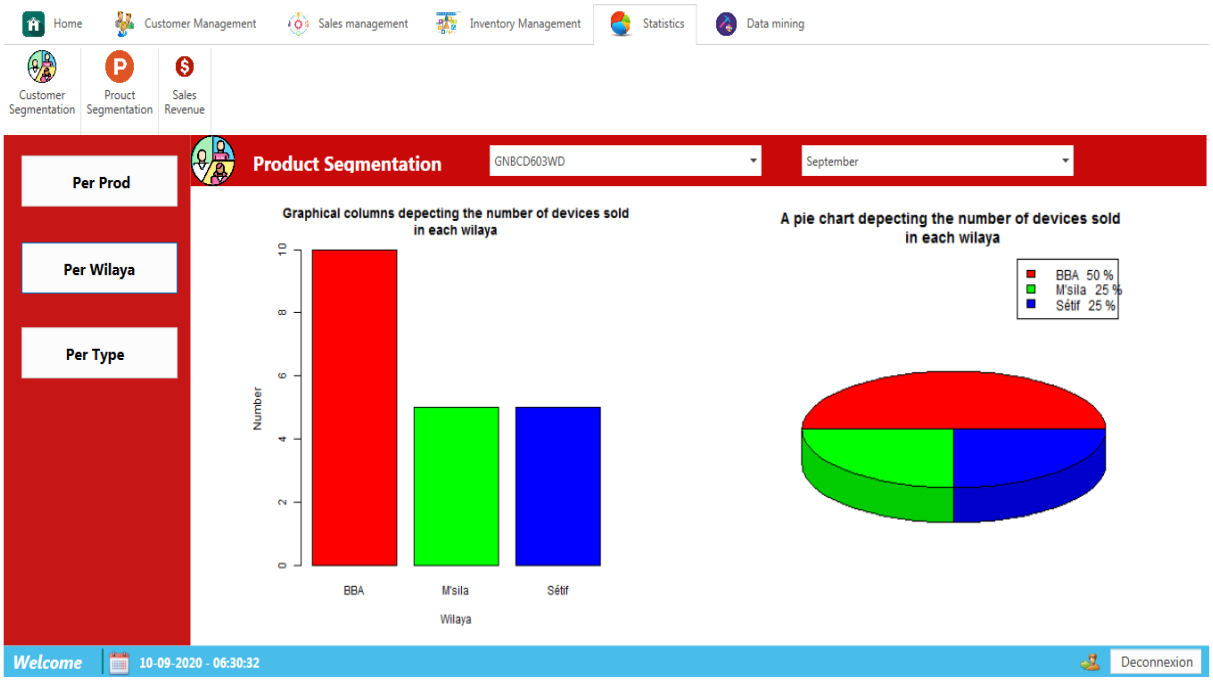


Figure 4.28: Classify the number of products sold in each wilaya

Classify the quantity of products sold by the day of the month is shown in the figure below:

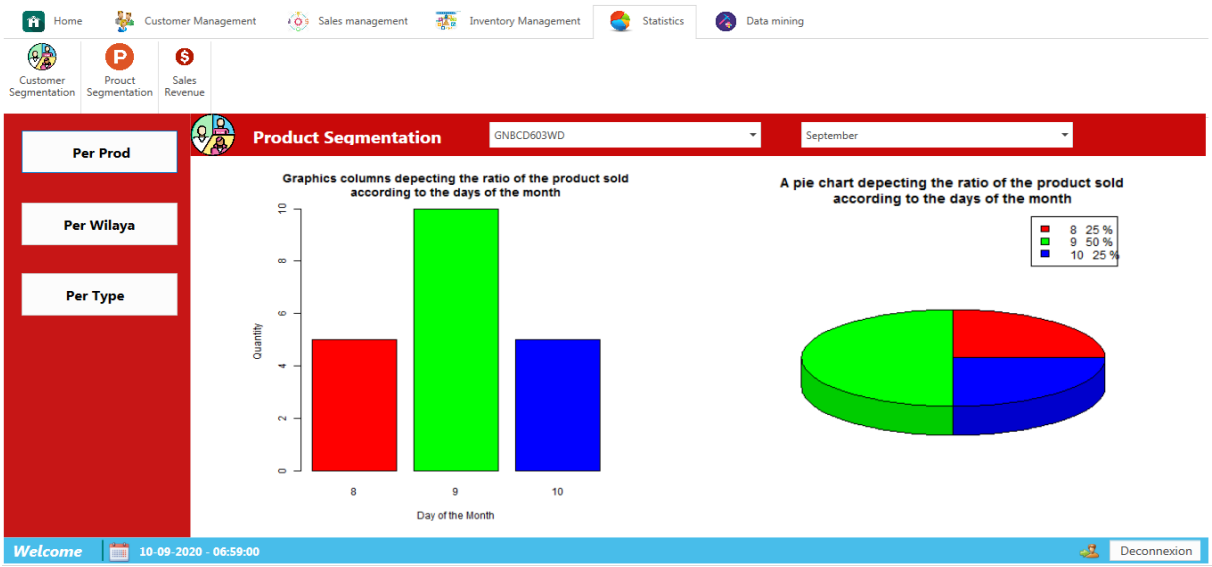


Figure 4.29: Classify the quantity of products sold by month

8.4.4. Data mining statistics:

In this section we can do dynamic statistics by using data mining techniques which are: classification, clustering, Association Rule and Regression.

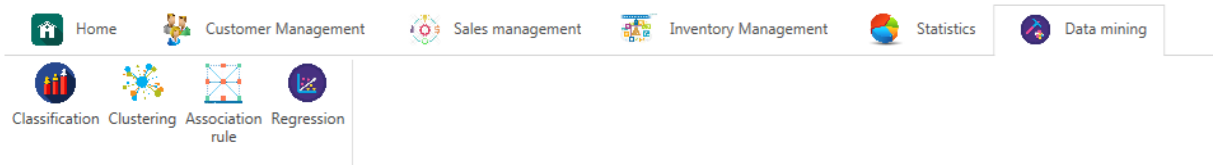


Figure 4.30: Dynamic statistics

8.4.4.1. Classification

The result of the classification algorithm is shown in the figure below:

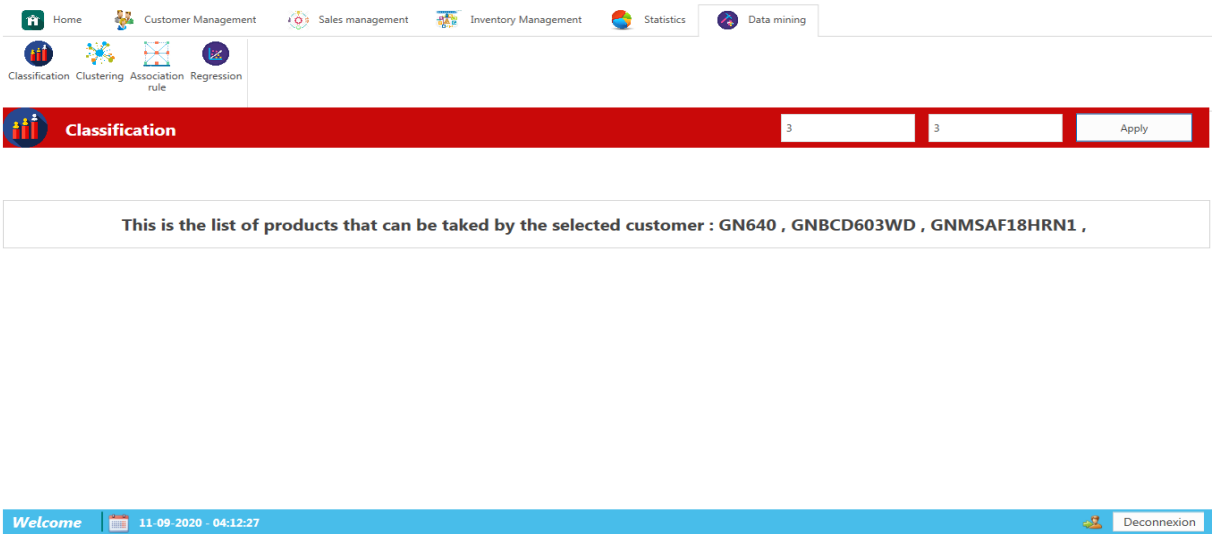


Figure 4.31: Classification result

8.4.4.2. Clustering

The result of the clustering is shown in the figure below:

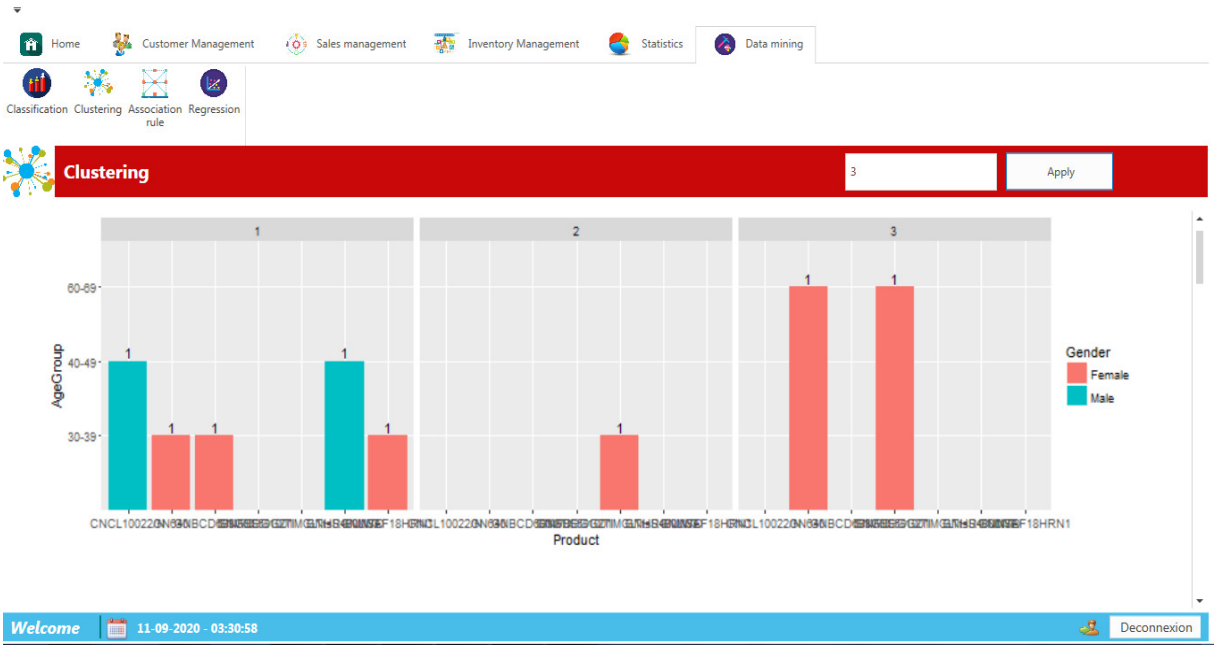


Figure 4.32: Clustering result

8.4.4.3. Association Rule

The result of association rule algorithm is shown in the figure below:

Figure 4.33: Association rule result

8.4.4.4. Regression

Predict the number of sales revenue for the next year is shown in the figure below:

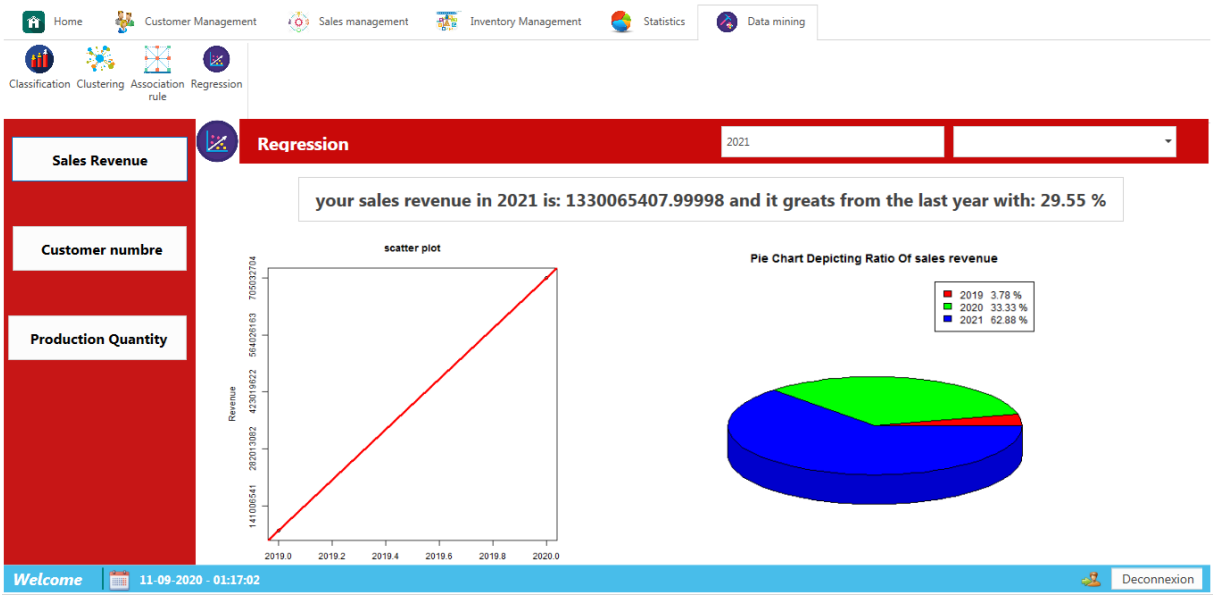


Figure 4.34: Applying regression algorithm to predict sales revenue

Predict the number of customer in the next year is shown in the figure below:

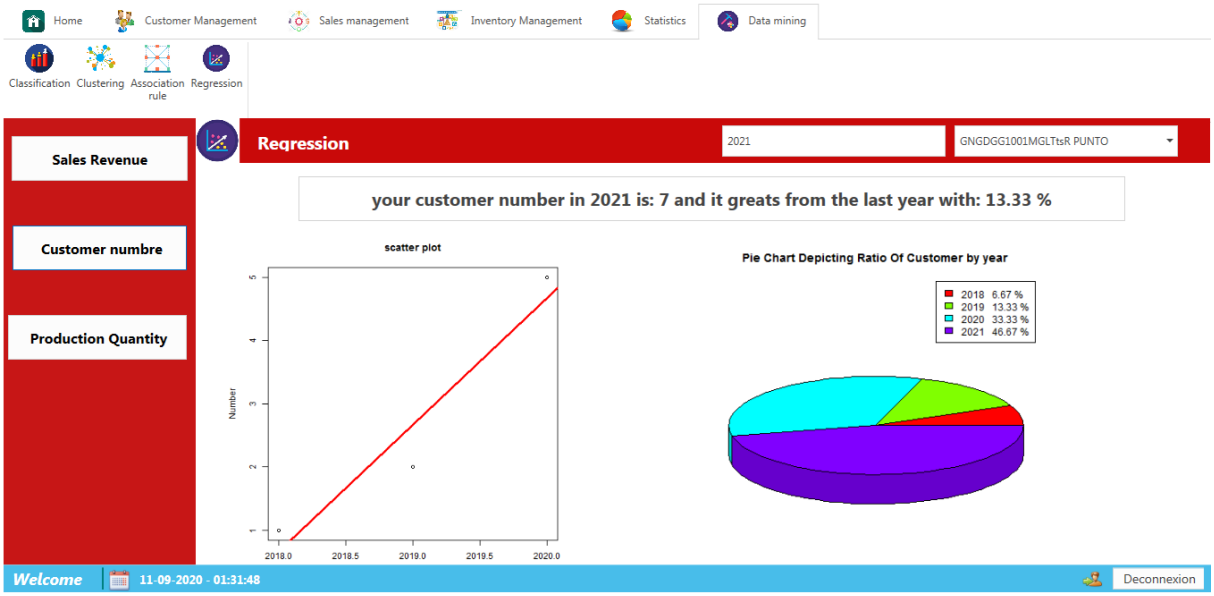


Figure 4.35: Applying regression algorithm to predict customer number

Predict the production quantity of product I shown in the figure below:

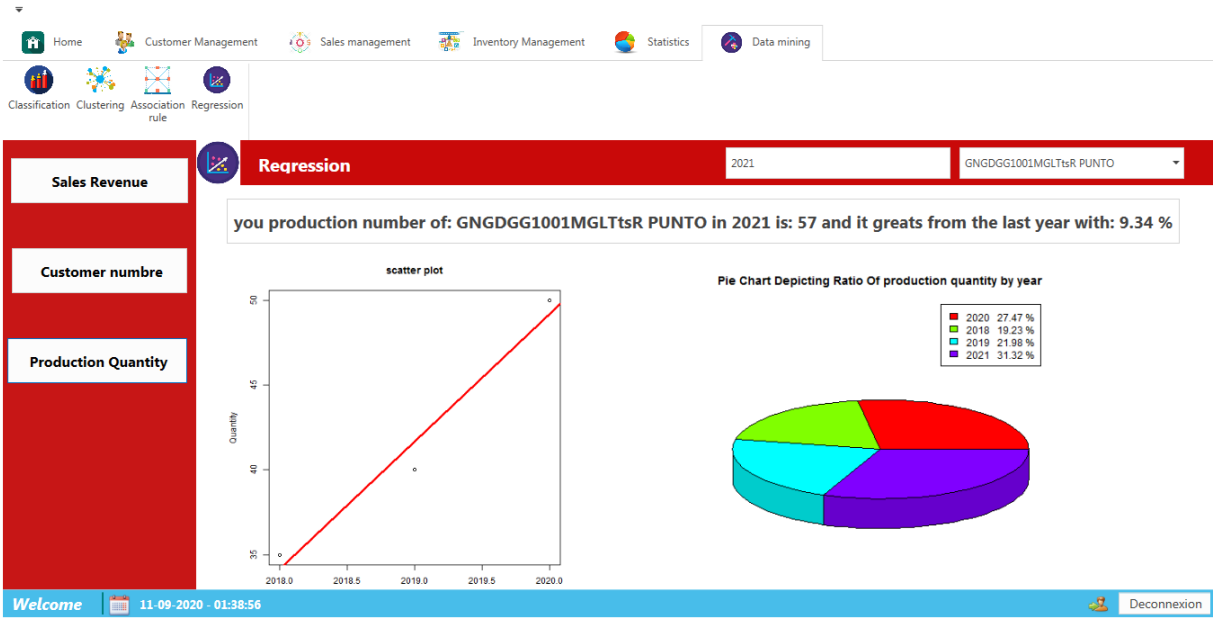


Figure 4.36: Applying Regression algorithm to predict production quantity

9. Conclusion

In this Chapter, we have presented our Relational Model resulting from the UML class diagram. We have also presented the different tools used to develop our application. Finally we have shown the different spots of the application through the presentation of some screenshots.

GENERAL CONCLUSION

General Conclusion

In economic context, data mining plays a pivotal role in helping organizations to understand their customers, anticipate their stocks and optimize their sale policies as well as other benefits which bring a considerable competitive advantage to the organization. All these advantages reinforce good decision making.

In this perspective we have developed to the Géant Electronics Company a customer management application equipped with a decision support tool. This application depends on the use of data mining techniques in order to analyze data and provide a predictive picture to the company in the coming years. The development process that we have adopted goes through three phases which are:

Firstly, we have modeled all the functions listed in the analysis phase by the UML modeling language.

Secondly, we used C# language in visual studio IDE to develop the principal spots in customer management, sales management and inventory management.

Thirdly, we used data mining techniques such as classification, clustering, association rules and regression to develop a decision support tool. In this phase we technically explore the R programming language in R studio IDE by ensuring the link between them through R.NET.Community.

For the improvement of this work in the future we suggest the following points:

- Taking into account all the necessary data circulating in the company.
- Taking into account all the necessary data circulating outside the company (competing companies).
- The addition of additional data mining modules to fully understand the market and understand the marketing process.

Finally, we hope that this work will contribute to the establishment of the culture of use of data mining in our Algerian companies.

Bibliography:

[1] Bruno Estevez. (2014, Novembre) Informatique Décisionnelle Opérationnelle.

[2] Jayanthi Ranjan, "Business Intelligence : concepts, components, techniques and benifits," *Theoretical and Applied Information Technology*, vol. 9, no. 1, pp. 63-64.

[3] Maria Miromind. SOFTWAREWORLD. [Online]. https://www.softwareworld.co/understanding-the-major-components-of-business-intelligence/

[4] cheick Sanogo. (2015, novembre) supinfo. [Online]. https://www.supinfo.com/articles/single/1182-vue-ensemble-architecture-system-bi

[5] Youssra Riyahi, "Business Intelligence: A Strategy for Business Development," *Economics and Management Studies (SSRG-IJEMS)*, vol. 4, no. 2, p. 02, septembre 2017.

[6] Jonas DeMuro and Brian Turner. (2019, Octobre) techradar. [Online]. https://www.techradar.com/best/best-bi-tools

[7] Timothy Chee et al., "Progress in Information & Communication Technology," in *BUSINESS INTELLIGENCE SYSTEMS: STATE-OF-THE-ART REVIEW AND CONTEMPORARY APPLICATIONS*, Malaysia, Janvier 2009, pp. 99-100.

[8] Julia Miles. (2019, septembre) my assignment services. [Online]. https://www.myassignmentservices.com/blog/benefits-and-components-of-business-intelligent-systems/

[9] Géraldine Graf and Julien Stern, "Customer Relationship Management," in *Le CRM analytique Les outils d'analyse OLAP et le Data Mining*, Fribourg, 2008, p. 13.

[10] Mahadev Ramageri Bharati, "DATA MINING TECHNIQUES AND APPLICATIONS ," *Computer Science and Engineering* , vol. 1, no. 4, p. 301, décembre 2010.

[11] Osmar R. Zaïane, Introduction to Data Mining, 1999.

[12] Said Gadri, *Multilingual Text Categorization*. M'sila, Alger: Noor publishing, 2017.

[13] Aruna J. Chamatkar and P.K Butey, "Importance of Data Mining with Different Types of Data Applications and Challenging Areas," *Engineering Research and Applications*, vol. 4, no. 5, p. 39, Mai 2014.

[14] Sethunya R Joseph, Hlomani Hlomani, and Keletso Letsholo, "Data Mining Algorithms: An Overview," *computers and technology*, vol. 15, no. 6, p. 6807, avril 2016.

[15] Sudhir M Gorade, Ankit Deo, and Preetesh Purohit, "A Study of Some Data Mining Classification Techniques," *International Research Journal of Engineering and Technology*, vol. 4, no. 4, p. 3112, avril 2017.

[16] S Neelamegam and E Ramaraj, "Classification algorithm in Data mining: An Overview," *International Journal of P2P Network Trends and Technology (IJPTT)*, vol. 3, no. 5, p.

2, Septembre 2013.

[17] Osmar R Zaïane. (1999) Data Classification.

[18] MaoFei Wenji and Wang Yue, "Cultural Modeling for Behavior Analysis and Prediction," in *in new advances in intelligent and security informatics*., Decembre 2012.

[19] Avinash Navlani. (2018, Decembre) DataCamp. [Online]. https://www.datacamp.com/community/tutorials/naive-bayes-scikit-learn

[20] Adi Bronshtein. (2017, avril) NoteWorthy. [Online]. https://blog.usejournal.com/a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7

[21] Dhilip Subramanian. (2019, Juin) towards data science. [Online]. https://towardsdatascience.com/a-simple-introduction-to-k-nearest-neighbors-algorithm-b3519ed98e

[22] Tavish Srivastava. (2018, Mars) Analytics Vidhya. [Online]. https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/

[23] BRIAN BEERS. (2020, féverier) Investopedia. [Online]. https://www.investopedia.com/terms/r/regression.asp

[24] Swati Gupta, "A Regression Modeling Technique on Data Mining," *International Journal of Computer Applications*, vol. 116, no. 9, p. 27, avril 2015.

[25] Tauqeer Khurram. (2018, Mars) Tech Funnel. [Online]. https://www.techfunnel.com/information-technology/regression-analysis-types/

[26] Jimmie Pearson. (2019, Juin) Urban Seminary. [Online]. http://www.urbanaseminary.org/types-of-regression-analysis/

[27] Pradeep Kumar Singh, "BVICAM 2016 IEEE," in *Clustering Techniques in Data Mining: A Comparaison*, Delhi, 2015, p. 410.

[28] Santosh Nirmal, "Comparative Study between K-Means and K-Medoids Clustering Algorithms," *International Research Journal of Engineering and Technology (IRJET)*, vol. 6, no. 3, p. 840, mars 2019.

[29] Osmar R Zaïane. (1999) Data Clustering.

[30] Alboukadel Kassambara. Data Novia. [Online]. https://www.datanovia.com/en/lessons/k-medoids-in-r-algorithm-and-practical-examples/

[31] Lior Rokach and Oded Maimon, Clustering Methods, 2005.

[32] Samarjeet Borah and Rupanka Bhuyan, "National Conference on Advancements in Information, Computer and Communication (AICC-2013)," in *A Survey of Some Density Based Clustering Techniques*, Rajasthan, India, 2013.

[33] Trupti A. Kumbhare and Santosh V Chobe, "An Overview of Association Rule Mining Algorithms," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 1, p. 928, 2014.

[34] Irina Tudor, Association Rule Mining as a Data Mining Technique, 2008.

[35] Margaret H Dunham, Yongqiao Xiao, Le Gruenwald, and Zahid Hossain, A SURVEY OF ASSOCIATION RULES.

[36] Thabet Slimani and amor Lazzez. Efficient Analysis of Pattern and Association Rule Mining Approaches.

[37] Said Sayad. [Online]. https://www.saedsayad.com/association_rules.htm

[38] Wide Skills00. [Online]. http://www.wideskills.com/data-mining-tutorial/data-mining-architecture

[39] chandan goopta. (2014, octobre) THENEWSTACK. [Online]. https://thenewstack.io/six-of-the-best-open-source-data-mining-tools/

[40] (2020, Avril) Guru99. [Online]. https://www.guru99.com/data-mining-tutorial.html

[41] P Rajkumar. (2014, Août) Big Data Made Simple. [Online]. https://bigdata-madesimple.com/14-useful-applications-of-data-mining/

[42] JDN. (2019, Janvier) JDN. [Online]. https://www.journaldunet.fr/business/dictionnaire-du-marketing/1197909-adv-administration-des-ventes-definition-traduction-et-synonymes/

[43] Manager Go. (2020, Septembre) Manager Go. [Online]. https://www.manager-go.com/marketing/

[44] LAURA LAKE. (2019, Octobre) thebalancesmb. [Online]. https://www.thebalancesmb.com/what-is-marketing-2296057

[45] M Prachi. (2019, juillet) the investors book. [Online]. https://theinvestorsbook.com/sales-management.html

[46] Samantha Mur. (2020, Mai) appvizer. [Online]. https://www.appvizer.fr/magazine/relation-client/support-client/service-apres-vente

[47] Sumon Deb, Muhammad Rashiduzzaman, and Zahid Hasan Showrov, "CUSTOMER RELATIONSHIP MANAGEMENT," DAFFODIL INTERNATIONAL UNIVERSITY, DHAKA, BANGLADESH, Bachelor Degree 2018.

[48] ALEKSANDRA DEWICKA, MICHAŁ TRZISZKA, and JOANNA ZIOMEK, "The CRM system and its role in raising the competitiveness of enterprises," *Annales H - Oeconomia*, vol. 53, no. 2, p. 18, octobre 2019.

[49] Ahmad M Zamil, "Customer Relationship Management: A Strategy to Sustain the Organization's Name and Products in the Customers' Minds," *European Journal of Social Sciences* , vol. 22, no. 3, p. 456, juillet 2011.

[50] Rozitta Chittaie, "CUSTOMER RELATIONSHIP MANAGEMENT AND BUSINESS STRATEGIES," *International Journal of Organizational Leadership*, vol. 1, no. 1, p. 14, 2012.

[51] tutorialspoint. Customer Relationship.

[52] B Bathelot. (2015, Janvier) definitions marketing. [Online]. https://www.definitions-marketing.com/definition/crm-analytique/

[53] Carena Marchi. (2020, Juin) MARKETING MANAGER. [Online]. https://www.mirabelsmarketingmanager.com/blog/analytical-crm

[54] (2011) Cahierdecharges.com. [Online]. http://www.cahierdescharges.com/blog/?p=75

[55] Zidane BENOUARET and Samir BOUDJAOUI, L'étude de la mise en place d'un projet CRM cas pratique TCHIN LAIT (CANDIA), 2015.

[56] OLIVIA LABARRE. (2019, Septembre) Investopedia. [Online]. https://www.investopedia.com/terms/e/erp.asp

[57] S Surbhi. (2015, septembre) Key Differences. [Online]. https://keydifferences.com/difference-between-crm-and-erp.html

[58] (2019, Avril) Gestisoft. [Online]. https://gestisoft.com/difference-entre-logiciel-erp-et-crm/

[59] Amer Wilson. (2019, Juin) ROLUSTECH. [Online]. https://www.rolustech.com/blog/data-mining-crm

[60] Gaurav Gupta and Himanshu Aggarwal, "Improving Customer Relationship Management Using Data Mining," *International Journal of Machine Learning and Computing*, vol. 2, no. 6, pp. 876,877, Decembre 2012.

[61] HARSH GUPTA. (2020, Juin) What After College. [Online]. https://whataftercollege.com/data-science/data-mining-crm/

[62] vishnu. (2019, Novembre) Dhyey Technologies. [Online]. http://dhyeytech.com/2019/11/15/data-mining-in-crm/

[63] Hamza LYAZIDI and Mohamed Amine MIHRAJE, "La mise en place et le paramétrage de l'ERP Odoo pour le service de la Production," Université Sidi Mohamed Ben Abdellah, mémoire 2015.

[64] Lucidchart Content Team. Lucidchart. [Online]. ttps://www.lucidchart.com/blog/types-of-UML-diagrams

[65] Ilhem EL OUNI, "Conception et Réalisation d'une Plateforme de Détection et Gestion des Appels d'Offre," mémoire master 2018.

[66] Sarah Taouli and Ikram Hamza Cherif, "Réalisation d'un site web de vente en ligne," Université Abou Bakr Belkaid-Tlemcen, Tlemcen, memoire license 2015.

[67] Stéphane Crozat, "Conception des bases de données I : Introduction," 2020.

[68] arezki moghrani and zoubir azzoug. (2009) memoireonline. [Online]. https://www.memoireonline.com/08/09/2577/m_Conception-et-realisation-dune-application-de-suivi-de-patients-dans-un-etablissement-hospitalier7.html#toc16

[69] Ines BEN KRAIEM, "Modélisation et intégration de données de capteurs/compteurs du SGE," rapport de stage 2017.

[70] Anass SAIKOUK and Abdellatif MABROUKI. (2015, juin) Application web de la Gestion des indemnités du transport du Algo Consulting Group.

[71] Rstudio. RStudio. [Online]. https://rstudio.com/products/rstudio/

[72] Lucas SOUMILLE, "Rapport de stage," institut universitaire de technologie, Marseille, rapport de stage 2014.

[73] Janice Friedman. (2019, Aout) csharp-station. [Online]. https://csharp-station.com/what-is-c-used-for/

[74] R. R. [Online]. https://www.r-project.org/about.html

[75] tutorialspoint. tutorialspoint. [Online]. https://www.tutorialspoint.com/entity_framework/entity_framework_overview.htm

[76] Entity Framework Tutorial. Entity Framework Tutorial. [Online]. https://www.entityframeworktutorial.net/code-first/what-is-code-first.aspx

[77] Microsoft. (2018, Janvier) Microsoft. [Online]. https://docs.microsoft.com/en-us/dotnet/desktop/wpf/getting-started/?view=netframeworkdesktop-4.8

[78] Jekyll. R.NET for users. [Online]. https://rdotnet.github.io/rdotnet/

ملخص

في وقتنا الحالي شهد ولازال يشهد العالم نمو كبير وسريع للأسواق والمؤسسات مما ولد منافسة شرسة هدفها الأسمى هو كسب رضا العملاء وضمان ولائهم وبالتالي النجاح والاستمرارية وفرض هيمنة الشركة في السوق المحلية ولما لا في الأسواق العالمية. إن إدراج تقنيات التنقيب عن البيانات في برنامج إدارة علاقة العملاء يعتبر وسيلة فعالة ومساعدة في تحسين وتقليص المسافة بين المؤسسة وعملائها, تسهيل العمل والمساهمة في اتخاذ القرارات الصحيحة وفي وقت وجيز. في هذه الأطروحة حاولنا تقديم صورة مفصلة عن هذا البرنامج ومختلف أقسامه الرئيسية بغية استعماله وإدخاله في تسيير شركة جيون إلكترونيك للمساهمة في تطورها ونجاحها.

الكلمات المفتاحية: تنقيب البيانات، إدارة علاقة العملاء، UML، لغة R ،شركة جيون.

Abstract:

At the present time, the world has witnessed and continues to witness a great and rapid growth of markets and institutions, which has generated fierce competition, whose ultimate goal is gaining customer satisfaction and loyalty, success and continuity and imposing the dominance of the company in the local market and why not in the global markets. The inclusion of data mining techniques in the customer relationship management program is an effective and helpful way to improve the distance between the enterprise and its customers, facilitate work and contribute to making the right decisions in a short time. In this thesis, we propose a CRM solution endorsed by Data Mining techniques for the company Géant Electronics in order to contribute to its development and success.

Key words: Data Mining, CRM, UML, R language, Géant Electronics

Résumé:

À l'heure actuelle, le monde a assisté et continue d'assister à une croissance forte et rapide des marchés et des organisations, ce qui a généré une concurrence féroce, dont le but ultime est de gagner la satisfaction des clients et donc le succès et la continuité et d'imposer la domination de l'entreprise sur le marché local et pourquoi pas les marchés mondiaux. L'inclusion des techniques d'exploration de données dans le programme de gestion de la relation client est un moyen efficace et utile pour réduire la distance entre l'entreprise et ses clients, de faciliter le travail et de contribuer à prendre les bonnes décisions en un temps réduit. Dans ce mémoire, nous proposons une solution CRM endossée par les techniques du Data Mining pour l'entreprise Géant Electronics afin de contribuer à son développement et à sa réussite.

Mots clés: Data Mining, CRM, UML, langage R, Géant Electronics.