Data-driven decision making

Hernán Darío Rojas Gutiérrez, José Alejandro Morales Bobadilla, Jayden Ortiz Diaz, Jorge Enrique Portella Cleves

National Unified Corporation - CUN
Bogotá, Colombia
hernan.rojas@cun.edu.co
jose.moralesb@cun.edu.co
jayden.ortiz@cun.edu.co
jorge_portella@cun.edu.co

I. PROBLEMS

In the business world there are a number of companies or organisations that do not have the technological applications and tools that allow them to transform the data to obtain clear and accurate information that will help in making decisions to meet their objectives, this happens because they do not have knowledge of Business Intelligence (BI); which would be very useful to solve this problem.

II. OBJECTIVES

- Provide basic information on the subject based on business intelligence, information warehousing and data mining.
- To provide information on the characteristics and application of BI in business decision-making.
- Show tools and techniques such as ETL for data that are fundamental in the decision making process of a company.

III. INTRODUCTION

Data is the most valuable asset of a company or organisation nowadays, for the management of such data there are several techniques used for the storage and analysis of these data, if the organisation chooses wrongly among the alternatives it could face a very costly problem. Based on the above information we will study a very important issue today in the business world, with the global economic growth has

also grown the world of technology and therefore organizations or companies also need to move forward with firm and fast pace how business evolves; Its weaknesses, strengths and the need to always be at the forefront of technological tools we will dive into the subject of business intelligence or also known as Business Intelligence (BI), Datawarehouse and Data Mining which is nothing more than a group of applications and tools that allow you to extract, transform and load some data to get to get information and knowledge in order to make a quick decision, accurate and efficient in the organization to achieve the objectives outlined.

Due to the fact that there are still companies or organisations that make blind decisions in their customer or strategic processes. In order to solve this problem we rely on data mining as an alternative to minimise risks, which can lead to major and valuable losses within an organisation. In the case of a private company where data mining is applied to study patterns of customer behaviour on their own parameters of location, consumption, etc.. And third party data. The search for information is profitable for the business administration.

Data mining is applied as a tool for the development of marketing tactics in competitive production and industrial sectors. This technology attempts to help perceive the attachment of databases. Data mining works on a preferential level looking for patterns, behaviours, orders or groupings that can create a model that allows us to better understand the concept and help in decision making so organisations rely on

different systems such as CRM, ERP and many others, but to move from just information to generate business intelligence must be centralised in a single place where you can run data analysis of all types to discover trends that help decision making that place can be mainly a Lake or a Warehouse.

CHAPTER I
WHAT IS BUSINESS INTELLIGENCE

Business Intelligence - Business Intelligence (BI) - is a group of applications and methods that allow me in an easy and fast practical way, to be able to take a quantity of collected data and to analyse them, this with the purpose of being able to obtain precise information of these and knowing these data to be able to arrive at a conclusion that allows the manager or administrator of a company or company to take the best decisions in favour of fulfilling the proposed objectives.

In the early days of technology advancement, talk of business intelligence was very premature and the vast majority of people had no idea what it was all about. Business intelligence is for people who want to get out of technological illiteracy and want to be at the forefront of applications or programmes that allow users to streamline their tasks in an optimal and efficient way.

Nowadays, Business Intelligence (BI) is a well-known term in the world, but even so, there are still many people who confuse the tasks that can be developed with business intelligence and end up doing the work with spreadsheets or other programs that do not allow you to do a quick and easy job as it is with the use of BI [1]. According to [2] this happens because there is still a wide gap between what people in companies really want and what they think or believe can be done with business intelligence tools.

We can affirm that business intelligence, rather than presenting a graph of the data we have collected,

what it seeks is to reach a conclusion, in other words, to be able to analyse that data to obtain information and that serves me to be able to make decisions in the company that go in depth with the fulfilment of goals or objectives that have been set.

Bi is not only for a specific area within the business Bi expands to several areas of companies or companies such as vendors, suppliers, customers, products and services among other areas that can be found within organizations, it is important to note that the process that makes Bi is quite broad and is divided into a few stages that we will talk about below.

One of the first stages of business intelligence is the participatory or collaborative process, as this allows Business Intelligence to deliver a continuous analysis of information over time, not just for a moment or a specific time, but to keep analysing the information constantly, which will allow then to obtain versatile information results that help you make organisational decisions within the company.

Another important stage of Business Intelligence is the exploration of the data or information we have collected. This exploration allows us to see and understand how our company, business or enterprise has performed.

The next stage in smart business is to analyse in that stage we can see how the relationships between the variables that we have within the analysis of the information have changed and how they have evolved and that will allow us to make some calculations, obtain some models or standards to find the formulas and to be able to analyse the customers, suppliers, services that work within the organisation we are analysing.

In this last stage of smart business we talk about data storage, such data is stored in a Datawarehouse in this space companies store information in tables for later retrieval in the fastest and most practical way possible that information is organised in records and each of the records has its attributes, so we can have organised information to analyse and meet the

objective of Business intelligence to meet the goals or objectives within organisations.

A. How important is Business Intelligence in the business world?

The technologies or tools used in business intelligence are of great importance in companies or enterprises because they allow me to understand more quickly the information obtained from the transformation of data and thus be able to know what was the cause of the performance or failure that the company or enterprise had at a certain time, which is significant at the time of decision making.

B. How does business intelligence work?

For the operation of business intelligence, this has within its set of applications a tool known as ETL which means extraction, transformation and loading of data; the idea is that by organising the data and studying the data the tool allows to load that information in a data warehouse which is called Datawarehouse, from there I can already know, organise and speed up the processing of information.

C. Who needs business intelligence?

BI is so important in organisations that the need is for everyone who wants to make decisions in an organisation, company or business to have a business intelligence model established in their organisation in order to be able to solve questions that arise when meeting their objectives.

D. Benefits of business intelligence

Well, one of the most important and perhaps the main benefit of Business intelligence is the great support when making decisions. Hence it is important to see how people who are in charge of the management of an entity or company learn to see what kind of information we are going to take and how we are going to analyse it, because with this and according to our experience and training, it will help us to make a decision that will place us as good or bad managers. Hence, being able to have a broad knowledge in our company from the collected data is a great benefit that Business intelligence offers us.

Of the benefits that business intelligence gives us there are several, of which we can highlight among them the reduction of expenses because with the tools of intelligent business we can see at what time or space is that we are spending more resources for the production or maintenance of the company, also the agility to do some activities in the same way we can find what processes or activities can be automated within the organization, the generation of more income this we can obtain it maintaining our clients, finding the formula to increase the profit in our products so that they are more profitable, to make that increase our clients that are faithful to our organisation this would be to help to the increase of the income of the company, in addition we can manage to see what type of markets, products, clients or suppliers we can have as strategy in our company or company that allow to increase the utility.

E. Components of business intelligence

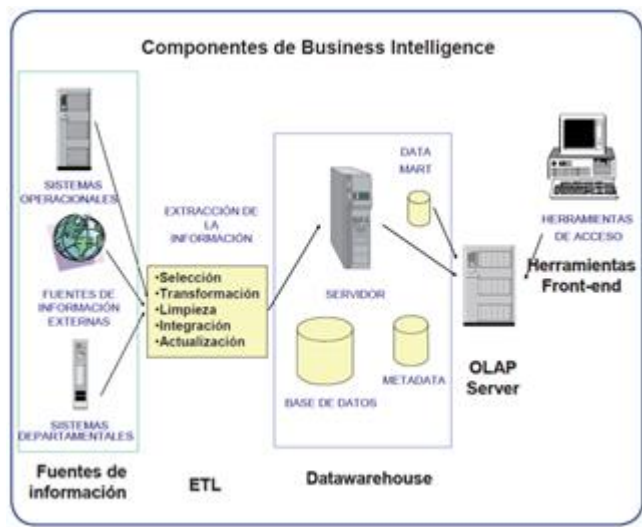In the following graph we can see the components of the Business Intelligence tool

Fig. 1. Components of Business Intelligence. Taken from Business Intelligence. https://basesdatoscms.files.wordpress.com/2012/10/resumen-businessintelligence.pdf

Sources of information: This component is constituted by all the parts from where I am going to extract the information or data, it can be from a system that I have located in a company's equipment or an external source, for example, let's say I have information in a cloud server from where I am going to obtain the data; all this is known as information source where I have hosted the data that I am going to analyse with the tools that make up the business intelligence.

- •   **ETL: usually** the data that we have stored as information of our company, are not ready to be able to analyse them, then the first thing we must do is the extraction that corresponds to download them to be able to manipulate them, then make the transformation that means, filtering, cleaning and organising, to then proceed to load the data in the Datawarehouse where they would be ready to be used with the business intelligence tools.

- •   **Datawarehouse:** It is the warehouse of our databases where data dictionaries are applied to store them and it is what allows me to maximise the performance of the space and also when I need the data I can obtain it in a faster and easier way.

- •   **OLAP server:** OLAP stands for online analytical processing, this engine or server allows me to perform calculations, queries and other tasks that I can execute with my stored data.

- •   **Front-end tools:** this component is the one that contains the tools to navigate and visualise the information that I have processed with the business intelligence tools. From here I can start to analyse the information.

In a Data Lake, all forms of data that have been generated throughout the company are divided, this includes data sources such as conversation logs, emails, images, audio, videos, databases among others, resulting in an accumulated amount of information, which needs to be organised and filtered in order to be analysed.

Data lakes handle a huge amount of information, the volumes are so high that traditional databases can take a long time to execute a single request, so it is important to have specialised hardware and invest heavily in the storage of the information, on the other hand and unlike the dark depths within the unstructured data of a Data Lake, a Data Warehouse has its analysis clean and its data sorted to extract information much faster.

CHAPTER II

DATAWAREHOUSE

Data warehouses can store information that has already been transformed. The function of the data mart is the analysis and filtering of data, this can be considered a subsection of datawarehouse for a very specific need of a department, unlike a warehouse and a data lake where the information is stored in a unique and centralised file, the data mart works with a differentiated and decentralised account of data, This provides a higher level of security for the organisation in general, as the unit using the data mart will only have access to the data previously loaded in its database, i.e. it does not see the data of the rest of the company. The same applies to

efficiency, workloads in an isolated environment do not use resources from other sectors or departments, in general terms.

1. PROPERTIES

Datawarehouse has a number of properties that are important for its architecture such as:

Separation: it is important that the data transaction and the data analysis process should be done separately.

Scalability: means that as information grows, the software and hardware of a datawarehouse architecture can be easily upgraded.

Extensibility: is obtained when the Datawarehouse architecture allows me to add new applications and technologies without having to analyse the whole system.

Security: this property is important as the datawarehouse contains a lot of important information.

Manageability: an important characteristic of the tools must be their management by the users, and in the Datawarehouse the management is not so complex.

2. LAYERS OF DATAWAREHOUSE STRUCTURE

Data sources layer: this layer represents the different types of data that are stored in the Datawarehouse, these can be in database, flat file, Excel file or other with which you can load data to perform operations among others.

Data extraction layer: in this layer is where data is extracted to be loaded into the Datawarehouse, here in this layer a superficial cleaning of the data can be done, but it is not yet transformed.

Testing area: this layer is where the debugging and transformation of the data in the Datawarehouse takes place.

ETL layer: this is the important layer in the transformation of the data because it is where they stop being data to make transactions and become data that I can analyse, in this layer is where the Datawarehouse takes more time because the data are cleaned.

Data storage layer: in this layer the transformed and cleaned data is already placed or stored.

Logical data layer: this layer is where you organise how you are going to handle the information later on, in other layers where you put the rules.

Data presentation layer: this layer is where the way in which the data will be presented to the end user is organised, in other words, it is the way in which the report of the data that the user wishes to see will be generated.

Metadata layer: this layer is where the information is characterised in terms of quality, availability, history and conditions; special metadata tools are used here.

System operations layer: this last layer provides information on how the Datawarehouse works, its performance, the history of users who have logged in, among others.

3. COMPONENTS

To understand how data warehousing works, we need a thorough understanding of its components, so let's start by defining the operational systems.

Operational   Systems

-Marketing

-Sales

-ERP

-scm

These would be the company's IT systems from which the information comes, but it can also come

from different departments such as sales, manufacturing, human resources or finance, and the information from these systems goes to an integration layer:

-integration

-ODS

-Staging   area

This is where the ETL data transformation and loading process takes place. In this integration layer, all the work is done to transform the data into a form in which it can be analysed.

4. DATA VAULT

A data vault is where the processed information from these operational systems is stored and where the raw processing and analysis of the data takes place.

5. DATA MART

It is where the content of the datawarehouse is separated to have it divided depending on the functions of the organisation, data mart is the smallest approach of the three and is oriented to short term projects. The choice between these three systems will depend on the present and future needs of the business, it is possible that many organisations will not need a data mart to operate successfully, in other words data marts are subsets of data within a data warehouse that contain information about a specific area of the company or business for example by department by function such as sales manufacturing support and could be said to be small data warehouses focused on a topic each managing its own data integrity and update which are queried using OLAP tools to be analysed with the rest of the data marts.

6. APPLICATIONS

Some of the applications of the datawarehouse include.

Market forecasting.

Market review which is predicting the flow of a market with historical information and detecting patterns over time.

Behavioural analysis.

Behavioural analysis of either customers or businesses according to specific parameters and thus classify their behaviour.

- • **Cost modelling and budgeting**.

Cost and budget modelling using aggregation functions and groupings can be used to analyse operating costs to make business improvements.

7. BENEFITS

- **Quality and consistency of information**

One of the benefits of creating a Data Warehouse in our company would be the quality and consistency of the information by having a centralised repository and integrity when adding new information, guaranteeing granularity and a unified representation of the data.

- **Speed of response**

We will also gain speed of response because the automated process allows the relevant departments to obtain immediate reports, without requiring ite's intervention.

- **Intuitive visualisation**

There are some business processes that benefit from the representation of these Data Warehouse toasts of information by means of a star schema.

- **Security**

The data warehouses that have been implementing this technological tool for many years are quite secure, as the information can be recovered at any time, but for this function to be completed, backups must first be programmed, the levels of access to the information, access to the lan or wan network, and user permissions.

8. ARCHITECTURE

The structure of a Datawarehouse may vary according to the number of data sources, but in conclusion, any Datawarehouse architecture has several layers as discussed below [3]:

Layer of data sources

    Data extraction layer

    -Testing area

    ETL Layer

    Data storage layer

    Logical data layer

    Data presentation layer

    Metadata layer

    System operations layer

In the following we are going to present two of the most important architectures for the construction of data warehouses. First of all, to provide some context, Bill Inmon and Ralph Kimbal are two of the most important exponents in the field of database analysis, both proposed interesting ideas about the construction and format of a data warehouse, In the case of Bill Inmon he proposed a top-down view of warehouses which consists of building the data warehouse from the top down, i.e. from the most complex to the simplest, and Ralph Kimbal proposed a reverse construction from the simplest to the most complex.

9. BILL INMON

Let's first analyse the architecture

Bill Inmon, this is called a top-down view.

Because to build the Data Warehouse is designed as a centralised unit initially defining all possible data and their relationships, it is necessary to know in advance all the components of the company, then, we start having the data sources, which are extracted from the operational systems of the company, and in

the integration layer the data extraction and transformation process is applied to transform them into a third normal form, i.e. the data are normalised, then the data are divided into themes to transform them into the respective data marts, and thus present the information in the presentation layer by means of reports, analysis and data mining.

-Features

Bill Inmon, proposes four characteristics that a data warehouse must meet.

Theme-oriented

The first is that it must be topic-oriented, i.e. the data corresponding to a specific topic or business process must be close and related in its tables.

Time-varying

 The following time-variant feature means that this data will change over time, but historical spaces will be kept in order to have all the information backwards from the changes of this information.

Non-volatile

 Non-volatile means that this information is permanently persistent, no previous information is lost.

-Integrated

It means that all the information of the organisation or company is centralised and can therefore be related to each other.

10. RALPH KIMBALL

Now we are going to look at Ralph Kimball's architecture which consists of a Bottom-Up view, i.e. a bottom-up view because the data marts.

Modular design

Data marts are first created for each department or business unit with their own independent analysis and operations.

- •   **It is done by means of a dimensional model of the data.**

It is a form of design that facilitates the outcome of queries on the end user.

- •   **Data warehouse**

For this step it must be done with a "bus" that connects all data warehouses.

So to understand it better we start from the bottom with the data sources that could come from operational systems or from other external applications.

- •   **Data Marts**

-Reports

-Applications

 This means that each department is responsible for the transformation of its data, after which the data is transformed into a dimensional model, usually with the star schema, in order to present the data in the presentation layer.

11.COMPARISON OF ARCHITECTURES

Knowing both architectures we can now make a small comparison between them to start with in the Bill Inmon architecture, we have that it requires more time in previous design, as we have to know the whole structure of the company and its processes, because everything is done centrally instead Ralph Kimbal architecture. We have a modular design, each department creates its own data marts independently, therefore it is not necessary to design all the data marts to start creating the data warehouses block by block. As for the representation of the data, in the case of data warehouses, it proposes a relational model in the third normal form. Similar to the transactional databases that optimise their data to generate results which are compatible with websites, they are quite secure and reliable, which is why they are used to manage transactional portals, such as bank transfers,   the traditional ones where their model is unfortunately not very efficient for analysis in its third normal form, where its design

process introduces many tables that consequently have and require many joins,   this type of operation can considerably lower performance. In the case of Ralph Kimbal having a dimensional model and which we will see later is designed specifically for data analysis and ease of reading so we might expect queries to be faster in addition in terms of updating the data and the Bill Inmon case.

This means that data from all data warehouses   are updated at the same time and are consistent in the case of Ralph Kimbal.

As each data marts is managed independently then their data is updated asynchronously this means that data marts can be updated before others and information than others that information may not be the most up to date at all times. As far as the process of loading and transformation of data is concerned, it is similar to Bill Inmon, the data is loaded all at the same time and its integrity is verified in the case of Ralph Kimbal's architecture. Each data marts takes care of its own data loading and transformation.

Finally, in terms of scalability and maintenance in the top-down architecture it is a bit complex because as everything is unified if we want to add tables or data we must maintain the integrity reference and this could be a bit difficult if the tables have many restrictions and many relationships between them instead.

In the Bottom-Up   architecture the maintenance depends on each department and if we want to add new information pertaining to a topic or a business process we only have to touch that specific data mart of the data Warehouse. Now we can proceed to see the steps for the design of a data Warehouse from this moment on we are going to assume that we are using Ralph Kimbal's architecture due to its ease and modularity is more seen in the practice of companies and therefore is the one that interests us the most. To begin the design.

12. DATA WAREHOUSE DESIGN

We need to understand the dimensional model as this is how we will represent the data in the dimensional

model. It consists of a discipline of modelling of alternative data to the model entity relation that contains the same information, but the data are packed in a symmetrical format whose goal are the facility of understanding the rapidity of consultations and the resistance to the change, this last one what it means is that what future changes to the structure of the database already be adding information or removing information modify anyone of less work and smaller manipulation of the database, the relational model is based on two components or tables which would be the tables of facts and the tables of dimensions. The tables that make a star structure are similar to a diagram of processes of each business that would come to make the set of activities defined within the company on a specific topic to deepen a little fact tables are the centre of the relational model, these contain measurements continuity of business and means numerical or additive values to which they can apply their matter their average that are the result of the analysis that we do on our business processes in addition these tables do not have a primary key. The second component of the dimension tables could be said to be the tables that accompany the facts from which the data is obtained. These are the ones that hold the information extracted from the transactional systems. This means that each of these dimensions has a unique primary key that is used to relate between the different facts and also contain fields that function as filters and groupers to apply aggregation functions. We will now see the step-by- step process of how to transform a database from an entity-relationship model to a dimensional model. The entity-relationship schema that could be for example an online shop that sells products to its customers through shipments and in turn has suppliers of manufacturing companies we can then see different entities and relationships that we are going to transform for that the first thing we must do is to separate the entity-relationship diagram into sub-diagrams. For each business process each business process could be almost every one of the relationships that are in this diagram so from it what we do is to discover the minimum amount of entities that are part of each of the relationships and so we

have little pieces of the diagram separated once completed the sub-items we move on to the next step. It would be to obtain the quantifiable values decades one of the quantifiable diagrams consists of data that we can extract from each one of these relations for example in the process of sending products to clients we could have the interest to calculate the time of shipment costs of shipment and weights of products those quantifiable values that we extract of cases of gram would come to be the facts to form the diagrams in Star.

Then once we have these facts the next step would be to transform the tables involved in each of the business processes into flat tables with unique, self-contained primary keys. Having the shipment facts we can reduce the product and customer lane tables into their simplest forms with the necessary data to be able to measure the shipment facts and then complete our first star schema when a dimension table is connected to facts from different business processes they are called a conformed dimension, This means that these dimensions are shared between different schemas and therefore must represent the same object, we return to the concept of "bus" architecture which was discussed earlier, in this case the "bus" architecture corresponds to a standard of dimensions between data marts to unify the data this means that if for example we have a dimension of products, which exists between different datamarts or between different departments.

In order to standardise the final product, but at the same time to be compatible in its entirety with all the data marts, with the objective that when the data Warehouse makes data marts to any of these products, its information will be compatible. Once the design of our data Warehouse has been completed we proceed to   load data into it, for this we have the process of extraction and transformation of data, this process has three parts:

Extracting data from and transnational bases

Processing and transport to their final data marts on the extraction side.

The upload is the result of the information search, where the requested data is displayed...

This means that operations of summation, averaging and some other functions are performed, as well as the maintenance of metadata. Then we have the data transformation process which includes cleaning and integration, this means that if we have the same data but in different formats in several databases, what we do is to transform a single meaning; for example if we have country names or numerations in different databases we transform all these names to a single language so to speak, and finally in the data transport stage, these are inserted into the columnar databases of the data marts, load tests and error management quality tests are performed and once this is completed we have the data ready to start performing analysis and interpretation of the data obtained. Therefore, our data facilitates the implementation of decision making, having a clear knowledge of the information, to better understand the subject it is necessary that we understand what data mining is.

CHAPTER III
DATA MINING

Data mining arises from databases, data repositories, statistics, machine learning, data visualisation, information search and retrieval and high-performance computing, to elaborate fundamental processes where a series of intelligent methods are applied to extract and reveal patterns in the data, by using algorithms such as   linear regression, logistic regression,   association, fuzzy logic, decision trees and neural networks among others,   which allow   to achieve historical and prospective knowledge for decision making in the area of study.

1. OBJECTIVES OF DATA MINING

Specifically, it is exploring and consolidating what is in the databases, and the architecture is client-server using parallel processes. And this facilitates the applicability of tools and techniques to achieve the extraction of data and data mining originates types of information that is useful for businesses such as

associations, sequences, classifications, groupings and forecasts that help us to make decisions.

2. DATA MINING APPLICATIONS

Data mining can be applied to different areas of people's activities that generate data, for example:

Medicine: how diseases and the effectiveness of procedures to patients.

Trade: shopper segmentation.

Banking: risk analysis.

Security: identification and biometrics and access.

Geology: soil data and satellite imagery

Environmental sciences: functioning of natural and/or artificial ecosystems.

Social sciences: Study of waves of public or private opinion.

Energy: measurement of the energy demand of each customer.

▪Tourism   : to establish socio-economic types of tourists.

▪Web : Behaviour of users accessing different websites.

Biology   : in the analysis of gene or protein sequences.

Telecommunications : calling patterns or network load types.

▪   Education: classification of students and

selection and encouragement of absenteeism.

We are currently seeing an exponential growth of artificial intelligence applications for decision making in companies, where they manage to optimise customer satisfaction, provide fast and efficient assistance, technical support where such applications allow computerised data analysis. It is of utmost importance that companies can have, read and understand all the information in an exponential way, that is why the implementation of these data

mining techniques obtain solutions through knowledge.

3. INTERPRETATION OF DATA MINING

We find an evaluation and interpretation phase in data mining as the verification if the results are consistent, once we have the model, we must proceed to its validation and verification of the conclusions obtained. There are the techniques of evaluation, simple validation, cross-validation and cross-validation with folds. Such data interpretation refers to the implementation of methods through which data are reviewed with the aim of reaching an understood conclusion and is a fundamental stage of the data process. While there are several types of different processes that are implemented in a function of the data itself. The common categories are quantitative and qualitative analysis within a statistical analysis. And the goal is to be able to analyse and interpret data useful for decision making. It is the way to set clear interpretation objectives and first perform an analysis that helps the company to focus and improve on strengths and weaknesses where a potential solution or improvement in the company's role can be recommended. For example, if you are conducting a performance study you can categorise the data according to what is associated with each outcome in general, such as employee productivity and learning in their specific job. Levels of measurement are also important and should be governed by the data and we find the nominal level, ordinal level, interval and ratio types of all three. Observations, documents, interviews.
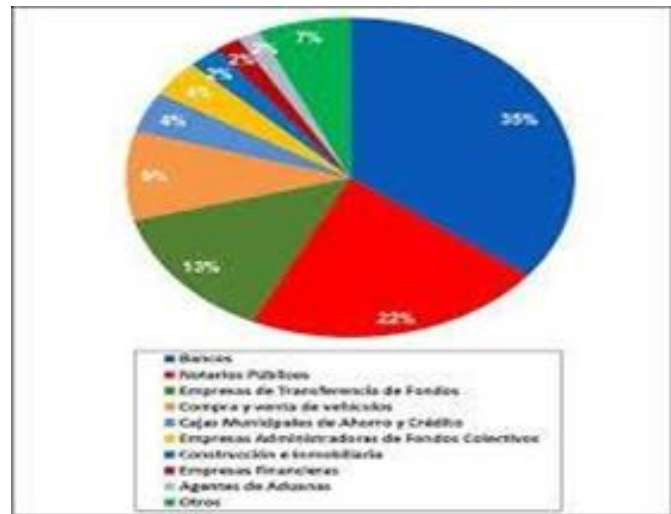

Figure 2. Interpretation of data

4. DATA MINING TOOLS

In researching the tools for mining

There are methods such as statistics, association rules, decision trees, graphical models, neural networks, genetic algorithms and fuzzy logic that can be used in data mining. In data mining we find tools that we can use such as the Python IDE Spyder libraries with the Anaconda Navigator IDE, with the aim of extracting documents as, for example, .txt or other formats. These Python libraries are free and help us to be operational.

5. DATA MINING APPLICATIONS AND TYPES

We have some development environments such as; Oracle Data Warehouse, SQL Server Enterprise and Sybase, PowerDesigner and Sybase IQ these are some of the applications that are used in data mining. Can't we ask what types of data are used in data mining? Depending on how the data is stored: unstructured, structured or semi-structured and relational databases, spatial databases, temporal databases, documentary databases, multimedia databases. We also find models that are predictive models and descriptive models that help us to identify patterns that allow us to examine the properties of the inspected data.

6. WORK ON DATA MINING

We find regression and classification which is predictive, there is also unsupervised classification which is descriptive, association rules, outlier detection and visualisation which are descriptive. As we see that data mining is at the heart of the knowledge discovery process as we can see in the following image;

Figure 3. Data mining information system

First, we find the preparation of the data that is in a database, we move to data mining and look for patterns; evaluation, interpretation and visualisation of the data and then insight. Then we move to a process of initial data, integration and collection, data preparation, moving on to modelling, an assessment, deployment and review that go in parallel with decisions.

7. HOW DOES DATA MINING HELP CREATE BUSINESS VALUE?

There are different types of analysis that can be obtained from big data. This means that according to each type of analysis we will have a different result and which data mining technique we apply to the type of problem of the company. Data mining specifically refers to the discovery of interesting and previously anonymous patterns. We find some stages such as the determination of the objectives that practically deals with the delimitation of the objectives, there is also the pre-processing of the data and it is the acquisition, reduction and transformation of the databases. Another stage is the determination of the model, which begins with statistical analysis of the data and leads to a graphic visualisation and finally the analysis of results, which consists of verifying whether the results obtained are coherent and analysing them with the results. One of the advantages is that it allows us to find hidden and unforeseen information and analyse large amounts of data very easy and interpret them and the main one is that it helps to make excellent tactical and strategic decisions to discover key information for company that requires and finally the models are approved and tested using statistical techniques, before being used with the forecasts that are achieved are confident and valid, from these build models quickly that gives us new business opportunities and save some costs to companies.

We also found some disadvantages in data mining such as; it can be difficult to collect the data, generating more work and debate in the selection. Costs require immense capital expenditure in the purchase of the technologies.

required to carry out the selection of such data. We conclude that data mining offers us better information management and forecasting of future events.

8. DECISION-MAKING METHODS

Associated with business activities, the method applied in data mining is the set of methodologies, applications and technologies that allow us to unify, cleanse and transform data from transactional systems and unstructured data internal and external

to the company and with structured data we can exploit them directly and respectively for their analysis and transformation into knowledge to give results and support decision making on the role of business or activity. Indicator technique we consider as a tool of unification through communication and dialogue that is created on the basis of the information between the different representatives of the process.

9. DECISION MAKING WITH DATA ANALYTICS

There is a problem with executives who are the decision makers and are in communication with data mining engineers and sometimes they do not understand the technical language we want to convey with the results extracted from the databases. Because companies thrive on data literacy, i.e. without data we could not advance in the economy and especially in business, that is why data-driven decision making is based on supporting their decisions on information that historically the company is storing that we will be organizing and then we will analyze it. And these bring benefits such as:

- More appropriate decisions.

- Reduction of administrative costs.

- More proactive personnel

- Continuous improvement in processes

- A clearer and more understandable   future.

Here it involves a data culture, which is not easy to adopt overnight, we have a collaborative process, corporate help and support in making data the centre of every meeting and empowering every individual in the usability of data and enabling reliable information and due process governance, that is to say that the data must go through a process of concentration, auditing that if that data has the integrity that is required for decision making and that meets the conditions of answering the questions that the business person wants to answer through that

data. In decision making we have a process; which is to identify your mission, second is to identify the sources of data that can help us with the approaches of the first step, then in the third step is to clean and organise the data, centralised and create tactics to preserve its quality, we move to the fourth step which is to model the data and generalise it. And finally it is to update the processes and repeat the cycle. Part of the process is also to identify the area of greatest impact, the roles and profiles of the human resources that will be involved in the project. Establish clear, understandable objectives, so that the group is aware of the process and the information, where we create hypotheses and define success metrics that will help us in the decision making process.

CONCLUSIONS

This article was very important for its development, as the relevant concepts of Business Intelligence are implicit in it and therefore the study of these tools is very useful to be able to carry out the best decision making in an organisation or company.

We see that the application of data mining is focused on the actions performed and processed by an organisation. The techniques and algorithms in data mining are defined according to the problem to be solved. There is still a gap in small, medium and large companies in understanding that new technologies help to improve the analysis of data and invest in improving these tools that will help make or predict future decisions to acquire expected results, in the same way the development of a Data Warehouse in the process of implementing a business intelligence solution and its implementation in data mining, is equally or even more important than the technology selected for its exploitation, because without a good model we can face very complex problems such as long response times, inconsistent information, problems to display information, so even if you have the most expensive technology in the world and trends, if you do not get the purpose of the Data Warehouse, and its correct design is almost a key to any Business Intelligence (BI) project will not be successful as expected.

REFERENCES

[1]   I. Ilabaca, "Business Intelligence," Retrieved from http://www.emb.cl/gerencia/articulo.mvc?xid=1769# .

[2]   C. Jose Miguel "Business Intelligence," Retrieved from http://www.emb.cl/gerencia/articulo.mvc?xid=1769# .

3]   PowerData, What you need to know about datawarehouse architecture.   2016. Retrieved from https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/lo-que-necesitas-saber-acerca-de-la-arquitectura-de-un-datawarehouse

[4]   2021, E. S. (4, 10, 2006). ABC del Data Warehouse.   Obtenido   de https://www.evaluandosoftware.com/abc-del-data-warehouse/

5]   Adolfo, L. A. (2012). Use of Data Warehouse Technology . Science Journal UNEMI, 1-7.

6]   EL CEDAZO. (15 June 2009). Story of an Old Computer Scientist. And the Data Warehouse became Business Intelligence. Retrieved from Historia de un Viejo Informático. And the Data Warehouse became Business Intelligence: https://eltamiz.com/elcedazo/2009/06/15/historia-de-un-viejo-informatico-y-el-data-warehouse-se-convirtio-en-business-intelligence/

7] BUSINESS   INTELLIGENCE: STATE OF THE ART. (4 of 2020). Scientia et Technica Year XVI, No 44, April 2010. Technological University of Pereira. ISSN 0122-1701, pp. 1-6.

8]   Sandoval, L. (4 November 2020). INFORMATICA WINDOW. Retrieved from OLAP Tools, What are they and What is their use?: https://ventanainformatica.com/software/herramientas-olap/#Herramientas_OLAP

9]   Tamayo1, M. (3 December 2006). Analysis of the MOLAP storage model versus analysis of the MOLAP storage model versus analysis of the MOLAP storage model. REVISTA INGENIERÕA E INVESTIGACI "N VOL. 26, 1-8.

10] Riquelme Santos, J. C., Ruiz, R., Gilbert, K. (2006). Data mining: Concepts and trends. Inteligencia Artificial: Revista Iberoamericana de Inteligencia Artificial,

10 (29), 11-18.

11] Orallo, J. H., Quintana, M. J. R., & Ramírez, C. F. (2004). Introduction to Data Mining. Pearson Educación.

12] Suárez, Y. R., & Amador, A. D. (2009). Data mining tools. Cuban Journal of Sciences

Informatics, 3(3-4), 73-80.

[13] Pérez López, C., & Santín González, D.

(2007). Data Mining. Techniques and tools:

techniques and tools. Editorial Paraninfo.

14] Durán, E., & Costaguta, R. (2007). Data mining to discover learning styles. Revista Iberoamericana de Educación, 42(2), 1-10.

15] Marcano Aular, Y. J., & Talavera Pereira, R. (2007). Data mining as a support for business decision making. Opción, 23(52), 104-118.

16] Miranda, M. A., & Guzmán, J. (2017). Analysis of university student attrition using data mining techniques. Formación universitaria, 10(3), 61- 68.

17] Febles Rodríguez, J. P., & González Pérez, A. (2002). Application of data mining in bioinformatics. Acimed, 10(2), 69-76.

18] Jaramillo, A., & Arias, H. P. P. (2015). Application of Data Mining Techniques to Determine Student Interactions in a Virtual Learning Environment. Revista Tecnológica-ESPOL, 28(1).

19] Alcover, R., Benlloch, J., Blesa, P., Calduch, M. A., Celma, M., Ferri, C., ... & Zúnica, L. R. (2007). Analysis of academic performance in the studies of

at the Polytechnic University of Valencia applying data mining techniques. Teruel, Spain. Available at:

http://bioinfo.   uib.   es/~ joemiro/aenui/procJenui/Jen2007/laalta. pdf.

 20] Castillo, G. M., & Campaña, M. E. M. (2010). Integration of Data Mining and Multiagent Systems: a field of research and

development. Information Science, 41(3), 53-56.

 21] Gorbea, P. S., 2013. Transdisciplinary trends in information metric studies and their relationship with information and knowledge management. Perspectivas em Gestáo & Conhecimento V. 3, N 1, pp.13-27.

 22] Coto, F., Cristaldo, P. R., Schab, E. A., Nuñez, J. P., Rivera, R. A., Retamar, M. S., ... & Bracco, L. J. (2018). Data mining and information visualization.

23] De Battista, A., Nuñez, J. P., Retamar, S., Bouzenard, D., & Herrera, N. E. (1970, November). Data mining applied to massive data. In XVIII Workshop of Researchers in Computer Science (WICC 2016, Entre Ríos, Argentina).