Business Intelligence and Machine Learning.
Integrated cloud solutions providing business insights for decision makers

Laura - Gabriela TĂNĂSESCU
The Bucharest University of Economic Studies, Romania
lauratanasescu@gmail.com

The aim of this paper is to present the latest trends in business intelligence and ways in which nowadays organizations can implement cloud technologies. This work is going to present challenges of the market, providers of integrated cloud business intelligences tools, advantages and disadvantages of moving to the cloud. A real life use case will argue the importance of taking advantage of data, as well as the necessity and the obvious benefits of having the right tools of transforming data into correct business decisions.

Keywords*: business intelligence, cloud computing, artificial intelligence, analytics, machine learning, innovation, data*

1 Introduction

Technology has evolved a lot in the last decade and nowadays we can even talk about a new paradigm related to cloud and how cloud technologies are going to influence organizations and their development.

In the same time, there are plenty of talks about a revolution related to data. How big data has been developing in recent years, how is going to challenge artificial intelligence our everyday work and what is the way in which organizations adopt business intelligence in order to gain insights from this data.

Therefore, the aim of this paper is to talk about the recent trends in technology, offering clear but relevant information about the most important concepts. In addition, the paper is going to provide an example about business intelligence applications in real life use cases, using cloud technologies from one of the top providers of cloud.

In the following chapters, this work is going to talk about all the theoretical concepts mentioned before in order to provide a clear image of the domain. Afterwards, it is presented an analysis of Oracle Corporation and its analytics solution in cloud, with advantages and disadvantages, competitors and benefits. Finally, a use case is going to be realized with this technology.

2 Cloud technology

This chapter is going to present some theoretical aspects related to cloud, as well as types of it and what are the advantages and disadvantages of using it.

2.1 Cloud computing

We can refer to cloud computing as the possibility to provision computing services with the help of the internet, services where we can include networks, software, servers, analytics and databases. All these cloud capabilities are used to offer a faster innovation and a more flexible way to use resources. [1]

A concept (**Fig. 1**) that comes with cloud technology is that the locations of the service used, the hardware, all the operating systems and also many more other details remain irrelevant to the final user. [2]

Practically, cloud providers offer services that enable the users to access, store or transmit file or applications on different remote servers as well as the power to access all the data using the internet. This being said, it is not required for any user to be in a specific place in order to gain access to it. [3]

Fig. 1. Cloud computing concept [4]

2.2 Cloud classification

Cloud computing is known as public or private cloud. The first one refers to those services that are offered to users for free over the internet. The second one provides access just for a number of people, offering services that are a system of networks. Here, we can also mention a third category, known as a hybrid cloud, that is a combination of the two described above.

Cloud computing cannot be seen as one piece of technology, but it is divided in three different services: software-as-a-service (SaaS), infrastructure-as-a-service (IaaS) and platform-as-a-service (PaaS). The first one is related to the part of the license for software applications that is offered to customers, license that is provided with a pay as you go model.

The second one provides customers the opportunity to practically rent infrastructure that includes servers, storage, operating systems and networks from any of the cloud providers of infrastructure.

The last one is especially designed in order to make it easier for developers to create web and mobile apps, without having the need to manage or set an environment and infrastructure for the development process. [5]

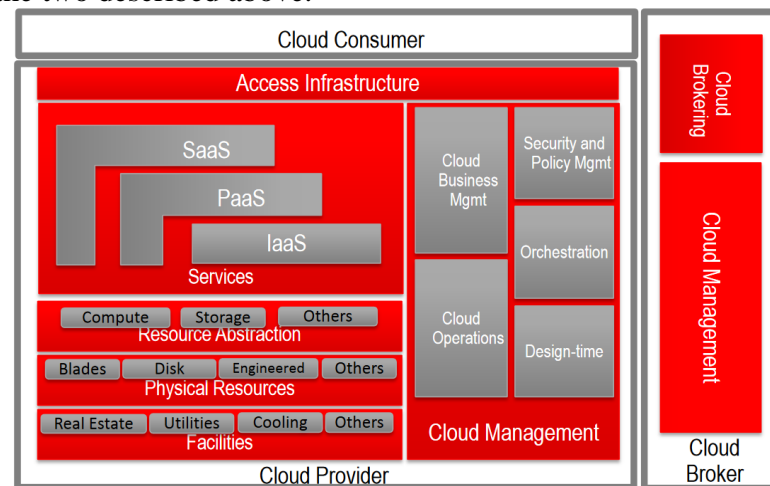In **Fig. 2** can be seen a complete and detailed architecture of cloud.

Fig. 2. Cloud architecture

2.3 Advantages and disadvantages

Looking at this new technology that impacts our life nowadays, it is important to talk about the benefits that come with it. So, the first one to mention is related to costs. The fact that cloud computing eliminates all the expenses that were coming with the hardware and the software, as well as with setting up and running all the data centers, it is needless to say that the overall costs are being reduced. The second one is effective for all the customers and also for the technical users and it is about speed. All the services are provided as self-service and on demand, as well as the fact that every service can be provisioned within minutes and without a great amount of knowledge. The third benefit that deserves being mentioned is related to performance that comes from the fact that cloud services run

on a worldwide network of secure data centers. In addition, the performance comes also from the upgrades that are regularly being made to the systems, making them always faster and more efficient.

Last, but not least, we must talk about security. There are plenty of systems that cannot assure a good security due to the lack of knowledge or the missing budgets for improvements, so all the technologies and policies that are offered from the cloud providers offer a very important and needed secured system for customers.

Having mentioned all the benefits from moving to cloud and accept the innovations, it is equally fair to also talk about the downsides. The most important one that can be identified is again related to security. Moving and working with sensible data to a cloud that runs on a different country, for example, can cause concerns.

There are also several regulations that are unclear when talking about whether or not some critical national data can be stored in another country where the data center physically is. So, this is a risk that an organization should consider when trying to adopt cloud technologies.

Furthermore, the fact that just one portal is used by multiple employers at the same time, manipulating data and making changes too, can cause damage to the overall course of work.

3   Business Intelligence

This chapter is going to present some aspects of business intelligence as known today, as well as how this technology is used in cloud.

3.1 BI concept in nowadays technology

Business intelligence can be defined as a software application that is realized in order to analyze, report and offer visualizations of data. The entire procedure that includes reporting data, analyzing it and also accessing all the sources are achieved by a business intelligence software. This concept covers multiples directions like applications, technologies, processes and tools, as well as practices of translating relevant conclusions. [1]

As mentioned above, business intelligence is, in fact, a process driven by and with data, where data storage and knowledge management make a combination that helps in the business decision process. [1]

BI technologies (**Fig. 3**) are used in order to help organizations achieve better decisions about the existent processes, requiring skills, relevant data and innovative technology. BI can be extended as a concept where it can include not only applications and tools, but also infrastructure and practices that enables those organizations to analyze information faster and better, optimize processes and formulate relevant conclusions and taking decisions. [1]

We can admit that a successful BI implementation should be focused on software development or hardware, but on the value that comes from information.

Taking this into considerations, it is important to understand the way data is created and used, what is the quality of that data, how it is constructed that system and the service levels.

So, coming to a point where a conclusion of business intelligence concept is needed for nowadays technology, we can best define it as a set of business data that is taken from multiple sources which are translated into information using different applications in order to support decision making and help organizations to achieve their needs. [6]
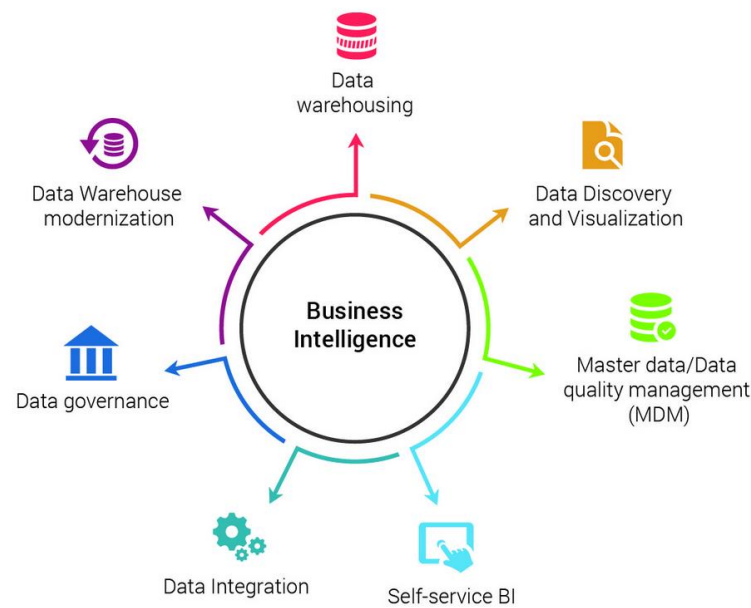
Fig. 3. Business Intelligence concept

3.2 BI in cloud

When using business intelligence solution in cloud computing environment, we should underline the great opportunities this combination can offer. Even though both of these technologies are at a starting point of their development, they are in trends for most of the organizations, having some difficulties to solve though. One of the main problems with these two is related to integration. Of course it is relevant to add here the costs that come from reorganization of processes and work, as well as from people trainings. Not only will these costs appear, but it is also possible that many employees will have a bad attitude towards change. Here it can be added the lack of resources to support these changes, the possibility of downsizing the targeted departments and also the uncertainty that comes with adopting new technology. Nevertheless, every risks and disadvantage that were mentioned about any cloud technologies can be translated to this combination of applications too.

3.3 Methods used to create a business intelligence system in cloud

In order to adopt such a system, there are several steps to follow so that the final result should be the one expected.

The first thing to consider is about data collections, where it is important to have the means of accessing and integrating all the places from where the data can be taken. In addition, an architectural model has to be proposed in order to have the best way to collect data.

Another step in this process is the validation part that also comes with reliability. So it is proposed to solve the reliability issues to ensure correct measurement accuracy and also the right measuring instrument used in the measuring process. Reliability is in fact used to phrase the measurement to which a metric provides correct results and no random errors.

The third part is about data preparation. In order to prepare data, we have to first collect all the data, combine it from all the sources where the data was found, structure it in order to be clear and organize it so that it can be easily analyzed. Analysis of the data comes with a process where statistical or logical techniques can be applied in order to illustrate, recap and evaluate data.

The last part and the one that also brings business value and insights is about data

analytics. The most common way used was descriptive analysis, where reports were added. After big data has started entering in the biggest companies on the market, the traditional business intelligence has changed due to speed and ways of storing. Therefore, predictive and normative analysis has emerged lately, the firs ones being in the spotlight as well.

The evolution of big data and analytics has affected the overall way of business intelligence delivery. Information needs to be quickly extracted from data, organizations being more and more concerned about normative and predictive analytics that include machine learning capabilities and rapid ways of building relevant visualizations.

4   Artificial intelligence

In this chapter, the paper presents some theoretical information, as well as a brief introduction in artificial intelligence. Moreover, it is going to be made the obvious relationship between data and artificial intelligence in nowadays businesses, as well as the important part named machine learning.

4.1 AI definition and structure

While it is mentioned before that business intelligence works with the aim of collecting, reporting and analyzing data, artificial intelligence comes with another approach that impacts data.

In fact, artificial intelligence enables computers to make their own decisions. Thus, we can define artificial intelligence as the ability of a machine or computer to learn and think like human's brain.

Artificial intelligence contains subfields like machine learning, neural network, deep learning, compute vision and natural language processing (**Fig.4**). Explicitly, machine learning is working to automate analytical model building. This field uses different methods like neural network, operations research and statistics so that to find hidden insights from data. [7] [8]
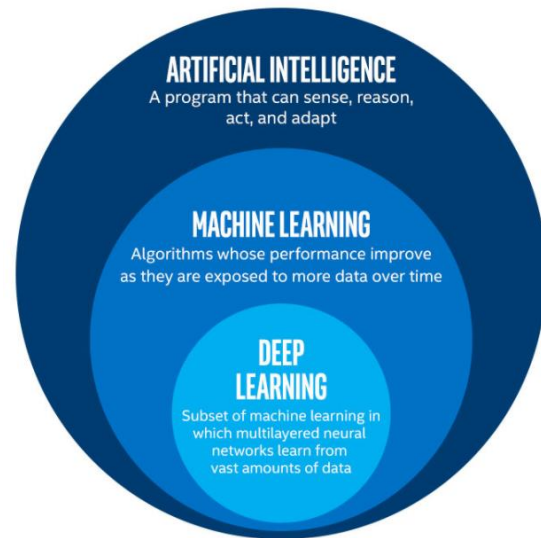

Fig. 4. Artificial intelligence and its categories

4.2 Data and AI

There are multiple sectors of economy that deal with huge amounts of data which are available in different formats and sources. This enormous amount known as big data is becoming available and easily accessible due to the progress of technology. Multiples data applications of machine learning are formed through complex algorithms build into a machine or computer. The code used creates a model that identifies the data and, after data, it is building predictions around it. The model is going to use parameters built in the algorithm in order to form patterns that are going to help the decision making process. When new data is added to the process, the algorithm used will adjust those parameters mentioned before in order to check if the patter has changed. However, the entire model should remain the same.

AI along with Machine Learning and Deep Learning present multiple technologies that are utilizing Tensor Process Unit (TPU) and Graphics Processing Unit (GPU).

4.3 AI applications in cloud

Apart from the visualizations of data that are done using a business intelligence software, we can talk separately about

what is the value that cloud brings to the machine learning component.

Therefore, there are many reasons to talk about regarding using machine learning in cloud, along with business intelligence. First of all, it is about the leverage and speed provided by the power of the GPUs that are needed to train different algorithms, without investing a lot in hardware. Moreover, the scale up and down capability make it efficient and easier for users to improve the power depending on the needs and measure a project have.

In addition, the new picture offered by cloud providers in terms of business intelligence and machine learning does not require advanced skills and lots of knowledge in data science and programming.

5 Oracle as a cloud provider of analytics platform

In the following lines, the paper introduces Oracle as a cloud provider, as well as an interesting player in data and analytics market for cloud users. Finally, this chapter also propose a demo of Oracle analytics platform that is going to demonstrate the benefits of using cloud for analyzing data.

5.1 Oracle Analytics Cloud

One of the top cloud providers that also comes with an analytics platform is Oracle, which is proposing a comprehensive tool in a unified platform, including data preparation for enterprise reporting, self-service visualizations, advanced analytics, self-learning analytics and machine learning integration on top. (**Fig.5**)
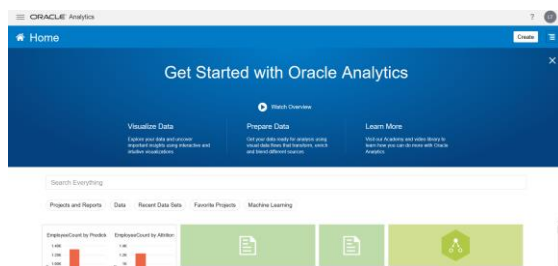

Fig. 5. Oracle Analytics Cloud interface

In the capabilities of this cloud platform we can enumerate data discovery, which helps users to easily collaborate with others, building intelligent analysis, machine learning models and statistical modeling.

Another thing to mention is related to the fact that developers can utilize interfaces that help them extend and customize all the analytics experiences in the flow.

It is very interesting the fact that in Oracle Analytics Cloud users can take data from any source, collaborate on project with others and explore real time data. Furthermore, unlike other providers that require the user to compromise between self-service, governed and centralized analytics, Oracle Analytics Cloud (OAC) solves this problems by offering a single solution that also incorporates Machine Learning and Artificial Intelligence.

Through the capabilities of OAC we find the data preparation enrichment that is built into the analytics platform. Another one is the business scenario modeling, a self-service engine for industry that helps in multidimensional and visual analyses. Moreover, we see here that proactive mobile that always learns from your work and offers contextual insights in daily activities. Last, but not least, is the enterprise reporting capability, the power of security and governance having a semantic layer which maps complex data into familiar business terms.

5.2 Augmented analytics – features of OAC

Keeping in mind the concepts mentioned above, we can converge business intelligence, artificial intelligence and more specific machine learning, into a term named augmented analytics.

We can see this concept as an evolution for the foundation build from analytics and business intelligence as well as big data, combining different and emerging technologies.

While business intelligence is about creating and finding data insights, AI and ML are about learning from different datasets in order to offer machine-driven decisions.

As it is known at the moment, a BI platform actually ingest a lot of data from multiples sources before anyone can prepare and reorder data.

An augmented analytics system is taking these latter steps and automates them using machine learning and artificial intelligence technologies. As an explanation, machine learning handles data preparation and artificial intelligence handles initial analysis.

Looking at the benefits of such a concept, we can tell that, in spite of those that are offered by multiples providers, there are some that offer a level of efficiency and accuracy that is possible due to computer processing. Thus, one of the most important aspect of augmented analytics includes accuracy. If the analysis is made by data scientists, there is likely that a mistake is going to occur. When using machine learning for that, these situations are eliminated from the beginning. [9]

Another thing to consider is speed. There are gaps that can appear when we first initiate a project using a BI platform like when we manually prepared data and also wait some time in order to receive an answer from different parties. Using augmented analytics, this process begins immediately, launching AI to cull the specific and needed data and also to begin the drilling down for the specific output needed for the project.

One more aspect to consider is the reduction of bias. Bias does not have to come as a personal shortcoming, but as a habit or a routine. Humans tend to revert to patterns so there can be a blind spot for data scientist that can lead to overlooked insights. In this case, computers and machines are going to work more efficient without inherent bias.

Last thing on this list is about the resources used. Augmented analytics can increase the resources by having them do more important things than some manual labor. So, for data scientist, it is going to mean more time to create different business problems and extract deeper insights form data. [9]

5.3 Oracle versus competitors

One of the advantages that Oracle Analytics Cloud has, as CEO of Red Pill Analytics said [8], is the fact that Oracle offers all in a single solutions. In fact, there is known that other providers offer multiple products in order to satisfy the same need and the problems is this process takes more time, resources and configurations before getting value from the investment.

Another thing to consider is the ability to scale up or down in order to adjust the resources, depending of the nature of workloads.

Also, its ability to offer not only visualizations, but a comprehensive view that helps the enterprise is considered important by another group director of Qubix International [8].

6   HR Attrition case study using OAC

Having all these concepts about business intelligence, artificial intelligence and cloud technology in mind, a small demo can be easily provided. An HR data set added in Oracle analytics cloud is going to be used so that to present the advantages and extended possibilities for data analysis.

6.3 Data   loading   and   hypothesis formulation

We are going to use a public data set about employees and some details about them, as well as staff attrition.

These data set contains details about age, department, hours worked, over time, distance from home, daily rate, education, employee satisfaction, gender, job level, job role, marital status, relationship satisfaction and years since promotion.

These variables are considered suitable in order to make an analysis for the

organization's employees (in order to see their satisfaction and problems based on work life balance, benefits and capabilities) as well as realizing an algorithm in order to predict whether or not an employee that we do not know anything about is going to leave the company or not.
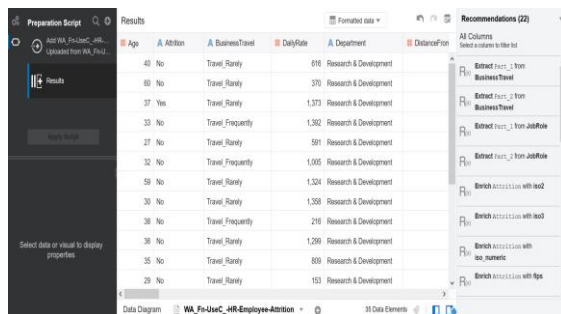

Fig. 6. Data loading menu for OAC

6.4 Data preparation

As mentioned before, Oracle Analytics Cloud, the tool used for the analysis, is offering intelligent recommendations and possibilities to arrange and filter data, as well as change a measure into an attribute and vice versa.
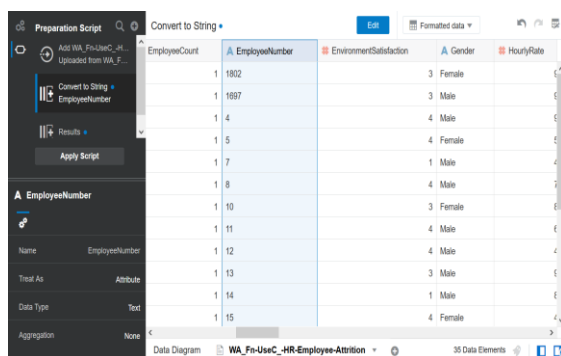
Fig. 7. Data preparation menu

There is also a part provided by some machine learning capabilities of the tool that is helping the users to enrich the actual data. Depending on this, we can add this recommendations or not.
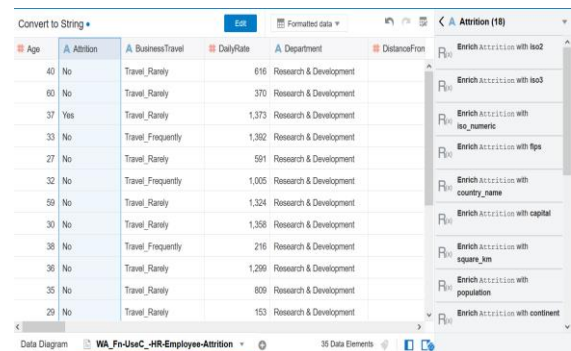

Fig. 8. Data enrichment capability

In the same time, the explanation mode that can be seen below is part of the augmented analytics. Thus, for a numeric variable like monthly income, we see that the tool offers us different graphics that are relevant for the analysis.
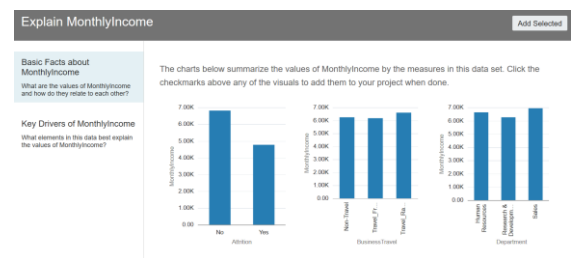

Fig. 9. Explanation mode for monthly income – basic menu

Finally, the same thing can be seen in key driver tab that shows us which are the variables that best explain monthly income (**Fig. 10**).
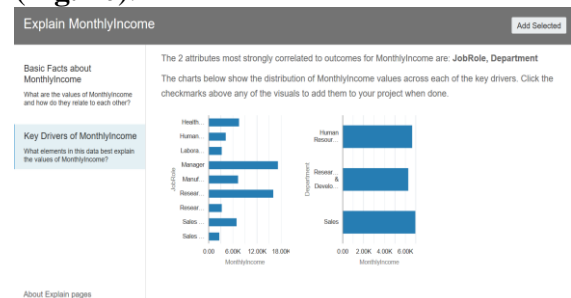

Fig. 10. Explanation mode for monthly income – key driver menu

6.5 Data visualization

One type of dashboard we can make is a general one that provides us with general information about that data we use.

32

*Business Intelligence and Machine Learning. Integrated
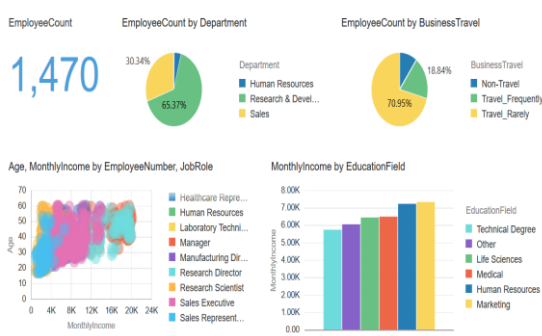cloud solutions providing business insights for decision makers*

Fig. 11. Data visualization for general
overview

Therefore, we can see in **Fig. 11** that the total number of people analyzed are 1470. The majority are working in research and development and after that sales. In addition, we see that 70% of the people travel rarely, while 19% travel frequently.

In the same time, if we look at the monthly income of the people, as well as the role they have and the age, we see that the greatest income is for research scientists and research directors, with age between 35 and 60. On the other hand, we have the minimal income values for sales representatives and laboratory technicians, where the start age begins with 20.

Last, but no least, we see that marketing provides the greatest monthly income, while a technical degree provides the smallest value when analyzing this organization.
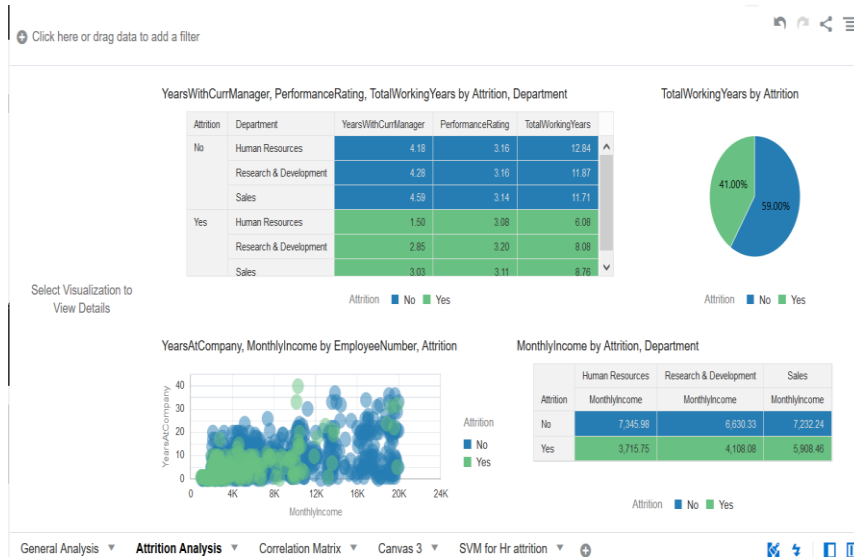

Fig. 12. Data visualization for attrition

The second dashboard (**Fig. 12**) provides information about attrition. We can see from the charts that most of the people that are going to leave the organization are not for many years with the current manager, that they have lower performance ratings and lower total working hours. This shows us that people who will stay within the organization have a history in it, they dedicate a lot of time to it and they perform.

In the same time, most of the people that are going to leave have a smaller income and they have been with the organization for little time than the others.

Nevertheless, we see that the younger employees are those that will leave the organization.

Another interesting thing to consider is the fact that the tool provides an instant visualization of correlation between variables like in the picture below.
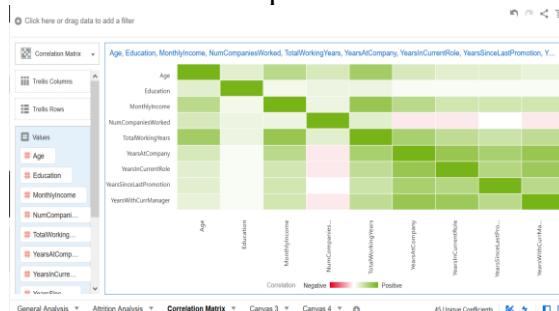

Fig. 13. Correlation matrix

Not only (**Fig. 13**) is this correlation easy to build because it does not request any technical knowledge, but it instantly provides useful information about our data like the fact that there are powerful and positive connections between age and total working years, monthly income and total working years, age and total working years.

On the other hand, we see some negative connections between years in current role and number of companies worked and number of companies worked and years with current manager.

6.4. Machine learning for HR use case

For the machine learning part of this project, we are going to build different machine learning algorithms of classification. After building them, an analysis is useful in order to decide which model is the best one for the use case and that model will be used for prediction.

For the HR attrition use case we are going to use support vector machine algorithm and Naïve Bayes algorithm, both useful for the binary type of classification.

For building this models, we are going to use the data flow that is available in Oracle Analytics Cloud. (**Fig. 14**)

As mentioned before, this is step does not require programming or very advanced technical knowledge.
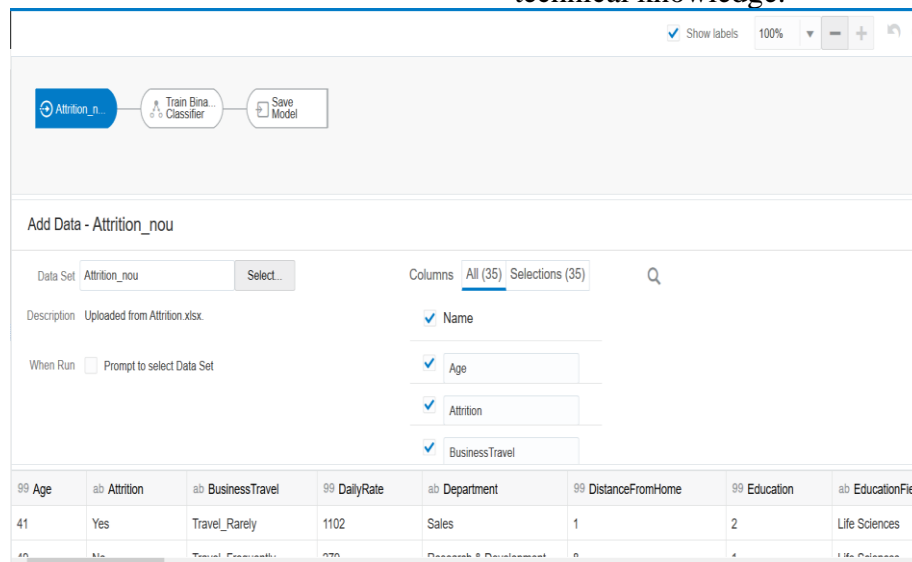

Fig. 14. Data flow menu

After applying the chosen algorithm to the data set, we obtain a model that is going to be inspected in order to see how good or bad that model is.
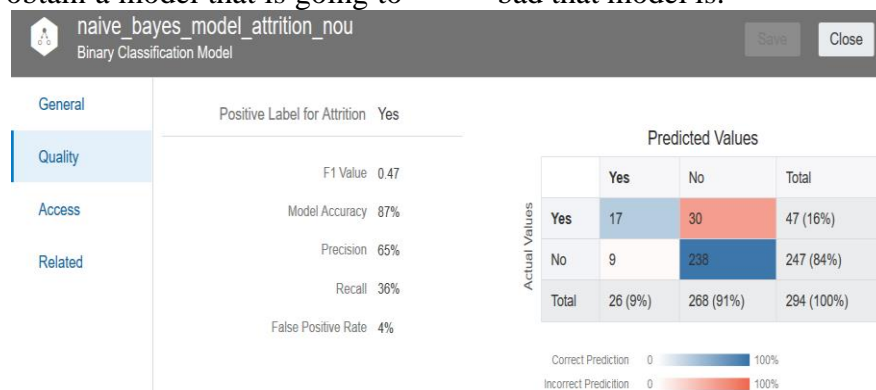

Fig. 15. Model analysis for Naïve Bayes

The first one is built for Naïve Bayes and we have an accuracy of 87%, with a precision of 65%, a recall 38% of and a false positive rate of 4%. (**Fig. 15**)
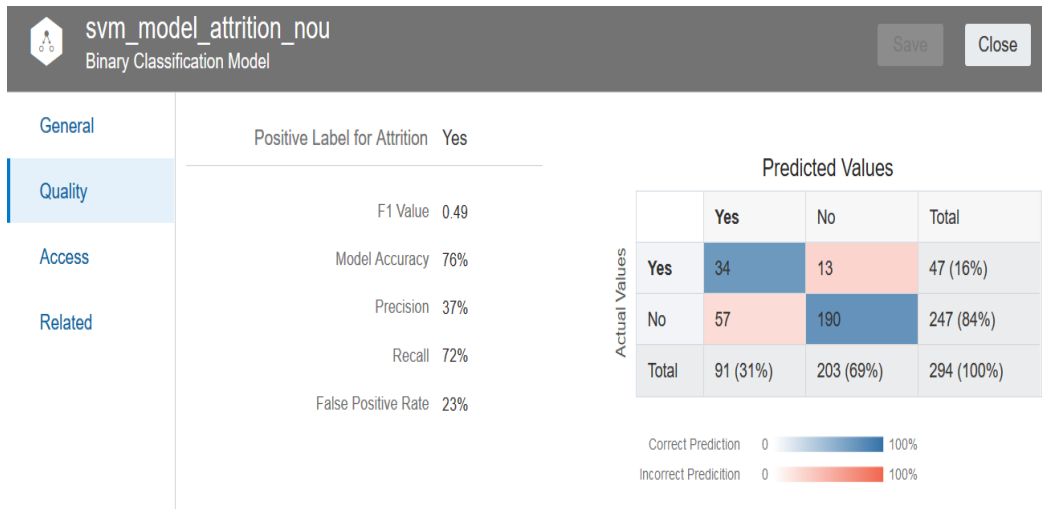

Fig. 16. Model analysis for SVM

For the second one we have an accuracy of 76%, a precision of 37%, a recall 72% and a false positive rate of 23%. (**Fig. 16**)

Therefore, in order to choose the best model, we are looking at the fact that the model is predicting attrition. In other words, we want to maximize the recall because we want to predict correctly all truly positive cases.

Lastly, a false negative for these models points out to the idea that we are wrongly going to conclude that a person is going to leave the company, fact that might decrease the chances to prevent a person from leaving.

That being said, we can use support vector machine in the next steps so that to predict new values for possible leavers.

Using the same data flow where, to the initial set of data, the support vector machine model is applied to the data, we are going to obtain a prediction that is visualized in the pictures below.
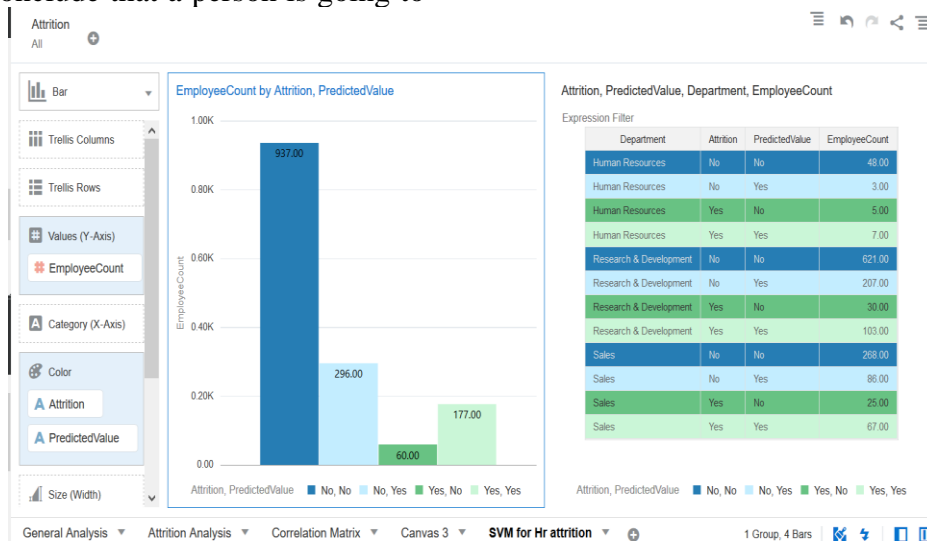

Fig. 17. Visualization for SVM model prediction

So, from the total number of people, we can see that most of them were predicted for attrition correctly. 937 have no for attrition and they were also predicted with no, 177 have yes and their prediction was yes too. The interesting part is that, like

remembered in the previous paragraph, 296 were mistaken, but for our analysis is better to think that 296 are going to leave, even though they are not. Finally, the smallest number is 60 for those that are going to leave and they were actually predicted as non-leavers. (**Fig. 17**)

In the next chart, we see the attrition and prediction for attrition grouped by department. (**Fig. 17**)

We see that most of the leavers are from research and development, as well as from sales. These two departments have also the biggest numbers for the false negatives, but the predictions are overall very good.

6.5 Use case results and proposal for improvements

We have seen through this use case the population that we analyzed. So, we have different people working in sales, research and human resources, that have experimented different levels of income, working years, type of managers, number of trainings, levels of job satisfaction, work life balance and more.

For these people, we have seen that those with lower levels of income, people that work for little time in the organization, that are little experience or that have been working for less time with a manager are going to be exposed to attrition.

So, in order to prevent this event, a machine learning algorithm of classification is being used so that to predict the possible employees that can leave. We have chosen the best one, more exactly the support vector machine that offered the minimal false negative rate.

With this algorithm we have predicted those employees that might leave and, the targeted organization can now address to them in order to find solution to the existing problems.

As seen from the data used, some of the solutions to propose might include a solid plan for development that includes levels of income, ways of promotion and trainings. Moreover, organizations should adapt to younger people that tend to leave

early when something is wrong, on contradiction to those that are older and that have spent many years in just one place.

7   Conclusions

First thing to mention in this final part of the paper is the fact that, using all the technologies presented before, a business problem was solved within days and with little technical knowledge.

All things being considered, a BI tool like Oracle Analytics provides us an integrated platform that is going to support the work from preparation till predicting future behavior, facts that are going to help business decision makers to act faster and better in their daily work.

Artificial intelligence and machine learning was useful not only for the suggestion area, the explanation mode or enrichment part, but also as providers of useful algorithms that can be applied right away.

Last, but not least, the cloud has offered multiple benefits in the entire work process. First of all, the permissions and roles part that helps more users to work on the same project or use it at the same time. One administrator can be responsible to create all these users and give them the right privileges. [4]

Second of all, the power offered by the cloud in order to run machine learning algorithms is very important. Considering the fact that a support vector machine classification can take a lot of resources, it is clearly an advantage when we can run in just one minute an entire algorithm in order to build a model and a prediction.

Lastly, the platform can be integrated with other solutions, it can take data from other applications, database or personal computer and everything is going to be in one place.

References

[1] A. Pyae, "Cloud Computing in Business Intelligence," Asia Pacific

University of Technology and Innovation, November 2018.

[2] https://www.zdnet.com/article/what-is-cloud-computing-everything-you-need-to-know-from-public-and-private-cloud-to-software-as-a/, "What is cloud computing?," Steve Ranger, 2018.

[3] https://www.investopedia.com/terms/c/cloud-computing.asp, "Investopedia Cloud Computing".

[4] A. Banafa, "Ten Myths about Cloud Computing," https://www.experfy.com/blog/ten-myths-about-cloud-computing, 2019.

[5] https://azure.microsoft.com/en-gb/overview/what-is-cloud-computing/, "Azure Microsoft - Cloud computing".

[6] https://www.saksoft.com/information-management-services/business-intelligence/, "Information management services".

[7] https://www.sas.com/en_us/insights/articles/big-data/artificial-intelligence-machine-learning-deep-learning-and-beyond.html, "SAS Big Data and Artificial Intelligence".

[8] R. Clayton, "Oracle Analytics Cloud Succeeds Where Others Fall Short," Oracle Blogs, 2019.

[9] https://blogs.oracle.com/analyticscloud/what-is-augmented-analytics-v2, "What is Augmented Analytics?," Oracle Blogs, 2019.