***Original Research Article***

**ARTIFICIAL LEARNING BASED ON SVM: APPLICATION TO CHURN ANALYSIS IN A TELECOMMUNICATION STRUCTURE**

**1. Abstract**

This paper is the result of our research on the analysis and classification of churning customers in a telecommunications company. These data, often heterogeneous and coming from various sources, require in-depth analysis as well as new storage and exploration paradigms to extract value from them.

In the telecommunications sector, companies accumulate large amounts of information about their customers, coming from multiple sources: social networks, telephone platforms, electronic messaging, open data, geolocation, and many others. The intelligent exploitation of this data allows to better understand user behavior and anticipate key phenomena, such as “ churn ” – i.e. customer unsubscription.

Churn is a major strategic issue for telecommunications companies, as customer loss leads to high costs related to new subscriber acquisition and reduced revenue. Thus, identifying customers at risk of churn and understanding the underlying factors are essential to implement preventive actions and build customer loyalty .

In this study, we propose a machine learning model based on support vector machines (SVM) to analyze and classify churning customers. This algorithm, recognized for its ability to handle complex and multidimensional data, is implemented using the LIBSVM library in C# language. The objective is to build a powerful predictive model to identify, with high accuracy, customers likely to leave the operator, in order to optimize retention strategies and maximize customer satisfaction.

*Keywords: Machine Learning, Support Vector Machines , churners , Clients,Telecommunications.*

**2. Introduction**

Every day, humanity generates an exponential amount of data. Faced with globalization and increasing competition, the information available via the Internet and the many connected objects continues to increase. This explosion of data, often heterogeneous and from diverse sources, poses major challenges in terms of storage, analysis and exploitation. Their efficient processing therefore becomes a fundamental issue, requiring new analytical approaches and innovative tools to extract usable knowledge.

In a context where decision-making has become crucial for business and institutional leaders, the effectiveness of these decisions relies on access to relevant information and the use of advanced technologies capable of transforming these masses of data into added value. Machine learning is now one of the pillars of information technology, providing powerful solutions for exploring vast volumes of data. Thanks to sophisticated algorithms, this discipline makes it possible to discover hidden correlations, classify individuals according to their behavior and predict trends, thus providing valuable assistance in strategic decision-making.

In this context, we are interested in the application of machine learning, in particular the support vector machine (SVM) algorithm, to analyze and classify the behavior of Airtel Congo subscribers . The objective is to understand the consumption habits of customers through the purchase of packages (calls, mobile data, minutes, SMS, etc.), the reception of calls and messages, as well as the use of the Mobile Money service. By exploiting this information, the predictive model aims to identify subscribers at risk of unsubscribing and anticipate their behavior, thus allowing the operator to implement appropriate strategies to improve customer loyalty and maximize revenues.

**3. Artificial Learning**

### 3.1. Type of learning and techniques used

### Several types of learning can exist but grouped into two according to the techniques used to explore the data, namely.

**3.1.1. Supervised learning or predictive technique**

In this technique, the classes (or individuals or learning data) are predetermined and the examples are known in advance and the system learns to classify according to a classification model . It is a technique that acts on the data while having additional information about this data. The predictive technique is divided into two groups: classification (target variable is qualitative) and prediction (target variable is quantitative).

**3.1.2. Unsupervised learning (in English clustering )**

In this analysis or technique, the system or operator has only examples, but no labels, and the number of classes and their nature have not been predetermined. No expert is required. The algorithm must discover the more or less underlying structure of the data on its own.

**3.1.3. Goals**

The goals of machine learning are given by two sciences. In modeling science, which consists of searching for underlying (hidden) regularities in observed data. Thus, we learn in this science to understand and explain and predict and decide. In adaptation science, which also consists of searching but this time for situation-action patterns by interaction with the world. From where, we learn here to react to problems that may arise and anticipate unexpected problems.

**3.1.4. Training data**

Training data is often divided into three categories:

* **The training set** or **training population** : constitutes the set of candidates or examples (images, attributes, DB, etc.) used to generate the learning model;
* The Test set consists of the candidates on which the learning model will be applied (to test and correct the algorithm);
* **The validation set** : can be used during training (as a subpopulation of the training set) in order to validate (integrate) the model and avoid overfitting.

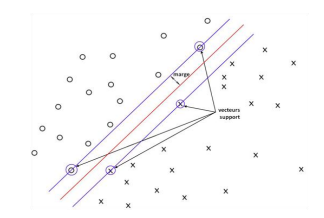
**4. SVM**

**4.1. Goals**

SVMs are used to solve the problems of discrimination, i.e. deciding which class a sample or group of samples belongs to, and [regression](https://fr.wikipedia.org/wiki/R%C3%A9gression) , i.e. predicting the numerical value of any variable. Solving both of these problems involves constructing a function that has an input vector mapped to an output : **.** They also aim to minimize the empirical and structural error while maximizing the geometric margin. In the case of a binary classification problem, SVMs aim to construct a decision function (separator) that will better separate the data and maximize the distance between two classes.

### 4.2. Operating principle

In the case of a binary classification problem, they construct a linear decision function (separator) which allows to better separate the data and maximize the distance between two classes.



H2

H

***H1***

X

**Y**

Vecteurs supports

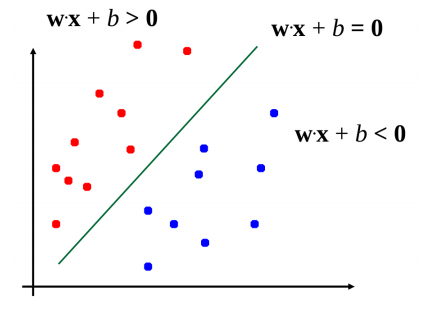
*Figure 1: Principle of the Wide Margin Separator (VMS)*

### 5. Perceptron Problem

Let us take the case of a linear classifier given by:

Where **s**

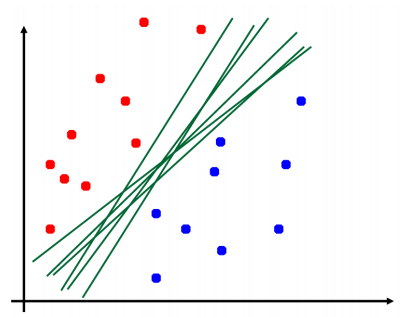
Graphically, it can be represented as follows:



*Figure 2: Simple Linear Classifier .*

But, since there are many possible choices for **w** and **b** :

The problem here will be to choose the best classifier among several possible classifiers , illustrated by the following graph:



*Figure 3: Presentation of several classifiers that can separate classes*

Several hyperplanes may be possible to separate a set of data as illustrated in the figure above, but SVMs seek a hyperplane that is optimal, that is to say the one that maximizes the margin, hence the name of Large Margin Separator. This is why, from the point of view of data separation, we distinguish linear SVMs (linear separator) and nonlinear SVMs (nonlinear separator):

### 5.1. Linear Support Vector Machines

Linear SVMs are when there is a linear function separating the dataset. However, they are approached in two ways, depending on whether the data is completely separable by the function or not.

**5.2. Separable cases (or Hard Margin SVM)**

Suppose we have a set of learning examples defined as follows:

: is a real vector, that is,

, represents the class of a data.

The linear separator or decision function defined by the SVM is given by:

, Or :

: is a vector perpendicular to the linear separator, called the weight vector;

: is the bais or a scalar;

: is a scalar product of the vectors and .

This linear separator does not provide vectors worth and , but we consider when the result ofis positive, the vector belongs to the label class and the vector belongs to the label class when this result is negative.

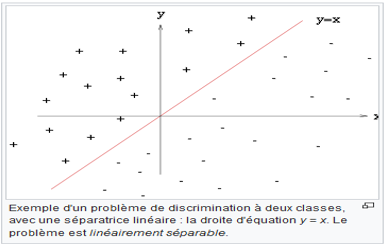
If we decompose the set *(* of training examples *)* into two subsets according to the value of and we therefore define:

And

And it is said thatis linearly separable, if there exists an and such that

And

Or simply:

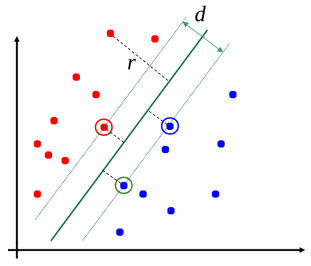
Let us suppose in a two-dimensional space in which two groups of points are distributed. These points are associated with a group: the points and the points . We can find an obvious linear separator in this case, the straight line with equation . Let us take the example of a two-class discrimination problem, with a linear separator the straight line , this problem is said to be linearly separable.

*Figure 4: Representation of a linearly separable problem*

To separate the samples, we will introduce the notion of the maximum margin which is the distance between the separation boundary and the closest [samples . The latter are called](zim://A/%C3%89chantillon_%28statistiques%29.html) *support vectors* . In SVMs, the separation boundary is chosen as the one that maximizes the margin.

**5.3. Concept of Margin**

We call margin ***d*** the distance between the 2 classes of samples or labels. It is this distance ***d*** that we would like to maximize ***[[1]](#footnote-1).*** The margin can also be defined as the distance between the hyperplane and the closest samples, the latter are called support vectors . In the separable case, we will consider the points closest to the separating hyperplane called support vectors . For any point in the space of examples, the distance to the separating hyperplane, represented by the figure below, is given by:



*H1*

*H2*

*H*

*Figure 5: Support Vectors and Distance to the Separating Hyperplane.*

**5.4. Margin Quantification**

However, it can be recalled that in a planthe distance from a pointrelative to a straight lineis given by:

Similarly, we can also deduce the distance from a vectorcompared withto the hyperplane **:**

Now for the points located on the hyperplanes H 1 and H 2 , that is to say the support vectors:

,

This implies that:

To limit the space of possibilities , we only consider the closest points, those which are located on the canonical hyperplanes, that is to saygiven by:

**;**

In this case, the margin to be maximized is:

Thus, the conditions for a good classification will be:

***5.5.* Determination of the maximum margin separator**

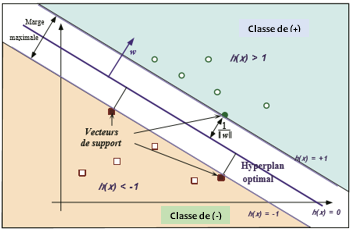
Maximizing the margin then comes down to findingAndsuch as:

is maximum , under the constraints of the equation:

Which can be combined into:

Equivalently, the problem can be written more simply as the minimization of:

The mathematical expression above is called the primal of the problem or primal formula.



*Figure 6: Representation of an optimal hyperplane.*

This minimization is possible under the so-called Karush -Kuhn-Tucker (KKT) conditions :

Let the Lagrangian be :

Kuhn-Tucker

The KKT conditions are then:

Moreover, the last condition implies that for any point not verifying **,** the is zero. The points that verify , are called support vectors. These are the points closest to the margin. They are supposed to be few in number compared to the set of examples. The decision function (or the hyperplane or the optimal separator) is given by:

***5.6.* Calculation of bias *(b)***

Not appearing in the dual problem, the bias ***b*** is calculated from a support vector:

.

Formore precision,we take the average of the support vectors which is given by:

**,**

Oris the set of support vectors.

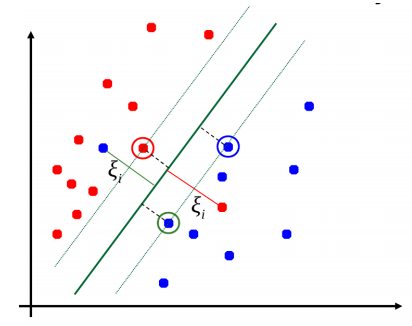
***4.7.* Classification of new data**

We saw above that the class of a data was given by the decision function signTo classify datawe have:

***4.7.1.* Non-separable cases (Soft Margin SVM)**

***has)* Problem formulation**

In case the data or samples are difficult to separate or not linearly separable, we try to add adjustment variables or to relax the constraints by introducing error terms.which will control the overshoot or which will take into account the classification errors or even the noises. A non-separable problem can be represented in the following way:



*Figure 7: Soft Margin Classification for a Nonlinearly Separable Problem.*

This problem is represented by the following mathematical expressions:

A datais well ranked if : **.** It is in the margin but well ranked if:and is misclassified in other cases.

So,indicates to what extent the datamaybeon the wrong side: ifis on the wrong side of the separator (hyperplane), the further it is from the separator, the moreis big.

So it is good superior to the number of classification errors.

***b)* Problem to be determined**

The problem here is therefore to find a hyperplane which maximizes the margin and which minimizes the permitted errors:

is a constant to control the trade-off between the number of classification errors and the width of the margin or an error penalization variable that trades off the size of the margin against misclassified data.

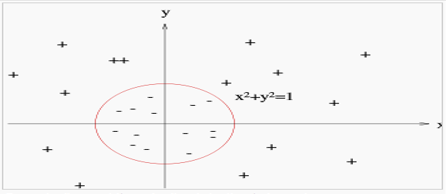
### 

### *6.* Nonlinear Support Vector Machines

#### 6.1. General idea

Suppose we are in a space where the data (samples or vectors) are not linearly separable. To separate these samples or these two groups of classes, no linear separator can properly separate them.

However, the problem is said to be nonlinearly separable, that is, there is no separating hyperplane that can separate the samples. And this can be represented by the following figure:

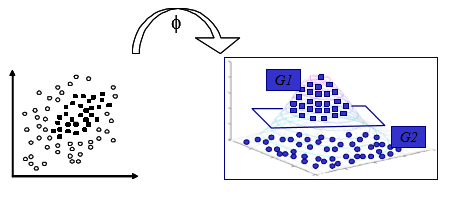


*Figure 8: Representation of a nonlinearly separable problem*

In this case, to separate these samples in order to remedy the problem of the absence of a linear separator, the principle of SVM consists in reconsidering the problem in a space of [dimension](https://fr.wikipedia.org/wiki/Dimension_d%27un_espace_vectoriel) greater than two or possibly of infinite dimension called redescription space or characteristic space.

In other words, the principle of SVM consists in determining a decision function which first passes through a transformation of the data space into another characteristic representation space, possibly of high dimension, where the data can become linearly separable.

In this new space, it is then likely that there exists a linear separator.



*Figure 9: Separation of elements in a 3-dimensional space of a nonlinear problem*

#### 6.2. Formulation

In the new data space, learning can occur:

, where and

Therefore, the optimization problem presented as follows:

#### 6.3. Classifying new data

From the above, we can deduce the decision function for classifying a newly given data.

To classify datawe have:

#### 6.4. Some kernel functions

For some characteristic spaces and associated applications, scalar products are easily calculable using specific functions, called kernel functions ( Kernel functions ) such as:

The interest of these functions is to make possible certain calculations of scalar products in the characteristic space without having to explicitly transform the data by the function , therefore without necessarily knowing this function .

Some kernel functions are:

* Linear:
* Polynomial: or
* RBF (Radial Basic Function):

## 

## 6.5. Evaluating the quality of a classifier

After building a prediction model, it is important to validate it, that is to say, to show whether the model is well built or not, by trying to estimate the classification errors it has made.

## 7. Application of SVMs

## 7.1. Presentation of variables

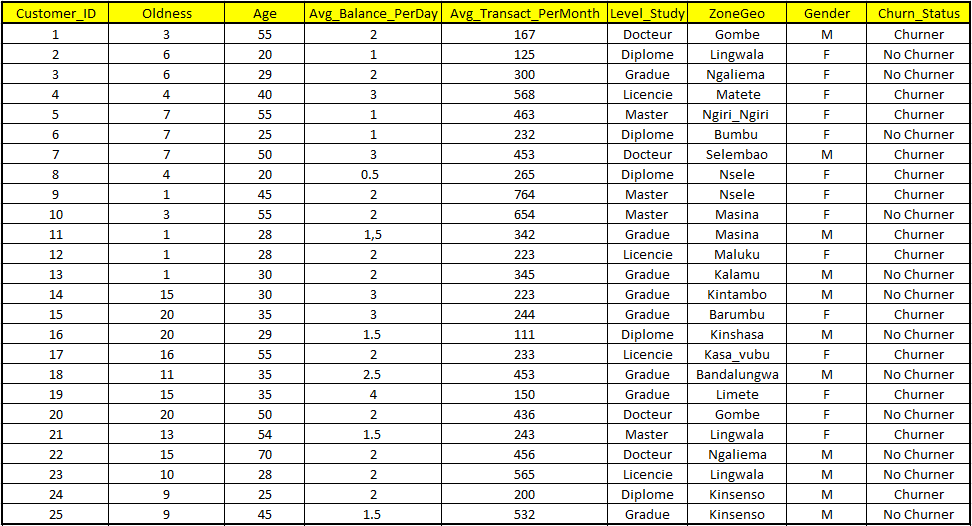
The dataset used for the implementation of our prediction model is based on the existing reality, within the telecommunications company named Airtel Congo; on the customer management policy, more precisely the customers who are candidates for churn , as described in the table below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No.** | **Variable designation** | **Variable meaning** | **Kind** | **Other choices** |
| 1 | Customer\_ID | Customer Identifier: Customer Identifier | Digital | - |
| 2 | Oldness | Oldness : Seniority of the Client | Digital | - |
| 3 | Age | Age of the Customer | Digital | - |
| 4 | Avg\_Balance\_PerDay | Average Balance Per Day: Average Balance per Day of the customer | Digital | - |
| 5 | Avg\_Trasanct\_Per\_Month | Average Transact Per Day: Average transaction per month of the customer. | Digital | - |
| 6 | Level\_Study | Level Study : Client's level of study | Text | - |
| 7 | Geographic Area | Customer geographic area | Text |  |
| 8 | Gender | Customer gender | Boolean | M or F |
| 9 | Churn\_Status | Customer churn status | Boolean | To Churner or Not to Churner |

*Table 1: Dictionary of variables.*

**7.3. Extracting training data**

Here we present the extract of the data that we will use for our training with the SVM:



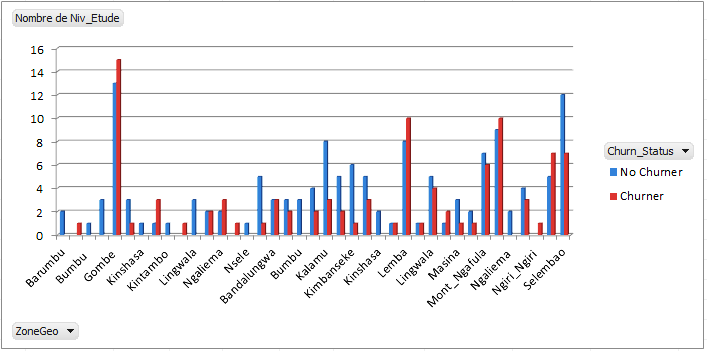
*Table 2: Extract of training data*

**7.4. Statistics of existing data**

As the algorithm has to predict based on existing data, we present some graphs expressing the behaviors of customers in different cases i.e. non-active customers (churner) and active customers (non churner). In other words, we present the two classes (Churner and Non Churner) of belonging to which new customers will be classified.

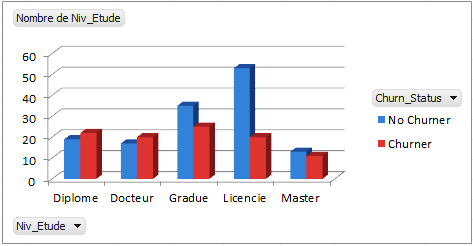
**7.4.1. According to the geographical area**

Depending on the geographical area, we present in red histogram, the customers who have churned or the lost customers and in blue, the non-churned customers or the active customers.



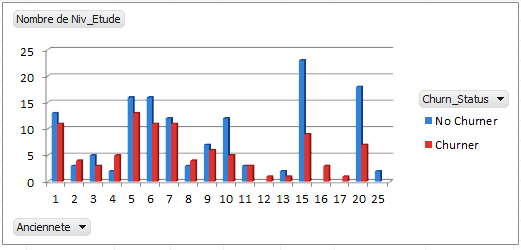
*Figure 10churner and non- churner customers by geographic area*

**According to the level of study:** active customers are represented in blue histogram and customers who have left or lost are represented in red histogram.



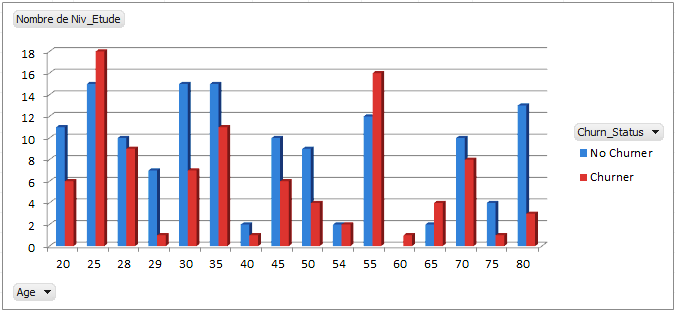
*Figure 11churner and non- churner clients according to the level of study*

**According to seniority:** Active customers are represented in blue histogram and lost customers in red histogram using the figure below:



*Figure 12churner and non- churner customers by seniority*

**According to age:** Customers, active customers are represented in blue histogram and lost customers in red histogram.



*Figure 13churner and non- churner customers by age*

**8. Introducing the Visual Studio 2015 Programming Environment**

To implement our system, we used the Visual Studio programming language version 2015 because it supports the SVM library ( LibSVM ).

**8.1. Data observation with SVMs**

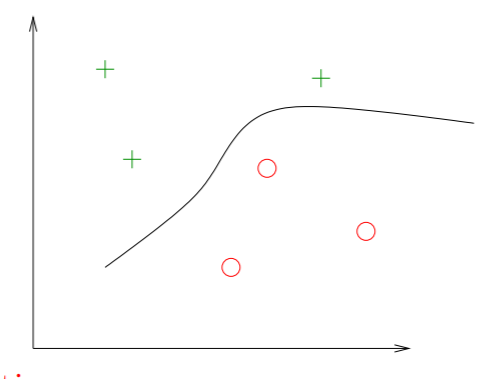
We are interested in a phenomenon (possibly non-deterministic) which, from a certain set of data: inputs , produces an output .

The goal is to find from the sole observation of a certain number of input- output pairs .

churn analysis , we want here to evaluate the risk of a subscriber leaving. Hence from training data, we find a decision boundary that separates the space into two regions (not necessarily connected).

### 8.2. Graphical illustration of the problem

We want to predict or classify the behavior of a new subscriber, from our training data, depending on whether the latter is a “ **Churner**» or “ **Non Churner ”. In our graph** churner behavior is represented by **+** (plus) and non- churner by (bullet).



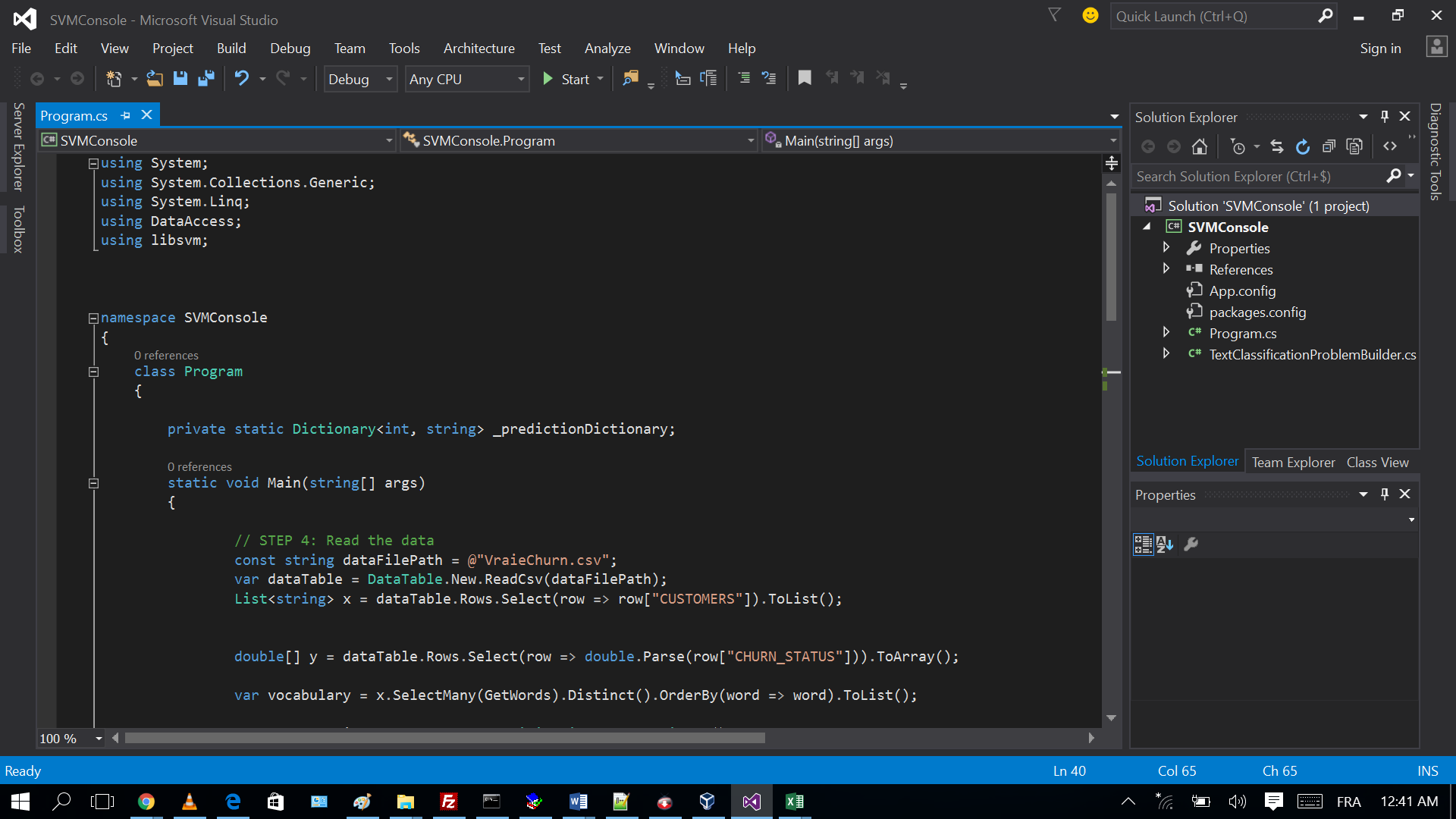
*Figure 14Churner and No Churner Customer Representation*

We find a decision boundary that separates the space into two (not necessarily connected) regions. We find that the hyperplane in this training example is optimal because the margins are maximal at all points.

The algorithm implemented in C# in console mode classifies the behavior of a new subscriber entered from the keyboard according to whether it is a or a . Below we present examples of the subscribers to be analyzed according to the characteristics of the training data.

**8.3. Presentation of the end of the C# code**

The interface below shows some bits of the C# source codes used to implement our project in console mode.



*Figure 15: Presentation of the end of the C# source codes*

**8.4. Prediction of new subscribers**

We will present two cases of new subscribers, each with its own ranking from the system.

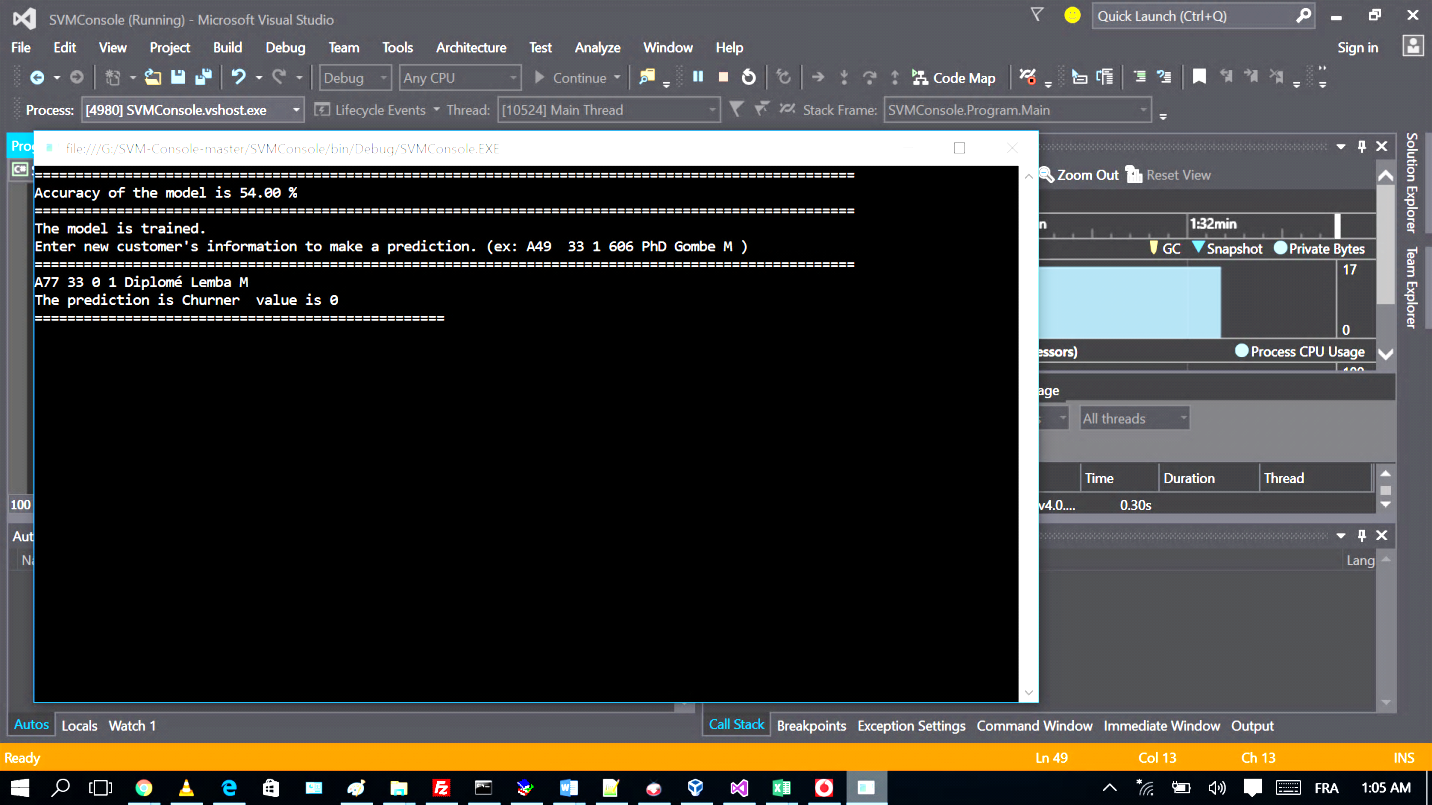
**First case: A. New subscriber**



*Table 3: Presentation of the first new subscriber*

**B. Classification from the system**

After having entered the information of the first new subscriber, we observe in the figure below that the algorithm has classified it in the " *Churner* " class, that is to say, given the behavior of the new subscriber entered in table 2 above, the algorithm classifies it among the non-active customers or among the customers who have lost.



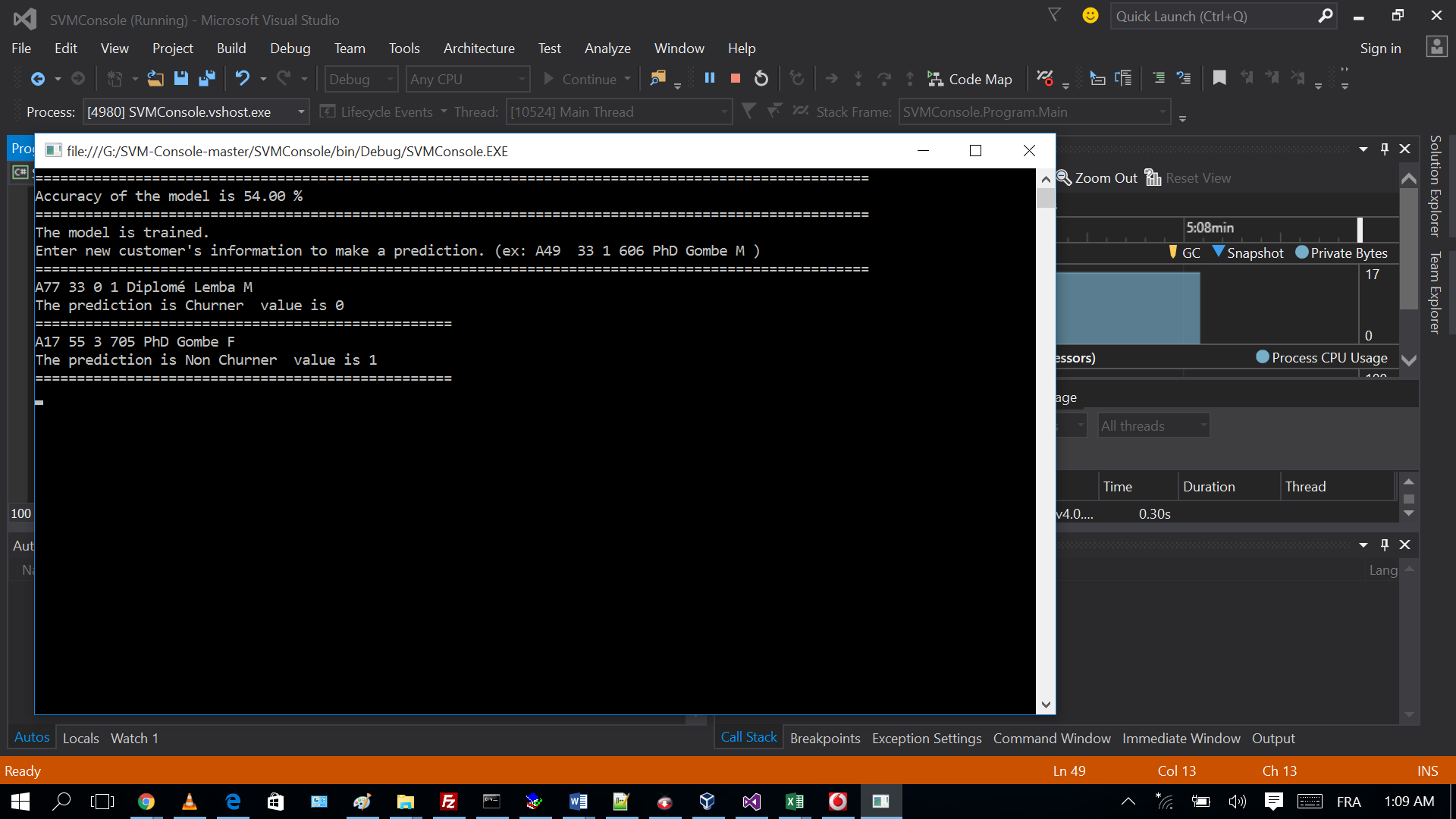
*Figure 16: Ranking of the first new subscriber with SVMs*

**Second case: A. New subscriber**

*Table 4: Presentation of the second new subscriber*

**B. Classification from the system**

After having introduced the information of the second new subscriber, we also observe in the figure below that the algorithm has classified this one in the class of, that is to say given the behaviors of the new subscriber introduced in table 3 above, the algorithm classifies him among the active customers.



*Figure 17: Ranking of the second new subscriber with SVMs*

**Conclusion**

At the end of this article, we discussed analyzing the "Churn" using artificial learning based on the SVM algorithm. Indeed, we have presented artificial learning which represents, in the field of artificial intelligence, a significant advance by offering methods and tools capable of processing and analyzing complex data. Based on various techniques such as supervised and unsupervised learning, it allows to discover hidden patterns and make accurate predictions. Support Vector Machines (SVM), in particular, illustrate the effectiveness of supervised approaches, by allowing an optimal separation of classes through the maximization of the margin.

**Reference**

1. Mitchell, T. (2006). An Introduction to Machine Learning . Antoine Cornuéjols .
2. Cornuéjols , A., Miclet , L., & Kodratoff , Y. (2003). Artificial Learning: Concepts and Algorithms (2nd ed .). Éditions Eyrolles .
3. Bishop, C. M. (2006). Pattern Recognition and Machine Learning . Springer.
4. Vapnik , V. N. (1998). Statistical Learning Theory . Wiley .
5. Hastie, T., Tibshirani , R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed .). Springer.
6. Cristianini , N., & Shawe -Taylor, J. (2000). An Introduction to Support Vector Machines and Other Kernel-based Learning Methods . Cambridge University Press .
7. Schölkopf , B., & Smola , A.J. (2002). Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond . MIT Press .
8. Cortes, C., & Vapnik , V. (1995). Support-vector networks. Machine Learning , 20(3), 273-297.
9. Zhang, H., & Schneider, J. (2010). Support Vector Machines: Theory and Applications . Springer.
10. Guyon , I., & Elisseeff , A. (2003). An introduction to variable selection. Journal of Machine Learning Research , 3, 1157-1182.
11. Liu, H., & Motoda , H. (2007). Computational Methods of Feature Selection . CRC Press .
12. Bennett, K. P., & Blue, J. (2000). Support vector machines: A new method for classification. Proceedings of the 2000 IEEE International Conference on Data Mining .
13. Friedman, J., Hastie, T., & Tibshirani , R. (2001). The Elements of Statistical Learning . Springer.
14. Rosasco , L., & Verri , A. (2004). Support Vector Machines for Function Approximation . Springer.
15. Koller , D., & Friedman, N. (2009). Probabilistic Graphical Models : Principles and Techniques . MIT Press.

1. Arnaud Revel, Vector Machines Support / Large Margin Separator [↑](#footnote-ref-1)