

## Research Article

### Designing A Method for Alcohol Consumption Prediction Based on Clustering and Support Vector Machines

Mendoza-Palechor Fabio, De la Hoz-Manotas Alexis, Morales-Ortega Roberto, Martínez-Palacio Ubaldo, Díaz-Martínez Jorge and Combita-Niño Harold

Department of Computer Science and Electronics, Universidad de la Costa, Barranquilla, Colombia

**Abstract:** In this study, an implementation of several data mining techniques is presented, including decision trees, Support Vector Machines (SVM), Bayesian Networks and K-Nearest Neighbor and their comparison using different evaluation metrics such as True Positive Rate (TpRate), False Positive Rate (FpRate) and Recall, with the dataset “STUDENT ALCOHOL CONSUMPTION”, that provides information of alcohol consumption in teenagers in Portugal. High alcohol consumption rate in teenagers in society, high schoolers and college students, has become a social problem with alarming data showing they start consuming alcohol between 10 and 14 years and this obviously has a huge impact in their behavior, especially with situations such as binge drinking. At the end of the study, the results found show that Support Vector Machines (SVM) have a better accuracy rate than other techniques used and corroborate that the proposed method it is quite efficient and highly precise for detection of students consuming alcohol, improving the results obtained in previous similar studies.

**Keywords:** Alcohol consumption, bayesian networks, data mining, Decision Trees (DT), K-Nearest Neighbors (KNN), Support Vector Machines (SVM)

## INTRODUCTION

Alcohol consumption in teenagers has become a serious health problem in current society, based on data provided by several studies in different countries, in Portugal (Bi *et al.*, 2013), the average age of starting consuming alcohol is around 14 years (73,5%) and 42,6% of them ingest alcohol usually, meanwhile in United States (Zuba *et al.*, 2012), 90% of alcohol consumption under 21 years involves binge drinking and 44% of college students participate in some sort of high risk consuming activities. In Chile (García *et al.*, 2009), high school alcohol consumption is 26,2% in metropolitan area, with a weekly alcohol intake with ages from 15 to 25 years, which clearly demonstrates that this situation is the same all around the world.

In this study, an implementation of several data mining techniques is presented, including decision trees, Support Vector Machines (SVM), Bayesian Networks and K-Nearest Neighbor comparing them using different evaluation metrics such as True Positive Rate (TpRate), False Positive Rate (FpRate) and Recall, with the dataset “STUDENT ALCOHOL CONSUMPTION” (Cortez and Silva, 2008), having information of alcohol consumption in teenagers in Portugal and also in Mendoza-Palechor *et al.* (2016).

Alcohol consumption in teenagers is quite a reality, nevertheless data mining has not been applied to similar consumption studies, due to lack of information in some countries, or priorities in other aspects of public health. Some of the previous studies found are:

In Bi *et al.* (2013), a study using two machine learning methods is presented, to analyze collected data from polls made in college students. Their methodology could identify effectively the daily dynamic alcohol consumption and the risk factors associated to it. For this, they proposed a Support Vector Machine (SVM) as classifier to establish a function for stress, state of mind and consumption expectancy, differentiating nightly intake days from regular days. After that, a mix of clustering analysis and feature classification was made to identify consumption patterns based on daily behavior of average intake and can detect risk factors associated to each pattern.

Zuba *et al.* (2012) demonstrates the use of feature selection with 1-norm support vector machines (SVM) to help classify college students between high risk and low risk alcohol drinkers and the risk factors associated to the first class. This approach could be used to help to detect early signs of addiction and dependence to alcohol in students.

Cortez and Silva (2008) makes an exposition of several data mining techniques, including neural

**Corresponding Author:** De la Hoz-Manotas Alexis, Department of Computer Science and Electronics, Universidad de la Costa, Barranquilla, Colombia

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

networks, decision trees and Naïve Bayes, their implementation in a study case with a dataset of alcohol consumption in teenagers and their possible relationship with personality. In their results, we can highlight that the neural network model is more accurate and these techniques has been widely used to analyze drug dependence and similar.

Pang *et al.* (2015) applies a multimodal analysis to identify alcohol consumption in minors as users of the Instagram social network, based on facial recognition using a tool called Face++ and exploring the tags assigned to each image with the objective of finding consumption patterns in terms of time, frequency and location. In the same way, they measured the penetration of alcohol brands to establish their influence in the consumption behavior of their followers. Their results were satisfactory and compliant with the polls made in the same audience, which can lead to use this approach to other domains of public health.

A study to establish the risks associated to alcohol consumption in teenager students in Chile was developed based on the information provided by the VII National Study of Drugs in School Population, applying the CART technique (Classification and Regression Trees) looking to identify sub-populations with differential risk in dangerous alcohol consumption (Villalón and Cuellar, 2013).

In Crutzen *et al.* (2015) a group of Dutch researchers studied the relationship between parental reports, teenager perception and parenting practices to identify drinkers with alcoholic episodes. They designed a binary classifier using alternating decision trees to establish the effectiveness of the results of exploring nonlinear relationships of data.

Montaño *et al.* (2014) presents an analysis of psychosocial variables about nicotine consumption in teenagers, applying several classification techniques such as RNA Multi-layer perceptron, radial basis functions and probabilistic networks, decision trees, logistic regression model and discriminant analysis. Thus, they discriminated successfully 78.20% of the data, which indicates that this approach can be used to predict and prevent similar addictive behavior.

Based in these results, we can safely say the model proposed would be an excellent tool to identify patterns and predict behaviors in alcohol drinks consumption for teenagers and college students and other health care areas as well.

## LITERATURE REVIEW

**Decision trees:** Decision trees is one of the data mining algorithms most widely used for classification and regression. Each interior node maps to an input variable and is divided in children nodes with the values taken by it. Each leaf or terminal node represents a value of

the output variable, for example, the specific class of a category variable for a classification problem and a certain real value of a continuous variable for regression problems.

During the learning process of a classification tree, samples in each interior node are divided in subsets based in an attribute and this process is repeated in each subset recursively, this is called “recursive partitioning”. Recursion ends when a subset in a node has the same target value, when the division does not improve the prediction and when the division it is not possible due limitations set by the user.

In each step, during the growth of the decision tree, one of the input variables is selected to divide samples. With the basis of the selected variable, the division point is determined using a test of attribute values. Impurity and entropy are the tests most used for classification trees (Kim, 2016).

**Support vector machines:** Support vector machines are decision algorithms that solve regression and classification problems in an efficient way due to their automatic learning. These are based in stochastic learning theory developed by Vapnik and Vapnik (1998), where they propose a mathematic model for regression and classification problems (Vapnik, 1995). Other authors mention that SVM is a margin classifier trained by a group of data formulated as feature vectors. SVM tries to find an optimal level that divides two different classes of feature vectors with a maximum margin (distance between the optimum hyperplane and the nearest vector). To classify an inseparable dataset, a nonlinear SVM is a feature vector projection in a high dimensional space using a kernel function, as a radial basis kernel function (Bakhtiarizadeh *et al.*, 2014).

Support Vector Machines (SVM) construction is based in the idea of transforming or projecting a dataset in any n dimension space to a higher dimension space applying a kernel function-Kernel Trick. From this new space, data will be used as a linear problem, solving the problem without considering the data dimensionality (Gutiérrez and Velandia, 2011).

The success with support vector machines relies in three fundamental advantages: first, their solid mathematical foundation. Second, the concept of structural risk minimization (Kecman, 2001; Cristianini and Shawe-Taylor, 2000), which translates in minimizing the probability of a wrong classification on new examples. This case usually appears when there is a lack of training data. The third advantage is based in the powerful tools and algorithms available to find a solution in a fast and efficient way (Moscovitz and Rengifo, 2010).

**Naive bayes:** The representation of probabilistic information using building networks, dates from the

first quarter of the XX century (Pearl, 2001) and Bayesian Networks are presented as an alternative to classic expert systems oriented to decision support and prediction under uncertainty in probabilistic terms (Cowell *et al.*, 1999). These networks are considered key statistical tools to represent associated uncertainties based in conditional independent relationships established between them.

A Bayesian network by Edwards (1998) and Edwards and Fasolo (2001), it is a structure composed by four levels. The higher level is formed by a set of variables represented as nodes and arrows that are related on influence terms. In the next level are levels or states, also known as space states (Nadkarni and Shenoy, 2001, 2004) that can map each of the variables of the model. In the third level, you can find a set of conditional probability functions, one for each node, that represent the occurrence probability of each state of the variable conditioned to the possible values of the variables that are determined by the variable value. Finally, in the lowest level, you would find a set of algorithms that allow the network to recalculate the probabilities assigned to each of the previous levels when some evidence exists in the model.

It is important to highlight that a Bayesian network is based in two elements, a qualitative and a quantitative dimension (Martínez *et al.*, 2003). The qualitative dimension is composed by graph theory and probability theory (Ríos, 1995).

Graph theory creates graphics models that represent the elements of the problem in a holistic sense, provided by Euler to solve the problem of the Königsberg bridges (Ríos, 1995). A Bayesian network is a graph and it is defined as  $G = (V, E)$ , where  $V$  is a finite set of vertex, nodes and variables and  $E$  is a subset of the set  $V \times V$  of ordered pairs of vertex called links or edges (Harary, 1969; Ronald, 1988). Besides, a Bayesian network is a particular type of graph called an acyclic directed graph (Spirtes *et al.*, 2000).

There are three fundamental elements of the quantitative dimension of a Bayesian network: the concept of probability, the Bayes theorem and the conditional probability functions. The probability can be understood as subjective, as the level of belief over a fact (Dixon and Pastor, 1970), this concept of probability is called Bayesian and comes from the principle of insufficient reason of uncertainty principle (Cowell *et al.*, 1999).

The Bayes theorem is deduced from the axiom that relates the probability of events intersection and conditional probability, which favors that can be worked efficiently with graphic models of probability propagation in terms of conditional dependency and independency (Cowell *et al.*, 1999). What makes a Bayesian network, is their ability to update the probabilities inside an acyclic directed graph based on conditional independency principles when evidence is incorporated to the model.

A Bayesian network needs a set of conditional probability functions, one for each variable or network node and they are used to apply the Bayes rule. Specifically, each network variable is defined by a conditional probability table where the values that variable can have are represented in conjunction with the values of the set of variables is dependent. In this sense, following Cowell *et al.* (1999).

**K-nearest neighbors.** The K-nearest neighbors is an algorithm used for data classification and regression. The algorithm stores all known cases and classifies them or assigns a property to new cases based on similar features (Sánchez *et al.*, 2016). This method must be one of the first options for a classification study, when there is little or none previous knowledge about the data distribution. The K-nearest neighbors method was developed given the need of making discriminant analysis when reliable parametric estimation of the probability density are unknown or difficult to establish (Fix and Hodges Jr, 1951).

This method assumes the data is found in a features space. Data can be scalar or multi-dimension vectors. All training data is composed of a set of vectors and tags associated to each of them. The method is retrograded and supervised (its training phase is made in different time to the testing phase) and its main argument is the distance between instances (Rodríguez *et al.*, 2013).

The KNN method compares the new instance to classify with the k-nearest known neighbors and depending to the similarities in the attributes of the new case, it will be in the class closest to the values of its own attributes and compliant with the concept of consistency heuristic.

The main difficulty of this method is to establish the value of k, since if it is too large, there is a risk of making a classification based on most the data and not to the similar ones and if the value is too low, there can be a lack of precision in the classification due to not having enough selected data as comparison instances (Mucherino *et al.*, 2009).

To mitigate this problem, several approaches to the method were proposed: to set the value of k, for example, 1-nn, where the first k-nearest neighbor found is used as comparison instance and k can be found taking a comparison radius, or using Voronoi diagrams (Rodríguez *et al.*, 2013).

One of the most relevant advantages of the k-NN method is that it can change radically its classification results, without changing its structure, only modifying the metric used to find the distance. The metric must be selected according to the problem that needs to be solved. The huge advantage of being able to change metrics is that can achieve entirely different results and the general algorithm of the method does not change, only the procedure to measure distances.

**Clustering:** Clustering is one of the most popular unsupervised learning methods, consists in building a set of physical or abstract objects in classes with similar objects (Han and Kamber, 2001). A good clustering will produce groups of high quality with high similarities intra-class and low similarities between classes. It can be used as an independent tool to better understand the data distribution or to serve as a pre-processing step for other tasks (Hastie *et al.*, 2001).

Clustering can also be used as a stand-alone process and supervised algorithm technique based in prototypes (Jain *et al.*, 1999) and several applications with non-vector data has been built. The implementation of clustering algorithms for unsupervised analysis of the data has become a useful tool to explore and to solve different problems in data mining. The clustering methods (Xu and Wunsch, 2005; Zhang and Zhou, 2009) has been used to

Table 1: Student alcohol consumption dataset attributes

No	Type	Variables	Values
1	Category	School-student's school	GP-Gabriel Pereira MS-Mousinho da Silveira
2	Category	Sex-student's sex	F-Female M-Male
3	Continuous	Age-student's age	>15 and <22
4	Category	Address-student's home address type	U-Urban R-Rural
5	Category	Pstatus-parent's cohabitation status	GT3-Mayor que 3 T-Living together A-Apart
6	Continuous	Medu-mother's education	0-None 1-Primary Education (4 <sup>th</sup> grade) 2-5 <sup>th</sup> to 9 <sup>th</sup> grade 3-Secondary Education 4-Higher Education
7	Category	Mjob-mother's job	Teacher Health Services At Home Other
8	Category	Fjob-father's job	Teacher Health Services At Home Other Reputation Course Other Father Other
9	Continuous	Traveltime-home to school travel time	1-less than 15 minutes 2-15 to 30 minutes 3 - 30 minutes to 1 hour 4-More than 1 hour
10	Continuous	Studytime-weekly study time	1-less than 2 hours 2-2 to 5 hours 3-5 to 10 hours 4-More than 10 hours
11	Continuous	Failures-number of past class failures	n if 1 <= n < 3, else 4
12	Category	Schoolsup-extra educational support	Yes No
13	Category	Famsup-family educational support	Yes No
14	Category	Paid-extra paid classes within the course subject	Yes No
15	Category	Activities-extra-curricular activities	Yes No
16	Category	Nursery-attended nursery school	Yes No
17	Category	Higher-wants to take higher education	Yes No
18	Category	Internet-Internet access at home	Yes No
19	Category	Romantic-with a romantic relationship	Yes No
20	Continuous	Freetime-free time after school	1 to 5
21	Continuous	Goout-going out with friends	1 to 5
22	Continuous	Health-current health status	1 to 5
23	Continuous	Absences-number of school absences	0 to 93
24	Continuous	G3-final grade	0 to 20
25	Continuous	Consal	Yes/No

provide solution to problems of many different contexts and disciplines.

### MATERIALS AND METHODS

**Dataset preparation:** For this study, the dataset “STUDENT ALCOHOL CONSUMPTION” was used, located in the Machine Learning Repository UCI (Cortez and Silva, 2008), with multi-variable data, which means that have continuous and categorized, counting with 32 attributes and 1044 records.

First, the preparation of the dataset was made, eliminating non-relevant variables and others with correlation higher than 0.7, due its effect in the processes of segmentation, classification and prediction.

Finally, 24 of the 32 attributes of dataset were chosen and an extra attribute was added, called “consal”, that indicates if a person consumes alcohol. To do this, the equation n° 1 proposed in Pagnotta and Amran (2016) was used, calculated with the original attributes found on the dataset:

$$consal = \frac{walc*2+dalc*5}{7} \quad (1)$$

In Table 1, the final list of the attributes used is shown.

**Experimentation:** The present study is based on the “STUDENT ALCOHOL CONSUMPTION” dataset, available from the Machine Learning Repository UCI. The purpose of this research is to compare several data mining techniques, which are commonly used in segmentation, classification and prediction tasks. The methodology used is described as follows.

After downloading and preparation of the dataset to be used in training and evaluating the proposed method, an analysis of the data mining tools available was made, considering factors such as the methods found in the Weka tool and their license of use. For the preparation of data, it was necessary to analyze the correlation rate between the variables, especially for those that generate noise and those with high correlation rate, so they would not be included in the processes of training and classification of the data. Also, a balancing process was needed, since the values of alcohol consumption in the dataset are unbalanced and this affects the model, specifically the learning process, having only a single class or situation. Next, you can find a screenshot of the initial state of the data, as shown in Fig. 1 and the data after the balancing process was performed, in Fig. 2.

After selection of the tool was made and the dataset was prepared, the selection of the clustering method to be used was developed, considering that users must be grouped in two categories: drinkers and non-drinkers. For this process, the Simple K-means method was used and the results obtained are shown in Table 2.

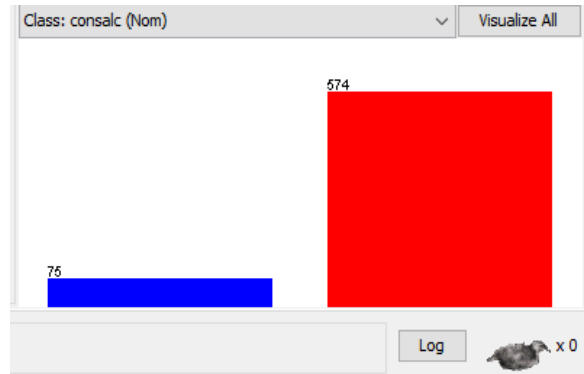


Fig. 1: Dataset before the balancing process

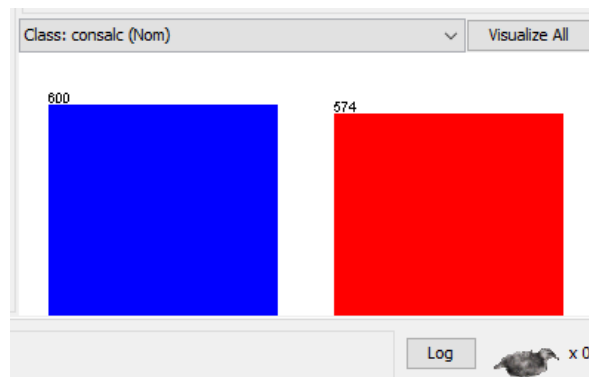


Fig. 2: Dataset after balancing process

Table 2: Data groups after clustering method

Assigned cluster	Numbers of items
Cluster 0	674 (57%)
Cluster 1	500 (43%)

In Table 2 you can see the number of users in each cluster, where cluster 0 are the people that are consuming alcohol and cluster 1, the people that does not consume alcohol.

After clustering of the data, the selection of the classification methods to be compared was made and those were: Decision Trees, Support Vector Machines, Naïve Bayes and Lazy IBK, with the metrics to evaluated: True Positive Rate (TP Rate), False Positive Rate (FP Rate), precision and recall, calculated with the equations in the Table 3.

The data distribution for the training and testing process was performed using cross validation, the tool selects a percentage of the data for training and another to testing with the proposed methods. After that, every data mining method chosen is tested and subject to comparison with the metrics previously selected.

Finally, after the best method was identified, a file test was implemented to check the results obtained with the best classifier selected. The document for testing created is shown in Fig. 3.

The methodology for the study is found in Fig. 4.

Table 3: Evaluation metrics for classification methods

Metrics	Formula
Precision	$\frac{True\ Positive}{True\ Positive + False\ Positive}$
Recall	$\frac{True\ Positive}{True\ Positive + False\ Negative}$
TpRate	$\frac{True\ Positive + False\ Negative}{True\ Positive + False\ Negative + False\ Positive}$
FpRate	$\frac{False\ Positive + True\ Negative}{True\ Positive + False\ Negative + False\ Positive}$

```

@attribute failures numeric
@attribute schoolsup numeric
@attribute famsup numeric
@attribute activities numeric
@attribute paid numeric
@attribute internet numeric
@attribute nursery numeric
@attribute higher numeric
@attribute romantic numeric
@attribute freetime numeric
@attribute goout numeric
@attribute health numeric
@attribute absences numeric
@attribute G1 numeric
@attribute G2 numeric
@attribute G3 numeric
@attribute consalc {si,no}

@data
1,14,0,0,1,4,3,2,2,0,0,1,0,0,0,0,1,1,0,3,4,3,4,0,11,11,?
1,17,0,0,0,1,3,1,2,4,0,0,1,0,0,1,0,1,0,3,3,3,2,9,11,11,?
0,17,0,0,0,1,3,1,2,4,0,1,0,0,0,1,1,1,0,3,2,3,6,12,13,12,?
0,15,0,0,0,4,1,1,3,2,0,0,1,1,0,1,1,1,1,2,2,5,0,14,14,14,?
0,16,0,0,0,3,4,1,2,4,0,0,1,0,0,0,1,1,0,3,2,5,0,11,13,13,?

```

Fig. 3: Test file

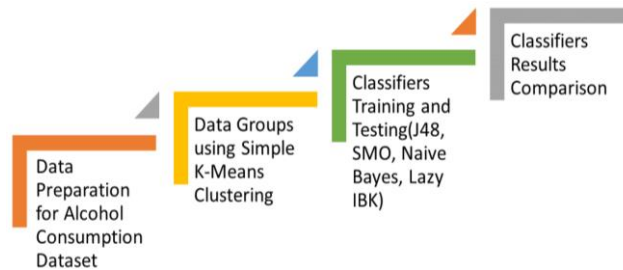


Fig. 4: Complete study methodology

**RESULTS AND DISCUSSION**

In the present study, the starting point was data preparation, process that must remove variables generating noise and irrelevant for the training and classification process of the model. After data preparation was complete, the clustering technique selection was needed to make the right segmentation of the data and this has a great impact in the classifiers subject to analysis. Next, you can find the results obtained with the processes mentioned in the methodology.

In Fig. 5, the results obtained by the decision trees algorithm J48 with the following percentages for each metric: 96.60% (TpRate), 4.70% (FpRate), 95.60% (Precision), 95.60 (Recall).

In Fig. 6, the results obtained by the support vector machines algorithm SMO with the following percentages for each metric: 98% (TpRate), 1.90% (FpRate), 98% (Precision), 98% (Recall).

In Fig. 7, the results obtained by the Naïve Bayes algorithm with the following percentages for each metric: 86.20% (TpRate), 13.20% (FpRate), 86.60% (Precision), 86.20% (Recall).

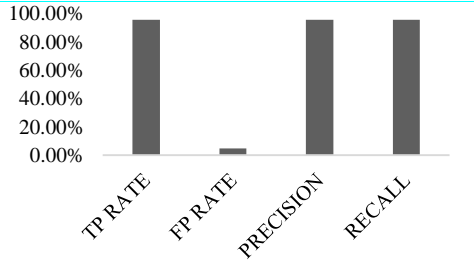


Fig. 5: J48 algorithm results

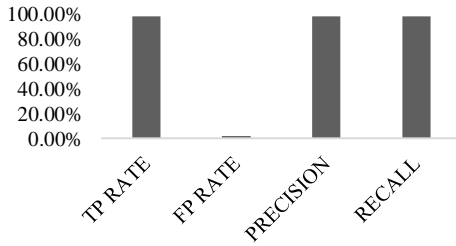


Fig. 6: SMO algorithm results

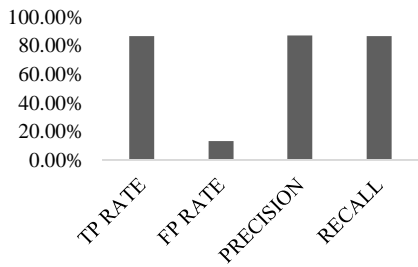


Fig. 7: Naïve bayes algorithm results

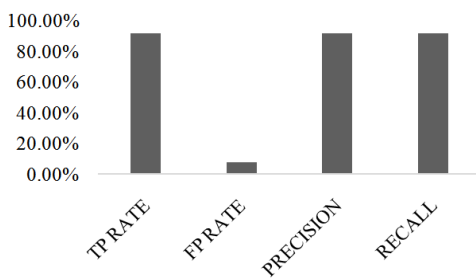


Fig. 8: Lazy IBK algorithm results

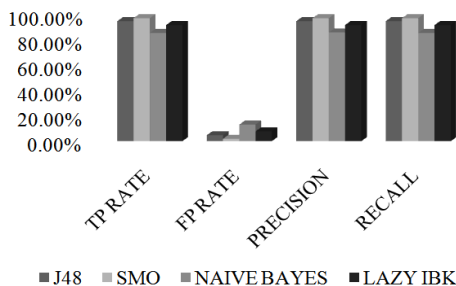


Fig. 9: Classifier algorithm results

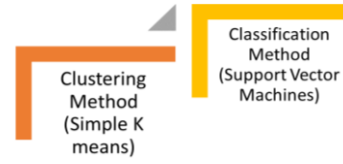


Fig. 10: Proposed data mining method

Table 4: Classifier algorithms results

METHOD	TP RATE (%)	FP RATE (%)	PRECISION (%)	RECALL (%)
J48	95.60	4.70	95.60	95.60
SMO	98.00	1.90	98.00	98.00
NAIVE	86.20	13.20	86.60	86.20
BAYES				
LAZY IBK	92.50	7.80	92.50	92.50

In Fig. 8, the results obtained by the Lazy IBK algorithm with the following percentages for each metric: 92.50% (TpRate), 7.80% (FpRate), 92.50% (Precision), 92.50% (Recall).

All the results obtained by classifiers are resumed in Table 4.

In Fig. 9, the comparison of all techniques subject to study are presented and with these results, the support vector machines methods obtain the best percentages in all the evaluated metrics, which clearly indicates that it can be used for classification processes in alcohol consumption evaluation per the data studied. The proposed method has the structure shown in Fig. 10.

## CONCLUSION

The present study had the objective of proposing a method to allow classification processes to identify alcohol consumption in students, using as a source the STUDENT ALCOHOL CONSUMPTION dataset, for this it was necessary to prepare the data, using the simple k-means method for segmentation and then applying data mining techniques such as: decision trees, support vector machines, bayesian networks and k-nearest neighbors. Using these techniques, the method with the best results was support vector machines with a TpRate of 98%, FpRate of 1.90%, Precision of 98% and Recall of 98%. The previous results indicate that the proposed method obtained a high percentage in all metrics evaluated and its considered an efficient and accurate method for the consumption of alcohol in students, improving the results obtained by other studies (Pagnotta and Amran, 2016), where the authors obtained only 92% using their approach.

## REFERENCES

Bakhtiarizadeh, M.R., M. Moradi-Shahrbabak, M. Ebrahimi and E. Ebrahimie, 2014. Neural network and SVM classifiers accurately predict lipid binding proteins, irrespective of sequence homology. *J. Theor. Biol.*, 356: 213-222.

- Bi, J., J. Sun, Y. Wu, H. Tennen and S. Armeli, 2013. A machine learning approach to college drinking prediction and risk factor identification. *ACM T. Intell. Syst. Technol.*, 4(4).
- Cortez, P. and A. Silva, 2008. Using data mining to predict secondary school student performance. In: Brito, A. and J. Teixeira (Eds.), *Proceeding of 5th Future Business Technology Conference (FUBUTEC, 2008)*. Porto, Portugal, April, pp: 5-12.
- Cowell, R.G., P. Dawid, S.L. Lauritzen and D.J. Spiegelhalter, 1999. *Probabilistic Networks and Expert Systems: Exact Computational Methods for Bayesian Networks*. Springer-Verlag, New York, pp: 324.
- Cristianini, N. and J. Shawe-Taylor, 2000. *An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods*. Cambridge University Press, New York.
- Crutzen, R., P.J. Giabbanelli, A. Jander, L. Mercken and H. de Vries, 2015. Identifying binge drinkers based on parenting dimensions and alcohol-specific parenting practices: Building classifiers on adolescent-parent paired data. *BMC Public Health*, 15(1): 747.
- Dixon, J.R. and V.M. Pastor, 1970. *Introducción a la Probabilidad: Texto Programado*. 1st Edn., Editorial Limusa, Mexico, D.F.
- Edwards, W., 1998. Hailfinder. Tools for and experiences with Bayesian normative modeling. *Am. Psychol.*, 53(4): 416-428.
- Edwards, W. and B. Fasolo, 2001. Decision technology. *Annu. Rev. Psychol.*, 52(1): 581-606.
- Fix, E. and J.L. Hodges Jr, 1951. Discriminatory analysis-nonparametric discrimination: Consistency properties. *Air Technical Index*, California University, Berkeley.
- García, E.G., R.J. López, J.J.M. Moreno, A.S. Abad, B.C. Blasco and A.P. Pol, 2009. La metodología del Data Mining. Una aplicación al consumo de alcohol en adolescentes. *Adicciones*, 21(1): 65-80.
- Gutiérrez, J.F.M. and L.F.M. Velandia, 2011. Pronóstico de Incumplimientos de Pago Mediante Máquinas de Vectores de Soporte: Una Aproximación Inicial a La Gestión Del Riesgo de crédito. Banco de la República, Bogotá, No. 677.
- Han, J. and M. Kamber, 2001. *Data Mining: Concepts and Techniques*. 2nd Edn., Morgan Kaufmann Publishers, San Francisco.
- Harary, F., 1969. *Graph Theory*. Addison-Wesley Publishing Company Inc., Reading, MA.
- Hastie, T., R. Tibshirani and J.H. Friedman, 2001. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, New York.
- Jain, A.K., M.N. Murty and P.J. Flynn, 1999. Data clustering: A review. *ACM Comput. Surv.*, 31(3): 264-323.
- Kecman, V., 2001. *Learning and Soft Computing Support Vector Machines, Neural Networks and Fuzzy Logic Models*. MIT Press, Cambridge, Mass.
- Kim, K., 2016. A hybrid classification algorithm by subspace partitioning through semi-supervised decision tree. *Pattern Recogn.*, 60: 157-163.
- Martínez, F., M.C. Díaz, M.T. Martín, V.M. Rivas and L.A. Ureña, 2003. Aplicación de redes neuronales y redes bayesianas en la detección de multipalabras para tareas IR. *Proceeding of the Artculo Presentado en las II Jornadas de Tratamiento y Recuperación de la Informacin*, Madrid.
- Mendoza-Palechor, F.E., P.P. Ariza-Colpas, J.A. Sepulveda-Ojeda, A., De-La-Hoz-Manotas and M. Piñeres Melo, 2016. Fertility Analysis Method Based on Supervised and Unsupervised Data Mining Techniques. *Int. J. Appl. Eng. Res.*, 11(21): 10374-10379.
- Montaño, J.J., E. Gervilla, B. Cajal and A. Palmer, 2014. Data mining classification techniques: An application to tobacco consumption in teenagers. *An. Psicol.*, 30(2): 633-641.
- Moscovitz, L.J. and P.R. Rengifo, 2010. Al interior de una máquina de soporte vectorial. *Rev. Cienc.*, 14: 73-85.
- Mucherino, A., P.J. Papajorgji and P.M. Pardalos, 2009. *Data Mining in Agriculture*. Vol. 34, Springer-Verlag, New York.
- Nadkarni, S. and P.P. Shenoy, 2001. A Bayesian network approach to making inferences in causal maps. *Eur. J. Oper. Res.*, 128(3): 479-498.
- Nadkarni, S. and P.P. Shenoy, 2004. A causal mapping approach to constructing Bayesian networks. *Decis. Support Syst.*, 38(2): 259-281.
- Pagnotta, F. and H.M. Amran, 2016. Using data mining to predict secondary school student alcohol consumption. Department of Computer Science, University of Camerino.
- Pang, R., A. Baretto, H. Kautz and J. Luo, 2015. Monitoring adolescent alcohol use via multimodal analysis in social multimedia. *Proceeding of the IEEE International Conference on Big Data (Big Data)*, pp: 1509-1518.
- Pearl, J., 2001. *Bayesian networks, causal inference and knowledge discovery*. Technical Report, Computer Science Department, Cognitive Systems Laboratory, University of California, Los Angeles.
- Ríos, S., 1995. *Modelización*. Alianza, Madrid.
- Rodríguez, J.E.R., E.A.R. Blanco and R.O.F. Camacho, 2013. Clasificación de datos usando el método k-*nn*. *Vínculos*, 4(1): 4-18.
- Ronald, G., 1988. *Graph Theory*. Benjamin/Cummings Publishing Co., Menlo Park, CA.
- Sánchez, A.S., F.J. Iglesias-Rodríguez, P.R. Fernández and F.J. de Cos Juez, 2016. Applying the K-nearest neighbor technique to the classification of workers according to their risk of suffering musculoskeletal disorders. *Int. J. Ind. Ergonom.*, 52: 92-99.



- Spirtes, P., C. Glymour and R. Scheines, 2000. Causation, Prediction and Search. 2nd Edn., MIT Press, Cambridge.
- Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer-Verlag, New York.
- Vapnik, V.N. and V. Vapnik, 1998. Statistical Learning Theory. Vol. 1. Wiley, New York.
- Villalón, M. and C. Cuellar, 2013. Adolescentes y consumo nocivo de alcohol. Chile 2009: Mirando a las políticas públicas. *Rev. Méd. Chile*, 141(5): 644-651.
- Xu, R. and D. Wunsch, 2005. Survey of clustering algorithms. *IEEE T. Neural Networ.*, 16(3): 645-678.
- Zhang, M.L. and Z.H. Zhou, 2009. Multi-instance clustering with applications to multi-instance prediction. *Appl. Intell.*, 31(1): 47-68.
- Zuba, M., J. Gilbert, Y. Wu, J. Bi, H. Tennen and S. Armeli, 2012. 1-norm support vector machine for college drinking risk factor identification. *Proceeding of the 2nd ACM SIGHIT International Health Informatics Symposium*, pp: 651-660.