https://repository.uaeh.edu.mx/revistas/index.php/huejutla/issue/archive

Ciencia Huasteca Boletín Científico de la Escuela Superior de Huejutla

Publicación semestral, Vol. 8, No. 15 (2020) 1-8

ISSN: 2007-493X

Application Method of Data Mining Using the K means Algorithm for the Determination of Stress Level in High School Students Using the Beck Depression Inventory

Aplicación de Minería de Datos Usando el Algoritmo K-Means en Alumnos de Bachillerato para determiner los Niveles de Depresión Mediante el Inventario de Depresión de Beck

Felipe J. Núñez-Cardenas [a], Eduardo Aguirre-Hernandez [b], Alain E. Guerrero-Zenil [b], Ana M. Felipe-Redondo [c]

Abstract:

In the present work a description of the application of data mining techniques on depression is made to high school students, in which the inventory of the Beck depression is applied as an evaluation tool, also by the algorithm K -Means the grouping of the data and obtain results with 6 groups, because the assessment tool are 6 possible diagnoses that are given according to the score obtained.

Keywords:

Data Mining, K-Means, Beck Depression Inventory

Resumen:

En el presente trabajo se realizará una descripción aplicando técnicas de minería de datos acerca de la depresión a los alumnos de educación media superior, en el cual se aplica como herramienta de evaluación el inventario de depresión Beck, así mismo mediante el algoritmo K-Means se realizará la agrupación de los datos y como resultados se busca obtener seis clusters, debido a que en la herramienta de evaluación son 6 los posibles diagnósticos que se dan de acuerdo con la puntuación obtenida.

Palabras Clave:

Mineria de Datos, K means, inventario de depression Beck

1. Introducción

Currently, students who are about to complete their high school education suffer from depression because of the harassment they suffer within the classroom, family, economic problems, etc., In addition to enter a higher education level, may present distrust or fear for to experience the same thing again, so students should be evaluated to help them.

In this document we will talk about the use of data mining tools applying the K-Means algorithm for the grouping of data, this to determine the level of depression in the students who graduated from high school in the High School of Huejutla belonging to the Autonomous University of the State of Hidalgo, using the Beck depression test as an evaluation tool.

Data mining is a set of data analysis techniques that allow you to extract patterns, trends and regularities to better describe and understand the data.

[a] Autor de Correspondencia, Universidad Autónoma del Estado de Hidalgo, Escuela Superior de Huejutla, https://orcid.org/0000-0002-2462-3654, email: felipe_nunez@uaeh.edu.mx

[b] Universidad Autónoma del Estado de Hidalgo, Escuela Superior de Huejutla, alumno de Ciencias Computacionales, emails: eduar030696@gmail.com y alaguerrero96@gmail.com

[c] Universidad Tecnológica de la Huasteca Hidalguense, Profesor de Tiempo Completo, email: ana.felipe@uthh.edu.mx

The Beck Depression Inventory is a paper and pencil self-report composed of 21 Likert-type items. The inventory initially proposed by Beck and his later versions have been the most used instruments to detect and evaluate the severity of depression, his items are not derived from any concrete theory about the construct means, but they describe the most frequent clinical symptoms of depression. the patients with depression. The test is preferably designed for clinical use, as a means to assess the severity of depression in adult patients and adolescents with a psychiatric diagnosis and with 13 years of age or older.

1.1 State of Art

In the year 2005 Sergio Valero Orea in his project that has by title Mining of data: prediction of the scholastic desertion by means of the algorithm of decision trees and the algorithm of the k nearer neighbors, in this project Orea looks through the mining of data to predict school desertion at the Izúcar de Matamoros Technology University, based on the analysis of the socioeconomic study data of the EXANI II, prepared by Ceneval, which has been applied since 2003. [8]

In the year 2004 in Peru, Cecilia Diaz in her work, which has the name Conditional factors of depression in metallurgical workers, in which she intends to look for predetermining factors in the development of the depression clinical picture, where she performs the study of 153 railway workers and operators of crane-bridge in a mining-metallurgical copper company, evaluated with the Hamilton scale, where I detect 25 probable cases of depression and who are examined by the psychiatrist. [3]

During the year 2007, Alcover R. in his project entitled, Analysis of academic performance in computer studies at the Polytechnic University of Valencia applying data mining techniques, makes an analysis of the academic performance of new students in the degree of Technical Engineering in Computer Systems of the University of Valencia, using techniques of data mining, which aims to determine the level of conditioning that exists in the performance of new students. [2]

In 2011, Claudia Xiomara Santos León, in her work named Emotional States and Physical Activity Levels at the General José Maria Córdova Military Cadet School, seeks to analyze the emotional states (anxiety / depression) and the level of physical activity, and its probable relations, of the students of the military school who present normal body weight and who are overweight, a quantitative, descriptive and analytical component study was carried out, with which 76 male and female students were evaluated. For the analysis of depression the inventory for the BECK depression was used, the questionnaire was applied collectively. To evaluate the anxiety, the state / trait anxiety scale (STAI) of Spielberger was used. [4]

Gerardo Ramiro Luna Guevara in 2016 in Mexico in his project called Department of Psychiatry and Mental Health. Faculty of Medicine, in which we talk about determining a profile in patients with some symptoms and then its application with a classification method, we used the EEG signal of patients with depression and define a statistical profile, we present the calculation of a profile from a series of high-dimensional signals from 9 healthy subjects and 14 subjects with depression. [5]

1.2 Data Mining

Data mining techniques have emerged from systems of inductive learning in computers, the main difference between them being the data on which the search for new knowledge is carried out.

Data Mining is a process that, through the discovery and quantification of predictive relationships in the data, allows to transform the available information into useful knowledge, it is one of the main ways of exploiting Data Warehouse.

Data Mining emerged as an integration of multiple technologies such as statistics, decision support, machine learning, database management and storage, and parallel processing. In order to carry out these processes, techniques from a variety of areas are applied, such as genetic algorithms, neural networks, decision trees, etc.

Data mining discovers relationships, trends, deviations, atypical behaviors, hidden patterns and trajectories, with the purpose of supporting decision-making processes with greater knowledge. Data mining can be located at the highest level of the evolution of the technological processes of data analysis.

Data mining has arisen from the potential of analyzing large volumes of information, in order to obtain summaries and knowledge that supports decision-making and that can build an experience based on the millions of detailed transactions that a corporation registers in its information systems. [7]

1.3 K-Means Algorithm

K-means in a simple unsupervised learning algorithm that it uses to solve grouping problems. It follows a simple procedure of classifying a set of data given in a series of groups, defined by the letter "k", which is set before proceeding. Clusters are positioned as points and all observations or data points are associated with the nearest cluster, are calculated. They are adjusted and then the process starts again using the new settings until a desired result is reached.

The K-Means algorithm is a method used for the analysis of clusters, especially in data mining and statistics. A goal is to divide a set of observations into a group series (k). It can be considered a method to discover to which group a certain object really belongs.

The algorithm:

1) A. K points are placed in the data space of the object that represents the initial group of centroids.
2) Each object or data point is assigned to the nearest k.
3) After assigning all the objects, the positions of the centroid k are recalculated.
4) Steps 2 and 3 are repeated until the positions of the centroids no longer move. [9]

1.4 Beck Depression Inventory

The Beck Depression Inventory, created by the psychiatrist, researcher and founder of Cognitive Therapy, Aaron T. Beck, is a self-administered questionnaire consisting of 21 multiple-choice questions. It is one of the most commonly used instruments to measure the severity of a depression. The most current versions of this questionnaire can be used in people over 13 years of age. [1]

It is based on the patient's descriptions of different items: mood, pessimism, feeling of failure, dissatisfaction, guilt, irritability, suicidal ideas, crying, social isolation, indecision, changes in physical appearance, difficulty in work, insomnia, fatigability, loss of appetite, weight loss, somatic concern and loss of libido.
Each item is valued from 0 to 3 and the diagnosis is based on the sum of the value of each item resulting in a total value of the sum and is diagnosed. [6]

PUNCTUATION	LEVEL OF DEPRESSION
1-10	States ups and downs are considered normal
11-16	Slight disturbance of the mood
17-20	Intermittent states of depression
21-30	Moderate depression
31-40	Severe depression
+40	Extreme depression

Table 1: Valoration

2. Methodology

The CRISP-DM methodology, which is the name of a standard inter-industrial process for data mining, is a proven method to guide your data mining work.

- The methodology includes descriptions of the normal phases of a project, the tasks required in each phase and an explanation of the relationships between the tasks.
- As a process model, CRISP-DM offers a summary of the life cycle of data mining.

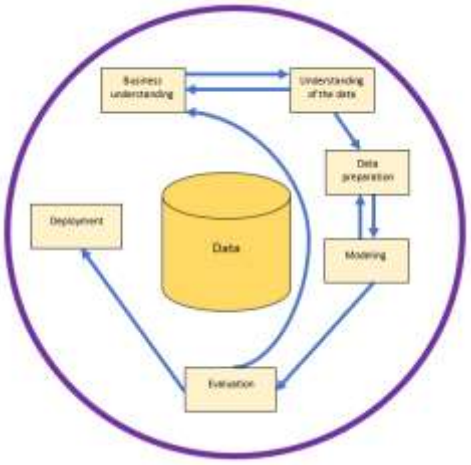


Figure 1: Methodology CRISP-DM

- Business understanding (Objectives and requirements from a non-technical perspective)
- Understanding of the data (Becoming familiar with the data keeping in mind the business objectives)
- Preparing the data (Get the view minable or dataset)
- Modeling (Apply data mining techniques to dataset)
- Evaluation (From the previous phase models to determine if they are useful to the needs of the business)
- Deployment (Exploit the usefulness of the models, integrating them in the decision-making tasks of the organization) [10]

3. Development

During this chapter the development of the project will be carried out applying the CRISP-DM methodology in each of its stages.

3.1 Business Understanding

Next, we will follow each of the tasks that comprise this first phase in the process of data mining, whose purpose is to determine the objectives and requirements of the project from a business perspective, to later become targets from the technical point of view and in a project plan.

3.1.1 Determine The Objetives Of The Business

The objective of the data mining that will be applied in this project is to make a description based on the data collected from the Beck inventory applied to high school students of the sixth semester of the Autonomous University of the State of Hidalgo. Evaluate the level of depression.

In reference to the situation of the business in the organization (Universidad Autónoma del Estado de Hidalgo) at the beginning of this project it can be said that the results obtained from the test applied to the students are available.

The objectives of the business, as already mentioned, is the description of the data resulting from the application of the beck inventory, in such a way that a detailed description of the results can be made by means of the k groups obtained by applying the k-means algorithm and what are the predominant similarities in the groups.

This description can be very useful for teachers and parents to detect depression problems in students and thus try to find out the factors that are influencing the student's behavior.

3.1.2 Business Success Criterio

From the business point of view, it is established as a success criterion the possibility of visualizing in what percentage the depression index is established and what are the relationships between the affected groups.

3.1.3 Determine The Objetives Of Data Mining

The objectives in terms of data mining are:

- Group the population according to similarity or relationship that exists between the data that make up the group. Visualize in what percentage the data was grouped.

- Identify those data that result with a high level of association or percentage.

3.1.4 Prepare the work plan

The project will be divided into the following stages to facilitate its organization and estimate the time it takes to complete it:

- **Stage 1**: Selection of the instrument for the evaluation of depression.

Estimated time: 1 Week

- **Stage 2**: Execution of the evaluation instrument.

Estimated time: 1 Week

- **Stage 3**: Preparation of the data (selection, cleaning) to facilitate the mining of data about them.

Estimated time: 1 Week

- **Stage 4**: Choice of the techniques of modeling and executing them on the data.

Estimated time: 1 Week

- **Stage 5**: Analysis of the results obtained in the previous stage.

Estimated time: 1 Week

- **Stage 6**: Production of reports with the results obtained according to the business objectives and established success criteria.

Estimated time: 1 Week

- **Stage 7**: Presentation of the results obtained.

Estimated time: 1 Week

3.2 Comprehension of data

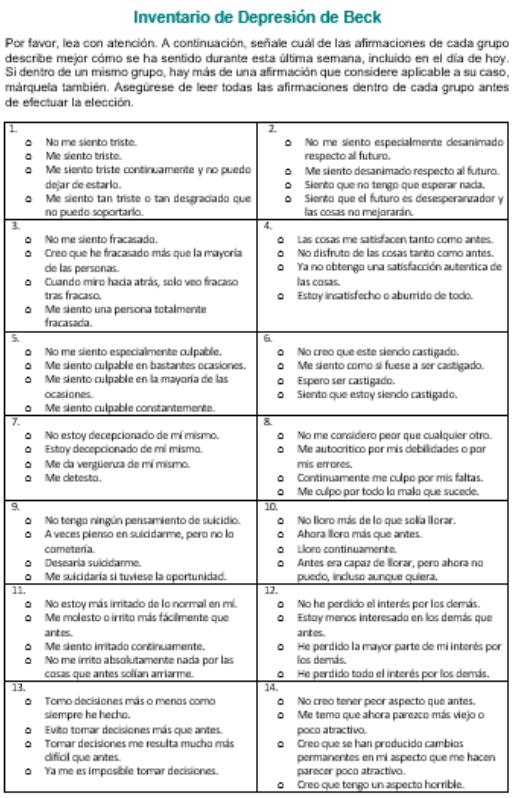
In this second phase of the CRISP-DM methodology, the initial data collection is performed in order to establish a first contact with the problem, become familiar with the data and find out its quality, as well as identify the most obvious relationships to formulate the first hypotheses.

3.2.1 Collect the data

The data used in this project are data referring to students of upper secondary education specifically to 116 students who study at the Huejutla Higher School, which include only one assessment based on Beck's depression inventory, which consists of 21 items and each Item can take 4 values from 0 to 3.

3.2.2 Description of data

The data was collected from the physical inventory of the Beck depression inventory, so that student was evaluated per student and a diagnosis was made, followed by the Beck depression inventory, which is the instrument applied to the students of upper secondary education.



Inventario de Depresión de Beck

Por favor, lea con atención. A continuación, señale cuál de las afirmaciones de cada grupo describe mejor cómo se ha sentido durante esta última semana, incluido en el día de hoy. Si dentro de un mismo grupo, hay más de una afirmación que considere aplicable a su caso, márquela también. Asegúrese de leer todas las afirmaciones dentro de cada grupo antes de efectuar la elección.

1. o No me siento triste. o Me siento triste. o Me siento triste continuamente y no puedo dejar de estarlo. o Me siento tan triste o tan desgraciado que no puedo soportarlo.	2. o No me siento especialmente desanimado respecto al futuro. o Me siento desanimado respecto al futuro. o Siento que no tengo que esperar nada. o Siento que el futuro es desesperanzador y las cosas no mejorarán.
3. o No me siento fracasado. o Creo que he fracasado más que la mayoría de las personas. o Cuando miro hacia atrás, solo veo fracaso tras fracaso. o Me siento una persona totalmente fracasada.	4. o Las cosas me satisfacen tanto como antes. o No disfruto de las cosas tanto como antes. o Ya no obtengo una satisfacción auténtica de las cosas. o Estoy insatisfecho o aburrido de todo.
5. o No me siento especialmente culpable. o Me siento culpable en bastantes ocasiones. o Me siento culpable en la mayoría de las ocasiones. o Me siento culpable constantemente.	6. o No creo que este siendo castigado. o Me siento como si fuese a ser castigado. o Espero ser castigado. o Siento que estoy siendo castigado.
7. o No estoy decepcionado de mí mismo. o Estoy decepcionado de mí mismo. o Me da vergüenza de mí mismo. o Me detesto.	8. o No me considero peor que cualquier otro. o Me autocritico por mis debilidades o por mis errores. o Continuamente me culpo por mis faltas. o Me culpo por todo lo malo que sucede.
9. o No tengo ningún pensamiento de suicidio. o A veces pienso en suicidarme, pero no lo cometería. o Desearía suicidarme. o Me suicidaría si tuviese la oportunidad.	10. o No lloro más de lo que solía llorar. o Ahora lloro más que antes. o Lloro continuamente. o Antes era capaz de llorar, pero ahora no puedo, incluso aunque quiera.
11. o No estoy más irritado de lo normal en mí. o Me molesto o irrito más fácilmente que antes. o Me siento irritado continuamente. o No me irrito absolutamente nada por las cosas que antes solían arriarme.	12. o No he perdido el interés por los demás. o Estoy menos interesado en los demás que antes. o He perdido la mayor parte de mi interés por los demás. o He perdido todo el interés por los demás.
13. o Tomo decisiones más o menos como siempre he hecho. o Evito tomar decisiones más que antes. o Tomar decisiones me resulta mucho más difícil que antes. o Ya me es imposible tomar decisiones.	14. o No creo tener peor aspecto que antes. o Me temo que ahora parezco más viejo o poco atractivo. o Creo que se han producido cambios permanentes en mi aspecto que me hacen parecer poco atractivo. o Creo que tengo un aspecto horrible.

Figure 2: Beck Depression Inventory

15. o Trabajo igual que antes. o Me cuesta un esfuerzo extra comenzar a hacer algo. o Tengo que obligarme mucho para hacer algo. o No puedo hacer nada en absoluto.	16. o Duermo tan bien como siempre. o No duermo tan bien como antes. o Me despierto una o dos horas de lo habitual y me resulta difícil volver a dormir. o Me despierto varias horas antes de lo habitual y no puedo volverme a dormir.
17. o No me siento más cansado de lo normal. o Me canso más fácilmente que antes. o Me canso en cuanto hago cualquier cosa. o Estoy demasiado cansado para hacer nada.	18. o Mi apetito no ha disminuido. o No tengo tan buen apetito como antes. o Ahora tengo mucho menos apetito o He perdido completamente el apetito.
19. o Últimamente he perdido poco peso o no he perdido nada. o He perdido más de 2 kilos y medio. o He perdido más de 4 kilos. o He perdido más de 7 kilos.	20. o No estoy preocupado por mi salud más de lo normal. o Estoy preocupado por problemas físicos como dolores, molestias, malestar de estómago o estreñimiento. o Estoy preocupado por mis problemas físicos y me resulta difícil pensar algo más. o Estoy tan preocupado por mis problemas físicos que soy incapaz de pensar en cualquier cosa.
21. o No he observado ningún cambio reciente en mi interés. o Estoy menos interesado por el sexo que antes. o Estoy mucho menos interesado por el sexo. o He perdido totalmente mi interés por el sexo.	**Gracias por contestar el test.**

Figure 3: Beck Depression Inventory

3.3 Data preparation

Once the data have been described, we proceed to prepare the initial data to be used in the process of Data Mining, ie create a file. arff to be able to use the WEKA tool, we will select the variables that we want to analyze and that are appropriate.

```
@relation test_beck

@attribute var1 numeric
@attribute var2 numeric
@attribute var3 numeric
@attribute var4 numeric
@attribute var5 numeric
@attribute var6 numeric
@attribute var7 numeric
@attribute var8 numeric
@attribute var9 numeric
@attribute var10 numeric
@attribute var11 numeric
@attribute var12 numeric
@attribute var13 numeric
@attribute var14 numeric
@attribute var15 numeric
@attribute var16 numeric
@attribute var17 numeric
@attribute var18 numeric
@attribute var19 numeric
@attribute var20 numeric
@attribute var21 numeric

@data
0,0,0,0,0,0,0,0,0,3,1,0,0,0,0,0,0,0,0,0,0
0,0,0,1,0,0,0,0,0,0,0,1,0,0,1,1,2,1,0,0,0
```

Figure 4: Code for the preparation of data in the WEKA tool

3.4 Modeling

In this phase of the methodology, the most appropriate technique will be chosen for the marked objectives of data mining. Then, and once a test plan has been made for the selected models, we will proceed to apply these techniques on the data to generate the model and finally we will have to evaluate if said model has met the success criteria or not.

3.4.1 Modeling technique for the project

As a modeling technique for the project, the K-Means algorithm will be used to carry out the grouping, which involves identifying groups in the data, for this the WEKA tool will be used.

- We upload the data to the WEKA tool.

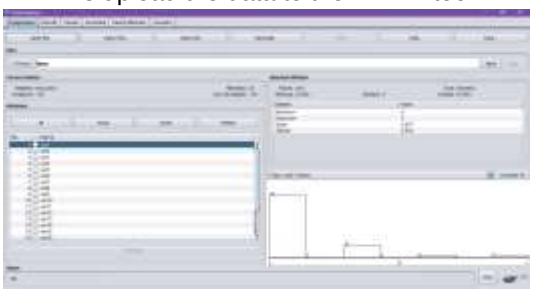


Figure 5: Loading data in the WEKA tool

- We selected cluster in WEKA.

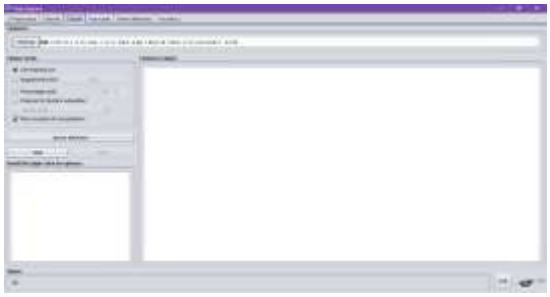


Figure 6: Cluster at WEKA

- We selected the K-Means algorithm.

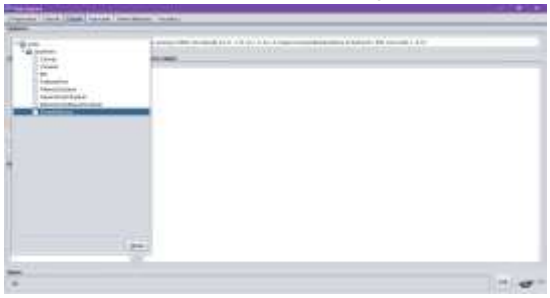


Figure 7: K-Means algorithm.

- We entered the number of clusters according to the possible diagnoses of depression.

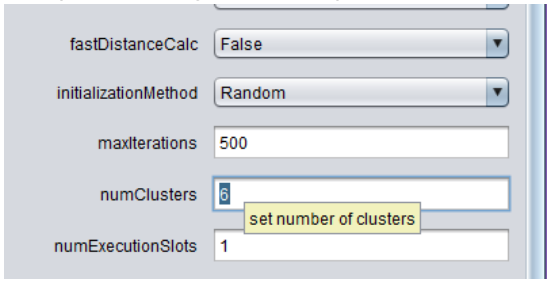


Figure 8: Number of clusters

- We started the WEKA tool.

Figure 9: Execution of the WEKA tool

- Results.

Cluster 0: 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
Cluster 1: 0,0,0,0,0,0,1,0,0,3,0,2,1,1,0,2,1,1,3,1,0
Cluster 2: 1,3,3,3,0,0,0,0,0,0,3,0,0,0,0,0,0,0,0,1,0
Cluster 3: 0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,1,1,0,0,0,0
Cluster 4: 0,0,0,1,0,0,0,1,0,0,0,0,2,0,0,1,0,0,0,1,0
Cluster 5: 2,0,0,2,0,0,0,1,0,0,1,0,0,1,1,1,3,0,0,1,0

0	73 (63%)	1	8 (7%)
2	1 (1%)	319	(16%)
4	11 (9%)	5	4 (3%)

3.5 Evaluation

From the business point of view, the description of the data obtained from the result of applying a depression evaluation test to high school students had been established as the main success criterion.

- We will proceed to describe the resulting clusters:

0	0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
1	0,0,0,0,0,0,1,0,0,3,0,2,1,1,0,2,1,1,3,1,0
2	1,3,3,3,0,0,0,0,0,0,3,0,0,0,0,0,0,0,0,1,0
3	0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,1,1,0,0,0,0
4	0,0,0,1,0,0,0,1,0,0,0,0,2,0,0,1,0,0,0,1,0
5	2,0,0,2,0,0,0,1,0,0,1,0,0,1,1,1,3,0,0,1,0

Table 2: Table of final clusters

In cluster 0, it can be observed that the randomly selected data correspond to a population of high school students with a diagnosis of depression evaluated according to the Beck depression inventory, "normal depression level".

Cluster 1 has data with characteristics that according to the evaluation test with a diagnosis of "mild disturbance of mood" with a probability of mutating to a diagnosis of "states of intermittent depression".

Cluster 2, its data correspond to standard characteristics of "mild disturbance of mood".

Cluster 3, its characteristics according to Beck's depression inventory are considered normal.

Cluster 4, according to the beck depression inventory are considered normal.

Finally, cluster 5, the data corresponding to this cluster are according to the invention of normal Beck depression, but with the probability of mutating to a level of depression of "mild disturbance of mood".

To obtain these results, 5 iterations were performed with the K-Means algorithm, this done as mentioned above in the WEKA tool.

Cluster	Population	%
0	73	63%
1	8	7%
2	1	1%
3	19	16%
4	11	9%
5	4	3%
Total:	116	99%

Table 3: Table of percentages of the clusters

In the previous table you can see that:

a) In cluster 0, of the total of the population 73 correspond to 63%.
b) In cluster 1, 8 samples correspond to 7%.
c) In cluster 2, of the total of population 1 corresponds to 1%.
d) In cluster 3, 19 correspond to 16%.
e) In cluster 4, 11 correspond to 9%.
f) In cluster 5, 4 correspond to 3%.

It can be said that according to the inventory of beck depression, the majority of the population's depression states are normal.

3.6 Deployment

This is the last phase of the CRISP-DM methodology and the objective of it is to expose the results obtained, to whom they serve.

As a result of the application of Beck's depression inventory, important information on this topic was obtained in upper secondary students, the K-Means algorithm was applied for the grouping of the data, 6 clusters were chosen because in the inventory of depression of Beck are 6 possible diagnoses that can be given according to the score obtained in the evaluation. Next, the percentage of the clusters remained.

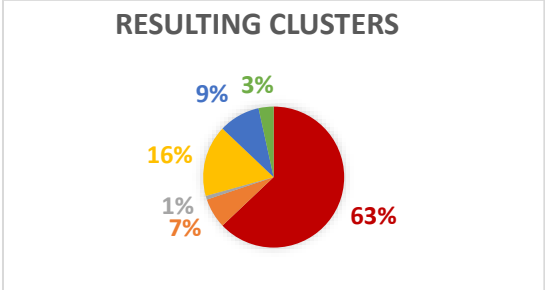


Figure 10: Graph of final clusters

According to the results obtained, it can be described that in the population of high school students 63% have a level of null or normal depression, while the remaining 37% have high depression.
These results help both parents and teachers to know what percentage of depression is related to students, this to have more control over what happens in their lives, and prevent situations such as suicide, dropping out of school, addictions, etc.
Next, it shows how the data is conformed in each cluster, where it can be seen that the response according to Beck's inventory predominates 0:

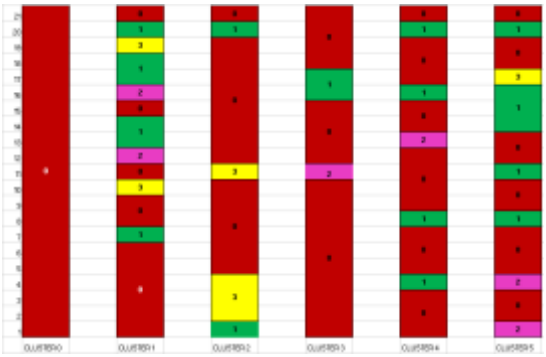


Figure 11: Conformation of the clusters

4. Conclusions and future work

During the development of the present work, a data mining course was taken where basic concepts were explained for the understanding of how it is carried out, as well as prediction algorithms such as the a priori and the K-Meas algorithm which is described. used in this work.

On the other hand, in order to carry out a data mining work, a search was made of the instrument for obtaining the data, where it was decided to use Beck's depression inventory, which was applied to 116 middle school students. superior, where there were problems due to the unsettling behavior of the students and the application of the instrument was complicated, during 3 days the

instrument was applied to different students completing the 116.

In addition, the methodology that was to be implemented was investigated, for the development of a data mining project, and in the course it was recommended to use the CRISP-DM methodology, which is completely focused on the development of mining projects.

Also, I want to mention that data mining is one of the most modern tools applied today and that the development of this work leaves us with a great understanding of how to work with big data and to be able to extract valuable information from them by means of techniques of data mining, it was possible to reach the desired objective, which was to describe or typify the level of depression in upper secondary students.

To finish the project, I would say that to give a continuation, the algorithm could be applied a priori to predict according to some characteristics those students who have some level of depression and thus act in time to help him.

Referencias

[1]Psyciencia. (19 de 08 de 2014). Obtenido de https://www.psyciencia.com/pdf-inventario-de-depresion-de-beck/

[2]Alcover, R. B. (2007). Análisis del rendimiento académico en los estudios de informática de la Universidad Politécnica de Valencia aplicando técnicas de minería de datos. . Valencia, España.

[3]Cecilia Díaz, A. R. (Marzo de 2004). Factores condicionantes de depresión en trabajadores metalúrgicos. Perú.

[4]Claudia Xiomara Santos León, S. M. (2011). Estados Emocionales y Nivel de Actividad Física en la Escuela Militar General José María Córdova.

[5]Guevara, G. R. (2016). Perfil estadistico en pacientes con depresion. México.

[6]guiasalud. (s.f.). Obtenido de http://www.guiasalud.es/egpc/depresion/completa/documentos/anexos/Anexo_9_Instrumentos_de_evaluacion_de_la_depresion.pdf

[7]Martínez, M. B. (s.f.). minería de datos. Puebla, Puebla, México.

[8]Sergio Valero Orea, A. S. (2005). Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos. Puebla, México.

[9]tachopedia. (s.f.). Obtenido de https://www.techopedia.com/definition/32057/k-means-clustering

[10]goicochea, a. (11 de 8 de 2009). CRISP-DM, Una metodología para proyectos de Minería de Datos. Obtenido de https://anibalgoicochea.com/2009/08/11/crisp-dm-una-metodologia-para-proyectos-de-mineria-de-datos/

.