

Measurement of Social Capital through Data Mining Techniques at Universidad ECCI

Medición de Capital Social en la Universidad ECCI a través de Técnicas de Minería de Datos

Manuel Fernando Cabrera Jiménez¹, Luz Stella García Monsalve¹, Olga Camila Hernández Morales¹,
Luz Adriana Suárez Suárez¹

¹Universidad ECCI, Bogotá, Colombia

Received: 28 Sep 2015

Accepted: 23 Feb 2016

Available Online: 19 Aug 2016

Abstract

By using Naive Bayes and C4.5 algorithms of data mining, this study of social capital at Universidad ECCI identifies the community's potential to generate positive associative relations that impact upon the quality of the citizen the Institution is delivering to the nation. A data collection instrument (survey) was applied that, from the cognitive and structural dimensions, permits characterizing the faculty's perception against elements like trust, norms, reciprocity, associativity, and cohesion regarding common objectives. Data were classified to identify the accumulated stock of social capital at the university to promote the generation of networks aimed to benefit the whole community, besides materializing the institutional mission and vision. As a result, the study identified low accumulated stock of social capital at the university; hence, it may be stated – in the first place – that the University faculty presents low levels of trust in all its participants, which affects construction of associativity strategies so that, from internal relationships, networks may be constructed to be part of the organizational assets; in the second place, the level of recognition and cohesion of the community tends to be at medium levels, a situation that leads to rethinking strategies that accomplish the materialization of the mission and vision in the construction of a much more committed community aware of its internal processes. In the third place, the study permits generating a diagnosis that reveals shortcomings in relation to participation in collegiate entities, where it is fundamental to involve the interests of the members of the community.

Keywords: Social capital, Data mining, Desertion.

Resumen

A partir del uso de los algoritmos de Naïve Bayes y C4.5 propios de la minería de datos, se aborda el estudio del capital social en la Universidad ECCI con el fin de identificar el potencial de la comunidad para generar relaciones asociativas positivas que impacten en la calidad del ciudadano que la institución entrega al país. Se aplicó un instrumento de recolección de datos (encuesta) que desde dos dimensiones (cognitiva y estructural) permitiera caracterizar la percepción de la comunidad docente frente a elementos como la confianza, las normas, la reciprocidad, la asociatividad y cohesión frente a objetivos comunes. Se clasificaron los datos con el propósito de identificar el stock acumulado de capital social en la universidad para impulsar la generación de redes orientadas al beneficio de toda la comunidad y además materializar la misión y la visión institucional. Como resultado del estudio se identificó un bajo stock acumulado de capital social en la universidad, por lo tanto es posible afirmar; en primer lugar, que la comunidad docente presenta bajos niveles de confianza en todos los actores de la Universidad, lo cual incide en la construcción de estrategias de asociatividad para que a partir de las relaciones internas se puedan construir redes que hagan parte de los activos organizacionales; en segundo lugar, el nivel de reconocimiento y cohesión de la comunidad tiende a estar en niveles medios, situación que conlleva al replanteamiento de estrategias que logren que la misión y la visión se materialicen en la construcción de una comunidad mucho más comprometida y conocedora de sus procesos internos. En tercer lugar, el estudio permite generar un diagnóstico que demuestra las falencias que se tienen en relación con la participación en entes colegiados en los cuales es fundamental involucrar los intereses de todos los miembros de la comunidad.

Palabras clave: Capital social, Minería de datos, Deserción.

*Corresponding Author.
E-mail: mcabreraj@eccu.edu.co

How to cite: Cabrera, M. F., García Monsalve, L. E., Hernández, O. C., Suárez Suárez, L. A., *Measurement of Social Capital through Data Mining Techniques at Universidad ECCI*, TECCIENCIA, Vol. 12 No. 21., 25-32, 2016, DOI: <http://dx.doi.org/>

1. Introduction

The process of measuring social capital in Colombia began from the development of a quantitative analysis sponsored by the National Planning Department (DNP, for the term in Spanish) in 1997; thereafter, other entities like the Bogotá Chamber of Commerce and the Antonio Restrepo Barco Foundation have conducted studies in 2005 and 2011. These studies on social capital carried out in the country have had the BARCAS methodology (Barometer of Social Capital by the World Bank) [1] as a base, which focused on the dimensions of: trust, validation of information sources, and social capital.

Results show low interpersonal trust – as well as low trust of institutions, high social inequity within cities, a growing unequal gap of living conditions between urban and rural populations, low credibility in communication media, social and political atomism, high clientelism, and low citizen trust in State management [2]. Based on this general background at the national level and starting from identifying the relevance of education in social transformation, this study seeks to characterize the perception level in the Universidad ECCI community in relation to the cognitive and structural dimensions of social capital, which will permit offering a limited study case as an example within the context of higher education.

1.1 Social capital and participants in higher education

Social capital is understood as a collective construct that emerged during the early 20th century, from the observation of social interaction in function of facilitating associative strategies that permit explaining social behaviors and potentiate education, community life, democracy, and economic growth in function of society's development. This concept has become quite popular within the context of social sciences due to the post-modern interaction of social structures; its contribution is considered a source of cohesion that allows generating relationships of citizens that articulated into networks generate skills and resources in a pre-established regulatory framework [3] [4].

Since the mid 1960's, as a result of economic development in Western society, organizations underwent important changes as a consequence of the new paradigms of knowledge and the economic and political dynamics of postwar processes permeating the new social relationships.

Scientists like Bordieu [5], Coleman [6], and Putnam [7] have addressed the study of social relationships in different cultural and economic settings and their incidence on human development from articulating them with other types of capital (natural capital, physical capital, human capital),

highlighting how social capital intrinsically evidences a potential as source of networks that permeate human relations from associativity to benefit all the members belonging to a network.

The World Bank (2000) classified social capital into cognitive and structural in function of their use; the first, is related to the individual's subjectivity and behavior supported on values, education, and cultural principles that permit reading the environment and interacting with it; the second, holds subjective and intangible components evidenced in the development of interpersonal relations supported by the individual's ideological forms and manifestations, religious beliefs, conception of values and attitudes, manifestations of emotions inherent to humans [8].

Each of the dimensions supports a series of interrelated components that permit evidencing an integrality that gives sense to social capital as construct of intangible capital; this integrality is founded on sharing principles and values, which are the foundations of the processes of positive associativity with community sense, development of civic life, and consolidation of trust at transverse level and with the social system [9].

Table 1 shows the variables of social capital proposed by the World Bank and which were kept in mind to conduct the research field work that supports this article.

1.2 Use of data mining within the scope of social relationships

Within the context of education, relationships generated among members of a community potentiate the consolidation of networks that seek collective benefit, with this being the principal characteristic of social capital. Some studies evidence the use of data mining within the context of social sciences [10], permitting automatic or semi-automatic exploration of large amounts of data to find repetitive patterns, regulations, and tendencies that explain the behavior of situations susceptible to being analyzed [11]. Data analysis was performed through Naïve Bayes [12] [13] and C4.5 [14] [15] data mining algorithms to generate indices to measure social capital at Universidad ECCI.

1.2.1 Naïve Bayes

This is a technique which takes as primary source that of the Thomas Bayes theorem, which can predict the probability of a case belonging to a given class; this method is based on the probability theory, uses frequencies to calculate conditional probabilities and, thus, generate predictions on new cases.

Table 1 Description of social capital variables by the World Bank. Source: web page World Bank Org. (2009).

Variable	Description
Participation in networks	Capacity to form new networks related between individuals and groups
Reciprocity	Altruism in the short term and personal interest in the long term
Trust	Willingness to accept risk. It is expected that individuals respond as support is offered
Social norms	Those who know these informal norms assume behaviors within a given social context; obvious need for legal and institutionalized actions
Proactivity	Development of social capital requires active and vindicating involvement of the citizens in a community.
Solidarity	Thoughts and attitudes of individuals to relate to others
Collective action and cooperation	Organized activities in a community to reach precise goals
Information and communication	Dissemination of information, promotion of dialogue among participants

Part of the assumption is that all the attributes are independent when the value of the class variable has been determined [16]. Equation (1) will be used to find the probability according to the Bayes theorem and perform the analysis procedure.

$$P(F_j | E_1 E_2 \dots E_m) = \frac{P(E_1 | F_j) P(E_2 | F_j) \dots P(E_m | F_j) P(F_j)}{P(E_1 E_2 \dots E_m)} \quad (1)$$

Here E and F are events and E_i are independent for a given F_j .

1.2.2 C4.5 algorithm

This algorithm based on equation (2) constructs a decision tree where its internal nodes are labelled as attributes, the branches protruding from each node represent tests for the values of the attribute, and the leaves identify the categories. It is based on the heuristic technique known as proportion of gain and permits considering all the possible tests that can divide the data set and selects the test with the highest information gain.

$$GR(A_i) = \frac{G(A_i)}{I(División A_i)} = \frac{G(A_i)}{-\sum_{j=1}^{nv(A_i)} \frac{n_{ij}}{n} \log_2 \left(\frac{n_{ij}}{n} \right)} \quad (2)$$

Here $I(División A_i)$ is the rupture information, and the proportion of gain is $\frac{Ganancia(S)}{I-división(S)}$.

1.3 Weka software tool

Some software tools support data mining processes, allowing management of large volumes of data, which is the case of Weka – developed in Waikato in New Zealand and which in this study shows results of the process of the Naïve Bayes and C4.5 algorithms (denominated J48 in the Weka tool).

Upon processing the data, the Weka tool generates a confusion matrix to verify the distribution of errors made by the classifier, where the VP (true positives) are instances correctly recognized by the system; said value corresponds to the proportion examples classified as class x, from among all the examples that truly have class x, that is, what amount of the class has been captured. In the confusion matrix, it is the value of the element of the diagonal divided by the sum of the row. The FN (false negatives) are positive instances and which the system says are not. The FP (false positives) are the negative instances, but the system says they are not. The VN (true negatives) are the negative instances and correctly recognized as such [17]. In addition, the Weka tool also obtains results, like: instances badly classified, instances well classified, Kappa statistic, absolute error, recall, precision, and F-Measure. Instances badly classified and well classified correspond to registries that were incorrectly or correctly classified. The Kappa statistic [18] measures the coincidence of the prediction with the real class.

Absolute error (mean absolute error): indicates the quality of the measurement. Indicates the mean error produced in each prediction.

Recall: measures the proportion of terms correctly recognized with respect to the total of real terms. This index is obtained through equation (3).

$$IR(M, A, B) = \frac{VP}{VP + FN} \quad (3)$$

Here $IR(M, A, B)$ is the Recall index for Medium, Low, and High.

Precision: measures the number of terms correctly recognized with respect to the total of terms predicted, whether true or false. In this article, this index is calculated based on equation (4).

$$IP(M, A, B) = \frac{VP}{VP+FP} \quad (4)$$

Here $IP(M, A, B)$ is the precision index for Medium, Low, and High.

F-Measure: this statistic combines Precision and Recall in a balanced harmonic average of the results obtained. Equation (5) is used to generate the index.

$$F(M, A, B) = \frac{2VP}{2VP+FP+FN} \quad (5)$$

Here $F(M, A, B)$ is the F-Measure for Medium, Low, and High.

2. Methods and materials

During the development of the project to measure social capital at Universidad ECCI, the researchers worked with data sources provided by the Institution and which correspond to professors hired on the first semester of 2015; these data generated a universe of 304 individuals from which a sample was calculated equivalent to 208 subjects surveyed. The structure of the instruments used to collect data was based on the methodology by the World Bank to measure social capital globally. Figure 1 shows the process used in the research and which is based on the typical steps of the knowledge extraction process.

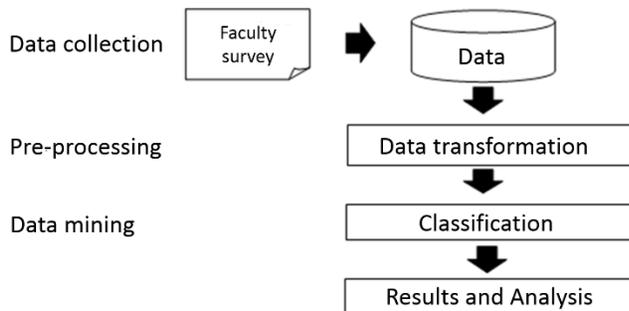


Figure 1 Method applied in identifying social capital at Universidad ECCI.

2.1 Data collection

Social capital, understood as a whole and comprising a structural dimension based on horizontal and vertical horizons regulated by norms as basis of associativity, and a cognitive dimension assumed from the values interiorized by the citizen, such as: trust, reciprocity, solidarity, and civic sense of that which is public.

Social capital identifies tensions created from the lack of associativity, which affects social realities that impact upon

the possibility of generating networks. In this sense, social capital is assumed as a means to increase positive associativity. The following presents the structure to design the instrument that permitted collecting primary information for its subsequent processing and analysis. Information was gathered integrated into a single data set from a 54-question survey (Table 2), which was answered by 208 professors from the Institution.

Table 2 Questions from the survey given to professors.

COGNITIVE SOCIAL CAPITAL	
Question	Type of response
What level of importance does trust have as a base of social relationships? Responses: 1, 2, 3. 1 being low and 3 high.	Ordinal scale rating
Do you consider that during the course of your life you have received education that allows you to better understand your environment in general? Responses: 1, 2, 3. 1 being low and 3 high.	Ordinal scale rating
To what degree are the following values promoted in the University? Team work, Tolerance, Empowerment, Leadership, Solidarity, Reciprocity, Respect for norms Responses: 1, 2, 3. 1 being low and 3 high.	Ordinal scale rating
Do you know the mission of your University? Yes, No, Does not know / Does not respond	Categorical polytomous variables
Would you agree with establishing an elective course on social capital at the University that promotes associativity, reciprocity, and equity based on trust? Yes, No, Does not know / Does not respond	Categorical polytomous variables
To what degree do you experience trust with the following players in general? Close friends, university classmates, Students, Administrative staff, Directors, Public servants related to education. Responses: 1, 2, 3. 1 being low and 3 high.	Ordinal scale rating
During the course of your educational and professional life, to what level have you excluded people with differences regarding the following factors? Socio-economic condition, Religious belief, Sexual Preference, Ethnic group, Physical disability, Another nationality. Responses: 1, 2, 3. 1 being low and 3 high.	Ordinal scale rating
During the course of your educational and professional life, to what level have you felt due to differences regarding the following factors? Socio-economic condition, Religious belief, Sexual Preference, Ethnic group, Physical disability, Another nationality. Responses: 1, 2, 3. 1 being low and 3 high.	Ordinal scale rating

Do you belong to groups in your University, such as? Sports groups, artistic groups, academic reflection groups, research groups, others. Responses: 1, 2, 3. 1 being low and 3 high.	Ordinal scale rating
Do you consider it important to participate as a voter in University government entities with professor participation in areas of? Support and solidarity brigades, Program committees, Faculty Council, Welfare committee, Election and representation to collegiate corps. Yes, No, Does not know / Does not respond	Ordinal scale rating
At what level of excellence do you consider those graduating from the University will be assessed as professionals within the work context? Responses: 1, 2, 3. 1 being low and 3 high.	Categorical polytomous variables
STRUCTURAL SOCIAL CAPITAL	
Question	Type of response
Is there evidence in the University of opportunities that permit being part of groups that promote development of projects at extracurricular level? Yes, No, Does not know / Does not respond	Categorical polytomous variables
Do you know the normativity that regulates internal relationships at the University? Yes, No, Does not know / Does not respond	Categorical polytomous variables
Do you consider that information with superior entities of the University is clear and flows within the framework of established guidelines? Yes, No, Does not know / Does not respond	Categorical polytomous variables
To what degree do you identify internal and external networks at the University that permit generating associative work with other universities? Responses: 1, 2, 3. 1 being low and 3 high.	Ordinal scale rating
Is your perception on quality in general at the University more positive now than a year ago? Responses: 1, 2, 3. 1 being low and 3 high.	Ordinal scale rating
Is leadership promoted amongst outstanding professors in the University at academic and innovative levels? Yes, No, Does not know / Does not respond	Ordinal scale rating
As professor, are you invited to participate in self-evaluation and improvement processes in the following areas? Institutional committee, at program level, faculty, institutional Yes, No, Does not know / Does not respond	Categorical polytomous variables
At what level do you consider the academic community in general is willing to participate in associative networks? Responses: 1, 2, 3. 1 being low and 3 high.	Ordinal scale rating
At what level do you consider the University is willing to support in the development of associative networks? Responses: 1, 2, 3. 1 being low and 3 high.	Ordinal scale rating

At what level are communication channels in general effective in the University? Responses: 1, 2, 3. 1 being low and 3 high.	Ordinal scale rating
--	----------------------

2.2 Pre-processing

This stage, generally, undertakes tasks, like: data cleaning and transformation of variables; data are separated to later apply the data mining algorithms. The study surveys considered questions with High, Medium, and Low responses and questions with Yes, No, Does not know/Does not respond responses, which were transformed into numerical values to generate an indicator that will facilitate measuring each set of data and from there generate the class (Low, Medium, High).

The transformation to numerical values was done based on the following formula:

$$C_i = \frac{P_i - P_{\min}}{P_{\max} - P_{\min}} \quad (6)$$

Where,

C_i is the estimated score per professor.

P_{\min} is the minimum value of the sum of the set of data.

P_{\max} is the maximum value of the sum of the set of data.

P_i is the sum of the scores obtained in the Likert scale [18].

To characterize the class, a division was conducted of equally spaced subclasses, that is, values between 0 and 0.33 correspond to the low social capital index; values above 0.33 and below 0.66 correspond to the medium social capital index; and values above 0.66 correspond to the high social capital index.

2.3 Data mining

As mentioned above, at this stage Naïve Bayes and C4.5 data mining algorithms were used to classify the level (Low, Medium, and High) of perception of trust of the professors from Universidad ECCI.

3. Results and Analysis

The research integrated social capital and data mining to characterize the perception level of the ECCI community in relation to the cognitive and structural dimensions.

Table 3 shows class instances, evidencing that the number of registries classified as high are: 8; the number of registries classified as low are 31; and the number of registries in the medium category are 169, which indicates that much of the population surveyed ignores the extent of the benefits of

social capital as a cohesion factor. This helps us to understand why the low and medium appreciation prevails in relation to conceiving it as an option of articulation and affinity to generate community.

Additionally, the study demonstrates that a low number of people value as high the extent of social capital and these results permit suspecting the causes endogenous to Universidad ECCI that lead to this situation, identified thus:

Firstly, a high level of lack of knowledge exist of the values that integrate the different social structures within the context of the University and based on social capital; secondly, lack of interest may be a factor that conditions community participation in endogenous processes; and thirdly, communication strategies do not achieve their goal within the ECCI academic community.

Table 3 Class instances per category.

Category	Value
Medium	169
Low	31
High	8

Upon processing the 208 registries through Naïve Bayes and C4.5 algorithms, Weka yielded the results shown in Table 4.

Table 4 Results obtained by the Naïve Bayes algorithm.

Factor	Naïve Bayes	C4.5
Correctly Classified Instances	189 (90.9%)	203 (97.6%)
Incorrectly Classified Instances	19 (9.1%)	5 (2.4%)
Kappa statistic	0.76	0.92
Mean absolute error	0.060	0.025
Relative absolute error	28.2%	11.9%

For the Naïve Bayes algorithm, the Kappa coefficient index is within the range from 0.61 and 0.80 with a considerable concordance rate, which indicates that the social capital accumulated in the institution does not reflect the capacity to construct solidarity relationships based on associativity that enhances the cognitive and structural dimensions of social capital.

The C4.5 algorithm, with a Kappa coefficient index within the range from 0.81 to 1, has an almost perfect concordance rate, which indicates this method's power under the parameters defined by the study participants, given each of their characteristics, the social capital index may be established with a very small margin of error.

The cognitive dimensions (values, principles) and structural dimensions (networks, norms) are assumed as components of a whole denominated as social capital, which requires coherent development and articulation, for example, a society with low levels of social values, tends to show lower recognition of the norms and, hence, shows increased levels of distrust in social relationships.

In this sense, the study showed that the ECCI community is prone to constructing relations based on high levels of recognition of norms, but not on their appropriation, a phenomenon not foreign to the behavior of the national context.

3.1 Confusion matrix

Confusion matrix of the Naïve Bayes algorithm. Table 5 shows that of the 208 registries considered for the experiment, 169 that correspond to 81.25% are Medium, 31 equivalent to 14.90% are Low, and 8 equivalent to 3.85% are High.

Table 5 Confusion matrix of the Naïve Bayes algorithm.

		Prediction			
		Medium	Low	High	
Class data	Medium	151	11	7	169
	Low	1	30	0	31
	High	0	0	8	8
		152	41	15	

The following analyzes each of the categories for the Naïve Bayes algorithm:

Table 6 Prediction of medium values with Naïve Bayes.

151 True positives (Medium correctly classified as Medium)	1 False positive (1 Low incorrectly classified as Medium)
18 False negatives (18 Medium incorrectly classified as Low and High)	38 True negatives (correctly classified as high and low, but not medium)

The classifier correctly identified 151 of the 169 Medium values (Table 6), equivalent to 89.35% (VP) and correctly identified 38 of the 39 different registries (High and Low) (97.4%, which are VN). Likewise, the classifier incorrectly identified 18 registries of the 169 with an error percentage of 10.65% (FN), and 1 of the 39 with an error percentage of 2.56% (FP).

The classifier correctly identified 8 of the 8 High (Table 7), equivalent to 100% (VP), and correctly identified 193 of the 200 different registries (Medium and Low) (96.50%, which are VN). Likewise, the classifier incorrectly identified 0 registries of the 8 with an error percentage of 0.0% (FN) and 7 of the 200 with an error percentage of 3.50% (FP).

Table 7 Prediction of high values with Naïve Bayes.

8 True positives (High correctly classified as High)	7 False positive (7 Medium incorrectly classified as High)
0 False negatives	193 True negatives (correctly classified as low and medium, but not high)

Table 8 Prediction of low values with Naïve Bayes.

30 True positives (Low classified correctly as Low)	11 False positives (11 Medium incorrectly classified as Low)
1 False negative (1 Low incorrectly classified as Medium)	166 True negatives (correctly classified as high and medium, but not low)

The classifier correctly identified 30 of the 31 Low (Table 8), equivalent to 96.77% (VP), and correctly identified 166 of the 177 different registries (High and Medium) (93.79%, which are VN). Likewise, the classifier incorrectly identified 1 registry of the 31 with an error percentage of 3.23% (FN) and 11 of the 177 with an error percentage of 6.21% (FP).

Confusion matrix of the C4.5 algorithm. Table 9 shows that of the 208 registries considered for the experiment, 169 – corresponding to 81.25% - are Medium, 31 equivalent to 14.90% are Low, and 8 equivalent to 3.85% are High.

Table 9 Confusion matrix of the C4.5 algorithm.

		Prediction			
		Medium	Low	High	
Class data	Medium	168	0	1	169
	Low	2	29	0	31
	High	2	0	6	8
		172	29	7	

The classifier correctly identified 168 of the 169 medium values (Table 10), equivalent to 99.41% (VP) and correctly identified 35 of the 39 different registries (High and Low) (89.74%, which are VN). Likewise, the classifier incorrectly identified one registry of the 169 with an error percentage of

0.59% (FN) and 4 of the 39 with an error percentage of 10.26% (FP).

Table 10 Prediction of medium values with C4.5.

168 True positives (Medium correctly classified as Medium)	4 False positives (2 Low and 2 High incorrectly classified as Medium)
1 False negative (1 Medium incorrectly classified as High)	35 True negatives (correctly classified as High and Low, but not Medium)

Table 11 Prediction of high values with C4.5.

6 True positives (High correctly classified as High)	1 False positive (1 Medium incorrectly classified as High)
2 False negatives (2 High incorrectly classified as Medium)	199 True negatives (correctly classified as Low and Medium, but not High)

The classifier correctly identified 6 of the 8 High values (Table 11), equivalent to 75% (VP) and correctly identified 199 of the 200 different registries (Medium and Low) (99.50%, which are VN). Likewise, the classifier incorrectly identified two registries of the eight with an error percentage of 25% (FN) and 1 of the 200 with an error percentage of 0.50% (FP).

Table 12 Prediction of low values with C4.5.

29 True positives (Low correctly classified as Low)	0 false positives
2 false negatives (2 Low incorrectly classified as Medium)	177 True negatives (correctly classified as High and Medium, but not Low)

The classifier correctly identified 29 of the 31 low values (Table 12), equivalent to 93.55% (VP), and correctly identified 177 of the 177 different registries (High and Medium) (100%, which are VN). Likewise, the classifier incorrectly identified 2 registries of the 31 with an error percentage of 6.45% (FN) and 0 of the 177 with an error percentage of 0% (FP).

3.2 Precision

The Naïve Bayes and C4.5 algorithms classify data primarily into Medium and Low, which can be interpreted as low levels of trust in their participants (Table 13).

Table 13 Results of Precision, Recall, and F-Measure with the Naïve Bayes algorithm.

Class	Precision		Recall		F-Measure	
	Naïve Bayes	C4.5	Naïve Bayes	C4.5	Naïve Bayes	C4.5
Medium	0.993	0.977	0.893	0.994	0.941	0.985
Low	0.7	1	0.968	0.935	0.833	0.967
High	0.533	0.857	1	0.75	0.696	0.8

4. Conclusions

The study conducted in Universidad ECCI allowed the identification of a low accumulated stock of social capital not foreign to the national reality; thereby, it may be stated – firstly – that the institution has low levels of trust in all its participants, which affects the construction of positive associativity strategies with community sense to, from the internal relationships, be able to build networks that are part of the organizational social assets; secondly, the level of the community’s recognition and cohesion tends to be at medium levels.

This situation leads to rethinking strategies that achieve the materialization of the institutional mission and vision in the construction of a community that is much more committed, cohesive, and knowledgeable of its internal processes. Thirdly, the study permits generating a diagnosis that demonstrates the drawbacks in relation to participation in internal collegiate bodies where it is fundamental to visualize the interests of all the members of the community.

When comparing the algorithms applied in the article, it was noted that the C4.5 Kappa confidence interval is higher than that of the Naïve Bayes, as well as when classifying with both algorithms it may be determined that of 208 registries processed, the Naïve Bayes algorithm obtained 189 coincidences and the C4.5 203; hence, C4.5 has 98% assertiveness and Naïve Bayes 91%.

Finally, the study permits identifying that education permeates and affects the construction of social capital and plays an integrating role of skills and knowledge that permit the transformation of a social conglomerate founded on principles of trust, reciprocity, recognition of norms, and participation in function of the community’s social benefit.

5. References

[1] Barcas: Barómetro de Capital Social del Banco Mundial. Banco Mundial. 1999.

[2] Foliaco Gamboa, Julio. Capital Social: importancia de las mediciones para Colombia. v.18, n. 2, p. 43-60. dic. 2013. ISSN 2422-5053.

Available in:
<<http://revistas.ufps.edu.co/ojs/index.php/respuestas/article/view/26>>.

[3] Paul S. Adler and Seok-Woo Kwon. Social Capital: Prospects for a new concept. *The Academy of Management Review*, Vol. 27, No. 1, pp. 17-40. 2002.

[4] Boxman, Paul M. De Graaf and Hendrik D. Flap. The impact of social and human capital on the income attainment of Dutch managers. *Social Networks Journal*. 1991.

[5] Bordieu, P. The forms of capital. In J.G. Richardson (ed.) *Handbook of theory and Research in Sociology of Education*. New York: Greenwood. 1986.

[6] Coleman, J. Social Capital in the creation of human capital. *American Journal of Sociology*. 1988.

[7] Putnam, R. The prosperous community: Social capital and public life. *The American Prospect*. 1993.

[8] Grootaert Christiaan, Bastelaer Thierry. Understanding and Measuring Social Capital. A synthesis of finding and recommendations from the social capital initiative. 2001.

[9] Atria Raúl, Siles Marcelo, Arriaga Irma, Robinson Lindon, Whiteford Scott. Capital social y reducción de la pobreza en América Latina y el Caribe: en busca de un nuevo paradigma. 2003.

[10] Kovanovic Vitomir, Joksimovic Srecko, Gasevic Dragan. What is the Source of Social Capital? The association between Social Network position and Social Presence in communities of inquiry. 2014.

[11] Mohammed J. Zaki, Wagner Meira Jr. Data mining and analysis. *Fundamental Concepts and Algorithms*. Cambridge University Press. ISBN: 9780521766333. 2014.

[12] Berger James O. *Statistical Decision Theory and Bayesian Analysis*. Second edition. Springer-Verlag. New York. 1985.

[13] Pacheco Samuel, Díaz Luis, García Rodolfo. El clasificador Naïve Bayes en la extracción del conocimiento de bases de datos. *Posgrado en Ingeniería de Sistemas*. FIME-UANL.2005.

[14] Quinlan, J. R. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.

[15] Duda, Richard. Hart, Peter. Stork, David. *Pattern Classification*. United States of America: John Wiley and Sons. 2001.

[16] Han Jiawei, Kamber Micheline. *Data Mining: Concepts and Techniques*. The Morgan Kaufmann Series in Data Management Systems, Jim Gray. Series Editor, Morgan Kaufmann Publishers. March 2006.

[17] Consulted 01 October 2015 from <http://www.cs.waikato.ac.nz/ml/weka/>

[18] Landis J, Koch G: The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-74.