Neutrosophic Sets and Systems
{Special Issue:Impact of neutrosophic scientific publication in Latin American context}, Vol. 44, 2021

University of New Mexico

Neutrosophy to Enrich Legal Opinion Mining

Wilson Eduardo Castro Núñez[1], Paul Alejandro Centeno Maldonado[2], Diego Xavier Chamorro Valencia[3] , Danilo Santiago Alvarado Ibarra[4], and Anthony Rafael Sotolongo León[5]

[1] Universidad Regional Autónoma de los Andes (UNIANDES). Transversal Central, Quevedo. Los Ríos. Ecuador. Email: uq.wilsoncastro@uniandes.edu.ec
[2] Universidad Regional Autónoma de los Andes (UNIANDES). Avenida Lizarzaburu. Riobamba, Chimborazo. Email: ur.lpaulcenteno@uniandes.edu.ec
[3] Universidad Regional Autónoma de los Andes (UNIANDES). Juan de Salinas 612-500, Ibarra. Imbabura. Ecuador. Email: ui.diegochamorro@uniandes.edu.ec
[4] Universidad Regional Autónoma de los Andes (UNIANDES). Km 5 ½ vía a Baños. Ambato. Tungurahua. Ecuador. Email: ua.daniloalvarado@uniandes.edu.ec
[5] OnGres, Santiago de Chile, Chile. Email: asotolongo@ongres.com

Abstract. The main characteristics of legal research lie in its complex operation and the articulation between science, technology, politics, and law. This is why it is claimed that the technology helps experts in processing information resulting from the sources of a legal investigation. Furthermore, the data mining process offers statistical techniques and pattern recognition and trend identification in the stored information. In this way, it facilitates its study and offers a significant contribution to the speed and efficiency required by today's users. Therefore, the main objective of the research is to develop a simple method for mining legal opinions enriched with Neutrosophy that contributes to legal scientific research. This multidisciplinary project allowed obtaining a model that integrates single value neutrosophic numbers to calculate polarity and statistical processing based on data clustering.

Keywords: opinion mining, legal research, SVN, clustering, Orange.

1. Introduction

Legal research is of vital importance for law professionals. In fact, according to what was stated by the authors in [1]:

> The dual system of professional training in the legal sciences is based on the premise of the investigation as an element that reinforces and improves the performance of the student, who incorporates this competence, to his daily performance as a resolution mechanism in the face of any problem situation (p. 3).

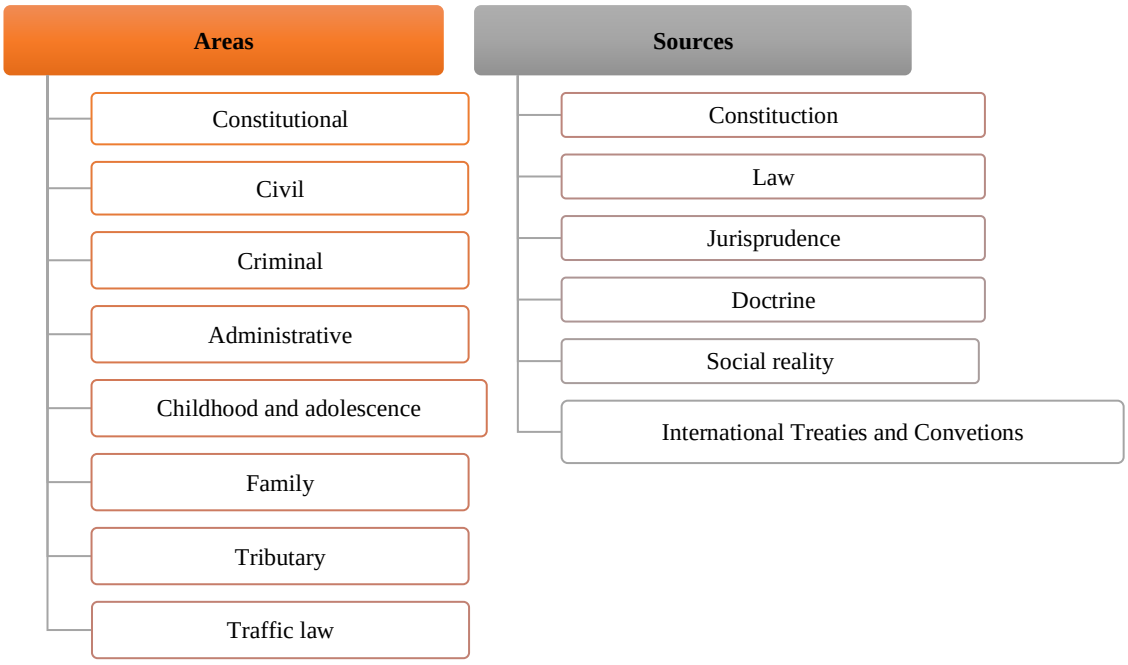

Figure 1: Main areas and sources of legal research.

To establish the theoretical and epistemological foundations of this type of research, the details stated in figure 2 must be taken into account since they are necessary to identify specific problems in law and legal sciences and propose solutions through legal scientific production. These models, methods, and paradigms have qualitative as a fundamental component, one of their primary applications being the analysis of legal texts associated with the sources and areas shown in Figure 1.

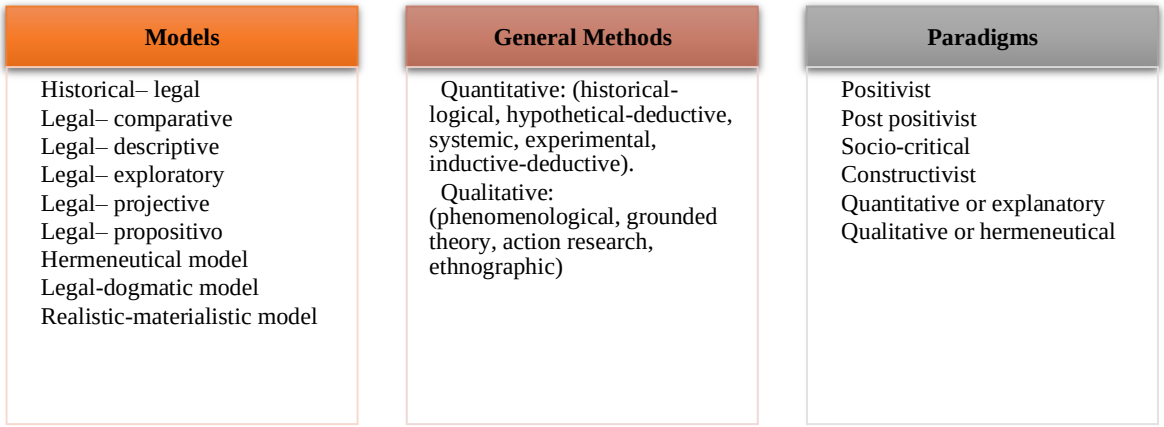

Figure 2: Models, methods and paradigms for legal scientific research.

According to [1], the main characteristics of legal research lie in its complex behavior and the articulation between science, technology, politics, and law, which leads to an updated view of it. That is why technology currently helps experts in processing scientific evidence [1] in all the areas in which it develops. Nowadays, this issue has been consistently consolidated because Artificial Intelligence has made Opinion Mining and Sentiment Analysis possible. It is worth remarking that in the consulted references, we could observe that the authors are divided regarding establishing the similarities and differences of these terms. Some say that it is not the same since the first focuses on the narrative while the second on pure emotions. However, all agree on the help provided by both techniques for legal scientific work [2-5].

Statistical techniques and techniques for pattern recognition and trend identification are used in the stored information through the data mining process. In this way, it facilitates its study and offers a great contribution to the speed and efficiency required by today's users [4, 6-9]. The foregoing is supported by the fact that opinion mining subtasks have evolved and now include [10]:

1. Subjectivity analysis: implies determining whether one of the topics for the blogging sub-community is text (neutral in feeling) or subjective (expresses a positive or negative sentiment or opinion).
2. Polarity analysis: includes the prediction of whether a text that has been established as subjective is positive or negative in its polarity
3. Degree of polarity: measures positive or negative polarity, in a subjective text.

It should be noted that among the advantages of using data mining is the ease of its implementation and its various applications. But the main difficulty in applying these models lies in the effort required in the evaluative processes of the different algorithms. However, as it was possible to verify in the previous paragraph, the polarity analysis is dichotomous, so applying it to a legal investigation where neutralities exist, is not convenient. Therefore, the authors of the current work agree on the need to incorporate a science such as Neutrosophy [11-20], since it analyzes this type of situation. Then all these criteria will be unified, while still simplifying the process to obtain cost reduction. Thus, this research will be carried out with free software so as not to have to pay excessive licenses [1-5, 10, 21].

According to the need to improve and humanize legal science researchers' work, we will develop a simple method for mining legal opinions enriched with Neutrosophy that contributes to legal scientific research. Which is exposed as the main objective of this article. Since it is established as a hypothesis of whether the text processing is optimized that allows analyzing opinions to obtain knowledge on a certain subject, it will be possible to contribute effectively to legal research. In addition, with the introduction of neutrosophic science, it will be possible to complement the opinion analysis system covering a broader field of variables that enrich the qualitative-quantitative research of any student of the social sciences in the legal field. For the fulfillment of this, in the first place, a theoretical framework will be established with the previous knowledge in section two. Subsequently, the method developed will be presented in section three and its respective application in the Results section. Finally, the conclusions reached and the bibliographic references of the research will be presented.

2 Previous knowledge

For the analysis of this topic within the present research, we start from a common thread of thought as illustrated in figure 3:

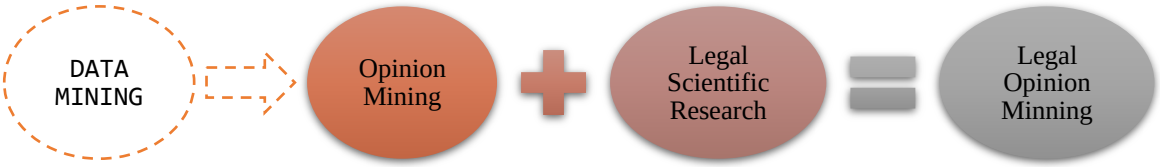

Figure 3: Relationship diagram.

2.1 Data mining

Data mining is based on statistics, database management, and artificial intelligence models and is closely related to computer science. Its purpose is to generate knowledge from databases [2-5, 10].

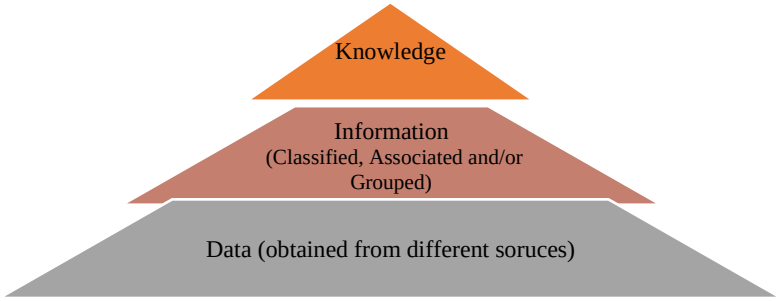

Figure 3: Phases of data mining. Adapted by the authors from [3].

In general, a Database Knowledge process consists of 5 stages[5]: data selection, data preparation or pre-processing, exploratory analysis, Data mining, interpretation of patterns, and evaluation of models. Obtaining the final DM model depends on the last two stages, which requires successive iterations on the exploratory analysis of the input data, and obtaining patterns and interpretation of results.[5].

2.2 Opinion mining

According to [10]:

> Opinion mining identifies the author's point of view on a topic rather than simply identifying the topic itself (…). Analyzing the text for your opinions can be extremely valuable to a legal researcher looking for a perspective on a legal topic or even information on a product or service. Organizations can also benefit from automatic opinion extraction by getting a timely picture of how their customers view their products or services, or more generally their names. The Web is a sprawling environment where customers seek or submit opinions that may be ripe for mining. An increasing number of customer views are posted on blogs (short for Weblogs). The content of these blogs can range from brief product reviews by consumers to elaborate essays on legal issues by law professors (p. 231).

3 Materials and methods

To develop the work method, a Legal-exploratory model will be used, where exploratory research based on the mining of legal options is used to describe the phenomena as they exist. In other words, it will be used to identify and obtain information on the polarities of opinions in the legal field. All this pursuing a quantitative or explanatory paradigm where the researcher describes what he has found based on his condition as an observer. The neutrosophic details on which the method will be based will be explained below.

3.1 Neutrosophy

Neutrosophy is a new branch of philosophy that studies the origin, nature and scope of neutralities created by Professor Florentin Smarandache. Its incorporation guarantees that the uncertainty of decision-making is considered, including indeterminacies where experts will issue their criteria evaluating linguistic and non-numerical terms, which constitutes the most natural form of measurement in human beings [13, 22-26]. Logic and neutrosophic sets, on the other hand, constitute a generalization of Zadeh's logic and fuzzy sets, and especially of Atanassov's intuitionist logic, with multiple applications in the field of decision-making and machine learning [13, 23, 27-29]. The truth value in the neutrosophic set is defined as follows [29-31]**:**

Definition 1 [13, 32, 33]: Be X a universe of discourse, a Neutrosophic Set (NS) is characterized by three membership functions, $u_A(x):, r_A(x):, v_A(x): X \rightarrow]^{-}0, 1^{+}[$ which satisfy the condition $-0 \leq \inf u_A(x) + \inf r_A(x) + \inf v_A(x) \leq \sup u_A(x) + \sup r_A(x) + \sup v_A(x) \leq 3^{+}$ $x \subset X$ $\forall u_A(x), r_A(x)$ and $v_A(x)$: denote the membership functions of true, indeterminate, and false of x in A, respectively, and their images are standard or non-standard subsets of $]^{-}0, 1^{+}[$. Let $N = \{(T, I, F): T, I, F \subseteq [0,1]\} n$ be a neutrosophic evaluation of a mapping of a group of formulas propositional to N, and for each sentence p:

$$\boldsymbol{v(p) = (T, I, F)} \quad (1)$$

To facilitate the practical application in real-world problems [7], the use of Single-Value neutrosophic Sets (SVNS) was proposed, through which it is likely to use linguistic terms to obtain greater interpretability of the results [8]. Let X be a universe of discourse, an SVNS A over X has the following form [9]:

$$\boldsymbol{A = \{\langle x, u_a(x), r_a(x), v_a(x)\rangle : x \in X\}} \quad (2)$$

Where $\boldsymbol{u_a(x): X \rightarrow [0,1], r_a(x): X \rightarrow [0,1]\ y\ v_a(x): X \rightarrow [0,1]}$

With $\boldsymbol{0 \leq u_a(x), r_a(x), v_a(x) \leq 3, \forall\, x \in X}$ (3)

The intervals $\boldsymbol{u_a(x), r_a(x)\ y\ v_a(x)}$ denote the memberships related to true, indeterminate and false from x in A, respectively[10]. For convenience reasons, a Single Value Neutrosophic Number (SVNN) is expressed as A = (a, b, c), where a, b, c$\in$ [0.1] and $0 \leq a + b + c \leq 3$.

Linguistic terms	SVNN numbers
Extremely good (EG)	(1,0,0)
Very very good (VVG)	(0.9, 0.1, 0.1)
Very good (VG)	(0.8,0,15,0.20)
Good (G)	(0.70,0.25,0.30)
Medium good (MDG)	(0.60,0.35,0.40)
Medium (M)	(0.50,0.50,0.50)
Moderately bad (MDB)	(0.40,0.65,0.60)
Bad (B)	(0.30,0.75,0.70)
Very bad (VB)	(0.20,0.85,0.80)
Very very bad (VVB)	(0.10,0.90,0.90)
Extremely bad (EB)	(0,1,1)

Table 1: Linguistic terms used. Source: [34].

Let A = (a, b, c) be a single-valued neutrosophic number, a score function S related to a single-valued neutrosophic value, based on the truth-membership degree, indeterminacy-membership degree, and false membership degree is defined [35]:

$$\boldsymbol{s(V_i) = 2 + T_i - F_i - I_j} \quad (4)$$

3.2 Designed method

For the development of the model, we based on what was exposed by the investigations of [2-10, 14-21, 30, 36-41] and obtained the following:

1. Define the subject area of interest
2. Identify external or internal sources of information
3. Data collection from identified sources
4. Formation of the corpus: the extracted data will be filtered and ordered to form the dataset destined for legal research (corpus), whether they are complete writings of an author, laws, among others. That is, an analysis of the extracted data is made to verify that they comply with the study standards.
5. Algorithms:
   5.1. Data transformation: a series of rules are applied to convert it to a specific format desired by the unit of analysis.
   5.2. Opinion classification: the previously identified words are taken to assign a polarity, that is, to define the associated sentiment according to the SVNN. To determine the polarity and therefore identify feelings or opinions, each word will be assigned its respective grammatical category (verb, adjective, adverb, name, etc.). Once done, the system will search the corpus and process only the polarity of the information collected.
6. Results display:
   For the output, visualization, and discussion of the results, the data will be clustered. This technique starts from a measure of proximity between individuals and, from there, they look for the groups of individuals most similar to each other, according to a series of measured variables [12, 42-49]. According to what was consulted, the K-mean method is very favorable for this, so its application is decided in this step.

However, to manage neutrality within the entire process, we added to this technique, a neutrosophic version according to [22, 42-45, 47-54]:

Definition 2: A partition $P = \{C_1, C_2, \dots, C_c\}$ is said to be a soft partition of data set X, if and only if it is true that: $\left(\forall x_i \in X, \forall C_j \in P\right) \leq \mu C_j(x_i) \leq 1$ and $\left(\forall x_i \in X, \exists C_j \in P\right)$such that $\mu C_j(x_i) > 0$. Where $\mu C_c(x_i)$ denotes the degree to which xi belongs to the cluster *Cj*

Definition 3: A Partition is a special soft partition when the sum of the degrees of membership of a specific point in all the clusters is equal to 1 as shown in equation 5.

$$\sum_j \quad \mu C_j(x_i) = 1, (\forall x_i \in X \tag{5}$$

Definition 4: A constrained soft partition is a partition that meets this additional condition. The Neutrosophic K-Means algorithm produces a constrained smooth partition and to do this the objective function J is extended in two ways: $\forall x_i \in X, \exists C_j \in P$ such that $\mu C_j(x_i) > 0$ where the degrees of neutrosophic membership of each data in each cluster are incorporated or; introducing an additional parameter that serves as exponent weight in the membership function, thus the extended objective function *Jm* is as shown in 6.

$$\mu c_1(x_1) = \frac{1}{\sum_{j=1}^{2} \quad \left[\frac{\|x_1 - v_1\|^2}{\|x_1 - v_j\|^2}\right]^2} \tag{6}$$

Where P is a fuzzy partition of the data set X formed by $\{C_1, C_2, \dots, C_k\}$ and the parameter m is a weight that determines the degree to which the partial members of a cluster affect the result. This refers to a similarity between the classical method and its neutrosophic extension since the latter also tries to find a good partition by searching for the prototypes vi in such a way that they minimize the objective function Jm and that in the same way, it must also search for the functions of membership μC_1 that minimize Jm. In addition to the method, equation 7 is established for the calculation of the initial membership functions of both clusters:

$$J_m(P, V) = \sum_{j=1}^{k} \quad \sum_{j_k \in X} \quad (\mu c_j(x_k))^m \|x_k - v_j\|^2 \tag{7}$$

The calculations are subsequently updated according to equation 8.

$$v_1 = \frac{\sum_{k=1}^{n} \quad (\mu C_1(x_k))^2 x_k}{\sum_{k=1}^{n} \quad (\mu C_1(x_k))^2} \tag{8}$$

3.3 Software for the processing of the method

Due to the benefits it offers, we chose Orange Software, which is a free software application for data mining and predictive analysis, under the GPL license. It is a powerful tool, but at the same time, it is friendly and intuitive and allows visual, fast and versatile programming for an exploratory data analysis. Developed in C ++, the Orange application is multiplatform, as Windows, Linux and Mac support it. It allows predictive modeling through classification trees, regression, logistics, Bayes classification, and association rules. In addition to data description methods, self-organizing maps, and clustering. Lastly, you have model validation and cross-validation techniques. It also has an easy, powerful, fast, and versatile visual programming component [4, 36].

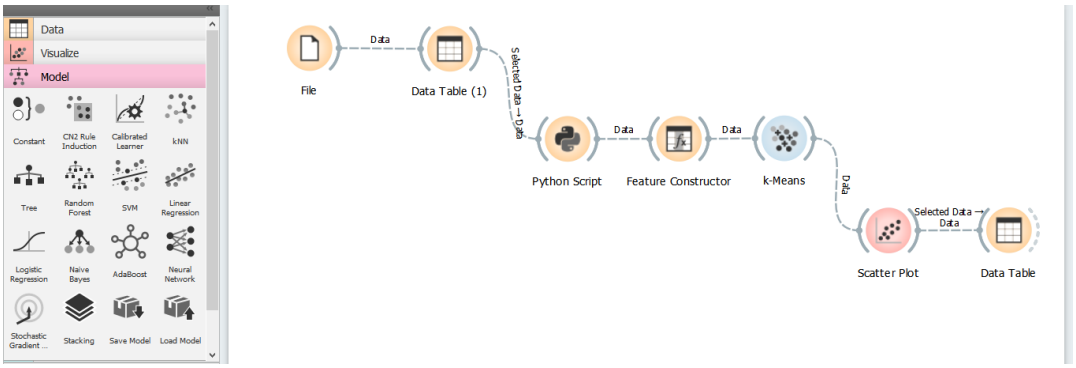

Figure 4: Workflow in Orange.

To start the process, the corpus is loaded through a file widget, then its polarity is processed by applying the SVNNs and the neutrosophic K-means. In the end, the results are shown in a Scatter Plot and a Data Table to better analyze the results: apply filters, count, sort, etc. Equation 4 is incorporated into the Feature Constructor for the deneutrosophication of SVNNs and equations 6, 7, and 8 through a Python Script.

4 Results

To illustrate the operation of the proposed model, a test experiment is carried out. Where it is proposed as a thematic field of interest to determine the legal opinions expressed by the students in a blog about the sentence in a judicial process of a femicide case. The data entry for the formation of the corpus was based on the comments made by said students on this particular topic. For the data transformation, words were chosen to form a dictionary where the neutrosophic numbers are assigned based on the polarity shown. Here is an example:

Corpus-polarity	Neutrosophic linguistic term	SVNN
Indication of total acceptance with the verdict: totally agree, very fair, exemplary	Extremely positive (EP)	(1,0,0)
Indication of an agreement but without showing empathy: Fair	Average (M)	(0.50,0.50,0.50)
Indication of revulsion with the verdict: appeal, unworthy, unfair	Extremely negative (EN)	(0,1,1)

Table 2: Association of the SVNN to the polarity of opinion.

Then the data is entered according to the flow that was programmed in Orange, obtaining the following:

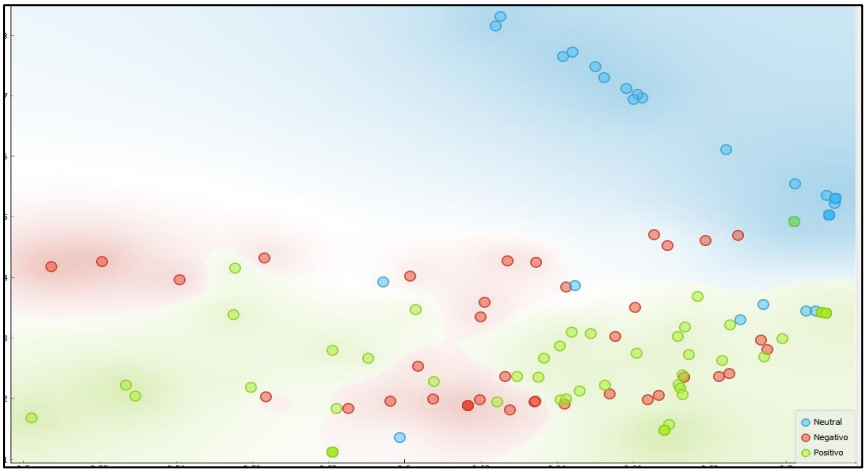

Figure 5: Results visualization. Source: Own elaboration (software output)

As could be observed for the experiment carried out, the neutral opinions show the highest level within the analyzed corpus, obtaining a higher prevalence of the positive ones over the negative ones. So it can be said that the general opinion among the students was rated as medium-positive. However, students may not have all the elements available to evaluate the process effectively. Therefore, it was decided to re-evaluate the opinions once all the elements of the judgment had been presented. They are listed below:

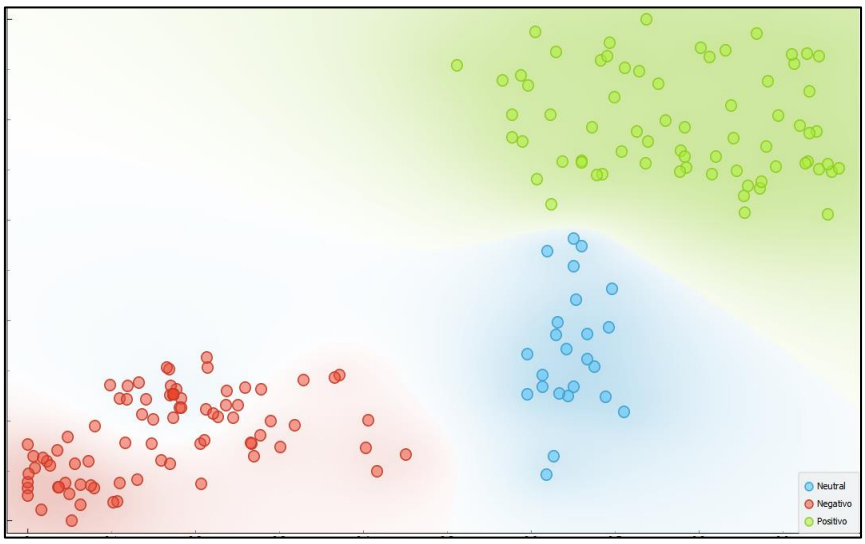

Figure 6: Second iterative round of the method. Source: Own elaboration (Orange software output).

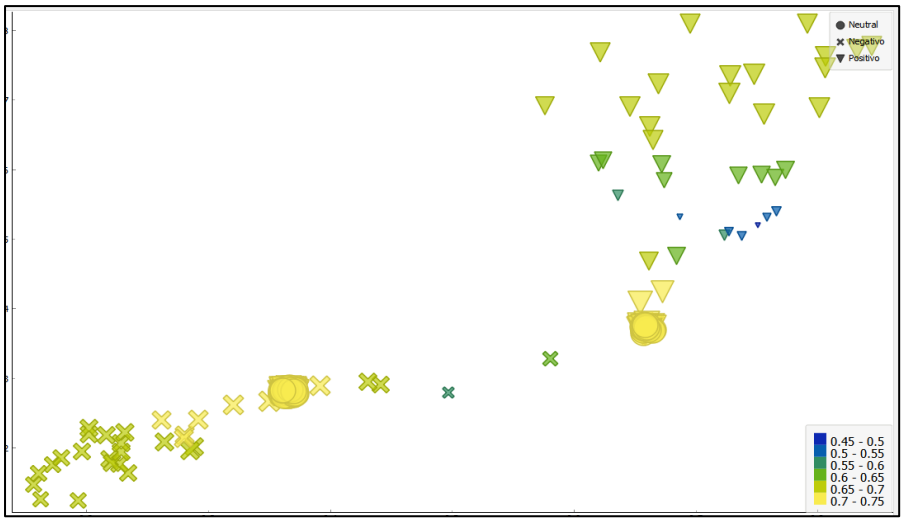

Figure 7: Second iterative round of the method. Source: Own elaboration (Orange software output).

It can be seen that the change of opinion is substantial. Although the experiment constitutes a test that could empirically determine and establish an "adequate" own judgment, it is necessary to have all the elements inherent to the case. However, each person depending on the affectation to the problem can develop more empathy than others and affect their point of view. Sensitive situation, so it is decided to expose the results to the students to obtain greater feedback on this subject in a legal debate for educational purposes.

Conclusion

It is worth remarking that in the consulted references, it could be observed that the authors are divided regarding the establishment of the similarities and differences of these terms. Some comment that there are differences since the first focuses on the narrative while the second on pure emotions. However, all agree on the help that both techniques provide for legal scientific work. Data mining is a discovery of knowledge that is carried out through the exploration and analysis of large volumes of data stored in the databases of the information unit, to discover significant correlations, new patterns and trends between the explored and analyzed data that facilitate decision-making to improve the services it provides.

The use of data mining processing tools fused with neutrosophy makes it possible to build very robust opinion mining systems. This will make it possible to carry out new sentiment or opinion analysis experiments focused on different domains of the social sciences, beyond legal scientific research. A project of this type is interdisciplinary and requires the assistance of specialists in the areas of Law and Data Mining as well as in the field of statistics. One of the complexities involved when obtaining an adequate model is understanding its internal functioning, hence the need for a combination of specialties. The achievement of synergy between the authors and collaborators allowed us to explore its components. Therefore, it can be said that this method is the result of a complex project based on the segmentation of knowledge to support decision-making, which must be properly interpreted.

For future lines of research, it would be interesting to identify the profiles of those who make the comments to combine them with other broader data such as gender, age range, false profiles, patterns in the messages, etc. It is necessary to combine them with a semantic analysis of the texts to extract more information, determining words and in which contexts they appear. It is suggested to carry out new studies that, with the use of these models contrasted with the knowledge of experts, allow to help strengthen and consolidate social research. As well as it is considered that it is of utmost importance to obtain feedback from the model to perfect it in search of continuous improvement and thus achieve an established algorithm within legal scientific research.

References

[1] J. Estupiñán Ricardo, V. M. Vaca Rosado., J. Piedra Fernández, and S. Mantilla Martínez. (2020) Importancia de la investigación jurídica para la formación de los profesionales del Derecho en Ecuador. *Revista Dilemas Contemporáneos: Educación, Política y Valores.* 1-25. Available: http://www.dilemascontemporaneoseducacionpoliticayvalores.com/

[2] E. Lace, "AI & Law: Legal Sentiment Analysis and Opinion Mining," *Medium*, 2020.

[3] C. J. Martínez Moncaleano and O. Palencia Fajardo, "Modelo de minería de datos para el análisis de la productividad y crecimiento personal en las mujeres emprendedoras: el caso de la Asociación las Rosas," *Suma de Negocios*, vol. 12, pp. 23-30, 2021.

[4] M. Fernandez-Morales and R. Bonilla-Carrión, "Bibliominería, datos y el proceso de toma de decisiones," *Revista Interamericana de Bibliotecología*, vol. 43, 2020.
[5] F. Medina-Quispe, W. Castillo-Rojas, and C. Meneses Villegas, "Métricas para el apoyo de la exploración visual de componentes en modelos de minería de datos," *Ingeniare. Revista chilena de ingeniería*, vol. 28, pp. 596-611, 2020.
[6] J. L. Rivero Pérez, Y. Peñate Santana, and P. H. Martínez López. (2016) Propuesta de plataforma de procesamiento de datos para Marketing directo. . *Revista Universidad y Sociedad*. 65-71. Available: http://rus.ucf.edu.cu/
[7] J. J. Martínez "Minería de opiniones mediante análisis de sentimientos y extracción de conceptos en Twitter," Máster en Ingeniería Informática, Facultad De Informática. Departamento de Sistemas Informáticos y Computación, Universidad Complutense de Madrid, Madrid, España, 2017.
[8] C. Busón. (2020) La minería de opinión para el análisis del discurso de odio en las redes sociales. Un estudio de caso sobre Paulo Freire en YouTube durante el periodo 2007-2019. . *Commons. Revista de Comunicación y Ciudadanía Digital*. 119-159.
[9] C. H. Miranda, J. Guzmán, and D. Salcedo. (2016) Minería de Opiniones basado en la adaptación al español de ANEW sobre opiniones acerca de hoteles. *Procesamiento de Lenguaje Natural*. 25-32.
[10] J. G. Conrad and F. Schilder. (2007) Opinion Mining in Legal Blogs. *Frank* 231-236.
[11] M. Leyva-Vázquez and F. Smarandache, *Inteligencia Artificial: retos, perspectivas y papel de la Neutrosofía*: Infinite Study, 2018.
[12] F. J. Cedeño Moran, M. P. Marcillo Sánchez, L. D. Roman Barrezueta, and C. N. Martillo Avilés. (2019) K- Means Neutrosófico para la segmentación de mercado. *Neutrosophic Computing and Machine Learning*.
[13] M. Leyva Vázquez and F. Smarandache, "Neutrosofía: Nuevos avances en el tratamiento de la incertidumbre," *Pons, Bruselas. ,* 2018.
[14] M. A. Q. Martinez, H. F. V. Espinoza, M. Y. L. Vazquez, and M. D. G. Rios, "Feasibility Analysis Proposal for an IoT Infrastructure for the Efficient Processing of Data in Agriculture, Case Study on Cocoa," *Revista Ibérica de Sistemas e Tecnologias de Informação*, pp. 413-426, 2020.
[15] M. Q. Martinez, A. B. Quirumbay, and M. L. Vazquez, "ESTUDIO CUALITATIVO DE RECONOCIMIENTO DE EMOCIONES EN TIEMPO REAL PARA ATENCIÓN AL CLIENTE UTILIZANDO DEEPLENS FACE DETECTION," *Investigación Operacional*, vol. 42, pp. 63-73, 2021.
[16] A. S. Molina, W. A. C. Calle, and J. D. B. Remache, "The application of Microsoft Solution Framework Software Testing using Neutrosophic Numbers," *Neutrosophic Sets and Systems*, vol. 37, pp. 267-276, 2020.
[17] A. L. S. Pillajo, M. A. C. Cabrera, R. A. D. Vásquez, and J. L. A. Espinoza, "Sistema experto para el diagnóstico y tratamiento de enfermedades y plagas en plantas ornamentales," *Universidad y Sociedad*, vol. 13, pp. 505-511, 2021.
[18] M. Tello Cadena, M. F. Jaramillo Burgos, and M. J. Jiménez Montenegro, "Neutrosophic Case-Based Reasoning Method to Determine the Profitability of the Tourism Sector in the City of Riobamba," *Neutrosophic Sets and Systems*, vol. 37, pp. 108-114, 2020.
[19] M. Y. L. Vazquez, "An Analysis of Deep Learning Architectures for Cancer Diagnosis," *Artificial Intelligence, Computer and Software Engineering Advances: Proceedings of the CIT 2020 Volume 1*, vol. 1326, p. 19, 2021.
[20] M. Y. L. Vazquez, B. S. M. Arteaga, J. A. M. López, and M. A. Q. Martinez, "Design of an IoT Architecture in Medical Environments for the Treatment of Hypertensive Patients," *Revista Ibérica de Sistemas e Tecnologias de Informação*, pp. 188-200, 2020.
[21] K. Mejía González, "Enriquecimiento del modelo basado en reglas VADER a través de lexicones," *Universidad Autónoma Del Estado De México Unidad Académica Profesional Tianguistenco. Ingeniería En Software* 2018.
[22] Q. Li, Y. Ma, and S. Zhu, "Single valued Neutrosophic clustering algorithm Based on Tsallis Entropy Maximization," *MDPI*, 2018.
[23] F. Samarandache, "Introduction to Neutrosophic Statistics," *Sitech & Education Publishing* 2014.
[24] F. Smarandache, "A Unifying Field in Logics: Neutrosophic Logic. Neutrosophy, Neutrosophic Set, Neutrosophic Probability: Neutrosophic Logic. Neutrosophy, Neutrosophic Set, Neutrosophic Probability: Infinite Study.," 2005.
[25] J. L. a. F. S. Salmerona, "Redesigning Decision Matrix Method with an indeterminacy-based inference process. Multispace and Multistructure. ," *Neutrosophic Transdisciplinarity (100 Collected Papers of Sciences)*, vol. 4, p. 151, 2010.
[26] M. Ali, Shabir, M., Smarandache, F., and Vladareanu, L., "Neutrosophic LA-semigroup Rings," *Neutrosophic Sets and Systems*, vol. 7, pp. 81-88, 2015.
[27] W. B. Vasantha, I. Kandasamy, and F. Smarandache, "Algebraic Structure of Neutrosophic Duplets in Neutrosophic Rings < Z U I>,< Q U I> and < R U I.> " *Neutrosophic Sets and Systems, ,* vol. 23, pp. 85-95, 2018.
[28] F. Smarandache, *Neutrosophy, a new Branch of Philosophy*: Infinite Study, 2002.
[29] A. Sleem, M. Abdel-Baset, and I. El-henawy, "PyIVNS: A python-based tool for Interval-valued neutrosophic operations and normalization," *SoftwareX*, 2020.
[30] F. Smarandache, M. A. Quiroz-Martínez, J. E. Ricardo, N. B. Hernández, and M. Y. L. Vázquez, "Application of neutrosophic offsets for digital image processing," *Investig. Operacional*, 2020.
[31] M. L. Vázquez, J. Estupiñan, and F. Smarandache, "Neutrosofía en Latinoamérica, avances y perspectivas," *Revista Asociación Latinoamericana de Ciencias Neutrosóficas. ISSN 2574-1101*, vol. 14, pp. 01-08, 2020.
[32] W. V. Kandasamy and F. Smarandache, "Fuzzy Neutrosophic Models for Social Scientists.," *Education Publisher Inc.*, (2013)
[33] W. B. V. a. F. S. Kandasamy, "Fuzzy cognitive maps and neutrosophic cognitive maps. ," *American Research Press.*, 2003.

[34] R. Şahin and M. Yiğider, "A Multi-criteria neutrosophic group decision making method based TOPSIS for supplier selection," *arXiv preprint arXiv:1412.5077.* , 2014.
[35] M. Leyva-Vázquez, M. A. Quiroz-Martínez, Y. Portilla-Castell, J. R. Hechavarría-Hernández, and E. González-Caballero, "A New Model for the Selection of Information Technology Project in a Neutrosophic Environment," *Neutrosophic Sets Syst., p. 344*.
[36] J. Demšar, "Orange: Data mining toolbox in Python," *J. Mach. Learn. Res*, 2013.
[37] H. Cho, J. Lee, and S. Kim, "Enhancing Lexicon-Based Review Classification by Merging and Revising Sentiment Dictionaries," *Technology, Yonsei University*, 2018.
[38] P. Gonçalves, M. Araújo, F. Benevenuto, and M. Cha, "Comparing and Combining Sentiment Analysis Methods," *UFMG*, 2018.
[39] A. Glebukh, "Procesamiento del lenguaje natural," *Sociedad Mexicana de Inteligencia Artificial* 2018.
[40] M. Alberich, "Procesamiento del Lenguaje Natural Guía Introductoria," 2018.
[41] C. E. Herrera Cucaita and J. R. Noreña Angel, "Implementación de un módulo de gestión de proyectos para el sistema AS/400," 2016.
[42] M. Rizwan Khan. (2018). *K Means Clustering Algorithm & its Application*. Available: https://medium.com/datadriveninvestor
[43] Z. E. Mamani Rodríguez, "Aplicación de la minería de datos distribuida usando algoritmo de clustering k-means para mejorar la calidad de servicios de las organizaciones modernas," Facultad de Ciencias Matemáticas, Unniversidad Nacional Mayor de San Marcos, 2015.
[44] M. A. Naoui, B. Lejdel, and M. Ayad, "Using K-means algorithm for regression curve in big data system for business environment," *Revista Cubana de Ciencias Informáticas*, vol. 14, pp. 34-48, 2020.
[45] B. B. Yousif, M. M. Ata, N. Fawzy, and M. Obaya, "Toward an optimized neutrosophic K-means with genetic algorithm for automatic vehicle license plate recognition (ONKM-AVLPR)," *IEEE Access*, vol. 8, pp. 49285-49312, 2020.
[46] L. Morissette and S. Chartier, "The k-means clustering technique: General considerations and implementation in Mathematica. ," *Tutorials in Quantitative Methods for Psychology.*, vol. 9, 2013.
[47] D. T. Pham, S. S. Dimov, and C. D. Nguyen, "Selection of K in K-means clustering," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 219, pp. 103-119, 2005.
[48] I. Dabbura, "K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks," 2018.
[49] J. Burkardt, "K-means clustering," *Virginia Tech, Advanced Research Computing, Interdisciplinary Center for Applied Mathematics*, 2009.
[50] E. Florido, F. Martínez-Álvarez, and J. Aznarte, *Metodología basada en minería de datos para el descubrimiento de patrones precursores de terremotos de magnitud media y elevada*, 2015.
[51] M. A. I. Salame Ortiz, B. C. Pérez Mayorga, D. E. Huera Castro, and T. S. Viteri Paredes, "Las técnicas de reproducción asistida en el Ecuador ¿hecho jurídico o acto jurídico?," *Uniandes EPISTEME. Revista digital de Ciencia, Tecnología e Innovación*, vol. 5, 2018.
[52] S. F. Chamba Jiménez, "Minería de Datos para segmentación de clientes en la empresa tecnológica Master PC," Unniversidad de Loja, 2015.
[53] J. C. Córdova Galarza, "Aplicación de técnicas de minería de datos para predecir la deserción de los estudiantes que pertenecen al colegio fisco-misional "San Francisco" de la ciudad de Ibarra"," Facultad de Sistemas Mercantiles, Universidad Regional Autónoma de Los Andes "UNIANDES", 2014.
[54] A. S. Ashour, A. R. Hawas, Y. Guo, and M. A. Wahba, "A novel optimized neutrosophic k-means using genetic algorithm for skin lesion detection in dermoscopy images," *Signal, Image and Video Processing*, vol. 12, pp. 1311-1318, 2018.

Received: March 14, 2021. Accepted**:** May 9, 2021