v. 3, n. 35, 2023

International Journal of Human Sciences Research

ANALYSIS OF FEELINGS IN EDUCATIONAL ENVIRONMENTS

Humberto J. Centurión Cardeña
Tecnológico Nacional de México, Campus Motul, Carretera Mérida-Motul

Danice D. Cano Barrón
Tecnológico Nacional de México, Campus Motul, Carretera Mérida-Motul

José L. Tamayo Canul
Tecnológico Nacional de México, Campus Motul, Carretera Mérida-Motul

Abstract: Social network popularity has made easy sharing ideas in a variety of settings, whether professional, educational or personal. However, when the amount of information exchanged grows, so does the complexity of following up on occurrences. This pilot project for Twitter seeks to identify the polarity of a comment using data from a test account that was monitor for a couple of months at a time. The NLTK library was used for human language analysis, which provides easy-to-use interfaces and lexical resources to classify, label, parse language elements, and perform semantic reasoning. The process included the creation of various word clouds from the original database, the identification of the most relevant words. Finally, the quality of the word selected as relevant as well as their combination were evaluated using a neural network to determine the polarization.
Keywords: Social networks / word cloud/ neural network / higher education / Twitter

INTRODUCTION

Social problems are those situations that collectively afflict the citizens of some region and that have various causes as their origin. These, according to Pérez (2017), generally have historical motivations that over the years have shaped society and have become problems that temporarily affect society. The common factors that social problems have is that they are difficult to overcome, because they are part of a social fabric that in many cases lives adapted to them.

The use of social networks, according to Del Valle (2017), is a phenomenon that has notable acceptance worldwide, so it is not surprising that young people are the ones who feel most attracted to these popular media, especially, because through these platforms they can easily communicate with their friends and family. The problem is that teenagers do not always use technology appropriately as part of a beneficial benefit, such as their education, but rather as entertainment.

The university identity is conceived as the set of cultural dimensions shared by the university community, with diversity and multiculturalism being the most important factors in the process of formation of the institutional identity, which is of utmost importance for the development of the person. within the formation of their academic identity (Almazán, 2014). Promoting this sense of belonging in educational institutions facilitates the permanence of active students and maintains a constant relationship with graduates, which fosters ties of collaboration and exchange. This sense of belonging is one of the most complex tasks of educational organizations, especially when they are young and are trying to make a name and prestige of their own.

This project is a first initiative to analyze the institution's social networks as a mechanism to strengthen the academic identity of its members, which consists of using neural networks to determine whether the comment issued is positive or negative, so that when it begins By working with the interactions between network members, the meaning of said comments can be determined with a certain level of confidence.

JUSTIFICATION

The popularity of social sites such as Facebook and Instagram have become a space with great global attendance; it cannot be denied that both are so popular that every day more people join cyberspace to be part of the technological era and thus share experiences. with their loved ones (Del valle, 2017).

Facebook was born as a space aimed exclusively at Harvard University students to get to know each other, later students from other institutions joined this platform and finally it was not only young university students but an entire planet; It is estimated that it has more than 1.35 billion active users worldwide, in addition to having 110 languages available (Molina, 2021).

In the same way, Instagram emerged and little by little it gained popularity, so much so that two years after its launch it had more than 100 million users, something unimaginable despite being "new" within social networks, it was created to exclusively share photos. and videos, but over time it has developed updates that attract the attention of users (Sánchez, 2013).

In the face of social networks like Facebook or Twitter, Instagram acted as a younger sister, but its potential is being deployed more and more. According to a study, on Instagram the average user follows an average of 134 people and 76% of users connect at least once a day. This community is very active: 85% of them publish more than two photos per week (Regil, 2014).

Among the mass of 150 million users, influential users (power users in English) rise, those who manage to group people around their personal brand. Thus, 33% of influential users have more than 100,000 followers on their account, which allows them to be attractive to companies that want to enhance their products (Grimaldi, 2017). Understanding social dynamics would support the academic and personal development of students, as well as manage conflicts efficiently.

THEORETICAL FRAMEWORK

Text mining is a specific branch of data mining that is related to information capture systems, machine learning, statistics and computational linguistics and attempts to obtain knowledge from unstructured text from emails, electronic documents., web pages and social media. It is estimated that 80% of an organization's information is stored in the form of documents. Without a doubt, this field of study is very broad, so techniques such as text categorization, natural language processing, information extraction and retrieval or machine learning, among others, support the analysis of social networks (Rayón, 2015).

According to Ballester (2012), the systems that support decision making stand out for their great benefits to data warehouses (data warehouses) and data markets (data marts) as they are known in their departmental version. These systems, well structured within Artificial Intelligence, convert the operational data of an organization into knowledge, they are a competitive tool that allows the data to be examined in a more strategic way, perform analysis and detect trends, model various scenarios, track critical measures, produce reports, reports and presentations more quickly and with easy, flexible and intuitive access to the information needed at all times. They also allow data mining and knowledge discovery tasks to be carried out in databases (KDD).

Data analysis has evolved as large volumes of data grow. As Joyanes (2013) mentions, business intelligence tools have been collecting the technologies of online analytical processing, reports and queries, visualization and, especially, data mining and innovative social mining in media data analysis. social that has been supported by sentiment and opinion analysis techniques, or opinion mining and sentiment mining as it is also known.

The avalanche of data that is generated, captured, stored and analyzed every day in organizations and companies and, therefore, individuals, has given rise to the new Big Data trend. For example, the millions of users who visit Facebook, the millions of tweets that are published daily, the millions of messages and conversations that take place through WhatsApp, in addition to the fact that there are more than 2.8 billion Internet users in the world. If this were not enough, we must add the data that is transferred between the billions of objects or things that communicate with each other. This explosion of large volumes of data does not stop growing and it seems exponentially.

One of the problems of recently created schools is the development of institutional identity and the students' sense of belonging. Many of them find the process of making students feel proud of their alma mater difficult, in addition to involving them in the activities designed for them is a challenge, not to mention that the feedback process does not always reveal the experiences and aspects positive or negative of the experience.

Social interactions in digital environments based on social networks are a source of countless information about the relationships that exist between the various actors in a school organization. Being able to characterize the communities and their main actors will allow us to act more efficiently to solve their needs and areas of opportunity.

Generating a formal analysis program of relationships through social networks will allow not only to understand the relationships between students but also the level of influence that certain people have in the community and to be able to work with them on positive leadership issues. Likewise, being able to measure the impact that various training campaigns can have on the community would be interesting, since one can learn to carry them out more efficiently.

STATE OF THE ART

In this sense, works such as Acevedo, Clorio, Zagal & García (2014) that propose an approach that integrates a web application, a semantic hierarchy based on WordNet - Affect Hierachy and a Naive Bayes classifier that identifies publications in the emotions joy, sadness and anger seek to identify beyond the positive or negative feeling. Five tests were carried out with a sample of 802 publications, different strategies were combined until reaching the definitive approach: using base classifier without emoticon processing, using Naïve Bayes classifier in filtering publications, classifier based on hierarchy of concepts based on WNA weighting, trained with selected publications achieves a performance of 63% correct classification.

Rosá, Chiruzzo, Etcheverry & Castro (2017) used classifiers based on SVM (Support Vector Machine) and Convolutional Neural Networks (CNN). The SVM model evaluated the polarity of a tweet, an SVM classifier trained with the various attributes was built. Regarding the CNN model, they considered two structural variants of neural network models: a single layer and another with several layers that receive the input independently and their outputs are concatenated. The results for the experiments carried out show that the hybrid model was the one that obtained the best results with 64.7% correctness.

In both cases, the aim is to identify the sentiment behind the tweet as a mechanism to determine its polarity and its subsequent analysis. In the case of this project, it is important to be able to make that distinction to be able to determine if the person refers to a positive feeling in some type of survey or open opinion and to be able to determine their level of integration into the educational entity to which they belong.

MATERIALS AND METHODS

The process of knowledge discovery in databases involves nine phases (Velarde, 2003), as can be seen in Fig. 1. A first phase begins with the selection of a set of data and focusing on the search in subsets of variables. and/or data samples where to carry out the discovery process. Once the selection is established, the data is cleaned and preprocessed, designing an appropriate strategy to handle noise, incomplete values, time sequences, and others. It concludes with the reduction of data and projections to reduce the number of variables to be considered.

Once the variables have been selected and cleaned of missing values, the task of identifying the discovery technique to use, for example, classification, grouping, regression, etc., continues, together with the selection of the algorithms to use. When you have these, you can carry out the data mining process and interpret the results.

These two important blocks of treatment and analysis are possibly repeated with other data, other algorithms, other goals and other strategies, to determine if they are the most efficient or if it is possible to improve the selection of variables.

Finally the information is incorporated into the system (usually to improve it), which may include resolving potential conflicts with existing knowledge.

Although fundamentally all social networks work in a similar way, technologically speaking, the interesting thing about each of them is the openness to access user information. In this sense, each platform has different levels of access. In Table 1 you can see a list of the main social networks and their access permissions to information, so one of the first challenges to overcome is the opening of permissions and respect for the right of users to provide their information. information for the use of the tool to be developed.

Social network	Authorizations
Facebook	Limited access, the app must be developed first to be directly approved by its team of experts
Instagram	It has two versions, one for novice users (who have not developed apps) and an expert when there is an app developed and approved by them.
Twitter	Platform open to capturing information from users and their interactions with others

Table 1. Permissions of the most common social networks

Due to the previous context, to carry out the preliminary tests it was decided to work with a Twitter database due to the openness that the API allows developers to obtain raw information to be analyzed because this is a first approach to the development of the tool, in later phases the social networks most used by the ITS Motul academic community must be considered.

A Twitter test account was monitored for a couple of months to obtain the information that was produced during that period of time, the result can be seen in Figure 2. The following fields were considered in the database: an identifier of the tweet (ID), the user it came from (the user name was protected), the date and time it was made, the text, the longitude and latitude of the tweet, whether the direction of the tweet was positive or negative and language. The experimental database contained more than 37,000 records.

The NLTK library was used for language analysis, built for Python to work with human language data and which provides easy-to-use interfaces and lexical resources to classify, tokenize texts, label, parse, and perform semantic reasoning between other things (NLTK Project, 2018). This library allowed not only word-by-word analysis, but also the exclusion of a series of common words that do not contribute to the analysis of the themes (Stop_Words), to which punctuation marks and prepositions were added.

ID	User	Date Tweet	Text	Latitude	Longitude	Feeling	Language
1.02E+18		2018-07-23T04:59:54.	Â¿Donde qu	21.22812	-89.6748995	Negative	Spanish
1.02E+18		2018-07-23T04:59:51.	@robbscr Ne	20.941035	-89.623611	Positive	Norwegian
1.02E+18		2018-07-23T04:59:30.	@Nessasaur	21.22812	-89.6748995	Positive	English
1.02E+18		2018-07-23T04:58:28.	Yo todos los	20.941035	-89.623611	Positive	Spanish

Figure 2. Database example

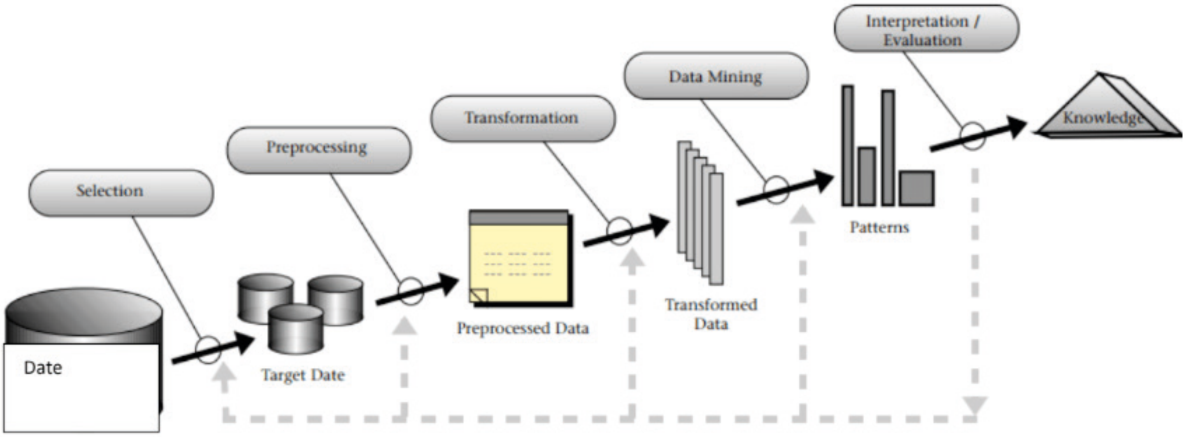


Figure 1. KDD Methodology

(a) (b)

(c) (d)

Figure 3. Example of the word clouds generated with the database, from the least filtered (a) to the most worked (d)

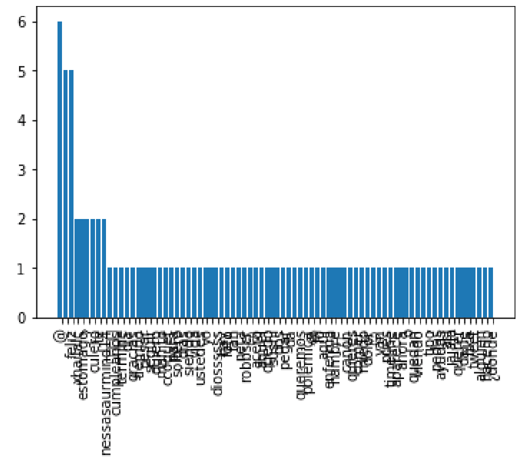
Once the filters were made on the words, three tools were used to determine the most common words in the text and to determine the themes on which we would work. The first is a word cloud generator, which allows you to identify the main words that are identified in the reviewed text (Mueller, 2013). The second was a histogram that displayed the 30 most repeated words in the text. This helped to understand not only the themes but also the possible combinations of words to use in the neural network. Finally, the third tool was a word counter to determine the most repeated words within the texts of the tweets.

These tools allowed us to have a first approach to the contents of the talks and identify the main themes. In order to establish the possible relationships between the words and whether the meaning of the comment was positive or negative, a neural network was run that allowed us to establish the general meaning of the comment.

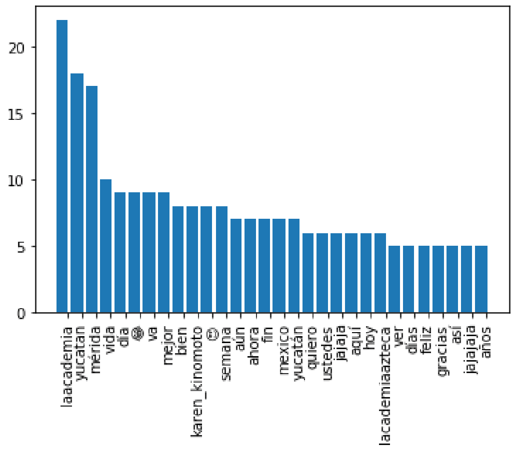
RESULTS AND DISCUSSION

In Figure 3 you can see the various word clouds generated from the original database using the text field. Figure 3 (a) represents the word cloud before filtering words and Fig. 3 (d) represents the cloud once all the words included in Stop_Words have been filtered, including the punctuation marks and prepositions that were identified as repetitive and that caused noise in the analysis.

In Figure 4 you can see two histograms, the first (a) is an initial histogram before word filtering and (b) is the result of delimiting the words to 30 to facilitate their identification. In addition to identifying the most repeated words, it was used to identify punctuation marks or words that were unnecessary to determine the meaning of the comments.



(a)



(b)

Figure 4. Histogram of most repeated words in comments

In Figure 5 you can see the list of the 30 most repeated words and the frequency with which they were found in the analyzed tweets. One of the main elements that this tool facilitated was identifying the different spelling errors that prevent a quick reading of the topics found in the tweet. For example "yucatan" and "yucatán", or "q". Likewise, expressions were identified that were identified as different but refer to the same thing, such as "hahahaha" and "hahaha." These elements must be considered in future applications to better clean the startup data.

[('mérida', 2265), ('yucatan', 1591), ('yucatán', 1243), ('today', 1147), ('I want', 1016), ('day', 1009), ('better', 981), ('life', 965), ('being', 946), ('like this', 941), ('good', 917), ('just', 878), ('thank you' ', 866), ('see', 848), ('mexico', 840), ('yuc', 828), ('always', 805), ('mexico', 781), ('hahaha', 725), ('now', 693), ('you', 677), ('home', 670), ('days', 613), ('going', 597), ('ago', 590), ('hahahaha', 560), ('here', 549), ('never', 538), ('do', 536), ('time', 524), ('years', 521), ('😍', 514), ('merida', 509), ('I'm going', 491), ('people', 474), ('things', 463), ('someone', 456), ('2 ', 443), ('lopezobrador_', 443), ('happy', 435), ('world', 435), ('tomorrow', 432), ('for', 416), ('love', 415), ('master', 413), ('can', 412), ('q', 412), ('time', 403), ('every', 401), ('can', 398)]

Figure 5. List of repeated words

Finally, this first step of pre-processing the information from the comments allowed us to identify a first set of words to use in the neural network scheme, with three being the initial selection: happy, life and love. With them, a database was generated that includes the original text of the tweet and, if it was considered positive or negative, the file.

Using the filtered data containing the keywords, it was used to generate a matrix that establishes the presence and absence of the keywords, as well as whether the tweet was considered positive or negative (see Figure 6). This file served as input to determine the combinations to train the neural network and based on the missing ones to test the effectiveness of the model. Although three words were used, it is intended to use more words to generate the training model.

happy	life	love	edo
0	1	0	1
1	0	0	1
0	1	0	1
0	1	0	1
0	1	0	1
0	1	0	1
0	1	0	1
0	1	0	1
0	1	0	1
0	1	0	1
0	0	1	1

Figure 6. Generated coding for the neural network

In Figure 7 you can see the output produced by the neural network, which not only concludes with the state of the tweet, but also coincides with the state established in the original file. At the end, the percentage of coincidence of all cases is added to identify the degree of effectiveness of the generated model, so that by adding or eliminating keywords the degree of coincidence can be identified and whether there is improvement or not when making the modifications.

Tweet status is positive ([
0.61756855]) do not match
Tweet status is positive ([0.61756855]) match
% match: 0.6154276131917272

Figure 7. Neural network output

CONCLUSION

The nature of each social network is different and not all of them generate the same type of information or in the same amount since while some like Facebook or Instagram base a large part of their content on photographs, others like Twitter or LinkedIn offer much more textual information. Regardless of the source of information, the valuable content we can extract from social networks to generate knowledge and the new emerging characteristics that data presents is posing unprecedented challenges and great opportunities for information analysis.

Although social networks can be or sound addictive, it must be noted that the essential factor in their operation is what they can mean for each individual, so we must not prejudice this means of communication but rather well teach that there are networks that are fun and highly functional for their training and academic and professional development.

It is important to mention that research in the analysis of social networks is one of the sources of greatest potential information to be processed and analyzed.

REFERENCES

Acevedo, C.; Clorio, R.; Zagal, R. & García, C. (2014). Arquitectura Web para análisis de sentimientos en Facebook con un enfoque semántico. *Research in computer Science*, 75, pp. 59 – 69.

Almazán, T. (2014). Escala de identidad Universitaria. Facultad de psicología, UNAM. Consultado el 24 de septiembre de 2018. Disponible en: http://www.academia.edu/21037961/Escala_de_Identidad_Universitaria

Álvarez, T. (2014). La adicción a Instagram. Websa100. Consultado el 17 de septiembre de 2018. Disponible en: https://www.websa100.com/blog/la-adiccion-a-instagram-en-una-infografia/

Del valle, J. (2017). Análisis del impacto del uso excesivo de las redes sociales Facebook e Instagram en el rendimiento académico de los estudiantes de la unidad educativa fiscal José Martínez Queirolo. Universidad de Guayaquil. Consultado el 20 de septiembre de 2018. Disponible: http://repositorio.ug.edu.ec/bitstream/redug/19386/1/empaste%20jeanina.pdf

Grimaldi, S. (2017). Sociología de Instagram. Consultado el 20 de septiembre de 2018. Disponible: https://noeotech.com/2017/01/22/sociologia-de-instagram/

Herrero, S. (2015). Adicción al Smartphone, análisis motivacional de uso entre nativos digitales. Opción, Año 31, No. Especial 4, 517-531, ISSN 1012-1587.

Hun, S., Nasridinov, A. y Ho, K. (2018). Information Flow Monitoring System. *IEEE Access*, 6, pp. 23820 – 23827.

Ministerio de educación, Cultura y Deporte. (2017). Educación bilingüe: tendencias educativas y conceptos claves. Centro Nacional de Innovación e investigación Educativa. Pág. 16 y 17.

Molina, D. (28 de Octubre de 2021). *Historia de Facebook: nacimiento y evolución de la red social de Mark Zuckerberg*. Obtenido de IeBS: https://www.iebschool.com/blog/auge-y-declive-de-un-imperio-llamado-facebook-redes-sociales/

Nawas, S. (2016). Creating a simple REST API with Slim Framework. Consultado el 22 de septiembre de 2018. Disponible en: https://www.cloudways.com/blog/simple-rest-api-with-slim-micro-framework/

Pérez, M. (2017). Los 15 problemas sociales de México actuales más graves. Consultado el 21 de septiembre de 2018. Disponible en: https://www.lifeder.com/problemas-sociales-mexico/

Rayón, A. (2015). Análisis de redes sociales: el poder de la teoría de grafos. Consultado el 16 de septiembre de 2018. Disponible en: https://blogs.deusto.es/bigdata/analisis-de-redes-sociales-el-poder-de-la-teoria-de-grafos/

Regil, L. (2014). *Cultura Digital Universitaria*. Barcelona: Universidad Autónoma de Barcelona.

Rochina, P. (2017). El análisis de redes sociales mediante la teoría de grafos. Consultado el 21 de septiembre de 2018. Disponible en: https://revistadigital.inesem.es/informatica-y-tics/teoria-grafos/

Rosá, A.; Chirruzo, L.; Etcheverry, M. & Castro, S. (2017). Análisis de sentimiento de tweets en español utilizando SVM y CNN. Consultado el 5 de noviembre de 2018. Disponible en: http://www.sepln.org/workshops/tass/2017/

Sánchez, E. (2013). Análisis del impacto de las redes sociales en el estilo de vida en los jóvenes entre 15 y 24 años de edad en México. Texcoco: Universidad Autónoma del Estado de México.