Sentiment analysis through twitter as a mechanism for assessing university satisfaction

Omar Chamorro-Atalaya[1], Dora Arce-Santillan[1], Guillermo Morales-Romero[2],
Primitiva Ramos-Salazar[3], César León-Velarde[4], Elizabeth Auqui-Ramos[5], Miguel Levano-Stella[6]
[1]Facultad de Ingeniería y Gestión, Universidad Nacional Tecnológica de Lima Sur, Lima, Perú
[2]Facultad de Ciencias, Universidad Nacional de Educación Enrique Guzmán y Valle, Lima, Perú
[3]Facultad de Tecnología, National University of Education Enrique Guzmán y Valle, Lima, Perú
[4]Departamento de Humanidades, Universidad Tecnológica del Perú, Lima, Perú
[5]Departamento de Ciencias, Universidad Peruana de Ciencias Aplicadas, Lima, Perú
[6]Facultad de Ingeniería, Universidad Privada del Norte, Lima, Perú

Article Info

ABSTRACT

Currently, the data generated in the university environment related to the perception of satisfaction is generated through surveys with categorical response questions defined on a Likert scale, with factors already defined to be evaluated, applied once per academic semester, which generates very biased information. This leads us to wonder why this survey is applied only once and why it only asks about some factors. The objective of the article is to demonstrate the feasibility of a proposal to determine the degree of perception of student satisfaction through the use of data science and natural language processing (NLP), supported by the social network twitter, as an element of data collection. As a result of the application of this proposal based on data science, it was possible to determine the level of student satisfaction, being 57.27%, through sentiment analysis using the Python library "NLTK"; Thus, it was also possible to extract texts linked to the relevant factors of teaching performance to achieve student satisfaction, through the term frequency and inverse document frequency (TF-IDF) approach, these being those linked to the use of tools of simulation in the virtual learning process.

Corresponding Author:

Omar Chamorro-Atalaya
Facultad de Ingeniería y Gestión, Universidad Nacional Tecnológica de Lima Sur
Lima, Perú
Email: ochamorro@untels.edu.pe

1. INTRODUCTION

In recent years, technological advances have allowed the generation and access to large volumes of data in different online environments, one of the most relevant being the social networks that include Facebook, Twitter or WhatsApp due to this the need for new paradigms arises that allow a better understanding of the dynamics of said transformations both of the ontological and of the epistemological [1], [2]. In this context the discipline of global studies originates with the purpose of better understanding the impact that globalization trends have on the different facets of our environment [3]. From this perspective, the scope and limits of new methodological tools based on data science and their possible impact on the discipline of global studies are explored [4]. Indeed, one of the most important transformations today is the exponential growth of data that opens the possibility of new ways to analyze reality [5]. Thus, the value of information no longer resides in concrete data, but in the way massive data is correlated to discover hidden

patterns, trends and associations [6]. Until now, the analysis of big data has been used in different disciplines of scientific research whose purpose has been to solve complex problems, in the fields of environment, education, health, transportation, national security and biomedicine [7].

Data science is made up of three areas, big data, which is used to process data, data mining, whose purpose is to find patterns and the visualization of data, whose purpose is to facilitate the understanding of information in a clear way and promote their socialization [8]. With regard to data mining, it is considered as the analysis of observed data sets, generally of large volume, with the aim of finding new relationships between variables, as well as the correct summary of said data sets in an understandable and comprehensive way useful [9]. For this reason, data mining is considered a fundamental step in the discovery of information patterns [10]. As indicated, data mining encompasses a series of modeling techniques oriented to different requirements of organizations, with respect to the educational sector, the use of these techniques can be seen in a series of works aimed at improving the quality of university education [11], [12]. Within data mining, there is text mining, which is the process of finding implicit patterns, previously unknown, that can be useful from a text repository [13]. Likewise, text mining consists of the search for regularities or patterns found in a text, from machine learning techniques, therefore, it is considered as one of the many branches of computational linguistics [14].

Text mining is embedded in the techniques of natural language processing (NLP), which is a field of computer science, artificial intelligence and linguistics that studies the interactions between computers and human language, through the syntactic, semantic, pragmatic and morphological analysis; These rules, in combination with the information stored in computational dictionaries, define the patterns to be recognized by means of images, text and voice [15], [16]. Thanks to the expansion of technology, in recent years there has been an exponential growth in subjective information available on the Internet. This phenomenon has sparked interest in sentiments analysis (AS), a natural language processing (NLP) task that identifies opinions related to an object [17]. All of this enriches information to the educational sector, in its search to improve the quality of the service, because nowadays the communication and interaction between students and teachers are social networks [18], [19]. The quality of the educational service has become a social requirement based on the educational policy implemented at the international level, globalization being a factor that inclines us to think that quality standards are established at the international level based on indicators of excellence in academic productivity, which are the results of scientific and technological research [20]. Given what has been described, a crucial ingredient to achieve quality is a great institutional capacity to adapt with initiative to new requirements and changes in its environment; that is, to function as a social and organizational system that combines homogeneity and homeostasis in a state of equilibrium, applying a series of self-regulated processes-related to organizational properties [21], [22].

In addition to the functions inherent to their mission, organizations play an important role in promoting efforts to improve quality in higher education through the transfer of knowledge and experience of quality processes and the practices to implement them [23]. One of the universal principles of quality management is customer focus. Educational institutions, in this search for opportunities to improve, have been identifying models to evaluate student satisfaction in tune with trends in quality management and performance excellence [24]. As indicated in [25], knowing the dimension of student satisfaction with the institution to which they attend will allow the identification of both positive and negative aspects, the latter being fundamental when determining strategies to improve educational quality. The purpose of analyzing student satisfaction is complex since multiple factors such as personal, social, academic, institutional or technological influence, among others [26]. According to [27], the key factor is related to the ability of teachers to offer a response according to the expectations and needs of the students. As indicated in [28] the performance of the university teacher is fundamental for the assessment of the quality of the educational service. Usually to measure student satisfaction, traditional tools and instruments such as student surveys and questionnaires are used. Nowadays, under this virtualization context, we seek to implement new methods and procedures in order to facilitate the use of digital tools [29]. Taking advantage of this way the benefits of social networks, and their relationship with self-learning, collaborative work and critical thinking, taking into account, in addition, that by this means students can share, respond, comment and discuss various information of the teaching process- learning [30].

In this sense, this article aims to identify the factors that influence the satisfaction of university students about their learning in virtual environments, through the sentiments analysis. In this way, it would be contributing to the improvement of the quality of the university service in relation to the factor of teaching performance, which, under the perception of the students obtained annually by the traditional questionnaires, needs to improve. Likewise, the analysis of big data in the field of education is a subject still little explored, which is why this article contributes to the debate on the contributions that data science offers us.

2.   LITERARY REVIEW

Text analysis, in which the texts are generated by tweets and other social networks, becomes even more complicated than the text coming from other more formal sources, or that go through a filter before being published, due to the freedom with which people write them, such as the use of abbreviations, incorrect writing, use of terms specific to each country and the inclusion of special characters, such as symbols, URLs, retweet entities, among others [31]. In this regard in [32], the author concludes that extraction tools are very useful when carrying out the process of collecting data from the web, to use these tools depends a lot on the structure of each web page and its data structure, In the case of Twitter, it is taken into account that the data is presented as the page progresses. Likewise, in an investigation related to opinion mining through twitter, it is concluded that the system for the analysis of opinions in the social network Twitter was successfully developed, extracting tweets, lemmatizing tokens, classifying and analyzing corpora and finally, display the results [33].

And it is that sources such as Twitter are recommended for the generation of opinion databases for two reasons, firstly because of the feasibility of the type of data (text) that allows analysis through NLP, unlike other networks in that its content consists of unstructured data (images or videos), and secondly by the volume of data generated that allows reaching minimum levels to perform complex analyzes [16]. In the same line of research in [14]. It is indicated that new tools have been created to facilitate access to the large amount of information that is generated daily, one of the most used at the organizational level is text mining, which offers the educational organization the possibility of exploring large amounts of texts, not organized in the form of data, establishing patterns and extracting useful knowledge in terms of educational management.

3.   METHOD
3.1. Descripcion of the data collected

The research work takes as a unit of analysis the comments or opinions expressed in the social network twitter by the students enrolled in the automatic process control course, of the professional school of mechanical and electrical engineering of the National Technological University of Lima Sur (UNTELS). The data acquisition was carried out through a non-experimental longitudinal type design, since the data is acquired in each class session. This collection period is from week 9 to week 13. It is non-experimental because no action is exercised on the students that stimulates an alteration or modification of their comments, so the data is treated in its natural condition.

It should be noted that the research project as a pilot test takes as a data collection period those 5 weeks, in order to obtain a first result and then apply this model to the analysis of the sixteen weeks of class that comprise an academic cycle, in the subjects of professional engineering schools in Peru. In Figure 1, the number of interactions that were generated in the twitter account during the study period is observed. It should be noted that the total impressions obtained in the 5 weeks of data acquisition were 455, and the number of interactions or comments represented by "tweets" was 220.
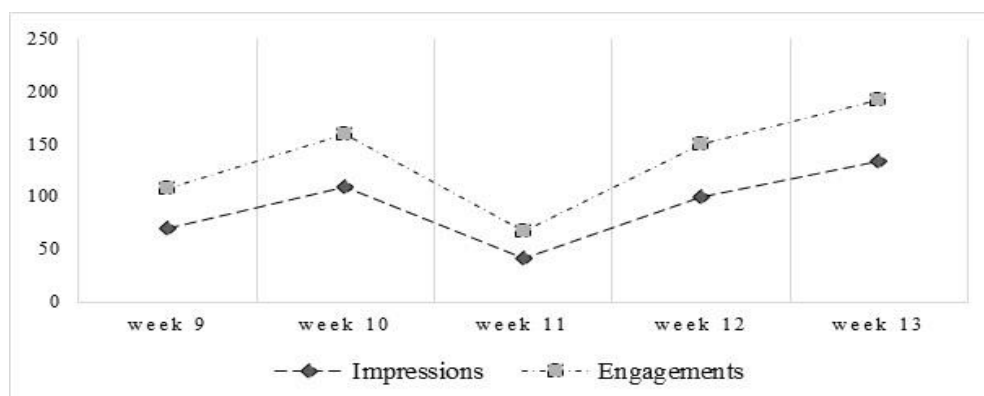

Figure 1. Interactions in the twitter account in the study period

3.2. Description of pre-processing, analysis of sentiments polarization and data extraction

In Figure 2, the model used for the identification of relevant factors of university student satisfaction is shown, under the sentiments analysis and the IF-IDF approach (Frequency of completion and frequency of

reverse document). The model shown specifies that from the data provided by the social network twitter in file format with the extension "csv", these will be conditioned to obtain the polarization of the comments made by the social network twitter through sentiments analysis. The conditioning or pre-processing consists of eliminating repeated texts, individual characters that are isolated, empty spaces that are found between text and text, as well as converting all texts to lower case.
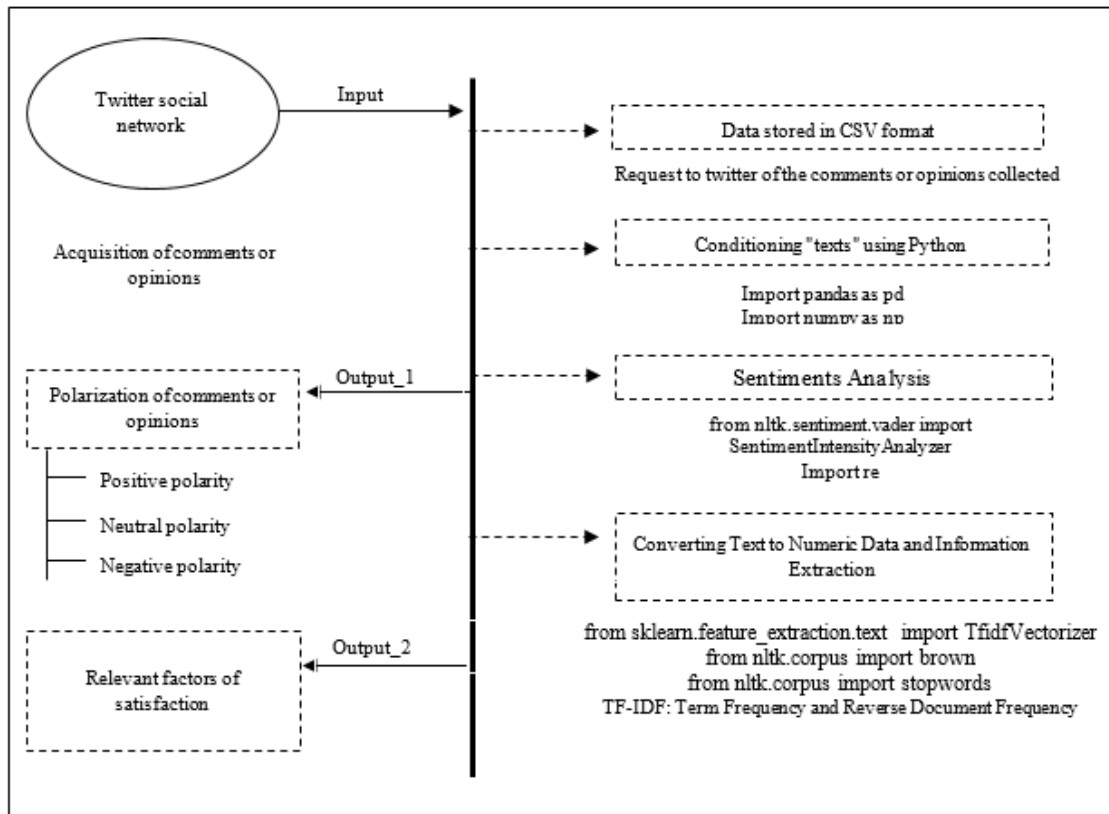

Figure 2. Model used to identify relevant factors of university student satisfaction

Next, the sentiments analysis of the tweets written in English is carried out, using Python libraries such as "nltk" and "vader". The result obtained will be the quantification of the feeling contained in each comment or opinion written by the student; the same ones that are classified in positive polarity, neutral polarity and negative polarity, representing output 1 of the model used; the same that will represent relevant information for decision-making in the improvement of the teaching-learning process of the process control subject. Table 1 specifies the ranges in which the "nltk" and "vader" libraries classify the sentiments contained in each tweet according to their polarity.

Table 1. Classification of sentiments according to their polarity through NLTK and vader

Sentiments polarity		
Positive	Neutral	Negative
Greater than 1	Equal to 0	Less tan 1

Likewise, in the proposed model it is observed that the last stage is related to the extraction of texts, through the TF-IDF approach, where TF represents the termination frequency and IDF represents the reverse document frequency; This indicator (TF-IDF) will allow selecting the texts that appear more in each of the comments or opinions and at the same time that appear less globally. Through this stage, the model allows identifying the relevant factors in university satisfaction with teaching performance in virtual teaching. This second output (output 2) of the proposed model will also provide the TF-IDF value, which represents the weight of a text "i" in a tweet "j". In (1), the relationship between TF-IDF, TF and IDF is shown, associating

it with the comments obtained through the social network twitter, where "i" and "j" represent the number of texts in a tweet and "j" Represents the number of tweets, obtained during the data collection period, where Tei represents the texs "i" and Twj represents tweet "j".

$$TF - IDF\ (Tei, Twj)\ =\ TF\ (Tei, Twj)\ x\ IDF\ (Tei, Twj) \tag{1}$$

3.3. Development of the proposed model through Python in the JupyterLab environment

Following the proposed model, Figure 3 shows the programming code that will allow loading or reading in the JupyterLab environment the file provided by twitter, called "tweets_process_control_ 29Dic21.csv". The "pandas" library will be imported into the program with which it will be possible to read and write data, in different formats or extensions, in this investigation the file has a "csv" extension; the "numpy" library will be used to create vectors or matrices in which it will seek to link the texts with their corresponding sentiments and polarity.

```python
import numpy as np

import pandas as pd

df=pd.read_csv("tweets_process_control_29Dic21.csv")
```

Figure 3. Script for reading twitter data in Python via JupyterLab

Once the data has been read, we proceed to carry out the conditioning or pre-processing of the texts contained in the comments on the twitter account, making use of the Python libraries ("nltk.sentiment.vader", "SentimentIntensityAnaly zer", "re"); it should be noted that these libraries are strictly linked to text conditioning, to then determine the level of sentiment. Once the texts had been conditioned, and all of them were standardized, the sentiments contained in each of the tweets, written by the students of the process control subject, was determined. In Figure 4 the programming code is shown, in which a column will be generated on the right side of the texts, called "sentiment" in which the results of the sentiments contained in each tweet will be stored. With this new data, a new matrix is generated containing the file "sentiments_data.csv". As evidence of the result obtained, Table 2 shows the sentiments value of the first 5 tweets; as mentioned in the paragraph, this data is stored in the column called "sentiment". From these results, it was possible to assign the corresponding polarity to each Tweet; and with this, in this stage, the satisfaction results will be obtained from the comparison of the number of comments with positive polarity and negative polarity.

```python
df= pd.read_csv("tweets_process_control_29Dic21.csv")

sid =SentimentIntensityAnalyzer()

df["sentimiento"]=df["Comments on the satisfaction of the teaching performance"].apply(lambda i: sid.polarity_scores(i)['compound'])

df.to_csv("sentiments_data.csv")

df.head()
```

Figure 4. Script to obtain the sentiments contained in each Tweet

Table 2. Result of the sentiments contained in each Tweet

Number Tweet	Comments on the satisfaccion of the teaching performance	Sentiment
0	It has been a good way to deepen the topic sho…	0.4404
1	I would have liked the problems proposed ih the…	0.2960
2	I found it interesting that the maximum oversh…	0.4019
3	The statements are the same as those learned …	0.0000
4	There are topics that were not clear in other …	-0.2924

Following the stages described in the proposed model, Figure 5 shows the programming code in which it seeks to extract the most relevant texts, contained in the sentiments tweets whose polarity is positive and negative, since these are closely related to the satisfaction and dissatisfaction of the teaching performance. The code makes use of libraries such as "stopwords" and "Brown" imported from the nltk.corpus library, the same ones that allow text processing; the "stopwords" contain a group of words that in simple terms have no meaning, so it will help to make the correct extraction of relevant "words". While the corpus "Brown" contains a number of representative words that allow them to be used by comparison to extract words that have a semantic and linguistic sense from any document, written in English or Spanish.

```python
from nltk.corpus import stopwords

from nltk.corpus import brown

vectorizer = TfidfVectorizer (max_features=10, min_df=10, max_df=0.7, stop_words=stopwords.words('english'))

processed_features = vectorizer.fit_transform(processed_features).toarray()
```

Figure 5. Script *para realizar la extracción de palabras, basado en el enfoque* TF-IDF

Also, as mentioned in the description of the proposed model, the approach used for the extraction of texts will be TF-IDF; so the "TfidfVectorizer" function is used. It should be noted that this function allows setting the parameter "max_features", which represents the number of "words" that appear most frequently and that must be extracted from all tweets with positive and negative polarity generated by the students. This also specifies the parameter "min_df", which establishes the minimum number of tweets in which the extracted words must be found; on the other hand, the parameter "max_df" is also specified, which represents the percentage of tweets in which at most the extracted texts should be found.

4.   RESULTS

Taking into account that the objective of the research is to describe the satisfaction of virtual learning from the sentiments analysis, contained in the comments or opinions of university students, carried out through the social network Twitter. In Table 3, it is shown shows the polarity result obtained through Python and its libraries linked with the natural language toolkit (NLTK). These results show that it is possible to obtain a benchmark on student satisfaction using text mining and natural language processing (NLP) techniques.

Table 3. Distribution of the polarity of the sentiments contained in the tweets

	Positive polarization	Negative polarization	Neutral polarization
Tweets numer	126	60	34
Percent	57.27%	27.27%	15.45%

Another aspect also considered as part of the research objectives is to determine the relevant factors of virtual learning satisfaction, based on predominant text extraction techniques, generated under the TF-IDF approach; which is why the following results are described. In Figure 6, the results of the word extraction are shown, in which in order to improve the performance and precision of the extraction it was defined that "max_features" is equal to 10, "min_df" equal to 10 and "max_df" equal to 0.7; In other words, the 10 words are extracted with the highest frequency, and that they appear in at least 10 tweets simultaneously and that they appear in at most 70% of all tweets. Also in Figure 6, the IDF levels are shown for each word extracted; in which their weights are evidenced with respect to a document or Tweet. The terms with the highest IDF were "simulink" and "pid", with a value of 3.1361; This result makes it possible to establish that the relevant factors that contribute significantly to the satisfaction of virtual learning are those related to the teaching performance in the class sessions that allow to achieve the understanding of the laboratory sessions through the use of simulation tools such as simulink and pid controllers (proportional-integrative-derivative).

Now considering the negative polarity tweets, in Figure 7 the results of the word extraction are shown, considering "max_features" is equal to 10, "min_df" is equal to 10 and "max_df" is equal to 0.7. In other words, the 10 words are extracted with the greatest frequency, and that these appear in at least 10 tweets simultaneously and that they appear in at most 70% of all tweets; this in order to establish the same level of performance and precision in the word extraction process. These extracted texts allow to establish that the relevant factors that generated dissatisfaction with virtual learning were linked to the laboratory sessions in which graphical representation and block diagrams were used to explain the class session; obviously, it was not possible for the students to understand in a meaningful way the use of these tools.

```
print(vectorizer.vocabulary_)

{'class': 1, 'interesting': 2, 'laboratory': 3, 'understand': 7, 'able': 0, 'us': 8, 'teacher': 6, 'use': 9, 'simulin
k': 5, 'pid': 4}

print (vectorizer.idf_)

[2.28883902 2.54835022 2.13061502 2.54835022 2.79966465 2.70869287
 2.89974811 2.89974811 3.13613689 3.13613689]

print(processed_features)

[[0.         1.         0.         ... 0.         0.         0.        ]
 [0.         1.         0.         ... 0.         0.         0.        ]
 [0.         0.         1.         ... 0.         0.         0.        ]
 ...
 [0.         0.         0.         ... 0.         0.         0.        ]
 [0.         0.         0.56195682 ... 0.         0.         0.82716657]
 [0.         1.         0.         ... 0.         0.         0.        ]]
```

Figure 6. Word extraction from tweets with positive polarity

```
[[0.         0.         0.         0.         0.        ]
 [0.         0.         0.         0.         0.        ]
 [0.         0.         0.         0.         1.        ]
 [0.         0.         0.         1.         0.        ]
 [1.         0.         0.         0.         0.        ]
 [0.39990794 0.43088679 0.46705111 0.46705111 0.46705111]
 [0.         0.         0.         0.         1.        ]
 [0.45226672 0.48730154 0.52820073 0.52820073 0.        ]
 [0.         0.         0.         0.         0.        ]
 [0.         0.         0.         0.         0.        ]
 [0.         0.         0.         0.         0.        ]
 [0.45226672 0.48730154 0.52820073 0.52820073 0.        ]
 [0.88037151 0.47428473 0.         0.         0.        ]
 [0.39990794 0.43088679 0.46705111 0.46705111 0.46705111]
 [0.         0.         0.         0.         1.        ]
 [0.53263021 0.57389038 0.62205698 0.         0.        ]
 [0.         0.         0.         0.         0.        ]
 [0.         0.         0.         0.         0.        ]
 [0.         0.         0.         0.         0.        ]
 [0.         0.         0.         0.         1.        ]
 [0.         0.         0.         1.         0.        ]
 [1.         0.         0.         0.         0.        ]
 [0.39990794 0.43088679 0.46705111 0.46705111 0.46705111]
 [0.         0.         0.         0.         1.        ]
 [0.45226672 0.48730154 0.52820073 0.52820073 0.        ]
 [0.         0.         0.         0.         0.        ]
 [0.         0.         0.         0.         0.        ]
 [0.45226672 0.48730154 0.52820073 0.52820073 0.        ]
 [0.88037151 0.47428473 0.         0.         0.        ]
 [0.39990794 0.43088679 0.46705111 0.46705111 0.46705111]
 [0.         0.         0.         0.         1.        ]
 [0.53263021 0.57389038 0.62205698 0.         0.        ]
 [0.         0.         0.         0.         0.        ]]
```

Figure 7. Word extraction from tweets with negative polarity

5. DISCUSSION

In relation to the degree or level of university student satisfaction regarding teaching performance in the virtual teaching-learning environment based on sentiment analysis techniques, whose opinions were collected from the social network twitter, it is established that it was possible to assess the degree of

satisfaction; associating it with positively polarized tweets, whose result is 57.27%. However, today the analysis of the data generated in the anniversary environment related to university satisfaction is generated by surveys with categorical response questions defined on a Likert scale, with already defined indicators to be assessed, applied once per academic semester which generates highly biased results; this is where the relevance of using data science, data mining and natural language processing is focused, with which it is possible to generate satisfaction results in real time from open comments, free opinions, in each class session and not just once or twice a year, as is traditionally done. The contribution of this study seeks to identify which factors should be improved by the teacher in order to have a preponderant effect on improving student satisfaction and improving academic quality.

In this regard, in [34] the authors manage to assess the perception of university students about academic activities based on sentiment analysis techniques, and more than focusing on the result obtained from satisfaction, I emphasize that in their research they use as an instrument to collect information data a survey with 3 categorical response questions and 4 open response questions; and here I disagree with the method used to assess satisfaction, since the survey presents already defined factors to be evaluated, which leads to questioning how these factors were identified. Secondly, the survey is applied on a single date, which leads to questioning and how it determined the moment in which the survey should be applied; these two conditions are those that when using social networks such as twitter, open opinion data can be collected, with a wide range of elements that the student considers relevant, in each class session, throughout the academic semester and its processing can be done in real time. Likewise, in [35] the author seeks to determine the level of satisfaction with online teaching using sentiment analysis, using a survey as a data collection technique, which again shows that many studies focus on the result without focusing in the disadvantages or weaknesses generated by applying traditional methods, without focusing on how data science and natural language processing can achieve a transformation of how today determine the perception of satisfaction in university institutions.

Regarding the identification of the relevant factors, the predominant TF-IDF text extraction approach was used, which was applied to the comments or opinions expressed by the students during different class sessions, through the social network twitter, managing to identify that the most relevant factors are linked to the use of simulation software for understanding the theoretical aspects, and it is understandable since being students of the school of mechanical and electrical engineering it is necessary to make use of these tools. And it is that due to the context of the health emergency, all the courses from the I to the X cycle are taught virtually; not all teachers use simulation software, that is where the relevance of changing the paradigms of satisfaction assessment is focused on very delimited questions of a survey, and on the contrary, techniques should be sought to process texts or comments issued freely by students, through data science. In this regard, in [36], [37] the authors make use of the social network twitter for data collection and through sentiment analysis they manage to evaluate university satisfaction in universities; in this case they manage to extract texts, with the purpose of identifying relevant factors in the evaluation process and achieving automated analysis; with which these works are close to the approach proposed in my research; agreeing on the method and technique of data processing.

6. CONCLUSION

In this article, the factors that influence university student satisfaction about their learning in virtual environments were identified, through sentiments analysis and the TF-IDF approach, showing that it is possible to obtain a reference point on student satisfaction using text mining and natural language processing (NLP) techniques. Given this, it was identified that the relevant factors that contribute significantly to the satisfaction of virtual learning are those related to the teaching performance in the use of simulation tools that allow understanding the class sessions. Once the research has been carried out, it is recommended to implement data science techniques and sentiments analysis relying on social networks in order to determine university student satisfaction with respect to the performance of the teacher and that higher institutions can make timely decisions for the same academic semester, and that these are not carried out after the end of the academic semester, as is currently done in many universities. For future work, it is recommended to expand the line of research by developing a predictive model of student satisfaction, with respect to the development of the teacher. Likewise, the results obtained could be compared with those obtained through traditional techniques and instruments for evaluating teacher performance satisfaction in order to establish the most optimal model to know the satisfaction of the students.

REFERENCES

[1]   D. Lemus-Delgado and R. Pérez-Navarro, "Data science and global studies: contributions and methodological challenges," *Colombia International*, no. 102, pp. 41-62, 2020, doi: 10.7440/colombiaint102.2020.03.

[2]   S. Stieglitz, M. Mirbadaie, B. Ross, and C. Neuberger, "Social media analytics – Challenges in topic discovery, data collection, and data preparation," *International Journal of Information Management*, vol. 39, pp. 156-168, Apr. 2018, doi: 10.1016/j.ijinfomgt.2017.12.002.

[3]   D. Gupta and R. Rinkle, "A study of big data evolution and research challenges," *Journal of Information Science*, vol. 45, no. 3, pp. 322-340, Jul. 2019, doi: 10.1177/0165551518789880.

[4]   C. Aradau and T. Blanke, "Politics of prediction: Security and the time/space of governmentality in the age of big data," *European Journal of Social Theory*, vol. 20, no. 3, pp. 373-391, Sep. 2016, doi: 10.1177/1368431016667623.

[5]   P. Dourish and E. Gómez, "Datafication and data fiction: Narrating data and narrating with data," *Big Data & Society,* vol. 5, pp. 1-10, Jul. 2018, doi: 10.1177/2053951718784083.

[6]   M. Sandeepkumar and P. Jayagopal, "A big data virtualization role in agriculture: A comprehensive review," *Walailak Journal of Science & Technology*, vol. 16, no. 2, Sep. 2019, pp. 55-70, doi: 10.48048/wjst.2019.3620.

[7]   D. Simpson, "The use of big data: Benefits, risks, and differential pricing issues," *Nueva York: Nova Science Publisher*, 2016.

[8]   D. Boyd and K. Crawford, "Critical questions for big data," *Information, Communication & Society*, vol. 15, no. 5, pp. 662-679, May. 2012, doi: 10.1080/1369118X.2012.678878.

[9]   D. J. Hand, H. Mannila, and P. Smyth, "Principles of data mining," *Cambridge: A Bradford Book-The MIT Press*, 2001.

[10] W. Graham, Williams, "Data mining with rattle and R. The art of excavating data for knowledge discovery," *Nueva York: Springer Science Business Media*, 2011.

[11] M. E. Meneses, "Big data, big challenges for social sciences," *Mexican Journal of Sociology*, vol. 80, no. 2, p. 57723, Jan. 2018, doi: 10.22201/iis.01882503p.2018.2.57723.

[12] N. Couldry and A. Powell, "Big data from the bottom up," *Big Data & Society,* vol. 1, pp. 1-5, Jul. 2014, doi: 10.1177/2053951714539277.

[13] R. Heiberger and J. Riebling, "Installing computational social science: Facing the challenges of new information and communication technologies in social science," *Methodological Innovations*, vol. 9, pp. 1-11, Mar. 2016, doi: 10.1177/2059799115622763.

[14] R. Kitchin and G. McArdle, "What makes big data, big data? exploring the ontological characteristics of 26 datasets," *Big Data and Society*, vol. 3, pp. 1-10, Feb. 2016, doi: 10.1177/2053951 716631130.

[15] M. Destéfano, "Language processing and informational semantics," *Philosophical Praxis*, no. 48, pp. 153-174, 2018, doi: 10.25100/pfilosofica.v0i48.7313.

[16] M. M. Echeverri and R. Manjarrés-Betancur, "Virtual academic assistant using technologies natural language processing cognitive," *Polytechnic Journal,* vol. 16, no. 31, pp. 85-96, Jan. 2020, doi: 10.33571/rpolitec.v16n31a7.

[17] J. Fernández, J. M. Gómez, E. Boldrini, and P. Martínez-Barco, "Sentiment analysis and opinion mining: the emotiblog corpus natural language processing," *Spanish Society for Natural Language Processing*, no. 47, pp. 179-187, 2011.

[18] P. M. Bigatel and S. Edel-Malizia, "Using the indicators of engaged learning online framework to evaluate online course quality," *TechTrends*, vol. 62, no. 31, pp. 58-70, Jan. 2017, doi: 10.1007/s11528-017-0239-4.

[19] D. Demissie and A. Rorissa, "The effect of information quality and satisfaction on a parent's behavioral intention to use a learning community management system," *Libri*, vol. 65, no. 2, pp. 143-150, Jun. 2015, doi: 10.1515/libri-2015-0019.

[20] A. R. Pérez and A. G. López, "Educational quality in graduate school: university autonomy and research," *Atenas*, vol. 1, pp. 147-155, 2018.

[21] C. Paradeise and Jean-claude Thoening, "In search of academic quality," *Palgrave Macmillan*, no. 51, pp. 1-238, 2015, doi: 10.1057/9781137298294

[22] B. A. Christie, K. K. Miller, R. Cooke, and J. G. White, "Environmental sustainability in higher education: What do academics think?," *Environmental Education Research*, vol. 21, no. 5, pp. 655–686, Mar. 2015, doi: 10.1080/13504622.2013. 879697.

[23] E. Alvarado-Lagunas, J. Luyando-Cuevas, and E. Picazzo-Palencia, "Students' perception of the quality of private universities in Monterrey," *Ibero-American Journal of Higher Education*, vol. 6, no. 17, pp. 58-76, 2015, doi: 10.22201/iisue.20072872e.2015.17.1088.

[24] H. Al-Samarraie, B. K. Teng, A. I. Alzahrani, and N. Alalwan, "E-learning continuance satisfaction in higher education: A unified perspective from instructors and students," *Studies in Higher Education*, vol. 43, 2017, doi: 10.1080/03075079.2017.1298088.

[25] N. Dioses, H. I. Morales, D. M. Díaz de Angulo, and A. Vasquez, "Satisfaction level in training quality at university level," *Telos*, vol. 23, no. 2, pp. 247-261, May. 2021, doi: 10.36390/telos232.04.

[26] P. Barrientos-Illanes, M. V. Pérez-Villalobos, J. Vergara-Morales, and A. Díaz-Mujica, "Influence of the perceived autonomy support, self-efficacy, and academic satisfaction in the intentions of permanence of university students," *Educare Electronic Journal*, vol. 25, no. 2, pp. 1-14, May. 2021, doi: 10.15359/ree.25-2.5.

[27] J. Vergara-Morales, M. Del Valle, A. Díaz and M. V. Pérez, "Adaptation of the academic satisfaction scale in Chilean university students," *Educational psychology*, vol. 24, no.2, pp. 99-106, Apr. 2018, doi: 10.5093/psed2018a15.

[28] R. Lent, D. Singley, H. Sheu, J. Schmidt, and L. Schmidt, "Relation of social-cognitive factors to academic satisfaction in engineering students," *Journal of Career Assessment*, vol. 15, no. 1, pp. 87-97, Feb. 2017, doi: 10.1177/106907 2706294518.

[29] K. G. Wilkins-Yel, M. L. Roach, T. J. G. Tracey, and N. Yel, "The effects of career adaptability on intended academic persistence: The mediating role of academic satisfaction," *Journal of Vocational Behavior*, vol. 108, no. 1, pp. 67-77, Oct. 2018, doi: 10.1016/j.jvb.2018.06.006.

[30] R. A. Salas and R. D. S. Rueda, "Impact of the social network Facebook on the higher educational process of the mathematics considering the science of data," *Journal of Social Sciences and Humanities*, vol. 28, no. 55, pp. 23-42, Dec. 2019, doi: 10.20983/noesis.2019.3.2.

[31] R. A. Cortez-Reyes, "Extraction of knowledge from texts obtained from Twitter," *Environment Journal*, no. 65, pp. 30-41, Jun. 2018, doi: 10.5377/entorno.v0i65.6048.

[32] C. E. Riofrio, A.C. Chóez, and J. Z. Gamboa, "Methods for extracting comments from the social network Twitter for use in natural language processing," *Pole of Knowledge Journal,* vol. 6, no. 11, pp. 104-123, Nov. 2021, doi: 10.23857/pc.v6i11.3257.

[33] B. P. Sánchez and L. P. Terán, "Opinion mining on tects on twitter," *Research in Computing Science*, vol. 148, no. 7, pp. 13-21, Apr. 2019, doi: 10.13053/rcs-148-7-1.

[34] G. E. Chanchi, M. A. Ospina, and M. E.Ospino, "Sentiment analysis of student perception systems engineering from the University of Cartagena (Colombia) on the academic activities developed during lockdown due to COVID-19," *Espacios Journal*, vol. 41, no. 42, pp. 247-259, Nov. 2020, doi: 10.48082/espacios-a20v41n42p21.

[35] J. Zhang, "Research on sentiment analysis and satisfaction evaluation of online teaching in universities during epidemic prevention," *Frontiers in Psychology*, vol. 12, pp. 2-7, Oct. 2021, doi: 10.3389/fp syg.2021.73 8776.

[36] A. Abdelrazeq, D. Janssen, C. Tummel, S. Jeschke, and A. Richer, "Sentiment analysis of social media for evaluating universities," *Proceedings of Second International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC2015)*, 2015.

[37] M. Korkmaz, "Sentiment analysis on university satisfaction in social media," in *2018 Electric Electronics, Computer Science, Biomedical Engineerings Meeting (EBBT),* Apr. 2018, doi: 10.1109/EBBT.2018.8391463.

BIOGRAPHIES OF AUTHORS

Omar Chamorro-Atalaya is an electronic engineer, graduated from the National University of Callao (UNAC), with a Master's degree in Systems Engineering and a doctoral student at the Faculty of Administrative Sciences at UNAC. Researcher recognized by CONCYTEC (National Council of Science, Technology and Technological Innovation – Peru). Research professor at the National Technological University of South Lima (UNTELS), he teaches courses on automatic process control and industrial automation, and design of control panels and electrical control. He can be contacted at email: ochamorro@untels.edu.pe.

Dora Arce-Santillan is an electrical mechanical engineer, with a Master's Degree in Electrical Engineering with a mention in Engineering Project Management at the National University of Callao (UNAC). She is the author of scientific articles indexed in Scopus and WoS. She reviewer of scientific articles in international journals indexed to Scopus. She can be contacted at email: yvonne.asantillan@gmail.com.

Guillermo Morales-Romero External Evaluator in University Higher Education at SINEACE. Doctor in Educational Sciences, Master in Systems Engineering, Master in Public Management, Master in Educational Management. Bachelor of Mathematics and Computer Science, Lawyer with specialist in computer auditing, computer security, X cycle Systems Engineering of UNFV, Professional with 24 years of experience in University Teaching in the careers of Engineering, Computer Science, Law, Mathematics Education, Statistics. Teacher in the different Postgraduate Schools of Public and Private Universities of the Country. Member of the Scientific Committee of the Scientific Journal: Ciencia & Sociedad Autonomous University Tomás Frías de Bolivia; Member of the Scientific Committee of the Multidisciplinary Magazine Ciencia Latina. Mexico. Member of the Lima Chamber of Commerce and the Lima Bar Association. Former specialist in the AGEBAT Area of UGEL 05 - MINEDU. He can be contacted at email: gmorales@une.edu.pe.

Primitiva Ramos-Salazar with completed doctoral studies in educational sciences, Master's degree in educational sciences, completed Master's studies in administration, Bachelor's degree in textile technology, with 23 years of experience in teaching technical-productive education and 7 years in university teaching. She can be contacted at email: pramos@une.edu.pe.

César León-Velarde ID SC P is Doctor in Education, Master in Education: Measurement, Evaluation and Accreditation of Educational Quality, Master in Educational Management, Graduate and Bachelor in Education, Bachelor in Sacred Theology. Professor of Philosophy and CDCSS. With more than 10 years of experience as a teacher in private and national universities in Peru. Recognized by a private university as a teacher of academic excellence and as one of the best teachers in that university. I love e-learning education in which I have been able to specialize over the years and obtain pleasing results and recognition. He can be contacted at email: c19593@utp.edu.pe.

Elizabeth Auqui-Ramos ID SC P has a Master's degree in the mention of University Teaching, with completed doctoral studies in the mention of Educational Sciences, studies in Industrial Engineering. She is fluent in English at an intermediate level, Italian and Quechua at a basic level, graduated in mathematics didactics and management and pedagogical innovation for Educational quality and competitiveness, with 11 years of experience in teaching at higher and university levels. She can be contacted at email: pcmaeauq@upc.edu.pe.

Miguel Levano-Stella ID SC P is an Electronic Engineer from the Ricardo Palma Private University (URP). Candidate for a Master's in Education with a mention in University Teaching from the Universidad Nacional Mayor de San Marcos (UNMSM). Full-time professor in the Computer Systems Engineering career at the Universidad Privada del Norte (UPN) Lima-Perú, he teaches courses on digital electronics, robotics, computer architecture and telecommunications. Research professor at UPN, dedicated to the investigation of educational methodologies in Engineering, solutions based on embedded systems. He can be contacted at email: miguel.levano@upn.edu.pe.