Opinion Mining of the Mexican Tourism Sector through Sets of Normalized N-Grams

Juan Carlos Rivas-Álvarez [1], René Arnulfo García-Hernández [1], Sergio Iván Medina-Martínez[1], Ana Mercedes Martínez-Ortiz[1], Néstor Hernández-Castañeda [1], José Esteban Ruiz-Melo[1], Ángel Hernández-Castañeda[1], Yulia Nikolaevna-Ledeneva[1].

[1] *Tianguistenco Professional Academic Unit, Autonomous University of the State of Mexico 1, San Pedro Tlaltizapan,, Tianguistenco, 52640, Mexico.*

Abstract

Opinion mining is a challenging problem as polarity should be predicted (negative to positive, between classes 1 to 5), from texts that express experiences from different people in the Mexican tourism sector. Those opinions are written by people from various regions with different terms. In this "working notes", we propose to extract all the terms in each class from one to four words (1...4-grams) as characteristics of the polarity. The first proposed method has a low computational complexity compared to other state-of-the-art methods because it only considers the terms of the original corpus. The second method performs a chain of translations of the opinions, from Spanish to other languages and back to Spanish, to obtain meanings and synonymous terms. In the experimentation, the Rest-Mex 2022 corpus integrates TripAdvisor opinions between 2002 and 2020.

Keywords

Rest-Mex 2022, opinion mining, polarity, attraction, tourism, sentiment analysis.

1. Introduction

Tourism in Mexico promotes national development; the natural and cultural wealth, in addition to its hospitality, has been a source of attraction for national and foreign tourists. Tourism represents an important source of foreign income for employment and regional progress. Among the main international destinations in 2017, Mexico was the sixth most visited country in the world with 39.3 million international travelers, and in 2019 (before the COVID-19 pandemic), it ranked seventh with 45 million tourists [14].

In 2021, according to the Anahuac Tourism Competitiveness and Research Center (CICOTUR) and the World Tourism Barometer of the United Nations World Tourism Organization (UNWTO), Mexico received 31.9 million tourists positioned it in the second rank between the top 10 countries with the highest number of international tourists [7].

Natural Language Processing (NLP) is a field from artificial intelligence and applied linguistics that studies the interactions through natural language between humans and digital machines [6]. Tourism opinion mining uses techniques of Natural Language Processing (NLP) to recover touristic activity, by offering tools to detect the problems in places of interest and businesses, based on the classification of opinions on online platforms, enhancing the recommendation of places from the best experiences documented by visitors, and offering useful data for improvement taking notice of unfavorable opinions.

Opinion mining about products or services, in general, offers valuable information that allows target audience understanding, in terms of their preference and needs. However, a dataset may contain imbalanced data, therefore not all the classes involved in the analysis, will be correctly represented,

IberLEF 2022, September 2022, A Coruña, Spain.
EMAIL: jrivasa001@alumno.uaemex.mx (J.C. Rivas-Álvarez); reagarciah@alumno.uaemex.mx (R.A. García-Hernández)

causing bias; this is clearly a problem because many machine learning algorithms are designed to maximize overall accuracy.

The 38th conference of the Spanish Society for Natural Language Processing, SEPLN 2022, offers a forum to present and share the latest research and developments in the field of Natural Language Processing (PNL) to the scientific community and companies in the sector. This conference counts with the track of Recommendation System for Mexican Tourism (REST-MEX) in its desire to address, through natural language processing, the social, cultural and economic phenomena of tourism. For this conference, REST-MEX called for a contest with three subtasks that involved a system of recommendations, sentiment analysis, and epidemiological traffic lights during the COVID-19 pandemic in Mexico [3].

The sentiment analysis subtask consisted of a classification task where the participant had to predict polarity and attraction of opinions issued by tourists who shared their opinion on TripAdvisor between 2002 and 2020 for representative places in Guanajuato, Mexico.

Opinion mining is a challenging problem as polarity should be predicted (negative to positive, between classes 1 to 5), from texts that express experiences from different people in the Mexican tourism sector. Polarity of each opinion is valued by an integer between [1, 5], where 1 represents the most negative polarity and 5 the most positive. Also, the type of Attraction about the opinion is considered: Attractive, Hotel and Restaurant. Polarity presents numerical values whose magnitude proportionally represents the positivity of the opinion, so it makes it possible to use a difference measure between continuous variables, such as the mean absolute error (MAE), to quantify the accuracy of a prediction.

The proposed methodology in this article found among the characteristics in the training corpus: language informality by the tourist who issued the opinion, lacking on grammar, subjectivity in the classification of the reviewed place, revealing an evident disparity in the criteria to define a language/polarity relationship between tourists. Likewise, this methodology pretends to be language independent, since the TripAdvisor site can contain opinions in any language; pursue a minimal computational complexity, compared to a state of the art dominated by machine learning algorithms, and modularity in stages that allow rethinking aspects of the methodology, taking advantage of the results obtained in previous stages.

2. Corpus analysis

The corpus lists 43,150 opinions shared on the TripAdvisor platform, divided into 70% for training and the remaining 30% for testing. Each row contains three columns:

1. Id (integer): The identifier of each record serves as a key for the results
2. Title (text): The title that the tourist gave to his opinion.
3. Opinion (text): The opinion issued by the tourist.

The training set consists of 30,212 opinions with unbalanced classes in both Polarity and Attraction, as shown in figure 1. Examples:

"Un callejón donde tienes que besar a tu amante por años de felicidad, en el amor es parte de un mito en esta ciudad especial. El callejón estrecho con escalones no es muy especial en sí mismo. Lo que lo hace especial es toda la historia a su alrededor."

Polarity: 5 (Very good)
Attraction: Attractive

"No alcanza nivel de restaurante, no disponen de un surtido básico de vinos y licores. La comida bien El servicio, pésimo."
Polarity: 3 (Neutral)
Attraction: Restaurant

"Pésimo servicio. Me ofrecieron una habitación que no me dieron el cuarto resultó estar la última torre del complejo prometiéndonos todo el lujo y la vdd estábamos lejísimos de todo. Para todos hay que tomar unos carritos de golf que están pésimo organizaros entonces esperamos a veces hasta 40 min. "

Polarity: 1 (Very bad)
Attraction: Hotel

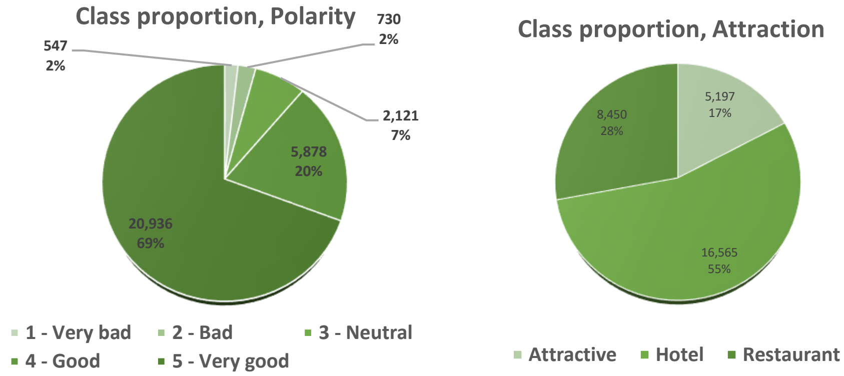

Fig. 1: Class proportion for Polarity and Attraction. Source: Own elaboration, based on the training set.

The training set is divided into two problems: polarity and attraction:

2.1. Analysis of the polarity problem

On the polarity problem, it is found that:

- On polarity class 1, there is at least one opinion with 3,144 words, at least one opinion with a minimum of 15 words, and an average of 147 words.
- In the case of polarity class 2, there is a behavior very close to polarity class 1, an average of 158 words.
- In class 4 and 5, there is at least one opinion with zero words, so it is understood that there are positive opinions without description, including only the title of the opinion.

Regarding the opinions length frequency for polarity: on class 1 there are 10 opinions of 47 words; there is an opinion with 3,144 words (the maximum of words for this polarity); the minimum length was observed only in one opinion (see figure 2), considering an interval of 46 to 47 words. Table 1 condenses the characteristics found for polarity.

It is worth to note that for polarity in class 2 there are 12 opinions with a length of 41 words, 10 opinions with 52 and 58 words, there is an opinion with the maximum of 1,677 words, the minimum length was observed only in one opinion (see figure 3), considering an interval of 41 to 42 words.

Table 1

Characteristics of the polarity training set.

Polarity	Opinions (percentage)	Quantity of words per opinion		
		Maximum	Minimum	Average
1	547 (1.81%)	3,144	15	147
2	730 (2.41%)	1,677	11	158
3	2,121 (7.02%)	3,481	8	142
4	5,878 (19.45%)	3,484	0	126
5	20,936 (69.29%)	2,700	0	93

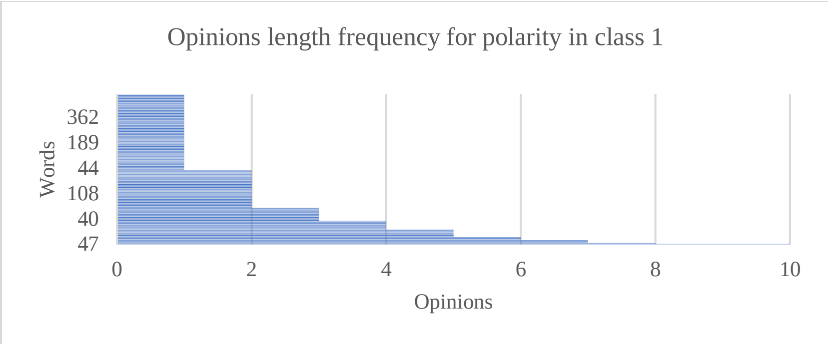

Fig. 2: Frequency of the quantity of words per opinion for polarity in class 1 based on the training set.

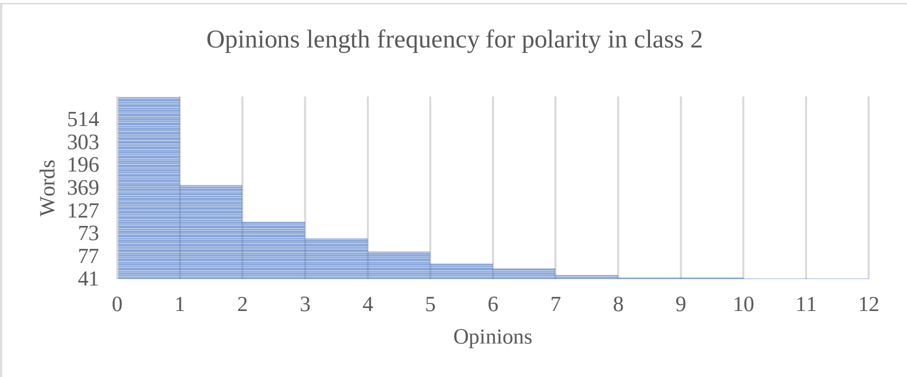

Fig. 3: Frequency of the quantity of words per opinion for polarity in class 2 based on the training set.

For polarity class 3, there are 13 opinions with a length of 144 words, and 12 opinions with 160 and 247 words. For both, the minimum and maximum lengths, there is only one opinion, respectively (see figure 4), considering an interval of 25 to 28 words.

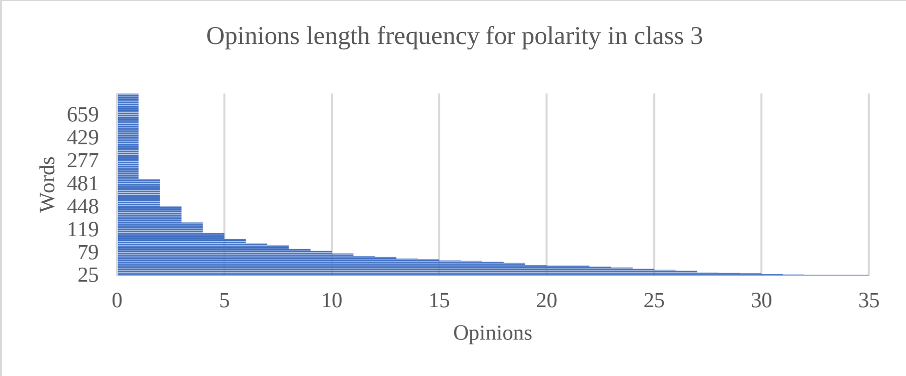

Fig. 4: Frequency of the quantity of words per opinion for polarity in class 3 based on the training set.

For polarity class 4, there are 109 opinions with length of 23 words, there are 97 opinions with 28 words, there is an opinion with no comment and an opinion with 3,484 words (see figure 5), considering an interval of 23 to 28 words.

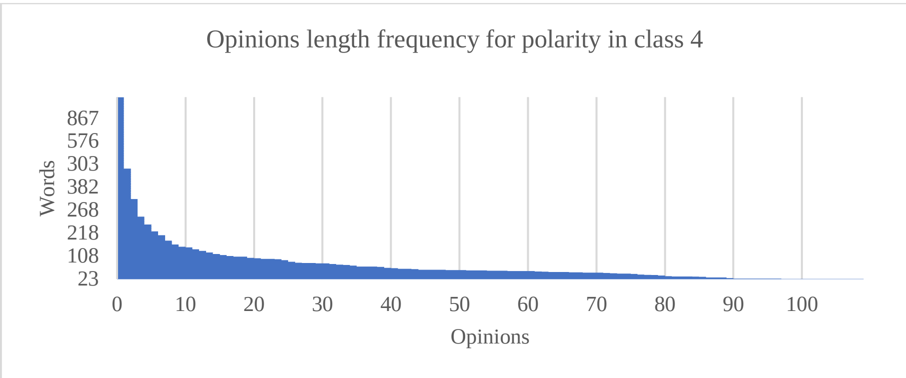

Fig. 5.: Frequency of the quantity of words per opinion for polarity in class 4 based on the training set.

For polarity class 5, there are 510 opinions with a length of 43 words, there are 489 opinions with 36 words, there is an opinion with no comment and an opinion with 2700 words (see figure 6), considering an interval of 42 to 43 words.

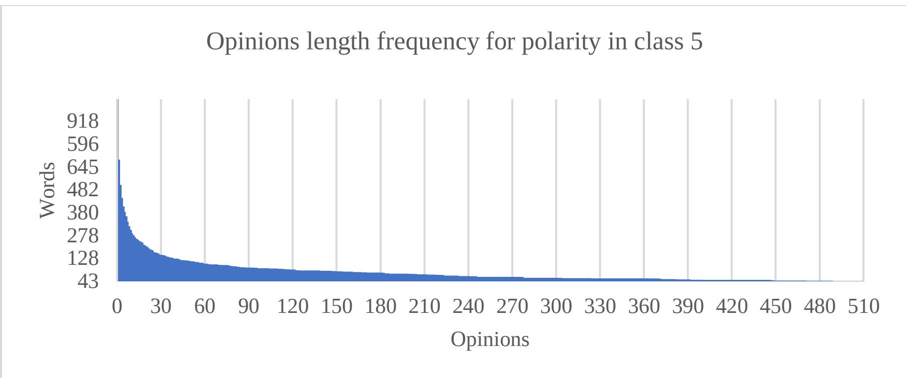

Fig. 6: Frequency of the quantity of words per opinion for polarity in class 5 based on the training set.

2.2. Analysis of the attraction problem

On the attraction problem, it is found that:

- In the case of the Hotel attraction, there is at least one opinion with 3484 words, at least one opinion with a minimum of 11 words and an average of 149.10 words.
- In the case of Restaurant attraction, there is at least one opinion with a maximum number of 637 words, there is at least one opinion with a minimum of 6 words and there is an average of 62.06 words.
- In the case of the Attractive attraction, it has a maximum of 670 words in at least one opinion, there is a minimum of 0 words in at least one opinion and an average of 41.43 words.

Table 2 summarizes the characteristics found for attraction.

Table 2

Characteristics of the polarity training set.

Attraction	Opinions	Opinions length		
		Maximum	Minimum	Average
Attractive	5,197 (17.20%)	3,484	11	149.10
Hotel	16,565 (54.82%)	637	6	62.06
Restaurant	8,450 (27.96%)	670	0	41.43

Regarding the opinions length frequency for attraction: on the hotel class, there are 427 opinions with 43 words, 362 opinions with 36 words, and one the largest opinion with 3,484 words (see figure 7).

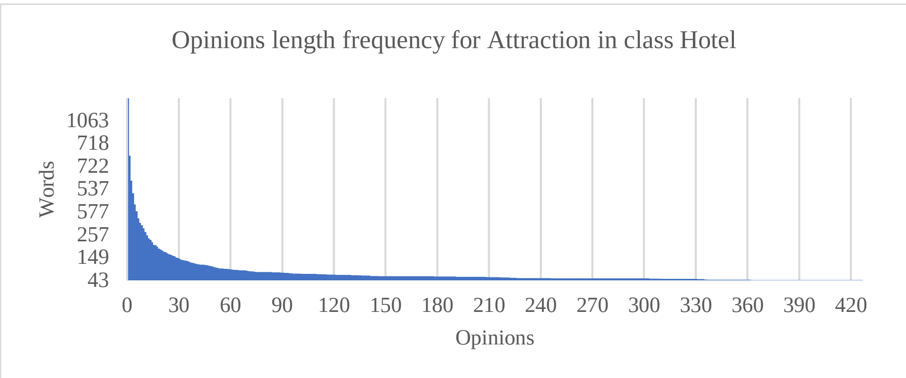

Fig. 7: Frequency of the quantity of words per opinion for Attraction in class Hotel based on the training set.

For the attractive class, there are 181 opinions with 23 words; 172 opinions with 20 words and one opinion with a maximum of 670 words. The minimum length was only observed in two opinions (see figure 8), considering an interval of 23 to 25 words.

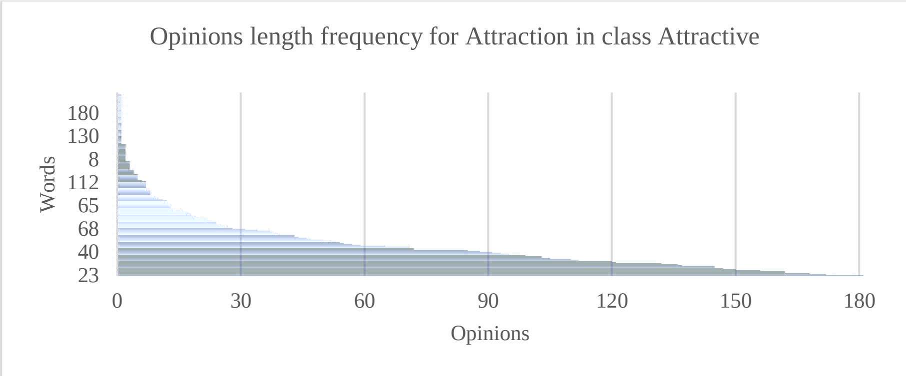

Fig. 8: Frequency of the quantity of words per opinion for Attraction in class Attractive based on the training set.

For the restaurant class, there are 198 opinions with a length of 21 words and 190 opinions with 23 words. For both, the minimum and maximum lengths, there is only one opinion, respectively (see figure 9), considering an interval of 20 to 23 words.

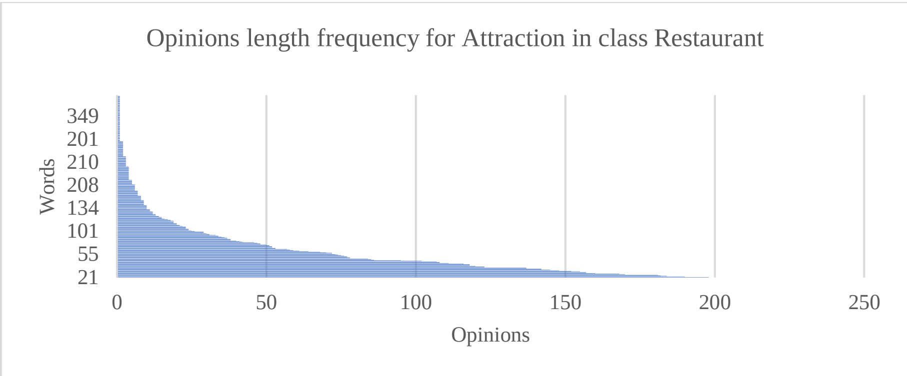

Fig. 9: Frequency of the quantity of words per opinion for Attraction in class Restaurant based on the training set.

In short, there are positive polarities without opinion, they only have a headline; the lowest average length is registered on polarity 5. In the attractive class there are opinions with zero words. The lowest average length is registered on the attractive class.

3. Proposed methodology

The proposed method is based on normalized frequency sets of n-grams per class, the n-grams were extracted with variations of n, considering discontinuity in the chain, which in the following will be called a "jump", e.g.:).

Original opinion: "*Un callejón donde tienes que besar a tu amante por años de felicidad, en el amor es parte de un mito en esta ciudad especial. El callejón estrecho con escalones no es muy especial en sí mismo. Lo que lo hace especial es toda la historia a su alrededor*."

Unigram: {"*Un*", "*callejón*", "*donde*", …, "*alrededor*",…}
Bigram without jump: {"*Un callejón*", "*callejón donde*", …, "*su alrededor*",…}
Bigram with jump 1: {"*Un donde*", "*callejón tienes*", "*donde que*", …}
Bigram with jump 2: {"*Un tienes*", "*callejón que*"}
Trigram without jump: {"*Un callejón donde*", …, "*a su alrededor*"}
Trigram with jump 1: {"*Un donde que*", "*callejón tienes besar*", …}

This approach does not use additional linguistic resources and due to its simplicity.

3.1. Preprocessing

The training sample of the corpus was subjected to a pre-processing in which ten dictionaries were obtained corresponding to the n-grams (unigrams; bigrams without jump and with jump 1, 2 y 3; trigrams without jump and with 1 and 2 jumps; tetragrams without jump and jump 1).

Each title and opinion was preprocessed by replacing line breaks (\n) with whitespaces, it was tokenized using white space as separators, the result obtained was stored in a list, respecting the position of the tokenized words for the generation of the n-grams, in each word the existence of punctuation marks was identified (!"#$%&'()*+, -./:;<=>?@[\]^_`{|}~¿¡), counting the repetitions, adding them to the unigrams' dictionary with their respective repetitions, and removing them from the original string. Since the string is devoid of punctuation marks; accents, umlauts, and other auxiliary signs have been removed.

After data cleaning, the ten dictionaries were created, from the analysis of the opinion and the title; these dictionaries record the repetitions of n-grams (frequency), after that, the repetition frequency of each word was normalized or n-gram, according to the following formula:

$$NFrequency(n\text{-}\ gram \in M) = \frac{\sum_{i=1}^{k} frequency(\text{n-gram}_i)}{Max(Frequency_{class})} \tag{1}$$

Where:
$NFrequency$ is the normalized frequency.
k is the instance of a n-gram in a particular $class$.
Max($\text{frequency}_{\text{class}}$) is the maximum frequency of an n-gram in a particular $class$.

So, the normalized n-gram frequency tables store the n-grams per class for the polarity task shown in table 3.

Table 3

Total of n-grams for each class in the Polarity task.

Class	1	2	3	4	5
Unigram	8,606	9,982	16,605	24,523	39,779
Bigram	43,133	56,230	116,899	217,963	421,204
Bigram jump 1	50,761	67,693	148,803	293,991	594,755
Bigram jump 2	53,810	72,176	161,200	324,254	662,362
Bigram jump 3	54,355	73,193	164,579	332,911	686,962
Trigram	69,612	95,707	228,140	486,857	1,071,660
Trigram jump 1	75,010	105,366	263,756	602,600	1,417,297
Trigram jump 2	75,325	106,297	270,387	631,819	1,525,822
Tetragram	77,094	108,570	275,837	636,908	1,522,496
Tetragram jump 1	77,362	110,060	286,887	693,698	1,748,065

Similarly, the normalized n-gram frequency tables present the total n-grams per class for the Attraction task shown in table 4.

Table 4

Total of n-grams for each class in the Attraction task.

Class	Attractive	Hotel	Restaurant
Unigram	12,075	44,667	17,941
Bigram	73,943	501,048	151,211
Bigram jump 1	93,200	716,048	202,132
Bigram jump 2	100,698	799,032	222,256
Bigram jump 3	102,553	830,149	228,596
Trigram	143,056	1,327,780	327,318
Trigram jump 1	166,096	1,791,950	400,174
Trigram jump 2	166,708	1,949,617	414,440
Tetragram	175,141	1,931,182	420,774
Tetragram jump 1	178,268	2,266,247	447,652

Each dictionary of the class considers the n-gram with the highest number of instances and divides the frequency for each n-gram. This process considered each one of the five classes of polarity and each one of the three classes of Attraction.

3.2. Class prediction

Prediction of the classes of polarity and attraction is gotten for each record of the training and test sets, both, the opinion and the title were processed, performing the same preprocessing described above for both characteristics, and searching on the sets of normalized frequencies of n-grams for each class, adding the normalized frequencies obtained in said search, and finally the class was chosen from the largest cumulative sum, so the quantification formula for each class is:

$$ClassScore = \sum_{i=1}^{k} NFrequency(n\text{-}\ gram \in M)_i \quad (2)$$

Where:
$NFrequency_i$ is the normalized frequency of the n-gram in the class.
k is the instance of the n-gram in the class.

3.3. Cross validation

The model was tested using the k-fold cross validation method, with k = 10; getting the sets or dictionaries of normalized frequencies of n-grams of order 1 (unigrams or terms), 2 (bigrams without jump and, jumps 1, 2 and 3), 3 (trigrams without jump and, jumps 1 and 2), 4 (tetragrams without jump and, jump 1), and 5 (pentagrams without jump) for 10 partitions of equal size in the corpus; associating the n-grams with the classes where they had instances.

After that, the polarity and attraction were inferred for the 9 remaining partitions in a total of 10 exercises, one for each partition, to finally average the results shown in figure 10:

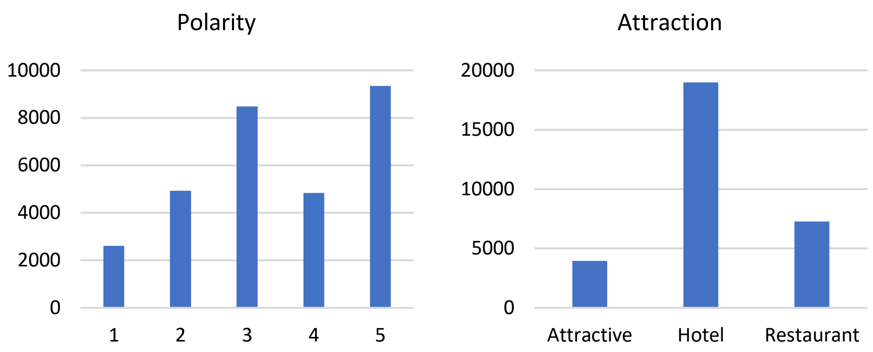

Fig. 10: Average of the results per partition using the 10-fold cross validation method in the unbalanced training set. Source: Own elaboration, based on the corpus' training set.

Likewise, the proposed methodology was tested considering data balance for the polarity classes, preparing a training subset of the original dataset, where 547 opinions of each of the 5 classes are considered and 10 partitions of 273 opinions are generated. After the cross-validation experiment, it is seen on the results shown on figure 11, that the prediction got a bias in polarity class 2, and its effects on the prediction of the attraction task:

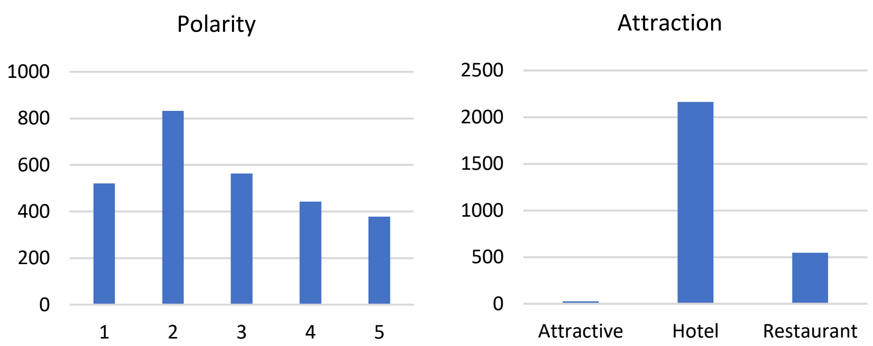

Fig. 11: Average results per partition using the 10-fold cross validation method with a balanced sample of the training set.

3.4. Measuring

Following the call published on the Rest-Mex 2022 website, the results were calculated as described below. Precision for each of the classes in polarity attraction is obtained using the formula in eq. 3:

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

Where:
TP = Total of True Positives.

FN = Total of False Negatives.

Likewise, the F1-Score measure is calculated for each of the classes in polarity and attraction using the formula in eq. 5:

$$F1\_Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{5}$$

From these values, a Macro-F1 measurement is obtained for attraction, using the formula in eq. 6:

$$Macro\_F1_k = \frac{F1_A(k) + F1_H(k) + F1_R(k)}{3} \tag{6}$$

Where, for the attraction case:

$F1_A$ = F1 Score for Attractive.

$F1_H$ = F1 Score for Hotel.

$F1_R$ = F1 Score for Restaurant.

For the case of polarity, the mean absolute error was obtained using the formula in eq. 7:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{7}$$

Finally, a rank measurement for polarity and attraction is obtained, using the formula in eq. 8:

$$Sentiment_{Res_k} = \frac{\frac{1}{1+MAE_k} + MacroF1_k}{2} \tag{8}$$

With this, the results obtained using the proposed methodology, for two exercises, are observed in Table 5 for the Polarity task and in Table 6 for the Attraction task.

Table 5

Measures for the Polarity task. Source: Rest-Mex 2022.

Measure	Method 2	Method 1
Final Rank	0.7035669036	0.6668485025
MAE	0.6361879734	0.9714020714
Accuracy	64.9868604112	56.2528984387
Macro F-measure	0.3426872166	0.2926745962
Macro Recall	0.3580585827	0.3395343446
Macro Precision	0.4309377624	0.4365770925
Class 1 F-measure	0.2077493800	0.1557767200
Class 2 F-measure	0.1723237600	0.1245640300
Class 3 F-measure	0.2638554200	0.2213629800
Class 4 F-measure	0.2389672800	0.1765693200
Class 5 F-measure	0.8305402400	0.7850999400
Class 1 Recall	0.1316614400	0.0875912400
Class 2 Recall	0.1084634300	0.0738770700
Class 3 Recall	0.2822164900	0.1932489500
Class 4 Recall	0.4383259900	0.4730215800
Class 5 Recall	0.8296255500	0.8699328000
Class 1 Precision	0.4921875000	0.7031250000
Class 2 Precision	0.4190476200	0.3968254000
Class 3 Precision	0.2477375600	0.2590497700
Class 4 Precision	0.1642591800	0.1085431300
Class 5 Precision	0.8314569500	0.7153421600

Table 6

Measures for the Attraction task. Source: Rest-Mex 2022.

Measure	Method 2	Method 1
Final Rank	0.7035669036	0.6668485025
Accuracy	82.0219508425	84.6885144535
Macro F-measure	0.7959570864	0.8264438096
Macro Recall	0.9133563075	0.8461729126
Macro Precision	0.7497692982	0.8284529852
Class Hotel F-measure	0.8597257600	0.8920363400
Class Restaurant F-measure	0.6948121600	0.7219225100
Class Attractive F-measure	0.8333333300	0.8653725700
Class Hotel Recall	0.7551694700	0.8485002500
Class Restaurant Recall	0.9867886200	0.8849699400
Class Attractive Recall	0.9981108300	0.8050485400
Class Hotel Precision	0.9978873200	0.9402816900
Class Restaurant Precision	0.5361678600	0.6096079500
Class Attractive Precision	0.7152527100	0.9354693100

4. Conclusions

The results obtained when compared with other methods used by other participants showed that the proposed methodology, due to its simplicity, does not obtain the best scores compared to more sophisticated systems that use more resources, but it can be positioned as a reference system, or baseline, for these tasks. The proposed methodology scored consistent results regardless of language and data balance; low computational complexity; and the use of normalized sets of n-grams allowed their reuse as modules that did not require regeneration, for alternative class prediction.

5. References

[1] J. Abreu, P. Mirabal, & A. Ballester-Espinosa. (2021). Cascade of biased two-class classifiers for multi-class sentiment analysis. CEUR Workshop Proceedings, 2943(September 2021), 185-191. http://hdl.handle.net/10045/117964.

[2] M. Á. Álvarez-Carmona, R. Aranda, S. Arce-Cardenas, D. Fajardo-Delgado, R. Guerrero-Rodríguez, A. P. López-Monroy, J. Martínez-Miranda, H. Pérez-Espinosa, A. Y. Rodríguez-González: Overview of Rest-Mex at IberLEF 2021?: Recommendation System for Text Mexican Tourism. Sociedad Española para el Procesamiento del Lenguaje Natural (2021). https://doi.org/10.26342/2021-67-14.

[3] M. Á. Álvarez-Carmona, R. Aranda, S. Arce-Cardenas, D. Fajardo-Delgado, R. Guerrero-Rodríguez, A. P. López-Monroy, J. Martínez-Miranda, H. Pérez-Espinosa, A. Y. Rodríguez-González: Overview of Rest-Mex at IberLEF 2022: Recommendation System for Text Mexican Tourism. Sociedad Española para el Procesamiento del Lenguaje Natural (2022).

[4] G. Carmona-Sánchez, Á. Carmona, & M. Álvarez-Carmona. (2021). Naive features for sentiment analysis on mexican touristic opinions texts. CEUR Workshop Proceedings, 2943(September 2021), 118-126.

[5] K. Dashtipour, M. Gogate, A. Gelbukh, A. Hussain. (2021). Sentiment Analysis To Adopting Transition Point Technique for Persian Sentiment Analysis. ICOTEN, Jun 2021, Glasgow, France. ⟨hal-03258136⟩. https://hal.archives-ouvertes.fr/hal-03258136/

[6] Decide4AI. (2019). Procesamiento del Lenguaje Natural (PLN o NLP): qué es y para qué se utiliza. mayo 9, 2022, de Decide Sitio web: https://decidesoluciones.es/procesamiento-del-lenguaje-natural-pln-o-nlp-que-es-y-para-que-se-utiliza/
[7] F. Madrid, G. Godínez, (2022). Turismo y pandemia en méxico. Centro de Investigación y Competitividad Turística Anáhuac (Cicotur). https://www.anahuac.mx/mexico/cicotur/sites/default/files/2022-05/TurismoYPandemia.pdf
[8] G. V. Medina, D. Irazu, & H. Farfas. (2021). DCI-UG participation at REST-MEX 2021: A Transfer Learning Approach for Sentiment Analysis in Spanish. CEUR-WS.org/Vol-2943 - Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021) (2021-09-16). http://ceur-ws.org/Vol-2943/
[9] E. R. Reyes, (2021). Techkatl: A sentiment analysis model to identify the polarity of mexican's tourism opinions. CEUR Workshop Proceedings, 2943(September), 171-178. http://ceur-ws.org/Vol-2943/restmex_paper7.pdf
[10] A. Romero-Cantón, & R. Aranda, (2021). Sentiment classification for mexican tourist reviews based on k-nn and jaccard distance. CEUR Workshop Proceedings, 2943(September), 179-184. http://ceur-ws.org/Vol-2943/restmex_paper8.pdf
[11] G. Sidorov, (2013). Non-linear construction of n-grams in computational linguistics: syntactic, filtered, and generalized n-grams. https://www.cic.ipn.mx/~sidorov/Nonlinear_construction_of_n_grams.pdf
[12] G. Sidorov, (2013). Syntactic Dependency Based N-grams in Rule Based Automatic English as Second Language Grammar Correction. International Journal of Computational Linguistics and Applications, 4(2), 169-188. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.644.907&rep=rep1&type=pdf%0Ahttps://www.ijcla.org/2013-2/
[13] G. Sidorov, F. Velasquez, E. Stamatatos, & A. Gelbukh. (2012). La sincronicidad y los códigos: hilos de la Práctica Docente Supervisada. InterSedes: Revista de Las Sedes Regionales, XIII(25), 82-96. https://revistas.ucr.ac.cr/index.php/intersedes/article/view/1991
[14] Staff Entorno y Futuro. (2022). 2022 se vislumbra como un oasis para el turismo en México. mayo 20, 2022, de Centro para el Futuro de las Ciudades Sitio web: https://futurociudades.tec.mx/es/2022-se-vislumbra-como-un-oasis-para-el-turismo-en-mexico.
[15] M. Toledo-Acosta, B. Martínez-Zaldivar, A. Ehrlich-López, E. Morales-González, D. Torres-Moreno, & J. Hermosillo-Valadez, (2021). Semantic representations of words and automatic keywords extraction for sentiment analysis of tourism reviews. CEUR Workshop Proceedings, 2943(September 2021), 149-164. http://ceur-ws.org/Vol-2943/restmex_paper5.pdf
[16] J. Vásquez, H. Gómez-Adorno, & G. Bel-Enguix. (2021). Bert-based approach for sentiment analysis of spanish reviews from tripadvisor. CEUR Workshop Proceedings, 2943(September 2021), 165-170. http://ceur-ws.org/Vol-2943/restmex_paper6.pdf