www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 16(1), Jan-April 2025, 132-139. ISSN: 2007-1558.
https://doi.org/10.61467/2007.1558.2025.v16i1.547

Improving Sentiment Polarity Identification on Twitter Using Metaclassifiers

Wendy Morales Castro[1], Rafael Guzman-Cabrera[1,], Tirtha Prasad-Mukhopadhyay[1], Armando Pérez-Crespo[1], Marco Bianchetti[1]*
[1] División de Ingenierías, Universidad de Guanajuato campus Irapuato-Salamanca, GTO, México.
*E-mail: guzmanc@ugto.mx

Abstract. The exponential growth of social networking platforms has led many researchers to focus on ways of mining information from them. In this paper, we will use texts from social media in conjunction with techniques of Natural Language Processing to design a system that helps business organizations to identify polarity indicators from customer feedback. In this paper, we analyze tweets related to perceptions of an airline company, and detect the polarity of such tweets, using preprocessing and processing techniques common to the area, and to later incorporate the same techniques, in a new methodology that consists of the incorporation of lexical resources (LR) and metaclassifiers to support the said task, thereby achieving a decision system with greater precision. In the present work, relevant results are reported in the area of NLP, making use of pre-processing and processing techniques known within the area, the main idea is to find the best classification scenario and increase the classification precision, for this the incorporation of lexical and metaclassifier resources was carried out.

Keywords: Metaclassifier, Lexical Resources, Classifier Scenarios

Article Info
Received November 13, 2024
Accepted March 12, 2025

1 Introduction

Recently, social media platforms have emerged as digital spaces in which a greater diversity of users publish and disseminate information spontaneously and voluntarily, for even wider outreach and impact that supersedes usage habits in the previous decades. This information represents an emerging challenging sector in the context of big data analytics. The information segments have become more than just a mean of communication. They have evolved in such a way that they now not only influence personal and social connections, but also influence the way business is done.

In this context Twitter definitely occupies a frontline position with 328 million users [1]. It is easy to see that, in order to carry out the automatic processing of the large number of opinions that are registered every day in this and other social networks, it is necessary to have tools that allow the tasks to be carried out automatically. Data mining is framed within the field of Natural Language Processing, and has been one of the most active processing areas since the early 2000s [2].

Again, the fundamental goal of sentiment analysis is to define automatic tools capable of extracting subjective information from natural language text, such as opinions or sentiments, in order to create structured and actionable knowledge that could be used by a decision-making system [3].

An opinion is a positive or negative evaluation about a product, service, organization, person or any other type of entity about which some sentiment can be expressed [4]. Due to the importance that sentiment analysis has for business and society, it has been extended from computer science to management and social sciences [5]. This is because, if opinions in the network are collected successfully and analyzed, they allow us to not only understand and explain many complex social phenomena, but also to predict them.

Within the state-of-the-art there are a large number of works, based on different techniques based on Sentiment Analysis, this because it is still an open research area, and can be implemented in various fields, from the political field [6], the social field [7]in the field of health [8] and in the field of valuation of services or products, task on which the present work focuses, some of the works found related to this area are presented in the following paragraphs.

Within [9] the authors implement sentiment analysis using opinions that customers made within Amazon, in order to detect the feeling of irony to verify whether these types of comments affect purchases made by other users. They make use of a corpus previously extracted by them, to then perform the preprocessing of the data, by using different values of n-grams and POS-grams, to then perform the classification with 3 types of classifiers Naïve Bayes (NB), Support Vector Machines (SVM), and Decision Tree (DT). Finally, they obtain an accuracy of 82.83% and define the values of the n-grams and Pos-grams with which the mentioned percentage is obtained. In [10] the authors present a novel approach to carry out the task of sentiment analysis on the Twitter platform, within the work to discover the sentiment, the opinion words are extracted which turn out to be a combination of adjectives together with verbs and adverbs. The authors perform a corpus-based method which is used to find the semantic orientation of adjectives and in turn make use of the dictionary-based method which is used to find the semantic orientation of verbs and adverbs. They continue with calculating the overall sentiment of the tweet by using linear equations incorporating emotion intensifiers, finally they calculate the score for the overall sentiment of the tweet and each of the tweets are classified as positive, neutral or negative based on the calculated score.

In the same vein, at work [11] the authors analyze data collected from samples with specific hashtags from Twitter to search for some items of interest such as computers, cars, and televisions, among others, to perform an analysis to effectively detect market intelligence and help support consumers' decision making. In [12], the authors exploit sentiment analysis to investigate to what extent Twitter comments affect movie promotions and sales. They determine this by estimating a dynamic panel data model.
As can be seen, Twitter has become one of the most widely used social networks for sentiment analysis.

Each researcher implements different pre-processing and processing methods, as appropriate to the requirement of the analysis. However, very little research focus on parsing lexical dictionaries and classifiers to perform the task of sentiment analysis. In the present work, our efforts are focused on the development of such a tool, using which would enable classification of Tweets. A corpus based on Twitter comments on Airlines such as American Airlines, JetBlue and United has been identified for our specific purpose. The following section presents the characteristics of the chosen dataset, as well as the methodology used to contribute to the solution to this problem, namely the development of a decision-making analytics as a tool for understanding or impacting consumer behavior.

2 Methodology

Figure 1 shows the general methodology proposed for this work. The main objective of this part of preparation was to find the best possible classification scenario with the best match in algorithms; for this task the methodology was separated into 2 phases: Phase one is created to obtain the 3 algorithms with the best performance in the classification task, with their respective preprocessing and processing stages, and to observe if there was a difference between the action of balancing the dataset and that of keeping it unbalanced among its classes. In phase two is kept the configuration that gave the best results in phase 1, incorporating lexical resources (LRs) to increase the precision, and finally, in order to try to increase the percentage of precision of the system, used two metaclassifiers.

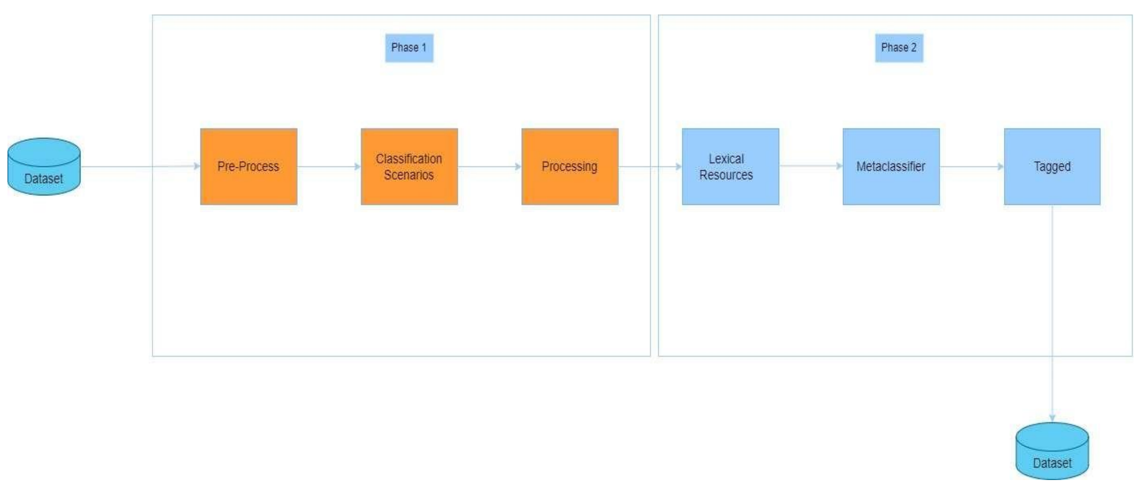

Fig. 1. Methodology diagram.

In the phase one for development the evaluation of the proposed methodology, the database 'Twitter US Airline Sentiment Analysis' was used. The corpus contains a set of Tweets published in February 2015 regarding the performance of some well-known airlines in the United States. The corpus is composed of 14,485 tweets manually labeled in three categories: Positive (2332 instances), Negative (9088 instances) and Neutral (3065 instances). In turn, some data preprocessing techniques, such as converting the data from uppercase to lowercase, removing stopwords, tokenization of the data, lemmatization and finally obtaining the data information gain (IG) [13], was executed on the data. In this part of the research method, 5 algorithms, whose outstanding results are widely reported in the sentiment analysis task, were also used in the processing stage: SVM [14], [15], Nearest Neighbors (KNN) [16], Logistic Regression (RL) [17], Naive Bayes (NB) [18] and a type of Decision Tree (J48) [19].
To obtain the values of the parameters of each classifier, two classification scenarios were used: Cross Validation Scenario (CV) [20] and Training and Test Set Scenario (TTS) [21]. In order to maximize the performance of the model of the given set. Within phase two, LRs were use, the LRs are a set of terms in which each term is associated with a sentiment. The label of the sentiment can be represented with polarity (Positive, Negative or Neutral) or with some numerical value (1,-1,0); this value reflects the affective likelihood of any behavioral reflexes or reception profile of the messages [22]. We also made use of a dataset containing positive and negative words called Sense. The dataset consists of 4699 positive words and 4722 negative words.

Another LR used was that of author Jeffrey Breen called Opinion Lexicon. This list was compiled over several years from work done by the authors in reference [23]. In their work, prompted by the demands of e-commerce, they sought to perform text mining and summarize the reviews of a product. Since hundreds or even thousands of reviews of a product were generated daily, their goal was to find the reviews with the greatest impact. The dataset is made up of 2005 positive words and a total of 4781 negative words.
Phase two of the methodology involved two Metaclassifiers too, Stacking metaclassifier [24] and Voting metaclassifier [25]. These metaclassifiers were used to improve prediction performance, because metaclassifiers combine multiple predictions from base classifiers to produce a final prediction, and this can help improve predictive performance, and compensate for individual weaknesses of base classifiers. In some cases, this can improve ranking performance compared to using a single algorithm. These metaclassifiers were also evaluated with the two classification scenarios mentioned above.

3 Results

The final values for each classifier were:

- CV: The set was created with 10 folds of data clustering.
- TTS: 80% of the set was assigned to system training and 20% to testing.
- SVM: c=1.0 with épsilon=1.0E-12
- RL: ridge parameter of 1.0E-8
- KNN: K=2
- DT(J48): c=0.25

The values of the classification scenarios were selected based on the size of the dataset, the amount of data available, the complexity of the models, and the computational resources available. With the 10 folds in the CV scenario, you have a good amount of data for training and testing in each iteration. Whereas for the TTS scenario, allocating 80% of the data to training ensures that there is enough data to train the model effectively, and equally 20% helps to evaluate performance reliably. Helping to avoid over- and under-fitting issues.

The values of the classifier parameters were based on the classification scenarios and in turn were varied until the final value that maximizes the performance of the model was obtained.

3.1 Experiment 1 (Phase I)

Results are presented using the unbalanced dataset with the five automatic classification methods: first at the baseline, using the data without any preprocessing, then with elimination of stopwords, and finally, lemmatization and the use of IG. Two classification scenarios are used: CV and TTS, as shown in Table 1.

Table 2 presents the results obtained for the Balanced database in the CV and TTS scenarios. For the balanced database, 2332 tweets from each class were used, because the class with fewer attributes had that number of tweets and it was chosen to balance all classes with this number of tweets.

Table 1. Precision results obtained from the unbalanced database. (CV & TTS).

CV-Unbalanced				
Algorithm	**Baseline**	**Dataset with StopWords Removal**	**Dataset with StopWords Removal and Lemmatization**	**Dataset with StopWords Removal, Lemmatization and IG**
NAIVE BAYES	-	0.611	0.603	0.648
SVM	**0.692**	**0.719**	**0.711**	**0.723**
J48	0.634	0.641	0.61	0.657
RL	-	0.718	0.708	0.721
KNN	0.579	0.5734	0.524	0.647
TTS-Unbalanced				
NAIVE BAYES	-	0.602	0.6	0.654
SVM	**0.682**	**0.718**	**0.702**	**0.729**
J48	0.627	0.644	0.605	0.666
RL	-	0.704	0.697	0.719
KNN	0.558	0.672	0.51	0.651

Table 2. Precision results obtained from the balanced database. (CV & TTS).

CV-Balanced				
Algorithm	**Baseline**	**Dataset with StopWords Removal**	**Dataset with StopWords Removal and Lemmatization**	**Dataset with StopWords Removal, Lemmatization and IG**
NAIVE BAYES	0.555	0.641	0.648	0.699
SVM	**0.617**	**0.711**	**0.702**	**0.741**
J48	0.575	0.633	0.66	0.703
RL	0.59	0.656	0.668	-
KNN	0.507	0.555	0.604	0.667
TTS-Balanced				
NAIVE BAYES	0.549	0.638	0.658	0.699
SVM	**0.63**	**0.694**	**0.7**	**0.747**
J48	0.546	0.6222	0.666	0.705
RL	0.588	0.634	0.628	-
KNN	0.508	0.536	0.591	0.681

As can be seen in the tables, the 3 algorithms with the best results in the classification of tweets were (SVM, RL, J48), these were used in the following experiments.

3.2 Experiment 2 (Phase II)

For this experiment we made use of the configuration with which the best results were obtained in experiment 1; to this configuration, we added the two LR mentioned above: the LR "Sense" and "Opinion Lexicon". First, the results obtained by using only the list of positive words are presented in Figure 2, then with only the list of negative words in Figure 3. Finally, both positive and negative lists of both LRs were combined, as shown in Figure 4.

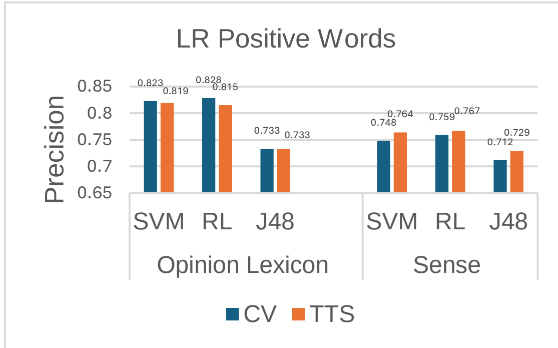

Fig. 2. Graph of the results of the LR of positive words for both lexical resources.

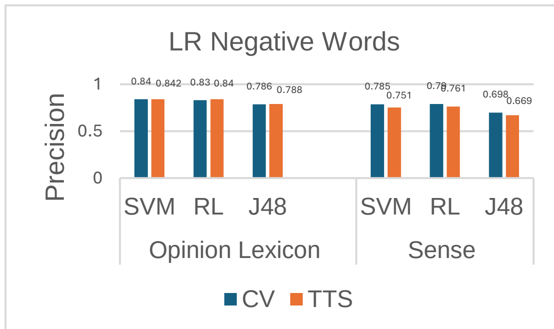

Fig. 3. Graph of the results of the LR of negative words for both lexical resources.

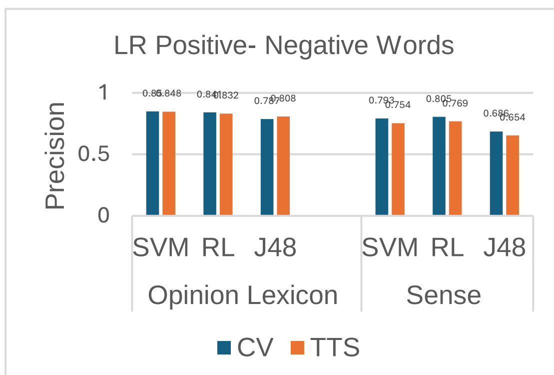

Fig. 4. Graph of the results of the LR of positive and negative words for both lexical resources.

As it can be seen, the best results were obtained by the LR list of negative words and the combination of the list of positive and negative words of the LR Opinion Lexicon, all reaching a precision of 84% - 85%, with an increment of up to 10% compared to the results obtained in phase 1,with the SVM algorithm. It has to be taken into account that within this experiment all the preprocessing stages were implemented and IG was used.

3.3 Experiment 3 (Phase II)

In the third phase, two meta-classifiers were implemented: Stacking and Voting, which used the best results of the previous experiment. For the Stacking type meta-classifier, a set of tests were performed with different algorithms (SVM, RL, KNN, Perceptron [26], J48, Naive Bayes) to choose the algorithm that was responsible for taking up the classifier inputs and estimating the output of the set of the 3 algorithms. And for the Voting algorithm, an analysis was performed that combined the 3 algorithms used to obtain a single one that could offer better performance. The methods that were used for the Voting metaclassifier were:

combination of majority of votes, calculating percentage of probability and the maximum and minimum probability. The goal was to perform a detailed analysis and make a comparison of precision between the two metaclassifiers used.
Within the results of the Stacking metaclassifier, the best result was obtained by the Perceptron algorithm, reaching a precision in the TTS scenario of 85.5%. This is shown in Figure 5. Also, in this figure are shown the results obtained by the Voting metaclassifier, and it can be seen that the most accurate value is given by the Voting metaclassifier of the Majority Vote type with a precision of 86.1% in the TTS scenario.

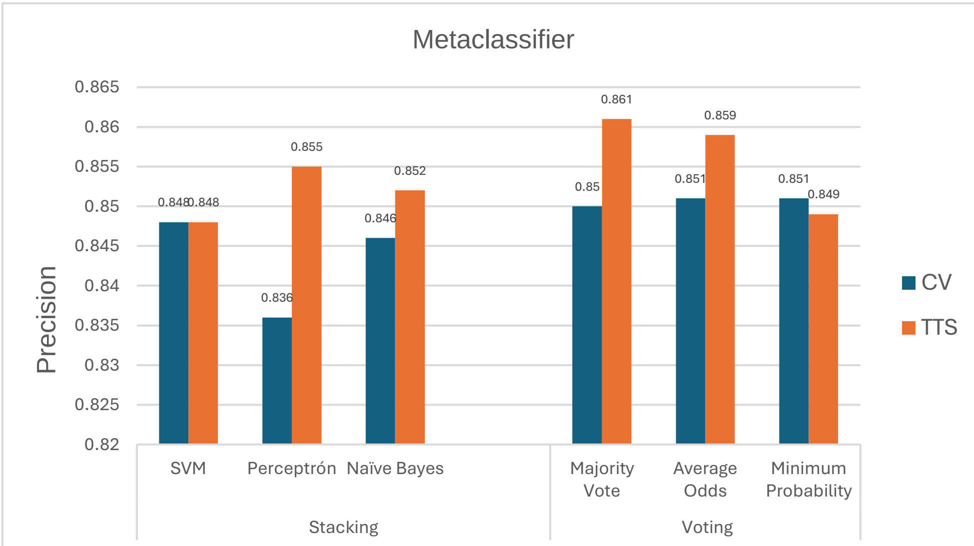

Fig. 5. Metaclassifier results.

4 Conclusions

The results obtained show the growth of the system throughout the experiments, improving precision from 50% (lowest percentage obtained) to 86% (highest percentage obtained). This was due to the types of preprocessing performed. It could be observed that the precision increased with the number of preprocessing that were applied to the corpus: 86% was achieved with the use of all types of preprocessing, with the obtaining of information gain, with the use of the LR and finally with the implementation of the metaclassifier, so that at this point it is concluded that the system designed in this thesis achieved the objectives set. On concluding the analysis of the experiments carried out, we obtained a clearly feasible model of a predictive system capable of automatically tagging Tweets downloaded in Real Time, following the implementation of the aforementioned methodology with exhaustive analysis, such as allowed us to identify the best possible scenario to carry out its identification.

The task of sentiment analysis has been growing exponentially in recent years due to the great demand that it has; this task is not trivial, such systems are designed to cover different areas within the taxonomy of feelings, i.e. the model can not only be used for the service industries, but in areas such as health, political psephology, social networking and advertising, among others.

References

[1] S. Biffignandi, A. Bianchi, and C. Salvatore, "Can Big Data provide good quality statistics? A case study on sentiment analysis on Twitter data," 2018, Accessed: Oct. 12, 2022. [Online]. Available: https://stats.oecd.org/glossary/detail.asp?ID=5150

[2] D. Bamman and N. A. Smith, "Contextualized Sarcasm Detection on Twitter," 2015, Accessed: Dec. 05, 2021. [Online]. Available: http://en.

[3] B. Priego Sánchez, D. Pinto, B. Priego Sánchez, and D. Pinto, "Idiom Polarity Identification using Contextual Information," *Computación y Sistemas*, vol. 22, no. 1, pp. 27–33, 2018, doi: 10.13053/CYS-22-1-2791.

[4] E. Cambria, B. Schuller, Y. X.-I. I. Systems, and U. 2013, "New avenues in opinion mining and sentiment analysis," *ieeexplore.ieee.org*, 2021, Accessed: Dec. 05, 2021. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6468032/

[5] L. Joyanes Aguilar, "Big Data: Análisis de grandes volúmenes de datos en organizaciones," *J Chem Inf Model*, vol. 1, p. 428, 2013.
[6] Caicedo Ortiz, "Análisis del sentimiento político mediante la aplicación de herramientas de minería de datos a través del uso de redes sociales." Accessed: Oct. 27, 2022. [Online]. Available: https://repository.javeriana.edu.co/handle/10554/20516
[7] T. I. Garcés Chaparro, "Análisis de sentimientos en redes sociales orientado a la percepción de la calidad de servicios de internet, redes móviles, tv cable y electricidad," 2019, Accessed: Oct. 27, 2022. [Online]. Available: https://repositorio.unab.cl/xmlui/handle/ria/17963
[8] D. Forero-Chaux and J. Vergara-Rojas, "Método automático para apoyar en la identificación de síntomas de depresión mediante el análisis en narrativa oral," 2021, Accessed: Oct. 29, 2022. [Online]. Available: https://repository.ucatolica.edu.co/handle/10983/26291
[9] A. Reyes and P. Rosso, "Making objective decisions from subjective data: Detecting irony in customer reviews," *Decis Support Syst*, vol. 53, no. 4, pp. 754–760, 2012, doi: 10.1016/j.dss.2012.05.027.
[10] A. Kumar and T. M. Sebastian, "Análisis de sentimiento en twitter," *... Revista de Problemas de Ciencias de la Computación (IJCSI)*, 2012, Accessed: Apr. 09, 2024. [Online]. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ce41114db57c2d58ce5592f274e2eede914bd9e9
[11] Y. M. Li and T. Y. Li, "Deriving market intelligence from microblogs," *Decis Support Syst*, vol. 55, no. 1, pp. 206–217, Apr. 2013, doi: 10.1016/J.DSS.2013.01.023.
[12] H. Rui, Y. Liu, and A. Whinston, "Whose and what chatter matters? the effect of tweets on movie sales," *Decis Support Syst*, vol. 55, no. 4, pp. 863–870, 2013, doi: 10.1016/j.dss.2012.12.022.
[13] P. G. Otero and M. G. González, "Técnicas de Procesamiento del Lenguaje Natural en la recuperación de información," 2012, Accessed: Feb. 13, 2024. [Online]. Available: https://minerva.usc.es/xmlui/handle/10347/17692
[14] C. González, G. de S. I.-G. de Sistemas, and undefined 2011, "SVM: Máquinas de vectores soporte," *infor.uva.es*, Accessed: Feb. 13, 2024. [Online]. Available: http://www.infor.uva.es/~calonso/MUI-TIC/MineriaDatos/SVM.pdf
[15] E. L.-T. de F. de G. en Matemáticas, undefined Universidad, and undefined 2017, "Introducción a las máquinas de vector soporte (SVM) en aprendizaje supervisado," *core.ac.uk*, Accessed: Feb. 13, 2024. [Online]. Available: https://core.ac.uk/download/pdf/289986621.pdf
[16] M. N. Vilema and R. M. Arias, "Predicción de clientes potenciales utilizando K vecino más cercano en el área de negocios de la Cooperativa Riobamba.," *45.184.102.148*, 2021, doi: 10.47187/perspectivas.4.1.150.
[17] I. M. Camejo, … G. A. M.-R. C., and undefined 2015, "Desempeño de algoritmos de minería en indicadores académicos: Árbol de Decisión y Regresión Logística," *scielo.sld.cu*, Accessed: Feb. 13, 2024. [Online]. Available: http://scielo.sld.cu/scielo.php?pid=S2227-18992015000400008&script=sci_arttext
[18] L. Sinaga and S. Suwilo, "Análisis de la clasificación y el algoritmo Naïve Bayes k-vecino más cercano en minería de datos," *Serie de Conferencias IOP: Materiales ...*, 2020, doi: 10.1088/1757-899X/725/1/012106.
[19] N. Bhargava, G. Sharma, R. Bhargava, and M. Mathuria, "Decision Tree Analysis on J48 Algorithm for Data Mining," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 6, p. 2277, 2013, Accessed: Feb. 13, 2024. [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka.
[20] J. Águila, M. Trabajo, and A. Solanas Gómez, "Aprendizaje supervisado en conjuntos de datos no balanceados con redes neuronales artificiales: métodos de mejora de rendimiento para modelos de clasificación," 2017, Accessed: Feb. 13, 2024. [Online]. Available: https://openaccess.uoc.edu/handle/10609/64768
[21] L. Pérez-Planells, J. Delegido, … J. R.-C.-R., and undefined 2015, "Análisis de métodos de validación cruzada para la obtención robusta de parámetros biofísicos," *ojs.upv.es*, Accessed: Feb. 13, 2024. [Online]. Available: http://ojs.upv.es/index.php/raet/article/view/4153
[22] G. Real-Flores, B. García-Mendoza, E. Calderón-Casanova, G. Ramírez-De-La-Rosa, and E. Villatoro-Tello, "Generación y enriquecimiento automático de recursos léxicos para el análisis de sentimientos Automatically Generating and Enriching Lexical Resources for Sentiment Analysis," *Computing Science*, vol. 134, pp. 163–176, 2017.
[23] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," 2004.
[24] I. Bonet, A. Rodríguez, M. García, R. G.-C. y Sistemas, and undefined 2012, "Combining classifiers for Bioinformatics," *scielo.org.mx*, vol. 16, no. 2, pp. 191–201, 2012, Accessed: Feb. 13, 2024. [Online]. Available: https://www.scielo.org.mx/scielo.php?pid=S1405-55462012000200006&script=sci_abstract&tlng=en

[25] A. Verdecia-Cabrera, … I. F.-B.-R. C., and undefined 2020, "Aprendiendo con meta-clasificadores a partir de flujos de datos no estacionarios," *scielo.sld.cu*, Accessed: Feb. 13, 2024. [Online]. Available: http://scielo.sld.cu/scielo.php?pid=S2227-18992020000400020&script=sci_arttext&tlng=en

[26] H. D. Block, "The perceptron: A model for brain functioning. I," *Rev Mod Phys*, vol. 34, no. 1, pp. 123–135, 1962, doi: 10.1103/REVMODPHYS.34.123.