Article

Bayesian Constitutionalization: Twitter Sentiment Analysis of the Chilean Constitutional Process through Bayesian Network Classifiers

Gonzalo A. Ruz ^{1,2,*} , Pablo A. Henríquez ³ and Aldo Mascareño ⁴

¹ Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Diagonal Las Torres 2640, Peñalolén, Santiago 7941169, Chile

² Center of Applied Ecology and Sustainability (CAPES), Santiago 8331150, Chile

³ Facultad de Economía y Empresas, Universidad Diego Portales, Av. Sta. Clara 797, Huechuraba, Santiago 8581169, Chile; pablo.henriquez@udp.cl

⁴ Centro de Estudios Públicos, Monseñor Sótero Sanz 162, Providencia, Santiago 7500011, Chile; amascareno@cepchile.cl

* Correspondence: gonzalo.ruz@uai.cl

Citation: Ruz, G.A.; Henríquez, P.A.; Mascareño, A. Bayesian

Constitutionalization: Twitter Sentiment Analysis of the Chilean Constitutional Process through Bayesian Network Classifiers.

Mathematics **2022**, *10*, 166. <https://doi.org/10.3390/math10020166>

Academic Editors: Victor Leiva and Liangxiao Jiang

Received: 6 December 2021

Accepted: 5 January 2022

Published: 6 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Constitutional processes are a cornerstone of modern democracies. Whether revolutionary or institutionally organized, they establish the core values of social order and determine the institutional architecture that governs social life. Constitutional processes are themselves evolutionary practices of mutual learning in which actors, regardless of their initial political positions, continuously interact with each other, demonstrating differences and making alliances regarding different topics. In this article, we develop Tree Augmented Naive Bayes (TAN) classifiers to model the behavior of constituent agents. According to the nature of the constituent dynamics, weights are learned by the model from the data using an evolution strategy to obtain a good classification performance. For our analysis, we used the constituent agents' communications on Twitter during the installation period of the Constitutional Convention (July–October 2021). In order to differentiate political positions (left, center, right), we applied the developed algorithm to obtain the scores of 882 ballots cast in the first stage of the convention (4 July to 29 September 2021). Then, we used *k*-means to identify three clusters containing right-wing, center, and left-wing positions. Experimental results obtained using the three constructed datasets showed that using alternative weight values in the TAN construction procedure, inferred by an evolution strategy, yielded improvements in the classification accuracy measured in the test sets compared to the results of the TAN constructed with conditional mutual information, as well as other Bayesian network classifier construction approaches. Additionally, our results may help us to better understand political behavior in constitutional processes and to improve the accuracy of TAN classifiers applied to social, real-world data.

Keywords: Bayesian networks; TAN classifiers; evolution strategy; Twitter data; constitution making

1. Introduction

On the afternoon of 18 October 2019, serious riots took place in different parts of Santiago, Chile. Within hours, the situation escalated into the most important Chilean political event of the 21st century, with massive and violent protests that lasted for about a month. In order to avoid the collapse of the democratic system, the government and political forces from across the ideological spectrum called for a peace agreement and a new constitution. This paved the way for the institutional organization of the demands through a new constitution [1,2].

After the reorganization of the electoral calendar due to COVID-19, the first session of the Constitutional Convention, in charge of writing the new Chilean Constitution, took place on 4 July 2021. The convention consisted of 155 members, with equal numbers of seats

for women and men, and with a quota of 17 seats reserved for native peoples. Politically, it was composed of 37 right-wing, 36 center-left, 54 left-wing, and 11 independent members (mostly left-wing members), as well as 17 representatives of native peoples.

The convention agenda provided three months for the creation of the rules of procedures and nine months for the elaboration of the substantive content of the constitution. Thus, the new constitution should be delivered by the convention in July 2022 and ratified (or rejected) by the people in a final plebiscite. On 29 September 2021, the convention approved its General Rules, and on 18 October 2021 (exactly two years after the 2019 riots), it began to discuss the substantive contents of the constitution.

In this article, we focus on analyzing the opinions posted on Twitter by the convention members during the constitutional process. In particular, we used a non-structured dataset composed of the Twitter posts of 135 out of the 155 members (20 did not have a Twitter account) of the Constitutional Convention between 4 July and 27 October 2021. This amounts to 37,814 tweets. In order to differentiate the members' political positions (left, center, right), we applied a nominate algorithm [3] to obtain the scores of 882 ballots cast in the first stage of the convention (4 July to 29 September 2021). This algorithm was originally developed for the analysis of U.S. congressional votes and has been applied in multiple political contexts. As a multidimensional scaling technique, its fundamental objective is to project high-dimensional data (voting patterns) to a low-dimensional space for the purpose of interpretation. In particular, this involves projection to a dimension with the voting behavior of a specific political group—in this case, the members of the Chilean Constitutional Convention. According to whether they vote in favor, against, or abstain, the convention members are given a score between -1 and 1 which positions them in a graphic space. Those who vote similarly are then placed closer in the distribution; when they differ, they are placed farther from each other. Based on this, it is assumed that those members situated below 0 hold a “left-wing” political position, while those who are positioned above 0 have a “right-wing” political position. Then, we used k -means to identify three clusters containing right-wing, center, and left-wing positions. These positions have to be considered to be relative to the political distribution within the convention—i.e., they depend on the voting behavior of the members. Members classified as “center” tended to be from center-left political parties (social democracy, socialist party), while the “left-wing” members of the convention were mainly members of the Communist Party, independents, and indigenous people.

For sentiment classification for each cluster, we were interested in using a classifier with interpretability capabilities and good classification performance; therefore, we explored different Bayesian network classifiers that have been successfully applied in other sentiment classification applications [4].

The rest of the paper is organized as follows. Section 2 presents a brief overview of Bayesian networks, Bayesian network classifiers, and recent sentiment analysis approaches. The Twitter data acquisition, description, pre-processing, sentiment classification, and proposed learning approach for the Bayesian network classifier is presented in Section 3. The results and analysis are discussed in Section 4. Finally, the conclusions of this work as well as future research are given in Section 5.

2. Background

2.1. Bayesian Networks

Bayesian networks were first introduced by Judea Pearl [5] for joint probability computation. These are probabilistic graphical models that represent discrete random variables and their conditional dependencies via a directed acyclic graph (DAG). Originally, Bayesian networks were constructed by hand using human domain knowledge, but then it became of interest to develop algorithms to learn the networks from data [6,7]. Learning Bayesian networks from data has two components: (1) the structure of the networks and (2) the parameters (conditional probability tables). Unfortunately, the work in [8] proved that learning Bayesian networks from data is NP-complete. Therefore, it has been of interest

to develop approximate learning approaches to simplify the learning process [9,10]. For (1), the search can be carried out through an evolutionary computation approach. Some examples include: the combination of genetic algorithms with particle swarm optimization (PSO) called Breeding Swarm to learn Bayesian networks [11], an artificial bee colony (ABC) algorithm for learning Bayesian networks [12], and learning Bayesian networks using a genetic algorithm [13,14]. A review of evolutionary computation approaches for learning Bayesian networks can be found in [15].

In classification (supervised learning) problems, when adopting a probabilistic approach the challenge is to effectively compute the posterior probability of the class variable Y_k (with $k = 1, \dots, K$) given an n -dimensional input data point $\mathbf{x} = (x_1, \dots, x_n)$. This can be achieved using the Bayes rule:

$$p(Y_k|\mathbf{x}) = \frac{p(Y_k)p(\mathbf{x}|Y_k)}{p(\mathbf{x})}. \quad (1)$$

In particular, it is of interest the numerator, which contains the product of the *a priori* probability of the class variable and the *likelihood*, which corresponds to the joint probability of the input features conditioned to the class variable. The computation of the *a priori* probability of the class variable is straightforward. It can be obtained through the relative frequency of the class variable values in the training set. Still, there are many approaches that can be used for computing the likelihood. One of these is to use Bayesian networks, giving way to Bayesian network classifiers [16]. Recent applications of Bayesian network classifiers include sentiment analysis using Twitter data [4], facial biotype prediction for orthodontic treatment planning [17], obstructive sleep apnea diagnosis [18], movie recommender systems [19], air quality forecasting [20], the detection of keratoconus (thinning of the cornea and irregularities of the cornea's surface) [21], fault diagnosis [22], breast cancer diagnosis [23], and the detection of hydrothermal alteration zones associated with epithermal gold mineralization [24].

Amongst the many versions of Bayesian network classifiers [25–30], the two most popular are the naive Bayesian network classifier [31] and the Tree Augmented Naive Bayes (TAN) classifier [16]. The naive Bayes version assumes conditional independence amongst the attributes given the class variable to compute the likelihood in (1). Therefore, there are no edges between the attributes, as shown in the example in Figure 1. Improvements to the naive Bayes include feature weighting techniques [32,33], boosted parameter learning [34], and the exact learning augmented naive Bayes classifier [35].

Figure 1. Example of the naive Bayesian network classifier.

Whereas TAN starts off by considering a fully connected network where the edges have weights, these are computed using the *conditional mutual information* between pairs of attributes. Then, the maximum weighted spanning tree (MWST) method, similar to Kruskal's algorithm [36], is applied to obtain a tree structure, leaving only $n - 1$ edges. Under this version of a Bayesian network classifier, each attribute will have one incoming edge from another attribute, except for the chosen root attribute node. Of course, just like the naive Bayesian network classifier, in the TAN model, in addition to the tree structure, each node will have an incoming edge from the class variable, as shown in the example of Figure 2.

Figure 2. Example of the TAN classifier.

The TAN model corrects the strong assumption of conditional independence established by the naive version. In theory, it should produce better results (accuracy) than the naive Bayes classifier. However, TAN suffers from some problems—one, in particular, is the ability to provide a correct estimate of the conditional mutual information. The conditional mutual information is a measure from information theory that corresponds to a generalization of the correlation, which is also capable of capturing linear and non-linear relationships. In particular, for two variables x_i and x_j , the conditional mutual information for these two attributes given the class is given by [37]:

$$I(x_i; x_j | Y_k) = \sum_{x_i, x_j, Y_k} \log \frac{P(x_i, x_j, Y_k)}{P(x_i | Y_k) P(x_j | Y_k)} \quad (2)$$

$I(x_i; x_j | Y_k)$ measures the information that x_j provides about x_i when the value of Y_k is known, and $I(x_i; x_j | Y_k)$ is non-negative. A main difficulty when working with mutual information or conditional mutual information is when there is not a sufficient number of training examples per class to estimate the joint probability distribution and the conditional distributions in (2) correctly, meaning that the resulting conditional mutual information estimate is not so reliable. This is critical, since TAN obtains its tree structure using $I(x_i; x_j | Y_k)$ as weights in the fully connected graph. From this, a natural question arises: can the weights of the network be learned directly from the data to obtain good classification results without having to estimate the conditional mutual information in (2)?

2.2. Tree Augmented Naive Bayes Classifiers

The TAN classifier is based on the seminal work of Chow and Liu [38], who presented a method to approximate the joint probability distribution of a set of variables, imposing a tree structure as a model of dependencies amongst the variables. The development starts off by using the Kullback–Leibler divergence (D_{KL}) as a measure to quantify how different one distribution is from another. Let $P(\mathbf{x})$ be the real probability distribution of $\mathbf{x}$ and let $P_a(\mathbf{x})$ be an approximation of the real distribution. Then, the matter of how close $P_a(\mathbf{x})$ is to $P(\mathbf{x})$ can be quantified through D_{KL} :

$$D_{KL}(P || P_a) = \sum_{\mathbf{x}} P(\mathbf{x}) \log \left(\frac{P(\mathbf{x})}{P_a(\mathbf{x})} \right). \quad (3)$$

The idea is to minimize (3). For this, Chow and Liu consider that P_a has a tree structure as a model of dependencies amongst the variables. With some algebraic manipulations and the use of information theory definitions, (3) can be rewritten as:

$$D_{KL}(P || P_a) = - \sum_{i=1}^n I(x_i, x_{j(i)}) + \sum_{i=1}^n H(x_i) - H(\mathbf{x}) \quad (4)$$

where $I(x_i, x_{j(i)})$ is the mutual information between variable x_i and its parent $x_{j(i)}$, $H(x_i)$ is the entropy of variable x_i , and $H(\mathbf{x})$ is the joint entropy of the variable set $\mathbf{x} = (x_1, \dots, x_n)$. By inspecting (4), one can see that $\sum_{i=1}^n H(x_i)$ and $H(\mathbf{x})$ are independent of the tree structure amongst the variables. Only $I(x_i, x_{j(i)})$ is dependent of the tree structure, and since it is a non-negative measure and there is a minus sign in front of the sum of the mutual

information between the edges of the tree structure, minimizing $D_{KL}(P||P_a)$ is equivalent to maximizing $\sum_{i=1}^n I(x_i, x_{j(i)})$. Therefore, the Chow–Liu tree procedure is as follows:

1. Build a fully connected weighted graph where the nodes are the variables/attributes of your problem and the weights of the edges are the mutual information between pairs of nodes.
2. Apply a maximum spanning tree algorithm to obtain a tree structure amongst the variables, such that the sum of weights is the maximum. Here, Kruskal’s algorithm can be used for this purpose.
3. Transform the undirected tree to a directed one by choosing a root variable and then setting the direction of all edges to face outward from it.

In the context of classification, we can partition the training set into subgroups containing only examples from the same class. Then, we can use the Chow–Liu tree algorithm to infer the joint probability of the input features (variables) for each class, and use this as the likelihood in (1) to see for which class the input data point obtains the highest posterior probability. An example of the Chow–Liu tree classifier for a two-class problem is shown in Figure 3.

Figure 3. Example of the Chow–Liu tree classifier.

The TAN classifier is an upgrade of the Chow–Liu tree algorithm, where instead of having to train a tree for each class, TAN uses conditional mutual information (2). Therefore, it is only necessary to train one tree for all the classes. The TAN procedure is as follows:

1. Build a fully connected weighted graph where the nodes are the variables/attributes of your problem and the weights of the edges are the conditional mutual information between pairs of nodes.
2. Apply a maximum spanning tree algorithm, to obtain a tree structure amongst the variables, such that the sum of weights is maximum. Here, Kruskal’s algorithm can be used for this purpose.
3. Transform the undirected tree to a directed one by choosing a root variable and then setting the direction of all edges to be outward from it.
4. Construct a TAN model by adding a vertex labelled y and adding an edge from y to each x_i .

As we can see, the resulting tree structure, and, therefore, the likelihood used in (1) to compute the posterior probability, depends of the weight values used in step (1). If we change the weights, this may lead to a different structure and thus a different posterior probability and eventually a different classification of an example. Therefore, in this work we propose an alternative learning method to infer the weights to use in step (1) of the TAN procedure, such that we obtain a high classification performance. For this, we adopt an optimization procedure based on the evolution strategy described in Section 3.6. Other improvements or variants of TAN include averaging TAN models (called ATAN) resulting

from considering each node as the root node (step 3 of the TAN procedure) [39], generalizing TAN from 1-dependence Bayesian network classifier to arbitrary k -dependence [40], and an attribute selective approach for TAN that builds a sequence of approximate models using only the top certain attributes [41].

2.3. Recent Sentiment Analysis Approaches

Although the focus of this paper is the use of Bayesian network classifiers in the context of sentiment analysis, another recent approach includes a stance classification system for Twitter data related to vaccination in Italy which uses term frequency-inverse document frequency (TF-IDF) for the representation of numerical features (the same as we do in this paper) and a support vector machine (SVM) with a linear kernel as the classification model [42]. When different data sources are combined (multimodal data) to carry out sentiment analysis, the important problem of missing values may appear, which is also known as the modality missing issue. In [43], the modality missing issue is addressed by proposing a modality distillation method based on a tensor fusion network. For the problem of intent classification for dialogue utterances, ref. [44] investigated several machine learning methods, concluding that SVM obtained the best results based on the macro-F1 metric. Ref. [45] presents a sentiment analysis study focusing on topics and emotions in video car reviews from YouTube; in this analysis, an SVM was used as the classification model. In [46], a review of major emotion models is conducted and a new version of the Hourglass model, which is a biologically inspired and psychologically motivated emotion categorization model for sentiment analysis, is proposed.

Rumors on Twitter are investigated in [47], where a two-stage methodology that combines a hierarchical long short-term memory (LSTM) network with the various classifiers—SVM, naive Bayes, decision trees, and multilayer perceptron—is presented. In [48], a stacked ensemble of shallow convolutional neural networks (CNNs) is proposed for tracking the moods, emotions, and sentiments of patients expressing an intake of medicine in Twitter. Sentiment and sarcasm classification are tackled in [49] via a multitask learning-based framework using a deep neural network, outperforming a state-of-the-art method based on a CNN. In [50], different deep-learning-based architectures are explored for multimodal sentiment classification, with the bc-LSTM model obtaining the best results. In [51], a new version of SenticNet called SenticNet 6 is proposed; this is built using an approach to knowledge representation that is both top-down and bottom-up via an ensemble of symbolic and subsymbolic AI tools that are applied to the problem of polarity detection from text.

A mathematical explanation of neural tensor network (NTN) is given in [52], as well as the inner link between NTN and other models—e.g., feedforward neural network and attention mechanism—from the perspective of Taylor's theorem. The empirical studies performed on sentiment analysis, named entity recognition and knowledge base completion tasks, provided some valuable insights into NTN. In particular, it was found that the 1st-order NTN achieves a competitive performance in most cases.

In cases where the text that has been analyzed contains more than one opinion target and therefore sentiment analysis is not treated as a binary classification problem, ref. [53] introduces a convolutional stacked bidirectional long short-term memory with a multiplicative attention mechanism for aspect category and sentiment polarity detection that treats the sentiment analysis as a multiclass classification problem. The proposed model was evaluated using the SemEval-2015 and SemEval-2016 datasets, outperforming state-of-the-art methods in aspect-based sentiment analysis.

3. Data and Methods

3.1. Twitter Data

Tweets from the 135 convention members with a Twitter account (out of the 155) were collected using the Twitter API between 4 July and 27 October 2021. This amounts to 37,814 tweets, including 497,715 total words, of which 33,832 are unique words.

To analyze these tweets, we grouped the Twitter accounts differentiating political positions (left, center, right) with the nominate algorithm [3]. We applied the algorithm to 882 ballots cast in the first stage of the convention (from 4 July to 29 September 2021). Then, we used k -means to identify three clusters containing left-wing (cluster 1), center (cluster 2), and right-wing positions (cluster 3). These positions have to be considered to be relative to the political distribution within the convention—i.e., they depend on the voting behavior of the members.

The nominate scores and the clusters found by k -means with $k = 3$ are shown in Figure 4.

Figure 4. Nominat score.

Table 1 presents the Twitter accounts and the number of tweets grouped by the their political positions (clusters).

Table 1. Clusters and Twitter accounts.

Clusters	Twitter Accounts	Number of Tweets
Cluster 1	@MillaburAdolfo @AmpueroAdriana @afc073 @AlondraCVidal @AlvinSM15 @BSepulvedaHales @Bastianlabbed20 @CamilaZarateZ @CaroDistrito1 @CaroVilchesF @CesarUribeA @CotaSanJuan @criordor @dbravosilva @DayyanaGonzalez @ElisaGiustinia1 @elisaloncon @ElsaLabrana @ErickaPortillaB @fersalinas333 @FranciscaArauna @MachiFrancisca1 @frandistrito14 @TiaPikachu @gloconstituyent @oasishernan @Hugo_Gutierrez_ @isabelgodoy @MamaniIsabella @ivannaolivares5 @JanisMeneses_D6 @lidiayagan @lisettevergarrar @loreto_vallejos @LuisJimenezC @manuconstituye @MarcoArellano29 @marcosbarrazag @CKawesqar @mariariveramit @MEQChile @medicanatyh @NatividadLlanq3 @NicolasFernand @RenatoGarinG @robertocedon @rkatrileo @FloresMadriaga @TiareHey @valemirandacc @vanessahoppe21 @BacianWilfredo	17,192

Table 1. Cont.

Clusters	Twitter Accounts	Number of Tweets
Cluster 2	@SquellaAgustin @amaya_alvez @AuroraDelgadoV @labeasanchez @BenitoBaranda @bessygallardoP @carolinasepud19 @CESARVALENZ @christianpviera @cgomezcas @ConySchon @damabarca @danielstingo @felipeharboe @fernando_atria @fchahin @gdominguez_ @giovannaroa @GuillermoNamor @IgnacioAchurra @Jaime_Bassa @JAVIERFUCHSLOC1 @jeniffermella @Jorgeabarcaxv @baradit @JuanjoMartinb @jjlalvarez @LorenaCesp_D23 @barcelobiobio21 @maluchayallego @Ma_joseOyarzun @MarielaSerey @MarioVa25830274 @Mati_Orellana_ @mdaza_abogado @MaxHurtadoR @BottoConstituy1 @patriciapolitz @PatoFdez @tia_paulina_vr @PedroMunozLeiva @Rmontero_ @rodrigo_logan @T_Pustilnick @tatiurru @tomaslaibe @YarelaAysen	14,502
Cluster 3	@amorenos @AlvaroJofre @angelica_tepper @arturozunigaj @bde lamaza @berfontaine @CarolCBown @clau_castrog @conihube @cmonckeberg @cetton15 @felipemena_ @geoconda_aysen @HarryJurgensen @HernanLarrain @arancibialmte @Ktymontealegre @LucianoErnest15 @LMayolB @ossandon_d12 @mcubillo sigall @margaritaleteli @CeciliaUbilla @martinarrau @pablotozoa @PatyLabraB @PaulinaVelosoM1 @PollyanaConsti1 @raulcelism @raneumannb @rvega_c @roci cantuarias @RodrigoAlvarez_ @ruggero_cozzi @ruth_uas @tere_marinovic	6120
Total	135	37,814

Finally, the time series representing the total number of tweets per hour and cluster are shown in Figure 5.

Figure 5. Time series representing the total number of tweets per hour.

3.2. Pre-Processing

Raw data are highly susceptible to inconsistency and redundancy. The pre-processing of tweets is necessary and includes the following points:

- Removing all URLs (e.g., www.xyz.com), hashtags (e.g., #topic), and targets (@user-name);
- Removing all punctuation, symbols, and numbers;
- Correcting the spellings and handling the sequence of repeated characters;
- Removing stop words;
- Removing non-Spanish tweets.

3.3. Term Frequency-Inverse Document Frequency (TF-IDF)

The TF-IDF method determines the relative frequency of words in a specific document through an inverse proportion of the word over the entire document corpus. For determining the value, the method uses two elements: TF, the term frequency of the term i in the document j , and IDF, the inverse document frequency of the term i . TF-IDF can be calculated as [54]:

$$a_{ij} = tf_{ij}idf_i = tf_{ij} \times \log_2 \frac{N}{df_i}, \quad (5)$$

where a_{ij} is the weight of term i in a document j , N is the number of documents in the collection, tf_{ij} is the term frequency of the term i in a document j , and df_i is the document frequency of the term i in the collection.

The resulting TF-IDF matrices for each cluster were, $17,192 \times 133$ for cluster 1, $14,502 \times 130$ for cluster 2, and 6120×172 for cluster 3. These will be the datasets used to train and test the classifiers.

3.4. Sentiment Analysis

In the previous subsection, we described how the input data matrices were constructed. Now, we need the class labels. These were obtained as follows.

Each tweet was labeled with a sentiment with two possible values: negative or positive. We used a list of English positive and negative opinion words or sentiment words (around 7000 words) [55], which we translated into Spanish. In this article, we first determined the sentiment polarity of each tweet by adapting the following measure [4]:

$$\text{Sentiment score} = \frac{\text{positive} - \text{negative}}{\text{positive} + \text{negative} + 2} \quad (6)$$

The sentiment score has the range $-1 < \text{Sentiment score} < +1$, where ‘positive’ represents the positive word count and ‘negative’ represents the negative word count in a tweet. Sometimes, we cannot decide the degree of emotionality of a tweet because the polarity measure fails to do so ($\text{Sentiment score} = 0$) or the positives and negatives cancel each other out. In this case, we cannot decide whether the tweet is positive, negative, or neutral. Therefore, we introduced the following definition:

$$C = \begin{cases} 1 \text{ (positive tweet)} & \text{if } \text{Sentiment score} \geq 0.1 \\ -1 \text{ (negative tweet)} & \text{if } \text{Sentiment score} < 0.1 \end{cases} \quad (7)$$

where C is the class label of the tweet being analyzed.

By applying the above methodology, we obtained the following class labels:

- Cluster 1: 12,298 negative tweets (−1) and 4894 positive tweets (1);
- Cluster 2: 9761 negative tweets (−1) and 4741 positive tweets (1);
- Cluster 3: 4518 negative tweets (−1) and 1602 positive tweets (1).

Usually, in many sentiment analysis classification models, neutral comments are not considered [56,57]. There are two main reasons for this: (1) these models focus on binary

classification—i.e., in the identification of positive versus negative opinions; (2) neutral comments lack information due to their ambiguity.

Neutrality means the absence of sentiment or no sentiment [58]. Neutral comments show an ambiguous weight of sentiment—i.e., they contain an equitable burden of positive and negative polarity [59].

Due to this, this category has been considered as noisy and is broadly excluded from many sentiment models [60–63]. However, some researchers have tackled the problem of neutrality: the authors in [57] propose taking into account neutral reviews in order to improve the classification results, and the authors in [64] propose the use of a multi-sentiment scale, 1–5 stars, to solve the problem of obtaining a wider range of sentiment representation. It is understood that detecting and removing noise can improve a model's performance [59].

In future work, we will consider studying the neutral tweets to improve the binary polarity classification using TAN classifiers.

3.5. Model Performance Evaluation

We evaluated four measures: precision (*Prc*), recall (*Rec*), accuracy (*Acc*), and F_1 -score. These measures are computed as follows:

$$Prc = \frac{TP}{TP + FP} \times 100 \quad (8)$$

$$Rec = \frac{TP}{TP + FN} \times 100 \quad (9)$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (10)$$

$$F_1\text{-score} = 2 \times \frac{Prc \times Rec}{Prc + Rec} \quad (11)$$

where *TP*, *TN*, *FP*, and *FN* denote true positive, true negative, false positive, and false negative, respectively.

3.6. An Evolution Strategy for Learning TAN Classifiers

An evolution strategy (*ES*) is an iterative (generational) procedure. In each generation, new individuals (offspring) are created from existing individuals (parents). The canonical versions of the *ES* are denoted by:

$$(\mu, \lambda) - ES \text{ and } (\mu + \lambda) - ES, \quad (12)$$

Here, μ denotes the number of parents and λ denotes the number of offspring. The parents are deterministically selected (i.e., deterministic survivor selection) from the multi-set of either the offspring, referred to as comma selection ($\mu < \lambda$ must hold), or both the parents and offspring, referred to as plus selection. Selection is based on the ranking of the individuals' fitness, taking the μ best individuals (also referred to as truncation selection). Figure 6 is a flowchart that briefly describes how the *ES* algorithm works. The interested reader is referred to [65] for more details on how *ES* works.

We propose to use the deterministic survivor selection (μ, λ) method to obtain weights for the TAN model, which yields good classification results without estimating the conditional mutual information.

Since the TAN learning procedure starts with a fully connected weighted graph, for a network with n nodes we need to define $m = n(n - 1)/2$ weights. Therefore, for the cluster 1 dataset with 133 attributes (words), we need to find 8778 weights (parameters). For the cluster 2 dataset, which has 130 attributes, we need to find 8385 weights, and for the cluster 3 dataset, with 172 attributes, we need to find 14,706 weights. To find the appropriate values for these weights for each dataset, we proceed independently as follows.

Figure 6. Flowchart of the evolution strategy.

An individual is a candidate solution coded as an m -dimensional vector containing the m weight values of a network. We generate an initial population of μ individuals by placing all the weight values uniformly at random in the unit hypercube for each individual. Next, we iterate over a fixed number of iterations of the algorithm. Each iteration first involves evaluating each candidate solution in the population.

Then, we calculate the fitness function using the fraction of correctly classified examples (in the training set), which is also known as the accuracy.

Next, we select the μ parents with the best scores, which are the highest scores in this case, as we are maximizing the objective function. We consider the following steps:

- The parent population contains $\mu = 10$ individuals.
- $\lambda = 20$ denotes the number of offspring generated in each iteration.
- Individuals die out after one iteration step (we use 1000 iterations), and only the offspring (the youngest individuals) survive to the next generation. In that case, environmental selection chooses μ parents from λ offspring.

These hyperparameters were chosen after trial and error simulations.

3.7. Experimental Setup

We compared the performance of the naive Bayesian network classifier (NB), original TAN, Averaged TAN (ATAN) [39] described in Section 2.2, and the proposed method (μ, λ) -TAN. In addition, we implemented four greedy hill-climbing algorithms for learning Bayesian network classifiers:

- Hill-climbing tree augmented naive Bayes (HC-TAN) [66].
- Hill-climbing super-parent tree augmented naive Bayes (HC-SP-TAN) [66].
- Backward sequential elimination and joining (BSEJ) [25].
- Forward sequential selection and joining (FSSJ) [25].

The HC-TAN and HC-SP-TAN algorithms start from a naive Bayes structure and add edges until there is no improvement in the network score. BSEJ starts from a naive

Bayes structure and adds augmenting edges; it then removes features from the model until no improvement in the network score is witnessed. On the contrary, FSSJ starts from a structure containing just the class node and proceeds by incorporating features and augmenting edges.

For each dataset (cluster), we run the data splitting procedure 10 times, with 70% used for training and 30% used for testing (the splitting was carried out randomly). For each run, we computed the four classification performance measures on the test set, after which the average and the standard deviation of each measure was reported.

All the simulations were carried out using the open-source R software environment for statistical computing.

4. Results and Analysis

The results, when considering the accuracy measure, are shown in Table 2. Overall, we notice that (μ, λ) -TAN outperforms TAN in the three datasets, particularly for the cluster 3 and cluster 1 datasets, where more than a 5% increase in accuracy score was obtained. When compared to NB, significant improvements were obtained for the three datasets—again, particularly for cluster 3 and cluster 1, where a more than 8% increase in accuracy score was achieved. Compared to the hill-climbing approaches, we notice that (μ, λ) -TAN outperforms HC-TAN and HC-SP-TAN for all the datasets. This also holds when compared to BSEJ and FSSJ. ATAN, an improvement of TAN, obtains similar results to HC-TAN, HC-SP-TAN, and BSEJ.

In general, we notice that the lowest performance for all the classifiers is for the cluster 3 dataset. This is mainly due to an overfitting phenomenon, since more weights need to be found than the existing amount of training examples. Nevertheless, (μ, λ) -TAN outperforms TAN in accuracy by 6%.

Table 2. Performance comparison of NB, TAN, ATAN, HC-TAN, HC-SP-TAN, BSEJ, FSSJ, and (μ, λ) -TAN in terms of accuracy in the test set.

Datasets	NB	TAN	ATAN	HC-TAN	HC-SP-TAN	BSEJ	FSSJ	(μ, λ) -TAN
Cluster 1	73.48 ± 0.642	76.46 ± 0.577	76.91 ± 0.399	76.81 ± 0.307	76.69 ± 0.426	76.94 ± 0.498	77.01 ± 0.278	81.64 ± 0.372
Cluster 2	74.32 ± 0.312	77.08 ± 0.579	76.91 ± 0.419	76.16 ± 0.733	76.46 ± 0.457	76.52 ± 0.491	76.88 ± 0.401	80.62 ± 0.737
Cluster 3	71.15 ± 1.271	73.63 ± 0.977	74.63 ± 0.841	74.15 ± 0.743	74.61 ± 0.854	74.18 ± 0.843	75.12 ± 0.783	79.63 ± 0.537

When looking at the precision measure results in Table 3, we notice that (μ, λ) -TAN outperforms all the other classifiers for all the datasets. Moreover, (μ, λ) -TAN was the only model that obtained precision values above 92% for the three datasets. This analysis also holds for the recall measure shown in Table 4 and the F_1 -score shown in Table 5, where (μ, λ) -TAN was the only classifier that obtained recall values above 81% and F_1 -score values above 86%, respectively, for the three datasets.

These results support our hypothesis that the TAN can infer better trees for classification purposes. This is achieved by replacing the computation of conditional mutual information with some other weights learned from the data specifically to enhance the classification performance.

Examples of the resulting networks using (μ, λ) -TAN are shown in Figure 7 for the cluster 1 dataset, Figure 8 for the cluster 2 dataset, and Figure 9 for the cluster 3 dataset. For better visualization, in these figures we omitted the node with the class variable and the edges from this node to all the other nodes.

Table 3. Performance comparison of NB, TAN, ATAN, HC-TAN, HC-SP-TAN, BSEJ, FSSJ, and (μ, λ) -TAN in terms of precision in the test set.

Datasets	NB	TAN	ATAN	HC-TAN	HC-SP-TAN	BSEJ	FSSJ	(μ, λ) -TAN
Cluster 1	84.38 $\pm$ 0.609	88.44 $\pm$ 0.361	89.79 $\pm$ 0.411	89.87 $\pm$ 0.591	89.73 $\pm$ 0.396	89.91 $\pm$ 0.571	90.12 $\pm$ 0.578	94.52 $\pm$ 0.422
Cluster 2	85.31 $\pm$ 0.782	87.77 $\pm$ 0.590	87.34 $\pm$ 0.504	89.53 $\pm$ 0.570	89.52 $\pm$ 0.519	89.69 $\pm$ 0.574	89.72 $\pm$ 0.549	92.58 $\pm$ 0.667
Cluster 3	81.53 $\pm$ 1.568	85.49 $\pm$ 1.302	86.01 $\pm$ 1.034	91.34 $\pm$ 0.684	90.99 $\pm$ 1.110	90.57 $\pm$ 0.815	91.89 $\pm$ 0.456	93.81 $\pm$ 0.795

Table 4. Performance comparison of NB, TAN, ATAN, HC-TAN, HC-SP-TAN, BSEJ, FSSJ, and (μ, λ) -TAN in terms of recall in the test set.

Datasets	NB	TAN	ATAN	HC-TAN	HC-SP-TAN	BSEJ	FSSJ	(μ, λ) -TAN
Cluster 1	79.68 $\pm$ 0.787	80.66 $\pm$ 0.622	80.81 $\pm$ 0.654	80.12 $\pm$ 0.311	80.03 $\pm$ 0.497	80.17 $\pm$ 0.299	80.65 $\pm$ 0.289	82.50 $\pm$ 0.205
Cluster 2	78.51 $\pm$ 0.539	80.13 $\pm$ 0.684	80.03 $\pm$ 0.578	78.11 $\pm$ 0.752	78.56 $\pm$ 0.568	78.50 $\pm$ 0.427	78.59 $\pm$ 0.389	81.03 $\pm$ 0.708
Cluster 3	79.88 $\pm$ 1.483	80.01 $\pm$ 1.021	79.92 $\pm$ 1.011	77.51 $\pm$ 0.867	78.23 $\pm$ 1.019	78.02 $\pm$ 0.961	79.01 $\pm$ 0.912	81.23 $\pm$ 0.608

Table 5. Performance comparison of NB, TAN, ATAN, HC-TAN, HC-SP-TAN, BSEJ, FSSJ, and (μ, λ) -TAN in terms of F_1 -score in the test set.

Datasets	NB	TAN	ATAN	HC-TAN	HC-SP-TAN	BSEJ	FSSJ	(μ, λ) -TAN
Cluster 1	78.78 $\pm$ 0.491	84.37 $\pm$ 0.414	84.48 $\pm$ 0.409	84.71 $\pm$ 0.218	84.61 $\pm$ 0.281	84.88 $\pm$ 0.389	84.98 $\pm$ 0.501	88.10 $\pm$ 0.225
Cluster 2	81.76 $\pm$ 0.308	83.77 $\pm$ 0.438	83.88 $\pm$ 0.439	83.43 $\pm$ 0.538	83.67 $\pm$ 0.335	83.72 $\pm$ 0.373	83.88 $\pm$ 0.343	86.42 $\pm$ 0.591
Cluster 3	80.69 $\pm$ 1.073	82.69 $\pm$ 0.727	82.56 $\pm$ 0.801	83.86 $\pm$ 0.515	84.12 $\pm$ 0.662	83.82 $\pm$ 0.615	85.12 $\pm$ 0.577	87.07 $\pm$ 0.356

Interpreting the Networks

Bayesian networks classifiers offer not only a quantitative mode of analysis but also a significant qualitative view of the data, particularly when dealing with Twitter communication.

For this qualitative approach to data, we analyze the three networks representing different political positions within the Chilean Constitutional Convention. Through the aforementioned procedure (see Section 3), we obtain a behavioral classification of the political attitudes of the members and not self-declarative positioning. In this classification (see Figure 4), we obtain a network for left-wing positions in cluster 1 (Figure 7), one for central positions in cluster 2 (Figure 8), and one for right-wing positions in cluster 3 (Figure 9).

Considering Figure 7, we observe several politically meaningful terms and directed edges for a rather identitarian left-wing view (independent movements, Communist Party, indigenous people). Terms such as “pueblo”, indígena (indigenous), mujeres (women), popular (popular), and octubre (October) reflect central concerns for the rather radical left-wing members of the convention. “October” refers to 18 October 2019, the day the popular uprising began. A very common self-description in left-wing Latin American politics consists of distinguishing between the elite and “pueblo”. In this case, the category of “pueblo” does not really mean “the people”, as in the U.S. American constitution, in which the term works as an inclusive category of citizenship. In Latin American semantic usage, “people” is a category that is construed for grouping historically discriminated-against actors [67]. It includes women, indigenous people, workers, among others—i.e., the “popular world” (as expressed in the network), which is opposed to the category of “elites”, made up of privileged, well-educated, modern citizens. This division of the world between “pueblo” and “elite” is rather common in Latin American left-wing populism (Argentina, Venezuela, Bolivia, Ecuador, Cuba) and constitutes the ideological basis for constructing a hegemonic strategy of political access to state power [68].

Figure 7. Examples of solutions achieved using (μ, λ) -TAN for the cluster 1 dataset.

In this case, it is worth mentioning three directed edges: (a) popular $\rightarrow$ Chile; (b) apoyo (support) $\rightarrow$ territorial, and apoyo (support) $\rightarrow$ social; (c) Mapuche $\rightarrow$ originarios; and (d) normas $\rightarrow$ indígena (indigenous) $\rightarrow$ convencionales (members of the convention). The directed edge (a) condenses the idea of a country (Chile) belonging to the popular classes, which is semantically combined with the directed edge (b), referring to the territorial and social support for the construction of this new order. The project of a popular nation began to grow after the critical events of 18 October 2019. Left-wing political parties aimed at reconstructing the events (protests, riots, violent demonstrations) as a popular demand for justice and political changes that would reverse the current state of affairs in which elites prevail, thereby transforming Chile into a “popular republic” (popular $\rightarrow$ Chile)—i.e., a country that belongs “to the pueblo”. In this scenario, indigenous people such as the Mapuche (from southern Chile) play a crucial role: as original inhabitants of the country, they represent the precolonial essence of the popular nation in left-wing politics [69]. For this reason, indigenous peoples have been granted reserved seats (escaños) in the convention (see the directed edge pueblo $\rightarrow$ escaños). Of the 155 seats, 17 have been exclusively granted to representatives of indigenous peoples, the majority of whom are Mapuche.

Considering Figure 8, in which a view of center-left politicians is represented (mostly from the social democracy and socialist party), we also observe meaningful terms and directed edges related to this political approach. Terms such as equidad (equity), diálogo (dialogue), democracia (democracy), transparencia (transparency), derechos (rights), ddhh (abb. for human rights), abd ciudadanía (citizenship) are classical concerns and principles

of social democratic parties in modern liberal democracies. The discourse concerning rights and human rights is the most general framework for this political view. According to this, the discourse around rights and human rights is not only an institutional discourse but also heuristic meaningful semantics through which people conceive their daily experience in the social world and articulate their demands of both the government and institutions [70]. This expresses the success of the modern narrative of rights and, at the same time, reinforces universal democratic values such as equity, dialogue, democracy, transparency, and citizenship—namely, the core of social democratic politics.

Figure 8. Examples of solutions achieved using (μ, λ) -TAN for the cluster 2 dataset.

According to this, in Figure 8, we observe some interesting directed edges: (a) conversando (conversation) $\rightarrow$ espacio (space) $\rightarrow$ diálogo (dialogue) $\rightarrow$ esperanza (hope), as well as the bifurcation diálogo (dialogue) $\rightarrow$ general $\rightarrow$ pueblos (people); (b) votación (ballots) $\rightarrow$ transparencia (transparency) $\rightarrow$ convención (convention); (c) trabajo (work) $\rightarrow$ constituyente (member of the convention) $\rightarrow$ pueblo (people) $\rightarrow$ derechos (rights); and (d) Chile $\rightarrow$ democracia (democracy) $\rightarrow$ públicas (policy). In this case, the network construction reveals a rather liberal conception of democracy in which, through spaces of conversation and dialogue, the hopes of the people can be addressed [71]. The transparency of democratic procedures is also mentioned in the network, as well as the (hard) work carried out by the Constitutional Convention for the rights of Chilean “pueblo” (people). Unlike what

happens in the case of the network of the radical left-wing members of the convention, the concept of “pueblo” must now be comprehended within a different semantic constellation that refers to the democratic conditions of citizenship, as expressed in the directed edge (a) and in the above-described relevant single terms. This is an inclusive and universal concept of “pueblo” understood as rights-based citizenship, which is common in U.S. American and European politics. Finally, the directed edge $\text{chile} \rightarrow \text{democracia}$ (democracy) $\rightarrow \text{públicas}$ (policy), refers to the policies (políticas públicas) that should implement rights that would lead to achieving citizenship. To that extent, policies are closely connected to democracy in the view of the center-left members of the convention.

Figure 9 represents the view of the right-wing members of the Chilean Constitutional Convention (liberal and social conservative parties). A rather common behavior in right-wing politics consists of constructing a dual vision of the political field. On the one hand, right-wing members affirm their own values, and, on the other, they express criticism of the principles and attitudes of left-wing politics [72]. The network mirrors this duality with terms affirming right-wing values, such as libertad (liberty), libre (free), expresión (expression), diálogo (dialogue), and bandera (flag), and terms referring to critical concerns, such as censura (censorship), izquierda (left), violencia (violence), and víctimas (victims). With these references, the right-wing members of the convention conduct an affirmative political strategy of their own positions and, at the same time, oppose the more left-wing stances.

Figure 9. Examples of solutions achieved using (μ, λ) -TAN for the cluster 3 dataset.

We find a similar situation when we observe the directed edges. We identify four relevant edges: (a) nueva (new) → censura (censorship); (b) nuevo (new) → negacionismo (denial); (c) expresión (expression) → plebiscitos (plebiscite) → free (libre); and (d) trabajar (work) → chilenos (Chilean) → acuerdo (agreement) → reglamento (regulation). Directed edges (a) and (b) oppose the proposals made by the left-wing members of the convention aimed at limiting freedom of expression (control of the media and the proposal of ‘denial by omission’, which implies the mandatory recognition of human rights violations every time the events of October 2019 are discussed); meanwhile, directed edges (c) and (d) express the affirming attitude of right-wing members regarding freedom of expression and the orientation towards agreements regarding the formulation of the convention’s rules of procedure.

As can be seen, the classification offers a meaningful and detailed view of the qualitative political dynamics of the members of the Chilean Constitutional Convention. It allows us to recognize distinctive semantic elements in left-wing, center, and right-wing politics that not only describe the static components of the process, but also the mutual references between political positions, as well as their substantive foundations (values, principles) and political strategies (self-affirmation, opposition, criticism).

5. Conclusions

These results confirmed our view that the tree structures obtained using conditional mutual information do not necessarily yield the best classifiers. Moreover, based on the number of features and training samples, the performance of the TAN classifier can be quite affected, as shown in the case of the cluster 3 dataset.

Another drawback when using conditional mutual information is that the resulting tree might not be reliable for interpretability purposes, a topic that is quite popular nowadays. Indeed, one of the advantages of using Bayesian network classifiers is the possibility of interpreting and understanding the classification process and the results by analyzing the network and how the features interact. For example, when analyzing the resulting network for the cluster 1 dataset, the TAN model achieved a mean accuracy value of 76 versus 81 when using (μ, λ) -TAN. Furthermore, even though the (μ, λ) -TAN model showed the lowest performance in cluster 3 (due to overfitting) with a mean accuracy of 79, it significantly outperformed the TAN model, which showed a mean accuracy of 73. Which network is more reliable for interpreting the results? Common sense warns that the model with the best performance should have a more reliable network, with interactions or dependencies that are more realistic than those in a model with a lower performance. This is something that could be tested in future research with artificially generated data.

Considering our qualitative interpretation of the networks, we can conclude that the resulting tree structures offer analytically interesting material with which to understand the political behavior and semantics of the members of the Chilean Constitutional Convention. These results present politically meaningful words and directed edges that allow for a differentiation between left-wing, center, and right-wing positions both ideologically (through principles and values) and in terms of political practices (self-affirmation, opposition, and criticism). The qualitative interpretation could even contribute to explaining the lower mean accuracy value of the (μ, λ) -TAN in cluster 3 (right-wing members of the Chilean Constitutional Convention), in addition to the overfitting. The composition of this network is more complex than that of others, as right-wing members apply a dual strategy of self-affirmation and criticism of left-wing positions, thus using terms with both a right-wing identity (self-affirmation) and a left-wing identity (for criticizing the left). Therefore, the model needs more training examples to increase its accuracy in this case. Further research should test this conjecture.

Here, we have used a simple evolution strategy in the model to learn the weights. Of course, other evolutionary computation techniques could also be used to test other hybrid algorithms. Future research should consider more benchmark datasets as well as industrial applications.

Author Contributions: Conceptualization, G.A.R., P.A.H. and A.M.; methodology, G.A.R., P.A.H. and A.M.; software, P.A.H.; validation, G.A.R., P.A.H. and A.M.; formal analysis, G.A.R., P.A.H. and A.M.; investigation, G.A.R., P.A.H. and A.M.; resources, G.A.R. and A.M.; data curation, P.A.H.; writing—original draft preparation, G.A.R., P.A.H. and A.M.; writing—review and editing, G.A.R., P.A.H. and A.M.; visualization, G.A.R., P.A.H. and A.M.; supervision, G.A.R., P.A.H. and A.M.; project administration, G.A.R., P.A.H. and A.M.; funding acquisition, G.A.R. and A.M. All authors have read and agreed to the published version of the manuscript.

Funding: The authors would like to thank ANID FONDECYT grant number 1180706 (G.A.R.), ANID FONDECYT grant number 1190265 (A.M., G.A.R.), and ANID PIA/BASAL FB0002 (G.A.R.) for financially supporting this research.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors would like to thank the research assistant Carolina Molinare for helping with the Twitter data recollection process.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Garcés, M. October 2019: Social Uprising in Neoliberal Chile. *J. Lat. Am. Cult. Stud.* **2019**, *28*, 483–491. [\[CrossRef\]](#)
2. Somma, N.M. Power cages and the October 2019 uprising in Chile. *Soc. Identities* **2021**, *27*, 579–592. [\[CrossRef\]](#)
3. Poole, K.T.; Rosenthal, H. A Spatial Model for Legislative Roll Call Analysis. *Am. J. Political Sci.* **1985**, *29*, 357–384. [\[CrossRef\]](#)
4. Ruz, G.A.; Henríquez, P.A.; Mascareño, A. Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. *Future Gener. Comput. Syst.* **2020**, *106*, 92–104. [\[CrossRef\]](#)
5. Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*; Morgan Kaufmann: San Francisco, CA, USA, 1988.
6. Cooper, G.F.; Herskovits, E. A Bayesian method for the induction of probabilistic networks from data. *Mach. Learn.* **1992**, *9*, 309–347. [\[CrossRef\]](#)
7. Heckerman, D.; Geiger, D.; Chickering, D.M. Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. *Mach. Learn.* **1995**, *20*, 197–243. [\[CrossRef\]](#)
8. Chickering, D.M. Learning Bayesian Networks is NP-Complete. In *Learning from Data: Artificial Intelligence and Statistics V*; Springer: New York, NY, USA, 1996; pp. 121–130.
9. Heckerman, D. *A Tutorial on Learning with Bayesian Networks*; Technical Report MSR-TR-95-06; Microsoft Research: Redmond, WA, USA, 1995.
10. Neapolitan, R.E. *Learning Bayesian Networks*; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2004.
11. Khanteymoori, A.R.; Olyae, M.H.; Abbaszadeh, O.; Valian, M. A novel method for Bayesian networks structure learning based on Breeding Swarm algorithm. *Soft Comput.* **2017**, *22*, 3049–3060. [\[CrossRef\]](#)
12. Ji, J.; Wei, H.; Liu, C. An Artificial Bee Colony Algorithm for Learning Bayesian Networks. *Soft Comput.* **2013**, *17*, 983–994. [\[CrossRef\]](#)
13. Larrañaga, P.; Poza, M.; Yurramendi, Y.; Murga, R.H.; Kuijpers, C.M.H. Structure learning of Bayesian networks by genetic algorithms: A performance analysis of control parameters. *IEEE Trans. Pattern Anal. Mach. Intell.* **1996**, *18*, 912–926. [\[CrossRef\]](#)
14. Zhang, C.; Cao, M.; Peng, B.; Zheng, S. Learning bayesian network by genetic algorithm using structure-parameter restrictions. In Proceedings of the 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), San Jose, CA, USA, 15–19 July 2013; pp. 1–5.
15. Larrañaga, P.; Karshenas, H.; Bielza, C.; Santana, R. A review on evolutionary algorithms in Bayesian network learning and inference tasks. *Inf. Sci.* **2013**, *233*, 109–125. [\[CrossRef\]](#)
16. Friedman, N.; Geiger, D.; Goldszmidt, M. Bayesian Network Classifiers. *Mach. Learn.* **1997**, *29*, 131–163. [\[CrossRef\]](#)
17. Ruz, G.A.; Araya-Díaz, P. Predicting Facial Biotypes Using Continuous Bayesian Network Classifiers. *Complexity* **2018**, *2018*, 4075656. [\[CrossRef\]](#)
18. Ferreira-Santos, D.; Rodrigues, P. Enhancing obstructive sleep apnea diagnosis with screening through disease phenotypes: Algorithm development and validation. *JMIR Med. Inform.* **2021**, *9*, e25124. [\[CrossRef\]](#)
19. Khurana, D.; Dhingra, S. An improved hybrid and knowledge based recommender system for accurate prediction of movies. In Proceedings of the 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom 2021), New Delhi, India, 17–19 March 2021; pp. 881–886.
20. Ramos-López, D.; Maldonado, A.D. Cost-Sensitive Variable Selection for Multi-Class Imbalanced Datasets Using Bayesian Networks. *Mathematics* **2021**, *9*, 156. [\[CrossRef\]](#)

21. Castro-Luna, G.; Martínez-Finkelshtein, A.; Ramos-López, D. Robust keratoconus detection with Bayesian network classifier for Placido-based corneal indices. *Contact Lens Anterior Eye* **2020**, *43*, 366–372. [\[CrossRef\]](#)
22. Amine Atoui, M.; Cohen, A. Fault diagnosis using PCA-Bayesian Network classifier with unknown faults. In Proceedings of the European Control Conference 2020 (ECC 2020), St. Petersburg, Russia, 12–15 May 2020; pp. 2039–2044.
23. Krishnakumar, N.; Abdou, T. Detection and Diagnosis of Breast Cancer Using a Bayesian Approach. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Cham, Switzerland, 2020; pp. 335–341.
24. Bolouki, S.; Ramazi, H.; Maghsoudi, A.; Pour, A.; Sohrabi, G. A remote sensing-based application of bayesian networks for epithermal gold potential mapping in Ahar-Arasbaran area, NW Iran. *Remote Sens.* **2020**, *12*, 105. [\[CrossRef\]](#)
25. Pazzani, M.J. Constructive Induction of Cartesian Product Attributes. In *Feature Extraction, Construction and Selection: A Data Mining Perspective*; Springer: Boston, MA, USA, 1998; pp. 341–354.
26. Provan, G.M.; Singh, M. Learning Bayesian Networks Using Feature Selection. In *Learning from Data: Artificial Intelligence and Statistics V*; Springer: New York, NY, USA, 1996; pp. 291–300.
27. Sahami, M. Learning Limited Dependence Bayesian Classifiers. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96), Portland, OR, USA, 2–4 August 1996; pp. 335–338.
28. Margaritis, D.; Thrun, S. Bayesian Network Induction via Local Neighborhoods. In *Advances in Neural Information Processing Systems 12*; Solla, S.A., Leen, T.K., Müller, K., Eds.; MIT Press: Cambridge, MA, USA, 2000; pp. 505–511.
29. Ruz, G.A.; Pham, D.T. Building Bayesian network classifiers through a Bayesian complexity monitoring system. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2009**, *223*, 743–755. [\[CrossRef\]](#)
30. Bielza, C.; Larrañaga, P. Discrete Bayesian Network Classifiers: A Survey. *ACM Comput. Surv.* **2014**, *47*, 5:1–5:43. [\[CrossRef\]](#)
31. Duda, R.O.; Hart, P.E. *Pattern Classification and Scene Analysis*; John Wiley & Sons: New York, NY, USA, 1973.
32. Jiang, L.; Zhang, L.; Li, C.; Wu, J. A Correlation-Based Feature Weighting Filter for Naive Bayes. *IEEE Trans. Knowl. Data Eng.* **2019**, *31*, 201–213. [\[CrossRef\]](#)
33. Jiang, L.; Zhang, L.; Yu, L.; Wang, D. Class-specific attribute weighted naive Bayes. *Pattern Recognit.* **2019**, *88*, 321–330. [\[CrossRef\]](#)
34. Jing, Y.; Pavlović, V.; Rehag, J.M. Efficient Discriminative Learning of Bayesian Network Classifier via Boosted Augmented Naive Bayes. In Proceedings of the 22nd International Conference on Machine Learning (ICML '05), Bonn, Germany, 7–11 August 2005; Association for Computing Machinery: New York, NY, USA, 2005; pp. 369–376. [\[CrossRef\]](#)
35. Sugahara, S.; Uto, M.; Ueno, M. Exact Learning Augmented Naive Bayes Classifier. *Entropy* **2018**, *23*, 1703. [\[CrossRef\]](#)
36. Kruskal, J.B. On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. *Proc. Am. Math. Soc.* **1956**, *7*, 48–50. [\[CrossRef\]](#)
37. Pham, D.T.; Ruz, G.A. Unsupervised training of Bayesian networks for data clustering. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2009**, *465*, 2927–2948. [\[CrossRef\]](#)
38. Chow, C.; Liu, C. Approximating discrete probability distributions with dependence trees. *IEEE Trans. Inf. Theory* **1968**, *14*, 462–467. [\[CrossRef\]](#)
39. Jiang, L.; Cai, Z.; Wang, D.; Zhang, H. Improving Tree augmented Naive Bayes for class probability estimation. *Knowl.-Based Syst.* **2012**, *26*, 239–245. [\[CrossRef\]](#)
40. Long, Y.; Wang, L.; Sun, M. Structure Extension of Tree-Augmented Naive Bayes. *Entropy* **2019**, *21*, 721. [\[CrossRef\]](#)
41. Chen, S.; Zhang, Z.; Liu, L. Attribute Selecting in Tree-Augmented Naive Bayes by Cross Validation Risk Minimization. *Mathematics* **2021**, *9*, 2564. [\[CrossRef\]](#)
42. Bechini, A.; Ducange, P.; Marcelloni, F.; Renda, A. Stance Analysis of Twitter Users: The Case of the Vaccination Topic in Italy. *IEEE Intell. Syst.* **2021**, *36*, 131–139. [\[CrossRef\]](#)
43. Peng, W.; Hong, X.; Zhao, G. Adaptive Modality Distillation for Separable Multimodal Sentiment Analysis. *IEEE Intell. Syst.* **2021**, *36*, 82–89. [\[CrossRef\]](#)
44. Schuurmans, J.; Frasincar, F. Intent Classification for Dialogue Utterances. *IEEE Intell. Syst.* **2020**, *35*, 82–88. [\[CrossRef\]](#)
45. Stappen, L.; Baird, A.; Cambria, E.; Schuller, B.W. Sentiment Analysis and Topic Recognition in Video Transcriptions. *IEEE Intell. Syst.* **2021**, *36*, 88–95. [\[CrossRef\]](#)
46. Susanto, Y.; Livingstone, A.G.; Ng, B.C.; Cambria, E. The Hourglass Model Revisited. *IEEE Intell. Syst.* **2020**, *35*, 96–102. [\[CrossRef\]](#)
47. Akhtar, M.S.; Ekbal, A.; Narayan, S.; Singh, V. No, That Never Happened!! Investigating Rumors on Twitter. *IEEE Intell. Syst.* **2018**, *33*, 8–15. [\[CrossRef\]](#)
48. Mahata, D.; Friedrichs, J.; Shah, R.R.; Jiang, J. Detecting Personal Intake of Medicine from Twitter. *IEEE Intell. Syst.* **2018**, *33*, 87–95. [\[CrossRef\]](#)
49. Majumder, N.; Poria, S.; Peng, H.; Chhaya, N.; Cambria, E.; Gelbukh, A. Sentiment and Sarcasm Classification With Multitask Learning. *IEEE Intell. Syst.* **2019**, *34*, 38–43. [\[CrossRef\]](#)
50. Poria, S.; Majumder, N.; Hazarika, D.; Cambria, E.; Gelbukh, A.; Hussain, A. Multimodal Sentiment Analysis: Addressing Key Issues and Setting Up the Baselines. *IEEE Intell. Syst.* **2018**, *33*, 17–25. [\[CrossRef\]](#)

51. Cambria, E.; Li, Y.; Xing, F.Z.; Poria, S.; Kwok, K. SenticNet 6: Ensemble Application of Symbolic and Subsymbolic AI for Sentiment Analysis. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management (CIKM'20), Virtual Event, Ireland, 19–23 October 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 105–114. [\[CrossRef\]](#)
52. Li, W.; Zhu, L.; Cambria, E. Taylor's theorem: A new perspective for neural tensor networks. *Knowl.-Based Syst.* **2021**, *228*, 107258. [\[CrossRef\]](#)
53. Kumar, A.J.; Trueman, T.E.; Cambria, E. A Convolutional Stacked Bidirectional LSTM with a Multiplicative Attention Mechanism for Aspect Category and Sentiment Detection. *Cogn. Comput.* **2021**, *13*, 1423–1432. [\[CrossRef\]](#)
54. Robertson, S. Understanding inverse document frequency: On theoretical arguments for IDF. *J. Doc.* **2004**, *60*, 503–520. [\[CrossRef\]](#)
55. Liu, B. Opinion Mining, Sentiment Analysis, and Opinion Spam Detection. Available online: <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html> (accessed on 1 November 2021).
56. Koppel, M.; Schler, J. Using neutral examples for learning polarity. In *International Joint Conference on Artificial Intelligence*; Lawrence Erlbaum Associates Ltd.: Edinburgh, Scotland, 2005; Volume 19, p. 1616.
57. Koppel, M.; Schler, J. The importance of neutral examples for learning sentiment. *Comput. Intell.* **2006**, *22*, 100–109. [\[CrossRef\]](#)
58. Liu, B. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*; Cambridge University Press: Cambridge, UK, 2020.
59. Valdivia, A.; Luzón, M.V.; Cambria, E.; Herrera, F. Consensus vote models for detecting and filtering neutrality in sentiment analysis. *Inf. Fusion* **2018**, *44*, 126–135. [\[CrossRef\]](#)
60. Pang, B.; Lee, L.; Vaithyanathan, S. Thumbs up? Sentiment classification using machine learning techniques. *arXiv* **2002**, arXiv:cs/0205070.
61. Wawre, S.V.; Deshmukh, S.N. Sentiment classification using machine learning techniques. *Int. J. Sci. Res. (IJSR)* **2016**, *5*, 819–821.
62. Da Silva, N.F.; Hruschka, E.R.; Hruschka, E.R., Jr. Tweet sentiment analysis with classifier ensembles. *Decis. Support Syst.* **2014**, *66*, 170–179. [\[CrossRef\]](#)
63. Díaz, F.; Henríquez, P.A. Social sentiment segregation: Evidence from Twitter and Google Trends in Chile during the COVID-19 dynamic quarantine strategy. *PLoS ONE* **2021**, *16*, e0254638. [\[CrossRef\]](#)
64. Pang, B.; Lee, L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *arXiv* **2005**, arXiv:cs/0506075.
65. Back, T. *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*; Oxford University Press: Oxford, UK, 1996.
66. Keogh, E.J.; Pazzani, M.J. Learning the structure of augmented Bayesian classifiers. *Int. J. Artif. Intell. Tools* **2002**, *11*, 587–601. [\[CrossRef\]](#)
67. Levitsky, S.; Roberts, K. *The Resurgence of the Latin American Left*; The Johns Hopkins University Press: Baltimore, Maryland, 2011.
68. Laclau, E.; Mouffe, C. *Hegemony and Socialist Strategy*; Verso Books: Brooklyn, NY, USA, 2014.
69. Gaitán-Barrera, A.; Azeez, G.K. Beyond recognition: Autonomy, the state, and the Mapuche Coordinadora Arauco Malleco. *Lat. Am. Caribb. Ethn. Stud.* **2018**, *13*, 113–134. [\[CrossRef\]](#)
70. Thornhill, C. The Sociology of Constitutions. *Annu. Rev. Law Soc. Sci.* **2017**, *13*, 493–513. [\[CrossRef\]](#)
71. Mackert, J.; Wolf, H.; Turner, B. *The Condition of Democracy: Volume 1: Neoliberal Politics and Sociological Perspectives*; Routledge: London, UK, 2021.
72. Wodak, R. *The Politics of Fear: What Right-Wing Populist Discourses Mean*; Sage: London, UK, 2015.