

Evaluation of classification techniques for identifying fake reviews about products and services on the internet

Avaliação de técnicas de classificação para identificação de comentários falsos sobre produtos e serviços na internet

Andrey Schmidt dos Santos¹, Luis Felipe Riehs Camargo¹, Daniel Pacheco Lacerda¹ 

¹Universidade do Vale do Rio dos Sinos – UNISINOS, Programa de Pós-graduação em Engenharia de Produção e Sistemas, São Leopoldo, RS, Brasil. E-mail: santos.andreys@gmail.com; feliperiehs@yahoo.com.br; dlacerda@unisinis.br

How to cite: Santos, A. S., Camargo, L. F. R., & Lacerda, D. P. (2020). Evaluation of classification techniques for identifying fake reviews about products and services on the internet. *Gestão & Produção*, 27(4), e4672. <https://doi.org/10.1590/0104-530X4672-20>

Abstract: With the e-commerce growth, more people are buying products over the internet. To increase customer satisfaction, merchants provide spaces for product and service reviews. Products with positive reviews attract customers, while products with negative reviews lose customers. Following this idea, some individuals and corporations write fake reviews to promote their products and services or defame their competitors. The difficulty for finding these reviews was in the large amount of information available. One solution is to use data mining techniques and tools, such as the classification function. Exploring this situation, the present work evaluates classification techniques to identify fake reviews about products and services on the Internet. The research also presents a literature systematic review on fake reviews. The research used 8 classification algorithms. The algorithms were trained and tested with a hotels database. The CONCENSO algorithm presented the best result, with 88% in the precision indicator. After the first test, the algorithms classified reviews on another hotels database. To compare the results of this new classification, the Review Skeptic algorithm was used. The SVM and GLMNET algorithms presented the highest convergence with the Review Skeptic algorithm, classifying 83% of reviews with the same result. The research contributes by demonstrating the algorithms ability to understand consumers' real reviews to products and services on the Internet. Another contribution is to be the pioneer in the investigation of fake reviews in Brazil and in production engineering.

Keywords: Fake reviews; Text classification; Knowledge discovery in databases; Text mining.

Resumo: Com a ascensão do comércio eletrônico, mais pessoas estão comprando produtos na internet. Para aumentar a satisfação, os comerciantes estão disponibilizando espaços para clientes comentarem sobre seus produtos e serviços. Produtos com comentários positivos atraem clientes, enquanto produtos com comentários negativos afastam clientes. Seguindo esta ideia, algumas pessoas e organizações estão escrevendo comentários falsos, a fim de promover ou de denegrir a imagem de um produto ou serviço. A dificuldade em encontrar estes comentários está na grande quantidade de informação disponível. Uma solução é o uso de técnicas e ferramentas da mineração de dados, em especial a classificação. Explorando esta

Received Feb. 25, 2018 - Accepted Nov. 9, 2018

Financial support: None.



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

situação, o objetivo deste trabalho é avaliar técnicas de classificação para identificar comentários falsos sobre produtos e serviços na internet. O trabalho também apresenta uma revisão de literatura sobre comentários falsos. A pesquisa utilizou oito algoritmos. Os algoritmos foram treinados e testados com uma base de dados de hotéis. O algoritmo CONCENSO apresentou o melhor resultado, com uma precisão de 88%. Após o primeiro teste, os algoritmos treinados classificaram comentários em outra base de dados de hotéis. Para comparar os resultados desta nova classificação, o algoritmo do Review Skeptic foi utilizado. Os algoritmos SVM e GLMNET apresentaram as melhores convergências com o algoritmo do Review Skeptic, classificando 83% dos comentários com o mesmo resultado. A pesquisa contribuiu ao demonstrar a habilidade dos algoritmos em entender a veracidade dos comentários sobre produtos e serviços na internet. Outra contribuição é ser o pioneiro na investigação de comentários falsos no Brasil e na engenharia de produção.

Palavras-chave: Comentários falsos; Classificação de texto; Descoberta de conhecimento em base de dados; Mineração de texto.

1 Introduction

With the e-commerce growth, more people are buying products over the internet. To increase customer satisfaction, merchants provide spaces for consumers to comment about the products and services they buy. With the largest number of customers using the internet, the amount of reviews that products receive grows (Hu & Liu, 2004). American Express affirms that 58% of users who read reviews on the Internet trust more on the product or service if there are positive reviews (Column Five, 2014). Products with more positive reviews attract customers, while products with negative reviews lose customers (Xie et al., 2012). Faced with this fact, some individuals and corporations write fake reviews (spam review, opinion spam) to promote their products and services or defame their competitors. Fake reviews make social media sources of lie and control information to mobilize people against a target (Liu, 2012).

The difficulty in finding fake reviews is in the large amount of information available. The human capacity to analyze and understand data is not enough to find the fake reviews. Another problem is the impact of fake reviews on the Internet users, the company network and the academy as is illustrated in the systemic structure (Andrade et al., 2006) of Figure 1 and its explanation in Table 1.

Table 1. Description of the systemic structure of fake reviews.

Cycle	Description
B1	The more fake reviews, the more database, the more academy after a while, more detection techniques, and after a while less fake reviews.
B2	The less true positive reviews, the more fake positive reviews, the more customers, the more sales, the more true reviews and the more true positive reviews.
B3	The less true negative reviews, the more fake negative reviews, the less customers, the less sales, the less true reviews, and the less true negative reviews.

B1: Balancing cycle number one; B2: Balancing cycle number two; B3: Balancing cycle number three.

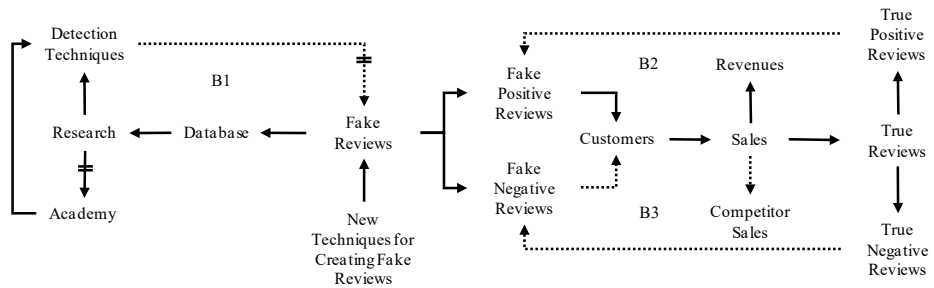


Figure 1. Systemic structure of fake reviews.

B1: Balancing cycle number one; B2: Balancing cycle number two; B3: Balancing cycle number three.

The identification of fake reviews on the Internet began in the literature from the Jindal & Liu (2007). Cheats in credit cards, money laundering, among others, have been researched for a longer time (Fuller et al., 2011). A problem similar to fake reviews is email spam, which has the purpose of sending unwanted information to a target (Cormack, 2008). A solution to this problem is to use computational techniques and tools of data mining (Fayyad et al., 1996), as the classification function (Wu et al., 2008) that separates data classes (Han & Kamber, 2006). An example of classification is the determination of the soccer team that undergraduate students belong based on certain characteristics, like city in which it lives and family soccer team.

The present work evaluates classification techniques to identify fake reviews about products and services on the Internet. This work offers techniques and algorithms for organizations to find fake reviews on the internet and understand the true opinions of their customers. The research also presents a systematic review of literature on fake reviews. Another contribution is to be the pioneer in the research of fake reviews in Brazil and in production engineering.

The work is segmented into six chapters. The second chapter presents the concept of knowledge discovery in databases. In sequence is detailed the identification of fake reviews. The methodological procedures present the work method. The fifth chapter presents the results found, while the last chapter presents the discussion of the results and the research conclusion.

2 Knowledge discovery in databases

Knowledge Discovery in Databases (KDD) is the process of discovering information from a large amount of data (Fayyad et al., 1996). KDD is applied for a variety of tasks, such as analyzing medical outcomes, detecting credit card fraud, and predicting customer buying behavior (Mitchell, 1999). Fayyad et al. (1996), Han & Kamber (2006) and Tan et al. (2009) present KDD processes with three common steps:

- Pre-processing: the removal of inconsistent noise and data, the combination of multiple data sources, the choose of relevant data, and the transformation of data into appropriate forms;
- Data mining: the extraction of data patterns with the support of statistical methods and machine learning;
- Post-processing: the interpretation of the patterns discovered in the data and the use of the information extracted in the form of knowledge.

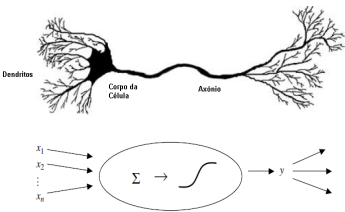
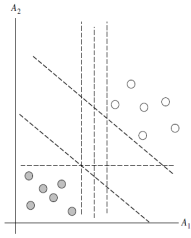
The two main functions of data mining are supervised learning (classification), and unsupervised learning (clustering) (Liu, 2007). The classification technique represents the relationship between the data and its classes, identified during the data training step (Ghosh et al., 2012). After the training, the algorithms are applied to determine the class of any new data (data testing step) (Tan et al., 2009). Using soccer teams as an example, 80% of the students (training data) of an undergraduate group inform their characteristics (family team, city of origin, etc.) and for which club they belong. With this information the algorithms are trained and then applied, without the information of the soccer team, to the other 20% of students in the class (testing data). In the end, the soccer teams classified by the algorithms are compared to the soccer teams informed by the 20% of the students. To ensure greater statistical validity, K-Fold Cross-Validation (Kohavi, 1995) is used to test algorithms with different partitions of the database (training data can become testing data and vice versa). To compare the effectiveness of the algorithms, the literature recommends the use of precision, recall and f-measure indicators (Tan et al., 2009; Weiss et al., 2010). Another measure of performance is the classification rate.

The main classification techniques are decision trees, neural networks, SVM (support vector machine), Naive Bayes classification and logistic regression. These techniques are detailed in Table 2 (Tan et al., 2009; Larose, 2005; Groth, 2000). Although each of these techniques uses a specific approach to classification, all perform the training based on data characteristics. An example of classification is the Akinator (2018), finding a character based on questions and answers about physical and behavioral characteristics.

Table 2. Description of data mining techniques.

Technique	Description	Example
Decision Tree	It has this name due to the appearance of a tree (Han & Kamber, 2006). It is constructed with the root, decision and leaf nodes, which are the questions, and the branches that are the answers (Larose, 2005). The algorithms occur in three steps (Groth, 2000): i) definition of dependent and independent variables from a data source; ii) examination of the impact of each variable on the result; and iii) definition of the variable that predicts the results of the other variables. The algorithms suggested by Jurka et al. (2013); were BAGGING (Breiman, 1996), RF (Liaw & Wiener, 2002) and TREE (Breiman et al., 1984).	<pre> graph TD A[Age?] -- "<25 years old" --> B[Student?] A -- ">25 years old" --> C[Credit rating?] B -- Yes --> D[Yes] B -- No --> E[No] C -- Bad --> F[No] C -- Good --> G[Yes] </pre>
Bayesian Classification	They predict the probability of the data belonging to certain classes. This technique is based on the Bayes' theorem and assumes that the data are independent of each other (Han & Kamber, 2006). Jurka et al. (2013) suggests the use of the BOOSTING algorithm (Freund & Schapire, 1997).	$P(H/X) = \frac{P(X/H) \cdot P(H)}{P(X)}$

Table 2. Continued...

Technique	Description	Example
Neural Networks	This technique simulates the functioning of the human brain, as a function of the large number of neurons, enabling learning based on experience (Larose, 2005). There are at least two types of neural networks: perceptron and multilayer (Tan et al., 2009). The algorithm SLDA (Blei & McAuliffe, 2010) is suggested by Jurka et al. (2013) to classify with this technique.	
SVM	The support vector machine raises training data to a higher dimension by looking for an optimal separation hyperplane (a greater distance separating different classes) (Feldman & Sanger, 2007; Han & Kamber, 2006). The SVM algorithm (Fan et al., 2005) is that indicated by Jurka et al. (2013).	
Logistic Regression	It is a special type of regression, which deals with categorical and independent variables (Groth, 2000). In binary classes, probabilities greater than 50% indicate the presence in the class " 1 " and probabilities less than 50% indicate the presence in the "0" class (Fuller et al., 2011). Jurka et al. (2013) suggests the use of the GLMNET algorithm (Friedman et al., 2010).	$y = b_0 + b_1x$

Usually the data is found in unstructured format, which makes it necessary to use text mining (Feldman & Sanger, 2007). This process uses techniques from the areas of information retrieval, natural language processing, information mining, and data mining (Ghosh et al., 2012).

The text mining process is composed of three steps (Weiss et al., 2010). The first step is the collection of documents and the creation of the database, called a corpus. The second step is the integration, cleaning and reduction of data, in which there is the conversion of all documents to the same format, to facilitate the next stages (Weiss et al., 2010). In this step there was tokenization, decomposing the texts into terms, called tokens (Hotho et al., 2005) (example: in the phrase "text mining has 3 steps" we have five tokens: i) text; ii) mining; iii) has; iv) 3; and v) steps). The stopwords removal eliminates irrelevant words (examples: a, o, one, from, to, with, etc.) (Han & Kamber, 2006). The stemming removes the word stem (example: reducing becomes reduce). After data cleaning, the data not removed are called the document collection dictionary (Hotho et al., 2005). The third stage of the text mining process transforms the data into numbers. There are three techniques diffused in the literature: i) Boolean, in which each term is considered present or absent in a document; ii) the frequency of terms (TF), in which the weight of a component in a document is the number of times it appears; and (iii) TF-IDF (inverted frequency), in which the weight of a component is determined by the product of the times number a term appears and the inverted frequency of terms (Liu, 2007).

3 Fake reviews identification

One of the challenges of the text mining is to identify fake reviews, as it is difficult to find them by reading them manually (Liu, 2012). The purpose of someone who writes fake reviews is to promote something or someone (hype spam) or to denigrate something or someone (defaming spam) (Liu, 2007). The first researches to find fake reviews looked for duplicates and near-duplicates, reviews with characteristics similar to each other (Jindal & Liu, 2007, 2008). After these researches, the subject grew in the literature (mainly American) according to Table 3.

Table 3. Synthesis of fake reviews in the literature.

Literature		Fei et al. (2013)	Lau et al. (2011)	Sharma & Lin (2013)	Malbon (2013)	Mukherjee et al. (2013a)	Mukherjee et al. (2013b)	Qian & Liu (2013)	Zhao et al. (2013)	Jindal & Liu (2008)	Mukherjee et al. (2011)	Mukherjee et al. (2012)	Ott et al. (2013)	Lu et al. (2013)	Ott et al. (2011)	Xie et al. (2012)	Jindal & Liu (2007)	Jindal et al. (2010)	Li et al. (2011)	Lappas (2012)	
Review types	Promote or denigrate	X	X	X		X	X			X				X	X	X	X	X	X	X	X
	Focus on features									X								X			
	Irrelevant									X								X			
Concentration areas	Fake promotion x good quality												X								
	Fake defamation x good quality																				
	Fake promotion x bad quality												X								
	Fake defamation x bad quality					X															
	Fake promotion x average quality												X								
	Fake defamation x average quality													X							
Way of acting	Individual							X				X									
	Group										X	X									
Approaches	Reviews		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	Authors		X			X				X	X	X	X					X			
	Product or service				X	X	X	X				X	X	X	X						X
Technique used	Decision tree																				
	Bayesian classification							X				X	X	X	X	X	X	X	X	X	X
	Neural network																				
	SVM	X	X			X	X	X			X	X	X	X	X						X
	Logistic regression			X						X	X	X	X				X		X		X

The classification of reviews types has three options: i) reviews that deceive readers or detection systems, promoting or denigrating something or someone; ii) reviews that

focus on other characteristics, such as brands, manufacturers and even sellers; and iii) advertising or irrelevant reviews, such as random texts (Jindal & Liu, 2008).

The classification of concentration areas is used according to the quality of the products and the type of reviews: fake promotion or fake defamation (Liu, 2007). Fake promotion reviews are often written by manufacturers or individuals who have economic interests' products and services (Liu, 2007). Reviews of fake defamation tend to be written by competitors (Liu, 2007).

The classification of the way of acting differentiates authors who work individually or in a group. By acting alone, the individual can post fake reviews as a single user, or as different users (Qian & Liu, 2013). Acting as a group, individuals work collaboratively, seeking to take control of reviews about a product or service (Liu, 2007). An alternative to find these groups is to rank the authors according to the probability of participating in groups (Mukherjee et al., 2011, 2012). There are differences in the behavior of individuals acting alone or in a group (Liu, 2007): i) the individual acting alone first builds a reputation as a good evaluator of products, posting real reviews. He then writes several fake reviews with different users. To confuse detection systems these individuals only write fake promotions or fake defamations, not both; ii) acting in a group, each individual submits reviews of the same polarity in the same product. These individuals write reviews at random intervals.

Another classification of fake reviews is by approaches. The first approach is to only analyze the reviews, relating to their ratings and their content. It is possible to identify non-trivial patterns (Sharma & Lin, 2013; Fei et al., 2013), look for different characteristics (Li et al., 2011), and analyze periods of postage concentration (Xie et al., 2012). The authors' approach consists of looking for abnormal behavior of the reviewers (Mukherjee et al., 2013a) and analyzing them from the point of view of time series (Jindal et al., 2010). The last approach is to compare information about the organizations and the products or services (Liu, 2007). There are works that use the approaches by reviews and authors at the same time, such as the search for factors that differentiate fake reviews and their respective authors (Lu et al., 2013) and search for the best technique for detecting fake reviews (Mukherjee et al., 2013b).

There are still works that do not fit into any of these classifications. The database creation, with 400 true positive reviews, 400 true negative reviews, 400 fake positive reviews, and 400 fake negative reviews (Ott et al., 2011, 2013; Review Skeptic, 2013). The evaluation of three classification systems (Ott et al., 2011): i) human, with people trained to determine the class of reviews; ii) based on psycholinguistics, with the terms of the reviews analyzed in relation to their meaning; and iii) text mining. The study of regulatory policies to avoid fake reviews, proposing an alliance approach to be added to existing laws (Malbon, 2013). The study of fake reviews from the customer's point of view, concluding that consumers tend to learn more from reviews of a product than from their own experience (Zhao et al., 2013). The last work found studies about fake reviews from the authors' point of view, creating a technique based on authenticity and the impact of the review (Lappas, 2012).

4 Methodological procedures

The work method used consisted of twelve steps. It was based on Saunders et al. (2009) and Cauchick et al. (2011) and is shown in Figure 2.

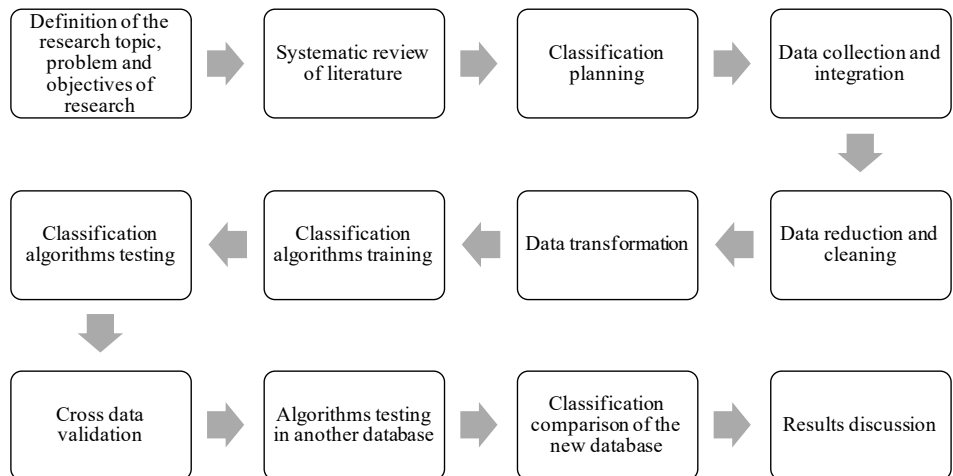


Figure 2. Work method.

The first step was the definition of an interest topic, the research problem and the objectives. Serious cases of fraud in texts on the Internet have been found. It was discovered that these frauds are called in the literature of fake reviews and are new problems and not studied in Brazil. Liu (2008) contributed to the definition of the real problem, since it has a brief introduction on the subject. Data mining classification techniques are used in the Brazilian literature in several areas, being recognized and appropriate to start the research in a new topic such as fake reviews. The second step was the systematic review of the literature in the research sources Google Scholar, Scopus, Scielo, Portal Capes and Ebscohost, in which works related to the topic were found.

The third step was the classification planning. We used 8 algorithms (Jurka et al., 2013): SVM, GLMNET, SLDA, BOOSTING, BAGGING, RF, TREE and CONCENSO. The CONCENSO use most of the classifications of other algorithms. The performance of the classification was evaluated in two ways: number of errors and correctness in classifying the reviews (classification rate), and the indicators of precision, recall and F-measure (Tan et al., 2009; Weiss et al., 2010). The performance of the classification was analyzed in relation to two dimensions: polarity (positive or negative review) and veracity (true or fake review).

The fourth step was the data collection and integration. The data was collected from Review Skeptic (2013), where a database of customer reviews of 20 most popular hotels in the city of Chicago/United States is available: Affinia Chicago; Hotel Allegro Chicago; Amalfi Hotel Chicago; Ambassador East Hotel; Conrad Chicago; Fairmont Chicago Millennium Park; Hard Rock Hotel Chicago; Hilton Chicago; Homewood Suites by Hilton Chicago Downtown; Hyatt Regency Chicago; InterContinental Chicago; James Chicago; Millennium Knickerbocker Hotel Chicago; Hotel Monaco Chicago; Omni Chicago Hotel; The Palmer House Hilton; Sheraton Chicago Hotel and Towers; Sofitel Chicago Water Tower; Swissotel Chicago; The Talbott Hotel. The integration of the data was performed in software R, since it is a language and an environment, which offers statistical and graphical tools, and that has open source (R Development Core Team, 2018). Originally each review collected was in a notepad file. With the use of R, these reviews were grouped in a Microsoft Excel file, with six columns: i) the identification of the review; ii) the hotel to which the review belongs; iii) the review;

iv) the polarity (0 for negative polarities and 1 for positive polarities) of the review; v) veracity (0 for true and 1 for fake) of the review; and vi) a random number. The random number was added to randomize the separation of reviews during data training to prevent patterns (eg, training the algorithms only in fake reviews).

The fifth step consisted of data reduction and cleaning, with the operations of stemming, removal of stopwords and tokenization. From this step, the operations occurred in the statistical software R with the use of the RTextTools package (Jurka et al., 2013). Accents, punctuations, and irrelevant characters have been removed. The size of the database has been reduced to facilitate data transformation. The sixth step was to transform the data into attribute-value tables using the TD-IDF technique (Liu, 2007).

The algorithms classification training was performed with 80% of the data for the training (Jurka et al., 2013; Kohavi, 1995). The reviews were sorted by the random number column and 80% of the data with lower random numbers (criterion chosen to separate the training data) were used. In the eighth step the trained algorithms were applied to the 20% of test data. Cross-validation of the data (Kohavi, 1995) tested the algorithms on 4 partitions (Jurka et al., 2013). The results of the 4 partitions were averaged and the algorithms were compared by the classification rate, precision, recall and f-measure indicators.

The tenth step was testing the algorithms in a Trip Advisor database (2018). The Trip Advisor database (2018) was chosen because of its importance in the hotel market and because no other databases were found with the classes known as Ott et al. (2011, 2013). We collected 100 English reviews from 10 hotels (10 reviews per hotel) with the highest popularity in the city of Porto Alegre in Brazil, as it is a tourist center in the south region, which presents international hotels. To perform the classification of the new database, the three best algorithms found in the previous step plus the CONCENSO were used. This number of algorithms facilitates the evaluation process, since there will be no analysis of indicators, since the true classes (true or fake) of the new database are unknown. Using the eight algorithms would generate an unnecessary discrepancy in the results.

The penultimate stage was to compare the performance of the classification of the new database. As the class of the Porto Alegre hotel base is unknown, the alternative was to compare the result of the classification with the algorithm of Myle Ott, available in Review Skeptic (2013). The performance of the comparison was evaluated by the similarity of results found (both algorithms classifying the reviews as true or fake). The last step was the discussion of the results found in the research.

5 Results

The eight algorithms SVM, GLMNET, SLDA, BAGGING, BOOSTING, RF, TREE and CONCENSO classified the test data from Ott et al. (2011, 2013) in true or fake using cross-validation with 4 data partitions. The algorithms were evaluated in twelve criteria: the percentage of correct answers in the classification of reviews; the percentage of correct answers in the classification of true reviews; the percentage of correct answers in the classification of fake reviews; the percentage of correct answers in the classification of positive reviews; the percentage of correctness in the classification of negative reviews; the percentage of correct answers in the classification of true positive reviews; the percentage of correct answers in the classification of true negative reviews; the percentage of correct answers in the

classification of fake positive reviews; the percentage of correct answers in the classification of fake negative reviews; precision; recall; and f-measure. The average of the result of the 4 data partitions for each criterion is illustrated in Figure 3 as a function of the algorithm used.

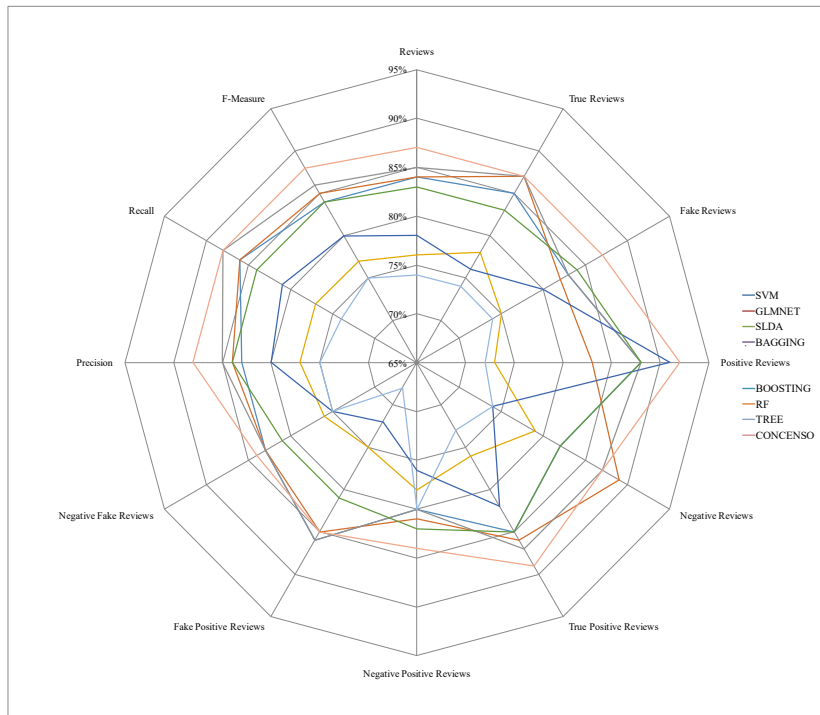


Figure 3. Results of the classification of the twelve criteria by algorithm.

The average of the algorithms results is shown in the following order: CONCENSO (87%), SLDA (85%), GLMNET (85%), SVM (84%), RF (83%), BOOSTING (79%), BAGGING (76%) and TREE (74%). The best result among the twelve criteria was 92% with the CONCENSO algorithm in the evaluation of positive reviews. Comparing with the other algorithms CONCENSO presented superior results in 7 of the 12 criteria.

The algorithms were tested in the Porto Alegre hotel database (Trip Advisor, 2018). We used the three algorithms (SVM, GLMNET, SLDA) with better results in the previous analysis plus the CONCENSO. The results of this new classification are illustrated in Figure 4 and Figure 5.

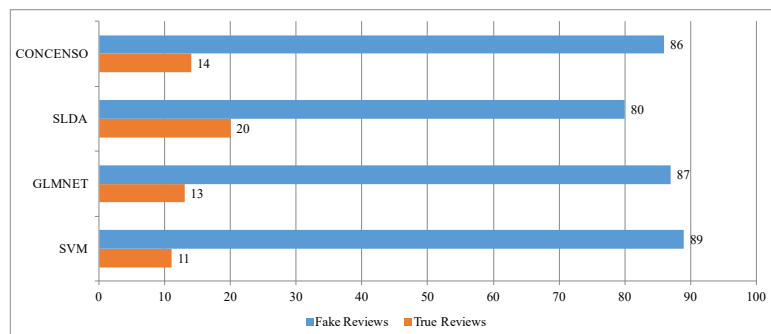


Figure 4. Results of the classification of the Porto Alegre hotel base by algorithm.

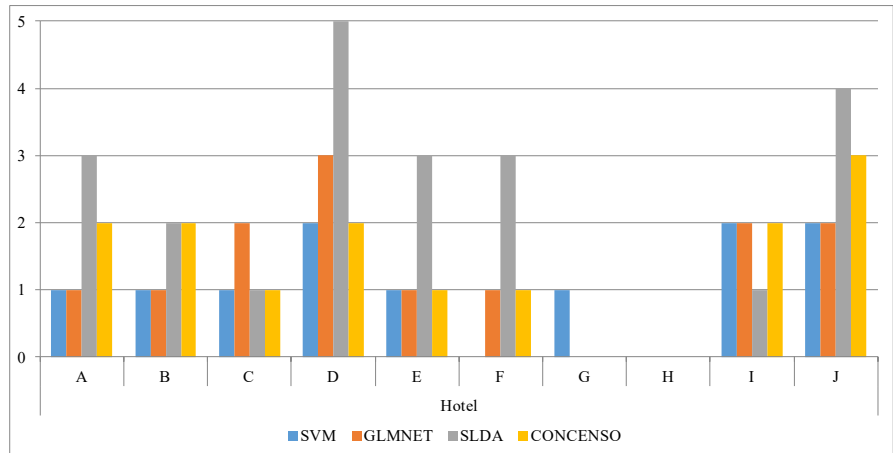


Figure 5. Results of the classification of fake reviews of the Porto Alegre base of hotels by algorithm and hotel.

The SLDA algorithm classified more reviews as fake. Note a descending order in the classification of fake reviews of the algorithms: SVM, GLMNET, CONCENSO and SLDA. In relation to hotels, none of the four algorithms found fake reviews in the Hotel H. The algorithm SLDA was the one that found the most fake reviews in the hotel D and J. Comparing the results of classification of the base of hotels of Porto Alegre with the classification accomplished in Review Skeptic (2013) a similarity was verified, as shown in Table 4.

Table 4. Comparison of the algorithms classification of the Porto Alegre base of hotels with Review Skeptic (2013).

Algorithms	Classification	Review Skeptic (2013)		Total
		True	Fake	
SVM	True	82	7	89
	Fake	10	1	11
	Total	92	8	100
GLMNET	True	81	6	87
	Fake	11	2	13
	Total	92	8	100
SLDA	True	73	7	80
	Fake	19	1	20
	Total	92	8	100
CONCENSO	True	79	7	86
	Fake	13	1	14
	Total	92	8	100

The SVM algorithm presented the greatest convergence with the Review Skeptic (2013) in the true reviews, whereas the GLMNET presented the biggest convergence in the fake reviews. Of the 92 reviews classified as true by the algorithm of Myle Ott, the SLDA converged only in 73, being the algorithm with the lowest similarity of results between the 4 algorithms.

6 Discussion and conclusion

This work evaluates classification techniques to identify fake reviews about products and services on the Internet. The research found 88% precision with the CONCENSO algorithm in the first database test (Ott et al., 2011, 2013). In the second test, the algorithms classified 100 hotel reviews in the Porto Alegre city (Trip Advisor, 2018). The algorithms SVM, SLDA, GLMNET and CONSENSO, for presenting the best results in the first test, were used. The algorithm that classified most reviews as true was the SVM with 89%, while the SLDA classified more reviews as fake with 20%. The algorithms were compared with Review Skeptic (2013). SVM and GLMNET classified 83% of reviews with the same class as the Myle Ott algorithm (Review Skeptic, 2013), showing similar results.

The results found are critical for business owners to separate true and fake reviews they receive about their products and services. Reading negative reviews helps managers analyze what customers do not like, enabling them to take corrective action. Reading positive reviews identifies the strengths of the products or services, and they cannot be lost. The research contributes by demonstrating the ability of algorithms to understand consumers' real reviews to products and services on the Internet. Separate what is fake and true is difficult, since people are controlled by feelings, emotions, and irrational thoughts that can influence their answers. The results also demonstrate the R software's ability to provide fast and consistent technology. Another contribution of the research was the systematic review of literature on fake reviews. It is expected that this work will be only the beginning of the research in fake reviews in Brazil. Brazilian businessmen may use techniques and algorithms to identify fake reviews for their products and services on the Internet. Brazilian researchers can create databases that facilitate the training of classification algorithms. For the production engineering the work contributes to: i) computational intelligence with the performance of algorithms for large amount of data; ii) information management dealing with reviews from several languages written on the internet every day; and iii) knowledge management by transforming this information into the decision-making of organizations.

This work needs to be understood in view of its limitations: i) the training was conducted with reviews in English, because the available database (Ott et al., 2011, 2013) is in this language; ii) the systematic review of the literature used the term fake review in Portuguese and fake review, spam review and opinion spam in English as presented by Liu (2007); iii) a Porto Alegre hotel base was used without the known classes since no other databases were found with the classes known in the literature; and iv) the veracity of the reviews of the Porto Alegre hotel base is not known, and for this reason it is not possible to evaluate the performance indicators of the classification. Considering this fact, the Review Skeptic (2013) was used as an alternative to evaluate the classification of the algorithms.

Recommendations for future work include: i) training algorithms to classify reviews in Portuguese; ii) to experiment emerging classification techniques in the literature; and iii) training algorithms to classify reviews from other databases, such as: industries, banks, retailers, universities and governments.

References

Akinator. (2018). Retrieved in 2018, October 26, from <https://pt.akinator.com/>

- Andrade, A. L., Seleme, A., Rodrigues, L. H., & Souto, R. (2006). *Pensamento sistêmico caderno de campo* (1. ed.). Porto Alegre: Bookman.
- Blei, D., & McAuliffe, J. (2010). Supervised topic models. In *Anais da 20ª Conferência Internacional de Sistemas de Processamento de Informações Neurais* (pp. 121-128). Vancouver: ACM.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140. <http://dx.doi.org/10.1007/BF00058655>.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees* (1st ed.). Wadsworth: Chapman & Hall.
- Cauchick, M. P. A., Fleury, A., Mello, C. H. P., Nakano, D. N., Lima, E. P., Turrioni, J. B., Ho, L. L., Morabito, R., Martins, R. A., Sousa, R., Costa, S. E. G., & Pureza, V. (2011). *Metodologia de pesquisa em engenharia de produção e gestão de operações* (2. ed.). Rio de Janeiro: Campus.
- Column Five. (2014). *Rave reviews: why do they matter most to local businesses*. Retrieved in 2018, October 26, from <https://www.columnfivemedia.com/work-items/infographic-rave-reviews-why-do-they-matter-most-to-local-businesses>
- Cormack, G. V. (2008). Email spam filtering: a systematic review. *Foundation and Trends in Information Retrieval*, 1(4), 335-455. <http://dx.doi.org/10.1561/15000000006>.
- Fan, R., Chen, P., & Lin, C. (2005). Working set selection using second order information for training support vector machines. *Journal of Machine Learning Research*, 6, 1889-1918.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11), 27-34. <http://dx.doi.org/10.1145/240455.240464>.
- Fei, G., Mukherjee, A., Liu, B., Hsu, M., Castellanos, M., & Ghosh, R. (2013). Exploiting burstiness in reviews for review spammer detection. In *Anais da 17ª Conferência Internacional de Mídias Sociais e Weblogs* (pp. 175-184). Cambridge: ACM.
- Feldman, R., & Sanger, J. (2007). *The text mining handbook* (1st ed.). Nova York: Cambridge University Press.
- Freund, Y., & Schapire, R. (1997). A decision theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139. <http://dx.doi.org/10.1006/jcss.1997.1504>.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1-22. <http://dx.doi.org/10.18637/jss.v033.i01>. PMID:20808728.
- Fuller, C., Biros, D., & Delen, D. (2011). An investigation of data and text mining methods for real world deception detection. *Expert Systems with Applications*, 38(7), 8392-8398. <http://dx.doi.org/10.1016/j.eswa.2011.01.032>.
- Ghosh, S., Roy, S., & Bandyopadhyay, S. (2012). A tutorial review on text mining algorithms. *International Journal of Advanced Research in Computer and Communication Engineering*, 1(4), 223-233.
- Groth, R. (2000). *Data mining: building competitive strategy* (2nd ed.). Nova Jersey: Prentice-Hall.
- Han, J., & Kamber, M. (2006). *Data mining: concepts and techniques* (3rd ed.). São Francisco: Elsevier.
- Hotho, A., Nürnbergger, A., & Paab, G. (2005). A brief survey of text mining. *Journal for Computational Linguistics and Language Technology*, 20(1), 19-62.
- Hu, M., & Liu, B. (2004). Mining opinion features in customer reviews. In *Anais da 19ª Conferência Nacional em Inteligência Artificial* (pp. 755-760). Palo Alto: ACM.

- Jindal, N., & Liu, B. (2007). Review spam detection. In *Anais da 16ª Conferência Internacional em World Wide Web* (pp. 1189-1190). Banff: ACM.
- Jindal, N., & Liu, B. (2008). Opinion spam and analysis. In *Anais da 8ª Conferência Internacional de Pesquisa na Web e Mineração de Dados* (pp. 219-230). Palo Alto: ACM.
- Jindal, N., Liu, B., & Lim, E. (2010). Finding unusual review patterns using unexpected rules. In *Anais da 19ª Conferência Internacional e Gestão do Conhecimento e da Informação* (pp. 1549-1552). Toronto: ACM.
- Jurka, T. P., Collingwood, L., Boydston, A. E., Grossman, E., & Atteveldt, W. (2013). RTextTools: a supervised learning package for text classification. *The R Journal*, 5(1), 6-12. <http://dx.doi.org/10.32614/RJ-2013-001>.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Anais da 14ª Conferência Mundial Internacional em Inteligência Artificial* (pp. 1137-1143). Montreal: ACM.
- Lappas, T. (2012). Fake reviews: the malicious perspective. In *Anais da 17ª Conferência Internacional em Aplicações de Linguagem Natural em Sistemas de Informação* (pp. 23-24). Groningen: ACM. http://dx.doi.org/10.1007/978-3-642-31178-9_3.
- Larose, D. (2005). *Discovering knowledge in data: an introduction to data mining* (1st ed.). Hoboken: Wiley.
- Lau, R. Y. K., Liao, S. Y., Kwok, R. C., Xu, K., Xia, Y., & Li, Y. (2011). Text mining and probabilistic language modeling for online review spam detection. *ACM Transactions on Management Information Systems*, 2(4), 2501-2530. <http://dx.doi.org/10.1145/2070710.2070716>.
- Li, F., Huang, M., Yang, Y., & Zhu, X. (2011). Learning to identify review spam. In *Anais da 22ª Conferência Mundial Internacional em Inteligência Artificial* (pp. 2488-2493). Barcelona: ACM.
- Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, 2(3), 18-22.
- Liu, B. (2007). *Web data mining* (2nd ed.). Nova York: Springer.
- Liu, B. (2008). *Opinion spam detection: detecting fake reviews and reviewers*. Illinois: University of Illinois. Retrieved in 2018, October 26, from <http://www.cs.uic.edu/~liub/FBS/fake-reviews.html>
- Liu, B. (2012). *Sentiment analysis and opinion mining* (1st ed.). Nova York: Morgan & Claypool Publishers. <http://dx.doi.org/10.2200/S00416ED1V01Y201204HLT016>.
- Lu, Y., Zhang, L., Xiao, Y., & Li, Y. (2013). Simultaneously detecting fake reviews and review spammers using factor graph model. In *Anais da 5ª Conferência Anual da ACM em Ciência WEB* (pp. 225-233). Paris: ACM. <http://dx.doi.org/10.1145/2464464.2464470>.
- Malbon, J. (2013). Taking fake online consumer reviews seriously. *Journal of Consumer Policy*, 36(2), 139-157. <http://dx.doi.org/10.1007/s10603-012-9216-7>.
- Mitchell, T. (1999). Machine learning and data mining. *Communications of the ACM*, 42(11), 30-36. <http://dx.doi.org/10.1145/319382.319388>.
- Mukherjee, A., Kumar, A., Liu, B., Wang, J., Hsu, M., Castellanos, M., & Ghosh, R. (2013a). Spotting opinion spammers using behavioral footprints. In *Anais da 19ª Conferência Internacional em Descoberta de Conhecimento e Mineração de Dados* (pp. 632-640). Chicago: ACM.
- Mukherjee, A., Liu, B., & Glance, N. (2012). Spotting fake reviewer groups in consumer reviews. In *Anais da 21ª Conferência Internacional em World Wide Web* (pp. 191-200): Lyon: ACM. <http://dx.doi.org/10.1145/2187836.2187863>.
- Mukherjee, A., Liu, B., Wang, J., Glance, N., & Jindal, N. (2011). Detecting group review spam. In *Anais da 20ª Conferência Internacional em World Wide Web* (pp. 93-94). Hyderabad: ACM.

- Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. (2013b). What yelp fake review filter might be doing. In *Anais da 17ª Conferência Internacional em Mídias Sociais e Weblogs* (pp. 409-418). Palo Alto: ACM.
- Ott, M., Cardie, C., & Hancock, J. (2013). Negative deceptive opinion spam. In *Anais da 21ª Conferência Internacional em World Wide Web* (pp. 497-501). Hyderabad: ACM
- Ott, M., Choi, Y., Cardie, C., & Hancock, J. (2011). Finding deceptive opinion spam by any stretch of the imagination. In *Anais do 49º Encontro Anual da Associação para Linguística Computacional: Tecnologias de Linguagem Humana* (pp. 309-319). Portland: ACM.
- Qian, T., & Liu, B. (2013). Identifying multiple userids of the same author. In *Anais da 11ª Conferência em Métodos Empíricos para Processamento de Linguagem Natural* (pp. 1124-1135). Seattle: ACM.
- R Development Core Team. (2018). *What is R?*. Vienna. Retrieved in 2018, September 22, from <https://www.r-project.org/about.html>
- Review Skeptic. (2013). *Review Skeptic is based on research at Cornell University that uses machine learning to identify fake hotel reviews with nearly 90% accuracy*. Retrieved in 2018, September 22, from <http://reviewskeptic.com/>
- Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students* (5th ed.). Londres: Prentice Hall.
- Sharma, K., & Lin, K. (2013). Review spam detector with rating consistency check. In *Anais da 51ª Conferência Regional da ACM Sudoeste* (pp. 341-346). Savannah: ACM. <http://dx.doi.org/10.1145/2498328.2500083>.
- Tan, P., Steinbach, M., & Kumar, V. (2009). *Introdução ao data mining* (1. ed.). Rio de Janeiro: Moderna.
- Trip Advisor (2018). *Hóteis em Porto Alegre*. Retrieved in 2018, September 22, from http://www.tripadvisor.com.br/Hotels-g303546-Porto_Alegre_State_of_Rio_Grande_do_Sul-Hotels.html
- Weiss, S., Indurkha, N., & Zhang, T. (2010). *Fundamentals of predictive text mining* (1st ed.). Nova York: Springer.
- Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z.-H., Steinbach, M., Hand, D. J., & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1), 1-37. <http://dx.doi.org/10.1007/s10115-007-0114-2>.
- Xie, S., Wang, G., Lin, S., & Yu, P. S. (2012). Review spam detection via temporal pattern discovery. In *Anais da 18ª Conferência Internacional em Descoberta do Conhecimento e Mineração de Dados* (pp. 823-831). Beijing: ACM.
- Zhao, Y., Yang, S., Narayan, V., & Zhao, Y. (2013). Modeling consumer learning from online product reviews. *Marketing Science*, 32(1), 153-169. <http://dx.doi.org/10.1287/mksc.1120.0755>.