Sentiment Analysis towards Actionable Intelligence via Deep Learning

Shorouq Fathi Eletter

Al Ain University, Al Ain, UAE

Abstract **- The exponential growth of unstructured data and the ability of businesses to utilize such data in decision-making have led to competitive advantages. The knowledge provided by analyzing unstructured data is crucial for product developers or service providers because it might affect the sustainability of the business. Sentiment analysis is used to gain an understanding of the attitudes, opinions, and emotions expressed within an online review. Naïve Bayes (NB), logistic regression (LR), decision trees (DT), deep learning (DL), and support vector machines (SVM) were used to build a classification model. In the data mining settings, the classification accuracy is the best metric to highlight the best classifier. The DL classifier outperformed other models in terms of accuracy rate. Classifying customers' feelings toward a product or service is critical for providing actionable insights. Utilizing such models will help to analyze huge volumes of reviews, saving both time and costs.**

Keywords **– Sentiment analysis, Text mining, Deep learning, Computational cost, Classification, Natural language processing (NLP).**

1. Introduction

The emergence of the information era has resulted in advancements in hardware and software technologies [1]. Historically, computers were utilized to process and automate the decision process of structured data using artificial intelligence techniques

DOI: 10.18421/TEM94-44
https://doi.org/10.18421/TEM94-44

Corresponding author: Shorouq Fathi Eletter,
Al Ain University, Al Ain, UAE.
Email: shorouq.eletter@aau.ac.ae

Received: 29 July 2020.
Revised: 24 September 2020.
Accepted: 06 October 2020.
Published: 27 November 2020.

while unstructured data were left for humans to process, using their intuition to make decisions manually.

The exponential growth of unstructured data and the ability of businesses to utilize such data in decision-making led to competitive advantages over rivals that lagged behind [2]. Nowadays, businesses are utilizing text analytics and text mining to make more informed decisions. In addition, a variety of available social media channels have contributed to the rapid growth of unstructured data. Individuals rely on such data to make decisions before they purchase a product or service, hotel booking, restaurant experience, etc. Business organizations have recently started taking advantage of the wealth of information on the internet to produce new knowledge as well as exploit available information [3]. Customers' reviews have become invaluable for business in the e-commerce and big data era [4]. The knowledge gleaned from analyzing unstructured data is crucial for product developers and service providers because it might affect the sustainability of a business.

Historically, social networking outlets such as Facebook have utilized recommendation systems to suggest connections for users. Music and media applications also utilize similar machine learning to make recommendations to numerous songs, videos, movies, etc., based on their users' previous experiences [5]. Nowadays, businesses are extracting valuable insights from the content generated on social media channels.

Providing online reviews also enhances potential consumers' inferences regarding a business's trustworthiness [6]. In the absence of knowledge or experience, individuals rely on word of mouth or others' opinions to make decisions. Thanks to an influx of such data resources from reviewers' recommendations, ratings, opinions, and feedback are available to individuals over the internet. In other words, they are guided by others' sentiments about the product or service. The process of identifying the polarity (i.e., positive or negative) of the expressed opinion is called opinion mining [3].

Sentiment analysis (opinion mining) is still a somewhat nascent arena in which businesses can utilize and extract opinions and subjectivity from the text [7]. Sentiment analysis aims to reveal the settled feeling or emotion behind textual data and to identify favorable and unfavorable attitudes, opinions, and emotions. Sentiment analysis is linked to text mining and natural language processing (NLP) [8]. The immensely huge volume of opinionated data recorded on social media platforms has boosted the rapid growth of the field [9]. Sentiment analysis is promoted as a fruitful tool to discover valuable insights from the immense volume of unstructured data [7]. "Tasks regarding the sentiment analysis rely on a combination of machine learning, information retrieval, and NLP techniques" [10].

Users post explicit opinions on social media. However, sometimes customers tend to write negative sentences in a positive way or vice versa. Given the lack of consistencies between customers' ratings and sentiment analysis evaluations, new approaches have to be developed to address the positive, negative, and neutral comments within sentiment analysis models. Such models can be utilized to uncover causes behind negative comments [11]. Machine learning models such as support vector machine (SVM), Naïve Bayes (NB), and k-nearest neighbor have also been successfully used for text classification [12]. Supervised classification methods like NB, logistic regression (LR), decision trees (DT), deep learning (DL), and SVM were used to build a classification model; utilizing such models will help analyze huge volumes of reviews, saving both time and costs.

2. Literature Review

Online reviews are critical for the success of online sales in the hospitality and tourism industry, in which customer satisfaction is crucial for the sustainability of the service [13]. An online review is a significant form of marketing strategy [14]. Individuals are social in nature; they learn from interactions that might impact their behaviors. Customer behavior is an eclectic field comprising active interactions and exchanges [15]. The impact of how people communicate with others is clarified by research on emotional mimicry and contamination [16]. However, communication platforms have changed with the increasing use of social media channels as the prominent channel of communication, leading individuals to share their personal experiences and express their opinions using informal language [17]. Reviewers convey their experiences in an explicit knowledge format. Therefore, individuals learn from online conversations and reviews of customers' experience and act accordingly.

As customers openly share their experiences, it may affect a business's sustainability [18]. Online reviews are a major emotional influential factor as they are customers' testaments shared after their experiences. The differential outcomes' effect of online reviews refers partly to readers' cognitive personalization. The review is perceived as truthful and powerful to sway over their future intentions, especially when sharing sense resonance with the reviewer [19]. At the same time, consumers are rational when they make decisions; they utilize principles to make procedural choices that are aligned with their needs. Their choices fulfill their needs relative to their personal objectives or self-interest [20]. Ultimately, an online recommendation is a directive act that strongly and emotionally affects consumers' choices [21].

3. Methodology

Sentiment analysis is a task that analyzes people's opinions, sentiments, and feelings toward things. Such analysis is based on personal faith and judgment bases that are subjective in nature and not initiated on rationale reasoning [22]. Sentiment classification aims to identify sentiments' polarity within a text [18]. Sentiment analysis is intended to determine the relative positivity or negativity of text [23]. The process involves the transformation of unstructured data into a numerical format that is then utilized by data-mining algorithms [24]. Sentiment classification is meant to transform text into meaningful groups.

Sentiment analysis is interdisciplinary in nature and leverages a variety of NLP techniques to classify the sentiment in the text [17]. A text-mining methodology based on NLP was used. NLP is closely connected to areas in artificial intelligence, computational linguistics, mathematics, and information science. The sentiment analysis tasks comprise the following steps: identify polarization by labeling the sentiment of the text as positive, negative, or neutral; use features to structure the text, known as aspect selection/extraction; and use one of the classifications (e.g., machine learning or lexicon approaches) to classify the text.

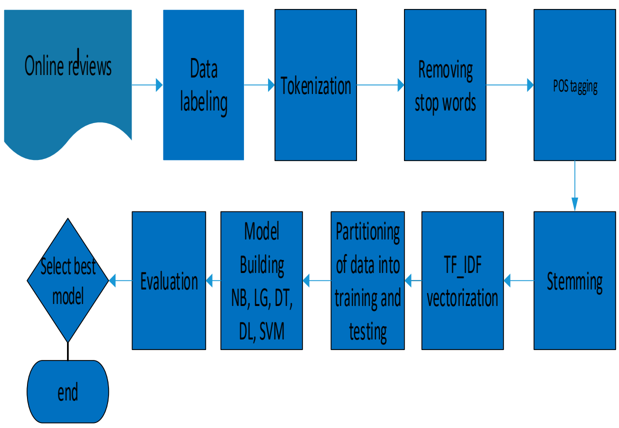

Figure 1. Text-mining methodology

A balanced dataset that contains 1,000 restaurant reviews was taken from the Academic Yelp dataset [25]. Reviews were labeled "1" or "0" based on n-grams depending on positive or negative feelings, attitudes, functioning, etc. For example, "a great touch," "service was very prompt," and "the food, amazing" were labeled as positive whereas "did not like it at all," "service sucks," and "we will never go again" represent negative sentiment. The dataset contained 500 positive and 500 negative sentiments. The reviews were labeled as positive or negative based on their semantic context. RapidMiner was used to extract and transform the text. The text was cleaned by removing numbers and punctuation as well as stop words; it was also transformed into lower case letters to decrease distinctions in the same words. Second, tokenization, which helps to chop up a sequence of characters into meaningful pieces (i.e., tokens) [26], was used to filter out certain characters, punctuation, etc. Tokenization transformed the text into attributes.

The stemming and lemmatization converted words into bases by decreasing inflectional and derivationally related forms. The tokens were represented in different grammatically categories that shares a common root form that is semantically related. Each token was then assigned a part of speech (verb, noun, adjective, etc.), known as part of speech (POS) tagging. N-grams involve generating groups of a set of tokens that might have more meaning than individual words or a stronger emotion. When generating n-grams, *n* refers to the number of words so there can be unigrams for one, bigrams for two, and trigrams for three [21].

Next, term frequency–inverse document frequency (TF–IDF) was used to calculate the numerical statistics in terms of the number of times a word occurred in a text and also reflect the significance of that word in the whole corpus, (Equation 1):

$$TF - IDF = \frac{n_t}{N} log \frac{k}{K_n} \qquad (1),$$

in which n_t is the number of times a word t is mentioned in a document (review), N is the number of words in the document, K represents the total number of documents, and K_n represents the number of documents that contain word t.

The dataset was divided into training and testing subsets, accounting for 70% and 30%, respectively. Five supervised classification methods (i.e., NB, LR, DT, DL, and SVM) were used to build a classification model. The metric used to assess the best in the data mining settings is the highest predictive accuracy (Equation 2) as well as the computational cost as a measure of efficiency [2]. In the data-mining settings, the classification accuracy (AC) (Equation 2) is the best metric to highlight the best classifier. Therefore, classification accuracy is used to highlight a proposed model to build a recommendation model. An overview of the used models is discussed next.

Table 1. Confusion Matrix

		Actual sentiments	
		Positive	Negative
Predicted sentiments	Positive	True positive (TP)	False positive (FP)
	Negative	False negative (FN)	True negative (TN)

$$AC = \frac{TP+TN}{TP+FN+TN+FP} \qquad (2)$$

NB is a probabilistic classifier based on Bayes' theorem, along with a strong independence assumption. The NB classifier assumes class conditional independence [27]. Despite its simplicity, NB has performed surprisingly well for text classification and also tends to perform best for certain problem classes with highly dependent features [28]. Furthermore, it treats each word independently, without considering the location of the term in the sentence. In this study, NB was used to calculate the probability of each term corresponding to a label in Equation 3:

$$p\ (class|features) = \frac{\mathrm{p(class)*p(features|class)}}{\mathrm{p(features)}} \quad (3),$$

in which *p(class)* is the prior probability of the class in the dataset, *P (features| class)* is the prior probability of a feature related to a class, and *p(feature)* is the prior probability of a feature that has occurred [29].

SVM, a non-probabilistic classification model, was first introduced to perform binary classification tasks. It builds a hyperplane that acts as a separator to designate the decision boundaries between data points with different classes (labels). The best hyperplane is one that can maintain the maximum distance between two support vectors of different classes [29]. SVM builds a hyperplane that maximizes the functional margin to the nearest training data points of any class to perform classification. Furthermore, as the margin increases, the generalization error of the classifier decreases. During training, the SVM algorithm classifies the training cases labeled "1" or "0" into one of the two categories to build a model. During the testing phase, the model is used to map new cases into that same space and to predict their belonging to a category based on the side of the gap on which they fall.

Deep learning (DL) has emerged as a powerful modeling technique due to its ability to produce state-of-the-art prediction results [9]. DL refers to constant learning of real-valued text representation using neural network approaches [30]. DL techniques are popular for classification tasks as they provide automatic feature extraction associated with richer representation capabilities and better performance compared to traditional techniques (e.g., surface methods) [2].

DL is the utilization of multiple layers of artificial networks to learn tasks, much like what the human brain does. A neural network comprises a number of fully connected units called neurons, which are analogous to the biological neurons [9]. DL emerged with powerful computational models that perform sentiment analysis tasks, including sentiment detection, polarity classification, and sentiment lexicon learning [30]. A typical feedforward neural network comprises three layers: input, hidden, and output layers of fully connected units called neurons [31]. A neuron is the basic computational component of the network. Weighted connections are used to link neurons across layers. The net picks up the relationship between inputs and output by adjusting the connection weights [9].

Deep learning represents words in vectors such that each vector denotes the frequency of a word in a document. Data are fed into the input layer such that vector a_i is given as input to a neuron, which is multiplied by a weighted matrix w. The dot product ($x_i = w.\ a_i$) is transferred to a hidden layer (embedded, assuming that the net has only one hidden layer), where an activation (transfer) function is applied. Similarly, the output of the hidden layer is multiplied by a weighted matrix and transferred to the output layer, which applies a transfer function (softmax) to produce a predicted output. Then the predicted output is compared with the actual value, and an error is calculated. During training, the error is propagated backward to adjust the weights until a loss function is minimized [32], and stochastic gradient descent via backpropagation is utilized to minimize the cross-entropy loss (loss function for softmax output; [9].

A decision tree (DT) is a valuable tool for data analysis and data classification [33]. DT is used successfully in sentiment analysis, where it performs well on large datasets. It is a tree-based classifier where a root node represents a feature and a leaf node represents the label (e.g., positive, or negative). DT is flow chart of nodes connected through branches. The top node is the root node. The leaf node at the end of the tree displays a label to the input. During training, DT uses a condition on the attribute value to divide the training data hierarchically. The condition is based on the presence or absence of a word. The tree recursively continues the division process until terminal nodes represent the small numbers of features that belong only to one class [29].

Logistic regression (LR) is model broadly used in classification problems. It is a linear regression in which the dependent variable is categorical (binary). The LR model is used to predict the probability of a binary outcome using a set of explanatory variables. LR is easy to interpret because of the additive linear combination of the independent variables.

4. Results and Discussion

Table 2 summarizes the classification results of the five models. As shown, the deep learning classifier correctly predicted 93.3% of the reviews. Decision trees, logistic regression, and SVM classifiers were able to correctly predict 89.3% of the reviews. On the other hand, the NB classifier predicted 86.0% of the cases correctly. The deep learning classifier scored the highest accuracy rate whereas the NB scored the lowest accuracy rate. However, deep learning scored the highest duration time for model building (761s), while decision trees scored the lowest duration (239s). The logistic regression, SVM, and NB scored 267, 351, and 364, respectively. As shown, deep learning outperformed the four classifiers used in sentiment classification. These results aligned with [34] findings that DL outperformed a number of machine learning algorithms in image processing and text mining.

Table 2. Classification Results

Model	AC	Computational Cost (second)
DT	89.3%	239
DL	93.3%	761
LR	89.3%	267
NB	86.0%	364
SVM	89.3%	351

The bar chart in Figure 2 shows the classification accuracy of the five models as a percentage. The higher the percentage achieved by a classifier, the more predictive ability it has. This figure indicates that deep learning ranked highest.

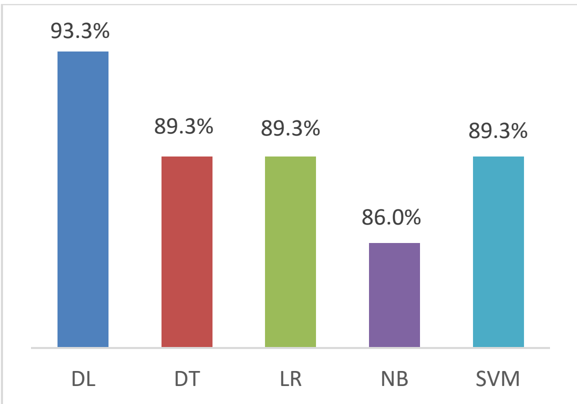

Figure 2. Classification accuracy of different models

The bar chart in Figure 3 shows the computational costs with training time. It is clear that a trade-off exists between classification accuracy and computational duration as deep learning needed 791 seconds whereas decision trees needed 239 seconds, logistic regression needed 267 seconds, SVM needed 351, and NB needed 364 seconds. [2] stated that computational costs (i.e., computational resources) might be one drawback of these approaches.

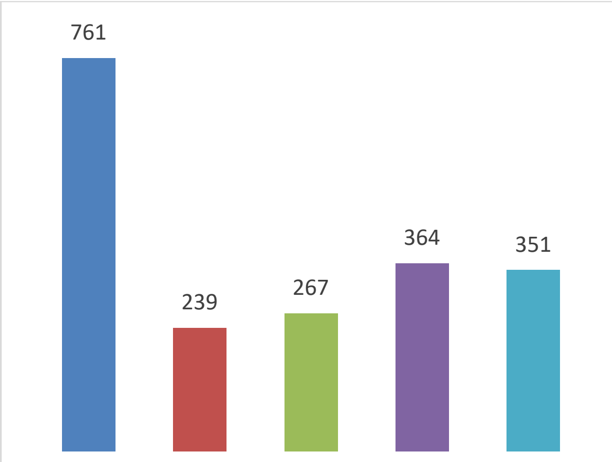

Figure 3. Total computational cost

On the other hand, the computational cost of deep learning was really very low compared to that of the manual analysis. The highest accuracy level was 93.3% for DL, [25] obtained a result of 91.25%. The current study's results related to deep learning were very good compared to those of [25]. [8] measured the classification accuracy of SVM using different datasets, including Yelp Restaurant reviews; the highest accuracy level was 83.5%.

5. Conclusions

This research aimed to identify the best classifier for restaurant reviews among five different classifiers. The analysis of customers' restaurant reviews provide management with valuable insights into what customers like so the restaurant continues to excel in that regard. Regarding what customers do not like, restaurant management should try to avoid it in the future to help management resolve any issues that can influence future reviews. In this way, new customers will have a different and more positive experience. The analysis herein can also help management monitor the functional and operational processes more closely compared to those for which this service was created. It will provide them with insights about where they are going so they can take actions to improve their processes.

Classifying customers' feelings toward a product or service is critical for providing actionable insights. However, manual classification of such reviews can be time consuming and labor intensive [9]. The deep learning classifier showed a higher classification accuracy level among the different models used. However, one of the limitations of using deep learning is its high computational cost.

Sentiment analysis (opinion mining) is highly dependent on the analysis of customers' reviews from social media controlled by platform owners. Such ownership gives them the right to publicize the information they want. As owners make money from the reviews, they might be biased and show only one side of the truth. [7] stated that social media platform owners are willing to share their data to certain limit which might impact sentiment analysis and whether the data can be legally mined. The complexity of human language in writing a positive sentence that implicitly reflects a negative feeling is a challenge for sentiment analysis. In addition, humans are good at communicating their feelings through body language and body language such facial expressions as well as verbal sentences, and it might be difficult for a machine to capture such feelings.

References

[1]. Bettencourt, L. A., Lusch, R. F., & Vargo, S. L. (2014). A service lens on value creation: marketing's role in achieving strategic advantage. *California management review*, *57*(1), 44-66.

[2]. Araque, O., Corcuera-Platas, I., Sánchez-Rada, J. F., & Iglesias, C. A. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. *Expert Systems with Applications*, *77*, 236-246. https://doi.org/10.1016/j.eswa.2017.02.002

[3]. Tsytsarau, M., & Palpanas, T. (2012). Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, *24*(3), 478-514. https://doi.org/10.1007/s10618-011-0238-6

[4]. Zhao, Y., Xu, X., & Wang, M. (2019). Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *International Journal of Hospitality Management*, *76*, 111-121.

[5]. McAuley, J., & Leskovec, J. (2013, October). Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM conference on Recommender systems* (pp. 165-172). https://doi.org/10.1145/2507157.2507163

[6]. Geetha, M., Singha, P., & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels-An empirical analysis. *Tourism Management*, *61*, 43-54.

[7]. Philander, K., & Zhong, Y. (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *International Journal of Hospitality Management*, *55*(2016), 16-24. https://doi.org/10.1016/j.ijhm.2016.02.001

[8]. Bhuvaneswari, K., & Parimala, R. (2017). Sentiment Reviews Classification using Hybrid Feature Selection. *Int. J. Database Theory Appl.*, *10*(7), 1-12. https://doi.org/10.14257/ijdta.2017.10.7.01

[9]. Zhang, F., Fleyeh, H., Wang, X., & Lu, M. (2019). Construction site accident analysis using text mining and natural language processing techniques. *Automation in Construction*, *99*, 238-248.

[10]. Schuller, B., Mousa, A. E. D., & Vryniotis, V. (2015). Sentiment analysis and opinion mining: on optimal parameters and performances. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *5*(5), 255-263. https://doi.org/10.1002/widm.1159

[11]. Valdivia, A., Luzón, M. V., & Herrera, F. (2017). Sentiment analysis in tripadvisor. *IEEE Intelligent Systems*, *32*(4), 72-77.

[12]. Das, M. K., Padhy, B., & Mishra, B. K. (2017, January). Opinion mining and sentiment classification: A review. In *2017 International Conference on Inventive Systems and Control (ICISC)* (pp. 1-3). IEEE.

[13]. Schuckert, M., Liu, X., & Law, R. (2015). Hospitality and tourism online reviews: Recent trends and future directions. *Journal of Travel & Tourism Marketing*, *32*(5), 608-621.

[14]. Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism management*, *29*(3), 458-468. https://doi.org/10.1016/j.tourman.2007.05.011

[15]. Mattila, A. S. (2004). Consumer behavior research in hospitality and tourism journals. *International Journal of Hospitality Management*, *23*(5), 449-457.

[16]. Fischer, A. H., & Van Kleef, G. A. (2010). Where have all the people gone? A plea for including social interaction in emotion research. *Emotion Review*, *2*(3), 208-211.

[17]. Fernández-Gavilanes, M., Álvarez-López, T., Juncal-Martínez, J., Costa-Montenegro, E., & González-Castaño, F. J. (2016). Unsupervised method for sentiment analysis in online texts. *Expert Systems with Applications*, *58*, 57-75.

[18]. Valdivia, A., Luzón, M. V., & Herrera, F. (2017). Sentiment analysis in tripadvisor. *IEEE Intelligent Systems*, *32*(4), 72-77.

[19]. Xia, L., & Bechwati, N. N. (2008). Word of mouse: the role of cognitive personalization in online consumer reviews. *Journal of interactive Advertising*, *9*(1), 3-13. https://doi.org/10.1080/15252019.2008.10722143

[20]. Faiers, A., Cook, M., & Neame, C. (2007). Towards a contemporary approach for understanding consumer behaviour in the context of domestic energy use. *Energy Policy*, *35*(8), 4381-4390. https://doi.org/10.1016/j.enpol.2007.01.003

[21]. Guerreiro, J., & Rita, P. (2020). How to predict explicit recommendations in online reviews using text mining and sentiment analysis. *Journal of Hospitality and Tourism Management*, *43*, 269-272.

[22]. Greco, F., & Polli, A. (2020). Emotional Text Mining: Customer profiling in brand management. *International Journal of Information Management*, *51*, 101934. https://doi.org/10.1016/j.ijinfomgt.2019.04.007

[23]. Annett, M., & Kondrak, G. (2008, May). A comparison of sentiment analysis techniques: Polarizing movie blogs. In *Conference of the Canadian Society for Computational Studies of Intelligence* (pp. 25-35). Springer, Berlin, Heidelberg.

[24]. Williams, T. P., & Gong, J. (2014). Predicting construction cost overruns using text mining, numerical data and ensemble classifiers. *Automation in Construction*, *43*, 23-29.

[25]. Kotzias, D., Denil, M., De Freitas, N., & Smyth, P. (2015, August). From group to individual labels using deep features. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 597-606). https://doi.org/10.1145/2783258.2783380

[26]. Rashid, J., Shah, S. M. A., & Irtaza, A. (2019). Fuzzy topic modeling approach for text mining over short text. *Information Processing & Management*, *56*(6), 102060.

[27]. Khairnar, J., & Kinikar, M. (2013). Machine learning algorithms for opinion mining and sentiment classification. *International Journal of Scientific and Research Publications*, *3*(6), 1-6.

[28]. Das, M. K., Padhy, B., & Mishra, B. K. (2017, January). Opinion mining and sentiment classification: A review. In *2017 International Conference on Inventive Systems and Control (ICISC)* (pp. 1-3). IEEE. https://doi.org/10.1109/ICISC.2017.8068637

[29]. Sharma, D., Sabharwal, M., Goyal, V., & Vij, M. (2020). Sentiment Analysis Techniques for Social Media Data: A Review. In *First International Conference on Sustainable Technologies for Computational Intelligence* (pp. 75-90). Springer, Singapore. https://doi.org/10.1007/978-981-15-0029-9_7

[30]. Tang, D., Qin, B., & Liu, T. (2015). Deep learning for sentiment analysis: successful approaches and future challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *5*(6), 292-303. https://doi.org/10.1002/widm.1171

[31]. Eletter, S. F., & Yaseen, S. G. (2017). Loan decision models for the Jordanian commercial banks. *Global Business and Economics Review*, *19*(3), 323-338.

[32]. Rojas-Barahona, L. M. (2016). Deep learning for sentiment analysis. *Language and Linguistics Compass*, *10*(12), 701-719.

[33]. Murthy, S. K. (1998). Automatic construction of decision trees from data: A multi-disciplinary survey. *Data mining and knowledge discovery*, *2*(4), 345-389.

[34]. Sagha, H., Cummins, N., & Schuller, B. (2017). Stacked denoising autoencoders for sentiment analysis: a review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *7*(5), e1212.