Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 185 (2021) 370–379

Procedia Computer Science

www.elsevier.com/locate/procedia

Complex Adaptive Systems Conference Theme: Big Data, IoT, and AI for a Smarter Future
Malvern, Pennsylvania, June 16-18, 2021

# Effect of Negation in Sentences on Sentiment Analysis and Polarity Detection

Partha Mukherjee[a]* , Youakim Badr[a], Shreyesh Doppalapudi[a], Satish M. Srinivasan[a], Raghvinder S. Sangwan[a], Rahul Sharma[a]

*The Pennsylvania State University, Great Valley, 30E. Swedesford Road, Malvern, PA-19355, USA*

**Abstract**

Sentiment analysis is one of the sub-domains of Natural Language Processing (NLP) that is of piqued interest in the research community. With the advent of e-commerce and social media, more and more customer opinions are being provided online in the written text form. Nowadays, sentiment analysis provides a way for companies to understand customer opinions towards products and services in a global marketplace. Negative sentences or using negations in sentences have a significant impact on sentiment polarity detection. Inappropriate processing of negations in leads to biases and misclassification of sentiments. In this paper, we provide a novel end-to-end sentiment analysis approach to handle negations, along with the inclusion of negation identification and negation scope marking. Our approach introduces a customized negation marking algorithm for explicit negation detection and perform experiments on sentiment analysis with different machine learning algorithms such as Naïve Bayes, Support Vector Machines, Artificial Neural Network (ANN), and Recurrent Neural Network (RNN) on sentiment analysis of Amazon reviews, specifically of cell phones. By evaluating the effect of the negation algorithm on the sentiment analysis tasks, the RNN achieved the best accuracy of 95.67% when combined with our negation marking processing, exceeding its accuracy without any identification of negative sentences. Further, our approach was applied to another dataset of Amazon reviews and demonstrated a significant improvement in the overall accuracy.



*Keywords: Sentiment Analysis; Deep Learning; Machine Learning; Natural Language Processing Negation Identification*

* Corresponding author. Tel : +1-610-648-3259
*Email address: pom5109@psu.edu*

1877-0509 © 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the Complex Adaptive Systems Conference, June 2021.
10.1016/j.procs.2021.05.038

## 1. Introduction

Natural Language Processing (NLP) is the branch of Artificial Intelligence (AI) that is concerned with empowering computers with capabilities to process and understand human language text. Sentiment analysis – also called opinion mining – is one particular sub-domain of NLP and becomes an integral part of analyzing peoples views on various issues posted on different social media platforms and online forums. Sentiment analysis is the field of study that analyses people's opinions, sentiments, evaluations, attitudes, and emotions from written languages [1]. With the more and more user-generated text being derived from the growing usage of Internet marketplaces and e-commerce websites, there is an overwhelming need to understand human sentiments. Sentiment analysis provides means for companies and businesses to understand the global marketplace and customer sentiment on a large scale. In addition, the advent of social networks has provided platforms for users to express their sentiment towards products and services more freely. Analyzing brand sentiments in social networks have provided a means for companies to evaluate their prospect in the market competition. This makes sentiment analysis a useful decision-making tool for companies.

With sufficient technical leverage, sentiment-mining techniques can be exploited for the creation and automated upkeep of reviews and opinion aggregation websites, in which opinionated text and videos are continuously gathered from the Web and not restricted to just product reviews, but also broader topics such as political issues and brand perception [2].

An essential aspect of sentiment analysis is the identification of negation in written text. The presence of the word negation can change the polarity of the text, and if not handled properly, it will affect the performance of the sentiment classification [3]. Traditional approaches to negation identification include the usage of the lexicon for negation matching and identification of explicit negative words such as not, but, never, none, and so on. Negation identification is essential not only for sentiment polarity classification but also for other fields in NLP, such as Name Entity Recognition [4], Emotion Mining [5], and Syntactic Parsing [6]

In this paper we develop an end-to-end pipeline for sentiment analysis using machine learning techniques, including the identification and demarcation of negation in online reviews. We use the Amazon reviews dataset to validate our end-to-end sentiment analysis approach [35]. Amazon, with its diverse product line, is one of the most comprehensive repositories of online reviews. This makes Amazon product reviews a favored choice among the research community for developing natural language processing techniques.

The remaining sections of this paper are organized as follows. In section 2, we highlight the advances made by previous researchers in handling negation in sentences. In section 3, we propose our methodology for preparing an end-to-end sentiment analysis pipeline. Section 4 presents and interprets the performance of various classifiers on the Amazon online reviews with and without the negation preprocessing. We conclude our work in section 5 and present future research direction.

## 2. Related Work

Sentiment analysis is one of the fastest-growing research areas in Computer Science, with around 7000+ published papers on the topic [7]. Methods of sentiment analysis can be categorized broadly as Machine Learning, Lexicon, and hybrid-based [8]. Previous research approaches focus on using various Supervised Machine Learning algorithms such as Support Vector Machines (SVM), Naïve Bayes Classifier, Random Forests, and Gradient Boosting Machines, unsupervised approaches through lexicons such as SentiWordNet [9], AFINN [10], VADER [11], sentiment lexicons [12-14] or hybrid approaches that combine Supervised and Unsupervised learning into Semi-supervised Learning [15].

Machine Learning is the predominant approach to Sentiment Analysis, especially in the cases where labeled data is already available. D'Souza and Sonawane [16] developed a method of training a Machine Learning model both actual reviews and the reviews reversed through a custom algorithm referred to as the Dual Sentiment Analysis (DSA). Amrani et al. [17] proposed a hybrid Machine Learning architecture, which is a combination of Random Forest and Support Vector Machine (SVM) architectures to address the problem of margin with hyperplanes using SVM and the sensitivity of Random Forests to the predictors. Shaheen et al. [18] compared the performance of multiple Machine Learning algorithms such as SVM, Naïve Bayes, Random Forests, Long Short Term Memory

(LSTM) networks, and Convolutional Neural Networks (CNN). Ziegelmayer and Schrader [19] evaluated the Sentiment Polarity Prediction using Partial Matching (PPM) and data compression models using PPM-like character n-gram frequencies. Mandhula et al.[20] proposed a hybrid topic modeling approach using the Latent Dirichlet Algorithm (LDA) and Fuzzy C-means to extract relevant keywords from Amazon reviews. They used a selective architecture based on Convolutional Neural Networks (CNN) to perform sentiment classification on these keywords. Chen et al. [21] explored the performance of sentiment analysis using Bi-Directional LSTM with conditional random fields (BiLSTM-CRF) for target sentence extraction and sentence type classification and a 1-D CNN to perform the sentiment classification on each type of sentence. Yanagimto et al. [22] proposed a new CNN architecture known as Gated Convolutional Neural Network (GCNN) and a self-attention mechanism to help understand the correspondence between the weights of a Neural Network and the words in raw reviews. This process develops a better understanding of complex Neural Networks models when used in Natural Language Processing applications.

One of the main reasons behind errors in sentence-level Sentiment Analysis is the inability to accurately determine the effect of negation on other words [23] Negation identification improves the sentiment classification accuracy by increasing the performance across the parts of sentences that would have different polarities when combined with the negation term. Related works attempt to address this problem through a combination of a rule-based approach and Machine Learning techniques. Sharif et al.[24] proposed a custom algorithm to calculate the polarity of a review considering the negations. They proposed a pipeline consisting of syntactic parser and polarity score calculation at sentence level through the use of the dependency table and average the polarity across multiple sentences of a review. Asmi and Ishaya [25] proposed the use of syntactic parser, polarity calculator (using SentiWordNet), the rules-based approach using Bag of Words (BoW) and dependency trees to identify and resolve the scope of negation in the text. Pandey et al. [26] proposed the technique of revaluating sentiment polarity classification after initial sentiment classification using a rule-based approach and utilizing the dependency parse tree for polarity calculation. Cruz et al. [27] modeled negation identification and scope definition as two consequent classification models using cues available in training documents. Chapman et al. [28] developed a regular expression based algorithm (NEGEX) that accurately identified the relevant negations from the narratives of medical records. Mukherjee et al. [29] proposed the negation parser, NegAIT, to identify morphological, syntactical, and double negations from medical corpora. They classify “easy” and “difficult” texts with very high accuracy and precision using different types of negations as the predictors.

Amazon reviews are convenient datasets for Sentiment Analysis. Prakoso et al. [30] evaluated the lexicon-based approach on sentiment analysis by various machine learning algorithms on Amazon Web reviews obtained from UCI Machine Learning. Fang and Zhan [31] performed a Sentiment Analysis on Amazon product reviews using both sentence-level and review-level categorization. Ejaz et al. [32] experimented with Decision Trees and Random Forest models along with n-grams count and a lexicon dictionary-based approach. Katic and Milicevic [33] compared the performance of different machine learning classifiers and vector representations of text for Amazon reviews. Haque et al. [34] used active learning to obtain labels for an unlabelled dataset and compared different supervised machine learning classifiers on large scale Amazon reviews.

Roughly speaking, these research contributions have focussed on various aspects of the Sentiment Analysis pipeline but never combine the negation types while analyzing the sentiments from the reviews of the Amazon users. This paper delineates and outlines an end-to-end machine learning framework combining many approaches from previous research, involving preprocessing of text, handling and marking of negation, feature extraction, and vector representation and building and tuning of deep learning models for sentiment polarity classification, demonstrating the working of this pipeline on Amazon reviews data.

## 3. Methodology

As depicted in Figure 1, our methodology relies on an end-to-end approach to handle negations in sentences, including multiple steps such as Data Collection and Preparations, Text preprocessing, Negation identification and marking, TF-IDF feature extraction, training, and evaluating our models, all of which are described below in Figure 1.

### 3.1. Data Collection and Cleaning

In order to validate our approach, we choose the Amazon reviews dataset obtained from Kaggle [35] and which contains information on different cellphones such as brand, image URL, number of reviews, average rating, etc. The dataset also contains information on each review, such as review title, review body, and verified user, and totaling up to 82,000 reviews of 792 different cellphones produced by 10 major brands. The attributes which are expected not to play a significant role in determining the sentiment, such as URL, image, review url, and name, were removed. Also, verified users were only considered in order to make the data more reliable, which left around 75,000 reviews for analysis and modeling. Table 1 represents the review corpus statistics where ASUS garners the least number of reviews. But it is to be noted that ASUS reviews have the highest values for all the features shown in Table 1. As there are close to 400 cellphone products of Samsung, the number of reviews posted by users for Samsung products are also highest among the brands followed by Apple.

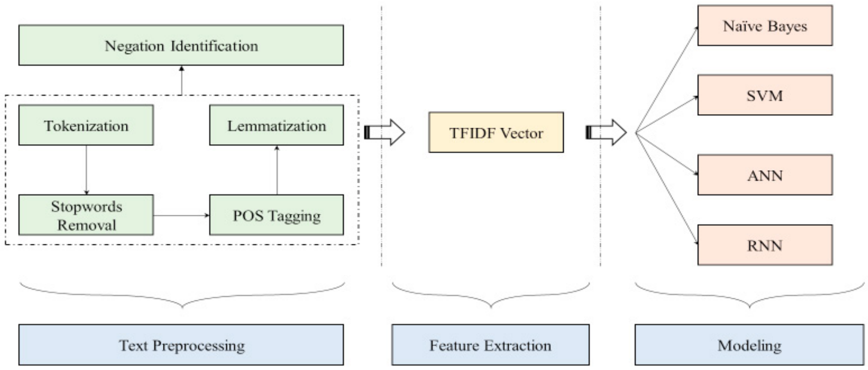


Fig. 1: Overview of Sentiment Analysis Approach enabling Negation Identification

Table 1. Review Corpus Statistics per Brand

| Brands | number_of_reviews | avg_word_count | avg_token_count | avg_sentence_count | avg_words_in_sentence |
|---|---|---|---|---|---|
| ASUS | 410 | 80.51 | 38.26 | 6.53 | 11.35 |
| Apple | 11534 | 28.64 | 13.67 | 3.04 | 9.24 |
| Google | 3413 | 52.47 | 24.9 | 4.78 | 10.1 |
| HUAWEI | 2757 | 52.52 | 26.05 | 4.81 | 9.47 |
| Motorola | 7606 | 51.15 | 24.28 | 4.85 | 9.24 |
| Nokia | 5010 | 58.88 | 27.75 | 5.35 | 9.25 |
| OnePlus | 436 | 48.16 | 22.91 | 4.42 | 9.39 |
| Samsung | 38694 | 37.23 | 17.37 | 3.83 | 8.92 |
| Sony | 2870 | 72.81 | 34.67 | 6.21 | 10.33 |
| Xiaomi | 2818 | 47.83 | 24.4 | 4.33 | 9.98 |

Now, each review is passed through the text preprocessing steps, as described in the next section

### 3.2. Text Pre-processing

Each review is passed through the steps of tokenization, stopwords removal, part of speech tagging, and lemmatization.

**Tokenization:** Tokenization is the method of separating a sequence set into individual entities, including words, keywords, phrases, symbols, and other items known as tokens. The tokens act as the input for various processes, such as parsing and text mining. Tokenization is applied to each review into sentences, and we have tokenized the sentence into words.

**Stopwords Removal**: Stop words are usually those items in a sentence that do not add information to the text and are used as fillers to complete sentences. In sentiment analysis, these words do not contribute to any particular sentiment.

**Part-of-speech Tagging**: Part of Speech tagging is the preliminary method of recognizing the functional role of a token within a sentence and is the critical step in any NLP pipeline. In general, parts of the speech include Nouns, Pronouns, Verbs, Adverbs, Adjectives, Conjunctions, and subcategories. For our case-study, the Penn Tree tagger was used for Part-of-Speech tagging.

**Lemmatization:** Lemmatization is the process of extracting the root form of a word from its various forms. For instance, if the tokens contain words such as playing, played, and plays, the lemmatization step will convert these words into their root form, i.e., play. Since a word represents a particular sentiment in any form, lemmatization is necessary to homogenize the representation of the reviews using different forms of the same word. We use WordNet lemmatizer, which requires the POS tag that was calculated in the previous step.

Table 2. Different types of negations

| | | |
|---|---|---|
| Morphological Negation | Identification | Through the usage of negation prefixes such as -ab,-dis,-il,-im,-in,-ir,-un. |
| | Scope of Negation | The scope of negation is limited to the word that is being negated. |
| Syntactical Negation | Identification | Through the use of full negative words such as no, not, never, no one, neither, none, etc. |
| | Scope of Negation | The scope of syntactical negation is limited by the next punctuation mark or is extended or delimited by a compound Verb. |
| Double Negation | Identification | Through the identification of morphological negation inside the scope of a syntactical negation |
| | Scope of Negation | The scope of double negation is limited to the word with the morphological negation inside a syntactical negation scope. No negation happens on this word. |
| Implicit Negation | Identification | Difficult to identify explicitly in a sentence. The identification is based on the position that each word occupies in the negation spectrum. |
| | Scope of Negation | The scope of negation applies to the specific word only, and the amount of negation also depends on the word by itself. |

Finally, we move onto the crucial step of negation identification and marking for scope detection. Table 2 exhibits different types of negations.

### *3.3. Negation Identification*

For the scope of our case-study, we only consider morphological and syntactical negations in product reviews. Morphological negations are stand-alone negations on words that are a result of appending prefixes such as –ab, -dis, -un to words. For example, in the sentence 'The audience disliked the actor leading the play,' the prefix dis is added the word liked to negate the meaning that the audience did not prefer the lead actor in the play. The scope of morphological negation is limited to the word it is negating.

Syntactical negations are intended to negate the meaning of a phrase inside a sentence. For example, in the sentence 'He was not supposed to come,' the explicit negation word not negates the meaning derived from the phrase: supposed to come. Syntactical are often identified through the use of explicit negation words such as not, never, neither, etc., and the end of the phrase defines their scope through punctuation.

Double negations are the effect of two negations canceling the effect of each other's contradiction. The most common type of double negation is through the combination of syntactical and morphological negations in the same sentence. For example, 'The price of the car is not insignificant,' not is the syntactical negation and insignificant is the morphological negation with the effect of canceling each other out. Though there are sentences where two syntactic negations are also present to signify double negation such as 'You are not the person I do not like'. For the purposes of our research, we are considering the former case of double negations.

Implicit negations are the negations in sentences are not as assertive or explicit as the negations mentioned above. Implicit negations bring a measure of uncertainty to the sentence phrase and in the process, bring about the negation of the phrase. For example, in the phrase, 'I doubt there is anything left', the word 'doubt' introduces the negation of the following words, but, the negation is not assertive and explicit. It just creates a measure of uncertainty and seeks to pull down the assertiveness of the sentence. Implicit negations are difficult to explicitly identify in a sentence and the amount and scope of negation depend on the word itself. In our research, we are not considering implicit negations.

Using the definitions of different negation types and the process to identify the negation and their scope in Table 2, we designed a pre-processing algorithm that aimed to identify words and phrases with explicit negation and replace the words with negation equivalent for their lemmatized form. This negation equivalent word has the '_NEG' suffix attached to the lemmatized form to help the model differentiate between a positive word and its negation use case.

A gold standard dataset developed for NegAIT [29] a parser for medical text simplification, was used for identification and root word extraction for morphological negations. More precisely, the corpus was used to identify the morphological negations while the syntactic negations were identified through the use of full negative words such as No, Not, None, etc. in a sentence or phrase. The scope of negation for morphological negations was limited to the word while that for syntactical negation was limited to the next punctuation. Stop words were ignored for negation as they did not add meaning to the negation in text and were removed after negation identification along with punctuation. Double negations are also handled in the algorithm through the use of the '_NEG' suffix.

### 3.4. Feature Extraction

In the post data cleaning phase, we apply the feature extraction methodology to convert the textual data to numerical data. To this end, we use TF-IDF (Term Frequency Inverse Document Frequency) based on the review corpus. The TF-IDF is defined by equation 1.

$$\textit{TF-IDF(word, doc)} = \textit{TF(word, doc)} * \textit{IDF(word)} \qquad (1)$$

Thus, two matrices must be determined in this process, one comprising the inverse document frequency of a word in the entire corpus and another containing the term frequency of each word in every document. The formulas for calculating these two terms are represented by equation 2 and 3 respectively:

$$\textit{TF(word, doc)} = \textit{Frequency of word} \in \textit{the document / Number of words} \in \textit{the document} \qquad (2)$$

$$\textit{IDF(word)} = \log(1 + \textit{Number of documents / Number of documents with word}) \qquad (3)$$

### 3.5. Classifiers

We developed and trained four separate machine learning models to investigate which Machine Learning model performs best on classifying cellphone reviews sentiments.

The Multinomial Naive Bayes (MNB) classifier is a family of simple probabilistic classifiers based on a standard assertion that, given the category variable, all features are independent of each other [36]. The MNB is used due to its simplicity in the training and classification phase.

In this paper, we have emphasized on classifying reviewer sentiments using an SVM classifier, a popular powerful tool for classification of vectors of real-valued features [37]. SVMs are a classification technique in Machine Learning that uses a method called a kernel to transform a data point space where the data is not linearly separable into a new space where it is, with allowances for erroneous classification [38].

We have used primarily two different neural network architecture for determining the sentiments of the reviews. An artificial neural network (ANN) is a mathematical model considering the architecture and components of the biological neural networks. They are regarded as the nonlinear statistical data modeling tools in which some patterns are recognized or identified using the intricate relationship between outputs and inputs.

Recurrent Neural Network (RNN) belongs to the class of Neural Networks whose relationships among neurons form a directed cycle. RNNs are useful for processing sequential information due to the use of its internal memory. RNN performs the same task for every sequence of input data, with each output relying on all previous calculations, which is like remembering the so far processed information [39].

### *3.6. SentiWordNet*

SentiWordNet is a publicly available lexical resource [9] that is hosted inside WordNet and provides three scores of positivity, negativity, and objectivity for words. SentiWordNet lexicon was used to aggregate the net polarity (i.e., the difference between positivity and negativity) for each word in a review. Finally, the net polarity score was used to determine the final polarity of the review.

In addition, the performance without the custom negation identification algorithm was also obtained to evaluate the effect of the negation identification algorithm on sentiment polarity prediction. To demonstrate the performance transferability of the approach, the classification approaches defined in this section were applied along with the negation algorithm to a new review dataset obtained from the prior research [34].

### *3.7. Implementation and Evaluation criterion*

With the minority negative sentiment class forming 25% of the dataset, there was a need for up-sampling the dataset to balance the class distribution. Up-sampling tends to cause some degree of overfitting to the minority class. However in this study, we managed to avoid a high degree of overfitting to negative sentiment class by training on the higher number of distinct positive sentiment reviews available in the dataset. Hence, the imbalanced dataset is up-sampled for reliable performance and is split between train and test set with a test split ratio of 0.2. SVM and Multinomial Naïve Bayes classifiers are trained on the training dataset, and the model performance is validated on the test set.

Similarly, we have trained the ANN and RNN model by taking TF-IDF as the feature matrix and the binary output of sentiments, i.e., positive and negative. We have used the feed-forward network for ANN and LSTM (Long-short term memory) architecture for RNN. We trained the LSTM network with 5 hidden layers 400, 200, 100, 50, and 10 units, respectively, taking ReLU as the activation function. Similarly, for the ANN model, the architecture contains 5 hidden layers with 1000, 500, 100, 60, and 20 units, respectively, using ReLU as the activation function. For both the models, the output layer contains 1 neuron with sigmoid activation function, a mini-batch size of 512, and the Adam optimizer (with a learning rate of 0.01) is used for the fast convergence of gradients. The optimal number of epochs is decided through an epoch plot of the training and validation loss. This resulted in the optimal number of epochs for the ANN model being 3 and 5 for the RNN model.

With this problem being a binary classification problem, we use accuracy, precision, recall, F-1 score, and area under the ROC curve as performance metrics to compare the performance of the different classifiers.

## 4. Results and Discussion

### *4.1. Performance Comparison between the machine learning models*

Table 3 compares the performance of different classifiers, along with the effect of the negation algorithm on the sentiment analysis. It can be seen that the combination of ANN classifier along with the negation identification is the best performing classifier at 95.67% accuracy, marginally ahead of RNN with negation combination at 95.30%. The performance of the classifier improves when the negation identification and marking happen beforehand. The difference is not much for ANN and RNN but is more pronounced for SVM and Naïve Bayes. Deep learning algorithms outperform traditional machine learning algorithms, in both cases:- a) with explicit negation marking and b) without explicit negation marking. All the machine learning approaches (with or without negation marking) significantly outperform the lexicon-based approach of SentiWordNet. These results show the superiority of the machine learning approach over lexicon-based approaches, especially when the labeled dataset is available. A limitation of the lexicon-based approach is the corpus over which is trained, making it difficult to generalize for datasets which differ from the corpus on which it was trained.

Table 3. Performance Comparison of Classifiers

| | Positive Class | | | Negative Class | | | | |
|---|---|---|---|---|---|---|---|---|
| Classifier | Precision | Recall | F-1 Score | Precision | Recall | F-1 Score | Accuracy | Area Under ROC Curve |
| RNN + Negation | 96.43% | 94.04% | 95.22% | 94.23% | 96.55% | 95.38% | 95.30% | 0.953 |
| ANN + Negation | 96.84% | 94.38% | 95.60% | 94.56% | 96.95% | 95.74% | 95.67% | 0.957 |
| SVM + Negation | 93.37% | 91.51% | 92.43% | 91.74% | 93.55% | 92.64% | 92.54% | 0.925 |
| Naïve Bayes + Negation | 90.92% | 89.40% | 90.15% | 89.66% | 91.15% | 90.40% | 90.28% | 0.903 |
| SentiWordNet | 60.23% | 85.59% | 70.70% | 75.46% | 43.94% | 55.54% | 64.68% | 0.648 |
| RNN | 95.19% | 93.79% | 94.49% | 93.93% | 95.30% | 94.61% | 94.55% | 0.945 |
| ANN | 94.81% | 93.22% | 94.01% | 93.39% | 94.94% | 94.15% | 94.08% | 0.941 |
| SVM | 91.86% | 88.45% | 90.13% | 88.95% | 92.23% | 90.56% | 90.35% | 0.903 |
| Naïve Bayes | 89.01% | 88.73% | 88.87% | 88.86% | 89.13% | 89.00% | 88.93% | 0.889 |

### *4.2. Performance Comparison on alternative reviews dataset*

Table 4 compares the performance of the classifiers and the negation algorithm on a different dataset [34]. This dataset contained 173,000 Amazon reviews on cellphones and was processed using the same preprocessing approach described in this paper. The imbalance in the dataset was also handled through upsampling to create a balanced dataset. Figure 2 delineates the accuracy of the proposed ANN and RNN models. The ANN model with negation gives out the highest accuracy of 96.32%, which is slightly more significant than the accuracy of RNN, i.e., 95.97% as shown in Figure 2.

Deep learning algorithms generalize well over the new dataset, and their performance remains the same while the performance of the traditional machine learning algorithms SVM and Naïve Bayes degrades. This means that the deep learning models are more stable in comparison with the traditional machine learning models in this case.

Table 4. Performance of Classifiers on New Dataset

| | Positive Class | | | Negative Class | | | | |
|---|---|---|---|---|---|---|---|---|
| Classifier | Precision | Recall | F-1 Score | Precision | Recall | F-1 Score | Accuracy | Area Under ROC Curve |
| RNN + Negation | 99.12% | 92.36% | 95.62% | 92.95% | 99.19% | 95.97% | 95.80% | 0.958 |
| ANN + Negation | 98.50% | 94.02% | 96.21% | 94.36% | 98.59% | 96.43% | 96.32% | 0.963 |
| SVM + Negation | 85.27% | 85.64% | 85.46% | 85.80% | 85.43% | 85.61% | 85.54% | 0.855 |
| Naïve Bayes + Negation | 80.78% | 83.99% | 82.36% | 83.60% | 80.32% | 81.93% | 82.15% | 0.822 |

## 5. Implications

We found the results interesting because the approach helps companies build sentiment models quickly using their own reviews data specific to a product or service in their set of products without losing out on performance. Irrespective of online or brick and mortar retailing, the consumer plays an essential role in the WoM branding of the product. The analysis of sentiments will decide the future of the brand. The online product reviews eventually enter social networking via WoM branding, which has a more significant magnitude on sales and ROI of the product. The customers' opinions and sentiments extracted from the product reviews can make or break the brands. methodologies that provide the exactness of perfections in measuring sentiments will help the producers/companies to offer the right response and solutions. This will enable companies to correct misunderstandings, review marketing strategies, fix problems, launching marketing campaigns, resolve issues, provide more features, and thus enhance consumers' engagement.

## 6. Conclusion

In this paper, we have proposed an end-to-end approach to sentiment analysis, which should help future developments in the field to modularize the process and develop sentiment analysis engines at scale for companies. In addition, we have described a custom negation marking algorithm to help classifiers identify negation in sentences. For this purpose, we have conducted experiments on two different datasets based on Amazon Product Reviews to determine the sentiments using four well-known classification models, such as SVM, Multinomial Naïve Bayes, ANN, and RNN. After pre-processing and feature extraction on the dataset, all four models utilized the generated TF-IDF matrix and computed the sentiments. Without negation marking, most of the explicit negations are lost during the pre-processing phase and implicated a loss of information which could be resolved by our approach. So, from our experiments, we have concluded that when traditional classifiers for text classification coupled with the negation identification, then the performance of the sentiment classifier increased.

We will consider implicit negations in our study of sentiment polarity detections in future. Additionally, future approaches will consider emotion identification in the text to quantify the market sentiment of the product.

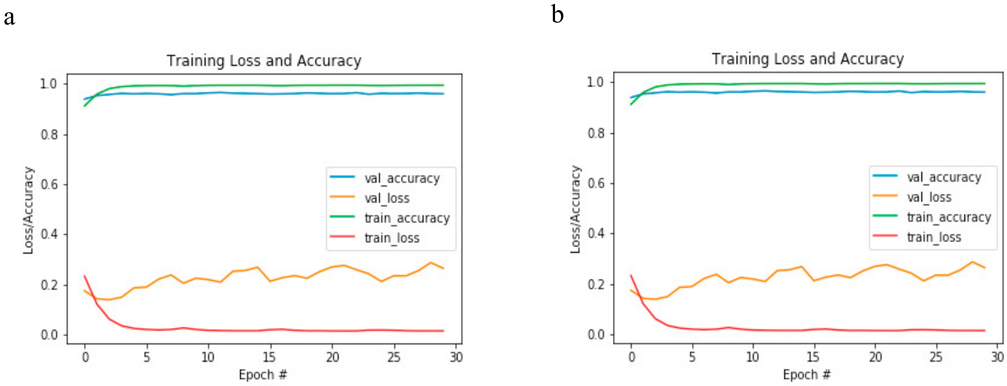


Fig. 2: Training & Validation Epoch plots to determine the number of optimal epochs: a) ANN, b) RNN

## References

[1] Liu, Bing. (2012) "Sentiment analysis and opinion mining." *Synthesis lectures on human language technologies,* **5(1)**: 1-167.

[2] Cambria, Erik. (2016) "Affective Computing and Sentiment Analysis." *IEEE Intelligent Systems,* **31(2)**: 102-107.

[3] Amalia, Rizkiana, Moch Arif Bijaksana, and Dhinta Darmantoro. (2018) "Negation handling in sentiment classification using rule-based adapted from Indonesian language syntactic for Indonesian text in Twitter." *Journal of Physics: Conference Series,* **971(1):** 012039.

[4] Peng, Yifan, Xiaosong Wang, Le Lu, Mohammadhadi Bagheri, Ronald Summers, and Zhiyong Lu. (2018) "Negbio: a high-performance tool for negation and uncertainty detection in radiology reports." *AMIA Summits on Translational Science Proceedings*: 188.

[5] Saini, Akriti, Nishank Bhatia, Bhavya Suri, and Shikha Jain. (2014) "EmoXract: Domain independent emotion mining model for unstructured data." *In Seventh International Conference on Contemporary Computing (IC3)*, 94-98, IEEE.

[6] Sohn, Sunghwan, Stephen Wu, and Christopher G. Chute. (2012) "Dependency parser-based negation detection in clinical narratives." *AMIA Summits on Translational Science Proceedings*: 1.

[7] Mäntylä, Mika V., Daniel Graziotin, and Miikka Kuutila. (2018) "The evolution of sentiment analysis—A review of research topics, venues, and top cited papers." *Computer Science Review,* **27**: 16-32.

[8] Kabir, Kabir, Mir Md. Jahangir Kabir, Shuxiang Xu, and Bodrunnessa Badhon. (2019) "An empirical research on sentiment analysis using machine learning approaches." *International Journal of Computers and Applications:* 1-9.

[9] Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani. (2010) "Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining." *Lrec*.

[10] Nielsen, Finn Årup. (2011) "A new ANEW: evaluation of a word list for sentiment analysis in microblogs." *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts: Big things come in small packages,* **718**: 93-98.

[11] Hutto, Clayton J., and Eric Gilbert. (2014) "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text." *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*, 216-225. AAAI.

[12] Kiritchenko, Svetlana, and Saif M. Mohammad. (2018) "Sentiment composition of words with opposing polarities." *Proceedings of NAACL-HLT*, 1102-1108, ACL.

[13] Kiritchenko, Svetlana, Saif Mohammad, and Mohammad Salameh. (2016) "Determining sentiment intensity of English and Arabic phrases." *Proceedings of the 10th international workshop on semantic evaluation (SEMEVAL-2016)*, 42-51.

[14] Agarwal, Apoorv, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca J Passonneau. (2011) "Sentiment analysis of twitter data." *Proceedings of the workshop on language in social media (LSM 2011)*, 30-38.

[15] Khan, Farhan Hassan, Usman Qamar, and Saba Bashir. (2017) "A semi-supervised approach to sentiment analysis using revised sentiment strength based on SentiWordNet." *Knowledge and information Systems,* **51(3)**: 851-872.

[16] D'souza, SStephina Rodney, and Kavita Sonawane. (2019) "Sentiment Analysis Based on Multiple Reviews by using Machine learning approaches." *3rd International Conference on Computing Methodologies and Communication (ICCMC)*, 188-193, IEEE.

[17] Al Amrani, Yassine, Mohamed Lazaar, and Kamal Eddine El Kadiri. (2018) "Random forest and support vector machine based hybrid approach to sentiment analysis." *Procedia Computer Science,* **127**: 511-520.

[18] Shaheen, Momina, Shahid M. Awan, Nisar Hussain, and Zaheer A. Gondal. (2019) "Sentiment analysis on mobile phone reviews using supervised learning techniques." *International Journal of Modern Education and Computer Science* **11(7)**: 32-43.

[19] Ziegelmayer, Dominique, and Rainer Schrader. (2012) "Sentiment polarity classification using statistical data compression models." *12th International Conference on Data Mining Workshop*s, 731-738, IEEE.

[20] Mandhula, Trupthi, Suresh Pabboju, and Narsimha Gugulotu. (2020) "Predicting the customer's opinion on amazon products using selective memory architecture-based convolutional neural network." *The Journal of Supercomputing,* **76**: 5923-5947.

[21] Chen, Tao, Ruifeng Xu, Yulan He, and Xuan Wang. (2017) "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN." *Expert Systems with Applications,* **72**: 221-230.

[22] Yanagimto, Hidekazu, Kiyota Hashimoto, and Makoto Okada. (2018) "Attention visualization of gated convolutional neural networks with self-attention in sentiment analysis." *International Conference on Machine Learning and Data Engineering (iCMLDE)*, 77-82, IEEE.

[23] Farooq, Umar, Hasan Mansoor, Antoine Nongaillard, Yacine Ouzrout, and Muhammad Abdul Qadir. (2017) "Negation Handling in Sentiment Analysis at Sentence Level." *Journal of Computers,* **12(5)**: 470-478.

[24] Sharif, Wareesa, Noor Azah Samsudin, Mustafa Mat Deris, and Rashid Naseem. (2016) "Effect of negation in sentiment analysis." *Sixth International Conference on Innovative Computing Technology (INTECH*), 718-723, IEEE.

[25] Asmi, Amna, and Tanko Ishaya. (2012) "Negation identification and calculation in sentiment analysis." *Second International Conference on Advances in Information Mining and Management*, 1-7.

[26] Pandey, Swastika, Santwana Sagnika, and Bhabani Shankar Prasad Mishra (2018). "A technique to handle negation in sentiment analysis on movie reviews." *International Conference on Communication and Signal Processing (ICCSP)*, 737-743, IEEE.

[27] Cruz, Noa P., Maite Taboada, and Ruslan Mitkov. (2016) "A machine-learning approach to negation and speculation detection for sentiment analysis." *Journal of the Association for Information Science and Technology,* **67(9)**: 2118-2136.

[28] Chapman, Wendy, Will Bridewel, Paul Hanbury, Gregory F. Cooper, and Bruce G. Buchanan. (2001) "A simple algorithm for identifying negated findings and diseases in discharge summaries." *Journal of biomedical informatics,* **34(5)**: 301–310.

[29] Mukherjee, Partha, Gondy Leroy, David Kauchak, Srinidhi Rajanarayanan, Damian Y. Romero Diaz, Nicole P. Yuan, T. Gail Pritchard, and Sonia Colina. (2017) "NegAIT: A new parser for medical text simplification using morphological, sentential and double negation." *Journal of biomedical informatics,* **69**: 55-62.

[30] Prakoso, Ardianda Aryo, Winantesa Yananta, and Arif Fitra Setyawan. (2018) "A Lexicon-Based Sentiment Analysis for Amazon Web Review." *International Seminar on Application for Technology of Information and Communication*, 503-508, IEEE.

[31] Fang, Xing, and Justin Zhan. (2015) "Sentiment analysis using product review data." *Journal of Big Data,* **2**: 5.1-5.14.

[32] Ejaz, Afshan, Zakia Turabee, Maria Rahim, and Shakeel Khoja. (2017) "Opinion mining approaches on Amazon product reviews: A comparative study." *International Conference on Information and Communication Technologies (ICICT)*, 173-179, IEEE.

[33] Katić, Tamara, and Nemanja Milićević (2018). "Comparing sentiment analysis and document representation methods of Amazon reviews." *16th International Symposium on Intelligent Systems and Informatics (SISY)*, 283-286, IEEE.

[34] Haque, Tanjim Ul, Nudrat Nawal Saber, and Faisal Muhammad Shah. (2018) "Sentiment analysis on large scale Amazon product reviews." *International Conference on Innovative Research and Development (ICIRD)*, 1-6, IEEE.

[35] Amazon Cell Phone Reviews, Kaggle, (2019), Available at: https://www.kaggle.com/grikomsn/amazon-cell-phones-reviews/version/1.

[36] Xu, Shuo. (2018) Bayesian Naïve Bayes classifiers to text classification. *Journal of Information Science*, **44(1)**: 48–59.

[37] Vladimir. (1998). *Statistical Learning Theory*. Wiley, Chichester, GB.

[38] Mullen, Tony, and Nigel Collier. (2004) Sentiment analysis using support vector machines with diverse information sources. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 412-418, ACL.

[39] Zhang, Lei., Shuai Wang, and Bing Liu. (2018) "Deep learning for sentiment analysis: A survey." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **8(4)**:1-25.