❐   4554

A Novel Hybrid Classification Approach for Sentiment Analysis of Text Document

Yassine Al Amrani[1], Mohamed Lazaar[2], Kamal Eddine El Kadiri[3]
[1,3]LIROSA Laboratory, Abdelmalek Essaadi University, Morocco
[2]New Technology Trends Team, Abdelmalek Essaadi University, Morocco

Article Info	ABSTRACT
	Sentiment analysis is a more popular area of highly active research in Automatic Language Processing. She assigns a negative or positive polarity to one or more entities using different natural language processing tools and also predicted high and low performance of various sentiment classifiers. Our approach focuses on the analysis of feelings resulting from reviews of products using original text search techniques. These reviews can be classified as having a positive or negative feeling based on certain aspects in relation to a query based on terms. In this paper, we chose to use two automatic learning methods for classification: Support Vector Machines (SVM) and Random Forest, and we introduce a novel hybrid approach to identify product reviews offered by Amazon. This is useful for consumers who want to research the sentiment of products before purchase, or companies that want to monitor the public sentiment of their brands. The results summarize that the proposed method outperforms these individual classifiers in this amazon dataset.

Corresponding Author:

Yassine Al Amrani,
LIROSA Laboratory,
Abdelmalek Esaadi University,
Tetuan, Morocco.
Email: alamraniyassine@gmail.com

1.   INTRODUCTION

Classification is the process wherein a class label is assigned to unlabeled data vectors. It can be categorized into supervised and un-supervised classification which is also known as clustering. In supervised classification learning is done with the help of supervisor i.e. learning through example. In this method, the set of possible class labels is known a priori to the end user [1]. Supervised classification can be subdivided into non-parametric and parametric classification. Parametric classifier method is dependent on the probability distribution of each class. Non-parametric classifiers are used when the density function is unknown. Examples of parametric supervised classification methods are Minimal Distance Classifier, Bayesian, Multivariate Gaussian, Support Vector machines and Decision Tree. Examples of non-parametric supervised classification methods are K- Nearest Neighbors, Euclidean Distance, Logistic Regression, Neural Network Kernel Density Estimation, Artificial Neural Network and Multilayer Perceptron.

Recently, multiple platforms are developing very interesting either in terms of volume of data or according to the number of users around the world, they offer users all the possibilities to express their opinions and to exchange their ideas with the others [2]. The sentiment analysis found in the form of comments, reviews and feedback and provides necessary information for various purposes [3]. These opinions or sentiments can be divided into two categories: positive and negative; or also categories of different rating points (e.g. 3 stars, 4 stars and 5 stars). The polarity of sentiments like "good" and "bad" also identify the sentiments either positive or negative [4]. Sentiment analysis is the part of the text mining that

attempts to define the opinions, feelings and attitudes present in a text or a set of text. It is particularly used in marketing to analyse for example the comments of the Net surfers or the comparatives and tests of the bloggers. It requires much more understanding of the language than text analysis and subject classification. Indeed, if the simplest algorithms consider only the statistics of frequency of occurrence of the words, it is usually insufficient to define the dominant opinion in a document. It is the process of determining the contextual polarity of the text, that is, whether a text is positive or negative [5].

The use of this analysis helps researchers and decision-makers better understand opinions and client satisfaction using sentiment classification techniques in order to automatically collect different perspectives on from various platforms. There has been a large amount of research in the area of sentiment classification. Traditionally most of it has focused on classifying larger pieces of text, like reviews (B. Pang, L. Lee, and S. Vaithyanathan., 2002). In this paper, a comparison of popular classifiers was performed to classify product reviews either positive or negative: Support Vector Machine, Random Forest and our approach Random Forest Support Vector Machine (RFSVM).

This paper presents a method to determine how sentiments can be classified using hybrid approach of Support Vector Machine and Random Forest. The paper provides the comparison with other existing technique, shows that the use of hybrid approach can improve the efficiency of sentiment analysis. The proposed hybrid approach gives better result as compare to the existing techniques. The rest of the paper is described as follows: Section 2 describe sentiment analysis system. Section 3 introduces applied algorithms in this field. Section 4 discusses proposed methods. Section 5 explain the results and analysis obtained. Section 6 presents the conclusion and future work for the proposed work.

2.   SENTIMENT ANALYSIS SYSTEM

To know the opinion of the other people was always an important information element during the decision process. Before making decisions, people are interested enormously in the opinions of the other people in different areas. They consult the opinions of the other consumers before making a purchase, or look at the opinions of the other people before seeing a film with the cinema or before buying a disc. Thanks to the internet we can discover the opinions and the experiments of very a large number of people who are neither our friends, nor the experts of fields, but of people who can have the same tastes that us, and thus their opinions can be very useful for us before making our choice and to have our own idea on a given subject. Today, more and more people are giving their opinion on different topics, these opinions are available to everyone on the internet.

According to the surveys [6], 81% of the users of the internet made at least once the online search on a product and approximately 80% of them declare that other people have a significant influence on their decision of purchase, which represents one very a large number of people. Approximately 30% provided an opinion on a product, on a service or on a person online via a marking system, which is not unimportant like number. For this reason, i.e. thanks to the interest which the users show for the opinions on the products and the services, the suppliers of the articles show very a great attention with the development of the marking systems [Hoffman (2008)]. With the explosion of platforms like the blogs, of the discussion forums, Peer-to-Peer network, and various other types of social media, the consumers have at their disposal a platform without precedent, of range and power, making it possible to share their experiments and to mark their opinion (positive or negative) on any product or service. The companies can meet the needs for the consumers by carrying out monitoring and analysis of the opinions to improve their product. Such a system will have firstly to collect opinions of the consumers and users in documents which show the subjective opinions and sentences. Sometimes, that is relatively easy, as in the cases of great sites where the opinions of the users are well structured such as for example Amazon.com.

Sentiment is a vision based on emotion rather than reason. It is a kind of subjective impression, not facts, also called the expression of sensitive feeling in art and literature. Sentiment Analysis is also a task of natural language processing and information extraction that aims to get the feelings of the writer expressed by positive or negative comments, questions and requests, by analyzing a great number of documents. Sentiment analysis is the computational technique for extracting, classifying, understanding and determining opinions expressed in various content. It focuses on identifying the opinion or sentiment that is held about an object. It uses natural language processing and computational techniques to automate the extraction or classification of sentiment from generally unstructured text [7].

In general, sentiment analysis aims to determine the state of mind of a speaker or a writer with respect to a subject or the overall tone of a document. Word of mouth is the process of passing information from person to another and plays an important role in clients' decision making about services or products. In business situations, Word of mouth involves consumers who share attitudes, opinions, products, or services with others. Word of mouth communication functions based on social networking [8].

In recent years, the massive increase in the use of internet and the exchange of public opinion are the engines of sentiment analysis today. The Web is an immense repository of structured and unstructured data. Analyzing this data to extract latent public opinion and sentiment is a difficult task. Sentiment analysis can be useful in online product reviews, recommendations, blogs, user's views of political candidates.

3.   APPLIED ALGORITHMS

To evaluate the performance of our approach, we chose to use two supervised learning algorithms: the random forest algorithm which is a classification algorithm that reduces the variance of the forecasts of a decision tree alone, thus improving their performances, and the Algorithm of Support Vectors Machines or Large Marginal Separators which is a binary classification method by supervised learning. These have been chosen because they are the machine learning algorithms that often give the best results for automatic classification of texts. Control flow of the system as shown in Figure 1.
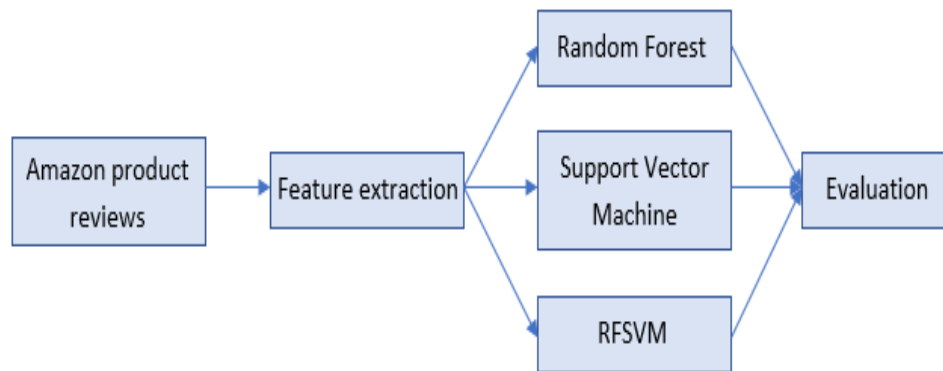
Figure 1. Control flow of the system

3.1. Random Forest

Random forest, which were formally proposed in 2001 by Leo Breiman and Adèle Cutler, are part of the automatic learning techniques. This algorithm combines the concepts of random subspaces and "bagging". The decision tree forest algorithm trains on multiple decision trees driven on slightly different subsets of data. Pictorial representation of random forest as shown in Figure 2.

The random forest is part of the family set methods that take the decision tree as an individual predictor, they are based on the methods of bagging, randomizing outputs and random subspace excusing boosting. This algorithm is one of the best among classification algorithms - able to classify large amounts of data with accuracy. It is an ensemble learning method for classification and regression that constructs a number of decision trees at training time and delivers the class that is the mode of the classes output by individual trees.

In random forest classification method, many classifiers are generated from smaller subsets of the input data and later their individual results are aggregated based on a voting mechanism to generate the desired output of the input data set. This ensemble learning strategy has recently become very popular. Before RF, boosting and bagging were the only two ensemble learning methods used. RF has been extensively applied in various areas including modern drug discovery, network intrusion detection, land cover analysis, credit rating analysis, remote sensing and gene microarrays data analysis etc... [9]

There are two ways to evaluate the error rate. One is to split the dataset into training part and test part. We can employ the training part to build the forest, and then use the test part to calculate the error rate. Another way is to use the Out of Bag (OOB) error estimate. Because random forests algorithm calculates the OOB error during the training phase, we do not need to split the training data. Random forest is ensemble of decision trees, which are based on information gain, the computation formula is presented as:

$$inf_{o_A}(D) = -\sum_{i=1}^{m} p_i \log_2 p_i$$
$$gain(A) = inf_o(D) - inf_{o_A}(D)$$

The step of random forest can be represented as:
a. Use bootstrap to extract k samples from the original training sets with N samples for k times,
b. Establish k decision trees,
c. Vote according to the classification results of all decision trees, the voting results called confidence score can be described as:

$$confidence\ score = tree_{number}(positive)/tree_{number}(total)$$
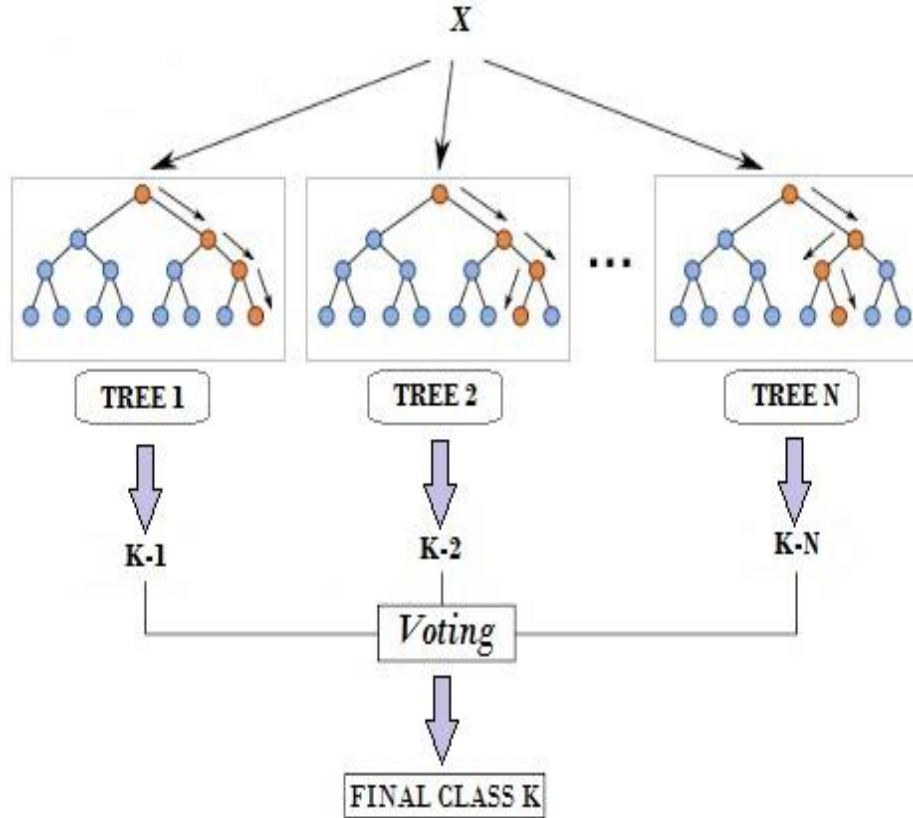
Figure 2. Pictorial representation of random forest

3.2. Support Vector Machine

The SVM method was introduced by Joachims [10], then used by Drucker [11], Taira and Haruno [12], and Yang and Liu [13]. The geometric SVM method can be considered as the attempt to find, among all the surfaces σ_1, σ_2, ... of a space of dimensions |T| which separates the positive learning examples from the negatives. The learning set is given by a set of vectors associated with their class of membership: $(X_1, y_1), (X_2, y_2), ..., (X_u, y_u), X_j \in R^n, y_j \in \{+1, -1\}$ with:

a. y_j represents the class of membership. In a two-class problem the first class corresponds to a positive answer $(y_j = +1)$, and the second class corresponds to a negative answer $(y_j = -1)$.
b. X_j represents the vector of the text number j of the training set.

The Support Vector Machine method separates the positive class vectors from the negative class vectors by a hyperplane defined by the following equation:

$$W \otimes X + b, \qquad W \in R^n, \qquad b \in R$$

For two classes of examples given, the goal of SVM is to find a classifier that will separate the data and maximize the distance between these two classes [14]. With SVM, this classifier is a linear classifier called hyper plan $f(x)$. In the following diagram, we determine a hyperplane that separates the two sets of points. Separation of two sets with separator as shown in Figure 3.
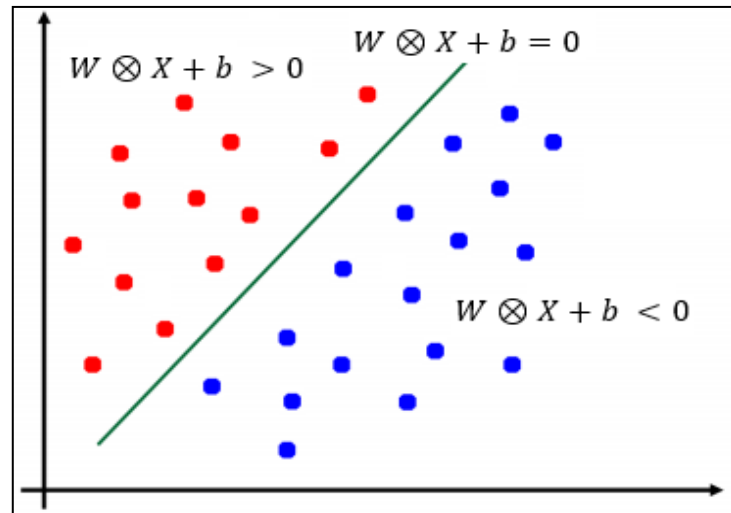
Figure 3. Separation of two sets with separator

In general, such a hyperplane is not unique [15]. The SVM method determines the optimal hyperplane by maximizing the margin: the margin is the distance between the positive labeled vectors and the negative labeled vectors. The learning set is not necessarily linearly separable, variables of gap ξ_j are introduced for all the X_j. These ξ_j take into account the error of classification, and must satisfy the following inequalities:

$$W \otimes X + b \geq 1 - \xi_j$$
$$W \otimes X + b \leq 1 + \xi_j$$

We have to minimize the following function of objective by taking into account these constraints: $\frac{1}{2}\|W\|^2 + C \sum_{j=1}^{u} \xi_j$. The first term of this function corresponds to the size of the margin and the second term represents the classification error, where u represents the number of vectors of the training set. Finding the previous objective function amounts to solving the following quadratic problem: finding the decision function h such that: $h(X) = sign(f(X))$ where:

$$f(X) = \sum_{i=1}^{m} \lambda_i y_i X_i \otimes X + b$$

$sign(x)$ represent the following function:
a.   if $x > 0$ then $sign(x) = 1$
b.   if $x < 0$ then $sign(x) = -1$
c.   if $x = 0$ then $sign(x) = 0$
y_j represent the class of membership,
λ_i represent the parameters to be found
$X_i \otimes X$ represent the scalar product of the vector Xi with the vector X.

The nearest points, which alone are used for determining the hyperplane, are called support vectors. Hyperplane of support vector machine as shown in Figure 4. It is obvious that there is a multitude of valid hyperplane but the remarkable property of the SVM is that this hyperplane must be optimal [16]. We are going to look for it thus more among the valid hyperplanes, the one who crosses "in the middle" points of both classes of examples. Intuitively, it comes down to looking for the "safest" hyperplane. Indeed, let us suppose that an example was not perfectly described, a small variation will not modify its classification if its distance in the hyperplane is big. Formally, this amounts to looking for a hyperplane whose minimum distance to the learning examples is maximum. This distance is called "margin" between the hyperplane and the examples. The optimal separator hyperplane is the one that maximizes the margin [17].

Intuitively, the fact of having a wider margin gets more security when classifying a new example. Moreover, if we find the classifier which behaves best with respect to the learning data, it is clear that it will

also be the one who will at best allow to classify the new examples. On the one hand Figure 5 shows us that with an optimal hyperplane, a new example remains classified well while it falls in the margin. On the other hand, we notice on the Figure 6 that with a smaller margin, the example sees itself badly classified.
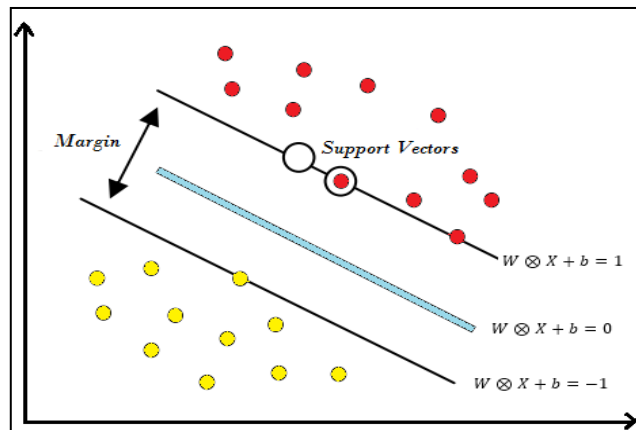

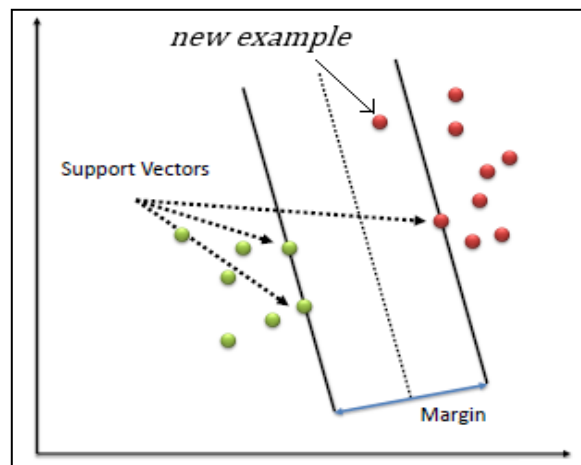
Figure 4. Hyperplane of support vector machine

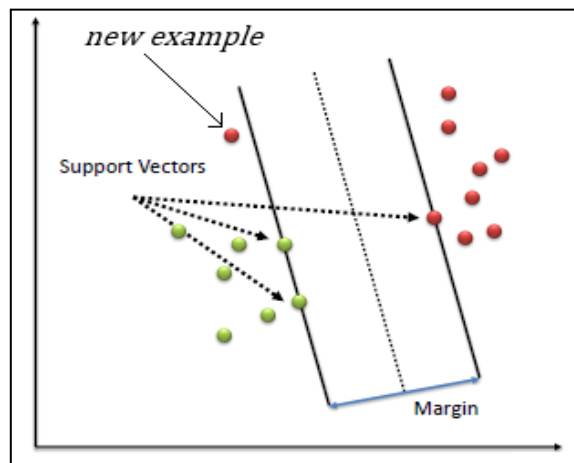Figure 5. Best hyperplane separator

Figure 6. Hyperplane with low margin

In general, the classification of a new unknown example is given by its position relative to the optimal hyperplane. for example, in the Figure 5, the new element will be classified in the category of red balls instead of green balls.

4.   PROPOSED METHOD

In this article, we propose a method which combines the power and the capabilities of Random Forest and Support Vector Machines at the same time for the supervised tasks to solve the problem of classification. Firstly, Random forest is an ensemble learning method that construct a number of decision trees at randomly selected features and predict the class of a test instance by voting of the individual trees. Support Vector Machine revolves around the notion of a margin-either side of a hyperplane that separates two classes.

Maximizing the margin and with this way creating the largest possible distance between the separating hyperplane and the instances on either side of it has been proven to reduce an upper bound on the expected generalization error. RF was not sensitive to input parameters; thus, we just used the default parameters for each classifier. The trained classifiers return scores between 0 and 1, these scores are then transformed to a binary state indicating 'negative' or 'positive'. For each combination, the existence of element is considered positive (P) or negative (N). Before turning to polarity, it may be interesting to identify whether the document corresponds to a subjective opinion or an objective fact. We would have a two-step analysis. Objectivity and subjectivity as shown in Figure 7.
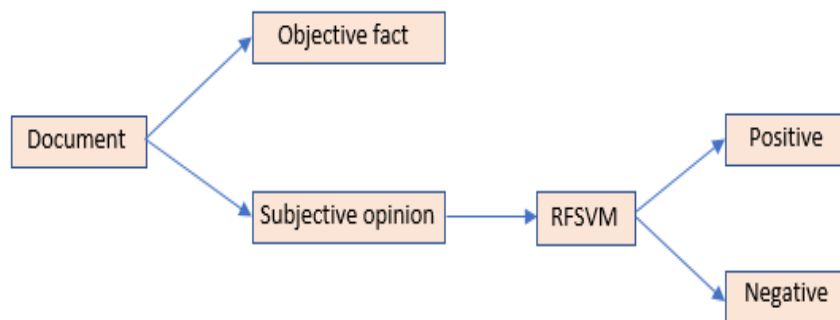

Figure 7. Objectivity and subjectivity

The notation of TP indicates True Positives: number of examples predicted positive that are actually positive, FP indicates False Positives: number of examples predicted positive that are actually negative, TN indicates True Negatives: number of examples predicted negative that are actually negative and FN indicates False Negatives: number of examples predicted negative that are actually positive.

The classification metrics considered for the sentiment analysis are Accuracy, Precision, Recall and F-Measure and these parameters are evaluated based on the calculated positivity and negativity of reviews by the proposed hybrid approach. The performance evaluation of classifiers is made according to the following formulas:

Report of the true positives. It corresponds to:

$$TP\ Rate = TP/(TP + FN)$$

It is thus the report between the number of positive instances classified well and the total number of elements which should be classified well. Report of the false positive one. He corresponds, symmetrically in the previous definition:

$$FP\ Rate = FP/(FP + TN)$$

The datum of the rates TP Rate and FP Rate allows to reconstruct the matrix of confusion for a given class. Precision is the report between the number of the true positive and the sum of the true positives and the false positive. A value of 1 expresses the fact that all the positive classified examples were really:

$$Precision = TP/(TP + FP)$$

Recall is the percentage of correct items that are selected. recall of 1 means that all the positive examples were found.

$$Recall = TP/(TP + FN)$$

Accuracy is a common measure for the classification performance and it's proportional of correctly classified instances to the total number of instances, whereas the error rate uses incorrectly classified rather than correctly.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN)$$

This quantity allows to group in a single number the performances of the classifier (for a given class) as regards Recall and the Precision:

$$F - Measure = (2 * Precision * Recall)/(Precision + Recall)$$

5. RESULTS AND DISCUSSIONS

To evaluate our approach, we used the "Amazon" dataset which contains 1000 instances divided into positive (500) and negative (500). We divided this data into two sets: a training set and a test set. In this article, Cross Validation method with fold value equal to 10 has been used for training and testing phases.

We will use some techniques that automatically extracts this data into positive or negative sentiments. By using the sentiment analysis, the customer can know the feedback about the product before making a purchase. Sentiment analysis is a type of natural language processing for tracking the mood of the public about a particular product.

5.1. Using Random Forest

Table 1 show the result obtained using the Random Forest algorithm. Looking at the results of Table 1, we notice that 820 reviews are correctly classified among 1000, and 180 reviews are misclassified. Figure 8 show the cost of random forest for class positive. Figure 9 show the cost of random forest for class negative.

Table 1. Cross Validation Results for Random Forest

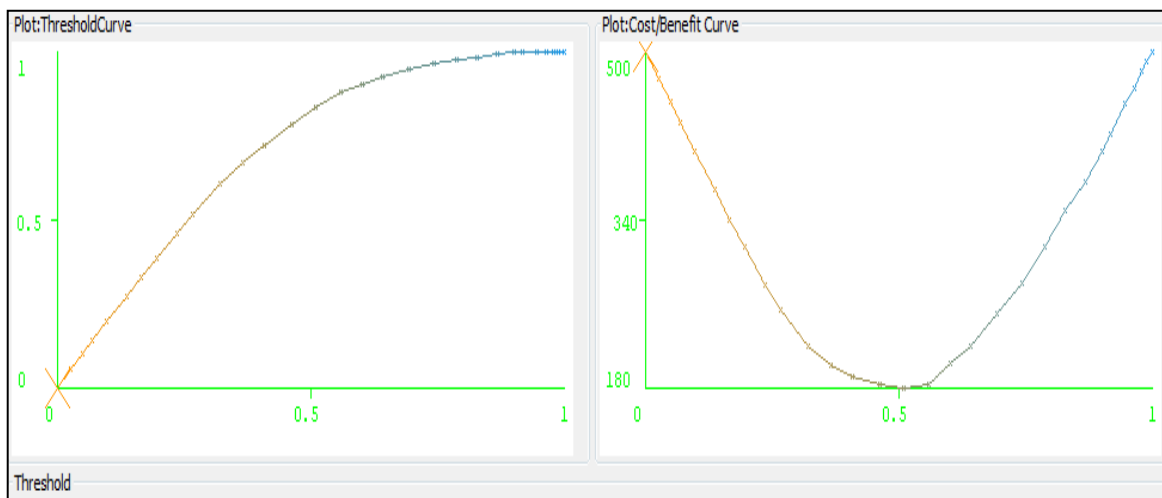|	Positive	Negative	Total
Positive	415	85	500
Negative	95	405	500
Total	510	490	1000

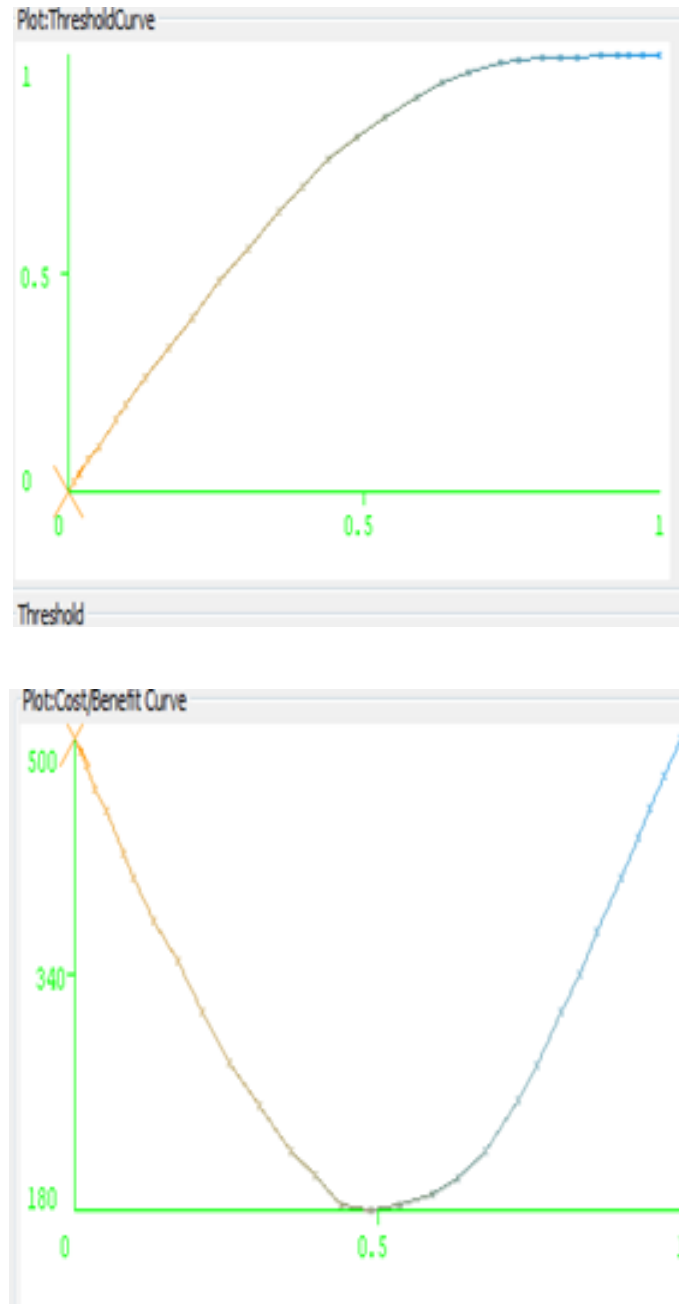Figure 8. Cost analysis of random forest algorithm for class positive

Figure 9. Cost analysis of random forest algorithm for class negative

5.2. Using Support Vector Machine

Table 2 show the result obtained using Support Vector Machine algorithm. Looking at the results of Table 2, we notice that 824 reviews are correctly classified among 1000, and 176 reviews are misclassified. Figure 10 show the cost of support vector machine for class positive. Figure 11 show the cost of support vector machine for class negative.

Table 2. Cross Validation Results for Support Vector Machine

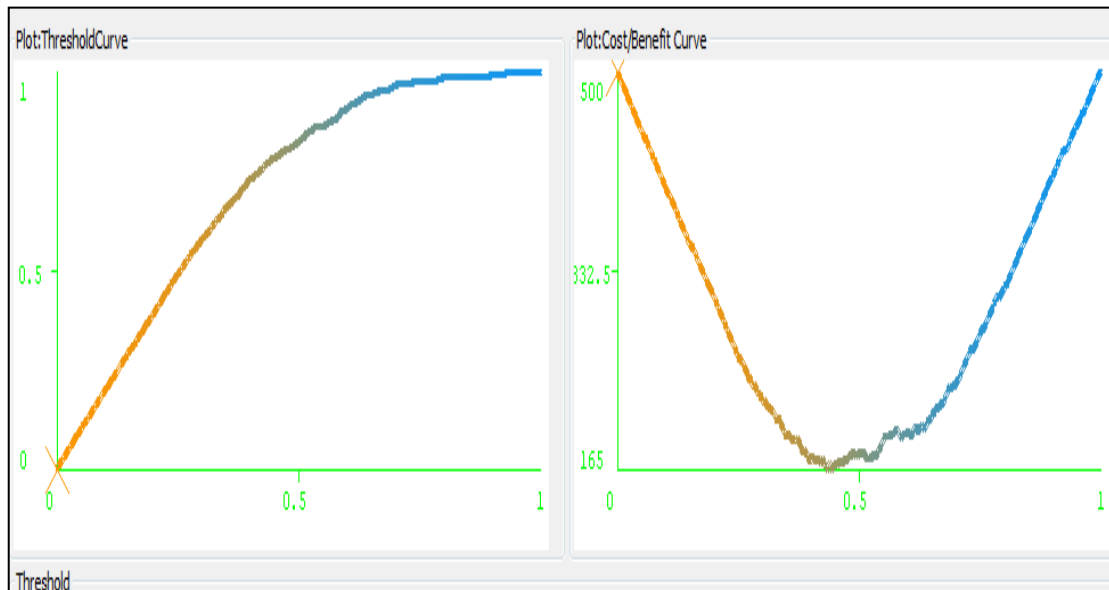|	Positive	Negative	Total
Positive	409	91	500
Negative	85	415	500
Total	494	506	1000

Figure 10. Cost analysis of support vector machine algorithm for class positive
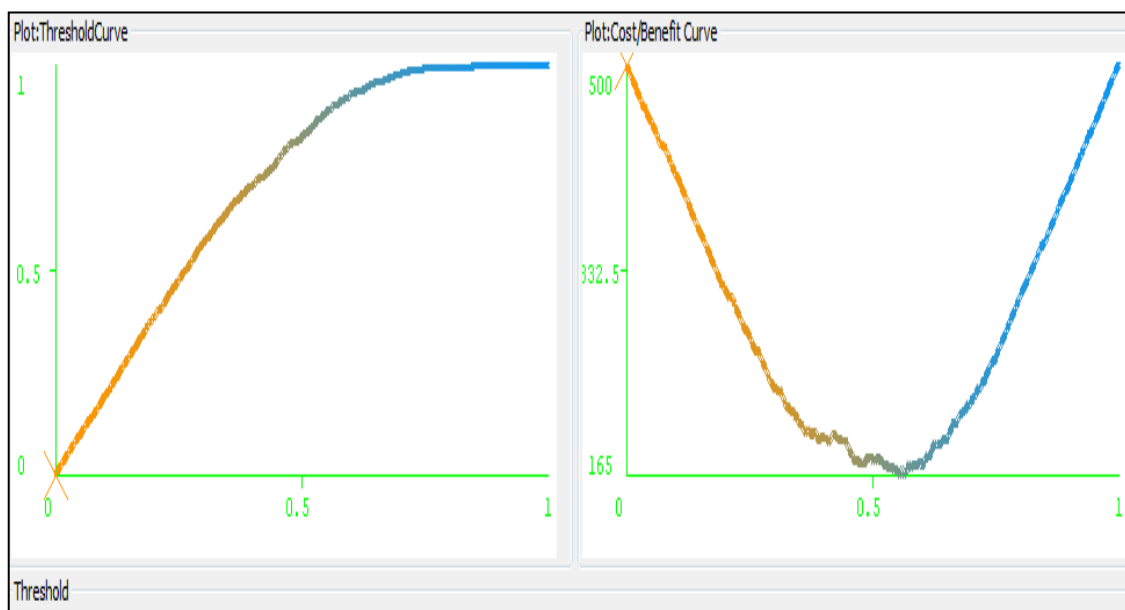

Figure 11. Cost analysis of support vector machine algorithm for class negative

5.3. Using Random Forest Support Vector Machine

Table 3 show the result obtained using our approach Random Forest Support Vector Machine algorithm (RFSVM). Looking at the results of Table 3, we notice that 847 reviews are correctly classified among 1000, and 153 reviews are misclassified. Figure 12 show the Cost of Random Forest Support Vector Machine for class Positive. Figure 13 show the Cost of Random Forest Support Vector Machine for class Negative.

Table 3. Cross Validation Results for RFSVM

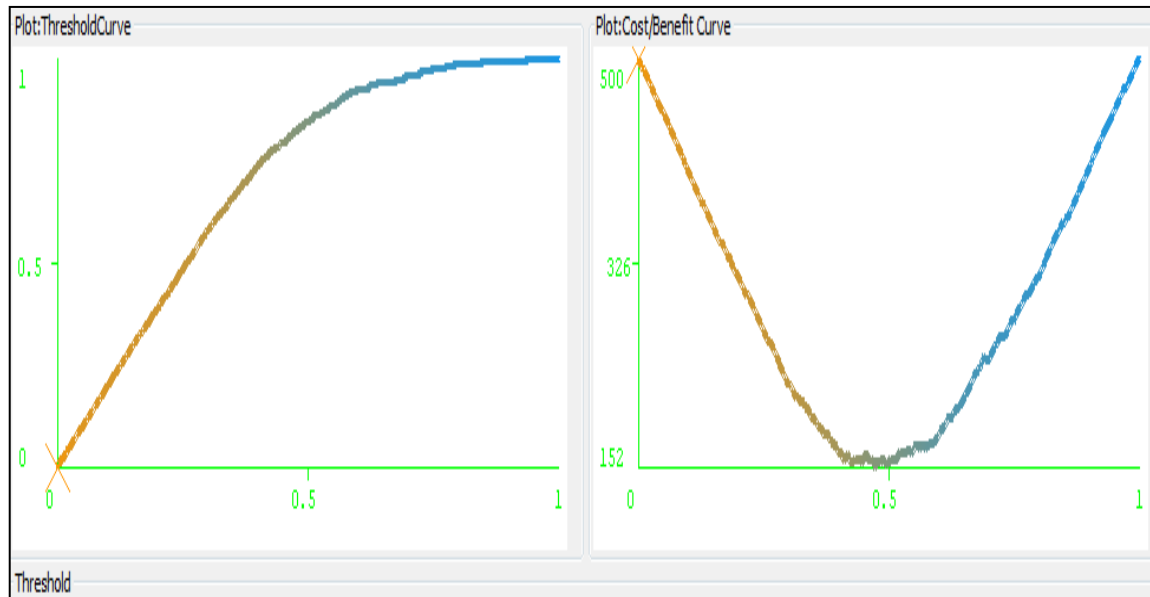|	Positive	Negative	Total
Positive	422	78	500
Negative	75	425	500
Total	497	503	1000

Figure 12. Cost analysis of random forest support vector machine for class positive
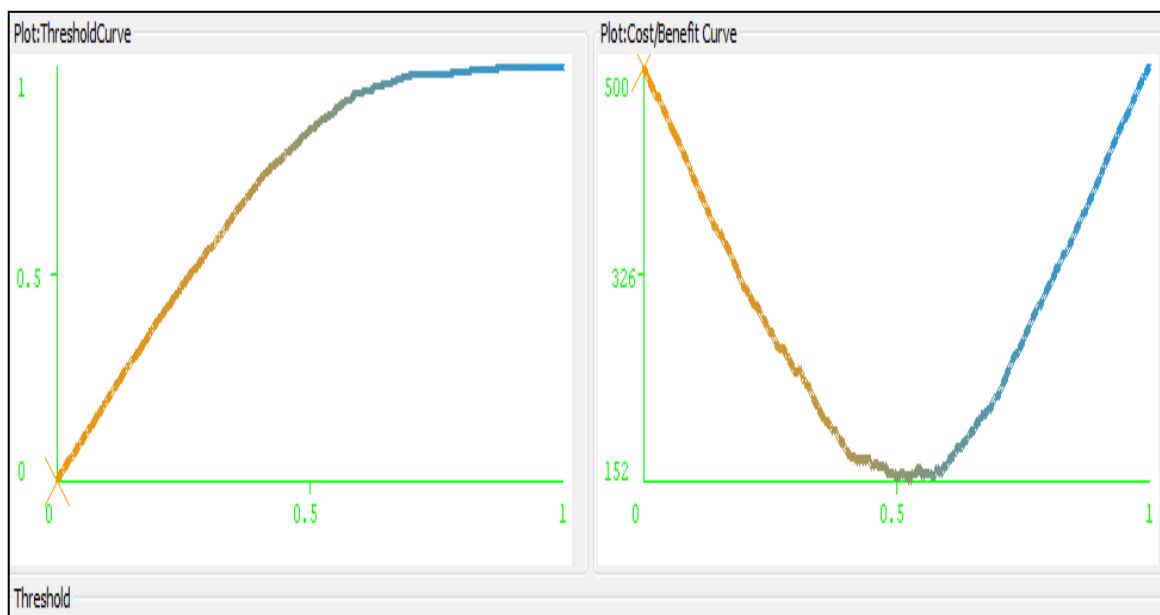

Figure 13. Cost analysis of random forest support vector machine for class negative

We can first discuss and compare the classification performance of each algorithm as well as our approach. The results obtained on the basis of learning with all the methods tested are summarized in Table 4.

Table 4. Representation of Results Obtained Using Proposed Approach and the other Algorithms.

	RF	SVM	RFSVM
Correctly Classified Instances	820	824	847
Incorrectly Classified Instances	180	176	153
Accuracy (%)	82.0	82.4	84.7
Time Taken to Build Model (s)	7.00	2.64	2.31

From Table 4, it is represented that the accuracy computed in the case of proposed method (RFSVM) is better as compared to random forest and support vector machine. Improving the algorithm in different ways could improve the results further. Number of correctly classified instances as shown in Figure 14. From Figure 14 it is evident that Random Forest Support Vector Machine (RFSVM) shows the best performance as compare to other studied algorithms.
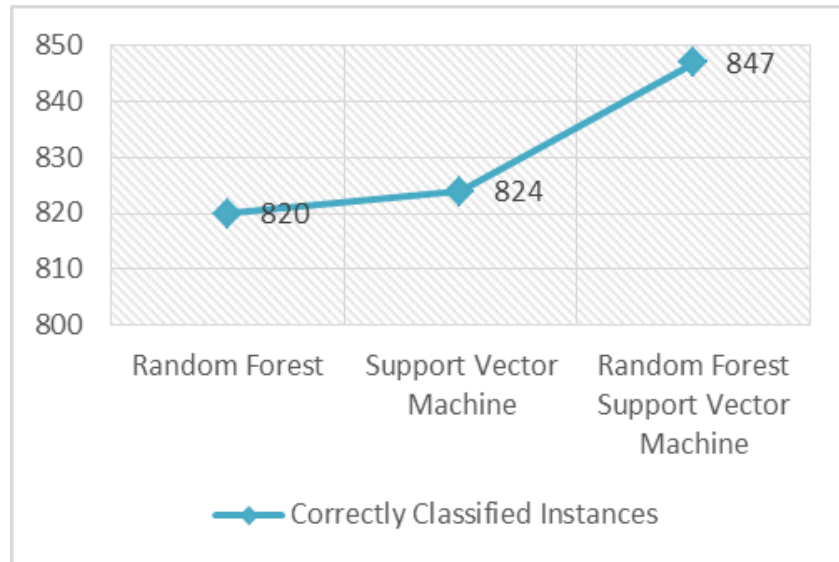

Figure 14. Number of correctly classified instances

For Table 5, we found that the precision, recall and f-measure for RFSVM were 84.7 %, 84.7 % and 84.7 %, respectively. The f-measure of RFSVM was higher than that of others algorithms, which meant RFSVM fitted better than these classifiers. The hybrid approach combines the advantage of both the Random Forest and Support Vector Machine. It is inheriting more accuracy using supervised machine learning approaches and providing good stability against the other algorithms.

Table 5. Number of Classified Instances

	RF		SVM		RFSVM	
	Positive	Negative	Positive	Negative	Positive	Negative
TP Rate	83.0	81.0	81.8	83.0	84.4	85.0
FP Rate	19.0	17.0	17.0	18.2	15.0	15.6
Precision	81.4	82.7	82.8	82.0	84.9	84.5
Recall	83.0	81.0	81.8	83.0	84.4	85.0
F-Measure	82.2	81.8	82.3	82.5	84.7	84.7
ROC Area	90.3	90.3	91.4	91.4	92.0	92.0

The paper considered the combination of supervised classification algorithms to product review data and also predicted the positive and negative reviews by people. The hybrid approach which contains the combination of Random Forest and Support Vector Machine produced better results on the basis of Accuracy, Precision, Recall and F- Measure. Random Forest approach improved the performance in the case of small reviews and Support Vector Machine improved the performance just in case of large reviews are working as a single hybrid approach.

6.   CONCLUSION

Although the results obtained are interesting and encouraging, many points are likely to be studied in future work to improve performance and achieve better results, such as the use of other classifiers and the experimentation of approaches different from the one proposed in this paper. In this work, we have compared Support Vector Machine, Random Forest and our approach (RFSVM) which are very suitable for generating rules in classification technique. From the experimental results, it is concluded that Random Forest Support

Vector Machine algorithm seems better than the other algorithms for product reviews dataset offered by Amazon.

REFERENCES

[1]   Jagtap V S, Pawar K. Analysis of different approaches to Sentence-Level Sentiment Classification. *International Journal of Scientific Engineering and Technology.* 2013; 2(3): 164-170.
[2]   Umadevi V. Sentiment analysis using weka. *International Journal of Engineering Trends and Technology.* 2014; 18(4): 181-183.
[3]   Xia R, Zong C, Li S. Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences.* 2011; 181(6): 1138–1152.
[4]   Prabowo R, Thelwall M. Sentiment analysis: A combined approach. *Journal of Informetrics.* 2009; 3(2): 143-157.
[5]   Liu B. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies.* 2012; 5(1): 1-167.
[6]   comScore/the Kelsey group. Online consumer-generated reviews have significant impact on offline purchase behavior. *Press Release.* 2007.
[7]   Dhokrat A, Khillare S, Mahender C N. Review on Opinion Mining for Fully Fledged System. *Indonesian Journal of Electrical Engineering and Informatics.* 2016; 4(2): 141-148.
[8]   Al Amrani Y, Lazaar M, El Kadiri K E. Sentiment Analysis Using Hybrid Method of Support Vector Machine and Decision Tree. *Journal of Theoretical and Applied Information Technology.* 2018; 96(7): 1886-1895.
[9]   Rodríguez-Galiano V F, Abarca-Hernández F, Ghimire B, Chica-Olmo M, Akinson P M, Jeganathan C. Incorporating Spatial Variability Measures in Land-cover Classification using Random Forest. *Procedia Environmental Sciences.* 2011; 3: 44-49.
[10] Joachims T. *Transductive inference for text classification using support vector machines.* Proceedings of the Sixteenth International Conference on Machine Learning. 1999: 200–209.
[11] Drucker H, Donghui W, Vladimir Vapnik N. Support Vector Machines for Spam Categorization. *IEEE transactions on neural networks.* 1999; 10(5): 1048-1054.
[12] Taira H, Haruno M. *Feature selection in svm text categorization.* Conference of the American Association for Artificial Intelligence (AAAI). 1999: 480-486.
[13] Yang Y, Liu X. *A re-examination of text categorization methods.* Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. ACM digital library. 1999: 42-49.
[14] Al Amrani Y, Lazaar M, El Kadiri K E. *Sentiment analysis using supervised classification algorithms.* Proceedings of the 2nd international Conference on Big Data, Cloud and Applications, ACM digital library. 2017; (61).
[15] Prasad S V S, Satya Savithri T, Murali Krishna Iyyanki V. Comparison of Accuracy Measures for RS Image Classification using SVM and ANN Classifiers. *International Journal of Electrical and Computer Engineering.* 2017; 7(3): 1180-1187.
[16] Hidayatullah A F, Ratnasari C I, Wisnugroho S. Analysis of Stemming Influence on Indonesian Tweet Classification. *TELKOMNIKA.* 2016; 14(2): 665–673.
[17] Hsun-Jung C, Ming-Te T. A support vector machine approach to CMOS-based radar signal processing for vehicle classification and speed estimation. *Mathematical and Computer Modelling.* 2013; 58: 438- 448.

BIOGRAPHIES OF AUTHORS

Yassine AL AMRANI is a PhD student in Abdemalek Essaadi University, Tetuan, Morocco. He is a Computer Sciences engineer, graduated in 2014 from National School of Applied Sciences, Abdelmalek Essaadi University. He got a DEUST on Mathematics Computer Sciences and Physics (MIP) in 2011 from Faculty of Sciences and Technologies, Moulay Ismail University.

Mohamed LAZAAR is a PhD doctor and professor of computer sciences at the National School of Applied Sciences of Tetuan (ENSATe). He is a supervisor of many Thesis, notably in the Machine Learning and Big Data. His research interest focused on machine learning and Big Data "applying Neural Networks, Markov Model, SVM and others to a practical problem in computer systems". He is the editor of the Proceedings of the International Conference on Big Data, Cloud and Applications (CEUR Workshop Proceedings 1580, ACM Proceeding ISBN 978-1-4503-4852-2), editor of the Special Issues (igi-global IJISSS and IJCAT Inderscience for BDCA'17 Conference) BDCA'17 Conference and IJCC Inderscience V(5) N(3) (2016) for BDCA'15). He is a founder of BDCA Conference.

Kamal Eddine ELKADIRI is a PhD doctor and professor of computer sciences at the Faculty of Sciences of Tetuan, Morocco. He is the Director of the ENSA School of engineers of Tetuan and the Director of LIROSA laboratory. He is part of many boards of international journals and international conferences.