Lexicon Based Sentiment Data Creation Technique for Product Protection (SDTCPP)

S. Geetha
Assistant Professor, Department of Computer Science,
Muthurangam Govt. Arts College, Vellore,Tamil Nadu, India
[1]geeth_s20@yahoo.com

Dr. R. Kaniezhil
Principal, MIT College of Arts & Science for Women, Musiri, Tamil Nadu, India
[2]kaniezhil@yahoo.co.in

Abstract - The market is pounded with the same hazardous material to users, specifically skin and healthcare. It also affects people both in terms of their health and economy. This makes it necessary to monitor products from user opinions for its protection from duplicates. Online users have increased rapidly and share their opinions on occasions, brands, individuals, products, and events occurring across the world. Millions of users participate in SNSs (Social Networking Sites) like Facebook, Twitters, Instagram and WhatsApp etc. Product reviews in public forums like SNSs can provide early clues on such duplicates in terms of allergies or adversities while using them. SA (Sentiment Analysis) can play a significant role in such surveillances. This paper proposes a framework for reviewing experiences in cosmetics in skincare using the Amazon reviews dataset. The proposed technique is based on safety lexicons which are trained for classifications from user sentiments.

1. Introduction

The global scourge of counterfeit drug and cosmetic products poses a significant threat to public safety. One strategy for combating counterfeit products is through the effective communication and tracking of early warning signals of product allergies, side or adverse effects, drug resistance and disease outbreaks [1]. E-commerce websites have gained popularity in providing the ability to shop from home and at discounted prices. Online users, not the only shop, but also share their opinions on SNSs. Millions of comments or reviews are generated on these sites, making it complicated for an organization to trace and track public opinion continually. This makes it imperative to study and classify these reviews to extract useful information from this data centre [2]. DM (Data Mining) techniques find their use of simplifying and extracting patterns from public sentiments. Thus, analyzing customer opinions is useful for business and has widened its scope in health care, sports, politics, recommendation system, and many other domains. Figure 1 depicts the SA Application Areas.

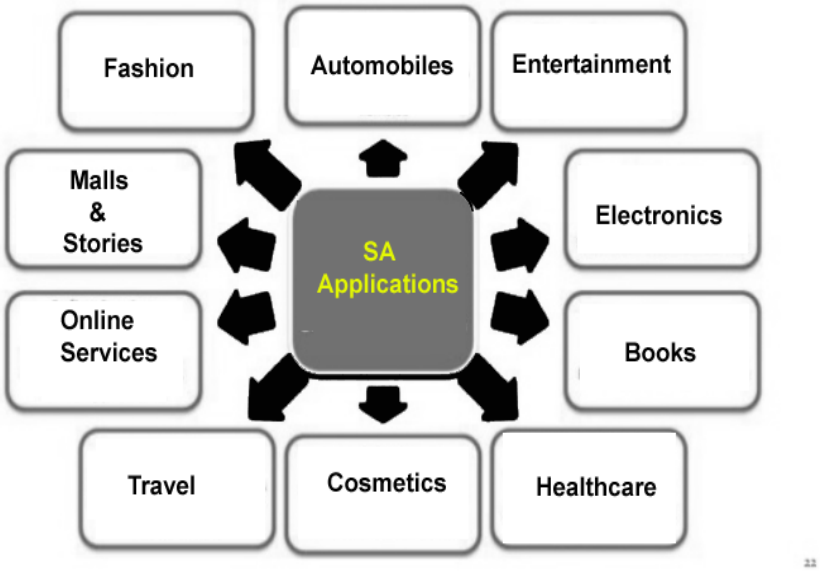

Fig. 1: SA Application Areas

A generic problem for organizations is automating classification of opinions when the dataset size increases. Moreover, many researchers have turned their opinion towards these areas by proposing methods that analyze for learnable outcomes [3]. OM (Opinion Mining) is a part of NLP (Natural Language Processing) that can extract personal information from texts. SA identifies public sentiments in three categories, namely Negative, Neutral & positive. This, however, is not done by just labelling words as positive or negative. Thus, any SA technique aims to analyze users' written reviews and classify them into positive/negative/neutral opinions [4]. Hence, this paper aims to perceive and decipher online users' opinions on skincare products that organizations can use to identify duplicates or ensure harmful products are not added in their kitbags [5]. Though there are many issues in classifying public opinion, the proposed scheme called SDTCPP (Sentiment Data Creation Technique for Product Protection) attempts to overcome them using a lexicon-based approach that prepared the classification data. The technique addresses the hazard of product counterfeiting with its proposed surveillance capable of harnessing and tracking online reported views and experiences of studied products' users. Users can apply SDTCPP, product manufacturers, regulatory and enforcement agencies to monitor brand or product sentiment trends to act in the event of a sudden or significant rise in negative sentiment. This paper is organized as follows. Section 2 lists related studies and section three explains the methodology used. Section 4 is the results section, while the paper concludes in section 5.

2. Related Review of Literature

SNSs have become the most useful information exchange tool of the 21st century. People from ages use SNSs to post messages, photos and videos about their daily activities. SNSs provide convenient and efficient ways of communicating and sharing information publicly. These sites are rapidly becoming information sources for early warning systems in public safety as most professionals utilize social media. At least 67% of people use social media tools for gathering information [6]. SA is an evolving field in research and includes many allied areas of computational research like ML, NLP, linguistics and text mining [7]. It is an analysis of sentiments, attitudes, emotions, subjectivity of thoughts from user's comments. SA was studied in detail by [8] who focused on SA, challenges, tasks, applications, and types. The primary tasks were polarity determinations, sentiment classifications/extractions and opinion summarizations. SA has been already applied in several different, non-security domains for monitoring and forecasting public opinions. In [2], the authors applied a domain-specific lexicon to classify hotel customer reviews into five-star categories. SA can majorly contribute to forecasting product satisfaction or product imitations or product benefits, in short, product protection [10-12]. The study in [13] presented a model for semantic word representation using a neural language model.

Both local and global words were used in semantic representations. Twitter data was used in [14] for SA, which included pre-processing and feature extraction. The study used SVM (Support Vector Machines) for classifying sentiments based on its polarity, which resulted in an accuracy of 80%. Continuing further, the study in [15] used SVM, NB (Naïve Bayes) and KNN (K-Nearest Neighbors) which was coded in python for improving the accuracy of classifications. Events were classified in [16] where their proposed scheme characterized topics from Twitter public sentiment like the U.S. Presidential election. A text classification model for assessing cyberbullying was presented in [17] which recognized different emotions like anger, embarrassment, empathy, fear, pride, relief, sadness from Twitter data. Correlations between public opinion and stock market sentiments were analyzed in [18]. The study used Twitter data to identify correlations by classifying messages into four different moods: calm, happy, alert and kind and based their predictions on the Dow Jones Industrial Average. Studies have also considered social media-based intelligent support systems in public safety. People with cancer were detected using a probabilistic model in [19]. The study predicted the disease's risk based on social ties, co-location from tweets. The study [20] monitored influenza's diffusion among the masses using SA while security informatics was addressed [21]. This work referred to the discovery of security-relevant data, awareness and predictive analysis. Their experiments were conducted on cyber incidents, public opinions, emerging topics/trends and possibilities of protests. Thus, the studies mentioned above demonstrate that social media and sentiment analysis have been considered in many different application domains.

2.1. *SDTCPP*

Lexicon-based Sentiment Analysis techniques are based on calculating polarity scores given to positive and negative words in a document. They can be broadly classified into Dictionary-based and Corpus-based, Dictionary-based methods create a database of positive and negative words from an initial set of words by including synonyms and antonyms while Corpus-based methods, on the other hand, obtains the dictionary from the initial set by the usage of statistical techniques. Application of lexicons is one of the main approaches to SA as it calculates sentiments from the semantic orientation of word or phrases in the text [22]. Lexicon methods use a dictionary of positive and negative words is required and positive or negative sentiment values assigned to each word. Different approaches to creating dictionaries have been proposed, including manual [23] and automatic [24] approaches. Thus, internet users can provide early clues about adverse effects through their Internet surfing. The work in [25] found in the visualization of CHFpatients.com forum that chat sentiments were used to measure the effectiveness of a drug by quantifying its side effects, particularly for the benefit of the forum members and their

physicians. SDTCPP aims to apply the same approach to drug/cosmetic product users, with intended beneficiaries as product users, manufacturers, regulatory and enforcement agencies. The proposed technique attempts to achieve product protections by harnessing Social Media users' experiences of popular drugs/cosmetics. The proposed methodology follows five stages: identifying Sentiments based on the lexicon, finding negative words, finding the intenseness of positive words, combining the previous steps through a function, and finally classifying to predict product protections. The figure depicts the SDTCPP Architecture

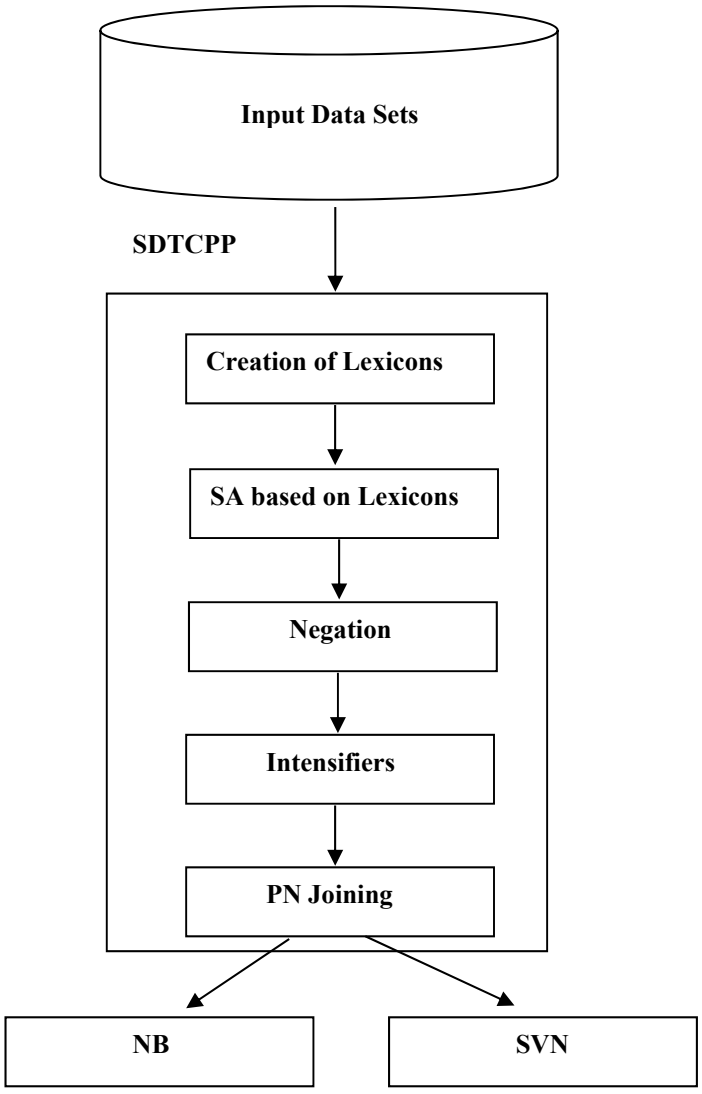

Figure 2: SDTCPP Architecture

a. **Lexicon based Sentiments:** The sentiment lexicon was constructed manually with 8000 words, including positive and negative words. Each word in the lexicon was assigned a value representing sentiment in −50 (most negative) to 50 (most optimistic). Empirical knowledge highlighted positive and negative words in a sentence. For example, the sentence "The Cosmetic was good but very oily" represents a positive message

Pb(positive|Wr) for positive words

(1)

Pb(negative|Wr) for negative words

b. that contains one positive (right) and one negative (oily) word which makes it difficult to say if it is positive or negative. SDTCPP sidelined this issue with a sentiment value for each word from the lexicon and estimated a conditional probability (denoted by Pb) as presented in the equation.

Pb(positive|Wr) for positive words

(1)

Pb(negative|Wr) for negative words

where # Pb – Probability and Wr – Total number of words

Each word's positive /negative probability was estimated and labelled with its frequency was computed from selected positive and negative messages using conditional probability depicted in equation (2).

$$\begin{aligned} &\text{Pb(positive|Wr) for positive words} = \text{Pb (positive} \cap \text{Wr)} = \#\,\text{Wrp}/\,\#\text{Wr} \\ &\text{Pb(negative|Wr) for negative words} = \text{Pb (negativee} \cap \text{Wr)} = \#\,\text{Wrn}/\,\#\text{Wr} \end{aligned} \tag{2}$$

where #Wrp – Number of Positive words, Wrn – Number of negative words, Pb – Probability and Wr – Total number of words. Equations (1) and (2) were applied recursively to estimate the probabilities.

c. **Negation:** Handling negation is typically done by reversing the lexicon item's polarity in a sentence [27]. SDTCPP used a negating function represented in equation (3) that calculated a negated word's value. The manual creation of negations included more negative words for identifying duplicates and product protection.

$$F_N(S) = \begin{cases} max\left\{\frac{S+100}{2}, 10\right\} & if\ S < 0 \\ min\left\{\frac{S-100}{2}, -10\right\} & if\ S > 0 \end{cases} \tag{3}$$

where F_N – Final Negation, S- Sentiment value from the lexicon. After identifying negatives in sentences, non-neutral words are searched for assigning positive/negative words using equation (3).

d. **Intensifiers:** SDTCPP uses intensifiers which can change sentiment scores of non-neutral words. They either increase or decrease the scores. These intensifiers are not a part of the sentiment lexicon, but when they appear in the text in a neighbourhood of positive or negative words, they are considered non-neutral assigned a score.

e. **P-N Joining:** After identifying polarity in words from sentence based on their local contexts is verified, a joining process for final sentiment values is used which sums polarity of individual words. Thus values of the sentences themselves range between -50 to +50. This function helps in modelling polarity scores of sentences. For example, "The Product is perfect (50)", "The product can be found easily (40)", "I would not advise this product as it is dangerous (-50)". Therefore, the overall positive/negative sentiment should be represented as a product of the average sentiment and a coefficient that's value depends on the number of positive/negative words. SDTCPP normalizes the sentences as depicted in equation (4)

$$\begin{aligned} F_P &= min\left\{\frac{A_P}{2 - \log(P \times W_{rP})}, 50\right\} \\ F_N &= max\left\{\frac{A_P}{2 - \log(P \times W_{rN})}, 50\right\} \end{aligned} \tag{4}$$

Where A_P - Average positive sentiment, A_N – Average negative sentiment, Wr_P – Positive word count, Wr_N – Negative word count. The use of logarithm models the relationship between positive/negative words count given by F_P and F^N - Average sentiments in a sentence. Most messages contained a minimum of three non-neutral words and hence an induced coefficient p for a sentiment score of 50 was assumed as three. . Ro find the exact value of this co-efficient, SDTCPP uses equation (5)

$$\frac{1}{2 - \log(P \times W_{rP})} \tag{5}$$

This normalization process results in two values ranging from 0–50 for total positive sentiments and -5 to 0 for total negative sentiments. Equation (6) is used to join polarized words in mixed sentiment sentences.

$$\begin{aligned} e_P &= min\left\{\frac{A_P}{2 - \log(3.5 \times W_{rP})}, 50\right\} \\ e_N &= max\left\{\frac{A_N}{2 - \log(3.5 \times W_{rN})}, -50\right\} \end{aligned} \tag{6}$$

Where, e_P - Pieces of evidence of positive sentiments, e_N- Evidence of negative sentiments. All values lesser than .5 were eliminated, and thus, values between 1 and −1 were considered based on equation (7).

$$\begin{aligned} e_P &= min\left\{\frac{A_P}{2 - \log(3.5 \times W_{rP})}, 1\right\} \\ e_N &= max\left\{\frac{A_N}{2 - \log(3.5 \times W_{rN})}, -1\right\} \end{aligned} \tag{7}$$

2.2 Classification

Since the final output of SDTCPP is acceptable to ML techniques, it is evaluated using ML classifiers NB and SVM. ML techniques use training to learn and then test on the learned model. It then predicts the direction of sentiment in documents. NB determines classes in input where the classification of the potential class c* in document d can be computed using equation (8):

$$e^* = max_c \ p(c/d) \tag{8}$$

and subsequent probabilities in outputs can be depicted as equation (9)

$$p(c_j/d_i) = \frac{p(c_j)\ p(c_j/d_j)}{p(d_i)} \tag{9}$$

Where $\boldsymbol{p(c_j|d_i)}$ the future probability of the class c_j of document $\boldsymbol{d_i}$. The probability of computed c_j *in equation (10)*

$$P(c_j) = \frac{N_i}{N} \tag{10}$$

Where N_i – count of documents classified in c_j, N – total documents in all classes and **P(d)** are the probability of document d.

SVM's are popular ML techniques and efficient in the classification of texts [28]. SVMs' minimize structural risks by dividing data points into two classes where support vectors play a significant role in selections. In a two-class problem, SVM's optimize decisions by separating hyper-plane between the data points. If X a set of feature vector $(x_1, y_1),\ldots, (x_n, y_n)$ where each point $x_i \in RN$ with label $y_i \in \{-1, +1\}$, where i = 1,…, n, the function $fn(x) = w.x_i + b$ identifies classes $y(x) = sign(fn(x))$ that is optimized using Equation (11)

$$\min_{w,b} \sum_{i=1}^{n} [1 - y_i(w.x_i + b)] + \frac{\lambda}{2}\|w\| \tag{11}$$

Where, λ – parameter for regularization parameter, x_i - feature vectors, $y_i \in \{-1,+1\}$, w- average hyper-plane vector and b: hyper-plane offset.

3. Experimental results

Data set used in the study was Stanford which contains positive and negative sentiments in equal proportions. The lexicon-based outputs were applied on 5 generic products, namely Facial Creams, Body Lotions, Deodorants, Health Drinks and Hair Oil. Python 3.9.1 was used for evaluations and outputs of lexicon-based dataset generations and subsequently for classifications using NB and SVM. TABLE 1lists the average values of products in its distribution of sentiments scores.

TABLE I - DISTRIBUTION OF SENTIMENT SCORES FOR THE 5 PRODUCTS

Product	Negative	Neutral	Positive
Facial Creams	10	23	23
Body Lotions	12	36	46
Deodorants	31	22	42
Health Drinks	6	12	12
Hair Oil	8	21	48

A total of 11000 reviews were used in the study. Table 2 lists a random preview of sentences in the dataset.

TABLE II - RANDOM PREVIEW OF Sentences

Positive sentences	Negative sentences
I look better gives you an advantage.	Very costly
I walk more after the energy drink	I am unsure of actual ingredients
The deodorant stays all-day	.Too many duplicates be careful
My hair has lesser dandruff	Hair oil is too smelly

Figure 3 depicts SDTCPP's snapshot of the manually created list of negative lexions, while positive words are depicted in Figure 4.

doubt,	UV rays	pigmented	lash clumps
dedicated	acne-prone	pores	lifeless
honey	affected area	processed	lined
untrustable	age	puffiness	smudge
useless	buildup	puffy	splotchy
examine	caky	skin-damaging	spots
unprofessional	duplicate	imitation	incompetent
wrong	chapped	exposure	sun
inconvenient	cracked	fade	tightness
note	creases	fading	tired skin
heavy	damaged	incompetent	troubled
cancellations	dark circles	flaking	sun burn
dirty	delicate	flat	wind
discoloration	detergents	formaldehyde-releasing age	wind-burned
drying	dirt	free radicals	wrinkles
dryness	old age	greasy	sun damage
dull	outdoors	harsh	sun-drenched
dullness	pain	heat	swollen
enlarged pores	phthalates	humidity	synthetic fragrances
mature	runs	imperfections	tight
scam	scarring	impurities	sweaty
oily	sensitive	inflamed	swelling
old	shadows	irritated	skin afflictions

Fig. 3 : SDTCPP Negative Lexicons

complex	reach
restrict	great
supplement	remarkable
unbeatable	evident
candy	climate
achievement	awesome
affordable	wonder
mess	deligh
perfection	exact
spotlight	excellent
amazing	exquisite
satisfied	acceptable
assistance	recommend

Fig. 4 : SDTCPP Positive Lexicons

SDTCPP's intensifiers were based on most frequently applied intensifiers divided into three types: reducers (decrease sentiment value by fifty percent) and weak, strong amplifiers (increase sentiment value by 70 to 100 percent) sentences. Positive/negative words identified are assigned sentiment value based on equation (3) like (Enjoyed:10, Hate: −20). This is contrary to polarity reversion for better accuracy in assigning values to negative words. Only sentences greater than an average value of 25 was considered. Halving values obtained in Equation 2) for ensuring high or low sentiments. Figure 5 depicts a snapshot of SDTCPP Intensifiers.

Benevolent	characterized by or expressing goodwill or kindly feelings.
Bewildered	completely puzzled or confused; perplexed.
Biting	sarcastic, having a biting or sarcastic tone.
Calm	free from excitement or passion; tranquil.
Candid	frank; outspoken
Desperate	having an urgent need, desire.
Detached	impartial or objective; disinterested; unbiased/ not concerned; aloof.
Ecstatic	in a state of ecstasy; rapturous.
Effusive	unreserved or unduly demonstrative.
Facetious	inappropriate; flippant
Frustrated	annoyed; discouraged

Fig. 5: SDTCPP Intensifiers

SDTCPP also normalizes data while combining sentence average and number of words to calculate the average. Initially, SVM, NB was applied as a baseline to the entire feature space. Seventy percent of the dataset was used for training and the balance used for testing by both the classifiers. A tenfold validation evaluated each algorithm to segment the dataset into ten equal sizes sub-samples. Moreover, their results were averaged to produce a single estimation. One implies selected feature vector and 30 tests were run. Figure 6 displays the output of the process.

NO	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	NB	SVM
1.	1	1	1	1										86.29	91.50
2.					1	1	1	1						88.81	88.88
3.					1	1	1	1	1	1				86.23	88.81
4.											1	1	1	84.32	91.55
5.	1	1	1	1	1	1	1	1						88.70	90.36
6.					1	1	1	1	1	1				87.54	90.16
7.									1	1	1	1	1	88.65	90.24
8.	1	1	1	1	1	1	1	1	1	1				88.88	92.17
9.					1	1	1	1	1	1	1	1	1	88.35	91.52
10.	1				1				1		1			86.74	91.22
11.		1				1				1		1		85.57	90.94
12.			1		1				1				1	86.08	91.10
13.				1		1	1	1						86.62	90.31
14.	1	1			1	1			1	1	1	1		87.24	90.10
15.			1	1	1	1			1	1		1	1	86.89	91.13
16.	1		1		1	1		1	1	1	1		1	87.44	91.20
17.	1		1	1	1	1	1		1	1	1	1		87.24	91.15
18.	1				1									82.66	90.60
19.		1						1						82.87	90.16
20.			1								1			83.30	90.54
21.				1		1								82.87	89.79
22.	1							1						82.76	90.84
23.		1										1		82.68	90.45
24.			1										1	86.48	89.79
25.	1				1				1					84.90	90.22
26.		1				1		1						85.14	90.22
27.			1		1						1			85.63	90.60
28.				1					1			1		87.74	90.01
29.	1				1								1	87.25	90.38
30.	1				1				1		1			84.06	90.38

Fig.6: B and SVM measures

The classifiers used in the study achieved more than 88% inaccuracy which was evaluated with F1-measure, where NB and SVM classifiers were applied as separate groups in the proposed methodology's output. Figure 7 depicts the F measure of the classifiers.

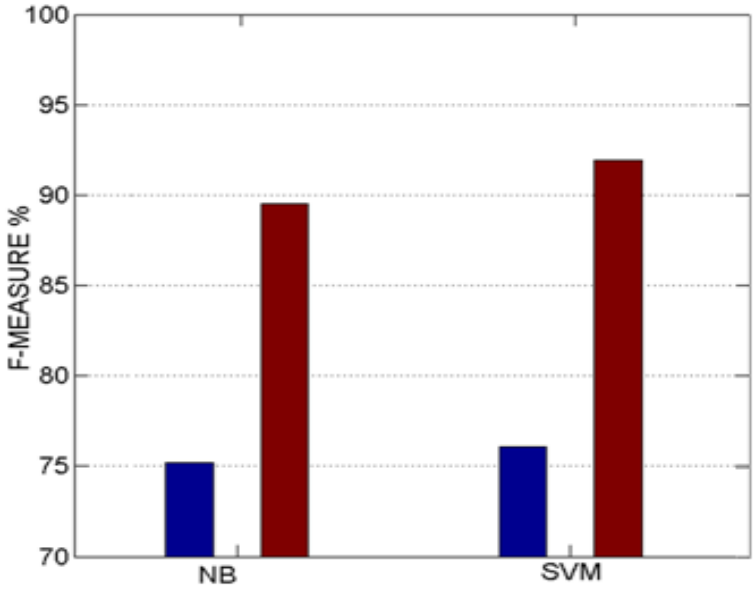

Fig. 7: F-measure scores of classifiers

Figure 7 results suggest that F-measures of lexicon-based result in an average accuracy of 90%. The classifier model's choice dramatically affects the quality of SA, and the value of SVM is 92% in terms f accuracy. The group with NB classifier showed lesser accuracy than SVM. Any model used in classification has to base its accuracy of correct predictions based on performance measures. F Score is the measure of accuracy and balances precision and the recall values.

4. Discussion

This paper proposes a Malay sentiment analysis classification model for improving classification performances based on the semantic orientation and machine learning approaches. First, a total of 2,478 Malay sentiment-lexicon phrases and words are assigned with a synonym and stored with the help of more than one Malay native speaker, and the polarity is manually allotted a score. Besides, four classification approaches (Naïve Bayes, SVM, SDTCPP and combination method) are used to evaluate Malay sentiment classification by using four subsets of features (presence of sentiment words and frequency, sentence level, sentiment words polarity features and subjective words conditional probability features). Finally, it highlights that the Malay sentiment analysis classification model enhances the classification performances with employing the four-classification approach (Naïve Bayes, SVM, SDTCPP and combined-classification approach). Experimental results show that the combination method, which combines various feature sets and classification algorithms, can achieve the best result with an F-measure value of 94.48%. It is a more efficient way to improve classification performances compared with the existing classifiers.

5. Conclusion and Future Work

This paper has presented a new lexicon-based classification of sentiments. In this approach, sentiments are normalized. A new evidence-based joining function was developed to improve classifiers' performance, as showed in classifier evaluations. This paper has demonstrated that the proposed technique can be used by ML techniques to infer sentiments over social media data suggesting views and experiences of drugs/cosmetic products by users. The framework harnessed users reviews using text mining and SA. The methodology's utility was probed by taking five product categories. NB and SVM classified the proposed technique's outputs. The framework utilized users views and experiences on cosmetic/drug products in their reviews. The study has demonstrated how to develop custom lexicon and train data for modelling using NB and SVM in classifications. The results achieved can be summarized as detailed below

- Public sentiments were assessed on given brands of cosmetic product
- Conversations of users were studied over a sample user population for assessing adverse effects and product counterfeiting, which were indicated more by negative sentiments and manually created negative sentiment lexicons.

Future work can be the application of lexicon bases on multilingual SA. The data can also be improvised in terms of developing standard lexicons for product protections by involving manufacturers. Deep Learning techniques can also be applied to the lexicon outputs.

References

[1] K. Dégardina, Y. Roggoaand and P. Margot, "Understanding and fighting the counterfeit medicine market," Journal of Pharmaceutical and Biomedical Analysis, vol. 87, pp. 167-175, January 2014.

[2] Priyank Pandey, Manoj Kumar, and Prakhar Srivastava. Classification techniques for big data: A survey. In Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on, pages 3625–3629. IEEE, 2016.

[3] Bhati, Reena. (2019). A Survey on Sentiment Analysis Algorithms and Datasets. Review of Computer Engineering Research. 6. 84-91. 10.18488/journal.76.2019.62.84.91.

[4] Bing Liu. Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press, 2015.

[5] C. Kaiser and F. Bodendorf, "Mining Patient Experiences on Web 2.0 - A Case Study in the Pharmaceutical Industry," in SRII Global Conference (SRII), California, 2012, pp. 139-145.

[6] LexisNexis® Risk Solutions (2012) Survey of law enforcement personnel and their use of social media in investigations.

[7] http://www.lexisnexis.com/investigations

[8] Ducange P, Fazzolari M, Petrocchi M and Vecchio M (2019), "An effective Decision Support System for social media listening based on cross-source sentiment analysis models", Engineering Applications of Artificial Intelligence, Vol. 78, pp. 71-85.

[9] Pang B, Lee L (2008) Opinion mining and sentiment analysis. Found Trends Inf Retr 2:1–135.

[10] Grabner D, Zanker M, Fliedl G, Fuchs M (2012) Classification of customer reviews based on sentiment analysis. In: proceeding of International Conference on Information and Communication Technologies in Tourism, pp 460–470

[11] Eeshita Biswas, K. Vijay-Shanker, Lori Pollock, "Exploring Word Embedding Techniques to Improve Sentiment Analysis of Software Engineering Texts",

[12] Tripathy A, Agrawal A, Rath SK (2016), "Classification of sentiment reviews using n-gram machine learning approach", Expert System Application" 57:117–126. https://doi.org/10.1016/j.eswa.2016.03.028

[13] Kataria S, Mitra P, Bhatia S (2010) , "Utilizing context in generative bayesian models for linked corpus." In: Paper presented at the 24th AAAI conference on artificial intelligence and the 22nd innovative applications of artificial intelligence conference, AAAI-10/IAAI-10, Atlanta, GA, United states, 11–15 July 2010, pp 1340–1345.

[14] HuangEH,SocherR,ManningCD,NgAY(2012),"Improvingwordrepresentationsviaglobalco ntext and multiple word prototypes".In: Paper presented at the 50 the annual meeting of the association for computational linguistics, ACL 2012, Jeju Island, Korea, Republic of, 8–14 July 2012, pp 873–882.

[15] Priyanka Tyagi, Sudeshna Chakraborty, R.C Tripathi, Tanupriya Choudhury, "Literature Review of Sentiment Analysis Techniques for Microblogging Site", Information Systems 50 ISSN: 2005-4289 IJDRBC

[16] Kim IC, Le DX, ThomaGR(2014) Automated method for extracting citation sentences from online biomedical articles using SVM-based text summarization technique. In: Paper presented at the IEEE international conference on systems, man, and cybernetics (SMC)2014 San Diego, CA, USA, 5–8 Oct 2014, pp 1991–1996.

[17] Xu J, Zhu X, Bellmore A (2012) Fast learning for sentiment analysis on bullying. In: Proceeding of International Workshop on Issues of Sentiment Discovery and Opinion Mining

[18] Mittal A, Goel A (2013) Stock prediction using Twitter sentiment analysis. In: Proceeding of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology

[19] Hu Y, Wang F, Kambhampati S (2013) Listening to the crowd: automated analysis of events via aggregated Twitter sentiment. In: Proceeding of International Joint Conference on Artificial Intelligence, pp 2540–2646.

[20] Lampos V, Bie TD, Cristianini N (2010) Flu detector—tracking epidemics on Twitter. In: Proceedings of European Conference on Machine Learning and Knowledge Discovery in Databases, pp 599–602

[21] Sadilek A, Kautz H, Silenzio V (2012) Predicting disease transmission from geo-tagged micro-blog data. In: Proceedings of AAAI Conference on Artificial Intelligence

[22] Glass K, Colbaugh R (2011) Web analytics for security informatics. In: Proceedings of European Intelligence and Security Informatics Conference, pp 214–219.

[23] Taboada M, Brooke J, Tofiloski M, Voll K, Stede M (2011) Lexicon-based methods for sentiment analysis. Comput Linguist J 267–307

[24] Tong RM (2001) An operational system for detecting and tracking opinions in online discussions. In: Working Notes of the SIGIR Workshop on Operational Text Classification, pp 1–6

[25] Turney P, Littman M (2003) Measuring praise and criticism: inference of semantic orientation from association. ACM Transact Inform Syst J 21(4):315–346

[26] B. Chee, K.G. Karahalios, and B. Schatz, "Social Visualization of Health Messages," in 42nd Hawaii International Conference on System Sciences, HICSS '09. , Big Island, 2009, pp. 1-10.

[27] Esuli A, Sebastiani E (2006) SentiWordNet: a publicly available lexical resource for opinion mining. In: Proceedings of language resources and evaluation (LREC)

[28] A, Bhayani R, Huang L (2009) Twitter sentiment classification using distant supervision. Technical Project, Stanford Digital Library Technologies Project

[29] Isa D, Lee LH, Kallimani V, Rajkumar R. Text document pre-processing with the Bayes formula for classification using the support vector machine. Knowledge and Data Engineering, IEEE Transactions on. 2008;20(9):1264–72.