PERCEPTIVE GENETIC ALGORITHM-BASED WOLF INSPIRED CLASSIFIER FOR BIG SENTIMENT DATA ANALYSIS

Ms. D. J. ANITHA MERLIN, Dr. D. VIMAL KUMAR

PhD Part Time Research Scholar, Department of Computer Science,
Nehru Arts and Science College, Coimbatore, India.
merlin.celestino@gmail.com

Associate Professor, Department of Computer Science,
Nehru Arts and Science College, Coimbatore, India.
drvimalcs@gmail.com

ABSTRACT

Big data analytics has a significant impact on real-time purchasing and e-commerce, and it's being used to boost sales and enhance consumer interactions. Customers increasingly rely on online marketing to find the best deals on high-quality goods. Social media activity on both sides of the transaction might reveal information about the customer's purchase experience and opinions about the business. Using Sentimental Analysis (SA), one can figure out how something makes you feel. SA examines people's ideas and feelings about a product. The importance of acquiring insight into customers' sentiments when purchasing things is reduced when SA is weak. Also, the influence of understanding the consumers' impression of a product is reduced. For sentiment analysis in big data, this research proposes a Perceptive Genetic Algorithm-based Wolf Inspired Classifier (PGAWIC). For sentiment analysis in large product review datasets, PGAWIC draws inspiration from wolf foraging behavior. Better optimization for classifying the sentiments is achieved using genetic algorithms. The proposed classifier is evaluated using MATLAB with the metrics namely precision accuracy and f-measure performance measures on a four-product review dataset. According to the results, the proposed classifier PGAWIC is more accurate than the current classifiers in classifying the sentiments.

Keywords: *Sentiment, Classification, Amazon, Wolf, Genetic*

1. INTRODUCTION

Sentiment analysis (SA) is the process of finding and classifying customer views on a specific subject via reviews. SA is a method for assessing a text's polarity. Business intelligence may be derived from software engineers and data scientists who design systems and apply the approach to complete texts, individual words, or parts thereof. SA is a key aspect of the core marketing process because it enables organizations to target their clients more effectively. SA helps businesses understand what their consumers think about their products and organization to improve customer service [1]. Using SA, it is possible to determine the customer's view and feel about the product. Social networking, opinion mining, political research, consumer feedback, and marketing are just a few of the advantages of SA. SA is a significant research area for multiple researchers, and it is used to analyze texts comments. The topic of SA has been resurrected in recent years because of advancements in digital data and machine learning. SA has become an essential business tool because it provides valuable information about consumers' thoughts and helps organizations make better marketing strategies and product development decisions. Individuals utilize SA to get insight into the public's perception of their brands, products, and services to customers. In addition, it provides businesses with actionable information that they can use to improve their operations [2], [3]. As a result, firms can better assess how they stack up against the competition. It's becoming increasingly difficult to keep up with the amount of data being generated today. Data contains a wealth of information that can be mined for new insights into the business [4].

The growing demand for SA is due to its ability to store data cost-effectively and efficiently. On the other hand, organizations are in a precarious position because of the exponential growth of data. Researchers must interpret and use unstructured data (i.e., review data), but doing so is not always straightforward [5]. If SA determines that a specific issue has a favorable connotation, it will approve its publication on social media sites. As a result of the application of SA in real-time problem resolution, AI companies are better able to deliver more personalized service and identify the core cause of difficulties. The annotated text collection will train classifiers using supervised-based learning methods, which will use n-grams, skip-grams and word vectors to represent each text. KNN, Naive Bayes and SVM are the three most prevalent classification methods [6], [7]. Researchers are changing their research direction to Latent Semantic Analysis and Deep Learning which are cutting-edge techniques. The applications of recent technology can be used to derive and interpret human emotions and sensations. Using opinion-mining, emotions like skepticism or sarcasm are modelled. Opinions may be gleaned from the SA to understand it better. All of the difficulties may be solved by manual instruction [8]. The automated method, which is primarily used to resolve issues within the domain itself, does not require any involvement from the user. Source materials such as papers, brief texts, lines from blogs and news articles, among others, may be analyzed using SA. More often than not, one must deal with vast text corpora and "formal language" in these contexts. To begin with, people submit messages from a wide range of technologies, such as tablets and smartphones, and then they build their own unique culture of terminology. As a result, tweets have a considerably higher incidence of spelling mistakes and jargon than communications from other domains [9].

Online consumers can discuss many other issues, unlike forums, newsfeeds, and many other sites that focus on a single topic. Classification is what we're trying to do with SA. According to the presence or absence of polarity (positive or negative), tweets may be classed as either polar (positive or negative) or neutral (neutral). This is similar to how huge texts are categorized [10], [11]. While some authors regard these six "universal" emotions as sentiments, others believe them to be emotions. This study argues that people's feelings can be either excellent or negative depending on their context [12].

Reviews of internet buying encounter many issues, including:

- **(i)** Negative and positive product reviews are far less prevalent than neutral ones.
- **(ii)** In contrast to other areas of SA (such as evaluations of hotels or restaurants), where most of the feedback is either excellent or negative, there are linguistically representing hurdles that come from feature engineering concerns.
- **(iii)** Product evaluations are frequently brief and don't convey much in feeling.

In opinion mining, adaptability is the most important factor in digesting opinionated content or evaluations offered in many formats. Opinion mining has been a prominent focus of real-time applications because of developments in opinion mining. People's comments and reviews are gathered over the internet and sorted into categories or summaries as needed. Product reviews on e-commerce websites are growing in number on a daily basis. No one has the time or patience to read every review. Thus, there is a pressing need to improve the classification of reviews.

The main objective of this research work is to propose a bio-inspired optimization-based classifier to achieve better classification accuracy. In this paper, Perceptive Genetic Algorithm-Based Wolf Inspired Classifier is proposed to classify the reviews present in amazon product review dataset more accurately than the current classifiers.

2. LITERATURE REVIEW

Current section discusses the related literatures that have the limitations of poor classification accuracy in mining the opinions about products that are sold in online.

"Auto Identification of Opinion" [13] is presented in the current study for detecting the sentiments. The binary with the ternary-based data classification is performed using the multi-class classification technique. The complexity and difficulty in the classification technique are expressed using various sentiments, and the performance is predicted to prove its efficiency. **"**Imputation Algorithm" [14] is used in the current study to reduce Twitter data size. The dimensionality of the feature vector is also reduced to a lesser extent. The classifier's performance is also evaluated and compared using other techniques to generate its findings. The robustness of the classifier is also measured using preprocessing method. **"**Domain-Specific Approach" [15] to measure the sentiments expressed for football discussion. Football datasets are used for the analysis of sentiments that are labelled. The sentiment lexicon is created automatically to develop a classifier for recognizing sentiments. An extensive experimental study is carried out to prove the performance of learning algorithms. "Target-Dependent Sentiment Classification" [16] is presented in the current study for building the language models in a pre-trained format. Target-based variations for the BERT base model were implemented in a large margin for classifying the sentiments. TD-BERT based model incorporated with target variations, and experiments are conducted. The efficiency of the study is measured using performance metrics. "Sentiment Similarity Analysis" [17] technique segregates the trust-based relationship among two different individuals. The similarity of sentiments is obtained for extracting the features. Based on the transitivity feature, the trust propagation is measured. The shortest path is used for defining the trust-based relationship among users. The accuracy is examined for generating the results to indicate its similarity.

"Multi-head Self-Attention Transformation networks" [18] is presented in the study for SA in aspect-based. The sentiments are analyzed using a self-attention mechanism along with dynamic target representation. Multi-Head target Specific self-attention mechanism is integrated along with fetching the transformation efficiently. The grammatical features are captured for conducting a series of experiments for better accuracy prediction. "Multi classification technique" [19] analyses consumer review sentiments based on a weighted model. Entity vocabulary for sentiment represents the similarity of sentiments that can express their meaning. Directed weighted links represent sentiment similarity among two nodes where the correlation is measured. Every path connects the links with the similarity of sentiments. The expression of sentiments constitutes a directed weighted model. Classification is also carried out by experimental analysis. "Fine-Tuned Multi-Label SA" [20] is presented in the current study for performing the code-switching of text to analyze the sentiments. Data augmentation is performed for under-sampling and ensemble learning techniques for fetching samples of balanced format. The classifiers are trained with multi-label attributes that predict the classifier's performance. The pre-trained models were selected wherein the fine-tuning tasks were carried out.

"Multi-task Learning framework" [21] is proposed in the present study with a sequence-based labelling problem. The aspects are predicted for every word for review. The cross-View Training method is applied for sequence learning with labelled and unlabeled reviews. Stacked bidirectional recurrent neural layers were used for learning the review representation. Moving-window Attentive Gated-Recurrent-Unit is also used to improve the learning representation and prediction accuracy. "Character-Level SA" [22] is used for handling the inadequate sentiments for overcoming the issue of word-based vector pre-training. The semantically related information is encoded for fetching emotion features to detect the polarity of shorter texts. Twitter datasets were used for evaluation for experimental study. "Sentiment Polarity" [23] is used in the present study for analyzing the public sentiments related to COVID-19. The textual data, statistical validation of the text-based data was incorporated in the framework for segregating the positive and negative sentiments from American Twitter data. Validation of the proposed technique is performed using experimental analysis. "Multidimensional Extra Evidence Mining" [24] is used in the proposed study for analyzing the image-based sentiments. The cross-modal sentimental semantics were enumerated with a soft voting-based sample refinement strategy. The Discriminant Correlation Analysis (DCA) is constructed for mining the cross-modal based semantics using image feature. Experimental analysis is carried out at the end of the study to prove its outperformance. Role of optimization started in evolving in all fields [25]–[35] including sentiment analysis.

"Sentiment Hashtag Embedding (SHE)" [36] is proposed for the current study for hashtag normalization, semantic similarity and topic modelling. The sentiment distribution and semantic representation were distorted using a Convolutional Neural Network-based classifier. The performance of the SHE technique is evaluated using the performance metrics. "Sentiment Prediction (SP)" [37] is implemented in the proposed study for developing the opinion of decision making in the analysis of sentiments. The information is mined for fetching text-based data, whereas the features are fetched to detect the sequence of time from various users. The textual data and public opinion-based text are merged to perform sentiment prediction. The text-based information is predicted for measuring the efficiency of the proposed method. "Density-based Clustering (DBS)" [38] was ensembled together to analyze the social network sentiments. The measurable clusters are fetched and exploited with their variants. Results were generated based on Twitter datasets, and the cluster's performance was generated.

3. PERCEPTIVE GENETIC ALGORITHM-BASED WOLF INSPIRED CLASSIFIER

There are several branches in machine learning (ML) and artificial intelligence (AI), but the artificial neural network (ANN) seems to be the most important one. This neuronal system can identify complicated patterns among processing layers. The ability of ANNs to learn from datasets are among the most widely recognized advantages of ANNs. Based on current knowledge of the brain's neural network, ANNs are cutting-edge information processing technologies. Input, output, and ANN layers (one or more) with units capable of transforming the inputs into something, the pooling layer is a part of a processing component. Many fields, such as advanced engineering, finance, and healthcare, have entirely relied on them for solving the issues that arise randomly. Untrained data may be approximated by ANNs, the essential feature. Variable weights are used to connect different processing units in an ANN design.

A Multi-Layer Perceptron (MLP) must have at least three layers which are (i) input layer, (ii) hidden layer(s), and (iii) output layer. Neurons indicate a computing unit that makes up each layer. It is like biological synapses in a computer system regarding the weights of interconnections. To reduce the inaccuracy, weights are calculated in the first iteration of the ANN performance, and when the process continues, they are tweaked to do so. Contrasting the outputs of the structure to the goal data is the best way of determining the performance mistake. Back-propagates the determined inaccuracy in an opposite direction so that the system may predict more suitable weights.

Eq.(1) represents the calculation of output for an individual neuron.

$$q = \sum g(SN + v) \tag{1}$$

where S indicates the input, N indicates the weight, v indicates the network bias, and the activation function is represented as g, and it is treated as Tangent-Sigmoid ($Tan - Sigm$) function

Eq.(2) defines the $Tan - Sigm$.

$$Tan - Sigm(p) = \prod \frac{2}{1 + h^{-2p}} - 1, \tag{2}$$

This new meta-heuristic biological technique, known as the grey wolf algorithm (GWO), was presented to accomplish an efficient optimization in classifying the sentiments in big product reviews. At the top of the food chain, Grey wolves travel in packs of 6–13 individuals (i.e., wolves). This has served as a model for the GWO's proliferation. Grey wolves in the wild tend to adhere to a rigid hierarchy. There are three females and two males in the alpha pack of wolves (d). The act of Hunting involves significant judgments made by them. To help the alpha wolves in their decision-making process, the beta (v) wolves are the next step below.

For the time being, their only option is to submit and follow the alpha's orders. In some instances, the v wolves are female, and their job is to keep the pack in balance. They are the best contenders to take the place of the alpha. Delta is the term given to the next tier of the flock (g). They are scouts, sentinels, etc., and perform the Hunting. Wolves in the final category, omega (n), are referred to as the weakest. They act as caretakers for the wolves. Some of the flock's battles may be witnessed when omega wolves aren't around. Grey wolves involve in Hunting; it is their primary mode of social interaction.

GW hunting consists of three steps:

(i) identifying, tracking, and approaching the prey
(ii) swarming the prey
(iii) assaulting the prey.

These two-contrasting social behaviors were taken into account while developing GWO's algorithm. Eq.(3) and Eq.(4) describe the mathematical modelling of encircling.

$$\vec{G} = \left|\vec{U}.\overline{P_m(f)} - \overline{P(f)}\right| + f, \tag{3}$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\vec{P}(f+1) = \frac{\overrightarrow{P_m(f)} - \vec{D}.\vec{G}}{f} \quad (4)$$

where $\vec{D}$ and $\vec{U}$ represents the coefficient vectors, $\overrightarrow{P_m}$ indicates the current location of prey, $\vec{P}$ represent the location of the swarm of wolves, $\vec{G}$ denotes the vector, and it is utilized to exact the overall location of wolves, f indicates the time consumed for every iteration. $\vec{D}$ and $\vec{U}$ can be mathematically expressed as Eq.(5) and Eq.(6).

$$\vec{D} = 2\vec{d}.\vec{b}_1 - \vec{d} \quad (5)$$

$$\vec{U} = 2.\vec{b}_2 \quad (6)$$

where $\vec{d}$ indicates the iteration's set of vectors, and it falls in [0,2]. $\vec{b}_1$ and $\vec{b}_2$ are the randomly generated vectors, and its result falls in [0, 1].

The δ uses a hunting strategy and pays attention to it only sometimes. To some extent, it can be said that the three participants (i.e., δ, γ and $\in$) in this kind of hunting task know more about where the prey would be located. These upgraded solutions may be used in all three sites, where the herd will automatically update their registration information and track their progress.

$$\vec{G}_\delta = \left|\vec{U}_1.\vec{P}_\delta - \vec{P}\right| \quad (7)$$

$$\vec{G}_\gamma = \left|\vec{U}_2.\vec{P}_\gamma - \vec{P}\right| \quad (8)$$

$$\vec{G}_\varepsilon = \left|\vec{U}_3.\vec{P}_\varepsilon - \vec{P}\right| \quad (9)$$

$$\vec{P}_1 = \vec{P}_\delta - D_1.\vec{G}_\delta \quad (10)$$

$$\vec{P}_2 = \vec{P}_v - D_2.\vec{G}_v \quad (11)$$

$$\vec{P}_3 = \vec{P}_g - D_3.\vec{G}_g \quad (12)$$

$$\vec{P}_{(f+1)} = \frac{\vec{P}_1 + \vec{P}_2 + \vec{P}_3}{3} \quad (13)$$

Once the prey stops moving, the GWs begin their attack. $|-2d.2d|$ is the range in which the vector of D indicates the random quantity. The $|D| < 1$ signifies that the grey wolves will have to come up with a better solution of attack.

The genetic algorithm (GA) has been extended using genetic programming (GP). GP is another classification of evolutionary algorithms (EAs). EA are looking for ways to solve multivariable functions that are based on the way nature changes. Darwin's idea of evolutionary processes influenced GA and GP, where both two differ in their focus on the development of computer strategies rather than real beings. There are four types of nodes in GP, which are: (i) root node (RoN), (ii) functional node (FuN), (iii) terminal node (TeN), and (iv) random node (RaN).

The technique for GP begins with the development of the first generation, which is generated randomly from sets of functionalities and endpoints that describe the nature of the issue. As a result, the provided fitness requirements for each tree will be examined, and a fitness value will be assigned accordingly. It would then be possible to breed better-fitting populations via cross-breeding and mutation operators to produce the following generation. Another operator just transfers a small section of the population over to a new system depending on the Roulette Wheel pick. The trees with greater fitness values are more easily selected using this selection technique. The crossover procedure will choose one RaN from each of the two trees that have been chosen once they have been selected. All linkages and nodes beneath that node are altered to construct new trees for the following generation.

Any node that has the same number of functions, terminals, and output connections is substituted for a RaN within a randomly chosen tree inside the process known as "mutation of a tree." A FuN may be replaced by another FuN, and a terminal node can be replaced by another terminal node. If three operators, namely (i) reproduction, (ii) crossover, and (iii) mutation used to create a new generation, then this operation will

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

keep repeating itself until the exact number of generations is generated.

4. DATASET AND PERFORMANCE METRICS

This research effort uses four significant product review datasets of Amazon [39]–[41], an online shopping website. The datasets used in this study are from four different domains of the amazon online shopping website: books, DVDs, electronics, and kitchen appliances. Table 1 indicates the count of records accessible in significant product review datasets.

Table 1. Dataset Size

Domain	Number of Records
Books	146294
DVDs	142885
Electronics	88127
Kitchen Appliances	72839

In order to calculate performance measures, several variables must be taken into account. Data mining measures include the following variables:

- ✓ $\boldsymbol{Tr\_Pos}$: Positive emotions that have been correctly detected as such are referred to as "true positives".
- ✓ $\boldsymbol{Tr\_Neg}$: Negative emotions that have been correctly detected as such are referred to as "true negatives".
- ✓ $\boldsymbol{Fl\_Pos}$: Positive emotions that have been incorrectly detected as such are referred to as "false positives".
- ✓ $\boldsymbol{Fl\_Neg}$: Negative emotions that have been incorrectly detected as such are referred to as "false negatives".

The performance metrics used to evaluate the classifiers are:

- ✓ **Precision:** It is the number of correctly identified positive samples relative to the total number of positive samples identified.
- ✓ **Classification Accuracy:** It is the ratio of correct predictions against the total input.
- ✓ **F-Measure:** It represents the measure of classifiers accuracy on the product review dataset.

5. RESULTS AND DISCUSSION

5.1 Precision Analysis

Figure 1 illustrates the percentage of precision attained by the proposed classifier $PGAWIC$ and the existing classifiers SP and DBS. The X-axis is marked with four big product review datasets from Amazon's online shopping site, and Y-Axis is marked with results of the classifier attained for the metric precision. From figure 1, it is clear to understand that $PGAWIC$ achieved better precision than SP and DBS. $PGAWIC$ identifies the sentiments using a modified grey wolf optimization strategy and performs classification using MLP. GA assists to perform classification more accurately, which leads to better precision. SP and DBS perform classification without any optimization strategies simply in a binary manner leading to poor precision. Data values of Figure 1 are provided in Table 1.

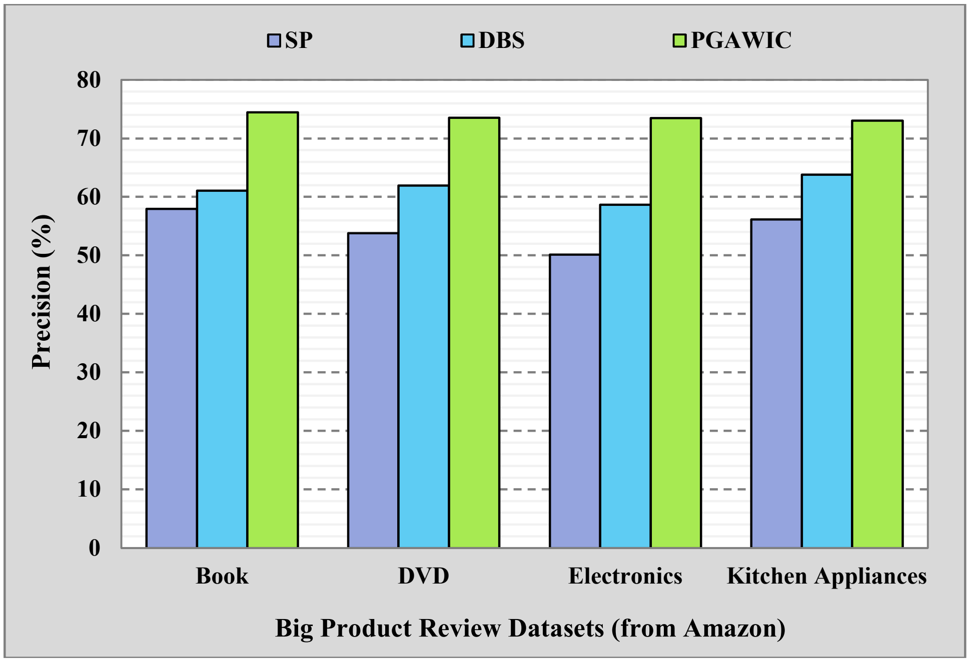

Fig 1. Classifers Vs Precision

Table 1. Precision Result Values

	SP	DBS	PGAWIC
Book	57.9749	61.06515	74.4345
DVD	53.78231	61.91358	73.53457
Electronics	50.10878	58.68488	73.49737
Kitchen Appliances	56.12543	63.81347	73.00705

5.2 Accuracy Analysis

Figure 2 shows the percentage of accuracy achieved by the proposed classifier PGAWIC, as well as the current classifiers SP and DBS, in the classification process. The Y-axis is indicated with the results of the classifier achieved for the metric accuracy, while the X-axis is marked with four large Amazon product review datasets. PGAWIC outperformed SP and DBS in terms of accuracy, as seen in Figure 2. PGAWIC uses different vector strategies for detecting the sentiments in reviews which makes them to achieve better accuracy in dross domain dataset. *SP* and *DBS* intends and perform classification in a binary manner which makes them to not to detect sentiments in a accurate manner leading to increase false positive and false negatives. Data values of Figure 2 are provided in Table 2.

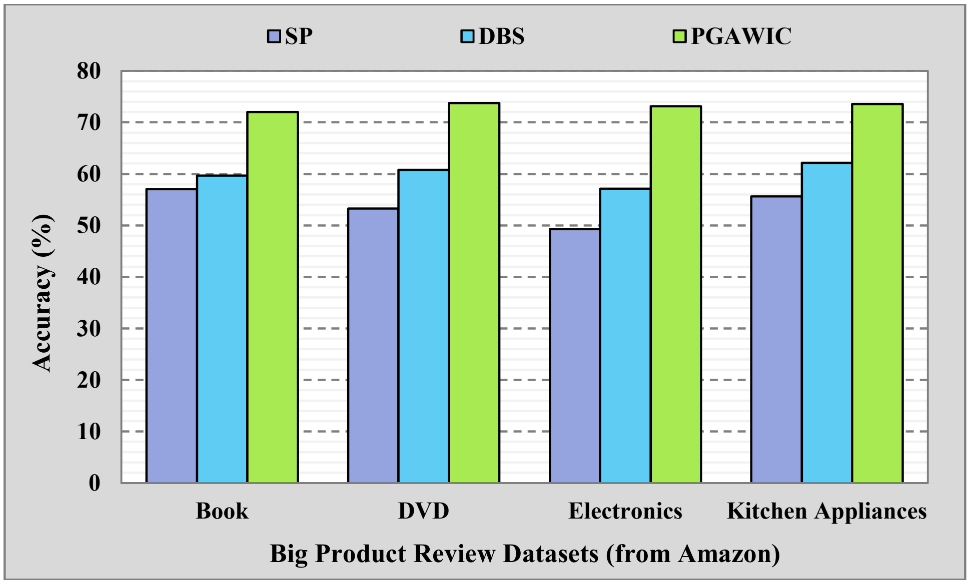

Fig 1. Classifers Vs Accuracy

	SP	DBS	PGAWIC
Book	57.06728	59.63402	72.02824
DVD	53.25192	60.78245	73.75162
Electronics	49.27548	57.10622	73.10245
Kitchen Appliances	55.65288	62.15489	73.56361

6.3 F-Measure Analysis

Figure 3 resembles the percentage of F-Measure achieved by the current classifiers SP and DBS and the proposed classifier PGAWIC. The Y-axis represents the percentage of F-Measure achieved by the classifiers. The X-axis is plotted with four big Amazon product review datasets. *GP* plays a significant role in PGAWIC in classifying the sentiments in big product review datasets. Along with optimization, GP acts as an added advantage in performing the classification in big product review datasets, leading to improved accuracy than the other classifiers. *SP* and *DBS* are designed to perform classification only for a specific dataset. So, the performance of SP and DBS gets down while performing classification with different big product review datasets. Data values of Figure 3 are provided in Table 3.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

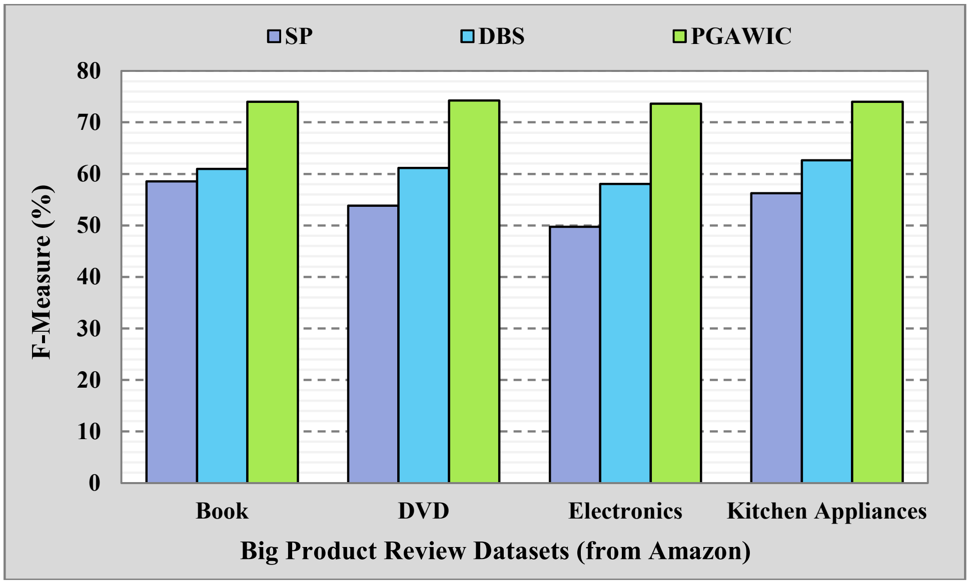

Fig 1. Classifers Vs F-Measure

	SP	DBS	PGAWIC
Book	58.54203	60.97192	73.97098
DVD	53.8351	61.16701	74.27058
Electronics	49.73011	58.0618	73.64934
Kitchen Appliances	56.26828	62.6568	74.03102

7. CONCLUSION

This paper has proposed a classifier, namely PGAWIC (Perceptive Genetic Algorithm-based Wolf Inspired Classifier), to classify sentiments in a large dataset of product reviews. Optimized features in PGAWIC make it easier to classify unstructured product evaluations from various domains. PGAWIC classifies sentiments according on their foraging behaviorpatterns. With the use of genetic programming, PGAWIC has become much more precise. Four benchmark datasets from Amazon are used to evaluate PGAWIC's performance. PGAWIC's average accuracy was determined to be 73.1%, whereas SP and DBS's were 53.8% and 59.9%, respectively. By concentrating on both visual and verbal feelings in product assessments and optimizing, this research might be enhanced in the future to boost accuracy of classification.

REFERENCES

[1] S. Kumar, K. De, and P. P. Roy, "Movie Recommendation System Using Sentiment Analysis From Microblogging Data," *IEEE Trans. Comput. Soc. Syst.*, vol. 7, no. 4, pp. 915–923, 2020, doi: 10.1109/TCSS.2020.2993585.

[2] F. Alattar and K. Shaalan, "A Survey on Opinion Reason Mining and Interpreting Sentiment Variations," *IEEE Access*, vol. 9, pp. 39636–39655, 2021, doi: 10.1109/ACCESS.2021.3063921.

[3] B. Peng, J. Wang, and X. Zhang, "Adversarial learning of sentiment word representations for sentiment analysis," *Inf. Sci. (Ny).*, vol. 541, pp. 426–441, 2020, doi: https://doi.org/10.1016/j.ins.2020.06.044.

[4] I. Fischer and H.-J. Steiger, "Toward automatic evaluation of medical abstracts: The current value of sentiment analysis and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

machine learning for classification of the importance of PubMed abstracts of randomized trials for stroke," *J. Stroke Cerebrovasc. Dis.*, vol. 29, no. 9, p. 105042, 2020, doi: https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.105042.

[5] A. E. de Oliveira Carosia, G. P. Coelho, and A. E. Antunes da Silva, "Investment Strategies Applied to the Brazilian Stock Market: A Methodology Based on Sentiment Analysis with Deep Learning," *Expert Syst. Appl.*, p. 115470, 2021, doi: https://doi.org/10.1016/j.eswa.2021.115470.

[6] P. Sánchez-Núñez, M. J. Cobo, C. D. L. Heras-Pedrosa, J. I. Peláez, and E. Herrera-Viedma, "Opinion Mining, Sentiment Analysis and Emotion Understanding in Advertising: A Bibliometric Analysis," *IEEE Access*, vol. 8, pp. 134563–134576, 2020, doi: 10.1109/ACCESS.2020.3009482.

[7] S. Kausar, X. Huahu, W. Ahmad, M. Y. Shabir, and W. Ahmad, "A Sentiment Polarity Categorization Technique for Online Product Reviews," *IEEE Access*, vol. 8, pp. 3594–3605, 2020, doi: 10.1109/ACCESS.2019.2963020.

[8] R. Vatrapu, R. R. Mukkamala, A. Hussain, and B. Flesch, "Social Set Analysis: A Set Theoretical Approach to Big Data Analytics," *IEEE Access*, vol. 4, pp. 2542–2571, 2016, doi: 10.1109/ACCESS.2016.2559584.

[9] Z. Jianqiang, G. Xiaolin, and Z. Xuejun, "Deep Convolution Neural Networks for Twitter Sentiment Analysis," *IEEE Access*, vol. 6, pp. 23253–23260, 2018, doi: 10.1109/ACCESS.2017.2776930.

[10] Z. Jianqiang and G. Xiaolin, "Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis," *IEEE Access*, vol. 5, pp. 2870–2879, 2017, doi: 10.1109/ACCESS.2017.2672677.

[11] J. Zeng, X. Ma, and K. Zhou, "Enhancing Attention-Based LSTM With Position Context for Aspect-Level Sentiment Classification," *IEEE Access*, vol. 7, pp. 20462–20471, 2019, doi: 10.1109/ACCESS.2019.2893806.

[12] J. D. Kinyua, C. Mutigwe, D. J. Cushing, and M. Poggi, "An analysis of the impact of President Trump's tweets on the DJIA and S&P 500 using machine learning and sentiment analysis," *J. Behav. Exp. Financ.*, vol. 29, p. 100447, 2021, doi: https://doi.org/10.1016/j.jbef.2020.100447.

[13] M. Bouazizi and T. Ohtsuki, "Multi-Class Sentiment Analysis in Twitter: What if Classification is Not the Answer," *IEEE Access*, vol. 6, pp. 64486–64502, 2018, doi: 10.1109/ACCESS.2018.2876674.

[14] M. Bibi *et al.*, "Class Association and Attribute Relevancy Based Imputation Algorithm to Reduce Twitter Data for Optimal Sentiment Analysis," *IEEE Access*, vol. 7, pp. 136535–136544, 2019, doi: 10.1109/ACCESS.2019.2942112.

[15] S. Aloufi and A. E. Saddik, "Sentiment Identification in Football-Specific Tweets," *IEEE Access*, vol. 6, pp. 78609–78621, 2018, doi: 10.1109/ACCESS.2018.2885117.

[16] Z. Gao, A. Feng, X. Song, and X. Wu, "Target-Dependent Sentiment Classification With BERT," *IEEE Access*, vol. 7, pp. 154290–154299, 2019, doi: 10.1109/ACCESS.2019.2946594.

[17] S. Zhang and H. Zhong, "Mining Users Trust From E-Commerce Reviews Based on Sentiment Similarity Analysis," *IEEE Access*, vol. 7, pp. 13523–13535, 2019, doi: 10.1109/ACCESS.2019.2893601.

[18] Y. Lin, C. Wang, H. Song, and Y. Li, "Multi-Head Self-Attention Transformation Networks for Aspect-Based Sentiment Analysis," *IEEE Access*, vol. 9, pp. 8762–8770, 2021, doi: 10.1109/ACCESS.2021.3049294.

[19] S. Zhang, D. Zhang, H. Zhong, and G. Wang, "A Multiclassification Model of Sentiment for E-Commerce Reviews," *IEEE Access*, vol. 8, pp. 189513–189526, 2020, doi: 10.1109/ACCESS.2020.3031588.

[20] T. Tang, X. Tang, and T. Yuan, "Fine-Tuning BERT for Multi-Label Sentiment Analysis in Unbalanced Code-Switching Text," *IEEE Access*, vol. 8, pp. 193248–193256, 2020, doi: 10.1109/ACCESS.2020.3030468.

[21] N. Li, C.-Y. Chow, and J.-D. Zhang, "SEML: A Semi-Supervised Multi-Task Learning Framework for Aspect-Based Sentiment Analysis," *IEEE Access*, vol. 8, pp. 189287–189297, 2020, doi: 10.1109/ACCESS.2020.3031665.

[22] H. Xia, C. Ding, and Y. Liu, "Sentiment Analysis Model Based on Self-Attention and Character-Level Embedding," *IEEE Access*, vol. 8, pp. 184614–184620, 2020, doi: 10.1109/ACCESS.2020.3029694.

[23] J. Samuel *et al.*, "Feeling Positive About Reopening? New Normal Scenarios From COVID-19 US Reopen Sentiment Analytics," *IEEE Access*, vol. 8, pp. 142173–142190, 2020, doi: 10.1109/ACCESS.2020.3013933.

[24] H. Zhang *et al.*, "Multidimensional Extra Evidence Mining for Image Sentiment Analysis," *IEEE Access*, vol. 8, pp. 103619–103634, 2020, doi: 10.1109/ACCESS.2020.2999128.

[25] J. Ramkumar and R. Vadivel, "Improved frog leap inspired protocol (IFLIP) – for routing in cognitive radio ad hoc networks (CRAHN)," *World J. Eng.*, vol. 15, no. 2, pp. 306–311, 2018, doi: 10.1108/WJE-08-2017-0260.

[26] J. Ramkumar and R. Vadivel, "CSIP—cuckoo search inspired protocol for routing in cognitive radio ad hoc networks," in *Advances in Intelligent Systems and Computing*, 2017, vol. 556, pp. 145–153, doi: 10.1007/978-981-10-3874-7_14.

[27] J. Ramkumar and R. Vadivel, "Meticulous elephant herding optimization based protocol for detecting intrusions in cognitive radio ad hoc networks," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 8, pp. 4549–4554, 2020, doi: 10.30534/ijeter/2020/82882020.

[28] R. Vadivel and J. Ramkumar, "QoS-Enabled Improved Cuckoo Search-Inspired Protocol (ICSIP) for IoT-Based Healthcare Applications," pp. 109–121, 2019, doi: 10.4018/978-1-7998-1090-2.ch006.

[29] J. Ramkumar and R. Vadivel, "Performance Modeling of Bio-Inspired Routing Protocols in Cognitive Radio Ad Hoc Network to Reduce End-to-End Delay," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 1, pp. 221–231, 2019, doi: 10.22266/ijies2019.0228.22.

[30] J. Ramkumar and R. Vadivel, "Bee inspired secured protocol for routing in cognitive radio ad hoc networks," *INDIAN J. Sci. Technol.*, vol. 13, no. 30, pp. 3059–3069, 2020, doi: 10.17485/IJST/v13i30.1152.

[31] J. Ramkumar and R. Vadivel, "Improved Wolf prey inspired protocol for routing in cognitive radio Ad Hoc networks," *Int. J. Comput. Networks Appl.*, vol. 7, no. 5, pp. 126–136, 2020, doi: 10.22247/ijcna/2020/202977.

[32] J. Ramkumar, R. Vadivel, and B. Narasimhan, "Constrained Cuckoo Search Optimization Based Protocol for Routing in Cloud Network," *Int. J. Comput. Networks Appl.*, doi: 10.22247/ijcna/2021/210727.

[33] J. Ramkumar and R. Vadivel, "Whale Optimization Routing Protocol for Minimizing Energy Consumption in Cognitive Radio Wireless Sensor Network," *Int. J. Comput. Networks Appl.*, vol. 8, no. 4, doi: 10.22247/ijcna/2021/209711.

[34] J. Ramkumar and R. Vadivel, "Intelligent Fish Swarm Inspired Protocol (IFSIP) For Dynamic Ideal Routing in Cognitive Radio Ad-Hoc Networks," *Int. J. Comput. Digit. Syst.*, vol. 10, no. 1, pp. 1063–1074, 2020, doi: http://dx.doi.org/10.12785/ijcds/100196.

[35] J. Ramkumar and R. Vadivel, "Multi-Adaptive Routing Protocol for Internet of Things based Ad-hoc Networks," *Wirel. Pers. Commun.*, pp. 1–23, Apr. 2021, doi: 10.1007/s11277-021-08495-z.

[36] L. G. Singh, A. Anil, and S. R. Singh, "SHE: Sentiment Hashtag Embedding Through Multitask Learning," *IEEE Trans. Comput. Soc. Syst.*, vol. 7, no. 2, pp. 417–424, 2020, doi: 10.1109/TCSS.2019.2962718.

[37] L. Li, Y. Wu, Y. Zhang, and T. Zhao, "Time+User Dual Attention Based Sentiment Prediction for Multiple Social Network Texts With Time Series," *IEEE Access*, vol. 7, pp. 17644–17653, 2019, doi: 10.1109/ACCESS.2019.2895897.

[38] H. Rehioui and A. Idrissi, "New Clustering Algorithms for Twitter Sentiment Analysis," *IEEE Syst. J.*, vol. 14, no. 1, pp. 530–537, 2020, doi: 10.1109/JSYST.2019.2912759.

[39] X. Li, P. Wu, C. Zou, H. Xie, and F. L. Wang, "Sentiment Lossless Summarization," *Knowledge-Based Syst.*, vol. 227, p. 107170, 2021, doi: https://doi.org/10.1016/j.knosys.2021.107170

.[40] P. R. Kanna and P. Pandiaraja, "An Efficient Sentiment Analysis Approach for Product Review using Turney Algorithm," *Procedia Comput. Sci.*, vol. 165, pp. 356–362, 2019, doi: https://doi.org/10.1016/j.procs.2020.01.038.

[41] W. Zhao *et al.*, "Weakly-Supervised Deep Embedding for Product Review Sentiment Analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 1, pp. 185–197, 2018, doi: 10.1109/TKDE.2017.2756658.