How to Cite:
Malathi, M., Thanamani, A. S., Padmapriya, P., Sharmila, S., Belwin, A. F., & Sherin, A. L. (2022). A new opinion mining method based on M-cuckoo with M-cat and SVM with NB classification algorithm. *International Journal of Health Sciences*, *6*(S2), 12776–12785. https://doi.org/10.53730/ijhs.v6nS2.8353

A new opinion mining method based on M-cuckoo with M-cat and SVM with NB classification algorithm

Malathi M
Research Scholar, Department of Computer Science, Nallamuthu Gounder Mahalingam College, Pollachi, Coimbatore, Tamilnadu, India
Corresponding author email: malathi.mca09@gmail.com

Dr. Antony Selvadoss Thanamani
Head & Associate Professor, Department of Computer Science, Nallamuthu Gounder Mahalingam College, Pollachi, Coimbatore, Tamilnadu, India

Padmapriya. P
Research Scholar, Department of Computer Science, Nallamuthu Gounder Mahalingam College, Pollachi, Coimbatore, Tamilnadu, India

S. Sharmila
Research Scholar, Department of Computer Science, Nallamuthu Gounder Mahalingam College, Pollachi, Coimbatore, Tamilnadu, India

Dr A.Finny Belwin
Research Scholar, Department of Computer Science, Nallamuthu Gounder Mahalingam College, Pollachi, Coimbatore, Tamilnadu, India

Dr A.Linda Sherin
Research Scholar, Department of Computer Science, Nallamuthu Gounder Mahalingam College, Pollachi, Coimbatore, Tamilnadu, India

***Abstract*---**A different point of view can help you make better decisions. People exchange useful knowledge with others thanks to the advent of social media via the Internet. This knowledge is utilised to coordinate, explore, and analyse for better decision making. Opinion Mining is a task in Natural Language Processing that involves defining and extracting information in order to analyse a person's attitude. The basic goal of opinion mining is to determine the polarity of a review at the phrase level, and whether the transmitted opinion is favourable or negative. Metaheuristic algorithms are currently widely utilised in data mining, engineering designs, image processing, and machine learning optimization. These algorithms are typically simple

and adaptable, straightforward to implement, and practical. Some fields remain undeveloped.

***Keywords*---**cuckoo search, opinion mining, swarm intelligence, classification, decision making, positive, negative.

Introduction

Nearly everyone's everyday life are influenced by social media. This allows visitors to express themselves regarding the products available on the e-commerce website. This process is referred to as "interpretation" or "research." It attempts to determine the writer's or speaker's mood or attitude toward the product, which could be good or negative. People's emotions, whether happy or negative, are referred to as feelings. Opinion mining, often known as sentiment analysis, is a computational linguistics and natural language processing (NLP) technique for recognising and extracting subjective information from source material [1]. As a result of the growing interest in e-commerce, which is also a frequent way of expressing and exchanging thoughts, emotions, and ideas, assessments and reviews are becoming more visible.

Literature Review

Sentiment analysis is a large and growing field of study. In recent years, a slew of studies have looked at the topic. Because of the enormous abundance of opinionated data on the Internet, it has quickly gained popularity. Using a variety of machine learning techniques, significant research has been done to investigate sentiment.Nonetheless, the approach's efficiency is frequently harmed by the downward trend in accuracy rates in these investigations. In order to overcome this barrier, Khalid Ait Hadi, Rafik Lasri, et al., presented an efficient methodology for sentiment mining in the context of large data in [7].The data is cleansed using a pretreatment data mining methodology before being used to choose the best features using a flexible greedy algorithm method called Carousel greedy, paired with a bio-inspired metaheuristic algorithm.The Cat Swarm Optimization Based Functional Link Artificial Neural Networks classifier is then used to classify the data, and the approach's performance is examined using experimental findings.

S. K. Lakshmanaprabu, K. Shankaret al., in [9], use a collecting method and a swarm-based improvement system to analyse high-recommendation web-based business sites.Client surveys of goods from web-based company locations with a few features were initially acquired, and then the features were grouped using a fuzzy c-means (FCM) grouping method for a less demanding approach.Also, the Dragonfly Algorithm (DA), which is unique to this study, detects ideal qualities of things on websites, and an advanced ideal feature-based placement technique will be used to determine, in the end, which web-based company webpage is the greatest and easiest to grasp.From the execution, the outcomes demonstrate the greatest exactness rate, that is, 94.56 percent compared with existing methods.

The importance of client ratings and reviews to service providers cannot be overstated. K. Sowmya, K. Monika, and colleagues used logistic regression, Naive Bayes, and SVM algorithms in [6]. These algorithms were applied on a data set containing 1500 customer reviews and ratings.When we look at the aforementioned three algorithms, we can see that logistic regression is 80.82 percent accurate, Naive Bayes is 67.6 percent accurate, and SVM is 80.80 percent accurate. When comparing the accuracy and performance of the above classification methods, logistic regression and SVM come out on top. Producing a summary of opinions based on product features is one of the most difficult problems in sentiment analysis of product reviews. [1] Arti Buche and Dr. M. B. Chandak surveyed and examined several strategies created for major tasks in opinion mining in this study.On the basis of our survey and analysis, we have presented an overview of what goes into designing a software system for opinion mining.

In [3,] Edison Marrese-Taylor, Juan D. Velasquez, and other researchers applied Bing Liu's aspect-based opinion mining approach to the tourist industry. Using this extension, we also present a method for exploring a new way to learn about consumer preferences for tourism products, particularly hotels and restaurants, by analyzing online reviews.An experiment is also carried out to evaluate our proposals, using TripAdvisor hotel and restaurant reviews. The findings revealed that online tourism product reviews include useful information about customer preferences that may be collected using an aspect-based opinion mining methodology.The proposed method was quite successful in determining the sentiment orientation of opinions, with a precision and recall rate of 90%. However, the algorithms were only able to extract 35% of the explicit aspect expressions on average.

Yang and Deb created the Cuckoo Search optimization technique in 2009. It's a tool for figuring out how to solve optimization challenges. It was inspired by the cuckoo bird, which deposits its eggs in the nests of other host birds. The initial basic impetus for the development of a novel optimization method is the cuckoo egg laying and breeding.The efficiency, accuracy, and convergence rate of this optimization algorithm are all improved. Ravi Kiran Varma P evaluated the cuckoo search technique as well as optimization and its difficulties in [10] Venkata Vijaya Geeta. Pentapalli. Various cuckoo search categories are discussed, as well as several cuckoo search applications.

Privacy-preserving data mining (PPDM) is a new way to dealing with privacy concerns that has arisen in the market. PPDM's goal is to develop data-mining techniques while minimizing the danger of misusing the data used to generate those schemes.G.K. Shailaja and C.V. Guru Rao devised a novel PPDM methodology in [4] that consists of two phases: data sanitization and data restoration. Before moving on to the next two phases, the association rules must first be extracted from the database.Key extraction is important in both the sanitization and restoration procedures, and it is chosen best using the Opposition Intensity-based Cuckoo Search Algorithm, which is a modified version of the Cuckoo Search Algorithm.Using the proposed sanitization and restoration techniques, four research issues such as concealment failure rate, information

preservation rate, and erroneous rule generation, as well as degree of alteration, are minimized.

In real-world situations, optimization strategies are crucial. They are utilized in numerous instances when decisions are made based on a random search. However, the user faces a significant issue in selecting the best optimization algorithm. [8] Manar Abdulkareem Al-Abaji discussed the Cuckoo Search Algorithm, which can be used to replace a variety of older procedures.In obtaining the solution relevant to the problem, Cuckoo search employs Levi flight technique based on Egg laying Radius. The CS optimization algorithm improves efficiency, accuracy, and the pace of convergence. Various cuckoo search categories are discussed, as well as several cuckoo search applications.Cuckoo search (CS) is a swarm-intelligence-based algorithm that has seen substantial improvements since its inception in 2009. Because of its simplicity and efficiency in tackling extremely non-linear optimization problems with real-world engineering applications, CS has many advantages.Iztok Fister Jr, Dusan Fister, et al., presents a timely evaluation of all state-of-the-art advances in the previous five years, including theoretical basis and research directions for future development of this strong algorithm in [5].

With the passage of time, the influence of social networking websites on people has grown dramatically. People are now communicating with one another in cyberspace and expressing their feelings in the form of comments on various social networking websites such as Twitter, Facebook, and Google Plus.In the world of video sharing, YouTube is regarded as the king. It is the world's largest video sharing repository, where people gather to give their opinions on videos in the form of comments.Asad Ullah Rafiq Khan, Madiha Khanet al., in [2] employed the Naive Bayes machine learning technique to perform multi-label classification to determine the sentiments of commenter’s for various possibilities. To reduce computing needs, we used the naive assumption that words surrounding keywords relating to a certain option would suffice to grasp the user's feelings.The developed classifier, which was based on the naive assumption, had a slightly poorer performance but required less computer resource.

Proposed Method

Because of the widespread usage of the World Wide Web, the Internet, and the rapid rise of social media, organisations feel compelled to do public opinion research in order to make decisions. However, in order to examine the polarity of viewpoints, precise intelligent knowledge must be filtered. As a result, automated opinion mining and sentiment analysis tools are necessary. Opinion mining technologies are used to gather ratings, ideas, political issues, and brand awareness automatically from the web. Sentiment analysis monitors, analyses, and measures public mood using natural language processing.

Opinion mining and sentiment analysis can be used by corporate intelligence systems to assess consumer impressions of their brand and tailor marketing efforts accordingly. In this study, the mRMR algorithm is merged with the Cuckoo Search method to create a new algorithm known as M-Cuckoo Search. The opinion mining procedure has three steps. Pre-processing is the first phase,

which involves cleaning. The second phase is feature extraction, which is critical for a successful outcome. Third step is classification where the result will be opinion like negative, neutral and positive. The Nave Bayes Classifier technique is utilised for classification in this study.

Data Pre-Processing

Pre-processing is one of the most important aspects of sentiment analysis. It will clean the dataset by lowering its complexity to prepare it for classification. Punctuation removal, letter replacement, text tokenization, and stop signs are all used in this study. Data Pre-Processing includes word removal and word stemming. Punctuation Removal — Most punctuation markers, such as commas and full stops, are ineffective for detecting polarity. Letter Replacement - In addition to English, several sorts of letters exist. As a result, some past study into the default type has replaced the numerous forms of each of these letters.

Tokenization of Text - Tokenization is a task in natural language processing that divides a string of words into semantically significant tokens. Phrase tokenization divides sentences within a document, while word tokenization splits words within a sentence. Sentence tokens are separated by stops, while word tokens are separated by blank spaces. For more intricate structures, such as words that regularly occur together or collocations, high-level tokenization can be used (e.g., New York). Here's an example of how word tokenization improves reading comprehension: Here's an example of how word tokenization improves comprehension: Customer service is exceptional!

Stop Word Removal - Stop word removal is an important stage in NLP text processing. It entails removing high-frequency keywords from a sentence that provide little or no semantic value, such as to, at, for, is, and others. For example, in the statement "Hello, I'm having issues logging in with my new password," stop words like "hello," "I," "am," "with," and "my" should be removed, leaving just the words that help us understand the ticket's topic: "trouble," "logging in," "new," and "password."
Stemming is a pre-process used in TM and NLP to extract the core forms of words. It is noted that the stemming approach is used to retrieve the root of a given word. The current variation in word morphology is a major source of worry for the IR. Duplicity of words is avoided, in addition to having words of the same origin. The stemming procedure is responsible for the improved results. The dictionary and storage space are both reduced because stemming is in charge of regulating the variance. Stemmers are stemming-aware algorithms. Stemming does not necessarily reduce a word to its most basic form; instead, it aims for a close match to the original term.

Feature Extraction

The function is defined first in the feature collection, then in the selection technique, and last, if necessary, in the extraction and reduction procedure. Feature identification involves recognising feature forms such as word frequency, term co-occurrence, part-of-speech, and opinion terms for identification purposes. The challenge of feature subset selection in classification is solved using a variety

of strategies. The main and often utilised procedures are knowledge gain, shared information, document frequency thresholding, x-test (CHI), and term power, to mention a few. The x statistics and data gain produce good results and are more successful in maximising classification outcomes, whereas the document frequency is efficient in terms of scalability and complexity.

Many evolutionary optimization strategies have been successfully investigated and implemented in the past and are now being examined in order to generate a more precise classification and lower the feature set, making it a difficult research subject. The most prevalent evolutionary optimization approaches used for feature selection are genetic algorithms, simulated annealing, gene expression programming, swarm algorithms, and others. In this study, the mRMR algorithm is merged with the Cuckoo Search method to create a new algorithm known as M-Cuckoo Search. The Amazon dataset is preprocessed for better results using the M-Cuckoo Search algorithm.

mRMR Algorithm

Algorithm 1 : mRMR (Minimum Redundancy Maximum Relevance) Algorithm

```
CHOSEN ← Ø
FS ← index list of DS
While FS!= NULL Do
        max eval ← -1
        idx ← NULL
For i ← 0 To length of FS Do
        tmp flist ← {CHOSEN ∪ FS [i]}
        tmp_eval ← mrmr (tmp_flist)    //evaluate given feature subset by mRMR measure
        IF tmp eval > max eval Then
                max eval ← tmp eval
                idx ← i
        End If
End For
Append FS [idx] To CHOSEN
Delete FS [idx] From FS
End While
Return CHOSEN
```

Unlike univariate feature selection methods, which return a subset of features without accounting for redundancy between the selected features, the minimum redundancy and maximum relevance (MRMR) based feature selection algorithm iteratively selects features that are maximally relevant to the prediction task and minimally redundant to the set of features already selected.The following Algorithm 1 shows an mRMR (Minimum Redundancy Maximum Relevance) algorithm

Modified Cuckoo Based Algorithm

The cuckoo search method is one of the approaches used to solve optimization problems. Cuckoo search optimization was proposed by Yang and Deb in 2009. The Cuckoo Search Algorithm was inspired by the natural cuckoo bird species. This method is used to solve both continuous and NP-hard problems. Several academics have tested this method on a variety of benchmark functions and compared it to other algorithms such as Functional Swarm Optimization and Genetic Algorithms, finding that the cuckoo search strategy outperforms PSO and GA. Cuckoo search is a meta-heuristic method named after the cuckoo bird, which belongs to the "Brood parasites" family of birds. It never builds its own nest and instead lays its eggs in another host bird's nest.

Algorithm 2 : Modified Cuckoo Search Algorithm

Input: Population of nests $\mathbf{x_i} = (x_{i1}, \ldots, x_{iD})^T$ for $i = 1 \ldots Np$, MAX_FE.

Output: The best solution xbest and its corresponding value $f_{\min} = \min(f(\mathbf{x}))$.

1: generate initial host nest locations();
2: $eval = 0$;
3: **while** termination condition not meet **do**
4: **for** $i = 1$ **to** Np **do**
5: $\mathbf{x}_i$ = generate_new_solution($\mathbf{x}_i$);
6: f_i = mrmr($\mathbf{x}_i$);
7: $eval = eval + 1$;
8: $j = \lfloor \text{rand}(0, 1) * Np + 1 \rfloor$;
9: **if** $f_i < f_j$ **then**
10: $\mathbf{x}_j = \mathbf{x}_i$; $f_j = f_i$; // replace j-th solution
11: **end if**
12: **if** rand$(0, 1) < p_a$ **then**
13: init_nest($\mathbf{x}_{worst}$);
14: **end if**
15: **if** $f_i < f_{\min}$ **then**
16: $\mathbf{x}_{best} = \mathbf{x}_i$; $f_{\min} = f_i$; // save the local best sol.
17: **end if**
18: **end for**
19: **end while**

The result is a new solution based on the existing one and the tweaking of those features. Each nest has one cuckoo egg in its most basic form, signalling a sequence of solutions in which each nest would have multiple eggs. Cuckoo search has been used successfully in structural engineering to solve scheduling and design optimization problems. Only a few examples include speech restructuring, task scheduling, and global optimization.

Opinion Mining

Optimization is a device modification strategy for making particular features operate more effectively or finding alternate output under specific restrictions by optimizing desired parameters and minimizing unwanted factors involved in the problem as much as possible. Maximizing is attempting to get successful outcomes at the lowest possible cost. Sentiment analysis and opinion mining are

techniques for identifying and extracting subjective data from text documents. It's a type of text analysis that employs machine linguistics and natural language processing to classify and extract feelings or opinions from text automatically (positive, negative, neutral, etc.).

Opinion mining is the process of identifying user opinions from input on a certain topic. Different methodologies are utilized for opinion mining and sentiment analysis. However, the nave Bayes classifier has been included. A prominent supervised machine learning algorithm is the NAIVE BAYES Classifier. It's the probabilistic classifier of Thomas Bayes. This categorization technique implies that the presence or absence of any characteristic in the file is unrelated to the presence or absence of any other feature. The Nave Bayes classifier argues that a file is a bag of words, and that the probability of a word in the file is independent of its location in the file and the presence of another word.The following algorithm 3 shows the Naïve Bayes Classification Algorithm.

Algorithm 3: Naïve Bayes Classification Algorithm

```
TRAINBERNOULLINB(C,D)
V ← EXTRACTVOCABULARY(D)
N ← COUNTDOCS(D)
for each c ∈ C
do N_c ← COUNTDOCSINCLASS(D,c)
prior[C] ← N_c / N
for each t ∈ V
do N_ct ← COUNTDOCSINCLASSCONTAININGTERM(D,c,t)
    cond prob[t][c] ← (N_ct +1) / (N_c + 2)
return V, prior, condprob

APPLYBERNOULLINB(C, V, prior, cond prob, d)
V_d ← EXTRACTTERMSFROMDOC(V,d)
for each c ∈ C
do Score[c] ← log prior[c]
        for each t ∈ V
        do if t ∈ V
        then score[c] += log cond prob[t][c]
         else score[c] += log(1-cond prob[t][c])
return arg max_cec score[c]
```

Result and Analysis

The Amazon appliances dataset is used in this experiment. The experiment was carried out in two stages. Punctuation Removal, Letter Replacement, Text Tokenization, Stop Word Removal, and Word Stemming are used to preprocess the dataset. After preprocessing, the Cuckoo Search algorithm is used to extract the feature set. Then, using the Naive Bayes Classifier algorithm, this result will be classified. The preprocessed dataset is then used to the suggested M-Cuckoo Search method in the second stage.

The output is then generated using the Naive Bayes classifier echnique. Finally, the results of two methods (Cuckoo using NB and M-Cuckoo using NB) are compared, and it is demonstrated that M M-Cuckoo and M-Cat with SVM and NB opinion mining achieved higher accuracy, precision, f-measure, and recall than Cuckoo based NB opinion mining.

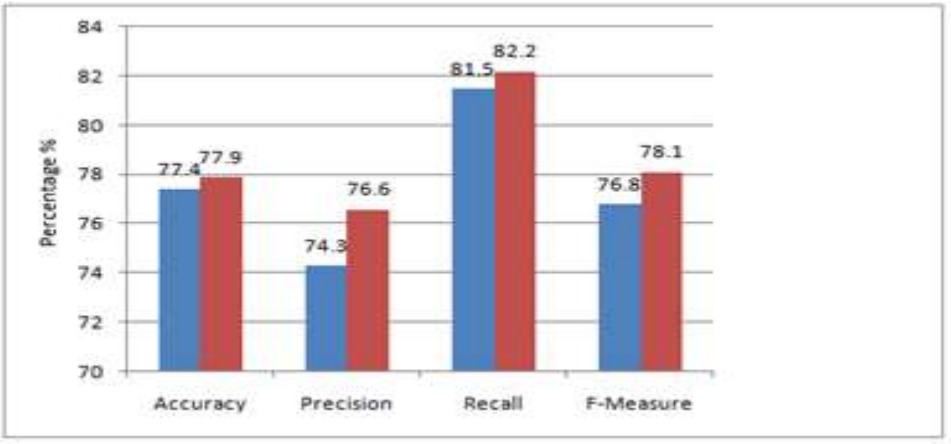

Figure 1: Result Comparison (M-Cuckoo and M-Cat with SVM and NB compared with Cuckoo +NB)

Figure 1. shows the result comparison of Cuckoo + NB based opinion mining and M-Cuckoo and M-Cat with SVM and NB based opinion mining. The result is based on four parameters such as Accuracy, Precision, Recall, F-Measure. M-Cuckoo and M-Cat with SVM and NB provides 0.5% more accuracy than Cuckoo based NB. On contrast of precision value M-Cuckoo and M-Cat with SVM and NB provides 2.3% more than Cuckoo based NB. As well as M-Cuckoo and M-Cat with SVM and NB provides 0.7% more on Recall value. Finally, F-measure of M-Cuckoo scored 1.3% more compared with Cuckoo. It seems proposed algorithm called M-Cuckoo and M-Cat with SVM and NB performs well compared with Cuckoo based NB on opinion mining.

Conclusion

We are now confronted with a vast amount of knowledge provided by individuals all over the world who share their perspectives on numerous topics as a result of the internet growth. Opinion mining is done in this study using a three-step technique. To reduce noise and make the dataset useable, the data is first preprocessed. Punctuation Removal, Letter Replacement, Text Tokenization, Stop Word Removal, and Word Stemming are all used to accomplish this. The second stage is to extract features. CS, which was inspired by cuckoo breeding behavior, is the best search algorithm. The CS algorithm is used in applications such as speaker recognition, job scheduling, wireless sensor networks, industry, flood forecasting, health care, shortest path in distributed systems, document clustering, and image processing.

This feature extraction is based on the M-Cuckoo Search (Modified Cuckoo Search) technique, which we proposed. Finally, we used the Naive Bayes classifier

technique to classify and create the results. As a result of recent improvements in the world of internet services, the value of a product review has increased. In this paper, we focus on the element of saving the user time and effort by recommending the best product. The Naive Bayes Classifier was used to do this, which accurately labels the feedback and averages it to give the product a final ranking. The outcome may be positive, negative, or neutral. When paired with Naive Bayes-based opinion mining, this M-CS technique produces the best results in terms of performance and computing time.

References

1. Arti Buche, Dr. M. B. Chandak, "Opinion Mining and Analysis: A Survey", International Journal on Natural Language Computing (IJNLC) Vol. 2, No.3, June 2013.
2. Asad Ullah Rafiq Khan, Madiha Khan et al., "Naïve Multi-label classification of YouTube comments using comparative opinion mining", Elsevier, Procedia Computer Science 82 (2016), pp.57-64.
3. Edison Marrese-Taylor, Juan D. Vel´asquez et al., "Identifying Customer Preferences about Tourism Products using an Aspect-Based Opinion Mining Approach", 17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems - KES2013, Science Direct.
4. G.K. Shailajaand C.V. Guru Rao, "Opposition Intensity-Based Cuckoo Search Algorithm for Data Privacy Preservation", J. Intell. Syst. 2020; 29(1): 1441–1452.
5. Iztok Fister Jr, Dusan Fister et al., "A comprehensive review of cuckoo search: variants and hybrids", Int. J. Mathematical Modelling and Numerical Optimisation, Vol. 4, No. 4, 2013, pp. 387-409.
6. K. Sowmya, K. Monika et al., "Customer Review Rating Analysis Using Opinion Mining", International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-8 Issue-7, May, 2019.
7. Khalid Ait Hadi, Rafik Lasri et al., "An Efficient Approach for Sentiment Analysis in a Big Data Environment", International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249-8958, Volume-8 Issue-4, April 2019.
8. Manar Abdulkareem Al-Abaji, "A Literature Review of Cuckoo Search Algorithm", Journal of Education and Practice, ISSN 2222-1735 (Paper) ISSN 2222-288X (Online), Vol.11, No.8, 2020.
9. S. K. Lakshmanaprabu, K. Shankar et al., "Ranking Analysis for Online Customer Reviews of Products Using Opinion Mining with Clustering", Hindawi, Complexity, Volume 2018, Article ID 3569351, 9 pages.
10. Venkata Vijaya Geeta. Pentapalli, Ravi Kiran Varma P, "Cuckoo Search Optimization and its Applications: A Review", International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 5, Issue 11, November 2016.