Consumer Feedback Sentiment Classification Improved Via Modified Metaheuristic Optimization Natural Language Processing

Stanislava Kozakijevic[1], Luka Jovanovic[2], Marko Mihajlovic[2], Milos Antonijevic[2], Ninoslava Jankovic[2], Branislav Radomirovic[2], Miodrag Zivkovic[2], Nebojsa Bacanin[2,*], and Ana Stoiljkovic[2]

[1]Faculty of Media and Communications, Singidunum University, Karadjordjeva 65, Belgrade 11000, Serbia

[2]Faculty of Informatics and Computing, Singidunum University, Danijelova 32, Belgrade 11000, Serbia

Abstract: This study investigates the synergy between the virtual and real-world economies through e-commerce, where seller reputation is critical in guiding consumer decisions. As traditional businesses shift towards online retail, user reviews become essential, offering feedback to both sellers and potential buyers. Sentiment analysis through machine learning (ML) techniques presents significant advantages for consumers and retailers alike. This research proposes a novel approach combining bidirectional encoder representations from transformers (BERT) embeddings with an optimized XGBoost classification model to enhance sentiment analysis performance. A modified metaheuristic algorithm, derived from the firefly algorithm (FA), is introduced to optimize the model. Testing on publicly available datasets demonstrates that models optimized by this algorithm achieved a peak accuracy of .881336. Further statistical analyses substantiate these improvements, and SHAP interpretation on the best-performing model identifies key features impacting model predictions, shedding light on factors driving customer sentiment insights.

Keywords: Optimization, Hybridization, Sentiment Analysis, Natural language processing, E-commerce.

1. INTRODUCTION

As the virtual world becomes more tightly knit with the everyday life of people, the conveniences of it are more apparent. One such convenience is online shopping, be it for household items, furniture, groceries, or clothes. Instead of spending hours walking around the stores available in a specific area, customers now often opt for browsing the websites of stores and ordering the items to be delivered. The obvious advantages of this style of shopping as opposed to the traditional is time and energy efficiency, as well as a wider availability of items not constrained by the area an individual is living in or the shops storage space. Some sellers now exist only online, as the lesser costs involved in space renting needed for the shops and the number of employees allow for more competitive prices appealing to many customers.

However, some drawbacks of the online shopping space are evident, especially when it comes to apparel. Namely, it can be hard for customers to judge the quality, fit, color, and sizing based purely on pictures and descriptions the brand offers. Customers often turn to the reviews of others who have purchased the items to solve this issue. Comments left in regards to clothing items can be extremely useful in helping potential customers make an informed decision to buy the right product for their needs. Other helpful comments can offer ways to style the items, what the weather or occasion the items are best suited for, which body types the items fit best, as well as how to best care for the item. The availability of helpful comments can be crucial in making a purchasing decision.

Additionally, comments left by customers provide helpful feedback to the apparel brand. The brand can infer from this revenue how satisfied the customers are, which is the best target group for their items, which features of the items are most popular, if there is any important information they have left out of their description as well as if the way they are marketing their product is true to the experience of customers. The comments can also help the brand pinpoint the dissatisfied customers and reach out to them, as well as adjust their product or product description accordingly. One of the results of globalisation is fast changing trends, especially in the fashion world. The trend cycles shorten as fashion pieces and styles become viral through social media, and the market gets over-saturated, so it can be hard to predict the trends for the upcoming season to both the customers and the brands. By analyzing comments about clothing pieces this prediction can be made more accurately.

The availability of large amounts of comments can be beneficial to both brands and users, however it can also cause clutter and make it harder to pinpoint the most useful comments. Some websites offer ways for customers to mark the most useful comments, however this still leaves comments about specific topics hard to find. Artificial intelligence (AI) has the capacity to quickly and accurately process large datasets, making it a good potential solution. One obstacle for this approach is the fact that AI models rely on numerical data for analysis. Thus comments in their original form,

*Address correspondence to this author at the Faculty of Informatics and Computing, Singidunum University, Danijelova 32, Belgrade 11000, Serbia; E-mail: nbacanin@singidunum.ac.rs

E-ISSN: 2409-9694/24

expressed in natural human language, cannot be subjected to analysis by AI models.

However, there are ways of adapting this kind of data for AI application. Natural language processing (NLP) is a field of study dedicated to the application of AI to human language, taking into account its complexity and tendency to evolve. NLP can take into account not only the language rules and meanings as from a dictionary, but also adapt to hidden meanings, new uses of words and slang words that are abundantly present in online communication. The adjustment of language to AI models is done through the encoding of non-numerical data into numerical data, with regards to the meaning it carries. This encoding can be done via the bidirectional encoder representations from transformers (BERT) [3] method, making it possible for AI to analyze the text from the reviews. The bidirectional nature of it allows for the model to read the text in both directions, making it easier and more accurate to understand the context of the sentence.

Due to the mentioned flexibility and intricacy of human talk, the challenge of analyzing it can be considered a nondeterministic polynomial time (NP-hard) problem. While it is possible to tackle this kind of a problem through trial-and-error methods, it is not efficient. Another method arises to replace the traditional one: metaheuristic optimization algorithms can be applied to the selection of hyperparameters, helping the AI model determine the importance of each variable to the outcome, and value it accordingly. There are many available metaheuristic optimization algorithms, since the field of application is so wide. In addition, the problems they tackle are complex and diverse which, according to the the no free lunch (NFL) theorem [29], makes it impossible to find a single algorithm that outperforms all others when applied to every single problem. Therefore a search for the algorithm best-suited to the task at hand is necessary.

- A proposal for combined BERT and ML classifier powered review sentiment analysis framework
- An introduction of a modified metaheursitic optimizer designed specifically to meet the demands of this study
- A comparative analysis between contemporary optimizer to establish a statistically significant improvement by the proposed modifications
- An game theory based approach is taken to evaluate feature importance of the best model bolstering future data collection

The reminding work is structure as per the following. Section 2 covers preceding works related to this study and discusses the observed literature gap. Section 3 discusses the proposed modified approach in detail and highlights some of the advantages. Sections 4 and Section 5 present and discuss the experimental setup and outcomes respectively. The work is concluded in Section 6 with proposals for future works.

2. RELATED WORKS

As a subfield of AI, NLP allows machines to understand human language, by translating it into language AI can handle: numerical data. By bridging this communication gap, NLP plays a crucial role in assuring more intuitive and seamless interactions with technology. NLP sees regular use in translation tools, virtual assistants, automated notifications and customer service chatbots. Beyond these uses, NLP also has a role in enabling voice issued commands, as well as sentiment analysis which can be used for market research.

The number of attention based algorithms has drastically increased in the past few years, and the Transformer algorithms have been in the limelight. The recurrent neural networks (RNN) [16], and Long-short term memory networks (LSTM) [25] have faced the problem of vanishing gradients. The longer more complex word sequences called for improvement in the NLP field. Research indicated that Generative Pre-trained Transformer (GPT-3) shows great results for non fine-tuned and not too specific tasks as the currently largest language model. BERT on the other hand, used for the Google search algorithm, shows promise on exactly those highly specific challenges [27]. By optimizing the BERT model, this study aims to explore the best overall model for the task of sentiment analysis in online reviews on apparel based products.

Metaheuristic algorithms are often inspired by natural processes and animal behaviors, drawing on the adaptive strategies that occur in nature to solve complex optimization problems. The term metaheuristic refers to a high level strategy made to guide the individual problem solving strategies the models use to the best available solutions. These algorithms are designed to find high-quality solutions to problems that are too difficult or time-consuming to solve using traditional methods. By imitating the behaviors seen in natureâ€”such as cooperation, mutation, or survival of the fittestâ€”these algorithms efficiently search vast solution spaces, often avoiding getting stuck in local optima, which is a common issue with simpler optimization techniques. They provide robust methods for solving real-world problems that require flexibility, adaptability, and efficiency. They are designed to search for optimal or near-optimal solutions in large, complex search spaces where traditional optimization techniques might struggle due to high dimensionality, non-linearity, or other complexities.

Examples of animal inspired algorithms include the bat algorithm (BA) [31], the firefly algorithm (FA) [32], and crayfish optimization algorithm (COA) [7]. Similarly, there are algorithms inspired by the laws of nature, such as genetic algorithm (GA) [18], particle swarm optimization (PSO) [10]. However, algorithms may also come from ideas from other fields, for example variable neighborhood search (VNS) [19]. Botox optimization algorithm (BOA) [6].

The usefulness of these optimizers is evident when examining previous works that utilize them. They have been successfully applied to the field of neuroscience, aiding in training models to detect various abnormalities [28, 11, 12]. In the field of cybersecurity they can be used for fraud detection and similar threats [4, 1, 20]. Predicting prices of various goods can be of help for the economic field [8, 17, 9]. Environmental studies can benefit from aid in predicting the generation and consumption of energy [33, 15, 26].

By creating an optimized ML model for a specific task, the speed and efficiency of solving the task are improved. Streamlining the learning path of automated processes helps in achieving this goal. This, in turn, aids the management of resources available, reduces cost, raises flexibility of automated processes, and raises the overall quality of the product or service through fostering a data-based decision making process. A good way of implementing the optimizers is to aid models in machine maintenance predictions, such as classifying if the sound made by a machine indicates a malfunction [2], or, specifically via NLP checking for logical mistakes in the code of software [21].

2.1. XGBoost

XGBoost (eXtreme Gradient Boosting) is a powerful machine learning algorithm known for its exceptional performance in both classification and regression tasks. By making use of an ensemble of weaker models, XGBoost can effectively improve the accuracy of prediction through a process called boosting. The algorithm iteratively builds and combines these models, with each one correcting the errors of the previous, thus leading to a strong and accurate predictive model. This technique has proven highly effective across various domains, from structured data problems to time series forecasting and beyond.

The success of XGBoost, however, is not only due to its boosting framework, but also due to the careful tuning of its hyperparameters, as previously mentioned. These parameters, such as learning rate, tree depth, and regularization terms, must be optimized to maximize the model's performance. In addition, XGBoost incorporates advanced features like regularization, shrinkage, and efficient handling of missing data, making it particularly adept at finding complex relationships between input features and target variables. The combination of these techniques allows XGBoost to consistently outperform many traditional machine learning algorithms.

When optimizing the XGBoost model, it's essential to focus on three key factors: processing speed, generalization ability, and predictive accuracy. These elements play a pivotal role in determining how well the model performs not only on the training data but also when applied to unseen data. Balancing these aspects is crucial for achieving a model that is both efficient and reliable, aligning with the broader goal of enhancing overall performance.

Best results can be gained by repeatedly adjusting the model. The fitness function of the XGBoost model can be seen in Eq. 1.

$$\mathrm{obj}(\Theta) = L(\Theta) + \Omega(\Theta), \tag{1}$$

in the equation, a total of the loss function and the regularization is presented. The symbol Θ represents the set of XGBoost hyperparameters, where $L(\Theta)$ signifies the loss function, and $\Omega(\Theta)$ denotes the regularization component. The term $\Omega(\Theta)$ is responsible for managing the complexity of the model. The mean squared error (MSE) is utilized as the chosen loss function.

$$\mathrm{L}(\Theta) = \sum_i (x_i - x_i)^2, \tag{2}$$

here, the value being predicted is marked by x_i. The term $\hat{x}_i$ represents the target variable value being predicted in every i iteration.

$$L(\Theta) = \sum_i [\, x_i \, ln\,(1 + e^{-\hat{x}_i}) + (1 - x_i) ln\,(1 + e^{\hat{x}_i})]. \tag{3}$$

The goal of this function is to determine which values are predicted and which are the actual values. If the overall loss function is brought to a minimum, the classification of values is more accurate.

2.2. Bidirectional Encoder Representations from Transformers

While all areas of life benefit from clear communication, spell checking is crucial in any field that requires strict and precise wording. When drafting legal documents, errors like incorrect wording or misplaced clauses can render the document ineffective and lead to significant legal risks. When writing code, errors such as typos or mistakes in the order of operations can not only cause the code to fail but also pose risks to the entire project. In medical documents, a mistake can endanger lives of patients. While basic grammar and spell-check tools can catch some of these mistakes, as legal documents become more complex, these tools often fall short.

Subtler issues, such as ambiguous phrasing or logical inconsistencies in argumentation, are rarely identified without human interference. While interpreters or compilers can help catch minor errors in some programming languages, the increasing complexity of code and databases can make this process both time-consuming and insufficient. More subtle issues, such as improper memory allocation, lack of protection, or logical flaws in the code, are often missed by traditional approaches. Advanced NLP models like BERT, which can understand the context and intent behind the text, could revolutionize both legal drafting and programming, and be applied to many similar cases.

BERT is a transformer-based machine learning model from the field of NLP, which introduces a new approach to the field. Unlike the previous NLP models which used unidirectional encoding, BERT employs a bidirectional approach meaning it takes into account the context of both the text preceding and following the target text. This allows it to capture deeper nuances and meanings within a sentence, making it highly effective for understanding the complexity of human language. In addition, BERT is the first representation model based on fine-tuning that achieves outstanding results across a wide range of sentence-level and token-level tasks. In this it surpasses numerous architectures specifically designed for individual tasks.

BERT models are trained using a masked language model approach, where random words in the text are hidden, and the model's task is to predict the missing words. This method helps the model grasp the structure and patterns of the language it is processing. When trained on a large dataset, BERT demonstrates transfer learning by applying the knowledge gained during training to new, unseen tasks. Once fully trained, BERT can be employed for a variety of applications, including text classification, summarization, and question-answering.

When addressing the issue of harassment in YouTube comments, BERT transforms each comment into a format suitable for machine learning analysis by leveraging its ability to comprehend context and deeper meanings. Using this underlying process, BERT breaks down comments into smaller, meaningful units while preserving the semantic information. By examining both the surrounding text and the comment itself, BERT can detect harassment. This way it can capture not only explicit insults and profanity but also subtler nuances like sarcasm, innuendos, and specific slang often used in YouTube comments.

3. METHODS

3.1. Original Firefly Algorithm

The FA [32] uses firefly light patterns as a model to tackle complicated problems. A single specimen's attractiveness can be gauged by the light's intensity, but the frequency and duration of the light's attraction to mates and prey change. All fireflies are always looking for the brightest firefly; in return, it roams the area at random because it has nothing drawing its interest.

Since light intensity needs to decrease with distance, as well as with respect to the surrounding environment and the objective function, it is important to optimize light intensity. First, a random arrangement of n fireflies is used. A brighter firefly is indicated by an increased function (Eq. 4). An objective function is applied to determine brightness. The symbol x_i in this case indicates the specimen i solution.

After creating and randomly distributing the firefly population, the brightness is determined by assessing an objective function. The following formula presents this process Eq. 4.

$$F_i = f(X_i), \tag{4}$$

The firefly i's location is denoted by X_i, and the value of the objective function is $f(X_i)$.

Each agent moves in the direction of the brighter firefly using an attraction model. The calculation for this movement looks like this:

$$X_i(t+1) = X_i(t) + \beta\, e^{-\gamma r_{ij}^2} (X_j(t) - X_i(t)) + \alpha\, \varepsilon_i(t), \tag{5}$$

where β_0 marks the level of attraction at $r = 0$. The Eq. (5) is predominantly replaced by Eq. (6):

$$\beta(r) = \beta_0 / \left(1 + \gamma \times r^2\right) \tag{6}$$

In iteration t, the momentary orientation of the agent j is denoted by r_{ij}, while the present position of the firefly is indicated by $X_i(t)$. β determines the attraction between i and j based on their distance from one another. The light absorption coefficient is expressed as γ, the degree of unpredictability as α, and the stochastic vector as $\varepsilon_i(t)$. Depending on the fitness function, brightness is changed, going up for higher function values and down for lower ones.

3.2. Modified FA

The baseline FA is well known for it powerful intensification search mechanisms. However, this mechanism can sometimes cause the optimizer to overly focus on a local optima, yielding overall less favorable scores in comparison to better balanced optimizer. This work explores the integration of an adaptive mechanism in to the FA in order to help better diversify the population during the optimization process, helping the FA algorithm overcome the observed premature convergence drawback.

To help derisive the population, a quasi-reflexive approach to generation is adopted. Quasi-reflexive opposites are generated form existing solutions as described in the following:

$$A_z^{qr} = rand(\frac{lb_z + ub_z}{2}, a_z) \quad (7)$$

where lb and ub denote lower and upper bounds of the search space and rad denotes a random value within the given interval. This mechanism enables quasi-reflexive learning (QRL) [22] to occur on a population level. The use of QRL is two fold. Initially, 50% of the population is generated using standard FA procedures, with the later half generated as quasi-reflexives to help boost diversification prior to the start of the optimization.

To help balance between diversification, and prevent the introduced mechanism form hindering intensification in latter iteration an adaptive approach is taken. In every iteration a total of NRS worst performing agents are removed form the populace. Empirically, the NRS value for this study was determined to give the best results when set to $NRS = 2$. Agents are replaced with new solutions generate by an adapted crossover mechanism borrowed form the GA [18]. The crossover mechanism simulates each parameter of an agent as a gene. Two parents are selected, and their genes combined to produce offspring agents. A visual description of the crossover mechanism is presented in Figure **1**.

To bolster diversification, in the initial 50% of iterations, solutions are generated as a crossover between a random agent in the population and the best performing agent. However, in later stages, offspring solutions are generated as offspring of the best performing agent and a quasi-reflexive opposite of the best solution bolstering intensification. The proposed approach is dubbed the Adaptive FA (AFA). The procedural code in presented in Alg **1**.

4. SIMULATION SETUP

Simulations to evaluate the proposed approach and optimize are conducted on a publicly available dataset

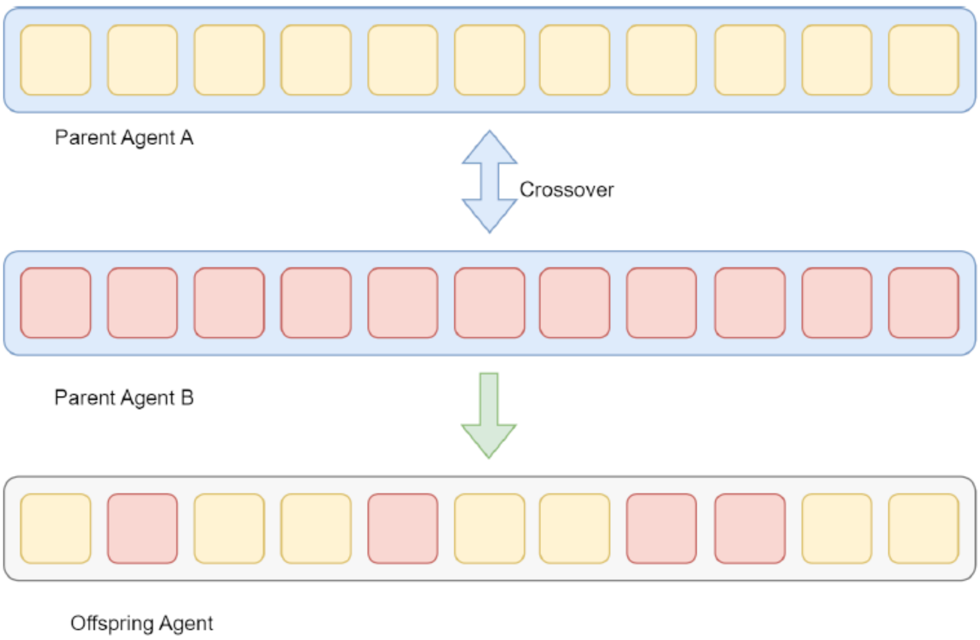

Figure 1: Crossover mechanisms diagram.

Algorithm 1 Introduced AFA procedural code.

Generate $\frac{1}{2}$ of agents for population P
Create quasi-reflexive opposites of current solutions
while $T > t$ **do**
  Assess agent quality based on objective function
  Update agent positioning using FA search
  Remove NRS worst agents form population
  if $t < \frac{T}{2}$ **then**
    for NRS **do**
      Generate new solutions using the best agent and a random agent as parents
  else
    for NRS **do**
      Generate new solutions using the best agent and a quasi reflexive opposite of the best agent

Table 1: XGBoost Parameter Ranges Subjected to Optimization

Parameter	Range
Learning Rate	.1 - .9
Minimum child weight	1 - 10
Subsample	.01 - 1.00
Col sample by tree	.01 - 1.00
Max depth	3 - 10
Gamma	.00 - .80

accessible of Kaggle [1]. Models are train of 70% and evaluated with the reminding 30% of data. A BERT encoding model is utilized to convert input data in to embeddings suitable for use as ML inputs. The BERT model is initialized using the training portion dataset to avoid data leakage. Optimizers are tasked with selecting the most suitable XGBoost parameters form a predefine subset presented in Table **1**. The constraints for each parameter are empirically determined.

Each algorithm is allocated a population size of five agents and allowed eight iterations to locate a solution. Independent implementations of the the FA [32], GA [18], VNS [19], PSO [10], BA [31], BOA [6], COA [7] are included in a comparative analysis. Independent implementation od each optimizer are made in Python in accordance with guidelines presented in each original work. A standard set of classification metrics [5] is tracked during optimization.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{10}$$

$$F1\_score = \frac{2 \cdot Precision \cdot Sensitivity}{Precision + Sensitivity} \tag{11}$$

where the TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative values, respectively. The optimization objective was the Cohen's kappa score, while error rate is logged as the indicator metric.

$$\kappa = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e} \tag{12}$$

where p_o represents an observed value while p_e is the expected outcomes.

$$Error\ Rate = 1 - Accuracy \tag{13}$$

Each optimizer is allocated a population of 10 agents, with a total of 15 iterations to improve population outcomes. All simulations are carried out

[1] https://www.kaggle.com/datasets/nicapotato/womens-ecommerce-clothing-reviews

Table 2: Classifier Objective Function Scores Over 30 Simulations

Algorithm	Best	Worst	Mean	Median	Std	Var
XG-AFA	**.577722**	**.558541**	**.564050**	**.562772**	**.005032**	**2.53E-05**
XG-FA	.565194	.536884	.550468	.549779	.007839	6.15E-05
XG-GA	.562080	.542756	.550133	.548736	.006559	4.30E-05
XG-VNS	.563378	.541527	.552962	.553332	.006010	3.61E-05
XG-PSO	.565016	.547274	.553292	.551617	.005740	3.29E-05
XG-BA	.560253	.538584	.548724	.549546	.006818	4.65E-05
XG-BOA	.570918	.549340	.556233	.552593	.007042	4.96E-05
XG-COA	.566845	.545540	.553991	.553103	.005894	3.47E-05

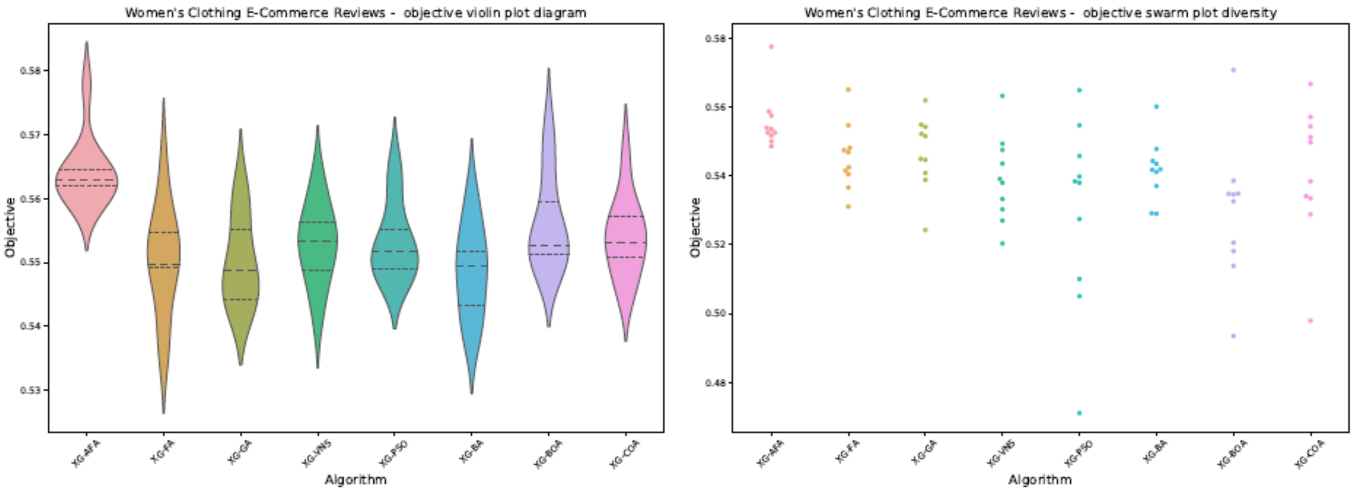

Figure 2: Classifier objective function distribution and swarm diagrams.

though 30 independent runs to account for randomness and provide a solid foundation for statistical evaluations.

5. EXPERIMENTAL OUTCOMES

Simulation outcomes in terms of objective function are provided in Table **2** followed by scores in terms of indicator function in Table **3**. Under both cases the introduced optimizer managed to yield the most favorable scores, outperforming competing optimizers. Models optimized with the introduced AFA attained the best score of .577722 in terms of objective and .118664 in terms of indicator scores. In terms of stability, in terms of objective function the introduced optimizer showcases the hugest rate of stability. Further stability comparisons are provided in Figure **2** in terms of objective and in Figure **3** for indicator scores.

Further details on optimizer abilities to avoid local optima traps and focus convergence towards a global optima are provided in Figure **4** for the objective and in Figure **5** for the indicator function in Figure **5**. In both cases the introduced optimizer locates a favorable outcomes by iteration 1. Other optimizers tend to stagnate prior to locating an equally promising area in the search space. In terms of indicator function, a drop in performance can be observed in several optimizers, including the introduced modified algorithm. This is to be somewhat expected, as the utilized dataset is not fully balanced, metrics that do not account for this, such as the error rate metric used as the indicator function, might not capture the subtle drop in performance should one class be preferred. Nevertheless, in terms of both metrics, the introduced optimizer showcases favorable performance, outperforming the baseline FA as well as other evaluated metaheuristics.

Comparisons between the best performing models considering standard classification metrics are presented in Table **4**. The introduced optimizer showcases the highest rate of accuracy

Table 3: Classifier Indicator Function Scores Over 30 Simulations

Algorithm	Best	Worst	Mean	Median	Std	Var
XG-AFA	**.118664**	**.122055**	**.121749**	**.121970**	.001373	1.88E-06
XG-FA	.119512	.125445	.123479	.124089	.001642	2.70E-06
XG-GA	.120698	.126123	.124258	.124428	.001872	3.51E-06
XG-VNS	.121377	.128835	.123919	.123326	.002227	4.96E-06
XG-PSO	.122563	.126293	.124462	.124936	.002157	4.65E-06
XG-BA	.123241	.124089	.123631	.123411	.001433	2.05E-06
XG-BOA	.119512	.124258	.122495	.122817	.001783	3.18E-06
XG-COA	.121207	.124767	.122817	.122648	**.001298**	**1.69E-06**

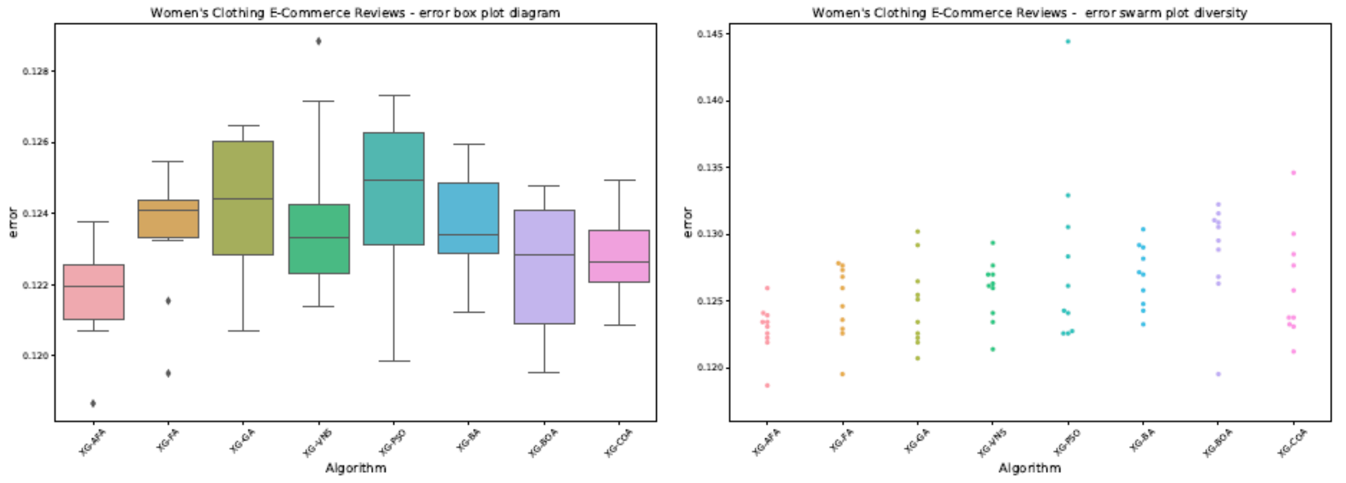

Figure 3: Classifier indicator function distribution and swarm diagrams.

scoring .881336. The model tuned by the AFA also demonstrates high precision for positive review identification. However the baseline FA also showcases good outcomes, showing a high precision and high recall rates for negative sentiment and positive sentiment identification. It is important to consider several metrics when comparing optimizers. Furthermore, only single instance comparisons are often considered insufficient to determine if one optimizer attains statistically significant outcomes in comparison to others. Therefore, further statistic validations are conducted later in this section as well.

Further details for the best performing AFA optimized model are provided in Figure **6** in the form of matrix and PR curves. The parameter selections for each of the best models tuned by optimizers included in the comparative analysis are provided in Table **5**.

5.1. Statistical Validation

Single instance comparisons are often insufficient for providing an in depth comparison between optimizers. The use of statistical comparisons is therefore mandated to ensure a sufficient compression is established [13]. In line with comparison guidelines two approaches can be adopted to support the statistical evaluation, parametric and non-parametric. To facilitate the use on parametric testing, three criteria must be met.

Independence is satisfied in this study, as each of the 30 independent extrusions is conducted using a new random seed. Levene's test [23] is conducted to establish the homoscedasticity criteria, and p-value of $.60$ suggesting conduction fulfillment. To establish the normality criteria the Shapiro-Wilk's test [24] is applied, with p-value outcomes provided in Table **6.** As computed p-values fall below the established $.05$ criteria normality cannot be established as the safe use of parametric testing cannot be justified. These observations are further enforced by the KDE diagrams of objective function outcomes attained during optimizations presented in Figure **7**.

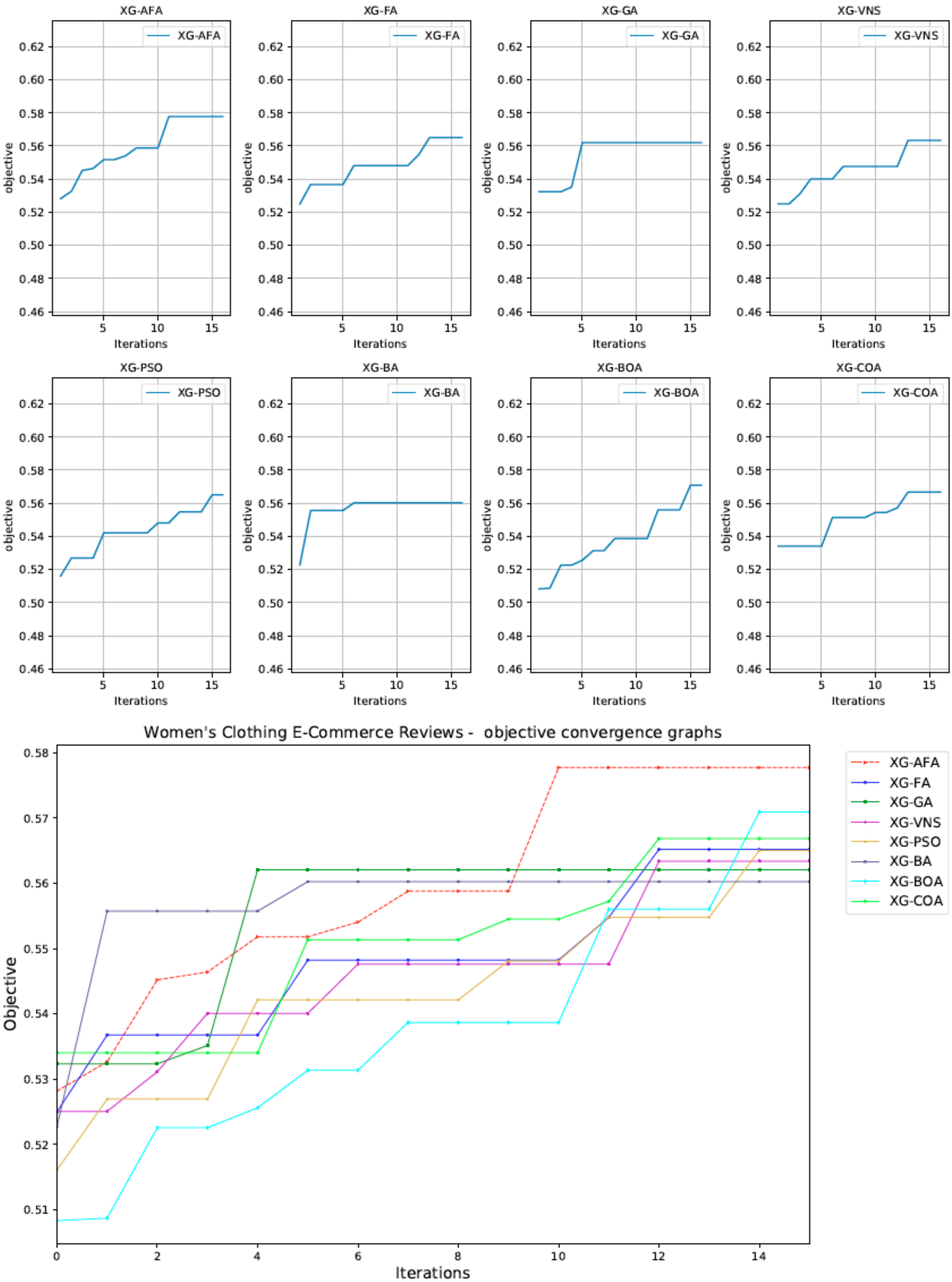

Figure 4: Objective function convergence diagrams.

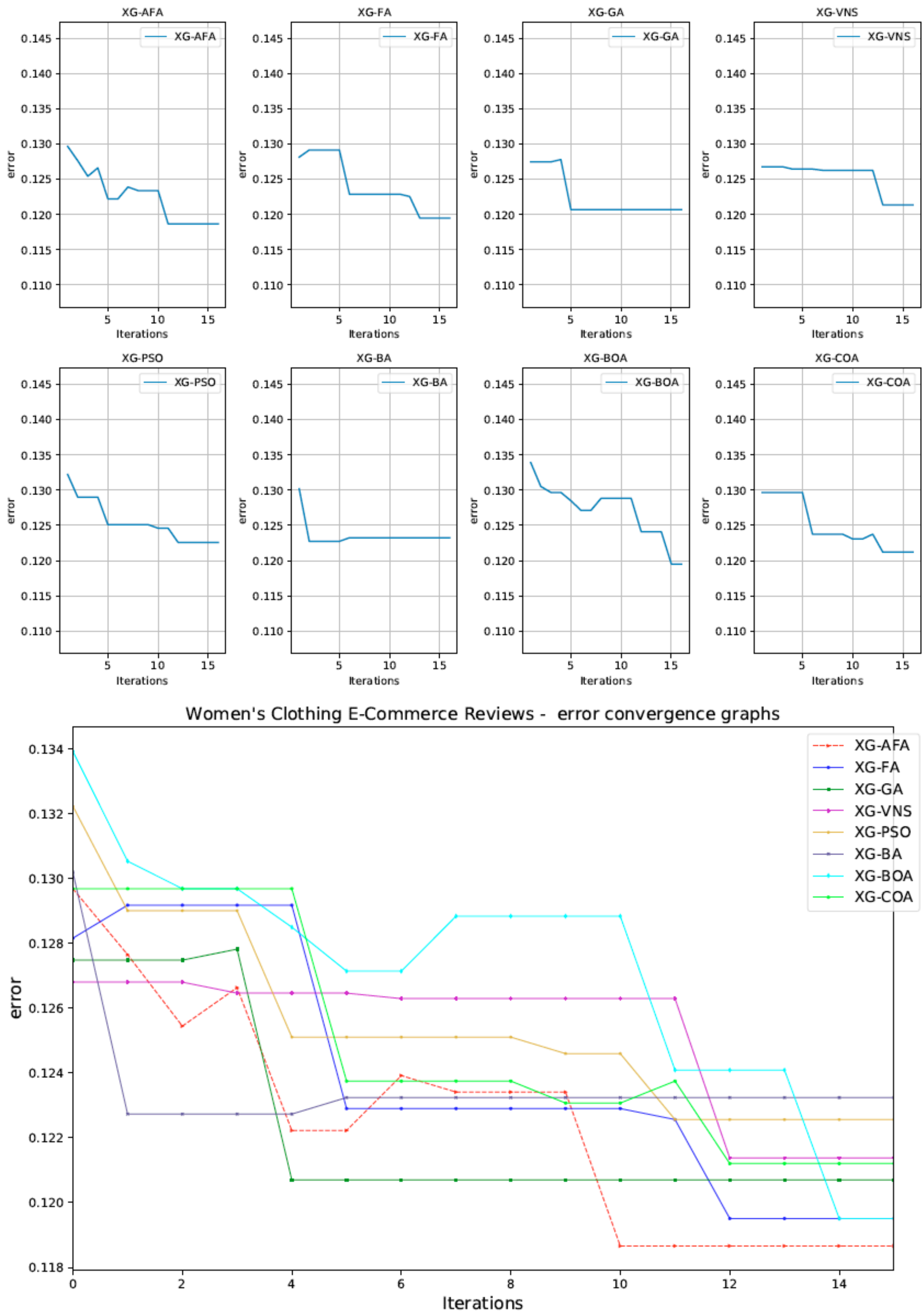

Figure 5: Indicator function convergence diagrams.

Table 4: Best Classifier Comparisons in Terms of Detailed Metrics

Approach	Metric	Negative	Positive	Accuracy	Macro avg.	Weighted avg.
XG-AFA	precision	.702938	**.914257**	**.881336**	.808598	**.875819**
	recall	**.602050**	.943431	**.881336**	**.772741**	**.881336**
	f1-score	**.648594**	**.928615**	**.881336**	**.788605**	**.877681**
XG-FA	precision	**.713457**	.909073	.880488	**.811265**	.873491
	recall	.573159	**.948819**	.880488	.760989	.880488
	f1-score	.635659	.928521	.880488	.782090	.875251
XG-GA	precision	.707710	.908946	.879302	.808328	.872342
	recall	.573159	.947368	.879302	.760264	.879302
	f1-score	.633368	.927760	.879302	.780564	.874211
XG-VNS	precision	.700337	.910343	.878623	.805340	.872144
	recall	.581547	.944675	.878623	.763111	.878623
	f1-score	.635438	.927191	.878623	.781315	.874123
XG-PSO	precision	.688985	.912528	.877437	.800756	.871866
	recall	.594595	.940323	.877437	.767459	.877437
	f1-score	.638319	.926217	.877437	.782268	.873850
XG-BA	precision	.689693	.910969	.876759	.800331	.870720
	recall	.586207	.941359	.876759	.763783	.876759
	f1-score	.633753	.925915	.876759	.779834	.872772
XG-BOA	precision	.705357	.911853	.880488	.808605	.874292
	recall	.589003	.945296	.880488	.767150	.880488
	f1-score	.641950	.928273	.880488	.785112	.876193
XG-COA	precision	.697137	.911841	.878793	.804489	.872788
	recall	.589935	.943017	.878793	.766476	.878793
	f1-score	.639071	.927167	.878793	.783119	.874764
	support	1073	4826			

Table 5: Parameter Selections Made by Each Optimizer for the Respective Best Performing Models

Method	Learning Rate	Min Child W.	Subsample	Col by Tree	Max depth	Gamma
XG-AFA	.867205	2.668964	.899447	.688523	5	.059692
XG-FA	.900000	3.780195	1.000000	.354832	4	.265936
XG-GA	.900000	1.000000	1.000000	.515288	4	.800000
XG-VNS	.900000	1.160622	1.000000	.538159	4	.000000
XG-PSO	.900000	1.735798	.693104	.205471	4	.416622
XG-BA	.900000	4.247616	1.000000	.675263	4	.800000
XG-BOA	.890182	5.404428	.973726	.659285	4	.510130
XG-COA	.900000	6.452072	.860321	.590119	4	.659784

Table 6: Shapiro-Wilk Outcomes Test P-Values for each Optimizer

Model	AFA	FA	GA	VNS	PSO	BA	BOA	COA
XGBoost	.032	.026	.037	.045	.039	.034	.031	.041

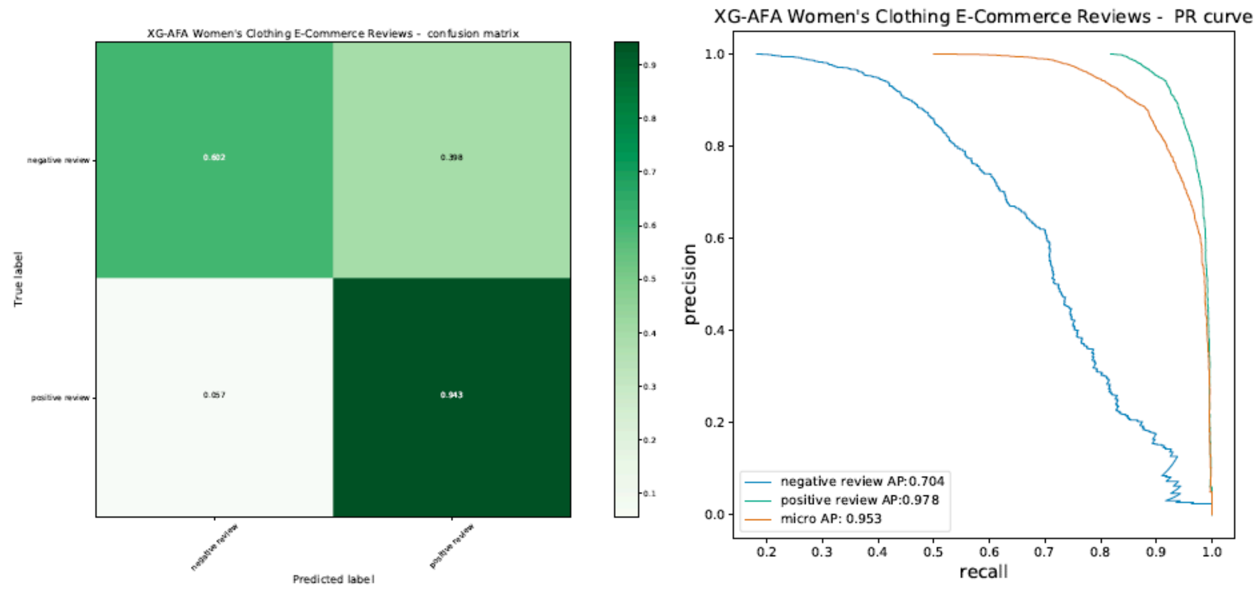

Figure 6: Best performing XG-AFA optimized model confusion matrix and PR plots.

In light of the normality criteria not being fulfilled, non-parametric testing is used. To that end significance is tested using the Wilcoxon signed-rank [30]. Comparisons between the introduced AFA and other algorithms is conducted with outcomes presented in Table **7**. Since the criteria of $\alpha = .05$ is surpassed in all test cases, we can conclude that the AFA optimizer shows statistically significant improvements over the other methods in the comparative analysis.

5.2. Best Model SHAP Interpretation

While optimizer comparisons in terms of statistical terms provide a valuable step in outcome validation, model interpretation provides an important step in detecting unintentional biasses present in model decisions or utilized data. With simpler models interpretation can be straight forward. However, more advanced systems require deeper understanding and more sophisticated evaluation techniques to be

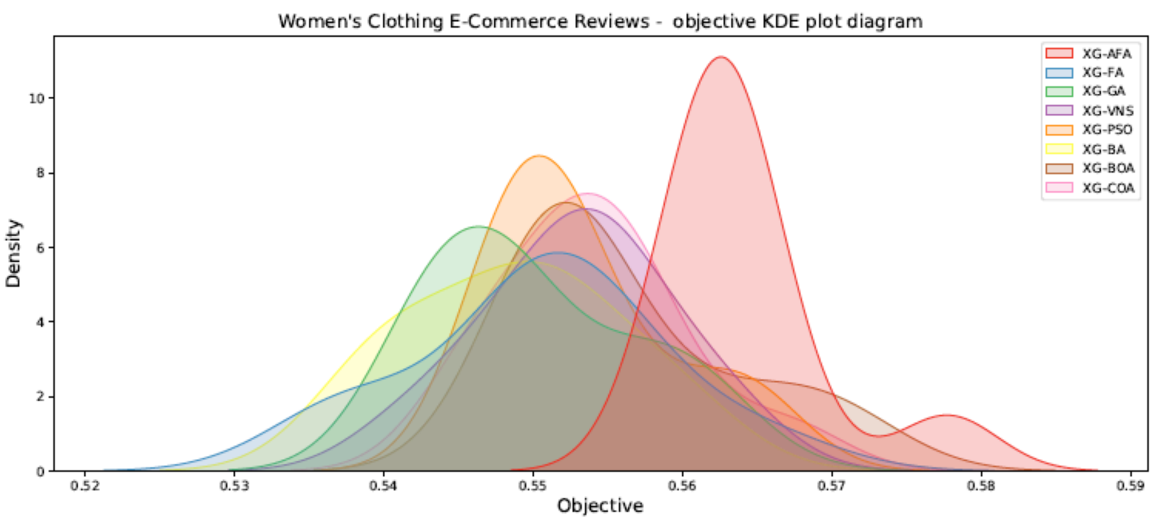

Figure 7: Objective function KDE

Table 7: Wilcoxon Signed-Rank Test Outcomes for Simulations

Model	FA	GA	VNS	PSO	BA	BOA	COA
AFA	.025	.037	.042	.039	.041	.034	.32

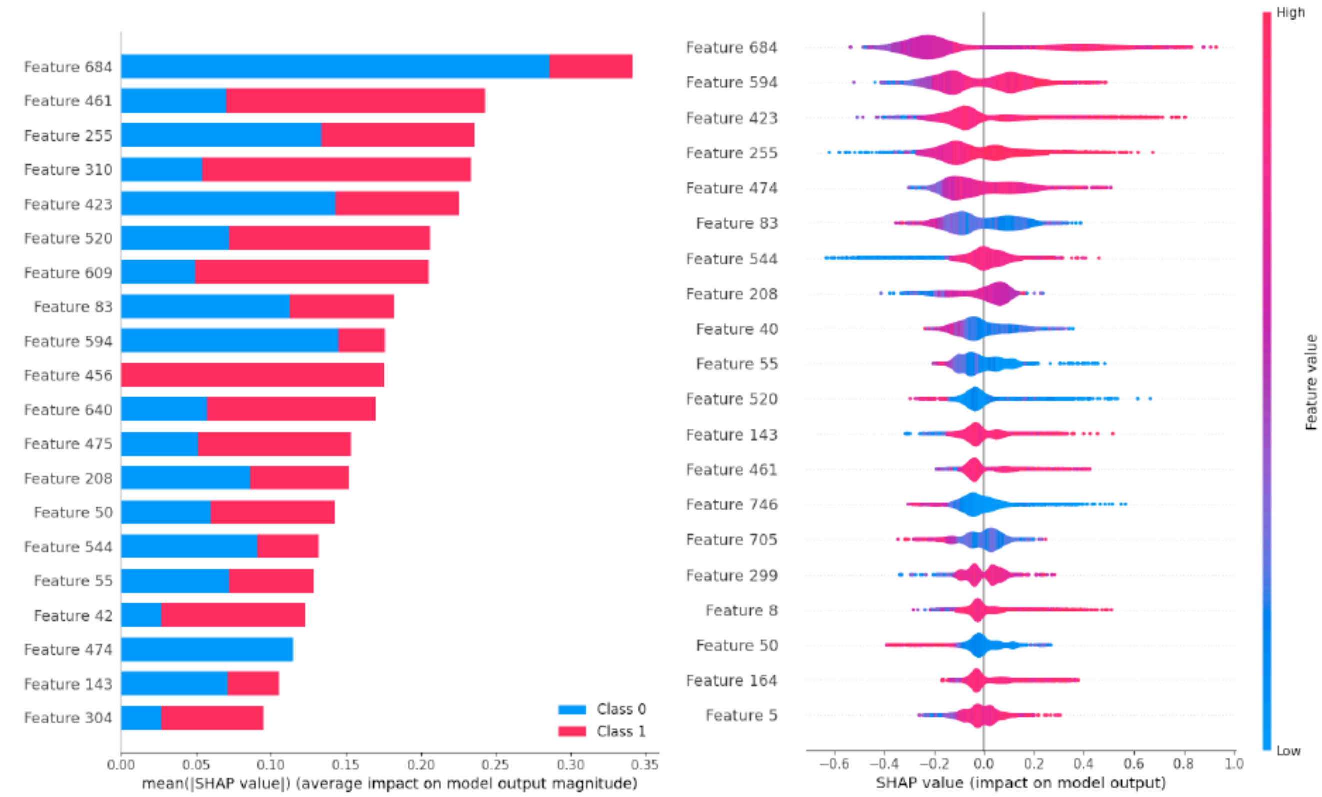

Figure 8: Best profiteering AFA optimized model SHAP interpretation.

employed. An approach with roots in game theory that has become increasingly popular is the use of SHAP [14] values for model feature impact interpretation. This work leverages the tree explainer to efficiently determine feature impacts of encoded BERT values on model decisions. Impacts of each feature are presented in Figure **8**.

6. CONCLUSION

The integration between the virtual and real world is exemplified in the adoption of e commerce. Many traditional business move towards retail towards online sales seller reputation plays an ever increasingly important role. User retrieves provide valuable feedback to the seller as well as other potential customers. Sentiment analysis using ML techniques hold significant potential to benefit both customers as well as retailers. This work explores combining BERT embedding with optimized XGBoost classification models. To handle optimization, a modified metaheursitic algorithm is introduced based on the baseline FA. The proposed approach is evaluated on publicly available data with the best performance attained by models optimized by the proposed optimizer with the best performing models attaining an accuracy score of .881336. Statistical evaluations are conducted to further enforce the observed improvements, and a SHAP interpenetration is conducted on the best model to determine feature importances for the best performing model.

It should be noted, that due to the heavy computational demands of BERT encoding, as well as optimization computational costs, only limited population sizes and relatively short optimization runs are tested in this study. Future works hope to expand on the proposed methodology, evaluation additional dataset and expanding on this study. Additionally, future applications of the proposed optimizer hope to address complex challenges in other fields as well.

REFERENCES

[1] Bacanin, N., Zivkovic, M., Hajdarevic, Z., Janicijevic, S., Dasho, A., Marjanovic, M., Jovanovic, L.: Performance of sine cosine algorithm for ann tuning and training for iot security. In: International Conference on Hybrid Intelligent Systems. pp. 302-312. Springer (2022). https://doi.org/10.1007/978-3-031-27409-1_27

[2] Bozovic, A., Jovanovic, L., Desnica, E., Radovanovic, L., Zivkovic, M., Bacanin, N.: Direct current motor malfunction detection through metastatic optimized audio analysis and classification. In: 2024 International Conference on Inventive Computation Technologies (ICICT). pp. 1471-1478. IEEE (2024). https://doi.org/10.1109/ICICT60155.2024.10544852

[3] Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: North American Chapter of the Association for Computational Linguistics (2019), https://api.semanticscholar.org/CorpusID:52967399

[4] Djuric, M., Jovanovic, L., Zivkovic, M., Bacanin, N., Antonijevic, M., Sarac, M.: The adaboost approach tuned by

sns metaheuristics for fraud detection. In: Proceedings of the International Conference on Paradigms of Computing, Communication and Data Sciences: PCCDS 2022. pp. 115-128. Springer (2023). https://doi.org/10.1007/978-981-19-8742-7_10

[5] Hossin, M., Sulaiman, M.N.: A review on evaluation metrics for data classification evaluations. International journal of data mining & knowledge management process 5(2), 1 (2015). https://doi.org/10.5121/ijdkp.2015.5201

[6] Hubalovska, M., Hubalovsky, S., Trojovsky, P.: Botox optimization algorithm: A new human-based metaheuristic algorithm for solving optimization problems. Biomimetics 9(3) (2024)., https://www.mdpi.com/2313-7673/9/3/137 https://doi.org/10.3390/biomimetics9030137

[7] Jia, H., Rao, H., Wen, C., Mirjalili, S.: Crayfish optimization algorithm. Artificial Intelligence Review 56(2), 1919–1979 (Nov 2023). https://doi.org/10.1007/s10462-023-10567-4

[8] Jovanovic, A., Dogandzic, T., Jovanovic, L., Kumpf, K., Zivkovic, M., Bacanin, N.: Metaheuristic optimized bilstm univariate time series forecasting of gold prices. In: International Conference on Data Science and Applications. pp. 221–235. Springer (2023). https://doi.org/10.1007/978-981-99-7862-5_17

[9] Jovanovic, L., Antonijevic, M., Zivkovic, M., Dobrojevic, M., Salb, M., Strumberger, I., Bacanin, N.: Long short-term memory tuning by enhanced harris hawks optimization algorithm for crude oil price forecasting (2024). https://doi.org/10.1016/bs.adcom.2024.01.002

[10] Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of ICNN'95 - International Conference on Neural Networks. vol. 4, pp. 1942–1948 vol.4 (1995). https://doi.org/10.1109/ICNN.1995.488968

[11] Kozakijevic, S., Salb, M., Elsadai, A., Mani, J., Devi, K., Sharko, A.D., Muthusamy, S.: Seizure detection via time series classification using modified metaheuristic optimized recurrent networks. Theoretical and Applied Computational Intelligence 1(1), 82-94 (2023). https://doi.org/10.31181/taci1120238

[12] Kumpf, K., Kozakijevic, S., Jovanovic, L., Cajic, M., Zivkovic, M., Bacanin, N.: A two layer hybrid approach for parkinsonâ€™s disease detection optimized via modified metaheuristic algorithm. In: International Conference on Innovations and Advances in Cognitive Systems. pp. 205-219. Springer (2024). https://doi.org/10.1007/978-3-031-69197-3_16

[13] LaTorre, A., Molina, D., Osaba, E., Poyatos, J., Del Ser, J., Herrera, F.: A prescription of methodological guidelines for comparing bio-inspired optimization algorithms. Swarm and Evolutionary Computation 67, 100973 (2021)., https://www.sciencedirect.com/science/article/pii/S2210650221001358. https://doi.org/10.1016/j.swevo.2021.100973

[14] Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. p. 4768â€“4777. NIPS'17, Curran Associates Inc., Red Hook, NY, USA (2017)

[15] Markovic, I., Krzanovic, J., Jovanovic, L., Toskovic, A., Bacanin, N., Petrovic, A., Zivkovic, M.: Flood prediction based on recurrent neural network time series classification boosted by modified metaheuristic optimization. In: International Conference on Advances in Data-driven Computing and Intelligent Systems. pp. 289-303. Springer (2023). https://doi.org/10.1007/978-981-99-9518-9_21

[16] Medsker, L., Jain, L.C.: Recurrent neural networks: design and applications. CRC press (1999). https://doi.org/10.1201/9781420049176

[17] Milicevic, M., Jovanovic, L., Bacanin, N., Zivkovic, M., Jovanovic, D., Antonijevic, M., Savanovic, N., Strumberger, I.: Optimizing long short-term memory by improved teacher learning-based optimization for ethereum price forecasting. In: Mobile Computing and Sustainable Informatics: Proceedings of ICMCSI 2023, pp. 125-139. Springer (2023). https://doi.org/10.1007/978-981-99-0835-6_9

[18] Mirjalili, S.: Genetic Algorithm, pp. 43-55. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-319-93025-1_4

[19] MladenoviÄ‡, N., Hansen, P.: Variable neighborhood search. Computers & Operations Research 24(11), 1097-1100 (1997)., https://www.sciencedirect.com/science/article/pii/S0305054897000312 https://doi.org/10.1016/S0305-0548(97)00031-2

[20] Petrovic, A., Antonijevic, M., Strumberger, I., Jovanovic, L., Savanovic, N., Janicijevic, S.: The xgboost approach tuned by tlb metaheuristics for fraud detection. In: Proceedings of the 1st international conference on innovation in information technology and business (ICIITB 2022). vol. 104, p. 219. Springer Nature (2023). https://doi.org/10.2991/978-94-6463-110-4_16

[21] Petrovic, A., Jovanovic, L., Bacanin, N., Antonijevic, M., Savanovic, N., Zivkovic, M., Milovanovic, M., Gajic, V.: Exploring metaheuristic optimized machine learning for software defect detection on natural language and classical datasets. Mathematics 12(18), 2918 (2024). https://doi.org/10.3390/math12182918

[22] Rahnamayan, S., Tizhoosh, H.R., Salama, M.M.: Quasi-oppositional differential evolution. In: 2007 IEEE congress on evolutionary computation. pp. 2229-2236. IEEE (2007). https://doi.org/10.1109/CEC.2007.4424748

[23] Schultz, B.B.: Levene's Test for Relative Variation. Systematic Biology 34(4), 449-456 (12 1985). https://doi.org/10.1093/sysbio/34.4.449

[24] Shapiro, S.S., Francia, R.S.: An approximate analysis of variance test for normality. Journal of the American Statistical Association 67(337), 215-216 (1972)., https://www.tandfonline.com https://doi.org/10.1080/01621459.1972.10481232

[25] Sherstinsky, A.: Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. Physica D: Nonlinear Phenomena 404, 132306 (2020). https://doi.org/10.1016/j.physd.2019.132306

[26] Stankovic, M., Jovanovic, L., Antonijevic, M., Bozovic, A., Bacanin, N., Zivkovic, M.: Univariate individual household energy forecasting by tuned long short-term memory network. In: Inventive Systems and Control: Proceedings of ICISC 2023, pp. 403-417. Springer (2023). https://doi.org/10.1007/978-981-99-1624-5_30

[27] Topal, M.O., Bas, A., van Heerden, I.: Exploring transformers in natural language generation: Gpt, bert, and xlnet. arXiv preprint arXiv:2102.08036 (2021).

[28] Toskovic, A., Kozakijevic, S., Jovanovic, L., Zivkovic, M., Bacanin, N., Antonijevic, M.: Anomaly detection in electroencephalography readings using long short-term memory tuned by modified metaheuristic. In: International Conference on Sustainable and Innovative Solutions for Current Challenges in Engineering & Technology. pp. 133-148. Springer (2023). https://doi.org/10.1007/978-981-97-0327-2_10

[29] Wolpert, D., Macready, W.: No free lunch theorems for optimization. IEEE Transactions on Evolutionary Computation 1(1), 67-82 (1997). https://doi.org/10.1109/4235.585893

[30] Woolson, R.F.: Wilcoxon signed-rank test (2005)., https://onlinelibrary.wiley.com https://doi.org/10.1002/0470011815.b2a15177

[31] Yang, X.S., He, X.: Bat algorithm: literature review and applications. International Journal of Bio-Inspired Computation 5(3), 141-149 (2013), https://www.inderscienceonline.com, pMID: 55093 https://doi.org/10.1504/IJBIC.2013.055093

[32] Yang, X.S., He, X.: Firefly algorithm: recent advances and applications. International journal of swarm intelligence 1(1), 36-50 (2013). https://doi.org/10.1504/IJSI.2013.055801

[33] Zivkovic, M., Jovanovic, L., Pavlov, M., Bacanin, N., Dobrojevic, M., Salb, M.: Optimized recurrent neural networks with attention for wind farm energy generation forecasting. In: 2023 16th International Conference on Advanced Technologies, Systems and Services in Telecommunications (TELSIKS). pp. 187–190. IEEE (2023). https://doi.org/10.1109/TELSIKS57806.2023.10316047

Received on 08-10-2024 Accepted on 31-10-2024 Published on 09-11-2024

https://doi.org/10.31875/2409-9694.2024.11.07