Soft Computing Techniques for Sentiment Analysis and Feature Selection

Ing. Raphael Kwaku Botchway, Ph.D.

Doctoral Thesis Summary

Tomas Bata University in Zlín

Tomas Bata University in Zlín
Faculty of Applied Informatics

Summary of Doctoral Dissertation

Soft Computing Techniques for Sentiment Analysis and Feature Selection

Soft computingové techniky pro analýzu sentimentu a výběr příznaků

Author:	**Ing. Raphael Kwaku Botchway, Ph.D.**
Degree programme:	Engineering Informatics
Degree course:	Engineering Informatics
Supervisor:	assoc. prof. Ing. Zuzana Komínková Oplatková, Ph.D.
External examiners:	prof. RNDr. PaedDr. E. Volná, Ph.D.
	prof. RNDr. Jiří Pospíchal, Dr.Sc.
	prof. RNDr. Ing. Miloš Šeda, CSc.

Zlín, September, 2023

Published by **Tomas Bata University in Zlín** in edition **Doctoral Thesis Summary.**

The publication was issued in the year 2023.

Klíčová slova: úhlová modulace, evoluční výpočetní techniky, výběr příznaků / atributů, optimalizace rojem částic, analýza sentimentu, sociální média

Keywords: angle modulation, evolutionary computation, feature selection, particle swarm optimization, sentiment analysis, social media

Full text of the doctoral thesis is available in the Library of TBU in Zlín.

ISBN 978-80-7678-193-1

ABSTRAKT

Pochopení významu sociálních médií v poslední době přitahuje akademickou pozornost. Jak kdysi řekl významný učenec, sociální média již nejsou pomíjivým pocitem nebo módou. Názory zákazníků vyjádřené na sociálních sítích mohou předávat důležité zprávy, které mohou podniky využít k budování pevných vztahů se zákazníky. S rostoucím využíváním sociálních médií mezi běžnou populací roste i jejich využití v obchodním světě, protože stále více firem využívá sociální média jako efektivní způsob, jak se spojit s mnoha klienty.

Navzdory rychlému přechodu od tradičních k sociálním médiím se firmy v této éře takzvaných velkých dat stále snaží plně porozumět potřebám a obavám svých zákazníků. Navíc schopnost rychle porozumět spotřebitelské komunikaci, aby management mohl reagovat včas a efektivně, zůstává klíčovou výzvou. Dále, velké množství nestrukturovaných dat a nedostatek praktických nástrojů pro analýzu nestrukturovaných dat tuto analýzu komplikuje.

Tato disertační práce představuje stručný přehled aplikací soft computing technik pro analýzu sentimentu a výběr příznaků. Zpočátku autor disertační práce využívá množství dat ze sociálních médií dostupných online k ovlivňování tím, že využívá techniky dolování textu k analýze obsahu generovaného uživateli z příspěvků na sociálních sítích (tweetů) na podporu spotřebitelského rozhodování a marketingové komunikace. Tento nestrukturovaný obsah vytvářený uživateli silně obsahuje slangy, slova s nesprávným pravopisem atd., což představuje výzvu pro výběr funkcí kvůli vágnosti, nepřesnosti a nejednoznačnosti, které jsou v něm obsaženy. V důsledku toho je implementováno řešení založené na metaheuristickém algoritmu Particle Swarm Optimization (PSO) pro optimální výběr textových prvků během analýzy sentimentu, aby se zvýšila přesnost predikce sentimentu.

Druhá část disertační práce kombinuje techniky evolučních výpočtů s úhlovou modulací pro řešení problému výběru příznaků (feature selection). Při hodnocení výkonnosti navržené techniky je použito osmnáct klasických datových sad strojového učení UCI. Zjištění potvrzují konkurenceschopnost a vynikající výkonnost navrženého přístupu při porovnání s jinými metaheuristickými metodami souvisejícími s prací, které jsou k dispozici v literatuře s tématem výběru příznaků. Další statistické testy rovněž potvrzují, že navrhovaná metoda je účinným nástrojem pro řešení binárních optimalizačních problémů v různých oblastech.

Klíčová slova: úhlová modulace, evoluční výpočetní techniky, výběr příznaků / atributů, optimalizace rojem částic, analýza sentimentu, sociální média

ABSTRACT

Understanding the significance of social media has attracted academic attention in recent times. As a prominent scholar once put it, social media is no longer a passing sensation or fad. Customer opinions expressed on social media can convey important messages that businesses can use to build strong relationships with customers. As social media usage among the general population grows, so are its uses in the business world as more businesses turn to social media as a cost-effective and efficient way to connect with many clients.

Despite the quick transition from traditional to social media, firms still struggle to fully comprehend the needs and concerns of their customers in this era of the so-called big data. Moreso, the ability to quickly comprehend consumer communications so that management can respond in a timely and effective manner remains a key challenge. Further, the huge amount of unstructured data and a scarcity of practical tools for analysing this unstructured data makes such analysis more complicated.

This dissertation presents a brief overview of the application of soft computing techniques for sentiment analysis and feature selection. Initially, the author of the dissertation utilizes the abundance of social media data available online as leverage by employing text mining techniques to analyze user-generated content from social media posts (tweets) to support consumer decision-making and marketing communications. This unstructured user-generated content heavily includes slang, misspelt words, etc... thereby presenting a challenge to feature selection due to the vagueness, imprecision, and ambiguity contained therein. Consequently, a metaheuristic-based solution using the Particle Swarm Optimization (PSO) algorithm for optimal text feature selection during sentiment analysis is implemented to enhance sentiment prediction accuracy.

The second segment of the dissertation combines evolutionary computation techniques with angle modulation to solve feature selection problems. Eighteen classical UCI machine learning datasets are employed in evaluating the performance of the proposed technique. The authors' findings confirm the competitive and superior performance of the proposed approach when juxtaposed with other work-related metaheuristics methods available in feature selection literature. Further statistical tests also confirm the proposed method as a potent tool for resolving binary optimization problems across different domains.

Keywords: angle modulation, evolutionary computation, feature selection, particle swarm optimization, sentiment analysis, social media

TABLE OF CONTENTS

1. INTRODUCTION

The functioning of our society has been significantly impacted by the internet [1], [2]. Users can impact other users individually and collectively through social media by sharing their opinions with them. Thus, social media acts as a natural laboratory to analyse new generation netizens' attitudes and access large-scale discussions in real-time [42].

The phrase "Sentiment Analysis" was first coined by Dave et al., in the year 2003 [41]. Sentiment analysis (SA), originally established as a natural language processing (NLP) text classification task, facilitates knowledge extraction from unstructured data abundantly available from social media sources for efficient decision-making. Research conducted by [3], [4] investigated customer sentiment expressed as either positive or negative emotional words in user-generated content (UGC) such as tweets.

Despite the quick transition from traditional to social media, firms still struggle to fully comprehend the needs and concerns of their customers in this era of the so-called big data. Moreso, the ability to quickly comprehend consumer communications so that management can respond in a timely and effective manner remains a key challenge facing businesses. Again, the huge amount of unstructured data and the scarcity of practical tools for analysing this (unstructured) data makes such analysis more complicated.

Soft Computing (SC) techniques are generally organized into five groups namely Machine Learning (ML), Evolutionary Computation (EC), Fuzzy Logic (FL), Probabilistic Reasoning (PR), and Neural Networks (NN) [5]. All SC approaches have one thing in common: they are capable of self-tuning, which means that they may learn from experimental data and approximate it to gain generalization power [11]. The use of SC techniques for sentiment analysis (SA) helps to transform unstructured social media data from sources such as Twitter into a structured data format for business intelligence purposes [6].

Consequently, the first segment of this dissertation utilizes the abundance of social media data available online as leverage to explore the use of soft computing techniques for sentiment analysis. The second part of the dissertation builds on the earlier segment by extending the use of evolutionary computation techniques to solve feature selection (FS) problems. In this phase, a metaheuristic-based solution using the Particle Swarm Optimization (PSO) algorithm for optimal subset text feature selection during sentiment analysis is implemented.

Furthermore, a low-level coevolutionary mixed hybrid approach is adopted to develop a new hybrid metaheuristic algorithm by hybridizing the GWO with the Angle Modulated PSO (AMPSO) [17], [20] for wrapper feature selection. The resulting hybrid version is the Angle Modulated GWOPSO (AMGWOPSO) [18], [23]. The proposed solution is evaluated and analyzed on different publicly available datasets to understand the benefits and drawbacks of the approach to formulate recommendations for future work.

2. THE STATE OF THE ART

This section begins by initiating a discussion on sentiment analysis which is a subfield of natural language processing (NLP). Guided by the aims of the dissertation, special attention is paid to approaches that help to understand and gain valuable insights from the vast amount of unstructured social media data (Twitter) available, as well as the role of soft computing techniques in sentiment analysis and feature selection.

Furthermore, a brief background and state-of-the-art solutions covering core concepts related to social media, NLP, sentiment analysis, evolutionary computation techniques (metaheuristics algorithms), and feature selection are presented in the sub-sections below.

2.1 SENTIMENT ANALYSIS APPROACHES

Sentiment Analysis (SA), also known as opinion mining (OM), is essentially an NLP activity that entails the identification of user sentiment, attitude, emotion, and opinion in natural language text.

According to [7], SA can be classified as machine learning-based or lexicon-based as shown in Fig. 2.1. A detailed description of the lexicon-based approach is detailed in the next section.

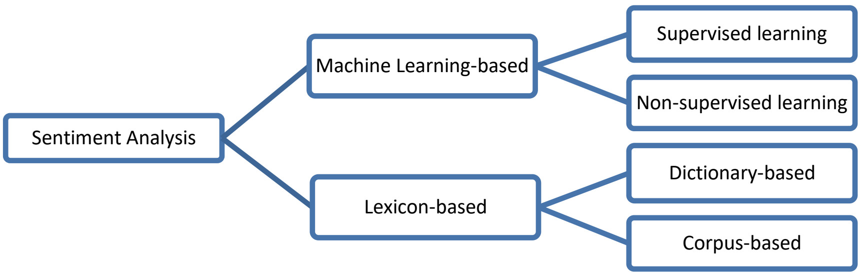

Fig. 2.1: Sentiment analysis approaches

2.1.1 LEXICON-BASED SENTIMENT ANALYSIS

In the unsupervised lexicon-based method, a sentiment lexicon is used to calculate the overall sentiment polarity of a text document based on the sum of the polarities of the individual words embedded in the text [43]. Lexical resources are typically employed in a variety of studies that use lexicon-based methods for unsupervised sentiment classification or analysis.

A standard sentiment lexicon includes attributes for each word, such as the polarity and intensity of the polarity. Hutto and Gilbert introduced VADER [9], a simple sentiment analysis model for social media sentiment analysis. VADER is an acronym for Valence Aware Dictionary sEntiment Reasoner [9]. As a pre-trained sentiment analyzer available via the Natural Language Toolkit (NLTK) library, it examines the lexical features of a document to compute a preliminary sentiment score and then applies five different rules based on general syntactic and grammatical conventions to modify the score.

2.1.2 MACHINE LEARNING TECHNIQUES

Machine Learning (ML) algorithms have demonstrated high performance in several application domains such as user behaviour analytics, NLP, computer vision, etc…[12]. However, the type of problem to be solved determines the choice of the ML algorithm and dataset needed. Usually, supervised ML algorithms utilize both a training set and a test set for sentiment classification.

Machine learning techniques such as Naive Bayes (NB) [15], Maximum Entropy (ME), K-Nearest Neighbour (k-NN), and Support Vector Machines (SVM) [13], [14] among several others are used for sentiment analysis tasks. However, pertinent literature in sentiment analysis suggests that SVM is the widely utilized learning algorithm by the research community [11] followed closely by k-NN and NB.

2.2 METAHEURISTIC ALGORITHMS

Metaheuristic algorithms (MAs) refer to optimization techniques with global search capabilities that provide (near) optimal solutions for optimization problems. These algorithms are simple, flexible, derivative-free, and can avoid local optimal. They begin the optimization process by generating random solutions and exhibit stochastic behaviour.

A key characteristic of MA is its remarkable ability to prevent premature convergence. Metaheuristic algorithms are classified into four groups based on their behaviour namely: evolution-based, swarm intelligence-based, human behaviour-based, and physics-based algorithms. A detailed explanation of these groups can be found in [16] [24]as shown in Fig. 2.2.

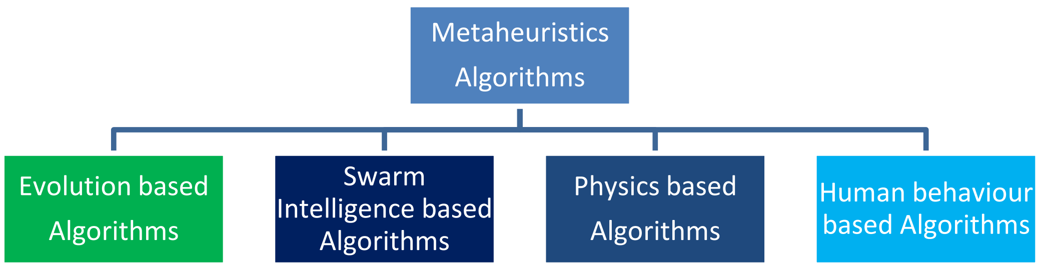

Fig. 2.2: Metaheuristic algorithm categories [24]

Sharma and Kaur report in [40] that metaheuristic techniques have been used more regularly to address various issues in robotics, education, and disease diagnostics (see Fig. 2.3). Nevertheless, sentiment analysis and fraud detection, however, are the least researched applications of metaheuristic algorithms.

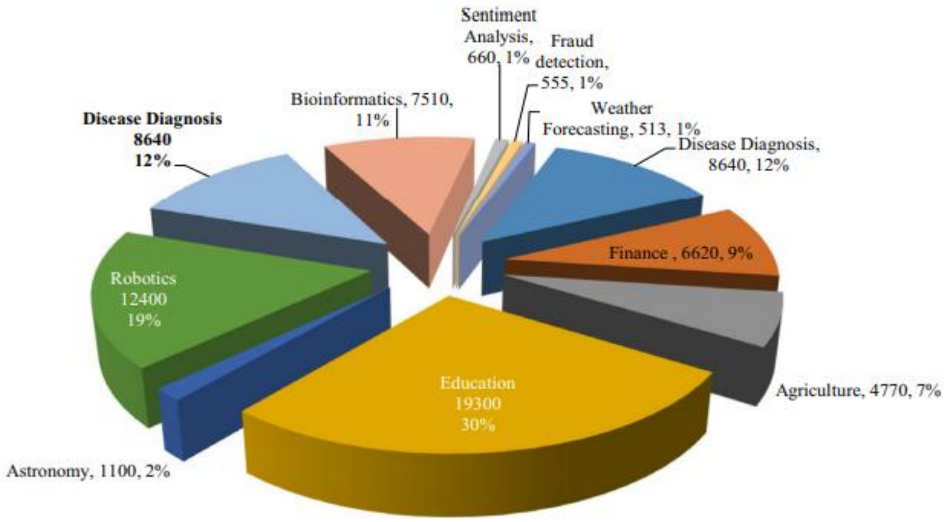

Fig. 2.3: Applications of metaheuristic techniques from various domains [40]

Indeed, it must be mentioned that some of these metaphor-based algorithms (e.g. GWO) have received their fair share of criticism from some researchers compared to some widely accepted ones like the genetic algorithm (GA), particle swarm optimization (PSO), differential evolution (DE), and others [45]. On the other hand, some studies proved that GWO also outperforms these algorithms in particular tasks [25]. Therefore, GWO was used in this thesis together with PSO which has been selected based on its impressive performance in solving numerous feature selection tasks and widespread application in various disciplines and competitions [44].

2.2.1 ANGLE MODULATED PARTICLE SWARM OPTIMIZATION (AMPSO)

The angle modulated PSO (AMPSO) is a discrete optimization method that utilizes the standard particle swarm optimizer to optimize binary problems without making any amendments to the PSO algorithm [19]. The coefficients of the trigonometric function indicated in equation (1) are optimized by the AMPSO using the PSO. This trigonometric function is also dubbed *generating function g* [20].

$$g(x) = \sin[2\pi(x-a)b\cos(2\pi(x-a)c)] + d \tag{1}$$

The coefficients a, b c, and d are defined as follows:

- a: regulates the horizontal shift.
- b: regulates the sine wave's frequency as well as the cos wave's amplitude.
- c: regulates the cos wave's frequency.
- d: regulates the vertical shift.

The *generating function's* shape is controlled by its coefficients due to the way it acts on the function’s displacement and frequency. When searching for optimal coefficients for g, the PSO, the location of a particle i is identified by $x_i = (a, b, c, d)$. Before a binary solution can be produced, the particle's position must be inserted into the function g. Following that, the function g is subsequently sampled periodically at $x = 1,2,3, \ldots, n_b$, where n_b defines the number of binary digits needed. By taking notice of the value of $g(x)$ for every sampling position, a binary solution $B \in \mathbb{B}^{n_b}$ is produced:

$$B_j = \begin{cases} 0 & if\ g(x) \leq 0 \\ 1 & otherwise \end{cases} \tag{2}$$

2.3 HYBRID METAHEURISTIC ALGORITHMS FOR FEATURE SELECTION

Search space exploration and optimal solution exploitation are two essential contradictory principles usually considered when utilizing metaheuristic algorithms. A good option will be to consider a hybrid technique (also referred to as the memetic method) which entails combining two or more metaheuristic algorithms to enhance the performance of the algorithms involved [22].

More recently in [18], a continuous hybrid GWO and PSO (GWOPSO) was proposed. In their work, the authors sought to improve the GWOPSO's ability to efficiently exploit optimal solutions and explore the search space using the PSO and GWO. The proposed Angle Modulated GWOPSO in this thesis generates a binary solution in the search space using equations (1) and (2), as opposed to the binary GWOPSO (BGWOPSO) [23], which employed a sigmoid function to update the locations of the search agents.

In this thesis, the AMGWOPSO is used as a wrapper for feature selection with the k-NN classifier. Several GWO characteristics distinguish it from other metaheuristic algorithms. GWO is easy to use, scalable, simple, and flexible [38], [39]. Apart from the GWO's unique characteristics, it was chosen for use in this thesis by the author because it was originally part of the hybrid metaheuristic algorithm that was modified using angle modulated PSO. Similarly, the k-NN classifier was selected due to its impressive track record in FS literature involving wrapper approaches coupled with its relative simplicity and speed during training and validation [24].

As the literature shows, the feature selection problem has two competing goals. That is, minimizing the number of features by removing redundant features in the feature set while maximizing the classification accuracy. To realize both objectives I adopted the fitness function in [25] shown below (3):

$$fitness\_am = \varphi\gamma(\mathrm{B}) + \mu\frac{|\mathrm{N}_f|}{|\mathrm{N}_\mathrm{T}|} \tag{3}$$

Where $|\mathrm{N}_f|$ depicts the size of attributes in the features subset, $|\mathrm{N}_\mathrm{T}|$, the size of attributes in the given dataset, $\varphi = [0,1]\ and\ \mu = (1 - \varphi)$ are parameters adapted from [25] while $\gamma(B)$ depicts the error rate of the k-NN classifier.

3. OBJECTIVES OF THE THESIS

This thesis employs the relevant text mining and evolutionary computation techniques for sentiment analysis and feature selection tasks.
Consequently, the main objective of this thesis along with the specific steps to be taken is formulated based on the above ideas:

1. Modification of the sentiment analysis pipeline to improve classification:
   a. To determine the best lexicon-based technique based on classification performance and also identify tweet (text) contents most illustrative of positive and negative value user contribution.
2. Design a metaheuristic-based solution for sentiment analysis using the binary PSO (BPSO) given its impressive performance in solving numerous feature selection tasks.
3. Develop a new Angle Modulated-based metaheuristic memetic method for wrapper feature selection. The proposed method utilizes the GWO and AMPSO (AMGWOPSO).
4. Test and evaluate the proposed metaheuristic and memetic method on some selected publicly available benchmark datasets from the University of California, Irvine (UCI) ML repository [21].

The following steps will be pursued to help achieve the aims captured in the thesis:

Literature review and analysis:

1. Of available published literature related to sentiment analysis and metaheuristic algorithms
2. Of current hybrid metaheuristic approaches and examine their potential modifications as it applies to feature selection.

Experiments are conducted:

1. To verify the effectiveness of the proposed approaches for sentiment analysis tasks and metaheuristic algorithm-based feature selection (see results section).
2. To test the newly developed hybrid AMGWOPSO.

Evaluation:

1. Of the selected lexicons for sentiment analysis using tweets crawled from the Twitter handle of an entity of interest (financial institution) to identify insights most illustrative of positive and negative value-user contribution (see section 5.1).
2. Applicability of the proposed metaheuristic technique for feature selection on some selected benchmark datasets (see sections 5.2 and 5.3).

4. WORKFLOW

A brief description of the sentiment lexicons utilized together with the sentiment analysis workflow adopted in this thesis is presented in this section.

4.1 SENTIMENT ANALYSIS WORKFLOW

The diagram in Fig. 4.1 shows the workflow adopted for sentiment analysis in this thesis. Details regarding the implementation of this workflow/pipeline are shown in the case study illustrated in section 5.1.

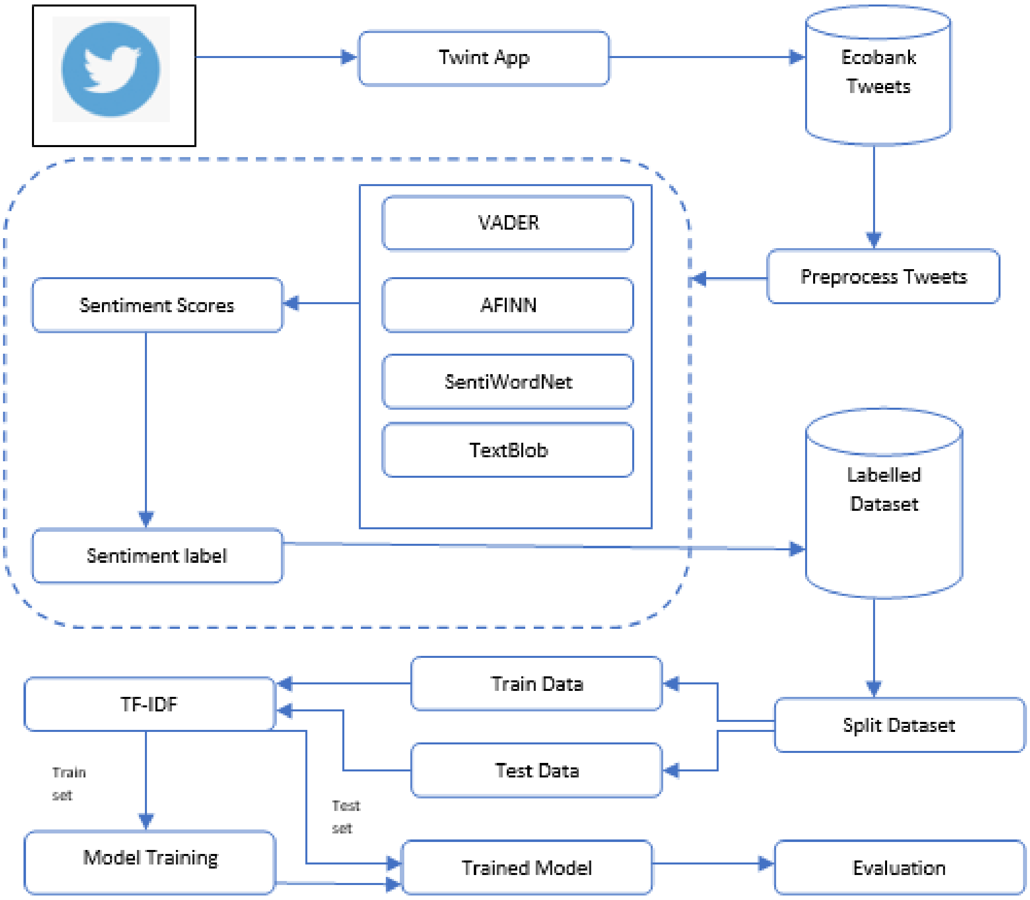

Fig. 4.1: Overview of the sentiment analysis workflow

4.2 FEATURE SELECTION WORKFLOW

The diagram below (see Fig. 4.2) illustrates the working mechanism of feature selection as utilized in this study.

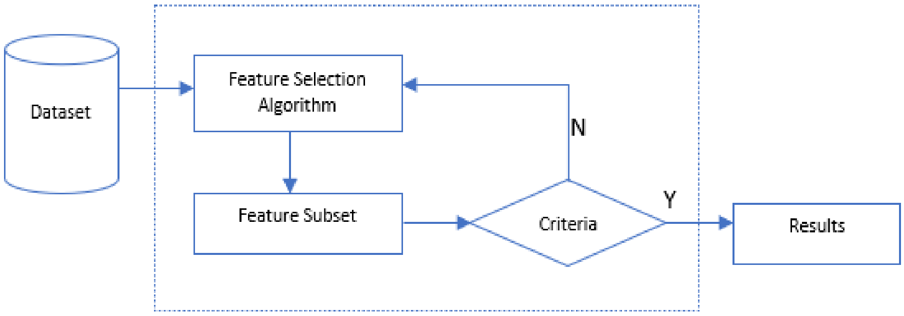

Fig. 4.2: Feature selection process [24]

5. RESULTS

In this section, the author of this thesis presents and discusses the results obtained in chronological order during the development of methods as part of the thesis preparation process. Specifically, some functional examples using the lexicon-based sentiment analysis approach and the metaheuristic algorithms for feature selection are provided in the following sub-sections.

5.1 DEDUCTIONS FROM A SUB-SAHARAN AFRICAN BANK: A SENTIMENT ANALYSIS APPROACH

The upsurge in social media websites has no doubt triggered a huge source of data for mining interesting expressions on a variety of subjects. In Sub-Saharan Africa, financial institutions are making the needed technological investments required to remain competitive in today's challenging global business environment. The methodology described in Chapter 4 was adopted and implemented in three phases as shown in Fig. 5.1 below. A detailed description of these processes can be found in [29]. The dataset used for the study consists of 7,730 English tweets collected between January 1, 2015, and December 31, 2019, using the "*ecobank*" keyword.

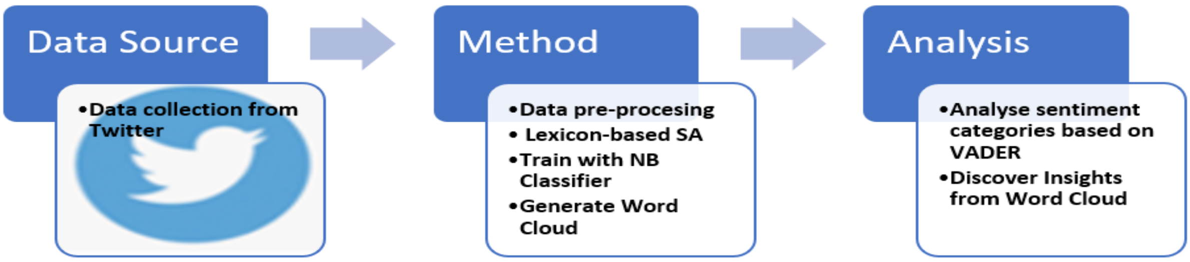

Fig. 5.1: Three-phase methodology deployed [29]

The study sought to answer three research questions (RQ):

- Which lexicon produces the best output that describes the opinions of bank clients on Twitter?

- What insights can be gained from the expressions garnered from the Ecobank group's Twitter handle?
- Since Fintech has come to stay, what lessons can the bank learn from social media to improve its service delivery?

Table *5.1* and Table 5.2 shows the classification of the Tweets by the various lexicons and some sample tweets respectively.

Table 5.1 Classification of Tweets by various lexicons

LEXICON	CLASSIFICATION		
	POSITIVE	NEGATIVE	NEUTRAL
VADER [9]	4317	1233	2180
AFINN [27]	4038	1567	2125
SentiWordNet[33]	2982	3148	1600
TextBlob [28]	3400	1159	3171

In response to research question 1, a sample of 773 tweets constituting 10% of the tweets is drawn from the dataset and manually labelled using two independent annotators from A.I.Lab research group at Faculty of Applied Informatics, Tomas Bata University in Zlín. Out of 773 tweets that were hand-labelled, 525 were classified as positive with 46 and 202 categorized as negative and neutral respectively. The inter-annotator agreement (Cohen’s Kappa) [30] reached 0.75 which indicates a reasonable agreement level. The author of this dissertation trained the labelled tweets of each lexicon using the Naive Bayes classifier to obtain the respective classification scores.

From Table 5.3, the VADER lexicon outperforms all the other lexicons used in this study. This indeed confirms what is stated in the literature [9] since it is accustomed to sentiments expressed on social media. Fig. 5.2 comprises 1,233 tweets expressing negative sentiments visualized as a word cloud after removing stop words. Hence, one can deduce that the tweets contained discussions regarding frustrations (pain points see Fig. 5.3) customers face concerning the usage of the e-banking solutions provided by the bank.

Table 5.2 Sample Tweets from the Ecobank Tweets dataset [29]

Tweet 1

Tweet 2

Table 5.3 Accuracy scores per lexicon using the NB classifier

LEXICON	Accuracy score (%)
VADER	70.1
AFINN	63.4
SentiWordNet	60.7
TextBlob	64.2

These revelations, when taken seriously, can provide valuable insights to the bank to make meaningful changes to boost their service delivery. This result in my view answers research questions 1 and 2 adequately.

Fig. 5.2: Negative sentiments word cloud   Fig. 5.3: Tweet illustrating a pain-point.

This work corroborates recent studies by Gregoire, Salle, and Tripp [31] which show that organizations such as financial service firms are increasingly interested in looking for better ways of improving the service experience of customers [32], [33].

5.2 TEXT-BASED FEATURE SELECTION USING BINARY PSO FOR SENTIMENT ANALYSIS

The author of this dissertation presents a metaheuristic-based approach for the optimal selection of features subset via the binary particle swarm optimization (BPSO) metaheuristic algorithm with the view to improve sentiment classification accuracy on the sentiment labelled sentences benchmark dataset. A detailed discussion of this work can be found in [37]. A brief description of the methodology adopted for this work is shown in Fig. 5.4. The author utilized the sentiment labelled sentences dataset available online via the UCI ML repository [26]. The dataset was created for the paper [34] by Kotzias et. al., in the year 2015.

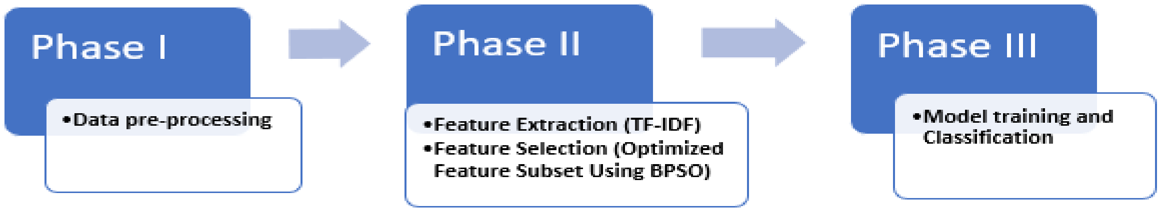

Fig. 5.4: Conceptual framework of the study

Table 5.4 Model accuracy scores

Method	Accuracy Score (%)	BPSO-based (%)	Accuracy Gain (%)
k-NN	68.66	69.57	1.0
NB	73.67	85.27	11.6
SVM	78.67	87.10	8.43

5.2.1 EXPERIMENTAL SETUP AND RESULTS

The study was conducted using Python 3.8 running on a Windows 10 Professional 64bit operating system machine furnished with an Intel Core i7 processor and 8 Gigabyte memory. The ML classifiers were used as a baseline whereas the proposed text feature selection method was designed and implemented using open-source Python libraries available via sklearn [35] and pyswarms [36].

The experiment is assessed using the Accuracy evaluation metric. The results of the text-based feature selection for sentiment classification using the optimized and non-optimized techniques on the sentiments labelled sentences dataset are presented in Table 5.4. From Table 5.4 the best score with BPSO was realized using SVM followed by the NB and k-NN on the sentiment labelled dataset. In conclusion, this work demonstrates the significance of text feature selection for sentiment classification from a metaheuristic perspective with the view to enhancing the classifier accuracy.

5.3 HYBRID BIO-INSPIRED FEATURE SELECTION USING ANGLE MODULATION

In this section, the grey wolf optimization algorithm (GWO) is combined with the angle modulated PSO (AMPSO) to create a new hybrid Angle Modulated GWO and PSO (AMGWOPSO) for wrapper feature selection with the k-NN classifier given the k-NN's implementation simplicity and popularity in hybrid metaheuristic literature [48]. The datasets utilised are shown in Table 5.5.

Table 5.5: UCI datasets used [21]

Number	Dataset	# Features	#Instances
1	Breastcancer	9	699
2	BreastEW	30	569
3	CongressEW	16	435
4	Exactly	13	1000
5	Exactly2	13	1000
6	HeartEW	13	270
7	IonosphereEW	34	351
8	KrVsKpEW	36	3196
9	Lymphography	18	148
10	M-of-n	13	1000
11	PenglungEW	325	73
12	SonarEW	60	208
13	SpectEW	22	268
14	Tic-tac-toe	9	958
15	Vote	16	300
16	WaveformEW	40	5000
17	WineEW	13	178
18	Zoo	16	101

5.3.1 METRICS FOR EVALUATION

To make certain that the experimental outcomes are stable and statistically relevant, the partitioned data is repeated 30 times with the following statistical metrics acquired from the validation data in each run. The Average Classification Accuracy, Average Feature Selection Size, Mean Fitness Function, Best Fitness Function, Worst Fitness function, and Average Computational Time metrics were adopted [23].

5.3.2 EXPERIMENTAL RESULTS AND DISCUSSION

A summary of average classification/features selected results obtained by AMGWOPSO based on the UCI benchmark datasets described earlier is depicted in Table 5.6. The results of the AMGWOPSO are verified against some state-of-the-art methods available in feature selection literature namely. From the results, AMGWOPSO performs better than all the other methods.

Table 5.6. Average classification and features results of proposed AMGWPSO compared to other related state-of-the-art algorithms.

Dataset	Average Accuracy					Average Features Selected				
	AMGWOPSO	AMPSO	BPSO	BWOA	WOASAT-2	AMGWOPSO	AMPSO	BPSO	BWOA	WOASAT-2
Breastcancer	**0.979**	0.967	0.967	0.957	0.970	**3.80**	6.20	5.70	6.38	4.20
BreastEW	0.965	0.926	0.935	0.955	**0.980**	12.50	16.80	16.60	23.80	**11.60**
CongressEW	**1.000**	0.934	0.938	0.929	0.980	**5.60**	7.20	6.80	10.20	6.40
Exactly	0.984	0.694	0.684	0.758	**1.000**	6.20	7.00	9.80	9.20	**6.00**
Exactly2	**0.764**	0.746	0.752	0.698	0.750	**2.40**	5.30	6.20	4.78	2.80
HeartEW	**0.868**	0.810	0.778	0.763	0.850	**4.60**	4.90	7.90	9.40	5.40
IonosphereEW	0.958	0.839	0.837	0.890	**0.960**	13.20	17.80	19.20	22.40	**12.80**

KrVsKpEW	0.975	0.952	0.958	0.915	**0.980**	19.30	19.80	20.80	24.20	**18.40**
Lymphography	**0.918**	0.678	0.689	0.786	0.890	**5.60**	6.80	9.00	10.80	7.20
M-of-n	**1.000**	0.852	0.857	0.854	**1.000**	**5.80**	6.60	9.10	6.02	6.00
PenglungEW	0.938	0.708	0.719	0.729	**0.940**	132.20	179.20	178.80	188.60	**127.40**
SonarEW	0.956	0.738	0.741	0.854	**0.970**	28.20	34.20	32.20	46.20	**26.40**
SpectEW	**0.891**	0.767	0.769	0.788	0.880	**8.80**	9.20	12.50	9.40	9.40
Tic-tac-toe	**0.805**	0.742	0.731	0.751	0.790	**5.60**	6.20	6.60	8.40	6.00
Vote	**0.974**	0.919	0.889	0.939	0.970	**4.20**	4.80	8.80	9.40	5.20
WaveformEW	**0.812**	0.748	0.758	0.713	0.760	**16.20**	23.20	22.70	33.60	20.60
WineEW	0.986	0.946	0.946	0.928	**0.990**	6.80	10.60	8.40	7.38	**6.40**
Zoo	**0.983**	0.824	0.830	0.965	0.970	**5.20**	6.20	9.70	8.80	5.60
Average	**0.931**	0.822	0.821	0.843	0.920	**15.90**	20.67	21.66	24.39	15.99

5.3.3 CONCLUSION OF ACHIEVED RESULTS

The author of this thesis can conclude that the impressive and better results obtained demonstrate an improvement in the exploitation and exploration abilities of the proposed hybrid AMGWOPSO by combining the GWO and AMPSO.
While this proposed AMGWOPSO represents the first attempt in introducing the concept of angle modulation into hybrid metaheuristics FS literature as far as the author can tell, this concept can further be experimented with other non-hybrid/hybrid metaheuristic algorithms to assess their efficacy and stability.

6. SUMMARY OF RESULTS AND DISCUSSION

The recent proliferation of new metaheuristic algorithms after the introduction of metaphor-based development has elicited what researchers refer to as the novel algorithm dilemma. However, issues such as an overly competitive and novel mindset, bio-inspired lingo, and algorithmic dualities among several others were highlighted by the authors in [45] as part of the defects identified in the new metaphor-based algorithms.

On the other hand, proponents of these new metaphor-based metaheuristic algorithms such as GWO cite the "*No Free Lunch*" doctrine in optimization as the main motivation for their metaheuristic algorithmic creativity [46]. Indeed, it is worth mentioning that while an algorithm may perform well on some datasets, its performance may suffer greatly when applied to a different dataset [47]. Thus, it is important to create novel or hybrid methods to optimally address a particular problem or set of problems. While it is important to acknowledge that the proliferation of these metaphor-based metaheuristics algorithms will not decline despite facing generic opposition, every effort must be made to avoid repetitive flaws usually associated with the new metaheuristic methods.

The aims of this dissertation were achieved.

1. Modification of the sentiment analysis pipeline to improve sentiment classification:

a. To determine the best lexicon-based technique based on classification performance and also identify tweet (text) contents most illustrative of positive and negative value user contribution.

The author adopted the sentiment analysis workflow in Chapter 4. The lexicons were selected based on the literature review and the author's interest. Results of this work showcasing the best lexicon based on classification as well as tweets contents most illustrative of positive and negative value-user contribution have since been published in a journal [29].

2. Design a metaheuristic-based solution for sentiment analysis using the binary PSO given the BPSO's impressive performance in feature selection.

A metaheuristic-based approach for optimal selection of features subset via the binary particle swarm optimization (BPSO) metaheuristic algorithm with the view to improve sentiment classification accuracy on the sentiment labelled sentences benchmark dataset was conducted by the author of this dissertation. The study results were presented at the International Conference on Electrical, Computer, and Energy Technologies (ICECET) held in Prague July 20-22, 2022, and have since been published in IEEE Xplore [37].

3. Develop a new Angle Modulated-based metaheuristic memetic method for wrapper feature selection. The proposed method utilizes the GWO AMPSO.

The author developed the hybrid Angle Modulated GWO and PSO (AMGWOPSO) for feature selection tasks (see Chapter 5). Possibilities for publishing the results in a reputable scientific journal will be explored and submitted in due course.

4. Test and evaluate the proposed metaheuristic and memetic method on some selected publicly available benchmark datasets from UCI Invine [21].

The statistical tests and evaluation processes were presented in the results described in Chapter 5.

7. CONTRIBUTION OF THESIS TO SCIENCE AND PRACTICE

As the literature suggests, the upsurge in social media websites has triggered a huge data source for mining interesting expressions on a variety of subjects. Social media data offers great insights for firms and prospective customers in general.

Adopting a data-driven approach for this work produced robust and generalizable outputs compared to conventional marketing approaches such as customer surveys, focus groups, and interviews. From a broader managerial perspective, the study findings can make firms responsive to customer needs and think strategically while focusing on areas of service provision that are vital to business growth. Theoretically, the study contributes to broadening the scope of online banking given the interplay of consumer sentiments via the social media channel.

For feature selection problems, maximizing the classifier performance and reducing the number of features to overcome the curse of dimensionality remains a key priority. It is in this light that a metaheuristic-based Binary Particle Swarm Optimization (BPSO) algorithm is utilized to demonstrate textual FS for effective sentiment analysis/classification on the UCI benchmark sentiment labelled sentences dataset. The results of the evaluation with and without the BPSO on the baseline models prove the superiority of the metaheuristic approach in text feature selection.

By taking inspiration from the BGWOPSO and creating a new (novel) hybrid AMGWOPSO, the concept of employing a trigonometric fitness function as a bit string generator is extended to hybrid metaheuristic algorithms. In other words, the resulting hybrid metaheuristic algorithm has embraced the angle modulation technique used in the domain of signal processing within the telecommunication industry [17]. Indeed, this ability to use the AMGWOPSO as a wrapper feature selection method for feature selection constitutes a major contribution to this work. In sum, the principal theoretical contributions of this new proposed hybrid AMGWOPSO are chronicled below:

- Introduction of angle modulation to the literature on memetic metaheuristic methods for feature selection.
- Propose a hybrid binary AMGWOPSO to solve binary optimization problems.
- Extend the concept of angle modulation from non-hybrid metaheuristic methods to the memetic metaheuristic paradigm.
- Testing and validating the proposed AMGWOPSO's performance on 18 UCI benchmark datasets and other selected metaheuristic algorithms for comparison.

8. CONCLUSION

Understanding the significance of social media has attracted academic attention in recent times. As a prominent scholar once put it, social media is no longer a passing sensation or fad. Customer opinions expressed on social media can convey important messages that businesses can use to build strong relationships with customers.

As a recap, the dual segmentation of the entire dissertation is organised into chapters as follows. The evolving nature of the social media landscape that has redefined the way our society works leading to an upsurge in sentiments analysis research as well as some concepts in soft computing techniques were covered in chapter one. The second chapter reviews (state-of-the-art) literature focusing on approaches that help in understanding and gaining valuable insights from the huge amount of unstructured social media data (Twitter) available as well as the role of soft computing techniques in sentiment analysis.

A summary of the thesis objectives is presented in chapter three. The next chapter describes the sentiment analysis and feature selection workflows adopted respectively. The essential steps involved in each of the workflows are highlighted. Chapter five showcases the relevant experimental results obtained and published (in a journal and a conference) by the author of this thesis where the following observations were made.

Besides enhancing the existing literature on social media analytics, the sentiment analysis approach adopted in this thesis demonstrates that adopting a data-driven approach produces robust and generalizable outputs compared to conventional marketing approaches such as customer surveys, focus groups, and interviews.

An implementation of text feature selection using the BPSO to enhance sentiment classification and analysis is also demonstrated. The results affirm the generalization power of SC methods given that social media data such as tweets, reviews, etc. serves as a good data source used in attesting the reasoning and search capabilities of SC techniques. Furthermore, the author of this thesis employs a new hybrid metaheuristic algorithm to solve feature selection tasks on eighteen selected UCI benchmark datasets.

The findings confirm the competitive and impressive performance of the AMGWOPSO when juxtaposed with other work-related metaheuristics methods available in feature selection literature. Further statistical tests also confirm AMGWOPSO as a potent technique for resolving binary optimization problems across different domains. Chapter six presents the contribution of the dissertation to science and industry with chapter seven concluding the work.

Despite the successes chalked by the novel hybrid method, the fixed amplitude of the generating function constitutes a drawback to the proposed approach given that it is a sine wave. In the future, the author of this thesis will consider modifying the amplitude of the generating function to potentially scale the effect of the vertical shift coefficient. While the authors' proposed AMGWOPSO represents the first attempt at introducing the concept of angle modulation into hybrid metaheuristics FS literature as far as the author can tell, this concept can be further experimented with other non-hybrid/hybrid metaheuristic algorithms to assess their efficacy and stability

BIBLIOGRAPHY

[1] DiMaggio, P., Hargittai, E., Neuman, W. R., & Robinson, J. P. (2001). Social implications of the Internet. *Annual review of sociology*, *27*(1), 307-336.

[2] Kumar, A., & Sebastian, T. M. (2012). Sentiment analysis on twitter. *International Journal of Computer Science Issues (IJCSI)*, *9*(4), 372.

[3] Park, E. (2019). Motivations for customer revisit behavior in online review comments: Analyzing the role of user experience using big data approaches. *Journal of Retailing and Consumer Services*, *51*, 14-18.

[4] Guo, J., Wang, X., & Wu, Y. (2020). Positive emotion bias: Role of emotional content from online customer reviews in purchase decisions. *Journal of Retailing and Consumer Services*, *52*, 101891.

[5] Sivanandam, S. N., & Deepa, S. N. (2007). *Principles of soft computing (with CD)*. John Wiley & Sons.

[6] Balas, V. E., & Fodor, J. (2013). *New Concepts and Applications in Soft Computing*. A. R. Várkonyi-Kóczy (Ed.). Springer.

[7] Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, *5*(1), 1-167.

[8] Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, *63*(1), 163-173.

[9] Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).

[10] Kumar, A., & Jaiswal, A. (2020). Systematic literature review of sentiment analysis on Twitter using soft computing techniques. *Concurrency and Computation: Practice and Experience*, *32*(1), e5107.

[11] Kumar, A., Dabas, V., & Hooda, P. (2020). Text classification algorithms for mining unstructured data: a SWOT analysis. *International Journal of Information Technology*, *12*(4), 1159-1169.

[12] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255-260.

[13] Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1996). Support vector regression machines. *Advances in neural information processing systems*, *9*.

[14] Zhang, J., Jin, R., Yang, Y., & Hauptmann, A. (2003). Modified logistic regression: An approximation to SVM and its applications in large-scale text categorization.

[15] Sulzmann, J. N., Fürnkranz, J., & Hüllermeier, E. (2007, September). On pairwise naive bayes classifiers. In *European Conference on Machine Learning* (pp. 371-381). Springer, Berlin, Heidelberg.

[16] Mohamed, A. W., Hadi, A. A., & Mohamed, A. K. (2020). Gaining-sharing knowledge based algorithm for solving optimization problems: a novel

nature-inspired algorithm. *International Journal of Machine Learning and Cybernetics*, *11*(7), 1501-1529.

[17] Proakis, J. G., & Salehi, M. (2002). Communication systems engineering Prentice-Hall. *Inc. Upper Saddle River, New Jersey*.

[18] Singh, N., & Singh, S. B. (2017). Hybrid algorithm of particle swarm optimization and grey wolf optimizer for improving convergence performance. *Journal of Applied Mathematics*, *2017*.

[19] Franken, C. J. (2005). *PSO-based coevolutionary game learning* (Doctoral dissertation, University of Pretoria).

[20] Pampara, G., Franken, N., & Engelbrecht, A. P. (2005, September). Combining particle swarm optimisation with angle modulation to solve binary problems. In *2005 IEEE congress on evolutionary computation* (Vol. 1, pp. 89-96). IEEE.

[21] Blake, C. (1998). UCI repository of machine learning databases.[Online]. Available: *http://www. ics. uci. edu/~ mlearn/MLRepository.html*.

[22] Mafarja, M. M., & Mirjalili, S. (2017). Hybrid whale optimization algorithm with simulated annealing for feature selection. *Neurocomputing*, *260*, 302-312.

[23] Al-Tashi, Q., Kadir, S. J. A., Rais, H. M., Mirjalili, S., & Alhussian, H. (2019). Binary optimization using hybrid grey wolf optimization for feature selection. *Ieee Access*, *7*, 39496-39508.

[24] Agrawal, P., Abutarboush, H. F., Ganesh, T., & Mohamed, A. W. (2021). Metaheuristic algorithms on feature selection: A survey of one decade of research (2009-2019). *IEEE Access*, *9*, 26766-26791.

[25] Emary, E., Zawbaa, H. M., & Hassanien, A. E. (2016). Binary grey wolf optimization approaches for feature selection. *Neurocomputing*, *172*, 371-381.

[26] Dua, D. and Graff, C. (2019). UCI Machine Learning Repository.[Online]. Available: [http://archives.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[27] Nielsen, F. Å. (2017). afinn project.

[28] Loria, S. (2018). Textblob Documentation. *Release 0.15*, *2*(8).

[29] Botchway, R. K., Jibril, A. B., Oplatková, Z. K., & Chovancová, M. (2020). Deductions from a Sub-Saharan African Bank’s Tweets: A sentiment analysis approach. *Cogent Economics & Finance*, *8*(1), 1776006.

[30] Pustejovsky, J., & Stubbs, A. (2012). Natural Language Annotation for Machine Learning: A guide to corpus-building for applications. " O'Reilly Media, Inc."

[31] Grégoire, Y., Salle, A., & Tripp, T. M. (2015). Managing social media crises with your customers: The good, the bad, and the ugly. *Business Horizons*, *58*(2), 173-182.

[32] Ernst & Young (2017). Customer Experience: Innovate Like a FinTech," Retrieved from https://www.ey.com/Publication/vwLUAssets/ey-gcbs-customerexperience/$FILE/ey-gcbs-customer-experience.pdf.

[33] Ibrahim, N. F., & Wang, X. (2019). Decoding the sentiment dynamics of online retailing customers: Time series analysis of social media. *Computers in Human Behavior*, *96*, 32-45.

[34] Kotzias, D., Denil, M., De Freitas, N., & Smyth, P. (2015, August). From group to individual labels using deep features. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 597-606).

[35] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

[36] Miranda, L. J. (2018). PySwarms: a research toolkit for Particle Swarm Optimization in Python. *Journal of Open Source Software*, *3*(21), 433.

[37] Botchway, R. K., Yadav, V., Komínková, Z. O., & Senkerik, R. (2022, July). Text-based feature selection using binary particle swarm optimization for sentiment analysis. In *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)* (pp. 1-4). IEEE.

[38] Faris, H., Aljarah, I., Al-Betar, M. A., & Mirjalili, S. (2018). Grey wolf optimizer: a review of recent variants and applications. *Neural computing and applications*, *30*(2), 413-435.

[39] Chantar, H., Mafarja, M., Alsawalqah, H., Heidari, A. A., Aljarah, I., & Faris, H. (2020). Feature selection using binary grey wolf optimizer with elite-based crossover for Arabic text classification. *Neural Computing and Applications*, *32*(16), 12201-12220.

[40] Sharma, M., & Kaur, P. (2021). A comprehensive analysis of nature-inspired meta-heuristic techniques for feature selection problem. *Archives of Computational Methods in Engineering*, *28*, 1103-1127.

[41] Dave, K., Lawrence, S., & Pennock, D. M. (2003, May). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web* (pp. 519-528).

[42] Singh, P., Dwivedi, Y. K., Kahlon, K. S., Sawhney, R. S., Alalwan, A. A., & Rana, N. P. (2020). Smart monitoring and controlling of government policies using social media and cloud computing. *Information Systems Frontiers*, *22*, 315-337.

[43] Fernández-Gavilanes, M., Álvarez-López, T., Juncal-Martínez, J., Costa-Montenegro, E., & González-Castaño, F. J. (2016). Unsupervised method for sentiment analysis in online texts. *Expert Systems with Applications*, *58*, 57-75.

[44] Macedo, M., Siqueira, H., Figueiredo, E., Santana, C., Lira, R. C., Gokhale, A., & Bastos-Filho, C. (2021). Overview on binary optimization using swarm-inspired algorithms. *IEEE Access*, *9*, 149814-149858.
[45] Sörensen, K. (2015). Metaheuristics—the metaphor exposed. *International Transactions in Operational Research*, *22*(1), 3-18.
[46] Ji, B., Lu, X., Sun, G., Zhang, W., Li, J., & Xiao, Y. (2020). Bio-inspired feature selection: An improved binary particle swarm optimization approach. *IEEE Access*, *8*, 85989-86002.
[47] Agushaka, J. O., Ezugwu, A. E., & Abualigah, L. (2022). Gazelle Optimization Algorithm: A novel nature-inspired metaheuristic optimizer. *Neural Computing and Applications*, 1-33.
[48] Alwajih, R., Abdulkadir, S. J., Al Hussian, H., Aziz, N., Al-Tashi, Q., Mirjalili, S., & Alqushaibi, A. (2022). Hybrid binary whale with harris hawks for feature selection. *Neural Computing and Applications*, *34*(21), 19377-19395.

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS, ACRONYMS, AND ABBREVIATIONS

AMPSO	Angle Modulated Particle Swarm Optimization
AMGWOPSO	Angle Modulated Grey Wolf Optimization Particle Swarm Optimization
BGWOPSO	Binary Grey Wolf Optimization Particle Swarm Optimization
BPSO	Binary Particle Swarm Optimization
BWOA	Binary Whale Optimization Algorithm
EC	Evolutionary Computation
FL	Fuzzy Logic
FS	Feature Selection
GWO	Grey Wolf Optimization
k-NN	K-Nearest Neighbour
MA	Metaheuristic Algorithm
ME	Maximum Entropy
ML	Machine Learning
NB	Naïve Bayes
NLP	Natural language Processing
NLTK	Natural language Toolkit
NN	Neural Network
OM	Opinion Mining
PR	Probabilistic Reasoning
PSO	Particle Swarm Optimization
PSOGWO	Particle Swarm Optimization Grey Wolf Optimization
SA	Sentiment Analysis
SC	Soft Computing
SVM	Support Vector Machine
TF-IDF	Term Frequency Inverse Document Frequency
UCI	University of California, Irvine
UGC	User Generated Content
VADER	Valence Aware Dictionary and sEntiment Reasoner
WOASAT-2	Whale Optimization Algorithm Simulated Annealing Tournament Selection 2

LIST OF PUBLICATIONS BY THE AUTHOR

Journals with IF:

[1] **Botchway, R. K**., Jibril, A. B., Oplatková, Z. K., Jasek, R., & Kwarteng, M. A. (2021). Decision science: a multi-criteria decision framework for enhancing an electoral voting system. *Systems Science & Control Engineering*, *9*(1), 556-569.

[2] Jibril, A. B., Kwarteng, M. A., **Botchway, R. K**., Bode, J., & Chovancova, M. (2020). The impact of online identity theft on customers' willingness to engage in e-banking transaction in Ghana: A technology threat avoidance theory. *Cogent Business & Management*, *7*(1), 1832825.

[3] **Botchway, R. K**., Jibril, A. B., Oplatková, Z. K., & Chovancová, M. (2020). Deductions from a Sub-Saharan African Bank's Tweets: A sentiment analysis approach. *Cogent Economics & Finance*, *8*(1), 1776006.

Journals (Scopus):

[1] Yadav, V., **Botchway, R. K**., Senkerik, R., & Oplatkova, Z. K. (2021, December). Robotic Automation of Software Testing From a Machine Learning Viewpoint. In *Mendel* (Vol. 27, No. 2, pp. 68-73).

Conference:

[1] Yadav, V., **Botchway, R. K**., Senkerik, R., & Komínková, Z. O. (2023, July). Robot Automation of Software Using Genetic Algorithm. *In 2023 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)* IEEE. Accepted.

[2] **Botchway, R. K**., Yadav, V., Komínková, Z. O., & Senkerik, R. (2022, July). Text-based feature selection using binary particle swarm optimization for sentiment analysis. In *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)* (pp. 1-4). IEEE.

[3] Yadav, V., **Botchway, R. K**., Senkerik, R., & Kominkova, Z. O. (2021, December). Robot Testing from a machine learning perspective. In *2021 International Conference on Electrical, Computer and Energy Technologies (ICECET)* (pp. 1-4). IEEE.

[4] Odei, M. A., Amoah, J., Jibril, A. B., **Botchway, R. K**., Naatu, F., & KorantwiI-Barimah, S. (2021, September). A Review of Barriers Facing Social Media Usage Among Firms in Less Digitalized Economies. In *European Conference on Innovation and Entrepreneurship* (pp. 677-XVII). Academic Conferences International Limited.

[5] Kwarteng, M. A., Ntsiful, A., **Botchway, R. K**., Pilik, M., & Oplatková, Z. K. (2020, December). Consumer Insight on Driverless Automobile Technology Adoption via Twitter Data: A Sentiment Analytic Approach. In *International Working Conference on Transfer and Diffusion of IT* (pp. 463-473). Springer, Cham.

[6]Kwarteng, M. A., Jibril, A. B., **Botchway, R. K**., Kwarteng, O. V., Pilik, M., & Chovancova, M. (2019, September). Assessing pre-purchase risk attributes towards used-products: evidence from e-shoppers in the Czech Republic. In *Proceedings of the 3rd International Conference on Business and Information Management* (pp. 15-20).

[7] **Botchway, R. K**., Jibril, A. B., Kwarteng, M. A., Chovancova, M., & Oplatková, Z. K. (2019, September). A review of social media posts from UniCredit bank in Europe: a sentiment analysis approach. In *Proceedings of the 3rd International Conference on Business and Information Management* (pp. 74-79).

AUTHOR'S PROFESSIONAL CURRICULUM VITAE

PERSONAL INFORMATION

Raphael Kwaku Botchway

TGM 3050,76001 , Zlin, Czech Republic

+420 774948141

botchway@utb.cz /ralph.botchway@gmail.com

Sex Male | Date of birth 20/06/1979 | Nationality Ghanaian

WORK EXPERIENCE

2004- JULY 2006

Electricity Company of Ghana
- Oracle Database Administrator

2006- 2010

National Investment Bank, Accra - Ghana
- Senior I.T Officer

2012- SEPT 2016

May-Awurade Adom Ventures, Accra - Ghana
- Operations Manager

APRIL 2017- SEPT 2017

Exxon Mobil Kft, Hungary
- Internship

EDUCATION AND TRAINING

April 2019-Date (Expected graduation: March 2023)

PhD Candidate

Faculty of Applied Informatics, Tomas Bata University in Zlin, Czech Republic

Thesis Title: Soft computing Techniques for Sentiment Analysis and Feature Selection
- Participated in faculty Internal Grant Agency (IGA) research projects

2010-2012

MSc Informatics & Systems Engineering

Czech University of Life Sciences in Prague, Prague-Czechia

1998-2002

BSc (Hons) Computer Science

Kwame Nkrumah University of Ghana & Technology, Kumasi - Ghana

STUDY INTERNERNSHIP

April 15, 2022 - July 15, 2022

Stellenbosch University, South Africa

Traineeship title : Angle modulated PSO for feature selection

Mother tongue(s) English & Akan

Other language(s)	UNDERSTANDING		SPEAKING		WRITING
	Listening	Reading	Spoken interaction	Spoken production	
English	C1	C1	C1	C1	C1
	.				

Levels: A1/2: Basic user - B1/2: Independent user - C1/2 Proficient user
Common European Framework of Reference for Languages

Communication skills

Good communication and listening skills
Team-work oriented

Organisational / managerial skills

Effectively led a team that deployed and managed Bancassurance product in NIB Ltd nationwide.

Computer skills

MATLAB, Python, SQL, Oracle PL-SQL, Oracle Database Administration

Honours and Awards

Czech Government/Government of Ghana Scholarship 2010

Soft computingové techniky pro analýzu sentimentu a výběr příznaků

Soft Computing Techniques for Sentiment Analysis and Feature Selection

Summary of Doctoral Dissertation

Published by: Tomas Bata University in Zlín

nám. T. G. Masaryka 5555, 760 01 Zlín, Czech Republic

Edition: Published electronically

Typesetting by: Ing. Raphael Kwaku Botchway, PhD.

This publication has not undergone any proofreading or editorial review.

First edition.

Publication year: 2023

ISBN 978-80-7678-193-1