Machine Learning and Sentiment Analysis: Examining the Contextual Polarity of Public Sentiment on Malaria Disease in Social Networks

Jelili Oyelade[1, 3, *], Efosa Uwoghiren[1, 3], Itunuoluwa Isewon[1, 3], Olufunke Oladipupo[1], Olufemi Aromolaran[1, 3] & Kingsley Michael[2]

[1]Department of Computer & Information Science, Covenant University, Ota
[2]Department of Accounting, Covenant University, Ota
[3]Covenant University Bioinformatics Research Cluster (CUBRe), Ota

Abstract

Malaria, a major deadly disease which is still a threat to human life's even though numerous efforts has been put to fight it, still affects over two hundred million people each year amongst which over a million individuals dies. Twitter happens to be an important and comprehensive source of information that is quite subjective to individual sentiments towards public health care. In this study, we extracted tweets from the social network twitter, we pre-processed the tweets extracted and built a model to fit our data using a machine learning approach for text classification to determine the contextual polarity of every tweet on the subject of malaria in the bid to harvest peoples' opinion towards malaria and understand how well research and recent development in the aid to tackle malaria has affected the opinions of the public towards the subject malaria. This study finds that tweets extracted, pre-processed and classified in this study were majorly classified as negative (-ve) due to the fact that tweets tweeted were majorly about different occurrence of death, misinformation and need for donations to save a life, hence a major awareness is needed.

Keywords: Sentiment Analysis, Machine Learning Technique, Malaria, Twitter, Data Mining.

1. Introduction

The dominance of malaria resistance to all identified anti-malarial drugs in current circulation has given rise to the increase of anti-malarial drug discovery research [1]–[3]. Hence, research towards the development of novel drug which would serve as effective solutions for malaria treatment are urgently needed [2]–[5] because despite the colossal efforts put in to fight malaria, the disease still affects up to over 200 million people every year amongst which close to half a million dies [1], [5]–[9]. Considering the declaration "Action and Investment to defeat Malaria 2016–2030 (AIM) – for a malaria-free world" [10], we have decided to examine public opinion towards the subject malaria. Opinions are central to almost all human activities because they are key influencers of our behaviours. As humans we love to find out what the public feels or think about a particular brand, topic or subject area this does not leave out finding out what people think about a particular disease and how research and recent developments are affecting the public opinion. With the proliferation of Web to applications such as micro-blogging, forums and social networks, reviews, comments, recommendations, ratings and feedbacks have been made very easy as users can generate content about virtually anything.Twitter, which is a micro blogging platform permits the fast exchange of personal ideas and thoughts, therefore allowing users to tweet messages of about 140 characters [11]–[13]. With the explosion of this user generated content, came the need for companies, service providers, social psychologists, analysts and researchers to mine and analyze these contents for different relevant uses, this is quite significant bearing in mind that tweets are often treated as facts and are cited in information outlets for example news media [14]. The research community and organizations are not left out in the need to find out how their research and recent developments are affecting the public opinion towards their discoveries and novel implementations in a particular subject area, hence, the reason for sentiment analysis. In this study, we have decided to examine if the ground-breaking research and recent developments on the subject malaria have really affected people's opinion and awareness on the subject malaria.

1.1 Natural Language Processing

Natural languages are those languages spoken or written by humans for the purposes of communication. Natural Language processing (NLP) can be defined as "a theoretically motivated range of computational techniques for analysing and representing naturally occurring texts at one or more levels, for the purpose of achieving human-like language processing of tasks and applications"[15], [16]. The field of NLP involves making computers to perform useful tasks with the natural languages for human use. The input and output of an NLP system can be either Speech or Written Text. Below are the different stages in natural language processing.

978-1-943436-11-8 / copyright ISCA, BICOB 2018
March 19-21, 2018, Las Vegas, Nevada, USA

Phonology: This deals with organizing sound systematically. Phonetics and phonology deal with the articulatory and acoustic properties of speech sounds, how they are produced, and how they are perceived, and the rules that govern them [17], [18].

Lexical Analysis:identifying and analysing the structure of words. Lexicon of a language means the collection of words and phrases in a language [19], [20].

Morphological Analysis:Thisrefers to the study of construction of words from primitive meaningful units. Since the meaning of each morphemes are the same across words human can break down an unknown word into constituent’s morphemes in order to understand its meaning [21], [22].

Syntactic Analysis: This involves determining the structural role of words in the sentence and in phrases. The words are transformed into structures that show how the words are related to each other. This requires the grammar and the parser. The output of this level of processing is a representation of the sentence that reveals the structural dependencies and relationships between the words [17], [18], [22].

Semantic Analysis:It is concerned with the meaning of words and how to combine words into meaningful phrases and sentences. It assigns meanings to natural language utterances. A semantic representation must be precise and unambiguous. It draws the exact meaning or the dictionary meaning from the text. The text is checked for meaningfulness. It is done by mapping syntactic structures and objects in the task domain [15], [20], [22].

Discourse Analysis: It deals with how the immediately preceding sentence can affect the interpretation of the next sentence. For example the word “it” in the sentence “she wanted it” depends upon the prior discourse context. The meaning of any sentence depends upon the meaning of the sentence just before it. In addition, it also brings about the meaning of immediately succeeding sentences [15], [17], [20], [21].

1.2 Text Mining

Text mining, also known as Intelligent Text Analysis or Knowledge-Discovery in Text (KDT), refers generally to the process of extracting interesting and non-trivial information and knowledge from unstructured text [23]–[26]. Text mining is now widely being applied to many domains, some of its application areas include: Sentiment analysis, Educational application, Security applications, biomedical applications, Digital humanities and Computational sociology etc.

1.3 Sentiment Analysis

Sentiment analysis gives room of harvesting opinions from reviews or expression of different users on a particular subject matter or product. This groups opinions into either negative, positive or neutral helping to determine the attitude or opinion of a particular writer or speaker with respect to certain topics [27]–[29].Sentiment Analysis is considered a classification process. 3 major classification levels makes up sentiment analysis which are the Sentence-level whose its objective is classifying sentiment expressed in subjective sentences as either negative or positive sentiments, the Document-Level whose its objective is classifying sentiment of the whole document as expressing as either negative or positive sentiments and the Aspect Level whose its objective is classifying sentiment with respect to the specific aspects of entities [30]–[35].

1.3.1 Applications of Sentiment Analysis

Sentiment analysis has been applied to various real world scenarios and a few has been considered in this study.

Reputation Monitoring: sentiment analysis has been applied to monitor the reputation of different brands on Facebook and twitter as they are the core of sentiment analysis in this domain [36]

Product and Service Reviews: sentiment analysis has been applied to the reviews of consumer products and services using websites that provide automated summaries of reviews about products and about their specific aspects [27], [37]

Result Prediction: sentiment analysis has been applied to predict the probable outcome of a particular event [38]. For instance, sentiment analysis can provide substantial value to candidates running for various positions enabling campaign track how voters feel about different issues and how they relate to the speeches and actions of the candidates [38]

Decision Making: sentiment analysis has been applied to help decision making process [39]. There are numerous news items, articles, blogs, and tweets about each public company. A sentiment analysis system can use these various sources, articles that discuss the companies and aggregate the sentiment about them as a single score that can be used by an automated trading system [29], [39]

1.3.2 Sentiment Classification

The major aim of sentiment classification is classifying documents or reviews into a definite number of predefined categories. The expressions of opinions in a sentence, document etc. could either be done using a scaling system or binaries (negative and positive).

Feature Selection in Sentiment Classification: Feature selection gives a better understanding of data by giving their important features. Feature selection is an important step in text categorization problems; not all the features in a document are required to classify it. Feature selection helps in removing unimportant or redundant words of text and thereby reduces the dimensionality of documents [29], [40].

1.3.3 Sentiment Classification Techniques

It can be roughly divided into machine learning approach and lexicon based approach, the lexical based approach relies on a sentiment lexicon, a collection of known and precompiled sentiment terms.

Lexicon-based approach: Opinion words are employed in many sentiment classification tasks. Positive opinion words are used to express some desired states, while negative opinion words are used to express some undesired states. There are also opinion phrases and idioms which together are called opinion lexicon [35], [38], [41], [42].

Machine learning (ML) Approach: Machine learning techniques first trains the algorithm with a training data set before applying it to the actual data set [43]. The text classification methods using machine learning approach can be roughly divided into supervised and unsupervised learning methods [43], [44].

2. Materials and Method

2.1 Phases of the Methodology

Represented in table 1 are the method and techniques used in different phases of this study.

Table 1: Method and techniques used in different phases

Phase	**Methods | techniques**	**Output**
Tweets Collection	Obtaining tweets from Streaming twitter API	Samples of relevant tweets.
Linguistic analysis	Lemmatization, stop words removal, harsh tags removal	Pre-processed tweets
Feature Extraction	Extracting feature vectors to identify sentiment polarity	Feature vectors
Sentiment Analysis	Implementation of the model to classify and identify sentiment polarity	Sentiment polarity, positive, Negative and Neutral

2.2 Design Considerations for the Sentiment Analysis Model

The approach for the development of our sentiment analysis model which involves tweets preprocessing, feature extraction, Machine learning text clustering and classification techniques is being described.

Tweets used were obtained through the Twitter streaming API using python. These sets of tweets serve as the input data for our model. These tweets were preprocessed by removing stop words and harsh tags. Thereafter, lemmatization, POS tagging, were performed to enable the efficient extraction of features for the classification of sentiments of the tweets. After these tweets have been preprocessed, feature vectors were extracted and these vectors were analyzed using machine learning algorithms of clustering and classification techniques, the algorithms to be used include Naïve Bayes and Support vector Machine algorithms.

2.3 System Architecture

This shows the formal description and representation in a way that supports reasoning about the structure and behaviour of the system. It comprises of the individual components and the way they work together to implement the sentiment analysis model. Figure 1 shows the System architecture and its various components.

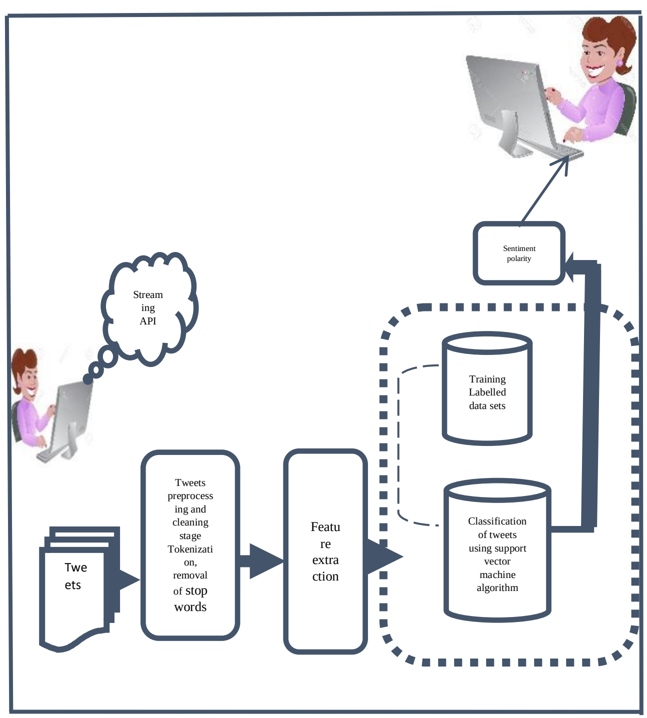

Figure 1: The system architecture of our study

2.4 Framework of the Sentiment Analysis Model

Here the different stages involved in the development of this study sentiment analysis model are comprehensively outlined. It explains the process and components for analysis, design and implementation of the model. Figure 2 shows the entire life cycle of this model which is in 4 major phases namely Tweet Pre-processing, Feature Extraction, Semantic Analysis, and Machine learning Classification.

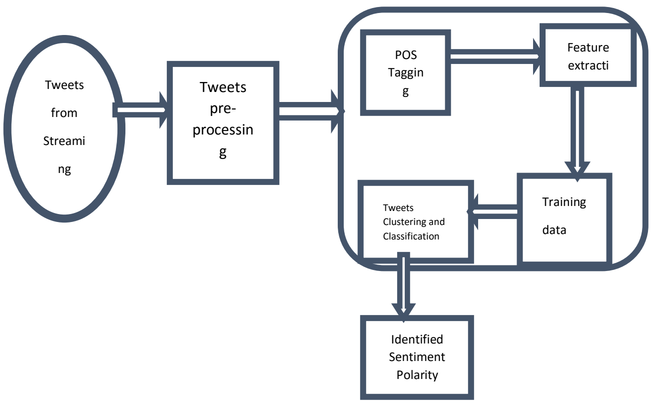

Figure 2: Framework of the Sentiment Analysis Model

2.4.1 Tweets Preprocessing

This is the first stage of the development process. Natural language text can't be processed directly; it must be pre-processed first in order to obtain accurate results at the end of this project. In light of that, the Pre-processing phase includes Tokenization, Language detection, Stop words removal. These processes would be explicitly explained below with corresponding examples.

Tokenization: We tokenized the text to make it easy to separate out other unnecessary symbols and punctuations, and leave out only those words that can add value to the sentimental polarity score of the text. For a sample input text says "Tunde said the food he bought is bad". Tokenizing divides the strings into lists of substrings known as the tokens.

Language detection: This enables only the detection of tweets in English since we are mainly interested in English text only. This is possible by using NLTK's language detection feature.

Stop words Removal: In this process, we removed very common words such as "all", "almost", "alone", etc. The reason for this is because their appearance in a tweet does not provide any useful information in classifying a tweet as positive, negative or neutral.

2.4.2 Part of speech (POS) tagging

POS tagging assigns a tag to each word in a text and classifies a word to a specific morphological category such as noun, verb, adjective, etc. POS taggers are efficient for explicit feature extraction in terms of accuracy.

2.4.3 Feature Extraction

The improved dataset after pre-processing has a lot of distinctive properties. The feature extraction method, extracts the adjective from the dataset. Later this adjective would be used to show the positive and negative polarity in a sentence which is useful for determining the opinion of individuals. This would be done using a Unigram model. Unigram model extracts the adjective and segregates it. It discards the preceding and successive word occurring with the adjective in the sentences. For above example, i.e. "painting ugly" through unigram model, only ugly is extracted from the sentence.

2.4.4 Tweets Classification and Identification of Sentiment Polarity

When using a machine learning approach, the ways features are selected are very important to the success rate of the classification. In this study, we used 2 major machine learning algorithms which are Support Vector Machine (SVM) Algorithm and Naïve Bayes (NB) Algorithm. SVM & NB algorithms are trained using the features generated which results into the training model.

3. Results

3.1 Tweets Collection

Twitter Application Programming Interface (API) which was implemented in python language was used to collect English-language tweets relating to malaria over a period of several weeks. The corpus consists of thousands of tweets from search tags like 'malaria', 'malaria Africa', 'malaria Mozambique', 'malaria Nigeria', etc. This task was performed by different experts.

Table 2: Sample of tweets collected

Sample of tweets collected	***Sample of tweets collected***
@Y1079FM @iambrownberry @deejayLoft @LukmanEvergreenHelo brown berry u make my sunday morning amazing wooow.cant s\xe2\x80\xa6 https://t.co/o2x1DPUT7z Currently sat under my	*@SandeshRishi @altnews_in @mepratap @timesofindia You are so right, the girl died because of Malaria !\xe2\x80\xa6 https://t.co/CibEtxBtk7 Do you know that hunger kills more people annually than AIDS, malaria and*

mosquito. net listening to TMS in Nigeria suffering with malaria! #bbccricket #TMS RT @fulelo: Extreme gardening to help tackle malaria https://t.co/ojaXL51W6J RT @Beanchesterr: Latest research has it that bank alerts can cure malaria. Extreme gardening to help tackle malaria https://t.co/ojaXL51W6J @GRadioRockstar Lmao and the you read through and find out "Big Daddy Juix was exposed to malaria" damn headlines	tuberculosis combined?\xe2\x80\xa6 https://t.co/51bh4ItiiD Is there a way to threat Malaria without taking pills or injections??? RT @scienmag: Chances of surviving malaria may be higher when host consumes fewer calories https://t.co/NCF8WH7JFP https://t.co/AAxOaotUz3 @Blackkout__ I got very sick at they little age... I was down with malaria 3+.
I was very OK but the stress was too much. And it's not like I did any tedious work. Just moving about caused malaria. RT @ISGLOBALorg: Beatriz Galatas @beagalatas: Working on #malaria elimination in #Mozambique https://t.co/esVX5jrSmX @Manhica_CISM @FundlaC\xe2\x80\xa6 After taking malaria meds i always think to myself "dont go to sleep, dont go to sleep..." https://t.co/egUI8rPWWe RT @PaulUithol: Second day of @sotmafrica! Kicking off with	Before he was diagnose of malaria in during the week and couldn't make it by weekend.So sad. I wish the Olajide's fortitude to bear the loss RT @PreventionTips: Malaria, Mosquitoes and Man: Prevention & Control https://t.co/9Ingffq6xk https://t.co/eMEM1cxZCP Expert wants increased awareness on malaria to save more\xc2\xa0lives https://t.co/0jl4tNTWsv https://t.co/iWnEODQh8y Don't forget to take your #Malaria pill!

3.2 Linguistic analysis

The tweets are pre-processed removing stop words, hash tags, duplications and languages that are not known to be English language hereby leaving us with the tweets needed for feature extraction. Below is a sample of the pre-processed tweets.

Table 3: Pre-processed tweets

Pre-processed tweets	***Pre-processed tweets***
extreme gardening to help tackle malaria latest research has it that bank alerts can cure malaria after taking malaria meds i always think to myself dont go to sleep, dont go to sleep researchers find that simple blood test predicts anemia risk after malaria treatment My barber's friend died last Friday fine boy, waiting for nysc. everything smooth he died of malaria	*Before he was diagnose of malaria in during the week and couldn't make it by weekend.so sad. i wish the olajide's fortitude to bear the loss wants increased awareness on malaria to save more 2tweetaboutit Africa is the epicentre of malaria. soon there will be an malaria epidemic in Europe*
a human trial for a malaria vaccine has achieved up to 100% protection against infection for at least 10 weeks mapbox using mapping and visualizations to fight malaria is there a way to threat malaria without taking pills or injections ?? chances of surviving malaria may be higher when host consumes fewer calories baygon is almost 2k which is hilarious bc it's technically cheaper for me to get malaria and treat it.	*world's first malaria vaccine will be given to thousands of babies in Africa Heavy rainfall + poor drainage = mosquitoes having the time of their life spreading malaria. indigenous knowledge systems and innovations in malaria control in Nigeria No wonder Nigeria is still fighting malaria. everywhere waterlogged as fuck Mosquitoes in Nigeria will give you malaria. Mosquitoes in Philippines will give you dengue fever, which is very deadly.*
malaria control in African schools dramatically cuts infection and reduces risk of anaemia u.s. malaria donations saved almost 2m African children i got very sick at they little age..i was down with malaria you can help us save lives by giving the gift of malaria treatment, which quickly restores	*antimalarials of unproven quality rampant in africa - sub-saharanafrica king mswati ii of Swaziland calls for increased domestic investments to eliminate malaria in Africa did you know: in Nigeria it is cheaper to get malaria and treat it than to buy insecticide touching needs donations to save children in Nigeria from malaria-please donate*

3.3 Feature Extraction

Below is an improved dataset after extracting the adjective from the pre-processed dataset which becomes useful when classifying the polarities into negative, neutral or positive polarities.

Table 4: Feature vectors

Sample feature vectors	**Sample feature vectors**
extreme gardening help	*diagnose malaria week make*

tackle malaria	*sad wish fortitude bear loss*
taking malaria meds think myself dont sleep dont sleep	*got sick little age malaria*
friend died Friday fine boy waiting nysc smooth died malaria	*calls strong regulation local production antimalarials*
malaria donations saved African children	*defeat malaria Africa*
chances surviving malaria higher host consumes fewer calories	*malaria vaccine given thousands babies Africa*
baygon hilarious bc technically cheaper malaria treat	*malaria death rate Africa fell*
	Gambia massive progress malaria elimination sight

3.4 Classifier Performance

Our classifier labelled tweet sentiment with an accuracy of about 72.41%. Importantly, no positively classified tweets were manually labelled as negative, and only 2% of the negatively classified tweets were manually labelled as positive by assigning a positive polarity to them. The misclassifications were predominantly for tweets with non-neutral sentiment classified as being neutral. As such, the overwhelming majority of misclassified tweets did not entail complete reversal of sentiment. Below are the accuracy of the five classifiers that were combined and used in this study.

Table 5b: Combinations of Naïve Bayes Classifiers

CombinationsofNaiveBayesClassifiers, fordatasetstraining	
Original Naive Bayes Algo accuracy percent	74.54819277108435
MNB_classifier accuracy percent	73.94578313253012
BernoulliNB_classifier accuracy percent	73.64457831325302

Table 5c: Combinations of Support Vector Machine Classifiers

CombinationsofSupportVectorMachineClassifiers, fordatasetstrainin	
LinearSVC_classifier accuracy percent	70.03012048192771
NuSVC_classifier accuracy percent	69.879518072289

3.4.1 F1-Score

The table 5a below shows the performance analysis of the corpus used for training the classifiers.

Precision Score = tp / (tp + fp)

Recall Score = tp / (tp + fn)

F1-Score: F1 = 2 * (precision * recall) / (precision + recall)

Where: tp = true positives, fp = false positives, fn= false negatives.

Table 5a: F1-score

	Precision	**recall**	**f1-score**	**support**
pos	0.87	0.86	0.87	1600
Neg	0.83	0.89	0.87	1600
Avg / total	0.85	0.88	0.87	3200

3.5 Sentiment Analysis

Every tweet after feature extraction was done separately on them after the model has been trainedwere assigned polarities based on the how it was classified by the model ranging from neutral (0.25, 0.5, 0.75, 1), to negative (0.25, 0.5, 0.75, 1), to positive (0.25, 0.5, 0.75, 1) respectfully. Below are samples of polarities assigned to tweets and a graph diagrammatically representing the polarities of tweets of a period of time?

Table 6: Polarity of tweets extracted

Sample of tweets polarity	**Sample of tweets polarity**
taking malaria meds think myself dont sleep dont sleep |neg|1.0	*diagnose malaria week make sad wish fortitude bear loss |neg|1.0*
friend died friday fine boy waiting nysc smooth died malaria |neg|1.0	*got sick little age malaria |neg|1.0*
baygon hilarious bc technically cheaper malaria treat |neg|1.0	*leaders adopt new strategic framework end aids tb malaria |pos|1.0*
American donations fight malaria Africa saved lives nearly million |pos|0.75	*antimalarials unproven quality rampant Africa scidevnet report malaria |neg|0.75*
people die malaria yearly reduce clean environment join movement |neg|1.0	*Gambia massive progress malaria elimination sight |neu|1*
children world die malaria |neg|0.75	*heavy rainfall poor drainage mosquitoes having time life spreading malaria |neg|1.0*

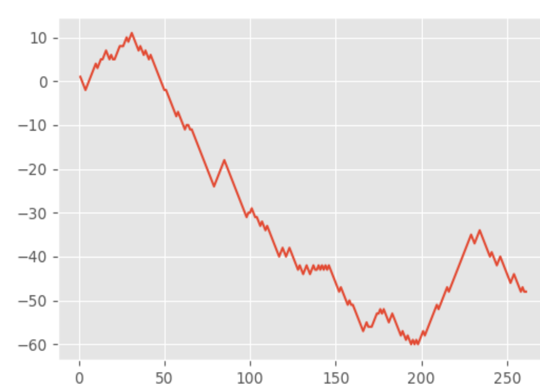

Figure 3: Graph to represent polarities of tweet

4. Discussion

In this study, we extracted tweets relating to malaria over a period of time from the social network twitter, we pre-processed the tweets extracted to eliminate unimportant tweets and redundancy, and we built a model to fit our data using machine learning approach for text classification to determine the contextual polarity of every tweet on the subject of malaria in a bid to harvest peoples opinion towards malaria and understand how well research and recent development in the aid to tackle malaria has affected the opinions of people towards the subject malaria. This study finds that though lots of ground breaking research are ongoing, awareness on malaria treatment and prevention needs to be on the increase and ground breaking research in this area needs to be communicated to the public appropriately through the appropriate authorities because tweets extracted, pre-processed and classified in this study were majorly classified as negative due to the fact that tweets tweeted were majorly about different occurrence of death, misinformation and need for donations to save a life. We hereby proposed that periodical analysis be done on the subject malaria, also expanding the source of data to closely monitor the awareness of the public and the opinions of the public on the subject malaria which would help benchmark how effective research beencarried out are affecting the public and their level of awareness on malaria prevention and treatments.

ACKNOWLEDGMENT

We acknowledge the support of Covenant University Center for Research, Innovation and Discovery, Covenant University, Ota, Nigeria.

Reference

[1] C. Ekenna, S. Fatumo, and E. Adebiyi, "In-silico evaluation of malaria drug targets," *Int. J.*, 2010.

[2] A. Golldack, B. Henke, B. Bergmann, and M. Wiechert, "Substrate-analogous inhibitors exert antimalarial action by targeting the Plasmodium lactate transporter PfFNT at nanomolar scale," *PLoS*, 2017.

[3] K. Sahan, C. Pell, and F. Smithuis, "Community engagement and the social context of targeted malaria treatment: a qualitative study in Kayin (Karen) State, Myanmar," *Malaria*, 2017.

[4] S. Hapuarachchi, S. Cobbold, and S. Shafik, "The Malaria Parasite's Lactate Transporter PfFNT Is the Target of Antiplasmodial Compounds Identified in Whole Cell Phenotypic Screens," *PLoS*, 2017.

[5] "WHO | World Malaria Report 2016," *WHO*, 2016.

[6] N. Dholakia, P. Dhandhukia, and N. Roy, "Screening of potential targets in Plasmodium falciparum using stage-specific metabolic network analysis," *Mol. Divers.*, 2015.

[7] C. Huthmacher, A. Hoppe, and S. Bulik, "Antimalarial drug targets in Plasmodium falciparum predicted by stage-specific metabolic network analysis," *BMC Syst.*, 2010.

[8] J. Teahton, "In vitro and in vivo anti-plasmodial activities of senna occidentalis roots extracts against plasmodium falciparum and plasmodium berghei," 2016.

[9] Segun Fatumo, Kitiporn Plaimas, Ezekiel Adebiyi, Rainer Konig (2011): Comparing metabolic network models based on genomic and automatically inferred enzyme information from Plasmodium and its human host to define drug targets in silico, Infection, Genetics and Evolution 11 (2011) 201–208

[10] Roll Back Malaria Partnership, "AIM 2016-2030 — Roll Back Malaria," 2016. [Online]. Available: http://www.rollbackmalaria.org/about-rbm/aim-2016-2030. [Accessed: 10-Jul-2017].

[11] D. A. Goff, R. Kullar, J. G. Newland, L. N, and M. G, "Review of Twitter for Infectious Diseases Clinicians: Useful or a Waste of Time?," *Clin. Infect. Dis.*, vol. 512, no. 10, pp. 126–9, Feb. 2015.

[12] A. J. Lazard, E. Scheinfeld, J. M. Bernhardt, G. B. Wilcox, and M. Suran, "Detecting themes of public concern: A text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat," *Am. J. Infect. Control*, vol. 43, no. 10, pp. 1109–1111, Oct. 2015.

[13] O. M and Y. S, "HIV/AIDS and the Millennium Development Goals: A Public Sentiment Analysis of World AIDS Day Twitter Chat," *Int. J. HIV/AIDS Res.*, pp. 129–132, Nov. 2016.

[14] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?," in *Proceedings of the 19th international conference on World wide web - WWW '10*, 2010, p. 591.

[15] R. J. P. N. N. N. Akshay Ghaisas, *International Journal of Engineering Research and Technology IJERT*, no. Vol.2-Issue 3 (March-2013). ESRSA Publ, 2012.

[16] G. G. Chowdhury, "Natural Language Processing,"

Annu. Rev. Appl. Linguist., vol. 37, no. 1, pp. 51–89, 2003.

[17] D. B. Pisoni and S. V Levi, “Some Observations on Representations and Representational Specificity in Speech Perception and Spoken Word Recognition 2 Conventional View of Speech 2.1 Background,” 2005.

[18] E. Hume and K. Johnson, “A model of the interplay of speech perception and phonology,” *Stud. interplay speech Percept. Phonol.*, vol. 55, pp. 1–22, 2001.

[19] M. Shapiro, “Style,” pp. 203–278, 2017.

[20] K. J. Patterson, “When Is a Metaphor Not a Metaphor? An Investigation Into Lexical Characteristics of Metaphoricity Among Uncertain Cases,” *Metaphor Symb.*, vol. 32, no. 2, pp. 103–117, Apr. 2017.

[21] M. Warsha, M. Choudhari, and M. Rinku Rajankar, “Introduction to Natural Language Processing With Python,” 2015.

[22] A. Tamrakar and D. Dubey, “Query Optimisation using Natural Language Processing,” *Int. J. Comput. Sci. Technol.*, vol. 3, no. 1, 2012.

[23] A.-H. Tan, “Text mining: The state of the art and the challenges,” in *Proceedings of the PAKDD 1999 Workshop on Knowledge Disocovery from Advanced Databases*, 1999, vol. 8, pp. 65–70.

[24] S. Vijayarani, J. Ilamathi, and M. Nithya, “Preprocessing Techniques for Text Mining - An Overview,” *Int. J. Comput. Sci. Commun. Networks*, vol. 5, no. 1, pp. 7–16, 2015.

[25] S. Anshika and G. Udayan, “Text Mining: A Burgeoning technology for knowledge extraction,” *Int. J. Sci. Res. Eng. Technol.*, vol. 1, no. 12, pp. 022–026, 2013.

[26] A. Brahme, “A Review of Knowledge Discovery Using Text Mining and Its Applications,” no. 5, pp. 291–295, 2015.

[27] Y. Zhang and P. Desouza, “Enhance the Power of Sentiment Analysis,” *Int. J. Comput. Autom. Control Inf. Eng.*, vol. 8, no. 3, pp. 7–12, 2014.

[28] E. Cambria, D. Das, S. Bandyopadhyay, and A. F. Editors, *A Practical Guide to Sentiment Analysis*, vol. 5. 2017.

[29] B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis,” *Found. Trends Inf. Retr.*, vol. 2, no. 1–2, pp. 1–135, 2004.

[30] N. N. Yusof, A. Mohamed, and S. Abdul-Rahman, “Reviewing classification approaches in sentiment analysis,” in *Communications in Computer and Information Science*, vol. 545, Springer, Singapore, 2015, pp. 43–53.

[31] R. S. Rahate and Emmanuel M, “Feature Selection for Sentiment Analysis by using SVM,” *Int. J. Comput. Appl.*, vol. 84, no. 5, pp. 975–8887, 2013.

[32] U. Aggarwal and G. Aggarwal, “Sentiment Analysis : A Survey,” *Int. J. Comput. Sci. Eng. Open Access Surv. Pap.*, vol. 5, no. 5, 2017.

[33] M. R. Osama, K. H. Ahmad, and A. A. Q. Dana, “Sentiment analysis as a way of web optimization,” *Sci. Res. Essays*, vol. 11, no. 8, pp. 90–96, Apr. 2016.

[34] A. M. Mostafa, “An Evaluation of Sentiment Analysis and Classification Algorithms for Arabic Textual Data,” *Int. J. Comput. Appl.*, vol. 158, no. 3, pp. 975–8887, 2017.

[35] W. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithms and applications: A survey,” *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014.

[36] K. Khan, B. Baharudin, A. Khan, and A. Ullah, “Mining opinion components from unstructured reviews: A review,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 26, no. 3, pp. 258–275, Sep. 2014.

[37] X. Fang and J. Zhan, “Sentiment analysis using product review data,” *J. Big Data*, vol. 2, no. 1, p. 5, Dec. 2015.

[38] L. Zhang and B. Liu, “Sentiment Analysis and Opinion Mining,” in *Encyclopedia of Machine Learning and Data Mining*, Boston, MA: Springer US, 2016, pp. 1–10.

[39] B. Shreya and P. Rupal, “A brief review of sentiment analysis methods,” *Int. J. Inf. Sci. Tech.*, vol. 6, 2016.

[40] H. Zhou, J. Guo, and Y. Wang, “A feature selection approach based on term distributions.,” *Springerplus*, vol. 5, p. 249, 2016.

[41] X. Ding, S. M. Street, B. Liu, S. M. Street, P. S. Yu, and S. M. Street, “A Holistic Lexicon-Based Approach to Opinion Mining,” *Wsdm’08*, 2008.

[42] M. Hajjouz, “Information and Knowledge Management Opinions Mining in Facebook,” vol. 6, no. 3, 2016.

[43] R. Shouval, O. Bondi, H. Mishan, A. Shimoni, R. Unger, and A. Nagler, “Application of machine

learning algorithms for clinical predictive modeling: a data-mining approach in SCT," *Bone Marrow Transplant.*, vol. 49, no. 3, pp. 332–337, Mar. 2014.

[44] Y. Baştanlar and M. Özuysal, "Introduction to machine learning," in *Methods in Molecular Biology*, vol. 1107, Humana Press, Totowa, NJ, 2014, pp. 105–128.