An Enhanced Feature Acquisition for Sentiment Analysis of English and Hausa Tweets

Amina Imam Abubakar[1]*, Abubakar Roko[2], Aminu Muhammad Bui[3], Ibrahim Saidu[4]
Department of Computer Science, Usmanu Danfodiyo University, Sokoto, Nigeria[1, 2, 3]
Department of ICT, Usmanu Danfodiyo University, Sokoto, Nigeria[4]

***Abstract*—Due to the continuous and rapid growth of social media, opinionated contents are actively created by users in different languages about various products, services, events, and political parties. The automated classification of these contents prompted the need for multilingual sentiment analysis researches. However, the majority of research efforts are devoted to English and Arabic, English and German, English and French languages, while a great share of information is available in other languages such as Hausa. This paper proposes multilingual sentiment analysis of English and Hausa tweets using an Enhanced Feature Acquisition Method (EFAM). The method uses machine learning approach to integrate two newly defined Hausa features (Hausa Lexical Feature and Hausa Sentiment Intensifiers) and English feature to measure classification performance and to synthesize a more accurate sentiment classification procedure. The approach has been evaluated using several experiments with different classifiers in both monolingual and multilingual datasets. The experimental results reveal the effectiveness of the approach in enhancing feature integration for multilingual sentiment analysis. Similarly, by using features drawn from multiple languages, we can construct machine learning classifiers with an average precision of over 65%.**

Keywords*—Multilingual sentiment analysis; sentiment analysis; social media; machine learning

I. Introduction

Social media have turned the web into a vast source of information that is generated by users about all kinds of topics. Twitter is considered one of the most popular and commonly used social media platform [1] where users communicate with each other, share their opinions and express their emotions (sentiments) in the form of convenient short blogs using limited words [2]. Due to the large volume of information, automated approaches that allow users to effectively interact with opinionated content [3] on the internet have been developed [4]. Such approaches form the field of sentiment analysis. Sentiment analysis represents the process of automatically extracting the sentiment orientation or polarity of an opinion on a specific object [5].

The majority of current sentiment analysis systems address a single language, usually English [4] [6]-[13] and analyzing sentiment in a single language increases the risks of missing essential information in texts written in other languages. However, Twitter users express their opinions in different languages such as Arabic, Spanish, German, French, and Hausa. This prompted the need for sentiment analysis systems that discover sentiment from a Twitter document made up of English and one other language. Such systems are called multilingual sentiment analysis systems. These systems are motivated in building sentiment analysis approaches for different languages [14]. While research on multilingual sentiment analysis has been done in several languages e.g English and Arabic [2], English, German, French and Portuguese [15], Italian, Spanish, French and German [16], Hindi, Telgu, and Tamil [17], none has been extended to Hausa language despite the popularity of the language as one of the most spoken language in Africa [18] and therefore, receive little attention in Natural Language Processing (NLP) task. Similarly, lack of NLP application for a language can deny its speakers the potential benefits of NLP technology and information access.

In this paper, multilingual sentiment analysis of English and Hausa tweets using an Enhanced Feature Acquisition Method (EFAM) is proposed. The method uses machine learning approach to integrate English feature and Hausa features to measure classification performance and to synthesize a more accurate sentiment classification procedure.

The main contribution of this study is the development of two newly defined Hausa features; Hausa Lexical Feature (HLF) and Hausa Sentiment Intensifiers (HSI). These features will determine if the frequency of Hausa words and Hausa intensifiers has any effect on a particular sentiment in a multilingual context.

The paper is organized as follows. Section 2 describes the related works, Section 3 discusses the proposed methodology for multilingual sentiment analysis, Section 4 describes the experiment, results and discussion, and Section 5 gives the conclusion and future work.

II. Related Work

Much research have been put into developing approaches for multilingual sentiment analysis of Tweets. These approaches are aimed at creating Twitter sentiment classification models using multiple languages.

The author in [19] proposed the use of emotion tokens for multilingual Twitter messages for English and non-English languages. The polarities of the tokens are labelled automatically based on their popular co-occurrences of emotions. Using a graph propagation algorithm, they construct a graph whose vertices are regular words and emotion tokens while the weight of edges gives a measure of co-occurrence. The comparative evaluations indicate that the emotion tokens are independent of the tweet for both English and non-English Twitter messages and achieve a better performance than the

*Corresponding Author

traditional semantic-based approach [20]. However, the propagation process assigns large positive scores for the majority of the tokens, and that negative scores do not contain many emotion tokens, resulting in a low recall rate on negative scores, especially for the English language.

The author in [15] examined the characteristics and feasibility of a language-independent, semi-supervised sentiment classification approach for tweets and use emoticons as noisy labels to generate training data from a completely raw set of tweets. Class probabilities for the polarities are calculated using logarithmic probabilities. The approach was evaluated in four different languages (English, German, French and Portuguese) that were manually annotated. They used a method similar to [21] to assign noisy polarity class labels to tweets based on the existence of positive or negative emotions. The evaluation performance for each of the 4 languages shows that the approach is less fit to classify some languages because of their structural differences. Therefore, unique impacts of different languages are needed for a proper classification approach.

The author in [16] presented a method to create a sentiment analysis system for tweets in English using tweets from SemEval 2013 [22] as a training and testing dataset. Using the Google machine translation system, the tweets were translated to four other languages; Italian, Spanish, French and German and are manually corrected to create gold standards for each target language. The result shows that the use of all the languages together improves the overall sentiment classification of sentiment in the data. While their system is found effective in the multilingual classification aspect, it, however, cannot eliminate the problem of translation errors due to differences in a language context.

The author in [14] analyzed a large set of manually labelled tweets to train sentiment classifiers in 13 European languages. The performance of these classifiers and the quality of human labelling are performed with the construction of automated classification models. The classification models depend much more on the quality and size of training data than on the type of model trained. While the performance of these models indicates that humans perceived the sentiment classes as ordered, it is, however, limited by the quality of the labelled data used.

The author in [17] proposed a sentiment analysis system of a very famous Indian movie *Baahubali2* using Twitter comments and posts. The authors use Hindi, Telgu, and Tamil languages which are converted to English language using Google translator. A classification algorithm was implemented for all the language datasets and processes each word in a tweet and store the score (positive, very positive, negative, very negative, and neutral) into Hadoop distributed file system. The proposed method effectively demonstrates the relation of positive, very positive, and neutral tweets which are strongly correlated with each other. However, the positive and very positive parts of the tweets are heavily influenced by the noises present in the dataset.

The author in [2] proposed a Vector Space Model (VSM) approach in handling tweets in both Arabic and English languages with different processing techniques applied. This approach is based on using the Term Frequency Inverse Document Frequency (TF-IDF) to generate the feature vector for the classification process. Experiments were performed on five datasets; two in Arabic and three in English and the performance of seven classification algorithms were analyzed. The experimental results reveal the effectiveness of the approach with a higher classification accuracy when applied to the English dataset than Arabic. However, extracting Arabic feature vectors from Arabic WordNet will have served as an additional feature for the Arabic dataset and thus, add classification performance.

Thoughtfully learning the literature, there is no existing work on multilingual sentiment analysis of Hausa language. This study is the first contribution on NLP for Hausa language.

III. Proposed Method for Multilingual Sentiment Analysis

This section describes the research workflow illustrated in Fig. 1 which comprises the dataset used (Twitter multilingual corpus and HWN), pre-processing methods, feature engineering, classification methods, and evaluation. The components are elucidated as follows.

A. Dataset Description

This study makes use of two resources for multilingual sentiment classification: Hausa WordNet lexical resource and Twitter multilingual corpus.

1) Hausa wordnet lexical resource: Hausa is a language spoken by more than 25 million people representing the original Hausa population [23]. The language stretches across the northern states of Nigeria, southern Niger and Hausa communities in Sudan. It is also spoken as a first language by scattered settlements throughout West Africa and as a second language by millions of non-Hausas in northern Nigeria and the northern parts of Benin, Togo, and Ghana [24]. However, despite the popularity of the language, there are no sufficient tools and resources for various NLP applications, hence, the development of Hausa WordNet (HWN). HWN [25] is a lexical resource for the Hausa language which extracts knowledge from a conventional Hausa dictionary and adopts a substructure of English and Hindi WordNets. It groups words based on different categories, introduces pronunciation, and uses close class categories to address the problem of missing pronunciation and coverage from existing WordNets.

2) Twitter multilingual corpus: The corpus for the study comprises Twitter pre-election data collected from a multilingual community (Nigeria). The dataset was collected using tweepy streaming API, preprocessed, and manually annotated by selected human annotators via a web-based interface. The corpus consists of 12,334 tweets which are both monolingual and multilingual. The monolingual tweets comprises: pure English language tweets and pure Hausa language tweets while the multilingual tweets comprise of the combination of English and Hausa tweets as shown in Table I.

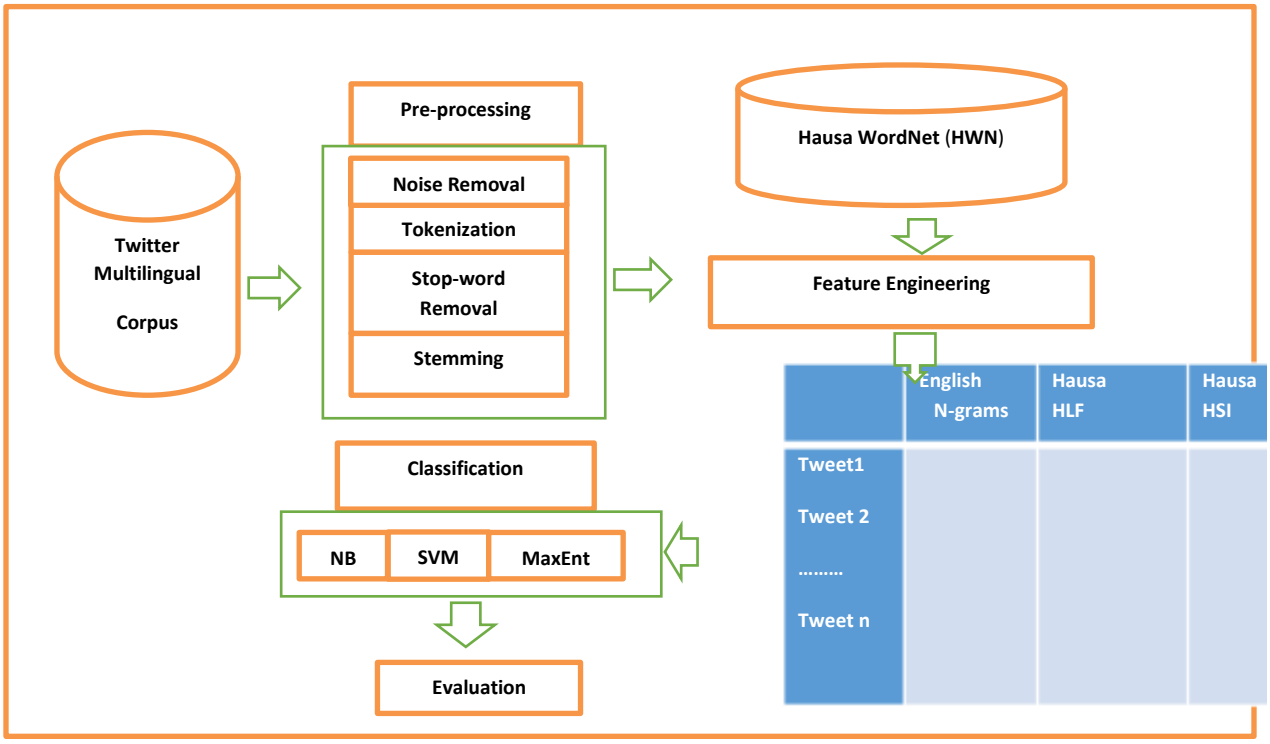
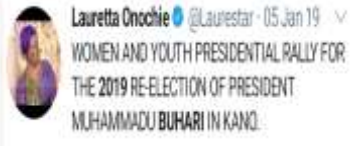


Fig. 1. Multilingual Sentiment Classification Workflow.

TABLE I. EXAMPLE OF MONOLINGUAL AND MULTILINGUAL TWEETS

Tweets	Language Classification
	Pure English (monolingual)
	Pure Hausa (monolingual)
	English and Hausa (multilingual)

B. Pre-Processing

The pre-processing task is an important step in sentiment classification. It is used to remove irrelevant parts from the data, as well as to transform the text to facilitate its analysis. The pre-processing step consists of four steps: noise removal, tokenization, stop word removal and stemming.

1) Noise removal: Removing noise in the text is one of the most essential pre-processing steps though it is highly domain-dependent. For example, Twitter contains a lot of noisy text such as a username (@username), retweets (RT), hashtags (#tag), mentions (@mentions), URL (web pages, web sites), emoticons (icons for smileys) and special characters ($^&*~ etc.). However, the streaming algorithm collected only the tweets without usernames, similarly, the dataset has no emoticons present. As a result, the noise removal process consists of removing all URLs, hashtags, retweets, and special characters from both pure English, pure Hausa, and Multilingual dataset. However, due to the user's constant informal practice of using social media, the only Hausa special character present is " ' " such as in "*ta'adda"*, " *'yanci"* and *"jama'a"*. Therefore, this character together with English special characters is removed.

2) Tokenization: Tokenization is the process of breaking a stream of text up into words, phrases, symbols or other meaningful elements called tokens. The list of tokens becomes an input for further processing such as sentiment classification. This process is very important in sentiment analysis of social media text because sentiment information can be sparsely and unusually represented [26]. The study implements word-level tokenization for all the three datasets. Each tweet will now be tokenized and split into words where each word needs to be captured and subjected to further pre-processing like stop word removal and stemming.

3) Stop-word removal: The removal of stop words in sentiment classification is necessary as the general idea of a text is retained in the absence of these words and also adds quality to the model. For the pure English dataset, NLTK was used in removing all English language stop-words, frequent words such as *the, and, of, a* and *is* that are present in almost

all tweets were removed. Whereas, in the pure Hausa dataset, stop words such as "*amma", " wannan", "yi", "za", "wata", "kuma", " cikin"*…, etc. were removed and this was done by searching for words in a pre-existing list of Hausa stop-words [1] . Similarly, for the multilingual dataset, the combination of pre-existing English and Hausa stop-words are filtered out from the dataset as they carry less discriminative power in analyzing sentiment in a multilingual context.

4) Stemming: Stemming is important in NLP as compound words are replaced by their morphological root. Hausa language stemmer [27] from the NLTK python library was used to perform the Hausa stemming process, for example, the word '*yan-maza* and *mazaje* share the same root word as *maza*. While for the English dataset, the porter English stemmer from NLTK python library was used to perform the word stemming process, for example, the words "*tester*", "*testing*", and "*tested*" all share the same root-word "*test*". At the end of this preprocessing phase, we obtain a set of a stemmed bag of words that represents the original feature vector.

C. Feature Engineering Approach

The study adopts two approaches for multilingual sentiment classification; the first approach is building a baseline method using the corpus to generate n-gram features. Whereas the second approach is improving the baseline approach using Term Frequency to generate the Hausa feature vectors for the classification process. The main objective in this approach is to determine whether the features introduced from Hausa language can improve sentiment classification accuracy.

1) The baseline approach: This approach uses Twitter's multilingual corpus to develop the English feature i.e. n-gram features as the baseline approach for sentiment classification.

a) N-gram Feature: An n-gram is a contiguous sequence of n words from a given piece of text. Typically, n-grams are the basic features used in supervised sentiment classification. The *n*-gram features used in the study are weighted unigrams, bigrams, and trigrams. The study did not explore higher-order *n*-grams to try and minimize the negative effect of high dimensionality.

Weighted n-gram features are generated from our datasets so as to assign weights to each gram in the feature vector to indicate their importance. Term Frequency-Inverse Document Frequency (TF-IDF) technique was implemented to generate weighted features. This technique is a statistical measure used to evaluate how important a word is to a document or corpus. The importance increases proportionally to the number of times a word appears in a document but is offset by the frequency of the word in the corpus [2].

TF-IDF is composed of two terms: Term Frequency (TF) and Inverse Document Frequency (IDF). TF measures how frequently a term (t) occurs in a document.

$$TF\ (t)=\frac{Number\ of\ term\ (t)\ in\ tweet}{Max\ (\ occurence\ of\ terms\ in\ tweets)} \quad (1)$$

IDF measures how important a term is.

$$IDF(t) = \frac{Total\ Number\ of\ tweets}{Number\ of\ tweets\ with\ term\ (t)} \quad (2)$$

Finally, the TF-IDF measure the product of TF and IDF, as follows

$$TF - IDF\ (t) = Tf\ (t) * IDF\ (t) \quad (3)$$

This means that larger weights are assigned to terms that appear relatively infrequent throughout the corpus, but very frequently in individual documents/tweets.

2) Developing hausa features: Developing newly defined features is an important task in sentiment classification and more generally in text classification. Thus, researchers along this line argue that selecting the right feature determines the overall performance of sentiment classification. To this end, the study developed the following Hausa features.

a) Hausa Lexical Feature (HLF): These are Hausa features generated from the co-occurrence of common words from both Hausa lexical resource (HWN) and multilingual twitter corpus. Therefore any word that occurs in both HWN and the corpus is now term as Hausa word. These common words will help us identify Hausa words from the multilingual tweets. Therefore, the feature vector is generated by finding the term frequency (*TF*) of Hausa words (*hw*) in a tweet and can be represented as follows:

$$TF(hw) = \frac{Number\ of\ hw\ in\ tweet}{Max\ (\ occurence\ of\ hw\ in\ tweets)} \quad (4)$$

$$HLF = TF(hw) \quad (5)$$

This approach will normalize the distribution of Hausa words in the corpus. Some users express themselves in Hausa and English when they are sad, angry or frustrated and some otherwise as shown in the following examples.

Example 1: To hell with this government babu komai sai zalunci.

Example 2: PMB is purely direct, bai iya manufunci ba, part of the reason why I support him.

Example 1 has a frequency of 4 Hausa words (*babu, komai, sai, and zalunci*), similarly, example 2 has 4 Hausa words (*bai, iya, manufunci, ba*). Therefore, this feature will determine if the frequency of Hausa words has any effect on a particular sentiment in a multilingual context.

b) Hausa Sentiment Intensifiers (HSI): These are Hausa words that emphasize or intensify sentiment. The study makes use of a dictionary of Hausa intensifiers developed purposely for this study. These intensifiers are generated from Hausa words and then manually annotated (as either positive, negative or neutral intensifiers) with a substantial inter-annotator agreement (Kappa= 0.8). The annotation exercise was conducted by 3 experts in the field of sentiment analysis, Hausa language, and Linguistics who were also proficient speakers of both English and Hausa languages and also able to

[1] https://github.com/stopwords-iso/stopwords-ha

comprehend social media contents. The resulted annotated words are then compared with their existing meaning from HWN to further verify their intensity. Therefore, this feature vector is generated by finding the term frequency of Hausa Positive Intensifiers *TF* (*hpi*), term frequency of Hausa Negative intensifiers *TF (hngi*), and term frequency of Hausa Neutral Intensifiers *TF (hni)* and normalise by their maximum occurrences in the document (tweets). This approach can be represented as follows:

$$TF(hpi) = \frac{Number\ of\ hpi\ in\ tweet}{Max\ (\ occurence\ of\ hpi\ in\ tweets)} \quad (6)$$

$$TF(hngi) = \frac{Number\ of\ hngi\ in\ tweet}{Max\ (\ occurence\ of\ hngi\ in\ tweets)} \quad (7)$$

$$TF(hni) = \frac{Number\ of\ hni\ in\ tweet}{Max\ (\ occurence\ of\ hpi\ in\ tweets)} \quad (8)$$

Table II shows some examples of Hausa words and their sentiment intensification.

TABLE II. SOME HAUSA SENTIMENT INTENSIFIERS

s/n	Hausa Words	Sentiment Intensity
1	Da kyau	Positive
2	Dodar	Positive
3	MashaAllah	Positive
4	Tayani	Neutral
5	Kajifa	Neutral
6	Anya	Negative
7	Tabdijam	Negative
8	Tirkashi	Negative

These words are clear indicators of sentiment when express in a context as shown in the following examples:

Example 3: PMB for the second tenure, anya kuwa?

Example 4: MashaAllah, the Kano rally was conducted, da kyau.

Example 3 has a frequency of only 1 negative intensifier (anya) while example 4 has a frequency of 2 positive intensifiers (mashaAllah, da kyau). Therefore, HSI feature was implemented to determine whether Hausa intensifiers have any effect on a particular sentiment in a multilingual context.

D. Machine Learning Methods: The Classification Algorithm

The pre-processed datasets were split into training (70%) and testing (30%) data. The training data are processed by the classification algorithms in Scikit-learn machine learning in Python [28]. A Tfidf Vectorizer is implemented on the datasets using sublinear_tf to reduce the bias generated by words that appear frequently. Extracted features from HLF were then appended with the baseline features (N-gram) to the training and validation data. Similarly, extracted features from HSI were appended with HLF and the baseline feature to the training and validation data using Numpy.

The vectors were trained on Naive Bayes (MultinomialNB), Support Vector Machine (SVM) and Maximum Entropy (MaxEnt) classifiers. The SVM regularization parameter (C) is set to 1e5 or 100000 (the larger the C the better the validation). MaxEnt regularization parameter (C) is set to 128 (smaller values specify stronger regularization) and then a prediction was generated, the accuracy of the prediction was then tested using accuracy_score.

These classification algorithms were used due to their simplicity, effectiveness and accurateness in a supervised learning classification process. As for the test data, the classification algorithms corresponding to the built model is applied, and thus classification results are obtained. A brief background about these classifiers is presented.

1) Naïve bayes classifier: Naïve Bayes (NB) is a probabilistic classifier that operates by building statistical models of classes from the training dataset. The study make use of a Naive Bayes model class c to represent a class and x for features calculated individually as shown in the formula.

$$P\ (c\backslash x) = \frac{P\ (x\backslash c)P\ (c)}{P\ (x)} \quad (9)$$

2) Support vector machine: Support Vector Machine (SVM) is highly effective and generally outperforms other classifiers at sentiment classification [26]. SVMs utilize hyperplanes to separate classes and seek a decision hyperplane represented by a support vector that separates the positive and negative training vectors of documents with maximum margin.

3) Maximum entropy: Maximum Entropy (MaxEnt) is a feature-based classifier that work on the idea that the most uniform model that satisfies a given constraint should be preferred [28]. In a two-class scenario, it is the same as using logistic regression to find a distribution over the classes. The model is represented by the following.

$$P\ \left(\frac{c}{d}, \lambda\right) = \frac{\exp(\Sigma_i \lambda_i\ f_i\ (c,d))}{\Sigma_d\ \exp\ (\Sigma_i\ \lambda_i\ f_i(c,d))} \quad (10)$$

The above classification algorithms were used to evaluate the feature set.

N-gram: Classifiers that evaluates only n-gram features.

N-gram + HLF: Classifiers that evaluates n-gram and HLF features.

N-gram + HLF + HSI: Classifier that evaluates n-gram, HLF and HSI features.

IV. EXPERIMENT

In this section, dataset characteristics, system setup, evaluation criteria and the results for evaluating the performance of the proposed approach is discussed.

1) Dataset characteristics: Several experiments were conducted on the Twitter multilingual corpus. The corpus comprises 12,405 tweets and each tweet has sentiment annotations on tweet level by 2 human annotators using sentiment classes positive, negative and neutral. The annotators' classes were aggregated to assign a sentiment to tweet, where tweet *t* has sentiment *S* if 2 annotators marked

the tweet with *S*; otherwise, the sentiment of *t* is conflicted. Similarly, if the sentiment of *t* is conflicted then t will be discarded.

$$t=S \text{ if } S1=S_{2 \text{ else}} \; t=\text{conflicted} \quad (11)$$

$$\text{If } t=\text{conflicted then } t=\text{discard} \quad (12)$$

Therefore, after removing all discarded tweets, the corpus now comprises 12,334 tweets; 4,623 of them considered as positive tweets, 6,531 as negative tweets, and the other 1,180 as neutral opinions. Similarly, the tweets are both monolingual and multilingual; the monolingual tweets comprise 1- Pure English language which comprises 10,900 tweets 2- Pure Hausa language with 244 tweets while the multilingual tweets comprise of the combination of English and Hausa language tweets with 1,190 tweets as shown in Table III.

2) System setup: The experimental setup was implemented with the following tools and environment:

- Windows 10 operating system.
- System specification of 30GB Hard disk space, 6GB RAM, and Intel ® core ™ processor @ 2.40GHz.
- Python programming language using Jupyter notebook is deployed which provides an easy and fast modelling workspace for the experiment.

3) Evaluation criteria: To evaluate the quality and usefulness of the classifiers and to efficiently integrate the feature set to synthesize a more accurate classification procedure, experimental results were sorted into the following: accuracy, precision, recall and F1. The contingency table below illustrates the arrangement of actual and predicted classification in a three-class problem (positive, negative, and neutral).

Table IV reports the counts of True Positives (TP), False Positives (FP), True Negative (TN), and False Negatives (FN) which are defined as follows:

- TP (A): TP is the number of positive tweets correctly classified as positive.
- FP (D + G): FP is the number of negative tweets falsely classified as positive.
- TN (E + I): TN is the number of negative tweets correctly classified as negative.
- FN (B + C): FN is the number of positive tweets falsely classified as negative.

Precision (P) for the three classes, positive, negative, and neutral is determined as follows:

$$P_{(Positive)}=A/(A+D+G) \quad (13)$$

$$P_{(Negative)}=E/(B+E+H) \quad (14)$$

$$P_{(Neutral)}=I/(C+F+I) \quad (15)$$

Recall (R) for the three classes, positive, negative, and neutral is determined as follows:

$$R_{(Positive)}=A/(A+B+C) \quad (16)$$

$$R_{(Negative)}=E/(D+E+F) \quad (17)$$

$$R_{(Neutral)}=I/(G+H+I) \quad (18)$$

The F1 for a class is given by the harmonic mean of the class precision and recall as follows:

$$F1=2TP/(2TP+FP+FN) \quad (19)$$

Similarly, accuracy is the number of correct predictions from all predictions made.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (20)$$

$$Accuracy = \frac{A+(E+I)}{A+(E+I)+(D+G)+(B+C)} \quad (21)$$

4) Experimental result and discussion: The performance of the classification algorithms using the feature set is analyzed. The Tables below represent the performance of the classification algorithms for the three classes evaluated by the metrics: accuracy, precision, recall and F1-score. The results were also classified based on datasets and each dataset was categorized based on the feature set used. However, the pure English dataset was evaluated using only the N-gram feature as there are no Hausa words in the dataset, and instantiating.

Hausa features in a non-Hausa dataset will yield an erroneous classification.

From Tables V, VI and VII above, bold font indicates the best performance on a dataset, and asterisk, *, indicates significant difference from the baseline, Ngram. The pure English dataset uses only the baseline features (Ngram) and achieves a little accuracy of 56% with SVM classifier. For the pure Hausa dataset, the best result is obtained using SVM classifier with an accuracy of 71% and for multilingual dataset the best result is obtained with Naïve Bayes classifier with an accuracy of 68%.

TABLE III. NUMBER OF CLASSES PER LANGUAGE

Classes	English	Hausa	Multilingual	Total
Positive	4,134	143	346	4,623
Negative	5,794	73	664	6,531
Neutral	972	28	180	1,180
Total	10,900	244	1,190	**12,334**

TABLE IV. CONTINGENCY TABLE

Classification	Positive	Negative	Neutral
Positive	**A**	**B**	**C**
Negative	**D**	**E**	**F**
Neutral	**G**	**H**	**I**

TABLE V. RESULTS FROM THE PURE ENGLISH TEST DATASET

Algorithm	Accuracy (%)	Positive (%) P R F1	Negative (%) P R F1	Neutral (%) P R F1
Naïve Bayes Ngram	55	49 82 61	66 59 62	29 02 04
SVM Ngram	56	52 77 62	63 62 63	43 12 19
MaxEnt Ngram	55	52 77 62	62 63 62	31 06 10

TABLE VI. RESULTS FROM THE PURE HAUSA TEST DATASET

Algorithm	Accuracy (%)	Positive (%) P R F1	Negative (%) P R F1	Neutral (%) P R F1
Naïve Bayes Ngram	50	45 80 58	64 24 35	52 50 51
Ngram+HLF	53	56 68 61	39 61 48	75 38 50
Ngram+HLF+HSI	**64***	58 70 64	73 73 73	65 50 57
SVM Ngram	57	59 64 62	60 52 56	54 58 56
Ngram+HLF	60	62 64 63	44 67 53	77 53 63
Ngram+HLF+HSI	**71 ***	74 67 70	65 77 71	72 69 71
MaxEnt Ngram	59	57 64 60	65 52 58	56 62 59
Ngram+HLF	60	58 64 61	48 67 56	77 53 63
Ngram+HLF+HSI	**68***	69 67 68	64 73 68	71 65 68

TABLE VII. RESULTS FROM THE MULTILINGUAL TEST DATASET

Algorithm	Accuracy (%)	Positive (%) P R F1	Negative (%) P R F1	Neutral (%) P R F1
Naïve Bayes Ngram	64	69 56 62	65 62 64	60 77 67
Ngram+HLF	**68***	63 71 66	69 68 69	75 67 71
Ngram+HLF+HSI	61	61 63 62	64 50 56	61 72 66
SVM Ngram	65	72 54 62	60 70 65	66 72 69
Ngram+HLF	**66**	62 67 64	68 64 66	69 68 68
Ngram+HLF+HIS	**66**	66 63 65	62 62 62	66 68 67
MaxEnt Ngram	64	71 55 62	63 66 64	61 74 67
Ngram+HLF	**66**	66 61 63	70 63 66	63 73 68
Ngram+HLF+HSI	65	69 63 66	65 60 63	62 73 67

The use of all feature set (Ngram+HLF+HSI) provides the best classification accuracy on pure Hausa dataset as shown in Fig. 2 while achieves little to no improvement on the multilingual dataset, this can be due to the higher frequency of Hausa words and Hausa intensifiers in Hausa dataset compared to the multilingual dataset. Similarly, the feature set (Ngram+HLF) provides the best classification accuracy and on the multilingual dataset.

The developed Hausa features from the two datasets (Hausa dataset and multilingual dataset) improve the accuracy of the baseline (with the exception of naïve Bayes classifier on multilingual dataset) using the 3 classifiers. Similarly, SVM is the best classification algorithm for all the datasets with 71% as shown in Fig. 3.

Furthermore, since there is no any existing work in Hausa language for direct comparison, the obtained Hausa dataset result is compared with result from Arabic dataset [2] using Arabic Sentiment Tweet Dataset (ASTD), we find the 2 results having equal accuracy of 68% when apply to Maximum Entropy (Logistic regression) while ASTD has a higher accuracy when apply on Naïve as shown in Table VIII. However, the SVM classifier in the proposed approach improves the accuracy to 71% and this can be due to its optimal margin gap between separating hyperplanes, thus, it is more robust in classification approaches.

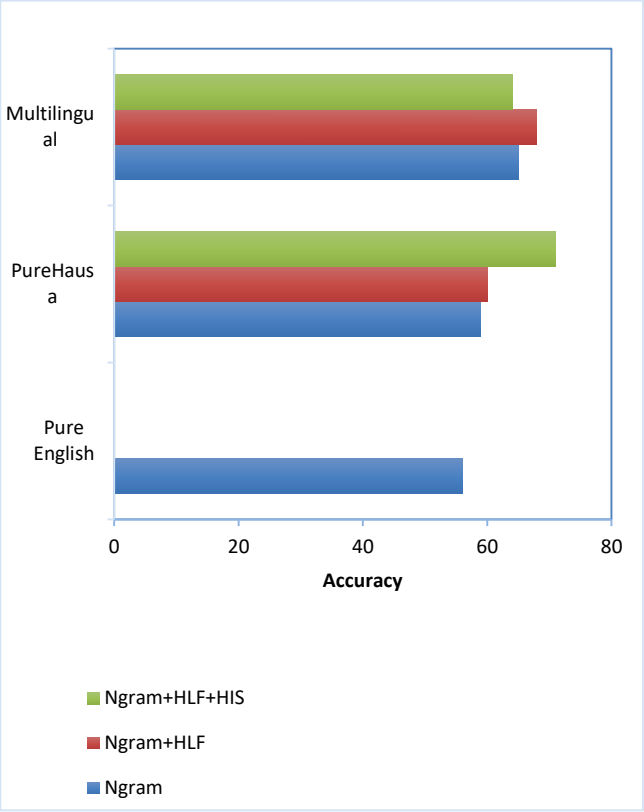


Fig. 2. Feature Set Accuracy Performance on Dataset.

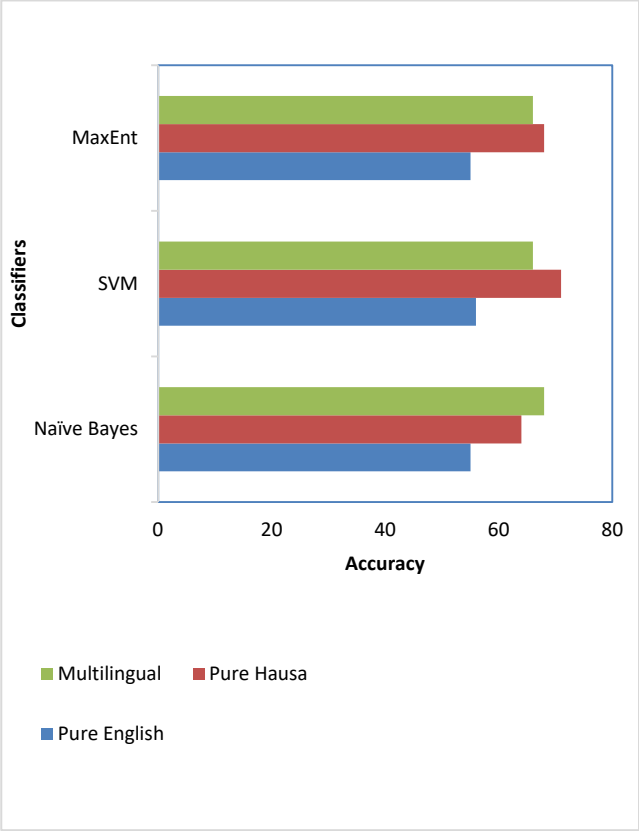


Fig. 3. Dataset Accuracy Performance on Classifiers.

TABLE VIII. ACCURACY COMPARISON USING HAUSA AND ARABIC DATASET

Dataset	Classifier	Accuracy
Proposed Approach		
Hausa	Naïve Bayes	64
	SVM	71
	MaxEnt	68
Elhadad et al., 2019		
Arabic [2]	Naïve Bayes	67
	SVM	NA
	Logistic Regression	68

V. CONCLUSION AND FUTURE WORK

The study proposed multilingual sentiment analysis of English and Hausa tweets using an Enhanced Feature Acquisition Method (EFAM). The method uses feature integration originating from two languages (English and Hausa) into a machine learning approach to multilingual sentiment analysis. We show that an enriched feature set provides effective modelling for sentiment classification of social media text. We achieved better classification performance using an SVM classifier and the use of all feature set (Ngram+HLF+HSI) provides the best classification accuracy on pure Hausa dataset while feature set (Ngram+HLF) provides the best classification accuracy on the multilingual dataset. Similarly, the results demonstrated that each of the newly defined feature set improves sentiment classification performance.

The pitfall of this study is that Term Frequency and Term frequency/Inverse document Frequency serve as a lexical level feature and thus tend to ignores the syntax and semantic of text. For future work, there are many avenues to pursue, including: 1- The use of statistical significant test such as T-test or ANOVA as supplementary technique. 2- Extending the proposed system against various languages other than Hausa or English. 3- Performance can be improve using a deep learning approach to automatically learn high-level features from the dataset by encoding sentiment information using Hausa and English word embedding methods.

REFERENCES

[1] S. L. Lo, E. Cambria, R. Chiong, and D. Cornforth, “Multilingual sentiment analysis : from formal to informal and scarce resource languages,” Artif. Intell. Rev., vol. 48, pp. 499–527, 2017.

[2] K. M. Elhadad, L. Fun, and G. Fayez, “Sentiment Analysis of Arabic and English Tweets,” in Web, Artificial Intelligence and Network, B. L, T. M, X. F, and E. T, Eds. Springer, Cham, 2019, pp. 334–348.

[3] N. Yadav, O. Kudale, S. Gupta, A. Rao, and A. Shitole, “Twitter Sentiment Analysis using Machine learning for Product Evaluation,” in IEEE nternational Conference on Inventive Computation Technologies (ICICT), 2020, pp. 181–185.

[4] A. Bakliwal, F. Jennifer, van der P. Jennifer, O. Ron, T. Lamia, and H. Mark, “Sentiment Analysis of Political Tweets: Towards an Accurate Classifier,” in Proceedings of the Workshop on Language in Social Media, 2013, pp. 49–58.

[5] B. Liu, "Synthesis Lectures on Human Language Technologies," Morgan and Claypool Publishers, 2012.

[6] A. Go, R. Bhayani, and L. Huang, “Twitter Sentiment Classification using Distant Supervision,” Standfort, 2009.

[7] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-Based Methods for Sentiment Analysis," Assoc. Comput. Linguist., vol. 37, no. 2, pp. 267–307, 2011.

[8] J. M. Chenlo and D. E. Losada, "A Machine Learning approach for Subjectivity Classification based on Positional and Discourse Features," in Multidiscilinary Information Retrieval, vol. 8201, L. M, K. E, and O. Loizides, F, Eds. Springer, Berlin, 2013, pp. 17–28.

[9] G. Vaitheeswaran and L. Arockiam, "Combining Lexicon and Machine Learning Method to Enhance the Accuracy of Sentiment Analysis on Big Data," Int. J. Comput. Sci. Inf. Technol., vol. 7, no. 1, pp. 306–311, 2016.

[10] M. Zubair, A. Khan, S. Ahmad, M. Qasim, and A. K. Khan, "Lexicon-enhanced sentiment analysis framework using rule-based classification scheme," PLoS One, pp. 1–22, 2017.

[11] A. Gupta, J. Pruthi, and N. Sahu, "Sentiment Analysis of Tweets using Machine Learning Approach," Int. J. Comput. Sci. Mob. Comput., vol. 6, no. 4, pp. 444–458, 2017.

[12] S. Joshi and D. Deepali, "Twitter Sentiment Analysis Systems," Int. J. Comput. Appl., vol. 180, no. 47, pp. 35–39, 2018.

[13] F. Zarisfi, S. Faramarz, and E. Esfandiar, "Solving The Twitter Sentiment Analysis Problem Based on a Machine Learning Based-Approach," J. Evol. Intell., vol. 13, pp. 381–398, 2020.

[14] I. Mozetic, J. Smailovi, and M. Grcar, "Multilingual Twitter Sentiment Classification : The Role of Human Annotators," PLoS One, vol. 11, no. 5, pp. 1–26, 2016.

[15] S. Narr, H. Michael, and S. Albayrak, "Language-Independent Twitter Sentiment Analysis," in In Proceedings of the Knowledge Discovery and Machine Learning., 2012.

[16] A. Balahur and M. Turchi, "Multilingual Sentiment Analysis using Machine Translation," Proc. 3rd Work. Comput. approaches to Subj. Sentim. Anal. Assoc. Comput. Linguist., pp. 52–60, 2012.

[17] N. Suri and T. Verma, "Multilingual Sentiment Analysis on Twitter dataset using Naive Bayes Algorithm," Sch. J. Eng. Technol., vol. 5, no. 9, pp. 473–477, 2017.

[18] A. S. Muhammad, M. M. Aliyu, and S. I. Zimit, "Towards the Development of Hausa Language Corpus," Int. J. Sci. Eng. Res., vol. 10, no. 10, pp. 1598–1604, 2019.

[19] A. Cui, M. Zhang, Y. Liu, and S. Ma, "Emotion Tokens : Bridging the Gap among Multilingual Twitter Sentiment Analysis," in Information Retrieval Technology, no. 7097, M. Saleem, K. Shaalan, F. Oroumchia, A. Shakery and H. Khalalfa, Eds. Springer, Berlin, 2011, pp. 238–249.

[20] S. Baccianella, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in Proceedings of the Seventh International Conference on Language Resource s and Evaluation (LREC 10), 2010, pp. 2200–2204.

[21] A. Pak and P. Patrick, "Twitter as a corpus for sentiment analysis and opinion mining," in Proceedings of the Seventh International Conference on Language resources and Evaluation (LREC' 10), 2010, pp. 1320–1326.

[22] P. Nakov, S. Rosenthal, Z. Kozareva, V. Stoyanov, A. Ritter, and W. Theresa, "SemEval-2013 task 2: Sentiment analysis in twitter," in Proceedings of the International Workshop on Semantic Evaluation, SemEval 13, 2013, pp. 312–320.

[23] B. Comrie and B. Comrie, "Hausa and the Chadic Languages," in The Worlds's Major Languages, Third., Taylor & Francis Group, 2018.

[24] R. M. Newman and P. Newman, "The Hausa Lexicographic Tradition," African Journals Online, vol. 11, pp. 263–286, 2001.

[25] A. Imam, A. Roko, A. Muhammad, and I. Sa'id, "Hausa WordNet : An Electronic Lexical Resource," Saudi J. Eng. Technol., vol. 4, no. 8, pp. 279–285, 2019.

[26] B. A. Muhammad, "Contextual Lexicon-based Sentiment Analysis for Social Media," Robert Gordon University, 2016.

[27] A. Bimba, I. Norisma, K. Norazlina, N. Nur, and L. Valiukas, "Stemming Hausa Text: Using affix-rulesmand reference lookup to stem words in Hausa language," Lang. Resour. Eval., vol. 50, no. 3, pp. 687–703, 2016.

[28] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.