# Metaphor Detection in Nigerian Poetry Using Machine Learning Techniques

[1]L. U. Bashir, [2]A. O. Olamiti and [3]I. T. Ayorinde
Department of Computer Science,
University of Ibadan,
Ibadan,
Nigeria.
*Emails: [1]latifahusain8@gmail.com; [2]aolamiti@yahoo.com ; [3]temiayorinde@yahoo.com*

**ABSTRACT**

***Metaphor detection in Nigerian poetry presents a unique challenge due to the rich cultural diversity and nuanced linguistic expressions inherent in Nigerian society. This study explores computational models for metaphor detection tailored to the complexities of Nigerian metaphors, aiming to bridge the gap in existing literature and foster a deeper understanding of cultural nuances in language. A labeled dataset comprising of 510 rows and an unlabelled dataset of 14,994 rows are utilized, with careful pre-processing and augmentation techniques employed to enhance data quality and model performance. Four machine learning algorithms which are Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR), and K Nearest Neighbour (KNN) Classifier are evaluated using both supervised and self-training semi-supervised learning approaches. The models are assessed based on accuracy, precision, recall, and F1-scores. For the supervised learning approach, SVM demonstrated the highest accuracy of 96% on the test set while for the self-training semi-supervised learning approach, logistic regression demonstrated the highest accuracy of 93% on the test set. The results of this study suggested that using more sophisticated imputation methods, collecting a larger and more diverse dataset with collaboration from linguists and cultural experts, validating the model across different cultural contexts, and exploring alternative machine learning algorithms will enhance the accuracy, reliability, and applicability of metaphor detection models in Nigerian poetry and broader Natural Language Processing (NLP) applications.***

**Keywords**: *Metaphor detection, Nigerian poetry, Supervised learning, Self-training semi-supervised learning*

**African Journal of Computing & ICT Reference Format:**

L. U. Bashir, A. O. Olamiti and I. T. Ayorinde (2024), Metaphor Detection In Nigerian Poetry Using Machine Learning Techniques, *Afr. J. Comp. & ICT*, Vol.17, No.2, pp. 1 - 23.

## 1. Introduction

Metaphors are used in language to make abstract concepts visualisable and to express ideas through implicit objects that relate to our environment and mental state [1]. Metaphors are expressions that draw a comparison between two dissimilar things, as distinguished from simile, an explicit comparison signaled by the words 'like' or 'as' [2]. They are pervasive and powerful linguistic devices that convey abstract, complex, or novel concepts by mapping them into more concrete, familiar, or simple ones [3]. Poetry on the other hand is an artistic form of expression that serves as a rich cultural reservoir that encapsulates the emotions, experiences, and identity of diverse communities [4]; it is one literary piece that has the usage of metaphor in abundance.

Metaphors have been a fundamental tool in poetry since ancient times, allowing poets to create vivid imagery, convey intricate emotions, and delve into abstract concepts. Some of the earliest instances of metaphors can be traced back to Homer's works like the Iliad and Odyssey where gods and heroes are described using metaphorical attributes such as Athena being "bright-eyed" or Achilles being "swift-footed" [5]. It continued to be embraced by poets across different cultures and eras; for example, Chinese poet Li Bai compared the life to "the spring wind" in his poem Before The Cask of Wine [6], while English poet William Shakespeare compared his love to "a summer's day" in "Sonnet 18" [7], and Emily Dickinson, compared hope to a bird with feathers in her poem "Hope is the thing with feathers" [8]. And even contemporary poets like the likes of Maya Angelou employed metaphors comparing "a caged bird" to a person who is oppressed in her poem, Caged Bird [9].

As metaphor intricately woven into the fabric of global literature, it plays a significant role in African and Nigerian poetry as well as an ancestral form of expression. People and poets utilize metaphors to articulate views on various facets of Nigerian society, culture, politics, and identity. Some examples of prominent Nigerian poets who use metaphors in their works are Wole Soyinka, Chinua Achebe, Niyi Osundare, J.P. Clark and many more others. For instance, in his poem titled Random Blues, Osundare depicts politicians as "a rag of leeches and lice," symbolizing their enrichment at the expense of citizens [10]. In Achebe's poem, Answer, the phrase "Crouching shadows" represents abandoned culture [11]. Similarly, Wole Soyinka characterizes corruption with cockroaches in his poem "Conversation at Night with Cockroach" [12].

Metaphors in Nigerian poetry encompass a range of categories and functions. In Otobotekere's poetry, metaphors are used to create a positive image of the Niger Delta ecology, with conceptual metaphors such as "The Niger river is a parent" and "Niger River Flow Is Entertainment" [13]. Stephen Kekeghe's Rumbling Sky reflects and represents the socio-political and religious vices in contemporary Nigeria, employing conceptual metaphors to capture these vices [14]. Niyi Osundare's poetry utilizes various types of metaphors, including animal metaphors, synthetic metaphors, organic metaphors, and telescoped metaphors, to question socio-political issues in Nigerian society [10]. Soyinka, Osundare, and Olafioye employ metaphors of corruption in their works to highlight the destructive nature of corrupt public officials and the need for cultural reorientation [15].

We could see from all these that metaphors remain an essential part of poetry and literary works. Understanding and interpreting them can vary across context, languages and cultures. With Natural Language Processing (NLP) tasks, researchers are constantly exploring ways to teach computers tounderstand and interpret languages in a more nuanced way and metaphor detection has been the effort they needed that could help improve various NLP applications, such as text understanding, information extraction, sentiment analysis and machine translation [16][17].

Metaphor detection is tasked to identify and analyse metaphorical expressions in a text and can be employed to automatically identify and analyse these metaphors by training models on annotated datasets. The basic meaning of words can be modelled and compared with their contextual meaning to detect metaphors. This approach aligns with the metaphor identification procedure (MIP), which emphasizes the contrast between contextual and basic meanings [18].Several studies have explored the application of machine learning techniques in detecting metaphors in various forms of literature and NLP. For instance, [19] explore metaphor detection in German expressionist poetry using computational methods like logistic regression, support vector machines, and neural networks but when it comes to Nigerian poetry specifically, there is a lack of research in this area.

The existing studies on metaphor detection in poetry often face limitations such as small datasets, incomplete literal annotations in existing datasets and reliance on word embedding when trained on low resource language data which might not fully represent language use and generalizability of the model due to cognitive cultural sensitivity. It is important to address these limitations and develop a robust metaphor detection framework specifically tailored to Nigerian poetry. By combining representation learning, supervised and unsupervised learning techniques, researchers can create a comprehensive framework for metaphor detection in Nigerian poetry; and utilized contextual word embeddings with external resources like dictionaries. Overall, machine learning can enhance the understanding and analysis of metaphors in Nigerian poetry, shedding light on the cultural and social nuances embedded within the language [20] Hence, this study is aimed at filling the above gap by creating a model that teaches computers, the Nigerian language in a more nuanced way (metaphorical way).

## 2. Related Works

[21] undertook a significant endeavour to enhance automatic metaphor detection and classification using deep contextualized word embeddings, bidirectional Long Short-Term Memories (LSTMs), and a multi-head attention mechanism. Their research was aimed at improving prediction accuracies and overcome training data shortages in modern deep learning-based NLP systems. Employing an end-to-end approach, they integrated sophisticated techniques to eliminate the need for complex, hand-crafted feature pipelines. Leveraging benchmark datasets such as TroFi and MOH-X, their method achieved notable performance metrics, including precisions ranging from 85.3% to 90.7%, recalls from 74.8% to 84.9%, F1-scores from 79.8% to 83.2%, and accuracies between 80.7% and 85.8%.

However, they acknowledged the scarcity of training data and proposed future extensions to other languages, additional information incorporation like POS or named-entity tags, and exploration of related tasks such as irony detection and sentiment analysis.

[22] addressed the optimal level of abstraction of semantic representations required for capturing and generalizing metaphorical mechanisms. Their research focused on computational processing of metaphors in NLP applications, employing an SVM classifier to predict metaphorical or literal expressions. Utilizing large-scale attribute-based semantic representations for metaphor identification, their method achieved F1-scores ranging from 0.73 to 0.77 on the TSV-TEST dataset, which comprised 100 literal and 100 metaphorical pairs. However, their study did not explore the generalizability of the proposed method to other languages or domains, relied on the availability of attribute-based semantic representations, and did not investigate the interpretability of the learned representations.

[22] investigated the stylistic and communicative functions of animal metaphors in the Yoruba language, aiming to understand their cultural significance and underlying motivations. He employed two-dimensional approach that encompasses stylistic and cultural dimensions, and then decomposed the semantic features of animals involved in metaphors into High Priority Semantic Markers (HPSM) and Low Priority Semantic Markers (LPSM). While the study did not involve specific datasets or machine learning models, it provided qualitative analyses, cultural insights, and stylistic interpretations of animal metaphors within the Yoruba cultural context. Limitations of the study included lack of quantitative analysis and cultural bias, which could be addressed to enhance the robustness and applicability of future research endeavours.

[24] proposed a novel unsupervised method for adjective-noun metaphor detection on low resource languages, Middle High German. They trained a neural network model using the combination of supervised and unsupervised learning methods. For the Supervised Learning aspect, the researchers initially trained a feed-forward neural network in a supervised manner to detect metaphors in adjective-noun pairs. This involved maximizing the cosine distance between word vectors of metaphorical pairs and minimizing the distance for non-metaphorical pairs. For the Unsupervised Learning, they developed a method to train the neural network in a zero-shot setting without any Labeled metaphor examples. This unsupervised approach involved generating artificial metaphor examples from existing datasets and use them to train the model. In summary, the two annotated datasets, TSV and POEMS, and they achieved average precision values of 0.90 and 0.82, respectively. The zero-shot approach on the GerDraCor dataset achieved an average precision values of 0.70 for TSV and 0.74 for poems indicates the effectiveness of the model in detecting metaphors without the need for any Labeled metaphor examples during training.

[25] automatically detect metaphors in poems written in the Misurata Arabic sub-dialect. The research utilizes LSTM model to detect metaphors based on word co-occurrence patterns in the poetic texts. The dataset utilized comprises verses from poems in the Misurata Arabic sub-dialect, written by six native poets and the model outperforms human annotators in detecting metaphors, achieving a higher accuracy score of 79% and also demonstrates an 80.7% accuracy in predicting metaphors from unseen blind data which showcased its robustness in handling

new instances. The limitation of the study involves limited dataset and generalizability to other Arabic ethnicity.

[26] were aimed to present a baseline approach to assess novelty in metaphors based on their annotations and to develop a methodology to distinguish between conventionalized and novel metaphors in an existing corpus, and to investigate the correlation between novelty and features used in metaphor detection. BiLSTM model with word embeddings and additional features such as the relative frequency of a token and Potential for Metaphoricity (POM) of verbs measure were used to predict the novelty score of a metaphor. The research used the VU Amsterdam Metaphor Corpus to create a new annotation layer to distinguish between conventionalized and novel metaphors. The Mean Absolute Error (MAE) for the configuration with embeddings was 0.166, and it improved to 0.163 when frequency and POM features were added. The study's limitation includes limited dataset and the annotation process may be subjected to individual annotator biases, and corpus may not be representative of all types of text.

In summary, these studies offer valuable insights into the methodologies, challenges, and implications of metaphor detection across linguistic and cultural contexts. By drawing upon their findings and addressing their limitations, we can refine our approach and contribute to the advancement of metaphor detection research within Nigerian poetry.

## 3. Methodology

The research design used for this study combines elements of data curation, model development, and evaluation to achieve its aim while leveraging on both labeled and unlabelled textual data to train and validate the computational model for metaphor detection in Nigerian poetry. Figure 1 depicts the architectural design of the main steps involved in the methodology

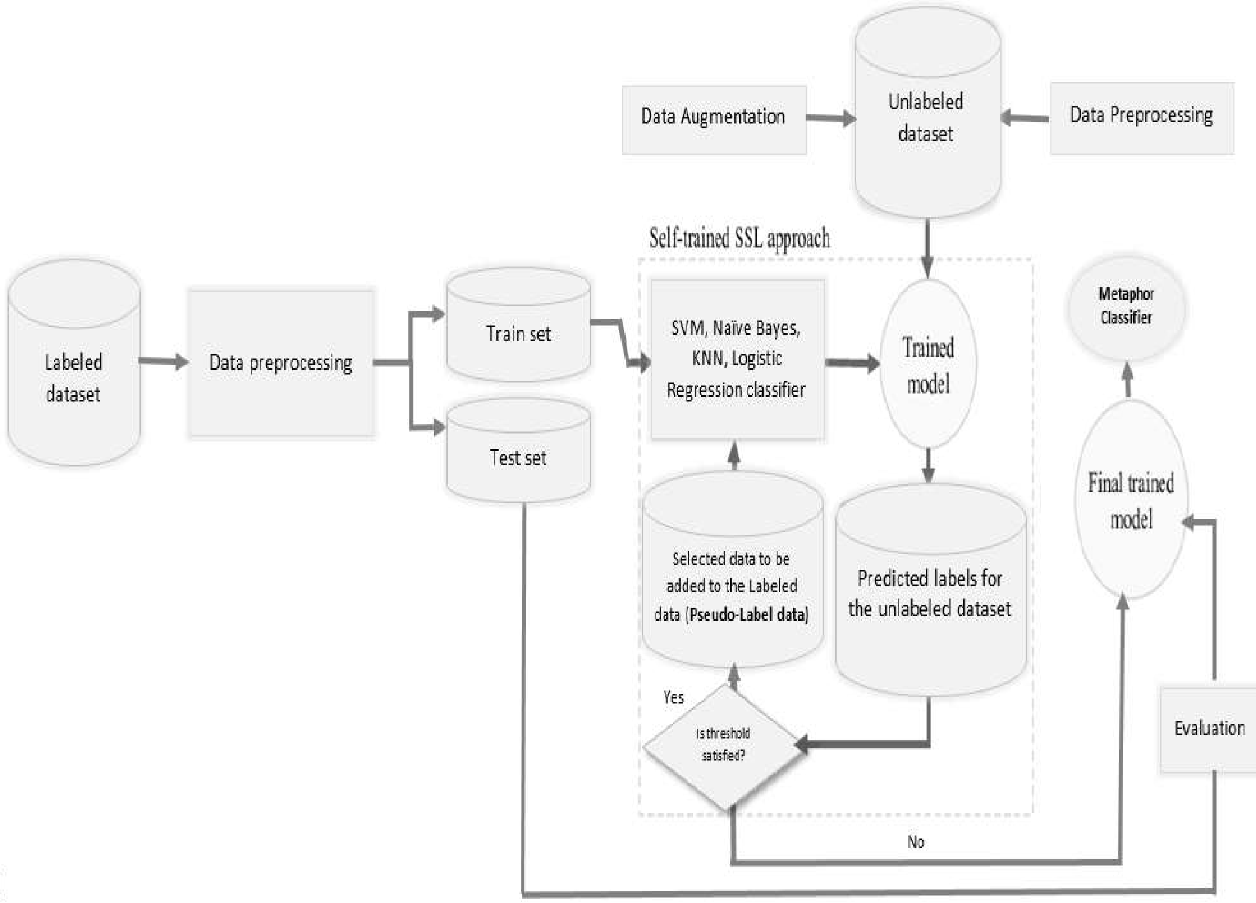



**Figure 1:** *The Architectural Design of the System*

### 3.1 Data Curation

Both labeled and unlabelled datasets were utilized in this study. The labeled dataset is sourced from a variety of repositories to capture the diversity of Nigerian poetry. These include published literature texts, online repositories of Nigerian poetry, contemporary literary journals and synthetic data generated to augment the dataset.

The literature texts include published works of Nigerian poetry books from renowned Nigerian poets, such as Wole Soyinka, Chinua Achebe, Niyi Osundare, Charles Akinsete and many others are curated to capture a wide range of poetic styles and themes prevalent in Nigerian literature [27][28][28][30]. The online sources, include digital libraries, poetry websites, literary forums and journals scoured from the web with recent works included to ensure the reflection of modern linguistic trends and cultural influences.[31][32][33][34].

In cases of data scarcity, the research employed synthetic data generation technique by utilizing ChatGPT AI to augment the dataset. Leveraging the structured nature of Artificial Intelligence (AI), a corpus of real poems was provided as input, prompting the system to produce more instances. This approach generated metaphorical expressions by extrapolating linguistic patterns and cultural motifs derived from existing poem samples.

Each data entry in the labeled dataset is meticulously annotated by domain experts to ensure accuracy and consistency. The annotation involves identifying and labeling metaphorical expressions within the poetry excerpts, assigning appropriate High Priority Semantic Markers (HPSM), Low Priority Semantic Markers (LPSM), and specifying the concepts being compared in each metaphorical instance.

To maintain data quality, a rigorous quality control process is implemented. This involves cross-validation by multiple annotators to verify the accuracy and consistency of annotations. Discrepancies are resolved through consensus discussions, ensuring the reliability of the labeled dataset.The unlabeled dataset consists of raw textual data extracted from various sources of Nigerian poetry, including literary collections and online archives. The data collection process involves the systematic compilation of poetry excerpts from diverse cultural and thematic contexts to capture the richness and complexity of Nigerian metaphors.Prior to model training, the raw textual data undergoes pre-processing steps to enhance readability and facilitate subsequent analysis. This includes tokenization, punctuation removal and lemmatization to standardize the text format that improves computational efficiency.

### 3.2 Size of Dataset

The dataset for metaphor detection in Nigerian poetry comprises both labeled and unlabelled data. The labeled dataset consists of approximately 1000 data entries, each annotated with relevant features such as excerpt, High Priority Semantic Markers (HPSM), Low Priority Semantic Markers (LPSM), Concepts Being Compared, and metaphorical Label. On the other hand, the unlabelled dataset comprises around 520 poems extracted from various sources of Nigerian poetry repositories as discussed in section 3.2, which was further pre-processed to about 14,000 excerpt entries.

### 3.3 Pre-processing

Poetry datasets, akin to many datasets encountered in real-world scenarios, frequently exhibit diverse forms of noise, discrepancies, and instances of missing data. These irregularities may arise from factors such as human errors during data collection processes or malfunctions in sensors. If left unattended, these challenges have the potential to markedly influence the efficacy of machine learning algorithms. Therefore, it becomes imperative to pre-process the dataset to mitigate these issues effectively. The ensuing sections delineate the various pre-processing steps employed to address these challenges. Several pre-processing steps were utilized and they are as follows:

#### 3.3.1 Text Cleaning

Regular expression was utilized to remove special characters, punctuation, and irrelevant symbols from the text data to ensure consistency and standardization.

#### 3.3.2 Tokenization

The study uses the Spacy's pre-trained model, 'en_core_web_lg' to split the text into individual tokens (words or phrases) to facilitate further analysis and feature extraction.

#### 3.3.3 Normalization

The study uses the Spacy's pre-trained model, 'en_core_web_lg' to convert all text to lowercase to reduce noise and improve model performance.

### 3.3.4 Vectorization

The study utilized an advanced vectorization technique provided by the Spacy pre-trained model, specifically 'en_core_web_lg'. This pre-trained model employs its own word embeddings, which are trained on extensive corpora and optimized for various NLP tasks. Through these embeddings, textual data is converted into dense numerical representations, enabling the capture of intricate semantic relationships between words within a continuous vector space. This approach proves advantageous for effectively capturing the nuanced context and meaning embedded within the excerpted poems.

### 3.4 Feature Engineering Techniques

Feature engineering plays a crucial role in enhancing the performance of machine learning models for metaphor detection. Techniques such as tok2vec, POS (part-of-speech) tagging, syntactic parser that are present in the Pre-trained spacy pipeline are utilized to capture linguistic patterns relevant to metaphor identification in Nigerian poetry.

### 3.5 Model Development

The model development employed a self-learning semi-supervised approach, a variant of semi-supervised learning. With a small Labeled dataset and a substantial amount of unlabelled data, this method iteratively trained on both sets, gradually enhancing performance through self-generated pseudo-labels. The process involved are as follows:

1. Training a supervised model on the initial labeled dataset (seed set).
2. Making predictions on unlabelled data using the trained model.
3. Adding predictions meeting probability threshold or k-best criteria to the pseudo-labelled set.
4. Combining labeled and pseudo-labelled data for training the next model iteration.
5. Repeating steps 2-4 until all the data are labelled or a stopping criterion is met.

Pseudo-labelled examples were merged with the original labeled dataset, forming an augmented training set for subsequent training iterations. The model iteratively refined its decision boundaries until convergence or reaching a predefined stopping criterion.

### 3.5.1 Machine Learning Classifiers

Multiple machine learning classifiers, including Decision Trees, Support Vector Machines (SVM), K-Neighbours and Logistic Regression classifiers were explored for their suitability in metaphor detection tasks. Each classifier is evaluated based on its performance metrics and ability to handle the complexities of Nigerian metaphors.

### 3.5.2 Model Evaluation and Validation

In this study, the performance of the classification models was assessed using a confusion matrix, Accuracy, Precision, Recall, and F1-score.

***a. Confusion Matrix***

A confusion matrix is a visual representation presenting the effectiveness of a classification model through a comparison between its predicted and true class values [35].

***b. Accuracy***

Accuracy serves as a widely utilized assessment measure for evaluating the effectiveness of a classification model. It is determined by dividing the number of correctly predicted instances by the total number of instances present in the dataset [35]. Mathematically, accuracy is represented as:

$$\text{Accuracy} = \frac{TruePositives + TrueNegatives}{TruePositives + TrueNegatives + FalsePositives + FalseNegatives}$$

#### c. Precision

Precision, a key metric in classification tasks, gauges the precision of positive predictions made by a model [35]. It's computed by dividing the count of true positive predictions by the total number of positive predictions (the sum of true positives and false positives). Put simply, precision signifies the ratio of accurately predicted positive instances out of all instances predicted as positive by the model. Mathematically, precision is represented as:

$$\text{Precision} = \frac{TruePositives}{TruePositives + FalsePositives}$$

#### d. Recall

Recall, alternatively termed Sensitivity, is a measurement employed in classification tasks to evaluate a model's capability in correctly identifying all pertinent instances [35]. It denotes the ratio of true positive instances correctly predicted by the model out of all actual positive instances. Mathematically, recall is represented as:

$$\text{Recall} = \frac{TruePositives}{TrusPositives + FalseNegatives}$$

#### e. F1-score

F1-score, is a metric that combines precision and recall into a single value to provide a balanced evaluation of a classification model's performance. It is particularly useful when there is an uneven class distribution in the dataset. The F-score is calculated as the harmonic mean of precision and recall, giving equal weight to both metrics [35]. The formula for calculating the F-score is:

$$\text{F1-Score} = \frac{2\ X Precision X Recall}{Precision + Recall}$$

These metrics were utilized to evaluate the performance of Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbours (KNN), and Logistic Regression classifiers. In evaluating the classifiers with the two classes (metaphorical '1' and non-metaphorical '0') individually, the confusion matrices tracks:

i. True Positives (TP): Correctly classified metaphorical instances ('1').
ii. False Positives (FP): Incorrectly classified as metaphorical ('1') when they're not.
iii True Negatives (TN): Correctly classified non-metaphorical instances ('0').
iv False Negatives (FN): Incorrectly classified as non-metaphorical ('0') when they're metaphorical.

Hence the confusion matrix facilitated the determination of the accuracy of each classification model.

### 3.4.1 Feature Extraction & Augmentation

In the labeled dataset, categorical features such as Excerpt, HPSM, LPSM, and Concepts Being Compared underwent manual annotation by an African linguistic expert. However, due to the absence of the HPSM, LPSM, and Concepts columns in the unlabelled dataset, and to avoid relying solely on the Excerpt column for the self-training model, the study employed a combination of rule-based methods and simple imputations to augment these columns.

For the Concepts Compared column, the study initially utilized SpaCy's large pre-trained model to extract the part-of-speech (POS) and vectors of words from the poem excerpts. This facilitated the development of a rule-based imputation approach based on the content of the Excerpt column to fill in missing values. Pairs of tokens were identified based on their POS types (adjective to noun pairs, verb to noun pairs, and noun to noun pairs), ensuring semantic dissimilarity through cosine similarity comparisons of their word vectors. A threshold of 0.5 controlled the level of allowed similarity.

Entries resembling metaphors but deemed non-metaphoric due to the presence of keywords 'Like' and 'As' in the HPSM and LPSM columns were augmented using a rule-based function. The respective columns were marked with 'the LIKE word present' and 'the AS word present' for HPSM, and 'direct comparison' for LPSM. Additionally, a simple imputation strategy was adopted to fill missing values with the placeholder "Unknown" for columns that did not meet the aforementioned conditions. Figure 2 depicts the Feature Extraction and Augmentation for Unlabelled dataset in the study

.

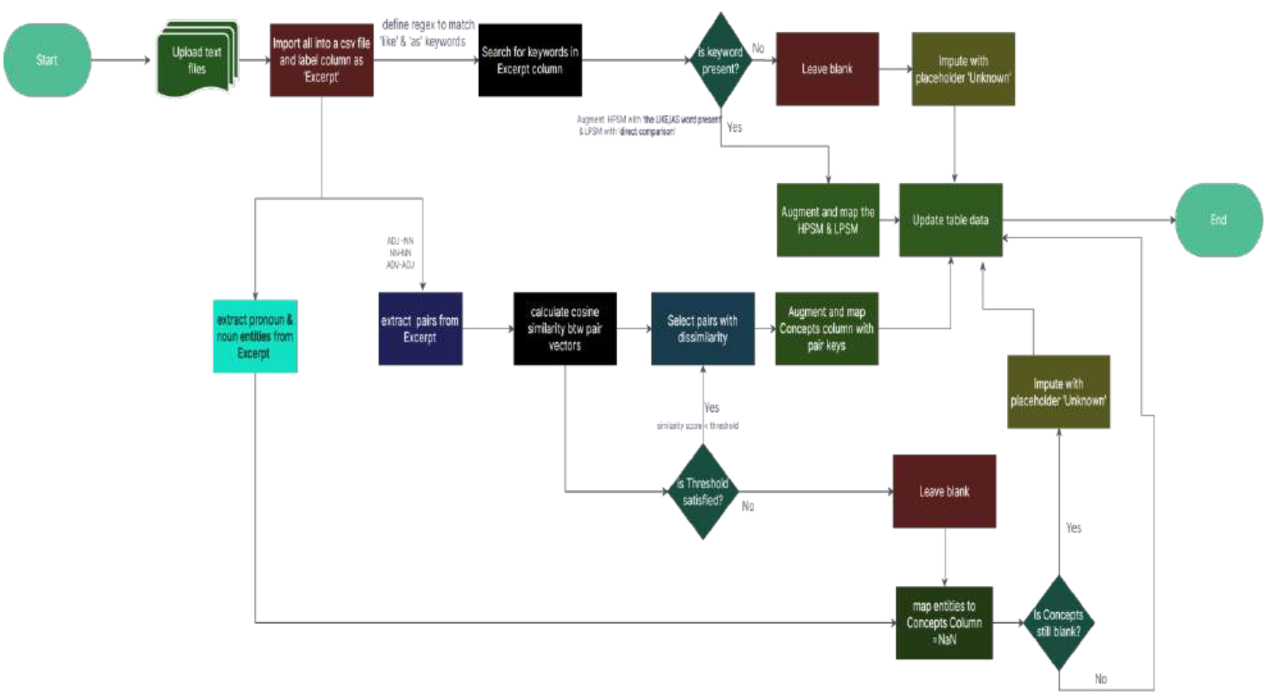



**Figure 2:***Feature Extraction and Augmentation for Unlabelled Dataset*

## 4. Implementation and Discussion of Results

This study utilized two datasets: a labeled dataset and an unlabelled dataset. The labeled dataset, sized at 57.1 KB, comprises of 510 rows and 5 columns. On the other hand, the unlabelled dataset, initially in .txt format consisting of 521 text files and totalling 556 KB, underwent transformation via Microsoft Excel Query Editor to become a CSV file.

Table 1 shows the labeled dataset and it contains all the relevant features for determining the present of metaphor. Figures 3 and 4 depicts the unlabelled datasets before and after transformation. Both labeled and unlabelled dataset were further pre-processed as discussed in section three. The labelled dataset was further split into train and test, the train set of the labeled observations was used to train the first supervised model.

**Table 1:***Label Dataset Samples*

| Excerpt | HPSM | LPSM | Concepts | Label |
|---|---|---|---|---|
| We shall not be visited by the vulture. | death | corpse, carcass | visit, vulture | 1 |
| Turn, turn, turn; the world knows little, of t... | passing time | moment | sun, huge mask | 1 |
| You are standing strong like the lobelia tree. | slim, strong, obsolete | but direct comparison | strong, lobelia tree | 0 |
| hush says the panting gun | threat | readiness | pant, gun | 1 |
| pour the white chalk on the floor | consult the oracle | voodoo | chalk, floor | 1 |

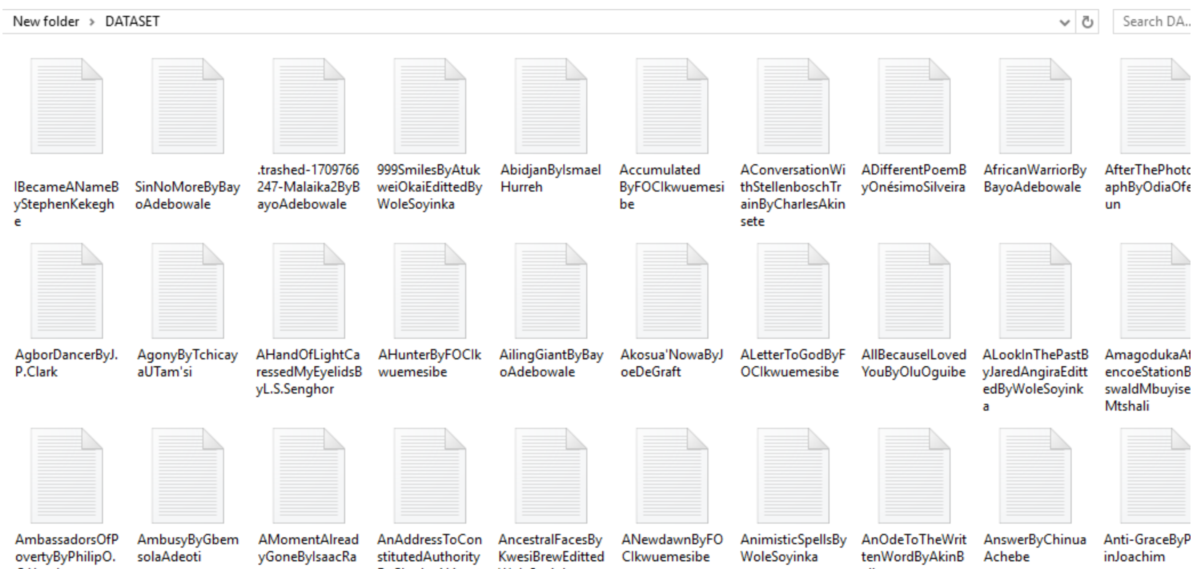



**Figure 3:** *Unlabelled Dataset in its raw form (in .txt files)*

| Excerpt |
|---|
| Master, if yesterday was dead, |
| I could still remember how I folded my knees |
| On the mocking face of the sand |
| To wash your filthy feet |
| My devotion dwarfed me, |

**Figure 4:***Unlabelled Dataset after Transformation*

Figure 5 depicts the head of first five (5) rows of the Unlabelled Dataset after Augmentation. The augmentation process involved integrating the HPSM, LPSM, and Concepts columns into the dataset using rule-based methods and simple imputations, as elaborated in Chapter Three. A threshold of 0.7 was utilized to regulate the semantic similarity and dissimilarity levels for the rule-based imputation. This was achieved by conducting cosine similarity comparisons of selected Part of Speech (POS) token pairs, including adjective-noun, verb-noun, and noun-noun pairs.

Additionally, a rule-based imputation was applied to identify entries resembling metaphors but determined as non-metaphoric due to the presence of keywords 'Like' and 'As' in the HPSM and LPSM columns. These entries were labelled accordingly: 'the LIKE word present' and 'the AS word present' for HPSM, and 'direct comparison' for LPSM.
For columns that did not meet these conditions, a simple imputation strategy was implemented, filling missing values with the placeholder "Unknown." After these augmentations the dataset expanded to 7.8 MB, featuring 14,994 rows and 4 columns.

| | Excerpt | HPSM | LPSM | Concepts |
|---|---|---|---|---|
| 0 | Master, if yesterday was dead, | Unknown | Unknown | Master, yesterday, |
| 1 | I could still remember how I folded my knees | Unknown | Unknown | I, I, my knees, |
| 2 | On the mocking face of the sand | Unknown | Unknown | mocking, face |
| 3 | To wash your filthy feet | Unknown | Unknown | filthy, foot |
| 4 | My devotion dwarfed me, | Unknown | Unknown | My devotion, me, |

**Figure 5:***The Unlabelled Dataset after Augmentation*

The study delved into the classification of metaphor detection in Nigerian poetry utilizing various machine learning models. Four models were employed: Support Vector Machine (SVM), Decision Tree, Logistic Regression, and K Nearest Neighbour (KNN) Classifier. Each model was assessed using the Semi-Supervised Learning (SSL) approach.

The initial Supervised Learning achieved an accuracy of 96% on the test set for SVM Classifier, it showed an excellent precision of 0.94, recall of 0.98, and F1-scores of 0.96 for metaphoric class 1, reflecting strong performance in metaphor detection. Logistic Regression Model achieved an accuracy of 95%, demonstrated a high precision of 0.94, recall of 0.96, and F1-scores of 0.95, indicating a robust performance in metaphor detection. For Decision Tree the model obtained an accuracy of 85%, precision of 0.81 recall of 0.90, and F1-scores of 0.85; while the accuracy is slightly lower compared to other models, it still demonstrated reasonable performance in metaphor detection. And for K Nearest Neighbor the model achieved an accuracy of 89%, displayed a solid precision of 0.91, recall of 0.86, and F1-scores of 0.88. Figure 6 depicts the classification reports for all four models used in the supervised learning approach.

---------- Decision Tree Classifier Model - Evaluation on Test Data ----------

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.90 | 0.81 | 0.85 | 53 |
| 1 | 0.81 | 0.90 | 0.85 | 49 |
| accuracy | | | 0.85 | 102 |
| macro avg | 0.86 | 0.85 | 0.85 | 102 |
| weighted avg | 0.86 | 0.85 | 0.85 | 102 |

---------- K Nearest Neighbors Classifier Model - Evaluation on Test Data ----------

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.88 | 0.92 | 0.90 | 53 |
| 1 | 0.91 | 0.86 | 0.88 | 49 |
| accuracy | | | 0.89 | 102 |
| macro avg | 0.89 | 0.89 | 0.89 | 102 |
| weighted avg | 0.89 | 0.89 | 0.89 | 102 |

---------- Logistic Regression Classifier Model - Evaluation on Test Data ----------

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.96 | 0.94 | 0.95 | 53 |
| 1 | 0.94 | 0.96 | 0.95 | 49 |
| accuracy | | | 0.95 | 102 |
| macro avg | 0.95 | 0.95 | 0.95 | 102 |
| weighted avg | 0.95 | 0.95 | 0.95 | 102 |

---------- Support Vector Machine Classifier Model - Evaluation on Test Data ----------

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.98 | 0.94 | 0.96 | 53 |
| 1 | 0.94 | 0.98 | 0.96 | 49 |
| accuracy | | | 0.96 | 102 |
| macro avg | 0.96 | 0.96 | 0.96 | 102 |
| weighted avg | 0.96 | 0.96 | 0.96 | 102 |

**Figure 6:** *Classification Reports for All Four Models in Supervised Learning Approach*

For Employing the Self-Training Semi-Supervised Learning approach, SVM achieved an accuracy of 80.39% on the test set. While the accuracy is slightly lower compared to the initial supervised SVM model, it still demonstrated reasonable performance, considering the utilization of both labeled and unlabelled data. The model demonstrated a slightly fair precision, recall and F1 scores for the metaphoric class 1, with precision of 0.78, a recall of 0.82 and F1-score of 0.80. The Decision Tree model achieved a final accuracy of 79.41%, slightly lower than the SVM model. It displayed comparable precision, F1-score and recall rates for both approaches, with precision of 0.80, a recall of 0.76 and a F1-score of 0.78, indicating its competency in metaphor detection but with a marginally lower accuracy compared to SVM in the SSL approach. The Logistic Regression model outperformed the other models with an impressive accuracy of 93.14% on the test set. It exhibited high precision, F1-score and recall rates for both classes, with precision of 0.92, a recall of 0.94 and F1-score of 0.93, indicating its robustness and efficiency in metaphor detection.

Lastly, the KNN Classifier achieved a final accuracy of 86.27%, displaying strong precision and recall rates for both classes. Its precision of 0.82, a recall of 0.92 and F1-score of 0.87 indicates its effectiveness in metaphor detection, although slightly lower than the Logistic Regression model.

Overall, the Logistic Regression model demonstrated the highest accuracy and robustness in metaphor detection among the models tested. However, all models showed promising results, indicating the feasibility of employing machine learning techniques for metaphor detection in Nigerian poetry. Figure 7 depicts the classification reports for all the four models utilizing self-trained semi-supervised learning approach.

```
---------- DecisionTreeClassifier - SSL Approach ----------
Final model accuracy on the test set: 0.7941176470588235
Classification report on the test set:
              precision    recall  f1-score   support

           0       0.79      0.83      0.81        53
           1       0.80      0.76      0.78        49

    accuracy                           0.79       102
   macro avg       0.80      0.79      0.79       102
weighted avg       0.79      0.79      0.79       102

Classes: [0 1]
Number of iterations: 101
```

```
---------- K-Neighbors Classifier - SSL Approach ----------
Final model accuracy on the test set: 0.8627450980392157
Classification report on the test set:
              precision    recall  f1-score   support

           0       0.91      0.81      0.86        53
           1       0.82      0.92      0.87        49

    accuracy                           0.86       102
   macro avg       0.87      0.86      0.86       102
weighted avg       0.87      0.86      0.86       102

Classes: [0 1]
```

```
---------- Logistic Regression Classifier - SSL Approach ----------
Final model accuracy on the test set: 0.9313725490196079
Classification report on the test set:
              precision    recall  f1-score   support

           0       0.94      0.92      0.93        53
           1       0.92      0.94      0.93        49

    accuracy                           0.93       102
   macro avg       0.93      0.93      0.93       102
weighted avg       0.93      0.93      0.93       102

Classes: [0 1]
Number of Iterations: [100]
```

```
---------- Support Vector Machine - SSL Approach ----------
Final model accuracy on the test set: 0.803921568627451
Classification report on the test set:
              precision    recall  f1-score   support

           0       0.82      0.79      0.81        53
           1       0.78      0.82      0.80        49

    accuracy                           0.80       102
   macro avg       0.80      0.80      0.80       102
weighted avg       0.80      0.80      0.80       102

Classes: [0 1]
Number of Iterations: [2964]
```

**Figure 7:***Classification Reports for All Models in the Self-Trained SSL Approach*

Using confusion matrix for visualization, we could see that in the supervised learning approach, the SVM model correctly classified 50 instances as positive (metaphors) and 48 instances as negative (non-metaphors). It made 3 false positive errors, where non-metaphors were incorrectly classified as metaphors, and 1 false negative error, where a metaphor was incorrectly classified as a non-metaphor. Overall, the SVM model achieved high accuracy and precision with minimal misclassifications. The Decision Tree model correctly classified 43 instances as positive and 44 instances as negative. It made 10 false positive errors and 5 false negative errors. While the model demonstrated reasonable performance, it had a higher rate of false positives compared to SVM. The KNN model correctly classified 49 instances as positive and 42 instances as negative. It made 4 false positive errors and 7 false negative errors. The model achieved high precision but had a slightly higher rate of false negatives compared to SVM. The Logistic Regression model correctly classified 50 instances as positive and 47 instances as negative. It made 3 false positive errors and 2 false negative errors. Similar to SVM, the Logistic Regression model demonstrated high accuracy and precision with minimal misclassifications. Figures 8 to 11 show the confusion matrix for all the four models in supervised approach.

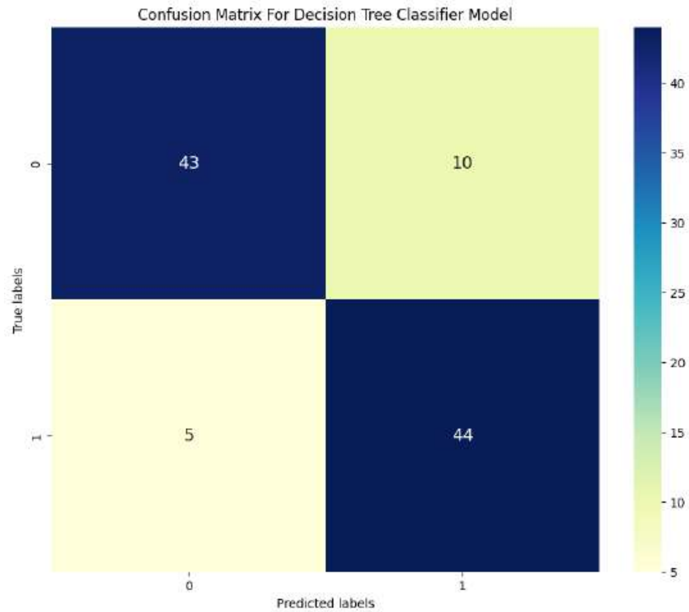



**Figure 8:***Confusion Matrix for Decision Tree in Supervised Approach*

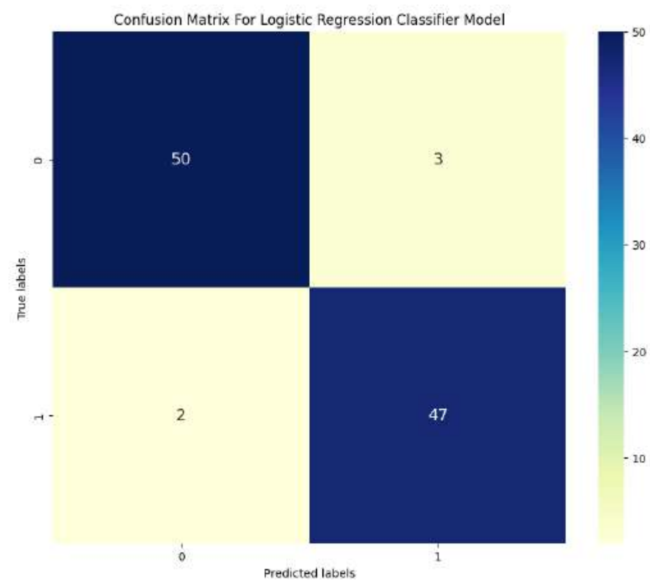



**Figure 9:***Confusion Matrix for Logistic Regression Classifier in Supervised Approach*

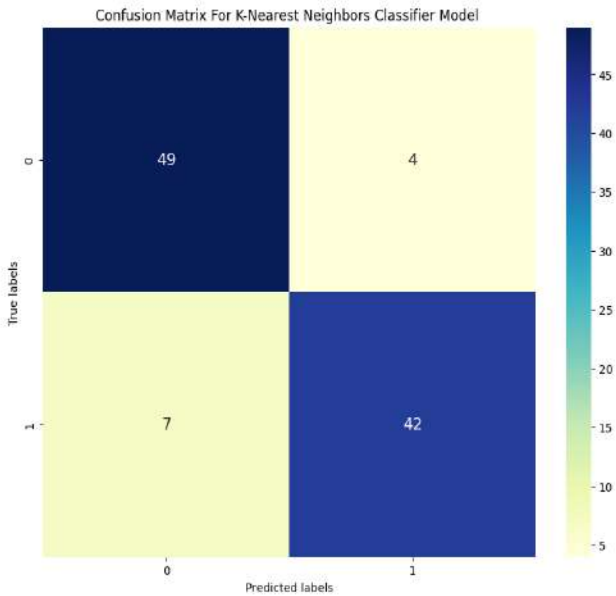
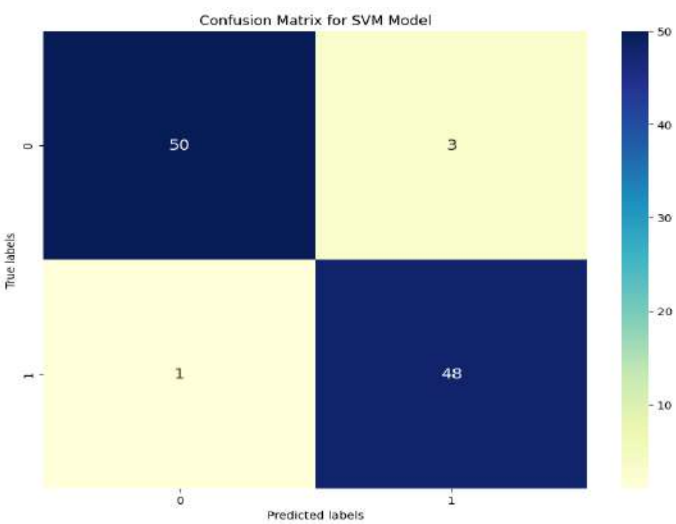



**Figure 10:***Confusion Matrix for KNN Classifier in Supervised Approach*




In the self-trained semi-supervised learning, the Decision Tree model trained correctly classified 44 instances as positive and 37 instances as negative. It made 9 false positive errors and 12 false negative errors. The model showed a higher rate of false negatives compared to the supervised learning approach. The KNN model demonstrated similar performance to the supervised learning approach, with 43 true positives, 45 true negatives, 10 false positives, and 4 false negatives. The Logistic Regression model, trained using self-training semi-supervised learning, achieved 49 true positives, 46 true negatives, 4 false positives, and 3 false negatives. The model showed slightly better performance compared to the supervised learning approach and the SVM model trained using self-training semi-supervised learning demonstrated comparable performance to the supervised learning approach, with 42 true positives, 40 true negatives, 11 false positives, and 9 false negatives. Figures 12 to 15 show the confusion matrix for all the four models in semi-supervised approach.

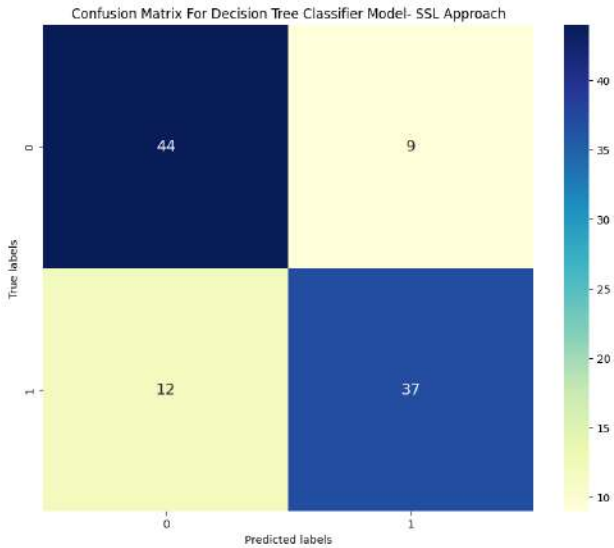



**Figure 12:** *Confusion Matrix for Decision Tree in Self-Trained Semi-Supervised Approach*

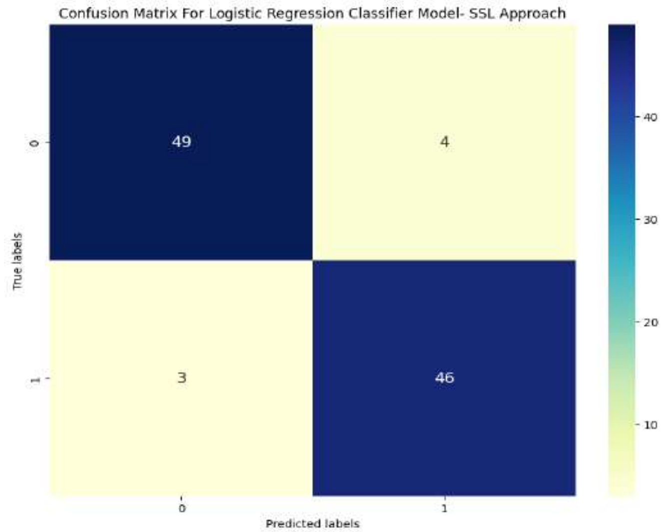



**Figure 14:** *Confusion Matrix for Logistic Regression Classifier in Self-Trained Semi-Supervised Approach*

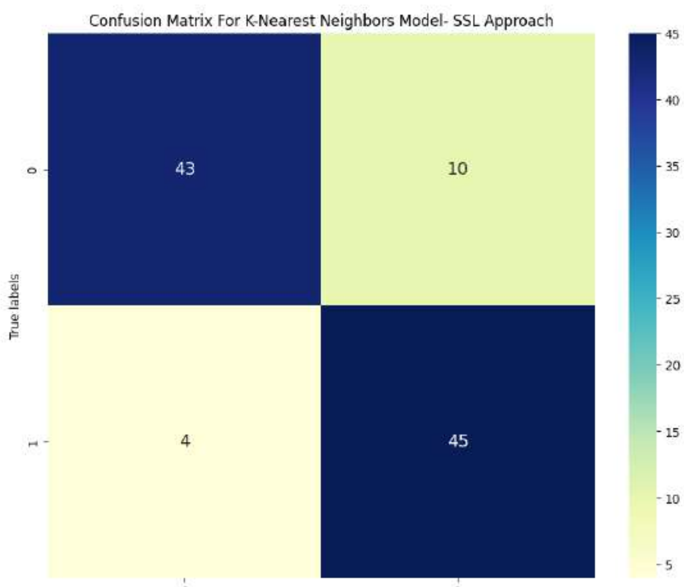



**Figure 13:** *Confusion Matrix for KNN Self-Trained Semi-Supervised Approach*

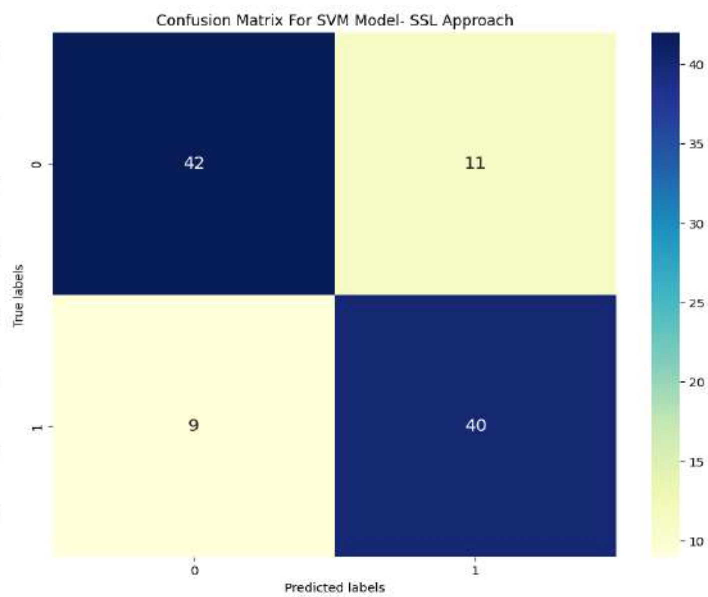



**Figure 15:** *Confusion Matrix for SVM in Self-trained Semi-Supervised Approach*

Table 2 gives the overall summary of both supervised and self-trained semi-supervised evaluation results for the Metaphoric Class using the four classifiers.

**Table 2:** *Supervised and Semi-Supervised Evaluation Result for Metaphoric Class Using Four Different Classifiers*

| Supervised Approach | | | | |
|---|---|---|---|---|
| **Classifier** | **Accuracy** | **Precision** | **Recall** | **F1-score** |
| **SVM** | 96% | 94% | 98% | 96% |
| **Decision Tree** | 85% | 81% | 90% | 85% |
| **KNN** | 89% | 91% | 86% | 88% |
| **Logistic Regression** | 95% | 94% | 96% | 95% |
| **Self-Train SSL Approach** | | | | |
| **SVM** | 80% | 78% | 82 % | 80% |
| **Decision Tree** | 79% | 80% | 76% | 78% |
| **KNN** | 86% | 82% | 92 % | 87% |
| **Logistic Regression** | 93% | 92% | 94% | 93% |

## 5. Limitation of the Study

This study encountered several limitations that may impact the robustness and generalizability of its findings. Firstly, the utilization of a simple imputation method, whereby missing values were replaced with "unknown" placeholders for High Priority Semantic Markers (HPSM) and Low Priority Semantic Markers (LPSM) in the unlabelled dataset, may not adequately capture the nuanced conceptual meanings of these markers; this could potentially lead to inaccuracies in metaphor detection.

Additionally, the study's reliance on a relatively small labeled dataset, comprising only 509 data entries, poses challenges in training and generalization. A larger dataset would provide a more diverse range of metaphorical expressions, thereby enhancing the model's accuracy and reliability. Lastly, Nigeria's rich ethnic diversity presents a significant challenge in capturing the full spectrum of metaphorical expressions. While the study aimed at exploring metaphors in Nigerian poetry, it may not fully encompass the richness and diversity of metaphors across various Nigerian's ethnic groups and languages, potentially limiting the model's applicability and its utility in broader Natural Language Processing (NLP) applications.

## 6. Conclusion

In summary the study addressed the challenge of metaphor detection in Nigerian poetry, a task complicated by the rich cultural diversity, data scarcity and the nuanced linguistic expressions inherent in Nigerian literature. By leveraging both labeled and unlabeled datasets, and employing a combination of supervised and self-training semi-supervised learning approaches, the research demonstrated the effectiveness of various machine learning models in detecting metaphors within Nigerian poetry.

The supervised learning approach, particularly the Support Vector Machine (SVM) classifier, achieved the highest accuracy of 96% on the test set, showcasing its robust capability in metaphor detection. Conversely, the self-training semi-supervised learning approach highlighted the potential of logistic regression, achieving an impressive accuracy of 93%. These results underscore the importance of employing sophisticated computational methods to capture the intricate and culturally specific metaphorical language used in Nigerian poetry.

One of the key findings of this research is the effectiveness of both SVM and logistic regression models, with SVM leading in supervised learning and logistic regression proving highly effective in semi-supervised learning. The careful pre-processing and augmentation techniques, including the use of synthetic data generation and feature engineering, significantly enhanced model performance. Moreover, the study emphasized the necessity of incorporating cultural and linguistic nuances into NLP models to improve their accuracy and relevance in specific literary contexts.

Despite these promising results, the study also identified areas for future improvement. Collecting a larger and more diverse dataset, employing more sophisticated imputation methods, and collaborating with linguists and cultural experts will further enhance the accuracy, reliability, and applicability of metaphor detection models. Additionally, validating the model across different cultural contexts and exploring alternative machine learning algorithms can provide deeper insights and broader applicability in the field of NLP.

In conclusion, this study underscores computational methods' potential in unravelling metaphorical expressions within Nigerian poetry. It contributes to knowledge by creating a model that understands the metaphoricity of the nuanced Nigerian expression.

**References**

[1] J. Li and H. Pang, "The Performance of Conceptual Metaphors in Different Language Systems," *Communications in Humanities Research*, vol. 3, no. 1, pp. 899–904, 2023. [Online]. Available: https://doi.org/10.54254/2753-7064/3/2022684

[2] T. Editors of Encyclopaedia Britannica, "Metaphor," *Encyclopedia Britannica*, 2024. [Online]. Available: https://www.britannica.com/art/metaphor

[3] P. Yunfei, "Conceptual Metaphor Analysis of the Great Gatsby," *Journal of Education and Educational Research*, pp. 46-49, 2023.

[4] X. Rong, E. Chersoni, Q. Lu, C.-R. Huang, W. Li, and Y. Long, "Lexical data augmentation for sentiment analysis," *Journal of the Association for Information Science and Technology*, vol. 72, no. 11, pp. 1467–1479, 2021. [Online]. Available: https://doi.org/10.1002/ASI.24493

[5] C. G. Passakos and B. de Raad, "Ancient personality: Trait attributions to characters in Homer's Iliad," *Ancient Narrative*, vol. 7, 2009.

[6] S. Gupta, "Before The Cask of Wine by Li Bai," 2020. [Online]. Available: https://poemanalysis.com/li-bai/before-the-cask-of-wine/; last retrieved 27th June, 2024.

[7] K. Doss, "Figurative Language in Sonnet 18 | Metaphor, Imagery & Others," Study.com, Feb. 21, 2023. [Online]. Available: https://study.com/academy/lesson/figurative-language-in-sonnet-18.html; last retrieved 27th June, 2024.

[8] O. Asad, "Hope is the Thing with Feathers by Emily Dickinson," 2016. [Online]. Available: https://poemanalysis.com/emily-dickinson/hope-is-the-thing-with-feathers/; last retrieved 27th June, 2024.

[9] A. Corfman, "Caged Bird by Maya Angelou," 2016. [Online]. Available: https://poemanalysis.com/maya-angelou/caged-bird/; last retrieved 27th June, 2024.

[10] O. D. Ogungbemi, "Metaphors as Discourse Strategies in Osundare's Poetry," *International Journal of Humanities and Social Sciences*, 2018. [Online]. Available: http://www.ijhcs.com/index.php/ijhcs/index

[11] A. Ifinedo, "Answer by Chinua Achebe," 2021. [Online]. Available: https://poemanalysis.com/chinua-achebe/answer/; last retrieved 27th June, 2024.

[12] O. Idoko, O. Celestine, and L. Nkeiruka, "Metaphor as Conceptual Constructs of Corruption and Identity in selected Wole Soyinka's Poems," 2022. [Online]. Available: https://ssrn.com/abstract=4065904; last retrieved 27th June, 2024.

[13] F. C. Ononye and C. Innocent, "There's still something positive about the niger delta ecology: Metaphor and

ideology in the niger delta poetic discourse," *Language and Literature*, vol. 32, no. 3, pp. 275–296, 2023.

[14] R. O. Maledo and E. O. Emama, "Metaphorising the Nigerian Space: A Critical Stylistic Study of Stephen Kekeghe's Rumbling Sky," *3L: Language, Linguistics, Literature*, vol. 28, no. 4, pp. 169–183, 2022. [Online]. Available: https://doi.org/10.17576/3L-2022-2804-12

[15] U. Okunrinmeta and O. O. Alabi, "A Cultural-Conceptual Analysis of Some Metaphors of Corruption in Nigerian Literature," *Asian Journal of Social Sciences & Humanities*, vol. 3, no. 3, 2014. [Online]. Available: www.ajssh.leena-luna.co.jp

[16] K. Maithili, N. Raja, et al., "A Survey (NLP) Natural Language Processing and Transactions on (NNL) Neural Networks and learning Systems," *E3S Web of Conferences*, vol. 430, 2023. [Online]. Available: https://doi.org/10.1051/e3sconf/202343 001148

[17] I. H. Chen, Y. Long, Q. Lu, and C. R. Huang, "Metaphor detection: Leveraging culturally grounded eventive information," *IEEE Access*, vol. 7, pp. 10987–10998, 2019. [Online]. Available: https://doi.org/10.1109/ACCESS.2019.2 892042

[18] G. Pragglejaz, "MIP: A Method for Identifying Metaphorically Used Words in Discourse," vol. 22, no. 1, 2007.

[19] I. Reinig and I. Rehbein, "Metaphor detection for German Poetry," in *Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019)*, pp. 149–160. [Online]. Available: http://www.deutschestextarchiv.de/

[20] A. Mouraz, A. Vale, and R. Rodrigues, "The Use of Metaphors in the Processes of Teaching and Learning in Higher Education," *www.iojes.net*, 2013.

[21] S. Aggarwal and R. Singh, "Metaphor Detection using Deep Contextualized Word Embeddings," arXiv, 2020. [Online]. Available: http://arxiv.org/abs/2009.12565

[22] L. Bulat, S. Clark, and E. Shutova, "Modelling metaphor with attribute-based semantics," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistic*, 2017, vol. 2, pp. 523-528. [Online]. Available: https://www.sketchengine.co.uk/xdocu mentation/wiki/Cor-; last retrieved 27th June, 2024.

[23] O. Adésolá, "The Yoruba Animal Metaphors: Analysis and Interpretation," *Nordic Journal of African Studies*, vol. 14, no. 3, 2005.

[24] F. Schneider, S. Sickert, P. Brandes, S. Marshall, and J. Denzler, "Metaphor Detection for Low Resource Languages: From Zero-Shot to Few-Shot Learning in Middle High German," 2022. [Online]. Available: https://github.com/cvjena/metaphor-detector; last retrieved 27th June, 2024

[25] A. Abugharsa, "Metaphor Detection in Poems in Misurata Arabic Sub-Dialect: An LSTM Model," Montclair State University, 2022. [Online]. Available: https://digitalcommons.montclair.edu/et d

[26] E.-L. Do Dinh, H. Wieland, and I. Gurevych, "Weeding out Conventionalized Metaphors: A Corpus of Novel Metaphor Annotations," in *Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 1412-1424. [Online]. Available:

http://www.ukp.tu-darmstadt.de

[27] W. Soyinka, *Poems of Black Africa*. Heinemann African Writers Series, 1975.

[28] C. Akinsete, *Do not preach to me*. Ibadan: Greenminds Media & Publishers, 2017.

[29] C. Akinsete, *Dance of a savage kingdom*. Minna: Amab books & publishing, 2020.

[30] F. Ademola, *Reflections Nigerian Prose & Verse*. Lagos: Africa Universities Press, 1965.

[31] PoemAnalysis, 2024. [Online]. Available: https://poemanalysis.com; last retrieved 27th June, 2024.

[32] AfricaWriter, 2024. [Online]. Available: https://www.africawriter.com; last retrieved 27th June, 2024.

[33] PoemHunter, 2024. [Online]. Available: https://poemhunter.com; last retrieved 27th June, 2024.

[34] Oleyede, 2024. [Online]. Available: https://www.oleyede.com; last retrieved 27th June, 2024.

[35] Vujović, Ž. (2021). Classification Model Evaluation Metrics. International Journal of Advanced Computer Science and Applications, 12(6), 599–606. https://doi.org/10.14569/IJACSA.2021.0120670