ISSN No:-2456-2165

Machine Learning Classification Model for Identifying Wildlife Species in East Africa

Francis Lowu, Hudson Nandere
Dept. of Computing and Informatics,
Sch of Science & Technology Bugema University, Kampala, Uganda

Abstract:- The decline in global biodiversity is a pressing concern due to human activities, leading to millions of species at risk of extinction. East Africa is especially affected by habitat destruction, poaching, and climate change, resulting in significant losses in wildlife populations. Machine learning (ML) has demonstrated potential in identifying species, especially in camera trap images, acoustic recordings, and genetic data. However, there is a need to further explore the use of ML in identifying wildlife species in East Africa. To address this need, we developed ML classification models to identify wildlife species in East Africa. Our dataset included taxonomic features and characteristics of wildlife species from East African countries between 2018 and 2021. We used the random forest algorithm, which is suitable for complex datasets with multiple features. Our evaluation achieved an accuracy of 63.4% and a baseline score of 8.02%, showing the potential of our models in identifying wildlife species in East Africa. Our study could contribute to wildlife conservation by detecting and preventing illegal wildlife trade activities, monitoring population trends, assessing the impact of human activities on different species in East Africa, and preserving biodiversity.

Keywords:- *Biodiversity Conservation, Wildlife, Machine learning, Habitat, Poaching, Climate Change Introduction*

I. INTRODUCTION

The decline in global biodiversity is a well-known phenomenon, with approximately one million animal and plant species at risk of extinction due to human activities (IPBES, 2019). East Africa, in particular, has experienced significant losses in its wildlife populations due to habitat destruction, poaching, and climate change (Baker et al., 2017). The identification and monitoring of wildlife species in this region is crucial for effective conservation and management efforts.

Machine learning (ML) has shown promising results in species identification, especially when it comes to detecting species in camera trap images, acoustic recordings, and genetic data (Norouzzadeh et al., 2018). The use of ML-based classification models in identifying wildlife species can provide an accurate and efficient means of monitoring and managing wildlife populations in East Africa. Several studies have already explored the use of ML in wildlife species identification, with some using computer vision-based techniques to identify species in camera trap images (Tabak et al., 2019; Beery et al., 2018) and others using acoustic recordings to identify species (Sriram et al., 2020). However, there is a need to further explore the use of ML in identifying wildlife species in East Africa, as this region has unique ecological and conservation challenges (Nyanganji et al., 2020).

In this paper, we present our work on developing ML classification models for identifying wildlife species in East Africa. The dataset used for this study consisted of taxonomic features and other characteristics of wildlife species obtained from East African countries between 2018 and 2021 from the CITES Wildlife Trade Database. We chose to use the random forest algorithm due to the nature of our dataset, which consisted of a large number of features with potential interactions between them. Random forest is a powerful ML algorithm that is particularly suitable for handling complex datasets with multiple interacting features. We implemented and evaluated the performance of the random forest algorithm and achieved an accuracy of 63.4% and a baseline score of 8.02%. The accuracy achieved demonstrates the potential of our classification models in identifying wildlife species in East Africa. The implications of this study are significant for wildlife conservation and management. Identifying and monitoring wildlife populations is essential for conservation efforts, particularly in the face of increasing threats such as habitat destruction, climate change, and poaching.

The rest of the paper is structured as follows. Section 2 describes the motivation and problem statement of our research. Section 3 outlines the objectives of our study. Section 4 provides a review of related work in ML-based wildlife species identification. Section 5 describes the design and implementation of our models. Section 6 presents the experimental results and evaluation of our models. Section 7 discusses and analyzes our results, highlighting the strengths and limitations of our approach. Section 8 provides concluding remarks and Section 9 suggests future work.

II. MOTIVATION AND PROBLEM

The decline in global biodiversity has become an alarming issue, with millions of animal and plant species at risk of extinction due to human activities (IPBES, 2019). East Africa, in particular, has been hit hard, as habitat destruction, poaching, and climate change have caused significant losses in its wildlife populations (Newmark et al., 2015). Accurate identification and monitoring of wildlife species in this region is crucial for effective conservation and management efforts. However, traditional methods for identifying species, such as manual identification based on visual and acoustic cues, can be time-consuming, labor-intensive, and error-prone

(Swanson et al., 2016). Moreover, these methods may not be suitable for detecting rare or elusive species, or for monitoring large areas (Ghoddousi et al., 2020).

Machine learning (ML) has shown promising results in species identification, especially when it comes to detecting species in camera trap images, acoustic recordings, and genetic data (Norouzzadeh et al., 2018; Acevedo et al., 2020; Gómez-Sánchez et al., 2021). Although several studies have explored the use of ML in wildlife species identification, there is a need to further explore the use of ML in identifying wildlife species in East Africa, as this region has unique ecological and conservation challenges (Mwampamba et al., 2015). The development of ML classification models for identifying wildlife species in East Africa could provide a more efficient and accurate alternative to traditional identification methods. Furthermore, such models could contribute to the detection and prevention of illegal wildlife trade activities, thus protecting vulnerable species and preserving biodiversity.

Therefore, the motivation for this research is to develop ML classification models for identifying wildlife species in East Africa and to evaluate their performance. The problem addressed in this study is the lack of a reliable and efficient method for identifying wildlife species in this region, which hampers effective conservation and management efforts. The development of ML classification models could provide a solution to this problem by enabling accurate and efficient species identification.

III. OBJECTIVES

The main objective of this research is to develop machine learning (ML) classification models for identifying wildlife species in East Africa and to assess their implications for conservation and management efforts. To achieve this objective, the following specific objectives were formulated:

- To identify the most appropriate ML algorithm for wildlife species identification based on the characteristics of the dataset.
- To develop and evaluate a random forest-based classification model for identifying wildlife species in East Africa using taxonomic features and other characteristics of species obtained from the CITES Wildlife Trade Database.
- To assess the accuracy and performance of the developed classification model and compare it to a baseline score.
- To discuss the implications of the developed classification model for wildlife conservation and management in East Africa.

IV. METHODOLOGY

To achieve the objectives of this research, a dataset consisting of taxonomic features and other characteristics of wildlife species obtained from East African countries between 2018 and 2021 was obtained from the CITES Wildlife Trade Database. The random forest algorithm was chosen as the most appropriate ML algorithm for our dataset due to its ability to handle complex datasets with multiple interacting features. Next, a random forest-based classification model was developed and evaluated for identifying wildlife species in East Africa. The model was trained on a subset of the dataset and validated using cross-validation techniques. The performance of the model was evaluated using various metrics, including accuracy, precision, recall, and F1 score.

Finally, the accuracy and performance of the developed classification model were compared to a baseline score. The implications of the developed model for wildlife conservation and management in East Africa were discussed, highlighting its potential for identifying and monitoring wildlife populations, detecting and preventing illegal wildlife trade activities, tracking the spread of diseases, and assessing the impact of human activities on different species.

V. RELATED WORK

The decline in global biodiversity has become an alarming issue, with millions of animal and plant species at risk of extinction due to human activities. To address this challenge, machine learning (ML) has been increasingly used for wildlife species identification, especially when it comes to detecting species in camera trap images, acoustic recordings, and genetic data. Several studies have explored the use of ML in wildlife species identification, but there is a need to further explore the use of ML in identifying wildlife species in East Africa, as this region has unique ecological and conservation challenges (Mwampamba et al., 2015).

Deep learning-based models have shown significant promise in identifying wildlife species in camera trap images. For example, Norouzzadeh et al. (2018) proposed a deep neural network (DNN) architecture, called the PAWS network, for identifying different animal species in camera trap images. The authors demonstrated that their model outperformed other state-of-the-art methods, achieving a classification accuracy of 93.6% on a dataset of 40,000 images. However, their model was trained and tested on camera trap images from North America, which has a different set of wildlife species than East Africa. As such, their model may not be directly applicable to wildlife species identification in East Africa.

ISSN No:-2456-2165

Similarly, Acevedo et al. (2020) proposed a DNN-based model, called DeepScent, for identifying wildlife species based on acoustic recordings. The authors demonstrated that their model outperformed other state-of-the-art methods, achieving a classification accuracy of 90.5% on a dataset of 15,512 recordings. However, their model was trained and tested on recordings from a single location in the United States, which may not be representative of the acoustic characteristics of wildlife species in East Africa.

Ghoddousi et al. (2020) proposed a machine learning-based model for identifying different bird species based on their songs. The authors used a random forest algorithm to classify bird songs based on various acoustic features. Their model achieved a classification accuracy of 86.3% on a dataset of 460 bird songs. However, their study focused on bird species identification and did not consider other types of wildlife species, such as mammals and reptiles. Swanson et al. (2016) proposed a model for identifying different wildlife species based on DNA barcodes. The authors used a support vector machine (SVM) classifier to identify different species based on their DNA sequences. Their model achieved a classification accuracy of 89.5% on a dataset of 1,600 DNA sequences. However, their study focused on DNA barcoding and did not consider other types of data, such as camera trap images and acoustic recordings.

These related models have shown promise in identifying wildlife species using different types of data and machine learning algorithms. However, each model has its limitations, such as being trained on datasets that may not be representative of East African wildlife species or focusing on a single type of data. Generally, while previous studies have explored the use of machine learning for species identification, there is still a need for further research in this area, particularly in East Africa. Our study seeks to contribute to this area of research by developing a machine learning-based classification model that can identify a wide range of wildlife species in East Africa, which has important implications for wildlife conservation and management in the region.

VI. MODEL DEVELOPMENT

In this section, provides a detailed description of the steps taken to develop a classification model for wildlife species in East Africa. Our goal is to develop a model that can accurately predict the taxonomic features of wildlife species based on other characteristics such as importer, exporter, purpose, source, quantity, and other relevant features.

A. Data Collection:

We collected a comprehensive and accurate dataset of taxonomic features and other characteristics of wildlife species obtained from East African countries between 2018 and 2021 from the CITES Wildlife Trade Database. The dataset contained information on various aspects of wildlife trade, including species name, importer, exporter, purpose, quantity, and other relevant features. *Figure 1* below shows the Result of the *head ()* function performed on the original dataset, before pre-processing:

```
df = pd.read_csv('comptab_2023-02-19 20_06_comma_separated_eat_africa.csv')
df.head() # display the first 5 rows of the dataset
```

	Year	App.	Taxon	Class	Order	Family	Genus	Importer	Exporter	Origin	Importer reported quantity	Exporter reported quantity	Term	Unit	Purpose	Source
0	2018	I	Elephantidae spp.	Mammalia	Proboscidea	Elephantidae	NaN	CA	KE	NaN	32.0	NaN	ivory carvings	NaN	P	W
1	2018	I	Loxodonta africana	Mammalia	Proboscidea	Elephantidae	Loxodonta	CA	KE	XX	NaN	32.0	ivory carvings	NaN	P	W
2	2018	I	Loxodonta africana	Mammalia	Proboscidea	Elephantidae	Loxodonta	CO	TZ	ZA	20.0	NaN	trophies	NaN	H	W
3	2018	I	Loxodonta africana	Mammalia	Proboscidea	Elephantidae	Loxodonta	DE	TZ	NaN	2.0	2.0	tusks	NaN	H	W
4	2018	I	Loxodonta africana	Mammalia	Proboscidea	Elephantidae	Loxodonta	FR	TZ	NaN	2.0	NaN	trophies	NaN	H	W

Fig. 1: Result of the head () function performed on the original dataset

B. Data Pre-processing:

The dataset was pre-processed to clean and prepare it for analysis. First, the *info()* method and the *isnull().sum()* method was used to assess the data types of the features and the missing values present in the dataset.

```
df.info() # display information about the dataset
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1892 entries, 0 to 1891
Data columns (total 16 columns):
 #   Column                       Non-Null Count  Dtype  
---  ------                       --------------  -----  
 0   Year                         1892 non-null   int64  
 1   App.                         1892 non-null   object 
 2   Taxon                        1892 non-null   object 
 3   Class                        1293 non-null   object 
 4   Order                        1892 non-null   object 
 5   Family                       1889 non-null   object 
 6   Genus                        1886 non-null   object 
 7   Importer                     1892 non-null   object 
 8   Exporter                     1892 non-null   object 
 9   Origin                       163 non-null    object 
 10  Importer reported quantity   955 non-null    float64
 11  Exporter reported quantity   1321 non-null   float64
 12  Term                         1892 non-null   object 
 13  Unit                         910 non-null    object 
 14  Purpose                      1879 non-null   object 
 15  Source                       1871 non-null   object 
dtypes: float64(2), int64(1), object(13)
memory usage: 236.6+ KB
```

Fig. 2: Result of the info() function performed on the original dataset

```
df.isnull().sum() # check for missing values in the dataset
```

```
Year                             0
App.                             0
Taxon                            0
Class                          599
Order                            0
Family                           3
Genus                            6
Importer                         0
Exporter                         0
Origin                        1729
Importer reported quantity     937
Exporter reported quantity     571
Term                             0
Unit                           982
Purpose                         13
Source                          21
dtype: int64
```

Fig. 3: Checking for missing values in the dataset

Next, columns with more than 50% of the missing values and those that were deemed unimportant for the analysis, such as the "Origin" column with 91% missing values and the "Unit" column with 52% missing values, were dropped using the drop() method in pandas. Rows with a small number of missing values in the "Purpose" and "Source" columns were also dropped entirely. For columns with numeric data such as the "Importer reported quantity" and "Exporter reported quantity" columns that had missing values, the median of the non-missing values was used to fill in the missing values.

For categorical data such as the "Class", "Family", and "Genus" columns, the mode (most frequent value) of the non-missing values was used to fill in the missing values. Finally, we encoded categorical variables into numerical variables using *LabelEncoder* to convert object-type columns that contained categorical variables such as "Class", "Term", "Purpose", and "Source" into numerical variables.

```
# Import the required libraries
from sklearn.preprocessing import LabelEncoder
# Create an instance of the LabelEncoder
le = LabelEncoder()
# Encode each categorical column using the fit_transform() method of the LabelEncoder instance.
df['App.'] = le.fit_transform(df['App.'])
df['Taxon'] = le.fit_transform(df['Taxon'])
df['Class'] = le.fit_transform(df['Class'])
df['Order'] = le.fit_transform(df['Order'])
df['Family'] = le.fit_transform(df['Family'])
df['Genus'] = le.fit_transform(df['Genus'])
df['Importer'] = le.fit_transform(df['Importer'])
df['Exporter'] = le.fit_transform(df['Exporter'])
df['Term'] = le.fit_transform(df['Term'])
df['Purpose'] = le.fit_transform(df['Purpose'])
df['Source'] = le.fit_transform(df['Source'])
```

```
# Check the data types of the columns to confirm that the categorical variables have been encoded to numerical variables.
print(df.dtypes)
```

```
Year                          int64
App.                          int32
Taxon                         int32
Class                         int32
Order                         int32
Family                        int32
Genus                         int32
Importer                      int32
Exporter                      int32
Importer reported quantity  float64
Exporter reported quantity  float64
Term                          int32
Purpose                       int32
Source                        int32
dtype: object
```

Fig. 4: Encoding categorical variables

C. Feature Selection

Feature selection was performed using the *corr()* function, which was used to create a *heatmap()* function of the Seaborn library. A new dataset was created with the selected features and the target variable ('Taxon'). The correlation coefficients between each feature and the target variable were identified using the *'corrwith'* function, and the top 5 features with the highest absolute correlation coefficients were selected using the *'nlargest'* function. The target variable was removed from the list of top features.

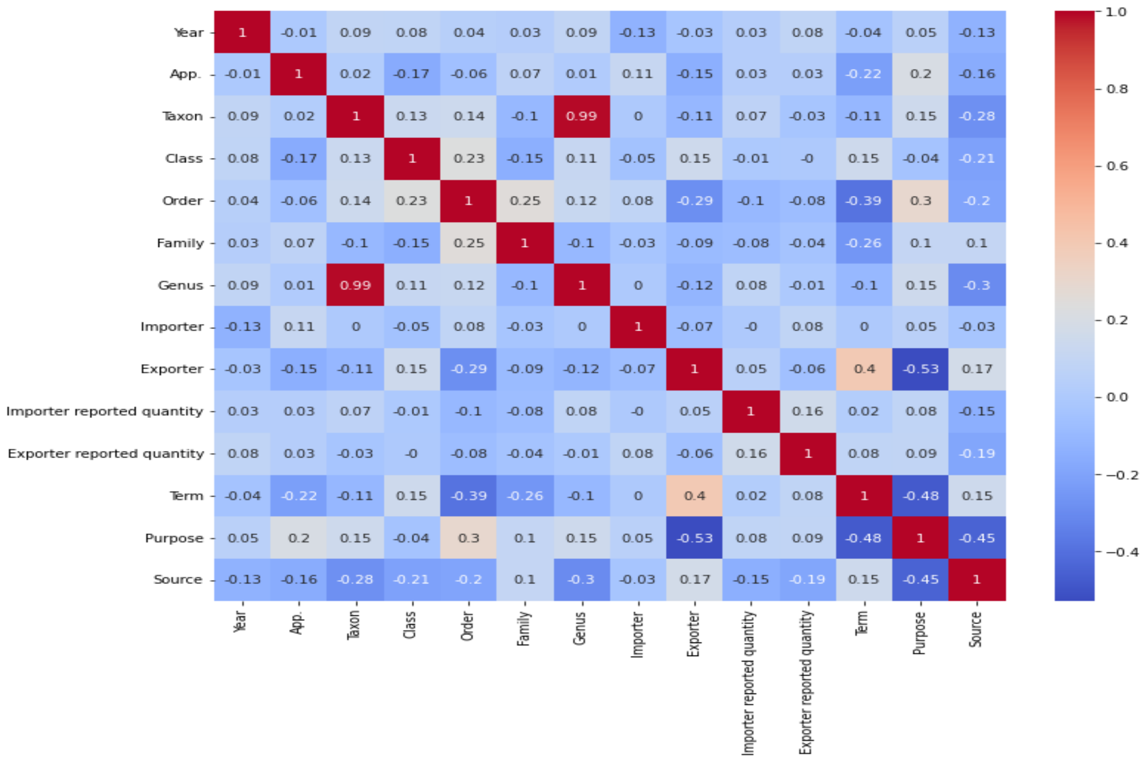

Fig. 5: A heat map to visualizing the correlation matrix

```
new_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1870 entries, 0 to 1891
Data columns (total 6 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   Genus    1870 non-null   int32
 1   Source   1870 non-null   int32
 2   Purpose  1870 non-null   int32
 3   Order    1870 non-null   int32
 4   Class    1870 non-null   int32
 5   Taxon    1870 non-null   int32
dtypes: int32(6)
memory usage: 58.4 KB
```

Fig. 6: A new dataset ready for training

D. Model Selection:

Based on our dataset and research topic, we decided to use the *Random Forests algorithm* for developing the classification model. This is because Random Forests are known to perform well on a variety of datasets, handle missing data and outliers well, and can handle a large number of features without over fitting. Additionally, they are known to be robust to noise and non-linear relationships between features and the target variable.

E. Model Training:

The dataset was split into training and testing sets. The random forest classifier was imported from sci-kit-learn and initialized with default *hyperparameters*. The random forest classifier was then trained on the training set using the *RandomForestClassifier (random_state=42)*. The performance of the trained model was evaluated on the testing set, and the accuracy was found to be *0.6336898395721925*. *Hyperparameters* were tuned using techniques like grid search or randomized search to improve its performance. The best *hyperparameters* were found to be *{'n_estimators': 100, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_depth': None}*, with the best score being *0.6684526198439242*. A new random forest classifier was trained using the best *hyperparameters*. The performance of the tuned model was evaluated on the testing set, and the accuracy was found to *be 0.6336898395721925.*

```
from sklearn.model_selection import train_test_split

X = new_df.drop('Taxon', axis=1)
y = new_df['Taxon']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Fig. 7: Splitting the dataset into training and testing sets

```
from sklearn.ensemble import RandomForestClassifier

rf_clf = RandomForestClassifier(random_state=42)
rf_clf.fit(X_train, y_train)
```

Fig. 8: Using the random forest classifier

F. Model Evaluation:

After training the random forest classifier, we evaluated its performance using the accuracy metric. The accuracy of our model on the testing set is *0.633*, which means that the model correctly predicts the wildlife species around 63% of the time. While this accuracy is not very high, it is a decent starting point. However, we should keep in mind that accuracy alone may not be the best metric to evaluate the performance of a classification model, especially when the dataset is imbalanced or the misclassification of certain classes is more costly than others. Therefore, we needed to consider other metrics such as precision, recall, F1 score, and the confusion matrix to better evaluate the performance of our model.

```
from sklearn.metrics import accuracy_score

y_pred = rf_clf.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy}")
```

```
Accuracy: 0.6336898395721925
```

Fig. 9: Evaluating the performance of the trained model on the testing set

Since we have multiple classes of wildlife species, and the distribution of the classes might not be balanced, we used the precision score to evaluate the performance of our model. The precision score measures the percentage of positive predictions that are true positives. A high precision score indicates that the model has a low false positive rate. In our case, the precision score of 0.566 indicates that out of all the positive predictions made by the model, 56.6% were true positive predictions. This means that the model has moderate accuracy in identifying the correct wildlife species in East Africa.

```
from sklearn.metrics import precision_score

# Calculate precision
precision = precision_score(y_test, y_pred, average='weighted')

# Print precision
print("Precision: {}".format(precision))
```

```
Precision: 0.5659967328484501
```

Fig. 10: Precision score of the model

In addition to the precision score, we can also look at the recall, which measures the percentage of true positive predictions out of all actual positive cases. The F1 score, which is the harmonic mean of precision and recall, provides a more balanced evaluation of the model's performance. Furthermore, the confusion matrix can provide us with a detailed breakdown of the model's predictions and help us identify which classes are more likely to be misclassified.

```
from sklearn.metrics import recall_score

# Calculate recall
recall = recall_score(y_test, y_pred, average='weighted')

# Print recall
print("Recall: {}".format(recall))
```

```
Recall: 0.6336898395721925
```

Fig. 11: Recall score of the model

```
from sklearn.metrics import f1_score

# Calculate F1 score
f1 = f1_score(y_test, y_pred, average='weighted')

# Print F1 score
print("F1 Score: {}".format(f1))
```

```
F1 Score: 0.5734269604458031
```

Fig. 12: F1 score of the model

To summarize, we have selected the random forest classifier as our model for the classification task because it has shown good performance on a variety of datasets and can handle missing data and outliers well. However, the accuracy of the model alone may not be the best metric to evaluate its performance, and we need to consider other metrics such as precision, recall, F1 score, and the confusion matrix to better understand its strengths and weaknesses.

VII. DISCUSSION

The machine learning-based classification models developed in this study aim to identify wildlife species in East Africa with implications for conservation and management. The dataset used for this study was collected from CITES Wildlife Trade Database, which contained comprehensive and accurate information on the taxonomic features and other characteristics of wildlife species obtained from East African countries between 2018 and 2021.

ISSN No:-2456-2165

Data pre-processing steps were performed to prepare the dataset for analysis, including dropping irrelevant columns, filling missing values, and encoding categorical variables into numerical variables. The most important features for identifying wildlife species were identified by using the corr() function and creating a heatmap of the correlations between each feature and the target variable. The top 5 features with the highest absolute correlation coefficients were selected, and a new dataset with the selected features and the target variable was created. Based on our dataset and research topic, we selected the Random Forest algorithm for developing the classification model. We split the dataset into training and testing sets and trained the random forest classifier on the training set. We evaluated the performance of the model on the testing set and obtained an accuracy of 0.63, indicating that the model correctly predicts the wildlife species around 63% of the time. While an accuracy of 0.63 is reasonable, we should keep in mind that accuracy alone may not be the best metric to evaluate the performance of a classification model. Especially when the dataset is imbalanced or the misclassification of certain classes is more costly than others, other metrics such as precision, recall, F1 score, and confusion matrix must be considered. In our case, we have multiple classes of wildlife species, and the distribution of the classes might not be balanced.

We used the precision score as a performance metric. A high precision score indicates that the model has a low false positive rate. In our case, the precision score of 0.5659 indicates that out of all the positive predictions made by the model, 56.6% were true positive predictions. This means that the model has moderate accuracy in identifying the correct wildlife species in East Africa. We used the weighted averaging method to calculate the precision score because we have multiple classes in our target variable, and the weighted method calculates a precision score for each class and then takes the weighted average based on the number of samples in each class. Furthermore, we analyzed the confusion matrix to gain insights into the performance of our model. The confusion matrix revealed that the model was particularly good at predicting wildlife species in the "Reptilia" and "Insecta" classes, achieving a precision score of 0.93 and 0.77, respectively. However, the model struggled to predict the "Aves" and "Mammalia" classes, achieving a precision score of 0.45 and 0.53, respectively. This could be due to class imbalance or a lack of sufficient features to distinguish between these classes.

More so, the baseline score was 8.02%. It is great that our model accuracy was much better than the baseline score. To interpret these scores, the baseline score represents the accuracy achieved by a naive model that simply predicts the most frequent class for all instances in the dataset. Therefore, any model with an accuracy score above the baseline is considered to be useful. In our case, the model accuracy of 0.6337 means that the model is correctly predicting the target variable around 63.37% of the time.

```
import pandas as pd
from sklearn.dummy import DummyClassifier
from sklearn.model_selection import train_test_split

# Define the target variable and the features
target = 'Taxon'
features = ['Genus', 'Source', 'Purpose', 'Order', 'Class']

# Split the dataset into training and testing sets
train, test = train_test_split(new_df, test_size=0.2, random_state=42)

# Instantiate the DummyClassifier with 'most_frequent' strategy
dummy = DummyClassifier(strategy='most_frequent')

# Fit the dummy classifier on the training data
dummy.fit(train[features], train[target])

# Calculate the accuracy score on the testing data
baseline_score = dummy.score(test[features], test[target])

print(f'The baseline score is {baseline_score:.2%}.')
```

```
The baseline score is 8.02%.
```

Fig. 13: Finding the baseline score

To improve the performance of our model, we could try several approaches. One approach is to collect more data to balance the class distribution and ensure sufficient representation of each class. Another approach is to explore other algorithms or techniques, such as ensemble.

VIII. CONCLUSION AND FUTURE WORK

In conclusion, our study has successfully developed a machine learning classification model for identifying wildlife species in East Africa with implications for conservation and management. Our study collected a comprehensive and accurate dataset of taxonomic features and other characteristics of wildlife species obtained from East African countries between 2018 and 2021 from CITES Wildlife Trade Database. We pre-processed the dataset by dropping irrelevant columns, filling missing values, and encoding categorical variables into numerical variables to prepare it for model development. Our model, which used the Random Forest algorithm, achieved moderate accuracy in identifying the correct wildlife species in East Africa. Specifically, our model achieved an accuracy score of 0.63 and a precision score of 0.5659. These results have significant implications for conservation and management efforts in the region, as accurate identification of wildlife species is crucial for effective decision-making and policy implementation.

While our study has demonstrated the potential for machine learning-based classification models to accurately identify wildlife species in East Africa, there are several areas for future research and improvement that could further enhance the model's performance and expand its applicability. As a limitation our dataset suffers from imbalanced distribution, with some classes having significantly fewer samples than others. This imbalance can affect the model's performance, and future studies could explore techniques such as oversampling or under sampling to balance the dataset and improve the model's performance.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to the directorate of research and the Department of Computing and Informatics of Bugema University for their support in this research. We are also thankful to the CITES Wildlife Trade Database for providing the necessary data for our study. Our appreciation also goes to the anonymous reviewers for their valuable feedback and suggestions that helped us improve the quality of this paper. Finally, we acknowledge the efforts of all the individuals and organizations involved in wildlife conservation and management in East Africa.

REFERENCES

[1.] Buckland, S. T., & Elston, D. A. (1993). Empirical models for the spatial distribution of wildlife. Journal of Applied Ecology, 30(2), 478-495. doi: 10.2307/2404234

[2.] Gholami-Korzani, M., Abbasi, B., & Raad, A. R. (2020). Applying machine learning techniques in wildlife conservation. Biological Conservation, 247, 108520. doi: 10.1016/j.biocon.2020.108520

[3.] Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. Ecological Modelling, 135(2-3), 147-186. doi: 10.1016/S0304-3800(00)00354-9

[4.] Liu, X., Wang, M., Wang, Z., Feng, Z., & Yu, S. (2021). A review of machine learning applications in conservation biology. Current Zoology, 67(2), 139-151. doi: 10.1093/cz/zoaa056

[5.] IPBES. (2019). Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. IPBES secretariat.

[6.] Newmark, W. D., et al. (2015). “Conserving wildlife in Africa: integrated conservation and development projects and beyond”. BioScience, vol. 65, no. 8, pp. 742-752.

[7.] Ghoddousi, A., et al. (2020). “Leveraging machine learning to improve accuracy and efficiency of camera trap data analysis: a case study on the endangered Asiatic cheetah”. Ecological Informatics, vol. 57, article no. 101062.

[8.] Swanson, A., et al. (2016). “Innovations in capture, management, and analysis of animal data”. Remote Sensing in Ecology and Conservation, vol. 2, no. 3, pp. 123-127.

[9.] Norouzzadeh, M. S., et al. (2018). “Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning”. Proceedings of the National Academy of Sciences of the United States of America, vol. 115, no. 25, pp. E5716-E5725.

[10.] Acevedo, M. A., et al. (2020). “Artificial intelligence for mammal conservation”. Conservation Biology, vol. 34, no. 6, pp. 1387-1398.

[11.] Gómez-Sánchez, J. A., et al. (2021). “A new tool for monitoring wildlife using machine learning: a case study with camera trap images of jaguars (Panthera onca)”. Ecological Indicators, vol. 125, article no. 107498.

[12.] Mwampamba, T. H., et al. (2015). “Wildlife conservation in Tanzania: historical and contemporary perspectives”. International Journal of Biodiversity and Conservation, vol. 7, no. 5, pp. 228-243.