ISyE 6740 – Fall2022
Final Report

Dangerous Animal Detection Through Audio Learning

Katherine Layton, Melirose Cleann M Liwag

ABSTRACT

Imagine you live on a couple acres in suburban America, it's 9:00PM and dark outside. You let your dog out for their nightly ritual and hear a cacophony of creatures nearby. Should you be concerned or delighted by the noises you hear? Now, imagine it's 5:30AM at the Dickerson Falls Trailhead and you hear an animal noise you do not recognize. Should you continue hiking? What if you had a way to record the creatures you hear to ease your mind or at a minimum to educate you on what creatures are in your immediate vicinity? Our research aims to compare audio files from various creatures to identify them and determine whether they are considered dangerous or not. Being able to identify an animal without seeing them, based purely on sound can help keep you and your pets safe.

1 Data Source

This dataset is from the Animal Sound Archive from a museum in Berlin. It contains around 120,000 sound recordings of various species and is available by request only. For our project, we decided to work with a subset of the overall data, namely the Carnivora animal group (for dangerous animals) and the Aves animal group (for bird species) https://www.museumfuernaturkunde.berlin/en/science/animal-sound-archive

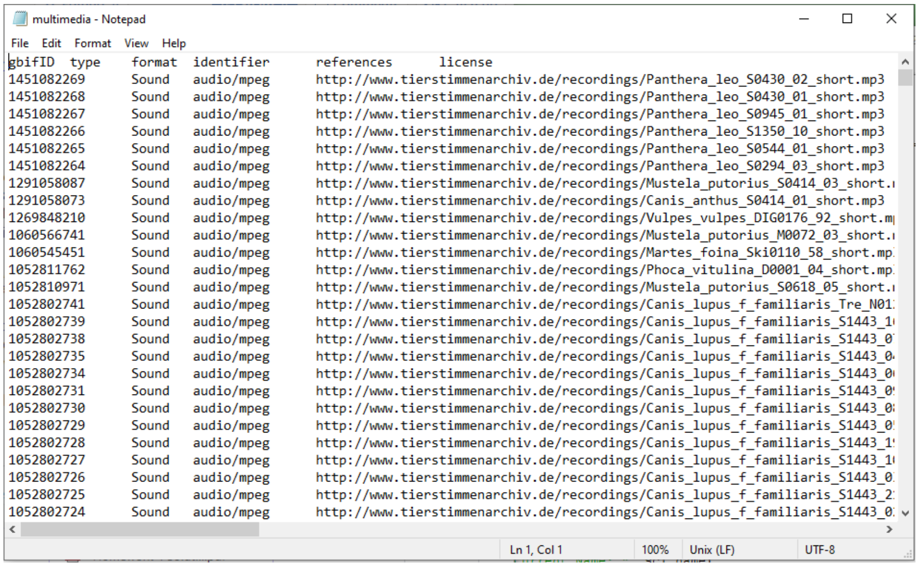
Carnivora Animal Group
GBIF.org (08 November 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.wxaq6n

Aves Animal Group
GBIF.org (08 November 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.svcpdb

1.1 Raw Data Set

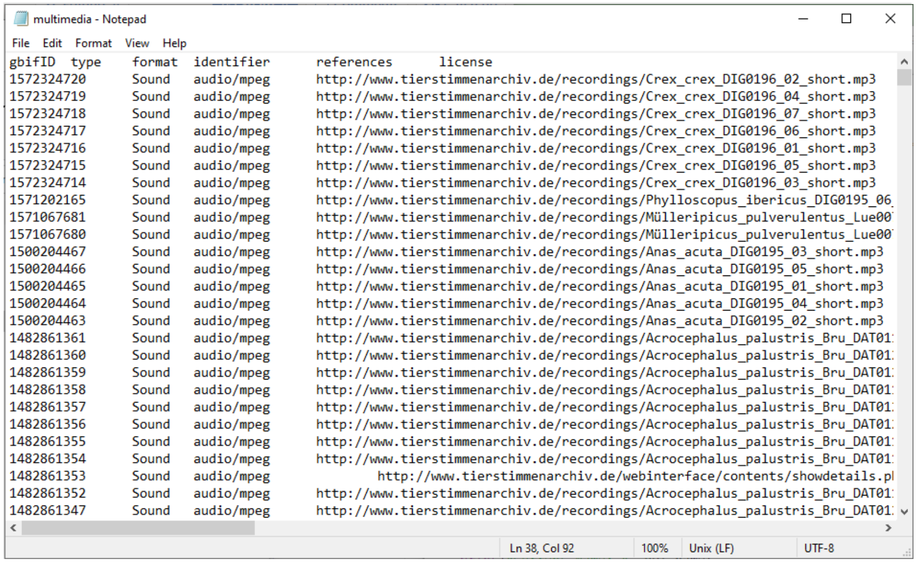
The downloaded data files came with multiple text files containing the data set we requested. Upon looking at the data files, some cleaning was required especially concerning the animal names within each animal group. The raw data set unfortunately does not include the common name of each animal which will make communicating our findings difficult for those who are not familiar with animal groups. Below is a screenshot of the downloaded data for the Carnivora animal ground and the Aves animal group respectively.

1.1.1 Carnivora



```
gbifID  type    format  identifier      references      license
1451082269      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Panthera_leo_S0430_02_short.mp3
1451082268      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Panthera_leo_S0430_01_short.mp3
1451082267      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Panthera_leo_S0945_01_short.mp3
1451082266      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Panthera_leo_S1350_10_short.mp3
1451082265      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Panthera_leo_S0544_01_short.mp3
1451082264      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Panthera_leo_S0294_03_short.mp3
1291058087      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Mustela_putorius_S0414_03_short.
1291058073      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Canis_anthus_S0414_01_short.mp3
1269848210      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Vulpes_vulpes_DIG0176_92_short.m
1060566741      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Mustela_putorius_M0072_03_short.
1060545451      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Martes_foina_Ski0110_58_short.mp
1052811762      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Phoca_vitulina_D0001_04_short.mp
1052810971      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Mustela_putorius_S0618_05_short.
1052802741      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Canis_lupus_f_familiaris_Tre_N01
1052802739      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Canis_lupus_f_familiaris_S1443_1
1052802738      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Canis_lupus_f_familiaris_S1443_0
1052802735      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Canis_lupus_f_familiaris_S1443_0
1052802734      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Canis_lupus_f_familiaris_S1443_0
1052802731      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Canis_lupus_f_familiaris_S1443_0
1052802730      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Canis_lupus_f_familiaris_S1443_0
1052802729      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Canis_lupus_f_familiaris_S1443_0
1052802728      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Canis_lupus_f_familiaris_S1443_1
1052802727      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Canis_lupus_f_familiaris_S1443_1
1052802726      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Canis_lupus_f_familiaris_S1443_0
1052802725      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Canis_lupus_f_familiaris_S1443_2
1052802724      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Canis_lupus_f_familiaris_S1443_0
```

1.1.2 Aves



```
gbifID  type    format  identifier      references      license
1572324720      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Crex_crex_DIG0196_02_short.mp3
1572324719      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Crex_crex_DIG0196_04_short.mp3
1572324718      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Crex_crex_DIG0196_07_short.mp3
1572324717      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Crex_crex_DIG0196_06_short.mp3
1572324716      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Crex_crex_DIG0196_01_short.mp3
1572324715      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Crex_crex_DIG0196_05_short.mp3
1572324714      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Crex_crex_DIG0196_03_short.mp3
1571202165      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Phylloscopus_ibericus_DIG0195_06_
1571067681      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Mülleripicus_pulverulentus_Lue00
1571067680      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Mülleripicus_pulverulentus_Lue00
1500204467      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Anas_acuta_DIG0195_03_short.mp3
1500204466      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Anas_acuta_DIG0195_05_short.mp3
1500204465      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Anas_acuta_DIG0195_01_short.mp3
1500204464      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Anas_acuta_DIG0195_04_short.mp3
1500204463      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Anas_acuta_DIG0195_02_short.mp3
1482861361      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Acrocephalus_palustris_Bru_DAT01
1482861360      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Acrocephalus_palustris_Bru_DAT01
1482861359      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Acrocephalus_palustris_Bru_DAT01
1482861358      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Acrocephalus_palustris_Bru_DAT01
1482861357      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Acrocephalus_palustris_Bru_DAT01
1482861356      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Acrocephalus_palustris_Bru_DAT01
1482861355      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Acrocephalus_palustris_Bru_DAT01
1482861354      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Acrocephalus_palustris_Bru_DAT01
1482861353      Sound   audio/mpeg              http://www.tierstimmenarchiv.de/webinterface/contents/showdetails.p
1482861352      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Acrocephalus_palustris_Bru_DAT01
1482861347      Sound   audio/mpeg      http://www.tierstimmenarchiv.de/recordings/Acrocephalus_palustris_Bru_DAT01
```

2 Methodology

2.1 Preprocessing

The website allowed for filtering only certain animal groups we wanted to use as well as grabbing only data containing audio files for each animal. The animal groups we decided to work with included the Carnivora group (1,089 data points) and the Aves group (13,849 data points). With this filtering system, the downloaded data set did not contain any missing audio values, however, further cleaning was required. Empty columns were removed, and additional columns were added for further convenience for us moving forward. As shown before, the raw data set itself was downloaded as a text file and thus was converted into a Pandas data frame for easier manipulation. The columns below were added to both data sets:

- *Scientific Name* – Contains the Scientific name of each animal within the group. Obtained through regular expressions while going through each audio URL name *(identifier* column)
- *Common Name* – Intuitively, using the common name of the animals within the group would be easily understandable by others. Unique scientific names will be collected from the previous column and an array of common names found on the web will be used to fill this column

The two data sets were concatenated to retain all information from both data sets and the resulting joined data frame was used to build our models for this project.

A script was created to batch download the 14,938 mp3 files from the web urls in the joined Pandas dataframe. Some files were not able to download from the website, resulting in 12,367 mp3s stored after script completion. The .mp3 files were then converted to .wav files for easier audio processing using Pydub library in python, keeping only the first 5 seconds of the audio to ensure a consistent duration resulting in only 12,263 files converted to .wav format. Features were extracted from the .wav files via the Librosa library in Python. The primary feature extracted and used for analysis was the MFCC (Mel-Frequency-Cepstral Coefficients). Librosa extracts the MFCCs by first slicing the signal into frames, then computes an estimate of power for each frame, applies a mel (log) transformation to the power spectrum, then performs a discrete cosine transformation (DCT) of the former. The resulting MFCC is shaped (n_mfcc's, n_frames) or in this case (20, 431). These MFCCs were stored in the row corresponding to the sample label as a Numpy array. The Melspectrogram was also extracted and used, which has similar steps to the above MFCC, except represents the spectrogram output on the mel scale rather than coefficients.

For the simplicity of this project, we are assuming all animals under the Carnivora group will be a danger to pets and others. All bird species under the Aves group will also be assumed to be harmless and are considered "not dangerous." Exploratory data analysis revealed that the dataset is highly unbalanced. Out of the 12,375 files that were converted appropriately, only ~1000 were from the Carnivora class, meaning the Carnivora class is underrepresented by the full data. Using the full Aves dataset would likely bias any model toward the Aves set, so it was decided that a subset of the Aves should be used. Within the two groups, Aves and Carnivora, there were numerous species. First, any species that had less than 20 or greater than 80 sound samples was removed from the full dataset. Then, random sampling was performed to extract a subset of the Aves data making the length of the Aves dataset closer to the length of the Carnivora dataset. The full dataset resulting dataset was 1274 samples consisting of almost equal parts Carnivora to Aves as well as an approximately equal number of species within each group.

The MFCC values varied significantly for each row, and the size of the MFCC features made modeling cumbersome. The MFCC values were first scaled and split into 90% training and 10% test sets due to the small sample size. Finally, the principal components were extracted using Sklearn PCA function and setting “n_components” to capture 95% of the explained variance.

3 Literature Review and Models Approach

The main method for processing and classifying audio files is essentially the same as image classification. An audio file is converted to a spectrogram or two which represents all frequencies in that sound file over time as an image. Sound consists of speed, amplitude, frequencies, and direction, but currently only frequency and amplitude are considered important in machine learning. Audio files are the manifestations of sound in electronic form; the waveform we see in any audio sample is the frequency over time. The spectrogram represents this audio file waveform, with the x-axis showing time and the y-axis, frequency in hertz. The color of the spectrogram reveals the amplitude in decibels. Any machine learning technique that can be applied to images can now be applied to audio with the spectrogram as an input. There are also several packages available in python for audio analysis such as Librosa and Pytorch Audio.

A literature review revealed that animal sound classification has been heavily studied for important work in biodiversity and climate change monitoring starting primarily with bird call classification. Lucio et al. kicked off the task with the use of spectrograms and an SVM classifier along with 10-fold cross-validation to classify 46 species with 77.65% accuracy [1]. Ramirez et al. explored the idea of using feature extraction techniques common to speech recognition models. To improve the bird species recognition task, Mel frequency cepstral coefficients (MFCC) and its counterpart (IMFCC) which captures the inverse or missing data from the MFCC algorithm, were compared determining that IMFICC yields better model performance [2]. More current work focuses on the use of deep learning or convolutional neural networks to achieve higher accuracy in detecting if a sound belongs to a specific species without relying on feature extraction such as Salamon et al. model achieving 96% accuracy [3].

This research improves upon the current state in the following innovative ways:

1. Attempts to classify audio sounds to determine if a species is dangerous
2. Uses custom dataset, consisting of 2 datasets merged and curated for this project
3. Explores use of different modeling techniques than what is typically used for state of the art species audio classification.

Formatted: Indent: Left: 0.25"

Following PCA, we tested several different models to attempt to classify if a species is dangerous or not dangerous based on the digital output of an audio file. PCA was important due to the high number of features present in the MFCC and Melspectrogram outputs, simply putting these into the model would be computationally expensive and time consuming. With PCA, we extract and use only the components needed to explain the greatest amount of variance.

In this paper, we applied various machine learning techniques to identify whether a creature is "dangerous" or "not dangerous" using image classification techniques learned from ISYE6740 on audio files. The intuition behind choosing models for this project was based on the similarities between audio processing and image processing explained in the literature reviewed. The Gaussian Mixture Model was selected due to its probabilistic nature of helping distinguish between groups in image segmentation. The project team manually applied the labels for "dangerous" and "not dangerous", with the idea that two distinct groups would appear in the model outputs as two distinct distributions. One-Class SVM was chosen as it works the best with binary classifications. Since this project is simply reduced to a "Dangerous" or "Not Dangerous" label, a binary classification type model was the best approach.

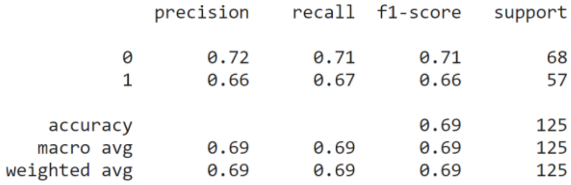
A Gaussian Mixture Model (GMM) was developed using sklearn Gaussian Mixture function with 2 components (dangerous, not dangerous), a tolerance of 1e-3, and k-means to initialize the weights. Random initialization of weights was also tested but did not perform as well as kmeans. The PCA output and the associated labels (0 for "Not Dangerous", 1 for "Dangerous") were used for the GMM model, the PCA output as an input and the labels for verifying the results against the actual labels in a classification report. The One-Class SVM model was developed using the sklearn OneClassSVM function with the default gamma value for the initial modelling. Further exploration of gamma values will be explained as well.

4 Evaluation and Final Results

The Gaussian Mixture Model was 69% accurate in classifying a "dangerous"(1) or "not dangerous" (0) species given the audio input. The results of the classification report are below:

Formatted: Font: Not Bold, No underline

Formatted: Centered



```
              precision    recall  f1-score   support

           0       0.72      0.71      0.71        68
           1       0.66      0.67      0.66        57

    accuracy                           0.69       125
   macro avg       0.69      0.69      0.69       125
weighted avg       0.69      0.69      0.69       125
```

The modeling correctly classifies 72% of the not dangerous and 66% of the dangerous audio samples as shown in the precision column and performs similarly for recall and f1-score. The support shows the number of occurrences of that class within the test dataset, which is a fairly small dataset with the 90/10 Train/test split established earlier. While the classification results of this model are promising, more data would need to be collected to establish the viability of the model. The under-sampling technique added some initial bias, and more bias is likely, due to the heavy number of bird species in the dataset. All of the "not dangerous" species are birds, which in hindsight could make this classification more of a "bird" or "not bird" detector. This makes the accuracy level achieved look less impressive due to the inherent bias.

Lasso Regression was also considered for modeling with the assumption that the data output of the spectrogram or waveform would be fairly sparse, but it turns out to not be the case. There are very few 0 values found in the feature extraction outputs for the audio files used in this project. As such, Lasso Regression was not considered for the final modeling.

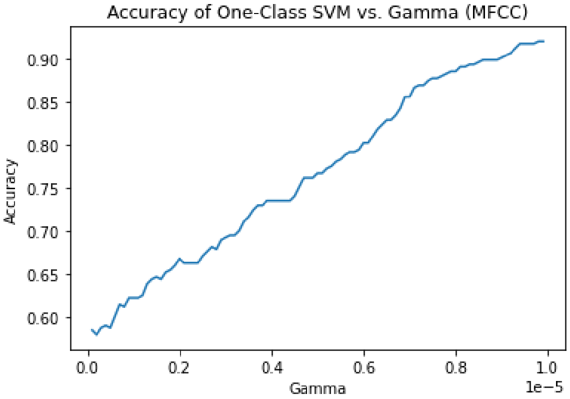
One final model we considered using is One-Class SVM. Prepared similarly to the previous GMM approach, however, PCA was tuned with n_components = 4 instead of using n_components = 0.95. We found that smaller dimensions provided better accuracy for the SVM model overall. We also changed the labels of each data point to be a binary component to make classification with SVM easier. With this, the spectrogram and label data were shuffled and partitioned to 70% training and 30% testing as we hoped to avoid some overfitting with an otherwise already small sample size. For the training data set, the data points that are labeled as "Not Dangerous", or simply reduced to the binary 0, are used to train the SVM where data points labeled as "Not Dangerous" are the inliers and everything else is an outlier. This model was built with both the MFCC and the Mel Spectrogram data sets to compare how well SVM performs in classifying different types of audio. The SVM resulted in an accuracy of 59.09% for the MFCC and 58.29% for the Mel Spectrogram.

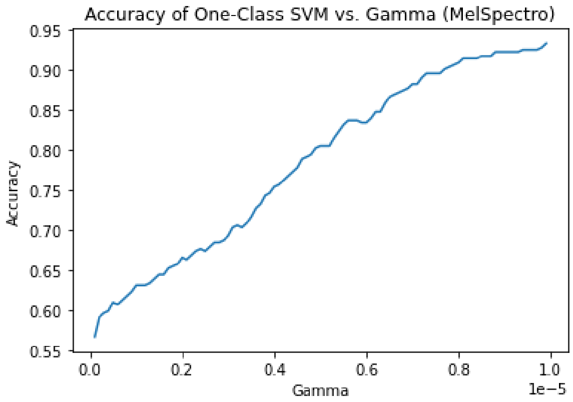
With such low accuracies from the SVM model, we decided to test different values of gamma with the RBF kernel for the SVM model. We tested with values 0.0000001 to 0.00001 and found that a gamma value of `9.800000000000001e-06` gave us the best accuracy result of 91.98% for the MFCC and a gamma value of `9.9e-06` gave us the best accuracy result of 93.32% for the Mel Spectrogram. We were surprised with the high accuracy of the SVM model overall but considering the classification to be a binary "Bird" or "Not Bird," it is reasonable that SVM was able to classify them easily. Furthermore, our data set was biased as the "Not Dangerous" data set is just birds and the "Dangerous" data set is filled with a variety of carnivorous animals which may have completely different audio features. Training the SVM model with this bias makes the near-perfect accuracy understandable in a way. It is also good to note that using a value of gamma higher than 0.00001 yielded perfect accuracy for both MFCC and the Mel Spectrogram. The relationship between the gamma value and the accuracy of the SVM is represented in the following figures:

Formatted: Font: (Default) Verdana, 11 pt

Formatted: Font: (Default) Verdana, 11 pt

Formatted: Font: 11 pt





Formatted: Font: +Body, 12 pt, Not Bold, No underline

5 Future Considerations

Below are some considerations for future research on this dataset and the danger detection mechanism.

- Converting audio to an input for machine learning turns out to be more of an art than a science. There are many different features that can be extracted, different ways to convert from .mp3 to .wav, and a plethora of parameters to consider when doing so. This research stuck to most of the defaults for the audio conversion and feature extraction tasks due to lack of domain knowledge in audio processing combined with a limited amount of time to learn due to the deadline of the project. For future research, more time should be spent figuring out what section of the audio to extract for the best results. For instance, this project extracted the first 5 seconds for simplicity, but it may be better to produce overlapping frames or extract a central region of the .wav output.
- As mentioned in preprocessing, this project classified all animals under the Carnivora as dangerous. All bird species under the Aves group were classified as "not dangerous." More time would need to be spent on establishing what constitutes as dangerous or not. An extra data set and column can be added to correctly identify dangerous and non-dangerous carnivores as well as dangerous and non-dangerous birds.
- Since all the "non-dangerous" group were types of birds, it is possible that the models are simply detecting "bird", "not bird". The bias inflicted with the dataset used in this research may tamper with models' performance to other tests or real-world datasets if creatures other than birds are counted as "not dangerous". It is recommended to collect more data and add more species to each category.
- A visual representation of the results of this project could be a nice addition. Also further expanding this project to something usable by the public could provide great value to outdoorsmen and rural people alike.

Formatted: Font: Verdana, 11 pt

6 Project Plan

The table below represents a high-level breakdown of the project plan and tasks. The team has contributed an equal amount of effort to the project and worked collaboratively on all tasks. This project was heavily frontloaded with a large amount of time required for cleaning/merging the two datasets from the web, acquiring the audio files from the URL, manipulating the files for ease of processing, and ensuring the data was in the correct format for machine learning model input.

Task	Responsible	Start	End	Status
Topic Identification	Meli/Katy	10/24/2022	10/28/2022	Completed
Request Data	Meli	11/4/2022	11/7/2022	Completed
Proposal Draft	Katy	11/5/2022	11/8/2022	Completed
Finalize Proposal	Meli/Katy	11/8/2022	11/9/2022	Completed
Data Cleaning	Meli	11/7/2022	11/15/2022	Completed
Extracting Features	Meli/Katy	11/15/2022	11/18/2022	Completed
EDA	Katy/Meli	11/18/2022	11/23/2022	Completed
Modeling	Katy/Meli	11/23/2022	11/30/2022	Completed
Report Draft	Katy	11/30/2022	12/4/2022	Completed
Final Report	Meli/Katy	12/4/2022	12/6/2022	Completed

7 CITATIONS

D. R. Lucio, Y. Maldonado and G. da Costa, "Bird species classification using spectrograms," 2015 Latin American Computing Conference (CLEI), 2015, pp. 1-11, doi: 10.1109/CLEI.2015.7359990.
https://ieeexplore.ieee.org/document/7359990/citations?tabFilter=papers#citations

A. D. P. Ramirez, J. I. de la Rosa Vargas, R. R. Valdez and A. Becerra, "A comparative between mel frequency cepstral coefficients (MFCC) and inverse mel frequency cepstral coefficients (IMFCC) features for an automatic bird species recognition system", *Proc. IEEE Latin Amer. Conf. Comput. Intell. (LA-CCI)*, pp. 1-4, Nov. 2018
https://ieeexplore.ieee.org/abstract/document/8625230

J. Salamon, J. P. Bello, A. Farnsworth and S. Kelling, "Fusing shallow and deep learning for bioacoustic bird species classification," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 141-145, doi: 10.1109/ICASSP.2017.7952134.
https://ieeexplore.ieee.org/document/7952134