CS6140: Machine Learning

Mini Project

Kshitij Verma (NUID -001305285) verma.ks@northeastern.edu
Vedant Jain (NUID -001305281) jain.ve@northeastern.edu

a) Project title – **Classification of Various Bird Species**

b) Names of the team members and their Northeastern University emails.

1) Vedant Jain - jain.ve@northeastern.edu
2) Kshitij Verma - verma.ks@northeastern.edu

c) Objectives and significance (1-2 paragraphs):
(a) Describe what the goal of the project is, why is it important, and your motivation for doing it.

Distinguishing and Identifying bird species is a difficult assignment regularly bringing about vague labels. Even professional bird watchers sometimes disagree on the species given an image of a bird. It is a troublesome issue that pushes the restrictions of the visual capacities for both the computers and humans. Albeit distinctive flying bird species share a similar fundamental arrangement of parts, however they can differ significantly in shape and appearance. Intraclass variance tends to be high because of variety in background and lighting and extraordinary deviance in pose (e.g., birds flying, swimming winged creatures, and roosted feathered creatures that are mostly blocked by branches).

- For many decades, ornithologists are facing problems in bird species identification. Ornithologists require studying all the details of birds such as their existence in the environment, their biology, their distribution, their ecological impact, etc. Our Project plans to utilize the power of Machine Learning to aid this problem and provide a model with higher accuracy for correctly identifying bird species.

d) Background (1-2 pages)

(a) Introduce all important concepts and background information.

- Fundamentally bird recognizable proof is done visually or acoustically. The primary visual segments include birds shape, its wings, size, present, shading, and so forth. In any case, while considering the parameters, Monthly season must be thought about because of the fact that bird wings change as indicated by their development. The acoustics segments contain the melodies and call that birds make. The imprints that recognize one bird from another are likewise helpful, for example, bosom spots, wing bars which are depicted as slender lines along the wings, eye rings, crowns, eyebrows. The shape of the beak is often an important aspect as a bird can recognize uniquely. The attributes of feathered creatures, for example, shape and stance are the generally used to recognize winged animals. For the most part specialists can recognize a bird from its outline since this trademark is hard to change. It can likewise be separated utilizing its tail. The tail can be perceived from various perspectives, for example, indented, long and pointed, or adjusted. At times legs are likewise utilized for perceiving a picture in design long, or short.

- By only considering a solitary or only one parameter won't yield a precise outcome. Along these lines, various parameters are to be considered so as to get suitable accurate results. The size of a bird in a picture fluctuates relying on components, for example, the goals, separation between the bird and the camera, and the central separation of the focal point. Hence, in light of a handy perception for an enormous number of pictures, pictures are separated based on shading which comprises different pixels. Inside and out it is discovered that more noteworthy the picture quality more noteworthy is its accuracy.

(b) Search the literature and describe previous work on this problem.

- Previous Works can be Majorly categorized into two approaches:
    1) Identification using Acoustic features or bird sounds.
    2) Identification using Visual Features

1) Identification using Acoustic features or bird sounds:

- Acoustic Approaches to the issue have centered upon several bird species classification challenges with closely related, but different, task descriptions have been held during the most recent couple of years. The interest and participation in these challenges have been high which indicates that these are relevant problems and that there is a need to solve them. The primary takeout from these difficulties are referenced as follows:

- In MLSP 2013 [2], which is the IEEE International Workshop on Machine Learning for Signal Processing (MLSP) declared a bird species classification challenge in the year of 2013. The test was to decide the entirety of the acoustically active bird species in each audio recording of a test set with a total of 19 different bird species., the winning solution were random forests trained on probabilities derived from template matching of species-specific spectrograms.

- The winning solution of NIPS4B 2013 [3] by winner used these results as a starting point but introduced an additional set of features statistically derived from the audio files. Winners also used a similar method to win the Bird CLEF 2015 challenge.

- However, during the Bird CLEF 2016 challenge, it was indicated that convolutional neural networks trained on spectral data computed from the sound recordings could outperform other state-of-the-art systems This thesis uses the work of Sprengel [4] as a starting point and a baseline, and explore the use of a new convolutional neural network method called deep residual neural networks as well as a new data augmentation technique called multiple-width frequency delta data augmentation.

2) **Identification using Visual Features:**

- The Scientific Papers which used Visual features approach are being mentioned here:

1) **Bird Species Categorization Using Pose Normalized Deep Convolutional Net**

- Author [1] proposes an architecture for fine-grained visual categorization that approaches expert human performance in the classification of bird species. Architecture first computes an estimate of the object's pose; this is used to compute local image features which are, in turn, used for classification. The features are computed by applying deep convolutional nets to image patches that are located and normalized by the pose, experiments advance state-of-the-art performance on bird species recognition, with a large improvement of correct classification rates over previous methods (75% vs. 55-65%).

2) **A Baseline for Large-Scale Bird Species Identification in Field Recordings**

- In 2016, Sprengel et al. [4] demonstrated the superior performance of convolutional neural networks (CNN) for the classification of bird sounds. Following that approach, the Author were able to improve the performance on a larger dataset containing 1500 different species with our 2017 Bird CLEF participation presenting an implementation of a streamlined workflow built on the most fundamental principles of visual classification using CNN.

3) **Bird Species Identification using Deep Learning**

- In this paper [5], instead of recognizing a large number of disparate categories, the problem of recognizing a large number of classes within one category is investigated that of birds.
  By using deep convolutional neural network (DCNN) algorithm an image converted into grayscale format to generate autograph by using tensor flow, where the multiple nodes of comparison are generated.

(c) If there exists previous work on the problem, describe what makes your work distinct or particularly interesting.

- Here we are focusing on using Visual features and we discerned that the main quintessence of the project was to comprehend the complexities of various Machine Learning Algorithms and to realize which algorithm gives great outcomes for which use case.

e) Methods (2-4 pages)
(a) Describe your data and how you will obtain it.

- We Used the Caltech-UCSD Birds 200 (CUB-200) image dataset which is annotated with 200 bird species (mostly North American). It was created to enable the study of subordinate categorization, which is not possible with other popular datasets that focus on basic level categories (such as PASCAL VOC, Caltech-101, etc.). The images were downloaded from the website Flickr and filtered by workers on Amazon Mechanical Turk. Each image is annotated with a bounding box, a rough bird segmentation, and a set of attribute labels.

(b) Describe your methodology and implementation. Be as detailed as necessary.

First of all, we used 10-fold cross validation technique to evaluate predictive models by partitioning the original sample into a training set to train the model, and a test set to evaluate it.

Our approach is to develop a 3 different baseline models:

1) Simple Models
2) Mid-Layer Models
3) Tree Based Model

1) Simple Model: We developed Simple models using **KNN and SVM Algorithms**

   As**, KNN** is a **non-parametric, lazy** learning algorithm with local approximation. Its purpose is to use a database in which the data points are separated into several classes to predict the classification of a new sample point. It is non-parametric as the model structure is determined from the data. Therefore, KNN probably should be one of the first choices for a classification when there is little or no prior knowledge about the distribution data.

   On the other hand, an **SVM model** classifies data by determining the optimal hyperplane that separates observations according to their class labels. The central concept is to accommodate classes separable by linear and non-linear class boundaries. Most classification problems require complex decision boundaries in order to make an optimal separation, i.e. correctly classify new observations in the test set based on observations in the training set.
   If a data distribution is essentially non-linear, the primary strategy is to transform the data to a higher dimension. In this case, the original data points of one class are mapped using mathematical functions known as “kernels.” The mapped data points allow the SVM to find the optimal line to separate classes, rather than constructing a complex curve.

   If training data is much larger than no. of features(m>>n), KNN is better than SVM. SVM outperforms KNN when there are large features and lesser training data.

2) Mid - Layer Model: We developed models using **Logistic Regression and Naive Bayes Algorithms.**

   Logistic regression is a machine learning modelling method to predict when the outcome variable is categorical. The basic principle is to maximize the likelihood and minimize the log cost function through the gradient descent with iterations. This helps in predicting the likelihood that the incident is going to happen or not with different parameters like AUC, prevalence, precision, recall.

   Naive Bayes is a classification technique which follows the rules of Bayes theorem. Classifier performs best with the categorical variables and for numerical variables with assumption to be normally distributed. The basic principle of naïve Bayes classifier works on the assumption that for a feature in a class is not related to any other feature. It means every single feature contributes independently to the probability of the output variable.

3) Tree Based Model: We'll be developing the model using **Decision Tree**

   The decision tree is a type of supervised learning method which works on a tree-like structure while making decisions considering the different features. Each node of the tree is subdivided into two more categories of Yes and No which represents some feature conclusion until there is a dead end. The simplest way to reach the decision which is mostly relatable to real life.

Looking at the results of these algorithms, we got a baseline for future techniques that are implemented. Next, we did **feature selection** using **Principal Component Analysis (PCA)** and trained models again on these reduced features. PCA chooses the best features from the combination of features, these features represent the highest variance proving important for the model training. We have used our own implementation of PCA and not the Scikit Version of PCA. All the models will be again trained on PCA reduced Dataset. The model giving the best accuracy on these reduced features will be our final model.

While, for improving the Decision Tree Model, we have used **Ensemble Techniques Random Forest and Gradient Boosting**. Random forest is a collaboration of many decision trees combined on a platform or bag. The concept of bagging is done to make a decision tree for different features all together. Concept of controlling the best split is by sub setting when features are higher in number. Through this, more unique trees are formed and reduce the correlation between trees. Gradient boost classifier is one of the highest rated machine learning technique and more frequently used technique. The algorithm used to improve the errors of the earlier used decision tree, which is included in the next tree to improve the learning rate. Learning rate decides how quick the error is fixed in every decision tree.

Finally, we have also trained a **Neural Network with 3 Hidden Layers** as with input layer as Convolution 2D with RELU Activation Function while used SoftMax Function as the output layer Activation Function. Used Stochastic Gradient Descent as the optimizer and Cross entropy for calculating loss.

(c) Describe evaluation strategy. Be as detailed as necessary.

1. AUC and ROC curve score

- AUC - ROC curve is a performance measurement for classification problems at various threshold settings. ROC is a probability curve and AUC represent degree or measure of separability. It tells how much model is capable of distinguishing between classes. Higher the AUC, better the model is at predicting 0s as 0s and 1s as 1s. An excellent model has AUC near to the 1 which means it has good measure of separability. A poor model has AUC near to the 0 which means it has the worst measure of separability.

2. Classification Accuracy:

- Classification accuracy is defined as the number of correct predictions divided by the total number of predictions, multiplied by 100.

(f) Results: -

-The project aims to quantify the qualitative description of different bird species using machine learning techniques and use it as an effective tool for bird species identification.

a)
Results obtained from Machine Learning models-

I. **KNN**
Area under ROC curve: 0.56315
Accuracy: 0.3318

Results with PCA:
Area under ROC curve: 0.59133
Accuracy: 0.3622

II. **SVM**
Area under ROC curve: 0.62243
Accuracy: 0.4482

Results with PCA:
Area under ROC curve: 0.68486
Accuracy: 0.4974

III. **Naïve Bayes**
Area under ROC curve: 0.50172
Accuracy: 0.2927

Results with PCA:
Area under ROC curve: 0.52113
Accuracy: 0.3165

IV. **Logistic Regression**
Area under ROC curve: 0.74283
Accuracy: 0.5961

Results with PCA:
Area under ROC curve: 0.76876
Accuracy: 0.6141

V. **Decision Tree**
Area under ROC curve: 0.52853
Accuracy: 0.3148

Results with PCA:
Area under ROC curve: 0.56284

Accuracy: 0.3358

VI. Gradient Boosting

Area under ROC curve: 0.65288
Accuracy: 0.4644

Results with PCA:
Area under ROC curve: 0.67121
Accuracy: 0.4897

VII. Random Forest

Area under ROC curve: 0.66017
Accuracy: 0. 4719

Results with PCA:
Area under ROC curve: 0.69328
Accuracy: 0.5102

Results obtained from Neural Network model-
Epoch 1/30
9823/9823 [==============================] - 563s 57ms/step - loss: 5.8412 - accuracy: 0.0093
Epoch 2/30
9823/9823 [==============================] - 562s 57ms/step - loss: 5.6294 - accuracy: 0.0323
Epoch 3/30
9823/9823 [==============================] - 549s 56ms/step - loss: 5.3013 - accuracy: 0.0644
Epoch 4/30
9823/9823 [==============================] - 535s 54ms/step - loss: 4.9671 - accuracy: 0.0940
Epoch 5/30
9823/9823 [==============================] - 541s 55ms/step - loss: 4.5976 - accuracy: 0.1306
Epoch 6/30
9823/9823 [==============================] - 537s 55ms/step - loss: 4.3474 - accuracy: 0.1545
Epoch 7/30
9823/9823 [==============================] - 541s 55ms/step - loss: 3.9680 - accuracy: 0.2030
Epoch 8/30
9823/9823 [==============================] - 539s 55ms/step - loss: 3.5207 - accuracy: 0.2612
Epoch 9/30
9823/9823 [==============================] - 531s 54ms/step - loss: 3.0738 - accuracy: 0.3241

```
Epoch 10/30
9823/9823 [==============================]  - 532s 54ms/step - loss: 2.6930 -
accuracy: 0.3905
Epoch 11/30
9823/9823 [==============================]  - 456s 46ms/step - loss: 2.3354 -
accuracy: 0.4597
Epoch 12/30
9823/9823 [==============================]  - 445s 45ms/step - loss: 1.9761 -
accuracy: 0.5312
Epoch 13/30
9823/9823 [==============================]  - 447s 46ms/step - loss: 1.6817 -
accuracy: 0.5919
Epoch 14/30
9823/9823 [==============================]  - 449s 46ms/step - loss: 1.4669 -
accuracy: 0.6441
Epoch 15/30
9823/9823 [==============================]  - 456s 46ms/step - loss: 1.2255 -
accuracy: 0.6952
Epoch 16/30
9823/9823 [==============================]  - 449s 46ms/step - loss: 1.0355 -
accuracy: 0.7381
Epoch 17/30
9823/9823 [==============================]  - 453s 46ms/step - loss: 0.8837 -
accuracy: 0.7750
Epoch 18/30
9823/9823 [==============================]  - 453s 46ms/step - loss: 0.7186 -
accuracy: 0.8144
Epoch 19/30
9823/9823 [==============================]  - 467s 48ms/step - loss: 0.6236 -
accuracy: 0.8348
Epoch 20/30
9823/9823 [==============================]  - 449s 46ms/step - loss: 0.5299 -
accuracy: 0.8594
Epoch 21/30
9823/9823 [==============================]  - 451s 46ms/step - loss: 0.4525 -
accuracy: 0.8761
Epoch 22/30
9823/9823 [==============================]  - 455s 46ms/step - loss: 0.3885 -
accuracy: 0.8937
Epoch 23/30
9823/9823 [==============================]  - 460s 47ms/step - loss: 0.3323 -
accuracy: 0.9075
Epoch 24/30
```

9823/9823 [==============================] - 454s 46ms/step - loss: 0.2754 -
accuracy: 0.9239
Epoch 25/30
9823/9823 [==============================] - 490s 50ms/step - loss: 0.2385 -
accuracy: 0.9337
Epoch 26/30
9823/9823 [==============================] - 499s 51ms/step - loss: 0.2065 -
accuracy: 0.9416
Epoch 27/30
9823/9823 [==============================] - 499s 51ms/step - loss: 0.2065 -
accuracy: 0.9481
Epoch 28/30
9823/9823 [==============================] - 489s 49ms/step - loss: 0.2038 -
accuracy: 0.9533
Epoch 29/30
9823/9823 [==============================] - 485s 48ms/step - loss: 0.2165 -
accuracy: 0.9593
Epoch 30/30
9823/9823 [==============================] - 499s 51ms/step - loss: 0.2065 -
accuracy: 0.9627

b)

Results Table

No.	**Algorithm**	**Without PCA**		**With PCA**	
		Accuracy	AUC	Accuracy	AUC
1.	KNN	0.3318	0.56315	0.3622	0.59133
2.	SVM	0.4482	0.62243	0.4974	0.68486
3.	Naïve Bayes	0.2927	0.50172	0.3165	0.52113
4.	Logistic Regression	0.5961	0.74283	0.6141	0.76876
5.	Decision Tree	0.3148	0.52853	0.3358	0.56284
6.	Gradient Boosting	0.4644	0.65288	0.4897	0.67121
7.	Random Forest	0. 4719	0.66017	0.5102	0.69328

Algorithm	**Epochs**	**Accuracy**
Neural Network (Using Keras)	30	0.9627

Observation-

We trained and tested our algorithms on the original data set by splitting it using 10-fold cross validation. In 10-fold cross-validation, the original sample is randomly partitioned into 10 equal size subsamples. Of the 10 subsamples, a single subsample is retained as the validation data for testing the model, and the remaining 9 subsamples are used as training data.

The tables show the results we observed on implementing algorithms as mentioned in the above section. We initially observed low accuracy with basic implementation of Naive Bayes, KNN and Decision Tree with fairly equivalent low AUC scores being obtained. This is due to the large number of features in training data. This is what we call curse of dimensionality while low AUC score suggests low class separation capability. However, these models when trained on the PCA dimensionally reduced data, improvement in the accuracy and AUC score is observed by 2-3 % approximately. That means the PCA is successful in Selecting important features having high variance but overall, these models didn't prove to be helpful.

Moreover, SVM yields better accuracy and AUC scores as compared to most of the other algorithm as it uses the Kernel Trick to find the best decision boundary. It is clear and more powerful way of learning non-linear functions. On the other hand, Logistic Regression performed best on the original as well as PCA applied data with highest accuracy. It performs better on removal of correlated attributes hence PCA plays an important role in regards to the performance of Logistic Regression.

Additionally, Ensemble methods like Random Forest and Gradient Boosting did not perform as good as expected on original as well as PCA applied data this is due to the fact the data is sparse, this is one downside of Random Forest.

On Training the Neural Network with 3 hidden layers on this data gives us outstanding accuracy of 96.27%. Because they have ability to learn non-linear and complex relationships, it generalizes better than any ML algorithm for this data.

We believe such an accuracy for a 200-class classification problem is fairly decent.

c)

Figure 1 shows the comparison of testing accuracy on Original Data vs Dimensionality reduced Data (PCA applied) for different learning methods that we implemented.

Figure 2 shows the comparison of Area under ROC score on Original Data vs Dimensionality reduced Data (PCA applied) for different learning methods that we implemented.

Figure-1

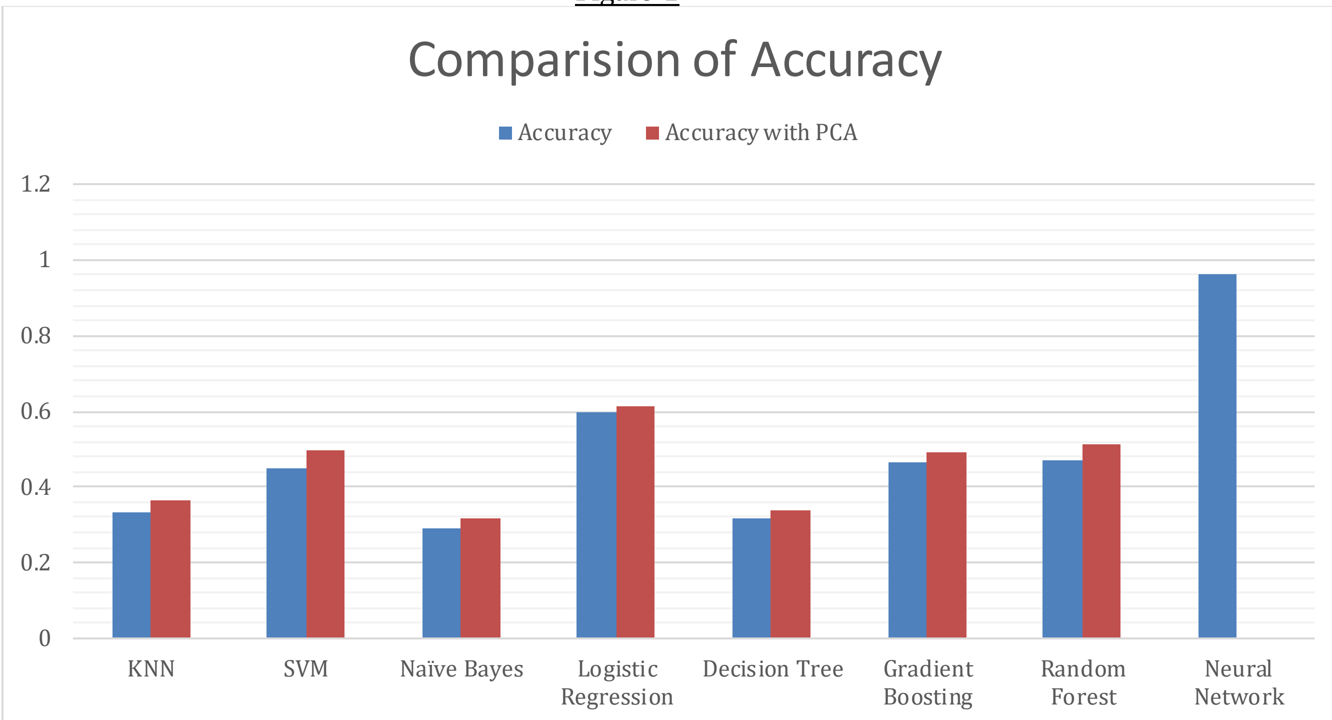
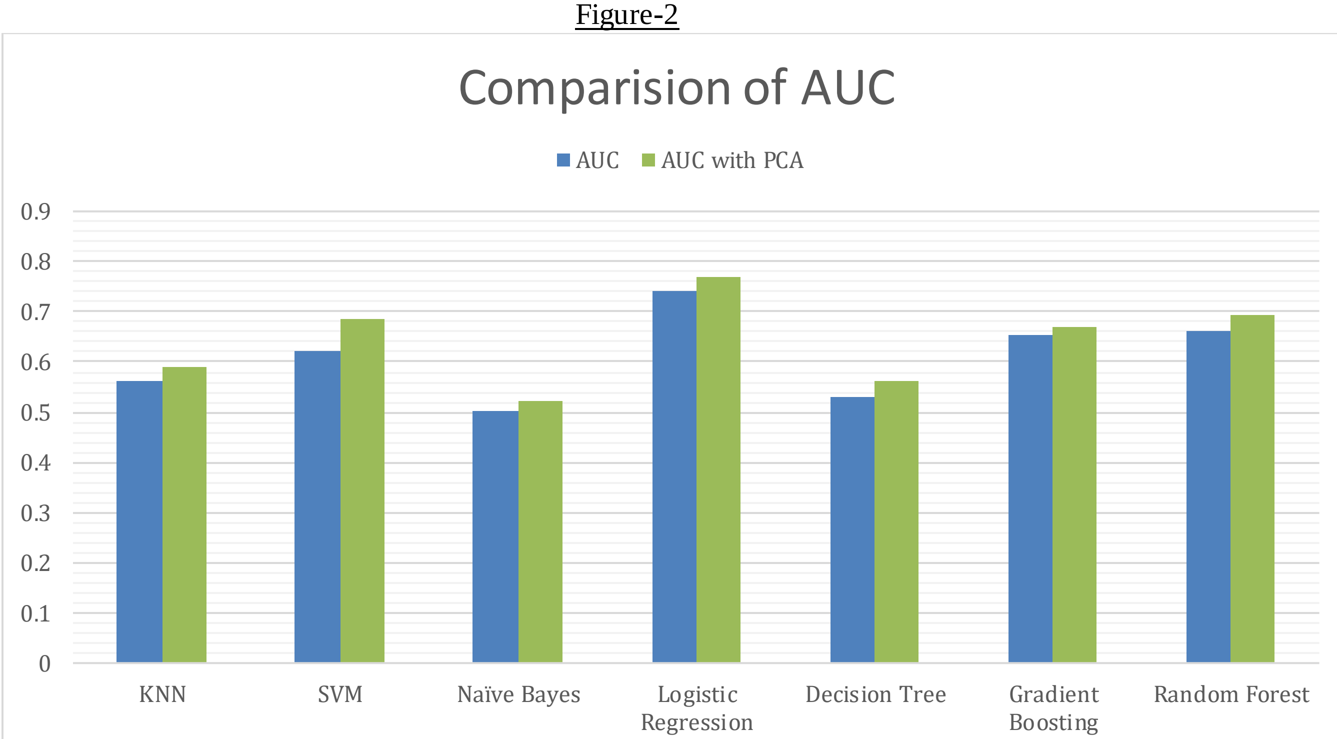


Figure-2



(g) Conclusions:

After analyzing all the models, we selected Neural Networks as our best model as it worked best on our data which has a high rate of sparsity and achieved a test accuracy of 96.2%. Also due to the high sparsity in data other Machine learning algorithms did not performed well as expected.

Additionally, the machine learning algorithm which performed well is Logistic Regression as it performed best on the original as well as PCA applied data with highest accuracy. It performs better on removal of correlated attributes hence PCA plays an important role in regards to the performance of Logistic Regression.

Principal Component Analysis did an impactful work by increasing the accuracy of every algorithm by 2-3% by removing the co-related features in the data resulting in less training time and reducing the overfitting of the model.

Future Work-

1. We implemented Neural Networks and when we ran it on our machine for just 3 hidden Neurons which took a significant amount of time for just 30 epochs. So, we can try to run Neural Networks on high performance computing machines.

2. Computer vision algorithms can be used for automatic feature extraction.

3. We can develop an Android/iOS application that identifies a bird in real time on clicking its photo.

h) Individual tasks (1-3 paragraphs)

Task was divided as-

1) Kshitij – I implemented Creation of all Baseline Models which include SVM, KNN, Naive Bayes, Logistic Regression and Decision Tree. Also, I created own implementation of 10-fold Cross Validation over the entire Dataset. Retrained all the models on Dimensionally Reduced Dataset obtained after using PCA and Evaluated them.

2) Vedant - I performed the task of Feature selection or dimensionality reduction using own implementation using Principal Component Analysis (PCA). Also implemented Ensemble Techniques Random Forest and Gradient Boosting on Data. Finally created the Neural Network Model with 3 hidden layers and Evaluated the same.

This project is a joint effort of two individuals as to achieve the desired results on such a dataset within the given timeframe, both individuals worked together coordinating with each other and helping one another whenever one faces a problem while doing their desired task.

g) References

(a) List books, scientific papers, web sites etc. that you referenced in the proposal body.

[1] Steve Branson et al. “Bird Species Categorization Using Pose Normalized Deep Convolutional Nets”. In: CoRR abs/1406.2952 (2014). URL: http://arxiv.org/abs/1406. 2952.

[2] Forrest Briggs, Yonghong Huang, Raviv Raich, Konstantinos Eftaxias, Zhong Lei, William Cukierski, Sarah Frey Hadley, Adam Hadley, Matthew Betts, Xiaoli Z. Fern, Jed Irvine, Lawrence Neal, Anil Thomas, Gabor Fodor, Grigorios Tsoumakas, Hong Wei Ng, Thi Ngoc Tho Nguyen, Heikki Huttunen, Pekka Ruusuvuori, Tapio Manninen, Aleksandr Diment, Tuomas Virtanen, Julien Marzat, Joseph Defretin, Dave Callender, Chris Hurlburt, Ken Larrey, and Maxim Milakov. The 9th annual MLSP competition: New methods for acoustic classification of multiple simultaneous bird species in a noisy environment. IEEE International Workshop on Machine Learning for Signal Processing, MLSP, 2013.

[3] Mario Lasseck. Bird song classification in field recordings: Winning solution for NIPS4B 2013 competition. Proc. of int. symp. Neural Information Scaled . . ., pages 1–6, 2013

[4] Elias Sprengel, Martin Jaggi, Yannic Kilcher, and Thomas Hofmann. Audio Based Bird Species Identification using Deep Learning Techniques. 2016.

[5] Prof. Pralhad Gavali, Ms. Prachi Abhijeet Mhetre, Ms. Neha Chandrakhant Patil, Ms. Nikita Suresh Bamane, Ms. Harshal Dipak Buva. Bird Species Identification using Deep Learning.

[6] https://datascience.aero/metrics-evaluating-machine-learning/

[7] https://towardsdatascience.com/feature-selection-using-random-forest-26d7b747597f