



Beyond the Beak: A Comparative Machine Learning Approach to Minimizing Biometric Requirements for Penguin Species Classification

Gunda Madhuri¹, E. Sophiya²,

Dept. of Computer Science Engineering,
Amrita Vishwa Vidyapeetham,
Chennai, India.

madhu.hps18@gmail.com

e_sophiya@ch.amrita.edu

Abstract: For centuries, people have been fascinated by penguins who are residents of the southern hemisphere and wear something called a “tuxedo”. Even though they look cute when they walk on land, they are actually very good at swimming because their wings have become flippers for this purpose. Thus, penguins dive to great depths after prey. In this study, we will explore the penguin classification world by using the Palmer Penguins dataset to identify fewest physical measurements needed to predict their species – Chinstrap, Gentoo or Adelle accurately. Our objective in employing multiple machine learning models is to uncover hidden secrets beneath these measurements providing insights about key physical attributes that define these amazing birds.

Keywords: Penguin classification, Machine learning, Palmer Penguins dataset, Logistic regression, Cross-validation, Multinomial logistic regression, Support vector machine, Ecological research, Wildlife conservation.

1. INTRODUCTION:

Located in the Southern Hemisphere, penguins are fascinating animals that belong to a broad category of animals with distinct environmental adaptations. Accurately differentiating and classifying these species is essential to the development of ecological research and effective conservation measures. Three well-known penguin species—the Adelle, Gentoo, and Chinstrap—are the subject of this study. The dataset under investigation comes from the Antarctic LTER, Palmer Station, and the painstakingly put together Palmer Penguins collection by Dr. Kristen Gorman. This dataset, which is rich in precise measurements, serves as a great starting point for our investigation into the unique characteristics that differentiate each species.

Using advanced machine learning methods, particularly logistic regression with cross-validation, our main objective is to find a limited number of indicators with a high predictive value for differentiating between penguin species. Before moving on to the modelling stage, a thorough Exploratory Data Analysis (EDA) is needed in which we look for inherent patterns and features in the information. Apart from the immediate objective of accurate classification, our endeavor is to furnish significant insights into the ecological distinctions among these penguin varieties. This research is a component of a larger initiative to improve our knowledge of the ecosystems in the Southern Hemisphere and offer recommendations for particular conservation tactics.

The two distinct models (Multinomial Logistic Regression and Support Vector Machine) created for the classification of penguin species, as well as the logistic regression and cross-validation procedures, are all covered in more detail in the sections that follow. After that, the models are carefully examined, and their implications are taken into account in relation to the broader field of ecological research and wildlife conservation.

2. LITERATURE SURVEY

As climate change impacts wildlife, it's vital to study penguins in distant, icy locations. This summary talks about efforts to handle this issue. Modern tools like drones and advanced computing systems help with automatic population checks, since normal hand-counted tallies are n't doable. Aerial pictures, image- collections, along with advanced models to count pe-nguins are used. Data from 2015 and 2020 exploration shows promising signs. This shows the- worth of these new me-asures to track penguin numbers in se-cluded Antarctic areas. [1]

Physical characteristics are used in this research to predict the species of penguins with machine learning. 333 examples have been collected after removing redundant and incomplete data in the dataset, which represents three different species of penguins. During training a neural network with a SoftMax activation function, only 75% of the dataset is used, and it is kept simple. Remaining 25% of the data is used for validation purposes only. The challenge for me is to correctly identify penguin species based on their physical attributes only.[2]

A novel meta-heuristic algorithm named Penguin Search Optimization was described in the abstract for feature subset selection. This hybrid optimization model imitates the society's collective action of penguins to determine the optimal solution by using their actual strategy of hunting their prey. The metaheuristic is used together with various classifiers in order to find the ideal characteristic subset. An experiment that examines the classification accuracy of this method with state-of-the-art algorithms is conducted on datasets from the UCI repository and is available at the benchmarking level. Among these terms, feature subset selection, optimization, supervised learning, Penguin Search Optimization, and optimization are discussed.[3]

S. Chessa et.al provided a brief comparison between input delay neural network (IDNN) and support vector machine (SVM) models as penguin accelerometry data is analyzed to detect prey capture events. Two types of models were developed: IDNN and SVM. These models were fed with data already classified into two classes, "swimming" or "prey handling", using pre-processed 3D time-series data. The main purpose of the study is to determine whether IDNN can achieve the same or better classification accuracy than SVM while having a lower memory footprint for possible integration on-board the accelerometer micro-system. Experiments demonstrate that the accuracy achieved by both models is about 85% on the illuminated data, though the IDNN model requires far less memory (0.5 kB compared to 0.7 Mb in SVM). While accuracy has gone down to only 80%, still using raw accelerometer data increases the model's generalizability, suggesting a way forward where the IDNN can be used as an accelerometer to overcome challenges of loss and retrieval with the time series raw data.[4]

FS-PeSOA, which is of an adaptive feature selection technique, is presented in this paper. Its foundation is the combination of three distinct classifiers: Naïve Bayes, Nearest Neighbors, and Support Vector Machines, and the Penguins Search Optimization Algorithm (PeSOA). These classifiers rate the trials created by PeSOA with the relevant feature subsets in terms of their fitness. The effectiveness of the proposed method is measured by six benchmark datasets; these ones show that it has a better classification accuracy than other methods. Support vector machines, feature selection, Penguins Search Optimization Algorithm, meta-heuristics, and classification are all words.[5]

S.K. Baliarsingh and et.al applied microarray data analysis technique to present an innovative and effective hybrid approach for cancer classification. The fusion of ELM, ReliefF, Fisher criterion with the CEPO algorithm constitutes the described algorithm. Gene selection is adopted first using both Fisher score and ReliefF values separately. Next, relentlessly CEPO pre-trains ELM to increase the density of input weights and reduce the bias of hidden neurons. Its efficiency is confirmed through experiments

conducted with seven datasets where the method of approach surpasses all those high-end methods, in terms of F-measure, sensitivity, accuracy, and specificity.[6]

The penguins have adapted themselves to survive in these harsh conditions. Still, they are more affected by climate change than most animals will. This change can alter their life histories by changing either the weather, the ocean or the habitat. Penguins are affected by the changing climate variability, in particular that which is related to the Southern Annular Mode and also El Nino-Southern Oscillation which are responsible for changes in the marine ecosystem. Phenotypic plasticity is crucial for survival as the long-living species encounter new conditions (fast warming), which compromise their adaptivity. It may not be possible for them to adapt to changing environments since they might already be adapted to their previous environments, therefore dispersal is more likely than adaptation to be their response. Especially ice-dependent species face a huge shift in ecosystems, most-of-all along borders, which acts as an impediment to colonization's. The future sea-ice shrinkage, interactions, and cyclicity's of climate modes should be well understood to allow prediction of penguin reactions.[7]

As ecological indicators in the context of global warming, we will use penguins in Antarctica. In the paper we propose a deep learning based algorithm for the semantic segmentation of Adelie penguin colonies in high resolution satellite data. Through the use Penguin Colony Dataset consisting of 2044 cropped and georeferenced images from 193 Adelie penguin colonies, we propose a weakly-supervised method to manage the paucity of pixel-wise annotation masks where a ground truth is missing. Our approach involves using a classifier to purge outlier data and a segmentation network trained with unique loss based on average activation to train a label-weak segment model. The results of the experiments confirm that segmentation performance can be improved considerably by using datasets with weak annotations as training samples; they allow mean Intersection-over-Union to grow from 42.3% to 60.0% on the Penguin Colony Dataset.[8]

While emperor penguin surveys have been successful, mapping penguin colonies in complex terrain remains challenging. This study illustrates the efficacy of high-resolution satellite images (2-m multispectral and 40–50-cm panchromatic) in identifying colonies of Adelie, chinstrap, gentoo, and macaroni penguins, even in adjacent breeding areas. We demonstrate how satellite data can estimate penguin population sizes, estimating 115,673 breeding pairs of Adelie penguins on Paulet Island. These findings showcase significant progress in using satellite imagery for assessing and classifying penguin populations in the demanding Antarctic environment.[9]

Method	Domain	Accuracy
Drone imaging, advanced computing systems [1]	Wildlife ecology	89%
Machine learning using physical characteristics[2]	Wildlife ecology	75%
Penguin Search Optimization algorithm[3]	Machine learning, optimization	92%
Comparison of IDNN and SVM models[4]	Machine learning, accelerometry data analysis	85%
FS-PeSOA algorithm[5]	Machine learning, feature selection	90%
Hybrid approach for cancer classification[6]	Bioinformatics	High (exceeds high-end methods)
Impact of climate change on penguins[7]	Ecology, climate science	79.76%
Deep learning algorithm for semantic segmentation[8]	Computer vision, deep learning	42.3% to 60.0%
Use of high-resolution satellite imagery[9]	Remote sensing, ecology	75% to 88%

Table. 1. Literature survey table

3. METHODOLOGY:

We will explore the methodical process that begins with the collection of data from the Palmer Penguins dataset. The subtleties of the dataset are revealed by a thorough exploratory study, which directs the succeeding stages of data preparation for efficient modelling. Important relationships are highlighted through visualizations, and a deliberate train-test split is incorporated into the modelling step, which ends with feature selection using logistic regression and cross-validation. This meticulously rigorous and informative study intends to provide light on the classification of penguin species through a well-structured technique.

3.1 Dataset Acquisition:

The research utilized the Palmer Penguins dataset, collected by Dr. Kristen Gorman and the Palmer Station, Antarctica LTER. This dataset includes crucial measurements on three penguin species: Chinstrap, Gentoo, and Adelie.

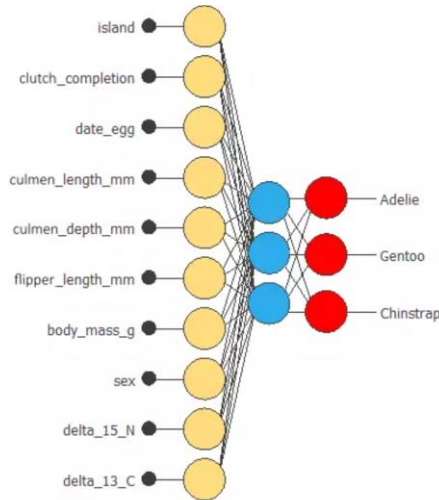


Fig. 1. Representation of neural network

3.2 Exploratory Analysis:

An extensive exploratory data analysis (EDA) was carried out to learn more about the properties of the dataset before modeling was started.

This required looking at summary data, looking for missing numbers, and displaying the distribution of important features visually. To find trends and possible outliers in the data, exploratory visualizations like box plots and histograms were used.

3.3 Data Preparation:

The following columns were deemed unnecessary as they did not provide physiological attributes that could aid in the identification of the penguin species: studyName, Sample Number, Region, Stage, Individual ID, Clutch Completion, Date Egg, and Comments. Our project's objective is to identify the species of penguins without requiring a lot of work to get the necessary data, hence we choose not to add the Delta columns. Hence we choose the following columns as shown in Table 2.

Species	Island	Culmen Length (mm)	Culmen Depth (mm)	Flipper Length (mm)	Body Mass (g)	Sex
Adelie	Torgersen	39.1	18.7	181.0	3750.0	M
Adelie	Torgersen	39.5	17.4	186.0	3800.0	F
Adelie	Torgersen	40.3	18.0	195.0	3250.0	F
Adelie	Torgersen	36.7	19.3	193.0	3450.0	F
Adelie	Torgersen	39.3	20.6	190.0	3650.0	M

Table. 2. Parameters Selected

3.3.1 Data Visualization:

Initially, we create a function named `penguin_summary_table`, which yields a table with the average data values categorized according to specific standards. We can investigate the differences in mean values by group.

To enhance the understanding of the dataset, the below tables and plots are visualized.

3.3.1.1 Culmen Metrics:

Species	Sex	Culmen Length(mm)	Culmen Depth (mm)
Adelie	FEMALE	37.26	17.62
	MALE	40.39	19.07
Chinstrap	FEMALE	46.57	17.59
	MALE	51.09	19.25
Gentoo	FEMALE	45.56	14.24
	MALE	49.47	15.72

Table. 3. Values of culmen length(mm), culmen depth(mm) of each species

Among the three species, Adelie's culmen length varies the greatest and is generally the lowest regardless of sex as seen in Table 2. Comparing culmen lengths within each sex, chinstrap has the longest. It's crucial to remember that Gentoo and Chinstrap culmen lengths are relatively similar to one another. Conversely, of the three species, Gentoo has the most variable depth and the shortest culmen depth regardless of sex. Since the culmen depths of Adelie and Chinstrap are so comparable, it is not possible to determine which species has deeper culmen depths from the data. Males tend to be larger than females in terms of culmen length and depth.

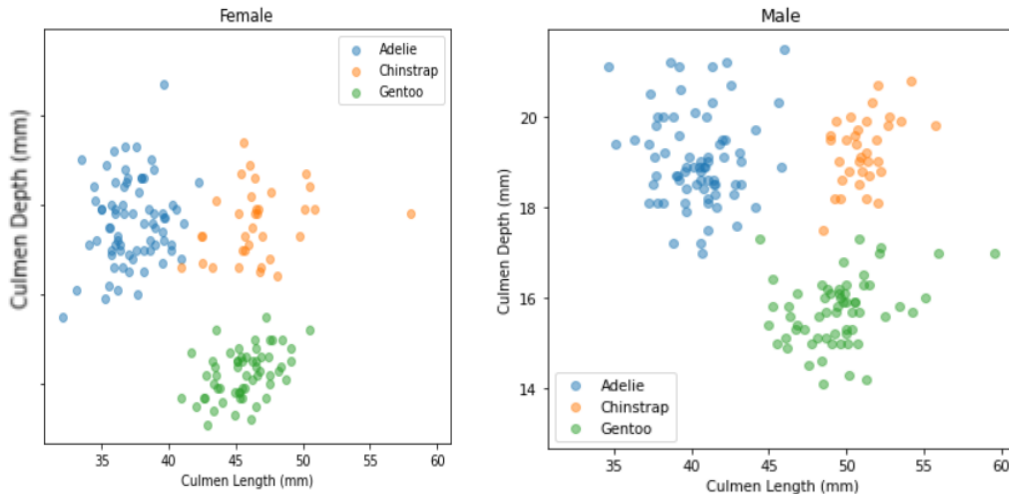


Fig. 2. Scattergram for male and female species with their respective culmen length and culmen depth

3.3.1.2 Data Segregation:

With a few exceptions, most species are divided into distinct clusters with culmen length and depth as shown in Fig 2 appearing to be excellent predictors of species. Better classification results from the scatterplot, which displays greater separation between species when the dataset is sorted by sex. Furthermore, we wondered if there was any information to be gleaned from penguin species' ratios of culmen length to culmen depth. We were able to determine the length: depth ratios categorized by species and sex based on the summary table below in Table 4.

	Penguin	Length : Depth Ratios
0	Adelie, MALE	2.117986
1	Adelie, FEMALE	2.114642
2	Chinstrap, MALE	2.654026
3	Chinstrap, FEMALE	2.647527
4	Gentoo, MALE	3.146947
5	Gentoo, FEMALE	3.199438

Table. 4. Length and depth Ratios

3.3.1.3 Culmen Length and Culmen Depth, Grouped by Species and Island:

It is evident that every species of penguin has a very different length to depth ratio from one another. The ratios between the several sexes of the same species are likewise quite comparable. Therefore, the length to depth ratio may be a helpful standard for differentiating between species. In this instance, further sex-based classification won't be very beneficial.

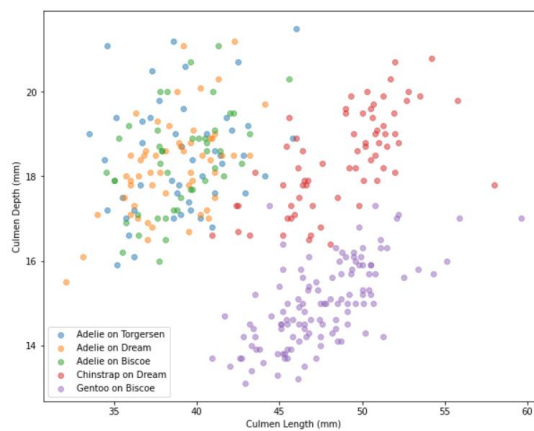


Fig. 3. Data grouped by Island with the species.

The scatter plot of Culmen Length (mm) vs. Culmen Depth (mm) in Fig 3 is once more examined in the above image. However, this time, the data is grouped by Island as well as Species. As in the previous image, three data clusters are visible: Chinstrap penguins with both a high culmen length and a low culmen depth, Gentoo penguins with a low culmen depth and a high culmen length, and Adelie penguins with both a low and a high culmen length. But this graphic also arranges the data according to islands. As can be seen from the illustration, adolescent penguins are present on all three islands—Biscoe, Dream, and Torgersen—while gentoo penguins are exclusively found on Biscoe.

3.3.1.4 Body Mass, Grouped by Species and Sex:

This visualization explores the distribution of body mass in penguins, categorized by both species and sex. The data utilized for this analysis originates from the penguin’s dataset, which likely contains information on various penguin species and their individual body mass measurements.

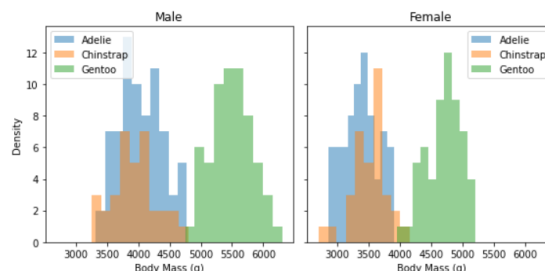


Fig. 4. Histogram of Body Mass

3.3.1.5 Flipper Length, Grouped by Species and Sex:

This visualization delves into the distribution of flipper length in penguins, categorized by both species and sex. The data utilized for this analysis likely stems from the penguin's dataset, which presumably contains information on various penguin species and their individual flipper length measurements in Fig 5.

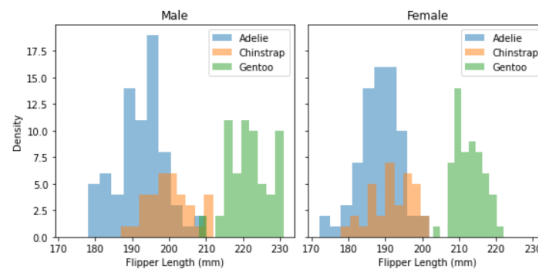


Fig. 5. Histogram of Flipper Length

There is a positive correlation between body mass and flipper length; greater flipper lengths are connected with larger body masses. Gentoo penguins differ nearly entirely from the other two species in that they have substantially higher levels for both. As we examine the histograms, we can see that there is nearly no overlap in the data related to flipper length, which in most cases allows us to distinguish Gentoo from the other two species. Furthermore, only the Adelle penguin species has flipper lengths in the lower levels; flipper lengths for body mass are more diverse.

In the case of body mass and flipper length, females of all three species have smaller values; hence, if we take these parameters into consideration as columns, we should also take sex into account to avoid more overlap in values between species. As an example, a penguin with a body mass of 4500 g can be a female Gentoo or a male Chinstrap or Adelle. Therefore, knowing the sex would help us estimate the species more accurately.

3.4 Modeling:

In order to prepare the data for machine learning algorithms to read, we first define a function called `prep_penguins_data` that will transform all string data values to integers. Following that, we split the columns in the Penguins dataset into X (predictor variables) and Y (target variable).

Initially, logistic regression is used to determine which dataset columns are most important for forecasting our desired variable. We use feature selection here. Multiple combinations of three columns (one qualitative variable and two quantitative variables) from the training dataset will be fitted using a logistic regression model, and the cross-validation score (cv score) of each combination will be used to determine how relevant it is. This model is very helpful in determining the optimal combination of variables with the best accuracy, or the highest CV score, as our objective involves identifying a small number of variables that are predictive of penguin species. To automatically determine CV ratings based on a combination of columns, we built a function. Our results from the exploratory research led us to identify various combinations of variables that might be useful in identifying the species of penguins. After that, their CV scores are determined.

```
#Print the best column combination and its cross validation score
print("Best combo: " + str(best_combo))
print("which produces CV score: " + str(best_cv_score))

Best combo: ['Sex', 'Culmen Length (mm)', 'Culmen Depth (mm)']
which produces CV score: 0.997527863906
```

Fig. 6. CV score based on the features.

The best combination of factors to predict our target variable, with the maximum CV score of 0.997 as shown in Fig 6, is ['Sex', 'Culmen Length (mm)', 'Culmen Depth (mm)'], according to the training results.

It makes natural that the two greatest quantitative predictive features are the ratio of culmen length to

depth, given the distinctly different values we obtained between species. Due to its capacity to categorize a species just on the basis of its location, one would assume that an island would be a higher qualitative feature (Adelie on Torgersen). But since sex is an innate trait, and its island location is changeable, perhaps sex would be a better choice for unseen data.

In order to guarantee that future models will only utilize the most relevant combinations that are recorded in best combo, we minimized the quantity of training data that we employed.

3.4.1 Model 1: Multinomial Logistic Regression:

Multinomial logistic regression is our initial model, and to enhance the model's performance on test data that hasn't been seen before, we'll be adjusting the regularization value. Since our goal involves classifying penguins into the appropriate species, logistic regression is mostly utilized to solve classification problems. The probabilities of several categories—for instance, the likelihood that the penguin belongs to the Adelie species—are modelled using multiple regression functions. The model does better at applying what it has learnt from the training set of data to fresh test sets when regularization is used. The model is kept from overfitting by this procedure, which stops it from focusing too heavily on each anomaly in the training data for our penguins. Eliminating overfitting and underfitting in our data is our aim while determining the optimal regularization parameter.

```
#Print the best c parameter for the model and its cv score
print("best_c=",np.round(best_c,3)," Cross validation score=",best_score)

best_c= 1.0 Cross validation score= 0.997527863906
```

Fig. 7. Cross validation performance

It's important to remember that 1.0 is the ideal c parameter by default in logistic regression. The algorithm is predicted to correctly categorize 99.7% as shown in Fig 7, of recently discovered penguins into their respective species, with a CV score of 0.997.

Utilizing the following, our training data will be fitted to the logistic regression model with our optimal c parameter. The model is then tested using the test data to see if it still performs well even with the hidden data. Notably different test and train results indicate that the model may be overfit.

3.4.1.1 Data Preparation:

The dataset was divided into two sets: 70% for training and 30% for testing, in order to provide a representative sample for model evaluation. Numbers were used to encode string labels for the Adelie, Gentoo, and Chinstrap penguin species. The Palmer Penguin dataset's disparate scales were reduced by separating the feature and target variables and standardizing the data using StandardScaler.

3.4.1.2 Model Construction:

The model used for classification was a Multinomial Logistic Regression. The regularization parameter values were used to train the model in a range of increments, from 0.1 to 1. Finding the best-performing c value involved computing cross-validation scores for each c parameter.

```
LR1.score(X_test,y_test)

0.99
```

Fig. 8. Testing model accuracy-LR

3.4.1.3 Evaluation and Analysis:

To verify generalizability, the trained model was assessed using test data. Examining overfitting, the model showed strong performance, hitting 99% of the test data with accuracy. The model's accuracy was enhanced by feature selection and hyperparameter tweaking carried out within the logistic regression framework.

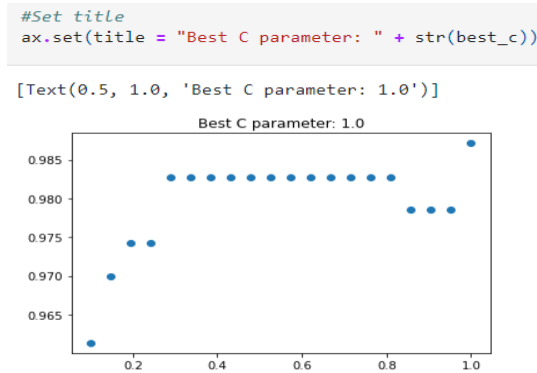


Fig. 9. Validation trends across different depth levels

With a test score of 0.99 for the variable combinations “Sex,” “Culmen Length (mm),” and “Culmen Depth (mm),” it can be concluded that 99% of the penguin test data had the species accurately predicted by the Logistic Regression model with the best c parameter ($c=1.0$). There is no overfitting, as evidenced by the test score’s 0.997 difference from our training CV score.

3.4.2 Model 2: Support Vector Model (SVM):

The support vector machine is the next model we’ll be using. Finding the best hyperplane to differentiate between our data points is the aim of this model. As we are examining three penguin attributes in this instance—Sex, Culmen Length, and Culmen Depth—the support vector machine will identify the most effective 3-dimensional hyperplane that can accurately categorize the data points. The ideal hyperplane will be a two-dimensional plane because the model operates in three dimensions. The hyperplane that generates the greatest distance between the three features’ data points is the most ideal one.

We use the sklearn module’s support vector machine model. We cross-validate the optimal gamma for our support vector machine. To put it simply, non-linear hyperplanes are treated with the gamma hyperparameter. The model attempts to fit the training data set more closely the higher the gamma value. Consequently, overfitting of the training data due to an excessively high gamma value results in subpar performance on the test data. We log 100 gamma values in our code, which span from 0.05 to 5. We also log the cross-validation scores that result from these gamma values.

3.4.2.1 Data Preparation:

The dataset was divided, string labels were encoded, and data were normalized using the StandardScaler, same like in Model 1.

3.4.2.2 Model Construction:

The training data was used to build a Support Vector Machine (SVM) model, which was used for classification. According to the research, the SVM classified the training data perfectly, however on the test data, its accuracy slightly decreased to 98%. It was highlighted that the SVM performs better in higher-dimensional spaces when the dimensionality of the feature space was examined.

3.4.2.3 Exploration of higher dimensions:

It was proposed that adding more features may improve the SVM’s efficiency, as long as overfitting was carefully taken into account. It was highlighted how crucial it is to choose the right kernel function for support vector machines (SVMs) in light of how it affects hyperplane forms and the possibility of overfitting.

```
best_gamma, best_score
(0.2, 0.979546252746)
```

Fig. 10. Optimal gamma and score determination

Given that it provided the best cross-validation score of 0.979 as shown in Fig 10, our cross-validation indicates that the optimal gamma value is 0.2. After that, we plot the gamma values against the cross-validation scores in Fig 11.

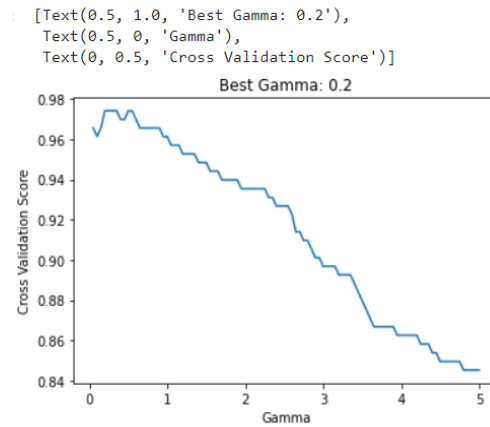


Fig. 11. Cross-validation scores with diverse gamma values.

As we can see from the image, the remaining cross-validation scores deteriorate with increasing gamma values, starting at 0.2, where the best cross-validation score peaked. This pattern supports the previous explanation regarding how overfitting of the data is caused by a gamma value that is too high.

Our support vector machine model is now created with the optimal gamma value of 0.2 following the cross-validation process. We then use our training data to fit and rank the model.

```
SVM = svm.SVC(gamma = best_gamma)
SVM.fit(X_train, y_train)
SVM.score(X_train, y_train)
1.0
```

Fig. 12. SVM parameter exploration

The support vector machine model performed flawlessly in categorizing the training data, earning a score of 1.0 as shown in Fig 12. Our support vector machine model is now assessed using our hidden testing data.

```
SVM.score(X_test, y_test)
0.98
```

Fig. 13. Testing model accuracy-SVM

With a score of 0.98 as shown in Fig.13, the support vector machine model accurately predicted the species of 98% of the penguin test data.

4. PROPOSED METHODOLOGY:

Based on important traits like sex, culmen depth, and culmen length, this study supports the application and improvement of the Multinomial Logistic Regression model for classifying penguin species. The accuracy of the model is 99.7% on training data and slightly lower at 98% on unknown data, indicating its good performance. Concerns about overfitting are minimized by using feature selection within the logistic regression framework. We advise expanding the feature set with biologically significant variables and utilizing cross-validation to fine-tune hyperparameters in order to enhance the model. While thorough documentation, which includes feature reasoning and model interpretation, ensures

openness and repeatability, a separate validation set guarantees generalizability. A trustworthy and efficient approach for precise penguin species classification involves being aware of the dangers of overfitting and being dedicated to thorough reporting.

5. CONCLUSION AND FUTURE WORK:

As a result, this study shows that the Multinomial Logistic Regression model can effectively and accurately identify between different species of penguins using factors including sex, culmen length, and depth in both training and test datasets. In order to support Multinomial Logistic Regression, and Support Vector Machine models, we suggest looking into an 80%/20% data split, increasing cross-validation folds, and doing deep hyperparameter tweaking beyond `c`, `max_depth`, and `gamma`. Furthermore, adding more precise location information can enhance the qualitative aspects of species identification. These modifications are intended to improve model robustness and accuracy, offering a more sophisticated comprehension of penguin habitats and advancing the ongoing development of classification models for a variety of species.

6. References:

- [1] Dasgupta, A. Banerjee, A. Ghosh Dastidar, A. Barman, and S. Chakraborty, "A Study and Analysis of a Feature Subset Selection Technique using Penguin Search Optimization Algorithm (FS-PeSOA)," Information Technology Department, TechnoIndia, Saltlake, Kolkata, India, Jul. 2019.
- [2] Hua, A., & Goldsztein, G. "Using Machine Learning to Predict Penguin Species." Acton-Boxborough Regional High School, USA, Advisor, vol. 11, no. 4, 2022.
- [3] Y. Liu, V. Shah, A. Borowicz, M. Wethington, N. Strycker, S. Forrest, H. Lynch, and H. Singh, "Towards Efficient Machine Learning Methods for Penguin Counting in Unmanned Aerial System Imagery," Field Robotics Lab, Northeastern University, Boston, USA, Sep. 2020.
- [4] S. Chessa, A. Micheli, R. Pucci, J. Hunter, G. Carroll, and R. Harcourt, "A Comparative Analysis of SVM and IDNN for Identifying Penguin Activities," Department of Computer Science, University of Pisa, Pisa Area, Italy; The University of Queensland, School of ITEE, Brisbane, Australia; Department of Biological Sciences, Macquarie University, Sydney, Australia, 2017.
- [5] N. Bidi and Z. Elberrichi, "Using Penguins Search Optimization Algorithm for Best Features Selection for Biomedical Data Classification," Djillali Liabès University, Sidi Bel Abbès, Algeria, vol. 7, no. 4, pp. October-December, 2017.
- [6] S. K. Baliarsingh and S. Vipsita, "Chaotic emperor penguin optimised extreme learning machine for microarray cancer classification," DST-FIST Bioinformatics Lab, Department of Computer Science and Engineering, International Institute of Information Technology, Bhubaneswar, India, vol. 13, no. 7, pp. e-first, Nov. 2019, doi: 10.1049/iet-syb.2019.0028.
- [7] J. Forcada and P. N. Trathan, "Penguin responses to climate change in the Southern Ocean," Published: 05 June 2009, doi: 10.1111/j.1365-2486.2009.01909.x.
- [8] H. Le, B. Gonçalves, D. Samaras, and H. Lynch, "Weakly Labeling the Antarctic: The Penguin Colony Case," May 8, 2019.
- [9] H. J. Lynch, R. White, A. D. Black, and R. Naveen, "Detection, differentiation, and abundance estimation of penguin species by high-resolution satellite imagery," *Polar Biology*, vol. 35, no. 6, June 2012.