Mapping Extinction:

A Spatial and Text Analysis of Endangered Species

Satomi Ito, Samyukta Vakkalanka, Jared Zirkes

Georgetown Data Science and Analytics

DSAN 5300

May 1, 2025

Abstract:

The global extinction crisis, driven by habitat loss, climate change, and human interference, poses a significant threat both now and in the years to come. This study aims to classify the extinction risk of animal species based on geographic, ecological, and unstructured text-based features. With the IUCN Red List, we constructed a series of predictive models to classify animals into IUCN threat classifications. We evaluated three classification models, a Random Forest, a Support Vector Machine (SVM), and a Neural Network, achieving the highest test accuracy with the Random Forest (87%). Further analysis revealed that human impact, species body mass, and geographic distribution are key predictors of extinction risk. In addition, we show that the neural network underperformed due to a strong class imbalance. Our findings highlight the utility of combining spatial, ecologic, and textual data to best understand and predict species vulnerability, informing conservation strategies in the face of growing pressures.

1. Introduction

The ongoing extinction crisis poses one of the most significant threats to global biodiversity. Driven by a complex web of factors, species are disappearing at rates far beyond natural background levels. According to the International Union for Conservation of Nature (IUCN), more than 28% of assessed species are currently threatened with extinction. As human activity continues to reshape the planet, understanding which species are most vulnerable, and why, is critical for designing effective conservation strategies.

Species extinction not only reduces ecological resilience but also affects human well-being, from destabilizing food webs to diminishing ecosystem services like pollution and climate regulation. Early identification of endangered species can enable targeted conservation interventions, but predicting risk is complicated by the diversity of species traits, regional environmental pressures, and data availability.

We aim to build predictive models that classify the level of endangerment for animal species, leveraging structured ecological data, spatial attributes, and unstructured text features. By integrating geographic, environmental, and semantic information, we seek not only to maximize predictive accuracy but also to gain deeper insights into the patterns and drivers of extinction risk.

1.1 Literature review

The extinction of species is occurring at an unprecedented rate, driven by an array of human-induced pressures including habitat destruction, climate change, overexploitation, pollution, and invasive species. Recent work describes this phenomenon as a “biological annihilation” with vertebrate population declines signaling the onset of a sixth mass extinction event (Ceballos et al., 2017). Biodiversity loss is not limited to outright extinctions but also includes widespread reductions in population sizes, which has cascading effects on ecosystems.

Understanding which species are most vulnerable is a core challenge in conservation science. Traditional approaches, such as expert-based risk assessments, are effective but time-consuming and limited by incomplete data. Machine learning models offer an alternative, with studies demonstrating that ecological and geographic traits can serve as strong predictors of extinction risk, even when detailed field data are missing (Bland et al., 2015).

Beyond species-species traits, recent research emphasizes the functional consequences of biodiversity loss. The extinction of threatened vertebrates will lead to highly idiosyncratic changes in functional diversity across ecosystems, with unpredictable impacts on ecosystem functioning (Toussaint et al., 2021). These findings highlight that the erosion of ecological roles may have even more profound and destabilizing effects than the loss of species numbers alone.

In terms of modeling methodology, prior comparative studies have found that parametric models like logistic regression offer transparency and interpretability, while nonparametric models, such as random forest and support vector machines (SVM), often yield higher predictive performance but at the cost of explainability (Bland et al., 2015). However, no model type consistently outperforms others across all conservation tasks; model choice largely depends on the dataset structure, quality, and the intended application of the predictions.

Building on these bodies of work, this project seems to predict species endangered by combining spatial attributes, ecological traits, and text-derived features. By systematically comparing multiple modeling approaches, we aim to enhance predictive accuracy while also uncovering meaningful patterns that contribute to extinction risk.

1.2 Research Questions

Regarding previous research on extinction risk factors, our project focuses on several key research questions.

- What are the biggest factors causing the endangerment of different animal species globally?

We aim to identify which ecological, geographic, and anthropogenic factors most strongly predict species' risk of endangerment.

- What groups of species are more likely to be endangered?

By analyzing taxonomic patterns, we aim to determine whether certain groups of animals are disproportionately affected by extinction threats.

- How does human development affect species' population security?

Our goal is to quantify the extent to which human development correlates with higher endangerment likelihood.

- How can predictive models support efforts to mitigate harm to wildlife?

Beyond building classification models, we aim to explore how machine learning outputs, such as feature importance, can inform real-world conservation strategies.

Through these research questions, our project seeks not only to enhance predictive accuracy but also to provide actionable insights into features driving extinction rates.

2. Data

2.1 Data Munging

Data for this project comes from two major sources. The first is the International Union for the Conservation of Nature (IUCN). The IUCN, which also labels species with their level of endangerment as explained below, also provides a dataset with general information about the species requested (*The IUCN Red List of Threatened Species*, n.d.). The majority of these features were either already numeric (such as latitude and longitude data) and required no cleaning, or categorical texts (like species family), and simply required one hot encoding. However, there was one column utilized that originates as a textual description of the threats facing the species. In order to convert the feature into a numeric, and therefore usable, feature, we fed the column through a BART-based model to calculate a sentiment. We set up a zero-shot classifier using the BART-MNLI model which categorizes a text as “optimistic”, “neutral”, or “pessimistic”. We then calculated the negativity score to be a classification of threat severity by computing the pessimistic minus the optimistic score. This feature was then included in the model.

The other major data source was a general dataset of animal species features from PanTHERIA1.0 (Jones, Kate E., et al). This dataset required only minimal munging as all the features used beyond the taxonomy that we used for merging data are numeric. Data used from this dataset included information such as animal size, litter size, and habitat + diet breadth. The final consideration for this dataset was that there were a number of animals and/or features with missing data, which were dropped from the dataset before further analysis.

2.2 Data Exploration

To understand the distribution of species characteristics and extinction risk factors, we conducted an exploratory data analysis across both structured and unstructured features. This section summarizes the key findings from spatial, categorical, and threat-based perspectives.

2.2.1 Threat Assessment and IUCN Classification Criteria

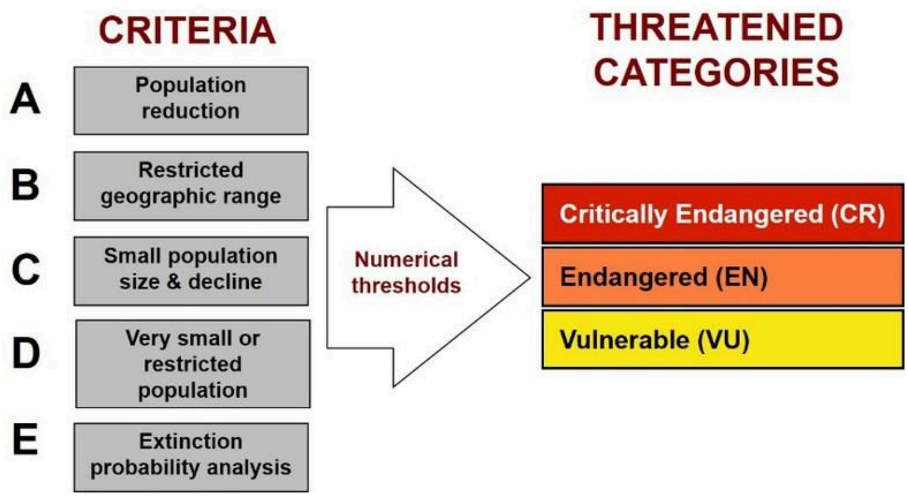


Figure 1: Threat categories assigned based on five key criteria and numeric thresholds

Before exploring the distribution of endangered species, it is important to understand how extinction risk is assessed. Figure 1 categorizes species based on five major criteria: population reduction, restricted geographic range, small population size and decline, very small or restricted populations, and extinction probability analysis. These criteria establish numerical thresholds that determine whether a species is classified as “Vulnerable (VU)”, “Endangered (EN)” or “Critically Endangered (CR)”.

2.2.2 Geographic Patterns of Endangered Species

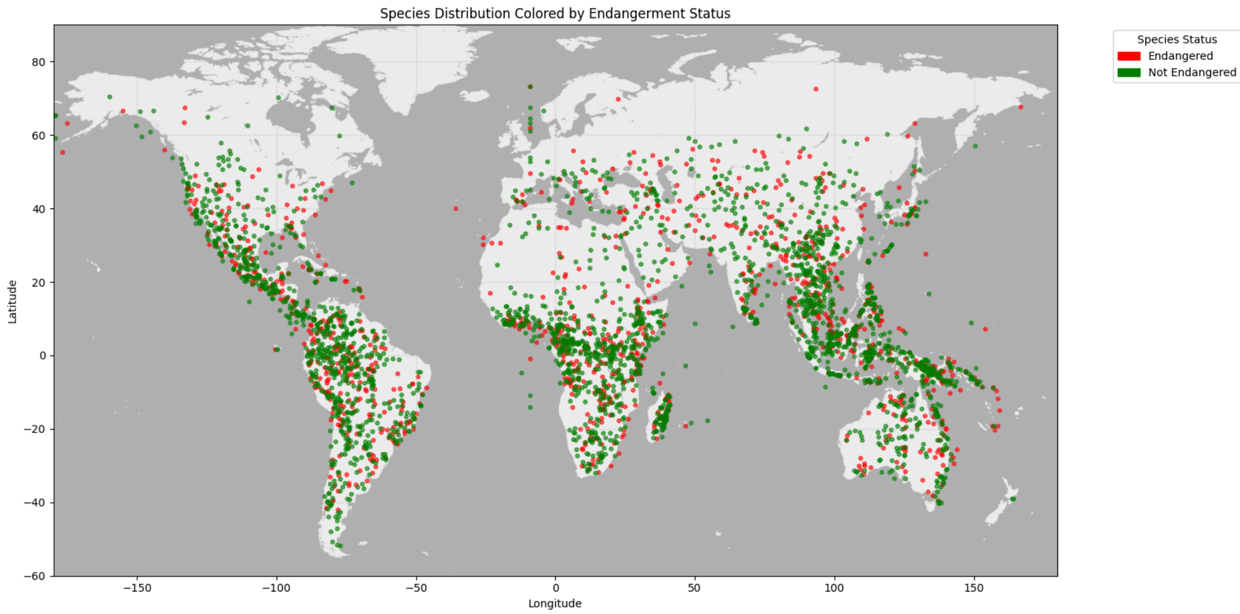


Figure 2: Global distribution of endangered (red) and non-endangered (green) species.

Figure 2 is a visualization of species locations, colored by endangerment status, which reveals distinct spatial patterns. Endangered species (red) appear highly concentrated in tropical and subtropical regions, particularly Central and South America, Central Africa Southeast Asia, and parts of Oceania. In contrast, species classified as non-endangered (green) are more distributed across regions. The map highlights that biodiversity hotspots coincide with areas of elevated extinction risk. These findings align with previous studies, which emphasize that tropical biomes face disproportionate conservation threats due to habitat loss, agricultural expansion, and human population graph (Ceballos et al., 2017; Toussain et al., 2021).

2.2.3 Endangerment by Biogeographic Realm

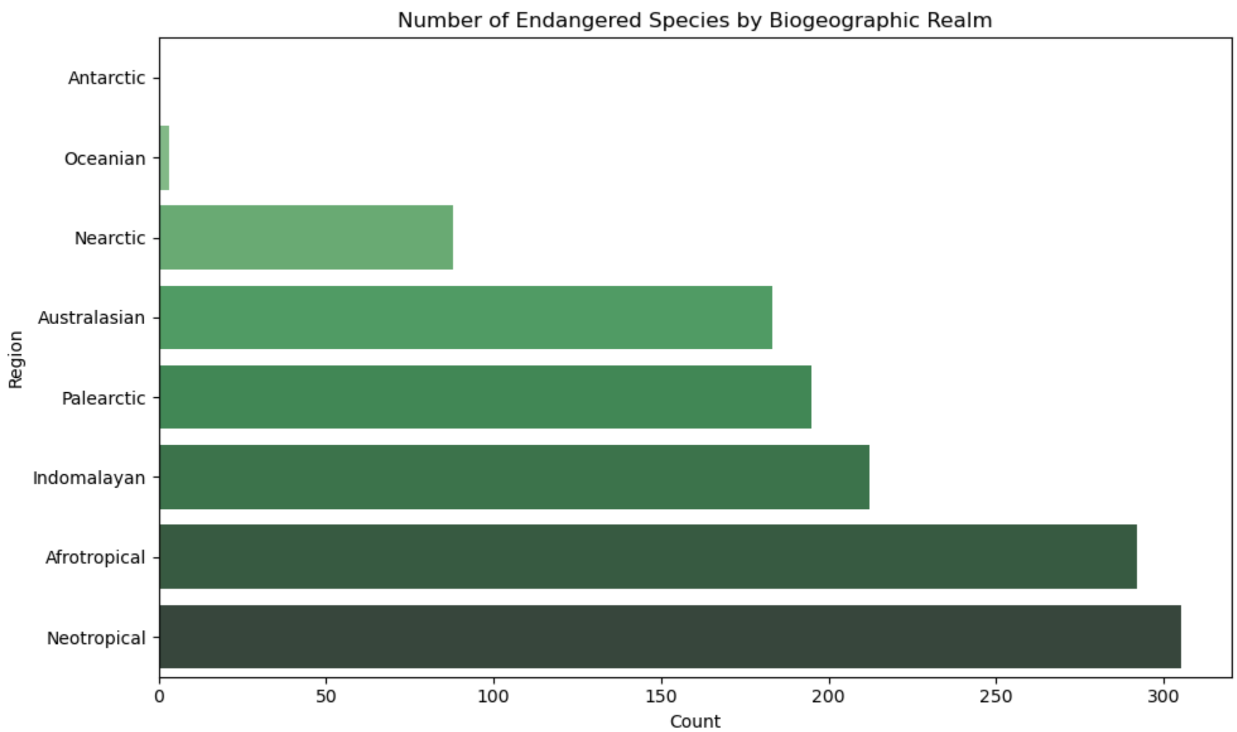


Figure 3: Number of endangered species by biogeographic realm.

Figure 3 summarizing the number of endangered species across biogeographic realms further supports the spatial analysis. The Neotropical and Afrotropical regions contain the highest number of endangered species, followed closely by the Indomalayan realm. Meanwhile, the Nearctic and Oceanian regions exhibit relatively fewer endangered species, and the Antarctic realm has near-zero representation. These patterns reflect differences in habitat diversity, human pressure, and conservation capacity across global regions.

2.2.3 Taxonomic Breakdown of Risk Categories

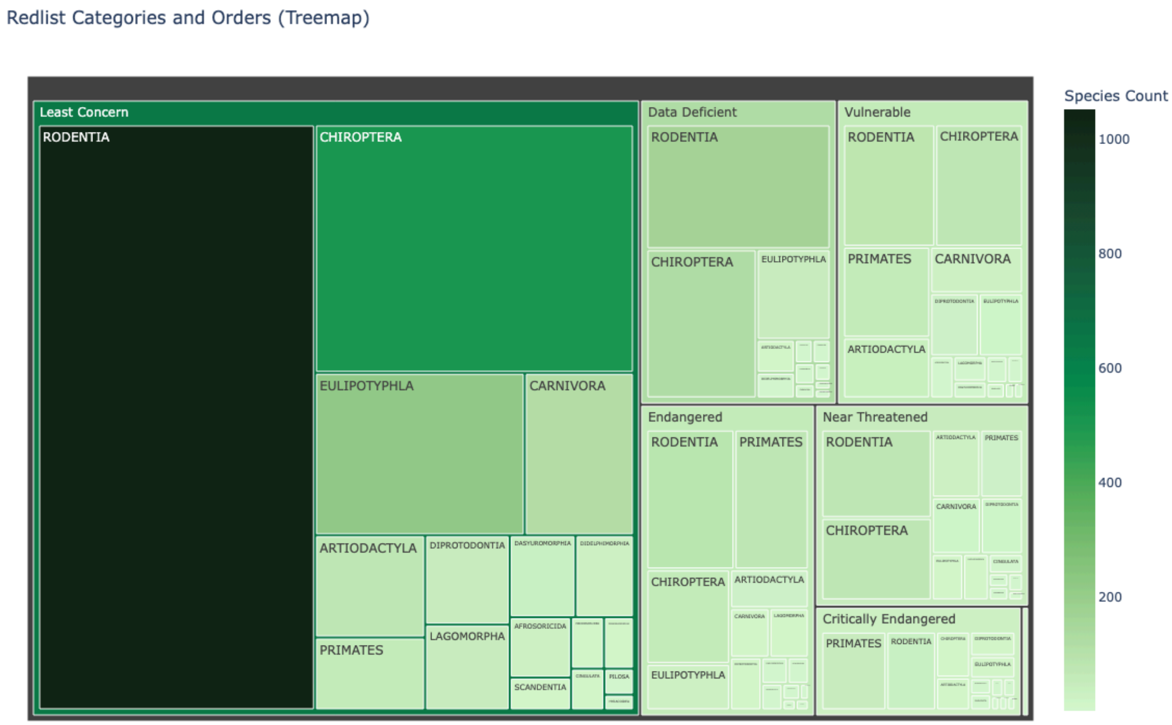


Figure 4: Treemap showing species count.

Figure 4 visualization illustrates how extinction risk varies across major taxonomic orders and IUCN Red List categories. Species classified as “Least Concern” are primarily concentrated among rodents (Rodentia) and bats (Chiroptera). In contrast, orders such as Primates and Artiodactyla exhibit a greater proportion of species in higher-risk categories such as “Endangered” and “Critically Endangered”. This taxonomic disparity highlights the vulnerability of certain groups, particularly large-bodied mammals and habitat specialists, to extinction pressures. This observation is consistent with previous findings that larger species with narrow ecological niches are more prone to decline (Bland et al., 2015).

3. Methods

To investigate the factors influencing the endangered status of species globally, we employed a comparative modeling study using three different approaches: support vector machines, random forests, and a neural network. Each model was trained on a range of biological and ecological features to predict species’ IUCN Red List status as described above.

3.1 Support Vector Machines (SVMs)

SVMs are a supervised learning method. They construct a hyperplane that can be used to separate the data points from different classes. We used a radial basis function (RBF) kernel to allow for non-linear decision boundaries, which we expected would better capture the complex relationships between species traits, environmental features, and endangered status.

They work well in classification tasks that have high-dimensional data. The mixture of biological, ecological, and conservation-based features in our dataset made SVMs a good choice to balance flexibility and generalizability. Additionally, SVMs with RBF kernels are robust to overlapping classes, which we noticed in this dataset while exploring the data. Because of this, this model can predict the species' statuses where the category boundaries may not be so clear.

3.1.1 Implementation Details

Prior to training, input features were standardized to have zero mean and unit variance because SVMs are sensitive to feature scaling. We performed hyperparameter optimization using a grid search over the penalty parameter C and the RBF kernel parameter γ using a fixed RBF kernel and class weights balanced to account for class imbalance. Model selection was based on the best mean F1 score from five-fold cross-validation. After identifying the best hyperparameters, we predicted probability estimates on the validation set and further fine-tuned the classification threshold. Rather than using the default 0.5 threshold for positive classification, we evaluated multiple thresholds between 0.2 and 0.8 and selected the one that maximized the F1 score on the test set. The final evaluation metrics were computed using this optimized threshold.

3.2. Random Forests

Random Forest is an ensemble learning method that constructs a group of decision trees during training and outputs the mode of their predictions for classification tasks. By putting together multiple trees, random forests aim to reduce the variance of the predictions and improve generalization, helping them avoid overfitting. They can also handle mixed data types and missing values, which makes them a good fit for this biological dataset.

We chose this model because of its ability to handle the large number of features that we had about the various species without requiring much feature engineering. It is built to handle noise and model complex, non-linear relationships. This is important given the diverse attributes in our dataset. Additionally, random forests provide feature importance measures which added interpretability into which traits most strongly influence endangered status.

3.2.1 Implementation Details

We conducted hyperparameter optimization using grid search with five-fold cross-validation. The hyperparameters tuned included the number of trees, the maximum depth of each tree, and the maximum features considered at each split. Model selection was based on maximizing the mean F1 score across folds. After identifying the best model, predictions were made on the test set, and final evaluation metrics were reported with additional threshold adjustments.

3.3 Neural Network

Simple feed-forward Neural Networks (ANNs) are a powerful prediction tool that stacks individual perceptrons into a network of 1 or more layer connections, before introducing a non-linear ‘activation’ function in order to inject non-linearities into the network. This complexity allows ANNs to excel at capturing both complex and non-linear relationships within data, which is why we opted to utilize one for this project.

3.3.1 Implementation Details

To train our neural network, we used the same cleaned dataset as outlined in sections 3.1 and 3.2. Additionally, in determining the optimal model to fit the dataset, we conducted hyperparameter tuning around both the number of hidden layers in the network and the number of nodes in each hidden layer. After identifying the optimal network based on the minimized cross-entropy, predictions were made on the test set, with final evaluation results available in the results section.

4. Results

To evaluate the ability of different models to predict the endangered status of species, we assessed the Support Vector Machine (SVM) and Random Forest classifiers using both cross-validation and independent test sets. This section presents the performance metrics, confusion matrices, and model insights for each approach, highlighting their strengths and weaknesses in handling the classification task.

4.1 SVM

Using an SVM with an RBF kernel, the best model achieved an F1 score of 0.860 during five-fold cross-validation. After additional threshold tuning, the best decision threshold was determined to be 0.43, improving the F1 score on the test set to 0.889. This step helped mitigate the effects of class imbalance. The final accuracy of the model on the test data was approximately 85%. Key performance metrics are summarized in Table 1.

Table 1: Performance Metrics for optimized SVM Model

Label	Precision	Recall	F1-Score
Not Endangered	0.77	0.72	0.74
Endangered	0.87	0.90	0.89

The confusion matrix is shown in Figure 5. Overall, the model correctly identified the majority of endangered species, but exhibited some false negatives, particularly for non-endangered cases.

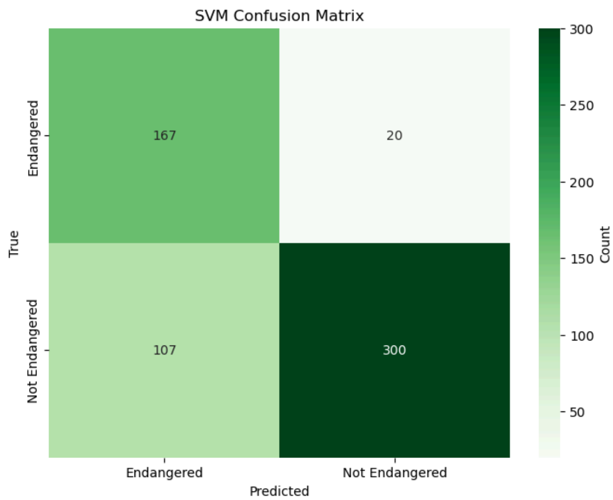


Figure 5: Confusion matrix for optimized SVM model

4.2 Random Forest

The random forest classifier that was tuned using a grid search achieved a best cross-validation F1 score of 0.906. The best hyperparameters were a max tree depth of 20, a max feature of log2, and 100 estimators. On the test set, the optimized model achieved an accuracy of approximately 87%. Key performance metrics are summarized in Table 2.

Table 2: Performance Metrics for optimized Random Forest Model

Label	Precision	Recall	F1-Score
Not Endangered	0.82	0.75	0.78
Endangered	0.89	0.93	0.91

The confusion matrix for the random forest model is presented in Figure 6. Similar to the SVM model, the random forest tended to predict endangered species more accurately. However, it produced a few more false positives than false negatives.

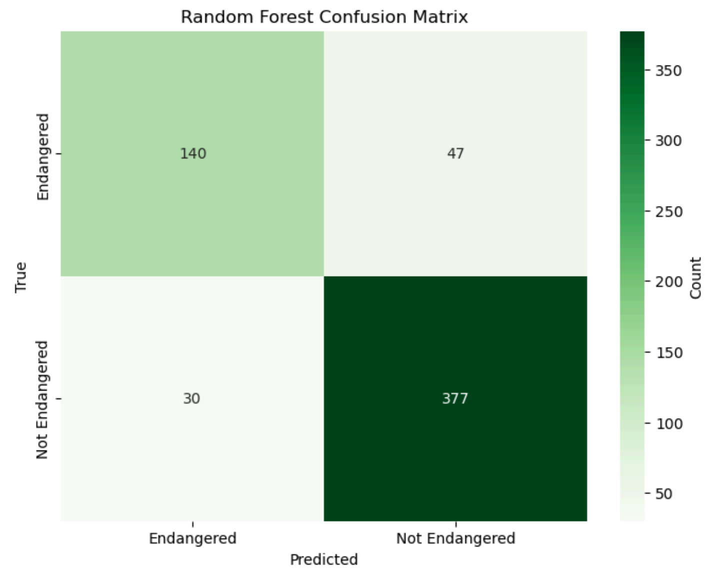


Figure 6: Confusion matrix for optimized Random Forest model

4.2.1 Feature Importance

Figure 7 shows the feature importance scores calculated from the trained Random Forest model.

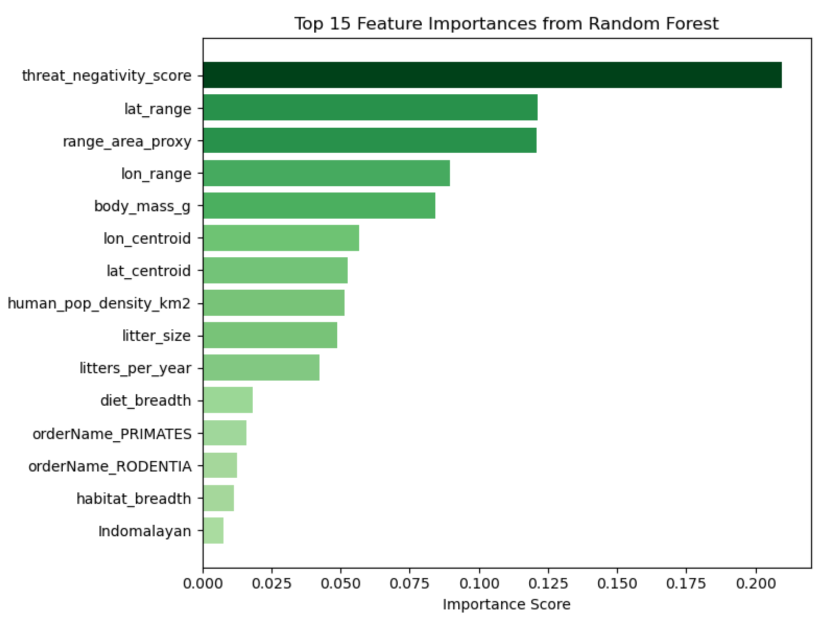


Figure 7: Feature importances from optimized random forest model

The top features include a combination of the details about the amount of space the species takes up in its habitat, physical features about the animals such as their body mass or litter size, and where in the world they are located as well as the human population density in these areas. The most important feature was the threat negativity score that we derived from the threat text column in the original dataset.

4.3 Neural Network

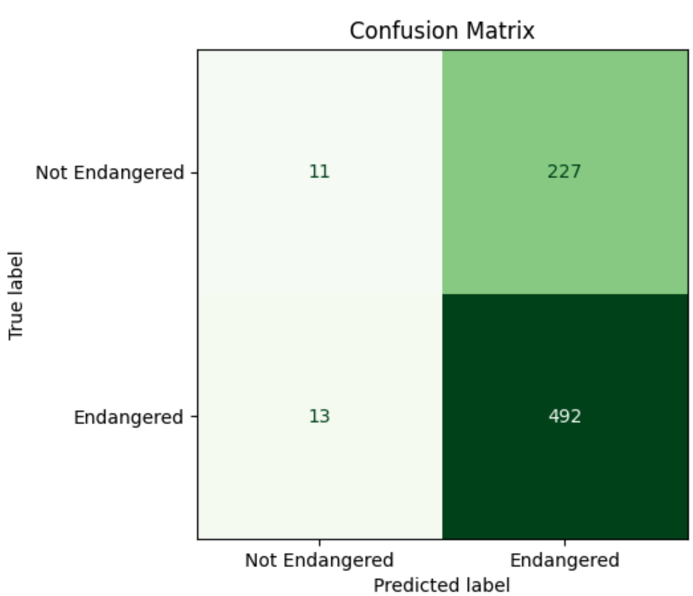
To make final testing predictions, we used a network with two hidden layers, with 128 and 64 nodes respectively, and ReLU activation functions across the whole network. In the validation stage, this model achieved an accuracy of 68.8% With this model trained, predictions were next made on the hold-out test dataset, with final evaluation metrics below:

Table 3: Performance Metrics for optimized Neural Network Model

Accuracy: 67%

Label	Precision	Recall	F1-Score
Not Endangered	0.68	0.97	0.80
Endangered	0.46	0.05	0.08

The confusion matrix for the network is also shown below:



It can be seen that the neural network model fails to predict endangerment as well as either of the previous random forest or SVM models. In particular, this is due to the fact that the model fails to overcome the significant class imbalance.

5. Discussion

This section discusses the implications of the results found above. Additionally, we compared the three models applied to the data and determined which one performed the best to predict the endangered status of the species in the dataset.

5.1 SVM

Our SVM model with an RBF kernel demonstrated strong predictive capability for species endangerment status, achieving an F1 score of 0.889 after threshold optimization. This performance suggests that SVMs can effectively capture the complex non-linear relationships between the biological and environmental features and conservation status. The threshold adjustment from the standard 0.5 to 0.43 improved performance, indicating that the class imbalance had a noticeable impact on this model. The threshold optimization is particularly important when working with conservation data where endangered species are typically less represented in datasets.

The model exhibited better performance in identifying non-endangered species with a recall of 0.90 compared to endangered species with a recall of 0.72. The difference in performance might be a result of the class imbalance in our dataset, where there were fewer endangered species. While our use of balanced class weights and threshold tuning helped address some of these effects, the model still showed a bias towards the majority class. In conservation applications, this type of bias needs to be addressed because failing to identify an endangered species might have more serious implications than false positives.

5.1.1 Limitations

The SVM model's training process involved hyperparameter tuning across C and gamma while maintaining the RBF kernel. While this was computationally intensive, it was important for maximizing model performance. The final model achieved a good balance between generalizability and overfitting as seen by the strong test set performance. A limitation of this approach is the decreased interpretability of the model. Since we used an RBF kernel, it is harder for there to be an intuitive interpretation of why certain species are classified as endangered. Also, SVMs don't scale as well to larger datasets so expanding this analysis to include more species and features would not be feasible.

5.2 Random Forest

The random forest model demonstrated excellent predictive capability with an F1 score of 0.91 for endangered species, slightly outperforming the SVM approach. With an overall accuracy of 87%, this method effectively captured the complex relationships between species characteristics and endangerment status. The optimal hyperparameters of a max tree depth of 20, log2 feature selection, and 100 estimators suggest that the model did not need to be overly complex to learn the patterns in the dataset without overfitting.

Similar to the SVM model, the random forest demonstrated stronger performance in correctly identifying the not endangered species with a recall of 0.93 compared to the endangered species with a recall of 0.75. However, this model did better in correctly identifying both classes, indicating that the random forest was particularly successful at capturing the features that are signals of endangerment. The confusion matrix shows that there is a higher false negative rate in this model compared with the SVM. 47 of the endangered species were classified to be not endangered. This increase in false negatives is more harmful to animals because species that need the protection might get misclassified as not endangered and won't get the attention of conservationists that they need to recuperate their populations.

5.2.1 Feature Importance Analysis

An advantage of the random forest model is its built-in feature importance metrics. This provides insights into the factors that are strongly associated with species endangerment. The threat negativity score was the most influential predictor, confirming that the text column of threat descriptions had substantial predictive information. This finding validates our novel approach of applying NLP techniques to IUCN textual data.

Geographic range indicators (lat_range, range_area_proxy, lon_range) were some of the most important predictors. This supports the idea that species with restricted ranges are more vulnerable to extinction (Chichorro et al.). The importance of body mass also appeared. Some further data exploration showed that a higher proportion of species with larger body mass were categorized as endangered. This could possibly be due to slower reproduction rates, greater resource requirements, and vulnerability to human pressures like hunting (Broom).

The significance of location-based features (lon_centroid, lat_centroid) and human population density highlights the human impact on endangerment, suggesting that geographic location and human proximity are key determinants of conservation status. Reproductive characteristics (litter size, litters per year) were also important predictors. This makes sense because species with lower reproductive rates are likely to be more vulnerable to population declines. Taxonomic indicators (Primates and Rodentia) demonstrated that certain taxonomic groups have distinctive endangerment patterns. Finally, the importance of diet breadth and habitat breadth could be related to the fact that species with more narrow, niche requirements might face greater extinction risks than generalists.

5.2.2 Limitations

The random forest model's ability to handle mixed data types was advantageous for our dataset which included continuous measurements, categorical features, and derived metrics.

Additionally, the model's robustness to outliers and ability to capture non-linear relationships made it well suited for this data containing complex interactions and non-normal distributions.

One limitation is that while the model provides feature importance scores, it does not directly model interaction effects like a parametric model would or provide easily interpretable decision rules like a simple decision tree. The relative features may also be influenced by correlations between predictors, such as body mass being correlated with reproductive rates. This may affect the interpretation of each feature's independent contribution.

5.3 Neural Network

The final neural network model fails to predict endangerment with the same capabilities as the other models presented. Accuracy is only marginally above the class-split random guesser, and classification metrics such as precision, recall, and F1 are either on par or significantly below those of the random forest and SVM. A closer examination of the confusion matrix quickly shows that this 'learning' failure is due to the class imbalance in the dataset, with the model predicting 'not endangered' for nearly 97% of all data points, and does not even carry a significant precision for the few endangered predictions.

As a result, the model does show a very elevated recall for the not endangered label. However, thinking critically about the context around this, this elevated recall is not very beneficial in a true conservation setting. With the endangered species being the animals in need of conservation, simply classifying every species as not endangered to elevate non-endangered precision involves lumping all the endangered species into this category, leaving them behind in any implemented conservation efforts.

5.3.1 Limitations

As described above, the final neural network model fails to overcome the class imbalance in the dataset. This is a common issue with neural networks, and so this result was not unexpected in our findings. By minimizing the cross entropy, the networks can tend to ‘optimize’ by pushing the majority or vast majority of predictions into the majority class. On the contrary, tree-based methods such as random forests, are often able to a better extent overcome these imbalances given sufficient data. Given that we see our random forest model perform significantly better than the neural network, we move forward in considering it a final model over the neural network.

5.4 Comparison of Models

Table 4 illustrates the comparison of the accuracy between the 3 models on the test data. Overall, the random forest model had the highest accuracy. This, in combination with its interpretability, generalizability, and relatively low false negative rate, makes it the best approach to modeling the endangerment status of a species given its physical and geographical features and its interactions with humans.

Table 4: Comparison of Accuracy across Three Models on Test Data

Model	Random Forest	SVM	Neural Network
Accuracy	0.87	0.84	0.68

6. Conclusion

Our study demonstrates the effectiveness of machine learning approaches in predicting species endangerment status using biological traits, geographical information, and human impact metrics. By comparing Support Vector Machines, Random Forests, and Neural Networks, we've identified that Random Forests offer the most accurate and interpretable framework for conservation risk assessment with an accuracy of 87% on test data.

6.1 Key Findings and Ecological Interpretation

The most influential predictors of endangerment status include threat negativity scores calculated from IUCN threat descriptions, geographic range indicators, body mass, and reproductive characteristics. These findings are in line with existing ecological theories of extinction vulnerability. Species with restricted geographic ranges face greater risk due to their limited habitat availability and increased vulnerability to natural disasters and other localized habitat destruction (Chichorro et al.). Similarly, species with larger body masses typically have smaller reproductive rates and longer generation times, making population recovery more challenging when faced with external pressures (Broom).

Our novel application of natural language processing to extract sentiment from threat descriptions proved particularly valuable, emerging as the strongest predictor in our models. This shows the complex nature of endangerment and suggests that descriptions written by conservation experts contain predictive power that can be quantified and incorporated into models through computational approaches.

6.2 Conservation Implications

As human development continues to expand into natural habitats, the likelihood of extinction for many species increases. Our findings regarding the importance of geographic range metrics and human population density highlight this relationship. The loss of habitat, combined with mounting pressures from pollution, climate change, agriculture, and invasive species, creates a compounding effect that places vulnerable animals at greater risk (Bull et al.).

Machine learning models like those developed in this study offer powerful tools for proactive conservation. Rather than waiting until species reach critically endangered status, these predictive frameworks could identify at-risk species before they go extinct. By analyzing patterns across biological traits and environmental conditions, conservation agencies can allocate resources more efficiently to species that are showing vulnerability indicators before their situations go too far.

To address these challenges effectively, we must take stronger steps to protect wildlife, reduce existing threats, and restore ecosystems where possible. Policymakers and corporations should be held accountable, and there is a critical need for regulations that limit harmful land development. Supporting sustainable and mixed-use development can help balance human needs with the urgent responsibility of biodiversity conservation (Kirk et al.).

6.2 Methodological Contributions and Limitations

Our comparative methodology highlights the value of evaluating multiple machine learning approaches when addressing conservation questions. While Neural Networks underperformed in our study, both SVMs and Random Forests demonstrated strong predictive power. The superior interpretability of Random Forests, coupled with their high accuracy, suggests they may be particularly valuable tools for conservation scientists who need information beyond predictions such as insights into the factors that are most influential when it comes to endangerment status.

The inclusion of natural language processing to quantify threat descriptions represents a novel methodological contribution that combines qualitative expert knowledge with quantitative prediction. However, our approach has limitations. The models cannot predict causality, only correlation, and their performance may change across different taxonomic groups or geographic regions. Additionally, data availability and quality remain persistent challenges in conservation science, with endangered species often being the least studied due to their rarity.

6.3 Future Directions

Future research should focus on expanding these models to include additional taxonomic groups, environmental variables, and threat metrics. Incorporating data collected over time could allow us to study how endangerment risk changes in response to conservation efforts or environmental shifts like climate change. Additionally, exploring the interaction effects between features could provide deeper insights into how various risk factors combine to influence extinction vulnerability. For example, understanding how body size interacts with habitat specialization or geographic range could show more specific patterns of endangerment risk.

As biodiversity loss increases globally, the need for evidence-based conservation grows more urgent. Our findings demonstrate that machine learning approaches can meaningfully contribute to conservation science by identifying endangered species more accurately and revealing the complex factors that contribute to extinction risk. By combining ecological theory with computational methods, we can develop more effective strategies to preserve the planet's biodiversity for future generations.

7. References

Broom, Douglas. “Humans Are Causing Larger Species to Go Extinct Faster.” World Economic Forum, 17 June 2019, www.weforum.org/stories/2019/06/humans-are-causing-larger-species-to-go-extinct-faster/.

Bull, J W, and M Maron. “How humans drive speciation as well as extinction.” Proceedings. Biological sciences vol. 283,1833 (2016): 20160600. doi:10.1098/rspb.2016.0600

Ceballos, Gerardo, et al. “Biological Annihilation via the Ongoing Sixth Mass Extinction Signaled by Vertebrate Population Losses and Declines.” *Proceedings of the National Academy of Sciences of the United States of America*, U.S. National Library of Medicine, 25 July 2017, pmc.ncbi.nlm.nih.gov/articles/PMC5544311/.

Chichorro, Filipe, et al. “A review of the relation between species traits and extinction risk.” Biological Conservation, vol. 237, Sept. 2019, pp. 220–229, https://doi.org/10.1016/j.biocon.2019.07.001.

IUCN. 2024. The IUCN Red List of Threatened Species. Version 2024-1. https://www.iucnredlist.org. Downloaded on 10 March 2025.

J;, Bland LM;Collen B;Orme CD;Bielby. “Predicting the Conservation Status of Data-Deficient Species.” *Conservation Biology : The Journal of the Society for Conservation Biology*, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/25124400/. Accessed 27 Apr. 2025.

Jones, Kate E., et al. "PanTHERIA: A Species-Level Database of Life History, Ecology, and Geography of Extant and Recently Extinct Mammals." *Ecology*, vol. 90, no. 9, 2009, p. 2648. https://esapubs.org/archive/ecol/E090/184/metadata.htm.

Kirk, Holly, et al. “Building biodiversity into the urban fabric: A case study in applying Biodiversity Sensitive Urban Design (BSUD).” Urban Forestry & Urban Greening, vol. 62, July 2021, p. 127176, https://doi.org/10.1016/j.ufug.2021.127176.

The IUCN red list of threatened species. (n.d.). IUCN Red List of Threatened Species. https://www.iucnredlist.org/search

Toussaint, Aurele, et al. “Extinction of Threatened Vertebrates Will Lead to Idiosyncratic Changes in Functional Diversity across the World.” *Nature News*, Nature Publishing Group, 27 Aug. 2021, www.nature.com/articles/s41467-021-25293-0.