MS-CC

# Eco Predict ML:Understanding and Predicting Ecosystem Disruption Impacting Mammals Using Machine Learning


Created By:
Juliet Alozie & Justin Freeman
Fisk University & Alabama State University
MS-CC Summer Research Internship


# Presentation Overview

- Introduction to ecosystem disruption and mammals
- Project goals
- Machine learning approach
- Results summary
- Methodology overview
- Key Takeaways and limitations
- Future directions
- Q&A

# Background Info

Did you know that polar bears rely on cold climates and sea ice to hunt? As global temperatures rise and weather patterns shift, ecosystems worldwide are being disrupted. These changes, like increasing heat, altered rainfall, and more extreme weather, can damage natural habitats and interfere with how species migrate, reproduce, and survive. While conservation scientists have traditionally used fieldwork and statistical models to monitor biodiversity loss, these methods often fall short in capturing the full scale and complexity of environmental change.

To address this, our project combined climatology, ecology, and machine learning to better understand how long-term climate variability affects mammal populations. Using the PanTheria dataset, we applied Random Forest models, an advanced machine learning approach that builds and combines multiple decision trees to uncover patterns between species traits, climate variables, and biodiversity loss. Our findings aimed to support more proactive conservation efforts and inform effective environmental policy planning.



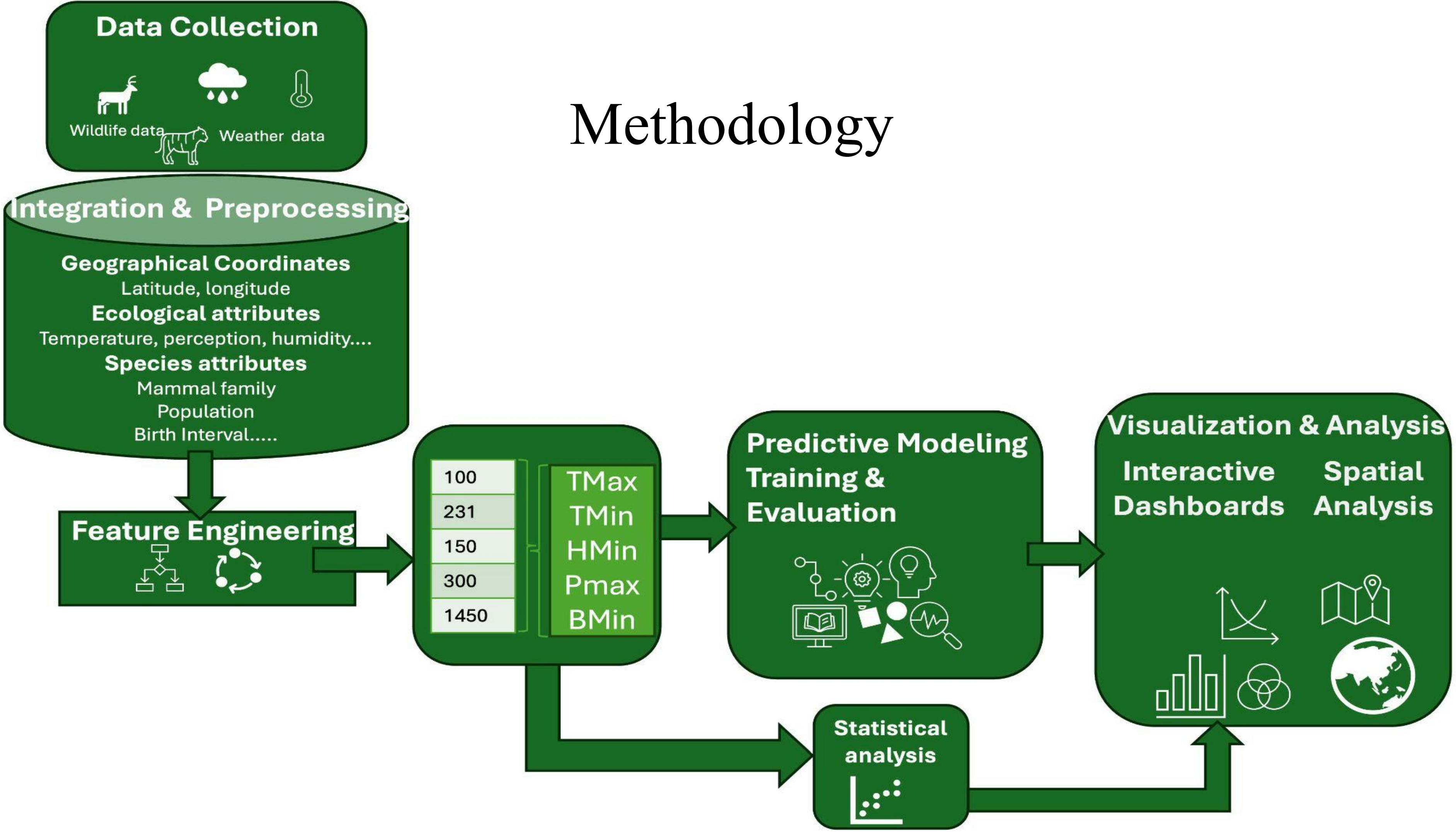

# Why This Research Matters

- Climate change, habitat loss, and land-use shifts increasingly threaten global biodiversity
- Mammals are sensitive indicators of environmental health
- Climate traits alone can predict risk with high accuracy
- Machine learning can help identify complex patterns, predict risks, and support conservation efforts

# Research Objectives

- Understand how long-term climate variability affects mammal species
- Transform complex climate and species data into actionable insights for conservation planning
- Integrate climate and mammal data using machine learning
- Train ML models to support environmental monitoring and conservation

# Methodology



# Data Collection

Collected mammal population dataset from **NOAA(National Oceanic and Atmospheric Administration)** and mammal tracking and population statistics from the **PanTheria** dataset. Those datasets contained trait data on over 5,000 mammal species. The collected data spans multiple years to account for both short-term fluctuations and long-term weather trends.

1. Assembled time-series data
2. Merged species and climate data

# Data Preprocessing

- ❖ Cleaned the dataset, imputed missing values, and removed outliers
  - ➢ Removes irrelevant data that can affect the scaling and interpretation of results
- ❖ Encoded species traits and normalized variables
  - ➢ Allows the model to process and puts all features on a common scale
- ❖ Ensured data consistency for ML input

# ML Algorithm Used-Random Forest

Applied the Random Forest Regression and Classification models  to analyze the relationship between mammal traits and ecosystem disruption variables

What it does:

- ❖ Handle complex ecological datasets
- ❖ Perform both classification and regression models
- ❖ Aggregates decision trees to reduce overfitting

# Training and validation strategy

Chosen for handling nonlinear ecological data with feature importance ranking

- ❖ Train/test split
  - ➢ Trains the model and evaluates its performance
- ❖ Performed fold cross-validation
  - ➢ Divides the data into multiple subsets and provides a more reliable metric
- ❖ Evaluated using accuracy, precision, recall, and F1-score (classification matrix)

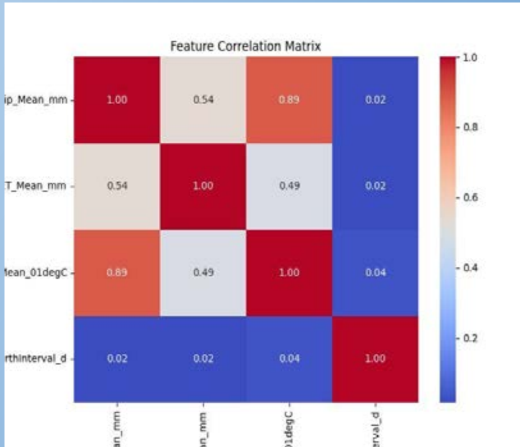
# Exploratory Data Analysis (EDA)

Explored the correlation between species traits and climate

- ❖ Regression analysis is used to quantify how specific atmospheric variables influence mammal survival, migration, and reproduction.
- ❖ Principal Component Analysis(PCA) is used to reduce dimensionality and identify the most significant weather variance factors.
- ❖ Time-series analysis is used to study long-term trends and survival analysis to assess the impact of extreme weather events on population viability.

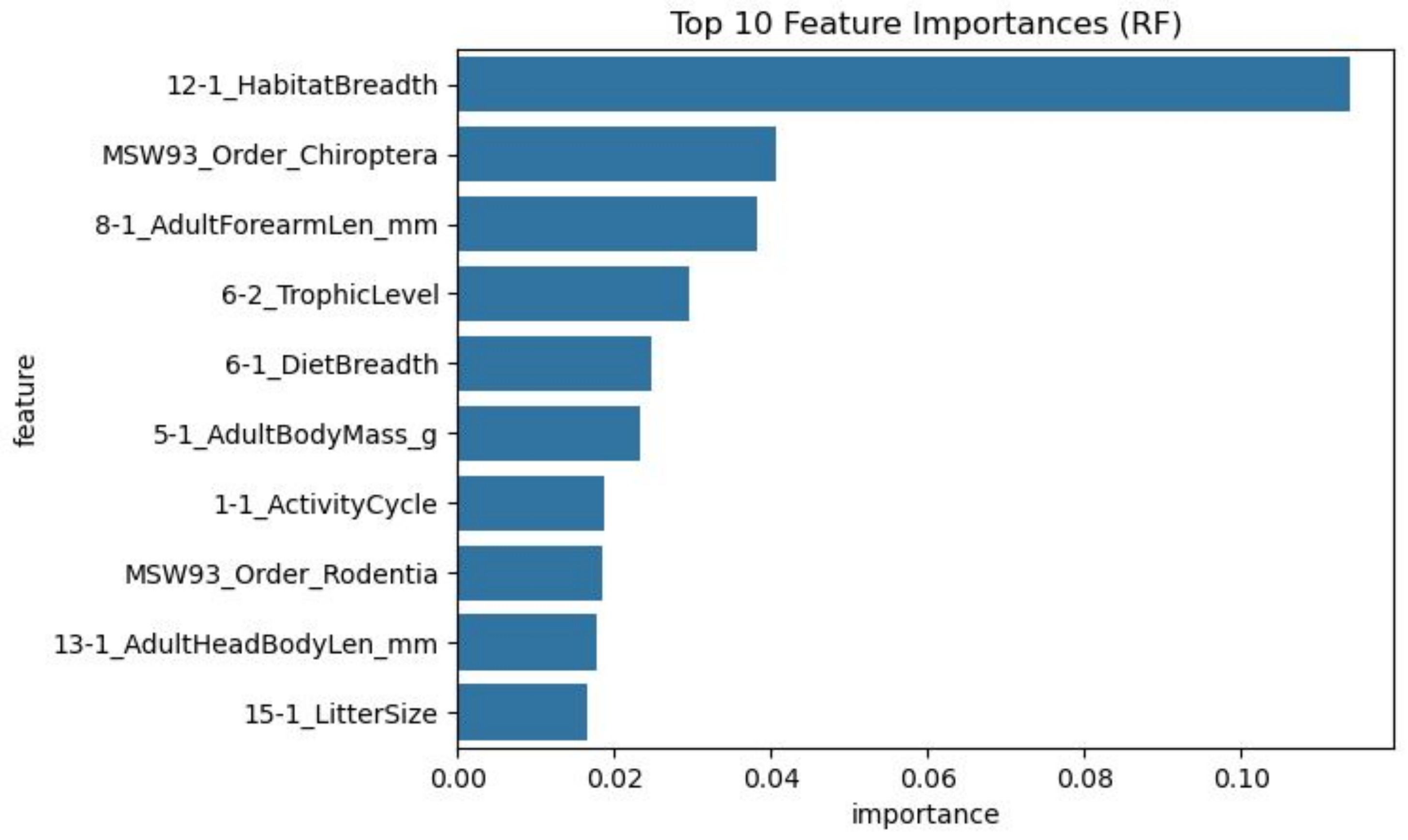
Mapped biodiversity vs. climate disruption

- ❖ Species sensitivity to environmental changes provides a strong predictive signal. EDA revealed that mammal biodiversity is tightly linked to specific weather variables



# Results

- Combining temporal weather data with species traits enables effective feature selection for predictive analytics in wildlife conservation.
- Random Forest Classifier was used for initial analysis, achieving a 60% accuracy score in predicting mammal species outcomes based on climate-related features.
- The classification matrix showed performance metrics across different mammal breeds, highlighting both strengths and limitations in prediction accuracy.
- The EcoPredictML framework showed potential as an AI-driven decision support tool, integrating species and atmospheric data for ecological insight.




This is a bar graph showing the most important features identified by the Random Forest model, such as the rainfall, and specific species traits.

It reveals which variables contributed most to accurate predictions in the model. This supports the claim that climate-based features are strong predictors of species vulnerability and ecosystem risk.

# Limitations/challenges

- Data Quality & Availability
  - Incomplete or sparse data weakens prediction accuracy and generalizability
- Class Imbalance in Model Training
  - Model favors common species, missing critical insights for endangered ones
- Spatial & Temporal Resolution Mismatch
  - Misaligned data timing or scale hides real ecological patterns and risks

# Future Direction

- Integrate satellite imagery and GPS tracking for deeper spatial analysis
- Address class imbalances to enhance classification accuracy
- Map geographic data to locate at-risk mammal habitats
- Refine features using correlation analysis and clustering

# Real_World Impact

- Bridges AI and Ecology
  - guiding targeted protection efforts
- Supports Biodiversity Monitoring
  - Identify ecological hotspots and track habitat disruption over time
- Informs Conservation Strategies
  - Empowers scientists and policymakers with data-driven tools for ecosystem resilience

# What We Learned

- The use of Python, Jupyter Notebook, and Google Colab for ML
  - pandas, sklearn, matplotlib, visualizing data
- ML Concepts and Tools
  - Exploratory Data Analysis, Random Forest Algorithm, Linear Regression
- How AI is influencing areas unrelated to computer science
  - Can be informative for environmental precautions, and climate change awareness
  - Human connection to changes in biodiversity

# Acknowledgments and closing

Funded by: MS-CC, Internet 2, NSF Awards #2137123 and #2234326

Mentorship by:

Dr. Sajid Hussain, Faculty Mentor

Asmah Muallem, PhD, Co-Mentor

Loni Taylor, MS, PMP, CETL, Co-Mentor

Special thanks to all or you and to the workshop coordinators and our cohort

# Q&A