# **Frog Occurrence Prediction: Final Value Case Report**

**Course:** BUAN 6390.003 – Analytics Practicum - S25

**Date:** April 15, 2025

## Vinayak Shanbhag

**Executive Summary**

This project aimed to predict frog species presence using 14 climate variables from the TerraClimate dataset. Leveraging machine learning techniques, our team developed a robust and reproducible classification model. After iterative optimization, the final Random Forest model achieved an F1 score of **0.8404**, recall of **0.9561**, and accuracy of **0.8176**—validated independently. The pipeline includes automated preprocessing, hyperparameter tuning, and integration with ecological data sources, making it suitable for real-world deployment across ecological, regulatory, and research domains.

**Business Application**

Our predictive frog occurrence model offers practical value across a wide spectrum of industries and organizations, transforming complex ecological data into actionable insights for:

1. **Government Wildlife Agencies – Biodiversity Conservation**

Frog species are known bioindicators of environmental health. Government agencies such as the U.S. Fish & Wildlife Service or Environment Canada can use this model to:

* Detect frog habitats across unmonitored regions using only latitude and longitude.
* Track biodiversity loss and assess the impact of climate change.
* Allocate conservation resources more efficiently.
* Replace exhaustive field surveys with accurate prediction-based planning, thereby improving habitat restoration ROI.

2. **Environmental Consulting & Land Use Planning – Ecosystem Risk Assessments**

Before large-scale projects like construction, mining, or urban development:

* Environmental consultants can use this model to evaluate the biodiversity sensitivity of a site.
* Urban planners and infrastructure developers can ensure compliance with ecological regulations by identifying protected zones.
* The model empowers baseline ecological risk assessments, guiding sustainable land-use decisions while preventing disruption to sensitive habitats.

3. **Academic & Research Institutions – Climate & Migration Studies**

This model provides a ready-to-use framework for ecologists, conservation biologists, and students to:

* Study amphibian behavior and migration patterns linked to climate fluctuations.
* Conduct longitudinal biodiversity studies by integrating temporal TerraClimate data across years or seasons.
* Analyze interactions between environmental factors and species presence with a scalable, replicable pipeline.

4. **NGOs & Global Conservation Bodies – Advocacy & Public Awareness**

**Organizations like WWF and IUCN can:**

* Leverage heatmaps from the model to visually highlight endangered zones.
* Influence environmental policies using data-driven biodiversity predictions.
* Mobilize community action and funding by raising awareness about high-risk habitats through visual storytelling.

5. **Insurance & Agribusiness – Ecological Risk Scoring**

In the environmental risk and agriculture sectors:

* Insurers can incorporate frog presence as a proxy for ecosystem health in their risk scoring algorithms.
* Agribusinesses can use predictions to identify pollution-sensitive zones, plan sustainable crop placement, or water resource management.

6. **Citizen Science & Ecotourism – Public Engagement**

The model’s simplicity (accepting only lat/lon inputs) allows for:

* Integration into mobile/web platforms for hobbyist researchers and field enthusiasts.
* Empowering citizen science projects that track frog populations in real-time.
* Enhancing eco-tourism by highlighting biodiversity-rich regions for photographers and nature travelers.

Each of these applications demonstrates the versatility and real-world relevance of our model, serving stakeholders from conservationists and researchers to business strategists and environmental regulators. With minimal inputs and no dependency on physical sampling, the model offers a fast, cost-effective, and scalable solution for ecological decision-making in a changing climate.

**Business Case**

Frogs are widely recognized as ecological indicators. Their presence (or absence) signals shifts in environmental health, water quality, and ecosystem stability. However, traditional monitoring is manual, expensive, and limited in coverage.

1. **Monetization:**

* **API as a Service**: Offer subscription access to real-time frog occurrence predictions.
* **Licensing**: License the model to NGOs, academic consortia, or biodiversity platforms.
* **Platform Integration**: Embed predictions in broader risk intelligence tools for land planners or regulators.

1. **Return on Investment ROI Consideration**:

* **Cost Avoidance**: Traditional ecological surveys cost ~$2,000/site. This model delivers instant predictions, cutting per-site cost by >90%.
* **Scalability**: Model runs quickly on cloud infrastructure, enabling country-wide risk mapping with minimal compute overhead.
* **Phase II Growth**: Expand into seasonal modelling, amphibian migration prediction, and integration with satellite data for global coverage.

**Improving Results**

Achieving our final results involved a series of deliberate, data-driven optimizations throughout the pipeline.

* **Data Merging Fix**:
  + We used direct coordinate merging after precisely extracting variables from the raster TIFF using rasterio and ensuring exact match without rounding. This led to a dramatic improvement in prediction accuracy and alignment with validation data.
* **Model Optimization**:
  + Hyperparameter tuning of Random Forest via cross-validation and RandomizedSearchCV significantly increased performance.
* **Feature Reduction**:
  + Feature correlation analysis helped reduce multicollinearity and overfitting by dropping redundant variables like pet, q, and ppt.
* **Final Model Evaluation:**
  + The final Random Forest model yielded:
    - Accuracy: **0.8176**
    - Precision: **0.7496**
    - Recall: **0.9561**
    - F1 Score: **0.8404**
* **Opportunities for Improvement & Future Development:**
  + **Temporal & Seasonal Modelling**: Expanding the dataset with multi-year and monthly climate variables would enable seasonality-aware predictions—capturing migration, breeding, and habitat dynamics more accurately.
  + **Feature Enrichment & Ecological Context**: Incorporating high-resolution environmental data such as vegetation indices (e.g., NDVI), soil type, and water body proximity can deepen ecological context, improve accuracy, and support more biologically informed predictions.
  + **Advanced Modelling Techniques**: Exploring ensemble stacking or deep learning approaches (such as convolutional networks for raster data) could uncover complex patterns and improve generalization across diverse ecological zones.
  + **Species-Level Classification**: Transitioning from binary to multi-class classification would allow prediction of specific frog species, increasing the biological and conservation value of the model.
  + **Scalability & Validation Integration**: Automating external validation using ecological survey templates and testing across different regions would enhance the model’s robustness and plug-and-play potential for field use.
  + **Model Transparency & Interpretability**: Adding explainability tools like SHAP or LIME would make the model’s predictions more transparent and build trust among both technical and non-technical stakeholders.

**Final Takeaway & Conclusion**

This project successfully delivers a reliable, fast, and scalable solution for ecological monitoring. The model has been rigorously validated and aligns closely with expert evaluations. It not only serves academic and ecological value but also has strong business potential in both public and private sectors. With a final F1 score exceeding **0.8404** on unseen data, the system is deployment-ready and supports the broader mission of climate-aware decision-making and conservation-first planning.