VALUE CASE

The EY Biodiversity Data Challenge

Predicting Frog Species Presence Using Machine Learning on Climate Data

Prepared By Group 6

SUMAN ANAND - SXA230187

SAI KUMAR - SXK220458

PRACHI RAJENDRA TRIVEDI - PXT230010

KAUSHIL MADHUBHAI MANGAROLIYA - KMM210014

Executive Summary

1. ## Problem Overview
   Frog populations are declining rapidly due to habitat destruction, pollution, and climate change. As sensitive bioindicators, frogs are critical for gauging ecosystem health. This project, developed for the EY Biodiversity Challenge, addresses this problem by predicting frog species presence using machine learning models trained on TerraClimate data in southeastern Australia.

2. ## Project Approach
   Using environmental predictors such as precipitation, evapotranspiration, radiation, vapor pressure, and drought indices, we constructed a structured machine-learning pipeline. Feature engineering, data balancing using Synthetic Minority Over-sampling Technique (SMOTE), and hyperparameter tuning led to a robust Random Forest model.

3. ## Results
   Our model achieved an F1 score of 82.44%, a significant improvement from the benchmark baseline. The score reflects balanced precision and recall, which are crucial for ecological classification, where false negatives can result in overlooked conservation zones.

4. ## Impact and Value
   The solution provides scalable, data-driven insights to support biodiversity conservation, policy decisions, ESG (Environmental, Social, and Governance) reporting, and climate resilience planning. The model's interpretability and performance make it deployable across ecological and business applications.

5. ## Key Takeaway
   This project demonstrates how climate and biodiversity data can be transformed into actionable intelligence, aiding a range of stakeholders in protecting biodiversity while meeting global sustainability goals.

Business Application

1. Conservation Strategy and Habitat Restoration

Beneficiaries

- NGOs like the World Wide Fund for Nature (WWF) and Conservation International would benefit by optimizing conservation investments and maximizing ecological return.
- International funding bodies can allocate resources with better targeting toward restoration projects with verified biodiversity potential.

Who Would Use the Analysis

- Field ecologists and restoration scientists are responsible for designing and executing ecosystem rehabilitation programs.
- Government-funded conservation projects aiming to align with national and global biodiversity goals.
- Environmental NGOs focused on amphibian preservation and habitat corridor connectivity.

How Results Could Be Used

- To identify zones with high potential for frog recovery and long-term amphibian survival.
- Prioritize degraded or fragmented habitats that are climatically suitable for frog species presence and resilience.
- Assess the impact of previous restoration initiatives by cross-verifying past presence data with model predictions.

Value of the Analysis

- Enables high-precision restoration planning and reduces dependency on exhaustive field sampling.
- Directly improves cost-efficiency by allowing NGOs to focus labor and funding in areas where the impact will be highest.
- Enhances coordination between government, nonprofit, and academic partners by offering a data-driven starting point.
- Rowley et al. (2020) highlight frog monitoring for habitat health; Global Forest Watch datasets can refine restoration overlays.

2. Government Environmental Policy & Planning

Beneficiaries

- Federal and regional departments of environment, forestry, and land management responsible for biodiversity action plans.
- Environmental regulatory bodies, such as Australia's Department of Climate Change, Energy, the Environment and Water.

Who Would Use the Analysis

- Urban planners, zoning officials, and ecological advisors evaluate the environmental impact of proposed development projects.
- Climate strategy units and ecological risk assessment teams work to meet biodiversity offset and habitat conservation regulations.

How Results Could Be Used

- To enhance the precision of Environmental Impact Assessments (EIAs) and integrate species presence data into infrastructure planning.
- Prioritize regulatory protection for sensitive habitats and refine biodiversity offset zones.

- Feed predictions into state biodiversity atlases and land-use policy models to monitor long-term change.

Value of the Analysis

- Increases transparency and defensibility of environmental regulations by grounding them in empirical, climate-driven predictions.
- Reduces time and cost in ecological evaluations and enhances stakeholder confidence in development decisions.
- Supports international reporting obligations under frameworks like the Convention on Biological Diversity and Sustainable Development Goal (SDG) 15.
- It closely aligns with Australia's Environment Protection and Biodiversity Conservation Act standards and national biodiversity offset guidelines.

3. Corporate ESG and Sustainability Reporting

Beneficiaries

- Companies in mining, agriculture, real estate, energy, and infrastructure aiming to de-risk biodiversity impacts.
- ESG investment firms and sustainable finance coalitions are seeking transparency in environmental metrics.

Who Would Use the Analysis

- Corporate sustainability teams, ESG reporting officers, and compliance managers are responsible for Carbon Disclosure Project (CDP)/ Taskforce on Nature-related Financial Disclosures (TNFD) disclosures.
- Independent auditors and ESG benchmarking firms who assess companies' biodiversity dependencies and risks.

How Results Could Be Used

- Support ESG due diligence by flagging project sites with high biodiversity sensitivity.
- Contribute to biodiversity Key Performance Indicators (KPIs) and carbon offset project validation.
- Inform nature-positive strategies such as ecological buffer zones and conservation funding partnerships.

Value of the Analysis

- Helps meet stakeholder expectations for transparency and accountability in nature-related risk disclosures.
- Strengthens market positioning in ESG rankings and appeals to green investors.
- Avoids reputational and operational risks associated with habitat encroachment or environmental non-compliance.
- Validates disclosures under frameworks like TNFD, CDP, and the EU Corporate Sustainability Reporting Directive (CSRD).

4. Climate Resilience and Disaster Planning

Beneficiaries

- Local councils, water authorities, and emergency planning departments managing climate adaptation.
- Environmental NGOs and consultants working on urban resilience and ecosystem-based disaster risk reduction.

Who Would Use the Analysis

- Civil engineers, infrastructure resilience planners, and climate adaptation strategists involved in ecological risk modeling.
- Government contractors design green infrastructure such as floodplains, bioswales, and buffer zones.

How Results Could Be Used

- Overlay habitat sensitivity with flood, drought, and wildfire risk maps to develop integrated nature-based solutions.
- Inform post-disaster recovery programs that incorporate ecological restoration alongside infrastructure repair.
- Prioritize climate adaptation investments in biodiversity hotspots most sensitive to climate variability.

Value of the Analysis

- Adds a biological early-warning layer to traditional climate risk maps.
- Enhances the effectiveness of flood control and water retention measures through nature-based infrastructure.
- Aligns climate resilience planning with ecosystem protection goals for greater co-benefits.
- Informed by the Commonwealth Scientific and Industrial Research Organization's (CSIRO) nature-based adaptation recommendations and the United Nations Environment Programme (UNEP) Nature-based Solutions toolbox.

5. Education, Citizen Science, and Awareness Campaigns

Beneficiaries

- K–12 and university students, educators, nonprofit environmental educators, and local conservation communities.
- Public outreach organizations and citizen science platforms focused on biodiversity education.

Who Would Use the Analysis

- Science teachers, curriculum designers, and educational NGOs developing nature-themed programs.
- Citizen science app developers and local engagement campaign managers encouraging biodiversity action.

How Results Could Be Used

- Convert frog presence predictions into educational maps, dashboards, and real-time biodiversity trackers.
- Empower students to explore climate-biodiversity relationships and validate local predictions through field observations.
- Enable collaborative learning by connecting local data with national biodiversity efforts.

Value of the Analysis

- Enhances ecological literacy and environmental awareness among youth and the public.
- Provides accessible tools to support community involvement in conservation and species monitoring.
- Strengthens the role of informal education and outreach in biodiversity protection strategies.
- Aligned with the National Academies' framework on participatory learning and citizen science.

Business Case

Target Market Segments

- Environmental Consulting Firms represent the most immediate and high-value market opportunity. These firms can integrate the predictive model into their environmental impact assessments and biodiversity monitoring services.
- Government Agencies are the key stakeholders, which include federal, state, and local departments involved in environmental protection, conservation, and urban planning. These agencies maintain multi-million-dollar budgets for environmental monitoring.
- Academic and Research Institutions & organizations are likely to adopt the model for both scientific studies and educational purposes. These institutions can serve as early adopters and credibility-enhancing reference clients.
- Private Land Developers and Resource Extraction Companies operating in real estate, mining, and natural resource sectors face growing regulatory requirements for environmental compliance. This model provides a cost-effective solution for ongoing biodiversity monitoring.

Monetization Strategy

- Launch a subscription-based biodiversity intelligence platform that provides predictive habitat insights accessible through Application Programming Interfaces and Geographic Information System integrated dashboards.
- Offer tiered pricing for NGOs, government bodies, and corporates, with advanced features for enterprise clients (e.g., real-time updates and custom reports).

Who Would Pay for It

- Conservation NGOs seeking high-impact restoration zones.

- Government agencies are responsible for compliance, risk analysis, and biodiversity planning.
- ESG-focused corporations need biodiversity metrics for regulatory and investment reporting.
- Educational institutions and citizen science platforms for curriculum integration and public engagement.

How It Could Be Marketed or Publicized

- Collaborate with biodiversity conferences and forums to showcase the model's application.
- Partner with environmental research labs and academic institutions for co-branded pilot studies.
- Launch an open-access version for educational outreach, with upsell options for analytics features.

Return on Investment (ROI)

- Expected to reduce manual biodiversity survey costs by up to 50%.
- Accelerates policy and ESG report generation with reliable, scalable data.
- Strengthens organizational reputation and stakeholder trust by enabling nature-positive actions.
- Offers first-mover advantage in a growing market for predictive ecological intelligence.

Competitive Advantage and Differentiation

- Domain-Specific Focus as the model is uniquely tailored to frog biodiversity as an ecological indicator, addressing a niche not well-covered by general species distribution models.
- Achieves high accuracy in amphibian habitat prediction, distinguishing it from more generalized ecological models.
- Leverages freely accessible climate data from TerraClimate, eliminating the need for costly proprietary datasets and reducing operational expenses.
- Designed to be adaptable to other geographic regions and amphibian species, enabling expansion without requiring fundamental architectural changes.

Improving Results

1. Additional Data Integration

- Integrate high-resolution satellite imagery (e.g., Normalized Difference Vegetation Index (NDVI), land cover data) to refine microhabitat classification.
- Include proximity to water bodies, soil moisture content, and elevation models to improve ecological context.
- Use acoustic monitoring data or citizen science audio inputs (e.g., frog calls) to improve presence validation.

2. Temporal Expansion

- Extend the dataset to cover a 10+ year period and include future climate projection scenarios (e.g., Coupled Model Intercomparison Project Phase 6 (CMIP6) models).
- Analyze seasonal trends more effectively using rolling temporal windows and interannual variability assessments.

3. Modeling Enhancements

- Implement hybrid models combining Convolutional Neural Networks (for spatial pattern recognition) and gradient boosting (for feature importance).
- Fine-tune model performance using ensemble techniques across multiple algorithms (e.g., XGBoost, Long Short-Term Memory (LSTM)).
- Incorporate SHapley Additive exPlanations (SHAP) values or Local Interpretable Model-Agnostic Explanations (LIME) to increase model interpretability for non-technical stakeholders.

4. Sampling and Labeling Improvements

- Redefine 'non-presence' zones based on ecological thresholds rather than assumed absence to reduce false negatives.
- Filter out urban or artificial surfaces using land cover classifications to reduce data noise.
- Perform iterative feedback loops with field experts to relabel ambiguous or disputed cases.

5. Field Validation and Community Integration

- Partner with the Australian Museum's FrogID initiative and other citizen science platforms to validate predictions.
- Integrate mobile tools where users can contribute real-time validation (e.g., frog sightings, photos, or calls).
- Use crowdsourced validations to continuously refine the model with locally informed corrections.

How These Would Improve Value

- Increased model accuracy (targeting >90% F1 score) would boost user confidence in decision-making contexts.
- Better interpretability and ecological realism enhance policy relevance and regulatory adoption.
- Continuous improvement through feedback loops aligns with agile conservation and adaptive planning goals.