

PROBLEM 1: BIOSCOUT ISLAMABAD

AI FOR COMMUNITY BIODIVERSITY & SUSTAINABLE INSIGHTS

Event Theme: AI for a Sustainable Future

Event Date: [May 16th, 2025]

Location Focus: Islamabad & Surrounding Regions (e.g., Margalla Hills National Park), Pakistan

1. Background: Biodiversity and Sustainability in Islamabad

Islamabad, nestled against the scenic Margalla Hills National Park, is a unique blend of urban and natural ecosystems, home to a rich diversity of flora and fauna. From vibrant birdlife and mammals in the hills to diverse plant species in green spaces like Shakarparian and Rawal Lake, this region's natural heritage is a treasure. However, urbanization and climate change pose significant threats to this biodiversity. Protecting these ecosystems aligns with the United Nations Sustainable Development Goals (SDGs), particularly SDG 15 (Life on Land) and SDG 13 (Climate Action). "BioScout Islamabad" leverages AI to empower communities to monitor and conserve local species, contributing to a sustainable future by fostering environmental awareness and data-driven conservation strategies.

2. The Challenge: Build "BioScout Islamabad"

Your challenge is to design and prototype "BioScout Islamabad," an AI-powered web platform that serves as a community-driven biodiversity database for Islamabad and surrounding ecological areas, featuring AI-assisted species identification and an intelligent Q&A system powered by Retrieval Augmented Generation (RAG).

The platform should enable community members to submit biodiversity observations and learn about local species and ecosystems through AI-driven tools. Imagine your platform being used by schools, hikers, and conservationists in Islamabad to track species, report rare sightings, and inspire real-world environmental stewardship. Your goal is to create a tool that not only collects data but also educates users and contributes to a sustainable future.

The focus is on innovative integration of existing AI models and datasets, not on training models from scratch. Given the 24-hour timeframe, prioritize a Minimum Viable Product (MVP) with core functionalities over advanced features.

3. Key Features & AI Integration Requirements

Your "BioScout Islamabad" platform prototype should include the following core components as an MVP:

A. Community Biodiversity Observation Hub:

- **Data Submission:** A simple user-friendly interface for community members to submit biodiversity observations, including:
 - Species name (common or scientific, if known).
 - Date and location of observation (text input or basic map pin within Islamabad and areas like Margalla Hills National Park).
 - Option to upload an image of the observed species.
 - A "Notes" field for additional details (e.g., behavior, habitat).
- **Data Viewing:** Display submitted observations as a list or on a simple interactive map of Islamabad and its environs (advanced filtering by species, date, or location is optional).
- **AI-Powered Species Identification (Assisted ID):**
 - When a user uploads an image, integrate a pre-trained image recognition model/API to suggest potential species.
 - Clearly indicate this is an AI suggestion (confidence scores optional).
 - Use freely available models/APIs (e.g., via Hugging Face, iNaturalist API, TensorFlow Hub – check terms of use). If API integration is a barrier, mock the output with hardcoded suggestions and explain the intended approach.
- **Basic Gamification:** Include a simple recognition mechanism, like a submission count or 'Top Observer' badge (can be mock data for the prototype), to encourage community participation.

B. RAG-Enhanced Biodiversity Q&A System:

- **Natural Language Queries:** Allow users to ask questions about local biodiversity in plain English (e.g., "What birds are common in Margalla Hills?", "Are there recent sightings of leopards near Rawal Lake?").
- **Knowledge Base for Retrieval:**
 - **Curated Documents:** Create a small knowledge base of 3-5 text snippets (in .txt or .md format) about flora and fauna relevant to Islamabad and Margalla Hills,

sourced from reliable online resources (a list of suggested sources like IWMB, WWF-Pakistan will be provided).

- **Observation Data Integration:** Include textual components of community-submitted observations (e.g., species names, notes) in the searchable data.
- **Retrieval Mechanism:** Implement a simple system to retrieve relevant text snippets from the knowledge base and observations based on the user's query (e.g., keyword search or basic semantic search using pre-trained embeddings).
- **Augmented Generation:** Use retrieved context to augment a prompt for a pre-trained Large Language Model (LLM) via a free API (e.g., Hugging Face Inference API, Google Gemini free tier – check availability). If API access is a hurdle, simulate the generation by preparing sample answers based on retrieved context and explain the intended integration.

4. Dataset Requirements (To be created by participants)

- **Sample Biodiversity Observation Dataset:**
 - Create a structured dataset (e.g., CSV format) with at least 10-15 fictional but plausible biodiversity observations from the Islamabad region (e.g., Margalla Hills, Rawal Lake).
 - Include fields like `observation_id`, `species_name`, `common_name`, `date_observed`, `location` (text or lat/long), `image_url` (placeholder), and `descriptive notes`. A template with 2-3 sample rows will be provided.
 - This dataset will initially populate and test your platform and RAG functionalities. Document your assumptions.
- **RAG Knowledge Base Documents:**
 - Curate 3-5 text snippets about biodiversity relevant to Islamabad and surroundings. A sample snippet will be provided as a starting point.

5. Technical Stack Guidance

- **Frontend:** Your choice (e.g., HTML, CSS, JavaScript, React, Streamlit for rapid prototyping).
- **Backend:** Your choice (e.g., Python with Flask/Django, Node.js).

- **Database:** Your choice (e.g., SQLite, Firebase, simple file-based storage for CSV).
- **AI Integration:**
 - **Species ID:** Use existing pre-trained models/APIs (e.g., iNaturalist API, Hugging Face models). Mock outputs if integration is a barrier.
 - **RAG:**
 - Embeddings (for semantic search, if attempted): Libraries like Sentence Transformers.
 - Similarity Search: Simple keyword matching or tools like scikit-learn's cosine_similarity.
 - LLMs (for generation): Free/developer-tier APIs (Hugging Face, Google Gemini). Simulate responses if needed.
- **Beginner Support:** For teams new to AI, prioritize simple tools. Tutorials and fallbacks (e.g., hardcoded AI suggestions, keyword search for RAG) will be provided. No model training from scratch is expected—focus on API integration or simulation.
- **Inspiration:** Look at platforms like iNaturalist for citizen science ideas, but your unique contribution is the locally focused knowledge base and AI-powered Q&A for Islamabad's sustainability.

6. Deliverables (24-Hour Hackathon)

- **Working Prototype:** A deployed or locally runnable version of "BioScout Islamabad" demonstrating the MVP features.
- **Source Code:** Access to complete source code (e.g., via Git repository).
- **Datasets:**
 - Sample biodiversity observation CSV file (10-15 entries).
 - Collection of 3-5 text snippets for the RAG knowledge base.
- **Presentation & Demo (5-7 minutes):**
 - Explain your platform's architecture and core features (submission, AI ID, Q&A).
 - Demonstrate community observation submission, AI-assisted ID, and RAG Q&A with example queries relevant to Islamabad's biodiversity.

- Discuss data sources and how the RAG system works (or would work with an LLM).
- Briefly (1-2 minutes) outline how "BioScout Islamabad" could scale or evolve to support conservation (e.g., integrating with national databases, partnering with NGOs).
- Highlight challenges and potential improvements.

7. Optional Enhancements for Extra Recognition

Implement features beyond the MVP for additional consideration under "Innovation & User Experience" and potential special mentions (e.g., "Most Creative Feature"):

- User accounts or advanced gamification for submitting observations.
- Community validation of AI-suggested species IDs.
- Advanced RAG (e.g., sophisticated embeddings or retrieval strategies).
- Multilingual support for queries/content (e.g., Urdu).
- Offline capabilities or Progressive Web App (PWA) for field use in areas like Margalla Hills.