**DSC 680 Applied Data Science**

**Chitramoy Mukherjee, T301 T302 2253 Winter 2025**

**Project-2 Milestone-1**

**Wildlife Protection of Endangered Species**

**Description:**

An endangered species is a type of organism that is threatened by extinction. Species become endangered for two main reasons: loss of habitat and loss of genetic variation. The conservation of endangered and rare wildlife is a primary part of preserving biodiversity. The purpose of this project is to intelligently monitor endangered and rare wildlife species by using Machine Learning models with animal data and computational resources from various parks in United States.

**Problem Statement:**

The rapid decline in populations of endangered species globally poses a critical threat to biodiversity and the stability of ecosystems. Traditional methods of monitoring and protecting these species are often labor-intensive, time-consuming, and prone to inaccuracies. There is an urgent need for innovative, scalable, and efficient solutions to identify, monitor, and protect endangered species effectively. Leveraging advancements in Artificial Intelligence (AI) and Machine Learning (ML), we propose developing a sophisticated AI/ML-based system to enhance the efforts of wildlife conservationists, researchers, and policymakers in safeguarding endangered species.

**Objectives:**

1. **Identification:** Develop AI/ML models capable of accurately identifying endangered species in their natural habitats using camera traps, drones, and satellite imagery.
2. **Monitoring:** Implement AI-powered real-time monitoring systems to track the movements, behaviors, and population dynamics of endangered species, providing valuable data for conservation efforts.
3. **Threat Detection:** Utilize machine learning algorithms to predict and identify potential threats to endangered species, such as poaching, habitat loss, and climate change impacts, allowing for timely and targeted interventions.
4. **Data Integration:** Create a centralized platform to integrate and analyze data from various sources, including field observations, sensor networks, and citizen science initiatives, to provide comprehensive insights into the status of endangered species.
5. **Policy Support:** Generate actionable insights and recommendations for policymakers and conservation organizations to formulate and implement effective conservation strategies based on data-driven evidence.

By harnessing the power of AI/ML, this project aims to revolutionize wildlife protection efforts, ensuring the survival and flourishing of endangered species for future generations.

**Datasets:**

Here I am using 3 datasets as below:

**File**: Species.csv

**Source**: Abigail Larion(n.d.), <https://www.kaggle.com/datasets/nationalparkservice/park-biodiversity/data?select=species.csv>

The species dataset contains plants and animal species information from different national parks of United States. This dataset contains 13 columns.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| Species ID | String | Unique ID. National Park service code for each species |
| Park Name | String | Park Name in which the species appears. |
| Category | String | Category of species like mammal, bird etcetera. |
| Order | String | The scientific order the species belongs to. |
| Family | String | The scientific family the species belongs to. |
| Scientific Name | String | Full scientific species name. |
| Common Names | String | Usual common name(s) for the species. Comma-delimited. |
| Record Status | String | record status from park |
| Occurrence | String | Current presence of that species |
| Nativeness | String | Whether the species is native to the area or a non-native/invasive. |
| Abundance | String | Commonality of sightings. |
| Seasonality | String | When the species can be found in the park.  Blank if the species is found there year-round. |
| Conservation Status | String | IUCN species conservation status. |

**Website**: National Park locations and areas

**Source**: <https://irma.nps.gov/NPSpecies>

This data source will be used to get parks information as below. Further information is provided under “Project Approach” section on how this data will be extracted.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| Park Code | String | Unique value. National Park code |
| Park Name | String | Name of the park. |
| State | String | US state(s) in which the park is located. Comma-separated. |
| Acres | Integer | Size of the park in acres. |
| Latitude | Integer | Latitude of the park (centroid). |
| Longitude | Integer | Longitude of the park (centroid). |

**API**: Animals API

**Source**: [Animals API - API Ninjas (api-ninjas.com)](https://api-ninjas.com/api/animals)

The Animals API provides interesting scientific facts on thousands of different animal species.

/v1/animals

**HTTP GET**: Returns up to 10 results matching the input name parameter.

**Parameters**: name (required) - common name of animal to search. This parameter supports partial matches (e.g. fox will match gray fox and red fox)

**Headers**: X-Api-Key (required) - lJCoL9STz0hLjJXG8gEYpQ==IGRnvx5k0MwcjfXR

**Example API call**:

Using Python script in Jupyter Notebook.

import requests  
name = 'fisher'  
api\_url = 'https://api.api-ninjas.com/v1/animals?name={}'.format(name)  
response = requests.get(api\_url, headers={'X-Api-Key': 'lJCoL9STz0hLjJXG8gEYpQ==IGRnvx5k0MwcjfXR'})  
if response.status\_code == requests.codes.ok:  
 print(response.text)  
else:  
 print("Error:", response.status\_code, response.text)

A screen shot of a computer

Description automatically generated

**Methods/Steps:**

Using the provided sources, here are specific analysis methods to complete the Wildlife Protection of Endangered Species project:

**1. Data Collection and Integration**

* **National Park Locations and Areas (Website):** Scrape or extract data on national park locations, areas, and relevant attributes. This data will provide important contextual information about the habitats and protected areas for endangered species.
* **Animals API:** Access and retrieve animal-related data such as species names, characteristics, habitats, and conservation statuses. This API data can be integrated with other sources to enhance the overall dataset.
* **Species.csv (File):** Import and preprocess the CSV file to extract species-specific data. This file may contain information on species names, scientific classifications, population sizes, threats, and conservation statuses.

**2. Geospatial Analysis**

* **Habitat Mapping:** Using GIS tools, map the locations of national parks and overlay them with species distribution data from the Animals API and Species.csv. This helps in identifying critical habitats and conservation areas.
* **Habitat Suitability Modeling:** Apply machine learning algorithms like MaxEnt to predict suitable habitats for endangered species based on environmental variables (e.g., climate, vegetation) extracted from the National Park data and Animals API.

**3. Species Identification and Monitoring**

* **Species Classification:** Develop and train machine learning models to classify different species based on attributes from the Animals API and Species.csv. Use this to identify species captured in images and videos from camera traps or drones within national parks.
* **Population Monitoring:** Utilize data from the Species.csv file to track population trends over time. Apply time-series analysis to detect patterns and predict future population changes.

**4. Threat Detection and Prediction**

* **Poaching and Illegal Activities:** Implement anomaly detection algorithms on data from the national parks to identify unusual activities or potential threats. Combine this with real-time monitoring using AI-powered surveillance systems.
* **Environmental Changes:** Analyze satellite imagery and remote sensing data to detect habitat changes and deforestation within national parks. Use this data to assess the impact of environmental changes on endangered species.

**5. Data Visualization and Reporting**

* **Interactive Maps:** Create interactive maps to visualize species distributions, habitats, and threats. Integrate data from the National Park locations, Animals API, and Species.csv to provide a comprehensive view.
* **Dashboards:** Develop dashboards to display key metrics and trends related to endangered species protection. Use data visualization tools to present insights and recommendations to conservationists and policymakers.

By utilizing these analysis methods, the project can effectively integrate data from various sources to monitor, protect, and conserve endangered species within national parks and their surrounding areas. If you have any specific requirements or need further details, feel free to let me know!

**Ethical Considerations:**

**Data Bias and Accuracy:** Inaccurate or biased data can lead to incorrect predictions, potentially diverting resources away from species in critical need or misinforming conservation strategies.

**Privacy Concerns:** Some datasets may include sensitive locations of endangered species. If this information is not handled correctly, it could lead to increased poaching or habitat destruction.

**Misuse of Predictions:** Predictions could be misinterpreted or misused by stakeholders, leading to harmful decisions or policies against the very species they aim to protect.

**Dependence on Technology:** Over-reliance on machine learning predictions could overshadow traditional conservation knowledge and practices, which are also crucial for species preservation.

**Resource Allocation:** Machine learning models might prioritize some species over others, influencing funding and conservation efforts and potentially neglecting species not deemed "important" by the model's criteria.

**Challenges/Issues:**

When working with data for endangered species in the context of machine learning several concerns and challenges may arise:

**Data Availability**: API data on endangered species can be scarce or incomplete due to the limited number of individuals and the difficulty in tracking and monitoring these species in their natural habitats. After running the API for a few animals, not every output provides all the information required in the final dataset.

**Temporal and Spatial Variability**: Endangered species data can be highly variable over time and space, requiring complex models to understand patterns of behavior, migration, and population dynamics.

**Impact of Climate Change**: The effects of climate change on species habitats, migration patterns, and population dynamics add another layer of complexity to modeling efforts.

**Integration with Conservation Strategies**: The ultimate goal is to aid conservation efforts, which requires models not only to predict but also to prescribe actionable strategies. This requires a deep understanding of the ecological, social, and economic

**References:**

Here are some additional sources that can be used to validate results and support the Wildlife Protection of Endangered Species project using AI/ML:

1. **Scientific Journals and Publications**
   * Sources: Journals such as "Conservation Biology," "Biodiversity and Conservation," and "Ecological Applications."
   * Description: These journals publish peer-reviewed research articles on wildlife conservation, habitat management, and the application of AI/ML in ecological studies. They provide credible scientific evidence and methodologies to support the project.
2. **International Union for Conservation of Nature (IUCN) Red List**
   * Source: IUCN Red List of Threatened Species
   * Description: The IUCN Red List is a comprehensive resource that provides information on the conservation status of species worldwide. It includes data on species populations, threats, and conservation measures, which can be used to validate species-specific results.
   * Website: IUCN Red List
3. **World Wildlife Fund (WWF) Reports**
   * Source: WWF
   * Description: WWF publishes reports and white papers on various aspects of wildlife conservation, including endangered species protection, habitat preservation, and the impact of climate change. These reports offer valuable insights and real-world case studies.
   * Website: World Wildlife Fund
4. **United Nations Environment Programme (UNEP)**
   * Source: UNEP
   * Description: UNEP provides reports, data, and policy recommendations on global environmental issues, including biodiversity and endangered species conservation. Their publications can help validate the project's methodologies and findings.
   * Website: UNEP
5. **Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES)**
   * Source: CITES
   * Description: CITES maintains a database of species that are regulated under international trade agreements. It provides information on species protection measures and trade regulations, which can be used to support conservation efforts.
   * Website: CITES
6. **National Geographic**
   * Source: National Geographic
   * Description: National Geographic publishes articles, documentaries, and research on wildlife conservation and endangered species. Their content often includes expert insights and field research that can complement the project's data and analysis.
   * Website: National Geographic

These sources offer credible and comprehensive information that can validate the results and support the Wildlife Protection of Endangered Species project using AI/ML.