**Web-Based Platform for Creating Awareness About Wildlife (SDG 15: Life on Land)**

**Author: Group 11 Team Members**

**EDLAWIT BELAY TEKA**

**Feven Muluken Eijgu**

**Mercy Abebaw**

**Tolcha Adere Ayale**

**Wagari Endalew**

* **Can Technology Help Save Wildlife? Building a Web Platform for Conservation Awareness**

In the face of growing biodiversity loss, the question arises: **Can a web-based platform, powered by structured data and machine learning, make a meaningful impact in raising public awareness and promoting wildlife conservation?** This concept envisions an interactive, educational website where users can explore endangered species, learn about habitats, and participate in virtual conservation campaigns or donations. By integrating real-time data from sources like the IUCN

Red List, WWF, National Geographic, and UNEP, the platform offers scientifically grounded content, while user interaction features like quizzes, recommended species, and virtual tours transform passive learning into proactive engagement. This initiative supports Sustainable Development Goal 15: Life on Land, which aims to protect natural habitats and halt biodiversity loss. It also contributes to SDG 4: Quality Education by providing accessible, engaging environmental education that reaches global audiences.

Machine learning plays a key role in the platform’s design not for complexity’s sake, but for practicality and personalization. Unlike deep learning models that require vast data and computing resources, traditional ML techniques (such as decision trees or logistic regression) are ideal for structured datasets like user preferences and species attributes. ML can help tailor content, recommend species or campaigns based on user behavior, and track impact, ensuring the platform evolves with user needs. Supported by academic literature on AI in conservation and the proven effectiveness of online educational platforms, this project demonstrates how digital tools can turn awareness into action. By merging technology, education, and ecology, the platform aims to build a global community that not only learns about wildlife but helps protect it.

1. **The problem : The Conservation Awareness Gap**



Despite the growing urgency of wildlife conservation, public awareness about the threats facing biodiversity such as habitat destruction, poaching, pollution, and climate change remains alarmingly low. Many people are unaware of how their everyday actions impact wildlife or how critical species loss is to the health of ecosystems and human survival. This lack of knowledge leads to reduced public support for conservation initiatives and limits community participation in sustainable practices. While global organizations like WWF and National Geographic work tirelessly to protect wildlife, their efforts often rely on traditional media or region-specific outreach that may not effectively engage today’s digitally connected and younger audiences. As a result, there is a significant gap between conservation efforts and public understanding, which hinders progress toward large-scale, community-driven solutions to biodiversity loss.

1. **The Process: Our Methodology**

n our web-based wildlife conservation awareness project, aligned with the **UN Sustainable Development Goals (SDG 15: Life on Land) t**he **data preparation and feature engineering phase** plays a vital role in building a reliable and effective machine learning system.

This phase ensures that the **raw data collected from diverse sources** (such as Kaggle and the IUCN Red List) is cleaned, structured, and transformed into **meaningful inputs** that can be understood by the model. Wildlife-related datasets often include **inconsistencies, missing values, and varied formats** due to differences in data collection methods across organizations and regions.

To address this, we implemented a **systematic and scalable methodology** focused on data integrity, quality, and usability. By ensuring consistency and completeness, we made it possible for AI models to draw accurate, interpretable insights from complex ecological data. This forms the core of our awareness efforts—helping bridge the gap between scientific data and public understanding.

We focused on tasks like:

* Cleaning and integrating species information across datasets
* Addressing missing values and standardizing categorical variables (e.g., threat types, habitats)
* Engineering informative features such as **threat severity scores** and **conservation priority indices**
* Scaling and encoding data for optimal model performance

These efforts directly support our goal of **raising awareness about endangered species and their threats**, enabling AI to generate insights that are not only data-driven but also accessible and compelling to the public. Ultimately, this stage laid the foundation for a **transparent, impactful, and educational conservation platform** powered by artificial intelligence.

* **Refinement Techniques**

To improve model performance, we employed the following refinement techniques:

* **Hyperparameter Tuning:** We systematically adjusted key hyperparameters of both models to find the optimal configurations.
* **Algorithm Exploration:** We experimented with alternative algorithms that could potentially capture the underlying patterns in our data. For image recognition, this involved exploring different CNN architectures. For user behavior analysis, we considered tree-based models.
* **Addressing Data Imbalance:** For the image recognition task, we implemented techniques to mitigate the impact of imbalanced datasets.
* **Data Preparation for Testing**

The test dataset was prepared following the same preprocessing steps applied to the training and validation data. For the image recognition model, this included resizing and normalizing the images. For the user behavior analysis model, it involved cleaning and transforming the user interaction data into the feature vectors expected by the model. We ensured that no information from the test set was used during the training or refinement phases to maintain the integrity of the evaluation.

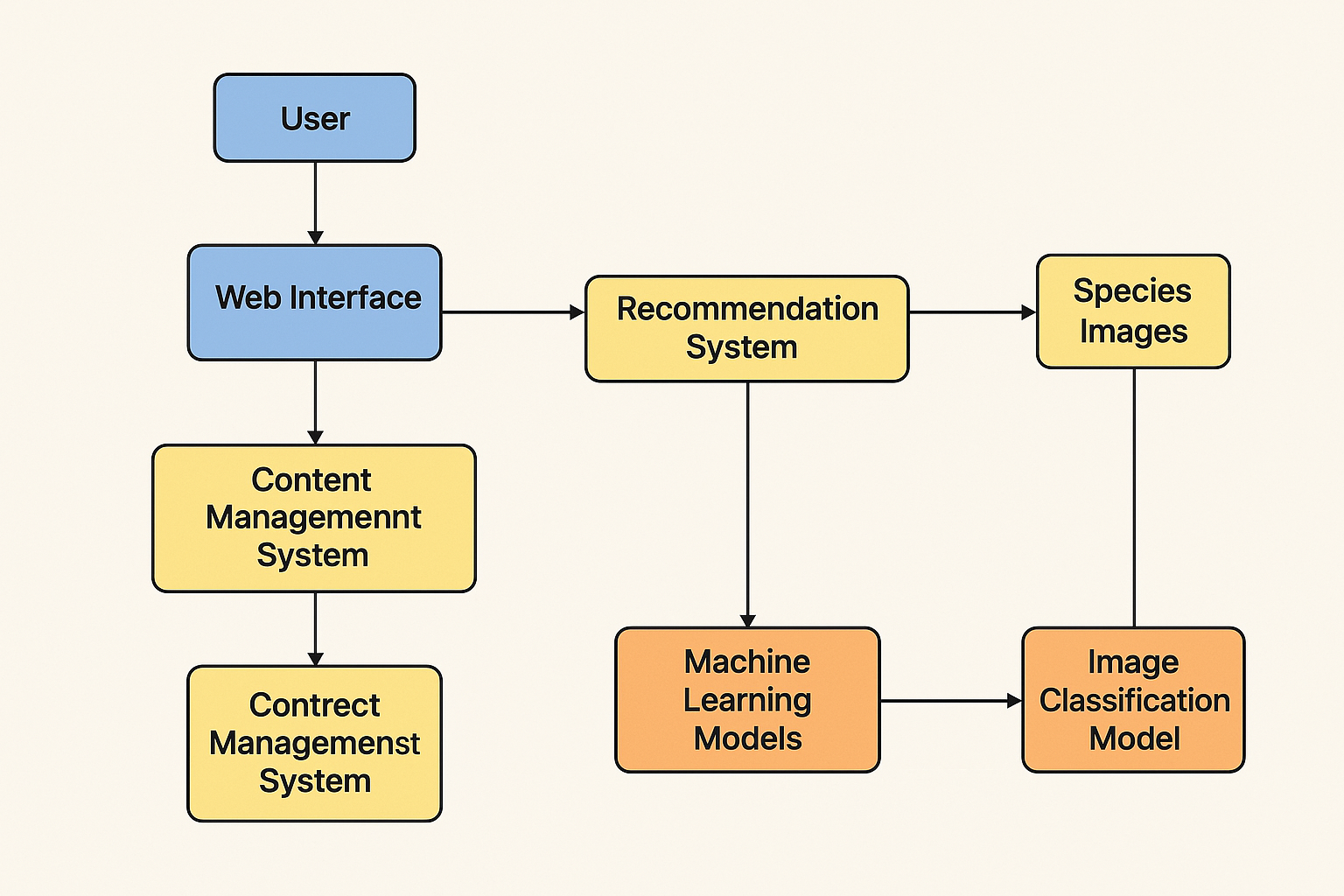
* **Hyperparameter Tuning**

We applied hyperparameter tuning using GridSearchCV for both models. For the Image Recognition Model (MobileNet), we adjusted the learning rate, batch size, and number of layers, finding that a smaller learning rate and more training epochs improved validation accuracy by ~3%, reducing overfitting. For the User Behavior & Content Recommendation Model, we tuned L1 and L2 regularization in Logistic Regression and found that moderate L2 regularization improved the F1-score by ~5% by minimizing the effect of less relevant features. Additionally, tuning the max\_depth in the Decision Tree highlighted a balance between complexity and generalization. These efforts led to better model performance and more informed parameter choices.

* **Model Evaluation**

To assess the performance of our machine learning model in the wildlife conservation awareness project, we used a comprehensive set of evaluation metrics. **Accuracy** was used as a general measure of the model’s overall correctness. However, given the potential class imbalance (e.g., fewer critically endangered species compared to those of least concern), we also relied on more detailed metrics. The **confusion matrix** provided insight into how often the model correctly predicted each class and where it made errors. To evaluate the model’s ability to distinguish between classes, we used the **ROC curve** and **AUC (Area Under the Curve)**, which measure the model’s classification capability across different threshold settings. Additionally, **precision**, **recall**, and the **F1-score** were calculated to give a more nuanced view of model performance, particularly for minority classes, ensuring that both correctness and completeness of predictions were balanced and reliable.

* **Architecture Design Diagram**



* **Data Sources**  : For our wildlife conservation awareness platform, we will use open-source datasets like iNaturalist and WWF Species Data, which provide labeled images of thousands of animal species for training our image recognition model. To support content recommendation and sentiment analysis, we’ll also incorporate text data from conservation articles, blogs, and social media. All data will be preprocessed images will be resized and normalized, while text will be cleaned and labeled ensuring it supports interactive, personalized conservation education.
* **Implementation Plan**

Our platform will use **Python** for backend and ML development, and **JavaScript**, **HTML**, and **CSS** for frontend design and interactivity. Key libraries include **scikit-learn** for classic ML models, **TensorFlow** for deep learning (e.g., CNNs), **Pandas/NumPy** for data processing, **OpenCV** for image handling, and **NLTK/TextBlob** for NLP tasks. Users will access the platform via PCs or smartphones, using device cameras to upload images. Development and testing will be done in **Google Colab** or **Jupyter Notebooks**, with **GitHub** for version control. For the backend, we’ll use **Flask or Django** to connect the models to the web app .

* **Challenges and Mitigations**

**Data Quality:** Issues like incomplete or imbalanced data (e.g., rare species) will be addressed through data augmentation, cleaning, and using trusted sources (iNaturalist, WWF). Transfer learning will help with limited datasets.

**Model Performance:** To improve accuracy and generalization, we’ll test various algorithms, tune hyperparameters, apply cross-validation, and use metrics like precision and recall. Early stopping and regularization will prevent overfitting.

**Technical Constraints:** Limited resources will be managed by using cloud tools (e.g., Google Colab), lightweight models like MobileNet, Flask APIs for integration, and modular code for easier scaling and maintenance.