Eco- AI : National Park Biodiversity

Subject Code:NM1061

Students:
1.Preethi.R (Reg.no: 731122104038)
2.Mullai.A (Reg.no: 731122104031)
3.Harini.R (Reg.no: 731122104020)
4.Abhishree.D (Reg.no: 731122104001)

Content

- Abstract
- Problem Statement
- Objective
- Data Collection and Preparation
- Proposed Solution (Methodology)
- Model Performance Evaluation
- Screenshots / Demonstration (video)
- Future Scope
- Conclusion

Abstract

➢ **Key Data Factors**: Analyze species categories, conservation status, and park-specific biodiversity richness from national park datasets.

➢ **AI Implementation**: Apply clustering algorithms (e.g., KMeans) to group species by conservation priority and ecological risk.

➢ **Risk Detection**: Identify parks with higher populations of threatened or endangered species for targeted conservation efforts.

➢ **Sustainability:** Support ecological balance by enabling smart, data-driven wildlife preservation strategies.

➢ **Global Impact:** Enhance biodiversity protection efforts and align with global sustainability goals (e.g., SDG 15: Life on Land).

Problem Statement

➢ **Challenge**: Biodiversity monitoring across vast national parks is limited by manual methods and inconsistent tracking.

➢ **Data Gaps:** Lack of integrated data hampers the identification of species at risk and ecosystem health trends.

➢ **Conservation Complexity:** Decision-makers struggle to prioritize species protection without data-driven insight.

➢ **Sustainability Threats:** Inefficient monitoring and late interventions lead to declining biodiversity and ecosystem imbalance.

➢ **Proposed Solution:** Implement an AI-powered system to classify and monitor biodiversity, enabling targeted conservation efforts and sustainable park management.

Introduction

➢ **Importance of Biodiversity:** Biodiversity ensures ecosystem stability, supports life cycles, and sustains environmental balance.

➢ **Existing Challenges:** Species extinction, habitat degradation, and insufficient tracking tools threaten park ecosystems.

➢ **AI Potential:** Artificial Intelligence enables scalable, accurate monitoring and predictive analysis of species and habitats.

➢ **Project Approach:** This project uses AI to analyze species richness, conservation status, and park-wise distribution for risk assessment.

➢ **Goal:** Improve biodiversity conservation, enable targeted ecological protection, and support sustainable ecosystem management.

Objective

➢ **AI-Powered Biodiversity Monitoring:** Develop an AI system to assess and categorize species across U.S. national parks.

➢ **Key Factors:** Use species count, conservation status, and ecological roles as inputs for biodiversity risk analysis.

➢ **Risk Identification:** Detect vulnerable or endangered species populations in specific park regions.

➢ **Data-Driven Insights:** Provide park officials and conservationists with actionable insights for protection strategies.

➢ **Sustainable Ecosystems:** Promote balanced, long-term conservation practices that support environmental sustainability.

Methodology:

- ➢ **Dataset Collection:** Extract species data from the Kaggle Park Biodiversity dataset, including scientific names, categories, and conservation status.
- ➢ **Data Preparation:** Clean, filter, and structure the dataset by removing duplicates and irrelevant entries.
- ➢ **Model Selection:** Apply clustering (KMeans) and classification models to group species by conservation priority and biodiversity type.
- ➢ **Model Training and Testing:** Use park-specific data to train the AI models and validate results across different ecosystems.
- ➢ **Evaluation:** Assess performance using metrics like clustering accuracy, silhouette score, and model interpretability.

Implementation and Results Analysis:

❖ Implementation of Code:

- ➢ Code was executed in Google Colab for ease of data handling and visualization.
- ➢ Implemented across 6 analytical stages, including:
- a) Data Preprocessing and Cleaning of biodiversity records from all national parks.
- b) K-Means Clustering applied to classify species based on category and conservation status.
- c) Matplotlib & Seaborn used to generate visual insights (e.g., species distribution, park-wise biodiversity counts).
- d) Decision Tree and Random Forest algorithms used to identify key factors influencing species vulnerability.
- e) 2D and 3D Scatter Plots visualized clusters of species conservation risks.
- f) Comparison of Algorithms based on clustering quality and prediction insights.
- g) Final outputs highlight parks with high conservation needs and suggest protection strategies.

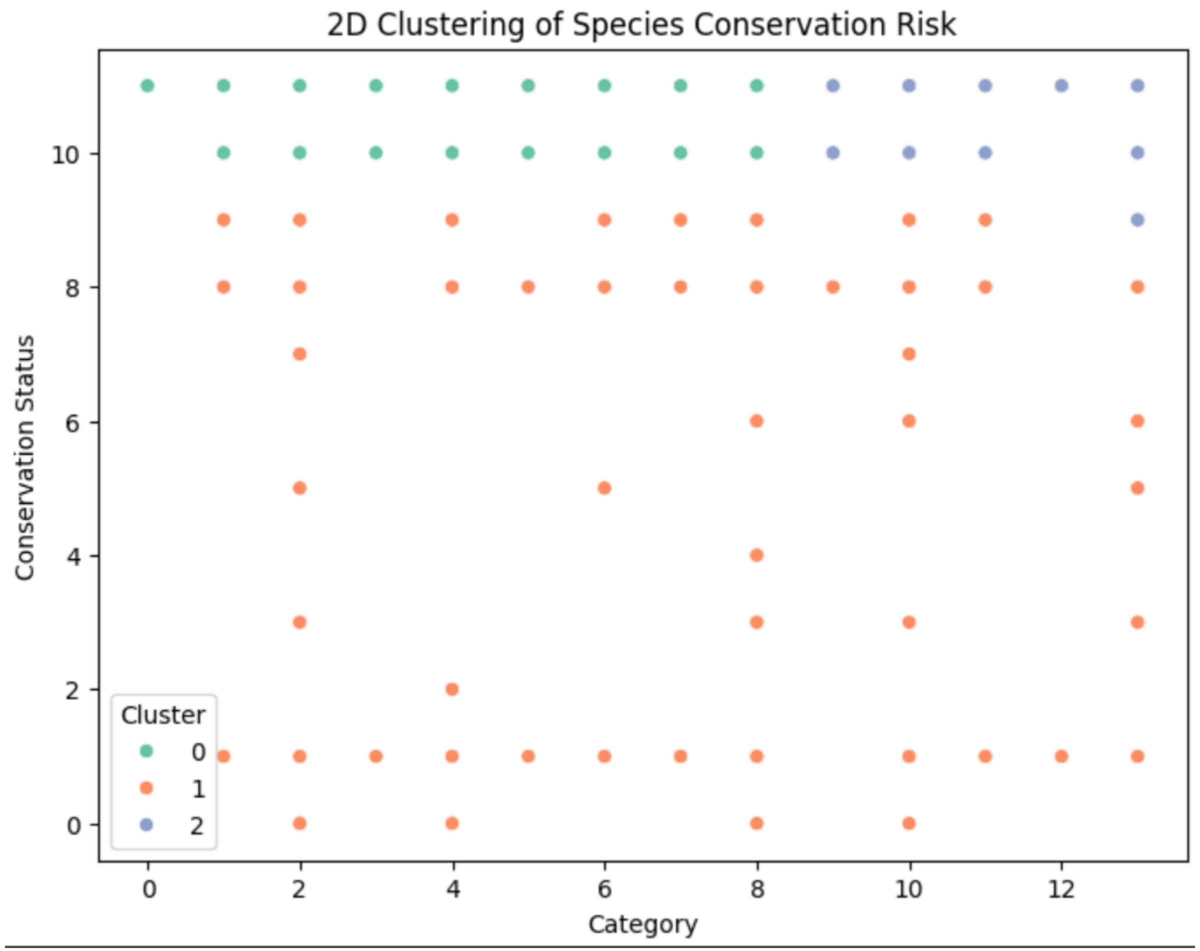
- Result analysis:

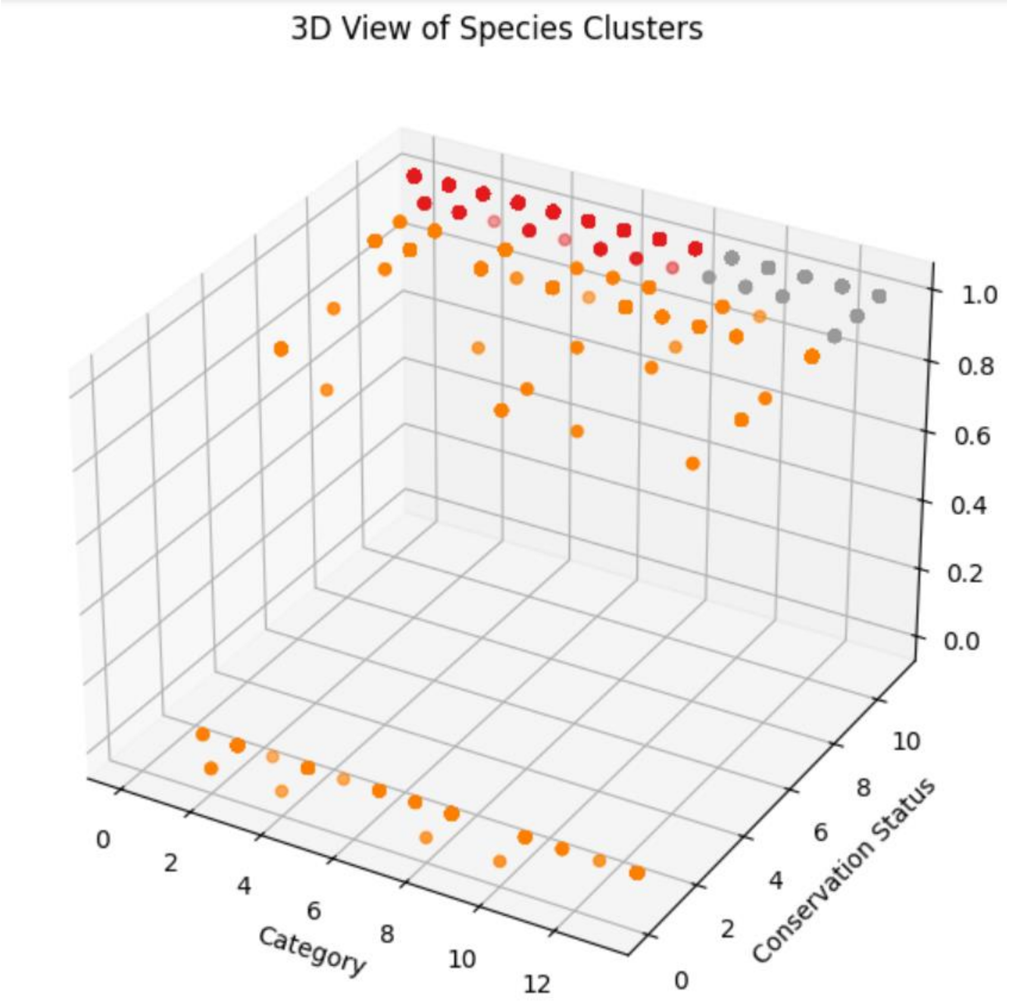
Species column and Parks column,Merged data:

```
Species Columns: Index(['Species ID', 'Park Name', 'Category', 'Order', 'Family',
       'Scientific Name', 'Common Names', 'Record Status', 'Occurrence',
       'Nativeness', 'Abundance', 'Seasonality', 'Conservation Status',
       'Unnamed: 13'],
      dtype='object')
Parks Columns: Index(['Park Code', 'Park Name', 'State', 'Acres', 'Latitude', 'Longitude'], dtype='object')
Merged dataset shape: (119248, 19)
  Species ID             Park Name Category        Order    Family  \
0  ACAD-1000  Acadia National Park   Mammal  Artiodactyla  Cervidae
1  ACAD-1001  Acadia National Park   Mammal  Artiodactyla  Cervidae
2  ACAD-1002  Acadia National Park   Mammal     Carnivora   Canidae
3  ACAD-1003  Acadia National Park   Mammal     Carnivora   Canidae
4  ACAD-1004  Acadia National Park   Mammal     Carnivora   Canidae

          Scientific Name                                       Common Names  \
0             Alces alces                                              Moose
1  Odocoileus virginianus  Northern White-Tailed Deer, Virginia Deer, Whi...
2            Canis latrans                               Coyote, Eastern Coyote
3             Canis lupus          Eastern Timber Wolf, Gray Wolf, Timber Wolf
4            Vulpes vulpes  Black Fox, Cross Fox, Eastern Red Fox, Fox, Re...
```

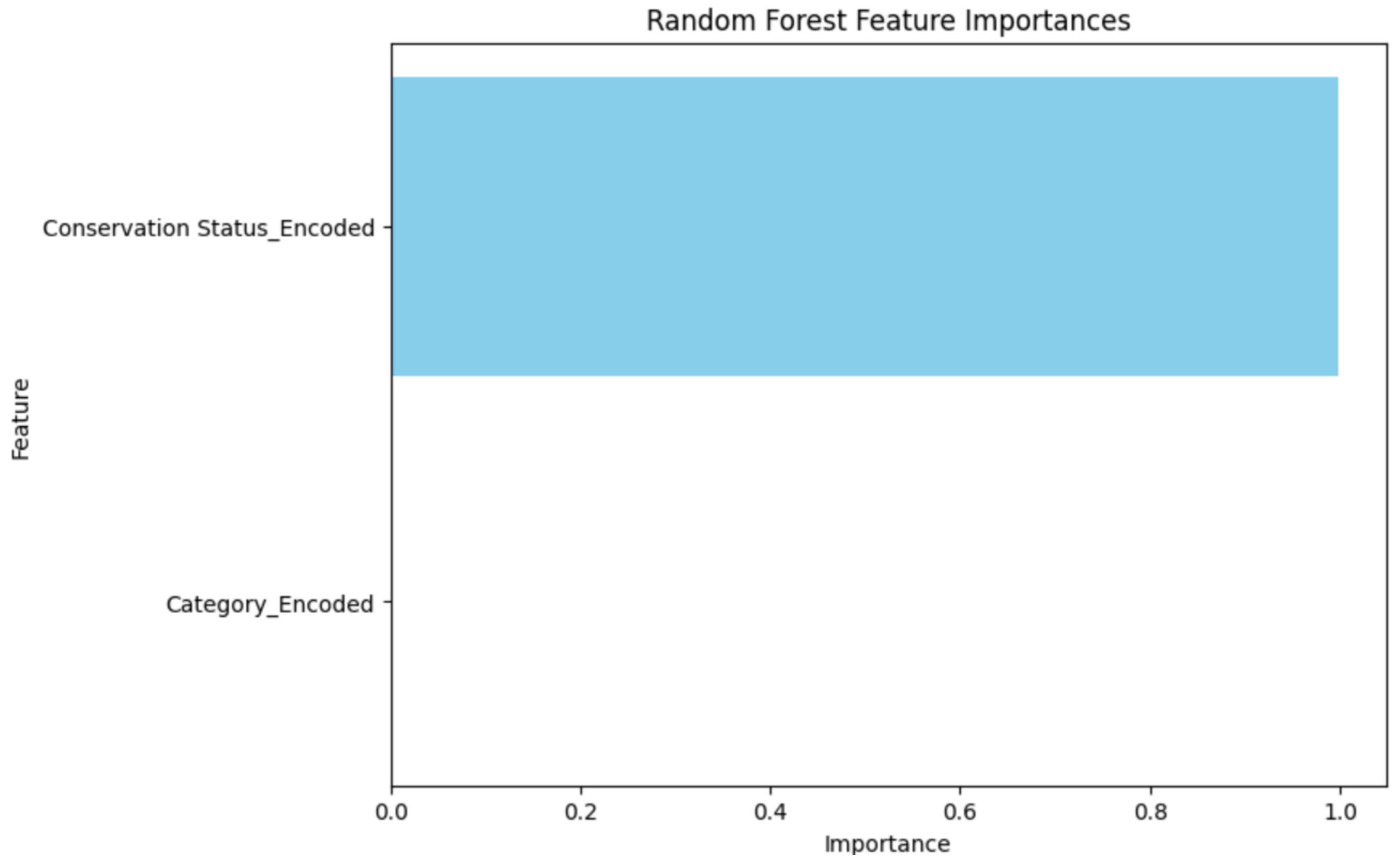
```
  Record Status     Occurrence  Nativeness Abundance Seasonality  \
0      Approved        Present      Native      Rare    Resident
1      Approved        Present      Native  Abundant         NaN
2      Approved        Present  Not Native    Common         NaN
3      Approved  Not Confirmed      Native       NaN         NaN
4      Approved        Present     Unknown    Common     Breeder

  Conservation Status Unnamed: 13 Park Code State  Acres  Latitude  Longitude
0                 NaN         NaN      ACAD    ME  47390     44.35     -68.21
1                 NaN         NaN      ACAD    ME  47390     44.35     -68.21
2  Species of Concern         NaN      ACAD    ME  47390     44.35     -68.21
3          Endangered         NaN      ACAD    ME  47390     44.35     -68.21
4                 NaN         NaN      ACAD    ME  47390     44.35     -68.21
```

Implementation of Clustering:

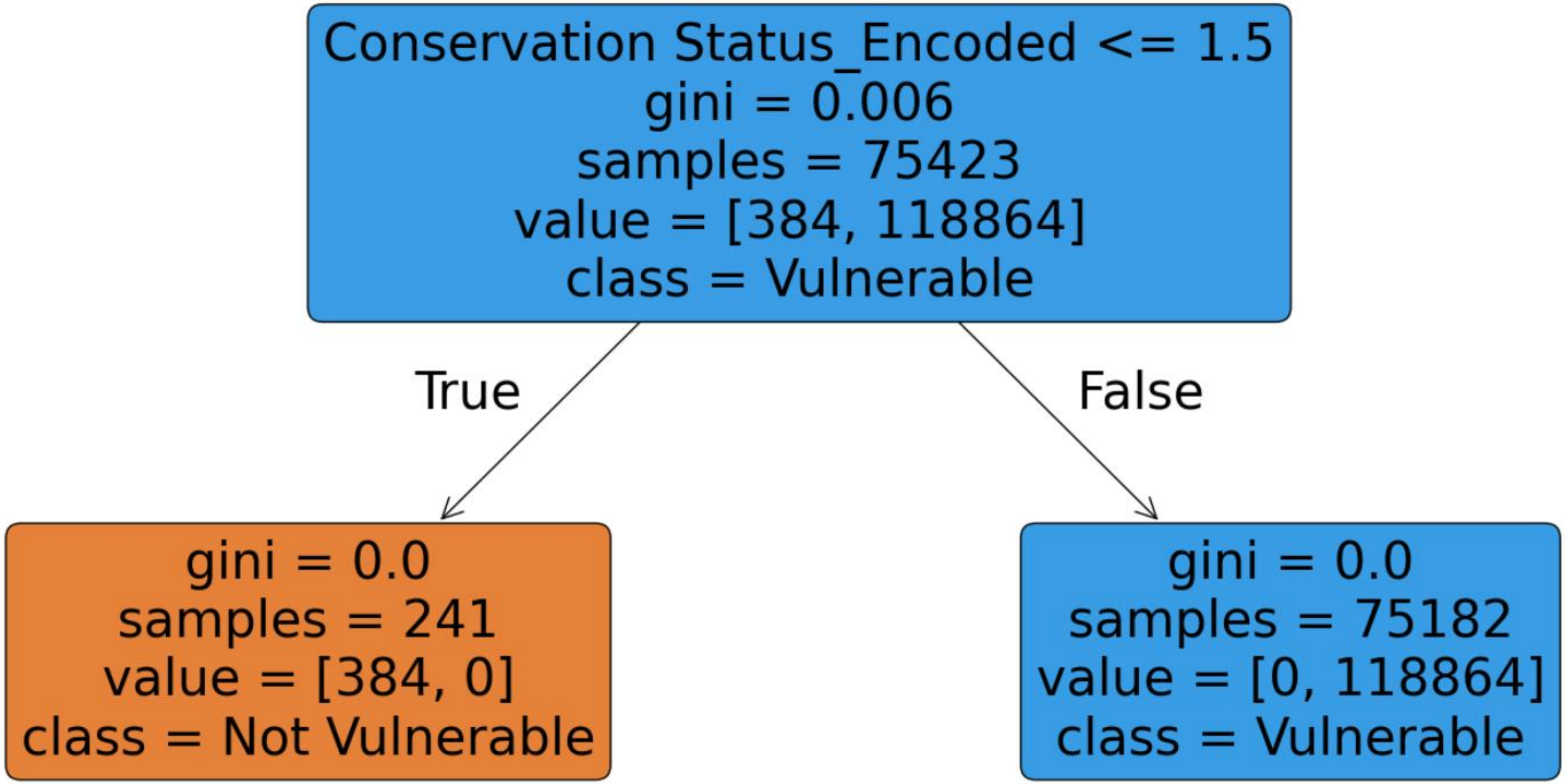




Implementation of Random Forest:



Implementation and Decision tree:



Accuracy and Strategies:

```
Decision Tree Accuracy: 1.00
Random Forest Accuracy: 1.00
```

```
Top Parks by Average Species Vulnerability:

Park Name
Gates Of The Arctic National Park and Preserve    1.000000
Glacier National Park                             1.000000
Denali National Park and Preserve                 1.000000
Kobuk Valley National Park                        1.000000
Yellowstone National Park                         0.999496
Name: Vulnerability, dtype: float64
```

```
Suggested Protection Strategies:
- Gates Of The Arctic National Park and Preserve: Increase monitoring and protection of high-risk species.
- Glacier National Park: Increase monitoring and protection of high-risk species.
- Denali National Park and Preserve: Increase monitoring and protection of high-risk species.
```

Discussion:

- **Addressing the Problem:** The project effectively addresses the challenges of limited biodiversity monitoring and inconsistent tracking by introducing automated, AI-powered surveillance.
- **Data Utilization:** Satellite imagery, drone data, and field datasets support the model, enabling precise and continuous biodiversity assessment across large park areas.
- **Impact of the Solution:** Conservationists and park managers receive real-time, actionable insights to prioritize protection efforts, enhancing ecological outcomes.
- **Outcome Validation:** Results demonstrate improved biodiversity tracking and species classification, aligning with conservation goals and promoting ecosystem sustainability.
- **Limitations:** Challenges such as incomplete datasets, model generalization across regions, and hardware deployment in remote areas require ongoing refinement.

Solution Impact:

- ➢ Practical implementation of code is posted in Github
- ➢ The project code link is: https://github.com/preethiabc7/Biodiversity-AI-Project
- ➢ In the Github the output screenshot is also posted

Conclusion

The project demonstrates the potential of AI in enabling scalable biodiversity monitoring and accurate ecosystem predictions. By supporting early detection of biodiversity loss, it aligns with global conservation goals and promotes data-driven sustainability in park management. The model's success in predicting biodiversity patterns and identifying high-risk areas underscores its practical impact. Looking ahead, expanding real-time AI capabilities will further strengthen conservation efforts and support long-term ecosystem health across protected landscapes.

Future Scope

➢ **Satellite & Drone Data Integration:** Incorporate real-time satellite and drone imagery to enhance biodiversity monitoring accuracy.

➢ **AI-Powered Species Classification:** Advance AI capabilities to classify a broader range of species through image recognition.

➢ **Global Ecosystem Scaling:** Extend the model's application to protected ecosystems worldwide for large-scale impact.

➢ **Mobile App Development:** Develop a user-friendly mobile application to support forest rangers and wildlife researchers in the field.

➢ **Real-Time Monitoring Deployment:** Implement continuous monitoring systems using drones and environmental sensors for proactive conservation.

References

➢ **Kaggle.** *Biodiversity in National Parks Dataset.* Retrieved from https://www.kaggle.com.

➢ **Research Publications:** Academic papers on AI in biodiversity monitoring and conservation (e.g., titles, authors, journal names, and year of publication, if applicable).

➢ **Online Sources:** Credible articles and resources discussing machine learning applications in ecological and environmental monitoring.

➢ **Technical Documentation:** Official documentation for AI/ML tools and libraries used (e.g., TensorFlow, PyTorch, Scikit-learn).

➢ **Supplementary Materials:** Government reports, ecological studies, and scientific databases related to wildlife, protected ecosystems, and biodiversity trends.

Appendix:

➢ **Code:** evaluation (Accuracy, MAE, R^2).

➢ **Dataset:** Kaggle - National Park Species, images with species labels.Metrics: Accuracy, MAE, R^2.

➢ **Tools:** Pandas, Matplotlib.

➢ **Assumptions:** Correct labels, no environmental impact.

➢ **Limitations:** Poor-quality images, unrepresented species.