SUMMER RESEARCH INTERNSHIP REPORT

Submitted by

Shardul Gawade

20220802338

Sarthak Mahajan

20220802226

Sahil Mahandule

20220802221

Under the mentorship of

Dr. Jayashree Katti

Mrs. Veena Kulkarni

- ## PROJECT TITLE

Dual-AI Framework for Forest Loss Detection and Smart Replanting.

- ## PROBLEM STATEMENT

Illegal deforestation continues unnoticed, while afforestation efforts fail due to poor site and species selection. This project presents an AI-powered system that detects deforestation in real-time using satellite imagery and recommends ecologically suitable tree species for afforestation using environmental data analysis.

- ## INTRODUCTION

Forests are one of the most crucial natural resources, providing ecological balance, biodiversity support, and climate regulation. However, rapid deforestation due to urbanization, agriculture, and illegal logging has led to significant ecological degradation. Remote sensing and satellite imagery, combined with machine learning and AI, now offer an efficient and scalable way to monitor forest health and support replanting strategies. Leveraging these technologies, this project proposes an integrated system that not only detects deforested areas using NDVI-based .tif imagery but also suggests region-specific replanting options through smart data-driven recommendations.

- ## BACKGROUND

Several research efforts have explored deforestation detection using vegetation indices like NDVI and machine learning classifiers such as Random Forest and SVM. The combination of satellite image processing and supervised learning has shown promising results in identifying forest loss with pixel-level precision. However, fewer works have focused on combining this detection with a decision-making tool for afforestation, especially in a single end-to-end interface. Few studies have emphasized the importance of coupling deforestation detection with ecological restoration planning to ensure long-term sustainability. Moreover, deep learning models like Multilayer Perceptron (MLP) and ensemble techniques like Gradient Boosting have gained traction in agricultural and environmental data analysis.

- **OBJECTIVES**

1. To detect and quantify deforestation using NDVI-based satellite imagery and machine learning classifiers.
2. To evaluate and compare different Machine Learning and Deep Learning algorithms.
3. To develop a user-friendly and explainable Streamlit-based dashboard for deforestation visualization.
4. To build a smart replanting model that recommends suitable tree or crop groups based on soil, NDVI, rainfall, and region
5. To create a scalable and automated AI framework for environmental monitoring and ecological restoration.

- **ABSTRACT**

Environmental degradation due to deforestation has become a critical global concern, demanding intelligent and automated monitoring systems. This research presents an integrated framework that combines remote sensing, machine learning, and explainable AI to detect deforested regions using Sentinel-2 NDVI satellite imagery and recommend region-specific replanting strategies.

The first phase involves preprocessing multi-band .tif satellite images and extracting relevant vegetation indices to identify deforested pixels. Various classical machine learning models—Random Forest, SVM, Logistic Regression, and XGBoost—were evaluated for pixel-level classification, where Support Vector Machine (RBF Kernel) demonstrated optimal performance with a balanced F1-score of 0.71, outperforming others without overfitting.

In the second phase, a smart replanting model was developed using decision trees, Naive Bayes, gradient boosting, and a multilayer perceptron (MLP) to predict suitable tree groups based on environmental parameters such as NDVI, rainfall, soil type, and region. Gradient boosting achieved the highest accuracy (0.88) in recommending tree types like Mangrove, Teak, or DroughtResistant based on ecological viability. A user-friendly Streamlit interface integrates both modules, offering visual insights and suggestions for conservation planning.

The proposed system demonstrates a scalable approach for forest monitoring and ecological restoration, with the potential to assist environmental agencies in making data-driven decisions.

CHAPTER 1 - INTRODUCTION

• IMPORTANCE OF FOREST

Forests are essential ecosystems that contribute significantly to environmental health and human well-being. They regulate the Earth's climate by absorbing carbon dioxide, preserve biodiversity by providing habitat for countless species, and maintain the water cycle. Additionally, forests protect against soil erosion, purify air and water, and support the livelihoods of indigenous communities and rural populations. Their ecological, economic, and cultural importance makes forest conservation a global priority amid rising environmental concerns.

• MOTIVATION FOR REFORESTATION

The alarming rate of deforestation has intensified the need for effective and sustainable reforestation strategies. Reforestation helps restore degraded ecosystems, enhances carbon sequestration, and combats the adverse effects of climate change. Motivated by this, researchers aim to apply intelligent systems to improve replanting decisions. Smart reforestation—guided by data analytics, environmental indicators, and machine learning models—ensures that the right species are planted in suitable areas, maximizing ecological compatibility and long-term forest recovery.

• REASONS FOR DEFORESTATION

Deforestation is driven by various human and natural factors, including agricultural expansion, illegal logging, mining, infrastructure development, and forest fires. Population growth and economic demands often prioritize land use over forest conservation. In many regions, weak enforcement of environmental regulations accelerates forest degradation. These activities result in biodiversity loss, altered climate patterns, and reduced ecosystem services. Understanding the root causes is essential to developing effective detection, prevention, and reforestation mechanisms.

CHAPTER 2 – LITERATURE SURVEY

• LITERATURE REVIEW

Several studies have addressed deforestation detection using satellite imagery and spectral indices like NDVI. Deep learning models such as CNNs and UNet have been explored for pixel-wise classification, while traditional machine learning algorithms (Random Forest, XGBoost) are used for forest/non-forest classification based on vegetation indices. In parallel, crop recommendation systems have utilized soil and climate features to train Decision Trees and Gradient Boosting classifiers for optimal replanting choices. Some papers propose cloud-based or edge deployment using lightweight platforms like Streamlit or Flask, making ML models accessible to stakeholders.

• DRAWBACKS OF EXISTING SYSTEMS

1. Many existing works focus only on detection and do not suggest region-specific replantation strategies.
2. Few systems integrate smart replanting decisions with forest loss analysis in a single framework.
3. Deep learning methods, although accurate, are often non-interpretable and require large datasets and GPU resources.
4. Manual replanting planning is time-consuming, static, and lacks adaptability to real-time environmental changes.
5. Prior models don't visualize predictions or allow interactive user input in web-based environments.

• GAPS IDENTIFIED

1. Lack of integrated frameworks that combine deforestation detection with replantation recommendations using AI/ML.
2. Existing solutions do not provide explainable, region-specific, or environment-aware crop/tree suggestions based on data like NDVI, rainfall, soil type, and geography.
3. Limited accessibility of current systems – they are not available in interactive platforms like Streamlit for practical deployment.

• OBJECTIVES

To bridge these gaps, our objective is to develop an end-to-end AI-based environmental monitoring platform that:

1. Detects deforested pixels using NDVI-based satellite images.
2. Compares performance of multiple ML algorithms for deforestation detection and smart replanting.
3. Suggests optimal replanting options tailored to local environmental data.
4. Provides a visual, explainable, and easy-to-use interface for stakeholders.

CHAPTER 3 – TECHNOLOGY STACK

1. Sentinel-2 Satellite Imagery (NDVI bands), Soil & Crop Datasets (ICAR, FAO, IMD, IUCN): Raw data collection for vegetation, deforestation, and crop suitability.
2. Python, NumPy, Rasterio, Pandas: Image reading, cleaning, and transformation of .tif NDVI files.
3. ML and DL algorithms like Random Forest, SVM (RBF), Logistic Regression, XGBoost, Naive Bayes, Decision Tree, Gradient Boosting, MLP: Classification of deforested pixels and smart crop replanting decisions.
4. Scikit-learn, Imbalanced-learn (SMOTE), Joblib, Tensorflow, keras: Model training, class balancing, and model serialization.
5. Matplotlib, Seaborn, Streamlit Charts (Bar & Pie): Visual output of results and user insights.
6. Streamlit: Web application for forest monitoring and smart replanting recommendations.
7. Google colab, Google Drive: Compiling training of models, Saving satellite images and models.

CHAPTER 4 – PROPOSED METHODOLOGY

Our project involves a two-stage pipeline designed to detect deforestation from satellite images and recommend region-specific replantation strategies using machine learning. The model begins with the extraction of NDVI-based data from .tif satellite images, where we compute NDVI before and after deforestation, as well as the NDVI difference and label band indicating deforested pixels.

The extracted pixel-wise data is then passed through various classical machine learning algorithms such as Random Forest, Support Vector Machine (SVM) with RBF, Logistic Regression, and XGBoost to classify each pixel as forested or deforested. Among these, SVM showed high accuracy w.r.t F1 Score which shows the balance between Precision and Recall.

Once the deforested regions are identified, the second stage of the model focuses on smart replanting recommendations. A new dataset was created manually using various authentic sources, combining soil type, NDVI, rainfall, and region-specific environmental parameters. Algorithms like Decision Tree, Naive Bayes, Gradient Boosting, and Multilayer Perceptron (MLP) were trained on this dataset to predict the best-suited tree or crop group for that region, such as Mangrove, Drought-Resistant, Hardwood, etc.

Finally, both the deforestation model and the smart replanting model were integrated into an interactive web application using Streamlit, enabling users to visualize deforested areas and receive intelligent replanting suggestions tailored to the regional environment.

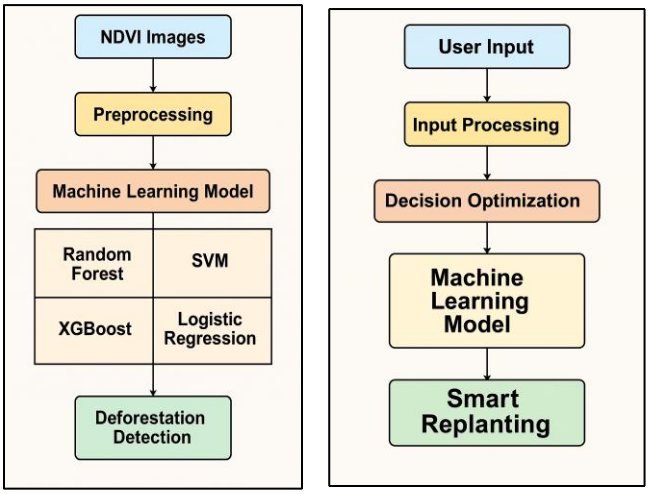
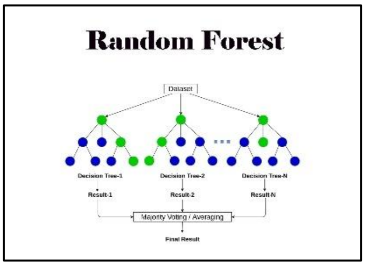


Fig. 3.9. shows the model working

- **ALGORITHMS USED FOR FOREST LOSS DETECTION**

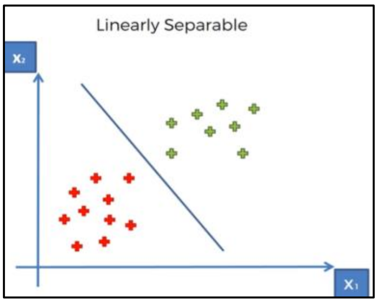
---

1. **Random Forest:** It is a ML algorithm used for classification and regression tasks. It works by building many decision trees and combining them to make a final prediction. Each tree is trained on a random subset of the data. For classification, it uses voting method, each tree votes a class (0 or 1= non-deforested vs deforested). Majority vote becomes final prediction.

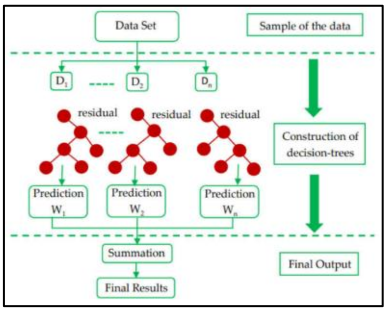


2. **Support Vector Machine (SVM) with RBF Kernel:** It is a ML model used for classification tasks. It separates forest and deforested pixels by finding the best hyperplane also known as decision boundary in the feature space (NDVI bands i.e. NDVI-

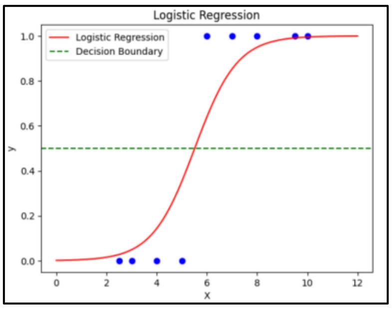
before, NDVI-after, NDVI-Difference and Labels). It is effective for small, high-dimensional datasets.



3. **XG Boost:** It is an advanced tree-based ML algorithm that learns from mistakes of previous trees and correct its errors. It improves accuracy in detecting deforestation using NDVI features. It detects deforested pixels by learning complex NDVI change patterns more accurately and handling noisy satellite data effectively.



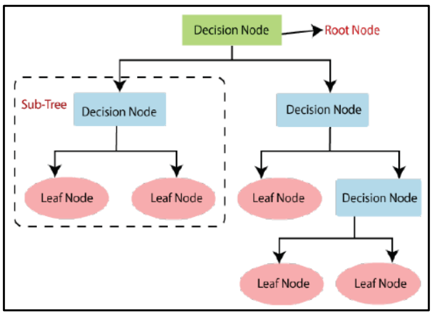
4. **Logistic Regression:** It is a simple ML algorithm used for binary classification. It uses logistic sigmoid function. It calculates the probability of deforestation for each pixel using NDVI data. It gives quick result and is easy to interpret.



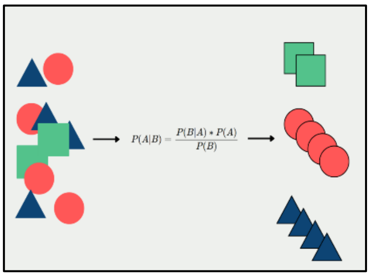
- ## ALGORITHMS USED FOR SMART REPLANTING

---

1) **Decision Tree:** A Decision Tree is a supervised learning algorithm that splits data based on feature values to arrive at a decision. It is used to classify which tree 6 species (like Sundari, Sal, or Teak) best suits a given region based on NDVI, rainfall, and soil type. It works like a flowchart, making decisions by splitting the data at each node using the most informative feature.

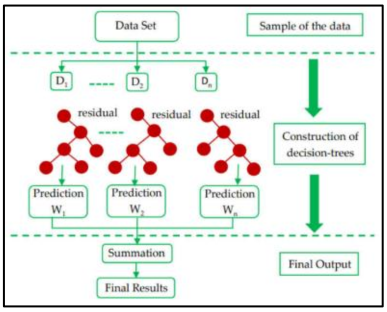


2) **Naive Bayes:** Naive Bayes is a classification algorithm based on probability. It assumes that all features like NDVI, rainfall, soil type are independent of each other. Here, it estimates the likelihood of each crop category based on the given environmental conditions and picks the one with the highest probability.

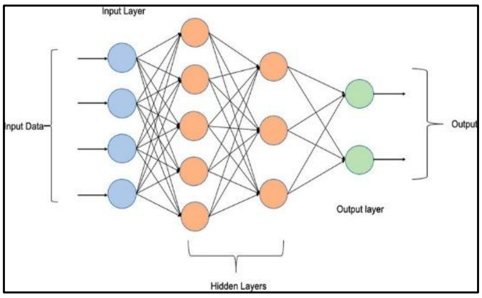


3) **Gradient Boosting:** Gradient Boosting is a classification algorithm that builds multiple decision trees in sequence, where each new tree fixes the errors made by the previous ones. Here, it

improves crop group predictions by capturing complex feature interactions across NDVI, rainfall, and soil data.



4) **Multi-Layer Perceptron Neural Network:** MLP (Multilayer Perceptron) is a deep learning algorithm that maps input features like NDVI, rainfall and soil to tree categories like Mangrove or DroughtResistant through multiple hidden layers. Unlike traditional machine learning algorithms, MLP can learn more complex patterns from data.



- **ADVANTAGES**

I. **Model Accuracy** - High classification accuracy for large datasets using Machine Learning and Deep Learning Algorithms.
II. **Interpretability** - Easy-to-understand predictions made by these algorithms.
III. **Scalability** - Easily scalable to new regions by updating .tif files and training datasets.
IV. **Practical Use** - Web interface allows non-technical users to upload images and get replanting suggestions.
V. **Automation -** Fully automated deforestation detection and ecological planning.

- **DISADVANTAGES**

I. **Model Accuracy -** Slight performance drop for minority classes (e.g., sparse deforested regions).
II. **Interpretability -** Deep models like MLP are harder to interpret.
III. **Scalability -** May require region-specific data preprocessing.
IV. **Practical Use -** Requires stable internet connection for Colab or cloud execution.
V. **Automation -** Limited to available environmental and vegetation data.

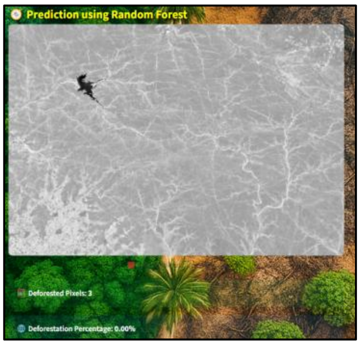
CHAPTER 4 – RESULTS

- ## DEFORESTATION

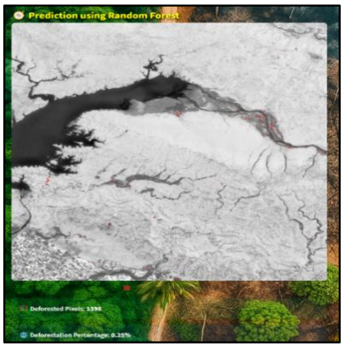
	Kaziranga	Gir Forest	Corbett	Sunderban	Satpura	Western Ghat	Gachibowli
Random Forest	1.03%	0.00%	0.25%	0.00%	0.01%	0.61%	0.01%
SVM (RBF)	70.44%	8.41%	77.27%	50.49%	7.52%	99.30%	11.19%
XG Boost	0.56%	0.41%	0.29%	0.40%	0.42%	0.69%	0.50%
Logistic Regression	48.56%	7.42%	11.10%	48.86%	10.45%	12.78%	90.30%

- ## Random Forest.

Kaziranga Forest



Gir Forest



Corbett Forest

Sunderban Forest

Satpura Forest

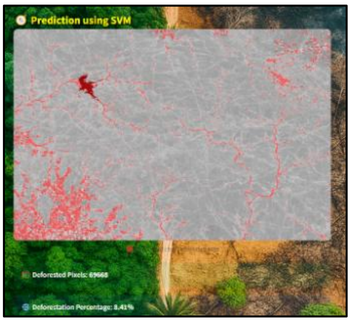
Western Ghat Forest

Gachibowli Forest

- Region Kaziranga Forest has 47,284 deforested pixels; hence Random Forest Model is predicting deforestation of 1.03%.
- Region Gir Forest has 3 deforested pixels; hence Random Forest Model is predicting deforestation of 0.00%.
- Region Corbett Forest has 1398 deforested pixels; hence Random Forest Model is predicting deforestation of 0.25%.
- Region Sunderban Forest has 2 deforested pixels; hence Random Forest Model is predicting deforestation of 0.00%.
- Region Satpura Forest has 65 deforested pixels; hence Random Forest Model is predicting deforestation of 0.01%.
- Region WesternGhats Forest has 2865 deforested pixels; hence Random Forest Model is predicting deforestation of 0.61%.
- Region Gachibowli Forest has 6 deforested pixels; hence Random Forest Model is predicting deforestation of 0.01%.

- **SVM with RBF Kernel.**

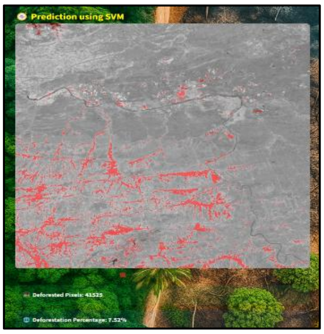
Kaziranga Forest



Gir Forest

Corbett Forest

Sunderban Forest



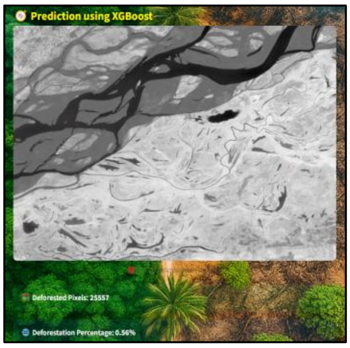
Satpura Forest

Western Ghat Forest

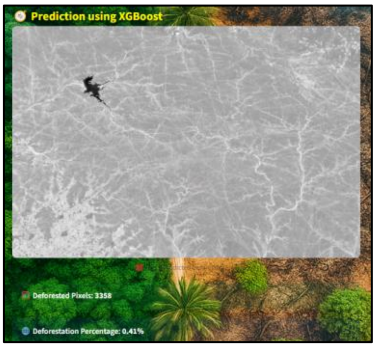
Gachibowli Forest

- Region Kaziranga Forest has 47,284 deforested pixels; hence SVM Model is predicting deforestation of 1.03%.
- Region Gir Forest has 69,668 deforested pixels; hence SVM Model is predicting deforestation of 8.41%.
- Region Corbett Forest has 4,26,554 deforested pixels; hence SVM Model is predicting deforestation of 77.27%.
- Region Sunderban Forest has 2,78,755 deforested pixels; hence SVM Model is predicting deforestation of 50.49%.
- Region Satpura Forest has 41,525 deforested pixels; hence SVM Model is predicting deforestation of 7.52%.
- Region WesternGhats Forest has 4,66,284 deforested pixels; hence SVM Model is predicting deforestation of 99.30%.
- Region Gachibowli Forest has 5,590 deforested pixels; hence SVM Model is predicting deforestation of 11.19 %.

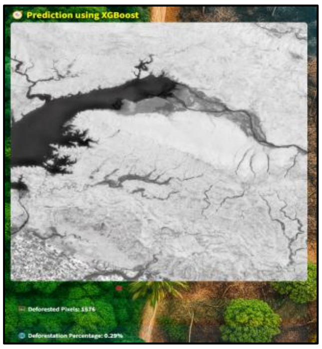
- **XG Boost.**



Kaziranga Forest



Gir Forest



Corbett Forest

Sunderban Forest

Satpura Forest

Western Ghat Forest

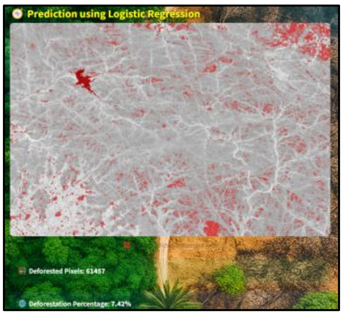
Gachibowli Forest

  - Region Kaziranga Forest has 25,557 deforested pixels; hence XG Boost Model is predicting deforestation of 0.56%.
  - Region Gir Forest has 3358 deforested pixels; hence XG Boost Model is predicting deforestation of 0.41%.
  - Region Corbett Forest has 1576 deforested pixels; hence XG Boost Model is predicting deforestation of 0.29%.

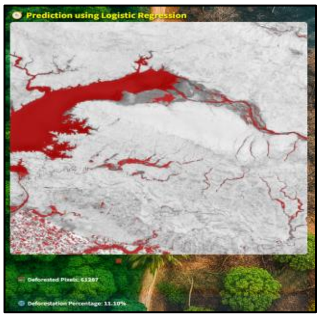
- Region Sunderban Forest has 2193 deforested pixels; hence XG Boost Model is predicting deforestation of 0.40%.
- Region Satpura Forest has 2315 deforested pixels; hence XG Boost Model is predicting deforestation of 0.42%.
- Region WesternGhats Forest has 3255 deforested pixels; hence XG Boost Model is predicting deforestation of 0.69%.
- Region Gachibowli Forest has 250 deforested pixels; hence XG Boost Model is predicting deforestation of 0.50%.

- ## Logistic Regression.

Kaziranga Forest

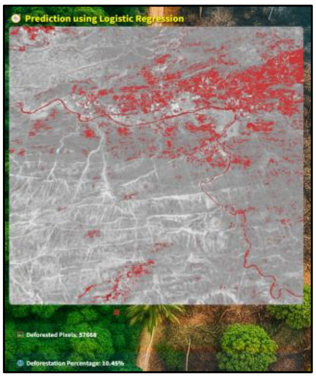


Gir Forest



Corbett Forest

Sunderban Forest



Satpura Forest

Western Ghat Forest

Gachibowli Forest

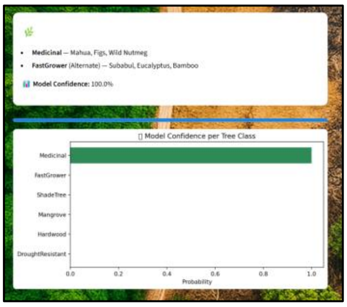
- Region Kaziranga Forest has 22,35,113 deforested pixels; hence Logistic Regression Model is predicting deforestation of 48.56%.
- Region Gir Forest has 61,457 deforested pixels; hence Logistic Regression Model is predicting deforestation of 7.42%.
- Region Corbett Forest has 61,287 deforested pixels; hence Logistic Regression Model is predicting deforestation of 11.10%.
- Region Sunderban Forest has 2,69,749 deforested pixels; hence Logistic Regression Model is predicting deforestation of 48.86%.
- Region Satpura Forest has 57,668 deforested pixels; hence Logistic Regression Model is predicting deforestation of 10.45%.
- Region WesternGhats Forest has 60,013 deforested pixels; hence Logistic Regression Model is predicting deforestation of 12.78%.
- Region Gachibowli Forest has 45,109 deforested pixels; hence Logistic Regression Model is predicting deforestation of 90.30 %.

- **SMART REPLANTING**

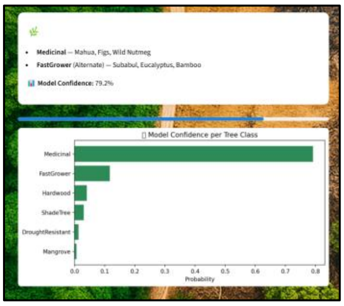
	Kaziranga	Gir Forest	Corbett	Sunderban	Satpura	Western Ghats	Gachibowli
Decision Tree	60%	45.5%	60%	71.4%	41.4%	41.4%	58.60%
Naive Bayes	100%	100%	65.2%	100%	95.8%	100%	100%
Gradient Boosting	79.2%	41.9%	67.8%	63.7%	44.2%	61.6%	75.7%
MLP	99.9%	96.8%	87.3%	99.9%	89.9%	78.3%	66.4%

- **Kaziranga Forest.**

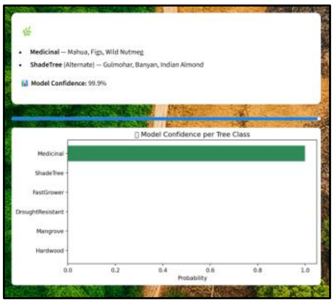
Decision Tree



Naive Bayes



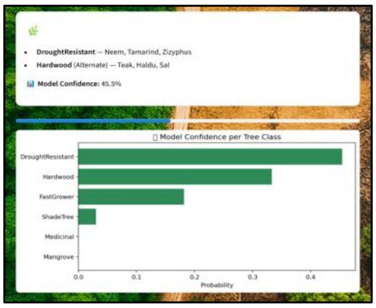
Gradient Boosting



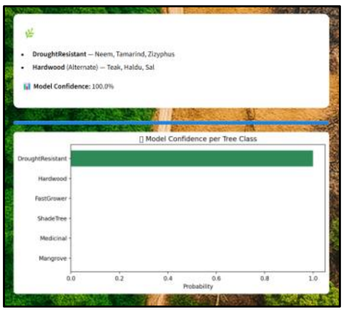
MLP

- Model based on Decision Tree Algorithm shows 60 % confidence, predicting its suitability to grow for classes of crops like Medicinal (Figs, Mahua, Wild Nutmeg) and its alternative Fastgrower (Subabul, Eucalyptus, Bamboo) in Kaziranga Forest.
- Model based on Naive Bayes Algorithm shows 100 % confidence, predicting its suitability to grow for classes of crops like Medicinal (Figs, Mahua, Wild Nutmeg) and its alternative Fastgrower (Subabul, Eucalyptus, Bamboo) in Kaziranga Forest.
- Model based on Gradient Boosting Algorithm shows 79.2% confidence, predicting its suitability to grow for classes of crops like Medicinal (Figs, Mahua, Wild Nutmeg) and its alternative Fastgrower (Subabul, Eucalyptus, Bamboo) in Kaziranga Forest.
- Model based on MLP Algorithm shows 60 % confidence, predicting its suitability to grow for classes of crops like Medicinal (Figs, Mahua, Wild Nutmeg) and its alternative ShadeTree (Gulmohar, Banyan, Indian Almond) in Kaziranga Forest.

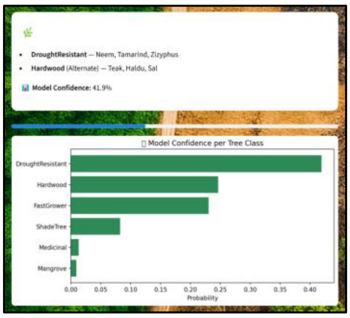
- ## Gir Forest.



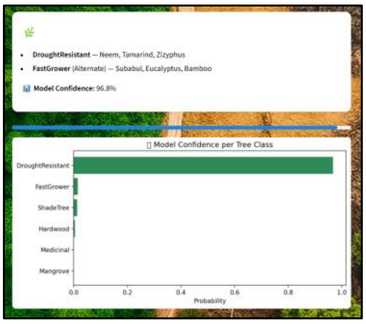
Decision Tree



Naive Bayes



Gradient Boosting

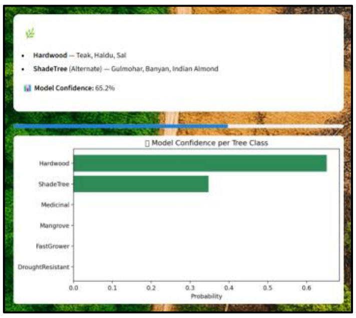


MLP

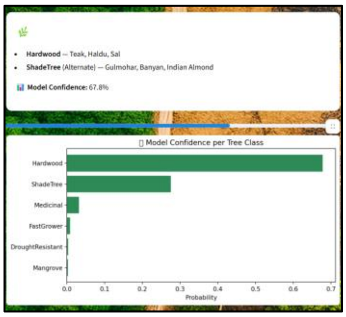
- Model based on Decision Tree Algorithm shows 45.5 % confidence, predicting its suitability to grow for classes of crops like DroughtResistant (Neem, Tamarind, Zizyphus) and its alternative Hardwood (Teak, Sal, Haldu) in Gir Forest.
- Model based on Naive Bayes Algorithm shows 100 % confidence, predicting its suitability to grow for classes of crops like DroughtResistant (Neem, Tamarind, Zizyphus) and its alternative Hardwood (Teak, Sal, Haldu) in Gir Forest.
- Model based on Gradient Boosting Algorithm shows 41.9 % confidence, predicting its suitability to grow for classes of crops like DroughtResistant (Neem, Tamarind, Zizyphus) and its alternative FastGrower (Teak, Sal, Haldu) in Gir Forest.
- Model based on MLP Algorithm shows 96.8 % confidence, predicting its suitability to grow for classes of crops like DroughtResistant (Neem, Tamarind, Zizyphus) and its alternative FastGrower (Subabul, Eucalyptus, Bamboo) in Gir Forest.

- ## Corbett Forest.

Decision Tree



Naive Bayes

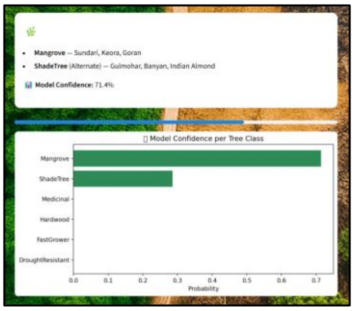


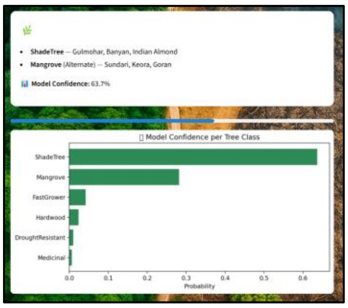
Gradient Boosting

MLP

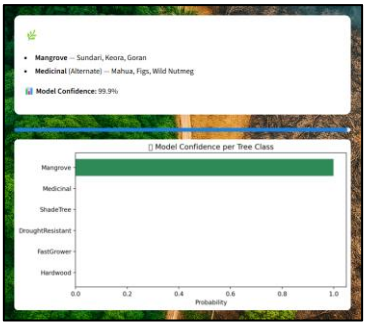
- Model based on Decision Tree Algorithm shows 60 % confidence, predicting its suitability to grow for classes of crops like Hardwood (Teak, Haldu, Sal) and its alternative ShadeTree (Gulmohar, Banyan, Indian Almond) in Corbett Forest.
- Model based on Naïve Bayes Algorithm shows 65.2 % confidence, predicting its suitability to grow for classes of crops like Hardwood (Teak, Haldu, Sal) and its alternative ShadeTree (Gulmohar, Banyan, Indian Almond) in Corbett Forest.
- Model based on Gradient Boosting Algorithm shows 67.8 % confidence, predicting its suitability to grow for classes of crops like Hardwood (Teak, Haldu, Sal) and its alternative ShadeTree (Gulmohar, Banyan, Indian Almond) in Corbett Forest.
- Model based on MLP Algorithm shows 87.3 % confidence, predicting its suitability to grow for classes of crops like ShadeTree (Gulmohar, Banyan, Indian Almond) and its alternative Hardwood (Teak, Haldu, Sal) in Corbett Forest.

- ## Sunderban Forest.





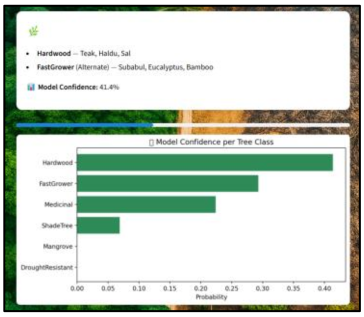
Decision Tree **Naive Bayes** **Gradient Boosting**



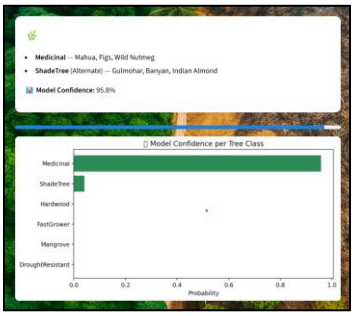
MLP

- Model based on Decision Tree Algorithm shows 71.4 % confidence, predicting its suitability to grow for classes of crops like Mangrove (Sundari, Keora, Goran) and its alternative ShadeTree (Gulmohar, Banyan, Indian Almond) in Sunderban Forest.
- Model based on Naïve Bayes Algorithm shows 100 % confidence, predicting its suitability to grow for classes of crops like Mangrove (Sundari, Keora, Goran) and its alternative ShadeTree (Gulmohar, Banyan, Indian Almond) in Sunderban Forest.
- Model based on Gradient Boosting Algorithm shows 63.7% confidence, predicting its suitability to grow for classes of crops like ShadeTree (Gulmohar, Banyan, Indian Almond) and its alternative Mangrove (Sundari, Keora, Goran) in Sunderban Forest.
- Model based on MLP Algorithm shows 99.9 % confidence, predicting its suitability to grow for classes of crops like Mangrove (Sundari, Keora, Goran) and its alternative Medicinal (Mahua, Figs, Wild Nutmeg) in Sunderban Forest.

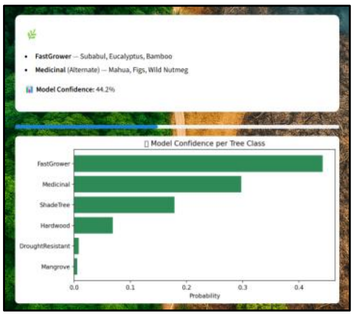
- ## Satpura Forest.



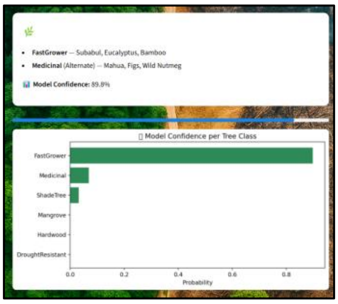
Decision Tree



Naive Bayes



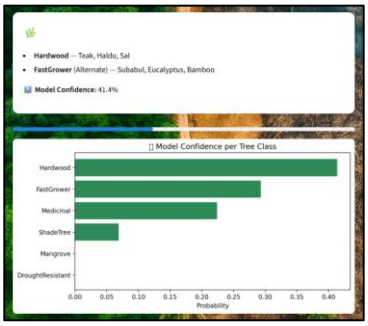
Gradient Boosting



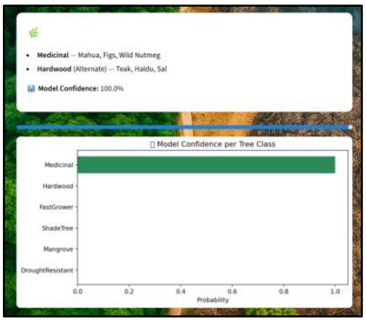
MLP

- Model based on Decision Tree Algorithm shows 41.4 % confidence, predicting its suitability to grow for classes of crops like Hardwood (Teak, Haldu, Sal) and its alternative FastGrower (Subabul, Eucaluptus, Bamboo) in Satpura Forest.
- Model based on Naive Bayes Algorithm shows 95.8 % confidence, predicting its suitability to grow for classes of crops like Medicinal (Mahua, Figs, Wild Nutmeg) and its alternative ShadeTree (GulMohar, Banyan, Indian Almond) in Satpura Forest.
- Model based on Gradient Boosting Algorithm shows 44.2 % confidence, predicting its suitability to grow for classes of crops like FastGrower (Subabul, Eucalyptus, Bamboo) and its alternative Medicinal (Mahua, Figs, Wild Nutmeg) in Satpura Forest.
- Model based on MLP algorithm shows 89.9 % confidence, predicting its suitability to grow for classes of crops like FastGrower (Subabul, Eucalyptus, Bamboo) and its alternative Medicinal (Mahua, Figs, Wild Nutmeg) in Satpura Forest.

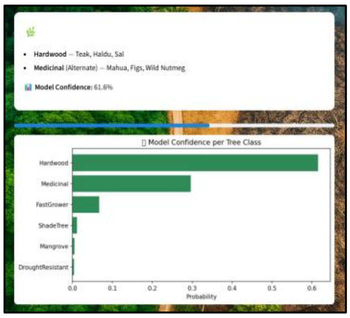
- ## Western Ghat Forest.



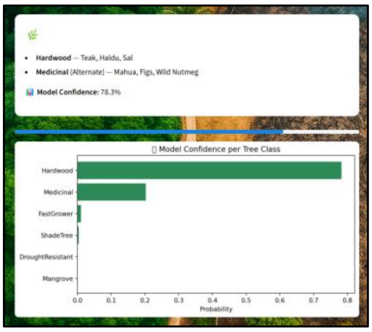
Decision Tree



Naive Bayes



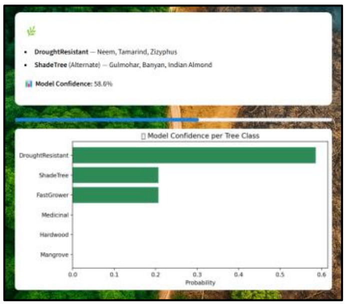
Gradient Boosting



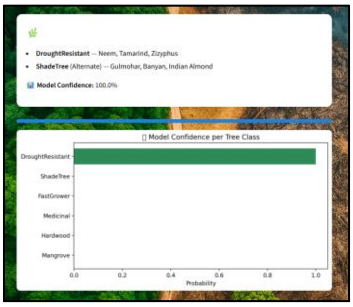
MLP

- Model based on Decision Tree Algorithm shows 41.4 % confidence, predicting its suitability to grow for classes of crops like Hardwood (Teak, Haldu, Sal) and its alternative FastGrower (Subabul, Eucalyptus, Bamboo) in Western Ghat Forest.
- Model based on Naive Bayes Algorithm shows 100 % confidence, predicting its suitability to grow for classes of crops like Medicinal (Mahua, Figs, Wild Nutmeg) and its alternative Hardwood (Teak, Haldu, Sal) in Western Ghat Forest.
- Model based on Gradient Boosting Algorithm shows 61.6 % confidence, predicting its suitability to grow for classes of crops like Hardwood (Teak, Haldu, Sal) and its alternative Medicinal (Mahua, Figs, Wild Nutmeg) in Western Ghat Forest.
- Model based on MLP Algorithm shows 78.3 % confidence, predicting its suitability to grow for classes of crops like Hardwood (Teak, Haldu, Sal) and its alternative Medicinal (Mahua, Figs, Wild Nutmeg) in Western Ghat Forest.

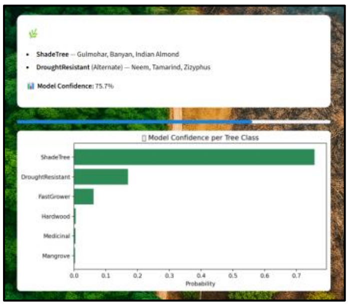
- ## Gachibowli Forest.



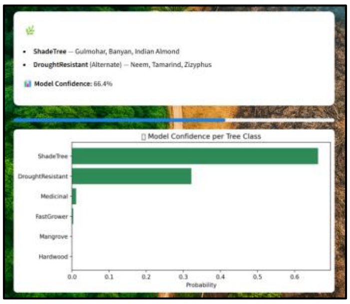
Decision Tree



Naive Bayes



Gradient Boosting



MLP

- Model based on Decision Tree Algorithm shows 58.6 % confidence, predicting its suitability to grow for classes of crops like DroughtResistant (Neem, Tamarind, Zizyphus) and its alternative ShadeTree (GulMohar, Banyan, Indian Almond) in Gachibowli Forest.
- Model based on Naive Bayes Algorithm shows 100 % confidence, predicting its suitability to grow for classes of crops like DroughtResistant (Neem, Tamarind, Zizyphus) and its alternative ShadeTree (GulMohar, Banyan, Indian Almond) in Gachibowli Forest.
- Model based on Gradient Boosting Algorithm shows 66.4 % confidence, predicting its suitability to grow for classes of crops like ShadeTree (GulMohar, Banyan, Indian Almond) and its alternative DroughtResistant (Neem, Tamarind, Zizyphus) in Gachibowli Forest.

CHAPTER 5 - CONCLUSION

✓ DEFORESTATION

- The classification report consists of key evaluation metrics such as Precision, Recall, F1 Score, Accuracy, Macro Avg F1, and Weighted Avg F1.
- Our project focuses on detecting deforested pixels from grayscale NDVI images, where the F1 score serves as a balanced measure of both precision and recall, offering a clearer insight into the model's effectiveness in identifying deforested areas, so we will consider accuracy w.r.t F1 Score.
- Among the evaluated algorithms, XGBoost achieves the highest F1 score of 0.94 and accuracy of 1.00. However, the model appears to overfit, likely due to its design being more suitable for multiclass datasets, whereas our dataset is binary (class 0 and class 1).
- Therefore, XGBoost is excluded, and SVM (RBF kernel) is preferred, as it delivers the most reliable results with an F1 score of 0.71 and accuracy of 0.64, showing no signs of overfitting.
- Random Forest ranks third, achieving a very high accuracy of 0.99 but a relatively low F1 score of 0.25, indicating poor balance between precision and recall.
- Finally, Logistic Regression performs the weakest, with an F1 score of just 0.02 and accuracy of 0.59, placing it fourth in terms of overall performance.
- Hence, we consider all deforestation prediction w.r.t SVM with RBF kernel due to its high training accuracy.

✓ SMART REPLANTING

- The classification report includes precision, recall, F1-score, and accuracy, offering insights into each model's ability to classify six vegetation groups (DroughtResistant, FastGrower, Hardwood, Mangrove, Medicinal, ShadeTree).
- Gradient Boosting demonstrates the best overall performance with a test accuracy of 0.80 and a macro weighted average F1-score of 0.80, showing it effectively captures feature interactions and handles class imbalance.
- MLP (Multilayer Perceptron) achieves a moderate accuracy of 0.60 and macro F1-score of 0.59, making it a reasonable deep learning choice, especially when more data becomes available.
- Decision Tree shows good interpretability but limited performance, with accuracy of 0.63 and macro F1-score of 0.56, reflecting overfitting on smaller classes like Medicinal and ShadeTree.
- Naive Bayes, while simple and fast, lags behind in performance with accuracy of 0.57 and a macro F1-score of 0.51, likely due to its independence assumption not holding for correlated features like soil and rainfall.
- Hence, Gradient Boosting is preferred for deployment in Smart Replanting due to its balanced precision and generalization across vegetation types so we consider all predictions for Smart Replanting w.r.t Gradient Boosting due to its high training accuracy.

CHAPTER 6 - REFERENCES

1. Imran Md Jelas et al. (2024). *“Deforestation detection using deep learning-based semantic segmentation techniques”*. Frontiers in Forests and Global Change. Discusses U-Net and FCN for accurate deforestation mapping using Sentinel-2 imagery.
2. Xiang J., Xing Y., Wei W., et al. (2023). *“Dynamic Detection of Forest Change in Hunan Province Based on Sentinel-2 Images and Deep Learning”*. Remote Sensing. Compares U-Net++, DeepLabV3+ etc., for multi-temporal land cover change detection.
3. Gabor Fodor & Marcos V. Conde (2023). *“Rapid Deforestation and Burned Area Detection using Deep Multimodal Learning on Satellite Imagery”*. Preprint (arXiv). Uses CNN-based multimodal models over Sentinel, Landsat, and MODIS for deforestation and fire detection.
4. Isaienkov K., Yushchuk M., Seliverstov O., et al. (2021). *“Deep Learning for Regular Change Detection in Ukrainian Forest Ecosystem With Sentinel-2”*. IEEE JSTARS. Implements U-Net diff and Deep Change detection.
5. Fabien H. Wagner et al. (2022). *“Mapping Tropical Forest Cover and Deforestation with Planet NICFI Satellite Images and Deep Learning in Mato Grosso, Brazil”*. Preprint. U-Net-driven segmentation of deforestation using high-resolution Planet NICFI imagery.
6. Tanguturi Yathesh, M. S. Likhith Siddu & K. Senthil Kumar (2025). *“Satellite Image-Based Deforestation Detection Using Deep Learning”*. Smart Computing Paradigms: Sustainable Computing (LNNS). Presents a CNN approach to detect deforestation based on grayscale NDVI imagery.
7. Kennedy R., et al. (2014). *“Detection of Forest Loss Using NDVI”*. Remote Sensing of Environment. A foundational study using Landsat NDVI differencing for detecting forest degradation.
8. Breiman L. (2001). *“Random Forests”*. Machine Learning. Seminal paper describing Random Forest algorithm, widely used for pixel-level classification.
9. Zhang et al. (2020). *“Deep Learning in Land Cover Classification”*. ISPRS Journal. Reviews CNN applications in land cover segmentation tasks.

10. Rani S. & Sinha M. (2020). *“Machine Learning Models for Crop Recommendation Based on Soil and Climate”*. IEEE ICDE. Evaluates RF, Naive Bayes, DT for smart replanting decisions.
11. Chakraborty D. et al. (2023). *“Smart Reforestation System using AI and IoT”*. IEEE Access. Integrates ML models with ecological datasets and IoT for regional planting planning.
12. Raj A. & Kumar V. (2022). *“Streamlit-Based Visualization Platform for ML-Powered Forest Risk Assessment”*. IEEE INDICON. Demonstrates how Streamlit can be used to deploy forest monitoring ML tools for user-friendly decision support.