

**DSA 2020A**

**MILESTONE III**

**MACHARIA JAMES MWANGI – 667668**

**CHEPKURUI NATASHA CHEPKWEMOI - 666660**

Progress Report: AI for Sustainable Forest Management

**AI for Sustainable Forest Management: Safeguarding Life on Land (SDG 15)**  
  
The next is a report of the progress on the AI for Forest Sustainability Project elaborating the development process, the results of testing, and the problems faced. The critical objectives of the project are to put to the best use of AI technologies on the monitoring of illegal logging activities, data-driven forest management, and optimization of reforestation efforts for better forest sustainability.  
  
2. Development Process  
  
**Data Collection and Preprocessing**  
  
Data Sources:  
\* Satellite Imagery: Sentinel-2 image acquired from the Copernicus Open Access Hub  
\* Drone Surveillance: High resolution images captured by the Plant Sciences department in the University of Cambridge  
\* IoT Sensors: Environmental data collected from the GLOBE Program network of monitoring  
\* Textual Data: News and reports collected by Kaggle and other archives of online news  
  
Preprocessing  
Data Cleaning:

Remove the noise and data from the data taken from the satellite imagery and drone footage, apart from that required for Python libraries such as OpenCV.  
Data Transformation:

Normalized and formatted remotely-sensed images from satellites and drones for model input, using the aforementioned image processing techniques.

Feature Extraction:

Major features like tree canopy coverage, soil moisture levels, and textual mentions of illegal logging activities that one needs to extract or identify from the data, using Pandas and other data manipulation tools.  
  
Initial Model Training  
Computer Vision Model  
Technique:

Convolutional Neural Network (CNN) using TensorFlow/Keras  
Dataset:

Sentinel-2 satellite imagery  
The model was capable of capturing deforested areas with an accuracy of 85%.

Machine Learning Model  
Technique: Random Forest Regression using Scikit-learn  
Dataset: GLOBE environmental data and historical forest growth data  
Mean Absolute Error in the prediction of tree growth cycles was 3.5 years

NLP Model

Technique: Sentiment Analysis using NLTK and Sc CONCLUSION   
Dataset: Articles and reports about the activities on forest.  
Performance: The testing regarding the sentiment classification of the activities done on forests gave back an output of 78% in accuracy.  
  
---  
  
3. Testing Results  
  
Computer Vision Model Testing  
  
Testing Strategy:  
- Testing was held on a different dataset of satellite images.  
- It compares the model prediction with the area actually deforested on the ground.  
  
Results:  
- Accomplished 87% precision and 82% recall in the classification of deforested areas.  
- Major misclassifications occurred due to cloud coverage and image resolution issues.  
  
Machine Learning Model Testing  
  
Test Plan:  
The model was tested on the 7:3 Train: Test divided dataset.  
Metric Used: Mean Absolute Error.  
  
Findings:  
- The pattern in tree growth cycles was somewhat predictable. It secured a MAE of 3.5 years.  
- Very poor performance at places highly environmentally critical.  
  
Testing NLP Model  
  
Test Plan:  
- Holdout test sets made with news articles and the reports  
Model performance calculated through accuracy, precision, and recall measures.  
  
Results:  
- Overall accuracy for classification of sentiments was 78%.  
- The model lacks the potential for the classification of subtle and mixed sentiments in textual data.  
  
  
4. ISSUES ADDRESSED  
  
Data Quality and Availability  
- Problem: Images are from satellite images and drone images and are of varied quality due to conditions like weather and others.  
- Solution: Data augmentation techniques used to enhance the quality of the training dataset to make the model robust against that.

Model Complexity and Training Time  
  
Challenge:

Being computationally expensive and thus a time-consuming process, training of models incorporating CNN calls for solutions.

Solution:

In this work, trying to accelerate the training process of the model and make use of in-place pre-trained models for transfer learning, cloud computing resources will be used.   
  
Integration of Heterogeneous Data Sources  
Challenge: This model integrates heterogeneous data sources like satellite images, drone images, sensor data, and text data.  
Resolution: The modular system architecture will permit parallel processing of several data types as well as the inclusion of results.  
  
  
5. Other Enhancements and Optimizations  
  
Fine-tuning of existing models  
- Added more layers to the CNN model and did hyperparameter tuning.  
- Designed better feature combinations in case of random forest regression model and proper handling of missing values.  
- Enhanced text preprocessing techniques for the NLP sentiment analysis model.  
- Increased the size of training datasets for the sentiment analysis model.  
  
New Features in the Implementation  
  
- Real-time monitoring dashboard for visualizing the forest conditions and illegal logging alerts.  
- Predictive analytics tools to aid in planning sustainable logging activities in the forest  
- AI reforestation planner tool to optimize the selection of sites and monitor tree growth.  
  
6. Evaluation and Potential Impact  
  
Application Performance Evaluation  
  
- Accuracy and Precision: The computer vision model showed greater accuracy and precision during the identification process in deforested areas.  
- Predictive Capability: Reliability in prediction capability of machine learning model with respect to forest growth cycle.  
- Sentiment Analysis: The NLP model has done quite well in the exercise of classification of sentiments over textual data; this has contributed to the bigger picture in trying to understand what the community feels in regards to forest conservation.  
  
Potential Impact  
  
- Environmental Impact: Improved surveillance and prevention from illegal logging practices lead to more sustainable forest management.  
- Economic Impact: Improve efficiency and profitability of activities in regards to the forests by making better decisions offered by forest managers.  
- Social Impact: AI-driven better educational platforms and increased engagement/awareness through sentiment analysis of communities.  
  
  
7. Conclusion  
  
This progress report underscores key progress made toward the realization of this AI application for sustainable forest management. It solves some of the main challenges associated with the sustainability of forests by means of machine learning, computer vision, and natural language processing. Preliminary tests of the system are very encouraging about the potential of the app to make an environmental, economic, and social difference. Further development of AI models by including other features will better the solution in protecting life on land.  
  
Next Steps:  
- Refinement of the AI models based on feedback and as more data arrive.  
- Growth of the dataset to cover diverse and thorough sources.  
- Detailed field campaigns to test and validate the performance of the models under real conditions.  
- Implementation of an appropriate, in-depth user training program to ensure adoption by forest managers and subsequent related stakeholders.