SEEING THE FOREST THROUGH THE TREES

Exploring Deforestation Trends in Colombia with Machine Learning

This study analyzes deforestation trends in Colombia (2010–2022) using interpretable machine learning. By combining a panel dataset with spatial and temporal analysis, it identifies key drivers—such as illicit coca cultivation, spatial marginality, and conflict—and reveals regional differences. The findings emphasize the need for region-specific deforestation policies that consider socio-economic and institutional complexities.

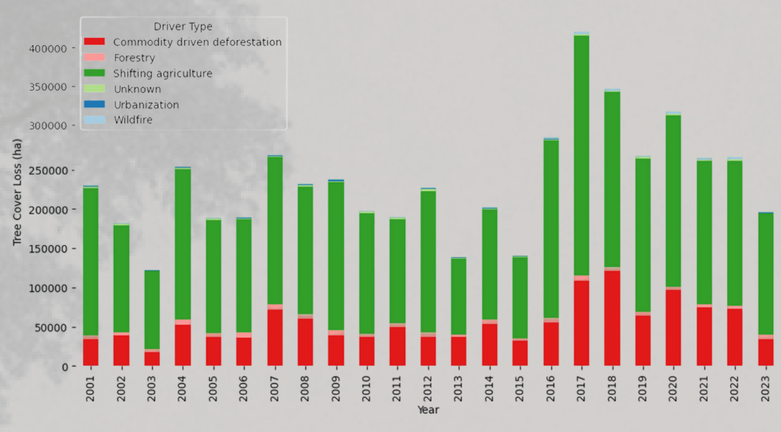
Background

Despite national policies like the 2020 Deforestation Control Policy and the 2016 Peace Agreement, forest loss remains high.

Regional studies show that deforestation dynamics vary widely and require localized responses.

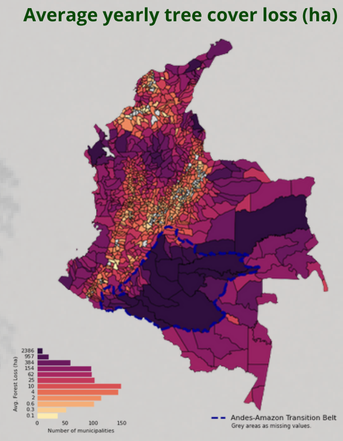
Special focus on the Andes–Amazon Transition Belt, since it is a critical deforestation and coca cultivation hotspot.

Tree cover loss (ha) in Colombia by type of driver based on the World Resources Institute data (2001-2023)

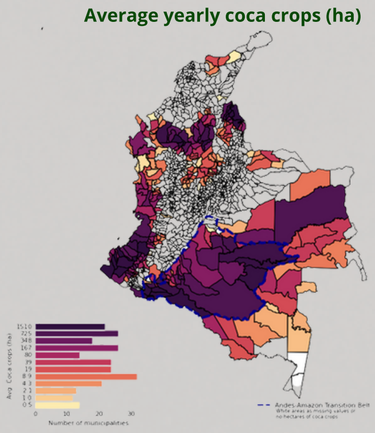


Average tree cover loss (ha) and coca crops cultivation (ha) by municipalities (2010-2022)

Average yearly tree cover loss (ha)



Average yearly coca crops (ha)



Literature Review

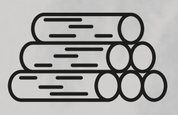
Direct Causes of Deforestation

Cattle ranching

Agriculture

Illicit Crops

Infrastructure and mining



Wood extraction

Fires/Natural disasters

Indirect Causes of Deforestation

Economic factors

Technology

Demographic characteristics

Conflict

Institutional/Political factors

Environmental factors

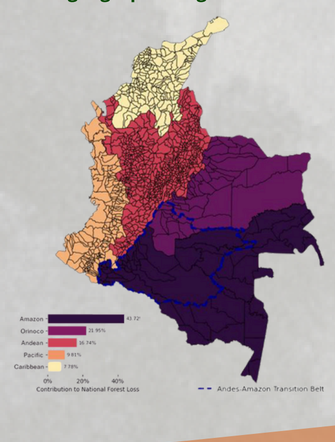
Culturall factors

Methodology

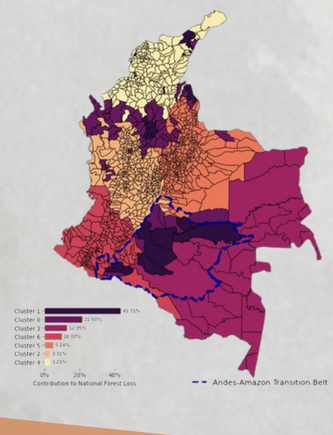
1. **Preprocessing of seven datasets from several sources, such as CEDE, DANE and DNP. Available data for 1,118 municipalities and 13 years (2010-2022).**
2. **Out-of-sample performance comparison of seven different machine learning algorithms based on four metrics: MSE, RMSE, MAE and R^2.**
3. **SHAP analysis for feature importance.**
4. **Spatiotemporal analysis for the best performing model for all available years and two types of regional grouping. Use of K-Means algorithm for spatial clustering based on deforestation trends.**

Average contribution (%) to the national tree cover loss by biographic regions and clusters (2010-2022)

Biogeographic Regions



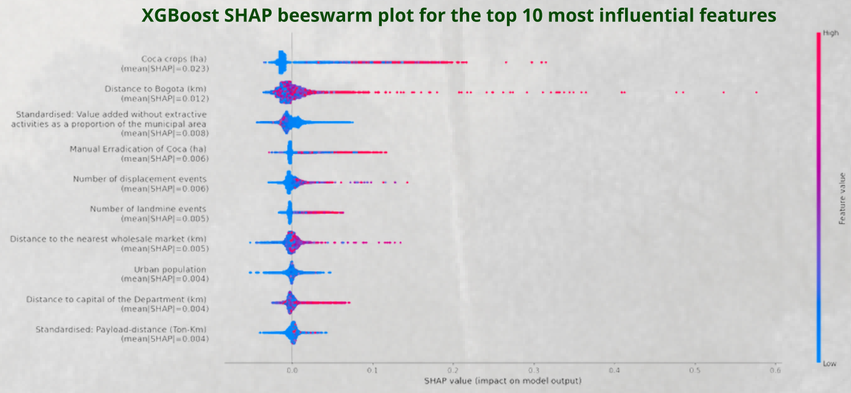
K-Means Clusters



Results

XGBoost outperformed all other models in predicting deforestation (R^2 = 0.696), highlighting the importance of modeling non-linear and complex interactions between drivers of forest loss.

XGBoost SHAP beeswarm plot for the top 10 most influential features

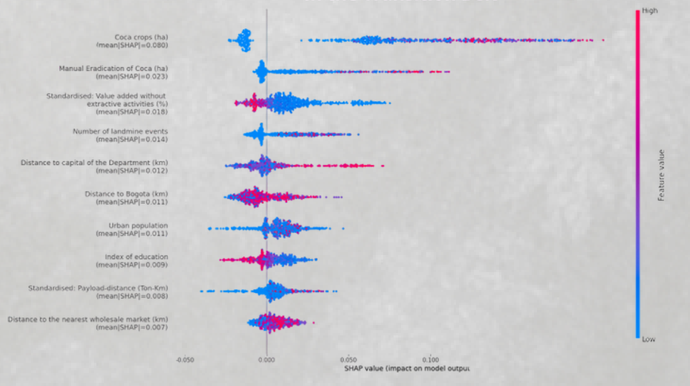


Main drivers of deforestation identified: illicit crops, remoteness (e.g., distance to Bogota or department capital), conflict related factors (e.g., displacement and landmine incidents), social and institutional factors (e.g., education and political participation), and economic diversification.

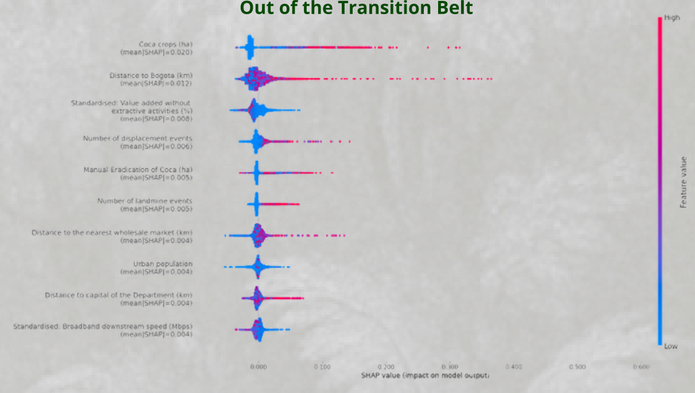
Regional analysis allows to see specific dynamics. in particular,. K-means clustering identified new high-deforestation hotspots not fully captured by traditional biogeographic zones, such as in the South Pacific region and the South of the Andes-Amazon Transition Belt.

Beeswarm plot for the Top 10 Features for the Andes-Amazon Transition Belt

In the Transition Belt



Out of the Transition Belt



From 2013 onward, manual coca eradication became a key driver, likely reflecting policy changes during peace negotiations with FARC. By 2022, agricultural performance (tons/hectare) emerged as a new top predictor, acting as a stabilizing factor against deforestation through improved land-use efficiency.

Recommendations

- **Leverage interpretable machine learning to support evidence-based decision-making and adaptive environmental management.**
- **Implement region-specific strategies that strengthen local governance, improve land-use monitoring, and target remote and conflict-affected areas, where institutional weakness and insecurity drive deforestation.**
- **Promote sustainable rural development by integrating forest conservation into economic planning, supporting alternative livelihoods in coca-producing areas, and encouraging high-productivity agriculture and economic diversification that avoids expansion into forested land.**

FULL TEXT

Milton Mier Santander| M.Mier@students.hertie-school.org | MDS 2025

Hertie School **Camilo Pedraza Jiménez| C.Pedraza@students.hertie-school.org| MDS 2025**