Final Summary: Deforestation Issue Analysis Using SVM

Objective

The goal of this project was to analyze global deforestation trends and predict the percentage of tree cover loss using Support Vector Machine (SVM). By identifying key factors such as illegal logging, corruption, policy enforcement, and economic indicators, the project aimed to provide actionable insights for reducing deforestation and improving environmental policy.

Summary with Insights

In this project, I used a Support Vector Machine (SVM) with a linear kernel to predict Tree Cover Loss (%) and analyze the major factors contributing to deforestation across different countries. After preprocessing the dataset, I trained the model and achieved the following performance metrics:

- Mean Absolute Error (MAE): 2.28%
- Mean Squared Error (MSE): 6.37
- Root Mean Squared Error (RMSE): 2.52%
- R-squared (R^2): 0.71

These results indicate moderate predictive accuracy for policy-level insights.

The analysis revealed that illegal lumbering incidents had the strongest direct impact on deforestation, followed by corruption in governance and expansion of agricultural land. In contrast, stricter deforestation policies and the presence of protected areas were associated with a clear reduction in forest loss. Interestingly, variables like CO2 emissions, population, and GDP showed indirect or regionally dependent effects, suggesting that economic development alone does not directly predict deforestation levels.

Based on the model results, I recommend:

- Strengthening forest protection laws and enforcement mechanisms
- Expanding the area under legal protection
- Targeting corruption within the forestry sector
- Promoting sustainable agricultural practices

These measures, if implemented strategically, could significantly curb tree cover loss in high-risk regions.

Overall, this project demonstrates how machine learning models like SVM can uncover the hidden drivers of deforestation and support better decision-making. Future extensions of this work could include experimenting with non-linear models (e.g., RBF kernel, Random Forests), time-series forecasting, or integrating spatial data (such as satellite imagery) for more advanced modeling and regional targeting.