**Readme: Forecasting Illegal Activities in Arabuko-Sokoke Forest**

Welcome to the Forecasting Illegal Activities in Arabuko-Sokoke Forest repository! This project aims to provide actionable insights to guide conservation interventions and law enforcement strategies by forecasting future illegal activities in the Arabuko-Sokoke Forest (ASF).

**Repository Contents**

1. **Jupyter Notebook (Illegal\_Activities\_Analysis.ipynb):**
   * This notebook contains a comprehensive analysis of illegal activities data in ASF, including data preprocessing, exploratory data analysis, time series analysis and modelling, and evaluation.
2. **README.md:**
   * This file provides an overview of the repository and instructions for running the code. It also includes information about the project structure, usage guide, references, and contributors.

**Introduction**

The Arabuko-Sokoke Forest (ASF), a UNESCO Man and Biosphere Reserve and a crucial biodiversity hotspot, faces persistent threats from illegal poaching and logging. These activities are primarily driven by the high demand for forest products from nearby urban centers such as Malindi, Watamu, Kilifi, and Mombasa, as well as the presence of over 200,000 people living at the forest boundary. These threats endanger the integrity of the forest ecosystem, its biodiversity, and the livelihoods of local communities dependent on its resources. Effective strategies are needed to combat illegal utilization of forest resources and safeguard the ecological and socio-economic value of ASF.

**General Objectives**

* To provide actionable insights to guide conservation interventions and law enforcement strategies with historical law-enforcement and biodiversity monitoring data.
* To develop forecasting models to forecast future illegal activities in the forest reserve and support proactive conservation efforts.

**Business and Data Understanding**

**Business Problem:** Providing actionable insights to guide conservation interventions and law enforcement strategies by forecasting future illegal activities in the Arabuko-Sokoke Forest (ASF).

**Implementation Overview:**

1. **Data Preparation:**
   * We used a dataset from a survey carried out by Arabuko-Sokoke Forest (ASF).
   * We parsed waypoint dates, dropped unnecessary columns and rows, created a function which categorized entries into poaching and logging, and renamed this new entry as illegal activities. We also filled missing values, took care of duplicates, and reset waypoint date as the index.
2. **Exploratory Data Analysis (EDA):**
   * The dataset was explored visualizing illegal activities over time. Temporal and seasonal analysis was carried out identifying trends and seasonality. Augmented Dickey-Fuller Test was employed which confirmed non-stationarity, differencing technique was applied before modeling.
3. **Time Series Modeling:**
   * Time series component was decomposed, auto-correlation, and partial auto-correlation function were analyzed.
   * Three Models were used: ARIMA, VAR, and XGBoost which were fitted and also used for forecasting.
4. **Feature Engineering and Modeling:**
   * Time series features were created, visualizing relationships between features and illegal activities, combining logging and poaching into a single target variable, training XGBoost regressor models, feature importance was also analyzed.
5. **Forecasting and Testing:**
   * Illegal activities were forecasted using trained models, forecasted values were against actual data, and model performance was evaluated using RMSE scores.

**Future Directions**

1. **Refinement of Machine Learning Models:** Further tuning of hyperparameters and exploration of additional algorithms to improve model performance.
2. **Integration of External Data:** Incorporation of external datasets (e.g., weather data, socio-economic indicators) to enhance predictive capabilities.
3. **Real-Time Monitoring System:** Development of a real-time monitoring system for early detection and prevention of illegal activities in ASF.
4. **Stakeholder Engagement:** Collaboration with conservation organizations, law enforcement agencies, and local communities to implement and validate predictive models in practice.

**Instructions**

1. Clone the repository to your local machine.
2. Open the Jupyter Notebook (Illegal\_Activities\_Analysis.ipynb) using Jupyter Notebook or JupyterLab.
3. Ensure that the necessary libraries (e.g., pandas, numpy, seaborn, statsmodels, xgboost) are installed.
4. Run each cell in the notebook sequentially to reproduce the analysis and results.

**Recommendations and Conclusions**