MACHING LEARNING PACKAGE ABSTRACT

TITLE: Algorithmic Trading

TEAM: Ashritha J S (23PW02)

Dhanya Shree (23PW05)

ALGORITHMS USED:

Random Forest Regressor: An ensemble learning method that builds multiple decision trees and averages their predictions. It is robust to overfitting, can capture complex non-linear relationships, and works well with noisy and high-dimensional financial data.

Gradient Boosting: A sequential ensemble technique where weak learners (usually shallow decision trees) are trained iteratively to correct the errors of previous models. It is highly effective for regression tasks and can model subtle patterns in financial time series.

ARIMA (AutoRegressive Integrated Moving Average): A statistical time series model that combines autoregression, differencing, and moving average components. It is suitable for forecasting linear trends and seasonal patterns in financial data.

Reinforcement Learning: A machine learning approach where an agent learns to make decisions by interacting with an environment and receiving feedback in the form of rewards. In finance, it can be applied for portfolio management, trading strategies, and dynamic decision-making under uncertainty.

DATASET:

Source: Yahoo Finance (yfinance Python library)

The dataset includes historical stock market data such as Open, High, Low, Close, and Volume for selected stocks or indices (e.g., NIFTY 50, S&P 500).
Derived features such as Moving Averages (MA), Relative Strength Index (RSI), MACD, and Volatility are used as input features for the model.

ABSTRACT:

The project aims to predict the next-day stock price movement (Up/Down) using machine learning. After training models on various technical indicators, both **SHAP (SHapley Additive exPlanations)** and **LIME (Local Interpretable Model-agnostic Explanations)** are applied to interpret the predictions. These explainability frameworks provide human-understandable insights such as:
"The model predicts the stock will go up because RSI < 30 and the 5-day Moving Average crossed above the 20-day Moving Average."
By leveraging SHAP and LIME together, the framework ensures that trading decisions are transparent, interpretable, and data-driven, rather than relying on black-box predictions. This enhances trust and actionable understanding for investors and analysts.

ASPECT:

Explainability:

The model focuses on the *explainability* aspect of machine learning by providing reasoning behind each prediction using SHAP values. This helps users and analysts understand why a particular trading decision or forecast is made, enhancing trust and interpretability in financial ML models.

Tech Stack:

- **Programming Language:** Python
- **Libraries:**
    - **yfinance** – for financial data collection
    - **pandas, numpy** – for data preprocessing and numerical computations
    - **sklearn** – for implementing Random Forest, Gradient Boosting, and evaluation metrics
    - **statsmodels** – for ARIMA modeling
    - **matplotlib, seaborn** – for data visualization and plotting results
    - **shap, lime** – for explainability and model interpretation
    - **gym / stable-baselines3** – for Reinforcement Learning simulation