

1. Strategy Overview

The strategy combines technical indicators with a **Decision Tree Classifier**, a machine learning model, to predict stock price movements and manage portfolio rebalancing. The strategy operates on a universe of equities, dynamically selecting assets based on fundamental data and trading signals. The goal is to allocate capital based on the prediction of future stock price movements relative to the S&P 500 (SPY) and manage risk using trailing stops.

This strategy uses the following key elements:

- **Technical Indicators:** Bollinger Bands, Fast and Slow SMAs.
- **Decision Tree Model:** A supervised learning algorithm that predicts whether a stock will outperform the SPY based on features derived from price, volume, and technical indicators.
- **Feature Extraction:** The model uses Bollinger Band position, SMA ratios, volume ratios, and price momentum as key features for prediction.

2. Technical Indicators Used

- **Bollinger Bands (BB):**
 - Used to measure volatility and identify overbought/oversold conditions. The upper and lower bands are calculated using a 20-period moving average and standard deviations.
 - **Upper Band:** A price above the upper band signals a potential overbought condition.
 - **Lower Band:** A price below the lower band signals a potential oversold condition.
- **Simple Moving Averages (SMA):**
 - **Fast SMA (50-period):** Used to track short-term price trends.
 - **Slow SMA (200-period):** Used to identify the longer-term trend.
 - A crossover between the fast and slow SMAs indicates a potential buy or sell signal (e.g., a bullish signal occurs when the fast SMA crosses above the slow SMA).

The combination of Bollinger Bands and SMAs provides the foundation for generating technical signals, which are then refined using predictions from the Decision Tree model.

3. Decision Tree Model

The **Decision Tree Classifier** is the machine learning core of the strategy. It predicts whether an asset will outperform the SPY based on various features derived from historical price and volume data.

Decision Tree Overview:

- **Decision Tree Classifier:** A decision tree is a supervised learning algorithm that splits data into branches based on the values of input features. Each branch represents a decision, and the model selects the best features to split the data, helping to predict an outcome.
- **Why Decision Tree:** The Decision Tree Classifier is ideal for this strategy because it handles complex, non-linear relationships between the input features (e.g., Bollinger Band position, SMA ratios) and the target (whether the stock will outperform SPY).

4. Training and Weight Optimization

The goal of training the Decision Tree model is to find the best splits in the input data that help classify whether a stock will outperform the SPY over the next 60 days. The process can be divided into two key phases: **training** and **testing**.

Training Phase:

- During training, the model processes **60 days of historical data** for each stock in the selected universe and compares the performance of each stock against SPY.
- The Decision Tree uses the following features:
 1. **Bollinger Band Position**: Measures how close the price is to the upper or lower Bollinger Band.
 2. **SMA Ratios**: The ratio of fast SMA to slow SMA, providing insight into short-term versus long-term trends.
 3. **Price Momentum**: The percentage change in price over the last 20 days.
 4. **Volume Ratios**: Compares average trading volumes over different time periods to detect volume surges.

The training process involves the following steps:

1. **Feature Scaling**: Features are standardized using **StandardScaler** to ensure all features have the same scale.
2. **Model Training**: The model is trained on historical data, using the 60-day performance of each stock relative to SPY as the target label. If a stock outperforms SPY, it is labeled as 1; otherwise, 0.

Testing Phase:

- After training, the Decision Tree is used to make predictions on new, unseen data during live trading. The model no longer updates its splits; it predicts whether a stock will outperform SPY based on current market conditions.

Key Hyperparameters:

- **Max Depth**: The tree is limited to a depth of 5 to avoid overfitting.
- **Feature Scaling**: All input features are scaled using **StandardScaler** to ensure consistency.

5. Feature Extraction for the Decision Tree

The machine learning model uses several features extracted from historical price and volume data:

1. **Bollinger Band Position**: Calculated as the position of the current price relative to the upper and lower Bollinger Bands.
2. **SMA Ratio**: The ratio of the fast SMA (50-period) to the slow SMA (200-period).
3. **Volume Ratio**: The ratio of the 30-day average volume to the overall average volume.
4. **Price Momentum**: The percentage change in the stock price over the past 20 days.

5. **Returns:** The past price returns and their volatility.

These features are used by the Decision Tree model to predict whether the stock will outperform SPY over the next 60 days.

6. Trading Logic and Rules

Once the Decision Tree model generates predictions, the strategy uses those predictions, in combination with technical indicators, to manage trades:

- **Long Entry Conditions:**
 - The Decision Tree predicts that the stock will outperform SPY.
 - The stock's price is above the upper Bollinger Band.
 - The fast SMA is above the slow SMA (indicating a bullish trend).
 - The overall market (SPY) is trading above its 200-day moving average.
 - If all conditions are met, the strategy enters a long position in the stock, allocating 5% of the portfolio to each position.
- **Stop-Loss Condition:**
 - A trailing stop is used to protect gains, set at 40% below the highest price achieved during the holding period.
- **Monthly retraining:**
 - The strategy does retraining the model and adjusting positions based on updated predictions and technical signals.

7. Predictions and Decision-Making

The Decision Tree model's predictions play a key role in the strategy's decision-making process:

- **Positive Prediction:** If the model predicts that a stock will outperform SPY, the strategy considers it for a long position, provided other technical signals (e.g., Bollinger Band and SMA crossover) align.
- **Negative Prediction:** If the model predicts underperformance, the stock is not selected for a position, or existing positions are liquidated.