Market Regime Prediction Using Random Forest Classifier

MIHIR NILESH HOLMUKHE

Introduction

Imagine standing at the edge of a vast forest, trying to understand the patterns of its wildlife. Some areas teem with life, others are quieter, and some have hidden paths where changes occur unexpectedly. Navigating the stock market feels much the same. Investors and analysts seek to classify the market into identifiable patterns—Bull, Bear, or Sideways—to make informed decisions. This project embarks on a similar journey, leveraging historical stock data to uncover these patterns and predict future trends. By employing the Random Forest Classifier, a powerful machine learning tool, we aim to illuminate the market's hidden pathways and provide actionable insights.

Objective

The primary goal of this project is to classify the stock market into three regimes based on historical data and predict future regimes. Accurately identifying these regimes is crucial as they provide insight into the market's current behaviour and likely future trends, enabling more informed decision-making.

The regimes are defined as:

- **Bull:** An upward trending market, characterized by consistent increases in stock prices. This phase often signals investor optimism and is associated with economic growth. It's a favourable environment for investors aiming for capital gains.
- **Bear:** A downward trending market, marked by significant declines in stock prices. This phase reflects pessimism among investors, often linked to economic slowdowns or crises. Understanding bear markets allows stakeholders to mitigate risks and develop strategies for preservation.
- **Sideways:** A relatively stable market with no significant trend in either direction. This phase indicates market indecision or balance between supply and demand. Identifying sideways markets helps investors avoid unproductive trades and focus on long-term opportunities.

Dataset

We utilized historical stock data for Apple Inc. (**AAPL**) obtained from the **Yahoo Finance API**. The dataset spans the period from **August 1, 2024, to October 31, 2024**. Key features engineered from the data include:

- Close Prices
- Moving Averages (6-day, 10-day, 20-day)
- Exponential Moving Averages (EMA50, EMA200)
- On-Balance Volume (OBV)

Libraries Used:

- Data Collection: yfinance
- Data Analysis: pandas, numpy
- Visualization: matplotlib
- Machine Learning: sklearn

Methodology

1. Feature Engineering

We selected the following technical terms to define and predict market behaviour:

- **Moving Averages (MA6, MA10, MA20):** Moving averages smooth out short-term price fluctuations, making it easier to identify trends. The different windows (6-day, 10-day, and 20-day) allow us to capture varying levels of price momentum, which is crucial for distinguishing between short-term and long-term trends.
- **Exponential Moving Averages (EMA50, EMA200):** EMAs prioritize recent data points, making them highly responsive to changes. EMA50 and EMA200 were chosen to capture medium and long-term trends, respectively, ensuring a well-rounded perspective on market momentum.
- **On-Balance Volume (OBV):** OBV integrates price and volume information, reflecting the intensity of buying and selling pressure. This feature was included because volume is often a precursor to price movements, providing early signals of regime changes.

These features were chosen because they are widely used by traders and analysts to interpret market movements, making them relevant and effective for our project.

2. Market Regime Classification

Market regimes were determined based on relationships between moving averages:

- **Bull:** MA6 > MA10 > MA20
- **Bear:** MA6 < MA10 < MA20
- **Sideways:** Any other configuration

3. Model Selection and Training

We chose the **Random Forest Classifier** for its:

- Ability to handle non-linear relationships.
- Resistance to overfitting due to ensemble learning.
- Robustness with a mix of numerical features.

The model was trained on historical data, with 90% of the data used for training and the last 10 days reserved for testing. The target variable was the classified market regime (Bull, Bear, Sideways).

To explain to a non-technical person why a **Random Forest Classifier** was chosen and what it does in this code, you could use the following analogy and explanation:

Why was Random Forest chosen?

Imagine you're trying to decide if a company is doing well based on advice from several experts. Instead of relying on just one expert, you gather a group of experts (a "forest") and ask each of them to give you their opinion based on what they know. Each expert looks at different parts of the company's performance (like revenue, customer feedback, and stock trends). You then combine their opinions to make a final decision.

The **Random Forest Classifier** is like this group of experts. It's a machine learning algorithm that builds many "decision trees," where each tree is trained to make predictions based on a subset of the data. By averaging the predictions of all the trees, Random Forest provides a more reliable and accurate outcome, reducing the risk of making mistakes caused by relying on just one decision tree.

What does it do in this project?

In this project, the goal is to predict whether the stock market is in a **Bull**, **Bear**, or **Sideways** regime (similar to understanding if the company is thriving, struggling, or just steady). Here's how Random Forest helps:

1. **Analysing Patterns:** It examines patterns in historical stock data, like the stock's closing prices, moving averages, and volume, to learn how these factors relate to the market regimes.

2. **Learning from the Past:** It trains itself using historical data to recognize what features (like high moving averages or changes in volume) typically correspond to Bull, Bear, or Sideways markets.
3. **Making Predictions:** Once trained, it predicts the market regime for unseen data—both recent historical data and even future dates (based on assumptions).
4. **Why Reliable?** Because it uses multiple "decision trees" and combines their results, the predictions are robust and less prone to errors caused by noise or anomalies in the data.

We chose Random Forest because it works like a team of experts. It's reliable, works well even when there's a lot of data, and can handle complex relationships between different factors in the stock market. This makes it a good choice for identifying patterns in stock prices and predicting market trends.

4. Future Predictions

Using the trained model, we predicted regimes for business days in **November and December 2024**. Features for future predictions were based on the last available data point, assuming static conditions.

Results

Historical Performance

The Random Forest model accurately classified recent market regimes, with predictions aligning closely with observed trends. Test data predictions were added to the dataset for validation.

Future Predictions

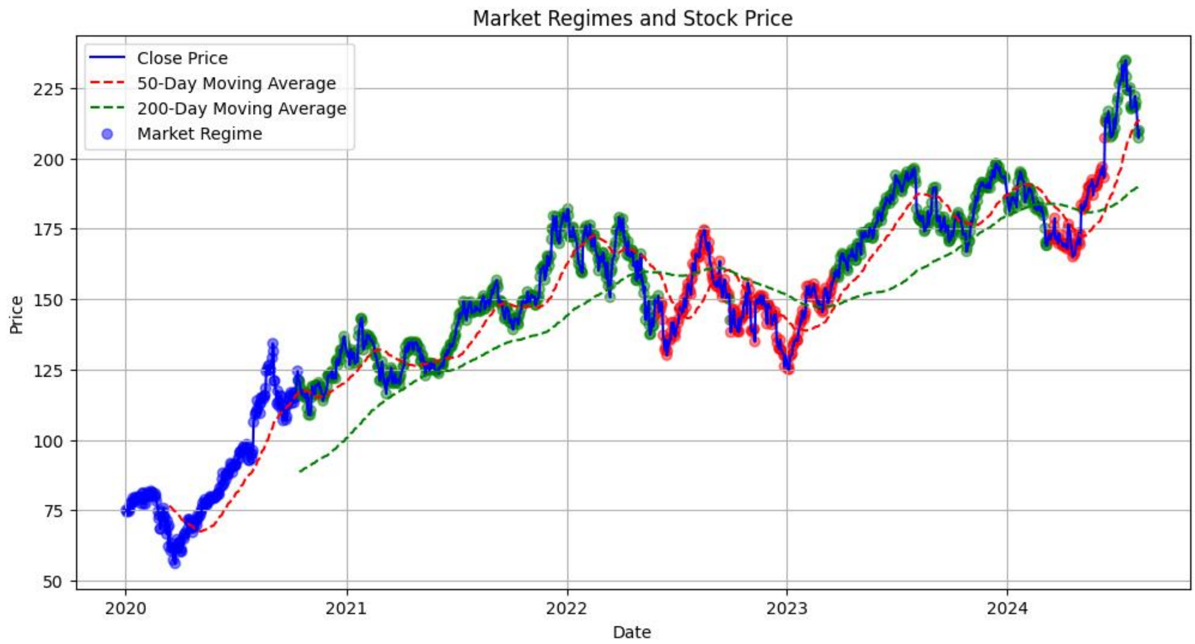
Predicted regimes for November and December 2024 suggest:

- **Bull markets** dominated early November.
- **Sideways markets** appeared in mid-November.
- Potential **Bear markets** emerging in late December.

Visualization

A comprehensive visualization was generated:

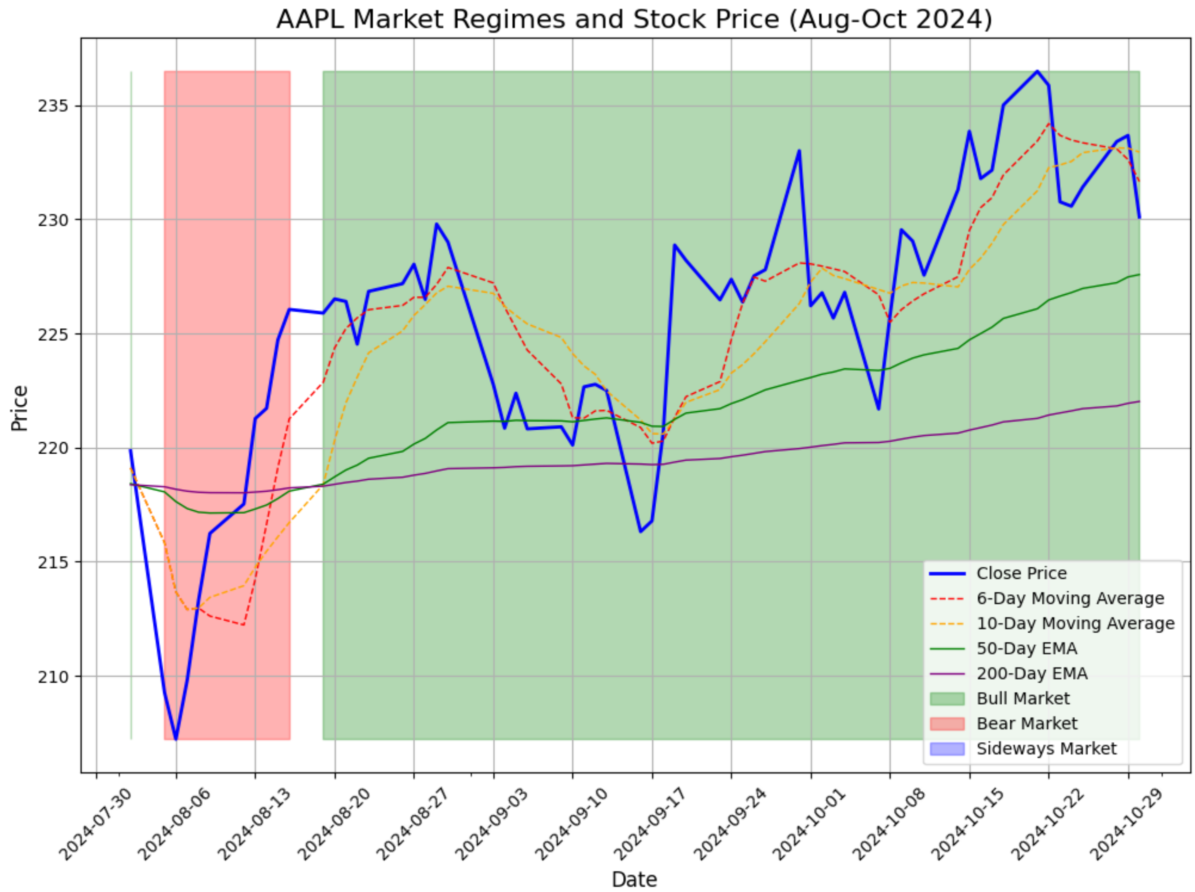
- Historical closing prices were plotted alongside predicted regimes.
- Future predictions were marked with distinct regime bands (Bull, Bear, Sideways).
- Color-coded regions highlight market trends, providing an intuitive representation.



A) <u>Market Regimes and Stock Price</u>

- **Details**:
  - It overlays the close price of a stock with 50-day and 200-day moving averages (MAs).
  - Market regimes (likely indicating bull, bear, or sideways markets) are represented as blue dots.
- **Insights**:

1. **Trend Identification**:
   - The stock has experienced a steady upward trend from 2020 to 2024, as indicated by the close price.
   - The 200-day MA provides a long-term trend line, while the 50-day MA shows shorter-term fluctuations.
2. **Market Regimes**:
   - The market alternates between bull and bear phases, with distinct periods of corrections or consolidations visible where price dips below the moving averages.
   - Regime changes coincide with crossing points of the 50-day and 200-day MAs (a golden or death cross).



B) <u>AAPL Market Regimes and Stock Price</u>

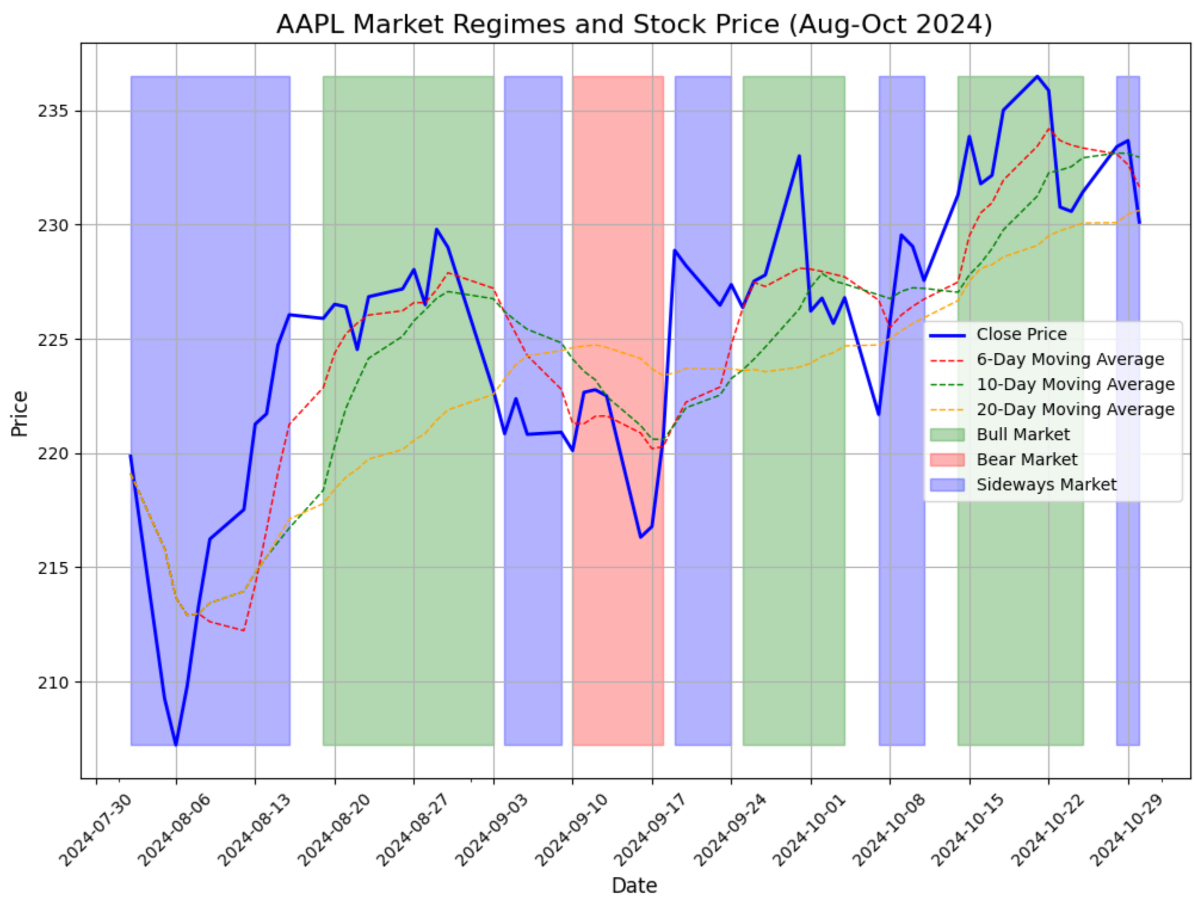
- **Details**:
  - Adds shaded regions for different market regimes:
    - **Green**: Bull market
    - **Red**: Bear market
    - **Blue**: Sideways market
  - Incorporates the 50-day and 200-day MAs to track trends.
- **Insights**:

1. **Bull Markets (Green Shaded Areas)**:
    - Clear uptrends occur during bull markets, with the price staying above both the 50-day and 200-day MAs.
    - These periods align with higher investor confidence and increased buying pressure.
2. **Bear Markets (Red Shaded Areas)**:

- Price falls significantly during bear markets, dipping below the moving averages.

3. **Sideways Markets (Blue Shaded Areas)**:

    - During sideways markets, the price oscillates within a range, with minimal directional movement.
    - This indicates indecision among market participants, often a precursor to larger trends.



C) <u>AAPL Market Regimes (Aug-Oct 2024)</u>

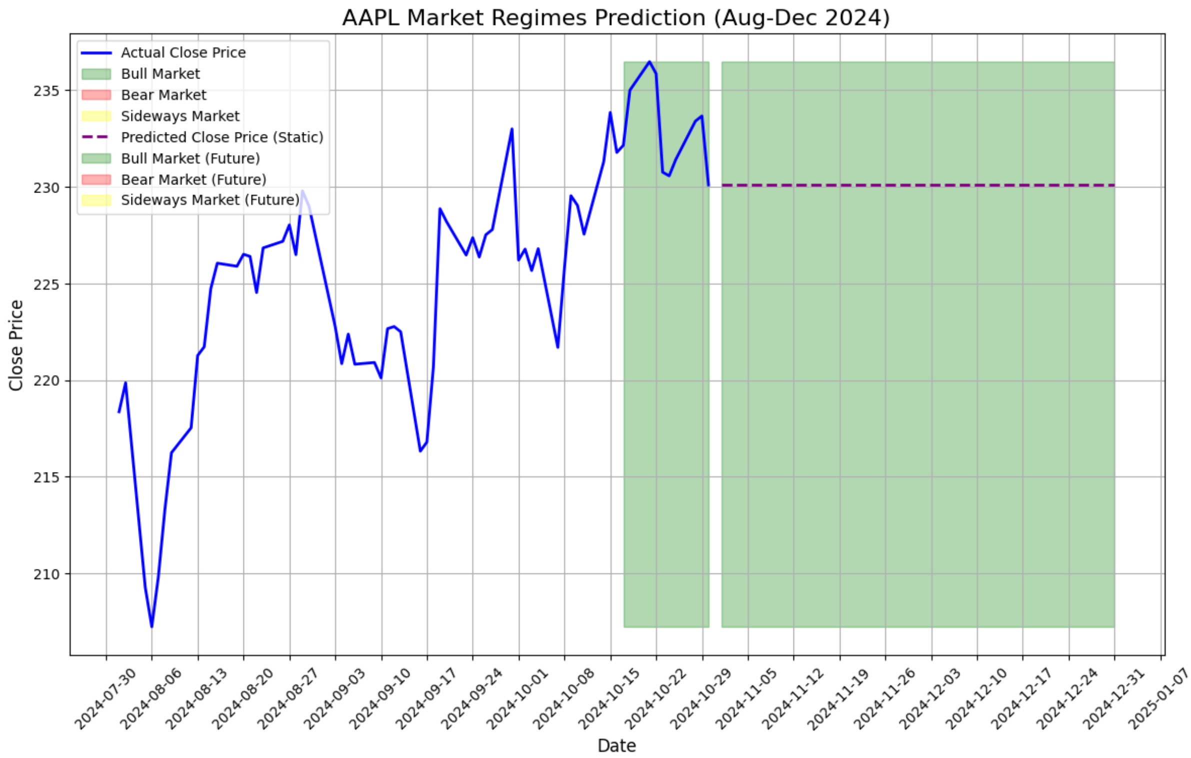
- **Details**:
  - A more granular view, incorporating additional MAs (6-day, 10-day, 50-day, and 200-day EMA).
  - Shaded regions show market regimes for a shorter period.
- **Insights**:

1. **Recent Bull and Bear Cycles**:
   - A sharp bear market (early August 2024) is followed by a sustained bull market (mid-August to October 2024).

2. **Short-Term Trends**:
   - The 6-day and 10-day MAs react quickly to price movements, providing early signals of trend changes.
   - The 50-day and 200-day EMAs provide long-term confirmation of sustained trends.

3. **Current Outlook (Oct 2024)**:

- The stock remains in a bull market as of October 2024, with price above all MAs.
- The convergence of shorter-term MAs suggests a potential slowdown or consolidation phase.



D) <u>AAPL Market Predictions (November and December)</u>

Here we can clearly see that our model has predicted that the apple stock will carry on a bullish trend in the month of November and December indicating that users or investors should buy more stocks of Apple(AAPL) to gain some profit by the end of December. The green colour in our graph indicates that the market will be following a bullish trend over the two months.

Stakeholders

This project serves a range of stakeholders, including:

- **Investors:** To identify optimal entry and exit points in the market.
- **Financial Analysts:** To support strategic decision-making with data-driven insights.
- **Portfolio Managers:** To adjust asset allocation based on predicted market regimes.
- **Academic Researchers:** To study the effectiveness of machine learning in financial predictions.

Conclusion

This project successfully demonstrated the ability to classify market regimes and predict future trends using a Random Forest Classifier. By leveraging historical data and advanced feature engineering, the model provides actionable insights that could guide investment strategies. Beyond the realm of investments, this predictive capability can assist individuals in their day-to-day financial planning and decision-making. For example, retail investors can use these insights to time their stock purchases or sales effectively. Business owners can anticipate market shifts to align their inventory or investment strategies with anticipated economic conditions. Additionally, financial advisors can offer tailored advice to their clients, ensuring strategies align with upcoming market trends. This approach empowers a wide range of people to make informed and confident decisions, enhancing financial literacy and stability in their daily lives.

Next Steps to be implemented!

- Extend the analysis to other stocks or indices.
- Evaluate the model using out-of-sample data for further validation.
- Incorporate external factors such as macroeconomic indicators or news sentiment.
