OPTI_WEALTH

STOCK PORTFOLIO RECOMMENDATION BASED ON RISK TOLERANCE AND CASH COMPOUNDING.

Authors:

1. Fanice Andayi
2. Jimmy Mumo
3. Kevin Riungu
4. Phyllis Kiruki
5. Cynthia Chelangat
6. Edwin Mwai

- BACKGROUND OF THE INDUSTRY

The investment landscape has evolved significantly over the years, driven by advancements in technology, changing market dynamics, and shifting investor preferences. Traditional investment approaches often relied on static asset allocation strategies that did not adapt to individual risk profiles or market conditions. However, with the advent of data analytics, machine learning, and financial modeling techniques, there is a growing opportunity to provide personalized investment advice tailored to investors' unique needs.'

- INTRODUCTION

In a world where data and algorithms drive financial decisions, envision having a personalized financial advisor at your fingertips. Instead of a magic wand, this system harnesses the power of data and algorithms to craft investment recommendations tailored specifically to you. Imagine inputting your risk tolerance and investment objectives, and presto! You receive customized portfolio suggestions designed not only to maximize returns but also to incorporate compounding strategies for long-term wealth growth. It's akin to having a financial genie fulfilling your wishes, but in the language of data and analytics!
Amidst the intricate landscape of financial markets today, investors grapple with the challenge of optimizing their portfolios to meet long-term financial objectives while effectively managing risk. Without personalized guidance, investors may find it daunting to navigate the array of investment choices available, potentially leading to

less-than-ideal outcomes. This highlights the necessity for a sophisticated recommendation system capable of analyzing individual risk preferences, investment timelines, and market dynamics to offer tailored portfolio suggestions.

- <u>PROBLEM STATEMENT</u>

1. Prevailing Circumstance

Currently, many investors lack access to personalized investment advice tailored to their risk tolerance levels and financial objectives. They often resort to generic investment strategies that may not align with their individual needs, leading to potential underperformance or excessive risk exposure.

2. Problem We're Trying to Solve

Our project aims to address this gap by developing an intelligent recommendation system that leverages machine learning algorithms and financial modeling techniques to provide personalized portfolio recommendations. By considering an investor's risk tolerance, investment horizon, and market conditions, we aim to optimize portfolio allocations and compounding strategies to maximize returns while mitigating risk.

3. How the Project Aims to Solve the Problem

Through data-driven analysis and advanced algorithms, our project seeks to empower investors with actionable insights that align with their unique financial goals and risk preferences. By harnessing historical stock price data, financial statements, and market indices, we aim to build a robust recommendation system capable of dynamically adjusting portfolio allocations and compounding strategies to optimize long-term wealth accumulation.

- <u>OBJECTIVES</u>

1. Main Objective:

The primary goal is to create a tailored recommendation system that enhances portfolio allocations and compounding strategies by considering individual risk tolerances and investment timelines.

2. Specific Objectives:

1.Employ machine learning algorithms to evaluate historical stock prices, financial records, and market indicators for risk assessment and the identification of appropriate investment prospects.
2.Create adaptive portfolio optimization methods that modify asset distributions according to evolving market dynamics and individual choices.
3.Integrate compounding tactics to optimize wealth growth over time and boost portfolio returns.

- <u>NOTEBOOK STRUCTURE</u>

1. Business Understanding
Provide a comprehensive overview of the investment landscape and the need for personalized investment recommendations
Clearly define the problem statement and the objectives of the project
2. Data Understanding
Identify and describe the relevant data sources, including historical stock prices, financial statements, and market indices
Understand the key features and characteristics of the data that will be used for the recommendation system
3. Data Cleaning
Perform necessary data cleaning and preprocessing steps to ensure the data is ready for analysis
Handle missing values, outliers, and any other data quality issues
4. Exploratory Data Analysis
Conduct in-depth analysis of the data to gain insights into the relationships between different variables
Identify patterns, trends, and potential drivers of investment performance
5. Data Preparation
Engineer relevant features from the raw data to be used as inputs for the machine learning models
Perform any necessary data transformations, scaling, or normalization
6. Modeling
Develop and train machine learning models, such as portfolio optimization algorithms and compounding strategies, to provide personalized investment recommendations
Experiment with different algorithms and techniques to find the most effective approach
7. Evaluation
Assess the performance of the recommendation system using appropriate metrics, such as risk-adjusted returns, portfolio diversification, and investor satisfaction
Identify areas for improvement and fine-tune the models accordingly
8. Conclusion, Recommendations, and Next Steps
Summarize the key findings and insights from the project
Provide actionable recommendations for implementing the personalized investment recommendation system
Outline the next steps and potential areas for future development and enhancement

BUSINESS UNDERSTANDING

As a prominent financial services provider in Kenya, we understand the crucial role of guiding clients towards informed investment decisions in a complex and volatile market setting. Clients seek tailored investment advice aligned with their financial objectives, risk preferences, and investment timelines. To meet these demands, our goal is to develop an advanced Stock Portfolio Recommendation System utilizing machine learning algorithms to analyze market data and optimize portfolio allocations. By integrating compounding strategies and dynamic optimization techniques, we aim to enhance long-term wealth accumulation for clients while fostering trust and loyalty. This initiative reflects our dedication to innovation, client-centricity, and market leadership in the financial services sector.

1. Stakeholders:

Investors: Individuals in search of personalized investment guidance tailored to their risk tolerance and financial goals.
Financial Advisors: Professionals seeking to strengthen client relationships by providing sophisticated portfolio recommendations based on data-driven insights.
Financial Institutions: Entities aiming to set themselves apart and attract clients through innovative investment solutions.

2. Metrics of Success:

Portfolio Return vs. Benchmark: Targeting a 3% annual outperformance against a relevant benchmark like the S&P 500.
Risk-adjusted Returns: Aiming for a Sharpe ratio exceeding 0.8 to ensure superior risk-adjusted performance compared to the benchmark.
Portfolio Diversification: Ensuring that at least 90% of portfolios meet diversification standards by being well-spread across various asset classes and industries.

DATA UNDERSTANDING

1.Data Source:

The dataset used for this project is obtained from Yahoo Finance through web scraping. It contains historical stock market data for the top 10 companies listed on the Nasdaq exchange, including Apple (AAPL), Starbucks (SBUX), Microsoft (MSFT), Cisco Systems (CSCO), Qualcomm (QCOM), Meta (META), Amazon.com (AMZN), Tesla (TSLA), Advanced Micro Devices (AMD), and Netflix (NFLX).

2. Data Size:

The dataset consists of 54,801 instances, each with 7 features. Each row represents the stock market data for a specific company on a given date. The features include 'Ticker', 'Date', 'Close/Last', 'Volume', 'Open', 'High', and 'Low'.

Dataset Columns and Description:

Column Name	Description
Ticker	The name of the company whose stock market data is recorded.
Date	The date in yyyy-mm-dd format representing the trading date.
Close/Last	The closing price of the stock at the end of the trading day.
Volume	The number of shares traded in the day.
Open	The opening price of the stock at the beginning of the trading day.
High	The highest price of the stock reached during the trading day.
Low	The lowest price of the stock reached during the trading day.

Relevance to the Project:

The dataset provides valuable insights into the historical performance of top companies listed on the Nasdaq exchange. It includes essential features such as stock prices (open, close, high, low) and trading volumes, enabling analysis of market trends, price movements, and trading activity over time. The 'Date' column serves as a temporal reference, allowing for time-series analysis and trend identification. By leveraging this dataset, the project aims to develop a portfolio recommendation system that utilizes historical stock market data to generate informed investment recommendations for users, aligning with the project's objective of providing valuable insights for investment decision-making.

DATA CLEANING

a) Importing necessary libraries for data manipulation, analysis, visualization, modeling, and evaluation. Here's a breakdown of the libraries that we will use;

- yfinance: Used for fetching historical stock market data from Yahoo Finance.
- pandas: Essential for data manipulation and analysis, providing data structures and functions to work with structured data.
- numpy: Used for numerical computations and operations on arrays and matrices.
- matplotlib.pyplot: Enables data visualization, particularly for creating plots and charts.
- seaborn: A library for advanced data visualization, offering a high-level interface for attractive and informative statistical graphics.
- train_test_split: From scikit-learn, used to split data into training and testing sets for machine learning models.
- LinearRegression: Also from scikit-learn, for implementing linear regression models.
- mean_squared_error: A metric from scikit-learn to evaluate the performance of regression models.
- datetime: Allows working with dates and times in Python.
- statsmodels.api: Useful for time series analysis and modeling.
- warnings: Enables the handling of warnings in Python.

By importing these libraries, the code sets up the environment for data retrieval, manipulation, visualization, modeling, and evaluation, laying the foundation for a comprehensive data analysis and modeling workflow.

b) We download the historical stock market data for the top ten companies from Yahoo Finance, specifically Apple (AAPL), Microsoft (MSFT), Amazon (AMZN), Starbucks (SBUX), Advanced Micro Devices (AMD), Meta (META), Tesla (TSLA), Cisco Systems (CSCO), Qualcomm (QCOM), and Netflix (NFLX). The data is fetched for the period from January 1, 2000, to April 24, 2024.

The download_top_ten_data function takes a list of tickers, a start date, and an end date as parameters. It downloads the data for each ticker using Yahoo Finance's API and then combines the data into a single DataFrame using the pd.concat function.

The resulting DataFrame, combined_data, contains the stock prices and trading volumes for the top ten companies over the specified period. The DataFrame has a MultiIndex with two levels: 'Ticker' and 'Date'. The columns include 'Open', 'High', 'Low', 'Close', 'Adj Close', and 'Volume'.

The code then resets the index of the DataFrame, converts the 'Date' column to datetime format, and drops the index column. The final output is a DataFrame with a datetime index and the same columns as before.

EXPLORATORY DATA ANALYSIS(EDA)

Exploratory Data Analysis (EDA) using univariate, bivariate, and multivariate analysis techniques, you gain significant insights into our data, using opening and closing stock prices.
a) Using Univariate: This technique allows us to examine individual variables in isolation, to provide insights into the trend analysis, seasonality and cyclicality, volatility and risk, correlation between opening and closing prices, the outliers and anomalies and finally comparison across the companies.
b) Using Bivariate: By exploring relationships between pairs of variables, bivariate analysis helps uncover correlations, dependencies, and interactions between two variables. Checking correlations between stock price and trading volumes, then we create a figure with two subplots, one for the distribution of stock prices and another for the distribution of trading volumes. Each subplot uses a kernel density estimation (KDE) plot to visualize the distribution of the respective variable across different companies. The KDE plots are stacked by company, allowing for a comparison of the distributions across companies. The plots provide a visual representation of the density of the data, helping to identify patterns and trends in the stock prices and trading volumes.
c) Using Multivariate: This involves examining relationships between three or more variables simultaneously. We investigated the trends over time for individual companies and across the entire dataset.
The top-performing companies based on stock price growth are:

TSLA (Tesla) with a growth rate of 6.72%

The top-performing companies based on trading volumes are:

TSLA (Tesla) with a trading volume of 337,112,476,700

These companies have demonstrated strong growth and trading activity, indicating their potential for future success. However, it's essential to consider various factors, including market trends, economic conditions, and company performance, before making investment decisions.

DATA PREPARATION

Calculation of various techniques thus indicates moving averages for stock data. The possible findings from the calculations include;

1. Moving Averages:

   Identifying trends and potential reversal points in stock prices using moving averages like the 30-day moving average.

   Analyzing the relationship between short-term and long-term moving averages to gauge momentum and trend strength.

2. Standard Deviation (Price Volatility):

   Assessing the volatility of stock prices using the standard deviation, which can help in risk management and identifying periods of high volatility.

3. Relative Strength Index (RSI):

   Evaluating overbought or oversold conditions in the market based on the RSI values.

   Identifying potential trend reversals or continuation patterns in stock prices.

4. Moving Average Convergence Divergence (MACD):

   Understanding the relationship between short-term and long-term moving averages to signal potential buy or sell opportunities.

   Analyzing momentum and trend changes in stock prices.

5. Stochastic Oscillator:

   Assessing the momentum and strength of stock price movements based on the Stochastic Oscillator values.

   Identifying potential trend reversals or continuation patterns.

6. Volume Moving Averages:

   Analyzing trends in trading volumes using moving averages to understand changes in market participation and interest.

7. Bollinger Bands:

   Identifying potential overbought or oversold conditions in stock prices based on the upper and lower Bollinger Bands.

   Assessing price volatility and potential breakout points.

8. Lagged Variables:

   Creating lagged variables like CloseLagged_1 to analyze the relationship between current and past stock prices.

   Understanding how past price movements may influence future price trends.

We can gain insights into stock price trends, volatility, momentum, and potential trading opportunities in the market. These findings can help in making informed investment decisions and developing trading strategies based on our historical stock data analysis.

MODELING