Passive Investing Analysis & Predictive Modelling

By Juan Moctezuma-Flores

What is the Purpose of this Project?

- The author's goal is to motivate other individuals to passively invest in index funds or exchange traded funds (ETFs) for long-term / low-fee monetary growth that comes with a low-level of risk.
- The purpose of the analysis & the prediction models is to mathematically & statistically prove that the funds or investment vehicles included in this project have the tendency to increase in value or price with respect to time.
- The author is not a financial advisor and although investing responsibly is highly encouraged for any individual, please do your own research and planning. Use the author's content and insights to support your investing decisions but please do not solely base your purchasing choices on these models.
- This project's GitHub includes:
  - Power Point presentation (Summary with URL links that lead to Jupyter Notebooks)
  - Financial analysis & visualization compiled using Python 3 / Jupyter Notebook:
    - Parts 1 through 3
  - Predictive models powered by Machine Learning (ML) using Python 3 / Jupyter Notebook:
    - Parts 4 through 6 (Random Forest Classifier)
    - Parts 7 through 9 (Simple Linear Regression)
  - Price Raw data CSV files – extracted from *Yahoo Finance*
  - Deliverables (Excel files):
    - Price analysis & visualization per fund (12 files in total)
    - Cleaned or processed historical data source (1 file)
    - Historical Data Storytelling (1 file)

Important note: Please skip slides 4 – 10 if you already have investment knowledge, as these slides contain basic descriptions, definitions, & concepts for beginners.

Basic Terms & Definitions

- Stocks are units of ownerships that companies sell to investors in the Stock Market (in this case – the U.S. Stock Market) throughout a brokerage or retirement account (for instance, *Charles Schwab* or *Fidelity*).
- The leading stock markets or stock exchanges are New York Stock Exchange (NYSE) and the Nasdaq.
- An investor is anyone that buys units at any given time to own a portion of that company and get a return in their investment.
- An investor can also sell those units at any given time and for whatever reason (stock's bad performance, necessity of cashing out, etc.).
- The term 'share' refers to the actual number of units or financial instrument that represents the ownership from a particular stock. Note that both terms can be used interchangeably.
  - For example, 'John Doe purchased 5 shares from Apple (in this case the company Apple has stock available at the Stock Market).
  - For example, 'Jane Doe invested in Amazon and Google stock (in this case we know that this investor bought shares from those companies).

What are Index Funds?

- Index funds are a big collection (or diversified bundle) of multiple stocks, bonds or other securities from different industries where investors buy ‘a little bit from everything’. Index funds can also be purchased through a brokerage account.
- Buying index funds allow investors to avoid buying individual stocks as well.
- Investors are safer with index funds because these have a lower level of risk (less chance of losing large sums of money compared to buying individual stocks). If one or few out of the many companies within the bundle performs poorly or goes bankrupt, losses won’t be as severe.
- For example, the Standard and Poor (S&P) 500 is the index of the largest publicly traded companies. You can’t invest directly into the index, but you can invest into the index fund – which will buy shares for the top 500 companies and will mirror the performance of the S&P 500.
- Note that there are ‘targeted’ index funds that track the index of specific sectors such as technology, pharmaceuticals, energy, etc. These could track hundreds of companies.

Stock vs Index Fund Example

Apple Stock

≠

Fictional ‘US Ten’
Index Fund

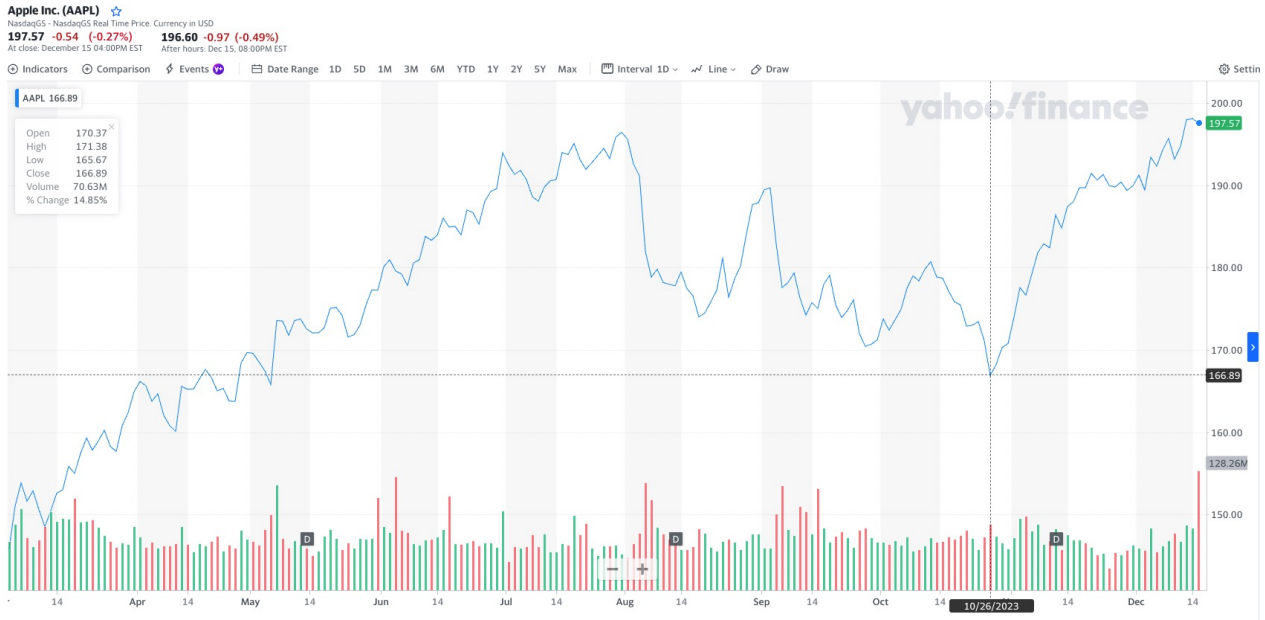
- Microsoft
- Apple
- Berkshire Hathaway
- Google
- Amazon
- Facebook
- Visa
- Chase
- Walmart
- Johnson & Johnson

Goals and Strategies

- The investors' goals are to increase monetary value (net worth) via dividends (payments given by a company's due to good performance / profit increase) or by simple letting the stock's value or price to increase with respect to time.
- There are different investment strategies:
  - Buy & Hold
  - Dividend Investor
  - Speculator
  - Trader
- Each strategy presents a different level of risk, time-frame, and return, and strategies may be combined, however, the author of this project focuses solely on long-term / low-risk / *Buy & Hold* strategy.

Important Notes

- The Stock Market is not a get rich quick scheme. Becoming a successful investor requires responsibility, time, strategic planning, research, and knowing your risk-tolerance level (not letting your emotions intervene).
- Remember that investors lose value or money when stocks are performing poorly. However, stock prices will always fluctuate (towards positive or negative), simply because a stock's value decreases in short period of time, it doesn't imply a long-term loss.



Graph of Apple's stock price fluctuation during 2023

Advantages & Disadvantages of Index Funds

Pros	Cons
• Passive Investing - You don't have to actively manage your funds. You simply buy shares and hold these for many years. You simply let these 'grow' by not doing anything. • Convenient for beginners & individuals that are not interested in actively investing or worrying about the macroeconomic conditions. • Dependable returns or dividends. • Many index funds have a low yearly expense ratio, and more tax efficient for long-term. • Diversification – which means that the risk gets balanced, and your portfolio experiences less volatility. • You don't need excessive cash to start investing.	• Lack of flexibility – you can't select which stocks to include or exclude on the S&P 500. • Index funds are not designed for short-term high rewards. Meaning that you can't 'beat the market' by passively investing. • Tracking error – an index fund will never perfectly track an index. • Not ideal for active investors or daily traders.

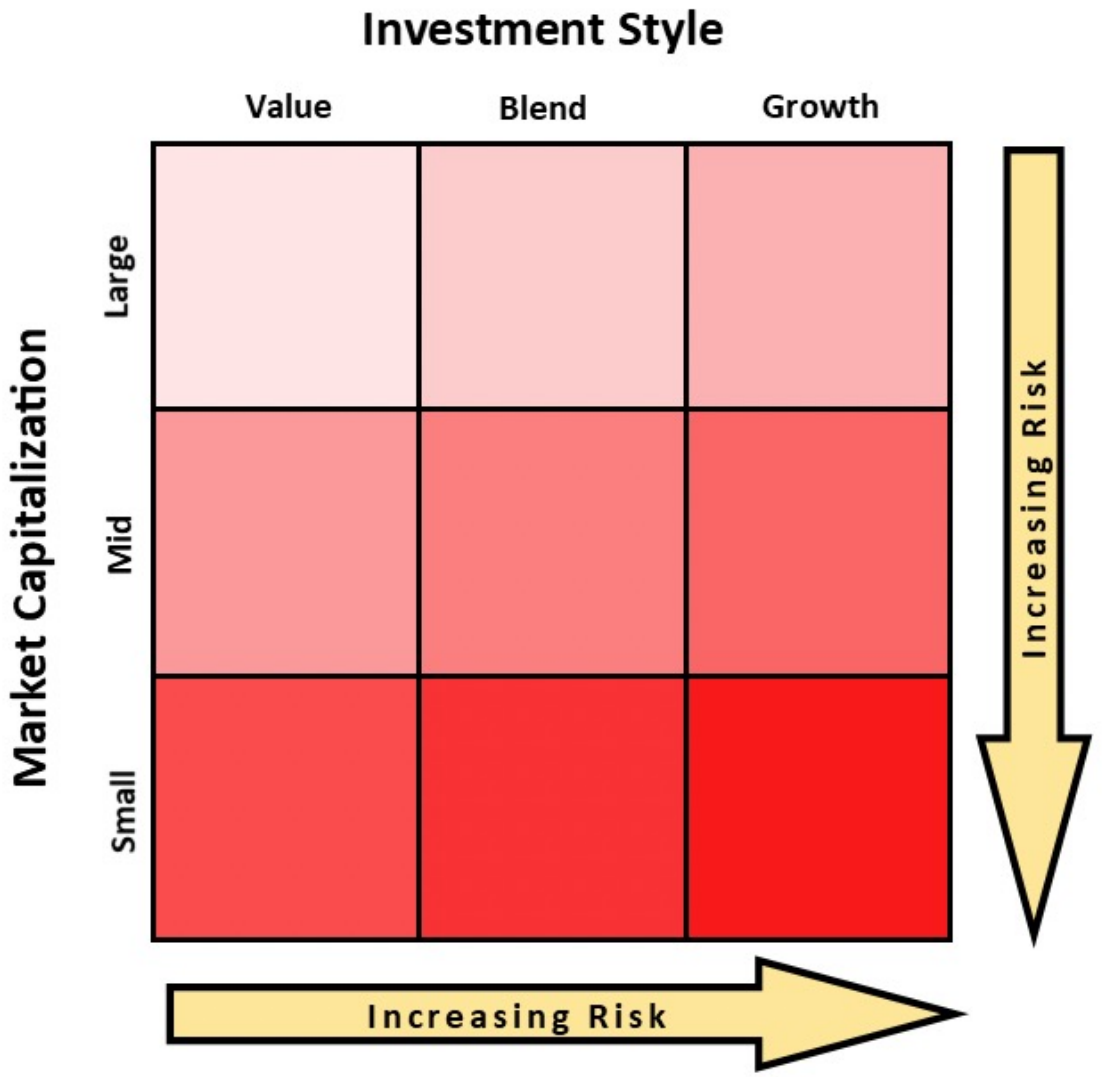
Variables to Keep in Mind When Purchasing Index Fund Shares

- Funds' annual expense ratio. The platform where you invest charges the investor a fee (typically small or affordable for index funds) for holding x number of shares of an ETF.
- Minimum initial investment. Some funds require investors to invest a large amount as minimum, but those funds or ETFs will be excluded in the analysis; As the goal is to promote passive investing for anyone.
- Transaction fees. Brokerage or retirement platforms charge investors transaction fees when buying or selling shares of certain funds.
- Knowing what are your financial goals are and investment time horizon.
- Knowing your risk tolerance (how much are you willing to invest in the market). It is tentative to sell shares when prices decrease. Although index funds are known for being low risk (but slow growth).
- Your return expectation. Some funds have a higher dividend yields.

How were the Index Funds (or ETFs) Selected for the Analysis?

- It is common for most people to select which index funds or stocks based on popularity or speculation – this is a bad strategy.
- The author decided to include the top 4 index funds / ETFs by fund size (in USD) or Assets Under Management (AUM) as in 2023, and by the following 3 style categories from the *Morningstar Style Box*:
  - Large Value: Funds that invest in large-sized companies that have slow growth and low valuations (low price ratios and high dividend yields).
  - Large Growth: Funds that invest in large-sized companies projected to grow fast. These tend to have high volatility, and high valuations (high price ratios & low dividend yields). These are known for being 'high risk but high reward'.
  - Large Blend: Is a combination of Value & Growth. These funds represent the overall U.S. Stock market and invest in across the spectrum of U.S. industries.
- Why do we exclude small and mid *Morningstar* sizes? These have less exposure to the U.S. Market Equity and hence less diversification (which implies a higher risk).

Morningstar Style Box



Funds or ETF per Category

Fund Category	Ticker	Name	Expense Ratio per Yr	Size / Assets Under Management
Large-Cap Blend	IVV	iShares Core S&P 500 ETF	0.03%	$483,801,000.00
	VOO	Vanguard Standard and Poor 500 ETF	0.03%	$403,534,000.00
	SPY	SPDR Standard and Poor 500 ETF Trust	0.09%	$377,876,000.00
	VTI	Vanguard Total Stock Market ETF	0.03%	$349,082,000.00
Large Value	IWD	Invesco QQQ Trust	0.19%	$53,382,900.00
	SCHD	Vanguard Growth Index Fund	0.06%	$52,257,000.00
	VYM	iShares Russell 1000 Growth ETF	0.06%	$50,907,700.00
	DGRO	iShares Core Dividend Growth ETF	0.08%	$25,155,500.00
Large Growth	QQQ	Invesco QQQ Trust	0.20%	$231,219,000.00
	VUG	Vanguard Growth Index Fund	0.04%	$104,787,000.00
	IWF	iShares Russell 1000 Growth ETF	0.19%	$81,087,700.00
	IVW	iShares S&P 500 Growth ETF	0.18%	$36,825,600.00

What are the Analyses About?

- The purpose is to numerically compare the funds from the 3 categories and discuss which investment vehicles are better from a passive-investment perspective.
- The statistical models for each analysis performed for the 3 fund categories check the following:

1. Average volume of shares or index funds' units traded with respect to time.
2. Highest/Lowest/Open index fund prices traded with respect to time.
3. Market Capitalization (Closing share price x number of outstanding shares).
4. Trends using 50-days & 200-days moving averages (MA) - Do shares' prices tend to increase in the long term?
5. Stock returns & volatility for annualized & daily returns – estimate risk percentage.
6. Correlation or mutual relationship among variables – for instance, are increasing share prices will cause the fund's trading volume to increase?

Large Blend Fund Category - Analysis Summary

- Due that the 4 ETFs have low volatility, favorable trend patterns (upward - increasing value as time passes), & overall low risk / great return ratio, any of these stocks would be worth buying and holding for many years. The only exception would be SPY. The reason being that it has the highest expense ratio among the 4 ETFs, despite having slightly better returns - on the long term, a high expense ratio is an important variable to keep in mind as as a passive investor looking for long-term growth.
- VOO & IVV follow an IDENTICAL price (open, high, low) therefore investing in either one won't make any difference and the same applies for long-term returns. Among VOO/IVV (both S&P 500 ETFs) & VTI (Total Stock Market), VTI has slightly more volatility, therefore IVV or VOO are less riskier investment options. Note that IVV, VOO & VTI have low expense-ratios (which is essential for 'buying & holding' strategies).
- In addition, these index funds are perfectly correlated! The reason being that these funds are NOT diversifying away (they all posses almost the same holdings). Although using a correlation coefficient on its own may not help to predict future stock returns, we know that these funds will move (or trend) in the same direction.
- **Based on these facts, the best long-term investment options (in order from greatest to least) within the 'Large Blend' fund category for passive investors are:**
    1. IVV or VOO - both are identical (investing in both at the same time is not recommended due to lack of diversification)
    2. VTI
    3. SPY
- See: https://github.com/Juan-Moctezuma/Passive_Investing_Analysis_-_Predictive_Modelling/blob/main/Part1-Large_Blend_Fund_Category-Financial_Analysis.ipynb

Large Value Fund Category - Analysis Summary

- Due that the 4 ETFs have low volatility, favorable trend patterns (upward - increasing value as time passes), & overall low risk / great return ratio, any of these stocks would be worth buying and holding for many years. The only exception would be IWD. The reason being that it has the highest expense ratio among the 4 ETFs, despite having slightly better returns - on the long term, a high expense ratio is an important variable to keep in mind as as a passive investor looking for long-term growth, because it adds up throughout the years. IWD would be a good investment for a short period of time.
- From a Market Cap perspective, SCHD has the best performance on the long term compared to VYM (2nd) & DGRO (3rd). Remember that SCHD, VYM & DGRO are dividend ETFs, and although DGRO is the least volatile and hence it has the least amount of risk, SCHD is currently (as in December 2023) a better investment as best performance (Market Cap) translates into a higher dividend yield. However, it is possible that in the future DGRO outperforms both SCHD & VYM, however historical data doesn't reflect that yet, maybe it will in a couple of years. Note that the 3 previously mentioned Dividend ETFs have low expense-ratios, which is essential for 'buying & holding' strategies.
- Note that when it comes to trending directions, since these 4 ETFs are highly correlated, these will move upward or downward depending on the economy (meaning that if there's a downturn - all of these will fall at the same time).
- **Based on these facts the best long-term investment options (in order from greatest to least) within the 'Large Value' fund category (High Yield Dividend Funds) for passive investors are:**

    1. SCHD
    2. VYM
    3. DGRO
    4. IWD

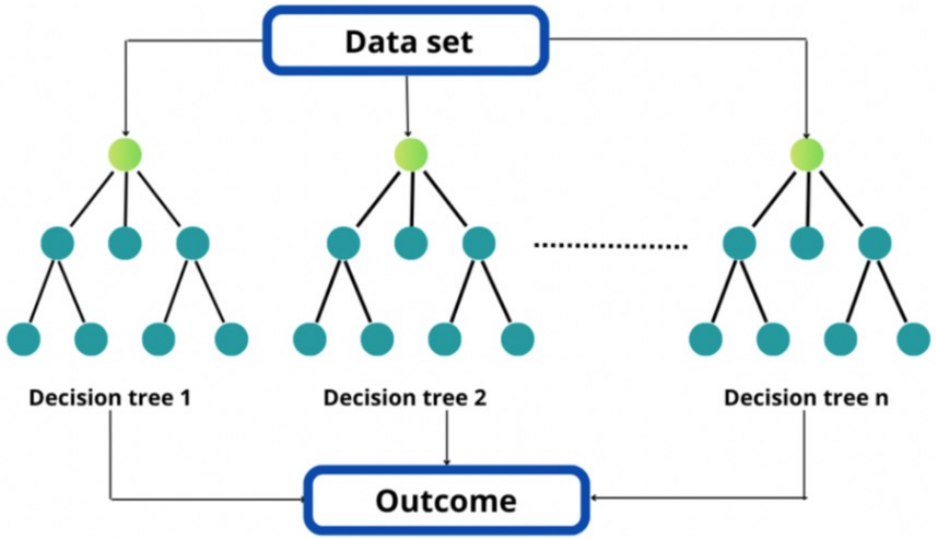
- See: https://github.com/Juan-Moctezuma/Passive_Investing_Analysis_-_Predictive_Modelling/blob/main/Part2-Large_Value_Fund_Category-Financial_Analysis.ipynb

Large Growth Fund Category - Analysis Summary

- The 4 ETFs have high volatility, and although these have favorable trend patterns (upward - increasing value as time passes), return ratio may vary, therefore only a few of these Growth Funds or ETFs would be worth buying and holding for many years. It's a bad strategy to purchase an excessive number of shares of Growth funds, unless you have a high-risk tolerance.
- When it comes to Market Cap, QQQ has the best performance on the long term compared to the rest, but that doesn't imply the risk will be low.
- In theory, QQQ & VUG should be the only options as these are the only that correlate positively between prices and volume, hence, you may at least get a 'hint' of the ETFs' outcome by looking at the trading volume, however, it is still bad strategy to only rely on positive correlation to make an investment decision. QQQ should be the most convenient ETF to hold, however, it has a high expense ratio of .20% which is high for long-term / passive investors. The next best choice should be VUG as it has an expense ratio of 0.04%. For the record, IWF & IVW have expense ratios of 0.19% & 0.18%.
- Note that when it comes to trending directions, since these 4 ETFs are highly correlated, these will move upward or downward depending on the economy (meaning that if there's a downturn - all of these will fall at the same time).
- **Based on these facts, the best long-term investment options (in order from greatest to least) within the 'Large Growth' fund category for passive investors are:**
    1. VUG
    2. QQQ
    3. IWF
    4. IVW
- See: https://github.com/Juan-Moctezuma/Passive_Investing_Analysis_-_Predictive_Modelling/blob/main/Part3-Large_Growth_Fund_Category-Financial_Analysis.ipynb

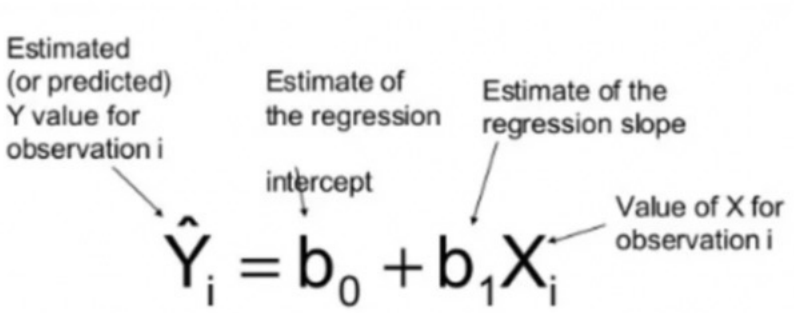
What is Random Forest Classifier?

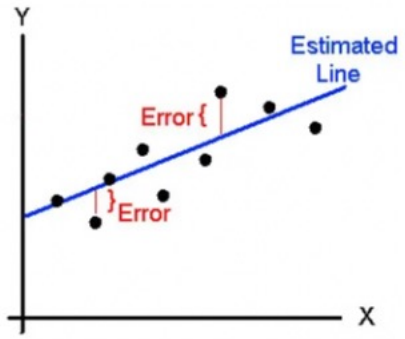
- A random decision forest is a machine learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time. Then when it comes to testing, it will try to perform a similar action.
- For classification tasks, the output of the random forest is the class selected by most trees. Therefore, the output of multiple decision trees reach a single result - in this case the output is the *probability* for share prices to increment in the future based on historical data patterns raging from 1 or 2 decades ago.



What is Simple Linear Regression?

- Linear regression is an algorithm that provides a linear relationship between an independent variable and a dependent variable to predict the outcome of future events.
- The regression model predicts the value of the dependent variable, which is the response or outcome variable being analyzed or studied & simulates a mathematical relationship between variables and makes predictions for continuous or numeric variables such as sales, salary, age, product price, etc.
- In this case, our prediction is based on the following question - Will share prices continue to increase on the long run?





All Fund Categories – Prediction Summary (Pt. 1)

- The random forest decision classifier / machine learning model was trained efficiently as it accurately predicted the odds of share price (from ‘close price’ & date data) increment in the future or 'tomorrow’.
- It is reasonable for the model to predict or output probabilities ranging between 50% & 60%, as realistically nobody (nor algorithms) can predict future prices with a high percentage of certainty.
- As seen in the financial analysis' Moving Average graphs - Jupyter Notebooks 1 through 3, Index Funds tend to increase in value with respect to time despite economic downturns or outlying events (such COVID Pandemic).
- The trained algorithm ‘notices’ the previously mentioned pattern and will attempt to make predictions based on historical data. On every category, multiple prediction ‘rounds’ were performed on testing data and only a few were inaccurate.

All Fund Categories – Prediction Summary (Pt. 2)

- The simple linear regression / machine learning model was trained efficiently as it accurately predicted the odds (or a trendline) of share price (from ‘close price’ & date data) increment in the future or 'tomorrow’.
- What this trained algorithm does is that it will try to fit in a ‘prediction’ line in-between the timeseries (testing) data; this prediction trendlines point ‘upward’ towards a *bull market* (Economic environment or times in which share prices are rising, hence encouraging investors to buy).
- Accuracy scores were higher than 70%, which is good & the models predict prices to continue increasing throughout the years, however, error scores are expected to be high as the market has plenty of ups & downs - therefore multiple data points' predictions are inaccurate.
- Remember that every observed index fund or ETF from this project is highly correlated, therefore it is statistically very likely for every fund (regardless of the Morningstar category) to get predicted as being in an ‘upward’ trend.

Linear Regression Model's Output for VOO (Historical Data from 2010 – 2023)

Please observe the red trendline from this sample graph.

All Fund Categories – Prediction Notebook Links

- Random forest decision classifier - Jupyter notebooks 4-6:
  - https://github.com/Juan-Moctezuma/Passive_Investing_Analysis_-_Predictive_Modelling/blob/main/Part4-Large_Blend_Fund_Category-Prediction_Random_Forest_Classifier_Model.ipynb
  - https://github.com/Juan-Moctezuma/Passive_Investing_Analysis_-_Predictive_Modelling/blob/main/Part5-Large_Value_Fund_Category-Prediction_Random_Forest_Classifier_Model.ipynb
  - https://github.com/Juan-Moctezuma/Passive_Investing_Analysis_-_Predictive_Modelling/blob/main/Part6-Large_Growth_Fund_Category-Prediction_Random_Forest_Classifier_Model.ipynb
- Simple linear regression model - Jupyter notebooks 7-9:
  - https://github.com/Juan-Moctezuma/Passive_Investing_Analysis_-_Predictive_Modelling/blob/main/Part7-Large_Blend_Fund_Category-Prediction_Linear_Regression_Model.ipynb
  - https://github.com/Juan-Moctezuma/Passive_Investing_Analysis_-_Predictive_Modelling/blob/main/Part8-Large_Value_Fund_Category-Prediction_Linear_Regression_Model.ipynb
  - https://github.com/Juan-Moctezuma/Passive_Investing_Analysis_-_Predictive_Modelling/blob/main/Part9-Large_Growth_Fund_Category-Prediction_Linear_Regression_Model.ipynb

Conclusion

- The financial (non-predictive) statistical models based on historical data went over average trading volumes, price fluctuation, market cap, trends using moving averages, stock returns & volatility, and correlation for blend, value & growth Morningstar categories (in which each category holds the top 4 largest index funds or ETFs based on assets' size). Although the initial analyses reflect patterns that look promising for investors, further modelling is needed to make data-driven decisions.
- Random forest classification M.L. models were trained (with historical data) and tested. Results from the model were realistic (50% - 60% chance of share prices to go up on the next day) because index funds tend to 'grow' in the long term although we can't accurately predict short-term price increments.
- Simple linear regression M.L. models were trained (with historical data as well) and tested. The models' predicted trendlines are projected towards a long-term upward direction. However, alike the random classification model, linear regression can't accurately predict short-term price increments either.
- **Is passive investing worth it?** Based on data from statistics and the outcomes from the M.L. models, it certainly supports the idea that responsibly buying shares from index funds is beneficial for the long-term financial wellbeing of any individual.

References Links

- https://www.analyticsvidhya.com/blog/2021/11/exploratory-data-analysis-on-uber-stocks-dataset/
- https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/
- https://copyassignment.com/random-forest-machine-learning/
- https://etfdb.com/etfs/size-and-style/
- https://github.com/shashimanyam/PREDICTING-STOCK-PRICES-WITH-LINEAR-REGRESSION/blob/master/Predicting%20Stock%20Prices%20with%20Linear%20Regression.ipynb
- https://www.investopedia.com/ask/answers/021015/how-can-you-calculate-volatility-excel.asp
- https://www.jrbcj.org/stylebox
- https://www.macroaxis.com/invest/stock-correlation
- https://www.marketcalls.in/python/build-a-correlation-matrix-using-python-pandas-and-seaborn.html
- https://www.merrilledge.com/article/company-size-why-market-capitalization-matters-ose#:~:text=Generally%2C%20market%20capitalization%20corresponds%20to,for%20less%20aggressive%20growth%20potential
- https://www.morningstar.com/content/dam/marketing/apac/au/pdfs/Legal/Stylebox_Factsheet.pdf
- https://www.morningstar.com/large-blend-etfs
- https://www.morningstar.com/large-growth-etfs
- https://www.morningstar.com/large-value-etfs
- https://portfolioslab.com/tools/stock-comparison
- https://www.programminginpython.com/stock-returns-and-volatility-analysis-with-python/
- https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-linear-regression/#:~:text=Linear%20regression%20is%20an%20algorithm,machine%20learning%20for%20predictive%20analysis
- https://statisticsbyjim.com/regression/mean-squared-error-mse/
- https://thecleverprogrammer.com/2021/06/22/r2-score-in-machine-learning/#:~:text=The%20R2%20score%20is%20a,predictions%20explained%20by%20the%20dataset