

Results

I have divided the results section into different modules for ease of understanding and stepwise clear implementation. The modules divided are mentioned below:

1. Equity Forecasting Model
2. Estimating Loss for Equities
3. Individual Risk Tolerance Calculator
4. Equity : Debt : Gold split Ratio calculator
5. Equity Suggester
6. Mutual fund Suggester & Expected Downside Risk Calculator
7. Recent Historical Gold Price Fluctuations
8. Portfolio Recommender
9. GUI for Investor interaction
10. Projected Portfolio Value Calculation and Plotting
11. Plotting graphs for Recommended Portfolio
12. GUI for Displaying Recommended Portfolio

Equity Forecasting Model

This module forecasts the future equities 'Open', 'High', 'Low' and 'Close' prices and stores the result in csv files for the equity suggester module. Based on the results of the different models used in the methodologies and experiments, the least error and the best results we were able to achieve was using the combination of Seasonal AutoRegressive Integrated Moving Average (SARIMA) and Generalized AutoRegressive Conditional Heteroskedasticity (GARCH), while the combination of SARIMA and Random forest ranked second and SARIMA and Light gradient boosting machine (LGBM) ranked third based on the evaluation metrics used.

As equity prediction is a complex task with loads of data trends, so rather relying on just one combination I combined all the four models to generate maximum accuracy and near-to result predictions. The model used for equity prediction is described in the steps below:

- After setting the date column as index, I separated the features necessary for the model, which is 'Open Price', 'High Price', 'Low Price' and 'Close Price' and added them to a new data frame for training to the model.
- For each feature, I trained the SARIMA model with parameters ($p=1$, $d=1$, $q=1$, $P=1$, $D=1$, $Q=1$ and $s=12$ (for monthly seasonality)).
- In the next step I generated the residuals for each feature from the above SARIMA model and trained all the features on GARCH model with parameters ($p=1$, $q=1$).

- For training the models LGBM and Random forest, I reshaped the residuals from the SARIMA and split the train test set as 80:20 with random state and trained all the features on this data with LGBM and Random Forest regressors.
- After training and saving the models, for forecasting I used `get_forecast` function to get the forecasts for SARIMA and GARCH models for the future 5 years (`forecast_period` = 1260 trading days).
- And for predicting on LGBM and Random forest models, I used the `predict` function with same `forecast_period` for each feature.
- For getting the final forecast I combined the predicted mean of SARIMA forecast, square root of GARCH forecast, LGBM forecasts and random forest forecasts for all the features and created a data frame for the same.

The above steps were added to a function and for providing the data, I looped through the metadata of all equities and divided the data based on its Market cap and after calling the function saved the forecasted files in the Predictions Data Folder for future use by the Equity Suggester Model.

Estimating Loss for Equities

Once we have got the predictions for all the equities future prices, the next step is to calculate the estimated loss associated with it. By calculating this we give the investor a picture of how actually their fund would perform if it went up or if it incurs losses then how much maximum loss would the portfolio suffer. This gives an estimate to the investor that how much risk is the investor willing to take in terms of returns he is expecting. It's important to understand the risk to reward ratio for every portfolio to invest smartly.

For calculating the estimated losses, I used the historical data of the equities and based on the period the user wants to invest got the volatility and calculated the maximum potential loss using risk metrics Value at Risk (VaR). The detailed steps are mentioned below:

- Using the metadata file I divided all the equities based on their Market cap and then for different time periods sliced the data for each equity.
- Using `pct_change()` calculated the percentage change for that time period on the Close prices of the equity.
- For calculations, I used the confidence level = 0.95(95%) which means that the probability of the portfolio value not going below certain calculated price is 95%.
- Then I calculated the tail threshold using the quantile of $(1 - \text{confidence level})$ of the `pct_change` values from the above.

- Then I filtered the returns which are above the `tail_threshold` and fit a Generalized Pareto Distribution (GPD) on the filtered returns to calculate the shape, loc and scale.
- Then using the `genpareto.ppf` function calculated the VaR using the shape, loc and scale from the previous steps.
- For each time period I calculated using the same way and stored in a different csv file to be used by Equity Suggester module later.

Individual Risk Tolerance Calculator

This module calculates the individual risk tolerance of an investor based on investor's financial situation and economic factors. The model I am using to calculate this is Gradient Boosting Machine (GBM) because GBM was able to provide the highest accuracy and least error out of the three regressor models I experimented before.

For understanding the investor's financial condition and adding it to the GBM to calculate the financial risk tolerance score and risk category, I put down a questionnaire to get inputs, these questions include:

- What's your Gender? – (M, F)
- Do you own a Vehicle? – (Y, N)
- Do you own a property? – (Y, N)
- Your Martial Status – ('Married', 'Single / not married', 'Civil marriage', 'Separated', 'Widow')
- What is your House Type? - ('House / apartment', 'With parents', 'Rented apartment', 'Municipal apartment', 'Co-op apartment', 'Office apartment')
- Your Age? – (int input age)
- Are you Employed? – (Y, N)
- Provide your Annual Income (in ₹) – (int input income)
- Provide your occupation from the options below – ('Managers', 'High skill tech staff', 'Accountants', 'Medicine staff', 'HR staff', 'IT staff', 'Sales staff', 'Core staff', 'Security staff', 'Cooking staff', 'Private service staff', 'Secretaries', 'Realty agents', 'Laborers', 'Drivers', 'Cleaning staff', 'Low-skill Laborers', 'Waiters/barmen staff')
- Provide your Education Qualification – ('Higher education', 'Academic degree', 'Secondary / secondary special', 'Incomplete higher', 'Lower secondary')
- How many Family Members (Dependants) do you have? – (int input family_members)
- What amount are you planning to invest (in ₹)? – (int input amount to invest)
- What is the term you want to invest for? – (1 month, 3 months, 3 months to 1 year, 1 to 3 years, 3 to 5 years)

After getting these inputs from the user, I processed the inputs by categorizing their occupation and education level and then calculated the cost of living (Annual income / Family members[dependants]). After processing these three fields, I mapped the inputs to its categorical form using the mapper_to_categorical function and then reshaped all the inputs to a 2D np array for predicted the output on the GBM regressor which I had trained during experiments and methodology section and saved as a pickle file. After predicting the financial score, I rounded it of to the nearest decimal.

For getting the risk category of the investor, I used the calculate_risk_category function and gave the financial risk tolerance score predicted as input and got the risk category on which the user will be suggested it's portfolio. This risk category will be further used by the Equity:Debt:Gold split ratio calculator to get the split ratio for the particular investor.

Equity : Debt : Gold split Ratio calculator

This module calculates the Equity : Debt : Gold split ratio for the investor based on its risk category it falls into using the previous module. The Equity : Debt : Gold spilt ratio determines what portion of the money the user wants to invest should be invested in Equity, Debt and Gold.

Equity which is investing in stock exchanges and Mutual funds which have their portfolio based on completely on Equities. Investing in equities can be highly risky because the stock market and exchanges are very volatile in nature and unpredictable. The market completely operates on demand and supply of that equity, as the demand rises people tend to buy more units of that stock and in turn increases the price of that equity, while lower demand results in people selling the units and resulting in stock price moving downwards.

Investing in debt means providing capital to large financial corporations, government or financial institutions in exchange of some fixed interest for a specific period. The interest you receive is after the maturity period and your capital would be secured during that period. This involves very less risk but also less returns, but it preserves your capital, so for investors seeking for less risk can invest more in debt then equity. There are lot of different ways to invest in Debt, most common one is Debt Mutual funds, where the capital is preserved and you get decent returns, some other ways to invest in debt are mentioned in the table below:

Investment_type	Expected Returns(%)	Risk
Fixed Deposits (FDs)	5 to 7	Low
Public Provident Fund (PPF)	7 to 8	Very Low
National Savings Certificates (NSC)	7 to 8	Very Low
Corporate Bonds	7 to 9	Medium to High
Government Bonds	6 to 7	Low
Treasury Bills (T-Bills)	4 to 6	Very Low
Bank Savings Schemes	4 to 6	Very Low
Non-Convertible Debentures (NCDs)	8 to 11	Medium to High
RBI Bonds	7 to 8	Very Low

Gold is recognized as a precious metal across the world, and over time it has prove that the prices of these metals always had an upper linear circuit, and mostly the fluctuation of gold usually is based on the country's economic factors, and big developing countries like India have always seen an upper circuit for gold and time on time the metal has prove it's worth. So, trading/investing in gold is a much more risk free investment than investing in equities, hence people falling into less risk category, should invest more in gold than equity to get good returns and hold for a longer period of time.

The Equity : Debt : Gold split ratio based on individuals risk tolerance is provided in the below table:

Risk_Category	Debt(%)	Gold(%)	Equity(%)	Equity Split (%)	
Low	70	20	10	Large Cap	5
				Mid Cap	0
				Small Cap	0
				Hybrid MF	5
Low to Medium	60	15	25	Large Cap	10
				Mid Cap	5
				Small Cap	0
				Hybrid MF	10
Medium	50	10	40	Large Cap	15
				Mid Cap	10
				Small Cap	5
				Hybrid MF	10
Medium to High	45	10	45	Large Cap	15
				Mid Cap	10
				Small Cap	10
				Equity MF	10
High	30	5	65	Large Cap	15
				Mid Cap	20
				Small Cap	25
				Equity MF	5
Very High	15	5	80	Large Cap	20
				Mid Cap	25
				Small Cap	30
				Equity MF	5

Equity Suggester

Using the predictions saved by the Equity forecasting model, this model suggests the equity for the investor based on their risk category, amount they wish to invest and the period for which they like to invest for.

The stepwise implementation of this model is mentioned below:

- Using the metadata file I looped through all the equities market cap wise and for each equity got the predicted close price at the end of the period the investor wants to invest.
- After rounding off the predicted price to 2 decimals, I got the close price at the end column of the historical dataset to get the return what the investor can expect.
- Then I got the return for all the equity by using the future close price and historical price and only calculated for those whose future close price was predicted in the positive integer.
- After getting the returns, I created a Data frame containing columns:

- Company Name – Name of the company
- Industry – the sector in which the company operates
- Symbol – NSE Symbol
- Market Cap – Market cap of the equity
- Ticker Symbol (BSE) – BSE ticker symbol
- Current Close Price (in ₹) – Close price at the end of the historical dataset
- Expected Close Price (in ₹) – Expected close price at the end of the requested period by the investor.
- Expected Return (in %) – Expected return based on the close prices above.
- Estimated Losses (in %) – Estimated losses, which were saved by the Estimated Losses module previously for the specific time_period
- After getting the values in a data frame, Based on the equity:debt:gold ratio, I split the columns by their market cap (large, mid and small) and sorted them in descending order based on their Expected returns and ascending order of Estimated Losses and kept a threshold based on the risk and Market cap of the equity.
- Next step I checked based on the sorted list based on the amount the investor wants to invest whether he/she will be able to minimum take one unit of that particular equity based on its current close price and also eliminated all the negative returns values and returned the three different data frames for different market cap of equities.

Based on the individuals risk category this model suggests which market cap equities the individual should target and what all equities in that market cap are expected to perform well for the period the user wants to invest their money for.

Mutual fund Suggester & Expected Downside Risk Calculator

Like Equity suggester module, this module suggests mutual funds based on the investors risk_category, amount they wish to invest and the period they want to invest for. The columns which the Mutual funds dataset has are:

- scheme_name – mutual fund scheme name
- min_sip – minimum amount required to start a SIP
- min_lumpsum - minimum lumpsum amount required to invest in the fund if the investor wants to not go with SIP
- expense_ratio - the management fees taken by the amc of the fund to manage the fund actively, the lower the better
- fund_size_cr – size of the fund in crore
- fund_age_yr – how long the fund have been active in years

- fund_manager - who is managing the fund
- sortino – measures the risk-adjusted return but only focusses on downside risk, the higher the better risk adjusted to returns
- alpha – measures the performance based on a benchmark interest, the higher the better, it means that fund has outperformed its benchmark
- sd – standard deviation represents the volatility of the fund, so for people able to endure more risk can be suggested with higher sd value
- beta – beta represents how much the fund moves with the market index like SENSEX, so a beta 1 represents if SENSEX goes up 1% the fund also moves 1%.
- beta-deviation – a new column which I created by subtracting 1 – beta in order to understand the fund movement.
- sharpe – similar to sortino ratio but this measure overall risk adjusted return rather than sortino which measures only the downside risk, higher the sharpe better the risk adjusted risk value.
- risk_level – represents the risk the fund possesses (1 – 6) (Low to very high risk)
- amc_name – Asset management company for that fund
- rating – fund rating value ranged from (0 – 5), higher the rating better the fund
- category – Equity/Debt/Hybrid fund
- sub_category – sub-category it falls into.
- returns_1yr – previous 1-year returns
- returns_3yr - previous 3-year returns
- returns_5yr - previous 5-year returns
- downside_risk – this is a new column which I created to store the Expected downside risk in percentage which a mutual fund has based on the different ratio it has. To calculate this, I replaced all the empty values from sortino ratio column with the mean and to calculate the downside risk the formula is:

Downside_risk = Previous_return – Average Government Bond Returns / sortino ratio

Here, previous return depends on the time which the user wants to invest and on an average I assumed Government bond rates (Risk free return) as 3 %.

For searching the top and most relevant mutual funds for the investor based on the above factors, the stepwise processing is mentioned below:

- I created a new column beta_deviation = 1 – beta to calculate the deviation for the fund to understand the movement of the fund based on the deviation.

- I converted all the risk_level from integer to systems risk categories i.e. ['Low Risk', 'Low to Medium Risk', 'Medium Risk', 'Medium to High Risk', 'High Risk', 'Very High Risk'] where it's equivalent risk_level is [1, 2, 3, 4, 5, 6] respectively.
- In the next step I sorted the fund based on the user's risk category and based on the time which the user wants to invest selected which column to target for the returns, (1yr, 3yr or 5yr).
- Based on the risk categories and time period I sorted the whole data frame based on specific columns ['expense_ratio', 'fund_size_cr', 'fund_age_yr', 'sortino', 'alpha', 'sd', 'beta_deviation', 'sharpe', 'rating', 'returns_1yr', 'downside_risk'] with each parameter ascending order as [True, False, False, False, False, True, True, False, False, False, True], for low risk categories and 12 months of yearly returns and similarly for all other factors based on the risk the individual can take.
- After sorting the data frame, I checked which funds the user will be able to invest based on the amount he/she wanted to invest and compared with the min_sip the fund has.
- The next step I split them into Debt, Equity and Hybrid funds based on the Equity Debt and Gold split ratio and renamed the data frame and returned only the columns which are essential for the investor, these include: ['Fund House', 'Fund Name', 'Expense Ratio(%)', 'Risk', 'Sub Category', 'Returns(1yr)', 'Returns(3yr)', 'Returns(5yr)', 'Fund Size(in Cr)', 'Expected Downside Risk(%)']

Using the above searching and sorting steps the expected mutuals funds are suggested to the investor specially tailored for his portfolio.

Recent Historical Gold Price Fluctuations

This module helps to get the current gold prices and fluctuations in the gold prices for the previous 3 months to give a clear picture about the investment to the investors. To get the current price and previous 3 months data I have used 3 different api calls from Quandl, Alphavantage and exchangerate api.

Quandl is used to get the current price of the gold being traded in Indian Rupee (INR) while Alphavantage api is used to get the historical gold prices in USD for previous 3 months from the current day and using the exchangerate api I get the current exchange rate for USD to INR and convert the same to get the fluctuations.

After converting the USD to INR, I subtract the current price received from the Quandl api with the latest price I the Alphavanatege api and using that subtraction factor I subtract all the values from the alphavantage api results with the subtraction factor in order to get the actual fluctuations in INR.

After getting the data frame I plot the data frame on a line chart using matplotlib's pyplot function and save the fig in a results directory for future use and display. This module function returns the current gold price and the gold historical data chart.

Portfolio Recommender

This is the main driver module; it helps in merging all the above modules and provide portfolio recommendation to the UI. The input from the UI is given to this module and based on the users input it communicates to all other modules mentioned above in order and provides all the necessary portfolio recommendations in different forms of data frames and the equity to debt to gold split ratio.

The order for calling the function is as follows:

1. Calculating the cost of living from user's input.
2. Putting user into assigned occupation and education categories based on user's occupation and education level.
3. Mapping the user input to categorical form and putting it into a numpy array and reshaping it for predicting the financial risk tolerance score.
4. Providing that array to the GBM model to predict the Financial Risk Tolerance score of the investor based on the np array of user's input.
5. Supplying the predicted financial risk tolerance score to get_risk_category function to get the Risk category of the individual.
6. Based on this risk category calculating the Equity to Debt to Gold split ratio of the investor for designing its portfolio based on its financial risk tolerance score.
7. From equity suggerer model get the suitable large, mid and small cap equities using the equity, debt, gold split ratio, time period for which the investor wishes to invest and the investing amount.
8. Get the debt, hybrid and equity mutual funds from the mutual fund suggerer using the same three factors from the above.
9. It calculates the current gold price and the gold chart from the current_gold_price function.
10. It converts the gold chart image to io.Bytes in order to print it on the html GUI.
11. Return all the data frames of the suggestions, equity debt gold ration and the gold price and charts to the GUI.

GUI for Investor interaction

For interacting with the investor and displaying portfolio analysis, I developed a simple GUI using python's Flask app and html css. I created two simple pages, one to get the users input and other to display portfolio results.

This module takes inputs from the investor and saves and passes them to the portfolio recommender module to get the portfolio analysis for the investor. The stepwise implementation is mentioned below:

- Using the HTML form and CSS I created an HTML form to take input from the investors on the various fields mentioned in the Individual Risk Tolerance calculator module, while some fields had options to select, and some were input fields.
- After getting the inputs once the user hits submit the flask app run and get all the values from the text boxes on html using get method and saves them into multiple variables.
- Later using these variables, I call the portfolio recommender model to calculate the portfolio for the investor and save those results into multiple data frames, to generate different graphs and provide them to the results GUI.

The results from the portfolio recommender are raw and in the form of python data frames so these data frames are later processed and converted into graphs and html tables to display in the Results section.

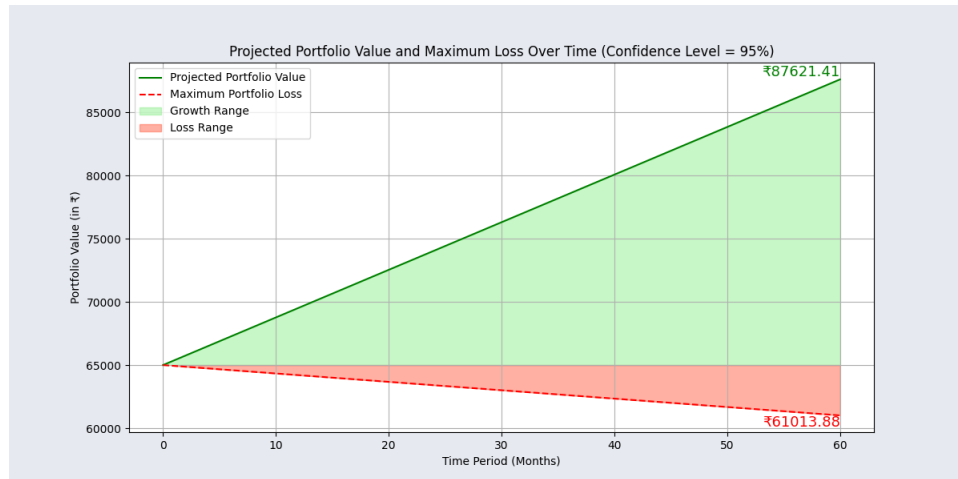
Projected Portfolio Value Calculation and Plotting

This module helps an investor to understand the associated risks with the suggested portfolio and returns he can expect in the form of graphs. The investor can see the risk to reward ratio of the portfolio and accordingly can plan to invest. This module actually calculates how much the value of the investor's portfolio would be after the certain time period and also how low could its portfolio go if things do not go as planned.

For calculating this I used the mean value of returns from the suggested portfolio for the risk as well as the return and calculated the projection of the portfolio. The stepwise implementation is mentioned below.

- For each suggested portfolio fields, if they are mutual funds then based on the time period, I calculated the mean for the Estimated Loss and Expected return and calculated the percentage based on the investing amount.
- Similarly, I calculated for all the equities suggested and summed up all the values.
- And after getting the portfolio projection I plotted them using line plot and filled with different colors to understand the risk and reward for the same. The graph below is an

example of an investor investing INR 65000, based on the returns for a period of 60 months it could go to INR 87,621 and the maximum portfolio loss is around INR 61013.88 with a confidence level of 95% which means that the probability of portfolio going lower than INR 61013.88 is 95%.

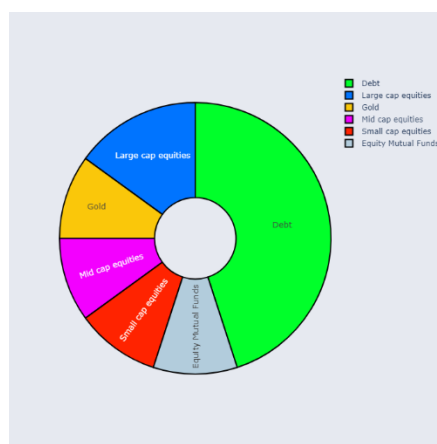


Plotting graphs for Recommended Portfolio

This module is a key module for displaying the results in a structured and understandable format for the investor to see how he/she can diversify their portfolio and what type of investments would be suitable for them. After getting the results from the portfolio recommender the data frames and outputs are provided to the generating graphs functions and the equity:debt:gold split ratio is been given to generate a over all pie chart on how the investor should diversify their portfolio based on their risk category.

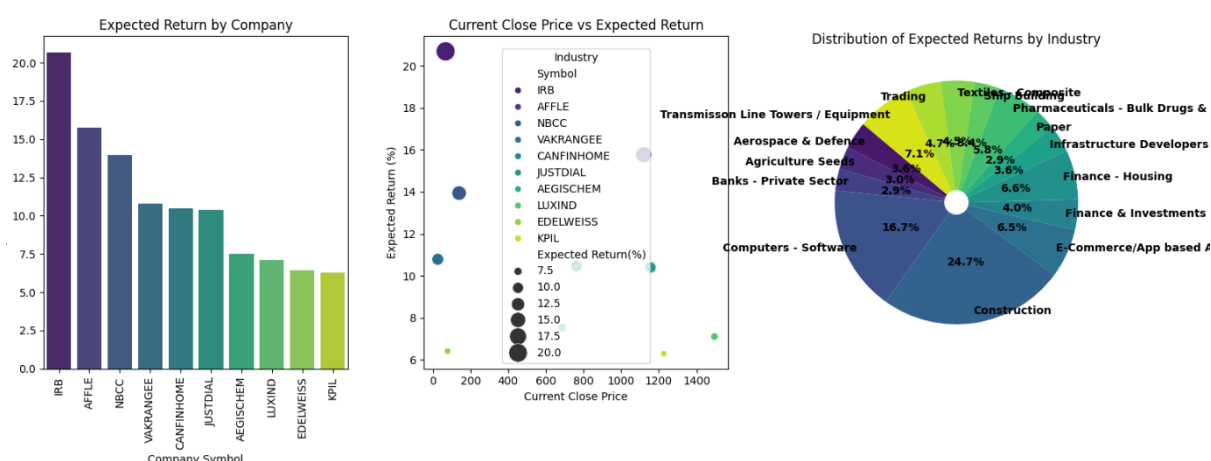
There are 4 different types of charts which I have used to represent all the outputs:

1. Pie Chart for Equity : Debt : Gold split ratio. Using the pie chart, I gave the investor an outlook of how his/her portfolio should look like based on its current financial position. Example of the pie chart is given below:

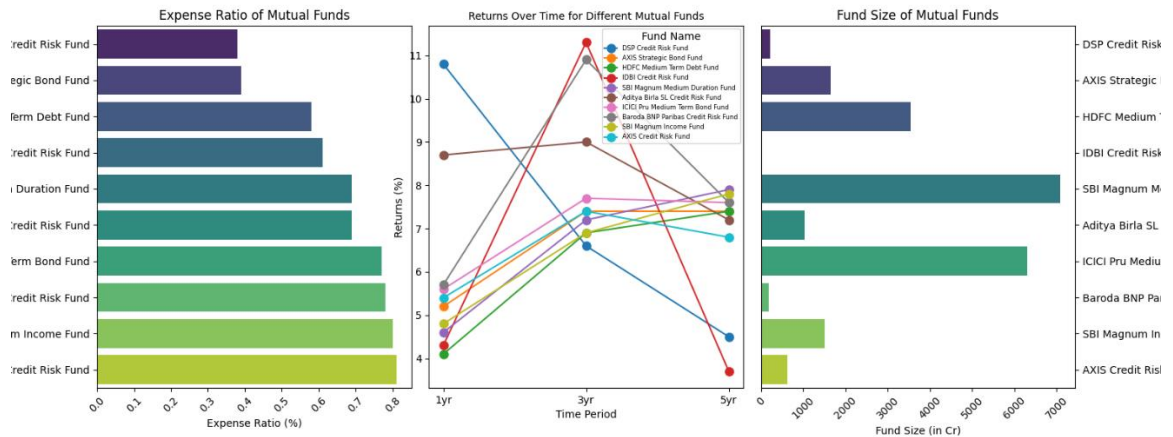


The above chart is an example for an individual who falls in Medium to High-risk category. From the pie chart we can see exactly how much percent of investing amount the investor should invest in Debt, Gold, Large Cap equities, Mid Cap equities, Small cap equities and Equity Mutual funds. The percentages can be seen on hover and if you don't want to invest in Equity mutual funds then you can just click on the Equity mutual funds legend on the right top corner and the recommendations will change accordingly. For creating the pie chart, I have used plotly.io and graph_objs packages.

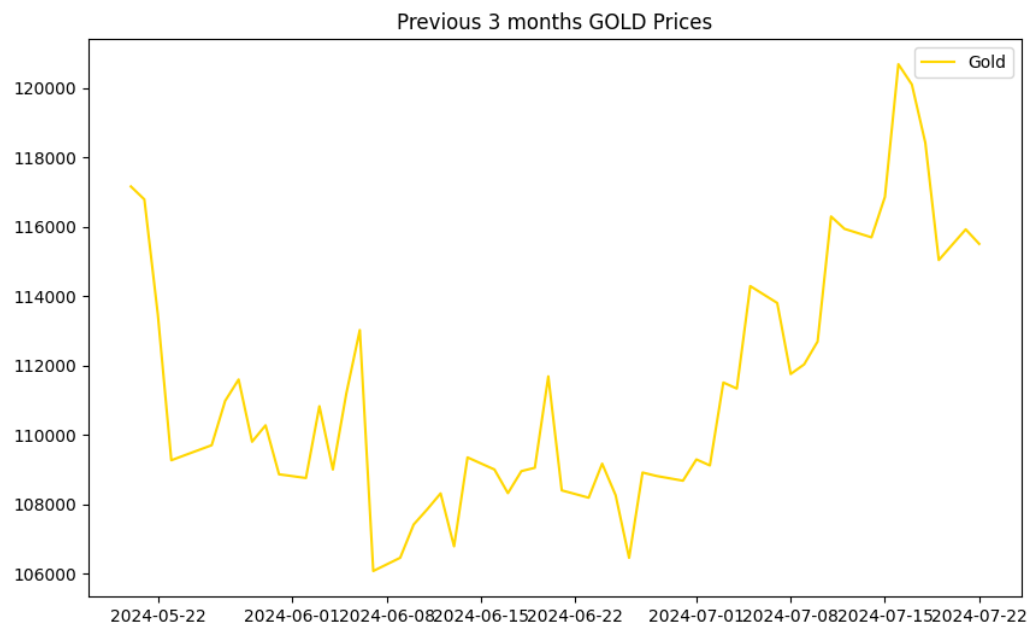
- Bar chart, Bubble chart and Pie chart to represent the equities. I used a sub plot of these three graphs to represents different types of insights from the equities recommended by the model. Bar chart would show the expected returns by company symbol, the bubble chart would show the current close price to expected return, in turn shows the risk to return reward and the pie chart would represent what Industry/sector would provide the maximum return. The chart below is an example of Small Cap equities recommended for Medium to High-risk category user, in similar way it will be generated for Large and Mid-cap equities as well.



- Bar chart and line chart to represent mutual funds. For representing the suggested mutual funds, I used a sub plot of bar and line chart, where the two bar charts were used one to represent the fund size and another to represent the funds expense ratio while the line chart was used to plot the returns a fund has provided in previous 1, 3 and 5 years. The chart below is an example of Debt Mutual funds suggested to an investor with Medium to High-risk category. In a similar way Hybrid and Equity mutual fund graphs are generated.



4. Line chart for Gold Price fluctuations. As previously discussed, the previous 3 months fluctuations was represented in a line chart to give the investor an idea of how the Gold price has been during previous days. The chart below is an example for the same:



GUI for Displaying Recommended Portfolio

The last module is used to display results on the web using flask and HTML CSS. After generating the graphs, the graphs and data frames are pre-processed in html format for displaying on the web. The web portal gives the investor a clear outlook of how its portfolio should be and gives options to alter. The steps to display results are given below:

- After generating all the graphs, we encode the graphs to Base 64 using 'b64encode' function from base64 for passing them as flask variable to display on the web.

- The next step is to convert all the data frames to html tables by using to_html function and providing the necessary html class for displaying.
- After which all these html tables and graphs are sent to the results page using flask's render template function and assigning appropriate variable names.
- For the display html page, I used bootstrap CSS for the tables and used various colour codes to represent various sections to make it easier and more understandable for the investor to relate.

Results

Investor Questionnaire page:

Inputs Given:

Gender: Male

Vehicle: No

Property: Yes

Marital Status: Married

House Type: Rented Apartment

Age: 38

Employed: Yes

Annual Income: ₹ 800,000

Occupation: IT Staff

Education: Academic degree

Family Member: 5

Amount to invest: ₹ 50000

Period to invest for: 3 months to 1 year

Questionnaire

What's your Gender?
Male

Do you own a Vehicle?
No

Do you own a property?
Yes

Your Marital Status
Married

What is your House Type?
Rented apartment

Your Age?
38

Are you Employed?
Yes

Provide your Annual Income(in ₹)

Questionnaire

127.0.0.1:5000

Your Age?
38

Are you Employed?
Yes

Provide your Annual Income(in ₹)
800000

Provide your occupation from the options below
IT staff

Provide your Education Qualification
Academic degree

How many Family Members(Dependants) do you have?
5

What amount are you planning to invest(in ₹)?
50000

What is the term you want to invest for?
3 months to 1 year

Submit

Results Page After submitting the form:

