IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 7 March 2023, accepted 20 March 2023, date of publication 29 March 2023, date of current version 5 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3263260

RESEARCH ARTICLE

Real-Time Portfolio Management System Utilizing Machine Learning Techniques

PRAKASH K. AITHAL[1], (Member, IEEE), GEETHA M.[1], (Member, IEEE), U. DINESH ACHARYA[1], BASRI SAVITHA[2], AND PARTHIV MENON[1]

[1]Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India
[2]Manipal Institute of Management, Manipal Academy of Higher Education, Manipal 576104, India

Corresponding author: U. Dinesh Acharya (dinesh.acharya@manipal.edu)

ABSTRACT There are 1641 companies listed on the National Stock Exchange of India. It is undoubtedly infeasible for a retail investor to invest in all the stocks. It is a well-known fact that the portfolio's return is an average return of all its constituent stocks, and risk will be less than or equal to the maximum risk of all the portfolio components. This paper is unique as it elaborates on the entire portfolio selection, optimization, and management process. Portfolio selection is accomplished through the K-Means algorithm. Optimization is achieved utilizing the genetic algorithm, and a sliding window is applied for portfolio management. Four different ways of portfolio calculation, namely, equally-weighted portfolio, global minimum variance portfolio, market cap-weighted portfolio, and maximum Sharpe ratio portfolio, are applied. The results depict that all three optimized portfolios outperform the Nifty index. The dataset for the study is obtained from globaldatafeeds.in.

INDEX TERMS Portfolio selection, portfolio optimization, portfolio management, real-time, K-means algorithm, metaheuristic algorithms, maximum sharpe ratio portfolio, global minimum variance portfolio, equally-weighted portfolio, sliding window.

I. INTRODUCTION

People invest in assets in the hope that they will bring future benefits. There are asset classes: treasury bills, bonds, stocks, gold, real estate, etc. People will invest in a combination of these asset classes. The present study considers stocks only.

A stock portfolio is a collection of stocks. The advantage of a portfolio is its return is the weighted return of its constituent stocks. The portfolio's risk will be less than the risks of its constituent stocks.

The portfolio selection is an NP-hard problem. Once it is decided how much money needs to be invested in stocks, the portfolio selection process begins.

Portfolio selection has two steps: The choice of stocks is made in the first step. The second step decides how much money goes into selected stocks. Portfolio optimization is carried over to adjust the money that is allocated to each chosen stock. Portfolio management is the process of continuously readjusting the funds allocated to the stocks to gain maximum profit at minimal risk [1].

The portfolio is managed in two different ways, namely, passive portfolio management and active portfolio management. In passive portfolio management, fund managers either invest their client's money in index funds or selected stocks. The built passive portfolio will not alter till its maturity. Passive portfolio managers assume that the markets are efficient. Contrary to passive portfolio management, active portfolio management involves investing in stocks and readjusting the portfolio whenever there is a change in market conditions. Active portfolio managers assume that there is a chance of making a profit whenever markets deviate from efficient form. Active portfolio managers do fundamental analyses, event studies, and technical analyses of the stock market [2].

There are numerous stocks traded every day. Every investor wishes to maximize their profit and minimize their risk. In Pre 1950s, people used to invest in individual stocks and were exposed to the risk of the instruments they invested in. Markowitz proved that risk could be reduced by investing

The associate editor coordinating the review of this manuscript and approving it for publication was Vlad Diaconita.

in a set of instruments known as a portfolio. The portfolio gives the average return of constituent instruments with reduced risk. Portfolio selection, optimization, and management became pioneer research fields. The computer-aided finance made the field even more interesting as it increased the stakeholder base, reduced the transaction cost, and so on [3].

The portfolio selection, optimization, and management are helpful for institutional investors as they can build and manage their portfolios. The proposed manuscript will help retail investors to build and manage their portfolios. The proposed work can be extended to any nation's stock market, and investors can make informed decisions on investing in the stock market. Fund managers can offer ETFs based on the proposed manuscript.

II. LITERATURE REVIEW

The modern portfolio theory put forth by Harry Markowitz in the early 1950s gave insight that instead of investing in a single stock, it is profitable to invest in a set of stocks known as a portfolio. Experts from several domains have investigated the construction and management of portfolios. The decision to buy, sell or hold the stocks in the portfolio is a tedious task. To select the instruments for the portfolio and to overcome the sector bias, the hierarchical clustering algorithm is used. The clustering is required to get the maximum returns from a portfolio. The clustering-based algorithm's volatilities are much smaller than the Markowitz Mean-Variance model [4], [5], [6]. The primary goal of any investor is to make a profit. Diversification and proper allocation of wealth are necessary to achieve the goal. Fund managers do it manually for their clients. A computer-based portfolio selection was presented [7]. It is impossible to manually analyze the stock movements and decide about buying, selling, or holding the stock. Portfolio optimization is usually accomplished through a genetic algorithm [8].There are three operators in a genetic algorithm: crossover, mutation, and inversion. A grouping genetic algorithm is applied by Chen et al. [9] for finding a group trading strategy portfolio. A group stock portfolio categorizes the stocks into different groups: banking stocks, IT stocks, Health stocks, and so on. A grouping genetic algorithm is applied to optimize the portfolio. If any stock within the sector is not performing, it is replaced with stock from the same sector [10]. Chen et al. have proposed an improvement to their earlier article by implementing the fuzzy grouping genetic algorithm [11]. A genetic algorithm is a nature-inspired algorithm. The genetic algorithm iterates over the population and finds the optimal solution. It applies crossover, mutation, and selection operations to the population to find the optimal solution [12]. The genetic algorithm and Particle Swarm Optimization algorithm are two renowned nature-inspired meta-heuristic algorithms. They reduce computational time and can manage realistic constraints efficiently [13]. The trading strategy involves stop-loss, take-profit points, and long-term portfolio management. The stop-loss and take-profit are required to tackle extreme profit or losses. The genetic algorithm is capable of identifying stop-loss and take-profit points [14]. A genetic algorithm was utilized to track the investment. The portfolio will move in which direction was studied [15]. A Jordan Elman network with a genetic algorithm was used for portfolio selection. The method uncovers the hidden patterns in the stock market [16] Investor sentiment is fed to the grouping genetic algorithm to optimize the portfolio. The Chen et al. have arrived at following rules [17]

1) Rule 1-On a day t, if a stock's Buy Sell Imbalance is Low and DayTrading is rising, then buy the stock.
2) Rule 2- On a day t, if a stock's Buy Sell Imbalance is High and DayTrading is rising, then sell the stock.

Mohammad Maholi Solin et al. have proposed an artificial neural network model to predict the stock prices and a genetic algorithm to optimize the portfolio. Their study reveals that the genetic algorithm outperforms the single-index model. The study involved 38 stocks from the Indonesian stock market: the data considered open, close, high, low, and volume [18]. An electricity market portfolio is established through the metaheuristic genetic algorithm. Metaheuristic genetic algorithm achieves similar results compared to the exact algorithm but with higher efficiency [19]. Gupta et al. categorize the stocks into three groups using a support vector machine. From each class, stocks are picked to form the portfolio utilizing the genetic algorithm [20]. For uncertain environments, one can not rely upon mean and variance alone. Skewness and Kurtosis need to be considered as well. In an uncertain environment, it is not sufficient to depend upon historical data; expert opinion must be taken into account [21]. Stock performance depends on economic features such as production power, service sector, industrialization, farming, natural resources, etc. Finance must reflect the current economy [22]. A genetic algorithm-based portfolio optimization technique was utilized to optimize the SPX500 stock market, and it outperforms the market by a significant margin [23].

The literature review shows that the metaheuristic genetic algorithm provides highly accurate results with greater efficiency. The genetic algorithm can be applied for portfolio selection and optimization.

To avoid overfitting by genetic algorithm sliding window is applied. The training window's length must be proportionate to the size of the testing window, say 12:1, 6:1, or 3:1 [24]

A multi-objective portfolio selection model using maximum entropy with a genetic algorithm was presented. The model achieved higher returns with more diversification due to the introduction of maximum entropy [25]. The benefit of including the real estate investment trust in the portfolio was studied using a genetic algorithm. The genetic algorithm outperformed the global minimum variance portfolio [26]. Diversification-aware portfolios were optimized using a genetic algorithm-Out of the three portfolios constructed, One of the portfolios underperformed. One performed marginally better, and one outperformed the benchmark [27]. A survey of different portfolio optimization methods was presented [28]. The stock price prediction application utilizing metaheuristics algorithms is an area yet to be explored [29]. The support of parallelism in genetic algorithms, simplicity, and efficiency makes it more suitable for portfolio optimization problems [30].

Behavioral finance is an interdisciplinary field involving psychology, event studies, economics, and sociology. Humans make their investment decisions based on emotions [31].

The reviewed literature does not cover all the phases of portfolio management. Only one metaheuristic algorithm was considered to optimize the portfolio in the reviewed literature. The validation of obtained results was missing in the literature.

The proposed work covers the entire portfolio management. Seven metaheuristic algorithms are compared. The constructed portfolio is validated. The manuscript presents all the phases of the portfolio, namely, portfolio selection, portfolio optimization, and portfolio management. The holistic approach is not present in the reviewed literature.

III. METHODOLOGY

The methodology is presented in subsections III-A and III-B with titles Models and Process, respectively.

A. MODELS

The following are the key concepts and models used in this article.

1) Equally-Weighted (EW) Portfolio
2) Market Cap Weighted (MCW) Portfolio
3) Global Minimum Variance (GMV) Portfolio
4) Maximum Sharpe Ratio (MSR) Portfolio
5) K-Means clustering Algorithm
6) Genetic Algorithm (GA)
7) Ant Colony Optimization (ACO) Algorithm
8) Particle Swarm Optimization (PSO) Algorithm
9) Firefly Algorithm
10) Artificial Bee Colony (ABC) Optimization Algorithm
11) Tabu Search
12) Simulated Annealing (SA)
13) Sliding Window.

1) EQUALLY-WEIGHTED PORTFOLIO

Equal weight is a proportional measuring method that assigns the same importance to every stock in a portfolio, index, or index fund. In such a system, stocks of smaller and lesser-known companies are given the same importance as stocks of larger and more important ones. An index with equally weighted assets must continuously buy and sell as shares of companies increase and decrease in price. Shares are bought when the prices of a security start to decline, and balance is restored where as shares are sold when the prices increases. Therefore, an equally weighted portfolio of assets implies that the capital invested is distributed equally among all the assets. This results in the creation of an equal weight index for each of the assets. Equal weight indices are also referred to as unweighted indices.

The performance of each asset in an equally weighted portfolio is responsible for the performance of the entire portfolio. The comparatively greater performance of equally-weighted portfolios can be attributed in part to its strategy of value investing. Value investing involves picking those stocks for the portfolio that appear to be undervalued and have the potential to grow in the future. Such assets, more often than not, give higher returns over a longer period.

Another defining feature of equally-weighted portfolios is their diversification. Diversification is a risk management strategy that combines a variety of assets in the portfolio. It works on the principle that having a mixture of different types of assets will lower the risk posed to one particular asset and yield higher returns over time.

Equally weighted portfolios also feature a higher portfolio turnover rate. Turnover rate is defined as the ratio in which assets in a fund are bought and sold by portfolio managers, which ultimately gives the percentage change of assets over a period and the consequent higher transaction costs and taxes. Additionally, when the market experiences prolonged declines during the bear market phase, equally-weighted portfolios are more vulnerable to sudden changes in value [32].

2) MARKET CAP WEIGHTED PORTFOLIO

A market capitalization-weighted portfolio, often shortened to a market-cap-weighted portfolio, consists of a portfolio where individual assets are included in values that correspond to their total market cap. The individual components of the portfolio are calculated by multiplying the outstanding shares (owned by individual shareholders, institutional block holdings, and company insider holdings) of a company by the current price of a single share, which corresponds to the current market price of the outstanding share.

The total value of such a portfolio keeps varying based on the price of the share in the market as well as the weight allocated to a particular asset in the portfolio. Market capitalization is a representation of the total market value of a firm's outstanding shares and the contribution of its shareholders.

In the market-cap-weighted portfolio, assets with a higher market cap will receive a higher percentage of the weight from the portfolio. Companies that fare comparatively poor in the market receive a lesser percentage of the weight from the total portfolio. This ensures that poorly performing stocks do not affect the overall performance of the overall index.

A big advantage of this type of index is that it gives access to a wide spectrum of companies. Even though this makes it vulnerable to frequent stock movements, the presence of larger companies might mean that the share values of such companies can give constant growth in value to the portfolio. Additionally, some companies have shares that are not completely public, forcing indices to use the technique of free float factor (the percentage of shares available for trading) for calculating prices and fluctuations.

In contrast, large companies having higher weights can have a disproportionate impact on the portfolio as a whole. Market-cap-weighted portfolio, hence, gives an almost unbiased automatic process for promoting stocks with higher values and negating ones with declining prices [32].

3) GLOBAL MINIMUM VARIANCE PORTFOLIO

Even with the existence of a variety of optimized portfolios, the returns that can be expected from a particular portfolio can be difficult to determine in most cases. Mathematically, the global minimum portfolio is the most efficient portfolio since it minimizes risks and maximizes returns from the investment. This occurs because the weights of this portfolio depend only on the return variance values and the covariances and not on the returns from the portfolio.

On a broader note, the global minimum variance portfolio is a subset of the efficient frontier representation. The efficient frontier is a set of optimal portfolios that give the best returns with the least risk, for any given level of return.

Portfolios lying below the efficient frontier are sub-optimal portfolios as they do not give enough returns for that level of risk, whereas those to the right are sub-optimal because they have a higher level of risk involved. Successfully optimizing a portfolio would imply that the portfolio will lie on the efficient frontier line.

The global minimum variance portfolio, hence, would lie to the leftmost end of the efficient frontier. It represents a portfolio having the least risk, by virtue of least standard deviation, but still provides optimal returns to the investor. Stocks in any minimum variance portfolio might be risky when held individually but balance each other out when held together under the same index.

Hence, diversification of investment is an apt concept to be applied to minimum variance portfolios in general. Investing in sectors with little correlation can lead to a less volatile portfolio, thus giving larger returns over a more extended period [32].

4) MAXIMUM SHARPE RATIO PORTFOLIO

The Sharpe ratio is used to help investors evaluate the return on investment compared to its risk. The ratio defines the average return in addition to the free risk rate upon the total risk involved.

The Sharpe ratio is calculated by subtracting the risk-free rate from the return of the portfolio and then dividing the result by the standard deviation of the portfolio. The standard deviation is a measure of the volatility of the portfolio. The Sharpe ratio hence is an efficient methodology to calculate risk-adjusted returns on an index.

The formula to calculate the Sharpe ratio is given in equation 1.

$$(R(p) - R(f))/s(p) \tag{1}$$

where R(p) indicates the return of the portfolio, R(f) indicates the risk-free rate of return, and s(p) denotes the standard deviation of the portfolio

Having a higher Sharpe ratio, generally above 1.0, indicates a better risk-adjusted performance. If portfolio analysis results in a negative Sharpe ratio, then it implies either that the risk-free rate is greater than the return of the portfolio or that the return of the portfolio is actually negative in value.

On similar lines, a well diversified portfolio would have a higher sharpe ratio compared to other portfolios having a lower level of diversification. Sharpe ratio can hence be used as a tool to identify and compare values of funds in two different categories. Shares that give similar returns but have different risk factors involved can be identified using this method.

One of the limitations of using the Sharpe ratio as a measure for analyzing risk-adjusted performance is that it relies on standard deviation as a critical parameter of calculation. It just assumes that returns are normally distributed. This may lead to erroneous results, faulty investments, and lesser returns over a longer time period [32].

5) K-MEANS CLUSTERING

Unsupervised learning algorithms analyze and create clusters in unlabeled datasets. These algorithms are widely used for exploratory data analysis and image recognition because of their ability to detect and group patterns on their own.

Clustering is an unsupervised data mining technique that groups data points into separate clusters, where the data points in a particular group have similar properties. More specifically, exclusive clustering, also known as hard clustering, implies that one particular data point can be part of only one particular cluster. The K Means clustering algorithm is one such simple and well-known exclusive clustering algorithm. Additionally, K Means clustering is also an example of partition clustering as the entire data is divided into non-hierarchical partitions by virtue of centroids.

In the K Means algorithm, all the data points are organized into K groups, where K represents the number of centroids present in the dataset. A centroid is an imaginary location depicting the center of the entire cluster. Every data point is then assigned to the nearest groups while at the same time keeping the centroids as small as possible.

The means in K Means refers to the averaging of data to find the centroid for every cluster. A larger K value in the algorithm indicates smaller groupings and more granularity, while a smaller K value indicates larger groups with less granularity.

The algorithm starts its execution with a randomly selected set of centroids. These centroids are used as starting points for the respective clusters. Further iterative calculations are performed on the dataset to optimize the positions of these centroids such that either of the following conditions is met:

- The positions of the centroids stop changing because they have been optimized
- The predetermined number of iterations has been completed. This leads to a successful clustering operation.

Even with all the advantages of K Means clustering, there are a few drawbacks to it. They are not suitable for noisy data and cannot identify non-convex clusters. The number of clusters to be formed also needs to be given in advance to the algorithm. Additionally, since it is an unsupervised learning algorithm, it takes a longer training time, has a higher risk factor of producing inaccurate results, and is ambiguous in the way it classifies clusters [33].

6) GENETIC PROGRAMMING

Genetic Programming is a type of evolutionary algorithm which is a subset of Machine Learning. They are a part of heuristic search algorithms that are based on the principle of natural selection and evolution. They are utilized to solve problems that humans cannot solve directly. Since they are free from human intervention, these algorithms and programming techniques are used in optimization problems and search algorithms.

A genetic program consists of generations that, in turn, consist of a population of individuals. Each individual entity represents a point in the search space. Each individual is represented by a string of characters, bits, or floats.

Cleverly exploiting random search, the genetic programming methodology generates a set of solutions which are in turn represented by sets of fixed-length strings. Each of these strings is referred to as a chromosome. From a set of best solutions, newer and improved solutions are created by reiterating the entire process. This process is repeated until either the best possible solution is obtained or the required number of iterations is completed.

Even though the algorithm appears to be random, there are three major operations that go into the genetic programming process.

Selection involves taking chromosomes with good fitness scores from the parents and passing them on to successive generations for fitter and better offspring. It is based on the idea of survival of the fittest.

The crossover operation involves actual mating between individuals. Crossover sites are chosen randomly, and better strings are created by exchanging genes at these locations.

The mutation is an operation that keeps running in the background. It facilitates a change in genes or chromosomes to allow for diversity in the population. It allows the algorithm to be on the lookout for newer possibilities for the optimal solution.

Genetic algorithms are more robust than traditional artificial intelligence algorithms as they do not converge or break under the presence of noise in the input stream. They also have a higher ability to optimize search problems. Genetic programming can be applied in the classification of data, automated bug fixing, ensuring network security by rapid identification and resolving of breaches, and as a supporting mechanism for various other machine learning methods like neural networks and recurrent neural networks [34].

7) ANT COLONY OPTIMIZATION ALGORITHM

Ant Colony Optimization (ACO) is classified under ant colony algorithms and applies specific metaheuristic optimizations as well. A probabilistic technique that reduces computational problems to finding the shortest path between graphs, it takes inspiration from the collective foraging behavior of ant colonies. It was first introduced by Marco Dorigo in 1992 [35] as a means to search for an optimal path in a graph. Eventually, it began to be applied to more numerical applications.

The crux of the algorithm is to observe and replicate the movement of ants from their nests in search of food. Ants tend to take the shortest path available to the food resource. Thus, they end up walking over the shorter path more often than on other paths. The presence and persistence of pheromones, chemicals that ants use to communicate, on these paths help determine the length of the path followed. The more ants march over a particular trail, the higher the pheromone density and the better the optimization. This mimicry of simulated ants finding the shortest path and leaving behind artificial pheromones by walking around a graph characterizes the ACO algorithm.

8) PARTICLE SWARM OPTIMIZATION ALGORITHM

Particle Swarm Optimization (PSO) is a population-based stochastic technique that optimizes a given problem by trying to improve a solution concerning the quality of the solution. PSO, first proposed by Kennedy and Eberhart [36], was initially aimed at simulating the social behavior of groups of organisms like flocks of birds or a school of fish.

Each particle is considered part of a research space containing a global minimum. None of the particles are aware of the exact location of this minimum but have a general idea of its location. Each particle is attributed with a certain velocity vector that is randomly chosen. The movement of the particles is affected by cognitive influence (towards its previous personal best position) and social influence (towards the best solution of the group).

The coefficients of inertia, cognitive and social influence control the levels of exploration and exploitation, which are important in deciding the solution to the problem. Exploration implies the ability of particles to move towards the best solutions found till a particular time. Exploitation, on the other hand, refers to the ability of particles to process the entire research space. Optimization using PSO often involves finding the right balance between exploration and exploitation.

9) FIREFLY ALGORITHM

The firefly algorithm is a metaheuristic inspired by the behavior of fireflies and their bioluminescence. It is believed that lesser brighter fireflies will be attracted to brighter ones. If there are no fireflies nearby, then a particular firefly moves randomly.

From the initial population of fireflies that exist, light intensity is formulated that is related to the function in question. An absorption coefficient controls the brightness of the light. The attractiveness of the fireflies varies with distance. As the fireflies move towards each other, newer solutions are computed, and the light intensity varies accordingly. In the end, the fireflies are ranked to find the firefly having the best value for the objective function.

This effectively subdivides the whole population into subgroups as each subgroup swarms around a local optima [29].

10) BEE COLONY OPTIMIZATION ALGORITHM

Bee Colony Optimization (BCO), a population-based algorithm, is inspired by the activity and movements of bees in nature. It belongs to the category of swarm intelligent algorithms, a subset of AI algorithms based on the study of actions of individuals in various decentralized systems. Artificial bees are programmed such that when acting together, they can solve complex combinatorial problems.

The algorithm has two alternating stages: the forward pass and the backward pass. In every forward pass, the bees explore newer search spaces (similar to how they search for nectar in newer locations). Here, a predefined number of moves is applied that helps to generate a new partial solution to the problem. Every backward pass consists of bees sharing all the information that was obtained during the forward pass (similar to how bees share information about nectar locations in the beehive).

These artificial bees have objective function values that characterize the quality of the solution. Based on a random probability, the bee decides whether to abandon the created partial solution or to recruit more bees to return to the created partial solution. It is observed that bees having a better objective function value have a higher probability of continuing its exploration. Every artificial bee generates a solution to the problem. The best possible solution is selected at the end [34].

11) TABU SEARCH

Tabu Search employs local search methods to optimize mathematical problems. Similar to a few other metaheuristics, this technique also considers the neighbors of a potential solution in the hope of discovering a better solution.

Tabu Search is different from other local search techniques in that it allows for worse states of neighbors to be accepted. This prevents the algorithm from being stuck at a local minima and helps it explore regions of the search space not visited earlier. Additionally, restrictions (or tabu) are applied on regions of the search space visited earlier, thus preventing the algorithm from considering the same solution repetitively.

There are different kinds of memory structures involved in tabu search.

- Short-term memory
  - It is based on how recent occurrences are handled and prevents the algorithm from going to recently visited search spaces
  - It is handled by a Tabu list
- Long-term memory
  - It is based on the frequency of occurrence and helps the algorithm explore new spaces
  - It is handled by the frequency memory

The algorithm is repeatedly invoked on a search space until a particular number of iterations has been achieved or the best solution is found [37].

12) SIMULATED ANNEALING

Simulated Annealing is a probabilistic optimization algorithm used to tune model parameters so that they do not get stuck at the local minimum while searching for an optimal global minimum value of a function. It draws inspiration from the physical annealing process, where a material is heated up until it reaches an annealing temperature and is cooled down slowly so that the required structure is obtained from the material.

At each interaction, the algorithm probabilistically determines whether or not to move the system to one of its neighboring states. The probability of moving from the current state to a new neighboring state is defined by the acceptance probability function and depends on the energies of the two states as well as temperature, which varies with time. The evaluation and acceptance of the neighbors from every move may result in the system temporarily accepting and moving to a worse state before moving on to find the global minimum [34].

13) SLIDING WINDOW

The sliding window algorithm is used to solve problems involving linear sequences. A window is formed that slides over the given data, thus covering different parts of the data at different points in time. The main aim behind this algorithm deals with converting two nested loops into one single loop. This reduces time complexity to linear time.

The sliding window technique has two basic requirements.

- There must be two pointers, say, L and R, that correspond to the left and right ends of the current working range
- The algorithm must be able to add new elements to the range when the right-most pointer is moved forward and must be able to delete or remove elements from the range when the leftmost pointer is moved forward.

Based on the size of the ranges, there are two types of sliding window techniques in practice. If the length of the sliding window remains fixed, it is known as the fixed-size sliding window technique. On the other hand, if the length of the sliding window is modifiable, it is called the flexible window size sliding technique.

This technique of applying a sliding window is beneficial in the fact that it prevents unnecessary iterations over a collection of data. It also gives a convenient way in order to analyze data in distinct sub-sections.

A major disadvantage of the sliding window technique of traversal of arrays or strings is the computational cost involved. Increasing the window size makes it faster but at the same time affects the accuracy of operation.

B. PROCESS

The actual stock data from globaldatafeeds.in is considered for the study. The data comprised of stock ticks data for a period of one month. The data consists of NSE NIFTY 100 stock prices. The comprehensive list of NIFTY-100 companies is depicted in table 1. The query for storing the historical data fields is depicted in figure 2. The instidentifier field identifies the instrument and open, high, low, and close fields depicts the open price, high price, low price and close price respectively. Traded quantity field stores the volume of trade for the particular instrument. The query for storing the real-time data fields into the timescale db is depicted in figure 3. This quantitative data, along with the qualitative data, is utilized to build the portfolio. The four different portfolios, namely, equally weighted, market-cap-weighted, global minimum variance, and maximum Sharpe ratio, are managed with the help of K-Means clustering and metaheuristic algorithms. The study considered forty-four financial indicators depicted in table 2. The seven financial indicators are chosen using principal component analysis. The seven significant financial indicators, namely, the corporate tax rate, GDP, current account to GDP, import, industrial production, prime lending rate, and tourist arrivals [38] are given as input to the system. The calendar of events data [39], tone of the annual report, stock prices, and volume are fed to the proposed system as depicted in the Fig. 1.

A stock portfolio is a collection of stocks. Portfolio selection and optimization for an active portfolio management technique are proposed. The proposed portfolio construction and optimization are based on quantitative and qualitative data. The quantitative data comprise macroeconomic indicators, stock prices, and stock volume. The qualitative data contains Tweets, news articles, the director's report, and auditor reports. The elbow method is applied for the ten-day sliding window to identify the best K of the K-means algorithm. The portfolio which maximizes the return is constructed. The following optimization methods are explored

1) Genetic Algorithm
2) Ant Colony Optimization Algorithm
3) Particle Swarm Optimization Algorithm
4) Firefly Algorithm
5) Artificial Bee Colony Optimization Algorithm
6) Tabu Search
7) Simulated Annealing.

The genetic algorithm with the minimum variance weights, and Sharpe weights objective functions are considered. For each of the objective functions, the portfolio returns are computed. The training period of six months and one year is utilized, and a testing period of one month is considered. K-means clustering partitions the N observations into K clusters. Each observation belongs to the cluster with the nearest mean. A genetic algorithm is a search-based optimization algorithm. The following parameters are utilized in GA as they gave the best results.

1) maximum number of iteration = 1000
2) population size = number of stocks
3) mutation probability = 0.5
4) elit ratio = 0.01
5) crossover probability = 0.06
6) parents portion = 0.3
7) crossover type = uniform
8) maximum iteration without improvement = 0.01.

Mutation probability is the allowed probability of change in population. Crossover probability is the probability that interpopulation change occurs. Elit ratio specifies how many entries need to be unaltered for the next generation. The uniform crossover ensures that the genes are chosen randomly, and there is no positional bias. The algorithm for portfolio optimization is given below.

The following parameters are utilized for the PSO.

1) nostalgia = 0.8
2) envy = 1.3
3) inertia weight = 0.6.

Since PSO was designed to find the global optima, the envy value considered is greater than the nostalgia value. The inertia weight decides the speed of the particle. To utilize the ACO, SA, and Tabu search in the model, The model was redesigned as a network. The historical portfolio returns and current day portfolio return are utilized to predict the future portfolio returns. In case of ACO, Euclidian distance is applied to obtain the optimal solution efficiently. SA and Tabu search required the input to be in the form of linear and binary quadratic model. ABC model is trained with the following parameters to gain the optimal results.

1) colony_size = 40
2) scouts = 0.5
3) iterations = 50
4) min_max = 'min'
5) nan_protection = True
6) log_agents = True.

As sliding window slides, the previous optimal portfolio is compared with the current optimal portfolio, and the best among the two is retained.

IV. RESULTS

The K-means clustering is utilized to group the stocks. The K-means algorithm with different numbers of stocks and different values of K is considered, and an optimal K-value is found utilizing the elbow method. The elbow plot for K = 5 and a total of 5 stocks is depicted in figure 4. The elbow plot for K = 100 and a total of 100 stocks is shown in figure 5. It is evident from the graph that as data size increases, the value of best-K also increases. The K-means clustering with optimal K = 5 for datasize of 10 stocks is given in the figure 6. In the figure X and Y axis are PCA axis.

To find the MSR a negtive sharp ratio is minimized. Weights of GMV portfolio is given in table 3 and weights of

TABLE 1. NIFTY100 instruments.

ABBOTINDIA	ACC	ADANIENT	ADANIPORTS	ALKEM
AMBUJACEM	APOLLOHOSP	ASIANPAINT	AUROPHARMA	AXISBANK
BAJAJ-AUTO	BAJAJFINSV	BAJAJHLDNG	BAJFINANCE	BANDHANBNK
BERGEPAINT	BHARTIARTL	BIOCON	BOSCHLTD	BPCL
BRITANNIA	CADILAHC	CIPLA	COALINDIA	COLPAL
DABUR	DIVISLAB	DLF	DMART	DRREDDY
EICHERMOT	GAIL	GLAND	GODREJCP	GRASIM
HAVELLS	HCLTECH	HDFC	HDFCAMC	HDFCBANK
HDFCLIFE	HEROMOTOCO	HINDALCO	HINDPETRO	HINDUNILVR
ICICIBANK	ICICIGI	ICICIPRULI	IGL	INDIGO
INDUSINDBK	INDUSTOWER	INFY	IOC	ITC
JSWSTEEL	JUBLFOOD	KOTAKBANK	LT	LTI
LUPIN	M&M	MARICO	MARUTI	MCDOWELL-N
MRF	MUTHOOTFIN	NAUKRI	NESTLEIND	NMDC
NTPC	ONGC	PEL	PETRONET	PGHH
PIDILITIND	PNB	POWERGRID	RELIANCE	SAIL
SBICARD	SBILIFE	SBIN	SHREECEM	SIEMENS
SUNPHARMA	TATACONSUM	TATAMOTORS	TATASTEEL	TCS
TECHM	TITAN	TORNTPHARM	UBL	ULTRACEMCO
UPL	VEDL	WIPRO	YESBANK	ZYDUSLIFE

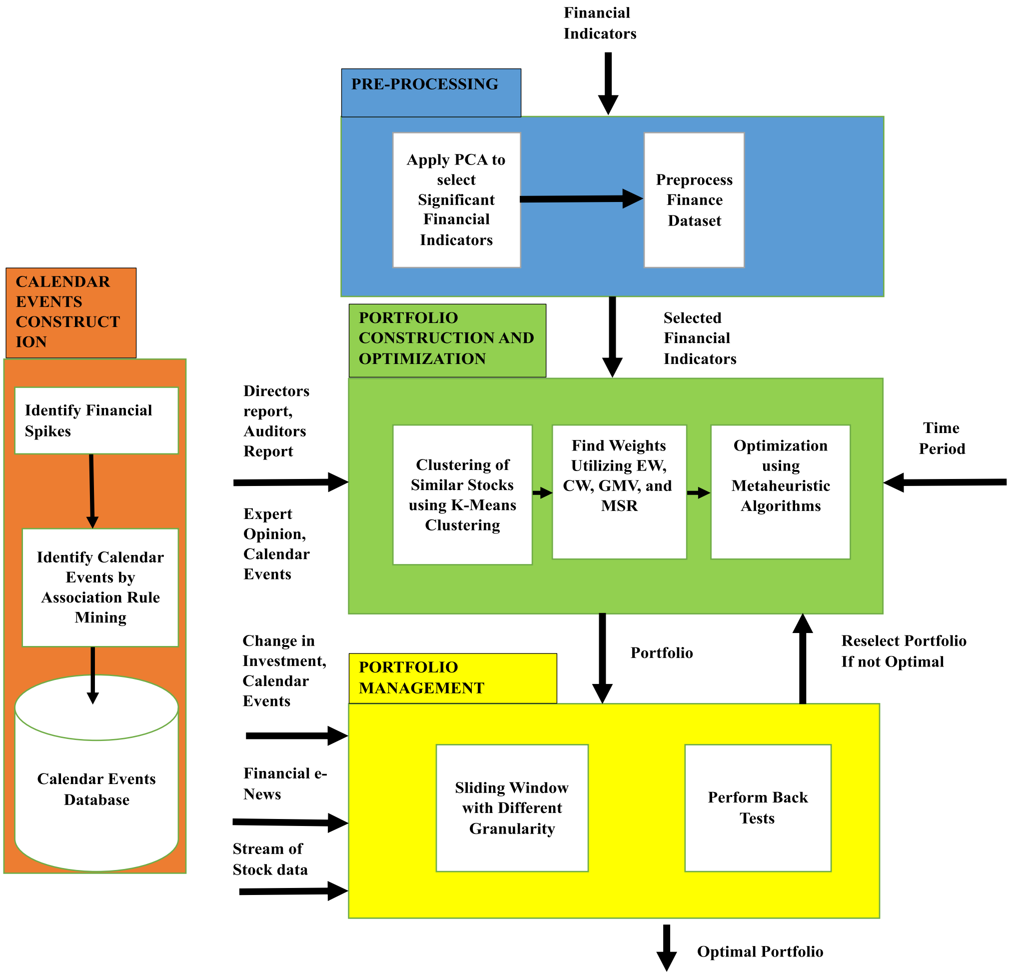

FIGURE 1. Overall architecture of proposed methodology.

MSR portfolio are given in table 4. The objective function of the GMV and MSR portfolios are depicted in figures 7 and 8 respectively.

The evolution of particles in a PSO algorithm is depicted in Figure 9. From Figure 9, it is clear that the global optima are attained as all the particles converge to value one. The

TABLE 2. List of macroeconomic indicators.

The Macroeconomic Indicators	
BALANCE-OF-TRADE	CAR SALES
CRUDE OIL PRODUCTION	EXPORT
INFLATION RATE	FDI
MANUFACTURING PMI	REMITTANCES
TOTAL EXTERNAL DEBT	WHOLE SALE INDEX SALE
MONEY SUPPLY M1	MONEY SUPPLY M2
CENTRAL-BANK-BALANCE-SHEET	CENTRAL-GOVT-BUDGET
CONSUMER PRICE INDEX	GDP
CHANGE-IN-INVENTORIES	FOREX
MONEY SUPPLY M3	PRODUCER PRICES
TREASURY BILL 91 DAYS YIELD	GOVT SPENDING
EXPORT PRICES	GOLD PRICES
CURRENT-ACCOUNT-TO-GDP	GOVT BUDGET VALUE
PRIME LENDING RATE	VAT
GOVT BOND 10 YEARS YIELD	SERVICE PMI
INFRASTRUCTURE OUTPUT	IMPORT
GROSS FIXED CAPITAL FORMATION	TOURIST ARRIVALS
INDUSTRIAL PRODUCTION	FOOD INFLATION
MANUFACTURING PRODUCTION	CORPORATE TAX RATE
TOTAL DISPOSABLE INCOME	UNEMPLOYMENT RATE
COMPETITIVE INDEX	CURRENT ACCOUNT
GOLD RESERVE	INTEREST RATE

```
CREATE TABLE json_data.dayhistory
(
   instidentifier character varying(10)
COLLATE pg_catalog."default" NOT NULL,
   lasttradetime bigint NOT NULL,
   quotationlot smallint NOT NULL,
   tradedqty bigint NOT NULL,
   openinterest smallint NOT NULL,
   open numeric NOT NULL,
   high numeric NOT NULL,
   low numeric NOT NULL,
   close numeric NOT NULL,
   CONSTRAINT dayhistory_pkey PRIMARY
KEY (instidentifier, lasttradetime)
)

TABLESPACE pg_default;

ALTER TABLE json_data.dayhistory
   OWNER to postgres;
```

FIGURE 2. Query to store historical data.

movements of Bees in a ABC algorithm are presented in10. From Figure 10, it is clear that, though the pattern of movement is different from the particles of the PSO algorithm, the result is that all the Bees converge to value one, thus attaining the global optima. The movement of ants in an ACO algorithm is depicted in Figure 11. From Figure 11, it is clear that the global optima are attained as all the ants move to value one. The solutions of ABC algorithm is depicted in figure 12. As can be observed from the figure 12 the ABC algorithm is successfully identifying the patterns.

The average and total return of individual stock of GMV portfolio is visualized in figures 13 and 14 respectively. The stock ticks data is fed to the built portfolio. The figure 15 depicts the real-time optimized portfolio returns against NSE

```
CREATE TABLE json_data.stockpkatwo
(
   exchange character varying(10) COLLATE pg_catalog."default"
NOT NULL,
   instrumentidentifier character varying(15) COLLATE
pg_catalog."default" NOT NULL,
   lasttradetime bigint NOT NULL,
   servertime bigint NOT NULL,
   averagetradedprice numeric NOT NULL,
   buyprice numeric NOT NULL,
   buyqty bigint NOT NULL,
   close numeric NOT NULL,
   high numeric NOT NULL,
   low numeric NOT NULL,
   lasttradeprice numeric NOT NULL,
   lasttradeqty bigint NOT NULL,
   open numeric NOT NULL,
   openinterest smallint NOT NULL,
   quotationlot double precision NOT NULL,
   sellprice numeric NOT NULL,
   sellqty bigint NOT NULL,
   totalqtytraded bigint NOT NULL,
   value numeric NOT NULL,
   preopen boolean NOT NULL,
   pricechange numeric NOT NULL,
   pricechangepercentage numeric NOT NULL,
   openinterestchange smallint NOT NULL,
   messagetype character varying(20) COLLATE
pg_catalog."default" NOT NULL
)

TABLESPACE pg_default;

ALTER TABLE json_data.stockpkatwo
   OWNER to postgres;
-- Index: stockpkatwo_lasttradetime_idx

-- DROP INDEX json_data.stockpkatwo_lasttradetime_idx;

CREATE INDEX stockpkatwo_lasttradetime_idx
   ON json_data.stockpkatwo USING btree
   (lasttradetime DESC NULLS FIRST)
   TABLESPACE pg_default;

-- Trigger: ts_insert_blocker

-- DROP TRIGGER ts_insert_blocker ON json_data.stockpkatwo;

CREATE TRIGGER ts_insert_blocker
   BEFORE INSERT
   ON json_data.stockpkatwo
   FOR EACH ROW
   EXECUTE FUNCTION _timescaledb_internal.insert_blocker();
```

FIGURE 3. Query to store real-time data in TimeScaleDB.

Input Stock prices, Tone
Output Optimal Portfolio

```
procedure Active Portfolio Optimization
    while sliding window Do
        for K = 2 to N Do
            Construct EW, MCW, GMV and MSR portfolios by chosing stocks from different K-Means cluster
            Construct GA with two objective functions (MV, SR weights)
            Compare portfolio returns with respect to K
        end for
        Find K which maximizes the portfolio return
        From Derived K
        Construct GA with two objective functions (MV, SR weights)
        Return four optimal portfolio
    end while
end procedure
```

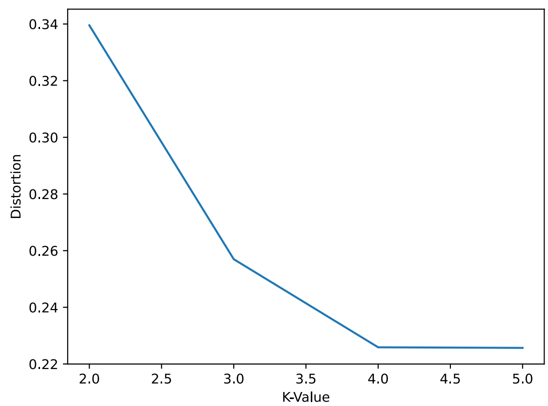

FIGURE 4. Optimal K for 5 stocks.

Nifty-100 returns. It can be easily observed that all the optimized portfolios are performing better than the Nifty-100

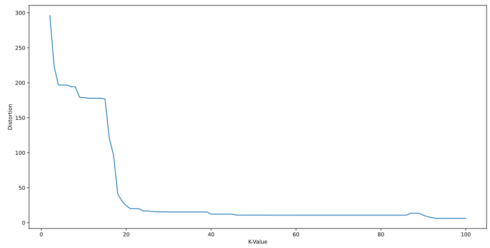

FIGURE 5. Optimal K for 100 stocks.

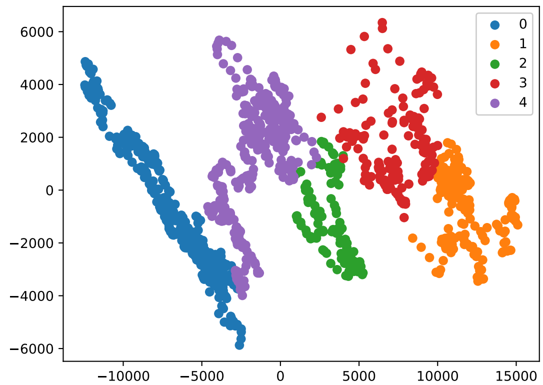

FIGURE 6. Sample KMEANS result with K = 5.

TABLE 3. Global minimum portfolio weights.

7.38323440e-02	4.88294573e-04	1.57975620e-04	3.44180702e-02
1.87396669e-04	2.33072631e-03	1.10943716e-07	9.47176190e-08
9.51431000e-08	5.23287140e-05	1.05973314e-07	1.04762204e-07
1.02095736e-07	9.45532645e-08	8.49421793e-08	8.77922450e-08
8.66861094e-08	6.73720365e-08	4.81186045e-08	3.79215934e-08
3.84622994e-05	7.47732877e-04	5.94359186e-03	4.26551275e-08
1.33445542e-05	6.51629774e-03	1.76676558e-04	4.37562027e-05
2.93616995e-02	1.98373673e-01	1.90855038e-02	2.57183959e-02
1.16430292e-01	1.65816857e-03	3.42325967e-03	4.92977302e-03
5.38895364e-03	8.96001645e-03	6.48036465e-04	1.14570387e-01
6.15277377e-05	1.81996630e-01	3.67136662e-02	5.80156357e-02
3.56690899e-03	1.84183925e-02	1.72433113e-02	3.45879518e-07
1.57130186e-07	1.82317472e-05	1.83626691e-03	6.02507826e-07
2.16036619e-07	5.85153484e-05	5.79173607e-05	2.47462791e-04
2.30294151e-04	8.76936841e-08	6.18178541e-08	4.41446399e-06
3.42144089e-06	1.41134160e-06	5.55263656e-08	5.83590318e-08
7.46712263e-08	1.23025434e-07	1.10862715e-07	1.08635001e-07
1.14957269e-07	1.04036243e-07	9.46094513e-08	9.77246249e-08
9.29291967e-08	9.23926603e-08	9.00336823e-08	9.07991609e-08
3.25265755e-04	2.32242423e-04	3.31572814e-03	2.50216068e-04
5.74401355e-03	8.72135321e-08	5.76452863e-06	8.00008253e-05
4.98844023e-05	2.23117058e-04	1.10345224e-07	1.08017229e-07
1.09138821e-07	8.53776814e-08	7.26220969e-08	4.58186792e-08
2.05121552e-08	7.81688215e-03	5.18357481e-03	1.66740790e-04
4.67143254e-03	2.71671915e-05	1.62654208e-05	8.48362091e-08

benchmark returns. The rolling return for four years and fifty year time periods is shown in figures 16 and 17 respectively. The rolling return gives an unbiased view of returns over a period. The six-month rolling return is plotted. Similarly, a rolling Sharpe ratio for a four and fifty-year time frame is depicted in figures 18 and 19 respectively. Whenever the Sharpe ratio is above one, the investment decision is wise, and the decision is terrible when the Sharpe ratio is below zero. Finally, a rolling beta for four and fifty years time horizon

TABLE 4. Maximum sharpe ratio portfolio weights.

1.32068236e-01	1.38635156e-02	2.86399821e-05	5.01093080e-05
2.16017320e-02	8.80531398e-05	1.04656931e-04	1.23476498e-04
7.36632397e-05	1.03110076e-04	1.42099233e-04	9.98089257e-05
4.62211465e-05	5.23287379e-06	9.30726379e-06	1.03693427e-04
6.97960384e-05	2.39125043e-05	1.05811190e-07	2.88891444e-07
1.03148776e-06	3.68100337e-04	4.98056219e-07	6.74166509e-04
1.16896483e-06	5.34546427e-07	3.73693771e-07	1.59890695e-07
3.05110529e-05	1.72986491e-01	2.79454153e-02	3.04081884e-02
5.62916910e-02	5.05423167e-02	1.84285660e-05	2.71548124e-05
3.92516169e-05	1.49283475e-05	7.36966844e-06	4.02580655e-02
1.76768261e-07	1.88814039e-01	9.71444703e-02	1.08884162e-01
6.42398644e-07	1.60292612e-05	2.82654833e-03	2.15818781e-05
1.82742728e-05	5.75997972e-05	8.41670306e-05	2.73851555e-05
2.75487729e-05	6.43116961e-05	3.22909783e-05	4.89989455e-05
2.13819016e-05	5.69900853e-05	4.13581932e-05	1.39142591e-08
1.00544501e-07	4.75104798e-07	3.39981774e-07	6.56230810e-06
4.55517797e-08	9.00216481e-05	7.51820263e-05	1.18110547e-04
1.40007183e-04	9.70118207e-05	5.05980354e-05	6.61380203e-05
9.52741755e-05	9.02033160e-05	3.74847746e-05	3.84807102e-05
1.79784217e-02	8.66040776e-08	5.88147253e-07	4.82740402e-04
5.94381284e-03	1.50852553e-03	1.26615123e-03	1.38102404e-05
4.68489996e-03	1.08745258e-04	1.26670186e-04	1.49032914e-04
2.16317415e-02	7.19418599e-05	7.71995157e-05	2.61100223e-05
1.18307029e-07	4.17410029e-07	1.29575583e-06	6.96350298e-07
1.97378007e-02	8.19116965e-03	5.87484027e-07	1.04112874e-05

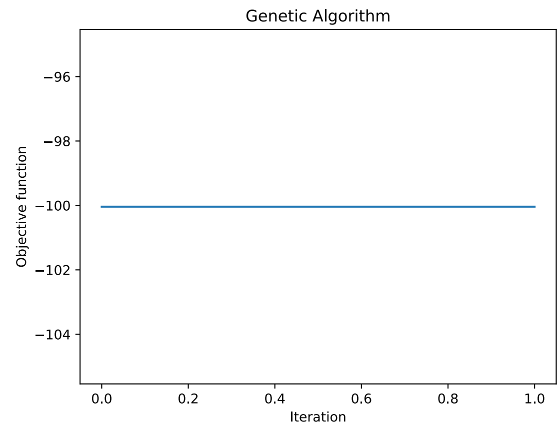

FIGURE 7. Objective function for GMV portfolio.

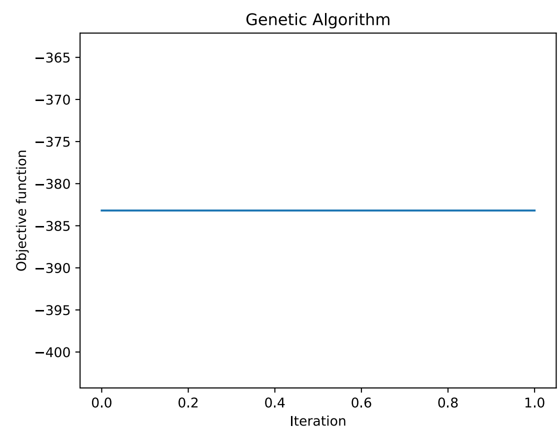

FIGURE 8. Objective function for MSR portfolio.

is illustrated in figures 20 and 21 respectively. It is evident from the figure that the built model is stable as beta is always

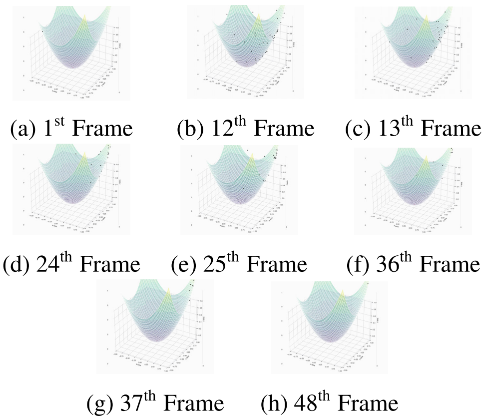

FIGURE 9. Evolution of PSO particles.

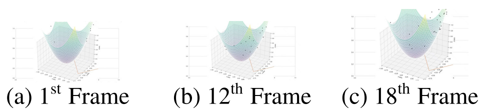

FIGURE 10. The bee movements.

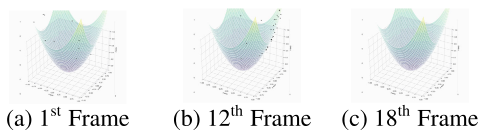

FIGURE 11. The ant movements.

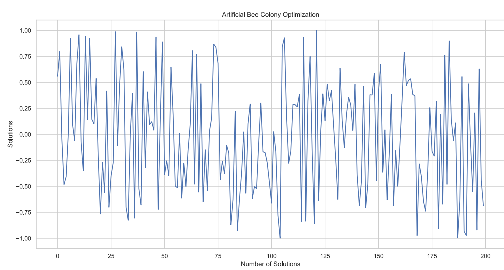

FIGURE 12. ABC algorithm solutions.

TABLE 5. Different portfolio returns.

Portfolio	Return in %	Risk in %	Risk to Reward Ratio
GMV	52.94	0.08	0.002
MSR	55.64	29.33	0.5
EW	50.93	631.49	10.97
CW	51	575.35	11.28

zero. The weights are optimized utilizing GA and the returns obtained with each portfolio is depicted in table 5. It is evident from table 5 that though all four portfolios are outperforming the Nifty-100 benchmark index, The risk to reward ratio of the four portfolios significantly differ. GMV portfolio is a clear winner with risk to reward ratio of 0.002. The close competitor is MSR portfolio, with risk to reward ratio of 0.5. The EW and CW portfolios lag with a risk-to-reward ratio of 10.97 and 11.28, respectively.

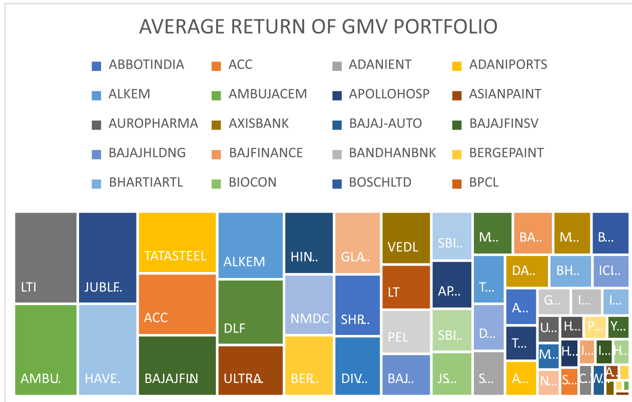

FIGURE 13. Average return of each stock under GMV portfolio.

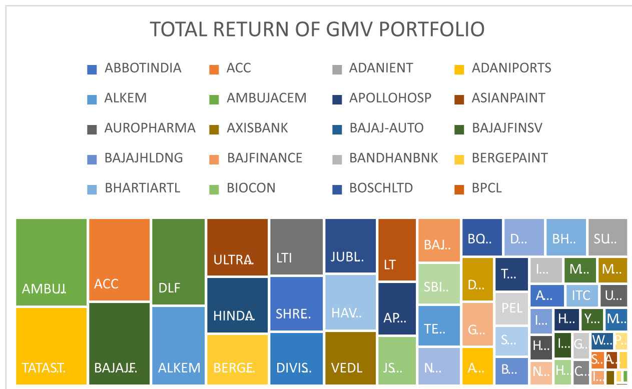

FIGURE 14. Total return of each stock under GMV POrtfolio.

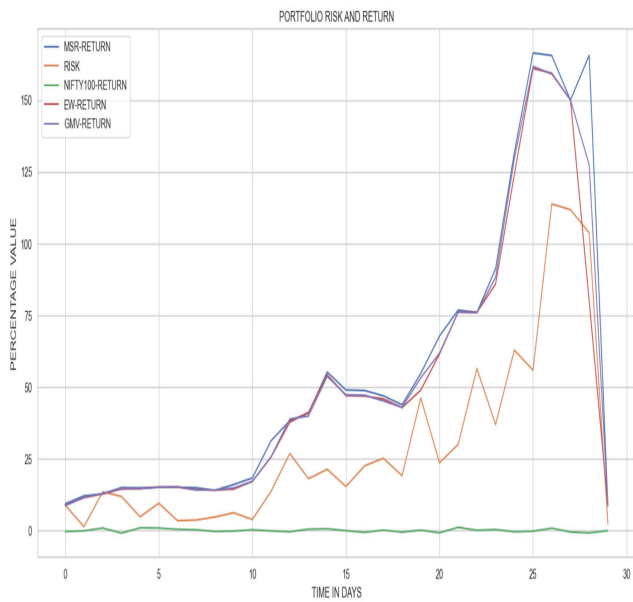

FIGURE 15. Real-time portfolio returns.

For the dataset under consideration, GA, ABC, ACO, and PSO algorithms outperform Simulated Annealing, Tabu search, and Firefly algorithms. GA is more efficient than the

TABLE 6. GMV portfolio returns for various metaheuristic algorithms.

Metaheuristic Algorithm	GMV Return in %	GMV Risk in %	GMV Risk to Reward Ratio
SA	20.38	0.07	0.0034
Tabu	20.18	0.07	0.0034
Firefly	20.95	0.06	0.0028
ACO	50.82	0.09	0.0017
ABC	52.88	0.08	0.0015
PSO	51.73	0.1	0.0019
GA	**52.94**	**0.08**	**0.0015**

TABLE 7. MSR portfolio returns for various metaheuristic algorithms.

Metaheuristic Algorithm	MSR Return in %	MSR Risk in %	MSR Risk to Reward Ratio
SA	25.8	18	0.69
Tabu	25.18	14	0.55
Firefly	25.96	13.85	0.53
ACO	50.88	28.21	0.55
ABC	53.88	29.11	0.54
PSO	53.11	29.71	0.56
GA	**55.64**	**29.33**	**0.52**

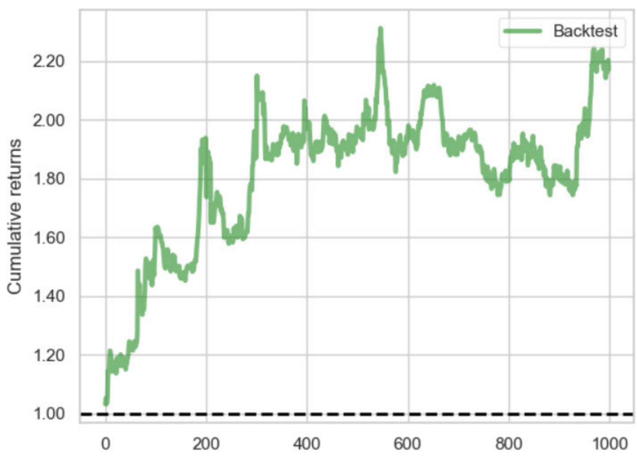

FIGURE 16. Rolling return for four years.

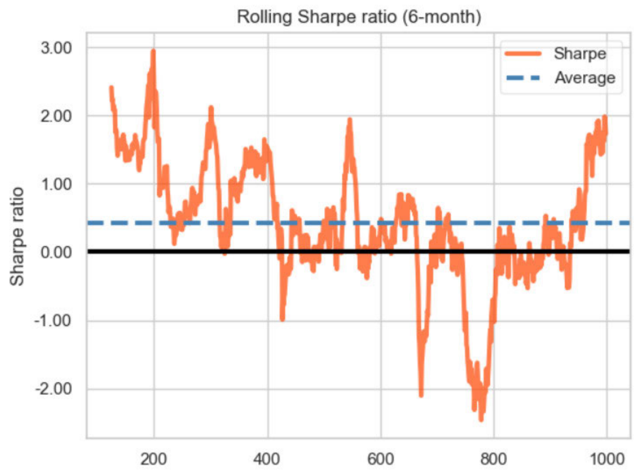

FIGURE 18. Rolling sharpe for four years.

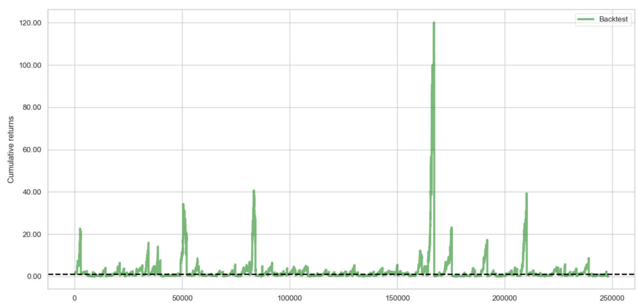

FIGURE 17. Rolling return for fifty years.

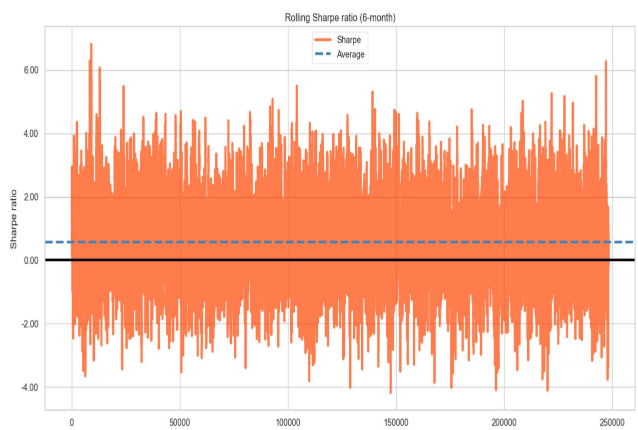

FIGURE 19. Rolling sharpe for fifty years.

ABC, ACO, and PSO algorithms. Simulated Annealing, Tabu search and Firefly algorithms gave portfolio return of 20% for GMV and 25.8% for MSR portfolio respectively. The rest of the algorithms results for GMV and MSR portfolios are depicted in Table 6 and 7 respectively. It is evident from Table 6 and 7 GMV portfolio outperforms all other portfoilos.

GA outperforms other metaheuristic algorithms as there is only risk and return to be optimized. Different metaheuristic

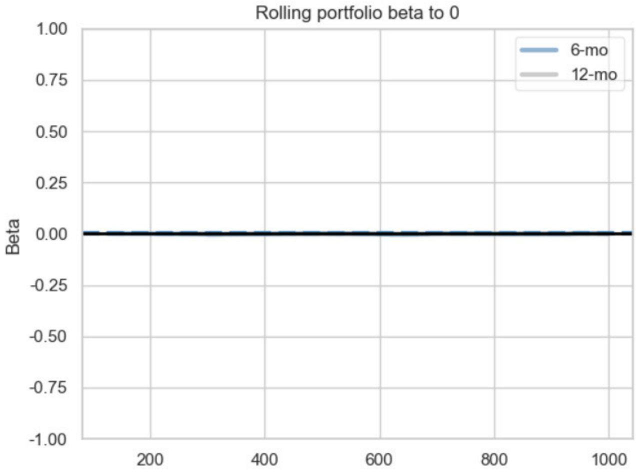

FIGURE 20. Rolling beta for four years.

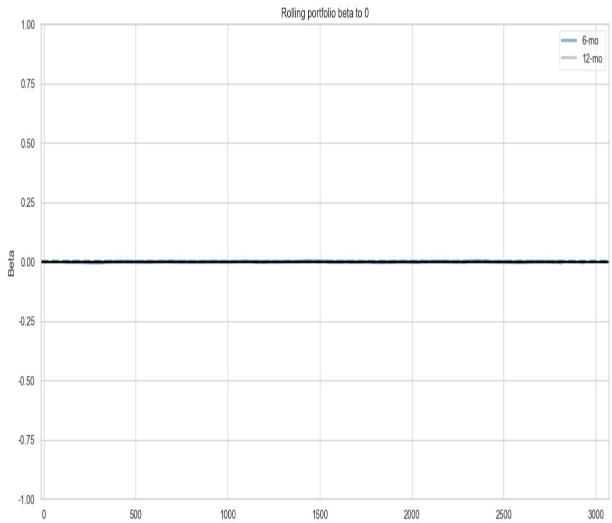

FIGURE 21. Rolling beta for fifty years.

algorithms will fair better if too many objective functions are to be optimized.

V. CONCLUSION

All the constructed portfolios outperform the Nifty-100 index. The genetically optimized global minimum variance portfolio outshines other portfolios. The optimized maximum Sharpe ratio portfolio gives the highest return, but the risk is high.

The seven metaheuristic algorithms are explored and it is observed that GA, ABC, ACO and PSO algorithms outperform SA, Tabu search, and Firefly algorithm. The GA algorithm is simple and efficient compared to ABC, ACO, and PSO algorithms.

The number of groupings increases with the increase in the number of stocks which implies the stocks are falling into different categories. The risk to reward ratio of the global minimum variance portfolio is the highest indicating an attractive investment opportunity. The built models are stable as the rolling beta is zero.

The key takeaways of this manuscript are portfolio selection and ranking with the K-Means algorithm, portfolio optimization through metaheuristic algorithms, and portfolio management with the help of a sliding window.

The deep neural network models can be applied. Current work can be extended to India's entire NSE and BSE stock markets. The number of clusters required for a varying number of stocks and the data size can be analyzed. In the future, the work can be extended to the stock market of other nations.

REFERENCES

[1] Z. Bodie, A. Kane, and A. J. Marcus, *Investments*. New York, NY, USA: McGraw-Hill, 2009.

[2] K. Kremnitzer, "Comparing active and passive fund management in emerging markets," Senior Honors thesis, Econ. Dept., Univ. California, Berkeley, Berkeley, CA, USA, 2012.

[3] G. Curtis, "Modern portfolio theory and behavioral finance," *J. Wealth Manage.*, vol. 7, no. 2, pp. 16–22, Jul. 2004.

[4] A. Arévalo, D. León, and G. Hernandez, "Portfolio selection based on hierarchical clustering and inverse-variance weighting," in *Computational Science ICCS 2019*, J. M. F. Rodrigues, P. J. S. Cardoso, J. Monteiro, R. Lam, V. V. Krzhizhanovskaya, M. H. Lees, J. J. Dongarra, and P. M. A. Sloot, Eds. Cham, Switzerland: Springer, 2019, pp. 315–325.

[5] A. A. Saji, J. J. M. Joseph, and B. S. Kumar, "Portfolio construction and testing using cluster analysis," *Asian J. Manage.*, vol. 11, no. 2, pp. 207–212, 2020.

[6] D. León, A. Aragón, J. Sandoval, G. Hernández, A. Arévalo, and J. Niño, "Clustering algorithms for risk-adjusted portfolio construction," *Proc. Comput. Sci.*, vol. 108, pp. 1334–1343, Jan. 2017.

[7] Y. S. Asawa, "Modern machine learning solutions for portfolio selection," *IEEE Eng. Manag. Rev.*, vol. 50, no. 1, pp. 94–112, Mar. 2022.

[8] S. Lim, M.-J. Kim, and C. W. Ahn, "A genetic algorithm (GA) approach to the portfolio design based on market movements and asset valuations," *IEEE Access*, vol. 8, pp. 140234–140249, 2020.

[9] C.-H. Chen, Y.-H. Chen, J. C.-W. Lin, and M.-E. Wu, "An effective approach for obtaining a group trading strategy portfolio using grouping genetic algorithm," *IEEE Access*, vol. 7, pp. 7313–7325, 2019.

[10] C.-H. Chen, C.-Y. Lu, T.-P. Hong, J. C.-W. Lin, and M. Gaeta, "An effective approach for the diverse group stock portfolio optimization using grouping genetic algorithm," *IEEE Access*, vol. 7, pp. 155871–155884, 2019.

[11] C.-H. Chen, B.-Y. Chiang, T.-P. Hong, D.-C. Wang, and J. C.-W. Lin, "An approach for diverse group stock portfolio optimization using the fuzzy grouping genetic algorithm," in *Intelligent Information and Database Systems*, N. T. Nguyen, D. H. Hoang, T.-P. Hong, H. Pham, and B. Trawiński, Eds. Cham, Switzerland: Springer, 2018, pp. 510–518.

[12] P.-Y. Kang, I.-C. Wu, and C.-H. Hsueh, "Applying hueristic algorithms to portfolio selection problem," in *Proc. Conf. Technol. Appl. Artif. Intell. (TAAI)*, Nov. 2015, pp. 323–329.

[13] D. G. Cortes, A. J. C. Jofre, and L. S. Martin, "Artificial intelligence based method for portfolio selection," in *Proc. Congreso Internacional de Innovación y Tendencias en Ingeniería (CONIITI)*, Oct. 2018, pp. 1–4.

[14] C.-H. Chen, Y.-H. Chen, V. G. Diaz, and J. C.-W. Lin, "An intelligent trading mechanism based on the group trading strategy portfolio to reduce massive loss by the grouping genetic algorithm," *Electron. Commerce Res.*, vol. 23, pp. 3–42, Mar. 2021.

[15] M. K. de Melo, R. T. N. Cardoso, and T. A. Jesus, "A genetic algorithm for investment tracking with stochastic model predictive control," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jun. 2021, pp. 1543–1550.

[16] C. Kumar and M. N. Doja, "An optimizing utility for portfolio selection based on optimal values computed using ANN, NSGA-II and machine learning technique," in *Proc. Int. Conf. Power Electron., Control Autom. (ICPECA)*, Nov. 2019, pp. 1–5.

[17] C.-H. Chen, T.-P. Hong, and S.-C. Chu, "Diverse group stock portfolio optimization based on investor sentiment index," in *Proc. Conf. Technol. Appl. Artif. Intell. (TAAI)*, Dec. 2017, pp. 100–103.

[18] M. M. Solin, A. Alamsyah, B. Rikumahu, and M. A. Arya Saputra, "Forecasting portfolio optimization using artificial neural network and genetic algorithm," in *Proc. 7th Int. Conf. Inf. Commun. Technol. (ICoICT)*, Jul. 2019, pp. 1–7.

[19] R. Faia, T. Pinto, Z. Vale, J. M. Corchado, J. Soares, and F. Lezama, "Genetic algorithms for portfolio optimization with weighted sum approach," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2018, pp. 1823–1829.
[20] P. Gupta, M. K. Mehlawat, and G. Mittal, "Asset portfolio optimization using support vector machines and real-coded genetic algorithm," *J. Global Optim.*, vol. 53, pp. 297–315, Jun. 2012.
[21] S. K. Mittal and N. Srivastava, "Mean-variance-skewness portfolio optimization under uncertain environment using improved genetic algorithm," *Artif. Intell. Rev.*, vol. 54, no. 8, pp. 6011–6032, Dec. 2021.
[22] B. Li, "Research on optimal portfolio of financial investment based on genetic algorithm," in *Proc. Int. Conf. Econ. Manage. Model Eng. (ICEMME)*, Dec. 2019, pp. 497–500.
[23] P. Sinha, A. Chandwani, and T. Sinha, "Algorithm of construction of optimum portfolio of stocks using genetic algorithm," *Int. J. Syst. Assurance Eng. Manage.*, vol. 6, no. 4, pp. 447–465, Dec. 2015.
[24] Y.-H. Chou, S.-Y. Kuo, and Y.-T. Lo, "Portfolio optimization based on funds standardization and genetic algorithm," *IEEE Access*, vol. 5, pp. 21885–21900, 2017.
[25] A. M. Rather, V. N. Sastry, and A. Agarwal, "Portfolio selection using maximum-entropy gain loss spread model: A GA based approach," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2014, pp. 400–406.
[26] F. Z. Habbab, M. Kampouridis, and A. A. Voudouris, "Optimizing mixed-asset portfolios involving REITs," in *Proc. IEEE Symp. Comput. Intell. for Financial Eng. Econ. (CIFEr)*, May 2022, pp. 1–8.
[27] T. Onek, "A genetic algorithm model for financial asset diversification," in *Proc. SoutheastCon*, Apr. 2019, pp. 1–6.
[28] M. K. Pareek and P. Thakkar, "Surveying stock market portfolio optimization techniques," in *Proc. 5th Nirma Univ. Int. Conf. Eng. (NUiCONE)*, Nov. 2015, pp. 1–5.
[29] P. Agrawal, H. F. Abutarboush, T. Ganesh, and A. W. Mohamed, "Metaheuristic algorithms on feature selection: A survey of one decade of research (2009–2019)," *IEEE Access*, vol. 9, pp. 26766–26791, 2021.
[30] S. Liu and C. Xiao, "Application and comparative study of optimization algorithms in financial investment portfolio problems," *Mobile Inf. Syst.*, vol. 2021, pp. 1–10, Nov. 2021.
[31] A. Antony, "Behavioral finance and portfolio management: Review of theory and literature," *J. Public Affairs*, vol. 20, no. 2, p. e1996, May 2020.
[32] Z. Bodie, A. Kane, A. J. Marcus, and P. Mohanty, *Investments*, 11th ed. India: McGraw-Hill.
[33] G. A. V. Pai and T. Michel, "Evolutionary optimization of constrained k-means clustered assets for diversification in small portfolios," *IEEE Trans. Evol. Comput.*, vol. 13, no. 5, pp. 1030–1053, Oct. 2009.
[34] S. Luke, *Essentials of Metaheuristics*, vol. 2. Raleigh, NC, USA: Lulu Raleigh, 2013.
[35] A. Colorni, M. Dorigo, and V. Maniezzo, "Distributed optimization by ant colonies," in *Proc. 1st Eur. Conf. Artif. Life*, vol. 142, Dec. 1991, pp. 134–142.
[36] R. Eberhart, "James. kennedy, particle swarm optimization," in *Proc. IEEE Int. Conf. neural Netw., Aust.*, vol. 1948, 1942.
[37] C. Yu, N. Lahrichi, and A. Matta, "Optimal budget allocation policy for Tabu search in stochastic simulation optimization," *Comput. Oper. Res.*, vol. 150, Feb. 2023, Art. no. 106046.
[38] P. K. Aithal, A. U. Dinesh, and M. Geetha, "Identifying significant macroeconomic indicators for Indian stock markets," *IEEE Access*, vol. 7, pp. 143829–143840, 2019.
[39] P. K. Aithal, U. D. Acharya, M. Geetha, and P. Menon, "Building a calendar of events database by analyzing financial spikes," *IEEE Access*, vol. 9, pp. 114192–114206, 2021.

PRAKASH K. AITHAL (Member, IEEE) is currently pursuing the Ph.D. degree with the Manipal Academy of Higher Education, Manipal, India. He is an Assistant Professor with the Department of Computer Science and Engineering, Manipal Academy of Higher Education, Manipal Institute of Technology, Manipal. He has presented several papers at national and international conferences. His work has been published in several international journals. His research interest includes data analytics in the financial domain.

GEETHA M. (Member, IEEE) received the Ph.D. degree from NITK Surathkal, in 2010. She is currently a Professor with the Department of Computer Science and Engineering, Manipal Academy of Higher Education, Manipal Institute of Technology, Manipal, India. She has presented several papers at national and international conferences. Her work has been published in several international journals. Her research interests include data mining and text mining in the healthcare and financial sectors.

U. DINESH ACHARYA received the Ph.D. degree from the Manipal Academy of Higher Education, Manipal, India, in 2008. He is currently a Professor with the Department of Computer Science and Engineering, Manipal Academy of Higher Education, Manipal Institute of Technology, Manipal. He has presented several papers at national and international conferences. His work has been published in several international journals. His current research interests include data analytics in the healthcare, agriculture, and financial sectors.

BASRI SAVITHA received the M.A. degree in economics and the M.B.A., H.G.Dip. (U.K.), M.Phil., Ph.D., and PGDHE degrees. She is currently an Associate Professor with the Manipal Institute of Management, Manipal Academy of Higher Education, Manipal, India. She has evaluated the impact of micro health insurance schemes in Karnataka for her Ph.D. study with the National Institute of Technology Karnataka. She has published more than 22 research articles in national and international journals. Her research interests include banking and insurance, and financial markets. She was awarded the Kalam Travel Grant to present her research paper from the Kalam Research Conference, University of Kentucky.

PARTHIV MENON is currently pursuing the degree with the Department of Computer Science and Engineering, Manipal Academy of Higher Education, Manipal Institute of Technology, Manipal, India. His research interest includes data analytics.