Expert Systems With Applications 179 (2021) 115060

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Robo-advisor using genetic algorithm and BERT sentiments from tweets for hybrid portfolio optimisation

Edmund Kwong Wei Leow [a,1], Binh P. Nguyen [b,*,2], Matthew Chin Heng Chua [a,3]

[a] Institute of Systems Science, National University of Singapore, 29 Heng Mui Keng Terrace, 119620, Singapore
[b] School of Mathematics and Statistics, Victoria University of Wellington, Kelburn Parade, Wellington 6140, New Zealand

ARTICLE INFO

Keywords:
BERT
Portfolio optimisation
Genetic algorithm
Social media sentiment analysis
Twitter

ABSTRACT

Robo-advisors are increasingly popular, with machine learning algorithms taking centre stage for researchers. However, classical financial theories and techniques, such as Constant Rebalancing (CRB) and Modern Portfolio Theory (MPT), can still be relevant by combining them with social media sentiments. In this study, we propose two novel models, namely Sentimental All-Weather (SAW) and Sentimental MPT (SMPT), which capture the up-to-date market conditions through Twitter sentiments via Google's Bidirectional Transformer (BERT) model. Genetic Algorithm was used to optimise the models for different objectives including maximising cumulative returns and minimising volatility. Trained on tweets and the United States stock data from August 2018 to end December 2019, and tested on an out-of-sample period from January 2020 to April 2020, our proposed models achieved superior performance in terms of common measures of portfolio performance including Sharpe ratio, cumulative returns, and value-at-risk, compared to the following benchmarks: buy-and-hold SPY index, MPT model, and CRB model for an All-Weather Portfolio.

1. Introduction

Innovations in financial technology (FinTech) are driving the banking sector, with more changes expected in the next couple of years than in the past two centuries. One of the most important changes is the introduction of advanced technologies, such as machine learning algorithms, to facilitate security trading and advisory services to investors, collectively termed robo-advisors (Phoon & Koh, 2018).

Abbreviation	Meaning
BERT	Bidirectional Transformer
CNN	Convolutional Neural Network
CRB	Constant Rebalancing, which is one of the methods for stock rebalancing where rebalancing follows a certain percentage for each asset in the portfolio
DNN	Deep Neural Network
ETF	Exchange Traded Funds
FinTech	Financial Technology
GA	Genetic Algorithm
HRP	Hierarchial Risk Parity
LR	Logistic Regression
LSTM	Long Short-Term Memory
MACD	Moving Average Convergence Divergence, which is a trend-following momentum indicator that shows the relationship between two moving averages of a security's price
MPT	Modern Portfolio Theory, also known as Mean–Variance Optimisation (MVO), which is one of the methods for stock rebalancing, where the expected return (mean) in a universe of assets is maximised given a constraint on its risk (variance), forming the Markowitz's Efficient Frontier
NLP	Natural Language Processing
RF	Random Forest
RL	Reinforcement Learning
SAW	Sentimental All-Weather which is Ray Dalio's All-Weather Portfoilo augmented by Twitter sentiments that is described in this paper
SMPT	Sentimental Modern Portfolio Theory which is Modern Portfolio Theory augmented by Twitter sentiments that is described in this paper
SPDR	Standard & Poor's Depositary Receipt, which are a family of ETFs traded in the United States

(continued on next page)

* Corresponding author.
E-mail addresses: e0401999@u.nus.edu (E.K.W. Leow), binh.p.nguyen@vuw.ac.nz (B.P. Nguyen), mattchua@nus.edu.sg (M.C.H. Chua).
[1] ORCID ID: https://orcid.org/0000-0003-1548-643X
[2] ORCID ID: https://orcid.org/0000-0001-6203-6664
[3] ORCID ID: https://orcid.org/0000-0002-5200-5079

https://doi.org/10.1016/j.eswa.2021.115060
Received 28 January 2021; Received in revised form 23 March 2021; Accepted 15 April 2021
Available online 22 April 2021
0957-4174/

(*continued*)

SPY	The trading symbol for SPDR S&P 500 trust, which is an ETF which trades on the NYSE Arca
SVM	Support Vector Machine
VADER	Valence Aware Dictionary and sEntiment Reasoner

The opportunity for robo-advisors is tremendous as it enables small investors to gain entry to low-fee automated wealth management, with low-minimum investment requirements. Another key advantage of robo-advisors is that they are less vulnerable to potential conflicts of interest. This is because they provide more transparent and significantly lower cost structures, compared to human financial advisors who are often prone to misguided incentive-based compensation schemes (Brenner & Meyll, 2020).

In recent years, several studies (Gomes & Selman, 2001) have shown that artificial intelligence can achieve better portfolio management performance than using existing strategies based on classical financial theories and traditional models, such as Modern Portfolio Theory (MPT) (Markowitz, 2013), and Hierarchical Risk Parity (HRP) (Lopez de Prado, 2016). However, the decisions of machine learning methods in a robo-advisor are difficult or impossible to explain, which can hamper users' trust in the system, and can lead to the rejection of the system (Rai, 2020). Most existing robo-advisors in the market also still utilise MPT, and is the method that fund managers are most familiar with Lam (2016).

At the same time, with the proliferation of social media, and advances in Natural Language Processing that has enabled the conversion of sentences into sentiments, in what is called sentiment analysis, there has been growing interest and research in this field. Because stock markets have been shown to be often driven by sentiments (Lasek & Lasek, 2015), there is opportunity to utilise sentiments from social media (such as Twitter) as a trading signal by forecasting possible future price to augment existing portfolio algorithms. For instance, Peterson (2016) built a market-neutral social media-based hedge fund that beat the S&P 500 by more than twenty-four percent.

In this paper, we propose a hybrid approach of using traditional portfolio techniques together with Twitter sentiments to improve portfolio performance, and introduce two models: Sentimental All-Weather (SAW) and Sentimental Modern Portfolio Theory (SMPT). The use of traditional portfolio techniques as a backbone will make it easier for clients and managers alike to understand, while the addition of Twitter sentiments will make it sensitive towards potential market dips and spikes.

2. Related work and theories

2.1. Stock selection and prediction

In creating a portfolio, it is common to adopt diversification to reduce total variance while maintaining expected return on investment, by allocating investments across different financial instruments. It is ineffective to merely invest in a large number of different assets, but to instead focus on minimising correlation between all the assets (Lekovic, 2018). Common traditional methods of accomplishing the above was by allocating across either different industries/sectors (health care, real estate, utilities, etc.) or different asset types (equity, bonds, commodity, etc.).

For each of these classes, stocks may be selected traditionally based on key technical and fundamental features. For long-term investments, fundamental analysis, which examines a company's management structure, industry position, competitors, income, growth potential, growth rate, and revenues to try to determine if it is a good value, is well-suited. For short-term investments, technical analysis, which focuses on patterns within stock charts to forecast volume trends and pricing in future, is often used.

In recent years, machine learning techniques have also been used for stock selection. Fu et al. (2018) used Genetic Algorithm (GA) to select 114 features out of 244 technical and fundamental features, and used a variety of machine learning algorithms including Random Forest (RF), Logistic Regression (LR), Deep Neural Network (DNN), and stacking of RF with DNN to classify stocks as either good or bad.

It is also common to use machine-learning models to predict stock trends and returns. Chen, Zhou, and Dai (2018) used Long Short-Term Memory (LSTM) for China stock index prediction with low-frequency data. Yang, Li, Chen, Cao, and Jiang (2019) used Convolutional Neural Network (CNN) and LSTM models with high-frequency price–volume data to predict the expected return rate on the current day, and select stocks with the highest expected yield at the opening to maximise total returns.

Shen, Jiang, and Zhang (2012) recognised and exploited the temporal correlation among global stock markets and various financial products to predict the next-day stock trend with the aid of Support Vector Machine (SVM), and created a practical trading model for the stock index.

The performance of price-prediction-based algorithms depends on the degree of prediction accuracy but future market prices are difficult to predict (Jiang, Xu, & Liang, 2017), so many have converted the problem to a Reinforcement Learning one. Li, Dagli, and Enke (2007) proposed actor-critic reinforcement learning-based system that could forecast short-term stock price movements. Deng, Bao, Kong, Ren, and Dai (2016) constructed a new model, Deep Direct Reinforcement Learning (DDRL), which combined direct reinforcement learning with deep recurrent neural network, for futures contract trading.

However, all of the above machine learning approaches typically seek to maximise returns, do not compare with traditional portfolio approaches such as Modern Portfolio Theory, and do not consider other factors such as differing objectives, risk appetites, and maximum draw-down, which we deem necessary in a robo-advisor.

2.2. Sentiments based on financial tweets for stock prediction and rebalancing

Due to the enormous interest and applicability in deriving automated sentiment or polarity analysis from text using natural language processing (NLP), there has been a large number of different methods in recent years. They can be split into two main methods, namely lexical-based and supervised machine learning methods.

Lexical methods typically use a predefined list of words, where each word is associated with a specific sentiment, but are very dependent on the context which they were created for Ribeiro, Araújo, Gonçalves, Benevenuto, and Gonçalves (2015). Some, such as Valence Aware Dictionary and sEntiment Reasoner (VADER) (Hutto & Gilbert, 2014), also combine the lexicon with processing of sentence characteristics to determine sentence polarity.

In recent years, deep learning models, particularly Google's Bidirectional Transformer (BERT) model (Sun, Huang, & Qiu, 2019; Xu, Liu, Shu, & Yu, 2019) and its variants, have gained unprecedented popularity and set new records in NLP tasks in the GLUE (General Language Understanding Evaluation) Benchmark.

In general, however, no single method always achieves the best prediction performance for all different datasets (Ribeiro et al., 2015). Therefore, whichever the method chosen, the model should be fine-tuned based on the specific dataset. Araci (2019) acknowledged that financial texts have a specialised language with unique vocabulary, and have a tendency to use vague expressions instead of easily identified negative/positive words, which make models trained on general corpora ill-suited for this purpose. Therefore, he fine-tuned BERT for financial data (FinBERT), and showed that it outperforms state-of-the-art machine learning methods for financial sentiment analysis datasets (Araci, 2019).

Commercialisation of financial sentiment analysis is also increasing, with sentiment analysis vendors such as Sentdex, PsychSignal, and

Accern, offering sentiment analysis as a service.

In utilising sentiments for stock analysis and prediction, there have been numerous research and studies. Zhang and Skiena (2010) used blogs and news data to derive company-specific sentiments which are then used to rank and shortlist top and bottom companies for a long-short strategy based on the sentiments. For social media sentiments, Bollen, Mao, and Zeng (2011) used a fuzzy neural network and showed that the public collective mood (happy, calm, anxiety) derived from Twitter are correlated to the value of the Dow Jones Industrial Index. A number of articles (Pagolu, Challa, Panda, & Majhi, 2016; Sul, Dennis, & Yuan, 2017) also found that public sentiments in tweets about specific firms were significantly related to stock market movements of the companies on subsequent days. More recently, it was also shown that strongly negative tweets from just an individual account (president Donald Trump) causes subsequent short-term reduction of market value of the company mentioned (Brans & Scholtens, 2020). However, in the above approaches, there was no performance comparison of prediction with any benchmarks, and the studies mostly focused on company-specific news and prediction.

2.3. Stock rebalancing

A portfolio's asset allocation determines the portfolio's risk and return characteristics. To maintain its original risk and return characteristics over time, the portfolio must be rebalanced either periodically at fixed intervals (e.g. monthly, semi-annually, annually) and/or at certain trigger thresholds (e.g. exceed 5% of desired allocation percent).

One of the simplest traditional methods for stock rebalancing is the Constant Rebalanced (CRB) portfolio, where rebalancing follows a certain percentage for each asset in the portfolio. A popular CRB is the 60/40 rule, which advocates that 60 percent of the portfolio be invested in potentially higher risk and historically higher return assets, such as stocks, and the other 40 percent in lower risk assets such as government bonds. Another popular CRB is the simplified All-Weather portfolio by Ray Dalio, which simplifies the original All-Weather risk parity approach of four equal-risk economic seasons into fixed-proportion allocation (30% stocks, 40% long-term bonds, 15% intermediate-term bonds, 7.5% gold, and 7.5% commodities) (Robbins, 2014).

Another method for stock rebalancing is Modern Portfolio Theory (MPT) or Mean–Variance Optimisation (MVO) (Markowitz, 2013), where the expected return (mean) in a universe of assets is maximised given a constraint on its risk (variance), forming the Markowitz's Efficient Frontier. However, in practice, it sometimes gave unintuitive weights that did not make sense to general investors, due to estimation errors in returns and variance (Xu, Chen, & Tsui, 2018). Hence, alternative returns and risk models were proposed by others. For instance, this was addressed by Black and Litterman (1991), with the Black-Litterman model, which uses a Bayesian approach to combine a prior estimate of returns with views on certain assets, so as to produce a posterior estimate of expected returns that is more intuitive (Idzorek, 2007). Similarly, instead of using the raw sample covariance matrix as the risk model, Ledoit and Wolf (2001) and Ledoit and Wolf (2004) proposed transforming it (shrinking) with a structured estimator to pull the most extreme coefficients toward more central values, systematically reducing estimation error when it matters most. These alternative returns and risk models could then be used in MPT optimisation for better performance.

A relatively new technique for stock rebalancing is Hierarchical Risk Parity (Lopez de Prado, 2016), introduced by Lopez de Prado, which performs hierarchical clustering on the covariance matrix of stock returns to find a diversified weighting by distributing capital to each cluster hierarchy. This gets around the issue of inverting the covariance matrix in classic MPT optimisation.

With regard to rebalancing, machine learning techniques have also been utilised. Reinforcement Learning (RL) is popular as a machine learning technique for portfolio management since the stock market price dynamics are difficult to predict using unsupervised and supervised learning paradigms. However, RL has the ability to capture the temporal difference in the signals and learn the market dynamics over time. Aboussalah and Lee (2020) used Stacked Deep Dynamic Recurrent Reinforcement Learning (SDDRRL) to construct real-time rebalancing optimal portfolio based on market conditions, so as to maximise Sharpe ratio, and showed that it achieves better performance than the MPT and risk parity models after 20 successive rounds for 10 selected stocks from different sectors of the S&P 500. Similarly, Liang, Chen, Zhu, Jiang, and Li (2018) used continuous reinforcement learning algorithms, namely Deep Deterministic Policy Gradient (DDPG), Proximal Policy Optimisation (PPO), and Policy Gradient (PG), for portfolio management in the China stock market. Although being used widely in portfolio management, deep reinforcement learning is often highly sensitive with an unstable performance, and tends to shift all holdings to one asset at a time (Liang et al., 2018).

However, in spite of their rising popularity, machine learning methods are often a black box, which makes it very difficult or even impossible to explain to clients how the algorithms decided on their portfolio weights when used in a robo-advisor, and this inscrutability can hamper users' trust in the system, especially in a robo-advisor where the investor's money is at stake, and can lead to the rejection of the system (Rai, 2020).

Therefore, in designing the robo-advisor, we wanted a portfolio algorithm that is.

- comprehensive to the average investor (should not be a black box; back-tested returns and risks should be visible)
- transparent – client should know what they are holding in the portfolio (i.e., holdings should not keep changing)
- adaptive to market conditions
- as simple as possible while being able to out-perform the market and traditional algorithms

Almost all prior work adopt an either-or approach, where they try to advocate a new model that is supposedly better than others. Here, we will attempt a hybrid approach, using traditional portfolio optimisation approaches as a baseline and augmenting portfolio performance using sentiments derived from Twitter. The usage of traditional portfolio optimisation as a backbone is advantageous as it is mostly familiar to investors and fund managers alike, and is easier to explain. By using Twitter sentiments as an insight into overall investor sentiments, we will also be pro-active towards potential market dips and spikes, and be able to increase our cumulative returns. An overview of the proposed approach is illustrated in Fig. 1. The decision points as well as key functions will be explained in the subsequent sub-sections.

3. Proposed approach

3.1. ETF selection

The first requirement is to choose the basket of Exchange-Traded Funds (ETFs) (universe). This is the "ETF selection" block in Fig. 1. Only the United States (US) ETFs will be considered as they are ideal for automated portfolios/ robo-advisors due to their broad diversification and low cost (Poterba & Shoven, 8781). There are many possibilities for ETF selection as mentioned in Section 2.1, including machine learning algorithms. For simplicity, we will consider only (i) one set of ETFs for sector-based diversification and (ii) another set for asset-based diversification. These can be easily switched out with more complex stock-selection algorithms in the overall flow if required.

The sector-based basket of ETFs comprises 11 SPDR sector ETFs (namely XLE, XLRE, XLF, XLV, XLC, XLI, XLY, XLP, XLB, XLK, XLU) which are ETFs of S&P 500 stocks grouped by Global Industry Classification Standard (GICS) sector classification to facilitate passive investment in specific sectors of the US economy, to create a diversified

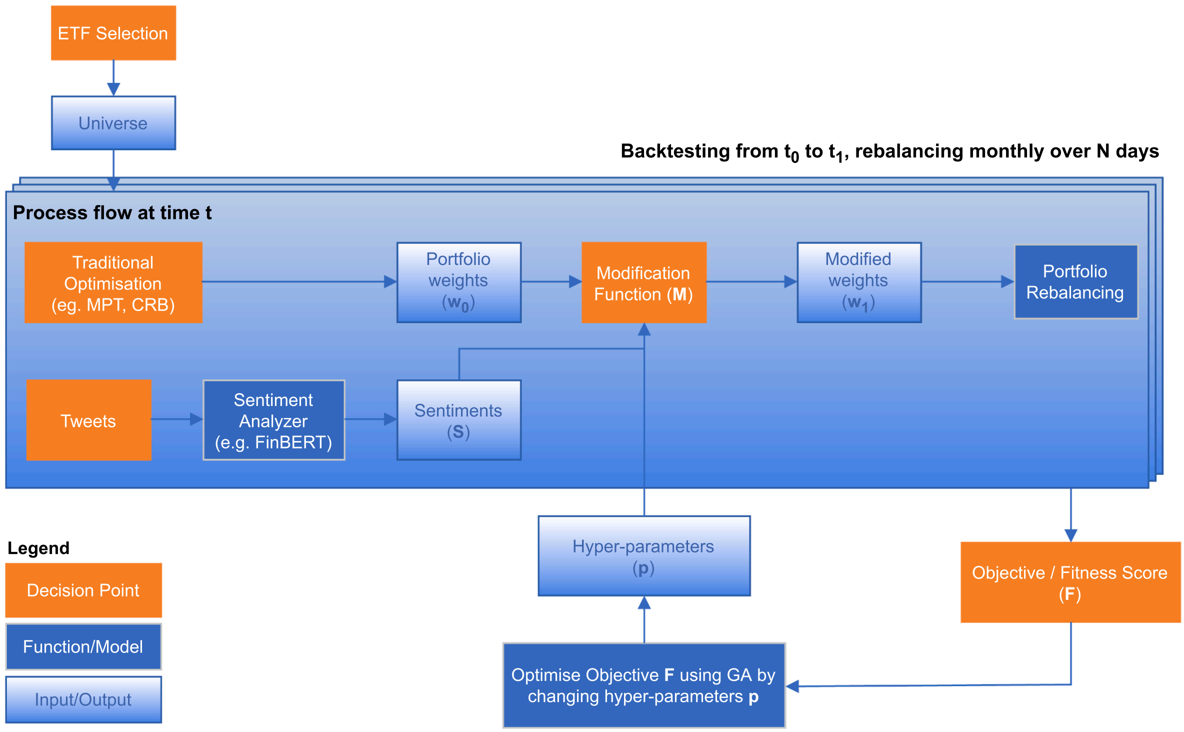


Fig. 1. Overview of proposed hybrid approach for portfolio optimisation.

portfolio across sectors (Table 1).

For asset-based diversification, we will consider the following asset classes as advocated by the All-Weather portfolio - equities, long-term bonds, intermediate-term bonds, gold, and commodities, which provide exposure to a variety of assets that perform differently across market environments. As there are many possible ETFs for each asset class, only one ETF will be taken as a representative of its asset class (Table 2).

3.2. Portfolio backtesting

Backtesting simulates a trading strategy to evaluate how effective the strategy might have been if it were traded historically. The results offer statistics to gauge the effectiveness of the strategy.

Daily time-series data are obtained for the selected ETFs from Yahoo Finance. Models are trained based on data from August 2018 to end December 2019, and backtested with out-of-sample data from January 2020 to April 2020. This allows the training and testing dataset to contain the market crash in December 2018 and the COVID-19 crash in January 2020, respectively (Fig. 2).

Commission for trades is assumed to be USD 0.005 per share, with minimum USD 1.00 per order. This follows the pricing structure of Interactive Brokers, the top stock broker in US.

Table 1
Symbols and sector types for SPDR ETFs (according to Select Sector SPDR).

Symbol	Select sector SPDR fund
XLC	Communication Services
XLY	Consumer Discretionary
XLP	Consumer Staples
XLE	Energy
XLF	Financials
XLV	Health Care
XLI	Industrials
XLB	Materials
XLRE	Real Estate
XLK	Technology
XLU	Utilities

Table 2
Representative symbols and asset types for the All-Weather portfolio, as suggested by PortfoliosLab.

Symbol	Asset Type
VTI	Equities
TLT	Long-term bonds
IEF	Intermediate-term bonds
GLD	Gold
DBC	Commodities

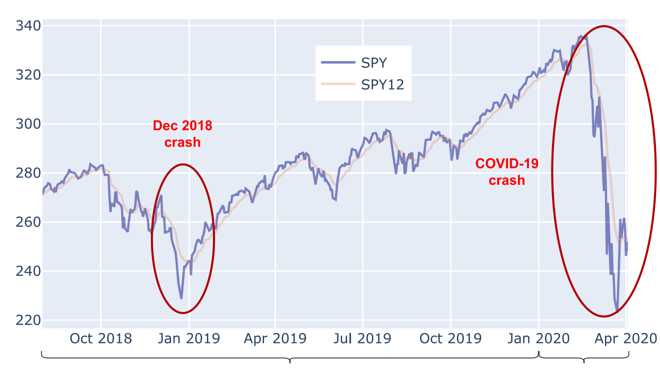


Fig. 2. SPDR S&P 500 (SPY) daily closing prices for the period of August 2018 to April 2020, as representative of US stock market.

3.3. Traditional optimisation and baselines

We consider the following traditional portfolio optimisation approaches, MPT and CRB, in the "Traditional Optimisation" block in Fig. 1. All rebalancing will be performed at the end of each month, and triggered only if target allocation differs from its current allocation by more than a threshold of 5%.

For CRB, the starting weights for the SPDR ETFs follow a naive equal-weighted allocation. For the All-Weather portfolio, the starting weights are the original recommended fixed-proportion allocation (30% stocks, 40% long-term bonds, 15% intermediate-term bonds, 7.5% gold, and 7.5% commodities) (Robbins, 2014).

For MPT, we use a simple mean historical returns for annualised expected returns as it is easily interpretable. Instead of using the sample covariance (S), Ledoit-Wolf shrinkage covariance matrix (Ledoit & Wolf, 2004) is used for the risk model, with an estimator given by Eq. 1, where F is the shrinkage target and δ is the shrinkage coefficient. Ledoit-Wolf shrinkage is used as it tends to pull the most extreme covariance coefficients towards more central values, thereby systematically reducing estimation error where it matters most (Ledoit & Wolf, 2004).

$$\delta F + (1-\delta)S, \quad 0 \leqslant \delta \leqslant 1 \tag{1}$$

All hybrid models will be baselined against a buy-and-hold strategy, as well as its corresponding traditional optimisation approach. Again, the traditional optimisation algorithms can be easily switched out with more complex variants in the overall flow if required.

3.4. Sentiments from Twitter data

Unlike most other work that focuses on company-specific tweets, we aim to capture the overall market sentiment, because we choose to decouple portfolio rebalancing from portfolio asset selection. This will enable us to have a fixed set of assets/holdings that will change only in terms of its percent allocation, so that our customers will be clear on what they are investing in (otherwise their holdings will likely fluctuate wildly with each rebalance). As such, we web scrape tweets from chosen specific accounts rather than filter based on mentions of companies. These accounts were chosen from various websites that recommended "Finance Twitter Accounts" to follow for financial insights, as well as Twitter accounts of news companies (such as CNBC).

For each of the chosen accounts, we web scrape the entire backtesting period, and convert each tweet into sentiments. We employ and contrast two methods: VADER (Hutto & Gilbert, 2014) and FinBERT (Araci, 2019). VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool specifically attuned to sentiments in social media, and uses a dictionary that maps lexical features to emotion intensities known as sentiment scores. The sentiment score of a text can be obtained by summing up the intensity of each word in the text. FinBERT is built by further training the BERT language model in the finance domain, using a large financial corpus and thereby fine-tuning it for financial sentiment classification as negative/positive words in finance can be different from general corpora (Araci, 2019).

Next, we perform post-processing, where tweets are converted to the correct timezone to correspond with the US trading market, and sentiments scores are aggregated to a per-day resolution. In order to reduce noise and outliers, we take the exponential moving average (EMA) over different spans.

Finally, we check for possible synchrony between stock data and tweet sentiments during volatile periods to shortlist suitable Twitter accounts that can represent the overall market sentiment by using Pearson coefficient with day-shifts.

3.5. Genetic Algorithm optimisation

We define a modification function (M) that takes in (i) EMA-weighted sentiment (S) and (ii) hyper-parameters (p), to transform traditional optimisation portfolio weight vector ($\overrightarrow{w_0}$) to modified weight vector ($\overrightarrow{w_1}$) at each rebalance interval (Fig. 1). For any given day, D, tweets and sentiments are used only up to $D-1$, to prevent look-ahead bias.

Applying Occam's razor, we start with the simplest possible modification functions and progressively add complexity. We calculate the Moving Average Convergence Divergence (MACD) of sentiments, S_{cd}, by subtracting the 26-period sentiment EMA from the 12-period EMA; and use it as a bullish signal if $S_{cd} > 0$ and as a bearish signal if $S_{cd} \leqslant 0$ (Eq. (2)).

$$S_{cd} = S_{12} - S_{26} \tag{2}$$

For CRB, some of the modification functions that were attempted are shown in Table 3. For MPT, an intuitive modification is to modify the expected volatility. For example, if the estimation is bullish, we should be able to take more risk such that our expected return along the efficient frontier will be higher. As such, instead of directly modifying the weights from MPT, we will first get the optimal volatility from MPT, modify it with a delta and then calculate the modified weights. This can be illustrated by the Markowitz Bullet in Fig. 3, using the All-Weather portfolio components. The Markowitz Bullet or efficient frontier shows the best possible expected level of return for its level of risk. A couple of volatility adjustments are attempted, and the final volatility adjustment, v_{adj}, is calculated by Eq. 3, where p_p is weight modifier when S_{cd} positive, p_n is weight modifier when S_{cd} negative.

$$v_{adj} = \begin{cases} v_{mpt} * (1 + S_{cd} * p_p), & S_{cd} > 0 \\ v_{mpt} * (1 + S_{cd} * p_n), & S_{cd} \leqslant 0 \end{cases} \tag{3}$$

Optimisation via GA is performed based on the simplest evolutionary algorithm presented by Bäck, Fogel, and Michalewicz, 2018, with a crossover rate of 0.5 and mutation rate of 0.2. For a given combination of "Decision Points" in Fig. 1, simulation is performed using a randomly generated set of weights over the backtesting period with rebalancing done at fixed intervals. At the end of the backtesting period, we derive a fitness score, F, which depends on our objective. We define three possible objectives – (i) maximise cumulative returns, (ii) maximise Sharpe Ratio, and (iii) minimise volatility, during the training period. These objectives correspond to different risk profiles for investors. An average investor might want to maximise his risk-adjusted returns by looking at the Sharpe Ratio, a conservative investor might be interested in a minimal volatility portfolio, and an aggressive investor might look only at the maximal cumulative returns. This fitness score is considered to be for an individual in GA, and the process is repeated for p population and n generations to form a single run. It is repeated for each combination of "Decision Points" with varying population size, number of generations and random seed.

4. Experimental results

We compare the various traditional portfolio optimisation techniques for our two different universes, and calculate cumulative returns, annual return, annual volatility, maximum drawdown, and Sharpe ratio. The Sharpe ratio of a portfolio is the ratio of mean risk-free excess return (over risk-free rate, R_f) to its standard deviation, σ, as given in Eq. (4). The maximum dropdown expresses the largest drop between a peak and valley in percentage, for the backtesting period.

Table 3
Sample modification functions for CRB in order of increasing complexity. S_n denotes EMA sentiment for n spans, $\overrightarrow{p}$ denotes the vector of hyper-parameters, with dimensions equivalent to $\overrightarrow{w}$.

Index	Equation
1	$\overrightarrow{w_1} = \overrightarrow{w_0} * S * \overrightarrow{p}$,
2	$\overrightarrow{w_1} = \overrightarrow{w_0} + S * \overrightarrow{p}$,
3	$\overrightarrow{w_1} = \overrightarrow{w_0} + \overrightarrow{p_p}$ if $S_{12} - S_{26} > 0$
	$\overrightarrow{w_1} = \overrightarrow{w_0} + \overrightarrow{p_n}$ if $S_{12} - S_{26} \leqslant 0$

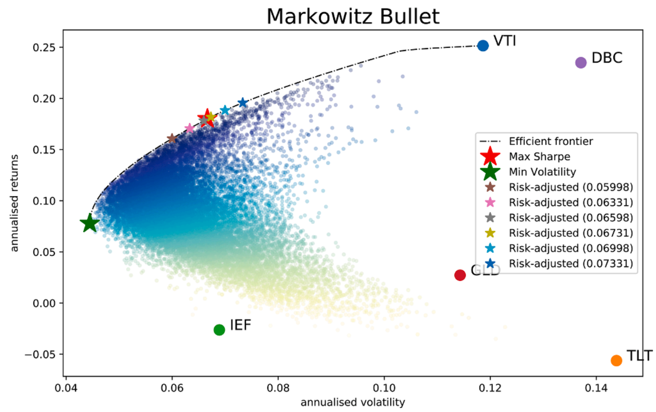
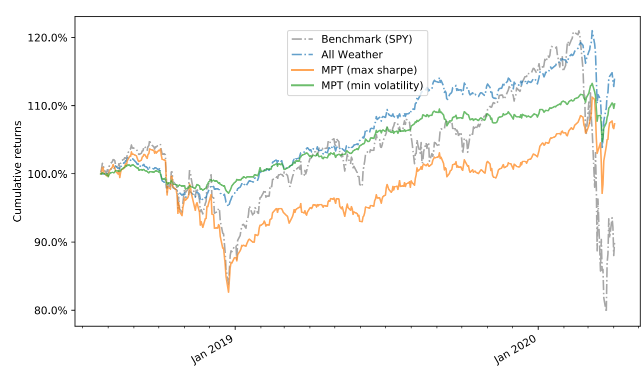


Fig. 3. Illustration of sentiment-adjusted MPT on Markowitz Bullet.

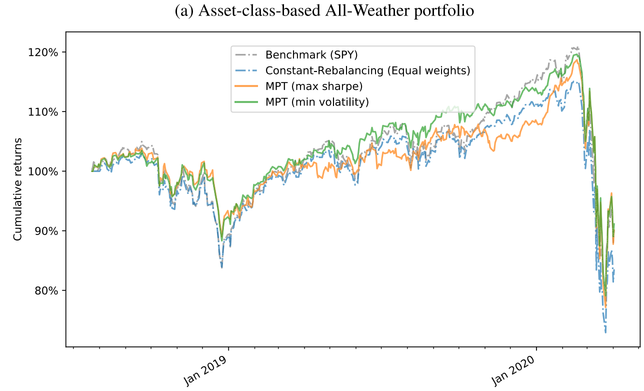
$$\text{Sharpe Ratio} = \frac{\overline{R_p} - R_f}{\sigma} \quad (4)$$

Using the asset-class-based portfolio, it can be seen that the traditional approaches are able to reduce maximum draw-down and annual volatility, and improve cumulative returns, as compared to buy-and-hold of the index, SPY (Fig. 4a, Table 4). In contrast, using a sector-based portfolio, all approaches moved similarly regardless of the optimisation technique (Fig. 4b), which suggested that it will not benefit much from optimisation. Hence, we decide to focus our efforts on only the asset-based portfolio for the subsequent experiments.

Tweets from various accounts are web scraped and converted to sentiments using VADER and FinBERT. In a random sampling of tweets, it is found that in most cases, FinBERT is able to score tweets more appropriately than VADER. Hence, while VADER is supposed to be suitable for social media, it is not good enough for finance related tweets. Some example tweets and sentiments are shown in Table 5.



(a) Asset-class-based All-Weather portfolio



(b) Sector-based SPDR sector ETFs portfolio

Fig. 4. Comparison of traditional approaches for the period of August 2018 to April 2020.

Table 4
Comparison of end performance for asset-class-based portfolio.

Metric	Baseline (SPY)	All-Weather	MPT (max Sharpe)	MPT (min volatility)
Cumulative returns	−11.26%	**14.09%**	7.35%	10.24%
Annual return	−6.78%	**8.07%**	4.27%	5.91%
Annual volatility	25.21%	8.64%	12.49%	**5.03%**
Max drawdown	−34.11%	−14.35%	−20.26%	**−7.56%**
Sharpe ratio	−0.15	0.94	0.4	**1.17**

We check for synchrony between daily stock data and the day-aggregated tweet sentiments during volatile periods, using Pearson correlation with time-shifting. In particular, it is found that tweet sentiments from CNBC (@cnbc) gave strong correlation with near-zero lag against SPY stock data. There is also some negative correlation between Donald Trump's (@realDonaldTrump) tweet sentiments and SPY (Fig. 5), which is similar to the findings by Brans and Scholtens (2020). Since @cnbc gives good correlation, the subsequent experiments are narrowed to using only their tweets.

As a proof of concept, we initially try a single-asset portfolio of only SPY and add a simple modification function, M, that takes CNBC sentiments into account, similar to Table 3, which we term "Sentimental SPY" portfolio. Even with a single asset, cumulative returns are improved compared to buy-and-hold SPY, as it is able to exit the market when sentiments are negative (Fig. 6a). This means that the portfolio holds a lot of uninvested cash during volatile period, which to the average investor may be overly conservative (Fig. 6b).

For a constant-rebalanced All-Weather portfolio, we use sentiments as trading signals to adjust allocation, which we term "Sentimental All-weather" (SAW) portfolio. Initially, we attempted to use the raw sentiment score in the modification function (index 1 and 2 of Table 3); but portfolio rebalancing was highly unstable, hence a simple delta was used based on the MACD sentiment (index 3).

An experiment is performed to gauge the effects of rebalancing at either "daily" or "weekly" intervals, and it is determined that optimisation results are similar. Therefore, for the rest of the simulations, we decide to use weekly rebalancing to reduce transaction costs.

We optimise parameters using GA over the training period of August 2018 to December 2019 for multiple runs, with weekly rebalancing, over three different objectives that represent different risk profiles (maximum Sharpe, maximum cumulative returns, minimum volatility) and compared them over the testing period of January 2020 to April 2020. From Fig. 7a, we can see that in all runs, SAW performed better than buy-and-hold SPY, and frequently out-performed original All-Weather portfolio, which lends credence to its feasibility as an improved portfolio algorithm.

With the same All-Weather portfolio, we perform MPT optimisation and modify the volatility as given in Eq. (3) to derive sentiment-adjusted weights, which we term "Sentimental MPT" (SMPT) portfolio, and likewise optimised parameters using GA. SMPT models tend to converge to certain performance that fared worse for the out-of-sample period (as seen from Fig. 7b), and only slightly better than MPT (Max Sharpe) for the entire period.

The performance metrics for out-of-sample data of (i) the best two SAW models for each objective, (ii) best SMPT model, and (iii) the baselines, are summarised in Table 6. In both SAW and SMPT, we considered only long positions. While short-selling is permitted in the US, and that could further increase portfolio performance, we wanted to show that even with a simple SAW and SMPT, considering only long positions, we could beat the benchmarks.

We conclude that SAW optimised for minimum volatility (SAW-MV, model MV200-5-244) is able to achieve improved annual volatility, maximum drawdown and slightly higher cumulative returns, as

Table 5
Some example tweets sentiment-analysed with VADER and FinBERT.

Tweet	VADER	FinBERT
JUST IN: U.S. weekly jobless claims surge to more than 6.6 million, vs. 3.1 million expected.	0	-0.0689
The World Health Organization has declared the coronavirus outbreak a global pandemic	0	-0.1286
Sometimes, the best thing to do is nothing. And according to Duke University behavioral economist @danariely, nothing is exactly what most investors should do during the coronavirus outbreak	0.6369	-0.0115
The Chinese government has deliberately underreported the total number of coronavirus cases and deaths in the country, the U.S. intelligence community reportedly told the White House	0.5267	-0.6744
European stocks just posted their worst quarter since 2002	-0.6249	-0.6510
Moody's cuts outlook on $6.6 trillion US corporate debt pile to 'negative'	-0.5719	-0.9144

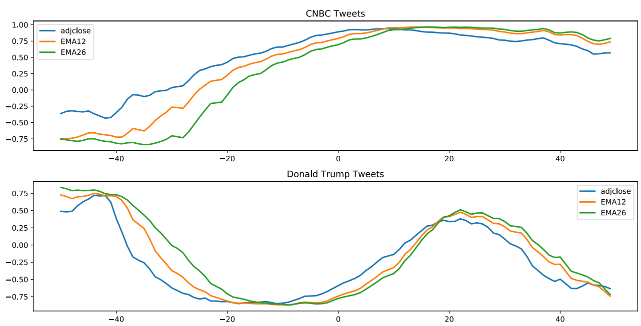
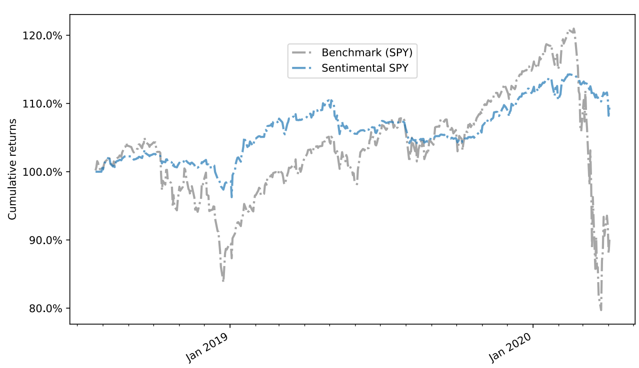
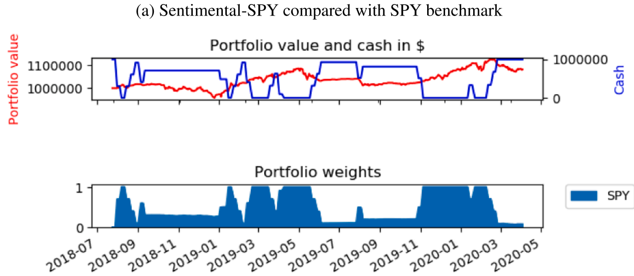


Fig. 5. Pearson coefficient against day-shifts for CNBC and Donald Trump's tweets.



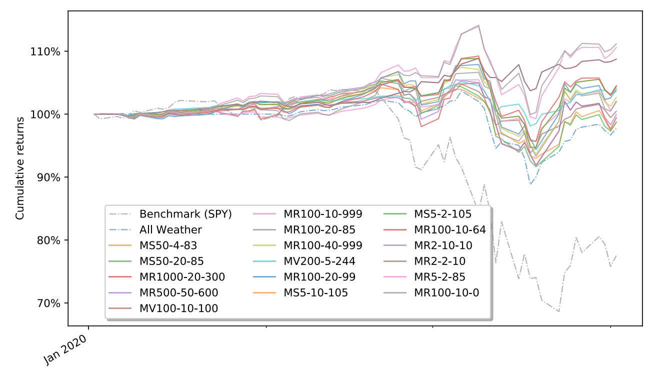
(a) Sentimental-SPY compared with SPY benchmark



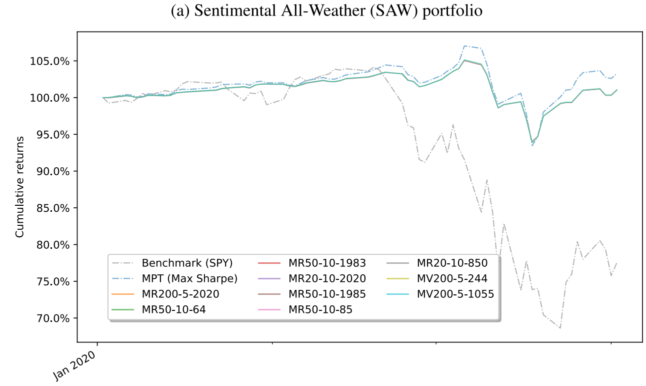
(b) Sentimental-SPY Portfolio allocation and value

Fig. 6. Sentimental-SPY Portfolio.

compared to the original All-Weather portfolio. In addition, SAW optimised for maximum cumulative returns (SAW-MR, model MR100-10-0) is able to achieve significantly higher cumulative returns and Sharpe ratio, but had slightly higher volatility.



(a) Sentimental All-Weather (SAW) portfolio



(b) Sentimental MPT (SMPT) portfolio

Fig. 7. Comparison of Sentiment-adjusted portfolios on out-of-sample data. Note that in (b), 11 series are plotted but all SMPT portfolios overlap.

5. Robo-advisor system

Django is chosen as the web development framework to build our robo-advisor as it is one of the most popular open-source free web development tools, has a short learning curve, and is catered for rapid development (Vincent, 2020). In order to accelerate development even further, Django project skeleton provided by Django Edge v2.2 is used. The prototype is deployed in Heroku.

We create a custom model class to store and retrieve each investor's details, such as asset transfers, gross asset value, and available cash in account. For portfolio details, we choose to use a Pickle object as the structure of the portfolios is complex.

Our robo-advisor has the following core functionalities – (i) Basic user management - sign up and log in, (ii) Summary page showing current account balance, earnings, portfolio asset value, etc., (iii) Add and withdraw funds (virtual funds, no actual interface with real money), (iv) Portfolio management – compare, buy and sell portfolios. The list of

Table 6
Comparison of performance for out-of-sample data for SAW and SMPT with baselines, on All-Weather portfolio.

	Baseline (SPY)	CRB	MPT (Max Sharpe)	MPT MR200-5-2020	MV100-10-100	**MV200-5-244**	MS5-10-105	MS5-2-105	MR100-20-99	**MR100-10-0**
Cumulative returns	−23.60%	−2.24%	3.29%	1.07%	1.34%	3.55%	5.00%	5.00%	6.14%	**11.57%**
Annual return	−65.36%	−8.53%	13.60%	4.29%	5.37%	14.73%	21.18%	21.18%	26.46%	**53.87%**
Annual volatility	55.05%	17.77%	18.01%	13.22%	13.09%	**9.40%**	18.93%	18.93%	23.90%	21.22%
Max drawdown	−34.11%	−14.35%	−12.70%	−10.17%	−10.52%	**−6.64%**	−12.94%	−12.94%	−12.51%	−12.45%
Daily value at risk	−7.30%	−2.27%	−2.21%	−1.65%	−1.63%	**−1.13%**	−2.30%	−2.30%	−2.91%	−2.49%
Sharpe ratio	−1.65	−0.41	0.8	0.38	0.46	1.51	1.11	1.11	1.1	**2.14**

Note: Objective functions are MV, MS and MR where MV - Minimum Volatility, MS - Maximum Sharpe Ratio, MR - Maximum (Cumulative) Returns. Only MPT MR200-5–2020 uses SMPT, the rest use SAW. Naming convention is [Objective][population size]-[number of generations]-[seed]. For example, MPT MR200-5–2020 would mean that we optimised based on MPT Maximum cumulative returns with population size of 200 and 5 generations with a seed of 2020.

portfolios is easily configurable from a spreadsheet, and includes SPDR sector ETFs, and All-Weather ETFs, SAW and SMPT portfolios. The overall user and system flows are shown in Fig. 8.

In order to categorise the portfolios based on risk, it is useful to use 99%-Value-at-Risk (VaR) shown in Eq. (5) which assumes returns are normally distributed. For example, a 99%-VaR of −1.5% will mean that there is a 99% chance of not losing more than 1.5% for the given time period. Alternatively, VaR can be calculated using the historical method. For simplicity, we classified risk into the following categories in Eq. (6) in the front-end.

$$99\% - \text{VaR} = \overline{R_p} - 2.326\sigma \tag{5}$$

$$\text{risk} = \begin{cases} \text{“low”}, & VaR > -10\% \\ \text{“medium”}, & VaR > -20\% \\ \text{“medium} - \text{high”}, & VaR > -30\% \\ \text{“high”}, & VaR \leqslant -30\% \end{cases} \tag{6}$$

6. Conclusion

Our proposed hybrid approach of using traditional portfolio techniques together with Twitter sentiments can improve portfolio performance, when optimised using GA for different objectives such as maximising cumulative returns or minimising volatility. The models proposed in this paper (SAW and SMPT) can be included as portfolios in a robo-advisor and easily deployed as an end-to-end system, using Fig. 8.

There are many simplification steps in this paper that may be expanded upon in future work. For example, in ETF selection, we can use more complicated algorithms to select the best performers for each asset class using rules or machine learning models. Sentiments can also be derived from a combination of different accounts, rather than just one, for more accurate daily sentiments. Our modification function can also be expanded upon to take in more inputs from other stocks or indices, such as the “fear” index (VIX), which has high readings when investors

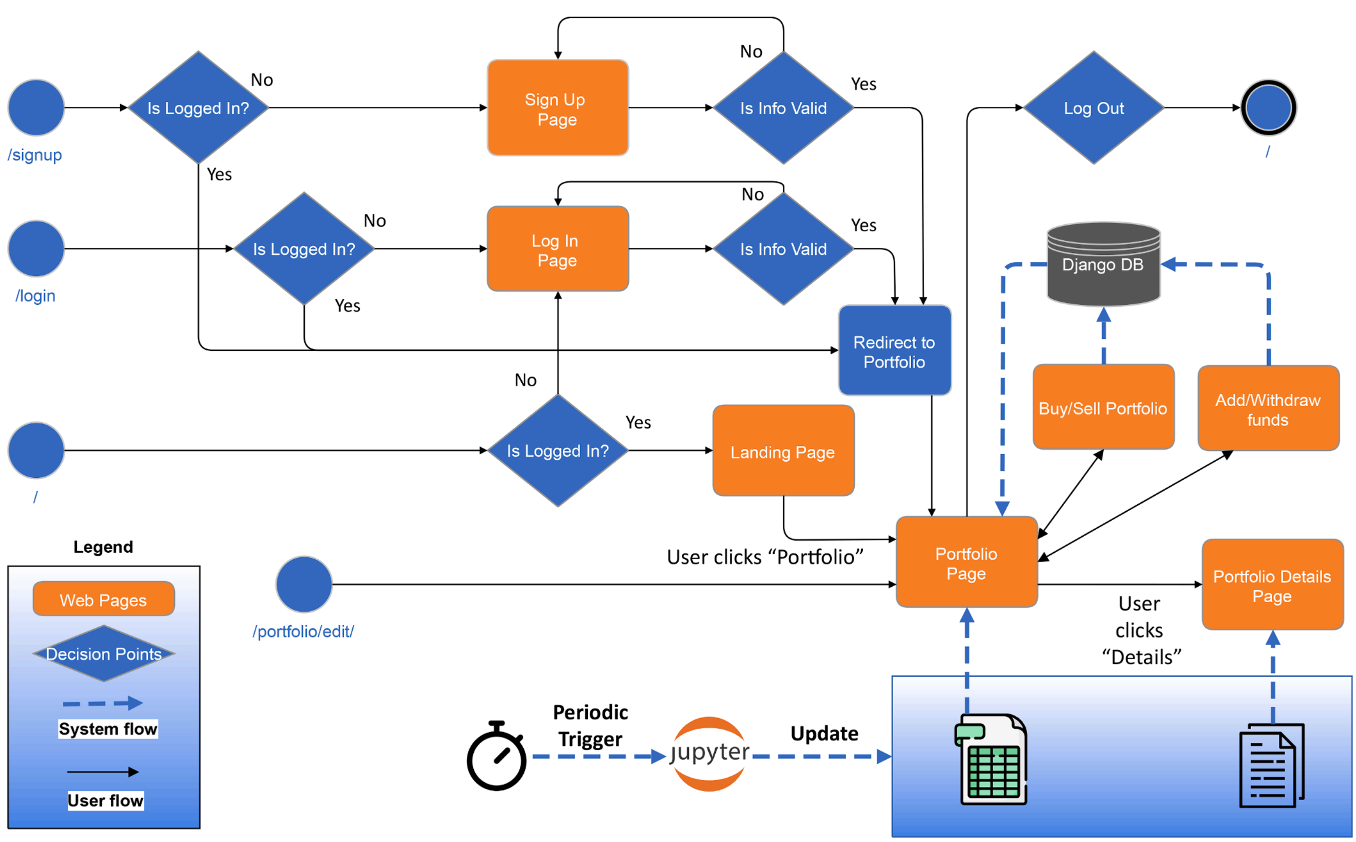


Fig. 8. User and system flow for Django robo-advisor app.

anticipate huge moves in the market. Other traditional algorithms such as HRP may also be added. In addition, while we only considered long positions, the algorithm and optimisation could be extended for short-selling as well, which would potentially increase performance.

CRediT authorship contribution statement

Edmund Kwong Wei Leow: Methodology, Data curation, Software, Formal analysis, Visualization, Validation, Writing - original draft. **Binh P. Nguyen:** Formal analysis, Visualization, Validation, Writing - review & editing. **Matthew Chin Heng Chua:** Conceptualization, Formal analysis, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to express their utmost gratitude and thanks to Leow Yan Gen Harris for sharing his financial knowledge through advisory and guidance in this work.

References

Aboussalah, A. M., & Lee, C.-G. (2020). Continuous control with stacked deep dynamic recurrent reinforcement learning for portfolio optimization. *Expert Systems with Applications, 140*. https://doi.org/10.1016/j.eswa.2019.112891, 112891.

Araci, D. (2019). *FinBERT: Financial sentiment analysis with pre-trained language models*.

Bäck, T., Fogel, D. B., & Michalewicz, Z. (2018). *Evolutionary computation 1: Basic algorithms and operators*. CRC Press.

Black, F., & Litterman, R. B. (1991). Asset allocation: Combining investor views with market equilibrium. *The Journal of Fixed Income, 1*(2), 7–18. https://doi.org/10.3905/jfi.1991.408013

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Brans, H., & Scholtens, B. (2020). Under his thumb the effect of president Donald Trump's Twitter messages on the US stock market. *PLoS One, 15*(3), 1–11. https://doi.org/10.1371/journal.pone.0229931

Brenner, L., & Meyll, T. (2020). Robo-advisors: A substitute for human financial advice? *Journal of Behavioral and Experimental Finance, 25*. https://doi.org/10.1016/j.jbef.2020.100275, 100275.

Chen, K., Zhou, Y., & Dai, F. (2018). A LSTM-based method for stock returns prediction: A case study of China stock market. In *Proceedings of the IEEE international conference on big data (Big Data 2015)* (pp. 2823–2824). IEEE. doi: 10.1109/BigData.2015.7364089.

Deng, Y., Bao, F., Kong, Y., Ren, Z., & Dai, Q. (2016). Deep direct reinforcement learning for financial signal representation and trading. *IEEE Transactions on Neural Networks and Learning Systems, 28*, 1–12. https://doi.org/10.1109/TNNLS.2016.2522401

Fu, X., Du, J., Guo, Y., Liu, M., Dong, T., & Duan, X. (2018). *A machine learning framework for stock selection*.

Gomes, C. P., & Selman, B. (2001). Algorithm portfolios. *Artificial Intelligence, 126*(1), 43–62. https://doi.org/10.1016/S0004-3702(00)00081-3

Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the 8th international AAAI conference on weblogs and social media (ICWSM 2014)* (pp. 216–225). AAAI/.

Idzorek, T. (2007). 2 - A step-by-step guide to the Black-Litterman model: Incorporating user-specified confidence levels. In S. Satchell (Ed.), *Forecasting expected returns in the financial markets* (pp. 17–38). Elsevier. ISBN 978-0-7506-8321-0, 2007, doi: 10.1016/B978-075068321-0.50003-0.

Jiang, Z., Xu, D., & Liang, J. (2017). *A deep reinforcement learning framework for the financial portfolio management problem*.

Lam, J. W. (2016). *Robo-advisors: A portfolio management perspective*. Tech. Rep., Yale College.

Lasek, M., & Lasek, J. (2015). Are stock markets driven more by sentiments than efficiency? *Journal of Engineering, Project, and Production Management, 6*. doi: 10.32738/JEPPM.201601.0005.

Ledoit, O., & Wolf, M. (2001). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance, 10*, 603–621. https://doi.org/10.1016/S0927-5398(03)00007-0

Ledoit, O., & Wolf, M. (2004). Honey, I shrunk the sample covariance matrixi shrunk the sample covariance matrix. *The Journal of Portfolio Management, 30*(4), 110–119. https://doi.org/10.3905/jpm.2004.110

Lekovic, M. (2018). Investment diversification as a strategy for reducing investment risk. *Economic Horizons, 20*(2), 169–184. https://doi.org/10.5937/ekonhor1802173L

Liang, Z., Chen, H., Zhu, J., Jiang, K., & Li, Y. (2018). *Adversarial deep reinforcement learning in portfolio management*.

Li, H., Dagli, C., & Enke, D. (2007). Short-term stock market timing prediction under reinforcement learning schemes. In *Proceedings of the IEEE international symposium on approximate dynamic programming and reinforcement learning (ADPRL 2007)* (pp. 233–240). IEEE. doi: 10.1109/ADPRL.2007.368193.

Lopez de Prado, M. (2016). Building diversified portfolios that outperform out of sample. *Journal of Portfolio Management, 42*(4), 59–69. https://doi.org/10.3905/jpm.2016.42.4.059

Markowitz, H. M. (2013). *Portfolio selection: Efficient diversification of investments*, Yale University Press, ISBN 0300013728, 2013.

Pagolu, V. S., Challa, K. N. R., Panda, G., & Majhi, B. (2016). *Sentiment analysis of twitter data for predicting stock market movements*.

Peterson, R. L. (2016). *Trading on sentiment: The power of minds over markets*. John Wiley & Sons.

Phoon, K., & Koh, F. (2018). Robo-Advisors and Wealth Management. *The Journal of Alternative Investments, 20*(3), 79–94. https://doi.org/10.3905/jai.2018.20.3.079

Poterba, J., & Shoven, J. B. (2002). *Exchange traded funds: A new investment option for taxable investors*. NBER Working Papers 8781, National Bureau of Economic Research Inc. URL: https://EconPapers.repec.org/RePEc:nbr:nberwo:8781.

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science, 48*(1), 137–141. https://doi.org/10.1007/s11747-019-00710-5

Ribeiro, F. N., Araújo, M., Gonçalves, P., Benevenuto, F., Gonçalves, M. A. (2015). SentiBench – a benchmark comparison of state-of-the-practice sentiment analysis methods.

Robbins, A. (2014). *Money: Master the Game: 7 simple steps to financial freedom*. Simon & Schuster UK, ISBN 978-1-47114-335-9.

Shen, S., Jiang, H., & Zhang, T. (2012). Stock market forecasting using machine learning algorithms. Tech. Rep., Stanford University.

Sul, H. K., Dennis, A. R., & Yuan, L. I. (2017). Trading on Twitter: Using social media sentiment to predict stock returns. *Decision Sciences, 48*(3), 454–488. https://doi.org/10.1111/deci.12229

Sun, C., Huang, L., & Qiu, X. (2019). Utilizing BERT for aspect-based sentiment analysis via. *Constructing Auxiliary Sentence*.

Vincent, W. S. (2020). Django for beginners: Build websites with Python and Django. *Still River Press*.

Xu, P., Chen, A., & Tsui, P. W. E. (2018). *Black-Litterman Model, Tech. Rep.* Berkeley: University of California.

Xu, H., Liu, B., Shu, L., Yu, P. S. (2019). *BERT post-training for review reading comprehension and aspect-based sentiment analysis*.

Yang, J., Li, Y., Chen, X., Cao, J., & Jiang, K. (2019). *Deep learning for stock selection based on high frequency price-volume data*.

Zhang, W. & Skiena, S. (2010). Trading strategies to exploit blog and news sentiment. In: *Proceedings of the fourth international AAAI conference on weblogs and social media (ICWSM 2010)* (pp. 375–378). AAAI.