Adv Syst Sci Appl 2023; **01**; 69-90
Published online at https://ijassa.ipu.ru.

A Novel Method for Predicting Technology Trends Based on Processing Multiple Data Sources

Nguyen Thanh Viet[1,2**], Alla Kravets[1,3]

[1)] *Volgograd State Technical University, Volgograd, Russia*

[2)] *Pham Van Dong University, Quang Ngai, Vietnam*

[3)] *Dubna State University, Dubna, Moscow region, Russia*

Abstract: In order to gain competing capability in conditions of quickly scientific changes, it is crucial to track the evolution of existing technologies and to explore promising and emerging technologies. Moreover, numerous previous studies showed that sudden changes in R&D and patents are actually correlated with great variations in the market profit of firms. For this reason, if stock prices of an enterprise keep uptrend, then the technologies developed by considered one will be likely to become promising innovations in future. In this paper, we proposed a method to predict technology trends based on processing multiple data sources by mining Web news, forecasting stock price trends of high-tech companies, and patent clustering analysis. Different from other studies, our proposed method promotes an idea of predicting technology trends by forecasting stock price trend using univariate and multivariate data preparation approaches, with the utilization of Bayesian optimization for exploring best hyperparameters of machine/deep learning models, also a new method for patent analysis. Besides, a program system was created for analyzing word burst detection, predicting trend of stock prices, and analyzing patent applications. After collecting patents of Samsung Electronics Co Ltd, as a case study, clustering analysis is implemented on extracted noun phrases to explore technology trends developed by the company. These technology trends have recently been confirmed by domain experts in their corresponding published articles. The obtained forecast precision is about 93.8%, which proves that the proposed method gains positive reliability.

Keywords: technology forecast, Web news, stock price, Samsung, Bayesian optimization, machine learning, deep learning, patent analysis, keyword extraction, clustering analysis.

1. RELATIONSHIP BETWEEN INNOVATION AND STOCK PRICES

The development of emerging technologies not only has affected current industrial production but also has generated promising manufacturing opportunities that impact significantly on social and economic factors. Exploring upcoming renovation tendencies of emerging technologies prematurely is essential for governments, research and development institutes, and industrial companies in managing strategies to achieve dominant advantages in business competitiveness. Moreover, numerous policy makers and organization managers are always conscious of the importance of forecasting the upcoming innovation tendencies of emerging technologies for their stable developments [18], [48].

On the other hand, over the industrial revolutions with speedy transformations in societies, capital assets are invested in positions as long as there are promising economic benefits. Technologies are forming various domains of the modern life. They enhance manufacture speed, and facilitate numerous aspects of life more effective. A domain that is being revolutionized by the progresses of technologies is the stock and financial market. There are different ways of how technologies have affected and making the present shape of the financial markets, and the future state as well.

[*] Corresponging author: vietqn1987@gmail.com

More specifically, in paper [54] 73 considered companies published declarations of blockchain application in period 2016-2019 were chosen as research objects, while the short-term event analysis method was adopted to examine quantitatively the influence of blockchain technology applications on market achievement of these companies. The results revealed that after their declarations of blockchain implementation in business, considered companies attained more obvious unusual profits, which showed that the financial market responded positively to the blockchain application. Concurrently, the event analysis method helped identify the effect of emerging technologies application on the company's stock values. For instance, a few authors examined empirically the effect of 3D printing application and cloud computing implementation on the stock revenue amount [19].

If stock prices indicate belief about future profits, then we can specify a correlation between innovation, which in case of successful implementation may impact positively on growth and profit of firms, and stock prices. In fact, there is a close relation between the volatility of stock prices and market shares when technology is still unpredictable through premature industrial evolution. In other words, the volatility of stock prices that reflects the uncertainty and risk attaches to technological innovation [25].

Furthermore, another group of studies that combines innovation to stock price refers to the relationship between patents and market values. The argument is that if patent analysis statistics hold facts about perspectives of technology opportunities, then they will be related to present shifts in market values, which are affected by the beliefs about future growing. Some authors investigate the correlation between the amount of granted patent applications of a firm, an estimation of the firm's expenditure in invented activities (R&D), and a measure of its inventive returns, i.e., stock market price. Their results showed that sudden shifts in R&D and patents are really correlated with substantial shifts in the firm market value.

For instance, Vitt et al. [49] investigated the relation between the patenting efforts of high-technology companies and the movement of their stock price. Also, the authors demonstrated correlation between the amount of patent applications along with the variety of their respective patent categories, and the market-adjusted stock revenues. Besides, the relation between those patent activity measures and volatility of the market-adjusted stock revenues, were confirmed completely. Herein, utilizing the moving window method, the scholars presented numerous lagged terms of patenting activity features in order to generate fitted models. The final results confirmed the essential effect of patenting efforts on stock price change direction and its significant volatility and drift statistics.

Moreover, Senarathne and Wei [35] stated that patent applications, backward patent citations, and issued patents affect trading volume to some extent, proposing that each of these parameters could be considered as absorptive capacity parameter when new information flow is correlated with economic innovation.

Likewise, experimental confirmation with stock price revealed that firms succeeded in investing in comprehensive innovation attained higher stock incomes. The paper [12] proposed a model that made clear the correlation between extent of firm innovation (or absorptive capacity) and stock incomes, the change direction of which presents prospect of firm's growth and profitability. The final results showed that the premise of ambiguity-aversion is absolutely significant in specifying higher incomes with comprehensive innovation presence, and the particular specification of desirable benefit formed the magnitude of the returns.

Lee [20] proved that the stock price of a company utilizing a certain technology would fluctuate along with the life cycle of the considered technology. Concretely, the author selected companies that typically supply augmented reality and were registered in Korea KOSDAQ market. Then these companies were grouped depending on the detailed technologies that make augmented reality. The author adopted the event study method to estimate the stock incomes in comparison with a predefined benchmark. The result stated that in the "Peak of Inflated Expectations" phase, the portfolios of all listed company utilizing augmented reality mostly presented greater incomes than the benchmark.

Additionally, the paper [11] investigated the correlation between patent activity of the top corporate international R&D investors and their market values. The research exploited a collection of over 1,250 public multinational corporations and their intellectual property rights – trademarks and patents – registered between 2005 and 2012. Finally, the study allows us to differentiate the effects of increasing market value from additional intellectual property rights.

In [5] the experimental evidence recommended that stock incomes in the emerging technology conditions showed high return volatility. The article analyzed the movement, time series attributes of the relationship between daily log returns and the extent of volatility disseminations from the emergent technologies conditions to the Spanish market portfolio, the Spanish banking sector, and the financial market of the EU countries.

On the other hand, recent studies [46], [47] present the current progress and existing problems of analyzing, forecasting technological trends and state that Web news is an essential data source for investigating overall public recognition of emergent technologies. Web news, which points out the latest changes of everyday life and publishes analyses throughout the world, has the attribute of real time update and becomes significant for exploring the upcoming technological development tendencies. After utilizing burst detection algorithm and clustering of burst terms, we obtained some prominent enterprises and main technology trends that occurred from January 2016 to January 2020 [46].

Subsequently, based on the above discussions, the authors putted forward a hypothesis: if we are able to predict the uptrend stock price of some high-tech enterprises, then the technologies developed by the enterprise will be most likely to become promising innovations in the future. Hence, we propose a method to predict technology trends by aggregating following steps: mining Web news to extract significant high-technology enterprises, forecasting stock price trends of selected ones, and patent analysis of considered enterprises with uptrend stock price. The general process of the method is depicted in figure 1.

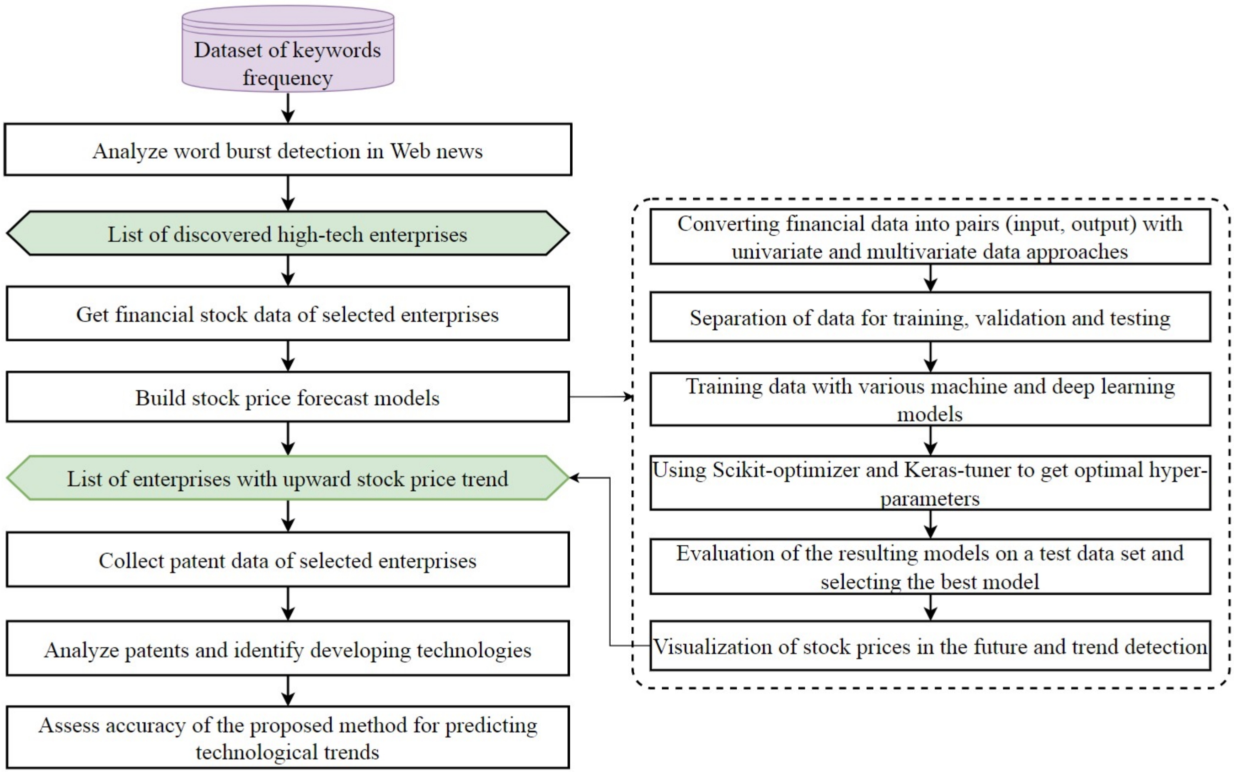


Figure 1. Overall workflow of proposed method for technology prediction

2. REVIEWING STUDIES ON STOCK PRICE PREDICTION

Stocks (shares) are financial product distinguished by flexible trading, great income and much uncertainty, which are preferred by a lot of investors. They can gain large revenues by correctly forecasting stock price movements. Stock price forecast is not new but a challenging and complicated problem, which is paid much attention from both data scientists and economists. In particular, if the goal is predicting the exact price values, then we consider it as regression problem, while if the goal is predicting the price direction, i.e., moving down or up, then we consider it as classification problem. Most of existing studies are investigating the daily forecast (5 or 10 days ahead) and only a few of them are investigating the forecast during one day, e.g., 15-min or hourly forecast. With the aim of creating powerful forecast model, both linear and machine learning methods have been adopted since the past few decades.

The primary attempts included fundamental analysis and technical analysis. The fundamental analysis estimates stock price depending on its native financial value, while the technical analysis considers mainly trends and charts. Besides, some works utilized technical indicators as additional input features for machine and deep learning models. Thereafter linear models were presented for solving stock price forecast, which mainly comprise ARIMA (autoregressive integrated moving average) [13] and GARCH (generalized autoregressive conditional heteroskedasticity) [6]. Nowadays, machine learning models are also utilized for predicting stock price with the success, e.g., Linear regression, Gradient boosting, Support vector machine, etc. [53], [31].

Recently various deep learning models have been recommended for solving this problem with great achievements [56], [10]. In terms of powerful capability processing big data and recognizing nonlinear correlations between input features and prediction output, for many years deep learning models have been providing better results than both linear and machine learning models on the problem of stock price forecast [15]. Some recent results suggest that the effectiveness of LSTM (long short-term memory) is better than non-time series and traditional machine learning models.

For instance, in [29] Nguyen and Yoon proposed a new effective framework, called deep transfer with related stock information, which made use of transfer learning and deep neural network. To do so, a base model adopting LSTM was trained with large dataset obtained from numerous stocks to initialize optimal training parameters. Then, this base model is fine-tuned by utilizing a small dataset from target stock along with numerous input features (derived from the correlation between stocks) to increase accuracy.

Moreover, Vo et al. [50] presented a Deep responsible investment portfolio model and compared the accuracy of LSTM, Bidirectional LSTM (BiLSTM) and gated recurrent unit (GRU) on the stock price forecast problem. Authors stated that the BiLSTM model reading the data backward once more would help enhance forecast performance, especially in predicting sequential data like financial time series.

Besides, Giang et al. [41] putted forward a method adopting LSTM along with deep concern in historical data of stocks and their nearest neighbors by similarity. Empirical results on 4 active stocks in the United State and 3 stocks in the Vietnamese stock market confirmed that the proposed model outperformed other methods in comparison (linear regression, random forest, Convolutional neural network – CNN, and vanilla LSTM).

Chen et al. [8] proposed a trend predictive model (TPM) relying on an encoder-decoder framework that forecasts the stock price movement tendency correctly. This model included two stages: in the first stage, a dual feature extraction method basing on numerous time periods was adopted to attain important information from the stock market. Differing from existing methods which extract significant features only at certain time points, the authors applied the CNN and PLR method to obtain both spatial short-term and temporal long-term features. Then, in the second stage of the TPM, an encoder-decoder framework basing on dual attention mechanism was utilized to pick out appropriate dual features and forecast the stock price movement tendency.

In study [22], by employing an eight-trigram feature engineering scheme of the candlestick samples between days, Lin et al. constructed a new ensemble machine learning framework for daily stock movement forecast, integrating normal candlestick charting with the most recent artificial intelligence (AI) methods. More specifically, various machine/deep learning methods were also adopted to forecast the closing stock price movement. Accordingly, this proposed framework helps select proper machine learning forecast method for all patterns basing on the trained outputs.

Additionally, the extensive surveys of Jiang [16] and Li [21] give latest reviews of recent works on deep learning models for stock price prediction. Authors not only categorize and introduce numerous data sources, many neural network models, and frequent used metrics for evaluation, but also the reproducibility and implementation.

In summary, econometric and machine learning models were the most frequently adopted methods in stock price forecast problem. Nevertheless, econometric models do not provide effective accuracy for nonlinear time series problems, while common machine learning models disregard much hidden knowledge evolving through times. Deep learning models can process effectively multi-period and time series data [41]. Concurrently, combining various deep learning models together has recently provided better performance than single model and is becoming the dominant approach for solving stock price forecast problems [16], [55].

3. PROPOSED METHOD AND EMPIRICAL RESULTS OF STOCK PRICE PREDICTION

In the course of the study, a program system was developed for analyzing the word burst detection in Web news, predicting the trend of stock prices, and patent analysis. The information-functional diagram of the system is shown in figure 2. This program system consists of 3 modules:

– word frequency analysis module;
– stock price forecast module;
– patent analysis module.

The program code is written in Python, with support of popular machine/deep learning libraries such as Scikit-learn, Keras, Pandas, Matplotlib, Seaborn, Numpy, spaCy, NLTK, Sentence Transformer, etc. All empirical runs were carried out in the interactive Google Colab environment.

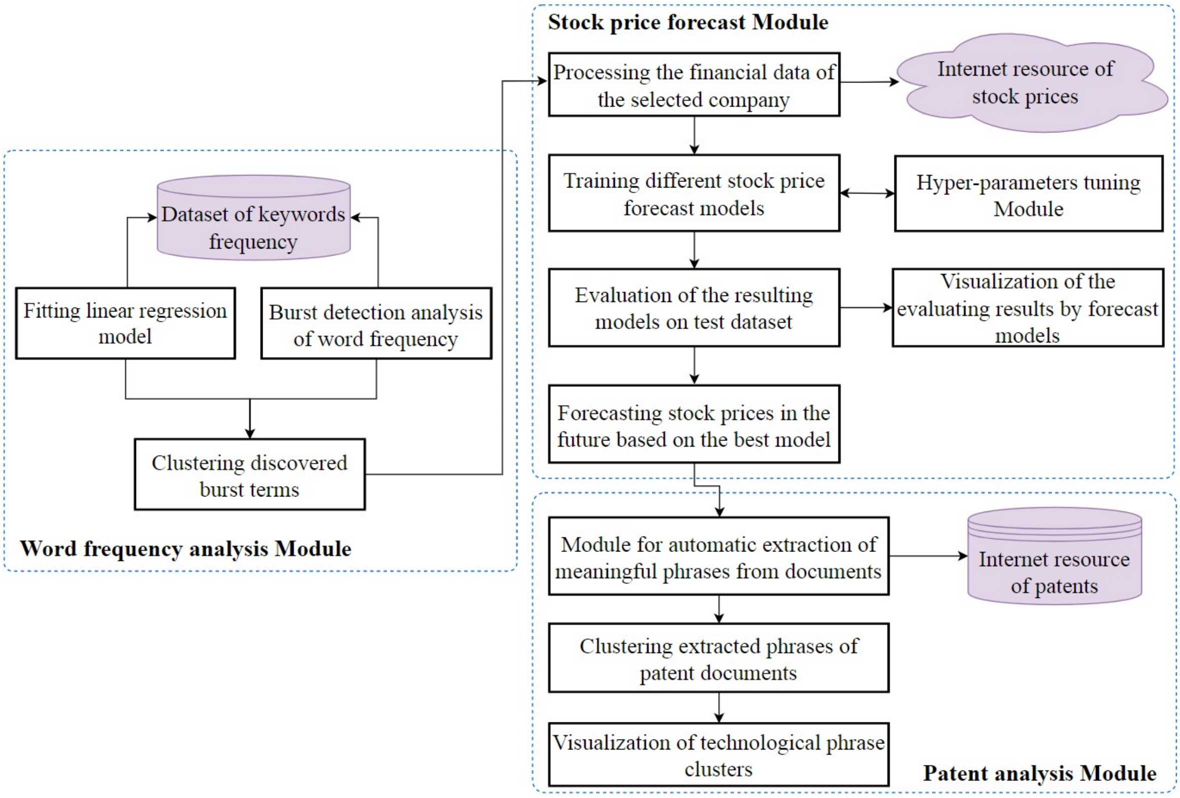


Figure 2. Information-functional diagram of the program system

Wherein the module for analyzing the words frequency in Web news were described specifically in our previous article [46]. This paper describes the functions of the stock price forecast module and patent analysis module. In this manner, our recent studies [46], [47] state that Web news is a crucial data source for investigating social recognition of emergent technologies. Also, after utilizing burst detection algorithm on Web news data and clustering of burst terms, we obtained some main technology trends that occurred from January 2016 to January 2020. Moreover, some of high-technology enterprises appeared among the explored burst terms is: Samsung, Huawei, Intel, Uber, Tesla, Apple, and Facebook. As a case study, the proposed method is applied to predict the stock price trend of Samsung Electronics Co Ltd.

3.1. Raw data

Available stock price data of Samsung (stock code 005930.KS) is downloaded from Yahoo Finance [33] in the period from January, 2000 to December, 2021 for further prediction model. The objective of this section is predicting the stock price overall trend (upward or downward) of Samsung in the year 2022 by adopting numerous machine learning and deep learning models. The raw data consists of all financial attributes that exist in a stock market, e.g., open/ close/ adjusted close/ high/ low prices, and trading volume. The downloaded dataset of Samsung stock price is demonstrated in figure 3.

	Date	Open	High	Low	Close	Adj Close	Volume
0	2000-01-04	6000.0	6110.0	5660.0	6110.0	4675.782715	74195000
1	2000-01-05	5800.0	6060.0	5520.0	5580.0	4270.191406	74680000
2	2000-01-06	5750.0	5780.0	5580.0	5620.0	4300.801270	54390000
3	2000-01-07	5560.0	5670.0	5360.0	5540.0	4239.581543	40305000
4	2000-01-10	5600.0	5770.0	5580.0	5770.0	4415.592773	46880000
...	...	...	...	...	...	...	...
5523	2021-12-24	80200.0	80800.0	80200.0	80500.0	80138.101563	12086380
5524	2021-12-27	80600.0	80600.0	79800.0	80200.0	79839.453125	10783368
5525	2021-12-28	80200.0	80400.0	79700.0	80300.0	79939.000000	18226325
5526	2021-12-29	80200.0	80200.0	78500.0	78800.0	78800.000000	19794795
5527	2021-12-30	78900.0	79500.0	78100.0	78300.0	78300.000000	14236700

Figure 3. Samsung stock price (currency in Korean Won) dataset from 2000 to 2021

3.2. Data preprocessing and preparation

There are 2 common approaches to prepare data for predicting stock price: univariate and multivariate data. For univariate input data we utilize autocorrelation to calculate how past stock prices affect their upcoming prices. Correlogram (ACF plot) is used to help determine whether our stock price series is presenting autocorrelation in overall, also at which point we can capture the pattern of values in the series for better future prediction. By observing ACF plot (figure 4) derived from Statsmodels Python package, we set lag window as 50, which means basing on Adjusted close price data of previous 50 days, the model needs to predict the adjusted close price in the 51th day (the next day). In univariate data preparation approach only the Adjusted close price column is used, while all remaining columns (open/ close/ high/ low prices, trading volume) are eliminated. This feature column is utilized as both input features (e.g., the historical prices in a look-back window) and prediction target (e.g., the adjusted close price of the next day).

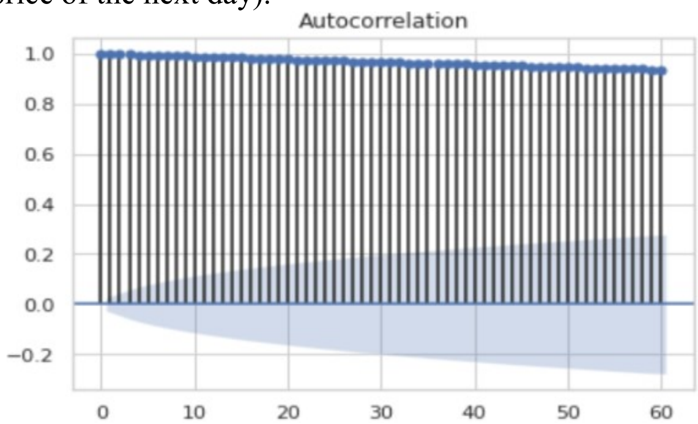


Figure 4. Correlogram (max 60 lags) of Samsung stock price

With respect to multivariate data approach, horizon time window is defined as the future time length needed for prediction by the model. Most of previous works consider long-term forecast horizon time windows, e.g., 7 days or 10 days [16], while fewer studies examine shorter horizon time windows such as 15 minutes or 1 day. However in this study, for multivariate data approach, we consider the horizon time window as 250, because it is the average number of working days within a year (the stock market is closed at weekends and on holidays). To do so, input data is financial data within a day (open/ adjusted close/ high/ low prices, trading volume), the developed model needs to predict the Adjusted close price in the next 250th day. The difference between 2 data preparation approaches is illustrated in figure 5 below.

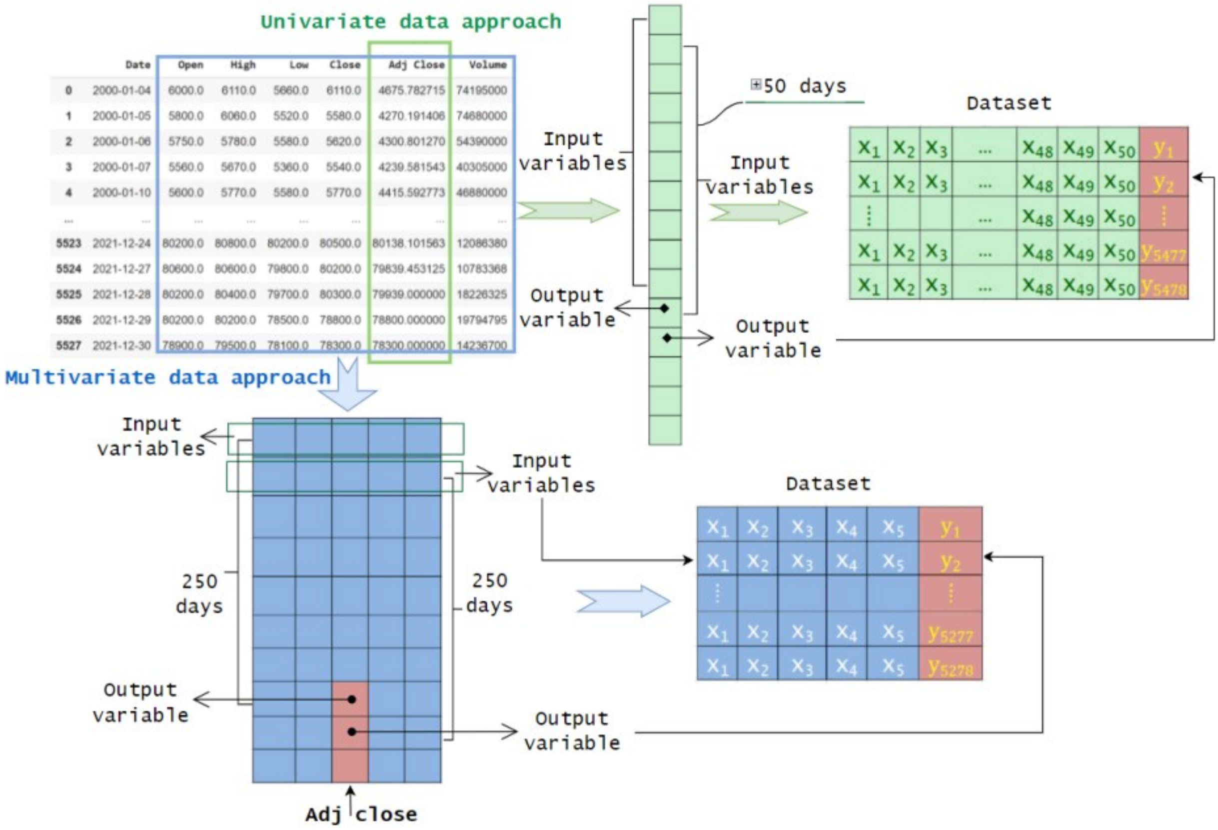


Figure 5. Univariate and multivariate data preparation methods

Additionally the resulting dataset is divided into a training set with 90% of the observations that are used to train the model (herein, the training set also contains 10% of the data for validation), and the remaining 10% for testing the model.

On the other hand, relating to input features of various scale values, feature normalization preprocess is utilized to assure that many machine/deep learning models can be trained effectively with performance and speed improvement. Hereby feature normalization is namely the process of rescaling the input feature values between 0 and 1.

3.3. Adopted machine/deep learning models for prediction

Furthermore, generally machine/deep learning models include two types of parameters (not for learning in training process): model design variables and hyperparameters. The model design variables consist of type of optimizer, loss function, and activation function, while hyperparameters are used to control the learning process. They may include number of layers, number of neurons, kernel size, dropout rate, and learning rate, etc. Thus, in order to select effective hyperparameters of the models, the Bayesian optimization (BO) is adopted with the utilization of Scikit-optimize [34] and Keras-tuner [30] libraries. The purpose of Bayesian optimization [52] is to find and adjust the hyperparameters of a given model for effective performance on the validation dataset. This optimization method is popularly considered to be superior to grid or random searches, because BO utilizes Gaussian processes to estimate an efficient probability distribution model of objective function. Hence, BO method can help identify better hyperparameter settings than a grid or random search within larger variable search space and fewer iteration cycles.

Besides, in this paper the authors employ popular machine learning models for regression, e.g., Extreme gradient boosting (XGB), Gradient boosting regression (GBR), Support vector regression (SVR), and k-nearest neighbor (KNN). On the other hand, these following models also provide high evaluation accuracy.

1. Linear regression (with Elastic Net regularization [37]). Linear regression is a classical linear model that fit the relation between the forecast target and the input variables, in which model parameters can be identified by the constrained least squares. While fitting linear regression model, the Elastic Net is a regularized regression method that combines linearly the L1 and L2 penalties of the Lasso and Ridge methods. In this study the search space of model hyperparameters is following: alpha $\in$ {0.01, 0.02, …, 0.99, 1}, l1_ratio is a random real number in range [0, 1].

2. Feedforward neural network (FFNN). It is the most basic kind of ANN (artificial neural network) where connections between the neuron nodes do not create cycle. In this network, the information flow passes only in one forward direction from the input nodes, through the hidden nodes and to the output nodes [40]. We define the search space of hyperparameters as following: number of layers $\in$ {1, 2, 3}, number of neurons $\in$ {100, 150, 200, 250, 300} for univariate data approach and $\in$ {200, 250, 300, 350, 400} for multivariate data, dropout rate $\in$ {0.01, 0.1}.

3. Convolutional neural network (CNN). The model is mainly regarded for processing 2D pictures, where each group of neurons named filter executes convolution operations to various subsequent regions of the input picture. Because the neurons share the same weight values, the amount of trained parameters is reduced compared to the densely connected FFNN. Afterwards max or average pooling operations can be employed for multiple times to reduce the original dimensions, until the final output is concatenated to a dense layer. By applying the convolutional and pooling operations to a single dimension, the model 1D CNN [9] is often used for time series prediction and classification. The search space of 1D CNN hyperparameters is following: number of filters $\in$ {150, 200, 250, 300, 350, 400} (univariate data) and $\in$ {300, 350, …, 750, 800} (multivariate data); kernel size $\in$ {2, 3, 4, 5} (univariate data) and $\in$ {2, 3, 4} (multivariate data); dropout rate $\in$ {0.01, 0.1}.

4. Recurrent neural network (RNN). Differing from FFNN, RNN is an ANN wherein links between the nodes generate cycles along a time sequence, which helps express time changing attributes. Nevertheless, standard RNN is affected by the vanishing and exploding gradient issue in training, namely, the gradients of node weights become diminishing or extending when the back propagation calculation is executed through multiple times. LSTM network [24] is RNN that resolved the vanishing/exploding gradient problems, where the hidden layers are replaced by recurrent gates. LSTM model has been utilized widely for many recent years in text analysis, emotion analysis, and speech recognition tasks, since the network restores its own memory of time sequences and can improve forecast accuracy. Thus in this paper we apply the search space of LSTM hyperparameters as following: for univariate data: number of units in the first layer $\in$ {50, 100, 150, 200}, number of units in the second layer $\in$ {50, 100, 150, 200}, for multivariate data: number of units $\in$ {200, 250, …, 500}; dropout rate $\in$ {0.01, 0.1}.

5. CNN~LSTM hybrid model. In feature engineering stage CNN network is often utilized to extract the most significant features, while LSTM has ability to memorize the time sequences with popular usage in time series analysis. According to these special characteristics of CNN and LSTM, a hybrid model based on CNN~LSTM is also constructed for solving stock price forecast. Moreover, the search space of hyperparameters in this hybrid model is a combination of search spaces in the abovementioned CNN and LSTM models.

3.4 Evaluation metrics

In this study *MAE, RMSE,* and R^2 are used as metrics to evaluate and compare the performance of the abovementioned forecast models. *MAE* (mean absolute error) computes the sum of absolute differences between true and predicted values, as shown in formula (3.1). *RMSE* (root mean square error) firstly calculates the average of sum of square magnitude of residuals caused by estimating predicted values, and then it gets the root value, as shown in formula (3.2). *MAE* and *RMSE* are measurements of closeness between the predicted and the actual values, which show the reliability of forecast model. Additionally, R^2 metric computes the linear correlation between 2 variables in comparison and helps eliminate the impact of dimensions on regression tasks, as shown in formula (3.3). We need to explore the best model that provides small *MAE, RMSE*, and high R^2. These 3 metrics are defined as follows:

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|y_{true}^{(i)} - y_{predict}^{(i)}\right| \tag{3.1}$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(y_{true}^{(i)} - y_{predict}^{(i)}\right)^2} \tag{3.2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m}\left(y_{true}^{(i)} - y_{predict}^{(i)}\right)^2}{\sum_{i=1}^{m}\left(y_{true}^{(i)} - \overline{y}\right)^2} \tag{3.3}$$

where regarding target variable of sample i, $y_{true}^{(i)}$ is the true value, $y_{predict}^{(i)}$ is the predicted value; $\overline{y}$ is the mean of true value of all samples; and m is the amount of samples in the dataset.

The following subsection presents results derived from different data preparation approaches and machine/deep learning models.

3.5. Empirical testing results

3.5.1. Univariate data approach

In this univariate data approach, basing on Adjusted close price data of previous 50 days, we have to explore the model that best predicts the Adjusted close price in the 51th day. Figure 6 demonstrates evaluation metrics of various adopted models. All positive R^2 values are also multiplied by 1000 for better visualization.

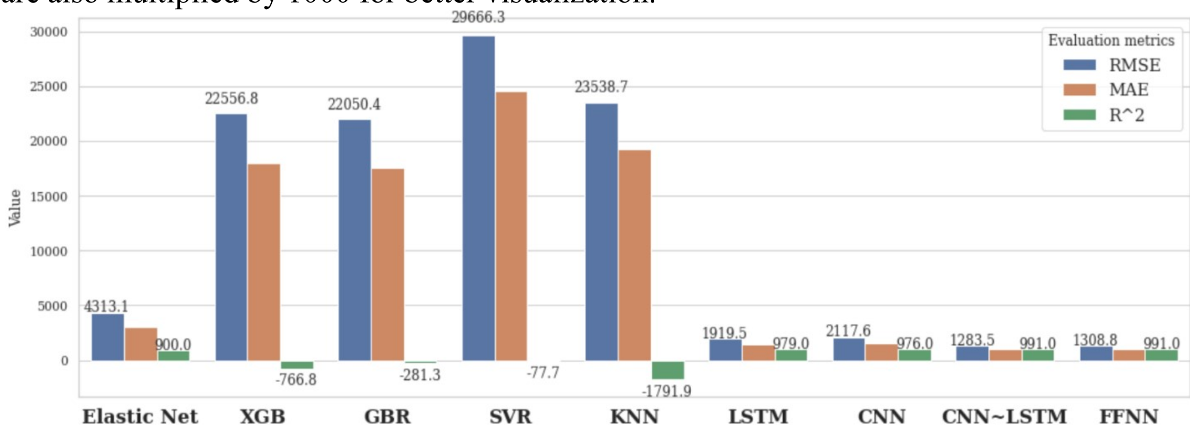
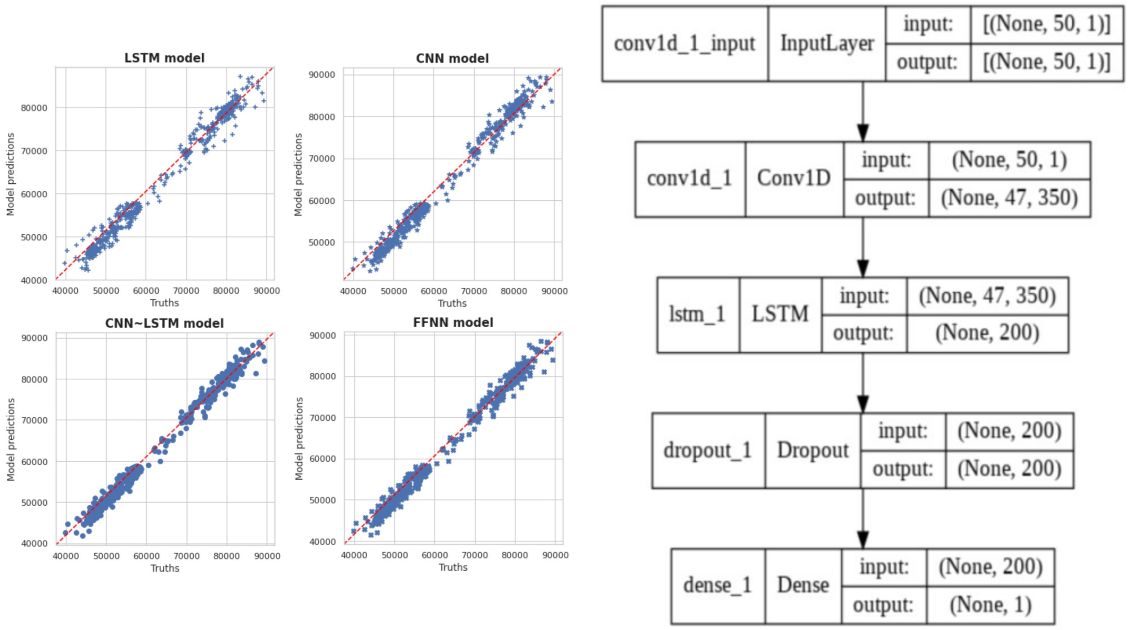


Figure 6. Evaluation metrics of various adopted models (univariate data)

From the figure 6 we can define that deep learning models, i.e., FFNN, CNN, LSTM, and CNN~LSTM, provide quite good forecast quality (low *MAE, RMSE*, and high R^2), as demonstrated by scatter plot in figure 7a. Basic purpose of scatter plot is displaying and observing relation between two numerical variables (true and predicted adjusted close price). The closer the points are to a diagonal (red dashed) line, the more precise forecast stock prices are.



(a) (b)

Figure 7. Evaluation of prediction quality: a) Scatter plot between truth values and predictions of the 4 best models; b) Structure of the best adopted model CNN~LSTM

Among them, CNN~LSTM model attains the best evaluation measures, i.e., *RMSE* = 1,283.49; *MAE* = 969.05; R^2 = 0.991. The final tuned hyperparameters are: filters = 350, kernels = 4, number of units = 200, dropout rate = 0.01. Also, the structure of the best model CNN~LSTM is presented in figure 7b.

3.5.2. Multivariate data approach

In multivariate data approach, we set horizon time window as 250 (the average number of working days within a year), as input data is financial data within a day (open/ adjusted close/ high/ low prices, trading volume), we need to select the model that best predicts Adjusted close price in the next 250th day. Figure 8 shows evaluation scores of various adopted models.

Rank	Model	RMSE	MAE	R^2	Best hyper-params
1	Elastic Net	14,622.6	11,393.0	-1.546	alpha=0.08, l1_ratio=1.00
2	CNN	19,520.0	16,835.3	-5.871	filters: 600, kernels: 3, drop_out: 0.1
3	FFNN	20,207.4	17,503.5	-7.096	num_layers: 2, units_1: 200, units_2: 100, drop_out: 0.1
4	LSTM	20,239.7	17,456.3	-9.395	LSTM_units: 350, drop_out: 0.01
5	XGB	27,723.8	23,115.5	-102	max_depth=10, n_estimators=221, min_child_weight=3, learning_rate=0.14, gamma=1.62
6	GBR	28,080.0	23,530.1	-127.24	max_depth=5, learning_rate=0.12, min_samples_split=87, min_samples_leaf=1, n_estimators=139
7	KNN	28,689.5	24,137.6	-130.95	n_neighbors = 27
8	SVR	48,136.9	45,855.9	-1131.50	C=10.0, gamma=scale, degree=3, kernel=rbf, epsilon=0.4
9	CNN~LSTM	66,451.3	65,055.7	-1200.2	filters: 200, kernels: 2, units = 300, dropout: 0.01

Figure 8. Evaluation scores of various adopted models (multivariate data)

From the above figure, we can identify the best model as Elastic Net, which provides the most optimal forecast testing metrics with respect to other models, i.e., *RMSE* = 14,622.6; *MAE* = 11,393; R^2 = –1.546. In this case, the test score is not as good as in univariate data approach, because the model needs to predict 250 days ahead with the financial input of one day; while in univariate data approach the model only has to forecast the stock price in the next day, basing on values of 50 previous days. Additionally, if we decrease the horizon time window, we can obtain better evaluation scores, for instance in case of time window as 50, then R^2 increases to 0.679, while $RMSE$ = 8,613.5; MAE = 6,761. However, as mentioned above, the goal of current research is to forecast the general upward/downward trend of stock prices, not the exact values. Thus, with utilization of the selected Elastic Net model, the upward trend is predicted quite precisely with test data, as shown in figure 9 below.

Figure 9. Comparison between true and predicted values by the best model Elastic Net

3.6. Samsung stock price prediction for 2022

From previous subsections, the best models for univariate and multivariate data have been chosen. Accordingly, for univariate data approach, the hybrid model CNN~LSTM gets the last 50 Adjusted close prices in 2021, and predicts the price in the next day. This derived value is appended to the end of input vector, which forms the input data for the next prediction phase. This forecast iteration is repeated 250 times. For the sake of clarity, the forecast algorithm from univariate data is illustrated in figure 10.

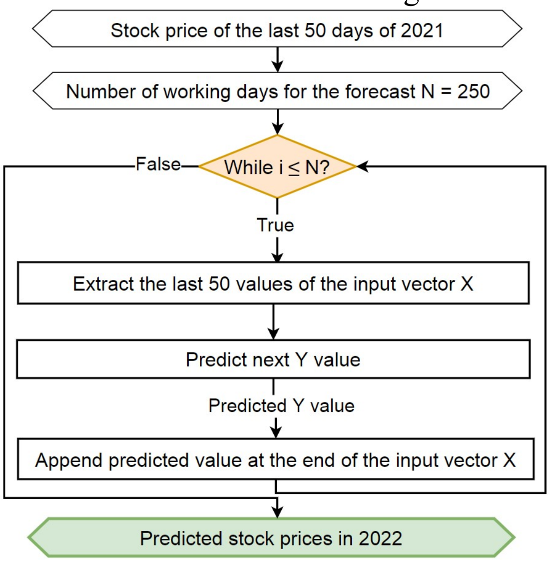


Figure 10. Sequential forecast algorithm from univariate data

In case of multivariate data, the best model Elastic Net gets the last 250 stock prices data in the dataset, each sample in this set provides predicted value in the next 250th day. Thus, with the input stock price data of 250 days from 29/12/2020 to 30/12/2021, the output predicted values will be approximately in date range of 2022. The predicted values of year 2022, derived from the 2 best models are demonstrated in figure 11.

Moreover, in order to check the trend created by the two models, authors also try to calculate another forecast values: it is the combination of these 2 lines, which means every new forecast value equals linear combination of univariate and multivariate forecast values with the coefficients 0.718 and 0.068, as shown in formula (3.4). These numbers are extracted after fitting linear regression model with true Samsung stock prices of the first 35 days in 2022. The new forecast values are also shown in the figure 11. Herein, the formula of new forecast future values (y_{fut}) basing on predicted values by multivariate (y_{mult}) and univariate data (y_{uni}) approaches is following:

$$y_{fut} = 0.718y_{mult} + 0.068y_{uni} \tag{3.4}$$

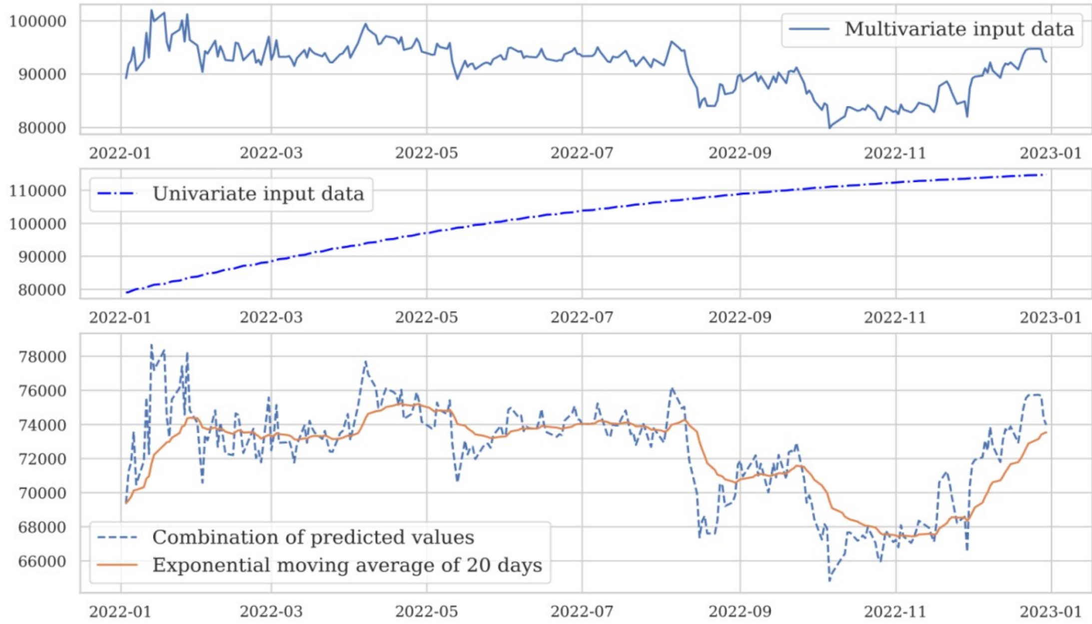


Figure 11. Samsung stock price prediction for 2022

Additionally, in order to check the general trend of Samsung stock prices in 2022 in comparison with the last 2 years, the authors depict all the last true and predicted values together in one figure 12. By observing this figure, we can conclude that generally stock prices of Samsung will keep the upward trend in some time periods of 2022.

Figure 12. The general trend of Samsung stock price

4. SAMSUNG PATENT ANALYSIS FOR INFERRING TECHNOLOGY TRENDS IN 2022

Arguments in section 1 have already stated that: if stock prices of an enterprise stay uptrend, then the technologies developed by the considered one will be likely to become promising innovations in future. After observing the predicted values, derived from proposed models, with high confidence we suppose that generally stock price of Samsung will keep the upward trend in some time periods of 2022, especially in comparison with 2020. For this reason, in this section we collect patent applications of Samsung, which were filed from 2020 to 2021, in order to predict technology trends in 2022. The metadata of Samsung patent applications was downloaded from the famous platform The Lens [3].

The filter query for obtaining Samsung patents is: *earliestPriorityDate = (2020-01-01 - 2021-12-31), Grouped by Simple Families, Document Type = (excl Design_right), Owner Name Exact = (Samsung Electronics Co Ltd)*. That means the earliest priority date (the earliest filing date in a family of patent applications) is from 01/01/2020 to 31/12/2021; and design right patents were also excluded. At the time of writing this paper (March, 2022), the query returns 1678 patent applications; accordingly, we can synthesize the most frequent CPC classification codes of Samsung patents, which are presented in figure 13.

Figure 13. Top CPC classification codes of developing technologies by Samsung

Among them, the top 5 CPC classification codes are:

1. G06N3/08: Computing settings based on biological models, which rely on physical entities operated by simulated intelligence to reproduce smart life arrangements;
2. H01L27: Devices made up of a majority of semiconductor or solid-state components formed in common substrate;
3. H01L25: Assemblies made up of a majority of individual semiconductor or solid state devices;
4. G06N3/0454: Computing settings established from biological models adopting integration of many neural networks;
5. G06F3: Input settings for transferring data to be processed by the computer; Output settings for transferring data from processing unit to output unit.

On the other hand, in order to understand technologies developed by Samsung, we devise a new method for analyzing patents with words clustering (figure 14).

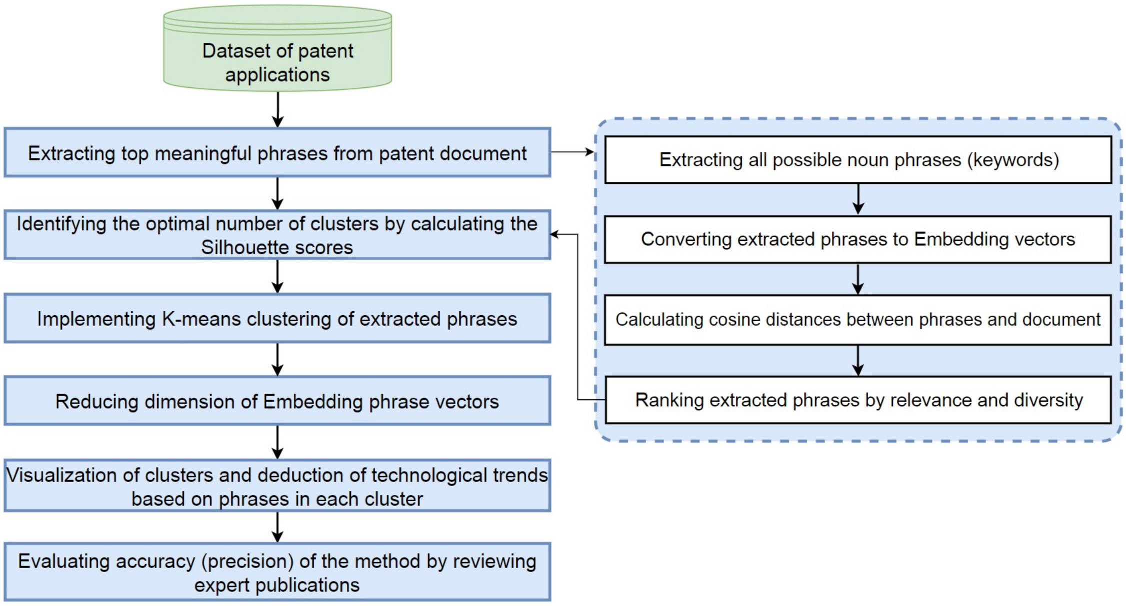


Figure 14. Method for analyzing patents with words clustering

Herein the stage “Extracting all possible noun phrases” consists of 3 steps:

– Part-of-speech tagging by utilizing spaCy library [14];

– Noun phrase identification: the longest possible noun phrases sequence of many consecutive words within a sentence such that the last word in the sequence is a noun and each of the other words is either a noun or an adjective;

– Eliminating plural forms by utilizing Natural Language Toolkit (NLTK) platform [28].

The pseudo-code of noun phrase extraction algorithm is presented in figure 15.

```
          Algorithm for extracting all noun phrases from the text
INPUT: Text and allowed part-of-speech tags allowed_postags:
ADJ (adjective), NOUN (noun), PROPN (pronoun)
OUTPUT: List of all noun phrases in singular form
________________________________________________________________
masked_words = []
while word in document do
    get POS tag of word
    if word POS tag is in allowed_postags or word == '-' do
        append word to masked_words
    else if previous word or next word == '-' do
        append word to masked_words
    else do
        append '.' to masked_words
    end
end
extracted_phrases = []
while index < masked_words length do
    get token word
    start_index = index
    if word != '.' doc
        start_index = last index that word not in {'.', 'ADJ'}
        if end_index > start_index + 1
            get phrase between (index and start_index)
            append phrase to extracted_phrases
        end
    end
    index = start_index + 1
end
lemmatized_phrases = []
while phrase in extracted_phrases do
    while token in phrase do
        get lemmatized word
    end
    join lemmatized words into phrase
    append pharase to lemmatized_phrases
end
```

Figure 15. Pseudo-code of noun phrases extraction algorithm

Then the authors adopt SentenceTransformers Python framework [36] in the stage "Converting extracted phrases to Embedding vectors" to convert extracted phrases into embedding vectors of 768 dimensions.

In the stage "Ranking extracted phrases by relevance and diversity" we follow the method Maximal marginal relevance, proposed by Maarten Grootendorst [17]. The idea of this method is minimizing redundancy (appearance of similar phrases) and maximizing the diversity of results in text summarization tasks (reducing data while increasing knowledge). Concretely, the method starts by selecting the phrase that is the most similar to the document. Then, it iteratively selects new promising phrases that are both similar to the document and different from the previously selected phrases by the diversity parameter, which is set to 0.3.

Herein authors select the most 3 meaningful phrases from each patent application, which consists of Title and Abstract, in the output of the stage "Extracting top meaningful phrases from patent documents". At the final result set, after eliminating duplicates we get 3496 unique noun phrases from 1678 patent applications of Samsung in 2020-2021.

Then the Silhouette score [38] is used to identify the optimal number of clusters, which is ranged from 80 to 120 (as a case study). Figure 16 shows that 112 clusters give the best Silhouette score.

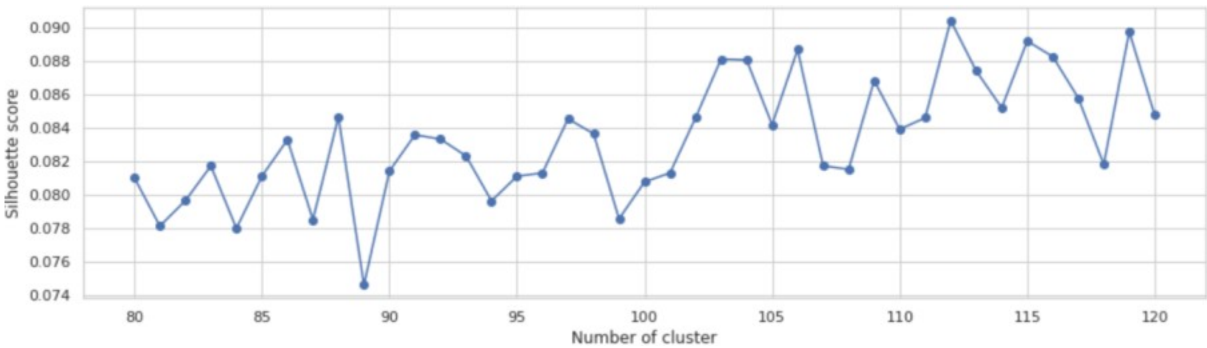


Figure 16. Silhouette analysis for optimal cluster number

For this reason, by using the K-means algorithm we divide 3496 extracted phrases into 112 clusters, herein 15 out of these 112 clusters are demonstrated in figure 17 (only some prominent phrases of each cluster are shown). Additionally, the Principle component analysis (PCA) method is adopted to convert these phrase vectors (768 dimensions) into 2D vectors for visualization.

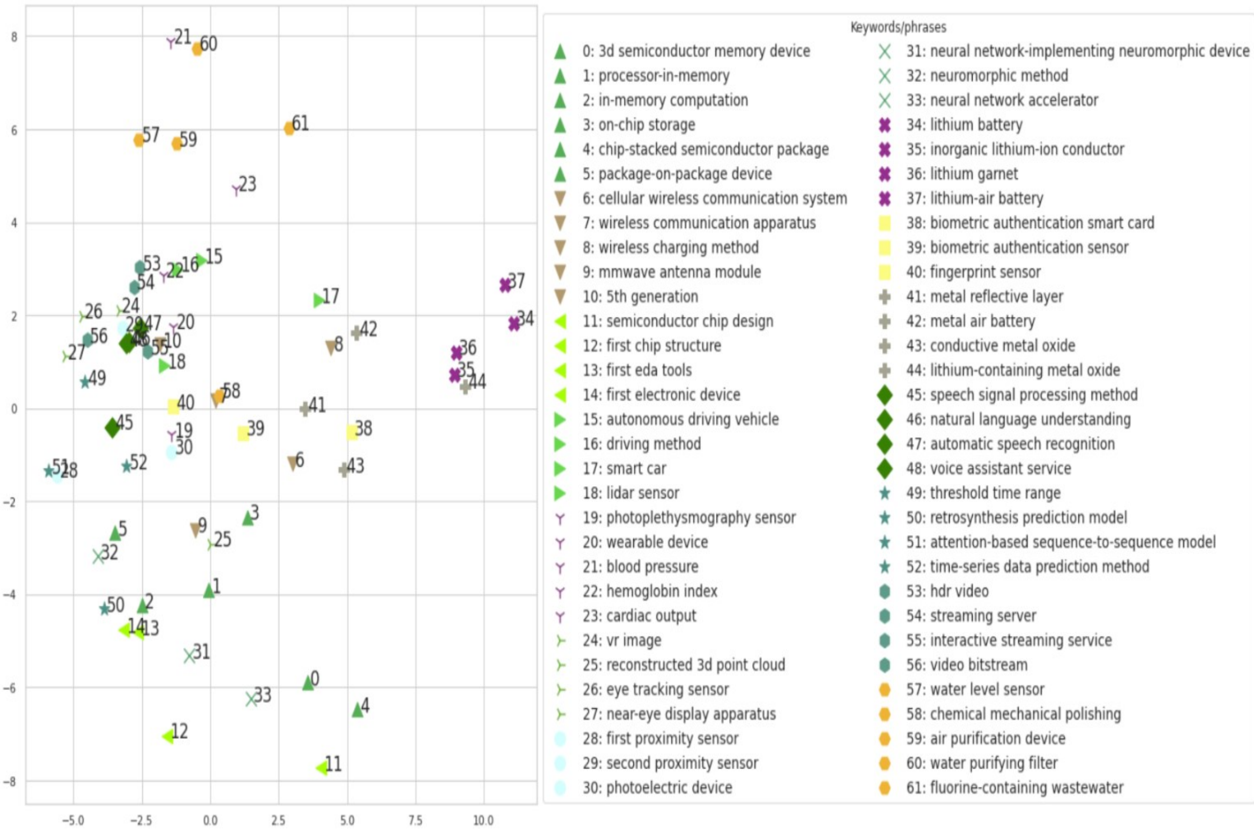


Figure 17. Visualization of 15 phrase clusters extracted from Samsung patent applications

Subsequently, from the obtained clusters we have inferred 16 technological sectors, which are considered as technology trends in 2022. Also, domain experts have recently proved these trends by their published articles, which are placed in corresponding following phrase clusters:

1) Semiconductor industry, Storage and computing integration, IC packaging (phrases: 3d semiconductor memory device, semiconductor package device, memory storage device, on-chip storage, semiconductor chip, semiconductor memory device, oxide semiconductor material, integrated circuit semiconductor device, in-memory operation unit, in-memory processing, processor-in-memory, in-memory computation, chip-stacked semiconductor package, package-on-package device, etc.): This can remove the delay and power consumption of data transfer, hundreds of times more efficient AI computing, hence it is especially appropriate for neural networks [4], [43].

2) Wireless technology, 5G network (phrases: cellular wireless communication system, wireless communication apparatus, wireless power transmitter, wireless charging method, mmwave antenna module, pre-5th-generation, 5th generation, etc.): 5G construction focuses on independent networking and millimeter-wave [4].

3) Electronic design automation – EDA (phrases: semiconductor chip design, first chip structure, first eda tools, first electronic device, etc.): AI will not only be the final formation of EDA development in future, but also the central solution to enhance the performance of the chip design in the upcoming years [4].

4) Autonomous driving technology (phrases: autonomous driving vehicle, automatic driving device, driving method, smart car, lidar sensor, etc.): In the future the car will certainly be an electromechanical intelligence. The current subsystems are integrated as much as possible will be a trend [4], [1], [45].

5) Bio-Sensing-PPG (photoplethysmography) technology (phrases: photoplethysmography sensor, wearable device, ear wearable device, blood pressure, systolic blood pressure, mean arterial pressure, hemoglobin index, cardiac output, blood vessel, pulmonary health parameter, blood compound concentration value, etc.): With upgrading algorithm, wearable devices are more and more likely to represent PPG-based bio-sensing functioning, which can determine not only heart rate, but also other parameters like body hydration, blood lipid, blood glucose, and blood oxygen levels in the future [45].

6) Metaverse – Eye Tracking and 3D Sensing (phrases: vr image, screen mirroring, virtual 3d graphic information, 3d navigation information, reconstructed 3d point cloud, eye tracking sensor, object recognition method, sensor-detected movement, trigger recognition, target tracking method, near-eye display apparatus, position sensor, etc.) [45].

7) Proximity sensor (phrases: first proximity sensor, second proximity sensor, photoelectric device, photoelectric conversion, etc.): The Proximity sensor market was evaluated at 3790.5 million USD in 2020 and is predicted to achieve 5761.12 million USD by 2026 and expand at 7.3% CAGR over the period 2021-2026 [45], [32].

8) Neuromorphic computing technology (phrases: neural network-implementing neuromorphic device, neuromorphic method, processor-implemented neural network, neural network accelerator, deep convolutional neural network, etc.): The prominent advantages of neuromorphic computing in comparison with usual procedures are fast executive speed, energy efficient, minimizing local failures, and the capacity to process and learn simply. The obvious potentials for AI is the greatest driving force for investing in neuromorphic computing [39].

9) Lithium-ion batteries – LIBs (phrases: lithium battery, inorganic lithium-ion conductor, lithium-air battery, lithium garnet, battery pack uptime, rechargeable battery, etc.). LIBs have been one of the leading energy storage solutions recently. The market share and application fields of LIBs have been increasing swiftly and continuously exhibiting a stable rising trend [23], [44].

10) Biometric industry (phrases: biometric authentication smart card, biometric authentication sensor, biometric authentication data, virtual enrollment fingerprint, fake fingerprint detection, fingerprint sensor, fingerprint verification, blended fingerprint image, etc.) [7].

11) Metal-air battery (phrases: metal reflective layer, metal air battery, conductive metal oxide, lithium-containing metal oxide, metal-semiconductor oxide, bump metal layer, pad metal pattern, metal interconnect pattern, etc.): The worldwide Metal-air battery market size was evaluated at 540.6 million USD in 2021 and is forecasted to reach 1303.5 million USD with a 13.4% CAGR by 2028 [26].

12) Natural language processing (phrases: speech signal processing method, natural language understanding, automatic speech recognition, audio processing block, voice assistant service, expressive text-to-speech system, voice recognition system): The

worldwide Natural language processing market is forecasted to reach at 18.78% CAGR during period 2017-2023 [27].

13) Time series intelligence software (phrases: forecasting model, threshold time range, retrosynthesis prediction model, attention-based sequence-to-sequence model, timing relationship indication, input prediction time, time-series data prediction method): [42].

14) Video streaming (phrases: video quality assessment method, hdr video, streaming server, interactive streaming service, video-based point cloud compression data, video bitstream, video stream, etc.) [2].

15) Water purifier and wastewater treatment technology (phrases: purifying filter, automatic water supply device, water level sensor, wet cleaning process, chemical mechanical polishing, air purification device, water purifier body, water purifying filter, clean water flow path, fluorine-containing wastewater, etc.) [51].

16) Plasma technology (phrases: plasma electrode, plasma processing system, local plasma cleaning process, plasma treatment apparatus, plasma generator, microwave plasma source, etc.).

Accordingly, 15 out of these 16 technology clusters have been approved and predicted to be quite significant in the near future, hence the evaluation for prediction accuracy (precision) can be estimated as 15/16 (~93.8%).

5. CONCLUSION

Prediction of technological development trends impact essentially on making strategic decisions. The early realization of potential emerging and promising technological trends helps enhance the market position and competitiveness of companies. For this reason, if innovation-oriented companies disregard emerging and promising technologies, they will not exploit the entire potentials and advantages of their own technologies or products.

A domain that is being revolutionized by the progresses of technologies is the stock and financial market. There are different ways of how technologies have affected and made the present shape of the financial markets, and the future state as well. Concretely, there is a strong relation between the volatility of stock prices and market shares when technology is still unpredictable through premature industrial evolution, while numerous previous studies showed that sudden shifts in R&D and patents are really correlated with substantial shifts in the firm market value. Likewise, experimental confirmation with stock prices revealed that firms succeeded in investing in comprehensive innovation attained higher stock incomes.

For this reason, if stock prices of an enterprise keep uptrend in some time periods in comparison with the last few years, then the technologies developed by considered one will be likely to become promising innovations in the future. Relying on this assumption, also existing progress and problems stated in the previous studies [46], [47], in this paper we proposed a novel method to predict promising technology trends based on processing multiple data sources by aggregating following steps: mining Web news to extract significant high-technology enterprises, forecast stock price trends of selected ones, and patent clustering analysis. Different from other studies, our proposed method promotes an idea of predicting technology trends by forecasting stock price trend, with the utilization of Bayesian optimization for exploring best hyperparameters of machine/deep learning models, also a new method of patent analysis. Moreover, authors combined using univariate and multivariate data preparation approaches for predicting stock price trend. After collecting patent applications of Samsung company (as a case study) and extracting most significant noun phrases of patent applications (consisting of titles and abstracts), the clustering method of noun phrases is utilized to explore technology trends developed by the enterprise. These technology trends have recently been confirmed by domain experts in their corresponding published articles. Thus, the obtained forecast accuracy (precision) is about 93.8%, which proves that the proposed method has gained positive reliability.

Further work will focus on improving the evaluation metrics (*RMSE*, *MAE*, R^2) by combining other relevant data sources (sentiment analysis in social networks, technological news, technical financial indicators), in order to enhance stock price prediction models using univariate and multivariate data approaches.

ACKNOWLEDGMENTS

The reported study was funded by RFBR, projects No. 19-07-01200 and 20-37-90092.

REFERENCES

1. *5 Trends to Watch in 2022: Autonomous Vehicles*. (2022). [Online] Available https://www.gtlaw.com/en/insights/2022/1/published-articles/5-trends-to-watch-in-2022-autonomous-vehicles
2. *9 Streaming Trends for 2022/2023: Top Predictions You Should Be Thinking About*. (2022). [Online] Available https://financesonline.com/streaming-trends/
3. *About The Lens.* (2022). [Online] Available https://about.lens.org/
4. *Top 10 Technology Trends in the Global Semiconductor Industry in 2022*. (2022). [Online] Available https://www.utmel.com/blog/categories/semiconductor/top-10-technology-trends-in-the-global-semiconductor-industry-in-2022
5. Arenas, L. & Gil-Lafuente, A. M. (2021). Impact of emerging technologies in banking and finance in Europe: A volatility spillover and contagion approach, *Journal of Intelligent & Fuzzy Systems*, **40**(2), 1903–1919 https://doi.org/10.3233/JIFS-189195
6. Aue, A., Horváth, L. & F. Pellatt, D. (2017). Functional Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Time Series Analysis*, **38**(1), 3–21 https://doi.org/10.1111/jtsa.12192
7. *Biometric Trends and Statistics to Keep an Eye on in 2022*. (2022). [Online] Available https://imageware.io/biometric-trends-and-statistics/
8. Chen, Y., Lin, W. & Wang, J. Z. (2019). A Dual-Attention-Based Stock Price Trend Prediction Model With Dual Features, *IEEE Access*, **7**, 148047–148058 https://doi.org/10.1109/ACCESS.2019.2946223
9. *Conv1D layer*. (2022). [Online] Available https://keras.io/api/layers/convolution layers/convolution1d/
10. Ding, G. & Qin, L. (2020). Study on the prediction of stock price based on the associated network model of LSTM, *International Journal of Machine Learning and Cybernetics*, **11**, 1307–1317 https://doi.org/10.1007/s13042-019-01041-1
11. Dosso, M. & Vezzani, A. (2020). Firm market valuation and intellectual property assets, *Industry and Innovation*, **27**(7), 705–729 https://doi.org/10.1080/13662716.2019.1685374
12. Grieco, D. (2018). Innovation and stock market performance: A model with ambiguity-averse agents, *Journal of Evolutionary Economics*, **28**(2), 287–303 https://doi.org/10.1007/s00191-017-0537-1
13. Hyndman, R. J. & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. Online: OTexts.
14. *Industrial-Strength Natural Language Processing*. (2022). [Online] Available https://spacy.io/
15. Ji, X., Wang, J. & Yan, Z. (2021). A stock price prediction method based on deep learning technology. *International Journal of Crowd Science*, **5**(1), 55–72 https://doi.org/10.1108/IJCS-05-2020-0012
16. Jiang, W. (2021). Applications of deep learning in stock market prediction: Recent progress, *Expert Systems with Applications*, **184** https://doi.org/10.1016/j.eswa.2021.115 537
17. *KeyBERT – Quickstart*. (2022). [Online] Available https://maartengr.github.io/

KeyBERT/guides/quickstart.html
18. Kravets, A., Gneushev, V., Biryukov, S., Skorikov, D. & Marinkin, D. (2020). Research of the LDA algorithm processing results on high-level classes of patents, *CEUR Workshop Proceedings* (Samara, Russia), 314–318.
19. Lam, H. K. S., Ding, L., Cheng, T. C. E. & Zhou, H. (2019). The impact of 3D printing implementation on stock returns, *International Journal of Operations & Production Management*, **39**(6/7/8), 935–961 https://doi.org/10.1108/IJOPM-01-2019-0075
20. Lee, B. (2018). The Relationship between Technology Life Cycle and Korean Stock Market Performance, *International Journal of Financial Studies*, **6**(4), 88 https://doi.org/10.3390/ijfs6040088
21. Li, A. W. & Bastos, G. S. (2020). Stock Market Forecasting Using Deep Learning and Technical Analysis: A Systematic Review, *IEEE Access*, **8**, 185232–185242 https://doi.org/10.1109/ACCESS.2020.3030226
22. Lin, Y., Liu, S., Yang, H. & Wu, H. (2021). Stock Trend Prediction Using Candlestick Charting and Ensemble Machine Learning Techniques With a Novelty Feature Engineering Scheme, *IEEE Access*, **9**, 101433–101446 https://doi.org/10.1109/ACCESS.2021.3096825
23. Liu, Y., Zhang, R., Wang, J. & Wang, Y. (2021). Current and future lithium-ion battery manufacturing, *IScience*, **24**(4), 102332 https://doi.org/10.1016/j.isci.2021.102332
24. *LSTM layer*. (2022). [Online] Available https://keras.io/api/layers/recurrent_layers/lstm/
25. Mazzucato, M. (2006). Innovation and Stock Prices: a Review of some Recent Work, *Revue de l'OFCE*, **97**(5), 159–179 https://doi.org/10.3917/reof.073.0159
26. *Metal-air Battery Market Trends, Business Overview, Key Players, Demand and Forecast to 2028*. (2022). [Online] Available https://www.marketwatch.com/press-release/metal-air-battery-market-trends-business-overview-key-players-demand-and-forecast-to-2028-2022-02-09
27. *Natural Language Processing market 2022 Growth Analysis by Key Players, Globally Effective Factors, Trends, Business Plans, and Forecast to 2030*. (2022). [Online] Available https://www.taiwannews.com.tw/en/news/4399824
28. *Natural Language Toolkit*. (2022). [Online] Available https://www.nltk.org/index.html
29. Nguyen, T.-T. & Yoon, S. (2019). A Novel Approach to Short-Term Stock Price Movement Prediction using Transfer Learning, *Applied Sciences*, **9**(22), 4745 https://doi.org/10.3390/app9224745
30. O'Malley, T., Bursztein, E., Long, J. & Chollet, F. (2019). *Keras Tuner*. [Online] Available https://keras.io/keras_tuner/
31. Parray, I. R., Khurana, S. S., Kumar, M. & Altalbe, A. A. (2020). Time series data analysis of stock price movement using machine learning techniques, *Soft Computing*, **24**(21), 16509–16517 https://doi.org/10.1007/s00500-020-04957-x
32. *Proximity sensor market - growth, trends, covid-19 impact, and forecasts (2022 - 2027)*. (2022). [Online] Available https://www.mordorintelligence.com/industry-reports/global-proximity-sensors-market-industry
33. *Samsung Electronics Co., Ltd*. (2022). [Online] Available https://finance.yahoo.com/quote/005930.KS/history?p=005930.KS
34. *Scikit-optimize. Sequential model-based optimization in Python*. (2020). [Online] Available https://scikit-optimize.github.io/stable/
35. Senarathne, C. W. & Wei, J. (2018). The impact of patent citation information flow regarding economic innovation on common stock returns: Volume vs. patent citations, *International Journal of Innovation Studies*, **2**(4), 137–152 https://doi.org/10.1016/j.ijis.2019.02.001
36. *SentenceTransformers Documentation*. (2022). [Online] Available https://www.sbert.net/
37. *Sklearn.linear_model.ElasticNet*. (2022). [Online] Available https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.ElasticNet.html

38. *Sklearn.metrics.silhouette_score*. (2022). [Online] Available https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html
39. *The global neuromorphic computing trends are set to dominate computer systems in 2022*. (2022). [Online] Available https://www.analyticsinsight.net/know-neuromorphic-computing-trends-to-leverage-in-2022/
40. *The Sequential model*. (2020). [Online] Available https://keras.io/guides/sequential_model/
41. Thu, H. G. T., Thanh, T. N. & Quy, T. Le. (2021). A Neighborhood Deep Neural Network Model using Sliding Window for Stock Price Prediction, *2021 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 69–74 https://doi.org/10.1109/BigComp51126.2021.00022
42. *Time Series Intelligence Software Market 2022: Booming Across The Globe By Share, Growth Size, Key Segments and Regional Forecast by 2027*. (2022). [Online] Available https://www.marketwatch.com/press-release/time-series-intelligence-software-market-2022-booming-across-the-globe-by-share-growth-size-key-segments-and-regional-forecast-by-2027-2021-12-16
43. *Top 10 Electronics Manufacturing Trends, Technologies & Innovations for 2022*. (2022). [Online] Available https://www.startus-insights.com/innovators-guide/electronics-manufacturing-trends
44. *Top 10 Renewable Energy Trends & Innovations in 2022*. (2022). [Online] Available https://www.startus-insights.com/innovators-guide/top-10-renewable-energy-trends-2022/
45. *TrendForce 2022 Infrared Sensing Application Market and Branding Strategies- Open up New Horizons*. (2022). [Online] Available https://www.ledinside.com/node/32494
46. Viet, N. T. & Gneushev, V. (2021). Analyzing and Forecasting Emerging Technology Trends by Mining Web News. In *Communications in Computer and Information Science*, 55–69 https://doi.org/10.1007/978-3-030-87034-8_5
47. Viet, N. T., Kravets, A. & Duong, Q. H. T. (2021). Data Mining Methods for Analysis and Forecast of an Emerging Technology Trend: A Systematic Mapping Study from SCOPUS Papers. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 81–101 https://doi.org/10.1007/978-3-030-86855-0_7
48. Viet, N. T. & Kravets, A. G. (2019). Analyzing Recent Research Trends of Computer Science from Academic Open-access Digital Library. *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, 31–36 https://doi.org/10.1109/SMART46866.2019.9117215
49. Vitt, C. A. & Xiong, H. (2015). The Impact of Patent Activities on Stock Dynamics in the High-Tech Sector. *2015 IEEE International Conference on Data Mining*, 399–408 https://doi.org/10.1109/ICDM.2015.95
50. Vo, N. N. Y., He, X., Liu, S. & Xu, G. (2019). Deep learning for decision making and the optimization of socially responsible investments and portfolio, *Decision Support Systems*, *124* https://doi.org/10.1016/j.dss.2019.113097
51. *Water Purifier Market: Global Industry Trends, Share, Size, Growth, Opportunity and Forecast 2022-2027*. (2022). [Online] Available https://www.researchandmarkets.com/reports/5530617/water-purifier-market-global-industry-trends
52. Wu, J., Chen, X. Y., Zhang, H., Xiong, L. D., Lei, H. & Deng, S. H. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimization, *Journal of Electronic Science and Technology*, **17**(1), 26–40 https://doi.org/10.11989/JEST.1674-862X.80904120
53. Xiao, C., Xia, W. & Jiang, J. (2020). Stock price forecast based on combined model of ARI-MA-LS-SVM, *Neural Computing and Applications*, **32**(10), 5379–5388 https://doi.org/10.1007/s00521-019-04698-5

54. Xu, M. (2021). The Impact of Blockchain Technology on Stock Price: An Emprical Study, *SHS Web of Conferences*, **96**, 1–5 https://doi.org/10.1051/shsconf/20219604008
55. Yang, C., Zhai, J. & Tao, G. (2020). Deep Learning for Price Movement Prediction Using Convolutional Neural Network and Long Short-Term Memory, *Mathematical Problems in Engineering*, **2020**(6), 1–13. https://doi.org/10.1155/2020/2746845
56. Zhao, J., Zeng, D., Liang, S., Kang, H. & Liu, Q. (2021). Prediction model for stock price trend based on recurrent neural network, *Journal of Ambient Intelligence and Humanized Computing*, **12**(1), 745–753 https://doi.org/10.1007/s12652-020-02057-0