# Data Science in Economics

**Saeed Nosratabadi** ( saeed.nosratabadi@phd.uni-szie.hu )
Faculty of Humanities and Social Sciences, Oxford Brookes University, Oxford OX30BP, UK
https://orcid.org/0000-0002-0440-6564

**Amir Mosavi**
Faculty of Humanities and Social Sciences, Oxford Brookes University, Oxford OX30BP, UK
https://orcid.org/0000-0003-4842-0613

**Puhong Duan**
College of Electrical and Information Engineering, Hunan University, Changsha 410082, China

**Pedram Ghamisi**
Exploration Devision, Helmholtz Institute Freiberg for Resource Technology, Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany

**Ferdinand Filip**
Department of Mathematics, J. Selye University, 94501 Komarno, Slovakia

---

**Systematic Review**

**Keywords:** data science, deep learning, ensemble machine learning models, economics, hybrid machine learning, marketing, stock market, stock price, prediction, economics, cryptocurrency, customer behavior

**Posted Date:** May 8th, 2020

**DOI:** https://doi.org/10.21203/rs.3.rs-27318/v1



---

# Data Science in Economics

**Saeed Nosratabadi [1], Amir Mosavi [1,2]*, Puhong Duan [3], Pedram Ghamisi [4], and Ferdinand Filip [5]**

[1] Faculty of Humanities and Social Sciences, Oxford Brookes University, Oxford OX30BP, UK
[2] Faculty of Civil Engineering, Technische Universität Dresden, 01069 Dresden, Germany;
[3] College of Electrical and Information Engineering, Hunan University, Changsha 410082, China; puhong_duan@hnu.edu.cn
[4] Exploration Devision, Helmholtz Institute Freiberg for Resource Technology, Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany; p.ghamisi@hzdr.de
[5] Department of Mathematics, J. Selye University, 94501 Komarno, Slovakia; filipf@ujs.sk

*Correspondence: a.mosavi@brooks.ac.uk;

Received: date; Accepted: date; Published: date

**Abstract:** This paper provides the state of the art of data science in economics. Through a novel taxonomy of applications and methods advances in data science are investigated. The novel data science methods and applications are investigated in four individual classes of deep learning models, hybrid deep learning models, hybrid machine learning, and ensemble models. Application domains include a wide range of economics research, from stock market, marketing, E-commerce, to corporate banking, and cryptocurrency. Prisma method, a systematic literature review methodology is used to ensure the quality of the survey. The findings revealed that the trends are on the advancement of hybrid models. On the other hand, based on the accuracy metric it is also reported that the hybrid models outperform other learning algorithms. It is further expected that the trends would go toward the advancements of sophisticated hybrid deep learning models.

**Keywords:** data science; deep learning; ensemble machine learning models; economics; hybrid machine learning

## 1. Introduction

Due to the remarkable progress in data analysis and use, the use of data science in various disciplines has been increasing exponentially. Like other disciplines, economics has benefited from the advancements of data science. Advancements of data science in economics have been progressive and have reported promising results in the literature. Several studies suggest that applications of data science in economics can be categorized and studied in various emerging technologies. Deep learning, hybrid learning models, and ensemble algorithms are among the most popular technologies. Machine learning (ML) algorithms provides the ability of learning from data and deliver in depth insight into problems. Researchers use machine learning models to solve various problems associated with economics. Notable applications of data science in economics are presented in Table 1. Deep Learning (DL) as an emerging field of machine learning, are today applied in many aspects of the today's society from the self-driving cars to image recognition, hazard prediction, and health informatics [1-3]. Literature includes several comparative studies to evaluate the performance of DL models with standard ML models, e.g., support vector machine (SVM), K-nearest neighbors (KNN), generalized regression neural networks (GRNN) in applications to economics. The evolution of DL methods progresses in a fast pace, and every day, many sections and disciplines are added to the number of users and beneficiaries of DL algorithms. On the other hand, the hybrid machine learning models which consist of two or more single algorithms are other emerging technology in data science used to increase the accuracy of the models [4]. Hybrid models can

be formed by combining two predictive machine learning algorithms or a machine learning algorithm and an optimization method to maximize the prediction function [5]. It is demonstrated that the hybrid machine learning models outperform the single algorithms and such an approach has improved the predication accuracy [6-8]. Ensemble machine learning algorithms are one of the supervised learning algorithms that use multiple learning algorithms to improve learning processes and increase predictive accuracy [4]. Ensemble models apply different training algorithms to enhance training and learning from data.

Table 1. Notable classic machine learnings

| Sources | Machine learning models | Objectives |
|---|---|---|
| Lee et al. [9] | Support Vector Regression (SVR) | Anomaly Detection |
| Husejinović [10] | Naive Bayesian And C4.5 Decision Tree Classifiers | Credit Card Fraud Detection |
| Zhang [11] | Improved BP Neural Network | Aquatic Product Export Volume Prediction |
| Sundar and Satyanarayana [12] | Multilayer Feed Forward Neural Network | Stock Price Prediction |
| Hew et al. [13] | Artificial Neural Network (ANN) | Mobile Social Commerce |
| Abdillah and Suharjito [14] | Adaptive Neuro Fuzzy Inference System (ANFIS) | E-Banking Failure |
| Sabaitytė et al. [15] | Decision Tree (DT) | Customer Behavior |

The body of literature embraced the review papers on the state-of-the-art of DL methods in different disciplines such as image recognition (e.g. [16]), animal behavior (e.g. [17]), renewable energy forecasting (e.g. [18]), and the review papers on hybrid methods in different various fields such as financial time series [19], solar radiation forecasting [20], FOREX rate prediction [21], and the review papers on ensemble methods in fields such as breast cancer [22], image categorization[23], electric vehicle user behavior prediction [24], and solar power generation forecasting [25].

Despite the fact that many researchers applied deep learning methods to address different problems in the field of economics (see Figure 1), these studies are scattered. However, no single study exists to provide a comprehensive view of the contributions of data science in economics related fields. Therefore, the current study is conducted to bridge this literature gap. Therefore, the main objective of this study is to investigate the advancement of data science in three parts: deep learning methods, hybrid deep learning methods, and finally ensemble machine learning techniques in economics related fields. In other words, the present study answers the following research questions: 1) What economics-related applications have data science contributed to? 2) what data science models are applied in each of these applications?

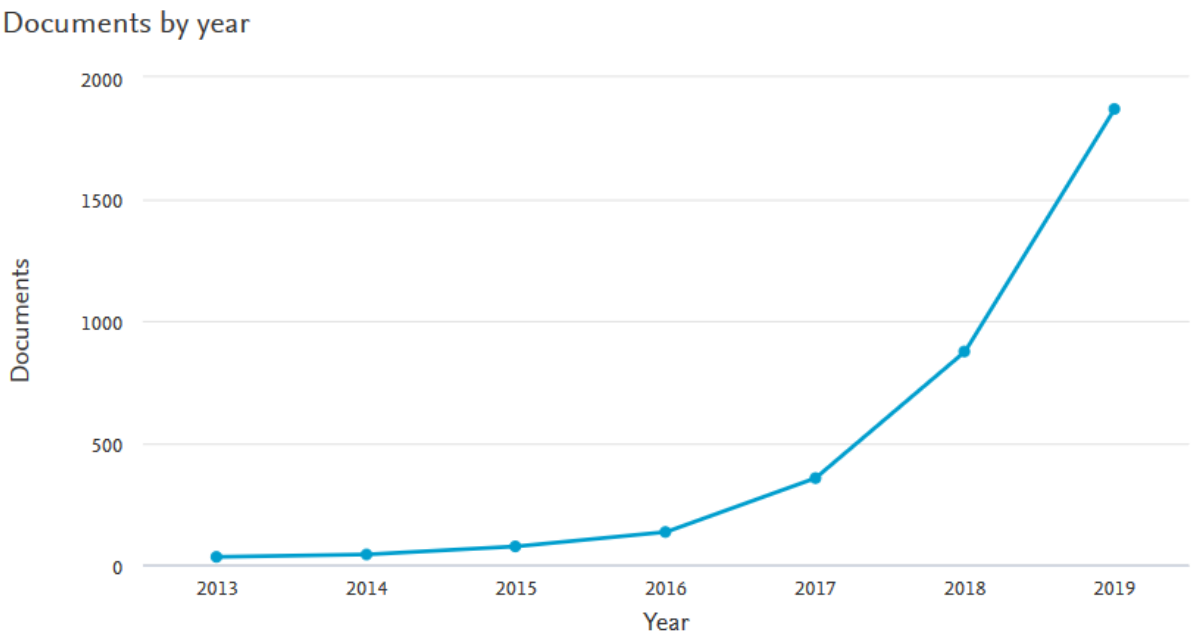



**Figure 1.** Increasing application of data science in Economics

**In the following, the method by which the database of this article was formed is described in Section 2. In Section 3, that is findings and discussion section, articles were categorized by sector and area of study, and by models. In this section, a taxonomy of application of data science in economics is presented as well. Finally, the findings of this study are analyzed, and the conclusions of the article are presented.**

## 2. Materials and Methods

**The current study applied Prisma, a systematic literature review approach, to find the most published articles applying data science methods for addressing an issue in a field related to Economics. Systematic literature review based on Prisma method includes four steps: 1) identification, 2) screening, 3) eligibility, 4) inclusion. In the identification stage, the documents are identified through initial search among the mentioned databases. In this study the review is limited to the original peer-review research articles published through Thomson Reuters Web-of-Science (WoS) and Elsevier Scopus. This review is limited to an article written in English. This step resulted in 217 articles. The screening step includes two stages on which, first, duplicate articles are eliminated. As a result, 135 unique articles moved to the next stage, where the relevance of the articles is examined on the basis of their title and abstract. The result of this step was 84 articles for further consideration. The next step of Prisma model is eligibility, in which the full text of articles was read by the authors and 57 of them considered eligible for final review in this study. The last step of the Prisma model is the creation of the database of the study, which is used for qualitative and quantitative analysis. The database of the current study comprises 57 articles and all the analysis in this study took place based on these articles. Figure 2 illustrates the steps of creating the database of the current study based on the Prism method.**

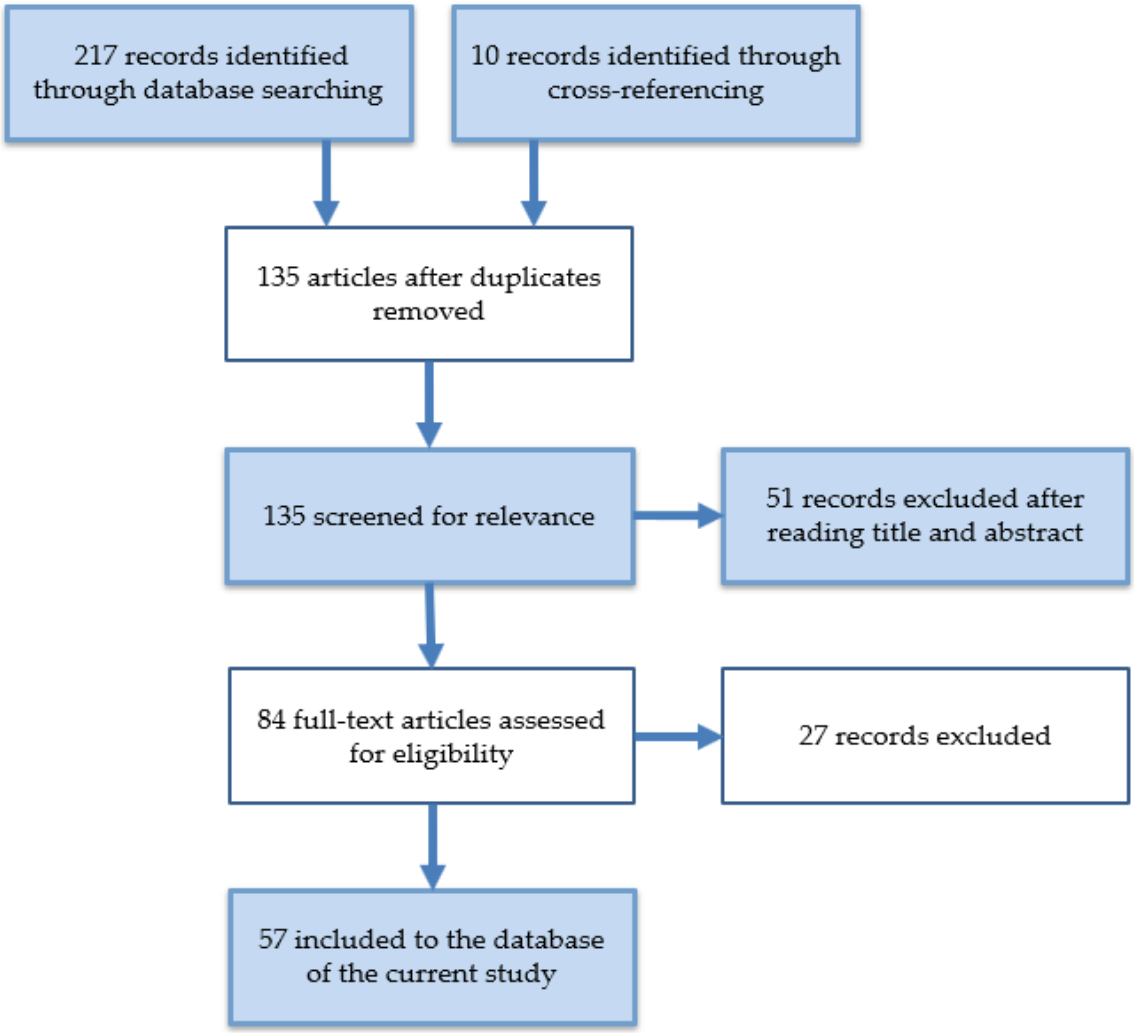



**Figure 2.** Diagram of systematic selection of the study database

## 3. Findings and Discussion

Figure 2 shows that the database of this study consists of 57 articles that are analyzed and categorized according to two criteria: 1) by research / application area and 2) by method type. Based on the review of articles by application, it was found that these articles were designed to address the issues of five different applications of Stock Market (with 37 articles), Marketing (with 6 articles), E-commerce (with 8 articles), Corporate Bankruptcy (with 3 articles), and Cryptocurrency (with 3 articles). In addition, the articles are analyzed by type of methods and it was revealed that 42 unique algorithms employed among 57 reviewed articles (see Figure 3) in which 9 of them used 9 single DL models (see Table 8), 18 HDL models (see Table 9), 7 hybrid machine learning models (see table 10), and 8 ensemble models (see table 11). The applications are first identified and then each of these methods is described in detail. In the following, the identified applications and each of these methods are described in detail.

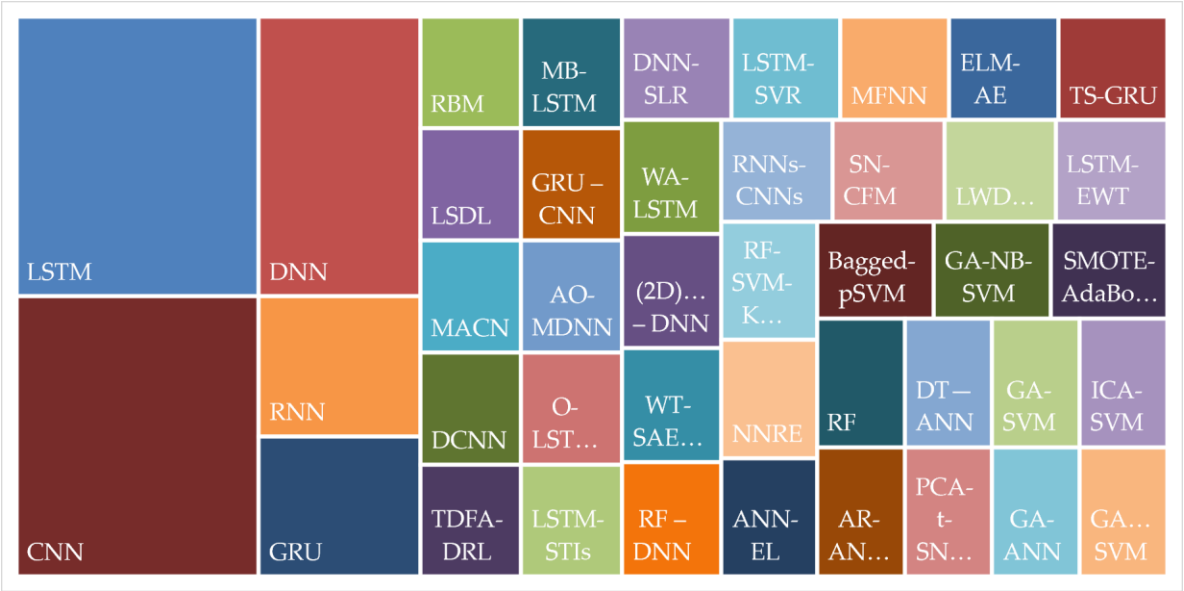



**Figure 3.** List of all deep learning and hybrid deep learning models applied in Economics related fields

### *3.1. Applications of data science in economics*

#### 3.1.1 Stock Market

Applying deep learning in the stock market has been more common than in other areas of economics as most of the research studies reviewed in the current study are classified in this category (37 out of 57). Table 2 provides a summary of the articles employed predictive models in stock market studies. This table presents the research objective, data source, and applied models of each article. Investment in the stock market is profitable, while higher the profit, higher the risk. Therefore, the investors always try to determine and estimate the stock value before any action. The stock value is often influenced by a sort of uncontrollable economic and political factors that make it notoriously difficult to identify the future stock market trends. On the one hand, the nature of stock market is so volatile and complex, and on the other hand, the financial time series data are so noisy and nonstationary. Thus, the traditional forecasting models are not reliable enough to predict the stock value and researchers are seeking new methodologies based on data science (DS) algorithms to enhance the accuracy of such predictions. Forecasting stock price has been the objective of 29 out of 37 articles. In addition, there are studies aimed at applying DS for the purpose of sentiment analysis (sentiment analysis refers to analyzing the context of texts to extracts subjective information) to find the future trends in the stock market. Besides, portfolio management, algorithmic trading (i.e., using a pre-programmed automated system for trading), automated stock trading, socially responsible investment portfolios, the S&P 500 index trend prediction, and exchange-trade-fund (EFT) options prices prediction were the objectives of other articles stimulating to employ DS methods. Financial time series have been the data source of all these studies, except for the studies aimed at sentiment analysis that used different data sources such as social media and financial news.

#### LSTM

Long short-term memory (LSTM) networks are a special kind of recurrent neural networks (RNN), which can overcome the main issue of RNN, i.e., vanishing gradients by using the gates to selectively retain the relevant information and discard the unrelated information. The structure of an LSTM neural network is shown in Figure 4, which is composed of a memory unit $C$, a hidden stat $h$ and three types of

gates. Specifically, for each step $t$, LSTM receives an input $x_t$ and the previous hidden state $h_{t-1}$. Then, it calculates activations of the gates. Finally, the memory unit and the hidden state are updated. The computations involved are described below:

$$f_t = \sigma(W_f x_t + w_f h_{t-1} + b_f)$$

$$i_t = \sigma(W_i x_t + w_i h_{t-1} + b_i)$$

$$O_t = \sigma(W_o x_t + w_o h_{t-1} + b_o)$$

$$C_t = f_t \square C_{t-1} + i_t \square \sigma_c(W_c x_t + w_c h_{t-1} + b_c)$$

$$h_t = O_t \square \tanh(C_t)$$

Where $W$, $w$ and $b$ denote the weights of input, weights of recurrent output, and biases. $f$, $i$ and $O$ represent the forget, input, and output gate vectors, respectively. $\square$ is the element-wise multiplication.

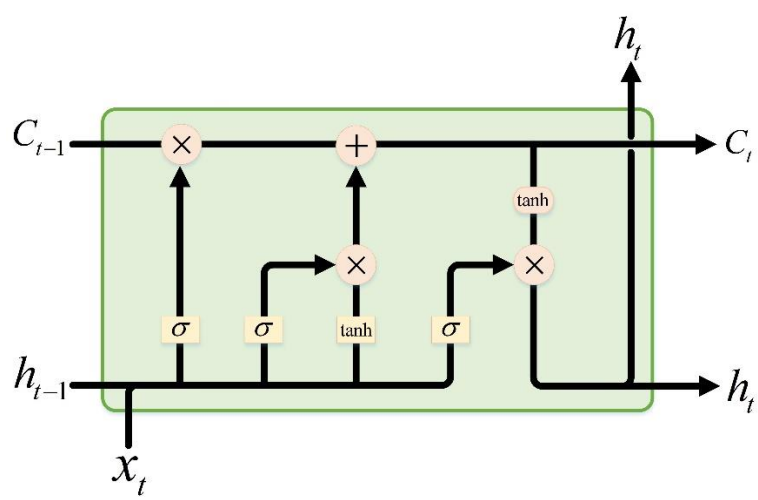



**Figure 4.** The structure of the LSTM network

Relying on LSTM algorithm, either a single long short-term memory (LSTM) or a hybrid model of LSTM, many researchers strived to forecast the stock value. Adapting LSTM algorithm, Moon and Kim [26] propose an algorithm to predict the stock market index and the stock market volatility. Fischer and Krauss [27] expand the LSTM networks to forecast out-of-sample directional movements in the stock market. The comparison between the performance of their model with the random forest (RF), deep neural network (DNN), logistic regression classifier (LOG) illustrates a remarkable outperformance of the LSTM model. Tamura et al. [28] introduce a two-dimensional approach to predict the stock values in which the financial technical indexes of Japanese stoke market entered as input data to the LSTM for the prediction and then the data on financial statements of the related companies are retrieve and added to the database. Wang et al. [29] tried to find the best model to predict financial time-series for portfolio management to optimize the portfolio formation. They compare the results of the LSTM against SVM, RF, DNN, and autoregressive integrated moving average model (ARIMA) and they realized that LSTM is more suitable for financial time-series forecasting in their study. Using LSTM, Fister et al. [30] design a

model for automated stock trading. They argue that the performance of LSTM is remarkably higher than the traditional trading strategies such as passive and rule-based trading strategies, in their case studies the German blue-chip stock and BMW in the period between 2010 and 2018 formed the data sources.

In addition, there are many evidences in the literature that hybrid LSTM methods also outperform the other single DL methods [31]. In the application of stock market, LSTM has been combined with different methods to develop a hybrid model. Tamura et al. [28], for instance, use LSTM to predict stock price and they report that the results of the accuracy test outperform the model in the literature, in their case study. Employing optimal long short-term memory (O-LSTM), Agrawal et al. [32] propose a model for the stock price prediction. Agrawal et al. [32] use correlation-tensor which is formed by stock technical indicators (STIs) to optimize the deep learning function. Agrawal et al. [32] integrate optimized LSTM with STIs and develop two predictive models one for price trend prediction and the other for taking Buy-Sell decision at the end of day. Integrating wavelet transforms (WT), stacked autoencoders (SAEs) and LSTM, Bao, Yue, and Rao [33] propose a new method to predict the stock price. According to Bao et al. [33], WT, firstly, eliminates noises to decompose the stock price time series. In the next stage, predictive features for the stock price are created by SAEs. And finally, the LSTM is applied to predict the next day's closing price based on the features generated through the previous stage. Bao et al. [33] claim that their model outperforms the state-of-the-art models in the literature in terms of predictive accuracy and profitability performance. To cope with non-linearity and non-stationary characteristics of financial time series, Yan and Ouyang [34] integrate wavelet analysis – LSTM (WA-LSTM) to forecast the daily closing price of the Shanghai Composite Index. Their results show that their proposed model outperformed multiple layer perceptron (MLP), SVM and KNN in finding the patterns in the financial time series. Vo et al. [35] use a multivariate bidirectional-LSTM (MB-LSTM) to develop a deep responsible investment portfolio (DRIP) model for the prediction of stock returns for socially responsible investment portfolios. They applied the deep reinforcement learning (DRL) model to retrain neural networks. Fang, Chen, Xue [36] develop a methodology to predict the exchange-trade-fund (EFT) options prices. Integrating LSTM model and support vector regression (SVR), They firstly develop two models of LSTM-SVR I and LSTM-SVR II where in LSTM-SVR I the output of LSTM and the final transaction price, buy price, highest price, lowest price, volume, historical volatility, and the implied volatility of the time segment, that considered as factors affecting the price, added as the input of SVR model. Whilst, in LSTM-SVR II, the hidden state vectors of LSTM and the seven factors affecting the option price considered as the SVR's inputs. They also compare the results with the LSTM model and the RF model.

Table 2. Application of data science algorithms in Stock Market

| Source | Modeling Methods | Data Source | Research Objective |
|---|---|---|---|
| Wang et al. [29] | LSTM Comparing with SVM, RF, DNN, and ARIMAs | Financial Time Series | Portfolio management |
| Lei et al. [37] | time-driven feature-aware and DRL | Financial Time Series | Algorithmic trading |
| Vo et al. [35] | Multivariate Bidirectional LSTM Comparing with DRL | Financial Time Series | Socially Responsible Investment Portfolios |
| Sabeena and Venkata Subba Reddy [38] | GRU – CNN | Financial Time Series | Stock Price Prediction |
| Das and Mishra [39] | Adam optimizer-MDNN | Financial Time Series | Stock Price Prediction |
| Go and Hong [40] | DNN | Financial Time Series | Stock Price Prediction |

| | | | |
|---|---|---|---|
| Agrawal et al. [32] | O-LSTM-STI | Financial Time Series | Stock Price Prediction |
| Gonçalves et al. [41] | CNN comparing with DNNC and LSTM | Financial Time Series | Stock Price Prediction |
| Moews et al. [42] | DNN-SLR | Financial Time Series | Stock Price Prediction |
| Song et al. [43] | DNN | Financial Time Series | Stock Price Prediction |
| Fang et al. [36] | LSTM-SVR comparing with RF and LSTM | Financial Time Series | Exchange-trade-fund (EFT) Options Prices Prediction |
| Long et al. [44] | MFNN (CNN and RNN) | Financial Time Series | Stock Price Prediction |
| Fister et al. [30] | LSTM | Financial Time Series | Automated Stock Trading |
| Rajesh [45] | RF, SVM and KNN | Financial Time Series | Stock Price Prediction |
| Moon and Kim [26] | LSTM | Financial Time Series | Stock Price Prediction |
| Sim Kim, and Ahn [46] | CNN comparing with ANN and SVM | Financial Time Series | Stock Price Prediction |
| Agrawal et al. [47] | LSTM-STIs | Financial Time Series | Stock Price Prediction |
| Tashiro et al. [48] | CNN | Financial Time Series | Stock Price Prediction |
| Sirignano and Cont [49] | LSDL | Financial Time Series | Stock Price Prediction |
| Weng et al. [50] | BRT comparing with NNRE, SVRE, and RFR | Financial Time Series | Stock Price Prediction |
| Preeti et al. [51] | ELM-AE comparing with GARCH, GRNN, MLP, RF and GRDH | Financial Time Series | Stock Price Prediction |
| Sohangir et al. [52] | CNN comparing with doc2vec and LSTM | Social media | Sentiment Analysis |
| Fischer and Krauss [27] | LSTM comparing with RF, DNN and LOG | Financial Time Series | Stock Price Prediction |
| Lien Minh et al. [53] | two-stream GRU | Financial news | Sentiment Analysis |
| Das et al. [54] | DNN | Financial Time Series | The S&P 500 Index Trend Prediction |
| Yan and Ouyang [34] | wavelet analysis with LSTM, comparing with SVM, KNN and MLP | Financial Time Series | Stock Price Prediction |
| Kim et al. [55] | MACN | Financial Time Series | Stock Price Prediction |
| Faghihi-Nezhad and Minaei-Bidgoli [56] | EL-ANN | Financial Time Series | Stock Price Prediction |
| Tamura et al. [28] | LSTM | Financial Time Series | Stock Price Prediction |
| Chong et al. [57] | DNN comparing with PCA, Autoencoder, and RBM | Financial Time Series | Stock Price Prediction |
| Dingli and Fournier | CNN | Financial Time Series | Stock Price Prediction |

| | | | |
|---|---|---|---|
| [58] | | | |
| Singh and Srivastava [59] | (2D)2PCA – DNN comparing with RBFNN | Financial Time Series | Stock Price Prediction |
| Bao et al. [33] | WT-SAEs-LSTM | Financial Time Series | Stock Price Prediction |
| Shekhar and Varshney [60] | GA-SVM | Financial Time Series | Stock Price Prediction |
| Ahmadi et al. [61] | ICA- SVM | Financial Time Series | Stock Price Prediction |
| Ebadati and Mortazavi [62] | GA-ANN | Financial Time Series | Stock Price Prediction |
| Johari et al. [63] | GARCH-SVM | Financial Time Series | Stock Price Prediction |

DNN

Deep neural network (DNN) is the other DL method, which is composed of multiple levels of nonlinear operations, and each layer only receives the connections from its previous training layer (as shown in Figure 5). Suppose $X$ be the input data, and $w_j$ be a filter bank. The multi-layer features of the DNN can be computed as:

$$f(X) = f_k(\cdots f_2(f_1(X; w_1); w_2)\cdots, w_k).$$

Currently, DNN has widely applied in the stock market to identify the trends and patterns among the financial time series data. Go and Hong [40] applied DNN method to predict the stock value. They firstly train the method by the time series data and then test and confirm the predictability of their model. Song et al. [43] develop DNN using 715 novel input-features to forecast the stock price fluctuation. They also compare the performance of their model with the other models that include simple price-based input-features. For predicting the stock market behavior, Chong, Han, and Park [57] examine the performance of DNN. They consider high-frequency intraday stock returns as the input in their model. They analyze the predictability of principal component analysis (PCA), autoencoder, and RBM. According to their results, DNN have good predictability with the information they receive from the residuals of the autoregressive mode. Whilst, applying the autoregressive model to the residuals of the network cannot contribute to the predictability of the model. In addition, Chong et al. [57] found out applying covariance-based market structure analysis to the predictive network remarkably increase the covariance estimation. Das et al. [54] use DNN to predict the future trends of the S&P 500 Index. Their results show that their model can poorly forecast the behavior of the underlying stocks in the S&P 500 index. They believe that randomness and non-stationarity are the reasons making hard the predictability of this index.

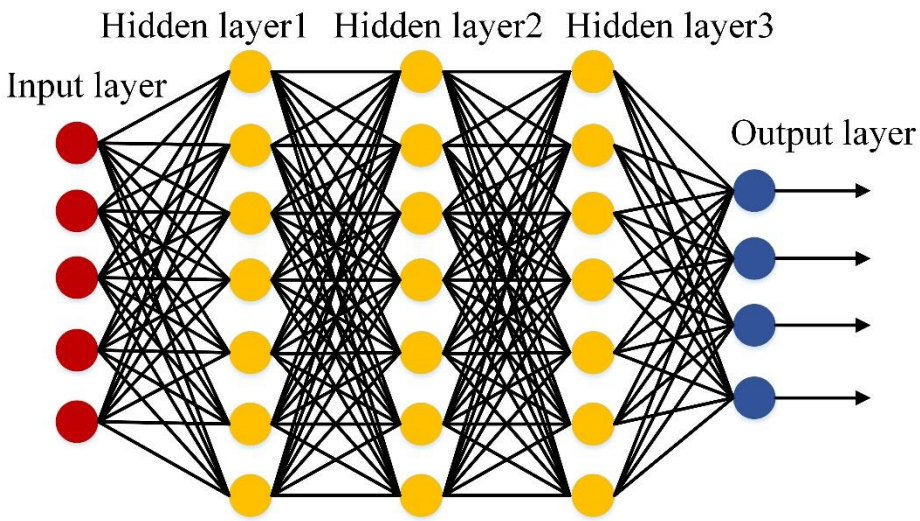



**Figure 5.** Structure of the deep neural network

Besides, hybrid methods that are constructed based on DNN have been reported very accurate in the financial time series data. Das and Mishra [39], for instance, propose an advanced model to plan, analyze and predict the stock value. They use a multilayer deep neural network (MDNN) optimized by Adam optimizer (AO) to find the patterns among the stock values. Moews et al. [42] propose a method to predict the behavior of the stock market, as a complex system with a massive number of noisy time series. Their model integrates DNN and stepwise linear regressions (SLR). Moews et al. [42] consider regression slopes as trend strength indicators for a given time interval. To predict the Google stock price, Singh and Srivastava [59] compare two integrated models 2-directional 2-dimensional principal component analysis-DNN ((2D)2PCA-DNN) and (2D)2PCA-radial basis function neural network (RBFNN). According to their results, (2D)2PCA-DNN model has a higher accuracy in predicting the stock price in their case study. They also compare their results with RNN model, and they reported that the predictability (2D)2PCA-DNN outperforms RNN as well.

CNN

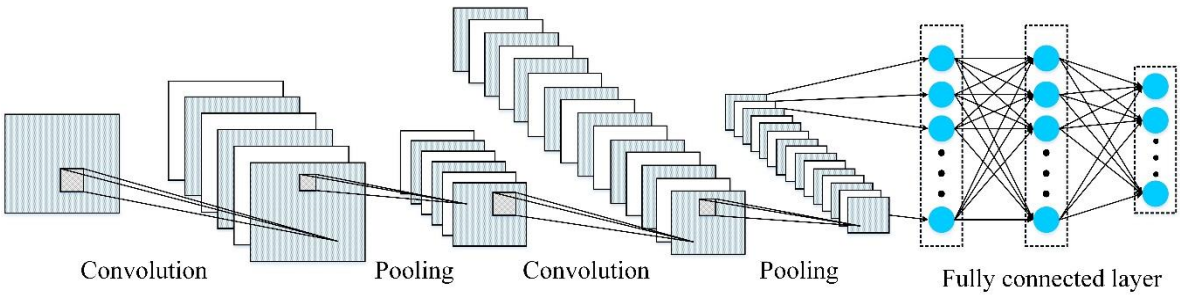



**Figure 6.** Structure of the CNN

Convolutional neural network (CNN) is one of the most popular methods in deep learning, which is widely applied in various fields [64-67], such as classification, language processing, and object detection. A classical structure of CNN is presented in Figure 6, which mainly consists of three components, i.e., convolution layers, pooling layers, and fully connected layers. Different layers have different roles in the training process. Here, those layers are briefly reviewed:

Convolutional layer: Convolutional layer is composed of a set of trainable filters, which is used to perform feature extraction. Suppose $X$ be the input data. There are $k$ filters in convolutional layers, and thus, the output of the convolutional layer can be denoting as follows:

$$y_j = \sum_i f(x_i * w_j + b_j), j = 1, 2, ..., k$$

where $w$ and $b$ are the weight and bias, respectively. $f(\cdot)$ denotes an activation function. * represents the convolution operation.

Pooling layer: In general, the pooling layer is used to decrease the dimensions of the obtained feature data and network parameters. Currently, max pooling and average pooling are the most widely used ways. Let $S$ be a $p \times p$ window size, the average pooling operation can be expressed as follows:

$$z = \frac{1}{N} \sum_{(i,j) \in S} x_{ij}$$

Where $x_{ij}$ indicates the activation value at $(i, j)$. $N$ is the total number of elements in $S$.

Fully connected layer: Following the last pooling layer, the fully connected layer is utilized to reshape the feature maps into 1-D feature vector, which can be expressed as:

$$Y = \sum_i f(WZ + b)$$

Where $Y$ and $Z$ denote the output vector and the input features. $W$ and $b$ represents the weight and bias of fully connected layer.

Recently, CNN has extensively applied by many researchers for predicting the stock values using the financial time series data. Sim et al. [46] develop a model to predict the stock price. They adapt CNN to develop their model. In their study, the result of comparing the predictive performance of CNN with ANN and SVM illustrated that the CNN proved a better performance in forecasting stock price. Tashiro et al. [48] firstly criticize the current models for the price prediction in the stock markets that these models ignore properties of market orders. Therefore, they come up with the CNN architecture integrating order-based features to predict the mid-price trends in the stock markets. Their results prove that adding the features of orders to the model has increased the accuracy of the model. Dingli and Fournier [58] applied the CNN to predict the future movement of stock prices. They report that the predictive accuracy of their model was 65% when they study the next month price and was 60% when they look for the next week price. Gonçalves et al. [41] compare the results of prediction of CNN, LSTM and deep neural network classifier (DNNC) for finding the best model to predict the price trends in the exchange markets. Their finding reveals that CNN, on average, has the best predictive power in their case studies. Sohangir et al. [52] compare the performance of several neural network models like CNN, LSTM, and doc2vec for sentiment analysis among the posts and opinions of the experts in StockTwits to predict the movements in the stock markets. Their results disclose that a CNN has had the highest accuracy to predict sentiment of experts in their case study.

To increase the accuracy of CNN, some researchers integrated CNN with other models and have proposed a new hybrid model. For example, integrating gated recurrent unit (GRU) and CNN, Sabeena et al. [38] introduce a hybrid DL model to predict financial fluctuations in the real-time stock market that is able to process the real-time data from online financial sites. To predict the price movement from financial time series samples, Long et al. [44] introduce an end-to-end model called multi-filters neural network (MFNN). They incorporate CNN and recurrent neural network (RNN) to construct the multi-filters structure of their model.

Other algorithms

In addition to LSTM, DNN, and CNN other DS methods are employed for prediction of stock value using time series data. For example, Sirignano and Cont [49] develop a LSDL model to study the USA market quotes and transactions. Their results disclose that there is a universal and stationary relationship between order flow history and the price trends. Kim et al. [55] develop multi-agent collaborated network (MACN) model to optimize financial time series data. They claim that the contribution of their model is that this model is able to share and generalize the experience by agent in the stock trading.

Besides, various other hybrid methods are applied by the researchers for financial time series data. For example, to predict the stock prices trends, Lien Minh et al. [53] develop the two-stream GRU (TS-GRU) network and Stock2Vec model. They believe that financial news and sentiment dictionary affect the stock prices therefore, their method trained to analyze the sentiments in the financial news and their relationship with the financial prices. Their finding supported the outperformance of their model in compare with the current models. Lien Minh et al. [53] also claim that Stock2Vec is very efficient in financial datasets. Lei et al. [37] with combining deep learning models and reinforcement learning models develop a time-driven feature-aware (TDFA) jointly deep reinforcement learning model (TFJ-DRL) for financial time-series forecasting in algorithmic trading. Preeti et al. [51] introduce an extreme learning machine (ELM)-auto-encoder (AE) model to find the patterns in the financial time series. They test the accuracy of their model by the time series data of gold price and crude oil price and also compare the results with those of generalized autoregressive conditional heteroskedasticity (GARCH), GRNN, MLP, RF and group method of data handling (GRDH). The result of mean square error (MSE) test proved that the performance of their model was higher than existing methods.

In addition to the hybrid deep learning models, four articles applied hybrid machine learning models to financial time series data. Shekhar and Varshney [60] integrate a hybrid model of genetic algorithm-SVM (GV-SVM) with sentiment analysis to predict the future of stock market. Using quantitative empirical analysis, they prove that the combination of sentiment analysis with GV-SVM increased the model's accuracy by 18.6% and reported the final model accuracy of about 89.93%. Ahmadi et al. [61] compare the performance of two hybrid machine learning models in prediction of timing of the stock markets, naming imperialist competition algorithm-SVM (ICA- SVM) and SVM-GA. Their results exposed that SVM-ICA had a higher performance in compared with SVM-GA in prediction of stock market trends for periods of 1–6 day. To predict stock prices using financial time series data, Ebadati and Mortazavi [62] apply a hybrid model by integrating GA-ANN where GA is employed to select ANN features and optimize parameters. Their study shows that this hybrid machine learning model has improved sum square error (SSE) (i.e. performance accuracy) by 99.99% and has improved time (i.e. speed accuracy) by 90.66%. Johari et al. [63] compare the accuracy performance of GARCH-SVM and GARCH-ANN models in the financial time series data for stock price forecasting. They showed that GARCH-SVM outperformed GARCH-ANN, SVM, ANN and, GARCH based on MSE and RMSE accuracy metrics in the prediction of stock price.

Rajesh et al. [45] use ensemble learning to predict the future stock trend. They applied heat map and ensemble learning to financial data of top 500 companies in S&P stock exchange. Their results show that the combination of RF, SVM and K-neighbors classifiers had the most accurate results and the accuracy of the proposed model was 23% higher than a single classifier labelling prediction model. Weng et al. [50] tried to design a financial expert system to forecast short term stock prices. To analyze the collected data and predict the stock prices, they employed four machine learning ensemble methods of neural network regression ensemble (NNRE), support vector regression ensemble (SVRE), boosted regression tree, and random forest regression (RFR). They, using Citi Group stock ($C) data, forecast the 1-day ahead price of 19 stocks from different industries. Weng et al. [50] claimed that boosted regression tree (BRT) outperformed other ensemble models with a considerable mean absolute percent error (MAPE) that was

better than what was reported in the literature. Faghihi-Nezhad and Minaei-Bidgoli [56] using ensemble learning and ANN proposed a two-stage model to predict the stock price. They firstly predict the direction of the next price movement and then they create a new training dataset to forecast the stock price. They use genetic algorithm (GA) optimization and particle swarm optimization (PSO) to optimize the results of each stage. They claimed that the accuracy of their model in prediction of stock price outperformed other models in the literature.

Reviewing the articles categorized in stock market category reveals that although the research objectives of the articles are different, they mainly utilize financial time series data (i.e. 35 out of 37 articles) and only 2 articles have used the financial news and social media as the data source to find the future trends in the stock market (see Table 3).

Table 3. Classification of Articles Using Data Science by Research Purpose and Data Source in the Stock Market Section

| Research Objective | Data Source | Number of documents |
|---|---|---|
| Stock Price Prediction | Financial Time Series | 29 |
| Sentiment Analysis | Financial News, Social Media | 2 |
| Portfolio management | Financial Time Series | 1 |
| Algorithmic trading | Financial Time Series | 1 |
| Socially Responsible Investment Portfolios | Financial Time Series | 1 |
| Automated Stock Trading | Financial Time Series | 1 |
| The S&P 500 Index Trend Prediction | Financial Time Series | 1 |
| Exchange-trade-fund (EFT) Options Prices Prediction | Financial Time Series | 1 |

3.1.2 Marketing

Studying the purpose of the articles disclosed that DS algorithms are used for the purpose of studying customer behavior and promotional activities, that is why these articles are classified in a group labeled marketing. As it is presented in Table 4, two studies have applied a single DL method and three of them used hybrid DL methods. In addition, these studies use various data sources such as customer time series data, case studies, and social media. For example, Paolanti et al. [68] employ deep convolutional neural network (DCNN) to develop a mobile robot, so called ROCKy, to analyze real-time store heat maps of retail store shelves for detection of shelf out of stock (SOOS) and promotional activities, during working hours. Dingli, Marmara, and Fournier [69] were looking for solutions to find the patterns and features among transactional data to predict customer churn within the retail industry. To do so, they compare the performance of CNN and restricted Boltzmann machine (RBM), and they found out the RBM attained outperformed in customer churn prediction.

Table 4. Application of data science algorithms in Marketing

| Source | Modeling Methods | Data Source | Research Objective |
|---|---|---|---|
| Ładyżyński et al. [70] | RF – DNN | Time Series data of Customers | Customer Behavior |
| Ullah et al. [71] | RF | Time Series data of Customers | Customer Behavior |
| Paolanti et al. [68] | DCNN | Primary Data | Detection of Shelf Out |

| | | | |
|---|---|---|---|
| | | | of Stock (SOOS) and Promotional Activities |
| Agarwal [72] | RNNs- CNNs | Social media | Sentiment Analysis |
| Shamshirband et al. [73] | SN-CFM | Social media | Customer behavior |
| Dingli et al. [69] | RBM | Primary Data | Customer behavior |

On the other hand, RF-DNN, RNN-CNN and similarity, neighborhood-based collaborative filtering model (SN-CFM) are hybrid models that researchers proposed to study the customer behavior. Ładyżyński et al. [70], for instance, using Random Forest (RF) and DNN methods and customers' historical transactional data propose a hybrid model that is able to predict the customers' willingness to purchase credit products from the banks. Ullah et al. [71] use the RF algorithm to prediction churn customers and use their result to formulate strategies for customer relationship management to prevent churners. Ullah et al. [71] explain that combination of churn classification utilizing the RF algorithm and customer profiling using k-means clustering increased their model performance. Agarwal [72] integrated RNN and CNN to develop a model for sentiment analysis. According to Agarwal [72], sentiment analysis is the best approach to get the customers' feedback. He tested his proposed model using the social media data and believes that the result of the sentiment analysis provides guidance to the business to improve the quality of their service and presents evidences for the startups to improve customer experience. Shamshoddin et al. [73] propose SN-CFM to predict the consumer preferences according to similarity of features of users and products that are acquired from the internet of things and social networks.

### 3.1.3 E-commerce

Another category emerged after reviewing the articles is labeled e-commerce where the authors employed data science models to address problems in e-commerce sector. A summary of these studies is presented in Table 5. Lei [74], based on the GRU model, designs a neural network investment quality evaluation model to support the decision-making related to investment in e-commerce. Their proposed model is able to evaluate different index data that provides a better picture to investors. Leung et al. [75] argue that the ability to handle orders and logistics management is one of the major challenges in e-commerce. Therefore, using hybrid autoregressive-adaptive neuro-fuzzy inference system (AR-ANFIS), Leung et al. [75] develop a prediction model for B2C e-commerce order arrival. According to their results, their proposed hybrid model has been successful to forecast e-order arrivals. Cai et al. [76] use deep reinforcement learning to develop an algorithm to address allocating impression problems in e-commerce websites, such as www.taobao.com, www.ebay.com, and www.amazon.com. In this algorithm the buyers are allocated to the sellers based on the buyer's impressions and the seller's strategies in a way that to maximize the income of the platform. To do so, they applied GRU and their finding shows that GRU outperform the deep deterministic policy gradient (DDPG). Ha, Pyo, Kim [77] apply RNN to develop deep categorization network (Deep CN) for item categorization in e-commerce. Item categorization refers to classifying the leaf category of items from their metadata. They use RNN to generate features from text metadata and categorizing the items accordingly. Xu et al. [78] design an advanced credit risk evaluation system for e-commerce platforms to minimize the transaction risk associated with buyers and sellers. To this end, they use a hybrid machine leaning model combined with decision tree—ANN (DT-ANN) and the results of their study show that this model has a high accuracy and outperform other hybrid machine learning models such as DT-Logistic Regression and DR-dynamic Bayesian network.

Selling products online has unique challenges, and data science has been able to provide solutions to address some of these challenges. In order to increase the buyer's trust in the quality of the products they are going to buy online, Saravanan and Charanya [79] design an algorithm by which the products are

categorized according to several criteria, including reviews and ratings of other users. Integrating a hybrid feature extraction method principle component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) with SVM using lexicon-based method, Saravanan and Charanya [79] propose a model that separates the products from the large collection of different products based on characteristics, the best product ratings and positive reviews. Wang, Mo, and Tseng. [80] use RNN to develop a personalized product recommendation system on e-commerce websites. The result of testing their model discloses the outperformance of RNN to K nearest neighbors (KNN). Wu and Yan [81] claim that the main assumption of the current production recommender models for e-commerce websites is that all the historical data of users are recorded while in practice many platforms fail to record such dada. Therefore, they come up with a list-wise DNN (LWDNN) to model the temporal online user behaviors and provide recommendations for the anonymous users.

Table 5. Application of data science algorithms in E-commerce

| Source | Modeling Methods | Data Source | Research Objective |
|---|---|---|---|
| Lei [74] | GRU | Financial Time Series | Investment Quality Evaluation Model |
| Leung et al. [75] | AR-ANFIS | Primary Data | Order Arrival Prediction |
| Cai et al. [76] | GRU | Customers Time Series | Impression Allocation Problem |
| Ha et al. [77] | RNN | Primary Data | Item Categorization |
| Xu et al. [78] | DT—ANN | Credit Data | Dynamic Credit Risk Evaluation |
| Saravanan and Charanya [79] | PCA- t-SNE-SVM | Primary Data | Product Recommendation |
| Wang et al. [80] | RNN | Primary Data | Product Recommendation |
| Wu and Yan [81] | LWDNN | Customers Time Series | Product Recommendation |

3.1.4 Cryptocurrency

The decision-making process related to investing on the cryptocurrencies is similar to investing to the stock market where the prediction of future value is very determinant and effective on the investment decisions. Applying machine learning and DL models to predict the trends of cryptocurrencies' prices is an attractive research problem which is emerging in the literature (see Table 6). Lahmiri and Bekiros [82], for example, applied deep learning methods for the prediction of price of cryptocurrencies including Bitcoin, Digital Cash and Ripple. They compare the predictive performance of LSTM and GRNN. Their finding discloses that the LSTM model has a better performance in their case studies in compare with GRNN. Altan, Karasu, and Bekiros [83] claim that integrating LSTM and empirical wavelet transform (EWT) improves the performance of LSTM in forecasting the digital currency price. In this study, Altan et al. (2019) test their proposed model using the Bitcoin, Ripple, Digital Cash and Litecoin time series data. Jiang and Liang [84] develop a CNN model to predict the price of Bitcoin as a cryptocurrency example. They train their proposed model by historical data of financial assets prices and output of their model was designed to be portfolio weights of the set.

Table 6. Application of data science algorithms in Cryptocurrency

| Source | Modeling Methods | Data Source | Research Objective |
|---|---|---|---|
| Lahmiri and | LSTM comparing | Financial Time Series | Cryptocurrencies |

| | | | |
|---|---|---|---|
| Bekiros [82] | with GRNN | | Price prediction Cryptocurrencies |
| Altana et al. [83] | LSTM-EWT | Financial Time Series | Price prediction Cryptocurrencies |
| Jiang and Liang [84] | CNN | Financial Time Series | Price prediction |

3.1.5 Corporate bankruptcy prediction

Corporate bankruptcy prediction has become an important tool to evaluate the future financial situation of the companies. Utilizing machine learning based methods is widely recommended to address bankruptcy prediction problems. To address corporate bankruptcy prediction, Chen, Chen, Shi [85] utilize bagging and boosting ensemble strategies and develop two models of Bagged- proportion support vector machines (pSVM) and Boosted-pSVM. Using datasets of UCI and LibSVM, they test their models and explain that ensemble learning strategies increased the performance of the models in of bankruptcy prediction. Lin, Lu, and Tsai [86] believe that finding the best match of feature selection and classification techniques improves the prediction performance of bankruptcy prediction models. Their results reveal that the genetic algorithm as the wrapper-based feature selection method and the combination of the genetic algorithm with the naïve Bayes and support vector machine classifiers had a remarkable predictive performance. Lahmiri et al. [87], to develop an accurate model for forecasting corporate bankruptcy, compare the performance of different ensemble systems of AdaBoost, LogitBoost, RUSBoost, subspace, and bagging. Their finding reveals that that AdaBoost model has been effective in terms of short time data processing and low classification error, and limited complexity. Faris et al. [88] investigate the combination of re-sampling (oversampling) techniques and multiple features election methods to improve the accuracy of bankruptcy prediction methods. According to their results, employing SMOTE oversampling technique and AdaBoost ensemble method using reduced error pruning (REP) tree provides reliable promising results to bankruptcy prediction. A summary of these research articles is presented in Table 7.

Table 7. Application of data science algorithms in corporate bankruptcy prediction

| Source | Modeling Methods | Data Source | Research Objective |
|---|---|---|---|
| Chen et al. [85] | Bagged-pSVM and Boosted-pSVM | UCI and LibSVM datasets | bankruptcy prediction |
| Lin et al. [86] | Genetic Algorithm with the Naïve Bayes and SVM classifiers | Australian credit, German credit, and Taiwan bankruptcy datasets | bankruptcy prediction |
| Lahmiri et al. [87] | AdaBoost | University of California Irvine (UCI) Machine Learning Repository | bankruptcy prediction |
| Faris et al. [88] | SMOTE- AdaBoost- REP Tree | Infotel database | bankruptcy prediction |

*3.2. Applied data science methods in the Economics*

3.2.1 Deep learning methods

DL models are structured based on the artificial neural network and the modern algorithms of deep learning rooted in the work of Hinton et al. [89]. Hinton et al. [89] develop, for the first time, a two-step approach in which the deep learning algorithms are firstly trained and then fine-tune the model in a back-through process. The advantage of DL models compared to other ML models is that DL models can effectively identify high-level features and representations (the outputs) from a large diverse data sample (inputs). Employing an unsupervised pre-training and a supervised fine-tuning approach, the deep learning models extract hierarchical features from the inputs to classify the patterns of the data [90]. **The ability of DL algorithms in prediction and finding the patterns among the raw data has grabbed the attraction of many researchers from various fields. Economics researchers have applied DL** models **for a variety of reasons such as stock price prediction (e.g. [26-28]) and forecasting consumer behavior (e.g.[69]). LSTM, CNN, and DNN are respectively the most applied DL models among the database of the study. LSTM is applied to stock price prediction [26-28], portfolio management** [29]**, automated stock trading** [30], and cryptocurrencies price prediction [82]. Among the reviewed papers, the LSTM method has only applied to find the patterns among financial time series data.

Similar to LSTM, the CNN algorithm is applied mainly for financial time series data to stock price prediction [41, 46, 48, 58] and cryptocurrencies price prediction [84]. CNN algorithm is also used for analyzing **social media** data for the purpose of **sentiment analysis** [52]. The DNN algorithm, likewise the LSTM, is only used to analyze financial time series data to predict stock prices [40, 43, 57] **and the S&P 500 Index trend prediction** [54]. The GRU algorithm, which is another DL model, is applied in the e-commerce section to analyze financial time series [74] and customers time series [76]. RNN is a DL algorithm applied to analyze primary data for item categorization [77] and product recommendation [80]. Large-scale deep learning (LSDL) and MACN are DL algorithms that are used to analyze financial time series data to predict stock price [49, 55]. Ultimately, it is found that DCNN and RBM applied for analyzing primary data for respectively promotional activities [68] and customer behavior forecasting [69] (see Table 8).

Table 8. List of single deep learning methods employed in Economics related fields

| Method | Applications | | | |
|---|---|---|---|---|
| | **Stock Market** | **Marketing** | **Cryptocurrency** | **E-Commerce** |
| LSTM | Moon and Kim [26], Fischer and Krauss [27], Tamura et al. [28], Wang et al. [29], Fister et al. [30], | - | Lahmiri and Bekiros [82] | - |
| CNN | Gonçalves et al. [41], Sim Kim, and Ahn [46], Tashiro et al. [48], Sohangir et al. [52], Dingli and Fournier [58], | - | Jiang and Liang [84] | - |
| DNN | Go and Hong [40], Song et al. [43], Das et al. [54], Chong et al. [57] | - | - | - |
| GRU | - | - | - | Lei [74], Cai et al. [76] |
| RNN | - | - | - | Ha et al. [77], Wang et al. [80] |
| LSDL | Sirignano and Cont [49] | - | - | - |
| MACN | Kim et al. [55] | - | - | - |
| DCNN | - | Paolanti et al. [68] | - | - |

| | | | | |
|---|---|---|---|---|
| RBM | - | Dingli et al. [69] | - | - |

3.2.2 Hybrid deep learning methods

Hybrid deep neural networks are architectures that apply generative and discriminative components at the same time. Hybrid models combines machine learning models or combine a machine learning model with an optimization model to improve the predictivity of the deep learning model [6]. The findings of the literature review in the study reveal that hybrid deep learning (HDL) models are widely applied in the field of economics. On the other hand, Figure 8 illustrates that the accuracy of such models is reported higher than the single DL models. The various HDL models used among the papers reviewed in this study are summarized in Table 9.

Table 9. List of hybrid deep learning models employed in Economic related fields

| Application | The hybrid method | Source |
|---|---|---|
| Stock Market | TDFA-DRL | Lei et al. [37] |
| | MB-LSTM | Vo et al. [35] |
| | GRU – CNN | Sabeena and Venkata Subba Reddy [38] |
| | AO-MDNN | Das and Mishra [39] |
| | OLSTM-STI | Agrawal et al. [32] |
| | DNN-SLR | Moews et al. [42] |
| | LSTM-SVR | Fang et al. [36] |
| | MFNN (CNN and RNN) | Long et al. [44] |
| | LSTM-STIs | Agrawal et al. [47] |
| | ELM-AE | Preeti et al. [51] |
| | TS-GRU | Lien Minh et al. [53] |
| | WA-LSTM | Yan and Ouyang [34] |
| | (2D)2PCA – DNN | Singh and Srivastava [59] |
| | WT-SAEs-LSTM | Bao et al. [33] |
| | RNNs- CNNs | Agarwal [72] |
| | SN-CFM | Shamshoddin, et al. [73] |
| E-commerce | LWDNN | Wu and Yan [81] |
| Cryptocurrency | LSTM-EWT | Altan et al. [83] |

3.2.3 Hybrid machine learning methods

Hybrid models combines machine learning models or combine a machine learning model with an optimization model to improve the predictivity of the machine learning model. Hybrid machine learning models are frequently employed among the reviewed articles in this study. Table 10 illustrates that 7 hybrid machine learning models are mainly proposed to stock price predictions (i.e. 4 out of 7) and three models have been developed to address problems related to e-commerce, namely order arrival prediction, dynamic credit risk evaluation, and product recommendation.

Table 10. List of hybrid machine learning models employed in Economic related fields

| Application | The hybrid method | Source |
|---|---|---|
| Stock Market | GA-SVM | Shekhar and Varshney 2020 |
| | ICA- SVM | Ahmadi et al. 2018 |
| | GA-ANN | Ebadati and Mortazavi 2018 |

| | | |
|---|---|---|
| | GARCH-SVM | Johari et al. 2018 |
| | AR-ANFIS | Leung et al. 2019 |
| E-commerce | DT—ANN | Xu et al. 2019 |
| | PCA- t-SNE-SVM | Saravanan and Charanya 2019 |

3.2.4 Ensemble machine learning algorithms

Ensemble machine learning algorithms or ensemble learning (EL) models use multiple learning algorithms to improve training processes and boost learning from data [4]. There are many evidences in the literature that ensemble classifiers and ensemble learning systems have been very effective in financial time series data. For instance, Rajesh et al. [45] combine of RF, SVM and K-neighbors classifiers, Weng et al. [50] use neural network regression ensemble, and Faghihi-Nezhad and Minaei-Bidgoli [56] apply a ANN-EL to predict stock values. Besides, Chen et al. [85], Lin et al. [86], Lahmiri et al. [87] respectively applied Bagged-pSVM and Boosted-pSVM models, genetic algorithm with the naïve Bayes and SVM, SMOTE- AdaBoost- REP Tree to predict corporate bankruptcy. Ullah et al. [71] also take advantage of EL strategies and use a RF algorithm to predict the churn customers. Table 11 summers these studies which have used an ensemble model.

Table 11. List of ensemble models applied in the database on the current study

| Application | The ensemble method | Source |
|---|---|---|
| Stock Market | RF-SVM-K-neighbors | Rajesh et al. [45] |
| | NNRE | Weng et al. [50] |
| | ANN-EL | Faghihi-Nezhad and Minaei-Bidgoli [56] |
| Corporate Bankruptcy | Bagged-pSVM and Boosted-pSVM | Chen et al. [85] |
| | Genetic Algorithm with the Naïve Bayes and SVM | Lin et al. [86] |
| | SMOTE- AdaBoost- REP Tree | Lahmiri et al. [87] |
| Marketing | RF – DNN | Ładyżyński et al. [70] |
| | RF | Ullah et al. [71] |

*3.3. Discussions on the taxonomy of data science advancements in Economics*

One of the contributions of this study is presenting the taxonomy of data science methods and applications in the economics. Indeed, the output of reviewing the literature is presented in Figure 7. This study presents the advancement of deep learning models, hybrid deep learning models, hybrid machine learning models, and ensemble machine learning models in economics. It is figured out that overall, 42 unique algorithms employed among 57 reviewed articles in which 9 of them used single DL models (see Table 8), 18 hybrid DL models (see Table 9), 7 hybrid ML (see Table 10), and 8 ensemble models (see table 11). It has also been revealed that these 42 models have been used for different purposes, including algorithmic trading, automated stock trading, bankruptcy prediction, cryptocurrencies price prediction, customer behavior, shelf out of stock, impression allocation problem, investment quality evaluation model, item categorization, portfolio management, product recommendation, sentiment analysis, socially responsible investment portfolios, stock price prediction. It is also disclosed that data science algorithms mainly applied for financial time series data to forecast stock prices in which LSTM model has been the most popular model for analyzing financial time series. CNN and DNN have been respectively the most applied algorithms among the reviewed articles in this study.

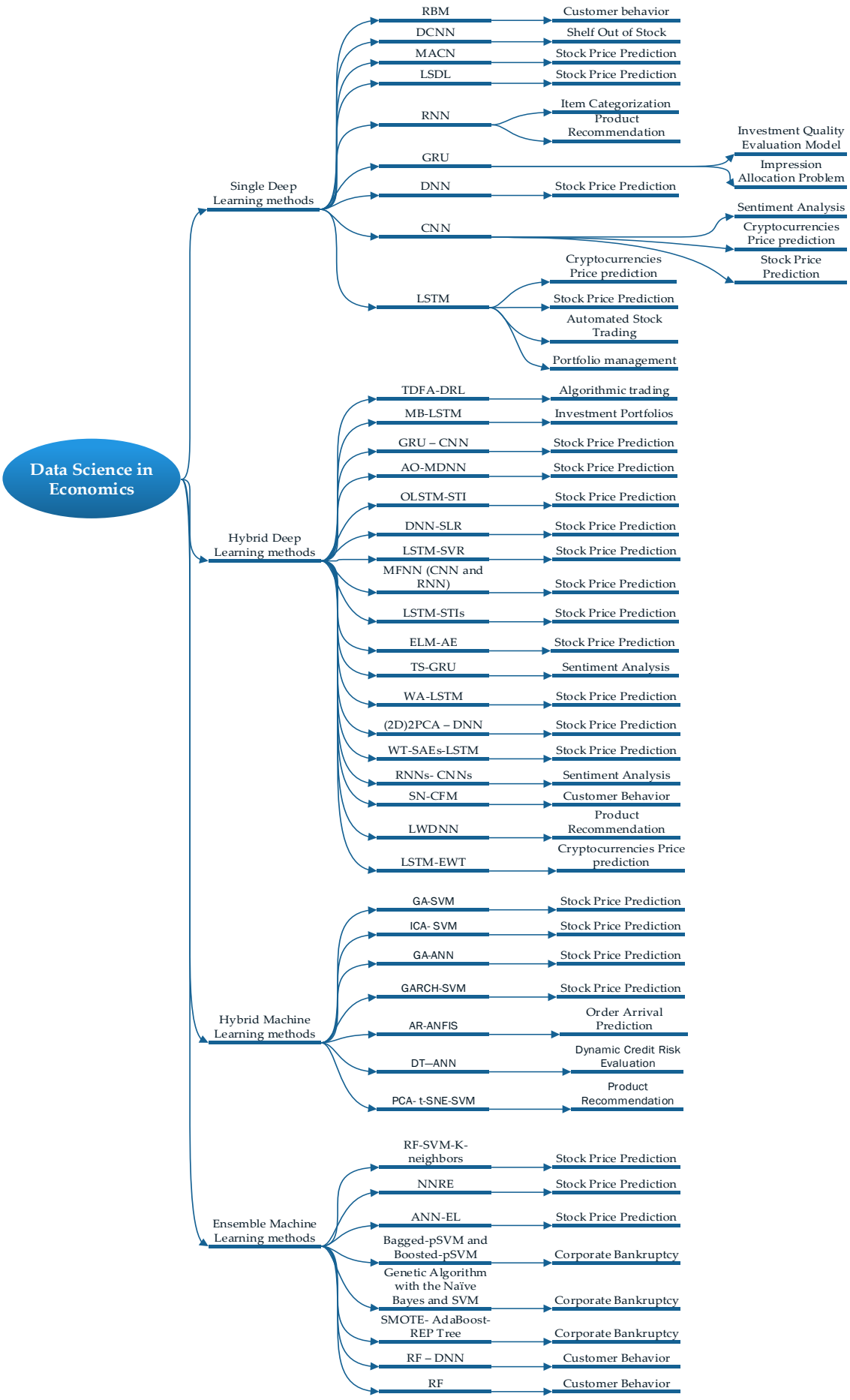

Data Science in Economics
Single Deep Learning methods
RBM
Customer behavior
DCNN
Shelf Out of Stock
MACN
Stock Price Prediction
LSDL
Stock Price Prediction
RNN
Item Categorization
Product Recommendation
GRU
Investment Quality Evaluation Model
Impression Allocation Problem
DNN
Stock Price Prediction
CNN
Sentiment Analysis
Cryptocurrencies Price prediction
Stock Price Prediction
LSTM
Cryptocurrencies Price prediction
Stock Price Prediction
Automated Stock Trading
Portfolio management
Hybrid Deep Learning methods
TDFA-DRL
Algorithmic trading
MB-LSTM
Investment Portfolios
GRU – CNN
Stock Price Prediction
AO-MDNN
Stock Price Prediction
OLSTM-STI
Stock Price Prediction
DNN-SLR
Stock Price Prediction
LSTM-SVR
Stock Price Prediction
MFNN (CNN and RNN)
Stock Price Prediction
LSTM-STIs
Stock Price Prediction
ELM-AE
Stock Price Prediction
TS-GRU
Sentiment Analysis
WA-LSTM
Stock Price Prediction
(2D)2PCA – DNN
Stock Price Prediction
WT-SAEs-LSTM
Stock Price Prediction
RNNs- CNNs
Sentiment Analysis
SN-CFM
Customer Behavior
LWDNN
Product Recommendation
LSTM-EWT
Cryptocurrencies Price prediction
Hybrid Machine Learning methods
GA-SVM
Stock Price Prediction
ICA- SVM
Stock Price Prediction
GA-ANN
Stock Price Prediction
GARCH-SVM
Stock Price Prediction
AR-ANFIS
Order Arrival Prediction
DT–ANN
Dynamic Credit Risk Evaluation
PCA- t-SNE-SVM
Product Recommendation
Ensemble Machine Learning methods
RF-SVM-K-neighbors
Stock Price Prediction
NNRE
Stock Price Prediction
ANN-EL
Stock Price Prediction
Bagged-pSVM and Boosted-pSVM
Corporate Bankruptcy
Genetic Algorithm with the Naïve Bayes and SVM
Corporate Bankruptcy
SMOTE- AdaBoost-REP Tree
Corporate Bankruptcy
RF – DNN
Customer Behavior
RF
Customer Behavior

**Figure 7.** Taxonomy of data science advancements in Economics

Twenty-five out of 42 models presented in Figure 6 are hybrid models that show the trend of using hybrid models in the economy. This may be because the predictive power of hybrid models is higher than that of single DL models. Therefore, the reported root-mean-squared error (RMSE), which is one of the accuracy metrics of machine learning models, for all these models is reviewed and presented in Figure 8. Lower the RMSE, higher the accuracy of the model [7]. According to Figure 8, the value of RMSE for the single DL models have been higher than the hybrid DL models that indicates the higher accuracy of hybrid DL models. This finding is consistent with the claims made in the literature (e.g. [6-8]) and it can also explain why the use of hybrid models has become trendy.

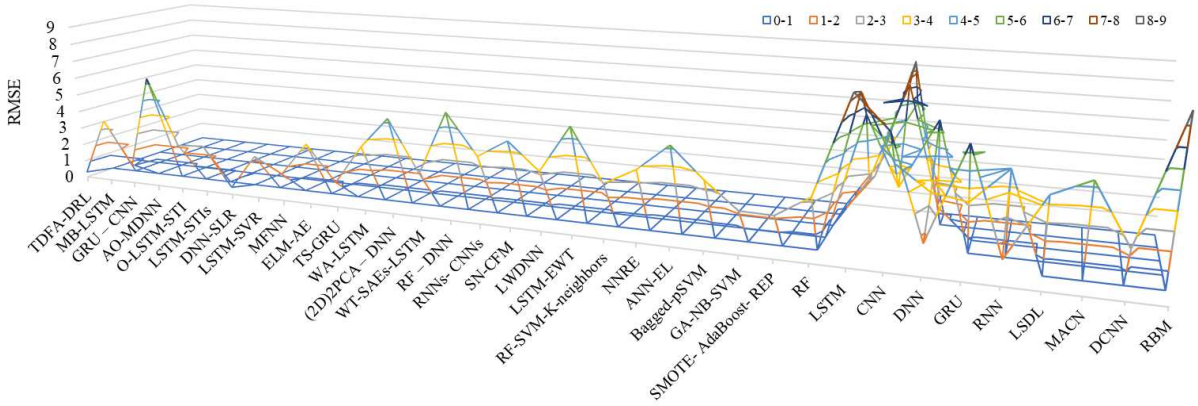



**Figure 8.** Comparison of root-mean-squared error values of hybrid deep learning models and deep learning

The use of data science to address economics-related issues is increasing dramatically. Because data science has made significant advances in providing solutions to these problems. Providing the state-of-the-art of data science methods depicts a clear image of the advancements of data science in the economics-related fields. Therefore, the main objective of this study was to investigate the advances of data science in economics through a novel taxonomy of applications and methods. To that end, the Prisma method, a systematic literature review methodology, applied to survey all the publications indexed in the Scopus and **Thomson Reuters Web-of-Science** databases. The review is limited to published documents written in English and used data science techniques to solve problems in one of the economics-related fields. Finally, fifty-seven papers were selected for further consideration. The finding revealed that five economics-related fields, namely stock market, marketing, E-commerce, corporate banking, and cryptocurrency, utilized data science solutions. On the other hand, it was found that deep learning models, hybrid deep learning models, hybrid machine learning models, and ensemble machine learning models were used in each of these applications. It disclosed that the trends are on advancement of hybrid models as 25 out of 42 models used among the reviewed articles have been hybrid models. According to Root-mean-square error accuracy metric, hybrid models had higher prediction accuracy than other algorithms, this can justify the prevalence of the use of hybrid models. The findings of this study are confined to these five applications related to the field of economics and cannot be generalized to other fields or disciplines. This research provides insights into the advancements of data science models and the application of these models in the field of economics, which provides guidance for researchers and practitioners in the field.

This research presents the state-of-the-art of advancements in data science technologies and methods and their application in economics. The findings of this study provide guidance to researchers and

practitioners in this field who, after familiarizing themselves with the latest developments in data science models, can select the appropriate model according to their application.

## 4. Conclusion

The use of data science techniques in economics and related fields is increasing day by day. The current study provides the latest developments in data science methods and their application in economics. The findings of this study show that, firstly, the advancement of data science algorithms was in four types of methods, constituting deep learning models, ensemble models, hybrid deep learning models, and hybrid machine learning models. These four type models were investigated in five different areas, namely 1) stock market, in which stock price prediction was the main goal of most of the papers, 2) marketing, in which the objectives of the papers were mostly to study the customer behavior, 3) corporate bankruptcy, 4) cryptocurrency, which is a new trend in the economic field where the researchers try to predict the price of digital moneys, and 5) e-commerce, in which the DS methods are applied mainly to increase the performance of e-commerce websites for example by item categorization and product recommendation. This study elaborately provides the algorithms of deep learning models used in the economics, data source, and purpose of data science use in each category. Findings reveal that the use of hybrid models has increased due to the higher prediction accuracy than single deep learning models. LSTM, CNN, and DNN models have been respectively the most applied models in the literature to analyze financial time series data and predict stock price. Fifty-seven articles were reviewed in this study and it was found that these fifty-seven articles used forty-two unique models. Twenty-five out of forty-two models were found to be hybrid models which illustrates the trend of using such models in the economics. The results of comparing RMSE values disclosed that hybrid models show a remarkable low error level, indicating that such models have higher prediction accuracy. Therefore, it is recommended to apply hybrid models to model and optimize objectives in all the five fields of studied in this research. This study provides state-of-the-art data science models in economics and generates insights for researchers and practitioners to choose the most appropriate model for their applications.

**Acronyms used in this study**

| Acronym | Explanation |
|---|---|
| (2D)2PCA | 2-Directional 2-Dimensional Principal Component Analysis |
| AE | Auto-Encoder |
| ANN | Artificial Neural Network |
| AO | Adam Optimizer |
| AR-ANFIS | Auto Regressive-Adaptive Neuro-Fuzzy Inference System |
| ARIMA | Autoregressive Integrated Moving Average Model |
| BRT | Boosted Regression Tree |
| CNN | Convolutional Neural Network |
| DCNN | Deep Convolutional Neural Network |
| DL | Deep Learning |
| DNN | Deep Neural Network |
| DNNC | Deep Neural Network Classifier |
| DRL | Deep Reinforcement Learning |
| DS | Data Science |
| DT—ANN | Decision Tree—Artificial Neural Network |
| EL | Ensemble Learning |

| | |
|---|---|
| ELM | Extreme Learning Machine |
| EWT | Empirical Wavelet Transform |
| GA-ANN | Genetic Algorithm—Artificial Neural Network |
| GARCH | Generalized Autoregressive Conditional Heteroskedasticity |
| GRDH | Group Method of Data Handling |
| GRNN | Generalized Regression Neural Networks |
| GRU | Gated Recurrent Unit |
| ICA | Imperialist Competition Algorithm |
| KNN | K-Nearest Neighbors |
| LOG | Logistic Regression Classifier |
| LSDL | Large-Scale Deep Learning |
| LSTM | Long Short-Term Memory |
| LWDNN | List-Wise Deep Neural Network |
| MACN | Multi-Agent Collaborated Network |
| MB-LSTM | Multivariate Bidirectional LSTM |
| MDNN | Multilayer Deep Neural Network |
| MFNN | Multi-Filters Neural Network |
| ML | Machine Learning |
| MLP | Multiple Layer Perceptron |
| NNRE | Neural Network Regression Ensemble |
| O-LSRM | Optimal Long Short-Term Memory |
| PCA | Principal Component Analysis |
| pSVM | Proportion Support Vector Machines |
| RBFNN | Radial Basis Function Neural Network |
| RBM | Restricted Boltzmann Machine |
| REP | Reduced Error Pruning |
| RF | Random Forest |
| RFR | Random Forest Regression |
| RNN | Recurrent Neural Network |
| SAE | Stacked Autoencoders |
| SLR | Stepwise Linear Regressions |
| SN-CFM | Similarity, Neighborhood-Based Collaborative Filtering Model |
| STI | Stock Technical Indicators |
| SVM | Support Vector Machine |
| SVR | Support Vector Regression |
| SVRE | Support Vector Regression Ensemble, |
| TDFA | Time-Driven Feature-Aware |
| TS-GRU | Two-Stream GRU |
| t-SNE | t-distributed stochastic neighbor embedding |
| WA | Wavelet Analysis |
| WT | Wavelet Transforms |

**Author Contributions:** Conceptualization, A.M.; investigation, S.N.; writing—original draft preparation, S.N.;

methodology, A.M.; writing—review and editing, A.M. and P.G.; visualization, S.N; validation, A.M.; supervision and controlling the results, P.G.

**Funding:**

**Acknowledgments:**

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Mittal, Shachi, Catalin Stoean, Andre Kajdacsy-Balla, and Rohit Bhargava. "Digital Assessment of Stained Breast Tissue Images for Comprehensive Tumor and Microenvironment Analysis." *Frontiers in bioengineering and biotechnology* 7 (2019): 246.
2. Stoean, Ruxandra, Dominic Iliescu, and Catalin Stoean. "Segmentation of Points of Interest During Fetal Cardiac Assesment in the First Trimester from Color Doppler Ultrasound." *arXiv preprint arXiv:1909.11903* (2019).
3. Stoean, Ruxandra, Catalin Stoean, Adriana Samide, and Gonzalo Joya. "Convolutional Neural Network Learning Versus Traditional Segmentation for the Approximation of the Degree of Defective Surface in Titanium for Implantable Medical Devices." Paper presented at the International Work-Conference on Artificial Neural Networks 2019.
4. Ardabili, Sina, Amir Mosavi, and Annamária R Várkonyi-Kóczy. "Advances in Machine Learning Modeling Reviewing Hybrid and Ensemble Methods." Paper presented at the International Conference on Global Research and Education 2019.
5. Torabi, Mehrnoosh, Sattar Hashemi, Mahmoud Reza Saybani, Shahaboddin Shamshirband, and Amir Mosavi. "A Hybrid Clustering and Classification Technique for Forecasting Short-Term Energy Consumption." *Environmental Progress & Sustainable Energy* 38, no. 1 (2019): 66-76.
6. Mosavi, Amir, and Mohammad Edalatifar. "A Hybrid Neuro-Fuzzy Algorithm for Prediction of Reference Evapotranspiration." Paper presented at the International Conference on Global Research and Education 2018.
7. Mosavi, Amir, Shahaboddin Shamshirband, Ely Salwana, Kwok-wing Chau, and Joseph HM Tah. "Prediction of Multi-Inputs Bubble Column Reactor Using a Novel Hybrid Model of Computational Fluid Dynamics and Machine Learning." *Engineering Applications of Computational Fluid Mechanics* 13, no. 1 (2019): 482-92.
8. Torabi, Mehrnoosh, Amir Mosavi, Pinar Ozturk, Annamaria Varkonyi-Koczy, and Vajda Istvan. "A Hybrid Machine Learning Approach for Daily Prediction of Solar Radiation." Paper presented at the International Conference on Global Research and Education 2018.
9. Lee, H., G. Li, A. Rai, and A. Chattopadhyay. "Real-Time Anomaly Detection Framework Using a Support Vector Regression for the Safety Monitoring of Commercial Aircraft." *Advanced Engineering Informatics* 44 (2020).
10. Husejinović, A. "Credit Card Fraud Detection Using Naive Bayesian and C4.5 Decision Tree Classifiers." *Periodicals of Engineering and Natural Sciences* 8, no. 1 (2020): 1-5.
11. Zhang, Y. "Application of Improved Bp Neural Network Based on E-Commerce Supply Chain Network Data in the Forecast of Aquatic Product Export Volume." *Cognitive Systems Research* 57 (2019): 228-35.
12. Sundar, G., and K. Satyanarayana. "Multi Layer Feed Forward Neural Network Knowledge Base to Future Stock Market Prediction." *International Journal of Innovative Technology and Exploring Engineering* 8, no. 11 Special Issue (2019): 1061-75.
13. Hew, J. J., L. Y. Leong, G. W. H. Tan, K. B. Ooi, and V. H. Lee. "The Age of Mobile Social Commerce: An Artificial Neural Network Analysis on Its Resistances." *Technological Forecasting and Social Change* 144 (2019): 311-24.
14. Abdillah, Y., and Suharjito. "Failure Prediction of E-Banking Application System Using Adaptive Neuro Fuzzy Inference System (Anfis)." *International Journal of Electrical and Computer Engineering* 9, no. 1 (2019): 667-75.

15. Sabaitytė, J., V. Davidavičienė, J. Straková, and J. Raudeliūnienė. "Decision Tree Modelling of E-Consumers' Preferences for Internet Marketing Communication Tools During Browsing." *E a M: Ekonomie a Management* 22, no. 1 (2019): 206-24.
16. Fujiyoshi, Hironobu, Tsubasa Hirakawa, and Takayoshi Yamashita. "Deep Learning-Based Image Recognition for Autonomous Driving." *IATSS Research* (2019).
17. Mathis, Mackenzie Weygandt, and Alexander Mathis. "Deep Learning Tools for the Measurement of Animal Behavior in Neuroscience." *Current Opinion in Neurobiology* 60 (2020): 1-11.
18. Wang, Huaizhi, Zhenxing Lei, Xian Zhang, Bin Zhou, and Jianchun Peng. "A Review of Deep Learning for Renewable Energy Forecasting." *Energy Conversion and Management* 198 (2019): 111799.
19. Durairaj, M., and B. H. Krishna Mohan. "A Review of Two Decades of Deep Learning Hybrids for Financial Time Series Prediction." *International Journal on Emerging Technologies* 10, no. 3 (2019): 324-31.
20. Guermoui, Mawloud, Farid Melgani, Kacem Gairaa, and Mohamed Lamine Mekhalfi. "A Comprehensive Review of Hybrid Models for Solar Radiation Forecasting." *Journal of Cleaner Production* (2020): 120357.
21. Pradeepkumar, Dadabada, and Vadlamani Ravi. "Soft Computing Hybrids for Forex Rate Prediction: A Comprehensive Review." *Computers & Operations Research* 99 (2018): 262-84.
22. Hosni, Mohamed, Ibtissam Abnane, Ali Idri, Juan M Carrillo de Gea, and José Luis Fernández Alemán. "Reviewing Ensemble Classification Methods in Breast Cancer." *Computer methods and programs in biomedicine* (2019).
23. Song, Xiangfa, LC Jiao, Shuyuan Yang, Xiangrong Zhang, and Fanhua Shang. "Sparse Coding and Classifier Ensemble Based Multi-Instance Learning for Image Categorization." *Signal processing* 93, no. 1 (2013): 1-11.
24. Chung, Yu-Wei, Behnam Khaki, Tianyi Li, Chicheng Chu, and Rajit Gadh. "Ensemble Machine Learning-Based Algorithm for Electric Vehicle User Behavior Prediction." *Applied Energy* 254 (2019): 113732.
25. AlKandari, Mariam, and Imtiaz Ahmad. "Solar Power Generation Forecasting Using Ensemble Approach Based on Deep Learning and Statistical Methods." *Applied Computing and Informatics* (2019).
26. Moon, K. S., and H. Kim. "Performance of Deep Learning in Prediction of Stock Market Volatility." *Economic Computation and Economic Cybernetics Studies and Research* 53, no. 2 (2019): 77-92.
27. Fischer, T., and C. Krauss. "Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions." *European Journal of Operational Research* 270, no. 2 (2018): 654-69.
28. Tamura, K., K. Uenoyama, S. Iitsuka, and Y. Matsuo. "Model for Evaluation of Stock Values by Ensemble Model Using Deep Learning." *Transactions of the Japanese Society for Artificial Intelligence* 33, no. 1 (2018).
29. Wang, W., W. Li, N. Zhang, and K. Liu. "Portfolio Formation with Preselection Using Deep Learning from Long-Term Financial Data." *Expert Systems with Applications* 143 (2020).
30. Fister, D., J. C. Mun, V. Jagric, and T. Jagric. "Deep Learning for Stock Market Trading: A Superior Trading Strategy?" *Neural Network World* 29, no. 3 (2019): 151-71.
31. Stoean, Catalin, Wiesław Paja, Ruxandra Stoean, and Adrian Sandita. "Deep Architectures for Long-Term Stock Price Prediction with a Heuristic-Based Strategy for Trading Simulations." *PloS one* 14, no. 10 (2019).
32. Agrawal, M., A. U. Khan, and P. K. Shukla. "Stock Price Prediction Using Technical Indicators: A Predictive Model Using Optimal Deep Learning." *International Journal of Recent Technology and Engineering* 8, no. 2 (2019): 2297-305.
33. Bao, W., J. Yue, and Y. Rao. "A Deep Learning Framework for Financial Time Series Using Stacked Autoencoders and Long-Short Term Memory." *PloS one* 12, no. 7 (2017).
34. Yan, H., and H. Ouyang. "Financial Time Series Prediction Based on Deep Learning." *Wireless Personal Communications* 102, no. 2 (2018): 683-700.
35. Vo, N. N. Y., X. He, S. Liu, and G. Xu. "Deep Learning for Decision Making and the Optimization of Socially Responsible Investments and Portfolio." *Decision Support Systems* 124 (2019).
36. Fang, Y., J. Chen, and Z. Xue. "Research on Quantitative Investment Strategies Based on Deep Learning." *Algorithms* 12, no. 2 (2019).
37. Lei, K., B. Zhang, Y. Li, M. Yang, and Y. Shen. "Time-Driven Feature-Aware Jointly Deep Reinforcement Learning for Financial Signal Representation and Algorithmic Trading." *Expert Systems with Applications* 140 (2020).
38. Sabeena, J., and P. Venkata Subba Reddy. "A Modified Deep Learning Enthused Adversarial Network Model to Predict Financial Fluctuations in Stock Market." *International Journal of Engineering and Advanced Technology* 8, no. 6 (2019): 2996-3000.

39. Das, S., and S. Mishra. "Advanced Deep Learning Framework for Stock Value Prediction." *International Journal of Innovative Technology and Exploring Engineering* 8, no. 10 (2019): 2358-67.
40. Go, Y. H., and J. K. Hong. "Prediction of Stock Value Using Pattern Matching Algorithm Based on Deep Learning." *International Journal of Recent Technology and Engineering* 8, no. 2 Special Issue 6 (2019): 31-35.
41. Gonçalves, R., V. M. Ribeiro, F. L. Pereira, and A. P. Rocha. "Deep Learning in Exchange Markets." *Information Economics and Policy* 47 (2019): 38-51.
42. Moews, B., J. M. Herrmann, and G. Ibikunle. "Lagged Correlation-Based Deep Learning for Directional Trend Change Prediction in Financial Time Series." *Expert Systems with Applications* 120 (2019): 197-206.
43. Song, Y., J. W. Lee, and J. Lee. "A Study on Novel Filtering and Relationship between Input-Features and Target-Vectors in a Deep Learning Model for Stock Price Prediction." *Applied Intelligence* 49, no. 3 (2019): 897-911.
44. Long, W., Z. Lu, and L. Cui. "Deep Learning-Based Feature Engineering for Stock Price Movement Prediction." *Knowledge-Based Systems* 164 (2019): 163-73.
45. Rajesh, P., N. Srinivas, K. Vamshikrishna Reddy, G. VamsiPriya, M. Vakula Dwija, and D. Himaja. "Stock Trend Prediction Using Ensemble Learning Techniques in Python." *International Journal of Innovative Technology and Exploring Engineering* 8, no. 5 (2019): 150-55.
46. Sim, H. S., H. I. Kim, and J. J. Ahn. "Is Deep Learning for Image Recognition Applicable to Stock Market Prediction?" *Complexity* 2019 (2019).
47. Agrawal, M., A. U. Khan, and P. K. Shukla. "Stock Indices Price Prediction Based on Technical Indicators Using Deep Learning Model." *International Journal on Emerging Technologies* 10, no. 2 (2019): 186-94.
48. Tashiro, D., H. Matsushima, K. Izumi, and H. Sakaji. "Encoding of High-Frequency Order Information and Prediction of Short-Term Stock Price by Deep Learning." *Quantitative Finance* 19, no. 9 (2019): 1499-506.
49. Sirignano, J., and R. Cont. "Universal Features of Price Formation in Financial Markets: Perspectives from Deep Learning." *Quantitative Finance* 19, no. 9 (2019): 1449-59.
50. Weng, Bin, Lin Lu, Xing Wang, Fadel M Megahed, and Waldyn Martinez. "Predicting Short-Term Stock Prices Using Ensemble Methods and Online Data Sources." *Expert Systems with Applications* 112 (2018): 258-73.
51. Preeti, A. Dagar, R. Bala, and R. P. Singh. "Financial Time Series Forecasting Using Deep Learning Network." In *Communications in Computer and Information Science*, 23-33: Springer Verlag, 2019.
52. Sohangir, S., D. Wang, A. Pomeranets, and T. M. Khoshgoftaar. "Big Data: Deep Learning for Financial Sentiment Analysis." *Journal of Big Data* 5, no. 1 (2018).
53. Lien Minh, D., A. Sadeghi-Niaraki, H. D. Huy, K. Min, and H. Moon. "Deep Learning Approach for Short-Term Stock Trends Prediction Based on Two-Stream Gated Recurrent Unit Network." *IEEE Access* 6 (2018): 55392-404.
54. Das, S. R., K. Mokashi, and R. Culkin. "Are Markets Truly Efficient? Experiments Using Deep Learning Algorithms for Market Movement Prediction." *Algorithms* 11, no. 9 (2018).
55. Kim, J. J., S. H. Cha, K. H. Cho, and M. Ryu. "Deep Reinforcement Learning Based Multi-Agent Collaborated Network for Distributed Stock Trading." *International Journal of Grid and Distributed Computing* 11, no. 2 (2018): 11-20.
56. Faghihi-Nezhad, Mohammad-Taghi, and Behrouz Minaei-Bidgoli. "Prediction of Stock Market Using an Ensemble Learning-Based Intelligent Model." *Industrial Engineering & Management Systems* 17, no. 3 (2018): 479-96.
57. Chong, E., C. Han, and F. C. Park. "Deep Learning Networks for Stock Market Analysis and Prediction: Methodology, Data Representations, and Case Studies." *Expert Systems with Applications* 83 (2017): 187-205.
58. Dingli, A., and K. S. Fournier. "Financial Time Series Forecasting - a Deep Learning Approach." *International Journal of Machine Learning and Computing* 7, no. 5 (2017): 118-22.
59. Singh, R., and S. Srivastava. "Stock Prediction Using Deep Learning." *Multimedia Tools and Applications* 76, no. 18 (2017): 18569-84.
60. Shekhar, S., and N. Varshney. "A Hybrid Ga-Svm and Sentiment Analysis for Forecasting Stock Market Movement Direction." *Test Engineering and Management* 82, no. 1-2 (2020): 64-72.
61. Ahmadi, E., M. Jasemi, L. Monplaisir, M. A. Nabavi, A. Mahmoodi, and P. Amini Jam. "New Efficient Hybrid Candlestick Technical Analysis Model for Stock Market Timing on the Basis of the Support Vector

Machine and Heuristic Algorithms of Imperialist Competition and Genetic." *Expert Systems with Applications* 94 (2018): 21-31.

62. Ebadati, O. M. E., and M. T. Mortazavi. "An Efficient Hybrid Machine Learning Method for Time Series Stock Market Forecasting." *Neural Network World* 28, no. 1 (2018): 41-55.
63. Johari, S. N. M., F. H. M. Farid, N. A. E. B. Nasrudin, N. S. L. Bistamam, and N. S. S. M. Shuhaili. "Predicting Stock Market Index Using Hybrid Intelligence Model." *International Journal of Engineering and Technology(UAE)* 7, no. 3 (2018): 36-39.
64. Ghamisi, Pedram, Bernhard Höfle, and Xiao Xiang Zhu. "Hyperspectral and Lidar Data Fusion Using Extinction Profiles and Deep Convolutional Neural Network." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10, no. 6 (2016): 3011-24.
65. Chen, Yushi, Hanlu Jiang, Chunyang Li, Xiuping Jia, and Pedram Ghamisi. "Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks." *IEEE Transactions on Geoscience and Remote Sensing* 54, no. 10 (2016): 6232-51.
66. Duan, Puhong, Xudong Kang, Shutao Li, and Pedram Ghamisi. "Multichannel Pulse-Coupled Neural Network-Based Hyperspectral Image Visualization." *IEEE Transactions on Geoscience and Remote Sensing* (2019).
67. Li, Shutao, Weiwei Song, Leyuan Fang, Yushi Chen, Pedram Ghamisi, and Jón Atli Benediktsson. "Deep Learning for Hyperspectral Image Classification: An Overview." *IEEE Transactions on Geoscience and Remote Sensing* 57, no. 9 (2019): 6690-709.
68. Paolanti, M., L. Romeo, M. Martini, A. Mancini, E. Frontoni, and P. Zingaretti. "Robotic Retail Surveying by Deep Learning Visual and Textual Data." *Robotics and Autonomous Systems* 118 (2019): 179-88.
69. Dingli, A., V. Marmara, and N. S. Fournier. "Comparison of Deep Learning Algorithms to Predict Customer Churn within a Local Retail Industry." *International Journal of Machine Learning and Computing* 7, no. 5 (2017): 128-32.
70. Ładyżyński, P., K. Żbikowski, and P. Gawrysiak. "Direct Marketing Campaigns in Retail Banking with the Use of Deep Learning and Random Forests." *Expert Systems with Applications* 134 (2019): 28-35.
71. Ullah, Irfan, Basit Raza, Ahmad Kamran Malik, Muhammad Imran, Saif Ul Islam, and Sung Won Kim. "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector." *IEEE Access* 7 (2019): 60134-49.
72. Agarwal, S. "Deep Learning-Based Sentiment Analysis: Establishing Customer Dimension as the Lifeblood of Business Management." *Global Business Review* (2019).
73. Shamshirband, S., J. Khader, and S. Gani. "Predicting Consumer Preferences in Electronic Market Based on Iot and Social Networks Using Deep Learning Based Collaborative Filtering Techniques." *Electronic Commerce Research* (2019).
74. Lei, Z. "Research and Analysis of Deep Learning Algorithms for Investment Decision Support Model in Electronic Commerce." *Electronic Commerce Research* (2019).
75. Leung, K. H., K. L. Choy, G. T. S. Ho, C. K. M. Lee, H. Y. Lam, and C. C. Luk. "Prediction of B2c E-Commerce Order Arrival Using Hybrid Autoregressive-Adaptive Neuro-Fuzzy Inference System (Ar-Anfis) for Managing Fluctuation of Throughput in E-Fulfilment Centres." *Expert Systems with Applications* 134 (2019): 304-24.
76. Cai, Qingpeng, Aris Filos-Ratsikas, Pingzhong Tang, and Yiwei Zhang. "Reinforcement Mechanism Design for E-Commerce." Paper presented at the Proceedings of the 2018 World Wide Web Conference 2018.
77. Ha, Jung-Woo, Hyuna Pyo, and Jeonghee Kim. "Large-Scale Item Categorization in E-Commerce Using Multiple Recurrent Neural Networks." Paper presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2016.
78. Xu, Y. Z., J. L. Zhang, Y. Hua, and L. Y. Wang. "Dynamic Credit Risk Evaluation Method for E-Commerce Sellers Based on a Hybrid Artificial Intelligence Model." *Sustainability (Switzerland)* 11, no. 19 (2019).
79. Saravanan, V., and C. Sathya Charanya. "E-Commerce Product Classification Using Lexical Based Hybrid Feature Extraction and Svm." *International Journal of Innovative Technology and Exploring Engineering* 9, no. 1 (2019): 1885-91.
80. Wang, Yue, Daniel Y Mo, and Mitchell M Tseng. "Mapping Customer Needs to Design Parameters in the Front End of Product Design by Applying Deep Learning." *CIRP Annals* 67, no. 1 (2018): 145-48.

81. Wu, Chen, and Ming Yan. "Session-Aware Information Embedding for E-Commerce Product Recommendation." Paper presented at the Proceedings of the 2017 ACM on conference on information and knowledge management 2017.
82. Lahmiri, S., and S. Bekiros. "Cryptocurrency Forecasting with Deep Learning Chaotic Neural Networks." *Chaos, Solitons and Fractals* 118 (2019): 35-40.
83. Altan, Aytaç, Seçkin Karasu, and Stelios Bekiros. "Digital Currency Forecasting with Chaotic Meta-Heuristic Bio-Inspired Signal Processing Techniques." *Chaos, Solitons & Fractals* 126 (2019): 325-36.
84. Jiang, Zhengyao, and Jinjun Liang. "Cryptocurrency Portfolio Management with Deep Reinforcement Learning." Paper presented at the 2017 Intelligent Systems Conference (IntelliSys) 2017.
85. Chen, Zhensong, Wei Chen, and Yong Shi. "Ensemble Learning with Label Proportions for Bankruptcy Prediction." *Expert Systems with Applications* 146 (2020): 113155.
86. Lin, Wei-Chao, Yu-Hsin Lu, and Chih-Fong Tsai. "Feature Selection in Single and Ensemble Learning-Based Bankruptcy Prediction Models." *Expert Systems* 36, no. 1 (2019): e12335.
87. Lahmiri, Salim, Stelios Bekiros, Anastasia Giakoumelou, and Frank Bezzina. "Performance Assessment of Ensemble Learning Systems in Financial Data Classification." *Intelligent Systems in Accounting, Finance and Management* (2020): 1-7.
88. Faris, Hossam, Ruba Abukhurma, Waref Almanaseer, Mohammed Saadeh, Antonio M Mora, Pedro A Castillo, and Ibrahim Aljarah. "Improving Financial Bankruptcy Prediction in a Highly Imbalanced Class Distribution Using Oversampling and Ensemble Learning: A Case from the Spanish Market." *Progress in Artificial Intelligence* (2019): 1-23.
89. Hinton, Geoffrey E, and Ruslan R Salakhutdinov. "Reducing the Dimensionality of Data with Neural Networks." *science* 313, no. 5786 (2006): 504-07.
90. Zhang, Qingchen, Laurence T Yang, Zhikui Chen, and Peng Li. "A Survey on Deep Learning for Big Data." *Information Fusion* 42 (2018): 146-57.

# Figures

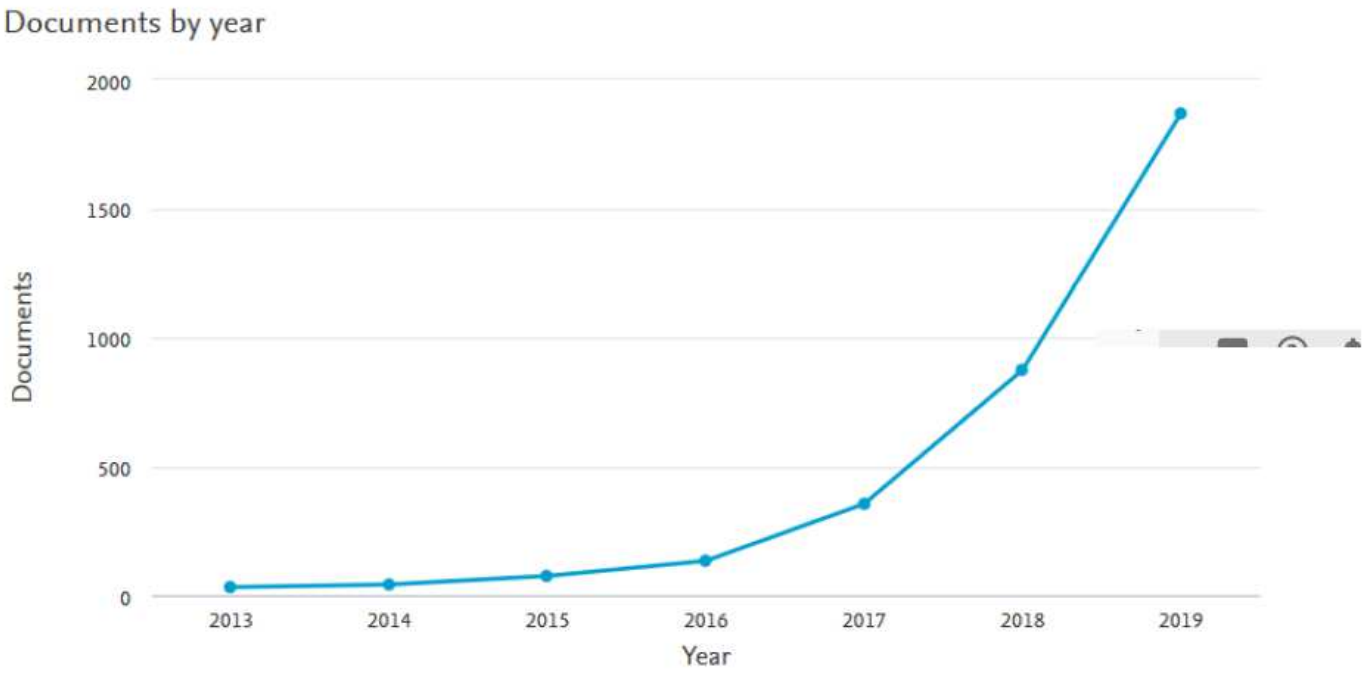



**Figure 1**

Increasing application of data science in Economics

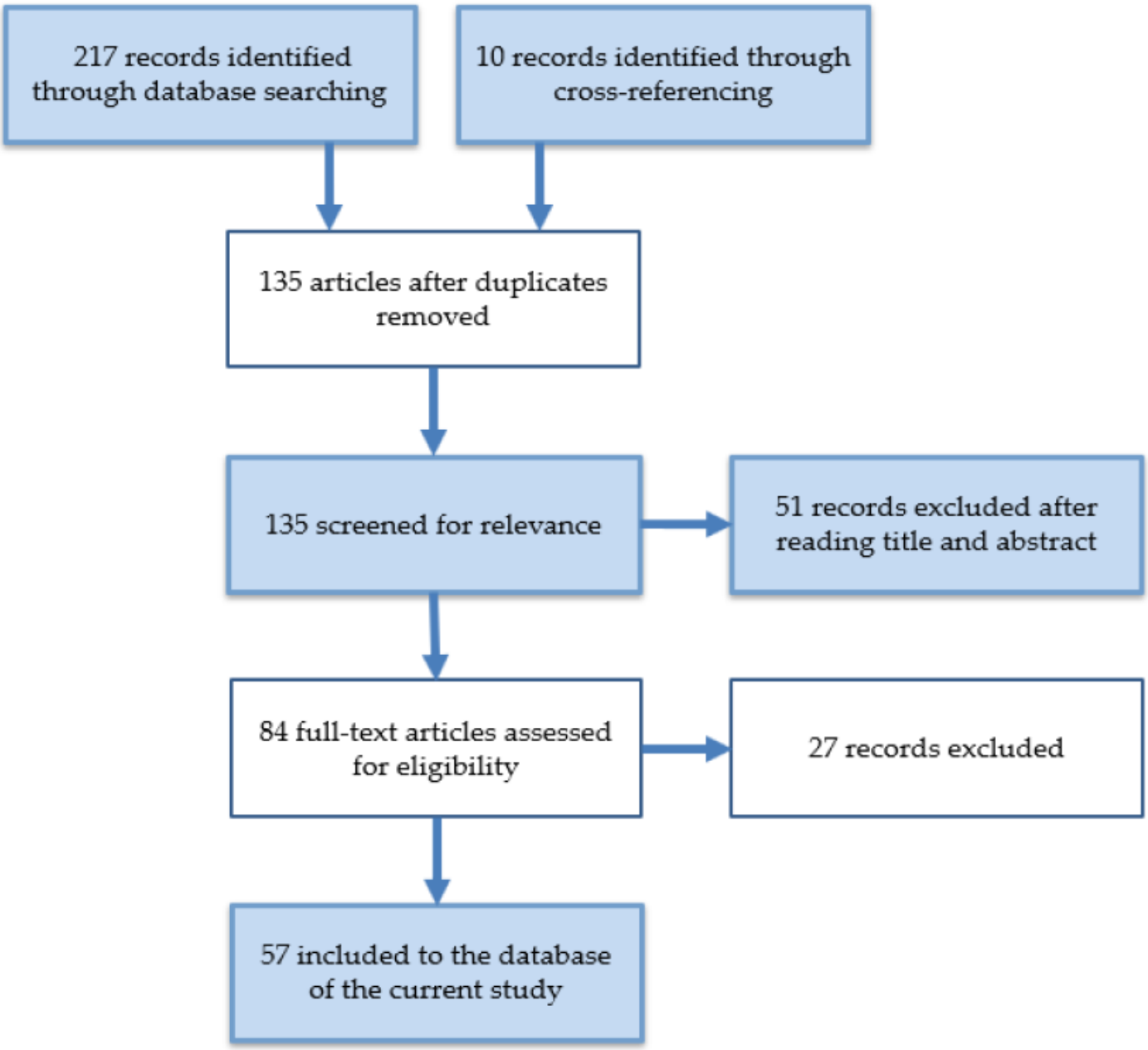



**Figure 2**

Diagram of systematic selection of the study database

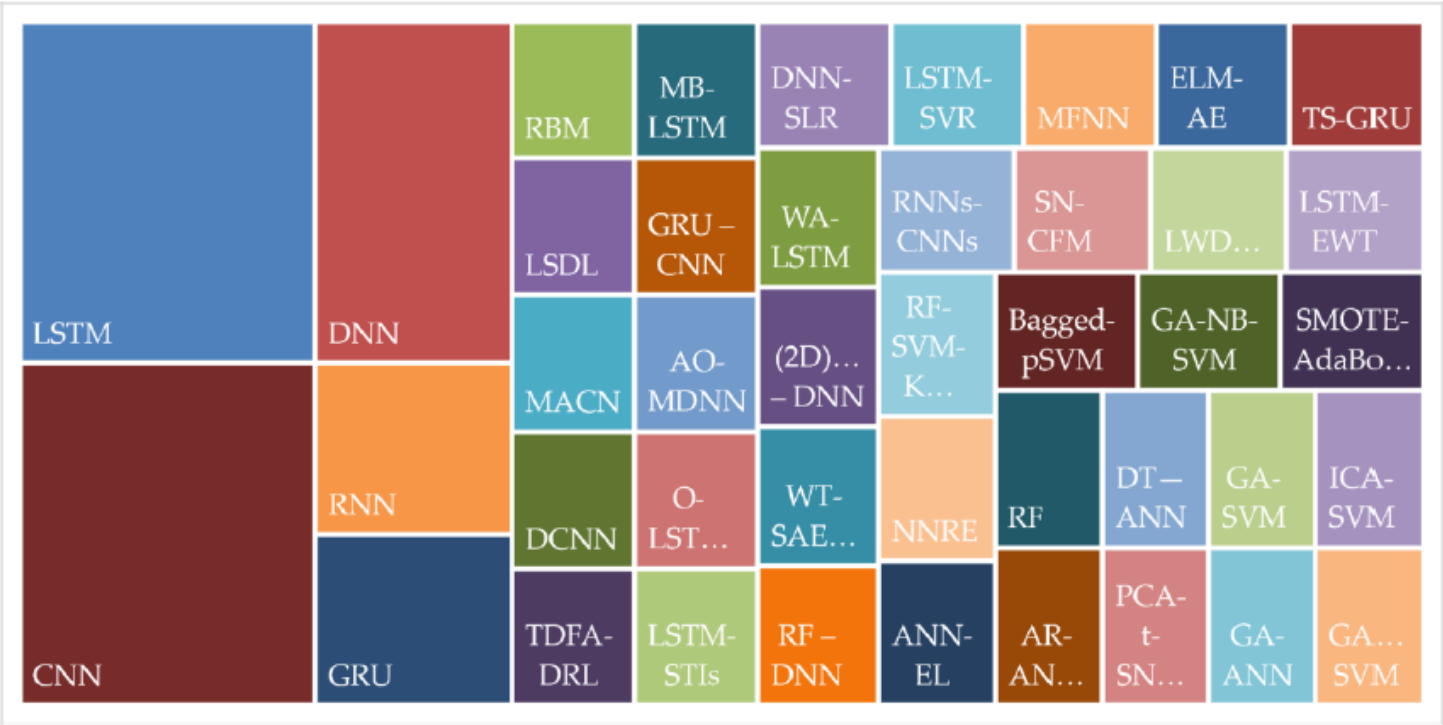



**Figure 3**

List of all deep learning and hybrid deep learning models applied in Economics related fields

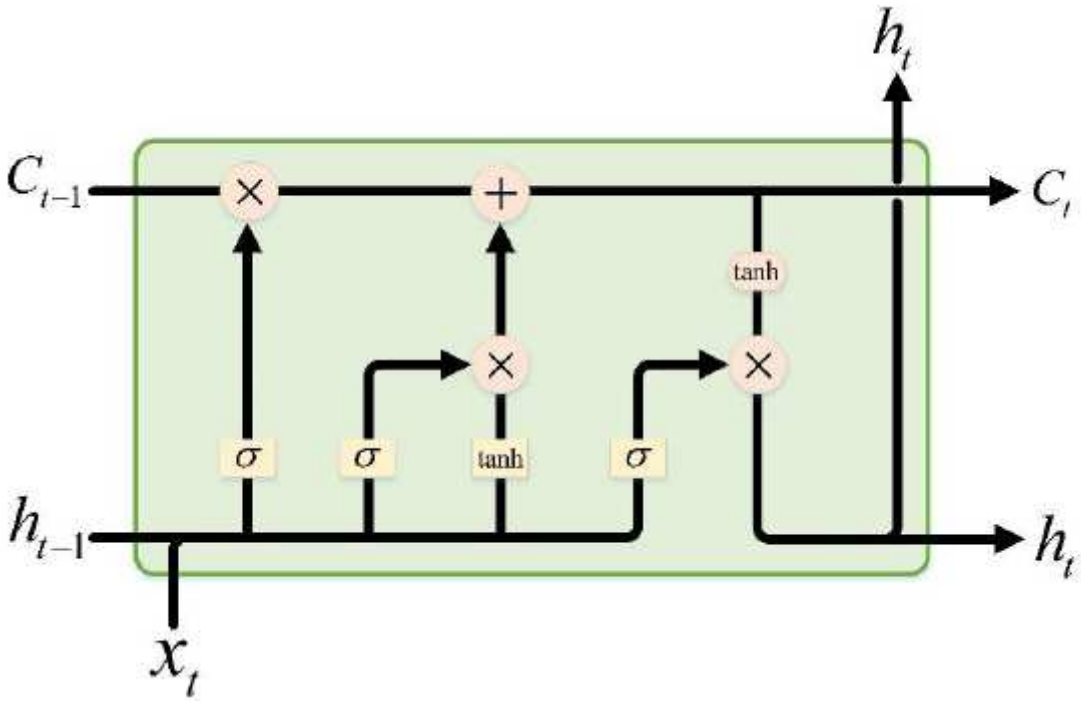



**Figure 4**

The structure of the LSTM network

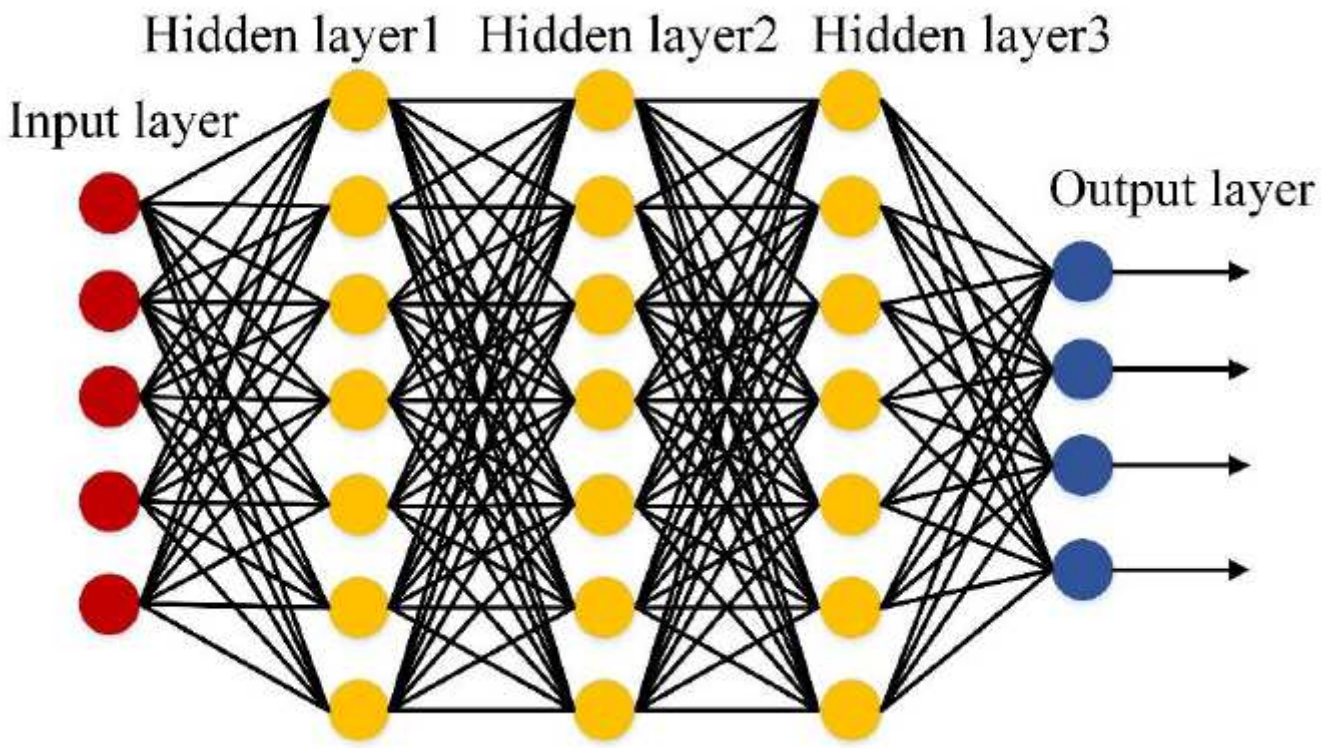



**Figure 5**

Structure of the deep neural network

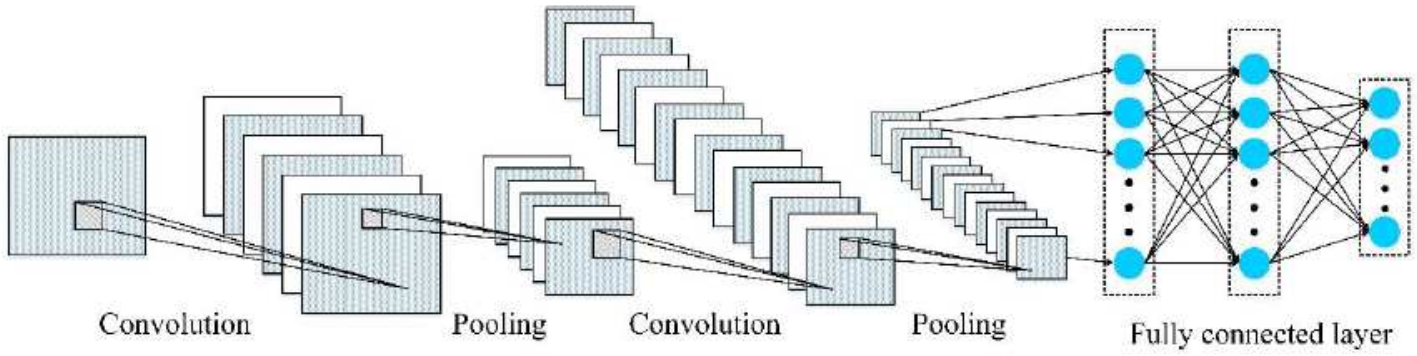
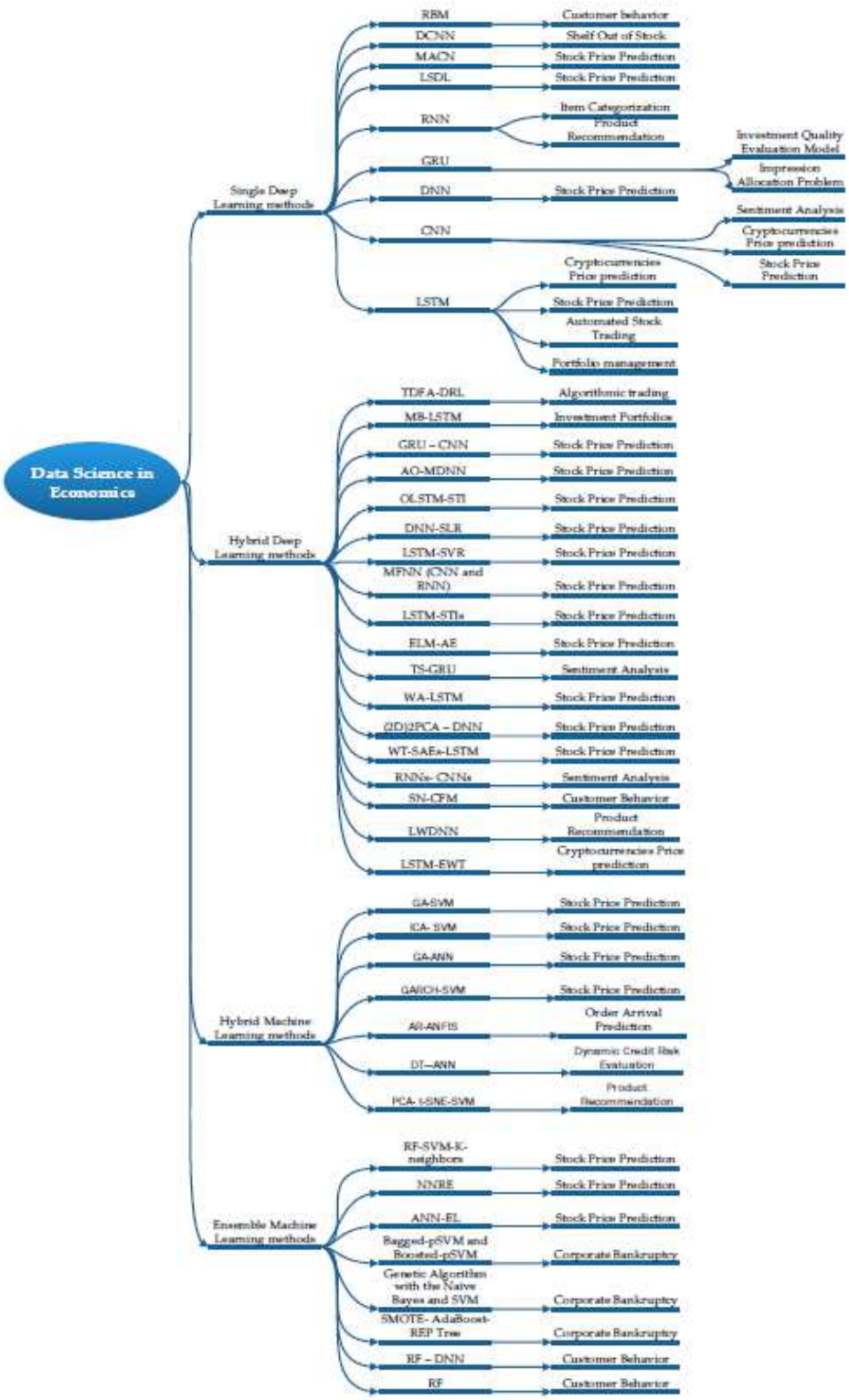


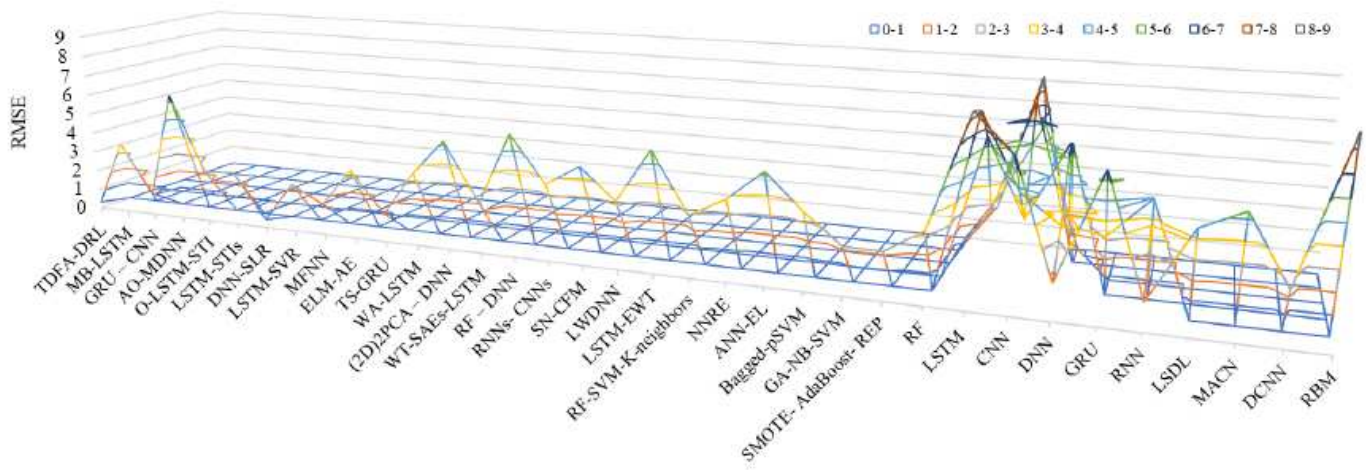

**Figure 6**

Structure of the CNN




## Figure 7

Taxonomy of data science advancements in Economics




## Figure 8

Comparison of root-mean-squared error values of hybrid deep learning models and deep learning