

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/352197415>

REVIEW OF STOCK PREDICTION USING MACHINE LEARNING TECHNIQUES

Conference Paper · June 2021

DOI: 10.1109/ICOEIS1242.2021.9453099

CITATIONS

6

READS

1,680

4 authors:



Ramkrishna Patel

National Institute of Technology, Kurukshetra

3 PUBLICATIONS 9 CITATIONS

[SEE PROFILE](#)



Vikas Choudhary

National Institute of Technology, Kurukshetra

3 PUBLICATIONS 9 CITATIONS

[SEE PROFILE](#)



Deepika Saxena

Goethe-Universität Frankfurt am Main

46 PUBLICATIONS 409 CITATIONS

[SEE PROFILE](#)



Ashutosh Kumar Singh

National Institute of Technology, Kurukshetra

345 PUBLICATIONS 3,126 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



7th International Conference on Smart Computing & Communications (ICSCC) @ Sarawak Malaysia [View project](#)



Doctor of Philosophy [View project](#)

REVIEW OF STOCK PREDICTION USING MACHINE LEARNING TECHNIQUES

RAMKRISHNA PATEL

Department of Computer
Application
NIT KURUKSHETRA
Haryana, India
pramkrishna514@gmail.com

VIKAS CHOUDHARY

Department of Computer
Application
NIT KURUKSHETRA
Haryana, India
vikaschoudhary2906@gmail.com

DEEPIKA SAXENA

Department of Computer
Application
NIT KURUKSHETRA
Haryana, India
13deepikasaxena@gmail.com

ASHUTOSH KUMAR
SINGH

Department of Computer
Application
NIT KURUKSHETRA
Haryana, India
ashutosh@nitkr.ac.in

Abstract—Stock prices change everyday by market forces (supply and demand). In recent years stock price prediction has been one of the most significant concern. Investors are investing on stock market on the basis of certain prediction. For prediction, stock market prices investors are applying some techniques and methods through which they get more profits and minimize their risks. Machine Learning methods are often used for the prediction of stock prices. This survey paper discusses various machine learning approaches (Supervised or Unsupervised) and methods through which the investors get to know the stock prices increase or decrease. It was done in five phases, such as data acquired, pre-processing of dataset, extraction of features, prediction of stock price using different techniques and display the result. In first phase, the data is collected from different social sites, historical data of companies. In second phase, the removal of incorrect, duplicate and dirt is done in pre-processing phase. In third phase, the reduction of data sets and the selection of useful data is done. In fourth phase, prediction is done using different machine learning techniques and approaches which is categorized as supervised and unsupervised learning techniques. Now, in last phase the accuracy is determined using different approaches.

Keywords—Stock prediction, Machine learning, sentimental analysis.

I. INTRODUCTION

The stock market is one of the most important part of every country, It fluctuates the economic growth of the country and also provide a huge effect on many fields, such as employment and industry etc. According to research only 10% of people of every country take a risk to invest in the stock market, it is very difficult to invest because of its dynamic nature, non- stationary, noisy and non-parametric nature [1]. Stock price change due to political problems, the financial economic crisis and many other factors affecting the market. Market forces adjust stock prices on a daily basis. Stock prices fluctuate as a result of supply and demand. The price of a stock increases when there are more people who want to buy it (demand) than there are people who want to sell it (supply). If more people wanted to sell a stock than buy it, the supply would exceed the demand, causing the price to drop [2]. Many researchers are consistently working on building an efficient, inexpensive, and feasible model which guides investors, newcomers, and shareholders in a better way that where to hold or invest their stock so that it can maximize the profitability with

lower risk. There are two categories for stock prediction technical and fundamental analysis. Technical analysis deals with the analysis of historical stock prices to forecast the stock prices. On the other hand, fundamental analysis usually deals on the unstructured dataset such as financial news and economic data. To achieve a better result or accuracy, most researchers concluded that both structured and unstructured data sets are used. Many machine learning techniques are used in stock price prediction like Naive Bayesian, Support Vector Machine, K-Nearest Neighbor, Artificial neural network, Random forest and the important branch of machine learning algorithm, Neural network(NN), Deep neural networking(DNN). The main aim of this research is to provide a detailed analysis using different methodologies of stock market forecasting to predict future trends and stock returns. This paper explains many of the current stock market prediction tools and techniques adapted to the recent research works [3]. An overview of how the stock price predicted from sentiments can be seen in Fig.1.

The next section of this paper contains the literature survey of the related work in this field. Section 3 provides us different exciting stock prediction approaches. Section 4 summary of existing work for stock prediction carries on with the comparative analysis of algorithms along with their accuracy, followed by Section 5 which contains the challenges. In the end, the paper is concluded along with the future advances needed in Section 6.

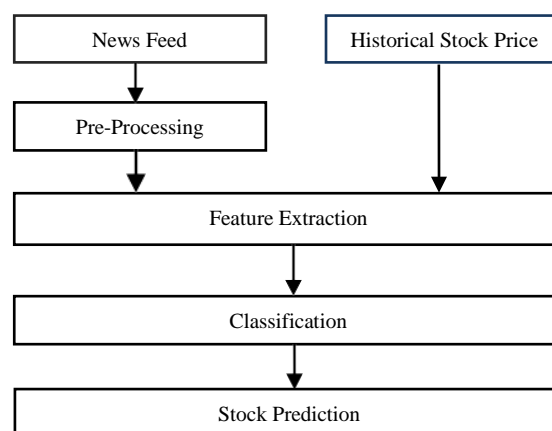


Fig.1 Operational flow of Stock Prediction.

II. LITERATURE SURVEY

Different researchers have used intelligent approaches and strategies in the stock market for decision making in recent year.

Bharathi et al. [3] have proposed a method for stock prediction in which a connection between stock market values and RSS news feed is developed by applying sentiment analysis algorithm such as Natural Language Processing (POS taggers, SSS algorithm), lexicon approach (i.e. Dictionary Based Approach). The authors have used the combination of two datasets Really Simple Syndication [RSS] news feeds and sensex points of ARBK from Amman stock exchange for effective prediction. Experimental output have yielded valuable results, increase in precision prediction of 14.43 percent, when Compared with the regular ID3, C4.5 algorithm.

Vijaya Kumar B P et al. [4] stated that inexpensive, efficient model to predict the stock price using sentiment analysis on social media data these are the crucial factors for applying the ability of machine learning for stock prediction. Main focus on the minimum difference between predicted value given by model and the everyday stock value. Author discussed about many techniques which are important for predicting stocks techniques are: SVM, LSTM, KNN, LINEAR REGRESSION.

Sunil Kumar Khatri et al. [5] have considered that artificial neural network used in problems involving computer tasks, research, similarity finding, and much more. Sentimental evaluation is analyzed in this paper on data extracted from Twitter. This research article carried out an analysis of the sentiments gathered from Yahoo and prepared artificial neural network with performance and market prices for the five largest I.T. Companies estimate the return on investment for the future.

Xiaodong Li a Haoran Xie et al. [6] have introduced a standardized system for forecasting stock price. By using the psychological dictionary of Harvard and Loughran-McDonald's financial sentiment dictionary to create the sentiment dimensions. In this paper, Experiments are performed on five years of historical Hong Kong Stock Exchange prices and news articles. The results show that (1) Models with sentiment analysis outperform the bag-of-words model at the individual stock, sector and index levels. (2) Models using sentiment polarity cannot make useful predictions. (3) there is a small difference between models using two separate sentiment dictionaries.

Aditya Bhardwaja et al. [7] found Sensex and Nifty to be two important measures for forecasting the stock market situation in India. By which shareholders and investors are able to understand the stock market situations that relies on the routine verification and checking of sensex point and Nifty forecast values. This paper analyses stock market sentiment by gathering Sensex and Nifty live server data at various time intervals that can be used to forecast stock

market state. Researchers make use of python scripting language which has a quick implementation context and this will assist the people (investors) what share capital should be invested in.

Ayman E. Khedr et al. [8] stated the stock market is seen as a treasure of investment information. In this model, The authors use two types of news content: market relevant news, business news and financial report written by financial analysts on stocks. Author first stage analysis, news statement to get the text polarity by using Naive Bayes and obtains 72.82% accuracy. In second stage author use both datasets news and historical stock data to forecast the stock prices using a K - NN algorithm and obtains 89% accuracy. Author says this model achieved the highest accuracy compared to other previous models. This model predicts the future stock market trend and obtained accuracy up to 89.80%.

Prof. G.S. Mate et al. [9] assumed that stock prediction and financial news are interlinked to each other. In this paper, Two types of data are gathered which are stock indices and news data from the NY Times Archive API. The Natural Language Toolkit package in python is most commonly used for sentiment analysis to identify emotions or actions by natural language processing. If the news is positive, then we can state that this news effect is good on the stock, so that more chances of stock prices are high. And if the news were bad, it might have an effect on the stock price to go down. Output of sentiment analysis is fed into the machine learning model to forecast stock prices.

Saloni Mohan et al. [10] have developed forecast models based on time series prediction models such as ARIMA, RNN, and Facebook Prophet. This paper forecasts stock prices using time series models and a mixture of neural networks and news posts. The dataset in this paper contains Regular prices of stock for S&P500 companies over five years, along with news articles linked to these businesses. Authors have achieved better results with RN.

Ghaith et al. [11] have focused on the prediction classification Model using Twitter Sentiment based on the HNBC (Hybrid Naive Bayes Classifiers). The authors divided their work into five phases – data collection (Twitter API), filtration and transformation (to obtain the relevant data), labelling (In which data polarity is specified and negative, positive or neutral values are assigned to the opinion of the individual concerned), classification (acceptable stock market patterns are established by the hybridization of Bayes Classifiers) and last one is performance and evaluation. The proposed work achieved accuracy = 90.38%.

Vaanchitha Kalyanaraman et al. [12] This paper presents a sentiment analysis on news articles to see the impact on stock prices. Dataset was from the Bing API. There is a specialized sentiment dictionary intended solely for the study of stock papers. Two separate machine learning algorithms were used for the data set and the performance of

the two was compared. There is a comparison of the projected effects with the real shift in stock prices.

Rakhi Batra et al. [13] considered SVM models for stock prediction because they can handle large feature spaces and are also effective for text categorization. Authors uses two dataset tweets related to Apple products and market index data from Yahoo Finance. The results of the proposed work are 75.22 percent accuracy of the training and 76.68 percent accuracy of the test. According to the authors, this work needs to improve the accuracy of the text classification test and the accuracy of the prediction if the size of the dataset is increased.

Dev Shah et al. [2] focused on some of the stock market forecasting research achievements and discussed some of the forecasting methods and technical, fundamental, short-term, long-term approaches that help predict stock prices. According to the author, a combination of statistical and machine learning techniques is likely to be more useful for stock prediction. The authors note that the market acts as a weighting machine with less noise and more predictability in the long run.

Antony et al. [14] considered that ANN (Artificial Neural Network) was a highly versatile, non-linear approach to forecasting stock prices for BSE-related companies (Bombay Stock Exchange). In this model, the author uses the opening price, high, low, closing price, and stock volume as a dataset. According to the author, if we want to increase the accuracy rather than increase the number of inputs, because the author observed that the error in the model was high when they took three inputs for prediction, and it was minimized when the five inputs and the accuracy of the predictions were higher.

Ryo Akita et al. [15] proposed Deep Learning Models and LSTM for prediction of stock prices. Authors approach predicts stock prices by using Regression for textual information and LSTM for numerical information. Paper results have shown that the distributed representation of text data is better than the numerical data and Bag-of-Words based methods, LSTM was effective for time series data impacts than other models.

Jing Zhang et al. [16] proposed a new stock market trend prediction method that can forecast financial performance as well as its interval of increase (or decrease) rate within predefined forecasting periods. According to this paper, the model uses an heuristic algorithm (unsupervised) to split raw user information from each stock into multiple clips of predetermined duration and classify them into four categories Up, Down, Flat, and Unknown. This proposed model outperforms some existing techniques.

III. STOCK PREDICTION APPROACHES

Pre-Processing and Feature Selection [3]: In machine learning, pre-processing is a crucial phase to get rid of the repetitive, outdated, and incorrectly formatted content. Dirty data may cause uncertainty in the third step. As a result, data pre-processing is carried out in order to clean the dataset.

Data Pre-processing phase resolves three major type of issues- Data cleaning, Data Transformation, Data Reduction. In feature selection phase of machine learning the reduction of datasets and selection of useful data is done. There are two major types of feature selection techniques: supervised and unsupervised methods. When developing a predictive model, feature selection is the method of reducing the number of input variables. The number of input variables should be reduced to reduce the computational cost of modeling and, in some cases, to increase the model's accuracy.

Natural Language Processing (NLP) [10]: opinion mining is a methodology use to determine the mood, opinion and emotion of an individual about and specific issues . To recognize and bring out subjective information it utilize the processing of natural language and text analysis in source material.

Dictionary based Approach [3] [4]: Dictionary-based methods use synonyms, antonyms and hierarchies WordNet to decide the meaning of the word.

Naïve Bayes [3]: Naive Bayes has been one of the easiest and most reliable classification machine learning technique that strengthens fast machine learning models that can make quick predictions. It is a probabilistic classifier, which implies that it predicts on the basis of the probability of an object.

Support Vector Machine (SVM) [17]: Support Vector Machine(also known as support-vector networks) is a set of related supervisory techniques used to classify and regress. SVM is a supervised machine learning models with associated learning algorithms that analyze data for classification and regression analysis in machine learning. SVMs, which are based on mathematical learning models, are one of the most reliable prediction methods [18].

LINEAR REGRESSION [8]: Linear Regression is a part of machine learning algorithm(supervised learning). It's performing a regression operation. It is often used to assess the relationship between variables and forecasting.

K-nearest neighbor (KNN) [7]: The method for information processing is k-nearest neighbor. In machine learning methods, K nearest neighbor is a simple algorithm where the categorization is accomplished by recognizing the closest neighbor to query examples and using those neighbors to determine the question class.

Artificial Neural Network (ANN) [19] [20]: is a component of a computer system designed to stimulate the gathering and processing of human brain information. It is the backbone of Artificial Intelligence (AI) and resolves issues that would prove impossible or difficult by human or statistical standards [21]. ANN is commonly use for regression as well as classification in various field [22].

Long short term memory (LSTM) [23]: is an RNN(recurrent neural network). LSTM networks used for forecasting based

on real time series data, classification of data processing of data [24].

IV. SUMMARY OF EXISITING WORK FOR STOCK PREDICTION.

Year	Literature reference	Future selection Approach	Classification (Approach)	Dataset	Result (accuracy)	Limitation
2013	Antony et al	ANN(Artificial neural network)	Supervised approach	BSE(Bombay stock exchange) and news article.	Improve the accuracy	Difficulties in transmitting the issue to the network.
2014	Vaanchitha Kalyanaraman et al.	Sentiment dictionary	Supervised learning	News articles	predicted sentiment using linear regression with normal equation.=54.54% and with Gradient Descent=81.81%	If the sentiment dictionary does not recognise a word, it must be taken in isolation.
2014	Xiaodong Li a Haoran Xie et al.	psychological dictionary of Harvard and Loughran-McDonald's financial dictionary.	Bag-of-words approach	news articles and historical Stock Exchange prices.	LMD achieves the best in both validation(0:5976) and testing(0:5527) when compared to six different approaches.	The sense of the word in the text is ignored by the approach. It pays little attention to the context in which it is used.
2015	Aditya Bhardwaja* et al.	Python scripting language.		Sensex data and Nifty data values.	Prediction of stock market prices.	Third-party frameworks and libraries are needed.
2016	Sunil Kumar Khatri1 et al.	NLP and Text Analysis (artificial neural network)	Supervised approach	Stock Twits	Value determined for each company with a mean square error	Lack of research and development. Low-resource language
2016	Ghaith Abdulsattar A. Jabbar Alkubaisi et al.	Hybrid Navie Bayes Classifiers.	Supervised approach	Twitter Dataset (Twitter API).	Accuracy = 90.38% .	This limits the applicability of this algorithm in real-world use cases. In certain cases, its predictions are incorrect.
2017	Shri Bharathi.Sv et al.	Natural Language Processing, lexicon approach(Dictionary Based Approach)	Unsupervised learning	Really Simple Syndication[RSS] news feeds and sensex points of ARBK	Accuracy improve 14.43% as compared to ID3, C4.5.	Since the input data is unknown and not labelled in advance, the results are less accurate.

2018	Rakhi Batra et al.	SVM(Support vector machine)	Supervised Approach.	Market index data and tweets from twitter	Accuracy up to 76.65%	Large datasets require a long training time.
2018	Jing Zhang et al.	Heuristic algorithm	Unsupervised learning	Stock Price Data	Improves the accuracy.	When approximate solutions are sufficient but exact solutions are computationally costly, heuristic algorithms are frequently used.
2019	Saloni Mohan1 et al.	Regression, Time series models, Neural Network and financial news articles	Supervised approach (Deep Learning).	Daily stock prices and news articles.	Improve the accuracy of stock price predictions.	Only LINEAR relationships are being considered.
2019	Dev. Shah et al.	combination of statistical and machine learning techniques	Supervised approach	Daily stock price and news articles.	Improve the accuracy.	Needs lot of time.
2019	Prof. G. S. Mate et al.	natural language processing	Unsupervised Machine learning	Stock indices and news data from the NY Times Archive API	Predict the stock prices	Machines have a hard time implementing NLP.
2019	Ayman E. Khedr et al	Naive bays K - NN algorithm.	Supervised approach	market relevant news, business news	Accuracy up to 89.80%.	In certain cases, its predictions are incorrect.
2020	VijayaKumar et al.	k-NN algorithm, regression analysis	Supervised approach	News articles and historical data	Accuracy up to 85%	Probability estimation would be zero if there are no appearances of a class label.
2021	Ryo Akita et al.	LSTM(long short term memory)	bag-of-words approach	miscellaneous events reported in newspapers	Predict the stock prices.	The sense of the word in the text is ignored by the approach. It pays little attention to the context in which it is used.

V. CHALLENGES:

(1). Stock Prediction is dynamic in nature.

This is due to the fluctuations in stock market which changes every day. To overcome this challenge we need to Re-train online our model periodically so that we get to know the prediction of every change.

(2) Re-training the model frequently to predict the stock market change.

There is an ambiguity that after how much period we need to Re-train our model so that we get optimum prediction.

(3) Selecting the appropriate machine learning model for the respective prediction.

Suppose If *the training data* is much larger than no. of feature in such case, KNN is better than SVM but if no. Of features is much larger than training data, then SVM outperforms KNN.

(4) Out of vast the real time amount of dataset sorting, pre-processing, features extraction is difficult task to evaluate.

VI. CONCLUSION AND FUTURE WORK:

The goal of our research study is to help stockbrokers and investors to invest capital in the stock market. Prediction plays vital role in the stock market, which is a very complicated and difficult process due to the political problems, the financial economic crisis and many other factors affecting the market. This survey paper discusses various machine learning approaches such as NLP, Linear Regression, KNN, SVM, LSTM, Artificial Neural Networking and many more. Merit of prediction model is to guides investors, newcomers, and shareholders in a better way that where to hold or invest their stocks so that it can maximize the profitability with lower risk. On the other hand, If the dataset contains fake news and inappropriate information then the prediction of stock price would be incorrect. Our plan for future would be to improve the accuracy of stock prediction by implementing the new method. We will prepare a model which will be more accurate than the existing one and it will also overcome the current limitations.

References

- [1] A. Sharma, D. Bhuriya and U. Singh, "Survey of Stock Market Prediction Using," *International Conference on Electronics, Communication and Aerospace Technology*, pp. 506-509, 2017.
- [2] D. Shah, H. Isah and F. Zulkernine, "Stock Market Analysis: A Review and Taxonomy of Prediction Techniques," *International Journal of Financial Studies*, pp. 2-22, 2019.
- [3] S. Bharathi and A. Gheetha, "Sentiment Analysis for Effective Stock Market Prediction," *International Journal of Intelligent Engineering and Systems*, pp. 142-153, 2017.
- [4] N. N. Reddy, V. K. B P and N. E. , "Prediction Stock Price Using Sentimental Analysis Through Twitter Data," *UTC from IEEE Xplore* , 2020.
- [5] S. K. Khatri and A. Srivastava, "Using Sentimental Analysis in Prediction of Stock Market Investment," *International Conference on Reliability, Infocom Technologies and Optimization*, pp. 566-563, 2016.
- [6] X. Li and J. Wang, "News impact on stock price return via sentiment analysis," *Knowledge Based System*, vol. 69, pp. 14-23, 2014.
- [7] A. Bhardwaj, Y. Narayan, P. and V. , "Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty," *Procedia Computer Science*, vol. 70, pp. 85-91, 2015.
- [8] A. E. Khedr and N. Yaseen, "Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis," *Intelligent Systems and Applications*, vol. 7, pp. 22-30, 2017.
- [9] G. S. Amidwar, R. Kulkarni and M. Muthya, "Stock Prediction Through News Sentiment Analysis," *Journal Of Architecture & Technology*, vol. 9, pp. 36-40, 2019.
- [10] S. Mohan, S. Mullapudi, S. Sammeta, . V. Parag and . D. C. Anastasiu, "Stock Price Prediction Using News Sentiment Analysis," *International Conference on Big Data Computing Service and Applications*, pp. 205-208, 2019.
- [11] G. Abdulsattar, S. S. Kamaruddin and H. Husni, "Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naive Bayes Classifiers," *Computer and Information Science*, vol. 11, pp. 52-64, 2018.
- [12] V. Kalyanaraman, S. Kazi , T. Rohan and S. Oswal , "Sentiment Analysis on News Articles for Stocks," *Asia International Conference Modelling Symposium*, pp. 10-15, 2014.
- [13] R. Batra and S. M. Daudpota, "Integrating StockTwits with Sentiment Analysis for better Prediction of Stock Price Movement," *International Conference on Computing, Mathematics and Engineering Technologies – iCoMET*, vol. 2, 2018.
- [14] A. V. Devadoss and T. A. A. Ligori, "Stock Prediction Using Artificial Neural Networks," *International Journal of Data Mining Techniques and Applications*, vol. 02, pp. 283-291, 2013.
- [15] R. Akita, A. Yoshihara, T. Matsubara and K. Uehara, "Deep Learning for Stock Prediction Using Numerical and Textual Information," *Authorized licensed use limited to: IEEE Xplore*, 2016.
- [16] J. Zhang, S. Cui, Y. Xu, Q. Li and T. Li, "A novel data-driven stock price trend prediction system," *Expert Systems With Applications*, pp. 60-69, 2018.
- [17] L. and M. Chi, "Using support vector machine with a hybrid feature selection method to the stock trend prediction," *Expert Systems with Applications* 36 , vol. 8, pp. 10896-10904, 2009.
- [18] Guo, . H. A. Zhang and W. Wang, "An accelerator for online SVM based on the fixed-size KKT window," *Engineering Applications of Artificial Intelligence* 92, 2020.
- [19] D. Saxena and A. K. Singh, "A proactive autoscaling and energy-efficient VM allocation framework using online multi-resource neural network for cloud data center," *Neurocomputing* 426, pp. 248-264, 2021.
- [20] D. Saxena and A. K. Singh, "Auto-adaptive learning-based workload forecasting in dynamic cloud environment," *International Journal of*

Computers and Applications , pp. 1-11, 2020.

- [21] D. Saxena, A. K. Singh, A. Mohan and J. Kumar, "Biphase adaptive learning-based neural network model for cloud datacenter workload forecasting," *Soft Computing*, pp. 1-18, 2020.
- [22] D. Saxena, A. K. Singh and R. Buyya, " An Online VM Prediction based Multi-objective Load Balancing Framework for Resource Management at Cloud Datacenter," *IEEE Transactions on Cloud Computing* , 2021.
- [23] G. D. and . L. QIN, "Study on the prediction of stock price based on the associated network model of LSTM.," *International Journal of Machine Learning and Cybernetics 11*, vol. 6, pp. 1307-1317, 2020.
- [24] J. Y. Yang and . Y. Liu, "Stock closing price prediction based on sentiment analysis and LSTM," *Neural Computing and Applications*, pp. 1-17, 2019.
- [25] T. S. Kumar , "Data Mining Based Marketing Decision Support System Using, Hybrid Machine Learning Algorithm" *Journal of Artificial Intelligence 2*, vol. 3, pp. 185-193, 2020.
- [26] Vivekanadam B. , "Analysis of Recent Trend and Applications in Block Chain Technology ," *Journal of ISMAC 2*, vol. 4, pp. 200-206, 2020.