DOI: https://doi.org/10.30837/MMP.2024.062

ALGORITHMS FOR DECISION SUPPORT IN RISK MANAGEMENT IN THE DEVELOPMENT OF INFORMATION-SENSITIVE SOCIALLY ORIENTED SYSTEMS

Khovrat A., Kobziev V.

The increase in the amount of data and the overall change in social and market processes are driving the transformation of basic management principles. This is especially true for risk management. Under non-deterministic conditions, when developing solutions for socially oriented systems, forecasting possible problems requires the use of modern data mining tools. Existing concepts cannot fully guarantee high efficiency in the face of social shifts exacerbated by a falsified information environment, for example, during discrediting campaigns against ideas proposed by businesses. At the same time, there is a problem with the speed of simple solutions and the high cost and security of using more complex cloud technologies. The current work is focused on considering modifications to simple decision support models and building a data processing algorithm to improve the accuracy and reliability of project forecasting. The proposed sequence of steps allows us to take into account the essence of instability and its impact on society, while taking into account the peculiarities of the information environment. Experimental comparison of various classical models with the proposed solution allows us to state the higher efficiency of the created algorithm, both in terms of accuracy and speed due to parallelization. This, in turn, opens the way to solving the problem of risk prediction for information-sensitive socially oriented systems

1 Introduction

One of the most important preparatory steps in project management in any field is to identify key risks and ways to mitigate them. In today’s environment, this task is becoming complex and requires processing a large amount of historical data. This state of affairs has given rise to the development and popularization of decision support systems (DSS), which can both aggregate information or make forecasts and, in some cases, provide recommendations on how to solve problems. Nevertheless, most DSSs, based on available reviews on specialized resources, are not effective enough in the face of uncertainty, especially when targeted projects relate to the activities of social groups, such as urban systems [1]. The intensification of social shifts over the past few years, which was reflected in the pandemic, and later in the Russian-Ukrainian war and several local armed conflicts, raises the issue of improving the algorithms that may underlie the described systems. At the same time, this problem is even more acute when building a highly intelligent infrastructure for various sectors of human life. Not least because of the impact of falsified information on human behavior. As an example, we can mention the campaign against the installation of a more modern Internet based on 5G technology, which is certainly necessary for the more reliable functioning of all the latest infrastructure solutions [2]. Another similar example is the discrediting of the "smart city" concept by linking it to the forest fires on the island of Maui [3].

Another important problem is that existing solutions either require significant cloud resources or are not fast enough, which can play a critical role under certain conditions. Although there are several large companies that provide their own facilities for deploying complex software products, the issue of their reliability and security remains unresolved [4].

Given this, it was decided to modify the existing algorithms of the SPPR in order to increase their effectiveness in managing risks for socially oriented systems. Such a modification should not only take into account the nature of social shifts or audience behavior, but also the likelihood of falsified information aimed at discrediting the solutions being developed, in particular those related to urban infrastructure. At the same time, the speed of these algorithms should be sufficient for their practical application To achieve this goal, the following tasks were identified

- review of existing solutions that underlie modern SPDS and identification of their key shortcomings;
- development of a data pre-processing algorithm that would allow to take into account changes in the behavior of the target audience and the impact of news on it;
- determination of the basic algorithm for forecasting the company's performance within the target project and opportunities for its further improvement;
- researching the possibilities and implementing the parallelization of the algorithm;
- determining the effectiveness of the proposed solution by conducting an experiment and solving a linear optimization problem.

We will gradually review the results of the above steps.

2 Theoretical basis

This section presents the results of the analysis of the key algorithms used in decision support systems and the identification of the features of target indicators of audience behavior, information environment, social shift and business environment.

2.1 Choosing the basic algorithm

In general, DSSs are usually divided according to two criteria [5]:

- the way of decision support: focused on data, knowledge, documents, communication, and models;
- method of interaction with the user: active, passive, combined.

Given that it was decided to focus on simple solutions that do not require significant hardware capacity and on the development of an algorithm for information processing, only data-oriented passive DSS will be considered in the following. The choice of the "passive" form is explained by the fact that the recommendation subsystem is a separate model and goes beyond the defined tasks.

In 2023, some of the most popular systems implemented in the risk management process are: Hyperproof, Soterion, Whistic [1].

After analyzing the feedback over the past 3 years and the official websites of the identified projects, the following features were identified

- the predictive algorithms are based on neural networks and/or autoregressive models. It is worth noting that some other DSSs are also capable of using a probabilistic approach, but it is slower and requires the use of cloud computing;
- there is a limited consideration of risks associated with the activities of the target audience of the systems being developed, in particular, problems with forecasting business activity during the COVID-19 pandemic were noticed;
- lack of consideration of the information environment in which the target project is being created.

The defined classes of basic predictive algorithms are relatively broad and require additional filtering. In particular, in the case of neural networks, we can consider:

- convolutional neural networks (CNNs) that use a convolutional function to reduce the dimensionality of input data;
- recurrent neural networks (RNN), which are based on reusing the results of the previous layer. More complex networks with long-term memory support can also be considered here;
- combined hybrid networks (generalized RCNN) that combine several simple models, providing higher accuracy.

At the same time, among autoregressive models, it is worth mentioning

- distributed lag autoregression and seasonal autoregression based on the use of basic features of time series;
- moving average autoregression and integrated moving average autoregression, which allow for the aggregation of related data.

An additional study of the problem of forecasting economic indicators by using international scientific papers allows us to limit the set of basic forecasting algorithms to a combined hybrid neural network and autoregressive integrated moving average. These models, although the slowest, have the highest accuracy of the proposed ones, they are also capable of parallelization and do not require significant hardware capacity [6].

2.2 Defining target indicators

In order to mitigate the risks that may accompany the development and implementation of a socially oriented program project, a set of key aggregate indicators should be identified. Based on the discussion in Section 1, the following general indicators can serve as

- social shift profile, a parameter that allows converting the uncertainty of conditions into a numerical form;
- target audience profile, which summarizes the behavior of the most influential targeted actors;
- business environment profile, which defines the specifics of implementing a particular project in the market context;
- information environment, an indicator that reflects the intensity of the impact of fabricated information on the target concept.

In order to form a mathematical representation, it is necessary to understand the specifics of each of the identified factors. The following indicators were formed based on the analysis of modern scientific publications and expert evaluation among 100 sociologists, engineers, managers and executives of Kharkiv, Lviv, Dnipro, Kyiv, Lisbon and Krakow.

When studying the concept of a social shift, also called a "social catastrophe", it was found that the most influential sub-indicators are the prevalence of the source of the shift, its duration (including the moment of the first information appearance), specifics for a particular area of activity, and the level of severity. The first two indicators are inherently objective numerical variables, while the other two reflect the subjective perception of the shift and require additional algorithms to be used in forecasting.

The profile of the target audience can be determined by taking into account the size of this audience, the market paradoxicality of the target decision (if it is subject to well-known neoclassical economic paradoxes), the degree of trust, and the general description of society. As in the first case, there is a mix of numerical objective indicators with textual indicators subjective to the target project.

The profile of the business environment focuses on the reaction of business entities to both the implementation of the proposed project solution in general and the social catastrophe in particular. To take this into account, it was decided to focus on indicators of financial stability of the economy (both global and local) and business readiness for emergencies. Although the latter indicator in most cases does not directly affect systems focused on the smart city concept, it allows for an adjustment of the objective assessment of a city's financial stability with the subjective perception of its internal counterparties.

The information environment, in the framework of the current work, refers to the intensity of the spread of fake news about the project topic, technological reforms, and other similar domains of knowledge. To be able to calculate this, it is necessary to understand the specifics of defining fabricated information. Since it was decided to focus on textual data, the following characteristics can be identified [7, 8]:

- the use of an unnatural number of rhetorical questions (contextual distortion of socially significant topics). Linguistic studies show that this type of speech construction is not often used in official business and journalistic styles for use by the media;
- lack of negative constructions to reduce the cognitive load in combination with pessimistic colors of the selected words. As an example, the replacement of the word "bad" with "catastrophe". It is worth noting here that profanity will be deliberately removed from further texts, as it complicates the process of analyzing emotional coloration;
- the use of appeals and encouragement in the wrong context and the use of an unreasonable number of pronouns. In this case, there is an imitation of a journalistic style of presentation;
- high frequency of using short sentences and words with grammatical errors.

These characteristics are not exhaustive, but it should be emphasized that the target texts for review, in addition to obscene vocabulary, will not include the mixing of several languages and the use of regional dialects. Such add-ons are beyond the scope of the tasks described.

In general, the field of determining the falsity of information is not new, as mentioned above. Several groups of European scientists have shown that machine learning algorithms based on both neural networks and more modern transformers or autoencoders require significant amounts of data to achieve an accuracy of more than 90% [9, 10]. However, this problem can be solved by using a balanced data set, as demonstrated by Ukrainian researchers.

Another well-known way to determine whether data has been falsified is to use graph models [11]. In their work, Harvard scientists have demonstrated their high efficiency in solving the problem of detecting fake accounts. However, this method will not allow to fully process textual information from news, or will require more significant capacity for pre-processing. A similar problem applies to algorithms that help detect spam. A research team of Chinese-American scientists has proved the possibility of effective use of Markov networks [12]. However, given the specifics of the field and the goal, their use is quite cumbersome and will require the use of cloud technologies. A similar problem has already been considered for the basic predictive algorithm [6].

3 Mathematical basis

This section will outline the key features of the proposed basic algorithms and target indicators of pre-processing.

3.1 Target algorithms

As noted above, the current paper considers two basic algorithms - RCNN and vector moving average autoregression. The vector nature of the latter is necessary for the possibility of simultaneous processing of several indicators.

In a simple CNN model, passing a filter allows to take into account the neighborhood of each element, but the specificity of the proposed indicators requires understanding a longer time period without a significant shift to the future. Thus, a significant context may be outside the filter of the CNN model. To avoid this problem, it was decided to combine RNN and CNN.

Although there are several ways to do this, this study will only consider an architecture that uses two neural networks in sequence. In other words, after convolution is performed, the result is not only concatenated, but sent to the layer with the recurrent neural network.

To be able to take into account the context to the fullest extent, it was decided to use a bidirectional recurrent neural network with support for both long-term and short-term memory rather than a simple RNN architecture. It is based on the use of hyperbolic tangents and sigmoidal lines, which avoid the problem of exploding and disappearing gradients by limiting the area of the resultant values. At the same time, the defined model allows to take into account the entire historical context. Thus, the RCNN architecture can be represented as shown in Figure 1.

As a result of the cross-validation testing, it was determined that the best key hyperparameters are the following values:

- the kernel size is 4;
- the step size is set at 1;
- based on the set step size, the parameter for adding insignificant zeros will not be applied;

- based on the specifics of the subject area, it was decided not to apply the offset parameter;
- the filter dimension is set to 5´5´3 (the last value of the dimension is determined by the number of target indicators).

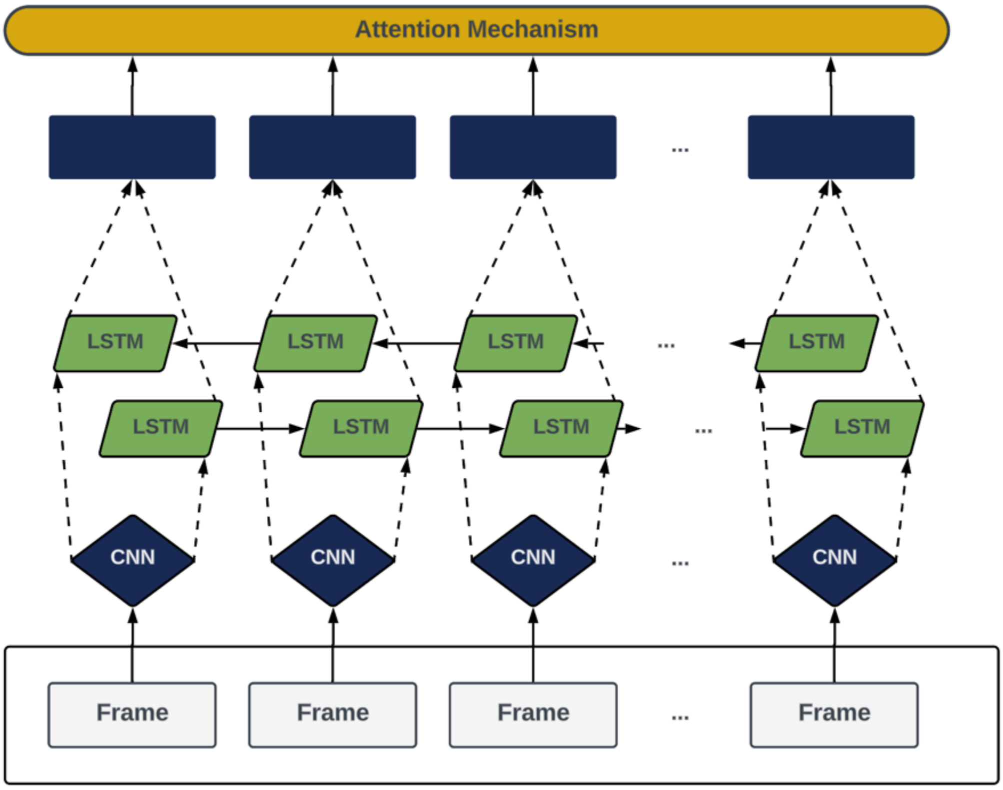

Fig. 1. Schematic of RCNN architecture

To understand the essence of the second basic algorithm, it is necessary to consider the features of vector autoregression (VAR) in general. It can be represented as follows:

$$\Phi_0 y_t = \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \Theta_0 u_t + \Theta_1 u_{t-1} + \ldots + \Theta_q u_{t-q}, \qquad (1)$$

where y_t – K-dimensional time series;

Φ_i, Θ_j – dimension matrices $K \times K$, $i = \overline{1, p}$, $j = \overline{1, q}$;

u_t – K-dimensional white noise vector with zero mean and the following non-degenerate covariance matrix $\Sigma = \mathbb{E}(u_t, u_t')$.

The above formula (1) shows that the classical family of VAR models forecasts only static variables. In order to take into account exogenous variables, we decided to use a modification of the error correction (abbreviated as EC). This adjustment is necessary when several endogenous variables have a common stochastic trend [13]. This is the case for the problem under consideration.

The general formula for the modified EC-VAR family of algorithms is as follows:

$$\Phi_0 \Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{p-1} \Psi_i \Delta y_{t-i} + \sum_{j=0}^{q} \Theta_j u_{t-j}, \tag{2}$$

where $\Pi = -\left(\Phi_0 - \Phi_1 - \ldots - \Phi_p\right)$; $\Psi_i = -\left(\Phi_{i+1} + \ldots + \Phi_p\right)$, $i = \overline{1, p-1}$.

The integration and mobility of the selected target model will provide the ability to take into account the neighborhood of the target element of the time series. In this case, the problem is similar to the one identified for CNN, as the family of autoregressive models does not provide for the full consideration of the historical context.

3.2 Target indicators

The Social Disruption Profile (SDP), as already mentioned, is divided into 4 indicators, each of which involves a separate processing algorithm.

The prevalence (SDO) will be determined by analyzing social networks with a search by keywords selected by the user of the model being developed. Determining the geolocation of the posts (where possible) will allow us to calculate the number of regions where the social disaster is mentioned. If this number reaches 10 (Europe, Asia, North America, Central America, South America, Australia, Oceania, North Africa, South Africa and Central Africa), the algorithm will indicate the prevalence at 100%, or 1.

After consulting with the expert group, it was decided to limit the maximum possible time of the active phase of the social shift to 365 days. If more time has passed since the first news about the described disaster, the duration indicator (SDD) will be set at 1, if less time has passed, it will be determined as a fraction of the selected maximum.

Sector-specific features (SDF) are processed using sentiment analysis to determine how prepared the entity responsible for a particular project decision is for a social disaster, while the indicator also covers the reaction of the population. The Ukrainian language is inherently polymorphic, which complicates the classical processing process. However, we will consider this a limitation for the proposed model. The general processing algorithm is as follows:

- cleaning the text description from words without a significant lexicographic load;
- creating a dictionary of key lemmas and finding the frequency characteristics of each word form;

- determination of the polarity of each word and correction of frequency values;
- aggregation and subsequent normalization of the obtained data in the range from 0 to 1, where 0 means the situation has a negative emotional description, 1 means a positive one.

The Severity Level (SDS) is a subjective numerical indicator set by the user of the developed algorithm in the range from 0 to 100 (integer values only). The final result is then normalized.

The four indicators are combined as follows:

$$SDP = \frac{SDO \times SDS \times SDD}{SDF} \quad . \tag{3}$$

The Target Audience Profile (TAP), in addition to the indicators outlined in the previous section, also takes into account the SDP indicator calculated using formula (3).

The maximum possible size of the target audience (TAS) was set at 100,000 people. It will be processed similarly to the SDD indicator.

The need to adjust behavior based on existing neoclassical economic paradoxes (TAX), which imply an increase in demand in a crisis, is a binary indicator (0 – no need to apply, 1 – need to apply). This indicator is determined by the model user, as well as the degree of trust (TAT), for which the limits are set similar to SDS.

The algorithm for processing the textual description of society (TAF) is similar to the above, but instead of determining the polarity of each lemma, the words are sorted according to the concept of emotions proposed by Robert Plattschik. The result is aggregated and normalized in the range from 0 to 1, where 0 means that negative emotions prevail in the target audience, and 1 means that positive emotions prevail. The general formula is as follows:

$$TAP = \left(\frac{TAF \times TAT \times (1 + TAX)}{TAS} \right)^{SDP} \quad . \tag{4}$$

The Business Environment Profile (BEP), according to the described methodology, consists of three indicators. At the same time, the SDP adjustment should also be taken into account.

The global economic system financial stability indicator (BWFS) is defined as a weighted average of three elements expressed in shares:

- changes in global GDP relative to the beginning of the social shift;
- changes in the S&P 500 index relative to the beginning of the social shift;
- changes in prices for basic energy resources.

According to the macroeconomic theory proposed by Friedman, when forecasting one's own business activities, it is necessary to consider the GDP of the country in which these activities are carried out. Although the solutions aimed at implementing the concept of "smart cities" are relatively isolated, the current level of globalization indicates the need to adjust the assessment of stability in relation to the world as a whole.

The energy price index has a direct impact on the logistics of any company and, accordingly, on the prices of their products. Since the data on products or the consumer price index will not necessarily be targeted in forecasting, it is necessary to take into account the impact of the company's reactions to the growth in the cost of contracts with counterparties, which is why this indicator is considered.

The local economic system stability indicator (BLFS) has similar elements to the BWFS, which are extrapolated to a limited area:

- changes in regional GDP relative to the beginning of the social shift;
- changes in the consumer price index relative to the beginning of the social shift;
- the rate of devaluation of the target currency.

The latter element is needed to take into account the extent to which the national currency has depreciated both against itself and against the world's most influential currencies – dollar, euro, yuan, yen, and pound sterling.

Business readiness (BR) is defined as follows:

$$BR = \frac{s_t^2 \times IAI \times FSI}{HHI}, \tag{5}$$

where N is the number of companies in the selected market;

s_t is the market share owned by company t;

FSI is the company's financial stability indicator;

IAI is the company's innovation activity indicator;

HHI is the Herfindahl–Hirschman index.

The overall indicator is calculated using the following formula:

$$BP = \left(BWFS \times BLFS \times BR\right)\frac{1}{SDP} \ . \tag{6}$$

The final indicator of the information environment (IEA) is a reflection of the number of fabricated news items related to the topic of the target urban decision or related topics relative to the total amount of information. The classification is carried out using another RCNN model, the parameters of which are defined as follows:

- the kernel size is 3;
- step size is set to 1;

- the parameter for adding insignificant zeros will not be applied, as well as the offset parameter, in order not to lose the context of the news;
- the filter dimension is set to 5´5´1.

The above four indicators serve as external data. To test the possibility of using them, we conducted a Granger causality test [14]. It was found that the correlation between the values of one variable and the past values of another allows us to consider them as external indicators.

4 Parallelization

To implement parallelization, it was decided to use MapReduce technology, which consists in dividing the original data set into separate nodes. Based on this idea, the key elements of the built model are the mapping and reduction functions. The most popular implementations are Spark-based and Hadoop-based. In this paper, we chose the second option, which has an additional pair of similar functions, but within each node, to speed up the interaction with databases [15]. This is a positive feature of the chosen approach, given the large amount of diverse information that needs to be processed.

The proposed solution can be presented as shown in Figure 2.

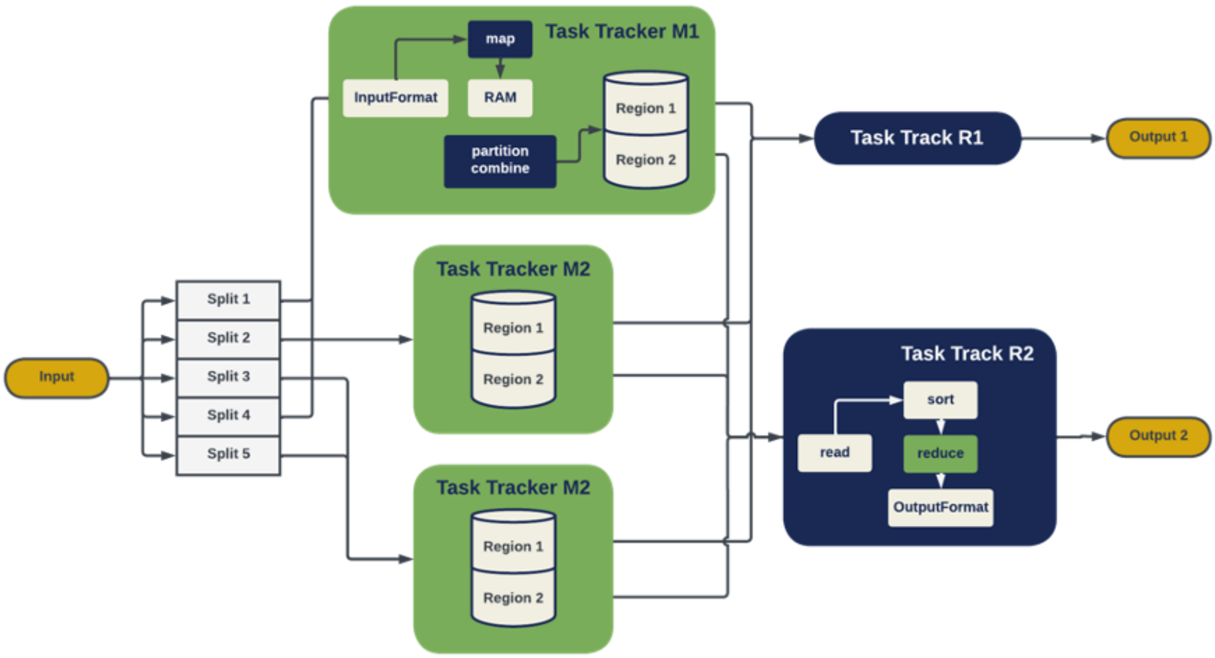

Fig. 2. Diagram of MapReduce technology based on Hadoop

For the RCNN architecture, the first step is the CNN layer. In this layer, the weights are iteratively adjusted by calculating their partial gradients after each set of training data is propagated through the network. Thus, parallelization during the training phase can be achieved by dividing the data into several chunks. Then, each data chunk is passed to multiple CNNs, and each CNN is trained independently in parallel. After that, the results are aggregated using a reducer to obtain the final results, which are then used to update the weights for the next iteration.

After the CNN layer is finished, the aggregated data is transferred to a bidirectional recurrent neural network. To speed up the process, you can divide the work of two neural networks between two nodes. In this case, the reduction function will actually serve as a function of aggregating the results of the two networks.

In the case of vector autoregression, although the overall calculation result depends on all the data, parallelization can be achieved by distributing the window load. The calculation of the integrated average can be performed on individual nodes and then aggregated. This will avoid the need to wait for the most time-consuming (in terms of CPU time) tasks to be completed.

5 Testing the approach

To test the effectiveness of the proposed approaches, the implementation was carried out on a stable environment. In terms of parallelization, the nodes copied the local hardware, and their number was set to 4.

The datasets for checking news falsification were created in-house based on the processing of news related to the implementation of an electronic ticket in Kharkiv and the introduction of the "smart city" concept in Kyiv. Data for forecasting with project targets were also generated semi-automatically. The following were considered as targets: the dynamics of expenses and income; the level of involvement of the target audience; and the efficiency of intermediate tasks.

The obtained values were combined into 3 separate data sets.

After expert evaluation, the following indicators were selected as key performance criteria:

- accuracy with an importance coefficient of 16;
- saving the time of the target algorithm (taking into account preprocessing) with an importance factor of 8;
- saving the minimum allowable amount of target data to achieve "accuracy" of more than 80% with an importance factor of 4.

The weighting factors for linear additive convolution are calculated based on the importance coefficients.

Since the issue at hand is prediction, not classification, accuracy will be measured using the normalized inverse root mean square error. The time savings will also take into account the parallelization proposed above, in order to offset the loss that accompanies the use of the data preprocessing algorithm. The minimum allowable volume savings are measured in terms of the number of elements in the time series and normalized relative to the extreme values. Figure 3 shows the results of 5 measurements of the target algorithm's time savings, rounded to whole seconds for normal attempts and to tenths in the case of the average value.

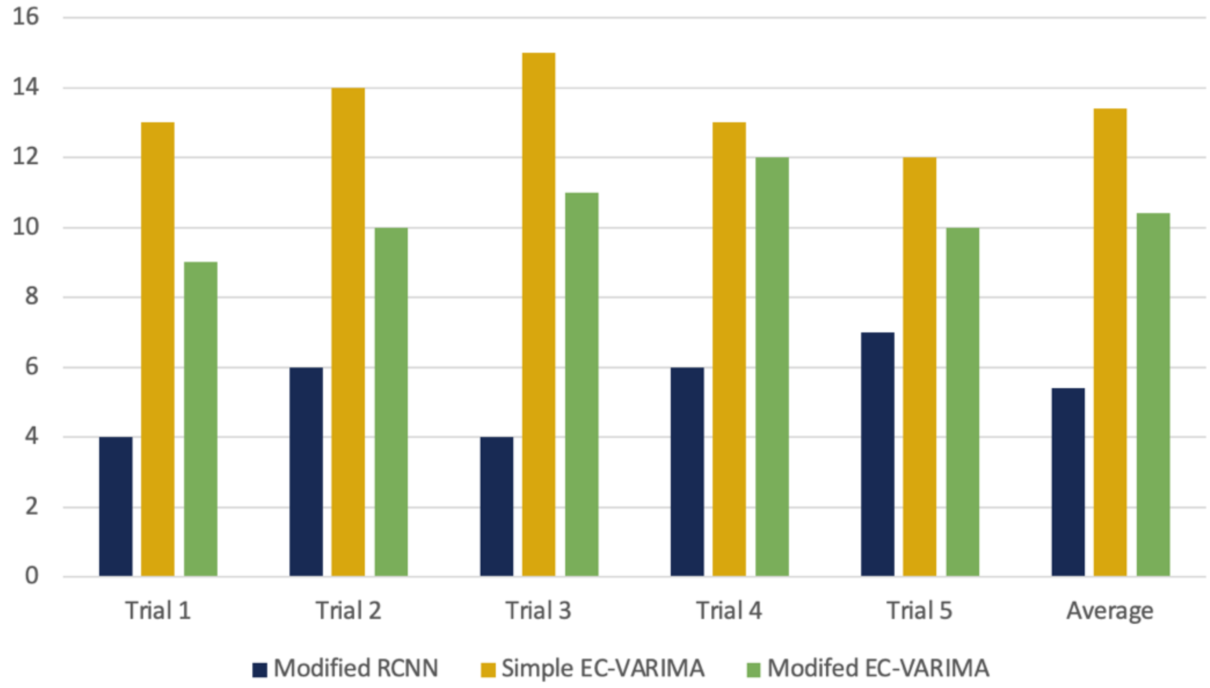

Fig. 3. Results in terms of time savings

As can be seen from Figure 3, the fastest algorithm is the simple EC-VARIMA, followed by the modified one. This result is achieved due to the parallelization of the model itself and all the preprocessing steps. For the accuracy indicator, the situation is different: the most accurate algorithm (as shown in Figure 4) is the modified RCNN. At the same time, one can notice the instability of the basic algorithms without taking into account external indicators. This corresponds to the hypothesis mentioned above.

The final indicator is data volume savings. It should be noted here that the two baseline models did not achieve the required result of the minimum accuracy result when gradually increasing to 500000 elements. Therefore, for these algorithms, the savings value is 0. For the modified RCNN, the minimum allowable value is 50,000 elements, and for the modified EC-VARIMA, it is 100,000.

These results can be presented in the form of the following Table 1, taking into account their normalization and rounding to the hundredths.

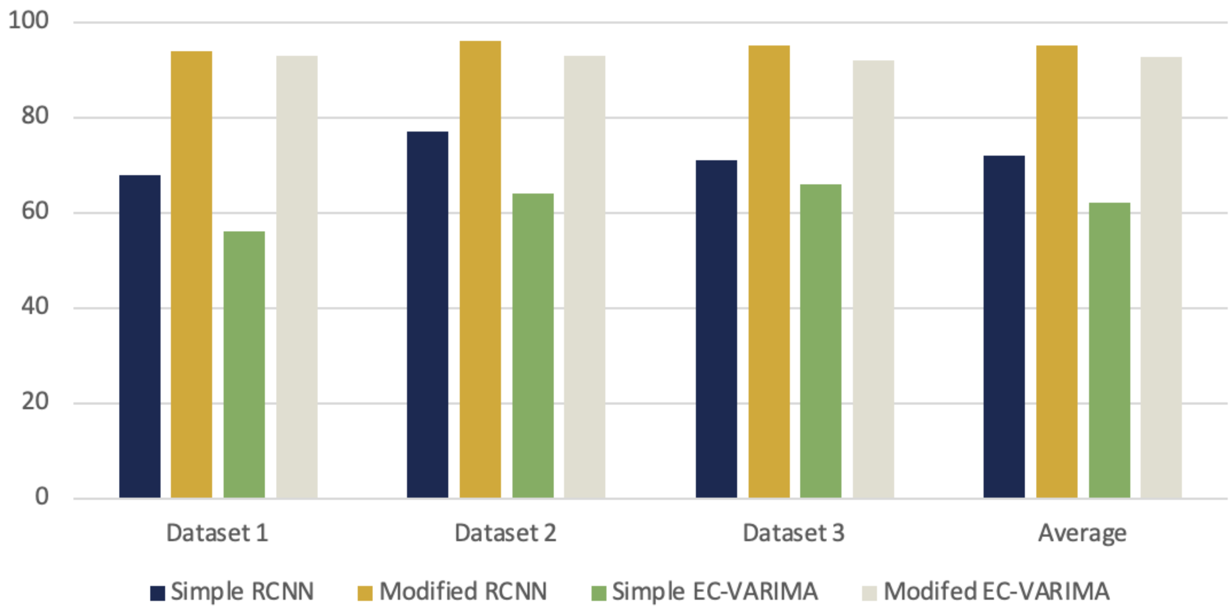

Fig. 4. The result in terms of "accuracy"

Table 1

Processed results of the experiment

Model	Accuracy	Time saving	Data saving
Simple RCNN	0.72	0.00	0.00
Simple EC-VARIMA	0.62	1.00	0.00
Modified RCNN	0.95	0.40	0.90
Modified EC-VARIMA	0.93	0.78	0.80

Based on the results obtained, we calculated the value of linear additive convolution with weighting coefficients. For Simple RCNN, the value was 0.41, for Simple EC-VARIMA – 0.64, for Modified RCNN – 0.79, and for Modified EC-VARIM – 0.87.

6 Conclusions

The analysis of the industry allowed us to identify a hybrid neural network consisting of a combination of recurrent and convolutional subnets, as well as vector autoregressive integrated moving average as simple models that do not require significant computing power. To be able to take into account the impact of social shifts on the process of developing and implementing targeted project solutions, 4 key indicators were identified

- profile of the social disaster
- profile of the target audience,;

- profile of the business environment (both global and local);
- the information environment, which reflects the intensity of the impact of fabricated information on the target concept.

The use of algorithms for their processing causes the problem of speed. To mitigate this problem, we used Hadoop-based MapReduce technology. Experimental results have shown that it has reduced the speed gap between simple and modified models. In addition to saving time, accuracy and saving the minimum allowable amount of data to achieve 80% accuracy were also considered as target indicators.

Taking into account the obtained values of linear additive convolution, it can be noted that the proposed approach to parallelization and pre-processing of external data gives the desired result, increasing the efficiency of using simple models. At the same time, the forecast accuracy is high for both modified algorithms. However, due to its simplicity, the modified EC-VARIMA is more effective given the selected set of indicators.

Thus, the approach described above allows for effective forecasting of the economic performance of information-sensitive socially oriented systems during emergencies. This, in turn, may allow for a timely response to crisis phenomena and changes in management policy.

References

1. G2, Best IT Risk Management Software, 2024. URL: https://www.g2.com/categories/it-risk-management
2. E. Flaherty, T. Sturm, E. Farries, The conspiracy of Covid-19 and 5G: Spatial analysis fallacies in the age of data democratization, Soc Sci Med. 293 (2022) no. 114546. DOI: 10.1016/j.socscimed.2021.114546
3. P. Marcelo, Conspiracy theories falsely tie Maui wildfires to 'smart cities' and tech conferences, AP News, 2023. URL: https://apnews.com/article/fact-check-maui-hawaii-wildfires-smart-cities-387327837046
4. U. A. Butt, R. Amin, M. Mehmood, H. Aldabbas, M. T. Alharbi, N. Albaqami, Cloud Security Threats and Solutions: A Survey, Wireless Personal Communications. 128 (2022) pp. 387–413. DOI: 10.1007/s11277-022-09960-z
5. K. Srinivas, Process of Risk Management, IntechOpen, 2019. DOI: 10.5772/intechopen.80804
6. S. Yakovlev, A. Khovrat, V. Kobziev, Using Parallelized Neural Networks to Detect Falsified Audio Information in Socially Oriented Systems, in: Proceedings of the International Scientific Conference "Information Technology and Implementation", IT&I '23, Kyiv, Ukraine, 2023, pp. 220–238. URL: https://ceur-ws.org/Vol-3624/Paper_19.pdf

7. R. Deng, F. Duzhin, Topological Data Analysis Helps to Improve Accuracy of Deep Learning Models for Fake News Detection Trained on Very Small Training Sets. Big Data and Cognitive Computing. 6 (3) (2022) no. 74. DOI: 10.3390/bdcc6030074
8. A. Choudhary, A. Arora, Linguistic feature based learning model for fake news detection and classification. Expert Systems with Applications. 169 (2021) no. 114171. DOI: 10.1016/j.eswa.2020.114171
9. R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, J. Ortega-Garcia, Deepfakes and beyond: A Survey of face manipulation and fake detection. Information Fusion. 10 (11) (2020) pp. 131–148. DOI: 10.1016/j.inffus.2020.06.014
10. M.A. Alonso, D. Vilares, C. Gómez-Rodríguez, J. Vilares, Sentiment Analysis for Fake News Detection. Electronics. 10 (11) (2021) no. 1348. DOI: 10.3390/electronics10111348
11. A. Breuer, R. Eilat, U. Weinsberg, Friend or Faux: Graph-Based Early Detection of Fake Accounts on Social Networks, in: Proceedings of the Web Conference, WWW '20, Association for Computing Machinery, Taipei, Taiwan, 2020, pp. 1287–1297. DOI: 10.1145/3366423.3380204
12. T. Xia, X. Chen, A Discrete Hidden Markov Model for SMS Spam Detection. Applied Science. 10 (14) (2020) no. 5011. DOI: 10.3390/app10145011
13. M. Akkaya, Vector Autoregressive Model and Analysis. Handbook of Reseach on Emerging Theories. Springer. (2021) pp. 197–214. DOI:10.1007/978-3-030-54108-8_8
14. T. Xia, X. Chen, Granger Causality: A Review and Recent Advances. Annual Review of Statistics and Its Application. 9 (2022) pp. 289-319. DOI: 10.1146/annurev-statistics-040120-010930
15. A. Khovrat, V. Kobziev, Using Recurrent and Convulation Neural Networks to Indentify the Fake Audio Messages, in: Proceedings of the International Conference on Methods and Systems of Navigation and Motion Control, MSNMC '23, IEEE, Kyiv, Ukraine, 2023, pp. 174–177. DOI: 10.1109/MSNMC61017.2023.10329236