

Decision support algorithm in the development of information sensitive socially oriented systems

Sergiy Yakovlev^{1,2}, Artem Khovrat³, Volodymyr Kobziev³ and Dmytro Uzlov²

¹ Lodz University of Technology, 90-924 Lodz, Poland

² Institute of Computer Science and Artificial Intelligence, V.N. Karazin Kharkiv National University, 4, Svobody, Sq., Kharkiv, 61022, Ukraine

³ Kharkiv National University of Radio Electronics, 14, Nauky, Ave., Kharkiv, 61166, Ukraine

Abstract

The increase in the amount of data and the general change in social and market processes lead to the transformation of basic management principles. Under nondeterministic conditions, when developing solutions related to socially orientated systems forecasting of possible problems requires the use of modern tools of intelligent data analysis. Existing ready-made concepts cannot guarantee high efficiency during social shifts exacerbated by a falsified information, that is, during discrediting campaigns directed against the ideas proposed by the business. At the same time, there is a problem with the speed of simple solutions and the high cost and safety of using more complex cloud technologies. The current work is focused on the modification of simple decision support models and the construction of a data reprocessing algorithm to increase the accuracy and reliability of project activity forecasting. The proposed sequence of steps includes the definition of four profiles – social shift, target audience, business environment and information environment. The second of them opens up the possibility of taking into account the features of the information environment in which the deployment of the developed system is proposed. It is based on an algorithm for detecting text fake news through the use of complex neural networks, and the principles of content analysis, which focus on three components – the profile of the emergency situation, the target audience, and the business environment. An experimental comparison of simple autoregressive and neural network with the proposed solution allows the authors to demonstrate the higher efficiency of the created algorithm, both in terms of accuracy and speed due to parallelisation based on MapReduce technology. This, in turn, paves the way for further test implementation in a real environment in order to expand the possibilities of forecasting and taking into account risks for information sensitive socially orientated projects.

Keywords

Behavioural analysis, data analysis, forecasting, neural networks, parallelisation, falsified news

1. Introduction

One of the most important preparatory stages in project management in any field is the identification of key risks and ways to mitigate them. Under modern conditions, this task becomes complex and requires the processing of a large amount of historical data. This state of affairs became the basis for the development and popularisation of decision support systems (DSS), which are capable of simply aggregating received information or making a forecast, as well as, in some cases, providing recommendations on the possibilities of solving problems. Despite this, most DSSs, based on available feedback on specialised resources, are not effective enough in uncertainty conditions, especially when the target projects are concerned with the activities of social groups, such as solutions for the development of "smart cities" [1]. The activation 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, actualises the issue of improving the algorithms that can be the basis of the described systems. At the same time, when building a highly intelligent social infrastructure

¹ProfIT AI 2024: 4th International Workshop of IT-professionals on Artificial Intelligence (ProfIT AI 2024), September 25–27, 2024, Cambridge, MA, USA

✉ sergiy.yakovlev@p.lodz.pl (S. Yakovlev); artem.khovrat@nure.ua (A. Khovrat); volodymyr.kobziev@nure.ua (V. Kobziev); dmytro.uzlov@karazin.ua (D. Uzlov);

ORCID 0000-0003-1707-843X (S. Yakovlev); 0000-0002-1753-8929 (A. Khovrat); 0000-0002-8303-1595 (V. Kobziev); 0000-0003-3308-424X (D. Uzlov)



© 2024 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

(such as urban solution), this problem is felt even more acutely. Not least because of the influence of falsified information on people's behaviour. As an example, the campaign against the installation of a more modern Internet based on 5G technology, which is definitely necessary for the more reliable operation of all the latest infrastructure solutions, can be mentioned [2]. Another similar example can be considered the discrediting of the "smart city" concept by linking it with forest fires on Maui [3].

An 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 the deployment of complex software products, the issue of their reliability and safety remains unresolved [4].

Taking into account the above, a decision was made within the framework of this work to modify existing DSS algorithms to increase their effectiveness in risk management for socially orientated systems. Such a modification should not only note the essence of social shifts or the behaviour of the audience, but also the possibility of the influence of falsified information aimed at discrediting the developed solutions. At the same time, the speed of these algorithms should be sufficient for their practical application.

To achieve the goal set, the following tasks were selected.

- review of the existing solutions that are at the basis of modern DSSs and identification of their key shortcomings;
- development of a data reprocessing algorithm that would allow taking into account the change in the behaviour of the target audience and the impact of fabricated news on it;
- determination of the basic algorithm for forecasting the company's activity indicators within the framework of the target project and possibilities for its further improvement;
- study of possibilities and implementation of a parallelised version of the algorithm;
- determining the effectiveness of the proposed solution by conducting an experiment and solving the problem of linear optimisation.

2. Domain exploration

Within the framework of this section, the results of the analysis of the key algorithms embedded in decision support systems and the determination of the characteristics of target indicators of audience behaviour, information environment, social shift, and business environment are presented.

2.1. Finding the basic algorithm

In general, it is customary to divide the DSS according to two criteria [5]:

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

Considering the fact that it is necessary to concentrate on simple solutions that do not require significant hardware capabilities and on the development of an information reprocessing algorithm, in the future only data-orientated passive DSSs will be considered. 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, the most popular systems implemented in the risk management process are Hyperproof, Soterion, Whistic [1].

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

- prognostic algorithms are based on solutions based on neural networks and/or autoregression models. It is worth noting that some other DSSs are also able to use the

probabilistic approach; however, it works more slowly and requires the use of cloud computing;

- There is limited consideration of the risks associated with the activities of the target audience of the developed systems; in particular, problems were observed with the forecast of business activity during the COVID-19 pandemic were noticed;
- lack of consideration of the information environment in which the target project is created, which is critical in the development of socially oriented solutions.

The defined classes of basic predictive algorithms are relatively broad and require additional filtering. In particular, in the case of neural networks, it is possible to consider the following.

- convolutional neural networks (CNN), which use the "convolution" function to reduce the dimensionality of input data;
- recurrent neural networks (RNNs), which are based on the reuse of the processing results of the previous layer;
- combined hybrid networks (generalised RCNN), which combine several simple models, providing higher accuracy.

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

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

An additional study of the problem of forecasting economic indicators using international scientific works allows limiting the set of basic forecasting algorithms to a combined hybrid neural network and autoregression of an integrated moving average. Although they are the slowest, they have the highest accuracy of those proposed; they are also capable of parallelisation and do not require significant hardware power.

2.2. Identification of target indicators features

To mitigate the risks that may accompany the development and implementation of a socially orientated software project, it is necessary to define a set of key aggregated indicators. Based on what was stated in section 1, such general indicators can serve as:

- social shift profile, a parameter that regulates the uncertainty of conditions conversion into a numerical form;
- profile of the target audience, summarising the behaviour of the most influential targeted subjects;
- a profile of the business environment, which determines the specifics of the implementation of a specific project in the market-conjunctural context;
- information environment, an indicator that reflects the intensity of the influence of fabricated information on the target concept.

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

When studying the concept of social shift, which is also called a "social disaster", it was found that the most influential subindicators are the prevalence of the source of the shift, its duration (taking into account the moment of the first information appearance), specifics for a specific field of activity, and the level of acuteness. The first two indicators are by their nature objective numerical variables; the other two reflect the subjective perception of the shift and require additional algorithms to be able to use them in forecasting.

The profile of the target audience can be determined taking into account the size of this audience, the market paradoxicality of the target decision (if it falls under the influence of known

neoclassical economic paradoxes), the degree of trust, and the general description of society. As in the first case, there is a mixing of objective numerical indicators with text indicators that are subjective for the target project.

The profile of the business environment focusses on the reaction of business entities both to the implementation of the proposed project solution in general and to the social disaster 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 emergency situations. Although the last indicator in most cases does not directly affect the systems focused on the social aspect, it allows one to adjust the objective assessment of the financial stability of the city with the subjective perception of its internal counterparties.

The information environment, within the context of current work, refers to the intensity of the spread of false news regarding the topic of the project, technological reforms, and other similar domains of knowledge. Since it was decided to focus on textual data, the following characteristic features can be determined [6, 7]:

- using an unnatural number of rhetorical questions (contextual distortion of socially significant topics). Conducted linguistic studies show that in official business, journalistic styles for use by mass media, this type of speech construction is not often used;
- absence of negative constructions to reduce cognitive load in combination with pessimistic colours of selected words. As an example, replace the word "trouble" with "catastrophe". Here, it is worth noting that profanity will be deliberately removed from further texts, as this complicates the process of analysing emotional colouring;
- using appeals and encouragements in an inappropriate context and using an unreasonable number of pronouns. In this case, there is an imitation of a journalistic presentation style;
- high frequency of using short sentences and words with grammatical errors.

These characteristics are not exhaustive, but it should be emphasised that the target texts for consideration, in addition to profanity, will also not consider the mixing of several languages and the use of regional dialects. Similar add-ons go beyond the described tasks.

In general, the field of determining the falsity of information is not new, as already noted above. When researching fake text news by several groups of European scientists, it was shown that machine learning algorithms built on the basis of both neural networks and more modern transformers or autoencoders require significant amounts of data to achieve an accuracy of more than 90% [8, 9]. However, this problem can be solved by using a balanced data set, which was demonstrated in the work of Ukrainian researchers.

Another well-known way to determine the fact of data falsification is the use of graph models [10]. In their work, scientists at Harvard demonstrated their high efficiency in solving the problem of detecting fake accounts. However, this method will not allow to fully process textual information from the news, or will require more substantial capacities for processing. A similar problem concerns algorithms that help detect spam. A Chinese-American scientific research group proved the possibility of an effective application of Markov networks [11]. However, taking into account the specifics of the area and the set goal, its application is quite cumbersome and will require the use of cloud technologies. A similar problem has already been considered for the basic prognostic algorithm [12].

3. Mathematical representation

Within the framework of this section, the key features of the proposed basic algorithms and target indicators of refinement will be presented.

3.1. Core algorithms

As mentioned above, the current work considers two basic algorithms – RCNN and vector autoregression of the moving average. Vectority of the latter is necessary for the possibility of simultaneous processing of several indicators.

In a simple CNN model, passing the filter allows to take into account the environment of each element; however, the specifics of the proposed indicators require an understanding of a longer time span without a fundamental shift into the future. Thus, the salient context may be outside the CNN model filter. To avoid this problem, it was decided to combine RNN and CNN.

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

To be able to take into account the context to the fullest extent, it was decided not to use a simple RNN architecture, but a bidirectional recurrent neural network with support for long- and short-term memory. It is based on the use of hyperbolic tangents and sigmoids, which avoid the problem of explosive and vanishing gradients by limiting the range of resulting values. At the same time, the defined model allows us to take into account the entire historical context. Thus, the RCNN architecture can be presented as shown in Figure 1.

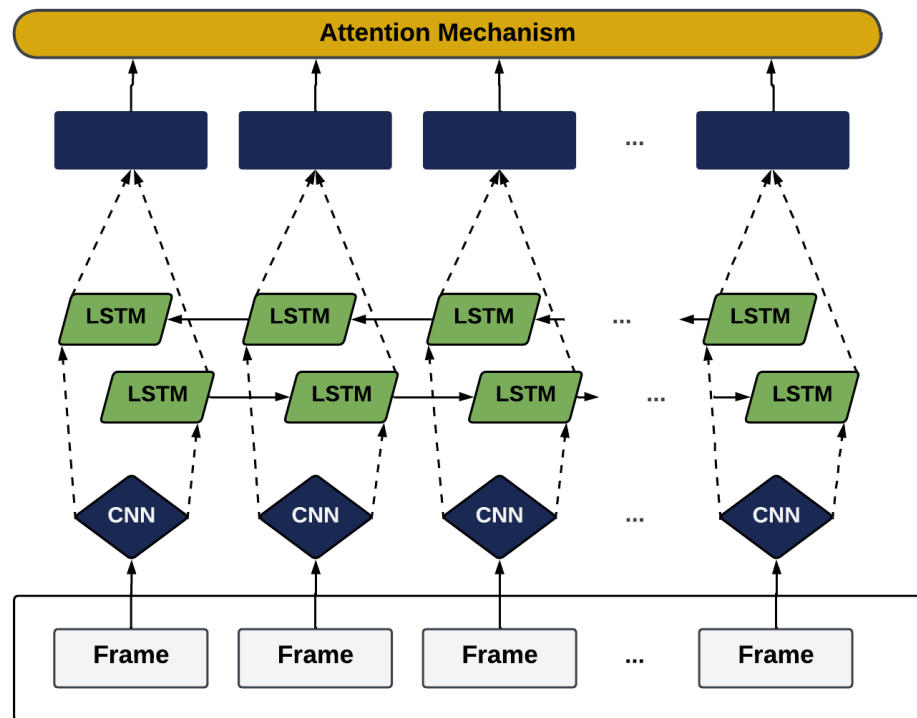


Figure 1: Scheme of the RCNN architecture.

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

- the core size is determined as 4;
- step size is defined as 1;
- based on the set step, the option of adding nonsignificant zeros will not be applied;
- based on the specifics of the subject area, it was decided not to apply the displacement parameter;
- the dimension of the filter is set to $5 \times 5 \times 3$ (the last value of the dimension is determined by the number of target indicators).

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 presented in the following form:

$$\Phi_0 y_t = \sum_{i=1}^p \Phi_i y_{t-i} + \sum_{j=0}^q \Theta_j u_{t-j}, \quad (1)$$

where y_t – K-dimensional time series; Φ_i, Θ_i – matrices $K \times K$, $i = \overline{1, p}$, $j = \overline{1, q}$; u_t – K-dimensional vector of white noise with zero mean and the following non-degenerate covariance matrix $\sum E(u_t, u_t')$.

It follows from the above formula (1) that the classic family of VAR models predicts only static variables. To take into account exogenous indicators, it was decided to use error correction modification (EC). A similar adjustment is necessary if 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 will have the following form:

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

where $\Pi = -\Phi_0 + \sum_{i=1}^p \Phi_i$; $\Psi_i = -\sum_{j=i+1}^p \Phi_j$, $i = \overline{1, p-1}$.

The integration and mobility of the selected target model will ensure the possibility of taking into account the target element of the time series. And in this case the observed problem is similar to the one defined for CNN, the possibility of taking into account the historical context in full within the framework of the family of autoregressive models is not provided.

3.2. Target indicators

The social disaster profile (SDP), as already mentioned, is divided into 4 indicators, each of which involves a separate processing algorithm.

The prevalence (SDO) will be determined using social network analysis with keyword searches selected by the user of the model being developed. Determining the geolocation of sending posts (where possible) will allow to count the number of regions in which 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 mark the prevalence as 100%, or 1.

After consultations 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 of the described disaster, the duration indicator (SDD) will be set to 1, if less, defined as a fraction of the division of the chosen maximum.

Features for a specific area (SDF) are processed using sentiment analysis, which allows to determine how much the entity responsible for a certain project decision is ready for a social disaster, at the same time, the indicator covers the reaction of the population. The Ukrainian language is inherently polymorphic, which complicates the classical processing process. However, this will be considered as a limitation of the proposed model. The general processing algorithm is as follows:

- cleaning of the textual description from words without 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 further normalisation of the obtained data within the range from 0 to 1, where 0 - the situation has a negative emotional description, and 1 - positive.

The severity level (SDS) is a subjective numerical indicator set by the algorithm user of the developed in the range of 0 to 100 (integer values only). After that, the final result is normalised.

These four indicators are combined as follows:

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

The profile of the target audience (TAP), in addition to the indicators outlined in the previous section, also takes into account the SDP indicator calculated according to formula (3).

The maximum possible size of the target audience (TAS) was established at the level of 100,000 people. Processing will be carried out similarly to the SDD indicator.

The need to correct behaviour based on existing neoclassical economic paradoxes (TAX), which predict an increase in demand in crisis situations, is a binary indicator (0 - no need to apply, 1 - need to apply). This indicator is defined by the user of the model, as well as the degree of confidence (TAT), for which limits are set similarly 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 Plutchick. The result is aggregated and normalised within the range from 0 to 1, where 0 – the target audience is dominated by negative emotions, 1 – positive. The general formula has the following form:

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

The business environment profile (BP), according to the described methodology, consists of three indicators. At the same time, the SDP correction must also be taken into account.

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

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

In the macroeconomic theory proposed by Friedman, when forecasting one's own business activity, it is necessary to consider the GDP of the country in which this activity is carried out. The current level of globalisation indicates the need to correct the assessment of stability in relation to the world in general.

The indicator of the level of energy resources directly affects the logistics of any company and, accordingly, the prices of their products. Since the data on products or the consumer price index will not necessarily be targeted when forecasting, it is necessary to take into account the impact of the company's reactions to the increase in the value of contracts with counterparties; for this purpose, the specified indicator is considered.

The indicator of stability of the local economic system (BLFS) has elements similar 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;
- rates of devaluation of the target currency.

The last element is needed to take into account how much the national currency depreciated, both relative to itself and relative to the world's most influential currencies, the dollar, euro, yuan, yen, and pound sterling.

Business emergency preparedness (BR) is defined as follows:

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

where N is the number of companies in the selected market; s_t is the market share belonging to company t ; FSI is an indicator of the company's financial stability; IAI is an indicator of the company's innovative activity; HHI – is the Gerfindahl-Girschman index.

The general indicator is calculated using the following formula:

$$BP = (BWFS \times BLFS \times BR)^{\frac{1}{SDP}} \quad (6)$$

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

- the core size is 3;
- the step size is defined at level 1;
- the parameter of adding insignificant zeros will not be applied, as well as the offset parameter, in order not to lose the context of the news;
- the dimension of the filter is set to $5 \times 5 \times 1$.

The four above indicators serve as external data. To verify the possibility of its use, the Granger causality test was performed [14]. It was established that the correlation between the values of one variable and the past values of another allows them to be considered external indicators.

4. Parallelisation of core algorithms

To implement parallelisation, it was decided to use the MapReduce technology, which consists of 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 in use are implementations based on Spark and Hadoop. In this work, the second option was chosen; it has an additional pair of similar functions, but inside 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.

Graphically, the proposed solution can be presented as shown in Figure 2.

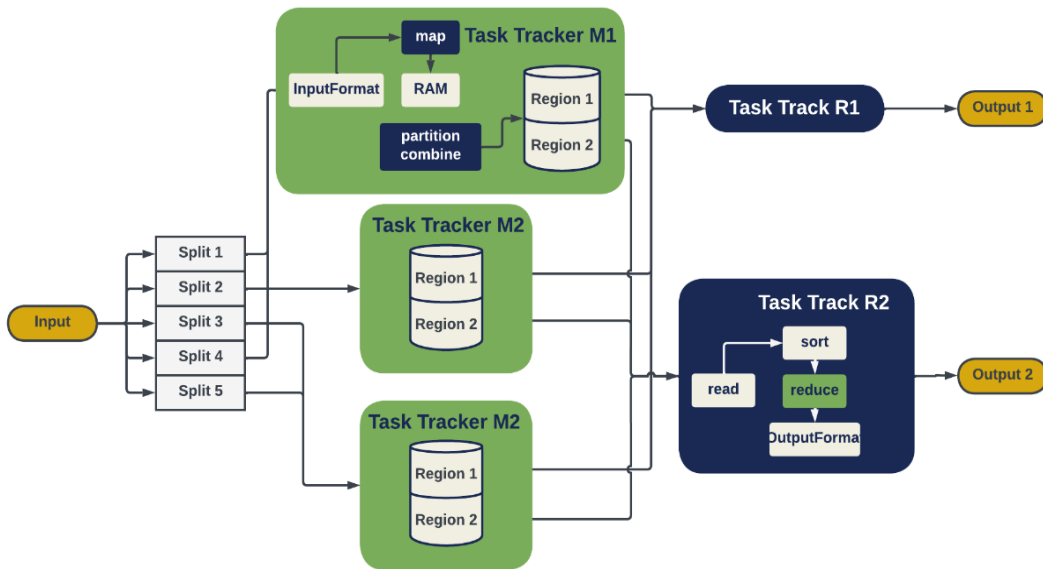


Figure 2: Scheme of the MapReduce architecture based on Hadoop.

For the RCNN architecture, the first step is the CNN layer. It iteratively adjusts the weights by computing their partial gradients after each set of training data is propagated through the network. Thus, parallelisation during the training phase can be achieved by dividing the data into several fragments. Each piece of data is then fed into multiple CNNs, and each CNN is trained independently in parallel. The results are then aggregated using a reducer to produce the final results, which are then used to update the weights.

After the CNN layer is completed, the aggregated data is transferred to a bidirectional recurrent neural network. To speed up the process, the execution process of two neural networks can be distributed between two nodes. In this case, the reduction function will actually serve as an aggregation function of the results of the two networks.

In the case of vector autoregression, although the overall result of the calculation depends on all the data, parallelisation can be carried out using window load distribution. The calculation of the integrated average can be performed on individual nodes and then aggregated. This will avoid the need to wait for the execution of the most expensive (from the point of view of processor time).

5. Experiment overview

To check the effectiveness of using the proposed approaches, the implementation was carried out in a stable environment. Regarding parallelisation, nodes copied local hardware, their number was set to 4.

The datasets for checking the falsification of news were created by our own hands on the basis of the processing of news related to the implementation of the electronic ticket in Kharkiv and the implementation of the "smart city" concept in Kyiv. Data for forecasting with project targets were also generated semi-automatically. The following were considered as target indicators: the dynamics of costs and revenues; the level of involvement of the target audience; performance indicator of intermediate tasks.

The resulting values were combined into three separate data sets.

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

- accuracy with an importance factor of 16;
- saving the time of the target algorithm with an importance factor of 8;
- saving the minimum permissible amount of target data to achieve "accuracy" of more than 80% with an importance factor of 4.

The weights for the linear convolution are calculated on the importance coefficients.

Since the problem is prediction rather than classification, the accuracy will be determined using the normalised inverse root mean square error. The saving of processing time will also take into account the parallelisation proposed above, in order to level the loss that accompanies the use of the data reprocessing algorithm. Savings of the minimum permissible volume are measured in the number of elements of the time series and are normalised relative to extreme values. Figure 3 shows the results of five time savings measurements of the target algorithm, rounded to whole seconds for normal attempts and to tenths for the average value (the simple RCNN approach was the slowest).

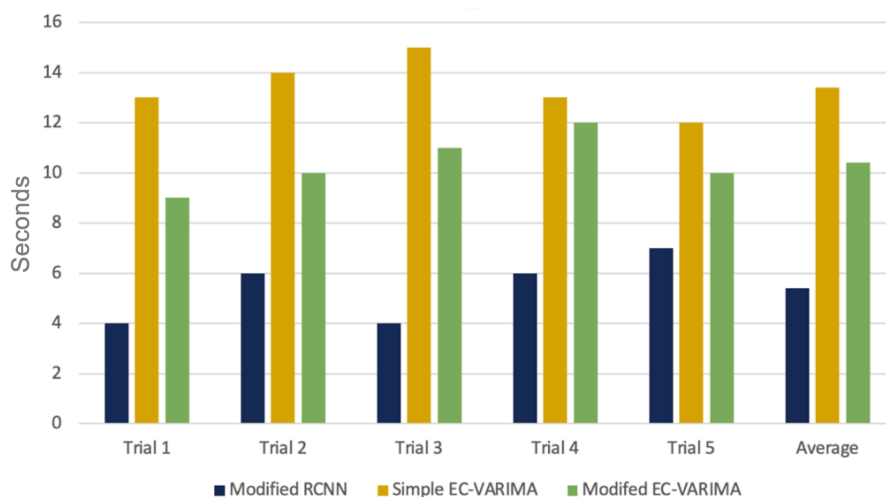


Figure 3: Results for time-saving indicator.

As can be seen in Figure 3, the fastest algorithm is the simple EC-VARIMA, followed by the modified one. The result obtained is achieved as a result of the parallelisation of the model itself and all steps of refinement. For the accuracy indicator, the situation is different; the most accurate algorithm (as can be seen in Figure 4) is the modified RCNN. At the same time, it is possible to notice the instability of the basic algorithms without taking into account external indicators. This corresponds to the hypothesis stated above.

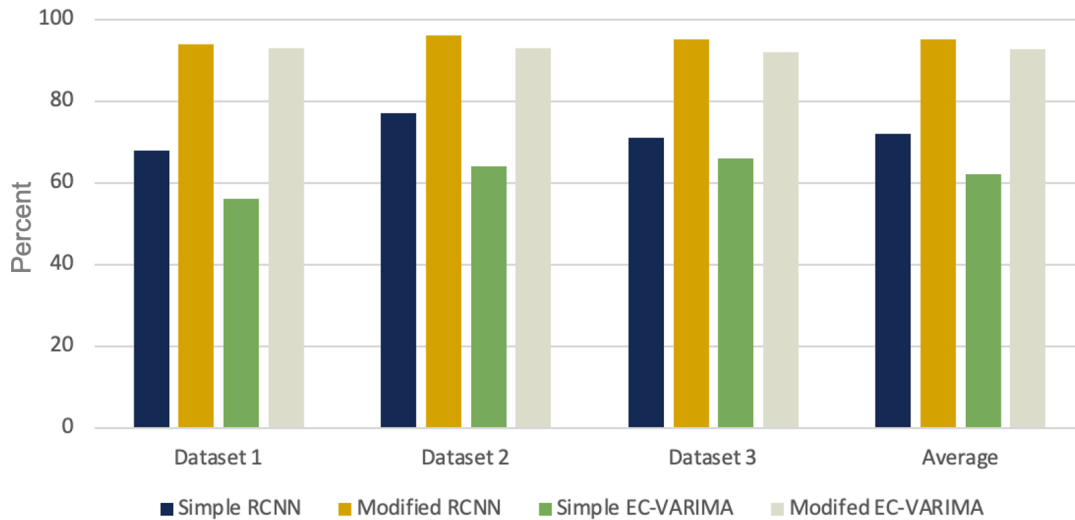


Figure 4: Results for the indicator of ‘accuracy’.

The final indicator is data volume savings. It should be noted here that the two baseline models did not achieve the desired minimum accuracy result when incrementally increasing to 500,000 elements. Therefore, for these algorithms, the saving 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.

The specified results can be presented in the form of the following Table 1, taking into account their normalisation and rounding to hundredths:

Table 1

Processed results of the experiment

Model	Accuracy	Time-Saving	Volume-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, the value of the linear additive convolution with weighting coefficients was calculated. 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 aim of the current work was to modify the existing basic algorithms embedded in decision support systems for socially orientated systems under nondeterministic conditions.

The conducted analysis of the industry made it possible to identify as simple models that do not require significant hardware power, a hybrid neural network from a combination of recurrent and convolutional subnetworks, as well as vector autoregression of an integrated moving average. To

be able to take into account the impact of social shifts on the process of development and implementation of targeted project solutions, four key indicators were defined:

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

The use of algorithms for their refinement causes the problem of speed. MapReduce technology based on Hadoop was used for its levelling. As the results of the experiments showed, it made it possible to reduce the gap in speed between simple and modified models. In addition to time-saving, precision and the minimum allowed amount of data to achieve an accuracy of 80% were also considered as target indicators.

Given the values obtained for linear additive convolution, it can be seen that the proposed approach to the parallelisation and reprocessing of external data gives the desired result, increasing the efficiency of the application of 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.

The efficiency value allows to proceed to the next stage of testing the proposed solution in the real environment of the development of information sensitive socially orientated projects and taking into account the problems of determining the fact of falsification of information and other external indicators.

Acknowledgements

The authors would like to thank the Armed Forces of Ukraine for the opportunity to write a valid work during the full-scale invasion of the Russian Federation on the territory of Ukraine.

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] 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.
- [7] 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.
- [8] 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.
- [9] 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.

- [10] 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.
- [11] T. Xia, X. Chen, A Discrete Hidden Markov Model for SMS Spam Detection. Applied Science. 10 (14) (2020) no. 5011. doi: 10.3390/app10145011.
- [12] 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.
- [13] M. Akkaya, Vector Autoregressive Model and Analysis. Handbook of Research 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 Convolution Neural Networks to Identify 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.