A deep neural network model for predicting the competitive score of social projects for community development

Anatoliy Tryhuba^{1,*,†}, Inna Tryhuba^{1,†}, Oksana Malanchuk^{2,†} and Anna Marmulyak^{3,†}

¹ Lviv National Environmental University, 1, V.Velykoho str., Dubliany-Lviv, 80381, Ukraine

² Danylo Halytsky Lviv National Medical University, 69, Pekarska str., Lviv, 79017, Ukraine

³ Lviv State University of Life Safety, 35, Kleparivska str., 79007, Lviv, Ukraine

Abstract

The purpose of the study is to substantiate a model for predicting the competitive score of social projects for community development with the optimization of the architecture of a deep neural network trained on the prepared data from the implementation of previous projects. The proposed methodology for predicting the competitive score of social projects for community development is based on the use of deep neural networks and includes ten stages. These stages are reflected in the developed research algorithm. The peculiarity of the proposed methodology is that the database is formed on the basis of the collected data on the results of competitions for local initiative projects in the region. In the course of the research, 6 types of modern deep neural network architectures (FNN, RNN, LSTM, GRU, CNN, RCNN) were used. To evaluate the models of a given deep neural network architecture, such metrics as MSE, MAE, R2 Score, RMSE, Inverse RMSE, and ms/step were used. It was found that the best accuracy rates are provided by the model based on RCNN. It showed the best results among all models. It has the lowest values of MSE=6.962, MAE=2.11, and RMSE=2.638, and the highest R2Score=0.43, which indicates a better ability to explain variability in the original data. We have optimized the basic RCNN model using 4 options: 1) increasing the number of layers and neurons; 2) using another optimizer; 3) using Dropout regularization; 4) changing the loss function. It is established that the RCNN model using Dropout regularization is the best for predicting the quantitative value of the competitive score of social community development projects. Relative to the baseline RCNN model, there was a decrease in MSE by 1.22% and a decrease in MAE by 2.51%. Further research should be conducted in the direction of developing a decision support system for planning social community development projects from the proposed RCNN model using Dropout regularization to predict the competitive score of social community development projects.

Keywords

model, deep neural network, forecasting, competition score, social projects

MoMLeT-2024: 6th International Workshop on Modern Machine Learning Technologies, May, 31 - June, 1, 2024, Lviv-Shatsk, Ukraine

^{*} Corresponding author.

[†]These authors contributed equally.

[⊡] trianamik@gmail.com (Anatoliy Tryhuba); trinle@ukr.net (Inna Tryhuba); oksana.malan@gmail.com (Oksana Malanchuk); anya.marmulyak@gmail.com (Anna Marmulyak)

O000-0001-8014-5661 (Anatoliy Tryhuba); 0000-0002-5239-5951 (Inna Tryhuba); 0000-0001-7518-7824 (Oksana Malanchuk); 0000-0002-7526-5850 (Anna Marmulyak)

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

1. Introduction

In today's world, social projects for community development play an important role in improving the lives and activities of their populations [1-3]. In order to efficiently allocate resources and support the most promising initiatives, it is important to have accurate methods for predicting the success of social projects submitted for competitions. In recent decades, artificial intelligence has become a key tool in the development of society [4-6]. It can radically change a number of industries, including the social sphere and civic activities. In particular, the use of deep neural networks for forecasting is becoming increasingly common among project managers who select social projects to be submitted to a competition. Such competitions of social projects of local initiatives are held in the territory of individual communities and regions [7].

Since the threshold for a passing score fluctuates from year to year, the forecast results are the basis for a preliminary assessment by both the developers of the competition proposals and project managers of the possibilities for obtaining a passing score or improving the competition ideas. To determine a rational model for predicting the competitive score of social projects for community development, 6 types of the most common deep neural networks were considered. Subsequently, the best of the obtained models was optimized according to various criteria [8-10]. Our paper presents the results of data collection and intellectual analysis of the results of competitions for local initiative projects in the Lviv region (Ukraine). Based on the results, we conducted a study. It concerns the justification of a rational model of a deep neural network. Its architecture was also optimized, which ensures accurate prediction of the competition score of social projects submitted to competitions.

Thus, the rationale for the deep neural network model for predicting the competitive score of social projects for community development is based on the collection and use of data on previous projects, including project description, budget, community capacity, and competition results. This data was used to train a neural network that subsequently provides accurate forecasting of the competition score of new social projects submitted for the competition. The resulting model is able to demonstrate high prediction accuracy on the test data set. This indicates that it is a valuable tool for assessing the prospects of social projects for community development.

2. Analysis of published data and problem setting

Today, researchers pay a lot of attention to the development of tools for project management [11-13], including for the management of social projects [14-16]. They have made a significant contribution to the development of science in this area. However, with the development of computational intelligence, research is carried out annually on the development of machine learning-based tools for various spheres of human life and activity [17-19]. At the same time, the attention of scientists is focused on the development of tools for project management based on computational intelligence technologies [20-22].

Taking into account the relevance of using both computational intelligence and the expediency of developing tools for project management, our work investigates the use of

deep neural networks to predict the competitive score of social community development projects. Justification of an effective model based on deep neural networks is an urgent task for both science and practice of holding competitions for local initiative projects in individual communities and regions.

Based on the analysis, it was found that there are many scientific papers devoted to this area of research. We have analyzed the most relevant scientific works in this area. Noteworthy are the works [23-25] that deal with predicting project success using neural networks. In these works, their authors used different computational models for forecasting. Each of these papers contains the results of model development and proof that these models improve the results compared to traditional machine learning methods such as logistic regression and decision trees.

In [26-28], it is proved that recurrent neural networks are important models in the field of deep learning. This network structure is used to recursively form complex deep networks with a simple structure. It is possible to add additional weights to the network, which ensures the creation of cycles in the network graph. In addition, it is possible to use information about long-distance dependencies, which ensures high prediction accuracy with sufficient data quality. The learning rate of recurrent neural networks cannot be improved, and the gradient gradually disappears. LSTM improves the nodes of the hidden layer of the RNN into special cellular structures that can perform better in longer sequences. Paper [26] investigates the cyclic neural network algorithm and its improved model based on LSTM, and builds a model for predicting athletes' performance based on a cyclic neural network that can be used to predict athletes' performance with high prediction accuracy.

There are also studies [29-32] that have been conducted to predict the success of crowdfunding projects based on the texts of online social welfare crowdfunding projects. By calculating the amount of information and analyzing the sentimental value of the text, the authors studied how the textual information of an interconnected social security crowdfunding project affects the success of the project. It is found that imperfect R-squared indices reflect that multiple linear regression models perform poorly with respect to this prediction. Furthermore, this paper tests and analyzes the prediction performance of four machine learning models, including a multiple regression model, a decision tree regression model, a random forest regression model, and an AdaBoost regression model.

Recently, several methods have been proposed to explain the predictions of recurrent neural networks (RNNs), including LSTM [33-36]. The goal of these methods is to understand the network's decisions by assigning a relevance to each input variable, such as a word, indicating the extent to which it influenced a particular prediction.

In existing works, some of the architectures of neural network models have not been compared with each other or evaluated only qualitatively. We propose to fill this gap by quantitatively comparing different neural network model architectures for predicting the competitive score of social community development projects. Using the model that will show the best results during the research, perform optimization for it using different options. Thus, the objective of this paper is to substantiate a qualitative deep neural network model for predicting the competitive score of social projects for community development. The model will be trained on a dataset containing information about previous projects, including project description, budget, and competition results.

3. The purpose and objectives of the study

The aim of the paper is to substantiate a model for predicting the competitive score of social community development projects based on an optimized deep neural network architecture trained on prepared data with the results of previous projects, including their description, budget characteristics, community capabilities, and competition results.

To achieve the presented goal, the following tasks should be solved:

1. to propose a research methodology and prepare data that will ensure the training of deep neural networks for predicting the competitive score of social projects for community development;

2. to substantiate and optimize the architecture of the deep neural network model for predicting the competitive score of social projects for community development and to evaluate its accuracy.

4. Research methodology and data preparation for training deep neural networks

The main stages of the research methodology for substantiating and optimizing the architecture of the deep neural network model for predicting the competitive score of social community development projects, as well as evaluating its accuracy indicators, are shown in Fig. 1.

The proposed algorithm involves 10 main stages, which include the collection and formation of a database of implemented social projects in the territory of communities based on data with competitive proposals of social projects in the territory of communities. In particular, we have collected data on the results of competitions for local initiative projects in the Lviv region during 2018-2021. They reflect the indicators of the implementation of social community development projects in the Lviv region, which were funded from the regional budget. We received 741 copies of data on implemented social projects (Lviv region, Ukraine) (Fig. 2), which are distributed by attributes: 1) year of project implementation (Year); 2) project number (No); 3) project registration number (Registration_number); 4) project name (Project_name); 5) name of the territorial community (Territorial_community); 6) name of the settlement (Settlement); 7) total project budget (Project_budget); 8) funds from the regional budget (Regional_budget); 9) funds from the district budget (District_budget); 10) funds from the territorial community (Public_budget); 11) sponsorship funds (Sponsorship_funds); 12) financial contribution (Financial_contrib); 13) non-financial contribution (Nonfinancial_contrib); 14) percentage of non-budgetary contribution (%_non-budgetary_contrib); 15) total score (Total_score); 16) expert score 1 (score_E1); 17) expert score 2 (score_E2); 18) Expert 3 score (score_E3); 19) Taxcapacity_index; 20) Score_taxability_index; 21) Final_score; 22) Results of project selection (Results).

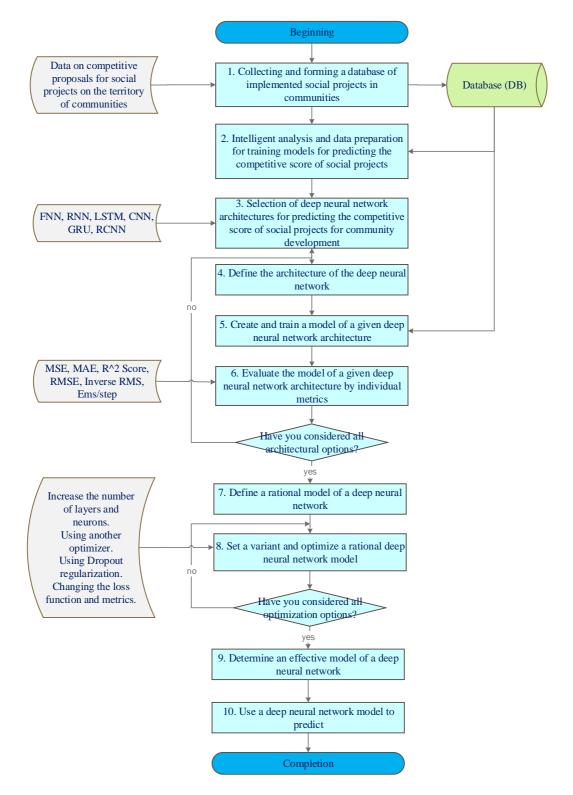


Figure 1: Algorithm for conducting research to substantiate and optimize the architecture of a deep neural network model for predicting the competitive score of social projects for community development.

	Year	No	Registration_number	Project_name	Territorial_community	Settlement	Project_budget	Regional_budget	District_budget	Public_budget	
0	2021	1	1218	Капітальний ремонт приміщень корпусу №1 Підгор	Сколівська	с. Підгородці	299603	149802	0	74601	
1	2021	2	942	Капітальний ремонт системи внутрішнього водопо	Меденицька	смт. Меденичі	299996	147498	0	76498	
2	2021	3	502	Капітальний ремонт приміщення спортивного залу	Судововишнянсь ка	м. Судова Вишня	208785	104392	0	61393	
3	2021	4	1578	Релаксуємо, навчаючись. Придбання обладнання т	Сколівська	м. Сколе	199530	99760	0	49666	
4	2021	5	938	Капітальний ремонт санвузлів Лучицького НВК "З	Сокальська	с. Лучиці	223229	111000	0	36678	

5 rows × 22 columns

Figure 2: A fragment of the database on the results of competitions for local initiative projects in the Lviv region.

The interactive software environment Jupyter Notebook was used to intelligently analyze data on the results of competitions for local initiative projects in the Lviv region. This made it possible to analyze the data and quickly experiment with the code and demonstrate the results. Based on the data analysis, it was found that some of them (Total_score, score_E1, score_E2, score_E3) duplicate the final score (Final_score) and were removed from the DataFrame (df). During the preliminary processing of the data in the df, we analyzed it for gaps and anomalies. In addition, the values of individual columns of the df were converted to floating point numbers (float64) and integers (int). To determine the type of settlement, we used a function based on a simple method based on the prefix:

 $df ['Settlement'] = df ['Settlement'].apply (lambda X : map_settlement(X)), (1)$

where df ['*Settlement*'] – a column in df with the name of the settlement (Settlement); X – current value of the column df ['*Settlement*'].

The proposed map_settlement function (1) is applied to the column df ['Settlement']. It takes a single value as an input parameter and uses it to determine the type of settlement. It returns an integer value (Fig. 3), which corresponds to the following types of settlements: 1) 0 - village (begins with "c."); 2) 1 - urban village (begins with "cMT."); 3) 2 - city (begins with "M."); 4) -1 - unknown type of settlement (does not correspond to any of the previous cases).

For preliminary data processing, we used the scikit-learn library of the Python programming language to analyze the results of the local initiative project competitions. As a result, we have prepared data that will provide training for deep neural networks to predict the competitive score of community development projects (Table 1).

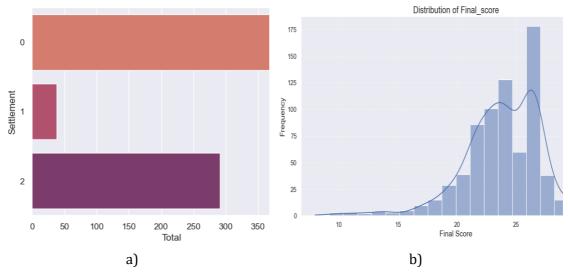


Figure 3: Histogram of the number of settlements of different types in which social community development projects have been implemented in Lviv Oblast (a) and the distribution of the final scores (b)

Table 1

A fragment of the prepared data for predicting the competitive score of social projects for community development based on the use of deep neural networks

Project	Settlement	Project_budget	Regional_budget	District budget	Public_budget	Sponsorship_funds	Financial_contrib	Nonfinancial_contrib	%_non- budgetary_contrib	Taxcapacityjndex	Score_taxabilityjndex	Final_score
_	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	Y1
0	0	299603	149802	0	74601	30367	15000	29833	25.1	0.7097	0.5	31.833
1	1	299996	147498	0	76498	76000	0	0	25.3	0.4156	1.5	31.500
2	2	208785	104392	0	61393	25000	18000	0	20.6	0.4144	1.5	31.167
270	2	199530	99760	0	49666	30034	20070	0	25.1	0.7097	0.5	31.167
271	0	223229	111000	0	36678	0	42000	33551	33.8	0.6075	1.0	31.000

To improve the efficiency of training deep neural network models, we scaled the data. To do this, we used normalization and minimum scaling of the data of individual df columns.

This ensured that the values X_i of the df columns were transformed so that all data in the range from 0 to 1 were obtained:

$$X'_{i} = \left(X_{i} - \min(X_{i})\right) / \left(\max(X_{i}) - \min(X_{i})\right), \qquad (2)$$

where X_i – initial values of the column df; $min(X_i)$ – is the minimum value of the indicator in the df column, $max(X_i)$ – the maximum value of the indicator in the column, X'_i – scaled value of the indicator.

We have determined the input parameters of the model for predicting the competitive score of social community development projects based on the use of deep neural networks. At the same time, we selected attributes that are closely related to the target feature (final score) due to the constructed correlation matrix. For each input parameter of the model, which are shown in Table 1, we determined their average value \bar{X}_i :

$$\bar{X}_{i} = \frac{1}{N} \sum_{i=1}^{N} X_{ij}, j = 1, m,$$
(3)

where N – number of social projects implemented, units.

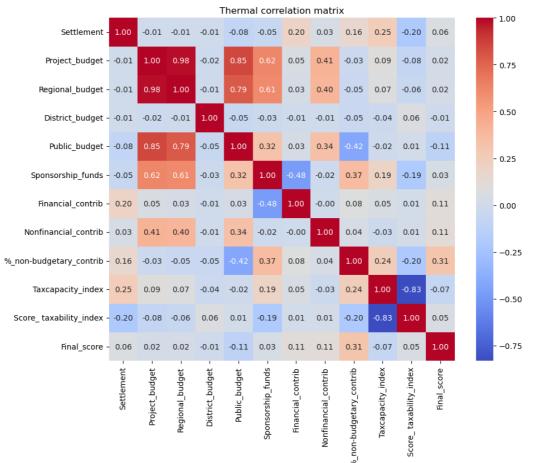


Figure 4: Correlation matrix between the input data and the target feature "Final_score".

The correlation matrix K_{ij} is constructed, the elements of which are determined using the formula:

$$K_{ij} = \frac{\operatorname{cov}(X_{ij}, Y_1)}{\sigma(X_{ij}), \sigma(Y_1)},$$
(4)

where $cov(X_{ii}, Y_1)$ – covariance between the input data X_{ii} and the target feature Y_1 .

The quantitative value $cov(X_{ii}, Y_1)$ is determined by the formula:

$$\operatorname{cov}(X_{ij}, Y_1) = \frac{1}{N-1} \sum_{l=1}^{N} (X_{li} - \bar{X}_i) (X_{lj} - \bar{X}_j), i, j = 1, m.$$
(5)

As a result of the calculations, a correlation matrix between the input data and the target feature "Final_score" was built (Figure 4).

The analysis revealed strong and weak correlations between the input data and the target feature "Final_score". Based on the analysis of this correlation matrix, we have an idea of the existing links between funding sources, characteristics of social projects and their final scores. The information obtained may be valuable for further analysis or model development.

5. The results of substantiation and optimization of the architecture of the deep neural network model for predicting the competitive score of social projects for community development

We begin the model justification with the choice of deep neural network architectures that will ensure the prediction of the competitive score of social community development projects. When choosing deep neural network architectures for predicting the competitive score of social community development projects, we took into account the features of the data and the task of prediction. We propose to use the following deep neural network architectures (Figure 5): 1) Feedforward Neural Network (FNN); 2) Recurrent Neural Network (RNN); 3) Long Short-Term Memory (LSTM); 4) Gated Recurrent Unit (GRU); 5) Convolutional Neural Network (CNN); 6) Recursive Neural Network (RCNN).

A linear neural network FNN is used to make predictions based on input features without depending on previous time steps or sequences. If the data has a complex nonlinear structure, FNN may be ineffective for modeling it. Recurrent neural network RNN provides processing of sequential data. However, RNNs can face the problem of vanishing gradient when training on long sequences, which leads to limitations in the model's ability to account for long-term dependencies. LSTM and GRU neural networks are advanced versions of RNNs that solve the vanishing gradient problem by providing special gates that allow the model to store and update information for a long time. They are especially effective for modeling long-term dependencies in sequential data. CNNs are commonly used for image processing, but they can also be useful for sequence analysis, as they can be transformed into a form that preserves spatial structure. Such a deep neural network is useful for analyzing relationships between attributes at different levels of abstraction. The convolutional neural network RCNN combines recursive and convolutional layers, which allows the model to

analyze sequences and structural relationships in the data. This is useful for modeling sequences that reflect the relationships between social project evaluations and their characteristics.

We used the Sequential model from the Keras library, which is a sequential neural network where layers are arranged one after the other. This is the simplest type of model that is well suited for many types of machine learning tasks. It is proposed to use deep neural network architectures with 3 to 7 layers with the ReLU activation function. In the last output layer, only one neuron is used, since the task of predicting the competitive score of social community development projects is a regression task. When compiling the models, the Adam optimizer is used and the root mean square error is used as a function of loss. The models are trained on training data for 100 epochs.

To evaluate the models of a given deep neural network architecture, we used the following metrics. An indicator for assessing the accuracy of regression models (Mean Squared Error), which is defined as the root mean square difference between the predicted and actual values of the competitive score of social community development projects:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \left(Y_i - \hat{Y}_i \right)^2.$$
 (6)

where Y_i – real value of the competition score of social projects for community development, points; \hat{Y}_i – predicted value of the competitive score of social projects for community development, points; N – number of data on social projects in the dataset, units.

An indicator for assessing the accuracy of regression models *MAE* (Mean Absolute Error), which is defined as the average value of the difference between the predicted and actual values of the competitive score of social projects for community development:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |Y_i - \hat{Y}|.$$
 (7)

The coefficient of determination (R^2Score) is used to assess the quality of a regression model. It shows how well the model reflects the variation in the dependent variable. This indicator is determined by the formula:

$$R^2 Score = 1 - \frac{SS_{res}}{SS_{tot}}.$$
(8)

where SS_{res} – sum of squared errors; SS_{tot} – total sum of squares.

The closer the value is R^2Score to 1, the better the model explains the variation in responses.

The next indicator shows the square root of the root mean square error (*RMSE*). It is used to measure the average size of regression model errors and is calculated by the formula

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2}.$$
 (9)

where Y_i – real value of the competition score of social projects for community development, points; \hat{Y}_i – predicted value of the competitive score of social projects for community development, points; N – number of data on social projects in the dataset, units.

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 64)	768
dense_1 (Dense)	(None, 32)	2080
dense_2 (Dense)	(None, 1)	33

Total params: 2881 (11.25 KB) Trainable params: 2881 (11.25 KB) Non-trainable params: 0 (0.00 Byte)

a)

Model:	"sequential_2	"
--------	---------------	---

Model: "sequential 4"

Istm (LSTM) (None, 11, 64) 16896 Istm_1 (LSTM) (None, 32) 12416 dense_4 (Dense) (None, 1) 33	Layer (type)	Output Shape	Param #
_ , , , , , , , ,	lstm (LSTM)	(None, 11, 64)	16896
dense_4 (Dense) (None, 1) 33	lstm_1 (LSTM)	(None, 32)	12416
	dense_4 (Dense)	(None, 1)	33

_____ Total params: 29345 (114.63 KB)

Trainable params: 29345 (114.63 KB) Non-trainable params: 0 (0.00 Byte)

Model: "sequential_1"

Layer (type)	Output Shape	Param #
simple_rnn (SimpleRNN)	(None, 11, 64)	4224
<pre>simple_rnn_1 (SimpleRNN)</pre>	(None, 32)	3104
dense 3 (Dense)	(None, 1)	33

Total params: 7361 (28.75 KB) Trainable params: 7361 (28.75 KB)

Non-trainable params: 0 (0.00 Byte)

b)

Model: "sequential_3"

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 9, 64)	256
max_pooling1d (MaxPooling1 D)	(None, 4, 64)	0
conv1d_1 (Conv1D)	(None, 2, 32)	6176
max_pooling1d_1 (MaxPoolin g1D)	(None, 1, 32)	0
flatten (Flatten)	(None, 32)	0
dense_5 (Dense)	(None, 1)	33

-----Total params: 6465 (25.25 KB) Trainable params: 6465 (25.25 KB)

Non-trainable params: 0 (0.00 Byte)

Model: "sequential_11"

d)

. =		
Layer (type)	Output Shape	Param #
gru (GRU)	(None, 11, 64)	12864
gru_1 (GRU)	(None, 32)	9408
dense_6 (Dense)	(None, 1)	33

e)

c)

------Total params: 22305 (87.13 KB) Trainable params: 22305 (87.13 KB) Non-trainable params: 0 (0.00 Byte)

Layer (type)	Output Shape	Param ‡
conv1d_2 (Conv1D)	(None, 9, 32)	128
max_pooling1d_2 (MaxPoolin g1D)	(None, 4, 32)	0
conv1d_3 (Conv1D)	(None, 2, 64)	6208
max_pooling1d_3 (MaxPoolin g1D)	(None, 1, 64)	0
flatten_1 (Flatten)	(None, 64)	0
dense_33 (Dense)	(None, 64)	4160
dense_34 (Dense)	(None, 1)	65
Total params: 10561 (41.25 k Trainable params: 10561 (41.	·	

Non-trainable params: 0 (0.00 Byte)

f)

Figure 5: Variants of the studied deep neural network architectures: a) Feedforward Neural Network (FNN); b) Recurrent Neural Network (RNN); c) Long Short-Term Memory (LSTM); d) Gated Recurrent Unit (GRU); e) Convolutional Neural Network (CNN); f) Recursive Neural Network (RCNN)

The indicator *RMSE* is measured in units of the original variable, i.e., in our case, in the competitive scores of social community development projects. The smaller the value RMSE , the better the model's predictions.

The inverse square root of the mean (*Inverse RMS*) is worthy of note: This is simply the inverse of *RMSE*.

Inverse
$$RMS = \frac{1}{RMSE}$$
. (10)

If *RMSE* it measures the average size of the prediction errors of the competition scores of social projects for community development, the inverse indicator *Inverse RMS*, measures how many times the average prediction value deviates from the actual value. In other words, the higher the value of *Inverse RMS*, the lower the value of the model error.

Based on the above-described variants of the architecture of the studied deep neural network architectures (Figure 5), models for predicting the competitive score of social community development projects were created. For each of them, metrics were determined using formulas (6-10), which made it possible to obtain their quantitative values, which are presented in Table 2.

<e< th=""><th colspan="9">estits of determining the accuracy of the created models on test data</th></e<>	estits of determining the accuracy of the created models on test data								
	Variant of the model	MSE	MAE	R2 Score	RMSE	Inverse RMSE	ms/step		
	FNN	7.736	2.143	0.366	2.781	0.360	3		
	RNN	7.600	2.237	0.377	2.757	0.363	2		
	LSTM	12.237	2.554	-0.003	3.498	0.286	5		
	CNN	7.671	2.163	0.371	2.770	0.361	2		
	GRU	12.234	2.553	-0.002	3.498	0.286	6		
	RCNN	6.962	2.110	0.430	2.638	0.379	2		

Table 2

Results of determining the accuracy of the created models on test data

Table 2 shows the results of numerical metrics for different model architectures that were evaluated for their effectiveness in predicting the competitive scores of community development projects. The results show that each of the models has its advantages and disadvantages in predicting the competitive scores of community development projects.

To display the training process of each model by the number of epochs performed, we have plotted the graphs of changes in the mean square value of the absolute error (MAE), which are shown in Figure 6.

As you can see from the graphs above, when the MAE value decreases on both the training and validation datasets, it means that the model is learning and performing better with each new epoch. However, if the MAE value on the training set continues to fall, while the MAE value on the validation set starts to rise or remains unchanged, this is a sign that the model is overtraining. The resulting graphs help to visualize the efficiency of model training and assess its overall performance.

First of all, we can refer to the MSE, MAE, and RMSE values to evaluate the accuracy of the predictions. In particular, the FNN (Feedforward Neural Network) and RNN (Recurrent Neural Network) models have similar MSE, MAE, and RMSE values, but the RNN model has a slightly better R²Score, which may indicate a better ability to explain variability in the original data. The models based on LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit) showed the worst performance among all models.

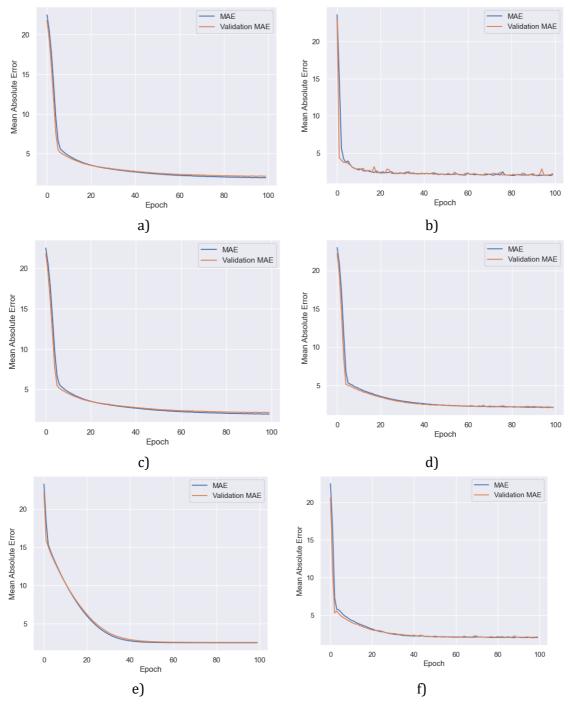


Figure 6: Graphs of changes in the mean-square value of the absolute error (MAE) during the training of models of the studied deep neural network architectures: a) Feedforward Neural Network (FNN); b) Recurrent Neural Network (RNN); c) Long Short-Term Memory (LSTM); d) Gated Recurrent Unit (GRU); e) Convolutional Neural Network (CNN); f) Recursive Neural Network (RCNN)

They have the highest MSE, MAE, and RMSE values, as well as a very low R²Score, which may indicate a poor ability to explain variability in the data. The model based on CNN (Convolutional Neural Network) showed acceptable results, but slightly worse than FNN and RNN. The model based on RCNN (Recurrent Convolutional Neural Network) showed the best results among all models. It has the lowest values of MSE=6.962, MAE=2.11, and RMSE=2.638, and the highest R²Score=0.43, which indicates a better ability to explain variability in the original data (Figure 7).

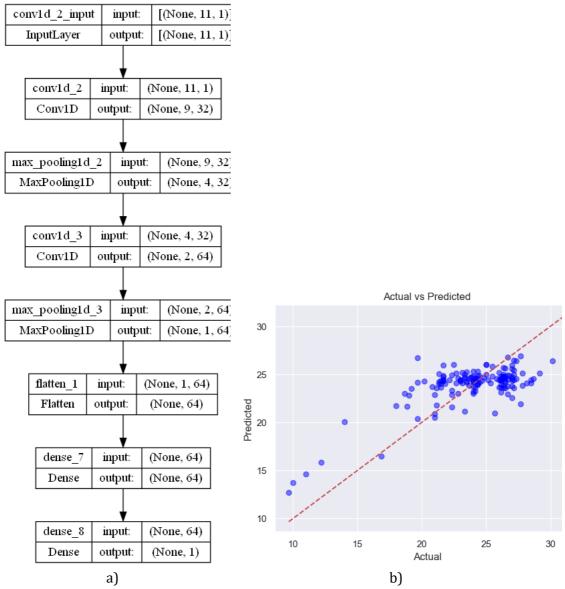


Figure 7: Rational architecture of the Recursive Neural Network (RCNN) deep neural network model (a) and scatter plot with identity line and color coding (b)

Thus, the best model for predicting the competitive scores of social community development projects, taking into account the numerical values of the presented metrics, is

the RCNN-based model. This model demonstrates the lowest values of prediction errors and the highest level of explanation of variability in the original data.

Analyzing the rational architecture of the RCNN deep neural network model, we see that it has a sequential stack of layers. The first layer Conv1D has 32 filters and a kernel size of 3. This layer is responsible for extracting features from the input data. After that, the MaxPooling1D layer is applied, which reduces the dimensionality of the original data by half. Then there is a second layer Conv1D with 64 filters and a kernel size of 3. It is also responsible for feature extraction. After that, the MaxPooling1D layer is used again, which reduces the dimensionality of the data by half. The data is flattened with Flatten for further input into the fully connected layer. The fully connected layer has 64 neurons and uses the relu activation function. The last layer is the output layer with one neuron without activation, as we are predicting a numerical value. The total number of parameters in the model is 10,561. This includes weights and offsets for each layer. All model parameters are training parameters. The resulting RCNN deep neural network model has a fairly simple structure with only convolutional and fully connected layers. This architecture is often used to process sequential data and it makes it possible to accurately predict the quantitative value of the competitive score of social community development projects.

Subsequently, we optimized the obtained rational deep neural network model to predict the quantitative value of the competitive score of social community development projects. It is proposed to optimize the model by 4 options: 1) increasing the number of layers and neurons; 2) using another optimizer; 3) using Dropout regularization; 4) changing the loss function. For each of the variants of model training, metrics are determined using formulas (6-10), which make it possible to obtain their quantitative values, which are presented in Table 3.

0		- J I -				
Model RCNN	MSE	MAE	R2 Score	RMSE	Inverse RMSE	ms/step
Basic	6.962	2.110	0.430	2.638	0.379	2
Increasing the number of layers and neurons Use of the RMSprop	7.380	2.133	0.395	2.717	0.368	1
optimizer	7.777	2.248	0.363	2.789	0.359	1
Use of Dropout regularization Changes in the loss	6.877	2.057	0.437	2.622	0.381	2
function	8.171	2.226	0.330	2.859	0.350	1

Table 3

Results of determining the accuracy of optimized RCNN models on test data

The basic RCNN model has good performance in terms of accuracy in predicting the quantitative value of the competitive score of community development projects. However, there is room for improvement. The proposed model using Dropout regularization shows the best results in terms of MSE, MAE, and R2 Score. It achieves the lowest mean square error and the highest coefficient of determination (R2 Score), which indicates the model's

better ability to explain variability in the original data. The model with the modified loss function has the worst performance among all models. This may indicate that the chosen loss function is not suitable for the given forecasting task. Regarding the training time of the models, the models with RMSprop optimizers and the modified loss function require less time for each step, but this is not a decisive factor, since the time difference is very small. Thus, taking into account the obtained quantitative values of the accuracy indicators of the optimized RCNN models, the model using Dropout regularization is the best for predicting the quantitative value of the competitive score of social community development projects.

For the deep neural network model using Dropout regularization, a graph of the dependence of actual values on the predicted values of the competitive score of social projects for community development was constructed (Fig. 8, a), as well as the distribution of model residuals (Fig. 8, b).

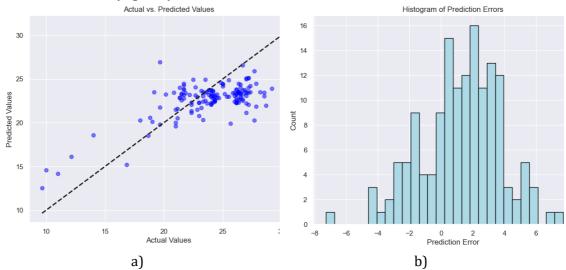


Figure 8: Graph of the dependence of actual values on the predicted values of the competitive score of community development projects (a) and the distribution of residuals of the deep neural network model using Dropout regularization (b)

The resulting graph of the dependence of the actual values on the predicted values of the competition score of social community development projects (Fig. 8, a) shows the actual values of the competition score on the abscissa axis. These values represent the actual values of the competition score during the implementation of social community development projects in Lviv Oblast. The ordinate axis shows the predicted values of the competition scores of community development projects, which were obtained using a deep neural network model with Dropout regularization. Each point on the graph represents a pair of actual and corresponding predicted values of the competition scores of community development projects. The more the point is shifted up and to the left, the greater the error in the prediction of the deep neural network model using Dropout regularization. The diagonal line (black dashed line) represents the ideal case when the actual and predicted values coincide. The points close to this line indicate accurate predictions of the competition scores of community development projects.

The constructed distribution of residuals of the deep neural network model using Dropout regularization displays on the abscissa axis the residuals, which is the difference between the actual and predicted values of the competition score. The ordinate axis shows the number of observations falling into the corresponding residual intervals. This distribution makes it possible to determine how evenly the model residuals are distributed. The ideal case is when the residuals of the competition scores of community development projects are distributed evenly around zero without significant deviations.

Both graphs presented in Fig. 8 are important for assessing the accuracy and efficiency of the model in predicting the competitive scores of social projects. The first graph allows us to visualize how well the model predicts the values, and the second graph helps us understand the distribution of model errors. It has been found that the vast majority of error deviations are between -2 and +4 points, which meets the requirements for the accuracy of forecasting the values of the competitive scores of social development projects.

6. Conclusions

1. The proposed methodology for predicting the competitive score of social projects for community development is based on the use of deep neural networks and includes ten stages. These stages are reflected in the developed research algorithm to substantiate and optimize the architecture of the deep neural network model for predicting the competitive score of social community development projects (Figure 1). The peculiarity of the proposed methodology is that the database is formed on the basis of the collected data on the results of competitions for local initiative projects in the region. Intelligent analysis of the collected data makes it possible to assess the relationship between the input characteristics and the competitive score of social community development projects. The described methodology is the basis for performing high-quality data preparation, which is the basis for justifying an effective deep neural network model for predicting the competitive score of community development projects.

2. Based on the developed methodology, as well as on the collected data on the results of competitions for local initiative projects in the Lviv region during 2018-2021, the data were prepared for training deep neural network models. The research used 6 types of modern deep neural network architectures (FNN, RNN, LSTM, GRU, CNN, RCNN). To evaluate the models of a given deep neural network architecture, such metrics as MSE, MAE, R²Score, RMSE, Inverse RMSE, and ms/step were used. It was found that the best accuracy rates are provided by the model based on RCNN (Recurrent Convolutional Neural Network). It showed the best results among all models. It has the lowest values of MSE=6.962, MAE=2.11 and RMSE=2.638, as well as the highest R²Score=0.43, which indicates a better ability to explain variability in the original data. We have optimized the basic RCNN model using 4 options: 1) increasing the number of layers and neurons; 2) using another optimizer; 3) using Dropout regularization; 4) changing the loss function. It is established that the RCNN model using Dropout regularization is the best for predicting the quantitative value of the competitive score of social community development projects. The use of Dropout regularization in the model brought a slight improvement in the R2 Score by 1.63%, which indicates a better ability of the model to explain the variance in the original data. At the same time, relative to the baseline RCNN model, there was a decrease in MSE by 1.22% and a decrease in MAE by 2.51%. Further research should be carried out in the direction of developing a decision support system for planning social community development projects based on the proposed RCNN model using Dropout regularization to predict the competitive score of social community development projects.

7. References

- [1] K. Cairampoma-Caro, J. Vargas-Florez, C. Romero-Izaga, Towards a Lean Construction toolbox to improve social projects management, Brazilian Journal of Operations and Production Management, 2022, 19(2), e20221396
- [2] A. Tryhuba, I. Kondysiuk, I. Tryhuba, O. Boiarchuk, A. Tatomyr, Intellectual information system for formation of portfolio projects of motor transport enterprises, CEUR Workshop Proceedings, 2022, 3109, pp. 44–52.
- [3] M. Odrekhivskyy, V. Pasichnyk, N. Kunanets, A. Rzheuskyi, G. Korz D. Tabachyshyn, The Use of Modern Information Technology in Medical and Health Institutions of Truskavets Resort, Proceedings of the 2nd International Workshop on Modern Machine Learning Technologies and Data Science (MoMLeT+DS 2020). Volume I: Main Conference. Lviv-Shatsk, Ukraine, June 2-3, 2020. pp.184-197.
- [4] T. Basyuk, A. Vasyliuk, V. Lytvyn, O. Vlasenko, Features of designing and implementing an information system for studying and determining the level of foreign language proficiency, Proceedings of the 4nd International Workshop on Modern Machine Learning Technologies and Data Science Workshop MoMLeT&DS 2022. Leiden, The Netherlands, November 25-26, 2022. pp.212-225.
- [5] Y.Vyklyuk, M. Radovanovic, V. Pasichnyk, N. Kunanets, P.Sydor, Forecasting of forest fires in portugal using parallel calculations and machine learning, Proceedings of the XVIII International Conference on Data Science and Intelligent Analysis of Information, June 4–7, 2018, Kyiv, Ukraine. pp.39-49.
- [6] A. Tryhuba, V. Boyarchuk, I. Tryhuba, O. Ftoma, R. Padyuka, M. Rudynets, Forecasting the Risk of the Resource Demand for Dairy Farms Basing on Machine Learning, Proceedings of the 2nd International Workshop on Modern Machine Learning Technologies and Data Science (MoMLeT+DS 2020). Volume I: Main Conference. Lviv-Shatsk, Ukraine, June 2-3, 2020. pp.327-340.
- [7] Project archive: Public budget of Lviv. URL: https://lviv.pb.org.ua/projects/archive/2?session
- [8] R. Peleshchak, V. Lytvyn, I. Peleshchak, O. Veres, Neural Network Architecture with Oscillatory Synaptic Blocks for Signal Processing of Spatially Spaced Multiband Radar Complex Lecture Notes in Electrical Engineering, 2023, 965 LNEE, pp. 367–374.
- [9] O. Malanchuk, A. Tryhuba, I. Tryhuba, R. Sholudko, O. Pankiv, A Neural Network Modelbased Decision Support System for Time Management in Pediatric Diabetes Care Projects. International Scientific and Technical Conference on Computer Sciences and Information Technologies, 2023.
- [10] R. Yurynets, Z. Yurynets, M. Grzebyk, M. Kokhan, N. Kunanets, M. Shevchenko, Neural Network Modeling of the Social and Economic, Investment and Innovation Policy of the

State, Proceedings of the 4nd International Workshop on Modern Machine Learning Technologies and Data Science Workshop MoMLeT&DS 2022. Leiden, The Netherlands, November 25-26, 2022. pp.252-262.

- [11] A. Tryhuba, I. Tryhuba, O. Ftoma, O. Boyarchuk, Method of quantitative evaluation of the risk of benefits for investors of fodder-producing cooperatives. International Scientific and Technical Conference on Computer Sciences and Information Technologies, 2019, 3, pp. 55–58, 8929788.
- [12] R. Ratushny, I. Horodetskyy, Y. Molchak, V. Grabovets, The configurations coordination of the projects products of development of the community fire extinguishing systems with the project environment. CEUR Workshop Proceedings, 2021, 2851, pp. 238–248.
- [13] N. Koval, I. Kondysiuk, V. Grabovets, V. Onyshchuk, Forecasting the fund of time for performance of works in hybrid projects using machine training technologies. CEUR Workshop Proceedings, 2021, 2917, pp. 196–206.
- [14] R. Ratushny, O. Bashynsky, V. Ptashnyk, Development and Usage of a Computer Model of Evaluating the Scenarios of Projects for the Creation of Fire Fighting Systems of Rural Communities. 11th International Scientific and Practical Conference on Electronics and Information Technologies, ELIT 2019 - Proceedings, 2019, pp. 34–39, 8892320.
- [15] A. Tryhuba, R. Ratushny, I. Tryhuba, N. Koval, I. Androshchuk, The model of projects creation of the fire extinguishing systems in community territories. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 2020, 68(2), pp. 419–431.
- [16] S. Tyshko, O. Lavrut, V. Vysotska, O. Markiv, O. Zabula, Y. Chernichenko, T. Lavrut, Compensatory Method for Measuring Phase Shift Using Signals Bisemiperiodic Conversion in Diagnostic Intelligence Systems, Proceedings of the 4nd International Workshop on Modern Machine Learning Technologies and Data Science Workshop MoMLeT&DS 2022. Leiden, The Netherlands, November 25-26, 2022. pp.144-154.
- [17] A. Tryhuba, V. Boyarchuk, I. Tryhuba, S. Francik, M. Rudynets, Method and software of planning of the substantial risks in the projects of production of raw material for biofuel, CEUR Workshop Proceedings, 2020, 2565, pp. 116–129.
- [18] N. Kunanets, O. Vasiuta, N. Boiko, Advanced technologies of big data research in distributed information systems. 14th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), vol. 3, 71–76, September 2019. DOI: 10.1109/STC-CSIT.2019.8929756
- [19] R. Peleshchak, V. Lytvyn, I. Peleshchak, V. Vysotska. Stochastic Pseudo-Spin Neural Network with Tridiagonal Synaptic Connections. SIST 2021 - 2021 IEEE International Conference on Smart Information Systems and Technologies, 2021, 9465998
- [20] A. Tryhuba, N. Koval, I. Tryhuba, O. Boiarchuk, Application of Sarima Models in Information Systems Forecasting Seasonal Volumes of Food Raw Materials of Procurement on the Territory of Communities. CEUR Workshop Proceedings, 2022, 3295, pp. 64–75.
- [21] S. Gupta, H. K. Verma, and D. Bhardwaj, Classification of diabetes using Naïve Bayes and support vector machine as a technique, Operations Management and Systems Engineering, Springer, Singapore, 2021.
- [22] T. Basyuk, A. Vasyliuk, V. Lytvyn, Mathematical Model of Semantic Search and Search Optimization, Proceedings of the 3rd International Conference on Computational

Linguistics and Intelligent Systems (COLINS-2019). Volume I: Main Conference, Kharkiv, Ukraine, April 18-19, 2019, Vol-2362: pp.96-105.

- [23] H. Youneszadeh, A. Ardeshir, M.H. Sebt, Predicting project success in residential building projects (Rbps) using artificial neural networks (ANNS), Civil Engineering Journal (Iran), 2020, 6(11), pp. 2203–2219
- [24] S. Voloshyn, R. Peleshchak, I. Peleshchak, V. Vysotska, Big Data Analysis for Multispectral Images Recognition Based on Deep Learning, International Scientific and Technical Conference on Computer Sciences and Information Technologies, 2021, 1, pp. 160–170.
- [25] R. Padyuka, V. Tymochko, P. Lub, Mathematical model for forecasting product losses in crop production projects. CEUR Workshop Proceedings, 2022, 3109, pp. 25–31.
- [26] Zhang, Y. Research on Athlete Performance Prediction Model Based on Recurrent Neural Network. IEEE Conference on Telecommunications, Optics and Computer Science, TOCS 2022, 2022, pp. 1292–1296.
- [27] K. Saini, N. Kumar, R. Kumar, B. Bhushan, Recurrent Neural Network for the Identification of Nonlinear Dynamical Systems: A Comparative Study, Lecture Notes in Electrical Engineering, 2024, 1099, pp. 345–353.
- [28] S. Anfilets, S. Bezobrazov, V. Golovko, A. Sachenko, M. Komar, R. Dolny, V. Kasyanik, P. Bykovyy, E. Mikhno, O. Osolinskyi, Deep multilayer neural network for predicting the winner of football matches. International Journal of Computing, 2020, 19(1), 70-77.
- [29] Bâra, A., Oprea, S.-V. Machine Learning Algorithms for Power System Sign Classification and a Multivariate Stacked LSTM Model for Predicting the Electricity Imbalance Volume. International Journal of Computational Intelligence Systems, 2024, 17(1), 80.
- [30] V. Turchenko, E.Chalmers, A. Luczak, A deep convolutional auto-encoder with pooling unpooling layers in Caffe. International Journal of Computing, 2019, 18(1), 8-31.
- [31] J. Petchhan, V. Phanichraksaphong, S. Doungtap, N. Tridet, Toward Project Success Forecasting in Reward-based Crowdfunding through Wide-and-Deep Computational Learning, 15th IEEE International Conference on Industry Applications, INDUSCON 2023 - Proceedings, 2023, pp. 1489–1493.
- [32] O. Bashynsky, I. Garasymchuk, D. Vilchinska, V. Dubik, Research of the variable natural potential of the wind and energy energy in the northern strip of the ukrainian carpathians, E3S Web of Conferences, 2020, 154, 06002.
- [33] C. Perez, K. Sokolova, Is There a Community of French Fitness Influencers on YouTube? A Social Network Analysis Approach to YouTube Engagement, ACM International Conference Proceeding Series, 2023, pp. 13–19.
- [34] J. Meng, C. Li, J. Tao, Y. Dong, J. Du, RNN-LSTM-Based Model Predictive Control for a Corn-to-Sugar Process, Processes, 2023, 11(4), 1080.
- [35] R. Li, X. Gong, B. Zhang, M. Li, H. Guo, Study in intelligent exam based on RNN and LSTM, Artificial Intelligence in Education and Teaching Assessment, 2022, pp. 63–67.
- [36] V. V. Morozov, O. V. Kalnichenko, O. M. Mezentseva, The method of interaction modeling on basis of deep learning the neural networks in complex it-projects. International Journal of Computing, 2020, 19(1), 88-96.