

Combining Forecasts Based on Time Series Models in Machine Learning Tasks

Irina Kalinina¹, Peter Bidyuk², Aleksandr Gozhyj¹ and Pavlo Malchenko¹

^{1.} *Petro Mohyla Black Sea National University, St. 68 Desantnykiv 10, Mykolaiv, Ukraine, 54000*

^{2.} *National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», 37, Prospect Beresteyskyi (former Peremohy), Kyiv, Ukraine, 03056*

Abstract

The article investigates the solution of the forecasting problem using the combination of basic forecasting models for machine learning tasks. Methods of combining forecasts have been studied. Simple mean, weighted averaging, and regression combining methods were considered. The conditions and features of using each method to improve forecast accuracy are defined. A methodology for building combined forecasts based on methods of combining forecast estimates has been developed. The methodology consists of the following stages: analysis and preliminary processing of the data set; division of prepared data into training and test samples; modeling and forecasting based on basic models; formation of weight coefficients of combined forecasts based on evaluations of the effectiveness of basic models; unit for combining and evaluating forecasts. The architecture of the forecasting information system based on time series models has been developed. The efficiency of building combined forecasts for solving machine learning tasks has been studied. Methods of combining forecasts were studied on data sets that characterize changes in the dynamics of share prices of three companies.

Keywords 1

Combined forecast, Simple averaging, Weighted averaging, Regression, Basic model, Time series, Forecast performance evaluation.

1. Introduction

Recently, machine learning technologies have taken a leading position in the market of intelligent solutions. One of the main tasks solved by machine learning technologies is forecasting. Improving the quality of predictive solutions is achieved by various methods and approaches. One approach is to use a combination of forecasts. Combinations of forecasts have become widespread in recent years and have become part of the main direction of research on improving the quality of forecast solutions. Combining and combining several predictions obtained on the basis of a single data set is now widely used to improve accuracy by integrating information obtained from different sources. This reduces the risk of determining one "best" forecast. Combination schemes have evolved from the historically first, simple, evaluation-free combination methods to complex methods involving time-varying weights, nonlinear combinations, correlations between components, and cross-training. They include a combination of point forecasts and a combination of probabilistic forecasts.

It is known that combining several forecasts obtained using different forecasting methods is often a better approach than identifying a single "best forecast". For time series, forecasts will be generated by a process determined by a specific functional form, due to the possibility of changing trends over

MoMLeT+DS 2023: 5th International Workshop on Modern Machine Learning Technologies and Data Science, June 3, 2023, Lviv, Ukraine

EMAIL: irina.kalinina1612@gmail.com (I. Kalinina); pbidyuke_00@ukr.net (P. Bidyuk); alex.gozhyj@gmail.com (A. Gozhyj); twink1337zhaba@gmail.com (P. Malchenko)

ORCID: 0000-0001-8359-2045 (I. Kalinina); 0000-0002-7421-3565 (P. Bidyuk); 0000-0002-3517-580X (A. Gozhyj); 0009-0002-5259-3752 (P. Malchenko)



© 2023 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)

time, seasonal components, structural shifts and the complexity of real data generation processes [1,2]. The choice of one best predictive model to approximate an unknown, in most cases, nonlinear, non-stationary process of data generation may be associated with three types of uncertainty: data uncertainty, parameter uncertainty, and model uncertainty [3,4]. Given these challenges, it is often better to combine multiple predictions to account for multiple components of the actual data generation process and to reduce uncertainty about model form and parameter specification. Combinations of forecasts are currently effectively used in various fields, such as Internet trade [5], economics [6], epidemiology [7], medicine [8] etc. There are different types of forecast combinations: linear and non-linear, with constant or time-varying parameters, and those that ignore or take into account correlations between individual forecasts. Despite a diverse set of schemes for combining forecasts, an unambiguously better way of combining has not been found [9-12]. And a simple averaging method often dominates complex weighting schemes that should be better. Therefore, three basic methods of combining forecast estimates are often used: based on simple average, weighted averaging, and regression.

Problem statement. The article is aimed at solving the following tasks: research on methods of combining forecasts; development of a methodology for building combined forecasts based on methods of combining forecast estimates; development of the architecture of the forecasting information system based on time series models and research on the effectiveness of building combined forecasts for solving machine learning tasks.

2. Materials and methods

This section reviews the methodology for constructing combined forecasts based on time series models. Methods of combining estimates of forecasts are studied. An example of the use of the developed methodology is offered and the effectiveness of combining forecasts is analyzed by comparing performance estimates.

2.1. Methods of combining estimates of forecasts

Methods of combining forecast estimates for solving machine learning tasks are built on the basis of simple averaging of forecasts, weighted combination of forecasts and regression, presented in Table 1 [13].

Table 1

Methods of combining estimates of forecasts

Methods of combining estimates of forecasts	Mathematical representation
Simple averaging	$\hat{y}_i^c = \sum_{n=1}^N \hat{y}_{ni} / N$
Weighted averaging	$\hat{y}_i^c = \sum_{n=1}^N w_n \hat{y}_{ni}, \sum_{n=1}^N w_n = 1$
Regression	$\hat{y}_i^c = \alpha + \sum_{n=1}^N \beta_{ni} \hat{y}_{ni}$

Averaging forecasts. For N forecasting methods, the combined forecast by the simple averaging method is determined by the following expression:

$$\hat{y}_i^c = \frac{\hat{y}_{1i} + \hat{y}_{2i} + \dots + \hat{y}_{Ni}}{N},$$

where \hat{y}_i^c is a combined forecast; $\hat{y}_{1i}, \hat{y}_{2i}, \dots, \hat{y}_{Ni}$ - forecasts obtained by various methods of machine learning.

The simple averaging method has the following application advantages:

- the weights of forecasts obtained by different methods are equal and cannot be evaluated;
- simple averaging significantly reduces variance and error by averaging the error of individual forecasts;
- the use of simple averaging when it is necessary to take into account the uncertainty of the weight estimate.

The average performance of simple averaging depends on model volatility and the variance ratio of forecast errors associated with different forecasting models [14,15].

If for two forecasting methods ($N=2$) there are forecast values \hat{y}_{1i} , \hat{y}_{2i} for the actual value y_i , then the combined forecast by the averaging method is determined:

$$\hat{y}_i^c = \frac{\hat{y}_{1i} + \hat{y}_{2i}}{2}.$$

Assuming that the individual forecasts are unbiased (which the forecasting method must ensure), then the combined forecast will also be unbiased. The error of the combined forecast is defined as the average error of individual estimates:

$$e_i^c = y_i - \hat{y}_i^c = y_i - \frac{\hat{y}_{1i} + \hat{y}_{2i}}{2} = \frac{e_{1i} + e_{2i}}{2},$$

where $e_{ji} = y_i - \hat{y}_{ji}$, and $E[e_{ji}] = 0$ and $Var[e_{ji}] = \sigma_j^2$, for $j=1,2$.

The error variance of the combined forecast is calculated:

$$\begin{aligned} Var[e_i^c] &= Var\left[\frac{e_{1i} + e_{2i}}{2}\right] = E\left[\frac{e_{1i} + e_{2i}}{2}\right]^2 = \frac{1}{4}E[e_{1i}^2 + 2e_{1i}e_{2i} + e_{2i}^2] = \\ &= \frac{1}{4}\{E[e_{1i}^2] + 2E[e_{1i}e_{2i}] + E[e_{2i}^2]\} = \\ &= \frac{1}{4}\left\{\sigma_1^2 + 2\frac{E[e_{1i}e_{2i}]}{\sigma_1\sigma_2} \cdot \sigma_1\sigma_2 + \sigma_2^2\right\} = \frac{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2}{4}. \end{aligned}$$

Thus, the variance of the forecast in the case of combining two separate forecasts is calculated by the expression:

$$\sigma_c^2 = \frac{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}{4}, \quad (1)$$

where ρ is the correlation coefficient between forecast errors. In the case when the forecast estimation errors are independent, i.e., $\rho = 0$, then formula (1) is simplified:

$$\sigma_c^2 = \frac{\sigma_1^2 + \sigma_2^2}{4}. \quad (2)$$

Assume that the variances of the two individual forecasts are independent, then the variance of the combined error will be significantly less than either of the two variances. For example, let $\sigma_1^2 = \sigma_2^2 = 144$, then

$$\sigma_c^2 = \frac{144 + 144}{4} = 72.$$

But even if there is a fairly high correlation between forecasting errors, the variance of the combined forecast error will be smaller than the variance of each method separately. For example, let $\sigma_1^2 = \sigma_2^2 = 144$ and $\rho = 0,8$, then

$$\sigma_c^2 = \frac{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}{4} = \frac{144 + 144 + 2 \cdot 0,8 \cdot 12 \cdot 12}{4} = 129,6.$$

Even in this situation, a decrease in the variance of the forecast error is observed after averaging the estimates obtained by the two methods. The situation changes in the case when the variances of individual errors differ greatly. For example, let $\sigma_1^2 = 144$, $\sigma_2^2 = 16$ and $\rho = 0,8$, then

$$\sigma_c^2 = \frac{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}{4} = \frac{144 + 16 + 2 \cdot 0,8 \cdot 12 \cdot 4}{4} = 59,2.$$

If the error variances are very different from each other and the possibility of a high correlation between the prediction errors cannot be ruled out, then simply averaging the results will not improve the prediction accuracy. Thus, simple averaging can be effectively applied in cases where the variances of individual forecasting errors are approximately equal or do not differ greatly in their values.

Weighted combination of forecasts. If there is no information regarding the characteristics of individual forecast assessments, then individual forecasts are assigned different weighting factors based on subjective or expert judgments. For two forecasting methods, the combined forecast is calculated using the expression:

$$\hat{y}_i^c = w_1 \hat{y}_{1i} + w_2 \hat{y}_{2i}, \quad (3)$$

where w_1, w_2 are weighting factors. Obviously, larger values of the weighting coefficients are assigned to those individual forecasts that have a smaller variance of errors. At the same time, for the correctness of the calculations, the following condition must be fulfilled: $w_1 + w_2 = 1$. Then expression (3) changes:

$$\hat{y}_i^c = (1 - w) \hat{y}_{1i} + w \hat{y}_{2i}.$$

The determination of forecast errors for specific models and processes occurs at the stages of machine learning. Or they are determined on the training sample. This makes it possible to objectively approach the problem of choosing weighting factors. Since models that give mesh sums of squared errors of forecasts generate better forecasts, it is necessary to take this measure as a basis for determining weighting factors.

The sum of squares of forecasting errors (for a historical forecast) has the form: $see = \sum_i^N e_i^2$. Then the expressions for the weighting coefficients of individual forecasts are:

$$w_1 = \frac{see_1^{-1}}{see_1^{-1} + see_2^{-1}}, \quad w_2 = \frac{see_2^{-1}}{see_1^{-1} + see_2^{-1}},$$

where see_1, see_2 – sums of squared errors for each of the methods used.

Let the sums of squared errors for the two forecasting methods be equal $see_1 = 144$, $see_2 = 16$, then the forecast weights are:

$$w_1 = \frac{144^{-1}}{144^{-1} + 16^{-1}} = \frac{0.0069}{0.0069 + 0.0625} = 0.0994,$$

$$w_2 = \frac{16^{-1}}{144^{-1} + 16^{-1}} = \frac{0.0625}{0.0069 + 0.0625} = 0.9006.$$

Thus, a greater weighting factor was assigned to a more accurate estimate of the forecast. At the same time, the condition $\sum w_i = 1$ is fulfilled, which is necessary to achieve the correct application of the method.

The regression method is a generalization of the variance-covariance method. It can be considered as an estimation of the parameters of a regression equation of the kind

$$\hat{y}_i^c = \alpha + \sum_{n=1}^N \beta_{ni} \hat{y}_{ni}.$$

The combined forecast when using the regression method is a linear combination of N forecasts. Coefficients α, β_{ni} are estimated by the method of least squares. If all forecasts are unbiased, then the coefficient α can be neglected. In this case, the values of the coefficients will converge with the estimates of the weight coefficients w_i from the previous method.

Thus, it is possible to draw a general conclusion that when forecasting processes of an arbitrary nature, it is necessary to apply both separate methods and a combination of forecast estimates calculated using different methods. At the same time, the weighting coefficients for individual assessments can be obtained in various ways, which also contributes to the search for a better option for forecasting. It is obvious that such approaches to forecasting are better implemented in appropriate information systems with automation of data processing functions, evaluation of structures and parameters of models and forecasts based on them.

2.2. Methodology of construction of combined forecasts

Based on the study of methods of combining forecasts, a methodology was developed, the structural diagram of which is presented in Figure 1.

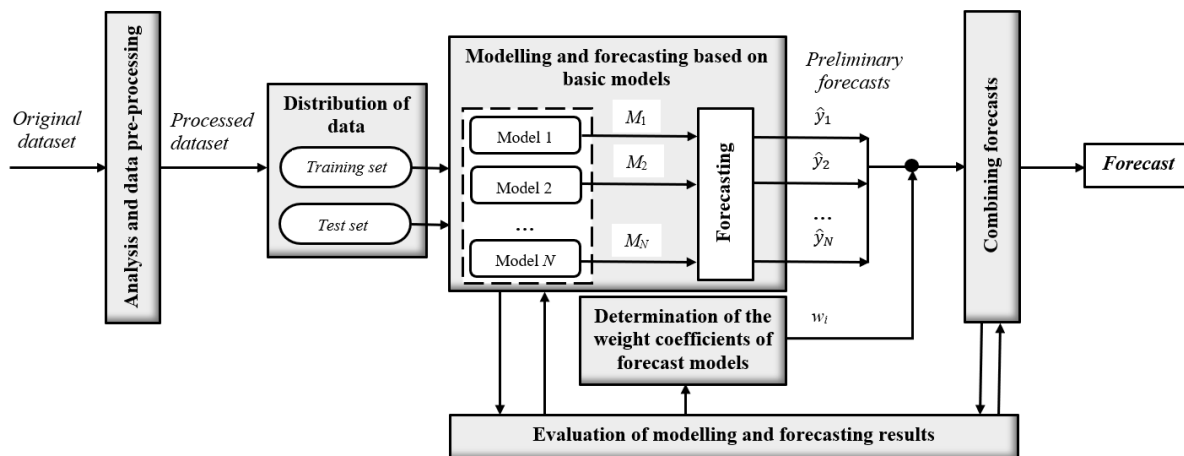


Figure 1: Structural scheme of the methodology for building combined forecasts based on time series

The structural diagram shows the methodology for building combined forecasts. The first stage of the methodology is the process of analysis and preliminary processing of the data set. At this stage, procedures are implemented: detection and processing of gaps in the data set, detection of anomalies, checking for the presence of nonlinearity, non-stationarity and their consideration, filtering and smoothing of data, etc. After this stage, the primary data set is fully prepared for the modeling process. At the second stage, the data set is divided into two parts: training and test. The next stage is modeling and forecasting based on basic models. Base models are built on the basis of selected methods. They are checked for adequacy using quality metrics, the values of which are transferred to the model evaluation results block. Preliminary forecasts are formed from the basic models. Assessments of the quality of models are the basis for the formation of weighting factors when combining forecasts. The final stage of the methodology is the stage of combining, at which the method of combining is determined and its effectiveness is determined. If an improvement in forecast accuracy is not found, it is necessary to return to the stage of forming basic models, or to change their number and type of combination. Such a structural scheme fully corresponds to the process of building combined forecasts for time series based on simple averaging of forecasts, weighted combination of forecasts and regression.

2.3. Implementation of combined forecasts

The proposed method. It was implemented as part of the forecasting information system. The architecture of the system is presented in fig. 2. The system consists of the following functional blocks: interface, data storage, data analysis and pre-preparation block, data set separation block for training and test, base model building block based on forecasting methods, combined forecast building block.

The building block of basic models contains components for assessing the quality of predictive models, which includes the coefficient of determination (R²), the Durbin-Watson criterion (DW), the Akaike information criterion (AIC), and the Bayesian information criterion (BIC). The combined forecast building block contains an evaluation procedure based on forecast quality metrics, which includes mean error (ME), root mean square error (RMSE), mean absolute error (MAE), mean percent error (MPE), mean absolute percent error (MAPE), mean absolute scaled error (MASE), root mean square scaled error (RMSSE), autocorrelation of errors at lag 1 (ACF1).

As an example of the application of the techniques of combining time series forecasts, the task of forecasting the share prices of three companies: *Amazon*, *Facebook* and *Google* are considered. The "shares" data set is loaded into the *information storage system*, which contains information about the value of companies at the time of closing of trades in the period from January 1, 2016 to May 26, 2019. These data were collected from the website <https://finance.yahoo.com/>.

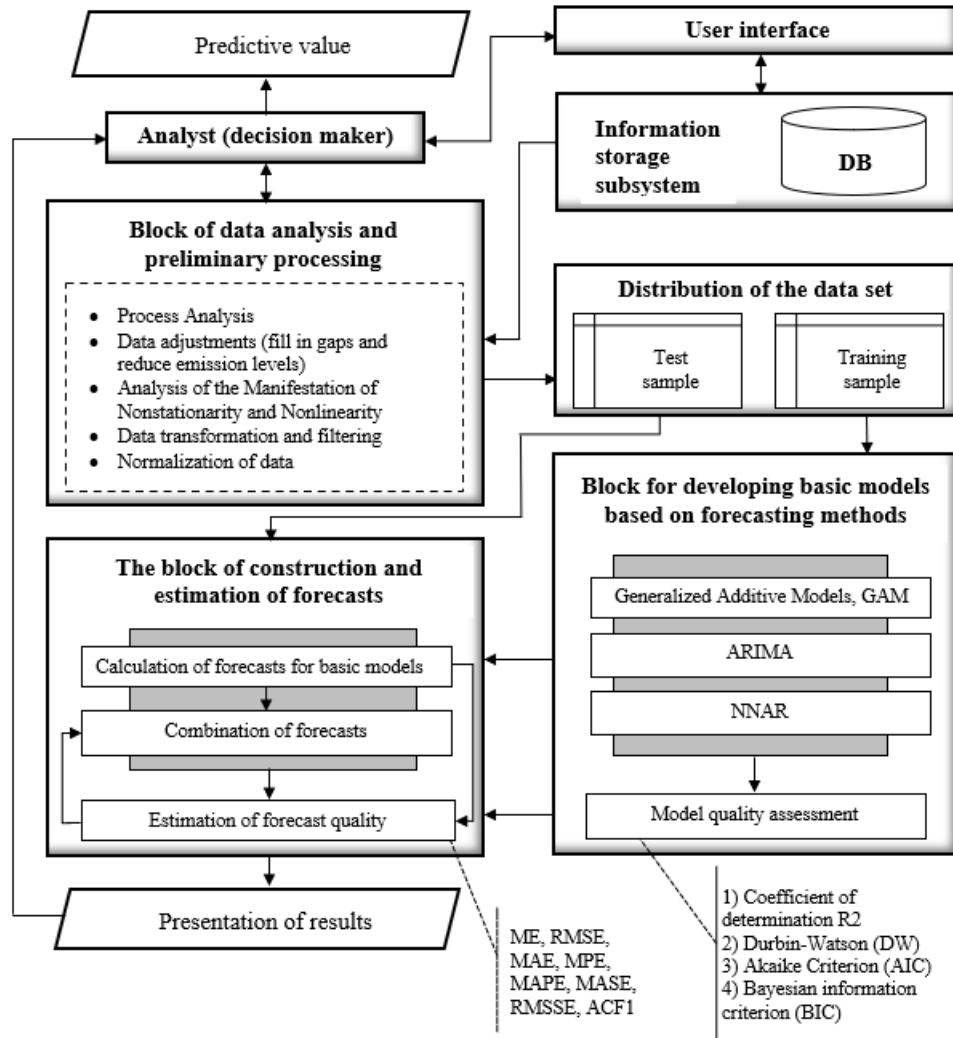


Figure 2: Architecture of forecasting information system based on time series models

After loading in the *analysis block and preliminary preparation of the data*, an analysis of the structure and types of data was first carried out, processing of missing values was performed. The data is characterized by irregular registration of observations, which leads to a large number of missing values and masking of possible seasonal fluctuations. This makes the task of forecasting quite difficult. A detailed analysis of gaps showed that their number in each data set is more than 30% with an average length of 2 consecutive gaps. To restore gaps in the time series, Kalman smoothing was used [16-19]. The gaps are filled without emissions, which is visualized in Figure 3.

Graphs (Fig. 3) of changes in the dynamics of share prices of three companies demonstrate the similarity of processes, which makes it possible to use the same types of forecast models. As a result of using a set of statistical tests (ADF, KPSS, PP), a conclusion was made about the non-stationarity of processes, which is reflected by the set of observed values of time series. The lack of stationarity of the process is also confirmed by the nature of the values of the sample autocorrelation functions ACF and PACF. Checking for non-linearity using a set of tests (terasvirta.test, white.test, Keenan.test, McLeod.Li.test, Tsay.test, tlr) revealed the presence of non-linearity.

An important condition for building reliable forecast models is a good understanding of the structure of time series. Decomposition of series using the STL method [20] made it possible to determine the main principles of modeling. First of all, it is necessary to take into account the dominant role of trend components present in the data, which represent non-linear and non-stationary behavior. There are also patterns that reflect the seasonal behavior of the data to be displayed in the models. However, their influence is insignificant. This confirms that the processes under investigation belong to the class of non-linear and non-stationary. For correct use in the process of modeling various types of models, the data in the sets were transformed using the Box-Cox transformation, differentiation, and normalized.



Figure 3: Stock price data for three companies after missing values are restored

In the *data set partitioning block*, before starting the process of building predictive models for each of the time series, the initial sets were divided into two parts: training and test samples. The last 10 observations were left as test samples, corresponding to a forecast horizon of 10 days for short-term forecasting.

ARIMA statistical models, models built on the basis of the method of fitting additive regression models (GAM) and forward propagation artificial neural networks (NNAR) are used as basic predictive models in the *modeling block*. These methods were chosen because of their ability to recognize complex patterns in time series.

ARIMA models are the result of a combination of three components: autoregressive (AR), integration (I), and moving average (MA). The Box-Jenkins algorithm [21,22] helps in choosing the best model based on the graphs of the autocorrelation function and the partial autocorrelation function. However, identifying the best model requires experience because a single data series may represent different models. However, compared to others, this methodology still differs in ease of use and especially in the accuracy of the models. Alternative ARIMA models were selected automatically and by manual selection. Automatic selection was based on the following methods: full search, quick search, search with smoothing of the input data set. Table 2 shows a comparison of ARIMA models by quality metrics for the *Amazon* time series.

Table 2

Comparison of ARIMA models on quality metrics for the *Amazon* time series

Model	Model structure (p, d, q) (P, D, Q)	AIC	BIC	R ²	DW
ARIMA 1	(0, 1, 0), (2, 0, 0)	10642.26	10662.72	0.998327	2.017735
ARIMA 2	(0, 1, 0), (0, 0, 2)	10642.32	10662.78	0.998327	2.017926
ARIMA 3	(0, 1, 2), (2, 0, 2)	-7016.39	-6975.46	0.9983226	1.992346
ARIMA 4	(0, 1, 1), (0, 0, 1)	10645.48	10665.94	0.9983226	1.997119

In Table 2, the ARIMA1 model is automatically generated on the basis of a full search, the ARIMA2 model is automatically generated on the basis of a quick search, the ARIMA3 model is

automatically generated with smoothing, and the ARIMA4 model is the best model for manual selection of parameters.

The basis for creating GAM forecasting models is the procedure for fitting additive regression models (GAM). The estimation of the parameters of the fitted model is performed using the principles of Bayesian statistics, either by the method of finding the posterior maximum, or by full Bayesian inference. For this, the Stan probabilistic programming platform is used. Preliminary data analysis revealed that the seasonal component of the time series consists of two parts: weekly and annual. But for a systematic approach, GAM will also use a monthly component, so the models were presented with the following set: monthly additive type (GAM 1); annual multiplicative (GAM 2); annual multiplicative and weekly additive (GAM 3). Table 3 shows a comparison of alternative GAM models by quality metrics for the *Amazon* time series.

Table 3
Comparison of GAM models on quality metrics for the *Amazon* time series

Model	R ²	DW
GAM 1	0.9918001	0.2011187
GAM 2	0.9909922	0.1840706
GAM 3	0.9938260	0.2595217

Artificial neural networks can be considered as a non-linear regression method. The main advantage of NN is the ability to model complex time series without prior knowledge of the data creation process. In addition, NMs are important when the communication function between input and output is unknown. The use of NN as a model in the task of forecasting time series has some peculiarities:

1. to remove the trend, the first differences are not applied to the input of the model;
2. since the previous values of the series act as explanatory variables, it is important to determine how many lags are essential for describing a specific process;
3. since the number of lags is limited, long-term trends are not simulated in such a model.

As a result of the experiments, it was possible to find a better architecture of NN (3, 10, 1). It is presented in Figure 4.

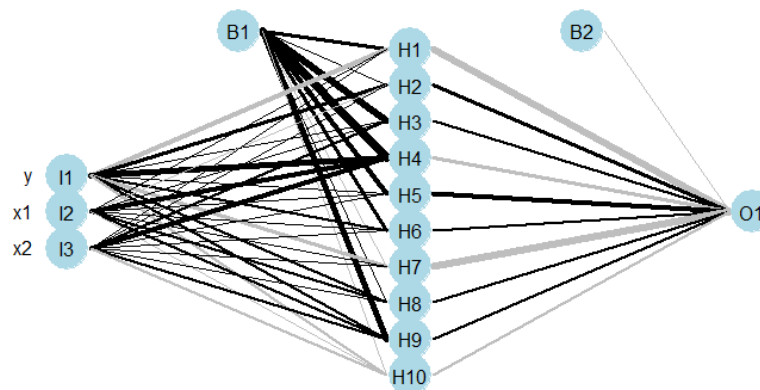


Figure 4: Neural network architecture for time series forecasting

Building block of combined forecasts. In the first step, the forecasts for each of the alternative models are calculated and evaluated, and the best one from the group of models is selected for combining. Table 4 shows the results of forecasting by various metrics for the *Amazon* time series. A fragment of the program code in the R language corresponding to the combining phase is presented in Figure 5.

```
comb_model <- amazon_price %>% model(prophet = prophet(y ~ season("year", 10,
  1, type = "multiplicative") + season("week", 10, 1, type="additive")),
  arima = ARIMA(y), nnar = nnet (y)) %>% mutate (combination = (prophet +
  arima + nnar) / 3)
```

Figure 5: Program code corresponding to the prediction combining phase

To increase the accuracy of the combined forecast, forecasting is performed on models with close variance values. The GAM model has a variance value that is significantly different from the variance of other models. Therefore, the GAM model was not considered in the next iteration of combining forecasts. Table 5 shows a comparison of forecast estimates for the *Amazon* time series for ARIMA, NNAR, and the combined model.

Table 4

Comparison of prediction performance scores for the *Amazon* time series

Model	ME	RMSE	MAE	MPE	MAPE	MASE	RMSSE
ARIMA	-60.5	67.9	60.5	-3.29	3.29	1.79	1.37
NNAR	-61.3	67.3	61.3	-3.34	3.34	1.82	1.36
GAM	-75.8	87.0	75.8	-4.13	4.13	2.25	1.76
Combination	-65.9	74.0	65.9	-3.58	3.58	1.95	1.50

Table 5

Comparison of prediction performance scores for the *Amazon* time series

Model	ME	RMSE	MAE	MPE	MAPE	MASE	RMSSE
ARIMA	-60.5	67.9	60.5	-3.29	3.29	1.79	1.37
NNAR	-61.3	67.3	61.3	-3.34	3.34	1.82	1.36
Combination	-60.9	60.8	54.2	-2.95	2.95	1.60	1.23

From the analysis of Table 5, it follows that the combined predictive model has the best quality indicators compared to the basic models [23,24]. A graphical representation of the prediction results using the combined model is shown in Figure 6. The 80% and 95% prediction intervals for each component and their combination are shown. Only the predictive part is shown.

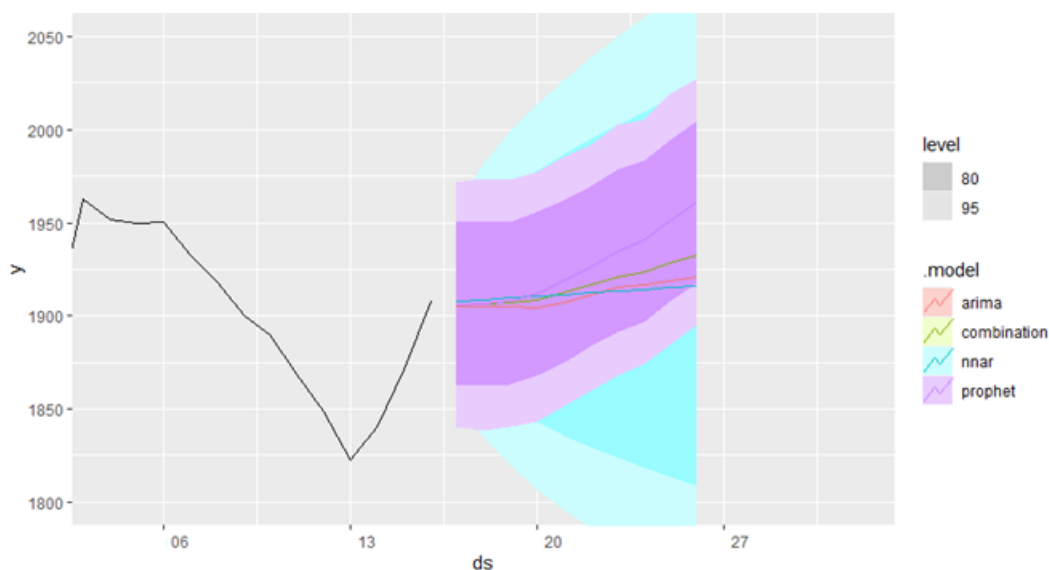


Figure 6: Graphical representation of forecasting results using the combined model

Similar results were obtained when creating combined predictive models for forecasting the dynamics of share prices of *Facebook* and *Google* companies included in the "shares" data set.

3. Conclusions

The solution to the problem of forecasting the prices of shares of commercial companies using a combination of basic forecasting models has been studied. Methods of combining forecasts based on

simple mean, weighted averaging, and regression were investigated. A methodology for building combined forecasts based on methods of combining forecast estimates has been developed. The methodology consists of the following stages: analysis and preliminary processing of the data set; division of prepared data into training and test samples; modeling and forecasting based on basic models; formation of weight coefficients of combined forecasts based on evaluations of the effectiveness of basic models; unit for combining and evaluating forecasts.

The architecture of the forecasting information system based on time series models has been developed. It has been confirmed that when forecasting processes of an arbitrary nature, it is necessary to use both separate methods and a combination of forecast estimates calculated using different methods. At the same time, the weighting coefficients for individual assessments can be obtained in various ways, which also contributes to the search for a better option for forming combined forecasts. Forecasting results using combined forecasts have been improved.

References

- [1] M. P. Clements, D. Hendry, Forecasting Economic Time Series. *Journal of the American Statistical Association* 95(450), 2000. DOI:10.1017/CBO9780511599286.
- [2] M. P. Clements, D. Hendry, Forecasting economic processes. *International Journal of Forecasting*, Vol. 14, Issue 1, 1998, pp. 111-131.
- [3] F. Petropoulos, N. Kourentzes, K. Nikolopoulos, E. Siemsen, Judgmental selection of forecasting models. Vol. 60, Issue1, 2018, pp. 34-46. <https://doi.org/10.1016/j.jom.2018.05.005>.
- [4] N. Kourentzes, G. Athanopoulos, Elucidate structure in intermittent demand series, Department of Econometrics and Business Statistics, Monash University, Working Paper 27/19, 2019, pp. 1-38.
- [5] Sh. Ma, R. Fildes, Retail sales forecasting with meta-learning. *European Journal of Operational Research*, 288(1), 2021, pp.1-39. DOI: 10.1016/j.ejor.2020.05.038.
- [6] K. A. Aastveit, B. Albuquerque, A. Anundsen, Changing supply elasticities and regional housing booms. Bank of England 2020, Staff Working Paper No. 844, pp. 1-53.
- [7] S. Ray, A. A. Abugable, J. Parker et al., A mechanism for oxidative damage repair at gene regulatory elements. Springer Nature Limited, 609 (7929), 2022, pp. 1038-1047. doi: 10.1038/s41586-022-05217-8.
- [8] P. Bidyuk, I. Kalinina, A. Gozhyj, Methodology of Constructing Statistical Models for Nonlinear Non-stationary Processes in Medical Diagnostic Systems. *IDDM'2020: 3rd International Conference on Informatics & Data-Driven Medicine, 2020, Data Stream Mining & Processing*, pp. 470-485. Växjö, Sweden. [CEUR-WS.org/Vol-2753/paper4.pdf](https://www.ceur-ws.org/Vol-2753/paper4.pdf). DOI: 10.1007/978-3-030-61656-4_32.
- [9] J. H. Stock, M. W. Watson, Combination Forecasts of Output Growth in a Seven-Country Data Set, *Journal of Forecasting*, Vol. 23, Issue 6, 2004, pp. 405-430. DOI: 10.1002/for.928.
- [10] J. Smith, K. F. Wallis, A Simple Explanation of the Forecast Combination Puzzle. *Journal of Oxford Bulletin of Economics and Statistics*, Vol. 71, Issue 3, 2009, pp. 331-355. <https://doi.org/10.1111/j.1468-0084.2008.00541.x>.
- [11] G. Claeskens, J. R. Magnus, A. L. Vasnev, W. Wang, The forecast combination puzzle: A simple theoretical explanation. *International Journal of Forecasting*, Vol. 32, Issue 3, 2016, pp. 754-762. <https://doi.org/10.1016/j.ijforecast.2015.12.005>.
- [12] F. Chan, L. Pauwels, Some theoretical results on forecast combinations. *International Journal of Forecasting*, Vol. 34, Issue 1, 2018, pp. 64-74. DOI: 10.1016/j.ijforecast.2017.08.005
- [13] A. C. B. Mancuso, L. Werner, A comparative study on combinations of forecasts and their individual forecasts by means of simulated series *Acta Scientiarum Technologia*, Vol. 41, 2019, Universidade Estadual de Maringá. DOI: <https://doi.org/10.4025/actascitechnol.v41i1.41452>.
- [14] A. Timmermann, Forecast Combinations, Chapter 04 in *Handbook of Economic Forecasting*, 2006, Vol. 1, pp. 135-196. [https://doi.org/10.1016/S1574-0706\(05\)01004-9](https://doi.org/10.1016/S1574-0706(05)01004-9).

- [15] X. Wang, R. J. Hyndman, F. Li, Y. Kang, Forecast combinations: an over 50-year review, Cornell University, 2022. arXiv:2205.04216v2 [stat.ME]. <https://doi.org/10.48550/arXiv.2205.04216>.
- [16] Y. Kim, H. Bang, Introduction to Kalman Filter and Its Applications. Open access peer-reviewed chapter, 2018. DOI: 10.5772/intechopen.80600.
- [17] Y. Pei, S. Biswas, D. S. Fussell, K. Pingali, An Elementary Introduction to Kalman Filtering. arXiv:1710.04055v5 [eess.SY] 27 Jun 2019.
- [18] T. Babb, How a Kalman filter works, in pictures. Bzarg. 2018, Accessed: 2018-11-30. <https://www.bzarg.com/p/how-a-kalman-filter-works-in-pictures/>.
- [19] A. V. Balakrishnan. Kalman Filtering Theory. Optimization Software, Inc., Los Angeles, CA, USA, 1987.
- [20] R. J. Hyndman, G. Athanasopoulos, Forecasting: principles and practice, 3rd edition, O’Texts: Melbourne, Australia. [OTexts. com/fpp3](https://www.otexts.com/fpp3). 2021.
- [21] S. J. Taylor, B. Letham, Forecasting at Scale. Journal “The American statistician, 2018, Vol.72, No.1, pp. 37-45. DOI: 10.1080/00031305.2017.1380080.
- [22] G. Box, G. Jenkins, Time Series Analysis: Forecasting and Control. San Francisco: Holden Day. 1970.
- [23] P. Bidyuk, A. Gozhyj, I. Kalinina, V. Vysotska, Methods for forecasting nonlinear non-stationary processes in machine learning. In: Data Stream Mining and Processing. DSMP 2020. Communications in Computer and Information Science. 2020, Vol. 1158, pp. 470–485. Springer, Cham, (2020). https://doi.org/10.1007/978-3-030-61656-4_32.
- [24] P. Bidyuk, I. Kalinina, A. Gozhyj, An Approach to Identifying and Filling Data Gaps in Machine Learning Procedures. International Scientific Conference “Intellectual Systems of Decision Making and Problem of Computational Intelligence” ISDMCI 2021: Lecture Notes in Computational Intelligence and Decision Making, 2021, pp. 164–176.