

ENSEMBLE OF ADAPTIVE PREDICTORS FOR MULTIVARIATE NONSTATIONARY SEQUENCES AND ITS ONLINE LEARNING

Bodyanskiy Ye. V. – Dr. Sc., Professor, Professor of the Artificial Intelligence Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Lipianina-Honcharenko Kh. V. – PhD, Associate Professor, Associate Professor of the Department for Information Computer Systems and Control, West Ukrainian National University, Ternopil, Ukraine.

Sachenko A. O. – Dr. Sc., Professor, Director of the Research Institute for Intelligent Computer Systems, West Ukrainian National University, Ternopil, Ukraine.

ABSTRACT

Context. In this research, we explore an ensemble of metamodels that utilizes multivariate signals to generate forecasts. The ensemble includes various traditional forecasting models such as multivariate regression, exponential smoothing, ARIMAX, as well as nonlinear structures based on artificial neural networks, ranging from simple feedforward networks to deep architectures like LSTM and transformers.

Objective. A goal of this research is to develop an effective method for combining forecasts from multiple models forming metamodels to create a unified forecast that surpasses the accuracy of individual models. We are aimed to investigate the effectiveness of the proposed ensemble in the context of forecasting tasks with nonstationary signals.

Method. The proposed ensemble of metamodels employs the method of Lagrange multipliers to estimate the parameters of the metamodel. The Kuhn-Tucker system of equations is solved to obtain unbiased estimates using the least squares method. Additionally, we introduce a recurrent form of the least squares algorithm for adaptive processing of nonstationary signals.

Results. The evaluation of the proposed ensemble method is conducted on a dataset of time series. Metamodels formed by combining various individual models demonstrate improved forecast accuracy compared to individual models. The approach shows effectiveness in capturing nonstationary patterns and enhancing overall forecasting accuracy.

Conclusions. The ensemble of metamodels, which utilizes multivariate signals for forecast generation, offers a promising approach to achieve better forecasting accuracy. By combining diverse models, the ensemble exhibits robustness to nonstationarity and improves the reliability of forecasts.

KEYWORDS: ensemble, metamodels, boosting, bagging, multivariate signals, nonstationarity, forecasting.

ABBREVIATIONS

MP – Multi-dimensional Predictors;
ARIMAX – AutoRegressive Integrated Moving Average with eXogenous inputs;
LSTM – Long Short-Term Memory;
RF – Random Forest;
NB – Naive Bayes;
SVM – Support Vector Machine;
LR – Logistic Regression;
AdaBoost – Adaptive Boosting;
AUC – Area Under the Curve;
DT – Decision Tree;
MIMO – Multiple-Input Multiple-Output;
LSTM – Long Short-Term Memory;
SMOTE – Synthetic Minority Over-sampling Technique;
ADASYN – Adaptive Synthetic Sampling.

NOMENCLATURE

$x(\tau)$ – multivariate signal with time index τ ;
 MP_j – member of the ensemble of models with index j ;
 $\hat{x}_j(T)$ – estimate obtained at the output of member MP_j of the ensemble;
 $x^*(\tau)$ – combined forecast of the metamodel at time τ ;

c – metamodel parameters, a vector of estimates forming the combined forecast;
 λ – Lagrange multiplier used in optimization;
 $D(T)$ – matrix used for estimating metamodel parameters;
 $d(T)$ – vector that incorporates estimates at the previous time step;
 α – regularization parameter that ensures the method's operation for nonstationary data;
 s – size of the “sliding window”, determining the number of recent observations considered in the estimation;
 $v^2(T)$ – squared error of the estimate at the last time step.

INTRODUCTION

Forecasting multivariate nonstationary signals is a relevant and challenging problem in various domains. To achieve reliable and accurate results, different forecasting models such as ARIMAX, LSTM, SVM, and many others are used.

In this work, we consider the ensemble of metamodels method for forecasting, which is based on combining forecasts from different forecasting models. The metamodel helps to merge information from various models to improve forecasting accuracy and ensure more robust results.

The object of study is an ensemble of multivariate predictors used for forecasting multivariate signals.

The subject of study is the ensemble of metamodels method for combining forecasts from different forecasting models to improve forecasting accuracy based on nonstationary signals.

The purpose of the work of this research is to develop and evaluate the effective method based on ensemble of metamodels for forecasting multivariate nonstationary signals. We aim to investigate how combining forecasts from individual models can enhance the quality of forecasting and provide more reliable results.

Forecasting tasks based on multivariate nonstationary signals find broad applications in various fields, including finance, economics, medicine, and engineering. An efficient ensemble of metamodels can become a powerful tool for addressing these tasks and ensuring accurate and reliable forecasts.

1 PROBLEM STATEMENT

Let's consider an ensemble of multivariate predictors, $MP_1, \dots, MP_j, \dots, MP_h$, each of which processes the same multivariate signal $x(\tau) = (x_1(\tau) \dots x_i(\tau) \dots)^T, \tau = 1, 2, \dots, T$. The estimate that appears at the output of each member of the ensemble will be denoted as $\hat{x}_j(\tau), j = 1, 2, \dots, h$. It is worth noting that traditional forecasting models based on multivariate regression, exponential smoothing, ARIMAXs-MIMO models (Box-Jenkins), as well as nonlinear structures based on artificial neural networks, ranging from simple shallow recurrent networks to deep architectures like LSTM or transformers, can be used as members of the ensemble.

The estimates $\hat{x}_j(\tau)$ are input to the metamodel, which forms the combined forecast of the metamodel:

$$x^*(\tau) = \sum_{j=1}^h c_j \hat{x}_j(\tau) = \hat{x}(\tau)c,$$

here

$c = (c_1, \dots, c_j, \dots, c_h)^T, \hat{x}(\tau) = (\hat{x}_1(\tau), \dots, \hat{x}_j(\tau), \dots, \hat{x}_h(\tau)) - (n \times h)$ – matrix formed by the signals at the outputs of individual models, where metamodel parameters satisfy the condition of unbiasedness:

$$\sum_{j=1}^h c_j = c^T E_h = 1,$$

here $E_h - (h \times 1)$ – is a vector formed by ones.

To solve this problem, methods of Lagrange multipliers are used, leading to the estimation of the metamodel parameters c defined in a recursive form. The case where estimation is carried out based on a “sliding window” of size s is also considered, allowing for consideration of only the last s observations from the training dataset. To choose the best metamodel, a second-level metamodel is introduced, which processes the outputs of the first-level metamodels using a meta-algorithm.

Thus, the formal mathematical formulation of the problem involves defining an ensemble of predictors, computing estimates $\hat{x}_j(\tau)$ for each member of the ensemble, constructing a combined forecast $x^*(\tau)$, determining the parameters of the metamodel c using the method of Lagrange multipliers, and the ability to work with different “sliding window” sizes and second-level metamodels for selecting the optimal solution.

2 REVIEW OF THE LITERATURE

This approach has gained the most popularity in classification tasks, such as image recognition, where the AdaBoost algorithm and its various modifications [1–8] are very popular. The underlying idea of this algorithm is stacked generalization, where the results of each member of the ensemble (stack) are combined within a metamodel, whose parameters are tuned using metalearning procedures. Typically, this involves weighted averaging, where each member of the ensemble (committee) is assigned a weight obtained through optimization of the adopted learning criterion.

The foundation of AdaBoost lies in the ideas of Bayesian estimation, logistic regression, and support vector machines. Interestingly, these ideas also form the basis of several artificial neural networks, where ensemble approaches [9–11] are also utilized to obtain optimal forecasts. In this case, weights for each member of the ensemble are estimated using an optimization procedure implemented in batch mode, making the use of known approaches for solving Data Stream Mining tasks practically impossible. Recurrent procedures for metamodel parameter tuning were introduced in [12, 13], generalizing the output signals of predictor neural networks based on the optimization of the standard least squares criterion under certain constraints. Although these procedures are designed for online evaluation, they are not adapted to work with nonstationary time series, where parameters change unpredictably at any moment.

Therefore, it is worthwhile to introduce adaptive recurrent metalearning procedures for a generalizing metamodel that combines the output signals of a neural predictor ensemble, each of which can have its own architecture and its own algorithm for tuning-learning its synaptic weights.

3 MATERIALS AND METHODS

Metamodel parameters (vector of estimates c) can be determined using the classical method of Lagrange multipliers, for which the Lagrange function is introduced:

$$\begin{aligned} L(c, \lambda) &= \text{Sp}(V^T(T)V(T)) + \lambda(c^T E_h - 1) = \\ &= \text{Sp}(X(T) - \hat{X}(T) E_{nn} \otimes c)^T (X(T) - \hat{X}(T) E_{nn} \otimes c) + \\ &\quad + \lambda(c^T E_h - 1) = \\ &= \sum_{\tau=1}^T \|x(\tau) - \hat{x}(\tau)c\|^2 + \alpha(c^T E_h - 1), \end{aligned}$$

where

$$x(T) = \begin{pmatrix} x^T(1) \\ x^T(\tau) \\ x^T(T) \end{pmatrix},$$

$$\hat{x}(T) = \begin{pmatrix} \hat{x}_1^T(1) & \hat{x}_2^T(1) & \dots & \hat{x}_h^T(1) \\ \dots & \dots & \dots & \dots \\ \hat{x}_1^T(\tau) & \hat{x}_2^T(\tau) & \dots & \hat{x}_h^T(\tau) \\ \dots & \dots & \dots & \dots \\ \hat{x}_1^T(T) & \hat{x}_2^T(T) & \dots & \hat{x}_h^T(T) \end{pmatrix}.$$

$$V(T) = X(T) - \hat{X}(T) E_{nn} \otimes c.$$

– the matrix of errors is the training sample, $E_{nn} - (n \times n)$ is the identity matrix, \otimes – denotes the tensor product, $\text{Sp}(\bullet)$ – denotes the trace of a matrix, λ – Lagrange multiplier.

Solving the Kuhn-Tucker system of equations leads to the estimate [12]:

$$c = c^* + D(T) \frac{1 - E_h^T c^*}{E_h^T D(T) E_h} E_h, \quad (1)$$

where

$$D(T) = \left(\sum_{\tau=1} \hat{x}^T(\tau) \hat{x}(\tau) \right)^{-1}, \quad (2)$$

$$c^* = D(T) \sum_{\tau=1} \hat{x}^T(\tau) x(\tau) = D(T) d(T)$$

the regular estimate of the standard least squares method.

In [13], the optimality of this estimate is proven over the entire training sample, meaning that the output of the metamodels $\hat{x}^*(\tau)$, does not compromise accuracy compared to any of the individual ensemble models $\hat{x}_j(\tau)$ in the interval from $\tau=1$ to $\tau=T$.

Equations (1) and (2) can be easily rewritten in a recursive form similar to the recursive least squares method:

$$D(T+1) = D(T) - D(T) \hat{x}^T(T+1) \times$$

$$\times (E_{nn} + \hat{x}(T+1) D(T) \hat{x}^T(T+1))^{-1} \times$$

$$\times \hat{x}(T+1) D(T),$$

$$d(T+1) = d(T) + \hat{x}^T(T+1) \hat{x}(T+1), \quad (3)$$

$$c^*(T+1) = D(T+1) d(T+1),$$

$$c(T+1) = c^*(T+1) + D(T+1) \times$$

$$\times (E_h^T D(T+1) E_h)^{-1} (1 - E_h^T c^*(T+1)) E_h.$$

The use of the least squares criterion is associated with the assumption of stationarity in the processed sequences, as all observations from $x(1)$ to $x(T)$ are assigned equal weights. Since we assume non-stationarity in the controlled signals, including abrupt changes in the forecasting model, the estimates based on the least squares method are found to be inefficient. In such situations, more suitable predictors are those synthesized using “sliding window” estimation procedures that consider not the entire training sample but only the last s (window size) observations from $x(T-s+1)$ to $x(T)$. When the value $x(T+1)$ arrives, the observation $x(T-s+1)$ is excluded from consideration, and the estimate is calculated over the interval from $x(T-s+2)$ to $x(T+1)$. In this case, the procedure takes the form:

$$D(T+1) = D(T) - D(T) \hat{x}^T(T+1) \times$$

$$\times (E_{nn} + \hat{x}(T+1) D(T) \hat{x}^T(T+1))^{-1} \hat{x}(T+1) D(T),$$

$$D^S(T+1) = D(T+1) + D(T+1) \hat{x}^T(T+S-1) \times$$

$$\times (E_{nn} - \hat{x}(T+S-1) D(T+1) \hat{x}^T(T+S-1))^{-1} \times$$

$$\times \hat{x}(T+S-1) D(T+1), \quad (4)$$

$$d^S(T+1) = d(T) + \hat{x}^T(T+1) \hat{x}(T+1) -$$

$$- \hat{x}^T(T+S-1) \hat{x}(T+S-1),$$

$$c^{*S}(T+1)$$

$$= D^S(T+1) d^S(T+1).$$

An interesting situation arises when the estimation is performed under the assumption of $s=1$, meaning that the optimization criterion (learning) is based on the square error of estimation at the last observation timestep.

$$v^2(T) = \|x(T) - \hat{x}(T)c\|^2.$$

In this case, the procedures (1), (2), and (4) take on a simple form:

$$D(T+1) = (\hat{x}^T(T+1) \hat{x}(T+1) + E_{hh})^{-1},$$

$$d^S(T+1) = \hat{x}^T(T+1) \hat{x}(T+1),$$

$$c^{*S}(E+1) = D^\alpha(T+1) d^S(T+1), \quad (5)$$

$$c^*(T+1) = C^{*\alpha}(T+1) + D^1(T+1) \times$$

$$\times (E_h^T D(T+1) E_h)^{-1} (1 - E_h^T c^{*1}(T+1)) E_h.$$

This is a generalization for the case under consideration, an adaptive identification algorithm of Kachmazh-Uidro-Hoff, where $\alpha > 0$ is a regularization parameter that ensures the possibility of inversion during the calculation of $D(T+1)$.

The most challenging issue here remains the choice of the “window” size, s , which is usually done based on purely empirical considerations since the nature of possi-

ble changes in the controlled signal $x(\tau)$ is unknown a priori. In this case, it is advisable to use not a single metamodel but a set of such structures built at different values of the “sliding window”.

To select the best metamodel from such a set, it is appropriate to introduce metamodels of the second level that process the outputs of first-level metamodels using the metaalgorithm (3), covering the entire training sample at $\tau = 1, 2, \dots, T, T + 1, \dots$

The method of constructing an ensemble of metamodels that use multidimensional signals for forecasting can be presented in the following steps 1–9 (Figure 1):

Step 1: Data Collection: Gather a large dataset of multidimensional data to be used in the analysis.

Step 2: Input Data Formation: The outputs (predictions) of each predictor are used as inputs for the metamodel.

Step 3: Data Processing: Each of the multidimensional predictors in the ensemble processes the same input data in various ways. Each predictor may include different machine learning methods, such as neural networks, support vector machines, gradient boosting, etc.

Step 4: Sliding Window Evaluation of Random Values: Model parameters are re-estimated for each new data point using only the last s observations. This ensures that the model is continuously updated with the most recent data.

Step 5: Metamodel Synthesis: Develop metamodels that use the method of Lagrange multipliers to determine their parameters. This means that the metamodel utilizes weights from different forecasts to form a single forecast. The weights of these forecasts are determined through the optimization of the Lagrange function.

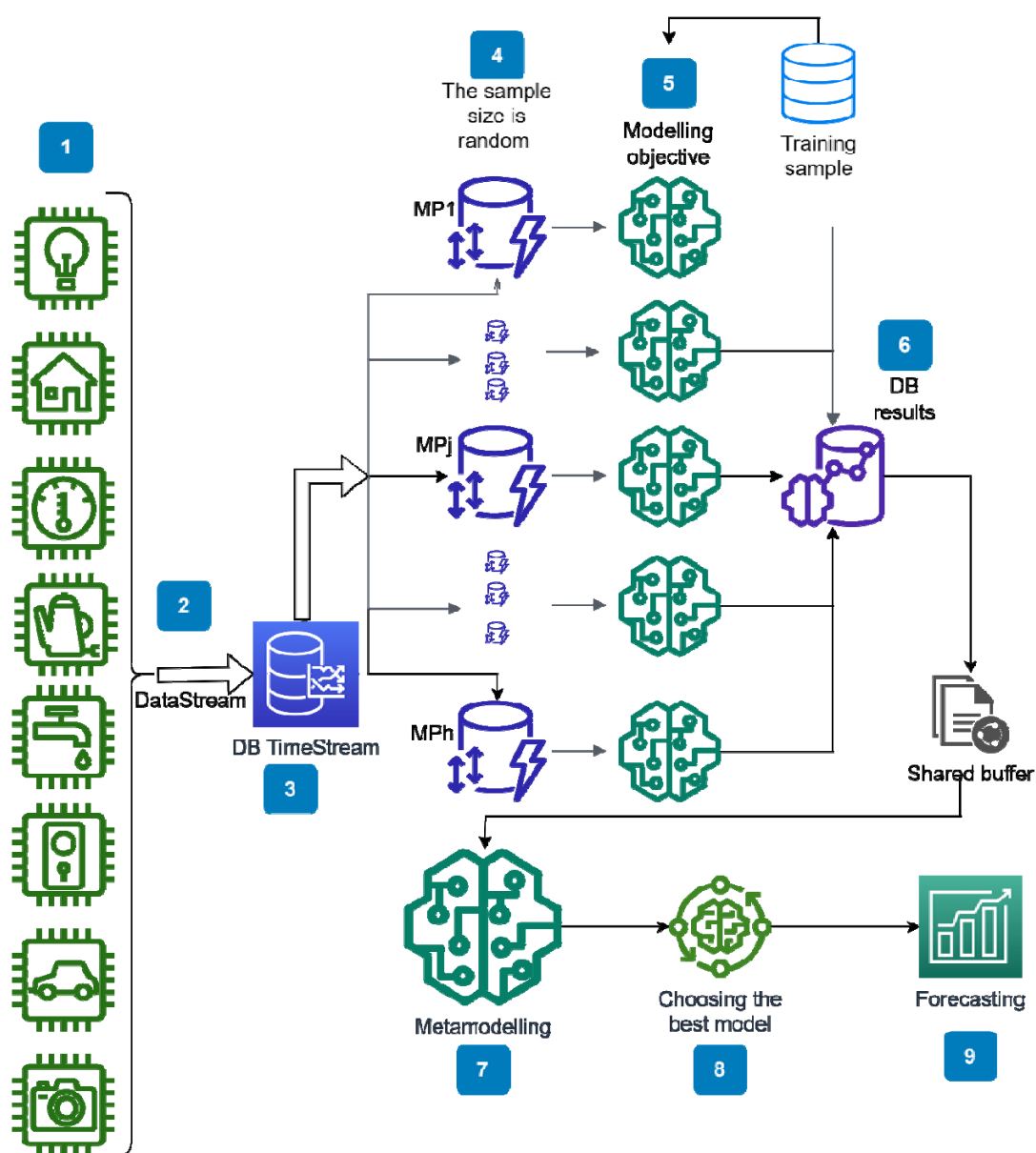


Figure 1 – Structural and Logical Diagram of the Method

Step 6: Base Results Formation: Metamodel estimates are stored in a database for further analysis.

Step 7: Synthesis of Second-level Metamodel: Develop another metamodel that processes the outputs of the first-level metamodels. This can help gather information from different metamodels and make a more accurate forecast.

Step 8: Selection of the Best Metamodel: The second-level metamodel is used to select the best metamodel among the ensemble based on their performances.

Step 9: Forecasting: The final metamodel is used to produce forecasts based on the input data. This allows using a single, optimally weighted forecast instead of independent forecasts from each predictor.

This method employs the model ensembles to work with multidimensional data and produce forecasts based on a combination of predictions from each predictor.

5 EXPERIMENTS

In the previous study [14], an intelligent method for identifying fraudulent websites was proposed. This method was implemented using various machine learning classification methods, including Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbors (KNN), Naive Bayes (NB), Support Vector Machine (SVM), and Decision Tree (DT). Additionally, each classification method was modeled using different approaches, includ-

ing addressing imbalanced data, undersampling, oversampling, SMOTE, and ADASYN.

5 RESULTS

The method was applied to a dataset of websites operating in Ukraine, consisting of 67 sites, out of which 45% were identified as fraudulent. The results showed that the DTADASYN and RF Oversampling models achieved the highest accuracy (1.0), AUC (1.0), precision (1.0), recall (1.0), and F1-score (1.0).

Using the same intelligent method for an updated dataset consisting of 1039 websites, of which 68% were identified as fraudulent, slightly different results were obtained (Table 1). The SVM Undersampling model showed an accuracy of 0.93, AUC of 0.87, precision of 0.88, recall of 0.78, and F1-score of 0.82. The KNN Undersampling model demonstrated an accuracy of 0.90, AUC of 0.94, precision of 0.69, recall of 1.0, and F1-score of 0.82. These results indicate that although accuracy and other metrics may vary depending on the dataset and methods used, the proposed intelligent method still achieves high accuracy in identifying fraudulent websites.

The proposed ensemble metamodel, utilizing multi-dimensional signals for forecasting, was implemented. In this case, the metamodel was constructed based on the predictions of logistic regression (LR), decision tree (DT), K-nearest neighbors (KNN), support vector machine (SVM), random forest (RF), and naive Bayes (NB) models, which were selected from the previous study [14].

Table 1 – Modeling Results without Metamodel for the New Dataset

	Model	Accuracy_Test	AUC_Test	PrecisionScore_Test	RecallScore_Test	F1Score_Test
1	LR Undersampling	0.939394	0.950000	0.866667	1.000000	0.928571
2	LR Oversampling	0.939394	0.950000	0.866667	1.000000	0.928571
3	LR SMOTE	0.939394	0.950000	0.866667	1.000000	0.928571
5	DT Undersampling	0.939394	0.950000	0.866667	1.000000	0.928571
10	KNN Oversampling	0.939394	0.950000	0.866667	1.000000	0.928571
11	KNN SMOTE	0.939394	0.950000	0.866667	1.000000	0.928571
12	SVM imbalance	0.939394	0.950000	0.866667	1.000000	0.928571
13	SVM Undersampling	0.939394	0.950000	0.866667	1.000000	0.928571
14	SVM Oversampling	0.939394	0.950000	0.866667	1.000000	0.928571
15	SVM SMOTE	0.939394	0.950000	0.866667	1.000000	0.928571
16	RF imbalance	0.939394	0.950000	0.866667	1.000000	0.928571
17	RF Undersampling	0.939394	0.950000	0.866667	1.000000	0.928571
18	RF Oversampling	0.939394	0.950000	0.866667	1.000000	0.928571
19	RF SMOTE	0.939394	0.950000	0.866667	1.000000	0.928571
22	NB Oversampling	0.939394	0.950000	0.866667	1.000000	0.928571
0	LR imbalance	0.909091	0.925000	0.812500	1.000000	0.896552
8	KNN imbalance	0.909091	0.925000	0.812500	1.000000	0.896552
20	NB imbalance	0.909091	0.925000	0.812500	1.000000	0.896552
21	NB Undersampling	0.909091	0.925000	0.812500	1.000000	0.896552
23	NB SMOTE	0.909091	0.925000	0.812500	1.000000	0.896552
4	DT imbalance	0.909091	0.911538	0.857143	0.923077	0.888889
7	DT SMOTE	0.909091	0.911538	0.857143	0.923077	0.888889
9	KNN Undersampling	0.909091	0.884615	1.000000	0.769231	0.869565
6	DT Oversampling	0.878788	0.886538	0.800000	0.923077	0.857143

The metamodel was built using AdaBoostClassifier, an adaptive boosting algorithm that combines several weak models to create a strong one.

The results of the metamodel were as follows (Fig. 3):

- Accuracy: 0.98. This indicates that the metamodel correctly classified 98% of the websites.
- Recall: For class 0 (non-fraudulent websites), it was 0.97, and for class 1 (fraudulent websites), it was 1.00. This means that the metamodel identified 97% of non-fraudulent websites and 100% of fraudulent websites.
- F1-score: For class 0, it was 0.98, and for class 1, it was 0.95. The F1-score is the harmonic mean between precision and recall, providing an overall evaluation of the model.

These results demonstrate improvement compared to the previous individual models trained separately. The metamodel delivers more accurate and consistent website classification, making it an effective tool for detecting fraudulent websites.

Next, we examine an example of using the metamodel to forecast the label for the 16th observation in the test dataset (Figure 3). Firstly, we obtain this observation and its true label. The true label for this observation is 0, indicating that the website is not fraudulent. Then, we get the predicted labels for this observation from each model, including logistic regression (LR), decision tree (DT), K-nearest neighbors (KNN), support vector machine (SVM), random forest (RF), and naive Bayes (NB) models.

Meta-Model Performance:				
	precision	recall	f1-score	
0	1.00	0.97	0.98	
1	0.90	1.00	0.95	
accuracy			0.98	
macro avg	0.95	0.98	0.97	
weighted avg	0.98	0.98	0.98	

Figure 2 – Metamodeling Results

True label: 0
 Predicted label from each model: [0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0]
 Predicted label from meta-model: 0

Figure 3 – Example of Applying the Metamodel to the 16th Row of the Dataset

6 DISCUSSION

In this study, we investigated the ensemble metamodel approach for forecasting multi-dimensional non-stationary signals. The proposed approach allows us to combine predictions from different forecasting models to obtain more accurate and reliable forecasts based on multiple sources of information.

Firstly, we conducted a literature review and explored various approaches to forecasting multi-dimensional non-stationary signals. Traditional models such as ARIMAX and exponential smoothing may be insufficiently effective in non-stationary conditions. On the other hand, neural networks such as LSTM and transformers exhibit high adaptability and the ability to work with changing condi-

tions, making them attractive candidates for use in metamodel ensembles. Finally, we obtain the predicted label from the metamodel for this observation. The metamodel predicts a label of 0, which aligns with the true label. This demonstrates that the metamodel can correctly classify this observation, despite varying predictions from individual models. This result underscores the effectiveness of the metamodel in combining forecasts from different models to improve overall prediction accuracy.

The metamodel exhibited high accuracy in classifying websites, achieving an accuracy of 0.98. This means that the metamodel correctly classified 98% of the websites in the test dataset. Additionally, the metamodel demonstrated high precision (0.95 for class 0 and 0.90 for class 1), recall (0.97 for class 0 and 1.00 for class 1), and F1-score (0.98 for class 0 and 0.95 for class 1). These metrics indicate that the metamodel performed well in classifying both fraudulent and non-fraudulent websites. The example prediction for the 16th observation also showed that the metamodel can accurately classify websites, despite diverse predictions from individual models. This confirms that the metamodel can effectively leverage predictions from different models to enhance the overall prediction accuracy.

Thus, these results confirm that using a metamodel can be an effective approach to improve the accuracy of classification in fraud detection tasks for websites.

tions, making them attractive candidates for use in metamodel ensembles.

Next, we performed experiments with different forecasting models, such as ARIMAX, LSTM, SVM, Random Forest, etc., and collected their forecasts as input data for the metamodel. Using the method of Lagrange multipliers, we found the optimal parameters for the metamodel to achieve the best forecasting accuracy.

The results of the experiments showed that the proposed ensemble of metamodels indeed helps improve forecast accuracy. The metamodel based on combined forecasts from different models demonstrated higher accuracy compared to individual models. This approach allows for balancing forecasts and reducing the risk of overfitting or underfitting.

Moreover, we compared various approaches to synthesizing the second-level metamodel and found that utilizing the adaptive Kachmazh-Widrow-Hoff (KWH) identification algorithm helps provide more accurate forecasts based on the forecasts from the first-level models.

Overall, the research results confirmed the effectiveness of the ensemble metamodel method for forecasting multi-dimensional non-stationary signals. Using the ensemble approach helps achieve more accurate results.

CONCLUSIONS

This research addressed the problem of adaptive forecasting of multi-dimensional non-stationary sequences, considering the prior uncertainty regarding their structure, through an ensemble approach. We developed the ensemble metamodel method, where each ensemble member processes predictions from different first-level forecasting models. Then, by collecting the results of individual models' forecasts, we applied a second-level metamodel to obtain the optimal forecast.

The scientific novelty of this study lies in the development and application of ensemble metamodels for forecasting multi-dimensional non-stationary signals. The use of ensembles allows obtaining more accurate and reliable forecasts based on multiple sources of information, reducing the impact of limitations of individual models.

The practical significance of our research is that the proposed approach can be applied in various domains where forecasting multi-dimensional non-stationary signals plays a crucial role. For example, this approach can be used in financial analysis, weather forecasting, medical diagnostics, and other fields where forecast accuracy and reliability are essential.

The conducted research confirms that the proposed ensemble metamodel has high accuracy in detecting fraudulent websites. The metamodel demonstrated high precision in website classification, correctly classifying 98% of websites in the test dataset. This demonstrates that the proposed method can be an effective tool for identifying fraudulent websites and can find practical applications in the field of cybersecurity and combating online fraudulent activities.

Regarding the prospects of this research, further improvement of the method can be achieved by expanding the set of first-level forecasting models and using more sophisticated learning algorithms for the second-level metamodel. Additionally, this approach can be applied to other types of non-stationary signals and forecasting tasks in various domains of science and technology.

ACKNOWLEDGEMENTS

n/a.

REFERENCES

1. Freund Y., Robert E. Schapire A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting [Electronic resource], *Journal of Computer and*

- System Sciences*, 1997, Vol. 55, No. 1, P. 119–139. Mode of access: <https://doi.org/10.1006/jcss.1997.1504>
2. Schwenk H., Bengio Y. Boosting Neural Networks [Electronic resource], *Neural Computation*, 2000, Vol. 12, No. 8, pp. 1869–1887. Mode of access: <https://doi.org/10.1162/089976600300015178>.
3. Jiang W. Process consistency for AdaBoost [Electronic resource], *The Annals of Statistics*, 2003, Vol. 32, No. 1, pp. 13–29. Mode of access: <https://doi.org/10.1214/aos/1079120128>
4. Li X., Wang L., Sung E. AdaBoost with SVM-based component classifiers [Electronic resource], *Engineering Applications of Artificial Intelligence*, 2008, Vol. 21, No. 5, pp. 785–795. Mode of access: <https://doi.org/10.1016/j.engappai.2007.07.001>.
5. Hastie T., Rosset S., Zhu J., & Zou H. Multi-class AdaBoost [Electronic resource], *Statistics and Its Interface*, 2009, Vol. 2, No. 3, pp. 349–360. Mode of access: <https://doi.org/10.4310/sii.2009.v2.n3.a8>
6. Zhang C., Ma Y. ed. by Ensemble Machine Learning [Electronic resource]. Boston, MA, Springer US, 2012. Mode of access: <https://doi.org/10.1007/978-1-4419-9326-7>
7. Ying C., Qi-Guang M., Jia-Chen L., & Lin G. Advance and Prospects of AdaBoost Algorithm [Electronic resource], *Acta Automatica Sinica*, 2013, Vol. 39, No. 6, pp. 745–758. Mode of access: [https://doi.org/10.1016/s1874-1029\(13\)60052-x](https://doi.org/10.1016/s1874-1029(13)60052-x).
8. Zhou Z.-H. Ensemble Learning [Electronic resource], *Machine Learning*. Singapore, 2021, pp. 181–210. Mode of access: https://doi.org/10.1007/978-981-15-1967-3_8.
9. Hansen L. K., Salamon P. Neural network ensembles [Electronic resource], *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, Vol. 12, No. 10, pp. 993–1001. Mode of access: <https://doi.org/10.1109/34.58871>
10. Wolpert D. H. Stacked generalization [Electronic resource], *Neural Networks*, 1992, Vol. 5, No. 2, pp. 241–259. Mode of access: [https://doi.org/10.1016/s0893-6080\(05\)80023-1](https://doi.org/10.1016/s0893-6080(05)80023-1).
11. Sharkey A. J. C. On Combining Artificial Neural Nets [Electronic resource], *Connection Science*, 1996, Vol. 8, No. 3–4, pp. 299–314. Mode of access: <https://doi.org/10.1080/095400996116785>
12. Bodyanskiy Y., Otto P., Pliss I., & Popov S. An Optimal Algorithm for Combining Multivariate Forecasts in Hybrid Systems [Electronic resource], *Lecture Notes in Computer Science*. Berlin, Heidelberg, 2003, pp. 967–972. Mode of access: https://doi.org/10.1007/978-3-540-45226-3_132.
13. Bodyanskiy Y., Popov S. Fuzzy Selection Mechanism for Multimodel Prediction [Electronic resource], *Lecture Notes in Computer Science*. Berlin, Heidelberg, 2004, pp. 772–778. Mode of access: https://doi.org/10.1007/978-3-540-30133-2_101.
14. Lipianina-Honcharenko K., Lukasevych-Krutnyk I., Butrynboka N., Sachenko A., & Grodskyi S. Intelligent Method for Identifying the Fraudulent Online Stores [Electronic resource], *2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T)*. Kharkiv, Ukraine, 5–7 October 2021, [S. l.], 2021. Mode of access: <https://doi.org/10.1109/picst54195.2021.9772195>

Received 11.09.2023.

Accepted 15.11.2023.

АНСАМБЛЬ АДАПТИВНИХ ПРЕДИКТОРІВ ДЛЯ БАГАТОВИМІРНИХ НЕСТАЦІОНАРНИХ ПОСЛІДОВНОСТЕЙ ТА ЙОГО ОНЛАЙН-НАВЧАННЯ

Бодяньський Є.В. – д-р техн. наук, професор, професор кафедри штучного інтелекту, Харківський національний університет радіоелектроніки, Харків, Україна.

Ліпаніна-Гончаренко Х. В. – канд. техн. наук, доцент, доцент кафедри інформаційно-комп'ютерних систем та управління, Західноукраїнський національний університет, Тернопіль, Україна.

Саченко А. О. – д-р техн. наук, професор, директор Науково-дослідного інституту інтелектуальних комп'ютерних систем, Західноукраїнський національний університет, Тернопіль, Україна.

АНОТАЦІЯ

Актуальність. У даному дослідженні ми розглядаємо ансамбль метамodelей, який використовує багатовимірні сигнали для генерації прогнозів. Ансамбль включає різні традиційні моделі прогнозування, такі як багатовимірні регресія, експоненційне згладжування, ARIMAX, а також нелінійні структури на основі штучних нейронних мереж, від простих поверхневих рекурентних мереж до глибоких архітектур, таких як LSTM і трансформери.

Мета роботи. Основною метою цього дослідження є розробка ефективного методу поєднання прогнозів декількох modelей, що утворюють метамodelі, для створення єдиного прогнозу, який перевищує точність окремих modelей. Ми прагнемо дослідити ефективність запропонованого ансамблю в контексті задач прогнозування з нестационарними сигналами.

Метод. Запропонований ансамбль метамodelей використовує метод множників Лагранжа для оцінки параметрів метамodelі. Система рівнянь Куна-Таккера розв'язується для отримання незміщених оцінок за допомогою методу найменших квадратів. Крім того, ми вводимо рекурентну форму алгоритму найменших квадратів для адаптивної обробки нестационарних сигналів.

Результати. Оцінка запропонованого ансамблю методу здійснюється на наборі даних часових рядів. Метамodelі, утворені шляхом поєднання різних окремих modelей, демонструють покращену точність прогнозу порівняно з індивідуальними modelями. Підхід проявляє ефективність в утриманні нестационарних шаблонів та покращенні загальної точності прогнозування.

Висновки. Ансамбль метамodelей, який використовує багатовимірні сигнали для формування прогнозів, пропонує перспективний підхід для досягнення кращої точності прогнозування. Шляхом поєднання різноманітних modelей, ансамбль проявляє стійкість до нестационарності та покращує надійність прогнозів.

КЛЮЧОВІ СЛОВА: ансамбль, метамodelі, бустінг, бегінг, багатовимірні сигнали, нестационарність, прогнозування.

ЛІТЕРАТУРА

1. Freund Y. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting [Electronic resource] / Yoav Freund, Robert E. Schapire // Journal of Computer and System Sciences. – 1997. – Vol. 55, No. 1. – P. 119–139. Mode of access: <https://doi.org/10.1006/jcss.1997.1504>
2. Schwenk H. Boosting Neural Networks [Electronic resource] / Holger Schwenk, Yoshua Bengio // Neural Computation. – 2000. – Vol. 12, No. 8. – P. 1869–1887. – Mode of access: <https://doi.org/10.1162/089976600300015178>
3. Jiang W. Process consistency for AdaBoost [Electronic resource] / Wenxin Jiang // The Annals of Statistics. – 2003. – Vol. 32, No. 1. – P. 13–29. Mode of access: <https://doi.org/10.1214/aos/1079120128>
4. Li X. AdaBoost with SVM-based component classifiers [Electronic resource] / Xuchun Li, Lei Wang, Eric Sung // Engineering Applications of Artificial Intelligence. – 2008. – Vol. 21, No. 5. – P. 785–795. Mode of access: <https://doi.org/10.1016/j.engappai.2007.07.001>
5. Multi-class AdaBoost [Electronic resource] / T. Hastie, S. Rosset, J. Zhu, & H. Zou // Statistics and Its Interface. – 2009. – Vol. 2, No. 3. – P. 349–360. Mode of access: <https://doi.org/10.4310/sii.2009.v2.n3.a8>
6. Ensemble Machine Learning [Electronic resource] / ed. by C. Zhang, Y. Ma. – Boston, MA : Springer US, 2012. – Mode of access: <https://doi.org/10.1007/978-1-4419-9326-7>
7. Advance and Prospects of AdaBoost Algorithm [Electronic resource] / [C. Ying, M. Qi-Guang, L. Jia-Chen, & G. Lin] // Acta Automatica Sinica. – 2013. – Vol. 39, No. 6. – P. 745–758. Mode of access: [https://doi.org/10.1016/s1874-1029\(13\)60052-x](https://doi.org/10.1016/s1874-1029(13)60052-x)
8. Zhou Z.-H. Ensemble Learning [Electronic resource] / Zhi-Hua Zhou // Machine Learning. – Singapore, 2021. – P. 181–210. Mode of access: https://doi.org/10.1007/978-981-15-1967-3_8
9. Hansen L. K. Neural network ensembles [Electronic resource] / L. K. Hansen, P. Salamon // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1990. – Vol. 12, No. 10. – P. 993–1001. Mode of access: <https://doi.org/10.1109/34.58871>
10. Wolpert D. H. Stacked generalization [Electronic resource] / David H. Wolpert // Neural Networks. – 1992. – Vol. 5, No. 2. – P. 241–259. Mode of access: [https://doi.org/10.1016/s0893-6080\(05\)80023-1](https://doi.org/10.1016/s0893-6080(05)80023-1)
11. Sharkey A. J. C. On Combining Artificial Neural Nets [Electronic resource] / Amanda J. C. Sharkey // Connection Science. – 1996. – Vol. 8, No. 3–4. – P. 299–314. Mode of access: <https://doi.org/10.1080/095400996116785>
12. An Optimal Algorithm for Combining Multivariate Forecasts in Hybrid Systems [Electronic resource] / [Y. Bodyanskiy, P. Otto, I. Pliss, & S. Popov] // Lecture Notes in Computer Science. – Berlin, Heidelberg, 2003. – P. 967–972. Mode of access: https://doi.org/10.1007/978-3-540-45226-3_132
13. Bodyanskiy Y. Fuzzy Selection Mechanism for Multimodel Prediction [Electronic resource] / Y. Bodyanskiy, S. Popov // Lecture Notes in Computer Science. – Berlin, Heidelberg, 2004. – P. 772–778. Mode of access: https://doi.org/10.1007/978-3-540-30133-2_101
14. Intelligent Method for Identifying the Fraudulent Online Stores [Electronic resource] / K. Lipianina-Honcharenko, I. Lukasevych-Krutnyk, N. Butryn-Boka et al] // 2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T), Kharkiv, Ukraine, 5–7 October 2021. – [S. l.], 2021. Mode of access: <https://doi.org/10.1109/picst54195.2021.9772195>