Forecasting Time Series in the Banking Sector Using a Machine Learning Pipeline

Olga Gorodetskaya
Department of data analisys and machine earning
Financial University under the Government of the Russian Federation
Moscow, Russia
OGorodetskaya@fa.ru

Yana Gobareva
Department of data analisys and machine earning
Financial University under the Government of the Russian Federation
Moscow, Russia
YGobareva@fa.ru

Mikhail Koroteev
Department of data analisys and machine earning
Financial University under the Government of the Russian Federation
Moscow, Russia
mvkoroteev@fa.ru

Abstract—**This article is devoted to the applications of machine learning methods for solving optimization problems in the banking sector. The literature analysis on the application of methods for forecasting time series in economics and finance is presented. A universal scenario for forecasting a large number of non-stationary time series in automatic mode has been developed. The use of the developed scenario for solving specific banking tasks to improve business efficiency, including optimizing demand for ATMs, forecasting the load on the call center and cash center, is considered. This article will be helpful for specialists dealing with the problem of forecasting economic time series and students and researchers due to a large number of links to systematic literature reviews on this topic.**

Keywords— machine learning, artificial neural networks, data mining, ATMs, time series forecasting, load forecasting, service optimization

I. Introduction

A large number of works are devoted to the application of machine learning methods for problems of forecasting time series. The most developed areas of application include stock market forecasting [1], trading systems development [2], practical examples of market forecasting [3] using machine learning models. Bahrammirzaee et al. [4] reviewed financial forecasting and planning and other financial applications using various artificial intelligence techniques such as artificial neural networks (ANNs), expert systems, and hybrid models.

Evolutionary algorithms, such as genetic algorithms, find many applications in forecasting economic and financial time series [5]; multipurpose evolutionary algorithms have been widely studied for various economic applications, including time series forecasting [6–8]. Li and Ma looked at implementations of ANN for stock price forecasting and some other financial applications [9]. Tkac et al. [10] investigated various implementations of ANN in financial applications, including for predicting stock prices. Recently Elmsili and Outtaj investigated applications of ANNs in economic and management research, including series forecasting [11].

There are a large number of systematic reviews on specific applications of financial time series forecasting. In this area, the dominance of the problem of forecasting financial and stock markets is traced. Several predictive stock market research reviews have been published based on various soft computing methods at different times [12-14]. Chatterjee et al. [15], Katarya, and Mahajan [16] focused on the ANN-based financial market predictive studies, while Hu et al. [17] focused on machine learning (ML) implementations for inventory forecasting and algorithmic trading models. In another application for time series forecasting, the researchers reviewed studies on forex forecasting using ANN [18] and various other soft computing methods [19].

Note that in all the variety of financial and economical applications of time series forecasting methods, insufficient attention is paid to more specific problems [20], such as forecasting the ATM load, which is the subject of this article. However, this does not mean that this problem has not been considered in the literature [21]. We note, for example, the works of M. Rafi, which are devoted to modeling ATM loading using classical statistical methods such as moving average autoregression [22], as well as the use of recurrent neural networks based on LSTM cells [23]. LSTM can be considered a classical method since it was created specifically for modeling time series. Therefore, in this problem, it also often finds its application, for example, in [24–26]. Also, classical and deep learning methods were applied to this problem, the performance comparison of which is given in [27,28]. However, analysis of the literature shows that classical machine learning methods such as regression analysis [29], support vector machines [30], dynamic programming [31], ARIMA [32], and gradient boosting [33] are used much more often.

Also, mention should be made of the research work carried out based on the database of the NN5 competition [34] since this competition is very closely related to our work. The dataset in this competition contains two years of daily cash withdrawals from 111 ATMs in the UK. The challenge is to predict ATM withdrawals for each ATM over the next 56 days.

The most efficient model from Andrawis et al. [35] is an ensemble of general Gaussian regression, neural network, and linear models and achieved a SMAPE of 18.95%. Venkatesh et al. [36] improved the prediction accuracy reported by

2021 14th International Conference Management of large-scale system development (MLSD) | 978-1-6654-1230-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/MLSD52249.2021.9600170

Andrawis et al. [35] by clustering similar ATMs and applying many popular neural networks. These studies show that using modern ML algorithms for forecasting time series shows better results than classical methods. Also, one of the advantages is taking into account the belonging of the ATM to a particular cluster. It allows improving the quality of the demand forecast.

In another work, Venkatesh et al. [37] further improved the results by simulating chaos in cash withdrawal time series and by referring to popular neural network architectures. However, the study did not include the effects of exogenous traits, such as the "day_of_week" dummy, whether it is a weekday or a weekend, as these traits significantly impact forecast accuracy. The type of day (work, weekend) is significant in the problem under study. Therefore, this study will consider the effect of the days of the week on the amount of demand.

None of the articles shows the importance of detecting and marking periodic anomalies in time series concerning problems of this kind. Detecting anomalies will become one of the critical points of research in this work - taking into account the influence of periodic anomalies associated with regularly recurring events. However, the article [38] examines the methodology for detecting anomalies in time series using the CUSUM method and shows the applicability to financial data (value of stock prices). The results obtained in the article for marking anomalies by this method will be applicable in this work.

II. Methods

A universal pipeline for forecasting time series was developed to solve this problem, consisting of the following steps:

A. *Formation of a Feature Space*

For any machine learning algorithm, a feature space is required. Extensive feature space typical for time series (Fig. 1):

One-hot encoding of calendar features (day of the week, month, weekend, holiday, reduced working day, and similar)

Lagged variables (time series values for previous days)

Rolling statistics grouped by calendar features (average, variance, minimum, maximum, and similar)

Events of massive payments (advance payment, salary, and similar).

The selection of the necessary parameters (window width in statistics, the number of lags, and similar) and the selection of significant features are performed for each time series individually using cross-validation in a sliding window (Fig. 2).

TimeStamp	Target	One-Hot-Encoding					Lags		Rolling Statistics				One-Hot-Encoding	
	y	пн	вт	ср	чт	пт	lag_1	lag_2	rolling_mean	rolling_std	rolling_mean_weekday	rolling_min_weekday	зарплата 15	праздник
2017-09-08	1273900.0	0	0	0	0	1	1118000.0	896500.0	653100.000000	391381.906412	513666.666667	306200.0	0	0
2017-09-09	344700.0	0	0	0	0	0	1273900.0	1118000.0	726185.714286	457375.871790	143900.000000	72500.0	0	0
2017-09-10	45200.0	0	0	0	0	0	344700.0	1273900.0	746514.285714	432713.001224	36466.666667	6200.0	0	0
2017-09-11	930000.0	1	0	0	0	0	45200.0	344700.0	740328.571429	443853.266723	843400.000000	443700.0	0	0
2017-09-12	43900.0	0	1	0	0	0	930000.0	45200.0	735400.000000	441132.391314	494900.000000	468100.0	0	0
2017-[illegible]	544300.0	0	0	1	0	0	43900.0	930000.0	664600.000000	511907.166063	546400.000000	319400.0	0	0

Fig. 1. Figure 1 Generated feature space.

B. *Anomaly Detection*

Anomaly detection plays a vital role in time series analysis and forecasting. Training on a time series with a large number of outliers can significantly reduce the prediction accuracy.

Anomalies can be classified into two types:

periodic anomalies;

anomalies that do not have a periodicity.

The first class can be used as an element of the feature space for further forecasting, while the second can exclude the sample.

CUSUM (Cumulative Sum) is a method that tracks variations in a process (time series). It is based on the accumulation of changes in the time series from the mean. As soon as the cumulative sum exceeds the specified threshold value, this point is marked as an anomaly. The detection process is carried out through iterative learning and prediction with a moving window [38]. The method's parameters are calculated on the training set, and anomalies are detected on the test set.

Mathematical interpretation:

$$S_i^{+}=\left(0, S_{i-1}^{+}+x_i-(\mu + drift * \sigma)\right)$$

$$S_i^{-}=\left(0, S_{i-1}^{-}+x_i-(\mu - drift * \sigma)\right)$$

$$Anomalies = \{1,\ S_i^{+} > threshold * \sigma \ or\ S_i^{-} < -threshold * \sigma \ \ 0, otherwise$$

where:

S_i^{+} - the upper limit of the cumulative sum;

S_i^{-} - the lower limit of the cumulative amount;

x_i - values of the time series in the test window;

μ - the average of the time series in the learning window;

σ - standard deviation of the time series in the learning window;

$drift$ is the number of standard deviations to summarize the changes;

$threshold$ - the number of standard deviations for the threshold value.

The CUSUM anomaly detection algorithm also helps to identify periodic emissions determined by the days of massive payments: salary, advance payment, pension, and its recursive application in a sliding window to extract significant and relevant signs of relevant events.

Correct marking of periodic anomalies is necessary for efficient generation of the feature space for the subsequent construction of a forecast.

A large number of anomalies noise the search for periodic values weaker in the signal. The first ones are marked following the transfers (weekends, holidays) and are removed from the original series. Then the search algorithm is rerun.

It was revealed that the peak values of the anomalies are on the 10th and 11th of each month, which can be interpreted as salary days. There is also a less pronounced anomaly on the 25th. This anomaly is explained by the fact that this is an advance, significantly less than the basic salary.

C. Feature Selection

Feature selection is performed in a greedy way. All possible subsets of the feature space are enumerated. In this case, all possible subsets of the feature space are considered and obtained 2n-1 options (without an empty set). It is worth noting that such a feature selection method is computationally laborious. Therefore, parallelization of computations for all processors is built into this method.

When selecting features, cross-validation was used for time series, the primary meaning of sequential movement, and the time series, increasing the training sample. In this case, the number of folds for dividing the sample is taken equal to 5. As a result, we obtain a histogram of the importance of features.

D. Building Models

Models based on ensembles of decision trees are used to build time-series forecasts: a random forest, gradient boosting over decision trees. The model is implemented in Python: sklearn, xgboost, lightgbm, catboost. The choice of the best modeling method is carried out for a given time series in cross-validation.

It should be noted that cross-validation for a time series is arranged differently than for other types of samples, where the sequence of observations is not essential. For most machine learning problems, it is customary to assert the independence of observations. Therefore, to split the sample into folds in cross-validation, random mixing of the sample elements is allowed. However, this assumption about the independence of observations is incorrect for time series since the subsequent values of the series depend on the previous ones. Therefore, the cross-validation process for time series is different. The model is trained at some interval of the series from the initial point to some value t. Then a forecast is made for the horizon n, for the interval of the series from t to t + n. The error on this fold is calculated.

Further, the training sample increases from the starting point of the series to the point t + n. Forecasting is carried out for the horizon n, that is, for the interval of the series from t + n to t + 2 × n. Furthermore, similar, until the end of the sample is reached. The number of folds is calculated as the number of row intervals n-that fit from the start to the end of the row. This approach is commonly referred to in the literature as "sliding window cross-validation".

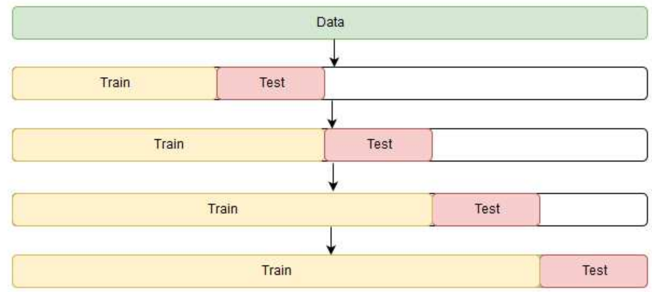

Fig. 2. Cross-validation in a sliding window.

E. Selection of Hyperparameters

It is necessary to perform a selection of hyperparameters to improve the quality of predictive models. F grid is set (Cartesian product of all fitted parameters) to select the optimal values of hyperparameters, and the selection method is carried out to select the combination that gave the best quality on the cross-validation sample. This approach has traditionally been used in machine learning and is called grid search. However, this approach is computationally laborious both in terms of time and memory, especially if the number of objects for forecasting is too large since it is required to reapply the model for the entire sample at each subsequent iteration. Also, this approach does not take into account the results of previous iterations.

Another approach to the selection of hyperparameters is the application of Bayesian optimization. It should be noted that this approach has shown itself well in practice. The essence of this approach lies in the fact that based on the experiments already carried out, the best hyperparameters are predicted and selected, and not just the given values of the hyperparameters are enumerated.

So, at the first step, let us assume that the quality of the model is a Gaussian process from hyperparameters. The initialization of hyperparameters is randomly selected.

At step n, the first step is to select the next point x_n:

$$x_n = \operatorname{argmax}_x \int_{f(\hat{x})}^{\infty} \left(f(x) - (\hat{x})\right) \times p \left(\frac{f}{x}\right)$$

where:

f (x) is our current guess;

$\hat{x}$ is the current optimal set of hyperparameters;

On step n, at the second stage, we update the obtained assumption about the function f. Thus, the process of searching for the optimal values of hyperparameters iteratively seeks improvements in their values in the most likely containing regions, taking into account the previous experiments.

III. Conclusion

One of the essential tasks of the banking sector is the task of predicting the optimal amount of loading cash into ATMs since with proper planning and management of cash flows; it is possible to minimize the costs of servicing cash flows, reduce operating costs, and increase the efficiency of using the ATM network, which will ultimately lead to the profitability of the bank.

The solution to this problem is possible using new information technologies based on machine learning, which means algorithms, methods, processes that allow a computer to draw conclusions based on data without following specific rules. These methods are based on the principle of training the model on the available data. Machine learning is used to analyze a large amount of information and identify implicit, hidden relationships between data.

After analyzing a large number of sources on the application of machine learning methods for problems of forecasting time series, we found that the most developed areas of application include forecasting the financial and stock markets, the development of trading systems. However, insufficient attention is paid to more specific problems, such as predicting the load of ATMs, which was the subject of the authors' research.

A literature review has shown that classical machine learning methods such as regression analysis, support vector machines, dynamic programming, ARIMA, and gradient boosting are much more commonly used. The authors' research led to the conclusion that the direction of using modern ML algorithms for forecasting time series shows results better than that of classical methods.

The authors also noted that scientific studies did not show the importance of detecting and marking periodic anomalies in time series concerning problems of this kind, which became one of the critical points of the study in this work - taking into account the influence of periodic anomalies associated with regularly recurring events.

After conducting a comparative analysis of models for forecasting time series, the authors concluded that there is no uniform approach to forecasting all types of time series. The methods used differ for stationary and non-stationary time series. The choice of a model is often based on a preliminary visual assessment of the data in question. If a time series has a trend, regression models show the best predictive ability, and exponential smoothing models show the best predictive ability for series with a pronounced seasonality.

Given the specificity of time series in the banking sector, such as their non-stationarity for the most part, as well as their large number, it is proposed to apply forecasting methods based on modern machine learning (decision trees, random forest, gradient boosting over decision trees, neural networks, and others).

The use of machine learning methods has made it possible to create a universal pipeline applicable to predicting a large number of time series for solving problems in the banking sector.

The universal pipeline for time series forecasting, developed by the authors, includes a sequence of performing the following stages: (i) formation of feature space, including One-hot encoding of calendar features, lag variables, sliding statistics grouped by calendar features, events of mass payments; (ii) detection of anomalies, allowing to identify of any periodic outliers for efficient generation of the feature space for the subsequent construction of a forecast; (iii)feature selection based on cross-validation for time series; (iv) building models based on ensembles of decision trees: random forest, gradient boosting over decision trees for making predictions; (v) selection of hyperparameters to improve the quality of predictive models.

Thus, the development of predictive models of non-stationary time series according to the above methodology allows saving labor costs for several weeks by automatically generating the feature space, selecting features, selecting the best model, and effectively selecting the hyperparameters. The built-in detection of anomalies improves the forecasting quality by 10% by identifying periodic anomalies and generating the corresponding features.

REFERENCES

[1] Aliev RA, Fazlollahi B, Aliev RR. Soft Computing in Electronic Business. Soft Computing and its Applications in Business and Economics. 2004. pp. 431-446. doi:10.1007 / 978-3-540-44429-9_9

[2] Dymowa L. Soft Computing in Economics and Finance. Intelligent Systems Reference Library. 2011. doi:10.1007 / 978-3-642-17719-4

[3] Kovalerchuk B, Vityaev E. Data Mining in Finance: Advances in Relational and Hybrid Methods. Springer Science & Business Media; 2006. Available: https://play.google.com/store/books/details?id=quDlBwAAQBAJ

[4] Bahrammirzaee A. A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. Neural Comput Appl. 2010; 19: 1165-1195. doi:10.1007 / s00521-010-0362-z

[5] Chen SH. Genetic Algorithms and Genetic Programming in Computational Finance: An Overview of the Book. Genetic Algorithms and Genetic Programming in Computational Finance. 2002. pp. 1-26. doi:10.1007 / 978-1-4615-0835-9_1

[6] Tapia MGC, Coello Coello CA. Applications of multi-objective evolutionary algorithms in economics and finance: A survey. 2007 IEEE Congress on Evolutionary Computation. 2007. doi:10.1109 / cec.2007.4424516

[7] Ponsich A, Jaimes AL, Coello Coello CA. A Survey on Multiobjective Evolutionary Algorithms for the Solution of the Portfolio Optimization Problem and Other Finance and Economics Applications. IEEE Transactions on Evolutionary Computation. 2013. pp. 321-344. doi:10.1109 / tevc.2012.2196800

[8] Aguilar-Rivera R, Valenzuela-Rendón M, Rodríguez-Ortiz JJ. Genetic algorithms and Darwinian approaches in financial applications: A survey. Expert Systems with Applications. 2015. pp. 7684-7697. doi:10.1016 / j.eswa.2015.06.001

[9] Li Y, Ma W. Applications of Artificial Neural Networks in Financial Economics: A Survey. 2010 International Symposium on Computational Intelligence and Design. 2010. doi:10.1109 / iscid.2010.70

[10] Tkáč M, Verner R. Artificial neural networks in business: Two decades of research. Applied Soft Computing. 2016. pp. 788-804. doi:10.1016 / j.asoc.2015.09.040

[11] Elmsili B, Outtaj B. Artificial neural networks applications in economics and management research: An exploratory literature review. 2018 4th International Conference on Optimization and Applications (ICOA).:2018.doi10.1109 / icoa.2018.8370600

[12] Nair BB, Mohandas VP. Artificial intelligence applications in financial forecasting - a survey and some empirical results. Intelligent Decision Technologies. 2014. pp. 99-140. doi:10.3233 / idt-140211

[13] Cavalcante RC, Brasileiro RC, Souza VLF, Nobrega JP, Oliveira ALI. Computational Intelligence and Financial Markets: A Survey and Future

Directions. Expert Systems with Applications. 2016. pp. 194-211. doi:10.1016 / j.eswa.2016.02.006

[14] Atsalakis GS, Valavanis KP. Surveying stock market forecasting techniques - Part II: Soft computing methods. Expert Systems with Applications. 2009. pp. 5932-5941. doi:10.1016 / j.eswa.2008.07.006

[15] Chatterjee A, Felix Ayadi O, Boone BE. Artificial Neural Network and the Financial Markets: A Survey. Managerial Finance. 2000. pp. 32-45. doi:10.1108 / 03074350010767034

[16] Katarya R, Mahajan A. A survey of neural network techniques in market trend analysis. 2017 International Conference on Intelligent Sustainable Systems (ICISS).:2017.doi10.1109 / iss1.2017.8389302

[17] Hu Y, Liu K, Zhang X, Su L, Ngai EWT, Liu M. Application of evolutionary computation for rule discovery in stock algorithmic trading: A literature review. Applied Soft Computing. 2015. pp. 534-551. doi:10.1016 / j.asoc.2015.07.008

[18] Hewamalage H, Bergmeir C, Bandara K. Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. International Journal of Forecasting. 2021. pp. 388-427. doi:10.1016 / j.ijforecast.2020.06.008

[19] Pradeepkumar D, Ravi V. Soft computing hybrids for FOREX rate prediction: A comprehensive review. Computers & Operations Research. 2018. pp. 262-284. doi:10.1016 / j.cor.2018.05.020

[20] Sezer OB, Gudelek MU, Ozbayoglu AM. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. Applied Soft Computing. 2020. p. 106181. doi:10.1016 / j.asoc.2020.106181

[21] Ekinci Y, Serban N, Duman E. Optimal ATM replenishment policies under demand uncertainty. Operational Research.:2019.doi10.1007 / s12351-019-00466-4

[22] Rafi M, Wahab MT, Khan MB, Raza H. ATM Cash Prediction Using Time Series Approach. 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET).:2020.doi10.1109 / icomet48670.2020.9073937

[23] Asad M, Shahzaib M, Abbasi Y, Rafi M. A Long-Short-Term-Memory Based Model for Predicting ATM Replenishment Amount. 2020 21st International Arab Conference on Information Technology (ACIT). 2020. doi:10.1109 / acit50332.2020.9300115

[24] Arabani SP, Komleh HE. The Improvement of Forecasting ATMs Cash Demand of Iran Banking Network Using Convolutional Neural Network. Arabian Journal for Science and Engineering. 2019. pp. 3733-3743. doi:10.1007 / s13369-018-3647-7

[25] Serengil SI, Ozpinar A. ATM Cash Flow Prediction and Replenishment Optimization with ANN. Uluslararası Muhendislik Arastirma ve Gelistirme Dergisi. 2019. pp. 402-408. doi:10.29137 / umagd.484670

[26] Rong J, Wang D. Research on Prediction of the Cash Usage in Banks Based on LSTM of Improved Gray Wolf Optimizer. Journal of Physics: Conference Series. 2021. p. 012031. doi:10.1088 / 1742-6596 / 1769/1/012031

[27] Vangala S, Vadlamani R. ATM Cash demand forecasting in an Indian Bank with chaos and deep learning. 2020. Available: http://arxiv.org/abs/2008.10365

[28] Hasheminejad SMH, Reisjafari Z. ATM management prediction using Artificial Intelligence techniques: A survey. Intelligent Decision Technologies. 2017. pp. 375-398. doi:10.3233 / idt-170302

[29] Rajwani A, Syed T, Khan B, Behlim S. Regression Analysis for ATM Cash Flow Prediction. 2017 International Conference on Frontiers of Information Technology (FIT).:2017.doi10.1109 / fit.2017.00045

[30] Jadwal PK, Jain S, Gupta U, Khanna P. K-Means Clustering with Neural Networks for ATM Cash Repository Prediction. Information and Communication Technology for Intelligent Systems (ICTIS 2017) - Volume 1. 2018. pp. 588-596. doi:10.1007 / 978-3-319-63673-3_71

[31] Ozer F, Toroslu IH, Karagoz P, Yucel F. Dynamic Programming Solution to ATM Cash Replenishment Optimization Problem. Intelligent Computing & Optimization. 2019. pp. 428-437. doi:10.1007 / 978-3-030-00979-3_45

[32] Khanarsa P, Sinapiromsaran K. Multiple ARIMA subsequences aggregate time series model to forecast cash in ATM. 2017 9th International Conference on Knowledge and Smart Technology (KST). 2017. doi:10.1109 / kst.2017.7886096

[33] Shcherbitsky VV, Panachev AA, Medvedeva MA, Kazakova EI. On the prediction of dispenser status in ATM using gradient boosting method. PROCEEDINGS OF THE INTERNATIONAL CONFERENCE OF COMPUTATIONAL METHODS IN SCIENCES AND ENGINEERING 2019 (ICCMSE-2019).:2019.doi10.1063 / 1.5137948

[34] Time Series Forecasting Competition for Neural Networks and Computational Intelligence - Homepage. [cited May 5 2021]. Available: http://www.neural-forecasting-competition.com/NN5/

[35] Andrawis RR, Atiya AF, El-Shishiny H. Forecast combinations of computational intelligence and linear models for the NN5 time series forecasting competition. Int J Forecast. 2011; 27: 672–688. doi:10.1016 / j.ijforecast.2010.09.005

[36] Venkatesh K, Ravi V, Prinzie A, Van den Poel D. Cash demand forecasting in ATMs by clustering and neural networks. European Journal of Operational Research. 2014. pp. 383-392. doi:10.1016 / j.ejor.2013.07.027

[37] Kamini V, Ravi V, Nagesh Kumar D. Chaotic time series analysis with neural networks to forecast cash demand in ATMs. 2014 IEEE International Conference on Computational Intelligence and Computing Research. 2014. doi:10.1109 / iccic.2014.7238399

[38] Fremdt S. Page's sequential procedure for change-point detection in time series regression. Statistics. 2015. pp. 128-155. doi:10.1080 / 02331888.2013.870568