RESEARCH ON THE APPLICATION OF TIME SERIES IN EARLY RISK DETECTION SYSTEMS FOR IT PROJECTS

Stanislav Kholodov
Kharkiv National University of Radio Electronics, Ukraine
Tetiana Fesenko
Kharkiv National University of Radio Electronics, Ukraine

Abstract

The article substantiates the relevance of implementing early risk detection systems in IT projects based on the analysis of time series formed from unstructured data of digital ecosystems and social media. The modern environment for implementing IT projects is characterised by high dynamism, non-linearity of information flows and a significant influence of subjective factors of stakeholders, which makes traditional monitoring methods insufficiently effective for identifying hidden threats.

The authors define the role of time series as 'leading indicators' that allow the transformation of unstructured user responses (number of mentions, tone of messages, frequency of incidents) into formalised indicators. A detailed analysis of the mathematical apparatus for decomposing time series into trend, seasonal, cyclical, and random components is conducted. Particular attention is paid to the interpretation of the random component as a source of signals about the occurrence of significant management events and anomalies.

The methodology for identifying models within the Box-Jenkins approach is considered. A description and visualisation of the autocorrelation (ACF) and partial autocorrelation (PACF) functions, which are key to determining the parameters of autoregression and moving average models, is provided. The use of ARIMA and SARIMA models for forming a baseline level of activity is justified. Deviations of actual data from this level are interpreted as early risk signals, which allows project management to move from reactive to proactive management.

A separate stage of the research is the analysis of the capabilities of deep learning methods, in particular Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures. It has been proven that these models, thanks to gate mechanisms, effectively solve the problem of long-term dependencies and noise filtering in social media data. The combination of classical statistical methods with neural network architectures allows the creation of adaptive monitoring systems that meet the requirements of PMBOK standards and Agile methodologies.

The results of the study confirm that the use of time series ensures the objectification of interaction with stakeholders and increases the manageability of IT projects by identifying latent risks in near real time. The practical significance of the work lies in the possibility of integrating the proposed models into information and analytical decision support systems for project managers and system analysts.

Keywords: IT project, risk management, time series, ARIMA, SARIMA, monitoring, stakeholders, LSTM, GRU.

Introduction. Modern IT projects are implemented in a dynamic environment characterised by rapidly changing requirements, the significant role of stakeholders, and considerable reputational risks [1–3]. Traditional monitoring methods often fail to identify hidden threats in a timely manner, making it critically important to implement early risk detection systems based on objective data from digital platforms.

Data from social media and digital ecosystems reflect the real reactions of users and customers to project results, acting as 'leading indicators' of potential problems. Information flows from social platforms are non-stationary, noisy and non-linear, requiring the use of time series mathematics to convert them into formalised indicators. The use of predictive models meets the requirements of A Guide to the Project Management Body of Knowledge (PMBOK), Agile, and others for continuous monitoring of key performance indicators (KPIs) and adaptive decision-making. The combination of classical statistical methods (AutoRegressive Integrated Moving Average / Seasonal AutoRegressive Integrated Moving Average, ARIMA/SARIMA) for forming a baseline with modern deep learning architectures (Long Short-Term Memory, LSTM; Gated Recurrent Unit, GRU) allows complex risk patterns to be identified in near real time. In this regard, the development of early risk detection systems based on time series analysis is relevant, as it allows unstructured information flows to be transformed into a strategic resource for improving the predictability and manageability of IT projects.

The purpose of this study is to justify the use of time series analysis and forecasting methods for creating early risk detection systems in IT projects. This involves transforming unstructured data flows from social media into formalised indicators for information and analytical support of monitoring and control processes within the framework of classical, flexible and hybrid approaches.

To achieve this goal, the following tasks must be accomplished:

- analyse the role of time series in the processes of monitoring and controlling project activities;
- identify opportunities for applying predictive models to manage risks and stakeholder expectations;
- justify the use of deep learning methods (LSTM, GRU) to detect anomalies and hidden risk signals in high-noise data from digital platforms.

Time series as a tool for monitoring and controlling IT projects. In the tasks of information and analytical support for IT project management, the time series Y_t is considered as a derivative indicator formed by aggregating and quantifying unstructured data from social media and digital platforms. The value Y_t It can reflect the number of mentions of the project or its individual functionality (in units of 'messages per time interval'), the average tone of messages (on a dimensionless scale, for example, from -1 to +1), or the frequency of complaints or incidents (the number of events per time interval t). The time interval t is determined by the requirements of the management task and can correspond to hours, days or weeks of observation. The use of time series provides: the formation of a baseline level of activity for monitoring deviations; the detection of early risk signals through residual component analysis; support for adaptability in Agile and hybrid approaches; objectification of stakeholder interaction based on quantitative indicators.

Decomposition of time series and interpretation of components. Time series serve as 'leading indicators' that allow moving from stating facts to proactive management, ensuring transparency and controllability of the project at all stages of its life cycle. The main task of such models is to form a mathematically sound forecast of the situation, which is the basis for managing stakeholder expectations and responding to threats in a timely manner.

To analyse the dynamics of indicators, the time series is decomposed into trend, seasonal, cyclical and random components. In the case of an additive decomposition model, the time series is described by equation (1), and in the case of a multiplicative model, by equation (2):

$$Y_t = R_t + S_t + C_t + E_t, \quad (1)$$

$$Y_t = R_t \times S_t \times C_t \times E_t. \quad (2)$$

where R_t - trend component reflecting long-term changes in the level of attention to the project or the level of stakeholder satisfaction, measured in the same units as the initial time series Yt (number of mentions or average tone level per unit of time);

S_t – seasonal component, which characterises regular periodic fluctuations in the indicator caused by sprints, releases or marketing activities;

C_t – the cyclical component characterises regular periodic fluctuations in the indicator associated with sprints, software releases or marketing activities, and is also expressed in the corresponding units of measurement of the series. Yt;

E_t– random (residual) component, contains irregular, unpredictable deviations from normal dynamics and is interpreted as a manifestation of significant management events; its values are also measured in units Yt.

Figure 1 shows an example of a non-stationary time series illustrating a non-stationary time series with a trend and quasi-periodic fluctuations characteristic of user activity on social media.

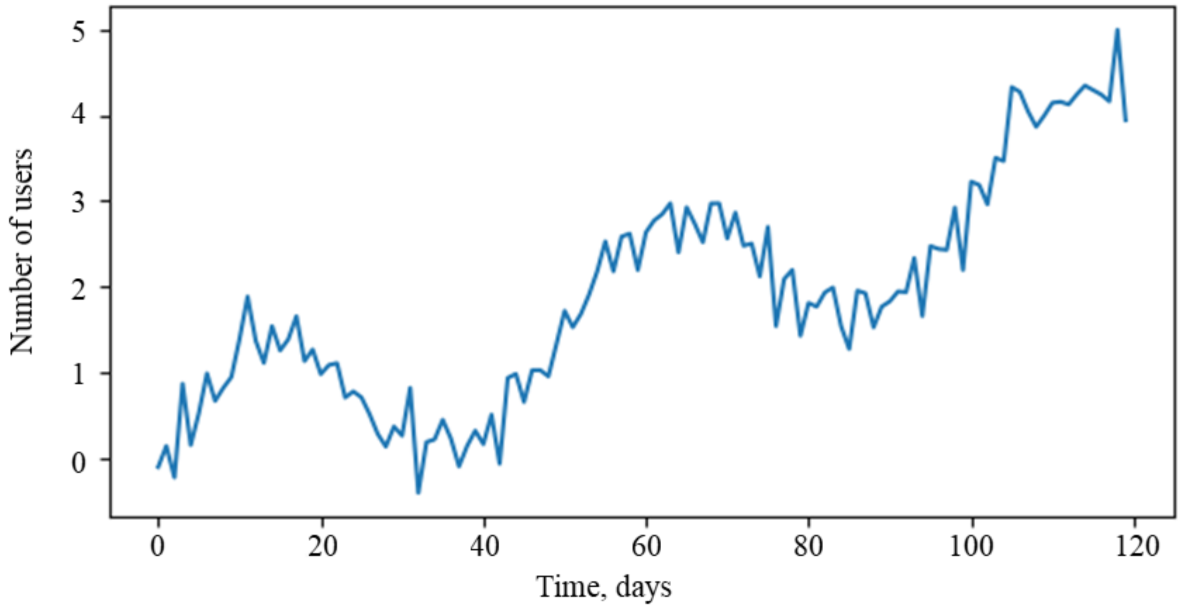

Fig. 1. Example of a time series with a trend and quasi-periodic fluctuations

Stationarity and model identification. Most classical forecasting methods require time series stationarity, which is formalised by constant values of mathematical expectation and autocovariance:

$$E(Y_t) = \mu\ , \tag{3}$$

$$Cov(Y_t, Y_{t-k}) = \gamma_k\ . \tag{4}$$

where Y_t – the value of the derivative time series at time t, reflecting a quantitative indicator of user activity or response (e.g., number of mentions, average message tone level, or frequency of incidents per time interval);

μ – mathematical expectation (mean value) of a stationary time series;

$Cov(Y_t, Y_{t-k})$ – covariance between the current value of a series and its lagged value

Y_{t-k} – the value of the derivative time series at time $t-k$, i.e. the value distant from the current one by k time intervals (lag k);

γ_k – autocovariance of the time series for lag k, which depends only on the lag value and not on the time moment t;

k – time lag, which determines the number of intervals between observations.

In the case of non-stationarity, differentiation is applied, which in AutoRegressive Integrated Moving Average (ARIMA) models corresponds to the integration parameter d. At this stage, it is advisable to analyse the autocorrelation and partial autocorrelation functions. The autocorrelation function (5) allows us to estimate the degree of linear relationship between the current value of the time series and its previous values with a lag k:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{Cov(Y_t, Y_{t-k})}{Var(Y_t)} \tag{5}$$

where ρ_k – autocorrelation coefficient for lag k, characterising the degree of linear relationship between the current value of the time series and its previous values;

k – lag (number of time intervals shifted backwards);

γ_k – autocovariance of the time series for lag k;

γ_0 – dispersion of the time series corresponding to autocovariance at zero lag;

$Var(Y_t)$ – dispersion of time series values;

Y_{t-k} – value of a time series with a shift of k time intervals.

Fig. 2 shows an example of the Autocorrelation Function (ACF) for a stationary time series. The ACF graph is characterised by a gradual monotonic decrease in autocorrelation coefficients with increasing lag, indicating a stable time dependence between successive observations. This form of the autocorrelation function is typical for series with an autoregressive structure and indicates the advisability of including the Augmented Reality (AR) component in the forecasting model. Analysis of the autocorrelation decay rate allows for a preliminary assessment of the complexity of the model and is used at the stage of identifying the parameters p (number of previous values in the series) and q (number of previous random errors) within the Box–Jenkins methodology [6].

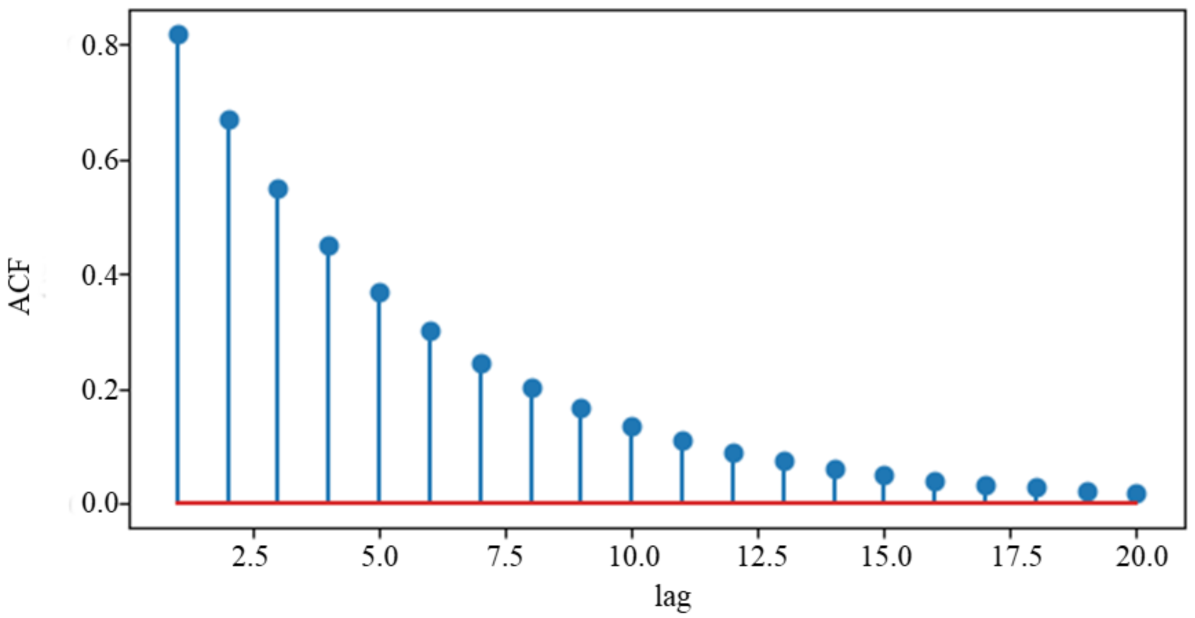

Fig. 2. Example of the autocorrelation function of the digital platform user activity indicator

The Partial Autocorrelation Function (PACF) reflects the direct relationship between Y_t and Y_{t-k} after eliminating intermediate influences and is used to determine the order of the autoregressive component p. Figure 3 shows an example of PACF for a stationary time series. It is noteworthy that for the first three lags k = 1, 2, 3 (which correspond to values of 2.5, 5.0 and 7.5 on the lag axis), the PACF values are 0.8, 0.5 and 0.2, respectively. Further, the PACF coefficient decreases sharply and approaches zero. The absence of significant partial autocorrelations for larger lags indicates that the direct influence of previous values of the time series is limited to the last few observations.

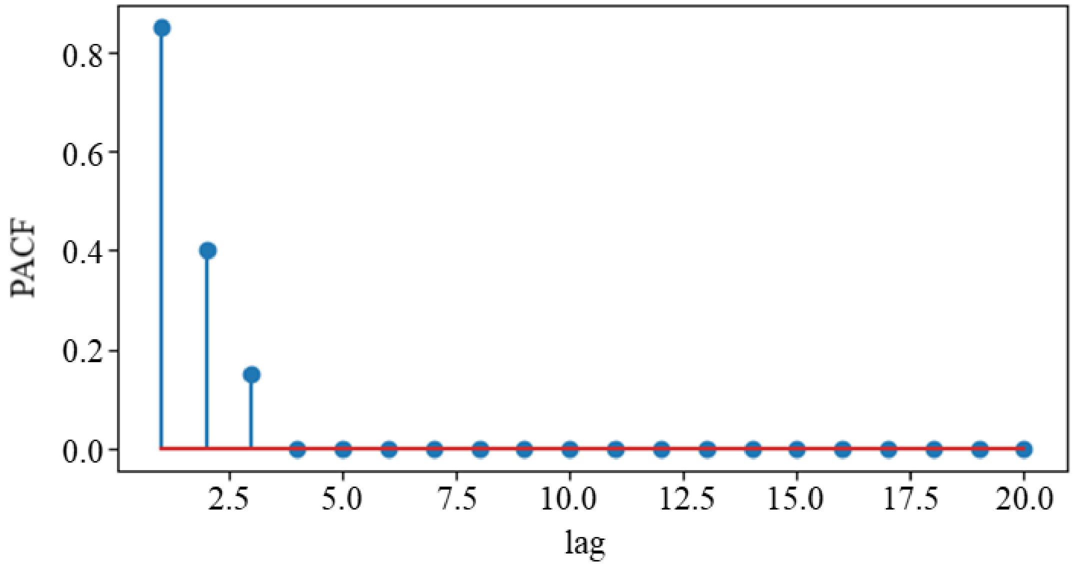

Fig. 3. Example of a partial autocorrelation function of the stakeholder response indicator over time

Formation of the baseline and detection of anomalies. Based on the identified parameters of the ARIMA model, namely the orders of the autoregressive component p, integration d and moving average q, determined based on the results of the analysis of autocorrelation (ACF) and partial autocorrelation (PACF) functions, an ARIMA model is constructed, which is used to form the baseline level of activity. Fig. 4 shows an example of baseline formation and reflects the expected dynamics of the indicator, taking into account the trend and seasonal components. Significant discrepancies between actual values and the forecast baseline are interpreted as abnormal events. Anomalies manifest themselves in the form of sharp peaks and troughs in activity and may be associated with crisis information events, increased reputational risks or changes in stakeholder expectations.

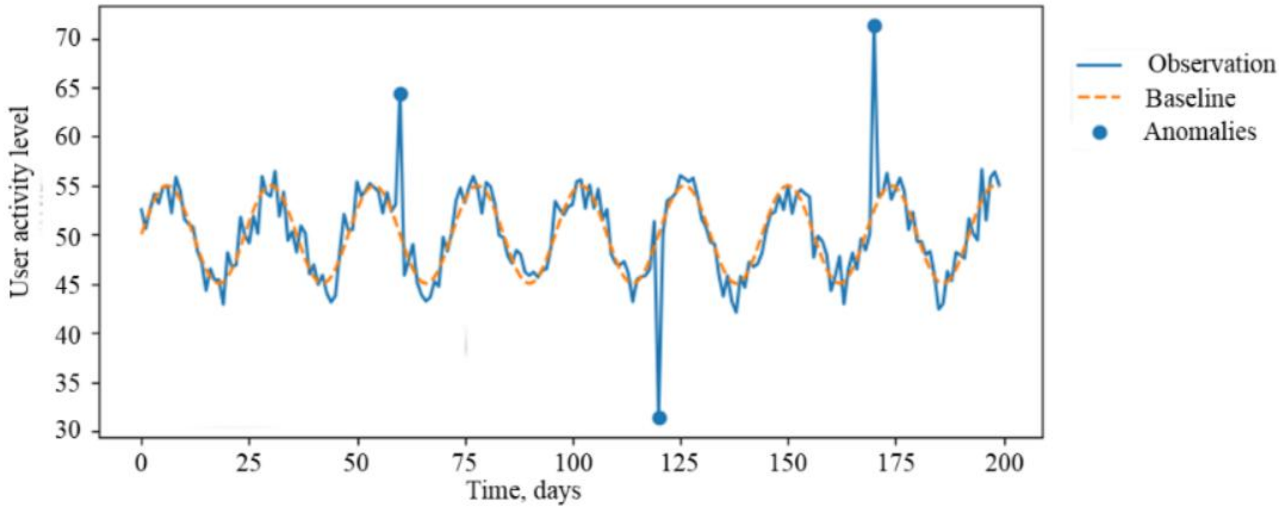

Fig. 4. Example of baseline activity according to the ARIMA model and deviations of actual values of the social media time series

Exponential smoothing methods. Simple exponential smoothing is used to forecast stationary time series without a pronounced trend or seasonality and is described by the relationship (6):

$$\widehat{Y_{t+1}} = \alpha\, Y_t + (1 - \alpha)\, \widehat{Y}_t \tag{6}$$

where Y_t – actual value of a time series at a point in time t;

$\widehat{Y}_t$ – smoothed (forecast) value of the indicator at a given point in time t;

$\widehat{Y_{t+1}}$ – the predicted value of a time series at the next point in time;

$\alpha \in (0;1)$ – smoothing parameter that determines the relative weight of the current observation compared to the previous forecast.

For time series containing a trend or seasonal component, extended Holt and Holt–Winters exponential smoothing models are used, which allow for changes in level and periodic fluctuations in indicators to be taken into account.

The autoregression model with integrated moving average ARIMA after differentiation looks like this (7):

$$\varphi(B)(1 - B)^{d} Y_t = \theta(B)\, \varepsilon_t \tag{7}$$

where Y_t – time series value at a point in time t;

B – lag operator for which the equality holds $BY_t = Y_{t-1}$;

d – integration order, which determines the number of differentiation operations required to achieve stationarity;

$\varphi(B)$ – polynomial of the autoregressive part of the model;

$\theta(B)$ – polynomial of the moving average;

ε_t – random error modelled by white noise with zero mathematical expectation and constant variance.

For time series with a pronounced seasonal structure, a generalized SARIMA model is used, which additionally takes into account seasonal autoregressive and moving components.

Deep learning methods. The basic recurrent neural network is described by the hidden state update equation (8):

$$h_t = f(W_{hh}\, h_{t-1} + W_{xh}\, x_t + b_h) \tag{8}$$

where h_t– hidden state of a neural network at a point in time t;

h_{t-1} – hidden state at the previous time step;

x_t – input feature vector at a point in time t;

W_h – weight matrix describing the recurrent connection between hidden states;

W_{xh} – weight matrix between the input vector and the hidden state;

b_h – displacement vector;

f – non-linear activation function.

Thanks to gate mechanisms, LSTM and GRU modifications are capable of taking long-term dependencies into account, filtering noise, and detecting hidden risk patterns, which makes them effective tools for early warning systems.

Conclusions. As a result of the study, the urgent scientific and practical problem of substantiating methods for analysing time series for early detection of IT project risks has been solved. The strategic role of time series analysis in modern management methodologies (PMBOK, Agile, Hybrid) has been proven. It has been established that the transformation of unstructured data from digital platforms into formalised time series allows a shift from the observation of retrospective facts to proactive monitoring. This ensures the objectification of deviation control through the formation of a mathematically justified baseline for project activity.

The diagnostic potential of ARIMA/SARIMA forecasting models and exponential smoothing methods has been determined. It has been shown that the analysis of autocorrelation (ACF) and partial autocorrelation (PACF) functions is a reliable tool for identifying the internal structure of stakeholder response processes. It has been found that the residual component of these models (E_t) acts as an indicator of abnormal events, allowing management to identify emerging risks at the stage of weak signals.

The feasibility of using deep learning architectures (LSTM, GRU) for processing high-noise data from digital ecosystems has been substantiated. It has been established that, thanks to gate mechanisms, these models effectively eliminate the problem of 'vanishing gradients' and reveal nonlinear patterns in social media dynamics that are not captured by classical statistical methods. The combination of ARIMA interpretability with LSTM approximation capabilities allows the creation of hybrid models with high predictive accuracy.

Prospects for further research lie in the development of algorithms for the automated adaptation of hybrid model parameters to the specifics of particular IT cases. A separate vector of development is the integration of the developed forecasting mechanisms into corporate decision support systems (DSS), which will automate the process of early warning of critical risks in the context of global digitalisation.

References

1. Yasinetskyi, O., Fesenko, T. (2025). IT project risk management: an analytical review of research. Bulletin of NTU "KhPI". Series: Strategic management, portfolio, program and project management, 1(10), 98–109. https://doi.org/10.20998/2413-3000.2025.10.6.
2. Fesenko, T., Fesenko, G. (2017). Developing gender maturity models to project and programme management. Eastern-European Journal of Interiorise Technologies, 1/3(85), 46–55. https://doi.org/10.15587/1729-4061.2017.28031.
3. Fesenko, T., Shakhov, A., Fesenko, G. (2017). Modeling of maturity of gender-oriented project management office. Eastern-European Journal of Interiorise Technologies, 5/3(89), 30–38. https://doi.org/10.15587/1729-4061.2017.110286.
4. Boiarchuk, S., & Tyshchenko, I. (2025). Models for Time Series Forecasting Using ARIMA and LSTM in Economics and Finance. Computer & Data Science, 7(1), 172–180. https://doi.org/10.23939/cds2025.01.172
5. Koçak, H. (2024). Time Series Prediction of Temperature Using Seasonal ARIMA and LSTM Models. Gazi Journal of Engineering Sciences, 9(3), 574–584. https://dergipark.org.tr/en/pub/gmbd/issue/82372/1299006
6. Iqbal, A. (2024). Time series forecasting and anomaly detection using deep learning. Computers & Chemical Engineering, 182, 108560. https://doi.org/10.1016/j.compchemeng.2023.108560
7. Hyndman R. J., Athanasopoulos G. Forecasting: Principles and Practice / R. J. Hyndman, G. Athanasopoulos. – 3rd ed. – Melbourne : OTexts, 2021. – URL: https://otexts.com/fpp3/
8. Box, George; Jenkins, Gwilym (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
9. "Deep learning for time series anomaly detection: a survey." (2023). ACM Computing Surveys. https://doi.org/10.1145/3691338
10. Ahmad, S., Lavin, A., Purdy, S., & Agha, Z. (2017). Unsupervised real-time anomaly detection for streaming data. Neurocomputing, 262, 134–147. https://doi.org/10.1016/j.neucom.2017.04.070