 ISSN Print 2786-5150
ISSN Online 2786-5169

ECONOMIC AND MATHEMATICAL MODELING AND INFORMATION TECHNOLOGIES IN ECONOMICS

UDC 330.45:519.237.8]:339.133.6]:[629.33:621.313] DOI https://doi.org/10.26661/2414-0287-2025-2-66-03

CONCEPTUAL MODEL FOR FORECASTING THE DYNAMICS OF ELECTRIC VEHICLE SALES VOLUMES ON THE GLOBAL MARKET

Kurkula S.G., Maksyshko N.K.
Zaporizhzhia National University
Ukraine, 69011, Zaporizhzhia, Universytetska str., 66
sergeysergey0093@gmail.com, maxishko@ukr.net
ORCID: 0000-0003-0717-0291, 0000-0002-0473-7195

Key words:
sales forecasting, time series, causal models, non-causal models, neural networks, clustering.

The modern electric vehicle market is characterized by rapid development and significant fluctuations driven by technological, economic, and political factors. Forecasting sales dynamics is critically important for the strategic planning of manufacturers, investors, and government institutions. However, existing forecasting methods often lack accuracy due to data heterogeneity, insufficient consideration of market-specific features, and influencing factors.

This paper proposes a conceptual model for forecasting electric vehicle sales dynamics in the global market. The model incorporates market clustering, time series analysis, and both causal and non-causal forecasting models. The model includes the following stages: data collection, clustering of markets based on dynamic characteristics, construction of causal (regression model using an MLP neural network) and non-causal (ARIMA, RNN, hybrid model) forecasting models, as well as evaluation of forecast quality and selection of the most relevant model.

The scientific novelty of the research lies in a comprehensive approach that combines modern forecasting methods, enabling improved prediction outcomes in an emergent economy. The proposed model demonstrates potential due to its use of hybrid methods and the integration of neural networks alongside traditional statistical approaches.

The research results may be valuable for manufacturers, investors, and government bodies in planning infrastructure projects, developing support policies, and assessing market trends. The model also opens up opportunities for further research in the field of forecasting dynamic markets.

КОНЦЕПТУАЛЬНА МОДЕЛЬ ПРОГНОЗУВАННЯ ДИНАМІКИ ОБСЯГІВ ПРОДАЖІВ ЕЛЕКТРОМОБІЛІВ НА СВІТОВОМУ РИНКУ

Куркула С.Г., Максишко Н.К.
Запорізький національний університет
Україна, 69011, м. Запоріжжя, вул. Університетська, 66

Ключові слова:
прогнозування продажів, часові ряди, каузальні моделі, некаузальні моделі, нейронні мережі, кластеризація.

Сучасний ринок електромобілів характеризується стрімким розвитком та значними коливаннями, зумовленими технологічними, економічними та політичними факторами. Прогнозування динаміки продажів є критично важливим для стратегічного планування виробників, інвесторів та державних установ. Однак існуючі методи прогнозування часто мають обмежену точність через неоднорідність даних, недостатнє врахування специфіки окремих ринків та факторів що на них впливають.

У цій статті запропоновано концептуальну модель прогнозування динаміки обсягів продажів електромобілів на світовому ринку, яка включає кластеризацію ринків, аналіз часових рядів, каузальні та некаузальні моделі прогнозування. Модель включає такі етапи: збір даних, кластеризацію ринків за характером динаміки, побудову каузальних (регресійна модель за допомогою нейронної мережі MLP) та некаузальних (ARIMA, RNN, гібридна модель) моделей, а також оцінку якості отриманих прогнозів та вибір найбільш релевантної моделі.

ISSN Print 2786-5150
ISSN Online 2786-5169

Наукова новизна дослідження полягає в комплексному підході, який поєдную сучасні методи прогнозування, що дозволить покращити результати прогнозів в умовах емерджентної економіки. Запропонована модель має потенціал завдяки використанню гібридних методів та використанню нейронних мереж наряду із традиційними статистичними підходами.
Результати дослідження можуть бути корисними для виробників, інвесторів та державних органів при плануванні інфраструктурних проектів, розробці політик підтримки та оцінці ринкових тенденцій. Модель також відкриває перспективи для подальших досліджень у галузі прогнозування динамічних ринків.

Statement of the problem

The modern world is undergoing an active transition to environmentally friendly technologies, among which electric vehicles hold a special place. The growing popularity of electric vehicles is driven by their environmental benefits, cost-effectiveness in operation, and support from governments across various countries. However, the global electric vehicle market is marked by significant differences in sales volume dynamics between countries. These differences depend on a range of factors such as government policy, the level of economic development, electricity and fuel prices, and the deployment of charging infrastructure.

Forecasting the sales dynamics of electric vehicles in the global market is an important tool for strategic planning, the development of effective public policies, and business decision-making. However, the accuracy of forecasts is often limited due to data heterogeneity, the unpredictability of changes in regulatory mechanisms, and the rapid pace of technological advancement. Therefore, there is a need to create adaptive models that consider both global and local factors influencing the market.

Currently, a significant portion of existing forecasting research focuses either on causal models, which account for specific influencing factors, or on non-causal models, which analyze time series. However, these methods are typically used separately and little attention is paid to integrating these approaches into a comprehensive forecasting framework to improve prediction accuracy. Additionally, the issue of clustering markets based on their dynamic characteristics is underexplored in the literature, despite its potential to significantly enhance forecast quality for each specific market group.

Thus, the development of a forecasting model for electric vehicle sales dynamics in the global market – one that combines time series analysis with factor analysis and takes into account the specific characteristics of different national markets – is a relevant and pressing task.

Analysis of recent studies and publications

Forecasting sales dynamics is one of the key tasks in the field of economic analysis and management. In modern scientific literature, various forecasting methods are widely covered, including time series analysis, causal models, machine learning, and combined approaches.

Among classical approaches for forecasting non-stationary time series, ARIMA models are widely used and thoroughly described in [1]. These models allow for predictions based on historical data; however, their limitations include the requirement of stationarity and dependence on a fixed model structure. Furthermore, studies [2, 3] explore extensions of classical models through the use of seasonal integration and adaptive filters.

Causal models focus on identifying and quantifying the factors that influence the variables being forecasted. Research [4] highlights the importance of analyzing correlations between economic, social, and infrastructural indicators for predicting sales volumes. In particular, factors such as government subsidies, fuel prices, and the availability of charging infrastructure play a decisive role in determining the dynamics of electric vehicle sales. Studies [5, 6] demonstrate that regression models can be effective in accounting for these factors, though they are limited by their assumption of linear relationships.

Recent research has devoted considerable attention to the use of machine learning methods for forecasting sales dynamics. Neural networks – especially recurrent neural networks (RNN) and multilayer perceptrons (MLP) – have proven to be effective tools for modeling nonlinear dependencies and handling large datasets. Studies [7, 8, 9] show that hybrid models combining neural networks with traditional time series analysis methods deliver higher forecasting accuracy compared to either approach used independently.

Another promising direction is the clustering of markets based on the similarity of their development dynamics. In works [10, 11], clustering algorithms are proposed that group economic agents in electric vehicle markets to build specialized models and analyze the resulting information. This enhances forecast accuracy and enables model adaptation to local conditions.

Despite significant progress in the field, several unresolved issues remain. In particular, there is insufficient research on the integration of market clustering methods with both causal and non-causal forecasting models. Additionally, further study is needed on the adaptation of hybrid models to the rapidly changing environment characteristic of the electric vehicle market.

Therefore, there is a need to develop a model that incorporates market clustering based on sales dynamics, factor analysis, and modern modeling methods to improve the accuracy of forecasts for electric vehicle sales dynamics.

Objectives of the article

The aim of this study is to develop a conceptual model for forecasting the dynamics of electric vehicle sales in the global market, which will enhance forecast accuracy through a multi-stage process structure. This structure integrates the following components: market clustering based on dynamic characteristics, time series analysis, factor analysis, and the application of various forecasting methods.

The model is designed to serve as a universal forecasting tool that enables adaptation to the specific features of different national markets. It aims to ensure effective use in both the private and public sectors for planning purposes in the context of the rapidly evolving electric vehicle market.

The main material of the research

To forecast the dynamics of electric vehicle sales in the global market, a conceptual model will be developed based on the following key stages:

a) Market clustering – to account for the specific characteristics of sales volume dynamics in different electric vehicle markets;

b) Development and use of causal models – to identify the impact of defined factors on sales dynamics;

c) Development and use of non-causal models – to capture the nature and characteristics of time series, enabling faster results based on historical data without delving into cause-and-effect relationships, particularly for datasets with strong temporal patterns;

d) Forecast generation using the developed models;

e) Comparative analysis of forecast accuracy and determination of the most effective forecasting method for different market clusters.

The structure of the model is illustrated in the diagram (Fig. 1). Each stage of the model plays a key role in the forecasting process.

Let us take a closer look at the main stages of forecast construction using the proposed model:

Stage 1. Data collection on electric vehicle sales volumes

At this initial stage, data is collected on the dynamics of electric vehicle sales across various countries. It is advisable to consider sources such as: official reports from electric vehicle manufacturers, statistical data from government agencies, and information from open databases such as EV Volumes [12], IEA [13], among others.

Additional variables should also be included, such as data on charging infrastructure development, subsidy levels, demographic and economic indicators, as well as electricity prices.

Stage 2. Market clustering based on sales dynamics

Once the data is collected, markets are clustered based on their sales dynamics. It is recommended to use the Pearson correlation metric for this purpose, as it allows grouping countries into clusters that show the highest correlation in their sales dynamics. This metric is calculated using the following formula [14]:

$$K(x, y) = \frac{\sum_{i=1}^{N} x_i y_i}{\sqrt{\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i^2}} \quad (1)$$

Stage 3. Modeling the Dynamics of Sales Volumes

The application of causal (cause-and-effect) and non-causal (statistical) models in forecasting the dynamics of sales volumes is appropriate, as each of these approaches has its own advantages and limitations. Using both approaches in parallel makes it possible to compensate for the shortcomings of individual methods and to improve forecasting accuracy.

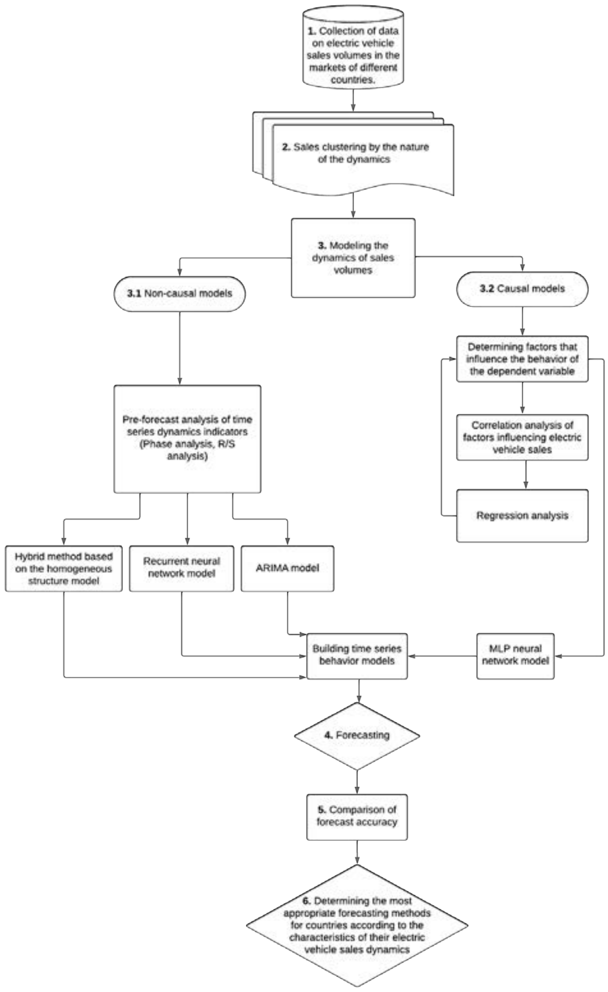

Fig. 1 – Scheme of the conceptual model of forecast construction

Non-causal models, such as ARIMA, recurrent neural network (RNN) models, or hybrid methods, are effective in detecting time dependencies, trends, and seasonality. However, they do not take into account external influencing factors. Conversely, causal models, such as regression analysis or multifactor neural network models, allow for the assessment of the impact of marketing, economic, or social factors, but may underestimate complex nonlinear relationships hidden in historical data.

The combined use of both approaches provides more stable and interpretable forecasts. This enables the generation of the most accurate results, especially under conditions of structural shifts or market instability.

3.1. Non-Causal Models

At this stage, the dynamic characteristics of time series are analyzed, followed by the forecasting of future values of sales volumes.

The following approaches are proposed for forecasting: ARIMA models, recurrent neural networks (RNN), and a hybrid method based on a homogeneous structure.

The construction of the proposed non-causal models requires pre-forecast analysis to identify the characteristics of the time series. Various time series analysis methods are used to detect nonlinear (chaotic) behavior in economic data [15].

In our opinion, for the proposed forecasting methods, it is advisable to use the following tools for studying nonlinear dynamics, namely: traditional R/S analysis (Rescaled Range Analysis) – Hurst exponent method, phase analysis, and recurrence analysis.

To obtain an overall assessment of the fractal properties of the time series (TS), we will apply the Hurst rescaled range algorithm [16].

It is known [16] that if a system produces a Hurst statistic *H* over a sufficiently long period, this indicates the result of interdependent events. The measure of such interdependence, as is well known, is the correlation coefficient. The influence of the present on the future can be expressed through the following correlation relationship:

$$C = 2^{2H-1} - 1, \quad (2)$$

where *C* is the correlation measure, *H* is the Hurst exponent.

The range of values of the Hurst exponent *H* is the interval [0;1]. The value of the exponent *H* allows us to divide (classify) all TS into three groups: 1) $H = 0,5$, 2) $0 \le H < 0,5$ and 3) $0,5 < H \le 1$.

Value $H = 0,5$ indicates a random TS: events are random and uncorrelated (in accordance $C = 0$). The present does not affect the future.

If $H \in (0,5;1]$, then the considered TS is persistent, or trend-resistant and is characterized by the effect of long-term memory. Events are more correlated, the closer the value of *H* is to unity (respectively, *C* also approaches unity or 100% correlation according to (2)).

Values of *H* within the interval $H \in [0;0,5)$ correspond to antipersistent time series. In a loose definition, antipersistence means a tendency to revert to the mean or, in other terms, frequent reversal (alternation of positive and negative increments) more often than in a random process. Thus, the Hurst exponent *H* is a key indicator for diagnosing the nature of the development of a system or process.

To test the validity of the results regarding the presence of long-term memory based on the value of the Hurst *H* index, we propose to use the test for random mixing of the levels of the TS.

Phase analysis is one of the effective methods for obtaining information about the nature of the dynamics of the system under consideration [17]. For a time series $X = \left(x(t), t = \overline{1,n}\right)$, this representation method is used to return from the observed state of the system to its previous state. This "return" is implemented by the method of time delays and is carried out by constructing a phase trajectory (phase portrait) of dimension ρ:

$$\Phi_\rho(X) = \left\{\left(x(t), x(t+1), ..., x(t+\rho-1)\right)\right\}, \quad t = \overline{1,n}, \quad (3)$$

which represents a set of points called "M-history" ($\rho \equiv M$). For any time series (TS), the set of all its *M*-histories defines the corresponding set of points in the pseudo-phase (or lag) space. In this case, when using the terms "phase portrait" or "phase trajectory," it is implied that neighboring points of set (2) are visually connected by straight or curved line segments for clarity.

A graphical representation of a system on the phase plane (or in phase space), where the coordinate axes correspond to the values of the system's variables (levels of the time series), is called the phase portrait of the system. The behavior of phase points over time, described by the phase trajectory, and the collection of such phase trajectories for any initial conditions form the phase portrait. The phase portrait is a mathematical method for representing the system's behavior and a geometric depiction of individual motions. It also reflects equilibrium states, periodic and chaotic motion of the phase point, the logic of the system's behavior, and its dependence on external and internal influences.

Objective information about the nature of the behavior of a dynamic process can be obtained through the observation of the time series *X*, based on Takens' theorem [18]: if the system generating the time series is *m*-dimensional and the inequality $\rho \ge 2m + 1$, then, in the general case, the phase trajectories will reproduce the dynamics of the studied system. There exists a diffeomorphism between the phase trajectories and the true data generated by the system. This result allows making conclusions about the behavior of the system based on observational data and, moreover, obtaining information for forecasting this behavior.

Analysis of the phase portrait allows us to determine the type and characteristic features of the dynamics of a particular system.

To determine the embedding dimension of a time series, the false nearest neighbor method is applied, as described in [19]. This method is based on the assumption that, with successive iterations, neighboring points of the phase trajectory remain sufficiently close. However, if the nearest points move away from each other, they are considered false nearest neighbors. The goal of the method is to select such an embedding dimension ρ for the time series at which the proportion of points with false neighbors is minimized.

Based on the calculated embedding dimension and lag parameters, recurrence plots of the time series are constructed.

The analysis of statistical characteristics of the recurrence plot allows determining measures of the complexity of the recurrence structure [20], namely:

recurrence rate (%REC),

determinism measure (DET),

average (ADL) and maximum (MDL) diagonal line lengths of the recurrence plot.

Based on the analysis of these statistical characteristics, it is possible to identify the presence of homogeneous processes with independent random values; processes with slowly varying parameters; periodic or oscillatory processes corresponding to nonlinear systems. Thus, the analysis of the recurrence surface enables the assessment of the characteristics of a nonlinear object using relatively short time series, facilitating prompt decision-making in the management of the object.

3.2 Causal Models. Preparation for forecasting using a causal model involves several key stages aimed at identifying and assessing factors that influence the dependent variable, as well as building a mathematical model of the relationships. The main stages of this process include determining the factors affecting the behavior of the dependent variable, correlation analysis, and regression analysis of these factors.

At the first stage, factors that may affect sales volumes are identified. Both theoretical approaches (economic models, expert assessments) and empirical methods (analysis of historical data, market trends) are used. Factors can be internal (prices, product assortment) or external (consumer income levels, social factors, competition, macroeconomic conditions). It is important at this stage to formulate hypotheses about the influence of each factor on the dependent variable and to determine potential interrelationships.

After identifying the factors, correlation analysis is conducted to evaluate the strength and direction of the relationship between each factor and the dependent variable. Correlation analysis is performed using either the Pearson correlation coefficient [21] or Spearman's rank correlation coefficient [22], depending on the nature of the data.

Speaking of the Pearson correlation, let us explain how it is calculated. Let *a* and *b* be two real random variables; then the Pearson correlation coefficient (PCC) is defined as [21]:

$$\rho(a,b) = \frac{E(a,\ b)}{\sigma_a \sigma_b} \tag{4}$$

where E(a, b) is the cross-correlation (covariance) between a and b, and $\sigma_a = E(a^2)$ i $\sigma_b = E(b^2)$ dispersions of signals *a* and *b*, respectively. Ultimately, it will be more convenient to work with the squared Pearson correlation coefficient (SPCC):

$$\rho^2(a,b) = \frac{E^2(a,\ b)}{\sigma_a^2 \sigma_b^2} \tag{5}$$

One of the most important properties of SPCC is that

$$0 \le \rho^2(a,b) \le 1 \tag{6}$$

SPCC indicates the strength of the linear relationship between two random variables *a* and *b*. If $p^2\ (a,\ b) = 0$, then *a* and *b* are uncorrelated. The closer the value of $p^2\ (a,\ b)$ is to *1*, the stronger the correlation between the two variables. If the two variables are independent, then $p^2\ (a,\ b) = 0$. However, the reverse is not true, since the PCC only detects linear dependencies between the two variables *a* and *b*. For nonlinear dependencies, the PCC may be zero.

The general classification and interpretation of correlations is as follows[24]:

– strong, or dense, with a correlation coefficient of $\rho > 0{,}70$;

– average at $0{,}50 < \rho < 0{,}69$;

– moderate at $0{,}30 < \rho < 0{,}49$;

– weak at $0{,}20 < \rho < 0{,}29$;

– very weak at $\rho < 0{,}19$.

That is, factors with moderate correlation and below clearly cannot be used for further model construction.

We know that simple linear correlation describes the relationship between two variables or phenomena; that is, when one of the variables changes, it causes a change in the other, whether an increase or a decrease. When both variables increase or decrease together, their relationship is positive. Conversely, when one variable decreases while the other increases, their relationship is negative.

A simplified definition of the Spearman rank correlation coefficient can be given as follows: it is a coefficient that expresses the strength and direction of the relationship between two phenomena. This relationship can be either positive or negative, and weak or strong. Spearman's correlation is used for rank correlation, where variable *A* has rank R_A and variable *B* has rank R_B. Assuming *d* represents the difference between the two ranks, i.e. $d = R_A - R_B$, then Spearman's correlation coefficient ρ is calculated by the following formula:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{7}$$

where *n* is the number of ordered pairs of observations.

The next step is to conduct regression analysis [23] to quantitatively assess the impact of independent variables on the dependent variable. Most often, the method of multiple linear regression is used, which allows estimating the contribution of each factor to the change in the predicted value. To build the regression model, regression coefficients are determined, which show how much the dependent variable changes with a change in each independent factor. Additionally, the statistical significance of the model is evaluated (F-test, coefficient of determination R^2) as well as the significance of individual factors (t-test, p-value).

For assessing the goodness of fit in multiple linear regression, the coefficient of determination or R^2 is a very straightforward tool and is most commonly used in practice. Although it is not recommended as the final tool for model selection, it provides an indication of how well the chosen explanatory variables predict the response [25]. However, besides statistical significance, it is necessary to confirm how the selected independent variable affects the dependent variable.

In the context of classical multiple linear regression, the coefficient takes values between 0 and 1. It is generally accepted that the closer the coefficient is to 1, the better the model. The coefficient of determination increases with the inclusion of predictors (independent variables) in the model. However, this does not necessarily mean that a model with more predictors is better than a model with fewer predictors. Therefore, the coefficient of determination should be used only as one of the metrics for evaluating the validity of the model.

The coefficient of determination is defined as follows [26]:

$$R^2 = 1 - \frac{V(y|x)}{V(y)} = 1 - \frac{\sigma^2}{\sigma_y^2}\ , \tag{8}$$

where $V(y) = \sigma_y^2$ – variance of a random variable *y*, $V(y|x) = \sigma^2$ – conditional variance of the dependent variable (variance of the model error).

To calculate the sample coefficient of determination, sample estimates of the values of the corresponding variances are used:

$$R^2 = 1 - \frac{\hat{\sigma}^2}{\hat{\sigma}_y^2} = 1 - \frac{RSS/n}{TSS/n} = 1 - \frac{RSS}{TSS}, \quad (9)$$

where $\sum_{t=1}^{n} e_t^2 = \sum_{t=1}^{n}(y_t - \hat{y}_t)^2$

– sum of squares of regression residuals, y_t, $\hat{y}_t$ – actual and estimated values of the explanatory (dependent) variable.

$TSS = \sum_{t=1}^{n}(y_t - \bar{y})^2 = n\hat{\sigma}_y^2$ – total sum of squares.

$\bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t$ – average value of observed (actual) data.

In the case of classical linear multiple regression (regression with a constant):

$TSS = RSS + ESS$, де $ESS = \sum_{t=1}^{n}(\hat{y}_t - \bar{y})^2$ – sum of squares explained. And as a result:

$$R^2 = \frac{ESS}{TSS} \quad (10)$$

As a result of data analysis and preparation, a causal model is formed that explains the influence of key factors on sales volumes and can be used to forecast future values of the dependent variable under certain conditions.

The construction of a causal model using regression based on neural networks can be performed using ready-made software products that have proven effective in similar studies [27, 28, 29, 30]. One example of such a product is Statistica 13. The multilayer perceptrons (MLP) used in this software are among the most popular neural network architectures today.

In this architecture, each neuron computes a weighted sum of its input data and passes it through a transfer function *f* to produce the output. For each neural layer in the MLP, there is also a bias term. The bias is a neuron whose activation function is constantly set to 1. Like other neurons, the bias connects to neurons in the upper layer through a weight often called the threshold. Neurons and biases are arranged in a layered feedforward topology. Thus, the network can be simply interpreted as an input-output model with weights and thresholds as free (adjustable) model parameters. Such networks can model functions of almost arbitrary complexity, with the number of layers and units in each layer defining the complexity of the function [31].

A schematic diagram of the MLP neural network [32] is shown in Fig. 2.

In this architecture, each neuron computes a weighted sum of its input data and passes it through a transfer function to generate the output. This method allows building a model that describes target (dependent) variables based on independent ones. In other words, when constructing this model, the interdependence of variables is taken into account, revealing the causal nature of the model. Such networks enable the creation of process (dynamic) models of almost arbitrary complexity, with the number of layers and units in each layer determined by the complexity of these processes [31].

Thus, the simultaneous application of causal and non-causal models is a promising approach in sales forecasting, as the final result benefits from the strengths of each model in its respective use case. This is especially relevant in a dynamic business environment, where underestimating external or internal factors can lead to significant forecasting errors.

Stage 4. Forecast Construction

At the forecasting stage, predictions are generated according to each model. However, the common steps in the forecasting process are as follows:

- splitting the data into training (80-90%) and testing (10-20%) sets;
- determining the model parameters;
- training the model (parameter optimization);
- forecasting future values;
- determining confidence intervals;
- comparing the forecast with actual data.

Stage 5. Comparison of Forecast Quality Results

At this stage, the accuracy of the obtained forecasts is evaluated, for example, using MAPE (Mean Absolute Percentage Error). Models are compared based on their performance within each cluster. The Mean Absolute Percentage Error (MAPE) is a measure of forecasting accuracy in statistics. It typically expresses accuracy as a ratio defined by the formula:

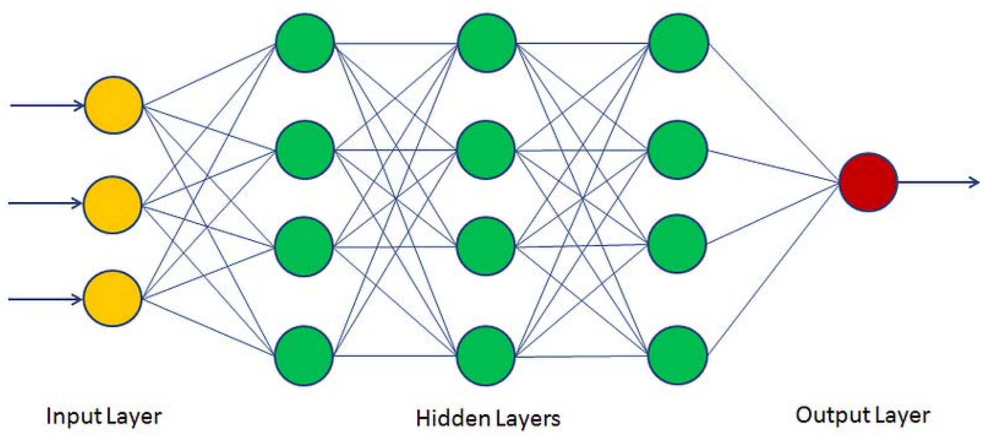

Fig. 2 – Diagram of an MLP neural network [32]

ISSN Print 2786-5150
ISSN Online 2786-5169

$$MAPE = 100\frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|, \qquad (11)$$

where A_t is the actual value, and F_t is the forecasted value. Their difference is divided by the actual value A_t. The absolute value of this ratio is summed over each forecasted time point and divided by the number of observations *n*.

Stage 6. Selection of the Best Forecasting Method Based on the evaluation results of forecast accuracy, the forecasting method that demonstrates the highest accuracy is selected. This method is considered the best fit for the specific market characteristics of the given cluster.

Conclusions

In the course of this research, a conceptual model for forecasting the dynamics of electric vehicle sales volumes in the global market was developed. The model incorporates market clustering, time series analysis, and both causal and non-causal forecasting approaches. The proposed model allows for consideration of the specific characteristics of different markets, which is a key factor in ensuring high forecast accuracy in a rapidly changing environment. Thus, the scientific novelty of this study lies in the integrated approach to forecasting that combines the aforementioned tools.

The relevance of the research is driven by the rapid development of the electric vehicle market, which is a crucial segment of the global automotive industry. The growing demand for electric vehicles is accompanied by significant fluctuations caused by technological, economic, and political factors. Accurate forecasting of sales dynamics is critically important for manufacturers, investors, governments, and other stakeholders to enable effective planning and decision-making.

The proposed model also demonstrates innovation through the use of modern techniques such as neural networks and hybrid forecasting methods, allowing for a substantial improvement in accuracy compared to traditional models.

The obtained results can be applied for strategic planning at the level of investors, governmental support programs, and infrastructure projects. Furthermore, the proposed model lays the groundwork for further scientific research in forecasting under dynamic market conditions.

References

1. Beveridge S., Oickle C. A Comparison of Box–Jenkins and objective methods for determining the order of a non-seasonal ARMA Model, Journal of Forecasting, Volume 13, Issue 5, September 1994. DOI:10.1002/for.3980130502.
2. Xilin Liu, Predicting Apples future stock price using ARIMA model, Theoretical and Natural Science 26(1), December 2023. DOI:10.54254/2753-8818/26/20241066
3. Vorobets I., Fryz M. Vykorystannia modelei ARIMA dlia prohnozuvannia chasovykh riadiv iz vlastyvistiu tsyklichnosti, Materialy VI Mizhnarodnoi studentskoi naukovo-tekhnichnoi konferentsii «Pryrodnychi ta humanitarni nauky. Aktualni pytannia», Ternopil: TNTU, 2023. URL: https://elartu.tntu.edu.ua/bitstream/lib/41432/2/122-123.pdf
4. Zhang Y, Zhong M, Geng N, Jiang Y Forecasting electric vehicles sales with univariate and multivariate time series models: The case of China. PLoS ONE 12(5), 2017. DOI:10.1371/journal.pone.0176729.
5. Yu R, Wang X, Xu X, Zhang Z. Research on Forecasting Sales of Pure Electric Vehicles in China Based on the Seasonal Autoregressive Integrated Moving Average–Gray Relational Analysis–Support Vector Regression Model. Systems. 2024; 12(11). DOI:10.3390/systems12110486.
6. Lingxiao T., Jia S. Predict the sales of New-energy Vehicle using linear regression analysis. E3S Web of Conferences, 2018. DOI: 10.1051/e3sconf/201911802076.
7. Afandizadeh, S., Sharifi, D., Kalantari, N. et al. Using machine learning methods to predict electric vehicles penetration in the automotive market. Sci Rep 13, 8345 (2023). DOI:10.1038/s41598-023-35366-3
8. Wu M, Chen W. Forecast of Electric Vehicle Sales in the World and China Based on PCA-GRNN. Sustainability, 14(4), 2022. DOI:10.3390/su14042206
9. Boshuai Q., Sigal K., Jie H. Exploring the Bev Sales Forecasting Under Dynamic Market Conditions In Jiangsu Province, China. Available at SSRN: https://ssrn.com/abstract=4813790 or http://dx.doi.org/10.2139/ssrn.4813790
10. Jahangir H., Gougheri S., Vatandoust B., Golkar M., Ahmadian A., Hajizadeh A., Plug-in Electric Vehicle Behavior Modeling in Energy Market: A Novel Deep Learning-Based Approach With Clustering Technique, IEEE Transactions on Smart Grid, vol. 11, no. 6, Nov. 2020, DOI: 10.1109/TSG.2020.2998072
11. Xiong Y., Wang B., Chu C., Gadh R. Electric Vehicle Driver Clustering using Statistical Model and Machine Learning, IEEE Power & Energy Society General Meeting (PESGM), Portland, OR, USA, 2018, DOI: 10.1109/PESGM.2018.8586132.
12. The EV-volumes, URL:https://www.ev-volumes.com/
13. The International Energy Agency, URL:https://www.iea.org/
14. Apache Mahout. Metrics to determine user similarity. URL:https://habr.com/ru/articles/188350/
15. Faggini M. Chaotic time series analysis in economics: Balance and perspectives Citation: Chaos: An Interdisciplinary Journal of Nonlinear Science 24, 2014, DOI:10.1063/1.4903797
16. Peters E., Fractal Market Analysis. Applying Chaos Theory to Investment and Analysis (John Wiley & Sons, Inc., New York, 1994).
17. Perepelitsa V.V., Maksyshko N.K., Analiz i prognozirovanie evolyuczii ekonomicheskikh sistem: problemy strukturirovaniya dannykh v usloviyakh neopredelennosti i predprognoznogo analiza. (Analysis and forecasting of

ISSN Print 2786-5150
ISSN Online 2786-5169

the economic systems evolution: problems of data structuring in conditions of uncertainty and pre-forecast analysis). (Lambert Academic Publishing GmbH& Co. KG, Saarbrucken, 2012).
18. Takens F. Detecting strange attractors in turbulence. Dynamical systems and turbulence, eds. D.Rand, L.Young. Berlin: Springer, Verlag, 1981.
19. Kennel, M. B., Brown, R., H. D. I.: Abarbanel Determining embedding dimension for phase-space reconstruction using a geometrical construction. Physical Review A 45(6), 1992. DOI:10.1103/PhysRevA.45.3403
20. Wallot S. Recurrence Quantification Analysis of Processes and Products of Discourse: A Tutorial in R, Discourse Processes, 54:5-6, 2017, DOI: 10.1080/0163853X.2017.1297921
21. Benesty, J., Chen, J., Huang, Y., Cohen, I. Pearson Correlation Coefficient. In: Noise Reduction in Speech Processing. Springer Topics in Signal Processing, vol 2. Springer, Berlin, Heidelberg, 2009. DOI:10.1007/978-3-642-00296-0_5
22. Ali Abd Al-Hameed, K. Spearman's correlation coefficient in statistical analysis. International Journal of Nonlinear Analysis and Applications, 13(1), 2022. DOI:10.22075/ijnaa.2022.6079
23. Chatterjee S., Ali Hadi S. Regression analysis by example, New York University, American University in Cairo, Actuarial Science Program, Cairo, Egypt. – Fifth edition, 2006, pp. 20-21.
24. Kotsiubynskyi V. Yu., Kyslytsia L. M. Osnovy modeliuvannia rynkovykh sytuatsii, Navchalnyi posibnyk, 2013, URL: https://web.posibnyky.vntu.edu.ua/fksa/12kocubynsky,kyslycia_osn_model_rynk_sytuac/zmist.html
25. Renaud O., Victoria-Feser M. A robust coefficient of determination for regression, Journal of Statistical Planning and Inference Volume 140, Issue 7, 2010. DOI: 10.1016/j.jspi.2010.01.008
26. Di Bucchianico A. Coefficient of Determination (R2). In Encyclopedia of Statistics in Quality and Reliability (eds F. Ruggeri, R.S. Kenett and F.W. Faltin). 2008. DOI:10.1002/9780470061572.eqr173
27. Kucher P., Yunkova O. Prohnozuvannia dynamiky rynku vitaminiv za dopomohoiu neiromerezh, Nauka i tekhnika sohodni № 3(17) 2023. DOI:10.52058/2786-6025-2023-3(17)-110-121
28. Yatsenko V. V., Hrytsenko K. H., Koibichuk V. V., Shtefan A. V. Neiromerezheve modeliuvannia ta prohnozuvannia aktualizatsii kibersportyvnoi industrii na svitovomu rivni, Visnyk Khmelnytskoho natsionalnoho universytetu, №2, 2021 DOI:10.31891/2307-5732-2021-295-2-289-295
29. Zoryna V., Yurynets V., Kruhliakova V. Neiromerezheve modeliuvannia yak instrument prohnozuvannia innovatsiinoho rozvytku ekonomiky Ukrainy, Aktualni problemy ekonomiky №6(180), 2016 URL:https://financial.lnu.edu.ua/wp-content/uploads/2017/10/ape_2016_6_51.pdf
30. Herasymenko V. A., Vasylenko V. V., Maiborodyna N. V., Kovalov A. V. Neiromerezheve prohnozuvannia strumu vytoku na osnovi tekhnolohichnykh parametriv, "Enerhetyka i avtomatyka", №3, 2022. DOI:10.31548/energiya2022.03.109
31. Technical documentation site for TIBCO products, URL:https://docs.tibco.com/pub/stat/14.0.0/doc/html/UsersGuide/GUID-BF0E25C0-2D6F-4DCA-9883-6168B21D3B09.html
32. Machine Learning Geek. URL:https://machinelearninggeek.com/multi-layer-perceptron-neural-network-using-python/