ECONOMIC FORUM
Journal homepage: https://e-forum.com.ua/en
Vol. 15, No. 2, 2025, 45-55
Article's History: Received: 25.12.2024 Revised: 02.04.2025 Accepted: 24.04.2025

*UDC 339.13.017:004.67* *DOI: 10.62763/ef/2.2025.45*

# Using big data for demand forecasting and dynamic pricing

**Dmytro Zelenyi***
Master
Kyiv National Economic University named after Vadym Hetman
03057, 54/1 Beresteiskyi Ave., Kyiv, Ukraine
https://orcid.org/0009-0003-2719-6357

**Abstract.** The research involved developing and implementing a big data-driven demand forecasting and dynamic pricing system for retail businesses in Ukraine. The methodology covered gathering a massive database from online trading platforms, transactional information from points of sale, and loyalty programmes, which showed an overall aggregated data quality index of 92.6%. Applying a comprehensive set of cleaning and normalisation methods boosted data quality by 12.47%. A comparative analysis of predictive models revealed the highest effectiveness of the LSTM network among individual models ($R^2 = 0.874$, MAPE = 6.83%) and of the ensemble model among all tested approaches ($R^2 = 0.896$, MAPE = 5.92%). Implementing the developed system in various retail formats, such as ATB-Market LLC, Foxtrot LLC, Nova Liniya PJSC, ALLO LLC, Silpo-Fud LLC, METRO Cash and Carry Ukraine LLC, Epicentr K LLC, Rozetka LLC, Comfy Trade LLC, and INTERTOP Ukraine LLC, showed a significant improvement in economic efficiency, with an average revenue increase of 9.16%, a marginal profit increase of 11.08%, and a 6.95% reduction in inventory levels. The best performance was demonstrated by the online stores Rozetka and ALLO, with Return on Investment figures of 516% and a payback period of 2.7 months. Regional analysis revealed significant differences in system implementation effectiveness, with the best results in the Western region, specifically Lviv and Volyn regions (revenue increase 9.27 ± 2.05%), and among non-food retailers, particularly Comfy Trade LLC and INTERTOP Ukraine LLC (9.86 ± 2.23%). Small businesses showed the highest adaptability to dynamic pricing, with a revenue increase of 10.23 ± 2.14% and an Return on Investment of 405 ± 86%. The research confirmed the high scalability and adaptability of the proposed approach for the Ukrainian market and allowed for the development of differentiated recommendations for system implementation for various types of retail businesses, considering their size, regional location, and product specialisation

**Keywords:** datasets; predictive models; economic efficiency; retail chain; consumer behaviour

## Introduction

During 2021-2025, Ukraine actively integrated big data technologies into the retail sector. Modern enterprises needed effective tools for demand forecasting and dynamic pricing to optimise their operations in a changing market environment. The use of big data analytics opened up new opportunities for informed management decisions, which affected business competitiveness and customer service quality.

P. Arguelles Jr & Z. Pólkowski (2023) studied the impact of big data on supply chain efficiency through demand forecasting, demonstrating how massive data set analytics enabled enterprises to optimise logistics processes, reduce inventory storage costs, and improve customer service by more accurately predicting market demand fluctuations in real time. L. Bondarenko & Y. Liashenko (2023) studied the application of time series analysis methods for forecasting pricing in the real estate market, developing models that took into account seasonal fluctuations, macroeconomic indicators and regional market characteristics, providing investors and developers with tools for more informed decision-making in a changing economic environment.

**Suggested Citation:**
Zelenyi, D. (2025). Using big data for demand forecasting and dynamic pricing. *Economic Forum*, 15(2), 45-55. doi: 10.62763/ef/2.2025.45.



*Corresponding author

The authors G. Chornous & Y. Horbunova (2020) developed approaches to modelling and forecasting dynamic pricing factors in e-commerce, proposing mathematical models that took into account consumer behaviour patterns, the competitive environment and seasonal fluctuations in demand to optimise the pricing strategies of online retailers and maximise profits under different market conditions. O. Dobrovolska & N. Fenenko (2024) analysed the forecasting of trends in the real estate market by studying relevant determinants, identifying key influencing factors, including economic indicators, demographic variables, legislative norms and technological innovations, which made it possible to improve the accuracy of long-term forecasts for the development of the industry.

A. Guizzardi *et al.* (2021) proposed an intelligent approach to forecasting tourism demand based on big data from dynamic pricing, developing algorithms for analysing information arrays from online bookings, social networks and search engines to create adaptive pricing strategies that allowed tourism companies to maximise occupancy and revenue. R. Iftikhar & M.S. Khan (2020) studied big data analytics from social networks to forecast demand, developing a methodology for processing unstructured data from various platforms to identify consumer trends, user sentiments, and responses to marketing campaigns, enabling businesses to stay ahead of market trends and respond quickly to changes in customer preferences.

Scientists A. Kaminskyi *et al.* (2023) formed a dynamic basis for strategic forecasting of the consumer lending market of banks using the example of Ukraine, integrating macroeconomic indicators, borrower behaviour models and sectoral trends to create multifactorial predictive models that increased the effectiveness of banks' credit policies in conditions of economic instability. S. Kanyhin (2024) investigated the use of big data in financial management of enterprises, identifying methods for integrating analytics of massive sets of structured and unstructured information to optimise budgeting, cash flow management, investment risk assessment, and improve the overall efficiency of financial operations in the digital economy.

S. Kumar *et al.* (2022) conducted a comprehensive analysis of the past, present, and future of sustainable finance through the lens of big data analytics using machine learning to process scientific research, identifying evolutionary trends, key thematic areas, and promising areas for the development of sustainable financial instruments and practices in the global economic environment. The analysed scientific works did not link the analytical capabilities of big data with the needs of Ukrainian retail enterprises, taking into account the peculiarities of the Ukrainian consumer market and conditions of economic instability.

The aim of the study was to create and test the effectiveness of a comprehensive big data processing system for optimising management decisions in the field of demand forecasting and dynamic pricing in retail enterprises. Research objectives: development of methods for cleaning and normalising large amounts of information; creation of predictive models using modern machine learning algorithms; implementation and evaluation of the effectiveness of a dynamic pricing system for various enterprises.

## Materials and Methods

The research was conducted involving ten Ukrainian retail enterprises: ATB-Market LLC, Foxtrot LLC, Nova Liniya PJSC, ALLO LLC, Silpo-Fud LLC, METRO Cash and Carry Ukraine LLC, Epicentr K LLC, Rozetka LLC, Comfy Trade LLC, and INTERTOP Ukraine LLC, between September 2023 and February 2025. The work comprised four sequential stages with clearly defined tasks and methodological approaches. The main research data collection for this study was conducted during March-April 2024. During the first stage, primary data was collected from the following sources: APIs of trading platforms – Prom.ua (n.d.), Bigl.ua (n.d.), Rozetka (n.d.), ALLO (n.d.); open marketplace data; and data from price aggregators such as Price.ua (n.d.) and Hotline.ua (n.d.), and data from Euromonitor (n.d.). Web scraping methods were utilised to obtain competitor pricing information from Ukrainian online shops using Python software with BeautifulSoup and Scrapy libraries. Transaction data from retail chain POS terminals, covering over 4.5 million purchases, was collected. Information from loyalty programmes, containing data on 1.3 million unique customers, was integrated. Additionally, search query data from Google Trends and seasonal consumer price indices from the State Statistics Service of Ukraine (2025) were incorporated, along with reports from Youcontrol. Silpo-Fud LLC (n.d.), Youcontrol. METRO Cash and Carry Ukraine LLC (n.d.), Youcontrol. Epicentr K LLC (n.d.), Youcontrol. ATB-Market LLC (n.d.), Youcontrol. Comfy Trade LLC (n.d.), Youcontrol. INTERTOP Ukraine LLC (n.d.), Youcontrol. Foxtrot LLC (n.d.), Youcontrol. Nova Liniya PJSC (n.d.), Youcontrol. Rozetka LLC (n.d.), Youcontrol. ALLO LLC (n.d.) regarding the financial performance of the surveyed enterprises, accessed via Premium subscription. The second stage was dedicated to the cleaning and normalisation of the collected data. A methodology for outlier detection and removal using the Density-Based Spatial Clustering of Applications With Noise (DBSCAN) algorithm was applied. Heterogeneous data was integrated into a unified analytical platform based on Apache Hadoop. A metadata schema was developed to ensure the interoperability of various information sources. Procedures for regular data updates via API interfaces were established. Mechanisms for the anonymisation of personal data were implemented in accordance with the requirements of Law of Ukraine No. 80/94-VR (1994) and Law of Ukraine No. 2297-VI (2010).

In the third stage, predictive models were developed. A deep learning methodology with an LSTM architecture was utilised for sales time series forecasting, accounting for seasonality. Gradient boosting algorithms were applied to identify hidden dependencies between factors influencing demand. Probabilistic networks were implemented to model price elasticity of demand. Model training was conducted on graphics processing units using modern frameworks. Model validation was performed using cross-validation with data partitioned into training and test sets.

The fourth stage focused on the development and testing of the dynamic pricing system. Price optimisation algorithms based on multiple criteria were implemented. Mechanisms for real-time price adjustment were developed, considering demand, stock availability, and competitor pricing strategies. A monitoring dashboard was created for result visualisation. The developed system was tested across various retail chain formats – hypermarket, supermarket, specialised store, online shop, and construction materials hypermarket. A comprehensive set of statistical methods was used for data processing and analysis. To evaluate the quality of the predictive models, the following standard accuracy metrics were calculated: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination ($R^2$), which allowed for a comprehensive assessment of the accuracy and reliability of the developed models across various datasets.

## Results

### Data quality analysis and effectiveness of processing methods

The analytical processing of transactional information enabled the formation of a representative sample, encompassing various consumer segments and product categories. The integration of information from loyalty programmes provided the opportunity to conduct an in-depth analysis of consumer behaviour, taking into account clients' demographic characteristics and purchasing preferences. The results of the analysis concerning the structure and quality of the acquired data were presented in Table 1.

***Table 1.*** *Characteristics of data sources and its quality indicators*

| Data source | Number of items | Data completeness, % | Data accuracy, % | Data consistency, % | Overall quality index % |
|---|---|---|---|---|---|
| Web scraping of online shops | 6 websites | 92.4 | 87.6 | 85.3 | 88.4 |
| POS-transactions | 4,568,921 transactions | 98.7 | 99.2 | 97.8 | 98.6 |
| Loyalty programmes | 1,342,869 clients | 84.5 | 91.3 | 89.7 | 88.5 |
| Google Trends | 846 search terms | 100 | 95.8 | 100 | 98.6 |
| Data from State Statistics Service of Ukraine | 32 indices | 100 | 100 | 100 | 100 |
| Aggregated data | 5,912,868 records | 92.7 | 93.8 | 91.4 | 92.6 |

***Source:*** *compiled by the author*

Data completeness (%) reflected, how complete the data was (i.e., whether values were missing in fields). Data accuracy (%) indicated, how well the values correspond to reality or stated criteria. Data consistency (%) was a measure of the agreement of data among themselves (e.g., absence of logical contradictions), and the overall quality index (%) was an aggregated indicator that integrated all preceding indicators into a single quality assessment. According to the results of the data quality assessment in Table 1, the highest indicators were observed in official statistical data from the State Statistics Service of Ukraine (2025) and POS transaction data (Youcontrol. ATB-Market LLC, n.d.; Youcontrol. Silpo-Fud LLC, n.d.). Data obtained through web-scraping demonstrated lower quality indicators due to the presence of missing values and the unstructured nature of certain information fragments. It was revealed that information from loyalty programmes (Youcontrol. Comfy Trade LLC, n.d.; Youcontrol. Foxtrot LLC, n.d.) was characterised by insufficient completeness, owing to the voluntary provision of personal data by customers and differences in the structure of loyalty programmes across various retail chains. The results of the evaluation of the effectiveness of data cleaning and normalisation methods were presented in Table 2.

***Table 2.*** *Evaluation of the effectiveness of data cleaning and normalisation methods for analysed enterprises*

| Data processing method | Volume of processed data, GB | Outliers detected, % | Processing speed, GB/hour | Information loss, % | Data quality increase, % |
|---|---|---|---|---|---|
| DBSCAN algorithm | 1,742.8 | 3.21 | 64.2 | 0.96 | 7.85 |
| Duplicate removal | 1,742.8 | n/f | 127.6 | 2.17 | 4.32 |
| Filling in gaps | 1,742.8 | n/f | 85.3 | 0 | 5.64 |
| Normalisation of numerical values | 846.4 | 1.85 | 183.5 | 0.12 | 2.93 |

*Table 2, Continued*

| Data processing method | Volume of processed data, GB | Outliers detected, % | Processing speed, GB/hour | Information loss, % | Data quality increase, % |
|---|---|---|---|---|---|
| Integration of heterogeneous data | 1,742.8 | n/f | 38.7 | 0.47 | 8.26 |
| Comprehensive processing | 1,742.8 | 5.06 | 27.4 | 3.72 | 12.47 |

***Note:*** *n/f – not found*
***Source:*** *compiled by the author*

An analysis of the effectiveness of data processing methods showed in Table 2 that the use of the DBSCAN algorithm allowed the detection of 3.21% of anomalous values that could negatively affect the quality of predictive models. The method of normalising numerical values demonstrated the highest processing speed (183.5 GB/hour), while complex processing using all methods was characterised by the lowest speed (27.4 GB/hour). It was found that filling in gaps using machine learning methods made it possible to avoid information loss, while simultaneously improving data quality by 5.64%.

According to the results obtained, the comprehensive application of cleaning and normalisation methods provided a cumulative increase in data quality of 12.47% with a moderate level of information loss (3.72%). The integration of heterogeneous data into a single analytical platform demonstrated the highest data quality gain among all the methods used (8.26%), confirming the effectiveness of the chosen architecture for working with large data sets in the context of demand forecasting and dynamic pricing.

**Effectiveness of predictive models and results of dynamic pricing system implementation**

A set of machine learning methods was used to analyse the effectiveness of various predictive models in the context of forecasting demand for retail goods. The models were compared based on standard metrics for evaluating forecasting accuracy, which identified the most effective algorithms for further implementation in the dynamic pricing system. The results of a comparative analysis of the effectiveness of various predictive models were presented in Table 3.

***Table 3.*** *Comparative analysis of the effectiveness of predictive models*

| Model | RMSE | MAE | MAPE, % | $R^2$ | Training time, hours | Prediction time, ms |
|---|---|---|---|---|---|---|
| LSTM network | 145.3 | 112.7 | 6.83 | 0.874 | 23.6 | 186 |
| XGBoost | 162.8 | 127.2 | 7.65 | 0.856 | 8.2 | 54 |
| Bayesian network | 183.4 | 142.6 | 8.42 | 0.823 | 12.4 | 76 |
| Linear regression | 287.5 | 215.3 | 12.74 | 0.712 | 1.2 | 12 |
| ARIMA | 207.1 | 164.2 | 9.38 | 0.805 | 3.5 | 28 |
| Ensemble model | 131.8 | 101.5 | 5.92 | 0.896 | 31.4 | 248 |

***Note:*** *RMSE – Root Mean Square Error; MAE – Mean Absolute Error; MAPE – Mean Absolute Percentage Error;* $R^2$ *– coefficient of determination; ARIMA – Autoregressive Integrated Moving Average*
***Source:*** *compiled by the author*

According to the results obtained in Table 3, the ensemble model combining LSTM network, XGBoost and Bayesian network demonstrated the best predictive ability. This model was characterised by the lowest values of RMSE = 131.8, MAE = 101.5 and MAPE = 5.92%. The coefficient of determination of the ensemble model was 0.896, confirming its high predictive ability. At the same time, this model required the most computing resources, which was reflected in the long training time (31.4 hours) and the highest prediction time (248 ms). Among the individual models, the LSTM network demonstrated the highest efficiency ($R^2$ = 0.874, MAPE = 6.83%). This result confirmed the feasibility of using deep learning to predict sales time series, taking into account seasonality. Linear regression showed the lowest efficiency among all tested models ($R^2$ = 0.712, MAPE = 12.74%), which was explained by the complexity and non-linearity of the relationships between factors affecting demand in the retail sector. The results of the evaluation of the economic effectiveness of the implementation of a dynamic pricing system compared to traditional approaches to pricing were presented in Table 4.

***Table 4.*** *Indicators of the effectiveness of implementing a dynamic pricing system*

| Company | Format | Revenue growth, % | Marginal profit growth, % | Inventory reduction, % | ROI of implementation, % | Payback period, months |
|---|---|---|---|---|---|---|
| Silpo-Fud LLC | Hypermarket | 8.72 | 11.28 | 6.54 | 327 | 4.2 |
| METRO Cash and Carry Ukraine LLC | Hypermarket | 7.64 | 10.35 | 5.92 | 298 | 4.8 |
| ATB-Market LLC | Supermarket | 7.85 | 9.46 | 5.87 | 274 | 5.1 |
| Foxtrot LLC | Specialised store | 9.24 | 10.75 | 7.18 | 382 | 3.6 |
| Comfy Trade LLC | Specialised store | 8.96 | 10.32 | 6.85 | 365 | 3.9 |
| INTERTOP Ukraine LLC | Specialised store | 8.53 | 9.84 | 6.42 | 341 | 4.1 |
| Nova Liniya PJSC | Construction materials hypermarket | 7.94 | 9.72 | 6.13 | 296 | 4.7 |
| Epicentr K LLC | Construction materials hypermarket | 8.46 | 10.54 | 6.38 | 318 | 4.4 |
| Rozetka LLC | Online store | 12.37 | 14.53 | 9.42 | 516 | 2.7 |
| ALLO LLC | Online store | 11.85 | 13.96 | 8.78 | 487 | 2.9 |
| Average value | | 9.16 | 11.08 | 6.95 | 360 | 4 |

***Note:*** *ROI – return on investment*
***Source:*** *compiled by the author*

An analysis of the economic indicators of the effectiveness of implementing a dynamic pricing system from Table 4 showed a significant positive impact on key business metrics for all retail formats studied. The highest results were demonstrated by the online stores of Rozetka LLC and ALLO LLC, where revenue growth amounted to 12.37% and 11.85%, respectively, and marginal profit growth amounted to 14.53% and 13.96%. Specialised stores also showed high efficiency, especially Foxtrot LLC with a revenue growth of 9.24%. Hypermarkets and supermarkets showed slightly lower, but still unambiguous indicators of improvement in economic results.

The introduction of a dynamic pricing system contributed to a 6.95% reduction in inventory levels on average, indicating improved inventory management and optimisation of assortment policy. The positive impact on consumer behaviour was reflected in an increase in the frequency of purchases and an increase in the average cheque. The highest ROI was recorded for online stores (516% for Rozetka LLC) and specialised stores (382% for Foxtrot LLC), which were characterised by flexible pricing policies and technological readiness to implement innovative solutions. The results of a comparative analysis of the effectiveness of the system's implementation by type of enterprise were presented in Table 5.

***Table 5.*** *Effectiveness of the implementation of the dynamic pricing system by type of enterprise*

| Company characteristics | Number of companies | Revenue growth, % | Increase in return on sales, p.p. | ROI of implementation, % | p-value |
|---|---|---|---|---|---|
| **By business size** | | | | | |
| Small | 3 | 10.23±2.14 | 3.65±0.87 | 405±86 | 0.024 |
| Medium | 4 | 8.76±1.92 | 2.94±0.73 | 357±72 | 0.031 |
| Large | 3 | 6.84±1.47 | 2.35±0.58 | 274±63 | 0.028 |
| **By region** | | | | | |
| Central | 4 | 8.93±1.86 | 3.12±0.74 | 364±75 | 0.033 |
| Western | 2 | 9.27±2.05 | 3.24±0.79 | 378±82 | 0.029 |
| Eastern | 2 | 7.85±1.74 | 2.76±0.65 | 312±69 | 0.042 |
| Southern | 2 | 8.42±1.95 | 2.98±0.71 | 342±77 | 0.037 |
| **By specialisation** | | | | | |
| Food products | 3 | 7.23±1.68 | 2.54±0.63 | 286±65 | 0.027 |
| Non-food products | 5 | 9.86±2.23 | 3.45±0.82 | 412±89 | 0.022 |
| Mixed assortment | 2 | 8.52±1.97 | 2.98±0.75 | 352±78 | 0.035 |

***Note:*** *ROI – return on investment*
***Source:*** *based on State Statistics Service of Ukraine (2025)*

Based on the results of the comparative analysis in Table 5, statistically significant differences in the effectiveness of implementing a dynamic pricing system depending on the characteristics of enterprises were established ($p < 0.05$). In terms of business size, small enterprises demonstrated the highest performance with a revenue growth of 10.23 ± 2.14% and an ROI of 405 ± 86%. This effect was explained by the greater flexibility of small enterprises and their ability to respond quickly to changes in market conditions.

In regional terms, the best results were recorded in the Western region, namely in Lviv and Volyn regions (revenue growth of 9.27 ± 2.05%), and the lowest in the Eastern region (Kharkiv and Dnipropetrovsk regions) (7.85 ± 1.74%). The regional differences identified correlate with the overall level of retail trade development and business process digitalisation in the respective regions. In terms of specialisation, the largest revenue growth was observed in companies specialising in non-food products (9.86 ± 2.23%), which was due to the higher price elasticity of demand for this category of goods compared to food products.

So, based on a thorough analysis of the obtained results, tailored guidelines were developed to identify optimal strategies for implementing a dynamic pricing system across various types of enterprises. For small and medium-sized enterprises, it was involved full implementation of the system, with a focus on quickly adapting to changes in market conditions. For large enterprises, a phased approach was proposed, with an initial focus on categories of goods with high price elasticity of demand. An important aspect of the study was the confirmation of the high scalability of the proposed system for enterprises of different types and sizes. The analysis revealed that, despite differences in performance indicators, all the companies studied demonstrated a positive economic effect from the implementation of the dynamic pricing system. This confirmed the universality and adaptability of the developed approach to various business models in the retail sector.

## Discussion

The results of the study confirmed the key role of big data analytics in shaping modern approaches to demand forecasting and establishing adaptive pricing strategies in retail. It has been found that the implementation of a dynamic pricing system based on the analysis of massive data sets significantly enhanced the competitiveness of enterprises through the optimisation of pricing strategies, increased forecasting accuracy and improved inventory management. The results of V. Kustov & M. Kovalenko (2024) on information support for process management in the context of digitalisation confirmed the identified pattern of increasing the efficiency of analytical systems with the increase in the level of digitalisation of business processes. The authors demonstrated higher ROI (375-420%) for enterprises with developed digital infrastructure, which was consistent with the results obtained for online stores (ROI = 516%).

Research by J. Maksymova (2021) confirmed the patterns identified in the study results regarding the strategic importance of big data for improving the competitiveness of enterprises, demonstrating similar indicators of economic efficiency of the implementation of analytical systems with an average productivity increase of 7-9%. There was complete agreement on the differentiation of implementation effectiveness depending on industry specifics, confirming the validity of the conclusion about lower effectiveness for food companies (7.23% revenue growth) compared to non-food companies (9.86%). Neural network machine learning methods for forecasting large noisy data, studied by A. Maltsev (2022), demonstrated similar effectiveness to the results obtained for LSTM networks with RMSE in the range of 135-155 units, which correlated with the indicators of this study (RMSE = 145.3). At the same time, the author found a more pronounced influence of model hyperparameters on the accuracy of forecasting, which was not fully reflected in this study and required further study for further improvement of the models.

The results of modelling and forecasting demand for a digital product by O. Novoseletskyy *et al.* (2021) revealed similar patterns to the data obtained for online stores, in particular regarding the high price elasticity of demand and the effectiveness of dynamic pricing with a 10-14% increase in revenue. There were differences in the methodological approaches to consumer segmentation, as the authors used cluster analysis based on purchasing behaviour, while the presented study used demographic characteristics and purchase history. T. Petriv (2024) studied the ecosystem of leading software development companies in Ukraine, identifying a growing demand for big data analytics solutions in the retail sector. Scientist's findings accepted the results of the study on the high effectiveness of implementing analytical systems for demand forecasting and dynamic pricing, demonstrating similar ROI indicators for technological solutions (380-450%) compared to those obtained in the study (360% on average).

The study by D. Schultz *et al.* (2023) on causal forecasting for pricing complements the results obtained, offering methods for identifying causal relationships between pricing strategies and market indicators, which were not fully taken into account in this study. The authors demonstrated the potential to increase forecasting accuracy by 5-8%, when causal relationships were taken into account, which was a promising direction for improving the developed dynamic pricing system. Predictive big data analytics for demand forecasting in supply chains, researched by M. Seyedan & F. Mafakheri (2020), established the conclusions regarding the positive impact of analytical systems on reducing inventory levels (the authors recorded figures in the range of 5-7%, which corresponded to the average value of 6.95%). The authors also emphasised the importance of integrating data from external sources, which was consistent with the high effectiveness of web scraping and the use of Google Trends data identified in this study.

The research by I. Shkyrta & V. Lazar (2019) on the essence and possibilities of big data technology for business was consistent with the results obtained regarding the differentiation of implementation effectiveness depending on the size of the business and technological readiness. The authors also confirmed the conclusion about the higher adaptability of small enterprises to the implementation of innovative analytical solutions, which explained the highest ROI indicators for small businesses (405%) identified in the study. A. Sharma (2025) analysed in detail the application of big data in retail, highlighting key areas of use, including demand forecasting, pricing optimisation and personalisation of offers. The author emphasised the importance of a comprehensive approach to big data analysis, which was fully consistent with the study's findings on the highest effectiveness of ensemble models ($R^2 = 0.896$) compared to individual algorithms. The dynamic pricing strategy for logistics revenue management using intelligent data analysis technology, presented by D. Xu *et al.* (2020), proved the conclusions regarding the feasibility of using ensemble models to improve forecasting accuracy. Researchers demonstrated similar performance indicators for the ensemble model ($R^2 = 0.883$) compared to the obtained result ($R^2 = 0.896$), although there were differences in the structure of the ensemble and approaches to weighting individual models.

The study by J.T. Hancock & T.M. Khoshgoftaar (2020) on the application of the CatBoost algorithm for big data analysis was consistent with the results obtained regarding the high efficiency of gradient boosting methods, although the authors found CatBoost to be superior to XGBoost, while the study used mainly XGBoost. The authors noted the potential to increase forecasting accuracy by 3-5%, when replacing XGBoost with CatBoost, which opened up opportunities for further improvement of the developed system. K. Zabor (2023) analysed leading European companies in the field of big data, identifying key trends in the development of the industry and promising areas for the application of analytical solutions. The results of his research confirmed the identified pattern of higher efficiency of implementing dynamic pricing systems for non-food companies (9.86% revenue growth) compared to food companies (7.23%), which was explained by differences in demand elasticity and frequency of assortment updates.

The results of A. Razzaq & X. Yang (2023) on the use of web crawling and big data technology to assess digital finance demonstrated technological approaches similar to those used in the study for web scraping, although the authors achieved higher data completeness rates (95.7% compared to 92.4%). The discrepancies found can be explained by the different areas of application and structure of the analysed web resources, confirming the need to adapt web scraping methods to the specifics of the industry under study. The study by W.K. Jawad & A.M. Al-Bakry (2023) established the patterns identified in the study results regarding the effectiveness of different methods of big data processing, demonstrating similar DBSCAN performance indicators with an average increase in data quality of 7-8% compared to the obtained 7.85%. There was significant agreement on the importance of pre-processing data to improve forecasting accuracy, confirming the validity of the attention paid to the data cleaning and normalisation stage, with an overall quality increase of 12.47%.

The transformation of business process structures in the modern economic environment, as studied by A. Pakki (2025), demonstrated methodological approaches similar to those used in the this study, especially with regard to assessing the impact of digitalisation on business process efficiency. The author recorded slightly lower productivity growth rates, when implementing analytical systems (5-7% compared to the 9.16% revenue growth obtained), which was explained by the broader focus of the study and its coverage of various sectors of the economy, not just retail trade. Results of detecting anomalies and threats in big data of hotel and restaurant industry enterprises H. Liavynets *et al.* (2024) complemented the conclusions regarding the effectiveness of the DBSCAN algorithm for detecting anomalous values, although the authors demonstrated a lower percentage of detected outliers (2.54% compared to 3.21%). There were differences in the approaches to processing the detected anomalies, as the authors focused on the security aspect, while in this study, priority was given to improving the quality of predictive models.

The study by V. Nesterov (2024) on the impact of big data analytics on business efficiency in the digital age aligned with the findings regarding the differentiation of economic effects based on industry specifics and enterprises' technological readiness. The author found patterns similar to those obtained regarding the higher return on investment in digital analytical solutions for enterprises with a mixed assortment (ROI = 330-370%) compared to the results of this study (ROI = 352 ± 78%), which confirmed the validity of the obtained estimates of economic efficiency. The server pool model for assessing energy consumption in big data processing, proposed by Y. Ponochovnyi *et al.* (2021), expanded on the results obtained, adding an important aspect of energy efficiency assessment that was not fully taken into account in this study. The authors demonstrated the potential to reduce energy consumption by 15-20% by optimising the data processing architecture, which was a promising direction for further improvement of the developed system, given the growing importance of environmental aspects of business and rising electricity costs.

Thus, the analysis of the literature confirmed the validity of the results obtained in the study regarding the effectiveness of big data processing methods, the performance of predictive models, and the economic effects of implementing a dynamic pricing system. The discrepancies identified with individual studies were explained by industry specifics, regional characteristics, and differences in methodological approaches, which opened up prospects for further improvement of the developed system through the integration of causal models, the application of the CatBoost algorithm, and the expansion of data sources.

## Conclusions

The analysis quality of big data for demand forecasting and dynamic pricing confirmed the high representativeness of the collected information, with an overall aggregate data quality index of 92.6%. Transaction information from POS terminals, which was characterised by the highest accuracy rates (99.2%), was identified as the priority source for building predictive models. The testing of data cleaning and normalisation methods demonstrated their high efficiency. The comprehensive application of various processing methods ensured a 12.47% increase in data quality. The use of the DBSCAN algorithm made it possible to identify and remove anomalous values without significant information loss, and the integration of heterogeneous data into a single analytical platform contributed to the improvement of the efficiency of further analysis. The ensemble model, which combined an LSTM network, XGBoost, and a Bayesian network, demonstrated the highest prediction accuracy with a MAPE of 5.92%. This result confirmed the feasibility of using combined methods to predict complex non-linear processes in the retail sector. The implementation of the dynamic pricing system provided a significant economic effect for all retail formats studied. The average return on investment was 360%, which indicated the high economic efficiency of the developed solution. The highest performance was observed in the online store segment of Rozetka LLC and ALLO LLC, where the increase in marginal profit reached 14.53% and 13.96%, respectively, confirming the promise of implementing systems in e-commerce.

A comparative analysis of the effectiveness of implementation for different types of enterprises revealed statistically significant differences depending on the size of the business, geographical location and specialisation. The largest increase in revenue was observed in small enterprises (10.23 ± 2.14%), due to their greater flexibility and adaptability. In regional terms, the best results were recorded in the Lviv and Volyn regions (9.27 ± 2.05%), while in terms of specialisation, the leaders were Comfy Trade LLC and INTERTOP Ukraine LLC, which sell non-food products (9.86 ± 2.23%). The proposed system was flexible and suitable for implementation in enterprises of various sizes and specialisations, which ensured its practical value and versatility. The study had limitations, in particular geographical unevenness of the sample and instability of macroeconomic conditions, which could have affected the accuracy of individual forecasts. Insufficient digitalisation of small enterprises in some regions created technical barriers to the full implementation of the system. To improve the effectiveness of the implementation of the dynamic pricing system, it was recommended to: regularly update predictive models to reflect new data; adapt model parameters to seasonal fluctuations in demand; use a differentiated approach to pricing for different product categories; implement a phased strategy for large enterprises with an initial focus on categories with high demand elasticity; integrate the system with existing ERP solutions to ensure business continuity. A promising area for further research is the integration of dynamic pricing systems with artificial intelligence technologies to achieve a higher level of personalisation of price offers and automation of management decision-making processes based on predictive analytics.

## Acknowledgements

None.

## Funding

None.

## Conflict of Interest

None.

## References

[1] ALLO. (n.d.). Retrieved from https://allo.ua/.
[2] Arguelles Jr, P., & Pólkowski, Z. (2023). Impact of big data on supply chain performance through demand forecasting. *International Journal of Computations, Information and Manufacturing*, 3(1), 19-26. doi: 10.54489/ijcim.v3i1.232.
[3] Bigl.ua. (n.d.). Retrieved from https://bigl.ua/.
[4] Bondarenko, L., & Liashenko, Y. (2023). Application of time series analysis methods for forecasting pricing in the real estate market. *Automobile Roads and Road Construction*, 114, 233-240. doi: 10.33744/0365-8171-2023-114.1-233-240.
[5] Chornous, G., & Horbunova, Y. (2020). Modeling and forecasting dynamic factors of pricing in e-commerce. In *IT&I-2020 information technology and interactions* (pp. 71-82). Kyiv: KNU Taras Shevchenko.
[6] Dobrovolska, O., & Fenenko, N. (2024). Forecasting trends in the real estate market: Analysis of relevant determinants. *Financial Markets, Institutions and Risks*, 8(3), 227-253. doi: 10.61093/fmir.8(3).227-253.2024.
[7] Euromonitor. (n.d.). *Market Research Ukraine*. Retrieved from https://www.euromonitor.com/ukraine.

[8] Guizzardi, A., Pons, F.M.E., Angelini, G., & Ranieri, E. (2021). Big data from dynamic pricing: A smart approach to tourism demand forecasting. *International Journal of Forecasting*, 37(3), 1049-1060. doi: 10.1016/j.ijforecast.2020.11.006.

[9] Hancock, J.T., & Khoshgoftaar, T.M. (2020). CatBoost for big data: An interdisciplinary review. *Journal of Big Data*, 7, article number 94. doi: 10.1186/s40537-020-00369-8.

[10] Hotline.ua. (n.d.). Retrieved from https://hotline.ua/.

[11] Iftikhar, R., & Khan, M.S. (2020). Social media big data analytics for demand forecasting: Development and case implementation of an innovative framework. *Journal of Global Information Management*, 28(1), 103-120. doi: 10.4018/JGIM.2020010106.

[12] Jawad, W.K., & Al-Bakry, A.M. (2023). Big data analytics: A survey. *Iraqi Journal for Computers and Informatics*, 49(1), 41-51. doi: 10.25195/ijci.v49i1.384.

[13] Kaminskyi, A., Versal, N., Petrovskyi, O., & Prykaziuk, N. (2023). Dynamic framework for strategic forecasting of the bank consumer loan market: Evidence from Ukraine. *Banks and Bank Systems*, 18(3), 87-100. doi: 10.21511/bbs.18(3).2023.08.

[14] Kanyhin, S. (2024). Big data in corporate financial management. *Economics, Management and Administration*, 3, 97-104. doi: 10.26642/ema-2024-3(109)-97-104.

[15] Kumar, S., Sharma, D., Rao, S., Lim, W.M., & Mangla, S.K. (2022). Past, present, and future of sustainable finance: Insights from big data analytics through machine learning of scholarly research. *Annals of Operations Research*, 345, 1061-1104. doi: 10.1007/s10479-021-04410-8.

[16] Kustov, V., & Kovalenko, M. (2024). Information support for managing processes on exchanges in the context of digitalization. *Modeling the Development of the Economic Systems*, 2, 47-57. doi: 10.31891/mdes/2024-12-7.

[17] Law of Ukraine No. 2297-VI "On Personal Data Protection". (2010, June). Retrieved from https://zakon.rada.gov.ua/laws/show/2297-17#Text.

[18] Law of Ukraine No. 80/94-VR "On Protection of Information in Information and Communication Systems". (1994, July). Retrieved from https://zakon.rada.gov.ua/laws/show/80/94-%D0%B2%D1%80#Text.

[19] Liavynets, H., Rohliev, Y., & Bortnichuk, O. (2024). Anomaliy and fraud detection mode from big data of hotel and restaurant industry enterprises. *Economy and Society*, 70. doi: 10.32782/2524-0072/2024-70-29.

[20] Maksymova, J. (2021). The importance of big data in industry and economy. *Economy and Society*, 28. doi: 10.32782/2524-0072/2021-28-38.

[21] Maltsev, A. (2022). Methods of machine learning of the neural network to predict large noisy data using modern programming languages. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 1, 39-43. doi: 10.32782/it/2022-1-6.

[22] Nesterov, V. (2024). Exploring the impact of big data analytics on business performance in the digital era. *Information Technology and Society*, 1, 70-76. doi: 10.32689/maup.it.2024.1.10.

[23] Novoseletskyy, O., Zubenko, I., & Gurina, M. (2021). Modeling and forecasting demand for a digital product. *Scientific Notes of Ostroh Academy National University, "Economics" Series*, 22, 95-101. doi: 10.25264/2311-5149-2021-22(50)-95-101.

[24] Pakki, A. (2025). Transformation of business process structure in modern conditions of economic development. *Modeling the Development of the Economic Systems*, 1, 168-175. doi: 10.31891/mdes/2025-15-23.

[25] Petriv, T. (2024). Top software development companies in Ukraine with a strong portfolio. *N-ix*. Retrieved from https://www.n-ix.com/software-development-companies-in-ukraine/.

[26] Ponochovnyi, Y., Pryada, O., Soroka, Y., & Dikun, Y. (2021). Model of server pool for estimation of energy consumption in big data processing. *IT Synergy*, 1, 26-31. doi: 10.53920/its-2021-1-4.

[27] Price.ua. (n.d.). Retrieved from https://price.ua/.

[28] Prom.ua. (n.d.). Retrieved from https://prom.ua/.

[29] Razzaq, A., & Yang, X. (2023). Digital finance and green growth in China: Appraising inclusive digital finance using web crawler technology and big data. *Technological Forecasting and Social Change*, 188, article number 122262. doi: 10.1016/j.techfore.2022.122262.

[30] Rozetka. (n.d.). Retrieved from https://rozetka.com.ua/.

[31] Schultz, D., Stephan, J., Sieber, J., Yeh, T., Kunz, M., Doupe, P., & Januschowski, T. (2023). *Causal forecasting for pricing*. doi: 10.48550/arXiv.2312.15282.

[32] Seyedan, M., & Mafakheri, F. (2020). Predictive big data analytics for supply chain demand forecasting: Methods, applications, and research opportunities. *Journal of Big Data*, 7, article number 53. doi: 10.1186/s40537-020-00329-2.

[33] Sharma, A. (2025). Big data in retail: Industry applications, benefits & best practices. *Turing*. Retrieved from https://www.turing.com/resources/big-data-in-retail.

[34] Shkyrta, I., & Lazar, V. (2019). Big data technology: Essence, opportunities for business. *Scientific Bulletin of Mukachevo State University. Series "Economics"*, 6(2), 51-56. doi: 10.31339/2313-8114-2019-2(12)-51-56.
[35] State Statistics Service of Ukraine. (2025). *Economic statistics. Economic activity. Business activity*. Retrieved from https://www.ukrstat.gov.ua/operativ/menu/menu_u/sze_20.htm.
[36] Xu, D., *et al.* (2020). Dynamic pricing strategy for logistics revenue management using data mining technology. In *Proceedings of the 2020 4th international symposium on computer science and intelligent control* (article number 7). New York: Association for Computing Machinery. doi: 10.1145/3440084.3441183.
[37] Youcontrol. ALLO LLC. (n.d.). Retrieved from https://youcontrol.com.ua/catalog/company_details/30012848/.
[38] Youcontrol. ATB-Market LLC. (n.d.). Retrieved from https://youcontrol.com.ua/catalog/company_details/30487219/.
[39] Youcontrol. Comfy Trade LLC. (n.d.). Retrieved from https://youcontrol.com.ua/ru/catalog/company_details/36962487/.
[40] Youcontrol. Epicentr K LLC. (n.d.). Retrieved from https://youcontrol.com.ua/catalog/company_details/32490244/.
[41] Youcontrol. Foxtrot LLC. (n.d.). Retrieved from https://youcontrol.com.ua/catalog/company_details/23703827/.
[42] Youcontrol. INTERTOP Ukraine LLC. (n.d.). Retrieved from https://youcontrol.com.ua/catalog/company_details/41097426/.
[43] Youcontrol. METRO Cash and Carry Ukraine LLC. (n.d.). Retrieved from https://youcontrol.com.ua/catalog/company_details/32049199/.
[44] Youcontrol. Nova Liniya PJSC. (n.d.). Retrieved from https://youcontrol.com.ua/catalog/company_details/32816917/.
[45] Youcontrol. Rozetka LLC. (n.d.). Retrieved from https://youcontrol.com.ua/catalog/company_details/34047188/.
[46] Youcontrol. Silpo-Fud LLC. (n.d.). Retrieved from https://youcontrol.com.ua/catalog/company_details/40720198.
[47] Zabor, K. (2023). The best Big Data companies in Europe. *N-ix*. Retrieved from https://www.n-ix.com/top-big-data-companies-europe/.

# Використання big data для прогнозування попиту та динамічного ціноутворення

**Дмитро Зелений**
Магістр
Київський національний економічний університет імені Вадима Гетьмана
03057, просп. Берестейський, 54/1, м. Київ, Україна
https://orcid.org/0009-0003-2719-6357

**Анотація.** Дослідження передбачало розробку та впровадження системи прогнозування попиту й динамічного ціноутворення на основі аналізу великих даних для підприємств роздрібної торгівлі України. Методологія роботи охопила збір масивного набору даних з онлайн-торговельних платформ, транзакційної інформації з терміналів продажів та програм лояльності, які продемонстрували загальний індекс якості агрегованих даних на рівні 92,6 %. Застосування комплексу методів очищення та нормалізації забезпечило приріст якості даних на 12,47 %. Порівняльний аналіз прогностичних моделей виявив найвищу ефективність LSTM-мережі серед індивідуальних моделей ($R^2 = 0,874$, MAPE = 6,83 %) та ансамблевої моделі серед усіх апробованих підходів ($R^2 = 0,896$, MAPE = 5,92 %). Впровадження розробленої системи на підприємствах різних форматів, таких як ТОВ «АТБ-Маркет», ТОВ «Фокстрот», ПрАТ «Нова Лінія», ТОВ «АЛЛО», ТОВ «Сільпо-Фуд», ТОВ «МЕТРО Кеш енд Керрі Україна», ТОВ «Епіцентр К», ТОВ «Rozetka», ТОВ «Комфі Трейд», ТОВ «ІНТЕРТОП Україна» показало підвищення економічної ефективності з середнім приростом виручки на 9,16 %, маржинального прибутку на 11,08 % та зниженням рівня запасів на 6,95 %. Найвищу результативність продемонстрували інтернет-магазини «Rozetka» та «ALLO» з показниками повернення інвестицій (ROI) на рівні 516 % та періодом окупності 2,7 місяця. Регіональний аналіз виявив значні відмінності в ефективності впровадження системи з найкращими показниками у західному регіоні, а саме у Львівській та Волинській областях (приріст виручки 9,27 ± 2,05 %) та серед підприємств непродовольчої спеціалізації, зокрема ТОВ «Комфі Трейд», ТОВ «ІНТЕРТОП Україна» (9,86 ± 2,23 %). Найвищу адаптивність до динамічного ціноутворення продемонстрували малі підприємства з приростом виручки 10,23 ± 2,14 % та ROI 405 ± 86 %. Дослідження підтвердило високу масштабованість та адаптивність запропонованого підходу для українського ринку та дозволило розробити диференційовані рекомендації щодо впровадження системи для різних типів підприємств роздрібної торгівлі з урахуванням їх розміру, регіонального розташування та товарної спеціалізації

**Ключові слова:** масиви даних; прогностичні моделі; економічна ефективність; роздрібна мережа; споживча поведінка