

Proceedings of the 2010 ACM Conference on Information Technology Education (Midland, Michigan, USA, October 7–9, 2010), 73-84.

3. C. Pahl, E. Holohan. «Applications of Semantic Web Technology to Support Learning Content Development», *Interdisciplinary Journal of E-learning and Learning Objects*, 5:1 (2009), 1-25.

4. Brusilovsky, Peter. (1999). *Adaptive and Intelligent Technologies for Web-based Education*. KI. 13. 19-25.

5. Brusilovsky, P., Peylo, C. (2003) *Adaptive and intelligent web-based educational systems*. *Int. J. Artif. Intell. Ed.* 13, 159-172.

## **СИСТЕМА ПРОГНОЗУВАННЯ ВАРТОСТІ ТОВАРІВ НА ОСНОВІ МОДИФІКОВАНОГО АЛГОРИТМУ ARIMA ДЛЯ ОФІСНИХ КОМП'ЮТЕРІВ**

*Мальченко П.О.* – магістрант, [twink1337zhaba@gmail.com](mailto:twink1337zhaba@gmail.com)

*Чорноморський національний університет імені Петра Могили*

*Іванов Г.О.* – к. т. н., доц., [ivanovgo0708@gmail.com](mailto:ivanovgo0708@gmail.com)

*Миколаївський національний аграрний університет*

*Україна, м. Миколаїв*

## **THE SYSTEM FOR FORECASTING THE COST OF GOODS BASED ON THE MODIFIED ARIMA ALGORITHM FOR OFFICE COMPUTERS**

*Malchenko P.* – master's student, [twink1337zhaba@gmail.com](mailto:twink1337zhaba@gmail.com)

*Black Sea National University*

*Ivanov G.* – Candidate of technical Sciences, Associate Professor,

[ivanovgo0708@gmail.com](mailto:ivanovgo0708@gmail.com)

*Mykolaiv National Agrarian University*

*Ukraine, Mykolaiv*

*The paper analyzes the use of modified ARIMA algorithm for office computers in systematic approach with other models like GAM creating an opportunity for predictive modeling on computers with old hardware. Forecasting was always a problematic question in different spheres of life and numerous methods and approaches are already in use. Product cost forecasting is an important topic in finance and economics, which involves effective forecasting of price dynamics that can minimize the risk of losses and maximize profits. Predicting prices is a difficult task due to complex time series models.*

**Keywords:** *forecasting, time series, machine learning, decomposition, autocorrelation, stationarity, models, metrics, R.*

**Актуальність дослідження** полягає у необхідності підвищення якості прогнозування, відборі кращих моделей для відповідних наборів даних. Це дозволить зменшити збитки, що несуть компанії у результаті коливання вартості за рахунок підготовки заздалегідь.

**Метою роботи** є підвищення якості короткострокового прогнозування вартості товарів для слабких комп'ютерів за рахунок використання модифікованого методу ARIMA.

**Висновок.** В результаті проведеного дослідження було отримано моделі, що, в середньому, дають покращення якості прогнозування на 2% для різних наборів даних.

Дані часового ряду – це послідовність числових спостережень, природно впорядкованих у часі. Останнім часом стрімкий розвиток інформаційних технологій призвів до ситуації, коли величезні обсяги даних накопичуються з великою швидкістю і фактично утворюють різноманітні часові ряди. Моделювання таких часових рядів є надзвичайно важливим і привертає увагу як практиків, так і дослідників. Однак це також вважається досить складною проблемою через багато складних характеристик, які часто присутні в часових рядах, таких як нерівномірності, нестабільність, тренди та шуми, тощо. Для моделювання часових рядів на основі їх поточної та минулої поведінки було розроблено низку методів.

Модель інтегрованої авторегресії середнього ковзного (ARIMA) і узагальнена адитивна модель (GAM) широко використовувалися для прогнозування часових рядів у сфері цін на біржі М'янми. Дані готуються для аналізу часових рядів шляхом виконання етапів попередньої обробки даних, таких як перетворення позначок часу, стаціонарна ідентифікація та стаціонарна обробка.

Наявні публікації на схожу тематику пропонують використання методів машинного навчання (моделі ARIMA [1], GAM, SVR, модифіковані та комбіновані моделі), штучного інтелекту, тощо.

Перевагою використання моделей ARIMA є їх універсальність і те, що їх можна налаштувати для прогнозування різних часових явищ [2, 3, 4, 5]. Завдяки своїй ефективності як класичного методу прогнозування він широко використовується, наприклад, для прогнозування якості води, робочих навантажень у хмарних додатках, індексу EBITDA для фінансових показників і короткострокових навантажень клієнтів.

Для розробки програмного забезпечення було використано середовище розробки R. R має значні можливості для здійснення статистичних аналізів, включаючи лінійну і нелінійну регресію, класичні статистичні тести, аналіз часових рядів (серій), кластерний аналіз і багато іншого. R легко розбудовується завдяки використанню додаткових функцій і пакетів доступних на сайті Comprehensive R Archive Network.

Для навчання моделі було використано набір даних, що представлений вартістю акцій компанії Amazon за 2016-2019 роки. Завантаживши дані, можна побачити, що це часовий ряд, визначений 3-змінними: дата, назва компанії, ціна.

Першим кроком у прогнозуванні є підготовка даних у правильному форматі. Цей процес може включати завантаження даних, визначення відсутніх значень, фільтрацію часових рядів та інші завдання попередньої обробки. Функціональність, яку надає `tsibble` та інші пакети в `tidyverse`, значно спрощує цей крок.

Багато моделей мають різні вимоги до даних; деякі вимагають, щоб ряди були в порядку часу, інші вимагають відсутності пропущених значень. Перевірка ваших даних є важливим кроком для розуміння їх характеристик, і її слід завжди робити перед оцінкою моделей.

Дані часових рядів можуть демонструвати різноманітні моделі, і часто корисно розділити часовий ряд на кілька компонентів, кожен з яких представляє основну категорію шаблону.

Коли ми розкладаємо часовий ряд на компоненти, ми зазвичай об'єднуємо тренд і цикл в один компонент тренд-цикл (часто для спрощення його називають трендом). Таким чином, ми можемо розглядати часовий ряд як такий, що складається з трьох компонентів: компонент трендового циклу, сезонний компонент і компонент залишку (що містить будь-що інше в часовому ряді).

Наступним кроком є утворення ACF та PACF. Коли дані ACF показують стаціонарність, а PACF майже повністю відповідають вигляду білого шуму – можна підбирати моделі. На цьому етапі і відображається зміна підходу. Засоби R дозволяють утворювати моделі автоматичним чином трьома методами: швидкий підхід, підхід повного перебору та нормований закритий підхід. Підхід повного перебору дозволяє обрати оптимальну модель за показниками AIC та BIC, але потребує потужний комп'ютер. Інші підходи дають гірші результати, тому ми фокусуємось на підході повного перебору, як на очікуваному результаті, а порівнюватимемо з швидким підходом, оскільки він також створений для роботи на слабких комп'ютерах. Модифікований алгоритм дозволяє отримати той самий результат, що й повний перебір, але витрачаючи менші ресурси. Розглянемо сам алгоритм.

### **Модифікований алгоритм ARIMA**

Крок 1. Визначити закономірності, побудувавши дані та виявивши незвичайні спостереження.

Крок 2. Використати перетворення Бокса-Кокса, щоб стабілізувати дисперсію (якщо необхідно).

Крок 3. Диференціювати дані, щоб вони були стаціонарними (за потреби).

Крок 4. Побудувати графік ACF диференційованих даних і спробувати визначити можливі моделі-кандидати.

Крок 5. Використати метрику AIC для визначення кращої моделі.

Крок 6. Перевірте залишки з моделі, побудувавши графік ACF залишків і виконавши тест портманто.

Крок 7. Чи виглядають залишки як білий шум? Якщо ТАК (розрахувати прогнози), НІ (перейти до кроку 4).

Використання цього алгоритму дозволяє отримувати найкращу модель ARIMA заощаджуючи ресурси комп'ютеру. Підключивши додатково інші моделі, наприклад GAM, отримуємо систему прогнозування, що обиратиме найкращу модель під відповідний набір даних. На основі попередніх досліджень було визначено, що завдяки декомпозиції часового ряду, що була виконана раніше, вдається досить точно визначити параметри моделей GAM, які і будуть використані у системному підході, якщо це необхідно.

Протестувавши запропонований підхід на різних наборах даних, було визначено, що його використання дозволяє покращити точність прогнозування приблизно на 2% у порівнянні з готовим рішенням у пакеті R.

Таблиця 1 – Порівняння результатів похибки прогнозування для стандартного та запропонованого рішення на різних наборах даних

Підхід	Набір даних				
	1	2	3	4	5
	Похибка прогнозування за метрикою MAPE, %				
Модифікований алгоритм	3,29	1,98	4,37	1,57	3,84
Стандартне рішення від R	5,27	4,02	6,11	3,64	5,95

### Л і т е р а т у р а

1. Othman M. S., Ghadeer J. M. M. A modified ARIMA model for forecasting chemical sales in the USA. *Journal of Physics: Conference Series*. 2021. doi:10.1088/1742-6596/1879/3/032008.
2. Alsharif M. H., Younes M. K., Kim J. Time Series ARIMA Model for Prediction of Daily and Monthly Average Global Solar Radiation. *Symmetry*. 2019. 11(2):240. <https://doi.org/10.3390/sym11020240>.
3. Nashirah A. B., Sofian R. Autoregressive Integrated Moving Average (ARIMA) Model for Forecasting Cryptocurrency Exchange Rate in High Volatility Environment: A New Insight of Bitcoin Transaction. *International Journal of Advanced Engineering Research and Science (IJAERS)*. 2017. Vol. 4. DOI:10.22161/ijaers.4.11.20.
4. Bata, M., Carriveau, R., Ting, D.SK. Short-term water demand forecasting using hybrid supervised and unsupervised machine learning model. *Smart Water* 5. 2020. <https://doi.org/10.1186/s40713-020-00020-y>.
5. Durdu Ö. F. A hybrid neural network and ARIMA model for water quality time series prediction. *Engineering Applications of Artificial Intelligence*. 2010. Vol. 23. 586-594. <https://doi.org/10.1016/j.engappai.2009.09.015>.