

МОДЕЛЬ ARIMA: ТЕОРІЯ ТА ОСОБЛИВОСТІ*Гордій Я.В.*

Модель **ARIMA** (Autoregressive Integrated Moving Average) – це статистична модель часових рядів, яка поєднує авторегресію (AR), інтегрування (I) та ковзне середнє (MA) для прогнозування майбутніх значень ряду [1]. Формально ARIMA позначається як $ARIMA(p, d, q)$, де p – порядок авторегресії, d – порядок різницювання (для досягнення стаціонарності ряду), q – порядок компонентного ковзного середнього. Авторегресійний компонент $AR(p)$ моделює залежність поточного значення від кількох попередніх (лагів), інтегрування порядку d означає використання різниць значень ряду (усунення тренду), а компонент ковзного середнього $MA(q)$ враховує вплив попередніх випадкових шумів (помилки) на поточне значення.

Такий підхід дозволяє ARIMA-моделі описувати широкий спектр поведінки часових рядів – від стаціонарних коливань навколо середнього до рядів зі зростаючим трендом або сезонністю (при розширенні до SARIMA). Наприклад, у разі сезонних коливань додаються сезонні параметри (SARIMA), що враховують повторювані патерни з певним періодом.

У аналітичних задачах модель ARIMA є однією з базових завдяки відносній простоті та інтерпретованості. ARIMA добре описує *лінійні* залежності в часі і часто слугує відправною точкою перед застосуванням більш складних методів.

Застосування ARIMA в аналітичних IT-системах

Інтелектуальні аналітичні платформи обробляють різнотипні часові ряди: не лише фінансові показники чи продажі, а й технічні метрики – навантаження на сервери, трафік мережі, кількість запитів до веб-служб, споживання ресурсів центрів обчислень тощо. Прогнозування таких показників дозволяє системам завчасно реагувати на можливі пікові навантаження, оптимізувати розподіл ресурсів і виявляти аномальні відхилення. Модель ARIMA успішно використовується в подібних задачах як компонент систем підтримки рішень. Зокрема, в дослідженнях з управління хмарними серверами ARIMA-модель застосовується для прогнозу використання CPU та пам'яті, виділяючи *лінійні* закономірності в цих метриках [2].

Переваги ARIMA в IT-аналітиці:

- **Універсальність:** модель не потребує великих обсягів навчальних даних і підходить для різних доменів
- **Швидкість:** оцінювання параметрів ARIMA та формування прогнозу виконується відносно швидко, що важливо для систем реального часу.
- **Інтерпретованість:** внесок AR, I, MA компонентів легко трактувати, що дає розуміння характеру поведінки показника.

Обмеження: ARIMA припускає лінійність та стаціонарність ряду (після різницювання). В реальності метрики можуть мати складні нелінійні патерни або зміни структури з часом. У таких випадках ARIMA може давати систематичну похибку, і доцільно розглянути її розширення (напр. SARIMA для сезонних ефектів) чи комбінувати з іншими підходами (напр. нейронними мережами для нелінійних залишків). Тим не менш, багато задач оперативного аналізу успішно вирішуються саме ARIMA-моделями, особливо коли йдеться про короткострокові прогнози на основі останніх тенденцій.

Приклад прогнозування різнотипних показників

Для ілюстрації застосування ARIMA обрано лише **один, але показовий** часовий ряд - **HTTP-трафік `elb_request_count`** (кількість запитів за 5 хв) із відкритого набору **Numenta Anomaly Benchmark (NAB)** [3]. Ряд має чітку добову циклічність та поодинокі аномальні сплески.

Етапи побудови моделі

1. **Попередній аналіз** – ADF-тест і огляд графіка виявили добову сезонність, тому застосовано сезонне різницювання з лагом 288 (1440 хв(доба) / 5 хв).
2. **Вибір параметрів (p,d,q)** – за графіками ACF/PACF та критерієм AIC підібрано **ARIMA (2, 1, 2)**.
3. **Навчання** – модель навчена на 2-добовій історії (576 відліків).
4. **Прогноз** – сформовано 24-годинний прогноз із 95 % довірчим інтервалом.



Рис. 1 метрика **elb_request_count**: фактичні значення за 2 доби (5-хв інтервал) та добовий прогноз ARIMA(2,1,2)

Модель достовірно відтворює добовий цикл і прогнозує денне зростання навантаження; 95 % інтервал охоплює основні коливання. Ключові метрики на тренувальному відрізку:

- **MAE (Середня абсолютна похибка) ≈ 210** запитів (приблизно рівний середньому рівню трафіку ≈ 210 запитів / 5 хв);
- **MAPE (Середня відносна похибка) $\approx 150\%$** — висока через кілька різких піків (наприклад, сплеск о 21:54), які лінійна ARIMA схильна недооцінювати.

ARIMA (2, 1, 2) забезпечує прийнятну точність для типових коливань трафіку й може слугувати базовим методом прогнозування, доповненим спеціалізованими модулями для пікових навантажень. Така гібридна схема одночасно мінімізує середню похибку й підвищує стійкість системи до рідкісних, але критичних аномалій.

Незважаючи на лінійність і вимогу стаціонарності, ARIMA залишається швидким та інтерпретованим інструментом короткострокового прогнозування у «розумних» ІТ-системах. На прикладі HTTP-трафіку вона підтвердила здатність точно відтворювати регулярні патерни і надавати корисні прогнози для проактивного масштабування ресурсів. У випадку ж аномальних сплесків ARIMA доцільно комбінувати з методами виявлення аномалій або нелінійними моделями, що забезпечує комплексну підтримку аналітичних рішень.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Thomas Vincent, Lisa Tagliaferri (2017). *A Guide to Time Series Forecasting with ARIMA in Python* 3. DigitalOcean. Режим доступу: <https://www.digitalocean.com/community/tutorials/a-guide-to-time-series-forecasting-with-arima-in-python-3>
2. K. Lalitha Devi, S. Valli (2023). *Time series-based workload prediction using the statistical hybrid model for the cloud environment*. **Computing**, 105:353–374. Режим доступу: <https://doi.org/10.1016/j.neucom.2017.04.070>
3. Ahmad S., Lavin A. et al. (2017). *Unsupervised real-time anomaly detection for streaming data*. **Neurocomputing**, 262:134–147. (Дані Numenta Anomaly Benchmark). Режим доступу: <https://github.com/numenta/NAB>

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**



**ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ
АСПЕКТИ РОЗРОБКИ
КОМП'ЮТЕРНИХ СИСТЕМ**

**ЗБІРНИК НАУКОВИХ ПРАЦЬ ЗА МАТЕРІАЛАМИ
VII ВСЕУКРАЇНСЬКОЇ НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ
СТУДЕНТІВ І АСПІРАНТІВ
*24 квітня 2025 року***

Київ 2025

УДК 004

Відповідальна за випуск: М.І. Лендел

Збірник наукових праць за матеріалами VII Всеукраїнської науково-практичної конференції студентів і аспірантів «ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ АСПЕКТИ РОЗРОБКИ КОМП'ЮТЕРНИХ СИСТЕМ '2025», 24 квітня 2025 року, НУБіП України, Київ. – 452 с. (електронне видання)

Відповідальність за зміст публікацій несуть автори.

Передрук матеріалів, а також використання їх будь-якій формі допускається лише з дозволу авторів