

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМ. І. СІКОРСЬКОГО»
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ**

Моделі для прогнозування характеристик трафіка цифрової реклами

Виконав: студент групи КА-61 м

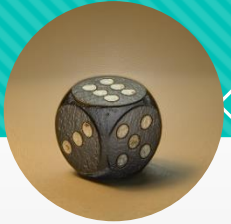
Логін В. В.

Науковий керівник:

д.т.н., професор Бідюк П. І.

Київ – 2018

Структурні елементи дослідження



Об'єкт дослідження: трафік цифрової реклами у формі статистичних даних.



Предмет дослідження: моделі та методи аналізу даних у формі часових рядів, методи прикладної статистики.



Мета роботи: побудова моделей часових рядів для прогнозування найважливіших характеристик трафіка цифрової реклами.



Методи дослідження: моделі часових рядів для прогнозування даних та порівняльний аналіз отриманих моделей.

Актуальність роботи

3

**Прогнозування
характеристик
рекламного
трафіка потрібне
для:**

- Моніторингу поточного стану рекламного трафіка;
- Оптимізації витрат;
- Проведення планування.

**Існуючі
проблеми:**

- Велика кількість метрик;
- Значна розрізненість даних в межах трафіка цифрової реклами.

Змістовне формулювання задачі

Деяка рекламна компанія здійснює посередницьку діяльність між рекламодавцями та видавцями. При цьому генерується значний об'єм трафіка цифрової реклами у формі статистичних даних.



За цими даними потрібно побудувати моделі для прогнозування ключових характеристик рекламного трафіка, а також провести порівняльний аналіз отриманих результатів. Це дасть змогу здійснити оптимізацію витрат у майбутньому.

Постановка задачі дослідження

5

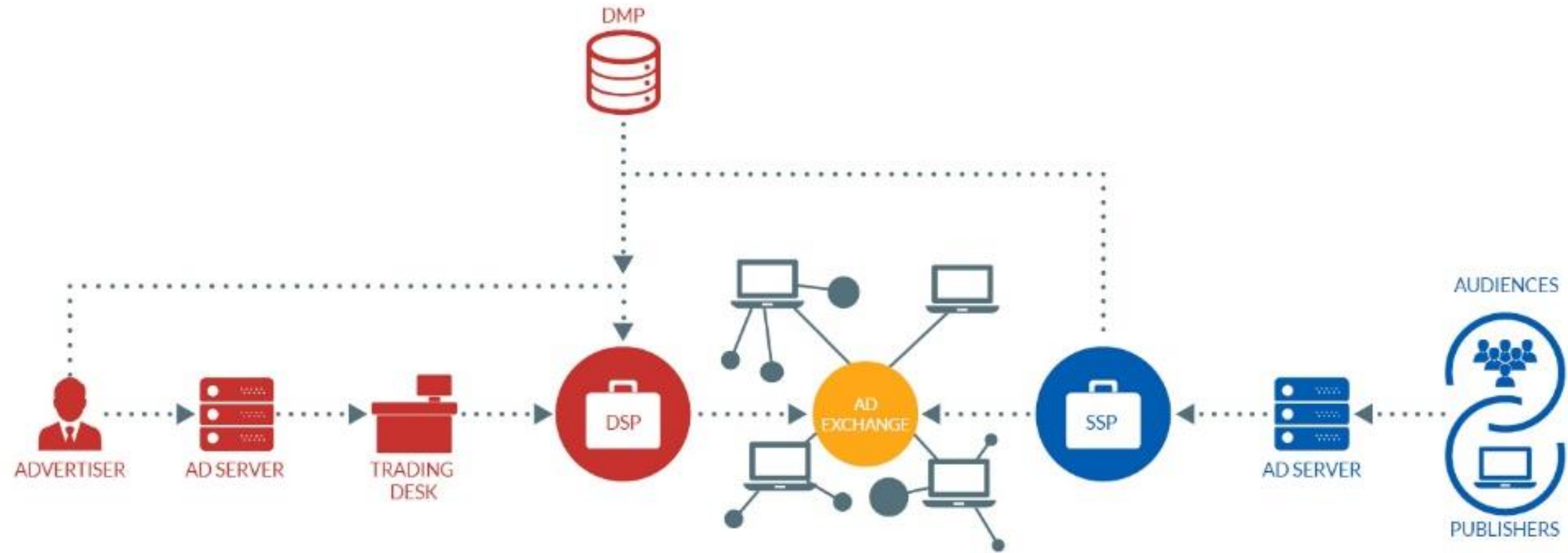
Виконати загальний огляд моделей часових рядів та вибрати ті, які відповідають поставленій задачі прогнозування характеристик трафіка цифрової реклами.

Побудувати моделі за допомогою обраних підходів з використанням статистичних даних.

Порівняти отримані моделі та вибрати найкращу.

Виробити рекомендації щодо вдосконалення у майбутньому розроблених підходів до розв'язання даної задачі.

Структура ринку цифрової реклами



Боротьба з невизначеностями у моделюванні за статистичними даними

Можливі невизначеності

- вплив зовнішніх випадкових збурень;
- вплив похибок вимірів;
- можливі пропуски вимірів;
- короткі вибірки.

Методи боротьби

- використання методів заповнення пропусків;
- використання цифрової та оптимальної фільтрації для боротьби з випадковими збуреннями та похибками вимірів.

Моделі, що використані в роботі

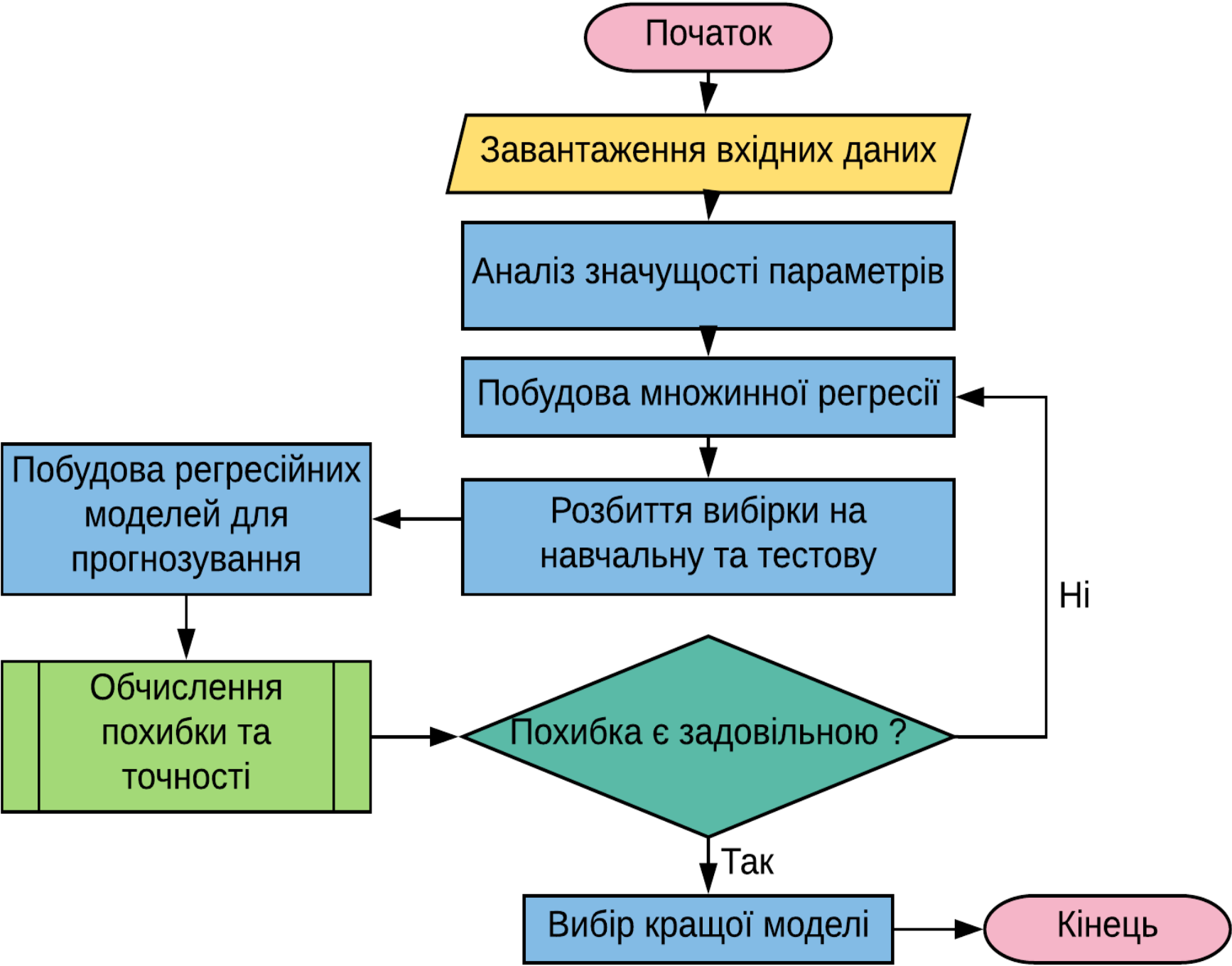
- AR
- ARMA
- ARIMA
- ARIMAX



Програмне забезпечення, використане в роботі



Блок-схема
розв'язання
задачі



Рівняння моделей прогнозу для доходу

- AR: $y(k) = 4234.48 + 0.85 \cdot y(k - 1)$
- ARMA: $y(k) = 4234.48 + 0.85 \cdot y(k - 1)$
- ARIMA: $y(k) = 1.25 \cdot y(k - 1) + 0.92 \cdot y(k - 2) - 1.48 \cdot v(k - 1) + 1.09 \cdot v(k - 2) - 0.19 \cdot v(k - 3)$
- ARIMAX: $y(k) = 1.25 \cdot y(k - 1) + 0.92 \cdot y(k - 2) - 1.48 \cdot v(k - 1) + 1.09 \cdot v(k - 2) - 0.19 \cdot v(k - 3) + 9.8 \cdot 10^{-3} \cdot x_1 + 9.15 \cdot 10^3 \cdot x_2 + 9.93 \cdot 10^4 \cdot x_3$

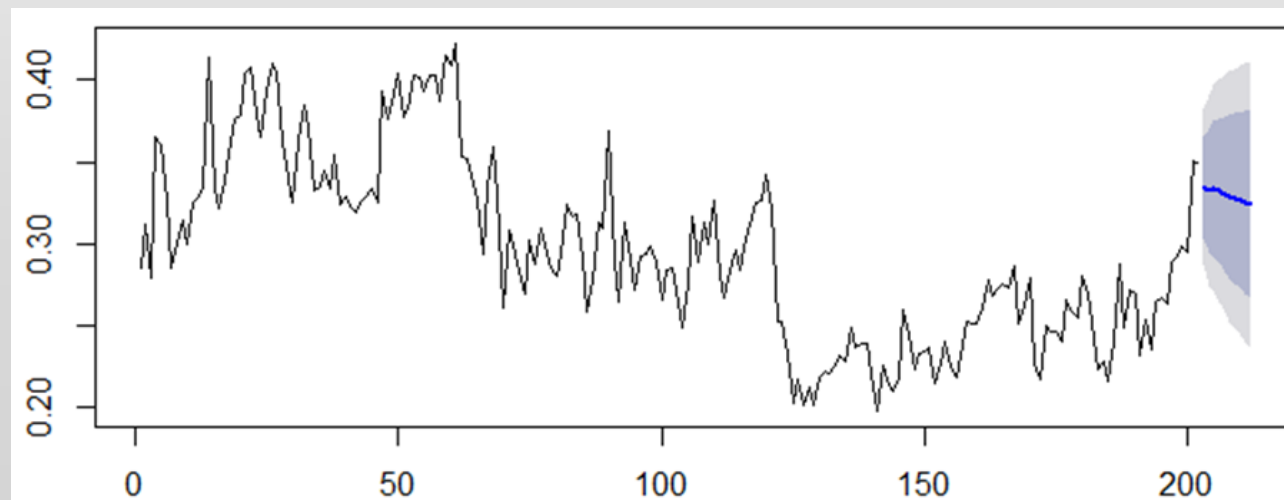
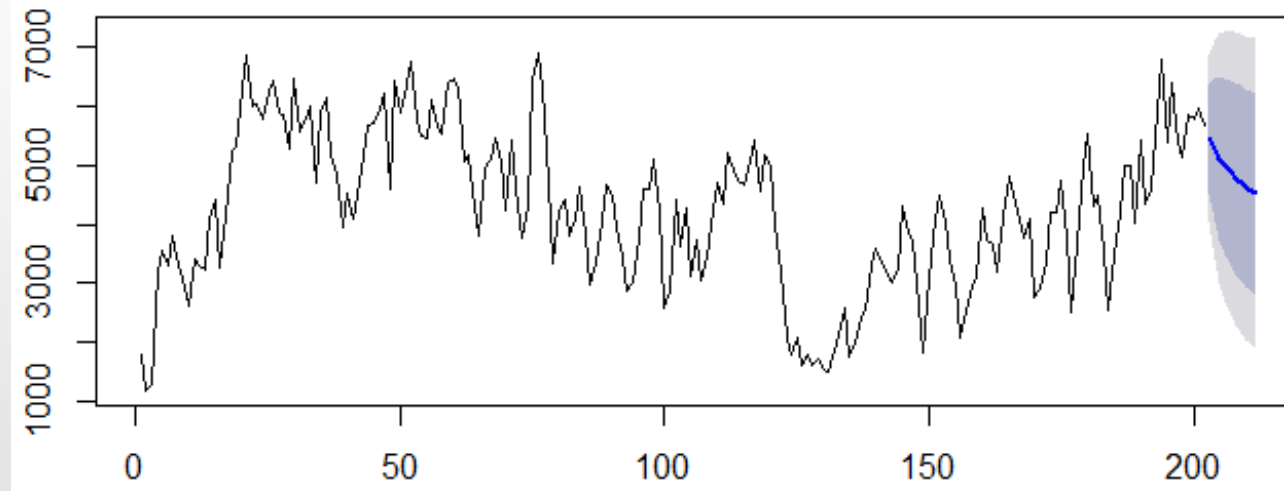
Рівняння моделей прогнозу для показника заповнюваності

- AR: $y(k) = 0.29 + 0.7 \cdot y(k - 1) + 0.012 \cdot y(k - 2) + 0.15 \cdot y(k - 3)$
- ARMA: $y(k) = 0.3 + 0.029 \cdot y(k - 1) + 0.9 \cdot y(k - 2) + 0.68 \cdot v(k - 1) - 0.4 \cdot v(k - 2) - 0.18 \cdot v(k - 3)$
- ARIMA: $y(k) = 1.13 \cdot y(k - 1) + 0.66 \cdot v(k - 1) - 0.44 \cdot v(k - 2) - 0.2 \cdot v(k - 3)$
- ARIMAX: $y(k) = 1.13 \cdot y(k - 1) + 0.66 \cdot v(k - 1) - 0.44 \cdot v(k - 2) - 0.2 \cdot v(k - 3) + 1.18 \cdot 10^{-5} \cdot x_1 + 4.1 \cdot 10^{-9} \cdot x_2 - 1.2 \cdot 10^{-7} \cdot x_3$

Результати моделювання

13

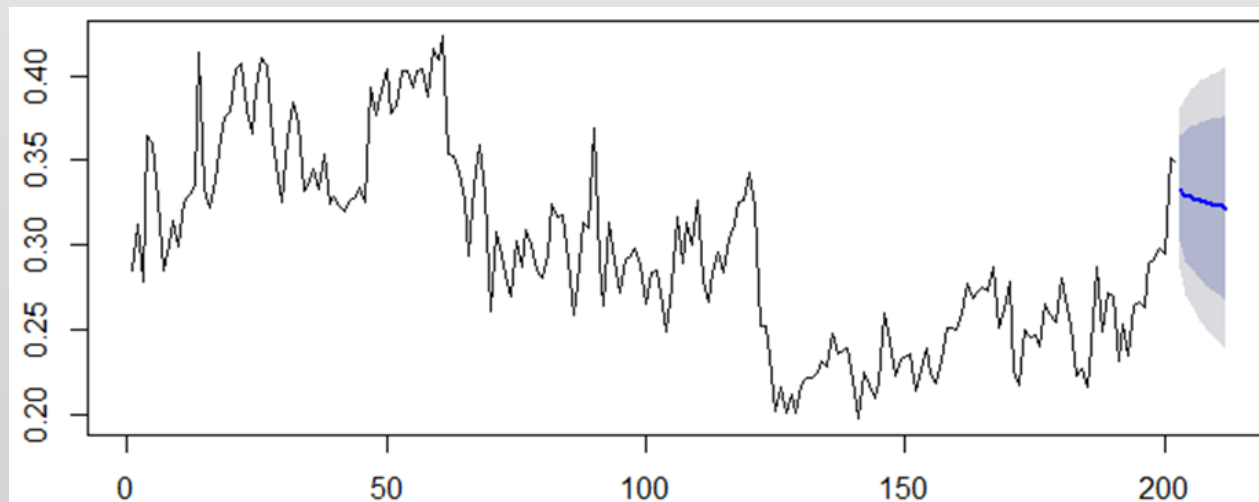
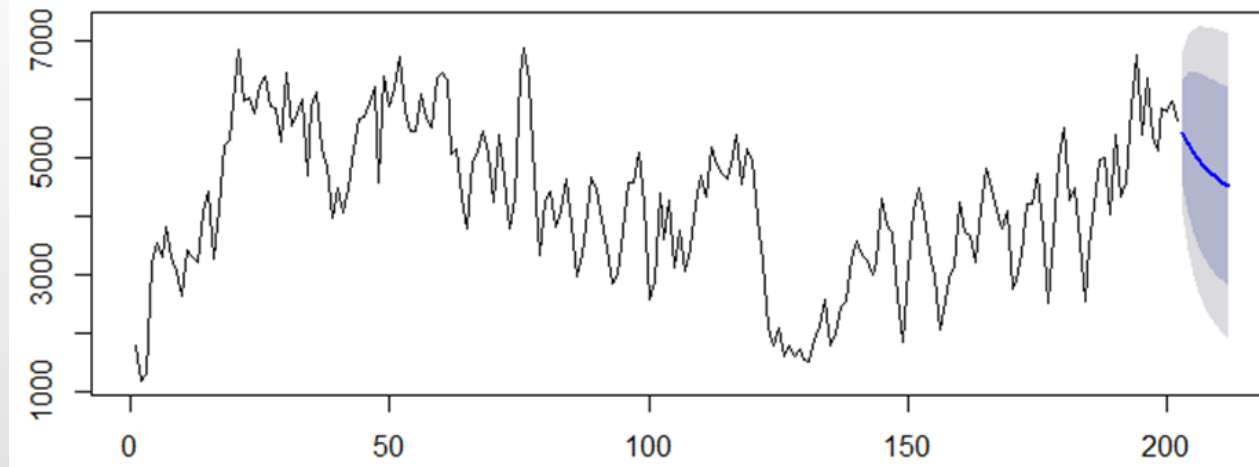
Прогноз на тестовій вибірці за моделлю AR для доходу та показника заповнюваності відповідно



Результати моделювання

14

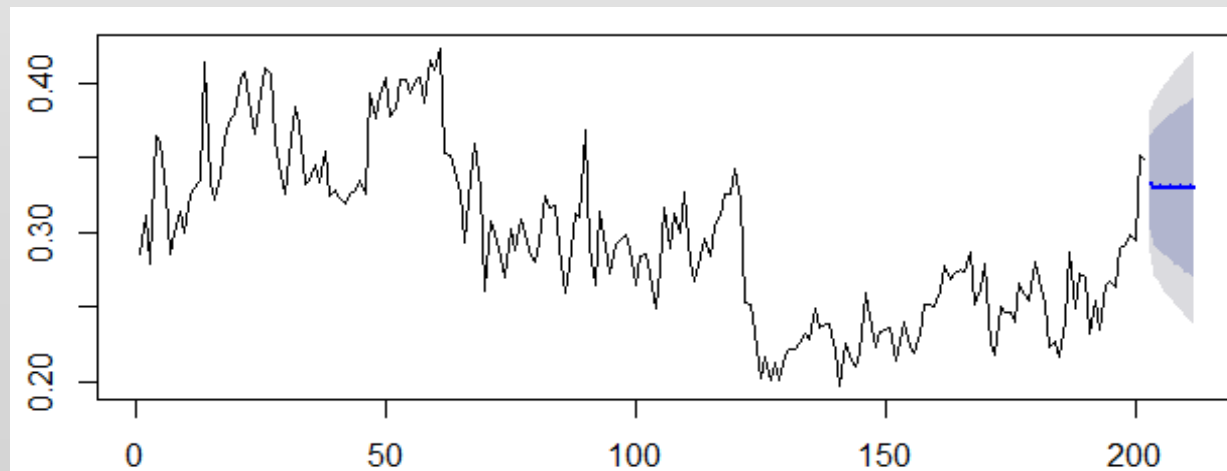
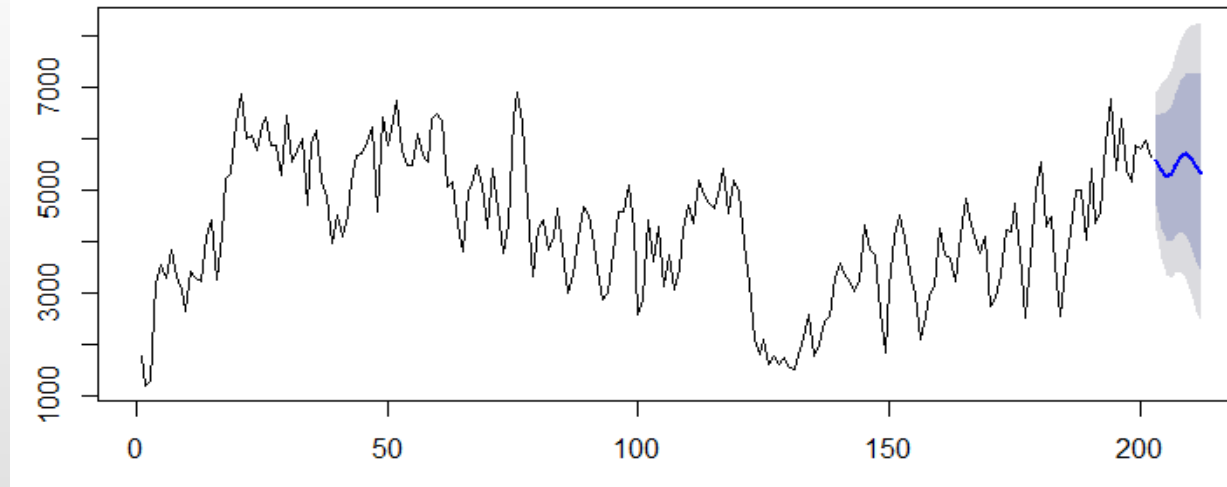
Прогноз на тестовій вибірці за моделлю ARMA для доходу та показника заповнюваності відповідно



Результати моделювання

15

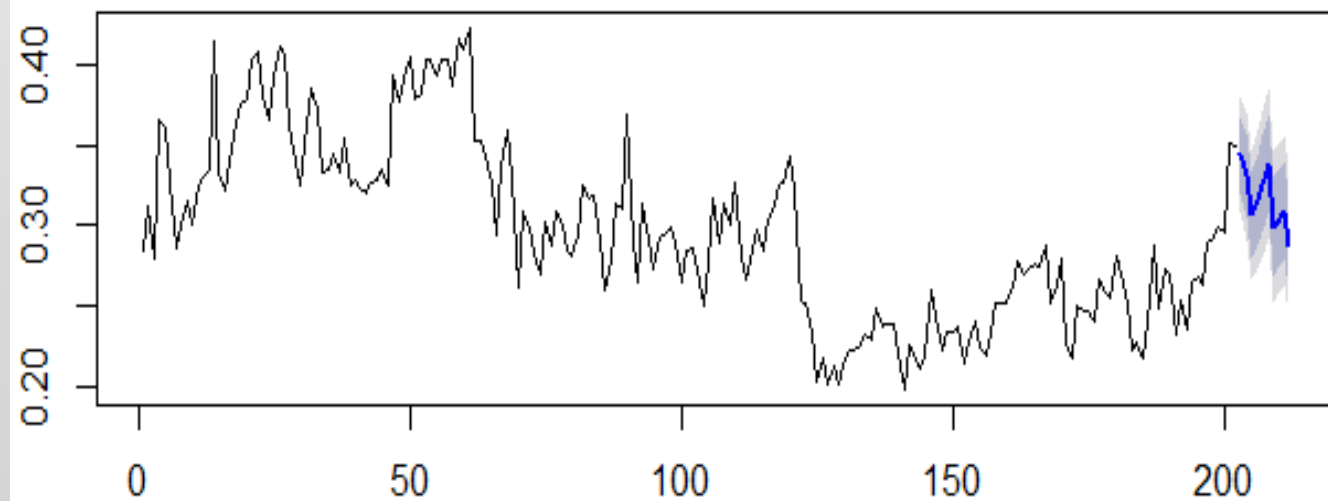
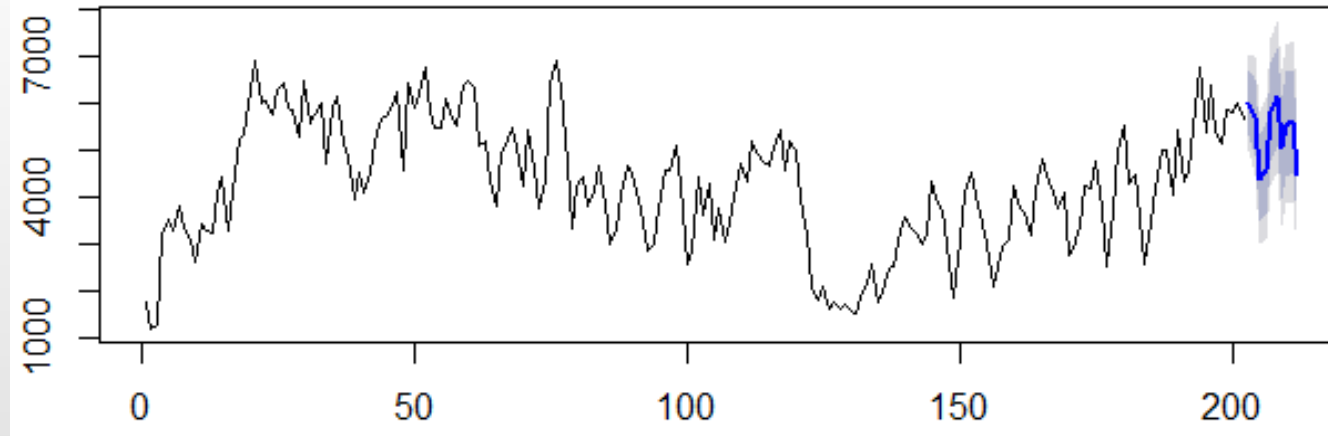
Прогноз на тестовій вибірці за моделлю ARIMA для доходу та показника заповнюваності відповідно



Результати моделювання

16

Прогноз на тестовій вибірці за моделлю ARIMAX для доходу та показника заповнюваності відповідно



Вибір кращої моделі

17

Таблиця 1 – Порівняння побудованих моделей для доходу за критеріями

Model	DW	R^2	AIC	BIC
AR	1.981992	0.714015	3233.09	3243.01
ARMA	1.981992	0.714015	3233.09	3243.01
ARIMA	1.981691	0.7522077	3202.2	3222.01
ARIMAX	2.007797	0.8531828	3118.26	3161.27

Вибір кращої моделі

18

Таблиця 2 – Порівняння побудованих моделей для доходу за похибками

Model	RMSE	MAE	MAPE	MASE
AR	710.3305	587.1062	15.69017	0.9687749
ARMA	710.3305	587.1062	15.69017	0.9687749
ARIMA	673.1706	536.188	13.8156	0.8847556
ARIMAX	509.021	373.7665	9.315525	0.6167463

Вибір кращої моделі

19

Таблиця 3 – Порівняння побудованих моделей для заповнюваності за критеріями

Model	DW	R^2	AIC	BIC
AR	1.993624	0.8180255	-924.14	-904.29
ARMA	2.003657	0.8186622	-922.83	-899.68
ARIMA	2.006298	0.818524	-919.87	-903.35
ARIMAX	2.001912	0.9000842	-1025.5	-972.56

Вибір кращої моделі

20

Таблиця 4 – Порівняння побудованих моделей
для заповнюваності за похибками

Model	RMSE	MAE	MAPE	MASE
AR	0.02374302	0.01780827	6.044972	0.9460523
ARMA	0.02370153	0.0177796	6.045356	0.9445291
ARIMA	0.02387146	0.01785158	6.081153	0.9483528
ARIMAX	0.01760031	0.01380587	4.715555	0.7334275

Вибір кращої моделі

Із наведених на попередніх слайдах таблиць видно, що найкращою моделлю для прогнозування характеристик рекламного трафіка є модель ARIMAX.

За результатами досліджень видно, що у даної моделі найкращі показники як за критеріями, так і за похибками.

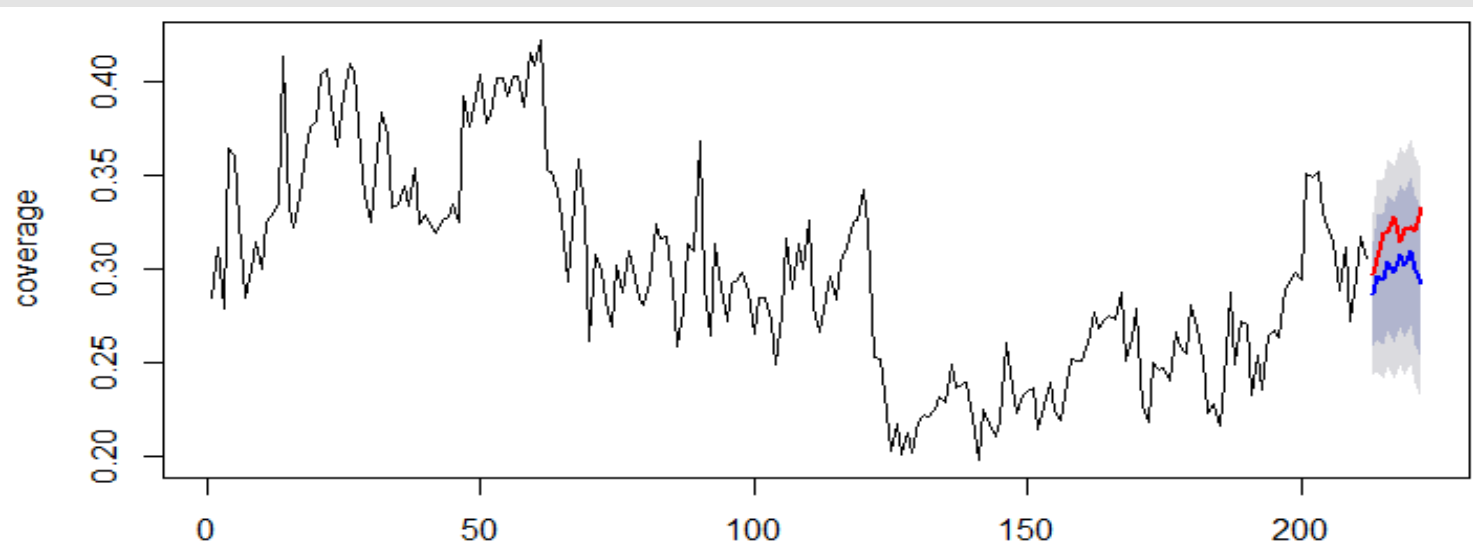
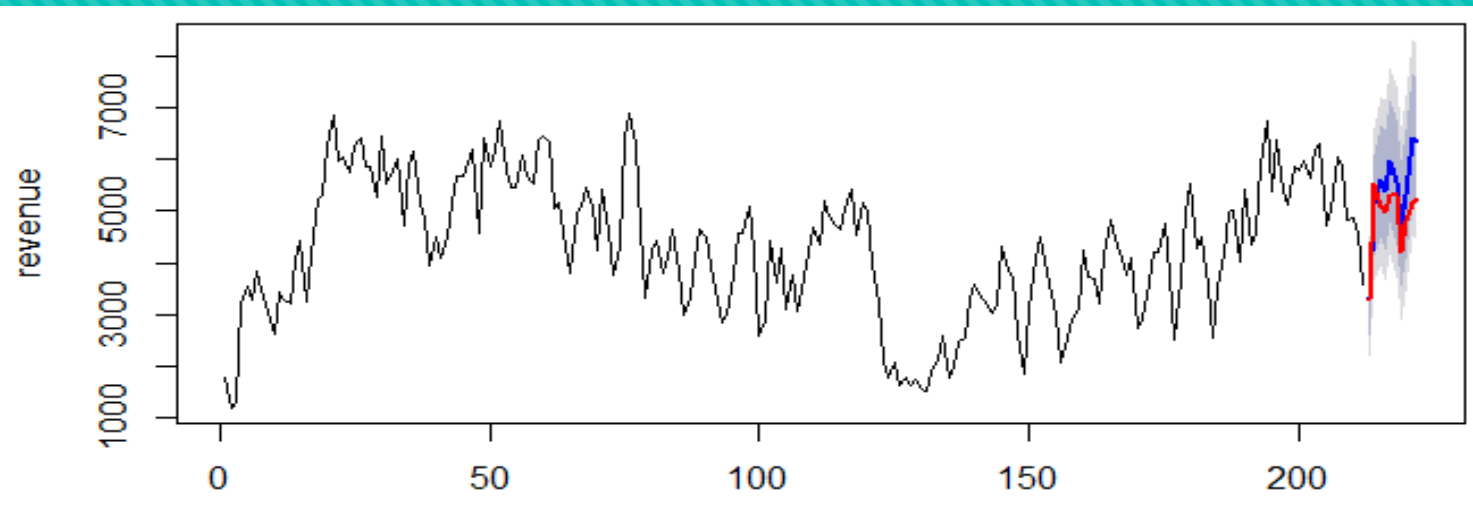
Якщо розглянути найважливіший вид похибок – MAPE, то можна зробити наступний висновок.
 $MAPE(ARIMAX)_{revenue} = 9.31 < 10$ та

$MAPE(ARIMAX)_{coverage} = 4.71 < 10$, тобто для обох показників похибка MAPE менша за 10%, що свідчить про високу точність моделей.



Прогноз за побудованими моделями

22



Наукова новизна

23



Побудовано нові моделі для поставленої задачі прогнозування характеристик трафіка цифрової реклами:

- AR;
- ARMA;
- ARIMA;
- ARIMAX.



При здійсненні прогнозування ключових характеристик трафіка цифрової реклами пропонується використовувати модель авторегресії та інтегрованого ковзного середнього з екзогенними змінними.



Побудовано моделі для прогнозування ключових характеристик трафіка цифрової реклами.



Побудовані моделі застосовано до набору реальних статистичних даних рекламної компанії-посередника, та проведений порівняльний аналіз побудованих 4 моделей:

- AR;
- ARMA;
- ARIMA;
- ARIMAX.



Рекомендовано використовувати, для подальшої побудови прогнозів, модель авторегресії та інтегрованого ковзного середнього з екзогенними змінними.

Перспективи для подальших досліджень

25

Побудова
нових
моделей для
прогнозування
характеристик
рекламного
трафіка.

Подальше
вдосконалення
побудованих
моделей з
метою
покращення
їхньої якості.

Дослідження
окремих
найбільш
значущих
сайтів із
трафіка
цифрової
реклами.

Наукові публікації

26



Наукова стаття «Models for forecasting parameters of digital advertising traffic», що буде опублікована у видавництві CEUR.



Тези «Моделі для прогнозування характеристик трафіка цифрової», що будуть опубліковані у збірнику тез конференції CAIT-2018

Впровадження результатів



Результати наукових досліджень з теми магістерської дисертації були використані у діяльності рекламної компанії «MEDIAWAYSS Sp. z o. o.»

Дякую за увагу !