

**Київський національний університет імені Тараса Шевченка**

**Економічний факультет**

**Кафедра економічної кібернетики**

**КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА**

**Прогнозування ремаркетингових кампаній з використанням  
BigQuery ML**

студента 4 курсу

спеціальності 051 «Економіка»

ОПП «Економічна кібернетика»

денної форми навчання

Кузнецова Нікити Сергійовича

Науковий керівник:

доктор економічних наук, професор

**Чорноус Галина Олександрівна**

Засвідчую, що в цій роботі немає запозичень із праць інших авторів без  
відповідних посилань

Студент \_\_\_\_\_

Роботу допущено до захисту перед ЕК  
рішенням кафедри економічної кібернетики  
від 12 червня 2025 р.,  
протокол № 15

**Завідувач кафедри:**

доктор економічних наук, професор

Ляшенко Олена Ігорівна \_\_\_\_\_

КИЇВ – 2025

## РЕФЕРАТ

Кваліфікаційна робота бакалавра містить: 46 ст., 2 рис., 3 табл., 28 джерел

**Ключові слова:** прогнозування, ремаркетинг, BigQuery ML, ARIMA, часові ряди, LTV, машинне навчання, маркетингові кампанії, Genesis Company

**Об'єкт дослідження:** процес прогнозування доходів від ремаркетингових кампаній.

**Мета дослідження:** розробити ефективний метод прогнозування доходів від ремаркетингових кампаній для когорт користувачів на 365 день з використанням моделі ARIMA в BigQuery ML

**Методи дослідження:** теоретичний аналіз, статистичний аналіз, математичне моделювання, програмування, експериментальний аналіз

**Наукова новизна, теоретична значимість дослідження:** розробка нового методу прогнозування доходів від ремаркетингових кампаній з використанням моделі ARIMA в BigQuery ML, який враховує специфіку рекламних мереж і адаптується до змін у даних через автоматичне перенавчання

**Практична цінність:** створення інструменту для маркетингових відділів, який забезпечує підтримку у прийнятті стратегічних рішень щодо планування ремаркетингових кампаній на тривалий горизонт

## RESUME

Taras Shevchenko National University of Kyiv,

Faculty of Economics, Department of Economic Cybernetics

Key words: forecasting, remarketing, BigQuery ML, ARIMA, time series, LTV, machine learning, marketing campaigns, Genesis Company

The graduation research of student Nikita Kuznetsov deals with forecasting revenues from remarketing campaigns using BigQuery ML

The work is interesting for marketing departments, data analysts, and companies involved in digital marketing

Pages 46, tables 3, bibliog. 28

## ЗМІСТ

<b><i>ВСТУП</i></b> .....	<b>5</b>
<b><i>РОЗДІЛ 1. ТЕОРЕТИЧНІ ВІДОМОСТІ ПРО ПРОГНОЗУВАННЯ LTV В МАРКЕТИНГОВИХ КАМПАНІЯХ</i></b> .....	<b>8</b>
1.1. Розкриття поняття LTV у контексті економічної проблематики .....	8
1.2. Методи прогнозування LTV.....	10
1.3. Виклики ремаркетингу та потреби в нових, адаптивних підходах.....	12
<b><i>РОЗДІЛ 2. ПОБУДОВА МОДЕЛІ ARIMA PLUS В СЕРЕДОВИЩІ BIGQUERY ML</i></b> .....	<b>16</b>
2.1. Підготовка даних та обробка викидів та аномалій.....	16
2.2. Коригування ефектів та виявлення та екстраполяція сезонності .....	19
<b><i>РОЗДІЛ 3. ПРОГНОЗУВАННЯ РЕМАРКЕТИНГОВИХ КАМПАНІЙ НА ПРИКЛАДІ ПІДПРИЄМСТВА “GENESIS COMPANY”</i></b> .....	<b>22</b>
3.1. Створення таблиці маркетингових каналів по відповідних платформах 23	23
3.2. Формування таблиць транзакцій для ремаркетингу для відповідних платформ.....	24
3.3. Динамічне створення, прогнозування та моніторинг ARIMA-моделей	29
3.4. Оптимізація ARIMA моделей при створенні прогнозів.....	33
<b><i>ВИСНОВКИ</i></b> .....	<b>38</b>
<b><i>СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ</i></b> .....	<b>42</b>

## ВСТУП

У сучасному цифровому світі ремаркетингові кампанії стали невід'ємною частиною маркетингових стратегій, спрямованих на підвищення лояльності клієнтів та максимізацію доходів від існуючих користувачів. Ремаркетинг дозволяє залучати тих, хто вже взаємодіяв із продуктом, через персоналізовані пропозиції та рекламу, що базується на попередній поведінці. Проте ключовим фактором успіху таких кампаній є ефективна оцінка повторного залучення трафіку та точність прогнозування поведінки користувачів, зокрема їхньої реакції на маркетингові стимули протягом тривалого часу, наприклад, 365 днів після встановлення додатку. Основна проблематика полягає в тому, що для нових когорт користувачів, визначених за датою інсталу, часто бракує історичних даних, що ускладнює прогнозування тенденцій доходів.

Традиційні методи прогнозування LTV часто виявляються недостатньо ефективними за відсутності достатньої кількості даних. Наприклад, наявний метод прогнозування за допомогою ланцюгових коефіцієнтів до 180 дня втрачає точність, якщо відсутні достатні дані для аналізу, як у нашому випадку. Це підкреслює потребу в інноваційних підходах, які здатні забезпечити прогнозовану оцінку трафіку, враховуючи динаміку виторгу когорт і дозволяючи прогнозувати на тривалий горизонт планування.

У цьому контексті використання BigQuery ML, хмарної платформи Google для машинного навчання, відкриває нові можливості. BigQuery ML дозволяє створювати та навчати моделі, зокрема ARIMA, безпосередньо в базі даних за допомогою SQL-запитів, що спрощує інтеграцію аналітики та прогнозування. Модель ARIMA (Autoregressive Integrated Moving Average) є потужним інструментом для роботи з часовими рядами, оскільки враховує історичні коефіцієнти та здатна адаптуватися до трендів і сезонності.

Практична доцільність цього дослідження полягає в розробці інструменту, який допоможе бізнесу оптимізувати маркетингові бюджети та

підвищити ROI (Return on Investment). Наукова актуальність обумовлена недостатньою вивченістю методів прогнозування часових рядів у контексті ремаркетингу. Відмінність пропонованого підходу від відомих полягає в інтеграції моделі ARIMA в BigQuery ML із механізмом автоматичного перенавчання за допомогою SQL у разі виявлення аномально поганих показників прогнозу, що забезпечує адаптивність і точність.

**Об’єкт та предмет дослідження.** Об’єктом нашого дослідження є процес прогнозування продажів маркетингових кампаній, а предметом – методи та алгоритми прогнозування LTV, які використовують часові ряди у свої технічній основі на базі хмарного середовища BigQuery ML.

**Мета дослідження.** Розробити ефективний метод прогнозування доходів від ремаркетингових кампаній для когорт користувачів на 365 день їхнього “життя” на продукті із використанням моделі ARIMA в BigQuery ML.

**Завдання дослідження:**

1. Вивчити наявний метод прогнозування ремаркетингових кампаній до 180 дня “життя” когорти.
2. Проаналізувати можливість застосування прогнозних часових рядів, зокрема моделі ARIMA, в контексті маркетингових задач.
3. Виявити параметри за допомогою, яких можна визначати ефективність моделі ARIMA PLUS в хмарному середовищі BigQuery ML.
4. Розробити системний алгоритм автоматичного перенавчання моделі у BigQuery у разі виявлення аномально поганих показників.
5. Оцінити ефективність розробленого методу на реальних даних і порівняти його з існуючими підходами.

**Методи дослідження.** Для виконання дослідження були застосовані такі методи, як: теоретичний аналіз, статистичний аналіз, економіко-математичне моделювання, програмування та експериментальний аналіз.

**Ступінь новизни роботи.** Наукова новизна полягає в розробці нового

методу прогнозування доходів від ремаркетингових кампаній із використанням моделі ARIMA в BigQuery ML, який враховує специфіку рекламних мереж і адаптується до змін у даних через автоматичне перенавчання. А практична новизна полягає в створенні інструменту для маркетингових відділів, який забезпечує підтримку у прийнятті стратегічних рішень щодо планування ремаркетингових кампаній на тривалий горизонт, що перевищує обсяг доступних історичних даних. Відмінність від відомих результатів полягає в інтеграції часових рядів і машинного навчання в єдиній платформі BigQuery ML, а також у впровадженні адаптивного механізму перенавчання, що не реалізовувався раніше.

**Джерела інформації.** Основними джерела для даного дослідження слугують внутрішні дані компанії та документація BigQuery ML підготовлена компанією Google.

**Структура роботи.** Ця дослідницька робота складається з вступу, трьох розділів, висновків та списку використаних джерел. Кожен розділ містить всебічний аналіз теми дослідження та сприяє загальному розумінню проблеми.

## РОЗДІЛ 1. ТЕОРЕТИЧНІ ВІДОМОСТІ ПРО ПРОГНОЗУВАННЯ LTV В МАРКЕТИНГОВИХ КАМПАНІЯХ

### 1.1. Розкриття поняття LTV у контексті економічної проблематики

У сучасному бізнес-середовищі конкуренція за увагу споживачів досягла безпрецедентних масштабів. Компанії змушені шукати нові способи оптимізації своїх маркетингових стратегій, щоб залишатися конкурентоспроможними. Одним із ключових інструментів у цьому процесі є Lifetime Value (LTV) – метрика, яка відображає загальний дохід, який компанія може отримати від клієнта протягом усього часу його взаємодії з продуктом чи послугою. LTV не просто показує прибутковість клієнта, а й слугує основою для стратегічного планування, розподілу маркетингових бюджетів, розробки програм лояльності та утримання клієнтів.

Прогнозування LTV на тривалий період, зокрема на 365 днів, є особливо актуальним у сфері ремаркетингу. Ремаркетинг передбачає повторне залучення клієнтів, які вже взаємодіяли з продуктом, через персоналізовані пропозиції та рекламу. Точне прогнозування LTV у цьому контексті дозволяє:

- Оптимізувати маркетингові бюджети. Компанії можуть заздалегідь планувати витрати, орієнтуючись на очікуваний дохід від клієнтів, і уникати надмірних інвестицій у кампанії з низьким потенціалом. Наприклад, якщо прогноз показує, що певна когорта клієнтів принесе лише 10% від загального доходу, компанія може перенаправити ресурси на більш прибуткові сегменти.
- Покращити утримання клієнтів. Розуміння довгострокової цінності клієнта допомагає розробляти стратегії для підвищення лояльності, такі як знижки чи персоналізовані пропозиції, що стимулюють повторні покупки.
- Підвищити ефективність стратегічного планування. Точні прогнози LTV дають можливість приймати рішення на основі даних, а не інтуїції, що є

критично важливим у нестабільних ринкових умовах, коли економічні фактори можуть швидко змінюватися.

Економічна проблема полягає в тому, що багато компаній досі використовують короткострокові підходи до оцінки ефективності маркетингових кампаній. Це призводить до недооцінки довгострокового потенціалу клієнтів. У ремаркетингу, де поведінка клієнтів є динамічною, а дати повернення платників можуть варіюватися від кількох днів до кількох місяців, відсутність точного прогнозу LTV на довгостроковий період може спричинити неефективний розподіл ресурсів і втрату прибутку. Наприклад, якщо компанія витрачає значні кошти на залучення нових клієнтів, не враховуючи, що існуючі клієнти з високим LTV можуть принести більший дохід за менших витрат, вона втрачає можливості для зростання.

Для точного обчислення та оптимізації LTV у контексті продуктів ІТ необхідно звернутися до юніт-економіки – аналізу фінансових метрик на рівні окремого користувача. Це дозволяє оцінити стійкість бізнес-моделі та зрозуміти, як ключові показники впливають на LTV. Основні метрики юніт-економіки включають:

- **Cost of Acquisition (CAC):** Витрати на залучення нового клієнта.
- **Average Revenue Per User (ARPU):** Середній дохід від одного користувача за певний період.
- **Retention Rate:** Відсоток користувачів, які продовжують взаємодіяти з продуктом.
- **Churn Rate:** Відсоток користувачів, які припиняють користуватися продуктом.

Ці метрики безпосередньо пов'язані з LTV. Наприклад, LTV можна розрахувати за формулою:

$$LTV = \frac{ARPU}{Churn\ rate}$$

Ця формула демонструє, як зростання ARPU або зниження Churn rate можуть суттєво збільшити LTV. Відповідно, навіть незначні зміни в цих метриках можуть мати масштабний вплив на загальну прибутковість бізнесу. Це підкреслює важливість постійного моніторингу ключових показників та вдосконалення користувацького досвіду для підвищення лояльності клієнтів. У результаті, точне розуміння структури LTV і його драйверів дає змогу ефективніше планувати маркетингові бюджети, підвищувати рентабельність інвестицій та забезпечувати сталий розвиток продукту.

Таким чином, потреба у прогнозуванні LTV на достатньо велику кількість днів (наприклад на 365 днів, як в нашому випадку) зумовлена необхідністю більш ґрунтовного підходу до прийняття рішень у маркетингу. Це вимагає розробки нових методів і підходів, які будуть адаптовані до специфіки ремаркетингу та враховуватимуть його унікальні особливості.

## 1.2. Методи прогнозування LTV

Існує декілька формул обрахунку LTV, але класична формула на якій базуються основні методи прогнозування має такий вигляд:

$$LTV = ARPU * Customer lifetime$$

де *ARPU* — середній дохід, який клієнт приносить за одиницю часу (наприклад, місяць або день); *Customer lifetime* — середня тривалість життя клієнта в періодах (наприклад, у місяцях або днях), тобто час, протягом якого клієнт залишається активним. Використовуючи цю формулу, можна доволі просто спрогнозувати прибуток, який клієнт принесе за обраний період життєвого циклу, за умови сталого ARPU.

Проте для прогнозування LTV на майбутні періоди ці базові формули можуть бути недостатніми — вони не враховують сезонність, зміну поведінки клієнтів, асиметрію в розподілі доходів або часову відкладеність у виторзі. Тому у практиці застосовуються більш гнучкі підходи. Існує кілька основних

методів прогнозування LTV, кожен із яких має свої переваги, обмеження та придатність залежно від типу бізнесу, структури даних та цілей аналітики.

Один із найпростіших і водночас широко використовуваних підходів — це метод ланцюгових коефіцієнтів. Його суть полягає в тому, щоб розглянути доходи когорт (груп користувачів, які зареєструвалися в один день або період) за фіксовані часові інтервали: 3, 7, 14, 30, 60, 90, 180 днів тощо. На основі цих періодів будуються співвідношення (коефіцієнти приросту доходу), які потім екстраполюються на довший горизонт, наприклад, 365 днів.

Алгоритм методу включає такі етапи:

1. Збір даних: фіксується кумулятивний виторг по кожній когорті за зазначені часові вікна (наприклад, на 3-й, 7-й, 14-й день тощо).
2. Розрахунок ланцюгових коефіцієнтів приросту доходу між періодами.
3. Екстраполяція на триваліший період. Якщо відомий виторг на день 3, можна отримати прогноз на 365 день як:

$$LTV_{365} = \text{Виторг}_3 * k_7 * k_{14} * \dots * k_{365}$$

де  $k$  це прирости доходу між періодами.

Цей метод дозволяє створити досить наочну і швидку модель для планування. Він часто використовується у продуктивній аналітиці, особливо на ранніх етапах, коли ще немає глибоких аналітичних ресурсів. Основною перевагою є простота реалізації та зрозумілість для бізнес-користувачів. Проте є й обмеження — метод добре працює лише при наявності стабільних історичних даних і регулярній поведінці клієнтів. У випадку різких змін у маркетингових каналах, поведінкових паттернах чи сезонності, точність значно знижується.

Більш математично обґрунтованим підходом є модель з використанням Beta та Gamma розподілів, також відома як BG/NBD-модель (Beta-Geometric/Negative Binomial Distribution). Вперше запропонована Fader & Hardie у 2005 році, ця модель широко використовується у транзакційних

бізнесах (роздріб, e-commerce), де кожен клієнт може здійснювати багато покупок упродовж життєвого циклу.

Суть методу полягає в моделюванні двох компонентів поведінки клієнта:

- Частота транзакцій (як часто клієнт здійснює покупки) — описується геометричним розподілом з параметрами Beta.
- Розмір транзакції — описується гамма-розподілом (Gamma), що дозволяє оцінити середній дохід від кожної покупки.

Формально модель можна представити так:

$$P(T) = \text{Beta}(\alpha, \beta) * \text{Gamma}(k, \theta)$$

де:  $\alpha, \beta$  — параметри, що відповідають за частоту взаємодій;  $k, \theta$  — параметри гамма-розподілу, які визначають середній дохід з транзакції.

Ця модель дозволяє прогнозувати як ймовірність активності клієнта в майбутньому, так і середню вартість кожного контакту з ним, що робить її особливо корисною для оцінки LTV у динаміці. Наприклад, компанія може заздалегідь визначити, з якою ймовірністю клієнт буде активний через 90 або 180 днів і який дохід він потенційно може принести. Проте для ремаркетингових сценаріїв, де клієнти можуть повертатися після довгих пауз або мають одиничні взаємодії, модель може втрачати точність. Також вона потребує ретельного калібрування, аналізу розподілів та наявності значного обсягу даних для адекватної побудови.

### 1.3. Виклики ремаркетингу та потреби в нових, адаптивних підходах

Сучасний цифровий маркетинг переживає фазу швидкої трансформації, спричиненої інтенсивним розвитком технологій, зміною поведінки споживачів і підвищеною конкуренцією на ринку. Одним із найбільш динамічних напрямів у цьому контексті виступає ремаркетинг. В умовах стрімкої зміни переваг споживачів і фрагментації ринку, ремаркетингові

кампанії потребують надзвичайно точного розуміння користувацької поведінки та здатності до оперативної адаптації маркетингових стратегій. Саме тому актуальним є впровадження адаптивних методів аналізу та прогнозування, які враховують складну економічну динаміку цього процесу.

Однією з головних проблем, з якими стикається бізнес у сфері ремаркетингу, є прогнозування lifetime value. Неправильна оцінка цього показника може призвести до систематичних неоптимальних перевитрат бюджету, некоректних прийнятих рішень та фінансових негараздів у компанії. У класичних підходах, таких як метод ланцюгових коефіцієнтів або модель BG/NBD (Beta-Geometric/Negative Binomial Distribution), прогнозування ґрунтується на припущенні про сталість поведінкових патернів користувачів і постійність ринкових умов. Проте реальна ситуація значно складніша: користувачі мобільних додатків демонструють різну частоту використання, нерегулярні шаблони транзакцій, різкий спад активності після встановлення застосунку, а також значну реакцію на зовнішні події (зміни у ціні реклами, сезонні фактори, кризи, зміни в екосистемі App Store/Google Play тощо).

Економічна природа цієї проблематики полягає у тому, що компанії, які інвестують у ремаркетинг, фактично приймають інвестиційне рішення із затримкою ефекту. Це означає, що вкладені кошти виявляють свою економічну результативність лише з часом, і саме тому прогноз LTV є своєрідним аналітичним аналогом оцінки прибутковості інвестиційного проекту. Помилки у прогнозуванні LTV еквівалентні неправильним оцінкам чистої теперішньої вартості або внутрішньої норми рентабельності у фінансовому аналізі. У такому випадку весь ремаркетинговий процес перетворюється на високоризикову діяльність із непередбачуваними фінансовими наслідками.

Крім того, проблема ускладнюється ще й тим, що звичні методи не здатні адаптуватися до нових даних у режимі реального часу. У динамічних

ринкових умовах, де поведінка користувачів може змінюватися щотижня або навіть щодня, необхідно використовувати моделі, які можуть автоматично враховувати ці зміни. Адаптивні підходи, які базуються на алгоритмах машинного навчання, таких як регресійні моделі, дерева рішень, градієнтний бустинг або нейронні мережі, здатні будувати прогнози з урахуванням великої кількості змінних і багатовимірних залежностей. Такі моделі демонструють гнучкість у врахуванні нових патернів взаємодії користувачів із додатком, зміни каналів трафіку, якості ремаркетингових креативів, реакції на цінні стимули тощо.

Однак, застосування вищезазначених алгоритмів також має певні обмеження. Вони потребують глибокого технічного супроводу, потужних обчислювальних ресурсів і спеціалізованих знань у галузі Data Science. До того ж, результативність складних моделей часто знижується в умовах нестабільної або недостатньої вибірки, а інтерпретованість результатів може бути обмеженою. У свою чергу, моделі часових рядів, зокрема ARIMA (AutoRegressive Integrated Moving Average), залишаються дієвим компромісом між точністю, адаптивністю та пояснюваністю результатів. Вони добре підходять для випадків, коли мета полягає у побудові прогностичних моделей із середньо- та довгостроковим горизонтом, наприклад, 365 днів LTV, як у цій роботі.

Особливу увагу варто звернути на економічні виклики, пов'язані з точністю прогнозів у періодах зростаючої ринкової нестабільності. Наприклад, коливання курсу валют, збільшення витрат на цифрову рекламу або зміни в законодавстві щодо захисту персональних даних безпосередньо впливають на вартість ремаркетингу та конверсії. В умовах таких ризиків стратегічна адаптація – тобто здатність аналітичної системи швидко підлаштовуватися під нову економічну реальність – стає критично важливою. Моделі, які мають можливість автоматичного перенавчання, дозволяють

уникнути ефекту «старіння» прогнозів і зберігати їх актуальність навіть при радикальних змінах у поведінці користувачів.

У контексті даного дослідження використання ARIMA-моделі у середовищі BigQuery ML виглядає обґрунтованим вибором з кількох причин. По-перше, ця модель добре зарекомендувала себе в умовах обмеженої історії спостережень і дозволяє проводити прогнозування з урахуванням сезонних коливань та трендових змін. По-друге, інтеграція з BigQuery ML відкриває можливість роботи з великими масивами даних без необхідності завантаження їх у зовнішнє середовище, що економить ресурси та пришвидшує аналітичний цикл. По-третє, можливість вбудованого автоматичного оновлення моделей забезпечує гнучкість та адаптивність без постійного залучення інженерів або аналітиків.

Отже, сучасні виклики ремаркетингу зумовлюють необхідність відмови від статичних, універсальних підходів до прогнозування користувацької цінності та переходу до адаптивних, динамічних систем, які базуються на моделюванні часових рядів та інтеграції з хмарними обчислювальними середовищами. Економічна ефективність таких систем обумовлюється не лише їхньою здатністю забезпечити точні прогнози, а й тим, що вони дозволяють стратегічно управляти ризиками та ресурсами компанії.

## РОЗДІЛ 2. ПОБУДОВА МОДЕЛІ ARIMA PLUS В СЕРЕДОВИЩІ BIGQUERY ML

Модель ARIMA\_PLUS у BigQuery ML є передовим інструментом для прогнозування часових рядів, який поєднує автоматизацію обробки даних із потужними методами моделювання. На відміну від класичної моделі ARIMA, ARIMA\_PLUS включає комплексний конвеєр, що охоплює підготовку даних, виявлення аномалій, коригування ефектів свят, розклад сезонності за допомогою STL (Seasonal and Trend decomposition using Loess) і моделювання тренду з автоматичним вибором параметрів ARIMA. Цей підхід дозволяє користувачам, навіть без глибоких знань у статистиці чи машинному навчанні, створювати надійні прогнози для бізнес-завдань, таких як прогнозування продажів, аналіз трафіку або оцінка попиту. Конвеєр ARIMA\_PLUS складається з таких основних етапів.

### 2.1. Підготовка даних та обробка викидів та аномалій

Підготовка даних є критично важливим етапом, який визначає якість прогнозів. ARIMA\_PLUS автоматизує цей процес, але розуміння його механізмів дозволяє оптимізувати моделювання. Частота даних (щоденна, щотижнева, щомісячна тощо) визначається автоматично на основі міток часу. BigQuery ML аналізує інтервали між сусідніми точками, щоб встановити регулярну сітку часу. Наприклад, для щоденних даних інтервал становить 24 години, для щотижневих — 7 днів. Якщо частота нерегулярна, система виконує перевибірку, що описано нижче. Нерегулярні інтервали між мітками часу є поширеною проблемою в реальних даних. BigQuery ML застосовує перевибірку до регулярної частоти, використовуючи методи агрегації (наприклад, середнє або суму) для надлишкових даних або інтерполяцію для пропущених. Наприклад, якщо дані надходять із нерівномірними інтервалами

(наприклад, 1 день, 3 дні, 2 дні), система створює рівномірну сітку з фіксованим інтервалом (наприклад, 1 день). Математично, для агрегації дубльованих значень у межах одного часового інтервалу використовується середнє:

$$y = \frac{1}{n} \sum_{i=1}^n y_{t,i}$$

де  $y_t$  – агреговане значення для часової мітки  $t$ ,  $y_{t,i}$  – окреме значення у часовому інтервалі  $t$ ,  $n$  – кількість записів, що припадають на цей інтервал.

Якщо для однієї мітки часу існує кілька значень (наприклад, через помилки введення даних), BigQuery ML обчислює середнє арифметичне. Це зменшує шум і забезпечує єдину точку даних для кожної мітки часу. Альтернативні методи, такі як вибір максимального чи мінімального значення, не використовуються, щоб уникнути спотворення тренду.

Пропущені значення заповнюються за допомогою лінійної інтерполяції, яка дозволяє зберегти локальний тренд ряду. Якщо значення  $y_{t_1}$  та  $y_{t_2}$ , то інтерполяція здійснюється за формулою:

$$y_t = y_{t_1} + \frac{(y_{t_2} - y_{t_1}) * (t - t_1)}{t_2 - t_1}$$

де  $y_t$  – інтерпольоване значення на момент часу  $t$ ,  $y_{t_1}, y_{t_2}$  – відомі значення до і після пропущеного моменту,  $t_1, t_2$  – відповідні часові мітки для цих значень. Цей метод зберігає локальний тренд і є стійким до невеликих прогалин у даних. Для великих прогалин (понад 10% даних) рекомендується попередня обробка поза BigQuery ML, наприклад, за допомогою експоненційного згладжування.

Порівняно з іншими платформами, такими як Python (бібліотеки pandas, statsmodels), BigQuery ML пропонує автоматизовану підготовку даних, що економить час, але обмежує гнучкість. Наприклад, pandas дозволяє

застосовувати складні методи заповнення пропусків (forward fill, backward fill, spline interpolation), тоді як BigQuery ML обмежується лінійною інтерполяцією. Однак автоматизація BigQuery ML компенсує це швидкістю обробки великих наборів даних (до 1 мільйона точок).

Викиди та аномалії можуть значно спотворити результати моделі, тому ARIMA\_PLUS включає механізми їх виявлення та коригування. BigQuery ML використовує статистичні методи, такі як міжквартильний розмах (IQR), для виявлення викидів. Виявлені викиди коригуються шляхом заміни на значення, отримані через інтерполяцію або згладжування. Раптові зміни рівня (level shifts), спричинені зовнішніми подіями (наприклад, зміна політики ціноутворення), виявляються за допомогою аналізу залишкової компоненти після розкладу STL. BigQuery ML коригує такі зміни, нормалізуючи дані в точках стрибка, щоб модель не інтерпретувала їх як частину тренду.

Обробка викидів є важливою, але водночас делікатною процедурою, оскільки викиди можуть відображати значущі події, наприклад, піки попиту внаслідок маркетингових кампаній або сезонних коливань. Видалення або згладжування таких значень може призвести до втрати цінної інформації.

Для оцінки впливу корекції викидів використовується середньоквадратична помилка (RMSE) між оригінальними  $y_t$  та скоригованими значеннями  $\hat{y}_t$ :

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$$

де  $y_t$  – фактичне значення в момент часу  $t$ ,  $\hat{y}_t$  – значення після обробки,  $n$  – кількість спостережень. Якщо RMSE перевищує прийнятний поріг (наприклад, 10% від стандартного відхилення ряду), це може свідчити про надмірну корекцію, і в таких випадках рекомендується ручна перевірка викидів для визначення їх економічної або прикладної значущості.

## 2.2. Коригування ефектів та виявлення і екстраполяція сезонності

Ефекти свят, такі як зростання продажів під час новорічних свят, враховуються за допомогою параметра HOLIDAY\_REGION. Наприклад, для США (HOLIDAY\_REGION='US') BigQuery ML додає регресори, які моделюють вплив свят, таких як День подяки чи Різдво. Математично, ефект свят додається до моделі як зовнішній регресор. Цей підхід підвищує точність прогнозів у періоди з аномальною активністю, але вимагає правильного вибору регіону свят.

Сезонність є ключовим елементом багатьох часових рядів, і ARIMA\_PLUS ефективно її обробляє. BigQuery ML застосовує STL-розклад для декомпозиції часового ряду:

$$Y_t = T_t + S_t + R_t$$

де  $Y_t$  – спостережуваний часовий ряд,  $T_t$  – тренд,  $S_t$  – сезонна компонента,  $R_t$  – залишкова компонента. STL використовує ітеративне згладжування Loess для оцінки компонент. Сезонна компонента, отримана за допомогою STL-декомпозиції, екстраполюється в майбутнє за допомогою моделі експоненційного згладжування (ETS). Зокрема, використовується адитивна модель Холта-Вінтерса, яка враховує рівень, тренд та сезонність. Модель описується наступними рівняннями:

- Рівень:  $l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1})$
- Тренд:  $b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}$
- Сезонність:  $s_t = \gamma(y_t - l_t) + (1 - \gamma)s_{t-m}$
- Прогноз на  $h$  кроків вперед:  $\hat{y}_{t+h} = l_t + hb_t + s_{t-m+(h \bmod m)}$

де  $y_t$  – фактичне значення на момент  $t$ ,  $l_t$  – рівень (сгладжене середнє),  $b_t$  – тренд (темп зміни рівня),  $s_t$  – сезонна компонента,  $m$  – період сезонності,  $\alpha, \beta, \gamma \in [0,1]$  – параметри згладжування,  $h$  – горизонт планування.

BigQuery ML автоматично вибирає оптимальну конфігурацію ETS на основі характеристик даних, що зменшує потребу в ручному налаштуванні. Порівняно з іншими методами декомпозиції, такими як класична адитивна чи мультиплікативна декомпозиція, STL є більш гнучким і стійким до викидів. Наприклад, класична декомпозиція припускає постійний період сезонності, тоді як STL може адаптуватися до змінних патернів. Однак STL вимагає більших обчислювальних ресурсів, що може бути обмеженням для дуже довгих часових рядів. Тренд, виділений за допомогою STL, моделюється з використанням ARIMA. Модель ARIMA(p, d, q) описує стаціонарний часовий ряд після диференціювання. BigQuery ML використовує алгоритм auto.ARIMA для автоматичного вибору параметрів (p, d, q). Алгоритм оцінює кілька кандидатних моделей і вибирає ту, що мінімізує критерій інформації Акаїке (AIC). Параметри обмежені: (p, q): від 0 до 5, (d): від 0 до 2. Параметр AUTO\_ARIMA\_MAX\_ORDER (за замовчуванням 3) обмежує суму (p + q), що впливає на кількість кандидатних моделей.

Для підвищення точності ARIMA\_PLUS можна внести такі модифікації:

1. Гібридна обробка викидів: Комбінувати IQR із методами на основі машинного навчання, такими як ізоляційний ліс, для точнішого виявлення аномалій.
2. Інтеграція зовнішніх регресорів: Використовувати ARIMA\_PLUS\_XREG для включення додаткових змінних, таких як маркетингові витрати чи погодні дані.
3. Адаптивний горизонт прогнозу: Автоматично визначати оптимальний горизонт прогнозу на основі волатильності даних.

Ці модифікації потребують додаткової реалізації поза стандартним функціоналом BigQuery ML, але можуть бути протестовані в Python або R.

ARIMA\_PLUS порівняно з іншими методами прогнозування, такими як Prophet (розроблений Facebook) і LSTM (глибоке навчання), має такі переваги:

- Простота використання: SQL-інтерфейс не вимагає знань програмування.
- Автоматизація: Усуває потребу в ручному налаштуванні параметрів.
- Масштабованість: Обробляє великі набори даних у хмарі.

Недоліки:

- Менша гнучкість порівняно з Prophet, який дозволяє налаштувати сезонність і регресори.
- Обмежена підтримка нелінійних трендів порівняно з LSTM.

Отже, підсумовуючи вище сказане модель ARIMA\_PLUS у BigQuery ML є потужним інструментом для прогнозування часових рядів, який поєднує автоматизацію, масштабованість і точність. Її конвеєр обробляє складні аспекти аналізу, такі як викиди, сезонність і ефекти свят, що робить її ідеальною для бізнес-додатків. Розуміння технічних деталей і методологічних аспектів дозволяє користувачам оптимізувати моделювання та адаптувати його до специфічних завдань.

### **РОЗДІЛ 3. ПРОГНОЗУВАННЯ РЕМАРКЕТИНГОВИХ КАМПАНІЙ НА ПРИКЛАДІ ПІДПРИЄМСТВА “GENESIS COMPANY”**

Компанія “Genesis” — це одна з провідних українських кофаундингових ІТ-компаній, яка спеціалізується на побудові глобальних технологічних бізнесів у партнерстві з провідними підприємцями Центральної та Східної Європи. Її місія полягає в розвитку високотехнологічних продуктів, орієнтованих на міжнародні ринки, із застосуванням сучасних підходів до аналітики, маркетингу та монетизації. На сьогоднішній день застосунки компанії завантажено понад 400 мільйонів разів, і щомісяця ними активно користуються мільйони людей по всьому світу. У межах мого досвіду роботи молодшим аналітиком у команді, що спеціалізується на соціальних комунікаціях, мені випала нагода взяти участь у проєкті, який мав за мету покращити ефективність ремаркетингових кампаній. Використовуючи існуючу модель прогнозування поведінки користувачів, мною було розроблено довгостроковий ремаркетинговий прогноз на 365 днів, що дозволяє більш точно планувати рекламні активності для повторного залучення аудиторії. Компанія працює за транзакційною бізнес-моделлю, що забезпечує чітке відстеження взаємодій користувачів із продуктом та подальших конверсій. Важливою перевагою є потужна інфраструктура даних, побудована на основі BigQuery, яка дозволяє ефективно зберігати, обробляти та аналізувати великі масиви інформації у реальному часі. Це середовище стало основною платформою для реалізації моделі прогнозування та побудови дашбордів для внутрішньої аналітики. Результати прогнозу будуть використані для оптимізації рекламних кампаній у мобільних додатках на платформах iOS та Android, з урахуванням географічного таргетування. Такий підхід дозволяє підвищити точність маркетингових активностей, зменшити витрати на залучення повторних користувачів та збільшити загальний рівень монетизації продукту.

### **3.1. Створення таблиці маркетингових каналів по відповідних платформах**

На початковому етапі дослідження було побудовано узагальнену аналітичну таблицю, яка описує активність користувачів за маркетинговими каналами із розподілом за платформами (Android та iOS) і географічними регіонами (Сполучені Штати — US, решта країн — Worldwide). Основною метою цієї таблиці є визначення наявності історичних даних для побудови прогнозних моделей, а також встановлення відповідних типів моделей, що можуть бути використані для кожної конкретної комбінації канал/платформа/географія.

Для реалізації цього підходу були використані дані про встановлення додатків, отримані з внутрішніх баз компанії, зокрема таблиці users, що містить інформацію про користувача, дату встановлення та пов'язаного партнера (канал), а також допоміжні таблиці з параметрами моделей прогнозування. У межах запити було здійснено групування користувачів за каналами (net), платформами та географією. Розраховувалися кількісні характеристики, такі як кількість унікальних користувачів, дата першого і останнього встановлення, а також тривалість активного періоду (різниця між останньою і першою датою встановлення).

Географію було нормалізовано до двох категорій — US та WW, що відповідає внутрішнім бізнес-поділам при аналізі ринків. Далі, на основі створеної агрегації, для кожного каналу здійснювалась перевірка наявності моделі прогнозування. Було використано два джерела моделей: базові ARIMA-моделі та моделі на агрегованому географічному рівні, що зберігаються в таблицях параметрів \*\_model\_params та \*\_geo\_params. Якщо відповідна модель уже наявна — її назву призначали як model\_name. Якщо ж явної моделі не існувало, але історія встановлень була меншою за 180 днів, тоді автоматично призначалась базова модель ARIMA залежно від платформи

та географії, наприклад: `arima_andr_geo_us`. Однією із можливих модифікацій даного процесу є можливість створення додаткових XREG моделей з зовнішніми регресорами для каналів, які прожили більше 30 днів, але менше 180 для оригінальних ARIMA\_PLUS моделей, але ця ідея не була широко розглянута командою.

Наступним кроком було визначення, чи доступні дані для побудови прогнозу з горизонтом у 365 днів. Для цього була сформована допоміжна вибірка користувачів, які мали встановлення в інтервалі між 181 і 365 днями до поточної дати. Якщо для відповідної комбінації канал/географія/платформа у цьому періоді були зафіксовані встановлення — канал відмічався як готовий до моделювання (Yes). Якщо дані були відсутні, але період активності був менше 180 днів, канал отримував позначення No (Geo/XREG Model), тобто прогноз може бути побудований лише з використанням агрегованих географічних даних. В іншому випадку коли канал не підпадає під дані умови (не було даних в даний проміжок часу, або інші причини) встановлюється також позначення No (Early install date).

У фінальній таблиці залишались лише ті записи, де останнє встановлення зафіксоване не раніше ніж 30 днів тому. Це дозволило отримати актуальний знімок стану маркетингових каналів і підготовки до довгострокового прогнозування. Результатом побудови цієї таблиці є компактна, але інформативна структура, що дозволяє бізнес-команді швидко ідентифікувати, для яких платформ та каналів вже існують моделі прогнозування, де вони потребують оновлення або побудови з нуля, а також чи є достатній обсяг даних для моделювання на горизонті в 365 днів.

### **3.2 Формування таблиць транзакцій для ремаркетингу для відповідних платформ**

У межах побудови аналітичної моделі прогнозування LTV особливу роль

відіграє етап формування транзакційних таблиць. Саме ці таблиці дозволяють зібрати, стандартизувати й систематизувати ключові фінансові параметри, пов'язані з поведінкою користувачів після встановлення мобільного додатку. У контексті ремаркетингу — тобто повторного залучення користувачів за допомогою рекламних кампаній — дані про транзакції мають особливу цінність, оскільки дають змогу оцінити ефективність витрачених ресурсів, виявити закономірності у поведінці залучених сегментів та спрогнозувати майбутні прибутки з урахуванням усіх супутніх витрат.

Побудова транзакційних таблиць — це багатокроковий процес, який включає агрегацію інформації з низки джерел. Серед них — внутрішні бази даних про платежі (`payment_transaction`), податкові таблиці (наприклад, `vat_rates_eu`), інформація про встановлення додатків, країну користувача, маркетингову атрибуцію (джерело трафіку, кампанія, ремаркетинг) та операційні витрати (комісії, хостинг). Важливою умовою є точна зв'язка між подіями встановлення додатку та подальшими транзакціями — лише таким чином можна уникнути змішування даних при повторному встановленні або оновленні пристрою. Для кожної з платформ — iOS та Android — формується окрема таблиця. Вона включає такі ключові атрибути:

- Дата встановлення (`install_date`) — початкова точка життєвого циклу користувача;
- Дата транзакції (`event_date`) — день, коли було здійснено платіж;
- Lifetime — кількість днів між встановленням і транзакцією;
- Сума транзакції — з урахуванням податкових коригувань та комісій;
- Кількість транзакцій — кількість унікальних подій оплати;
- Країна користувача — визначає валюту та ставку ПДВ;
- Ознака ремаркетингу — ідентифікує користувачів, які були залучені вдруге;
- Платформа — Android або iOS, оскільки політики обробки транзакцій

відрізняються;

- Витрати на хостинг — інфраструктурні витрати на обслуговування користувача.

Ці дані формують структурований набір, що може бути використаний для статистичного аналізу, побудови time-series моделей та подальших ML-процедур з оцінки LTV.

Платформа iOS вимагає врахування специфіки політики Apple щодо комісій, зберігання транзакцій та оподаткування. Основним джерелом даних є таблиці `ltv.partner_costs` та `payment_transaction`. У цих таблицях фіксуються всі транзакції користувачів після інсталу, включаючи час, суму, джерело трафіку, платіжну систему та ідентифікатор пристрою.

Найскладнішим завданням є коректна фільтрація транзакцій, щоб виключити дублікати, що виникають через повторні інстали. Для цього в розрахунок потрапляють лише ті транзакції, які сталися після певного інсталу і до наступного інсталу того ж користувача. Це дозволяє уникнути ситуації, коли одна транзакція рахується кілька разів у різних життєвих циклах користувача.

Також важливим є врахування ПДВ. Якщо країна користувача відома (визначається за IP або billing-даними), застосовується відповідна податкова ставка. Це дозволяє точніше визначити фактичний дохід, оскільки ПДВ є непрямим податком, який не включається у чистий прибуток компанії.

Витрати на хостинг розраховуються з урахуванням тарифів, що діють на відповідну дату. Наприклад, для червня 2023 року може застосовуватись ставка `host_fee_ios = 0.1`, а для попередніх періодів — константа 0.09 або динамічне значення з таблиці `cost_inst`. Комісія Apple також змінюється у часі: якщо до середини 2023 року вона становила 15%, то після зміни політики вона може зрости до 30%, залежно від типу застосунку, доходу розробника та території.

Платформа Android (через Google Play) має свої особливості. Основна логіка побудови таблиці є подібною: транзакції зіставляються з інсталами, обраховується lifetime, визначається джерело користувача та ознака ремаркетингу. Проте, у Google комісійна політика є гнучкішою. Наприклад, зміна ставки комісії до 30% для певних категорій додатків відбулася у липні 2024 року, і це повинно бути враховано при розрахунках доходу.

Так само, як і для iOS, витрати на хостинг обраховуються на основі коефіцієнтів з таблиці `cost_hosts_koef`. У разі відсутності специфічного значення, застосовується середній коефіцієнт по місяцю. Це дозволяє утримувати узгодженість у витратах при змінах трафіку. Маркер `rm = 1` ставиться тим, хто був залучений через ретаргетинг (Google Ads, Meta, Apple Search Ads тощо) навіть після неактивності або деінсталяції юзера. Після первинного формування таблиць виконується ретельна перевірка повноти транзакцій. Часто трапляється ситуація, коли певна транзакція була облікована в системі платежів, але не пройшла фільтрацію під час джойнінгу з таблицею інсталів. Це може бути спричинено затримками синхронізації, некоректною прив'язкою до пристрою або тимчасовими технічними збоями. Щоб уникнути втрати важливих даних, реалізується механізм додаткової вставки таких записів. Тобто, усі комбінації “`user_id — event_date`”, які є у фінальній транзакційній таблиці, але відсутні після агрегації, додаються окремим запитом. Це гарантує цілісність і репрезентативність фінального датасету.

На завершальному етапі усі зібрані та оброблені дані формуються в єдину таблицю прогнозування LTV. Вона є основним джерелом для побудови моделей, які оцінюють очікувану цінність користувача протягом певного періоду. У фінальній таблиці фіксуються такі поля: унікальний id користувача; дата інсталу; дата транзакції; lifetime (у днях); платформа; регіон/країна (геолокація); дохід (`net revenue`); витрати на хостинг; ремаркетинговий маркер; додаткова метаінформація для аналітики (наприклад, `source_id`, `campaign_id`);

уніфікована дата транзакції (*formatted\_date*) — використовується для побудови часових рядів.

Поле *formatted\_date* є ключовим компонентом для застосування методів аналізу часових рядів у прогнозуванні LTV. Воно виконує роль уніфікованої часової координати, що дозволяє зрівняти всіх користувачів незалежно від дати їх фактичної установки додатку.

Суть цього підходу полягає в наступному: за початкову точку відліку береться мінімальна дата інсталу серед усіх користувачів, яка умовно вважається "першим днем життя" системи. До цієї дати додається значення *lifetime* — кількість днів, що минули від моменту інсталяції до певної транзакції користувача. Таким чином, кожна транзакція "переміщується" у спільний масштаб часу, де  $formatted\_date = мінімальна\_дата\_інсталу + lifetime$ . Цей підхід дозволяє:

- стандартизувати часові ряди для всіх користувачів;
- агрегувати транзакції у єдиному тимчасовому просторі;
- ефективно застосовувати моделі прогнозування LTV (BG/NBD, Gamma-Gamma, XGBoost, LSTM тощо), які залежать від структури часового ряду.

Умовно кажучи, всі користувачі "починають життя" одночасно, і подальші події моделюються у порівнянному часовому масштабі. Це суттєво підвищує точність аналізу та дозволяє краще виявляти закономірності поведінки платників у різні моменти їх життєвого циклу. Формування транзакційних таблиць є не просто технічним завданням, а фундаментальним етапом у побудові аналітичної інфраструктури, що дозволяє оцінювати ефективність бізнес-моделі, оптимізувати маркетингові бюджети та підвищувати рентабельність.

### 3.3 Динамічне створення, прогнозування та моніторинг ARIMA-моделей

Для полегшення розпису практичних обрахунків ми будемо саме акцентувати увагу на платформі iOS, але зазначаючи, що ідентичні методи адаптуємо до Android платформи. Першим кроком у процесі є формування набору параметрів моделей, які зберігаються в таблиці `ios_model_params` (рис. 3.1). Ця таблиця виконує роль керуючого шару – у ній визначено назви моделей, платформу, географічний регіон, а також діапазон допустимих значень прогнозу (`lower_bound` та `upper_bound`). Замість жорстко закодованого підходу, логіка побудована таким чином, що створення моделей відбувається у циклі: кожна конфігурація з таблиці параметрів зчитується як елемент масиву `model_params_ios`, після чого динамічно формується SQL-запит, який виконує створення моделі.

Рис. 3.1. Таблиця “ios\_model\_params” в середовищі BigQuery ML

Model_Name	Platform	Geography	Lower_Bound	Upper_Bound	Forecast_Model
arima_ios_ww_fb	Facebook	Worldwide	2	1500	forecast_value
arima_ios_us_fb	Facebook	United States	5	2800	forecast_value
arima_ios_wwlow	Snapchat	Worldwide	5	2000	forecast_value
arima_ios_us_mol	Moloco	United States	30	2900	forecast_value
arima_ios_us_sn	Snapchat	United States	30	6700	forecast_value
arima_ios_ww_ap	Apple Search Ads	Worldwide	50	3700	forecast_value
arima_ios_us_tt	TikTok	United States	50	2300	forecast_value
arima	Adwords	United States	50	1700	forecast_value
arima_ios_us_org	Organic	United States	50	12400	forecast_value
arima_ios_ww_tt	TikTok	Worldwide	50	3800	prediction_interval_lower_bound
arima_ios_ww_nc	Nocost	Worldwide	50	2000	prediction_interval_lower_bound
arima_apple_search_ios	Apple Search Ads	United States	50	16100	prediction_interval_lower_bound

*Джерело: середовище BigQuery ML компанії “Genesis”*

Моделі ARIMA створюються на основі таблиці `Forecasting_iOS`, яка містить агреговані дані про транзакційну активність користувачів у форматі: дата транзакції (`Formatted_Date`), платформа, географія та дохід. Створені моделі використовують такі опції:

- MODEL\_TYPE='ARIMA\_PLUS' – вказує на застосування вдосконаленої версії ARIMA з автоматичним визначенням параметрів.
- TIME\_SERIES\_TIMESTAMP\_COL='Formatted\_Date' – вказує, що прогноз будується на основі згладженого часового ряду.
- TIME\_SERIES\_DATA\_COL='Revenue' – цільова змінна для прогнозування – дохід.
- TIME\_SERIES\_ID\_COL=['Geography','Platform'] – дозволяє будувати незалежні моделі для кожної унікальної пари платформа-регіон.
- FORECAST\_LIMIT\_LOWER\_BOUND / UPPER\_BOUND – встановлює обмеження для діапазону прогнозу, що запобігає генерації нереалістичних значень. Фінальне прогнозоване значення встановлюється за допомогою масштабованого логіт-перетворення, яке має формулу:

$$y = \log\left(\frac{x - a}{b - x}\right)$$

де  $a$  – FORECAST\_LIMIT\_LOWER\_BOUND,  $b$  – FORECAST\_LIMIT\_UPPER\_BOUND

Обмеження по lifetime – 180 днів – не є довільним, а зумовлене специфікою нашої задачі: ми формуємо прогнози на основі когорт користувачів, для яких вже накопичено повні фактичні дані до 180-го дня життя. Це забезпечує надійну емпіричну базу для побудови моделі. У той же час цільовим горизонтом прогнозування є 365-й день, тож модель ARIMA має за допомогою аналізу патернів першої половини життєвого циклу екстраполювати поведінку користувачів на другу половину. Таким чином, 180 днів – це оптимальний баланс між достатнім обсягом історичних даних і необхідною глибиною прогнозу.

Таблиця 3.1. Параметри моделей ARIMA\_PLUS

geo	net	p	d	q	has_drift	log_probability	AIC	variance	seasonal_periods	has_holiday_effect	has_spikes_and_dips	has_step_changes
United States	Apple Search Ads	0	2	2	0	112	-218	0.01	[WEEKLY]	0	1	0
United States	Adwords	0	1	3	1	-166	342	0.39	[WEEKLY]	0	1	0
United States	Facebook	1	1	2	1	-79	169	0.15	[WEEKLY]	0	0	0

*Джерело: складено на основі розрахунків автора*

Наступним етапом є безпосередній запуск прогнозів на побудованих моделях. Прогнозування виконується одразу для кількох моделей за допомогою команди `ML.FORECAST`, яка використовується у вигляді об'єднання (`UNION ALL`) для усіх релевантних моделей. Для кожної з них прогноз здійснюється на горизонті 180 днів при довірчому інтервалі 90%. Кожен прогнозований рядок містить:

- дату прогнозу (`forecast_timestamp`);
- значення прогнозу (`forecast_value`);
- нижню межу інтервалу довіри (`prediction_interval_lower_bound`);
- ідентифікатор моделі (`Forecast_Model`).

Отримані результати об'єднуються з фактичними транзакційними даними з таблиці `Forecasting_iOS`, що дозволяє здійснити порівняльний аналіз між прогнозованими значеннями та реальним доходом. Результати зберігаються у таблицю `ios_forecasting_monitoring` (табл.3.2), де щодня зберігається зріз з:

- датою архівації (`archive_date`);
- регіоном і платформою;
- загальним прогнозованим значенням доходу;
- фактичним доходом.

Таблиця 3.2. Приклади прорахунків в таблиці `ios_forecasting_monitoring`

<code>archive_date</code>	<code>geography</code>	<code>platform</code>	<code>forecast</code>	<code>revenue</code>
07.03.2025	United States	Adwords	5080	4512
07.03.2025	United States	Apple Search Ads	44668	37190
07.03.2025	United States	Facebook	12577	11262

*Джерело: складено на основі розрахунків автора*

Для забезпечення контролю якості моделей створена окрема підсистема оцінки їх точності. Логіка полягає в обчисленні співвідношення між реальним доходом та прогнозом для кожної моделі (`revenue / forecast`). Якщо дане співвідношення виходить за межі допустимого діапазону, модель автоматично маркується як така, що має "екстримальне відхилення".

Математично це реалізовано через порівняння значення `ratio` (відношення `revenue` до `forecast`) із середнім значенням `mean` та стандартним відхиленням `std_dev`, які розраховуються по всій таблиці моніторингу. Якщо:

- співвідношення сильно відрізняється від середнього (більше ніж на три стандартних відхилення),
- або значення менше 0.5 (прогноз завищений вдвічі),
- або більше 2 (прогноз занижений вдвічі),

тоді таку модель вважають нестабільною. Всі випадки таких відхилень

зберігаються у таблиці `ios_bad_performance`.

Особливої уваги заслуговує додатковий контроль: якщо протягом останніх 7 днів поспіль для певної моделі фіксуються екстремальні відхилення, вона потрапляє у спеціальну таблицю `ios_bad_table`, яка сигналізує про необхідність перегляду параметрів моделі або перетренування на нових даних. Таким чином, підтримується баланс між автоматизацією процесу моделювання та якісним контролем його результатів.

### **3.4 Оптимізація ARIMA моделей при створенні прогнозів**

При додаванні моделі до таблиці `ios_bad_table`, вона потрапляє в оптимізаційну процедуру, яка дозволяє автоматично покращувати показники моделі. Зважаючи на велику кількість моделей, що використовуються для різних комбінацій географічного розташування та маркетингових сіток, було реалізовано гнучку процедуру автоматизованої оптимізації моделей ARIMA у BigQuery, яка враховує змінні вхідні параметри, проводить навчання моделей, оцінює їхню якість та забезпечує моніторинг на рівні продакшн.

Ключовим принципом реалізації є те, що моделі не створюються разово й статично. Замість цього, для кожної моделі заздалегідь визначається низка гіперпараметрів, яка описується через два значення: нижню межу (lower bound) та верхню межу (upper bound). Ці параметри визначають мінімальні межі прогнозованих значень, як ми вже зазначали раніше. Для різних моделей (тобто комбінацій географії та платформи) передбачаються різні правила генерації параметрів. Наприклад, деякі моделі передбачають лише зміну нижньої межі при фіксованій верхній, тоді як інші дозволяють варіювати обидві межі одночасно, що генерує ширший простір комбінацій.

При кожній комбінації параметрів моделюється окрема ARIMA модель, результати прогнозування, якої порівнюються з наявними неповними фактичними даними. На основі цього розраховується середня абсолютна

похибка (MAE), яка виступає головним критерієм якості моделі. Оцінка MAE проводиться окремо для двох сценаріїв:

- Основний (central) — базується на прогнозованому значенні доходу.
- Консервативний (pessimistic) — базується на нижній межі довірчого інтервалу.

Такий підхід дозволяє порівняти сценарії з різним ступенем ризику: основний прогноз відображає найімовірнішу траєкторію доходу, тоді як песимістичний дозволяє побачити, наскільки може просісти дохід у найгіршому випадку. Після отримання значень MAE для всіх комбінацій параметрів для кожної моделі виконується автоматичний вибір найкращої конфігурації. Найнижче значення MAE серед усіх варіантів вважається оптимальним, і зберігається відповідна пара меж (lower bound / upper bound) разом із типом обраного сценарію (основного або консервативного). Ця інформація централізовано фіксується у таблиці параметрів моделей, яка використовується для наступних кроків. Описана процедура є повністю автоматизованою та масштабованою. Додавання нової моделі (наприклад, для нового ринку або нового типу користувачів) вимагає лише внесення її параметрів до відповідного списку. Уся інша логіка — від створення параметрів до вибору найкращого варіанту — виконується автоматично.

Після перерахунку моделей з процедури настає фінальний етап систематизованого обрахунку прогнозів з моделі ARIMA\_PLUS. Основна мета заключного процесу — виявити потенційні відхилення фактичних доходів (revenue) від прогнозованих значень (forecast) та скоригувати прогнози за допомогою множників, сформованих на базі розподілу фактичних/прогнозованих співвідношень. На початку етапу виконується очищення таблиці ios\_monitoring\_adjust від записів за поточну дату архівації. Це гарантує, що всі подальші дані будуть оновлені актуальними значеннями, без накладення настарілих результатів. Далі виконується вставка нових

значень, які базуються на логіці перерахунку скоригованих прогнозів (adjusted forecast).

Центральним елементом процесу є розрахунок коефіцієнта ratio, який визначає співвідношення між фактичним доходом і прогнозом:  $revenue / forecast$ . Це дозволяє зрозуміти, наскільки реальні показники відрізняються від очікуваних. Після отримання масиву коефіцієнтів ratio, система розраховує міжквартильний розмах (IQR) — це різниця між 75-м і 25-м процентилем розподілу. Використання IQR дозволяє виявити межі нормальних коливань у співвідношеннях. Якщо ratio виходить за межі цієї зони, це може свідчити про аномалію в прогнозі.

На базі цих розрахунків формується динамічний множник (multiplier) для коригування прогнозу. Він базується на піднятті ratio до ступеня, що враховує величину IQR. Такий підхід дозволяє не просто масштабувати прогноз, а робити це з урахуванням того, наскільки сильне відхилення є від "нормального" розподілу значень.

Далі, для кожного випадку здійснюється перевірка умов (check). Якщо ratio менше 25-го перцентиля або якщо фактичні доходи перевищують прогноз, тоді прогноз не змінюється, тобто множник не застосовується. Це робиться для того, щоб уникнути зайвого коригування в ситуаціях, де доходи вже перевищили очікування або співвідношення є нижчим за "норму". Насамкінець, обчислюється фінальне значення скоригованого прогнозу: якщо перевірка (check) дорівнює нулю (тобто необхідне коригування), то прогноз масштабується за допомогою обчисленого множника, інакше лишається без змін. Результати кінцевого обчислення прогнозованого значення та їх порівняння з фактичним наведені в таблиці 3.3.

Таблиця 3.3. Результати відпрацювання моделей

geography	platform	MAE	RMSE
United States	TikTok	35	71
United States	Facebook	54	53
United States	Apple Search Ads	154	143

*Джерело: складено на основі розрахунків автора*

З метою моніторингу результатів прогнозу, розроблено інтерактивний дашборд в Looker Studio, який дозволяє оцінити ефективність роботи моделей (рис. 3.2).

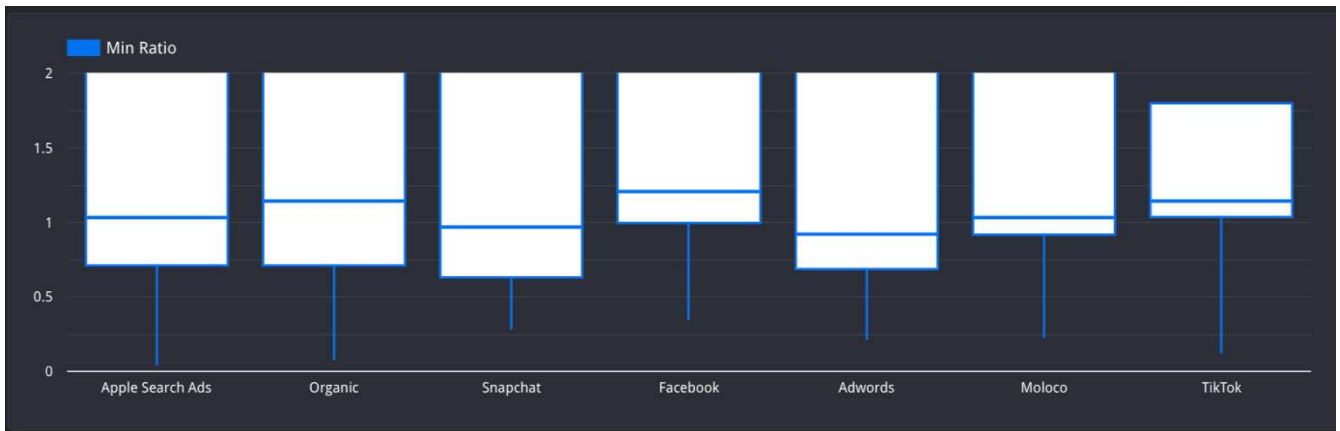


Рис 3.2. Boxplot-графіки відношення прогнозованих значень до фактичних  
*Джерело: складено на основі розрахунків автора*

Кінцевим результатом прогнозування є поєднання методу ланцюгових коефіцієнтів з результатами роботи моделі ARIMA\_PLUS. Перший крок — видалення старих записів з таблиці Coef\_GP, де комбінація географії та платформи присутня у параметрах активних моделей прогнозування (ios\_model\_params). Це забезпечує цілісність оновленої інформації, уникаючи накладення застарілих коефіцієнтів.

Після цього система здійснює розрахунок нових коефіцієнтів. Для цього спочатку агрегуються дані про доходи користувачів за певні періоди: 0, 3, 7, 14, 30, 60, 90 та 180 днів з моменту їхнього залучення. Для кожного періоду

обчислюється сумарна вартість транзакцій. На базі цих значень розраховуються коефіцієнти зростання, які показують, наскільки зростає сума доходу між зазначеними періодами. Наприклад,  $Coef_3$  — це співвідношення між доходами за 3 дні і доходами за 0 днів, а  $Coef_{30}$  — між 30 та 14 днями. Особливе значення має коефіцієнт  $Coef_{360}$ , який враховує прогнозовані значення на 180 днів у майбутньому, отримані за допомогою ARIMA-моделей. Він обчислюється як співвідношення між сумою доходів за попередні 180 днів та сумою, отриманою після додавання прогнозованого доходу на 180 днів. Цей коефіцієнт дає змогу оцінити очікуваний приріст у випадках, коли частина даних ще не реалізована фактично. Прогнозовані значення на 180 днів вперед формуються за допомогою об'єднання декількох ARIMA-моделей, що належать до різних каналів залучення користувачів. Наприклад, моделі  $arima\_ios\_us\_fb$ ,  $arima\_ios\_ww\_ar$ ,  $arima\_ios\_us\_org$  тощо відповідають за специфіку різних платформ або географій. Об'єднання результатів з усіх цих моделей дозволяє отримати повне уявлення про сумарний прогноз.

Кожен прогноз може бути модифікований множником (*multiplier*), якщо в моніторинговій таблиці відповідного регіону та платформи встановлений  $check = 1$ . У цьому випадку до суми прогнозів додається множник, який дозволяє врахувати аномальні зміни в трендах, які не були зафіксовані в попередніх періодах. Інакше — використовуються «чисті» прогнози з моделей.

В результаті отримані коефіцієнти записуються в таблицю  $Coef\_GP$ , і стають основою для подальшого розрахунку прогнозних значень доходу користувачів у залежності від їхнього «віку» в системі та методом залучення. Завдяки автоматизації на базі BigQuery, ця модель забезпечує високу точність, масштабованість і практичну цінність, дозволяючи не лише передбачати майбутні доходи, але й адаптувати маркетингові стратегії для максимізації прибутковості в конкурентному середовищі.

## ВИСНОВКИ

Дана кваліфікаційна робота присвячена створенню ефективного, автоматизованого та масштабованого методу прогнозування доходів від ремаркетингових кампаній для користувачів мобільних додатків на горизонті 365 днів. У роботі було застосовано модель ARIMA\_PLUS у середовищі BigQuery ML, що дозволило поєднати можливості часових моделей із перевагами хмарної обробки великих даних. Основною метою дослідження було вирішення прикладної задачі прогнозування довгострокової цінності користувачів (LTV) з урахуванням динаміки їхньої поведінки, специфіки маркетингових каналів та необхідності оптимізації бюджетів кампаній.

Розгляд існуючих підходів до прогнозування LTV показав обмеження традиційних методів у контексті ремаркетингу. Метод ланцюгових коефіцієнтів виявився чутливим до нестачі історичних даних у нових когортах, що призводило до зростання похибки при прогнозуванні понад 180 днів. Модель BG/NBD, хоч і ефективна для оцінки регулярних транзакцій, демонструвала знижену точність в умовах ремаркетингу, де повторна активність користувача має більш спорадичний характер. Це підкреслює потребу в побудові методів, які є одночасно адаптивними, автоматизованими та здатними до самооновлення.

Запропонований метод базувався на моделі ARIMA\_PLUS, інтегрованій у BigQuery ML. Основна перевага цього підходу полягала у можливості здійснювати складний економетричний аналіз часових рядів безпосередньо у хмарному середовищі, що дозволяє обробляти значні обсяги транзакційних даних у режимі реального часу. Модель ARIMA\_PLUS дозволила врахувати лінійні тренди, сезонність, а також статистичні аномалії. Автоматичний підбір параметрів (order) та можливість застосування додаткових регресорів значно підвищили адаптивність системи. Суттєвою інновацією стало впровадження процесу автоматичного моніторингу якості моделей. Система регулярно

аналізує похибки прогнозів (MAE), порівнює їх з пороговими значеннями та ініціює перенавчання моделей у разі виявлення деградації якості. У рамках процедури гіперпараметричної оптимізації враховувалися також сценарії з різними ризиковими припущеннями — песимістичними та базовими. Кожна ARIMA-модель прогнозувала діапазон (upper/lower bound), а обрана конфігурація фіксувалася як найкраща за критерієм мінімального MAE. Завдяки цьому забезпечувалося не лише підвищення точності, а й можливість сценарного аналізу результатів.

Практична апробація розробленого методу була здійснена на базі реальних даних мобільного бізнесу компанії Genesis. Прогнози будувалися окремо для платформ iOS та Android у розрізі регіонів (зокрема, США та решта світу), що дало змогу враховувати різницю в поведінці користувачів на різних ринках. Для кожної комбінації каналів (наприклад, TikTok, Facebook, Apple Search Ads) здійснювався окремий аналіз, який включав побудову понад 1000 варіантів ARIMA-моделей з різними параметрами. Отримані результати показали високу ефективність методу. Зокрема, значення MAE для iOS у США склали 35 (TikTok), 54 (Facebook) і 154 (ASA), що є прийнятними для стратегічного прогнозування. Крім того, аналіз варіативності прогнозів у межах upper/lower bound дозволив сформувати сценарії ризику, що є важливим інструментом для фінансового планування та ухвалення рішень. Це також зменшило ймовірність хибної алокації бюджетів, коли через переоцінку очікуваних доходів зростає ризик негативного ROI.

Метод був протестований у динамічному середовищі, де дані змінювалися щодня. Завдяки регулярному моніторингу система виявляла ситуації, коли точність прогнозу знижувалась нижче критичного рівня, і запускала процедуру перенавчання з новими параметрами. Такий підхід дозволив не лише підтримувати актуальність моделей, а й виявляти структурні зміни у поведінці користувачів, зокрема сезонні зміни або вплив зовнішніх

подій. Сама логіка моделювання була реалізована у SQL, що спрощує інтеграцію в існуючі бізнес-процеси компанії. Крім того, структура рішення забезпечує його масштабованість. Додавання нової когорти, ринку або джерела трафіку відбувається автоматично через таблиці конфігурацій, без необхідності переписування логіки моделі. Це особливо важливо в умовах зростання обсягів даних та кількості каналів комунікації з користувачами. Робота також передбачає можливість підключення зовнішніх регресорів у майбутньому — наприклад, витрат на маркетинг або макроекономічних індикаторів.

Робота має виразний науковий вклад у напрямі застосування економетричних моделей у практичних задачах цифрового маркетингу. Основними теоретичними здобутками є розробка адаптивного механізму перенавчання часових моделей, формалізація сценарного підходу до аналізу доходів, інтеграція методів моніторингу якості прогнозів у хмарне середовище BigQuery ML та концептуалізація LTV-прогнозування як задачі послідовного статистичного оновлення. Таким чином, дослідження не лише створює корисний інструмент для бізнесу, але й розширює прикладне застосування ARIMA-моделей у нових сферах.

Запропонований метод має високу практичну цінність. Він може бути використаний маркетологами для планування бюджетів, аналітиками для оцінки ефективності кампаній, продуктовими менеджерами для аналізу життєвого циклу користувачів, а також стратегічними менеджерами для сценарного планування доходів. Подальше вдосконалення передбачає інтеграцію зовнішніх факторів, таких як витрати на рекламу чи сезонні івенти, розширення моделі на нові платформи, автоматизацію звітності за допомогою візуальних дашбордів та порівняння з альтернативними моделями, зокрема LSTM або Prophet. Запропонована система є гнучкою основою для подальших досліджень. Перспективними напрямками можуть стати вивчення впливу

частоти взаємодій у ремаркетингу на прогнозовані доходи, інтеграція мультиваріантних часових моделей із залежними рядами та побудова системи раннього попередження про потенційні провали кампаній. Також доцільно розглянути можливість поєднання статистичних моделей із методами штучного інтелекту для підвищення надійності прогнозів.

Таким чином, поставлена в роботі задача — розробка методу прогнозування доходів ремаркетингу на 365 днів — була повністю вирішена. Метод демонструє високий рівень точності, гнучкості та адаптивності. Він поєднує найкращі риси класичних економетричних моделей і сучасних хмарних інструментів обробки даних. Це дозволяє не лише приймати більш обґрунтовані маркетингові рішення, але й сформувати аналітичну інфраструктуру, здатну підтримувати зростання бізнесу в умовах високої конкуренції цифрових продуктів. Робота створює надійну платформу для подальших досліджень у сфері прикладної аналітики, зокрема для прогнозування CLV, оцінки ефективності кампаній і стратегічного планування на основі сценаріїв ризику.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. The CREATE MODEL statement for ARIMA\_PLUS models URL: <https://cloud.google.com/bigquery/docs/reference/standard-sql/bigqueryml-syntax-create-time-series>
2. BigQuery: Demystifying ARIMA\_PLUS vs ARIMA\_PLUS\_XREG for Time Series Forecasting URL: <https://medium.com/google-cloud/bigquery-demystifying-arima-plus-vs-arima-plus-xreg-for-time-series-forecasting-08a068c0b20c>
3. Predicting Customer Life Time Value (CLTV) via Beta Geometric / Negative Binominal Distribution (BG/NBD) and Gamma Gamma Model URL: <https://medium.com/geekculture/predicting-customer-life-time-value-cltv-via-beta-geometric-negative-binominal-distribution-59be07ac30bd>
4. Sales Prediction Based on ARIMA Time Series and Multifactorial Linear Model URL: [https://www.researchgate.net/publication/369462420\\_Sales\\_Prediction\\_Based\\_on\\_ARIMA\\_Time\\_Series\\_and\\_Multifactorial\\_Linear\\_Model](https://www.researchgate.net/publication/369462420_Sales_Prediction_Based_on_ARIMA_Time_Series_and_Multifactorial_Linear_Model)
5. Forecasting: Principles and Practice, Rob J Hyndman and George Athanasopoulos URL: <https://otexts.com/fpp2/>
6. Методи прогнозування LTV для транзакційного методу монетизації в застосунку iOS URL: <https://ela.kpi.ua/items/359971d2-3419-44a4-89e2-e97cb5d740e4>
7. Customer Lifetime Value Prediction: Part 1 — Heuristics, Probabilities, and Machine Learning URL: <https://medium.com/@chenycy/customer-lifetime-value-prediction-part-1-heuristics-probabilities-and-machine-learning-a31468090008>
8. Forecasting LTV with SQL and Excel for E-Commerce URL:

<https://segment.com/docs/guides/how-to-guides/forecast-with-sql/>

9. BigQuery documentation URL: <https://cloud.google.com/bigquery/docs>
10. Genesis IT компанія URL: <https://www.gen.tech/>
11. SQL Stored Procedures URL:  
[https://www.w3schools.com/sql/sql\\_stored\\_procedures.asp](https://www.w3schools.com/sql/sql_stored_procedures.asp)
12. Моделі масштабування бізнесу на засадах юніт-економіки URL:  
<https://ela.kpi.ua/items/5b173c03-8da4-43ec-ac72-da60fa1b6e99>
13. Прогнозування відтоку користувачів за допомогою машинного навчання  
URL: <https://ela.kpi.ua/bitstreams/57f15d1d-75c0-4e77-8679-faa9032633d5/download>
14. Analyzing the Usability, Performance, and Cost-Efficiency of Deploying ML Models on BigQuery ML and Vertex AI in Google Cloud URL:  
<https://dl.acm.org/doi/abs/10.1145/3694860.3694863>
15. Machine learning in google cloud big query using sql URL:  
[https://www.researchgate.net/profile/Ravi-Kashyap-10/publication/375525277\\_Machine\\_Learning\\_in\\_Google\\_Cloud\\_Big\\_Query\\_using\\_SQL/links/654dd685ce88b87031d8dd2e/Machine-Learning-in-Google-Cloud-Big-Query-using-SQL.pdf](https://www.researchgate.net/profile/Ravi-Kashyap-10/publication/375525277_Machine_Learning_in_Google_Cloud_Big_Query_using_SQL/links/654dd685ce88b87031d8dd2e/Machine-Learning-in-Google-Cloud-Big-Query-using-SQL.pdf)
16. Machine Learning with BigQuery ML: Create, execute, and improve machine learning models in BigQuery using standard SQL queries URL:  
[https://books.google.com.ua/books?hl=en&lr=&id=8PMuEAAAQBAJ&oi=fnd&pg=PP1&dq=BigQuery+ML&ots=BX2IaLq7KY&sig=YaFjeVBWkKCPphoDA7SeSMJU0Q0&redir\\_esc=y#v=onepage&q=BigQuery%20ML&f=false](https://books.google.com.ua/books?hl=en&lr=&id=8PMuEAAAQBAJ&oi=fnd&pg=PP1&dq=BigQuery+ML&ots=BX2IaLq7KY&sig=YaFjeVBWkKCPphoDA7SeSMJU0Q0&redir_esc=y#v=onepage&q=BigQuery%20ML&f=false)
17. Turbo remarketing by e-commerce start-ups: Converting abandoned carts into sales  
URL: <https://www.researchgate.net/profile/Naveen-Kumar-Ranganathan->

[2/publication/381527010\\_Turbo\\_Remarketing\\_by\\_E-Commerce\\_Start-ups\\_Converting\\_Abandoned\\_Carts\\_into\\_Sales/links/668a12d20a25e27fbc2f9cb9/Turbo-Remarketing-by-E-Commerce-Start-ups-Converting-Abandoned-Carts-into-Sales.pdf](https://publication/381527010_Turbo_Remarketing_by_E-Commerce_Start-ups_Converting_Abandoned_Carts_into_Sales/links/668a12d20a25e27fbc2f9cb9/Turbo-Remarketing-by-E-Commerce-Start-ups-Converting-Abandoned-Carts-into-Sales.pdf)

18. Економічне прогнозування: навчальний посібник URL: [https://elartu.tntu.edu.ua/bitstream/lib/25820/1/%D0%9D%D0%B0%D0%B2%D1%87%D0%B0%D0%BB%D1%8C%D0%BD%D0%B8%D0%B9\\_%D0%9F%D0%BE%D1%81%D1%96%D0%B1%D0%BD%D0%B8%D0%BA\\_%D0%95%D0%BB%D0%B0%D1%80%D1%82%D1%83.doc](https://elartu.tntu.edu.ua/bitstream/lib/25820/1/%D0%9D%D0%B0%D0%B2%D1%87%D0%B0%D0%BB%D1%8C%D0%BD%D0%B8%D0%B9_%D0%9F%D0%BE%D1%81%D1%96%D0%B1%D0%BD%D0%B8%D0%BA_%D0%95%D0%BB%D0%B0%D1%80%D1%82%D1%83.doc)
19. A Comparison of Machine Learning Classifiers for Evaluation of Remarketing Audiences in E-Commerce URL: <https://dergipark.org.tr/en/download/article-file/1573085>
20. Cross-Generational Impacts of Digital Remarketing: An Examination of Purchasing Behaviours among Generation Z and Generation Y URL: <https://doi.org/10.57125/FEL.2023.06.25.06>
21. Google Remarketing URL: <https://mailchimp.com/marketing-glossary/google-remarketing/>
22. Rethinking retargeting with machine learning: a technical quest URL: <https://medium.com/incentro/rethinking-retargeting-with-machine-learning-a-technical-quest-bd7ee956da6>
23. The Ultimate Guide to Remarketing: Understanding Tracking Consent, Cookies, and Google Ads URL: <https://www.jentis.com/en/article/remarketing-explained/>
24. Machine Learning and Remarketing URL: <https://www.slideshare.net/slideshow/machine-learning-and-remarketing/89083697>

25. A Comparison of Machine Learning Classifiers for Evaluation of Remarketing Audiences in E-Commerce URL: [https://www.researchgate.net/publication/353624302\\_A\\_Comparison\\_of\\_Machine\\_Learning\\_Classifiers\\_for\\_Evaluation\\_of\\_Remarketing\\_Audiences\\_in\\_E-Commerce](https://www.researchgate.net/publication/353624302_A_Comparison_of_Machine_Learning_Classifiers_for_Evaluation_of_Remarketing_Audiences_in_E-Commerce)
26. Чорноус Г.О. *Моделювання процесу прийняття управлінських рішень в соціально-економічних системах на основі інтелектуального аналізу даних*: дис. ... д-ра екон. наук: 08.00.11 / Чорноус Галина Олександрівна ; М-во освіти і науки України, Київ. нац. ун-т ім. Тараса Шевченка. – Київ, 2015. – 485 с. + 7 арк. дод.
27. Комашко О., Рябий М., Распутна Л. SARCN-моделі та прогнозування в умовах невизначеності: теорія і практика // Банківська справа. – 1998. – С. 47–52.
28. Черняк О.І., Ставицький А.В., Баженова О.В., Шебаніна О.В. *Економетрика: підручник [для студентів вищих навчальних закладів]* / О.І. Черняк, О.В. Комашко, А.В. Ставицький, О.В. Баженова ; КНУТШ. – Київ: Київський університет, 2010. – 359 с. – ISBN 978-966-439-236-2.

## Календарний план виконання кваліфікаційної роботи бакалавра

№	Етапи роботи	Терміни виконання	Відмітка керівника про виконання
1	Вибір теми кваліфікаційної роботи	До 20 січня	Виконано вчасно
2	Розробка та затвердження завдання кваліфікаційної роботи	До 20 лютого	Виконано вчасно
3	Формування теоретичного розділу	До 20 березня	Виконано вчасно
4	Збір і попередня обробка статистичних даних (транзакційні дані, встановлення взаємозв'язків між таблицями)	До 20 квітня	Виконано вчасно
5	Побудова прогнозів з використанням BigQuery ML	До 20 травня	Виконано вчасно
6	Побудова автоматичного моніторингу результатів моделей	До 20 травня	Виконано вчасно
7	Аналіз результатів і формування висновків	До 20 травня	Виконано вчасно
8	Оформлення основної частини кваліфікаційної роботи відповідно до вимог	До 1 червня	Виконано вчасно
9	Подання роботи до попереднього захисту	До 1 червня	Виконано вчасно
10	Внесення правок за результатами попереднього захисту	До 13 червня	Виконано вчасно
11	Редагування, перевірка на плагіат, підготовка до захисту	До 17 червня	Виконано вчасно
12	Захист роботи	23 червня	Виконано вчасно

**Науковий керівник:** Черноус Галина Олександрівна

**Студент:** Кузнецов Нікіта Сергійович