Київський національний університет імені Тараса Шевченка

Економічний факультет

Кафедра економічної кібернетики

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

«Прогнозування ефективності рекламних кампаній на українському ринку Telegram на основі складових рекламних креативів»

студентки 4 курсу
спеціальності 051 «Економіка»
ОПП «Економічна кібернетика»
денної форми навчання
Ковальчук Мілани Юріївни

Науковий керівник:
доктор економічних наук, професор
Ляшенко Олена Ігорівна

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 05 червня 2024 р., протокол № 15

Завідувач кафедри: доктор економічних наук,
професор Ляшенко Олена Ігорівна _____

КИЇВ – 2024

РЕФЕРАТ

Кваліфікаційна робота бакалавра містить: 41 сторінку, 30 рисунків, 1 таблиця, 38 джерел, додатки.

Ключові слова: рекламні кампанії, Telegram, ринок цифрової реклами, рекламні креативи, аналіз ефективності, маркетингові стратегії, кореляційний аналіз, регресійний аналіз, KNIME Analytics.

Об’єкт дослідження: рекламні кампанії на українському ринку Telegram.

Мета дослідження: прогнозування ефективності рекламних кампаній на основі складових рекламних креативів у месенджері Telegram.

Методи дослідження: кореляційний аналіз, регресійний аналіз, аналіз великих даних, використання аналітичної платформи KNIME Analytics.

Наукова новизна, теоретична значимість дослідження: розробка методики прогнозування ефективності рекламних кампаній, яка включає аналіз складових рекламних креативів. Вперше в українських дослідженнях використовується дана методика для ринку Telegram, що дозволяє глибше розуміння залежностей між елементами рекламних креативів та ефективністю реклами.

Практична цінність: результати дослідження можуть бути використані маркетинговими агенціями для оптимізації стратегій рекламних кампаній в Telegram, забезпечуючи більшу залученість аудиторії та ефективність реклами.

RESUME

Taras Shevchenko National University of Kyiv,

Faculty of Economics, Department of Economic Cybernetics

The graduation research of student Kovalchuk Milana deals with the prediction of advertising campaign effectiveness based on the components of advertising creatives in the Telegram messenger within the Ukrainian market. This work explores the dynamics of digital marketing strategies through the utilization of correlation and regression analysis, extensive data analytics, and KNIME Analytics as a tool for analytical processing.

Key words: advertising campaigns, Telegram, digital marketing market, advertising creatives, effectiveness analysis, marketing strategies, correlation analysis, regression analysis, KNIME Analytics.

The object of research is advertising campaigns on the Ukrainian Telegram market. The purpose of the research is to predict the effectiveness of advertising campaigns by analyzing the components of advertising creatives in Telegram. The study employs methods such as correlation analysis, regression analysis, big data analytics, and the use of the KNIME Analytics platform.

The scientific novelty and theoretical significance of the research lie in the development of a methodology for predicting the effectiveness of advertising campaigns, which includes the analysis of the components of advertising creatives. This approach is first of its kind in Ukrainian research applied to the Telegram market, enabling a deeper understanding of the relationships between the elements of advertising creatives and the effectiveness of advertising.

The practical value of the research is significant as the results can be utilized by marketing agencies to optimize advertising campaign strategies in Telegram, ensuring greater audience engagement and advertising effectiveness.

Pages: 41, tables: 1, bibliog.: 38, append included.

ЗМІСТ

ВСТУП 5

РОЗДІЛ 1 РОЗМІЩЕННЯ РЕКЛАМИ В TELEGRAM 7

1.1. Месенджер Telegram в Україні 7

1.2. Розміщення реклами в Telegram 9

1.3. Аналітика реклами в Telegram 13

Висновки до розділу 1 16

РОЗДІЛ 2 АНАЛІТИКА ТА ПІДГОТОВКА БАЗИ ДАНИХ РЕКЛАМНИХ ПОСТІВ 17

2.1. Методологія CRISP-DM 17

2.2. KNIME Analytics як альтернатива IBM SPSS Modeler 20

2.3. База даних рекламних постів 22

2.4. Кореляційний аналіз 29

Висновки до розділу 2 33

РОЗДІЛ 3 ПРОГНОЗУВАННЯ ЕФЕКТИВНОСТІ РЕКЛАМИ В TELEGRAM В УКРАЇНІ 34

3.1. Підготовка даних до моделювання 34

3.2. Моделювання ефективності рекламного допису 35

3.3. Оцінка точності моделювання 40

Висновки до розділу 3 43

ВИСНОВКИ 44

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 46

ДОДАТКИ 51

ВСТУП

Сучасний світ цифрових технологій постійно еволюціонує, призводячи до змін у способах спілкування та розповсюдження інформації. Один з яскравих прикладів такої еволюції — популярність месенджера Telegram в Україні, який використовується не тільки для особистого спілкування, але й як потужний маркетинговий інструмент. Це робить актуальним дослідження механізмів і факторів ефективності рекламних кампаній у цій платформі.

Об’єктом дослідження є рекламні кампанії в месенджері Telegram.

Предмет дослідження — компоненти рекламних креативів, що впливають на ефективність рекламних кампаній.

Мета дослідження — прогнозування ефективності рекламних кампаній на основі аналізу складових рекламних креативів. Для досягнення цієї мети були визначені наступні завдання:

1. Визначення шляхів використання месенджеру Telegram в Україні.
2. Дослідження впливу нового законодавства України на бізнес в Telegram.
3. Аналіз рекламних креативів, розміщених в Telegram на початку 2024 року.
4. Виявлення залежності між компонентами креативів і залученістю аудиторії каналу до рекламного допису за допомогою кореляційного аналізу.
5. Розробка моделі для прогнозування ефективності рекламних кампаній на платформі KNIME Analytics.
6. Спростувати або підтвердити висунуті гіпотези про впливовість компонентів рекламного допису напряму на результативність реклами.

Методи дослідження включають кореляційний та регресійний аналіз, які допоможуть з'ясувати взаємозв'язки між різними факторами і спрогнозувати результативність рекламних кампаній.

Наукова новизна роботи полягає в розробці методики прогнозування ефективності рекламних кампаній з використанням специфічних характеристик контенту в Telegram. Практична новизна полягає у можливості застосування розробленої методики агенціями для оптимізації маркетингових стратегій.

Результати роботи були представлені на конференції Шевченківська весна 2024 «Стратегії економічного зростання: погляд у майбутнє для України» і передані для використання в майбутньому масштабному дослідженні Telegram за 2024 рік від сервісу Telemetrio [3, 21].

Інформаційна база дослідження складається з 30 тисяч рекламних постів українського ринку Telegram, наданих сервісом аналітики Telemetrio [17].

Структура роботи включає вступ, три основних розділи, висновки та список використаних джерел з додатками. Кожен розділ поділений на змістовні підпункти та містить проміжні висновки.

РОЗДІЛ 1
РОЗМІЩЕННЯ РЕКЛАМИ В TELEGRAM

1.1. Месенджер Telegram в Україні

В умовах стрімкого розвитку цифрового маркетингу і на тлі росту популярності соціальних медіа, Telegram в Україні став не лише майданчиком для обміну повідомленнями, а й ефективним інструментом для розміщення реклами. Інтерес до каналів в Telegram породив бізнес, де власники цих каналів заробляють десятки тисяч доларів на рекламі. Telegram став справжнім полем битви за думки та увагу людей, що підтверджує його значення та вплив.

Згідно з дослідженням компанії Internews, у 2023 році Telegram користувалися 72% українців. Порівняно з 2021 роком, коли частка користувачів становила лише 20%, популярність месенджера зросла більш ніж утричі. Це значне зростання спостерігалося на тлі військових дій, коли Telegram став ключовим інструментом для швидкого отримання та поширення інформації [2, 12].

Наразі серед 100 найпопулярніших Telegram каналів в Україні, топ 10 каналів на рис. 1.1 стосуються ЗМІ, громадських діячів, політиків чи державних структур. Однак лише один офіційний канал "ТСН новини" входить у топ-20. Велика частина каналів в топі є анонімними, що свідчить про різноманітність інформаційного контенту на платформі [6].

#	Назва	Підписники
1	Труха⚡Укра... @truexanewsua	2 590 141
2	Николаевски... @vanek_nikolaev	2 371 531
3	Украина Сей... Канал закритий	1 581 599
4	Лачен пише @lachentyt	1 490 695
5	Times of Ukra... @times_of	1 432 731
6	Реальна Війн... @voynareal	1 367 039
7	Инсайдер ... @insiderukr	1 330 473 44% 55%
8	Реальний Киї... @kievreal1	1 178 209
9	Україна Оп... @uaonlii	1 167 924 42% 57%
10	Всевидящее ... @okoo_ua	1 163 606

Рис. 1.1. Топ 10 Telegram каналів України станом на травень 2024 року

Джерело: розроблено автором на основі [6]

Гучний розголос серед Telegram спільноти отримав зареєстрований 25 травня 2024 року законопроєкт у Верховній Раді України. Він покликаний встановити чіткі рамки для регуляції платформ спільного доступу до інформації, включаючи Telegram. Цей законопроєкт створює баланс між свободою вираження поглядів та захистом національних інтересів, що важливо в умовах зовнішнього інформаційного впливу [14, 20].

Пропонуємо розібратися у 3 його головних аспектах та оцінити їх вплив на ринок реклами:

1. Введенні обмежень на діяльність медіа, які належать або фінансуються резидентами держави-агресора, є ключовим для забезпечення інформаційної безпеки України. Такі заходи не впливають на більшість рекламних кампаній у Telegram, оскільки вони здебільшого проводяться українським бізнесом і адміністрацією каналів, які працюють на внутрішньому ринку.

2. Запровадженні вимоги про наявність представництва в Україні для провайдерів платформ також сприяє зміцненню законності та відповідальності на ринку. Це не лише полегшує взаємодію з юридичними особами, але й сприяє прозорості та відкритості бізнесу. Ініціатива не становить загрози для рекламних кампаній, які здійснюються резидентами України, адже вони вже діють відповідно до українських законів та норм.

3. Зобов'язання провайдерів платформ розкривати структуру власності та джерела фінансування підвищує довіру споживачів та забезпечує вищу ступінь прозорості у сфері медіа. Це також важливий крок у боротьбі з дезінформацією та зовнішнім впливом, що особливо важливо в сучасних умовах.

Ці заходи навряд чи вплинуть на рекламні кампанії, що проводяться в Telegram, за винятком тих, що прямо пов'язані з державними структурами або фінансуються з-за кордону.

1.2. Розміщення реклами в Telegram

Репутація Telegram як надійного джерела новин сприяла зростанню інтересу до рекламних можливостей серед відомих брендів та місцевих підприємців, які почали активно взаємодіяти з адміністраторами Telegram каналів.

Економічне значення месенджера в Україні посилюється завдяки медійному впливу та великі аудиторії каналів. Основним джерелом доходу для таких каналів стає реклама, вартість якої може варіюватись від 80 до 1000 гривень за тисячу переглядів, як зазначає Олексій Мартиновський, власник медіа-холдингу в Telegram. Рівень доходу залежить від охоплення аудиторії, з яким на якісному каналі може досягати 25-30% [22].

Таким чином, канал у Telegram з аудиторією у мільйон підписників, де один пост здатен зібрати 250-300 тисяч переглядів, може заробити не менше 25 тисяч гривень за одне рекламне розміщення. Згідно з коментарем Олексія Мартиновського, найнижча ціна за рекламу на каналах з мільйонною аудиторією становить 100 гривень за тисячу переглядів на день. Канали з мільйоном або більше підписників можуть заробляти мінімум 30 тисяч доларів щомісяця. Він також зазначає, що середній місячний дохід таких каналів коливається від 50 до

70 тисяч доларів, а в часи виборчих кампаній чи інших значущих подій доходи можуть перевищувати 100 тисяч доларів [22].

Значні кошти витрачаються і на підтримку та розвиток каналів, включно з обслуговуванням, підтримкою команди та просуванням. Власники каналів часто інвестують від 25 до 50% доходів у залучення нових підписників. Витрати на персонал залежать від розміру каналу та процесів. Великі сітки каналів, такі як "Труха" із аудиторією 8 мільйонів, можуть мати щонайменше сотню працівників, тоді як деякі мільйонники обходяться кількома особами в штаті [18, 22].

Що стосується інвестицій у розвиток каналів, для залучення 100 тисяч підписників потрібно близько 1 мільйона гривень, що становить приблизно 10 гривень за одного підписника. Щоб створити мільйонний канал, зараз необхідно інвестувати до 750 тисяч доларів, у порівнянні з попередніми роками, коли вистачало і половини цієї суми [22].

Популярність Telegram також збільшила кількість рекламодавців, включаючи великі українські бренди як Comfy, АЛЛО, Сушия, Фокстрот, Сільпо, що використовують платформу як додатковий канал для реклами. У грудні 2023 року представники Епіцентр, Prom, Uklon та Rozetka навіть виступили на першій офлайн конференції про Telegram в Києві [13].

Рис. 1.2. Топ українських брендів, що розміщують рекламу в Telegram

Джерело: розроблено автором на основі [3]

Переважна більшість рекламодавців у Telegram складається з інших Telegram каналів, які прагнуть збільшити свою аудиторію. Їхня ціль – це набрати більше аудиторії і залучити вже великі бренди до замовлення реклами в їх каналах.

Також серед рекламодавців можна знайти локальні компанії, які розміщують оголошення в регіональних каналах у пошуках нових клієнтів. Чимало рекламних звернень надходить від організаторів освітніх курсів та заходів. Водночас, індустрії, такі як казино та букмекерські контори, що часто стикаються з проблемами у пошуку трафіку, також активно використовують Telegram.

У месенджері Telegram розрізняють два головних типи реклами: зовнішня реклама та офіційна реклама відома як «Telegram Ads».

Зовнішня реклама в Telegram вбудовується безпосередньо в потік контенту каналів або груп, виглядаючи як стандартні повідомлення, що їх може формувати адміністратор каналу. Цей тип реклами часто використовується для промоції товарів, послуг чи інших медіа ресурсів, і її успішність великою мірою залежить від того, наскільки ефективно реклама інтегрована в контекст каналу та наскільки вона залучає аудиторію.

На рис. 1.3. зображена реклама акцій тижня мережі магазинів «Сільпо». Реклама була розміщена о 10 ранку в новинному каналі мільйоннику «Телеграмна служба новин». Ця реклама була придбана за фіксовану ціну та без видалення через певний проміжок часу. Станом на лютий 2024 фіксована ціна за рекламний пост в цьому каналі складає 37 тисяч гривень.

Рис. 1.3. Приклад зовнішньої реклами Сільпо в Telegram каналі «Телеграмна служба новин»

Джерело: розроблено автором на основі [16]

Офіційна реклама, відома як Telegram Ads, це рекламні продукти, пропоновані самим Telegram. Цей сервіс був оголошений Павлом Дуровим у листопаді 2021 року і включає спонсовані оголошення, які не базуються на даних користувачів для націлювання, а розміщуються виключно в публічних каналах, що відповідають певній тематиці та мають понад 1000 учасників. Ці оголошення з'являються у стрічці постів каналу при певних діях користувача. Ці реклами керуються централізовано і забезпечують широке охоплення, хоча і мають менше можливостей для індивідуалізації в порівнянні з зовнішньою рекламою [11, 35].

На рис. 1.4. представлена реклама спільноти дизайнерів «Behance». Таку рекламу можуть бачити лише користувачі у яких немає Telegram Premium

акаунту при першому приєднанні до публічного каналу чи публікації каналом нового посту.

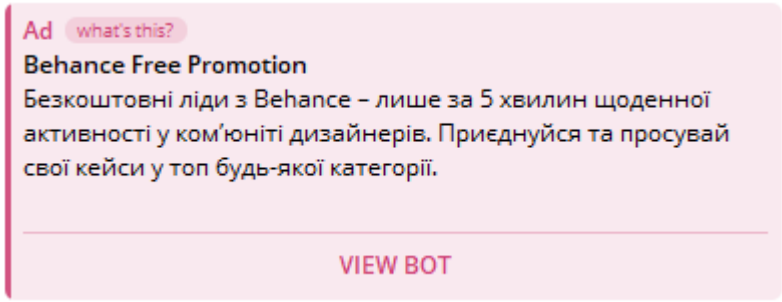

Рис. 1.4. Приклад реклами Behance розміщеної через офіційну платформу Telegram Ads в Telegram каналі «Creative Practice»

Джерело: розроблено автором на основі [5]

У цьому дослідженні ми зосередимо нашу увагу на прогнозуванні ефективності зовнішньої реклами. Обрано саме цей тип реклами, оскільки він вимагає більшої креативності та глибокої взаємодії з аудиторією. Зовнішня реклама також має численні компоненти, які можуть бути аналізовані та оптимізовані, зробивши її особливо цікавою для дослідження ефективності рекламних кампаній.

1.3. Аналітика реклами в Telegram

При аналітиці реклами рекламодавці часто користуються спеціалізованими аналітичними платформами, такими як Telemetrio. Цей сервіс надає важливу інформацію про канал, включаючи кількість активних підписників, перегляди та репости публікацій, а також динаміку зростання аудиторії каналу. Така інформація є критично важливою на ринку, де існують практики штучного створення каналів та імітації активності для продажу реклами новачкам без досвіду [17].

На ринку реклами в Telegram сформувалися власні метрики для виміру ефективності рекламних кампаній. Основні показники включають перегляди постів, кількість підписок на рекламований канал, а також вартість залучення одного нового підписника. Як зазначає Єгор Зінченко, власник мережі регіональних Telegram каналів та блогу «Адмінські Угіддя», вартість підписника прямо залежить від якості рекламного посту [10].

Основним показником вартості реклами є охоплення поста протягом перших 24 годин. Рекламодавці часто висувають вимогу не перевищувати вартість 120 CPM (Cost Per Mille, в нашому контексті — вартість за тисячу показів). Вартість CPM може значно варіюватися в залежності від тематики каналу: наприклад, в новинах, історії та гумористичних каналах вона може досягати до 150 CPM, тоді як у криптовалютних каналах — до 3500 CPM, в бізнес- та маркетингових каналах — 1000-1200 CPM. Такі варіації підкреслюють необхідність детального вивчення цільової аудиторії та вибору відповідного каналу для максимізації рекламного бюджету [10].

Основа рекламної кампанії в Telegram це:

1. Рекламний креатив: текст, емодзі, хештеги, медіа.
2. Час та дата публікації.
3. Тип розміщення допису: з видаленням чи без, на 24 чи на 48 годин, час знаходження в «топі».
4. ER посту – коефіцієнт залученості аудиторії.

Здатність прогнозувати ефективність розміщення реклами є ключовою, і існує багато гіпотез про те, які компоненти можуть впливати на основний показник рекламної активності – залучення аудиторії (ER). У цій роботі ми спробуємо підтвердити або спростити гіпотези, чи наступні фактори, що контролюються адміністратором, впливають на ефективність реклами:

a. Загальна кількість символів у пості: впливає на зміст і зрозумілість повідомлення.

b. Час публікації посту: згідно з дослідженнями аналітичного сервісу Telemetrio, найпопулярніший час для публікації реклами в Україні — це 10:00 ранку [3].

c. День тижня публікації посту: найпопулярніший день для розміщення реклами, згідно з Telemetrio — вівторок та середа [3].

d. Кількість картинок та GIF-зображень у пості: вплив на візуальну привабливість.

e. Кількість відео та аудіо файлів та їх тривалість: може вплинути на час утримання уваги аудиторії.

f. Кількість емодзі та хештегів у пості: важливі для привернення уваги та підвищення взаємодії.

g. Формат цільового посилання: може бути вставлене повністю у текст, інтегроване в слово або вшите в кнопку під постом.

Крім цього, існують більш загальні критерії якості для рекламного посту, такі як його рівень креативності та відповідність тематиці каналу. На жаль, ми не можемо точно виміряти ці аспекти, оскільки вони суб'єктивні за своєю природою. Проте ми будемо враховувати їх під час формулювання висновків щодо роботи.

Крім вказаних залежних від адміністратора каналу компонентів, існують і незалежні фактори, які є результатами рекламної активності:

1. ER посту: відображає рівень залученості аудиторії.
2. Кількість репостів: вказує на розповсюдження контенту.
3. Кількість реакцій: показник взаємодії з постом.

На рис. 1.5. продемонстровані елементи вже опублікованого рекламного креативу.

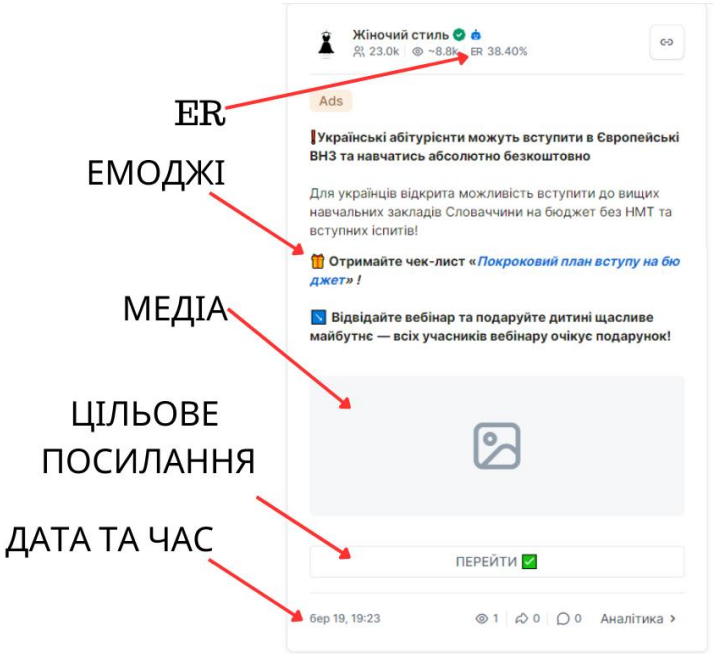

Рис. 1.5. Елементи рекламного креативу

Джерело: розроблено автором на основі [15]

Висновки до розділу 1

У розділі 1 кваліфікаційної роботи бакалавра, присвяченому аналізу розміщення реклами в месенджері Telegram в Україні, ми розглянули динамічний розвиток цифрового маркетингу та стрімке зростання популярності месенджеру у контексті збільшення українських користувачів з 20% у 2021 році до 72% у 2023 році.

Платформа стала важливим майданчиком не лише для спілкування, але й для рекламної діяльності, з вартістю рекламних публікацій, яка може коливатись від 80 до 1000 гривень за тисячу переглядів, залежно від охоплення аудиторії та активності. Таким чином, визначаючи розміщення реклами в Telegram як високоприбуткову справу.

У підсумку, хочеться наголосити на важливості визначити фактори, які можуть покращити результативність рекламних кампаній в Telegram як для власників каналів так і для рекламодавців. Саме таке завдання поставлено нами для розділів 2 та 3.

РОЗДІЛ 2
АНАЛІТИКА ТА ПІДГОТОВКА БАЗИ ДАНИХ РЕКЛАМНИХ ПОСТІВ

2.1. Методологія CRISP-DM

Тип розміщення допису залежить від специфіки аудиторії в каналі, а от з приводу оформлення рекламного креативу та часу його розміщення існує дуже багато «порад від експертів» та гіпотез. Однак усі вони безпідставні, що ми і мали на меті виправити. Для цього ми пройдемося по класичним етапам методології CRISP-DM.

CRISP-DM (Cross-Industry Standard Process for Data Mining) - це стандартний процес для дослідження даних, який документований і доступний для всіх. Він включає опис типових фаз проекту з дослідження даних, завдань у складі кожної з цих фаз і пояснення взаємозв'язків між завданнями, що дозволяє використовувати методи інтелектуального аналізу даних. CRISP-DM описується в термінах ієрархічної моделі процесу [4, 34].

Модель включає набір завдань, які розглядаються на чотирьох рівнях абстракції - від загальних до конкретних як на рис. 2.1. На верхньому рівні процесу Data Mining визначаються кілька фаз розробки, кожна з яких включає кілька загальних завдань, що становлять другий рівень ієрархії. Цей рівень є загальним, оскільки завдання, що його складають, не враховують особливостей конкретної сфери застосування. Третій рівень - це спеціалізовані завдання, які включають опис кроків, необхідних для адаптації загальних завдань до вирішення конкретних проблем. На четвертому рівні представлені конкретні дії, рішення і результати, що зустрічаються в процесі Data Mining.

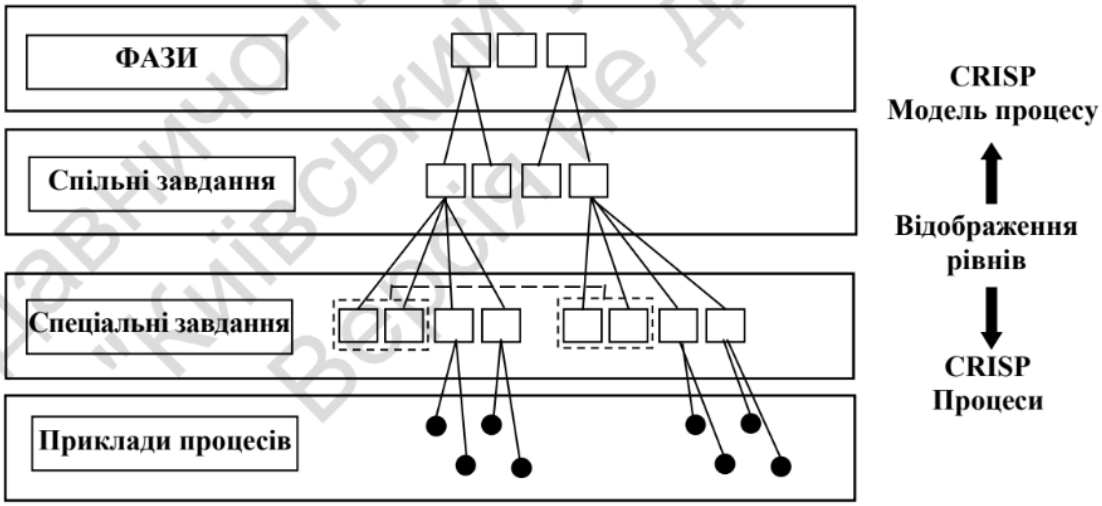

Рис. 2.1. Чотири рівні методології CRISP-DM

Джерело: розроблено автором на основі [4]

Розглянемо модель CRISP-DM на рис. 2.2. Вона складається з шести фаз, а стрілки позначають найважливіші зв'язки між ними. Послідовність цих фаз не є строгою, а зовнішні стрілки, що мають циклічний характер, відображають спіралеподібний процес розробки проектів Data Mining.

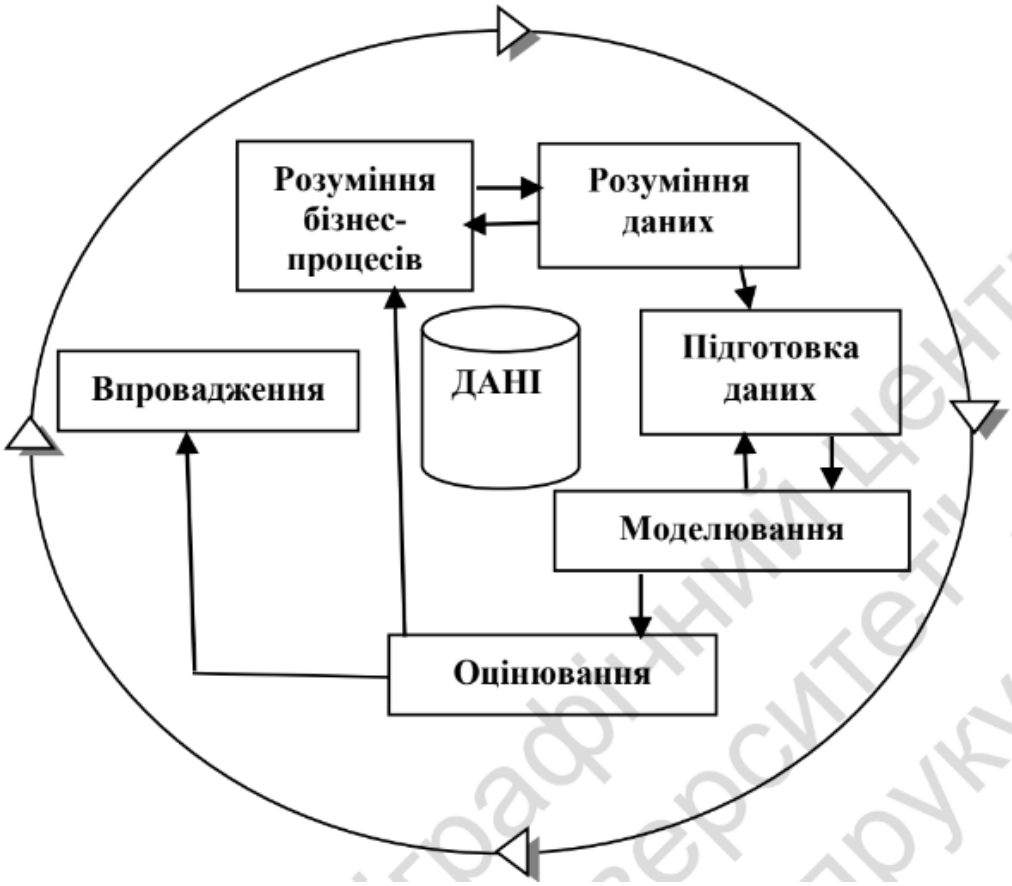

Рис. 2.2. Шість фаз моделі CRISP-DM

Джерело: розроблено автором на основі [4]

Під час цієї фази розуміння бізнес-процесів (Business Understanding) визначаються бізнес-цілі проекту, проводиться оцінка ситуації, встановлюються цілі дослідження даних та створюється план проекту. Важливо надати належну увагу цілям проекту з точки зору бізнесу, визначити, які саме дані потрібно проаналізувати та сформулювати стратегію досягнення цих цілей. Саме це і описано у вступній частині цієї роботи та першому розділі.

Фаза розуміння даних (Data Understanding) починається зі збору початкових даних, опису даних, їхнього дослідження та перевірки якості. Ця фаза спрямована на вивчення вихідних даних, виявлення проблем якості даних та ідентифікацію цікавих аспектів для подальшого аналізу. Цій фазі присвячений розділ 2.3 та 2.4.

Фаза підготовки даних (Data Preparation) охоплює весь комплекс дій, пов'язаних з підготовкою даних для подальшого аналізу. Сюди входять вибір даних, їх очищення, конструювання, інтеграція та форматування. Так у розділі 3.1. ми розділяємо дані на тестову та тренувальну множини та нормалізуємо їх, позбавляючись від викидів.

Фаза моделювання (Modeling) має на меті вибір оптимального методу побудови моделей та налаштування їхніх параметрів для отримання оптимальних рішень. На цьому етапі здійснюється вибір методу моделювання, генерація тестового проекту, побудова моделей та оцінка їхньої ефективності. Використані регресійні моделі та особливості моделювання на основі бази даних рекламних постів містяться у розділі 3.2.

Фаза оцінювання (Evaluation) дозволяє остаточно оцінити модель перед її впровадженням, щоб переконатися у досягненні бізнес-цілей. Тут проводиться оцінка результатів, перегляд процесу та визначення подальших кроків. Таким чином основою для розділу 3.3. стало оцінювання точностей побудованих моделей та формування проміжних висновків до роботи.

Фаза впровадження (Deployment) необхідна для організації знань та їх подання у зручному для користувача форматі. На цьому етапі здійснюється планування впровадження, моніторинг та обслуговування, створення

підсумкового звіту та підсумковий огляд проекту. Ця фаза реалізована паралельно з фазою моделювання, адже одразу після тренування моделі був запущений і процес тестування за рахунок поділу бази даних на дві множини.

Співпрацюючи з сервісом аналітики Telemetrio ми зібрали базу даних, що складається з 30 тисяч рекламних креативів українського ринку Telegram. Для ефективної обробки цієї інформації і досягнення поставлених цілей, ми розробили і пройшли шість фаз відповідно до методології CRISP-DM , що зображені на рис. 2.3.

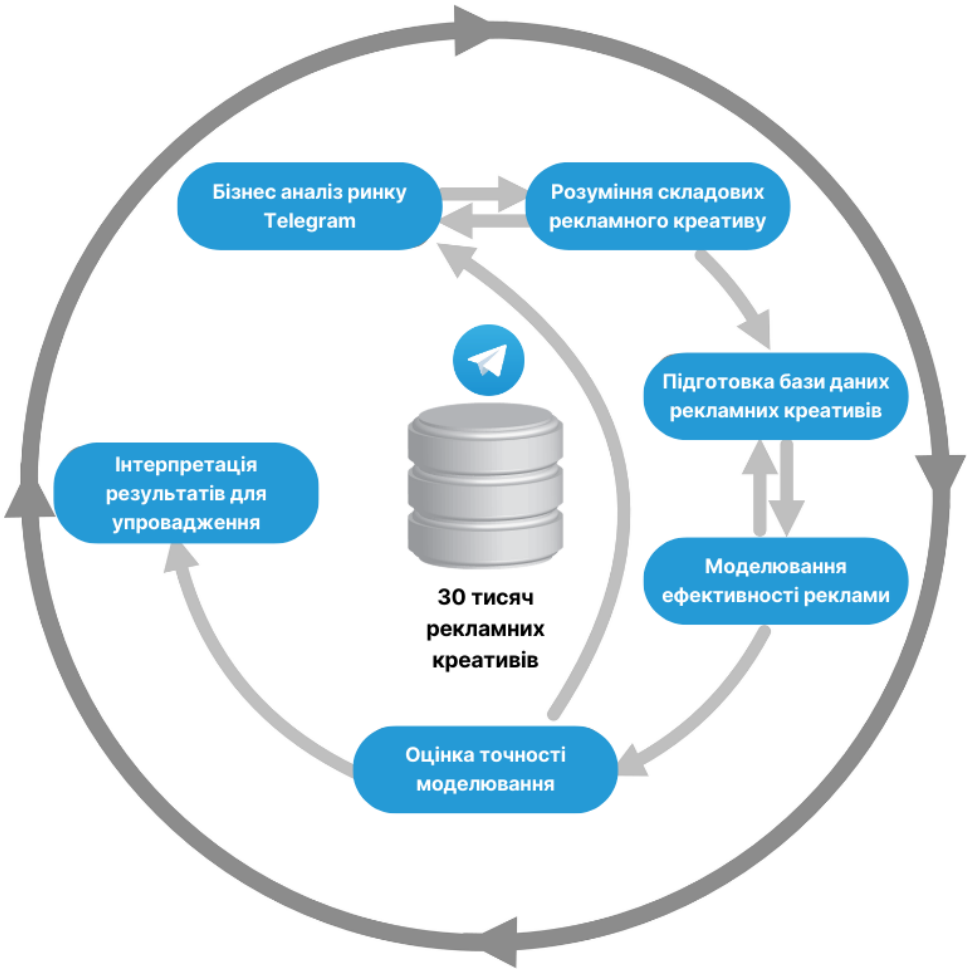

Рис. 2.3. Фази виконання завдань роботи за методологією CRISP-DM

Джерело: розроблено автором

2.2. KNIME Analytics як альтернатива IBM SPSS Modeler

Протягом навчального процесу на спеціальності економічна кібернетика ми користувалися IBM SPSS Modeler для аналізу даних. Проте, станом на 2024 рік, ця програма та її інтерфейс, що зображений на рис. 2.4., вважається застрілим.

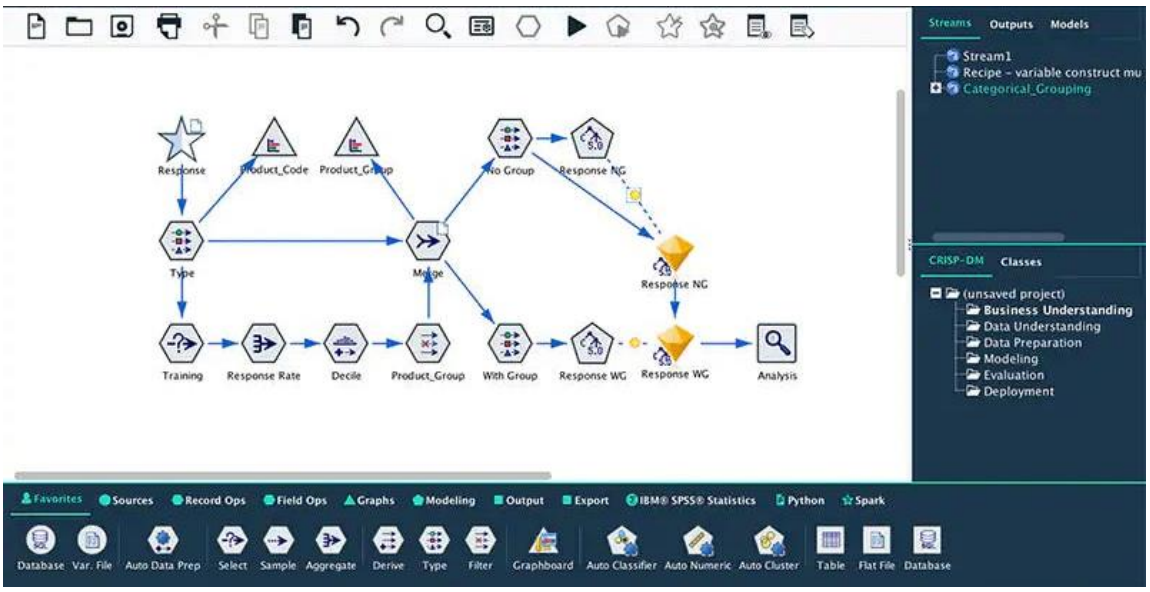

Рис. 2.4. Інтерфейс платформи «IBM SPSS Modeler»

Джерело: розроблено автором на основі [4]

Базовий функціонал платформи доступний лише протягом першого місяця використання, після чого доступ до раніше створених проєктів припиняється. При роботі з базами даних, що містять більше 5 тисяч рядків, навіть найпростіші лінійні моделі завантажуються протягом 5 хвилин, а для моделей класифікації нейронних мереж на звичайних офісних комп’ютерах може знадобитися від 8 до 10 годин.

У порівнянні з IBM SPSS Modeler, KNIME Analytics видається більш сучасним та доступним інструментом для аналізу даних. Його інтерфейс є інтуїтивно зрозумілим та зручним для користувачів будь-якого рівня, що можна побачити на рис. 2.5. Однією з переваг KNIME є те, що він безкоштовний, що робить його доступним для широкого кола користувачів.

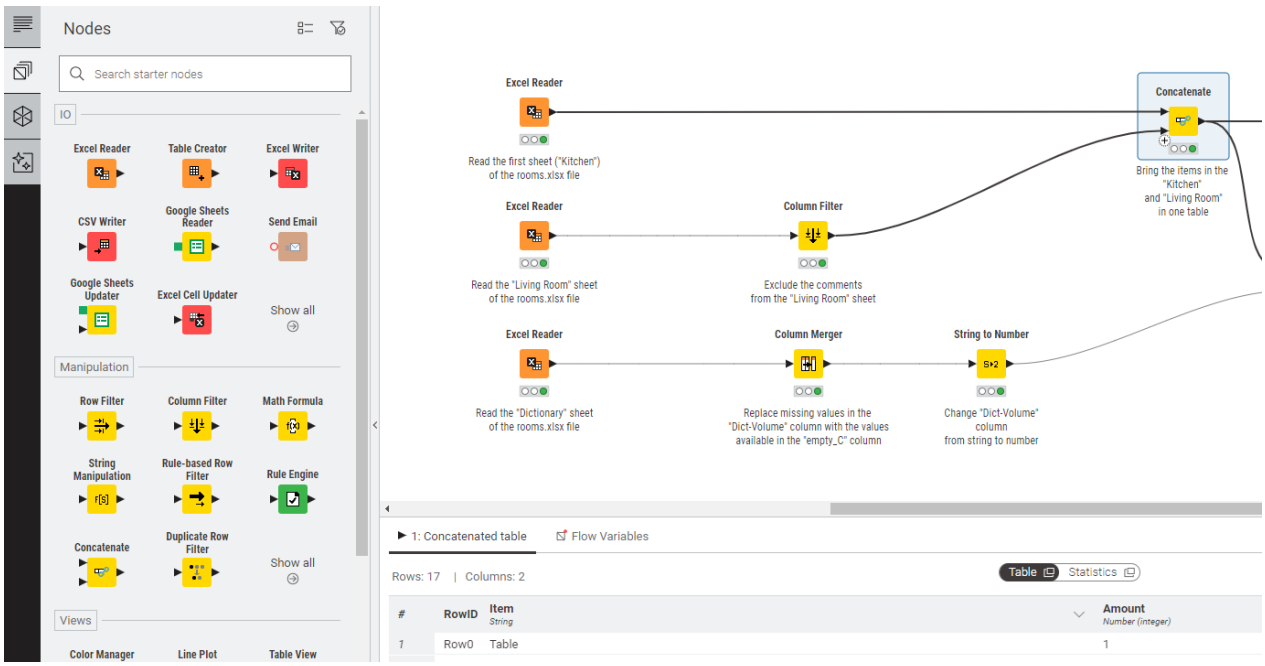

Рис. 2.5. Інтерфейс платформи «KNIME Analytics»

Джерело: розроблено автором на основі [1]

Окрім цього, KNIME також забезпечує своїх користувачів багатофункціональними можливостями та різноманітними інструментами для аналізу даних. Блог платформи містить велику кількість корисних статей, порад та інструкцій, що допомагають користувачам максимально використовувати потенціал програми для вирішення різних завдань аналізу даних [26].

У цьому контексті KNIME Analytics виступає як краща альтернатива. KNIME надає широкий спектр нод (вузлів), які дозволяють проводити різноманітний вивід, аналіз, інтерпретацію, візуалізацію даних [27].

Наприклад, для прогнозування ефективності рекламних кампаній на українському ринку Telegram на основі складових рекламних креативів, ми використовуватимемо такі вузли, як "Data Reader" для завантаження даних, "Data Preprocessing" для очистки та підготовки даних, "Modeling" для побудови прогностичних моделей та "Evaluation" для оцінки їх точності.

2.3. База даних рекламних постів

Використовуючи базу даних з 30 тисяч рекламних дописів українських Telegram каналів, надану сервісом аналітики Telemetrio, ми дослідимо як різні

параметри креативу впливають на один із ключових показників успішності рекламного допису — коефіцієнт залученості (ER).

Коефіцієнт залученості (Engagement Rate) важливий, адже він вимірює середню кількість переглядів одного поста у відношенні до загальної кількості підписників на каналі. На прикладі інфографіки на рис. 2.6., що є типовою для рекламодавців, можна оцінити потенційну віддачу від рекламного допису. Саме на цей показник свою увагу звертають рекламодавці при закупівлі реклами.

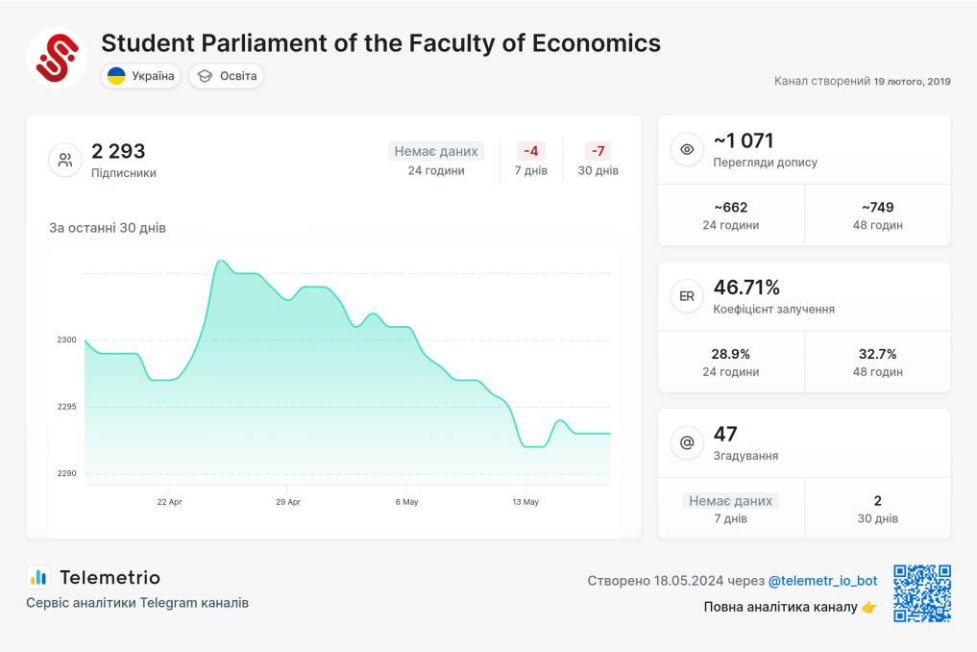

Рис. 2.6. Статистика картинкою каналу студпарламенту Економічного факультету КНУ ім. Т. Шевченка

Джерело: розроблено автором на основі [19]

Детально розглянемо наступні фактори, що за нашим припущенням впливають на ER:

1. Тип медіаконтенту: аналізуємо ефективність різних форматів, включаючи окремі зображення, колажі, короткі та довгі відео, аудіофайли та GIF-зображення.

2. Кількість символів у тексті: дослідимо, як довжина тексту корелює з ER посту.

3. Використання емодзі: з'ясуємо оптимальну кількість емодзі для досягнення збалансованого рівня залучення.

4. Формат цільових посилань: порівняємо ефективність прямих URL, посилань, інтегрованих у текст, та кнопок з посиланнями.

5. Час та день публікації: визначимо, які часові рамки та дні тижня є найбільш ефективними для публікації, щоб забезпечити максимальну залученість.

Таким чином, наша початкова база даних складається з 30 099 рядків, що відповідає кількості креативів та з 20 стовпців: id каналу, id допису, кількість підписників на каналі, дата публікації посту, час публікації, день тижня, текст посту, сумарна кількість переглядів посту і відповідно відсоток залученості (ER), кількість репостів та реакцій, кількість символів у пості, кількість емодзі, хештегів, фото, відео, аудіо, gif-файлів, загальна тривалість відео та аудіо матеріалів в секундах та тип рекламного посилання. Структуру бази даних можна переглянути у Додатку А.

Для аналізу бази даних рекламних дописів українських Telegram каналів у середовищі KNIME Analytics, використовуючи ноду «Excel Reader» для завантаження даних, ми застосовуємо 2 типи аналітичних нод для обробки інформації.

Використання ноди «Value Counter» у KNIME Analytics дозволяє нам аналізувати кількість унікальних значень у стовпчику. Завдяки цьому ми визначаємо унікальну кількість каналів, яка дорівнює 3 746. Це кількість каналів, пости яких були ідентифіковані як рекламні за допомогою унікальної моделі штучного інтелекту і були включені до досліджуваної бази в 30 099 креативів.

Інша нода «Statistics» здійснює обчислення статистичних характеристик, включно з мінімальними та максимальними значеннями, середнім значенням, стандартним відхиленням, дисперсією, медіаною, загальною сумою, кількістю пропущених значень, а також загальною кількістю рядків у всіх числових стовпчиках, паралельно здійснюючи підрахунок всіх номінальних значень разом із їх частотами входження. Нижче представлені основні дані та висновки, що були згенеровані на основі результатів аналізу нодою нашої бази даних.

Найчастіший час для публікації рекламних дописів — 8 ранку. Для візуального представлення також був побудований графік на рис. 2.7. За графіком складно визначити в який саме час були опубліковані пости з найвищим значенням ER, але можна висунути гіпотезу про 2 хвилі: до обіду та після.

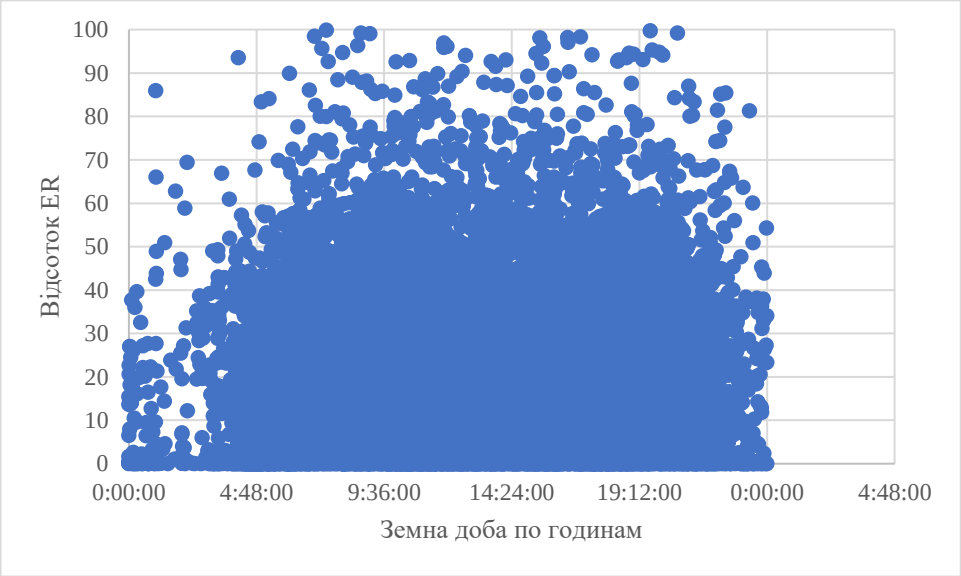

Рис. 2.7. Графік співвідношення параметрів «час публікації реклами» та значення відсотку ER постів

Джерело: розроблено автором

За статистичними даними, середа виявилася найбільш популярним днем для публікацій, що відповідає результатам дослідження від Telemetrio за грудень 2023 року, які вказують на високу ймовірність зустрічі з рекламою у вівторок та середу, з середньою кількістю публікацій близько 11 тисяч [3].

З четверга кількість публікацій знижується, досягаючи мінімуму в понеділок наступного тижня з приблизно 7.5 тисячами рекламних дописів [3]. За графіком на рис. 2.8. поки складно сказати про кореляцію дня тижня з досліджуваний показником ER.

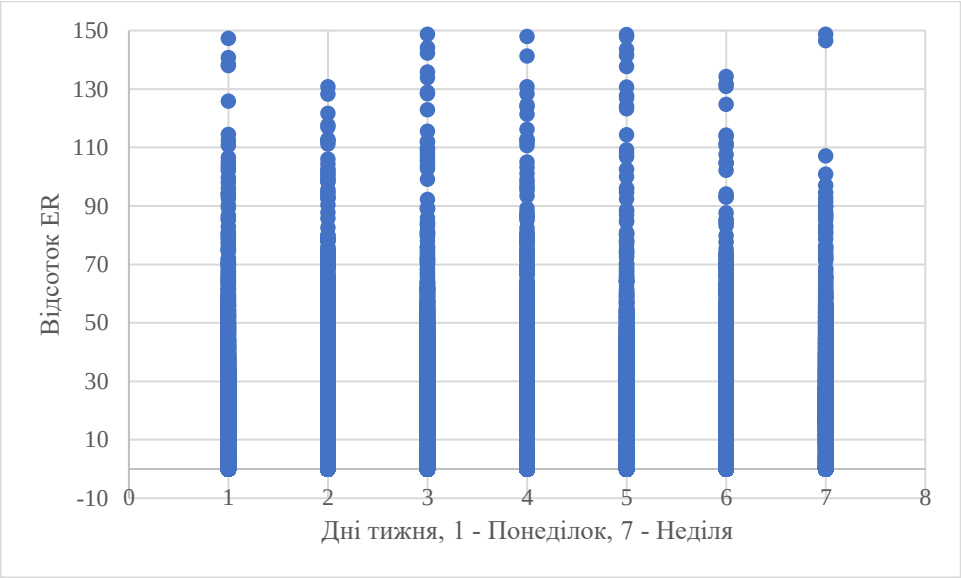

Рис. 2.8. Графік співвідношення параметру «день тижня публікації поста» до значення відсотку ER

Джерело: розроблено автором

Переважна більшість постів має довжину 108 символів. За графіком на рис. 2.9. можемо висунути гіпотезу, що пости до 1000 символів мають вищий показник ER ніж пости від 1 тисячі символів і більше.

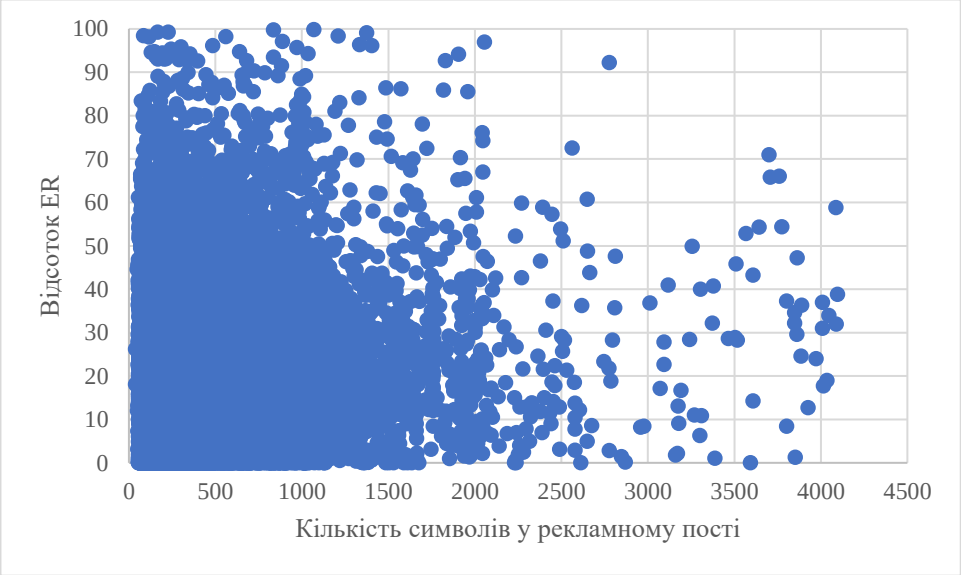

Рис. 2.9. Графік співвідношення параметру «кількість символів у рекламному пості» до значення відсотку ER

Джерело: розроблено автором

Типова кількість емодзі в постах — 2, тоді як хештеги зазвичай відсутні. На рис. 2.10. та 2.11. представлено співвідношення кількості емодзі та хештегів в

пості до їх відсотку залучення. 88% постів досліджуваних постів не містять хештегів взагалі.

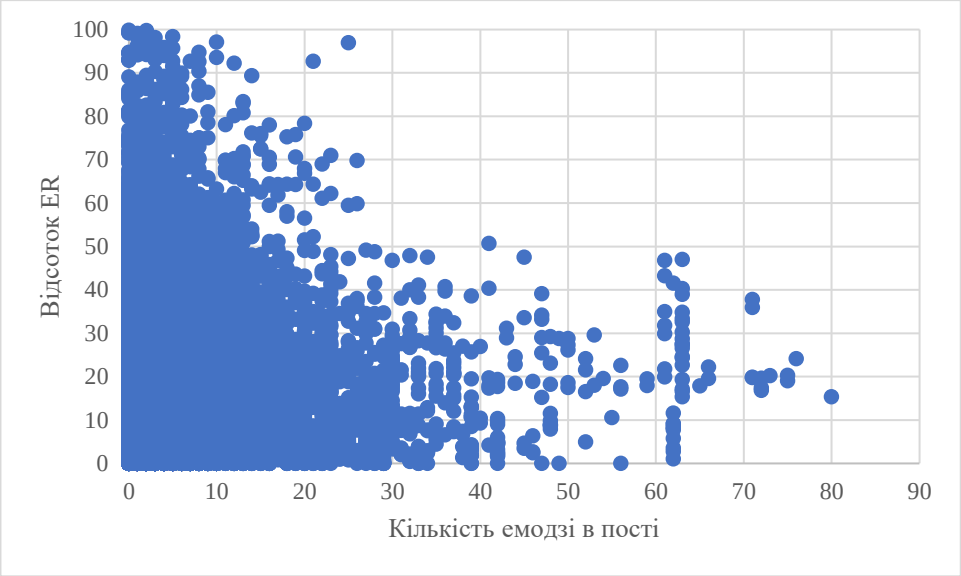

Рис. 2.10. Графік співвідношення параметру «кількість емодзі у пості» до значення відсотку ER

Джерело: розроблено автором

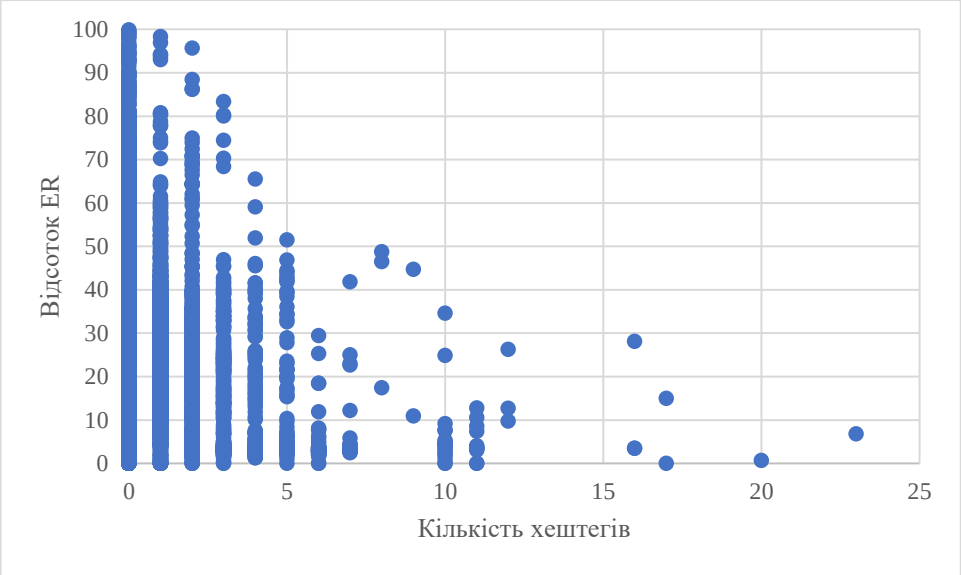

Рис. 2.11. Графік співвідношення параметру «кількість хештегів у пості» до значення відсотку ER

Джерело: розроблено автором

У 99% дописів немає аудіо, то ж ми не беремо цей показник у дослідження. Фотографії присутні у кількості одна або жодної; gif-зображення також відсутні.

Співвідношення кількості фото та відео з відсотком залученості можна переглянути на рис. 2.12 та 2.13.

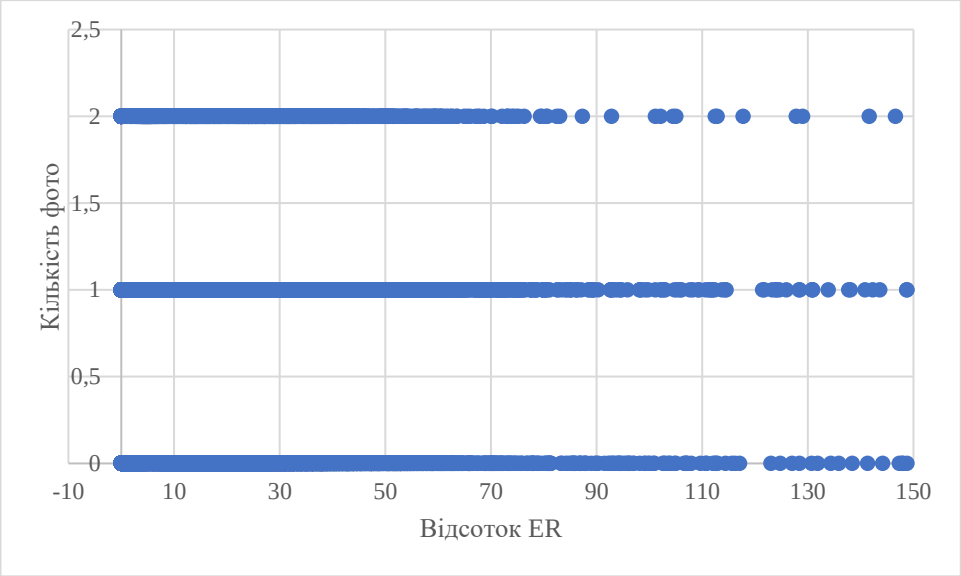

Рис. 2.12. Графік співвідношення параметру «кількість фото у рекламному пості» до значення відсотку ER

Джерело: розроблено автором

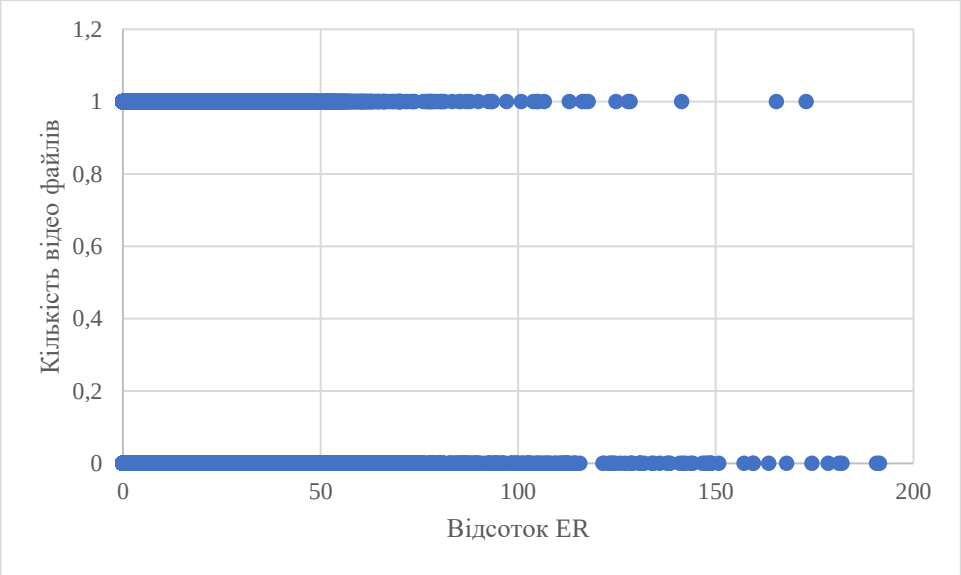

Рис. 2.13. Графік співвідношення параметру «кількість відео у рекламному пості» до значення відсотку ER

Джерело: розроблено автором

Цільові посилання зазвичай вмонтовані в кнопки під постами. Це відповідає типу посилання №3 і таких 13 331 пост. Вшивають посилання у слово майже на

стільки часто – 11 462 постів типу №2. Тип посилання №1, коли посиланням вставляються без редагування у текст, у 2 рази менше від типу №2 – 5 305 постів.

Кореляцію з відсотком ER можна переглянути на графіку на рис. 2.14. та висунути гіпотезу про вищий коефіцієнт залученості для посилання тип №2.

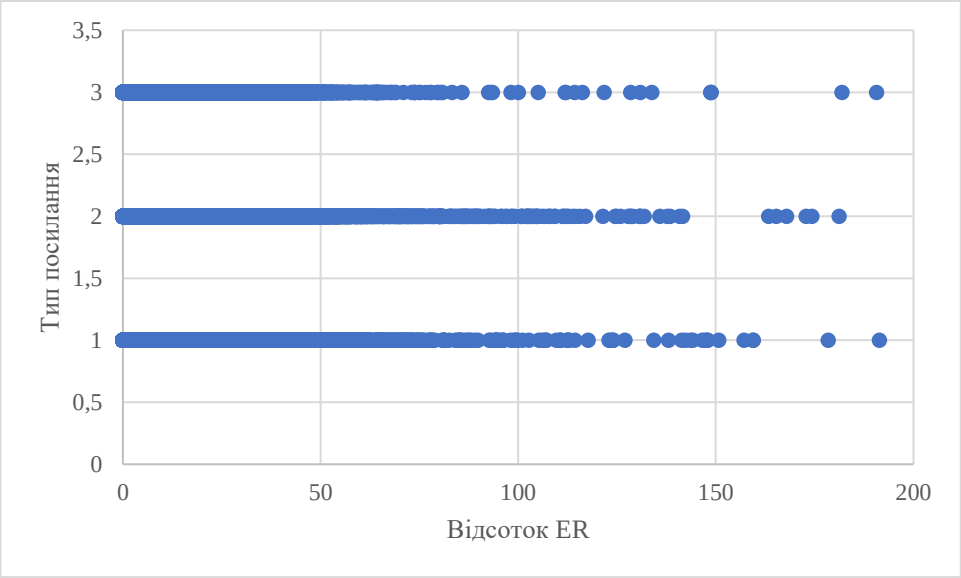

Рис. 2.14. Графік співвідношення параметру «тип цільового посилання» до значення відсотку ER

Джерело: розроблено автором

2.4. Кореляційний аналіз

За допомогою аналітичних нод KNIME Analytics ми проаналізуємо спочатку кореляцію складових рекламного допису з відсотком його залученості. Таким чином, ми ставимо собі за мету визначити підходящі елементи посту, на які може впливати власник каналу, для прогнозування значення ER. Для визначення взаємозалежності показників, ми обрали лінійну та рангову кореляційні моделі.

Лінійна кореляція, відома також як коефіцієнт кореляції Пірсона, використовується для вимірювання ступеня лінійної залежності між двома змінними, де вона вказує, наскільки зміна однієї змінної лінійно відповідає зміні іншої змінної. Цей коефіцієнт припускає, що обидві аналізовані змінні є кількісними, що дозволяє встановити, чи зростає або зменшується одна змінна у відповідь на зміну іншої [1].

Відповідно до документації KNIME Analytics обчислення лінійної кореляції здійснюється за формулою (2.1) [7, 28].

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad (2.1)$$

де n – розмір вибірки;

x_i, y_i – окремі точки вибірки з індексом і;

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ – (середнє значення вибірки); і аналогічно для $\bar{y}$.

Налаштування ноди для визначення лінійної кореляції представлене на рис. 2.15. З огляду на особливість лінійної кореляційної моделі, ми перевили значення часу та дня тижня публікації в кількісне значення.

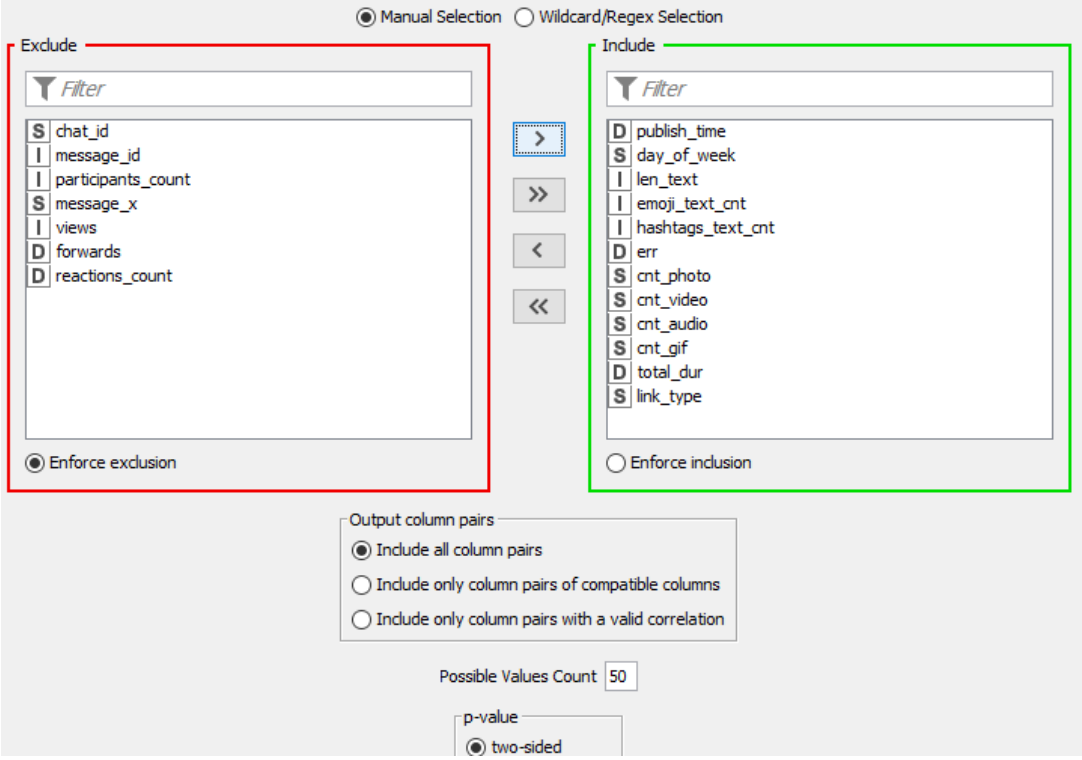

Рис. 2.15. Налаштування ноди «Лінійна кореляція»

Джерело: розроблено автором

Результати кореляційного аналізу представлені на рис. 2.16. Згідно кольорового позначення жоден з показників не корелює суттєво з ключовим показником залучення ER. Лише 2 показники мають додатнє значення, а саме «кількість символів у пості» - 0,084 та загальна тривалість аудіо/відео матеріалів доданих до посту – 0,095.

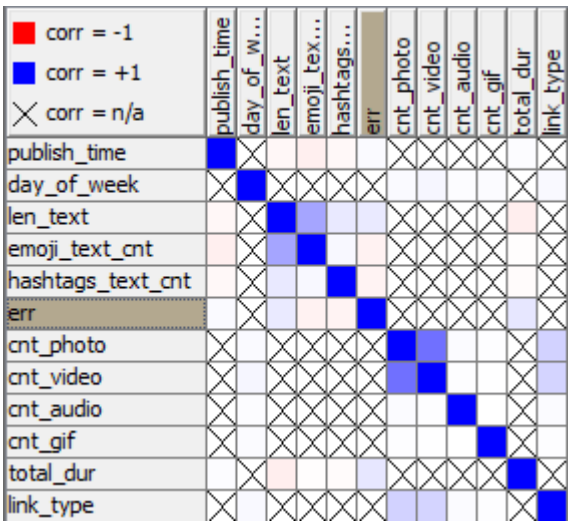

Рис. 2.16. Лінійна кореляційна матриця

Джерело: розроблено автором

Наступна вже рангова кореляція, представлена коефіцієнтами кореляції Спірмена та Кендалла, призначена для вимірювання монотонної залежності між двома змінними, дозволяючи оцінити, наскільки добре можливо описати зв'язок між ними за допомогою монотонної функції, незалежно від того, чи є цей зв'язок лінійним. Корисність цього підходу полягає в тому, що він застосовний не тільки до кількісних, але й до порядкових (рангових) змінних, роблячи його універсальним інструментом для аналізу даних, де точний тип залежності між змінними не є критичним.

Коефіцієнт Спірмена підходить для безперервних та дискретних змінних, у тому числі порядкових і тому нода застосовує формулу (2.2) у випадках, коли у вибірці з’являються такі дані [9, 32]. У нашому випадку це стосується днів тижня та часу публікації.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{2.2}$$

де n – розмір вибірки;

x_i, y_i – окремі точки вибірки з індексом i;

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ – (середнє значення вибірки); і аналогічно для $\bar{y}$.

Для вимірювання сили зв'язку між двома вимірюваними величинами використовується гамма- та тау-коефіцієнт кореляції Кендала за формулою (2.3) [8].

$$r = \frac{s_1 - s_2}{\frac{1}{2}n(n-1)} \tag{2.3}$$

де n – розмір вибірки;

s_1 – кількість узгоджених пар (якщо $x_i > x_j$ та $y_i > y_j$ або якщо $x_i < x_j$ та $y_i < y_j$)

s_2 – кількість неузгоджених пар (якщо $x_i > x_j$ та $y_i < y_j$ або якщо $x_i < x_j$ та $y_i > y_j$)

Налаштування ноди для визначення рангової кореляції представлене на рис. 2.17.

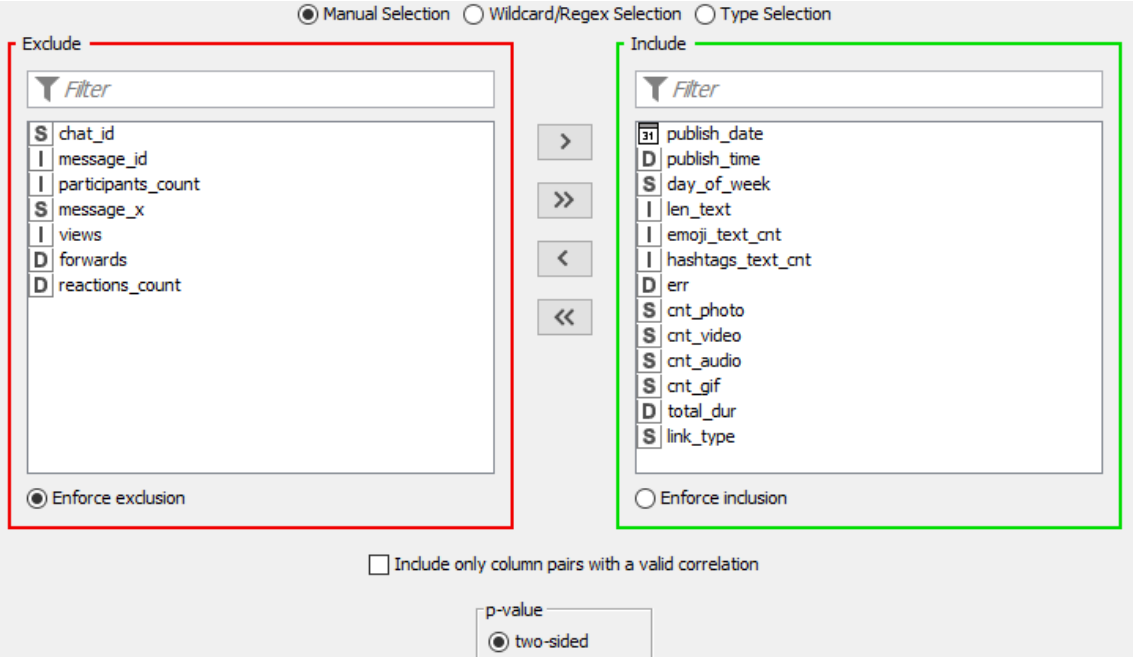

Рис. 2.17. Налаштування ноди «Рангова кореляція»

Джерело: розроблено автором

Результати рангового кореляційного аналізу представлені на рис. 2.18. Згідно кольорового позначення жоден з показників не корелює сильно з ключовим показником залучення ER. Лише 4 показники мають додатнє значення, а саме «кількість символів у пості» - 0,1, кількість фото та відео відповідно 0,085 та 0,092 та загальна тривалість аудіо/відео матеріалів доданих до посту – 0,1.

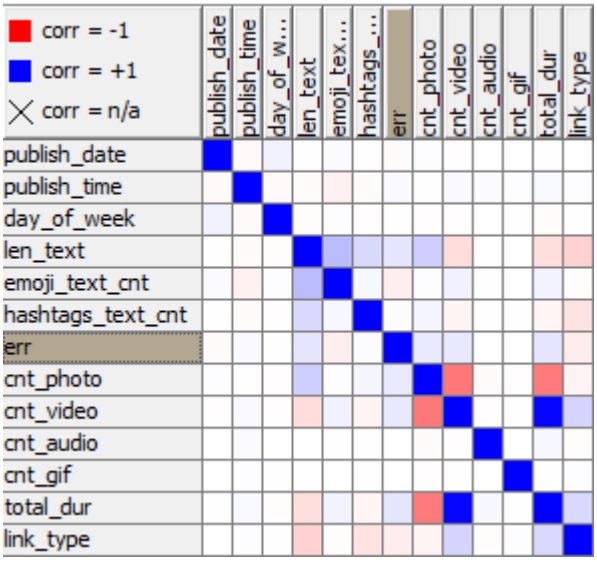

Рис. 2.18. Рангова кореляційна матриця

Джерело: розроблено автором

Висновки до розділу 2

У другому розділі кваліфікаційної роботи бакалавра виконано аналітичну підготовку та обробку бази даних рекламних постів з використанням методології CRISP-DM та інструментарію KNIME Analytics, що забезпечує всебічний аналіз даних.

Через широкий спектр нод в KNIME, таких як "Data Reader", "Data Preprocessing", "Modeling" та "Evaluation", здійснено детальну обробку 30 тисяч рекламних дописів українських Telegram каналів. Аналіз показав, як різні параметри, включаючи тип медіа контенту, кількість символів, використання емодзі, формати цільових посилань, час і день публікації, співвідносяться з коефіцієнтом залученості (ER).

Результати ж кореляційного аналізу, виконаного за допомогою лінійної та рангової кореляції, вказують на слабкий зв’язок між більшістю вивчених параметрів і показником залученості. Таким чином, для подальшого прогнозування показнику ER ми залишаємо такі 8 факторів як час публікації посту, день тижня, кількість символів, емодзі, хештегів, фото, відео та тип цільового посилання.

РОЗДІЛ 3
ПРОГНОЗУВАННЯ ЕФЕКТИВНОСТІ РЕКЛАМИ В TELEGRAM В УКРАЇНІ

3.1. Підготовка даних до моделювання

Увесь процес аналізу даних, їх підготовки до моделювання та побудова самих моделей зображений в Додатку Б. Кожен процес розміщений у блоці та відповідно підписаний. Пройдемося по двом важливим підготовним етапам: розподіл бази на тренувальну й тестову множини та нормалізацію даних.

На базі даних із 30 тисяч постів виконано поділ на тестову та тренувальну множини у співвідношенні 20% до 80% відповідно, використовуючи ноду «Partitioning». Згідно з налаштуваннями ноди, зображеними на рисунку 3.1, ми використали метод випадкового відбору (Draw randomly), що дозволяє рівномірно і випадково розподілити пости між множинами без прив’язки до конкретних атрибутів. Такий метод забезпечує неупереджений вибір даних для тренувальної та тестової множин, важливий для об'єктивної оцінки моделі.

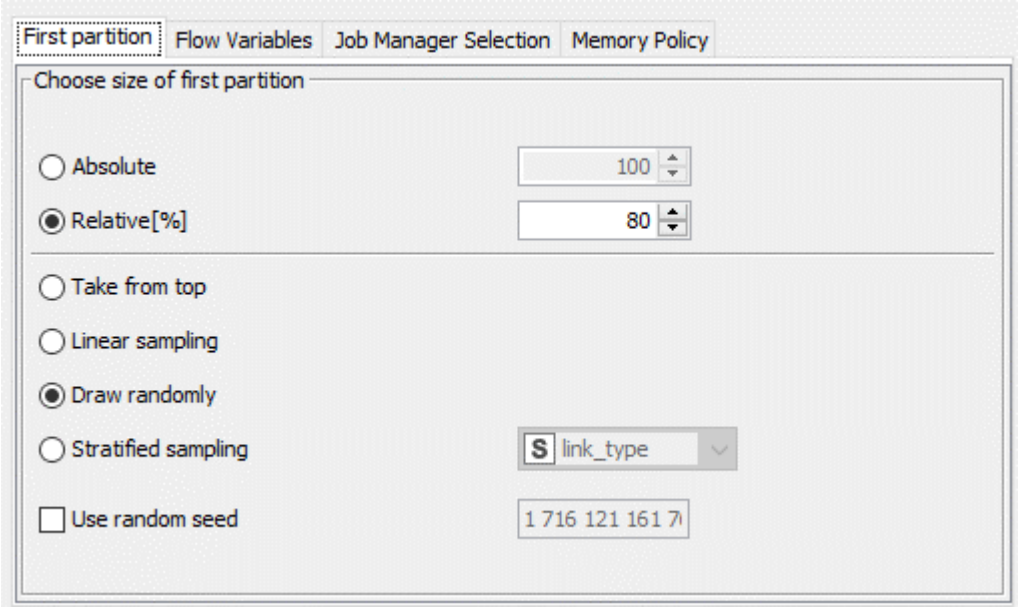

Рис. 3.1. Налаштування ноди для розподілу бази даних на тренувальну та тестову множини

Джерело: розроблено автором

Наступним кроком у підготовці даних стала нормалізація для підвищення точності моделі, яка передбачала видалення аномалій у даних. Наша база даних була попередньо очищена від пустих рядків, однак важливим кроком залишалося

видалення викидів, таких як надмірно довгі пости, пости з великою кількістю емодзі, або з численними хештегами, які були ідентифіковані в розділі 2.3 за допомогою графічного аналізу.

З метою корекції таких викидів було використано налаштування в ноді для вибору аномалій, представлені на рисунку 3.2. Викиди були відфільтровані шляхом заміни на найближчі допустимі значення, що дозволяє уникнути спотворення моделі вибраними екстремальними даними.

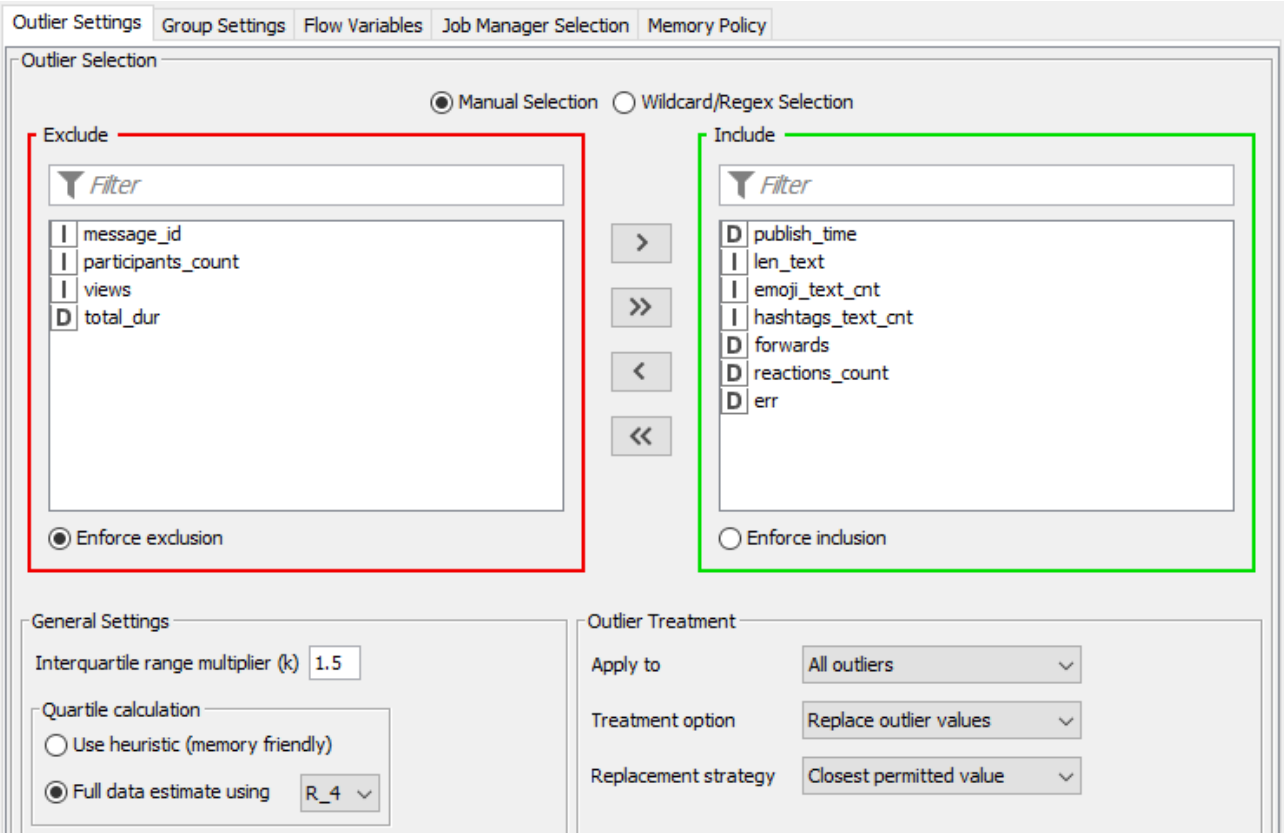

Рис. 3.2. Налаштування ноди для нормалізації даних

Джерело: розроблено автором

Таким чином, ми отримали 2 ноди «Normalizer» та « Normalizer (Apply)» для тренування та застосування моделі відповідно. Множина даних для тренування моделі після нормалізації склала 24078 постів та 6020 креативів для тестування.

3.2. Моделювання ефективності рекламного допису

Загалом тестувалися на платформі усі доступні для побудови регресії моделі, однак лише 4 з 7 показали адекватні результати та відповідали нашим запитам аналізувати як числові так і текстові дані та прогнозувати числове значення ER.

Кількість факторів для прогнозування відсотку ER– 8: день тижня та час публікації креативу; кількість символів, емодзі та хештегів, фото та відео в пості; тип цільового посилання (3 типи).

На рис. 3.3. червоним позначені 3 моделі, що виявилися не придатними для наших параметрів і показали від'ємні результати точності прогнозування, в той час як зеленим позначені 4 моделі, які ми беремо в роботу та будемо покращувати їх показники.

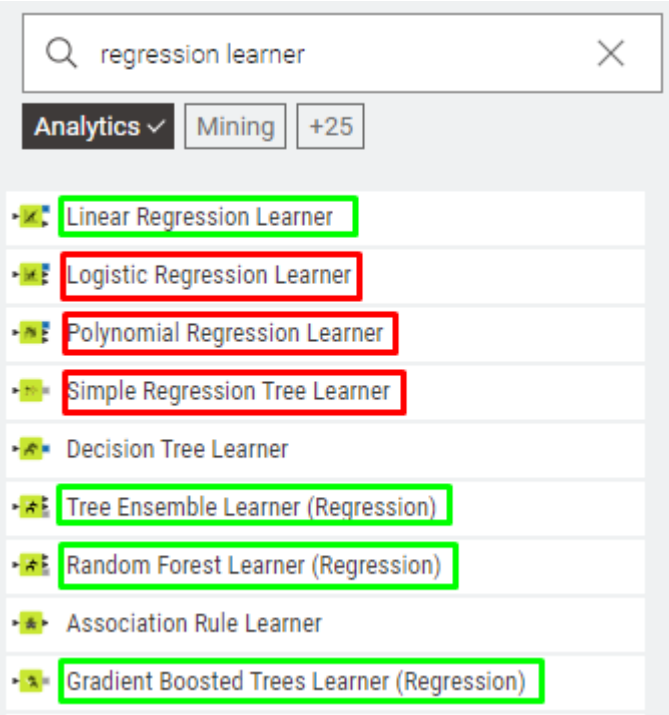

Рис. 3.3. Моделі KNIME Analytics, що доступні для побудови регресій

Джерело: розроблено автором

1. Лінійна регресія (Linear Regression):

Проста модель, яка передбачає лінійний зв'язок між залежною та незалежними змінними. Модель намагається підібрати найкращу пряму лінію, що описує цей зв'язок. Використовується, коли змінні мають приблизно лінійну залежність [29, 33].

На рис. 3.4. представлене налаштування моделі, де ми визначили прогнозований показник (ER) та 8 вищеописаних факторів. Для пропущених значень в комірках ми зазначили логіку «ігнорувати пропущені значення».

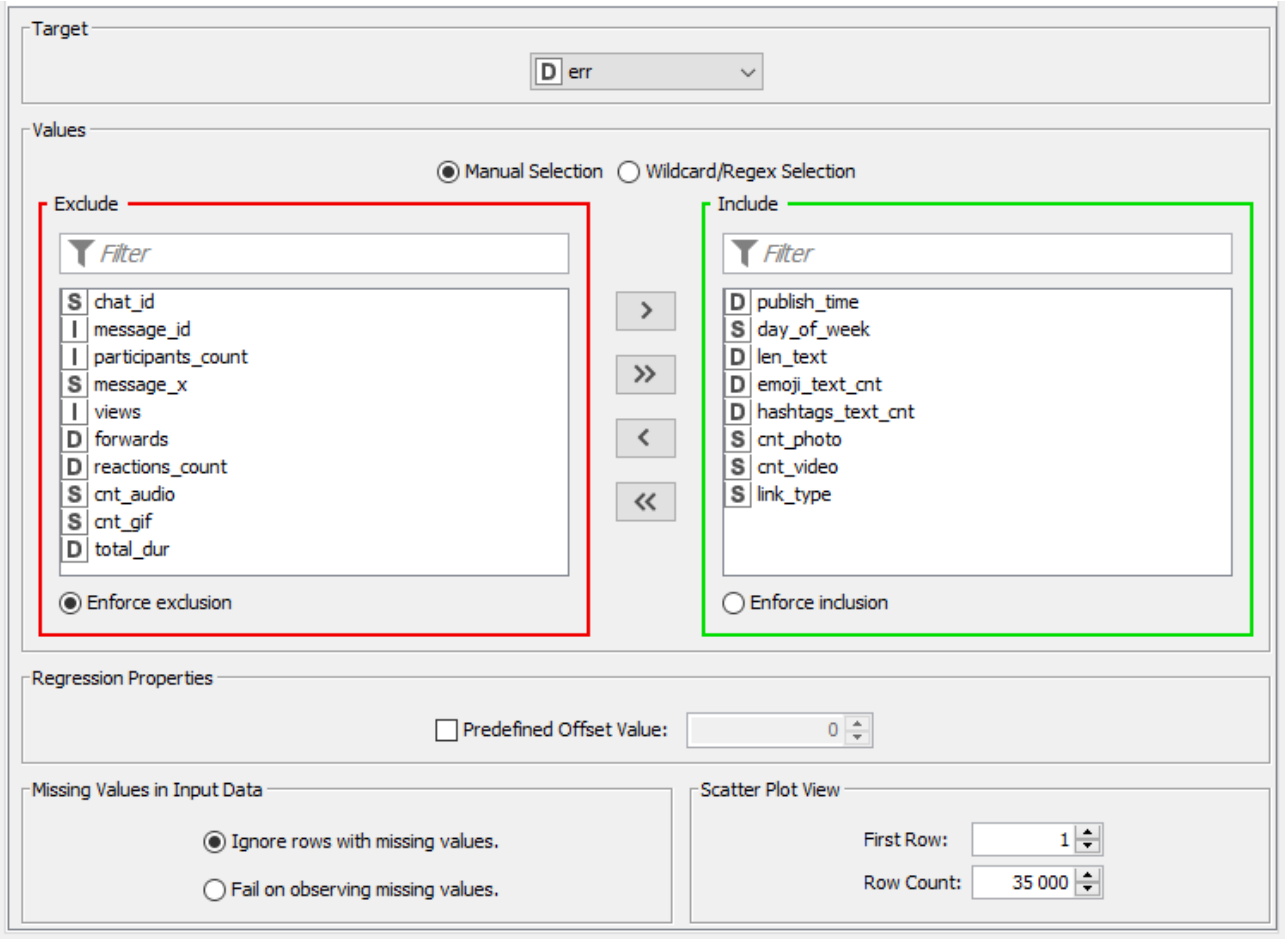

Рис. 3.4. Налаштування моделі «Лінійна регресія»

Джерело: розроблено автором

2. Дерево ансамблю регресії (Tree Ensemble Regression):

Цей метод використовує ансамбль (групу) рішеннях дерев для покращення точності прогнозування. Він об'єднує прогнози з декількох рішеннях дерев, щоб отримати загальний прогноз, який зазвичай точніший, ніж прогноз будь-якого окремого дерева [23, 37, 38].

На рис. 3.5. представлене налаштування моделі з цільовим показником та атрибутами з приміткою «ігнорувати колонки без доменного значення».

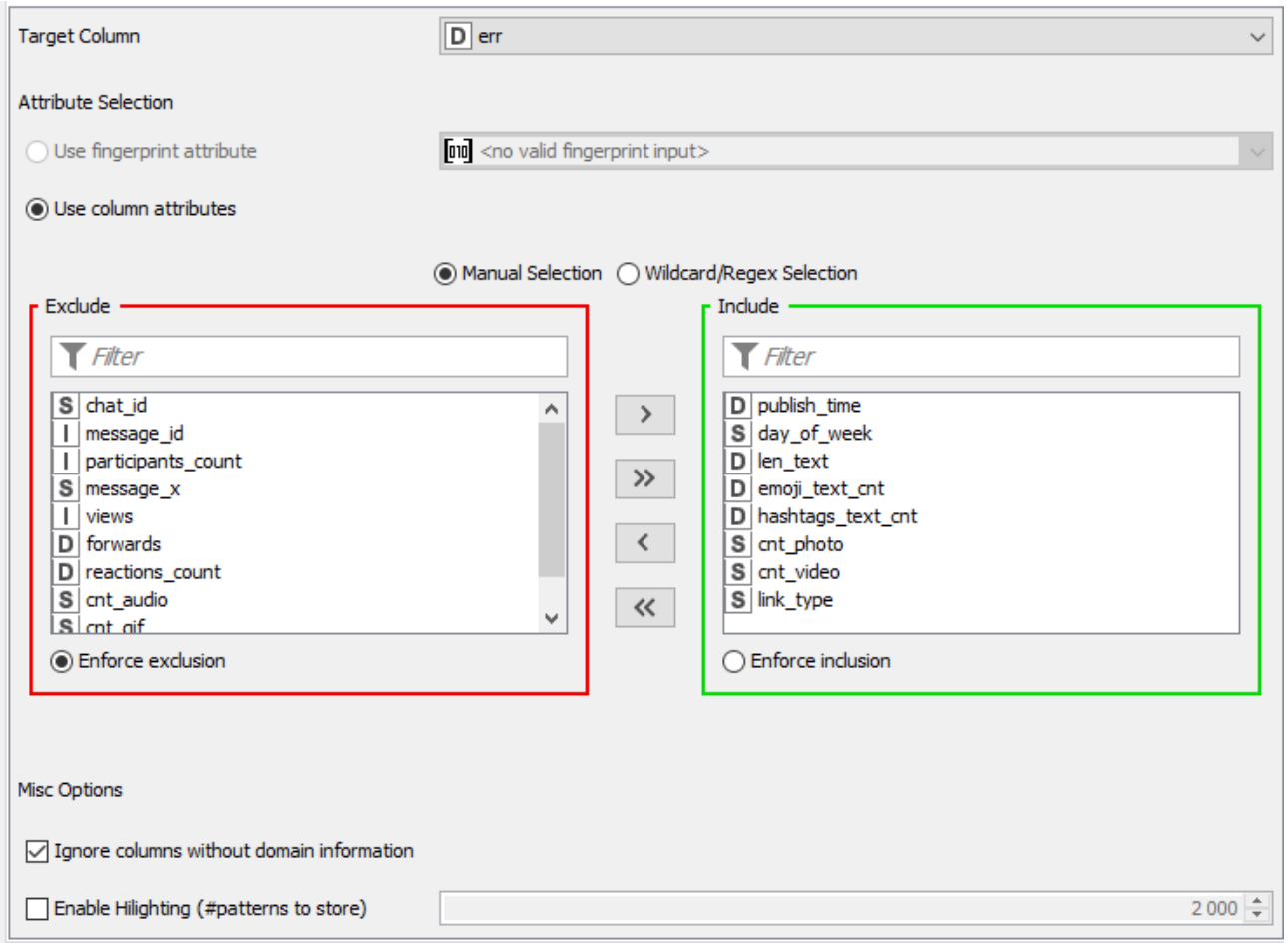

Рис. 3.5. Налаштування моделі «Дерево ансамблю регресії»

Джерело: розроблено автором

3. Випадковий ліс (Random Forest Regression):

Це спеціалізований тип ансамблю дерев, де кожне дерево будується з випадково вибраних підмножин даних і змінних, з метою зменшення перенавчання і підвищення точності прогнозування. Випадковий ліс зазвичай надає дуже високу точність прогнозів і може добре працювати навіть на дуже складних даних [30, 31].

На рис. 3.6. представлене налаштування моделі, де додаткових параметрів моделювання ми не додали у кінцевому варіанті, адже без цього модель давала кращі результати з якими можна буде ознайомитися в наступному підрозділі 3.3.

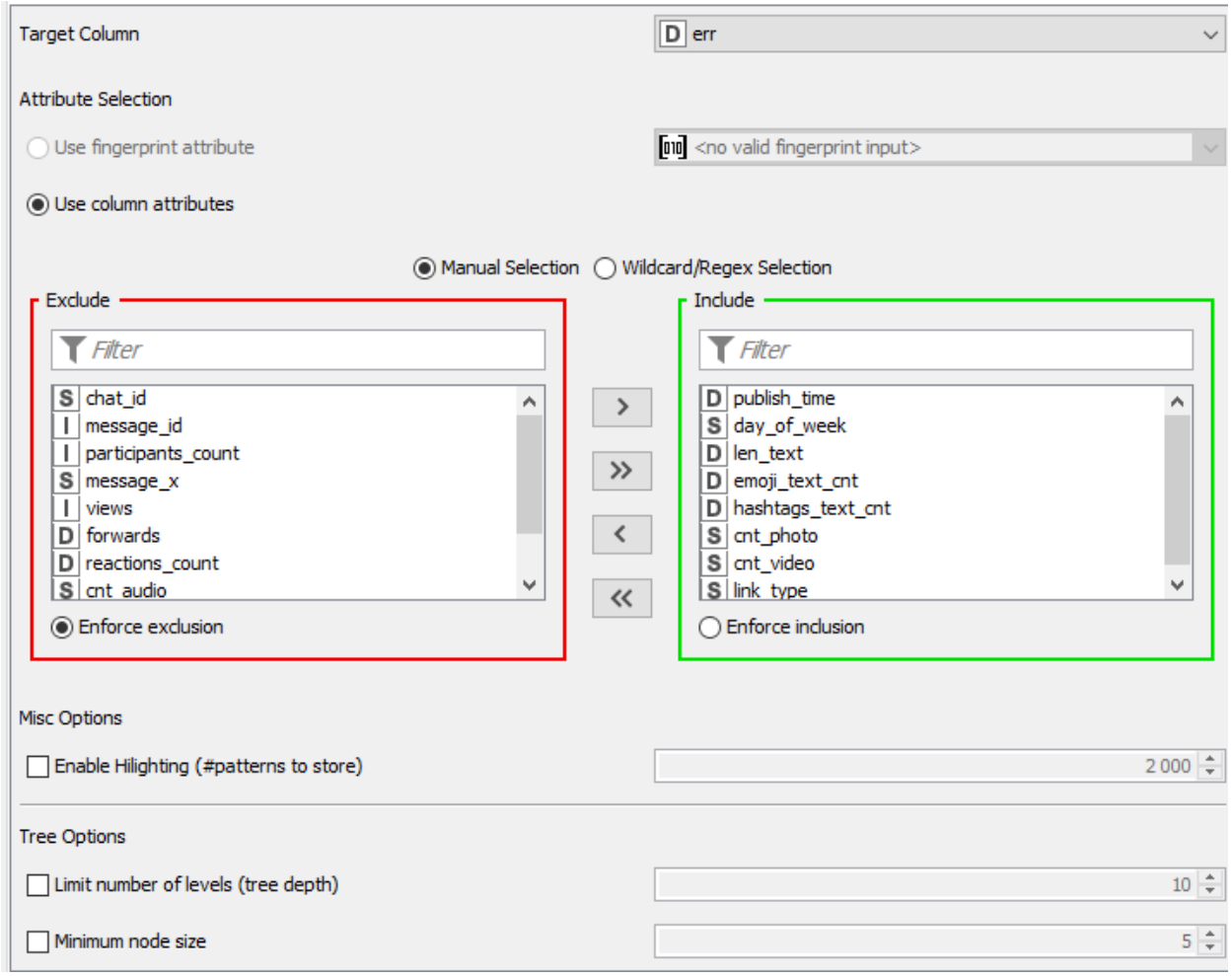

Рис. 3.6. Налаштування моделі «Випадковий ліс»

Джерело: розроблено автором

4. Градієнтний бустинг дерев (Gradient Boosted Trees Regression):

Цей метод також використовує ансамбль рішеннях дерев, але додає дерева послідовно, де кожне нове дерево намагається виправити помилки попередніх. Градієнтний бустинг налаштовується через штраф за помилки, що дозволяє точно налаштовувати модель для досягнення високої точності прогнозування [23, 24, 25].

На рис. 3.7. в налаштування моделі ми особливу увагу приділили параметру «Ліміт на кількість рівнів». Це кількість (глибина) розгалужень дерев (гілок) моделі. Чим більше рівнів тим довше будується модель та аналізуються дані. Ми досягли найкращого рівня точності моделі при глибині – 5. Це значення також відповідає кількості атрибутів, які беруться моделлю для прогнозування цільового параметру – ER.

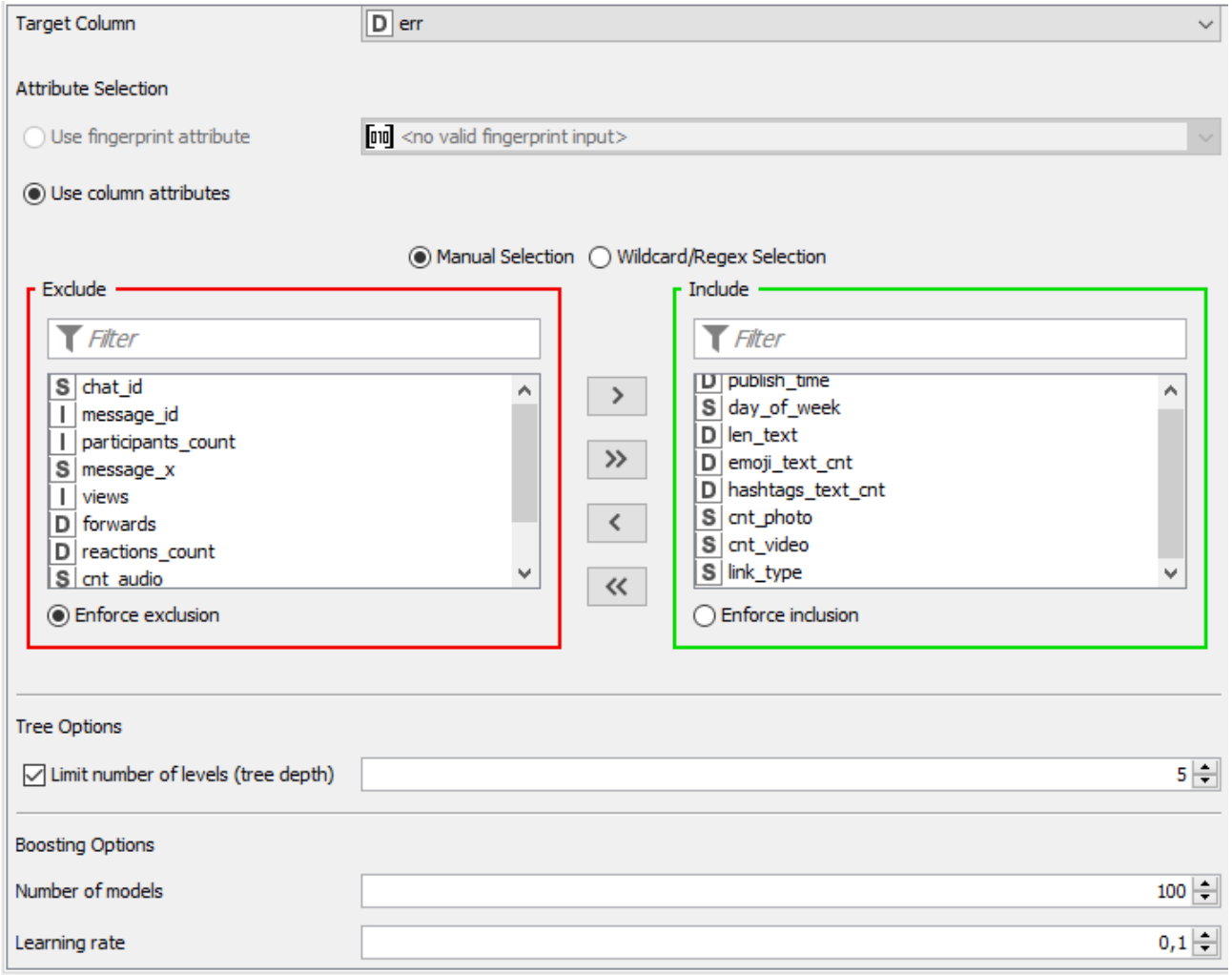

Рис. 3.7. Налаштування моделі «Градієнтний бустинг дерев»

Джерело: розроблено автором

Після запуску нод типу «Learner», що відповідають за тренування моделі, ми переходимо до тестування натренованої моделі за допомогою нод типу «Predictor». Ці ноди тестування моделі на тих 20% даних, що ми виділили в розділі 3.1., не потребують налаштувань. Після застосування натренованої моделі на тестових даних, ми переходимо до аналізу результатів.

3.3. Оцінка точності моделювання

Нода «Scorer» у KNIME Analytics обраховує декілька статистичних показників, які допомагають оцінити, наскільки добре модель прогнозування впоралася з завданням. Вузол використовує реальні значення (ri) та прогнозовані моделлю значення (pi) для розрахунку наступного [36]:

• R^2 (коефіцієнт детермінації) - показує, яка частка змінності реальних даних пояснюється моделлю. Чим ближче значення до 1, тим краще модель відображає

реальні дані. Важливо, що R^2 може бути і від'ємним, що говорить про дуже погане прогнозування.

• Середня абсолютна помилка - середнє значення абсолютних різниць між реальними та прогнозованими значеннями. Чим менше це число, тим точніше ваша модель.

• Середньоквадратична помилка - середнє з квадратів різниць між реальними та прогнозованими значеннями. Цей показник чутливий до великих помилок, тому велике значення може вказувати на наявність помилкових прогнозів з великою різницею.

• Корінь з середньоквадратичної помилки - це просто квадратний корінь з попереднього показника. Він повертає помилку до вихідних одиниць вимірювання і теж допомагає оцінити точність моделі.

• Середня сигнована різниця - показує середнє значення різниць між прогнозованими та реальними значеннями, дозволяючи оцінити, чи схильна модель переоцінювати чи недооцінювати дані.

• Середня абсолютна відсоткова помилка - це середнє з абсолютних значень відсоткових різниць між реальними та прогнозованими значеннями. Цей показник дає уявлення про помилки відносно розміру самих даних.

• Виправлений R^2 - модифікована версія R^2, яка враховує кількість прогнозованих змінних та обсяг даних. Він дає більш точну оцінку для моделей з великою кількістю змінних.

В зведеній табл. 3.1. можемо ознайомитися з цими показниками по кожній моделі.

Таблиця 3.1.

Показники точності прогнозування ER чотирма регресійними моделями

№	Model	Linear Regression	Tree Ensemble	Random Forest	Gradient Boosted Trees
1	R^2	0,089	0,164	0,164	0,182
2	Mean absolute error	10,481	10,019	10,013	9,755
3	Mean squared error	176,778	162,132	162,159	158,631
4	Root mean squared error	13,296	12,733	12,734	12,595
5	Mean signed difference	-0,114	-0,033	-0,043	-0,451
6	Mean absolute percentage error	353,546	345,460	345,321	314,582
7	Adjusted R^2	0,088	0,163	0,163	0,181

Джерело: розроблено автором

Лінійна регресія показує найнижчий R^2 зі значенням 0,089, що вказує на незадовільне пояснення моделлю варіабельності реальних даних. Це відображається і в інших показниках, таких як високі значення помилок (MAE, MSE, RMSE). Цікаво, що лінійна регресія має найменшу середню сигновану різницю, що вказує на її здатність досягати збалансованих прогнозів без сильного ухилу в переоцінюванні або недооцінюванні.

Моделі Tree Ensemble та Random Forest показують значно кращі результати з R^2, що дорівнює 0,164 для обох, а також мають порівняно нижчі помилки за всіма трьома метриками (MAE, MSE, RMSE) у порівнянні з лінійною регресією. Ці моделі також демонструють кращу здатність управляти помилками, оскільки

їх MAPE значно нижчі, ніж у лінійної регресії. Значення середньої сигнованої різниці вказують на деяку тенденцію до недооцінки даних.

Gradient Boosted Trees виокремлюються з найвищим R^2 (0,182), що свідчить про найкращу здатність пояснювати варіативність даних. Ця модель також має найменші помилки за всіма категоріями (MAE, MSE, RMSE) та найнижчий MAPE, що вказує на її високу точність та стабільність в прогнозуванні. Однак, згідно зі значенням середньої сигнованої різниці, ця модель має найсильнішу тенденцію до переоцінки даних.

Висновки до розділу 3

У розділі 3 ми розглянули процес підготовки даних, моделювання та оцінювання ефективності рекламних постів у Telegram на прикладі бази з 30 тисяч постів. Процес розподілу на тренувальну та тестову множини забезпечив об'єктивний вибір даних, а нормалізація допомогла виключити аномалії, зокрема від аномально довгих постів або таких, що містять надмірну кількість емодзі та хештегів.

Під час моделювання було випробувано чотири різні типи регресійних моделей. Результати показали, що найефективнішою є модель Gradient Boosted Trees з $R^2 = 0.182$, хоча навіть цей показник залишається нижчим за порогове значення 0.3, що вважається індикатором адекватної точності моделі.

Така низька загальна точність підтверджує нашу гіпотезу про те, що аналізовані фактори — час публікації, день тижня, кількість символів, наявність емодзі, хештегів, фото, відео, і тип цільового посилання — значно не впливають на залученість користувачів (ER). Це вказує на потребу розгляду інших, можливо більш суб'єктивних або творчих аспектів рекламних креативів, які могли б вплинути на ефективність рекламних кампаній в месенджерах.

ВИСНОВКИ

У цій кваліфікаційній роботі було виконано глибоке дослідження різних аспектів рекламних кампаній в Telegram. Основна увага зосереджена на аналізі складових рекламних креативів та їх впливу на ефективність реклами.

У першому розділі ми розглянули динаміку розвитку Telegram як ефективної платформи для реклами, підкреслюючи стрімке збільшення його популярності серед українських користувачів з 2021 по 2023 рік. Станом на 2023 рік платформа стала важливим майданчиком не лише для спілкування, але й для ведення бізнесу, з високою вартістю рекламних публікацій, що варіюється від 80 до 1000 гривень за тисячу переглядів, залежно від охоплення та активності аудиторії. Значний вплив на ефективність реклами мають фактори, які визначаються контентом та контекстом рекламних креативів, тому розуміння цих аспектів є ключовим для оптимізації рекламних стратегій.

Другий розділ зосереджувався на аналітичній підготовці та обробці даних, використовуючи методологію CRISP-DM та інструмент KNIME Analytics, що дозволило провести всебічний аналіз 30 тисяч рекламних постів. Аналіз дозволив виявити, як різні параметри (тип медіаконтенту, кількість символів, використання емодзі, формати цільових посилань, час і день публікації) співвідносяться з коефіцієнтом залученості (ER). Результати кореляційного аналізу вказали на слабкий зв'язок між більшістю вивчених параметрів і показником залученості, що вказує на складність і багатофакторність впливу рекламних креативів на поведінку аудиторії.

У третьому розділі був описаний процес підготовки даних, моделювання та оцінювання ефективності рекламних постів. Попри детальне дослідження і застосування різних регресійних моделей, загальна точність моделі залишалася низькою ($R^2 = 0.182$), що підтверджує гіпотезу про те, що аналізовані фактори значно не впливають на залученість користувачів. Це відкриває можливості для подальшого дослідження більш суб'єктивних або творчих аспектів рекламних креативів.

У рамках проведеного дослідження було зроблено унікальний аналіз, який наразі не має аналогів у відкритих наукових джерелах. Вперше було систематично досліджено вплив складових рекламних креативів на ефективність рекламних кампаній у Telegram на українському ринку. Це ставить нашу роботу на передній план в області досліджень цифрового маркетингу в месенджерах, заповнюючи істотний пробіл у розумінні, як саме візуальні та текстові складники рекламного допису в Telegram впливають на залученість аудиторії. Досі подібні дослідження обмежувалися загальними оцінками ефективності маркетингових кампаній без занурення в специфіку Telegram, що робить наш підхід піонерським.

Практична цінність цього дослідження полягає у виявленні та роз'ясненні обмеженої впливовості об'єктивних факторів, які може контролювати адміністратор при створенні рекламного допису, таких як час публікації, день тижня, використання емодзі та інших візуальних елементів. Результати дослідження наштовхую на думку про важливість найму кваліфікованих фахівців для написання текстів, які б привертали увагу та стимулювали б продажі. Ця інформація була передана менеджерам аналітичного сервісу Telemetrio для подальшого глибокого аналізу і використання в їхньому щорічному дослідженні рекламних кампаній в Telegram, що сприятиме поліпшенню стратегій контент-маркетингу в месенджері [3].

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Документація Knime Analytics [Електронний ресурс] – Режим доступу до ресурсу: https://docs.knime.com/.

2. Дослідження рівня довіри українців до Telegram для отримання новин під час війни [Електронний ресурс] // Internews. – 2023. – Режим доступу до ресурсу: https://internews.in.ua/uk/news/ukraintsi-znachnoiu-miroiu-pokladaiutsia-na-telegram-kanaly-dlia-otrymannia-novyn-pid-chas-viyny/.

3. Дослідження українського Telegram 2023 [Електронний ресурс] // Telemetrio. – 2023. – Режим доступу до ресурсу: https://telemetr.io/uk/research-2023.

4. Інтелектуальний аналіз даних у бізнесі з використанням IBM SPSS Modeler : навчальний посібник / О. І. Черняк, Г. О.Чорноус . ВПЦ «Київський університет», 2020. – 263с.

5. Канал «Creative Practice» в Telegram [Електронний ресурс]. – 2024. – Режим доступу до ресурсу: https://t.me/creativepracticeua.

6. Каталог Telegram каналів України [Електронний ресурс] // Telemetrio. – 2024. – Режим доступу до ресурсу: https://telemetr.io/uk/catalog/ukraine.

7. Коефіцієнт кореляції Пірсона [Електронний ресурс] – Режим доступу до ресурсу: https://uk.wikipedia.org/wiki/%D0%9A%D0%BE%D0%B5%D1%84%D1%96%D1%86%D1%96%D1%94%D0%BD%D1%82_%D0%BA%D0%BE%D1%80%D0%B5%D0%BB%D1%8F%D1%86%D1%96%D1%97_%D0%9F%D1%96%D1%80%D1%81%D0%BE%D0%BD%D0%B0.

8. Коефіцієнт кореляції рангу Кендала [Електронний ресурс] – Режим доступу до ресурсу: https://uk.wikipedia.org/wiki/%D0%9A%D0%BE%D0%B5%D1%84%D1%96%D1%86%D1%96%D1%94%D0%BD%D1%82_%D0%BA%D0%BE%D1%80%D0%B5%D0%BB%D1%8F%D1%86%D1%96%D1%97_%D1%80%D0%B0%D0%BD%

D0%B3%D1%83_%D0%9A%D0%B5%D0%BD%D0%B4%D0%B0%D0%BB%D0%B0.

9. Коефіцієнт кореляції рангу Спірмена [Електронний ресурс] – Режим доступу до ресурсу: https://uk.wikipedia.org/wiki/%D0%9A%D0%BE%D0%B5%D1%84%D1%96%D1%86%D1%96%D1%94%D0%BD%D1%82_%D0%BA%D0%BE%D1%80%D0%B5%D0%BB%D1%8F%D1%86%D1%96%D1%97_%D1%80%D0%B0%D0%BD%D0%B3%D1%83_%D0%A1%D0%BF%D1%96%D1%80%D0%BC%D0%B5%D0%BD%D0%B0.

10. Коноплицький С. Як заробляють на Telegram каналах в Україні [Електронний ресурс] / Сергій Коноплицький // SPEKA. – 2023. – Режим доступу до ресурсу: https://speka.media/yak-zaroblyayut-na-telegram-kanalax-v-ukrayini-py22k9?utm_source=google&utm_medium=cpc&utm_campaign=21107681931&gad_source=1.

11. Кузнецова А. Telegram Ads — як працює офіційна платформа для реклами [Електронний ресурс] / Анастасія Кузнецова // Webpromo. – 2023. – Режим доступу до ресурсу: https://web-promo.ua/ua/blog/telegram-ads-yak-pracyuye-oficijna-platforma-dlya-reklami-gajd-po-stvorennyu-i-zapusku-ogoloshen/.

12. Опитування USAID-Internews щодо споживання медіа [Електронний ресурс] // Internews. – 2023. – Режим доступу до ресурсу: https://internews.in.ua/wp-content/uploads/2023/10/Ukrainski-media-stavlennia-ta-dovira-2023r.pdf.

13. Перша конференція про Telegram в Україні \"ERA Telegram\" [Електронний ресурс] // Telemetrio. – 2023. – Режим доступу до ресурсу: https://eratelegram.com/uk.

14. Проект Закону про внесення змін до деяких законів України щодо регулювання діяльності платформ спільного доступу до інформації, через які поширюється масова інформація [Електронний ресурс] // Верховна Рада України. – 2024. – Режим доступу до ресурсу: https://itd.rada.gov.ua/billinfo/Bills/Card/43884.

15. Рекламний креатив в каналі «Жіночий стиль» [Електронний ресурс]. – 2024. – Режим доступу до ресурсу: https://telemetr.io/uk/ads-posts?period=ADS_SEARCH_PERIOD_30_DAYS&country=ukraine.

16. Рекламний пост Сільпо в каналі «Телеграмна служба новин» [Електронний ресурс]. – 2024. – Режим доступу до ресурсу: https://t.me/c/1172830166/66950.

17. Сервіс аналітики Telegram каналів «Telemetrio» [Електронний ресурс]. – 2024. – Режим доступу до ресурсу: https://telemetr.io/uk.

18. Сітка Telegram каналів "Труха" [Електронний ресурс] – Режим доступу до ресурсу: https://telemetr.io/en/net/trukhaukrayina.

19. Статистика Telegram каналу "Student Parliament of the Faculty of Economics" [Електронний ресурс] – Режим доступу до ресурсу: https://telemetr.io/uk/channels/1135580767-spef_knu.

20. У Раді зареєстрували законопроєкт про регулювання Telegram в Україні [Електронний ресурс] // УКРІНФОРМ. – 2024. – Режим доступу до ресурсу: https://www.ukrinform.ua/rubric-society/3844545-u-radi-zareestruvali-zakonoproekt-pro-reguluvanna-telegram-v-ukraini.html.

21. Шевченківська весна 2024. Стратегії економічного зростання: погляд у майбутнє [Електронний ресурс]. – 2024. – Режим доступу до ресурсу: https://econom.knu.ua/2024/02/03/%D1%88%D0%B5%D0%B2%D1%87%D0%B5%D0%BD%D0%BA%D1%96%D0%B2%D1%81%D1%8C%D0%BA%D0%B0-%D0%B2%D0%B5%D1%81%D0%BD%D0%B0-2024-%D1%81%D1%82%D1%80%D0%B0%D1%82%D0%B5%D0%B3%D1%96%D1%97-%D0%B5%D0%BA%D0%BE%D0%BD/.

22. Як виріс Telegram в Україні і чи зможе влада вплинути на месенджер [Електронний ресурс] // Українська правда. – 2024. – Режим доступу до ресурсу: https://www.epravda.com.ua/publications/2024/04/8/712099/.

23. Bhati B. S. Ensemble-Based Approach For Intrusion Detection Using Extra Tree Classifier. In Intelligent Computing In Engineering / B. S. Bhati, C. S. Rai. // Springer Singapore. – 2020. – С. 213–220.

24. Gradient Boosted Trees Learner (Regression) [Електронний ресурс] – Режим доступу до ресурсу: https://hub.knime.com/knime/extensions/org.knime.features.ensembles/latest/org.knime.base.node.mine.treeensemble2.node.gradientboosting.learner.regression.GradientBoostingRegressionLearnerNodeFactory2.

25. Gradient Boosted Trees Predictor [Електронний ресурс] – Режим доступу до ресурсу: https://hub.knime.com/knime/extensions/org.knime.features.ensembles/latest/org.knime.base.node.mine.treeensemble2.node.gradientboosting.predictor.classification.GradientBoostingClassificationPredictorNodeFactory3.

26. KNIME Analytics Community Hub [Електронний ресурс] – Режим доступу до ресурсу: https://hub.knime.com/.

27. KNIME Base nodes [Електронний ресурс] – Режим доступу до ресурсу: https://hub.knime.com/knime/extensions/org.knime.features.base/latest.

28. Linear Correlation [Електронний ресурс] // KNIME Analytics – Режим доступу до ресурсу: https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.preproc.correlation.compute2.CorrelationCompute2NodeFactory.

29. Linear Regression Learner[Електронний ресурс] – Режим доступу до ресурсу: https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.mine.regression.linear2.learner.LinReg2LearnerNodeFactory2.

30. Random Forest Learner (Regression) [Електронний ресурс] – Режим доступу до ресурсу: https://hub.knime.com/knime/extensions/org.knime.features.ensembles/latest/org.knime.base.node.mine.treeensemble2.node.randomforest.learner.regression.RandomForestRegressionLearnerNodeFactory.

31. Random Forest Predictor [Електронний ресурс] – Режим доступу до ресурсу: https://hub.knime.com/knime/extensions/org.knime.features.ensembles/latest/org.knime.base.node.mine.treeensemble2.node.randomforest.predictor.classification.RandomForestClassificationPredictorNodeFactory2.

32. Rank Correlation [Електронний ресурс] // KNIME Analytics – Режим доступу до ресурсу: https://hub.knime.com/knime/extensions/org.knime.

features.stats/latest/org.knime.base.node.stats.correlation.rank2.RankCorrelationCompute2NodeFactory.

33. Regression Predictor [Електронний ресурс] – Режим доступу до ресурсу: https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.mine.regression.predict3.RegressionPredictorNodeFactory2.

34. Rüdiger W. “CRISP-DM: Towards a Standard Process Model for Data Mining.” Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining / W. Rüdiger, J. Hipp. – С. 29–39.

35. Sarkar S. Telegram’s new sponsored messages feature [Електронний ресурс] / Shloke Sarkar // GizChina. – 2021. – Режим доступу до ресурсу: https://www.gizchina.com/2021/11/19/telegram-announces-new-sponsored-messages-feature/.

36. Scorer Node Overview [Електронний ресурс] // KNIME Analytics – Режим доступу до ресурсу: https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.mine.scorer.accuracy.AccuracyScorer2NodeFactory.

37. Tree Ensemble Learner (Regression) [Електронний ресурс] – Режим доступу до ресурсу: https://hub.knime.com/knime/extensions/org.knime.features.ensembles/latest/org.knime.base.node.mine.treeensemble2.node.learner.regression.TreeEnsembleRegressionLearnerNodeFactory.

38. Tree Ensemble Predictor [Електронний ресурс] – Режим доступу до ресурсу: https://hub.knime.com/knime/extensions/org.knime.features.ensembles/latest/org.knime.base.node.mine.treeensemble2.node.predictor.classification.TreeEnsembleClassificationPredictorNodeFactory2.

ДОДАТКИ

Додаток А

Структура бази даних рекламних постів в Telegram за січень-лютий 2024 року

day_of_week	message_x	len_text	emoji_text_cnt	hashtags_text_cnt	views	forwards	reactions_count	err	cnt_photo	cnt_video	cnt_audio	cnt_gif	total_dur	link_type
Tuesday	𝐽𝑢𝑖𝑐𝑦 **AEST**	38	5	0	5938	0,0007303	0,001308502	18,0695	0	1	0	0	107,9	B
Friday	Прем’єра 😍 (	38	5	0	7180	0,0091274	0,000803213	26,21395	0	0	1	0	0	B
Saturday	Кияночка тепе	49	2	0	12339	0,0034295	0,008736147	44,54352	0	0	0	0	0	B
Wednesday	🔷ТЕРМІНОІ	50	4	0	3796	0,0000000	6,31074E-05	11,97779	0	0	0	0	0	C
Wednesday	Сет вправ на к	51	2	0	5781	0,0006603	0,000546448	13,16257	0	1	0	0	6	B
Wednesday	Задовбали ці Т	52	4	0	13403	0,0026858	0,015570817	45,56674	0	1	0	0	9,767	B
Tuesday	🔵ВИ ЄДИН	52	2	0	6070	0,0000000	0,000204588	7,30498	0	0	0	0	0	C
Friday	На Кіровоград	52	2	0	3488	0,0000180	0	3,144212	0	0	0	0	0	B
Thursday	пиздить кохан	53	4	0	4938	0,0002754	0,001652589	22,66801	0	1	0	0	21,634	B
Tuesday	Бажаєте розмі	53	1	0	5654	0,0002296	0	25,96436	0	0	0	0	0	C
Tuesday	🔵ВИ ЄДИН	53	2	0	10574	0,0000558	0,000641866	29,50911	0	0	0	0	0	C
Friday	🔵ВИ ЄДИН	53	2	0	13581	0,0000310	0,000247793	42,06597	0	0	0	0	0	C
Friday	🔵ВИ ЄДИН	53	2	0	6346	0,0001148	0,000401699	36,41685	0	0	0	0	0	C
Tuesday	Бажаєте розмі	53	1	0	3067	0,0000000	0	28,10925	0	0	0	0	0	C
Tuesday	Понаїдалися т	54	1	0	14263	0,0000725	0,001994078	17,2373	0	1	0	0	9,467	B
Sunday	Хвастаюсь ♡	54	2	0	54	0,0000000	0	0,263492	1	0	0	0	0	B
Wednesday	🔷 УВАГА! Т	54	3	0	3858	0,0000718	0,000718236	27,70955	0	0	0	0	0	C
Thursday	Всі п🔞рно ві	54	2	0	2444	0,0000728	0	17,7836	0	0	0	0	0	C
Wednesday	🔷МАСОВИЙ	55	4	0	21597	0,0000000	0,000375785	15,60736	0	1	0	0	4,867	C
Wednesday	🔵Причина тр	55	2	0	10237	0,0000000	0,002933806	31,28476	0	0	0	0	0	B
Monday	❎ ВІДБІЙ Т	55	2	0	9481	0,0000000	0,001042665	15,94437	0	0	0	0	0	B
Tuesday	Повідомляємо	55	1	0	6759	0,0000000	0	41,13566	0	0	0	0	0	C
Tuesday	Трусики обтян	55	1	0	4822	0,0000883	2,52172E-05	6,079863	0	0	0	0	0	B
Tuesday	🔷 УВАГА !!	55	4	0	403	0,0000000	0,000511866	1,875291	0	0	0	0	0	C
Tuesday	Повідомляємо	55	1	0	6210	0,0000000	6,08606E-05	37,79441	0	0	0	0	0	C
Tuesday	🔊Цікаво де с	55	2	0	22434	0,0000000	0,002372181	40,31629	1	0	0	0	0	C
Wednesday	Змінили свій н	55	2	0	3328	0,0000000	0	2,999982	0	0	0	0	0	B
Sunday	Повідомляємо	55	1	0	7575	0,0000000	6,08606E-05	46,10188	0	0	0	0	0	C
Monday	☆УВАГА☆	55	5	0	1400	0,0000352	0,000598992	4,932878	0	0	0	0	0	B
Saturday	▫️🔳 ЗАПРОІ	55	1	0	2315	0,0000645	0,001354227	14,92874	0	0	0	0	0	B
Thursday	ТУТ БЕЗКОЦ	55	0	0	1	0,0000000	0	0,008318	0	0	0	0	0	B
Monday	Попередньо Б	55	2	0	13566	0,0005195	0,006845547	41,45835	0	0	0	0	0	B
Wednesday	🏎 Зліт 3-х F1	55	2	0	8736	0,0050408	0,004719005	46,84685	0	0	0	0	0	B
Wednesday	💎 Добірка ш	56	2	0	1255	0,0003983	0,001832377	9,998407	0	1	0	0	11,535	B
Sunday	В Індії розваж	56	1	0	16906	0,0004351	0,001728201	20,43145	0	1	0	0	24,634	B
Wednesday	🔞 Ранок у Па	56	3	0	14617	0,0067828	0,011011556	61,19997	0	1	0	0	7,3	B
Monday	⚪ Відбій три	56	2	0	17219	0,0000000	0,002286654	22,12016	0	0	0	0	0	B
Wednesday	🔵 Всі підклю	56	1	0	306	0,0000000	3,39167E-05	1,037851	0	0	0	0	0	B
Saturday	!!!!!!!!!1️⃣ ✅ I	56	7	0	2965	0,0000000	0,000421692	3,572332	0	0	0	0	0	C

Додаток Б

Робоча область KNIME Analytics для здійснення прогнозування ефективності реклами в Telegram

Календарний план виконання кваліфікаційної роботи бакалавра

№	Етапи роботи	Терміни виконання	Відмітка керівника про виконання
1	Розробка та затвердження завдання кваліфікаційної роботи бакалавра	До 22 листопада 2023 року	
2	Ознайомлення з темою та формування цілей дослідження	До 15 січня 2024 року	
3	Збір та аналіз літератури	До 1 лютого 2024 року	
4	Формування теоретичної бази дослідження	До 10 лютого 2024 року	
5	Визначення методології дослідження	До 20 лютого 2024 року	
6	Збір даних та підготовка вхідних даних	До 1 березня 2024 року	
7	Визначення кореляційних зв’язків між атрибутами реклами	До 19 березня 2024 року	
8	Розробка моделі для прогнозування значення ефективності реклами	До 1 квітня 2024 року	
9	Аналіз отриманих результатів	До 15 квітня 2024 року	
10	Написання основних розділів роботи	До 10 травня 2024 року	
11	Редагування та доопрацювання тексту роботи	До 31 травня 2024 року	
12	Подання роботи на попередній захист	До 3 червня 2024 року	

Науковий керівник: Ляшенко Олена Ігорівна

Студентка: Ковальчук Мілана Юріївна

ЗАВДАННЯ

на кваліфікаційну роботу бакалавра

студентки 4 курсу спеціальності 051 «Економіка», ОПП «Економічна кібернетика»

Ковальчук Мілани Юріївни

1. Тема роботи: Прогнозування ефективності рекламних кампаній на українському ринку Telegram на основі складових рекламних креативів.
2. Термін завершення роботи: 31 травня 2024 року
3. Попередній захист роботи: 3 червня 2024 року
4. Об’єкт дослідження: рекламні кампанії в месенджері Telegram.
5. Предмет дослідження: компоненти рекламних креативів, що впливають на ефективність рекламних кампаній.
6. Мета дослідження: прогнозування ефективності рекламних кампаній на основі аналізу складових рекламних креативів.
7. Завдання дослідження:

   7.1. Визначення шляхів використання месенджеру Telegram в Україні.

   7.2. Дослідження впливу нового законодавства України на бізнес в Telegram.

   7.3. Аналіз рекламних креативів, розміщених в Telegram на початку 2024 року.

   7.4. Виявлення залежності між компонентами креативів і залученістю аудиторії каналу до рекламного допису за допомогою кореляційного аналізу.

   7.5. Розробка моделі для прогнозування ефективності рекламних кампаній на платформі KNIME Analytics.

   7.6. Спростувати або підтвердити висунуті гіпотези про впливовість компонентів рекламного допису напряму на результативність реклами.

Науковий керівник: доктор економічних наук, професор Ляшенко Олена Ігорівна

Студентка: Ковальчук Мілана Юріївна

Затверджено на засіданні кафедри економічної кібернетики

протокол № 4 від 22.11.2023 р.