

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

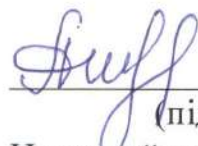
Економічний факультет

Кафедра статистики, інформаційно-аналітичних систем і демографії

КВАЛІФІКАЦІЙНА МАГІСТЕРСЬКА РОБОТА


СТАТИСТИЧНЕ МОДЕЛЮВАННЯ ТЕЛЕВІЗІЙНИХ РЕЙТИНГІВ

Студентки II курсу
спеціальності 051 Економіка
за освітньою програмою
«Економічна аналітика та
статистика»
денної форми навчання
Аліни Настасяк



(підпис)

Науковий керівник:
д.е.н., проф.
Антоніна Єріна



(підпис)

Роботу допущено до захисту на засіданні ЕК рішенням кафедри
статистики, інформаційно-аналітичних систем і демографії, протокол № 12
від «12» травня 2022 року

Завідувач кафедри



д.е.н., проф. Н.В. Ковтун

Київ – 2022

РЕФЕРАТ

Магістерська робота містить: 56 ст., 18 рис., 5 табл., 36 джерел, 2 додатки.

Ключові слова: телевізійний рейтинг, рекламний рейтинг, ТБ реклама, часовий ряд, кореляційний аналіз, регресійний аналіз, дерево рішень, штучні нейронні мережі, статистичні моделі, прогнозування.

Об'єкт дослідження: ринок телевізійної реклами, рекламні рейтинги, процеси та фактори, що впливають на формування рекламних рейтингів.

Мета дослідження: створення статистичної моделі, яка дасть змогу прогнозувати телевізійні рейтинги, для ефективного розміщення реклами.

Методи дослідження: порівняльний та графічний аналіз, кореляційний аналіз, регресійний аналіз, штучні нейронні мережі, аналіз часових рядів.

Наукова новизна, теоретична значимість дослідження: у роботі розроблена статистична модель, яка дає змогу прогнозувати рекламні рейтинги. Також зібрано теоретичний матеріал щодо процесів та факторів, що є впливовими на формування рекламних рейтингів.

Практична цінність: побудована в роботі статистична модель може бути використана медіа агентствами при розробці рекламних кампаній, а також медіа холдингами при розміщенні реклами в ефірі задля уникнення перевиконання або недовиконання плану розміщення.

RESUME

Taras Shevchenko National University of Kyiv,
Faculty of Economics, Department of Statistics, Information and Analytical
Systems and Demography

Key words: television rating, advertising rating, television advertisement, time series, correlation analysis, regression analysis, decision tree, artificial neural networks, statistical models, forecasting.

The graduation research of student Alina Nastasiak “Statistical modeling of TV ratings” deals with building model which creating a statistical model that will allow predicting television ratings for the efficient use of television time..

The author investigated theoretical aspects of advertising rating and its modeling. To construct optimal model are used correlation analysis, regression analysis, graphical analysis, time series analysis, market-based comparative approach and artificial neural networks.

The work is interesting for media agencies and media holdings.

Pages 56, tables 5, bibliog. 36, append. 2.

ЗМІСТ

ВСТУП	5
РОЗДІЛ I. Характеристики ринку телевізійної реклами	8
1.1 Суть та роль телевізійної реклами	8
1.2 Український ринок ТВ реклами	10
РОЗДІЛ 2. Методи дослідження медіа ринку та телевізійних рейтингів	13
2.1 Методологія формування телевізійних рейтингів.....	13
2.2 Порівняльна характеристика методів аналізу та прогнозу телевізійних рейтингів.....	15
2.3 Загальна методика проведення дослідження телевізійних рейтингів.	18
РОЗДІЛ 3. Прогнозування рекламних рейтингів	19
3.1 Опис бази даних дослідження рекламних рейтингів	19
3.2 Статистична модель прогнозування рейтингів	24
3.3 Обґрунтування результатів та практичне застосування	35
ВИСНОВКИ	36
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	39
ДОДАТКИ	42

ВСТУП

Актуальність теми дослідження обумовлена тим, що точний прогноз телевізійних рейтингів безпосередньо впливає на результати рекламних кампаній. Від цього залежить дохід, репутація медіа агентств та холдингів, їх просування та сприйняття потенційними споживачами на ринку.

У 2021 році бюджети на телерекламу у світі зросли до \$169 млрд і займають майже третину від усіх інших медіа-інвестицій [24]. Медіа холдинги та агентства завжди працюють з прогнозними значеннями Avg при плануванні рекламних кампаній. При заключенні контракту між рекламодавцем та рекламістом обговорюється сума та об'єми рекламних рейтингів, які необхідні клієнту. Неякісний прогноз може спричинити порушення домовленостей і медіа холдинги будуть змушені сплачувати штраф, тому вкрай важливо розробити модель, яка б давала найбільш точний прогноз.

Дослідженню впливу прогнозів телевізійних рейтингів на діяльність підприємства присвячені роботи таких українських вчених: Архіпової Т.В., Бець М.Т., Почепцова Г.Г., Цитовського А., Коломійця В.П., Крепак А. [3,4,6, 20].

Дослідженням телевізійного бізнесу займалися такі іноземні вчені: Бодрійяр Ж., Вебб П., Дейян А., Міллер К.Х., Рей М., Рівз Р. [29,36].

Метою є розробка статистичної моделі, яка дасть змогу прогнозувати телевізійні рейтинги, для оптимальної реалізації планів розміщення реклами.

Для вирішення визначеної мети поставлено наступні **завдання**:

- Дослідити ринок телевізійної реклами;
- Проаналізувати сутність та роль телевізійної реклами;
- Охарактеризувати види реклами на телебаченні;
- Дослідити методологію формування телевізійних рейтингів;
- Визначити показники, що впливають на величину телевізійних рейтингів;
- Порівняти методи та підходи для прогнозування телевізійних рейтингів;

- Побудувати статистичну модель прогнозування рейтингів;
- Визначити практичні шляхи використання отриманої моделі для медіа агентств та холдингів;

Предметом дослідження є сукупність теоретичних, методологічно-організаційних та практичних аспектів прогнозування рейтингів телевізійної реклами.

Об'єктом дослідження є ринок телевізійної реклами, рейтинги рекламних блоків – *avr%* та процеси і фактори, що впливають на їх формування та динаміку росту.

При проведенні наукового дослідження були використані наступні **методи**: індукція, дедукція, прогнозування, порівняльний аналіз, регресійний аналіз, кореляційний аналіз, узагальнення, методи економічного та статистичного аналізу та інші загальнонаукові методи.

Інформаційною базою написання роботи були нормативно-правові акти України з питань реклами та телебачення, навчально-методична, довідкова економічна література, матеріали фахових періодичних видань, статистичні довідники, інтернет - ресурси, статистичні матеріали Державної служби статистики України, Національного банку України, Міністерства фінансів України та статистичної служби Європейського союзу.

Дана робота складається зі вступу, трьох розділів, висновків, списку використаних джерел та додатків.

У вступі обґрунтовано актуальність даної теми, визначено об'єкт та предмет дослідження, інформаційна база написання роботи.

У першому розділі «Характеристика ринку телевізійної реклами» розкрито сутність, роль та функції телевізійної реклами, проведено порівняльний аналіз видів реклами, виокремлено переваги та недоліки використання реклами на телебаченні для підприємств. Проведено аналіз ринку телевізійної реклами в Україні та його учасників.

У другому розділі «Методи дослідження медіа ринку та телевізійних рейтингів» висвітлено методологію формування телевізійних рейтингів,

реалізовано порівняльний аналіз методів для прогнозування рекламних рейтингів, а також висвітлено загальну методику проведення дослідження.

У третьому розділі «Прогнозування рекламних рейтингів» було описано базу даних для дослідження, проведено однофакторний та багатофакторний аналіз, що включав в себе регресійний аналіз та аналіз з використання штучних нейронних мереж структури рекламних рейтингів – *avr%*. На основі отриманих результатів побудовано статистичну модель та наведено шляхи використання результатів для медіа холдингів та агентств.

Загальний обсяг випускної кваліфікаційної роботи викладено на сторінках друкованого тексту. Робота містить 5 таблиць, 18 рисунків, 2 додатки. Список використаних джерел налічує 36 найменувань.

РОЗДІЛ I. ХАРАКТЕРИСТИКИ РИНКУ ТЕЛЕВІЗІЙНОЇ РЕКЛАМИ

1.1 Суть та роль телевізійної реклами

Телевізійна реклама є потужним інструментом маркетингових комунікацій підприємницьких структур з потенційним споживачем їхньої продукції чи послуг. На сьогоднішній день компанії витрачають від 3% до 10% річного прибутку на просування своїх товарів чи послуг через телебачення. Від рекламної кампанії залежить репутація фірми, її сприйняття потенційними споживачами та майбутнє на ринку.

До переваг телевізійної реклами можна віднести:

- *Комплексний вплив на телеглядача.* Телереклама одночасно включає в себе зображення і звук, що задіює зорову і слухову пам'ять.
- *Масовість впливу.* Одна трансляція рекламного ролика забезпечує контакт з тисячами споживачів. Телебачення дозволяє застосовувати вплив на цільову аудиторію рекламодавця.

До недоліків використання телевізійної реклами відносять:

- Висока вартість трансляції у телевізійному ефірі;
- Висока вартість організації зйомок ролика;
- Складність розробки креативного рекламного ролика;
- Складність внесення змін до змісту рекламного повідомлення;
- Телеглядачі часто переключають канали, уникаючи рекламних оголошень;
- Складність розробки рекламної кампанії;
- Високий рівень конкуренції, адже на ринку велика кількість аналогічних продуктів.
- Насиченість реклами на телебаченні. Дослідження, які ще у 1979 році описали Пітер Вебб та Майкл Рей, доводять, що насичення рекламою, насамперед на телебаченні, зменшує ефект кожного ролика на телеглядача. [36].

Функції телевізійної реклами:

- Економічна функція. Реклама, що стимулює попит на товар і сприяє зростанню прибутку.
- Інформаційно функція. Реклама доносить певну заздалегідь підготовлену інформацію до населення.
- Соціально-культурна функція. Реклама сприяє зміцненню певних естетичних норм.
- Інтеграційно-організаторська функція. Реклама сприяє організації суспільного життя в певних сферах.
- Контролююча функція. Контролює процеси створення переваг груп споживачів до товарів різного асортименту. [36]

На рекламному ринку виділяють 2 основних види розміщення реклами:

- **Спонсорство (непряма реклама)** - добровільна матеріальна, фінансова, організаційна та інша підтримка фізичними та юридичними особами будь-якої діяльності з метою популяризації виключно свого імені, найменування, свого знаку для товарів і послуг.
- **Пряма реклама** – інформація рекламного характеру, яка доноситься до населення через прямий контакт з ним. Іншими словами, це рекламні прояви, які ми бачимо по телевізору між ТВ-програмами, це фото, зображення, опис товарів чи послуг, які доносяться до користувача в аудіо, відео чи текстовому форматі. [1]

1.2 Український ринок ТВ реклами

Всеукраїнська рекламна коаліція щорічно проводить дослідження стану медіа ринку. У табл. 1.1, наведені обсяги діяльності медіа ринку у 2021 році і відсоток зміни показників 2021 року до 2020 року. Також, було розраховано частку ринку, що займає кожен із видів реклами.

Таблиця 1.1

Стан медіа ринку в Україні

Платформи розповсюдження реклами	Підсумки 2020 р. млн. грн	Підсумки 2021 р. млн. грн	Відсоток зміни 2021 до 2020	Відсоток від всього ринку
ТВ-реклама	12 175	13 642	12,05%	43,93%
Реклама в пресі	1 466	1 599	9,07%	5,15%
Радіо реклама	717	855	19,25%	2,75%
ООН Media (Out Of Home Media)	3 159	4 098	29,72%	13,20%
Реклама в кінотеатрах	20	26	30,00%	0,08%
Digital Медіа реклама	6 980	10 833	55,20%	34,89%
Всього рекламний медіа ринок	24 517	31 053	26,66%	100,00%

Джерело: складено автором на основі [11]

У 2021 році спостерігається зростання для всіх платформ розповсюдження реклами. Найбільші темп приросту виявила Digital Медіа реклама (55,2%), на третину зросла реклама в кінотеатрах та реклама в кінотеатрах ООН Media. Незважаючи на високі темпи зростання рекламного медіа ринку на цих платформах, найбільшу частку ринку (43,9%) посідає ТВ реклама. Всеукраїнська рекламна коаліція прогнозувала збільшення ТВ реклами у 2022 році на 14%. [11] У зв'язку з воєнним станом в країні будь-яка реклама на телебаченні була призупинена, тому, очевидно, даний прогноз нереальний. Але,

через 2 місяці після впровадження воєнного стану частина рекламодавців повертаються в ефір – тобто, ринок починає функціонувати.

Ринок телевізійної реклами, за своєю структурою, функціонує аналогічно до звичайного ринку. Медіа холдинги формують на ринку пропозицію, надаючи, ефірний час, тобто рекламні блоки з відповідними рейтингами, а медіа агентства та населення – формують попит, зокрема, медіа агентства купують необхідні об'єми телевізійних рейтингів для розміщення рекламних продуктів своїх клієнтів. Тобто, для медіа холдингу вкрай важливим є коректний прогноз рейтингів, адже медіа агентства не платитимуть додаткові кошти у разі перевиконання плану і вимагатимуть компенсацію у разі його не виконання.

Рекламні відділи медіа холдингів працюють B2B («бізнес для бізнеса»), оскільки вони орієнтуються на інші бізнес одиниці в лиці медіа агентств. Медіа агентства, також працюють B2B, адже вони розробляють та забезпечують реалізацію рекламних кампаній для підприємств, які, в свою чергу, працюють B2C («бізнес для споживачів») - прагнуть реалізувати свої товари чи послуги, орієнтація йде на кінцевого споживача продукту - населення.

На території України функціонують наступні медіа холдинги:

- Група «1+1 медіа»
- Група «Інтер»;
- Група «Старлайт медіа»;
- Група «Україна»;
- Інші

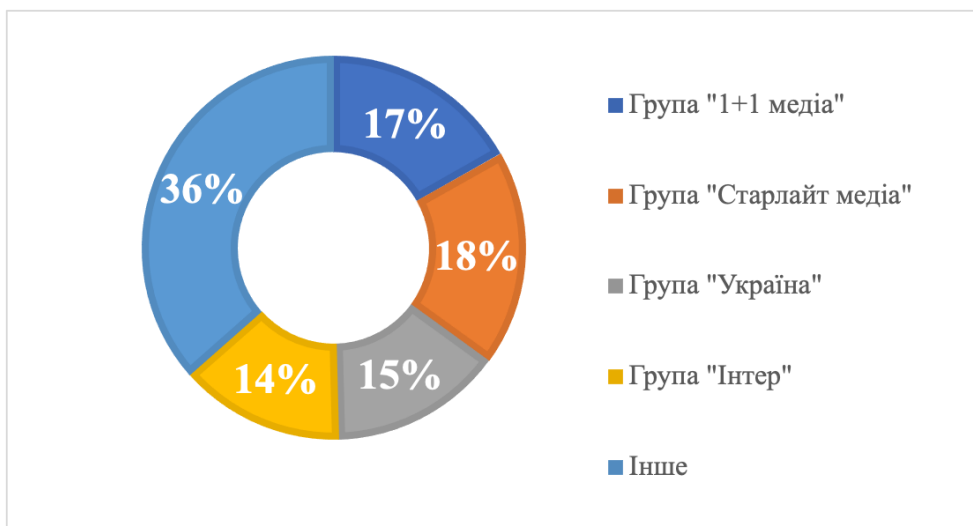


Рис. 1.1 Частка ринку медіа холдингів України станом на 2021 рік по аудиторії: 4+ вся Україна

Джерело: систематизовано автором на основі [5,13-15,18]

Порівняльна характеристика для медіа холдингів наведена у табл. 1.2.

Таблиця 1.2

Порівняльна характеристика для медіа холдингів

Фактор/Група	«1+1 медіа»	«Україна»	«Старлайт медіа»	«Інтер»
Дата заснування	1995	2010	2009	2005
Наявність міжнародних каналів	+	+	+	+
Телеканали	10	7	6	9
Найприбутковіший телеканал	1+1	Україна	СТБ	Інтер
Технічне покриття територією України	98%	92,4%	98%	94,3%

Джерело: складено автором на основі [11,13-15,18].

З табл. 1.2, бачимо, що найдовше на ринку працює 1+1 медіа, розпочавши свою діяльність у 1995 році. Найбільшою кількістю телеканалів володіє «1+1 медіа». Усі представлені медіа холдинги мають у наявності міжнародні телеканали. Щодо технічного покриття на території України, то лідерами є «1+1 медіа» та «Старлайт медіа» з показником 98%. Найбільшою часткою ринку володіє «Старлайт медіа».

РОЗДІЛ 2. МЕТОДИ ДОСЛІДЖЕННЯ МЕДІА РИНКУ ТА ТЕЛЕВІЗІЙНИХ РЕЙТИНГІВ

2.1 Методологія формування телевізійних рейтингів

Щоб дізнатись ефективність реклами, потрібно розуміти, яка кількість населення контактувала з рекламним повідомленням, для цього використовують телевізійні рейтинги. Телевізійний рейтинг – це відсоток конкретної групи населення, яка переглядала певний фрагмент телевізійного контенту чи реклами в певний момент часу [16] .

З 2014 р. оператором ТБ панелі є компанія Nielsen. Nielsen Holdings plc - це глобальна компанія, яка надає послуги з вимірювання та поліпшення бізнес-результатів і є ключовим джерелом інформації про те, що споживачі дивляться і купують. Засновником є Артур Нільсен. Холдинг функціонує з 1923 року. Компанія Nielsen працює більш ніж у 100 країнах світу. Штаб-квартира знаходиться у Нью-Йорку, США. Тобто Nielsen здійснює вимірювання телеаудиторії за допомогою спеціальних пристроїв електронного сканування комунікативної поведінки респондентів.

Як відомо, в основі будь-якого дослідження знаходиться база даних. Чим більша база даних, тим більша ймовірність охопити всі можливі соціальні класи населення. Зібрати таку велику вибірку завдяки опитуванню – неможливо, саме тому було створено піплметр або ТБметр. ТБметр фіксує інформацію про переглянуті телепередачі, час, впродовж якого здійснювався перегляд. На підставі даних, що збираються за допомогою цієї технології, формуються такі статистичні показники, як, наприклад, рейтинг і доля ефірної події чи телеканалу. Можна також проаналізувати соціально-демографічний профіль аудиторії.

Панель дослідження української телеаудиторії налічує 2 840 домогосподарств, з них 1 900 представляють міста з населенням понад 50 тисяч і 940 – невеликі міста (менше 50 тисяч) та села. Обрані домогосподарства за своїми соціально-демографічними та технічними характеристиками

відтворюють структуру населення України [2]. У кожному відібраному домогосподарстві на кожен працюючий і використовуваний телевізор встановлюється ТБметр (піплметр), за допомогою якого відбувається автоматична реєстрація усього перегляду телевізійного контенту. Уся інформація, отримана з піплметрів, щодня проходить серію перевірок за різними критеріями якості. Методика перевірок і допустимі значення параметрів відповідають стандартам BARB (Broadcasters Audience Research Board, Бюро по дослідженню аудиторії мовлення) [17]. Потім інформація потрапляє до бази даних, що доступна користувачам проекту. Завдяки цим даним експерти в області досліджень і програмінгу формують рейтинги для телевізійних проявів.

Результати вимірювання використовують телеканали, рекламні агенції та рекламодавці. На підставі цих даних експерти в області досліджень і телевізійного програмінгу визначають телевізійні та рекламні рейтинги, які відображають глядацький інтерес певних груп населення до того чи іншого фрагменту телевізійного прояву у визначений часовий проміжок. Саме на основі рейтингів можна виявити зміни в перевагах телеаудиторії [3].

2.2 Порівняльна характеристика методів аналізу та прогнозу телевізійних рейтингів

Існує безліч методів аналізу, які можна застосовувати для прогнозування часових рядів. У даній роботі висвітлено 6 моделей.

Розпочнемо з однофакторного аналізу. До даного типу аналізу відносяться моделі у основі яких лежить лише один показник.

У роботі були побудовані наступні моделі даного типу аналізу:

- Модель «Наївного прогнозу». Ми припускаємо, що значення в майбутньому періоді буде такими ж як було в попередньому. При застосуванні даного методу аналізу варто враховувати сезонність даних. Переваги “наївного” методу полягають у тому, що вони прості у використанні та мають можливість генерувати прогнози на основі коротких попередніх спостережень, коли довші історичні ряди даних недоступні. Недоліком є те, що модель не враховує жодних можливих причинно-наслідкових зв'язків, які лежать в основі прогнозованої змінної. Дану модель не варто застосовувати для часових рядів, які мають трендову змінну.
- Експоненційне згладжування, тобто при кожній наступній ітерації враховуються всі попередні значення ряду, але ступінь врахування зменшується за експоненційним законом. Даний метод включає в себе модель Холта-Вінтерса, яка є дуже потужним алгоритмом прогнозування, незважаючи на те, що є однією з найпростіших моделей. У основі лежить спосіб моделювання трьох аспектів часового ряду: типове значення (середнє), тенденція у часі (тренд) та циклічний повторюваний шаблон (сезонність). [32]
- LSTM - мережі довгої короткострокової пам'яті. LSTM – це особливий різновид архітектури рекурентних нейронних мереж. LSTM-модулі розроблені спеціально, щоб уникнути проблеми довготривалої залежності, запам'ятовуючи значення як на короткі,

так і довгі проміжки часу. Це обумовлено тим, що LSTM-модуль не використовує активаційну функцію всередині своїх рекурентних компонентів. Таким чином, значення, що зберігається, не розмивається в часі і градієнт не зникає при використанні методу зворотного поширення помилки в часі при тренуванні мережі. Ключові компоненти LSTM-модуля: стан осередку та різні фільтри. Про стан осередку можна говорити як про пам'ять мережі, яка передає відповідну інформацію по всьому ланцюжку модулів. Таким чином, навіть інформація з ранніх тимчасових кроків може бути отримана на пізніших, нівелюючи ефект короткочасної пам'яті. У міру того, як відбувається навчання, стан комірки змінюється, інформація додається або видаляється зі стану комірки структурами, які називаються фільтрами. Фільтри контролюють потік інформації на входах та на виходах модуля на підставі певних умов. Вони складаються з шару сигмоїдальної нейронної мережі та операції поточкового множення. Сигмоїдальний шар повертає числа у діапазоні $[0; 1]$, які позначають, яку частку кожного блоку інформації слід пропустити далі через мережу. Множення на це значення використовується для пропуску або заборони потоку інформації всередину та назовні пам'яті. Наприклад, вхідний фільтр контролює міру входження нового значення пам'ять, а фільтр забування контролює міру збереження значення пам'яті. Вихідний фільтр контролює міру того, якою мірою значення, що знаходиться в пам'яті, використовується при розрахунку вихідної функції активації. [31,34]

З багатофакторних моделей для дослідження були обрані наступні:

- Лінійна регресія – це метод моделювання впливу набору показників на цільову змінну. При використанні лінійної регресії взаємозв'язок між даними моделюється за допомогою лінійних функцій, а невідомі

параметри моделі оцінюються за вхідними даними. До переваг можемо віднести простоту побудови моделі та аргументації результатів. Недоліком є те, що модель може виявити лише лінійну залежність. [32]

- Деревя рішень - використовуються як для задач регресії, так і для класифікації. До переваг віднесемо: здатність ефективно опрацьовувати масиви даних із великою кількістю змінних, нечутливість до масштабування значень ознак, якісна обробка безперервних і дискретних ознак [30].
 - Extra Trees Regressor – тип ансамблевих моделей. Цей клас реалізує мета-оцінювач, який відповідає низці рандомізованих дерев рішень на різних підвбірках набору даних і використовує усереднення для покращення точності прогнозування та запобіганню перенавчанню моделі.
 - Random Forest – алгоритм машинного навчання. Основна ідея полягає у використанні великого ансамблю дерев рішень, кожне з яких саме по собі дає невисоку якість класифікації, але за рахунок їх великої кількості отримуємо якісний результат.

Мова програмування Python останнім часом все частіше використовується для аналізу даних як у науці, так і комерційній сфері. Цьому сприяє простота мови, а також велика різноманітність відкритих бібліотек. Виділимо основні переваги використання цієї мови програмування:

- Відкритий код. Python є доволі простою мовою програмування, яка швидко розвивається, тому якщо дослідника цікавлять функції які ще не реалізовані, то він може розробити власний код для цього і загрузити його для інших користувачів
- Гнучкість. Python легко застосовувати до будь-якого роду задач – чи то аналіз даних, чи то створення веб-сайту.

2.3 Загальна методика проведення дослідження телевізійних рейтингів.

Розробка статистичної моделі для прогнозування рекламних рейтингів проводиться в декілька етапів:

1. Визначення показників, що мають прямий чи опосередкований вплив на рекламний рейтинг. За допомогою кореляційного аналізу можна виявити міру лінійного впливу на цільову змінну. Також варто побудувати графіки, це надасть змогу прослідкувати залежність між показниками.
2. Збір бази даних. Необхідно забезпечити достатній розмір бази даних, для проведення дослідження.
3. Обробка бази даних. На даному етапі необхідно перевірити дані на наявність пропущених значень, дублікатів, а також проконтролювати типи даних для змінних.
4. Описова статистика, яка забезпечить розуміння базових закономірностей, досліджуваних даних.
5. Пошук функціональної форми статистичної моделі експериментальним шляхом.
6. Оцінювання якості прогнозу.
7. Інтерпретація моделі.
8. Шляхи та умови використання отриманої, у ході дослідження, моделі у діяльності медіа агентства.

РОЗДІЛ 3. ПРОГНОЗУВАННЯ РЕКЛАМНИХ РЕЙТИНГІВ

3.1 Опис бази даних дослідження рекламних рейтингів

База даних, що використовується у ході дослідження складається з 14 424 даних. Досліджуваний період – 2 роки: 01.01.2018 – 31.12.2019. Періодичність даних – 1 година.

Для дослідження було обрано такі параметри:

- $Tvr\%$ - *Total TV rating* – Співвідношення сумарного часу перегляду всього ТБ з максимально можливим часом перегляду всього ТБ, за заданий період. Формула для розрахунку [21]:

$$Tvr\% = \frac{\text{Сума контактів з ТБ}}{\text{Тривалість періода * обсяг цільової аудиторії}} * 100$$

- $Avr\%$ - *Advertising Average Audience* – рейтинг рекламного прояву.

Співвідношення сумарного часу перегляду події з максимально можливим часом перегляду цієї події. Таким чином, при розрахунку показника враховується тривалість перегляду кожним глядачем. Формула для розрахунку [21]:

$$Avr\% = \frac{\text{Сума контактів з подією}}{\text{Тривалість події * обсяг цільової аудиторії}} * 100$$

Розробляючи рекламну кампанію, медіа агентства прагнуть максимізувати показник $Avr\%$. Тому, необхідно дослідити його структуру для розробки оптимального плану розміщення реклами, який надасть змогу в коротший проміжок часу охопити необхідну для рекламодавця аудиторію.

- Часові проміжки. Протягом доби досліджувався проміжок часу з 06:00:00 до 26:00:00, тобто до 02:00:00 години ночі. Таке позначення часу пов'язано з тим, що телевізійна доба триває до третьої години ночі наступної доби. Тобто, наприклад 06:00:00 03.05.2019 – 03:00:00 03.05.2019 – вважається однією добою. Також, варто відмітити, важливість використання часових проміжків при аналізуванні структури телевізійних рейтингів. Як згадувалось раніше, добу розділяють на 2 основні частини:

- Prime time – проміжок часу з 7:00:00-10:00:00 та 18:00:00-23:00:00 для будніх днів та вся доба вихідного дня, але інколи даний проміжок часу може бути зміщеним – залежить від медіа холдингу.

- Off-prime time – весь інший час, що не увійшов до prime time.

У базі даних, даний показник є булевою змінною: 1 – Prime time, 0 – Off-prime time. Рис. 3.1 підтверджує факт, що prime-time користується більшою популярністю серед телеглядачів.

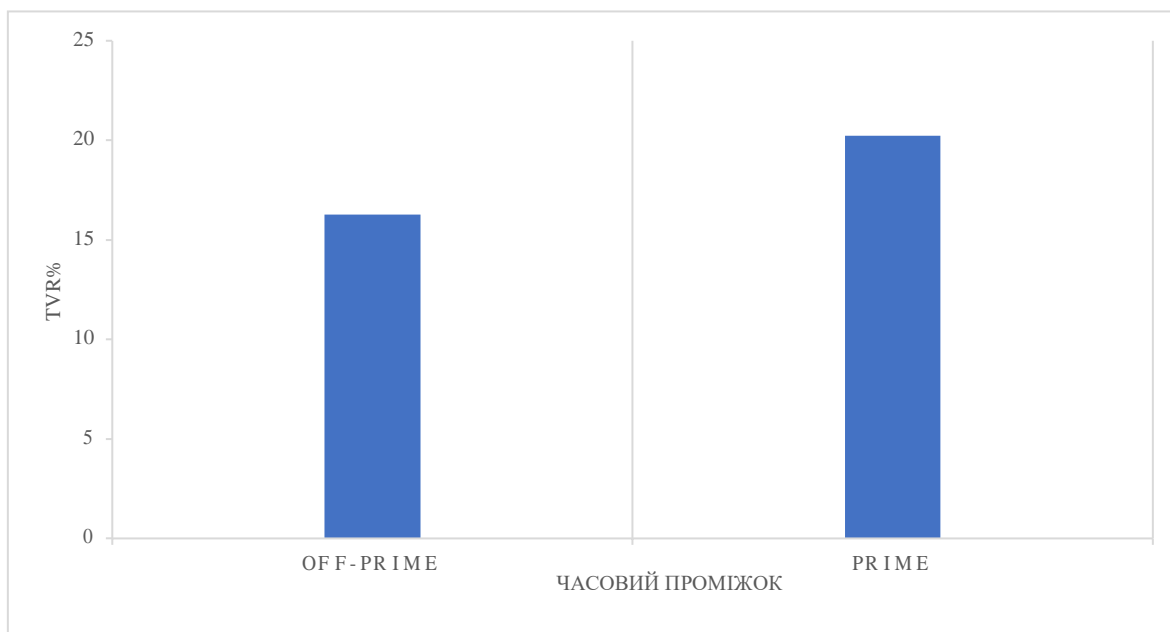


Рис. 3.1 Зміна показника TVR% в залежності від часового інтервалу

Джерело: побудовано автором на основі власної бази даних.

- День тижня. Було виділено будні дні: з понеділок по п'ятницю та вихідні дні: субота – неділя. Судячи, з характеристики, що наведена про часові проміжки, бачимо, що будні дні належать до off-prime time, а вихідні до prime time, а отже день тижня має вплив на величину рейтингів. У базі даних, день тижня є булевою змінною: 1 – вихідний день, 0 – будній день.

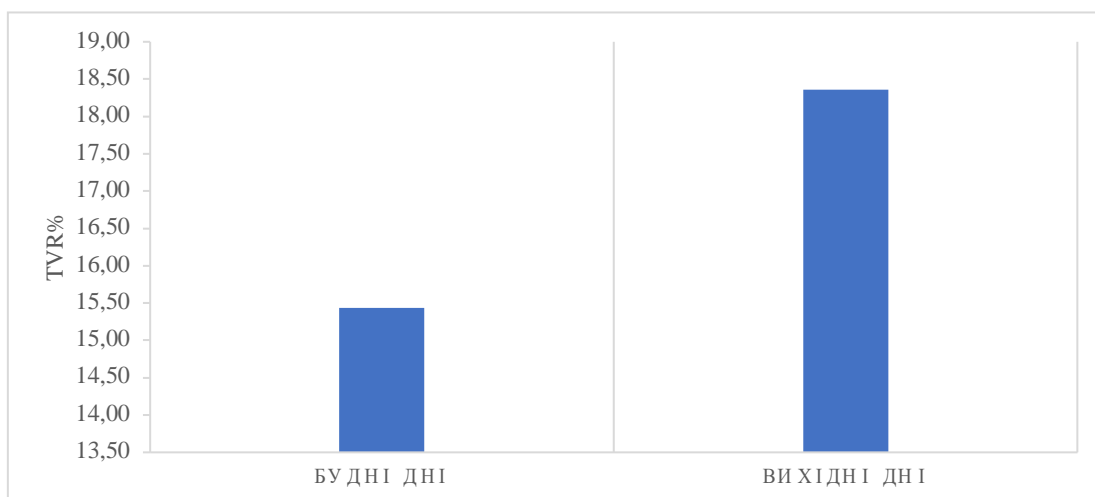


Рис. 3.2 Зміна показника TVR% в залежності від дня тижні

Джерело: побудовано автором на основі власної бази даних.

- Сезон року. Очевидним є вплив пори року на телевізійний рейтинг. З рис. 3.2, бачимо, що наявний спад від зими до літа, і, відповідно ріст від літа до зими. Сезон року набуває чотирьох значень:
 - 1 – зима;
 - 2 – весна;
 - 3 – літо;
 - 4 – осінь;

На рис. 3.3 зображена динаміка показника телевізійних рейтингів в залежності від пори року.

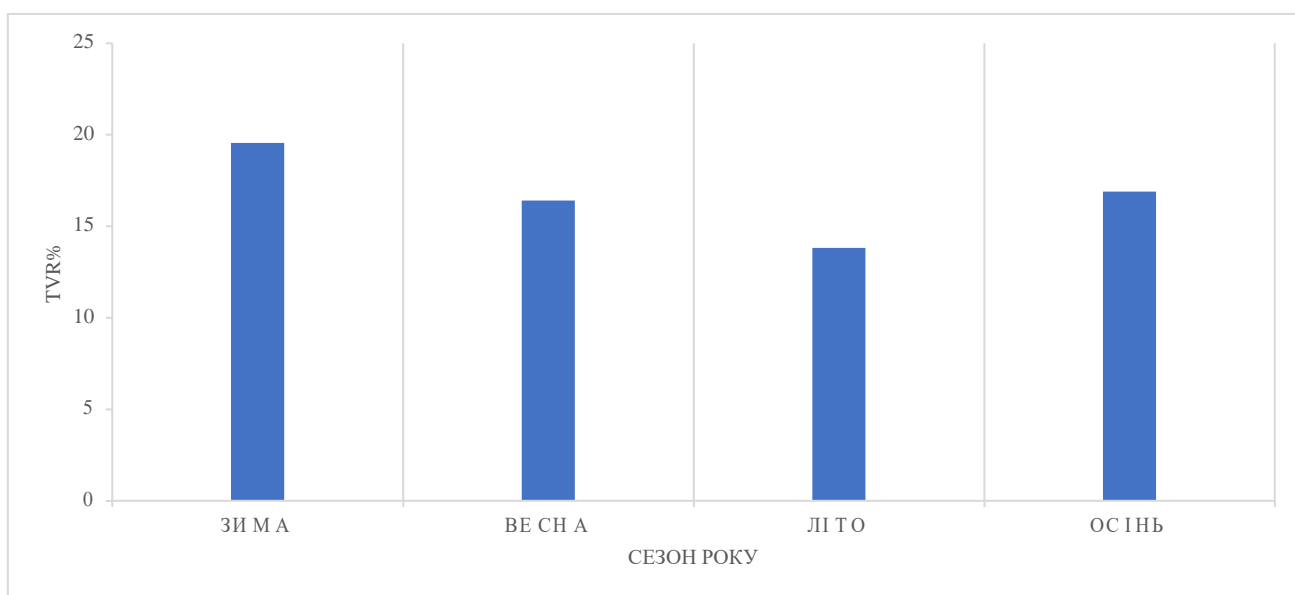


Рис. 3.3 Зміна показника TVR% в залежності від часового інтервалу

Джерело: побудовано автором на основі власної бази даних.

- Температура повітря. Даний показник характеризує температуру повітря в градусах Цельсія. З рис. 3.4 чітко прослідковується залежність між $T_{vr}\%$ та температурою. При збільшенні температури $T_{vr}\%$ зменшується і навпаки.

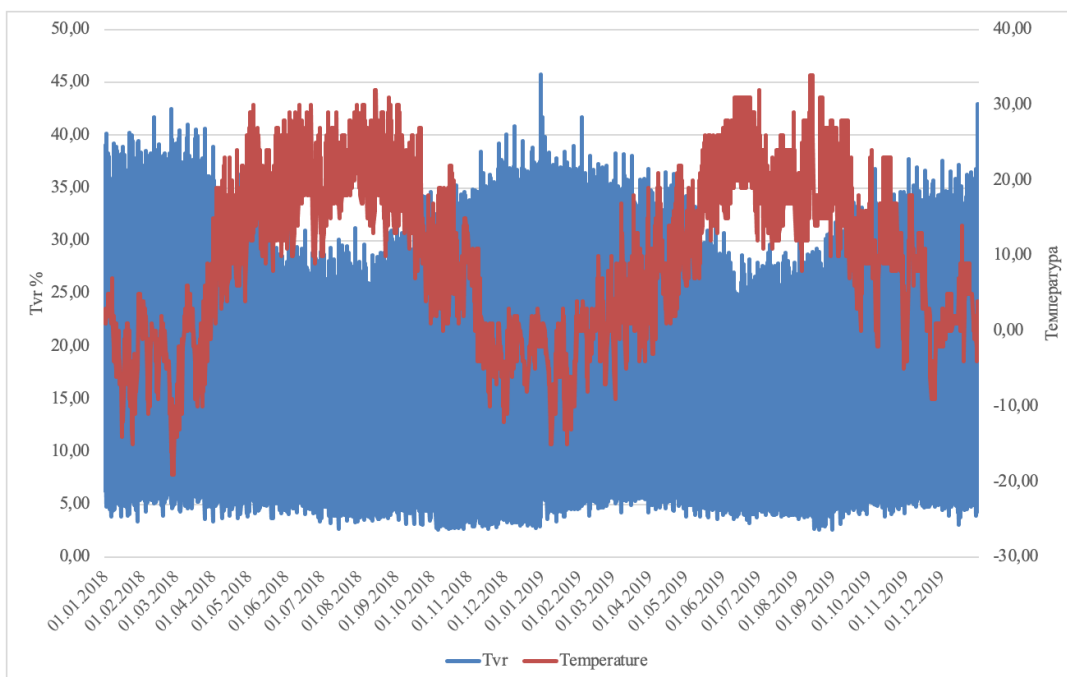


Рис. 3.4 Взаємозалежність показника $T_{vr}\%$ та температури повітря

Джерело: побудовано автором на основі власної бази даних.

- Погодні явища. До погодні явищ було віднесено: сніг, заметіль, дощ, грозу та мряку. Існує думка, що в сонячні дні телеглядачі між двома альтернативами: дивитись телевизор чи провести час на вулиці – обирають другий варіант, тобто прогулянку. Графічне підтвердження проілюстроване на рис. 3.5. Показник «погодні явища» є булевою змінною і приймає значення 1 – у разі наявності будь-яких погодних явищ, і відповідно 0 – при їх відсутності [22].

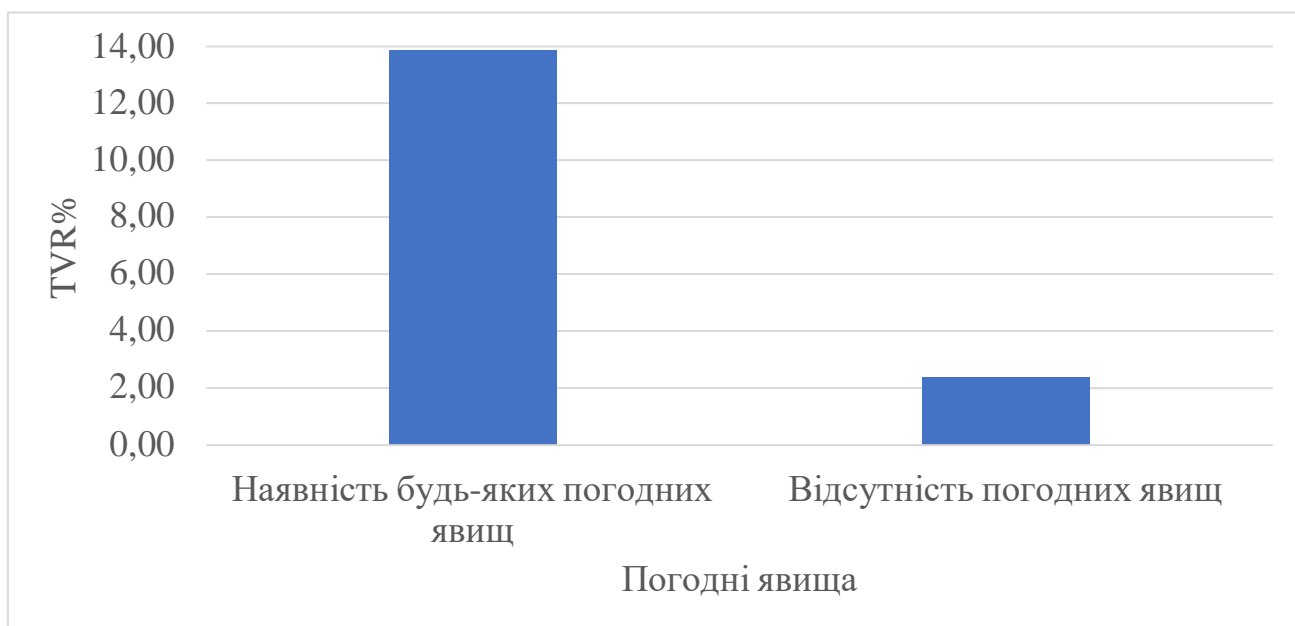


Рис. 3.5 Вплив погодних явищ на телеперегляд

Джерело: побудовано автором на основі власної бази даних.

3.2 Статистична модель прогнозування рейтингів

Першим кроком дослідження є підготовка бази даних для подальшого аналізу. На цьому етапі було створено dummy змінні для таких показників: наявності будь-яких природних явищ, сезону року та тип дня, а також видалено колонки, які не несуть ніякої інформації. У кінцевому результаті ми отримали таблицю з 14 змінними і 14 423 спостереженнями. (Додаток А)

Наступний крок – описова статистика.

Таблиця 3.1

Описова статистика бази даних

	Time type	Avr	Tvr	Temp.	autumn	spring	summer	winter	mist	nothing	rain	snow
count	14423	14423	14423	14423	14423	14423	14423	14423	14423	14423	14423	14423
mean	0,58	1,45	16,26	10,29	0,25	0,25	0,25	0,25	0,01	0,88	0,05	0,06
std	0,49	0,99	8,66	10,82	0,44	0,43	0,43	0,43	0,10	0,33	0,21	0,25
min	0	0	2,59	-19	0	0	0	0	0	0	0	0
25%	0	0,81	9,56	1	0	0	0	0	0	1	0	0
50%	1	1,19	14,19	11	0	0	0	0	0	1	0	0
75%	1	1,8	21,7	20	1	0	0	0	0	1	0	0
max	1	11,52	45,72	34	1	1	1	1	1	1	1	1

Джерело: побудовано автором [Додаток Б]

Дана таблиця наводить інформацію про кількість спостережень, середнє значення, стандартне відхилення, квартилі, максимальне та мінімальне значення для кожного показника. Ми бачимо, що розподіли для різних сезонів року є дуже схожими, що не дивно, адже рік рівномірно поділений на 4 періоди. Tvr значно більший за Avr - це підтверджує припущення, що телеглядачі під час реклами перемикають канали. Протягом досліджуваного періоду показник температури повітря коливався від -19 градусів за Цельсієм до 34.

Далі проведемо кореляційний аналіз, щоб виявити лінійну залежність між змінними. З рис.3.6 бачимо, що існує висока кореляція між Avr та Tvr, оскільки чим більш рейтинговою є програма, тим більша ймовірність, що телеглядачі не перемикають телеканал. Також бачимо середній рівень лінійної залежності

між Avr та змінною Time_type, що описує часові проміжки, які належать до prime time або off prime time.

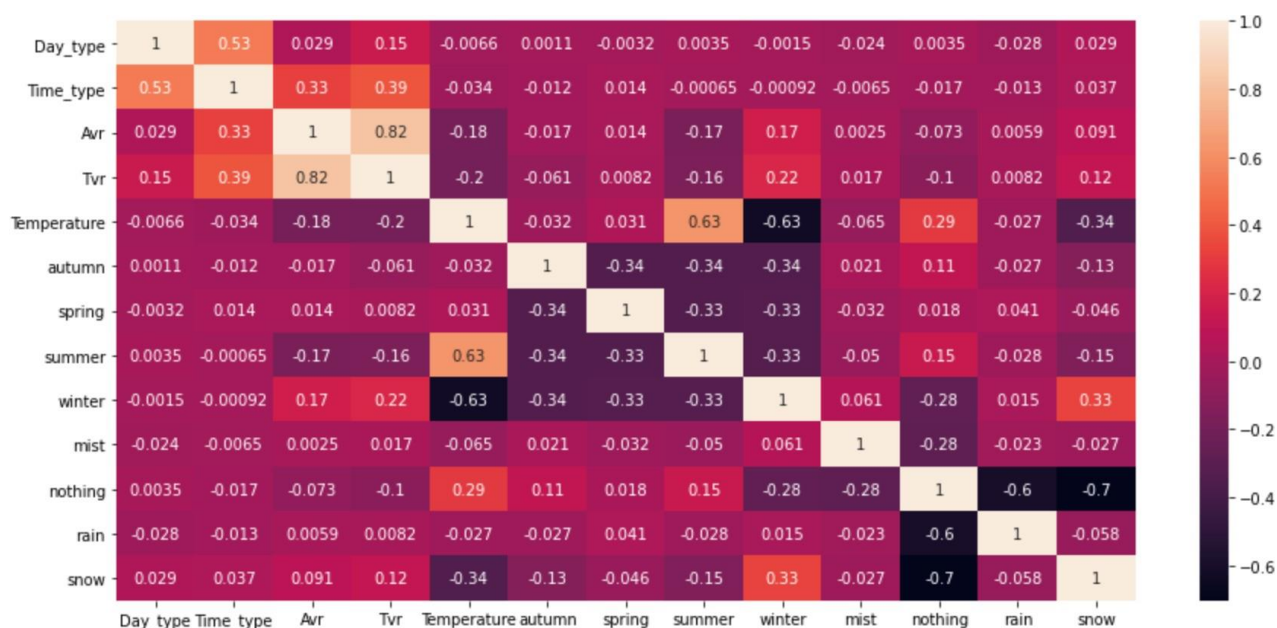


Рис. 3.6 Кореляційний аналіз для досліджуваних факторів

Джерело: побудовано автором на основі власної бази даних [Додаток Б] .

Наступним кроком є декомпозиція часового ряду. Декомпозиція часових рядів передбачає розгляд ряду як комбінації компонентів ряду, тенденції, сезонності та шуму. Декомпозиція надає корисну абстрактну модель про часові ряди та для кращого розуміння проблем під час аналізу та прогнозування часових рядів [33] .

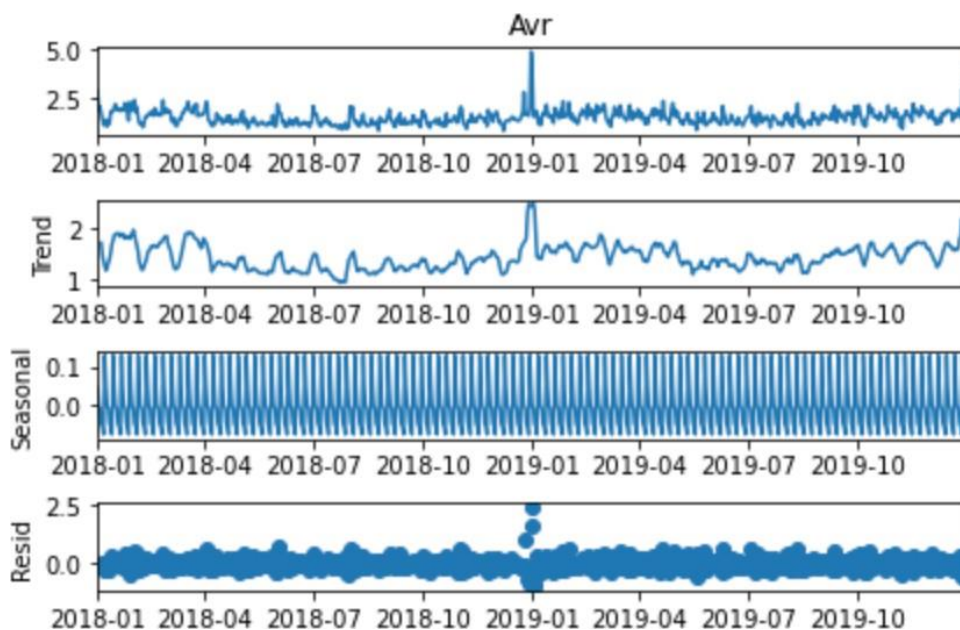


Рис. 3.7 Декомпозиція часового ряду

Джерело: побудовано автором на основі власної бази даних [Додаток Б] .

З рис.3.7 можемо зробити висновки про відсутність тренду. Також бачимо доволі чітку сезонність. Для більш детального дослідження перегрупуємо щогодинні дані в місячні та кварталні, порахувавши середні значення.

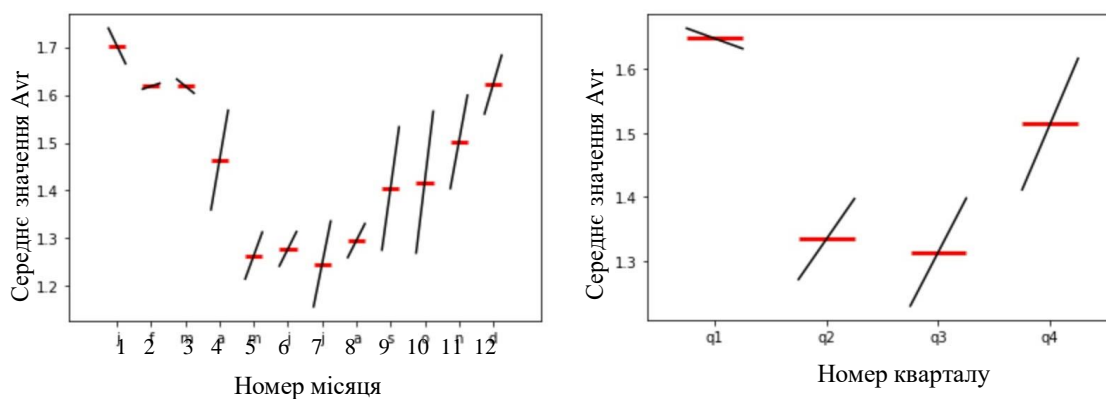


Рис. 3.8 Представлення даних з місячною та кварталною періодичністю

Джерело: побудовано автором на основі власної бази даних [Додаток Б] .

З рис. 3.8 бачимо, що з квітня місяця середнє значення Avg зменшується, і збільшується з вересня. Аналогічна ситуація з щоквартальними даними - для другого та третього кварталу показник середнього рекламного рейтингу менший ніж для першого та четвертого. Тобто, можемо зробити висновок про наявність річної сезонності.

Перевіримо часовий ряд на стаціонарність за допомогою тесту Дікі-Фулера [33] . Результати наведені в табл. 3.2:

Таблиця 3.2

Результати тесту Дікі-Фулера

Показник	Значення
ADF test statistic	-9.5319
p-value	$2.8756 \cdot 10^{-16}$
Count of lags used	42.0000
Count of observations	14380.0000
critical value (1%)	-3.4308
critical value (5%)	-2.8617
critical value (10%)	-2.5669

Джерело: побудовано автором [Додаток Б] .

Отже, досліджувані дані є стаціонарними, адже $p\text{-value} < 0.05$.

Для побудови моделей необхідно поділити базу даних на тестову та навчальну вибірку. Оскільки ми працюємо з часовими рядами, то необхідно це робити рівномірно. Таким чином 01.12.2019-31.12.2019 належатиме до тестової вибірки, а весь інший період до навчальної, відповідно. Побудовані моделі робитимуть прогноз на один місяць.

Розпочнемо з однофакторного аналізу. Перша модель – модель «наївного» прогнозу, яка ґрунтується на умові, що рекламні рейтинги в наступному періоді будуть дорівнювати продажам попереднього періоду. Оскільки дані мають річну сезонність, то ми припускатимемо, що рейтинги в грудні місяці 2019 року будуть такими ж, як були у грудні 2018 року.

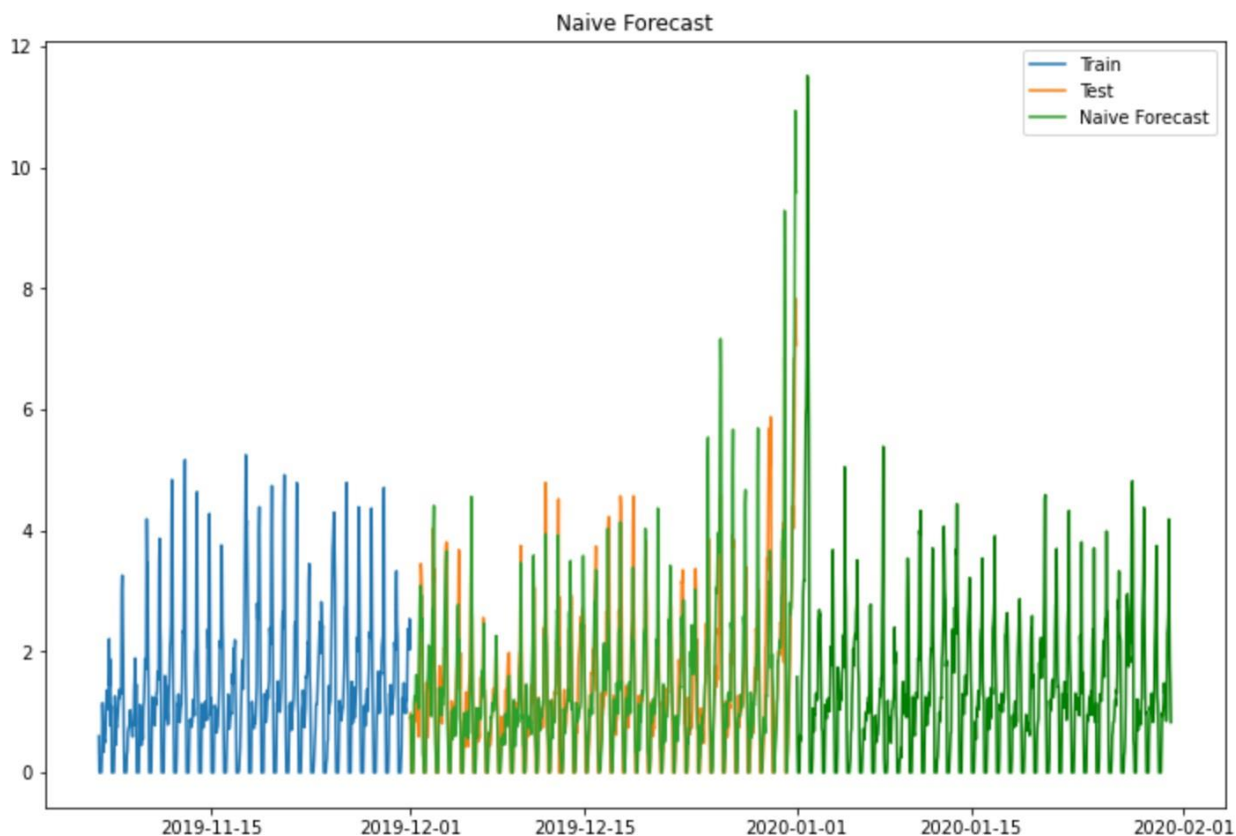


Рис. 3.9 «Наївний» прогноз

Джерело: побудовано автором на основі бази даних [Додаток Б] .

З рис.3.9 бачимо, що криві для тестових та прогнозних даних є дуже близькими один до одної. Для оцінки прогнозу було розраховано RMSE (root mean square error) = 0.752, при цьому середнє значення вибірки становить 1.34. RMSE менше за середнє значення Avg, тобто можемо зробити, що дана модель дає непогані результати і є сенс її використовувати при прогнозування рекламний рейтингів.

Розглянемо модель Холта-Вінтерса, яка є одним із методів прогнозування з використанням експоненційних згладжувань. Холт-Вінтерс – це спосіб моделювання трьох аспектів часового ряду: середнього, тренду та сезонності. За допомогою Python будемо модель з використанням бібліотеки ExponentialSmoothing [32,33] .

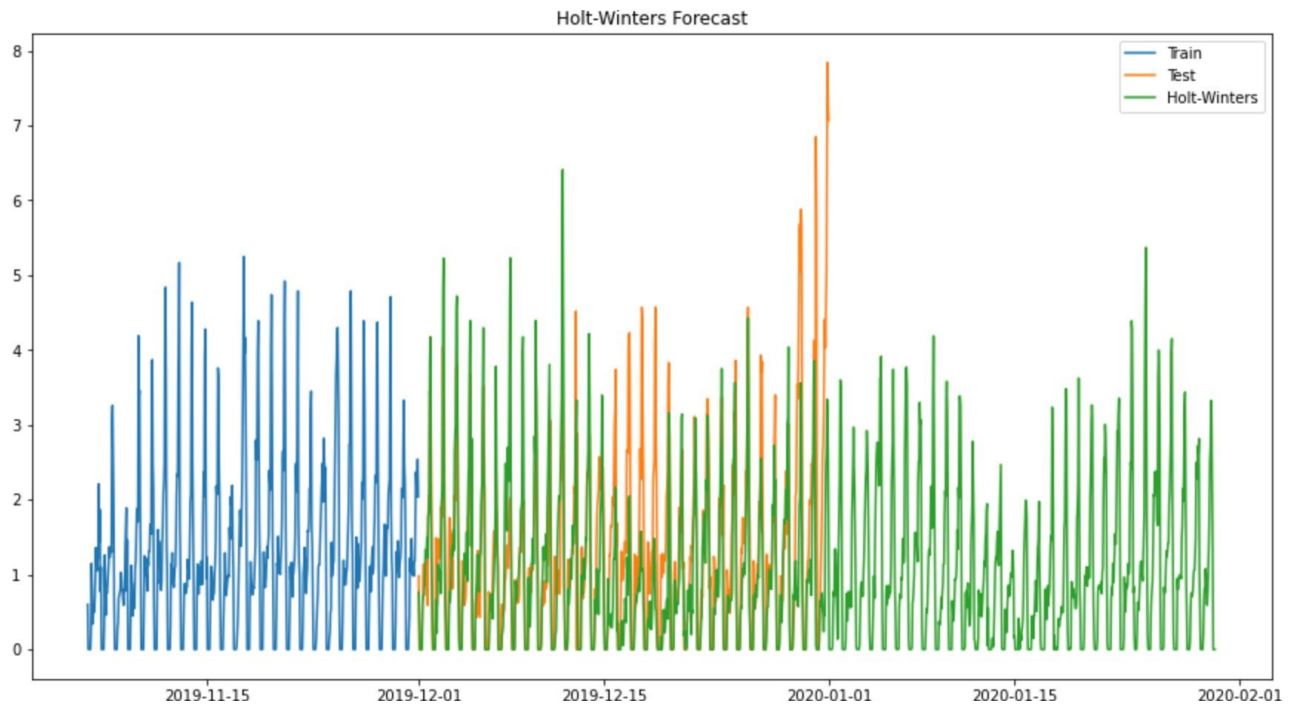


Рис. 3.10 Прогноз з використанням моделі Холта-Вінтерса

Джерело: побудовано автором на основі бази даних [Додаток Б] .

Отримали значення RMSE та середнього для досліджуваної вибірки: $RMSE=0.975$, середнє= 1.337. Бачимо, що модель «наївного» прогнозу дала значно кращі результати.

Далі розглянемо багатофакторний аналіз та додамо в аналіз змінні температури, сезону та типу дня.

Спочатку розпочнемо з лінійної регресії. Використовуючи метод `LinearRegression` отримуємо наступне рівняння:

$$\begin{aligned}
 y = & -0.0363 + -0.2171 \cdot Day_type + 0.1139 \cdot Time_type + 0.0896 \cdot Tvr \\
 & - 0.0028 \cdot Temperature + 0.0419 \cdot autumn + 0.0118 \cdot spring \\
 & - 0.0241 \cdot summer - 0.0295 \cdot winter - 0.0919 \cdot mist + 0.0677 \\
 & \cdot nothing + 0.0268 \cdot rain - 0.0026 \cdot winter
 \end{aligned}$$

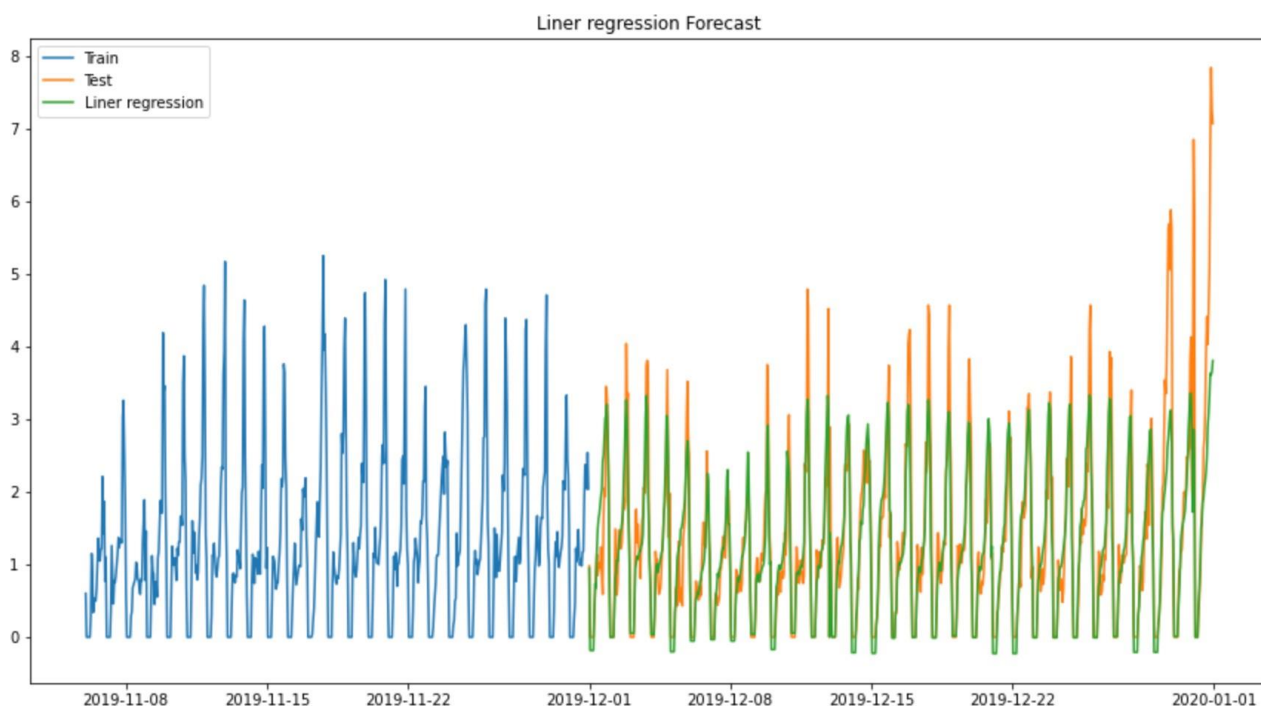


Рис. 3.11 Прогноз з використання лінійної регресії

Джерело: побудовано автором на основі бази даних [Додаток Б] .

З рисунку 3.11 бачимо, що значення прогнозу є близькими до тестової вибірки. Значення RMSE становить 0.59716.

Існує велика кількість моделей, які можна застосувати для прогнозування часових рядів, щоб не будувати всі можливі варіанти було застосовано автоматизоване машинне навчання (AutoML), яке відноситься до методів автоматичного виявлення ефективних моделей для завдань прогнозного моделювання. За допомогою бібліотеки ТРОТ було виявлено найбільш ефективну модель, яку варто застосувати до нашої бази даних - регресор додаткових дерев (extra-trees regressor) з наступними характеристиками:

- `bootstrap=True` – це означає, що для побудови одного дерева використовується лише частина набору даних;
- `max_features=0.7` - кількість параметрів, які слід враховувати при пошуку найкращого розділення;
- `min_samples_leaf=4` - мінімальна кількість вибірок, яка повинна бути у вузлі листка;

- `min_samples_split=3` - мінімальна кількість вибірок, необхідних для розділення внутрішнього вузла;
- `n_estimators=100` – кількість «дерев рішень» в «лісі». [35]

Таким чином, отримали наступні результати:

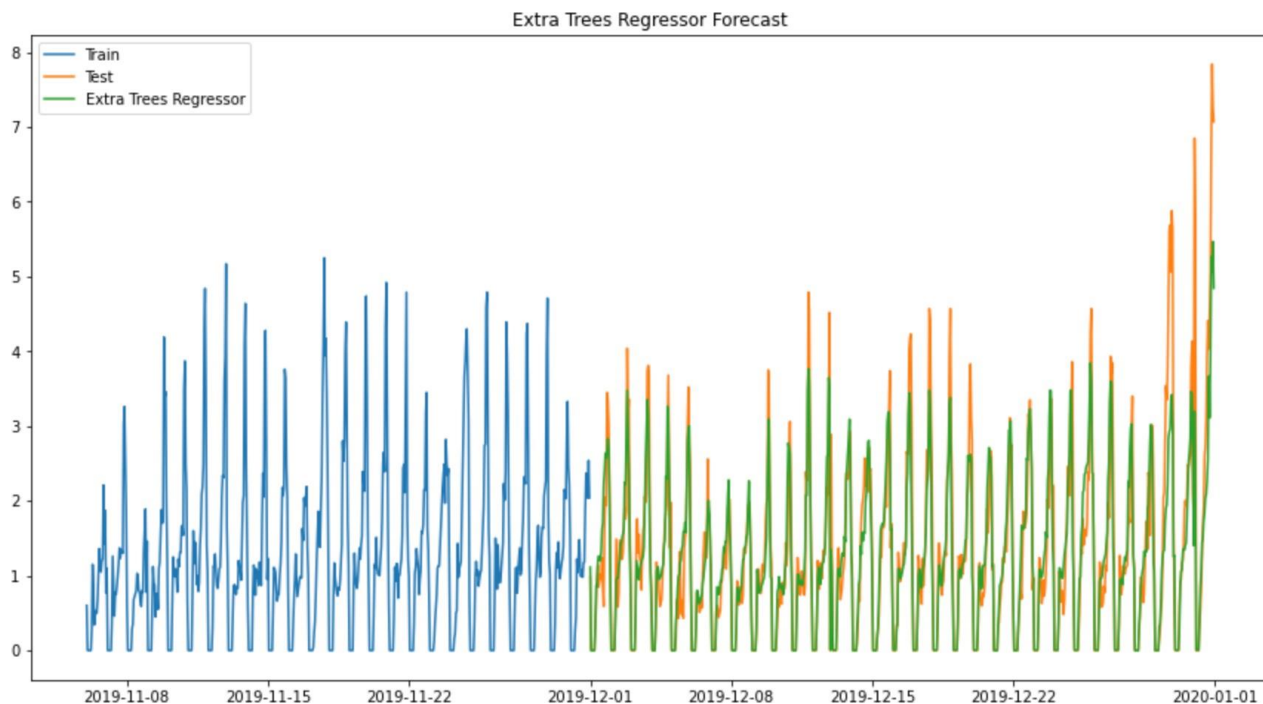


Рис. 3.12 Прогноз з використанням Extra Trees Regressor

Джерело: побудовано автором на основі бази даних [Додаток Б] .

$RSME = 0.529$. Можемо зробити висновки, що модель придатна для прогнозування рейтингів.

Далі застосуємо бібліотеку XGBoost , яка базується на алгоритмі дерева рішень і реалізує методи градієнтного бустингу. Бібліотека найбільш ефективна при побудові моделей передбачення на структурованих великих даних, є вбудована регуляризація, а також за допомогою безлічі налаштувань можна покращувати якість прогнозування моделі за найкоротші терміни, адже є можливість паралельної обробки [32] . Після написання відповідного коду, отримали наступні результати:

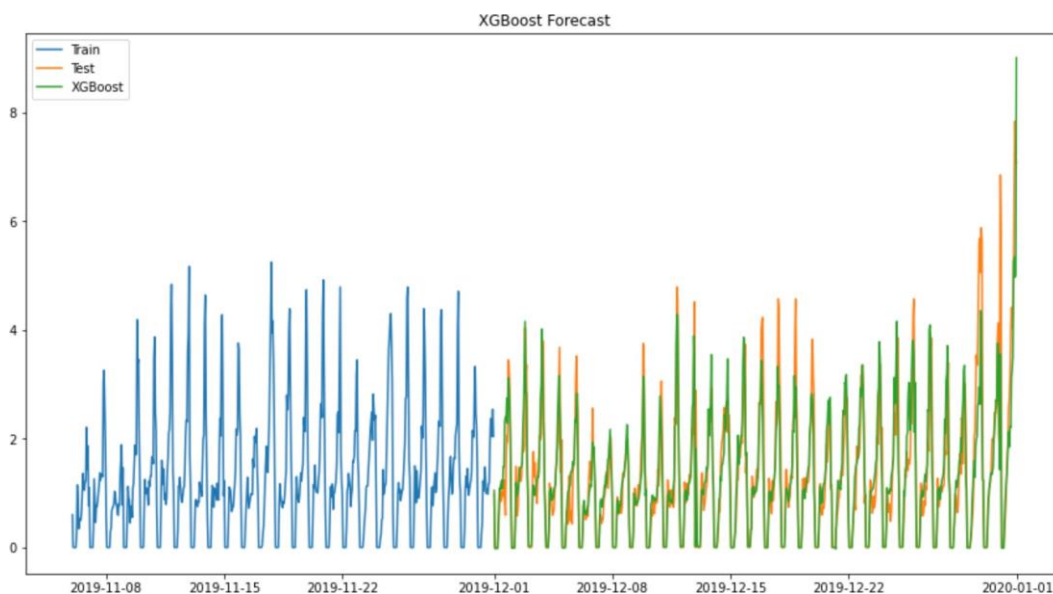


Рис. 3.13 Прогноз з використанням XGBoost

Джерело: побудовано автором на основі бази даних [Додаток Б] .

RMSE = 0.552, що більше на 0.023 п. порівняно з попереднім прогнозом, де використовували Extra Trees Regressor, але дана модель значно краще спрогнозувала ріст рейтингу в останньому періоді.

Також за допомогою даного пакета є можливість оцінити рівень важливості параметрів на рекламний рейтинг:

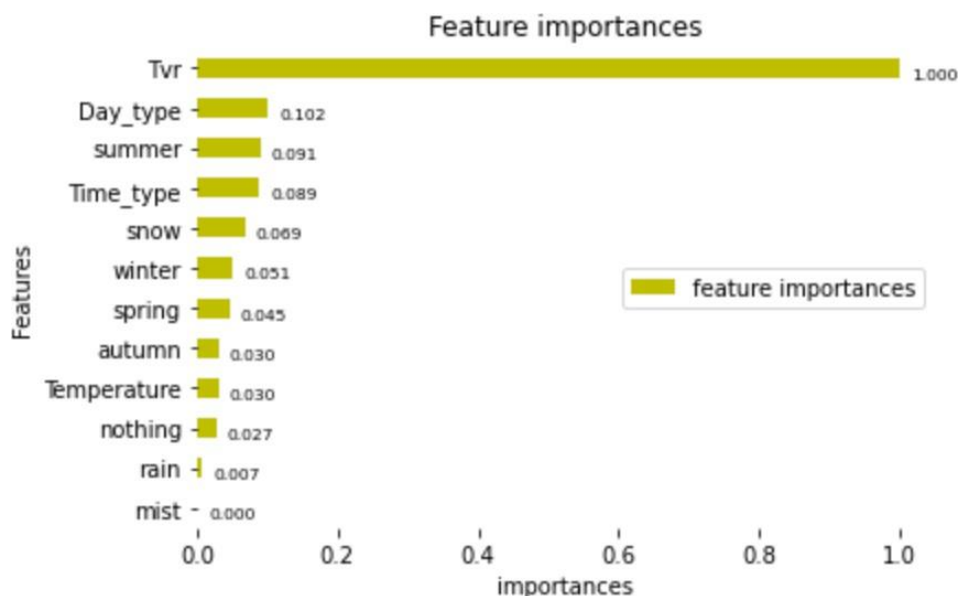


Рис. 3.14 Гістограма важливості показників

Джерело: побудовано автором на основі бази даних [Додаток Б] .

Якщо з бази даних видалити найменш важливі фактори, а саме наявність дощу, мряки та відсутність будь-яких погодніх явищ, то результати моделі частково покращяться, адже RMSE зменшиться та становитиме 0.546.

Наступний метод, який був реалізований- дерево рішень (Random Forest Regressor).

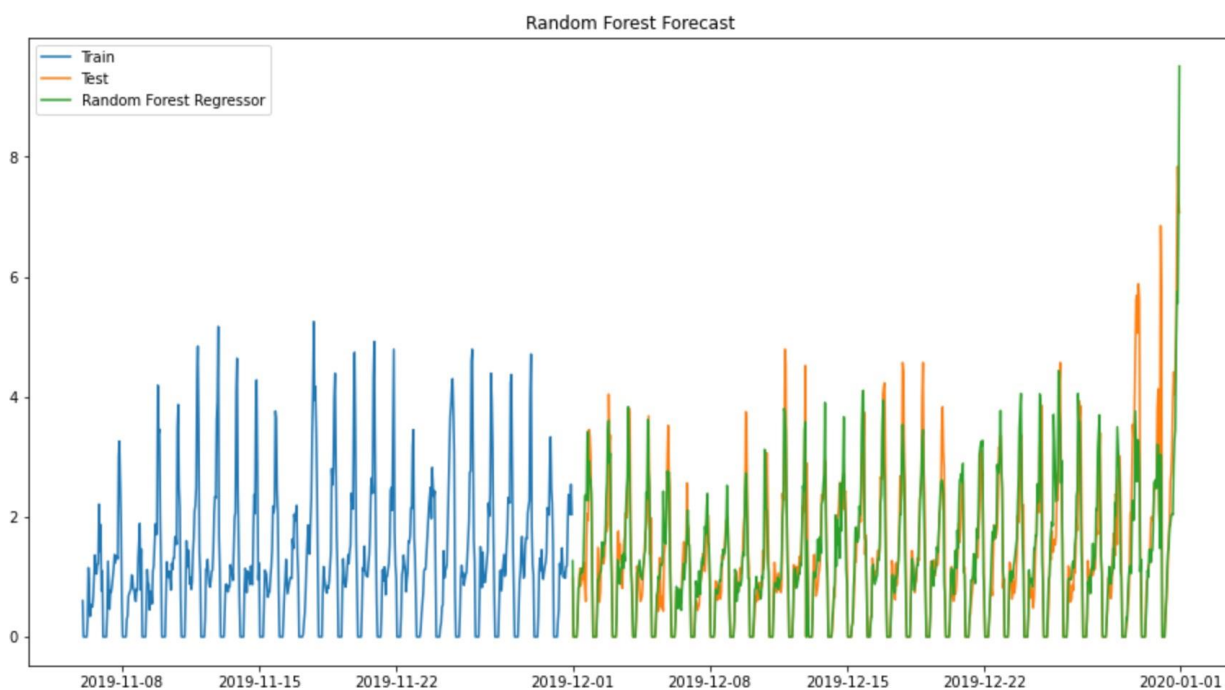


Рис. 3.15 Прогноз з використанням Random Forest Regressor

Джерело: побудовано автором на основі бази даних [Додаток Б] .

$RMSE = 0.577$. Отже можемо зробити висновки, що модель придатна для прогнозування рейтингів.

І останній крок дослідження це побудова нейронної мережі LSTM (Long short-term memory). Експериментальним шляхом, було виявлено, що найкращою є модель, у якій рекламні рейтинги залежать лише від свого попереднього значення. Тобто було змінено структуру даних таким чином, що перші 120 значення масиву впливають на 121 значення і так до кінця бази даних. Перед початком моделювання ми виконали нормалізацію даних за допомогою масштабування, використовуючи бібліотеку MinMaxScaler. Нейронна мережа має наступну структуру:

```
Model: "sequential"
```

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 120, 200)	161600
dropout (Dropout)	(None, 120, 200)	0
lstm_1 (LSTM)	(None, 120, 100)	120400
dropout_1 (Dropout)	(None, 120, 100)	0
lstm_2 (LSTM)	(None, 120, 50)	30200
dropout_2 (Dropout)	(None, 120, 50)	0
lstm_3 (LSTM)	(None, 15)	3960
dense (Dense)	(None, 1)	16

```

Total params: 316,176
Trainable params: 316,176
Non-trainable params: 0

```

Рис. 3.16 Структура нейронної мережі

Джерело: побудовано автором на основі бази даних [Додаток Б] .

Як видно на рис. 3.16, перший шар містить 200 нейронів, наступні шари мають по 100, 50 та 15 нейронів відповідно. Також в нейронну мережу було додано техніку Dropout, яка запобігає перенавчанню моделі. Як результат можна отримати одне значення на виході моделі, яке власне і буде прогнозованим рекламним рейтингом для певного періоду часу.

Після навчання моделі і побудови прогнозу були отримані результати, які представлені на рис. 3.17 Значення RMSE становить 0.506 і являє собою найкращий показник, яким ми отримували в процесі дослідження.

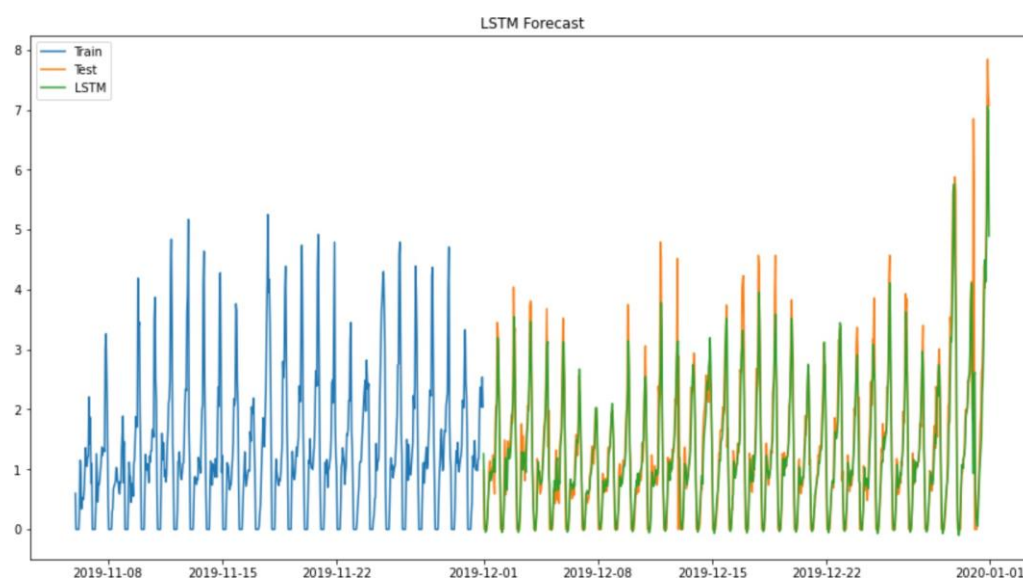


Рис. 3.17 Прогноз з використанням нейронної мережі LSTM

Джерело: побудовано автором на основі бази даних. [Додаток Б]

3.3 Обґрунтування результатів та практичне застосування

В процесі дослідження побудовано 8 моделей. У таблиці 3.1 наведені значення RMSE для кожного дослідження.

Таблиця 3.3

RMSE моделей

Модель	RMSE
«Наївна» модель	0.752
Модель Холта-Вінтерса	0.975
Лінійна регресія	0.597
Extra Trees	0.529
XGBoost	0.552
XGBoost 2	0.546
Random Forest	0.577
LSTM	0.506.

Джерело: побудовано автором на основі проведеного аналізу

Як свідчать дані табл. 3.1, нейронні мережі дають найкращий результат прогнозу. До переваг використання побудованої моделі LSTM варто додати те, що вона базується лише на минулих значеннях Avr , тобто не потрібно досліджувати поведінку інших показників. Отримана статистична модель не є громіздкою та не вимагає великої кількості ресурсу для її побудови. Для отримання більш точних результатів можна будувати прогноз лише на тиждень і постійно добавляти оновлену інформацію про рекламні рейтинги, тим самим збільшуючи базу даних.

Модель LSTM можуть застосовувати медіа холдинги, які безпосередньо забезпечують реалізацію рекламних проявів в ефірі, для більш точного прогнозу. А також дана модель може мати практичне застосування в медіа агентствах, які самостійно розміщують рекламу в телеефірі.

Для покращення результатів можна спробувати змінити структуру нейронної мережі, зменшуючи або збільшуючи кількість рівнів та нейронів.

ВИСНОВКИ

У ході дослідження рекламних рейтингів та розробки статистичної моделі для їх прогнозування були зроблена наступні висновки та пропозиції:

Ринок медіа реклами продовжує розвиватись. На сьогоднішній день компанії витрачають від 3% до 10% річного прибутку на просування своїх товарів чи послуг через телебачення. Відрекламної кампанії залежить репутація фірми, її сприйняття потенційними споживачами та майбутнє на ринку.

1. Рекламний ринок України зростає швидкими темпами. Відносний приріст об'єму ринку за останній рік становить 12.05%. На території України діє 4 найбільших медіа холдингів: «1+1 медіа», «Старлайт медіа», «Україна», «Інтер». – в залежності від розвитку яких, можна охарактеризувати розвиток медіа ринку.
2. Не дивлячись на те, що реклама на телебаченні характеризується високою вартістю, складністю створення роликів та розробки планів розміщення кампаній, вона все ж таки займає найбільшу частку медіа ринку – 43.93%. Це пояснюється тим, що ні один із видів реклами не забезпечують такої масовості та комплексного впливу на споживачів як телевізійна реклама.
3. Виділяють два основних види реклами: пряма реклама та спонсорство, які в свою чергу, мають широкий спектр підвидів, кожен з яких володіє, відмінними від інших, властивостями.
4. Панель дослідження української телеаудиторії компанії Nielsen налічує 2 840 домогосподарств, які за своїми соціально-демографічними та технічними характеристиками відтворюють структуру населення України. У кожному домогосподарстві на телевізор встановлюється ТБметр, за допомогою якого відбувається автоматична реєстрація усього перегляду телевізійного контенту. Уся інформація, отримана з піпметрів, проходить серію перевірок, після чого потрапляє до бази даних, доступної користувачам

проекту. Завдяки цим даним експерти в області досліджень і телевізійного програмінгу формують телевізійні рейтинги, які характеризують ефективність реклами.

5. Кореляційний аналіз виявив сильну лінійну залежність між Avg та Tvr та середню кореляцію між Avg та змінною $Time_type$, що описує часові проміжки, які належать до *prime time* або *off prime time*.
6. Декомпозиція часового ряду візуально продемонструвала відсутність тренду у даних та наявність річної сезонності. Для додаткового дослідження сезонності було побудовано графіки з місячною та квартальною періодичністю. Таким чином нам вдалось виявити зменшення Avg у літні періоди та збільшення у зимовий.
7. За допомогою тесту Дікі-Фулера було підтверджено гіпотезу про стаціонарність часового ряду.
8. «Наївний» прогноз – це найпростіший вид прогнозу, який показав непогані результати на досліджуваній базі даних. Даний метод є простий у розумінні та застосуванні, але не враховує зміни показників з часом.
9. Модель «Холта-Вінтерса» показала найгірший результат серед усіх досліджуваних методів. Оскільки ми прогнозували останній місяць року це завдало додаткових труднощів, адже даний період значно відрізняється від інших, тому що рейтинги новорічної ночі є найвищими в році.
10. У ході дослідження за допомогою лінійної регресії була отримана функція залежності Avg від значень температури, сезону та типу дня.
11. Застосування бібліотеки ТРОТ допомогло виявити модель, яка б давала якісний прогноз.
12. *Extra Trees Regressor* показав один з найкращих результатів прогнозування. Складність використання моделі може бути при налаштуванні параметрів.

13. За допомогою бібліотеки XGBoost було побудовано гістограму важливості показників, яка показала, що найбільший вплив на Avg мають Tvr, Day type і dummy змінна summer. Найменш важливими змінними були mist, rain і nothing, які характеризують наявність певного роду погодних явищ. Ранжування важливості змінних дало змогу удосконалити модель, виключивши з дослідження показників, що не несуть необхідної інформації.
14. Random Forest Regressor показав непогані результати при прогнозування останніх днів місяця, на відміну від Extra Trees Regressor. Але початок та середина місяця тестової вибірки та прогнозні дані мають набагато більшу різницю ніж при використанні методу «додаткових дерев».
15. Нейронна мережа LSTM показала найкращі результати прогнозування рекламних рейтингів. Оскільки даний метод має властивість до запам'ятовування поведінки часового ряду в попередні періоди, то прогнозування рейтингів для останнього місяця року та новорічної ночі, абсолютно не виявилось проблемою, на відміну від моделі Холта-Вінтерса, як згадувалось вище.
16. Для медіа холдингів є вкрай важлива висока точність прогнозованих значень рекламних рейтингів, адже у разі перевищення домовлених меж для об'ємів рейтингів, вони будуть вимушені сплачувати штрафи. Також в точності прогнозу зацікавлені медіа агентства, які самостійно розміщують клієнта на телеканалах медіа холдингу. Тому отримана у ході дослідження статистична модель буде вкрай корисною у процесі прогнозування.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Про рекламу: Закон України від 03.07.1996 № 271/96-ВР, зі змінами та доповненнями [Електронний ресурс]. – Режим доступу: <https://zakon.rada.gov.ua/laws/show/270/96-%D0%B2%D1%80>
2. Про телебачення та радіомовлення: Закон України від 21.12.1993 № 3760-ХІІ, зі змінами та доповненнями [Електронний ресурс]. – Режим доступу: <https://zakon.rada.gov.ua/laws/show/3759-12>
3. Архіпова Т.В. Особливості купівлі рекламного простору на телебаченні. Формування ринкової економіки: зб. Наук. пр. ДВНЗ «КНЕУ ім. В. Гетьмана». К.: КНЕУ, 2009, Вип. 22. С. 359-370
4. Бець М.Т. Проблеми розвитку ринку телевізійної реклами в Україні / Бець М.Т., Тимець О.А., Козак П.З.. // Науковий вісник НЛТУ. – 2010. – №21. – С. 117–124.
5. Звіт із корпоративної діяльності. // 1+1 media. – 2019. – №4. – С. 1–42.
6. Крепак А. Телевізійна реклама: зважені рейтинги у медіаплануванні / Крепак А.. // Вісник КНТЕУ. – 2017. – С. 62–72.
7. Курбанова А. Телевізійна реклама в Україні / Курбанова А. // Na chasi. – 2018. – [Електронний ресурс]. – Режим доступу: <https://nachasi.com/2018/11/23/videoreklama-v-ukrayini/>.
8. Лук'янець Т.І. Рекламний менеджмент / Лук'янець Т.І.. – Київ, 2003. – 440 с. – (КНЕУ).
9. Настасяк А.Я. Дослідження впливу факторів на телевізійні рейтинги для розміщення реклами / Настасяк А.Я., Федоренко І.К.. // Київський національний університет імені Тараса Шевченка. – 2020. – С. 44–45.
10. Настасяк А.Я. Модельовання рейтингів телевізійної реклами з використанням нейронних мереж / Настасяк А.Я., Єріна А.М.. //XVIII Міжнародна науково-практична конференція «Нові джерела та методи поширення даних у статистиці»– 2020.
11. Офіційний веб-сайт Всеукраїнської рекламної коаліції [Електронний ресурс]. – Режим доступу: <https://vrk.org.ua/ad-market/>

12. Офіційний сайт Державної служби статистики України [Електронний ресурс]. – Режим доступу: <http://www.ukrstat.gov.ua/>
13. Офіційний веб-сайт Індустріального телевізійного комітету [Електронний ресурс]. – Режим доступу: <http://www.itk.ua/ua/>
14. Офіційний веб-сайт медіа групи «Старлайт медіа» [Електронний ресурс]. – Режим доступу: <https://slm.ua/ru/>
15. Офіційний веб-сайт медіа групи «Україна» [Електронний ресурс]. – Режим доступу: <https://mgukraine.com/ru/>
16. Офіційний веб-сайт компанії Nielsen [Електронний ресурс]. – Режим доступу: <https://www.nielsen.com/ua/uk/>
17. Офіційний веб-сайт Broadcasters' Audience Research Board [Електронний ресурс]. – Режим доступу: <https://www.barb.co.uk/>
18. Офіційний веб-сайт 1+1 media [Електронний ресурс]. – Режим доступу: <https://media.1plus1.ua/>
19. Піплметр [Електронний ресурс] // Вікіпедія. – 2018. – Режим доступу до ресурсу: <https://goo.su/13hY>.
20. Почепцов Г. Теорія комунікації / Почепцов Г.. – Київ: Ваклер, 1999. – 656 с.
21. Руководство пользователя Markdata Media Workstation. // GFK. – 2007.
22. Сайт «Метеопост». Архів метеоданих. [Електронний ресурс]. – Режим доступу: <https://meteorpost.com/weather/archive/>
23. Смартфон переграв ящик. Що відбувається з ринком реклами в Україні — хто, кому і за що платить. // НВ Бізнес. – 2020. [Електронний ресурс]. – Режим доступу: <https://nv.ua/ukr/biz/markets/rinok-reklami-pidsumki-i-prognozi-internet-proti-televizora-novini-ukrajini-50067589.html>
24. Состав.юа. У 2021 витрати на рекламу у світі виростуть на 5.8%, а digital вперше займе половину всіх глобальних витрат [Електронний ресурс] / Состав.юа. – 2021. – Режим доступу до ресурсу: <https://sostav.ua/publication/u-2021-vitrati-na-reklamu-u-sv-t-virostut-na-5-8-a-digital-vpershe-zajme-87705.html>.

25. Федоренко І.К. Дослідження операцій в економіці / Федоренко І.К., Черняк О.І. – Київ: Знання, 2005. – 417 с.
26. Черняк О.І., Комашко О.В., Ставицький А.В., Баженова О.В. Економетрика. ВПЦ „Київський університет”. 2010. 359 с.
27. Черняк О.І., Ставицький А.В., Черноус Г.О.. Системи обробки економічної інформації: К.: Знання, 2006. 447 с.
28. Bergmeir, C., Hyndman, R. J., & Benítez, J. M. (2016). Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation. *International Journal of Forecasting*, 32(2), 303–312.
29. Carl H. Miller. Beer and Television: Perfectly Tuned In / Carl H. Miller. // All About Beer Magazine. – 2002.
30. Pablo Aznar. What is the difference between Extra Trees and Random Forest? [Електронний ресурс] / Pablo Aznar // QuantDare. – 2020. – Режим доступу до ресурсу: <https://quantdare.com/what-is-the-difference-between-extra-trees-and-random-forest/>.
31. P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, “Long short term memory networks for anomaly detection in time series,” in *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, vol. 89, 2015.
32. Rob J Hyndman. *Forecasting: Principles and Practice* / Rob J Hyndman, George Athanasopoulos. – Australia: Texts, 2018. – ((2nd ed)).
33. Scikit-learn Machine Learning in Python [Електронний ресурс] – Режим доступу до ресурсу: <https://scikit-learn.org/stable/>.
34. Steven Elsworth. *Time Series Forecasting Using LSTM Networks: A Symbolic Approach* / Steven Elsworth, Stefan Guttel. – 2020.
35. TPOT API [Електронний ресурс] – Режим доступу до ресурсу: <http://epistasislab.github.io/tpot/api/>.
36. Webb P. Effects of TV clutter / Webb P., Ray M.. // *Journal of Advertising Research*. – 1971. – №19. – С. 7–12.

ДОДАТКИ

Додаток А

Зразок обробленої бази даних

Date_Time	Day type	Time type	Avr	Tvr	Temp.	autumn	spring	summer	winter	mist	nothing	rain	snow
2018-01-01 06:00:00	0	0	1,17	8,15	1	0	0	0	1	0	0	1	0
2018-01-01 07:00:00	0	1	0,7	6,27	1	0	0	0	1	0	1	0	0
2018-01-01 08:00:00	0	1	0,55	10,1	1	0	0	0	1	0	1	0	0
2018-01-01 09:00:00	0	1	0,83	19,94	2	0	0	0	1	0	1	0	0
2018-01-01 10:00:00	0	0	1,51	23,44	2	0	0	0	1	0	1	0	0
2018-01-01 11:00:00	0	0	1,7	26,72	2	0	0	0	1	0	1	0	0
2018-01-01 12:00:00	0	0	2,39	28,11	3	0	0	0	1	0	1	0	0
2018-01-01 13:00:00	0	0	1,87	28,13	3	0	0	0	1	0	1	0	0
2018-01-01 14:00:00	0	0	1,91	28,48	3	0	0	0	1	0	1	0	0
2018-01-01 15:00:00	0	0	1,66	31,87	3	0	0	0	1	0	1	0	0
2018-01-01 16:00:00	0	0	2,99	30,17	3	0	0	0	1	0	1	0	0
2018-01-01 17:00:00	0	0	4,31	33,48	3	0	0	0	1	0	1	0	0
2018-01-01 18:00:00	0	1	4,19	35,39	3	0	0	0	1	0	1	0	0
2018-01-01 19:00:00	0	1	4,45	37,24	3	0	0	0	1	0	1	0	0
2018-01-01 20:00:00	0	1	6,24	39,07	3	0	0	0	1	0	1	0	0
2018-01-01 21:00:00	0	1	6,8	38,14	3	0	0	0	1	0	1	0	0
2018-01-01 22:00:00	0	1	5,87	30,58	3	0	0	0	1	0	1	0	0
2018-01-01 23:00:00	0	0	5,4	16,67	3	0	0	0	1	0	1	0	0

Python код статистичного аналізу

```
#import libraries

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import warnings

import seaborn as sns

from statsmodels.tsa.ar_model import AR,ARResults

from statsmodels.graphics.tsaplots import month_plot,quarter_plot

from statsmodels.tsa.seasonal import seasonal_decompose

from sklearn.metrics import mean_squared_error

from math import sqrt

from statsmodels.tsa.holtwinters import ExponentialSmoothing

from statsmodels.tsa.stattools import adfuller

from sklearn.linear_model import LinearRegression

from tpot import TPOTRegressor

import xgboost as xgb

from xgboost import XGBRegressor

from sklearn.ensemble import RandomForestRegressor

import math

import keras

import tensorflow
```

```
from keras.models import Sequential

from keras.layers import Dense

from keras.layers import LSTM

from sklearn.preprocessing import MinMaxScaler

from sklearn.metrics import mean_squared_error

from keras.preprocessing.sequence import TimeseriesGenerator

from sklearn.preprocessing import MinMaxScaler

from keras.models import Sequential

from keras.layers import Dense

from keras.layers import LSTM

from keras.layers import Dropout

warnings.filterwarnings('ignore')

#read data and check it

df=pd.read_excel('data.xlsx')

df.head()

#change type for date variable

df.Date_Time=pd.to_datetime(df.Date_Time)

df.index=df.Date_Time

df.asfreq(freq='H')

#create dummy variables

df.Day_type=np.where(df.Day_type=="Workday",0,1)

df.Time_type=np.where(df.Time_type=="Prime time",1,0)

df=pd.concat([df, pd.get_dummies(df.Season)], axis=1)
```

```

df=pd.concat([df, pd.get_dummies(df.Weather)], axis=1)

#drop not useful columns

df.drop(columns=['Channel','Season','Weather','Date','Time','TimeBand','Shr','Date_Time'],axis=1,inplace=True)

#descriptive statistics

df[["Time_type",'Avr','Tvr','Temperature','autumn','spring','summer','winter','mist','nothing','rain','snow']].describe()

#correlation

f, ax = plt.subplots(figsize=(15, 7))

ax=sns.heatmap(df.corr(),annot=True)

# month_plot and quarter_plot

dfq = df['Avr'].resample(rule='MS').mean()

ax=month_plot(dfq)

dfq = df['Avr'].resample(rule='Q').mean()

quarter_plot(dfq);

#seasonal decompose

dfq = df['Avr'].resample(rule='D').mean()

seasonal_decompose(dfq, model='add').plot();

# train test split

train=df.loc[:"2019-01-12 06:00:00"]

test=df.loc["2019-01-12 06:00:00":]

x_train=pd.DataFrame(train.Avr)

```

```

x_test=pd.DataFrame(test.Avr)

#Naive forecast

dfq = pd.DataFrame(df['Avr'].resample(rule='H').mean())

dfq['Avr']=np.where(dfq['Avr'].isna()==True,0,dfq['Avr'])

dfq=pd.Series(dfq['Avr'])

train=dfq.loc[:"2019-12-01 00:00:00"]

train=train[:-1]

test=dfq.loc["2019-12-01 00:00:00":]

dd= np.asarray(train)

y_hat = pd.DataFrame(test.copy())

y_hat['naive'] = train.loc["2018-12-01 00:00:00":"2018-12-31 23:00:00"].values

plt.figure(figsize=(12,8))

plt.plot(train.index[-600:], train[-600:], label='Train')

plt.plot(test.index, test, label='Test')

plt.plot(y_hat.index,y_hat['naive'], label='Naive Forecast')

plt.plot(pd.date_range(start="2020-01-01",end="2020-01-31",
freq="H"),train.loc["2019-01-01 00:00:00":"2019-01-31 00:00:00"].values,c='g')

plt.legend(loc='best')

plt.title("Naive Forecast")

plt.show()

rms = sqrt(mean_squared_error(test.values, y_hat.naive))

print("RMSE: ", rms)

print("Mean value Avr: ", test.mean())

```

```

#Holt_winters

hw_fit = ExponentialSmoothing(train ,seasonal_periods=8136, seasonal='add').fit()

hw_pred = hw_fit.forecast(744+700).clip(lower=0)

plt.figure(figsize=(15,8))

plt.plot(train.index[-600:], train[-600:], label='Train')

plt.plot(test.index, test, label='Test')

plt.plot(hw_pred.index, hw_pred, label='Holt-Winters')

plt.legend(loc='best')

plt.title("Holt-Winters Forecast")

plt.show()

rms = sqrt(mean_squared_error(test.values, hw_pred[:744].values))

print("RMSE: ", rms)

print("Mean value Avr: ", test.mean())

# Dickey-Fuller Test

def adf_test(series,title=""):

    print(f'Augmented Dickey-Fuller Test: {title}')

    result = adfuller(series.dropna(),autolag='AIC') # .dropna() handles differenced data

    labels = ['ADF test statistic','p-value','# lags used','# observations']

    out = pd.Series(result[0:4],index=labels)

    for key,val in result[4].items():

        out[f'critical value ({key})']=val

    print(out.to_string())

    if result[1] <= 0.05:

```

```

print("Strong evidence against the null hypothesis")

print("Reject the null hypothesis")

print("Data has no unit root and is stationary")

else:

    print("Weak evidence against the null hypothesis")

    print("Fail to reject the null hypothesis")

    print("Data has a unit root and is non-stationary")

adf_test(df['Avr'])

#liner regression

dfq = pd.DataFrame(df.resample(rule='H').mean())

dfq['Avr']=np.where(dfq['Avr'].isna()==True,0,dfq['Avr'])

dfq['Tvr']=np.where(dfq['Tvr'].isna()==True,0.05,dfq['Tvr'])

dfq['Time_type']=np.where(dfq['Time_type'].isna()==True,0,dfq['Time_type'])

dfq['nothing']=np.where(dfq['nothing'].isna()==True,1,dfq['nothing'])

dfq['mist']=np.where(dfq['mist'].isna()==True,0,dfq['mist'])

dfq['rain']=np.where(dfq['rain'].isna()==True,0,dfq['rain'])

dfq['snow']=np.where(dfq['snow'].isna()==True,0,dfq['snow'])

for i in range(dfq.shape[0]):

    if pd.isna(dfq.Temperature[i]):

        dfq.Temperature[i]=dfq.Temperature[i-1]

        dfq.Day_type[i]=dfq.Day_type[i-1]

        dfq.autumn[i]=dfq.autumn[i-1]

        dfq.spring[i]=dfq.spring[i-1]

```

```
dfq.summer[i]=dfq.summer[i-1]

dfq.winter[i]=dfq.winter[i-1]

train=dfq.loc[:"2019-12-01 00:00:00"]

train=train[:-1]

test=dfq.loc["2019-12-01 00:00:00":]

y_train=train['Avr']

train.drop(['Avr'],axis=1,inplace=True)

x_train=train

y_test=test['Avr']

test.drop(['Avr'],axis=1,inplace=True)

x_test=test

regressor = LinearRegression()

regressor.fit(x_train,y_train)

print(regressor.intercept_)

print(regressor.coef_)

y_pred = regressor.predict(x_test)

plt.figure(figsize=(15,8))

plt.plot(y_train.index[-600:], y_train[-600:], label='Train')

plt.plot(y_test.index, y_test, label='Test')

plt.plot(y_test.index,y_pred, label='Liner regression')

plt.legend(loc='best')

plt.title("Liner regression Forecast")

plt.show()
```

```
rms = sqrt(mean_squared_error(y_test.values, y_pred))

print("RMSE: ", rms)

print("Mean value Avr: ", y_test.mean())

#TPOT

tpot = TPOTRegressor(generations=10, population_size=15, verbosity=2)

tpot.fit(x_train, y_train)

tpot_best = tpot.fitted_pipeline_

tpot_best.fit(x_train, y_train)

tpot_valid_pred = tpot_best.predict(x_test)

plt.figure(figsize=(15,8))

plt.plot(y_train.index[-600:], y_train[-600:], label='Train')

plt.plot(y_test.index, y_test, label='Test')

plt.plot(y_test.index, tpot_valid_pred, label='Extra Trees Regressor')

plt.legend(loc='best')

plt.title("Extra Trees Regressor Forecast")

plt.show()

rms = sqrt(mean_squared_error(y_test.values, tpot_valid_pred))

print("RMSE: ", rms)

print("Mean value Avr: ", y_test.mean())

#XGBoost

def_xgboost_params = dict(objective='reg:linear', nthread=-1, seed=24, n_jobs=-1)

def_xgboost_model = XGBRegressor(**def_xgboost_params)

def_xgboost_model.fit(x_train, y_train)
```

```

def_xgboost_valid_pred = def_xgboost_model.predict(x_test)

plt.figure(figsize=(15,8))

plt.plot(y_train.index[-600:], y_train[-600:], label='Train')

plt.plot(y_test.index, y_test, label='Test')

plt.plot(y_test.index,def_xgboost_valid_pred, label='XGBoost')

plt.legend(loc='best')

plt.title("XGBoost Forecast")

plt.show()

rms = sqrt(mean_squared_error(y_test.values, def_xgboost_valid_pred))

print("RMSE: ", rms)

print("Mean value Avr: ", y_test.mean())

#Feature importances

fi_def_xgboost = pd.DataFrame({'feature': x_train.columns.tolist(), 'feature
importances': def_xgboost_model.feature_importances_})

fi_def_xgboost.sort_values(by=['feature importances'],ascending=False,
inplace=True)

fi_def_xgboost['feature'].replace({'active': 'feature'}, inplace=True, regex=True)

fi_def_xgboost.set_index('feature', inplace=True)

minmaxscaler = MinMaxScaler()

fi_def_xgboost[['feature importances']] =
minmaxscaler.fit_transform(fi_def_xgboost[['feature importances']])

fi_def_xgboost

ax = fi_def_xgboost.sort_values(by='feature importances').tail(20).plot(kind='barh',
color = 'dimgray')

```

```

for p in ax.patches:

    ax.annotate(str("%.3f" % p.get_width()), #str(p.get_width()), (p.get_x() +
p.get_width(), p.get_y()), xytext=(5,0), textcoords='offset points',
horizontalalignment='left', fontsize=7.5)

for spine in plt.gca().spines.values():

    spine.set_visible(False)

plt.title('Feature importances')

plt.xlabel('importances')

plt.ylabel('Features')

plt.legend(loc="center right")

plt.show()

#Drop not usefule features

x_train2=x_train[['Day_type','Time_type','Tvr','Temperature','autumn','spring','summer',
'winter','snow']]

x_test2=x_test[['Day_type','Time_type','Tvr','Temperature','autumn','spring','summer',
'winter','snow']]

def_xgboost_model.fit(x_train2, y_train)

def_xgboost_valid_pred = def_xgboost_model.predict(x_test2)

plt.figure(figsize=(15,8))

plt.plot(y_train.index[-600:], y_train[-600:], label='Train')

plt.plot(y_test.index, y_test, label='Test')

plt.plot(y_test.index,def_xgboost_valid_pred, label='XGBoost')

plt.legend(loc='best')

```

```

plt.title("XGBoost Forecast")

plt.show()

rms = sqrt(mean_squared_error(y_test.values, def_xgboost_valid_pred))

print("RMSE: ", rms)

print("Mean value Avr: ", y_test.mean())

# Random Forest Regressor

def_rf_params = dict(random_state=42, n_jobs=-1)

def_rf_model = RandomForestRegressor(**def_rf_params)

def_rf_model.fit(x_train, y_train)

def_rf_valid_pred = def_rf_model.predict(x_test)

plt.figure(figsize=(15,8))

plt.plot(y_train.index[-600:], y_train[-600:], label='Train')

plt.plot(y_test.index, y_test, label='Test')

plt.plot(y_test.index,def_rf_valid_pred, label='Random Forest Regressor')

plt.legend(loc='best')

plt.title("Random Forest Forecast")

plt.show()

#LSTM

sc = MinMaxScaler(feature_range = (0, 1))

train=dfq.loc[:"2019-12-01 00:00:00"]

train=train[:-1]

test=dfq.loc["2019-12-01 00:00:00":]

y_train=train['Avr']

```

```
train.drop(['Avr'],axis=1,inplace=True)

x_train=train

y_test=test['Avr']

test.drop(['Avr'],axis=1,inplace=True)

x_test=test

training_set=pd.DataFrame(y_train)

training_set_scaled = sc.fit_transform(training_set)

X_train = []

y_train = []

for i in range(120, 16776):

    X_train.append(training_set_scaled[i-120:i, 0])

    y_train.append(training_set_scaled[i, 0])

X_train, y_train = np.array(X_train), np.array(y_train)

X_train = np.reshape(X_train, (X_train.shape[0], X_train.shape[1], 1))

regressor = Sequential()

regressor.add(LSTM(units = 200, return_sequences = True, input_shape =
(X_train.shape[1], 1)))

regressor.add(Dropout(0.2))

regressor.add(LSTM(units = 100, return_sequences = True))

regressor.add(Dropout(0.2))

regressor.add(LSTM(units = 50, return_sequences = True))

regressor.add(Dropout(0.2))

regressor.add(LSTM(units = 15))
```

```

regressor.add(Dense(units = 1))

regressor.compile(optimizer = 'adam', loss = 'mean_squared_error')

regressor.fit(X_train, y_train, epochs = 10, batch_size = 24)

train=dfq.loc["2019-12-01 00:00:00"]

train=train[:-1]

y_train=train['Avr']

inputs=np.array(list(y_train[-120:].values)+list(y_test.values))

inputs = inputs.reshape(-1,1)

inputs = sc.transform(inputs)

X_test = []

for i in range(120, 744+120):

    X_test.append(inputs[i-120:i, 0])

X_test = np.array(X_test)

X_test = np.reshape(X_test, (X_test.shape[0], X_test.shape[1], 1))

predicted_ratings= regressor.predict(X_test)

predicted = sc.inverse_transform(predicted)

train=dfq.loc["2019-12-01 00:00:00"]

train=train[:-1]

test=dfq.loc["2019-12-01 00:00:00":]

plt.figure(figsize=(15,8))

plt.plot(train.index[-600:], train['Avr'][-600:], label='Train')

plt.plot(y_test.index, y_test, label='Test')

plt.plot(y_test.index, predicted_ratings, label='LSTM')

```

```
plt.legend(loc='best')  
  
plt.title("LSTM Forecast")  
  
plt.show()  
  
rms = sqrt(mean_squared_error(y_test.values, predicted_stock_price))  
  
print("RMSE: ", rms)  
  
print("Mean value Avr: ", y_test.mean())
```