

УДК 004.896

DOI <https://doi.org/10.32782/IT/2024-1-10>

**Іван ПІНЦАК**

аспірант факультету інформаційних технологій, ДВНЗ «Ужгородський національний університет», вул. Підгірна, 46, м. Ужгород, Україна, 88000

ORCID: 0000-0003-2084-1934

**Бібліографічний опис статті:** Пінцак, І. (2024). Використання машинного навчання та аналізу даних для прогнозування тенденцій у електронній комерції. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 1, 80–88, doi: <https://doi.org/10.32782/IT/2024-1-10>

## ВИКОРИСТАННЯ МАШИННОГО НАВЧАННЯ ТА АНАЛІЗУ ДАНИХ ДЛЯ ПРОГНОЗУВАННЯ ТЕНДЕНЦІЙ У ЕЛЕКТРОННІЙ КОМЕРЦІЇ

Дана стаття розглядає застосування машинного навчання та аналізу даних як ключових інструментів для прогнозування тенденцій у електронній комерції. Вона висвітлює потенціал цих технологій у вирішенні завдань прогнозування на основі великої кількості даних, а також їхню ефективність у виявленні складних зв'язків та розвитку прогностичних моделей. **Мета роботи.** Стаття ставить за мету дослідити, які можливості надають ці технології для виявлення складних зв'язків та аналізу даних з метою прогнозування майбутніх тенденцій, що допомагає організаціям та галузям, зокрема електронній комерції, ефективніше планувати свою діяльність, а також вирішувати різноманітні проблеми. У даній статті використовується **методологія**, спрямована на аналіз зв'язку між конверсіями з посилань у публікаціях на веб-сайті та кількістю замовлень. Підхід передбачає застосування регресійного аналізу в програмному середовищі R-Studio. Перш ніж провести аналіз, дані піддаються перевірці на гетероскедастичність, автокореляцію та нормальність розподілу залишкового ряду. Використано методи, такі як метод Глейзера для виявлення гетероскедастичності та підхід Дарбіна-Ватсона для оцінки автокореляції. Після підтвердження адекватності даних за вищезазначеними критеріями, застосовується лінійна регресія для встановлення залежності між кількістю покупок та кількістю відвідувань веб-сайту. Результати аналізу показують високий рівень залежності між цими показниками, з конверсією сайту на рівні 20%. **Наукова новизна** полягає в обґрунтуванні теоретико-методологічних підходів, спрямованих на вдосконалення процедур оптимізації та зменшення використання маркетингових ресурсів українським бізнесом з метою збільшення електронних продажів, а також в оцінці фінансового ефекту впровадження запропоноване моделювання на практиці. **Висновки.** Результати дослідження продемонстрували, що розвиток значного масиву цифрових даних у поєднанні з технологією їх обробки та зберігання, а також наявність математичних інструментів для оцінки цих даних призводить до підвищення ефективності бізнес діяльності. Методи обробки та аналізу даних включають штучні нейронні мережі, машинне навчання, просторовий аналіз, кластерний аналіз, регресійний аналіз, класифікацію та інтелектуальний аналіз даних.

**Ключові слова:** електронна комерція, машинне навчання, прогнозування, регресійний аналіз.

**Ivan PINTSAK**

Postgraduate student of the Faculty of Information Technology, SHEI "Uzhhorod National University", 46, Pidhirna Str., Uzhhorod, Ukraine, 88000, [official@uzhnu.edu.ua](mailto:official@uzhnu.edu.ua)

ORCID: 0000-0002-5190-841X

**To cite this article:** Pintsak, I. (2024). Vykorystannya mashynnoho navchannya ta analizu danykh dlya prohnzuvannya tendentsiy v elektronniy komertsiiy [The use of machine learning and data analysis for predicting trends and patterns in various industries]. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 1, 80–88, doi: <https://doi.org/10.32782/IT/2024-1-10>

## THE USE OF MACHINE LEARNING AND DATA ANALYSIS FOR PREDICTING TRENDS IN E-COMMERCE

This article explores the application of machine learning and data analysis as key tools for predicting trends in e-commerce. It highlights the potential of these technologies in addressing predicting tasks based on large datasets and their effectiveness in uncovering complex relationships and developing predictive models. **Aim of the study.** The article aims to investigate the capabilities of these technologies in detecting intricate connections and analyzing data to predict future trends. This assists organizations and industries, including e-commerce, in more effectively planning their activities and addressing diverse challenges. The methodology employed in this article focuses

on analyzing the correlation between link conversions in website publications and the number of orders. The **approach** involves using regression analysis in the R-Studio software environment. Before conducting the analysis, the data undergoes checks for heteroskedasticity, autocorrelation, and normality of residual distribution. Methods such as the Glejser method for detecting heteroskedasticity and the Durbin-Watson approach for autocorrelation assessment are utilized. Following confirmation of data adequacy based on the aforementioned criteria, linear regression is applied to establish the relationship between the number of purchases and website visits. The analysis results demonstrate a high level of dependence between these indicators, with a website conversion rate of 20%. **The scientific novelty** lies in the justification of theoretical and methodological approaches aimed at improving optimization procedures and reducing the use of marketing resources by Ukrainian businesses to increase online sales, as well as in assessing the financial effect of implementing the proposed modeling in practice. **Conclusions.** The results of the study demonstrated that the development of a significant amount of digital data in combination with the technology of its processing and storage, as well as the availability of mathematical tools for evaluating this data leads to an increase in the efficiency of business activities. Data processing and analysis methods include artificial neural networks, machine learning, spatial analysis, cluster analysis, regression analysis, classification, and data mining.

**Key words:** E-commerce, machine learning, forecasting, regresearch analysis.

Машинне навчання – це підгалузь штучного інтелекту, яка зосереджена на розробці алгоритмів і моделей, які дозволяють комп'ютерам вчитися на основі даних і вдосконалювати свою продуктивність у виконанні конкретних завдань без явного програмування.

Під загальним уявленням про аналіз даних розуміється процес збору, очищення, організації та інтерпретації великого обсягу даних з метою виявлення цінних знань, закономірностей та патернів. Аналіз даних може включати застосування різних методів, таких як статистичний аналіз, машинне навчання, аналіз текстів, класифікація тощо, залежно від конкретних цілей дослідження.

Важливість прогнозування тенденцій полягає в тому, що воно допомагає галузям або організаціям адаптуватися до майбутніх змін і використовувати знання про тенденції для прийняття обґрунтованих стратегічних рішень. Наприклад, прогнозування тенденцій у фінансовій галузі може допомогти банкам і фінансовим установам прогнозувати ризики та визначати оптимальні інвестиційні можливості.

Використання машинного навчання та аналізу даних у прогнозуванні тенденцій стало надзвичайно потужним інструментом. Алгоритми машинного навчання можуть аналізувати великі обсяги структурованих та неструктурованих даних, що дозволяє точніше визначати залежності і патерни. Це робить прогнози більш об'єктивними та надійними, що в свою чергу допомагає уникнути потенційних помилок у прийнятті рішень.

Областю цього дослідження є електронна комерція. Запуск нового методу збору та аналізу даних може вплинути на організацію, оскільки результат може бути позитивним або навпаки. Платформи електронної комерції збирають великі обсяги даних та зберігають їх у своїх центрах обробки даних. Вони не розглядають

це як перевагу для своїх ділових можливостей, таких як аналіз даних та їх моделей за останні роки. Наприклад, усі дані клієнтів з реєстрації, історії пошуку, продажів, чатів зберігаються на сервері і будуть використовуватися лише у разі проблеми з існуючими даними. Зрозуміло, чому вони не хочуть, щоб інша компанія аналізувала їхні дані через проблеми з конфіденційністю, але вони також можуть створити свою власну команду для аналізу даних, які можуть бути вигідними.

**Метою** цієї статті є вивчення важливості та застосування машинного навчання та аналізу даних для прогнозування тенденцій та патернів у різних галузях. Стаття ставить за мету дослідити, які можливості надають ці технології для виявлення складних зв'язків та аналізу даних з метою прогнозування майбутніх тенденцій, що допомагає організаціям та галузям, зокрема електронній комерції, ефективніше планувати свою діяльність, а також вирішувати різноманітні проблеми.

Досліджувана галузь витрачає сотні мільйонів доларів на рекламу, соціальні мережі, сортування захищених даних та багато іншого, щоб збільшити продажі, але вони не розуміли, що за допомогою машинного навчання вони зможуть випередити своїх конкурентів. Машинне навчання – це велика деревоподібна гілка, яка має безліч спеціалізацій, таких як датамайнінг, штучний інтелект, доповнена реальність та прогнозування (Mitchell, T., 1997; Hastie, T., 2009). У цьому дослідженні розглядається лише аспект прогнозування з допомогою машинного навчання та аналізу даних (Goodfellow, I., 2016; Bishop, C. M., 2006; Zhao, K., 2017).

Завдяки можливості прогнозування з використанням алгоритму машинного навчання для електронної комерції можна виявляти будь-які приховані шаблони, викиди, точки інтересу та багато іншого. Це дозволяє електронній

комерції мати можливість правильно визначити важливі деталі у кожному аспекті. Вони зможуть використовувати всі свої дані, такі як кількість придбаного продукту, категорії продуктів, спосіб оплати, процентна ставка, тривалість доставки та місцезнаходження клієнта, щоб краще зрозуміти, як покращити свої продажі та керувати ними.

Якщо платформа електронної комерції зможе прогнозувати свої продажі на наступний рік, місяць чи день, вони зможуть приймати ефективніші бізнес-рішення. Вони також зможуть відстежувати тенденції у своїх продажах, якщо щорічно проводиться будь-який фестиваль або захід. Вони також зможуть відстежувати свої запаси, щоб усі товари були у достатній кількості, що дозволить уникнути перетоварювання чи нестачі товару, оскільки вони зможуть отримати приблизну оцінку покупок, які можуть статися. Мало того, вони також зможуть краще відстежувати свої фінанси та робити розумні покупки, а також мати належний бюджет упродовж усієї комерційної діяльності.

### 1. Методи машинного навчання

Машинне навчання – це підгалузь штучного інтелекту, яка дозволяє комп'ютерам вчитися на основі даних і покращувати свою продуктивність у виконанні завдань без явного програмування.

#### 1. Навчання з учителем (*Supervised Learning*):

– Пояснення поняття навчання з учителем: це вид машинного навчання, де алгоритм навчається на основі вхідних даних та відповідних міток або правильних відповідей.

– Застосування: навчання з учителем широко використовується у задачах класифікації та регресії. Наприклад, визначення, чи є електронний лист спамом чи неспамом, або прогнозування ціни на нерухомість на основі історичних даних.

#### 2. Навчання без учителя (*Unsupervised Learning*):

– Пояснення поняття навчання без учителя: це вид машинного навчання, де алгоритм навчається на неозначених даних без міток або правильних відповідей.

– Застосування: навчання без учителя використовується для кластеризації даних та виявлення прихованих залежностей. Наприклад, групування споживачів за спільними характеристиками для розвитку бізнес-стратегій.

#### 3. Півнаглядване навчання (*Semi-Supervised Learning*):

– Пояснення поняття півнаглядваного навчання: це підхід, який комбінує навчання з учителем та навчання без учителя, використовуючи дані з мітками та без міток.

– Застосування: півнаглядване навчання ефективно використовується там, де отримати повний набір міток є дорогим або складним завданням. Наприклад, в задачах обробки медичних зображень, де надійні експертні мітки можуть бути обмеженими.

Прикладами методів аналізу даних для виявлення патернів та закономірностей є методи регресії, алгоритми класифікації, методи кластеризації, нейронні мережі, аналіз тексту та обробка природної мови. Розглянемо кожен з них більш детально.

#### *Методи регресії:*

– Пояснення поняття регресії: регресія є методом машинного навчання, що дозволяє встановлювати залежність між залежною та незалежною змінними і використовується для прогнозування числових значень.

– Застосування: регресія широко використовується для прогнозування цін на товари, прогнозування доходів, попиту на продукцію та інших числових параметрів.

#### *Алгоритми класифікації:*

– Пояснення поняття класифікації: класифікація є методом машинного навчання, де алгоритм навчається розподіляти дані в категорії або класи на основі характеристик даних.

– Застосування: класифікація використовується для розпізнавання образів, діагностики медичних захворювань, виявлення шахрайства в фінансових транзакціях та багатьох інших завдань.

#### *Методи кластеризації:*

– Пояснення поняття кластеризації: кластеризація є методом машинного навчання, де дані групуються в класи або кластери на основі схожості між ними.

– Застосування: кластеризація використовується для сегментації клієнтів у маркетингових дослідженнях, групування новинних статей за тематикою, класифікації геномних даних у біоінформатиці та інших задач.

#### *Нейронні мережі:*

– Пояснення поняття нейронних мереж: нейронні мережі – це алгоритми, інспіровані структурою людського мозку, що використовуються для виявлення складних залежностей в даних.

– Застосування: нейронні мережі успішно використовуються для обробки зображень, мовних даних, аналізу тексту, рекомендаційних систем та багатьох інших сфер.

#### *Аналіз тексту та обробка природної мови:*

– Пояснення поняття аналізу тексту та обробки природної мови: це галузь машинного навчання, що дозволяє аналізувати та розуміти людську мову.

– Застосування: аналіз тексту та обробка природної мови використовуються для автоматичної категоризації текстів, аналізу настроїв у соціальних мережах, перекладу мов та інших застосувань.

В свою чергу, прогнозування продажів є важливим завданням, яке має виконувати електронна комерція, і прогнозування може вплинути на процес прийняття бізнес-рішень. Мало того, що, маючи прогноз продажів для платформи електронної комерції, вони можуть краще зрозуміти своє фінансове становище для управління робочою силою та подальшого покращення управління своїм ланцюжком поставок.

Грунтуючись на (Zhao, K., Wang, C., 2017) і (Bandara, K., 2019), прогноз продажів дозволяє платформі електронної комерції мати більш точні та надійні прогнози, які допоможуть їм у плануванні запасів, конкурентоспроможних цінах та стратегії своєчасного просування.

Згідно з (Li, M., Ji, S., Liu, G., 2018), прогноз продажів електронної комерції дозволяє зрозуміти життєвий цикл платформи електронної комерції як її продажу та зростання, стабільність, спад і те, як на продаж впливають короткострокові цілі продукту, такі як просування, ціноутворення, сезон та рейтинг в Інтернеті.

Згідно з дослідженням, проведеним (Zhao, K., & Wang, C., 2017), використовувався алгоритм згортової нейронної мережі (CNN) для прогнозування продажів в електронній комерції. Це дослідження було проведено для усунення виявленого обмеження, яке полягало в тому, що метод вимагав індивідуальної ручної розробки функцій для конкретних сценаріїв, що складно та забирає багато часу і вимагає великої кількості експертних знань. Однак мета цього дослідження полягала в тому, щоб визначити, чи може цей підхід автоматично отримувати ефективні функції та забезпечувати прогнозування продажів на основі вилучених функцій, як згадувалося в (Zhao, K., & Wang, C., 2017). Основним алгоритмом, який використовувався для дослідження, був алгоритм CNN для прогнозування продажів. Однак для цілей порівняння у дослідженні були обрані також алгоритми ARIMA, DNN, TL та WD, щоб отримати найточніші результати для прогнозування продажів. Використовувалися також зменшення та перенесення ваги зразка, методи навчання для подальшого підвищення точності прогнозування, які довели свою високу ефективність у експериментах. Результати дослідження показали, що MST модель ARIMA має найвище середнє значення, однак алгоритм CNN досяг своєї мети, оскільки він може автоматично

отримувати ефективні функції та виконувати прогнозування продажів із використанням своїх функцій.

Грунтуючись на дослідженнях, проведених (Bandara, K., 2019) і (Li, M., Ji, S., Liu, G., 2018), можна дійти до висновку, що використання також алгоритмів нейронної мережі є досить доцільним. Обидва ці алгоритми нейронної мережі мають свій власний підхід, у якому в дослідженні 2018 року використовується нелінійна авторегресивна нейронна мережа (NARNN), а в дослідженні 2019 року проводилося використання рекурентних нейронних мереж (RNN) та мереж із довготривалою короткочасною пам'яттю (LSTM). Алгоритм роботи реалізовувався у межах спеціальної нейронної мережі. Ці дослідники використали власний алгоритм цього підходу для прогнозування продажу та попиту в електронній комерції. Проблема, про яку йдеться у дослідженнях, аналогічна: це складність у визначенні різних моделей попиту/продажів між продуктами та наявними кореляціями. Ціль обох досліджень полягала в тому, щоб створити систематичну структуру попередньої обробки для подолання проблем в умовах електронної комерції, а також створити основу для прогнозування. Алгоритм, який використовувався для порівняння цих досліджень був ARIMA. Обговорення результатів дослідження 2018 показало, що помилка прогнозу для NARNN становить 0,1016, а для ARIMA – 0,1389, що показує, що NARNN має нижчу частоту помилок порівняно з ARIMA. Результати Дослідження 2019 року також показують, що LSTM має більш низьке середнє значення та медіану порівняно з ARIMA.

Прогнозування продажів зазвичай робиться з використанням найпоширенішого методу – аналізу часових рядів. Аналіз часових рядів включає функцію авторегресії, яка допомагає будь-якому типу прогнозного аналізу.

Відповідно до (Pavlyshenko, B. M., 2019) дослідження – використовувалися моделі машинного навчання для прогнозування часових рядів продажів. У ньому згадується, що прогнозування продажів є сучасним методом бізнес-аналітики. Також згадується (Li, M., Ji, S., Liu, G., 2018), що модель ARIMA має найкращий підхід до продуктивності прогнозування при аналізі часових рядів. Основна проблема, заявлена в цьому дослідженні (Pavlyshenko, B. M., 2019), полягає в тому, що для даних тимчасових рядів потрібні великі дані для відображення сезонності, а великі дані про транзакційні продажі можуть мати багато відсутніх даних та викидів. Потім в цих даних потрібно буде враховувати

безліч різних факторів, які можуть вплинути на продаж. Мета цього аналізу тимчасових рядів у тому, щоб об'єднати різні алгоритми тимчасових рядів та щоб підвищити точність цього прогнозу. У дослідженні вибрали п'ять алгоритмів: ExtraTree, ARIMA, RandomForest, Lasso і NeuralNetwork, які є алгоритмами часових рядів. На основі результатів тестування помилок прогнозування, ExtraTree має найвищу помилку перевірки в порівнянні з іншими, а нейронна мережа має найнижчу помилку перевірки, що робить його одним з кращих алгоритмів для прогнозування.

Аналіз іншого дослідження (Elias, N., & Singh, S., 2018) показав, що було проведено дослідження щодо прогнозування продажів Walmart з використанням алгоритмів машинного навчання. Ключ до цього дослідження було зроблено шляхом реалізації декількох різних алгоритмів класифікації в даних про продаж з усіх магазинів Walmart по всій території Сполучених Штатів. Проблема, виділена у цьому дослідженні, полягала у створенні конкурентного порівняльного аналізу, щоб знайти найкращий алгоритм. Дослідник вибрав 3 різних алгоритми для порівняння та протестував їх, використовуючи оцінку MAE R<sup>2</sup> Score. Методами, які використовувалися для цього дослідження, були Випадковий ліс та Підвищення градієнта. Результати цього дослідження показують, що Випадковий ліс є найкращим алгоритмом, який набрав мінімальну кількість балів в оцінці MAE (1979,4) та високий бал R<sup>2</sup> (0,94), який показав високу точність порівняно з іншими.

## 2. Основні результати дослідження

Reckitt Benckiser, одна з найбільших світових корпорацій, виробляє товари для дому, фармацевтичні засоби, предмети гігієни та здоров'я. На українському ринку споживачів Reckitt Benckiser активно використовує цифрові маркетингові технології, зокрема розміщує рекламний контент на веб-сайтах соціальних мереж, для збільшення електронних продажів і повернення нових клієнтів. Зібрана за допомогою рекламних заходів інформація може виявитися надзвичайно корисною для бізнесу, допомагаючи розуміти поведінку споживачів та визначити оптимальні маркетингові стратегії.

Важливо зрозуміти, що маркетинг та машинне навчання взаємодоповнюють одне одного. За допомогою машинного навчання аналізуються рентабельність маркетингових витрат, а маркетинг може застосовувати стратегії машинного навчання для оптимізації електронної комерції. Розрахунок впливу маркетингових комунікацій на продажі та адаптація

маркетингових пропозицій – дві універсальні проблеми маркетингу, на які можна легко відповісти, використовуючи методології машинного навчання.

Ефективність рекламної кампанії клієнта в мережі Instagram може бути порівняна з динамікою вторинних продажів продукції Reckitt Benckiser на сайті клієнта за допомогою аналітичних інструментів машинного навчання, таких як регресійний аналіз.

Промо-активність відбувається наступним чином: на сторінці клієнта в соцмережі Instagram розміщується інформація про хвилю розпродажів певного асортименту бренду. Публікація містить посилання на інтернет-магазин мережі, де користувачі можуть переглянути всі доступні акційні товари та зробити покупки.

Регресійна модель може оцінити кожен показник окремо з високим ступенем точності, навіть якщо існує синергетичний вплив між набором показників, таких як спеціальні акції та реклама. Тому рекомендується використовувати регресійну модель, якщо компаніям необхідно знати ефективність кожного маркетингового каналу та формату маркетингу для оптимізації маркетингового бюджету.

У дослідженні було використано симульовану базу даних (табл. 1), де приведено щоденні дані про кількість відвідувань сайту та подальших замовлень на акційні товари.

В першу чергу, важливо встановити, чи існує зв'язок між конверсіями з посилань у публікаціях на веб-сайті та кількістю замовлень. Для аналізу цієї статистики дуже корисним інструментом є регресійний аналіз, розроблений для програмного середовища R-Studio.

У регресійній моделі використовуються наступні параметри і змінні (Черняк О. І., 2009):

- Невідомі параметри, позначені як  $\beta$  і які представляють собою скаляри або вектор.
- Незалежні змінні,  $X$ .
- Залежна змінна,  $Y$ .

Регресійна модель співвідносить  $Y$  до функції по  $X$  та  $\beta$  :

$$Y = f(X, \beta)$$

В даному випадку залежною змінною  $Y$  буде кількість продажів, незалежною  $X$  – кількість відвідувань сайту інтернет-магазину.

Перед аналізом статистичних даних необхідно перевірити їх на наявність гетероскедастичності, автокореляції та нормальності залишкового розподілу, серед інших критеріїв. Для виявлення гетероскедастичності був використаний метод Глейзера, який базується на регресії абсолютних значень залишків, що відповідає

Таблиця 1

**Кількість відвідувачів сайту інтернет-магазину в січні 2024 року – статистичні дані**

Дата	Кількість переходів на сайт	Кількість купівель	Дата	Кількість переходів на сайт	Кількість купівель
01.01.2024	554	70	17.01.2024	694	102
02.01.2024	630	75	18.01.2024	722	77
03.01.2024	674	85	19.01.2024	736	82
04.01.2024	667	77	20.01.2024	607	75
05.01.2024	690	72	21.01.2024	680	55
06.01.2024	572	63	22.01.2024	654	82
07.01.2024	354	16	23.01.2024	675	70
08.01.2024	479	20	24.01.2024	727	97
09.01.2024	633	59	25.01.2024	623	67
10.01.2024	605	67	26.01.2024	729	112
11.01.2024	1059	107	27.01.2024	655	102
12.01.2024	1374	115	28.01.2024	626	75
13.01.2024	768	72	29.01.2024	670	62
14.01.2024	793	105	30.01.2024	610	50
15.01.2024	688	92	31.01.2024	674	67
16.01.2024	652	95	Сума	21274	2365

регресії найменших квадратів як певній функції від (Черняк О. І., Комашко О. В., Ставицький А. В., & Баженова О. В., 2009). Результати цієї перевірки представлені на рис. 1.

Оцінка коефіцієнта детермінації  $R^2$  визначає придатність моделі за допомогою цього методу. Оскільки  $R^2$  є малим та не близьким до одиниці ( $R^2 = 0.0099$ ), модель не є адекватною, а отже гетероскедастичність даних відсутня.

Коли функція корелює з собою після незначного зміщення незалежної змінної, цей явища відоме як автокореляція або автокореляційна функція. Це може виявити закономірності, такі як періодичність, і важливо для аналізу набору даних. Для цього аналізу ми використали підхід Дарбіна-Ватсона, який показаний на рис. 2.

Автокореляція за даним методом присутня тоді, коли  $p$ -value менше 0,05. В даному

```
Call:
lm(formula = gl)

Residuals:
    Min       1Q   Median       3Q      Max
-12.940  -8.112  -3.798   10.210   20.252

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.345528   9.940747   1.644   0.122
visits      -0.004478   0.011974  -0.374   0.714

Residual standard error: 10.88 on 14 degrees of freedom
Multiple R-squared:  0.009892, Adjusted R-squared:  -0.06083
F-statistic: 0.1399 on 1 and 14 DF, p-value: 0.714
```

**Рис. 1. Перевірка на гетероскедастичність**

```
Durbin-Watson test

data: regression_model
DW = 0.76174, p-value = 0.001454
alternative hypothesis: true autocorrelation is greater than 0
```

**Рис. 2. Перевірка даних на автокореляцію**

випадку  $p\text{-value} = 0.0015$ , отже автокореляція присутня.

В якісній моделі часових рядів важливо, щоб у залишках був присутній незалежний набір без систематичних компонентів і лише з випадковим шумом. Це означає, що автокореляційна функція залишків не повинна демонструвати жодної періодичності. Тому необхідно проводити ретельний аналіз залишків, оскільки вони є важливою складовою для визначення адекватності моделі. Недоліки в якості моделі можуть виникнути, якщо залишки мають періодичну складову або постійно розсіюються, наприклад, якщо вони виявляються від'ємними в першій половині ряду та близькими до нуля в другій. Важливо відзначити, що в аналізі часових рядів ретельний аналіз залишків є ключовим. Процес оцінки моделі передбачає, що залишки мають нормальний розподіл і не корелюють між собою (Колесников, С. А., Яковлева, І. І., 2021).

Для визначення того, чи є розподіл залишків нормальним, одним із найкращих методів є використання критерію Шапіро-Вілка. Цей критерій дозволяє перевірити гіпотезу про те, що випадкова величина розподілена нормально, вивчивши дані з таблиці. Форма нормального розподілу залишків представлена на рис. 3:

Shapiro-wilk normality test

```
data: residuals
W = 0.95027, p-value = 0.6717
```

**Рис. 3. Перевірка даних на нормальність розподілу залишків**

Оцінюється результат тесту також за  $p\text{-value}$ . Оскільки  $p\text{-value}$  більше 0,05, тому залишки даної моделі нормально розподілені.

```
Call:
lm(formula = Purchases ~ Website_Visits, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-36.709 -10.066  -1.610   9.318  31.679

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.11871   14.73966   0.822   0.418
Website_Visits  0.09445    0.01836   5.144 1.7e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.01 on 29 degrees of freedom
Multiple R-squared:  0.4771,    Adjusted R-squared:  0.4591
F-statistic: 26.46 on 1 and 29 DF,  p-value: 1.701e-05
```

**Рис. 4. Результат моделі регресійного аналізу**

Після перевірки даних за допомогою вищезазначених статистичних стандартів ми можемо переходити до регресійного аналізу бази даних. Для проведення цього аналізу використовувався метод лінійної регресії (рис. 4).

Моделю є адекватною, тому що коефіцієнт детермінації наближається до одиниці –  $R^2 = 0.4771$ , а  $p\text{-value} < 0.05$ . Як наслідок, існує дуже високий рівень залежності між кількістю покупок і кількістю відвідувань веб-сайту для інтернет-магазину.

Приймаючи лінійний характер залежності даних двох статистичних показників, коефіцієнт  $b$  в даній регресії характеризує метрику інтернет-маркетингу – конверсію продажів. А отже конверсія сайту, згідно моделі, становить  $b = 0.2000$ , або 20%.

### 3. Результати та обговорення

У даному дослідженні застосовувався метод регресійного аналізу для аналізу впливу рекламних ініціатив на електронні продажі в інтернет-магазині. Таким чином, можна дійти до наступних висновків:

1. Результати дослідження підтверджують, що рекламний контент на веб-сайтах соціальних мереж має значний вплив на обсяг електронних продажів. Це підтверджує важливість ефективної маркетингової стратегії для збільшення обсягу продажів в інтернет-магазині.

2. Виявлено, що витрати на широкий спектр публікацій можуть бути неефективними, оскільки тип публікацій безпосередньо не впливає на кількість проданих товарів. Таким чином, доцільно зосередити фінансові ресурси на більш ефективні маркетингові канали, які приносять найбільший прибуток.

3. Раніше проведені дослідження показують, що ефективність рекламних кампаній

у соціальних мережах залежить від цільової аудиторії та типу продукту чи послуги.

Отже, для успішної оптимізації маркетингових ініціатив та максимізації електронних продажів необхідно враховувати отримані результати дослідження та належним чином розподіляти фінансові ресурси для досягнення максимального ефекту.

**Висновок.** Дослідження з використання машинного навчання та аналізу даних підтверджує велике значення цих технологій у вирішенні глобальних викликів, з якими стикається сучасне суспільство. Застосування машинного навчання та аналізу даних допомагає вирішувати складні проблеми, здійснювати прогнози та розробляти стратегії для сталого розвитку. За допомогою машинного навчання можна здійснювати аналіз великих обсягів даних, виявляти патерни та залежності, що дозволяє зробити більш обґрунтовані рішення у різних галузях, таких як медицина, екологія, економіка тощо. Машинне навчання допомагає прогнозувати майбутні тенденції та патерни на основі історичних даних. Це дозволяє зробити передбачення щодо змін у різних сферах, таких як зміна клімату, зростання попиту на ресурси, зміни ринкових умов та інше. Застосування машинного навчання та аналізу даних допомагає вирішувати важливі глобальні проблеми, такі як забруднення довкілля, зміна клімату, енергетична криза, медичні проблеми, технологічні інновації тощо. Застосування машинного навчання у медицині допомагає покращити діагностику та лікування різних захворювань, що збільшує шанси на одужання та покращує якість життя пацієнтів. Машинне

навчання є ключовим фактором для розвитку нових технологій і інновацій у різних сферах. Воно сприяє створенню автономних систем, інтелектуальних роботів, автоматизованих процесів та багато іншого.

Таким чином, результати дослідження щодо використання машинного навчання для прогнозування бізнес-рішень електронної комерції показали, що існує безліч різних методів машинного навчання, які можна використовувати при прогнозуванні бізнес-рішень у майбутньому. Для вибору найбільш відповідного методу машинного навчання для цілей поставленої задачі доцільно провести попередньо тестовані обраних альтернатив на конкретних наборах вихідних даних з метою визначення переважно варіантів рішень. Як рішення вибирається модель, що має найкращий діапазон передбачення, де прогнозоване значення та фактичне значення майже збігаються. Потім інтегрувати обраний метод у розробку рекомендаційної системи. Результати цього дослідження на одному конкретному наборі вихідних даних показали перевагу Випадкового лісу над Підвищенням градієнта та моделі SARIMA над ARIMA. Однак перевага була незначна, що говорить про можливість застосування всіх розглянутих методів.

Усі ці аспекти підтверджують важливість цих технологій у вирішенні глобальних викликів та покращенні якості життя нашого суспільства. Продовження досліджень та інновацій у використанні машинного навчання та аналізу даних є ключовими для досягнення успіхів у вирішенні складних проблем, що стоять перед сучасним світом.

#### ЛІТЕРАТУРА:

1. Mitchell T. M. Machine Learning McGraw-Hill International. – 1997. ISBN: 0070428077.
2. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning: data mining, inference, and prediction. – New York : springer, 2009. Т. 2. С. 1–758. ISBN: 978-0387848570.
3. Goodfellow I., Bengio Y., Courville A. Deep learning. – MIT press, 2016. ISBN: 978-0262035613.
4. Bishop C. M. Pattern recognition and machine learning //Springer google schola. 2006. Т. 2. С. 645–678. ISBN: 978-0387310732.
5. Zhao K., Wang C. Sales forecast in e-commerce using convolutional neural network //arXiv preprint arXiv:1708.07946. 2017. Access mode: <https://doi.org/10.48550/arXiv.1708.07946>.
6. Bandara K., Shi P., Bergmeir C., Hewamalage H., Tran Q., Seaman B. Sales demand forecast in e-commerce using a long short-term memory neural network methodology //Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part III 26. – Springer International Publishing, 2019. С. 462–474. Access mode: <https://doi.org/10.48550/arXiv.1901.04028>.
7. Li M., Ji S., Liu G. Forecasting of Chinese E-commerce sales: an empirical comparison of ARIMA, nonlinear autoregressive neural network, and a combined ARIMA-NARNN model //Mathematical Problems in Engineering. 2018. Т. 2018. С. 1–12. Access mode: <https://doi.org/10.1155/2018/6924960>.
8. Pavlyshenko B. M. Machine-learning models for sales time series forecasting // Data. 2019. Т. 4. №. 1. С. 15. Access mode: <https://doi.org/10.3390/data4010015>.



9. Elias N., Singh S. Forecasting of Walmart sales using machine learning algorithms //Research paper, Dept. of Electronics & Comm. Engineering, BMS Inst. of Technology & Management, Bangalore, India. – 2018. Access mode: <https://medium.com/@auggieheschmeyer/forecasting-walmart-sales-using-machinelearning-models-3bf38f6c533>
10. VanderPlas J. Python data science handbook: Essential tools for working with data. – « O'Reilly Media, Inc.», 2016. ISBN: 978-1491912058.
11. Жук О. Ю., Гольченко О. О. Методи прогнозування та аналізу даних. Київ: Видавничий Дім «Слово/Книга». 2018.
12. Колесников С. А., Чередніченко В. В., Яковлева І. І. Прогнозування та аналіз даних. Київ: Видавничий центр КНЕУ. 2018.
13. Колесников С. А., Яковлева І. І. Машинне навчання та аналіз даних. Київ: Видавничий центр КНЕУ. 2021.
14. Мартиненко В. Г., Кадук О. В. Прогнозування та технічний аналіз на фінансових ринках. Київ: Видавничий центр КНЕУ. 2019.
15. Черняк О. І., Комашко О. В., Ставицький А. В., Баженова О. В. Економетрика: підручник. Київ: Видавничо-поліграфічний центр «Київський національний університет імені Тараса Шевченка». 2009.

#### REFERENCES:

1. Mitchell, T. (1997). Machine Learning. McGraw-Hill Education. ISBN: 0070428077.
2. Hastie, T., Tibshirani, R. & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer. ISBN: 978-0387848570.
3. Goodfellow, I., Bengio, Y. & Courville, A. (2016). Deep Learning. MIT Press. ISBN: 978-0262035613.
4. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer. ISBN: 978-0387310732.
5. Zhao, K. & Wang, C. (2017). Sales forecast in e-commerce using convolutional neural network. 8 p. Access mode: <https://doi.org/10.48550/arXiv.1708.07946>.
6. Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q. & Seaman, B. (2019). Sales demand forecast in e-commerce using a long short-term memory neural network methodology. In Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part III 26 (pp. 462-474). Springer International Publishing. Access mode: <https://doi.org/10.48550/arXiv.1901.04028>.
7. Li, M., Ji, S., & Liu, G. (2018). Forecasting of Chinese E-commerce sales: an empirical comparison of ARIMA, nonlinear autoregressive neural network, and a combined ARIMA-NARNN model. Mathematical Problems in Engineering, 2018, 1–12. Access mode: <https://doi.org/10.1155/2018/6924960>.
8. Pavlyshenko, B. M. (2019). Machine-learning models for sales time series forecasting. Data, 4(1), 15. Access mode: <https://doi.org/10.3390/data4010015>.
9. Elias, N. & Singh, S. (2018). Forecasting of Walmart sales using machine learning algorithms. Research paper, Dept. of Electronics & Comm. Engineering, BMS Inst. of Technology & Management, Bangalore, India. Access mode: <https://medium.com/@auggieheschmeyer/forecasting-walmart-sales-using-machinelearning-models-3bf38f6c533>
10. VanderPlas, J. (2016). Python Data Science Handbook. O'Reilly Media. ISBN: 978-1491912058.
11. Zhuk, O.Y. & Gol'chenko, O.O. (2018). Metody prohnozuvannya ta analizu danykh [Methods of forecasting and data analysis]. Kyiv: Publishing House «Word/Book» [in Ukrainian].
12. Kolesnikov, S. A., Cherednichenko, V. V. & Yakovleva, I. I. (2018). Prohnozuvannya ta analiz danykh [Forecasting and data analysis]. Kyiv: Publishing Center of KNEU [in Ukrainian].
13. Kolesnikov, S.A. & Yakovleva, I.I. (2021). Mashynne navchannya ta analiz danykh [Machine learning and data analysis]. Kyiv: Publishing Center of KNEU [in Ukrainian].
14. Martynenko, V.G. & Kaduk, O.V. (2019). Prohnozuvannya ta tekhnichnyy analiz na finansovykh rynkakh [Forecasting and technical analysis on financial markets]. Kyiv: Publishing Center of KNEU [in Ukrainian].
15. Chernyak, O.I., Komashko, O.V., Stavitsky, A.V. & Bazhenova, O.V. (2009). Ekonometryka: pidruchnyk [Econometrics: textbook]. Kyiv: Publishing and Printing Center «Kyiv National University named after Taras Shevchenko» [in Ukrainian].