



Звіт про оригінальність

● Оцінка схожості

% 6

● Ризик плагіату

СЕРЕДНІЙ

👤 Olga Kagalo 🕒 2025-06-19 23:04

Посилання на звіт: 10mDc / Посилання користувача: qEAc



Ось вона – Ваша звіт про оригінальність!

Ми раді повідомити, що перевірка вашого документа завершена, і результати вже готові! Наші алгоритми старанно працювали, щоб знайти збіги в наших базах даних.

На наступних сторінках ви знайдете результати перевірки:

Бали

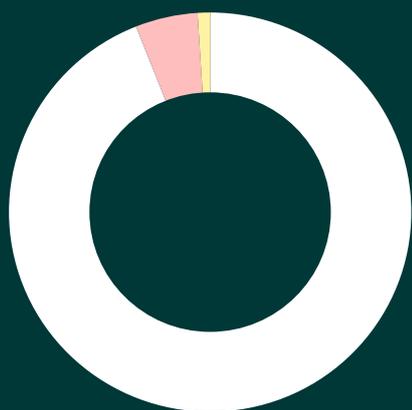
Збіги

Посилання

Ваш документ було перевірено за такими джерелами:

- База даних інтернет-джерел
- База даних наукових статей
- Глибока перевірка (наш вдосконалений алгоритм)

Бали



● Збіги тексту	5%
● Перефразування	1%
● Цитований текст	0%
● Неправильне цитування	0%
● Збігів не знайдено	94%

Ризик плагіату

СЕРЕДНІЙ

Ризик плагіату вказує, як збіги тексту розподілені по документу. Вищий ризик виникає, коли збіги з'являються близько один до одного, наприклад, у тому самому абзаці або розділі.

Оцінка схожості

% **6**

Оцінка схожості показує, скільки слів або символів у вашому документі збігаються з текстами інших документів, включаючи перефразовані тексти або неправильні цитати.

Збіги

1 Теоретичні основи прогнозування продажів засобами машинного навчання

1.1 Поняття прогнозування продажів та його значення для бізнесу

Прогнозування — це процес оцінки майбутніх подій на основі аналізу історичних даних і поточних тенденцій. У контексті підприємницької діяльності прогнозування є критично важливим інструментом для прийняття стратегічних і тактичних рішень. Однією з найважливіших сфер застосування прогнозування в бізнесі є прогнозування обсягів продажів — ключового показника, від якого залежать прибутковість, ефективність виробництва, планування запасів, логістика, маркетинг і фінансова політика підприємства.

Продаж — це кінцева мета будь-якої комерційної діяльності. Від його обсягів залежить, з одного боку, дохід компанії, а з іншого — необхідність ресурсів, закупівель, виробничих потужностей, складів, транспортних витрат та ін. Тому прогнозування продажів — це основа для ефективного управління всією операційною системою підприємства.

Прогнозування продажів — це процес формування припущення щодо майбутнього обсягу реалізації товарів або послуг **15** на основі аналізу історичних даних, сезонних трендів, ринкових змін та інших факторів. Це не просто передбачення, а раціонально обґрунтоване рішення, яке базується на наукових методах: статистичних, аналітичних, економетричних, а останнім часом — і на основі алгоритмів машинного навчання.

Прогнозування продажів виконує кілька ключових функцій:

Фінансове планування — дає змогу розрахувати доходи, витрати, податки, грошові потоки.

Планування виробництва — визначає необхідні обсяги виготовлення товару.

Управління запасами — допомагає уникнути як дефіциту, так і надлишкових складських залишків.

Оптимізація логістики — дозволяє краще планувати маршрути, поставки, пакування.

Маркетингова стратегія — дає змогу коригувати кампанії залежно від очікуваного попиту.

У бізнесі використовуються різні види прогнозів залежно від завдань і горизонту планування:

Короткострокове прогнозування (від кількох днів до 3 місяців) — використовується для щоденних операцій, управління складом, оперативного планування закупівель.

Середньострокове прогнозування (від 3 до 12 місяців) — необхідне для сезонного планування, управління ресурсами, маркетингової активності.

Довгострокове прогнозування (1–5 років і більше) — потрібне для стратегічного розвитку, відкриття нових ринків, розширення виробництва.

Методи прогнозування умовно поділяються на:

Якісні (експертні) методи:

метод експертних оцінок;

метод сценарного аналізу;

метод Дельфі;

анкетування споживачів;

аналіз фокус-груп.

Кількісні (математичні) методи:

трендовий аналіз;

метод **19** ковзної середньої;

19 експоненційне згладжування;

19 регресійний аналіз;

методи прогнозування часових рядів;

алгоритми машинного навчання.

Раніше більшість підприємств поклалися на статистичні підходи, але з розвитком

цифрових технологій з'явилися нові інструменти — Data Science і машинне навчання, які дозволяють побудувати складні моделі з високою точністю.

Недооцінка або переоцінка обсягів продажів може мати серйозні наслідки:

Надлишковий запас призводить до додаткових витрат на зберігання, втрати від псування товару.

Дефіцит товару — до втрати продажів і репутації, незадоволених клієнтів.

Неправильне планування персоналу — або надлишок робочої сили, або її нестача в пікові періоди.

Фінансові втрати — через нераціональне використання ресурсів, закупівель або реклами.

При недостатньому прогнозі компанія втрачає можливість реалізації; при надмірному — несе витрати на зберігання. Рисунок 1.1 демонструє вплив точності прогнозу на ефективність управління запасами.

Рисунок 1.1 – Вплив точності прогнозу на ефективність управління запасами

У реальному житті дані продажів схильні до:

Сезонних коливань (зростання на свята, зниження влітку тощо);

Циклічних змін (залежно від економічних фаз);

Довгострокових трендів (наприклад, перехід на онлайн-торгівлю).

Успішне прогнозування повинно враховувати всі ці аспекти. Машинне навчання дозволяє моделювати складні взаємозв'язки між факторами, включаючи приховані шаблони.

На практиці прогнозування продажів може бути реалізоване як:

Внутрішній аналітичний відділ, який використовує Excel, Power BI, або Python.

Інтегровані CRM/ERP-системи, де вбудовані алгоритми прогнозування.

Спеціалізовані веб-застосунки, що працюють із базами даних, API, інтерактивною візуалізацією.

Розробка власного веб-застосунку дозволяє адаптувати систему під конкретні потреби підприємства, що є перевагою над готовими рішеннями.

Інтеграція системи прогнозування дозволяє керівництву:

точно визначити оптимальний асортимент;

змінювати цінову політику залежно від очікуваного попиту;

зменшувати обсяги закупівель у низький сезон;

збільшувати виробництво в пік продажів;

визначати, коли і скільки запускати рекламу.

Це безпосередньо впливає на прибутковість та конкурентоспроможність підприємства.

Попри велику кількість методів, прогнозування супроводжується певними складнощами:

Наявність "шуму" в даних;

Відсутність повних історичних даних;

Різкі зовнішні впливи (економічні кризи, війна, пандемія);

Зміни в поведінці споживачів;

Проблеми з точністю моделей при нестабільному попиті.

Саме тому застосування гнучких алгоритмів машинного навчання у веб-застосунку є сучасною відповіддю на ці виклики.

1.2 Методи прогнозування: класичні та сучасні підходи

Прогнозування продажів — це міждисциплінарна задача, яка поєднує елементи статистики, економіки, управління, інформатики та математичного аналізу. Методи, які застосовуються для вирішення таких задач, поділяються на:

Класичні (статистичні) методи — базуються на аналізі часових рядів, регресії, ковзних середніх, експоненційному згладжуванні тощо.

Сучасні (обчислювальні, інтелектуальні) методи — включають алгоритми машинного навчання, нейронні мережі, ансамблеві моделі, глибоке навчання.

Кожна група **7** методів має свої переваги та обмеження, і вибір методу залежить від специфіки даних, цілей прогнозування та технічних ресурсів підприємства.

Класичні методи прогнозування залишаються актуальними в багатьох сферах завдяки своїй простоті, пояснюваності результатів та ефективності при малих обсягах даних.

Метод ковзної середньої (Moving Average, MA). Цей метод ґрунтується на усередненні попередніх значень для згладжування коливань і виявлення трендів. Формула простої ковзної середньої:

$$\hat{y}_t = (y_{t-1} + y_{t-2} + \dots + y_{t-n}) / n \quad (1.1)$$

\hat{y}_t — прогнозоване значення на момент часу t ,

n — довжина вікна усереднення.

Переваги:

Простота реалізації.

Добре працює для згладжених даних.

Недоліки:

Не враховує сезонність.

Чутливість до різких змін у тренді.

Метод експоненціального згладжування (Exponential Smoothing). На відміну від MA, цей метод надає більшу вагу останнім значенням:

$$\hat{y}_t = \alpha y_{t-1} + (1 - \alpha) \hat{y}_{t-1} \quad (1.2)$$

α — коефіцієнт згладжування ($0 < \alpha < 1$).

Переваги:

Краще реагує на нові зміни.

Можна розширити до моделей Holt та Holt-Winters для тренду і сезонності.

Регресійний аналіз дозволяє знайти функціональну залежність між змінними. Наприклад, зв'язок між кількістю продажів і ціною, рекламною активністю чи погодними умовами.

$$y = \beta_0 + \beta_1 x + \epsilon \quad (1.3)$$

Переваги:

Інтерпретованість моделей.

Можна враховувати декілька факторів.

Недоліки:

Не враховує складних нелінійних залежностей.

Сучасні підходи базуються на алгоритмах машинного навчання (ML), які **6** здатні автоматично виявляти складні закономірності в даних.

6 Лінійна регресія (ML-реалізація). Попри свою простоту, лінійна регресія є однією з основних моделей у ML і використовується як базовий еталон для порівняння.

Переваги:

Висока швидкість навчання.

Інтерпретованість коефіцієнтів.

Недоліки:

Погано працює при сильній нелінійності чи кореляції змінних.

Дерева рішень (Decision Trees) – це ієрархічна структура, яка розбиває дані на підмножини за ознаками.

Переваги:

Не потребує масштабування даних.

Інтуїтивно зрозуміла модель.

Недоліки:

Схильність до перенавчання (overfitting).

Random Forest – це ансамбль із **6** багатьох дерев рішень. Кожне дерево працює на випадковому підмножині даних, а результат — середнє або голосування.

Переваги:

Висока точність.

Стійкість до перенавчання.

XGBoost. Один із найпотужніших алгоритмів градієнтного бустингу. **8** Активно використовується в змаганнях Kaggle та промислових системах.

Переваги:

Висока швидкість.

Потужна обробка аномалій і відсутніх значень.

Нейронні мережі — це потужні математичні моделі, що імітують принципи **8** роботи людського мозку. Вони здатні навчатися з даних, виявляючи складні, багатовимірні, нелінійні залежності, які часто не вдається описати традиційними статистичними методами. Завдяки своїй гнучкості, ці моделі ефективно застосовуються в задачах прогнозування, класифікації, розпізнавання образів, обробки природної мови та багатьох інших сферах.

Серед різновидів нейронних мереж, що застосовуються для прогнозування, виділяють кілька основних архітектур:

MLP (багатошарові перцептрони) — загальні прогностичні мережі.

RNN/LSTM/GRU — спеціалізовані для обробки часових рядів.

Модель приймає на вхід вектор історичних продажів, обробляє в кількох шарах і дає прогнозне значення (рисунок 1.2).

Рисунок 1.2 – Схема багатошарової нейронної мережі

Порівняльний аналіз методів прогнозування дає змогу обґрунтовано вибрати найбільш доцільний підхід залежно від наявних даних, цілей бізнесу та технічних можливостей. У таблиці 1.1 наведено основні характеристики найбільш уживаних методів.

Таблиця 1.1 – Порівняльний аналіз методів

Метод

Переваги

Недоліки

Коли застосовувати

MA / EMA

Прості, швидкі

Ігнорують контекст

Початковий аналіз

Регресія

Пояснюваність

Погано з нелінійністю

Коли мало змінних

Decision Tree

Інтуїтивні

Overfitting

Простий поділ сегментів

Random Forest

Стабільні

Важче інтерпретувати

Для структурованих даних

XGBoost

Точні, гнучкі

Складність налаштування

Промислові проекти

LSTM

Враховує порядок

Більше даних потрібно

Часові ряди з сезонністю

Проблеми та виклики при виборі методу:

Наявність та якість даних — багато алгоритмів не працюють при значній кількості

пропущених значень або шуму.

Обчислювальні ресурси — складні моделі вимагають потужного обладнання.

Тлумачення результатів — деякі моделі (нейронні мережі) складно пояснити керівництву.

Масштабованість — важливо, щоб модель легко інтегрувалась у веб-застосунок.

У межах цієї дипломної протестовані кілька 10 алгоритмів, зокрема:

10 лінійна регресія;

10 дерева рішень;

10 Random Forest;

XGBoost;

LSTM (експериментально).

На основі точності (RMSE, MAE), швидкості та зручності інтеграції буде обрано один основний метод для реалізації у веб-застосунку.

1.3 Основи машинного навчання: supervised, unsupervised, reinforcement learning

Машинне навчання (англ. Machine Learning, ML) — це ключова підгалузь штучного інтелекту (AI), яка забезпечує системам здатність автоматично навчатися з досвіду (даних) і покращувати продуктивність без явного програмування. Якщо штучний інтелект є загальним поняттям, що охоплює всі форми інтелектуального поведінки комп'ютерних систем, то машинне навчання — це механізм, за допомогою якого ці системи опановують нові навички.

У класичних IT-системах програміст визначає усі правила поведінки. У ML же основна логіка системи формується автоматично — модель виводить правила, які найкраще пояснюють залежності в 10 даних. Це дозволяє вирішувати складні задачі, де ручне програмування було б надто громіздким або взагалі неможливим.

Сучасні компанії працюють у середовищі, де обсяг даних стрімко зростає. Ручна обробка інформації стає неможливою, тому все більше підприємств впроваджують ML-рішення для:

персоналізованих рекомендацій (наприклад, Amazon, Netflix),

динамічного ціноутворення,

прогнозування попиту,

виявлення шахрайства (fraud detection),

оптимізації логістики,

автоматизації обслуговування клієнтів (чат-боти).

Прогнозування продажів — одна з найочевидніших і найпрактичніших задач, де ML демонструє переваги: швидкість, гнучкість, адаптивність до змін.

Кожна ML-система проходить схожий цикл:

Збір даних: витяг інформації з баз, API, CSV-файлів, сенсорів тощо.

Обробка (preprocessing): очищення, нормалізація, усунення аномалій.

17 Інженерія ознак (feature engineering): створення нових змінних, які краще описують дані.

Розбиття на вибірки: зазвичай 70% — навчання, 15% — перевірка, 15% — тестування.

Навчання моделі: підбір параметрів.

Оцінка: аналіз помилок, візуалізація, метрики.

Розгортання (deployment): інтеграція моделі у веб-застосунок чи інший сервіс.

Машинне навчання класифікується за способом взаємодії з даними на три основні типи: навчання з учителем, без учителя та з підкріпленням.

Навчання з учителем (Supervised Learning) – це найбільш поширений тип навчання. Його суть: модель має доступ до навчального набору, де кожен запис містить вхідні значення (ознаки) і правильну відповідь (мітку). Ключова мета: навчитися зіставляти $X \rightarrow Y$, щоб на нових даних передбачити Y .

Задачі:

Класифікація: розподіл об'єктів по категоріях (купити/не купити, "гарячий" чи "холодний" лід).

Регресія: прогноз числових значень (прибуток, кількість продажів).

Приклади:

Прогноз курсу акцій.

Виявлення спаму в електронній пошті.

Оцінка кредитоспроможності клієнтів.

Поширені алгоритми:

Лінійна і 11 логістична регресія

11 K-Nearest Neighbors (KNN)

Support Vector Machines (SVM)

11 Дерева 11 рішень і 14 Random Forest

14 XGBoost, LightGBM

14 Нейронні мережі (MLP, CNN)

14 Навчання без учителя (Unsupervised Learning). У цьому випадку немає міток. Модель намагається знайти приховані структури в даних. Такі методи дозволяють виявляти закономірності, групи схожих об'єктів або знижувати розмірність даних для подальшого аналізу.

Задачі:

Кластеризація: об'єднання подібних об'єктів (сегменти покупців).

Зменшення розмірності: PCA, t-SNE для візуалізації.

Аномалії: виявлення підозрілих транзакцій.

Застосування:

Розподіл клієнтів на групи.

Пошук схожих продуктів.

Стиснення даних.

Поширені алгоритми:

K-means

DBSCAN

PCA 11 (метод головних компонент)

11 Autoencoders (нейронні мережі)

Навчання з підкріпленням (Reinforcement Learning). Модель (агент) вивчає послідовність дій, які максимізують довгострокову винагороду. Кожна дія отримує оцінку (нагороду або штраф), і з часом модель вчиться приймати найвигідніші рішення.

Застосування:

Штучний інтелект у грі Go, шахах, StarCraft

Управління запасами

Оптимізація маршрутів доставок

Алгоритми:

Q-learning

Deep Q Networks (DQN)

Actor-Critic

Обмеження – висока складність, потреба в симуляції середовища.

Для кращого розуміння класифікації методів машинного навчання було наведено схему (рисунок 1.3), яка візуально демонструє основні підходи — з учителем, без учителя та з підкріпленням.

Рисунок 1.3 – Класифікація типів машинного навчання

У сучасній розробці ML-проектів застосовуються інструменти та бібліотеки продемонстровані в таблиці 1.2.

Таблиця 1.2 – Інструменти для машинного навчання

Інструмент

Опис

Scikit-learn

Класика ML. Регресії, класифікації, дерева.

Pandas, Numpy

Маніпуляція з даними.

TensorFlow, PyTorch

Глибоке навчання, нейромережі.

XGBoost, LightGBM

Гнучкі градієнтні бустинги.

Keras

Високорівнева бібліотека для NN.

MLflow

Відстеження експериментів.

Streamlit / Flask / FastAPI

Розгортання моделей у веб.

ML-модель завжди має бути перевірена на точність. Основні метрики продемонстровані в таблиці 1.3.

Таблиця 1.3 – Метрики оцінки точності моделей машинного навчання

Метрика

Формула

Інтерпретація

MAE (Середня абсолютна помилка)

$$MAE = (1/n) * \sum |y_i - \hat{y}_i|$$

Середнє абсолютне відхилення між прогнозом і реальними значеннями

RMSE (Корінь середньоквадратичної помилки)

$$RMSE = \sqrt{(1/n) * \sum (y_i - \hat{y}_i)^2}$$

Більше штрафує великі помилки, чутливий до викидів

R² (Коефіцієнт детермінації)

$$R^2 = 1 - (\sum(y_i - \hat{y}_i)^2 / \sum(y_i - \bar{y})^2)$$

Відображає частку поясненої варіації в залежній змінній

MAPE (Середня абсолютна відносна помилка)

$$MAPE = (100\% / n) * \sum |(y_i - \hat{y}_i) / y_i|$$

Помилка у відсотках; погано працює з нульовими значеннями

Приклад оцінки на Python:

```
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score  
  
mae = mean_absolute_error(y_true, y_pred)  
  
rmse = mean_squared_error(y_true, y_pred, squared=False)  
  
r2 = r2_score(y_true, y_pred)
```

Проблеми та виклики машинного навчання:

Overfitting (перенавчання): модель "зазубрює" дані, але погано працює на нових.

Underfitting: модель надто спрощена, не бачить закономірностей.

Пропущені дані, шум.

Неетичне використання: дискримінація, порушення приватності.

Тлумачення: деякі моделі, як нейромережі, "чорні скриньки".

Перспективи розвитку ML:

AutoML: автоматичний підбір моделей.

Explainable AI (XAI): прозорість рішень моделей.

Federated Learning: навчання без передачі даних.

ML на краю (Edge AI): запуск моделей на смартфонах, IoT.

1.4 Огляд алгоритмів, що застосовуються для прогнозування

4 У сучасному світі дані стали однією з найцінніших ресурсів, а здатність витягати з них корисну інформацію перетворилася на ключову конкурентну перевагу. В межах завдань прогнозування продажів головною метою є побудова такої моделі, яка зможе з

високою точністю передбачити майбутній обсяг реалізації товару або послуги. Для цього **1** використовуються різноманітні алгоритми машинного навчання, кожен з яких має свої особливості, переваги, недоліки та сфери застосування.

Вибір алгоритму залежить від багатьох факторів: кількості та якості наявних даних, складності залежностей між змінними, бажаної точності прогнозу, обчислювальних ресурсів, а також вимог до швидкості навчання й інтерпретованості моделі. **1** У цьому розділі буде розглянуто найпоширеніші алгоритми, що довели свою ефективність у задачах прогнозування, зокрема у прогнозуванні продажів.

Лінійна регресія є одним із найпростіших і найпоширеніших методів прогнозування. Вона передбачає, що залежність між незалежними змінними (ознаками) та залежною змінною (прогнозованим значенням) є лінійною. Іншими словами, кожна незалежна змінна вносить лінійний внесок у результат.

Модель має вигляд:

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n + \epsilon(1.4)$$

де y — цільова змінна (обсяг продажів), x_1, x_2, \dots, x_n — ознаки, β — коефіцієнти моделі, ϵ — випадкова помилка.

Основна перевага лінійної регресії — простота реалізації та інтерпретації. Вона дозволяє точно зрозуміти, як кожна змінна впливає на прогноз. Водночас вона погано справляється зі складними, нелінійними залежностями, що часто зустрічаються в реальних бізнес-даних. У таких випадках ефективнішими є складніші алгоритми.

Дерева рішень є більш гнучкими порівняно з лінійною регресією. Вони створюють ієрархічну структуру, в якій дані розділяються за умовами на кожному рівні. Алгоритм ітеративно визначає, яка ознака і яке значення цієї ознаки найкраще поділяє дані на групи з мінімальною різницею в межах груп і максимальною різницею між групами.

Модель є легко інтерпретованою і працює добре при наявності як числових, так і категоріальних змінних. Проте дерева рішень мають серйозний недолік — схильність до перенавчання. Якщо дерево не обмежити по глибині чи кількості вузлів, воно буде «зазубрювати» навчальні дані, втрачаючи здатність до узагальнення.

Random Forest — це ансамблевий метод, який будує велику **9** кількість дерев рішень і **9** об'єднує їх прогнози. Для **9** кожного дерева вибирається випадкова підмножина ознак і випадкова вибірка даних із заміною. Результат — середній (у випадку регресії) або найбільш популярний (у випадку класифікації) прогноз усіх дерев.

Цей підхід дозволяє значно зменшити ризик перенавчання і забезпечує високу

точність. Random Forest добре працює навіть із "брудними" даними, має вбудовані механізми оцінки важливості ознак і здатен працювати з великою кількістю змінних. Основним недоліком є втрата прозорості — пояснити, чому модель зробила певний прогноз, значно складніше.

XGBoost — один із найефективніших сучасних алгоритмів. Це реалізація методу бустингу, в якому дерева рішень будуються послідовно, і кожне нове дерево коригує помилки попереднього. XGBoost використовує градієнтне зменшення функції втрат, регуляризацію і спеціальні методи для уникнення перенавчання.

Переваги XGBoost:

Висока точність прогнозу;

Гнучкість налаштувань;

Швидкість завдяки оптимізації багатоядерності;

Підтримка відсутніх значень, вагових коефіцієнтів.

Недолік — складність у налаштуванні параметрів, потреба в глибокому розумінні процесу для отримання максимального результату.

XGBoost сьогодні є стандартом де-факто в задачах прогнозування, особливо там, де якість прогнозу критично важлива, а структура даних складна.

LSTM — це тип рекурентної нейронної мережі (RNN), що здатна працювати з часовими рядами, зберігаючи залежності на довгих відрізках часу. На відміну від звичайних, LSTM враховує порядок подій і запам'ятовує значення в часі.

Цей підхід є особливо ефективним для прогнозування послідовностей, таких як продажі за днями або тижнями. LSTM дозволяє моделювати сезонність, тренди, вплив попередніх значень на поточний прогноз. Проте модель вимагає значної кількості даних, довгого часу навчання та складної архітектури.

Окрім наведених, варто згадати ще кілька алгоритмів, **4** що можуть бути корисними в прогнозуванні:

KNN (k-найближчих сусідів): проста модель, яка прогнозує на основі схожих прикладів з історії. Не потребує навчання, але погано масштабується.

Support Vector Regression (SVR): побудована на основі гіперплощин, має добру узагальнюючу здатність, але складна для великого числа ознак.

ARIMA/Prophet: класичні статистичні **1** моделі для аналізу часових рядів, які залишаються актуальними для простих, добре структурованих даних.

Порівняльний аналіз алгоритмів продемонстрований в таблиці 1.4.

Таблиця 1.4 – Порівняльний аналіз алгоритмів

Алгоритм

Точність

Інтерпретованість

Стійкість до шуму

Потреби в даних

Лінійна регресія

Середня

Висока

Низька

Низькі

Decision Tree

Середня

Висока

Низька

Середні

Продовження таблиці 1.4 – Порівняльний аналіз алгоритмів

Random Forest

Висока

Низька

Висока

Середні

XGBoost

Висока+

Низька

Висока

Середні

LSTM

Висока+

Низька

Середня

Високі

У межах цього дослідження передбачається побудова моделі прогнозування продажів на основі реальних або змодельованих даних підприємства. Особливістю є наявність часової залежності, що вимагає врахування динаміки попередніх значень. Після попереднього тестування кількох алгоритмів буде обрано той, який забезпечить найкращий баланс точності, швидкості роботи і зручності реалізації у веб-застосунку.

На етапі прототипування будуть використані такі підходи:

Лінійна регресія (базова модель),

Random Forest (еталонна ансамблева модель),

XGBoost (основна модель для глибокого тестування),

LSTM (експериментальна для розширеної версії веб-застосунку).

2 Аналіз **1** предметної області та постановка задачі

2.1 Характеристика умовного підприємства

У цій роботі розглядається умовне підприємство, яке здійснює діяльність у сфері роздрібної торгівлі товарами широкого вжитку. Обрання умовного об'єкта дозволяє створити універсальні методичні рекомендації та розробити рішення, які можуть бути застосовані в різних компаніях з подібними бізнес-процесами, незалежно від галузі чи

регіональних особливостей. Такий підхід особливо корисний для формування загальних принципів побудови систем прогнозування продажів із використанням **1** алгоритмів машинного навчання.

1 Умовне підприємство має розгалужену мережу реалізації продукції, що включає як фізичні торгові точки, так і онлайн-канали продажу. Відсутність конкретної назви і детальної організаційної структури дозволяє сфокусуватися на аналізі основних бізнес-процесів, які характерні для більшості подібних суб'єктів господарювання.

Основною метою діяльності підприємства є максимізація обсягів продажів при оптимальному використанні ресурсів, забезпеченні високого рівня сервісу для клієнтів та ефективному управлінні товарними запасами. **16** Високий рівень конкуренції на ринку змушує компанію звертати особливу увагу на оперативне планування і точне прогнозування попиту.

Діяльність підприємства відзначається комплексністю впливу внутрішніх і зовнішніх чинників. На внутрішньому рівні це організація закупівель, управління логістикою і складськими запасами, ведення бухгалтерського і аналітичного обліку. Зовнішні фактори включають ринкові тенденції, сезонність попиту, маркетингові акції, поведінку споживачів, економічні умови, а також вплив подій соціального чи політичного характеру.

Особливістю даного підприємства є наявність великого обсягу історичних даних про продажі, що містять інформацію за різними параметрами: часовими періодами, категоріями товарів, регіонами збуту, каналами реалізації. Це дає змогу застосувати сучасні методи аналізу, які базуються на алгоритмах **1** машинного навчання, для підвищення точності прогнозування.

Прогнозування продажів є одним з ключових елементів стратегічного управління, яке дозволяє:

Передбачати обсяги попиту на різні товари в різні періоди часу;

Оптимізувати обсяги закупівель, щоб уникнути як дефіциту, так і надлишку товарів;

Планувати виробничі та логістичні процеси з урахуванням прогнозованого попиту;

Формувати маркетингові стратегії, направлені на стимулювання продажів у конкретні періоди;

Управляти фінансовими потоками і бюджетом компанії більш ефективно.

Без застосування надійних моделей прогнозування **18** підприємство ризикує втратити

значну частину прибутку через неефективне планування запасів, збільшення витрат на зберігання, а також через втрачені продажі внаслідок дефіциту товару.

Ринок, на якому функціонує підприємство, характеризується підвищеною конкуренцією і значною динамікою змін. Поведінка споживачів є складним і мінливим процесом, який залежить від безлічі факторів, таких як сезонні свята, економічна ситуація, зміни в перевагах, появи нових технологій, а також акції та знижки конкурентів.

Ці фактори створюють складний фон, на якому змінюються обсяги продажів. Тому класичні прості методи прогнозування часто виявляються недостатньо ефективними, і є необхідність у використанні більш складних моделей, які здатні враховувати багатомірні залежності та адаптуватись до змін.

Умовне підприємство використовує сучасні інформаційні системи **1** для збору, зберігання і обробки даних. Ці системи охоплюють різні напрямки:

Бази даних продажів і замовлень;

CRM-системи для управління взаємодією з клієнтами;

ERP-системи для управління ресурсами;

Інструменти аналітики та звітності.

Наявність великого обсягу структурованих даних є одним із ключових чинників для **1** застосування машинного навчання в прогнозуванні, оскільки якість і обсяг даних безпосередньо впливають на точність і надійність моделей.

Аналіз предметної області дозволяє визначити ряд викликів, з якими стикається підприємство:

Нестабільність попиту через сезонність і зовнішні фактори;

Потреба врахування великої кількості змінних у моделюванні;

Недосконалість і неповнота даних, що вимагає їхньої попередньої обробки;

Необхідність забезпечення швидкості отримання прогнозів для оперативного прийняття рішень;

Вимоги до інтуїтивності інтерфейсів, що використовують менеджери.

2.2 Аналіз наявних даних продажів: структура, обсяг, проблеми

Для побудови ефективної системи прогнозування продажів критично важливо

детально проаналізувати наявні дані. Цей розділ присвячений вивченню структури, обсягу та якості даних, а також ідентифікації основних проблем, які можуть впливати на точність і надійність майбутніх прогнозів.

Дані про продажі представляють собою багатовимірний набір інформації, що містить деталі щодо операцій, які відбулися в різні часові проміжки (рисунок 2.1). Основні компоненти структури включають:

Часові характеристики: дати і час здійснення продажу, що дозволяє аналізувати динаміку у часових рамках — днях, тижнях, місяцях, сезонах.

Товарна категорія: інформація про вид товару, бренди, характеристики (модель, колір, технічні параметри).

Географія продажів: місце реалізації — конкретні магазини, регіони, онлайн-платформа.

Обсяги продажів: кількість реалізованих одиниць, сума угоди.

Цінові параметри: ціна продажу, знижки, акційні пропозиції.

Клієнтська інформація: у разі наявності — демографічні дані, історія покупок.

Також можуть бути додаткові ознаки — канал продажів, метод оплати, інформація про постачання тощо.

Рисунок 2.1 – Структура наявних даних про продажі

Обсяг історичних даних суттєво впливає на можливості для навчання **1 моделей машинного навчання**. Умовне підприємство накопичило дані за декілька останніх років, що становить приблизно мільйони записів у базі. Такий великий обсяг дозволяє використовувати як прості статистичні методи, так і складні алгоритми глибокого навчання.

Висока якість даних є запорукою коректних результатів. Аналіз даних виявив низку проблем, типових для реальних бізнес-систем:

Пропуски і неповні записи: деякі поля, особливо у старих записах, можуть бути неповними або відсутніми.

Аномальні значення: випадки різко відхилених показників, які не відповідають загальним тенденціям (наприклад, продажі з надзвичайно великим обсягом).

Дублікати: повторні записи одних і тих самих операцій.

Розбіжності у форматах: різний формат зберігання дат, кодування товарів.

Проблеми зі зміною категорій: реорганізація товарних груп за період часу може призвести до неоднорідності.

Наступний графік (рисунок 2.2) демонструє наявність викидів у даних — надзвичайно великих чи маленьких обсягів продажів, які виходять за межі типового діапазону.

Рисунок 2.2 – Приклад виявлення аномальних значень у даних

Ці проблеми потребують спеціальних процедур очищення, нормалізації та підготовки даних до подальшої обробки.

Для прогнозування особливо важливо виділити ті параметри, які мають найбільший вплив на обсяги продажів. Це дозволить не лише покращити точність моделей, а й підвищити їхню інтерпретованість.

Аналітика показує, що серед ключових змінних — сезонність, категорія товару, канал продажів, поточні акції та ціни. Також важливо враховувати зовнішні фактори, які можуть бути додані як додаткові параметри (погода, економічні індикатори).

Проведений аналіз свідчить про необхідність ретельної **1** роботи з даними перед їх використанням для навчання моделей. Зокрема, рекомендовано:

Запровадити процедури регулярного очищення і валідації даних.

Стандартизувати формати збереження інформації.

Впровадити моніторинг якості даних.

Використовувати методи виявлення і корекції аномалій.

2.3 Постановка задачі прогнозування: мета, вхідні/вихідні параметри

Після детального аналізу **1** предметної області та наявних даних важливим кроком у розробці системи прогнозування є чітке формулювання самої задачі. Визначення мети, а також параметрів, які використовуватимуться на вході та очікуються на виході, є фундаментом для побудови ефективної моделі і коректної реалізації веб-застосунку.

Головною метою прогнозування є забезпечення підприємства інформацією про очікувані обсяги продажів у майбутньому з достатньою точністю, щоб оптимізувати управлінські рішення щодо закупівель, запасів, виробництва та маркетингу. Точне прогнозування дозволяє зменшити втрати від надлишкових запасів або дефіциту товарів, підвищити рівень обслуговування клієнтів і, як наслідок, збільшити

прибутковість.

Загалом мета прогнозування можна сформулювати так:

Підвищити ефективність планування ресурсів підприємства через точне передбачення обсягів продажів у задані часові проміжки.

Забезпечити адаптивність бізнесу до сезонних і ринкових змін.

Задача прогнозування формалізується як задача регресії, де **1** на основі історичних даних та супутніх параметрів слід побудувати модель, що прогнозує кількість продажів y_t на момент часу t .

$$y_t = f(X_t) + \epsilon_t \quad (2.1)$$

???????? — прогнозований обсяг продажів у момент часу

???????? — вектор вхідних параметрів (ознаковий простір), що включає історичні дані і додаткові фактори,

???? — функція прогнозування, яку визначає модель машинного навчання,

???????? — випадкова помилка прогнозу.

Для формування вхідного вектора ????????? використовуються різні типи даних, що впливають на обсяг продажів. Основні з них:

Історичні продажі: обсяги реалізації за попередні періоди (дні, місяці),

Сезонні фактори: місяць, день тижня, святкові дні,

Категорії товарів: тип продукції, бренд, модель,

Цінові характеристики: поточна ціна, знижки, акції,

Маркетингові активності: реклама, промо-кампанії,

Зовнішні фактори: погода, економічні індикатори,

Інші: регіон продажу, канал збуту (онлайн чи офлайн).

Це багатовимірний вхідний простір, який **1** моделі машинного навчання аналізують для виявлення закономірностей і побудови прогнозу.

Вихідним параметром є безпосередньо прогнозований обсяг продажів — числова величина, що показує очікувану кількість товарів, які буде реалізовано в певний

момент або період часу. В залежності від задачі, прогноз може бути:

Короткостроковим: прогноз на наступний день чи тиждень,

Середньостроковим: прогноз на місяць,

Довгостроковим: прогноз на квартал або рік.

У постановці задачі важливо враховувати деякі обмеження та припущення:

Дані про продажі є достатньо повними і репрезентативними для навчання моделі,

Зовнішні фактори впливають на продажі лінійно або у вигляді взаємодій, які можна виявити за допомогою ML,

Прогнози будуть використовуватись **12** як допоміжний інструмент прийняття рішень, а не як єдиний критерій,

Існує можливість регулярного оновлення моделі з новими даними для підтримки актуальності прогнозів.

Для оцінки якості побудованої моделі визначаються метрики точності прогнозів. Серед них:

MAE (Mean Absolute Error): **2** середнє абсолютне відхилення,

2 RMSE (Root Mean Squared Error): корінь середньоквадратичної помилки,

R^2 (коефіцієнт детермінації): частка варіації, пояснена моделлю.

Оптимальна модель має мінімізувати помилки і максимізувати пояснену дисперсію.

2.4 Вибір методів обробки даних: нормалізація, очищення, агрегація

Обробка даних є ключовим етапом у побудові системи прогнозування, що значною мірою визначає якість та точність майбутніх моделей машинного навчання. На практиці сирі дані часто містять помилки, пропуски, різномірні формати, а також можуть мати різний масштаб. Відповідна попередня обробка дозволяє підготувати дані таким чином, щоб мінімізувати шум і виявити корисні закономірності.

Очищення — це процес усунення помилок, пропусків і аномалій у даних, що забезпечує їхню якість і коректність.

Основні етапи очищення:

Виявлення і заповнення пропусків: Пропуски в даних можуть виникати через технічні збої або неповне збирання інформації. Для заповнення застосовують методи інтерполяції, середні значення, або **2** моделі на основі сусідніх записів.

Видалення дублікатів: Повторні записи спотворюють статистику і призводять до перекручування прогнозів.

Виявлення і корекція аномалій: Аномальні значення (викиди) можуть бути викликані помилками або реальними, але рідкісними подіями. Їх або видаляють, або замінюють на розумні значення залежно від контексту.

Уніфікація форматів: Вирівнювання форматів дат, назв категорій, числових значень для забезпечення однорідності.

Нормалізація — це трансформація даних у єдиний масштаб, що дозволяє уникнути переважання змінних із великим діапазоном значень під час навчання моделей.

Основні методи нормалізації:

Мін-макс нормалізація: масштабування даних у діапазон [0, 1] за формулою:

$$x' = (x - x_{\min}) / (x_{\max} - x_{\min}) \quad (2.2)$$

Z-нормалізація (стандартизація): перетворення даних так, щоб середнє було 0, а стандартне відхилення — 1:

$$x' = (x - \mu) / \sigma \quad (2.3)$$

Даний етап особливо важливий для моделей, чутливих до масштабу ознак, таких як нейронні мережі або методи k-ближчих сусідів.

Агрегація полягає у формуванні узагальнених показників на основі детальних даних. Вона допомагає зменшити обсяг інформації, зробити її більш осмисленою і полегшити моделювання.

Приклади агрегації:

Об'єднання продажів за днями у тижневі або місячні обсяги,

Підсумовування обсягів за категоріями товарів,

Формування показників середнього чека або кількості покупок за період.

Правильний вибір ступеня агрегації впливає на здатність моделі виявляти тренди і сезонність.

В межах розробки системи прогнозування планується застосувати комплексний підхід до обробки даних:

Очищення із застосуванням як простих методів (заповнення пропусків середнім), так і більш складних (аналіз аномалій з використанням статистичних методів).

Нормалізація з вибором оптимального методу залежно від моделі: для лінійної регресії і ансамблевих моделей нормалізація необов'язкова, натомість для нейронних мереж — критично необхідна.

Агрегація буде здійснюватися на різних рівнях для побудови моделей із різним горизонтом прогнозування.

2.5 Визначення критеріїв оцінки ефективності прогнозу

Для оцінки якості моделей прогнозування надзвичайно важливо застосовувати відповідні метрики, які дозволяють кількісно виміряти точність і надійність отриманих результатів. Використання чітко визначених критеріїв дає змогу не лише порівнювати різні моделі між собою, а й контролювати якість прогнозів у процесі їхньої експлуатації.

У задачах регресії, якими є прогнозування кількості продажів, найпоширенішими метриками точності є:

1. **3** Середня абсолютна помилка (MAE — **2** Mean Absolute Error).

2 MAE обчислює середнє абсолютне відхилення прогнозованих **3** значень від фактичних. Вона дає уявлення про середню величину помилки в тих же одиницях, що й прогнозовані дані.

$$MAE = (1/n) * \sum |y_i - \hat{y}_i| \quad (2.4)$$

y_i **2** — фактичне значення,

2 \hat{y}_i — прогнозоване значення,

n — кількість спостережень.

Переваги MAE — простота інтерпретації і стійкість до аномальних значень.

2. **3** Корінь середньоквадратичної помилки (RMSE **3** — **3** Root Mean Squared Error)

3 RMSE є **13** квадратним коренем середнього квадрата різниці між фактичними і прогнозованими значеннями. Вона чутливіша до великих помилок, оскільки квадратичний член посилює вплив значних відхилень.

$$\text{RMSE} = \sqrt{(1/n) * \sum (y_i - \hat{y}_i)^2} \quad (2.5)$$

RMSE часто використовується, коли важливо уникнути великих похибок, що можуть суттєво вплинути на бізнес-процеси.

3. **2** Коефіцієнт детермінації (R^2)

2 R^2 показує, **5** яку частку варіації залежної змінної пояснює модель. Значення R^2 варіюються від 0 до 1 (іноді можуть бути від'ємними, якщо модель дуже погана).

$$R^2 = 1 - (\sum (y_i - \hat{y}_i)^2 / \sum (y_i - \bar{y})^2) \quad (2.6)$$

\bar{y} **2** — середнє значення фактичних **13** даних.

5 Чим ближче R^2 **5** до 1, тим краще модель описує дані.

4. Середня абсолютна відносна помилка (MAPE — **2** Mean Absolute Percentage Error)

2 MAPE вимірює **13** середню абсолютну похибку у відсотках, що дозволяє порівнювати точність прогнозів на різних масштабах.

$$\text{MAPE} = (100\% / n) * \sum |(y_i - \hat{y}_i) / y_i| \quad (2.7)$$

Недоліком MAPE є погана робота з нульовими або дуже малими фактичними значеннями.

Вибір конкретних метрик оцінки залежить від особливостей задачі і бізнес-вимог.

13 Для задач прогнозування продажів найбільш релевантними є MAE і RMSE, оскільки вони чітко відображають розмір похибок у фактичних одиницях виміру. R^2 допомагає оцінити якість пояснення даних моделлю.

MAPE може бути використаний для додаткового аналізу, особливо коли потрібно порівнювати ефективність прогнозів між різними товарами або категоріями.

Посилання

Це джерела виділених збігів у вашому документі. Кожен збіг позначено темно-зеленим числом, яке відповідає вказаному тут джерелу. Джерела впорядковані за схожістю — чим вищий бал, тим сильніше збіг.

#	Джерело	%
1	krs.chmnu.edu.ua	1.1%
2	ir.stu.cn.ua	0.7%
3	core.ac.uk	0.4%
4	warn-erasmus.eu	0.3%
5	duikt.edu.ua	0.3%
6	dspace.wunu.edu.ua	0.2%
7	duikt.edu.ua	0.2%
8	eir.nuos.edu.ua	0.2%
9	essuir.sumdu.edu.ua	0.2%
10	metod.vntu.edu.ua	0.2%
11	duikt.edu.ua	0.2%
12	ela.kpi.ua	0.2%
13	ela.kpi.ua	0.2%
14	nubip.edu.ua	0.2%
15	rau.ua	0.2%
16	dspace.wunu.edu.ua	0.1%
17	javascript.org.ua	0.1%
18	econom.knu.ua	0.1%
19	dspace.znu.edu.ua	0.1%



Дякуємо, що перевірили
свій документ за допомогою
Plag!