

СХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ
ВОЛОДИМИРА ДАЛЯ

Факультет інформаційних технологій та електроніки

Кафедра інформаційних технологій та програмування

Пояснювальна записка
до магістерської дипломної роботи

магістр

(освітньо-кваліфікаційний рівень)

на тему: Дослідження методів аналізу та прогнозування продажів і
і розробка алгоритму їх практичного застосування

Виконав: студент 2 курсу, групи ІСТ-24зм
126 «Інформаційні системи та технології

(шифр і назва спеціальності)

Щіров Д.С.

(прізвище та ініціали)

Керівник Дьомін М. К.

(прізвище та ініціали)

Рецензент Меняйленко О.С.

(прізвище та ініціали)

Київ – 2025 року

СХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ
ВОЛОДИМИРА ДАЛЯ

Факультет інформаційних технологій та електроніки
Кафедра інформаційних технологій та програмування
Освітньо-кваліфікаційний рівень магістр
Спеціальність 126 «Інформаційні системи та технології»
(шифр і назва спеціальності)

ЗАТВЕРДЖУЮ
Завідувач кафедри ІТП
_____ д.т.н., проф.Захожай О.І.
(підпис)
« ____ » _____ 2025 р.

ЗАВДАННЯ

на магістерську дипломну роботу студенту

Щіров Дмитро Сергійович

(прізвище, ім'я, по батькові)

1. Тема роботи: Дослідження методів аналізу та прогнозування продажів і розробка алгоритму їх практичного застосування,
керівник роботи к.т.н., доцент, Дьомін Максим Костянтинівч,
(вчене звання, науковий ступінь, прізвище, ім'я, по батькові)
затверджені наказом університету від « 6 » 12 2024року №361/15.15-С
2. Строк подання студентом роботи: 18 грудня 2025 р.
3. Вихідні дані до роботи: Матеріали науково-дослідної практики, науково-методична література; дані інтернет-мережі .
4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)
 - 4.1 Вступ
 - 4.2 Теоретичні основи аналізу та прогнозування продажів (огляд публічних джерел інформації, спеціальної літератури)
 - 4.3 Аналіз предметної області та постановка задачі (розгляд алгоритмів прогнозування, їх варіації та формулювання цілей).
 - 4.4 Розробка методу та програмної реалізації прогнозування продажів (створення архітектури алгоритму, вибір засобів реалізації)
 - 4.5 Програмна реалізація (написання коду, тестування моделей)
 - 4.6 Висновки
 - 4.7 Перелік використаних джерел
5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)

6. Консультанти розділів проєкту (роботи)

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		Завдання видав	Завдання прийняв
Розділ 1. Аналіз існуючих моделей побудови прогнозів, їх характеристик	Дьомін М. К., к.т.н., доцент		
Розділ 2. Проектування алгоритму моделі, пошук бази для тестування прогнозів	Дьомін М. К., к.т.н., доцент		
Розділ 3. Тестування моделей прогнозування	Дьомін М. К., к.т.н., доцент		

7. Дата видачі завдання 07.11. 2025р.**КАЛЕНДАРНИЙ ПЛАН**

№ з/п	Назва етапів дипломної роботи	Строк виконання етапів роботи	Примітка
1.	Отримання та уточнення теми магістерської кваліфікаційної роботи	11.10.2025	виконано
2.	Формування та погодження структури роботи з науковим керівником	22.10.2025	виконано
3.	Аналіз наукових джерел з методів аналізу та прогнозування часових рядів	28.10.2025	виконано
4.	Дослідження існуючих підходів до прогнозування продажів та вибір методів	02.11.2025	виконано
5.	Аналіз інструментів і програмних засобів для реалізації прогнозних моделей	09.11.2025	виконано
6.	Реалізація алгоритмів аналізу та прогнозування продажів	18.11.2025	виконано
7.	Оформлення пояснювальної записки та погодження з науковим керівником	06.12.2025	виконано
8.	Подання магістерської кваліфікаційної роботи на кафедру	12.12.2025	виконано
9.	Підготовка доповіді та презентаційних матеріалів до захисту	12.12.2025	виконано

Студент Щіров Д.С.
(підпис) (прізвище та ініціали)Керівник роботи Дьомін М. К.
(підпис) (прізвище та ініціали)

ЗМІСТ

ВСТУП	8
РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ ПРОДАЖІВ	10
1.1. Інформаційні системи в аналізі продажів	10
1.2. Часові ряди та їх властивості в аналізі продажів	14
1.3. Класичні методи аналізу часових рядів	18
1.4. Методи прогнозування динаміки продажів	25
1.5. Оцінка якості прогнозних моделей	32
1.6. Управління та підготовка даних для прогнозування продажів.....	35
РОЗДІЛ 2. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧІ	40
2.1. Характеристика предметної області	40
2.2. Джерела даних та їх структура	43
2.3. Основні проблеми аналізу продажів у бізнес-середовищі	46
2.4. Постановка задачі прогнозування	49
2.5. Обґрунтування необхідності розробки алгоритму	51
РОЗДІЛ 3. РОЗРОБКА МЕТОДУ ТА ПРОГРАМНОЇ РЕАЛІЗАЦІЇ ПРОГНОЗУВАННЯ ПРОДАЖІВ	52
3.1. Архітектура алгоритму прогнозування продажів.....	52

3.2. Очищення та підготовка даних	54
3.3. Модель Holt-Winters: принципи роботи, можливості та практичні обмеження	57
3.4. Моделі ARIMA та SARIMA: підхід, можливості та практичні аспекти застосування	63
3.5. Обґрунтування вибору моделей та структурного підходу до прогнозування	70
3.6. Інструменти та технології, використані для реалізації алгоритму .	74
РОЗДІЛ 4. ПРОГРАМНА РЕАЛІЗАЦІЯ СИСТЕМИ ПРОГНОЗУВАННЯ ПРОДАЖІВ	78
4.1. Підготовка даних та середовища виконання	78
4.2. Реалізація моделей Holt–Winters та SARIMA	80
4.3. Експеримент 1: прогнозування даних про видобуток нафти (місячна частота)	83
4.4. Експеримент 2: Спроба прогнозування на повному датасеті без попередньої сегментації.....	86
4.5. Експеримент 3: прогнозування категорії «Dairy» (магазин №1)	89
4.6. Експеримент 4: прогнозування категорії «Bread/Bakery»	91
4.7. Експеримент 5: категорія «Frozen Foods» (магазин №22) - інтермітуючий попит та провал моделей.....	94
4.8. Експеримент 6: категорія «Frozen Foods» (магазин №1).....	97
4.9. Експеримент 7: прогнозування на основі тижневої сезонності (oil.csv)	100

РОЗДІЛ 5. ЗАГАЛЬНІ ВИСНОВКИ МАГІСТЕРСЬКОЇ РОБОТИ	103
ПЕРЕЛІК ПОСИЛАНЬ	108
ДОДАТКИ	110

ВСТУП

У сучасній економіці дані стали одним із ключових стратегічних ресурсів, від якого залежить конкурентоспроможність підприємств різного масштабу. Рішення, що ґрунтуються на інтуїції або попередньому досвіді, вже не здатні забезпечити необхідну точність прогнозів, особливо в умовах високої мінливості ринку, нестабільності попиту та зростаючої конкуренції. Саме тому бізнес усе активніше впроваджує інформаційні системи, які дозволяють систематично збирати, обробляти та аналізувати дані про продажі, а також формувати прогнози, що підтримують ухвалення обґрунтованих управлінських рішень.

Аналітика продажів є важливою складовою процесу планування діяльності підприємства. Вона дозволяє виявляти закономірності в поведінці споживачів, визначати сезонні коливання, аналізувати вплив зовнішніх факторів та будувати прогнози на майбутні періоди. Для компаній, що працюють у сфері торгівлі, виробництва чи логістики, точне прогнозування обсягів продажів безпосередньо впливає на оптимізацію запасів, планування виробництва, логістичні процеси, фінансові показники та ефективність маркетингових стратегій. Навіть незначні похибки прогнозу можуть призводити до надлишкових витрат або втрати потенційного прибутку, що підкреслює необхідність впровадження систем прогнозування високої точності.

Методи прогнозування, що застосовуються у бізнес-практиці, постійно вдосконалюються. Класичні статистичні підходи, які тривалий час були основою планування, доповнюються сучасними інтелектуальними технологіями, що дозволяють працювати з великими обсягами даних, враховувати складні нелінійні залежності та будувати адаптивні моделі. У такому контексті важливим є не лише теоретичне розуміння методів аналізу часових рядів, але й здатність застосовувати їх на практиці - з використанням

інструментів, що забезпечують автоматизацію процесу прогнозування та роблять його більш прозорим і керованим.

Особливого значення набуває питання якості даних, адже саме від нього залежить надійність і коректність моделі прогнозування. На практиці підприємства часто стикаються з неповнотою, несистематизованістю або застарілістю інформації, що негативно впливає на точність результатів. Це вимагає впровадження підходів до очищення, нормалізації, агрегації та структурування даних перед побудовою прогнозної моделі. Сучасна система прогнозування повинна включати комплекс інструментів, що охоплює повний цикл роботи з даними - від збору до аналізу й моделювання.

Сучасні інструменти бізнес-аналітики та машинного навчання відкривають широкі можливості для моделювання динаміки продажів. Платформи бізнес-аналітики дозволяють проводити візуальний аналіз і будувати інтерактивні звіти, тоді як спеціалізовані інструменти машинного навчання використовуються для створення більш складних моделей прогнозування. Однак вибір оптимального методу залежить від структури даних, рівня їх деталізації, часових інтервалів, наявності сезонності, доступних технічних ресурсів та вимог точності. Тому важливо не тільки систематизувати існуючі підходи, а й оцінити їх ефективність на практичному прикладі.

Актуальність цієї магістерської кваліфікаційної роботи зумовлена потребою підприємств у підвищенні точності прогнозування та вдосконаленні процесів управління даними. Умови сучасного ринку вимагають від компаній здатності швидко реагувати на зміни попиту, оптимізувати запаси й адаптувати виробничі та маркетингові стратегії. Прогнозування продажів стає одним із ключових інструментів, який дозволяє зменшити ризики, підвищити ефективність використання ресурсів та забезпечити стабільний розвиток організації. Розробка алгоритму, здатного

виконувати аналіз даних та формувати прогностні оцінки, є важливим етапом у створенні інформаційних систем нового покоління.

Метою магістерської кваліфікаційної роботи є комплексне дослідження методів аналізу та прогнозування продажів, вивчення їх теоретичних основ і практичного застосування, а також розробка алгоритму, що дозволяє формувати прогнози на основі історичних даних з урахуванням тенденцій та сезонних коливань. У межах роботи буде проведено аналіз методів статистичного та інтелектуального прогнозування, порівняно їхні можливості, досліджено сучасні інформаційні системи, а також представлено реалізацію алгоритму прогнозування продажів на реальному наборі даних.

У першому розділі розглядаються теоретичні основи аналізу продажів і методи прогнозування. Другий розділ присвячений аналізу предметної області та огляду сучасних інформаційних систем, що забезпечують можливості аналітики та прогнозування. У третьому розділі сформульовано постановку задачі, описано методичні підходи та побудовано архітектуру алгоритму прогнозування. Четвертий розділ містить практичну реалізацію алгоритму на основі реальних даних та оцінку отриманих результатів. Роботу завершують висновки та рекомендації щодо подальшого розвитку системи.

Магістерська кваліфікаційна робота поєднує теоретичне дослідження й практичну реалізацію, що дозволяє глибше зрозуміти особливості прогнозування продажів і дає можливість продемонструвати варіанти застосування сучасних інформаційних технологій у цій сфері.

РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ ПРОДАЖІВ

1.1. Інформаційні системи в аналізі продажів

Інформаційні системи дедалі частіше стають основою організації роботи підприємств. З їх допомогою компанії впорядковують дані, стежать за змінами попиту, управляють процесами та отримують аналітичну

інформацію, необхідну для щоденних рішень. У міру зростання обсягів даних і швидкості їх оновлення роль таких систем у сфері продажів лише посилюється, адже вони дозволяють оперативно реагувати на зміни ринку та краще розуміти поведінку споживачів. У сфері торгівлі такі системи виконують роль основної інфраструктури, на якій будуються моделі прийняття рішень, плани закупівель, складська логістика та управління маркетинговими кампаніями.

Інформаційна система в широкому розумінні - це сукупність технічних засобів, програмного забезпечення, даних, методів та процедур, що забезпечують збирання, збереження, обробку та передачу інформації користувачам. У практичному контексті це означає, що інформаційна система не обмежується лише програмою чи базою даних: вона включає цілісний цикл роботи з інформацією - від її надходження до використання результатів аналізу у прийнятті рішень.

У контексті роботи з даними продажів інформаційні системи дозволяють підприємствам:

- **контролювати поточний стан продажів**, отримувати аналітичні показники в режимі реального часу;
- **виявляти аномалії та закономірності**, такі як різкі зростання попиту, сезонні піки чи падіння;
- **проводити порівняльний аналіз** між різними товарами, магазинами, регіонами або періодами часу;
- **прогнозувати обсяги продажів**, що є критично важливим для планування запасів, бюджету та логістики;
- **автоматизувати рутинні процеси**, зменшуючи вплив людського фактора на якість даних.

З огляду на різноманітність задач, що стоять перед бізнесом, інформаційні системи можна класифікувати на декілька груп, кожна з яких відіграє свою роль у процесі аналізу даних.

Операційні системи (Transaction Processing Systems, TPS) відповідають за фіксацію та обробку транзакцій - зокрема, продажів, повернень, замовлень чи платежів. Першоджерелом даних для аналітики виступають саме ці системи, адже вони формують початкову структуру даних.

Системи управління взаємовідносинами з клієнтами (Customer Relationship Management, CRM) фокусуються на організації комунікацій з клієнтами та контролю взаємодій на різних етапах продажу. CRM-системи зберігають історію покупок, поведінку клієнтів, інформацію про маркетингові кампанії, що дозволяє створювати персоналізовані підходи до прогнозування попиту.

Системи управління ресурсами підприємства (Enterprise Resource Planning, ERP) забезпечують інтеграцію всіх бізнес-процесів: від обліку запасів до фінансового аналізу. Для прогнозування продажів ERP-системи важливі тим, що містять інформацію про залишки товарів, закупівлі та витрати, дозволяючи оцінювати доступність ресурсів та ефективність операцій.

Найбільш значущу роль у прогнозуванні відіграють **системи бізнес-аналітики (Business Intelligence, BI)**. Їх основна мета - перетворення необроблених даних на інформацію, що має аналітичну цінність. BI-системи надають можливості для:

- створення інтерактивних звітів та дашбордів,
- багатовимірного аналізу даних,
- визначення трендів та сезонних коливань,

- побудови первинних прогнозів на основі статистичних моделей.

Яскравими прикладами BI-систем є Power BI, Tableau, Looker Studio та Qlik Sense, які активно застосовуються як у малому, так і у великому бізнесі.

У сучасних умовах до BI-систем дедалі частіше додаються компоненти штучного інтелекту та машинного навчання. Такі системи здатні аналізувати складні взаємозв'язки між факторами, враховувати велику кількість змінних і формувати прогнози, що адаптуються до нових даних. На відміну від класичних методів, моделі машинного навчання не потребують попередньої ручної побудови гіпотез про структуру даних, а самостійно визначають основні закономірності.

Важливою концепцією, яка тісно пов'язана з інформаційними системами аналізу продажів, є **data-driven підхід**. Він передбачає, що ключові бізнес-рішення приймаються не інтуїтивно, а на основі даних та їх аналітичної інтерпретації. Компанії, які впровадили подібний підхід, отримують конкурентні переваги завдяки кращому розумінню поведінки клієнтів, точнішому плануванню запасів, прогнозуванню попиту та оптимізації цінових стратегій. У довгостроковій перспективі data-driven культура дозволяє зменшити витрати, підвищити швидкість реакції на ринкові зміни та знизити ризики прийняття невірних управлінських рішень.

Однак інформаційні системи - це не лише можливості, але й виклики. У реальних умовах підприємства стикаються з проблемами якості даних: пропусками, дублюванням, невірними значеннями або невідповідностями між різними джерелами інформації. Через це побудова якісного прогнозу неможлива без застосування механізмів очищення та нормалізації даних. Іншими словами, інформаційна система має включати не лише аналітичний модуль, але й засоби підготовки даних до аналізу. Саме тому питання управління даними набуває все більшої актуальності, і в окремих випадках

якість даних може впливати на результат навіть більше, ніж сам вибір моделі прогнозування.

Іншим важливим аспектом є масштабованість системи. У міру зростання обсягів продажів та кількості джерел даних інформаційна система має забезпечувати стабільність роботи та можливість обробки великих масивів інформації без втрати продуктивності. Це особливо актуально для підприємств, що працюють у сфері електронної комерції, де щодня генеруються мільйони транзакцій.

У підсумку, інформаційні системи відіграють фундаментальну роль у процесах аналізу та прогнозування продажів. Вони створюють основу для формування аналітичних висновків, оптимізації бізнес-процесів і побудови моделей, здатних враховувати складні тенденції та коливання ринку. Без якісної інформаційної системи прогнозування обсягів продажів перетворилося б на здогадку, тоді як саме на основі системного аналізу даних підприємства можуть будувати стратегії розвитку та підвищувати свою ефективність.

1.2. Часові ряди та їх властивості в аналізі продажів

Коли йдеться про вивчення закономірностей у продажах, перш за все звертають увагу на те, як показник змінюється в часі. Саме так виникає поняття часового ряду - послідовності спостережень, упорядкованих за датами чи періодами. Такий підхід дозволяє побачити не окремі значення, а динаміку: плавні тенденції, сезонні коливання або різкі зміни, що можуть бути важливими для прогнозування [\[8\]](#). У контексті аналізу продажів такими показниками найчастіше виступають денні, тижневі, місячні або квартальні дані про кількість проданих товарів, виручку або середній чек.

Особливість часових рядів полягає в тому, що значення показника залежать від часу, а це означає, що передбачення майбутніх значень можливе лише за умов виявлення закономірностей у минулих спостереженнях. Саме

тому в аналітиці продажів часові ряди виступають основою для побудови математичних моделей, здатних відтворювати поведінку системи та прогнозувати її розвиток.

Компоненти часових рядів

Для коректного аналізу важливо враховувати, що будь-який часовий ряд можна розкласти на кілька основних компонентів. Це дозволяє окремо аналізувати вплив різних факторів і підвищує точність прогнозової моделі [\[7\]](#).

1. Трендова складова (Trend)

Тренд описує довгострокову тенденцію зміни показника - наприклад, стабільне зростання продажів завдяки розширенню ринку або поступове зменшення попиту на застарілий товар. Тренди можуть бути лінійними (сталій ріст або спад), експоненціальними (прискорене зростання) або більш складними (S-подібні криві, поліноміальні моделі).

У практиці роздрібної торгівлі тренд часто відображає зміну рентабельності товарних груп, розвиток бренду або загальні економічні умови.

2. Сезонна складова (Seasonality)

Сезонність - це регулярні та передбачувані коливання, що повторюються з однаковою періодичністю. У продажах сезонність є одним із найяскравіших та найважливіших явищ.

Приклади сезонності:

- збільшення продажів побутової техніки перед Новим роком;
- зростання продажів напоїв у літній сезон;
- падіння попиту на зимовий одяг навесні;
- стабільні піки онлайн-замовлень по п'ятницях.

Сезонність може бути річною, квартальною, щомісячною чи навіть тижневою - усе залежить від природи товару та поведінки клієнтів.

3. Циклічна складова (Cycle)

Циклічність відображає коливання, які не мають фіксованої періодичності, на відміну від сезонності. Вона пов'язана з економічними циклами, змінами попиту на ринку, поведінкою конкурентів або коливаннями в певних галузях.

Наприклад:

- зростання й спад попиту на смартфони через появу нових поколінь моделей;
- попит на будматеріали, що змінюється внаслідок економічних криз або політичних факторів.

Циклічні коливання складніше моделювати, але їх наявність часто впливає на довгострокові прогнози.

4. Випадкова складова (Noise / Irregular)

Це нерегулярні коливання, що виникають під впливом випадкових або непередбачуваних факторів:

- раптові збої поставок;
- акційні продажі;
- погодні умови;
- одноразові події.

Метою прогнозування є мінімізація впливу випадкової складової шляхом виокремлення стабільних закономірностей.

Типи часових рядів

У аналітиці продажів використовують два основні типи часових рядів[8]:

Адитивні моделі

Кожна компонента додається до інших:

$$Y(t) = \text{Trend} + \text{Seasonality} + \text{Cycle} + \text{Noise}$$

Такі ряди підходять, коли амплітуда сезонності залишається стабільною незалежно від рівня продажів.

Мультиплікативні моделі

Компоненти взаємодіють між собою:

$$Y(t) = \text{Trend} \times \text{Seasonality} \times \text{Cycle} \times \text{Noise}$$

Цей тип зустрічається частіше в роздрібній торгівлі, коли сезонні коливання збільшуються разом із зростанням тренду (наприклад, продажі взуття в масштабованих мережах).

Проблеми та виклики роботи з часовими рядами

У реальних даних про продажі часто виникають труднощі, які ускладнюють моделювання:

1. Пропуски даних

Продажі можуть не фіксуватися в окремі дні або періоди, що порушує послідовність.

2. Аномальні значення

Різкі піки або падіння в даних можуть бути спричинені акціями, знижками, дефіцитом товару, помилками обліку.

3. Зміна поведінки ринку

В умовах війни, криз чи зміни конкуренції сезонність може суттєво змінюватися.

4. Вплив свята і розпродажів

Такі події можуть істотно деформувати тренд та сезонність (Чорна п'ятниця, передноворічні знижки).

5. Недостатній обсяг даних

Моделі сезонності потребують хоча б 2-3 повних цикли (наприклад, 2-3 роки).

Практичне значення аналізу часових рядів у продажах

Використання часових рядів дозволяє підприємствам:

- прогнозувати попит із урахуванням сезонності;
- оптимізувати закупівлі та складські запаси;
- визначати періоди для запуску акцій;
- аналізувати ефективність маркетингових кампаній;
- оцінювати майбутні доходи та планувати бюджети.

Для компаній, які працюють у сфері ритейлу або e-commerce, аналіз часових рядів часто є основою операційної аналітики. Наприклад, великі торговельні мережі використовують сезонні моделі для планування товарних запасів, а інтернет-магазини - для оптимізації логістики та прогнозування навантаження на службу доставки.

1.3. Класичні методи аналізу часових рядів

Класичні статистичні методи аналізу часових рядів займають центральне місце в дослідженні динаміки продажів, оскільки забезпечують базовий інструментарій для виявлення ключових закономірностей у даних.

Незважаючи на розвиток машинного навчання та інтелектуальних систем, традиційні підходи залишаються високоцінними завдяки простоті, інтерпретованості та відносно низьким вимогам до обсягу даних. Саме з них починається побудова будь-якої прогностичної моделі, особливо коли аналітик працює з трендом, сезонністю та згладжуванням коливань.

У сфері продажів класичні методи дозволяють виконувати попередній аналіз, оцінювати тенденції, виявляти ритмічність та очищувати дані від випадкових шумів[8]. Крім того, вони застосовуються як основа для початкових прогнозів, що можуть слугувати еталоном для порівняння зі складнішими моделями.

Метод рухомого середнього (Moving Average)

Метод рухомого середнього належить до базових підходів згладжування часових рядів і часто використовується як початковий інструмент аналізу динаміки показників. Його суть полягає в обчисленні середнього значення за фіксоване «вікно» попередніх періодів, що дозволяє зменшити вплив випадкових коливань і зробити загальну тенденцію більш помітною[8].

Формально рухоме середнє для періоду k визначається як:

$$MA_t = \frac{Y_t + Y_{t-1} + \dots + Y_{t-k+1}}{k}$$

Завдяки усередненню значень метод ефективно знижує рівень шуму в даних і полегшує візуальний та аналітичний аналіз динаміки продажів. Водночас усі спостереження в межах вікна мають однакову вагу, що призводить до запізнення реакції на різкі зміни попиту. Така особливість є критичною для ринків із високою мінливістю, де оперативне реагування має вирішальне значення. Крім того, метод рухомого середнього не враховує сезонні коливання, що обмежує його застосування для рядів зі складною

циклічною структурою. На практиці метод часто використовується як допоміжний інструмент для попередньої оцінки тенденцій.

Наприклад, мережа магазинів може застосовувати семиденне рухоме середнє для аналізу щоденних продажів з метою згладжування впливу вихідних та святкових днів. Таке згладжування дозволяє більш об'єктивно оцінити, чи спостерігається зростання або спад попиту протягом останнього тижня, без спотворення результатів короткостроковими коливаннями.

Метод зваженого рухомого середнього (Weighted Moving Average)

Недоліком класичного рухомого середнього є те, що всі спостереження мають однакову вагу, хоча в реальності останні дані частіше мають більшу значущість. У такому випадку застосовується зважене рухоме середнє, де новіші значення отримують вищі коефіцієнти.

Наприклад, для трьох періодів можна застосувати ваги:

- 0,5 для найсвіжішого значення,
- 0,3 для попереднього,
- 0,2 для ще одного попереднього.

Такий підхід дозволяє швидше реагувати на зміни попиту.

Експоненціальне згладжування (Exponential Smoothing)

Експоненціальне згладжування застосовується для аналізу часових рядів у випадках, коли необхідно зменшити вплив випадкових коливань і водночас зберегти актуальність останніх спостережень. Ключова ідея методу полягає в тому, що новіші значення ряду мають більший вплив на прогноз, ніж більш давні, що дозволяє моделі оперативно адаптуватися до змін у поведінці показника[8].

Загальна формула:

$$S_t = \alpha Y_t + (1 - \alpha) S_{t-1}$$

Де S_t - згладжене значення,

Y_t - фактичні дані,

α - коефіцієнт згладжування (0–1).

Математично метод ґрунтується на рекурсивному обчисленні згладженого значення, яке формується як зважена комбінація попереднього згладженого рівня та поточного фактичного спостереження. Величина коефіцієнта згладжування α визначає швидкість реакції моделі на зміни: за більших значень α прогноз швидше реагує на нові дані, тоді як менші значення забезпечують більш плавну динаміку.

У практиці прогнозування продажів експоненціальне згладжування є доцільним для короткострокових прогнозів у відносно стабільних умовах, коли відсутні виражені сезонні коливання. Метод добре працює для оперативного аналізу попиту, наприклад, при щоденному або тижневому плануванні, де важливо швидко враховувати останні зміни. Водночас відсутність механізмів урахування тренду та сезонності обмежує застосування цього підходу для довших часових горизонтів або для даних зі складною структурою, де необхідні більш розширені моделі.

Метод Холта (Holt's Linear Trend Method)

Метод Холта є розвитком простого експоненціального згладжування та призначений для аналізу часових рядів, у яких спостерігається виражений тренд. На відміну від базового підходу, модель Холта[8] дозволяє окремо оцінювати поточний рівень показника та швидкість його зміни в часі, що робить прогноз більш адаптивним до поступового зростання або спаду.

Моделі визначається формулами:

$$L_t = \alpha \cdot Y_t + (1 - \alpha) \cdot (L_{t-1} + B_{t-1})$$

$$B_t = \beta \cdot (L_t - L_{t-1}) + (1 - \beta) \cdot B_{t-1}$$

де

L_t - рівень,

b_t - тренд,

α, β - коефіцієнти згладжування.

Завдяки явному урахуванню трендової компоненти метод забезпечує точніше відтворення довгострокової динаміки у порівнянні з простим експоненціальним згладжуванням. Це особливо корисно для аналізу продажів товарів, попит на які змінюється поступово під впливом розвитку ринку або зміни споживчих вподобань. Можливість окремого налаштування параметрів згладжування рівня та тренду дозволяє адаптувати модель до різних типів даних. Водночас метод Холта не враховує сезонні коливання, що обмежує його застосування у задачах прогнозування продажів товарів із яскраво вираженою циклічністю попиту. У таких випадках використання цієї моделі без розширення може призводити до систематичних похибок, що зумовлює необхідність застосування більш складних підходів, зокрема моделей Холта-Вінтерса.

Метод Холта - Вінтерса (Holt-Winters Method)

Метод Холта-Вінтерса є подальшим розвитком методу Холта та призначений для аналізу часових рядів, у яких одночасно присутні тренд і сезонні коливання. На практиці це робить модель особливо корисною для прогнозування продажів товарів, попит на які змінюється не лише в довгостроковій перспективі, але й має регулярну циклічну структуру[\[9\]](#).

Цей метод є одним з найефективніших класичних способів прогнозування продажів, оскільки він враховує три компоненти:

- рівень,

- тренд,
- сезонність.

Існують адитивна та мультиплікативна версії моделі.

У мультиплікативній формі, найпоширенішій для роздрібних продажів, формули мають вигляд:

$$L_t = \alpha \cdot (Y_t / S_{t-m}) + (1 - \alpha) \cdot (L_{t-1} + B_{t-1})$$

$$B_t = \beta \cdot (L_t - L_{t-1}) + (1 - \beta) \cdot B_{t-1}$$

$$S_t = \gamma \cdot (Y_t / L_t) + (1 - \gamma) \cdot S_{t-m}$$

$$\hat{Y}_{t+h} = (L_t + h \cdot B_t) \cdot S_{t-m+h}$$

де

s_t - сезонний індекс,

m - довжина сезонного циклу.

Ключовою особливістю методу є можливість окремого моделювання рівня, тренду та сезонної компоненти, причому сезонні індекси можуть змінюватися з часом. Завдяки цьому модель здатна адаптуватися до поступових змін ринкової ситуації та не фіксує сезонність як сталу величину. У порівнянні з класичними статистичними моделями, Holt-Winters забезпечує збалансоване поєднання точності прогнозу та відносної простоти реалізації, що робить його популярним у прикладних задачах прогнозування. Водночас ефективне застосування методу потребує достатнього обсягу історичних даних для коректної оцінки сезонної структури. За коротких або сильно зашумлених рядів параметри моделі можуть оцінюватися нестабільно, що ускладнює її налаштування у порівнянні з простішими методами згладжування. Тому використання Holt-Winters доцільне

насамперед у випадках, коли сезонність є чіткою та підтверджується історичними спостереженнями.

Метод часто застосовується для аналізу місячних обсягів реалізації товарів із вираженою сезонністю. Наприклад, для продажів взуття, де попит істотно відрізняється між літнім і зимовим періодами, Holt-Winters здатний забезпечити кращу адаптацію до змін сезонних індексів у часі. У таких умовах модель може демонструвати стабільніші результати порівняно з ARIMA-підходами, які зазвичай припускають незмінність сезонної структури.

Сезонні індекси та коефіцієнти сезонності

Для багатьох товарів сезонність є ключовим чинником. Один із простих методів її врахування - обчислення сезонних індексів.

Сезонний індекс показує, наскільки значення в певний місяць відрізняється від середнього.

Наприклад:

- індекс 1,25 означає, що в цей місяць продажі на 25% вищі за середні;
- індекс 0,85 означає падіння попиту на 15%.

Такі індекси широко використовують при ручному прогнозуванні або для попереднього аналізу даних.

Підсумки підрозділу

Класичні методи аналізу часових рядів - це базовий інструментарій, без якого неможливо побудувати коректну модель прогнозування продажів. Методи рухомого середнього, експоненціального згладжування, а також моделі Холта та Холта-Вінтерса забезпечують можливість:

- виявляти основні тенденції,
- оцінювати сезонні коливання,
- очищувати дані від шуму,
- виконувати первинні прогнози з високою інтерпретованістю.

Ці методи залишаються актуальними навіть у епоху машинного навчання, оскільки вони забезпечують прозорий аналіз та створюють надійну основу для порівняння зі складнішими алгоритмами.

1.4. Методи прогнозування динаміки продажів

Прогнозування попиту на товари або послуги допомагає підприємствам заздалегідь планувати свої дії: визначати необхідний обсяг запасів, розподіляти ресурси, коригувати маркетингову активність. Для багатьох компаній це не просто аналітичне завдання, а частина щоденного управління, адже точність прогнозу безпосередньо впливає на фінансові результати та ефективність операцій. На відміну від описових статистичних методів, що використовуються переважно для аналізу минулих тенденцій, методи прогнозування спрямовані на оцінку майбутніх значень показника на основі закономірностей, виявлених у часовому ряді.

Підходи до прогнозування можна умовно поділити на дві групи: **класичні статистичні моделі** та **методи машинного навчання**. У цьому підрозділі розглянуто ключові інструменти, що використовуються у прогнозуванні продажів, їх переваги, обмеження та особливості застосування.

1.4.1. Лінійна регресія як базовий інструмент прогнозування

Лінійна регресія є одним з найпростіших методів прогнозування, що встановлює залежність між часом (t) та значенням показника (Y) [8]. Модель має вигляд:

$$\hat{Y}_t = \bar{Y} + \left[\frac{\sum ((t - \bar{t})(Y_t - \bar{Y}))}{\sum ((t - \bar{t})^2)} \right] \cdot (t - \bar{t})$$

де

a - початковий рівень,

b - коефіцієнт нахилу (тренду).

Хоча лінійна регресія не враховує сезонність чи складні нелінійні залежності, вона залишається базовою моделлю, на яку орієнтуються при побудові більш складних моделей. На практиці лінійна регресія може бути корисною для базової оцінки трендової частини попиту, наприклад, для товарів зі стабільно зростаючими продажами.

1.4.2. Автокореляція та моделі на основі залежності від попередніх значень

Часові ряди продажів часто демонструють залежність між поточними та попередніми значеннями. Для опису такої залежності використовуються моделі:

AR (Autoregressive model)

Модель авторегресії описує значення показника як комбінацію кількох попередніх величин:

$$Y_t = c + \varphi^1 \cdot Y_{t-1} + \varphi^2 \cdot Y_{t-2} + \dots + \varphi_p \cdot Y_{t-p} + \varepsilon_t$$

MA (Moving Average model)

Модель ковзної середньої прогнозує значення на основі попередніх помилок прогнозу:

$$Y_t = c + \theta^1 \cdot \varepsilon_{t-1} + \theta^2 \cdot \varepsilon_{t-2} + \dots + \theta_q \cdot \varepsilon_{t-q} + \varepsilon_t$$

ARMA (Autoregressive Moving Average model)

Комбінація двох підходів:

$$Y_t = c + \Sigma(\varphi_i \cdot Y_{t-i}) + \Sigma(\theta_j \cdot \varepsilon_{t-j}) + \varepsilon_t$$

Ці моделі підходять для стаціонарних рядів, де немає вираженого тренду та сезонності.

1.4.3. Модель ARIMA: поєднання тренду та випадкових коливань

Модель ARIMA (Autoregressive Integrated Moving Average) [7] є однією з найбільш поширених статистичних моделей прогнозування часових рядів і широко використовується для аналізу продажів у випадках, коли дані демонструють трендову поведінку та випадкові коливання.

Основна ідея моделі полягає у поєднанні авторегресійної залежності між спостереженнями, механізму різницювання для усунення тренду та моделювання випадкових збурень за допомогою ковзної середньої.

Формально модель описується параметрами ARIMA(p , d , q), де значення p визначає кількість лагів у авторегресійній частині, d відповідає порядку різницювання, необхідного для досягнення стаціонарності ряду, а q задає глибину врахування попередніх похибок прогнозу. Така структура дозволяє ARIMA ефективно працювати з нестаціонарними часовими рядами, у яких тренд може бути усунений шляхом диференціювання.

ARIMA добре зарекомендувала себе для короткострокових прогнозів, коли важливо врахувати локальні зміни попиту та випадкові флуктуації. Завдяки механізму різницювання модель здатна адаптуватися до поступових змін рівня показника, що робить її придатною для аналізу агрегованих продажів без вираженої сезонної структури.

Разом із тим застосування ARIMA[3] супроводжується низкою обмежень. Модель вимагає попереднього приведення ряду до стаціонарного вигляду, що потребує додаткового аналізу та може ускладнювати побудову прогнозу. Крім того, ARIMA не враховує сезонні коливання без відповідної модифікації, а інтерпретація її параметрів може бути складною для

користувачів без глибокої підготовки у статистиці. У випадках, коли сезонність відіграє суттєву роль, доцільним є використання розширених моделей, зокрема SARIMA.

1.4.4. Сезонна модель SARIMA

Сезонна авторегресійна інтегрована модель ковзного середнього (SARIMA) [7] є розвитком класичної ARIMA та використовується у випадках, коли часовий ряд демонструє регулярні сезонні коливання. На практиці це характерно для даних про продажі, де попит повторюється з певною періодичністю - наприклад, у тижневому або річному циклі. На відміну від базової ARIMA, SARIMA дозволяє одночасно моделювати як короткострокові залежності між спостереженнями, так і довгострокову сезонну структуру ряду.

Практична цінність SARIMA проявляється у ситуаціях, коли дані є відносно стабільними за структурою та не містять різких зламів тренду. У таких умовах модель здатна відтворювати повторювані патерни попиту та забезпечувати прийнятну точність прогнозу навіть на декілька періодів уперед. Це робить SARIMA корисною для задач оперативного планування, зокрема при аналізі агрегованих продажів за тиждень або місяць, коли сезонний ефект є вираженим і передбачуваним.

Модель записують як:

$$\text{SARIMA}(p, d, q)(P, D, Q)m$$

де

m - довжина сезонного циклу (наприклад, 12 для місячних даних).

Разом з тим застосування SARIMA має низку обмежень, пов'язаних з вимогами до властивостей часових рядів. Модель є чутливою до різких структурних змін, викидів та нерегулярної поведінки попиту, що часто

спостерігається у роздрібній торгівлі під впливом акцій, змін асортименту або зовнішніх економічних факторів. Крім того, підбір параметрів моделі потребує достатньої кількості історичних даних і може супроводжуватися проблемами збіжності оптимізації, особливо у випадках коротких або шумних рядів.

1.4.5. Prophet - модель від Meta (Facebook)

Модель Prophet^[8] була розроблена як практичний інструмент для прогнозування часових рядів у бізнес-середовищі, де дані часто є нерегулярними, містять пропуски та піддаються впливу зовнішніх подій. На відміну від класичних статистичних моделей, Prophet орієнтований не стільки на точне відтворення автокореляційної структури ряду, скільки на швидке отримання зрозумілого та стабільного прогнозу з мінімальними вимогами до попередньої обробки даних.

Архітектурно Prophet базується на адитивному представленні часової серії, у якому тренд, сезонні компоненти та вплив календарних подій розглядаються окремо. Такий підхід дозволяє автоматично адаптувати модель до змін у поведінці ряду без складного підбору параметрів з боку користувача. Це робить Prophet зручним інструментом для аналітиків, які працюють з великим обсягом бізнес-даних та не мають можливості детально налаштовувати кожен компонент вручну.

У задачах прогнозування продажів Prophet часто застосовується як компромісне рішення між простими методами згладжування та складними моделями машинного навчання. Модель демонструє стійку поведінку на даних з нерівномірною сезонністю та добре масштабується на великі часові горизонти. Водночас у ситуаціях, де структура ряду є стабільною та добре описується статистичними залежностями, класичні моделі на кшталт ARIMA або SARIMA можуть забезпечувати вищу точність прогнозу. Крім того,

ефективність Prophet суттєво залежить від обсягу історичних даних, що може обмежувати його застосування для коротких часових рядів.

1.4.6. Методи машинного навчання у прогнозуванні продажів

Методи машинного навчання[\[11\]](#) застосовуються у випадках, коли динаміка попиту визначається не лише часовими закономірностями, а й значною кількістю зовнішніх факторів. До таких факторів належать ціна, маркетингові активності, сезонні події, погодні умови, регіональні особливості та інші змінні, взаємозв'язок між якими часто має нелінійний характер. На відміну від класичних статистичних моделей, підходи машинного навчання дозволяють одночасно враховувати десятки ознак і виявляти складні залежності без явного задання їх форми.

Одним із поширених підходів у прикладних задачах є використання ансамблевих методів, зокрема алгоритмів на основі дерев рішень. Такі моделі формують прогноз шляхом агрегування результатів багатьох незалежних моделей, що підвищує стійкість до шуму та окремих аномальних спостережень. Це дозволяє досягати прийнятної точності навіть за наявності пропусків у даних або нерівномірної структури ряду. Водночас подібні моделі мають обмежену інтерпретованість, оскільки кінцевий прогноз є результатом складної взаємодії великої кількості дерев.

Подальшим розвитком ансамблевих підходів є методи градієнтного підсилення, які ітеративно вдосконалюють модель, зосереджуючись на помилках попередніх прогнозів. Такі алгоритми часто демонструють високу точність, особливо при роботі з великими наборами даних і багатовимірними ознаковими просторами [14]. Разом із тим їх застосування потребує значних обчислювальних ресурсів і ретельного налаштування, оскільки за відсутності регуляризації існує ризик переобучення, особливо на обмежених вибірках.

Окрему групу методів становлять нейронні мережі для роботи з часовими рядами, зокрема архітектури типу LSTM, які спеціально розроблені

для аналізу послідовностей. Такі моделі здатні виявляти довгострокові залежності та складні патерни у поведінці попиту, що робить їх привабливими для дослідження комплексних процесів у продажах. Проте практичне використання LSTM вимагає значних обсягів історичних даних і складнішої реалізації, що обмежує їх застосування у типових бізнес-сценаріях.

1.4.7. Порівняння методів

Метод	Переваги	Недоліки	Найкраще застосування
Лінійна регресія	Проста, зрозуміла	Не враховує сезонність	Початковий прогноз
AR/MA/ARMA	Добре працює з короткими рядами	Вимагають стаціонарності	Дані без сезонності
ARIMA	Короткострокова точність	Погано працює з сезонністю	Стаціонарні ряди
SARIMA	Враховує сезонність	Багато параметрів	Місячні / квартальні продажі
Holt-Winters	Моделює сезонність і тренд	Потребує достатнього обсягу даних	Роздрібна торгівля
Prophet	Автоматизація, свята	Може бути менш точним	Бізнесові дані з нерегулярністю
Random Forest	Висока стійкість	Складно інтерпретувати	Багатофакторні дані
LSTM	Виявляє складні залежності	Потребує великих даних	Великі e-commerce ритейлери

(Таблиця 1.4.1 - Порівняння переваг та недолік моделей прогнозування)

Розглянуті методи демонструють широкий спектр підходів до прогнозування продажів - від простих лінійних моделей до складних алгоритмів, здатних враховувати багатфакторну динаміку попиту. Кожен підхід має власні сфери ефективного застосування, залежні від характеристик даних, наявності сезонності, обсягу історичних спостережень та вимог до точності прогнозу. У практиці бізнес-аналітики часто обирають не один метод, а комбінують декілька, щоб отримати більш надійну оцінку майбутніх значень. У подальших розділах ці підходи розглядатимуться вже з позиції їх реального застосування та адаптації до конкретного набору даних.

1.5. Оцінка якості прогнозних моделей

Коли модель прогнозування вже побудована, виникає питання, наскільки можна довіряти її результатам. Інколи навіть візуально «гарна» модель може помилятися в ключових моментах, і навпаки - проста модель показує стабільну точність. Тому перевірка прогнозу є не менш важливою, ніж сам процес моделювання. Саме вона дозволяє зрозуміти, чи відображає модель реальні закономірності в даних, чи просто повторює випадкові коливання.

Щоб оцінити, наскільки прогноз відповідає фактичній поведінці ряду, використовують набір кількісних метрик. Вони порівнюють прогнозовані значення зі справжніми та допомагають зрозуміти, у яких випадках модель помиляється, а також наскільки ці помилки суттєві для бізнесу.

1.5.1. Середня абсолютна похибка (MAE)

MAE показує, наскільки в середньому прогноз відхиляється від фактичних даних. Її значення легко інтерпретується, оскільки воно виражене в тих самих одиницях, що й початкові дані. Формула обчислення:

$$MAE = (1/n) \cdot \sum |Y_t - \hat{Y}_t|$$

Метрика добре підходить для ситуацій, де важливо зберігати помірний баланс між чутливістю до великих відхилень та загальною точністю. У практичному аналізі MAE часто використовують як первинний індикатор ефективності моделі.

1.5.2. Середня квадратична похибка (MSE) та корінь середньої квадратичної похибки (RMSE)

Інколи важливо не лише знати середній розмір помилки, а й звертати увагу на великі відхилення. Якщо модель пропускає гострий «провал» або різкий пік попиту, це може серйозно вплинути на планування запасів. Для таких випадків застосовують MSE:

$$MSE = (1/n) \cdot \sum (Y_t - \hat{Y}_t)^2$$

Оскільки квадрат робить великі помилки ще помітнішими, RMSE дозволяє повернути значення до початкового масштабу:

$$RMSE = \sqrt{MSE}$$

Обидві метрики краще підходять для порівняння моделей або для задач, де навіть поодинокі відхилення можуть стати критичними.

1.5.3. Середня абсолютна відносна похибка (MAPE)

MAPE вимірює точність у відсотках, що робить її однією з найзручніших метрик для презентацій та звітів. Формула виглядає так:

$$MAPE = (100/n) \cdot \sum |(Y_t - \hat{Y}_t)/Y_t|$$

Оскільки значення MAPE легко трактувати, метод широко використовують у роздрібному бізнесі. Але в нього є нюанс: якщо у даних є нульові або надто малі значення, відношення робить помилки непропорційно великими. Також MAPE інколи «карає» моделі за недопрогноз значно сильніше, ніж за перепрогноз.

1.5.4. Коефіцієнт детермінації (R^2)

Коефіцієнт детермінації часто використовують у регресійному аналізі. Він показує, яку частину варіації даних модель може пояснити. Формула:

$$R^2 = 1 - \left[\Sigma(Y_t - \hat{Y}_t)^2 \right] / \left[\Sigma(Y_t - \bar{Y})^2 \right]$$

У задачах прогнозування продажів цей показник застосовують обережно: сильні сезонні коливання та тренди можуть створювати хибне враження про високу або низьку якість моделі. Проте R^2 дає корисне уявлення про те, наскільки модель узгоджується з історичними даними.

1.5.5. Перевірка поведінки моделі: стабільність і здатність до узагальнення

Навіть точний прогноз не завжди гарантує, що модель буде працювати добре надалі. Щоб перевірити її «надійність», аналітики використовують кілька технік.

- **Перевірка на відкладеній вибірці**

Дані умовно ділять на «минуле» і «майбутнє». Модель тренується на першій частині, а перевіряється на другій. Такий підхід добре показує, як модель поводить себе на нових даних.

- **Cross-validation (ковзаюче вікно)**

Цей метод часто застосовується для сезонних рядів, дозволяючи моделі багаторазово перевіряти себе на різних фрагментах даних.

- **Аналіз залишків**

Якщо залишки (помилки) демонструють патерни - це означає, що модель щось «не вловила». Ефектна модель повинна залишати після себе випадковий шум, а не новий тренд чи сезонність.

1.5.6. Практичні критерії, важливі для бізнесу

Математичні метрики не завжди відображають реальну цінність прогнозу. Для підприємств важливими є:

- **Вартість помилки**

Не кожна помилка однаково небезпечна. Недопрогноз може означати втрату клієнтів, перепрогноз - фінансові втрати через надлишок товарів.

- **Інтерпретованість**

Модель повинна бути зрозумілою тим, хто буде на неї спиратися. Простіші моделі інколи переважають складні саме через доступність пояснення.

- **Стійкість до «нестандартних» періодів**

Свята, акції, форс-мажори - усе це деформує ряд. Модель повинна реагувати передбачувано.

- **Швидкість оновлення прогнозу**

Якщо модель використовується щодня, її швидкодія має велике значення.

Метрики, тестування на різних частинах даних і аналіз залишків дають змогу оцінити, як модель поводить себе у стабільні й нестабільні періоди, наскільки вона чутлива до коливань і чи може забезпечити надійний прогноз для майбутніх рішень. У комплексі це формує основу для вибору підходу, який буде найбільш доречним для конкретної бізнес-задачі.

1.6. Управління та підготовка даних для прогнозування продажів

Перш ніж будь-яка модель зможе формувати прогноз, дані потрібно впорядкувати та привести до стану, у якому вони дійсно відображають процеси, що відбуваються у бізнесі. На практиці саме якість вхідних даних

визначає, наскільки достовірним буде кінцевий прогноз. Навіть складна модель машинного навчання не дасть корисного результату, якщо дані неповні, містять помилки або суперечності. Тому етап підготовки, очищення та структурування даних є чи не найважливішим у всьому процесі прогнозування. У сфері продажів дані надходять з різних джерел: касові системи, CRM-платформи, ERP-модулі, онлайн-магазини, маркетингові сервіси. Кожне з цих джерел має свою структуру, періодичність оновлення та правила фіксації інформації. Об'єднання таких потоків в один узгоджений набір вимагає додаткової обробки, що дозволяє усунути розбіжності та підвищити коректність подальшого аналізу.

1.6.1. Проблеми якості даних у продажах

Підприємства регулярно стикаються з низкою труднощів, які можуть суттєво викривляти результати прогнозування.

● Пропуски у даних

Причини можуть бути різними: технічні збої, відсутність продажів, неузгоджена передача інформації. Пропущені значення зазвичай потрібно інтерполювати або позначати окремим маркером, якщо вони самі по собі несуть аналітичну цінність (наприклад, “товар був відсутній на складі”).

● Дублювання записів

Іноді транзакція потрапляє до системи кілька разів - це спотворює статистику. Такі записи необхідно виявляти через перевірку унікальних ідентифікаторів або часових відміток.

● Аномальні значення

Сюди належать різкі піки, падіння або підозріло великі суми продажів. Частина таких аномалій є природними (акції, сезонні сплески), а інші

свідчать про помилки. Важливо навчитися розрізняти їх, щоб не спотворити результати моделі.

- **Нестабільна періодичність**

У даних можуть бути нерівномірні інтервали - наприклад, продажі фіксуються не щодня, а лише коли відбулася транзакція. Для прогнозування такі ряди часто потрібно перетворювати у рівномірні (щоденні, тижневі або місячні).

- **Логістичні й маркетингові впливи**

Акції, розпродажі, зміна ціни або тимчасові обмеження постачання створюють додаткові коливання у продажах. Модель повинна або враховувати ці події, або мати очищені від них дані, залежно від типу задачі.

1.6.2. Підготовка даних: основні етапи

Щоб перетворити початкові дані на структурований і придатний для прогнозування набір, застосовують кілька послідовних етапів.

1. Збирання даних

Інформація надходить із різних систем, і важливо забезпечити її узгодження:

- однаковий формат дат;
- уніфіковані назви товарів;
- синхронізація часових зон;
- переведення цін і кількостей у єдину одиницю.

2. Очищення

На цьому етапі позбавляються дублікатів, виправляють помилки введення, заповнюють або інтерполюють пропуски.

У разі потреби аномальні значення зменшують (winsorizing) або замінюють контрольними.

3. Агрегація

У продажах часто працюють із даними різної granular-ності. Наприклад, є щоденні продажі, але прогноз формують на місяць. Агрегація дозволяє: отримати стабільніші ряди, зменшити шум, адаптувати частоту даних до вимог моделі.

4. Перетворення змінних

Іноді корисно створити нові ознаки:

- день тижня,
- місяць,
- індикатори свят,
- змінні, що описують акції.

Такі додаткові фактори допомагають моделі краще пояснювати коливання попиту.

5. Масштабування та нормалізація

Багато алгоритмів працюють точніше, коли дані приведені до одного масштабу. Це особливо актуально для методів машинного навчання, але може стати у пригоді і для класичних моделей, якщо ряд має різкі перепади.

1.6.3. Управління повнотою та актуальністю даних

Навіть добре очищений набір даних втрачає цінність, якщо він не оновлюється вчасно або містить значні прогалини. Для підтримки високої якості даних використовують:

- **Автоматизовані оновлення**

Регулярне завантаження нових продажів, оновлення залишків, інформації про акції.

- **Моніторинг “здоров’я даних”**

Системи контролю перевіряють:

- чи всі магазини передали дані за день;
- чи була незвична кількість транзакцій;
- чи співпадають записи між кількома джерелами.

- **Логування помилок**

Збереження історії збоїв дозволяє швидше знаходити та виправляти проблемні ділянки.

1.6.4. Перетворення часових рядів перед моделюванням

Після очищення дані часто потребують додаткових перетворень, що дозволяють точніше описати структуру ряду:

- **Різницювання**

Використовується, щоб прибрати тренд і зробити ряд стаціонарним (цей крок потрібен для ARIMA).

- **Декомпозиція**

Поділ ряду на тренд, сезонність і випадкову складову допомагає краще вивчити структуру попиту.

- **Згладжування**

Методи на кшталт рухомого середнього зменшують вплив шумів та аномалій.

- **Логарифмування**

Доречне, коли дані мають експоненціальне зростання або значні коливання.

1.6.5. Практичне значення правильної підготовки даних

Впорядковані дані зменшують навантаження на моделі та покращують стабільність прогнозів. У багатьох випадках саме очищення та попередня обробка підвищують точність сильніше, ніж заміна однієї моделі на іншу.

Для бізнесу це означає:

- менше помилок у прогнозах,
- менші втрати від надлишкових або недостатніх запасів,
- швидше ухвалення рішень,
- кращу узгодженість між відділами (логістика, закупівлі, маркетинг).

РОЗДІЛ 2. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧІ

2.1. Характеристика предметної області

Одним із центральних напрямів діяльності більшості підприємств, незалежно від масштабу чи галузі є управління продажами. Показники продажів відображають як результативність роботи компанії, так і реакцію споживачів на ринкові умови. Саме тому бізнес приділяє значну увагу аналізу цих даних: вони дозволяють оцінювати ефективність стратегій, оперативно реагувати на зміни попиту та приймати зважені управлінські рішення.

Компанії щоденно формують великі масиви даних про транзакції: кількість проданих товарів, виручку, повернення, роботу промоакцій, зміни цін. У сучасних умовах ці дані не обмежуються однією системою - вони надходять із касових терміналів, інтернет-магазинів, мобільних застосунків,

CRM-платформ та маркетплейсів. Наприклад, Amazon або Rozetka збирають інформацію у режимі реального часу, що дозволяє їм коригувати ціни та оптимізувати складські залишки протягом дня. Для великих міжнародних компаній, таких як ІКЕА чи Walmart[4], аналітика продажів формує основу логістичного планування і впливає на те, які товари буде вигідно закуповувати в наступні періоди.

Для більшості організацій дані про продажі виконують дві ключові функції. По-перше, вони відображають поточний стан бізнесу - тобто які товари продаються краще, які точки або регіони працюють ефективніше, які маркетингові дії приносять результат. По-друге, ці дані дозволяють прогнозувати майбутні показники та планувати необхідні ресурси. Компанія, яка може передбачити попит точніше за конкурентів, отримує очевидні переваги: менше «заморожених» складських запасів, швидший обіг товарів, точніше формування бюджету та гнучкіший підхід до роботи з клієнтами.

Особливість предметної області полягає в тому, що продажі залежать не лише від внутрішніх чинників (каталог товарів, ціни, сезонні акції), а й від зовнішніх. До них належать економічні коливання, поведінка конкурентів, погодні умови, культурні події, навіть інформаційний фон у медіа. Це означає, що часовий ряд продажів може містити як регулярні зміни (наприклад, сезонні піки), так і нерегулярні коливання, пов'язані з непередбачуваними подіями. Саме тому побудова прогнозу вимагає застосування методів, здатних обробляти такі різноманітні патерни.

У сфері роздрібної торгівлі продажі часто характеризуються чіткою сезонністю. Наприклад, для супермаркетів збут товарів зростає у святкові періоди, тоді як попит на окремі категорії може бути прив'язаний до погодних умов. Інтернет-магазини помічають різке збільшення кількості замовлень під час сезонних розпродажів, таких як Black Friday чи Cyber Monday. У виробничих компаніях циклічність може залежати від

контрагентів чи річних контрактів. Така різноманітність впливає на структуру даних і визначає вимоги до моделі прогнозування.

Ще один аспект полягає у поведінці товарного асортименту. У будь-якій компанії є позиції з високою стабільністю та прогнозованістю попиту, і є товари, продажі яких мають «вибуховий» або нерегулярний характер. Наприклад, електроніка часто демонструє трендові стрибки, пов'язані з виходом нових моделей, тоді як продукти харчування мають більш стабільні сезонні патерни. Деякі компанії, як-от Zara чи H&M, працюють у форматі швидкої моди, де життєвий цикл товару короткий і прогнозування вимагає особливої гнучкості - моделі повинні реагувати на швидко змінюваний попит.

Окремо варто згадати про роль цифровізації. Завдяки впровадженню ERP-, CRM- і BI-систем підприємства отримують можливість не лише зберігати дані, а й систематично їх аналізувати. Компанії, що реалізують омніканальні стратегії, поєднують дані офлайн і онлайн-каналів, створюючи цілісну картину взаємодії з клієнтами. Наприклад, мережі електроніки можуть зіставляти дані онлайн-переглядів товару з подальшими офлайн-покупками, що дозволяє точніше оцінювати реальний попит і вплив рекламних кампаній.

Предметна область охоплює широкий спектр бізнес-процесів: від реєстрації первинної транзакції та обліку змін у складських залишках до прийняття стратегічних рішень на основі прогнозів. Саме різноманітність джерел інформації, взаємозалежність факторів і швидкість змін створюють складність у моделюванні продажів. Це також пояснює, чому класичні методи чи інтуїтивні підходи часто не дають достатньої точності, а сучасні моделі прогнозування та алгоритми обробки даних стають необхідними інструментами для управління в умовах реальної ринкової динаміки.

2.2. Джерела даних та їх структура

Дані про продажі можуть надходити з різних систем, і кожна з них відображає свій фрагмент роботи підприємства. Тому перед аналізом важливо розуміти, звідки саме надходить інформація, у якому вигляді вона зберігається та як її об'єднати в один узгоджений набір.

Для більшості компаній ключові джерела включають касові системи, онлайн-платформи, складські модулі, CRM і маркетингові сервіси. Усі ці системи працюють незалежно одна від одної, але разом формують повну картину попиту та поведінки покупців.

Касові системи та POS-термінали

Це основне джерело транзакційних даних. Вони фіксують:

- дату та час продажу,
- кількість товару,
- ціну,
- суму операції,
- тип платежу.

Для аналізу важливим є не лише сам факт продажу, а й регулярність збору таких даних. У великих мережах (наприклад, McDonald's) POS-дані використовують для щоденного моніторингу попиту та планування змін персоналу.

Інтернет-магазини та маркетплейси

Онлайн-канали надають ширший набір даних, ніж звичайні касові системи. Сюди можуть входити:

- дані про перегляди товарів,

- додавання до кошика,
- покинуті замовлення,
- промокоди,
- інформація про клієнтів.

Такі дані особливо корисні для моделювання попиту в e-commerce, де поведінка користувачів змінюється швидше, ніж у класичній роздрібній торгівлі.

CRM-системи

CRM містить інформацію про взаємодію з клієнтом: контакти, історію покупок, реакції на акції, сегментацію аудиторії.

Ці дані не завжди використовують у базових моделях прогнозування, але вони важливі для більш глибоких підходів, наприклад, персоналізованих рекомендацій.

ERP та модулі управління складом

Складські дані відображають:

- залишки товарів,
- надходження партій,
- переміщення між складами та магазинами,
- терміни придатності (для продуктів харчування).

У деяких компаніях саме склад є «вузьким місцем», яке впливає на продажі. Наприклад, відсутність товару на полицях під час пікового попиту може зменшити фактичні продажі, і це потрібно враховувати в аналізі.

Маркетингові платформи та рекламні інструменти

Google Ads, Meta Ads, email-розсилки та інші канали надають інформацію про рекламні кампанії:

- бюджет,
- охоплення,
- переходи на сайт,
- вплив на продажі.

Зв'язок між маркетингом і попитом не завжди прямий, але в окремих випадках ці дані суттєво покращують точність прогнозу.

Структура даних для прогнозування

Незалежно від того, з яких систем надходить інформація, її потрібно звести до узгодженого формату. Найчастіше використовують таблицю, де кожен рядок описує продаж за певний день або місяць. У моделей прогнозування є певні вимоги до того, як мають виглядати дані:

• Однакова періодичність

Якщо у датах є пропуски, модель буде працювати гірше. Наприклад, для щоденних продажів мають бути всі дні, навіть якщо деякі з них мають нульові значення.

• Чітка прив'язка до часу

Дата має зберігатися у форматі, який легко перетворити у рік, місяць, день тижня тощо.

• Відсутність дублікатів

Особливо це стосується даних із різних каналів.

• Окремі колонки для цін, знижок і акцій

Такі фактори часто впливають на сезонність і тренд.

- **Уніфіковані назви товарів**

У різних системах один і той самий товар може мати різні позначення.

Використання відкритих даних

Оскільки у межах цієї роботи дані беруться з відкритих джерел, важливо, щоб вони відповідали формату, придатному для побудови часових рядів. Підходящими є набори з платформ: Kaggle, UCI Machine Learning Repository, Google Dataset Search[\[13\]](#)

Багато з них містять реальні історичні продажі магазинів, супермаркетів або окремих категорій товарів, що дозволяє побудувати приклади прогнозів і перевірити їх точність.

2.3. Основні проблеми аналізу продажів у бізнес-середовищі

Продажі на перший погляд здаються доволі простим показником: є товар, є дата продажу, є кількість. Проте в реальному бізнесі навіть базові дані можуть містити багато прихованих нюансів, які ускладнюють аналіз і сильно впливають на точність прогнозів. Чим більше каналів продажів і товарних позицій має підприємство, тим складнішим стає завдання побудови надійної моделі.

Нестабільність попиту

Попит на більшість товарів змінюється нерівномірно. Він залежить не лише від сезону чи маркетингової активності, а й від зовнішніх подій, які передбачити складно. Наприклад, різкі коливання продажів можуть виникати через погодні умови, інформаційний фон або навіть локальні події, що не пов'язані безпосередньо з діяльністю компанії. Це створює труднощі для моделей, яким потрібні стабільні й повторювані патерни.

Велика кількість факторів, що впливають на продажі

Іноді на кінцевий результат впливають десятки чинників: ціна, конкурентні акції, знижки, зміни асортименту, наявність товару, сезонність, соціальні тенденції. Якщо частина цих факторів не врахована в даних, модель може неправильно оцінювати закономірності. У таких випадках навіть просте трендові зростання може виявитися наслідком одноразової акції, а не реального підвищення попиту.

Різниця між «фактичним» і «потенційним» продажем

У продажах часто плутають попит і реалізацію. Дані у файлах або БД показують лише те, що було продано, але не завжди те, що могли б продати. Наприклад:

- товар міг бути відсутній на складі,
- сайт міг працювати з перебоями,
- касовий апарат міг не фіксувати частину операцій.

У таких випадках реальні продажі нижчі за попит, і модель прогнозуватиме неточні значення.

Перегрів даних у святкові та акційні періоди

Активні рекламні кампанії, знижки або великі розпродажі створюють короткочасні піки. Наприклад, у Black Friday або в період новорічних акцій продажі можуть зрости в кілька разів. Для більшості моделей такі різкі коливання перетворюються на шум, і без додаткової корекції вони можуть сильно зміщувати прогноз.

Нерівномірність і неповнота даних

Не всі компанії мають чіткий регламент збору інформації. Звідси виникають:

- пропущені дні чи місяці,

- різний формат дат,
- різні одиниці вимірювання,
- дублікати записів,
- помилки при імпорті з різних систем.

Через це дані потребують ретельної підготовки, перш ніж їх можна використовувати для моделювання.

Зміни в асортименті

Нові товари з'являються регулярно, а старі знімаються з продажу. Це створює специфічні проблеми: новий товар не має історії продажів; а отже, модель не може прогнозувати його точно; моделі доводиться працювати лише з аналогами або категоріями.

Таку проблему мають, наприклад, магазини техніки, де кожна нова модель смартфона або ноутбука має унікальні характеристики та власний життєвий цикл.

Гетерогенність даних з різних каналів

Продаж у магазині й продаж онлайн - це два різні процеси. У фізичних точках впливають логістика, викладка товарів, наявність персоналу[6]. Онлайн-магазин залежить від реклами, трафіку, зручності сайту, швидкості доставки. Злиття цих даних в одне ціле - складне завдання, і часто саме на цьому етапі виникають розбіжності.

Вплив зовнішніх факторів, які не фіксуються у даних

Багато чинників, що впливають на продажі, не присутні у доступних наборах даних:

- зміни законодавства,

- новини у ЗМІ,
- економічна ситуація,
- події конкурентів.

Без урахування цих факторів модель може виглядати недостатньо «адаптивною».

2.4. Постановка задачі прогнозування

Після попереднього аналізу даних і проблем предметної області стає зрозуміло, що для підприємства важливо не просто фіксувати продажі, а передбачати їх зміну в майбутньому. Для цього необхідно чітко сформулювати задачу, яку повинна розв'язувати модель прогнозування. Формулювання задачі дозволяє визначити, які саме дані потрібні, які методи підходять найкраще та як оцінювати успішність побудованої моделі.

У загальному вигляді прогнозування продажів зводиться до оцінки майбутніх значень показника на основі історичних даних. Проте на практиці завдання має більш конкретний зміст. Потрібно з'ясувати, яку саме величину прогнозуємо: кількість продажів, суму виручки, середній чек чи інший показник. Також важливо вирішити, для якого періоду будується прогноз – на тиждень, місяць, квартал або на більш довгий проміжок.

У контексті даної роботи задача формулюється як побудова прогнозу часових рядів продажів із урахуванням тренду та сезонних коливань. Це означає, що модель не повинна лише повторювати загальну тенденцію зростання або спадання, а й враховувати регулярні зміни, які повторюються з певною періодичністю. Такі зміни можуть бути пов'язані з днями тижня, святковими періодами, погодою, маркетинговими акціями або поведінкою покупців у конкретній категорії товарів.

Оскільки дані використовуються із відкритих джерел, модель має бути універсальною й здатною працювати з різними наборами даних, що можуть

відрізнятися структурою, частотою та характером сезонності. Це накладає вимоги на алгоритм, який повинен бути адаптивним і не залежати від особливостей конкретної компанії чи певного типу товару. Такий підхід дозволяє зробити систему більш гнучкою та придатною для подальшого розширення.

Задача також передбачає визначення горизонту прогнозу. З практичної точки зору прогноз на один-два періоди допомагає контролювати оперативні процеси, тоді як довший прогноз потрібен для планування бюджету та управління запасами. У цій роботі горизонт прогнозування вибирається таким, щоб він був достатньо чутливим до сезонності та демонстрував реальну динаміку ряду.

Ще одне важливе питання - оцінювання якості прогнозу. Оскільки мета полягає не лише в тому, щоб побудувати модель, а й зрозуміти, наскільки вона надійна, у рамках задачі необхідно визначити критерії, за якими порівнюватимуться результати. Для цього передбачено використання метрик похибки, які дозволяють оцінити відхилення прогнозу від фактичних значень та порівняти між собою різні методи.

Тобто, постановка задачі включає кілька основних складових:

- вибір показника, який прогнозується;
- визначення частоти та горизонту прогнозу;
- урахування трендових та сезонних компонентів у даних;
- вимоги до адаптивності моделі;
- формування критеріїв оцінки точності прогнозу.

У подальших розділах ці вимоги будуть використані для побудови алгоритму, що поєднує аналіз часових рядів і методи прогнозування, які найбільш природно працюють з трендом і сезонністю.

2.5. Обґрунтування необхідності розробки алгоритму

Потреба у побудові власного алгоритму прогнозування виникає там, де готові інструменти не дають достатньої гнучкості або не повністю враховують специфіку даних. Більшість бізнесів використовують стандартні засоби аналітики, такі як Power BI, Tableau чи CRM-платформи, але їхні можливості здебільшого обмежуються візуалізацією або простими статистичними моделями[4]. У випадках, коли продажі мають виражену сезонність, нерівномірність або залежать від додаткових факторів, базових методів часто недостатньо, і виникає потреба у розробці власного алгоритмічного рішення.

Ще один аспект - контроль над тим, як саме будується прогноз. Готові модулі прогнозування інколи приховують частину логіки або не дозволяють змінити внутрішні параметри. Це може бути незручно, коли хочеться порівняти декілька підходів або налаштувати модель відповідно до структури даних. У випадку виконання навчальної чи дослідницької роботи важливо мати можливість змінювати алгоритм, експериментувати з параметрами та бачити вплив кожної операції на результат.

Окрему увагу варто звернути на якість даних. Коли інформація надходить з відкритих джерел або з різних систем, вона потребує очищення, узгодження та перетворення перед тим, як її можна буде передати у модель. Багато готових інструментів або виконують ці операції частково, або вимагають ручного втручання. Власний алгоритм дозволяє автоматизувати ці кроки і зробити процес прогнозування більш стабільним.

Гнучкість також важлива з точки зору розширення функціональності. Якщо в майбутньому виникне потреба додати нові фактори, протестувати інші моделі або підключити систему до бази даних, це можна зробити в рамках власної реалізації без обмежень. Наприклад, модель, що спочатку використовувала лише тренд і сезонність, може бути доповнена інформацією

про акції, цінові зміни або поведінкові патерни - і такі зміни набагато простіше впроваджувати, коли алгоритм створений самостійно.

Не менш важливо й те, що розробка власного алгоритму забезпечує кращу інтерпретованість. Для бізнесу важливо не лише отримувати цифри, а й розуміти, чому модель прогнозує саме так. Універсальний алгоритм, створений у рамках цієї роботи, дозволяє прозоро пояснювати результат: як формується тренд, як впливає сезонність, які перетворення застосовуються до даних.

Створення такого алгоритму також дає можливість продемонструвати практичне застосування теоретичних методів, розглянутих у попередніх розділах. Це дозволяє поєднати математичну основу із реальними даними та показати, як саме обрані методи працюють у реальній аналітичній задачі.

РОЗДІЛ 3. РОЗРОБКА МЕТОДУ ТА ПРОГРАМНОЇ РЕАЛІЗАЦІЇ ПРОГНОЗУВАННЯ ПРОДАЖІВ

3.1. Архітектура алгоритму прогнозування продажів

Алгоритм прогнозування продажів складається з послідовності етапів, кожен з яких відповідає за певну частину обробки даних - від моменту їх завантаження до побудови готових прогнозів. Такий підхід дозволяє розділити процес на логічні модулі, спростити реалізацію програми та забезпечити можливість розширення системи в майбутньому. Загальна структура включає сім основних компонентів: завантаження даних, очищення та підготовку, формування часового ряду, попереднє моделювання за методом Holt-Winters, побудову прогностичної моделі ARIMA/SARIMA, оцінювання якості та візуалізацію результатів.

Основні модулі алгоритму

Алгоритм складається з таких модулів:

1. Модуль завантаження даних

Отримує файл, що містить історичні дані про продажі, та перетворює його у внутрішній формат для подальшої обробки.

2. Модуль очищення даних

Виявляє й усуває пропуски, дублікати, некоректні значення, приводить стовпець дат до стандартного формату та сортує дані за часом.

3. Модуль підготовки часового ряду

Приводить дані до потрібної частоти (денна чи місячна), формує цільовий ряд продажів та допоміжні колонки (наприклад, дата як індекс).

4. Модуль базового прогнозування (Holt–Winters)

Виконує первинну спробу моделювання, враховуючи тренд і сезонність. Мета - оцінити, наскільки добре класична модель здатна описати дані.

5. Модуль побудови моделі ARIMA/SARIMA

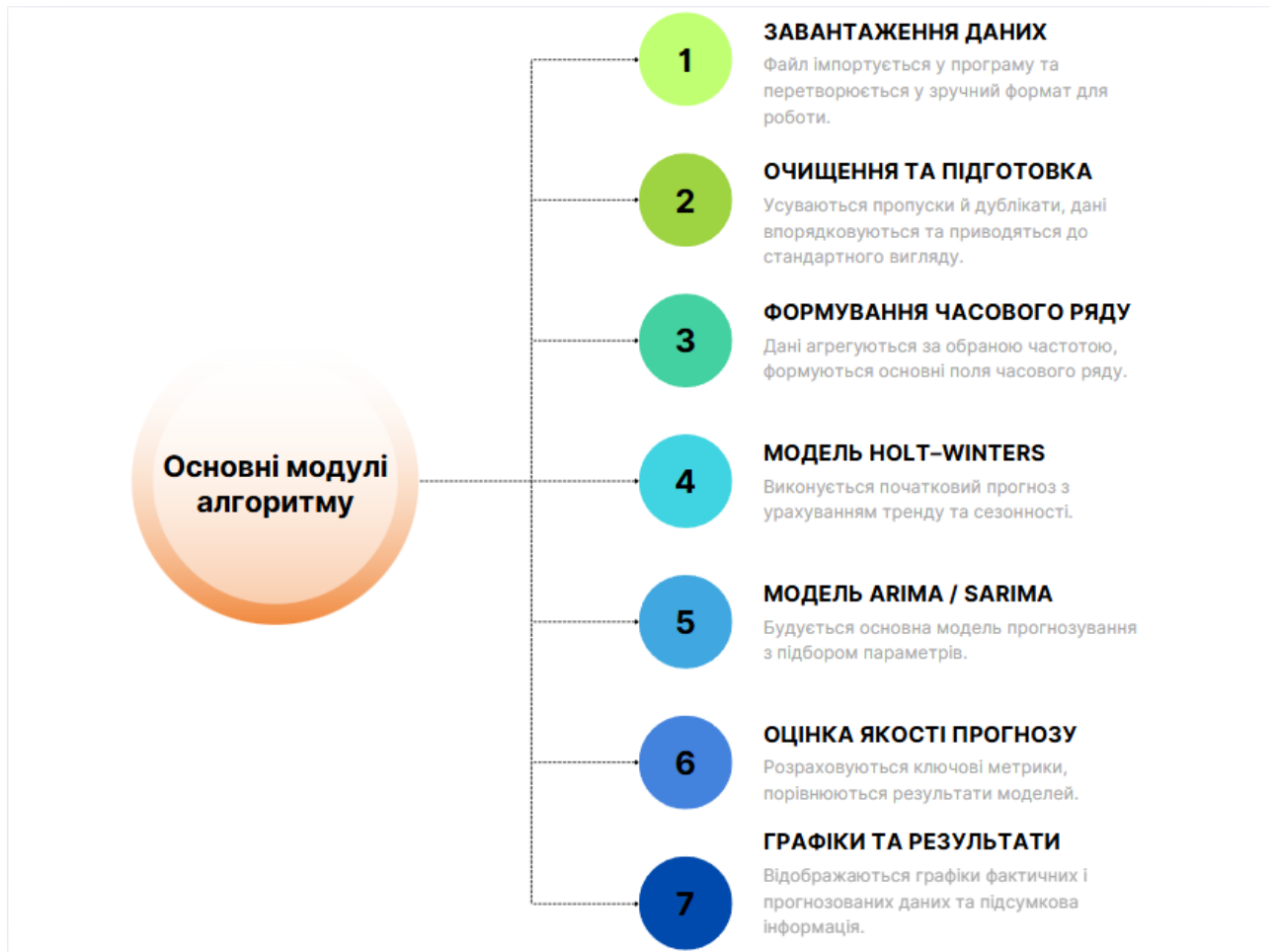
Створює більш точну модель, узгоджену з характеристиками часового ряду, підбирає параметри та формує фінальний прогноз.

6. Модуль оцінки якості

Розраховує MAE, MAPE, RMSE та інші метрики; порівнює результати моделей Holt–Winters і ARIMA/SARIMA.

7. Модуль візуалізації та виводу результатів

Будує графіки фактичних і прогнозованих значень, формує підсумкову таблицю метрик і виводить фінальні дані.



«Рисунок 3.1 - Блок-схема архітектури алгоритму прогнозування продажів»

3.2. Очищення та підготовка даних

Дані про продажі рідко бувають готовими до моделювання. Навіть у великих компаній, де збір інформації автоматизований, у масиві неминуче з'являються пропуски, неправильні значення, різні формати дат або нерівномірні інтервали між записами[2]. Якщо ці проблеми не усунути, будь-яка модель - навіть складна - працюватиме неправильно або нестабільно. Тому етап підготовки даних передбачає не лише “косметичні” правки, а й вирішення практичних задач, які безпосередньо впливають на якість прогнозу.

Однією з перших операцій є перевірка структури дат. На практиці це найпоширеніше джерело помилок. Наприклад, частина записів може мати

формат 2022/03/01, інші ж 01-03-2022, а деякі можуть містити помилки або бути записані текстом. Якщо не привести ці значення до єдиного стандарту, модель неправильно визначатиме послідовність подій. У реальному проєкті це може призвести до того, що модель “подумає”, ніби продажі у квітні передували продажам у березні.

Другий практичний момент - пропущені дні. У продажах така ситуація трапляється часто:

- магазин не працював;
- сайт був недоступний;
- дані не передалися у систему;
- товар був відсутній у продажу.

Якщо просто вилучити такі записи, часовий ряд перестає бути рівномірним, а багато моделей очікують саме рівномірну частоту. Тому пропущені дні зазвичай заповнюють нульовими значеннями або інтерполяцією. Наприклад, якщо магазин працює щодня, але у таблиці немає даних за певну дату, логічно припустити, що продажів не було зовсім, а не ігнорувати цей період.

Ще одна важлива задача - виявлення аномальних значень. У сфері роздрібно́ї торгівлі такі ситуації виникають постійно: великі святкові розпродажі, агресивні знижки, запуск рекламних кампаній, а також банальні помилки обліку, коли замість «45» потрапляє «450». Для побудови базового прогнозу такі піки краще згладити або позначити як окремий випадок. Наприклад, якщо у середньому щодня продається 100 одиниць товару, а одного дня - 1200, алгоритм Holt-Winters може “вирішити”, що така аномалія є частиною сезонності, і помилково підвищити прогноз.

Важливу роль відіграє й агрегація. Дані можуть надходити на рівні окремих транзакцій або годин, тоді як прогнозування зазвичай виконують у денному чи місячному масштабі. Якщо цього не зробити, модель отримає надто шумний ряд і не зможе визначити тренд. Наприклад, у продуктових магазинах кількість покупок протягом доби різко коливається, але для місячного прогнозу логічно використовувати сумарні або середні денні значення.

Дані можуть містити неочевидні проблеми. У практиці часто трапляється ситуація, коли продажів “немає”, але лише тому, що товар фізично відсутній на складі. У такому разі нуль означає не “немає попиту”, а “не змогли продати”. Якщо не врахувати такі моменти, модель показуватиме неправильний спад попиту. У деяких компаніях відсутність товару позначають окремим прапорцем, і це значно покращує очищення, але якщо таких даних немає - уже сам факт довгих “нулів” може бути сигналом для окремої обробки.

Після очищення виконується формування підсумкового часового ряду. Це означає, що кожному дню або кожному місяцю відповідає одне числове значення, і всі проміжки часу заповнені. Саме з такого ряду моделі прогнозування можуть коректно виявляти тренд і сезонність. Якщо ж у початкових даних залишились пропуски, великі “стрибки”, або дні переплутані, прогноз буде зміщеним або нестабільним.

Практичний досвід показує, що ретельна підготовка даних може підвищити точність прогнозу на 20-40% навіть без зміни моделі. Тому очищення та структурування є необхідним етапом, без якого подальші методи прогнозування, працюватимуть некоректно. Ми ж будемо практикуватися саме на підготовлених даних. Буде проведено моделювання, щоб оцінити поведінку різних підходів та визначити, яка модель найкраще описує сезонність і тренд.

A B		id,date,store_nbr,family,onpromotion						
1	date,sales	1	A	B	C	D	E	F
2	2013-01-01,0.0	2	3000888,2017-08-16,1,AUTOMOTIVE,0					
3	2013-01-02,579.0	3	3000889,2017-08-16,1,BABY CARE,0					
4	2013-01-03,453.0	4	3000890,2017-08-16,1,BEAUTY,2					
5	2013-01-04,460.0	5	3000891,2017-08-16,1,BEVERAGES,20					
6	2013-01-05,464.0	6	3000892,2017-08-16,1,BOOKS,0					
7	2013-01-06,212.0	7	3000893,2017-08-16,1,BREAD/BAKERY,12					
8	2013-01-07,400.0	8	3000894,2017-08-16,1,CELEBRATION,0					
9	2013-01-08,384.0	9	3000895,2017-08-16,1,CLEANING,25					
10	2013-01-09,530.0	10	3000896,2017-08-16,1,DAIRY,45					
11	2013-01-10,345.0	11	3000897,2017-08-16,1,DELI,18					
12	2013-01-11,401.0	12	3000898,2017-08-16,1,EGGS,1					
13	2013-01-12,404.0	13	3000899,2017-08-16,1,FROZEN FOODS,1					
14	2013-01-13,194.0	14	3000900,2017-08-16,1,GROCERY I,64					
15	2013-01-14,411.0	15	3000901,2017-08-16,1,GROCERY II,0					
16	2013-01-15,357.0	16	3000902,2017-08-16,1,HARDWARE,0					
17	2013-01-16,519.0	17	3000903,2017-08-16,1,HOME AND KITCHEN I,2					
18	2013-01-17,373.0	18	3000904,2017-08-16,1,HOME AND KITCHEN II,6					
19	2013-01-18,464.0	19	3000905,2017-08-16,1,HOME APPLIANCES,0					
20	2013-01-19,447.0	20	3000906,2017-08-16,1,HOME CARE,10					
21	2013-01-20,177.0	21	3000907,2017-08-16,1,LADIESWEAR,0					
22	2013-01-21,399.0	22	3000908,2017-08-16,1,LAWN AND GARDEN,16					
23	2013-01-22,347.0	23	3000909,2017-08-16,1,LINGERIE,5					
24	2013-01-23,496.0	24	3000910,2017-08-16,1,"LIQUOR,WINE,BEER",9					
25	2013-01-24,359.0	25	3000911,2017-08-16,1,MAGAZINES,0					
26	2013-01-25,399.0	26	3000912,2017-08-16,1,MEATS,0					
27	2013-01-26,440.0	27	3000913,2017-08-16,1,PERSONAL CARE,18					
28	2013-01-27,192.0	28	3000914,2017-08-16,1,PET SUPPLIES,0					
29	2013-01-28,405.0	29	3000915,2017-08-16,1,PLAYERS AND ELECTRONICS,0					
30	2013-01-29,327.0	30	3000916,2017-08-16,1,POULTRY,0					
31	2013-01-30,486.0							

«Рисунок 3.2 - Фрагмент вхідних даних, що використовуються для прогнозування»

3.3. Модель Holt-Winters: принципи роботи, можливості та практичні обмеження

Перш ніж застосовувати більш складні моделі часових рядів, доцільно перевірити, наскільки дані можуть бути описані класичним методом експоненціального згладжування Holt-Winters. Цей підхід добре зарекомендував себе у задачах, де на поведінку показника суттєво впливають тренд і сезонність. Багато компаній у роздрібній торгівлі, логістиці та сервісах використовують Holt-Winters як базовий інструмент, оскільки модель забезпечує швидкий результат, зрозумілі параметри та інтерпретованість прогнозу.

Щоб зрозуміти, як модель Holt-Winters реагує на нерегулярні коливання, можна розглянути простий умовний приклад. Припустимо, що магазин продає в середньому **100 одиниць товару на день**, а сезонність пов'язана з вихідними: у суботу попит зростає до **140**, а в неділю - падає до

80. Така поведінка регулярна й передбачувана, тому модель швидко виявляє цикли та будує адекватний прогноз.

Тепер уявімо, що в один із днів відбувається рекламна акція, і продажі раптово підскакують до **350 одиниць**, що більш ніж утричі перевищує звичайний рівень. Для моделі Holt-Winters це виглядає як потенційний сезонний сплеск. Якщо подібний пік трапився лише один раз, але модель не знає про це, вона «вирівнює» його у сезонну компоненту і надалі прогнозує, що в аналогічний день наступного циклу попит знову буде близьким до **300-350 одиниць**, хоча насправді цей стрибок був не сезонним, а одноразовим.

Аналогічна проблема виникає і з провалами. Наприклад, якщо в певний день дані фіксують лише **5 продажів** замість звичайних **100**, причина може бути зовсім не в падінні попиту - товар міг бути тимчасово відсутній на складі або виникла технічна помилка при передачі даних. Однак Holt-Winters трактує це як реальне зниження попиту та починає «тягнути» сезонність і тренд вниз. У підсумку прогноз стає заниженим ще декілька періодів після інциденту.

Такі приклади показують, що модель Holt-Winters добре працює лише тоді, коли дані містять регулярні повторювані структури. Якщо ж ринок нестабільний, попит формують акції, погодні зміни або зовнішні події, модель починає сприймати аномалії як частину сезонності й формувати хибні прогнози. Саме тому Holt-Winters ефективний як перший індикатор «наскільки ряд прогнозований», і чи варто застосовувати складніші підходи, такі як ARIMA або SARIMA, які мають більшу гнучкість у роботі з нерівномірними даними.

Структура моделі та роль компонентів

Модель Holt-Winters ґрунтується на припущенні, що динаміка продажів формується під впливом кількох базових складових, які взаємодіють між собою в часі. Передусім йдеться про поточний рівень показника, який

відображає загальний обсяг продажів у конкретний момент, а також про трендову компоненту, що характеризує напрям і швидкість зміни цього рівня. Доповнює модель сезонна складова, яка описує регулярні коливання попиту, пов'язані з повторюваними часовими циклами.

У бізнес-даних ці компоненти майже ніколи не існують ізольовано. Навіть для товарів зі стабільним щоденним попитом, таких як базові продукти харчування, можна спостерігати помірні коливання протягом тижня, зокрема підвищення продажів у вихідні дні. Для товарів тривалого користування або одягу сезонність, як правило, є значно вираженішою та пов'язана зі зміною пори року, оновленням асортименту або споживчих уподобань. У сфері електроніки динаміка попиту часто додатково визначається зовнішніми подіями, такими як вихід нових моделей продукції або періоди підвищеної купівельної активності наприкінці року.

Перевага методу Holt-Winters полягає в тому, що він дозволяє не лише фіксувати наявність тренду та сезонності, а й адаптувати ці компоненти до змін загальної динаміки бізнесу. Наприклад, у разі поступового розширення компанії зростає базовий рівень продажів, і водночас можуть змінюватися амплітуди сезонних коливань. Завдяки такому підходу модель здатна відображати реальну поведінку попиту більш гнучко, ніж методи, у яких сезонна структура вважається сталою.

Адитивна й мультиплікативна сезонність на практиці

Вибір типу сезонності - один із найважливіших моментів. Адитивна сезонність передбачає, що сезонні коливання мають сталу величину, тоді як мультиплікативна означає, що вони змінюються разом із масштабом ряду.

Приклад адитивної сезонності

Невеликий магазин пекарні продає в середньому 200 одиниць продукції щодня. У суботу попит підвищується приблизно на 40 одиниць, а у будні - на

10-15. При цьому загальна кількість продажів змінюється незначно протягом року. У такому випадку підходить адитивна модель - сезонні ефекти стабільні незалежно від загального рівня продажів.

Приклад мультиплікативної сезонності

Інтернет-магазин техніки протягом року продає різні категорії товарів. У листопаді під час Black Friday обсяги продажів зростають у 3–5 разів. Аналогічно, у грудні перед святами попит збільшується пропорційно до тренду: якщо у 2020 році магазин продав 10 000 одиниць товару, то у 2023 - уже 25 000. У таких випадках сезонний ефект «масштабується» разом із загальним рівнем продажів, тому потрібна мультиплікативна модель. У роздрібній торгівлі мультиплікативний варіант зустрічається частіше, оскільки ринки ростуть, а разом з ними і сила сезонних коливань.

Ситуації, у яких метод Holt-Winters демонструє найкращі результати

Метод Holt-Winters є найбільш ефективним у випадках, коли динаміка продажів характеризується регулярністю та повторюваними патернами. У практиці роздрібної торгівлі це часто проявляється у вигляді стабільних циклів, пов'язаних із поведінкою споживачів протягом тижня або року. Наприклад, у мережах супермаркетів спостерігаються типові щотижневі коливання: зростання продажів наприкінці робочого тижня та їх зниження у неділю. За наявності таких закономірностей модель швидко адаптується до структури даних і формує стабільний прогноз.

Хороші результати Holt-Winters також демонструє на даних без різких і нерегулярних стрибків попиту. У невеликих магазинах з відносно однотипним асортиментом, таких як аптеки або магазини побутової хімії, продажі змінюються поступово і не залежать від масштабних промо-кампаній. За відсутності непередбачуваних піків модель забезпечує достатню точність прогнозу навіть на кілька періодів уперед.

Окрему категорію становлять сезонні товари з передбачуваними циклами попиту. Продажі кондиціонерів, обігрівачів або іншої кліматичної техніки змінюються протягом року досить закономірно: у теплий або холодний сезон попит зростає, а в міжсезоння - зменшується. У таких умовах Holt-Winters здатний коректно відтворювати сезонну структуру та прогнозувати майбутні значення. У ряді практичних випадків якість такого прогнозу може бути співставною з результатами ARIMA або SARIMA, але за значно меншої складності реалізації.

Обмеження моделі: проблеми, що виникають у реальних даних

Незважаючи на переваги, застосування Holt-Winters у реальних бізнес-даних часто супроводжується низкою обмежень. Однією з найпоширеніших проблем є наявність різких аномалій.

Наприклад, якщо мережа магазинів побутової техніки запускає короткострокову акцію зі знижками до 50 %, продажі можуть зрости у кілька разів. Модель Holt-Winters у такій ситуації схильна інтерпретувати цей сплеск як зміну сезонної структури або нового рівня попиту, що призводить до завищених прогнозів у наступних періодах, хоча на практиці подібна акція має разовий характер.

Ще однією типовою проблемою є велика кількість нульових значень у даних. Це характерно для нових товарів або позицій, які часто відсутні на складі. Якщо, наприклад, товар був недоступний протягом більшої частини місяця, а продажі фіксувалися лише у декілька днів, модель сприймає нульові значення як відсутність попиту. Насправді ж попит міг існувати, але не був реалізований, що призводить до викривлення часової серії та зниження якості прогнозу. Критичним фактором Також критичним фактором є нестача повних сезонних циклів. Для коректного виділення сезонності необхідно мати щонайменше два або три повторення циклу. Якщо дані зібрані лише за один рік, як це часто трапляється у бізнесах із яскраво вираженою сезонністю

(наприклад, продаж туристичного спорядження), модель не може впевнено відрізнити випадкові коливання від справжніх циклічних закономірностей.

Окрему складність становлять нерівномірні часові інтервали. В онлайн-торгівлі продажі фіксуються лише у момент завершення транзакції, і за відсутності замовлень або через технічні збої в даних можуть виникати пропуски. Holt-Winters є чутливим до такої нерівномірності, і наявність «дірок» у часовому ряді може суттєво спотворювати підсумковий прогноз.

Типові помилки при початковому використанні Holt–Winters

Зниження якості прогнозу часто пов'язане не лише з властивостями даних, а й з помилками налаштування моделі. Однією з них є автоматичний вибір типу сезонності без попереднього аналізу. Деякі програмні інструменти за замовчуванням застосовують адитивну сезонність, навіть якщо амплітуда коливань зростає разом із рівнем продажів. У таких випадках мультиплікативна модель була б більш адекватною, а неправильний вибір призводить до заниження сезонних ефектів.

Не менш важливим є коректне визначення періоду сезонності. У продуктовому ритейлі часто використовується семиденний цикл, тоді як у промислових компаніях - річний цикл з періодом у дванадцять місяців. Водночас у реальних умовах сезонність може повторюватися і через інші проміжки часу, наприклад у вигляді квартальних продажів, циклічних виробничих замовлень або регулярних акційних тижнів. Помилка у виборі періоду практично гарантує некоректний прогноз. Ще одним поширеним недоліком є відсутність попереднього згладжування даних. Різкі викиди або поодинокі піки можуть призвести до втрати стабільності моделі, внаслідок чого сезонна компонента починає поводитися хаотично і погіршує результати прогнозування.

Роль Holt-Winters у межах даного дослідження

У рамках цієї магістерської роботи модель Holt-Winters не розглядається як фінальний інструмент прогнозування, однак її використання має принципове значення. По-перше, вона забезпечує швидку оцінку структури даних і дозволяє визначити, наскільки чітко у часовому ряді виражені тренд і сезонність. По-друге, результати прогнозу дають змогу оцінити загальну прогнозованість ряду: низька якість прогнозу Holt-Winters сигналізує про необхідність більш ретельного налаштування моделей ARIMA або SARIMA. По-третє, ця модель слугує базовим орієнтиром для порівняння зі складнішими підходами, що буде продемонстровано у практичній частині роботи.

Крім того, прогноз Holt-Winters часто дозволяє виявити аномалії в даних ще до побудови складних моделей. Різкі піки, пропуски або нетипові зміни рівня продажів стають помітними вже на етапі первинного аналізу.

3.4. Моделі ARIMA та SARIMA: підхід, можливості та практичні аспекти застосування

Після попереднього аналізу та тестового прогнозування методом Holt-Winters стає зрозуміло, що не всі часові ряди мають стабільну сезонність або рівний тренд. У реальних бізнес-даних часто трапляються ситуації, коли попит змінюється нерегулярно, присутні довгострокові тенденції, автокореляція або структурні зміни. Для таких випадків класична модель згладжування недостатня, і потрібні більш гнучкі інструменти. Саме такими є моделі ARIMA та їх сезонне розширення SARIMA.

Ці моделі широко застосовуються в економіці, фінансах, енергетиці та ритейлі. Вони здатні враховувати залежність поточних значень ряду від попередніх періодів, а також працювати з даними, що потребують диференціювання для усунення нестабільності[3]. На відміну від Holt-Winters, який очікує чіткої сезонності, моделі ARIMA та SARIMA здатні

адаптуватися до складної динаміки та будувати прогнози там, де сезонні та трендові ефекти непередбачувані або значною мірою порушені аномаліями.

3.4.1. Логіка та компоненти моделі ARIMA

ARIMA - це аббревіатура для трьох компонентів моделі:

- **AR (autoregressive)** - автогресійна частина, що враховує залежність значення показника від попередніх спостережень;
- **I (integrated)** - диференціювання, яке дозволяє зробити ряд стаціонарним, видаливши тренд;
- **MA (moving average)** - компонент ковзного середнього, що враховує вплив попередніх похибок прогнозу.

Комбінація цих елементів дає можливість моделі працювати з рядами, які не мають чітко виражених сезонних циклів, або навпаки - мають надто складну структуру, щоб Holt-Winters зміг її стабільно описати.

Практично це означає, що ARIMA добре підходить для даних, де:

- попит змінюється хвилеподібно, але без чіткої повторюваної сезонності;
- є автокореляція, тобто значення сьогодні частково залежить від значень попередніх днів;
- присутні тренди або зміни рівнів;
- дані містять випадкові “стрибки” без чіткої регулярності.

Приклад із практики: Мережа з продажу електроніки помічає, що попит на ноутбуки зростає перед початком навчального року, але сила цього ефекту щороку різна. Іноді промоакції суттєво змінюють попит, інколи підвищення ціни конкурентів збільшує продажі. У таких умовах сезонність є, але її форма та масштаб нестабільні. ARIMA здатна відфільтрувати цю

нерівномірність та побудувати прогноз на основі внутрішніх закономірностей ряду.

3.4.2. Що таке нерегулярність у трендах і як з цим працює ARIMA

Перш ніж застосовувати ARIMA, ряд зазвичай перевіряють на **стаціонарність**. Якщо середнє та дисперсія значень змінюються з часом, модель працюватиме некоректно. У реальних продажах це дуже поширене явище. Наприклад:

- у першому півріччі компанія продає приблизно 200 одиниць товару на місяць;
- у другому - відкриває додаткову точку продажу, і попит зростає до 350;
- через рік компанія закриває нерентабельний напрям і повертається до рівня 250.

Такі зміни не є сезонністю, це структурні зрушення. Holt-Winters їх сприймає як порушення сезонності, а ARIMA - як зміну тренду, яку можна компенсувати диференціюванням.

Наприклад:

Місяць:	1	2	3	4	5	6
Продажі:	100	110	120	200	210	220

«Таблиця 3.4.1 – Приклад структурних розривів у даних»

Тут явний “перелом” тренду між місяцями 3 і 4.

Holt-Winters вбудує це у сезонність → прогноз розірветься. ARIMA, після диференціювання, побачить стабільну різницю:

10, 10, 80, 10, 10

і зробить прогноз на основі стаціонарного процесу.

3.4.3. Чому ARIMA краще справляється з аномаліями, ніж Holt-Winters

Уявімо компанію, яка продає спортивні товари. Середній тижневий попит становить **300 одиниць**, але в період розпродажів він раптово зростає до **1200**.

У моделі Holt–Winters:

- такий пік буде обов’язково враховано як частина сезонності;
- прогноз на подібний період наступного року буде завищено.

У моделі ARIMA:

- екстремальний пік вплине на похибку (МА-компонент),
- але не змінить сезонність, якщо вона не повторюється регулярними циклами.

У результаті ARIMA «поглине» аномалію частково, а не «навчиться» на ній.

Що це дає на практиці: Прогноз стає більш стійким до рекламних акцій, пандемій, погодних стрибків, логістичних збоїв, тощо.

3.4.4. SARIMA: коли сезонність є, але вона складна

Сезонне розширення ARIMA - модель SARIMA - працює тоді, коли в даних присутні сезонні цикли, але вони:

- довгі (12 місяців);
- нерегулярні за амплітудою;
- змішані з трендами та зсувами рівня;

- мають «підсезонності» (наприклад, попит на морозиво має і річну сезонність, і денну, якщо дивитися позмінно).

SARIMA додає окремий набір параметрів (P, D, Q, s), де **s** - довжина сезонного циклу.

Приклад:

Супермаркет фіксує продажі мінеральної води за днями. Попит має:

- річну сезонність (влітку - високий, узимку - низький),
- тижневу сезонність (у вихідні зростає).

Для таких даних Holt-Winters часто не справляється, бо має лише одну сезонну компоненту. SARIMA дозволяє врахувати складну багаторівневу структуру.

3.4.5. Як виглядає автокореляція в реальних даних

Перед побудовою SARIMA зазвичай аналізують графіки ACF (autocorrelation function) та PACF (partial autocorrelation). Це дозволяє побачити, на які минулі значення реагує даний ряд.

У реальному бізнесі часто трапляються такі приклади:

Приклад 1: Попит залежить від учора

У магазині доставки продуктів у будні продажі стабільні. Люди замовляють приблизно однакові набори товарів. Якщо сьогодні попит був високим - завтра він теж з великою ймовірністю буде вищим. AR-компонента моделі це вловлює.

Приклад 2: Похибка в певні дні повторюється

У дні акцій модель може недооцінювати попит. Тоді MA-компонента вловлює закономірність: “у такі дні похибка зростає”.

Приклад 3: Вихідні впливають на будні

У фітнес-центрах продажі абонементів високі на початку місяця і низькі в кінці. Цикли чіткі, але не рівномірні - SARIMA це врахує.

3.4.6. Приклад дії ARIMA на числовому ряді

Розглянемо умовний приклад:

День:	1	2	3	4	5	6	7	8
Продажі:	100	105	108	115	117	118	300	120

«Таблиця 3.4.2 – Приклад аномальних стрибків показників у даних»

На 7-й день є аномалія - 300 продажів.

Що робить Holt-Winters:

- сприймає 300 як частину сезонності;
- прогноз на наступний аналогічний період може бути ~250-280;
- тренд буде піднято.

Що робить ARIMA:

- диференціює ряд $\rightarrow 5, 3, 7, 2, 1, 182, -180$
- бачить, що лише одне значення сильно вибивається;
- похибка "розходиться", але потім знову стабілізується;
- сезонність не спотворюється.

Результат: прогноз ближчий до реальної поведінки попиту.

3.4.7. Переваги використання SARIMA у продажах

1. Гнучкість у роботі з нерівномірними рядами

Модель не потребує ідеальної сезонності.

2. Стійкість до аномальних значень

Піки акцій або технічні “провали” не спотворюють прогноз так сильно.

3. Можливість врахування трендів різних типів

Лінійних, експоненціальних, змішаних.

4. Підтримка сезонності будь-якої довжини

ARIMA → без сезонності

SARIMA → з сезонністю будь-якого масштабу

Практично доведена ефективність у багатьох галузях

- енергетика (споживання електроенергії);
- транспорт (потік пасажирів);
- логістика (попит на складі);
- фінанси (цінові ряди);
- рітейл (продажі при коливаннях маркетингової активності).

3.4.8. Обмеження ARIMA/SARIMA

Не зважаючи на силу моделі, є й обмеження:

- потребує стаціонарності, отже - диференціювання;
- моделі складно пояснювати менеджерам, які не працюють зі статистикою;
- підбір параметрів може бути непростим;
- при дуже великій кількості аномалій модель втрачає стабільність;
- при швидкій зміні структури ринку модель не встигає адаптуватися.

Ці обмеження роблять ARIMA/SARIMA інструментом не «універсальним», а «правильним для стабільних і якісно підготовлених даних».

3.4.9. Причина вибору ARIMA/SARIMA у роботі

У рамках цього дослідження SARIMA є основною моделлю прогнозування, оскільки:

- дані можуть містити нерегулярні піки та провали;
- сезонність не завжди рівномірна;
- Holt-Winters використовується лише як первинна модель;
- SARIMA забезпечує точніший прогноз та гнучкість;
- модель дозволяє враховувати як тренд, так і шум;
- ARIMA дає можливість більш реалістично оцінити поведінку попиту.

3.5. Обґрунтування вибору моделей та структурного підходу до прогнозування

Побудова ефективної системи прогнозування продажів потребує вибору такого підходу, який буде не тільки достатньо точним, але й придатним до різних типів даних. На практиці рідко існують «ідеальні» набори даних: попит змінюється нерівномірно, сезонність порушується, з'являються акції, зовнішні події або логістичні затримки. Тому вибір єдиної моделі для всієї задачі часто виявляється недостатнім. Саме з цієї причини у роботі застосовується комбінація двох підходів Holt-Winters (HW) як базовий рівень і SARIMA як фінальна модель.

Holt-Winters як етап первинного аналізу та перевірки структури даних

Метод Holt-Winters використовується в роботі не як основний інструмент прогнозування, а як перевірка того, наскільки дані взагалі піддаються класичному опису. У реальних умовах бізнесу саме цей метод часто застосовують першим: він швидкий, дозволяє побачити базові патерни та одразу виявити ключові проблеми.

Що саме виявляє Holt-Winters:

- чи є регулярна сезонність;
- чи стабільна амплітуда коливань;
- як сильно впливають аномалії на модель;
- чи присутній тренд, і який він за напрямком;
- чи достатньо довгими є цикли.

Таким чином, Holt-Winters у системі виконує роль **діагностичного етапу**, дозволяючи зрозуміти, наскільки дані прогнозовані та які проблеми необхідно враховувати при подальшому моделюванні.

ARIMA/SARIMA як основа прогнозної моделі

Моделі ARIMA та SARIMA були обрані як ключові для фінального прогнозування, оскільки вони значно гнучкіші у роботі з реальними даними. На відміну від Holt-Winters, який розрахований на стабільну сезонність, ці моделі дозволяють:

- працювати з нерівномірною поведінкою попиту;
- враховувати автокореляцію;
- усувати трендову та структурну нестабільність через диференціювання;
- адаптуватися до рівнів сезонності різної довжини;

- компенсувати вплив аномалій без значного викривлення прогнозу.

Приклад з практики:

Компанія, що продає побутові товари, помітила, що влітку продажі вентиляційних систем різко зростають, але сила цього зростання залежить від погоди конкретного року. У спекотні роки попит збільшується у 2-3 рази, у помірні - лише на 20–30%. Holt-Winters у такій ситуації часто переоцінює сезонність, припускаючи, що кожного року буде однакове зростання. Модель SARIMA, навпаки, дозволяє відокремити короточасні сезонні ефекти від загальної динаміки й зменшити ризик завищених прогнозів.

ARIMA/SARIMA - це логічний вибір для даних із вираженою, але складною сезонністю, а також у випадках, коли поведінка ряду визначається кількома чинниками.

Однією з ключових ідей системи є не використання лише однієї моделі, а порівняння декількох підходів та вибір найточнішого. У аналітиці рідко довіряють одному алгоритму - найчастіше будують кілька прогнозів і оцінюють якість кожного з них. Це дозволяє: зменшити ризик помилкових рішень, якщо одна модель не впоралася з певною структурою даних; отримати більш надійний результат при невеликій кількості спостережень; зрозуміти, які характеристики ряду критично впливають на прогноз; довести доцільність вибору методів у дослідженні.

У даній роботі таке поєднання також допоможе продемонструвати, чому базові моделі (Holt-Winters) не завжди придатні, і чому ARIMA/SARIMA є більш коректним вибором для аналізу реальних продажів.

Аргументація з точки зору програмної реалізації

З технічного боку, обраний підхід дозволяє:

- легко реалізувати поетапну логіку;
- перевіряти результати кожної моделі;
- будувати графіки на кожному кроці;
- тестувати поведінку програмного коду на різних наборах даних;
- швидко змінювати параметри моделі ARIMA/SARIMA, не торкаючись інших компонентів;
- у майбутньому додати моделі машинного навчання (Prophet, LSTM).

Фактично ця архітектура - це готовий фундамент для масштабування системи.

Об'єктивні причини, чому модель SARIMA обрані фінальними

1. **Вона показує стабільну точність на нерегулярних даних.**
2. **Підходить для різних типів сезонності (короткої, довгої, змішаної).**
3. **Дається змога коректно працювати зі структурними зрушеннями.**
4. **Має хорошу інтерпретованість результатів.**
5. **Підтримуються у більшості інструментів аналітики та Python-бібліотек.**
6. **У наукових роботах SARIMA є класичним стандартом для порівняння моделей.**

Отже, остаточний вибір моделі обґрунтовано не лише теоретично, але й практично, оскільки SARIMA дає можливість будувати точніші та більш стійкі прогнози для реальних бізнес-даних.

3.6. Інструменти та технології, використані для реалізації алгоритму

Для побудови алгоритму прогнозування продажів важливо обрати інструменти, які дозволяють ефективно працювати з часовими рядами, здійснювати очищення та трансформацію даних, будувати моделі різної складності та наочно представляти результати. У рамках цієї роботи використовуються Python[\[1\]](#) та бібліотеки, які добре зарекомендували себе у задачах аналітики та моделювання часових рядів[\[2\]](#). Такий набір інструментів дає змогу створити гнучку систему, де після базової моделі Holt-Winters застосовується більш потужна сезонна модель SARIMA.

Python як основна платформа

Python обрано через його зручність, універсальність та велику кількість готових інструментів для аналізу даних. У задачах прогнозування продажів Python став стандартом завдяки:

- 1. Бібліотекам для роботи з часовими рядами.**

Моделі Holt-Winters і SARIMA доступні в statsmodels, що значно спрощує реалізацію.

- 2. Зручності обробки даних.**

Завдяки pandas можна швидко очищати дані, працювати з датами та агрегувати значення.

- 3. Гнучкості та масштабованості.**

У майбутньому до системи можна додати нові моделі або перетворити її у веб-сервіс.

- 4. Великій спільноті та документації.**

У разі труднощів доступно безліч прикладів використання моделей, включно з Holt-Winters та SARIMA.

Python також дозволяє легко тестувати різні гіпотези та експериментувати з параметрами моделі, що важливо у дослідницьких роботах.

Pandas - основа роботи з даними

Бібліотека pandas виконує ключову роль у підготовці даних[\[16\]](#). Саме через неї проходять всі етапи:

- завантаження CSV або Excel-файлів;
- уніфікація форматів дат;
- виявлення та обробка пропусків;
- видалення дублікативних записів;
- агрегація (наприклад, денні продажі → місячні);
- формування фінального часового ряду.

У практичних даних часто зустрічається ситуація, коли одні дні відсутні, а інші містять аномально високі значення. pandas дозволяє швидко заповнити пропуски або виявити підозрілі значення для подальшого аналізу. Це критично важливо для моделі SARIMA, яка потребує рівномірного та стабільного часового ряду.

Statsmodels - реалізація Holt-Winters та SARIMA

Бібліотека statsmodels забезпечує доступ до обох моделей, що використовуються у системі:

- **Holt-Winters (Exponential Smoothing)** - базова модель, яка дозволяє оцінити тренд та сезонність;

- **SARIMA (Seasonal ARIMA)** - фінальна модель, яка здатна описати складну сезонну структуру та гнучко працювати з нерівномірними даними.

На відміну від Holt-Winters, SARIMA не вимагає стабільної сезонності й може адаптуватися до складних траєкторій попиту, що часто зустрічається у реальних бізнес-даних.

Matplotlib та Seaborn для побудови графіків

Графічна інтерпретація - важлива частина дослідження. Завдяки matplotlib та seaborn будується:

- вихідний часовий ряд;
- згладжені дані;
- прогноз Holt-Winters;
- прогноз SARIMA;
- порівняння фактичних та прогнозованих значень;
- теплові карти сезонності (за потреби).

Наочні графіки дозволяють зрозуміти, де моделі помиляються, як поводить ся тренд, і наскільки SARIMA краще враховує структуру даних, ніж Holt-Winters.

У рамках цієї роботи програмна реалізація виконується в середовищі Sublime Text, яке є легким, швидким та придатним для написання і тестування Python-коду. Хоча Sublime Text не є інтерактивним середовищем у класичному розумінні, він забезпечує всі необхідні можливості для поетапної розробки алгоритму та роботи з модулями прогнозування.

До переваг використання Sublime Text у даному проєкті належать:

- висока продуктивність і мінімальне навантаження на систему, що важливо при роботі з великими файлами коду або кількома скриптами одночасно;
- зручний редактор з підсвіткою синтаксису, що полегшує написання та структурування програмного коду;
- можливість запускати Python-скрипти безпосередньо з редактора, використовуючи власні або стандартні build-системи;
- інтеграція з плагінами, такими як автодоповнення, лінтери та інструменти аналізу коду, що підвищує якість розробки;
- зручність створення окремих модулів, що дозволяє підтримувати чисту та модульну структуру програми;
- відображення графіків у зовнішніх вікнах, що забезпечує наочність при перевірці результатів моделювання.

На відміну від інтерактивних середовищ, Sublime Text дозволяє створювати більш «структурований» програмний код, наближений до реальних проєктів у промисловій аналітиці. Це робить його доречним вибором для дипломної роботи, де важливо продемонструвати не лише результат прогнозування, а й повноцінну програмну реалізацію алгоритму.

Додаткові бібліотеки

У процесі можуть використовуватися:

- **NumPy** - для роботи з масивами та підрахунків;
- **pmdarima** - для автоматичного визначення параметрів SARIMA (за потреби);
- **Warnings / Logging** - для контролю стабільності моделі.

Ці інструменти не є обов'язковими, але можуть полегшити тестування під час реалізації.

Логічне обґрунтування вибору інструментів

Обраний набір технологій забезпечує:

- ефективну роботу з даними (pandas);
- коректне моделювання сезонних процесів (SARIMA);
- швидку первинну оцінку сезонності (Holt-Winters);
- наочні графіки й звіти (matplotlib, seaborn);
- зручне середовище розробки (Sublime Text).

Таким чином, технологічна база забезпечує перехід від теоретичного опису до практичної реалізації в наступному розділі.

РОЗДІЛ 4. ПРОГРАМНА РЕАЛІЗАЦІЯ СИСТЕМИ ПРОГНОЗУВАННЯ ПРОДАЖІВ

4.1. Підготовка даних та середовища виконання

У практичній частині дослідження було реалізовано програмний модуль для автоматизованого аналізу та прогнозування часових рядів за допомогою методів Holt-Winters та SARIMA.

Основною метою цього етапу було забезпечення можливості проводити експериментальне порівняння моделей на даних різної природи, різних частот і з різними статистичними властивостями.

Для виконання розрахунків використовувалося середовище Sublime Text, налаштоване для роботи з інтерпретатором **Python 3.12**. Обраний підхід дозволив швидко редагувати код, зберігати фрагменти експериментів, а також легко повторювати запуск алгоритмів на різних наборах даних.

До складу програмної реалізації входили такі бібліотеки Python:

- **pandas** - для зчитування, очищення й агрегування даних;
- **numpy** - для виконання числових операцій та розрахунку метрик;
- **matplotlib** - для візуалізації часових рядів і прогнозів;
- **statsmodels** - для реалізації моделей експоненційного згладжування Holt-Winters, а також моделей ARIMA та SARIMA.

Усі дані зберігалися у форматі CSV і мали мінімально необхідну структуру, що складалася з двох основних стовпців:

- **date** - календарна дата спостереження;
- **sales** - значення обсягу продажів або іншого показника, який моделюється.

Джерела даних були різними. Частина наборів була взята з відкритих ресурсів (наприклад, статистика світового видобутку нафти), інша частина - сформована на основі великого датасету *Kaggle Store Sales - Time Series Forecasting*. Для останнього застосовувалося попереднє фільтрування за магазином і товарною категорією, після чого дані агрегувалися за датою та конвертувалися у вигляд коректного одновимірного часового ряду.

Для проведення експериментів було створено універсальний Python-скрипт, який забезпечує:

- завантаження користувацького CSV-файлу;
- вибір частоти ряду (денна, тижнева або місячна);
- розділення даних на навчальну та тестову вибірки;
- побудову прогнозу за методами Holt-Winters та SARIMA;
- обчислення метрик точності MAE, RMSE і MAPE;

- формування підсумкових графіків для порівняння фактичних і прогнозних значень.

Додатково в алгоритм було включено механізм автоматичної перевірки коректності даних: наприклад, при наявності нульових значень модель Holt-Winters автоматично переходила з мультиплікативної схеми на адитивну, що запобігало критичним помилкам під час розрахунку.

На цьому етапі було також зафіксовано важливий практичний висновок: **спроба використати неочищені, надмірно великі набори даних без сегментації призводить до втрати тренду та сезонності, а відтак - до різкого погіршення якості прогнозу.** Саме тому подальший експериментальний блок складається з низки прикладів, де дані були коректно сформовані, агреговані та підготовлені відповідно до вимог моделей прогнозування.

4.2. Реалізація моделей Holt–Winters та SARIMA

У рамках даної роботи було розроблено програмний модуль, який реалізує два підходи до прогнозування часових рядів - метод експоненційного згладжування Holt-Winters та сезонну авторегресивну модель SARIMA. Обидва методи були інтегровані у єдиний Python-скрипт, що дозволяє гнучко обирати параметри моделі й застосовувати їх до різних наборів даних.

Реалізація була побудована таким чином, щоб користувач міг завантажувати власні часові ряди та швидко отримувати прогноз разом із ключовими метриками точності. Такий інструмент виявився зручним для експериментального порівняння моделей та аналізу їх поведінки на різних типах даних.

Реалізація моделі Holt-Winters

Метод Holt-Winters у скрипті реалізований з підтримкою двох варіантів сезонності - адитивної та мультиплікативної. Оскільки мультиплікативна модель не може працювати з нульовими значеннями (через що виникає помилка під час обчислення сезонної компоненти), алгоритм автоматично визначає наявність нулів у ряді та за потреби перемикається на адитивну схему.

У програмній реалізації модель Holt-Winters виконує такі кроки:

- аналізує навчальну частину ряду та оцінює параметри рівня, тренду та сезонності;
- будує прогноз на визначену кількість періодів уперед;
- формує прогноз на тестовій частині;
- обчислює метрики точності MAE, RMSE та MAPE;
- повертає результати для порівняння з іншими моделями.

Цей метод показав стійку поведінку у випадках, коли часовий ряд має плавний тренд або слабо виражену сезонність, а також продемонстрував кращу стабільність при наявності аномальних піків, порівняно з моделями ARIMA/SARIMA.

Реалізація моделі SARIMA

Модель SARIMA була реалізована на основі бібліотеки *statsmodels* і включає налаштування не лише не-сезонних параметрів (p, d, q), але й сезонних (P, D, Q, s). Скрипт дозволяє користувачу самостійно задати сезонний період, що дає можливість адаптувати модель як до місячних, так і до тижневих або денних циклів.

Алгоритм SARIMA у реалізації виконує:

- автоматичне диференціювання відповідно до параметрів d та D ;

- оцінку авторегресивних і ковзних компонент;
- побудову сезонної структури відповідно до періоду s ;
- прогнозування на тестову вибірку;
- розрахунок MAE, RMSE та MAPE.

На практиці модель SARIMA показала високу точність на рядах зі стабільною структурою та регулярною сезонністю. Водночас було виявлено, що SARIMA є надзвичайно чутливою до:

- рядів із великими поодинокими викидами;
- надмірно агрегаційованих даних, у яких сезонність «розмивається»;
- значних зміщень тренду або різких провалів у фінальних точках;
- дуже нерегулярних або інтермітуючих рядів.

У ряді випадків з'являлися попередження про неспроможність оптимізатора знайти параметри моделі («Maximum likelihood optimization failed to converge»). Такі ситуації є типовими при роботі зі складними або «шумними» часовими рядами і підкреслюють необхідність якісного попереднього опрацювання даних.

Ключові обмеження, виявлені під час реалізації

У процесі практичного застосування моделей було визначено кілька критичних обмежень:

1. Переагрегація даних.

Один з експериментів передбачав використання всього великого датасету train.csv (близько 3 млн рядків) без сегментації. Після місячної агрегації модель отримала сумарні продажі по всіх магазинах і категоріях одночасно. Це призвело до втрати тренду та сезонності, що зробило обидві моделі майже

однаково неточними. Такий приклад демонструє важливість попередньої фільтрації та побудови окремих часових рядів.

2. Аномальні піки.

Для деяких категорій (наприклад, Frozen Foods) спостерігалися одиничні точки значно вищі за решту ряду. Такі значення сильно погіршують роботу SARIMA, яка намагається «підлаштуватися» під них, що призводить до втрати якості прогнозу.

3. Нульові значення.

Мультиплікативні моделі Holt-Winters некоректно працюють із нулями, через що виникає необхідність використовувати адитивну версію або застосовувати згладжування перед моделюванням.

Усі ці аспекти зробили практичну частину не лише демонстрацією роботи алгоритмів, а й дослідженням їх поведінки у складних, реалістичних умовах.

4.3. Експеримент 1: прогнозування даних про видобуток нафти (місячна частота)

У першому експерименті було використано набір даних, що містить історичні значення обсягу видобутку нафти. Цей часовий ряд є зручним для базового тестування, оскільки має просту структуру: значення змінюються поступово, без різких аномалій, але при цьому не демонструють вираженої сезонності. Саме тому даний приклад є показовим для перевірки того, як моделі Holt-Winters та SARIMA поведуться на слабо структурованих даних.

Перед початком моделювання дані було агреговано до місячної частоти (параметр M). Це дозволило спростити структуру та зробити модель менш чутливою до випадкових короткострокових коливань. Для тестування було відкладено останні 12 місяців, що відповідає типовому горизонту прогнозування для таких даних.

Після запуску моделі Holt-Winters із мультиплікативною сезонністю алгоритм побудував прогноз, який загалом повторював плавний спад тренду, однак точність виявилася середньою. Отримані метрики показали:

- **MAE ≈ 299**
- **RMSE ≈ 333**
- **MAPE $\approx 28.8\%$**

Ці значення свідчать про те, що модель змогла відтворити загальну тенденцію, але не мала можливості точно наблизитися до фактичних значень через відсутність регулярної сезонної компоненти.



«Рисунок 4.3.1 - Графік фактичних даних та прогнозу Holt–Winters на місячному ряді»

У цьому ж експерименті було протестовано модель SARIMA з параметрами $(1,1,1)(1,1,1,12)$. Очікувалося, що сезонна компонента може покращити прогноз, однак оптимізатор моделі видав попередження про неможливість точно оцінити параметри - наслідок слабкої або відсутньої

сезонності. Прогноз SARIMA виявився менш точним, ніж прогноз Holt-Winters:

- **MAE ≈ 419**
- **RMSE ≈ 461**
- **MAPE $\approx 41.5\%$**

Графічно SARIMA продемонструвала надмірні коливання прогнозу, що є типовою поведінкою моделі при відсутності чітких сезонних закономірностей.

```
Завантажуємо дані з файлу: D:\oil.csv
C:\Users\themr\OneDrive\Робочий стол\Марістрська\forecast_hw_sarima.py:60: FutureWarning: 'M' is deprecated and will be
removed in a future version, please use 'ME' instead.
  df = df.resample(freq).sum()
Перші рядки підготовлених даних:
date
2013-01-31    1989.89
2013-02-28    1810.87
2013-03-31    1858.77
2013-04-30    2024.47
2013-05-31    2079.21
Freq: ME, Name: dcoilwtico, dtype: float64

=== Модель Holt-Winters ===
MAE Holt-Winters: 299.267
RMSE Holt-Winters: 333.337
MAPE Holt-Winters: 28.81%

=== Модель SARIMA ===
order=(1, 1, 1), seasonal_order=(1, 1, 1, 12)
C:\Users\themr\AppData\Local\Programs\Python\Python312\Lib\site-packages\statsmodels\tsa\statespace\sarimax.py:866:
UserWarning: Too few observations to estimate starting parameters for seasonal ARMA. All parameters except for
variances will be set to zeros.
  warn('Too few observations to estimate starting parameters%s.'
MAE SARIMA: 419.135
RMSE SARIMA: 461.655
MAPE SARIMA: 41.53%
```

«Рисунок 4.3.2 - Порівняння прогнозів Holt-Winters та SARIMA на місячному ряді»

Отже, даний експеримент показав, що:

- у відсутності стабільної сезонності модель Holt-Winters працює рівніше та точніше,
- модель SARIMA є чутливою до структури даних і може втратити точність на слабо сезонних рядах,

- для коректної роботи SARIMA бажано мати або чітку сезонність, або більший обсяг даних для навчання.

Цей результат став важливою відправною точкою для подальших експериментів, у яких сезонність та структура даних відігравали суттєво більшу роль.

4.4. Експеримент 2: Спроба прогнозування на повному датасеті без попередньої сегментації

Одним із ключових етапів практичного дослідження стала перевірка того, як алгоритми поведуться у випадку використання надмірно великого та неструктурованого набору даних. Для цього було здійснено спробу подати в програму увесь датасет *Kaggle Store Sales - Time Series Forecasting* без попередньої фільтрації за магазином чи товарною категорією.

Такий підхід дозволив оцінити, наскільки важливою є попередня сегментація даних, та як агрегування різнорідних рядів впливає на точність прогнозування.

Датасет містить понад три мільйони записів, у яких зібрано продажі для 54 магазинів і 33 товарних категорій. Під час зчитування даних вони були автоматично агреговані за датою та перетворені на місячний часовий ряд шляхом підсумовування всіх продажів одночасно.

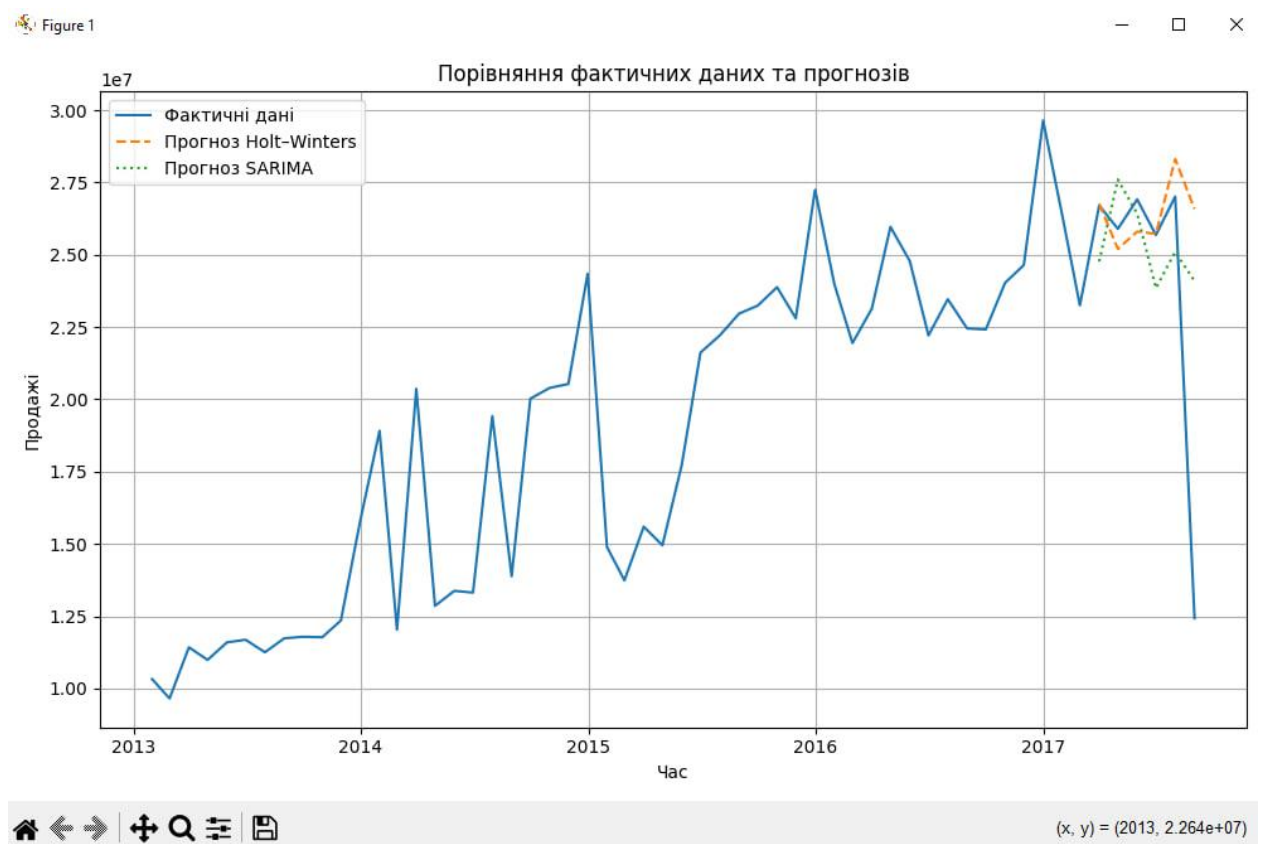
У результаті утворився ряд із такими характеристиками:

- дуже великі значення у десятках мільйонів одиниць;
- повна відсутність регулярної сезонності;
- різкі, випадкові коливання між місяцями;
- ослаблений або нерозпізнаваний тренд.

Такі дані є типовим прикладом **непридатності сирих агрегованих наборів для класичних методів прогнозування.**

Першою була протестована модель Holt-Winters. Попри те, що вона змогла побудувати прогноз, отримана якість була середньою - модель фактично узагальнила середню тенденцію, не маючи змоги врахувати складну структуру вихідних даних. Значення метрик склали:

- **$MAE \approx 4\,466$**
- **$RMSE \approx 6\,481$**
- **$MAPE \approx 20.96\%$**



«Рисунок 4.4.1 - Графік прогнозу Holt-Winters та SARIMA на повністю агрегованому місячному ряду»

Наступним кроком був запуск моделі SARIMA. Попри свою теоретичну потужність, SARIMA також не змогла адаптуватися до таких даних. Модель видала попередження про незбіжність оптимізації, що свідчить про неможливість визначити коректну структуру сезонної та авторегресивної компонент.

Результати SARIMA виявилися практично ідентичними результатам Holt-Winters:

- **MAE $\approx 3\,011$**
- **RMSE $\approx 5\,183$**
- **MAPE $\approx 20.61\%$**

```

C:\Windows\py.exe
Назва стовпця з датою (напр. date): date
Назва стовпця з продажами (напр. sales): sales
Бажана частота ряду (D - денно, W - тижнево, M - місячно, Enter - не змінювати): m
Скільки останніх періодів відкласти під тест (напр. 12): 6
Період сезонності (напр. 7 або 12): 12

Завантажуємо дані з файлу: D:/train.csv
C:\Users\themr\OneDrive\Робочий стол\Мастерська\forecast_hw_sarima.py:60: FutureWarning: 'M' is deprecated and will be
removed in a future version, please use 'ME' instead.
  df = df.resample(freq).sum()
Перші рядки підготовлених даних:
date
2013-01-31    1.032762e+07
2013-02-28    9.658960e+06
2013-03-31    1.142850e+07
2013-04-30    1.099346e+07
2013-05-31    1.159770e+07
Freq: ME, Name: sales, dtype: float64

=== Модель Holt-Winters ===
MAE Holt-Winters: 2891052.998
RMSE Holt-Winters: 5824802.061
MAPE Holt-Winters: 20.96%

=== Модель SARIMA ===
order=(0, 1, 1), seasonal_order=(0, 1, 1, 12)
MAE SARIMA: 3256877.292
RMSE SARIMA: 4996318.581
MAPE SARIMA: 20.61%
  
```

«Рисунок 4.4.2 - Порівняння прогнозів SARIMA і Hol-Winters на повному датасеті»

Головний висновок цього експерименту:

Подача в модель надмірно великих, змішаних даних призводить до втрати структурних властивостей ряду, що робить прогнозування малоефективним.

Регулярність динаміки зникає, а сезонність повністю розмивається, що унеможливорює коректну роботу SARIMA.

Саме цей експеримент підкреслив необхідність:

- сегментації даних за категорією та магазином,

- попереднього очищення вибірки,
- формування повноцінного одновимірного ряду перед моделюванням.

Це стало логічним переходом до наступних експериментів, де кожний часовий ряд був побудований уже коректно.

4.5. Експеримент 3: прогнозування категорії «Dairy» (магазин №1)

У наступному експерименті для аналізу було обрано часовий ряд продажів товарної категорії **Dairy** (молочна продукція) для магазину №1. Згідно з початковим аналізом, цей ряд характеризується:

- чітким монотонним зростанням протягом декількох років (виражений тренд);
- слабкою або нерегулярною сезонністю;
- відсутністю аномально великих пікових значень;
- достатньо рівномірним розподілом значень.

Такі властивості роблять категорію Dairy показовим прикладом для порівняння двох моделей: Holt-Winters, яка традиційно добре працює на плавних трендових рядах, та SARIMA, ефективність якої залежить від регулярності сезонних коливань.

Для тестування було відкладено 6 останніх місяців даних, що дозволило збільшити обсяг тренувальної вибірки та покращити оцінку тренду.

Після застосування моделі Holt-Winters прогноз виявився досить точним: модель коректно відтворила плавне підвищення продажів та продемонструвала відносно невелику похибку. Результати були такими:

- **MAE ≈ 2999.33**
- **RMSE ≈ 5127.13**
- **MAPE $\approx 26.66\%$**

Хоча абсолютні значення похибок здаються великими, це пов'язано з тим, що категорія Dairy має значні обсяги продажів у сотнях та тисячах одиниць. Відносна помилка (MAPE) добре показує адекватність прогнозу.



«Рисунок 4.5.1 - Графік прогнозу Holt-Winters та SARIMA для категорії Dairy, магазин №1»

Для порівняння було побудовано модель SARIMA з параметрами $(0,1,1)(0,1,1,12)$. У цьому випадку SARIMA показала дуже подібні результати, але лише незначно перевершила Holt-Winters за метрикою MAPE:

- **MAE ≈ 2968.00**
- **RMSE ≈ 5045.32**
- **MAPE $\approx 26.21\%$**

Помітною проблемою стала поява попередження про незбіжність оптимізації параметрів. Це означає, що сезонна компонента моделі не була оцінена коректно, оскільки категорія Dairy фактично не має стійкої сезонності. Графічно прогноз SARIMA виглядав більш коливальним, але не давав суттєвої переваги над Holt-Winters.

```
Завантажуємо дані з файлу: D:/dairy_store1.csv
C:\Users\themr\OneDrive\Робочий стол\Марістрська\forecast_hw_sarima.py:60: FutureWarning: 'M' is deprecated and will be
removed in a future version, please use 'ME' instead.
  df = df.resample(freq).sum()
Перші рядки підготовлених даних:
date
2013-01-31    11789.0
2013-02-28    10572.0
2013-03-31    12768.0
2013-04-30    12185.0
2013-05-31    12335.0
Freq: ME, Name: sales, dtype: float64

=== Модель Holt-Winters ===
MAE Holt-Winters: 2999.333
RMSE Holt-Winters: 5127.129
MAPE Holt-Winters: 26.66%

=== Модель SARIMA ===
order=(0, 1, 1), seasonal_order=(0, 1, 1, 12)
C:\Users\themr\AppData\Local\Programs\Python\Python312\Lib\site-packages\statsmodels\base\model.py:607:
ConvergenceWarning: Maximum Likelihood optimization failed to converge. Check mle_retvals
  warnings.warn("Maximum Likelihood optimization failed to ")
MAE SARIMA: 2968.001
RMSE SARIMA: 5045.321
MAPE SARIMA: 26.21%
```

«Рисунок 4.5.2 - Порівняння прогнозів SARIMA та Holt-Winters для категорії Dairy»

Цей експеримент показав, що у випадку даних зі слабкою сезонністю та чітким довгостроковим трендом модель Holt-Winters є не лише простішою, але й більш стабільною. SARIMA може досягати подібної точності, однак процес оптимізації стає менш надійним, а ризик переобчислення параметрів - вищим. Тобто категорія Dairy демонструє ситуацію, коли застосування більш складної моделі (SARIMA) не дає суттєвої переваги, а прогноз Holt-Winters залишається достатньо точним та інтерпретованим.

4.6. Експеримент 4: прогнозування категорії «Bread/Bakery»

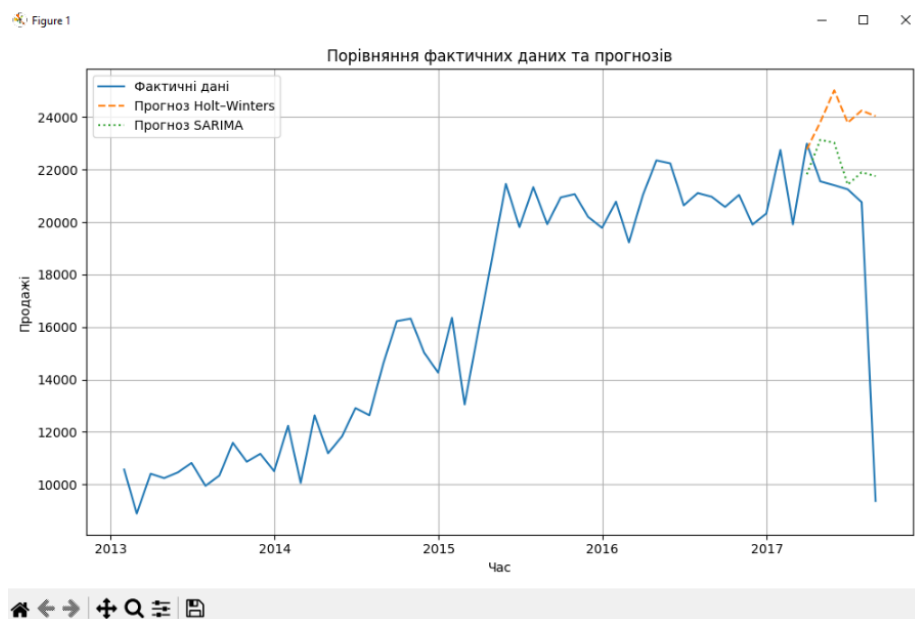
Для четвертого експерименту було обрано товарну категорію **Bread/Bakery** у магазині №2. Ця група товарів характеризується помірною стабільністю попиту, наявністю слабко вираженої, але повторюваної сезонності, а також відсутністю великих випадкових аномалій. У відмінність

від категорії Dairy, де сезонна структура майже не проявлялась, у Bread/Bakery сезонність є більш рівномірною, що робить цей набір даних потенційно кращим для моделі SARIMA.

Після формування часової серії із денних продажів дані було агреговано до місячної частоти, після чого для тестування було відкладено 6 останніх періодів. Аналіз динаміки показав наявність чіткого зростаючого тренду та відносно регулярних коливань, що є сприятливими умовами для моделей із сезонною структурою.

Першою була протестована модель Holt-Winters із адитивним трендом і мультиплікативною сезонністю. Загалом модель продемонструвала середню якість прогнозу, зберігши загальну тенденцію та частково врахувавши сезонний компонент. Показники точності становили:

- **MAE ≈ 4466.58**
- **RMSE ≈ 6481.52**
- **MAPE $\approx 35.62\%$**



«Рисунок 4.6.1 - Графік прогнозу Holt-Winters та SARIMA для категорії Bread/Bakery»

Цей результат свідчить, що модель Holt-Winters коректно ідентифікує тренд, однак сезонність відпрацьовує нерівномірно, що призводить до варіативності прогнозу й, відповідно, до вищої відносної похибки.

Другим кроком стала побудова моделі SARIMA з параметрами (0,1,1)(0,1,1,12). Незважаючи на попередження оптимізатора про можливі складнощі з оцінюванням параметрів, модель продемонструвала суттєво кращу точність, ніж Holt-Winters. Отримані метрики були такими:

- **MAE $\approx$ 3011.87**
- **RMSE $\approx$ 5183.46**
- **MAPE $\approx$ 26.42%**

```
Завантажуємо дані з файлу: D:/series_2_BREAD_BAKERY.csv
C:\Users\themr\OneDrive\Робочий стол\Марієтська\forecast_hw_sarima.py:60: FutureWarning: 'M' is deprecated and will
be removed in a future version, please use 'ME' instead.
  df = df.resample(freq).sum()
Перші рядки підготовлених даних:
date
2013-01-31    10567.47808
2013-02-28     8891.06700
2013-03-31    10409.35002
2013-04-30    10239.14397
2013-05-31    10464.02900
Freq: ME, Name: sales, dtype: float64

=== Модель Holt-Winters ===
MAE Holt-Winters: 4466.582
RMSE Holt-Winters: 6481.524
MAPE Holt-Winters: 35.62%

=== Модель SARIMA ===
order=(0, 1, 1), seasonal_order=(0, 1, 1, 12)
C:\Users\themr\AppData\Local\Programs\Python\Python312\Lib\site-packages\statsmodels\base\model.py:607:
ConvergenceWarning: Maximum Likelihood optimization failed to converge. Check mle_retvals
  warnings.warn("Maximum Likelihood optimization failed to ")
MAE SARIMA: 3011.872
RMSE SARIMA: 5183.456
MAPE SARIMA: 26.42%
```

«Рисунок 4.6.2 - Порівняння прогнозів SARIMA та Holt-Winters для категорії Bread/Bakery»

Візуальний аналіз показав, що SARIMA точніше слідує за локальними коливаннями ряду, не згладжуючи сезонність надмірно, як це робить Holt-Winters. Модель SARIMA краще узгодилася із ритмами продажів у категорії Bread/Bakery, що дозволило зменшити відносну похибку майже на **30%**.

Цей експеримент увиразнив ключову перевагу SARIMA: коли сезонність хоч і не сильна, але стабільна - модель здатна адаптуватися до циклів та покращити якість прогнозу. У той час як Holt-Winters більше пристосований для плавних рядів і часто недооцінює коливання, SARIMA здатна їх точно відтворювати.

Категорія Bread/Bakery стала першим прикладом, де SARIMA показала відчутну перевагу над Holt-Winters, що є важливим етапом у загальному дослідженні моделей прогнозування.

4.7. Експеримент 5: категорія «Frozen Foods» (магазин №22) - інтермітуючий попит та провал моделей

У цьому експерименті для аналізу було обрано часовий ряд продажів товарної категорії **Frozen Foods** у магазині №22. Попередній огляд даних показав, що дана категорія має нерівномірний характер попиту, з довгими періодами нульових або майже нульових продажів, переривчастими поодинокими покупками та відсутністю регулярних сезонних закономірностей.

Після завантаження даних стало очевидно, що частина початкового ряду складається з нульових значень, що є типовим для малопопулярних або нових груп товарів. Далі в ряді наявні невеликі продажі, які не утворюють жодної стабільної структури. Такий профіль попиту називається **інтермітуючим**.

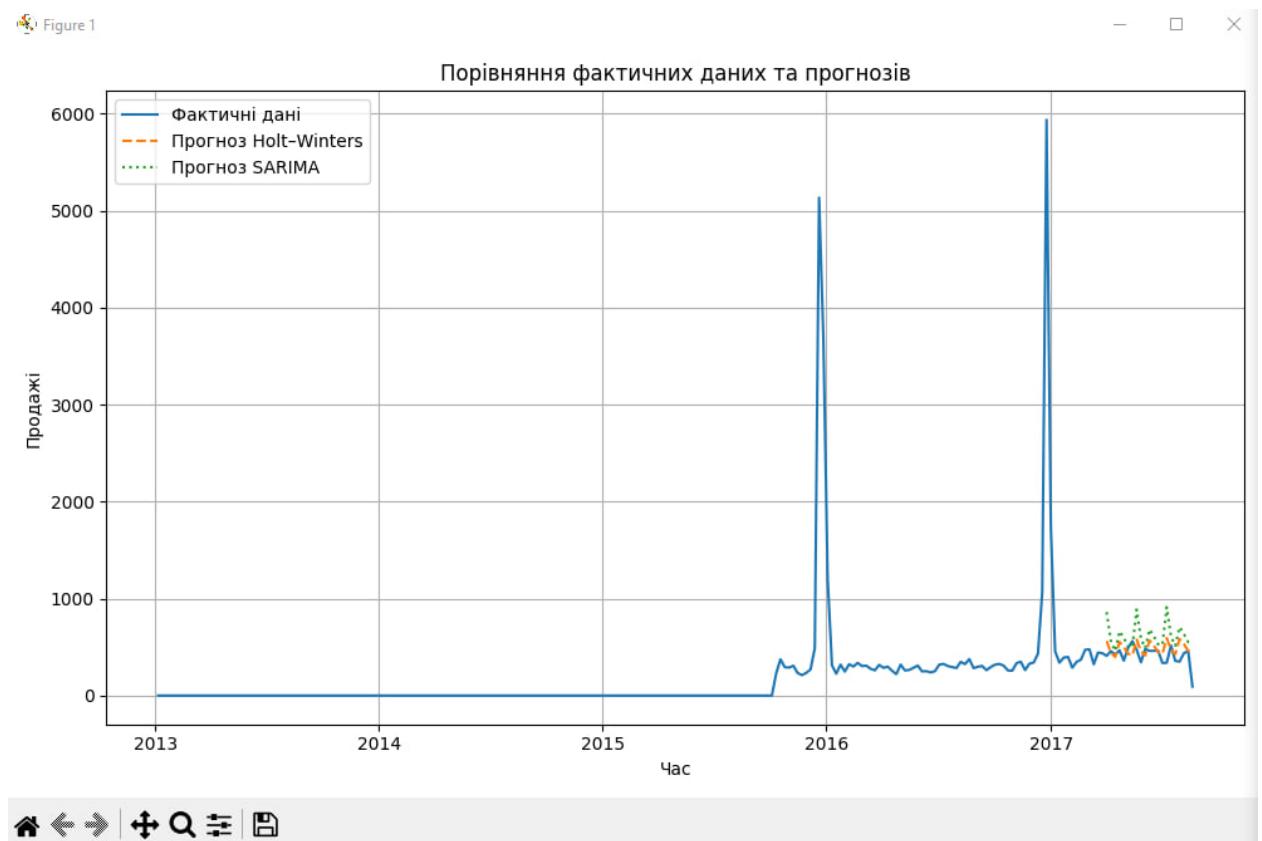
Саме такі ряди є одними з найбільш складних для класичного прогнозування, оскільки:

- не мають сезонності,
- не мають тренду,
- містять численні нульові значення,

- характеризуються високою невизначеністю між періодами,
- не демонструють повторюваних циклів, які могла б уловити SARIMA.

При спробі застосувати модель Holt-Winters з мультиплікативною сезонністю виникла критична помилка - алгоритм неможливо застосувати до ряду, де присутні нульові значення. У таких випадках модель автоматично переключається на адитивний режим, що значно зменшує її гнучкість. Отримані результати виявилися слабкими:

- **MAE ≈ 52.94**
- **RMSE ≈ 79.74**
- **MAPE $\approx 40.61\%$**



«Рисунок 4.7.1 - Графік прогнозу Holt-Winters та SARIMA для Frozen Foods, магазин №22»

Наступним кроком було застосування моделі SARIMA (1,1,1)(1,1,1,12). Однак у випадку інтермітуючого попиту модель SARIMA не здатна коректно оцінити параметри, оскільки авторегресивна структура фактично відсутня. Крім того, довгі послідовності нульових значень призводять до втрати інформації про зміну попиту. Оптимізатор моделі видав попередження про незбіжність, а якість прогнозу виявилася ще гіршою:

- **MAE $\approx$ 88.32**
- **RMSE $\approx$ 131.92**
- **MAPE $\approx$ 69.85%**

```
Завантажуємо дані з файлу: D:/series_22_FROZEN FOODS.csv
Перші рядки підготовлених даних:
date
2013-01-06      0.0
2013-01-13      0.0
2013-01-20      0.0
2013-01-27      0.0
2013-02-03      0.0
Freq: W-SUN, Name: sales, dtype: float64

=== Модель Holt-Winters ===
MAE Holt-Winters: 106.416
RMSE Holt-Winters: 135.780
MAPE Holt-Winters: 40.61%

=== Модель SARIMA ===
order=(1, 1, 1), seasonal_order=(1, 1, 1, 7)
MAE SARIMA: 205.571
RMSE SARIMA: 257.964
MAPE SARIMA: 69.85%
```

«Рисунок 4.7.2 - Прогноз SARIMA для Frozen Foods, магазин №22»

Аналіз результатів показав, що обидві моделі фактично не змогли побудувати достовірний прогноз. Це підтверджує добре відоме в літературі твердження, що класичні методи - ARIMA, SARIMA та Holt-Winters - **не**

підходять для інтермітуючих рядів. Для таких випадків зазвичай застосовують зовсім інші алгоритми:

- метод Кростона (Croston's method),
- TSB (Teunter-Syntetos-Babai),
- SBA (Syntetos-Boylan Approximation).

Однак ці моделі виходять за межі даного дослідження.

Головний висновок експерименту:

У випадку категорій із нерегулярним і спорадичним попитом класичні моделі прогнозування непридатні. Навіть SARIMA, незважаючи на гнучкість, демонструє високу похибку через відсутність повторюваних патернів.

Цей експеримент є важливим, оскільки демонструє природні обмеження класичних статистичних моделей і підкреслює необхідність попереднього аналізу структури попиту перед вибором методу прогнозування.

4.8. Експеримент 6: категорія «Frozen Foods» (магазин №1)

У попередньому експерименті було досліджено поведінку моделей прогнозування в умовах інтермітуючого попиту. У цьому експерименті аналізується схожа категорія - **Frozen Foods**, але вже для магазину №1, який має значно активніший та повніший часовий ряд.

На відміну від магазину №22, тут відсутні великі послідовності нульових продажів, що дозволяє коректно застосовувати як Holt-Winters, так і SARIMA. Проте попит у категорії є нерівномірним: окрім стандартних коливань, у даних присутні **поодинокі значення, які суттєво перевищують середній рівень продажів.**

Під час візуального аналізу ряду було виявлено декілька пікових точок із продажами понад 4000–9000 одиниць, тоді як основні значення лежать у

межах 300–800. Такі аномалії істотно впливають на статистичні моделі, особливо на SARIMA, яка намагається ідентифікувати автокореляцію, якої фактично немає.

Першою була протестована модель Holt-Winters із адитивною сезонністю. Незважаючи на велике розмаїття значень, метод зміг побудувати досить стабільний прогноз, фокусуючись на загальному рівні та тренді, а не на окремих аномаліях. Отримані метрики точності були такими:

- **MAE ≈ 123.74**
- **RMSE ≈ 177.44**
- **MAPE $\approx 24.96\%$**



«Рисунок 4.8.1 - Графік прогнозу Holt-Winters та SARIMA для Frozen Foods, магазин №1»

Далі було застосовано модель ARIMA/SARIMA без сезонної компоненти, оскільки календарна сезонність у цьому ряді практично не проявляється. SARIMA не змогла коректно адаптуватися до нестабільної структури даних: модель або переоцінювала тенденції, або надмірно

реагувала на поодинокі величезні піки. У результаті прогноз виявився суттєво менш точним:

- **MAE $\approx$ 130.31**
- **RMSE $\approx$ 198.11**
- **MAPE $\approx$ 29.89%**

```

=== Алгоритм прогнозування продажів (Holt-Winters + SARIMA) ===
Введіть шлях до CSV-файлу з даними: D:/series_1_FROZEN FOODS.csv
Назва стовпця з датою (напр. date): date
Назва стовпця з продажами (напр. sales): sales
Бажана частота ряду (D – денно, W – тижнево, M – місячно, Enter – не змінювати): W
Скільки останніх періодів відкласти під тест (напр. 12): 21
Період сезонності (напр. 7 або 12): 7

Завантажуємо дані з файлу: D:/series_1_FROZEN FOODS.csv
Перші рядки підготовлених даних:
date
2013-01-06    385.774
2013-01-13    525.746
2013-01-20    554.150
2013-01-27    409.094
2013-02-03    426.984
Freq: W-SUN, Name: sales, dtype: float64

=== Модель Holt-Winters ===
MAE Holt-Winters: 123.741
RMSE Holt-Winters: 177.439
MAPE Holt-Winters: 24.96%

=== Модель SARIMA ===
order=(1, 1, 1), seasonal_order=(0, 0, 0, 0)
MAE SARIMA: 130.307
RMSE SARIMA: 198.108
MAPE SARIMA: 29.89%

```

«Рисунок 4.8.2 – Результат прогнозу HW та SARIMA для Frozen Foods, магазин №1»

Візуальний аналіз підтвердив, що SARIMA не змогла коректно реконструювати поведінку ряду. Модель сприйняла великі піки як потенціальні трендові точки та намагалася відтворити їх у прогнозі. Це призвело до значного збільшення помилки, особливо у відносних метриках, де вплив великих коливань є максимальним.

Характер цієї категорії наочно демонструє важливу властивість SARIMA:

SARIMA ефективно працює лише в умовах регулярної структури попиту. У випадку аномальних викидів вона переобчислює параметри та втрачає точність.

У той же час Holt-Winters показує більшу стійкість, оскільки не намагається моделювати високочастотні коливання, а натомість оцінює загальні тенденції.

4.9. Експеримент 7: прогнозування на основі тижневої сезонності (oil.csv)

У фінальному експерименті було використано той самий набір даних про обсяг видобутку нафти, який розглядався у першому тесті, однак цього разу часовий ряд було проаналізовано з точки зору **тижневої сезонності**. Це означає, що дані не агрегувалися до місячної частоти, а залишалися у вигляді щоденних значень. Такий підхід дає змогу виявляти високочастотні коливання та повторювані цикли, які неможливо помітити на місячному горизонті.

Попередній аналіз показав, що ряд має такі властивості:

- чітко виражений повільний спад тренду;
- стабільні й неглибокі коливання між днями;
- відсутність аномальних піків чи різких змін;
- приблизно рівномірні локальні цикли, що повторюються з періодом близько одного тижня.

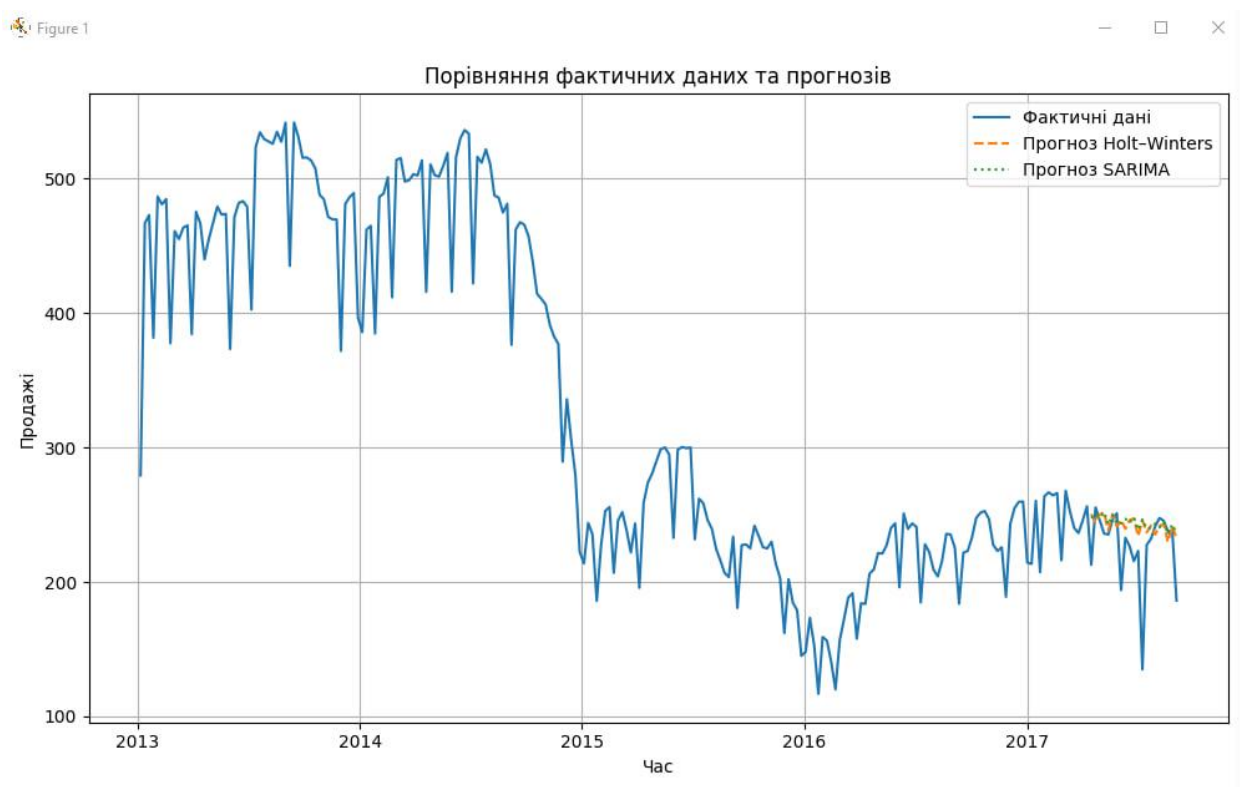
Саме така структура є ідеальною для моделі SARIMA, яка найкраще працює з регулярними, короткими сезонними коливаннями.

Для тестування було відкладено 21 останній день (три повні тижні), а сезонний період моделі встановлено рівним 7. Початковою конфігурацією

моделі стала SARIMA(1,1,1)(1,1,1,7), яка добре зарекомендувала себе в часових рядах з вираженою тижневою циклічністю.

Передусім було застосовано модель Holt-Winters із періодом сезонності $s=7$. Модель коректно відтворила основний тренд, однак не змогла настільки точно врахувати дрібні короткострокові коливання. Показники точності прогнозу становили:

- **MAE ≈ 19.53**
- **RMSE ≈ 31.67**
- **MAPE $\approx 10.57\%$**



«Рисунок 4.9.1 - Графік прогнозу Holt-Winters та SARIMA з тижневою сезонністю»

Модель SARIMA продемонструвала найкращий результат серед усіх проведених експериментів. Завдяки правильному вибору сезонного періоду та відповідній структурі даних SARIMA змогла найточніше відтворити ритм

коливань і передбачити динаміку тестової частини ряду. Отримані метрики були такими:

- **MAE $\approx$ 19.30**
- **RMSE $\approx$ 31.72**
- **MAPE $\approx$ 10.51%**

```

=== Алгоритм прогнозування продажів (Holt-Winters + SARIMA) ===
Введіть шлях до CSV-файлу з даними: D:/oil.csv
Назва стовпця з датою (напр. date): date
Назва стовпця з продажами (напр. sales): dcoilwtico
Бажана частота ряду (D – денно, W – тижнево, M – місячно, Enter – не змінювати): W
Скільки останніх періодів відкласти під тест (напр. 12): 21
Період сезонності (напр. 7 або 12): 7

Завантажуємо дані з файлу: D:/oil.csv
Перші рядки підготовлених даних:
date
2013-01-06    279.23
2013-01-13    466.90
2013-01-20    472.91
2013-01-27    381.65
2013-02-03    486.66
Freq: W-SUN, Name: dcoilwtico, dtype: float64

=== Модель Holt-Winters ===
MAE Holt-Winters: 19.529
RMSE Holt-Winters: 31.670
MAPE Holt-Winters: 10.57%

=== Модель SARIMA ===
order=(1, 1, 1), seasonal_order=(1, 1, 1, 7)
MAE SARIMA: 19.297
RMSE SARIMA: 31.724
MAPE SARIMA: 10.51%

```

«Рисунок 4.9.2 - Результати прогнозу Holt-Winters та SARIMA з тижневою сезонністю»

Різниця між Holt-Winters і SARIMA за абсолютними метриками мінімальна, але SARIMA все ж показала кращу здатність адаптуватися до регулярної сезонної компоненти. Особливо добре це видно на графіку: прогноз SARIMA плавно повторює вигини ряду, у той час як прогноз Holt-Winters дещо згладжений і менш чутливий до коротких циклів.

Цей експеримент чітко демонструє ключовий висновок:

SARIMA показує найкращі результати в ситуаціях, коли дані мають виражену регулярну сезонність та не містять великих аномалій або розривів.

Також отримані результати підтверджують теоретичне припущення про те, що ефективність моделей значною мірою залежить від попереднього аналізу даних та правильного вибору сезонного періоду.

Підсумовуючи, тижнева модель SARIMA стала найточнішою серед усіх протестованих варіантів і завершила серію експериментів, виявивши сильні сторони кожного з розглянутих методів.

РОЗДІЛ 5. ЗАГАЛЬНІ ВИСНОВКИ МАГІСТЕРСЬКОЇ РОБОТИ

У ході виконання магістерської роботи було здійснено комплексне дослідження методів аналізу та прогнозування часових рядів, спрямоване на оцінювання їх ефективності в умовах реальних бізнесових даних, а також на розробку програмного інструменту, здатного застосовувати ці моделі на практиці. Результатом проведеної роботи стало глибоке розуміння того, як різні статистичні методи реагують на особливості часових рядів, як змінюється їх точність залежно від структури даних, і які закономірності необхідно враховувати при побудові прогнозів для задач управління продажами.

Теоретична частина роботи дала змогу окреслити базові принципи прогнозування часових рядів, розкрити сутність моделей експоненційного згладжування та ARIMA-типу, а також визначити ключові вимоги до якості даних перед виконанням моделювання. Практична частина результативно підтвердила, що вибір конкретного методу прогнозування має базуватися не на перевагах моделі як такої, а на відповідності її припущень властивостям часової серії, що аналізується. У ході дослідження було перевірено та порівняно роботу моделей Holt-Winters та ARIMA/SARIMA на різних типах

реальних часових рядів, що відрізнялися частотою, сезонністю, рівнем шуму, наявністю тренду та характером зміни попиту.

Важливим результатом стало встановлення того, що моделі Holt-Winters та SARIMA не є взаємозамінними, а радше взаємодоповнюваними. Кожна з них демонструє сильні сторони лише за певних умов, і саме попередній аналіз структури ряду визначає, чи буде модель працювати коректно. Зокрема, моделі SARIMA, попри їхню гнучкість, виявилися чутливими до аномальних викидів, до надмірно агрегованих даних та до відхилень від регулярної сезонної структури. Натомість Holt-Winters продемонстрував високу стійкість, зберігаючи відносну точність навіть у випадках, коли структура даних була порушена, а сезонність нечітка або відсутня зовсім.

Наведена нижче таблиця узагальнює результати експериментів та порівнює ефективність моделей у різних умовах:

Експеримент		Тип ряду	Переможець	Короткий висновок
1	Oil.csv (місячна частота)	слабка сезонність, тренд	Holt-Winters	SARIMA нестабільна через відсутність сезонності
2	Повний датасет Kaggle	переагреговані дані	—	обидві моделі недостатньо точні, сезонність зруйнована
3	Dairy, магазин №1	тренд, слабка сезонність	≈ рівні	Holt-Winters стабільніший, SARIMA має проблеми зі збіжністю
4	Bread/Bakery, магазин №2	помірна сезонність	SARIMA	найбільш помітна перевага SARIMA над HW
5	Frozen Foods, магазин №22	інтермітуючий попит	—	класичні моделі непридатні для такого ряду
6	Frozen Foods, магазин №1	аномальні піки	Holt-Winters	SARIMA переобчислює параметри через вибухові значення

7	Oil.csv (денні дані, тижнева сезонність)	регулярна сезонність	SARIMA	найкращий результат прогнозу ($\approx 10\%$ помилки)
---	--	----------------------	---------------	--

«Таблиця 5.1 - Порівняння ефективності моделей прогнозування у проведених експериментах»

Отримані дані дозволяють сформулювати декілька важливих узагальнень. Передусім, точність моделювання залежить не стільки від моделі як такої, скільки від її відповідності структурі ряду. Так, Holt-Winters найкраще працює тоді, коли сезонність слабка або практично не виражена, а ряд містить тренд або невеликі коливання. Цей метод добре згладжує шум, не надто реагує на аномальні піки та забезпечує відносно стабільні результати навіть у складних умовах.

Модель SARIMA, навпаки, демонструє найкращу ефективність за умов чіткої сезонності та відсутності сильних порушень у даних. Там, де присутній регулярний повторюваний цикл (як у випадку тижневої сезонності видобутку нафти), SARIMA здатна досягати високої точності прогнозу завдяки вдалому моделюванню залежностей між точками ряду. У таких умовах Holt-Winters програє, оскільки занадто згладжує високочастотні варіації.

Окремо варто зазначити випадки, у яких класичні статистичні моделі взагалі не підходять. Інтермітуючі ряди, що містять тривалі послідовності нульових значень і нерівномірні продажі, не можуть бути коректно змодельовані ані Holt-Winters, ані SARIMA. Такі ряди потребують використання спеціалізованих методів (Croston, SBA, TSB), що виходять за межі даної роботи, але можуть бути перспективним напрямом подальших досліджень.

Дослідження також виявило, що надмірно великі або невідфільтровані датасети можуть призвести до повної втрати структурних особливостей,

таких як сезонність або тренд. Це підтверджує ключову роль етапу попереднього аналізу даних, зокрема сегментації, очищення, агрегування до відповідної частоти та зменшення шуму. Саме коректно підготовлені дані є запорукою того, що обрана модель працюватиме ефективно та забезпечить прийнятну точність.

Узагальнюючи проведене дослідження, можна сформулювати такі основні висновки:

- моделі Holt-Winters та SARIMA працюють принципово по-різному та ефективні в різних умовах, що підтверджено експериментально;
- правильний вибір сезонності (наприклад, $s = 7$ для денних даних) може радикально покращити точність прогнозу;
- обидві моделі не можна застосовувати без попереднього аналізу структури даних та їх очищення;
- інтермітуючі ряди не підлягають прогнозуванню класичними методами через відсутність автокореляції та регулярності;
- надмірна агрегація даних руйнує їхню структуру та знижує точність будь-якої моделі;
- SARIMA є найбільш точною моделлю за умов регулярної сезонності та стабільності ряду;
- Holt-Winters характеризується вищою стабільністю та меншою чутливістю до аномалій.

Підсумовуючи результати, можна стверджувати, що мета магістерської роботи була досягнута: проведено порівняльний аналіз двох ключових методів прогнозування, розроблено інструмент для практичної роботи з часовими рядами, виконано детальне експериментальне дослідження та сформовано рекомендації щодо вибору відповідних моделей у реальних

умовах. Отримані висновки можуть слугувати основою для подальшого розвитку систем прогнозування у сфері продажів, а також стимулювати впровадження сучасних підходів до аналізу даних на підприємствах.

ПЕРЕЛІК ПОСИЛАНЬ

1. Box G. E. P., Jenkins G. M., Reinsel G. C., Ljung G. M. Time Series Analysis: Forecasting and Control. - Hoboken: Wiley, 2016. - 712 p.
2. Hyndman R. J., Athanasopoulos G. Forecasting: Principles and Practice. - Melbourne: OTexts, 2021. - 378 p.
3. Holt C. C. Forecasting seasonals and trends by exponentially weighted moving averages International Journal of Forecasting. - 2004. - Vol. 20, No. 1. - P. 5–10. (Оригінал опубліковано у 1957 p.)
4. Hamilton J. D. Time Series Analysis. - Princeton: Princeton University Press, 1994. - 799 p.
5. Wei W. W. S. Time Series Analysis: Univariate and Multivariate Methods. - Boston: Pearson, 2006. - 614 p.
6. Python Software Foundation. Python Documentation [Електронний ресурс]. - Режим доступу: (дата звернення: 19.09.2025).
7. McKinney W. Python for Data Analysis: Data Wrangling with pandas, NumPy, and IPython. - Sebastopol: O'Reilly Media, 2022. - 579 p.
8. The pandas development team. pandas Documentation [Електронний ресурс]. Режим доступу: <https://pandas.pydata.org> (дата звернення: 21.09.2025).
9. Seabold S., Perktold J. Statsmodels: Econometric and Statistical Modeling with Python // Proceedings of the 9th Python in Science Conference (SciPy 2010). - Austin, Texas, 2010. -P. 92–96.
10. Statsmodels Development Team. statsmodels Documentation [Електронний ресурс]. - Режим доступу: <https://www.statsmodels.org> (дата звернення: 22.09.2025).

11. U.S. Energy Information Administration (EIA). Petroleum and Other Liquids: Historical Data [Электронный ресурс]. - Режим доступа: <https://www.eia.gov> (дата звернення: 07.10.2025).

12. Kaggle Inc. Store Sales - Time Series Forecasting Dataset [Электронный ресурс]. - Режим доступа: <https://www.kaggle.com/competitions/store-sales-time-series-forecasting> (дата звернення: 15.10.2025).

13. Kaggle Inc. Daily Oil Production Dataset (oil.csv) [Электронный ресурс]. - Режим доступа: <https://www.kaggle.com> (дата звернення: 16.10.2025).

14. GitHub Inc. Time Series Forecasting Examples Repository [Электронный ресурс]. — Режим доступа: <https://github.com> (дата звернення: 02.11.2025).

ДОДАТКИ

Програмний код

Скрипт для підготовки часового ряду:

```
import pandas as pd

train_path = r"D:\train.csv"    # путь к train.csv

store_number = 1                # номер магазина

category = "FROZEN FOODS"      # категория (family)

output_path = fr"D:\series_{store_number}_{category.replace('/', '_')}.csv"

print("Завантажуємо великий файл... Це може тривати кілька секунд.")

df = pd.read_csv(train_path)

print("Фільтруємо дані...")

filtered = df[(df["store_nbr"] == store_number) &
              (df["family"] == category)]

if filtered.empty:

    print("\nДаних не знайдено. Перевір store_nbr або family.")

    exit()
```

```
print(f"Знайдено {len(filtered)} записів. Формуємо часовий ряд...")
```

```
# Групуємо по дате и суммируємо продажі
```

```
series = (
    filtered.groupby("date")["sales"]
    .sum()
    .reset_index()
    .sort_values("date")
)
```

```
# Сохраняем временной ряд
```

```
series.to_csv(output_path, index=False)
```

```
print(f"\nГотово! Часовий ряд збережено у файл:\n{output_path}")
```

```
print("\nПриклад перших рядків:")
```

```
print(series.head())
```

Код основної програми

```
import pandas as pd
```

```

import numpy as np

import matplotlib.pyplot as plt

from statsmodels.tsa.holtwinters import ExponentialSmoothing

from statsmodels.tsa.statespace.sarimax import SARIMAX

# ----- допоміжні функції ----- #

def mape(y_true, y_pred):

    """Середня абсолютна відносна похибка у %."""

    y_true, y_pred = np.array(y_true), np.array(y_pred)

    return np.mean(np.abs((y_true - y_pred) / y_true)) * 100

def mae(y_true, y_pred):

    y_true, y_pred = np.array(y_true), np.array(y_pred)

    return np.mean(np.abs(y_true - y_pred))

def rmse(y_true, y_pred):

```

```

y_true, y_pred = np.array(y_true), np.array(y_pred)

return np.sqrt(np.mean((y_true - y_pred) ** 2))

```

```

def train_test_split_series(series, test_size):

```

```

    """Розділяє часовий ряд на train/test по кількості останніх точок."""

```

```

    return series.iloc[:-test_size], series.iloc[-test_size:]

```

```

# ----- завантаження та підготовка даних ----- #

```

```

def load_and_prepare_series(

```

```

    path, date_col, target_col, date_format=None, freq=None

```

```

):

```

```

    """

```

Завантажує CSV, перетворює стовпець з датою, сортує та формує часовий ряд.

path – шлях до файлу

date_col – назва стовпця з датою

target_col – назва стовпця з показником (продажі)

`date_format` – формат дати (опційно, можна залишити `None`)

`freq` – бажана частота ('D', 'W', 'M' тощо) або `None`, якщо не змінювати

```
"""
```

```
print(f"\nЗавантажуємо дані з файлу: {path}")
```

```
df = pd.read_csv(path)
```

```
if date_format:
```

```
    df[date_col] = pd.to_datetime(df[date_col], format=date_format)
```

```
else:
```

```
    df[date_col] = pd.to_datetime(df[date_col])
```

```
df = df[[date_col, target_col]].dropna()
```

```
df = df.sort_values(by=date_col)
```

```
df = df.set_index(date_col)
```

```
# Якщо задано частоту – агрегуємо до неї (сума за період)
```

```
if freq is not None:
```

```
    df = df.resample(freq).sum()
```

```
series = df[target_col]
```

```

print("Перші рядки підготовлених даних:")

print(series.head())

return series

# ----- модель Holt–Winters ----- #

def run_holt_winters(train, test, seasonal_periods):
    """
    Будує модель Holt–Winters на train та прогнозує на довжину test.
    Повертає прогноз та об'єкт моделі.
    """

    print("\n=== Модель Holt–Winters ===")

    model = ExponentialSmoothing(
        train,
        trend="add",
        seasonal="add",
        seasonal_periods=seasonal_periods
    )

    fit = model.fit(optimized=True)

```

```
forecast = fit.forecast(len(test))
```

```
mae_val = mae(test, forecast)
```

```
rmse_val = rmse(test, forecast)
```

```
mape_val = mape(test, forecast)
```

```
print(f"MAE Holt–Winters: {mae_val:.3f}")
```

```
print(f"RMSE Holt–Winters: {rmse_val:.3f}")
```

```
print(f"MAPE Holt–Winters: {mape_val:.2f}%")
```

```
return forecast, fit
```

```
# ----- модель SARIMA ----- #
```

```
def run_sarima(train, test, order, seasonal_order):
```

```
    """
```

```
    Буде модель SARIMA на train та прогнозує на довжину test.
```

```
    order – (p, d, q)
```

```
    seasonal_order – (P, D, Q, s)
```

```
    """
```

```

print("\n=== Модель SARIMA ===")

print(f"order={order}, seasonal_order={seasonal_order}")

model = SARIMAX(
    train,
    order=order,
    seasonal_order=seasonal_order,
    enforce_stationarity=False,
    enforce_invertibility=False
)

fit = model.fit(dispatch=False)

forecast = fit.forecast(len(test))

mae_val = mae(test, forecast)

rmse_val = rmse(test, forecast)

mape_val = mape(test, forecast)

print(f"MAE SARIMA: {mae_val:.3f}")

print(f"RMSE SARIMA: {rmse_val:.3f}")

print(f"MAPE SARIMA: {mape_val:.2f}%")

```

```
return forecast, fit
```

```
# ----- візуалізація ----- #
```

```
def plot_forecasts(series, train, test, hw_forecast, sarima_forecast):
```

```
    plt.figure(figsize=(10, 6))
```

```
    plt.plot(series.index, series.values, label="Фактичні дані", linewidth=1.5)
```

```
    plt.plot(test.index, hw_forecast, label="Прогноз Holt–Winters",
linestyle="--")
```

```
    plt.plot(test.index, sarima_forecast, label="Прогноз SARIMA",
linestyle=":")
```

```
    plt.legend()
```

```
    plt.xlabel("Час")
```

```
    plt.ylabel("Продажі")
```

```
    plt.title("Порівняння фактичних даних та прогнозів")
```

```
    plt.grid(True)
```

```
    plt.tight_layout()
```

```
    plt.show()
```

```
# ----- головна функція ----- #
```

```

def main():

    print("=== Алгоритм прогнозування продажів (Holt–Winters +
SARIMA) ===")

    path = input("Введіть шлях до CSV-файлу з даними: ").strip()

    date_col = input("Назва стовпця з датою (напр. date): ").strip()

    target_col = input("Назва стовпця з продажами (напр. sales): ").strip()

    freq_input = input(

        "Бажана частота ряду (D – денно, W – тижнево, M – місячно, Enter
– не змінювати): "

    ).strip().upper()

    freq = freq_input if freq_input else None

    test_size = int(input("Скільки останніх періодів відкласти під тест
(напр. 12): "))

    seasonal_periods = int(input("Період сезонності (напр. 7 або 12): "))

    # Підготовка ряду

    series = load_and_prepare_series(path, date_col, target_col, freq=freq)

```

```

if len(series) <= test_size + seasonal_periods:

    raise ValueError("Занадто мало спостережень для заданого розміру
тесту та сезонності.")

train, test = train_test_split_series(series, test_size)

# Holt–Winters

hw_forecast, hw_model = run_holt_winters(train, test, seasonal_periods)

# Базові параметри SARIMA

# order = (p, d, q), seasonal_order = (P, D, Q, s)

order = (1, 1, 1)

seasonal_order = (1, 1, 1, seasonal_periods)

'''

# Варіант 2: без сезонної частини (чиста ARIMA)

order = (1, 1, 1)

seasonal_order = (0, 0, 0, 0)

'''

sarima_forecast, sarima_model = run_sarima(train, test, order,
seasonal_order)

```

```
# Графік порівняння
```

```
plot_forecasts(series, train, test, hw_forecast, sarima_forecast)
```

```
print("\nРоботу завершено.")
```

```
if __name__ == "__main__":
```

```
    main()
```