

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Чорноморський національний університет
імені Петра Могили
Факультет комп'ютерних наук
Кафедра інтелектуальних інформаційних систем

ДОПУЩЕНО ДО ЗАХИСТУ

Завідувач кафедри інтелектуальних
інформаційних систем, д-р техн. наук, проф.

_____Ю. П. Кондратенко

«_____»_____2024 р.

КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

**ІНТЕЛЕКТУАЛЬНА СИСТЕМА ПРОГНОЗУВАННЯ ЦІН
НА АГРАРНУ ПРОДУКЦІЮ В УКРАЇНІ НА БАЗІ
ANDROID-ЗАСТОСУНКУ**

Спеціальність 122 «Комп'ютерні науки»

122 – КРМ – 601.21810126

Виконав студент 6-го курсу, групи 601

_____ *А. О. Буданов*

«20» лютого 2024 р.

Керівник: канд. техн. наук, доцент

_____ *І. О. Калініна*

«20» лютого 2024 р.

Миколаїв – 2024

Чорноморський національний університет ім. Петра Могили
Факультет комп'ютерних наук
Кафедра інтелектуальних інформаційних систем

Освітньо-кваліфікаційний рівень **магістр**

Галузь знань **12 «Інформаційні технології»**

(шифр і назва)

Спеціальність **122 «Комп'ютерні науки»**

(шифр і назва)

ЗАТВЕРДЖУЮ

Завідувач кафедри інтелектуальних
інформаційних систем, д-р техн. наук, проф.

_____ Ю. П. Кондратенко

« ____ » _____ 20__ р.

ЗАВДАННЯ

на виконання кваліфікаційної роботи

Буданову Андрію Олеговичу

(прізвище, ім'я, по батькові)

1. Тема кваліфікаційної роботи «Інтелектуальна система прогнозування цін на аграрну продукцію в Україні на базі Android-застосунку».

Керівник роботи Калініна Ірина Олександрівна, кандидат техн. наук, доцент.

Затв. наказом Ректора ЧНУ ім. Петра Могили від «19» жовтня 2023 р. № 201

2. Строк подання студентом роботи 19 лютого 2024 р.

3. Вхідні (початкові) дані до роботи: часові ряди цін на аграрну продукцію в Україні. Очікуваний результат: моделі прогнозування часових рядів з відображенням результату в застосунку Android.

4. Зміст пояснювальної записки (перелік питань, які потрібно розглянути):

- аналіз сучасного стану ринку аграрних культур в Україні;
- огляд та опис компонентів часового ряду;
- огляд існуючих методів прогнозування часових рядів;
- побудова моделей прогнозування ARIMA та Хольта-Вінтерса;
- розробка системи прогнозування цін на аграрну продукцію на базі Android-застосунку.

5. Перелік графічного матеріалу: презентація.

6. Завдання до спеціальної частини: Опис мікрокліматичних умов робочого приміщення.

• 7. Консультанти:

Розділ	Прізвище, ініціали та посада консультанта	Підпис
Спеціальна частина з охорони праці	Григор'єва Л. І. доктор біологічних наук, професор	
Методична частина	Калініна І. О. кандидат технічних наук, доцент	

Керівник роботи кандидат технічних наук, доцент Калініна І. О.
(наук. ступінь, вчене звання, прізвище та ініціали)

_____ (підпис)

Завдання прийнято до виконання Буданов А. О.
(прізвище та ініціали)

_____ (підпис)

Дата видачі завдання « 31 » жовтня 2023 р.

КАЛЕНДАРНИЙ ПЛАН виконання кваліфікаційної роботи магістра

Тема: Інтелектуальна система прогнозування цін на аграрну продукцію в Україні на базі Android-застосунку

№	Найменування роботи	Початок	Закінчення	Примітки
1	Визначення керівника і теми КРМ. Подання заяви на затвердження теми КРМ	31.10.2023	10.11.2023	Виконано
2	Отримання завдання на виконання КРМ	11.11.2023	11.11.2023	Виконано
3	Складання календарного плану на період виконання КРМ	12.11.2023	12.11.2023	Виконано
4	Огляд літератури за темою дослідження	14.11.2023	26.11.2023	Виконано
5	Проходження передатестаційної практики, збір та аналіз матеріалів до КРМ	27.11.2023	23.12.2023	Виконано
6	Аналіз предметної області та розробка технічного завдання. Моделювання результатів	25.12.2023	12.01.2024	Виконано
7	Опис фахової частини КРМ, зокрема дослідження публікацій щодо моделювання прогнозів часових рядів, огляд існуючих моделей прогнозування часових рядів, методів подачі інформації інструментами системи Android для вирішення поставленої задачі, реалізація обраних технологій з аналізом отриманих результатів	13.01.2024	25.01.2024	Виконано
8	Розробка спеціальної частини з охорони праці та методичної частини	26.01.2024	02.02.2024	Виконано
9	Перший попередній захист КРМ на засіданні комісії кафедри	29.01.2024	29.01.2024	Виконано
10	Корегування роботи за результатами попереднього захисту	30.01.2024	05.02.2024	Виконано
11	Доробка та остаточне оформлення КРМ	06.02.2024	11.02.2024	Виконано
12	Другий попередній захист КРМ на засіданні комісії кафедри	12.02.2024	12.02.2024	Виконано
13	Подання КРМ, її електронної копії та інших документів (відгуку, рецензії) до захисту	19.02.2024	20.02.2024	Виконано
14	Захист КРМ перед екзаменаційною комісією (ЕК)	27.02.2024	27.02.2024	Виконано

Розробив студент Буданов А. О.
(прізвище та ініціали)

(підпис)

Керівник роботи кандидат технічних наук, доцент, Калініна І. О.
(наук. ступінь, вчене звання, прізвище та ініціали)

(підпис)

«09» листопада 2023 р.

АНОТАЦІЯ

до кваліфікаційної роботи магістра
студента групи 601 ЧНУ ім. Петра Могили

Буданова Андрія Олеговича

на тему: **“ІНТЕЛЕКТУАЛЬНА СИСТЕМА ПРОГНОЗУВАННЯ ЦІН НА АГРАРНУ ПРОДУКЦІЮ В УКРАЇНІ НА БАЗІ ANDROID-ЗАСТОСУНКУ”**

Актуальність даного дослідження полягає у підвищенні ефективності прогнозування цін на аграрну продукцію в Україні, шляхом дослідження та виявлення кращих моделей прогнозування для окремих аграрних культур, що надасть корисну підтримку прийняття рішень для різних учасників ринку, включаючи інвесторів, фермерів та агропромислових компаній. Розробка Android-застосунку позитивно відображає використання сучасних технологій у фінансовій галузі та допомагає користувачам отримувати оперативну інформацію незалежно від доступу до стаціонарних ЕОМ, сприяючи цифровій трансформації аграрного та фінансового сектору.

Об’єктом є процес дослідження статистичних даних цін на аграрну продукцію в Україні.

Предметом дослідження та розробка методів та алгоритмів побудови моделей прогнозування і користувацького інтерфейсу в системі Android.

Метою дослідження є створення інтелектуальної інформаційної системи прогнозування цін на аграрну продукцію на базі Android-застосунку.

В результаті виконання роботи було досліджено дві моделі прогнозування часових рядів, а саме: ARIMA та модель Хольта-Вінтерса. Проаналізована робота алгоритмів, що реалізують моделі, визначені основні їх переваги та недоліки, а також розроблено програмне забезпечення, в якому реалізовані відповідні методи, та створено Android застосунок, що реалізує інтерфейс користувача для роботи з отриманими даними.

Дана робота складається з чотирьох розділів. Кожен розділ відповідно присвячений: опису предметної сфери, моделям і методам, використаним у магістерській роботі, моделюванню і проектуванню системи прогнозування цін,

програмній реалізації та тестуванню отриманих результатів, аналізу отриманих результатів, охороні праці, методичній частині магістерської роботи. Загальний обсяг роботи – 66 сторінок. Кваліфікаційна робота магістра містить 37 рисунків, 3 таблиць і посилання на 22 літературних джерел.

Ключові слова: алгоритм, ARIMA-модель, модель Хольта-Вінтерса, Android-застосунок, часовий ряд, інтелектуальна система, інформаційна система.

ABSTRACT

to the master's qualification work by the student of the group 601 of Petro Mohyla
Black Sea National University

Budanov Andrii

“INTELLIGENT SYSTEM FOR FORECASTING PRICES FOR AGRICULTURAL PRODUCTS IN UKRAINE BASED ON THE ANDROID APPLICATION”

The relevance of this research lies in improving the efficiency of forecasting agricultural product prices in Ukraine by exploring and identifying the best forecasting models for specific agricultural crops. This will provide valuable support for decision-making for various market participants, including investors, farmers, and agro-industrial companies. The development of an Android application positively reflects the use of modern technologies in the financial sector and helps users access timely information regardless of access to stationary computers, contributing to the digital transformation of the agricultural and financial sectors.

The object of the study is the process of researching statistical data on prices for agricultural products.

The subject of the research is the development of methods and algorithms for building forecasting models in the Android system.

The aim of the research is to create an intelligent information system for forecasting prices for agricultural products based on Android.

As a result of the work, two time series forecasting models were investigated: the ARIMA model for non-seasonal time series and the Holt-Winters model for time series with trend and seasonality. The functionality of the algorithms implementing the models was analyzed, their main advantages and disadvantages were identified. Additionally, software was developed, incorporating the relevant methods, and an Android application was created to implement the user interface for working with the obtained data.

This thesis consists of four sections, each dedicated to: describing the subject area, the models and methods used in the master's thesis, modeling and designing the price forecasting system, software implementation and testing of the obtained results, analyzing the results, occupational safety, and the methodological part of the master's thesis. The total volume of the work is 66 pages. The master's qualification paper includes 37 figures, 3 tables, and references to 22 literary sources.

Keywords: algorithm, ARIMA model, Holt-Winters model, Android application, time series, intelligent system, information system.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ.....	4
ВСТУП.....	5
1 АНАЛІЗ ПРЕДМЕТНОЇ СФЕРИ.....	6
1.1 Аналіз агрокультурного експорту, як частини економіки України	6
1.2 Часовий ряд.....	7
1.3 Аналіз часових рядів	11
1.4 Стаціонарність часового ряду.....	12
1.5 Функція автокореляції та функція правдоподібності.....	14
1.6 Аналіз вибору системи Android для реалізації інтерфейсу користувача	15
Висновок до розділу 1	16
2. МЕТОДИ ТА МОДЕЛІ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ	18
2.1 Метод прогнозування часового ряду ARIMA.....	18
2.2 Модель Хольта.....	21
2.3 Метод Хольта-Вінтерса.....	23
Висновок до розділу 2	24
3 МОДЕЛЮВАННЯ ТА ДОСЛІДЖЕННЯ ОТРИМАННИХ РЕЗУЛЬТАТІВ	26
3.1 Формування набору даних	26
3.2 Пошук аномальних значень у часових рядах.....	28
3.3 Побудова моделі прогнозування ARIMA.....	37
3.4 Побудова моделі прогнозування Хольта-Вінтерса.....	48
Висновок до розділу 3	49
4 ПРОГРАМНА РЕАЛІЗАЦІЯ	51

4.1	Опис основних технологій	51
4.2	Огляд допоміжних бібліотек для побудови моделей прогнозування ..	52
4.3	Опис допоміжних бібліотек для побудови користувацького інтерфейсу Android-застосунку.....	54
4.4	Налаштування локального серверу для передачі інформації між R та Kotlin	56
4.5	Розробка інтерфейсу користувача Android-застосунку.....	58
	Висновок до розділу 4.....	61
	ВИСНОВКИ	63
	ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ	64

ПЕРЕЛІК СКОРОЧЕНЬ

ARIMA	Autoregressive Integrated Moving Average
ACF	Auto Correlation Function (автокореляційна функція)
AIC	Akaike Information Criterion (інформаційний критерій Акайке)
ARMA	Autoregressive moving-average model
PACF	Partial Auto Correlation Function
HW	Holt-Winters model

ВСТУП

Сільськогосподарський сектор в сучасному світі, особливо в Україні є ключовим елементом економіки, і ціни на сільськогосподарську продукцію, зокрема на аграрну продукцію, є важливим фактором, який впливає на економічну стійкість країни та сільськогосподарських виробників. Однак непередбачувані коливання цін на зернові культури можуть стати викликом для планування виробництва, ресурсного управління та прийняття стратегічних рішень в агропромисловому комплексі.

З урахуванням динамічної природи ринку цін на аграрну продукцію та їхнього залежного від часу характеру, аналіз часових рядів стає ключовим інструментом для прогнозування майбутніх цінових тенденцій. У даній роботі розглядаються різні методи аналізу часових рядів та варіанти їх застосування для прогнозування цін на аграрну продукцію. Вибір оптимального методу аналізу є важливою задачею для точного та ефективного прогнозування.

У роботі було розглянуто традиційні методи аналізу часових рядів, такі як ARIMA, метод Хольта, а також метод Хольта-Вінтерса. Кожен метод було детально проаналізовано щодо його суті, переваг та обмежень.

Вивчення цих методів дозволить визначити оптимальний підхід для прогнозування тренду ціноутворень аграрної продукцію, що в свою чергу сприятиме розробці інструментів для ефективного управління ризиками та прийняття обґрунтованих рішень в агросекторі.

1 АНАЛІЗ ПРЕДМЕТНОЇ СФЕРИ

1.1 Аналіз агрокультурного експорту, як частини економіки України

Аграрний сектор економіки України є міжнародної візитівкою держави, завдяки значному внеску України у світовий ринок продовольства. Зараз цей сектор становить майже 1/4 обсягу валового внутрішнього продукту (ВВП) і відзначається високим рівнем виробництва. Останні роки відзначаються спеціалізацією та концентрацією в аграрному виробництві, що сприяє розвитку експорту. Цей успіх аграрного сектору обумовлений не лише високим рівнем виробництва, але й сприятливими законодавчими ініціативами. Україна приєдналася до міжнародних аграрних ринків, що сприяє зростанню економічних можливостей. Це свідчить про стратегічну важливість аграрного сектору для економічного розвитку країни.

За даними Державної служби статистики, у довоєнний період частка продукції агро-промислового комплексу в загальній структурі експорту товарів з України значно зросла. У 2014 році вона складала 31%, а в 2021 році вже досягла 41%, що в еквіваленті становить 27,8 млрд доларів США. У 2022 році цей показник ще зріс до 53%, а за перше півріччя 2023 року — до 61%. Цей стрімкий ріст може бути пояснений активною підтримкою зовнішніх партнерів України, які сприяють стабілізації її економіки, зокрема, через підтримку сектору експорту.

Згідно з чинною Стратегією експортної політики України [1], основними її цілями є підвищення конкурентоспроможності продукції та розширення товарної номенклатури експорту; диверсифікація ринків збуту, підтримка і розвиток існуючих ринків; просування українського бренду та інформаційно-аналітичне забезпечення аграрного експорту. Востаннє стратегія була відредагована 12 травня 2023 року. Однак, через постійні зміни ситуації з аграрним експортом, деякі аспекти залишаються недостатньо висвітленими та потребують подальшого уточнення та адаптації.



Рисунок 1.1 – Структура експорту аграрних культур в Україні за січень-жовтень 2023

За оцінками науковців Інституту аграрної економіки, обсяги експорту в 2023 році могли б стати рекордними для України за всю історію незалежності. Однак внаслідок втрат логістики і виробництва через військову агресію, експорт товарів зменшився на 35% порівняно з 2022 роком.

1.2 Часовий ряд

Часовий ряд - це послідовність даних, зібраних, виміряних чи обчислених у різних моментах часу. Такий ряд представляє собою велику кількість спостережень, зібраних в різні моменти часу t , і може включати денні, щотижневі, місячні або інші вимірювання. Такий набір даних позначається, як Z_t . Часові ряди зазвичай використовуються для аналізу та прогнозування явищ, які змінюються в часі, таких як ціни на товари, температура, фінансові показники, обсяги продажів тощо.

На часовий ряд загалом впливають три основні компоненти, які можуть бути виділені зі спостережуваних даних. Ці компоненти включають: тренд, сезонні варіації, випадкові коливання.

$$y_t = tr_t + s_t + r_t,$$

де tr_t – це тренд, s_t – це складова, що називається сезонністю, r_t – це випадкові зміни на часовому ряді.

Тренд - це загальна тенденція до збільшення, зменшення або стагнації числового ряду протягом тривалого періоду часу. Тренд є довгостроковою характеристикою часового ряду. Наприклад, ряд показників, що стосуються зростання чисельності населення, кількості осіб, що мають відношення до будівництва в місті тощо, мають тенденцію до зростання, тоді як низхідна тенденція може бути пов'язана з рівнем смертності, епідеміями і т.д. Класифіковано декілька видів трендів часового ряду, наприклад – детермінований тренд. Якщо висловити детермінований тренд за допомогою полінома будь-якого порядку відносно часу:

$$y(k) = a_0 + a_1 * k + a_2 * k^2 + \dots + a_p * k^n + \varepsilon(k), \quad E[\varepsilon(k)] = 0,$$

то визначення прогнозу тренду зводиться до підстановки в це рівняння бажаного значення часу k у рівняння, що виражає детермінований тренд у вигляді полінома будь-якого порядку, та застосування безумовного математичного очікування. Прогнозування значень тренду вважається довгостроковим, оскільки детермінований тренд вказує на довгострокові зміни в процесах. Обмеження на випадковий процес $E[\varepsilon(k)] = 0$ необхідне для коректного застосування методів оцінювання моделей, а також для подальшого аналізу якості оцінок прогнозів. Це вимога буде виконана автоматично, якщо обраний метод оцінювання параметрів моделі відповідає відомим статистичним характеристикам модельованого процесу.

Для моделювання і передбачення стохастичного тренду можна використовувати, наприклад, відоме рівняння випадкового кроку з перетином.

$$y(k) = a_0 + y(k - 1) + \varepsilon(k),$$

розв'язок якого має вигляд:

$$y(k) = a_0 + k * a_0 + \sum_{i=1}^k \varepsilon(i),$$

де $\sum_{i=1}^k \varepsilon(i)$ описує випадковий елемент тренду. Цю компоненту також називають нерегулярною.

Сезонні коливання в часовому ряді відображають зміни, які відбуваються протягом конкретного сезону. Наприклад, в осінній та зимні періоди може спостерігатися збільшення продажу аграрної продукції, тоді як восени та влітку попит зменшується. Це пов'язано із особливостями розглядаємого товару, так як, наприклад, зерно сіють весною та збирають восени. Для підприємців, власників магазинів і виробників важливо враховувати ці сезонні варіації при плануванні діяльності на майбутнє. Наприклад, на рис. 1.2 зображено коливання цін на кукурудзу, де можна побачити, що пікового значення вона досягає у осінній сезон.

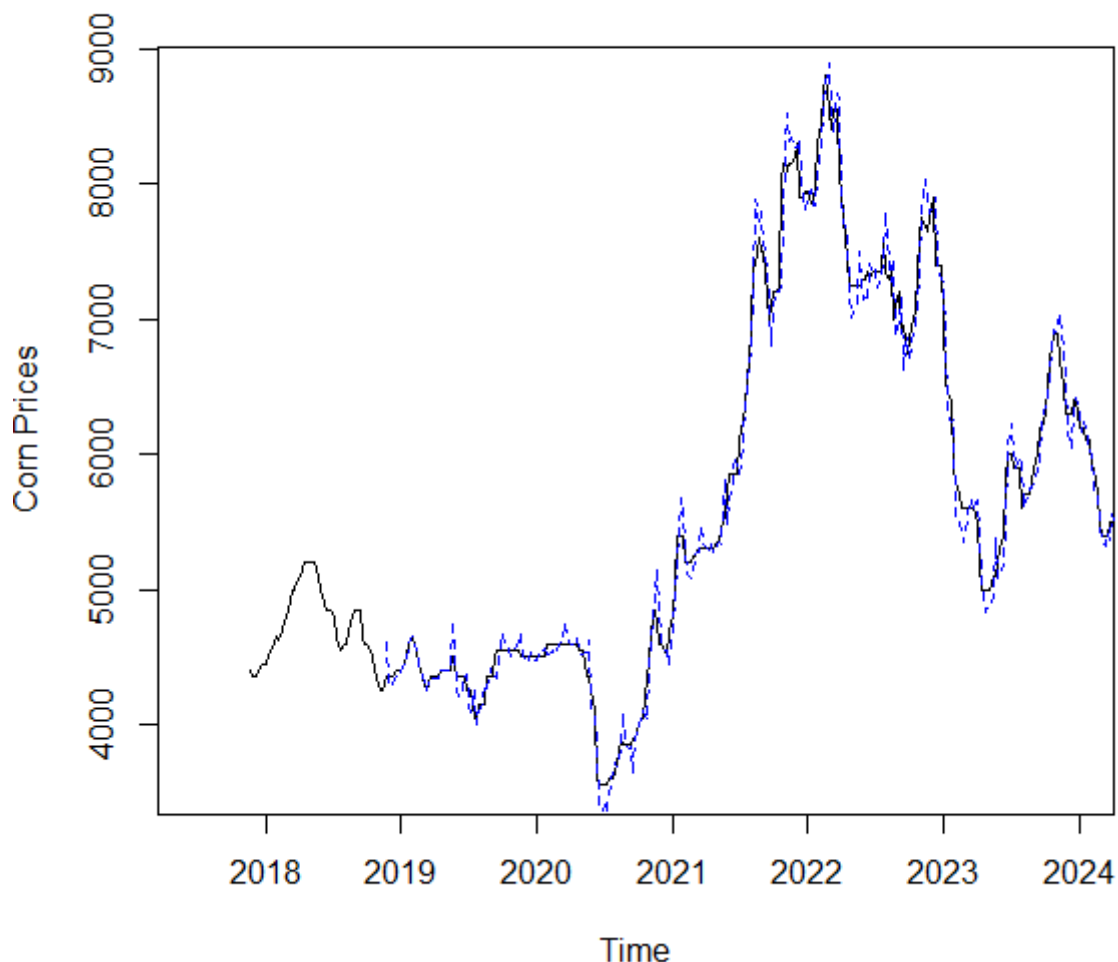


Рисунок 1.2 – графічне зображення кривої цін кукурудзи, згідно з даними «Держстату»

Циклічні зміни в часовому ряді відображають середньострокові зміни, обумовлені обставинами, які повторюються в циклах. Такі цикли мають тривалість більше ніж один рік, зазвичай два чи більше. Багато економічних і фінансових часових рядів характеризуються циклічними коливаннями, які можуть впливати на ринкові умови та бізнес-процеси.

Розуміння сезонних і циклічних варіацій допомагає підприємцям і аналітикам краще адаптувати свої стратегії до змін у попиті та умовах ринку, покращуючи таким чином планування та прийняття рішень.

Важливою характеристикою часового ряду є випадкові зміни в часовому ряді, що є частиною варіацій, яку не можна пояснити або передбачити за допомогою тренду, сезонності або циклічності. Ці зміни представляють собою непередбачувані випадкові впливи або помилки, які можуть виникати з різних причин. Розглядаючи агрокультурну сферу, можна виділити наступні випадкові чинники: зміни в кліматі, такі як надмірні опади, суховії, або інші погодні умови, можуть впливати на врожайність і викликати випадкові коливання цін; природні катастрофи, такі як повені, землетруси, лісові пожежі або урагани, можуть призвести до значних збитків для сільськогосподарської продукції; випадкові події на ринку, такі як фінансові кризи, політичні нестабільності, торгові та військові конфлікти або зміни у валютних курсах; зміни в сільськогосподарських технологіях, використання нових видів насіння, агрохімікатів чи обладнання можуть впливати на врожайність; запаси продукції на ринку та їх рівень можуть визначати споживання та ціни; розповсюдження епідемій або захворювань серед сільськогосподарських культур; рішення уряду, такі як субсидії, тарифи, експортні обмеження чи інші форми регулювання ринку; зміни в світових тенденціях виробництва та споживання, такі як зростання населення, зміни в дієтичних звичках чи зростання вимог до біоенергетики. На рис. 1.3 упорядковані дані з ресурсу «УкрАгроКонсалт», де видно впливання випадкових подій, у даному випадку вторгнення в Україну [2], на ціни різних аграрних культур.



Рисунок 1.3 – графік коливання цін на агрокультури, в Україні із поясненням впливу випадкових подій на ціни

Розуміння та врахування цих факторів дозволяє аналітикам будувати більш точні моделі з використанням часових рядів та аналізувати і передбачати цінові коливання в агросекторі, щоб приймати відповідальні рішення в умовах невизначеності. Важливо вміти виявляти та мінімізувати вплив випадкових змін при аналізі часового ряду. Це може включати в себе застосування таких методів, як: згладжування, фільтрації шуму, виявлення аномалій. Використання статистичних методів для виявлення аномалій і видалення випадкових впливів може покращити точність аналізу.

1.3 Аналіз часових рядів

Аналіз часових рядів - це метод дослідження та вивчення часових рядів. Цей підхід дозволяє виявляти закономірності, тренди, цикли та інші характеристики, що можуть бути важливими для розуміння поведінки процесів часових рядів. Аналіз часових рядів використовує різноманітні методи, такі як статистичні техніки, моделювання, спектральний аналіз та інші, для отримання інсайтів та прогнозування майбутніх значень.

Стаціонарність вказує на те, що статистичні властивості ряду, такі як середнє значення та дисперсія, залишаються сталими впродовж часу. Це передбачає відсутність тренду, сезонності та циклічності у даних. В стаціонарному ряді відсутній загальний тренд у певному напрямку. Тренд може бути горизонтальним або коливатися навколо фіксованого рівня. Якщо в ряду присутні сезонні та циклічні зміни, то їхні параметри (амплітуда, тривалість) залишаються сталими. Більшість аналітичних методів передбачення базуються на припущенні стаціонарності. Моделі, побудовані для стаціонарних рядів, можуть бути більш надійними та точними, що дозволяє краще розуміти динаміку даних та легше виявляти основні закономірності. Не всі часові ряди є стаціонарними, і іноді необхідно застосовувати методи диференціювання чи інші техніки, щоб зробити їх стаціонарними перед застосуванням аналітичних методів.

1.4 Стаціонарність часового ряду

Для кожного часового ряду можна визначити ряд числових характеристик. За методами математичної статистики, для часового ряду, як і для будь-якого набору даних, обчислюють основні моменти. У контексті аналізу часових рядів ключовими моментами є математичне сподівання, дисперсія та коваріація. Математичне сподівання часового ряду $\{Y_t\}$ є функція:

$$\mu_t = E y_t = \int_{-\infty}^{+\infty} x dF_t(x),$$

де $F_t(x) = P(y_t < x)$ – це функція розподілу $y_t, t = \overline{1, T}$.

Коваріація часового ряду $\{Y_t\}$ визначається, як:

$$\text{cov}(y_t, y_{t-j}) = E(y_t - \mu_t)(y_{t-j} - \mu_{t-j}), \quad j = 1, 2, \dots$$

Дисперсія часового ряду визначається за формулою:

$$D(y_t) = E(y_t - \mu_t)^2$$

Часовий ряд вважається стаціонарним, якщо:

- математичне сподівання $\mu_t = \mu < \infty$ для всіх $t = \overline{1, T}$;
- дисперсія $D(y_t) = \gamma_0 < \infty$ для всіх $t = \overline{1, T}$;

- коваріація j -порядку $cov(y_t, y_{t-j}) = \gamma_0 < \infty$ для всіх $t = \overline{1, T}$, $j = 1, 2, 3, \dots$

Отже, у стаціонарного часового ряду очікувані значення та дисперсія повинні залишатися постійними та обмеженими для кожного періоду часу, а коваріація будь-якого порядку повинна мати постійну та обмежену величину в усі моменти часу.

Оцінками цих моментів часового ряду є статистики:

- математичне сподівання

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T y_t;$$

- дисперсія

$$\hat{\mu}_0 = \frac{1}{T-1} \sum_{t=1}^T (y_t - \hat{\mu})^2$$

- коваріація j -порядку

$$\hat{\mu}_j = \frac{1}{T-j} \sum_{t=j+1}^T (y_t - \mu_t) (y_{t-j} - \hat{\mu})$$

Подані вище значення можуть значно відрізнятися для різних часових рядів. Розробка методів для аналізу такої різноманітної інформації була б важкою. Таким чином, була розроблена класифікація часових рядів залежно від їх числових характеристик. Однією з найважливіших категорій є поняття стаціонарності. Зрозуміло, що жоден ряд, що відображає реальну інформацію про процес у часі, не може бути ідеально стаціонарним. Але якщо для конкретного часового ряду умови стаціонарності виконуються з певним наближенням, то для його аналізу можна використовувати різноманітні методи аналізу та прогнозування стаціонарних часових рядів.

1.5 Функція автокореляції та функція правдоподібності

Автокореляція вказує на взаємозв'язок між значеннями ряду у різні моменти часу. Автокореляційна функція (далі АСФ) вимірює ступінь подібності між значеннями ряду для різних відставань. Формально це виражається як кореляція між значеннями Y_t та Y_{t-k} де k - це відставання.

$$p_j = \frac{cov(y_t, y_{t-j})}{D(y_t)}$$

Цей коефіцієнт вказує на ступінь взаємозалежності між спостереженнями, розташованими на відстані j періодів. Якщо побудувати графік залежності p_j , отримаємо корелограму. Вона відображає певну криву, що демонструє, як змінюється взаємовплив між спостереженнями залежно від часового інтервалу між ними. Розглядаються кілька можливих сценаріїв:

- випадковий процес: якщо маємо справу з абсолютно випадковим процесом із значущою кількістю даних, то для усіх ненульових значень j коефіцієнт автокореляції p_j буде дорівнювати 0;
- короткотермінова залежність: у стаціонарних процесах часто спостерігається короткотермінова залежність між спостереженнями, виражається тим, що кілька перших коефіцієнтів автокореляції мають відносно великі за абсолютним значенням в порівнянні з усіма іншими, які наближаються до нуля. Якщо перші коефіцієнти є позитивними, то існує позитивна автокореляція; якщо ж перші коефіцієнти автокореляції постійно змінюють свій знак, то спостерігається негативна автокореляція;
- нестационарні процеси: у випадку часових рядів, що мають виразно виражений компонент тренду, значення p_j не будуть зближатися до нуля, навіть при великих j . Таким чином, зазвичай визначення функції автокореляції використовують тільки для стаціонарних часових рядів;
- сезонні коливання: при наявності сезонних коливань корелограма, як правило, буде відображати зменшення амплітуди хвиль з однаковою частотою.

Підкресливши сезонний компонент, функція автокореляції реально відобразить взаємозалежність спостережень у часі.

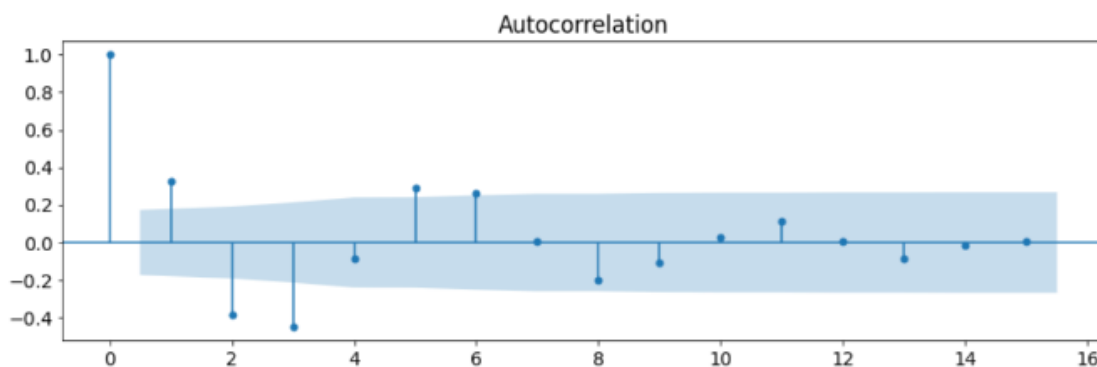


Рисунок 1.4 – Приклад графіку автокореляції

Графік АСF дозволяє візуально оцінити кореляційну структуру часового ряду. На графіку вісь x відповідає відставанню, а вісь y представляє кореляційні коефіцієнти. Значення автокореляційних коефіцієнтів знаходяться в діапазоні від -1 до 1. Значення близькі до 1 вказують на сильну позитивну кореляцію, близькі до -1 - на сильну негативну кореляцію, а значення близькі до 0 - на відсутність кореляції. Періодичність у графіку АСF може вказувати на наявність сезонності або інших циклічних патернів у даних. Якщо значення АСF близьке до нуля для більшості лагів, це може свідчити про відсутність систематичної кореляції в даних. Графік АСF може допомагати в ідентифікації моделей для прогнозування часового ряду. Зміни в АСF можуть вказувати на необхідність застосування різних методів згладжування, диференціації та інших.

1.6 Аналіз вибору системи Android для реалізації інтерфейсу користувача

Використання системи Android для розробки користувацького інтерфейсу є вигідним за кількома причинами. Android - одна з найпопулярніших операційних систем для мобільних пристроїв, що забезпечує широкий охоплення аудиторії користувачів.

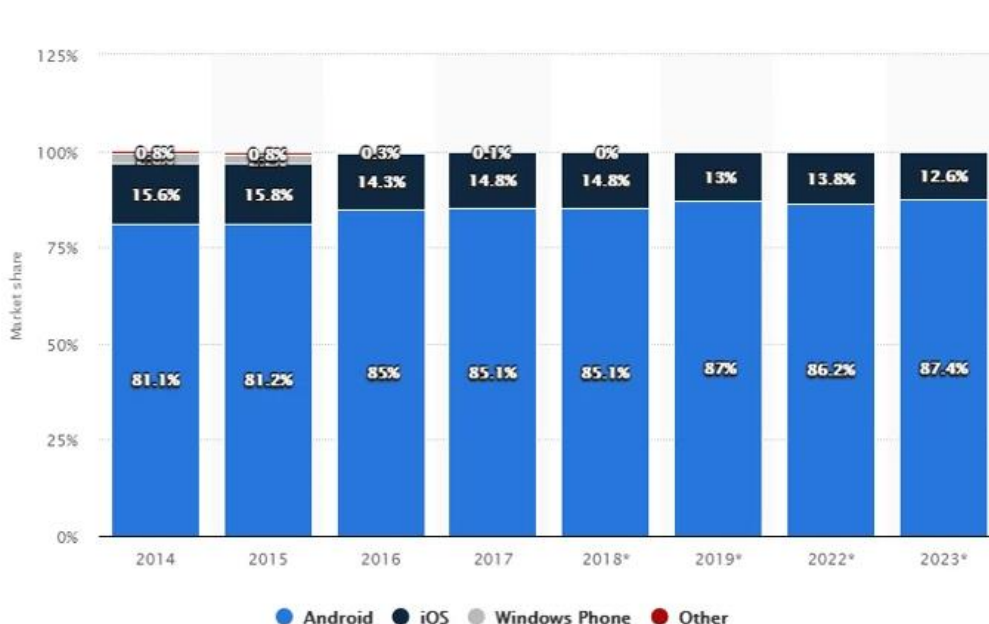


Рисунок 1.5 – Діаграма користувачів пристроїв системи Android

Будучи платформою з відкритим кодом, Android дозволяє швидке виявлення помилок і розвиток нових функцій завдяки активній спільноті розробників. Він також працює на різних типах пристроїв, адаптуючи інтерфейс для різних екранів. Розробка для Android надає більше можливостей для налаштувань і персоналізації інтерфейсу користувача і забезпечує розширені можливості інтеграції. Додатково, Android підтримує широкий спектр мов і регіональних налаштувань, що робить його ідеальним для проектів з міжнародною аудиторією.

Висновок до розділу 1

У першому розділі було розглянуто поняття часових рядів та їх компонентів. Це послідовність значень, вимірюваних або спостережених в послідовні моменти часу, що грають важливу роль у багатьох галузях, таких як економіка, фінанси, метеорологія та інші, оскільки надають можливість аналізу і прогнозування динаміки явищ з плином часу.

Компоненти часових рядів включають тренд, сезонні коливання та випадкові зміни. Тренд представляє собою загальний напрям часового ряду, сезонні коливання пов'язані з циклічними змінами в даних, а стохастичні зміни

відображають випадкові або непередбачувані фактори. Вивчення цих компонентів та їх взаємодії дозволяє здійснювати аналіз та прогнозування, що є важливим для прийняття обґрунтованих рішень в прогнозуванні аграрної продукції.

Також було розглянуто обґрунтування вибору платформи Android для розробки інтерфейсу користувача. Android є широко поширеною та популярною платформою, що забезпечує доступ до великої аудиторії користувачів. Відкритий код і наявність інструментів розробки дозволяють ефективно створювати інтерфейси, адаптовані для різних пристроїв. Зручність інтеграції, розширені можливості персоналізації та гнучкість вибору мов і регіональних налаштувань роблять Android вигідним вибором для розробки інтерфейсу застосунку, що відображає результати прогнозування.

2. МЕТОДИ ТА МОДЕЛІ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

2.1 Метод прогнозування часового ряду ARIMA

Метод середнього значення є одним з простих методів прогнозування часових рядів і використовується для отримання базового прогнозу, який припускає, що майбутні значення будуть подібні до середнього арифметичного минулих значень. Він не враховує випадкові події, що можуть негативно вплинути на результат прогнозування. Прогноз на кожний майбутній період розраховується, як середнє арифметичне всіх попередніх значень часового ряду. Математично цей метод можна представити наступним чином:

Дано часовий ряд: Y_1, Y_2, \dots, Y_t , то прогноз буде розрахований, як

$$Y_{t+1} = \frac{1}{t} \sum_{i=1}^t Y_i,$$

де Y_{t+1} – це прогнозоване значення для періоду $t + 1$, а t – це кількість попередніх значень

Метод середнього значення добре підходить для короткострокових прогнозів, коли можна припустити, що тенденція чи сезонні зміни залишаються сталими, але не адаптується до змін у структурі часового ряду та не враховує тенденції чи сезонність. Його простота дозволяє швидко отримати прогнози, але його точність може бути обмеженою у порівнянні з більш складними методами. Його можна використати, якщо відомі щоденні дані про продажі за останній місяць, то прогноз на наступний день можна отримати, просто взявши середнє арифметичне всіх денних значень. Метод середнього значення використовується як базова лінія для порівняння та визначення ефективності інших складніших методів прогнозування. Це може бути корисним у випадках, коли структура даних ще не повністю зрозуміла або коли необхідно швидко отримати простий прогноз.

Метод згладжування використовується для вирівнювання коливань в часових рядах та виділення основного тренду, який може слугувати основою для подальшого прогнозування. Цей метод ставить за мету зменшити або вилучити

шум та крайні коливання, щоб виділити загальний характер ряду. У методі експоненційного згладжування використовується зважена сума попередніх значень з ваговими коефіцієнтами, які експоненційно зменшуються. Ці коефіцієнти забезпечують більшу вагу останнім значенням. Математично можна представити наступною формулою:

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

Де α – це ваговий коефіцієнт. Цей метод застосовується для відображення та прогнозування тренду та сезонності.

В методі ковзного середнього для кожного моменту часу розраховується середнє значення з певної кількості попередніх точок. Формула, за якої він розраховується, має наступний вигляд:

$$\frac{1}{n} \sum_{i=1}^n Y_{t-i}$$

Де n – це розмір проміжку ковзного середнього. Цей метод використовується для вирівнювання шумів та виділення тренду.

Серед ключових відмінностей цих двох методів можна визначити, що для методу ковзного середнього потрібно вибрати розмір вікна, тоді як метод експоненційного згладжування використовує параметр згладжування α . Метод експоненційного згладжування може бути адаптивним, оскільки кожен новий пункт додається з новим ваговим коефіцієнтом. Слід зазначити, що деякі методи комбінують ковзне середнє з експоненційним згладжуванням для підвищення ефективності прогнозування, так як додавання сезонності та інших компонентів може поліпшити точність прогнозу.

Статистичний метод аналізу та прогнозування часових рядів ARIMA поєднує в собі три основні компоненти: авторегресію (AR), інтеграцію (I) та ковзне середнє (MA). Він дозволяє моделювати складні структури часових рядів, включаючи тренди, сезонність та інші закономірності. Авторегресія вказує на те, що поточне значення часового ряду залежить від його попередніх значень. У моделі ARIMA це

виражається у вигляді лінійної комбінації попередніх значень ряду. Інтеграція вказує на те, що для створення ряду стаціонарним, необхідно використовувати різницю між значеннями ряду. Це може включати перший рядок диференціювання. Ковзне середнє включає у себе вагову суму попередніх значень шуму часового ряду. Воно враховує систематичні помилки у попередніх точках. ARIMA(p, d, q) - це загальний вигляд моделі, де p, d і q - це параметри, що відповідають за авторегресію, інтеграцію та ковзне середнє відповідно, математично модель приймає вигляд формули:

$$Y'_t = \phi_1 Y'_{t-1} + \phi_2 Y'_{t-2} + \dots + \phi_p Y'_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

Де ϕ – це коефіцієнти авторегресії, ε – це шумові компоненти, θ – це коефіцієнти ковзного середнього. Для підбору оптимальних параметрів можна використовувати критерій Байєса. Для побудови моделі визначений ряд вимог, а саме:

- ряд має бути стаціонарним, тобто не повинен містити систематичних змін у часі;
- визначаються параметри p, d, q за допомогою статистичних методів або дослідницької практики;
- модель піддається оцінці за допомогою навчання на частині даних та перевірки на іншій частині;
- після побудови моделі можна використовувати її для прогнозування майбутніх значень ряду.

На мій погляд використання моделі ARIMA для прогнозування цін на агрокультуру є вигідним з кількох поглядів, оскільки ця методика може забезпечити ефективні та точні результати в аналізі та передбаченні цінових тенденцій в аграрному секторі. ARIMA може автоматично враховувати відстрочення попередніх значень та їх ваги, що дозволяє моделі адекватно реагувати на попередні цінові рухи. Авторегресійна складова ARIMA відображає залежність поточних цін на агрокультуру від їх минулих значень, враховуючи при цьому вагові коефіцієнти, це дозволяє більш точно моделювати взаємозв'язок між

різними періодами часу. Можливість інтеграції ARIMA з узагальненими методами для врахування сезонності, наприклад SARIMA, робить цю модель ще більш гнучкою та придатною для аналізу різних циклічних коливань у цінах на агрокультуру, таких як сезонні зміни в урожаї чи споживацькому попиті. Ця модель надає зручний метод оцінки впливу факторів, таких як зовнішні події, економічні та кліматичні умови, на цінову динаміку. Здатність моделі реагувати на зміни в економічному середовищі робить її корисною для розуміння та передбачення поведінки ринку агрокультур.

Таблиця 2.1 - Кроки реалізації методу ARIMA

Крок	Дія
1	Проаналізувати ряд на визначення стаціонарності. Побудувати ACF та PACF графіки.
2	Якщо модель не стаціонарна, то виконати диференціювання або логарифмування
3	Визначити модель ARIMA
4	Оцінити модель
5	Проаналізувати результати
6	Виконати прогноз

2.2 Модель Хольта

Метод Хольта застосовується для прогнозування часових рядів, зокрема тих, де спостерігається тенденція до зростання або зменшення значень. Цей метод також використовується у випадках, коли дані не мають повного циклу і неможливо виділити сезонність (наприклад, у випадку неповного року або прогнозування за місяцями).

Зазвичай, якщо часовий ряд має виражену тенденцію до зростання або зменшення, схоже на експоненційне згладжування, розумно виділити тренд. Метод

Хольта включає два ключових коефіцієнти: коефіцієнт згладжування ряду і коефіцієнт тренду. Для розрахунку прогнозу застосовують кілька етапів:

– на першому етапі проводиться розрахунок експоненційно-згладженого ряду. Цей етап виявляється критичним для відокремлення можливих випадкових факторів від справжнього впливу на дані. У початковому періоді даних експоненційно-згладженого ряду значення збігається з першим значенням самого ряду. За формулою проводиться обчислення цього ряду:

$$L_t = k * Y_t + (1 - k) * (L_{t-1} - T_{t-1})$$

де, L_t – це згладжена величина за поточний період;

k – це коефіцієнт згладжування ряду;

Y_t – це поточні значення ряду;

L_{t-1} – це згладжена величина за попередній період;

T_{t-1} – це значення тренду за попередній період.

– на другому етапі визначається тренд та його значення, що розраховується в межах наступної формули. Коефіцієнт згладжування підбирається вручну та може приймати значення від 0 до 1, але для першого періоду воно завжди дорівнює 0.

$$T_t = b * (L_t - L_{t-1}) + (1 - b) * T_{t-1}$$

де, T_t – це значення тренду за поточний період;

b – це коефіцієнт згладжування тренду;

L_t – це експоненційно згладжена величина за поточний період;

L_{t-1} – це експоненційно згладжена величина за попередній період;

T_{t-1} – це значення тренду за попередній період.

– на наступному етапі з використанням розрахованого експоненційно-згладженого ряду та значень тренда, отриманих за наступною формулою, застосовуємо метод Хольта. Важливо відзначити, що чим менший коефіцієнт, тим сильніше враховується вплив тривалих тенденцій, у той час як значення, близьке до одиниці, акцентує останні зміни в ряді.

$$Y_{t+p} = L_t + p * T_t$$

Де, Y_{t+p} – це прогноз методом Хольта на період p ;

L_t – це експоненціально згладжена величина за останній період;

p – це порядковий номер періоду, на який робимо прогноз;

T_t – це тренд за останній період.

2.3 Метод Хольта-Вінтерса

Даний метод є модифікацією методу експоненційного згладжування та враховує експоненційний тренд та адитивну сезонність. Модель складається з рівняння прогнозу та трьох наступних рівнянь згладжування: для тренду (b_t), для адитивної сезонної компоненти та для рівня (l_t), з відповідними параметрами згладжування α , β та γ . Для врахування сезонної компоненти існують два типи рівнянь: адитивний та мультиплікативний.

Використовують адитивний метод, коли сезонні коливання залишаються приблизно сталими на протязі всього ряду, у той час як мультиплікативний метод використовують, коли сезонні коливання змінюються пропорційно рівню ряду. При адитивному методі сезонна компонента виражається в абсолютних значеннях у масштабі спостережуваного ряду, і в рівнянні рівня ряду вона коригується шляхом віднімання сезонної компоненти. Адитивний метод має наступну формулу:

$$y_{t+h|t} = l_t + hb_t + S_{t+h-m(k+1)}$$

$$l_t = \alpha(y_t - S_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1})$$

$$b_t = \beta(l_t + l_{t-1}) + (1 - \beta)b_{t-1}$$

$$S_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)S_{t-1}$$

де, m – це довжина періоду;

k – це ціла частина від $\frac{h-1}{m}$, що гарантує, що оцінка сезонних індексів буде отримана з останнього року з вибірки;

α, β, γ – це параметри згладжування.

Рівняння мультиплікативного методу має вигляд:

$$\hat{y}_{t+h|t} = (l_t + hb_t)S_{t+h-m(k+1)}$$

$$l_t = \alpha \frac{y_t}{(y_t - S_{t-m})} + (1 - \alpha)(l_{t-1} + b_{t-1})$$

$$b_t = \beta(l_t + l_{t-1}) + (1 - \beta)b_{t-1}$$

$$S_t = \gamma \frac{y_t}{l_{t-1} + b_{t-1}} + (1 - \gamma)S_{t-1}$$

Рівняння тренду тут таке саме як і в методі Хольта, тобто вказує на зважене значення тренду в момент часу t та оцінку тренду в момент часу $t - 1$. Рівняння рівня показує зважене значення між сезонно скоригованими результатами та прогнозу без урахування сезону на момент часу t . Рівняння сезону є зваженим значенням між поточним значенням сезонного індексу та значенням сезонного індексу рік назад.

У цій моделі всі значення залежать від попередніх, і тому на початковому етапі виникає питання, звідки брати початкові значення. Зазвичай початкові оцінки компонент отримуються шляхом усереднення всіх вхідних даних. Параметри згладжування вибираються так, щоб мінімізувати похибку прогнозу.

Переваги цього методу включають той факт, що для використання експоненційного згладжування необхідна невелика кількість даних, хоча визначення відповідних параметрів становить значну частину процесу створення прогнозу. Іноді цілком доцільно використовувати вбудовані статистичні пакети мови прогнозувань для виділення тренду, сезонності та похибок з часового ряду та подальшої роботи з ними.

Висновок до розділу 2

Розглянуті методи ARIMA, метод Хольта і метод Хольта-Вінтерса є популярними і ефективними інструментами для прогнозування часових рядів. Кожен з них має свої особливості та відмінності, що можуть бути використані в побудові моделей прогнозування аграрних культур в Україні.

До переваг методу ARIMA можна віднести наступне: здатність моделювати широкий спектр часових рядів, включаючи ті, де присутня складна структура або тренд. До недоліків можна віднести те, що моделювання вимагає стаціонарності

ряду, що може стати важкою задачею. Підбір оптимальних параметрів може вимагати багато часу та експертного досвіду.

До переваг методу Хольта можна віднести те, що метод ефективний для моделювання часових рядів з трендом та зручний у ситуаціях, коли тренд може бути лінійним. Із недоліків виділяється те, що метод не враховує сезонність та може недостатньо точно моделювати дані зі складною структурою.

Метод Хольта-Вінтерса відзначається тим, що він враховує сезонність і тренд та дозволяє працювати з даними, що мають виражені циклі та сезонні коливання. Недоліки у тому, що моделювання потребує великої кількості параметрів, що може ускладнити моделювання та підбір, а також те, що він може бути чутливим до початкових значень та потребує обширного набору даних для точних прогнозів.

3 МОДЕЛЮВАННЯ ТА ДОСЛІДЖЕННЯ ОТРИМАННИХ РЕЗУЛЬТАТІВ

3.1 Формування набору даних

Для аналізу та прогнозування цін на аграрну продукцію, було використано дані, які були отримані від інформаційно-аналітичного агентства "АПК-Інформ". Це агентство відоме своєю надійністю та авторитетністю у сфері аграрної статистики та аналітики в Україні. Їхні дані є важливим інструментом для бізнесу, фермерів, інвесторів та інших учасників аграрного ринку для прийняття обґрунтованих рішень. На основі отриманих даних були створені графіки цін на різні види агрокультур, що дозволило провести детальний аналіз компонентів часового ряду у ціновій динаміці.



Рисунок 3.1 – Динаміка цін на експортну сою, згідно ресурсу «АПК-Інформ»

Ресурс «АПК-Інформ» надає користувачам графіки цін на обрану агрокультуру. Для створення часового ряду, потрібно було отримати дані, що мають структуру «Дата-Значення». На жаль, ресурс «АПК-Інформ» не надає такої інформації у відкритому доступі, тому було прийнято рішення виконати парсинг даних. Парсинг (від англійського "parse") - це процес аналізу даних, зазвичай у структурованому форматі, для вилучення конкретної інформації або перетворення її у більш зручний для обробки вид. Розроблений наступний алгоритм:

– вибір аграрної культури. Перший крок полягає у виборі конкретної аграрної культури для аналізу. До обраних культур увійшли: ячмінь(експорт, USD),

Кафедра інтелектуальних інформаційних систем
«Інтелектуальна система прогнозування цін на аграрну продукцію в Україні на базі Android-застосунку»

кукурудза (експорт, UAH), соя (експорт, UAH), насіння соняшника (експорт, UAH), олія соняшникова(експорт, USD), продовольча пшениця II класу (експорт, UAH);

– для отримання даних про цінову динаміку використовується ресурс "АПК-Інформ". Здійснюється отримання HAR-файлу, який містить JSON документ із відомостями про дату та ціни на вибрану аграрну культуру;



Рисунок 3.2 – HAR-файл, який містить JSON файл з шуканими даними

– використовуючи ресурс «HAR File Web Scraper», проводиться аналіз та вилучення необхідного JSON файлу із HAR-файлу. Цей етап важливий для подальшої обробки та конвертації даних;

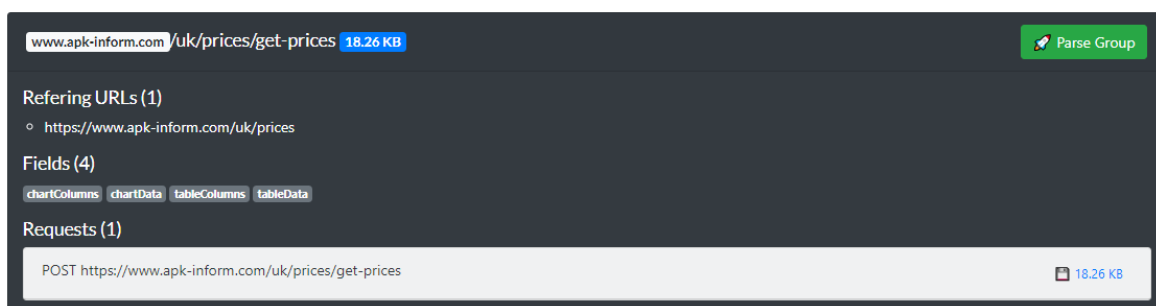


Рисунок 3.3 – Вилучення JSON-файлу з HAR-файлу за допомогою «HAR File Web Scraper»

– дані з отриманого JSON файлу конвертуються у XML-таблицю відповідно до заданого формату «Дата-Значення». Це спрощує подальший аналіз та дозволяє зручно представляти дані.

2	10.01.2014	4250
3	17.01.2014	4250
4	24.01.2014	4250
5	31.01.2014	4350
6	07.02.2014	4550
7	14.02.2014	4650
8	21.02.2014	4650
9	28.02.2014	4900
10	07.03.2014	5050

Рисунок 3.4 – Створена XML-таблиця на прикладі даних вилучених із JSON-файлу цін на сою.

Після застосування алгоритму для кожної із обраних культур, можна проаналізувати отримані дані. Аналіз наведено у таблиці 3.1

Таблиця 3.1 – Аналіз отриманих даних про ціни на агрокультури в Україні.

Назва культури	Початкова дата спостереження	Остання дата спостереження	Кількість спостережень
Ячмінь	02.01.2009	09.02.2024	701
Кукурудза	08.12.2017	09.02.2024	364
Соя	03.01.2014	16.02.2024	466
Насіння соняшника	03.01.2014	26.01.2024	525
Олія соняшникова	03.01.2014	16.02.2024	490
Пшениця	18.01.2019	16.02.2024	389

3.2 Пошук аномальних значень у часових рядах

Інтелектуальний аналіз даних (з англійської Data mining) використовується для виявлення нової значущої інформації з великого обсягу даних. З урахуванням постійного збільшення цих обсягів та наростання важливості отриманих результатів аналізу, проблема виявлення аномалій у даних стає особливо актуальною. Результати аналізу без попереднього виключення аномальних випадків можуть значно викривати зміст даних.

Виявлення аномалій полягає в пошуку неочікуваних значень або патернів у потоках даних. Аномалія, будь то викид, помилка, відхилення або виключення, представляє собою відхилення в поведінці системи від стандартного (очікуваного) режиму. Причини для аномальних значень в сфері агрокультур можуть бути різноманітні, включаючи технічні неполадки, несприятливі погодні умови, зовнішнє втручання та інші. На сьогодні існує значна кількість алгоритмів виявлення аномалій, спроектованих для різних типів даних.

Один зі статистичних методів, що дозволяє перевірити дані на наявність аномальних значень, - це метод "сигм" [13]. Аналіз викидів у даних за допомогою методу "сигм" дозволяє виявити аномальні значення в нестационарних рядах з розподілом, близьким до нормального. Основою цього методу аналізу є розрахунок середнього значення ряду та середньоквадратичного відхилення. Формула для розрахунку середнього значення ряду має вигляд:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (3.1)$$

Де \bar{x} – середнє значення ряду;

n – кількість елементів вибірки;

x_i – елемент вибірки i .

Формула для розрахунку середньоквадратичного відхилення має наступний вигляд:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad (3.2)$$

Де σ – це середньоквадратичне відхилення ряду.

Суть даного методу полягає в тому, що будь-які значення, що відрізняються від середнього більше, ніж на два середньоквадратичних відхилення, вважаються потенційними аномаліями. Поріг визначення аномалій (T) задається за допомогою формули:

$$T = x_i \pm 2 * \sigma, \quad (3.3)$$

Якщо аномалій виявилось дуже багато, то є сенс збільшити поріг та задати його дорівнюючим $3 * \sigma$, або $4 * \sigma$.

Застосовуючи правило сигм для перевірки на аномалії щотижневих значень часових рядів цін на аграрну продукцію, описаних у таблиці 3.1, отримаємо наступні значення.

Часовий ряд ціни попиту на ячмінь для експорту, складається з 701 спостережень за період з 02.01.2009 по 09.02.2024.

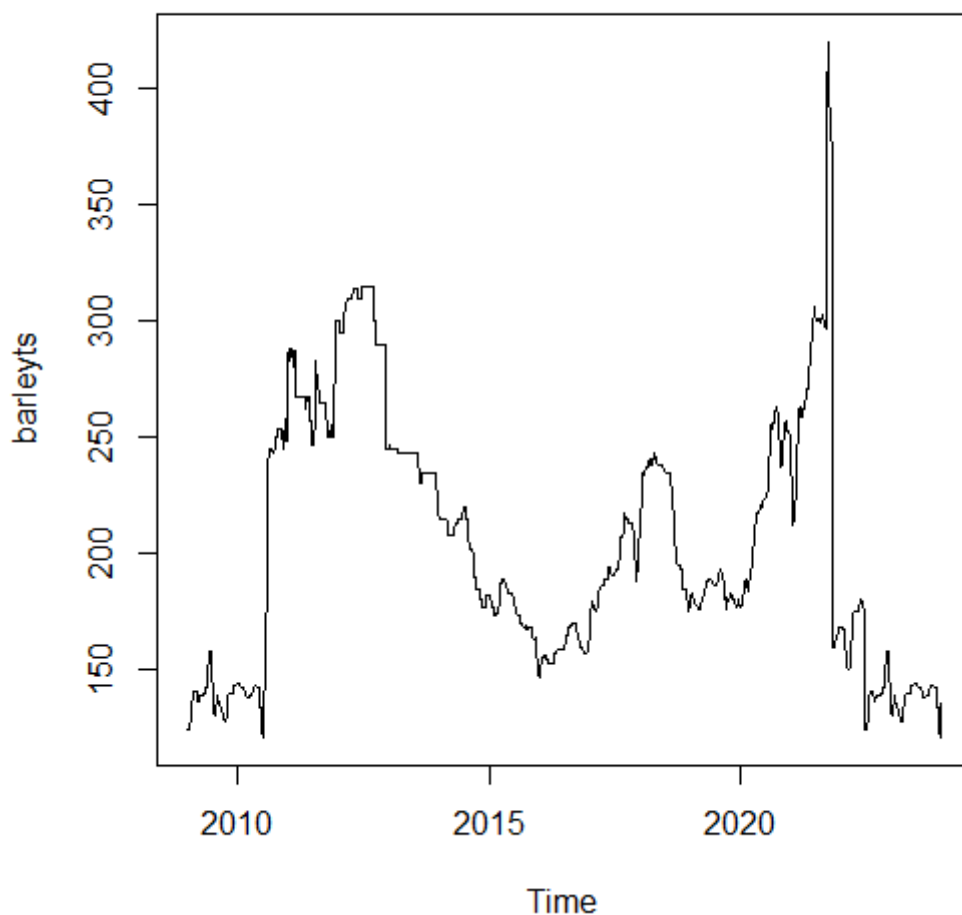


Рисунок 3.1 – Динаміка цін на ячмінь в Україні, експорт

За формулами (3.1) та (3.2) розрахуємо математичне сподівання та середньоквадратичне відхилення часового ряду.

$$\bar{x}_{\text{ячмінь}} = 203.43 \text{ } USD / T$$

$$\sigma_{\text{ячмінь}} = 54.59$$

За формулою (3.3) розрахуємо поріг визначення аномалії:

$$T_{\text{ячмінь}} = 203.43 \pm 3 * 54.59 = [39.66 ; 367.2] \quad (3.4)$$

З 701 спостережень часового ряду 6 спостережень виходять за межі порогу, ці відхилення будуть замінені на порогові значення з формули (3.4). Ці аномалії можна пояснити невизначеністю експортних цін через російське повномасштабне вторгнення в Україну 24 лютого 2022 року [2].

```
> plot(cornnts$days, type="l", col="blue", lty="n", las=1)
[1] "2022-03-18 UTC" "2022-03-25 UTC" "2022-04-01 UTC" "2022-04-08 UTC" "2022-04-15 UTC" "2022-04-22 UTC"
> |
```

Рисунок 3.2 – Визначені аномалії часового ряду цін на ячмінь.

Часовий ряд ціни попиту на кукурудзу для експорту, складається з 364 спостережень за період з 08.12.2017 по 09.02.2024

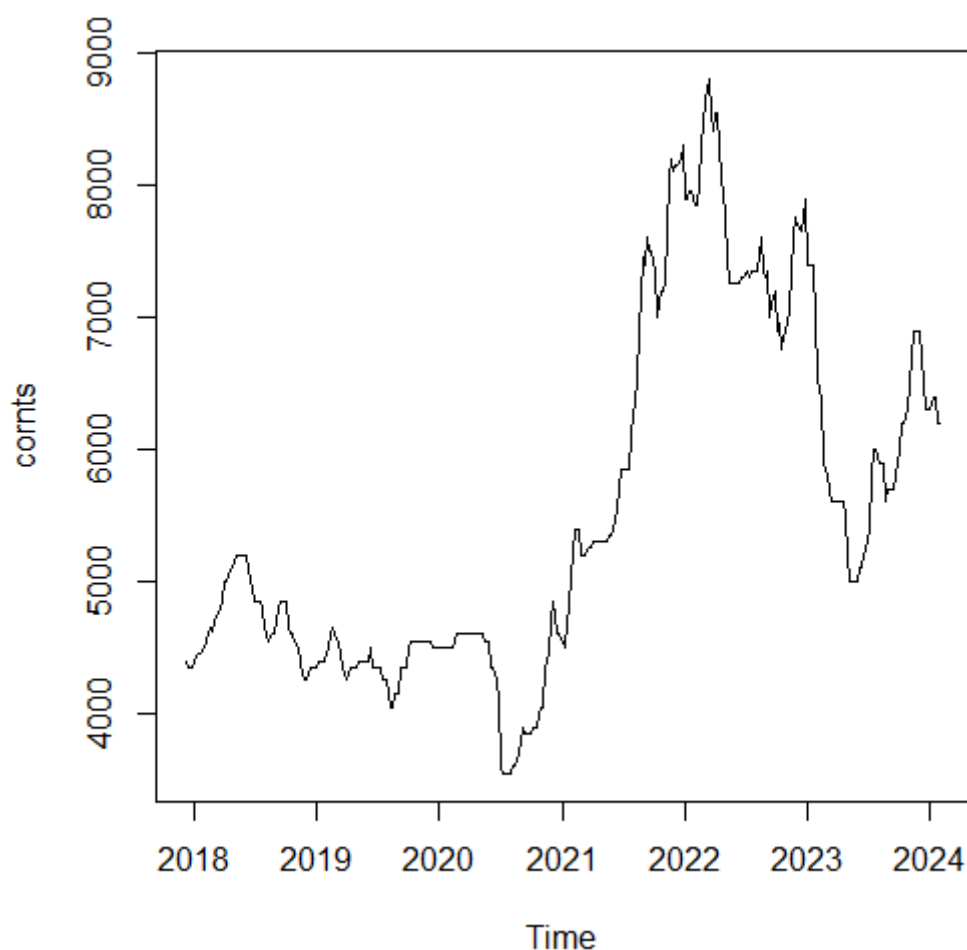


Рисунок 3.3 - Динаміка цін на кукурудзу в Україні, експорт

За формулами (3.1) та (3.2) розрахуємо математичне сподівання та середньоквадратичне відхилення часового ряду.

$$\bar{x}_{\text{кукурудза}} = 5532.78 \text{ ₴UAH/т}$$

$$\sigma_{\text{кукурудза}} = 1335.72$$

За формулою (3.3) використовуючи правило $3 * \sigma$ розрахуємо поріг визначення аномалії:

$$T_{\text{кукурудза}} = 5532.78 \pm 3 * 1335.72 = [1525.62 ; 9539.94]$$

Всі 364 значень часового ряду не виходять за межі порогу, тому аномалії для часового ряду цін на кукурудзу України не виявлено.

Часовий ряд ціни попиту на сою для експорту, складається з 466 спостережень за період з 03.01.2014 по 16.02.2024

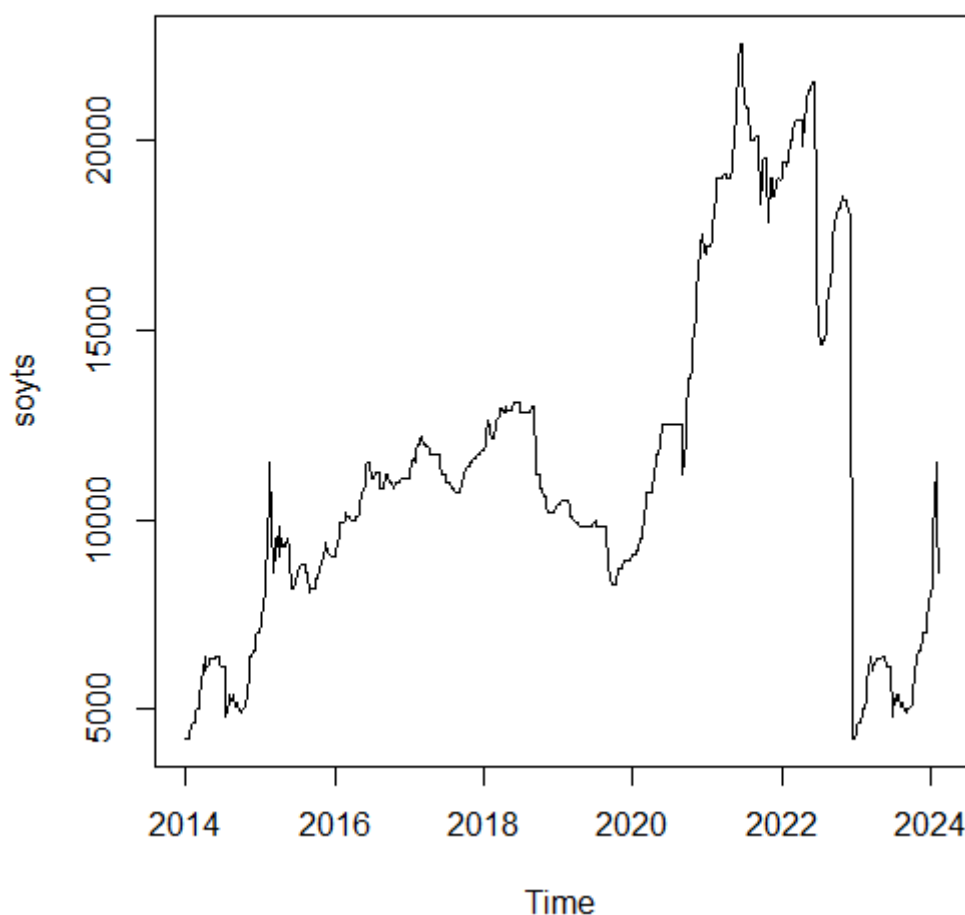


Рисунок 3.4 - Динаміка цін на сою в Україні, експорт

За формулами (3.1) та (3.2) розрахуємо математичне сподівання та середньоквадратичне відхилення часового ряду.

$$\bar{x}_{\text{соя}} = 11335.39 \text{ ₴UAH/т}$$

$$\sigma_{\text{соя}} = 4539.06$$

За формулою (3.3) використовуючи правило $2 * \sigma$ розрахуємо поріг визначення аномалії:

$$T_{\text{соя}} = 11335.39 \pm 2 * 4539.06 = [2257.27 ; 20413.51]$$

За межі верхнього порогу виходять 26 значень, що є більше 5% від загального обсягу спостережень. Використаємо правило $3 * \sigma$, щоб підняти поріг та зменшити частку аномалій від кількості спостережень.

$$T_{\text{соя}} = 11335.39 \pm 3 * 4539.06 = [2281.62 ; 24952.57]$$

Всі 466 значень часового ряду не виходять за межі порогу, тому аномалії для часового ряду цін на кукурудзу України не виявлено.

Часовий ряд ціни попиту на насіння соняшника для експорту, складається з 525 спостережень за період з 03.01.2014 по 26.01.2024

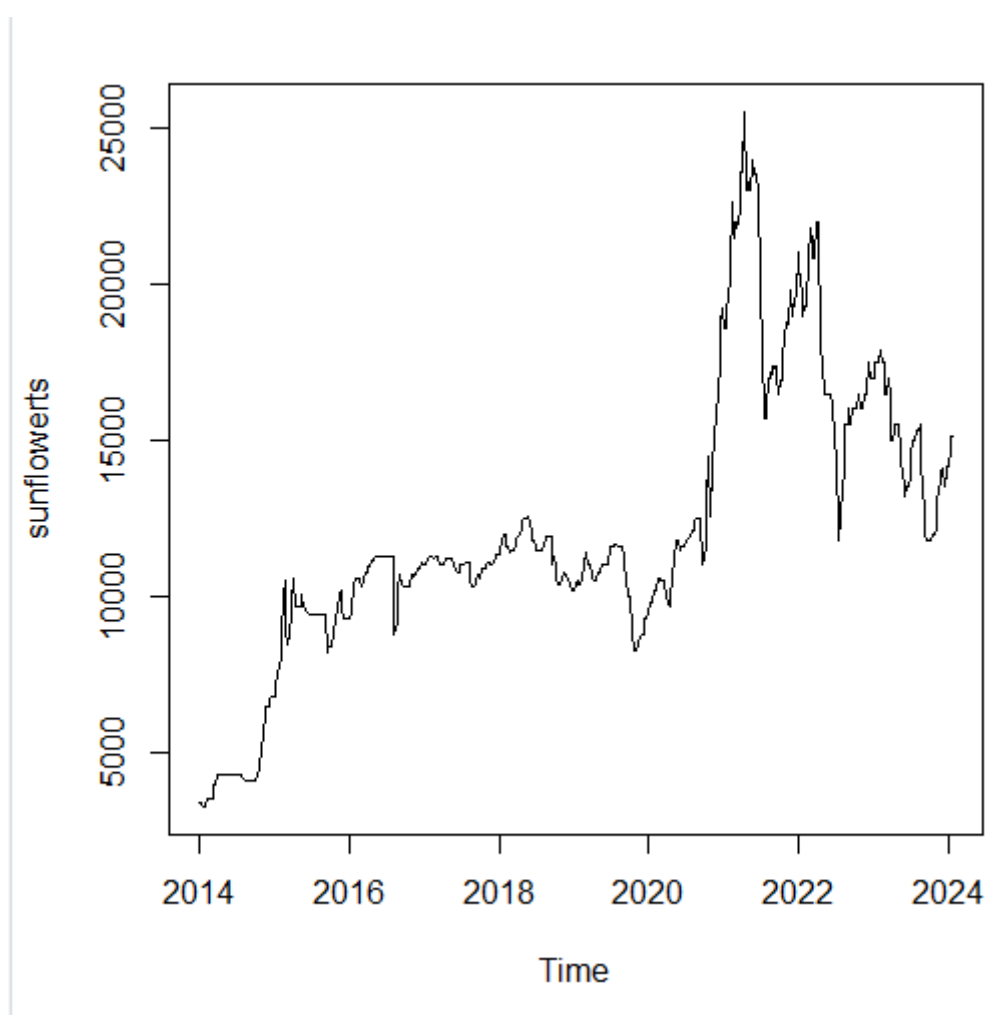


Рисунок 3.5 - Динаміка цін на насіння соняшника в Україні, експорт

За формулами (3.1) та (3.2) розрахуємо математичне сподівання та середньоквадратичне відхилення часового ряду.

$$\bar{x}_{\text{насіння_соняшника}} = 12138.45 \text{ €UAH/т}$$

$$\sigma_{\text{насіння_соняшника}} = 4397$$

За формулою (3.3) використовуючи правило $3 * \sigma$ розрахуємо поріг визначення аномалії:

$$T_{\text{насіння_соняшника}} = 12138.45 \pm 2 * 4397 = [3344.45 ; 20932.45]$$

За межі верхнього порогу виходять 28 значень, а за межі нижнього 4 значення, що є більше 5% від загального обсягу спостережень. Використаємо правило $3 * \sigma$, щоб підняти поріг та зменшити частку аномалій від кількості спостережень.

$$T_{\text{насіння_соняшника}} = 12138.45 \pm 3 * 4397 = [1052.55 ; 25329.45]$$

За межі верхнього порогу виходить 1 значення, що є прийнятною величиною.

Часовий ряд ціни попиту на олію соняшника для експорту, складається з 490 спостережень за період з 03.01.2014 по 16.02.2024.

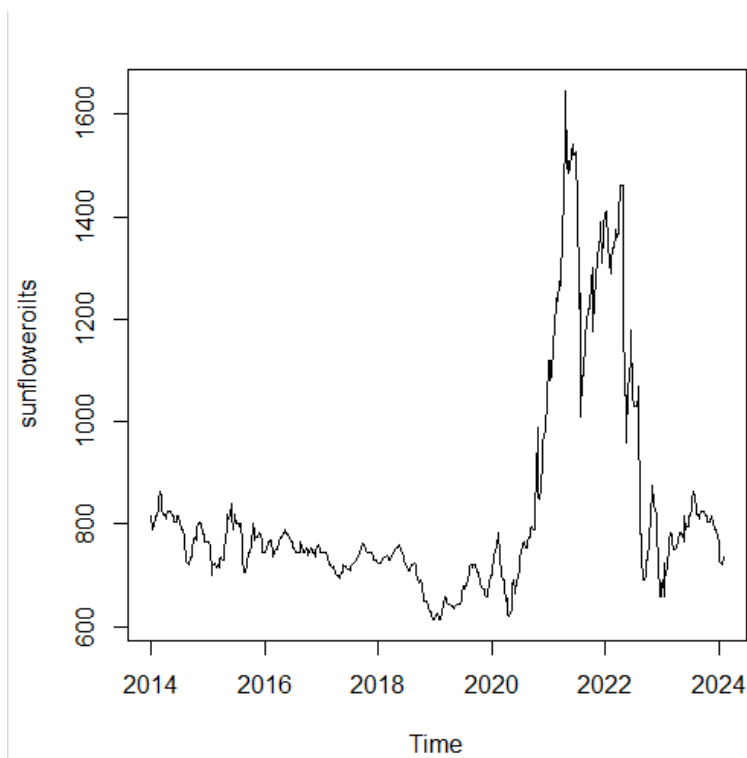


Рисунок 3.6 – Динаміка цін на соняшникову олію в Україні, експорт

За формулами (3.1) та (3.2) розрахуємо математичне сподівання та середньоквадратичне відхилення часового ряду.

$$\bar{x}_{\text{олія_соняшника}} = 831.22 \text{ USD/т}$$

$$\sigma_{\text{олія_соняшника}} = 212.18$$

За формулою (3.3) розрахуємо поріг визначення аномалії:

$$T_{\text{олія_соняшника}} = 831.22 \pm 2 * 212.18 = [406.85 ; 1255.6]$$

З 490 спостережень часового ряду 46 спостережень виходять за межі порогу, що є 9.3% від загальної кількості спостережень. Використаємо правило $3 * \sigma$, щоб підняти поріг та зменшити частку аномалій від кількості спостережень.

$$T_{\text{олія_соняшника}} = 831.22 \pm 3 * 212.18 = [194.66 ; 1467.8] \quad (3.5)$$

З 490 спостережень часового ряду 12 спостережень виходять за межі порогу, ці відхилення будуть замінені на порогові значення з формули (3.5). Ці аномалії можна пояснити несприятливими погодними умовами 2020 року, через що продукт аномально подорожчав в усьому світі [14].

[1] "2021-03-05 UTC" "2021-03-12 UTC" "2021-03-19 UTC" "2021-03-26 UTC" "2021-04-02 UTC" "2021-04-09 UTC" "2021-04-16 UTC" "2021-04-23 UTC" "2021-04-30 UTC"
[10] "2021-05-07 UTC" "2021-05-14 UTC" "2021-05-21 UTC"

Рисунок 3.7 – Аномальні значення часового ряду динаміки цін на олію соняшника

Часовий ряд ціни попиту на пшеницю 2 класу для експорту, складається з 389 спостережень за період з 18.01.2019 по 16.02.2024.

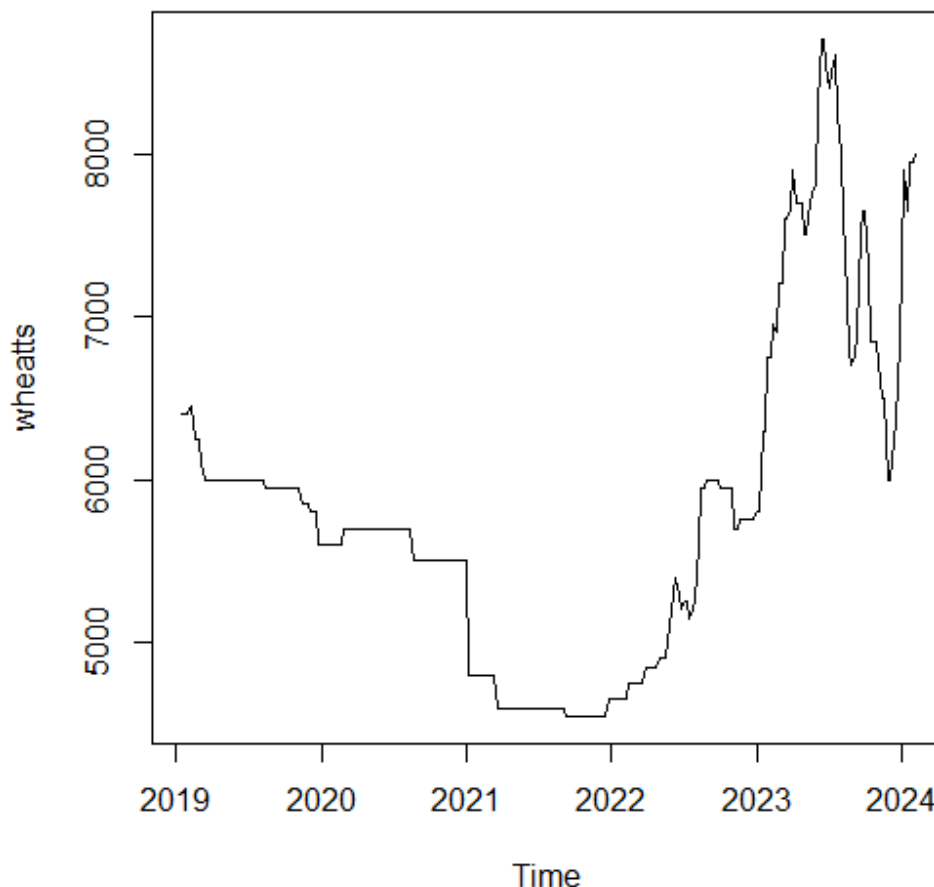


Рисунок 3.8 – Динаміка цін на пшеницю 2 класу в Україні, експорт

За формулами (3.1) та (3.2) розрахуємо математичне сподівання та середньоквадратичне відхилення часового ряду.

$$\bar{x}_{\text{пшениця}_2} = 5811.55 \text{ €UAH/т}$$

$$\sigma_{\text{пшениця}_2} = 1011.95$$

За формулою (3.3) розрахуємо поріг визначення аномалії:

$$T_{\text{пшениця}_2} = 5811.55 \pm 2 * 1011.95 = [3787.63 ; 7835.46]$$

З 389 спостережень часового ряду 41 спостережень виходять за межі порогу, що є 10.6% від загальної кількості спостережень. Використаємо правило $3 * \sigma$, щоб підняти поріг та зменшити частку аномалій від кількості спостережень.

$$T_{\text{пшениця}_2} = 5811.55 \pm 3 * 1011.95 = [2775.68 ; 8847.42] \quad (3.7)$$

З 389 спостережень часового ряду 3 спостереження виходять за межі порогу, ці відхилення будуть замінені на порогові значення з формули (3.7). Ці аномалії можна пояснити нестабільною динамікою ринку.

[1] "2021-11-26 UTC" "2022-01-21 UTC" "2022-01-28 UTC"

Рисунок 3.9 - Аномальні значення часового ряду динаміки цін на пшеницю
2 класу

Таблиця 3.2 – Результати аналізу часового ряду на аномальні значення

Назва культури	Кількість спостережень	Аномальні значення верхнього порогу	Аномальні значення нижнього порогу
Ячмінь, експорт	701	6	0
Кукурудза, експорт	364	0	0
Соя, експорт	466	0	0
Насіння соняшника, експорт	525	1	0
Соняшникова олія, експорт	490	12	0
Пшениця 2 класу, експорт	389	3	0

3.3 Побудова моделі прогнозування ARIMA

Для аналізу поведінки часових рядів з вираженими коливаннями та для створення математичних моделей, які описують цю поведінку, широко використовують лінійну стохастичну модель авторегресії та ковзного середнього ARIMA, або авторегресії та проінтегрованого ковзного середнього ARIMA. Ця модель пов'язує поточне значення змінної, що досліджується, зі значеннями цієї ж змінної у попередні моменти часу, а також з поточним та попередніми значеннями залишків моделі.

На основі щотижневих даних з 02.01.2009р. по 09.02.2024р. побудуємо прогноз ціни на ячмінь в Україні, експорт [9]. Часовий ряд складається з 701

значень, і жодне з них не було виключено з розгляду під час перевірки на аномалії. Розділимо часовий ряд на навчальну та тестову вибірки, де до навчальної вибірки увійдуть перші 681 значень часового ряду (~97%), а до тестової - останні 20 (~3%). Обираємо значення моделі $p[0; 3], d[0; 1], q[0; 3]$. Використавши формулу значення інформаційного критерію Акаїке(AIC) серед альтернативних моделей обираємо кращу модель, у якої значення буде мінімальним.

$$AIC = \ln\left(\frac{SS_e}{n}\right) + \frac{2}{n} * (r - 1) \quad (3.8)$$

Де, n – число спостережень;

r – число параметрів моделі.

```
ARIMA(0,1,0) : 5102.874
ARIMA(1,1,0) : 5076.112
ARIMA(0,1,1) : 5078.289
ARIMA(0,1,0) : 5100.867
ARIMA(2,1,0) : 5077.373
ARIMA(1,1,1) : 5076.312
ARIMA(2,1,1) : 5079.28
ARIMA(1,1,0) : 5074.101
ARIMA(2,1,0) : 5075.357
ARIMA(1,1,1) : 5074.296
ARIMA(0,1,1) : 5076.278
ARIMA(2,1,1) : 5077.258
```

Рисунок 3.10 – Визначення оптимальних параметрів моделі ARIMA для прогнозування цін на ячмінь, за AIC

Згідно AIC, оптимальними моделям виявились ARIMA(1,1,0) і ARIMA(1,1,1). Серед цих моделей обираємо ту, яка дасть кращий результат оцінки точності прогнозу на тестовому наборі даних. Для оцінки використовується середня абсолютна похибка у відсотках (MAPE).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - \hat{Y}_t|}{Y_i} \quad (3.9)$$

Де, Y_i – це фактичне значення ціни за період, що розглядається,

\hat{Y}_t - це значення прогнозованої моделі за період, що розглядається,

n - кількість періодів.

За результатами розрахунків на тестовій вибірці (20 періодів) отримуємо, що значення MAPE для моделі ARIMA(1,1,0) дорівнює 1.83%, а для моделі ARIMA(1,1,1) – 1,84%. Обираємо модель з найменшим значенням MAPE – модель ARIMA(1,1,0). Коефіцієнти цієї моделі: $p = 0.1221$

Графік залишків коефіцієнтів автокореляції показує хаотичне розміщення, деякі залишки виходять за межі довірчого коридору, але можна припустити, що вони незначні, так як це аномальні значення, про які пояснювалось раніше.

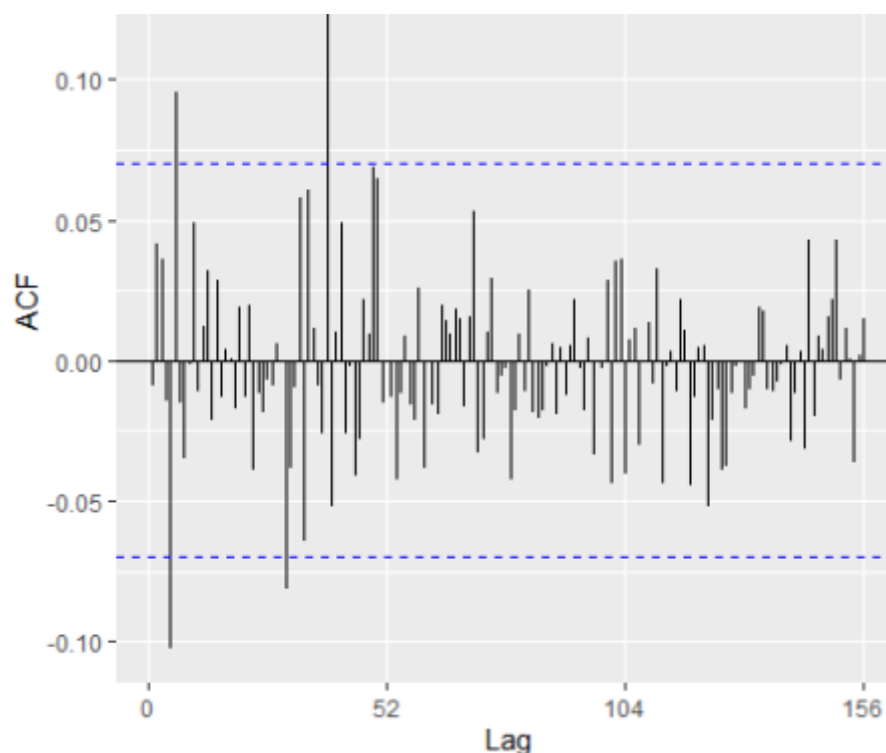


Рисунок 3.11 – Графік ACF залишків моделі для ціни на ячмінь.

Прогнозовані значення для 5 майбутніх періодів розраховано у RStudio за допомогою функцій бібліотеки forecast та наведено на рисунку 3.12

```
Forecast  Lo 80  Hi 80  Lo 95  Hi 95
137.7144 129.6974 145.7314 125.4535 149.9753
138.2407 125.7554 150.7260 119.1461 157.3354
138.3428 122.4256 154.2599 113.9996 162.6859
138.3626 119.6013 157.1238 109.6697 167.0554
138.3664 117.1335 159.5993 105.8934 170.8393
```

Рисунок 3.12 – Параметри прогнозованих значень ціни на ячмінь в Україні, експорт за моделлю ARIMA(1,1,0).

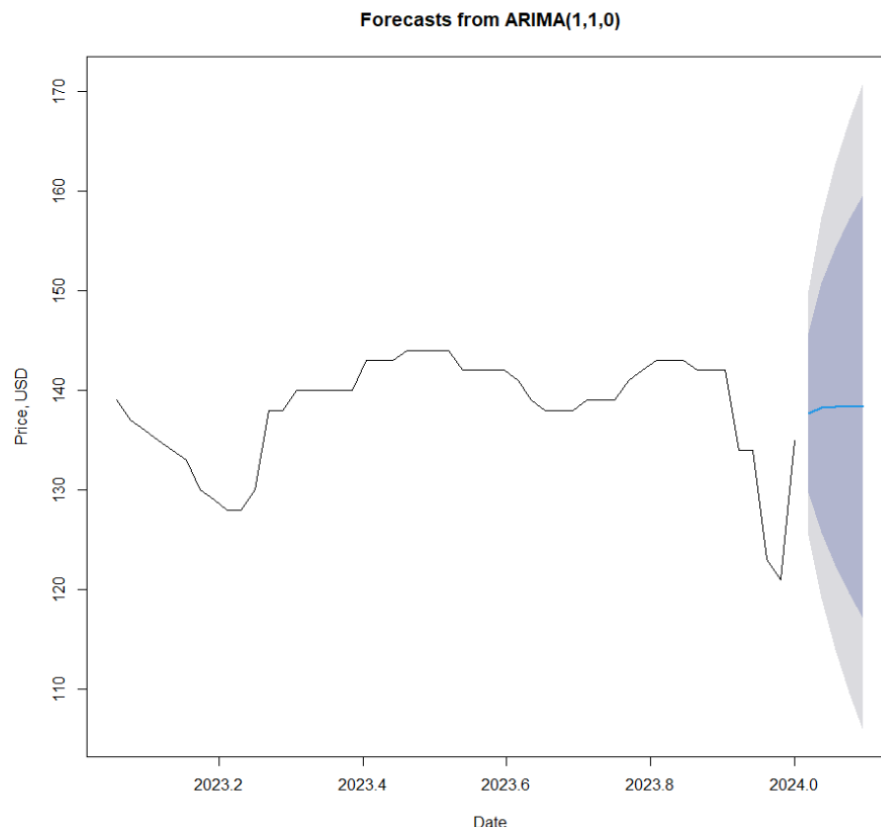


Рисунок 3.13 – Графік прогнозованих значень ціни на ячмінь

Сірим показана нижня межа довірчого інтервалу, 80%, темно-синім виділена верхня межа довірчого інтервалу, 95%.

Аналіз залишків на білий шум за допомогою тесту Льюнга-Бокса показує, що $p - value = 0.8 > 0.05$, враховуючи що за нульову гіпотезу береться те, що залишки це білий шум, то нульова гіпотеза приймається.

```
data: arimafk_barley$residuals
x-squared = 0.062232, df = 1, p-value = 0.803
```

Рисунок 3.14 – Тест Льюнга-Бокса, де нульова гіпотеза це «Залишки після прогнозу є білим шумом».

На основі щотижневих даних з 08.12.2017 по 09.02.2024. побудуємо прогноз ціни на кукурудзу в Україні, експорт [9]. Часовий ряд складається з 364 значень, і жодне з них не було виключено з розгляду під час перевірки на аномалії. Розділимо часовий ряд на навчальну та тестову вибірки, де до навчальної вибірки увійдуть перші 353 значень часового ряду (~97%), а до тестової - останні 11 (~3%).

Обираємо значення моделі $p[0; 4], d[0; 1], q[0; 4]$. Використавши формулу (3.8) серед альтернативних моделей обираємо кращу модель, у якій значення буде мінімальним.

```
ARIMA(0,1,4) : 4013.03
Best model: ARIMA(0,1,4)
```

Рисунок 3.15 – Параметри моделі прогнозування та АІС цін на кукурудзу в Україні, експорт, ARIMA

Згідно АІС, оптимальна модель ARIMA(0,1,4) з параметрами $q_1 = 0.2632, q_2 = 0.1253, q_3 = 0.2936, q_4 = 0.1623$, Для оцінки використовується формула (3.9), MAPE = 1.5%. Прогнозуємо ціну на наступні 5 періодів, та отримаємо результати на рисунку 3.15 та рисунку 3.16.

Forecast	Lo 80	Hi 80	Lo 95	Hi 95
6179.767	6014.275	6345.259	5926.669	6432.865
6104.313	5837.693	6370.934	5696.553	6512.074
6073.318	5721.347	6425.288	5535.025	6611.610
6077.349	5628.605	6526.093	5391.054	6763.644
6077.349	5534.640	6620.058	5247.347	6907.351

Рисунок 3.15 – Результати прогнозу цін на кукурудзу в Україні, експорт, на 5 періодів.

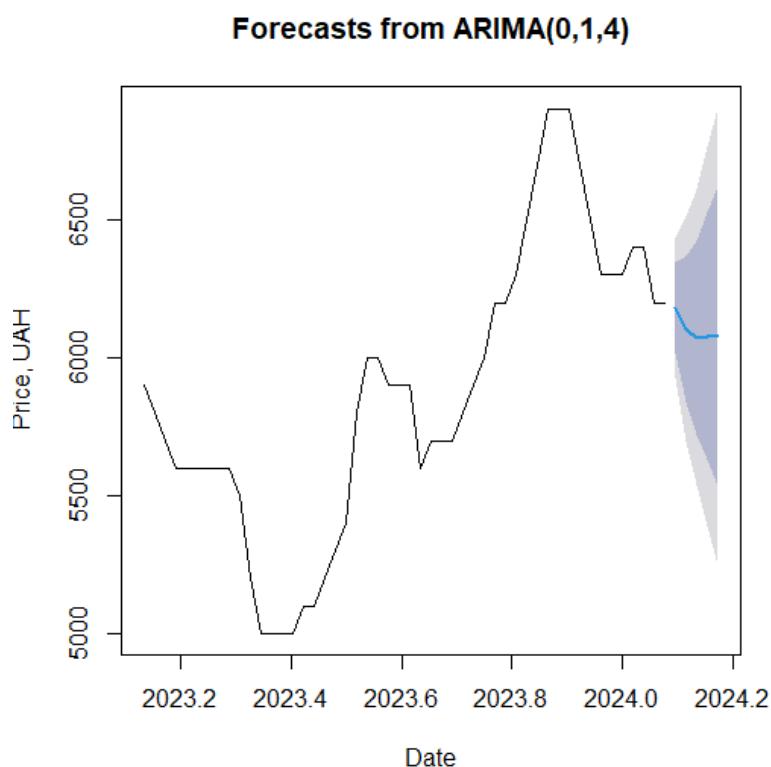


Рисунок 3.16 – Графік прогнозованих значень цін на кукурудзу в Україні, експорт, моделі ARIMA(0,1,4)

Графік залишків коефіцієнтів автокореляції показує хаотичне розміщення, деякі залишки виходять за межі довірчого коридору, але можна припустити, що вони незначні, так як це аномальні значення, про які пояснювалось раніше. Аналіз залишків на білий шум за допомогою тесту Льюнга-Бокса показує, що $p - value = 0.46 > 0.05$, враховуючи, що за нульову гіпотезу береться те, що залишки це білий шум, то нульова гіпотеза приймається.

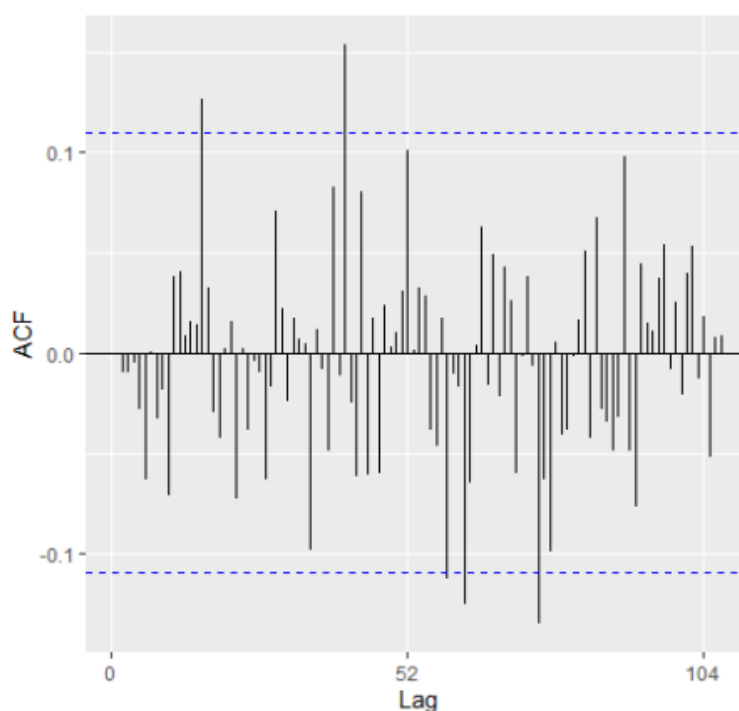


Рисунок 3.17 - Графік ACF залишків моделі для ціни на кукурудзу

На основі щотижневих даних з 08.12.2017 по 09.02.2024. побудуємо прогноз ціни на сою в Україні, експорт [9]. Часовий ряд складається з 466 значень, і жодне з них не було виключено з розгляду під час перевірки на аномалії. Розділимо часовий ряд на навчальну та тестову вибірки, де до навчальної вибірки увійдуть перші 461 значень часового ряду (~99%), а до тестової - останні 5 (~1%).

Обираємо значення моделі $p[0; 4], d[0; 1], q[0; 4]$. Використавши формулу (3.8) серед альтернативних моделей обираємо кращу модель, у якій значення буде мінімальним.

ARIMA(0,1,1) : 8881.869
Best model: ARIMA(0,1,1)

Рисунок 3.18 – Параметри моделі прогнозування та AIC цін на сою в Україні, експорт, ARIMA

Згідно AIC, оптимальна модель ARIMA(0,1,1) з параметрами $q_1 = 0.0443$, Для оцінки використовується формула (3.9), MAPE = 3.2%. Прогнозуємо ціну на наступні 5 періодів, та отримуємо результати на рисунку 3.19.

Forecast	Lo 80	Hi 80	Lo 95	Hi 95
8780.589	7345.347	10215.83	6585.576	10975.60
8780.589	6864.819	10696.36	5850.671	11710.51
8780.589	6482.670	11078.51	5266.225	12294.95
8780.589	6155.577	11405.60	4765.979	12795.20
8780.589	5864.951	11696.23	4321.505	13239.67

Рисунок 3.19 – Результати прогнозу цін на сою в Україні, експорт, на 5 періодів.

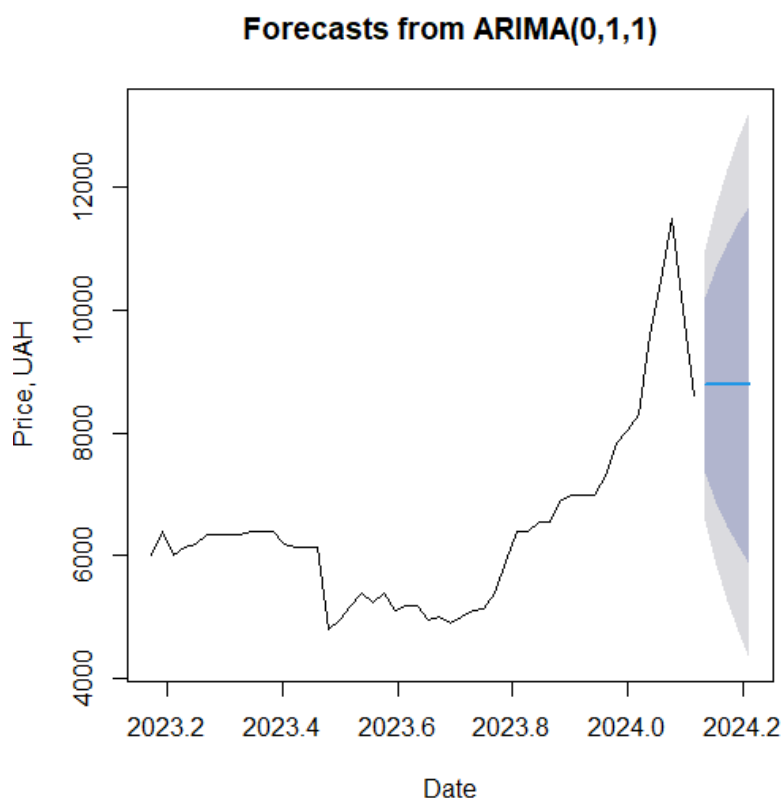


Рисунок 3.20 – Графік прогнозованих значень моделі ARIMA(0,1,1)

Графік залишків коефіцієнтів автокореляції показує хаотичне розміщення, деякі залишки виходять за межі довірчого коридору, але можна припустити, що вони незначні, так як це аномальні значення, про які пояснювалось раніше. Аналіз 2024 р.

залишків на білий шум за допомогою тесту Льюнга-Бокса показує, що $p - value = 0.826 > 0.05$, враховуючи, що за нульову гіпотезу береться те, що залишки це білий шум, то нульова гіпотеза приймається.

На основі щотижневих даних з 03.01.2014 по 26.01.2024. побудуємо прогноз ціни на насіння соняшника в Україні, експорт [9]. Часовий ряд складається з 525 значень, і жодне з них не було виключено з розгляду під час перевірки на аномалії. Розділимо часовий ряд на навчальну та тестову вибірки, де до навчальної вибірки увійдуть перші 519 значень часового ряду (~99%), а до тестової - останні 5 (~1%).

Обираємо значення моделі $p[0; 4]$, $d[0; 1]$, $q[0; 4]$. Використавши формулу (3.8) серед альтернативних моделей обираємо кращу модель, у якій значення буде мінімальним.

Згідно оцінки AIC, оптимальні моделі: ARIMA(2,1,0) та ARIMA(2,1,1). Для оцінки використовується формула (3.9) і обираємо модель з найнижчим значенням, для ARIMA(2,1,0) MAPE = 2.12%, для ARIMA(2,1,1) MAPE = 2.13%. Прогнозуємо ціну на наступні 5 періодів, використовуючи модель ARIMA(2,1,0) з параметрами та отримаємо результати на рисунку 3.21 та рисунку 3.22.

Forecast	Lo 80	Hi 80	Lo 95	Hi 95
15133.14	14465.96	15800.31	14112.77	16153.50
15137.10	14135.51	16138.70	13605.29	16668.91
15140.32	13851.01	16429.63	13168.49	17112.15
15141.04	13609.70	16672.37	12799.06	17483.01
15141.39	13397.99	16884.78	12475.09	17807.68

Рисунок 3.21 – Результати прогнозу цін на насіння соняшника в Україні, експорт, на 5 періодів.

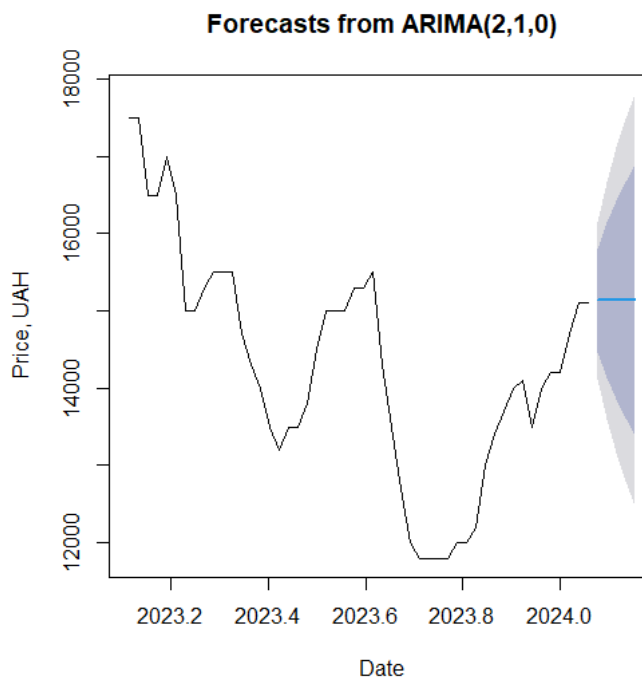


Рисунок 3.22 – Графік прогнозованих значень цін на насіння соняшника в Україні, експорт, моделі ARIMA(2,1,1)

На основі щотижневих даних з 03.01.2014 по 16.02.2024. побудуємо прогноз ціни на олію соняшникову в Україні, експорт [9]. Часовий ряд складається з 490 значень, і жодне з них не було виключено з розгляду під час перевірки на аномалії. Розділимо часовий ряд на навчальну та тестову вибірки, де до навчальної вибірки увійдуть перші 484 значень часового ряду (~99%), а до тестової - останні 5 (~1%).

Обираємо значення моделі $p[0; 4]$, $d[0; 1]$, $q[0; 4]$. Використавши формулу (3.8) серед альтернативних моделей обираємо кращу модель, у якій значення буде мінімальним.

Згідно оцінки AIC, оптимальна модель: ARIMA(2,1,2) з параметрами $p_1 = 0.3454$, $p_2 = 0.5547$, $q_1 = 0.3$, $q_2 = 0.7148$ Для оцінки використовується формула (3.9), для ARIMA(2,1,2) MAPE = 1.73%. Прогнозуємо ціну на наступні 5 періодів, використовуючи модель ARIMA(2,1,2) та отримаємо результати на рисунку 3.23 та рисунку 3.24.

Forecast	Lo 80	Hi 80	Lo 95	Hi 95
736.2042	693.6233	778.7852	671.0822	801.3263
736.4089	674.8274	797.9904	642.2282	830.5897
735.8116	655.2348	816.3885	612.5799	859.0433
735.4918	638.7768	832.2068	587.5790	883.4046
735.7127	626.8977	844.5276	569.2946	902.1307

Рисунок 3.23 – Результати прогнозу цін на олії соняшника в Україні, експорт, на 5 періодів.

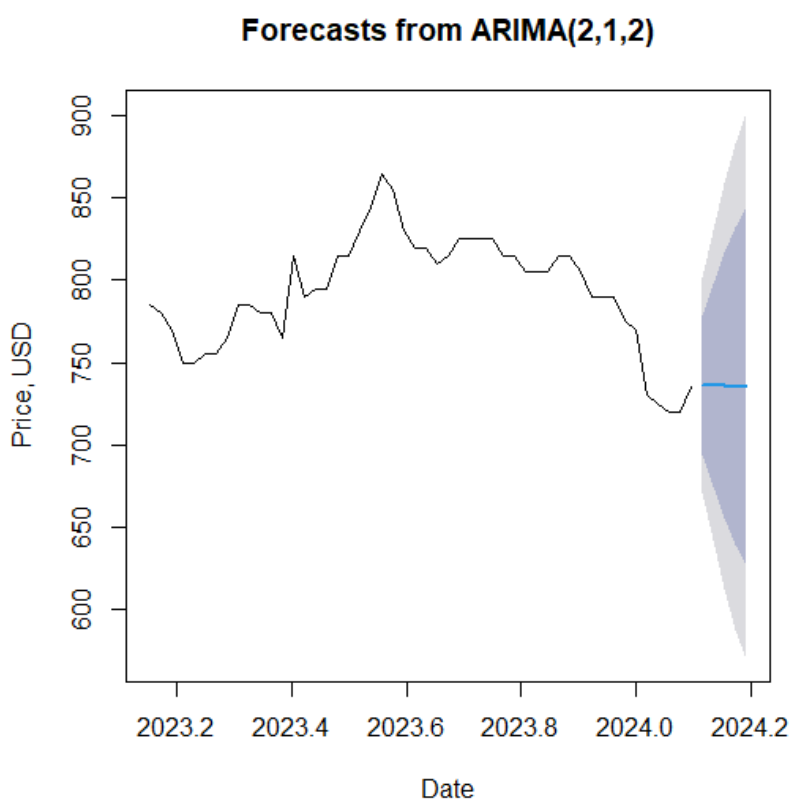


Рисунок 3.24 – Графік прогнозованих значень цін на олію соняшника в Україні, експорт, моделі ARIMA(2,1,2)

На основі щотижневих даних з 18.01.2019 по 16.02.2024. побудуємо прогноз ціни на пшеницю 2 класу в Україні, експорт [9]. Часовий ряд складається з 389 значень, і жодне з них не було виключено з розгляду під час перевірки на аномалії. Розділимо часовий ряд на навчальну та тестову вибірки, де до навчальної вибірки увійдуть перші 383 значень часового ряду (~98%), а до тестової - останні 5 (~2%).

Обираємо значення моделі $p[0; 4], d[0; 1], q[0; 4]$. Використавши формулу (3.8) серед альтернативних моделей обираємо кращу модель, у якій значення буде мінімальним.

Згідно оцінки AIC, оптимальна модель: ARIMA(2,1,0) з параметрами $p_1 = 0.2298$, $p_2 = 0.1766$. Для оцінки використовується формула (3.9), для ARIMA(2,1,0) MAPE = 1.9973%. Прогнозуємо ціну на наступні 5 періодів, використовуючи модель ARIMA(2,1,0) та отримуємо результати на рисунку 3.25 та рисунку 3.26.

Forecast	Lo 80	Hi 80	Lo 95	Hi 95
8011.490	7837.218	8185.762	7744.964	8278.016
8022.963	7746.731	8299.194	7600.503	8445.422
8027.629	7652.161	8403.096	7453.401	8601.857
8030.728	7567.927	8493.528	7322.936	8738.520
8032.264	7490.642	8573.885	7203.925	8860.603

Рисунок 3.23 – Результати прогнозу цін на пшеницю 2 класу в Україні, експорт, на 5 періодів.



Рисунок 3.22 – Графік прогнозованих значень цін на пшеницю 2 класу в Україні, експорт, моделі ARIMA(2,1,0)

3.4 Побудова моделі прогнозування Хольта-Вінтерса

Перед тим, як будувати модель прогнозування Хольта-Вінтерса, необхідно створити часовий ряд із отриманих даних. Перш ніж робити щось щодо прогнозування, нам потрібно повідомити R, що ці дані є часовим рядом. Для цього ми створюємо об'єкт часового ряду. Є два критичних параметра вхідних даних, які потрібно надати функції — частота та початок.

По-перше, ми повинні визнати, що функція, яка створює об'єкт часового ряду, не є розумною. Незважаючи на те, що набір даних має дуже читабельні значення, схожі на дату, потрібно було вручну повідомити йому частоту рядків набору даних, тобто скільки рядків даних є за рік. У поточній роботі використовується набір щотижневих даних, а отже частота для усіх наборів даних у часовому ряді буде дорівнювати 52.

Після створення часового ряду, можна провести декомпозицію основних його елементів, серед яких: тренд, сезонність та випадковий шум. На прикладі даних із таблиці 3.2 розкладемо на компоненти часовий ряд цін на соняшникову олію.

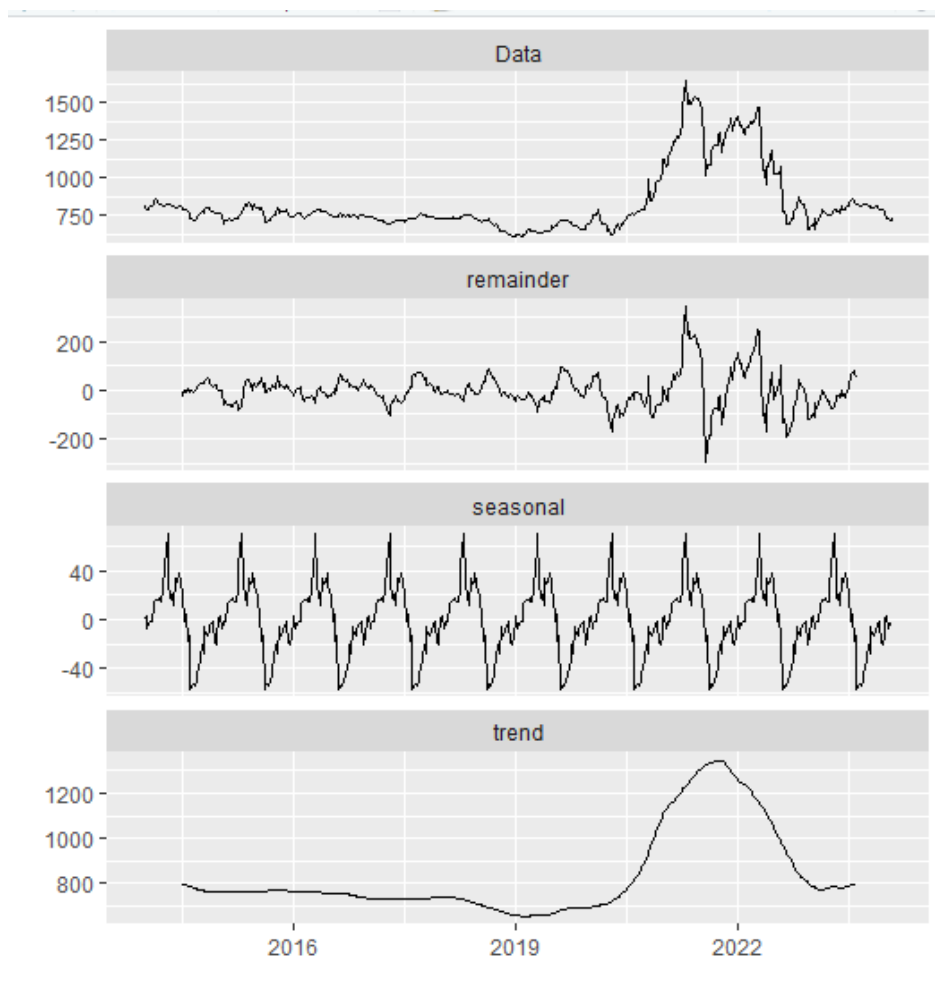


Рисунок 3.23 – Компоненти часового ряду цін на соняшкову олію

Наступним кроком необхідно визначити вагу α, β, γ . Через автоматизовану бібліотеку отримаємо $\alpha = 0.6217975$, $\beta = 0.028571$, $\gamma = 0.582106$.

Можна побудувати прогноз, виходячи з отриманих даних.

fit	upr	lwr
739.6745	813.9642	665.3849
739.5533	841.5142	637.5923
732.1140	855.6985	608.5296
731.7100	873.6614	589.7586
730.0984	888.2985	571.8982

Рисунок 3.24 – Прогноз моделі Хольта-Вінтерса ціни на соняшкову олію на наступні 5 періодів.

Висновок до розділу 3

В третьому розділі дипломної роботи було проведено детальний аналіз методів здобування наборів даних, а також їхнього подальшого аналізу на предмет

виявлення аномальних значень. Застосовані методи дозволили ефективно підготувати дані для подальшого використання в моделях прогнозування.

Перший етап включав у себе вибір та збір репрезентативних наборів даних, що відіграє важливу роль у точності та релевантності прогнозних моделей. Використані методи забезпечили зручний та ефективний спосіб здобування необхідних даних, що стало фундаментом для подальшого аналізу та прогнозування.

Далі, проведено аналіз наборів даних на предмет аномальних значень. Цей етап є ключовим у виключенні спотворень та аномалій, які можуть негативно впливати на якість прогнозу. Застосовані методи аналізу аномалій сприяли виявленню та виправленню неправильних даних, що забезпечило стабільність та достовірність подальших досліджень.

Далі, в розділі було розглянуто побудову моделей прогнозування, таких як ARIMA та Хольта-Вінтерса. Використання цих моделей дозволило ефективно моделювати та прогнозувати часові ряди з урахуванням їхньої структури та динаміки. Отримані результати свідчать про адекватність та точність обраних моделей в контексті досліджуваного явища.

4 ПРОГРАМНА РЕАЛІЗАЦІЯ

4.1 Опис основних технологій

Існує декілька мов програмування, інструменти яких, використовують розробники для побудови Android-застосунків. Kotlin — одна з найпопулярніших, яка використовується для створення мобільних додатків операційної системи Android. Вона була розроблена з орієнтацією на гнучкість, дозволяючи розробникам писати код, який буде працювати на будь-якій машині, незалежно від архітектури чи платформи, завдяки використанню JVM — віртуальної машини Java.

Для побудови користувацького інтерфейсу було обрано XML (eXtensible Markup Language), що є мовою розмітки, яка використовується для представлення та обміну даними між програмами. У розробці мобільних додатків, зокрема для Android, XML широко використовується для опису користувацького інтерфейсу.

Android XML Layouts використовують у поєднанні з XML для визначення структури та вигляду елементів інтерфейсу користувача в додатках для платформи Android. Основна мета використання XML у цьому контексті - розділення даних та логіки додатка від його представлення. Такий підхід дозволяє легко маніпулювати інтерфейсом, змінювати його вигляд та структуру без модифікації джерела даних або логіки програми. Android XML Layouts надає необхідні інструменти та парадигми для роботи із проектуванням користувацького інтерфейсу. Розмітка XML надає декларативний підхід до визначення, як повинен виглядати інтерфейс. Взаємодія між XML та Android XML Layouts працює таким чином, що розробники описують структуру та властивості елементів, а система Android відповідає за їх реалізацію. Перевагою XML для розробки інтерфейсу користувача є тим, що розмітка має текстовий формат, що легко читається та редагується. Це полегшує спільну роботу розробників та дизайнерів. Вона також надає інструменти для подальшого інтегрування дизайнерських рішень у застосунок, так як завдяки XML,

дизайнер може визначити вигляд інтерфейсу, а розробник може працювати над логікою без втручання в розмітку.

Для побудови моделей прогнозування необхідно було спеціалізоване середовище, R - це мова програмування та середовище для статистичного аналізу та візуалізації даних, яка визначається своєю високою потужністю та гнучкістю. Її популярність серед дослідників та аналітиків визначається кількома ключовими особливостями. R пропонує широкий спектр пакетів і бібліотек, спрямованих на роботу з часовими рядами. У побудові моделей прогнозування для поточної роботи були використані такі пакети, як `forecast`, `xts`, `tseries`, що надають інструменти для різних методів прогнозування та аналізу часових рядів та будуть у подальшому розглянуті.

Щоб виконати дослідницьку роботу для побудови та аналізу моделей прогнозування, середовище RStudio надає розширені інструменти візуалізації, наприклад, `ggplot2`. Це дозволяє створювати високоякісні графіки для аналізу та візуалізації часових рядів, полегшуючи сприйняття паттернів та трендів. Завдяки розширеним статистичним функціям, R дозволяє аналізувати та визначати характеристики часових рядів, такі як сезонність, тренди та аномалії. Мова дозволяє легко маніпулювати часовими маркерами, полегшуючи процес моделювання та прогнозування.

Використовуючи середовище RStudio, можна бути впевненим у активній спільноті користувачів і розробників, R сприяє обміну знаннями та розв'язанню проблем, а також постійному вдосконаленню інструментів прогнозування часових рядів. Це допоможе підтримувати застосунок та впроваджувати нові інструменти для роботи із наявними часовими рядами.

4.2 Огляд допоміжних бібліотек для побудови моделей прогнозування

Бібліотека “`tseries`” в мові програмування R призначена для роботи з аналізом та моделюванням часових рядів. Ця бібліотека містить набір функцій та

інструментів, що допомагають досліджувати та аналізувати характеристики часових рядів, такі як тренди, сезонність та інші особливості.

Основні можливості бібліотеки “tseries” включають усі необхідні інструменти для створення моделі прогнозування:

- бібліотека надає функції для проведення різних статистичних тестів, таких як тести на стаціонарність, що допомагають визначити, чи є часовий ряд стаціонарним;
- вона має інструменти для визначення автокореляційних коефіцієнтів, які вказують на наявність кореляцій між значеннями часового ряду в різних моментах часу;
- бібліотека дозволяє виявляти та аналізувати сезонні компоненти в часових рядах, що допомагає в розумінні повторюваних паттернів;
- бібліотека дозволяє використовувати моделі ARIMA (autoregressive integrated moving average) для прогнозування часових рядів;
- вона надає інструменти для вивчення спектральних характеристик часових рядів.

Для побудови моделей ARIMA була використана бібліотека “forecast”. У мові програмування R вона виступає як потужний інструмент для прогнозування часових рядів. Вона дозволяє аналізувати та моделювати динаміку даних в часі, надаючи різноманітні функції для різних методів прогнозування.

Зокрема, бібліотека дозволяє використовувати експоненційне згладжування для оцінки трендів та сезонності, а також застосовувати ARIMA-моделі для більш складних часових рядів, що і було досягнуто у прогнозуванні цін на аграрну продукцію. Можливість автоматичного вибору оптимальних параметрів моделі спрощує процес роботи із складними даними, а також допомагає швидше обробляти дані для будування багатьох прогнозів.

Одна з ключових переваг forecast - це здатність прогнозувати інтервали невизначеності, що дозволяє оцінити ступінь впевненості у прогнозі. Це корисно

при роботі з реальними даними, де невизначеність може впливати на точність прогнозу.

Бібліотека “forecast” також надає інструменти для автоматичної візуалізації результатів прогнозування, допомагаючи виявляти параметри для моделі ARIMA у ході визначення оптимальних параметрів моделі для різних агрокультур. Така можливість важлива для зрозуміння та інтерпретації прогнозів на стадії розробки моделі.

Бібліотека “stats” в мові програмування R є важливою частиною стандартного набору інструментів та містить різноманітні статистичні функції для аналізу та моделювання даних. Її роль в побудові моделі Хольта-Вінтерса визначається можливістю застосування методів експоненційного згладжування для моделювання тренду та сезонності в часових рядах.

Метод Хольта-Вінтерса був використаний для побудови моделей прогнозування для таких продуктів агрокультур, що визначаються вираженим трендом і сезонністю. Бібліотека надає необхідні інструменти для виконання цього аналізу, включаючи функції для використання методів експоненційного згладжування, що входять до складу моделі Хольта-Вінтерса. Ці функції дозволили визначити оптимальні параметри згладжування, такі як коефіцієнти згладжування для тренду, сезонності та рівня. Це було важливо для адаптації моделі до особливостей конкретного часового ряду та досягнення найкращої точності у прогнозуванні. Бібліотека також враховує можливість автоматичного вибору оптимальних параметрів, що полегшує процес роботи з різними типами даних та рядами. Використання цієї бібліотеки для побудови моделі Хольта-Вінтерса стало ефективним інструментом у процесі роботи з прогнозування.

4.3 Опис допоміжних бібліотек для побудови користувацького інтерфейсу Android-застосунку

Для побудови Android-застосунку необхідно було знайти спосіб поєднати інструменти мови R та Kotlin. Було знайдено рішення у створенні серверу, що

виконує роль буфера між обома середовищами для передачі даних. Для отримання даних середовища R, Kotlin використовує інструменти бібліотеки OkHttp.

Бібліотека OkHttp є потужним інструментом для роботи з HTTP-з'єднаннями в мові програмування Kotlin, і вона володіє численними особливостями, що роблять її важливим компонентом для розробки поточного мобільного застосунку. Однією з ключових характеристик OkHttp є її гнучкість у використанні для синхронних та асинхронних HTTP-запитів, що дозволяє ефективно взаємодіяти з мережею у різних сценаріях. Здатність встановлювати та управляти пулом з'єднань робить його ефективним інструментом для оптимізації ресурсів мережі. OkHttp також володіє вбудованою підтримкою кешування, що дозволяє автоматично зберігати та використовувати локальні копії відповідей, покращуючи швидкодію та зменшуючи трафік даних. Можливість легко додавати заголовки та параметри до запитів робить OkHttp дуже зручним для конфігурації HTTP-взаємодії. Крім того, бібліотека дозволяє ефективно передавати файли на сервер та завантажувати їх, що наразі не імплементовано у додаток. Окрім цього, OkHttp включає ряд функціональностей для безпечної роботи з HTTPS, обробки сертифікатів та перехоплення та модифікації запитів та відповідей, що робить його гнучким та надійним інструментом для виконання мережевих операцій.

Є необхідність обробляти отримані дані прогнозування для їх графічного зображення за потребою користувача. Для графічного зображення даних була використана бібліотека Kandy. Це бібліотека для мови програмування Kotlin, яка надає гнучкий та ідіоматичний DSL для створення різних типів графіків, використовуючи різні бібліотеки візуалізації. Метою використаної бібліотеки було спростити та прискорити процес створення графіків з великою кількістю параметрів і налаштувань, без необхідності вивчення довгих документацій.

Для зберігання інформації, що була переглянута користувачем, стояло завдання вирішити питання кешування. Для цього була використана бібліотека Room. Бібліотека Room є важливим інструментом для роботи з базами даних в Android-додатках, розроблена компанією Google. Її основна мета - спростити

роботу з базами даних SQLite та забезпечити зручний інтерфейс для розробників Android.

Однією з ключових можливостей Room є можливість визначати об'єкти-сутності з використанням анотацій, що дозволяє легко відображати структуру бази даних. Крім того, вона забезпечує створення DAO (Data Access Object) для виконання основних операцій з базою даних, таких як додавання, читання, оновлення та видалення даних. Room вбудовує підтримку LiveData та RxJava, що робить його ідеальним інструментом для роботи з асинхронними операціями та відслідковування змін в реальному часі. Це дозволяє розробникам легко реалізовувати реактивний підхід до роботи з даними. Одна з важливих переваг Room - це його вбудована підтримка міграцій, що дозволяє автоматично оновлювати базу даних при зміні схеми, зберігаючи при цьому існуючі дані. Компіляція SQL-запитів на етапі розробки дозволяє виявляти помилки та покращує ефективність взаємодії з базою даних. Загалом, Room є потужним інструментом для спрощення роботи з базами даних в Android-додатках, надаючи зручний та ефективний спосіб роботи з локальними даними.

4.4 Налаштування локального серверу для передачі інформації між R та Kotlin

Для обробки запитів користувача та зберігання результатів роботи моделі прогнозування, необхідно було використовувати виділений сервер. Циркуляція даних у такій системі виглядає наступним чином:

- на локальному комп'ютері ініціюється виконання batch-файлу для виконання скриптів, в якому реалізована модель обробки даних. під час виконання скрипт генерує обчислені дані;
- обчислені дані автоматично передаються на виділений сервер через мережеві протоколи забезпечуючи безпечний та ефективний обмін інформацією;

- користувач інтерактивно взаємодіє з інтерфейсом користувача, вводячи необхідний запит або взаємодіючи з відповідними елементами інтерфейсу користувача;
- запит користувача передається на сервер через відповідний протокол зв'язку, такий як http;
- сервер отримує запит від користувача і визначає, які дані необхідно передати далі. запит подається до локального комп'ютера, на якому раніше виконувався скрипт і обчислювалися дані;
- локальний комп'ютер обробляє отриманий запит, виконує необхідні дії з використанням раніше обчислених даних та повертає результати на сервер;
- сервер отримує результати від локального комп'ютера та надсилає їх користувачеві у вигляді відповіді на його початковий запит. користувач отримує результати свого запиту через інтерфейс користувача.

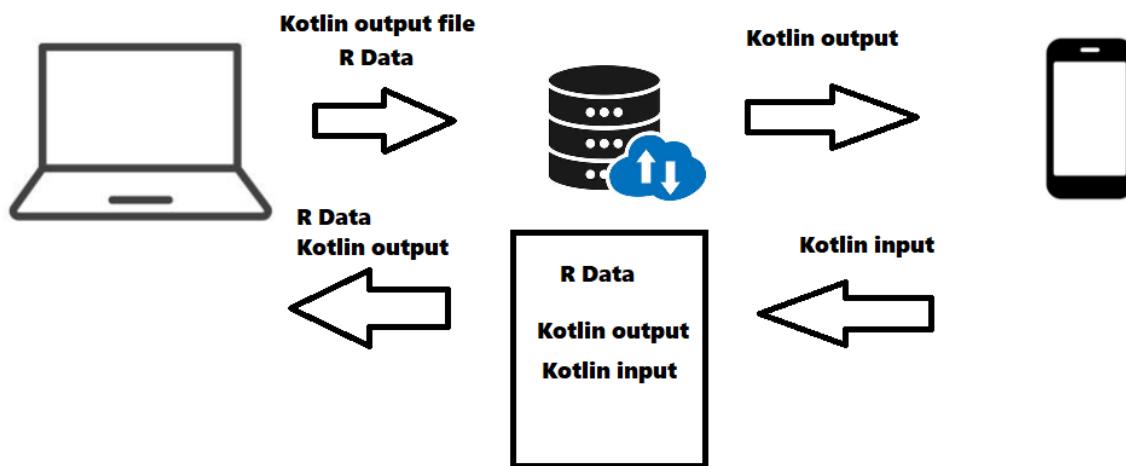


Рисунок 4.1 – схема роботи системи

Щоб налаштувати сервер, треба було переконатись, що на локальному комп'ютері ввімкнена система передачі даних та є можливість використовувати протокол FTP.

Для перевірки роботи серверу, створюється каталог проекту, та додається текстовий документ для перевірки передачі даних на сервер.

[\[To Parent Directory\]](#)

2/11/2024	5:10 PM	22372	Barley.xlsx
2/16/2024	4:42 PM	6186	BarleyPredicted.xlsx
2/11/2024	11:19 PM	22095	BarleyTest.xlsx
2/12/2024	6:14 AM	16188	Corn.xlsx
2/16/2024	4:42 PM	6186	CornPredicted.xlsx
12/25/2023	2:54 PM	12	test.txt
2/16/2024	4:42 PM	6186	Wheat.xlsx
2/16/2024	4:42 PM	6186	WheatPredicted.xlsx

Рисунок 4.2 – Тест працездатності файлового серверу.

4.5 Розробка інтерфейсу користувача Android-застосунку

Розробка користувацького інтерфейсу Android - це творчий та важливий процес, який спрямований на створення ефективної та приємної взаємодії користувача з додатком. При проектуванні інтерфейсу для Android слід було враховувати особливості платформи, забезпечуючи високий рівень зручності та інтуїтивності.

Дизайн інтерфейсу включає в себе розміщення елементів, колірну палітру та графічний дизайн, щоб забезпечити однорідний та привабливий вигляд додатку. Важливо було використовувати стандартизовані компоненти Android, такі як кнопки, поля введення та меню, для забезпечення консистентності інтерфейсу.

Ергономіка важлива для створення дружнього інтерфейсу. Розміщення елементів повинно було бути логічним та інтуїтивним, забезпечуючи легкий доступ до основних функцій додатку. Важливо також було уникати перенасичення інтерфейсу зайвою інформацією, щоб не перенавантажувати користувача.

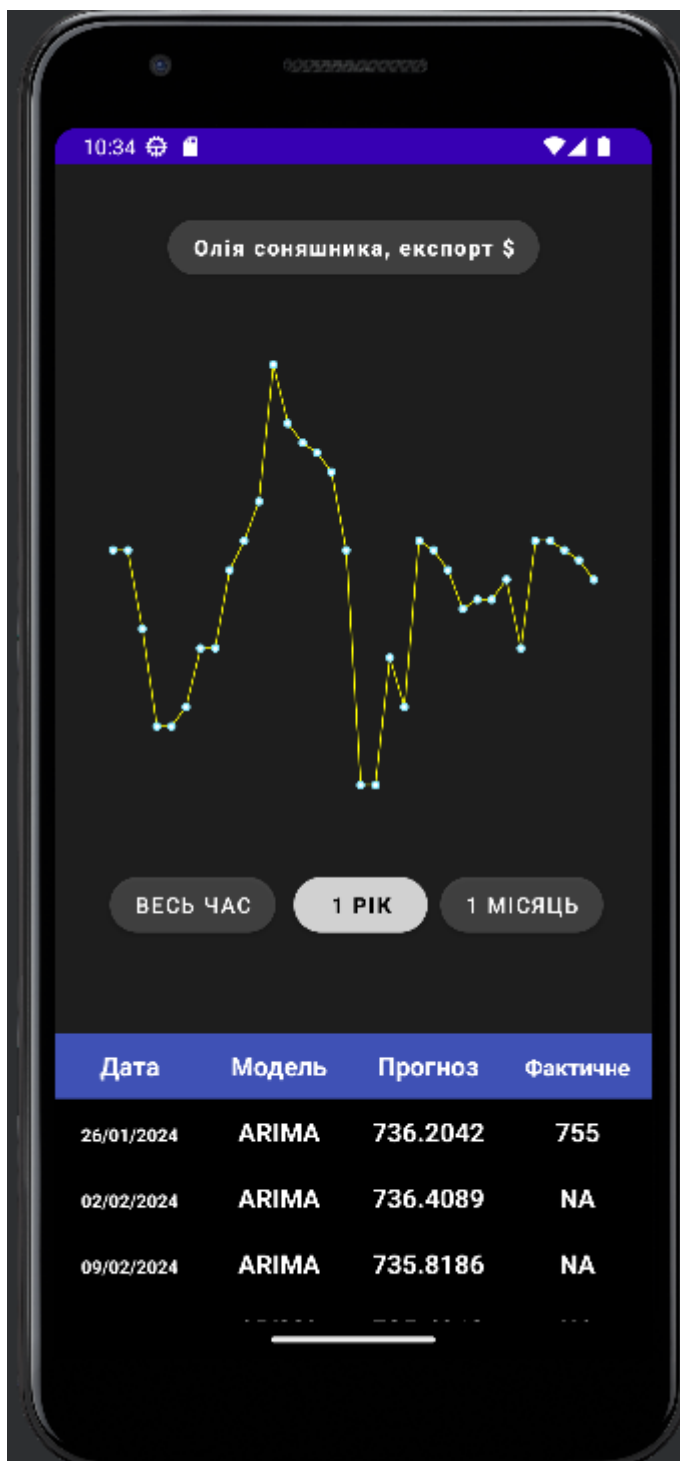


Рисунок 4.3 – Інтерфейс користувача Android-застосунку

На рисунку 4.3 зображений розроблений інтерфейс користувача. Верхнім елементом програми є кнопка, що відображає поточну досліджувану агрокультуру. Вона відображає назву агрокультури та знак \$ або ₴, що показує у яких грошових одиницях відображаються показані значення.

Далі на екрані зображений графік поточної агрокультури, що має три рівня налаштування зображення даних: за весь час, за 1 рік, за 1 місяць. Графік виконаний у мінімалістичному стилі, та має інтерактивний елемент у вигляді точок на графіку, натиснувши на які, буде відображений момент часу та ціна у цей момент. Елемент відображення графіку здатний до збільшення зображення жестами користувача.

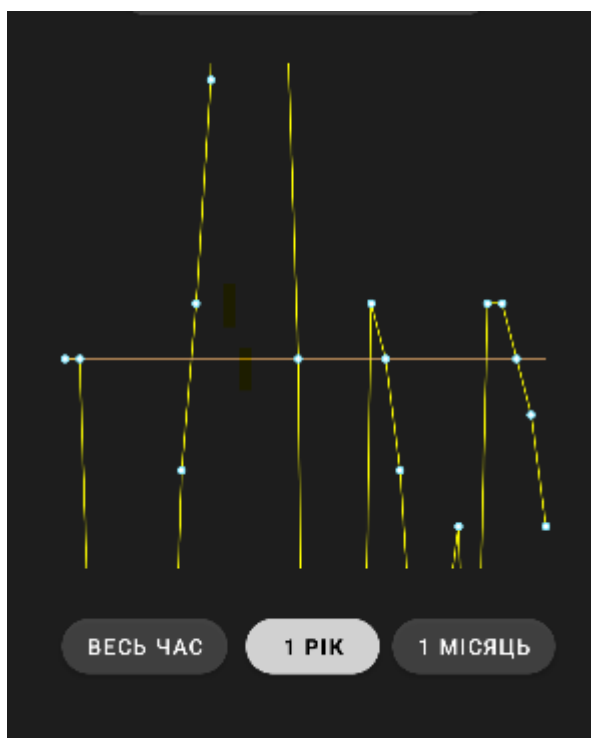


Рисунок 4.4 – Властивість елемента графіка до збільшення зображення.

Наступним елементом інтерфейсу є таблиця із прогнозами. Вона складається з 4 стовпців та 10 рядків. Перші 5 рядків відображають прогнози моделі ARIMA, наступні 5 рядків відображають прогнози моделі Хольта-Вінтерса. У першому стовпці розташована дата прогнозу у форматі ДД/ММ/РР. Другий стовпець містить інформацію про модель прогнозування. Третій стовпець містить опосередковане значення прогнозу між інтервалами впевненості High 95 та Low 95. Останній стовпець містить значення ціни агрокультури, якщо воно є у базі даних, та NA, якщо даних щодо ціни ще немає.

Дата	Модель	Прогноз	Фактичне
26/01/2024	ARIMA	736.2042	755
02/02/2024	ARIMA	736.4089	NA
09/02/2024	ARIMA	735.8186	NA
16/02/2024	ARIMA	735.4918	NA
23/02/2024	ARIMA	735.7127	NA
26/01/2024	HW	739.6742	755
02/02/2024	HW	739.5533	NA
09/02/2024	HW	732.1140	NA
16/02/2024	HW	731.7100	NA
23/02/2024	HW	730.0984	NA

Рисунок 4.5 – Відображення таблиці з даними прогнозів

Висновок до розділу 4

У четвертому розділі кваліфікаційної роботи було детально розглянуто використані бібліотеки при розробці моделей прогнозування та користувацького інтерфейсу для Android-застосунку. Цей етап виявився ключовим у забезпеченні функціональності та ефективності даного дослідження, а також у створенні зручного інтерфейсу для користувачів.

При побудові моделей прогнозування, було використано відомі та ефективні бібліотеки, які дозволили реалізувати алгоритми ARIMA та Хольта-Вінтерса з високою точністю та швидкістю. Використання таких бібліотек виявилось не лише економією часу, а й гарантією високого рівня надійності та стабільності прогнозів цін на аграрну продукцію в Україні.

У процесі розробки користувацького інтерфейсу для Android-застосунку, було взято до уваги важливість інтуїтивності та привабливого дизайну.

Використання відповідних компонентів та правильне їх розташування допомогли створити зручний і легко засвоюваний інтерфейс для користувачів.

Додатково, розгорнутий опис взаємодії елементів програмної реалізації, налаштування ftp-серверу та компонентів користувацького інтерфейсу, дозволив детально освітити аспекти реалізації дослідження.

В цілому, результати четвертого розділу вказують на вдалу реалізацію проекту, де ефективно використані бібліотеки, правильний дизайн інтерфейсу, а також детальна програмна реалізація сприяли успішному виконанню поставлених завдань та досягненню цілей кваліфікаційної роботи магістра.

ВИСНОВКИ

В ході роботи над кваліфікаційною роботою магістра на тему "Інтелектуальна система прогнозування цін на аграрну продукцію в Україні на базі Android-застосунку", було проведено високоінформативний аналіз часових рядів цін на аграрну продукцію. Враховуючи важливість прогнозування для аграрного сектору України, метою було виявлення найефективніших методів прогнозування. У результаті цього етапу, було детально проаналізовано моделі ARIMA та Хольта-Вінтерса, встановивши їхню придатність та ефективність для прогнозування цінових тенденцій.

Спроектвані та налаштовані моделі прогнозування виявилися корисними у контексті цінових динамік аграрної продукції, надаючи точні та адекватні прогнози. Це стало можливим завдяки глибокому розумінню вибраних моделей та їхній адаптації під конкретні умови українського аграрного ринку.

У сфері програмної реалізації розроблено багаторівневу структуру, яка включала в себе вивчення та використання відомих бібліотек для моделювання прогнозів. Додаток для Android був розроблений з урахуванням вимог ефективності та зручності взаємодії, інтегруючи результати прогнозування цін.

Одним із ключових аспектів дослідження було ретельне вивчення та аналіз існуючих бібліотек, які використовуються для розробки прогнозуючих моделей. Це дозволило забезпечити не лише правильність прогнозів, а й оптимальність їхньої обчислювальної складності.

Android-застосунок був розроблений із метою реалізації інтуїтивно зрозумілого інтерфейсу, що сприяє легкій навігації та використанню користувачем.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Костецький Я. І. Новітня парадигма розвитку аграрного сектору України: дис. ... д-ра екон. наук : 08.00.03. Тернопіль : ТНЕУ, 2019. 473 с. URL: <http://dspace.tneu.edu.ua/handle/316497/38556> (дата звернення: 09.01.2024).
2. Огляд збитків від війни в сільському господарстві України. URL: <https://efaidnbmnnnibpcajpcgclefindmkaj/> https://kse.ua/wp-content/uploads/2022/06/Damages_report_issue1_ua-1.pdf (дата звернення: 15.01.2024).
3. Abraham, B. (1981). Missing observations in time series, *Commun. Stat.*, A10, 1643—1653
4. Engle, R. F. And Granger, C. W. J. (1987). Co-integration and error correction: representation, estimation, and testing, *Econometrica*, 55, 251—276.
5. Koopmans, L. H. (1974). *The Spectral Analysis of Time Series*, Academic Press, New York.
6. Ljung, G. M. (1993). On outlier detection in time series, *J. R. Stat. Soc.*, B55, 559—567.
7. Robinson, E. A. (1967). *Multichannel Time Series Analysis*, Holden-Day, San Francisco, CA.
8. Zakoian, J. M. (1994). Threshold heteroscedastic models, *J. Econ. Dyn. Control*, 18, 931—955.
9. Ціни. АПК-Інформ. Інформаційно-аналітичне агентство «АПКІнформ». URL: <https://www.apk-inform.com/uk/prices> (дата звернення 24.01.2024).
10. V. Chandola, A. Banerjee, V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, vol. 41(3), 2009, pp. 1–58.
11. Н. Huang. Rank Based Anomaly Detection Algorithms. *Electrical Engineering and Computer Science – Dissertations*, 2013, 331.

12. Pankratz Alan Forecasting with Univariate Box – Jenkins Models: Concepts and Cases. Wiley, 1983. 576 p. DOI: 10.1002/9780470316566
13. Bilan Y, Gavurova B., Gedek S., Tkacova A. The Composite Coincident Indicator (CCI) for business cycles. Acta Polytechnica Hungarica. 2017. Vol. 7, p. 71-90. URL: https://www.uni-obuda.hu/journal/Bilan_Gavurova_Stanislaw_Tkacova_78.pdf
14. Росляков А.А. Формування та регулювання цін на сільськогосподарську продукцію на ринку України. Ефективна економіка. 2023. № 9. URL: <http://www.economy.nayka.com.ua/?op=1&z=2356>
15. Bidyuk P., Prosyankina-Zharova T., Terentiev O. Modelling Nonlinear Nonstationary Processes in Macroeconomy and Finances. Advances in Computer Science for Engineering and Education. ICCSEEA 2018. Advances in Intelligent Systems and Computing / Ed. Hu Z., Petoukhov S., Dychka I., He M. Cham : Springer, 2019. Vol. 754. P. 735–745. URL: http://doi.org/10.1007/978-3-319-91008-6_72
16. Holt-Winter`s seasonal smoothing URL: <https://otexts.com/fpp2/holtwinters.html>
17. Бідюк П. І., Гожий О. П., Проектування комп'ютерних інформаційних систем підтримки прийняття рішень , 2010, 335 с.
18. Івахненко А. Г. Індуктивний метод самоорганізації моделей складних систем: навч. посібник. Київ: Наукова думка, 1982. 245 с.
19. Баклан І.В., Селін Ю.М., Шулькевич Т.В. Математичні моделі прогнозування часових рядів різної природи // Вістн. Херсонського національного техн. ун-та. - Херсон: ХНТУ, 2014. - Вид. 3 (50).– С.213- 218
20. Поставки CIF, FOB та CPT: як формуються ціни на зерно? Agravery. Аграрне інформаційне агентство. 2021. 26 листоп. URL: <https://agravery.com/uk/posts/show/postavki-cif-fob-ta-cpt-ak-formuutsa-cini-nazerno>
21. E. Andres et al., “An adaptive Theory of Computation online course in ActiveMath”, in Proc. 5th Int. Conf. Computer Science & Education, 2010, pp. 317—322. doi: 10.1109/ICCSE.2010.5593624
22. Murat Y. Expert Android Studio / Y. Murat, D. Onur., 2016.

23. Android Emulator [Електронне джерело]. – Режим доступу: URL <http://developer.android.com/tools/help/emulator.html>.