

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

« ____ » _____ 2022 р.

БАКАЛАВРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

**ІНФОРМАЦІЙНА СИСТЕМА ПРОГНОЗУВАННЯ ФІНАНСОВИХ
ПОКАЗНИКІВ ПІДПРИЄМСТВА**

Спеціальність 122 «Комп'ютерні науки»

122 – БКР – 401.21810318

Виконала студентка 4-го курсу, групи 401

_____ *О. О. Петроченко*

« » червня 2022 р.

Керівник: канд. техн. наук, доцент

_____ *І. О. Калініна*

« » червня 2022 р.

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

Рівень вищої освіти **бакалавр**
Спеціальність **122 «Комп'ютерні науки»**
(шифр і назва)
Галузь знань **12 «Інформаційні технології»**
(шифр і назва)

ЗАТВЕРДЖУЮ

Завідувач кафедри інтелектуальних
інформаційних систем, д-р техн. наук, проф.
_____ Ю. П. Кондратенко
« ____ » _____ 2022 р.

З А В Д А Н Н Я
на виконання кваліфікаційної роботи

Видано студенту групи 401 факультету комп'ютерних наук Петроченко Олександрі Олегівні.

1. Тема кваліфікаційної роботи «Інформаційна система прогнозування фінансових показників підприємства».

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

Затв. наказом Ректора ЧНУ ім. Петра Могили від «17» __12__ 2021 р. № 318.

2. Строк представлення кваліфікаційної роботи студентом «__» _____ 2022 р.

3. Вхідні (початкові) дані до роботи: історичні дані вартості акцій компанії за 4 роки.

Очікуваний результат: інформаційна система прогнозування вартості акцій компанії.

4. Перелік питань, що підлягають розробці (зміст пояснювальної записки):

- аналіз існуючих методів аналізу та прогнозування часових рядів;
- огляд існуючих технологій для вирішення задачі;
- розробка інформаційної системи прогнозування часових рядів визначеними методами із застосуванням однієї з розглянутих технологій;

5. Перелік графічного матеріалу: презентація.
6. Завдання до спеціальної частини: «Оцінка умов праці співробітників та проектування плану евакуації у разі пожежі».
7. Консультанти розділів роботи.

Розділ	Прізвище, ініціали та посада консультанта	Підпис
Спеціальна частина з охорони праці		

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

(підпис)

Завдання прийнято до виконання Петроченко О. О.
(прізвище та ініціали)

(підпис)

Дата видачі завдання « 25 » _____ листопада _____ 2022 р.

АНОТАЦІЯ

Бакалаврської кваліфікаційної роботи студентки групи 401 ЧНУ ім. Петра Могили Петроченко Олександри Олегівни на тему «Інформаційна система прогнозування фінансових показників підприємства».

Фондовий ринок є важливою частиною ринку капіталу, але процесам на ньому властиві постійні зміни, тому завжди є **актуальним** проводити їх аналіз та спробувати прогнозувати їх.

Об'єктом дослідження є щоденна вибірка вартості акцій компанії.

Предметом дослідження є існуючі моделі та методи прогнозування фінансових процесів фондових ринків.

Метою та завданням роботи є проаналізувати та дослідити методи прогнозування часових рядів, змоделювати власний прогноз вартості акцій компанії. Розробити інформаційну систему прогнозування часових рядів.

У дипломній роботі розглянуто поняття часового ряду та існуючі методи його прогнозування, таких як АРІКС, МГУА. Розроблено інформаційну систему обробки вхідних даних та прогнозування вартості акцій компанії на мові програмування R використовуючи пакет Shiny для графічного інтерфейсу.

Пояснювальна записка складається із трьох розділів загальним об'ємом 72 с.

Дипломна робота містить 72 с., 30 рис., 3 табл., 2 додатки, 30 джерел.

Ключові слова: *часовий ряд, прогнозування, АР, АРКС, АРІКС, МГВА, лінійна регресія, R, Prophet.*

ABSTRACT

Bachelor's degree work of a student of group 401 of Petro Mohyla Black Sea National University Petrochenko Oleksandra on the topic "Information system of forecasting financial indicators of the company"

The stock market is an important part of the market of capital, but the processes in it are constantly changing, so it is always important to analyze and try to predict them.

The object of the thesis is the daily change in the value of the company's shares.

The subject of research is financial models and methods of forecasting stock markets.

The purpose and task of the work are to analyze and investigate the methods of forecasting time series, and to model their own forecast of the value of the company's shares. Develop an information system for forecasting time series.

In the thesis, the concept of time series and such methods of its forecasting as ARIMA and GMDH were considered. An information system for processing input data and forecasting the value of shares in R programming language with the help of the Shiny package for the graphical interface has been developed.

The explanatory note consists of three sections with a total volume of 72 p.

The thesis contains 72 p., 30 pictures, 3 tables, 2 addings, 30 sources.

Keywords: *time series, forecast, AR, ARMA, ARIMA, GMDH, linear regression, R, Prophet* .

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ.....	3
ВСТУП.....	5
1 МОНІТОРИНГ ФІНАНСОВИХ ПРОЦЕСІВ ТА ПРИКЛАДІВ	
ЇХ МОДЕЛЮВАННЯ. ПОСТАНОВКА ЗАДАЧІ.....	6
1.1 Основні задачі та методи прогнозування ринку цінних паперів	6
1.2 Огляд існуючих платформ прогнозування часових рядів	10
1.3 Огляд публікацій.....	15
1.4 Висновки до розділу та постановка задачі	16
2 МОДЕЛІ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ.	
СЕРЕДОВИЩА СТАТИСТИЧНИХ ОБЧИСЛЕНЬ	18
2.1 Опис моделей та підходів прогнозування	18
2.2 Огляд існуючих технологій для прогнозування часових рядів.....	27
Висновки до розділу	31
3 ПРОЄКТУВАННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ ПРОГНОЗУВАННЯ	
3.1 Вибір даних для моделювання.....	32
3.2 Побудова моделей	34
3.3 Аналіз отриманих результатів	40
3.4 Керівництво користувача	41
Висновки до розділу 3	48
4 ОХОРОНА ПРАЦІ	49
ВИСНОВКИ.....	62
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	63
ДОДАТОК А Норми штучного та природного освітлення	
виробничих приміщень.....	66
ДОДАТОК Б Лістинг програми.....	67

ПЕРЕЛІК СКОРОЧЕНЬ

АР – авторегресія

АРКС – авторегресія ковзного середнього

АРИКС – авторегресія інтегрованого ковзного середнього

МГВА – метод групового врахування аргументів

ІС – інформаційна система

ARIMA – autoregressive integrated moving average

GMDH – Group method of data handling

LTSM – Long short-term memory

СКП – середньоквадратична похибка

АКФ – автокореляційна функція

Пояснювальна записка
до кваліфікаційної роботи

на тему:

**«ІНФОРМАЦІЙНА СИСТЕМА ПРОГНОЗУВАННЯ ФІНАНСОВИХ
ПОКАЗНИКІВ ПІДПРИЄМСТВА»**

Спеціальність 122 «Комп'ютерні науки»

122 – БКР – 401.21810318

Виконала студентка 4-го курсу, групи 401

_____ Петроченко О. О.

(підпис, ініціали та прізвище)

« ___ » _____ 2022 р.

Керівник: канд. техн. наук, доцент _____

(наук. ступінь, вчене звання)

_____ Калініна І. О.

(підпис, ініціали та прізвище)

« ___ » _____ 2022 р.

Миколаїв – 2022

ВСТУП

Ринок цінних паперів є частиною ринку капіталу, де здійснюється купівля та продаж акцій (цінних паперів). Акції дозволяють їх утримувачу отримувати дивіденди (деяку частину прибутку), компенсації та іноді брати участь в управлінні підприємством. Але економічно-фінансовим системам властиві постійні зміни за об'єктивних та суб'єктивних факторів, а отже, ціна акцій постійно змінюється, що може призвести до фінансових збитків власнику акцій. Тому перед будь-якими людьми, які залучені у фондовий ринок, постає задача передбачення його поведінки, що стає у нагоді власне компанії, її інвесторам та пересічним утримувачам акцій. Правильний прогноз вартості активів зменшить інвестиційний ризик для інвестора та захистить від волатильності ринку.

Об'єктом дослідження є щоденна вибірка вартості акцій компанії PepsiCo за 4 роки (з 15.05.19pp. по 15.05.2022 pp.). Джерелом є провайдер фінансової інформації Yahoo!Finance.

Предметом дослідження є існуючі моделі та методи прогнозування фінансових процесів фондових ринків.

Метою та завданням роботи є проаналізувати та дослідити методи прогнозування часових рядів, змоделювати власний прогноз вартості акцій компанії. Розробити інформаційну систему прогнозування на мові R у середовищі RStudio.

1 МОНІТОРИНГ ФІНАНСОВИХ ПРОЦЕСІВ ТА ПРИКЛАДІВ ЇХ МОДЕЛЮВАННЯ. ПОСТАНОВКА ЗАДАЧІ

1.1 Основні задачі та методи прогнозування ринку цінних паперів

Прогнозування є основою всієї системи управління економіки та фондового ринку зокрема. Розвиток системи альтернативних прогнозів дозволить істотно зменшити ризики.

1.1.1 Опис фондового ринку та його значення

Фондовий ринок є однією з найважливіших ланок економіки будь-якої країни, включаючи Україну. Хоча ринок цінних паперів України почав формуватися в 1991 році, він все ще знаходиться на стадії становлення і характеризується пошуком форм та інструментів для визначення поведінки основних гравців в обміні конкретних товарів – цінних паперів, тобто акцій та облігацій. Акція наділена певною вартістю, є неподільною, прибутковою, бере участь в обігу та не має строку. Для вдосконалення механізму фондового ринку необхідно проводити якісний аналіз ринкової ситуації, відстежувати та прогнозувати ринкові зміни, знаходити нові шляхи вдосконалення, тобто проводити економічне прогнозування.

У економічному прогнозуванні існує 2 загальних підходи: суб'єктивні та засновані на моделях. Суб'єктивні зазвичай спираються на власні розмірковування експерта, а отже, не слідують строгим правилам. Але економіст все ж повинен мати під собою базоване підґрунтя.

Дослідники зазвичай розглядають ризики (події або умови, які можуть спричинити відмінність реальних результатів від їх початкових оцінок). Ці ризики допомагають проілюструвати обґрунтування, використане для отримання остаточних прогнозних даних. Під час підготовки економічних прогнозів, щоб підвищити точність, використовується найрізноманітніша інформація.

Економісти вибирають, які змінні важливі для матеріалу, про який йде мова. Вони можуть використовувати статистичний аналіз історичних даних, щоб визначити очевидний зв'язок між конкретними незалежними та залежними змінними. Спеціалісти проводять статистичні тести та розробляють статистичні моделі, щоб визначити, які зв'язки найкраще описують або прогнозують поведінку досліджуваних змінних.

1.1.2 Поняття часового ряду та підходи для його аналізу

Загальні відомості

Однією із структур представлення історичних даних є часовий ряд, що представляє собою набір точок, у яких індексами виступають рівновіддалені дати. Для часового ряду для можуть бути характерні закономірності, що повторюються через окремі проміжки часу, що називається сезонністю. Часові ряди легко представити графічно. Тривіальним прикладом часового ряду є курс валют (рис. 1.1). За його допомогою можна розробити модель для прогнозування майбутньої ціни.

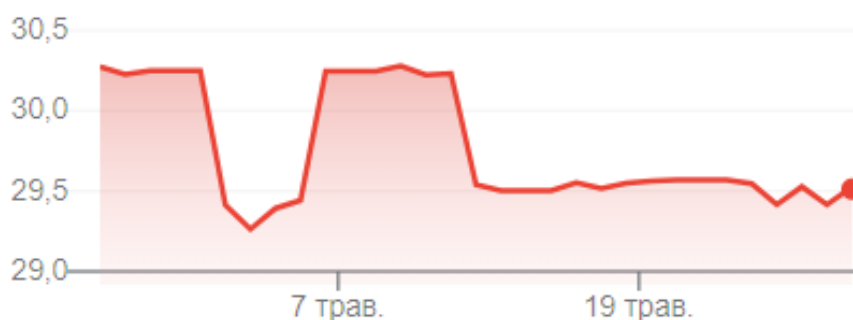


Рисунок 1.1 – часовий ряд курсу долара США до гривні

Велика кількість прикладних наук (економетрика, статистика, кліматологія, прогнозування погоди тощо) використовують часові вимірювання, отже можна використовувати часові ряди та аналізувати їх, тобто структурувати дані та прогнозувати їх.

Більшість фінансових процесів є нестационарними, а значить їх характер може змінюватися з часом – вони мають тренд. Найпростіший приклад стаціонарного процесу – білий шум. Тренд можна визначити як поточне середнє значення. Він визначається у процесі дослідження експериментальних даних, може бути детермінованим (показник довгострокових змін) та стохастичним (індикатор короткострокових та середньострокових змін). У процесі побудови прогнозу застосовують трендову та сезонну компоненти, що є незалежними одна від одної. Якщо ряд має змінну дисперсію, то він називається гетероскедастичним. Одна із задач аналізу даних полягає у знаходженні закону, за яким вона змінюється.

«Процеси з трендами або зі змінною дисперсією особливо характерні для нестійкої швидкоростучої перехідної економіки» [1], якою є економіка України.

Економічні та фінансові процеси можна назвати нелінійними. Це означає, що процес може істотно змінювати напрямок під впливом різних факторів. Нелінійність може проявлятися під величезним впливом одних факторів, а до факторів іншої природи бути байдужою.

У сучасному світі економічні процеси мають надзвичайно високі темпи змін, тобто значення змінних можуть швидко збільшуватися або зменшуватися. При цьому стаціонарні та нестационарні процеси можуть мати сезонну компоненту, яка посилює динаміку (швидкість зміни) досліджуваного процесу.

Існує дві загальних моделі побудови прогнозів на основі часових рядів: мультиплікативна та адитивна. За мультиплікативної моделі компоненти виражені коефіцієнтами, для оцінки явища використовується їх добуток (1.1). За адитивної компоненти сумуються (1.2).

$$x(t) = T(t) \times S(t) \times C(t) \times E(t), \quad (1.1)$$

$$x(t) = T(t) + S(t) + C(t) + E(t), \quad (1.2)$$

де, $T(t)$, $S(t)$, $C(t)$, $E(t)$ – компоненти часового ряду.

Перевірка на стаціонарність

Нульовою гіпотезою у статистиці називають загальне припущення (зазвичай позначається H_0) про те, що між двома явищами, що вимірюються, не існує зв'язку; гіпотеза, що її заперечує, називається альтернативною (H_1). Існують тести, що перевіряють нульову гіпотезу нестационарності часового ряду. До них належать:

- Тест Дікі – Фулера.
- Тест Квятковського – Філіпса – Шмідта – Шіна (KPSS).
- Тест Філіпса – Перрона (PP).

Всі ці тести розраховані так, що якщо їх вихідне значення $p > 0,05$, тоді часовий ряд називається нестационарним.

Фільтрування даних

Іноді у прогнозованого явища відсутні або не спостерігаються стійкі закономірності, але присутні зашумленість, зайва інформація, завеликі коливання. Для запобігання цьому застосовуються методи згладжування або фільтрації даних, які замінюють фактичні значення на прогнозні, більш підходящі для точного прогнозування.

Досить ефективним та поширеним є рекурсивний метод Кальмана, що використовується при розрахунках оптимальних оцінок залишків і побудови короткострокових прогнозів на основі обраних моделей. Фільтрування відбувається у 2 кроки. Спочатку відбувається екстраполяція, тобто побудова прогнозу поточного значення на основі попереднього. На другому отримані значення корегуються використанням середнього зваженого.

Також є популярними такі методи фільтрування [2]:

1. Медіанне. Масив N останніх значень сортується за зростанням, відфільтрованим буде значення посередині.

2. Метод ковзного середнього. Вихідне значення є середнім на певному ковзному інтервалі, тобто на вихідне значення впливає N останніх значень:

$$y_t = \frac{1}{N} \sum_{i=0}^{N-1} x_{t+1}, \quad (1.3)$$

де x_t – вхідний сигнал,

y_t – вихідний сигнал,

N – кількість точок інтервалу.

3. Еспоненційне згладжування. На відфільтровані значення впливають усі попередні значення. Записується за допомогою рівняння:

$$y_t = \theta x_t + (1 - \theta)y_{t-1}, \quad (1.4)$$

де x_t – обчислене значення у момент t ,

θ – коефіцієнт фільтрації, $[0;1]$,

y_t – відфільтроване значення у момент t .

4. Поліноміальне.

Підготовка даних

На практиці зібрані дані можуть бути спотворені, неправильно введені та оброблені, зашумлені чи взагалі відсутні. Тому постає задача заміни таких пропусків на приблизні значення.

Існують такі підходи до розв'язання цієї задачі:

- Заповнення найбільш частим або середнім значенням.
- Ігнорування пропущених значень.
- Метод найближчих сусідів.
- Метод випадкового лісу.
- Заміна за допомогою бустингу.

1.2 Огляд існуючих платформ прогнозування часових рядів

Існує досить багато готових програмних рішень з обробки, аналізу та прогнозування часових рядів. Нижче розглянуто деякі програми, що доступні на платформі Windows.

1.2.1 GMDH Shell

GMDH Shell – це професійне програмне забезпечення на основі нейронної мережі, яке вирішує проблеми прогнозування часових рядів. У системі присутні графічне представлення даних, побудова імітаційних моделей, аналітика, візуалізація даних, прогнозування.

Програма підтримує формат файлів .csv та .xlsx, після вибору файлу з даними пропонується обрати шаблон: кластеризація, класифікація, згладжування, регресійна модель, різні види прогнозів. Після цього програма починає обчислення, у випадку вибору шаблону «Time series forecast», отримується графік часового ряду зі змодельованими та зпрогнозованими даними, смугою впевненості (рис 1.2). Прогнозовані дані можна продивитися як у таблиці, так і на самому графіку (рис. 1.3).



Рис. 1.2 – вікно програми із результатами прогнозування



Рис. 1.3 – зпрогнозовані значення

Окрім самих значень, програма також обраховує інтервал довіри, коефіцієнт кореляції, відхилення, діаграми по залишкам.

Під час користування програмою було виявлено такі її переваги:

- 14-ти денний безкоштовний період;
- інтуїтивно зрозумілий та сучасний інтерфейс;
- підтримка багатьох мов;
- широкий функціонал.

А також деякі недоліки:

- робота із командним рядком для кращої кастомізації доступна тільки у бізнес-версії;
- переважно призначається для не дуже великих об'ємів даних, більший об'єм суттєво впливає на швидкодію та швидкість програми.

1.2.2 NCSS 2022

NCSS надає повну та просту у використанні колекцію сотень статистичних і графічних інструментів для аналізу та візуалізації даних.

NCSS містить набір інструментів для часових рядів і прогнозування, такі як ARIMA, спектральний аналіз, декомпозиція, експоненціальне згладжування та кореляції. Кожен часовий ряд і процедура прогнозування прості у використанні та перевірені на точність.

На рис. 1.4 наведено вікно роботи з набором даних. Підтримуються основні формати зберігання даних: .txt, .csv, .xlsx, а також присутній власний формат .NCSS.

The screenshot shows the NCSS 9 Data software interface. The main window displays a data table with the following columns: Date, Intel_Close, Intel_Volume, C4, C5, C6, C7, C8, C9, C10, and C11. The data is as follows:

	1	2	3	4	5	6	7	8	9	10	C1
Name	Date	Intel_Close	Intel_Volume	C4	C5	C6	C7	C8	C9	C10	C11
Data Type	DateTime	General	General	General	General	General	General	General	General	General	General
Format	d										
1	3/15/4503	65	11242.2								
2	3/16/4503	65	16689.9								
3	3/17/4503	62.75	14613.3								
4	3/18/4503	63	8009								
5	3/21/4503	63.875	6441.8								
6	3/22/4503	65.25	7664.5								
7	3/23/4503	66.75	8330.3								
8	3/24/4503	66.25	7983								
9	3/25/4503	66.5	8767.1								
10	3/28/4503	67	6266.4								
11	3/29/4503	67.25	8915.3								
12	3/30/4503	67.125	8833								
13	3/31/4503	67.125	8709.7								
14	4/1/4503	66.25	9603								
15	4/4/4503	65	21185.2								
16	4/5/4503	62.375	16006.5								
17	4/8/4503	61.125	11832.4								
18	4/11/4503	60.875	9168.1								
19	4/12/4503	59.75	17729.3								
20	4/13/4503	60.875	11500.7								
21											

Рис. 1.4 – інтерфейс NCSS

Присутні такі переваги:

- інтуїтивно-зрозумілий інтерфейс;
- відео-гайди та підтримка від виробників;
- зручна довідкова система;
- широкий функціонал роботи з даними та вибір їх візуалізації.

А також недоліки:

- функціонал безкоштовної версії обмежений у порівнянні із платною;
- є проблеми із експортом та імпортом даних;
- на ринку присутні більш потужні програми за меншу ціну.

1.2.3 SAS

Система SAS містить набір програмних пакетів, з яких варто приділити увагу SAS Enterprise Miner. Цей стандартний компонент SAS призначається для аналізу

інформації для подальшого прийняття рішень. SAS Enterprise Miner містить потужний інструмент для обробки та пошуку шаблонів у текстових даних, пакет сучасних статистичних методів та зручний інтерфейс для візуалізації даних. Тож мета пакету: оперування, аналіз та візуалізація даних.

Дані можуть бути завантажені із файлів різних типів, а також введені вручну.

Під час обробки даних вони перевіряються на наявність помилок, виправляються, перетворюються на зручний вигляд, перевіряються на правильність і зберігаються для подальших маніпуляцій. У процесі аналізу будуються прогнозні моделі, застосовуються статистичні методи обробки даних. Під візуалізацією розуміється створення графіків, діаграм для відображення результатів аналізу

SAS Enterprise Miner підтримує .txt, .csv, .dat формати, є інтеграція з базами даних Oracle, Teradata.

За допомогою графічного інтерфейсу користувачі можуть проходити всі етапи процесу аналізу даних: вибір джерел даних, дослідження та модифікації, моделювання та оцінка якості моделей, подальше використання отриманих моделей. Головне вікно Enterprise Miner показано на рисунку 1.5.

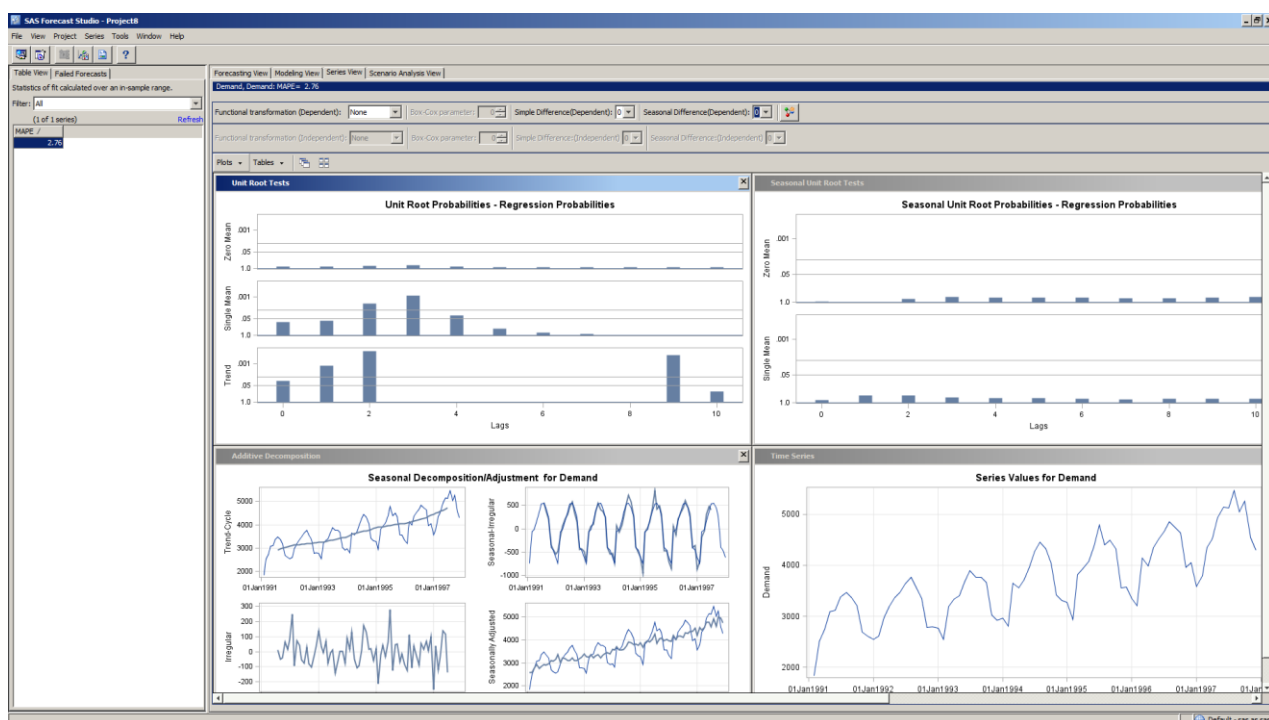


Рис. 1.5 – інтерфейс програми

Було знайдено такі переваги:

- простий синтаксис, швидке навчання початківців;
- інструментарій для роботи з кластерами;
- підтримка баз Oracle та Teradata;
- зручний формат вихідних даних (особливо таблиць);

Серед недоліків можна виділити:

- професійне використання передбачає покупку програмного продукту, що обмежує вивчення мови студентами;
- вихідні тексти багатьох виконуваних алгоритмів SAS не є публічними, отже вивчення роботи мови сильно обмежена.

1.3 Огляд публікацій

С. Наміні та А. Намін [3] досліджували, які методи прогнозування дають найкращі прогнози із меншою похибкою і вищою точністю: модель АРІКС чи нейронна мережа LSTM. Автори використали історичні дані ціни акцій декількох компаній, точніше кореговану вартість на час закриття біржі. Кожен набір даних часових рядів був розділений на дві підмножини: навчальні і тестові набори даних, де 70% було використано для навчання, а решта 30% – для перевірки точності моделей. Ці дві методики були впроваджені та застосовані до набору фінансових даних та результати показали, що LSTM перевершує ARIMA. Точніше, алгоритм на основі LSTM покращив прогноз на 85% у середньому порівняно з ARIMA. Крім того, стаття повідомляє про відсутність покращень при зміні кількості епох.

Сяо та Сю [4] також порівнювали роботу традиційної ARIMA та LSTM використовуючи дані вартості акцій компанії S&P. Перші 90% використовувались для навчання моделі, а 10% для передбачення. Далі був створений гібрид ARIMA-LSTM. Точність прогнозування гібридної моделі ARIMA-LSTM виявилася кращою, ніж інші фінансові моделі. Результати тесту функції втрат показали, що MSE, MAE та RMSE гібридної моделі ARIMA-LSTM менші, ніж інші контрольні

моделі. Тому модель ARIMA-LSTM можлива для прогнозування коефіцієнта кореляції оптимізації портфеля цінних паперів.

Цай і Ван [5] провели дослідження, де спробували передбачити ціни акцій за допомогою дерев прийняття рішень і штучних нейронних мереж. Вони взяли набір даних із тайванського фондового ринку, взяли до уваги фундаментальні, технічні і макроекономічні індекси. Гібрид дерева рішень та штучної нейронної мережи, навчений за даними фондової біржі, показав продуктивність F-score 77%. Ефективність до 67% F-score показали моноалгоритми.

1.4 Висновки до розділу та постановка задачі

В першому розділі було описано розвиток економічних процесів та їх характер. Більшість економічних процесів є нестационарними, особливо у країнах із перехідною економікою. Було оглянуто поняття часового ряду, його властивостей, методів його підготовки до подальшого аналізу. Були приведені існуючі програми із статистичного аналізу процесів різної природи, що виражені у формі часового ряду. Також можна зробити висновок, що тема прогнозування часових рядів є популярною серед дослідників та актуальною у економічних обчисленнях.

Існують різні моделі та методи прогнозування нелінійних нестационарних процесів, які будуть розглянуті у наступному розділі. Наявні інструменти дають можливість розв'язувати задачі моделювання і прогнозування з високою ефективністю, але, при цьому, виникає потреба будувати власні програмні засоби для виконання нестандартних функцій і підвищення якості кінцевих результатів моделювання і прогнозування.

Постановка задачі:

1. Обрати набір даних вартості акцій однієї або декількох компаній за певний проміжок часу.

2. Побудувати моделі вибраних часових рядів та зробити їх короткостроковий прогноз.

3. Розробити програмний пакет для із реалізацією цих моделей.
4. Провести експерименти порівняння реальних даних та обчислених.

Вимоги до програмного застосунку, що буде розроблено:

1. Програма повинна мати сучасний мінімалістичний, зрозумілий дизайн.
2. Дані для обробки програми завантажуються із файлу, має бути вибір колонок із потрібним набором даних.
3. Дані із файлу відображуються на екрані, розрахунки, побудова моделей та функцій відбувається після вибору поля даних та натискання кнопки початку розрахунку.
4. Значення окремих констант моделювання, де можливо, можуть бути змінені користувачем. Модель при цьому оновлюється автоматично.

2 МОДЕЛІ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ. СЕРЕДОВИЩА СТАТИСТИЧНИХ ОБЧИСЛЕНЬ

Існує багато підходів рішення задачі прогнозування часових рядів. У цьому розділі описуються деякі математичні моделі, а також програмні середовища для створення власних обчислень.

2.1 Опис моделей та підходів прогнозування

2.1.1 Моделі регресійного аналізу

Фінансові та економічні процеси можна представити у вигляді кількісних характеристик, між якими потрібно знайти залежність, модель залежності може виражатися у вигляді певної функції. Пошук цієї функції є задачею регресійного аналізу, на відміну від кореляційного, де з'ясовується міцність зв'язку.

Цілями регресійного аналізу є:

- 1) Прогнозування значення одної залежної змінної від іншої або інших.
- 2) Визначення ступеня детермінованості залежної змінної незалежною або незалежними змінними.
- 3) Визначення зв'язку між окремими незалежними змінними та варіацією залежної.

«Регресійна модель – функція незалежної величини та коефіцієнтів з включеними випадковими змінними» [6].

Залежна змінна зазвичай описується як сума значень деякої моделі та незалежних змінних. Залежно від характеру розподілу залежної змінної робиться припущення – гіпотеза породження даних. Для її підтвердження або спростування проводиться аналіз залишків.

Регресійні моделі поділяються на:

- однопараметричні виду $y = f(x)$:
 - лінійні моделі $y = \beta_0 + \beta_1 x$;

- нелінійні моделі (степенева, квадратична, експоненційна тощо);
- багатопараметричні виду $y = f(x_1, x_2, \dots, x_n) + \varepsilon$:
 - лінійні моделі $y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$;
 - моделі, що є нелінійними за змінними та параметрами.

Лінійна множинна регресія відноситься до багатопараметричних регресій та набуває виду:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \varepsilon, \quad (2.1)$$

де y – залежна змінна,

x_1, x_2, x_n – незалежні змінні,

$\beta_0, \beta_1, \beta_n$ – коефіцієнти регресії,

ε – випадкова похибка.

У випадку декількох (i) спостережень модель записується або у матричному вигляді або для кожного спостереження за формулою:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_n x_{in} + \varepsilon_i \quad (2.2)$$

При цьому ε_i не повинна мати систематичного зміщення, що вирішується, коли у рівнянні регресії присутня константна величина. Незалежні змінні мають бути не випадковими.

Також класична модель лінійної регресії додатково може відповідати таким вимогам щодо її специфікації та експериментальних даних:

- $\forall i \neq j E(u_i, u_j | x_i) = 0$ (відсутність кореляції залишків);
- $\forall i E(u_i^2 | x_i) = \sigma^2$ (гомоскедастичність, тобто незалежність значення дисперсії випадкової похибки від номера спостереження).

Втім, більшість реальних моделей, як-от фондові ціни чи розміри продажів, є надто складними для лінійної регресії. А отже необхідні комплексні методи для передбачення їх майбутніх значень.

2.1.2 Моделі авторегресії

Для опису стаціонарних економічних процесів використовуються авторегресивні моделі. В економіці, в основі побудови авторегресивних моделей лежить важлива властивість ряду економічних явищ і процесів – взаємозалежність між рівнями одного і того ж ряду. Умова нормальності розподілу даних не є обов'язковою при побудові їх економетричної моделі [7].

Відмінною рисою стаціонарних часових рядів є те, що їх розвиток відбувається за постійних стабільних умов без видимої тенденції, тому імовірнісний характер ряду залишається незмінним у часі. Функція розподілу стаціонарного часового ряду також залишається постійною протягом зсуву в часі.

Модель авторегресії

«Авторегресійні моделі описують процес, в якому поточне значення залежить від декількох попередніх значень цього ж процесу» [8]. Наступна модель буде називатися AR(p) – авторегресія p порядку, її можна представити так:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t, \quad (2.3)$$

де t – номер процесу, ε_t – білий шум, не корельований з лагами y_t .

Лагові змінні – пояснювальні змінні, що беруться у регресійній моделі із запізненням у часі. Модель авторегресії можна переписати (2.5), якщо ввести лаговий оператор L (2.4):

$$\alpha(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p, \quad (2.4)$$

$$\varepsilon_t = y_t(1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p) = \alpha(L)y_t \quad (2.5)$$

Для авторегресій різного порядку будуть різні умови стаціонарності. Так, AR(1):

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t, \quad (2.6)$$

де $cov(\varepsilon, y_{t-1}) = 0$, $\alpha_1 = \alpha$.

Якщо записати дисперсію з двох сторін, отримається $\sigma_y^2 = +\alpha\sigma_y^2 + \sigma_\varepsilon^2$, а отже умовою стаціонарного процесу буде $|\alpha| < 1$.

Для AR(2) умовою стаціонарності буде $\alpha_1 \pm \alpha_2 < 1, -1 < \alpha_2 < 1$.

Залишки

Залишки в моделі часового ряду – це те, що залишається після підігнання (fitting) моделі. Для багатьох (але не для всіх) моделей часових рядів залишки дорівнюють різниці між спостереженнями та відповідними підігнаними значеннями.

$$e_t = y_t - \hat{y}_t \quad (2.7)$$

Залишки є корисними для перевірки того, чи адекватно модель охопила інформацію в даних. Добре підібрана модель прогнозування дає залишки з такими властивостями:

- залишки некорельовані. Якщо між залишками є кореляції, то в залишках залишається інформація, яку слід використовувати при обчисленні прогнозів;
- залишки мають нульове середнє. Якщо залишки мають середнє, відмінне від нуля, то прогнози є упередженими.

Модель ковзного середнього

Ковзне середнє (КС або МА (moving average)) є іншим інструментом аналізу часових рядів, він полягає в обчисленні середнього підмножини значень. Розмір цієї підмножини може бути як сталим, так і змінним [9].

КС описує лінійну залежність процесу від декількох (q) його попередніх значень (ε).

$$y_t = \varepsilon_t - \beta_1\varepsilon_{t-1} - \dots - \beta_q\varepsilon_{t-q}, \quad (2.8)$$

Якщо позначити оператор ковзного середнього як $\beta(L) = 1 - \beta_1 L - \beta_2 L^2 - \dots - \beta_q L^q$, то модель прийме вид:

$$y_t = \beta(L)\varepsilon_t \quad (2.9)$$

Процес КС(q) буде стаціонарним без накладання обмежень на параметри β . Математичне сподівання цього процесу буде дорівнювати 0, а дисперсія $\sigma_0^2 = (1 + \beta_1^2 + \dots + \beta_q^2)\sigma_\varepsilon^2 = C\sigma_\varepsilon^2$.

Модель авторегресії ковзного середнього

Змішаний процес авторегресії та ковзного середнього позначається АРКС(p,q) або ARMA (p,q) (2.10–2.11). Така модель виникла, щоб оцінювати характеристики часового ряду за допомогою меншої кількості параметрів.

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \dots - \beta_q \varepsilon_{t-q}, \quad (2.10)$$

$$\beta(L)\varepsilon_t = \alpha(L)y_t \quad (2.11)$$

де $\alpha(L)$ – оператор авторегресії,

$\beta(L)$ – оператор ковзного середнього.

Бокс і Дженкінс запропонували виділити клас нестаціонарних рядів, які можна привести до стаціонарного вигляду типу ARMA за допомогою взяття послідовних різниць. Якщо після взяття d послідовних різниць ряд зводиться до стаціонарного, то далі для прогнозування можна застосувати комбіновану модель авторегресії і ковзного середнього, що позначається як ARIMA (p, d, q) [10]. Формалізовано модель можна представити у вигляді:

$$\Delta^d X_t = c + \sum_{i=1}^p a_i \Delta^d X_{t-1} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t, \quad (2.12)$$

де c, a_i, b_j – параметри моделі,

$\Delta^d = 1 - L$ – оператор різниці значень часового ряду d порядку.

За методологією Бокса-Дженкінса, підбір ARIMA-моделі для конкретного ряду спостережень має чотири етапи:

- проводиться ідентифікація моделі, вибір такої, що найкраще підходить процесу, що розглядається;
- оцінювання моделі, тобто використання регресійних методів для отримання оцінок параметрів, що включені у модель;
- здійснюється тестування моделі – перевірка на адекватність за допомогою тестів на нормальний розподіл (тест Жака – Бера), автокореляцію залишків (тест Дарбіна – Уотсона), на якість специфікації моделі (F-тест);
- модель використовується для побудови прогнозу.

Під час побудови ARIMA(p, d, q) моделі часового ряду необхідно прагнути мінімізації числа її параметрів. Це правило відоме як «принцип економії» та полягає у перевазі простої моделі складнішою. Параметри моделей типу ARIMA оцінюються з урахуванням коефіцієнтів автокореляції вихідного процесу. Зі збільшенням числа параметрів у моделі для визначення їх значень необхідно використовувати більше число вибірових коефіцієнтів автокореляції як вихідні дані (з великими кроками). Точність їх оцінки зі зростанням кроку падає, що знижує надійність оцінок коефіцієнтів моделей тимчасових рядів високих порядків та якість цих моделей.

Приведення ряду до стаціонарного

ARIMA-модель вимагає, щоб досліджуваний ряд був стаціонарним, що в реальності буває далеко не завжди. Але є способи приведення ряду до стаціонарного процесу. Можна віднести такі:

- взяття кінцевих різниць: $X_t = \Delta Y_t = Y_t - Y_{t-1}$, де X_t – перша різниця.
 $Z_t = \Delta X_t = X_t - X_{t-1}$ – друга різниця;
- логарифмування ряду: $X_t = \ln(Y_t)$;

- логарифмування ланцюгових індексів: $X_t = \ln\left(\frac{Y_t}{Y_{t-1}}\right) = \ln Y_t - \ln Y_{t-1}$;
- розрахунок темпів росту: $X_t = \left(\frac{Y_t}{Y_{t-1}}\right)$;
- розрахунок темпів приросту: $X_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}} = \left(\frac{Y_t}{Y_{t-1}}\right) - 1$;

Послідовність значень вихідного ряду Y_t вважається ARIMA(p, d, q) моделлю, якщо послідовність значень проінтегрованого ряду $\Delta^d Y = (\Delta^d Y^t)$ утворює ARMA(p, q)-модель.

2.1.3 Моделі на основі методу групового урахування аргументів

Іншим досить ефективним методом моделювання та прогнозування систем є метод групового урахування аргументів. Він використовується для аналізу багатопараметричних даних, розпізнавання образів, визначення взаємозв'язків між даними.

Суть методу полягає у побудові моделей, їх перебору та вибору найкращої згідно з мінімумом середньоквадратичної похибки (СКП). Моделями можуть виступати лінійні, нелінійні, ймовірнісні функції та кластеризація. На кожному наступному кроці модель ускладнюється, а отже і збільшується точність моделювання. Як і у звичайному регресійному аналізі, дані поділяються на перевірочну та навчальну частини, де перша використовується для оптимізації коефіцієнтів рівняння регресії, а от друга – для оцінки ступеня регулярності за величиною СКП [12].

Основним результатом МГВА є те, що із зашумленими даними з малими розмірами вибірки та великою кількістю пропусків, мінімум критерію покаже нефізичні моделі з простішою структурою та більшою точністю, ніж фізичні моделі.

Зазвичай використовується така базова модель:

$$y = a + \sum_{i=1}^m b_i x_j + \sum_{i=1}^m \sum_{j=1}^m c_{ij} x_i x_j, \quad (2.13)$$

$$y = \omega_0 + \omega_1 x_i + \omega_2 x_j + \omega_3 x_i^2 + \omega_4 x_j^2 + \omega_5 x_i x_j, \quad (2.14)$$

де y – залежна змінна, x_i, x_j – факторні змінні, m – кількість факторів,
(a, b_i, c_i), ($\omega_0, \omega_1, \omega_2, \omega_3, \omega_4, \omega_5$) – вектори вагових коефіцієнтів.

Вектори вагових коефіцієнтів обчислюються за допомогою можливих коефіцієнтів регуляризації $\lambda = 0, 0.01, 0.02, 0.08, \dots, 10.24$ для кожної моделі. Після цього вхідні дані розділяють на перевірочну (70–80%) та навчальну (20–30%) частини та алгоритм йде таким чином:

- 1) Після уточнення можливих коефіцієнтів регуляризації, на навчальній вибірці обчислюються вагові коефіцієнти для кожного з них.
- 2) Далі для кожного коефіцієнта регуляризації визначають зпрогнозовані значення перевірочної вибірки та СКП (внутрішній критерій) (2.19).
- 3) Знаходжується коефіцієнт регуляризації, що дає найменший середній квадрат похибок.
- 4) Зовнішнім критерієм оцінки якості моделі часто виступає прогнозний СКП.
- 5) Обирається певний набір моделей (зазвичай 2–10) із найменшим значенням прогнозного СКП та алгоритм повторюється.

2.1.4 Критерії якості оцінок прогнозів

Точність моделі характеризується тим, наскільки близьке обчислене значення до фактичного спостережуваного значення на період апроксимації. Зазвичай моделі з меншими відмінностями між фактичними і розрахунковими значеннями краще відображають досліджуваний процес. Для характеристики близькості використовуються описові статистики середнього квадратичного відхилення, коефіцієнту детермінації, середньої відносної похибки апроксимації, максимального відхилення тощо [13].

Статистично, точність прогнозів можна оцінити за допомогою так званих ретроспективних прогнозів. Для цього інформація розділяється на дві частини. Перша частина, що охоплює давні спостереження, використовується для оцінки

параметрів побудованої моделі, друга частина є реалізацією прогнозу. Отримана помилка прогнозування характеризує точність застосованої методики прогнозування.

У наступних показниках y_i буде означати фактичне значення, f_i – прогнозоване, $e_i = y_i - f_i$ – значення похибки прогнозу, n – розмір тестового набору.

1) Середня похибка:

$$ME = \frac{1}{n} \sum_{i=1}^n e_i \quad (2.15)$$

2) Середня абсолютна похибка:

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (2.16)$$

3) Середня відсоткова похибка:

$$MPE = \frac{1}{n} \sum_{i=1}^n \frac{e_i}{y_i} \times 100\% \quad (2.17)$$

4) Середня відсоткова абсолютна похибка:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|e_i|}{|y_i|} \times 100\% \quad (2.18)$$

5) Середній квадрат похибок:

$$MSE = \frac{1}{n} \sum_{i=1}^n e_i^2 \quad (2.19)$$

6) Коефіцієнт невідповідності Тейла:

$$U = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n |e_i|^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2 + \frac{1}{n} \sum_{i=1}^n f_i^2}} \quad (2.20)$$

Якщо $U=1$, прогнозовані і фактичні ряди не є корельованими, тому модель не може дати якісний прогноз. Якщо ж $U=0$, то прогнозовані значення ряду співпадають з реальними, а отже модель вважається ідеальною.

2.2 Огляд існуючих технологій для прогнозування часових рядів

2.2.1 Пакет Matlab

Це зручний інструмент для роботи з математичними матрицями, роботи з алгоритмами, створення робочих оболонок з програмами на інших мовах програмування. Хоча продукт розроблений для чисельних обчислень, з програмним забезпеченням Maple можна використовувати спеціальні інструменти, що робить його повною системою для роботи з алгеброю.



Рисунок 2.1 – значок пакету Matlab

У пакеті реалізовано формат даних «timetable», який з'яє рядок даних із датою та/або часом. Крім того, у timetable'ы доступні функції узгодження, комбінування та виконання обчислень з даними з мітками часу.

Додатковий пакет Time Series Analysis toolbox можна використовувати для [14]:

1. Стохастичної обробки сигналів.
2. Ідентифікації авторегресивної моделі.
3. Обчислення ентропії часового ряду.
4. Згладжування, передбачення, фільтрація.
5. Тест Гурвіца та інші.

Включено кілька критеріїв (AIC, BIC, FPE, MDL, SBC, CAT, PHI) для авторегресивної моделі. Крім того, набір інструментів містить швидку версію методу Юла-Уокера для оцінки параметрів авторегресії, функцію автоковаріації

(ACovF), функцію автокореляції (ACF), часткову ACF (PACF) та деякі інші корисні функції. Версія 2.40 (і вище) містить швидкі алгоритми для перевірки поліномів; реалізовані ACovF і алгоритми Левінсона-Дурбіна.

2.2.2 Мова Python

Python є інтерпретованою високорівневою об'єктно-орієнтованою мовою програмування із строгою динамічною типізацією. Розроблена Гвідо ван Россумом у 1990 році. Високорівневі структури даних разом з динамічною семантикою та динамічним зв'язуванням роблять її привабливим для швидкої розробки ПЗ, а також є засобом об'єднання наявних компонентів. Python підтримує модулі та пакети модулів.

Для прогнозування часових рядів в Python доступні такі підходи, як ARIMA, ARCH тощо [15]. Однак у 2017 році команда Core Data Science (Facebook) випустила нову бібліотеку для роботи з часовими рядами – Prophet.

Навчання проводиться методом `.fit()`, будується передбачення за допомогою `.predict()`. Для навчання необхідно передати DataFrame колонками:

- `ds` – час, формат, поле повинно бути `datetime`;
- `y` – змінна в числовому форматі.

Щоб отримати передбачення, потрібно передати на новий DataFrame, що містить стовпець `ds`. Для цього в бібліотеці є функція `make_future_dataframe`, яка приймає параметри `periods`, для якого треба отримати прогноз, і частота часового ряду `freq`.

Prophet дозволяє вказувати дати свят. Також можна регулювати ступінь їх впливу. Пакет є досить популярним та ефективним у побудові моделей прогнозування.

2.2.3 Мова R та середовище RStudio

R є мовою програмування та програмним середовищем для статистичних обчислень, аналізу та графічного представлення даних. Розвиток R відбувався під значним впливом двох мов програмування: мови програмування S та семантикою успадкування від Scheme. Незважаючи на деякі принципові відмінності, більшість програм, написаних мовою програмування S, працюють у середовищі R. Мова має велику підтримку спільноти та різноманітні бібліотеки для аналізу даних.

Аналіз часових рядів у R часто супроводжується проблемами, що пов'язані з необхідністю представлення даних у вигляді об'єктів таких традиційних класів, як `ts`, `zoo` або `xts`. На жаль, ці класи погано підходять для сучасних наборів даних, які часто характеризуються нерегулярною реєстрацією спостережень у часі, наявністю кількох змінних різних типів, кількох групуючих змінних тощо. Крім того, традиційні формати представлення часових рядів суперечать принципам організації та зберігання “tidy data” і, як наслідок, ускладнюють аналіз та моделювання за допомогою широко використовуваних сьогодні інструментів групи `tidyverse`. Для вирішення цих проблем група дослідників розробила новий формат, реалізований у пакеті `tsibble`. Він характеризується такими властивостями [16]:

- дані зберігаються у табличному вигляді;
- у таблиці повинні бути присутніми як мінімум два стовпці – зі значеннями кількісної змінної, що спостерігається в часі, і з упорядкованими за зростанням часовими відмітками (`index`);
- крім того, до таблиці можуть входити одна або кілька групуючих змінних (`key`) – значення цих змінних вказують на належність кожного спостереження до відповідного часового ряду;
- будь-яке спостереження в таблиці можна унікально ідентифікувати за поєднанням значень індексувальної та групуючих змінних.

Мова R має широке коло пакетів обробки та аналізу часових рядів. Наприклад, одним з найбільш популярних є пакет `forecast`, в якому реалізовані як класичні (експоненційне згладжування, модель Хольта-Вінтерса, ARIMA тощо), так і нещодавно розроблені методи прогнозування (моделі для згрупованих часових рядів, рядів з кількома сезонними компонентами тощо). Важливий недолік полягає в тому, що всі реалізовані в `forecast` методи мають власні параметри налаштування, і навіть досвідчені аналітики не застраховані від вибору неправильного методу і/або набору параметрів для вирішення завдання.

R також підтримує пакет прогнозування часових рядів `prophet`. Він багато в чому позбавлений зазначених вище недоліків `forecast` та інших подібних пакетів і дозволяє створювати точні моделі в (напів-)автоматичному режимі.

Необхідність виявлення викидів або аномалій у часових рядах часто виникає у таких ситуаціях, як моніторинг стану устаткування, відстеження несподіваних коливань на ринку цінних паперів, облік показників стану здоров'я пацієнтів тощо. Для вирішення цього спеціально розроблено пакет `anomalize`.

Для більш зручної розробки було створено RStudio, що є вільним та відкритим інтегрованим середовищем розробки для R.

Існує дві версії RStudio: RStudio Desktop, яка працює як звичайний застосунок, і RStudio Server, який дозволяє отримати доступ до RStudio, що працює на віддаленому сервері Linux, через веб-інтерфейс. Дистрибутив RStudio Desktop доступний для Windows, OS X і Linux.

Для полегшення створення власних інтерактивних веб-застосунків з графічним інтерфейсом існує програмне середовище Shiny. Веб-застосунки, розроблені в Shiny, можна запускати зі спеціального хмарного сервера `shinyapps.io`, локального сервера користувача або платного сервера Rstudio.



Рисунок 2.2 – значок середовища Shiny

Висновки до розділу

Було розглянуто декілька методів моделювання часових рядів: на основі регресійних моделей та групового урахування аргументів. Перевагами регресійних моделей є гнучкість, простота та швидкість прогнозування. Недоліками ж є складності із визначенням параметрів та неможливість моделювання нелінійних явищ. До переваг моделей на основі групового урахування аргументів можна віднести більшу універсальність та точність прогнозування, але при цьому метод має меншу швидкість роботи.

Також було розглянуто основні середовища та мови розробки власної ІС. Серед перелічених було обрано мову програмування R та середовище Shiny через великі можливості статистичних обчислень, аналізу та різноманітність бібліотек.

3 ПРОЄКТУВАННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ ПРОГНОЗУВАННЯ

В цьому розділі описується процес проєктування та розробки інформаційної системи з прогнозування часових рядів. Програма написана на мові R, реалізовано моделі, що були описані у розділі 2. Програма може бути застосована для будь-якого набору даних.

3.1 Вибір даних для моделювання

3.1.1 Підбір економічних даних

Передбачити, як буде працювати фондовий ринок, є однією з найскладніших речей. Існує дуже багато факторів, які беруть участь у передбаченні – фізичні фактори та психологічні, раціональної та ірраціональної поведінки тощо. Усі ці аспекти в поєднанні роблять ціни акцій нестійкими і дуже важкими для передбачення з високим ступенем точності.

Серед історичних даних вартості акцій треба розрізняти декілька їх видів:

- Open – ціна на час відкриття біржі;
- High – найвища ціна за день;
- Low – найнижча ціна за день;
- Close – ціна на момент закриття біржі;
- Adj. Close – скоригована ціна на момент закриття після урахування корпоративних дій;
- Volume – кількість цінних паперів, що торгувалися цим днем.

Зазвичай для побудови моделей і подальшого прогнозування рекомендується обирати дані Close [17], але створена ІС є універсальною, де можна вибрати будь-який стовпець даних. Перший стовпець зазвичай називається Date та містить часові відмітки спостережень, тому спеціально вибирати його не потрібно.

Назву колонки з потрібними даних можна вибрати із віджету `selectInput`, що, як і всі інші віджети у пакеті Shiny, створений за допомогою пакету Twitter Bootstrap.

```
selectInput("selData", label = 'Choose data column', choices = c())
```

Головну панель ІС, де реалізовано завантаження файлу, вибору даних та кнопки початку розрахунків зображено на рисунку 3.1.

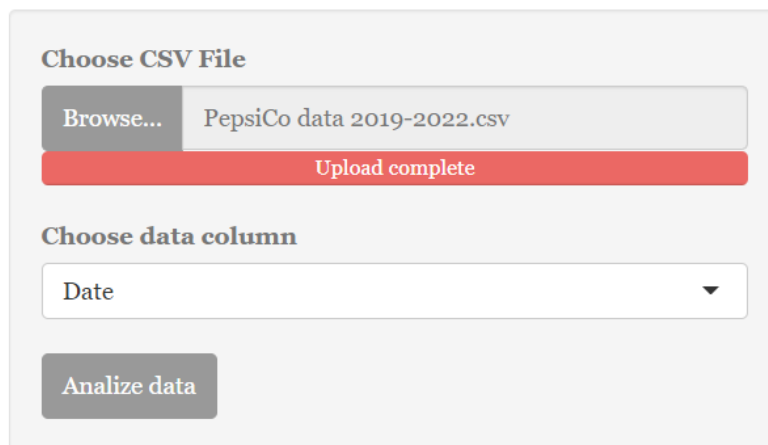


Рисунок 3.1 – панель вибору файлу та колонки з даними

3.1.2 Формат даних .csv

У деяких розглянутих у першому розділі програмах є можливість ввести дані вручну, але набір даних може бути завеликим для такого способу, тому буде більш зручним завантажувати готовий файл у ІС.

Як видно на рисунку 3.1, у ІС завантажується файл формату .csv. У RShiny це доступно за допомогою віджету `fileInput`, що, як і всі інші віджети, створений за допомогою пакету Twitter Bootstrap. Віджет спеціально налаштовано на прийом файлів формату .csv.

```
fileInput("file1", "Choose CSV File", accept = ".csv")
```

CSV є одним із варіантів представлення табличних даних, де поля відокремлюються комою та переходом на новий рядок. І хоча формат не є кінцево стандартизованим, він є дуже простим, і тому широко підтримується табличними процесорами, системами керування базами даних та мовами програмування.

Робочі дані вартості акцій компанії PepsiCo було взято з Yahoo!Finance [18], одного із найбільш популярних провайдерів фінансової інформації, він дозволяє завантажувати історичні дані у форматі .csv, тому було вирішено впровадити підтримку саме цього формату даних у ІС.

3.2 Побудова моделей

3.2.1 Розробка інтерфейсу

Загалом, структура shiny-застосунку така:

1) Опис UI елементів у частині `ui <- fluidPage()`, `fluidPage` – функція, що дозволяє автоматично підлаштуватися під розміри вікна браузера користувача.

2) Основна програмна частина описується у функції `server <- function(input, output, session){}`.

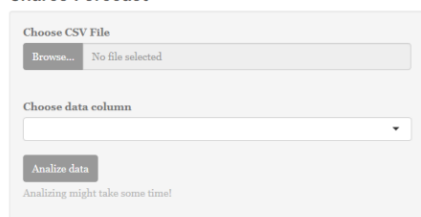
3) Виклик основної функції `shinyApp(ui = ui, server = server)`.

Всередині `fluidPage` було додано функції `sidebarLayout()` та `mainPanel()` для бокової та основних панелей інтерфейсу. У інтерфейсі використано такі функції:

- `fileInput` – завантаження файлу до системи.
- `selectInput` – вибір із списку.
- `actionButton` – кнопка.
- `textOutput` – звичайний показ тексту.
- `plotOutput` – показ графіку.
- `verbatimTextOutput` – показ тексту у консольному стилі.

Деякі елементи розроблені за допомогою проєкту Twitter Bootstrap. Кожен елемент має свій ідентифікатор, тому стає досить легким звертатися до них із серверної функції. Також доступна підтримка деяких HTML5 елементів, наприклад `<p>`, `<h1>`–`<h5>`. З такими інструментами легко створювати інтерактивні інтерфейси, як на рис. 3.2.

Shares Forecast



Data Decomposition ARIMA forecast GMDH forecast Prophet forecast

Рисунок 3.2 – головна сторінка ІС

3.2.2 Декомпозиція часового ряду

У вкладці декомпозиції відбувається показ часового ряду у вигляді графіку. Після завантаження файлу дані треба перетворити у формат `tsibble` із однойменного пакету (`as_tsibble(dataset, key=NULL, index=Date)`). `Tsibble` є форматом даних, де у кожного спостереження обов'язково мають бути присутні індекс (змінна з притаманним упорядкуванням від минулого до теперішнього) та ключ (набір змінних, які визначають одиниці спостереження в часі). Кожну одиницю спостережень слід вимірювати через загальний інтервал, якщо вони розподілені регулярно [19].

Для декомпозиції ряду треба перетворити його у формат `time-series`, що представляє собою вектор або матрицю, за допомогою функції `ts()` (`timeSeries <- ts(data=dataset, [...], frequency = 30)`). Розкладання часового ряду на сезонні, тенденційні та нерегулярні компоненти відбувається за допомогою функції `stl()` (`stlData <- stl(timeSeries[,1],s.window="periodic")`), результат можна побачити на рис. 3.3.

Decomposition of data

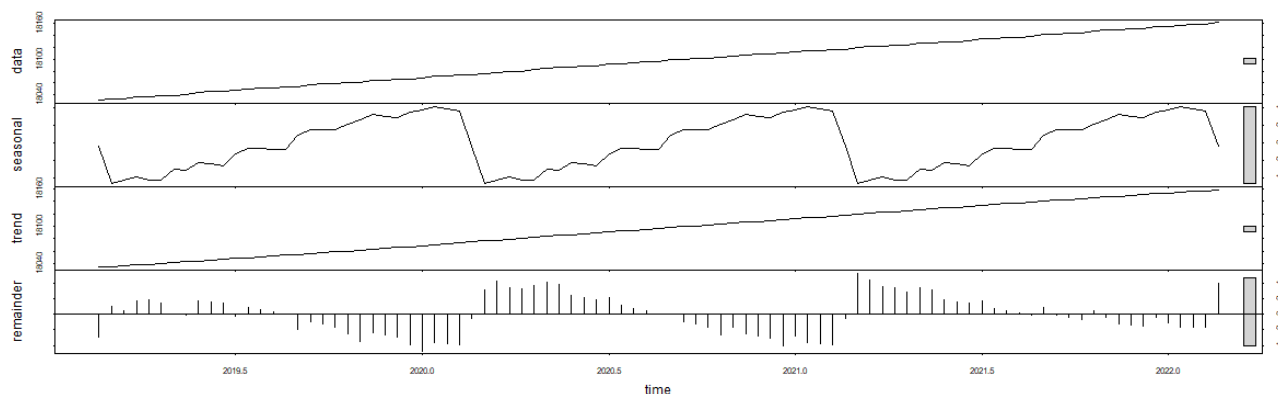


Рисунок 3.3 – графік декомпозиції ряду

Компоненти часового ряду показані на трьох нижніх графіках (trend – тренд, seasonal – компонента сезонності, remainder – залишки).

Також у вкладці декомпозиції можна провести тести на стаціонарність: Дікі – Фулера (ADF), KPSS та тест Філіпса – Перрона (PP); приклад коду: `adf.test(timeSeries[,1])`. Результати тестувань виводяться у вікна типу `verbatimOutput()` (рис. 3.4).

Tests on stationarity

Augmented Dickey-Fuller Test

```
data: timeSeries[, 1]
Dickey-Fuller = -2.5907, Lag order = 3, p-value = 0.3359
alternative hypothesis: stationary
```

KPSS Test for Level Stationarity

```
data: timeSeries[, 1]
KPSS Level = 1.6242, Truncation lag parameter = 3, p-value = 0.01
```

Phillips-Perron Unit Root Test

```
data: timeSeries[, 1]
Dickey-Fuller Z(alpha) = -36.481, Truncation lag parameter = 3, p-value = 0.01
alternative hypothesis: stationary
```

Рисунок 3.4 – приклад виконання тестувань

3.2.3 Побудова моделі ARIMA

Пакет `forecast` надає методи та інструменти для відображення та аналізу одновимірних часових рядів, включаючи експоненціальне згладжування за допомогою моделей простору станів і автоматичне моделювання ARIMA. Для побудови моделі використовується функція `auto.arima` (`arimaMod <- auto.arima(ts(filledData)[,2]))`, що приймає об'єкт формату `time-series`. Далі прогнозування – функція `forecast()`, що працює з багатьма різними типами вхідних даних (`forecast1 <- forecast(arimaMod, h = input$slideIn4)`). Зазвичай вона використовує часовий ряд або модель часового ряду як основний аргумент і створює прогноз вказаним способом. Вона завжди повертає об'єкти класу `forecast`.

У результаті цей об'єкт можна показати у вигляді графіку (рис. 3.5).

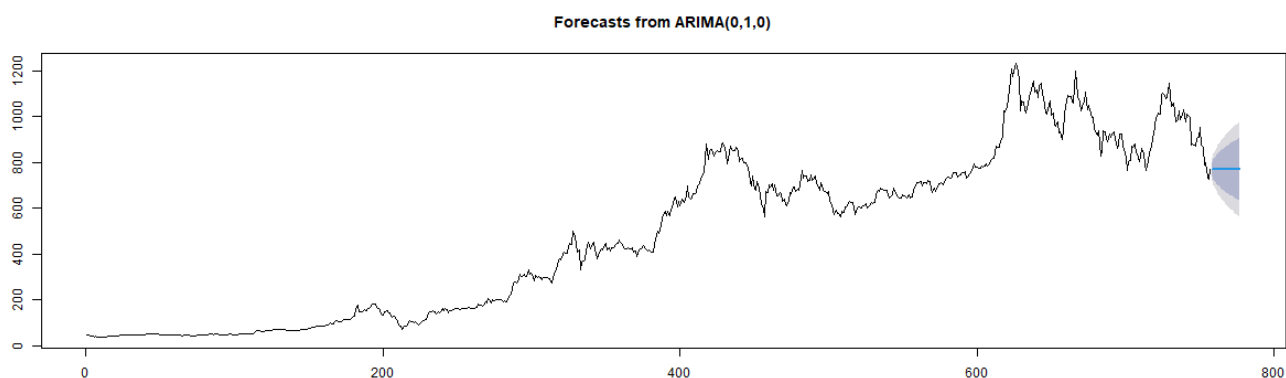


Рисунок 3.5 – прогнозування `auto.arima`

Для перегляду залишків моделі використовується функція `checkresiduals()` (рис. 3.6). На графіку можна побачити звичайний графік залишків моделі, АКФ залишки (автокореляційна функція) та гістограма. За допомогою цих графіків можна судити, наскільки добре підігнана модель під набір даних.

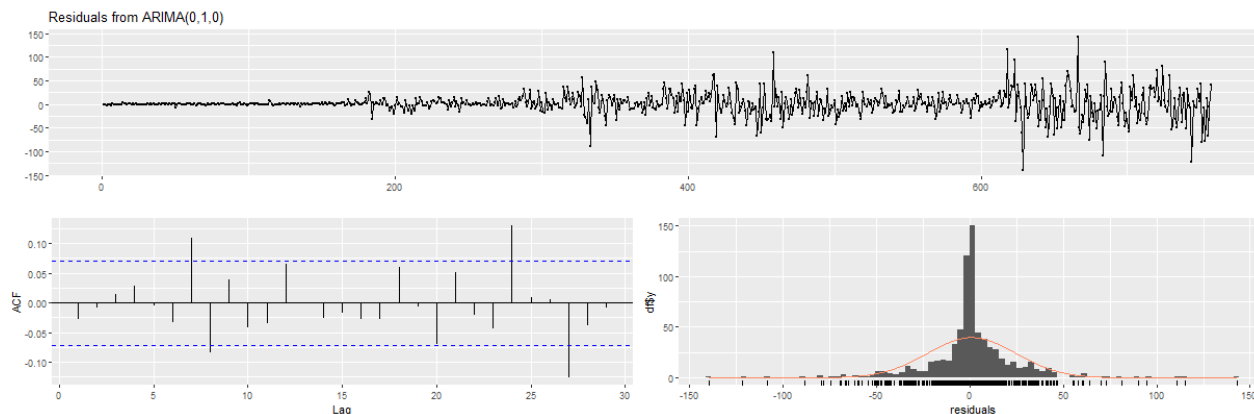


Рисунок 3.6 – залишки моделі

Для розрахунку похибок прогнозу використовується функція `accuracy()` із бібліотеки `Metrics`.

3.2.4 Побудова моделі GMDH

Реалізація МГВА була здійснена за допомогою пакету `GMDH`, що дозволяє проводити моделювання досить легко та швидко [21]. Використано функцію `fcast()` (`out = fcast(ts(filledData$y), weight = 0.9, method = "GMDH", f.number = number, layer = 1)`). Так як МГВА є методом на основі нейронних мереж, то можна вибрати кількість шарів мережі. Недоліком цієї функції є те, що, так як МГВА призначена для короткострокових прогнозів, не дозволяється визначати кількість зпрогнозованих інтервалів (днів) більше п'яти. Функція дозволяє вивести зпрогнозовані значення в окрему змінну (рис.3.7).

Predicted values

```
[1] 865.4319 865.3846 865.0972 864.8074 864.5224
```

Error values

```
      [,1] [,2]
[1,] "mse" "0.017071"
[2,] "mape" "0.018520072482309"
[3,] "mae"  "0.125246185117356"
```

Рисунок 3.7 – результати моделювання за МГВА

Підрахунок похибок (MSE, MAPE, MAE) відбувається за допомогою стандартних функцій R.

3.2.5 Прогнозування за допомогою Prophet

В основі процедури Prophet – адитивна регресійна модель з чотирма основними компонентами:

- 1) логістичний або кусково-лінійний тренд кривої зростання. Prophet автоматично виявляє зміни трендів, вибираючи з даних точки зміни;
- 2) річний сезонний компонент, що моделюється з використанням ряду Фур'є;
- 3) щотижневий сезонний компонент з використанням фіктивних змінних;
- 4) список важливих свят, що надається користувачем.

У розробленій ІС моделювання відбувається за допомогою функції prophet(), де спочатку вводиться значення кількості точок зламу (що може вплинути на точність), змінюється сезонність. Потім додається набір даних, і за допомогою функцій make_future_dataframe() (створення таблиці з датами) та predict() відбувається побудова прогнозової моделі. Результат моделювання зображено на рис. 3.8.

Об'єкт функції predict() – це звичайна таблиця, в якій зберігаються значення декількох розрахованих на основі моделі величин, включаючи компоненти моделі, передбачені значення відгуку, а також верхні і нижні межі довірчих інтервалів відповідних величин.

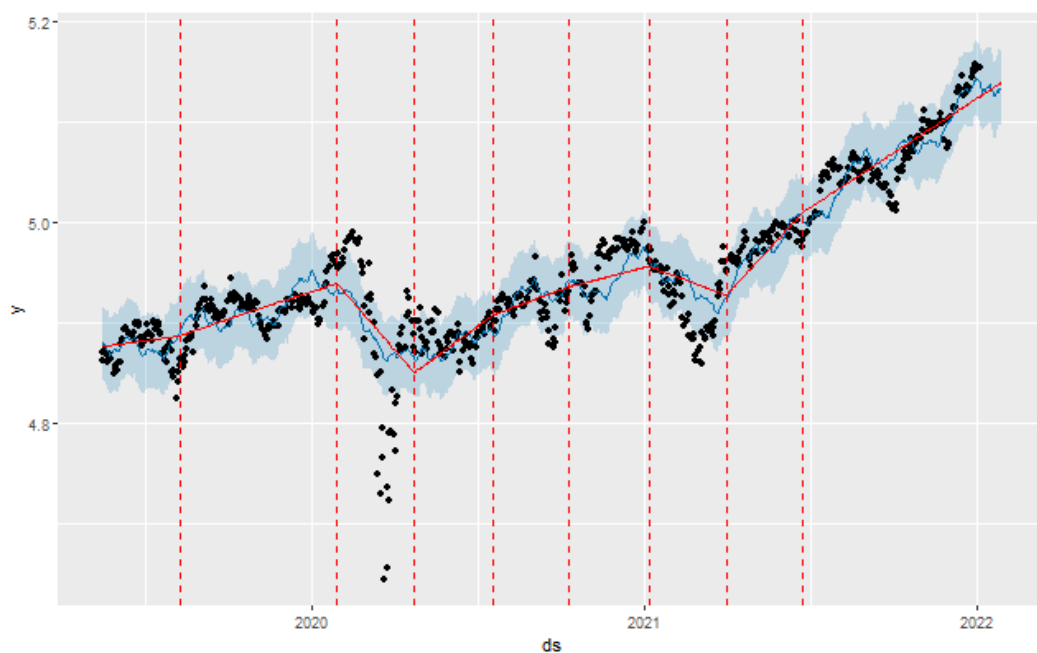


Рисунок 3.8 – графік зпрогнозованих значень із точками зламу

Розкладання ряду на тижневу, річну сезонність та трендову компоненту здійснюється за допомогою функції `prophet_plot_components(M1, forecastM1)`, що зображено на рис. 3.9. Тип сезонності – мультиплікативний.

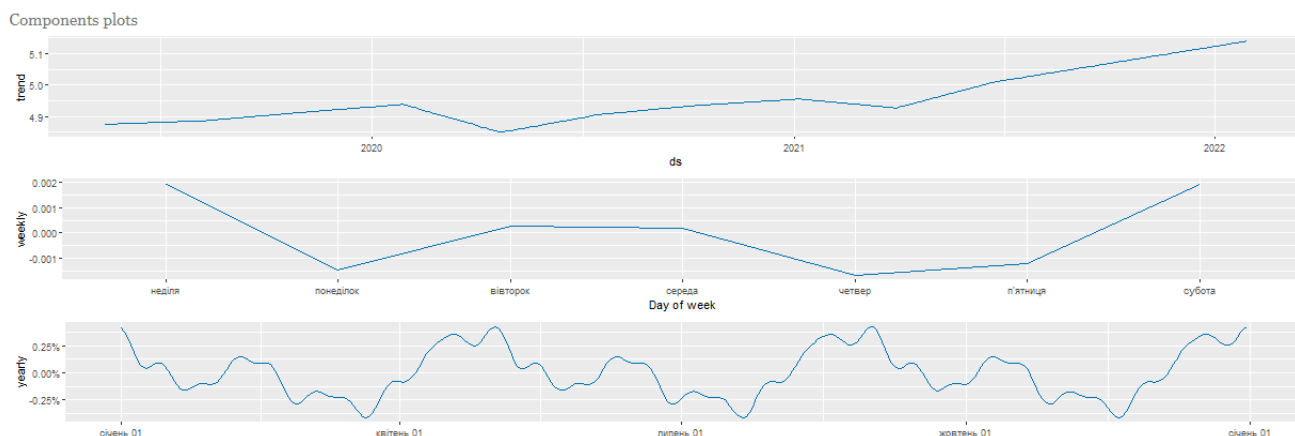


Рисунок 3.9 – компоненти часового ряду

3.3 Аналіз отриманих результатів

Було зпрогнозовано вартість акцій компанії PepsiCo на 5 днів трьома раніше переліченими способами. Результати та похибки прогнозування представлено у таблиці 3.1.

Таблиця 3.1 – результати прогнозування та їх похибки

Модель та значення	ARIMA	GMDH	Prophet
Похибка	172.00, 172.5, 172.5, 172.5, 172.5	170.52, 170.61, 170.70, 170.80, 170.9	171.08, 170.43, 170.30, 170.90, 171.03
MSE	3.836960	0.000069	0.008285
MAPE	0.911393	0.001207	0.045116
MAE	1.280364	0.006204	0.009151

Можна побачити, що модель методу групового урахування аргументів виявилася більш точною за всіма показниками, модель Prophet ліпшою за АРІКС, але гіршою за МГВА. Точність моделі Prophet може суттєво змінитися залежно від кількості точок зламу, також швидкість її розрахунків досить низька, наприклад, вищенаведений розрахунок проводився близько двох хвилин.

Модель АРІКС виявилася найгіршою, але так сталося тому, що значення порядків p (АР), d (інтегроване), q (КС) були підібрані неправильно. Функція `auto.arima()` визначає ці порядки автоматично, але дозволяє їх змінювати вручну. Можна зробити висновок, що інформаційна система має можливість розширення своїх функцій у вигляді дозволу кастомізації цих параметрів користувачем.

3.4 Керівництво користувача

3.4.1 Умови виконання програми

Мінімальні системні вимоги, необхідні для роботи системи:

- Операційна система: Windows 7 / 8 / 10.
- Процесор: 1,8 ГГц.
- Відеоадаптер: 2 ГБ з дозволом на 1280x1024.
- Оперативна пам'ять: 2 ГБ ОЗУ.

3.4.2 Виконання програми

Застосунок знаходиться у публічному доступі за посиланням alexapetrochenko.shinyapps.io/shareforecasttest і не потребує встановленої RStudio,

лише Інтернет зв'язок. Shinyapps.io – це платформа, яка дозволяє легко ділитися своїми застосунками [23].

Головна сторінка програми зображена на рисунку 3.10.



Рисунок 3.10 – головна сторінка програми Shares Forecast

Спочатку потрібно завантажити файл формату .csv за допомогою форм зліва та обрати стовпчик із даними, що потрібно спрогнозувати (рис. 3.11). Відразу після завантаження файлу справа з'явиться таблиця із даними для більш зручного вибору (рис. 3.12). Можна змінювати кількість рядків, що відображаються та шукати конкретні значення.

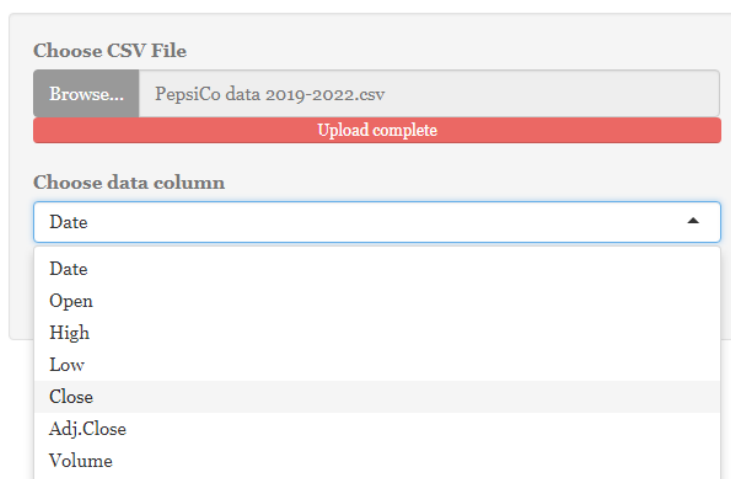


Рисунок 3.11 – приклад завантаження файлу та вибору стовпця

Shares Forecast

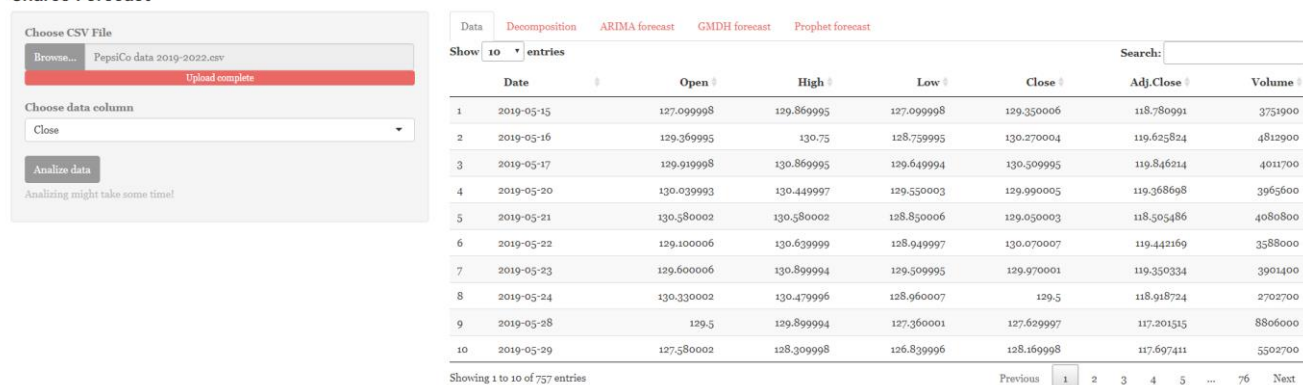


Рисунок 3.12 – показ даних

Застосунок складається із п'яти вкладок:

- Показ даних.
- Декомпозиція.
- Прогнозування за допомогою АРІКС моделі.
- Прогнозування за допомогою МГВА моделі.
- Прогнозування за допомогою пакету Prophet.

Після натискання кнопки «Analyze data» у вкладці декомпозиції з'являться графік обраного набору даних, графік декомпозиції на складові тренду, циклічної, сезонної та нерегулярної компонент, а також результати 3-х тестів на стаціонарність (рис. 3.13).

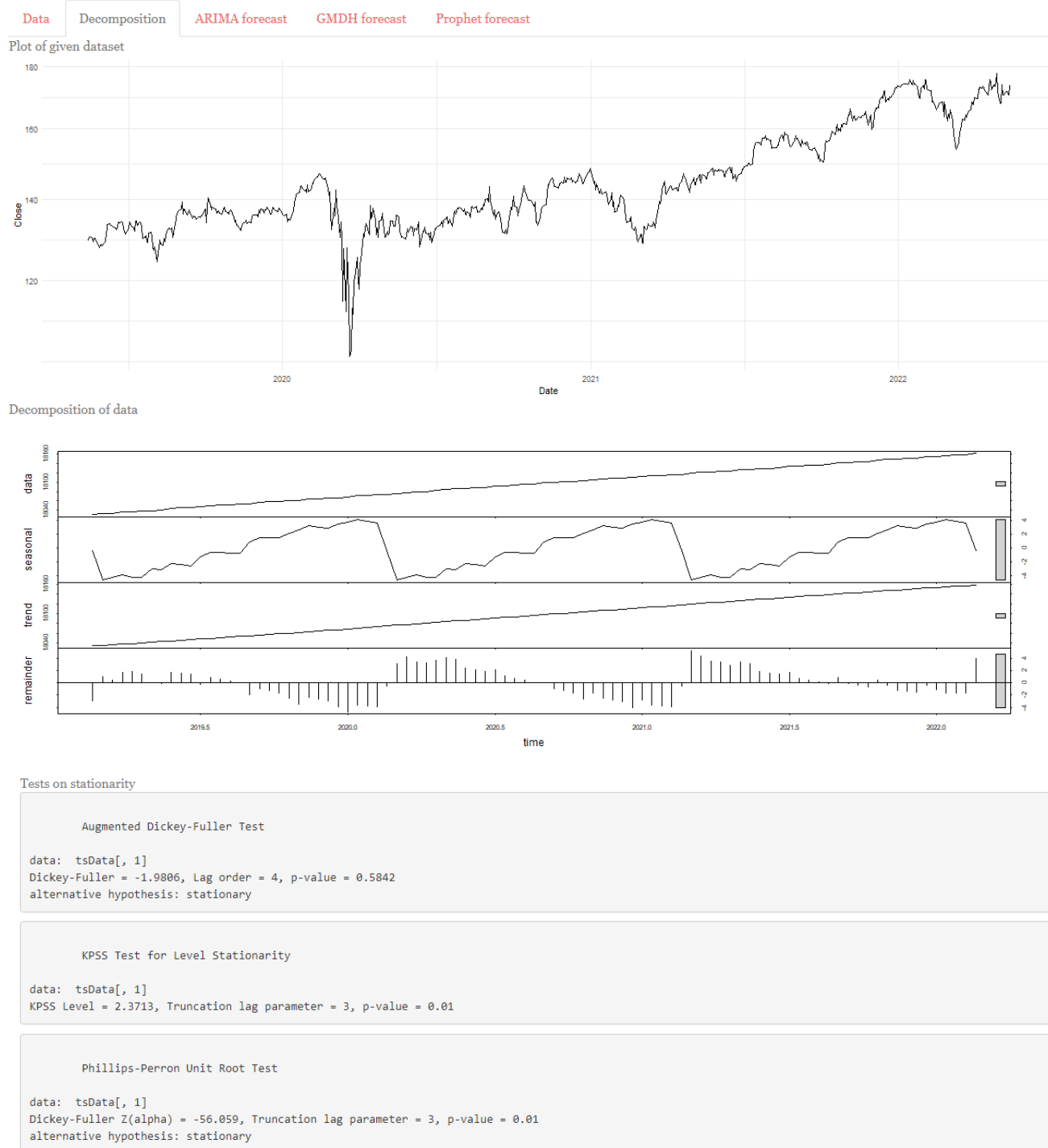


Рисунок 3.13 – вкладка декомпозиції часового ряду

У наступній вкладці доступне прогнозування за допомогою АРІКС моделі. Спочатку потрібно налаштувати кількість днів прогнозування, потім натиснути «Forecast». Як результат, отримується графік дійсних та зпрогнозованих значень, графіки залишків та похибки прогнозування (рис. 3.14).

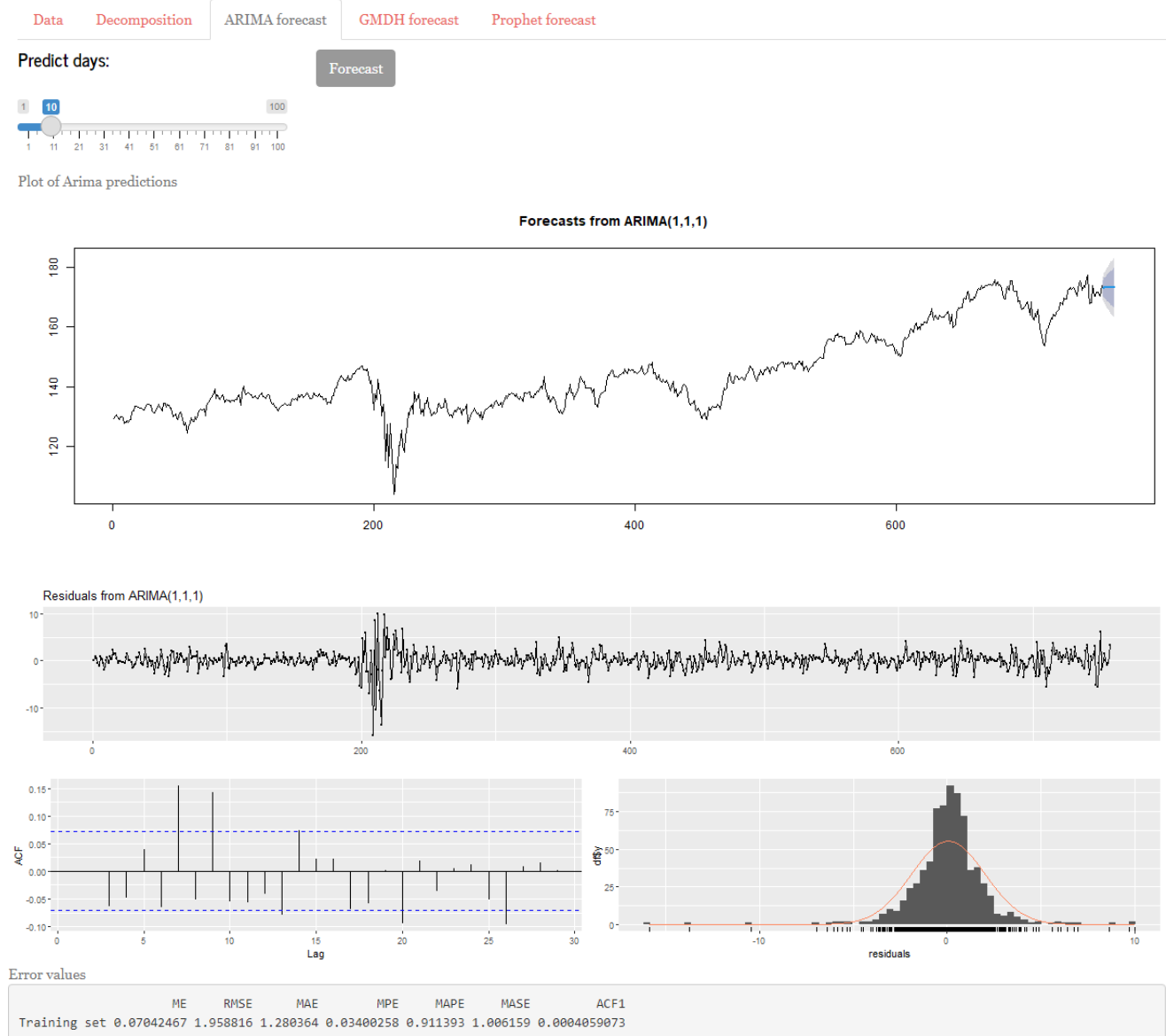


Рисунок 3.14 – вкладка побудови ARIMA-моделі

У вкладці прогнозування за допомогою МГВА можна налаштувати кількість днів, що потрібно спрогнозувати, після натискання кнопки «Forecast» з'являться поля зпрогнозованих значень та похибок прогнозування MSE, MAPE, MAE (рис. 3.15).

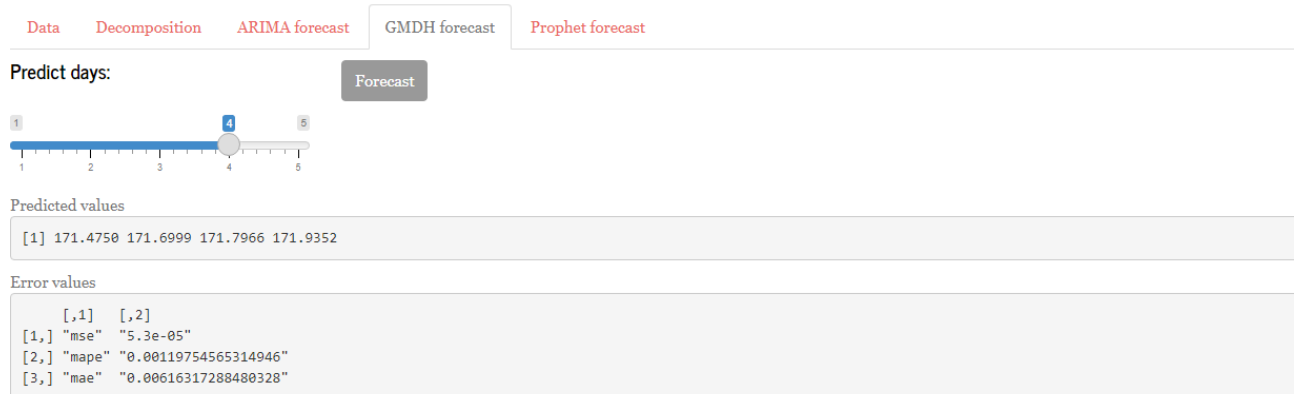


Рисунок 3.15 – вікно прогнозування за МГВА

І остання вкладка – прогнозування за допомогою пакету Prophet. Є можливість вибору кількості точок зламу та днів, що потрібно зпрогнозувати. Натисніть кнопку «Forecast» для початку побудови прогнозної моделі, це може зайняти деякий час.

You can change some parameters

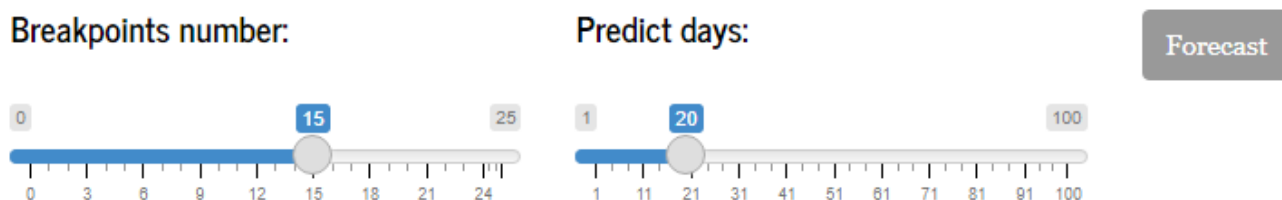
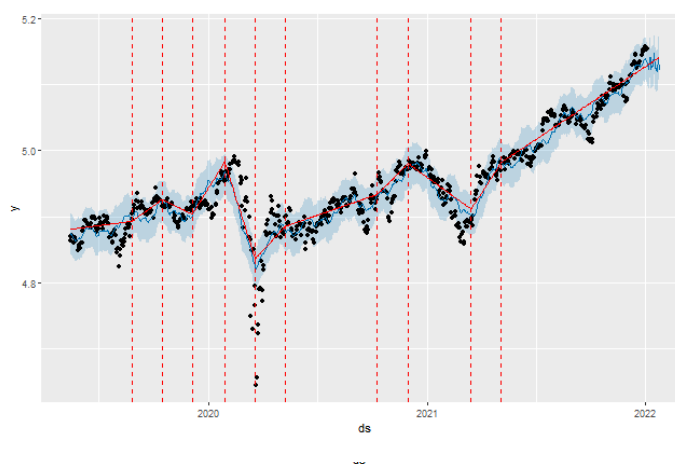


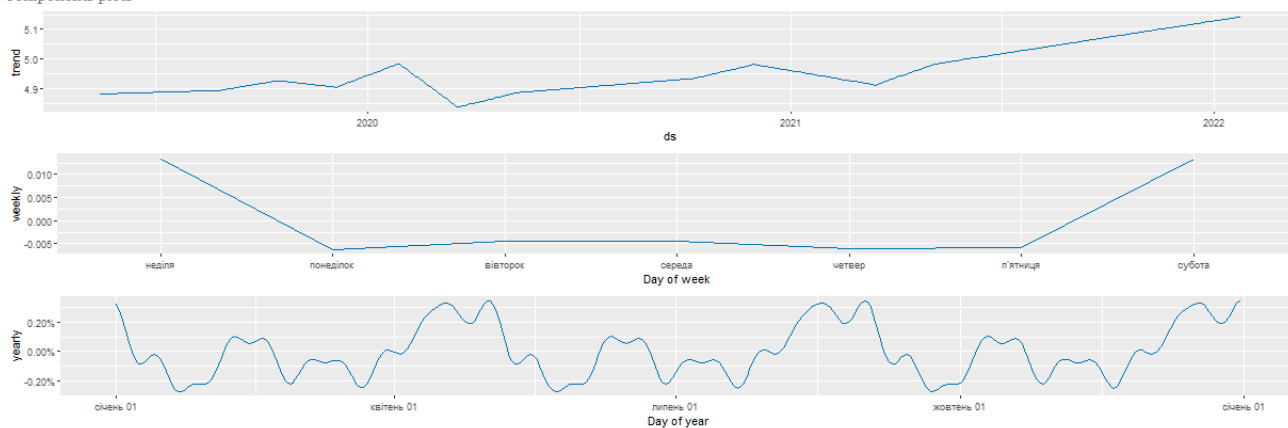
Рисунок 3.16 – зміна параметрів моделі

Після завершення прогнозування вам стануть доступними графіки дійсних та прогнозованих значень, графіки компонент моделі, список зпрогнозованих значень та значення похибок обчислення (рис. 3.17).

Prediction plot



Components plots



Predicted Values

```
[1] 169.0522 168.2672 167.9347 171.0112 171.0265 167.8728 168.4673 168.7445 168.6945 168.8285 172.0487 171.8458 168.2033 168.1979
[15] 167.9004 167.4413 167.4192 170.7484 170.9129 167.8076
```

Error values

horizon	mse	mae	mape
1 20 days	0.002745095	0.04111715	0.008323165

Рисунок 3.17 – вікно зпрогнозованих значень за допомогою пакету Prophet

Помилкою, з якою може зтовкнутися користувач, може бути вибір неправильного стовпця даних для прогнозу, наприклад, стовпчик із датами (рис. 3.18). Оберіть інший стовпчик та натисніть «Analyze data».



Рисунок 3.18 – помилка вибору неправильного стовпчика

Також не рекомендується натискати послідовно кнопки початку прогнозів – тільки після закінчення виконання минулого.

Висновки до розділу 3

Для прогнозування вартості акцій та інших їх показників рекомендовано брати їх вартість на момент закриття біржі, без урахування корпоративних дій або з ними.

Було проведено аналіз вартості акцій компанії PepsiCo за 4 роки, у результаті розроблено інформаційну систему з моделювання та прогнозування вартості акцій на основі моделей:

- моделі авторегресії інтегрованого ковзного середнього;
- моделі на основі МГВА;
- моделі на основі пакету Prophet.

У випадку прогнозування за допомогою МГВА та Prophet дані попередньо логарифмуються.

Моделі МГВА та Prophet мають достатньо якісні оцінки прогнозів, але модель ARIMA може налаштована для більш точних прогнозів. В цілому, моделі можна використовувати для прогнозування фінансових або економічних показників.

Спеціальний розділ

ОХОРОНА ПРАЦІ

до кваліфікаційної роботи

на тему:

«ІНФОРМАЦІЙНА СИСТЕМА ПРОГНОЗУВАННЯ ФІНАНСОВИХ ПОКАЗНИКІВ ПІДПРИЄМСТВА»

Спеціальність 122 «Комп'ютерні науки»

122 – БКР – 401.1810318

Виконав студент 4-го курсу, групи 401

Петроченко О. О.

(підпис, ініціали та прізвище)

«__» _____ 2022 р.

Консультант _____ *ст. викладач* _____

(наук. ступінь, вчене звання)

Макарова О. В.

(підпис, ініціали та прізвище)

«__» _____ 2022 р.

Миколаїв – 2022

4 ОХОРОНА ПРАЦІ

Забезпечення комфортних і безпечних умов праці є важливим обов'язком роботодавців. Середовище, в якому працює людина, безпосередньо впливає на її здоров'я, самопочуття, а отже, на працездатність і продуктивність. Система управління охороною праці сприяє виконанню вимог щодо зниження допустимого рівня шкідливих і шкодливих факторів у виробничому середовищі працівників, у тому числі і в ІТ-сфері [26].

Основною метою розділу є створення безпечних і здорових умов праці на робочому місці або у виробничому приміщенні. **Завдання:** оцінити та розрахувати достатність природного освітлення в офісі та розробити план евакуації робітників.

У першій частині розділу оцінюються умови праці співробітників ТОВ «StatTech», проводяться розрахунки тесту природного освітлення в окремих виробничих приміщеннях, а також надаються рекомендації щодо покращення для керівництва компанії.

У другій частині розділу розглядаються заходи щодо забезпечення безпеки працівників у надзвичайних ситуаціях, пов'язаних із порушенням вимог пожежної безпеки.

4.1 Оцінка умов праці іт-фахівців у приміщенні ТОВ «StatTech»

4.1.1 Опис виробничого приміщення, робочих місць, обладнання та умов праці

Робоча кімната компанії (ТОВ «StatTech») знаходиться на 3-му поверсі 3-поверхового будинку, що розташований за адресою м. Миколаїв, пр. Центральний 122. Загальний план виробничого приміщення зображено на рис. 4.1.

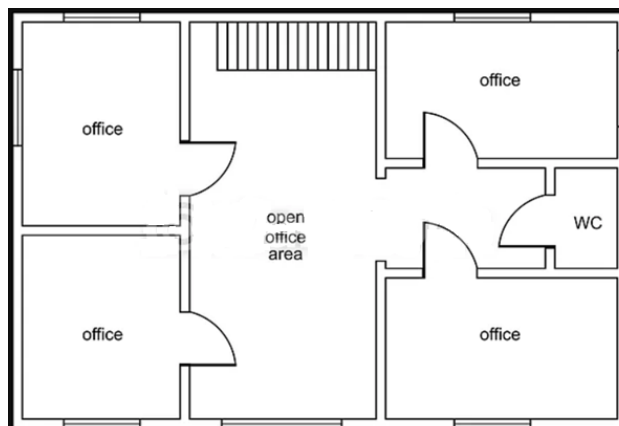


Рисунок 4.1 – План робочого приміщення компанії «StatTech»

Довжина кімнати – 8.5 м, ширина – 7 м. Стіни кімнати гладкі, кремово-білі, висотою 3.5 м, товщиною 0.2 м. Стеля – свіжопобілена. Підлога вкрита лінолеумом коричневого кольору. Природне освітлення у офісному приміщенні бокове, одностороннє, 3 вікна розміром $s \times d = 1,8 \times 2,4$ м. Вікна металопластикові потрійні склопакети, мають рами з вертикальними регульованими жалюзі.

Робоча кімната зображена на рис. 4.2.

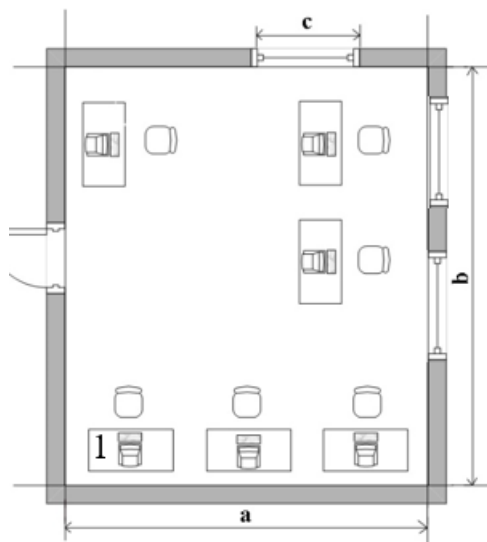


Рисунок 4.2 – План робочої кімнати та розташування робочих місць

У відділі працює 6 осіб, відповідно, знаходиться 6 комп'ютерних столів. Столи комп'ютерні, довжиною 1,5 м, шириною 0,6 м та висотою 0,75 м коричневого кольору. Кожне робоче місце працівника обладнано: столом, стільцем, монітором, системним блоком, клавіатурою, мишею.

За спостереженнями зазначається, що на робочому місці номер 1 протягом майже всього робочого дня задіяно штучне освітлення. Тому є доцільним виконати перевірочний розрахунок природного освітлення у приміщенні із зазначеним робочим місцем.

Геометрію досліджуваного приміщення, розташування робочого місця, протилежної будівлі наведено на рис. 4.3–4.4.

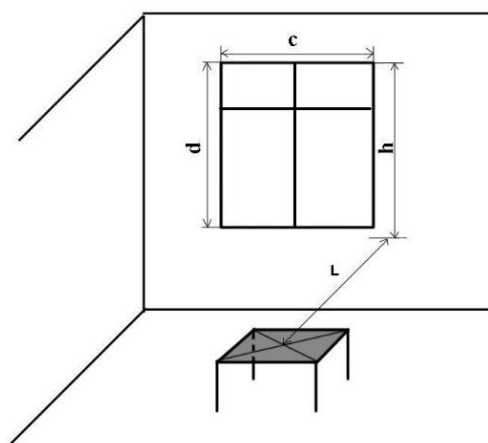


Рисунок 4.3 – Розташування еталонної робочої поверхні

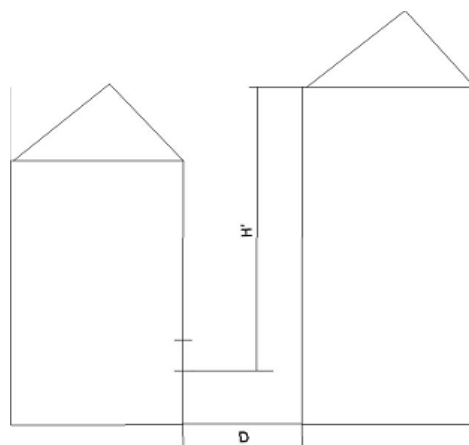


Рисунок 4.4 – Розташування протилежної будівлі

У таблиці нижче наведено чисельні значення розмірів, представлених на рис. 4.3–4.4 та інші вихідні дані, які є необхідними для виконання перевірочного розрахунку природного освітлення в обраному виробничому приміщенні.

Таблиця 4.1 – Вихідні дані для розрахунку природного освітлення у виробничому приміщенні

№	Параметри	Позначення	Розмірність	Значення
1	Довжина приміщення	a	м	8.5
2	Ширина приміщення	b	м	7.0
3	Висота приміщення	H	м	3.5
4	Ширина вікна	c	м	1.8
5	Висота вікна	d	м	2.4
6	Кількість вікон	n_e	шт.	3
7	Висота верхнього краю вікна відносно умовної робочої поверхні	h	м	2.5
8	Відстань розрахункової точки (робочої поверхні) до зовнішньої стіни	l	м	5.2
9	Висота карнізу протилежної будівлі відносно підвіконня	H'	м	7.5
10	Відстань до затіняючої будівлі	D	м	42.0

Працівники офісу увесь день працюють при штучному освітленні. Відповідно до цього є необхідність оцінити, чи достатньо природного освітлення у приміщенні у денний час робочого дня.

4.2.2 Оцінка природнього освітлення у виробничому приміщенні

Природне освітлення – освітлення приміщень денним світлом, що потрапляє через вікна в зовнішніх захищаючих конструкціях приміщення. Природне

освітлення характеризується тим, що змінюється в широких межах залежно від часу дня, пори року і ряду інших чинників [27].

Перевірочний розрахунок виконується в такій послідовності:

1) *Нормативне значення коефіцієнта природного освітлення для III поясу світлового клімату e_N^{III} , %.*

Визначається відповідно до ДБН В.2.5-28-2006 [28, табл. 1]. Для зорових робіт середньої точності при найменшому розмірі об'єкта розпізнавання 0.5-1 мм при боковому освітленні (так званий IV розряд зорової роботи):

$$e_n = 1.5 \%$$

2) *Коефіцієнт світлового клімату m .*

Для Миколаївської області при орієнтації світлових прорізів у зовнішніх стінах будівель на схід відповідно до рекомендацій [28]:

$$m = 0.9.$$

4) *Коефіцієнт сонячності клімату c .*

Для світлових отворів (вікон) в зовнішніх стінах будівель, розташованих у IV поясі світлового клімату та зорієнтованих по азимуту в діапазоні 136... 225 градусів відповідно до рекомендацій [28, табл. 3]:

$$c = 3.1.$$

5) *Нормоване значення коефіцієнта природного освітлення для розрахункових умов e_n , %:*

$$e_N = e_n \cdot m;$$

$$e_N = 0,015 \cdot 0,85 \cdot 100\% = 1.275\%.$$

6) *Коефіцієнт запасу, що використовується при розрахунку природного освітлення κ_3 .*

Відповідно до рекомендацій [28]: $\kappa_3 = 1.3...1.5$. Прийнято $\kappa_3 = 1.3$.

7) Геометричні співвідношення, що характеризують виробниче приміщення та розташування робочого місця в ньому: a/b , b/h , l/b .

$$a/b = 8,5 / 7 = 1.21;$$

$$b/h = 7 / 2,5 = 2.8;$$

$$l/b = 5,2 / 7 = 0.74.$$

8) Світлова характеристика вікна η_v .

Визначається відповідно до рекомендації [28, табл. 4] $\eta_v = f(a/b, b/h)$. Враховуючи, що $\eta_e = f(a/b, b/h)$, в ряді випадків слід виконати двомірну лінійну інтерполяцію.

$y \backslash x$	2	2.8	3
1.5	13		15
1.21		?	
1	16		18

$$\eta_e = f(x, y).$$

$$f(x, y_1) = f(x_1, y_1) + (x - x_1) \frac{f(x_2, y_1) - f(x_1, y_1)}{x_2 - x_1} = 13 + (2.8 - 2) \frac{15 - 13}{3 - 2} = 14.6$$

$$f(x, y_2) = f(x_1, y_2) + (x - x_1) \frac{f(x_2, y_2) - f(x_1, y_2)}{x_2 - x_1} = 16 + (2.8 - 2) \frac{18 - 16}{3 - 2} = 17.6$$

$$f(x, y) = f(x, y_1) + (y - y_1) \frac{f(x, y_2) - f(x, y_1)}{y_2 - y_1} = 14.6 + (1.21 - 1.5) \frac{17.6 - 14.6}{1 - 1.5} = 16.3$$

Отже, $\eta_e = 16.3$.

9) Коефіцієнт світлопропускання матеріалу τ_1 .

Визначається згідно рекомендацій [28, табл. 6] відповідно до потрібного склопакета металопластикових вікон: $\tau_1 = 0.75$.

10) Коефіцієнт, що враховує втрати світла у віконній рамі τ_2 .

Визначається згідно рекомендацій [28, табл. 6]. Для рам металопластикових вікон: $\tau_2 = 0.76$.

11) Коефіцієнт, що враховує втрати світла у несучих конструкціях τ_3 .

При боковому освітленні згідно рекомендацій [28]: $\tau_3 = 1$.

12) Коефіцієнт, що визначає втрату світла у сонцезахисних конструкціях τ_4 .

Для внутрішніх регульованих жалюзі та штор з рекомендації [28]: $\tau_4=1$.

13) Загальний коефіцієнт світлопропускання $\tau_{заг}$

$$\tau_{заг} = \tau_1 \cdot \tau_2 \cdot \tau_3 \cdot \tau_4 = 0,75 \cdot 0,78 \cdot 1 \cdot 1 = 0,585.$$

14) Розрахункове значення середньозваженого коефіцієнта відбивання внутрішніх поверхонь виробничого приміщення $\rho_{сер}$.

Відповідно до рекомендацій [28] $\rho_{сер}$ слід приймати 0.3 для виробничих приміщень.

$$\rho_{сер} = 0.3.$$

15) Площа підлоги виробничого приміщення $S_{підл}$, m^2 .

$$S_{підл} = a \cdot b;$$

$$S_{підл} = 8.5 \cdot 7 = 59.5 \text{ м}^2.$$

16) Коефіцієнт, що враховує підвищення коефіцієнта природного освітлення за рахунок світла, яке відбивається від внутрішніх поверхонь приміщення r_1 .

Згідно з даними [28, табл. Л. 5] $r_1 = f(\rho_{сер}, a/b, b/h, l/b)$.

$y \backslash x$	1	1.21	2
0.7	1.5		1.3
0.74		?	
0.9	1.85		1.5

$$r_1 = f(x, y).$$

$$f(x, y_1) = f(x_1, y_1) + (x - x_1) \frac{f(x_2, y_1) - f(x_1, y_1)}{x_2 - x_1} = 1.5 + (1.21 - 1) \frac{1.3 - 1.5}{2 - 1} = 1.46$$

$$f(x, y_2) = f(x_1, y_2) + (x - x_1) \frac{f(x_2, y_2) - f(x_1, y_2)}{x_2 - x_1} = 1.85 + (1.21 - 1) \frac{1.5 - 1.85}{2 - 1} = 1.78$$

$$f(x, y) = f(x, y_1) + (y - y_1) \frac{f(x, y_2) - f(x, y_1)}{y_2 - y_1} = 1.46 + (0.74 - 0.7) \frac{1.78 - 1.46}{0.9 - 0.7} = 1.52$$

Отже, $r_1 = 1.52$.

17) Відношення відстані проміж протилежними будівлями до висоти карнизу протилежного будинку над підвіконням D/H^I .

$$D/H^I = 42 / 7.5 = 5.6.$$

18) Коефіцієнт, що враховує вплив протилежної будівлі на освітленість у виробничому приміщенні $K_{б\gamma\delta}$.

Визначається згідно [28, табл. Л. 2] як $K_{б\gamma\delta} = f(D/H^I)$.

$$K_{б\gamma\delta} = 1.0.$$

19) Площа вікон, що необхідна для забезпечення нормованої природної освітленості у виробничому приміщенні S_{ϵ} , м².

$$S_{\epsilon} = \frac{e_N \cdot k_3 \cdot \eta_{\epsilon} \cdot S_{\text{нідл}} \cdot k_{б\gamma\delta}}{\tau_{\text{заг}} \cdot r_1 \cdot 100}.$$

$$S_{\epsilon} = \frac{1.275 \cdot 1.2 \cdot 16.3 \cdot 59.5 \cdot 1}{0.585 \cdot 1.52 \cdot 100} = \frac{1483.8}{67.86} = 16.68.$$

20) Розрахуємо дійсну площу вікон виробничого приміщення.

$$S_{\text{д}} = c \cdot d \cdot n_{\text{в}};$$

$$S_{\text{д}} = 1.8 \cdot 2.4 \cdot 3 = 12.96.$$

Були розраховані необхідна ($S_{\text{в}}$) та дійсна ($S_{\text{д}}$) площа вікон для виробничого приміщення. Дійсна площа є меншою за необхідну ($12.96 < 16.68$), а отже природного освітлення, яке відображається нормованим значенням пр. осв. e_N , недостатньо для виробничої діяльності у розглянутому приміщенні відповідно до санітарно-гігієнічних вимог. Крім того, для покращення ситуації з освітленням можна освітлити підлогу, стіни та стелі, що збільшить їх коефіцієнт відбиття [29].

4.2 Порядок дій персоналу «StatTech» у надзвичайній ситуації пов'язаній з пожежою. Шляхи евакуації людей

4.2.1 Порядок дій персоналу ТОВ «StatTech» у випадку пожежі

Надалі наведено рекомендації щодо порядку дій персоналу фірми у випадку пожежі.

У разі виявлення ознак горіння кожний співробітник зобов'язаний [30]:

- варто негайно повідомити про це телефоном пожежну охорону. Необхідно назвати адресу об'єкта, вказати кількість поверхів будівлі, місце виникнення пожежі, обстановку на пожежі, наявність людей, а також повідомити своє прізвище;

- по можливості вжити заходів до евакуації людей, гасіння пожежі та збереження матеріальних цінностей;

- якщо пожежа виникла на підприємстві, слід повідомити про неї керівника чи відповідну компетентну посадову особу та/або чергового по об'єкту;

- у разі необхідності викликати інші аварійно-рятувальні служби.

Посадова особа об'єкта, що прибула на місце пожежі, зобов'язана:

- перевірити, чи викликана пожежна охорона, довести подію до відома власника підприємства;

- у разі загрози життю людей негайно організувати їх евакуацію, використовуючи для цього наявні сили й засоби;

- видалити за межі небезпечної зони всіх працюючих, не пов'язаних з ліквідацією пожежі;

- припинити роботи в будівлі, крім робіт, пов'язаних із заходами по ліквідації пожежі;

- здійснити в разі необхідності вимкнення електроенергії (за винятком систем протипожежного захисту), зупинення транспортуючих пристроїв, агрегатів;

- слід перевірити ввімкнення оповіщення людей про пожежу, установок пожежогасіння, протидимового захисту;

- по можливості організувати зустріч підрозділів пожежної охорони, подати їм допомогу у виборі найкоротшого шляху для під'їзду до осередку пожежі та під'єднання до джерел води;

- одночасно з гасінням пожежі організувати евакуацію і захист матеріальних цінностей;
- забезпечити дотримання техніки безпеки працівниками, які беруть участь у гасінні пожежі.

4.2.1 План евакуації підприємства ТОВ «StatTech»

У разі виникнення пожежі на початковій стадії виділяється тепло, токсичні продукти згорання, можливі обвалення конструкцій. Тому слід врахувати необхідність евакуації людей у короткий строк.

Виходи вважаються евакуаційними, якщо вони ведуть [30]:

- з приміщень першого поверху назовні безпосередньо або через коридор, вестибюль, сходову клітку;
- з приміщень будь-якого поверху, крім першого, в коридори, що ведуть на сходову клітку (в тому числі через хол); при цьому сходові клітки повинні мати вихід назовні безпосередньо або через вестибюль, відділений від прилеглих коридорів перегородками з дверима;
- з приміщень в сусіднє приміщення на цьому ж поверсі, забезпечене виходами, вказаними вище.

Як правило, план евакуації вивішується на видних місцях будівлі та приміщення (в коридорах та проходах, біля виходів з приміщення). План евакуації підприємства «StatTech» зображено на рис. 4.5.

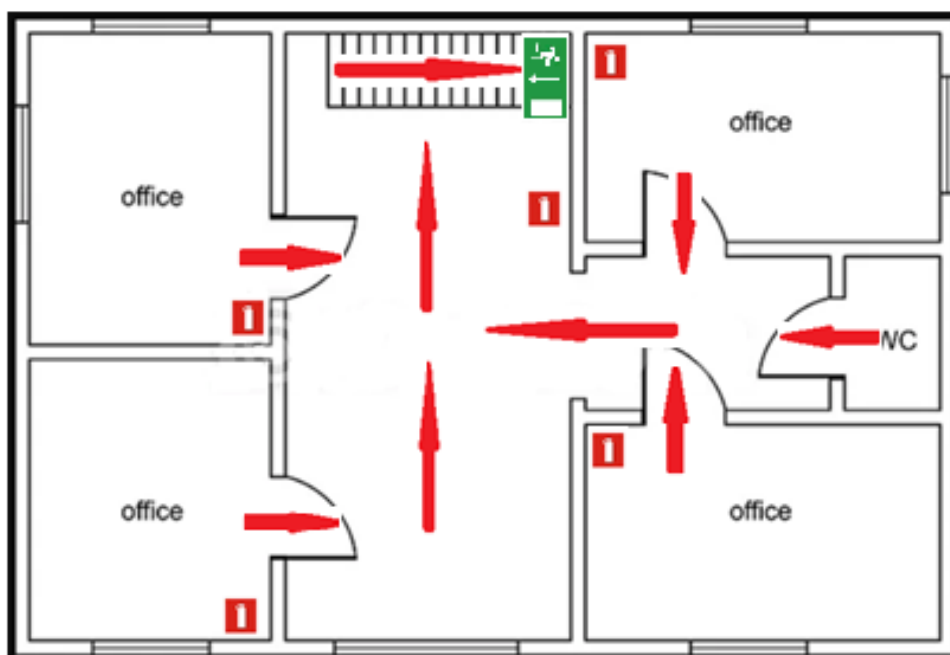




Рисунок 4.5 – План евакуації підприємства ТОВ “StatTech”

План евакуації використовує позначки, що зазначені у таблиці 4.1.

Таблиця 4.1 – Позначки плану евакуації

Позначення	Опис
1	Місце знаходження вогнегасника
	Шлях до евакуаційних виходів
	Евакуаційний вихід

Кількість працівників офісу складає до 12 робітників, при цьому підприємство має 1 вихід на випадок евакуації. Загальна кількість вогнегасників рівна 5, що є допустимим для розмірів даного приміщення.

Двері евакуаційних виходів відкриваються доволі легко з приміщення. Виходи нічим не заблоковані, що дає змогу у разі евакуації вільно пересуватися. Вікна виробничого приміщення не мають глухих ґрат. У кожному приміщенні на видних місцях плани евакуації, встановлені таблички із зазначенням порядку виклику пожежної охорони, знаки місць розміщення первинних засобів пожежогасіння.

Висновки до розділу охорони праці

У першому розділі даної роботи аналізуються основні фактори виробничого середовища та трудового процесу на виробництві ТОВ «StatTech». Ці фактори були визначені для відповідності санітарно-гігієнічним вимогам. Виняток становить природне освітлення. Запропоновано відповідний захід для покращення: замінити внутрішню поверхню на таку, що має найбільш можливий коефіцієнт відбиття.

Друга частина присвячена аналізу небезпеки для людини в умовах пожежонебезпечних ситуацій. Одним з основних способів захисту є своєчасне і швидке виведення або евакуація людей із небезпечної зони, тобто евакуація. Розглянуто правила евакуації з приміщень і будівель, вимоги до будівель і споруд в контексті евакуації, вимоги до шляхів і виходів евакуації. Було доведено, що приміщення ТОВ «StatTech» відповідають нормам та вимогам щодо протипожежних заходів.

ВИСНОВКИ

В ході виконання даної роботи було розглянуто фондовий ринок та його значення в економіці та тему часових рядів як одного із способів представлення вартості акцій.

Розглянуто існуючі публікації та наукові матеріали на тему прогнозування вартості акцій. Описано існуючі аналоги системи.

У другому розділі описано існуючі методи прогнозування часових рядів, їх властивості та послідовність виконання. Також розглянуто існуючі технології, за допомогою яких можна аналізувати та прогнозувати часові ряди, особливу увагу приділено мові R.

Опис реалізованих функцій ІС виконано у третьому розділі, написано керівництво користувача.

Мета цієї роботи досягнута, було проаналізовано декілька методів прогнозування часових рядів та розроблено відповідну інформаційну систему.

ІС відповідає усім поставленим вимогам, функціонал застосунку є простим та зрозумілим, визначено його слабкі місця. Також можливе подальше удосконалення системи, зокрема підвищення точності прогнозування окремих моделей.

Кожне завдання реалізовано та поставлені цілі досягнуті.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Качаловський А. С. Методи і моделі прогнозування лінійних та нелінійних нестационарних процесів. *International Scientific Journal*. 2016. № 5. URL : <https://www.inter-nauka.com/uploads/public/14655658092305.PDF> (дата звернення: 23.05.2022).
2. Бідюк П. І. Аналіз часових рядів : навч. посіб. / П. І. Бідюк, В. Д. Романенко, О. Л. Тимощук. Київ : Політехніка, 2010. 317 с.
3. Sima Siami-Namini, Akbar Siami Namin. Forecasting Economic and Financial Time Series: ARIMA vs. LSTM, 2018 – 19 p.
4. Daiyou Xiao, Jinxia Su, "Research on Stock Price Time Series Prediction Based on Deep Learning and Autoregressive Integrated Moving Average", *Scientific Programming*, Article ID 4758698, 12 pages, 2022. <https://doi.org/010.1155/2022/4758698> (дата звернення: 24.05.2022).
5. Wong F. S., Wang P. Z., Goh T. H., & Quek B. K. Fuzzy neural systems for stock selection. *Financial Analysts Journal*. 1992. № 4. P. 47–52.
6. Інтелектуальний аналіз даних: методичні вказівки до виконання комп'ютерних практикумів з навчальної дисципліни «Інтелектуальний аналіз даних». Частина-1. «Кореляційний та регресійний аналіз медичних даних». / Уклад.: д.б.н., с.н.с. Є. А. Настенко, к.т.н. В. С. Якимчук, к.т.н. О. К. Носовець. – К.: НТУУ «КПІ ім. І. Сікорського», 2017. – 51 с.
7. Катренко А. В. Системний аналіз об'єктів та процесів комп'ютеризації : навч. посіб. – Львів : «Новий світ-2010», 2011. – 424 с.
8. Андрусенко Ю. О. Аналіз основних моделей прогнозування часових рядів. Збірник наукових праць Харківського національного університету Повітряних Сил. 2020. № 3 (65). С. 91–96.
9. Chris Chatfield (1996). *The Analysis of Time Series, an Introduction*. Ed. 5-th. Chapman & Hall/CRC. p. 33.

10. Трегуб А.В., Трегуб И.В. Методика построения модели Arima для прогнозирования динамики временных рядов. *Лесной вестник*. №5, 2011. С. 179–183.

11. Кратович П. В. Нейронные сети и модели Arima для прогнозирования котировок. *Программные продукты и системы*. №1, 2011, с. 95–98.

12. Дяченко П. В., Максимов А. Є. Побудова моделі прогнозу вартості акцій на основі використання МГВА // Інформаційні технології в освіті, науці і техніці : *V Міжнар. наук.-практ. конф. 21–23 трав. 2020 р., м. Черкаси : тези*. Черкаси : ЧДТУ, 2020. С. 58–60.

13. Прогнозування та аналіз часових рядів : метод. вказівки до практичних занять та самостійної роботи студентів спец. 051 «Економіка» освітньої програми «Економічна кібернетика», «Економічна аналітика» / Укл. : М. Є. Юрченко. Чернігів : ЧНТУ, 2018. 88 с.

14. Документація Matlab. Режим доступу: <https://www.mathworks.com/help/matlab/> (дата звернення: 29.05.2022).

15. Python 3.10.5 documentation. URL: <https://docs.python.org/3/> (дата звернення: 26.05.2022).

16. Мастицкий С. Э. Анализ временных рядов с помощью R : электронная книга. 2020. URL: <https://ranalytics.github.io/tsa-with-r> (дата звернення: 28.05.2022).

17. Джерело фінансової інформації Investopedia. <https://www.investopedia.com/terms/c/closingprice.asp> (дата звернення: 29.05.2022).

18. Yahoo Finance : Web. 2022. ULR : <https://finance.yahoo.com/> (дата звернення: 28.05.2022).

19. Chatfield Chr. The Analysis of Time Series. An Introduction with R [Text] / Chr. Chatfield, H. Xing. – [7th ed.] – Taylor & Francis Group, LLC, 2019. – 414 p.

20. Hyndman R. J., & Khandakar Y. Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, 2008. № 27 (3). P. 1–22. URL: <https://doi.org/10.18637/jss.v027.i03> (дата звернення: 27.05.2022).

21. Dag, O., Yozgatligil, C. GMDH: An R Package for Short Term Forecasting via GMDH-Type Neural Network Algorithms. *The R Journal*, 8:1, 2016, p. 379–386.
22. Shumway R.H. Time Series Analysis and Its Applications. With R examples [Text] / R.H. Shumway, D.S. Stoffer. – [3d ed.] – Springer Science+Business Media, LLC, 2011. – 202 p.
23. Shinyapps.io. URL: <https://docs.rstudio.com/shinyapps.io/> (дата звернення: 28.05.2022).
24. Бідюк П. І., Гожий О. П., Коршевнік Л. О. Комп'ютерні системи підтримки прийняття рішень : навч. посіб. Миколаїв : Вид-во ЧДУ ім. Петра Могили, 2012. 380 с.
25. Томашевський В. М. Моделювання систем. Київ : ВНУ, 2015. 352 с.
26. Методичні вказівки до самостійної роботи з дисципліни «Безпека праці в індустрії інформаційних технологій» для студентів усіх спеціальностей заочної форми навчання / Упоряд: Г.В. Пронюк, Т.Є. Стиценко, Н.М. Сердюк. – Харків: ХНУРЕ, 2017. – 32 с.
27. Практикум із охорони праці. Навчальний посібник / В. Ц. Жидецький, В. С. Джигирей, В. М. Сторожук та ін.; за ред. В. Ц. Жидецького. – Львів: Афіша, 2000. – 352 с.
28. Природне і штучне освітлення: ДБН В.2.5-28-2006.Режим доступу : URL : <http://kbu.org.ua/assets/app/documents/dbn2/95.1.%20ДБН%20В.2.5-28-2006.%20Природне%20і%20штучне%20освітлення.pdf> (дата звернення: 27.05.2022).
29. Фабрика Мебелі. Проектування приміщення [Електронний ресурс]. – Режим доступу : <http://8marta.planningwiz.com>.
30. Желібо Є. П., Заверуха Н. М., Зацарний В. В. Безпека життєдіяльності. – К., 2002. – 162 с.

ДОДАТОК А

Норми штучного та природного освітлення виробничих приміщень (витяг з ДБН В.2.5-28-2006) [28, фрагмент табл. 1]

Характеристика зорової роботи	Найменший або еквівалентний розмір об'єкта розрізнення, мм	Розряд зорової роботи	Підрозряд зорової роботи	Контраст об'єкта з фоном	Характеристика фону	Штучне освітлення					Природне освітлення		Суміщене освітлення			
						Освітленість, лк			сукупність нормованих величин показника осліпленості і коефіцієнта пульсації		КПО, е _н %					
						при системі комбінованого освітлення		при системі загального освітлення			при верхньому або комбінованому освітленні	при боковому освітленні			при верхньому або комбінованому освітленні	при боковому освітленні
						всього	у т.ч. від загального									
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
Середньої точності	Більше 0,5 до 1,0	IV	а	Малий	Темний	750	200	300	40	20	4	1,5	2,4	0,9		
			б	Малий Середній	Середній Темний	500	200	200	40	20						
			в	Малий Середній Великий	Світлий Середній Темний	400	200	200	40	20						
			г	Середній Великий Великий	Світлий Світлий Середній	—	—	200	40	20						
Малої точності	Більше 1,0 до 5	V	а	Малий	Темний	400	200	300	40	20	3	1	1,8	0,6		
			б	Малий Середній	Середній Темний	—	—	200	40	20						
			в	Малий, Середній Великий	Світлий Середній Темний	—	—	200	40	20						
			г	Середній Великий Великий	Світлий Світлий Середній	—	—	200	40	20						

ДОДАТОК Б**Лістинг програми**

```
library(shiny)
library(shinythemes)
library(lubridate)
library(dplyr)
library(DT)
library(tsibble)
library(ggplot2)
library(tseries)
library(prophet)
library(imputeTS)
library(GMDH)
library(Metrics)
library(forecast)
library(rsconnect)
ui <- fluidPage(
  fluidPage(theme = shinytheme("journal"), titlePanel("Shares Forecast"),
    sidebarLayout(
      sidebarPanel(
        fileInput("file1", "Choose CSV File", accept = ".csv"),
        selectInput("selData", label = 'Choose data column', choices = c('Upload a
file', 'Upload a file')),
        actionButton("go", "Analyze data"), helpText("Analizing might take some time!")
      ),
      mainPanel( tabsetPanel(
        tabPanel("Data", DT::dataTableOutput("contents")),
        tabPanel("Decomposition",
        textOutput("dataPlot"), plotOutput("plot1"),
```



```

textOutput("decomp"), plotOutput("plot2"),
textOutput("tests"), verbatimTextOutput("adfOut"),
verbatimTextOutput("kpssOut"), verbatimTextOutput("ppOut")
    ),
    tabPanel("ARIMA forecast",
    fluidRow(
    column(3, h4("Predict days:"), sliderInput("slideIn4", "", min = 1, max =
100, value = 5)),
        column(3, h4(""), actionButton("arimaStart", "Forecast"))),
        textOutput("plotA"), plotOutput("plot5"), plotOutput("plot7"),
verbatimTextOutput("resid"),
textOutput("errotOutA"), verbatimTextOutput("foreArima")
    ),
    tabPanel("GMDH forecast",
    fluidRow(
    column(3, h4("Predict days:"), sliderInput ("slideIn3", "", min = 1, max = 5,
value = 4)),
        column(3, h4(""), actionButton("gmdhStart", "Forecast"))),
    textOutput("gmdhSign"), verbatimTextOutput("gmdhF"),
    textOutput("errorOutT"), verbatimTextOutput("errorsOut3")
    ),
    tabPanel("Prophet forecast",
    textOutput("signProph"),
        fluidRow(
    column(3, h4("Breakpoints number:"), sliderInput("slideIn1", "", min = 0, max =
25, value = 5)),
        column(3, h4("Predict days:"), sliderInput("slideIn2", "", min = 1,
max = 100, value = 4)),
        column(3, h4(""), actionButton("prophetStart", "Forecast"))),

```

```

textOutput("predSign"), plotOutput("plot3"),
textOutput("compon"), plotOutput("plot4"),
textOutput("predV"), verbatimTextOutput("predVal"),
textOutput("errorOut"), verbatimTextOutput("errorsOut1")
) ) ))))

#Server
server <- function(input, output, session){
  #get data
  mydata <- reactive({
    inFile <- input$file1
    if (is.null(inFile)) return(NULL)
    data<-read.csv(inFile$datapath, header = TRUE)
    data })
  #table render
  output$contents<-DT::renderDataTable({DT::datatable(mydata())})
  #get columns
  observeEvent(input$file1, {
    mytable <- read.csv(input$file1$datapath)
    updateSelectInput(session, "selData", choices = colnames(mytable))
  })
  #get time series object
  shareObj <- reactive({
    sharePrices <- mydata()
    str(sharePrices)
    sharePrices $Date=ymd(sharePrices $Date)
    data1 <- input$selData
    return(sharePrices %>% dplyr::select(Date, data1))
  })
#DECOMPOSITION PAGE

```

```

observeEvent(input$go, {
  dataset <- shareObj()
  output$dataPlot <- renderText("Plot of given dataset")
  output$decomp <- renderText("Decomposition of data")
  output$tests <- renderText("Tests on stationarity")
  #valuable define
  as_tsibble(dataset, key=NULL, index=Date)
  headers <- names(dataset)
  #plot output
  output$plot1 <- renderPlot({
    ggplot(data = dataset)+geom_line (aes_string(x = headers[1], y = headers[2])) +
scale_y_log10()+ theme_minimal()
  })
  max_Date <- max(dataset$Date)
  min_Date <- min(dataset$Date)
  timeSeries <- ts(data=dataset, end=c(year(max_Date),
month(max_Date)),start=c(year(min_Date), month(min_Date)), frequency = 30)
  #decomposition
  output$plot2 <- renderPlot({
    stlData <- stl(timeSeries[,1],s.window="periodic")
    plot(stlData)  })
  #tests
  output$adfOut <- renderPrint({
    adf.test(timeSeries[,1])  })
  output$kpssOut <- renderPrint({
    kpss.test(timeSeries[,1])  })
  output$ppOut <- renderPrint({
    pp.test(timeSeries[,1])
  }) })

```

```

#data preparation
prepData <- reactive({
  dataset <- shareObj()
  filledData <- na_mean(dataset, option = "mean")
as_tsibble(filledData, key=NULL, index=Date)
  names(filledData) <- c("ds", "y")
  return(filledData)
})
#PROPHET PAGE
observeEvent(input$go, {
  output$signProph <- renderText("You can change some parameters")
})
observeEvent(input$prophetStart, {
  output$predSign <- renderText("Prediction plot")
  output$compon <- renderText("Components plots")
output$errorOut <- renderText("Error values")
  output$predV <- renderText("Predicted Values")
  filledData <- prepData()
  shareTest <- filledData %>% mutate(ds=as.Date(ds), y=log(y)) %>%
    tail(90) %>% as.data.frame()
  shareTrain <- filledData %>% mutate(y=log(y)) %>%
    slice(1:(n()-90)) %>% as.data.frame()
  M1 <- prophet(n.changepoints = input$slideIn1, yearly.seasonality = FALSE)
  M1 <- add_seasonality(M1, name = 'yearly', period = 365.25/3, fourier.order = 10,
mode = "multiplicative")
  M1 <- fit.prophet(M1, shareTrain)
  futureDf <- make_future_dataframe(M1, periods = input$slideIn2)
  forecastM1 <- predict(M1, futureDf)
  output$plot3 <- renderPlot({

```

```

plot(M1, forecastM1)+add_changepoints_to_plot(M1)
})
output$predVal <- renderPrint({
exp(tail(forecastM1$yhat, input$slideIn2))
})
output$plot4 <- renderPlot({
  prophet_plot_components(M1, forecastM1)
})
M1_cv <- cross_validation(M1, initial = 280, period = input$slideIn2, horizon =
input$slideIn2, units = "days")
M1_perf <- performance_metrics(M1_cv,metrics = c("mse", "mae", "mape"),
rolling_window = 1)
output$errorsOut1 <- renderPrint({
  M1_perf %>% head()
}) })
#ARIMA PAGE
observeEvent(input$arimaStart, {
  output$errotOutA <- renderText("Error values")
  output$plotA <- renderText("Plot of Arima predictions")
  filledData <- prepData()
  shareTrain <- filledData %>% mutate(y=log(y)) %>%
  slice(1:(n()-90)) %>% as.data.frame()
  arimaMod <- auto.arima(ts(filledData)[,2])
  forecast1 <- forecast(arimaMod, h = input$slideIn4)
  output$plot5 <- renderPlot({
    plot(forecast1)
  })
  output$plot7 <- renderPlot({
    checkresiduals(arimaMod)
  })

```

```

})
output$foreArima <- renderPrint({
  accuracy(arimaMod)  }) })
test1 <- function(x, input){
  test <- tail(log(x), n = input)  }
train1 <- function(x, input){
  train <- ts(log(x)[1:(length(log(x))-input)], frequency = input)  }
#GMDH PAGE
observeEvent(input$gmdhStart, {
  output$gmdhSign <- renderText("Predicted values")
output$errorOutT <- renderText("Error values")
  filledData <- prepData()
  number <- input$slideIn3
  out = fcast(ts(train1(filledData$y, number)), weight = 0.9, method = "GMDH",
f.number = number, layer = 1)
  forecastTotal<- as.numeric(out$mean)
  output$gmdhF <- renderPrint({
    exp(forecastTotal)  })
  output$errorsOut3 <- renderPrint({
    cbind(c("mse", "mape", "mae"),
      c(
        round(mse(forecastTotal, test1(filledData$y, number)), 6),
        mape(forecastTotal, test1(filledData$y, number)),
        mae(forecastTotal, test1(filledData$y, number))))
    }) }) }
shinyApp(ui = ui, server = server)

```