

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Харківський національний університет радіоелектроніки
Центр післядипломної освіти
Кафедра Програмної інженерії

КВАЛІФІКАЦІЙНА РОБОТА

Пояснювальна записка

другий (магістерський)
(рівень вищої освіти)

Дослідження методів машинного навчання при розробці програмного
забезпечення для формування портфелю інвестора

Виконав:
студент 2 курсу, групи ІПЗдм-21-2
Ласіцин О.В.
(прізвище, ініціали)

Спеціальність 121 – Інженерія програмного
забезпечення
(код і повна назва спеціальності)

Тип програми Освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Керівник доц. Назаров О.С.
(посада, прізвище)

Допускається до захисту
Зав. кафедри

(підпис) З.В. Дудар
(прізвище, ініціали)

2023 р.

Харківський національний університет радіоелектроніки

Факультет _____ Центр післядипломної освіти _____
(повна назва)

Кафедра _____ Програмної інженерії _____
(повна назва)

Рівень вищої освіти _____ другий (магістерський) _____

Спеціальність _____ 121 – Інженерія програмного забезпечення _____
(код і повна назва)

Тип програми _____ освітньо-наукова програма _____
(освітньо-професійна або освітньо-наукова)

Освітня програма _____ Інженерія програмного забезпечення _____
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

«__» _____ 2023р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

студента _____ Ласіцина Олексія Володимировича _____
(прізвище, ім'я, по батькові)

1. Тема роботи «Дослідження методів машинного навчання при розробці програмного забезпечення для формування портфелю інвестора» затверджена наказом університету від «03» квітня 2023 р. № 83 Стз
2. Термін подання студентом роботи до екзаменаційної комісії «08» травня 2023 р.
3. Вихідні дані до роботи методи машинного навчання при розробці програмного забезпечення для формування портфелю інвестора
4. Перелік питань, що потрібно опрацювати в роботі мета роботи, аналіз предметної галузі і постановка задачі, методи машинного навчання при розробці програмного забезпечення для формування портфелю інвестора

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Аналіз предметної галузі	25.01.23 – 19.02.23	<i>Виконано</i>
2	Огляд існуючих методів та алгоритмів	20.02.23 – 10.03.23	<i>Виконано</i>
3	Алгоритми захисту ПЗ	11.03.23 – 25.03.23	<i>Виконано</i>
4	Підготовка пояснювальної записки	26.03.23 – 15.04.23	<i>Виконано</i>
5	Спецчастина	15.04.23 – 17.04.23	<i>Виконано</i>
6	Підготовка презентації та доповіді	18.04.23 – 07.05.23	<i>Виконано</i>
7	Попередній захист	19.05.23	<i>Виконано</i>
8	Нормоконтроль, рецензування	01.05.23 – 15.05.23	<i>Виконано</i>
9	Занесення диплома в електронний архів	20.05.23	<i>Виконано</i>
10	Допуск до захисту у зав. кафедри	22.05.23	<i>Виконано</i>

Дата видачі завдання 23 січня 2023 р.

Студент _____
(підпис)

Керівник роботи _____ доц. Назаров О.С.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ / ABSTRACT

Кваліфікаційна робота магістра містить: 66 с., 17 рис., 8 табл., 27 джерел.

ARIMA, ARMA, GARCH, RANDOM FOREST, TARCH, XGBOOST, КЛАСИФІКАТОР, МАРКОВІЦ, НЕЙРОННА МЕРЕЖА, ПОРТФЕЛЬНА ОПТИМІЗАЦІЯ, ПРОГНОЗУВАННЯ, ЦІНИ АКЦІЙ

Мета даної роботи полягає у дослідженні фінансових та економічних прогнозів за допомогою аналізу фінансових даних і методів машинного навчання з метою максимізації портфеля інвестора.

Об'єкт дослідження: фінансові часові ряди цін акцій на фондовому ринку.

Методи дослідження: методи аналізу часових рядів у дослідженнях Критерії Шарпа, модель Марковіца, радіальні базисні функції, випадковий ліс, ARIMA, AR, GARCH, TARCH, XGboost і нейронна мережа.

Актуальність теми: Сьогодні існує безліч варіантів як пасивного, так і активного фінансового збагачення. Інвестування грошей для отримання пасивного доходу є однією з багатьох можливостей. Незважаючи на широкий спектр перспектив фінансового інвестування, важливо ретельно вибирати інвестиційну мету. У зв'язку з цим нині є і, ймовірно, завжди буде інтерес до дослідницької галузі оптимізації портфеля інвесторів.

Програмну систему реалізовано за допомогою середовища розробки Jupyter Notebook та мови програмування Python.

Отримані результати: створено підхід для оптимізації портфеля, заснований на передбаченні фінансових даних, розподілених як часові ряди.

ARIMA, ARMA, GARCH, RANDOM FOREST, TARCH, XGBOOST, CLASSIFIER, MARKOVITZ, NEURAL NETWORK, PORTFOLIO OPTIMIZATION, FORECASTING, STOCK PRICES

The purpose of this work is to study financial and economic forecasts using financial data analysis and machine learning methods in order to maximize the investor's portfolio.

Research object: financial time series of share prices on the stock market.

Research methods: methods of time series analysis in research Sharpe Criterion, Markowitz model, radial basis functions, random forest, ARIMA, AR, GARCH, TARARCH, XGboost and neural network.

Relevance of the topic: Today, there are many options for both passive and active financial enrichment. Investing money for passive income is one of many possibilities. Despite the wide range of financial investment prospects, it is important to choose an investment goal carefully. Because of this, there is now, and probably always will be, interest in the research field of investor portfolio optimisation.

The software system is implemented using the Jupyter Notebook development environment and the Python programming language.

Results obtained: an approach for portfolio optimisation based on the prediction of financial data distributed as time series has been created.

Я, Ласіцин Олексій Володимирович, студент групи ППЗдм-21-2 здобувач вищої освіти на другому (магістерському) рівні, кафедра програмної інженерії, заявляю: моя кваліфікаційна робота на тему Дослідження методів машинного навчання при розробці програмного забезпечення для формування портфелю інвестора, що буде представлена до ЕК для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в електронному архіві відкритого доступу EIArKhNURE. Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомлений (а) з діючим положенням «Про протидію академічному плагіату в ХНУРЕ», згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

ЗМІСТ

ВСТУП.....	8
1 ДОСЛІДЖЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ	9
1.1 Інвестиції.....	9
1.1.1 Види інвестицій	9
1.2 Фондовий ринок та його структура.....	10
1.2.1 Фондовий ринок	12
1.2.2 Основні учасники фондового ринку.....	12
1.2.3 Типи акцій на ринку	13
1.3 Процес ціноутворення на ринку та фактори впливу.....	14
1.4 Фінансові ризики та їх види	15
1.5 Портфельна теорія.....	16
1.5.1 Пасивні і активні інвестиції.....	16
1.5.2 Модель Марковіца як приклад сучасної портфельної теорії	17
1.5.3 Модель Шарпа	19
1.5.4 Поняття межі ефективності	20
1.5.5 Прибутковість та ризикованість портфелів інвесторів.....	21
1.5.6 Основні типи інвестиційних портфелів.....	23
1.5.7 Основні висновки теорії портфельних інвестицій	24
1.6 Постановка задачі портфельної оптимізації	24
Висновки до розділу.....	25
2 ВИКОРИСТАННЯ МАШИННОГО НАВЧАННЯ ДЛЯ ПОБУДОВИ ПОРТФЕЛЮ ІНВЕСТОРА.....	27
2.1 Методи машинного навчання.....	27
2.1.1 Методи попередньої роботи з даними.....	27
2.1.2 Регресійні методи	28
2.1.3 Нейронні мережі	30
2.1.4 Ансамблеві методи	32
2.2 Основні характеристики фінансового часового ряду	34
2.2.1 Williams %R.....	34
2.2.2 Індекс відносної сили.....	35
2.2.3 Стохастичний осцилятор	36
2.3 Методи оцінки ефективності роботи моделей.....	37

Висновки до розділу.....	41
3 АНАЛІЗ ОБЧИСЛЮВАЛЬНИХ ЕКСПЕРИМЕНТІВ ТА ОГЛЯД	
ПРОГРАМНОГО ПРОДУКТУ	42
3.1 Вхідні дані	42
3.2 Підбір моделей для прогнозування доходностей і волатильностей часових рядів	45
3.3 Побудова і прогнози щотижневих трендів методами машинного навчання	53
3.4 Модель Марковіца для побудови оптимального портфелю акцій	56
Висновки до розділу.....	58
ВИСНОВКИ.....	60
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ.....	61
ДОДАТОК А Перелік джерел посилання за науковими напрямками керівника та науковців кафедри програмної інженерії	64
ДОДАТОК Б Звіт результатів перевірки на унікальність тексту.....	65
ДОДАТОК В Слайди презентації.....	66
ДОДАТОК Г Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008:2015.....	74
ДОДАТОК Д Наукові публікації.....	75

ВСТУП

Як і фондова біржа, ринок зараз перебуває у досить крихкій ситуації. Таким чином, одна з ключових завдань при роботі з інвестиціями – це вивчення ринкових умов і потенціалу вибору стратегії для оптимального вибору. На сьогоднішній день такий вид угод є способом заробітку як великого бізнесу, так і юридичних осіб або людей з помірними фінансовими оборотами.

З появою криптовалют, які є менш стабільними та мають більш непередбачувану поведінку, ніж інші валюти, завдання аналізу поведінки ринку стало ще більш важливим.

Для історичного аналізу портфеля інвестора були застосовані перевірені макроекономічні моделі та методи. Вони засновують свої рішення на оцінці волатильності активів і ризику.

У роботі пропонується взяти до уваги дві стратегії: метод формування портфеля, який використовуватиме ансамблевий та регресійний методи відбору активів, що дозволить розробити варіативний підхід до формування портфеля та дозволить вибрати стратегії на основі переваг інвестора. У документі також пропонується взяти до уваги підхід, який використовує методи машинного навчання для більш детального аналізу історичних даних.

1 ДОСЛІДЖЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ

У цьому розділі будуть розглянуті терміни інвестиції, фондовий ринок, портфель інвестора та фінансовий ризик.

1.1 Інвестиції

Інвестиція – це акт отримання грошей, майна, цінних паперів тощо від осіб, які пізніше будуть називатися інвесторами. Оскільки кредитні рахунки потрібно якимось чином надавати в банк і оплачувати вчасно, це має велике значення для людей, в яких зроблено інвестиції, через обурення. Інвестиційні кошти, з іншого боку, часто надаються або на безвідшкодовній основі, або за відсотковою ставкою, і вони будуть відшкодовані лише за рахунок прибуткових підприємств. Таким чином, інвестування схоже на покупку kota в мішку.

Ми можемо вжити необхідних заходів, щоб відреагувати, якщо проект стає збитковим, і ліквідувати його.

Інвестиційна діяльність – це фінансові рішення та дії, пов'язані з потенційним майбутнім доходом, спрямовані на підвищення прибутковості.

1.1.1 Види інвестицій

Виходячи з особистих і професійних інтересів фізичних і юридичних осіб в інвестиційних діях, можна виділити такі види інвестицій:

- меркантильні (особливий прибуток інвестора, основна мета цього виду – отримання прибутку);
- некомерційні отримання прибутку (мета – отримання соціального ефекту);
- асоційовані (пріоритети та цілі, які будуть реалізовані в майбутньому та сприятимуть іншим цілям інвестора).

Далі наведено підсумок того, що було згадано досі:

Як категорія економічних відносин між суб'єктами, інвестиції включають дії, що характеризуються довгостроковими фінансовими вливаннями, фактичним аудитом основних фондів, науково-технічних розробок, обладнання, рівня

кваліфікації спеціалістів. Ці дії в кінцевому підсумку підвищують ефективність підприємства і, як наслідок, отримують віддачу від вкладених коштів, у більш широкому сенсі. Подібні заходи фінансової допомоги тривають до тих пір, поки бізнес не буде самостійно отримувати прибуток і більше не потребуватиме участі партнерів, що означає виконання іншої стратегічної мети, поставленої інвесторами перед початком інвестування.

З правових положень чинного законодавства ми можемо зробити висновок, що інвестиційні активи сьогодні можуть мати такі форми: грошові (банківські активи, цінні папери), матеріальні (різні права власності на рухоме та нерухоме майно), різні форми інтелектуальної власності та так далі (авторське право, патенти).

1.2 Фондовий ринок та його структура

Однією з сфер ринку капіталу, де випускаються, купуються та продаються цінні папери, є фондовий ринок (ринок цінних паперів).

Активний ринок – це ринок, на якому відбувається багато частих значних операцій із певним типом цінних паперів або предметів.

Ідея фондового ринку, яка визначає діяльність для потенційної торгівлі цінними паперами, є абстрактною. Фондова біржа, на відміну від фондового ринку, є підприємством, яке сприяє здійсненню торговельних операцій і встановленню потенційних клієнтів – покупців і продавців цінних паперів. Проте ідея регульованого (інституціоналізованого) фондового ринку пов'язана з фондовою біржею.

Фондовий ринок – це абстрактна сутність, створена для полегшення торгівлі цінними паперами. Вона значно відрізняється від фондової біржі, оскільки остання є компанією, головною метою якої є знайомство клієнтів один з одним. Фондовий ринок можна порівняти з фондовою біржею, оскільки він де-факто служитиме тій самій меті, якщо його адміністративно оформити як вид офіційної діяльності.

Акціонер володіє акціями, а всі акціонери, які володіють акціями, колективно володіють корпорацією. Відсоток власності компанії розраховується на основі

кількості акцій, придбаних одним власником, по відношенню до загальної кількості акцій у власності. Як правило, невелика група людей з найбільшою кількістю спільних ресурсів приймає рішення про те, як адмініструвати такий спільний ресурс. Контролюючим акціонером є будь-яка особа, яка накопичила значну кількість акцій.

Фондові ринки бувають різних форм і розмірів. Розглянемо основні категорії:

- за поширеністю;
- за стадією обігу, включаючи первинні (угоди, укладені вперше основними підприємствами) і вторинні (угоди, побудовані на основі первинних, які згодом будуть реалізовані з цінними паперами);
- за емітентом – Залежно від емітента різні цінні папери мають різні ринки.

Теми РЦП:

- емітенти займаються фактичним випуском отриманих цінних паперів;
- регулюючі організації, які створюють платформи та організують вимоги для підписання контрактів;
- посередники або особи, які є посередниками та координують угоди або здійснюють посередницьку діяльність, пов'язану з цінними паперами;
- інвестори – це ті, хто робитиме внесок у покупку;
- учасники інфраструктури – це ті, хто, в тому числі, працює над розвитком інформаційної та інформаційної інфраструктури.

Основними атрибутами запропонованої функціональності є:

- заходів щодо контролю за доступністю кредитів та обігом грошей;
- допомога в переміщенні капіталу між різними секторами. сектори економіки;
- сприяння ефективному використанню інвестованого капіталу;
- сприяння в інформуванні клієнтів щодо інфраструктури.

1.2.1 Фондовий ринок

Фондовий ринок, також відомий як ринок цінних паперів, – це ринок, який дозволяє купувати та продавати акції, а також капіталовкладення та інвестиції.

Ринок, на якому акумулюються операції з певним типом товарів, які потім включаються в більші фінансові операції і фактично є частиною більш широкого фондового ринку, називається активним ринком.

Ідея фондового ринку є суто фіктивною та використовується для визначення потенційної торгової діяльності. Фондова біржа, на відміну від фондового ринку, є бізнесом, який сприяє торгівлі та встановленню потенційних клієнтів, які є покупцями та продавцями цінних паперів.

Проте ідея регульованого (інституціоналізованого) фондового ринку еквівалентна ідеї фондової біржі.

Акціонер, який володіє акцією фірми, є власником цієї частки, яка дорівнює кількості акцій у власності як відсоток від усіх акцій компанії.

Як приклад, особа або компанія, яка володіє 100 000 акцій фірми, яка має один мільйон акцій в обігу, володітиме 10% компанії. Більшість сучасних компаній мають мільйони або мільярди акцій.

1.2.2 Основні учасники фондового ринку

Основними гравцями на ринку цінних паперів є:

– емітент цінних паперів – уповноважений державою орган, юридична або фізична особа (останні лише у випадку, передбаченому законом). Обов'язки, що впливають з умов емісії, виконують емітенти, які випускають цінні папери за власний рахунок.

– торговці цінними паперами – суб'єкти, які можуть здійснювати такі види діяльності, пов'язані з випуском і розповсюдженням цінних паперів.

Основними цілями роботи учасників є:

– процес організації підписки торговця цінними паперами на цінні папери від імені, за рахунок і за дорученням емітента називається випуском цінних

паперів. У цьому випадку торговець цінними паперами може, відповідно до емітента, взяти на себе зобов'язання у разі незавершеного розміщення цінних паперів і викупити невипущені цінні папери в емітента.

– брокерство з цінними паперами – це група видів діяльності, пов'язаних із купівлею та продажем цінних паперів, зокрема за допомогою третьої сторони з обов'язком виступати в якості посередника.

– комерційна робота та інші завдання, пов'язані з цінними паперами та контрактами, які виконуються за рахунок посередника.

Торгівля цінними паперами не допускається:

- 1) власні емісійні цінні папери;
- 2) акції, в яких він прямо чи опосередковано володіє більш ніж 5% випущеного акціонерного капіталу.

1.2.3 Типи акцій на ринку

Звичайні акції та привілейовані акції є двома основними категоріями акцій.

Привілейовані акції зазвичай не мають права голосу, що не дозволяє звичайним акціонерам брати участь у зборах компанії (таких як річні загальні збори або загальні збори), коли мають бути прийняті рішення щодо виборів директорів або призначення аудиторів.

Для отримання дивідендів та активів у разі ліквідації привілейовані акції фірми мають перевагу над звичайними акціями. Оскільки їх ринкова вартість і обсяг торгів у багато разів перевищують вартість привілейованих акцій, термін «акції» найчастіше використовується для позначення звичайних акцій.

Права голосу за звичайними акціями є іншим способом їх класифікації.

Хоча ідея, що лежить в основі звичайних акцій, полягає в тому, що всі вони повинні мати однакові права голосу – один голос на акцію, – деякі підприємства мають два або більше класів акцій, кожен з яких має окремий набір прав голосу.

Акції класу В із «субординованим голосом» іноді мають лише один голос на акцію, але акції класу А у випадку подвійного класу можуть мати десять голосів на акцію. Існують угоди про акції з двома або більше класами, які передбачають

контроль засновника над ситуацією, стратегічним напрямком та інноваційним потенціалом.

1.3 Процес ціноутворення на ринку та фактори впливу

Існує багато способів оцінити акції на фондовому ринку, але найпопулярнішим є аукціон, де покупці та дилери роблять ставки та пропонують купівлю та продаж. Ціна, за якою покупець бажає купити, вказується його ставкою, тоді як ціна продавця показується його ставкою. Торгівля відбувається, якщо обидві заявки рівні та прийняті сторонами.

Загальний ринок складається із значної кількості трейдерів та інвесторів, які можуть мати різні думки щодо вартості конкретної акції та, відповідно, різні ціни, за якими вони були б готові її придбати чи продати. Протягом торгового дня акція зазнає сотень угод, оскільки трейдери та інвестори втілюють свої наміри в життя, купуючи та/або продаючи її. Біржа пропонує платформу, на якій можна легко проводити операції, з'єднуючи покупців і продавців акцій. Щоб звичайна особа отримала доступ до цих бірж, потрібен біржовий брокер. Посередником між покупцем і продавцем є біржовий маклер. Найпоширеніший спосіб отримати біржового брокера – відкрити рахунок у авторитетного роздрібного брокера.

Нині для визначення цін на цінні папери на ринку використовуються наведені нижче методи:

- експертний метод, який передбачає використання обґрунтованих суджень експерта щодо рівня реальної вартості конкретної інвестиції або вартості конкретної біржової операції;
- аналітичний підхід – ціноутворення здійснюється з використанням ретельного вивчення стану ринку, різних форм спостереження та аналізу поточних факторів;
- статистичні методи включають побудову математичних моделей, керування законами теорії ймовірностей, виконання регресійного аналізу тощо;
- за допомогою нормативно-параметричного (бального) підходу обрана група якостей, пов'язаних з нашими активами, буде оцінюватися професіоналами

в рамках обраної системи оцінки. Їм буде присвоєно бал на основі їх значущості, який покаже рівень впливу однієї або кількох характеристик, викладених у вибраних одиницях (балах);

– за балансовим методом кожна акція оцінюється шляхом прирівнювання її до еквівалентної кількості реальних фінансових або фізичних активів компанії (певний відсоток, однаковий для кожної акції). Таким чином, вартість акцій представлятиме фактичні грошові прибутки, отримані компанією;

– економіко-математичний метод: техніка відповідного моделювання, яка використовує графічні або логічні моделі, які працюють на основі встановлених комп'ютерних програм для моделювання динамічних моделей або даних з використанням динамічних моделей або даних. Моделі, які мають справу зі встановленням зв'язків і імітацією змін цін у відповідь на динамічні зміни в списку визначених факторів, включені в основу.

1.4 Фінансові ризики та їх види

Виділяють такі категорії фінансового ризику:

- а) кредитний ризик – це категорія, за якою торгові конкуренти не виконують договірних зобов'язань внаслідок фінансових втрат. Управління кредитним ризиком стало надзвичайно важливим, оскільки воно має значний вплив на інтереси компаній або окремих осіб. Великі підприємства часто використовують спеціалізований персонал для оцінки кредитоспроможності кожного партнера по транзакції з метою розрахунку потенційних прибутків і збитків;
- б) ризик ліквідності – це тип фінансового ризику, спричиненого неможливістю швидко продати фінансові активи, цінні папери або фізичні об'єкти за ціною, близькою до ринкової, і, що більш важливо, без зазнавання великих збитків. Генезис цього явища пов'язаний із незавершеністю ринку та інформаційною асиметрією;
- в) зміни процентних ставок, які не є вигідними для фінансових установ та інвесторів, є причиною процентного ризику. Цей ризик є значним джерелом

- доходу та капіталу для фінансових організацій. Проте надмірний процентний ризик негативно впливає на капітальну базу та надлишок установи;
- г) валютний ризик – це ризик, викликаний змінами валютних курсів;
 - д) операційний ризик – це ймовірність зміни вартості, що означає, що реальні збитки можуть відрізнятися від прогнозованих внаслідок зовнішніх подій або неефективних або несправних внутрішніх систем;
 - е) галузевий ризик – це потенційна можливість збитків у результаті конкретних змін у роботі галузі, які відбуваються протягом її життєвого циклу;
 - ж) політичний, правовий або регуляторний ризик;
 - и) ринковий ризик – це ризик, який стосується активів у вибраному фінансовому портфелі та існує або може стати можливим. Ці небезпеки є результатом змін обмінних курсів валют, змін ринкових цін на активи, погіршення вартості цінних паперів тощо;
 - к) системний ризик означає можливість нестабільності фінансової системи;
 - л) стихійні катастрофи чи інші кризи.

1.5 Портфельна теорія

1.5.1 Пасивні і активні інвестиції

Диверсифікований портфель використовується в пасивних інвестиціях, які мають нижчий загальний рівень ризику, ніж окремі акції в сукупності. Оскільки його можна легко адаптувати до будь-якої моделі фінансового інвестування, цей метод інвестування є чудовим вибором для всіх типів інвесторів з різними фінансовими обмеженнями. Однак навіть невеликих ризиків можна уникнути за допомогою правильної інвестиційної стратегії.

Основний висновок із цього полягає в тому, що створення пасивного доходу є трудомістким і точним процесом. Перед побудовою фінансового портфеля інвестора першочерговими завданнями є проведення дослідження ринку, аналіз цільових цілей інвестора, отримання історичної інформації про характеристики ринку та вибір відповідних стратегічних інвестиційних цілей.

Сьогодні пасивний дохід має низький рівень ризику. Більш агресивні та небезпечні форми інвестування є прямо протилежними цьому.

1.5.2 Модель Марковіца як приклад сучасної портфельної теорії

Модель Гаррі Марковіца середини 20 століття все ще використовується для визначення методу інвестування. Він носить ім'я автора і широко використовується в сучасну епоху.

Стратегія автора передбачає вибір певної кількості інвестиційних цілей із різними рівнями ризику та прибутковості, що дозволить коригувати майбутні доходи та реінвестиції.

Модель має дві основні мети. Однак через протиріччя між цими цілями необхідно провести диверсифікацію шляхом придбання цінних паперів іншої вартості.

Кореляція між доходами різних активів у портфелі слугує основою моделі. Це виправдано зв'язком між поведінкою часових рядів активів і самими їхніми активами. Тому для нас буде надзвичайно важливо з'ясувати величину та напрямок їх кореляції один з одним. Це стохастична міра, ця кореляція.

Модель Марковіца є розумним вибором, коли стан фондового ринку залишається незмінним, якщо потрібно створити портфель різнорідних активів. Іншим недоліком моделі є те, що вона робить припущення, що прогнозована прибутковість цінних паперів буде такою ж, як середня прибутковість, яка спостерігалася історично. Через це має сенс використовувати його, коли фондовий ринок є стабільним, а активи портфеля мають тривалий термін зберігання.

Щоб створити інвестиційний портфель, Гаррі Марковіц розробив методологію, відому як теорія портфеля Марковіца (аналіз середньої дисперсії, який є підходом, заснованим на аналізі очікуваних середніх значень і варіацій випадкових змінних). Він спрямований на те, щоб зробити найкращий можливий вибір активів, враховуючи необхідне співвідношення прибутковості та ризику.

Середньозважене значення доходів від кожного титулу цінних паперів – це те, як обчислюється прибутковість портфеля.

$$r_p = \sum_{i=1}^n w_i \cdot r_i \quad (1.1)$$

де w_i – відсоток інструменту в портфелі;
 r_i – прибутковість інструменту.

Стандартне відхилення доходності певного інструменту використовується для вимірювання його ризику. Під час розрахунку загального ризику портфеля (через коваріацію та кореляцію, показники взаємозв'язку) життєво важливо враховувати загальну зміну ризику кожного інструменту та їх взаємозалежність.

$$\sigma_p = \sqrt{w_j * w_j * V_{ij}} = \sqrt{\sum_{i=1}^n w_i^2 * \sigma_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n w_i * w_j * k_{ij} * \sigma_i * \sigma_j} \quad (1.2)$$

де σ_i – стандартне відхилення прибутковості інструменту;
 k_{ij} – коефіцієнт кореляції між і-м та j-м інструментами;
 V_{ij} – коваріація між доходністю і-го та j-го фінансових інструментів;
 n – кількість фінансових інструментів у портфелі.

У результаті, враховуючи негативну кореляцію інструментів, добре розроблений портфель забезпечує значне зниження як систематичного (ринкового) ризику, так і ризику, пов'язаного з конкретним інструментом.

Для складання портфоліо вирішено питання оптимізації. Оскільки сума часток активів дорівнює одиниці, і ці частки є позитивними, використання позичених грошей не передбачається в основній формі.

$$\left\{ \begin{array}{l} \sqrt{\sum_{i=1}^n w_i^2 * \sigma_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n w_i * w_j * k_{ij} * \sigma_i * \sigma_j} \rightarrow \min \\ \sum_{i=1}^n w_i * r_j * > r_p \\ \sum_{i=1}^n w_i = 1 \\ W_i \geq 0 \end{array} \right. \quad (1.3)$$

Збільште прибуток, припускаючи відомі рівні ризику:

$$\left\{ \begin{array}{l} \sum_{i=1}^n w_i * r_i \rightarrow \max \\ \sqrt{\sum_{i=1}^n w_i^2 * \sigma_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n w_i * w_j * w_{ij} * \sigma_i * \sigma_j} < \sigma_p \\ \sum_{i=1}^n w_i = 1 \\ W_i \geq 0 \end{array} \right. \quad (1.4)$$

Ми знижуємо ризик, зберігаючи прийнятний мінімальний рівень прибутковості.

1.5.3 Модель Шарпа

Модель Шарпа враховує ідею мінімізації умов згаданої суми прибутку на кожен окремий актив. Початкова мета полягає в тому, щоб брати до уваги лише ті моделі, де потенційний прибуток можна розглядати як компенсацію поточного ризику. Прибуток за довгостроковими контрактами з довгостроковим терміном погашення (понад 10 років) розглядається як основа для стратегії березового ведмедя. Для аналізу великої кількості активів, які представляють велику частку ринку, ця модель буде найбільш корисною.

Однак призначення має один недолік: ми повинні передбачити як прибутковість ринку в цілому, так і норму безризикового прибутку, яка також буде врахована без її коливань і, отже, не повністю враховує невизначеність. Крім того, багато залежить від того, як розвивається зв'язок між безризиковою частиною прибутку та прибутком фондового ринку.

Основним питанням, яким займається теорія портфельних інвестицій, є процедура поступового вибору, створення та, зрештою, вибору найкращого

портфеля. Щоб підвищити точність і визначити найкращі рішення, часто переглядається та вивчається величезна кількість активів. Однак це ускладнює проблему, класифікуючи її як багатовимірну.

1.5.4 Поняття межі ефективності

Цей підхід забезпечить опуклу межу ефективності, яка представляє межу колекції ефективних портфелів. Ми стверджуватимемо, що оптимальний результат буде точно посередині створеного ефективного фронту, оскільки він забезпечить збалансований результат між толерантністю до ризику та прогнозованою прибутковістю, враховуючи розумну техніку побудови оптимального результату та цільові функції, які використовуються для цього. Більше теоретично успішних портфелів буде створено, оскільки ми беремо на себе більший ризик, але вони також матимуть більшу ймовірну оцінену ймовірність збитку. Через це інвестори зазвичай тяжіють до більш поміркованих стратегій, у яких значні суми їхніх коштів вкладаються в активи з низьким рівнем ризику, а наявність високоризикових активів компенсується значною часткою тих, які запропонують стабільний прибуток із високою ймовірністю успіху.

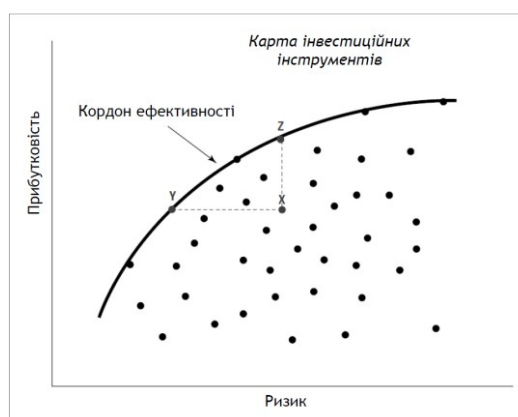


Рисунок 1.1 – Кордон ефективності

Ми можемо побачити взаємозв'язки, згадані вище, на рисунку вище (див. рис. 1.1), який представляє межу ефективності:

- більш ризиковані інвестиції приносять більший прибуток;

– низькі прибутки не передбачають значних ризиків.

У результаті наш пошук ідеальних портфоліо призводить до пошуку ідеальних засобів і, зрештою, до пошуку оптимізації.

1.5.5 Прибутковість та ризикованість портфелів інвесторів

Ризик і прибутковість мають тісний зв'язок у фінансовому менеджменті. Ризик – це ймовірність зазнати збитків або заробити менше, ніж очікувалося. Ступінь зміни самого активу служить проксі для ризику (ризикованості) активу. Традиційно для вимірювання фінансового ризику використовується волатильність різноманітних активів.

Створення портфеля інвестора – набору основних акцій, визначених експертами як найбільш надійні та відповідні для поточних інвестицій – є частою практикою. Основною метою цієї стратегії є зниження ризику під час фінансового інвестування, оскільки деякі акції принесуть більший прибуток для оцінюваної сторони, тоді як інші акції призведуть до грошових втрат. Наша здатність отримувати прибуток і компенсувати частину збитків (бажано всі) залежить від того, наскільки добре побудований наш портфель. Менш ризиковані акції будуть вибрані для пасивного раннього інвестування, що зменшить потенційні втрати та збільшить потенційні прибутки.

Дохідність кожної акції в портфелі додається разом, щоб отримати загальну дохідність:

$$r_p = \sum_{i=1}^n x_i \times r_i \quad (1.5)$$

де x_i – показник долі i -того активу в портфелі;

r_i – прогнозована прибутковість i -го активу;

n – загальна кількість власних активів акцій у портфелі.

Міра ризику портфеля, також відома як його ризикованість, часто є розрахунком дисперсії або відхилення від математичного очікування всіх активів, які складають портфель.

Вимірювання ризику окремих активів портфеля можна використовувати для визначення ризику портфеля, однак через обчислення середнього арифметичного цей зв'язок не виражений і натомість більш спотворений.

Ризик дійсно зменшується зі збільшенням кількості акцій у портфелі, але він ніколи не буде нульовим. Щоб це було здійсненним, має існувати значна кількість акцій або з негативним коефіцієнтом кореляції, або з нульовим коефіцієнтом кореляції.

Склад і структуру портфеля необхідно змінити, щоб змінити очікувану прибутковість і ризик.

Залежно від цілей інвестора існує безліч методів формування портфеля. Дві основні стратегії, які розглядаються, це довгострокова торгівля та короткострокова торгівля. У цьому сценарії враховуються прогнозовані доходи за різні часові рамки, і рівень ризику портфеля також буде різним. Здійснюючи короткострокові інвестиції, метою є максимізація прибутку при збереженні певного рівня толерантності до ризику; при здійсненні довгострокових інвестицій мета частіше полягає в мінімізації ризику.

Інвестор та його погляд на визначення ризику визначають, чи є портфель оптимальним чи неоптимальним. Для визначення ризику застосовується ідея функції байдужості.

Він демонструє ризики та прибуток пакета портфелів з однаковою вагою, щоб проілюструвати зв'язок між ризиком і доходністю. Привернути увагу до особливостей конкретного портфоліо буде простіше за допомогою графічного зображення.

1.5.6 Основні типи інвестиційних портфелів

Залежно від інвестора та того, як він визначає ризик, портфель буде оптимальним або неоптимальним. Ризик визначається за допомогою ідеї функції байдужості.

Відображаючи ризики та прибутковість пакета портфелів з однаковою вагою, він ілюструє, як пов'язані ризик і прибутковість.

Тому вони функціонують, використовуючи такі ідеї:

- а) консервативний. Усі акції в такому портфелі, який також містить золото, нерухомість та інші активи, є найбільш стабільними. Цільова функція мінімізації визначатиме рівень ризику, а власник портфеля вирішуватиме рівень дохідності та її межі. Оскільки консервативний портфель є відносно дорогим, він не обов'язково є найрозумнішим вибором для тих, хто хоче отримати пасивний дохід без початкових інвестицій;
- б) помірний. Завдання виконується одночасно, намагаючись зменшити ризики та максимізувати дохідність. Вони часто намагаються спрямувати загальний ризик портфеля до середньоринкового рівня. Це досягається за допомогою компенсаторних механізмів, які збалансовують помірно прибуткові акції з низьким ризиком і високоризикові акції з високим рівнем волатильності;
- г) агресивний. Усі компанії в агресивному інвестиційному портфелі мають потенціал бути високоприбутковими і ризикованими інвестиціями, і ними потрібно керувати обережно та в короткостроковій перспективі. Такий портфель, швидше за все, буде невдалим при довгостроковому підході, оскільки акції з високою прибутковістю та високим ризиком ідеально підходять для стратегії пасивного збагачення та вимагають передових можливостей управління активами.

1.5.7 Основні висновки теорії портфельних інвестицій

Основні висновки такі:

- портфелі, які задовольняють одну з умов –максимізація прибутків при фіксованих ризиках або мінімізація ризиків при фіксованих доходах – належать до колекції ефективних портфелів;
- зі складеного списку найкращих можливостей інвестор повинен вибрати відповідний портфель;
- для знаходження найкращих портфелів застосовується теорія оптимального фронту, причому найкращі портфелі – ті, точки яких знаходяться на кривій байдужості інвестора;
- диверсифікація зменшує ризики за рахунок зменшення функції прибутковості дисперсії портфеля, яка буде меншою за загальну дисперсію всіх акцій у портфелі, оскільки вони матимуть відмінну від нуля кореляцію;
- дохідність цінних паперів до ринкового індексу є ринковою моделлю;
- ринкова модель має спорадичні помилки та помилки, які є результатом неадекватних пояснень елементів моделі;
- загальний ризик ринкової моделі визначатиметься тим, наскільки вона ризикована;
- диверсифікація може значно знизити загальний ризик портфеля.

1.6 Постановка задачі портфельної оптимізації

Щоб визначити структуру портфеля, яка пропонує найвищий прибуток за певного рівня ризику, необхідно подолати наступну проблему:

$$\{x_{opt}\} = \{x\} \mid r \rightarrow \max, \beta = \text{cont} \quad (1.6)$$

Припускаючи, що доходність портфеля:

$$r = (r_{min} = \sum_{i=1}^N x_i r_i; \tilde{r} = \sum_{i=1}^N x_i \tilde{r}_i; r_{max} = \sum_{i=1}^N x_i r_{2i}) \quad (1.7)$$

де $(r_{1i}, \tilde{r}_i, r_{2i})$ – прибутковість i -го цінного паперу.

Перед нами постає таке питання оптимізації:

$$\begin{aligned} r &= \sum_{i=1}^N x_i \bar{r}_i \rightarrow \max \\ \beta &= \text{const} \\ \sum_{i=1}^N x_i &= 1 \quad x_i \geq 0 \quad i = \overline{1, N} \end{aligned} \quad (1.8)$$

Розглянемо випадок, коли значення ризику коливається від 0 до 1 ($0 < \beta < 1$).

Ця обставина можлива, оскільки це можна спостерігати, коли

$$\sum_{i=1}^N x_i r_{i1} \leq r^* < \sum_{i=1}^N x_i \bar{r}_i \quad (1.9)$$

або так

$$\sum_{i=1}^N x_i \tilde{r}_i \quad \sum_{i=1}^N x_i r_{i1} \leq r^* \leq \sum_{i=1}^N x_i r_{i2} \quad (1.10)$$

Таким чином, структура бажаного ідеального портфеля відповідатиме вектору $x = \{x_i\} \quad i = \overline{1, N}$. Роблячи це, ми сформулюємо проблему визначення ідеального портфолію.

Висновки до розділу

Розглянуто поняття портфеля, фінансового ризику, його призначення при створенні інвестиційного портфеля, наукова класифікація. Описується ідея ринку та оцінюється його статус в Україні сьогодні. Робимо висновок, що моделі з використанням великих даних більш ефективні для визначення ситуації на фондовому ринку та прогнозування щодо неї.

Таким чином, ми визначимо мету як легку та швидку оцінку стану ринку та створення інвестиційних портфелів, що має бути в змозі зробити наступна система підтримки прийняття рішень. Щоб оцінити їх ефективність традиційним методом Марковіца, було вирішено розробити систему підтримки прийняття рішень з використанням ряду математичних моделей.

2 ВИКОРИСТАННЯ МАШИННОГО НАВЧАННЯ ДЛЯ ПОБУДОВИ ПОРТФЕЛЮ ІНВЕСТОРА

2.1 Методи машинного навчання

Методи машинного навчання часто використовуються для вирішення проблем прогнозування та прийняття рішень, що допомагає отримати кращі результати. У цьому проекті ми працюватимемо з представленнями часових рядів історичних фінансових даних.

2.1.1 Методи попередньої роботи з даними

Метод експоненційного згладжування. Це трапляється часто. Він використовується особливо, коли важливо працювати без всеосяжних і надійних знань про моделі розвитку процесу. Браун вперше запропонував техніку експоненціального згладжування, яку Холт пізніше розширив. У своєму дослідженні Гарднер розглядає теоретичні та емпіричні обґрунтування вибору параметра згладжування. Він також взяв до уваги альтернативні моделі експоненціального згладжування для тренду та сезонності (відсутність, адитивність та мультиплікативність) (відсутня, лінійна тенденція, експоненціальна, затухла).

Крім того, типовим є подвійне експоненціальне згладжування.

Нижче наведено опис метода подвійного експоненціального згладжування:

$$\begin{aligned} S_k &= \alpha \times y(k) + (1 - \alpha) \times (S_{k-1} - b_{k-1}), \\ b_k &= \beta \times (S_k - S_{k-1}) + (1 - \beta) \times b_{k-1}, \end{aligned} \quad (2.1)$$

де α і β – параметри згладжування $0 < \alpha < \beta, 0 < \beta < 1$.

Лінійне нормування даних. Нормалізація даних передбачає обмеження діапазону змін значень ознак конкретними, істотними межами (наприклад, від 0 до 1). При застосуванні численних багатовимірних статистичних методів нормалізація є важливим першим кроком у процесі обробки даних. Це особливо

вірно, якщо змінні вимірюються на шкалах із різко різними розмірами (наприклад, від міліметрів до кілометрів).

Якщо для опису зв'язку між початковим і кінцевим діапазонами можна використовувати лінійну функцію, то нормалізація є лінійною (значення ознак у всьому діапазоні змінюються на однакову величину).

Коли значення змінної X рівномірно покривають діапазон, вибирається лінійна нормалізація. Нормалізація не повинна зосереджуватися на екстремальних (граничних) значеннях, якщо в даних є нечасті аномалії, які значно перевищують типовий розкид:

$$\tilde{x}_i = \frac{x_i - x_{i,min}}{x_{i,max} - x_{i,min}} \quad (2.2)$$

Він є нелінійним, якщо використовується нелінійна функція, наприклад експонента (значення ознак, розташованих у різних частинах діапазону, змінюються по-різному).

2.1.2 Регресійні методи

Ми зможемо використовувати методи аналізу стаціонарності та сезонності, а також прогнозувати моделі фінансової вартості, такі як ARMA та ARIMA, на основі яких ми виберемо наші дані для ідеального портфеля, щоб передбачити поведінку часових рядів фінансових значень.

ARMA. Однією з математичних моделей, що використовуються в статистиці для оцінки та прогнозування стаціонарних часових рядів, є авторегресійна модель ковзного середнього (ARMA). Модель ковзного середнього (MA) і модель авторегресії (AR) є двома простішими моделями часових рядів, які узагальнюються моделлю ARMA.

Модель ARMA (p, q) – це метод генерації часових рядів, де p і q – цілі числа, що вказують порядок моделі:

$$X_t = c + \varepsilon_t + \sum_{i=1}^p a_i X_{t-i} + \sum_{i=1}^q \beta_i \varepsilon_{t-i} \quad (2.3)$$

де c – фіксоване значення (константа);

$\{\varepsilon_t\}$ – білий шум;

коефіцієнти - дійсні числа, змінними середніми коефіцієнтами та авторегресійними коефіцієнтами.

ARIMA. Модель аналізу часових рядів під назвою ARIMA (інтегроване ковзне середнє з авторегресією) об'єднує авторегресію та підхід середньої змінної. Це більш складна форма моделі ARMA, яка використовується для нестационарних часових рядів, які стаціонарні шляхом взяття відмінностей від вихідних рядів певного порядку (інтегрований часовий ряд). Відмінності порядку d будуть чутливими до моделі ARMA (p, q) у моделі ARIMA (p, d, q).

Модель ARIMA (p, d, q), яка є нестационарною, для часового ряду $X\{t\}$ буде розрахована таким чином:

$$\Delta^d X_t = c + \sum_{i=1}^p a_i \Delta^d X_{t-i} + \sum_{i=1}^q b_i \varepsilon_{t-i} + \varepsilon_t \quad (2.4)$$

де c, a_i, b_j – параметри моделі,

ε_t – стаціонарний ряд.

Δ^d – оператор різниці часових рядів d -порядку (послідовне взяття d різниць першого порядку – спочатку з часового ряду, потім з отриманих різниць першого порядку, потім з різниць другого порядку тощо).

GARCH, TARЧН. Узагальненою авторегресійною моделлю умовної гетероскедастичності є модель GARCH.

Це означає, що така модель імовірно передбачає, що дисперсія даного часового ряду буде не константою, а скоріше змінною, і що, отже, він також буде описаний іншим часовим рядом, який може бути описаний іншим правилом.

Модель GARCH передбачає, що закон процесу авторегресії ковзного середнього описуватиме дисперсію помилки.

Модель поширення волатильності, в якій очікується, що значення математичного очікування – середнього ряду – будуть постійними, точно описується моделлю GARCH (p, q).

Модель ARCH, яка має багато тих самих припущень, що й TARЧН. Він використовує абсолютні значення для моделювання волатильності. Потужність за замовчуванням 2 відповідає процесам дисперсії, які розвиваються у квадратах, отже, ця модель визначається за допомогою $power = 1,0$.

2.1.3 Нейронні мережі

Багатошарова штучна нейронна мережа Одним із найсучасніших підходів до завдань прогнозування та класифікації є багатошарова нейронна мережа.

Сьогодні ми можемо сказати, що методи глибокого навчання, які включають багатошарові нейронні мережі, добре працюють в аналізі частотних рядів і обробці зображень.

Математики повинні вибрати найкращу архітектуру нейронної мережі, збирати адекватні дані та отримувати менш конфіденційні дані, щоб досягти хороших результатів.

Типовим входом до нейронної мережі є дані, які потім обробляються перед тим, як певний рівень або нейрон буде активовано відповідно до параметрів, необхідних для його роботи.

Кожен нейрон шару з'єднаний із нейроном під ним і виробляє імітований синаптичний сигнал, який допоможе визначити зв'язок між даними до того, як це станеться насправді.

Нейронна мережа має велику чутливість до масштабування та трансформації даних, які будуть згенеровані в майбутньому, оскільки це цілісна структура.

Традиційна нейронна мережа надсилає дані одним шляхом і не має механізму пам'яті.

Однак ці нейронні мережі використовуються рідше через їх зниження ефективності. Сьогодні більш поширеними є нейронні мережі, що включають пам'ять або механізм зворотного поширення помилок.

Перенавчання, коли дані мають високу точність на навчальній вибірці, але надто низьку на навчальній вибірці, є одним із недоліків моделі нейронної мережі.

Перехресна перевірка даних, більш складний дизайн і менша кількість епох – це шляхи вирішення цієї проблеми.

Наразі ми знаємо про проблему перетренованості з багатьма шарами; точність нейронних мереж з більш ніж п'ятнадцятьма шарами швидко падає і утворюються неправильні залежності.

Хоча зараз нейронні мережі використовуються для створення програм часових рядів, їхня ефективність нині поступається більш традиційним підходам, які точніше обробляють значення даних, враховуючи сезонні коливання, закономірності тощо.

Радіально-базисна нейронна мережа. Нейронна мережа прямого розповсюдження сигналу, відома як мережа радіальної базисної функції, включає проміжний (прихований) шар радіально-симетричних нейронів. Згідно з деяким нелінійним законом, такий нейрон змінює відстань між заданим вхідним вектором і його відповідним «центром» (зазвичай функцією Гауса).

Будь-яка функція $f(x)$, яка залежить виключно від відстані між x і заданою точкою в просторі X , називається радіальною.

Тепер, коли ми представили метрику, давайте визначимо наші радіальні функції.

Формула для нормального розподілу (Гауса) з діагональною коваріаційною матрицею виглядає так:

$$p_j(x) = N_j e^{-\frac{1}{2p_j(x, \mu_j)}} \quad (2.5)$$

Значення щільності в точці x зростає зі зменшенням відстані до центру. У результаті можна розглядати щільність як функцію того, наскільки близько вектор до фіксованого центру.

Використовуючи методику ЕМ, навчання спрощено до реконструкції щільності для кожного класу. Центри дисперсії компонентів є результатом навчання.

Цікаво зауважити, що коли ми оцінюємо дисперсії, ми фактично обираємо метрику, яка буде застосована для визначення відстані між центрами.

Найбільш ефективну кількість компонентів у суміші визначають за допомогою методики для кожного класу.

2.1.4 Ансамблеві методи

Випадковий ліс. Для більшості завдань модель ансамблю випадкового лісу є добре відомою моделлю машинного навчання. Зараз він часто використовується, оскільки він генерує дані з високим ступенем точності та працює з різними типами даних, віддаючи перевагу моделям регресії, які більш чутливі до типу даних і зв'язків між ними. Декілька простих моделей, у даному випадку дерев рішень, об'єднані для підвищення ефективності алгоритму в Random Forest, який є переважно ансамблевою моделлю. Він заснований на техніці запуску, яка змішує набори випадково згенерованих моделей, сприяючи підвищенню загальної стабільності моделі, знижуючи дисперсію та запобігаючи перенавчанню.

У цій ситуації для роботи з деревами рішень використовується ряд критеріїв. У цій статті враховано критерій Джіні. Це статистичний критерій, який показує, наскільки віддалені один від одного шари різних груп. Його частіше

використовують як соціально-економічний барометр для відображення соціального розшарування.

Тепер буде продемонстровано алгоритм випадкового лісу.

Входи:

- а) n пояснювальних змінних з відповідними класами складають набір T ;
- б) набір функцій F ;
- в) розділення критерію C ;
- г) всього Z дерев.

Кінцевим результатом є ліс дерев Z . Ініціалізація `TreeSet (EmptySet)`.

Для i між $(0, Z)$:

- а) намалювати зразок X за допомогою T ;
- б) набір дерев $T (X, F, C)$;
- в) включіть T у `TreeSet`;
- г) повернути `TreeSet`; Кінець.

Через нелінійні та лінійні зв'язки індикаторів ринкової ціни акцій, які можуть ускладнити використання деяких моделей, як-от нейронних мереж тощо, наші дані будуть сильно залежати один від одного на основі аналізу часових рядів і призначення прогнозів.

Gradient boosting – це техніка машинного навчання на основі ансамблю, заснована на завантаженні. Він використовує дерева рішень як прості моделі, подібні до моделей випадкового лісу. Градієнтне підсилення ітераційно об'єднує слабких «навчань» в одного потужного. Прикладом цього є регресійна модель найменших квадратів, де середня квадратична помилка мінімізована, намагаючись «навчити» модель приймати значення у формі.

Слабкі елементарні моделі оцінюються, а потім покращуються на кожному ітераційному кроці шляхом додавання нових модифікованих моделей, параметри яких будуть вибрані за допомогою підходу градієнтного спуску. Вирішальною концепцією є забезпечення бажаних результатів для наступної моделі, щоб зменшити помилку. Ці модифікації моделей забезпечать швидку підгонку необхідних параметрів і хорошу точність моделі. Основною відмінністю між цією

методологією та іншими є включення як методу градієнтного спуску, так і традиційної стратегії посилення.

2.2 Основні характеристики фінансового часового ряду

Для розвитку фундаментальних можливостей роботи з моделями машинного навчання ми будемо використовувати фундаментальні технічні індикатори та прийоми технічного аналізу.

Технічні індикатори – це функції, які побудовані відповідно до статистичних індикаторів торгівлі (ціна, обсяг торгів тощо), аналіз подій яких покликаний визначити, чи змінюється поточний ринковий тренд. Технічні аналітики використовують перевірку технічних індикаторів, щоб вирішити, чи відкривати (збільшувати) чи закривати (зменшувати) позиції.

Ми враховуємо такі категорії технічних показників:

- а) індикатори трендів використовуються для підтримки трендів – осцилятори передвіщають зміни в тренді.
- б) канальні індикатори – будуть використані наступні технічні індикатори.

2.2.1 Williams %R

Williams %R – це технічний індикатор, який використовує положення поточної ціни закриття в діапазоні між максимумом і мінімумом попередніх періодів, щоб визначити, перекуплений чи перепроданий актив.

Різниця між поточною ціною закриття та найвищою ціною (максимум high) за попередні періоди, поділена на різницю між найвищою ціною (максимум high) і найнижчою ціною (низько low) за попередні періоди, дає значення індикатора Williams%R у кожній точці в час:

$$\%R_t = \frac{close_t - \max_{i \in [t-n, t]}(high_i)}{\max_{i \in [t-n, t]}(high_i) - \min_{i \in [t-n, t]}(low_i)} * 100 \quad (2.6)$$

де $\%R_t$ - значення індикатора Williams %R в момент часу t ;

low_i – найнижча ціна періоду i ;

$high$ – найвища ціна періоду i ;

$close$ – ціна закриття в момент t ;

100 – коефіцієнт приведення.

2.2.2 Індекс відносної сили

Індикатор технічного аналізу під назвою Relative Strength Index (RSI) оцінює силу тенденції та ймовірність того, що вона може змінитися.

Простота інтерпретації RSI надає йому широкого поширення. Шаблони технічного аналізу, такі як «голова і плечі», «верх» та інші, які може намалювати програма, можна оцінити разом із графіками цін.

RSI розраховується з використанням як позитивних (U), так і негативних (D) цінових рухів. Якщо сьогоднішня ціна закриття вища за вчорашню, кажуть, що день «зростає».

U і D дорівнюють 0, якщо вчорашня та сьогоднішня ціни закриття рівні. Після цього модифіковане експоненціальне ковзне середнє з періодом N використовується для згладжування значень U і D.

Таким чином, початковий розрахунок має так звану «відносну міцність». (англ. Relative Strength, RS):

$$RS = \frac{EMA[N] \text{ of } U}{EMA[N] \text{ of } D} \quad (2.7)$$

RS також використовується для розрахунку самого RSI:

$$RSI = 100 - \frac{100}{(1+RS)} \quad (2.8)$$

2.2.3 Стохастичний осцилятор

Індикатор для технічного аналізу, відомий як стохастичний осцилятор, відображає статус поточної ціни відносно діапазону цін протягом певного періоду часу в минулому. Він кількісно виражений у відсотках.

За словами Джорджа Лейна, творця індикатора, основний принцип полягає в тому, що ціна закриття наступного таймфрейму має тенденцію зупинитися поблизу попередніх максимумів у разі висхідного цінового тренду (зростаючого тренду). У разі тенденції до зниження ціни ціна закриття наступного інтервалу зазвичай зупиняється близько до попередніх мінімумів.

По суті, індикатор показує, як в межах параметрів зазначеного таймфрейму ціна закриття поточного періоду відрізняється від значень попередніх періодів.

Індикатор складається з двох ліній:

– %D – повільний стохастик (пунктирна лінія, додатково усереднений графік);

– %K – швидкий стохастик (суцільна лінія, основний графік).

Формула для розрахунку:

$$\%K_t = \frac{C_t - L_n}{H_n - L_n} * 100 \quad (2.9)$$

де C_t – ціна закриття поточного періоду,

H_n – найвища ціна за останні n періодів,

L_n – найнижча ціна за останні n періодів,

%D – ковзне середнє значення щодо %K з коротким часом усереднення.

Можна застосувати декілька методів усереднення (просте усереднення, експоненціальне, згладжене, зважене).

Нижче наведено найбільш типові пояснення для графіка стохастичного осцилятора:

а) купуйте, коли лінія графіка індикатора (%K або %D) спочатку опускається

нижче попередньо визначеного рівня (часто 20%), перш ніж піднятися вище нього. Продавайте, якщо лінія графіка індикатора спочатку перетинає певний поріг (часто 80%), перш ніж опуститися нижче нього;

- б) купуйте, коли лінія %K перетинає лінію %D. Якщо лінія %K перетинає лінію %D, продайте;
- в) визнайте розбіжності, наприклад, коли ціни досягають ряду нових максимумів, але стохастичний індикатор не може піднятися вище своїх попередніх максимумів, що вказує на те, що вам слід продати;
- г) коли індикатор зростає, перетин 80% сприймається як показник того, що ціна може припинити підйом або навіть почати різко падати. Коли індикатор падає, 20% перетин розглядається як натяк на те, що падіння, швидше за все, закінчилося або навіть почалося.

2.3 Методи оцінки ефективності роботи моделей

ROC AUC score. Показником продуктивності для проблем класифікації на різних порогових рівнях є крива AUC-ROC.

Крива ймовірності відома як ROC, а ступінь роздільної здатності відомий як AUC. Це показує, наскільки добре модель може розрізняти різні класи.

Модель прогнозує 0 класів проти 0 і 1 клас як 1 краще, коли AUC більша.

Співвідношення TPR до FPR, з TPR на осі Y і FPR на осі X, використовується для малювання кривої ROC.

TPR означає чутливість класифікатора. визначається співвідношенням $TP / (TP + FP)$, де FP означає хибно позитивні твердження, а TP представляє правильно визначені істинні твердження.

Співвідношення $TP / (TP + FN)$, де FN означає хибнонегативні результати, представляє специфічність.

Односпецифічна характеристика – FPR.

R-squared. Невизначеність залежної змінної виражається статистично R-квадратом, який визначається з набору незалежних змінних.

Ця метрика в основному вказує, наскільки сильно дисперсія різних змінних залежить одна від одної. Тобто дисперсія моделі буде повністю пояснена значенням R^2 , рівним 1.

Статистична метрика під назвою R-квадрат дозволяє нам працювати з ідеями оцінок невизначеності в конкретних даних.

Коли використовується для аналізу фінансових даних, оцінка R-квадрат зазвичай розуміється як зміни технічного індексу. При аналізі руху цінних паперів, наприклад, значення, також відоме як коефіцієнт детермінації, відображає швидкість зростання ціни залежно від руху індексу цін.

Ми можемо визначити, наскільки невизначеною є модель, що описує процес, дивлячись на її R-квадрати. Оцінка «один» означає, що модель повністю пояснює і що цільова змінна сильно залежить від залежних змінних. З іншого боку, нуль означає повну неоднозначність і той факт, що значення класифікації або прогнозу є абсолютно випадковим.

RMSE та MSE. Стандартними показниками того, наскільки добре моделі регресії передбачають значення цільових даних або змінних, є показники RMSE та MSE. Зазвичай вони використовуються для початкової оцінки методологій, хоча й не беруться до уваги.

Помилка, також відома як MSE, є середнім значенням кватилів розбіжності між очікуваними та фактичними результатами. У процесі функціонування моделі вона представляє функцію ризику, яка, у свою чергу, відображає очікувані збитки. Ми можемо стверджувати, що MSE характеризує дисперсію прогнозованих результатів, оскільки він характеризує стандартне відхилення оцінки. Обрана модель буде тим оптимальнішою, чим ближче значення до нуля. Значення пропонують інформацію як про дисперсію, так і про зміщення оцінки. має ті самі одиниці вимірювання, що й прогнозовані дані.

Квадратний корінь оцінки MSE відомий як RMSE. Це середня квадратична помилка, або, альтернативно, ми можемо сказати, що це стандартне відхилення середнього квадратичного відхилення. Іншими словами, він відображає відстань

між помилкою та правильним значенням або відстань від початкової точки до справжньої точки.

Цей тип помилки можна обчислити за допомогою різноманітних відстаней, а не лише евклідових. Також розглядаються перетворення за допомогою поліномів і афінів. Подібно до початкових цільових вхідних даних, MSE також має ті самі розміри оцінки.

Оскільки прогнозовані точки будуть лежати вздовж лінії результуючої регресії, коли стандартні записи використовуються як вхідні дані метрики, існує прямий зв'язок із коефіцієнтом кореляції.

Інформаційний критерій Акаїке. Інформаційний критерій Акаїке використовується лише під час вибору між кількома статистичними моделями. розроблений Хіроцугу Акаїке в 1971 році як «інформаційний критерій» і викладений ним у статті 1974 року.

Його було створено з метою оцінки якості прогнозів моделі на тестовій вибірці за умови модифікації моделі методом максимальної правдоподібності. Іншими словами, метою було розрахувати переобладнання моделі. Для ряду виняткових обставин Акаїке вдалося отримати відповідну оцінку за допомогою теорії інформації (включаючи відстань Кульбака-Лейблера).

$$AIC = \frac{2k}{n} + \ln \frac{SS_{res}}{n} \quad (2.10)$$

де k – кількість параметрів моделі.

Байєсівський інформаційний критерій (BIC). Шварц запропонував байєсівський інформаційний критерій (BIC) у 1978 році; в результаті його часто називають критеріями Шварца (SC). Це найпоширеніший варіант AIC, який базується на байєсівському підході:

$$BIC = SC = k \ln n - 2l \quad (2.11)$$

де L – є найбільшим значенням функції ймовірності моделі.

Моделі з нижчими значеннями BIC зазвичай вибирають, коли потрібен вибір серед багатьох моделей. BIC є як зростаючою функцією k , так і зростаючою функцією дисперсії помилки. Іншими словами, кількість пояснювальних змінних і непояснена зміна залежної змінної підвищують значення BIC. Однак нижчий BIC не завжди означає, що одна модель є кращою за іншу. BIC є просто евристикою, оскільки вона передбачає наближення. Відмінності в BIC, зокрема, ніколи не слід розглядати як модифіковані фактори Байєса.

Важливо пам'ятати, що BIC можна використовувати лише для порівняння розрахункових моделей за умови, що числові значення $[b]$ залежної змінної є узгодженими для всіх моделей. На відміну від порівняння моделей за допомогою F-критерію або тесту співвідношення правдоподібності, порівнювані моделі не потрібно вкладати.

MAE, MAPE. Одним із найбільш широко використовуваних показників продуктивності для вимірювання точності прогнозу є середня абсолютна відсоткова помилка (MAPE).

Індивідуальна абсолютна сума похибки, поділена на попит, відома як MAPE. Він представляє середню частоту помилок.

Прогноз MAPE KPI дуже фантастичний.

Незважаючи на низьку точність показника, він загальновідомий серед керівників компаній. Як ви можете бачити з розрахунку, MAPE ділить кожну помилку окремо на попит, роблячи його спотвореним: MAPE піддається значному впливу великих помилок у періоди низького попиту. У результаті оптимізація MAPE дасть нетиповий прогноз, який, ймовірно, є надто оптимістичним щодо попиту.

Чудовим КРІ для вимірювання точності прогнозу є середня абсолютна помилка (MAE). Його назва говорить про те, що це середнє значення абсолютних помилок.

Нездатність КРІ адаптуватися до середнього попиту є однією з його головних вад. Ви не можете судити про те, чи є щось відмінним чи поганим, якщо вам повідомляють, що MAE для цього конкретного товару дорівнює 10. Якщо ваш типовий попит становить 1000, це, очевидно, неймовірно. На жаль, ця точність є досить низькою, якщо середній попит дорівнює 1. Для вирішення цієї проблеми типово розділити MAE на середній попит, щоб отримати відсотки.

Висновки до розділу.

Часовий ряд біржових котирувань використовується для ілюстрації історичних фінансових даних. Ми створимо такі моделі за допомогою методів машинного навчання та інструментів для технічного аналізу:

- моделі машинного навчання, які прогнозують поведінку котирувань і вибирають найкращі інвестиції, використовуючи дані часових рядів і технічні індикатори як вхідні дані.

- модель Марковіца, яка вибере оптимальні інвестиції для портфеля інвестицій на основі інформації про ризики.

Щоб максимізувати значення цільової функції при мінімізації ризиків акції, ми вирішимо питання прогнозування та лінійного програмування.

3 АНАЛІЗ ОБЧИСЛЮВАЛЬНИХ ЕКСПЕРИМЕНТІВ ТА ОГЛЯД ПРОГРАМНОГО ПРОДУКТУ

3.1 Вхідні дані

Дані Agilent Technologies, Inc., Align Technology, Inc. і Amgen Inc. були використані в цьому дослідженні за п'ятирічний період.

Діаграма ціноутворення за 5 років для цих компаній показана на рисунку 3.1. На графіку та в інших місцях використовуються такі позначення:

- AMGN – Amgen Inc.
- ALGN – Align Technology, Inc.
- A – Agilent Technologies, Inc.



Рисунок 3.1 – Ціни акцій компаній за п'ятирічний період.

Для отримання даних було використано API-з'єднання бібліотеки Yahoo Finance.

Для розробки використовувалася мова та середовище програмування Python.

Найкращий портфель обирається за допомогою методів, які вимагають прогнозування як волатильності, так і доходності. Моделі, засновані на часових рядах, використовуються для прогнозування та моделювання доходів і

волатильності. Для прогнозування волатильності використовуються підходи типу ARMA та його варіанти. Для задачі прогнозування врожайності використовуються моделі гетероскедастичності типу GARCH та її варіації. Давайте спочатку розглянемо наш часовий ряд. На рисунку 3.1 показано, як виглядають часові ряди. Крім того, графіки функції автокореляції та частково функції автокореляції для даних кожної компанії включені на рисунки 3.2–3.4 для візуального огляду даних часових рядів.

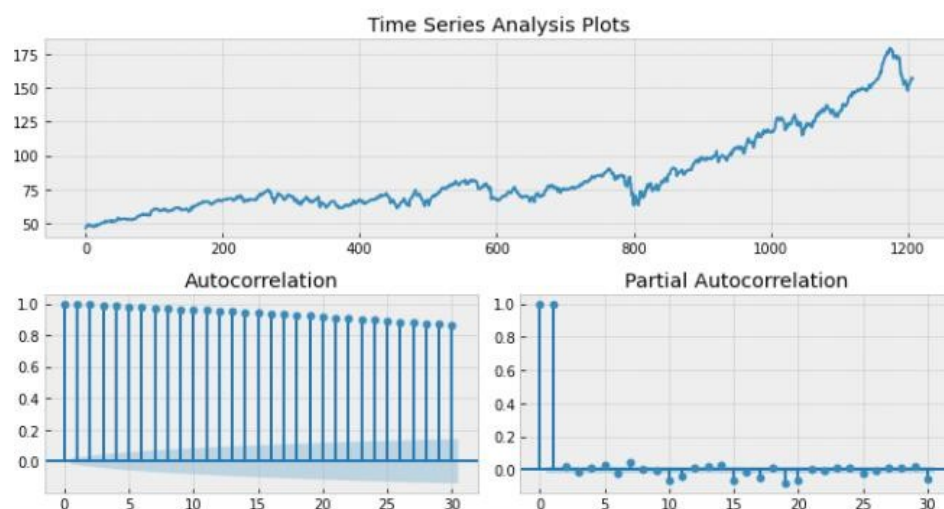


Рисунок 3.2 – Графіки ACF і PACF, часові ряди для Agilent Technologies, Inc.

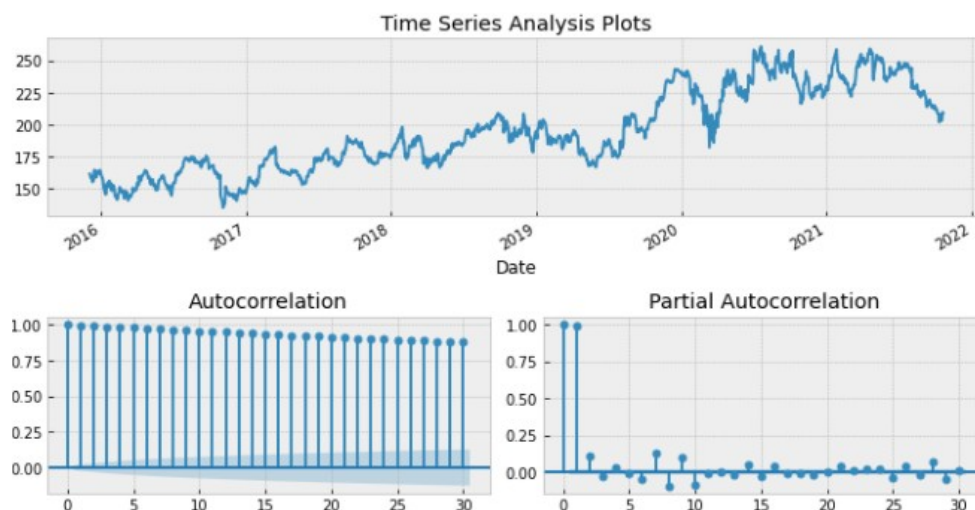


Рисунок 3.3 – Графіки ACF і PACF, часові ряди для Align Technology, Inc.

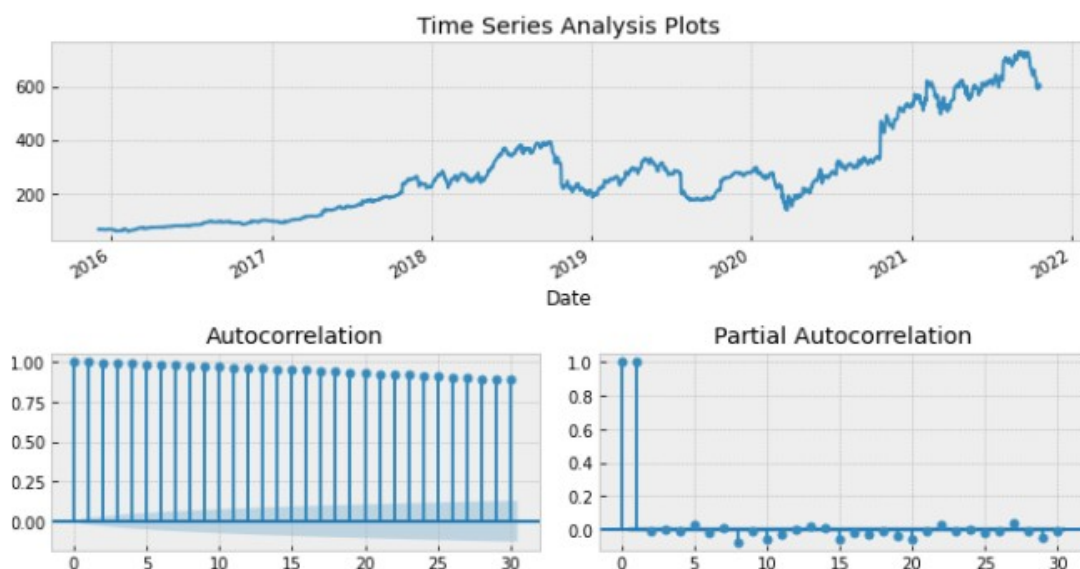


Рисунок 3.4 – Графіки ACF і PACF, часові ряди для Amgen Inc.

Можна зробити припущення щодо значних лагів для параметрів ковзної середньої та авторегресії для кожного з часових рядів. Ковзне середнє за ACF матиме значне відставання більше 30, а параметр авторегресії за ACF можна вважати рівним 2 для першого часового ряду. Відповідно до відповідних графіків ACF і PACF для кожної компанії, припущення будуть порівняти для другого і третього часового ряду.

Параметри ковзного середнього для кожного з трьох часових рядів можуть мати значення більше 30, тому оптимальна модель визначається шляхом порівняння моделей з різними значеннями параметрів.

Оскільки прогнозування волатильності є одним із етапів створення портфеля, важливо переконатися, що часові ряди є гетероскедастичними, тобто їх дисперсії змінюються з часом. Для цього використовується тест Енгла. Значення ймовірностей, розрахованих за цим критерієм, для всіх трьох сценаріїв становить 0, що не перевищує встановленого порогу 0,05. Таким чином, нульова гіпотеза про те, що дисперсія часового ряду є постійною протягом часу, спростовується.

3.2 Підбір моделей для прогнозування доходностей і волатильностей часових рядів

Давайте почнемо з роздумів про прогнозування прибутковості часових рядів. Було проведено дослідження, щоб визначити найкращі моделі прогнозування для кожного часового ряду для всіх трьох компаній.

Перед початком дослідження часові ряди для кожної з компаній спочатку були розкладені, а також оцінена наявність сезонності та тенденції.

Ми тестуватимемо моделі, які враховують сезонність, оскільки кожен часовий ряд має сезонну складову. Модель ARIMA враховує сезонність. Отже, давайте порівняємо моделі ARIMA з різними значеннями параметрів для кожного часового ряду, щоб побачити, яка з них найкраща. Давайте подумаємо про результати в таблиці 3.1.

Таблиця 3.1 – Оцінки моделі ARIMA компанії Agilent Technologies, Inc. для прогнозування прибутковості з різними параметрами.

	R2	THEIL	MAPE	MAE
ARIMA(1, 1, 4)	0,6145	0,9136	8,4535	2,63563
ARIMA(1, 1, 10)	0,7396	0,722	4,34134	0,63253
ARIMA(2, 1, 4)	0,8734	0,22603	1,71784	0,00631
ARIMA(2, 1, 10)	1,23436	5,4674	0,8925	0,6891
ARIMA(2, 1, 15)	0,01353	2,4631	0,41234	0,8014

Модель ARIMA(2, 1, 4) є найкращою моделлю для цього часового ряду, згідно з характеристиками таблиці 3.1.

Для компанії Agilent Technologies, Inc. на рисунку 3.5 показано візуалізацію справжньої та очікуваної прибутковості за допомогою підходу ARIMA(2, 1, 4). На рисунку 3.5 видно, що отримана модель робить досить точні прогнози.

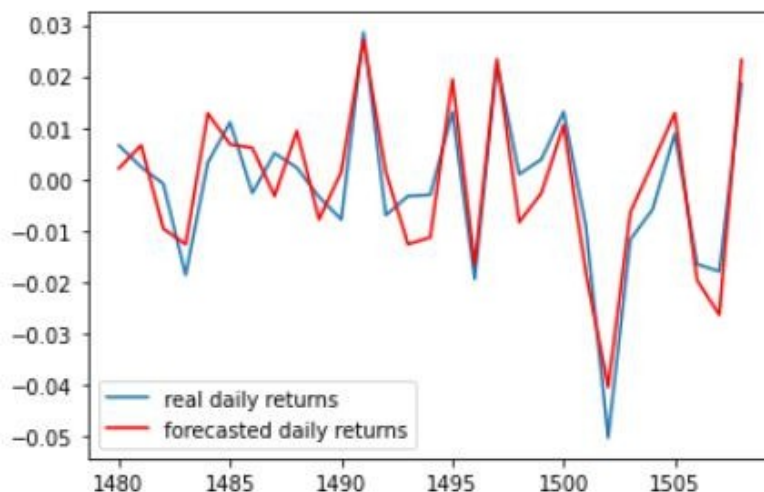


Рисунок 3.5 – Графік компанії Agilent Technologies, Inc. справжньої і спрогнозованої прибутковості методом ARIMA(2, 1, 4)

Висновки компанії Align Technologies, Inc. наведені в таблиці 3.2.

Таблиця 3.2 – Оцінки моделі ARIMA з використанням різних параметрів для прогнозування прибутковості Align Technology, Inc.

	R2	THEIL	MAPE	MAE
ARIMA(1, 1, 2)	0,7322	0,4645	6,4668	0,22463
ARIMA(1, 1, 8)	0,6413	0,5769	8,4345	0,83434
ARIMA(2, 1, 2)	0,8436	0,1786	3,2973	0,00738
ARIMA(2, 1, 8)	0,7447	0,2987	5,3465	0,01432
ARIMA(2, 1, 15)	0,5434	0,8322	10,427	1,63563

Модель ARIMA(2, 1, 2) є найбільш прийнятною для цього часового ряду, згідно з результатами, наведеними в таблиці 3.2.

Для компанії Align Technologies, Inc. на рисунку 3.6 показано графік істинної та очікуваної прибутковості з використанням підходу ARIMA(2, 1, 2). Графіки на рисунку 3.6 демонструють точність моделі, яка є досить високою.

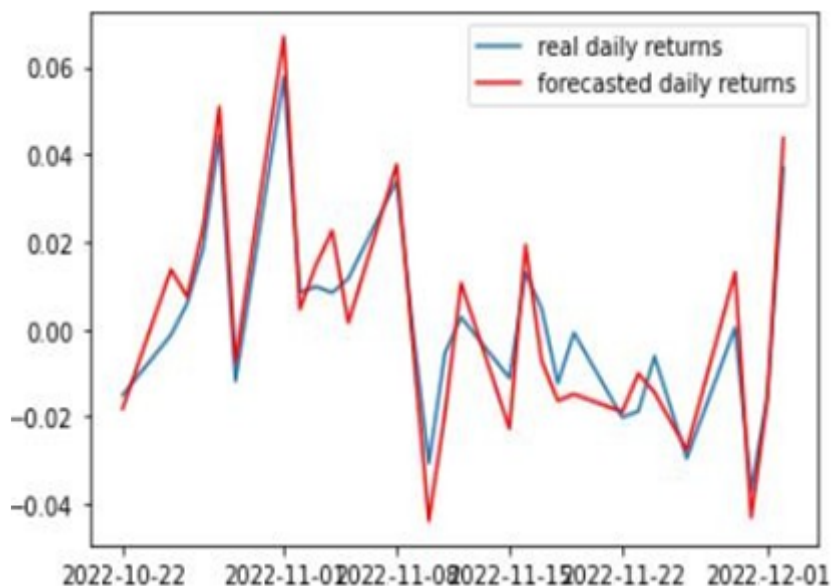


Рисунок 3.6 – Графік для компанії Align Technology, Inc. справжньої і спрогнозованої прибутковості методом ARIMA(2, 1, 2)

Розглянемо результати моделі ARIMA з використанням різних параметрів для прогнозування прибутковості Amgen Inc., які представлені в таблиці 3.3.

Таблиця 3.3 – Оцінки моделі ARIMA для оцінки прибутковості Amgen Inc. за різними параметрами

	R2	THEIL	MAPE	MAE
ARIMA(1, 1, 3)	0,8542	0,2456	0,9455	0,0148
ARIMA(1, 1, 8)	0,7145	0,5431	3,4517	0,9465
ARIMA(2, 1, 3)	0,9344	0,1336	0,5848	0,0026
ARIMA(2, 1, 8)	0,7825	0,3523	1,4666	0,0345
ARIMA(2, 1, 15)	0,6443	0,6345	4,565	1,9562

Результати в таблиці 3.3 дозволяють зробити висновок, що ARIMA є ідеальною моделлю для цього часового ряду (2, 1, 3). Крім того, на рисунку 3.7 показано графік прогнозів, зроблених за цією технікою, і оригінальні дані часових рядів для Amgen Inc.

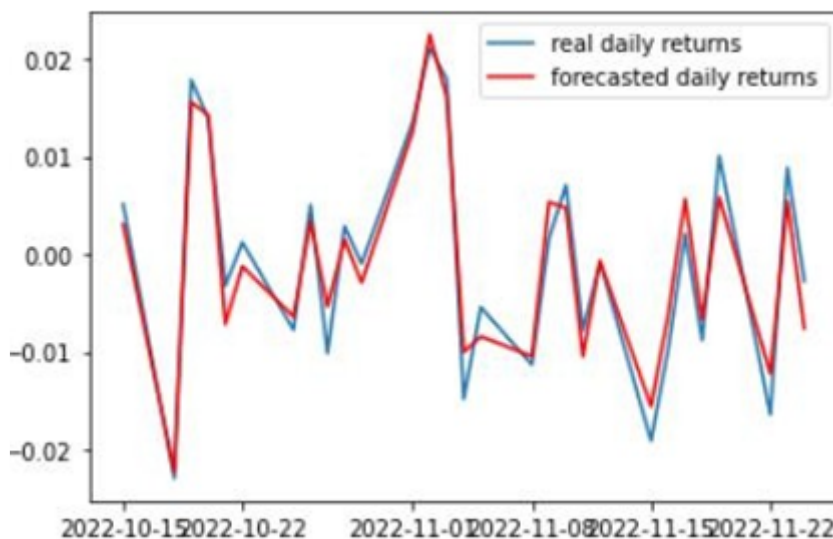


Рисунок 3.7 – Графік для компанії Amgen Inc. справжньої і спрогнозованої методом ARIMA(2, 1, 3)

Для корпоративних повернень було виявлено найкраще описані моделі. Давайте тепер подумаємо про проблему пошуку найкращих моделей для передбачення волатильності.

Оскільки всі три часові ряди є гетероскедастичними, як було зазначено в попередньому реченні, ми можемо використовувати гетероскедастичні моделі ARMA та їх варіації для пошуку прогнозів волатильності цих часових рядів.

Давайте визначимо важливі лаги для параметрів моделі кожної організації з графіків на рисунках 3.8-3.10.

ACF і PACF для Agilent Technologies, Inc. часові ряди залишків показані на рисунку 3.8.

Сценарій схожий на сценарій прогнозування доходності, оскільки параметр для ковзного середнього може дорівнювати 1, але параметр для авторегресії може бути більше 12. Іншими словами, щоб визначити найкраще рішення, ви повинні зважити свої варіанти.

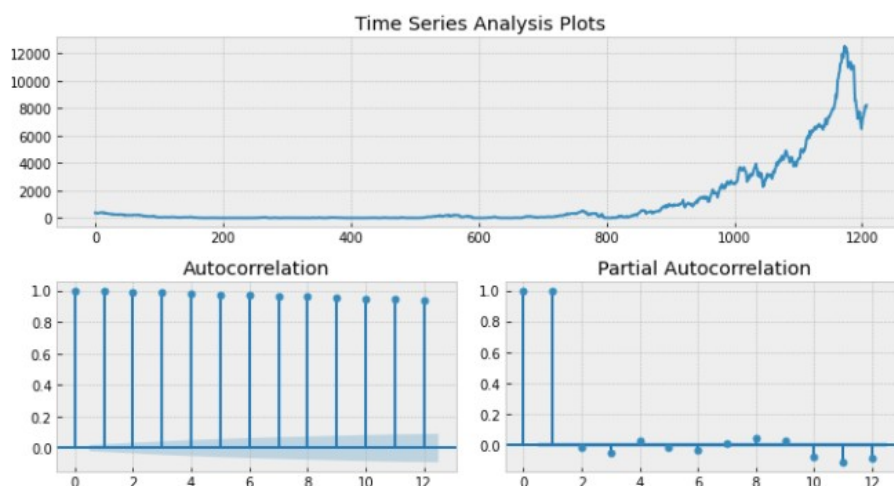


Рисунок 3.8 – Графіки ACF і PACF компанії Agilent Technologies, Inc. для квадратів залишків акцій

На рисунку 3.9 показано ACF і PACF часові ряди залишків Align Technology, Inc.

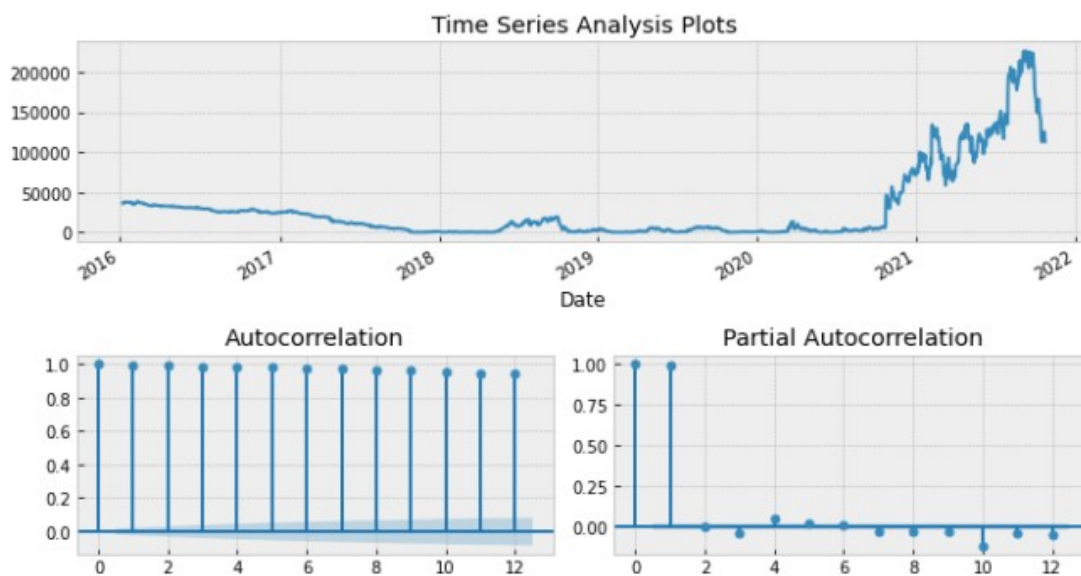


Рисунок 3.9 – Графіки ACF і PACF компанії Align Technology, Inc. квадратів залишків часового ряду

Обставини такі ж, як і для залишкових квадратів часових рядів від Agilent Technologies, Inc. Параметри необхідно змінити, оскільки параметр ковзного середнього дорівнює 1, а значення авторегресії може бути будь-яким числом, більшим за 12.

Діаграми ACF і PACF для квадратів залишків часових рядів компанії Amgen Inc. відображені на рисунку 3.10.

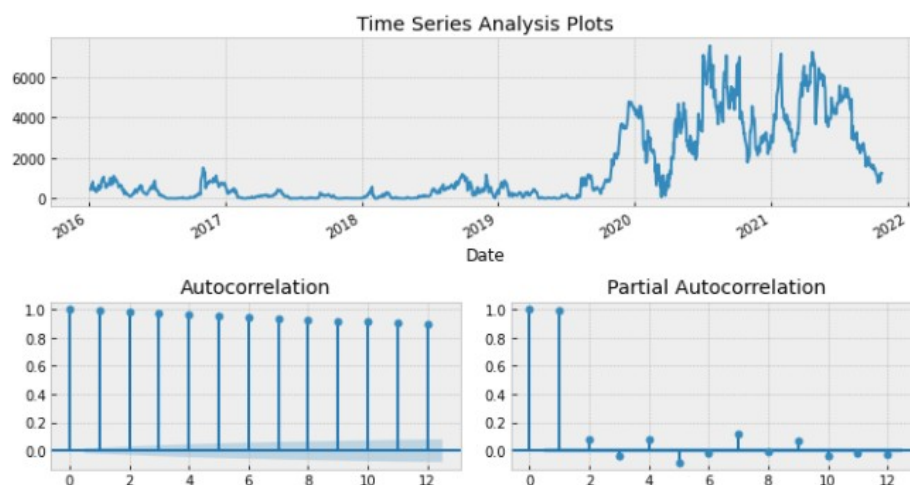


Рисунок 3.10 – Графіки ACF і PACF компанії Amgen Inc. для квадратів залишків часового ряду

Поточний екземпляр можна порівняти з двома попередніми. Іншими словами, ми вибираємо параметр для авторегресії та встановлюємо параметр 1 для ковзного середнього для цього часового ряду.

Щоб вибрати найкращі гетероскедастичні моделі для оцінки волатильності, ми використовуємо моделі GARCH і TARЧН.

Результати моделей GARCH і TARЧН з різними параметрами для часових рядів усіх трьох підприємств наведено в таблицях 3.4–3.6.

Таблиця 3.4 – Характеристики моделей GARCH і TARЧН для компанії Agilent Technologies, Inc. з різними параметрами

	R2	THEIL	MAPE	MAE
GARCH(1, 1)	0,6223	0,6735	1,8834	3,5718
TARCH(1, 1, 1)	0,7531	0,4576	0,7653	1,5785
GARCH(13, 1)	0,8316	0,1752	0,2457	0,7523
TARCH(13, 1, 1)	0,9222	0,0726	0,1912	0,3267
GARCH(20, 1)	0,6453	0,6223	1,5465	2,5863
TARCH(20, 1, 1)	0,7813	0,2566	0,5726	1,2784

Згідно з результатами таблиці 3.4, модель TARЧН(13, 1, 1) є найефективнішою моделлю для прогнозування волатильності цього часового ряду.

Таблиця 3.5 – Характеристики моделей GARCH і TARЧН для компанії Align Technology, Inc. з різними параметрами

	R2	THEIL	MAPE	MAE
GARCH(1, 1)	0,6422	0,6422	1,4564	8,5462
TARЧН(1, 1, 1)	0,7236	0,3255	0,7463	4,4694
GARCH(10, 1)	0,8344	0,1355	0,2451	3,5478
TARЧН(10, 1, 1)	0,9093	0,0667	0,1749	2,6838
GARCH(20, 1)	0,6521	0,4143	0,9423	6,2455
TARЧН(20, 1, 1)	0,7452	0,3566	0,8644	4,9864

Ми можемо сказати, що TARЧН(10, 1, 1) є найбільш підходящою моделлю для цієї ситуації на основі особливостей моделей у таблиці 3.5.

Таблиця 3.6 – Характеристики моделей GARCH і TARЧН для компанії Amgen Inc. з різними параметрами

	R2	THEIL	MAPE	MAE
GARCH(1, 1)	0,5731	0,8475	4,9951	2,7246
TARЧН(1, 1, 1)	0,7415	0,4713	1,8523	0,0363
GARCH(14, 1)	0,8418	0,3571	0,9562	0,0146
TARЧН(14, 1, 1)	0,9124	0,1475	0,8704	0,002
GARCH(20, 1)	0,6342	0,7258	4,7522	1,4378
TARЧН(20, 1, 1)	0,7145	0,4974	2,6352	0,761

Властивості моделей GARCH і TARЧН з різними параметрами для Amgen Inc. відображені в таблиці 3.6. Ця таблиця показує, що третє завдання прогнозування волатильності часових рядів може бути виконане за допомогою моделі TARЧН(14, 1, 1).

Моделі TARЧН продемонстрували найкращі результати в кожному з трьох випадків прогнозування волатильності для трьох компаній.

Графіки справжнього та прогнозованого TARЧН рівнів волатильності для відповідних компаній показані на рисунках 3.11–3.13.

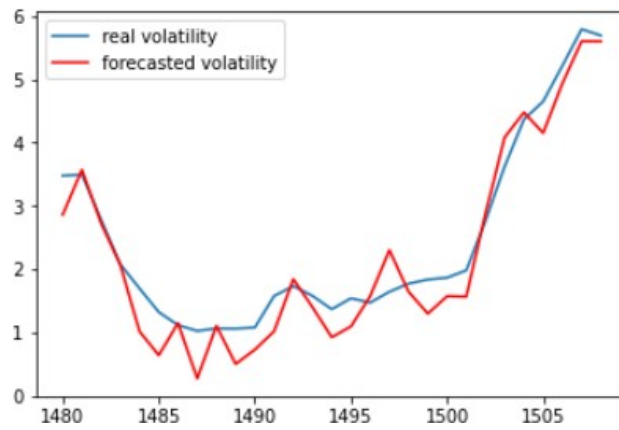


Рисунок 3.11 – Графік спрогнозованої волатильності для компанії Agilent Technologies, Inc. методом TARЧ(13, 1,1)

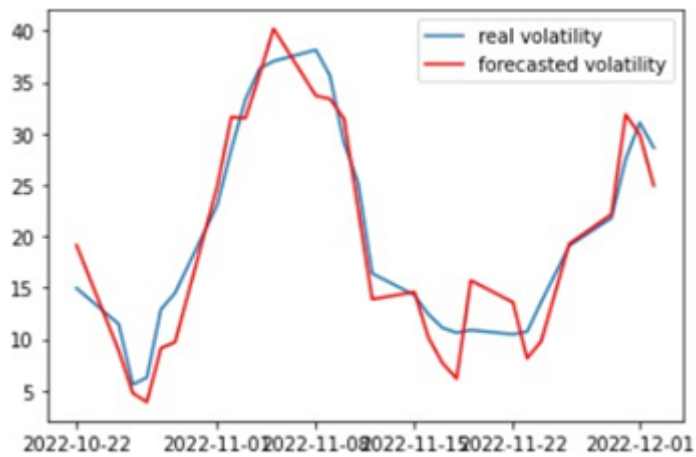


Рисунок 3.12 – Графік спрогнозованої волатильності для компанії Align Technology, Inc. методом TARЧ(10, 1,1)

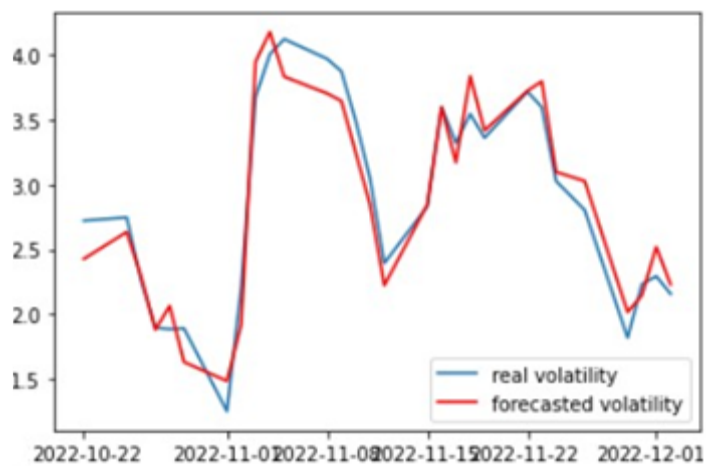


Рисунок 3.13 – Графік спрогнозованої волатильності для компанії Amgen Inc. методом TARЧ(14, 1,1)

Щоб передбачити прибутковість часових рядів і волатильність часових рядів трьох різних компаній, були розроблені моделі. Потім на основі цих прогнозів буде створено найкращий портфель акцій.

3.3 Побудова і прогнози щотижневих трендів методами машинного навчання

Щоб уважніше спостерігати за поведінкою ринку і краще розуміти поведінку ринку, було вирішено побудувати тижневі тренди.

Наведені нижче дані використовувалися для створення тестових і базових даних під час роботи з тенденціями:

- High;
- Low;
- Open;
- Close Adj;
- Close;
- Volume;
- n_low;
- RSI_21;
- %K;
- %D;
- STOCHk_14_3_3;
- STOCHd_14_3_3;
- CUMPCTRET_1;
- CUMLOGRET_1;

Це розраховані технічні показники та біржові дані.

В якості індикатора зміни тренду за попередній тиждень обрано індикатор tr_sgn . Було визначено, що це знак $sign(t_i - t_{i-7})$. Результатом став часовий ряд індикаторів ціни закриття (Close) за останній період.

Для цієї мети були обрані такі моделі, як випадковий ліс, градієнтне групування та багатошарові нейронні мережі з шарами, що містять радіальні базисні функції.

Ці моделі були обрані, оскільки вони найбільш ефективно обробляли фінансові дані.

Перед тестуванням моделей на двох екземплярах даних було завершено роботу з нормалізації даних.

Розглянемо, як дані на тепловій карті на рисунку 3.14 співвідносяться між собою.

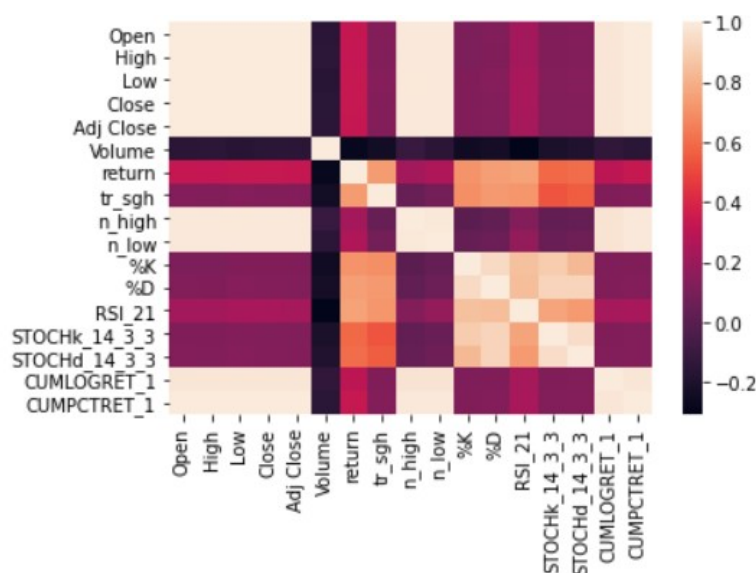


Рисунок 3.14 – Теплограма коефіцієнта кореляції Пірсона

Хоча нейронна мережа була дуже чутливою до трансформації та мала низьку продуктивність на ненормалізованих даних через різну розмірність даних, із сигналами проміжних результатів, які застрягли в точках локальних екстремумів, ансамблі показали менш значні відмінності в продуктивності нормалізованих і ненормалізованих наборів даних. Ситуація була покращена завдяки роботі над архітектурою та перетвореннями даних, але результати прогнозів залишалися нерівними і, швидше за все, не могли бути покращені шляхом вибору багатошарового перцептрона як основної моделі.

Оскільки розподіл цільових даних був нерівномірним, була також проведена робота, щоб збалансувати вхідні ваги для моделі випадкового лісу ансамблю. Таким чином було досягнуто найвищого рівня точності моделі.

Найкраща продуктивність моделі ROC AUC (з використанням ненормалізованих даних) наведена в таблиці 3.7.

Таблиця 3.7 – Результати моделі на ненормованих даних.

	NN	XGboost	RandomForest
Дані 1 компанії	0	0.77	0.79
Дані 2 компанії	0	0.88	0.85
Дані 3 компанії	0	0.95	0.88

Таблиця 3.8 підсумовує найкращі результати моделей ROC AUC (нормалізовані дані):

Таблиця 3.8 – Результати моделі на нормованих даних.

	NN	XGboost	Random Forest
Дані 1 компанії	0,34	0.89	0.884
Дані 2 компанії	0,26	0.846	0.917
Дані 3 компанії	0,63	0.95	0.93

Як ми бачимо, нормалізовані дані зазвичай дають кращі результати для даних.

На відміну від даних нейронної мережі, які вказують на те, що обраний тип нейронної мережі, ймовірно, не підходить для даного завдання, Random Forest і XGboost забезпечили приблизно рівні точності.

Хоча цей тип нейронної мережі використовується для завдань класифікації та регресії фінансових даних, особливо з шарами, що містять функції радіального базису, вибрані спеціально для роботи з фінансовими часовими рядами, ймовірно, що висока кореляція даних один з одним не дозволяла для покращення результатів навіть шляхом рекомбінації шарів нейронної мережі та підгонки параметрів.

У світлі результатів ми оптимізуємо наш майбутній портфель, використовуючи результати, передбачені моделлю Random Forest.

3.4 Модель Марковіца для побудови оптимального портфелю акцій

Ми розглянемо типову проблему оптимізації портфелю. Задача лінійної оптимізації може бути використана для реалізації моделі Марковіца.

Цільовими функціями, які беруться до уваги, є максимізація функції повернення та мінімізація загального ризику ринкового портфелю акцій.

Нам потрібно буде обмежити рівень ризику в обох проблемах.

Ми розглянемо проблему визначення найкращих вагових коефіцієнтів для кожного бізнесу, які відображають пропорцію акцій у кожному портфелі, і ми розглянемо альтернативу, коли всі акції розглядаються як безризикові та, отже, включені до портфелю з частка 0,333 ($1/n$ у нашому випадку дорівнюватиме $1/3$, оскільки ми беремо до уваги порцію з 3 часток)

Ми почнемо з розгляду традиційного сценарію, коли ми повинні з'ясувати, як зменшити загальний ризик портфелю, який визначається порівнянням ризиків кожної окремої акції. Оптимізація за критерієм Шарпа є одним із них. На рисунку 3.15 нижче показано теплову карту прибутковості кореляції вибраних акцій одна з одною. Друга і третя акції мають більшу асоціацію, як можна помітити.

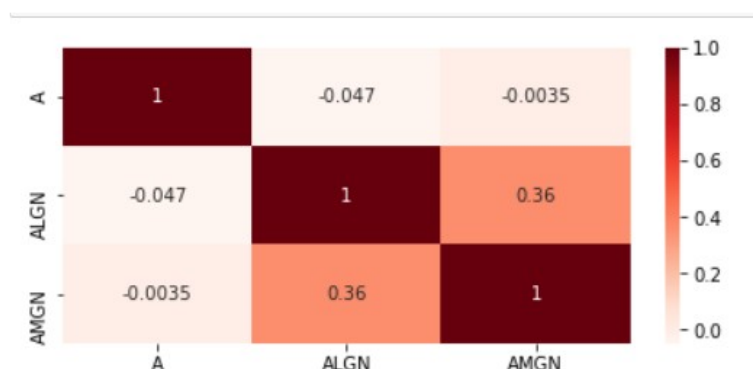


Рисунок 3.15 – Теплова карта кореляція між даними залишків компаній

Потім, використовуючи однакове зважування всіх акцій у так званому безризиковому портфелі, ми обчислимо графік дохідності.

Цільова функція максимізації прибутку при стабільному значенні волатильності буде використана для створення третього портфеля.

Зазначені вище методи аналізу часових рядів також використовувалися для прогнозування волатильності вибраних акцій протягом прогнозованого періоду.

Отже, прогнозовані дані будуть єдиним входом для задачі лінійної оптимізації.

Для пошуку оптимальних значень була використана функція мінімізації з пакету `scipy`, яка використовує техніку пошуку значень `SLSQP`, метод найменших квадратів для знаходження оптимального значення.

Оскільки максимальний випуск спочатку був обернений, на виході буде вектор трьох ідеальних ваг для параметрів разом із оберненим оптимальним значенням вибраної функції.

Подивіться на останній графік на рисунку 3.16, який показує три різні типи портфоліо.

`Prognozed` – портфель, отриманий з використанням прогнозованих даних на обраному рівні загальної волатильності портфеля) (портфель, отриманий на основі прогнозованих даних на обраному рівні загальної волатильності портфеля).

`No_risks` – портфель, в якому акції є безризиковими; у нашому прикладі передбачається, що розподіл портфеля дорівнює $1/n = 1/3$.

`Classical_sharp` – мінімізує параметр оцінки критерію Шарпа.

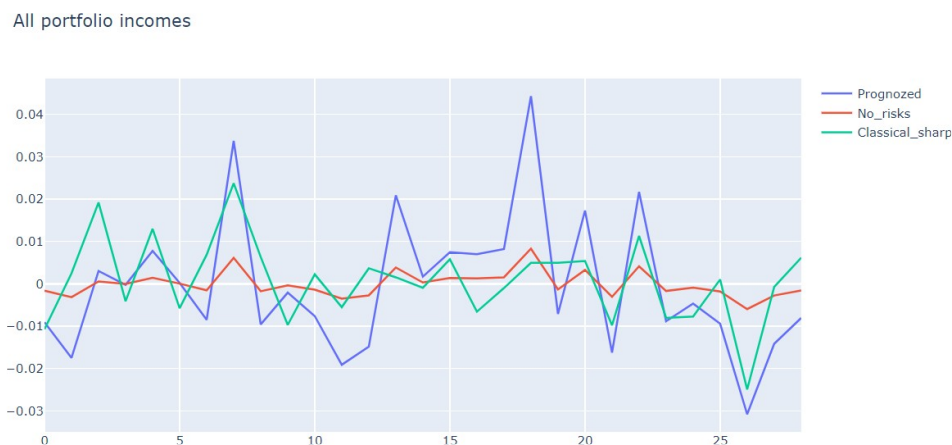


Рисунок 3.16 – Графік розподілу прибутковості портфелів

Графік показує, що прибутковість цих портфельів буде коливатися з різною амплітудою.

Зрозуміло, що очікуваний портфель матиме більші піки доходності, що покращить результати короткострокової стратегії продажу. З іншого боку, модель, заснована на коефіцієнті Шарпа, матиме менший спред і, як наслідок, менші потенційні прибутки з меншим ризиком.

Модель, що використовує акції, які вважаються умовно безризиковими та мають однакову вагу, матиме нижчі рівні прибутків і збитків.

Загалом можна зробити висновок, що стратегія максимізації прибутку, дотримуючись обмежень на толерантність до ризику, буде більш розумною для даної інвестиції.

Висновки до розділу

Проаналізувавши часові ряди, що містять фінансові дані, і застосувавши їх для оптимізації портфеля, ми можемо стверджувати наступне.

Спочатку були проведені обстеження та дослідження для визначення сезонних залежностей при дослідженні стаціонарності. Для прогнозування щоденних балансів використовувалася модель ARIMA з незалежно обраними параметрами. TARCN був обраний для прогнозування волатильності (фінансового ризику). Діаграми ACF і PACF перевірялися для вибору параметрів, але через складність цього підходу було вирішено покладатися на методологію для визначення найкращих затримок для параметрів, одночасно тестуючи моделі для інших пов'язаних параметрів. Завдяки цьому процесу ми змогли отримати ряд прогнозованих щоденних залишків для кожного з трьох бізнесів, які ми вибрали, а також ряд прогнозів волатильності на наступні 30 робочих днів.

Після цього ми переглянули дані наприкінці фінансового дня та знайшли додаткові характеристики, такі як технічний індекс RSI, осцилятори та технічний індекс Williams R. Додатковим визначенням цілі був напрямок зміни доходності за тиждень для кожного запису, що відображає поточну тенденцію. Для вибраних

моделей дані були збалансовані та нормалізовані для підвищення точності. Прогнозування виявило тенденції на наступні 30 робочих днів.

Таким чином, помноживши щоденні залишки на нашу тенденцію, ми досягли очікуваної прибутковості. Після отримання висновків ми можемо приступати до оптимізації портфеля. Після порівняння трьох методів, які були обрані з використанням рівномірно розподілених ваг, критерію Шарпа та прогнозованих даних із фіксованим ступенем ризику, можна сказати, що, враховуючи використані вхідні дані, модель машинного навчання є найприбутковішою та найризикованішою, модель Шарпа менш ризикована і її слід враховувати для більш обережних інвестицій у майбутньому, але модель з рівномірним розподілом ваг є найбільш прибутковою та безризиковою.

ВИСНОВКИ

Метою цього дослідження було краще зрозуміти складність створення фінансово-економічних прогнозів. Наведено стратегії вирішення проблеми оптимізації портфоліо та порівняння з методами машинного навчання.

Для цього завдання ми використали Python і його бібліотеки машинного навчання та візуалізації даних, зокрема Sklearn, Matplotlib, Pandas, Numpy та TensorFlow. Робочим середовищем був Jupyter Notebook.

Було взято до уваги три випадки, і можна зробити висновок, що модель, створена за допомогою методів машинного навчання, є як найприбутковішою, так і найризикованішою з використанням вхідних даних. Ця модель мала рівномірно розподілені ваги, вибрані за допомогою критерію Шарпа та зіставлені на основі прогнозованих даних із фіксованим ступенем ризику. Навпаки, рівномірно розподілена модель показала низький ризик і низьку прибутковість, але більш стабільну, ніж інші моделі, і, отже, кращу. Це можна взяти до уваги для більш обережних майбутніх інвестицій.

Для презентації було розроблено програмний продукт на основі Python, який можна швидко інтегрувати як у веб-сервіс, так і в бот Telegram.

Стартап-проект розглядався в останньому розділі та над ним працювали. Через низький рівень конкуренції та зручності стартапу було зроблено висновок, що він може закріпитися на ринку, що зробить його привабливим для інвесторів.

В майбутньому можна провести додаткові дослідження щодо додаткових активів для оптимізації портфеля та одночасної оптимізації кількох портфелів. А також використання абсолютно нових, покращених статистичних методів, створених у майбутньому.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Stock Portfolio Optimization Using a Combined Approach of Multi Objective Grey Wolf Optimizer and Machine Learning Preselection Methods / URL: <https://www.hindawi.com/journals/cin/2022/5974842/> (дата звернення: 20.03.2023).
2. Richard O. Michaud, Robert O. Michaud. Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation (Financial Management Association Survey and Synthesis / Publisher : Oxford University Press, 2008 – 144 p.
3. Steven Achelis. Technical Analysis from A to Z, 2nd Edition / Publisher : McGraw Hill, 2013 – 400 p.
4. Поважний О. С, Орлова Н. С, Свечкіна А. Л. Цінні папери і фондовий ринок: Навч. посібник / Львів: Магнолія, 2010. – 361 с.
5. David F. Swensen. Pioneering Portfolio Management: An Unconventional Approach to Institutional Investment, Fully Revised and Updated / Publisher : Free Press, 2009 – 432 p.
6. What's Wrong with Multiplying by the Square Root of Twelve / URL: <https://studylib.net/doc/7921569/what-s-wrong-with-multiplying-by-the-square-root-of> (дата звернення: 21.03.2023).
7. Jack Johnston, John Dinardo. Econometric methods / Publisher : McGraw-Hill/Irwin, 1996 – 531 p.
8. Ринок цінних паперів / URL; <https://core.ac.uk/download/84826115.pdf> (дата звернення: 22.03.2023).
9. Сучасний стан та перспективи розвитку ринку похідних цінних паперів в Україні / https://economyandsociety.in.ua/journals/13_ukr/212.pdf (дата звернення: 23.03.2023).
10. Modern Portfolio Theory: What MPT Is and How Investors Use It / URL: <https://www.investopedia.com/terms/m/modernportfoliotheory.asp> (дата звернення: 24.03.2023).

11. Бідюк П. І. Часові ряди: моделювання і прогнозування / Київ: ЕКМО, 2003 – 144 с.
12. Луців Б. Л. Банківська дальність у сфері інвестицій / Тернопіль: Карбланш, 2001 – 320 с.
13. Time Series Forecasting: Definition, Applications, and Examples / URL: <https://www.tableau.com/learn/articles/time-series-forecasting> (дата звернення: 24.03.2023).
14. Structural analysis of the methods of regression model construction by time series of observations / URL: https://www.researchgate.net/publication/296271231_Structural_analysis_of_the_methods_of_regression_model_construction_by_time_series_of_observations (дата звернення: 25.03.2023).
15. Mark J. L. Orr Introduction to Radial Basis Function Networks / Edinburgh: Centre for Cognitive Science, 1996 – 67 p.
16. Simon Haykin. Neural Networks and Learning Machines / Publisher : Pearson, 2008 – 936 p.
17. A Random Forest Guided Tour. URL: <https://erwanscornet.github.io/pdf/reviewforest.pdf> (дата звернення: 25.03.2023).
18. Stochastic Oscillator: What It Is, How It Works, How To Calculate / URL: <https://www.investopedia.com/terms/s/stochasticoscillator.asp> (дата звернення: 26.03.2023).
19. William Brock, Josef Lakonishok, Blake LeBaron, Simple Technical Trading Rules and the Stochastic Properties of Stock Returns / New York: S&P Global, 2017 – 64 p.
20. Leo Breiman. Random Forests / Berkeley: Statistics Department, University of California, 2001 – 33 p.
21. Informed trading, information uncertainty, and price momentum / URL: <https://www.sciencedirect.com/science/article/abs/pii/S0378426612000787> (дата звернення: 26.03.2023).

22. S.Basak, Khaidemn L., Saha S., Kar S. Predicting the direction of stock market prices using random forests / New-York: Wrox Press, 2016 – 178 p.

23. 11 Important Model Evaluation Techniques Everyone Should Know / <https://www.datasciencecentral.com/7-important-model-evaluation-error-metrics-everyone-should-know/> (дата звернення: 27.03.2023).

24. Order selection for same-realization predictions in autoregressive processes / URL: <https://projecteuclid.org/journals/annals-of-statistics/volume-33/issue-5/Order-selection-for-same-realization-predictions-in-autoregressive-processes/10.1214/009053605000000525.full> (дата звернення: 28.03.2023).

25. Smelyakov, K., Honchar, Y., Bohomolov, O., Chupryna, A. Machine Learning Models Efficiency Analysis for Image Classification Problem // CEUR Workshop Proceedings, 2022, 3171, pp. 942–959.

26. Nazarenko, D., Afanasieva, I., Golian, N., Golian, V. Investigation of the deep learning approaches to classify emotions in texts // CEUR Workshop Proceedings, 2021, 2870, pp. 206–224.

27. Sharonova, N., Kyrychenko, I., Tereshchenko, G. Application of Big Data methods in E-learning systems // 5th International Conference on Computational Linguistics and In-telligent Systems (COLINS-2021), Kharkiv, Ukraine, April 22-23, 2021. – CEUR Workshop Proceedings, 2021, 2870, Volume I, pp. 1302-1311.