**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**ІМЕНІ ТАРАСА ШЕВЧЕНКА**

**Економічний факультет**

**Кафедра економічної кібернетики**

****

**КУРСОВА РОБОТА**

**з теми**

**Створення системи управління інвестиційним портфелем на основі машинного навчання**

Виконав студент 2-го курсу групи ЕК-2

Христік Павло Вікторович

Науковий керівник: д.е.н., професор Ставицький Андрій Володимирович

Київ-2023

Зміст

[ВСТУП 3](#_Toc135673199)

[1 ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ОСНОВИ УПРАВЛІННЯ ІНВЕСТИЦІЙНИМ ПОРТФЕЛЕМ 4](#_Toc135673200)

[**1.1. Фондовий ринок та принципи його роботи** 4](#_Toc135673201)

[**1.2. Поняття інвестиційного портфелю** 5](#_Toc135673202)

[**1.3. Принципи створення інвестиційного портфелю** 7](#_Toc135673203)

[2 МАШИННЕ НАВЧАННЯ ПРИ АНАЛІЗІ ФОНДОВИХ РИНКІВ 11](#_Toc135673204)

[**2.1. Основні методи машинного навчання** 11](#_Toc135673205)

[**2.2. Вибір моделей для аналізу.** 14](#_Toc135673206)

[**2.3. Порівняння якості роботи моделей.** 16](#_Toc135673207)

[3 РОЗРОБКА МОДЕЛІ ОПТИМАЛЬНОГО ІНВЕСТИЦІЙНОГО ПОРТФЕЛЮ 19](#_Toc135673208)

[**3.1. Вибір даних та навчання моделі** 19](#_Toc135673209)

[**3.2. Динамічне управління портфелем на прикладі ІС «Моно Інвест»** 22](#_Toc135673210)

[**3.3. Розробка системи динамічного управління портфелем.** 23](#_Toc135673211)

[ВИСНОВКИ 26](#_Toc135673212)

[СПИСОК ДЖЕРЕЛ: 27](#_Toc135673213)

ВСТУП

У наш час інвестиції це один із способів швидкого заробітку. Інвестування потребує все більш точного прогнозу та аналізу, що саме буде з акціями у наступні тижні, місяці а то й роки. Дана галузь є дуже швидкорозвиваюча, та займає все більше й більше місця у наших життях. Саме тому машинне навчання та його методи дуже сильно допомагають з цим завданням.

Актуальність теми дослідження. В умовах постійного прогресу, та зміни загальної ситуації на фондовому ринку, прийняття рішення про те, які акції потрібно обирати для оптимального інвестування, стає дедалі складнішим. Створення моделі на основі машинного навчання повинна вирішити це питання, і допомогти користувачам у вирішенні цього питання.

Метою даної курсової є дослідження акцій, їх загальний аналіз, та використання машинного навчання у створенні моделі, яка допоможе для користувача програми обрати оптимальний інвестиційний портфель у відповідності до його бажань.

Завдання дослідження:

* Аналіз фондового ринку, ознайомлення з різновидами інвестиційних портфелів, способами взаємодій акцій одна з одною.
* З’ясувати, як саме створюється інвестиційний портфель.
* Вибір моделей машинного навчання для дослідження.
* З’ясування основних метрик оцінки моделей машинного навчання.
* Створення оптимального інвестиційного портфелю за результатами моделі.

Ключовим аспектом даної роботи є аналіз тенденцій розвитку відповідних акцій за допомогою машинного навчання, розбивка цих акцій у різні кластери, що допоможе для різних користувачів, як нових у цій сфері, так і більш досвідчених, краще зорієнтуватись, куди можна інвестувати свої кошти.

Об'єкт дослідження: система управління інвестиційним портфелем, та фондовий ринок.

Предмет дослідження: розробка та застосування методів машинного навчання для побудови системи управління інвестиційним портфелем.

Методи дослідження: машинне навчання, а саме кластеризаційні моделі.

Наукова та/або практична новизна роботи: наукова новизна даної роботи полягає в застосуванні методів машинного навчання саме у такій сфері. Дослідження буде спрямоване на створення оптимального розподілу капіталу для своїх користувачів.

Структура роботи: дана робота складається з вступу, трьох основних розділів, висновку та списку використаних джерел.

1. ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ОСНОВИ УПРАВЛІННЯ ІНВЕСТИЦІЙНИМ ПОРТФЕЛЕМ

**1.1. Фондовий ринок та принципи його роботи**

Фондовий ринок, або ринок цінних паперів, - це сегмент ринку капіталів, де здійснюються емісія, розміщення, купівля та продаж цінних паперів.

Для будь-якої країни, що прагне досягти економічного зростання, важливим етапом економічних перетворень є створення сучасної інфраструктури, ефективної інвестиційної системи та розвитку фондового ринку.

Для визначення фондового ринку також варто додати, хто саме бере участь у операціях в межах фондового ринку.

Учасниками фондового ринку є: [1]

- біржові брокери, які діють як посередники між фондовими біржами та інвесторами, купуючи та продаючи акції;

- інвестори;

- інвестиційні банкіри, які представляють приватні компанії.

У фондовому ринку основну роль відіграють самі цінні папери, над якими і відбуваються всі операції. Серед них:

* облігації
* ф’ючерси
* акції
* опціони
* паї біржових фондів, тощо.

Основна суть роботи фондового ринку заключається в тому, що в ньому суб’єкти здійснюють обмін, купівлю та інші фінансові операції з цінними паперами. Робота фондового ринку неможлива без зовнішньої регуляції, в Україні, наприклад, цим займається Національна комісія з цінних паперів та фондового ринку.[2]

Саме для організації різних процесів в середовищі фінансових операцій й існує фондовий ринок. У даних організацій, які керують всіма процесами, що відбуваються всередині фондового ринку, та регулюють його, за мету стоїть правильне створення та дотримання сурових правил щодо операцій, які відбуваються на фондовому ринку.[3]

У роботі фондового ринку існують фактори, які впливають на якість роботи самої системи, такі як:

* Попит і пропозиція;
* Фірмові фактори;
* Настрої інвесторів;
* Курси валют, особливо популярних (долар, євро, єна);
* Політична обстановка. [4]

**1.2. Поняття інвестиційного портфелю**

Інвестиції в наш час є дуже вигідним способом збереження або примноження власного капіталу, яким користуються багато успішних осіб. Але далеко не всі інвестиції є рентабельними. Також цей процес потребує конкретних знань та навичок, та високого ризику. В інвестиціях можуть відігравати роль багато різносторонніх факторів, серед яких: раціональність, емоції, та навіть певні мистецькі навички і довіра. Інвестування саме у фінансові цінні папери зараз вважається одним із найкращих способів інвестування своїх заощаджень. Нині, рідко можна зустріти інвесторів, які вкладають усі свої заощадження в один конкретний цінний папір. Замість цього, вони вкладають у певні групи цінних паперів, і саме ці групи цінних паперів складають поняття інвестиційного портфелю. Створення портфелю допомагає знизити ризики високих втрат коштів, адже він формується з різноманітних цінних паперів.

Процес керування інвестиційним портфелем охоплює всі процеси, які пов’язані зі створенням, підтримкою, та аналізом портфелю. Це різноманітний процес, який починається з проблеми вибору з великої кількості цінних паперів. Вибір інвестора залежить від характеристик ризику та доходності окремих цінних паперів. А в комбінації цінних паперів, все стає більш ускладнено, адже можна комбінувати різні папери, в різній кількості, які мають свої унікальні характеристики та можливості. Обираючи портфель, потрібно враховувати не лише ризики самих акцій, а й зовнішнє економічне положення. У міру того як економічне та фінансове середовище постійно змінюється, характеристики ризику та доходності окремих цінних паперів, а також портфелів також змінюються. Це вимагає періодичного перегляду та перегляду інвестиційних портфелів інвесторів. Інвестор інвестує свої кошти в портфель, очікуючи отримати хороший прибуток, який відповідає ризику, який йому доводиться нести. Наступним необхідно виміряти прибуток, отриманий від портфеля, і оцінити результативність портфеля. Для всіх процесів потрібно використовувати аналітичні методи оцінки портфелю. [5, с. 160]

Кожен інвестор ставить перед собою певні цілі, які він бажає досягнути, при створенні свого інвестиційного портфелю. До них належать такі:

- максимізація прибутку. Це, мабуть, найважливіша мета управління портфелем. Адже кожен інвестор хоче збільшити вартість своїх інвестицій.;

- приріст власного капіталу;

- баланс суміші активів.[6]

На поняття інвестиційного портфелю також впливає людський фактор. Одні люди хочуть ризикнути та зірвати «великий куш», інші вирішують зайняти більш безризикову позицію, і, наприклад, купити акції стабільніших компаній. Саме різноманіття цих поглядів і сформували різні типи інвестиційних портфелів. Їх існує дуже багато, і всі мають свої специфічні ознаки, але варто описати 5 основних. Серед них:

- Агресивний портфель. З назви зрозуміло, що у цьому портфелі, інвестор бере на себе більші ризики, а, отже, основною ціллю є високий прибуток. Часто в це поняття входять відносно нові компанії, які лише запустились, та прагнуть швидко розростись за рахунок зовнішнього інвестування.

- Захисний портфель. Протилежність минулому портфелю, оскільки тут акції є більш стабільними, та майже безризиковими. До цього відносять стабільні компанії, які навіть в період криз не страждають, адже займають основні ніші для виживання людей.

- Портфель доходів. Даний тип портфелю більш зосереджений на отримання дивідендів. Він схожий на захисний портфель, але тут обираються акції з вищою прибутковістю, наприклад, нерухомість.

- Спекулятивний портфель. У даному випадку, спекулятивний портфель схожий на агресивний портфель, але в нашому випадку ставиться ставка на успішне майбутнє компанії куди інвестуються кошти. Саме в далеку перспективу.

- Гібридний портфель. Останнім виступає гібридний портфель, суть якого заключається в тому, що він об’єднує всі попередні. Дуже гнучкий, балансний, і не загострений під якусь конкретну тактику.[7]

**1.3. Принципи створення інвестиційного портфелю**

Головним завданням інвестиційного портфелю є принесення прибутку інвестору. Для виконання цього завдання потрібно створити такий оптимальний інвестиційний портфель, з відповідним рівнем ризикованості, та рівнем прибутку. Процес підбору цінних паперів для портфоліо називають формування або створення інвестиційного портфелю. Процес відбору портфоліо складається з чотирьох основних кроків:

Крок 1: Визначення активів, які будуть розглянуті для створення портфеля.

Крок 2: Створення необхідних вхідних даних для вибору портфоліо; це передбачає оцінку очікуваної прибутковості, відхилень для всіх розглянутих активів.

Крок 3: Окреслення ефективного портфеля.

Крок 4: Враховуючи рівень терпимості інвестора до ризику, вибір оптимального портфеля з точки зору: активів, які будуть зберігатися; та пропорцію наявних коштів, які мають бути виділені кожному. [8 с. 220-221]

Крок 5: Оцінка ефективності. На цьому етапі відбувається обчислення реалізованої дохідності портфеля і зіставлення одержаного результату з обраним базисним показником. За базисного показника може бути обраний будь-який загальновідомий фондовий індекс.[9]

Для побудови портфелю існує два основних підходи. Перший це традиційний спосіб, у ньому потреби інвестора щодо доходу та приросту капіталу переоцінюються, і для задоволення потреб інвестора вибираються відповідні цінні папери. Загальною практикою традиційного підходу є оцінка всього фінансового плану особи. Другий спосіб, це сучасний. Його суть заключається в тому, що портфелі будуються таким чином, щоб максимізувати очікуваний прибуток для певного рівня ризику. Він розглядає створення портфеля з точки зору очікуваного прибутку та ризику, пов’язаного з отриманням очікуваного прибутку.

Розпочнемо з традиційного. Традиційний підхід в основному має справу з двома основними рішеннями, а саме:

- Визначення цілей портфоліо

- Вибір цінних паперів для включення в портфель.

Більш у подробицях цей метод зображено на рис. 1.1.

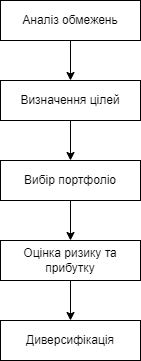


Рис. 1.1 Кроки в традиційному підході до створення інвестиційного портфелю[10]

1. До аналізу обмежень належать:

* Потреби в доході. Потреба в доході означає, що для інвестора потрібний стабільний та постійний дохід, для формування та підтримки інвестиційного портфелю.
* Ліквідність. Потреба інвестиції в ліквідності є індивідуальним поняттям. Якщо інвестор віддає перевагу високій ліквідності, тоді кошти слід інвестувати у високоякісні короткострокові боргові зобов’язання, такі як фонди грошового ринку, комерційні папери та акції, тобто ті, які широко і швидко продаються.
* Безпека основної вартості. Для інвестора варто враховувати при інвестиції, як давно компанія в яку інвестуються кошти знаходиться на ринку.

2. Загальні цілі були описані раніше, але можна додати, що прибуток, якого вимагає інвестор, і ступінь ризику, який він готовий прийняти, залежать від обмежень. Цілі портфеля варіюються від прибутку до приросту капіталу.

3. Вибір портфелю. Вибір портфеля залежить від різних цілей інвестора.

4. Оцінка ризику та прибутку. Процес аналізу ризику та прибутку при створенні портфеля базується на певних припущеннях. Однією з них є те, що інвестори віддають перевагу вищим прибуткам від цінних паперів, що вимагає більшого ризику. Здатність отримати більший прибуток залежить від оцінки інвестором ризику та готовності взяти на себе певні ризики, такі як процентна ставка, купівельна спроможність, фінансовий і ринковий ризики. Оцінюючи різні ступені ризику, інвестор може створити портфель.

5. Диверсифікація. Після визначення відповідної суміші активів і аналізу пов’язаних ризиків і прибутків завершальним кроком в управлінні портфелем є диверсифікація. Хоча фінансовий ризик можна мінімізувати шляхом інвестування у високоякісні облігації, ці цінні папери, як правило, пропонують слабкий захист від інфляції. Навпаки, акції, як правило, забезпечують кращий захист від інфляції, але більш сприйнятливі до фінансових ризиків. Щоб збалансувати фінансовий ризик і ризик купівельної спроможності, високоякісні конвертовані цінні папери можуть бути хорошим варіантом. Диверсифікація портфеля повинна базуватися на потребах інвестора в доходах і рівні толерантності до ризику. [10]

Тепер перейдемо до модерного. Традиційний підхід – це комплексний фінансовий план для особи. Він враховує індивідуальні потреби, такі як житло, страхування життя, пенсійні плани, тощо. Але ці підходи до фінансового планування не застосовуються в модерному підході, або підході Г. Марковіца. Г. Марковіц приділяє більше уваги процесу підбору портфоліо. Акції не вибираються на основі потреби в доході. Вибір ґрунтується на аналізі ризику та прибутку. Прибуток включає ринковий дохід і дивіденди. Інвестор потребує прибуток, і він може бути у формі ринкового доходу або дивідендів. Наприклад, зі списку акцій, що котируються на фондовій біржі, інвестор вибирає приблизно якусь групу акцій із 10 або 15 акцій. Для цих акцій буде розраховано очікуваний дохід і ризик. Передбачається, що інвестор має на меті максимізувати очікуваний прибуток і мінімізувати ризик. Крім того, передбачається, що інвестори підуть на ризик у ситуації, коли вони отримають за це відповідну винагороду. Це означає, що люди віддадуть перевагу портфелю з найвищим очікуваним прибутком за певного рівня ризику. У сучасному підході останнім кроком є процес розподілу активів, тобто вибір портфеля, який відповідає вимогам інвестора. Основною вимогою від інвесторів, за Г. Марковіцем, це є ризик.[10]

1. МАШИННЕ НАВЧАННЯ ПРИ АНАЛІЗІ ФОНДОВИХ РИНКІВ

**2.1. Основні методи машинного навчання**

Машинне навчання є поняттям широким та багатогранним. Завдяки ньому наше життя стає простіше, рішення приймаються швидше, оптимізуються різні сфери людської діяльності. Машинне навчання це отримання різних знань із наших даних. Саме завдяки машинному навчанню ми отримуємо наші рекомендації різноманітного контенту на різних платформах, таких як: YouTube, Instagram, Spotify і т.д. Навіть реклама, яка завжди з’являється у вашому смартфоні, побудована на принципах машинного навчання, адже саме завдяки цьому методу аналізу даних можна зрозуміти, які продукти у відповідності до користувача можуть йому сподобатись.[11, с. 1]

Саме машинне навчання та вміння користуватись його методами, та правильної праці з ним потребує багатьох навичок. Можна виокремити основні, до них належить: математика, теорія ймовірностей, статистика, високий рівень програмування, іноземна мова(у випадку якщо з нею проблеми), та загальне розуміння принципів машинного навчання. Всі ці знання потрібні для уникнення різноманітних помилок, і правильного функціонування моделі, які будуть будуватись у майбутньому.[12]

Оскільки машинне навчання доволі широка галузь, то у залежності від різноманіття завдань, розрізняють відповідно їх типи виконання. Існують два основних види задач: навчання з вчителем (так званий Supervised Learning), і протилежний типаж навчання без вчителя (так званий Unsupervised Learning).

Концептуально суть навчання з вчителем полягає в тому, що в цьому випадку ми маємо розуміння, який результат повинен бути на виході. Для невеликого прикладу для кращого розуміння ми маємо невеликий датасет, де ми маємо різні фрукти, та їх розміри, і ми вже зарання знаємо які фрукти з якими характеристиками. Ми можемо навчити модель, яка буде казати, чи буде цей фрукт бананом, чи апельсином в залежності від розмірів. Відповідно навчання з вчителем поділяється на дві інші категорії: класифікація, та регресія.

Завдання класифікації полягає в тому, що на основі даних, нам потрібно сказати, якусь конкретну відповідь, наприклад, «так/ні», чи ось як у минулій задачі віднести до певної групи наше значення. Для вирішення цього типу задач, можна використати цілу низку методів, серед яких: [13]

k-Nearest Neighbors. Суть цього методу полягає в тому, що ми беремо K кількість сусідів, і у відповідності до кількості цих сусідів рахуємо відстані до них. Перевагою цього методу є його простота, та висока точність при праці з великим train набором даних. З мінусів можна виділити знаходження оптимального значення K, що може зайняти багато часу, і також довгота виконання, адже якщо даних багато, то для нової точки потрібно обраховувати відстань до всіх інших точок, для винесення правильного рішення про призначення точки до відповідної групи.[14]

Decision Trees. Щоб передбачити мітку класу запису за допомогою дерев рішень, ми починаємо з кореня дерева та перевіряємо значення атрибута. Потім ми порівнюємо його з атрибутом відповідного запису. На основі цього порівняння ми переходимо до відповідної гілки та переходимо до наступного вузла, і так далі до поки дійдемо до листків, і всі дані будуть класифіковані. Основним мінусом є поняття overfitting тобто перенавчання, адже ми можемо дійти до такої ситуації, що у нас, наприклад, в кінцевому листку буде по 1-2 значенням, що не є допустимим. Плюсом є простота розуміння цього методу, та відсутність необхідності скейлингу та нормалізації даних, що для інших методів є досить важливим поняттям.[15]

Naive Bayes. Як видно з назви, за основу в цьому методі взято теорему Байєса, суть якої полягає в тому, що вона описує ймовірність події на основі попередніх знань про умови, які можуть бути пов’язані з подією. Простіше кажучи, теорема Байєса припускає, що наявність певної ознаки в класі не пов’язана з наявністю будь-якої іншої ознаки. З плюсів можна виразити швидкість, вміння працювати з багатокласовими даними, та зручність використання при вирішенні проблем з текстом. [16]

SVM (Support vector Machines). Логіка роботи цього методу полягає в тому, що ми беремо кожен елемент даних і представляємо його як точку в просторі з великою кількістю вимірів. Кожна ознака предмета отримує свою координату у нашому уявному просторі ознак, які наший набір даних володіє. Потім ми намагаємося знайти найкращий рядок, щоб розділити елементи на різні групи на основі їхніх характеристик. Ми хочемо, щоб цей рядок був якомога зрозумілішим, щоб було легко визначити, до якої групи належить кожен елемент. Опорні вектори — це просто координати індивідуального спостереження. Класифікатором SVM виступає межа, яка найкраще розділяє два класи. Цей метод підходить у більшості випадків, коли кількість просторів (фіч) є доволі високою, і він є доволі зручним для системи комп’ютера, адже використовується система векторів, яка простіше сприймається системою. З мінусів можна виділити його рідкість у використанні, адже SVM не використовується у випадках коли даних багато, та коли є шуми. [17]

Тепер про регресію. Регресія, це ситуація, коли ми маємо два набори даних, Y-набір, це той набір який нам потрібно передбачити, та X-набір, той який ми будемо брати за основу, так звані предиктори. Існує ціла низка методів, для створення регресійної моделі, які в залежності від ситуації, можна по різному використовувати.

Лінійна регресія. Лінійна регресія шукає найкращу лінію спуску між X та Y наборами, для того, щоб охарактеризувати дані найкращим чином. Для лінійної регресії існують підтипи, які залежать від кількості ознак, з якими ми працюємо. У випадку простої лінійної регресії, це взаємозалежність між однією Y змінною, та однією X змінною. Множинна лінійна регресія описує ситуацію коли ознака Y є лише одна змінна, а X від двох і більше ознак. Багатофакторна лінійна регресія описує схожу до минулого випадку модель, але ознак Y може бути більше двох.

Але далеко не завжди дані ідеально підходять під лінію. Їх потрібно описувати різного роду кривими. З цим може допомогти поліноміальна регресія. Її основна відмінність від регресії полягає в тому, що кожна наступна X-ова ознака підноситься до більшого степеню. Через це, й отримується крива. Але, якщо модель використає високий степінь, то через це, може виникнути проблема перенавчання моделі. [18]

Протилежністю до supervised learning є unsupervised learning, у якому нам потрібно встановити класи, або групи, і вже від цього відштовхуватись та будувати модель. Навчання без вчителя використовують в трьох основних випадках, до них відносяться: кластеризація та зменшення розмірності.

Кластеризація, це поняття, що є дуже схожим до класифікації, але у нас зарання невідомо як кількість груп, так і їх назви. Основним методом кластеризації є метод K-means. Алгоритм K-means — це ітераційний алгоритм, який намагається розділити набір даних на попередньо визначені K окремі підгрупи, які не перекриваються, де кожна точка даних належить лише одній групі. Він намагається зробити точки, які всередині кластеру, якомога подібнішими, водночас зберігаючи кластери максимально несхожими один на одного. Алгоритм призначає точки даних кластеру таким чином, щоб сума квадратів відстані між точками даних і центром кластера була мінімальною. [19]

Зменшення розмірності використовують у випадку складності візуалізацій, наприклад, через викиди, чи дуже велику кількість ознак даних. Методика зменшення розмірності допомагає для моделі уникати перенавчання через багатогранність ознак, які можуть спантеличувати нашу модель. Фундаментальним методом, який використовують для зменшення розмірності це аналіз основних компонентів, або PCA. Цей підхід передбачає перетворення даних за допомогою лінійного процесу, який створює нове представлення, генеруючи набір основних компонентів. Перший головний компонент – це напрямок, який охоплює найбільшу варіацію в наборі даних. Тим часом, другий головний компонент також визначає максимальну дисперсію в даних, але він абсолютно не пов’язаний з першим головним компонентом, роблячи його напрям перпендикулярним або під прямим кутом до першого компонента. Цей цикл повторюється для кожного виміру, створюючи новий головний компонент, ортогональний до попередніх і фіксує найбільшу кількість варіацій у даних.[20]

Машинне навчання допомагає в багатьох випадках, коли потрібно проаналізувати великі масиви даних, знайти патерни, переглянути динаміку.

**2.2. Вибір моделей для аналізу.**

У цій роботі створюється модель, для керування інвестиційним портфелем, та надання користувачу різного роду порад, з правильної оптимізації його портфелю, у залежності його бажань. Тому основним тут будуть регресія та кластеризація. Регресія потрібна для самих даних, які у нас є безперервною змінною. Кластеризація потрібна для встановлення різного роду корисної інформації, як-от: стійкість до ризику чи інвестиційні цілі інвестора.

Для правильної роботи моделі, потрібно пройти декілька кроків, для того, щоб зрозуміти, яка модель найкраще підходить у нашій ситуації. З розумінням, допоможе рис. 2.1:

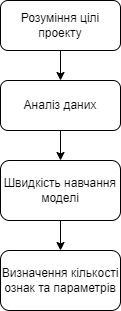


Рис. 2.1. Кроки для вибору моделі машинного навчання. [21]

Пройдемо по кожному з пунктів, і першим є розуміння цілі проекту. Зрозумівши, що саме ви хочете бачити як результат проекту, які дані вносите, і т.д. доволі легко можна розпочати створення власної моделі, адже завдяки цьому, є можливість швидко зрозуміти, саме яка модель підходить під ваші дані, і вже від цього відштовхуватись, та переходити до другого пункту. [21]

Другим пунктом є аналіз даних. До цього відноситься попередня обробка даних, тобто нормалізація, стандартизація, заповнення пропущених значень та інші методи. Завдяки цьому пункту, можна визначити також різного роду метрики, подивитись на дані під іншим кутом, зрозуміти глибше, які гіперпараметри можна обрати в майбутньому, чи виконати інші дії. У моїй моделі, буде аналіз декількох акцій відомих, та менш відомих компаній, і відповідно їх аналіз. [21]

Наступним швидкість навчання моделі. Завжди є можливість використати умовний GridSearchSV, який з однієї сторони виконає все правильно, але з іншої для перевірки правильного вибору параметрів, дана модель може перебрати тисячі, а то й десятки тисяч наборів параметрів, відповідно й створити таку ж кількість моделей. Це не є оптимально, тому варто відповідально ставитись до цього процесу. Варто додати, що для хорошої роботи моделі потрібно не забувати про техніки кросс-валідації, які допоможуть уникнути проблем з неправильно розподіленими вибірками, та неправильно розподіленим набором даних. [21]

Визначення оптимальної кількості параметрів та ознак. Через велику кількість і параметрів і ознак, наша модель може «перевчитись», тому потрібно бути обережним. Також, варто не забувати, що для досягнення високої точності, варто для моделі давати трішки більше часу на навчання, адже саме час, може стати фактором, який і визначить точність вашої моделі. Занадто мало часу може призвести до значних помилок у роботі, відповідно неправильних результатів. Для моєї моделі, теж використовується різний набір параметрів, для оптимальної роботи програми.[21]

Можна перебрати різну кількість моделей, щоб знайти оптимальну, з відповідними результатами. Тому не завжди варто зациклюватись на одній моделі, варто спробувати інші, можливо інші методи оптимізації, чи оцінки якості моделі, все залежить від ситуації.

**2.3. Порівняння якості роботи моделей.**

Для різного роду моделей існує багато способів визначити їх точність. Щоб правильно порівняти якості роботи різних моделей, потрібно обрати один або більше показників, і вже від цього відштовхуватись.

У двох схожих методів машинного навчання, кластеризації та класифікації, існує ціла низка методів порівняння та загальної оцінки роботи моделей. У випадку класифікації всі методи опираються на так звану «матрицю плутанини», де описуються всі результати моделі, з яких можна робити висновки. Її загальний вигляд складається так: по головній діагоналі елементи, які передбачились правильно(True positive, True negative), а всі інші елементи, ті, що неправильно(False positive, False negative). З цих даних, можна обрахувати 3 основні оцінки, серед яких:

* Точність(Accuracy). Її суть полягає в тому, що ми всі правильно передбачені значення ділимо на загальну кількість значень.
* Чіткість(Precision). Даний показник показує, якою є частка істинно позитивнх значень між тих, які належать до одного класу.
* Відтворення(Recall). Схожий з минулим показник, але в цьому випадку рахується як частка істинно позитивних, на суму істинно позитивних та хибно негативних. [22]

У випадку кластеризації, потрібно використовувати інші методики, адже у нас немає можливості переглянути зарання, чи до правильного кластеру ми віднесли відповідне значення, чи ні. Тут потрібно дивитись, якщо модель відокремлює несхожі чи віддалені дані, або поєднує близькі, то значить що модель вчиться правильно. Існує 2 методи, щоб визначити оцінку якості роботи моделі кластеризації.

Перший це коефіцієнт Силуету, і для його визначення потрібні два значення: значення A - середня відстань між значенням та всіма іншими точками в тому ж кластері, B - середня відстань між значенням та всіма іншими точками в наступному найближчому кластері. Загальна формула виглядає наступним чином:

Значення варіюється між -1 та 1, де -1 – це неправильна кластеризація, 1 – кластери розподілені правильно.

І другим показником для визначення точності моделі є індекс Данна. Індекс Данна дорівнює мінімальній відстані між кластерами, поділеній на максимальний розмір кластера. Даний метод не є популярним, адже при великих кластерах, та відповідно їх близькості даних індекс є дуже високим. [23]

Для регресій використовують іншого формату методи, адже дані є безперервними змінними. Основними серед них є:

* Середньоквадратична похибка (MSE, RMSE). Показує загальну похибку в даних, те на скільки воно відійшло від даних. Загальна формула:

* Оцінка . Дана оцінка показує дисперсію, яка виходить з моделі, на основі її харакетристик. Загальна формула: . [22]

Також варто не забувати за зовнішню оцінку даних. Порівняти зовнішній вигляд регресії, чи зменишити розмірність кластеру, і подивитись під різними кутами, знайти різного роду закономіроності і т.д., ці всі та інші дії точно не завадять, для покращення та налаштування моделі. Прийняття правильного рішення у таких випадках є дуже важливим кроком у побудові правильної моделі.

3 РОЗРОБКА МОДЕЛІ ОПТИМАЛЬНОГО ІНВЕСТИЦІЙНОГО ПОРТФЕЛЮ

**3.1. Вибір даних та навчання моделі**

Кожна з моделей має початок з даних. Саме дані визначають, наскільки конкретна модель буде якісною, та видавати потрібні результати. У випадку зі створенням моделі оптимального інвестиційного портфелю, у нас будуть використовуватись дані компаній, які входять до S&P500.

Для цієї моделі було прийнято рішення використовувати саме не ціни на конкретні акції, а їх прибутков, тобто зміну цін на ці акції у відсотках. Окрім основних відсоткових показників, також було додано поточну ціну на фондовому ринку на 1 акцію відповідної компанії.

У нашому випадку стоїть задача кластеризації даних, тобто знаходження прихованих паттернів у датасеті. Для цього було використано один із основних методів кластеризації це KMeans. Основною причиною використання саме цієї моделі стала простота та зрозумілість параметрів, зміна кластерів (у багатьох моделях автоматично знаходиться оптимальна кількість кластерів), та якість самої моделі, адже на етапі тестування різного роду моделей кластеризації з моїми даними, неозброїним оком було видно, що цей датасет підходить лише для KMeans.

Загалом дані, які будуть використані у процесі моделювання зображені на рис. 3.1.

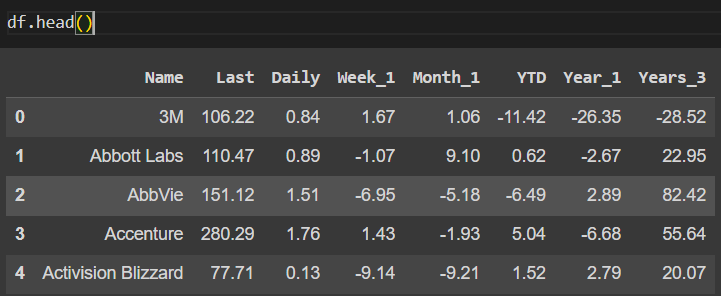


Рис. 3.1. Дані для моделі [25]

Тут можна побачити, що стовпчик “Name” відповідає назві компанії, “Last” – остання ціна на фондовому ринку на цю акцію, “Daily”, “Week\_1”, “Month\_1”, “YTD”, “Year\_1”, “Year\_3” – відсоткова зміна ціни на акцію.

Для даної моделі були використані саме ці значення, тому що:

1. У цій таблиці для звичайного користувача немає нічого нового, є назва акції, її остання ціна та відсоткові зміни за різні періоди.
2. Ці дані легше проходять через скейлер. Якби було використано не відсоткові зміни а звичайні ціни, то модель би могла некоректно працювати через поганий скейлинг даних.
3. Зручний спосіб візуалізації. Було б складно показати якість роботи моделі, якби не були відсоткові зміни вартості моделей.

Перш ніж створювати модель, потрібно провести загальну очистку даних, прибрати N/A значення, змінити типи даних, для коректної роботи моделі. В основному, була проблема лиш з типажем даних та різного роду знаками (+, -, %) у видаленні.

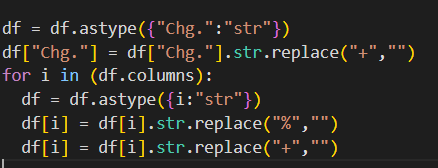


Рис. 3.2. Код до загальної зміни даних. [25]

У ході тестування моделі, виникли деякі сумніви стосовно якості роботи самої моделі. Якщо обирати всі дані, і їх внести в одну модель, були проблеми з якістю роботи самої моделі. Тому, було прийнято рішення розбити нашу модель на 3 шматочки, для деяких часових проміжків моделі, задля покращення ефективності роботи моделі. Було розбито дані наступним чином:

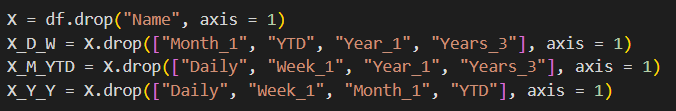


Рис. 3.3. Код до розбивки даних на частини. [25]

Як можна побачити, то я розбив на частини Daily та Week\_1, Month\_1 та YTD, і остання частина це річна зміна, тобто Year\_1, Years\_3. Дані готові до внесення у модель, але перед цим потрібно провести ще один етап підготовки даних.

Для коректної роботи моделі також варто не забувати про скейлинг даних, тобто масштабування. Дані з різними дисперсіями, погано проходять через модель, понижують її продуктивність. Для цього, зробимо зовнішнє порівняння нашої моделі через скейлер “StandardScaler” та “MinMaxScaler”. Загалом, з рис. 3.4 можна помітити, що у випадку MinMaxScaler оцінка буде вища, завдяки невеликим відстаням від кластерів, але при цьому, один із кластерів дуже сильно просідає саме в кількості представників, тому у нашому випадку краще використовувати саме StandardScaler, який допомагає моделі більше.

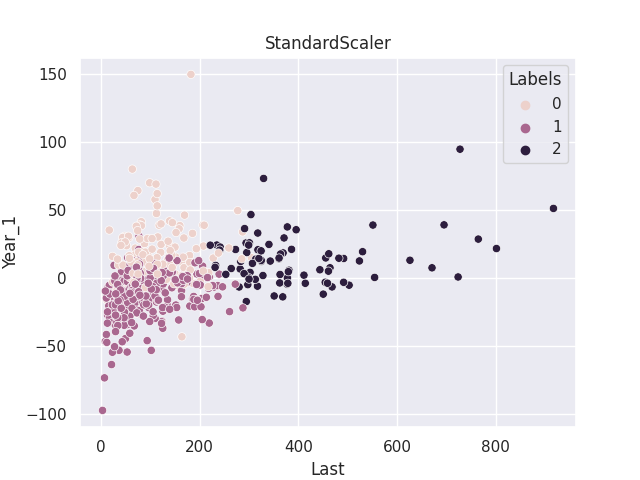
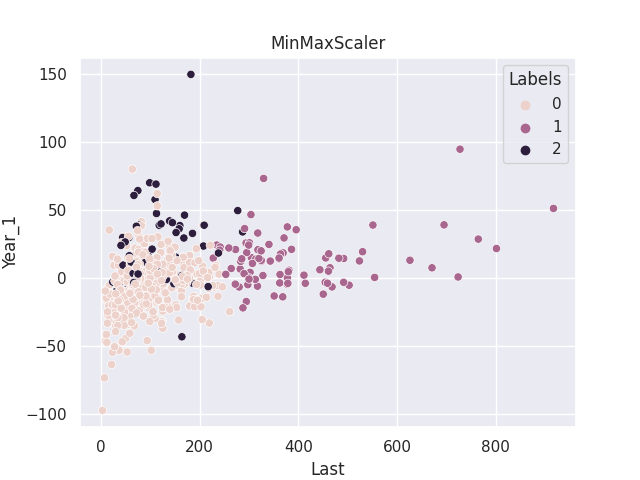


Рис. 3.4. Порівняння різних методів масштабування у моделі. [25]

Загалом код для самої моделі виглядає наступним чином:

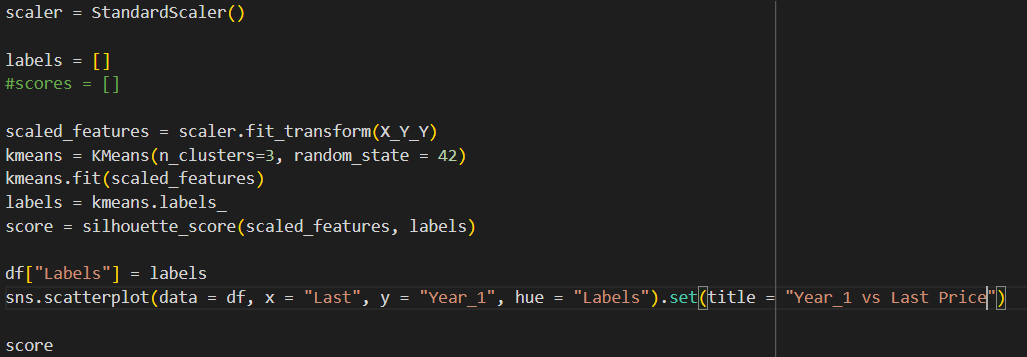


Рис. 3.5. Код для моделі. [25]

Загальні результати моделі заносяться в датасет, який буде використовуватись у самій програмі, для допомоги створення оптимального інвестиційного портфелю.

Після перевірки моделі, загальна таблиця виглядає наступним чином:



Рис. 3.6. Датасет після обробки даних, та результатів моделі. [25]

У нашому випадку, LpLi значить Low price Low income, LpHi – Low price High income, HpHi – High price High income.

Загалом точність моделі становить 0.55. Дане значення є коефіцієнтом Силуета. Дана оцінка не показує ось так напряму точність кластеризації, а в загальному дає розуміння про розподіл кластерів. Навіть за умови високого коефіцієнту Силуета може виникнути ситуація, що один із кластерів має приблизно 5 значень, коли інші більше 200. У такому випадку значення може бути високим, але модель вже буде некоректною. Але у нашому випадку оцінка моделі є непоганою.

**3.2. Динамічне управління портфелем на прикладі ІС «Моно Інвест»**

«Моно Інвест» - програма, створена засновниками Monobank, основною спеціалізацією якої, є створення та зручне редагування портфелю інвестицій користувача. Загальний інтерфейс виглядає як на рисунку 3.7.

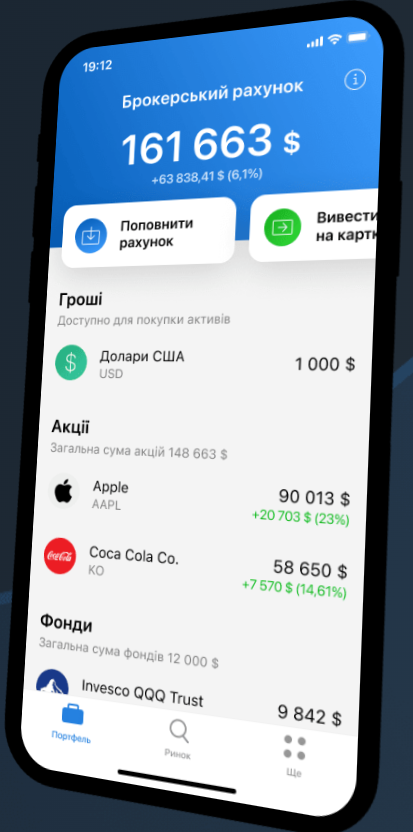


Рис. 3.7. Інтерфейс програми «Моно інвест». [24]

Можна помітити, що тут є загальний баланс рахунку, ваш персональний портфель, кнопка «Ринок», де можна шукати нові акції, зручно їх сортувати, переглядати останні тенденції на фондовому ринку. Також меню «Ще» для додаткових операцій, прочитання різного роду документації та допомоги роботи з програмою. [24]

Загалом оптимізація буде відбуватись за рахунок знаходження відповідних акцій з найвищою дохідністю за відповідний період, та з відповідним кластером. Завдяки цим параметрам користувач зможе встановити рентабельність купівлі даних акцій, чи можливий високий ризик.

**3.3. Розробка системи динамічного управління портфелем.**

З результатів кластеризації були відокремлені декілька видів акцій. З їх допомогою користувач може орієнтуватись, чи акції були дорогі, чи дешеві та не приносять прибутку. В етапі планування інтерфейсу програми було прийнято до уваги деякі пункти з програми «Моно інвест», що допомогло створити більш якісну та зручну програму.

Загальний вигляд системи зображений на рис. 3.8.

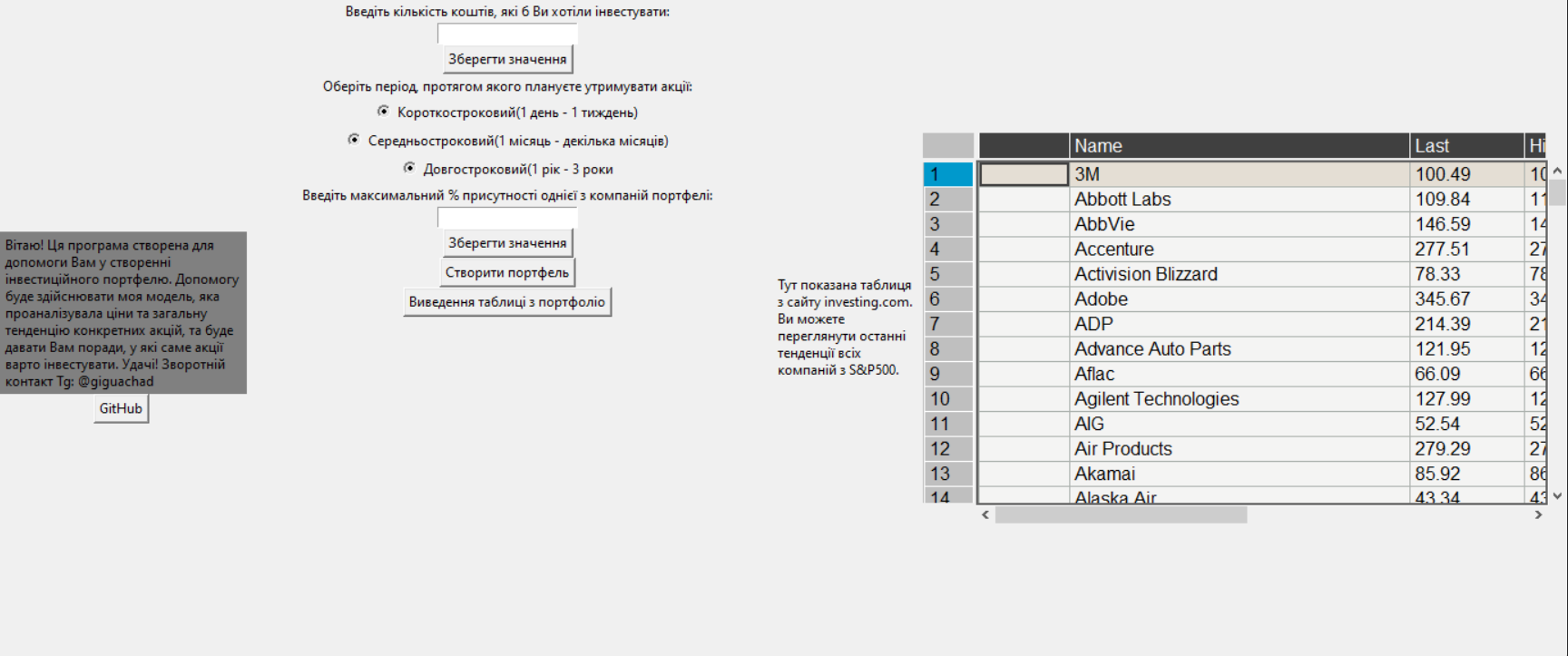


Рис. 3.8. Інтерфейс програми. [25]

Загалом дашборд розбитий на 3 колонки, кожна з яких дає певну інформацію. У першій колонці знаходиться загальна інформація про програму, та посилання на GitHub. У другій колонці знаходиться основний функціонал програми, тобто введення суми, яку бажає користувач інвестувати, період, протягом якого будуть утримуватись акції, та % присутності однієї з компаній у портфелі(у грошовому еквіваленті). З рис. 3.9 видно, що після введення значень виводиться таблиця акцій, у якій видно назви акцій, лейбли з моделі, та кількість акцій, яку варто придбати, на думку програми.

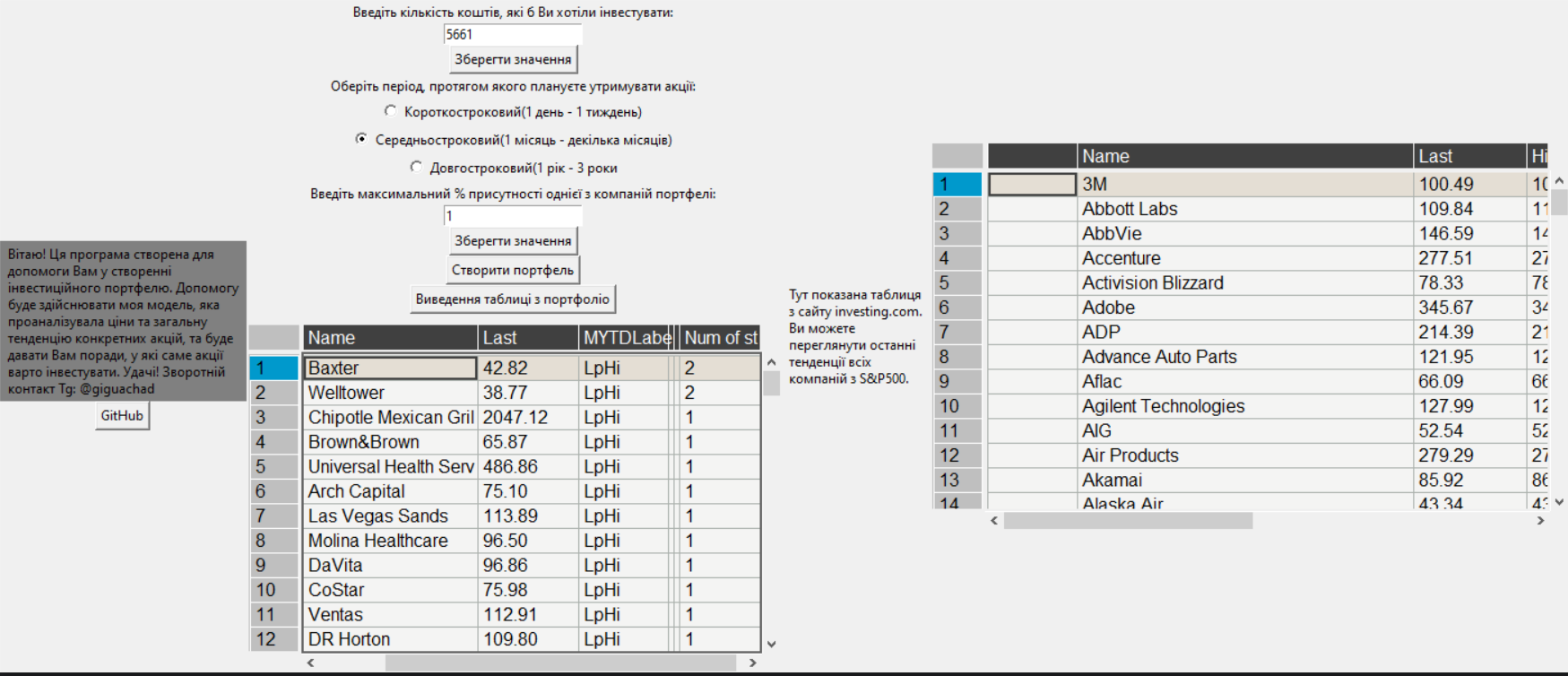


Рис. 3.9. Інтерфейс програми після калькуляцій. [25]

Остання ціна динамічно береться з сайту investing.com, що дозволить користувачеві швидко знаходити потрібні йому дані, та правильній роботі підрахунку вартості акцій, які будуть брати участь у даному портфелі.

Третя колонка це таблиця, яка показує тенденцію декількох останніх днів, відсоткові зміни у вартості акцій за останній день. Перевагою таких таблиць є сортування, можна швидко сортувати та знаходити для себе нові акції.

Варто не забувати, що це лише навчальна програма, і її результатам довіряти на 100% не варто. Вона показує те, як машинне навчання може допомогти кластеризувати та спростити завдання пошуку нових акцій, та менеджменту портфелю інвестицій.[25]

ВИСНОВКИ

У даній курсовій роботі було проведено аналіз фондового ринку, його загальний стан, та застосування методів машинного навчання для прогнозування. Завдяки машинному навчанню була побудована модель кластеризації, для ранжування акцій.

Фондовий ринок є широким поняттям, що включає у себе багато різних сфер діяльності суспільства. Фондовий ринок також слугує плацдармом для інвесторів, адже саме тут відбуваються основні операції з цінними паперами. Існує багато видів інвестиційних портфелів, кожен з них слугує для відповідних тактик.

З дослідження методів створення інвестиційного портфелю можна зробити висновок, що існують різні підходи до його формування, та етапи, яких потрібно дотримуватись для оптимізації процесу.

Вибір моделі є одним із основним завданням цієї курсової роботи. Їх існує багато, але для даної задачі було прийнято рішення обрати модель кластеризації, для кластерування наших даних, та знаходження різних зв’язків, та різних підходів до можливих корективів інвестиційного портфелю. Спробувавши декілька моделей кластеризації, було прийнято рішення обрати модель KMeans. Вона краще у взаємодії з користувачем, саме завдяки зручному встановленню параметрів, своєю концептуальною простотою методу визначення самих кластерів, та самими результатами моделі у цьому випадку кластеризації. Якість роботи інших моделей у цьому випадку була низькою, через відсутність у багатьох методах кластеризації встановлення параметру кількості груп, що було одним із основних показників, який визначав якість роботи та загальну оцінку моделі.

Вибір метрики оцінювання якості роботи моделі напряму залежить від обраної моделі. Для регресії можна використати оцінку , для класифікації існує ціла низка способів обрахувати оцінку. У цій роботі використовувався метод кластеризації для створення оптимального інвестиційного портфелю за допомогою машинного навчання, тому для цього був використаний коефіцієнт Силуета. Оцінка цієї моделі становить 0,55. У коефіцієнті Силуета значення варіюються від -1 до 1, де -1 значить що кластеризація проведена дуже погано, тобто відстані від кластерів дуже низькі, і формально кластер лягає один на одного, а 1, відповідно, дані прокластеризовані добре. Цей результат не є найвищим, адже по ходу роботи, було видно, що розміри третього кластеру час від часу відрізнялись, у залежності від скейлеру, чи інших факторів. І чим менше один із кластерів, тим вища оцінка, адже відстань між кластерами падає.

Створення інвестиційного портфелю відбувається за тими результатами, що дало машинне навчання, яке прокластеризувало всі дані у відповідності їх дохідності та ціни за різні періоди. Сам портфель створюється автоматично, лише потрібно задати деякі з парамтерів, а саме: кількість грошей, які бажає користувач інвестувати; період, протягом якого будуть утримуватись кошти; та відсоток присутності однієї з компаній у портфелі.

Машинне навчання дало для цієї задачі нове русло, адже завжди у такій сфері оцінювали більш з математичної точки зору, у перспективі. Завдяки цій моделі на акції можна подивитись під інакшим кутом, зробити нові висновки про рентабельність інвестування у дані акції, та створити для себе оптимальний інвестиційний портфель.

Даною роботою можуть користуватись інвестори будь-якого рівня досвіду, адже тут показано більш свіжий погляд на дану проблему, описано як модель машинного навчання бачить дану ситуацію зі своєї перспективи, та на основі чого робить відповідні висновки.

Дану задачу, я вважаю, було виконано лише частково, адже тут не використовувалась якась періодична динаміка для конкрентого аналізу, а розривні періоди, мається на увазі, наприклад, дані Daily і Week\_1, і вже від них відштовхувалась модель. Також, варто не забувати, що це лише модель машинного навчання, економічна складова у цьому плані є невисокою, адже модель не враховує багато факторів із зовнішнього світу, взаємодію покупців акцій, різного роду криз і т.д.

З результатів дослідження видно, що на фондовий ринок впливає багато факторів, і найголовнішим з них є час. З часом змінюється ринок, суспільство та технології. Вибір оптимального інвестиційного портфелю завжди був проблемним питанням. Дана модель, яка побудована на основі машинного навчання, допоможе знайти свій оптимальний інвестиційний портфель.

Отже, дана курсова робота має важливий внесок для створення нових моделей в економіці, особливо моделей машинного навчання, що дозволяють вирішувати проблеми різних роду та масштабів.

СПИСОК ДЖЕРЕЛ:

1. What Is the Stock Market, What Does It Do, and How Does It Work? URL: <https://www.investopedia.com/terms/s/stockmarket.asp#toc-how-the-stock-market-works> (дата звернення: 29.03.2023)
2. Що таке фондовий ринок? URL: <https://ffin.ua/blog/faq/investytsii/post/shcho-take-fondovyi-rynok> (дата звернення: 28.02.2023)
3. How Does the Stock Market Work? URL: <https://www.investopedia.com/articles/investing/082614/how-stock-market-works.asp#toc-what-is-a-stock> (дата звернення 17.03.2023)
4. Factors Affecting Stock Market URL: https://www.adityabirlacapital.com/abc-of-money/factors-affecting-stock-market
5. DK24: SECURITY ANALYSIS AND PORTFOLIO MANAGEMENT URL: <https://www.msuniv.ac.in/Download/Pdf/d1277b116ac84d8> (дата звернення 06.03.2023)
6. Goals of Portfolio Management URL: <https://study.com/academy/lesson/goals-of-portfolio-management.html>
7. Types of Portfolio Investment. URL: https://www.angelone.in/knowledge-center/share-market/portfolio-types (дата звернення: 19.03.2023)
8. Financial and Investment Management School of Distance Education Bharathiar University, Coimbatore – 641 046. URL: <https://www.acecollege.in/CITS_Upload/Downloads/Books/1045_File.pdf>
9. 1.3. Процес формування портфеля цінних паперів. Портфельний підхід. Принципи формування портфеля цінних паперів URL: <https://buklib.net/books/22466/> (дата звернення: 06.03.2023)
10. Розділ 6 УПРАВЛІННЯ ПОРТФЕЛЕМ ФІНАНСОВИХ ІНВЕСТИЦІЙ Портфель фінансових інвестицій і принципи його формування URL: <https://posibniki.com.ua/post-upravlinnya-portfelem-finansovih-investiciy-portfel-finansovih-investiciy-i-principi-yogo-formuvannya> (дата звернення 30.03.2023)
11. Introduction to Machine Learning with Python by Andreas C. Müller and Sarah Guido URL: <https://www.nrigroupindia.com/e-book/Introduction%20to%20Machine%20Learning%20with%20Python%20(%20PDFDrive.com%20)-min.pdf> (дата звернення 21.03.2023)
12. Top 5 Essential Prerequisites for Machine Learning URL: <https://intellipaat.com/blog/prerequisites-for-machine-learning/#:~:text=In%20short%2C%20machine%20learning%20requires,algebra%2C%20and%20knowledge%20of%20programming> (дата звернення 21.03.2023)
13. 4 Types of Classification Tasks in Machine Learning URL: <https://machinelearningmastery.com/types-of-classification-in-machine-learning/> дата звернення (21.03.2023)
14. K-Nearest Neighbor(KNN) Algorithm for Machine Learning URL: <https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning> дата звернення (21.03.2023)
15. Decision Tree Algorithm, Explained URL: <https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html> дата звернення (21.03.2023)
16. Naïve Bayes Classifier Algorithm URL: <https://www.javatpoint.com/machine-learning-naive-bayes-classifier> дата звернення (21.03.2023)
17. Learn How to Use Support Vector Machines (SVM) for Data Science URL: <https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/> дата звернення (21.03.2023)
18. Regression in Machine Learning: What It Is and Examples of Different Models URL: <https://builtin.com/data-science/regression-machine-learning> дата звернення (21.03.2023)
19. K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks URL: <https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a> дата звернення (21.03.2023)
20. What is unsupervised learning? URL: <https://www.ibm.com/topics/unsupervised-learning#:~:text=the%20next%20step-,What%20is%20unsupervised%20learning%3F,the%20need%20for%20human%20intervention>. дата звернення (21.03.2023)
21. How to Choose the Right Machine Learning Algorithm: A Pragmatic Approach URL: <https://labelyourdata.com/articles/how-to-choose-a-machine-learning-algorithm#step_1_understand_your_project_goal> дата звернення (25.03.2023)
22. Evaluating a machine learning model. URL: <https://www.jeremyjordan.me/evaluating-a-machine-learning-model/> (дата звернення 25.04.2023)
23. Quick Guide to Evaluation Metrics for Supervised and Unsupervised Machine Learning URL: <https://www.analyticsvidhya.com/blog/2020/10/quick-guide-to-evaluation-metrics-for-supervised-and-unsupervised-machine-learning/> (дата звернення 25.04.2023)
24. Інвестиції Monobank URL: <https://www.monobank.ua/invest?lang=uk> (дата звернення 13.05.2023)
25. Project URL: <https://github.com/giguachad/portfolio/tree/main> (дата звернення 16.05.2023)