КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

Факультет інформаційних технологій

Кафедра технологій управління

Спеціальність 122 – Комп'ютерні науки
Освітня програма «Інформаційна аналітика та впливи»

КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

на тему:

«Інформаційний аналіз та короткострокове прогнозування вартості акцій компаній із найбільшою капіталізацією»

Студента 2-го курсу групи ІАВ-21

Мельниковича Ярослав Руслановича

(прізвище, ім'я, по батькові)

Науковий керівник:

Кандидат технічних наук

(науковий ступінь, вчене звання)

Хлевний Андрій Олександрович

(прізвище, ім'я, по батькові)

(підпис студента)	*(дата)*	*(підпис)*

Попередній захист:

(Висновок: "До захисту в Екзаменаційній комісії")

Завідувач кафедри технологій управління _____ Морозов В.В. _____

(підпис) (прізвище, ініціали) (дата)

Київ – 2022

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

Факультет інформаційних технологій

Кафедра технологій управління
Освітній рівень Магістр
Спеціальність 122 Комп'ютерні науки
Освітня програма Інформаційна аналітика та впливи

ЗАТВЕРДЖУЮ
Завідувач кафедри
професор Морозов В.В.

_____
«____» _____20___року

ЗАВДАННЯ
НА ВИКОНАННЯ КВАЛІФІКАЦІЙНОЇ РОБОТИ

Студент: ***Мельникович Ярослав Русланович***

Група: ІАВ-21

1. **Тема дипломної роботи**: «Інформаційний аналіз та короткострокове прогнозування вартості акцій компаній із найбільшою капіталізацією».

   Затверджена протоколом від "17" листопада 2021 р. №___.

2. **Строк подання студентом готової роботи -** "__ " _____ 20__ р.
3. **Цільова установка та вихідні дані до роботи:** Дослідження, аналізування та прогнозування акційних даних компаній застосування даного результату в цілях інвестиційного вкладу із подальшою можливістю росту для аудиторії.
4. **Зміст роботи**: Пошук, дослідження даних компаній за становищем добових вартостей акцій, аналіз даних, прогнозування росту на короткостроковий період при умові мінімального збитку та подальшого росту із капітал-інвестиційного вкладу.

5. **Перелік графічного матеріалу**: математичні моделі, шкали інфографіки, інформаційні та статистичні графіки, таблиці даних.

6. **Календарний план виконання роботи**:

Етапи виконання дипломних робіт	**Термін виконання**
1. Вибір теми дипломної роботи	20.10.2021
2. Наказ про затвердження тем дипломних робіт та призначення наукових керівників	17.11.2021
3. Формування предметної галузі та об'єкту дослідження.	25.01.2022
4. Розробка плану дипломної роботи і його погодження з науковим керівником	28.02.2022
5. Написання І розділу дипломної роботи	14.03.2022
6. Написання ІІ розділу дипломної роботи	15.04.2022
7. Написання ІІІ розділу дипломної роботи	16.05.2022
8. Підготовка висновків і пропозицій	16.05.2022
9. Попередній захист дипломної роботи	19.05.2022

Дата видачі завдання "__" _____20__р.

Керівник роботи: ктн. ас. Хлевний А. О._____

(підпис)

Завдання прийняв до виконання:

студент групи ІАВ-21 Мельникович Я.Р._____ (підпис)

АНОТАЦІЯ

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

Факультет інформаційних технологій

Кафедра технологій управління

Освітня програма «Інформаційна аналітика та впливи»

Дипломна робота магістра Мельниковича Ярослава Руслановича

Тема роботи – «Інформаційний аналіз та короткострокове прогнозування вартості акцій компаній із найбільшою капіталізацією».

Мета дипломної магістерської роботи – аналіз та розробка моделі прогнозування з метою поліпшити та розширити інвесторам можливості бачення на реалії стану акційного ринку.

Об’єкт дослідження – сфери прогнозування акцій, акційні датасети у розрізі щоденних вартостей акцій компаній.

Предмет дослідження – інформаційні засоби, інструменти, моделі та методи управління акційними даними при аналізі та прогнозуванні їх вартостей на короткостроковий період.

Наукова новизна роботи – модель з графічним прогнозуванням вартості акцій компанії, котра базується на основі реальних отриманих даних, що сформовано для поліпшення українським та міжнародним інвесторам за метою заробітку в інвестиційних фондах.

Методи дослідження – інформаційні аналіз, математичні та рекурентні моделі для вирішення даної задачі, статистичні методи та інструменти для обробки інформації, інструменти для візуалізації даних, середовище збереження даних.

Практичне значення – реалізація за програмним середовищем із задачею прогнозування вартості акцій компаній у короткостроковий період, що базується на мові програмування Python.

Дипломна робота складається зі вступу, основної частини, яка включає три розділи, висновків та списку використаних джерел та додатків. Всього налічує 95 сторінок та перелік посилань з 30 джерела на 4 сторінках, який містить 37 рисунків, 1 таблицю, 16 формул.

Ключові слова: датасети, нейронні мережі, фондовий ринок, інвестори, інвестиції, прогнозування, дані, інформація, акції, компанії, цінні папери, вартість, фондовий індекс.

ПЕРЕЛІК ВИКОРИСТАННИХ СКОРОЧЕНЬ

LSTM(МДКП) – Мережі довгої короткострокової пам'яті;

RNN(РНМ) – Реккурентна нейрона мережа;

CAC 40 – Cotation Assistée en Continu;

Nikkei 225 – Nikkei 225 Stock Average;

S&P 500 – Standard & Poor's;

DJIA – Dow Jones Industrial Average

REIT – Real Exchange Investment Trusts.

ЗМІСТ

АНОТАЦІЯ....4

ПЕРЕЛІК ВИКОРИСТАННИХ СКОРОЧЕНЬ....6

ВСТУП....8

РОЗДІЛ 1. ПОСТАНОВКА ЗАДАЧІ ТА АНАЛІЗ РІШЕННЯ....11

1.1 Огляд предметної сфери та постановка задачі.....11

1.2 Аналіз принципу формування вартості акції компанії.....16

1.2.1 IT....18

1.2.2 Охорона здоров'я....20

1.2.3 Фінансовий сектор....22

1.2.4 Комунальні послуги....23

1.2.5 Товари другої потреби....24

1.2.6 Промисловий сектор....25

1.2.7.1 Товари першого споживання....26

1.2.7.2 Товари другого споживання....27

1.2.8 Нерухомість....28

1.2.9 Нафтогазовий сектор....29

1.2.10 Сировинний сектор....30

1.2.11 Комунальні послуги....31

1.3 Визначення показників для вирішення поставленої задачі.....36

1.4 Висновки після першого розділу.....41

РОЗДІЛ 2. УПРАВЛІННЯ ПРОЕКТОМ ОБРОБКИ ІНФОРМАЦІЇ....42

2.1 Уточнення необхідних показників для реалізації рекурентної нейронної мережі та LSTM моделі із додатковим залученням формул розрахунку акцій.....42

2.2 Засоби обробки інформації.....53

2.3 Розробка проекту проведення аналітичної діяльності.....55

2.4 Висновки після другого розділу....56

РОЗДІЛ 3. РЕАЛІЗАЦІЯ МЕТОДІВ ОБРОБКИ ІНФОРМАЦІЇ НА ОСНОВІ АКЦІЙНИХ ДАНИХ КОМПАНІЙ....57

3.1 Визначення та візуалізація вхідних даних акційних компаній на вибір.....57

3.2 Представлення поетапної обробки та операції із даними засобами обробки інформації.60

3.3 Візуалізація результату аналізу.....66

3.4 Висновки після третього розділу.....75

ВИСНОВОК....76

ПЕРЕЛІК ВИКОРИСТАНИХ ІНФОРМАЦІЙНИХ ДЖЕРЕЛ....78

ДОДАТОК....82

ВСТУП

Актуальність дослідження

Щоденно, світ змінюється у дві сторони, де кожний із цього дня має можливість стати історичним або бути як звичайним регулярним днем, як для людей так і для соціальних інформаційних новин. На даний момент, людство зараз на порозі розвитку як у космічних, технологічних, електронно-комерційних, підприємств, інформаційних та інвестиційних галузях. До уваги можна виділити середньостатистичну людину, котра більше 10 років відала та працює в одній із цій сфері і кожній людині, хотілось би мати мрію/ї, без мрій по суті людина не може співіснувати у соціальному суспільстві. Виходячи з цього, для досягнення таких мрій є декілька запроваджених в людському житті термінів морально-етичним ресурсам яких, стає – час та простір із особовим матеріальним фактором – гроші.

Беручи до уваги той факт, що в основному людство не стоїть на місці, та кожного дня проходять крізь руки та мережі багато цифрових інформаційних даних котрі були не так вже й давно запровадженні, більшу частину впливових людей із поставленими власними цілями можна сказати запровадило або простіше, вирішило на цьому непогано заробляти. Оскільки ми знаємо, що акції компаній ще було реалізовано в фізичному еквівалентному обсязі у 1602 році.[4] Поступово за такий короткий час як 400 років дані – «цінні папери» - перейшли до режиму як, торгівля фантиками, бо окрім цього висловлення нічого не може прийти до думки. Як один із елементів даного заробітку на даних може бути як інвестиції, в будь що, в нашому випадку елементом даного дослідження являється як один із чинних та стабільних можливостей не виходячи із власного будинку як – заробітку з інвестицій.

Інвестиції зумовлені тим, що дана частка особового гаманця людини віддається на ризикову, можна сказати як азартну можливість підняти свою фінансову допоміжну еквівалентність в розрізі від тієї суми котрі були інвестовані, це може бути як підприємства, компанії із володінням окремою конкретною часткою природних ресурсів, політичні гілки і тощо.

Акція – з кожним днем ці слова все частіше зустрічаються в медіа і звучать в побуті українців, сама акція являє собою як частку у компанії або долю пакету.

Інвестор, який купив акцію або акції, стає співвласником бізнесу, він може претендувати на частину його прибутку та майна. Чим більший прибуток отримає емітент, тим більше заробить акціонер. Емітент – це особа, юридично відповідальна, яка несе від імені або від імені публічно-правової(в нашому чині з огляду дослідження від лиця компанії або підприємства), зобов'язання перед власниками цінних паперів щодо здійснення прав, закріплених випущеними цінними бумагами. З іншого боку, якщо компанія збанкрутує, то й вартість акцій впаде до нуля, тоді інвестор втратить гроші. Простіше кажучи, акція — це частка в володінні бізнесом.

Тим часом за останні двадцять або ж тридцять років курси акцій у компаній гігантів на початку їх заснування – виросли майже в декілька мільйони разів, а світові акційні фонди та брокери масово скуповують та продають акції компаній і вкладають в їх розвиток мільйони доларів із можливою часткою бути співвласником.

Завдяки відповідальним, професіоналам, які знаються на факторних питаннях як ЗМІ, людський фактор, надійні джерела інформації які мають те, що не можна розповідати у ЗМІ, бачення наперед і тощо, що для спростування окремих механізмів було реалізовано як один із допоміжних інструментів як нейронні мережі - це сама існуюча електронна версія людського мозку, здатна швидше і краще вирішити будь-яке завдання, що володіє необмеженим потенціалом і взагалі, володіє активним функціоналом навчання.

Було одне «але». До пори до часу, нейронні мережі мешкали тільки в заповідних математичних пакетах, в дуже низько-рівневому вигляді, тішачи математиків і вчених графіками в **MatLab-ax, Mapple, Mathcad-ax** та **Statistica**.

Нейронні мережі не є машиною для пророкувань. Нейронні мережі – це свого роду дуже хороші апроксиматори. Вважається, що нейронна мережа може

апроксимувати практично що завгодно. При одній лише умові - якщо це «щось» піддається апроксимації.

І тут-то початківець дослідник і попадається на гачок когнітивних спотворень. Перша і головна помилка в тому, що історичні дані по котируваннях здаються чимось більшим, ніж просто статистикою. За ним можна намалювати стільки трикутників і стрілок пост-фактом, що тільки сліпому при погляді на це, буде неочевидно, що в цьому всьому є якась логіка, на яку просто не вдалося вчасно звернути увагу. Але яку, можливо, пізнає Машина.

Мета дипломної магістерської роботи – аналіз та розробка моделі прогнозування з метою поліпшити та розширити інвесторам можливість бачення на реалії стану акційного ринку.

Об'єкт дослідження – сфери прогнозування акцій, акційні датасети у розрізі щоденних вартостей компаній.

Предмет дослідження – інформаційний засоби, інструменти, моделі та методи управління акційними даними при аналізі та прогнозуванні їх вартостей на короткостроковий період.

Наукова новизна роботи – модель з графічним прогнозуванням вартості акцій компанії, котра базується на основі реальних отриманих даних, що сформовано для поліпшення українським та міжнародним інвесторам за метою заробітку в інвестиційних фондах.

Методи дослідження – інформаційні аналіз, математичні та рекурентні моделі для вирішення даної задачі, статистичні методи та інструменти для обробки інформації, інструменти для візуалізації даних, середовище збереження даних.

Практичне значення – реалізація за програмним середовищем із ціллю прогнозування вартості акцій компаній у короткостроковий період, що базується на мові програмування Python.

РОЗДІЛ 1. ПОСТАНОВКА ЗАДАЧІ ТА АНАЛІЗ РІШЕННЯ

1.1 Огляд предметної сфери та постановка задачі.

Фондова біржа – оформлений організаційно, чинний ринок постійного обігу, купівлі та продажу, на якому здійснюється торгівля цінними паперами товариством акціонерів формуванню їх біржового курсу та відтворює свою діяльність відповідно до чинного законодавства, статуту та правил фондової біржі. В кожній країні із ринковою економікою є національна фондова біржа. Фондові біржі можуть бути некомерційними організаціями, відкритими чи закритими акціонерними товариствами, державними чи недодержавними організаціями. В світі налічується більше ніж 250 фондових бірж, серед яких у 16-ти – капіталізація перевищувала 1 трильйон доларів. У Світову федерацію бірж входять 63 фондові біржі. Брати участь у торгах на біржі можуть фізичні чи юридичні особи, або ті та інші. Більшістю бірж розраховуються за фондовим індексом, обчислювання з урахуванням цін певної групи цінних паперів.

Фондовий індекс – зведений індекс, що обчислюється на основі цін певної групи цінних паперів – "індексного кошика". При розрахунку індексу початкове (базове) значення може бути сумою цін або прирівнюватися до довільного числа (наприклад, 100 або 1000). Для забезпечення сумісності вартості часто множать на **додаткові коефіцієнти**. Тому абсолютні значення індексів не є важливими. Важливе значення мають зміни індексу з часом, що дозволяє судити про загальний напрямок руху цін в індексному кошику, незважаючи на те, що ціни акцій всередині «індексного кошика» можуть змінюватися в різному напрямку. Залежно від принципу покладеного в основу вибору цінних паперів для індексу, він може відображати вартісну динаміку групи цінних паперів, об'єднаних за якоюсь ознакою (наприклад, висока, середня, мала капіталізація акцій) обраного сектору ринку (наприклад, телекомунікації), або широкого ринку акцій загалом.

Фондовий індекс не є єдиним цінним папером, тому зазвичай говорять про значення або рівень індексу, але не про ціну. Водночас фондові індекси часто є основою однойменних похідних фінансових інструментів (індексних

ф'ючерсів або опціонів), які використовуються для інвестиційних та спекулятивних цілей або для хеджування ризиків. У цьому значення індексу сприймається як вартість цього інструмента.[1]

За даними агентства Dow Jones & Co. Inc., на кінець 2004 року у світі налічувалося 2316 фондових індексів.

Основні види цінних паперів, які обертаються на біржі, — це **акції** й **облігації**, державні боргові зобов'язання, а також похідні від них. Процедура допуску їх до котування на фондовій біржі називається лістингом (допуск цінних паперів до торгів на біржі), вона здійснюється згідно з конкретними вимогами до їх емітента. Цінні папери, які пройшли процедуру лістингу і дозволені до купівлі-продажу, називаються фондовими цінностями (активами). Торгівля ведеться або партіями, на певну суму, або за видами, при цьому договори укладаються «заочно», без наявності на біржі самих цінних паперів.

Акції - вид цінних паперів, що являє собою право та свідоцтво про власності на певну частку капіталу акціонерного товариства та надає його власнику певні права, зокрема: право участі в управлінні товариством, а в разі ліквідації - на частину залишкової вартості підприємства.

Існує декілька видів акцій:

- За черговістю отримання доходу та обсягу на права власника

1. Проста акція являє собою свідоцтво та право власності на певну частку (за наявності на ті рішення загальних зборів) та брати участь у управлінні товариством.
2. Привілейована Привілейована акція — цінний папір, який передбачає фіксований дохід та обмежену участь власника у справах компанії. Привілейовані акції аналогічні до облігацій: вони дають майже гарантований дохід. Утримувачі привілейованих акцій отримують строго фіксовану суму чи відсоток. Обсяг дивідендів залежить від прибутку підприємства, на відміну звичайної акції. Привілейовані акції надають стабільний дохід. Ще одна особливість відхилення

привілейованих акцій від звичайних. Тримач звичайної акції може брати участь в управлінні компанією та голосувати на зборах акціонерів. Його права, як правило, набагато ширші за ті, які надають власникам привілейованих акцій. Втім, у ключових питаннях власники привілейованих акцій мають право голосу. Так, вони можуть накласти вето на рішення про злиття чи поглинання компанії.

- За характером розпорядження:

1. Іменні акції - цінні папери, видані на ім'я конкретної особи і пройшли реєстрацію в реєстрі товариства акціонерів. Іншими словами, на цих акціях має бути вказано ім'я їхнього власника. Вказана особа – правноправний член ради акціонерів компанії.
2. Акція на пред'явника – власником вважається фактичний утримувач. Передача права власності здійснюється шляхом простого її вручення. Акціонерне товариство жодним чином не фіксує обіг. Дивиденти видаються пред'явнику акцій або на підставі відповідного документа депозитарію, що засвідчує кількість акцій даного власника, що знаходяться на збереженні. У книзі реєстрації акцій фіксується лише загальна кількість випущених акцій на пред'явника. Чинне українське законодавство не допускає існування акцій на пред'явника (з 2005 року).[9]

Облігації – емісійний цінний папір, що засвідчує внесення його власником грошових коштів та підтверджує зобов'язання відшкодувати йому номінальну вартість цього цінного паперу з виплатою певного доходу або передати йому майно, надати послуги. Облігації – один із найдавніших цінних паперів та способів залучення грошових коштів чи покриття дефіциту бюджету.

Види облігацій:

- Державні облігації – цінний папір, емітентом якого виступає держава в особі уряду чи уповноваженого ним органу, випускається для погашення дефіциту державного бюджету;

- Корпоративні облігації – цінний папір, який випускають органи місцевої влади з метою покриття дефіциту місцевого бюджету або фінансування інфраструктури проектів;
- Муніципальні облігації – цінний папір, який випускають юридичні особи для фінансування своєї діяльності.[10]

Величезними біржами світу є Нью-Йоркська, Лондонська і Токійська, на них обертаються акції декількох тисяч великих корпорацій, які складають більш 75 % всього фінансового капіталу, що обертається в світі. Американський, англійський, японський і тайванський біржові ринки цінних паперів — найбільші в світі за сумою цін акцій, що котируються, і за обсягом укладених договорів.

Наприкінці назви фондових індексів може стояти число, що відображає кількість акціонерних компаній, на підставі яких розраховується індекс: CAC 40, Nikkei 225, S&P 500, DJIA.

Фондовий індекс Доу-Джонса (DJIA) – один із основних фондовий індекс цінних паперів (акцій) 30 найбільших американських підприємств, створених у 1896 році редактором The Wall Street Journal та співзасновником Dow Jones & Company Чарльзом Доу. Індекс був створений для зведення інформації щодо акцій індустріальних підприємств на американській біржі цінних паперів. Він залишається разом із Dow Jones Transportation Average найстарішим індексом цінних паперів у США.

В основі індексу Доу-Джонса — теорія Доу про аналіз ринку за допомогою середніх показників котирувань промислових та транспортних акцій. Вважається, що тенденція ринку в цілому позитивна, якщо один із цих середніх показників піднімається вище за попередній локальний пік, за яким настає аналогічний зростання іншого показника. Коли ж обидва показники знижуються нижче за попередній локальний мінімум, це підтверджує загальну тенденцію до спаду. Цю теорію покладено в основу прогнозування майбутніх змін на фондовому ринку.[8]

Фондовий індекс **S&P 500** – у кошик якого включено 500 обраних торгуються на фондових біржах США громадських підприємств, мають

найбільшу капіталізацію. Список належить компанії Standard & Poor's і нею складається.[7]

Фондовий індекс **Nikkei 225** – один із найважливіших фондових індексів Японії. Індекс обчислюється як просте середнє арифметичне цін акцій 225 найбільш активних торгів компаній першої секції Токійської фондової біржі.[6]

Фондовий індекс **CAC 40** – є найважливішим фондовим індексом Франції. Індекс обчислюється як середнє арифметичне виважене капіталізації значення цін акцій 40 найбільших компаній, акції які торгуються на біржі Euronext Paris.[5]

Вартість цінних паперів на біржі формується внаслідок подвійного аукціону чи ранжирування заяв у межах ціни закупівлі (нижня межа) та ціни продажу (верхня межа). Кожен договір передбачає новий аукціон. Біржові ціни фіксуються кілька разів на день і щодня публікуються у пресі, так само як і інформація про укладені договори. Щорічні інформаційні збірники про фондову біржу випускаються Всесвітньою федерацією фондових бірж та Федерацією бірж Європейського Співтовариства.

Основне завдання даної кваліфікаційної магістерської роботи дослідити інформаційні показники (дані) вартості акцій компаній та надати прогнозуючий графік.

Для реалізації поставленого завдання необхідним є вирішення таких задач, як:

1. Аналіз статистичних існуючих джерел із даними, незалежних джерельних авторів із варіативністю на подальший розвиток акцій та прогнозування падіння чи росту акційних фондів, бірж тощо.
2. Відокремлення основних факторів впливу на вартість акцій компаній для повного розуміння.
3. Збір актуальних даних в інформаційному просторі.
4. Демонстрація та аналіз статистичних даних.
5. Розробка моделі на основі нейронної мережі із даними для

прогнозування.

6. Демонстрація результатів дослідження.

1.2 Аналіз принципу формування вартості акції компанії.

Основні принципові показники секторів бізнесу/компанії/галузі є :[26]

1. **Інформаційні технології.**
2. **Охорона здоров'я.**
3. **Фінансовий сектор.**
4. **Комунікаційні послуги**
5. **Товари другої потреби.**
6. **Промисловий сектор.**
7. **Споживчі товари першої потреби.**
8. **Нерухомість.**
9. **Енергетика, тобто нафтогазовий сектор.**
10. **Сировинний сектор.**

Список з 10 секторів повністю покриває виробничий ланцюжок товару, який умовно можна розділити на п'ять етапів: [20]

- Видобуток та первинна переробка.
- Виробництво.
- Транспортування.
- Зберігання.
- Поширення та споживання.

У капіталістичній моделі діє правило: чим ближче у виробничому ланцюжку ми знаходимося до кінцевого споживача, тим вища рентабельність. Тобто чим далі ми уникаємо видобутку, тим глибший ступінь переробки сировини і тим вища додана вартість.

Це умовно ділить світ на два економічні ділянки: ядро і периферію, тобто розвинені країни і ті, що розвиваються, технологічні і сировинні. Цей поділ був особливо наочним у колоніальний час.

Не дивно, що найбагатша країна зараз США, де домінує ідеологія споживання. Саме американці знаходяться у найбільш технологічному та розважальному рівні.

А в периферійних економіках відбувається придушення технологічних галузей та внутрішнього виробництва. Крім того, видобуток та експорт сировини — це діяльність із спадним відштовхувальним фактором: розширення виробництва дає дедалі менший приріст доходу. Тому ресурсні країни часто приречені бовтатися на периферії: так зване сировинне прокляття.

Потрібно розуміти **ризики**, пов'язані з **вкладенням** на ринки, що розвиваються.

Циклічність говорить нам про те, що показники компанії пов'язані з етапами бізнес-циклу – зазвичай це 10 років. Циклічні компанії йдуть за тенденцією в економіці, за всіма приливами та відливами — відповідно такі акції більш ростуть: вони агресивно дорожчають у період розширення економіки, а потім падають під час її охолодження.

1.2.1 IT

У сектор входять компанії, пов'язані з розробкою технологічних продуктів: електроніки, програмних забезпечень, комп'ютерів та компонентів. Це найбільший сектор в індексі S&P 500, що становить чверть його капіталізації.

У IT-секторі виділяють шість галузей:

1) Програмне забезпечення. Розробники ПЗ для бізнесу та споживачів: корпоративні програми, системи управління базами даних, відеоігри, освітні програми.

2) IT-послуги. Компанії, що надають такі послуги, як консалтинг, обробка даних, аутсорсинг.

3) Напівпровідники: Intel, Microchip Technology, Nvidia, Qualcomm.

4) Устаткування зв'язку. Виробники обладнання для локальних мереж, маршрутизаторів, комутаторів та телекомунікаційних компонентів.

5) Апаратні засоби та компоненти. Компанії, що виробляють мобільні телефони, ПК, периферійні пристрої та компоненти: материнські плати, аудіо- та відеокарти, монітори, клавіатури та принтери.

6) Електронне устаткування. Постачальники електронних елементів, таких як трансформатори, конденсатори, резистори, котушки та пристрої зв'язку. Виробники лазерів, сканерів штрих-кодів, систем безпеки. Дистриб'ютори, що постачають технологічне обладнання іншим компаніям.

Технологічний сектор - основний драйвер зростання американського ринку в останнє десятиліття, він показав за цей час зростання майже 300%. Будь-який великий бізнес залежить від технологічних рішень та інновацій, потребує автоматизації, створення корпоративного програмного забезпечення, управління своїми логістичними системами, захисту своїх

даних. Споживчі товари, такі як персональні комп'ютери, мобільні пристрої, побутова техніка, телевізори постійно вдосконалюються і з новими функціями продаються споживачам.

Sector	Last	10 Year	10 Year Price Range	
S&P 500 Index	3,900.79	+199.44%	1,266.74	4,818.62
S&P 500 Information Technology	2,282.62	+414.12%	432.79	3,113.22
S&P 500 Health Care	1,492.11	+256.74%	408.52	1,672.39
S&P 500 Consumer Discretionary	1,116.01	+228.03%	326.90	1,699.37
S&P 500 Financials	547.15	+192.78%	178.59	693.52
S&P 500 Industrials	765.40	+158.01%	285.75	912.42
S&P 500 Materials	518.27	+141.41%	205.69	580.07
S&P 500 Consumer Staples	730.82	+114.65%	340.27	849.55
S&P 500 Utilities	361.84	+100.35%	170.86	392.38
S&P 500 Communication Services	194.22	+39.88%	130.82	289.48
S&P 500 Energy	616.28	+29.70%	172.45	738.71
S&P 500 Real Estate	261.69	N/A	N/A	N/A

Рисунок 1.2.1 – Загальний вигляд секторів за 10 років на платформі barchart.com

1.2.2 Охорона здоров’я

Це компанії, які надають медичні послуги, які займаються дослідженнями у цій сфері, виробники медичного приладдя.

У секторі виділяють шість галузей:

1) Медичне обладнання. Виробники спеціалізованого обладнання, інструментів, ортопедичних приладів, діагностичних апаратів - рентгенівських, МРТ, - а також витратних матеріалів, таких як лікарняне приладдя та шприци.

2) Медичні послуги. Сюди входять дистриб'ютори, що розповсюджують обладнання оптом, та будь-які медичні установи, включаючи лікарні, будинки для людей похилого віку, реабілітаційні центри та лікарні для тварин. Сюди відносять компанії, що надають послуги пацієнтам: центри діалізу та лабораторних аналізів.

3) Медичні технології. Компанії, що надають медичним організаціям IT-послуги: спеціалізоване ПЗ, консалтинг та інтернет-сервіси.

4) Біотехнології. Компанії, які займаються дослідженнями, розробкою та просуванням продуктів генної інженерії. Включно з компаніями, що надають білкові методи лікування проти таких захворювань, як рак або діабет.

5) Фармацевтика. Компанії, що займаються дослідженнями, розробкою та виробництвом фармацевтичних препаратів, включаючи ветеринарні.

6) Обладнання та послуги для досліджень. Постачальники аналітичних інструментів, витратних матеріалів, послуг із клінічних випробувань для фармацевтичних та біотехнологічних компаній.

Сектор охорони здоров'я — один із найбільших та найскладніших в економіці США, на його частку припадає майже п'ята частина загального

валового внутрішнього продукту. Прогресуюче старіння населення розвинених країн сприяє зростанню попиту цьому сегменті.

Але тут є специфічні ризики, пов'язані з випробуваннями нових препаратів, конкуренцією, закінченням патентів, діями регуляторів, тому акції біотехнологічних компаній зазвичай дуже мінливі.

Сектор вважається нециклічним і не так сильно залежить від стану економіки, адже люди не вибирають, коли на нього хворіти. Тому Healthcare проявляє себе краще за ринок у фазі заходу сонця і просідає не так сильно під час рецесії.

1.2.3 Фінансовий сектор

Компанії сектору надають фінансові послуги бізнесу та роздрібним клієнтам. Це банки, інвестиційні та страхові організації. Фінансовий сектор – третій за розмірами в індексі S&P 500. Акції фінансових компаній – популярний актив у портфелі інвестора, більшість виплачують високі дивіденди.

Сектор включає такі галузі:

1) Банки та установи, що спеціалізуються на іпотеці.
2) Страхування. До цієї групи входять страхові компанії, брокери та компанії з перестрахування. Перестрахування — особливий вид страхування, до якого вдаються страхові компанії, щоб краще керувати своїми ризиками.. Перестрахування захищає страхову компанію від банкрутства у разі форс-мажорів чи великих катастроф.
3) Споживчі фінанси: постачальники споживчих послуг, як особисті позики, кредитні картки, лізинг автомобілів, ломбарди, управління інвестиціями, брокерські послуги.
4) Ринок капіталу. Це фінансові ринки, де формується попит і пропозиція на середньостроковий та довгостроковий позичковий капітал.
5) Різні фінансові послуги. Група фінансових установ, які не вписуються в інші категорії, у тому числі агенції кредитного рейтингу та фондові біржі.

Виділяють ще дві галузі — іпотечні REIT та іпотечні фінанси, — але їхня вага надто незначна.

Оскільки фінансовий сектор надає кредити для бізнесу та споживачів, його статки часто відображають стан справ в економіці. Якщо фінансові компанії слабкі, то економіка йде на спад.

Коли ставки низькі, бізнес починає активно кредитуватися для нових капітальних проектів та інвестицій. Тому фінансовий сектор добре почувається на ранній стадії ділового циклу.

Помірне підвищення відсоткових ставок позначається на секторі сприятливо, але відсоткові ставки, що швидко зростають, можуть стати негативним фактором, тому що можуть призвести до зниження попиту на кредитування. Крім цього негативний вплив на фінансовий сектор може вплинути на державне регулювання, якщо воно посилює бюрократичні процедури. І навпаки, при пом'якшенні умов фінансові послуги стають доступнішими і сектор процвітає.

1.2.4 Комунальні послуги

Наймолодший і найсуперечливіший сектор, який в індексі займає трохи більше 10%. Виділяють п'ять галузей:

1) Різні бездротові та дротові послуги, включаючи організацію інтернет-зв'язку по оптоволоконних мережах, передачу голосів і даних.
2) Бездротовий та мобільний зв'язок.
3) Медіа. Компанії, що надають маркетингові та рекламні послуги, радіо- та ТВ-холдинги, видавництва, публікуючі книги, газети, журнали у друкованому та електронному вигляді.
4) Розваги. Компанії, які виробляють, транслюють та розповсюджують розважальні продукти: телешоу, фільми та музику.
5) Інтерактивні послуги: соціальні мережі, пошукові системи, компанії, що створюють та розповсюджують контент через свої платформи.

Зазвичай телекоми — це захисні папери нециклічного характеру. Діяльність компаній переважно орієнтована на внутрішній попит, часто вони працюють за підпискою, що забезпечує постійний виторг. Це дозволяє компаніям стабільно виплачувати високі дивіденди.

Довгий час телевізор існував у S&P 500 як окремий сектор, займаючи найменшу його частку - близько 2%. Але 21 вересня 2018 відбулася реорганізація індексу і телеком став галуззю в рамках новоствореного Communication Services.

Сюди ж перекочували кілька десятків провідних компаній із IT, зокрема Facebook з Google, та з Consumer Discretionary: CBS, Disney, Netflix, Comcast. Реорганізацію провели, щоб збалансувати частки секторів і знизити капіталізацію сектору IT, що сильно розрісся.

Реорганізація призвела до того, що сектор, насправді, став циклічним. Тому складно виділити чіткі закономірності його поведінки стосовно етапів бізнес-циклу. А при складанні портфеля галузь телекому, як і раніше, варто розглядати як окремий сектор.

1.2.5 Товари другої потреби

До циклічних товарів та послуг відносяться речі **другої потреби** та багато товарів тривалого користування: автомобілі, меблі, предмети розкоші, брендовий одяг, ресторани, готелі, авіаквитки. Споживачі активно користуються ними, коли економіка процвітає, але якщо купівельна спроможність падає, від них відмовляються насамперед. Відповідно, доходи циклічних компаній у цей час скорочуються. Якщо спад затяжний, а компанія має слабкі фундаментальні показники, стає можливим її банкрутство.

Нециклічні компанії, навпаки, демонструють стійкий виторг незалежно від економічної ситуації, тому що надають товари та послуги першої необхідності. Навіть у найважчі часи люди продовжують використовувати продукти харчування, медикаменти, тютюн, алкоголь, електроенергію, воду та газ.

Тому сектори **споживчих товарів першої необхідності**, електроенергетики та охорони здоров'я можна віднести до захисних: вони консервативніші і не відчувають різких коливань.

1.2.6 Промисловий сектор

Компанії сектору надають промислові продукти та послуги. Сектор тісно взаємопов'язаний із сировинним, але між ними є різниця: по-перше, продукція промислового сектору не є сировиною. По-друге, у промисловому секторі велика кількість компаній надають сервіс, тоді як у сировинному секторі зосереджені виробники товарів.

Промисловий сектор ділиться на чотирнадцять галузей, які утворюють три великі галузеві групи.

Capital Goods, капітальні товари, тобто ті, що не продаються кінцевим споживачам, а задіяні у виробництві інших продуктів. Це оборонне, промислове та сільськогосподарське обладнання, інжиніринг, цивільне будівництво, торговельні послуги. Сюди входять такі галузі:

1) Авіакосмічна та оборонна: цивільні та військові літаки, оборонна електроніка, радари, космічні прилади;
2) Будівельні матеріали та обладнання: скловолокно, вентиляційні системи, двері - крім пиломатеріалів, цементу та інших матеріалів, що потрапляють у сировинний сектор;
3) Будівництво та інженерія: нежитлові споруди, мости тощо. Сюди не входять забудовники житлових будівель, вони належать до Consumer Discretionary;
4) Електричне обладнання: турбіни, генератори, кабелі, дроти;
5) Промислові конгломерати: підприємства з діяльністю у трьох чи більше індустріях;
6) Тяжка техніка: сільськогосподарська, промислова, преси, ліфти, верстати;
7) Торгові компанії та дистриб'ютори;
8) Комерційні та професійні послуги. Тут дві галузі:

8.1) Комерційні послуги. Учасники цієї галузі надають іншим компаніям офісні, друковані, охоронні послуги, а також послуги харчування, прибирання та вивезення сміття. Іншими словами, це постачальники послуг B2B – бізнес для бізнесу;

8.2) Рекрутингові агенції, консалтинг та дослідження для бізнесу;

Перевізники, куди входять такі галузі:

9) Вантажні авіаперевезення та логістика;

10) Пасажирські авіалінії;

11) Водні перевезення - вантажні та пасажирські. Сюди не належать круїзні компанії;

12) Наземні та залізничні перевезення;

13) Транспортна інфраструктура: компанії, що керують аеропортами та морськими терміналами, дорогами, мостами, тунелями тощо.

Промисловий сектор економічно чутливий, він найкраще показує себе на ранній фазі ділового циклу, коли нарощуються темпи промислового виробництва, а фазі рецесії стає аутсайдером нарівні з Real Estate і IT. Винятком може бути оборонна промисловість, де бюджет та замовлення верстаються на кілька років уперед.

1.2.7.1 Товари першого споживання

Сектор споживчих товарів включає компанії, що виробляють товари першої необхідності. Це базові продукти, від яких люди не можуть відмовитись навіть у важкому фінансовому становищі.

У секторі виділяють шість галузей:

1) Роздрібна торгівля товарами першої необхідності: продуктовий рітейл та аптеки, а також дистриб'ютори, що не працюють із

кінцевими споживачами.

2) Їжа: виробники будь-якого типу продуктів від сировинних до упакованих товарів.
3) Напої, алкогольні та безалкогольні.
4) Тютюнова продукція.
5) Товари для дому та побуту: продукція нетривалого користування, наприклад мило, миючі засоби, паперові рушники та серветки.
6) Продукти для особистої гігієни: підгузки, шампуні та косметика.

Сектор краще за інших почувається на заході сонця і під час рецесії. Відрізняється стабільністю, не циклічний. Акції зазвичай дивідендні, із стійким, але не вражаючим зростанням. Must have для довгострокового дивідендного портфеля з низькою мінливістю, особливо в очікуванні рецесії.

1.2.7.2 Товари другого споживання

Компанії сектору пропонують товари другої необхідності, і навіть послуги, яких можна тимчасово відмовитися при економічному спаді. Це авто, меблі, побутова техніка, готелі, ресторани, кіно, телебачення, одяг.

Виділяють такі галузі:

1) Автокомпоненти, шини та ін.
2) Автомобілі та мотоцикли.
3) Житло та товари для дому: меблі, побутова техніка та інша продукція тривалого користування.
4) Товари для відпочинку та спорту, іграшки, велосипеди та ін.
5) Одяг та дорогі аксесуари: виробники одягу, взуття, сумок, аксесуарів, але не обов'язково продають їх безпосередньо споживачеві.
6) Готелі, ресторани, відпочинок. Сюди входить готельний бізнес, казино,

турагентства, круїзні лінії, фітнес-центри та тематичні парки.

7) Різні споживчі послуги: освітні, юридичні та дизайнерські послуги, аукціони, весільні та похоронні бюро.
8) Дистриб'ютори та оптові продавці перерахованих вище товарів.
9) Інтернет-рітейл, продаж за каталогами, магазин на дивані.
10) Багатопрофільна роздрібна торгівля, універмаги
11) Спеціалізований рітейл: магазини вузького спрямування, що представляють конкретний тип товару або бренд.

Сектор циклічний і безпосередньо залежить від стану економіки. Найкраще він показує себе на ранній фазі ділового циклу, а на заході сонця просідає.

Коли настають важкі часи, споживачі затягують пояси, відкладаючи відпустки та другорядні придбання. І навпаки, коли економіка зміцнюється, люди відчувають упевненість та витрачають гроші. Зростання заробітної плати та збільшення кредитування сприяють цьому. При сприятливій кон'юнктурі сектор зростає швидше за індекс.

1.2.8 Нерухомість

Це один з маленьких секторів, але він може стати чудовим інструментом диверсифікації портфеля інвестора через низьку кореляцію з іншими секторами, а також засобом захисту від інфляції.

У секторі виділяють дві галузі:

1) Компанії, що пропонують різні послуги у сфері нерухомості: управління, оренда, девелопмент, продаж.
2) REIT (Real Exchange Investment Trusts) - компанії, що інвестують у нерухомість: офісні будівлі, житлові будинки, торгові центри, готелі,

лікарні та будинки для людей похилого віку. Дохід вони отримують від зростання вартості нерухомості, так і від здачі власності в оренду. За законом REIT зобов'язуються розподіляти між акціонерами не менше 90% прибутку в обмін на податкові пільги та розподіл між акціонерами, тому ці емітенти мають високу дивідендну прибутковість. По суті, купуючи акції REIT, інвестор вкладає в нерухомість тільки з високою ліквідністю і низьким порогом входу.

Як і нафтогазовий сектор, сектор нерухомості - відмінний хеджуючий інструмент від інфляції. А ще основним джерелом доходів тут служить орендна плата, яка, як правило, слідує за загальним зростанням цін.

Сектор економічно чутливий, тому поганий під час рецесії, а найкраще поводиться на ранній фазі ділового циклу.

1.2.9 Нафтогазовий сектор

У **нафтогазовий сектор** входять підприємства, що функціонують у геологічній розвідці, видобутку, первинній переробці, транспортуванні, зберіганні нафти, газу та іншого споживаного палива. А також компанії, що пропонують супутні послуги та бурове обладнання.

У секторі виділяють дві галузі:

- Обладнання та сервіс.
- Нафта, газ та інші енергоносії.

Суспільство не може функціонувати без палива, тому ціна нафти дуже корелює з інфляцією і часто є її основним драйвером. І навпаки, щойно купівельна спроможність грошей слабшає, зростає цінність матеріальних активів, таких як нерухомість, сировинні товари, вуглеводні (компоненти нафти та газу).

Нафтогазові акції – це хороший балансуючий інструмент для портфеля з великою кількістю високодивідендних акцій та облігацій із фіксованим купоном.

Склалось, що нафтогазовий сектор — один із найгірших на ранньому етапі бізнес-циклу, коли економіка лише відроджується. Це пов'язано з низькою інфляцією та низькими цінами на сировину. З іншого боку, нафта — це ресурс, який живе за своїми законами. Тут багато політики, сентименту та маніпуляцій, тому інструмент важко піддається прогнозуванню. Якийсь військовий конфлікт може різко підштовхнути ціну, а надлишок нафти та економічне уповільнення веде до здешевлення на деяких світових ділянках, що ми й спостерігаємо зараз. Здавалося б, настільки стратегічна сировина має бути максимально стабільною і передбачуваною, але на практиці вона дуже мінлива: ціна бареля з 20$ у 1999 році піднялася до 140$ до 2008 року, а через рік знизалась до 50$. І навість зараз перебуває у нестабільному стані..

1.2.10 Сировинний сектор

Компанії із **сировинного** сектору займаються металами, хімічними та лісовими виробництвами, а їх типові продукти – добрива, пластик, фарби, клей, бетон, алюміній, деревина. Це матеріали, які активно використовуються у будівництві, та на які покладаються інші покладаються інші промисловості. Тому сектор дуже корелює з будівельними галузями. Ще цей сектор дуже залежить від котирувань на товарних біржах.

У сировинному секторі виділяють п'ять галузей:

Гірничодобувна промисловість: Геологічна розвідка, видобуток, первинна обробка металів та мінералів – це:

- Алюмінієва руда, вона ж боксит.
- Солі, фосфати, кольорові метали: мідь, свинець, нікель, титан та цинк.
- Фосфати, що використовуються у виробництві добрив та засобів

для чищення.

- Дорогоцінні метали та мінерали, у тому числі платина, золото та срібло.
- Сталь, включаючи видобуток коксівного вугілля, що використовується у виробництві сталі.

Таблиця 1

Хімічні продукти	Матеріали для будівництва	Тара та упаковка	Папір та продукти з дерева
Пластмаси, синтетика, нафтохімічні продукти	цемент	пляшки	-
Спеціальні добавки, клей, покриття, полімери	пісок	металеві банки	-
Добрива, пестициди та інша сільськогосподарська хімія	цегла	картонні коробки	-
Промислові гази такі як азот, водень, гелій	вапно	ковпачки	-
Інша хімія	гіпс	пробки та ін	-

З одного боку, данні підгалузі, наприклад металурги, болісно реагують на промисловий спад, але з іншого — деякі хімічні та пакувальні підгалузі вплетені у процеси захисних секторів.

1.2.11 Комунальні послуги

Сюди входять п'ять галузей:

1) Електроенергетика. Компанії, що генерують та розподіляють

електроенергію, включаючи атомні електростанції. Часто діяльність компаній жорстко регулюється через стратегічну важливість, і в них значна частка держави.

2) Постачальники газу. Сюди не враховують компанії, які займаються геологічною розвідкою та видобутком - це прерогатива нафтогазового сектору.

3) Гідроенергетика, водопостачання для побутового та комерційного використання. Включає очисні споруди. Вода - природна монополія, тому водоканали зазвичай знаходяться під муніципальним керуванням.

4) Диверсифікована енергетика: мультисервісні компанії, що пропонують широкий спектр комунальних послуг. Часто це холдинги, які володіють поруч з дочірніми підприємствами.

5) Відновлювана енергетика. Ці компанії виробляють та розподіляють електроенергію від відновлюваних джерел, таких як сонячна, геотермальна, вітрова енергія. Сюди не включені виробники сонячних батарей та технологічних елементів, вони входять до сектору IT.

Сектор вважається захисним, тому що комунальні компанії протягом усього бізнес-циклу мають стабільний виторг. Акції зазвичай не мінливі, компанії платять непогані дивіденди. А ще індустрія відновлюваної енергетики є дуже перспективною.

Сектор комунальних послуг добре проявляє себе під час кризи, а в період розгону економіки, на ранньому етапі циклу, зазвичай втрачає популярність, тому що інвестори віддають перевагу більш агресивним та зростаючим активам.

У компаній-комунальників є кілька суттєвих мінусів.

Великі борги. Зазвичай у комунальників розгорнута і дорога інфраструктура, а через неї велика заборгованість на балансах. Це робить компанії сектору особливо чутливими до підвищення відсоткової ставки: якщо позики стають дорожчими, це веде до збільшення операційних витрат.

Такі організації потребують ефективного та раціонального управління. Так як їм необхідний постійний приплив коштів для модернізації та капітальних витрат, вони часто вдаються до позик та розміщення облігацій. Занадто високе навантаження може зіпсувати кредитний рейтинг компанії, а це ускладнить отримання нових позик.

Компанії працюють на основі довгострокових угод про придбання електроенергії за фіксованою ставкою. Якщо витрати на виробництво зростають, компанії продовжують продавати комунальні послуги за узгодженою ставкою, що знижує їхню маржу.

Компанії піддаються ретельному нагляду і можуть легко підняти тарифи, збільшивши прибуток. Доводиться працювати у жорстко регульованих рамках.

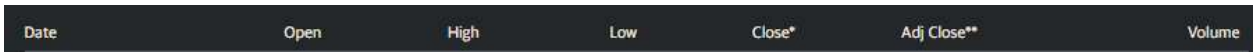


Рисунок 1.2.2 – Загальний вигляд атрибутів акцій на фондовому ринку платформи

Де, атрибути стовпців – це, Data(Дата), Open(Відкриття акції на фондовому ринку), High(Найвища ціна, за конкретний період), Low(Найнижча ціна, за конкретний період), Close(Закриття акції на фондовому ринку), Adj Close(збереження останньої незавершеної/завершеної операції в момент обробки та закриття біржі), Volume – Скоригована ціна закриття(являє собою ціну закриття після коригування для усіх застосованих розподілів і дивідентів).

Рисунок 1.2.3 – Загальний вигляд веб-додатку на фондовому ринку finviz.com

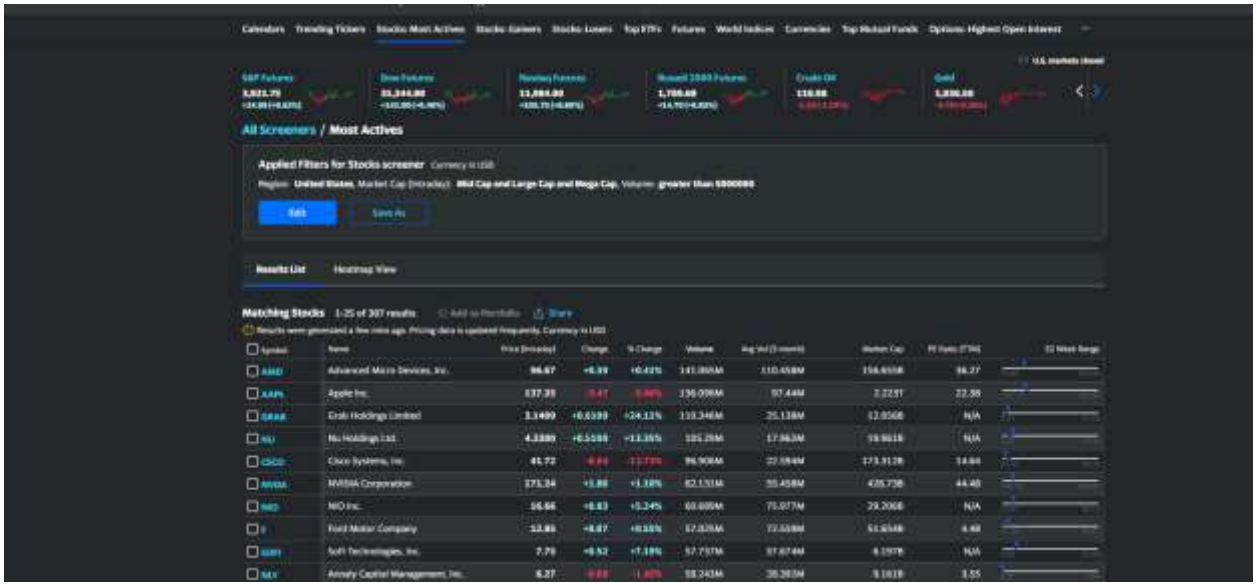


Рисунок 1.2.4 – Загальний вигляд веб-додатку на фондовому ринку finance.yahoo.com

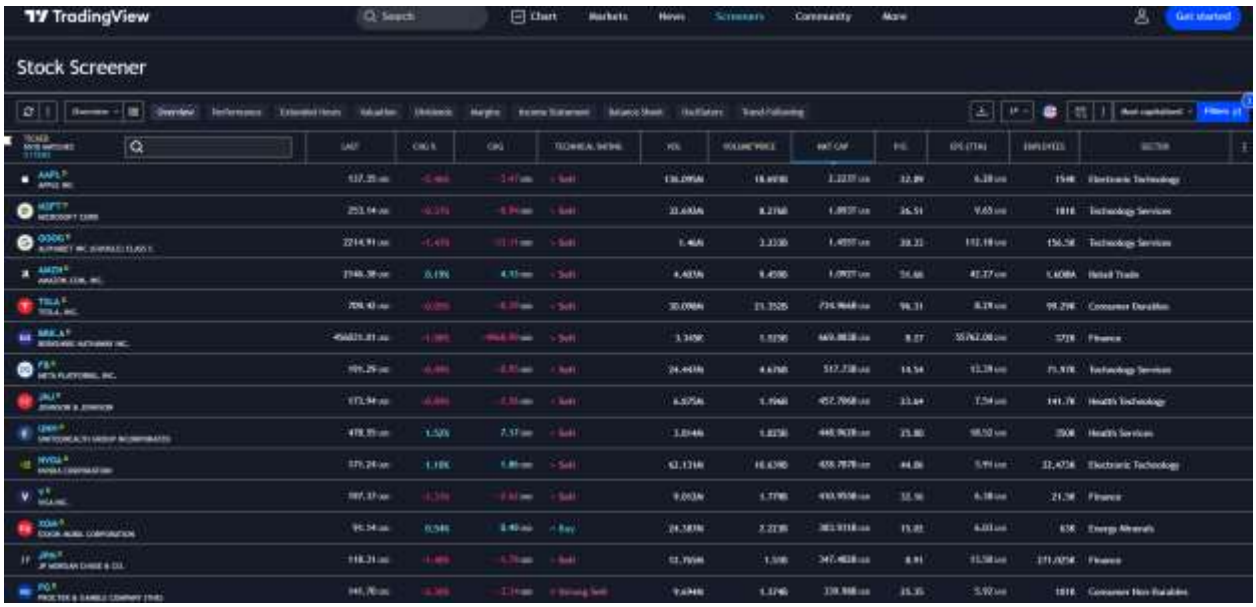


Рисунок 1.2.5 – Загальний вигляд веб-додатку на фондовому ринку tradingview.com

Дані платформи є одними із популярніших додатків для перегляду усіх можливих інтернет-бізнесових операцій, як криптовалюта, акції та вартості на природні ресурси – такі як, нафта, газ, золото, залізо і тощо.

1.3 Визначення показників для вирішення поставленої задачі.

Ціни акцій визначаються на маркетплейсах, де пропозиція продавця відповідає попиту покупців. Але чи замислювалися ви коли-небудь про те, що рухає фондовим ринком, тобто які фактори впливають на ціну акцій? На жаль, немає чіткого рівняння, яке б точно вказувало нам, як буде поводитися ціна акції. Тим не менш, ми знаємо деякі речі про сили, які рухають акції вгору або вниз. Ці ключові категорії поділяються на декілька категорії: фундаментальні фактори, технічні фактори та економічні фактори і тощо.

Попит

Фактори попиту, які можуть впливати на ціни акцій, включають новини та результати діяльності компанії, економічні фактори, тенденції в галузі, настрій ринку та несподівані події, такі як стихійні лиха або всесвітня криза.

Попит дає вартість акцій. Якщо немає попиту на акції компанії, вони просто не будуть мати цінності або звичайної рентабельності .

Будь-які новини навколо компанії — не/очікувані, можуть спричинити колосальні зміни цін акцій. Наприклад, звіт про прибутки, який показує конкретний показник прибутку за квартал або місяць, дає зрозуміти, що, запуск нового продукту, може знаходитись в не або досягаємій цілі, зазначеної в плані керівників і може призвести до коливань попиту та цін на акції. Навіть гуманітарні або природні кризи можуть спричинити збої в бізнесі та збільшити борг компанії, тобто зменшити попит або збільшити попит в розрізі масової скупівлі товару в критичний момент.

Економічні фактори

Економічні фактори, включаючи зміни відсоткових ставок, фінансові перспективи та інфляцію, впливають на ціни акцій. Якщо відсоткова ставка та інфляція зростають, а економічні перспективи слабшають, попит, як правило, зменшиться, а ціна акцій, ймовірно, знизиться. Якщо буде схоже, що економіка буде розвиватися, ціни на акції можуть зрости. Інвестори можуть купувати більше акцій, думаючи, що побачать майбутні прибутки та вищі ціни

на акції. Якщо економічні перспективи невизначені, інвестори можуть зменшити наплив купівель та інвестицій до своїх покупок або почати продавати їх в цілому.

Тенденції галузі

Галузеві тенденції часто визначають ціну акцій, оскільки компанії в одній галузі часто працюють подібним чином і піддаються однаковому тиску. Отже, коли галузь процвітає, попит на акції в цьому конкретному секторі часто збільшується, підвищуючи ціни на акції. Також можливе збільшення попиту на акції однієї компанії, якщо у конкурента погані показники.

Ринкові настрої

Ринкові настрої відносяться до психології учасників ринку, як окремо, так і колективно. Це, мабуть, найбільш неприємна категорія. Ринкові настрої часто суб'єктивні, упереджені та вперті. Наприклад, ви можете зробити тверде судження про майбутні перспективи зростання акцій, і майбутнє може навіть підтвердити ваші прогнози, але тим часом ринок може короткозорко зупинятися на одній новині, яка тримає акції штучно на високому або низькому рівні. І іноді можна довго чекати в надії, що інші інвестори помітять фундаментальні фактори із подальшим розвитком акції.

Відносно, нова сфера поведінкових фінансів досліджує настрій ринку. Воно починається з припущення, що ринки, мабуть, неефективні велику частину часу, і цю неефективність можна пояснити психологією та іншими дисциплінами соціальних наук. Ідея застосування соціальних наук до фінансів була повністю узаконена. Багато ідей у поведінкових фінансах підтверджують очевидні підозри: інвестори схильні переоцінювати дані, які легко спадають на думку; що багато інвесторів більш болісно реагують на збитки, задоволено на еквівалентні прибутки; і що інвестори, як правило, продовжують робити помилятись.

Деякі інвестори стверджують, що можуть отримати вигоду з теорії поведінкових фінансів. Для більшості, однак, ця категорія є достатньо новою, щоб служити «всеохоплюючою», де зберігається все, що ми не можемо

пояснити або долучити.

Різний приорітет

Різні типи інвесторів залежать від різних факторів. Короткострокові інвестори та трейдери, як правило, враховують і можуть навіть надавати пріоритет технічним факторам. Довгострокові інвестори віддають перевагу фундаментальним і визнають, що технічні фактори відіграють важливу роль. Інвестори, які твердо вірять у фундаментальні принципи, можуть примиритися з технічними силами із таким популярним аргументом: технічні фактори та настрої ринку часто переважають у короткостроковому періоді, але фундаментальні фактори встановлять ціну акцій у довгостроковій перспективі. Тим часом ми можемо очікувати більш захоплюючих змін у сфері поведінкових фінансів, тим паче, що традиційні фінансові теорії не можуть пояснити все, що відбувається на ринку.

Технічні фактори

Все було б простіше, якби ціни на акції встановлювали лише фундаментальні фактори. Технічні фактори – це сукупність зовнішніх умов, які змінюють попит і пропозицію акцій компанії. Деякі з них опосередковано впливають на основи. Наприклад, економічне зростання опосередковано сприяє зростанню доходів.

До технічних факторів можна віднести наступні категорії як:

1) **Інфляція**

Ми згадували про це раніше як вхідну інформацію в коефіцієнт оцінки, але інфляція також є величезною силою руху з технічної точки зору. Історично низька інфляція мала сильну зворотну кореляцію з оцінками (низька інфляція призводить до високих кратних, а висока інфляція — до низьких множників). З іншого боку, **дефляція**, як правило, погано впливає на акції, оскільки зазнає втрату цінової вартості та усієї влади для та над компанією.

2) **Економічна сила ринку та аналогів**

Акції компаній, як правило, співвідносяться з ринком і з аналогами в секторі чи галузі. Деякі відомі інвестиційні фірми стверджують, що поєднання

загальних рухів ринку та сектору — на відміну від індивідуальних показників компанії — визначає більшу частину руху акцій. (Дослідження показало, що на економічні/ринкові фактори припадає 90 відсотків.) Наприклад, раптово негативний прогноз щодо однієї роздрібної акції часто завдає шкоди іншим роздрібним акціям, оскільки «провина через асоціацію» гальмує попит у всьому секторі.

3) Замінники

Компанії конкурують за інвестиційні долари або євро з іншими класами активів на глобальній арені. До них належать корпоративні облігації, державні облігації, товари, нерухомість та іноземні акції. Зв'язок між попитом на міжнародні акції та їх замінниками важко визначити, але це має важливу роль.

4) Випадкові транзакції

Випадкові транзакції — це купівля або продаж акцій, які мотивуються чимось іншим, ніж віра у внутрішню вартість акцій. Ці операції включають виконавчі інсайдерські операції, які часто заплановані заздалегідь або обумовлені цілями особистого портфеля покупця. Інший приклад — брокери, які купують або закривають акції, щоб хеджувати інші інвестиції. Хоча ці операції можуть не являти собою офіційні «голоси», подані за чи проти акцій, вони впливають на попит і пропозицію, а отже, можуть змінювати ціну.

5) Демографічні показники

Було проведено деякі важливі дослідження щодо демографії інвесторів. Значна частина цього стосується цих двох динамік:

- Інвестори середнього віку, найбільші заробітки, які схильні інвестувати на фондовому ринку;
- Старші інвестори, які схильні виходити з ринку, щоб задовільнити власні потреби для виходу на пенсію;

Гіпотеза полягає в тому, що чим більша частка інвесторів середнього віку серед населення, що інвестує, тим більший попит на акції та вищі коефіцієнти оцінки.

6) Тенденції

Часто акція просто рухається відповідно до короткострокового тренду. З одного боку, акція, яка рухається вгору, може набирати обертів, оскільки «успіх породжує успіх», а популярність підсилює акції. З іншого боку, акція іноді веде себе протилежним чином у тренді і робить те, що називається поверненням до середнього.

На жаль, оскільки тенденції змінюються в обох напрямках і більш очевидні заднім числом, знання про те, що акції є «модними», не допомагає нам передбачити майбутнє.

7) Ліквідність

Ліквідність – важливий і іноді недооцінений фактор. Це стосується того, який інтерес з боку інвесторів привертає конкретна акція. Акції Wal-Mart, наприклад, є високоліквідними і, таким чином, дуже реагують на суттєві новини; середня компанія з малою капіталізацією менша. Обсяг торгів є не тільки показником ліквідності, але це також функція корпоративних комунікацій (тобто ступінь, до якої компанія привертає увагу спільноти інвесторів).

Акції з великою капіталізацією мають високу ліквідність — за ними добре відстежують і здійснюють великі операції. Багато акцій малої капіталізації страждають від майже постійної «знижки на ліквідність», оскільки їх просто немає на екранах радарів інвесторів.

1.4 Висновки після першого розділу.

В першому розділі було поставлено до даної магістерської кваліфікаційної роботи за задачу, дослідити інформаційні показники даних вартості акцій компаній та надати прогнозуючий графік.

Було проаналізовано принцип роботи фондових та біржових ринків продажу акцій. Та визначено загальну кількість галузевих факторів, які мають вплив на акції компаній, котрі можуть бути потребуючими або зв’язані через сировину, окремих видів ресурсів, матеріалів тощо.

РОЗДІЛ 2. УПРАВЛІННЯ ПРОЕКТОМ ОБРОБКИ ІНФОРМАЦІЇ

2.1 Рекурентна нейронна мережа та LSTM модель із додатковим занесенням математичних формул із залученням прогнозування акцій.

Рекурентна нейрона мережа – являє собою один із видів, де зв’язки між елементами утворюють направленні послідовні кроки.

Навчання РНМ аналогічне із навчанням звичайної нейронної мережі. В ньому може використовуватись алгоритм зворотного поширення помилок. Це використовується в тому випадку, коли подібні йому параметри використовуються на усіх часових періодах в мережі, результат залежить на кожному виході не тільки від розрахунків теперішнього кроку, а й від минулих часових кроків.

Для повного розуміння із подальшим аналізом нашої майбутньої роботи із пакетами даних акцій компаній слід застосовувати такі підрозділи науки як Нейроні Мережі або ж Машинне навчання.

В моєму випадку було обрано нейрону мережу LSTM, що являє собою зв'язки між рекурентними нейронними мережами із можливістю зберігати дані на довгий час.

РНМ – додає пам’ять до штучної нейронної мережі, але реалізувати дану пам’ять виходить коротким кроком – де на кожному кроці навчання, інформація в пам’яті поринає до нової та через декілька ітерацій повністю перезаписується. [14]

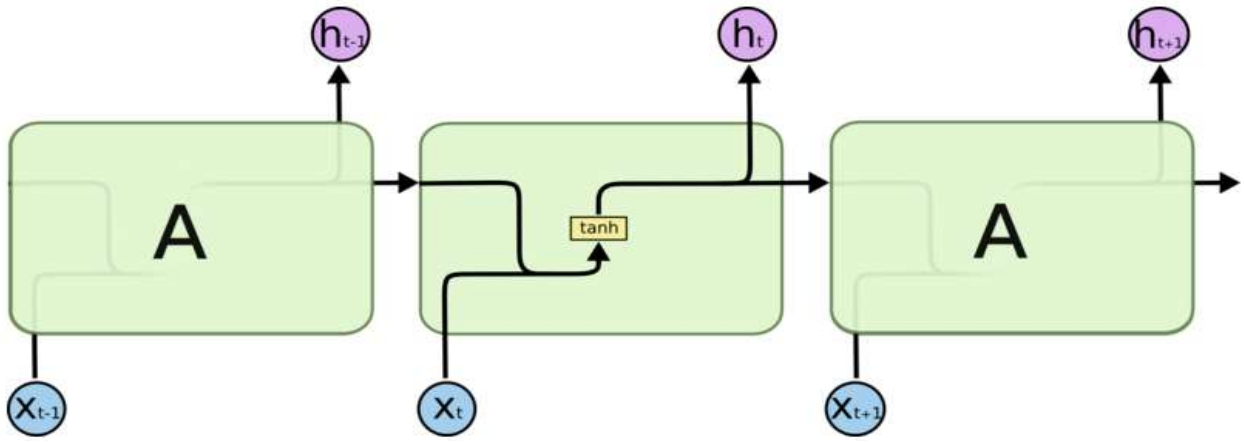


Рисунок 2.1.1 Схема слою рекурентної мережі.

Ключем до цієї можливості є те, що LSTM-модуль не використовує функції активації всередині своїх рекурентних компонентів. Таким чином, значення, що зберігається, не розмивається в часі.

LSTM-модулі розроблені спеціально, щоб уникнути проблеми довготривалої залежності, запам'ятовуючи значення як на короткі, так і довгі проміжки часу. LSTM-модулі часто групуються в блоки, що містять різні LSTM-модулі. Подібний пристрій характерний для «глибоких» багатошарових нейронних мереж та сприяє виконанню паралельних обчислень із застосуванням відповідного обладнання.

LSTM-блоки містять три або чотири «вентилі», які використовуються для контролю потоків інформації на входах та на виходах пам'яті даних блоків. Ці вентилі реалізовані як логістичної функції для обчислення значення діапазоні [0; 1]. Розмноження цього значення використовується для часткового допуску чи заборони потоку інформації всередину і назовні пам'яті. Наприклад, «вхідний вентиль» контролює міру входження нового значення пам'ять, а «вентиль забування» контролює міру збереження значення пам'яті. «Вихідний вентиль» контролює міру того, якою мірою значення, що знаходиться в пам'яті, використовується при розрахунку вихідної функції активації для блоку.[15]

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \text{ (1)}$$

Перший крок у LSTM – визначити, яку інформацію можна викинути зі стану комірки. Це рішення приймає сигмоїдний шар, званий шаром фільтра забування (forget gate layer). Він дивиться на h_{t-1} та x_t повертаючи число від 0 до 1 для кожного числа зі стану осередку. 1 означає "повністю зберегти", а 0 - "повністю викинути". Якщо ж число буде мати кому, число залишається для подальших ітерацій.

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \text{ (2)}$$

Наступним кроком є вирішити, яка нова інформація буде зберігатися в стані комірки. Цей етап складається з двох частин. По-перше, сигмоїдний шар під назвою «Вхід шару вхідного шару») визначає, які значення слід оновити.

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \text{ (3)}$$

Потім він створює вектор нових цінностей - кандидатів, які можна додати до стану елементу.

$$c_t = f_t * c_{t-1} + i_t * \sigma_c(W_c x_t + U_c h_{t-1} + b_o) \text{ (4)}$$

Наступним кроком ми перемножуємо старий стан на $f_t * c_{t-1}$.Потім додаємо $i_t * \sigma_c$. Це нові значення елементів, помножені на t – на скільки ми хочемо оновити кожне з значень стану.

$$h_t = o_t * \sigma_h(c_t) \text{ (5)}$$

Зрештою, потрібно вирішити, яку інформацію ми хочемо отримувати на виході. Вихідні дані будуть засновані за нашим станом елементу, до них будуть використані деякі фільтри. Спочатку ми застосовуємо сигмоїдний шар, який вирішує, яку інформацію зі стану осередку ми виводитимемо. Потім значення стану осередку проходять через tanh-шар, щоб отримати на виході значення діапазону від -1 до 1, котрі у свою чергу перемножуються з вихідними значеннями сигмоїдного шару, що дозволяє виводити тільки необхідну інформацію.

x_t – вхідний вектор;

h_{t-1} – вихідний вектор;

c_t – вектор стану;

W, U, b – матриці параметрів і вектор;

i_t, f_t, o_t – вектори вентелів;

f_t – вектор вентеля забування, вага запам'ятовування старої інформації;

o_t – вектор вхідного вентиля, вага отримання нової інформації;

i_t – вектор вихідного вентиля, кандидат на вихід;

σ_h – на основі гіперболічного тангенсу, але в роботі із оглядовими ділянками;

σ_g – на основі сигмоідів;

σ_c – на основі гіперболічного тангенсу.

Для мінімізації загальної помилки LSTM на всій множині тренувальних послідовностей, ітеративний градієнтний спуск такий як метод зворотного поширення помилки розгорнутий у часі може бути використаний для зміни кожного з ваг пропорційно його похідної залежно від величини помилки. Головною проблемою градієнтного спуску для стандартних рекурентних нейронних мереж є те, що градієнти помилок зменшуються з експоненційною швидкістю зі збільшенням тимчасової затримки між важливими подіями.

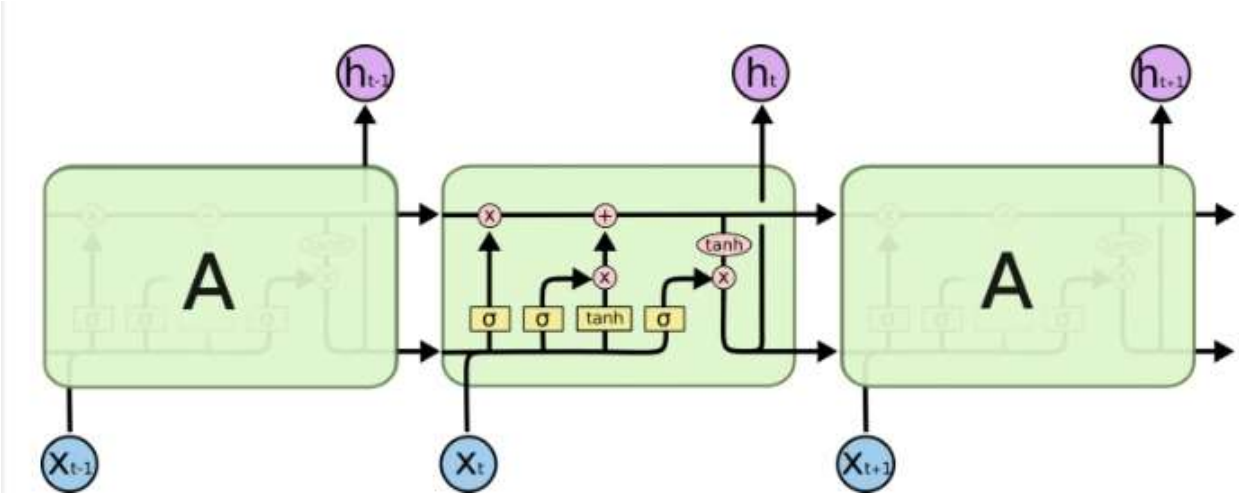


Рисунок 2.1.2 Стандартна модель LSTM із 4-ма взаємопов'язаних шарів передачі інформації.

Курс акції — її ціна на продаж, яка в умовах вільного ринку не є постійною величиною і яка встановлюється або на фондовій біржі, або на позабіржовому ринку. Суто теоретично: вона (ціна продаж акції) прямо пропорційна величині дивіденду і обернено пропорційна до рівня позичкового відсотка (тобто доходу, одержуваного вкладниками банку) - акція продається за таку суму грошей, яка при її поміщенні до банку принесе дохід не нижче дивіденду.

На практиці ціна акції формується, переважно, за принципом "попит-пропозиція/година експонування пропозиції-попиту".[11] Ціна акції постійно коливається навколо своєї нормальної величини. Під час зростання виробництва, коли прибуток, а разом з ним дивіденти зростають, курс акції підвищується. Коли відбувається спад економіки, дивіденди скорочуються, що

впливає на курс акції до різкого його зниження. Таким чином ринок цінних паперів сигналізує про загальне «здоров'я» економіки. На ціну акції теоретично має вплив інфляція (проте, як і на справедливу вартість будь-якого іншого фінансового інструменту).

Документ, за яким акціонерне товариство проводитиме виплату дивідендів, називається «купонний лист» (існування останнього, власне, не є обов'язковим).

$$U = \frac{D}{B} * 100\% \ (6)$$

U — *Курс акції;*

D — *Дивіденд;*

B — *Банківський відсоток;*

Формування вартості

Зрозуміти принцип попиту та пропозиції просто; набагато складніше розібратися, що утворює попит. Коливання ціни показує уявлення інвесторів щодо важливості компанії на фондовому ринку, але як вони визначають цю важливість? На їхнє судження, безумовно, впливає поточна дохідність: тобто, скільки прибутку приносить компанія. Але інвестори часто виходять за межі цифр. Тому вкладення інвесторів нерідко показують як поточну цінність компанії, так і перспективи зростання, якого від фірми чекають у майбутньому.

Існують спеціальні кількісні методи, використовувані для оцінки подібних перспектив, - їх називають "моделями дисконтування дивідендів" (DDM). "Дисконтування" у цьому контексті означає певну ставку – тобто кількість грошей, яку інвестору потрібно заплатити саме зараз за право отримувати майбутній прибуток.

Завдяки своїй прямолінійності однією з найпопулярніших моделей стала модель зростання Гордона, розроблена у 1960-х роках американським економістом Мірон Гордон.

Рівняння цієї моделі можна виразити так: (ставка дисконтування – темп

зростання) / (кількість дивідендів) = поточна вартість.

Або дробом:

$$P = \frac{D_1}{(r-g)} \quad (7)$$

P — поточна вартість;

g — темпи зростання майбутнього прибутку;

r — норма прибутку протягом року, підтримує акціонерний капітал всієї компанії;

D_1 — величина прибутку наступного року.

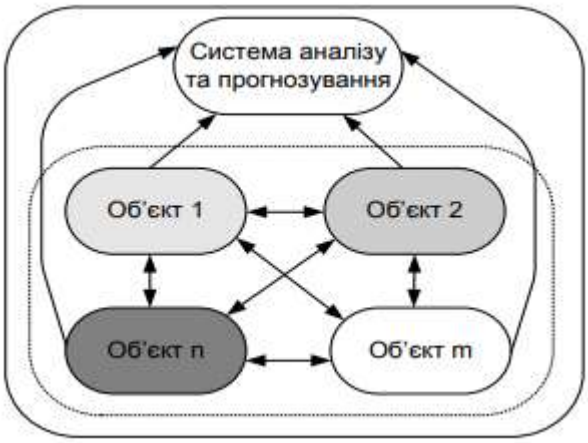


Рисунок 2.1.3 Схема рівноправного обміну даними

Ковзні середні

Ковзні середні вважаються одним з найпопулярніших і найпростіших індикаторів технічного аналізу.

Існує безліч видів ковзних середніх (математики називають ковзними середніми всі функції, які працюють за подібною ознакою), але їх поєднує одне - всі вони працюють із середніми значеннями, причому значення будь-якої ковзної середньої визначається зі значення вихідної функції за попередній період.

Ковзні середні - це індикатори, що запізнюються, які згладжують коливання і шум, тим самим допомагають трейдерам бачити поточний тренд фінансових активів.[22]

Типи ковзних середніх:

Проста ковзна середня (Simple moving average або SMA) - середнє арифметичне ціни паперу за певний період.

$$SMA = \frac{\sum_{i=1}^{n} PRICE_i}{n} \quad (8)$$

n – кількість одиничних періодів;

$PRICE_i$ – поточне значення ціни, в налаштуваннях індикатора можна вибрати такі варіанти: high, low, open, close, median price.

$$mp = \frac{h+l}{2} \quad (9)$$

mp – median price, середня ціна;

h – high price, найвища ціна;

l – low price, найнижчя ціна.

$$tp = \frac{h+l+c}{3} \quad (10)$$

tp – typical price, типова ціна

C – close, вартість за якою було зачинено акцію по фіксованій ціні.

$$wc = \frac{h+l+c+c}{4} \quad (11)$$

Wc - $weighted\ close$, середнє зважена від c.

У простої ковзної середньої багато недоліків:

- запізнення (на жаль, майбутнє туманно)
- всі значення мають один і той також пріоритет, незалежно від їхньої давності.

Але останні дані більш важливі для оцінки активу, ніж старі і вони повинні більш впливати на ковзну середню. Щоб реалізувати таку модель, вигадали експоненційну ковзну середню (Exponential moving average або EMA).

$$EMA = PRICE_i * F + (1 - F) * EMA_{i-1} (12)$$

EMA_{i-1}- минуле значення;

F - фактор згладжування (частка використання значення цін). Коефіцієнт F вибирається довільним чином у межах від 0 до 1;

При такому розрахунку останнім даним приділяється більше уваги, тому експоненційна ковзна середня більше реагує на зміни цін.

Ще два типи: це об'ємно-залежна ковзна середня (Volume adjusted moving average або VMA) і згладжена ковзна середня (Smoothed moving average або SMMA). Формули:

$$VMA(i) = \frac{\sum_{i=1}^{n} P(i) * V(i)}{\sum_{i=1}^{n} V(i)} (13)$$

$V(i)$ – значення обсягу угод;

$P(i)$ – поточне значення ціни;

$$SMMA(i) = \frac{SMMA_{i-1} * (n-1) + PRICE_i}{n} (14)$$

$SMMA_{i-1}$ – попереднє значення;

Відмінності:

Якщо експоненційна ковзна середня ставила нові значення пріоритетніших за старі, то об'ємно-залежна ковзна середня дає більшу вагу тим значенням ціни, за якими були максимальні обсяги угод.

Згладжена ковзна середня, як видно за назвою, згладжує свої значення і йде більш плавно.

Кореляція

Кореляція - це один з основних термінів теорії ймовірності, що показує міру залежності між двома та більш випадковими величинами. Ця залежність виражається через коефіцієнт кореляції. Коефіцієнт кореляції набуває значень від -1 до +1. Що значення коефіцієнта кореляції, то більше вписувалося залежність між величинами. Кореляція буває позитивною та негативною.

Кореляція - це статистичний взаємозв'язок двох або кількох випадкових величин (або величин, які можна з деяким допустимим ступенем точності вважати такими). У цьому змін значень однієї чи кількох із цих величин супроводжують систематичній зміні значень інших величин. Математичною мірою кореляції двох випадкових величин служить кореляційне відношення, або коефіцієнт кореляції. У разі, якщо зміна однієї випадкової величини не веде до закономірної зміни іншої випадкової величини, але приводить до зміни іншої статистичної характеристики цієї випадкової величини, то подібний зв'язок не вважається кореляційним, хоча є статистичним.

Випадкові величини X і Y можуть бути залежними, або незалежними. Якщо ці величини залежні, то зміна значень однієї або декількох з цих величин ведуть до зміни значень іншої або інших величин.

У випадку, якщо зв'язок виявиться суттєвим, то необхідно скористатися методом регресійного аналізу. Регресійний аналіз дозволяє визначити характер зв'язку та надалі побудувати її математичну модель.

$$r = \frac{\sum(x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 * \sum(y_i - \bar{y})^2}} \quad (15)$$

r – коефіцієнт кореляції;

x_i – значення х-змінної у вибірці;

$\bar{x}$ – середнє значення х-змінної;

y_i – значення у-змінної у вибірці;

$\bar{y}$ – середнє значення у-змінної.

Регресійний аналіз

Регресійний аналіз (regression analysis) – це метод вивчення статистичного взаємозв'язку між однією залежною кількісною залежною змінною від однієї або кількох незалежних кількісних змінних. Залежна змінна в регресійному аналізі називається результуючою, а змінні фактори - предикторами або перемінними.

Взаємозв'язок між середнім значенням результуючої змінної та середніми значеннями предикторів виявляється у вигляді рівняння регресії. Рівняння регресії – математична функція, яка підбирається на основі вихідних статистичних даних залежної та пояснювальної змінних. Найчастіше використовується лінійна функція. І тут говорять про лінійному регресійному аналізі.

Регресійний аналіз дуже тісно пов'язаний із кореляційним аналізом. У кореляційному аналізі досліджується напрям та тіснота зв'язку між кількісними змінними. У регресійному аналізі досліджується форма залежності між кількісними змінними. Тобто. Практично обидва способи вивчають ту саму взаємозв'язок, але з різних сторін, і доповнюють одне одного. Насправді кореляційний аналіз виконується перед регресійним аналізом. Після доказу наявності взаємозв'язку методом кореляційного аналізу можна висловити форму зв'язку з допомогою регресійного аналізу.[27]

Мета регресійного аналізу – за допомогою рівняння регресії передбачити очікуване середнє значення результуючої змінної.[28]

Основні завдання регресійного аналізу:

- визначення виду та форми залежності;
- оцінка параметрів рівняння регресії;
- перевірка значущості рівняння регресії;
- перевірка значущості окремих коефіцієнтів рівняння;

- побудова інтервальних оцінок коефіцієнтів;
- дослідження показників точності моделі;
- побудова точкових та інтервальних прогнозів результуючої змінної.

Як і кореляційний аналіз, регресійний аналіз відображає лише кількісні залежності між змінними. Причинно-наслідкові залежності регресійного аналізу не відображають. Гіпотези про причинно-наслідковий зв'язок змінних повинні формулюватися і обґрунтовуватися виходячи з теоретичного аналізу змісту явища, що вивчається.

$$y = a + b * X \quad (16)$$

y – залежна змінна;

a – вільний показник (зсув по осі OY)

b – кутовий коефіцієнт. Він вказує на поведінку кривої (зменшується або зростає, кут між віссю)

X – незалежна змінна (пояснювальна).

2.2 Засоби обробки інформації.

Після отримання даного файлу або ж файлів, його потрібно обробити розробленим скрипт-програмою на мові програмування Python. Даний скрипт являє собою функціональну програму за для:

1) Загального переліку елементів, стовбців та строків масивів даних у датасеті;
2) Форматування в тому вигляді який нам потрібен(відпрацювання циклу із розпізнаванням невідомих елементів(таких як, лапки, дефіз, двокрапки, зірочки і т.д та їх заміни на пустий елемент);
3) Перевірка елементів датасету до відповідності формату даних, integer, char, float.
4) Занесення його до реєстру пам’яті нейронної мережі із подальшими цілями поставленої задачі;

Одним із найбільш затребуваних інструментів при використанні Python для розробки фінансових програм є пакет Pandas.[29] Він потрібен вже на початку, але в міру поглиблення в процес розробки знадобляться і такі пакети як NumPy, SciPy, Matplotlib. Пакет pandas-datareader дозволяє отримувати дані із таких джерел, як Google, Yahoo! Finance або Світовий банк.

Seaborn – бібліотека для візуалізації даних.

cv2 — неофіційні попередньо вбудовані пакети OpenCV для Python лише для ЦП.

Keras.layers - це високорівневий API для побудови та навчання моделей, що включає першокласну підтримку для TensorFlow - специфічної функціональності, такої як eager execution, функцій tf. data і Estimators.

Пакет sklearn. preprocessing надає кілька загальних службових функцій та класів перетворювачів для перетворення необроблених векторів ознак на уявлення, що більше підходить для наступних оцінювачів.

Matplotlib.pyplot - це повна бібліотека для створення статичних, анімованих та інтерактивних візуалізацій на Python. Matplotlib робить легкі речі легкими, а важкі-можливими.

- Створення якісних графіків;
- Можливість зробити інтерактивні фігури, які можна масштабувати, панорамувати, оновлювати;
- Налаштування візуального стилю та макету.;
- Експорт у багато форматів файлів;
- Вбудовуваний в JupyterLab і графічні інтерфейси користувача;
- Використання багатого набіру сторонніх пакетів, створених на Matplotlib.

2.3 Розробка проекту проведення аналітичної діяльності.

В рамках даної дипломної роботи розробка проекту включає в себе варіант вирішення задачі прогнозування акцій компаній.

В реалізації даної поставленої задачі не було залучено команду із н-ої кількості людей (або ж людський ресурс), так як існують обмеження в ресурсах, таких як: гроші, людський фактор, багато часу на реалізацію та створення команди, окреме приміщення при потребі, тощо.

По етапне розглядання проекту:

1) Цілі:
   - Реалізація прогнозування над акціями компаній;
   - Створення корисного матеріалу дослідження;
   - Долучення даної інформації до читачів, в вільних інтернет джерелах, у розрізі дослідження;
2) План був внесений до постановки задачі.
3) В моменту виконання даного проекту повинно використовуватись знання аналітичної діяльності.
4) В момент закриття проектної діяльності, аналітик повиннен зробити висновок із поставленими цілями, які ресурси були витрачені, тощо.

2.4 Висновки після другого розділу

В даному розділі було зазначено, що для подальшої реалізації задачі прогнозування повинно бути використані РНМ та LSTM модель, із застосунком, регресійного аналізу, кореляції, ковзної середньої та самого розрахунку акції.

Визначено основні пакетні інструментарії мови програмування Python із застосунком прогнозування та графічно-візуалізаційного методу відображення на моніторі користувача із подальшим самостійним інформаційним аналізом та донесення до інвестора даної графічної картини.

Визначено основні цілі та користі даного задуму.

РОЗДІЛ 3. РЕАЛІЗАЦІЯ МЕТОДІВ ОБРОБКИ ІНФОРМАЦІЇ НА ОСНОВІ АКЦІЙНИХ ДАНИХ КОМПАНІЙ

3.1 Визначення та візуалізація вхідних даних акційних компаній на вибір.

На розгляді було узято одну із популярних американських компанії, але проаналізовано усі – Amazon, Google, Microsoft, Apple. Набір даних має складати собою такі параметри прийняті до нормим за колонками як: «Відкриті», «Нижчі», «Високі», «Закриті», «Обсяг», «Ім’я компанії», «Скоригована ціна закриття(являє собою ціну закриття після коригування для усіх застосованих розподілів і дивідентів)».

Було виконаний процес пошуку інформації із таких джерел як Kaggle, **Numerai**, DrivenData і т.д для подальшого отримання та збору. На жаль, більшість із них містили застарілі дані для даних компаній і було вирішено узяти приближені дати до теперішнього часу, бо фактично, фізично їх отримати може бути можливим у випадку власного занесення або написання автономного бота із заповненням ключових полей у датасеті із теперішніми часовими показниками в цей момент.

Apple – компанія-гігант, яка є відомою на світ своїм логотипом та назвою корпорації, виробляє персональні та планшетні комп'ютери, аудіоплеєри, смартфони, програмне забезпечення. І також дана платформа не стоїть на місці і розвиває галузь у кіноіндустрії.[19]

Date	Close/Last	Volume	Open	High	Low
05/13/2022	$147.11	113,990,900	$144.59	$148.105	$143.11
05/12/2022	$142.56	182,602,000	$142.77	$146.2	$138.8
05/11/2022	$146.5	142,689,800	$153.5	$155.45	$145.81
05/10/2022	$154.51	115,366,700	$155.52	$156.74	$152.93
05/09/2022	$152.06	131,577,900	$154.925	$155.83	$151.49
05/06/2022	$157.28	116,124,600	$156.01	$159.44	$154.18
05/05/2022	$156.77	130,525,300	$163.85	$164.08	$154.95
05/04/2022	$166.02	108,256,500	$159.67	$166.48	$159.26
05/03/2022	$159.48	88,966,530	$158.15	$160.71	$156.32
05/02/2022	$157.96	123,055,300	$156.71	$158.23	$153.27
04/29/2022	$157.65	131,747,600	$161.84	$166.2	$157.25
04/28/2022	$163.64	130,216,800	$159.25	$164.515	$158.93
04/27/2022	$156.57	88,063,190	$155.91	$159.79	$155.38

Рисунок 3.1.1 Датасет компанії Apple.

Microsoft – компанія-гігант, одна з найбільших транснаціональних компаній з виробництва програмного забезпечення для різноманітних обчислювальної техніки — персональних комп'ютерів, ігрових приставок, КПК, мобільних телефонів та іншого. Розробник найбільш широко розповсюдженої на даний момент у світі програмної платформи. [18]

Date	Close/Last	Volume	Open	High	Low
05/16/2022	$261.5	32,550,930	$259.955	$265.82	$255.78
05/13/2022	$261.12	34,925,090	$257.35	$263.04	$255.35
05/12/2022	$255.35	51,033,800	$257.69	$259.88	$250.02
05/11/2022	$260.55	48,975,900	$265.68	$271.36	$259.3
05/10/2022	$269.5	39,336,420	$271.69	$273.75	$265.07
05/09/2022	$264.58	47,725,950	$270.06	$272.36	$263.32
05/06/2022	$274.73	37,780,320	$274.805	$279.25	$271.27
05/05/2022	$277.35	43,260,390	$285.54	$286.35	$274.34
05/04/2022	$289.98	33,599,320	$282.59	$290.88	$276.73
05/03/2022	$281.78	25,978,610	$283.96	$284.13	$280.15
05/02/2022	$284.47	35,151,110	$277.71	$284.94	$276.22
04/29/2022	$277.52	37,073,920	$288.61	$289.88	$276.5
04/28/2022	$289.63	33,646,570	$285.185	$290.98	$281.456
04/27/2022	$283.22	63,477,690	$282.1	$290.97	$279.16

Рисунок 3.1.2 Датасет компанії Mircosoft.

Date	Close/Last	Volume	Open	High	Low
05/16/2022	$2288.9	1,299,522	$2299.11	$2323.4	$2277.79
05/13/2022	$2321.01	1,751,924	$2290.66	$2357.5	$2272.1001
05/12/2022	$2256.88	2,691,809	$2227.55	$2285.9	$2196.49
05/11/2022	$2272.05	1,876,737	$2264.73	$2327.2882	$2264.73
05/10/2022	$2287.9	1,995,019	$2310.15	$2325	$2258.01
05/09/2022	$2250.22	2,040,143	$2265	$2301.5418	$2240.0101
05/06/2022	$2314.93	1,985,479	$2303.69	$2351.43	$2280.31
05/05/2022	$2330.11	2,292,029	$2404.08	$2420.78	$2300.115
05/04/2022	$2445.22	2,495,764	$2340.63	$2457.09	$2302.32
05/03/2022	$2346.68	1,248,368	$2328.61	$2368.8363	$2320.69
05/02/2022	$2331.66	1,776,651	$2268.1	$2334.91	$2252
04/29/2022	$2282.19	2,133,800	$2334.69	$2358.395	$2276.29
04/28/2022	$2370.45	2,444,232	$2328.3	$2392.99	$2282.3
04/27/2022	$2285.89	4,591,029	$2289.385	$2344.84	$2254.7301

Рисунок 3.1.3 Датасет компанії Google.

Google – компанія-гігант, корпорація у складі холдингу Alphabet, що інвестує в інтернет-пошук, хмарні обчислення та рекламні технології. Google підтримує та розробляє ряд інтернет-сервісів та продуктів. Вона вважається однією з технологічних компаній «великої четвірки» поряд з Amazon, Apple та Microsoft. [16]

Amazon – компанія-гігант, котра являється однією із найбільших у світі на ринкових платформах електронної комерції та публічно-хмарних обчислень за виручкою та ринковою капіталізацією. Дана платформа вже на теперішній час не тільки підкорила та зарекомендувала себе на ринку, а і зростає в інших галузях, таких як: відеоігри, кіноіндустрія і тощо. [17]

Date	Close/Last	Volume	Open	High	Low
05/16/2022	$2216.21	3,728,287	$2262	$2279.85	$2207.0827
05/13/2022	$2261.1	4,684,228	$2181.38	$2263.68	$2156
05/12/2022	$2138.61	6,601,257	$2055	$2215.615	$2048.11
05/11/2022	$2107.44	5,485,182	$2162.07	$2203.12	$2088.57
05/10/2022	$2177.18	5,271,691	$2225	$2252.85	$2143.42
05/09/2022	$2175.78	6,406,219	$2226.25	$2280	$2159.14
05/06/2022	$2295.45	6,212,995	$2297	$2381.01	$2261.625
05/05/2022	$2328.14	7,219,561	$2460	$2469.98	$2301.45
05/04/2022	$2518.57	5,537,268	$2472	$2520	$2383.66
05/03/2022	$2485.07	3,956,668	$2481.07	$2524.41	$2456.5
05/02/2022	$2490	7,439,449	$2448.02	$2493.365	$2367.5
04/29/2022	$2485.63	13,633,090	$2596.98	$2615.22	$2432.5
04/28/2022	$2891.93	5,865,847	$2843.56	$2918.75	$2806
04/27/2022	$2763.34	3,566,815	$2803.83	$2838.97	$2715.67

Рисунок 3.1.4 Датасет компанії Amazon.

3.2 Представлення поетапної обробки та операції із даними засобами обробки інформації.

На початку реалізації даної задачі нам потрібно імпортувати в середовище програмування мови Python, інструментарій, для коректної роботи із даними датасетів котрі зображені на Рисунках 3.1.1 – 3.1.4. Інструментарій можна побачити на Рисунку 3.2.1.

```
import os
import pandas_datareader as pdr
import numpy as np
import matplotlib.pyplot as plt
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import LSTM
import tensorflow as tf
import math
from sklearn.metrics import mean_squared_error
from numpy import array

import seaborn as sns
from keras.models import Sequential
from keras.layers import Dense, LSTM, Dropout
from keras.layers import *
```

Рисунок 3.2.1 Імпорт бібліотек для реалізації задачі в середовищі мові програмування Python.

Наступним етапом, ми виконуємо операцію створення DataFrame в середовищі програмування мови Python.

```
df = pd.read_csv('../input/apple-stock-price-history/AAPL Historical Data.csv', sep=',')
df['Date'] = df['Date'].apply(lambda x: to_strftime(x))
df = df.sort_values('Date').reset_index(drop=True)
```

Рисунок 3.2.2 Початок операцій із датасетами.

На даному етапі, вказуємо конкретні назви датасетів та сортуємо їх по датами.

```
tech_list = ['AAPL', 'GOOG', 'MSFT', 'AMZN']

end = datetime.now()
start = datetime(end.year - 1, end.month, end.day)

for stock in tech_list:
    globals()[stock] = yf.download(stock, start, end)
```

Рисунок 3.2.3 Початок операцій із датасетами.

На Рисунку 3.2.4 До DataFrame Price створюємо окремі умови для початку роботи нейронної мережі.

```
df['Price'] = df['Price'].astype(float)

num_shape = 1900
window = 60

train = df.iloc[:num_shape, 1:2].values
test = df.iloc[num_shape:, 1:2].values
df_price = np.vstack((train, test))

print(train.shape)
print(test.shape)
```

Рисунок 3.2.4 Підготовка даних до видоутворення та тестування моделі.

На рисунку 3.2.5 використовується частина коду для початку тренування моделей та задання умов реалізації моделі.

```
    for i in range(window_size, len(data)-prediction_days_in_future):
        x.append(scaled_data[i-window_size:i, 0])
        y.append(scaled_data[i+prediction_days_in_future-1, 0])

    x = np.array(x).reshape(len(x), window_size, 1)
    y = np.array(y)

    n_training_samples = int(len(y) * TRAIN_SIZE)
    X_train = x[:n_training_samples, :]
    y_train = y[:n_training_samples]
    X_test = x[n_training_samples:, :]
    y_test = y[n_training_samples:]

    return X_train, y_train, X_test, y_test, n_training_samples

def plot_results(train, output):
    plt.figure(figsize=(16,6))
    plt.title('Model')
    plt.xlabel('Date', fontsize=18)
    plt.ylabel('Close Price USD ($)', fontsize=18)
    plt.plot(train['Close'])
    plt.plot(output[['Close', 'Predictions']])
    plt.legend(['Train', 'Val', 'Predictions'], loc='lower right')
    plt.show()
```

Рисунок 3.2.5 Обробка даних та виконання операції прогнозування.

На даному етапі на Рисунках 3.2.6 – 3.2.7, виконується операція додавання першого шару LSTM з функцією активації сигмоїдної форми та регулювання даних під тестові та тренувальні дані із додаванням після вихідного шару.

```
LSTM_model = Sequential()
LSTM_model.add(LSTM(units = 50, return_sequences = True, input_shape = (X_train.shape[1], 1)))
LSTM_model.add(Dropout(0.2))

LSTM_model.add(LSTM(units = 50, return_sequences = True))
LSTM_model.add(Dropout(0.2))

LSTM_model.add(LSTM(units = 50, return_sequences = True))
LSTM_model.add(Dropout(0.2))

LSTM_model.add(LSTM(units = 50))
LSTM_model.add(Dropout(0.2))

LSTM_model.add(Dense(units = 1))
LSTM_model.summary()
```

Рисунок 3.2.6 Будова LSTM моделі.

```
def fit_model(days_in_future, window_size):
    print(f"Predicting for {days_in_future} days in future...")

    X_train, y_train, X_test, y_test, n_training_samples = convert_data(days_in_future, window_
size)
    y_test_original = scaler.inverse_transform(y_test.reshape(-1, 1))
    print("Train samples:", len(y_train), "Test samples:", len(y_test))

    model = get_model(window_size)
    history = model.fit(
        x=X_train,
        y=y_train,
        epochs=15,
        batch_size=64,
        validation_data=(X_test, y_test),
        callbacks=[EarlyStopping(monitor='val_loss', patience=3)]
    )
    predictions = model.predict(X_test)
    predictions = scaler.inverse_transform(predictions)

    rmse = np.sqrt(np.mean(((predictions - y_test_original) ** 2)))
    print("RMSE:", rmse)
    mape = mean_absolute_percentage_error(y_test_original, predictions)
    print("Mean Absolute Percentage Error:", mape)

    error = (np.abs(predictions[-1] - y_test_original[-1]) / y_test_original[-1])[0]
    print("Final Prediction Error:", error)

    train = apple[:n_training_samples+window_size+days_in_future-1]
    output = apple[n_training_samples+window_size+days_in_future-1:-1]
    output['Predictions'] = predictions
    plot_results(train, output)
```

```
LSTM_predict = sc.inverse_transform(LSTM_model.predict(X_test))

def metrics(*args):
    for model in args:
        for key, value in model.items():

            diff = value - test
            print(f'MSE of {key} is {round(np.mean(diff**2), 2)}')
            print(f'MAE of {key} is {round(np.mean(abs(diff)), 2)}')
            print(f'RMSE of {key} is {round(np.sqrt(np.mean(diff**2)), 2)}')
```

Рисунок 3.2.7 Прогнозування за LSTM моделю.

На даному етапі відбувається початкова реалізація програмного коду для візуалізації статистичних даних із дата-сетів із подальшою реалізацією в наступному пункті розділу 3.3, та за Рисунками 3.2.7 – 3.2.11.

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 253 entries, 2021-05-17 to 2022-05-16
Data columns (total 7 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Open          253 non-null    float64
 1   High          253 non-null    float64
 2   Low           253 non-null    float64
 3   Close         253 non-null    float64
 4   Adj Close     253 non-null    float64
 5   Volume        253 non-null    int64
 6   company_name  253 non-null    object
dtypes: float64(5), int64(1), object(1)
memory usage: 15.8+ KB
```

Рисунок 3.2.7 Основна інформація датасетів із колонок та типу даних.

```
plt.figure(figsize=(15, 6))
plt.subplots_adjust(top=1.25, bottom=1.2)

for i, company in enumerate(company_list, 1):
    plt.subplot(2, 2, i)
    company['Adj Close'].plot()
    plt.ylabel('Adj Close')
    plt.xlabel(None)
    plt.title(f"Closing Price of {tech_list[i - 1]}")

plt.tight_layout()
```

Рисунок 3.2.8 Операція із даними для утворення графіку на закриття угод придбання акцій.

```
plt.figure(figsize=(15, 7))
plt.subplots_adjust(top=1.25, bottom=1.2)

for i, company in enumerate(company_list, 1):
    plt.subplot(2, 2, i)
    company['Volume'].plot()
    plt.ylabel('Volume')
    plt.xlabel(None)
    plt.title(f"Sales Volume for {tech_list[i - 1]}")

plt.tight_layout()
```

Рисунок 3.2.8 Операція із даними для утворення графіку загального продажу акцій на торгових фондах.

```
ma_day = [10, 20, 50]

for ma in ma_day:
    for company in company_list:
        column_name = f"MA for {ma} days"
        company[column_name] = company['Adj Close'].rolling(ma).mean()

fig, axes = plt.subplots(nrows=2, ncols=2)
fig.set_figheight(8)
fig.set_figwidth(15)

AAPL[['Adj Close', 'MA for 10 days', 'MA for 20 days', 'MA for 50 days']].plot(ax=axes[0,0])
axes[0,0].set_title('APPLE')

GOOG[['Adj Close', 'MA for 10 days', 'MA for 20 days', 'MA for 50 days']].plot(ax=axes[0,1])
axes[0,1].set_title('GOOGLE')

MSFT[['Adj Close', 'MA for 10 days', 'MA for 20 days', 'MA for 50 days']].plot(ax=axes[1,0])
axes[1,0].set_title('MICROSOFT')

AMZN[['Adj Close', 'MA for 10 days', 'MA for 20 days', 'MA for 50 days']].plot(ax=axes[1,1])
axes[1,1].set_title('AMAZON')

fig.tight_layout()
```

Рисунок 3.2.9 Операція із даними для утворення графіку ковзних середніх цін акцій.

```
for company in company_list:
    company['Daily Return'] = company['Adj Close'].pct_change()

fig, axes = plt.subplots(nrows=2, ncols=2)
fig.set_figheight(8)
fig.set_figwidth(15)

AAPL['Daily Return'].plot(ax=axes[0,0], legend=True, linestyle='--', marker='o')
axes[0,0].set_title('APPLE')

GOOG['Daily Return'].plot(ax=axes[0,1], legend=True, linestyle='--', marker='o')
axes[0,1].set_title('GOOGLE')

MSFT['Daily Return'].plot(ax=axes[1,0], legend=True, linestyle='--', marker='o')
axes[1,0].set_title('MICROSOFT')

AMZN['Daily Return'].plot(ax=axes[1,1], legend=True, linestyle='--', marker='o')
axes[1,1].set_title('AMAZON')

fig.tight_layout()
```

Рисунок 3.2.10 Операція із даними для утворення графіку із середнього щоденного прибутку з них.

```
plt.figure(figsize=(12, 7))

for i, company in enumerate(company_list, 1):
    plt.subplot(2, 2, i)
    company['Daily Return'].hist(bins=50)
    plt.ylabel('Daily Return')
    plt.title(f'{company_name[i - 1]}')

plt.tight_layout()
```

Рисунок 3.2.11 Операція із даними для утворення гістаграми ковзних середніх цін акцій.

3.3 Візуалізація результату аналізу.

На Рисунку 3.3.1 – за графіком річних акцій компаній, ми можемо зрозуміти, порівнюючи їх самих компаній, що кожна компанія починаючи з 2022 року, зазнала обвалу в ціні акцій. Але ми можемо спостерігати, що Apple на початку першої половини 2021 мав гарний підйом на декілька десятків доларів, Google, Amazon – мали обвал акцій стабільно із середини 2021 року, проте як Microsoft був стабільніший у середині 2021 року, але ми можемо наглядно продивитись що в останні місяці 2021 року, вартість акцій впала і скоріше за все, через випуск нової операційної системи Windows 11. З приводу падіння акцій Apple, можна лише гадати через що, або ж через презентації новинок від даної компанії або інших факторів, потрібно не забувати що компанія Apple в цьому тренді є однією із найбагатших компаній. Також умовами цих падінь може бути ще й судові справи або інші фактори впливу.

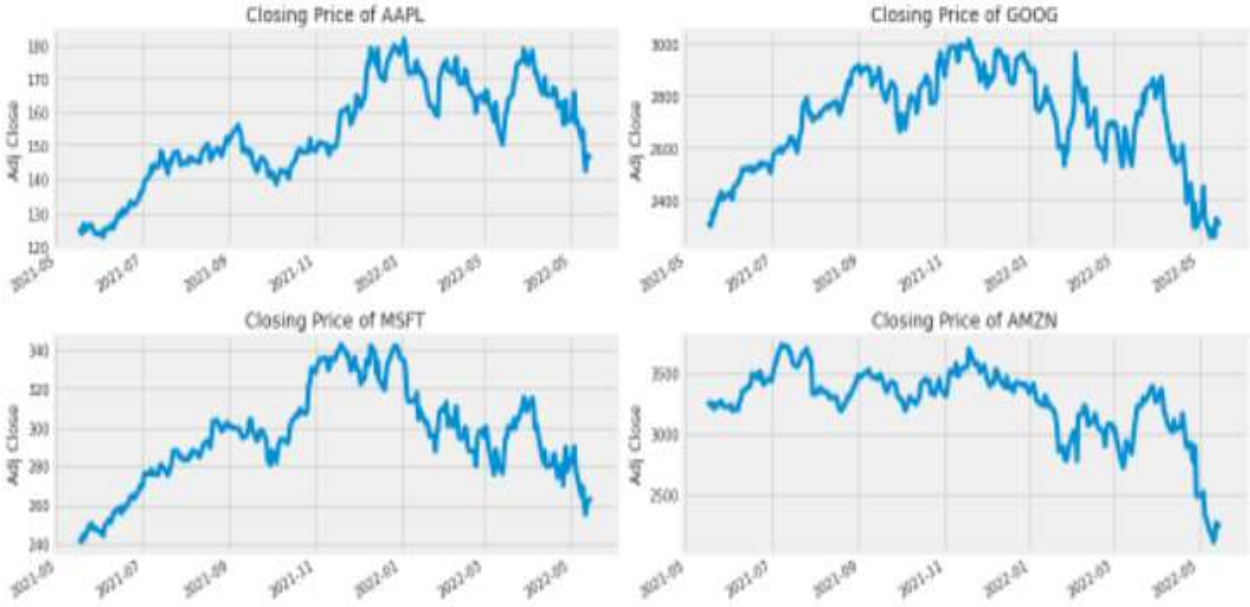


Рисунок 3.3.1 Загальний річний графік акцій компаній із закриттям угод на остаточну вартість купівлі.

На Рисунку 3.3.2 слід зазначити коливання графічної метрики, в свою чергу на такий попит може означати декілька факторів. Apple – на початку 2021 року був трохи стабільним із продажу акцій не беручи до уваги другу половину року літо/осінь – купівля акцій пішла на спад, тим самим на Рисунку

3.3.1 можна провести паралель, що ціна акції падала, і її не так багато і купляли. Прогрес даного графіку ми можемо спостерігати мабуть вже в грудні та на початку 2022 року, де ціна впала на декілька пунктів але має стрімкий скачок із послідовним різьким обвалом попиту наприкінці. Google на перший погляд має стабільний продаж акцій в якісь конкретні періоди, бо відбуваються на графіки не визначенні скачки продажу. Microsoft – в середині 2021 року був стабільним, на початку 2022 року можна спостерігати за одноразовим стрибком і все. Amazon – проведення паралелі із Google також має декілька, але значно менше скачків купівлі акцій.

Рисунок 3.3.2 Загальний річний графік компаній із продажу акцій на торгових фондах.

Рисунок 3.3.3 Загальний річний графік акцій компаній із ковзних середніх цін акцій.

На Рисунку 3.3.4 можна спостерігати загальну картину компаній та обвал вартості акцій. Більш за все просаджується компанії Google та Microsoft.

Тепер, коли ми провели базовий аналіз, давайте зануримося трохи глибше. Зараз ми збираємося проаналізувати ризики акції. Для цього нам потрібно буде уважніше розглянути щоденні зміни акції, а не лише її абсолютну вартість. До даної операції доходів взяли за основу акції Apple.

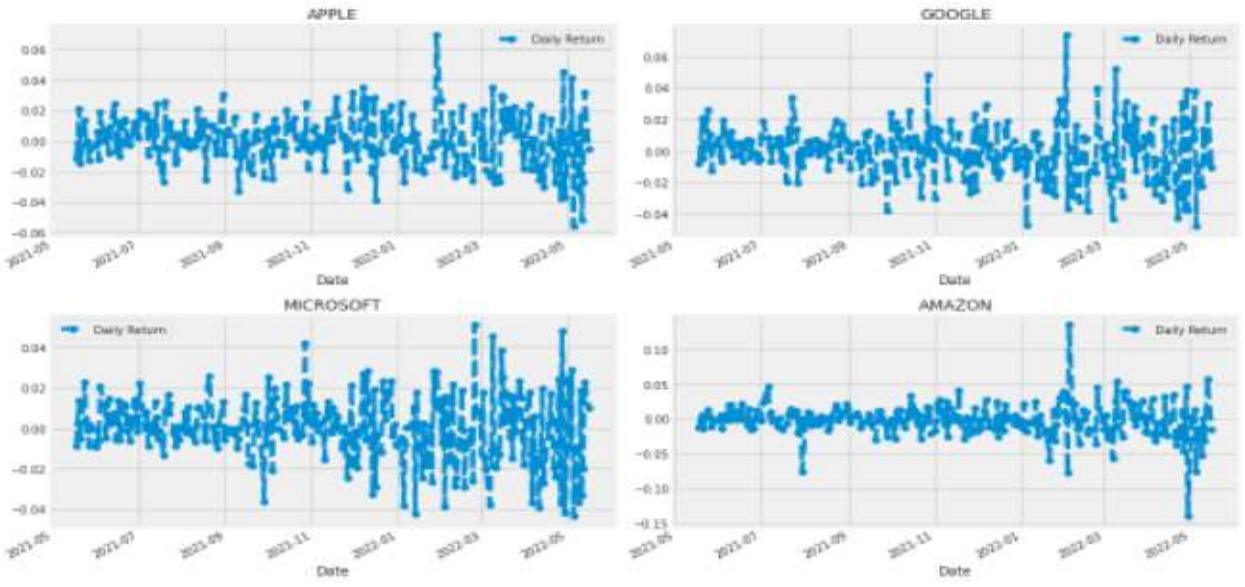


Рисунок 3.3.4 Загальний річний графік акцій компаній із середнього щоденного прибутку з них.

Тепер поглянемо на середню щоденну прибутковість за допомогою гістограми. Будемо використовувати пакет seaborn, щоб створити як гістограму. В лідерах з компаній прибуток більше отримує Amazon.

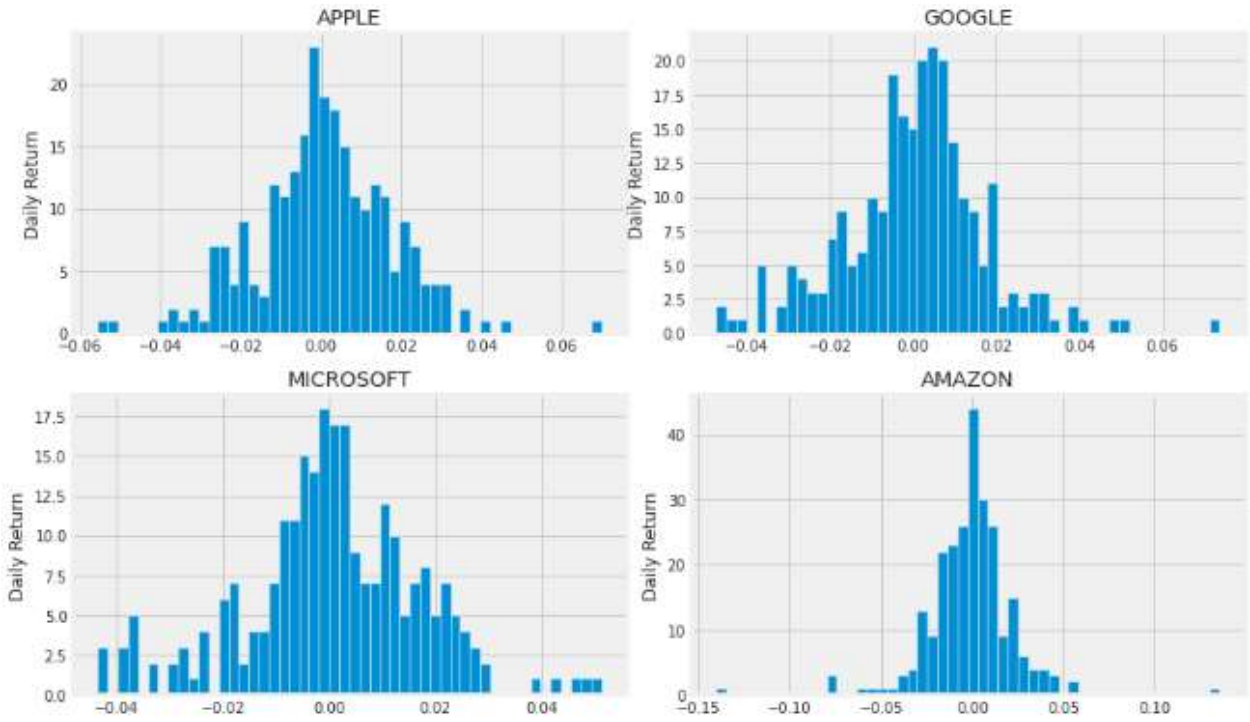


Рисунок 3.3.5 Гістограма прибутковості.

Нижче ми бачимо всі взаємозв'язки щоденних прибутків між усіма акціями. Гістограма показує цікаву кореляцію між щоденними доходами Google і Amazon. Було б цікаво дослідити це порівняння окремих осіб. Хоча простота звичайного виклику через функцію sns.pairplot(), ми також можемо використати sns.PairGrid() для повного контролю над фігурою, включно з тим, які графіки розташовані по діагоналі, верхньому трикутнику та нижньому трикутнику. Нижче наведено приклад використання всієї потужності seaborn для досягнення цього результату.

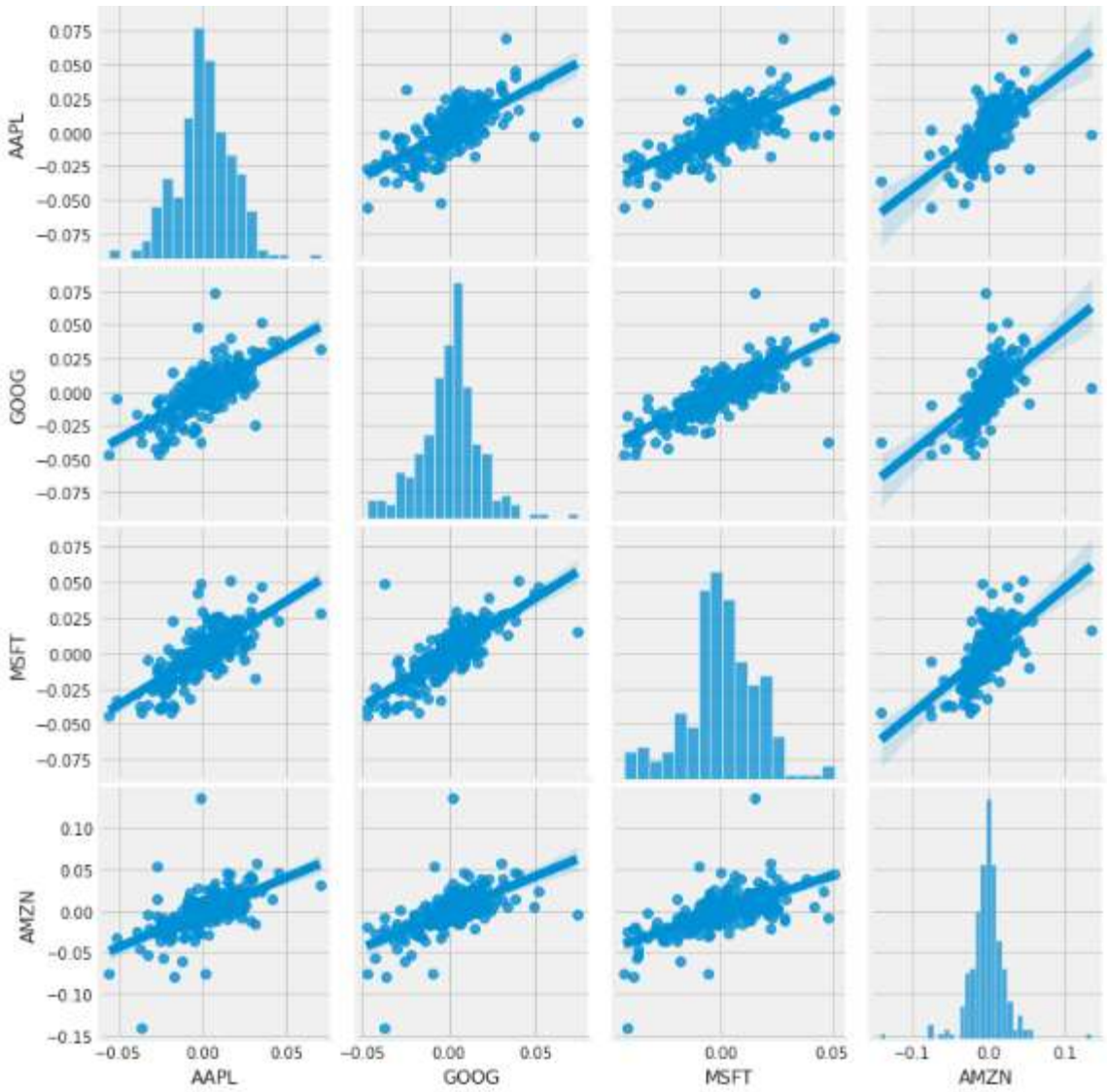


Рисунок 3.3.6 Гістограма прибутковості через пакет seaborn.

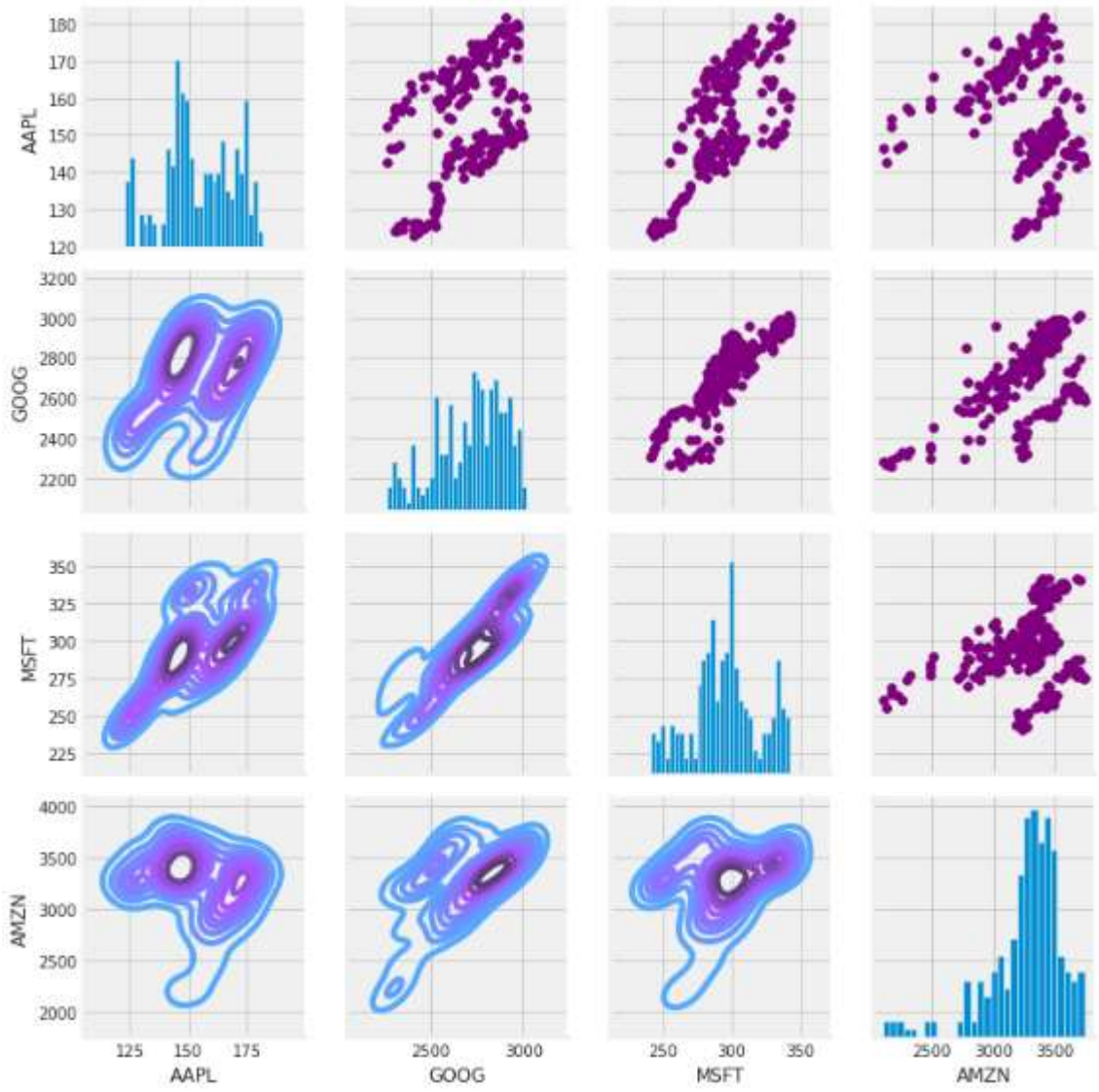


Рисунок 3.3.7 Гістограма прибутковості через пакет seaborn.

Нарешті, ми можемо створити графік кореляції, щоб отримати фактичні числові значення між щоденними значеннями прибутку акцій. Порівнюючи ціни закриття, ми бачимо цікаві відносини між Microsoft і Apple на наступному графіку.

Рисунок 3.3.8 Графік кореляції денного прибутку.

Візуально можна побачити, що Microsoft і Amazon мали найсильнішу кореляцію щоденної прибутковості акцій. Також цікаво побачити, що всі технологічні компанії позитивно корелюють.

Оцінка ризику при інвестицій в акції даних компаній та коефіцієнта прибутку. На данному Рисунку 3.3.9 ми спостерігаємо, що Amazon знаходиться у топі по риску та знаходиться в зоні негативного коефіцієнті за очікувальним прибутком, через те що останні декілька місяців акції не були стабільні.

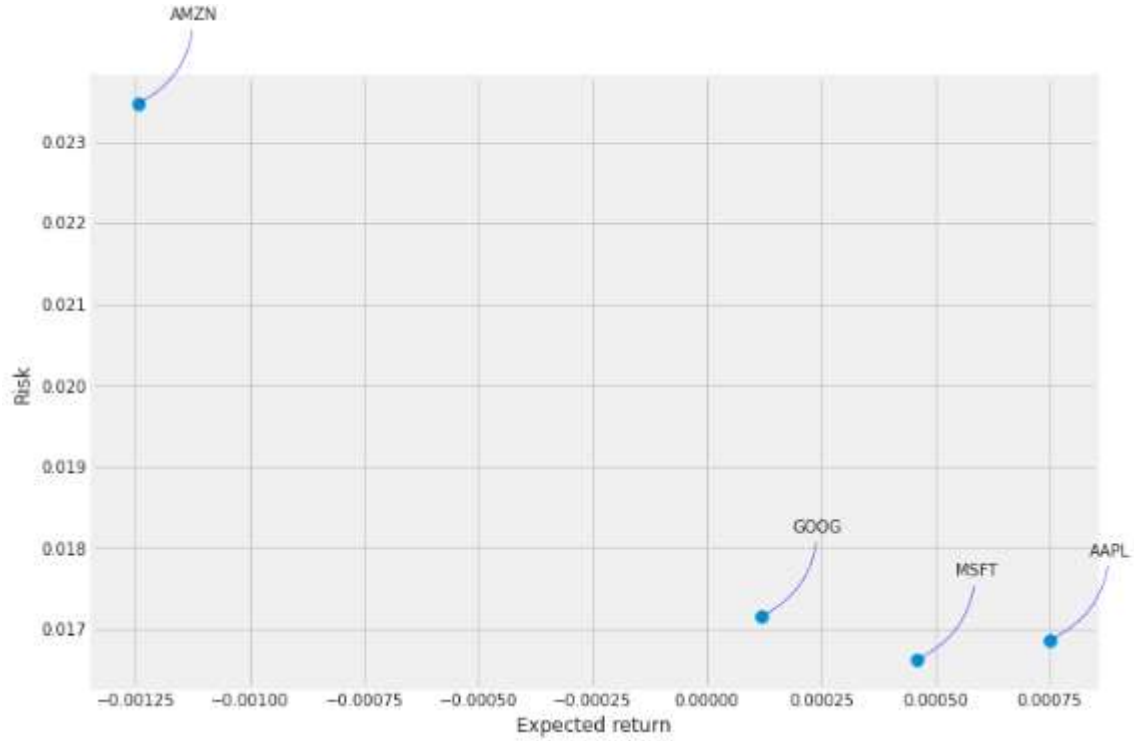


Рисунок 3.3.9 Оцінка ризику та коефіцієнта прибутку.

Прогнозування закриття курсу акцій компанії Amazon.

Для даної операції нам потрібен датасет із більшою кількістю даних, для більш приближеного результату.

Date	High	Low	Open	Close	Volume	Adj Close
2012-01-03	179.479996	175.550003	175.889999	179.029999	5110800	179.029999
2012-01-04	180.500000	176.070007	179.210007	177.509995	4205200	177.509995
2012-01-05	178.250000	174.050003	175.940002	177.610001	3809100	177.610001
2012-01-06	184.649994	177.500000	178.070007	182.610001	7008400	182.610001
2012-01-09	184.369995	177.000000	182.759995	178.559998	5056900	178.559998
...	...	...	...	...	...	...
2022-05-10	2252.850098	2143.419922	2225.000000	2177.179932	5271700	2177.179932
2022-05-11	2203.120117	2088.570068	2162.070068	2107.439941	5485200	2107.439941
2022-05-12	2215.610107	2048.110107	2055.000000	2138.610107	6601300	2138.610107
2022-05-13	2263.679932	2156.000000	2181.379883	2261.100098	4676700	2261.100098
2022-05-16	2279.000000	2207.082764	2262.000000	2228.350098	2613553	2228.350098

2610 rows × 6 columns

Рисунок 3.3.10 Датасет акцій компанії Amazon, за 10 років.

Спостерігаючи за графіком на Рисунку 3.3.11 можна побачити пік вартості акції Amazon приходить на 2021 рік.

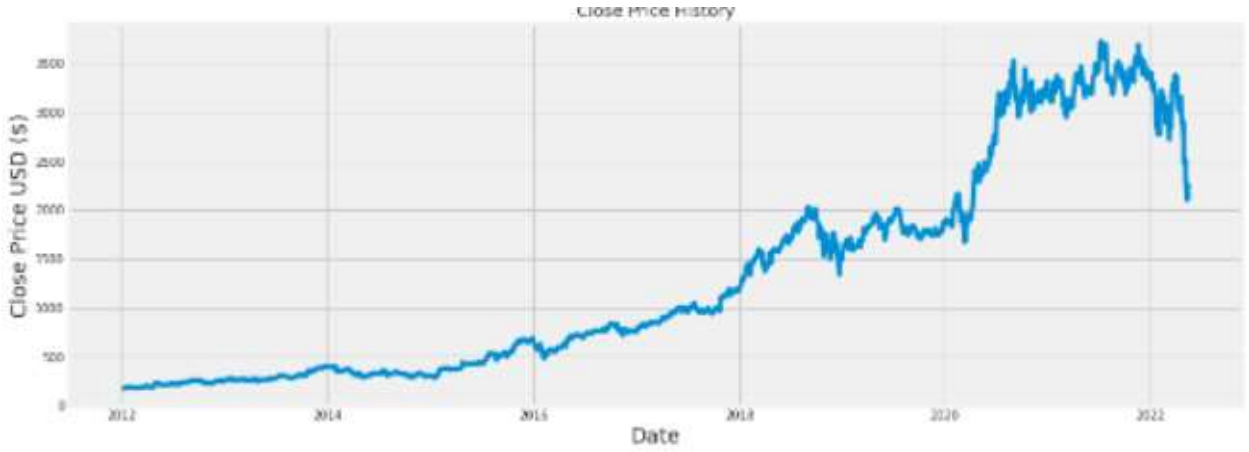


Рисунок 3.3.11 Загальний десяти-річний графік акцій компанії Amazon із закриттям угод на остаточну вартість купівлі.

На Рисунку 3.3.11 можна спостергіати що за даними котрі були отримані на протязі 10 років за графіком прогнозування ціни акції компанії Amazon, будуть падати і далі.

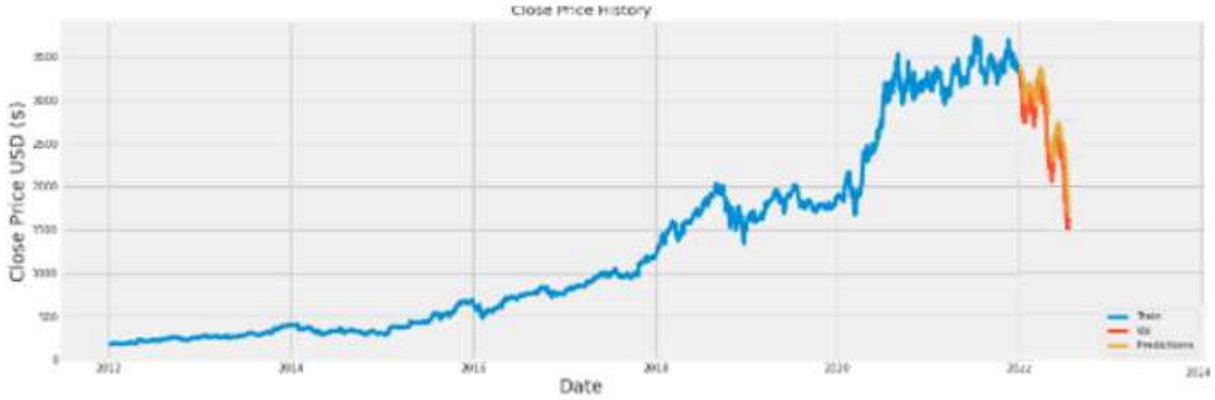


Рисунок 3.3.11 Загальний графік прогнозування акцій компанії Amazon із закриттям угод на остаточну вартість купівлі.

3.4 Висновки після третього розділу.

В даному розділі ми досягли усіх поставлених задач, а саме, було розроблено алгоритм для виконання прогнозування акцій компанії Amazon на основі отриманих даних із інформаційних джерел.

Було розроблено візуалізації на основі яких можна робити детальний інформаційний аналіз. Усі компанії котрі були долучені до аналізу, в підсумку реалізовувались візуально-графічно разом, для більш детального спостереження графіків, їх збігів, вартостей, ефективностей.

Чому саме Amazon, на мою думку для інвестора буде корисним знати, більше інформації з приводу компанії котра має найбільшу вартість на ринку за одну акцію компанії. Даний бренд розвивається в стримінгових, геймерських та кіно галузях, що насамперед є гарною інвестицією реалізації розвитку компанії в інших галузях, та бути конкуренто спроможним відстоювати свої інтереси в напрямку розвитку, тощо. Адже наприклад за участю Amazon Prime, було вже придбано більше 50 сценарних проектів і на очікуванні вже існують деякі відзняті, при можливості граних рейтингів це може вплинути на ріст акцій компанії та продажу товарів з основної галузі інтернет-комерції та отримання додаткового прибутку від переглядів та придбання підписки на Amazon Prime. Але попри все, не потрібно забувати, що інвестору ні в якому разі не потрібно інвестувати тільки в одну компанію, бо це може бути не так прибутково як хочется йому. Беручи до уваги, що зараз світ не стоїть на місці і більше в тренди входять технології, я б рекомендував робити додаткові інвестиції в військову, космічну та інших галузях, наприклад: “Nasa”, “Tesla”, “AeroVironment”. Остання компанія являє собою військове підприємство із масштабним випуском дронів-камікадзе «SwitchBlade».

ВИСНОВОК

Підсумовуючі нашу роботу, відкрито можна сказати, що усі поставленні за мету цілі були досягнуті та було досліджено питання з приводу прогнозування акції компаній на короткостроковий період, було проаналізовано компанії-гіганти такі як Amazon, Apple, Microsoft, Google.

Було визначено вхідні та вихідні дані, розроблено та використано інструментарії із середовища мови програмування Python, для обробки даних, збору даних, створення LSTM моделі, побудова показових, аналітичних та кореляційних графіків.

Але у кінці усього впадає до думки читачеві, для кого ця інформація буде корисною, хто нею буде або може користуватись, чо можна покладатись на дане прогнозування та бути впевненим що воно справдиться, чи воно буде постійно влучати у 100% вірогідність даного прогнозу у майбутньому та справджуватись, а не спростовуватись.

По-перше, ще людство не винайшло технології із пророкуванням тих чи інших вартостей на будь що, тому треба розраховувати, що даний прогноз містить як одну із імовірних випадків та варіативностей, котрих може бути безліч, так як теорію вірогідності, ніхто не відміняв.

По-друге, даний проект можна так назвати, не несе відповідальності за втрати та кошти інвесторів, бо за мету не було поставлено за задачу – як, заробітку на цьому із 100% поверненням грошей чи інших звичайних махінацій, як один із моментів дослідження та варіанту розвитку акції дана робота себе виправдовує та позиціонує себе в розрізі дослідження.

По-третє, із початковою задумкою теми магістерської роботи, являло себе за метою, дати можливість бачення можливої вірогідності даного прогнозу на одну із компаній-гігантів людям, інвесторам, брокерам, трейдерам, тощо.

Вартість компанії Microsoft – 2,565.90 триліон .

Вартість компанії Apple – 3,003.40 триліон .

Вартість компанії Amazon – 1,700.40 триліон .

Вартість компанії Google – 2,120.28 триліон .

Чому було обрано компанію-гігант Amazon із переліку порівнюючи інші компанії? По-перше, беручи до уваги той факт, що Amazon являється одним із найбільших компаній по вартості за одну акцію, то за мету було поставлено й дослідити, проаналізувати та виконати прогнозування для даної компанії із цілю можливого капіталовкладення. По-друге, дивлячись на графіки на протязі року за нашими даними, Amazon показував себе більш стабільно з травня 2021 року по грудень, а після нового 2022 року акції компанії Amazon почали дешевшати, факторів впливу може бути безліч, наприклад через публікації статистики компанії з приводу їх дохідності, в цьому документі загальний показник відобразив основне – зниження прибутку компанії, тим самим дана інформація поширилася в інтернеті як вірусне кумедне відео котре набуває масштабних трендів та відверто кажучи відбулася ланцюжкова психологічна паніка людей, котра була зумовлена тим, що компанія втрачає прибуток, та може занепасти і збанкрутувати, такою думки може бути більшість людей через нестабільний емоційно-психологічний фон. Такий фактор впливу можна назвати людським і його дуже багато де можна знайти.

Тож відповідаючи на питання, кому ця інформація може бути корисною, відповідь для неї – усім охочим та бажаючим дізнатись світ інвестицій та поринути у шторм великих обсягів даних.

ПЕРЕЛІК ВИКОРИСТАНИХ ІНФОРМАЦІЙНИХ ДЖЕРЕЛ

1. Фондовий індекс [Електронний ресурс] – Режим доступу до ресурсу: https://ru.wikipedia.org/wiki/%D0%A4%D0%BE%D0%BD%D0%B4%D0%BE%D0%B2%D1%8B%D0%B9_%D0%B8%D0%BD%D0%B4%D0%B5%D0%BA%D1%81
2. Financial Encyclopedia [Електронний ресурс] – Режим доступу до ресурсу: https://www.invest-rating.ru/financial-encyclopedia/chto-vlijaet-na-cenu-akcii/
3. Investment factors [Електронний ресурс] – Режим доступу до ресурсу: https://www.getsmarteraboutmoney.ca/invest/investment-products/stocks/factors-that-can-affect-stock-prices/
4. The History of Start of the Stock Market [Електронний ресурс] – Режим доступу до ресурсу: A Brief History of the Stock Market | SoFi
5. CAC 40 [Електронний ресурс] – Режим доступу до ресурсу: https://ru.wikipedia.org/wiki/CAC_40
6. Nikkei_225 [Електронний ресурс] – Режим доступу до ресурсу: https://ru.wikipedia.org/wiki/Nikkei_225
7. S&P 500 [Електронний ресурс] – Режим доступу до ресурсу: https://ru.wikipedia.org/wiki/S%26P_500
8. DJIA [Електронний ресурс] – Режим доступу до ресурсу: https://uk.wikipedia.org/wiki/%D0%9F%D1%80%D0%BE%D0%BC%D0%B8%D1%81%D0%BB%D0%BE%D0%B2%D0%B8%D0%B9_%D1%96%D0%BD%D0%B4%D0%B5%D0%BA%D1%81_%D0%94%D0%BE%D1%83-%D0%94%D0%B6%D0%BE%D0%BD%D1%81%D0%B0
9. Акція [Електронний ресурс] – Режим доступу до ресурсу: https://uk.wikipedia.org/wiki/%D0%90%D0%BA%D1%86%D1%96%D1%8F_(%D1%86%D1%96%D0%BD%D0%BD%D0%B8%D0%B9_%D0%BF%D0%B0%D0%BF%D1%96%D1%80)

10. Облігація [Електронний ресурс] – Режим доступу до ресурсу:https://uk.wikipedia.org/wiki/%D0%9E%D0%B1%D0%BB%D1%96%D0%B3%D0%B0%D1%86%D1%96%D1%8F

11. Ціна акції і фактори впливу на її ціну [Електронний ресурс] – Режим доступу до ресурсу: http://bizkiev.com/content/view/392/205/

12. Від чого залежить акції компанії? [Електронний ресурс] – Режим доступу до ресурсу: https://ffin.ua/blog/articles/investopediia/post/vid-choho-zalezhyt-vartist-aktsii-kompanii

13. Фундаментальний аналіз [Електронний ресурс] – Режим доступу до ресурсу: https://www.investopedia.com/terms/f/fundamentalanalysis.asp

14. Рекурентні нейронні мережі[Електронний ресурс] – Режим доступу до ресурсу: https://neerc.ifmo.ru/wiki/index.php?title=%D0%A0%D0%B5%D0%BA%D1%83%D1%80%D1%80%D0%B5%D0%BD%D1%82%D0%BD%D1%8B%D0%B5_%D0%BD%D0%B5%D0%B9%D1%80%D0%BE%D0%BD%D0%BD%D1%8B%D0%B5_%D1%81%D0%B5%D1%82%D0%B8

15. LSTM [Електронний ресурс] – Режим доступу до ресурсу: https://ru.wikipedia.org/wiki/%D0%94%D0%BE%D0%BB%D0%B3%D0%B0%D1%8F_%D0%BA%D1%80%D0%B0%D1%82%D0%BA%D0%BE%D1%81%D1%80%D0%BE%D1%87%D0%BD%D0%B0%D1%8F_%D0%BF%D0%B0%D0%BC%D1%8F%D1%82%D1%8C

16. Google [Електронний ресурс] – Режим доступу до ресурсу: https://ru.wikipedia.org/wiki/Google_(%D0%BA%D0%BE%D0%BC%D0%BF%D0%B0%D0%BD%D0%B8%D1%8F)

17. Amazon [Електронний ресурс] – Режим доступу до ресурсу: https://ru.wikipedia.org/wiki/Amazon

18. Microsoft [Електронний ресурс] – Режим доступу до ресурсу: https://ru.wikipedia.org/wiki/Microsoft

19. Apple [Електронний ресурс] – Режим доступу до ресурсу: https://ru.wikipedia.org/wiki/Apple

20. Stock market sector [Електронний ресурс] – Режим доступу до ресурсу: https://journal.tinkoff.ru/sectors/

21. The Birth of Stock Exchanges [Електронний ресурс] – Режим доступу до ресурсу: The Birth of Stock Exchanges (investopedia.com)

22. Ковзні середні [Електронний ресурс] – Режим доступу до ресурсу: https://red-circule.com/articles/skolsyashie-srednie

23. Чому акції Amazon сьогодні знову падають? [Електронний ресурс] – Режим доступу до ресурсу: https://www.fool.com/investing/2022/05/02/why-amazon-stock-is-falling-again-today/

24. Чому акції Amazon знову падають? [Електронний ресурс] – Режим доступу до ресурсу: https://www.fool.com/investing/2022/05/03/why-amazon-stock-plummeted-24-last-month/

25. Чому акції Google знову падають? [Електронний ресурс] – Режим доступу до ресурсу: https://www.fool.com/investing/2022/04/28/alphabet-stock-fell-after-q1-report-should-you-buy/#:~:text=Alphabet%20(GOOG%20%2D1.29%25),per%20share%20and%20revenue%20estimates.

26. The Top 25 Stocks in the S&P 500 [Електронний ресурс] – Режим доступу до ресурсу: https://www.investopedia.com/ask/answers/08/find-stocks-in-sp500.asp

27. Correlation analysis [Електронний ресурс] – Режим доступу до ресурсу: https://www.questionpro.com/features/correlation-analysis.html

28. What Do Correlation Coefficients Positive, Negative, and Zero Mean? [Електронний ресурс] – Режим доступу до ресурсу: https://www.investopedia.com/ask/answers/032515/what-does-it-mean-if-

correlation-coefficient-positive-negative-or-zero.asp

29. Statistical Analysis in Python using Pandas [Електронний ресурс] – Режим доступу до ресурсу: https://medium.com/analytics-vidhya/statistical-analysis-in-python-using-pandas-27c6a4209de2

30. Numpy [Електронний ресурс] – Режим доступу до ресурсу: https://numpy.org/doc/stable/reference/routines.statistics.html

ДОДАТОК

Lstm model prediction

```
# Create a new dataframe with only the 'Close column
data = df.filter(['Close'])
# Convert the dataframe to a numpy array
dataset = data.values
# Get the number of rows to train the model on
training_data_len = int(np.ceil( len(dataset) * .95 ))

training_data_len
2480
# Scale the data
from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler(feature_range=(0,1))
scaled_data = scaler.fit_transform(dataset)

scaled_data
array([[8.71895264e-04],
       [4.44384968e-04],
       [4.72512283e-04],
       ...,
       [5.52015525e-01],
       [5.86466568e-01],
       [5.77255434e-01]])
# Create the training data set
# Create the scaled training data set
train_data = scaled_data[0:int(training_data_len), :]
# Split the data into x_train and y_train data sets
x_train = []
y_train = []

for i in range(60, len(train_data)):
    x_train.append(train_data[i-60:i, 0])
    y_train.append(train_data[i, 0])
    if i<= 61:
        print(x_train)
        print(y_train)
        print()

# Convert the x_train and y_train to numpy arrays
x_train, y_train = np.array(x_train), np.array(y_train)
```

 # Reshape the data
x_train = np.reshape(x_train, (x_train.shape[0], x_train.shape[1], 1))
 # x_train.shape

[array([0.0008719 , 0.00044438, 0.00047251, 0.00187879, 0.0007397 ,
0.00095908, 0.00083533, 0. , 0.00070033, 0.0016116 ,
0.00379977, 0.00520886, 0.00421884, 0.00285756, 0.00311351,
0.00333851, 0.00489105, 0.00546762, 0.00456197, 0.00520605,
0.00099284, 0.00162847, 0.00330476, 0.00202786, 0.00232318,
0.002686 , 0.00254537, 0.00270287, 0.00440447, 0.00432291,
0.00240193, 0.00112502, 0.00184785, 0.00178035, 0.00130784,
0.00083252, 0.00090002, 0.00073127, 0.00221349, 0.00105753,
0.00115596, 0.00094784, 0.00121784, 0.00145128, 0.00220505,
0.00329351, 0.00235974, 0.00209817, 0.00243568, 0.00178035,
0.00239068, 0.00256506, 0.00269725, 0.0046126 , 0.00444385,
0.00463229, 0.0053748 , 0.00757704, 0.00829987, 0.00709609])]
[0.008066423825883726]

[array([0.0008719 , 0.00044438, 0.00047251, 0.00187879, 0.0007397 ,
0.00095908, 0.00083533, 0. , 0.00070033, 0.0016116 ,
0.00379977, 0.00520886, 0.00421884, 0.00285756, 0.00311351,
0.00333851, 0.00489105, 0.00546762, 0.00456197, 0.00520605,
0.00099284, 0.00162847, 0.00330476, 0.00202786, 0.00232318,
0.002686 , 0.00254537, 0.00270287, 0.00440447, 0.00432291,
0.00240193, 0.00112502, 0.00184785, 0.00178035, 0.00130784,
0.00083252, 0.00090002, 0.00073127, 0.00221349, 0.00105753,
0.00115596, 0.00094784, 0.00121784, 0.00145128, 0.00220505,
0.00329351, 0.00235974, 0.00209817, 0.00243568, 0.00178035,
0.00239068, 0.00256506, 0.00269725, 0.0046126 , 0.00444385,
0.00463229, 0.0053748 , 0.00757704, 0.00829987, 0.00709609]), array([0.000
44438, 0.00047251, 0.00187879, 0.0007397 , 0.00095908,
0.00083533, 0. , 0.00070033, 0.0016116 , 0.00379977,
0.00520886, 0.00421884, 0.00285756, 0.00311351, 0.00333851,
0.00489105, 0.00546762, 0.00456197, 0.00520605, 0.00099284,
0.00162847, 0.00330476, 0.00202786, 0.00232318, 0.002686 ,
0.00254537, 0.00270287, 0.00440447, 0.00432291, 0.00240193,
0.00112502, 0.00184785, 0.00178035, 0.00130784, 0.00083252,
0.00090002, 0.00073127, 0.00221349, 0.00105753, 0.00115596,
0.00094784, 0.00121784, 0.00145128, 0.00220505, 0.00329351,
0.00235974, 0.00209817, 0.00243568, 0.00178035, 0.00239068,
0.00256506, 0.00269725, 0.0046126 , 0.00444385, 0.00463229,
0.0053748 , 0.00757704, 0.00829987, 0.00709609, 0.00806642])]
[0.008066423825883726, 0.007475784544801765]

```
from keras.models import Sequential
from keras.layers import Dense, LSTM

# Build the LSTM model
model = Sequential()
model.add(LSTM(128, return_sequences=True, input_shape= (x_train.shape[1], 1)
))
model.add(LSTM(64, return_sequences=False))
model.add(Dense(25))
model.add(Dense(1))

# Compile the model
model.compile(optimizer='adam', loss='mean_squared_error')

# Train the model
model.fit(x_train, y_train, batch_size=1, epochs=1)
2420/2420 [==============================] - 101s 41ms/step - loss: 0.0
019
<keras.callbacks.History at 0x7f4644d1b950>
# Create the testing data set
# Create a new array containing scaled values from index 1543 to 2002
test_data = scaled_data[training_data_len - 60: , :]
# Create the data sets x_test and y_test
x_test = []
y_test = dataset[training_data_len:, :]
for i in range(60, len(test_data)):
    x_test.append(test_data[i-60:i, 0])

# Convert the data to a numpy array
x_test = np.array(x_test)

# Reshape the data
x_test = np.reshape(x_test, (x_test.shape[0], x_test.shape[1], 1 ))

# Get the models predicted price values
predictions = model.predict(x_test)
predictions = scaler.inverse_transform(predictions)

# Get the root mean squared error (RMSE)
rmse = np.sqrt(np.mean(((predictions - y_test) ** 2)))
rmse
227.09821776549228
# Plot the data
train = data[:training_data_len]
valid = data[training_data_len:]
```

```
valid['Predictions'] = predictions
# Visualize the data
plt.figure(figsize=(16,6))
plt.title('Model')
plt.xlabel('Date', fontsize=18)
plt.ylabel('Close Price USD ($)', fontsize=18)
plt.plot(train['Close'])
plt.plot(valid[['Close', 'Predictions']])
plt.legend(['Train', 'Val', 'Predictions'], loc='lower right')
plt.show()
```

```
/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:4: SettingWithCopy
Warning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable
/user_guide/indexing.html#returning-a-view-versus-a-copy
  after removing the cwd from sys.path.
```

Train model

```
# Ratio of sample to use for training vs. testing
TRAIN_SIZE = 0.9
def get_model(window_size):
    model = Sequential()
    model.add(LSTM(128, return_sequences=True, input_shape=(window_size, 1))
)
    model.add(LSTM(64, return_sequences=False))
    model.add(Dense(25))
    model.add(Dense(1))
    model.compile(optimizer='adam', loss='mean_squared_error')
    return model

model = get_model(60)
model.summary()
```

```
2022-05-03 10:54:06.048230: I tensorflow/stream_executor/cuda/cuda_gpu_execu
tor.cc:937] successful NUMA node read from SysFS had negative value (-1), but t
here must be at least one NUMA node, so returning NUMA node zero
2022-05-03 10:54:06.173330: I tensorflow/stream_executor/cuda/cuda_gpu_execu
tor.cc:937] successful NUMA node read from SysFS had negative value (-1), but t
here must be at least one NUMA node, so returning NUMA node zero
2022-05-03 10:54:06.174122: I tensorflow/stream_executor/cuda/cuda_gpu_execu
tor.cc:937] successful NUMA node read from SysFS had negative value (-1), but t
here must be at least one NUMA node, so returning NUMA node zero
2022-05-03 10:54:06.175274: I tensorflow/core/platform/cpu_feature_guard.cc:14
2] This TensorFlow binary is optimized with oneAPI Deep Neural Network Librar
y (oneDNN) to use the following CPU instructions in performance-critical operatio
ns:  AVX2 AVX512F FMA
To enable them in other operations, rebuild TensorFlow with the appropriate comp
iler flags.
2022-05-03 10:54:06.175596: I tensorflow/stream_executor/cuda/cuda_gpu_execu
tor.cc:937] successful NUMA node read from SysFS had negative value (-1), but t
here must be at least one NUMA node, so returning NUMA node zero
2022-05-03 10:54:06.176302: I tensorflow/stream_executor/cuda/cuda_gpu_execu
tor.cc:937] successful NUMA node read from SysFS had negative value (-1), but t
here must be at least one NUMA node, so returning NUMA node zero
2022-05-03 10:54:06.176929: I tensorflow/stream_executor/cuda/cuda_gpu_execu
tor.cc:937] successful NUMA node read from SysFS had negative value (-1), but t
here must be at least one NUMA node, so returning NUMA node zero
2022-05-03 10:54:08.055622: I tensorflow/stream_executor/cuda/cuda_gpu_execu
tor.cc:937] successful NUMA node read from SysFS had negative value (-1), but t
here must be at least one NUMA node, so returning NUMA node zero
2022-05-03 10:54:08.056564: I tensorflow/stream_executor/cuda/cuda_gpu_execu
tor.cc:937] successful NUMA node read from SysFS had negative value (-1), but t
here must be at least one NUMA node, so returning NUMA node zero
```

```
2022-05-03 10:54:08.057260: I tensorflow/stream_executor/cuda/cuda_gpu_execu
tor.cc:937] successful NUMA node read from SysFS had negative value (-1), but t
here must be at least one NUMA node, so returning NUMA node zero
2022-05-03 10:54:08.058608: I tensorflow/core/common_runtime/gpu/gpu_device.
cc:1510] Created device /job:localhost/replica:0/task:0/device:GPU:0 with 15403
MB memory:  -> device: 0, name: Tesla P100-PCIE-16GB, pci bus id: 0000:00:04.
0, compute capability: 6.0
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (None, 60, 128)           66560
_________________________________________________________________
lstm_1 (LSTM)                (None, 64)                49408
_________________________________________________________________
dense (Dense)                (None, 25)                1625
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 26
=================================================================
Total params: 117,619
Trainable params: 117,619
Non-trainable params: 0
_________________________________________________________________
```

```
def convert_data(prediction_days_in_future, window_size):
    x = []
    y = []

    # Fetch 'WINDOW_SIZE' values as x and the value 'PREDICTION_TARGET' d
ays later as y
    # Slide this window across all days to maximise values
    for i in range(window_size, len(data)-prediction_days_in_future):
        x.append(scaled_data[i-window_size:i, 0])
        y.append(scaled_data[i+prediction_days_in_future-1, 0])

    x = np.array(x).reshape(len(x), window_size, 1)
    y = np.array(y)

    n_training_samples = int(len(y) * TRAIN_SIZE)
    X_train = x[:n_training_samples, :]
    y_train = y[:n_training_samples]
    X_test = x[n_training_samples:, :]
    y_test = y[n_training_samples:]
```

```
    return X_train, y_train, X_test, y_test, n_training_samples

def plot_results(train, output):
    plt.figure(figsize=(16,6))
    plt.title('Model')
    plt.xlabel('Date', fontsize=18)
    plt.ylabel('Close Price USD ($)', fontsize=18)
    plt.plot(train['Close'])
    plt.plot(output[['Close', 'Predictions']])
    plt.legend(['Train', 'Val', 'Predictions'], loc='lower right')
    plt.show()

def fit_model(days_in_future, window_size):
    print(f"Predicting for {days_in_future} days in future...")

    X_train, y_train, X_test, y_test, n_training_samples = convert_data(days_in_fut
ure, window_size)
    y_test_original = scaler.inverse_transform(y_test.reshape(-1, 1))
    print("Train samples:", len(y_train), "Test samples:", len(y_test))

    # Fit model
    model = get_model(window_size)
    history = model.fit(
        x=X_train,
        y=y_train,
        epochs=15,
        batch_size=64,
        validation_data=(X_test, y_test),
        callbacks=[EarlyStopping(monitor='val_loss', patience=3)]
    )

    # Get the models predicted price values
    predictions = model.predict(X_test)
    predictions = scaler.inverse_transform(predictions)

    # Get the root mean squared error (RMSE)
    rmse = np.sqrt(np.mean(((predictions - y_test_original) ** 2)))
    print("RMSE:", rmse)

    # Mean absolute percentage error
    mape = mean_absolute_percentage_error(y_test_original, predictions)
    print("Mean Absolute Percentage Error:", mape)

    # Final prediction error
```

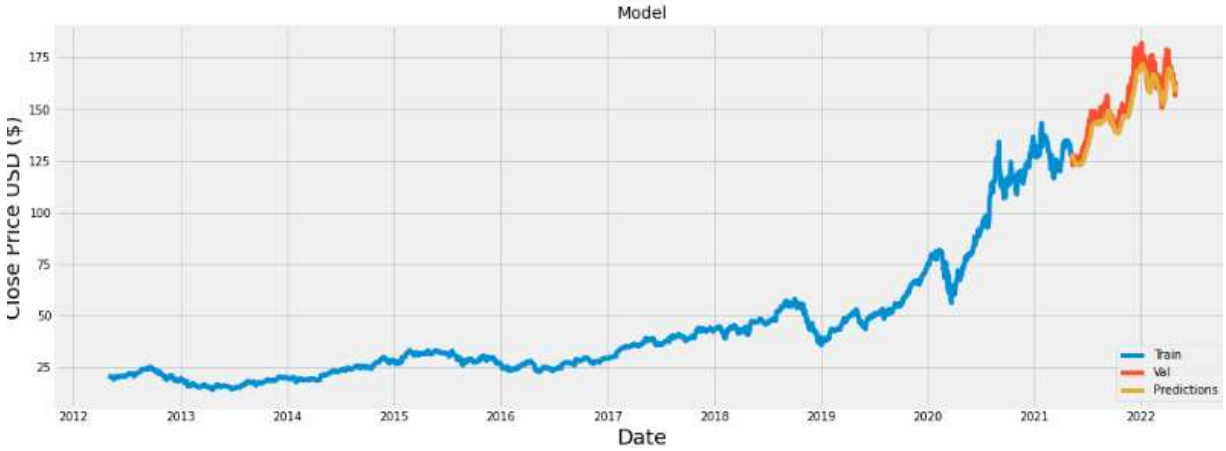
```
    error = (np.abs(predictions[-1] - y_test_original[-1]) / y_test_original[-1])[0]
    print("Final Prediction Error:", error)

    # Plot results
    train = apple[:n_training_samples+window_size+days_in_future-1]
    output = apple[n_training_samples+window_size+days_in_future-1:-1]
    output['Predictions'] = predictions
    plot_results(train, output)

fit_model(days_in_future=1, window_size=60)
Predicting for 1 days in future...
Train samples: 2211 Test samples: 246
2022-05-03 10:54:09.490376: I tensorflow/compiler/mlir/mlir_graph_optimization
_pass.cc:185] None of the MLIR Optimization Passes are enabled (registered 2)
Epoch 1/15
2022-05-03 10:54:12.872583: I tensorflow/stream_executor/cuda/cuda_dnn.cc:369
] Loaded cuDNN version 8005
35/35 [==============================] - 6s 26ms/step - loss: 0.0066 - v
al_loss: 0.0024
Epoch 2/15
35/35 [==============================] - 0s 8ms/step - loss: 2.9837e-04
- val_loss: 0.0010
Epoch 3/15
35/35 [==============================] - 0s 8ms/step - loss: 1.9484e-04
- val_loss: 0.0010
Epoch 4/15
35/35 [==============================] - 0s 8ms/step - loss: 1.9111e-04
- val_loss: 9.7768e-04
Epoch 5/15
35/35 [==============================] - 0s 8ms/step - loss: 1.8038e-04
- val_loss: 9.5286e-04
Epoch 6/15
35/35 [==============================] - 0s 8ms/step - loss: 1.7439e-04
- val_loss: 9.7527e-04
Epoch 7/15
35/35 [==============================] - 0s 9ms/step - loss: 1.7496e-04
- val_loss: 9.5797e-04
Epoch 8/15
35/35 [==============================] - 0s 9ms/step - loss: 1.6759e-04
- val_loss: 8.8263e-04
Epoch 9/15
35/35 [==============================] - 0s 8ms/step - loss: 1.6002e-04
- val_loss: 8.5036e-04
Epoch 10/15
```

35/35 [==============================] - 0s 8ms/step - loss: 1.6718e-04 - val_loss: 0.0016
Epoch 11/15
35/35 [==============================] - 0s 8ms/step - loss: 1.8488e-04 - val_loss: 0.0012
Epoch 12/15
35/35 [==============================] - 0s 8ms/step - loss: 1.5718e-04 - val_loss: 0.0017
RMSE: 6.884617458480209
Mean Absolute Percentage Error: 0.03503596897549511
Final Prediction Error: 0.03455420119763405
/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:38: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

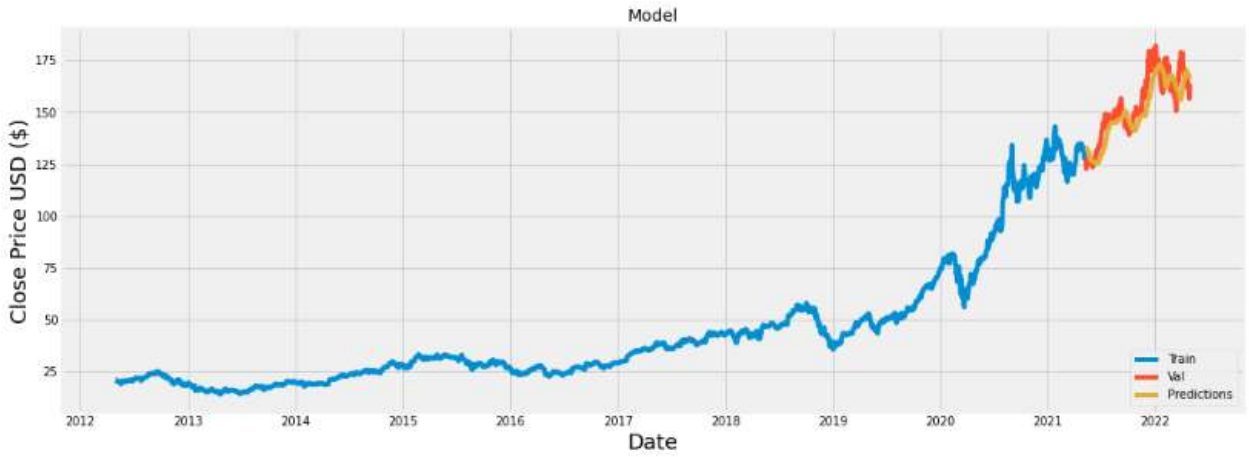
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view



Different date scopes

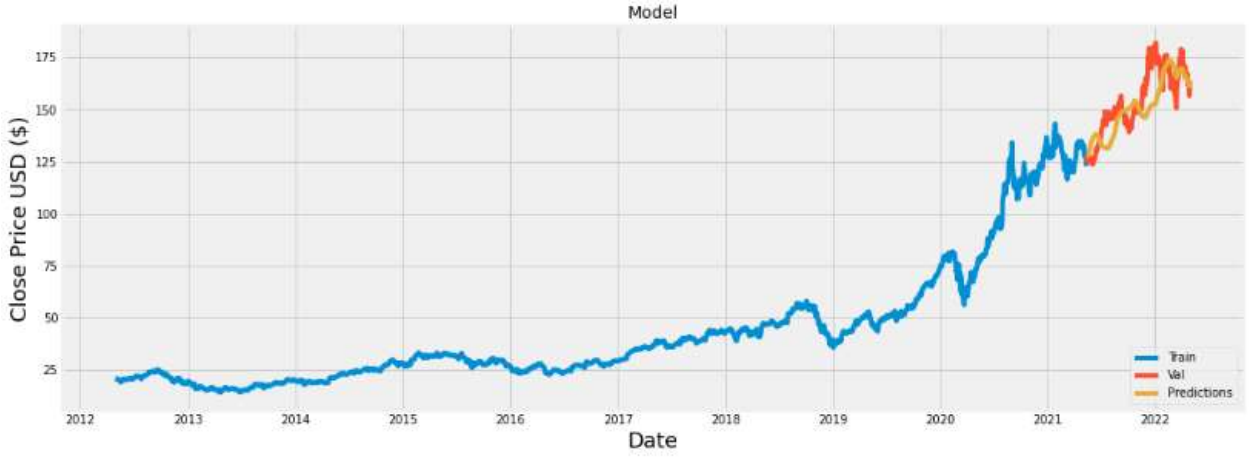
fit_model(days_in_future=7, window_size=60)
Predicting for 7 days in future...
Train samples: 2205 Test samples: 246
Epoch 1/15
35/35 [==============================] - 3s 34ms/step - loss: 0.0051 - val_loss: 0.0031
Epoch 2/15
35/35 [==============================] - 0s 8ms/step - loss: 4.8795e-04 - val_loss: 0.0021
Epoch 3/15
35/35 [==============================] - 0s 12ms/step - loss: 3.9110e-04 - val_loss: 0.0021
Epoch 4/15
35/35 [==============================] - 0s 13ms/step - loss: 4.0244e-04 - val_loss: 0.0021
Epoch 5/15
35/35 [==============================] - 0s 12ms/step - loss: 3.8607e-04 - val_loss: 0.0025
RMSE: 8.349175110481081
Mean Absolute Percentage Error: 0.042596585395891826
Final Prediction Error: 0.013735148113477872
/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:38: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy



fit_model(days_in_future=30, window_size=60)
Predicting for 30 days in future...
Train samples: 2185 Test samples: 243
Epoch 1/15
35/35 [==============================] - 3s 24ms/step - loss: 0.0065 - val_loss: 0.0093
Epoch 2/15
35/35 [==============================] - 0s 8ms/step - loss: 0.0015 - val_loss: 0.0041
Epoch 3/15
35/35 [==============================] - 0s 8ms/step - loss: 0.0013 - val_loss: 0.0042
Epoch 4/15
35/35 [==============================] - 0s 8ms/step - loss: 0.0013 - val_loss: 0.0039
Epoch 5/15
35/35 [==============================] - 0s 8ms/step - loss: 0.0013 - val_loss: 0.0060
Epoch 6/15
35/35 [==============================] - 0s 8ms/step - loss: 0.0013 - val_loss: 0.0071
Epoch 7/15
35/35 [==============================] - 0s 8ms/step - loss: 0.0012 - val_loss: 0.0042
RMSE: 10.95066439562247
Mean Absolute Percentage Error: 0.05502790896745705
Final Prediction Error: 0.01957132723957744
/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:38: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

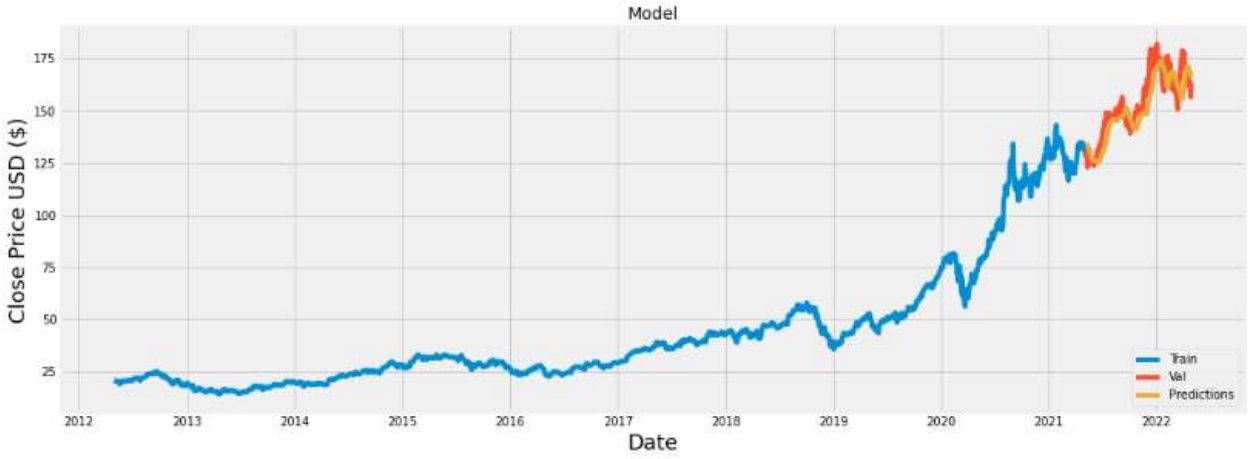
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy



Different window sizes

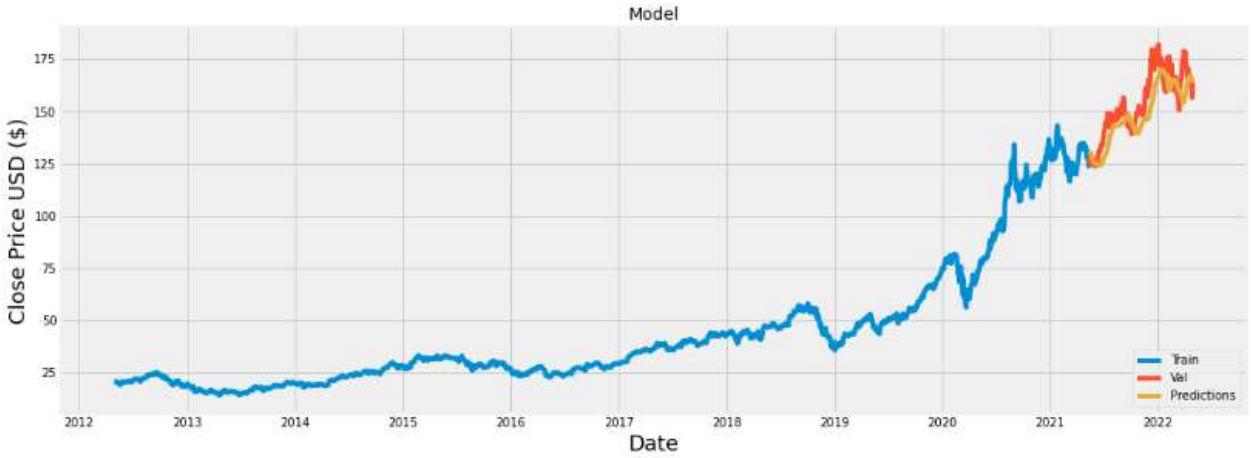
```
fit_model(days_in_future=7, window_size=30)
Predicting for 7 days in future...
Train samples: 2232 Test samples: 249
Epoch 1/15
35/35 [==============================] - 3s 21ms/step - loss: 0.0049 - v
al_loss: 0.0028
Epoch 2/15
35/35 [==============================] - 0s 6ms/step - loss: 4.4075e-04
- val_loss: 0.0030
Epoch 3/15
35/35 [==============================] - 0s 6ms/step - loss: 4.3322e-04
- val_loss: 0.0024
Epoch 4/15
35/35 [==============================] - 0s 6ms/step - loss: 4.0691e-04
- val_loss: 0.0026
Epoch 5/15
35/35 [==============================] - 0s 6ms/step - loss: 4.2287e-04
- val_loss: 0.0021
Epoch 6/15
35/35 [==============================] - 0s 6ms/step - loss: 4.1189e-04
- val_loss: 0.0034
Epoch 7/15
35/35 [==============================] - 0s 6ms/step - loss: 4.0511e-04
- val_loss: 0.0020
Epoch 8/15
35/35 [==============================] - 0s 6ms/step - loss: 5.2555e-04
- val_loss: 0.0044
Epoch 9/15
35/35 [==============================] - 0s 6ms/step - loss: 4.1338e-04
- val_loss: 0.0020
Epoch 10/15
35/35 [==============================] - 0s 6ms/step - loss: 4.0367e-04
- val_loss: 0.0023
RMSE: 8.090417358045938
Mean Absolute Percentage Error: 0.04116769623494656
Final Prediction Error: 0.013986632800932387
/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:38: SettingWithCop
yWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable
/user_guide/indexing.html#returning-a-view-versus-a-copy
```



fit_model(days_in_future=7, window_size=90)
Predicting for 7 days in future...
Train samples: 2178 Test samples: 243
Epoch 1/15
35/35 [==============================] - 3s 27ms/step - loss: 0.0045 - val_loss: 0.0024
Epoch 2/15
35/35 [==============================] - 0s 12ms/step - loss: 4.5109e-04 - val_loss: 0.0035
Epoch 3/15
35/35 [==============================] - 0s 11ms/step - loss: 6.3689e-04 - val_loss: 0.0025
Epoch 4/15
35/35 [==============================] - 0s 11ms/step - loss: 4.1874e-04 - val_loss: 0.0033
RMSE: 9.58127100562642
Mean Absolute Percentage Error: 0.04841007724234518
Final Prediction Error: 0.0007406536419915462
/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:38: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy


Model
Close Price USD ($)
175
150
125
100
75
50
25
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Date
Train
Val
Predictions