

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ  
СІКОРСЬКОГО»  
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ  
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

На правах рукопису  
УДК 004.942:519.216.3

До захисту допущено  
В. о. завідувача кафедри ММСА

О.Л.Тимошук

«\_\_\_» \_\_\_\_\_ 2020 р.

**Магістерська дисертація**

на здобуття ступеня магістра за спеціальністю 122 Комп'ютерні науки  
на тему: «Система підтримки прийняття рішень для прогнозування  
нелінійних нестационарних процесів у фінансах»

Виконав:

студент II курсу, групи КА-93 мп  
Гуць Євгеній Віталійович \_\_\_\_\_

Керівник: професор кафедри ММСА,  
д.т.н., професор,  
Бідюк П. І. \_\_\_\_\_

Рецензент: декан ФІОТ  
КПІ ім. Ігоря Сікорського,  
д. т. н., професор,  
Теленик С. Ф. \_\_\_\_\_

Засвідчую, що у цій магістерській дисертації  
немає запозичень з праць інших авторів  
без відповідних посилань

Студент \_\_\_\_\_

Київ  
2020

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ  
СІКОРСЬКОГО»  
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ  
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

Рівень вищої освіти — другий (магістерський)  
Спеціальність — 122«Комп'ютерні науки»

ЗАТВЕРДЖУЮ

В. о. завідувача кафедри ММСА

О. Л. Тимошук

«\_\_\_» \_\_\_\_\_ 2020 р.

### ЗАВДАННЯ

на магістерську дисертацію студенту Гуцю Євгенію Віталійовичу

**1. Тема дисертації:** «Система підтримки прийняття рішень для прогнозування нелінійних нестационарних процесів у фінансах», науковий керівник дисертації Бідюк Петро Іванович, д.т.н., професор, затверджені наказом по університету № 3182-с від 02.11.2020.

**2. Термін подання студентом дисертації:** 14 грудня 2020 р.

**3. Об'єкт дослідження:** нелінійні нестационарні процеси у фінансах, представлені статистичними даними стосовно їх розвитку.

**4. Предмет дослідження:** ймовірно-статистичні та методи інтелектуального аналізу даних для моделювання і прогнозування розвитку нелінійних нестационарних процесів; множини критеріїв для аналізу адекватності моделей та оцінювання якості прогнозів.

**5. Перелік завдань, які потрібно розробити:**

- 1) дослідити знайдений набір фінансових даних;
- 2) проаналізувати існуючі методи прогнозування нелінійних нестационарних процесів та виконати огляд використовуваних комп'ютерних систем для прогнозування;
- 3) дослідити методи та підходи до моделювання та прогнозування нелінійних та нестационарних процесів;
- 4) пошук даних для застосування в програмі;

5) вивчити та зрозуміти методики моделювання та прогнозування нелінійних нестационарних процесів та їхню структуру;

6) виконати побудову математичних моделей на реальних статистичних даних та отримати практичні результати, виконати аналіз результатів;

7) розробити стартап-проект виведення на ринок результатів дослідження;

8) розробити концептуальні висновки за результатами наукового дослідження

#### **6. Орієнтовний перелік графічного (ілюстративного) матеріалу:**

1). Графіки, створені розробленою програмою (рис.);

2). Приклади результатів побудованих моделей (рис.);

3). Таблиці у розділі стартап-проекту та порівняльні таблиці результатів;

4). Слайди презентації.

**7. Дата видачі завдання:** 01 вересня 2020 р.

#### **Календарний план**

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації
1.	Концептуальний вступ дисертації. Формулювання об'єкта, предмета, цілі, завдань, новизни, практичної значущості результатів	06.09.2020 — 13.09.2020
2.	Перший розділ. Огляд літературно-інформаційних джерел. Характеристика існуючих проблем	14.09.2020 — 27.09.2020
3.	Другий розділ. Огляд математичних методів прогнозування нелінійних нестационарних процесів	28.09.2020 — 11.10.2020
5.	Третій розділ. Формування структури моделі прогнозування та основних її етапів	12.10.2020 — 21.10.2020
6.	Четвертий розділ. Реалізація системи та отримання практичних результатів	22.10.2020 — 06.11.2020
7.	П'ятий розділ. Стартап-проект	07.11.2020 — 14.11.2020
8.	Концептуальні висновки. Перспективи розвитку отриманих рішень	15.11.2020 — 22.11.2020

Студент

Є.В. Гуць

Науковий керівник дисертації

П.І. Бідюк

## РЕФЕРАТ

Магістерська дисертація: 118с., 44 рис., 36 табл., 1 додаток, 18 джерел.

НЕЛІНІЙНІ НЕСТАЦІОНАРНІ ПРОЦЕСИ, ФІНАНСОВІ ПРОЦЕСИ,  
АКЦІЇ КОМПАНІЙ, ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ, РЕКУРЕНТНІ  
НЕЙРОННІ МЕРЕЖІ, ПРОГНОЗУВАННЯ

Дана робота присвячена вивченню методів прогнозування та методик побудови моделей нелінійних нестационарних процесів, а саме методу рекурентних нейронних мереж та застосування його на статистичних наборах фінансових даних.

Об'єктом дослідження є нелінійні нестационарні процеси, подані у вигляді статистичних даних стосовно ціноутворення акцій великих світових технологічних компаній.

Предметом дослідження є ймовірно-статистичні та методи інтелектуального аналізу даних для моделювання і прогнозування розвитку нелінійних нестационарних процесів, множини критеріїв для аналізу адекватності моделей та оцінювання якості прогнозів.

Метою магістерської роботи є вивчення основних засад побудови моделей та прогнозування з використанням нейронної мережі LSTM, а також розробка програмного продукту для отримання практичних результатів та порівняння роботи алгоритму на різних наборах фінансових даних.

## ABSTRACT

The master thesis consist of 118 pages, 44 images, 36 tables , 1 appendice, 18 sources.

The theme: «Decision support system for forecasting nonlinear non-stationary processes in finance».

NONLINEAR PROCESSES, FINANCIAL PROCESSES, COMPANY STOCKS, INTELLECTUAL ANALYSIS OF DATA, RECURRENT NEURAL NETWORKS, FORECASTING

This work is devoted to the study of forecasting methods and methods of constructing models of nonlinear non-stationary processes, the recurrent neural networks and its application on statistical data sets of financial data.

The object of the work is nonlinear non-stationary processes, presented in the form of statistical data on the pricing of shares of major global technology companies.

The subject of research is probabilistic-statistical and data mining methods for modeling and forecasting the development of nonlinear non-stationary processes, sets of criteria for analyzing the adequacy of models and assessing the quality of forecasts.

The purpose of the thesis is to study the main principles of models construction and forecasting using the neural network LSTM, as well as the development of a software product for obtaining practical results and comparing the operation of the algorithm on different sets of financial data.

## ЗМІСТ

ВСТУП.....	8
<b>РОЗДІЛ 1: АКТУАЛЬНІСТЬ ПРОБЛЕМИ ДОСЛІДЖЕННЯ ТА АНАЛІЗ ІСНУЮЧИХ СИСТЕМ .....</b>	<b>10</b>
1.1 Аналіз актуальності проблеми моделювання і прогнозування нелінійних нестационарних процесів (ННП) у фінансах .....	10
1.2 Існуючі методи моделювання і прогнозування ННП у фінансах .....	11
1.2.1 Методи інтелектуального аналізу даних.....	12
1.2.2 Основні етапи інтелектуального аналізу даних .....	13
1.3 Деякі сучасні системи для моделювання і прогнозування ННП.....	14
1.4 Опис знайдених фінансових даних, що будуть використані у роботі .....	18
1.5 Висновки до розділу і постановка задачі дослідження.....	21
<b>РОЗДІЛ 2: ВИБРАНІ МЕТОДИ І ПІДХОДИ ДО МОДЕЛЮВАННЯ І ПРОГНОЗУВАННЯ НЕЛІНІЙНИХ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ В ЕКОНОМІЦІ ТА ФІНАНСАХ .....</b>	<b>23</b>
2.1 Найпростіші математичні моделі для нелінійних нестационарних процесів.....	23
2.1.1 Модель авторегресії (AR – Autoregressive model).....	23
2.1.2 Модель ковзного середнього (MA – Moving Average) .....	24
2.1.3 Модель авторегресії з ковзним середнім (ARMA - Autoregressive Moving Average).....	26
2.1.4 Лінійна регресія.....	27
2.1.5 Метод опорних векторів .....	27
2.2 Рекурентні нейронні мережі .....	29
2.2.1 Недоліки рекурентних нейронних мереж .....	31
2.2.2 Мережа LSTM.....	33
2.2.3 Покроковий розбір мережі LSTM.....	35
2.3 Байєсівські мережі для моделювання нелінійних нестационарних процесів .....	37
2.3.1 Теорема Байєса як інструмент формування ймовірного висновку БМ .....	39
2.4 Висновки до розділу 2 .....	40
<b>РОЗДІЛ 3: МЕТОДИКА МОДЕЛЮВАННЯ ГЕТЕРОСКЕДАСТИЧНИХ ПРОЦЕСІВ .....</b>	<b>41</b>
3.1 Методика створення моделей гетероскедастичних процесів .....	41
3.2 Множини критеріїв для аналізу якості даних, адекватності моделей та якості оцінок прогнозів .....	43
3.3 Тести для аналізу даних на лінійність та гетероскедастичність .....	46
3.4 Висновки до розділу 3 .....	49
<b>РОЗДІЛ 4: ПОБУДОВА СППР ТА ВИКОНАННЯ ОБЧИСЛЮВАЛЬНИХ ЕКСПЕРИМЕНТІВ.....</b>	<b>50</b>

	7
4.1 Вибір інструментальної платформи для реалізації системи.....	50
4.2 Попередній аналіз та візуалізація даних.....	51
4.3 Використання нейронної мережі LSTM для прогнозування акцій компаній.....	62
4.4 Аналіз результатів, отриманих іншими моделями, та порівняння з LSTM. ....	66
4.5 Висновки по роботі.....	68
<b>РОЗДІЛ 5: РОЗРОБКА ВЛАСНОГО СТАРТАП ПРОЕКТУ .....</b>	<b>70</b>
5.1 Сутність та особливості стартап проектів .....	70
5.2 Формування команди стартапу.....	73
5.3 Розробка продукту для стартапу .....	75
5.4 Розроблення бізнес-моделі стартапу.....	78
5.5 Маркетингове планування стартапу .....	80
5.6 Бізнес-план стартап проекту.....	85
5.7 Патентний пошук.....	88
5.8 Правові аспекти реалізації стартапів, інтелектуальна власність та патентування .....	90
5.9 Інвестиційна та фінансове забезпечення стартап-проекту .....	103
<b>ПЕРЕЛІК ПОСИЛАНЬ .....</b>	<b>106</b>
<b>ДОДАТОК А ЛІСТИНГ ПРОГРАМИ .....</b>	<b>108</b>

## ВСТУП

Фінансові та економічні процеси є невід'ємною частиною нашого повсякденного життя. Тому дослідження та вивчення природи нелінійних нестационарних процесів вимагає все більшої уваги науковців.

Великі обсяги даних, які людству вдалось накопичити за часи свого існування, потребують відповідної обробки для прийняття певних рішень. Математичне моделювання все більше набуває популярності для ефективної організації управління певними суб'єктами господарювання та економічними спільнотами, тому що вірність прийнятих рішень суттєво залежить від якості прогнозування їх наслідків. Саме через це рішення, які ми приймаємо сьогодні, повинні бути обґрунтовані та базуватись на достовірних оцінках можливого розвитку процесів та подій у майбутньому.

Досить часто в даній роботі зустрічається термін «прогноз». Він перекладається з грецької як «знання наперед» та означає результат наукового дослідження певних перспектив розвитку якогось процесу. Для прогнозування використовують математичні методи та моделі, які можуть бути достатньо складними і нетривіальними. Внаслідок широкого розповсюдження персональних комп'ютерів, простих для розуміння та використання, а головне, ефективних програм прогнозування стало швидким і досить ефективним механізмом аналізу.

Для знаходження закономірностей в масивах даних використовуються методи інтелектуального аналізу даних. Вони базуються на використанні однієї з двох технологій: машинного навчання (Machine Learning) чи візуалізації (візуальне подання інформації). Байєсівські мережі відомі тим, що об'єднують у собі одразу дві технології.

Байєсівські мережі є типом ймовірнісної графічної моделі, що використовує байєсівський висновок для обчислення ймовірностей. Ці мережі



прагнуть моделювати умовну залежність і, отже, причинно-наслідковий зв'язок, подаючи умовну залежність ребрами в орієнтованому графі.

Сьогодні для побудови моделі широко використовуються нейронні мережі. Цікавим підвидом є рекурентні нейронні мережі, які містять зворотні зв'язки та можуть зберігати інформацію. Ця властивість дає змогу прогнозувати, спираючись на контекст, що дозволяє ефективно вирішувати цілий ряд задач: розпізнавання мовлення, моделювання голосу, переклад, розпізнавання зображень, прогнозування часових рядів та інші.

Прогнозування є одним із найпоширеніших завдань ІАД. Зокрема, при стратегічному плануванні і складанні бюджету необхідно прогнозувати багато параметрів з урахуванням численних взаємозв'язаних сезонних, регіональних, загальноекономічних та інших чинників. Досить актуальним є пошук нових або комбінування наявних методів прогнозування, які дають можливість отримати більш точну та адекватну модель, що дозволяє приймати ефективні управлінські рішення.

## РОЗДІЛ 1: АКТУАЛЬНІСТЬ ПРОБЛЕМИ ДОСЛІДЖЕННЯ ТА АНАЛІЗ ІСНУЮЧИХ СИСТЕМ

### 1.1 Аналіз актуальності проблеми моделювання і прогнозування нелінійних нестационарних процесів (ННП) у фінансах

Моделювання процесів та прогнозування широко використовується під час прийняття рішень. Під прогнозуванням розуміється передбачення майбутнього стану тієї чи іншої системи, зважаючи на дані з минулого та припущення відносно майбутнього.

Прогнозування відіграє досить важливу роль у роботі будь-якого підприємства, адже коли ми можемо приблизно оцінити, що очікує нас у близькому майбутньому, можна змінити напрямок розвитку підприємства, методи, які використовуються, інколи і вберегти від банкрутства.

Майже всі банківські установи користуються програмами для прогнозування ризиків при кредитуванні, проте більшість таких систем не завжди роблять правильний прогноз, що, зважаючи на великі грошові потоки, може перерости у серйозні збитки.

Вимоги сучасного бізнесу змушують переходити від простого статистичного аналізу до більш складного інтелектуального аналізу даних, який покликаний забезпечити побудову моделі для опису інформації і в кінцевому результаті призвести до необхідних висновків.

Інтелектуальним аналізом даних, скорочено ІАД, називається механізм обробки даних та виявлення в них ключових ознак, який дозволяє приймати ефективні рішення. Основні засади інтелектуального аналізу даних відомі досить давно, проте з появою досить об'ємних масивів інформації вони набули особливої актуальності. Ці зміни вплинули на зростання популярності методів інтелектуального аналізу даних, зокрема, тому, що інформація стає все більш різноманітною і всеохоплюючою, а не лише об'ємною.

Процес аналізу даних та побудови моделі здебільшого є ітеративним, оскільки потрібно відшукати та виявити різні відомості. Необхідно також розуміти, як пов'язати ті чи інші дані для забезпечення оптимального результату.

## 1.2 Існуючі методи моделювання і прогнозування ННП у фінансах

Алгоритм в інтелектуальному аналізі даних – це набір евристик та обчислень, який на основі цих даних створює модель. Алгоритм працює таким чином: спочатку аналізуються наявні дані, а потім здійснюється пошук певних закономірностей. Результат цього аналізу використовується ітеративно для створення оптимальних параметрів моделі інтелектуального аналізу даних. Згодом ці параметри застосовуються до всього набору даних з метою виявлення придатних до використання закономірностей.

Можна видалити наступні форми моделей:

- набір кластерів, що описують зв'язки варіантів з набору даних;
- дерево рішень, яке передбачає результат і описує вплив на нього різних факторів;
- математична модель, що здійснює прогнозування;
- набір правил, що здійснюють групування об'єктів в транзакціях, та аналіз їх ймовірностей.

Здебільшого вибір правильного алгоритму для використання в певній аналітичній задачі є досить непростим завданням. Можлива ситуація, коли використання різних алгоритмів для однієї і тієї ж задачі призведе до отримання різних результатів, крім того, деякі алгоритми можуть видавати більше одного типу результатів.

Наприклад, алгоритм дерева прийняття рішень можна використовувати не лише для прогнозування, а також для зменшення кількості стовпців в

наборі даних, оскільки він може виявляти стовпці, що не впливають на кінцеву модель інтелектуального аналізу даних.

### 1.2.1 Методи інтелектуального аналізу даних

#### А) Класифікація

Основна задача методу – встановити, до якого з класів відносяться ті чи інші дані, при цьому множина класів має бути заздалегідь визначена. Кожен клас володіє певними властивостями, що характеризують його об'єкти. Наприклад, в задачі виявлення кредитоспроможності клієнта банківський працівник оперує двома класами: «кредитоспроможний» і «некредитоспроможний». Віднести клієнта до однієї із груп допомагає аналіз ряду характеристик: віку, рівня доходів, місця роботи та інші. Це означає, що задача інтелектуального аналізу даних зводиться до того, щоб визначити один із параметрів об'єкта, аналізуючи значення інших його параметрів. З точки зору математики це виглядає так: слід визначити значення залежної змінної «кредитоспроможність», що може набувати значення «так» або «ні» за відомими значеннями незалежних змінних, таких як вік, рівень доходу, місце роботи та інші.

Для класифікації в ІАД використовуються нейронні мережі, дерева рішень, опорні вектори, метод к-середніх, алгоритми покриття та інші.

#### Б) Кластеризація

Метод, на перший погляд дуже схожий на класифікацію. По суті є логічним його продовженням, узагальненням задачі класифікації, коли набір класів заздалегідь невизначений. Його ще називають кластерним аналізом. Кластеризація відрізняється від класифікації тим, що для проведення аналізу не вимагається мати виокремлену залежну змінну. Задачі розв'язується на початкових етапах дослідження, коли про дані мало що відомо. Її розв'язання

допомагає ґрунтовно усвідомити дані і в цьому сенсі задача кластеризації є описовою. Після визначення кластерів застосовуються інші методи інтелектуального аналізу даних з метою встановлення, що означає таке розбиття і чим воно обумовлене.

Наприклад, цим методом користуються навіть малюки, розрізняючи об'єкти за принципом схожості-несхожості. Саме так вони здатні відрізнити ляльку від машинки, чоловіків від жінок, кішок від собак. Іншим прикладом є задача сегментації ринку, в основі якої лежить гіпотеза, що всі клієнти різні, проте за схожістю їх можна розділити на певну кількість груп.

#### В) Прогнозування

Цим методом часто користуються підприємці, аналізуючи наявні дані, що дозволяє будувати прогнози на майбутнє. До того ж, чим детальніші архівні дані і чим більший часовий інтервал їх охоплює, тим точнішими будуть результати. Цей метод часто використовується для передбачення потреби у фінансуванні підприємства, оцінки попиту на товари чи послуги, прогнозування ринку збуту продукції. Якщо, наприклад, власник квіткового магазину хоче визначити, скільки квітів потрібно замовити на свято, він має проаналізувати цифри відповідних продажів за останні кілька років.

### 1.2.2 Основні етапи інтелектуального аналізу даних

Аналітики виділяють наступні етапи інтелектуального аналізу даних:

А) Розуміння бізнесу: визначення мети проекту і вимог з боку бізнесу. Згодом ці знання втілюються в постановку задачі і попередній план по досягненню мети проекту.

Б) Розуміння даних: починається із збору даних і має на меті їх освоєння. Для цього необхідно виявити проблеми з якістю даних і сформулювати гіпотези про приховані закономірності цих даних.

В) Підготовка даних: метою є отримання результуючих наборів даних, що будуть використовуватись при моделюванні з різноформатних даних. При цьому відбувається відбір таблиць, записів та атрибутів, а також конвертація і очистка даних для моделювання.

Г) Моделювання: до даних застосовуються різноманітні методики, будуються моделі, їх параметри оптимізуються.

Д) Оцінка: передбачає оцінювання якості моделі перед її втіленням та виявлення бізнес-завдань, що були недостатньо вивчені.

Е) Розгортання: фінальний етап, що полягає у створенні звіту, або автоматизації процесу аналізу даних для вирішення бізнес-задач.

### 1.3 Деякі сучасні системи для моделювання і прогнозування ННП

На сьогоднішній день розроблено багато готових пакетних рішень для роботи із нелінійними нестационарними процесами. Серед них існують як безкоштовні (пробні та навчальні) версії, так і платні пакети для комерційного використання у великих підприємствах та корпораціях.

Досить популярним програмним продуктом є EViews. Ця програма пропонує академічним дослідникам, корпораціям, державним установам та студентам доступ до масивних статистичних часових рядів, прогнозування та моделювання за допомогою інноваційного, простого у використанні об'єктно-орієнтованого інтерфейсу (рис. 1.1).

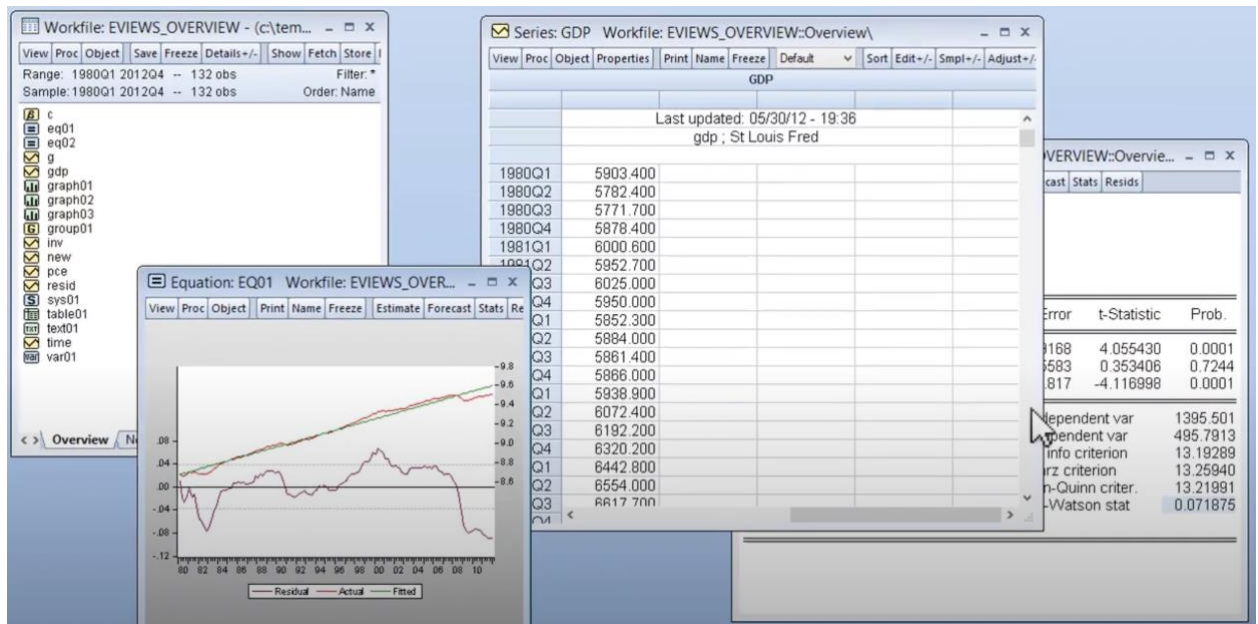


Рисунок 1.1 – Інтерфейс EViews.

Потужним аналітичним інструментом є SAS («Statistical Analysis System»). Він включає в себе набір статистичних програм, розроблений Інститутом SAS для управління даними, вдосконаленої аналітики, багатовимірного аналізу, бізнес-аналітики, розслідування злочинів та прогнозної аналітики (рис. 1.2).

Згідно з дослідженням агентства IDC компанія SAS лідирує в сегменті засобів поглибленої аналітики, займаючи 30,5 відсотків ринку (удвічі більше, ніж найближчий конкурент). У їх засобах поглибленої аналітики реалізовані передові інноваційні алгоритми, що дозволяють вирішувати найскладніші завдання, знаходити оптимальне рішення та виявляти можливості, які могли б залишитися непоміченими.

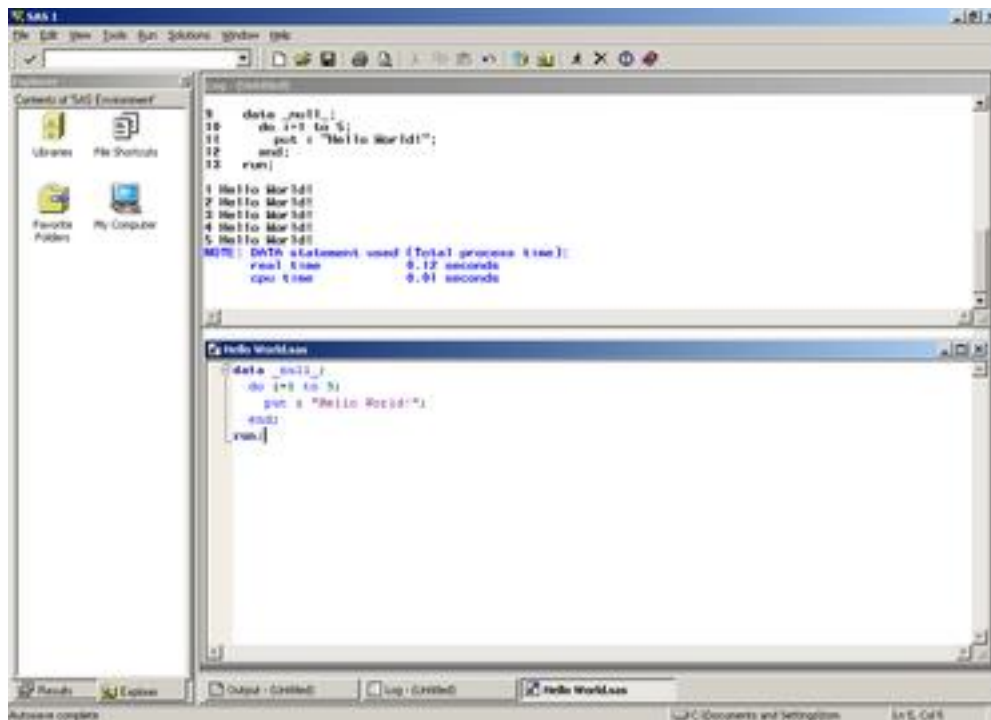


Рисунок 1.2 – Інтерфейс SAS 9 на Microsoft Windows.

Компанія DELL розробила власний статистичний пакет STATISTICA, який реалізує функції аналізу даних, управління даними, видобутку даних та їх візуалізації із залученням статистичних методів (рис. 1.3). Пакет має широкі графічні можливості, дозволяє виводити інформацію у вигляді різних типів графіків (включаючи наукові, ділові, тривимірні і двомірні графіки в різних системах координат, спеціалізовані статистичні графіки - гістограми, матричні, категоріальні графіки та ін.), всі компоненти графіків налаштовуються. У травні 2017 року Statistica була придбана компанією TIVCO.





Рисунок 1.3 – Інтерфейс ПЗ STATISTICA.

Простим для розуміння і, водночас, зручним у використанні є програмний модуль Deductor, який орієнтований на аналітиків, а не на програмістів (рис. 1.4). В системі логіка аналізу реалізується за рахунок візуальної побудови сценаріїв, без написання програмного коду, за допомогою всього 4-х майстрів. При побудові сценаріїв можна довільним чином комбінувати методи очищення, попередньої обробки, Data Mining від простих формул до складних алгоритмів. В Deductor реалізовані всі сучасні технології аналізу структурованих даних: Data Warehouse, ETL, OLAP, Data Mining. Це дозволяє на базі однієї платформи реалізувати повний аналітичний цикл: від очищення і консолідації до прогнозування і оптимізації.

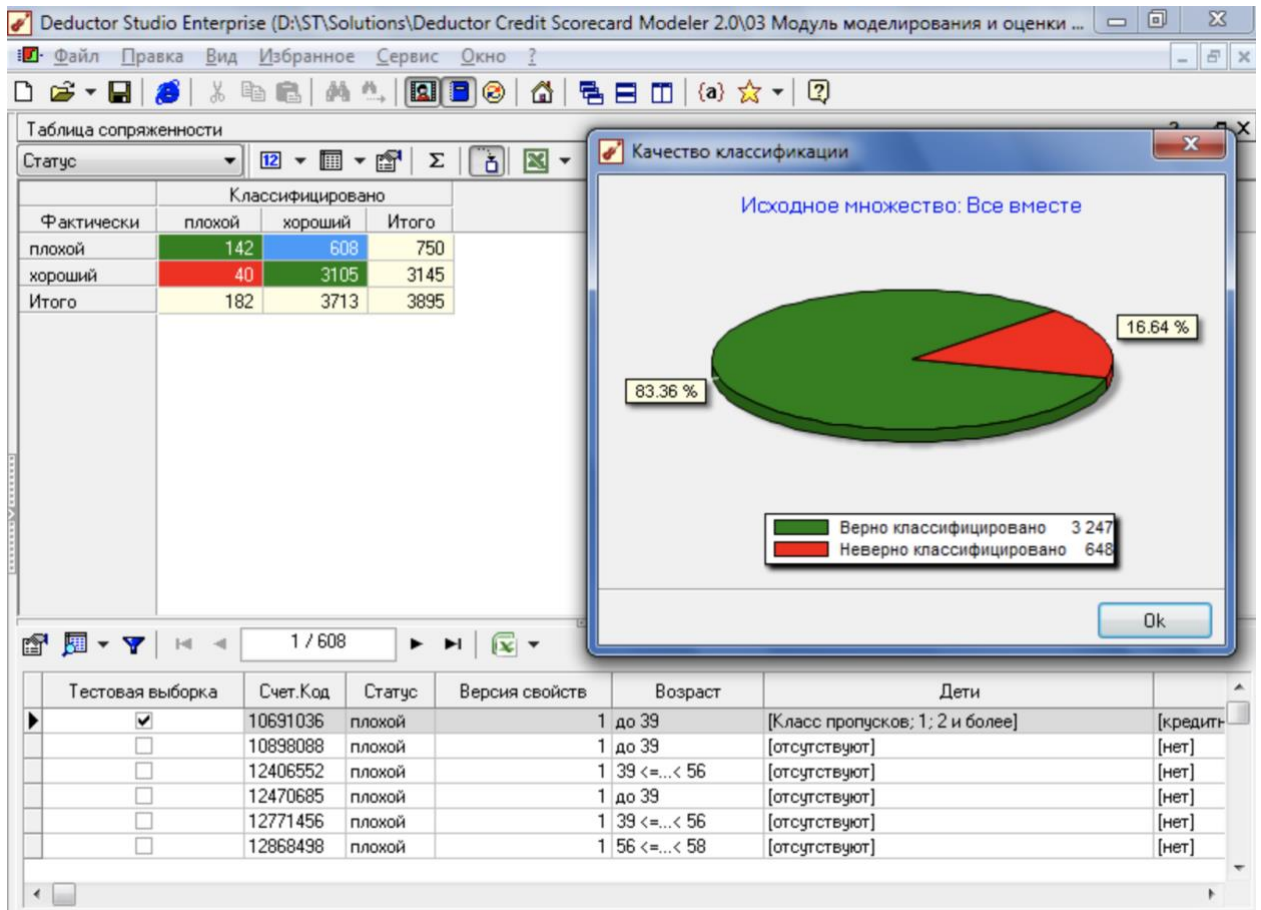


Рисунок 1.4 – Интерфейс Deductor Studio.

#### 1.4 Опис знайдених фінансових даних, що будуть використані у роботі

У період технічного прогресу спостерігається ріст впливу великих корпорацій, які впевнено займають лідируючі місця на світовому ринку та мають мільярдні прибутки. Ціноутворення акцій світових гігантів є дуже цікавим для вивчення процесом, який, як правило, є нестационарним, оскільки містить тренд та змінну у часі дисперсію.

Американський сервіс «Yahoo! Finance» надає досить детальні дані по темам, пов'язаним з бізнесом, фінансами та економікою. Довідкова інформація включає в себе котирування і рейтинги цінних паперів, прес-релізи та фінансові звіти компаній [1, 2, 3, 4].

Для роботи я обрав 4 відомі компанії: Apple, Google, Microsoft, Amazon, дані яких, на мою думку, є досить цікавими для вивчення і моделювання. Прогнозування ціни акцій, можливо, дасть змогу побачити, яка з компаній стане новим світовим лідером, а яка втратить свої позиції на ринку.

Набір даних містить наступні колонки:

- «Date» - дата торгів у форматі YYYY-MM-DD;
- «High» - максимальна ціна, яку люди заплатили за акції;
- «Low» - мінімальна ціна, яку люди заплатили за акції;
- «Open» - це перша ціна торгів, зафіксована, коли ринок відкрився;
- «Close» - це остання ціна торгів, зафіксована, коли ринок закрився;
- «Volume» - цей показник показує загальну кількість акцій, що торгуються за день;
- «Adj Close» - відкоригована ціна закриття з урахуванням дивідендів та поділів;
- «company\_name» - назва компанії.

На рисунках 1.5 – 1.7 наведено графічні подання даних у декількох можливих варіантах.

	High	Low	Open	Close	Volume	Adj Close	company_name
<b>Date</b>							
2020-10-06	3182.000000	3090.000000	3165.000000	3099.959961	5086900.0	3099.959961	AMAZON
2020-10-07	3200.000000	3132.389893	3135.000000	3195.689941	4309400.0	3195.689941	AMAZON
2020-10-08	3233.290039	3174.989990	3224.989990	3190.550049	3174100.0	3190.550049	AMAZON
2020-10-09	3288.989990	3197.830078	3210.000000	3286.649902	4907900.0	3286.649902	AMAZON
2020-10-12	3496.239990	3339.550049	3349.939941	3442.929932	8364200.0	3442.929932	AMAZON
2020-10-13	3492.379883	3424.219971	3467.989990	3443.629883	5744700.0	3443.629883	AMAZON
2020-10-14	3464.879883	3340.000000	3447.000000	3363.709961	5828900.0	3363.709961	AMAZON
2020-10-15	3355.879883	3280.000000	3292.010010	3338.649902	5223400.0	3338.649902	AMAZON
2020-10-16	3399.659912	3160.000000	3363.229980	3272.709961	6446500.0	3272.709961	AMAZON
2020-10-19	3329.000000	3203.570068	3299.610107	3211.219971	3636082.0	3211.219971	AMAZON

Рисунок 1.5 – Останні 10 значень датасету для компанії Amazon.



Рисунок 1.6 – Графіки динаміки зміни значень колонки «Adj Close» для 4 обраних компаній за останні 5 років.

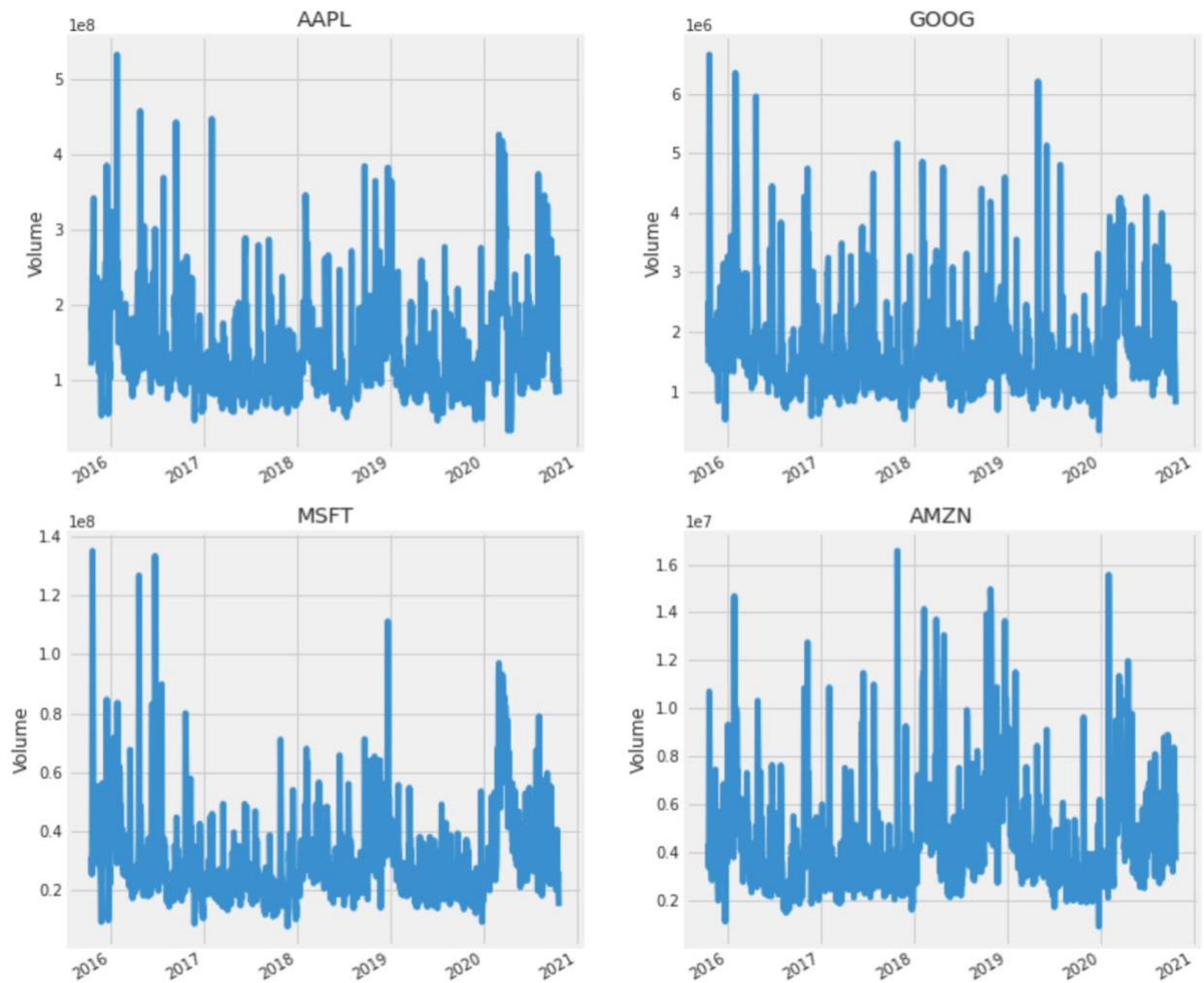


Рисунок 1.7 – Графіки динаміки зміни значень колонки «Volume» для 4 обраних компаній за останні 5 років.

### 1.5 Висновки до розділу і постановка задачі дослідження

У сучасному світі будь-які процеси можуть бути описані математично. Нинішній рівень фінансової сфери людської діяльності вимагає високої точності обчислень та прогнозів. Для того, аби мати змогу вести подальший аналіз процесу та прогнозувати його результати, слід відтворити процес у вигляді математичної моделі, що повністю відображає його поведінку. Математичні моделі можуть бути різноманітного характеру: лінійні або нелінійні, стаціонарні або нестаціонарні. Звісно, в природі найбільш

розповсюджені нелінійні нестационарні процеси. Тому в даній роботі розглядається матеріал, що стосується моделювання та прогнозування нелінійних нестационарних процесів.

У цьому розділі також було розглянуто основні підходи обробки інформації, методи та етапи інтелектуального аналізу даних та особливості нелінійних нестационарних процесів та їх моделей. Також продемонстровано перелік найбільш популярних сучасних систем для моделювання та прогнозування ННП.

Постановка задачі:

а) проаналізувати існуючі методики та проблеми моделювання фінансових нелінійних нестационарних процесів;

б) розглянути найбільш поширені методи моделювання і прогнозування нелінійних нестационарних процесів;

в) застосувати різні методи моделювання та прогнозування на обраних наборах даних і виконати порівняльний аналіз отриманих результатів;

г) виробити рекомендації щодо можливостей подальшого вдосконалення розробленої системи підтримки прийняття рішень.

## РОЗДІЛ 2: ВИБРАНІ МЕТОДИ І ПІДХОДИ ДО МОДЕЛЮВАННЯ І ПРОГНОЗУВАННЯ НЕЛІНІЙНИХ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ В ЕКОНОМІЦІ ТА ФІНАНСАХ

### 2.1 Найпростіші математичні моделі для нелінійних нестационарних процесів

#### 2.1.1 Модель авторегресії (AR – Autoregressive model)

Спочатку розглянемо модель авторегресії (AR), оскільки вона є однією з найбільш простих та зрозумілих. Основна ідея полягає в тому, що ми моделюємо значення в момент часу  $t$  як лінійну функцію, залежну від його попередніх значень і деякого випадково розподіленого доданка. Розглянута модель AR(N) описується наступною формулою:

$$y(k) = a_0 + \sum_{i=1}^N a_i y(k-i) + b_1 k + \dots + b_m k^m + \varepsilon(k);$$

У цій формулі  $a_i$  - коефіцієнти авторегресії,  $y(k)$  - досліджуваний ряд,  $k = 1, 2, 3, \dots$  - дискретний час,  $t = kT_s$ , де  $T_s$  – час виміру спроби, а  $N$  - порядок моделі, який, зазвичай значно менший від довжини серії. Шумовий доданок або залишок,  $\varepsilon$  у даній формулі, зазвичай вважається гаусівським білим шумом. Як бачимо, у цій моделі присутня стаціонарність, оскільки прослідковується тренд порядку  $m$  (ступінь  $k$ ).

Для визначення коефіцієнтів авторегресії використовується цілий ряд методів. Найбільш поширеним є метод найменших квадратів, який базується на рівняннях Юла-Уокера, котрі можна записати в матричній формі:

$$\begin{pmatrix} 1 & r_1 & r_2 & r_3 & \dots & r_{N-1} \\ r_1 & 1 & r_1 & r_2 & \dots & r_{N-2} \\ r_2 & r_1 & 1 & r_1 & \dots & r_{N-3} \\ r_3 & r_2 & r_1 & 1 & \ddots & r_{N-4} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ r_{N-1} & r_{N-2} & r_{N-3} & r_{N-4} & \dots & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ \vdots \\ a_N \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ \vdots \\ r_N \end{pmatrix}$$

### 2.1.2 Модель ковзного середнього (MA – Moving Average)

У статистиці метод ковзного середнього широко застосовується для «згладжування» функції шляхом створення серії середніх для різних підмножин повного набору даних. Перший елемент ковзного середнього отримують, взявши середнє значення початкового фіксованого набору елементів підмножини набору чисел. Тоді підмножина визначається шляхом «переміщення вперед», тобто виключаючи перше число серії і включаючи наступне значення в підмножині. Модель MA(q) описується наступною формулою:

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

Ковзне середнє зазвичай використовується у роботі з даними часових рядів для згладжування короткострокових коливань і виявлення довгострокових тенденцій або циклів. Наприклад, дана модель часто використовується в технічному аналізі фінансових даних, таких як ціни на акції, прибутки або обсяги товарообігів. Вона також використовується в економіці для визначення валового внутрішнього продукту, зайнятості працівників або інших макроекономічних показників. З точки зору математики, ковзне середнє значення є видом згортки і тому його можна



розглядати як приклад фільтра низьких частот, що використовується для обробки сигналів.

Існує декілька різновидів ковзного середнього:

- Просте ковзне середнє (SMA – Simple Moving Average)

У цій моделі значення шукається як незважене середнє попередніх  $n$  показників. Для набору  $a_M, a_{M-1}, a_{M-2}, a_{M-3}, \dots, a_{M-(n-1)}$  справедлива така формула розрахунку простого ковзного середнього:

$$\overline{a_M} = \frac{a_M + a_{M-1} + a_{M-2} + a_{M-3} + \dots + a_{M-(n-1)}}{n}$$

- Сукупне ковзне середнє (CMA - Cumulative Moving Average)

У сукупному ковзному середньому подана інформація утворює впорядкований потік даних, і користувач хоче отримати середнє значення всіх даних до поточної вхідної точки. Наприклад, інвестор хоче, щоб середня ціна всіх біржових операцій була з точністю до поточного часу. Формула розрахунку для  $n$  даних виглядає так:

$$CMA_n = \frac{x_1 + \dots + x_n}{n}$$

Для розрахунку  $n + 1$  значення можна використати попередньо обчислене значення:

$$CMA_{n+1} = \frac{x_{n+1} + n \cdot CMA_n}{n + 1}$$

- Зважене ковзне середнє (WMA - Weighted Moving Average)

Зважене ковзне середнє розраховується подібно до простого ковзного середнього, за винятком того, що даним, розташованим у різних місцях

ковзного вікна, надаються певні вагові коефіцієнти. У аналізі фінансових даних часто як ваги використовують спадну арифметичну прогресію:

$$WMA_M = \frac{na_M + (n-1)a_{M-1} + \dots + 2a_{M-n+2} + a_{M-n+1}}{n + (n-1) + \dots + 2 + 1}$$

### 2.1.3 Модель авторегресії з ковзним середнім (ARMA - Autoregressive Moving Average)

Авторегресія з ковзним середнім (ARMA) - це модель прогнозування, в якій до даних одночасно застосовуються обидва попередні методи авторегресії (AR) і ковзного середнього (MA). Наприклад, ARMA(p, q) складається із p елементів авторегресії і q – ковзного середнього. Ця модель містить AR(p) та MA(q):

$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{i=1}^q \theta_i \varepsilon(k-i) + b_1 k + \dots + b_m k^m + \varepsilon(k);$$

Модель ARMA є механізмом, що забезпечує розуміння і прогнозування майбутніх значень обраної серії. Частина AR відповідає за регресування змінної на власних попередніх значеннях. Інша частина, тобто MA, передбачає моделювання помилки, як лінійної комбінації помилок, які відбуваються одночасно і в різні часи в минулому. Як бачимо, у цій моделі також чітко прослідковується стаціонарність, оскільки має місце тренд порядку  $m$  (ступінь  $k$ ).

#### 2.1.4 Лінійна регресія

Лінійна регресія є однією із найпростіших моделей для прогнозування нелінійних нестационарних процесів, що використовує лінійний підхід до моделювання взаємозв'язку між скалярною величиною (або залежною змінною) та однією або кількома пояснювальними змінними (або незалежними змінними).

Математично лінійна регресія записується наступним чином:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u,$$

де  $y$  – залежна змінна,  $(x_1, x_2, \dots, x_k)$  – незалежні пояснювальні змінні,  $u$  – випадкова похибка з нульовим математичним сподіванням.

Переваги:

- простота реалізації;
- використовується для прогнозування числових значень.

Недоліки:

- схильна до перенавчання;
- не може бути використаною, коли зв'язок між незалежною та залежною змінними є нелінійним.

#### 2.1.5 Метод опорних векторів

Метод опорних векторів (Support Vector Machine, SVM) також може бути використаний як метод регресії, зберігаючи при цьому всі основні властивості, що характеризують алгоритм. Регресія опорних векторів (Support Vector Regression, SVR) використовує ті ж принципи, що і SVM для

класифікації, лише з невеликими відмінностями. Однак головна ідея завжди одна: мінімізувати помилки, індивідуалізуючи гіперплощину, яка максимізує запас, враховуючи те, що частина помилки допускається (рис. 2.1).

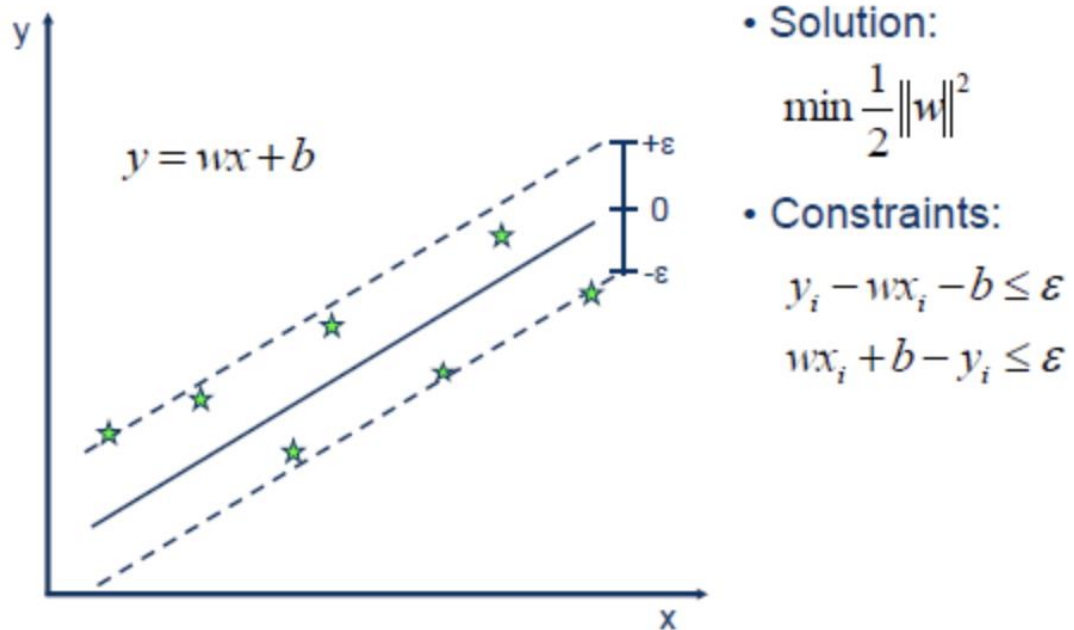


Рисунок 2.1 – Головна ідея регресора опорних векторів.

Лінійний регресор опорних векторів описується формулою:

$$y = \sum_{i=1}^N (\alpha_i - \alpha_i^*) \cdot \langle x_i, x \rangle + b$$

Нелінійний регресор використовує функцію ядра для перетворення даних у простір ознак більшої розмірності, щоб зробити можливим лінійне розподілення (рис. 2.2).

$$y = \sum_{i=1}^N (\alpha_i - \alpha_i^*) \cdot \langle \varphi(x_i), \varphi(x) \rangle + b$$

$$y = \sum_{i=1}^N (\alpha_i - \alpha_i^*) \cdot K(x_i, x) + b$$

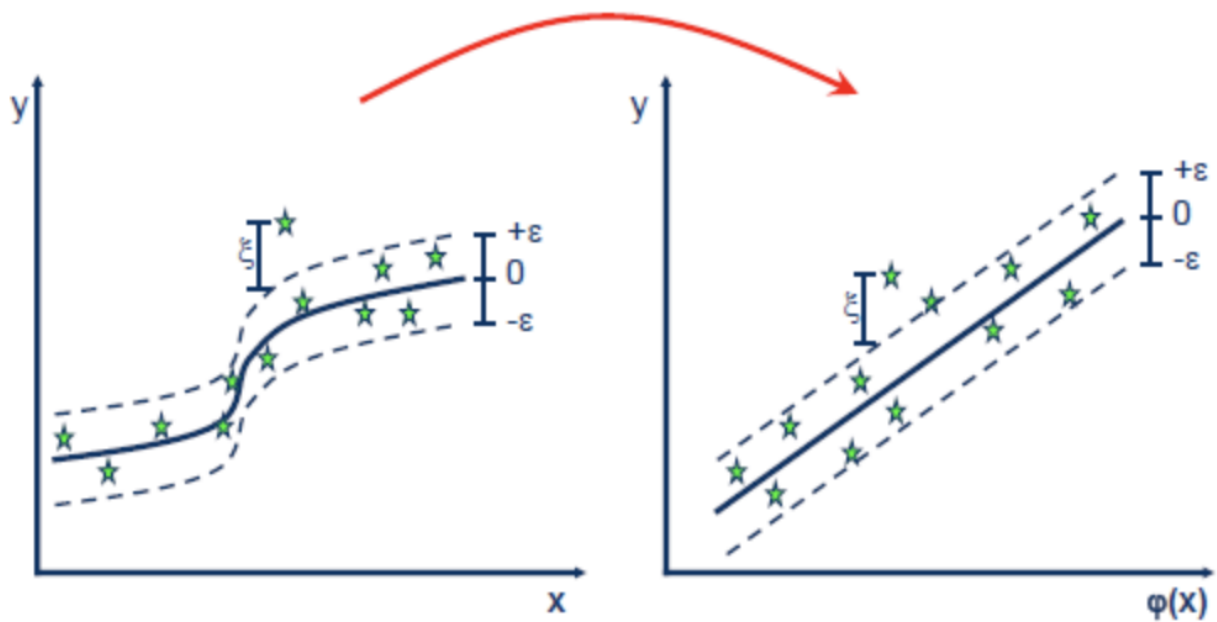


Рисунок 2.2 – Використання функції ядра робить можливим лінійне розподілення.

Можуть бути використані різні функції ядра, як, наприклад, лінійне, поліноміальне, сигмоїд та інші.

Переваги:

- ефективний у просторах великої розмірності;
- добре працює, коли кількість розмірностей перевищує кількість зразків.

Недоліки:

- не завжди добре працює на великому наборі даних;
- низька продуктивність, коли дані містять шум.

## 2.2 Рекурентні нейронні мережі

Людський мозок працює так, що людина думає послідовно, тобто не починає думати кожний наступний момент із початку. Читаючи будь-який текст, ми розуміємо зміст та значення кожного слова, відштовхуючись від

контексту. Традиційні нейронні мережі не володіють цією властивістю і це їх головний недолік. Уявимо, наприклад, що нам стоїть задача класифікувати події, що відбуваються у фільмі. Незрозуміло, чи могла б зробити це звичайна мережа, не беручи до уваги попередні кадри.

Для вирішення розглянутої проблеми були придумані рекурентні нейронні мережі (Recurrent Neural Networks, RNN), які містять зворотні зв'язки та можуть зберігати інформацію. На рисунку 2.3 показано схематична будова таких мереж. Фрагмент А отримує на вхід  $x_t$  і повертає  $h_t$ . Наявність оберненого зв'язку дає змогу передавати інформацію від одного шару мережі до іншого.

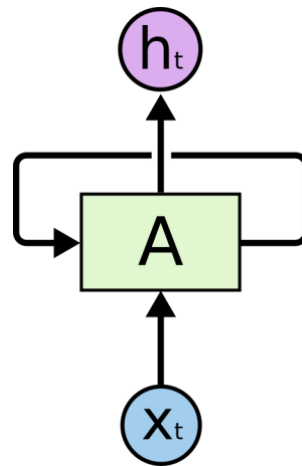


Рисунок 2.3 – Рекурентна нейронна мережа.

Для кращого розуміння рекурентну нейронну мережу можна зобразити у розгорнутому, більш звичному вигляді, що не містить циклів (рис. 2.4). У даному поданні елементи є копією однієї і тієї ж самої мережі, кожна з яких передає інформацію наступній копії.

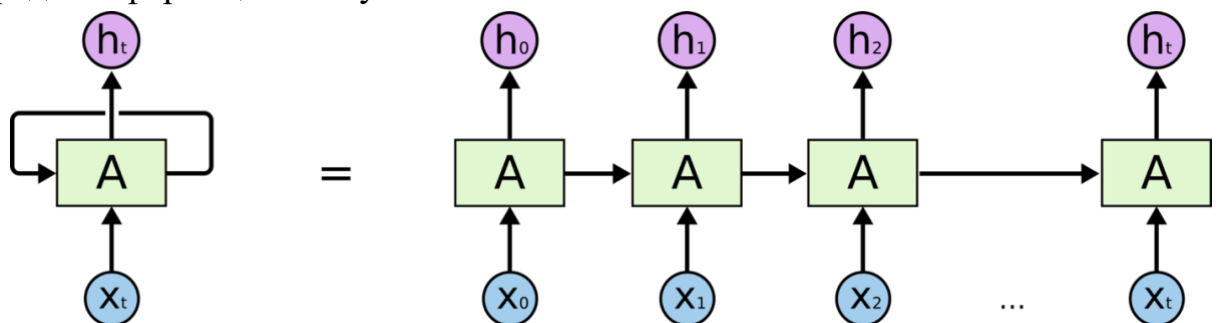


Рисунок 2.4 – Рекурентна нейронна мережа в розгорнутому вигляді.

Останні кілька років РНМ користуються великим попитом для вирішення цілого ряду задач: розпізнавання мовлення, моделювання голосу, переклад, розпізнавання зображень, прогнозування часових рядів та інші. В більшості випадків використовуються модифіковані мережі, як, наприклад, LSTM (Long short-term memory). Про них більш детально описано далі в роботі.

### 2.2.1 Недоліки рекурентних нейронних мереж

Незважаючи на всі переваги РНМ, існують кілька проблем, пов'язаних з їх використанням. Іноді для виконання задачі нам потрібно тільки недавня інформація. Розглянемо, наприклад, мережу, яка прогнозує наступне слово на основі попередніх. У випадку, коли прогнозоване слово очевидне і явно впливає із контексту, відстань між актуальною інформацією і місцем, де вона знадобилась, невелика (рис. 2.5), тому можна з легкістю навчити РНМ, використовуючи інформацію із минулого.

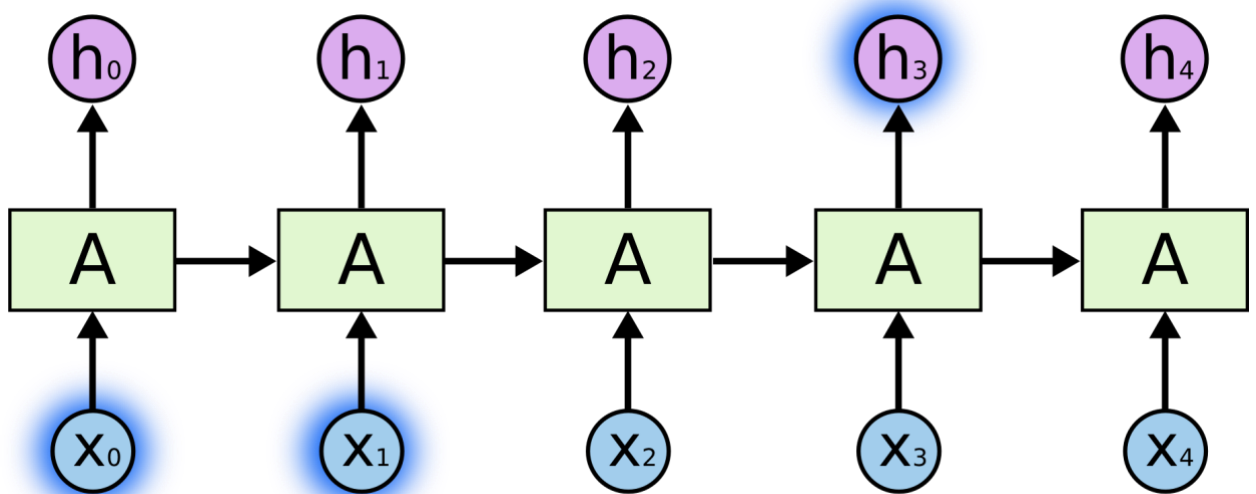


Рисунок 2.5 – РНМ з невеликою відстанню між вузлами.

Але можуть виникнути випадки, коли нам потрібно більше контексту. Допустимо, нам необхідно передбачити останнє слово у тексті «Я живу в Україні. Я впевнено розмовляю українською». Найближчі слова підказують, що останнім словом буде мова, але щоб визначити яка, нам потрібно дістатись слова «Україні» із більш далекого минулого. Таким чином, різниця між активною інформацією і точкою її застосування може стати дуже великою (рис. 2.6). В результаті, з мірою росту розмірів мережі втрачається властивість зв'язувати інформацію.

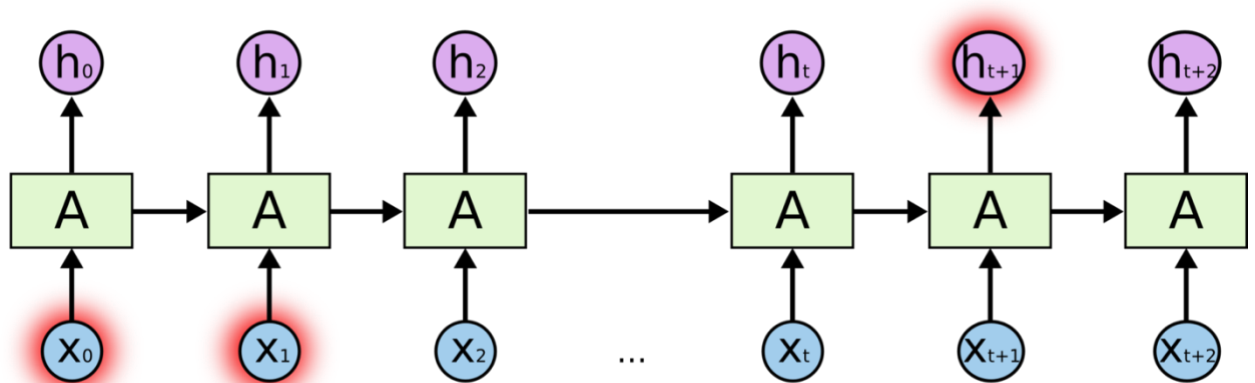


Рисунок 2.6 – РНМ з великою відстанню між вузлами.

Теоретично, такої проблеми виникати не повинно. Людина може обережно підбирати параметри мережі для розв'язування задачі такого типу. На практиці виявилось, що навчити РНМ таким параметрам неможливо. Це питання досліджували Зеп Хохрайтер (1991) та Іошуа Бенджіо (1994). Вони знайшли доведення того, що це зробити вкрай важко.



### 2.2.2 Мережа LSTM.

Long short-term memory (Довга короткострокова пам'ять), скорочено LSTM – підвид рекурентних нейронних мереж, у якому реалізовано можливість навчання довгострокових зв'язків.

Будь-яка рекурентна нейронна мережа має вигляд ланцюга модулів нейронної мережі, що повторюються. У звичайній РНМ модуль має просту будову і складається, наприклад, з одного шару із функцією активації  $\tanh$  (рис 2.7).

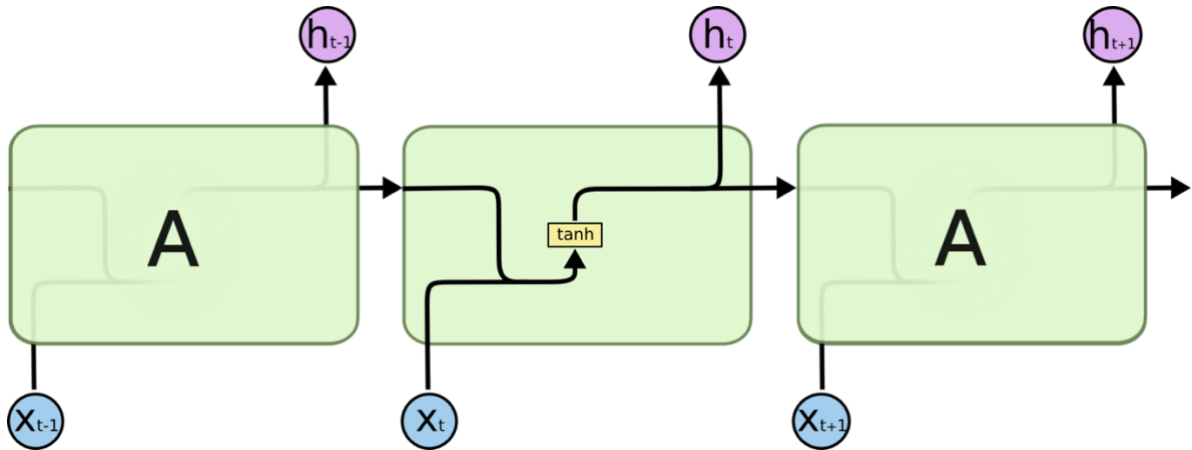


Рисунок 2.7 – Ланцюг звичайної РНМ, що складається із одного шару.

Структура LSTM також є ланцюгом, але модулі мають більш складну структуру. Замість одного шару, як у звичайній РНМ, модуль LSTM містить чотири шари, які взаємодіють між собою (рис. 2.8).

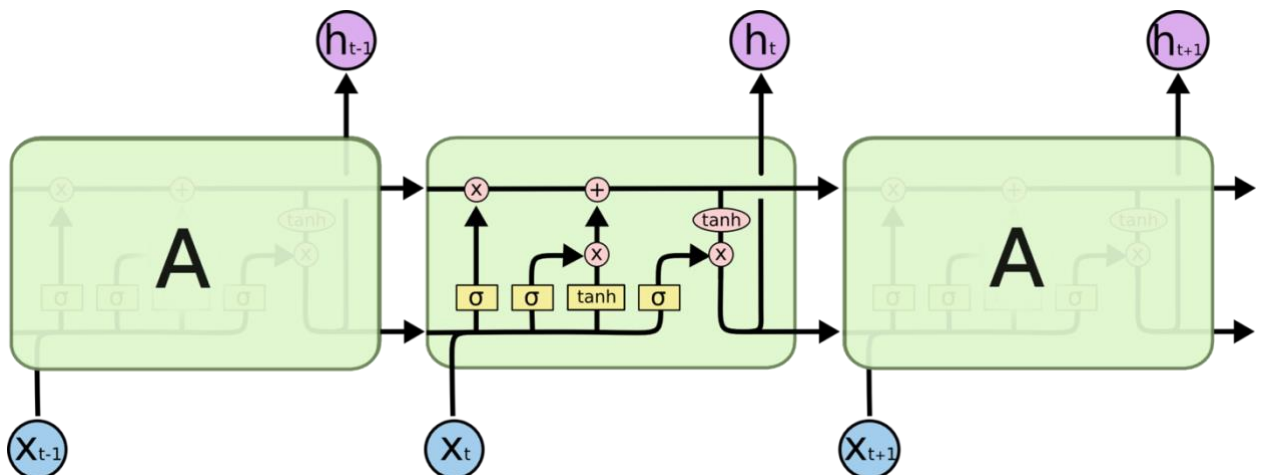


Рисунок 2.8 – Ланцюг LSTM, що складається із чотирьох шарів.

Основним компонентом LSTM є стан комірки (cell state) – горизонтальна лінія у верхній частині схеми (рис. 2.9). Вона проходить через всю комірку і бере участь лиш у деяких лінійних перетвореннях. Інформація також може не зазнавати жодних перетворень, проходячи через неї.

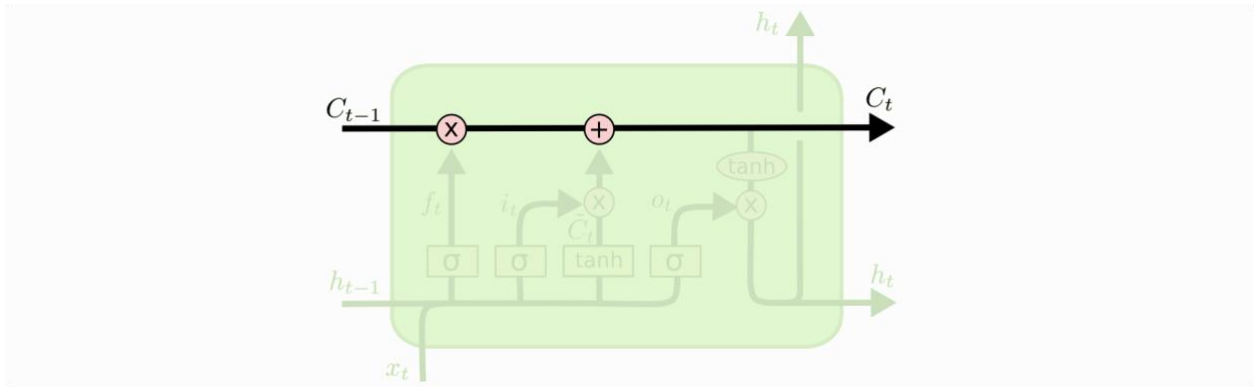


Рисунок 2.9 – Стан комірки.

Проте, LSTM може видаляти інформацію зі стану комірки, цей процес регулюється фільтрами (gates) (рис. 2.10). Фільтри дозволяють пропускати інформацію на підставі деяких умов. Вони складаються з шару сигмоїдальної нейронної мережі і операції поелементного множення.

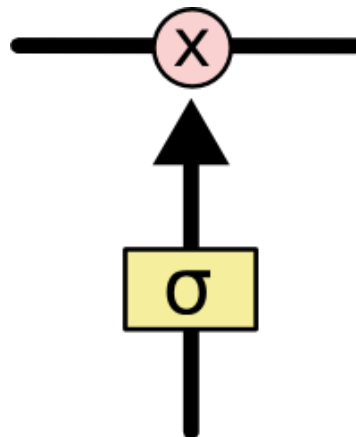


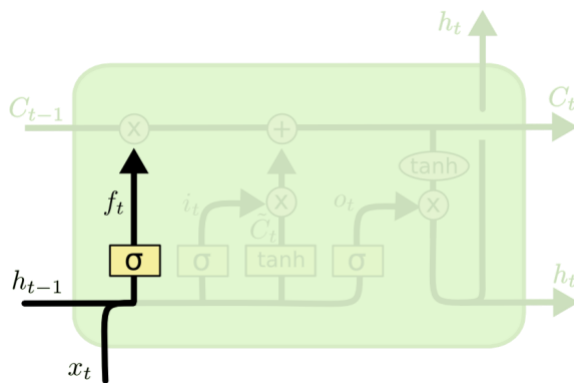
Рисунок 2.10 – Фільтр РНМ.

Сигмоїдальний шар повертає числа від нуля до одиниці, які позначають, яку частку кожного блоку інформації слід пропустити далі по мережі. Нуль в даному випадку означає "не пропускати нічого", одиниця - "пропустити все".

У LSTM існують три таких фільтра, що дозволяють захищати і контролювати стан комірки.

### 2.2.3 Покроковий розбір мережі LSTM.

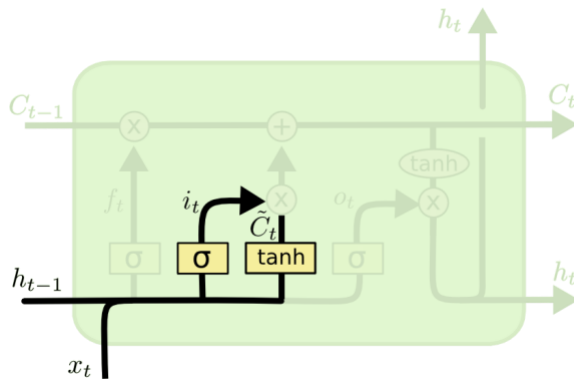
Першим кроком слід визначити, яка інформація у стані комірки є непотрібною і її можна викинути. Це рішення приймає сигмоїдальний шар, що називається «шаром фільтра забування» (англ. forget gate level). Його функція полягає в тому, щоб приймати на вхід  $h_{t-1}$  та  $x_t$  і повертати значення від 0 до 1 для кожного числа із стану комірки  $C_{t-1}$ , де 1 означає «повністю зберегти», а 0 – «повністю забути» (рис. 2.11).



$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Рисунок 2.11 – Фільтр забування.

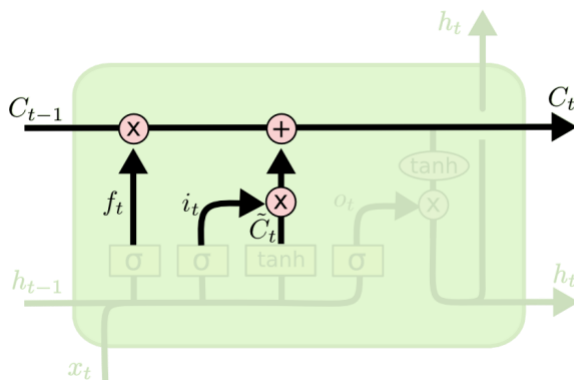
Наступним кроком слід вирішити, яка нова інформація буде зберігатись. Даний етап складається із двох частин. Спочатку елемент під назвою «шар вхідного фільтра» (англ. input layer gate) визначає які значення слід оновити. Після чого шар  $\tanh$  буде вектор значень нових кандидатів  $\tilde{C}_t$ , які додадуться в стан комірки (рис. 2.12).

Рисунок 2.12 – Шари входного фільтра та  $\tanh$ .

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

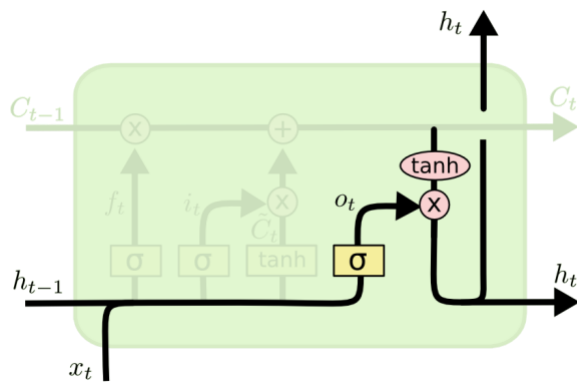
Після цього потрібно замінити старе значення стану комірки  $C_{t-1}$  на нове  $C_t$ . Для цього слід помножити старе значення на  $f_t$ , щоб забути те, що ми викинули, і додати  $i_t \cdot \tilde{C}_t$  – це кандидати на нові значення, помножені на число, що вказує, наскільки ми хочемо оновити кожне із значень стану (рис. 2.13).



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Рисунок 2.13 – Крок заміни старого значення комірки на нове.

Останнім кроком потрібно вирішити, яку інформацію передати на вихід. Вихідні дані базуються на нашому стані комірки із застосуванням деяких фільтрів. Сигмоїдальний шар вирішує, яку інформацію із стану комірки ми будемо виводити. Потім значення стану комірки переходять через шар  $\tanh$ , щоб отримати на виході значення в діапазоні від -1 до 1, і перемножуються з вихідними значеннями сигмоїдального шару, що дає змогу виводити лише потрібну інформацію (рис. 2.14).



$$o_t = \sigma (W_o [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh (C_t)$$

Рисунок 2.14 – Формування вихідного значення мережі.

### 2.3 Байєсівські мережі для моделювання нелінійних нестационарних процесів

Байєсівські мережі – перспективний ймовірнісний інструмент для моделювання складних ієрархічних процесів (статичних і динамічних) з невизначеностями довільного характеру. Мережа будується у вигляді спрямованого ациклічного графа, який відображає причинно-наслідкові зв'язки між вузлами (змінними) процесу, що досліджується.

Термін “байєсівська мережа” (Bayesian Network) був презентований американським вченим Джуді Перлом у 1985 році з метою узагальнення трьох аспектів:

- об'єктивної природи вхідних даних;
- отримання достовірної інформації при застосуванні теореми Байєса;
- ідеї застосування аналізу причин та наслідків, запропонованої в 1763 році у посмертній роботі англійського священика Томаса Байєса.

Байєсівські мережі з'явилися на стику двох наук:

- теорії ймовірностей;
- теорії графів (розділ дискретної математики) [6].

Використання байєсівських мереж як інструменту інтелектуального аналізу даних передбачає розв'язання двох математичних задач, таких як: побудова структури Байєсівської мережі та виведення ймовірнісного висновку.

Перша задача є проблемою нелінійної поліноміальної складності. Використовуючи рекурентну формулу Робінсона можна обрахувати кількість всіх можливих нециклічних моделей, які потрібно проаналізувати. Проте, як виявилось на практиці, враховуючи обмеження обчислювальних ресурсів повний перебір моделей можна здійснити лише для мережі з кількістю вузлів не більш ніж з 7.

Задача виведення ймовірнісного висновку є складною і відноситься до задач прийняття рішень. Існуючі методи побудови байєсівської мережі і формулювання висновку вимагають трудомістких обчислень, тому актуальними є пошук і розробка методів, що дозволяють зменшити обчислювальну складність при моделюванні процесів різної природи мережами Байєса.

Формально Байєсівська мережа визначається парою  $\langle G, B \rangle$ , в якій перша складова  $G$  є ациклічним графом, що відповідає змінним. Друга –  $B$  є множиною параметрів, що визначають мережу. Компонентами  $B$  є параметри  $\Theta_{x^i | \text{parent}(X^i)} = P(x^i | \text{parent}(X^i))$  для кожного можливого значення  $x^i$  із  $X^i$  і  $\text{parent}(X^i)$  із множини батьківських вершин  $X^i$  в графі  $G$ . Повна спільна ймовірність БМ визначається за формулою [9]:

$$P_B(X^1, \dots, X^N) = \prod_{i=1}^n P_B(X^i | \text{Parent}(X^i))$$

### 2.3.1 Теорема Байєса як інструмент формування ймовірнісного висновку БМ

Для обрахунку ймовірності одночасної появи двох незалежних подій  $A$  і  $B$  використовують наступну формулу:

$$p(A, B) = p(A) \cdot p(B)$$

У випадку залежності подій формула дещо змінюється, оскільки настання однієї з подій впливає на іншу:

$$p(A, B) = p(A) \cdot p(B|A) \quad (2.1)$$

Враховуючи комутативність операції, попередню формулу можна продовжити:

$$p(A, B) = p(A) \cdot p(B|A) = p(B) \cdot p(A|B)$$

З формули (2.1) випливає формула теореми Байєса для двох подій:

$$p(B|A) = \frac{p(B) \cdot p(A|B)}{p(A)}$$

У літературі зустрічається наступне визначення теореми Байєса:

Нехай випадкова подія  $A$  може відбуватися тільки разом із однією з гіпотез  $H_1, H_2, \dots, H_n$ , які утворюють повну групу подій деякого стохастичного експерименту. Нехай відомі апріорні, тобто ще до проведення експерименту, ймовірності гіпотез  $p(H_i)$ . В результаті проведеного експерименту відбулася подія  $A$ . Її поява впливає на ймовірності гіпотез. Постає питання: як

змінюються ймовірності гіпотез після того, як відбулася подія  $A$  і, отже, відбулась одна з гіпотез  $H_i$ ,  $i=1, 2, \dots, n$ ? Обчислення апостеріорних, тобто після дослідних, імовірностей гіпотез  $P(H_i/A)$  здійснюється за формулою Байєса

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{\sum_{k=1}^n P(H_k)P(A|H_k)}, \quad i = 1, 2, \dots, n.$$

Вищезгадана теорема активно використовується при формуванні ймовірнісного висновку на основі мережі Байєса.

## 2.4 Висновки до розділу 2

Даний розділ присвячений опису основних математичних моделей, що можуть бути застосовані до нелінійних нестационарних процесів, а саме: авторегресія, авторегресія з ковзним середнім, лінійна регресія та регресор опорних векторів. Більш детальніше розглянуто метод рекурентних нейронних та байєсівських мереж, алгоритми їх побудови.

Особливий інтерес викликають рекурентні нейронні мережі, які містять зворотні зв'язки та можуть зберігати інформацію. Ця властивість дає змогу прогнозувати інформацію, спираючись на контекст, що дало змогу ефективно вирішувати цілий ряд задач: розпізнавання мовлення, моделювання голосу, переклад, розпізнавання зображень, прогнозування часових рядів та інші. Розглянуто модифікацію LSTM, яка є великим кроком у розвитку рекурентних нейронних мереж.



## РОЗДІЛ 3: МЕТОДИКА МОДЕЛЮВАННЯ ГЕТЕРОСКЕДАСТИЧНИХ ПРОЦЕСІВ

### 3.1 Методика створення моделей гетероскедастичних процесів

Для початку ознайомимось із моделлю УАРУГ (узагальнена авторегресійна модель гетероскедастичного процесу). Умовна дисперсія в даному випадку матиме вигляд АРКС (авторегресії з ковзним середнім). Нехай похибки моделі обчислюються за формулою:

$$\varepsilon(k) = v(k)\sqrt{h(k)},$$

де  $v(k)$ - білий шум з дисперсією  $\sigma_v^2 = 1$ ;

$$h(k) = a_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i) + \sum_{i=1}^p \beta_i h(k-i), \quad (3.1)$$

де  $h(k)$  - умовна дисперсія.

Оскільки  $v(k)$  визначено, як білий шум, то має місце наступне співвідношення:

$$E[\varepsilon(k)] = E\{v(k)\sqrt{h(k)}\} = 0.$$

Оскільки дисперсія  $h(k)$  залежна від часу  $E_{k-1}[\varepsilon^2(k)] = h(k)$  та  $E_{k-1}[v^2(k)] = 1$ , то для її скінченності потрібно виконання умови, щоб корені характеристичного рівняння для (3.1) лежали всередині одиничного кола. Отже, можна зробити висновок: моделі УАРУГ властиво те, що умовна дисперсія збурень, які діють на систему, є процесом авторегресії з ковзним середнім.

Припустимо, що  $\{y(k)\}$  – це процес авторегресії із ковзним середнім. Тоді існують такі варіанти побудови моделі:

1. Вдалось побудувати модель АРКС, похибки якої мають природу білого шуму.
2. Не вдалось побудувати потрібну модель, тоді для аналізу дисперсії і визначення наявності гетероскедастичності слід створити модель УАРУГ, використавши автокореляційну функцію квадратів залишків.

Використавши співвідношення  $E_{k-1}[\varepsilon^2(k)] = h(k)$ , рівняння (3.1) можна записати у вигляді

$$E_{k-1}[\varepsilon^2(k)] = a_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i) + \sum_{i=1}^p \beta_i h(k-i). \quad (3.2)$$

Бачимо, що рівняння (3.2) має подібну структуру до рівняння АРКС  $(q, p)$  для  $\{\varepsilon^2(k)\}$ . Отже, методику побудови гетероскедастичних процесів можна сформулювати у вигляді наступних кроків:

1. Перед початком роботи з даними слід виконати попереднє опрацювання. Це може бути видалення порожніх значень, нормування, логарифмування або інші операції в залежності від якості та походження даних. Також слід дослідити чи змінюється дисперсія з часом, для цього слід провести тести на гетероскедастичність. Якщо в обраних даних присутній тренд, то потрібно також застосувати методи для видалення тренду. Досить багато інформації про характер даних можна отримати, візуально представивши їх.
2. Побудувати модель АР $(p)$  та АРКС $(p, q)$ . Для знаходження порядків моделі  $p$  та  $q$  використовуються АКФ і ЧАКФ. Також слід обчислити  $\{\hat{\varepsilon}^2(k)\}$  – ряд квадратів залишків та  $\hat{\sigma}_\varepsilon^2 = \frac{1}{N-1} \sum_{k=1}^N \hat{\varepsilon}^2(k)$  – вибіркова дисперсія збурень, де  $N$  – число залишків, отриманих з моделі АР або АРКС.

3. Для обчислених раніше квадратів залишків слід розрахувати та побудувати графік вибіркової автокореляції:

$$\rho(S) = \frac{\sum_{k=S+1}^N [\hat{\varepsilon}^2(k) - \hat{\sigma}_\varepsilon^2][\hat{\varepsilon}^2(k-S) - \hat{\sigma}_\varepsilon^2]}{\sum_{k=1}^N [\hat{\varepsilon}^2(k) - \hat{\sigma}_\varepsilon^2]}.$$

Існування ненульових із статистичної точки зору значень  $\rho(S)$  свідчить про наявність процесу типу АРУГ або УАРУГ.

4. Побудувати математичну модель УАРУГ (або АРУГ):

$$h(k) = a_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i) + \sum_{i=1}^p \beta_i h(k-i). \quad (3.3)$$

Процес буде гетероскедастичним, коли хоча б один коефіцієнт  $\alpha_i$  буде значимим.

5. На завершальному етапі модель слід уточнити. Після того, як з моделі (3.3) було отримано ряд  $\{\hat{\varepsilon}_1(k)\}$  потрібно згенерувати новий ряд  $\{y_1(k)\}$ , у якому  $y_1(k) = y(k) - \hat{\varepsilon}_1(k)$ . Уточнена модель АР чи АРКС будується з використанням ряду  $\{y_1(k)\}$ . Процес уточнення може повторюватись за потребою.

3.2 Множини критеріїв для аналізу якості даних, адекватності моделей та якості оцінок прогнозів

Щоб оцінити якість роботи певного методу побудови ймовірнісного висновку, можна використати такі величини, як середньоквадратична похибка, KL-відстань або квадратична відстань Хеллінджера (Hellinger) [10].

Середньоквадратична похибка  $MSE$ :

$$MSE = \frac{1}{\sum_{X^{(i)} \in X \setminus E} \text{card}(A^{(i)})} \cdot \sum_{X^{(i)} \in X \setminus E} \left( \sum_{\forall x^{(i)} \in X^{(i)}} (P(x^{(i)}|e) - \hat{P}(x^{(i)}|e))^2 \right).$$

$KL$  - відстань  $D_K$  між значенням ймовірності  $P(X^{(i)}|e)$  та оцінкою  $\hat{P}(X^{(i)}|e)$ :

$$D_K(P(X^{(i)}|e), \hat{P}(X^{(i)}|e)) = \sum_{\forall x^{(i)} \in A^{(i)}} \left( P(x^{(i)}|e) \cdot \log \left( \frac{P(x^{(i)}|e)}{\hat{P}(x^{(i)}|e)} \right) \right).$$

$KL$  - відстань  $D_K$  всієї БМ:

$$D_K(P, \hat{P}) = \frac{1}{\text{card}(X \setminus E)} \cdot \sum_{X^{(i)} \in X \setminus E} D_K(P(X^{(i)}|e), \hat{P}(X^{(i)}|e)).$$

Квадратична відстань Хеллінджера  $D_H$  між значенням ймовірності  $P(X^{(i)}|e)$  та оцінкою  $\hat{P}(X^{(i)}|e)$ :

$$D_H(P(X^{(i)}|e), \hat{P}(X^{(i)}|e)) = \sum_{\forall x^{(i)} \in A^{(i)}} \left( \sqrt{P(x^{(i)}|e)} - \sqrt{\hat{P}(x^{(i)}|e)} \right)^2.$$

Квадратична відстань Хеллінджера  $D_H$  всієї БМ:

$$D_H(P, \hat{P}) = \frac{1}{\text{card}(X \setminus E)} \cdot \sum_{X^{(i)} \in X \setminus E} D_H(P(X^{(i)}|e), \hat{P}(X^{(i)}|e)).$$

В представлених формулах  $X = \{X^{(1)}, \dots, X^{(N)}\}$  – сукупність всіх вершин графа БМ  $G$ , де кожна  $j$ -та вершина мережі ( $j = 1, \dots, N$ ) має

$A^{(j)} = \{0, 1, \dots, \alpha^{(j)} - 1\}$  ( $\alpha^{(j)} \geq 2$ ) станів, запис  $card(A^{(j)})$  означає потужність множини  $A^{(j)}$  (кількість елементів з яких вона складається),  $E \subset X, E = e$  – множина подій,  $P(X^{(i)}|e)$  – значення ймовірності вершини  $X^{(j)}$  за умови настання події  $E = e$ ,  $\hat{P}(X^{(i)}|e)$  – значення оцінки ймовірності.

Часто в літературі можна зустріти RMSE – квадратний корінь із середньоквадратичної похибки:

$$RMSE = \sqrt{MSE}.$$

MAPE – середня абсолютна похибка у відсотках:

$$MAPE = \frac{1}{N} \sum_{t=1}^N \frac{|Z(t) - \hat{Z}(t)|}{Z(t)} \cdot 100\%,$$

MAE – середня абсолютна похибка:

$$MAE = \frac{1}{N} \sum_{t=1}^N |Z(t) - \hat{Z}(t)|,$$

ME – середня похибка:

$$ME = \frac{1}{N} \sum_{t=1}^N (Z(t) - \hat{Z}(t)),$$

SD – стандартне відхилення:

$$SD = \sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{Z}(t) - ME)^2},$$

де  $Z(t)$  – фактичне значення часового ряду, а  $\hat{Z}(t)$  – прогнозоване значення.

Слід також зауважити, що точність прогнозування є величиною, прямо протилежною до помилки прогнозування. Якщо помилка прогнозування велика, то точність, відповідно, буде малою. Тому має місце наступна формула:

$$\text{Точність прогнозування у \%} = 100\% - \text{MAPE}.$$

Коли говорять про прогнозування, частіше визначають значення саме помилки прогнозу, тобто величину  $MSE$ ,  $MAPE$  або  $MAE$ . Але слід розуміти, що якщо  $MAPE = 5\%$ , то точність прогнозування 95%.

### 3.3 Тести для аналізу даних на лінійність та гетероскедастичність

Щоб побудувати модель, слід визначити характер даних. Для цього необхідно здійснити перевірку на гетероскедастичність. В літературі наведено декілька популярних тестів на гетероскедастичність [9, 11, 12], які мають багато спільного. Найбільш зручним та простим у використанні є наступний тест.

Спрощений тест на гетероскедастичність складається із чотирьох кроків:

1. Спочатку слід оцінити авторегресію:

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k)$$

Можна використовувати і більш високий порядок, наприклад, другий чи третій.

2. Побудувати ряд  $\{\varepsilon^2(k)\}$ , використавши залишками від оцінювання моделі.
3. Оцінити регресію:

$$\varepsilon^2(k) = a_0 + \alpha_1[0,4\varepsilon(k-1) + 0,3\varepsilon(k-2) + 0,2\varepsilon(k-3) + 0,1\varepsilon(k-4)].$$

4. Проаналізувати коефіцієнт  $\alpha_1$ . Якщо він є значимим, тобто відмінним від нуля, то модель, отримана для  $\varepsilon^2(k)$ , описує гетероскедастичний процес.

Зручність розглянутого тесту полягає в тому, що слід оцінити лише коефіцієнти  $\alpha_1$  і  $a_0$ , а решта коефіцієнтів (0,4; 0,3; 0,2; 0,1) є відомими. Для визначення того, чи є  $\alpha_1$  ненульовим, можна використати t-статистику (теорію перевірки статистик) для обраного коефіцієнта. Перевагами цього тесту є досить прості обчислення, що використовуються і під час аналізу лінійних моделей.

Сформулюємо постановку задачі прогнозування та оцінювання авторегресійних гетероскедастичних моделей АРКС( $p, q$ ), беручи за основу максимізацію функції правдоподібності.

На вхід подається набір значень числового ряду  $\{y(1), y(2), \dots, y(N)\}$ .

Для класу псевдолінійних та лінійних моделей АРУГ та АР( $p$ ) має вигляд:

$$y(k) = a_0 + \sum_{i=0}^p a_i y(k-i) + \varepsilon_1(k),$$

$$\varepsilon_1^2(k) = \alpha_0 + \alpha_1[\beta_1\varepsilon_1^2(k-1) + \beta_2\varepsilon_2^2(k-2) + \dots + \beta_m\varepsilon_m^2(k-m)] + \varepsilon_2(k),$$

де  $\varepsilon_1(k)$  – ряд випадкових величин із змінною дисперсією;

$\varepsilon_2^2(k)$  – ряд випадкових величин, що характеризуються наступним виразом:

$$E[\varepsilon_2(k)] = 0, E[\varepsilon_2(k)\varepsilon_2(j)] = \begin{cases} \sigma^2, & k = j, \\ 0, & k \neq j. \end{cases}$$

Необхідно знайти:

1. Оцінку вектора параметрів  $\theta_1 = [a_0 \ a_1 \ \dots \ a_p]^T$ , мінімізуючи критерій

$$J = \sum_{k=1}^N [y(k) - \hat{y}(k)]^2,$$

де  $\hat{y}(k) = \hat{a}_0 + \sum_{i=1}^p \hat{a}_i y(k-i) + \varepsilon_1(k)$ .

2. Оцінку вектора  $\theta_2 = [\alpha_0 \ \alpha_1 \ \beta_1 \ \dots \ \beta_m]^T$ , максимізуючи функцію правдоподібності

$$L(\theta_2) = \log(f_{y_1}(y_{1m}; \theta_2)) \prod_{k=2}^N (f_{y_k|y_{k-1}}(y_{km}|y_{(k-1)m}; \theta_2)).$$

3. Оцінку прогнозу дисперсії на основі оціненої моделі  $\hat{\varepsilon}^2(k+s|k) = E_k[\varepsilon^2(k)]$ ,  $s = 1, 2, \dots$ , використовуючи всі знайдені дані на момент часу  $k$ , із умови

$$J_1 = E[\hat{\varepsilon}^2(k+s), \hat{\varepsilon}^2(k+s)] \rightarrow \min.$$



### 3.4 Висновки до розділу 3

Процеси із змінною дисперсією, які ще називають гетероскедастичними, досить часто зустрічаються в повсякденному житті, у тому числі і у фінансовому сегменті. Змога досліджувати дані процеси та прогнозувати значення майбутніх величин, опираючись на минулі, допомагає у прийнятті вірних рішень та збереженні капіталу. Для найбільш точних результатів потрібно побудувати правильну математичну модель, методики побудови якої детально описані в даному розділі.

Побудувавши модель, слід також звернути увагу на її адекватність та точність прогнозу. Основні критерії та оцінки якості наведені в розділі. Серед найбільш популярних критеріїв якості можна виділити середньоквадратичну та середню абсолютну похибки. Остання з яких має відсотковий аналог, який досить зручно використовувати, коли порівнюються похибки роботи моделі на даних із суттєвою різницею між значеннями.

## РОЗДІЛ 4: ПОБУДОВА СППР ТА ВИКОНАННЯ ОБЧИСЛЮВАЛЬНИХ ЕКСПЕРИМЕНТІВ

### 4.1 Вибір інструментальної платформи для реалізації системи

Для виконання програмної частини роботи було обрано мову програмування Python. Вибір цієї мови не випадковий, оскільки Python набуває все більшої популярності, особливо якщо потрібно опрацьовувати, візуалізувати чи перетворювати дані. Існує багато готових прикладних бібліотек, які допомагають підключити різноманітні додаткові можливості, уникнути зайвих помилок та зробити код більш лаконічним. Значною перевагою мови Python також є її простота, як у використанні, так і у вивченні. Синтаксис зрозумілий одразу, мінімум зайвих різноманітних дужок, лише звичні англійські слова та табуляція. Обрана мова є кросплатформеною, що дає змогу з однаковим успіхом працювати на різних операційних системах, таких, як Windows 10, Mac OS X, Android, IOS та інші.

Для виконання поставленої у роботі задачі, були використані наступні бібліотеки:

- pandas – простота і зручність у роботі з даними, представленими у вигляді дата фреймів;
- numpy – швидкі математичні операції, робота з матрицями і масивами даних та інше;
- sklearn – різні методи інтелектуального аналізу даних, метрики та критерії оцінок якості моделей;
- keras – відкрита бібліотека для роботи з неймережами, компактність і та модульність у використанні та розширенні;
- matplotlib – побудова 2D графіків різної складності;

- seaborn – широкі можливості візуалізації даних, вбудовані теми оформлення графіків, побудова статистичних характеристик часових рядів, добре працює із структурами pandas та numpy;
- datareader – використовується для зчитування біржових даних напряму із сервісів yahoo finance;
- statsmodels – проведення статистичних тестів та дослідження статистичних даних.

## 4.2 Попередній аналіз та візуалізація даних

Як вже зазначалось у пункті 1.4, для моделювання в даній роботі було обрано процес ціноутворення акцій чотирьох великих міжнародних компаній: Apple, Google, Microsoft, Amazon. Для подальшої роботи слід обрати, яка характеристика нас найбільше цікавить і для якої ми будемо будувати моделі. Обраний датасет складається із наступних колонок: «High», «Low», «Open», «Close», «Volume», «Adj Close».

Досить цікавим для аналізу є поле «Adj Close», що показує скориговану ціну закриття. На перший погляд, здається, що у нас дві колонки для суми закриття, але насправді є вагома різниця між «Close» та «Adj Close». Перша вказує на ціну акцій, яка була зафіксована у кінці періоду, у нашому випадку, у кінці торгового дня. Однак ця величина не може бути використана для порівняння всіх цін акцій безпосередньо між собою, оскільки протягом тривалого періоду часу, безперечно, були різноманітні корпоративні дії, які впливали на ціну акцій. Двома основними з них є розбиття та дивіденди. Для порівняння цін на акції протягом життя акцій, ціни мають бути скориговані, щоб отримати значення скоригованого наближення. Це означає, що ціни коригуються назад у часі з дати, коли відбулася подія, що на них впливає.

Дослідимо ковзне середнє для обраної величини. Ковзне середнє фактично є звичайним середнім арифметичним цін на певному часовому інтервалі, який називають вікно. Таким чином, фільтруючи дрібні коливання, зможемо побачити головну тенденцію ринку. Для цього побудуємо декілька графіків із різними вікнами 10, 20 та 50 днів (рис. 4.1 – 4.4).

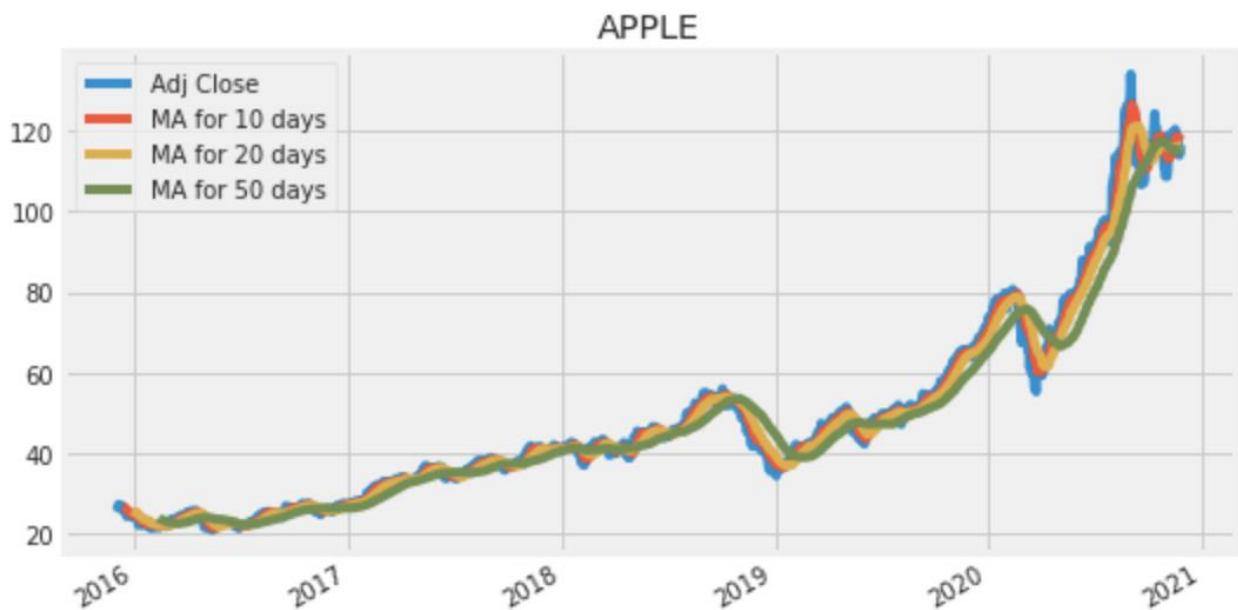


Рисунок 4.1 – Ковзне середнє з вікнами 10, 20 та 50 днів для компанії Apple.

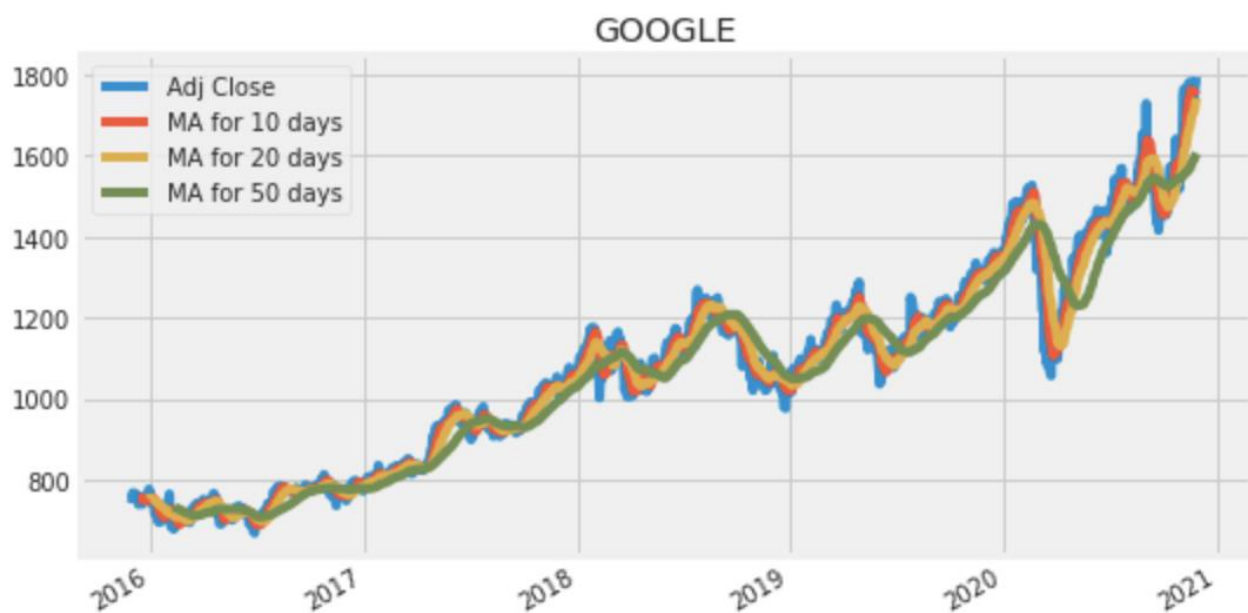


Рисунок 4.2 - Ковзне середнє з вікнами 10, 20 та 50 днів для компанії Google.

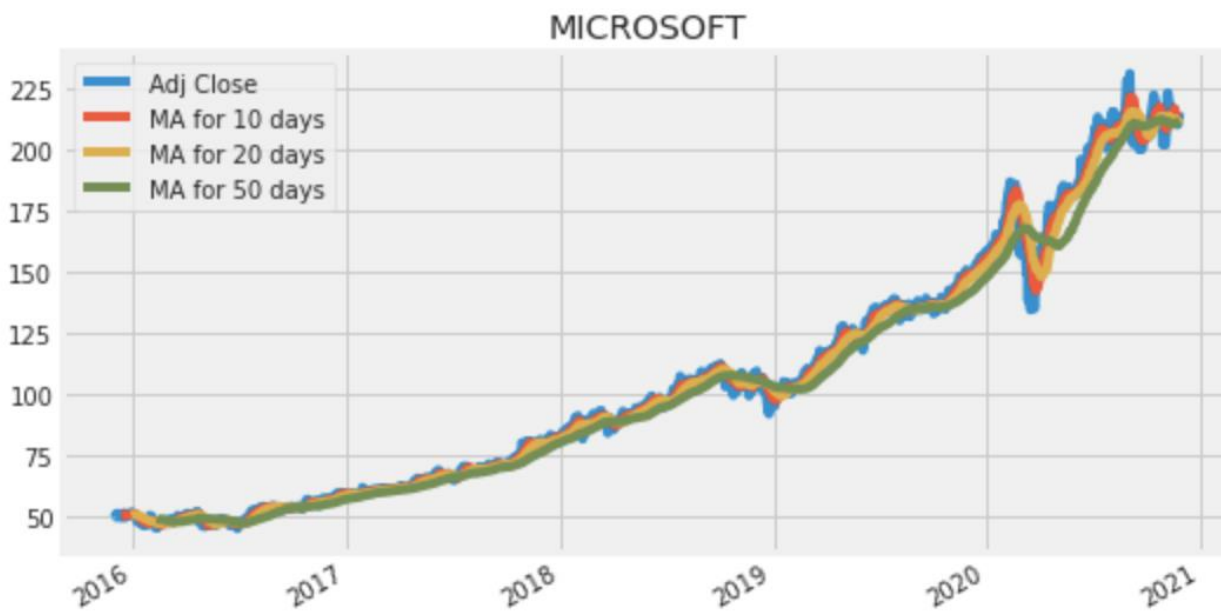


Рисунок 4.3 - Ковзне середнє з вікнами 10, 20 та 50 днів для компанії Microsoft.



Рисунок 4.4 - Ковзне середнє з вікнами 10, 20 та 50 днів для компанії Amazon.

Ковзне середнє – це базовий підхід, щоб візуально оцінити дані у згладженому вигляді. Далі здійснено більш глибокий аналіз, для цього поглянемо не на значення змінної, а на її щоденні коливання. На графіках (рис. 4.5 – 4.8) зображено денні зміни колонки «Adj Close» у відсотковому співвідношенні.

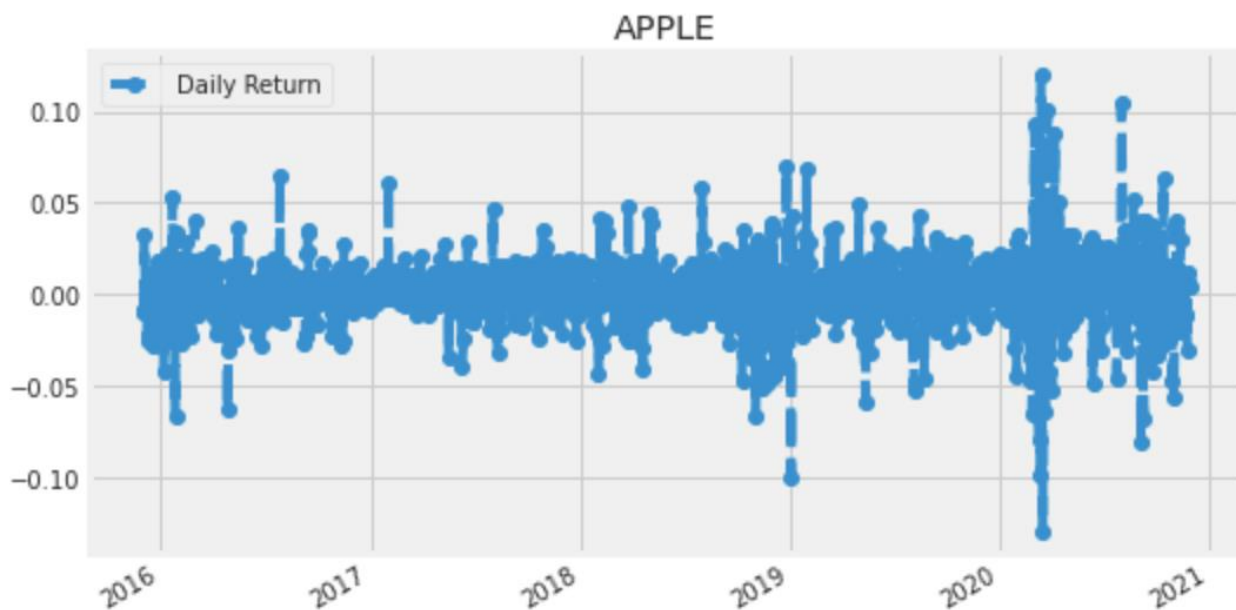


Рисунок 4.5 – Відсоткова зміна значень «Adj Close» відносно попереднього дня для компанії Apple.

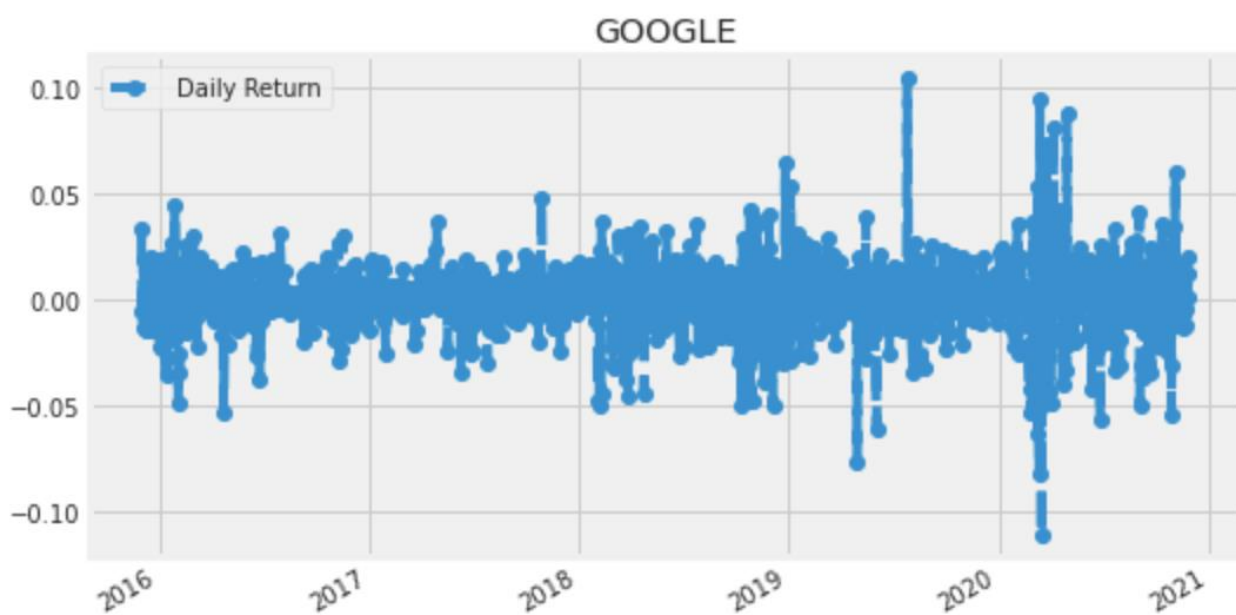


Рисунок 4.6 - Відсоткова зміна значень «Adj Close» відносно попереднього дня для компанії Google.

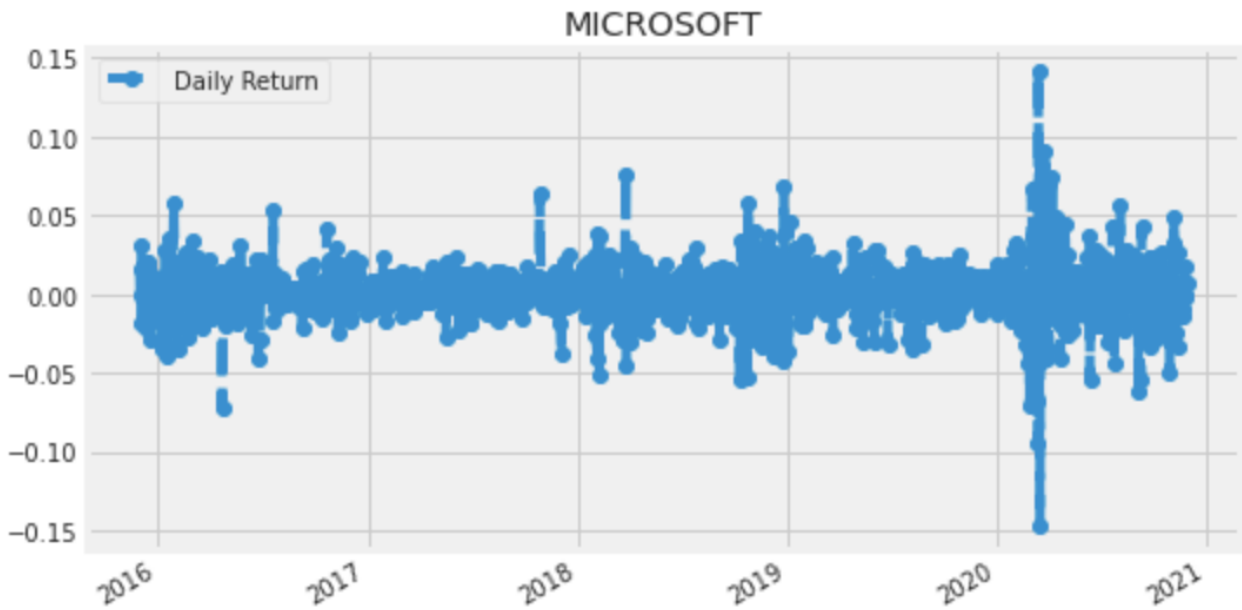


Рисунок 4.7 - Відсоткова зміна значень «Adj Close» відносно попереднього дня для компанії Microsoft.

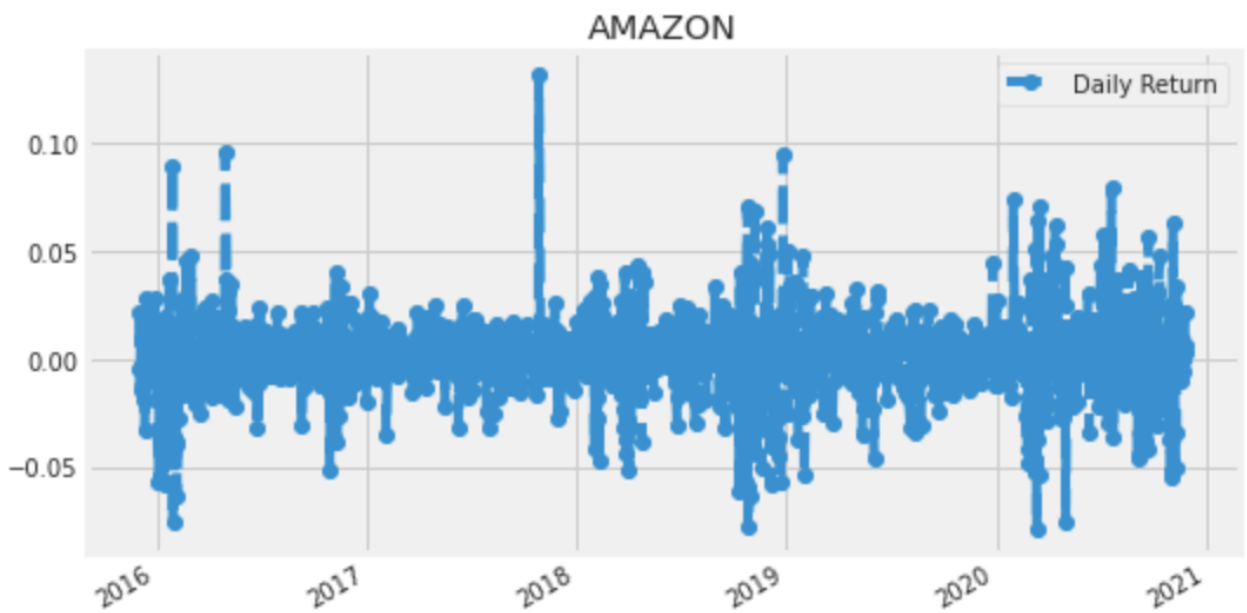


Рисунок 4.8 - Відсоткова зміна значень «Adj Close» відносно попереднього дня для компанії Amazon.

Отримані значення також можна відобразити у вигляді гісторграм (рис. 4.9).

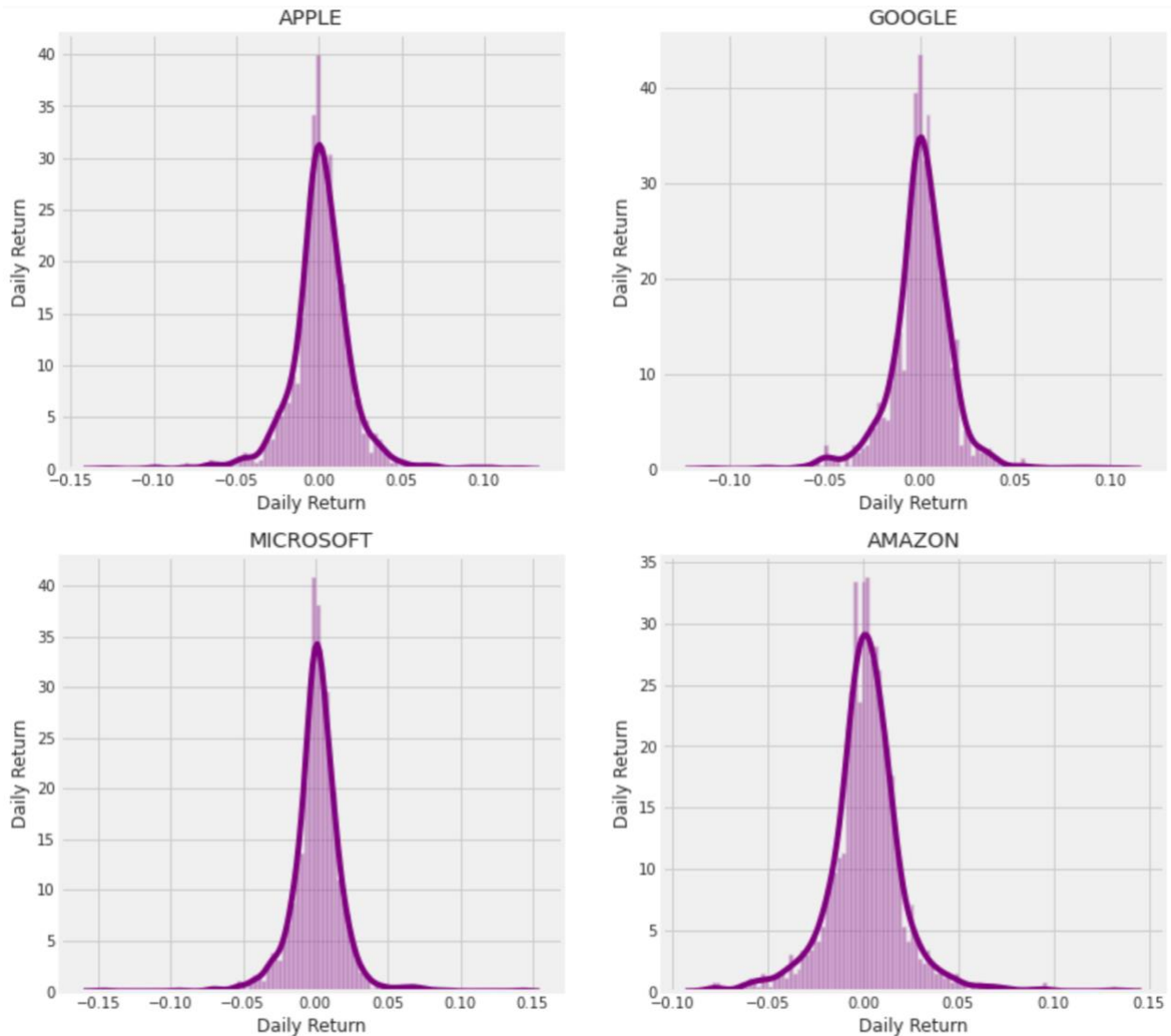


Рисунок 4.9 - Відсоткова зміна значень «Adj Close» у вигляді гістограми.

Тепер проаналізуємо значення «Adj Close» для всіх чотирьох компаній разом. Глянувши на 5 останніх значень із таблиці (рис. 4.10) бачимо, що числові значення для різних корпорацій мають різний порядок, тому порівнювати їх між собою буде неправильно. Тому перетворимо їх, використовуючи відсоткову зміну відносно попереднього значення, як робили раніше в розділі (рис. 4.11).



<b>Symbols</b>	<b>AAPL</b>	<b>GOOG</b>	<b>MSFT</b>	<b>AMZN</b>
<b>Date</b>				
<b>2020-11-20</b>	117.339996	1742.189941	210.389999	3099.399902
<b>2020-11-23</b>	113.849998	1734.859985	210.110001	3098.389893
<b>2020-11-24</b>	115.169998	1768.880005	213.860001	3118.060059
<b>2020-11-25</b>	116.029999	1771.430054	213.869995	3185.070068
<b>2020-11-27</b>	116.589996	1793.189941	215.229996	3195.340088

Рисунок 4.10 – Останні значення колонки «Adj Close» для обраних компаній.

<b>Symbols</b>	<b>AAPL</b>	<b>GOOG</b>	<b>MSFT</b>	<b>AMZN</b>
<b>Date</b>				
<b>2020-11-20</b>	-0.010958	-0.012319	-0.009557	-0.005653
<b>2020-11-23</b>	-0.029743	-0.004207	-0.001331	-0.000326
<b>2020-11-24</b>	0.011594	0.019610	0.017848	0.006349
<b>2020-11-25</b>	0.007467	0.001442	0.000047	0.021491
<b>2020-11-27</b>	0.004826	0.012284	0.006359	0.003224

Рисунок 4.11 – Останні значення денної зміни колонки «Adj Close» для обраних компаній.

Далі порівняємо значення денної зміни компаній і визначимо як вони корелюють між собою.

Спочатку переглянемо як корелюють значення компанії між собою на прикладі компанії Microsoft (рис. 4.12). Такий вид графіку називається скатерограма. У цьому випадку вона показує лінійну залежність.

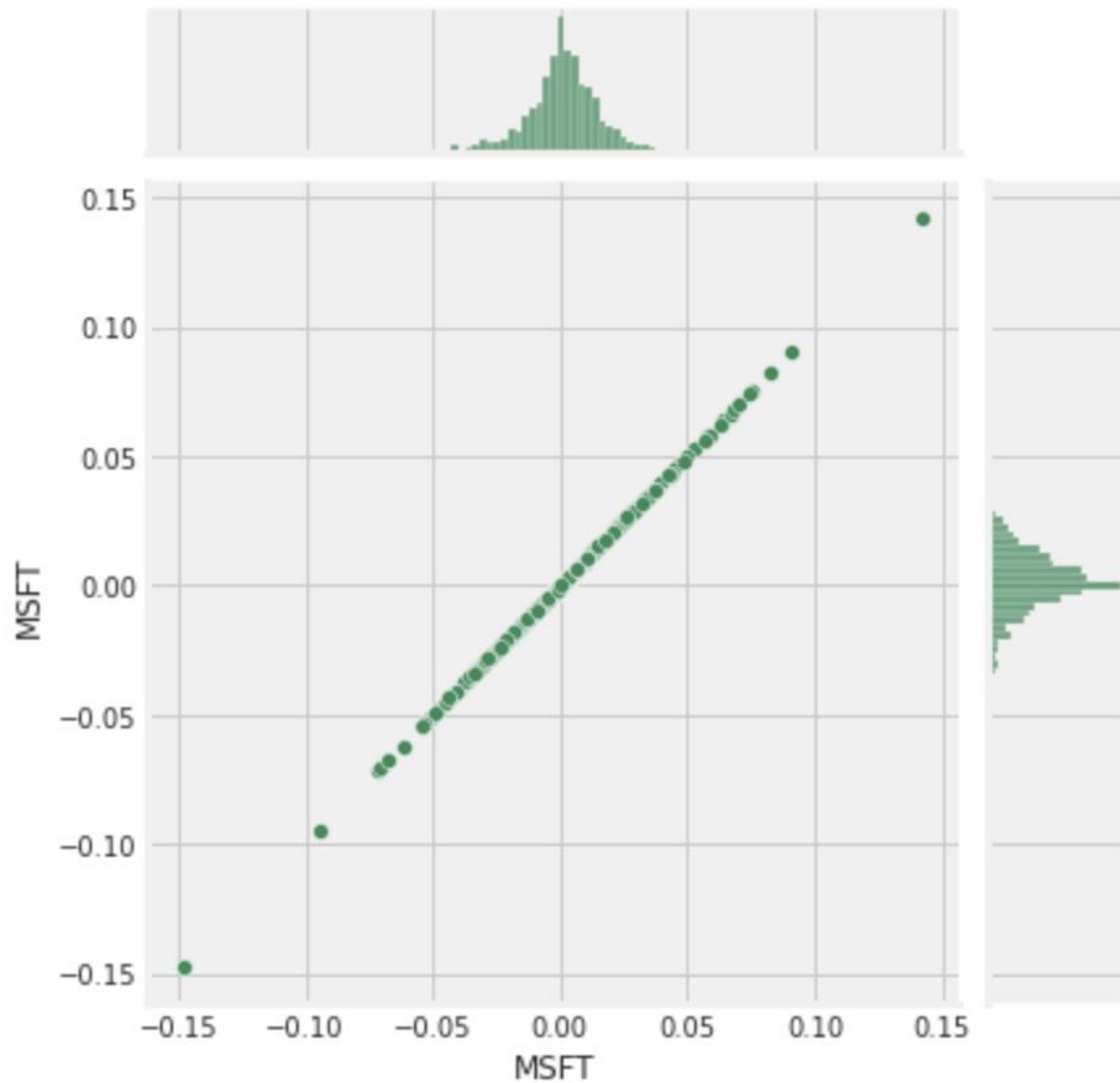


Рисунок 4.12 – Кореляція значень компанії Microsoft між собою.

Таку скатерограму можна використовувати і для порівняння денних змін різних компаній (рис. 4.13). Бачимо, що дані досить добре і позитивно корелюють між собою, на що вказує те, що більшим значенням абсциси відповідають переважно більші значення ординати.

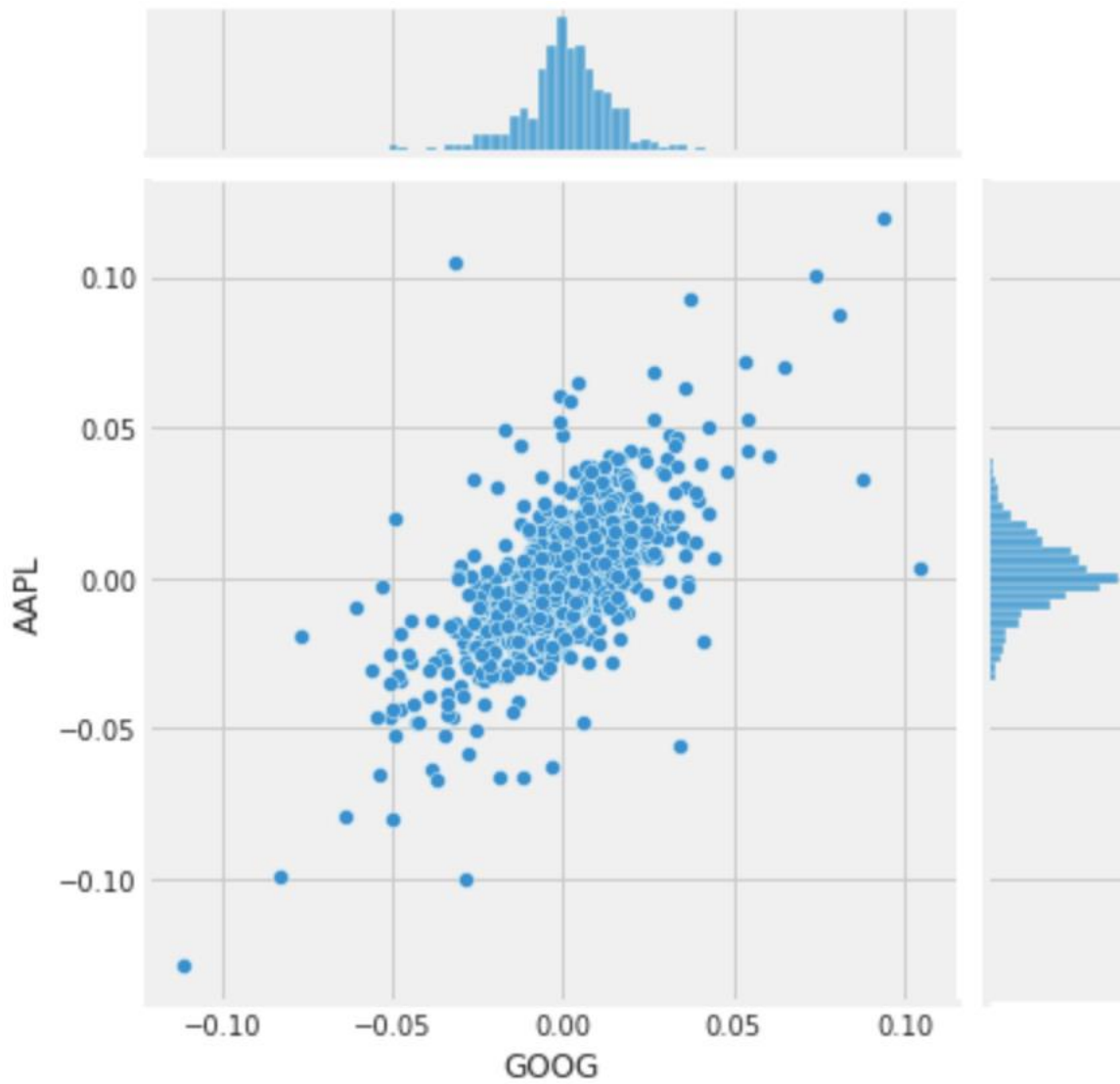


Рисунок 4.13 – Кореляція значень денних змін для компаній Apple та Google.

Мова Python та її бібліотеки Pandas та Seaborn дають можливість здійснити подібну візуалізацію для всіх можливих комбінацій елементів, у нашому випадку компаній (рис. 4.14).

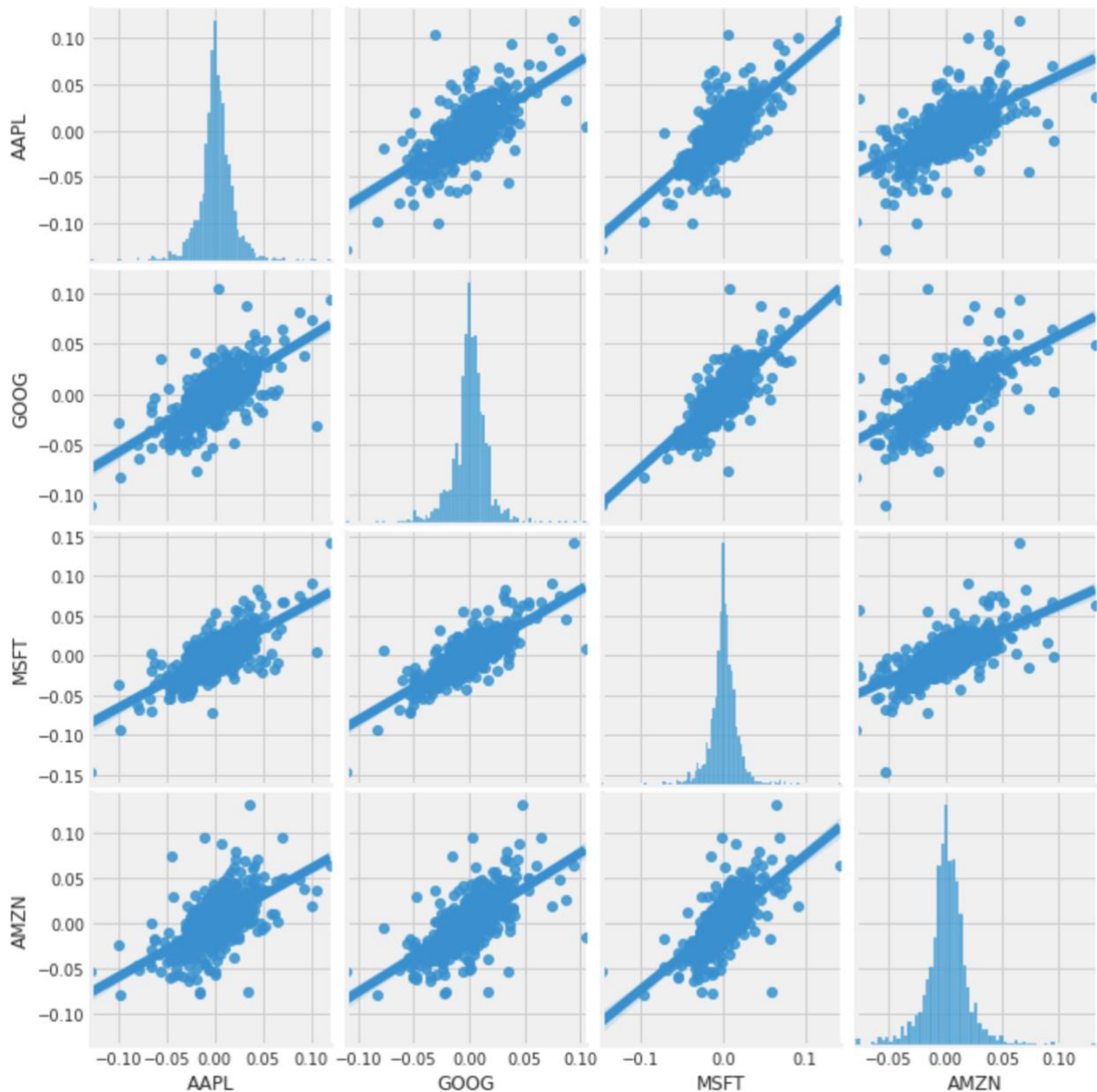


Рисунок 4.14 – Кореляція значень для всіх компаній.

На попередньому рисунку видно, як корелюють значення денної зміни «Adj Close» для всіх компаній. У більшості випадків простежується позитивна кореляція. Схожа поведінка акцій компаній може бути пов'язана із глобальними впливами, які є спільними для всіх, такі як, наприклад, зміна курсу валют, вибори президента Америки (головні офіси знаходяться в Америці), пандемія та інші.

Цікавою є залежність між Apple та Amazon, оскільки лінія, утворена на відповідному графіку, нахилена під досить гострим кутом до осі абсцис.

Крім аналізу візуального відображення корелюючих величин, зручно також дослідити числові значення (рис. 4.15, 4.16).

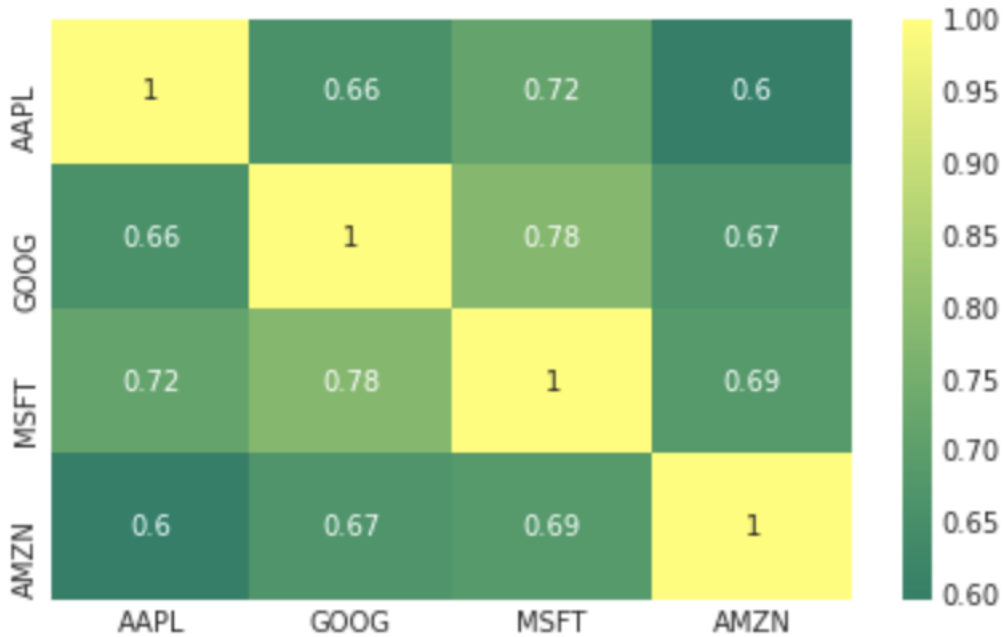


Рисунок 4.15 – Кореляційна матриця для денної зміни колонки «Adj Close».

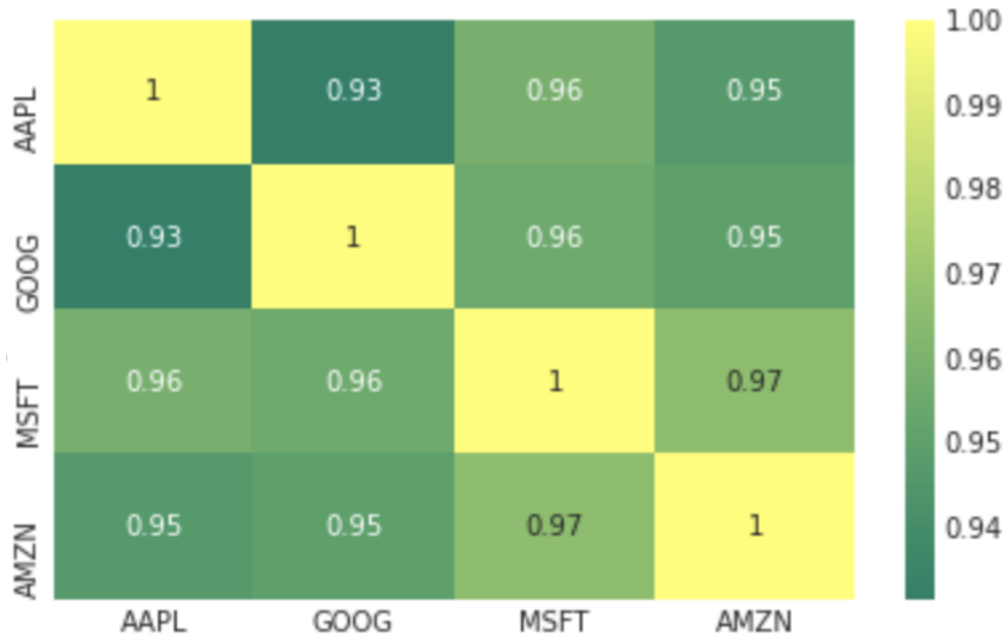


Рисунок 4.16 – Кореляційна матриця для значень колонки «Adj Close».

Як бачимо, наші припущення щодо залежності денної зміни «Adj Close» між Apple та Amazon підтвердились і, дійсно, значення кореляції між ними дорівнює 0.6, що є мінімальним значенням з таблиці, а отже, між ними

присутній найменший зв'язок. З рисунку 4.16 можна зробити висновок, що Microsoft та Amazon найсильніше коригують між собою і всі технологічні компанії мають позитивну кореляцію.

### 4.3 Використання нейронної мережі LSTM для прогнозування акцій компаній

Створимо і навчимо рекурентну нейронну мережу LSTM і перевіримо її роботу на даних закриття акцій компаній. Створена мережа складеться з 4 шарів (рис. 4.17), використовує оптимізатор Adam та середньоквадратичну помилку, як функцію втрат.

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 60, 50)	10400
lstm_1 (LSTM)	(None, 50)	20200
dense (Dense)	(None, 25)	1275
dense_1 (Dense)	(None, 1)	26

Рисунок 4.17 – Структура мережі.

Обрано діапазон даних з 1 січня 2015 року до листопада 2020 року. З них 80% тренувальна вибірка, 20% - валідаційна.

Результати роботи нейронної мережі на 4-ох різних датасетах зображено на рисунках 4.18 – 4.21. Синім кольором позначено тренувальну вибірку, червоним – валідаційну і помаранчевим – спрогнозовані значення.



Рисунок 4.18 – Результати прогнозування на даних компанії Apple.

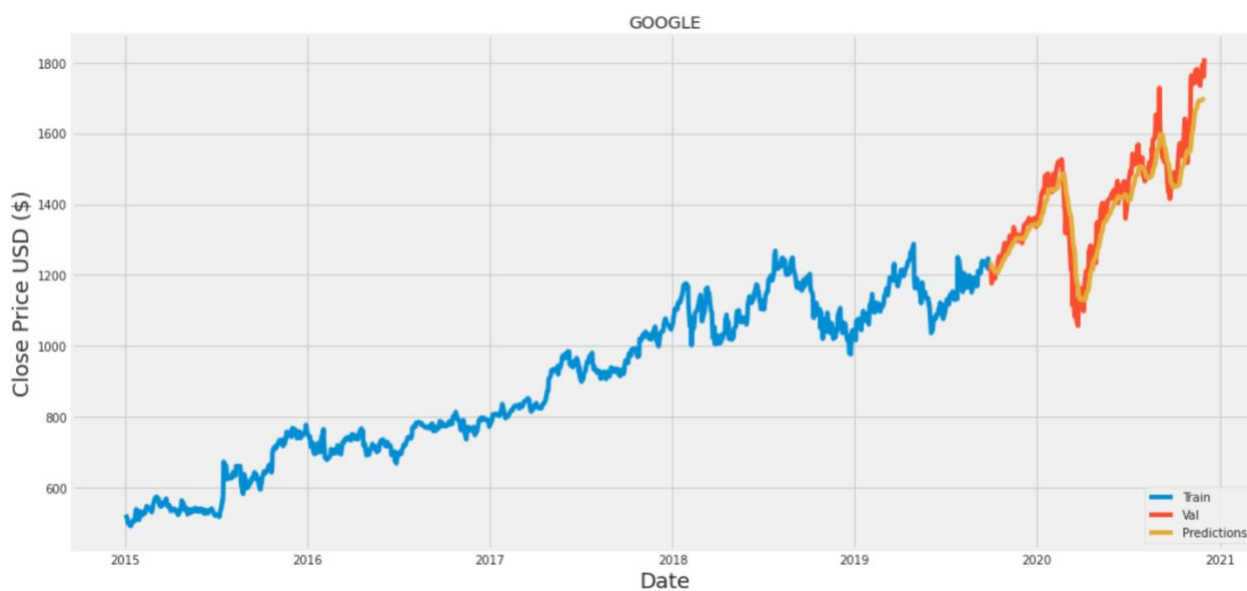


Рисунок 4.19 – Результати прогнозування на даних компанії Google.

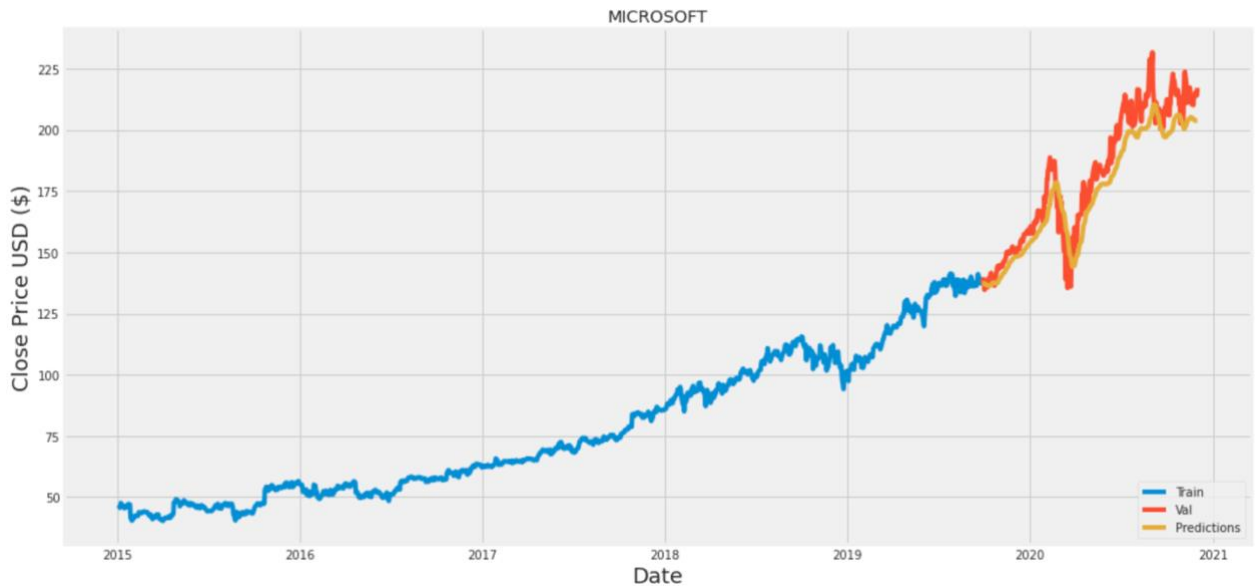


Рисунок 4.20 – Результати прогнозування на даних компанії Microsoft.

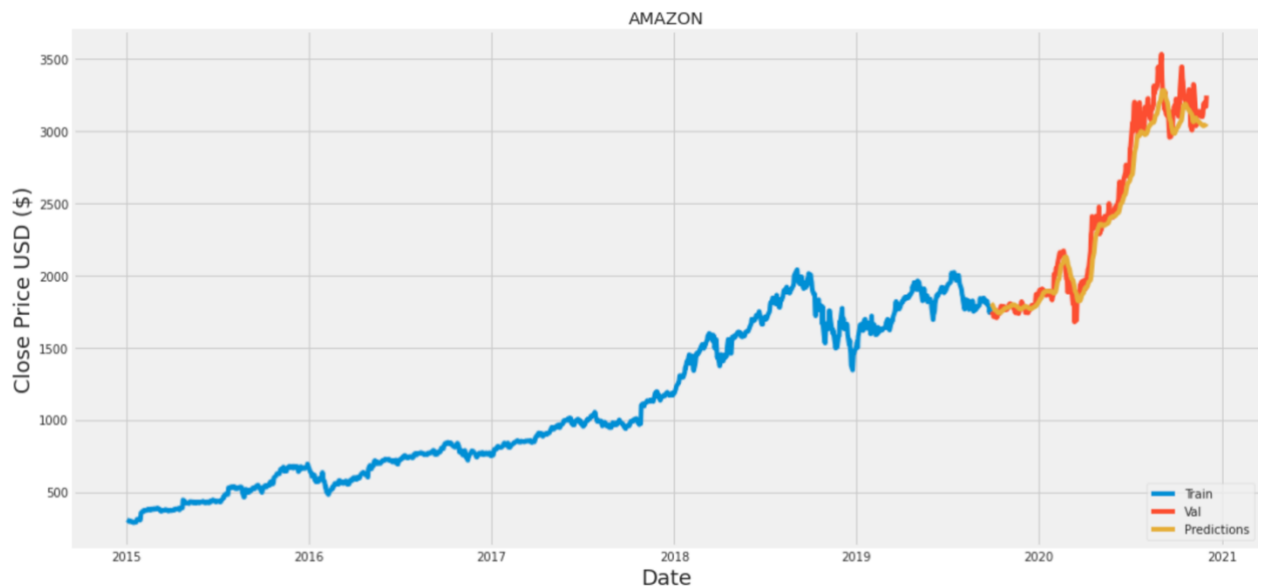


Рисунок 4.21 – Результати прогнозування на даних компанії Amazon.

Із графіків видно, що мережа досить добре справляється із задачею прогнозування даних. Значення, отримані шляхом прогнозування, є досить близькими до реальних. Але лише однієї візуальної оцінки недостатньо для формулювання висновку щодо роботи моделі та порівняння з іншими моделями, для цього слід порахувати оцінки якості моделей, такі, як RMSE, MAE, MAPE (таб. 4.1).



Таблиця 4.1 – Оцінки якості моделі.

Компанія	Оцінка			Точність прогнозу, %
	RMSE	MAE	MAPE	
Apple	3.916	2.838	3.173	96.827
Google	59.408	46.652	3.26	96.74
Microsoft	10.032	8.399	4.52	95.48
Amazon	120.989	89.236	3.449	96.551

Отримали досить гарний результат точності прогнозу, в середньому, на рівні 96%. Помітно, що отримані значення оцінок RMSE та MAE досить сильно різняться між собою. Щоб зрозуміти, у чому полягає причина такої відмінності, слід поглянути на таблицю 4.2, у якій наведені деякі статистичні характеристики обраних даних.

Таблиця 4.2 – Статистичні характеристики даних колонки Close.

Компанія	Характеристика			
	Максимум	Мінімум	Середнє	Медіана
Apple	133.948898	20.994093	48.579072	41.933216
Google	1812.310059	668.260010	1072.901074	1076.525024
Microsoft	231.045105	44.822552	104.764886	95.267551
Amazon	3531.449951	482.070007	1514.366984	1582.059998

Обчислені статистичні характеристики вказують на значну різницю у вартостях акцій компаній. Так, наприклад, максимальне значення для компанії Apple є майже у 4 рази меншим за мінімальне для Amazon і майже у 5 разів – за мінімальне для Google, а мінімум для Microsoft знаходиться приблизно на тому ж самому рівні, що і медіана для Apple.

Така різниця у вхідних даних знаходить своє відображення і в оцінках якості прогнозу, адже чим більшим є значення, тим і, відповідно, більшою є

похибка. У цьому випадку показовою є оцінка MAPE, оскільки вона відображає значення похибки, подане у відсотках.

#### 4.4 Аналіз результатів, отриманих іншими моделями, та порівняння з LSTM.

Для порівняння результатів, отриманих нейронною мережею, спробуємо використати лінійну регресію та метод опорних векторів для прогнозування цін закриття акцій компанії Apple на 30 днів. Для цього відкинемо від датасету значення за останні 30 днів, щоб потім порівняти їх із отриманим прогнозом.

Дані розділимо на тренувальну вибірку (80%) та тестувальну (20%). Навчивши моделі на тренувальній вибірці та перевіривши на тестувальній, отримали оцінки моделей, вказані у таблиці 4.3.

Таблиця 4.3 – Коефіцієнт детермінації ( $R^2$ ) моделей.

Модель	$R^2$
Linear Regression	0.93
Support Vector Machine (Regressor)	0.95

Отримані результати вказують на те, що моделі вийшли досить гарними, тому можна переходити до прогнозування. Результати прогнозування можна зображені на графіках (рис. 4.22 – 4.23).

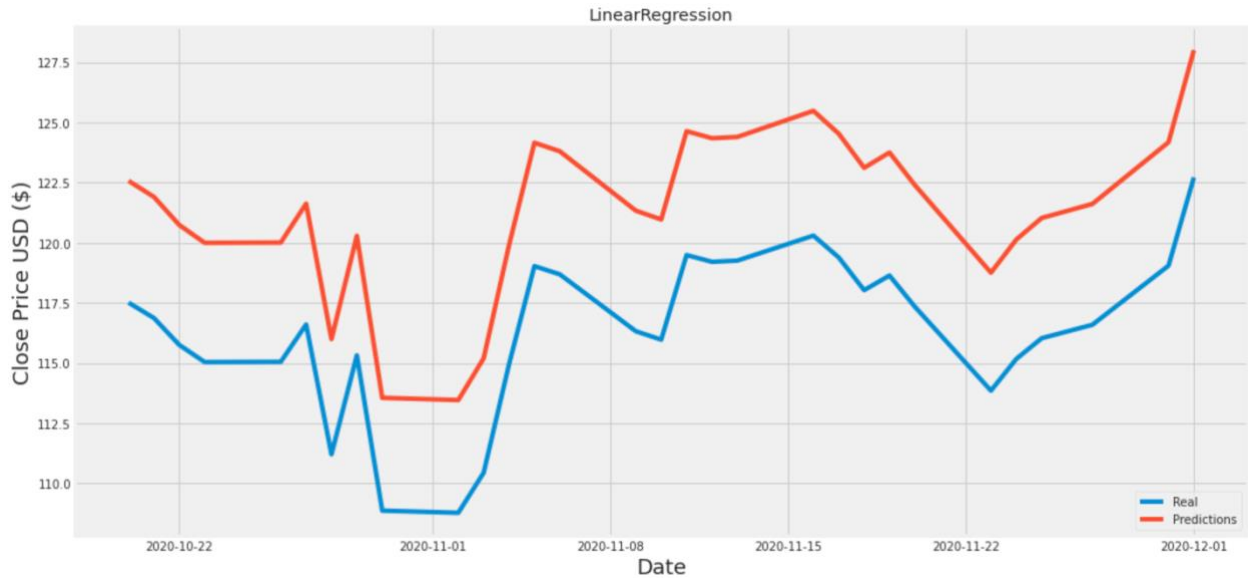


Рисунок 4.22 – Результати прогнозування на 30 днів з використанням лінійної регресії.

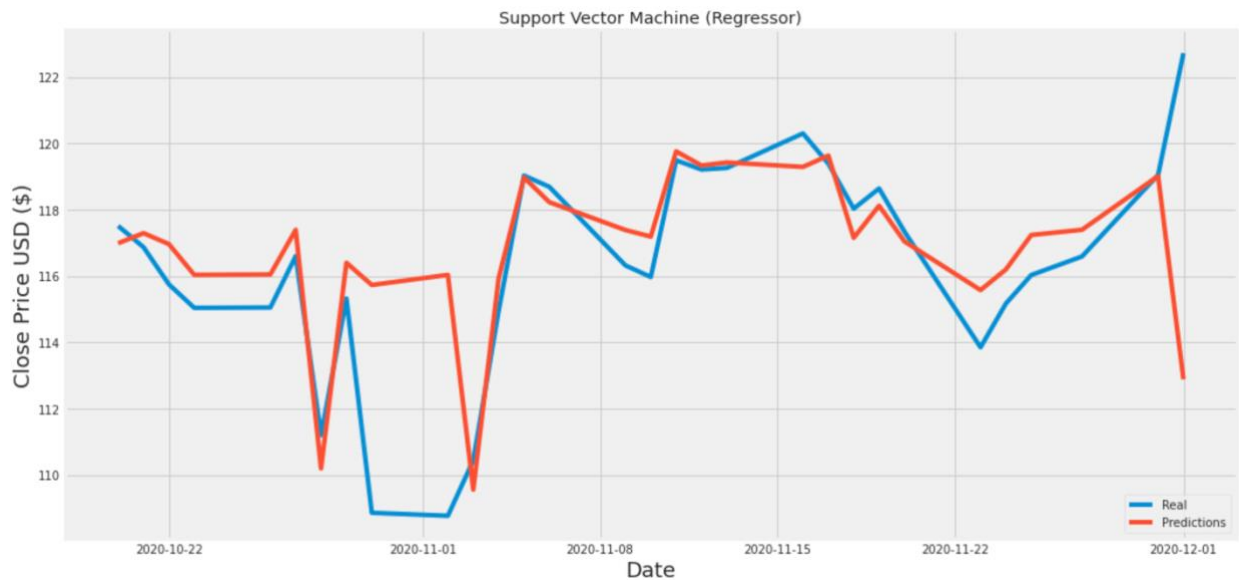


Рисунок 4.23 – Результати прогнозування на 30 днів з використанням методу опорних векторів.

Параметри оцінки якості моделей відображені у таблиці 4.4

Таблиця 4.4 – Оцінки якості моделей.

Модель	Оцінка			Точність прогнозу, %
	RMSE	MAE	MAPE	
Linear Regression	6.849	5.767	5.023	94.977
Support Vector Machine (Regressor)	4.026	3.033	2.639	97.361

Моделі досить добре впорались із поставленою задачею та показали добрий результат прогнозування. Слід зауважити, що якість роботи обраних методів, на відміну від нейронної мережі, сильно залежить від періоду на який ми прогнозуємо. Спробувавши виконати прогнозування на 100 днів, було виявлено значне погіршення якості моделей, в результаті чого  $R^2$  було на рівні 0.8, а точність прогнозу близько 70% для обох моделей.

#### 4.5 Висновки по роботі

У роботі було розглянуто результати роботи програми, написаної на мові Python, для моделювання та прогнозування нелінійних нестационарних процесів у фінансовій сфері. Виконано попередній аналіз та візуалізацію даних. Здійснено навчання рекурентної нейронної мережі LSTM та перевірено її роботу на значній за обсягом тестовій вибірці. Виконано прогнозування на 4 різних датасетах, що відповідають акціям великих технологічних компаній Apple, Google, Microsoft та Amazon. Обраховано критерії якості прогнозування мережі на кожному наборі даних. Для порівняння також було створено моделі лінійної регресії і регресії на основі методу опорних векторів та здійснено прогнозування на 30 календарних днів.

Отримані результати занесено у відповідні таблиці та здійснено порівняльний аналіз оцінок.

Щодо подальшого вдосконалення програми можна виділити кілька напрямків:

1. додавання підтримки нових моделей та оцінок;
2. розширення списку можливих задач, тобто додавання можливості виконувати класифікацію, кластерізацію та інше;
3. реалізація зручного інтерфейсу користувача з можливістю роботи як у веб додатку, так і локально на ПК чи смартфоні.

## РОЗДІЛ 5: РОЗРОБКА ВЛАСНОГО СТАРТАП ПРОЕКТУ

### 5.1 Сутність та особливості стартап проектів

Для створення стартап-проекту, перш за все, необхідно мати оригінальну ідею. Якщо ідеї ще немає, існують різні методи генерації інноваційних ідей. На основі використання морфологічного методу оберіть оптимальну ідею продукту.

Наший варіант виробу повинен виконувати функції інтуїтивно зрозумілого для кінцевого користувача інтерфейсного програмного забезпечення для моделювання та короткострокового прогнозування демографічних процесів з використанням авторегресійних методів, методу групового врахування аргументу та методу експоненційного згладжування. Усі основні параметри та проміжні рішення сформовані у таблиці 5.1.

Таблиця 5.1 – Морфологічна карта.

Основні параметри	Проміжні рішення				
	1-ше	2-ге	3-є	4-те	5-те
Сфера застосування даних	Демографічні процеси	Фінансові процеси (курси валют, значення ВВП і т.д.)	Страхові випадки та кредити (ймовірні ризики та історія клієнта)	Медична галузь	Інше
Формат вхідних даних	Excel таблиця	Csv файл	База даних	Хмарне сховище	Інше
Методи прогнозування	Нейронна мережа	Байєсівська мережа	Регресійний аналіз	Методи Інтелектуального Аналізу Даних	Інше
Тип програми	Веб додаток	Локальна версія для ПК	Мобільний додаток	Програма у форматі VR	Інше
Мова програмування	Python	Eviews	SAS	Matlab	Інше
Тип ліцензій	Безкоштовна	Комерційна	Випробувальна версія	Студентська	Інше

Також наведемо план оновлення у вигляді таблиці 5.2

Таблиця 5.2 – План оновлення.

№ з/п	Запитання	Відповідь
1	2	3
1	Частиною яких систем є продукт?	Аналітичні пакети програмних засобів
2	Які функції надсистеми може виконувати продукт? Як їх з ним пов'язати?	Прогноз макроекономічних показників у світі, динаміка демографічних процесів населення країни, кредитоспроможності клієнта банку
3	Чи можна розділити продукт на частини?	Базу даних можна зберігати локально або на загальному сервері, коли обчислювальні процеси виконуються у хмарі, а інтерфейс на девайсі користувача
4	Чи можна об'єднати (агрегувати) кілька елементів продукту в один?	Безперечно для комфортного та ефективного користування кінцевим продуктом буде краще зберігати всі елементи продукту в одному місці, для ефективного виконання обчислювальних процесів
5	Яким має бути ідеальний продукт?	Продукт має приймати на вхід файли будь-якого типу, а також враховувати специфікацію та особливості даних, зручний та інтуїтивно зрозумілий користувацький інтерфейс, низьковартісні обчислювальні ресурси
6	Що відбудеться, якщо вилучити цей продукт? Чим його можна замінити?	Готовий продукт на ринку аналітики такий, як Eviews, Deductor, ghjlerwsz SAS
7	Яким цей продукт був у минулому?	Eviews – незручний користувацький інтерфейс, відсутня можливість завантаження вхідних файлів певний типів, низька обчислювальна швидкість, не врахування особливостей нестационарних, нелінійних процесів та інші особливості
8	На розвиток яких функцій було спрямоване удосконалення продукту?	Сучасний користувацький інтерфейс, інші платформи використання, а також врахування особливостей процесів
9	Які функції залишилися «недорозвиненими»?	Додати композицію методів для більш точного прогнозування
10	Як можна зараз розвинути ці функції?	Доступне послідовне використання методів прогнозування, що веде за собою додаткові часові витрати та матеріальні



FutureViews – веб додаток для завантаження, аналізу та прогнозування нестационарних, нелінійних, динамічних процесів у різних сферах застосування. Використовуючи статистичні дані, які нам необхідно проаналізувати користувач їх підвантажує та застосовує різноманітні математичні операції для аналізу та прогнозування.

Етапи та продукти в послідовності заміщення можна розглянути в наступній таблиці 5.3

Таблиця 5.3 – Синхронізація завдань

Етапи	Продукти (послідовність заміщення)	
Назва	Збереження даних	Прогнозування на N кроків
Вчора	Локально файли	Обчислення з використанням регресійних моделей
Сьогодні	Бази даних, csv, Excel	Нейронні мережі
Завтра	Хмарне сховище	Штучний інтелект
Післязавтра	Персональне хмарне сховище із VR візуалізацією даних	Future View

## 5.2 Формування команди стартапу

Визначення важливості факторів щодо їх вкладу у створення та реалізацію стартапу наведені у таблиці 5.4

Таблиця 5.4 – Оцінка факторів реалізації стартапу

Фактор	Вага (важливість)
Ідея	6
Підготовка бізнес плану	7
Компетентність	8
Залученість і ризику	3
Обов'язки	5

Шкала оцінки (0-10)

Оцінювання особистого внеску кожного партнера у створення та реалізацію стартапу наведені в таблиці 5.5.

Таблиця 5.5 – Оцінка факторів особистого внеску

Фактор	Гуць Євгеній	Тертичний Роман
Ідея	6	6
Підготовка бізнес плану	7	9
Компетентність	9	6
Залученість і ризику	2	4
Обов'язки	8	8

Шкала оцінки (0-10)

Визначення дольової участі у стартап проєкті кожного учасника зображені в таблиці 5.6.

Таблиця 5.6 – Оцінка дольової участі

Фактор	Вага (важливість)	Гуць Євгеній	Тертичний Роман
Ідея	6	6	6
Підготовка бізнес плану	7	7	9
Компетентність	8	9	6
Залученість і ризику	3	2	4
Обов'язки	5	8	8
Разом	402	203	199
Процент	100%	50,5%	49,5%

Шкала оцінки (0-10)

### 5.3 Розробка продукту для стартапу

#### Розроблення MVP

1-й MVP: Меню прогнозування (рис. 5.1), де користувач завантажує статистичні дані у форматі .csv, за необхідністю перевіряє їх на стаціонарність за допомогою кнопки «Тест Діккі-Фуллера», за допомогою радіокнопки можна обрати модель і її порядок, та за потреби побудувати АКФ та ЧАКФ, вибравши власноруч кількість лагів. І, власне, зробити прогноз на визначену кількість років, натиснувши кнопку ОК.

Рисунок 5.1 – Меню прогнозування.

2-й MVP: Візуалізації побудови моделей для порівняння їхньої ефективності на реальних статистичних даних на часовому проміжку (рис. 5.2). Яку в подальшому можна використовувати для визначення оптимального порядку моделі авторегресійного прогнозування.

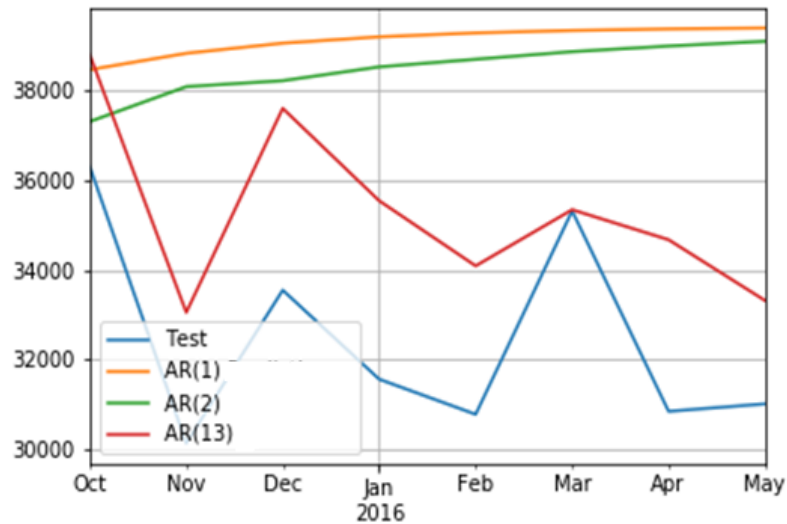


Рисунок 5.2 – Візуалізація побудованих моделей.

3-й MVP: Візуалізація методів Автокореляційної функції, а також часткової автокореляційної функції необхідної для обробки та вибору порядку моделей при натисканні однієї з кнопок програмного продукту (рис. 5.3).

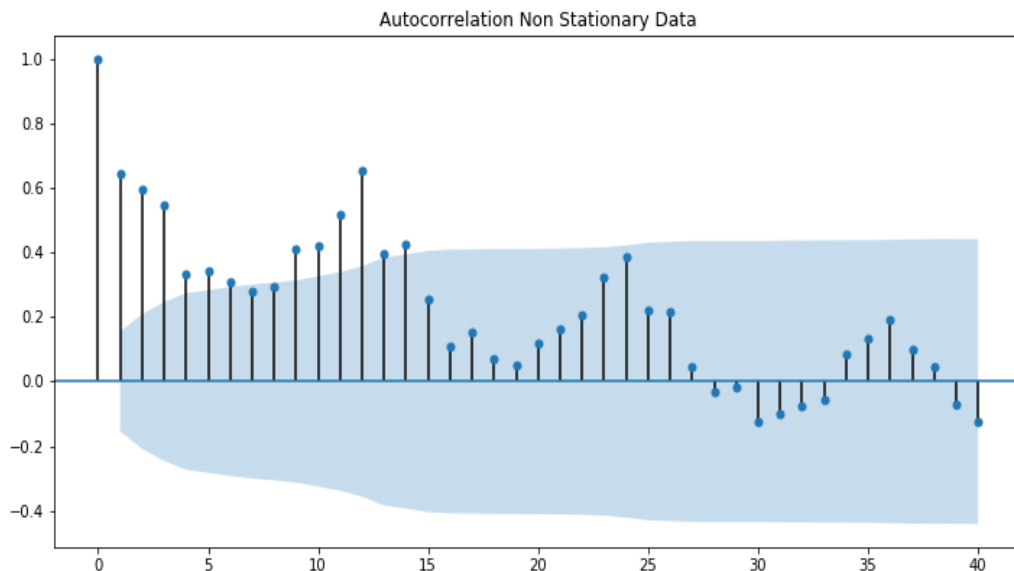


Рисунок 5.3 – Візуалізація методів Автокореляційної функції.

4-й MVP: Застосування роботи мереж Байєса для побудови моделі, а також для прогнозу нелінійних, нестационарних процесів у фінансовій сфері, а саме аналізу кредитоспроможності позичальників в банках (рис. 5.4).

Bayesian Network Diplom 2019

Please, select data file:

Correct predicted 9 form 10  
Accuracy 90.0%

Please, set bayesian network parameters:

Nodes correlation: 0.10

Result node name:

Test sample size: 10

Рисунок 5.4 – Вікно побудови мереж Байєса.

## 5.4 Розроблення бізнес-моделі стартапу

Бізнес-модель «Canvas» із всіма ключовими партнерами, видами діяльності, цілісною позицією, взаємовідносинами з клієнтами та споживчими сегментами цілісно відображено в таблиці 5.7.

Таблиця 5.7 – Бізнес-модель «Canvas»

Ключові партнери	Ключові види діяльності	Ціннісна пропозиція	Взаємовідносини з клієнтами	Споживчі сегменти
<ul style="list-style-type: none"> <li>- державні служби статистики;</li> <li>- фондові біржі (NYSE, NASDAQ, D-J, ASX);</li> <li>- Финам фінансовий портал;</li> <li>- спеціалізовані пакети статистичних даних (Yahoo Finance; Google Finance; Federal Reserve Bank of St. Louis (FOREX))</li> </ul>	<ul style="list-style-type: none"> <li>- Розробка ПЗ</li> <li>- Вирішення проблеми актуальності статистичних даних та їхня повнота;</li> <li>- Робота відділу підтримки;</li> </ul>	<ul style="list-style-type: none"> <li>- Актуальні статистичні дані та високоточний прогноз на будь-який проміжок часу;</li> <li>- Можливість безкоштовного випробувального терміну продукту;</li> <li>- Підтримка продукту протягом всього етапу його використання;</li> <li>- Зручний та інтуїтивно зрозумілий користувацький інтерфейс</li> </ul>	<ul style="list-style-type: none"> <li>- Щоденна робота відділу Support 24/7 в телефонному режимі, а також за допомогою чат-ботів</li> <li>- Створення спеціалізованих спільнот</li> <li>- Автоматизоване обслуговування</li> </ul>	<ul style="list-style-type: none"> <li>- Державні департаменти і різних сфер діяльності (фінансова, економічна, соціальна);</li> <li>- Міжнародні організації;</li> <li>- Підприємства, що потребують статистичні дані та прогноз зміни тенденцій, від яких залежить їхній бізнес</li> </ul>
	<p><b>Ключові ресурси</b></p> <ul style="list-style-type: none"> <li>- MySQL/MongoDB/Oracle</li> <li>- Android SDK</li> <li>- Microsoft</li> </ul>		<p><b>Канали збуту</b></p> <ul style="list-style-type: none"> <li>- Реклама в соціальних мережах</li> <li>- Реклама на спеціалізованих платформах для ведення таргетного бізнесу</li> <li>- Особистий зв'язок та розсилка промо з державними установами</li> </ul>	

Далі буде наведено таблицю 5.8, що містить інформацію про структуру витрат та потоки надходження доходу

Таблиця 5.8 – Структура витрат та потоки надходження доходу

<b>Структура витрат</b>	<b>Потоки надходження доходу</b>
<p>з переважною увагою до цінності</p> <ul style="list-style-type: none"> <li>- експлуатація серверів БД</li> <li>- ефект диверсифікації</li> <li>- Залучення користувачів</li> </ul>	<ul style="list-style-type: none"> <li>- Ліцензії програмного забезпечення;</li> <li>- Продаж актуальних статистичних даних;</li> <li>- Технічна підтримка та обслуговування продукту протягом всього циклу</li> </ul>

Бізнес-модель «Lean Canvas» із всіма ключовими партнерами, видами діяльності, цілісною позицією, ваємовідносинами з клієнтами та споживчими сегментами цілісно відображено в таблиці 5.9.

Таблиця 5.9 – Бізнес-модель «Lean Canvas»

<b>Проблема</b> <b>1</b>	<b>Рішення</b> <b>3</b>	<b>Унікальна торгова пропозиція</b> <b>2</b>	<b>Прихована перевага</b> <b>7</b>	<b>Споживчі</b> <b>1</b>
<p>1) Відсутність безкоштовних платформ прогнозування</p> <p>2) Відсутність підтримки протягом всього циклу співпраці</p> <p>3) Відсутність надання гарантії точності спрогнозованих процесів</p>	<ul style="list-style-type: none"> <li>- Виористання альтернативних безкоштовних платформ обслуговування процесів</li> <li>- Довгострокові договори та відділ підтримки</li> <li>- Запатентовані метрики оцінки прогнозу</li> </ul>	<ul style="list-style-type: none"> <li>- Підтримка продукту протягом всього етапу його використання</li> <li>- Наявна гарантія точності прогнозу та актуальності даних, що базується на запатентованих аналітичних оцінках</li> <li>- Зручний та інтуїтивно зрозумілий користувацький інтерфейс</li> </ul>	<ul style="list-style-type: none"> <li>- Сфокусованість на специфіку бізнесу</li> <li>- Унікальність методів вирішення проблеми</li> <li>- Регулярні бонуси за успішну співпрацю, що проявлятимуться згідно особливостей кожного партнера</li> </ul>	<ul style="list-style-type: none"> <li>- Малий та середній бізнес в різних сферах</li> <li>- Державні та міжнародні установи</li> <li>- Дослідницькі та наукові центри</li> </ul>

## Продовження таблиці 5.9

<p style="text-align: center;"><b>Ключові метрики</b></p> <p style="text-align: center;"><b>6</b></p> <ul style="list-style-type: none"> <li>- Оцінка повноти наданих статистичних даних та їх актуальність</li> <li>- Оцінка точності побудови моделі та її прогнозу, як короткострокового так і довгострокового</li> <li>- Стійкість моделі до впливу зовнішніх чинників в процесі</li> </ul>	<p style="text-align: center;"><b>Канали збуту</b></p> <p style="text-align: center;"><b>4</b></p> <ul style="list-style-type: none"> <li>Реклама в соціальних мережах</li> <li>- Реклама на спеціалізованих платформах для ведення таргетного бізнесу</li> <li>- Особистий зв'язок та розсилка промо з державними установами</li> </ul>
<p style="text-align: center;"><b>Структура витрат</b></p> <p style="text-align: center;"><b>5</b></p> <ul style="list-style-type: none"> <li>- з переважною увагою до цінності <ul style="list-style-type: none"> <li>- експлуатація серверів БД</li> <li>- ефект диверсифікації</li> <li>- Залучення користувачів</li> </ul> </li> </ul>	<p style="text-align: center;"><b>Потоки надходження доходу</b></p> <p style="text-align: center;"><b>8</b></p> <ul style="list-style-type: none"> <li>- Ліцензії програмного забезпечення;</li> <li>- Продаж актуальних статистичних даних;</li> <li>- Технічна підтримка та обслуговування продукту протягом всього циклу</li> </ul>

## 5.5 Маркетингове планування стартапу

Маркетингове планування включає в себе ряд задач: характеристика потенційних клієнтів (таб. 5.10), визначення ринкових можливостей та загроз (таб. 5.11), ступеневий аналіз конкуренції на ринку (таб. 5.12) та аналіз конкуренції в галузі за М. Портером (таб. 5.13).



Таблиця 5.10 - Характеристика потенційних клієнтів стартап-проекту.

№ п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
1	ПЗ для швидкого та точного прогнозування як короткострокового так і довгострокового	- Державні департаменти; - Міжнародні організації; - Підприємства	Сфера діяльності та відповідно сфера застосування статистичних даних	-актуальність статистичних даних; -точність та якість прогнозу;
2	Зручний та інтуїтивно зрозумілий користувацький інтерфейс	- Державні департаменти; - Міжнародні організації; - Підприємства	1) Простота використання ПЗ 2)Високоточність результату	- Акцент на зручності та простоті інтерфейсу ПЗ - Якість та точність прогнозу

Таблиця 5.11 – Визначення ринкових можливостей і загроз.

Параметри оцінки	Можливості	Загрози
1. Конкуренція	Кращі результати прогнозування за короткий проміжок часу та низьку ринкову ціну	Можлива робота у від'ємну маржинальну дохідність та з кастомними клієнтами
2. Якість та точність прогнозу	Отримання високоточного результату прогнозу	Можлива велика ресурсозатратність в обчислювальних можливостях
3. Робота з клієнтами вузької сфери діяльності	Унікальність власної розробки, яка є дефіцитною на ринку	Перепрофілювання суто на кастомного клієнта, з чого слідує відсутність універсальності та довготривалість проекту

Таблиця 5.12 – Ступеневий аналіз конкуренції на ринку.

<b>Особливості конкурентного середовища</b>	<b>В чому проявляється дана характеристика</b>	<b>Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)</b>
1. Вказати тип конкуренції - Чиста	Можливість незалежної конкуренції між компаніями та встановлення самостійної ціни, що не контролюється ринковими цінами	Можливість прояву унікальності продукту та його інновація без встановлених стандартів
2. За рівнем конкурентної боротьби - Національна	Конкуренція на міжнародному ринку	Врахування особливостей локалізації клієнта, а не диктування своїх методів
3. За галузевою ознакою - Міжгалузева	Застосування ПЗ у різних сферах діяльності	Можливість легко інтегруватись у такі сфери як економіка, екологія, медицина, політика
4. Конкуренція за видами товарів: - Товарно-родова	Конкуренція між різноманітними видами товарів, що можуть виконувати схожі функції	Більш багатофактурна сфера застосування
5. За характером конкурентних переваг - Цінова/нецінова	Конкурентна перевага і цінова і нецінова. Якість продукту та її цінова категорія	Цікаві пропозиції для заохочення менш бюджетних компаній
6. За інтенсивністю - Не марочна	Спеціальної марки програмний продукт не має	Адаптація під інші виробництва призводить іноді до зміни функцій програми

Таблиця 5.13 – Аналіз конкуренції в галузі за М. Портером.

	<b>Прямі конкуренти в галузі</b>	<b>Потенційні конкуренти</b>	<b>Постачальники</b>	<b>Клієнти</b>	<b>Товари замітники</b>
<b>Складові аналізу</b>	- HYDE; - NYSE; - Google Finance	-Статистика ООН; - Finam	HYDE LTD Corp., MADA ltd Comm	- Державні департамент; - Міжнародні організації; - Підприємства	Програмні продукти створені дослідницькими центрами
<b>Висновки:</b>	HYDE – вузькопрофільне середовище збереження даних; NYSE – не розглядає можливість прогнозування екологічних процесів Google Finance – високозатрат	Висока конкуренція на ринку, що має різноманітні шляхи розвитку	Фінансова залежність від спонсорів, що має сезонний характер та може впливати на роботоспроможність компанії	Зацікавлені в простоті використання та отриманні високоякісного результату	Дуже обмежений функціонал, що може впливати на якість результату

Обґрунтування факторів конкурентоспроможності наведено у таблиці 5.14.

Таблиця 5.14 – Обґрунтування факторів конкурентоспроможності.

№ п/п	Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)
1	Унікальність користувацького інтерфейсу	На ринку дуже мало компаній випускають зручний та зрозумілий у використанні інтерфейс, щоб персонал без додаткових туторіалів та навчальних лекцій не інтегрував інтерфейс в робочий процес
2	Математична та статистична база	Використання найновіших методик розрахунку прогнозу з урахуванням особливостей різноманітних процесів
3	Актуальність даних	Використання актуальних статистичних даних, що оновлюються якнайшвидше, залежно від характеру процесів та специфіки бажаного результату

SWOT – аналіз проекту можна переглянути у таблиці 5.15.

Таблиця 5.15 – SWOT- аналіз стартап-проекту.

<p style="text-align: center;"><b>Сильні сторони:</b></p> <ul style="list-style-type: none"> <li>- Унікальні підходи до математичного розрахунку;</li> <li>- Високоякісні технології;</li> <li>- Великий спектр застосування;</li> <li>- Інтуїтивно зрозумілий інтерфейс</li> </ul>	<p style="text-align: center;"><b>Слабкі сторони:</b></p> <ul style="list-style-type: none"> <li>- Молода компанія на ринку;</li> <li>- Відсутність налагоджених зв'язків з клієнтом;</li> <li>- Відсутність початкового фінансування;</li> </ul>
<p style="text-align: center;"><b>Можливості:</b></p> <ul style="list-style-type: none"> <li>- Заохочення малого та середнього бізнесу;</li> <li>- Робота висококваліфікованих спеціалістів;</li> <li>- Інтегрування процесів прогнозування в інші сфери діяльності;</li> <li>- Іноваційні підходи та цікаві пропозиції</li> </ul>	<p style="text-align: center;"><b>Загрози:</b></p> <ul style="list-style-type: none"> <li>- Можливість поглинання конкурентами;</li> <li>- Неактуальність на локальному ринку;</li> <li>- Втрата стартової компенсації</li> </ul>

## 5.6 Бізнес-план стартап проекту

Базова стратегія проекту наведена у таблиці 5.16.

Таблиця 5.16 – Визначення базової стратегії розвитку.

№ п/п	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку
1	Промо-розсилка в зацікавлені державні департамент та дослідні центри, а також пропозиції малому та середньому бізнесу.	Випробувальні та безкоштовні версії ПЗ на обмежений час та для кастомних клієнтів. Цікаві пропозиції в напрямку підтримки власного продукту на стороні клієнта	Максимальна рекламна компанія та оховат цільової аудиторії. Запровадження ПЗ та демонстрація роботи. Підтримка та оновлення ПЗ протягом всього життєвого циклу.

Базова стратегія конкурентної поведінки наведена у таблиці 5.17.

Таблиця 5.17 – Визначення базової стратегії конкурентної поведінки.

№ п/п	Чи є проект «першопрохідцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки*
1	Ні, але має унікальні ідеї	Орієнтована на пошук нових партнерів, шляхом зацікавлення в новому підході до ведення їхнього бізнесу	Основні характеристики будуть незмінні, оскільки вони базуються на математичних законах, а підходи до рішення будуть змінюватись	Стратегією будуть особливі пропозиції та інноваційні підходи

Ключові переваги концепції потенційного товару представлено у таблиці 5.18.

Таблиця 5.18 – Визначення ключових переваг концепції потенційного товару.

№ п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
1	Отримання актуальних даних	Високоточні статистичні дані в режимі реального часу	Оновлення даних якнайшвидше в режимі реального часу залежно від специфіки даних та поставлених бізнес-цілей
2	Високоякісний довгостроковий/короткостроковий прогноз	Побудований прогноз, який відповідає усім критерія оцінки	Використання найновіших методик розрахунку прогнозу з урахуванням особливостей різноманітних процесів та поставленого бізнес-плану користувача
3	Інтуїтивно зрозумілий інтерфейс	Простота та лаконічність зовнішньої оболонки ПЗ, яку бачить та використовує користувач	На ринку дуже мало компаній випускають зручний та зрозумілий у використанні інтерфейс, щоб персонал без додаткових туторіалів та навчальних лекцій не інтегрував інтерфейс в робочий процес

Три рівні моделі товару наведені у таблиці 5.19

Таблиця 5.19 – Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові
I. Товар за задумом	Сервіс прогнозування
II. Товар у реальному виконанні	Сервіс надання актуальної статистичної інформації та побудова на основі неї високоякісного короткострокового/довгострокового прогнозу враховуючи особливості обраних динамічних процесів; Сучасний зрозумілий інтерфейс для візуалізації та запуску усіх обраних процесів
III. Товар із підкріпленням	Можливість автономного виконання прогнозування на вказаний період часу та тригерні запуски функціоналу залежно від особливостей поведінки нестационарних нелінійних процесів у різноманітних сферах діяльності

Система збуту товару описана у таблиці 5.20.

Таблиця 5.20 – Формування системи збуту

№ п/п	Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
1	Державні організації міжнародного рівня	Виняткові права на постачання товару з відсутністю посередників та дотримання норм чинного законодавства	канал нульового рівня	Розсилка промо в держоргани відповідальним особам
2	Малий та середній бізнес	Виняткові права на постачання товару з відсутністю посередників та дотримання комерційної таємниці про співробітництво	канал нульового рівня	Розсилка промо на офіційну робочу електронну пошту

Концепція маркетингових комунікацій представлена у таблиці 5.21.

Таблиця 5.21 – Концепція маркетингових комунікацій

№ п/п	Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
1	Пошук безкоштовних або низьковартісних варіантів продукту	Спеціалізовані групи та канали зв'язку в мережі інтернет	Дешевизна	Запропонувати вигідну та цікаву пропозицію з подальшою співпрацею	Реклама, що містить демонстративний характер особливостей нашого продукту
2	Шляхом укладання взаємовигідних угод/договорів з двостороннім обміном	Державні канали зв'язку з потенційними партнерами	Надійна та довготривала співпраця	Представити основні переваги над конкурентами	Офіційне звернення у вигляді листа, що містить основні пункти співпраці

## 5.7 Патентний пошук

Результати патентного пошуку відображені у таблиці 5.22.

Таблиця 5.22 – Результати патентного пошуку для ідеї стартап-проекту

№	Країна	Шифр документа	ПБ автора ідеї	Назва документа і короткий опис ідеї
1	USA	US2016110812A1	Mun Johnathan	“Project economics analysis tool” - The present invention is in the field of corporate finance, corporate capital investments, economics, math, risk analysis, simulation, decision analysis, and business statistics, and relates to the modeling and valuation of investment decisions under uncertainty and risk within all companies, allowing these firms to properly identify, assess, quantify, value, diversify, and hedge their corporate capital investment decisions and their associated risks.
2	Canada	CN201710156129A	Electric Power Res Institute State Grid Jibei Electric Power Company Limited	“Method and device for predicting number of customer complaints” - The present invention relates to data processing technology, in particular to a method and a device for predicting the number of customer complaints.



## Продовження таблиці 5.22

3	USA	US2008002389W	Mordecai David K A	<p>“system and method for dynamic path- and state-dependent stochastic control allocation” - The present invention relates to transactions, and, more particularly, to mitigating risk and optimizing gain between multiple parties in multiple transactions in involving state-dependent and path-dependent conditions where outcomes maybe jointly interdependent between parties and across transactions.</p>
4	USA	US201414211112A	Mun Johnathan	<p>“Qualitative and quantitative modeling of enterprise risk management and risk registers” - The present invention is in the field of corporate finance, corporate capital investments, economics, math, risk analysis, simulation, decision analysis, and business statistics, and relates to the modeling and valuation of investment decisions under uncertainty and risk within all companies, allowing these firms to properly identify, assess, quantify, value, diversify, and hedge their corporate capital investment decisions and their associated risks. Specifically, the present invention looks at starting from a comprehensive qualitative risk register and moving the analysis into the realms of quantitative risk modeling, simulation, and optimization.</p>

## 5.8 Правові аспекти реалізації стартапів, інтелектуальна власність та патентування

### 1. Резюме

Найменування проекту: FutureViews;

Характеристика організації:

- ТОВ “Ф’ючер В’юз”.
- Товариство з обмеженою відповідальністю.
- кількість співробітників: 2.
- контактні дані: futureviews\_ltd@gmail.com, +380978987215.
- Тертичний Роман Віталійович, 22, Head of Department Analytics.

FutureViews - “start-up” проект, реалізований як мобільний додаток для Android та IOS, а також веб-сайт для створення власних короткострокових або ж довгострокових динамічних процесів базованих на обраних статистичних даних.

Персонал:

1) Тертичний Роман Віталійович — Data Analytic;

2) Гуць Євгеній Віталійович - Web Developer.

- На сьогоднішній день існують сервіси для обробки статистичних даних та створення короткострокових або ж довгострокових динамічних процесів, проте вони є або вузьконаправленими тільки на певної сфери застосування або ж їх використання є багатозатратним та немає підтримки сервісу на постійній основі.
- Використання сучасних технологій для обробки актуальних статистичних даних в будь-якій сфері діяльності.
- Власні ресурси відсутні.

- Вихід на міжнародний рівень після двох років роботи, вихід на перше місце серед конкурентів.
- Потреба інвестиції у серверну частину проекту, інвестиції будуть повертатися за допомогою преміум версії підписки на сервіс, а також реклами.
- Використання ліцензій для випробувальних термінів для поширення технологій в наукових центрах та лабораторіях.
- Застосування хмарних середовищ Google для зберігання та обробки статистичної інформації з подальшим впровадженням в технологію.
- Половина середнього та великого бізнесу використовують статистичні дані для планування своєї стратегії та розподілення ресурсів.
- Самостійне створення необхідних сервісів, які є аналогами Google Analytics та поступовий перехід виключно на власний сайт.

## 2. Опис підприємства

Команда складається з двох осіб розробників, та має комерційний досвід у веб-технологіях, а саме PHP, Angular JS, React JS; розробці мобільних додатків (Kotlin, Swift), і роботи з базами даних Oracle, MySQL.

## 3. Опис продукту

FutureViews – веб додаток для завантаження, аналізу та прогнозування нестационарних, нелінійних, динамічних процесів у різних сферах застосування. Використовуючи статистичні дані, які нам необхідно проаналізувати користувач їх підвантажує та застосовує різноманітні математичні операції для аналізу та прогнозування.

Розкажіть про асортимент вашої продукції, перерахуйте основні найменування ваших товарів і послуг, якщо їх занадто багато, об'єднайте в групи

Наш продукт - це комплект веб-сервіс (futuresviews.com) + мобільний додаток Android + мобільний додаток iOS.

Опишіть функціональне призначення продукції, для яких цілей вона призначена. Які потреби задовольняє, які проблеми допомагає вирішити.

Оскільки наш сайт має два глобальні підходи до використання даних та отримання бажаного результату з мінімальними затратами, користувач має свободу вибору. Перше - це зручний веб-сервіс, який поєднує в собі хмарне середовище, де можна знайти безліч корисного(найактуальніші статистичні дані структуризовані на вибір користувача відповідно його вимог з дуже зручною панеллю пошуку та голосовим асистентом), а також додати власноруч власні дані, які адаптуються до необхідної моделі. Перевага сервісу в тому, що користувачеві не потрібно зберігати всі етапи обробки та проміжні результати на власному пристрої, для цього створюється власний кабінет, де його дані будуть надійно захищені відповідно до всесвітньої технології Security Personal Data Storage. Друге – це можливість швидкого доступу до Ваших даних через мобільні додатки, що реалізовані як на Android та і на iOS. Інтуїтивно зрозумілий інтерфейс та можливість прив'язки до власного кабінету користувача спростить Вашу роботу та допоможе зберегти дорогоцінні хвилини в доступу до актуальної інформації по бізнесу.

Уявіть основні технічні характеристики вашого продукту (Не перевантажуйте читача спеціалізованими термінами).

Отже, наш веб-сервіс створений за допомогою Angular JS та в основі має одну з найпопулярніших та найнадійніших нерелятивних баз даних Oracle, основні методи та моделі високоякісного прогнозу реалізовані на Python. Та мають онлайн доступ до швидкісного оновлення даних, що прив'язані до Загальної Світової Статистичної Бази (ЗССБ).

Наведіть приклади використання продукції.

- вибір статистичних даних із категорій запропонованих веб-сервісом;
- завантаження власних статистичних даних;
- вибір адаптації статистичних даних;

- вибір параметрів для виконання прогнозу;
  - вибір типу оцінки якості моделювання та прогнозування.
- Перерахуйте основні етапи виробництва вашого продукту.
- побудова бази даних;
  - розробка веб-сервісу;
  - розробка додатку Android/ iOS;
  - розробка хмарного сховища;
  - підтримка та розробка нового функціоналу на всіх платформах.

Наскільки ваш продукт є універсальним або унікальним для кожного клієнта, в чому це проявляється.

Будь-який користувач може налаштувати інтерфейс на свій лад. Вибір платформи для роботи з програмою, яка інтегрується з усіма пристроями за допомогою власного кабінету.

Наскільки ваші товари або послуги відповідають прийнятим стандартам

Дана продукція повністю відповідає вимогам законодавства, менталітету та особливостям ведення тамтешнього бізнесу країни, де застосовується наш продукт. Торгової марки наразі немає. Дизайн сайту та додатків створений для надшвидкого та інтуїтивного використання. Продукт є унікальним не лише в Україні, а також у решті світу. Він є інтуїтивно простим у використанні, людина будь-якого сфери діяльності користуватися всіма функціями без особливої підготовки.

Товар та послуги повністю відповідають стандартам.

На якій стадії знаходиться продукт в даний час (ідея, робочий проект, дослідний зразок, серійне виробництво і т.д.)

Продукт знаходиться на стадії розробки додатку.

Вкажіть, чи є можливості для подальшого розвитку продукту

Планується перевести продукт на вищий рівень реалізації в плані покращення хмарного середовища з елементами автоматизації за допомогою Штучного інтелекту для полегшення роботи.

Проаналізуйте продукцію конкурентів, яка існує на ринку, відповідно до наступних критеріїв

На даний момент, конкурентами на нашому ринку організації, що займаються прогнозуванням вузьких сфер діяльності.

Що відзначають споживачі в своїх відгуках про ваш продукт. Якщо є письмові відгуки, або рекомендації, приведіть їх в додатку.

Сьогодні продукт має схвальні відгуки на рівні університетських установ та викладачів, які допомагали з рецензією.

#### 4. Аналіз ринку

Слід провести pest-аналіз ринку, що наведено у таблиці 5.23.

Таблиця 5.23 – Форма для проведення pest - аналізу

Групи чинників	Події/чинники	Небезпека/можливість	Вірогідність події або прояву чинника	Важливість чинника або події	Вплив на підприємство	Програма дій
Політичні	1 Державний вплив у галузь	2	1	1	1	Робота з державними сервісами
	2 Зміна законів України	1	1	2	2	Переобладнання роботи
Економічні	1 Ринок і торговельні цикли	3	2	3	3	Дослідження ринку
	2 Витрати на підтримку	1	1	2	3	Дослідження фінансової частини
Соціальні	1 Вплив ЗМІ	1	1	1	1	Робота з ЗМІ
	2 Демографічні зміни	2	2	3	3	Реклама, та робота з ЗМІ
Технологічні	1 Нові продукти	2	3	2	3	Впровадження нових технологій, цінова політика
	2 Розвиток технологій	2	3	3	2	Впровадження нових технологій

Результати SWOT – аналізу проекту наведені у таблиці 5.24.

Таблиця 5.24 – SWOT – аналіз стартап-проекту

<p><b>Сильні сторони:</b></p> <ul style="list-style-type: none"> <li>- Унікальні підходи до математичного розрахунку;</li> <li>- Високоякісні технології;</li> <li>- Великий спектр застосування;</li> <li>- Інтуїтивно зрозумілий інтерфейс</li> </ul>	<p><b>Слабкі сторони:</b></p> <ul style="list-style-type: none"> <li>- Молода компанія на ринку;</li> <li>- Відсутність налагоджених зв'язків з клієнтом;</li> <li>- Відсутність початкового фінансування;</li> </ul>
<p><b>Можливості:</b></p> <ul style="list-style-type: none"> <li>- Заохочення малого та середнього бізнесу;</li> <li>- Робота висококваліфікованих спеціалістів;</li> <li>- Інтегрування процесів прогнозування в інші сфери діяльності;</li> <li>- Іноваційні підходи та цікаві пропозиції</li> </ul>	<p><b>Загрози:</b></p> <ul style="list-style-type: none"> <li>- Можливість поглинання конкурентами;</li> <li>- Неактуальність на локальному ринку;</li> <li>- Втрата стартової компенсації</li> </ul>

Матриця оцінки загальноекономічних факторів впливу середовища на проект представлена у таблиці 5.25.

Таблиця 5.25 – Матриця оцінки загальноекономічних факторів впливу зовнішнього середовища на реалізацію стартап-проекту

Групи та окремі фактори		Оцінка залежності бізнесу від впливу зовнішніх факторів*				Сприятливі можливості	Загрози та небезпеки
		1	2	3	4		
Макроекономічні фактори	Посилення інфляції		2			Немає	Немає
	Зростання доходів населення та підвищення купівельної спроможності			3		Збільшення цін	Конкуренція
	Посилення диференціації доходів населення		2			Немає	Немає
	Зниження процентних ставок на банківський кредит		2			Немає	Немає
	Нестабільність макроекономічної ситуації			3		Немає	Немає
	Зростання економіки і розширення місткості ринку	i			4		Збільшення клієнтів і цін

## Продовження таблиці 5.25.

Елементи конкурентного середовища	Ринкові фактори			4	Збільшення цін	Конкуренція
	Вплив конкурентів			5	Впровадження нових технологій	Втрата клієнтів
	Вплив споживачів			4	Немає	Конкуренція
	Вплив постачальників	2			Немає	Немає
Соціально-демографічні характеристики	Зміни смаків споживачів, посилення рівня диференціації споживчих запитів та поява вільних ринкових сегментів, ін.		3		Отримання нових користувачів	Втрата користувачів та партнерів
Зміни в системі державного регулювання	Зміни в оподаткуванні	2			Немає	Немає
	Запровадження ліцензування та обов'язкової стандартизації продукції і послуг		3		Втрата деяких конкурентів	Втрата партнерів
	Посилення державного контролю та вимог		3		Немає	Втрата партнерів
	Посилення корупції	1			Немає	Немає
	Регулювання ціноутворення			4	Немає	Втрата користувачів
	Зміни в державних замовленнях	1			Немає	Немає
Галузеві фактори	Посилення конкуренції			5	Немає	Втрата клієнтів
	Ускладнення доступу до сировини та ресурсів			4	Немає	Втрата партнерів і користувачів
	Ускладнення доступу до споживачів			4	Немає	Втрата клієнтів
	Поява товарів-замінників			5	Немає	Втрата клієнтів
	Зміна рівня цін		3		Кількість користувачів	Конкуренція
Політичні		2			Немає	Немає
Науковотехнічні				4	Упровадження нових технологій	Конкуренція
Природні		1			Немає	Немає
Ментальні		2			Немає	Немає



## 5. План збуту та маркетингу

Планується два види отримання прибутку від створеного проекту. Перший з них це отримання коштів за куплені ліцензії різних типів, серед яких є й безкоштовні ліцензії, що покликані на заохочення до співпраці науково-дослідницьких центрів та перспективних студентів технічних вишів, а також безпосередньо категорія Преміум ліцензії, які і становитимуть вагому нішу нашого доходу. І другим видом доходу, буде монетизації кількості скачувань додатків на пристрої через Google Play та AppStore.

Вартість продукції буде розраховано з урахуванням витрат на підтримку серверної частини проекту, зарплати за розробку, оплати доставки та реклами. Види стратегій можна переглянути у таблиці 5.26.

Таблиця 5.26 – Види стратегій залежно від ціни та якості продукції

<b>Якість</b>	<b>Ціна</b>		
	<b>Висока</b>	<b>Середня</b>	<b>Низька</b>
<b>Висока</b>	Висока ціна та якість товару	Зацікавлення покупця завдяки високій якості товару і середній ціні	Завоювання ринку, збільшення частку ринку
<b>Середня</b>	Мінімізуємо витрати на етапі впровадження	Встановлення середніх цін на товар середньої якості	Завоювання ринку, збільшення частку ринку
<b>Низька</b>	Несе загрозу втратити в майбутньому покупців	Несе загрозу втратити в майбутньому покупців	Встановлення низької ціни на товари низької якості

Наразі є план продавати понад 1000 преміум версій додатка у місяць, який наведено у таблиці 5.27.

Таблиця 5.27 – План-продажів

Найменування показників	Періоди (по місяцях)												Всього за рік
	1	2	3	4	5	6	7	8	9	10	11	12	
Преміум версія додатку													
* обсяг ліцензій	10	20	30	40	50	70	90	100	120	150	150	170	1000
* ціна	\$ 500 -6,83	\$ 1000	\$ 1500	\$ 2000 1,11	\$ 2500 1,32	\$ 3500 1,43	\$ 4500	\$ 5000 1,58	\$ 6000 1,59	\$ 7500	\$ 7500	\$ 8500	\$50000
* виручка	\$ 400 -6,83	\$ 800	\$ 1200	\$ 1600 1,11	\$ 2000 1,32	\$ 2800 1,43	\$ 3600	\$ 4000 1,58	\$ 4800 1,59	\$ 6000	\$ 6000	\$ 6800	\$ 40000

Наразі цільовою аудиторією є малий та середній бізнес, проте в перспективі є заохочення великого бізнесу в різних галузях.

Для початку потрібно \$8000 для створення та підтримку сайту, мобільних додатків, юридичне оформлення.

Буде організована таргетна реклама в Інтернеті.

#### 6. Виробничий план

- Купівля ліцензії на користування бази даних Oracle
- побудова інтерфейсу додатків та веб-сервісу на Angular JS;
- Масове
- Додаткові функції для покращення роботи з користувачем за допомогою Штучного інтелекту

## 7. Організаційний план

Профіль посади представлено у таблиці 5.28.

Таблиця 5.28 – Профіль посади

Критерій	Зміст
Основна освіта	Вища освіта
Додаткова освіта, спеціалізація	Комп'ютерна спеціалізація
Необхідний досвід роботи	Щонайменше 2 рік
Завдання	Підтримка веб сайту та додатків
Знання	Oracle, Python, PHP, C#, Java Script, Android dev, Swift dev
Навички, вміння, ділові якості	Робота у команді
Особистісні якості	Вміння працювати з різними ПО
Мотивація (що можемо запропонувати)	Робота з дому

Стимулами будуть:

матеріальні: заробітна плата; постійний оклад, премії за досягнуті показники ефективності;

нематеріальні: можливість роботи, похвала, позиція в колективі та компанії, авторитет.

## 8. Фінансово-економічний план

На разі залучення інвесторів та власні кошти є орієнтованими джерелами фінансування нашого проекту. План інвестицій відображено у таблиці 5.29.

Таблиця 5.29 – Планування загальних інвестицій по проекту

Група та вид інвестицій	Сума
Першопочаткові інвестиції, в т.ч.:	\$2000
Розробка веб-сервісу	\$10000
Розробка додатку під Android	\$4000
Розробка додатку під IOS	\$5000
Юридичні витрати	\$1000
Витрати на підтримку сайту та програм	\$5000
Витрати на рекламу	\$2000
<b>ВСЬОГО ІНВЕСТИЦІЙ ПО ПРОЕКТУ:</b>	<b>\$29000</b>

Заходити щодо можливих ризиків наведені у таблиці 5.30.

Таблиця 5.30 – Заходи щодо упередження та реагування на ризики

№	Ризик	Індикатори настання	Заходи щодо упередження	Заходи щодо реагування	Відповідальний
1	Некоректність роботи	Несприйнятність результатів	Відповідні оцінки моделювання	Вибір інших параметрів	Тертичний Роман
2	Зависання сайту	Припинення роботисервісу	Тестування	Вихід нових версій	Гуць Євгеній

Розрахунок ефетивності проекту представлено у таблиці 5.31.

Таблиця 5.31 – Розрахунок ефективності проекту, що аналізується

Показник	2020	2021	2022	2023
1. Сума інвестицій	\$29000	-	-	-
2. Виручка від реалізації	-	\$18500	\$16000	\$25400
3. Витрати на експлуатацію проекту	-	\$3200	\$2400	\$2900
4. Амортизаційні відрахування	-	\$320	\$320	\$520
5. Ставка дисконту	-	18%	18%	18%
6. Грошові потоки	-	\$16000	\$14000	\$24000
7. Дисконтовані грошові потоки	-	\$13500	\$10000	\$14000
8. Дисконтовані грошовий потік з початку експлуатації проекту	-	\$13500	\$24000	\$38000
9. Дисконтовані грошові потоки за ставкою дисконту 40%	-	\$12000	\$7000	\$8500
10. Дисконтовані вигоди, тис. грн.	-	\$11500	\$12000	\$16000
11. Дисконтовані витрати, тис. грн.	-	\$3000	\$2000	\$19000

#### Розрахунок ефективності проекту

1. Сума інвестицій у проект становить \$29000.
2. Дисконтовані грошові потоки в результаті реалізації проекту становитимуть за 2020-2022рр. \$38000.
3. Чиста теперішня вартість проекту  $\$38000 - \$29000 = \$9000$ . Оскільки,  $NPV > 0$ , інвестиційний проект є вигідним для підприємства інвестора. За 3 роки функціонування проекту грошовий потік не лише задовольняє очікування інвестора у відношенні щодо одержання доходу, а й перевищують очікувані доходи на \$9000.
4. Термін окупності інвестицій. Застосуємо алгоритм розрахунку інвестицій:

$$TO = (t_0 - 1) + \frac{CI - \sum PV_{(t_0-1)}}{PV_{t_0}}$$

де  $t_0$  - номер першого року, у якому досягається умова ,

CI – сума інвестицій у проект.

PV – дисконтовані грошові потоки.

Отже, для даного прикладу,  $TO = (3-1) + (29000 - 24000) / 24000 = 2,208$  роки

5. Внутрішня норма рентабельності. При ставці дисконту 40%, NPV дорівнює \$26000. Отже,  $IRR = 0,18 + (38000 * (0,4 - 0,18) / (38000 - 26000)) = 0,89$ , або при ставці 89% сумарні дисконтовані вигоди дорівнюють сумарним дисконтованим витратам. Тобто IRR є ставкою дисконту, при якій NPV проекту дорівнює нулю.

6. Коефіцієнт вигід/витрат.  $\$40000 / \$7000 = 5,71$ . За \$1 теперішньої вартості вкладених коштів у проект підприємство отримає \$5,71 теперішньої вартості доходу.

7. Індекс прибутковості  $\$38000 / \$29000 = 1,31$  Отже,  $PI > 0$  і проект є ефективним.

## 5.9 Інвестиційна та фінансове забезпечення стартап-проекту

### Інвестиційна пропозиція

1. Назва проекту та відомості про організацію

Назва стартап-проекту: FutureViews

Галузева приналежність стартап-проекту: ІТ

Форма участі інвестора в стартап-проекті: надання інвестицій за договором простого товариства (договору про спільну діяльність)

Місце реалізації стартап-проекту: Gorizont1, м. Київ

Передбачувана дата початку реалізації стартап-проекту: 01.07.2020

Повне найменування організації / підприємства (партнера по проекту / ініціатора проекту): ТОВ “Ф’ючер В’юз”

Форма власності: приватна

Дата реєстрації: 01.04.2020

Поточні фінансові результати: -29000\$

Опис проекту (ідея/передумови, стан та етапи проекту, очікувані результати):

Ідея, передумови

FutureViews – веб додаток для завантаження, аналізу та прогнозування нестационарних, нелінійних, динамічних процесів у різних сферах застосування. Використовуючи статистичні дані, які нам необхідно проаналізувати користувач їх підвантажує та застосовує різноманітні математичні операції для аналізу та прогнозування.

Етапи:

1. Створення сайту, запуск маркетингової кампанії
2. Створення мобільних додатків
3. Розвиток сайту та додатків, розробка нового функціоналу
4. Вихід на міжнародний ринок

Результат - міжнародний проект масового форкасту, статистична платформа по набору даних

## 2. Характеристика ринку планованої до випуску продукції

### 2.1. Характеристика запланованої до випуску продукції:

- найменування та опис продукції: веб-сервіс futureviews.com, мобільний додаток FutureViews;
- основні споживачі: малий та середній бізнес у різних сферах діяльності та без кордонних меж;
- зацікавлені державні та науково-дослідницькі інститути у роботі з талановитою молоддю.

### 2.2. Опис поточного стану внутрішнього ринку планованої до випуску продукції:

- молодь України, від 15 до 35 років - 11 млн;
- кількість зацікавлених видів бізнесів більше 100 типів.

### 2.3. Основні показники внутрішнього ринку планованої до випуску продукції:

- річний обсяг промислового виробництва продукції, в поточних цінах: у 2021 році;
- рентабельність продажів продукції (в%): 45.7 (у 2021 році);
- ступінь концентрації (приблизна сумарна частка ринку п'яти найбільших гравців, в%): 100.

## 3. Перспективність і конкурентні переваги проекту

### 3.1. Опис наявної інфраструктури:

- організації доставки Нова пошта, Укрпошта.

### 3.2. Стратегічні переваги:

- ненасичений ринок пропозиції;
- легкий вихід на міжнародну арену (організація доставки за межами України, підтримка серверів та трафіку сервісу).



#### 4. Потреба у фінансуванні

Загальні інвестиційні витрати за стартап-проектом, всього: 29000\$, в тому числі капітальні витрати: 2200\$.

Потреба у фінансуванні проекту за рахунок інвестора: є, адже на розкрутку сервісу знадобиться багато часу, і потрібні будуть кошти на підтримку, доки сайт не почне приносити прибуток

#### 5. Попередні фінансові показники проекту

Виручка, без ПДВ (після виходу на проектну потужність): \$24000

Простий /динамічний термін окупності: 2.2 роки

Внутрішня норма прибутковості (IRR): 89 %

Чиста поточна вартість (NPV): \$29000.

Ставка дисконтування: 18 %

## ПЕРЕЛІК ПОСИЛАНЬ

1. Дані акцій компанії Apple на ресурсі Yahoo Finance. URL: <https://finance.yahoo.com/quote/AAPL>.
2. Дані акцій компанії Google на ресурсі Yahoo Finance. URL: <https://finance.yahoo.com/quote/GOOG>.
3. Дані акцій компанії Microsoft на ресурсі Yahoo Finance. URL: <https://finance.yahoo.com/quote/MSFT>.
4. Дані акцій компанії Amazon на ресурсі Yahoo Finance. URL: <https://finance.yahoo.com/quote/AMZN>.
5. Бідюк П. І., Романенко В. Д., Тимошук О. Л. Аналіз часових рядів (навчальний посібник). Київ: Політехніка, 2010. 317 с.
6. Згуровський М. З., Бідюк П. І., Терентьев О. М. Байєсівські мережі в системах підтримки прийняття рішень. Київ: ТОВ Видавниче Підприємство «Едельвейс», 2015. 300 с.
7. Бідюк П.І., Коршевнюк Л.О. Проектування комп'ютерних інформаційних систем підтримки прийняття рішень: навч. посіб. / Київ: ННК «ІПСА» НУТУ «КПІ», 2010. 340 с.
8. Бидюк П.И., Терентьев А.Н., Гасанов А.С. Построение и методы обучения Байесовских сетей. *Кибернетика и системный анализ*. 2005. № 4. С. 133 – 147.
9. Бідюк П.І., Половцев О.В. Аналіз та моделювання економічних процесів перехідного періоду. Київ: НТУУ «КПІ», 1999. 230 с.
10. Бидюк П.И., Баклан И.В., Рифа В.Н. Системный подход к построению регрессионной модели по временным рядам. *Системні дослідження та інформаційні технології*. 2002. № 3. С. 114 – 131.
11. Enders W. Applied econometric time series. New York: John Wiley & Sons, Inc., 1995. 434 p.

12. Johnston J., DiNardo J. *Econometric methods*. New York: McGraw—Hill, Inc., 1997. 530 p.
13. Chatfield C. *Time series forecasting*. London: Chapman & Hall, 2000. 267 p.
14. Fryzlewicz, P., Van Belleghem, S. & von Sachs, R. Forecasting non-stationary time series by wavelet process modelling. *Ann Inst Stat Math* 55. 2003. P. 737–764.
15. Gers F. A., Schmidhuber J., Cummins F. Learning to Forget: Continual Prediction with LSTM. *Neural Computation* 12 (10). 2000. P. 2451–2471.
16. Гуць, Є. В. Моделі нелінійних нестационарних процесів у фінансах : дипломна робота ... бакалавра : 6.050101 Комп'ютерні науки / Гуць Євгеній Віталійович. Київ, 2019. 101 с.
17. Understanding LSTM Networks URL: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
18. Introduction to Bayesian Networks URL: <https://towardsdatascience.com/introduction-to-bayesian-networks-81031eed94e>.

## ДОДАТОК А ЛІСТИНГ ПРОГРАМИ

```
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style('whitegrid')
plt.style.use("fivethirtyeight")
%matplotlib inline

# For reading stock data from yahoo
from pandas_datareader.data import DataReader

# For time stamps
from datetime import datetime
tech_list = ['AAPL', 'GOOG', 'MSFT', 'AMZN']

end = datetime.now()
start = datetime(end.year - 5, end.month, end.day)

for stock in tech_list:
    globals()[stock] = DataReader(stock, 'yahoo', start, end)
company_list = [AAPL, GOOG, MSFT, AMZN]
company_name = ["APPLE", "GOOGLE", "MICROSOFT", "AMAZON"]

for company, com_name in zip(company_list, company_name):
    company["company_name"] = com_name
```

```

df = pd.concat(company_list, axis=0)
df.tail(10)
AAPL.describe()
AAPL.info()
plt.figure(figsize=(12, 8))
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"{tech_list[i - 1]}")
plt.figure(figsize=(12, 8))
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"{tech_list[i - 1]}")
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()
df.groupby("company_name").hist(figsize=(12, 12));

```

```
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()
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()
```

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

```
for i, company in enumerate(company_list, 1):
```

```
    plt.subplot(2, 2, i)
```

```
    sns.distplot(company['Daily Return'].dropna(), bins=100, color='purple')
```

```
    plt.ylabel('Daily Return')
```

```
    plt.title(f'{company_name[i - 1]}')
```

```
closing_df = DataReader(tech_list, 'yahoo', start, end)['Adj Close']
```

```
closing_df.tail()
```

```
tech_rets = closing_df.pct_change()
```

```

tech_rets.tail()

sns.jointplot('MSFT', 'MSFT', tech_rets, kind='scatter', color='seagreen')
sns.jointplot('GOOG', 'AAPL', tech_rets, kind='scatter')
sns.pairplot(tech_rets, kind='reg')

sns.heatmap(tech_rets.corr(), annot=True, cmap='summer')

sns.heatmap(closing_df.corr(), annot=True, cmap='summer')

df = DataReader('AAPL', data_source='yahoo', start='2015-01-01',
end=datetime.now())

plt.figure(figsize=(16,8))
plt.title('Close Price History')
plt.plot(df['Close'])
plt.xlabel('Date', fontsize=18)
plt.ylabel('Close Price USD ($)', fontsize=18)
plt.show()

#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) * .8 ))

training_data_len

#Scale the data
from sklearn.preprocessing import MinMaxScaler

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

```



```

scaled_data
#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
from keras.models import Sequential
from keras.layers import Dense, LSTM

#Build the LSTM model
model = Sequential()
model.add(LSTM(50, return_sequences=True, input_shape=
(x_train.shape[1], 1)))

```

```
model.add(LSTM(50, 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)
model.summary()

#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)))
print('rmse=' + str(rmse))

mae = np.mean(np.abs(predictions - y_test))
print('mae=' + str(mae))

mape = np.mean( np.abs((predictions - y_test) / y_test)) * 100
print('mape=' + str(mape))

# Plot the data
train = data[:training_data_len]
valid = data[training_data_len:]
valid['Predictions'] = predictions

# Visualize the data
plt.figure(figsize=(16,8))
plt.title('AMAZON')
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()

from sklearn.linear_model import LinearRegression
from sklearn.svm import SVR
from sklearn.model_selection import train_test_split

df = AAPL[['Close']]
print(df.tail())
forecast_out = 30 #'n=30' days
```

```

df['Prediction'] = df[['Close']].shift(-forecast_out)
print(df.tail())
# Convert the dataframe to a numpy array
X = np.array(df.drop(['Prediction'],1))

#Remove the last '30' rows
X = X[:-forecast_out]
print(X)
# Convert the dataframe to a numpy array
y = np.array(df['Prediction'])
# Get all of the y values except the last '30' rows
y = y[:-forecast_out]
print(y)
# Split the data into 80% training and 20% testing
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
# Create and train the Support Vector Machine (Regressor)
svr_rbf = SVR(kernel='rbf', C=1e3, gamma=0.1)
svr_rbf.fit(x_train, y_train)
# Testing Model: Score returns the coefficient of determination R^2 of the
prediction.
# The best possible score is 1.0
svm_confidence = svr_rbf.score(x_test, y_test)
print("svm confidence: ", svm_confidence)
# Create and train the Linear Regression Model
lr = LinearRegression()
# Train the model
lr.fit(x_train, y_train)
lr_confidence = lr.score(x_test, y_test)
print("lr confidence: ", lr_confidence)
x_forecast = np.array(df.drop(['Prediction'],1))[:-forecast_out:]

```

```
print(x_forecast)

# Print linear regression model predictions for the next '30' days
lr_prediction = lr.predict(x_forecast)
print(lr_prediction)

# Print support vector regressor model predictions for the next '30' days
svm_prediction = svr_rbf.predict(x_forecast)
print(svm_prediction)

valid_close = np.array(df[['Close']][-forecast_out:])

# Get the root mean squared error (RMSE)
rmse = np.sqrt(np.mean(((valid_close - lr_prediction) ** 2)))
print('rmse=' + str(rmse))

mae = np.mean(np.abs(valid_close - lr_prediction))
print('mae=' + str(mae))

mape = np.mean( np.abs((valid_close - lr_prediction) / valid_close) * 100)
print('mape=' + str(mape))

# Get the root mean squared error (RMSE)
rmse = np.sqrt(np.mean(((valid_close - svm_prediction) ** 2)))
print('rmse=' + str(rmse))

mae = np.mean(np.abs(valid_close - svm_prediction))
print('mae=' + str(mae))

mape = np.mean( np.abs((valid_close - svm_prediction) / valid_close) * 100)
print('mape=' + str(mape))

# Plot the data
valid = df[['Close']][-forecast_out:]
valid['Predictions'] = lr_prediction
```

```
# Visualize the data
plt.figure(figsize=(16,8))
plt.title('LinearRegression')
plt.xlabel('Date', fontsize=18)
plt.ylabel('Close Price USD ($)', fontsize=18)
plt.plot(valid[['Close', 'Predictions']])
plt.legend(['Real', 'Predictions'], loc='lower right')
plt.show()

# Plot the data
valid = df[['Close']][-forecast_out:]
valid['Predictions'] = svm_prediction

# Visualize the data
plt.figure(figsize=(16,8))
plt.title('Support Vector Machine (Regressor)')
plt.xlabel('Date', fontsize=18)
plt.ylabel('Close Price USD ($)', fontsize=18)
plt.plot(valid[['Close', 'Predictions']])
plt.legend(['Real', 'Predictions'], loc='lower right')
plt.show()
```