КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

ІМЕНІ ТАРАСА ШЕВЧЕНКА

Факультет комп’ютерних наук та кібернетики
Кафедра системного аналізу та теорії прийняття рішень

Кваліфікаційна робота на здобуття ступеня бакалавра

за спеціальністю 124 Системний аналіз

на тему:

ПРОГНОЗУВАННЯ ЦІНИ НА МЕТАЛИ ЗА ДОПОМОГОЮ РЕГРЕСІЇ

Виконала студентка 4-го курсу
Широконос Софія В’ячеславівна

Науковий керівник:
асистент, кандидат технічних наук
Махно Михайло Федорович

Засвідчую, що в цій роботі немає запозичень з праць інших авторів без відповідних посилань.

Студент

Роботу розглянуто й допущено до захисту на засіданні кафедри системного аналізу та теорії прийняття рішень

«01» червня 2023 р.,
протокол № 13
Завідувач кафедри
О. Г. Наконечний ____________
(підпис)

Київ - 2023

РЕФЕРАТ

Обсяг роботи 53 сторінки, 56 ілюстрацій, 7 використаних джерел.

Об'єктом дослідження є моделі ARIMA, випадкового лісу та лінійної регресії.

Метою нашої роботи є створення моделей для прогнозування цін на такі дорогоцінні метали, як срібло, золото, платина, паладій. Для цього ми використаємо авторегресійні моделі, моделі лінійної регресії та моделі випадкового лісу.

Ми ставимо перед собою завдання проаналізувати результати передбачення за допомогою вищеописаних моделей строком на один місяць, для визначення кращих моделей для цієї мети. Також, важливим є визначення зв'язку між цінами на популярні енергоресурси та цінами на дорогоцінні метали

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ЗНАЧЕНЬ

AR - авторегресія

MA - ковзне середнє

ARMA - авторегресія та ковзна середня

ARIMA - інтегрована модель авторегресії - ковзного середнього

LR – лінійна регресія

RF - випадковий ліс

ВСТУП

Аналіз ролі дорогоцінних металів, і, насамперед, золота у світовій економіці, сьогодні є досить актуальним, оскільки за останні кілька десятиліть світовий ринок золота істотно змінився, як за функціональними, так і за структурними параметрами. Такі зміни відбулися в рамках глобального процесу «демонетизації» золота як основного грошового активу.

Зміни її грошових функцій і властивостей відбулися внаслідок трансформації світової валютно-фінансової системи, геополітичної напруженості та розвитку нових технологій.

Золото втратило низку своїх функцій грошового активу: воно перестало бути звичайним еквівалентом у порівнянні вартості всіх інших товарів, воно не використовується як критерій формування валютних курсів, оскільки національні валюти більше не мають офіційний золотий вміст.

Золото також втратило функцію світових грошей, які тепер є фіатною валютою – штучною міжнародною валютою. Проте золото та інші дорогоцінні метали набули нових функцій, у тому числі функції незамінної сировинної складової в багатьох галузях світової економіки, а також зміцнили свої властивості надійного та ліквідного інвестиційного активу та засобу збереження ліквідних активів.

РОЗДІЛ 1. ТЕОРЕТИЧНА ЧАСТИНА

1. Моделі ARIMA

1.1. Модель AR

У моделі авторегресії ми прогнозуємо цільову змінну, використовуючи лінійну комбінацію попередніх значень змінної. Термін «авторегресія» вказує на те, що це регресія змінної відносно самої себе. Таким чином, авторегресійну модель порядку p ми можемо записати як:

$y_t = c + k_1 y_{t-1} + k_2 y_{t-2} + \cdots + k_{p-1} y_{t-p+1} + k_p y_{t-p} + \varepsilon_t$, де ε_t—білий шум.

Якщо немає інших даних, вважається, що білий шум має розподіл, близький до нормального. Така модель має назву AR(p) модель, що означає «авторегресійна модель порядку p». Авторегресійні моделі є надзвичайно гнучкими в обробці широкого діапазону різних шаблонів часових рядів. До переваг моделі AR належить також її простота та досить невисока схильність до перенавчання.

Суттєвим недоліком моделі AR є її відносно невелика точність та потреба у досить великій кількості даних для навчання. Окрім того, ця модель, гарно працює лише при невеликих значеннях дисперсії часового ряду—і дає значно гірші результати для часових рядів з великою дисперсією.

Також, модель AR не спроможна будувати адекватний прогноз більше, аніж на один крок. У нашій роботі будуть розглянуті моделі AR(p) для p від 1 до 10, та виявлено найкраще значення параметру p для кожного з чотирьох дорогоцінних металів: золота, срібла, платини та паладію.

1.2. Модель MA

Замість того, щоб використовувати минулі значення змінної прогнозу в регресії, як це робить модель авторегресії AR(p), модель ковзного середнього MA(q) використовує попередні похибки передбачення у регресійній моделі виду:

$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_{p-1} \varepsilon_{t-p+1} + \theta_p \varepsilon_{t-p}$, де ε_t—білий шум.

Звісно, у дійсності ми не спостерігаємо значення ε_t, тому модель MA(q) не є регресією у її звичайному значенні. Цікаво, що кожне значення y_t можна розглядати як зважене ковзне середнє кількох попередніх похибок передбачення.

1.3. Модель ARMA

Модель авторегресії та ковзного середнього об'єднує у собі моделі AR та MA. За рахунок об'єднання обох підходів, ця модель дозволяє будувати більш точний прогноз ціною трохи більших витрат на обчислення. Фактично, для моделі ARMA(p,q) має місце формула, утворена поєднанням формул AR(p) та MA(q):

$$y_t = c + k_1 y_{t-1} + k_2 y_{t-2} + \cdots + k_{p-1} y_{t-p+1} + k_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_{p-1} \varepsilon_{t-p+1} + \theta_p \varepsilon_{t-p}.$$

Саме модель ARMA є найбільш розповсюдженою з моделей класу ARIMA. Існує думка, що процеси ARMA це практично все, що потрібно знати про стаціонарні процеси. Є теорема, котра каже, що будь-який стаціонарний процес можна подати у вигляді авторегресії з нескінченною кількістю лагів. Для ARMA процесу можна вибрати p та q досить великими, і вибрати коефіцієнт перед ігреками та перед епсилонами таким чином, щоб максимально наблизити прогнозовані значення до фактичних. Саме тому ARMA (p, q) моделі достатньо, щоб пояснити будь-який стаціонарний процес.

1.4. Модель ARIMA

ARIMA розшифровується як інтегрована модель авторегресії та ковзного середнього. Вона є розширенням моделей ARMA для нестаціонарних часових рядів, які можна зробити стаціонарними взяттям різниць деякого порядку від вихідного часового ряду. Модель ARIMA(p, d, q) означає, що різниці часового ряду порядку d підкоряються моделі ARMA(p, q).

Цікавою особливістю моделі ARIMA(p, d, q) є те, що маніпулюючи її компонентами можна отримати всі вищеописані моделі. Так, модель ARIMA(p, 0, 0) еквівалентна моделі AR(p). Модель ARIMA(0, 0, q) еквівалентна моделі MA(q). А модель ARIMA(p, 0, q) повністю еквівалентна моделі ARMA (p, q).

Слід зазначити, що у написаній нами програмі коефіцієнти для кожної конкретної моделі можуть бути лише строго додатними.

2. Лінійна регресія

Лінійна регресія є основним і найбільш широко використовуваним типом прогнозного аналізу. Загальна ідея регресії полягає в дослідженні двох речей: чи добре набір змінних предикторів справляється з прогнозуванням кінцевої (залежної) змінної, та які змінні, є значущими предикторами змінної результату, і яким чином вони, вказуючи на величину та знак бета-оцінок, впливають на змінну результату? Ці регресійні оцінки використовуються для пояснення зв’язку між однією залежною змінною та однією або кількома незалежними змінними. Найпростіша форма рівняння регресії з однією залежною та однією незалежною змінними визначається формулою $y = c + b*x$, де y = оцінка оцінки залежної змінної, c = константа, b = коефіцієнт регресії та x = оцінка на незалежна змінна.

У статистиці лінійна регресія — це лінійний підхід для моделювання зв’язку між скалярною відповіддю та однією або декількома пояснювальними змінними (також відомими як залежні та незалежні змінні). Випадок однієї пояснювальної змінної називається простою лінійною регресією; для більш ніж одного процес називається множинною лінійною регресією. Цей термін відрізняється від багатовимірної лінійної регресії, де прогнозується кілька корельованих залежних змінних, а не одна скалярна змінна.

У лінійній регресії співвідношення моделюються за допомогою лінійних функцій прогнозування, невідомі параметри моделі яких оцінюються на основі даних . Такі моделі називаються лінійними. Зазвичай умовне середнє значення відповіді, задане значеннями пояснювальних змінних (або предикторів), вважається афінною функцією цих значень; рідше використовується умовна медіана або інший квантиль. Як і всі форми регресійного аналізу, лінійна регресія фокусується на умовному розподілі ймовірностей відповіді з урахуванням значень предикторів, а не спільного розподілу ймовірностей усіх цих змінних, що є областю багатофакторного аналізу.

Лінійна регресія була першим типом регресійного аналізу, який ретельно вивчався та широко використовувався в практичних додатках. Це пояснюється тим, що моделі, які лінійно залежать від своїх невідомих параметрів, легше підібрати, ніж моделі, які нелінійно пов’язані з їхніми параметрами, і тому, що статистичні властивості отриманих оцінок легше визначити.

Лінійна регресія має багато практичних застосувань. Більшість програм належить до однієї з наступних двох широких категорій:

- Якщо метою є передбачення, прогнозування або зменшення помилок, лінійна регресія може бути використана для підгонки прогностичної моделі до спостережуваного набору даних значень відповіді та пояснювальних змінних. Після розробки такої моделі, якщо зібрані додаткові значення пояснювальних змінних без відповідного значення відповіді, підібрану модель можна використовувати для прогнозування відповіді.
- Якщо мета полягає в тому, щоб пояснити варіацію змінної відповіді, яку можна віднести до варіації пояснювальних змінних, лінійний регресійний аналіз можна застосувати для кількісної оцінки сили зв’язку між відповіддю та пояснювальними змінними, і, зокрема, для визначення того, чи деякі пояснювальні змінні можуть взагалі не мати лінійного зв’язку з відповіддю, або щоб визначити, які підмножини пояснювальних змінних можуть містити надлишкову інформацію про відповідь.

Моделі лінійної регресії часто підбираються за допомогою підходу найменших квадратів , але вони також можуть бути підігнані іншими способами, наприклад шляхом мінімізації «відсутності відповідності» в іншій нормі (як у випадку з регресією найменших абсолютних відхилень) або шляхом мінімізації штрафного версія функції вартості за методом найменших -квадратів, як у ридж-регресії (L_2 - штраф за норму) і ласо (L_1 - штраф за норму). І навпаки, підхід найменших квадратів можна використовувати для підгонки моделей, які не є лінійними. Таким чином, хоча терміни "найменші квадрати" і "лінійна модель" тісно пов'язані, вони не є синонімами.

Стандартні моделі лінійної регресії зі стандартними методами оцінки роблять низку припущень щодо змінних предиктора, змінних відповіді та їх взаємозв'язку. Було розроблено численні розширення, які дозволяють пом'якшити кожне з цих припущень (тобто звести до слабшої форми), а в деяких випадках повністю усунути. Зазвичай ці розширення роблять процедуру оцінки складнішою та трудомісткішою, а також можуть вимагати більше даних для створення так само точної моделі.

Нижче наведемо основні припущення, зроблені стандартними моделями лінійної регресії зі стандартними методами оцінки:

- Слабка екзогенність. По суті, це означає, що змінні предиктора x можна розглядати як фіксовані значення, а не як випадкові змінні . Це означає, наприклад, що передбачається, що змінні предиктора не містять помилок, тобто не містять помилок вимірювання. Хоча це припущення є нереалістичним у багатьох налаштуваннях, відмова від нього призводить до значно складніших моделей помилок у змінних .
- Лінійність. Це означає, що середнє значення змінної відповіді є лінійною комбінацією параметрів (коефіцієнтів регресії) і змінних предиктора. Зверніть увагу, що це припущення є набагато менш обмежувальним, ніж може здатися на перший погляд. Оскільки змінні предиктора розглядаються як фіксовані значення, лінійність насправді є лише обмеженням на параметри. Самі змінні предиктора можна довільно трансформувати, і фактично можна додати декілька копій однієї базової змінної предиктора, кожна з яких

трансформується по-різному. Ця техніка використовується, наприклад, у поліноміальній регресії , яка використовує лінійну регресію, щоб відповідати змінній відповіді як довільному поліномуфункція (до заданого ступеня) змінної предиктора. З такою великою гнучкістю такі моделі, як поліноміальна регресія, часто мають «занадто велику потужність», оскільки вони, як правило, переобладнують дані. Як наслідок, зазвичай необхідно використовувати певний вид регулярізації, щоб запобігти виходу необґрунтованих рішень у процесі оцінювання. Типовими прикладами є хребтова регресія та ласо-регресія. Також можна використовувати байєсівську лінійну регресію, яка за своєю природою більш-менш несприйнятлива до проблеми переобладнання.

•Постійна дисперсія (гомоскедастичність). Це означає, що дисперсія помилок не залежить від значень змінних предикторів. Таким чином, мінливість відповідей для заданих фіксованих значень предикторів є однаковою незалежно від того, наскільки великими чи малими є відповіді. Часто це не так, оскільки змінна, середнє значення якого велике, зазвичай матиме більшу дисперсію, ніж змінна, середнє значення якого мале. Наприклад, людина, чий дохід за прогнозами становитиме 100 000 доларів США, може легко мати фактичний дохід 80 000 або 120 000 доларів США, тобто стандартне відхиленняприблизно 20 000 доларів США, тоді як інша особа з прогнозованим доходом 10 000 доларів США навряд чи матиме таке саме стандартне відхилення у 20 000 доларів США, оскільки це означатиме, що їхній фактичний дохід може коливатися від −10 000 до 30 000 доларів США. Відсутність гомоскедастичності є називають гетероскедастичністю. Щоб перевірити це припущення, можна перевірити графік залишків проти прогнозованих значень (або значень кожного окремого предиктора) на предмет «ефекту віяла» (тобто збільшення або зменшення вертикального розкиду під час руху зліва направо на графіку). Графік абсолютних чи квадратичних залишків у порівнянні з прогнозованими значеннями (або кожним предиктором) також можна перевірити на тенденцію чи кривизну

• Гетероскедастичність . Наявність гетероскедастичності призведе до використання загальної «середньої» оцінки дисперсії замість оцінки, яка враховує справжню структуру дисперсії. Це призводить до менш точного (але у випадку звичайних найменших квадратів, не зміщені) оцінки параметрів і зміщені стандартні помилки, що призводить до оманливих тестів і інтервальних оцінок. Середня квадратична помилка для моделі також буде неправильною. Різні методи оцінки, включаючи зважені найменші квадрати та використання узгоджених із гетероскедастичністю стандартних помилок, можуть обробляти гетероскедастичність у досить загальний спосіб. Методи байєсівської лінійної регресії також можна використовувати, коли припускається, що дисперсія є функцією середнього. У деяких випадках також можливо вирішити проблему, застосовуючи перетворення до змінної відповіді.

3. Модель випадкового лісу

Модель регресії випадкового лісу побудована на базі моделей дерев рішень. Дерева рішень використовуються як для регресії, так і для проблем класифікації. Вони візуально перетікають як дерева, звідси й назва, а у випадку регресії вони починаються з кореня дерева та слідують за розбиттям на основі змінних результатів, доки не буде досягнуто листкового вузла та отримано результат.

Загальний метод випадкових лісів рішень був вперше запропонований Хо в 1995 році. Хо встановив, що ліси дерев, розбитих похилими гіперплощинами, можуть отримати точність, коли вони ростуть, не страждаючи від перенавчання, доки ліси випадково обмежені, щоб бути чутливими лише для вибраних розмірів функції. Подальша робота в тому ж напрямку показала, що інші методи розбиття поводяться подібним чином, якщо вони випадковим чином змушені бути нечутливими до деяких розмірів ознак. Варто зауважити, що спостереження більш складного класифікатора (більшого лісу), що стає більш точним майже монотонно, різко суперечить загальноприйнятій думці, що складність класифікатора може зрости лише до певного рівня точності, перш ніж постраждає від перенавчання. Пояснення стійкості методу лісу до перенавчання можна знайти в теорії стохастичного розрізнення Клейнберга.

На ранній розвиток поняття випадкових лісів Бреймана вплинула робота Аміта та Джемана, які представили ідею пошуку випадкової підмножини доступних рішень під час розбиття вузла в контексті вирощування одного дерева . Ідея випадкового вибору підпростору від Хо також мала вплив на проектування випадкових лісів. У цьому методі вирощується ліс дерев, і вводиться варіація серед дерев шляхом проектування навчальних даних у випадково вибраний підпростір перед підгонкою кожного дерева чи кожного вузла. Нарешті, ідея рандомізованої оптимізації вузла, де рішення на кожному вузлі вибирається за допомогою рандомізованої процедури, а не детермінованої оптимізації, була вперше представленаТомас Г. Дітеріх.

Правильне введення випадкових лісів було зроблено в статті Лео Бреймана. У цьому документі описано метод побудови лісу некорельованих дерев за допомогою процедури, подібної до CART, у поєднанні з рандомізованою оптимізацією вузлів і пакетуванням . Крім того, ця стаття поєднує в собі кілька компонентів, деякі з яких були раніше відомі, а деякі — нові, які складають основу сучасної практики випадкових лісів, зокрема: використання вихідної помилки як оцінки помилки узагальнення, вимірювання важливості змінної через перестановку.

Дерева рішень є популярним методом для різних завдань машинного навчання. Існує думка, що навчання дерева найближче відповідає вимогам щодо використання готової процедури для інтелектуального аналізу даних, оскільки воно є інваріантним щодо масштабування та різних інших перетворень значень ознак, стійкий до включення нерелевантних функцій і створює моделі, які можна перевірити. Однак вони рідко бувають точними.

Зокрема, дерева, які вирощуються дуже глибоко, мають тенденцію вивчати дуже нерегулярні шаблони: вони переповнюють свої навчальні набори, тобто мають низький зсув, але дуже високу дисперсію. Випадкові ліси

— це спосіб усереднення кількох глибоких дерев рішень, навчених на різних частинах одного навчального набору з метою зменшення дисперсії. Це відбувається за рахунок невеликого збільшення зміщення та деякої втрати інтерпретації, але загалом значно підвищує продуктивність кінцевої моделі.

Ліси схожі на об’єднання зусиль алгоритму дерева рішень. Спільна робота багатьох дерев покращує продуктивність одного випадкового дерева. Хоча ліси не дуже схожі, вони дають ефект k-кратної перехресної перевірки.

РОЗДІЛ 2.ПРАКТИЧНА ЧАСТИНА

1. Огляд даних

У цій роботі було використано часові ряди з щомісячними даними про ціну в період з 01.02.1997 по 01.09.2020 на такі ресурси: нафту, вугілля, природний газ, срібло, золото, платину, паладій. Графіки їх цін можна побачити на рисунках 1.1.-1.7.

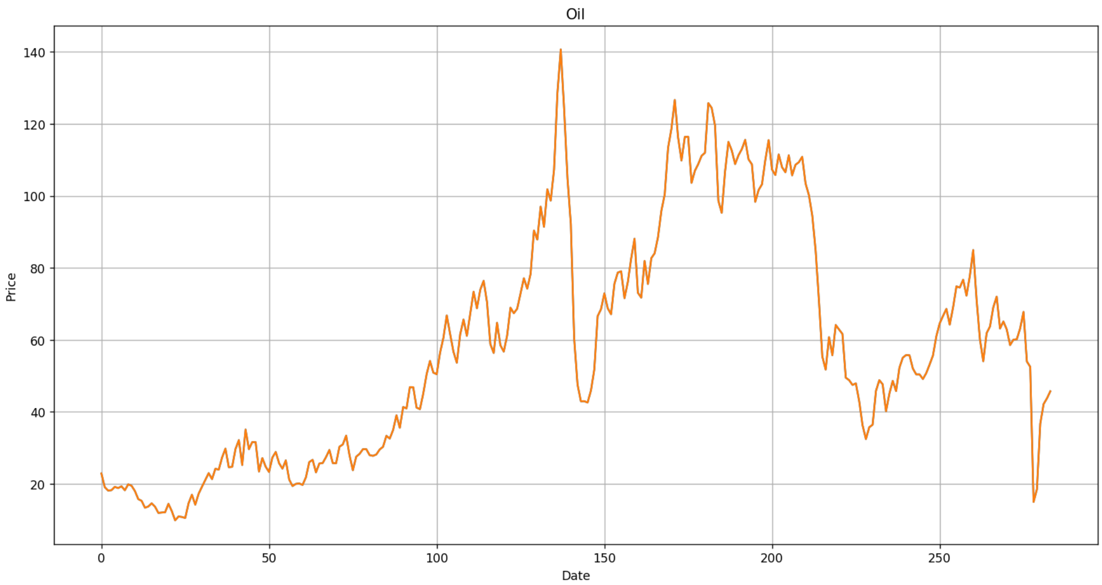

Рис. 1.1. Графік цін на нафту

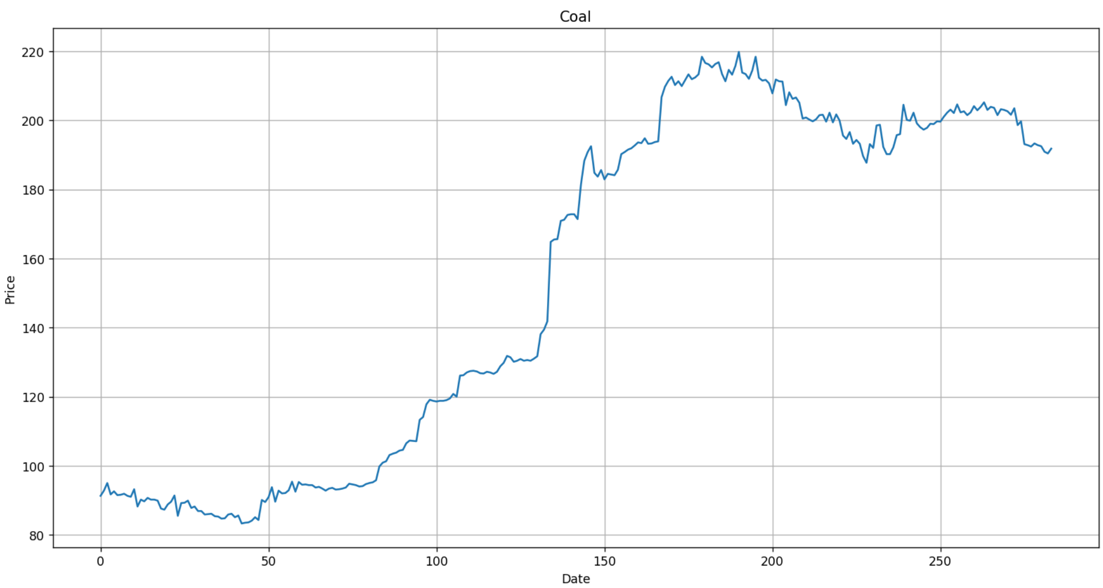

Рис. 1.2. Графік цін на вугілля

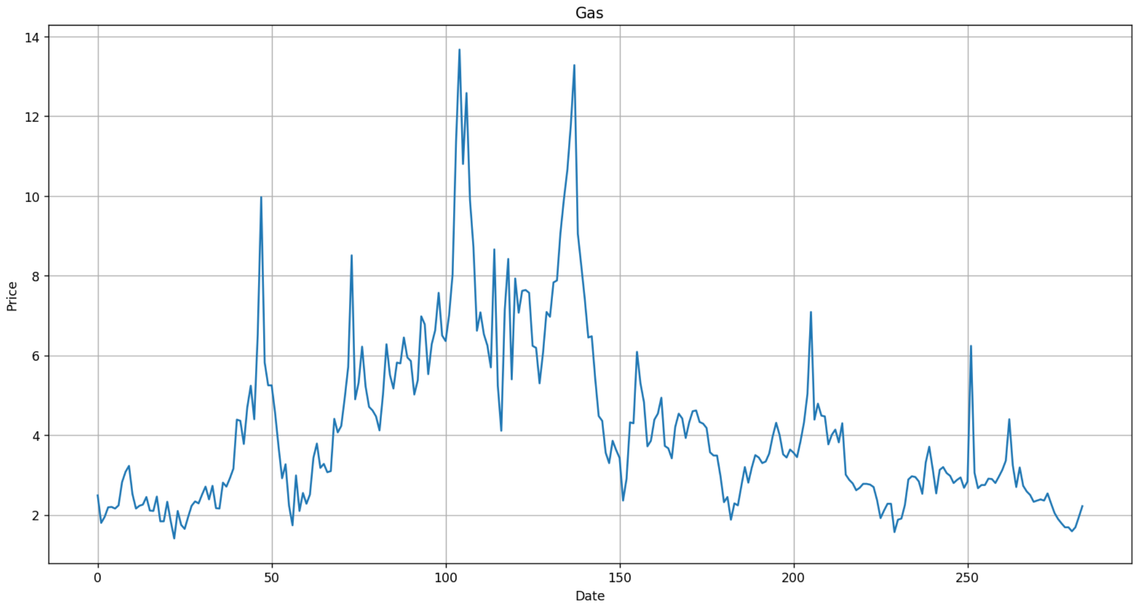

Рис. 1.3. Графік цін на природний газ

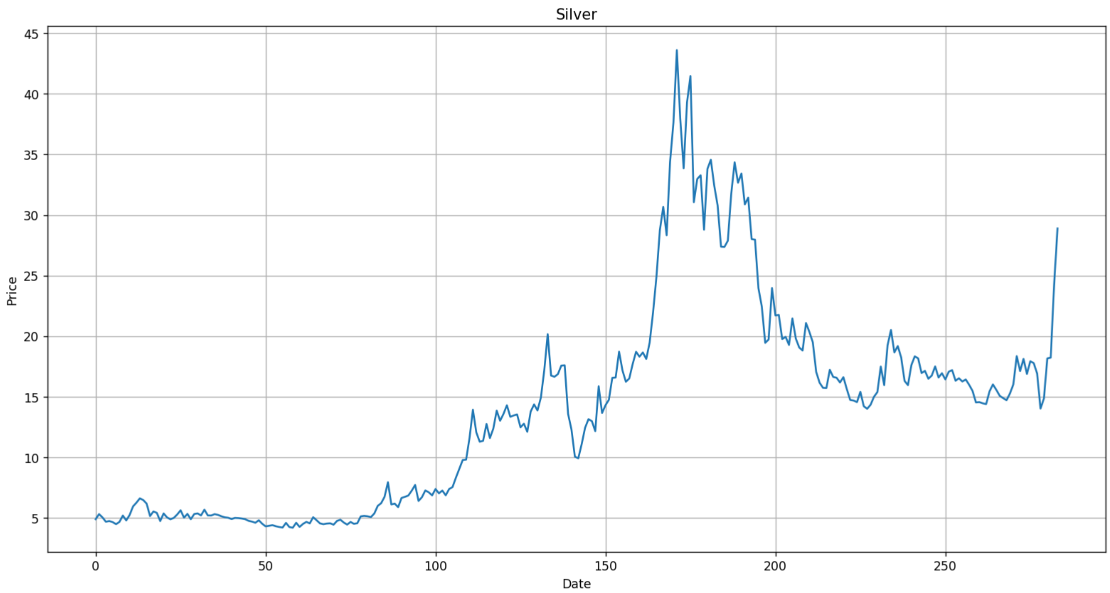

Рис. 1.4. Графік цін на срібло

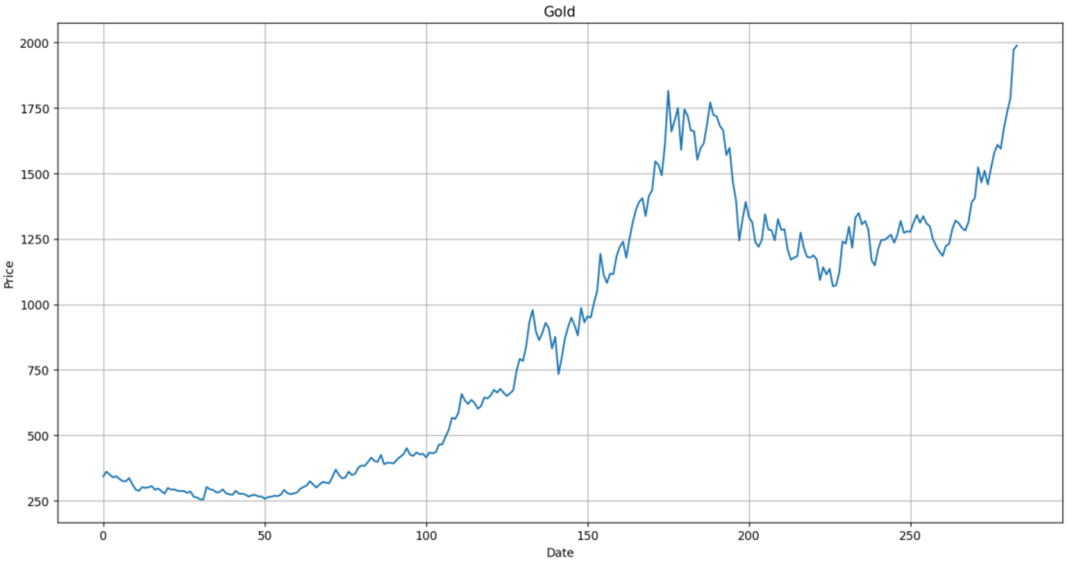

Рис. 1.5. Графік цін на золото

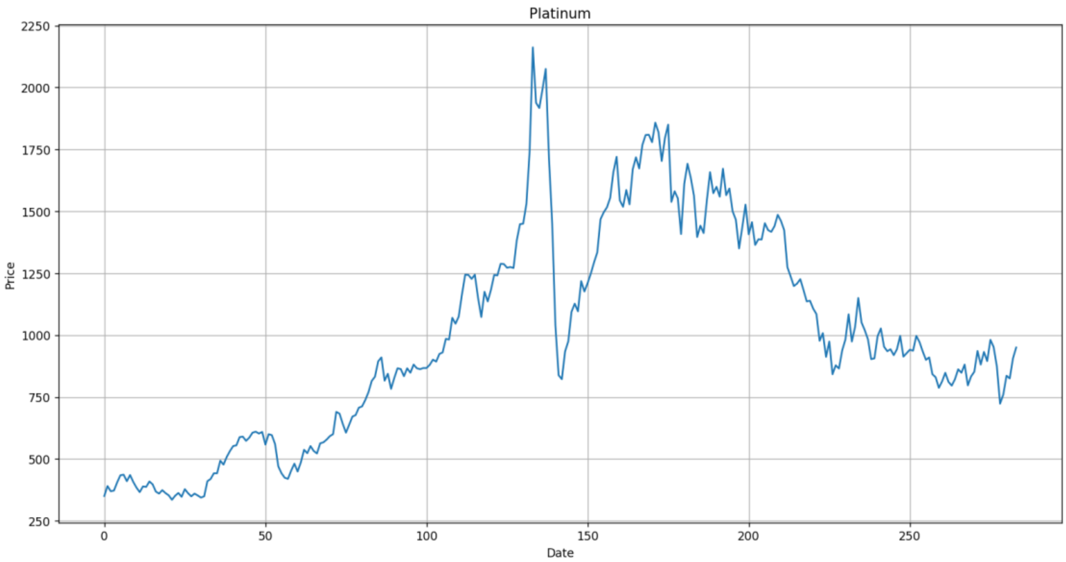

Рис. 1.6. Графік цін на платину

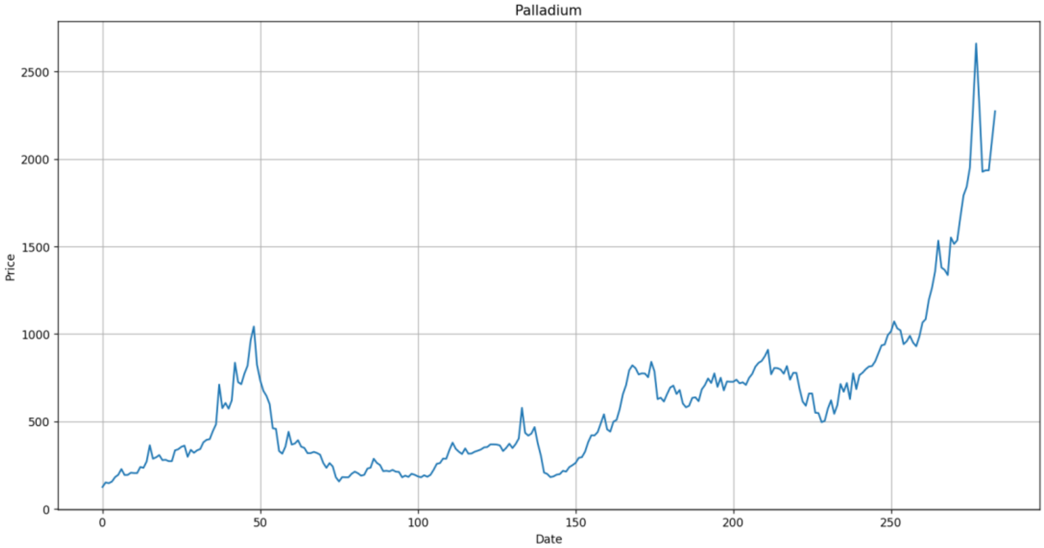

Рис. 1.7. Графік цін на паладій

Нафта, вугілля та природній газ належать до класу енергетичних ресурсів і використовуються лише для моделей лінійної регресії та випадкового лісу. Срібло, золото, платина та паладій належать до класу дорогоцінних металів, а їх ціни будуть цільовими для моделей передбачення.

2. Передбачення ціни срібла

2.1. Передбачення ціни срібла за допомогою моделей класу ARIMA

Спочатку розглянемо передбачення ціни срібла на основі моделей класу ARIMA. Програмою було здійснено підбір значень найкращих параметрів p, d, q методом підбору (від 1 до 5), де вибір кращої моделі здійснювався на основі критерію Акаїке. Результат такого підбору можна побачити на рисунку 2.1.

model	p	d	q	AIC
AR(3)	3.0	NaN	NaN	1021.860940
MA(5)	NaN	NaN	5.0	1212.027346
ARMA(5, 4)	5.0	NaN	4.0	1017.865249
ARIMA(5, 1, 4)	5.0	1.0	4.0	1010.512627

Рис. 2.1. Значення найкращих параметрів p, d, q

Оцінка ефективності моделей, виражена у середній похибці, максимальному відхиленні, сумарній абсолютній похибці та коефіцієнті детермінації, відображена на рисунку 2.2.

	AR	MA	ARMA	ARIMA
MEAN RESIDUAL	1.867120	2.439694	1.893240	1.849440
MAX RESIDUAL	6.512011	10.397355	6.197197	6.494804
SUM RESIDUAL	28.006801	36.595407	28.398597	27.741603
DETERMINATION COEFFICIENT	0.358470	0.815503	0.433216	0.405084

Рис. 2.2. Оцінка ефективності моделей ARIMA

Таким чином, найкращий коефіцієнт детермінації має модель MA, найменше максимальне відхилення має модель ARMA, а найменше середня та сумарна абсолютна похибка має модель ARIMA. Графіки передбачень моделями класу ARIMA для срібла можна побачити на рисунках 2.3-2.6.

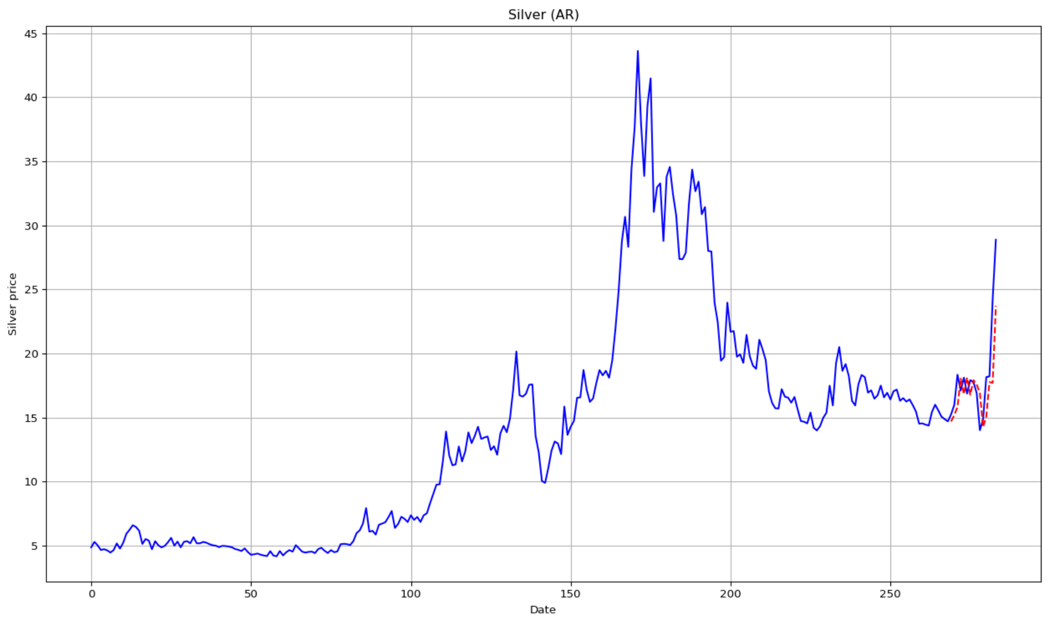

Рис. 2.3. Передбачення ціни срібла за допомогою моделі AR

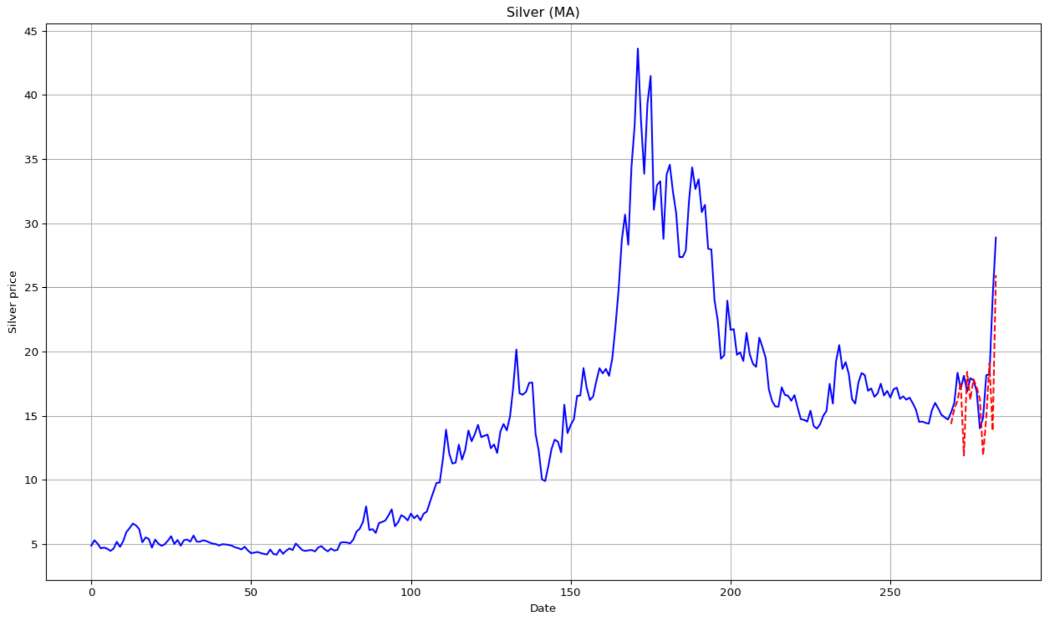

Рис. 2.4. Передбачення ціни срібла за допомогою моделі MA

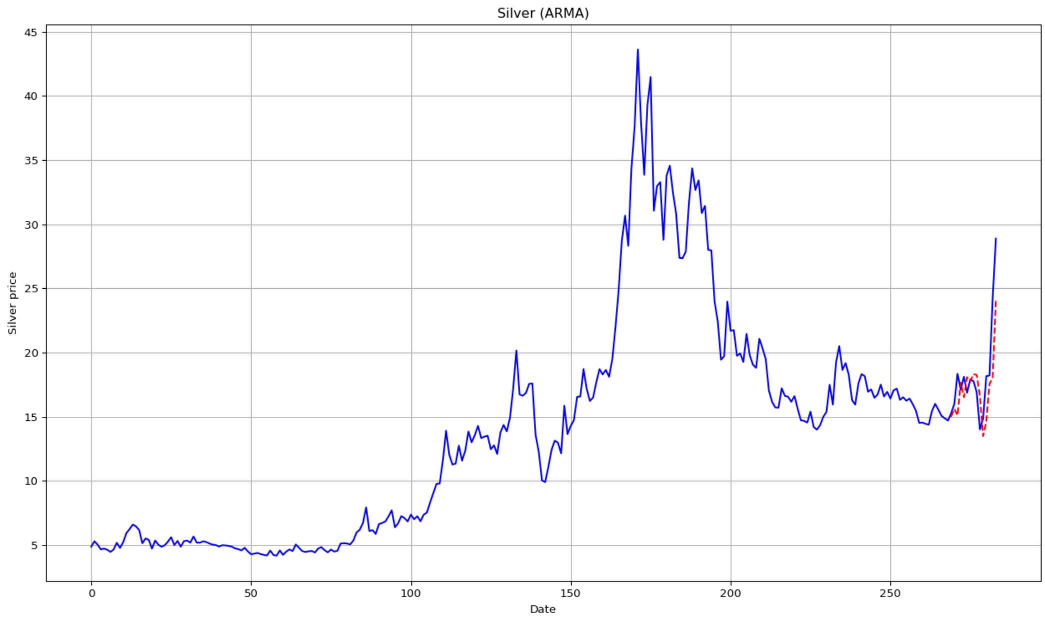

Рис. 2.5. Передбачення ціни срібла за допомогою моделі ARMA

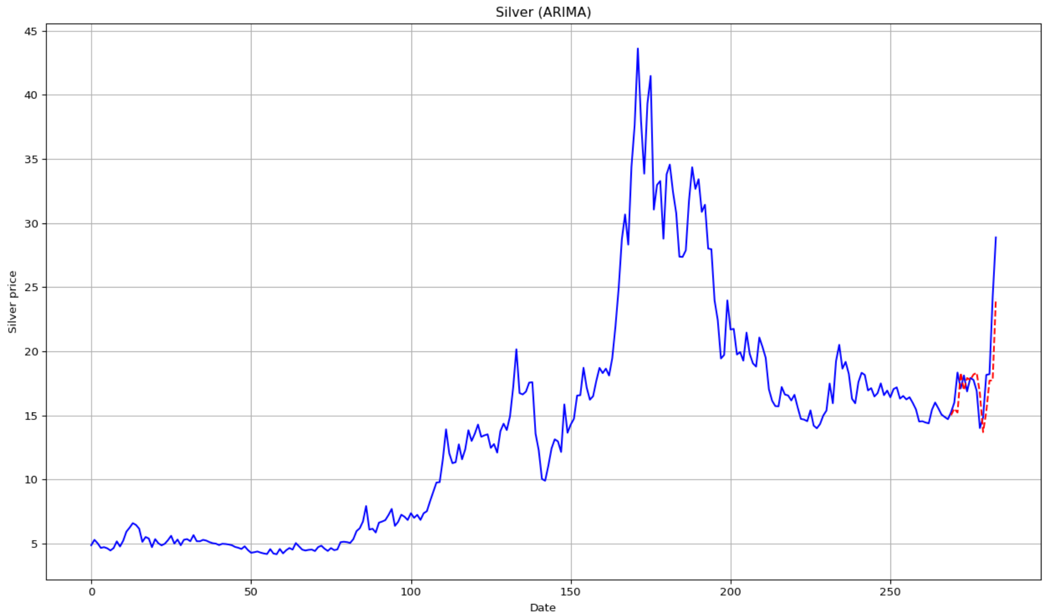

Рис. 2.6. Передбачення ціни срібла за допомогою моделі ARIMA

2.2. Передбачення ціни срібла за допомогою лінійної регресії

Для передбачення ціни срібла за допомогою лінійної регресії було застосовано три можливих набори параметри. Перший, під назвою енергетичні ресурси, передбачає включення до моделі лінійної регресії цін на енергетичні носії (а саме, нафту, вугілля та газ) за попередній місяць. Другий, під назвою метали, передбачає включення до моделі лінійної регресії цін на дорогоцінні метали (а саме, срібло, золото, платину, паладій) за попередній місяць. Третій, під назвою все, передбачає включення до моделі лінійної регресії цін як на енергетичні ресурси, так і на дорогоцінні метали за попередній місяць. Оцінка ефективності моделей лінійної регресії, виражена у середній похибці, максимальному відхиленні, сумарній абсолютній похибці та коефіцієнті детермінації, відображена на рисунку 2.7.

	energy resources	metal	all
MEAN RESIDUAL	3.554113	1.859603	2.314190
MAX RESIDUAL	13.815964	6.235115	7.734486
SUM RESIDUAL	53.311693	27.894050	34.712845
DETERMINATION COEFFICIENT	0.592957	0.383515	0.303092

Рис. 2.7. Оцінка ефективності моделей лінійної регресії

За коефіцієнтом детермінації найкращий результат у моделі з енергетичними ресурсами, однак найменшу абсолютну, середню та сумарну абсолютну похибку має лінійна модель з включенням лише металів. Графіки передбачень за допомогою лінійної регресії для срібла можна побачити на рисунках 2.8-2.10.

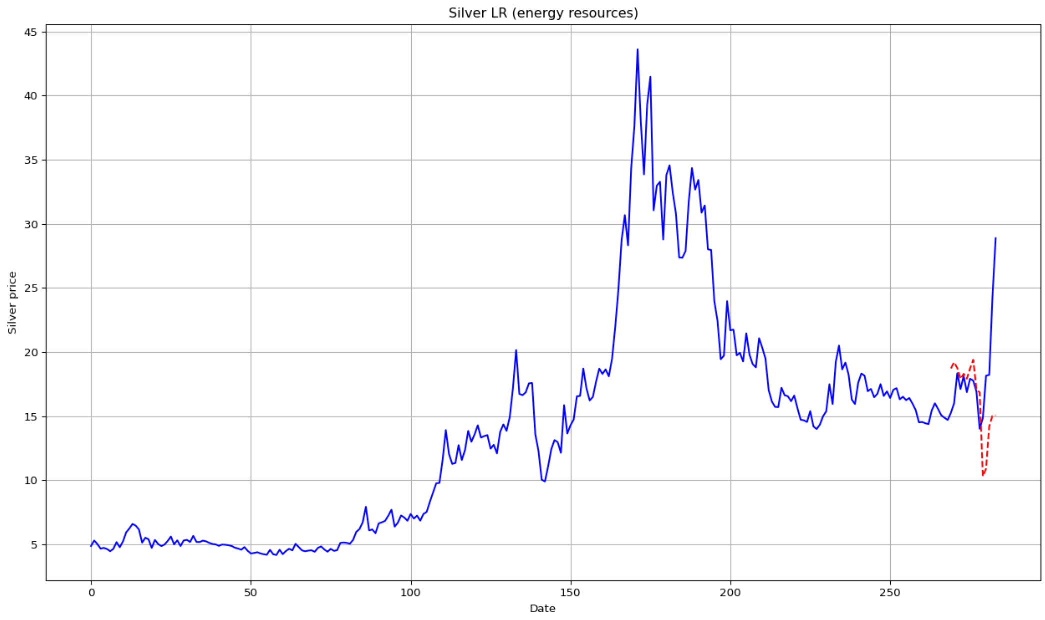

Рис. 2.8. Передбачення ціни срібла за допомогою лінійної регресії на основі енергетичних ресурсів

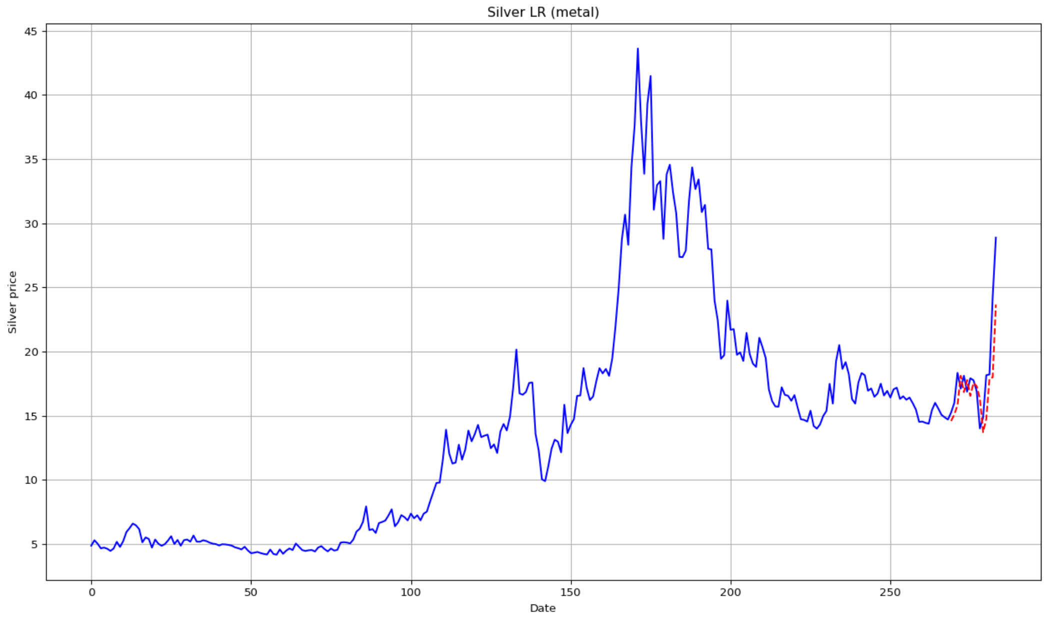

Рис. 2.9. Передбачення ціни срібла за допомогою лінійної регресії на основі металів

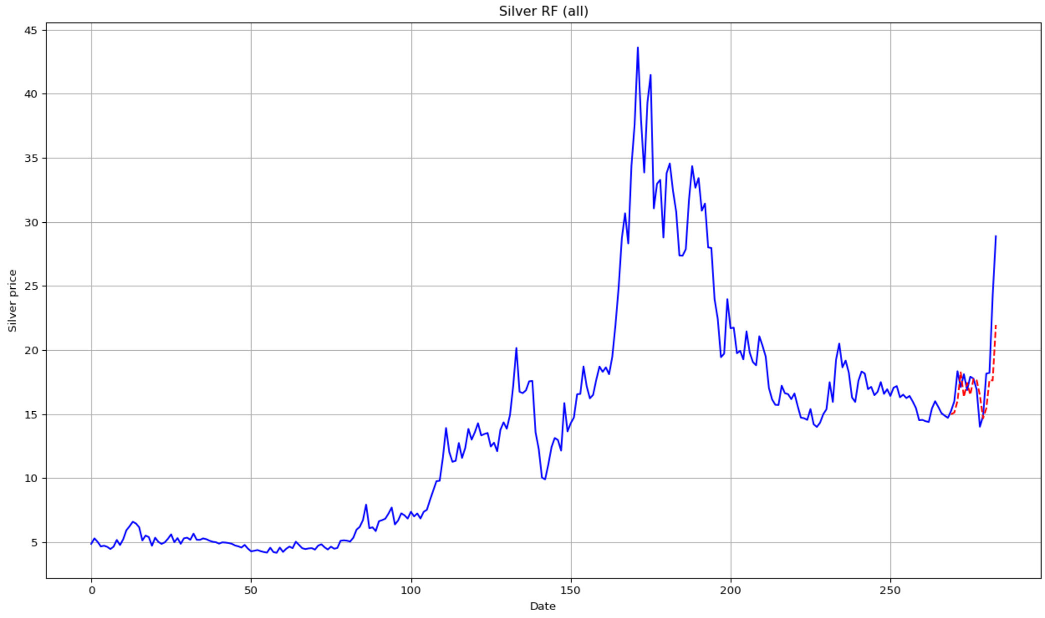

Рис. 2.10. Передбачення ціни срібла за допомогою лінійної регресії на основі енергетичних ресурсів та металів

2.3. Передбачення ціни срібла за допомогою моделі випадкового лісу

Для передбачення ціни срібла за допомогою моделі випадкового лісу було застосовано три можливих набори параметри. Перший, під назвою енергетичні ресурси, передбачає включення до моделі лінійної регресії цін на енергетичні носії (а саме, нафту, вугілля та газ) за попередній місяць. Другий, під назвою метали, передбачає включення до моделі лінійної регресії цін на дорогоцінні метали (а саме, срібло, золото, платину, паладій) за попередній місяць. Третій, під назвою все, передбачає включення до моделі випадкового лісу цін як на енергетичні ресурси, так і на дорогоцінні метали за попередній місяць. Оцінка ефективності моделей випадкового лісу, виражена у середній похибці, максимальному відхиленні, сумарній абсолютній похибці та коефіцієнті детермінації, відображена на рисунку 2.11.

	energy resources	metal	all
MEAN RESIDUAL	2.432457	1.936890	1.904597
MAX RESIDUAL	12.380100	7.319700	6.925700
SUM RESIDUAL	36.486850	29.053350	28.568950
DETERMINATION COEFFICIENT	0.028792	0.206404	0.229772

Рис. 2.11. Оцінка ефективності моделей випадкового лісу

За коефіцієнтом детермінації найкращий результат у моделі з енергетичними ресурсами, однак найменшу абсолютну, середню та сумарну абсолютну похибку має лінійна модель з включенням і металів, і енергетичних ресурсів. Графіки передбачень за допомогою моделі випадкового лісу для срібла можна побачити на рисунках 2.12-2.14.

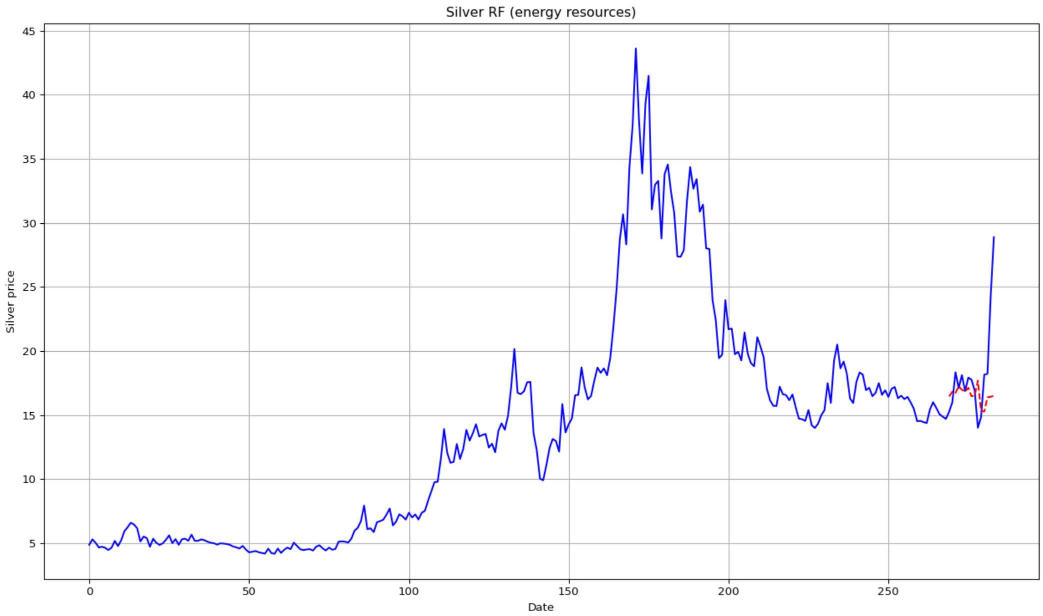

Рис. 2.12. Передбачення ціни срібла за допомогою моделі випадкового лісу на основі енергетичних ресурсів

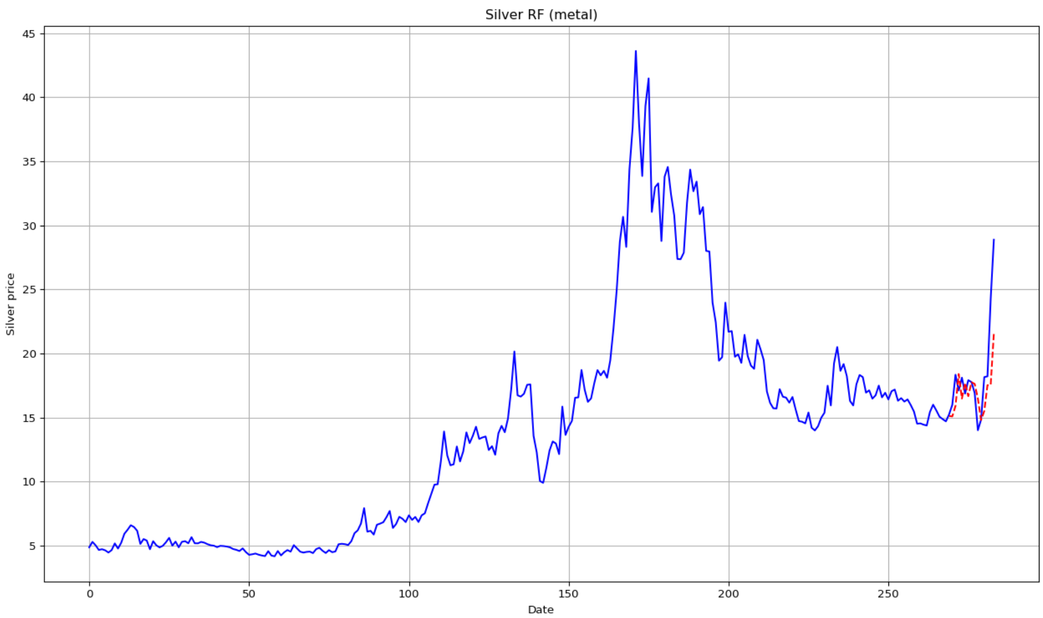

Рис. 2.13. Передбачення ціни срібла за допомогою моделі випадкового лісу на основі металів

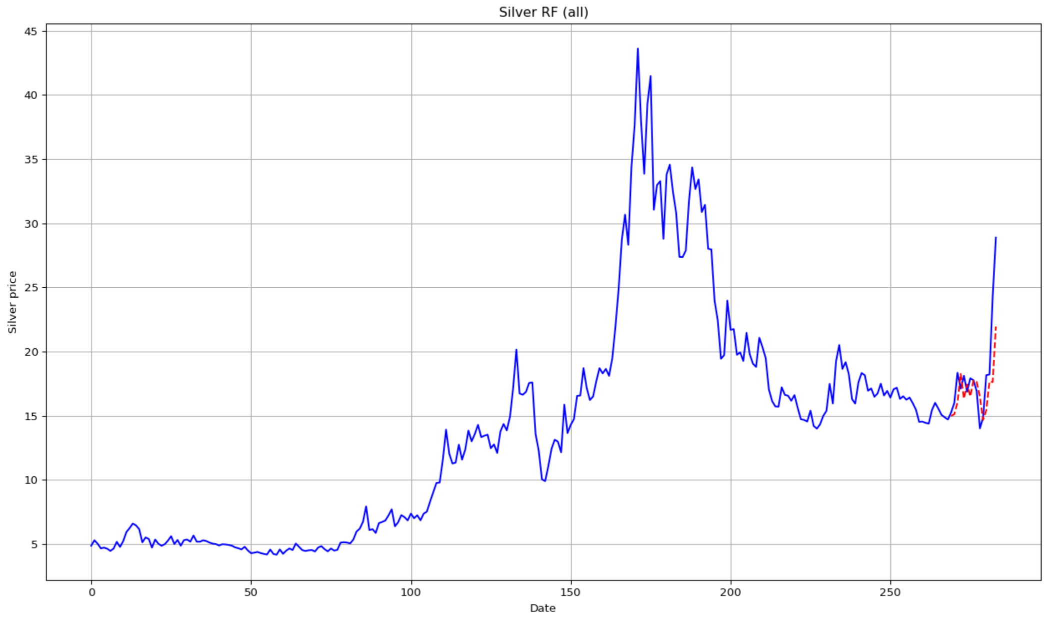

Рис. 2.14. Передбачення ціни срібла за допомогою моделі випадкового лісу на основі енергетичних ресурсів та металів

2.4. Аналіз результатів передбачення ціни срібла

Звернувши увагу на рисунки 2.2, 2.7, 2.11, ми можемо зробити висновок, що найкращою моделлю за коефіцієнтом детермінації є модель MA, де він найближчий до одиниці. Однак, за максимальним відхиленням найкращою є модель ARMA. За середнім відхиленням та сумарною абсолютною похибкою найкращою моделлю є модель ARIMA. Таким чином, моделі класу ARIMA виявились найкращими для передбачення щомісячної ціни срібла.

3. Передбачення ціни золота

3.1. Передбачення ціни золота за допомогою моделей класу ARIMA

Спочатку розглянемо передбачення ціни золота на основі моделей класу ARIMA. Програмою було здійснено підбір значень найкращих параметрів p, d, q методом підбору (від 1 до 5), де вибір кращої моделі здійснювався на основі критерію Акаїке. Результат такого підбору можна побачити на рисунку 3.1.

model	p	d	q	AIC
AR(2)	2.0	NaN	NaN	2842.615502
MA(5)	NaN	NaN	5.0	3220.668540
ARMA(3, 5)	3.0	NaN	5.0	2834.682441
ARIMA(3, 3, 5)	3.0	3.0	5.0	18.000000

Рис. 3.1. Значення найкращих параметрів p, d, q

Оцінка ефективності моделей, виражена у середній похибці, максимальному відхиленні, сумарній абсолютній похибці та коефіцієнті детермінації, відображена на рисунку 3.2.

	AR	MA	ARMA	ARIMA
MEAN RESIDUAL	64.795490	193.666600	63.287686	146.820133
MAX RESIDUAL	192.727566	360.889275	167.756509	1390.050000
SUM RESIDUAL	971.932351	2904.999005	949.315293	2202.302002
DETERMINATION COEFFICIENT	0.808697	0.777758	0.825852	5.571734

Рис. 3.2. Оцінка ефективності моделей ARIMA

Таким чином, найкращою за всіма показниками є модель ARMA. Графіки передбачень моделями класу ARIMA для золота можна побачити на рисунках 3.3-3.6.

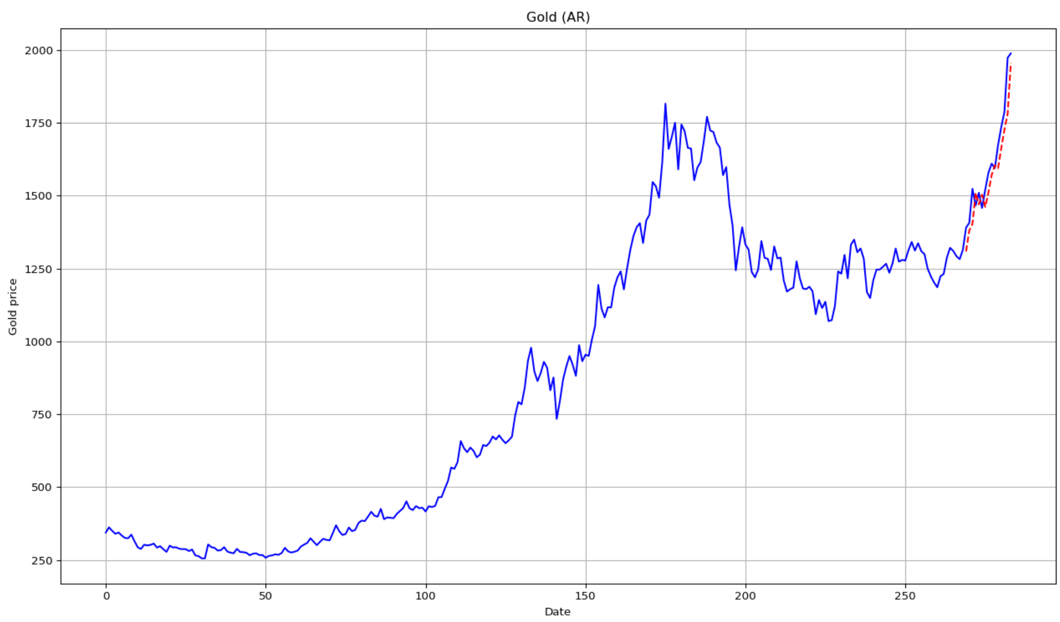

Рис. 3.3. Передбачення ціни золота за допомогою моделі AR

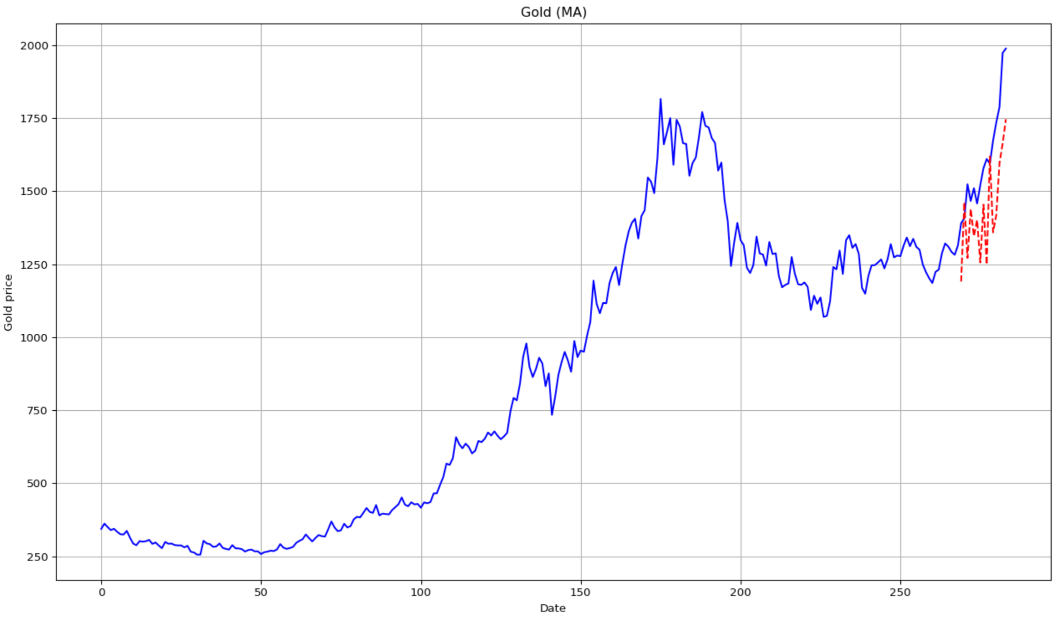

Рис. 3.4. Передбачення ціни золота за допомогою моделі MA

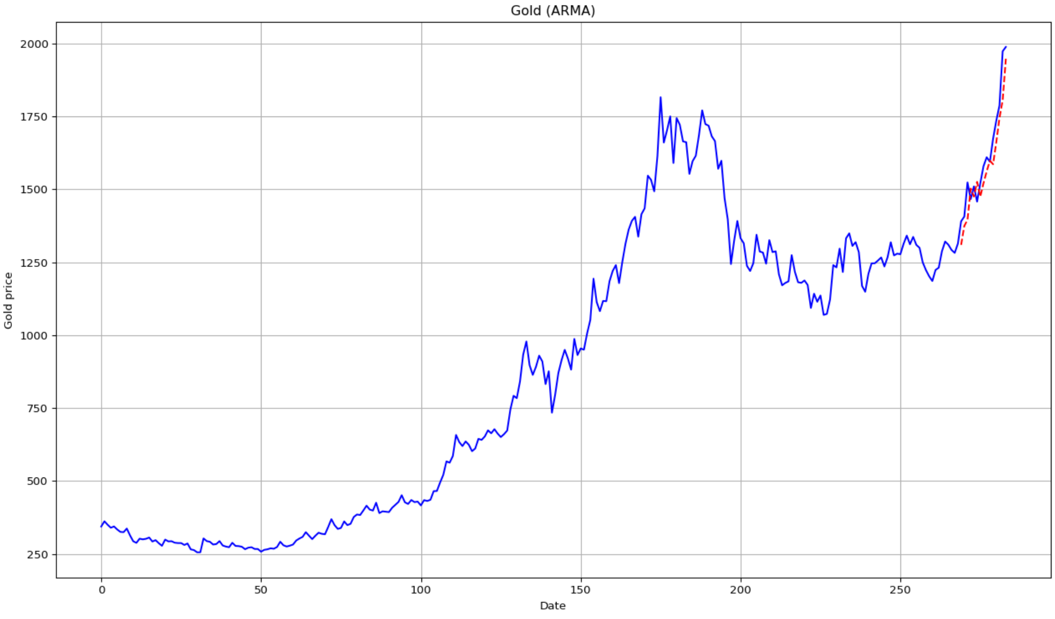

Рис. 3.5. Передбачення ціни золота за допомогою моделі ARMA

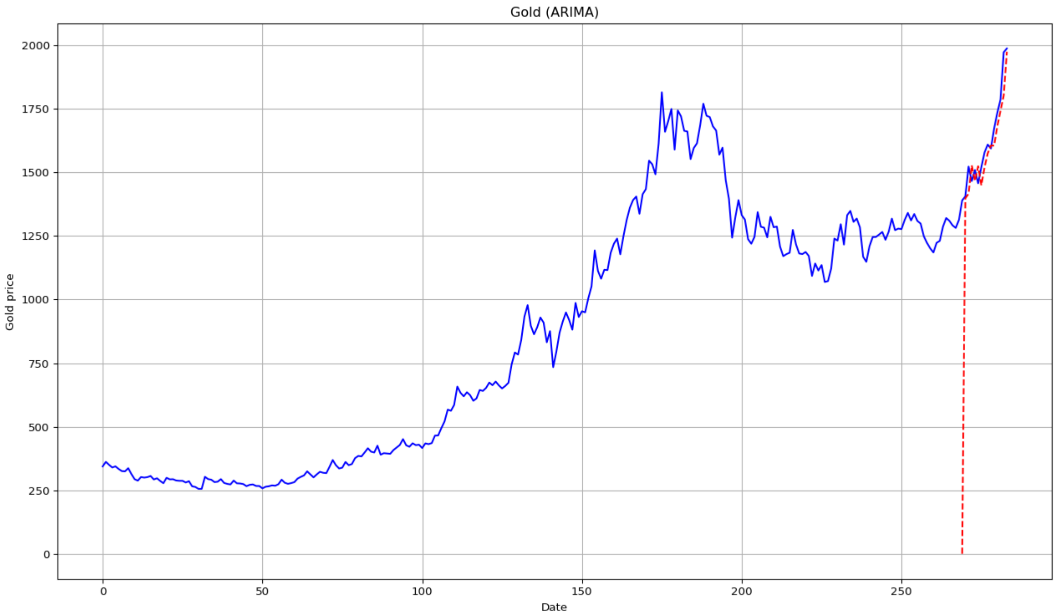

Рис. 3.6. Передбачення ціни золота за допомогою моделі ARIMA

3.2. Передбачення ціни золота за допомогою лінійної регресії

Для передбачення ціни золота за допомогою лінійної регресії було застосовано три можливих набори параметри. Перший, під назвою енергетичні ресурси, передбачає включення до моделі лінійної регресії цін на енергетичні носії (а саме, нафту, вугілля та газ) за попередній місяць. Другий, під назвою метали, передбачає включення до моделі лінійної регресії цін на дорогоцінні метали (а саме, срібло, золото, платину, паладій) за попередній місяць. Третій, під назвою все, передбачає включення до моделі лінійної регресії цін як на енергетичні ресурси, так і на дорогоцінні метали за попередній місяць. Оцінка ефективності моделей лінійної регресії, виражена у середній похибці, максимальному відхиленні, сумарній абсолютній похибці та коефіцієнті детермінації, відображена на рисунку 3.7.

	energy resources	metal	all
MEAN RESIDUAL	402.568228	68.495533	101.832176
MAX RESIDUAL	842.947268	198.277060	275.471225
SUM RESIDUAL	6038.523424	1027.432992	1527.482639
DETERMINATION COEFFICIENT	0.161553	0.811399	0.603431

Рис. 3.7. Оцінка ефективності моделей лінійної регресії

Найкращою за всіма показниками моделлю є модель зі включенням цін лише на дорогоцінні метали. Графіки передбачень за допомогою лінійної регресії для золота можна побачити на рисунках 3.8-3.10.

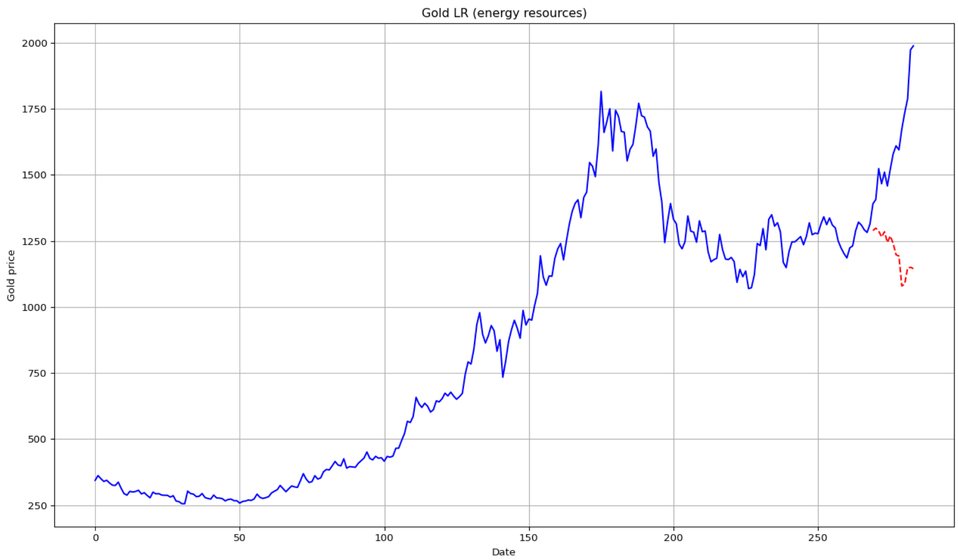

Рис. 3.8. Передбачення ціни золота за допомогою лінійної регресії на основі енергетичних ресурсів

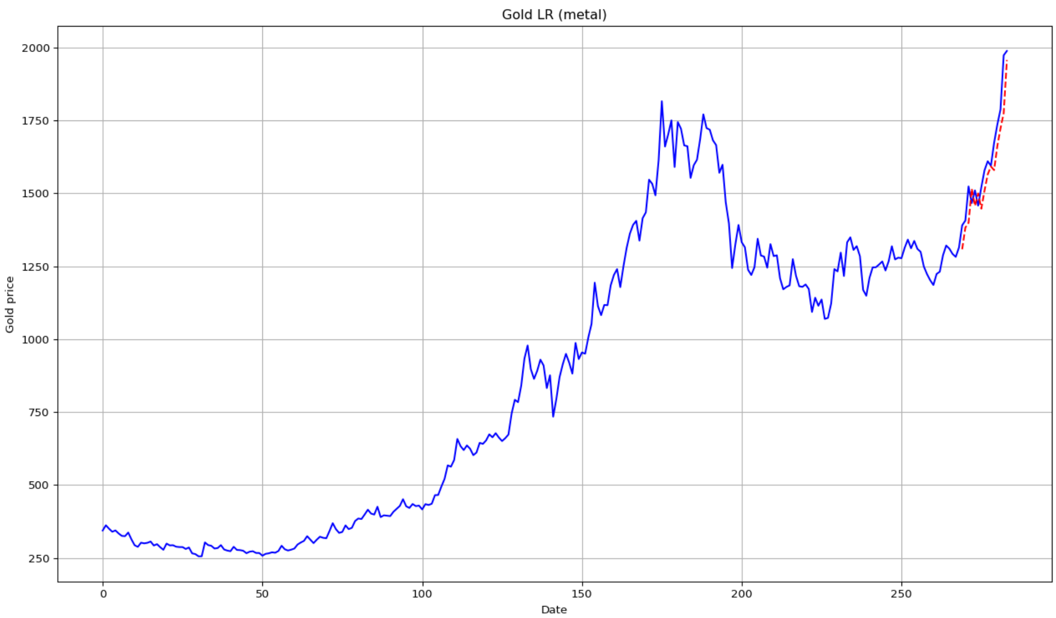

Рис. 3.9. Передбачення ціни золота за допомогою лінійної регресії на основі металів

Рис. 3.10. Передбачення ціни золота за допомогою лінійної регресії на основі енергетичних ресурсів та металів

3.3. Передбачення ціни золота за допомогою моделі випадкового лісу

Для передбачення ціни золота за допомогою моделі випадкового лісу було застосовано три можливих набори параметри.

Перший, під назвою енергетичні ресурси, передбачає включення до моделі лінійної регресії цін на енергетичні носії (а саме, нафту, вугілля та газ) за попередній місяць.

Другий, під назвою метали, передбачає включення до моделі лінійної регресії цін на дорогоцінні метали (а саме, срібло, золото, платину, паладій) за попередній місяць.

Третій, під назвою все, передбачає включення до моделі випадкового лісу цін як на енергетичні ресурси, так і на дорогоцінні метали за попередній місяць.

Оцінка ефективності моделей випадкового лісу, виражена у середній похибці, максимальному відхиленні, сумарній абсолютній похибці та коефіцієнті детермінації, відображена на рисунку 3.11.

	energy resources	metal	all
MEAN RESIDUAL	378.10410	169.701833	167.279800
MAX RESIDUAL	819.16200	441.459500	449.140000
SUM RESIDUAL	5671.56150	2545.527500	2509.197000
DETERMINATION COEFFICIENT	0.05184	0.222146	0.233984

Рис. 3.11. Оцінка ефективності моделей випадкового лісу

За коефіцієнтом детермінації, середнім відхиленням та сумарною абсолютною похибкою найкращий результат у моделі з енергетичними ресурсами та металами, однак найменшу максимальну похибку має лінійна модель з включенням металів. Графіки передбачень за допомогою моделі випадкового лісу для золота можна побачити на рисунках 3.12-3.14.

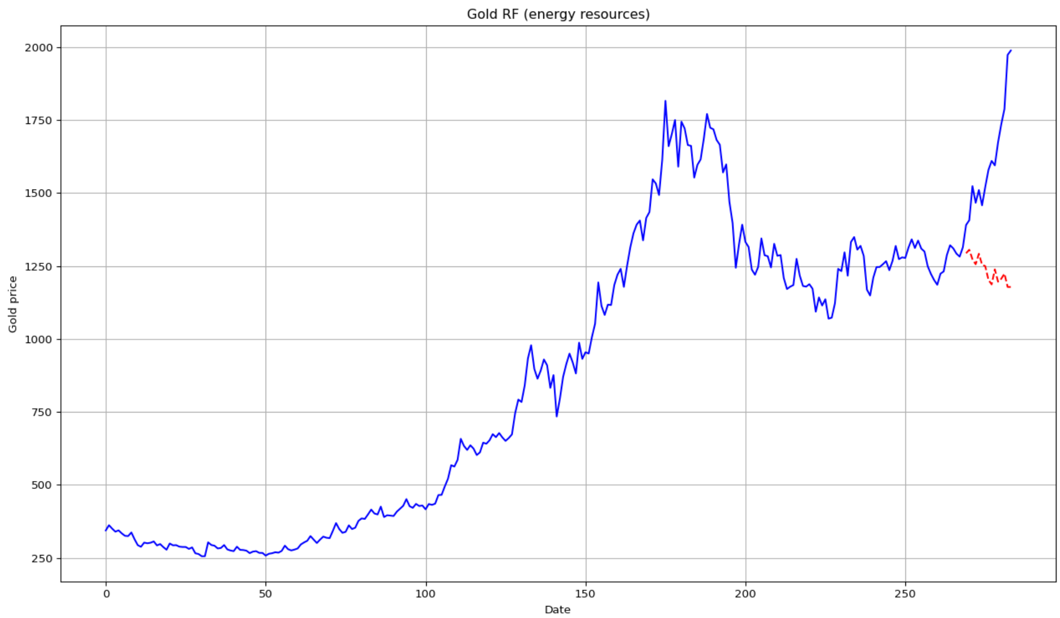

Рис. 3.12. Передбачення ціни золота за допомогою моделі випадкового лісу на основі енергетичних ресурсів

Рис. 3.13. Передбачення ціни золота за допомогою моделі випадкового лісу на основі металів

Рис. 3.14. Передбачення ціни золота за допомогою моделі випадкового лісу на основі енергетичних ресурсів та металів

3.4. Аналіз результатів передбачення ціни золота

Звернувши увагу на рисунки 3.2, 3.7, 3.11, ми можемо зробити висновок, що найкращою моделлю за всіма показниками є модель ARMA. Таким чином, саме моделі класу ARIMA виявились найкращими для передбачення щомісячної ціни золота.

4. Передбачення ціни платини

4.1. Передбачення ціни платини за допомогою моделей класу ARIMA

На початку розглянемо передбачення ціни платини на основі моделей класу ARIMA. Програмою було здійснено підбір значень найкращих параметрів p, d, q методом підбору (від 1 до 5), де вибір кращої моделі здійснювався на основі критерію Акаїке. Результат такого підбору можна побачити на рисунку 4.1.

	model	p	d	q	AIC
0	AR(2)	2.0	NaN	NaN	3109.456003
1	MA(5)	NaN	NaN	5.0	3275.696682
2	ARMA(3, 3)	3.0	NaN	3.0	3109.113720
3	ARIMA(2, 2, 5)	2.0	2.0	5.0	3088.856414

Рис. 4.1. Значення найкращих параметрів p, d, q

Оцінка ефективності моделей, виражена у середній похибці, максимальному відхиленні, сумарній абсолютній похибці та коефіцієнті детермінації, відображена на рисунку 4.2.

	AR	MA	ARMA	ARIMA
MEAN RESIDUAL	59.993897	80.342550	56.487441	61.176452
MAX RESIDUAL	142.327134	277.192018	139.417758	145.027466
SUM RESIDUAL	899.908454	1205.138256	847.311619	917.646779
DETERMINATION COEFFICIENT	1.092304	2.805190	1.067638	1.214448

Рис. 4.2. Оцінка ефективності моделей ARIMA

Таким чином, за всіма параметрами є найкращою модель ARMA. Графіки передбачень моделями класу ARIMA для платини можна побачити на рисунках 4.3-4.6.

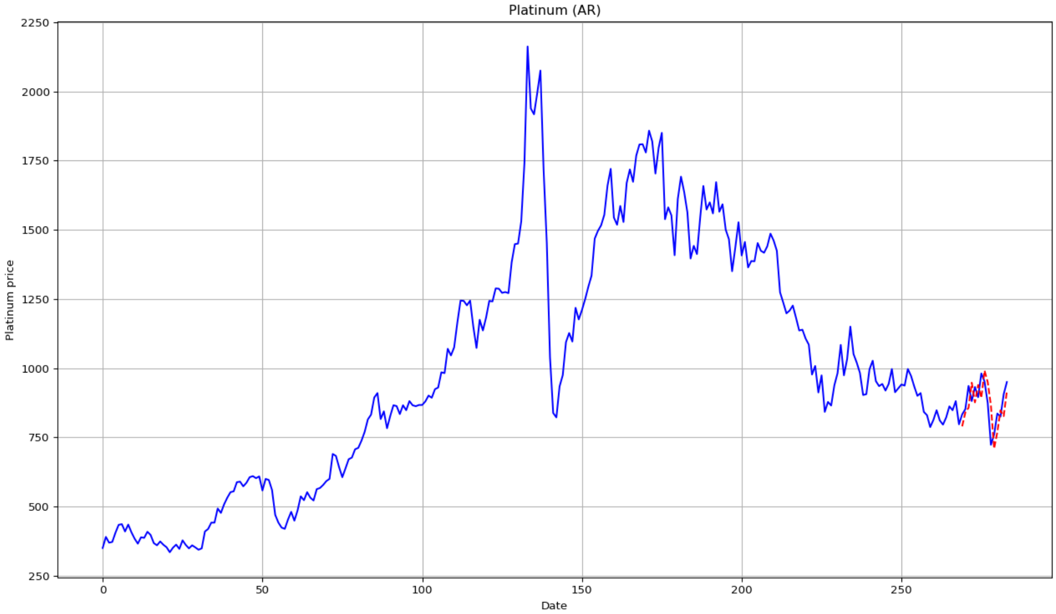

Рис. 4.3. Передбачення ціни платини за допомогою моделі AR

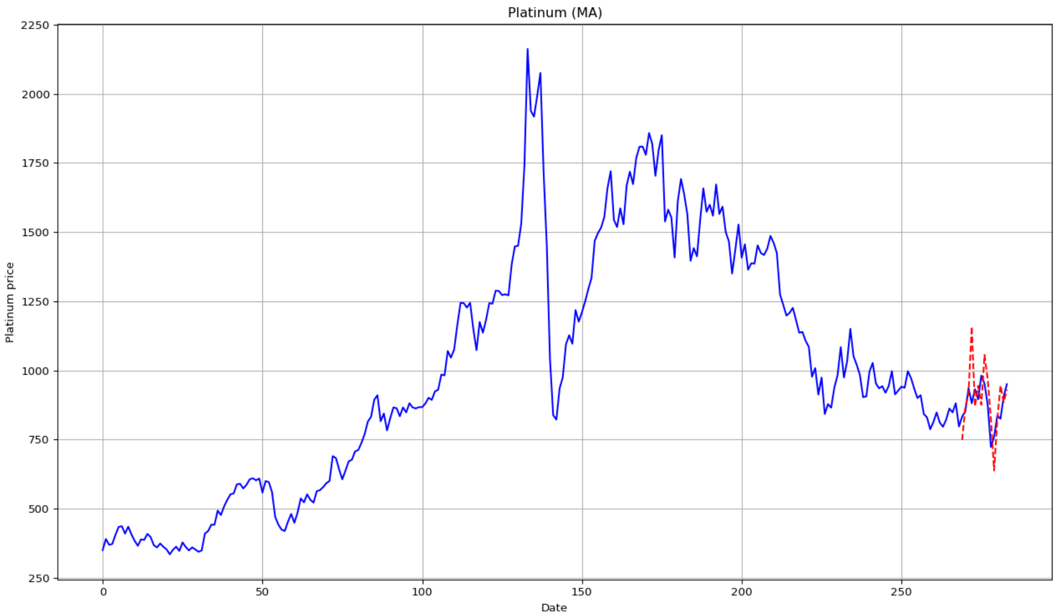

Рис. 4.4. Передбачення ціни платини за допомогою моделі MA

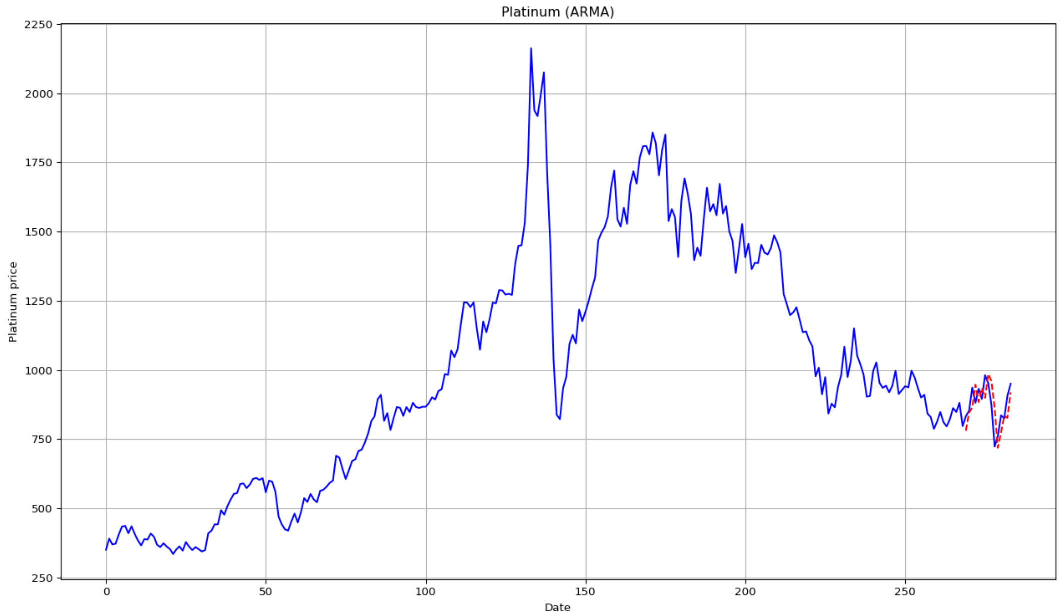

Рис. 4.5. Передбачення ціни платини за допомогою моделі ARMA

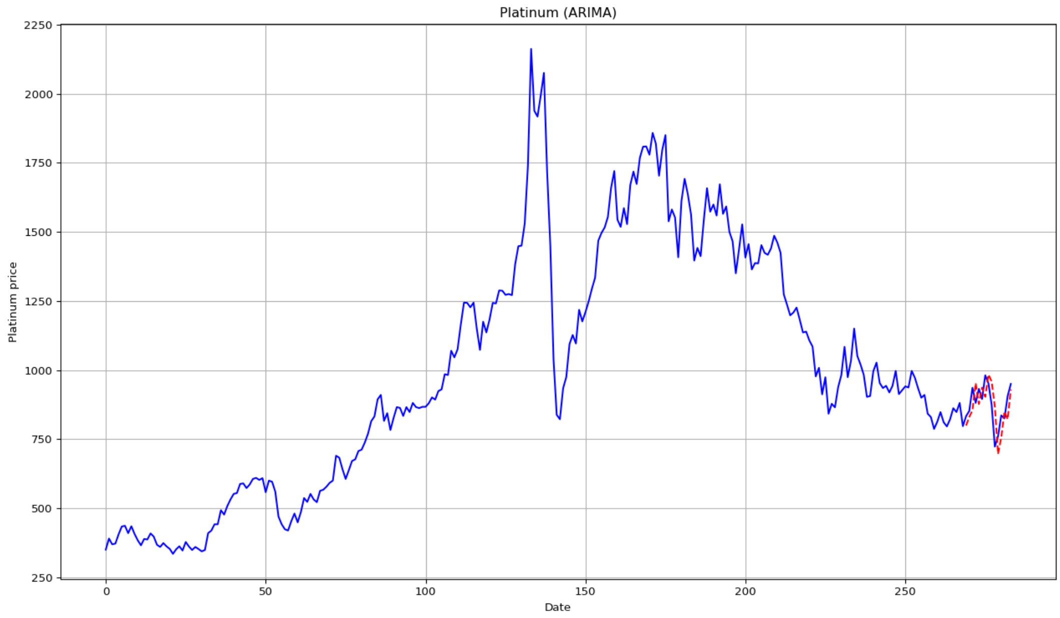

Рис. 4.6. Передбачення ціни платини за допомогою моделі ARIMA

4.2. Передбачення ціни платини за допомогою лінійної регресії

Для передбачення ціни платини за допомогою лінійної регресії було застосовано три можливих набори параметри.

Перший, під назвою енергетичні ресурси, передбачає включення до моделі лінійної регресії цін на енергетичні носії (а саме, нафту, вугілля та газ) за попередній місяць.

Другий, під назвою метали, передбачає включення до моделі лінійної регресії цін на дорогоцінні метали (а саме, срібло, золото, платину, паладій) за попередній місяць.

Третій, під назвою все, передбачає включення до моделі лінійної регресії цін як на енергетичні ресурси, так і на дорогоцінні метали за попередній місяць.

Оцінка ефективності моделей лінійної регресії, виражена у середній похибці, максимальному відхиленні, сумарній абсолютній похибці та коефіцієнті детермінації, відображена на рисунку 4.7.

	energy resources	metal	all
MEAN RESIDUAL	144.549063	64.990630	58.164669
MAX RESIDUAL	247.606338	126.351389	120.076925
SUM RESIDUAL	2168.235945	974.859445	872.470030
DETERMINATION COEFFICIENT	5.508401	0.928066	0.817476

Рис. 4.7. Оцінка ефективності моделей лінійної регресії

За коефіцієнтом детермінації найкращий результат у моделі лише з металами, однак найменшу абсолютну, середню та сумарну абсолютну похибку має лінійна модель з включенням енергетичних ресурсів та металів. Графіки передбачень за допомогою лінійної регресії для платини можна побачити на рисунках 4.8-4.10.

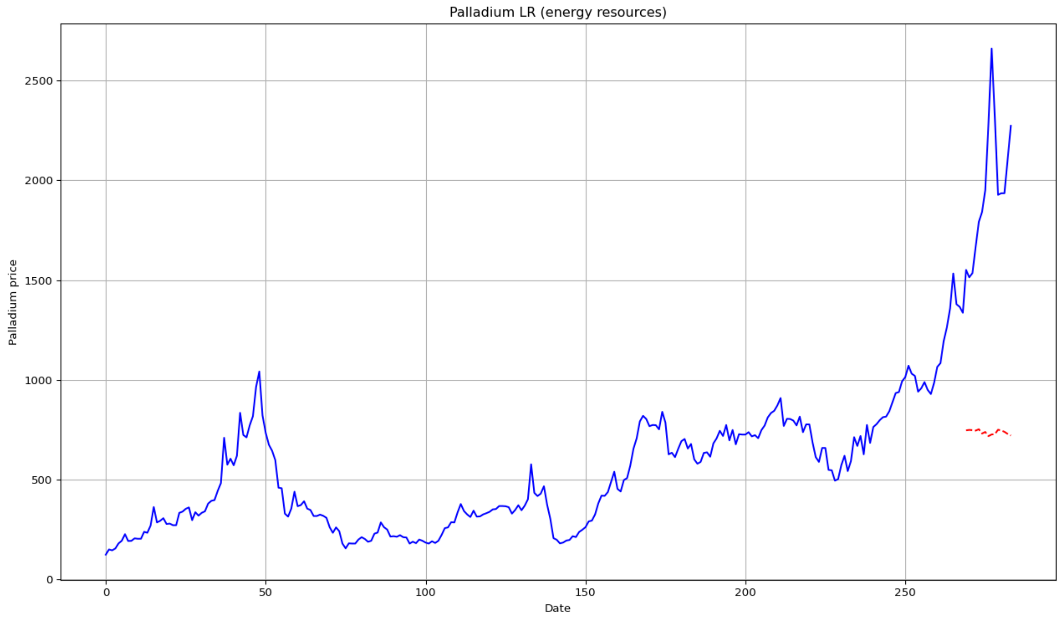

Рис. 4.8. Передбачення ціни платини за допомогою лінійної регресії на основі енергетичних ресурсів

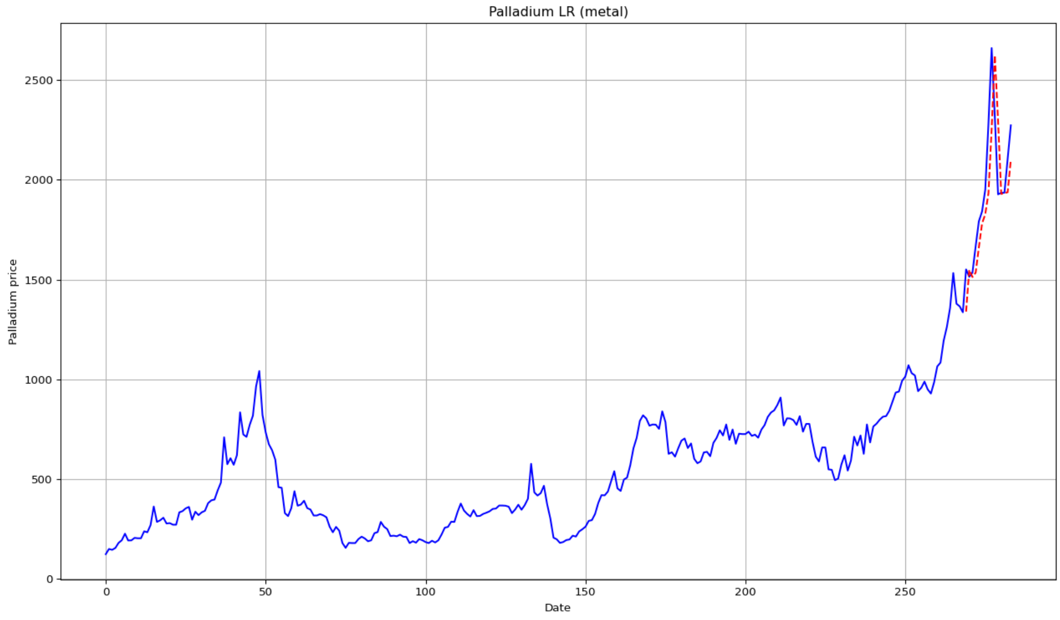

Рис. 4.9. Передбачення ціни платини за допомогою лінійної регресії на основі металів

Рис. 4.10. Передбачення ціни платини за допомогою лінійної регресії на основі енергетичних ресурсів та металів

4.3. Передбачення ціни платини за допомогою моделі випадкового лісу

Для передбачення ціни платини за допомогою моделі випадкового лісу було застосовано три можливих набори параметри.

Перший, під назвою енергетичні ресурси, передбачає включення до моделі лінійної регресії цін на енергетичні носії (а саме, нафту, вугілля та газ) за попередній місяць.

Другий, під назвою метали, передбачає включення до моделі лінійної регресії цін на дорогоцінні метали (а саме, срібло, золото, платину, паладій) за попередній місяць.

Третій, під назвою все, передбачає включення до моделі випадкового лісу цін як на енергетичні ресурси, так і на дорогоцінні метали за попередній місяць.

Оцінка ефективності моделей випадкового лісу, виражена у середній похибці, максимальному відхиленні, сумарній абсолютній похибці та коефіцієнті детермінації, відображена на рисунку 4.11.

	energy resources	metal	all
MEAN RESIDUAL	122.758667	53.384667	49.543667
MAX RESIDUAL	311.500000	141.410000	145.985000
SUM RESIDUAL	1841.380000	800.770000	743.155000
DETERMINATION COEFFICIENT	1.301567	0.611805	0.404144

Рис. 4.11. Оцінка ефективності моделей випадкового лісу

За коефіцієнтом детермінації найкращий результат у моделі з енергетичними ресурсами, однак найменшу абсолютну, середню та сумарну абсолютну похибку має лінійна модель з включенням металів та енергетичних ресурсів. Графіки передбачень за допомогою моделі випадкового лісу для срібла можна побачити на рисунках 4.12-4.14.

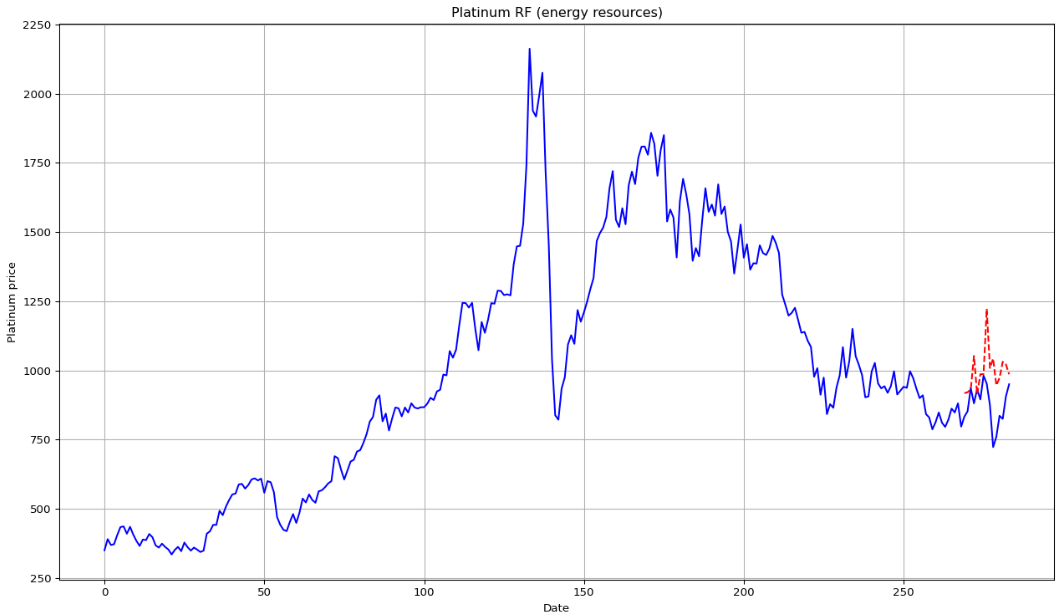

Рис. 4.12. Передбачення ціни платини за допомогою моделі випадкового лісу на основі енергетичних ресурсів

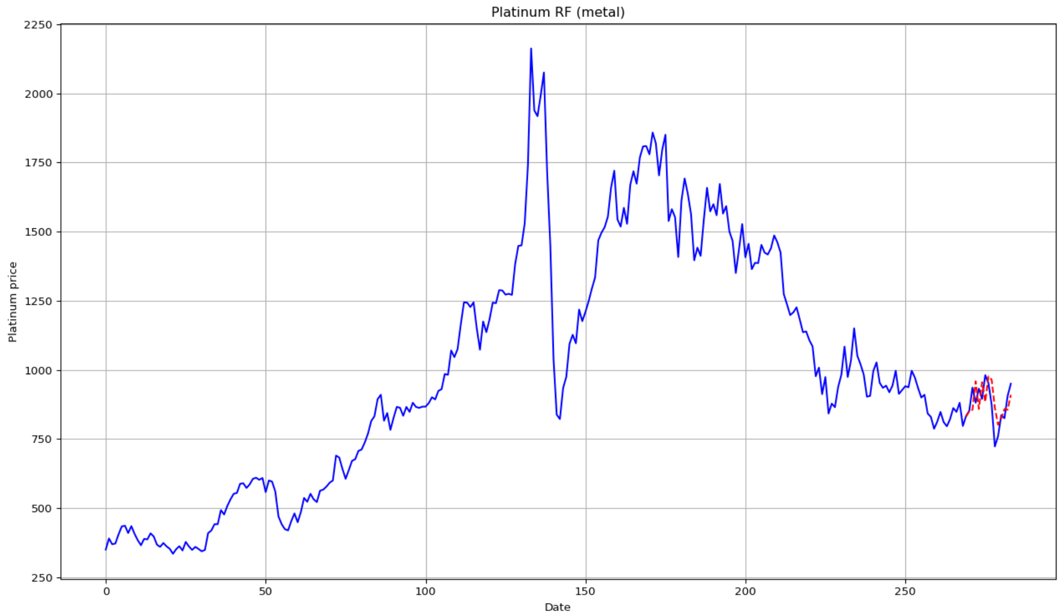

Рис. 4.13. Передбачення ціни платини за допомогою моделі випадкового лісу на основі металів

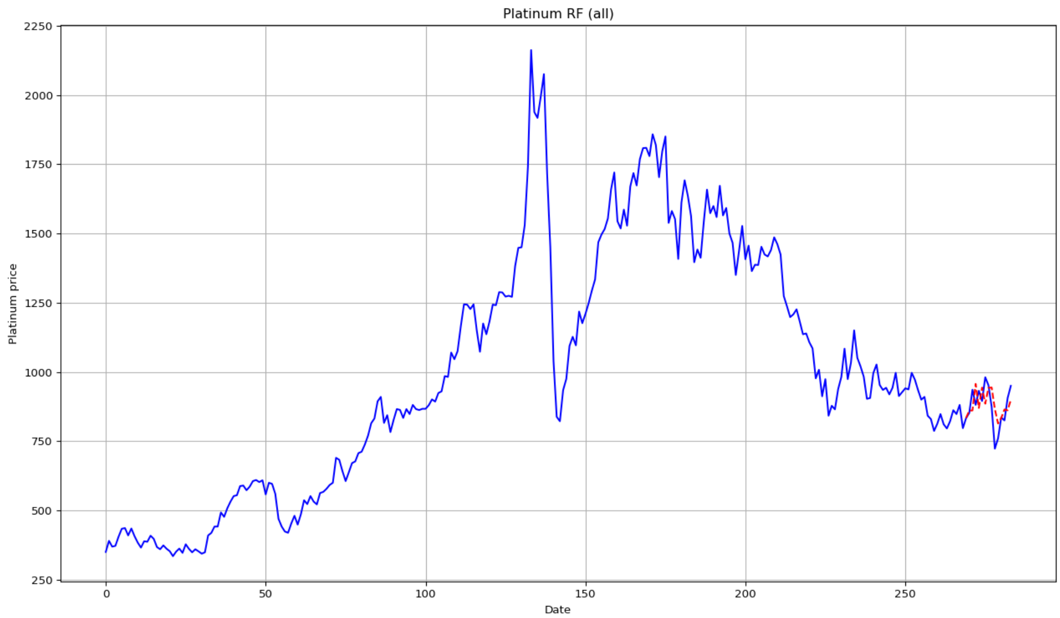

Рис. 4.14. Передбачення ціни платини за допомогою моделі випадкового лісу на основі енергетичних ресурсів та металів

4.4. Аналіз результатів передбачення ціни платини

Звернувши увагу на рисунки 4.2, 4.7, 4.11, ми можемо зробити висновок, що найкращою моделлю за коефіцієнтом детермінації є модель ARMA. За показником максимального відхилення, найкращою моделлю виявилась модель лінійної регресії з урахуванням цін на енергетичні ресурси та метали. За середньою та сумарною абсолютною похибками найкращою виявилась модель випадкового лісу з урахуванням цін на енергетичні ресурси та метали. Таким чином, для прогнозування ціни платини непогано показали себе всі моделі.

5. Передбачення ціни паладію

5.1. Передбачення ціни паладію за допомогою моделей класу ARIMA

Спершу розглянемо передбачення ціни паладію на основі моделей класу ARIMA. Програмою було здійснено підбір значень найкращих параметрів p, d, q методом підбору (від 1 до 5), де вибір кращої моделі здійснювався на основі критерію Акаїке. Результат такого підбору можна побачити на рисунку 5.1.

model	p	d	q	AIC
AR(1)	1.0	NaN	NaN	2932.247260
MA(4)	NaN	NaN	4.0	3141.761852
ARMA(3, 2)	3.0	NaN	2.0	2929.156225
ARIMA(2, 2, 3)	2.0	2.0	3.0	2905.908475

Рис. 5.1. Значення найкращих параметрів p, d, q

Оцінка ефективності моделей, виражена у середній похибці, максимальному відхиленні, сумарній абсолютній похибці та коефіцієнті детермінації, відображена на рисунку 5.2.

	AR	MA	ARMA	ARIMA
MEAN RESIDUAL	168.556382	366.500334	181.498459	186.254061
MAX RESIDUAL	383.092420	853.514105	413.317384	486.319639
SUM RESIDUAL	2528.345734	5497.505008	2722.476883	2793.810910
DETERMINATION COEFFICIENT	1.149974	1.267504	1.185541	1.380924

Рис. 5.2. Оцінка ефективності моделей ARIMA

Таким чином, найкращою за всіма параметрами є модель AR. Графіки передбачень моделями класу ARIMA для паладію можна побачити на рисунках 5.3-5.6.

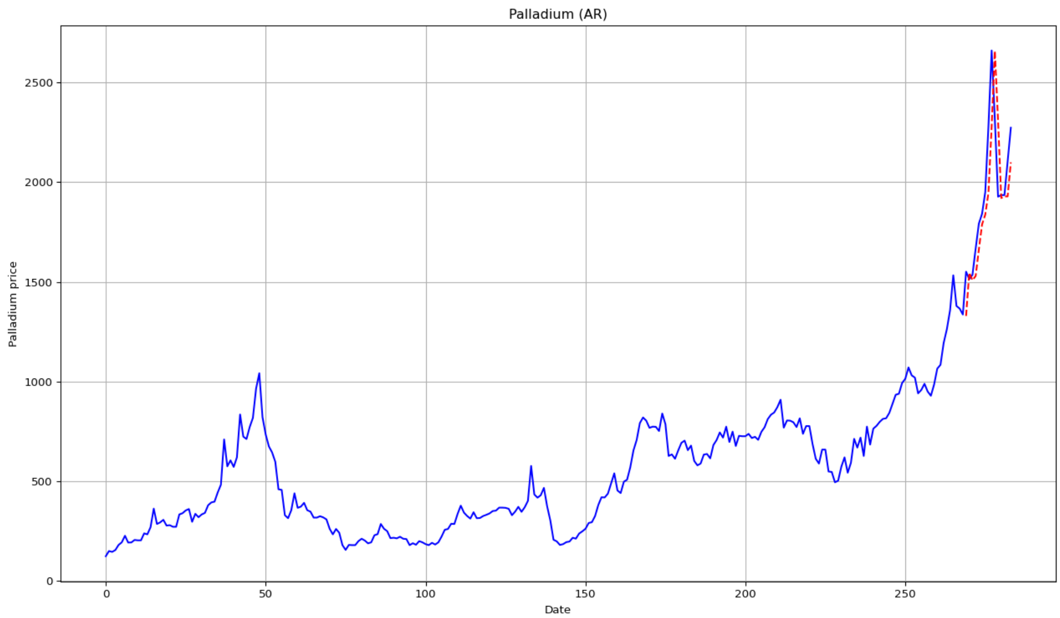

Рис. 5.3. Передбачення ціни паладію за допомогою моделі AR

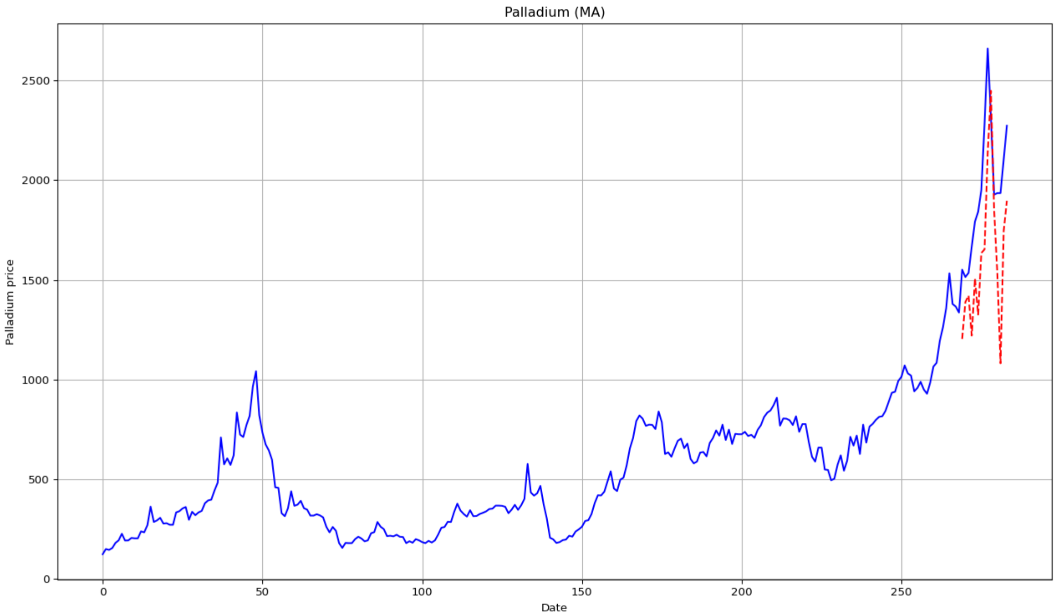

Рис. 5.4. Передбачення ціни паладію за допомогою моделі MA

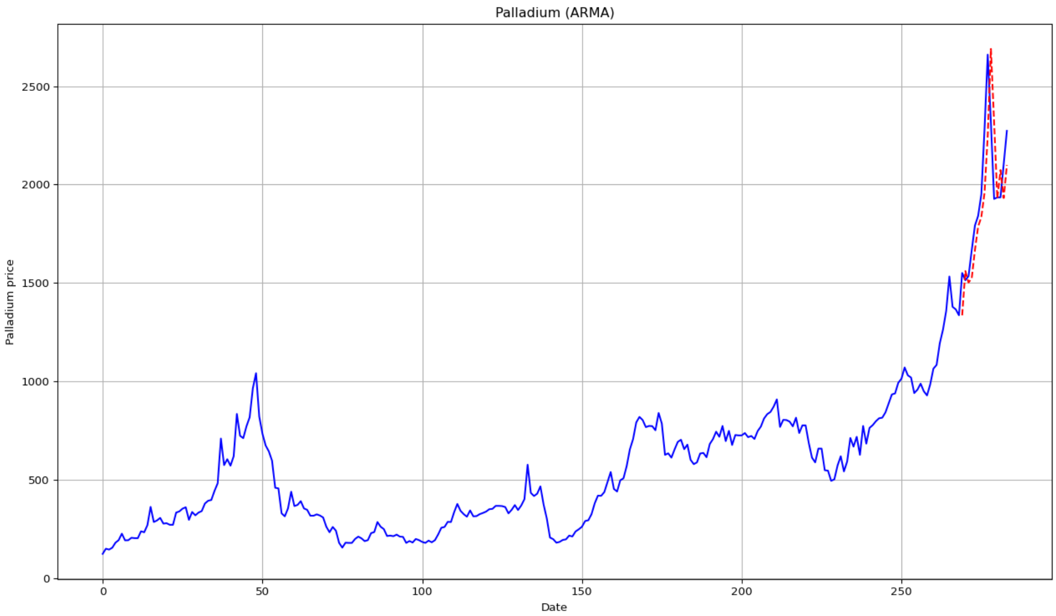

Рис. 5.5. Передбачення ціни паладію за допомогою моделі ARMA

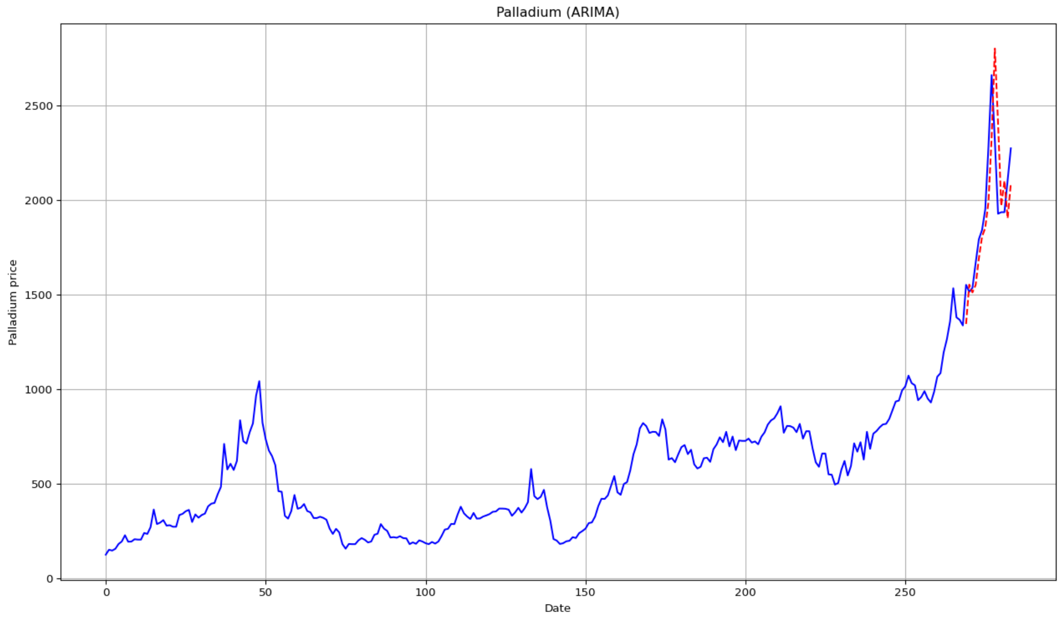

Рис. 5.6. Передбачення ціни паладію за допомогою моделі ARIMA

5.2. Передбачення ціни паладію за допомогою лінійної регресії

Для передбачення ціни паладію за допомогою лінійної регресії було застосовано три можливих набори параметри.

Перший, під назвою енергетичні ресурси, передбачає включення до моделі лінійної регресії цін на енергетичні носії (а саме, нафту, вугілля та газ) за попередній місяць.

Другий, під назвою метали, передбачає включення до моделі лінійної регресії цін на дорогоцінні метали (а саме, срібло, золото, -здбплатину, па-адій) за попередній місяць.

Третій, під назвою все, передбачає включення до моделі лінійної регресії цін як на енергетичні ресурси, так і на дорогоцінні метали за попередній місяць.

Оцінка ефективності моделей лінійної регресії, виражена у середній похибці, максимальному відхиленні, сумарній абсолютній похибці та коефіцієнті детермінації, відображена на рисунку 5.7.

	energy resources	metal	all
MEAN RESIDUAL	1213.704357	166.315247	162.582082
MAX RESIDUAL	1933.471729	405.423922	394.052289
SUM RESIDUAL	18205.565353	2494.728698	2438.731233
DETERMINATION COEFFICIENT	0.001225	1.085981	1.112068

Рис. 5.7. Оцінка ефективності моделей лінійної регресії

За коефіцієнтом детермінації найкращий результат у моделі з металами, однак найменшу абсолютну, середню та сумарну абсолютну похибку має лінійна модель з включенням енергетичних ресурсів та металів. Графіки передбачень за допомогою лінійної регресії для срібла можна побачити на рисунках 5.8-5.10.

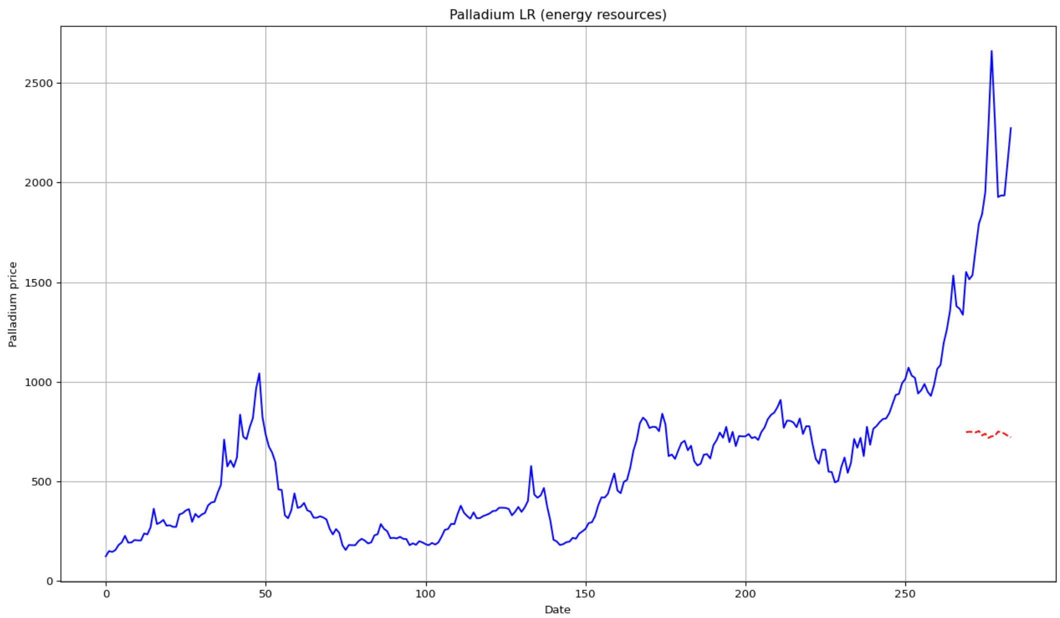

Рис. 5.8. Передбачення ціни паладію за допомогою лінійної регресії на основі енергетичних ресурсів

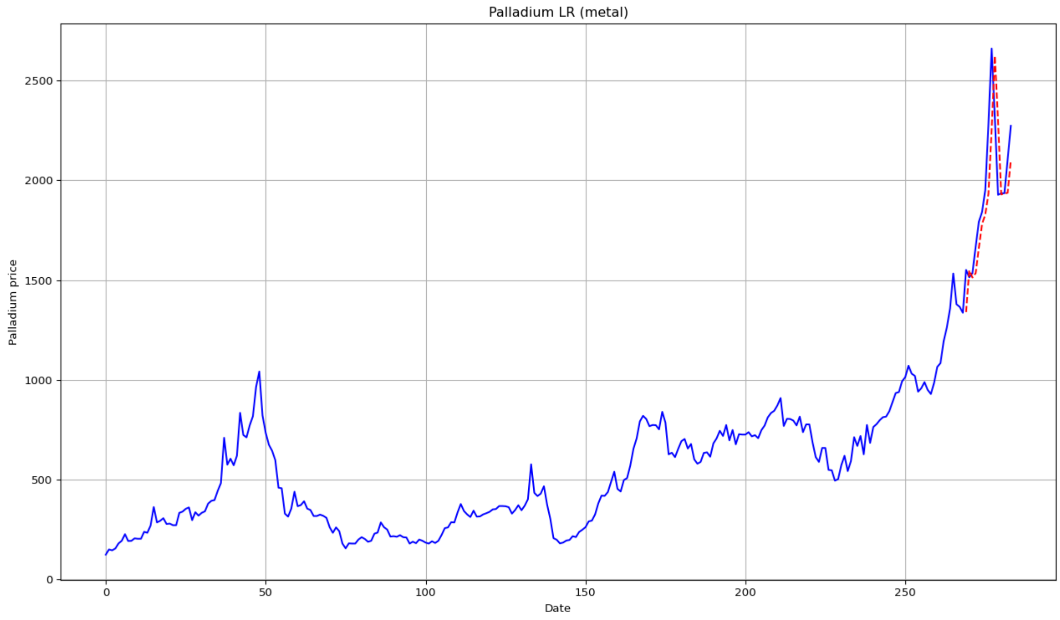

Рис. 5.9. Передбачення ціни паладію за допомогою лінійної регресії на основі металів

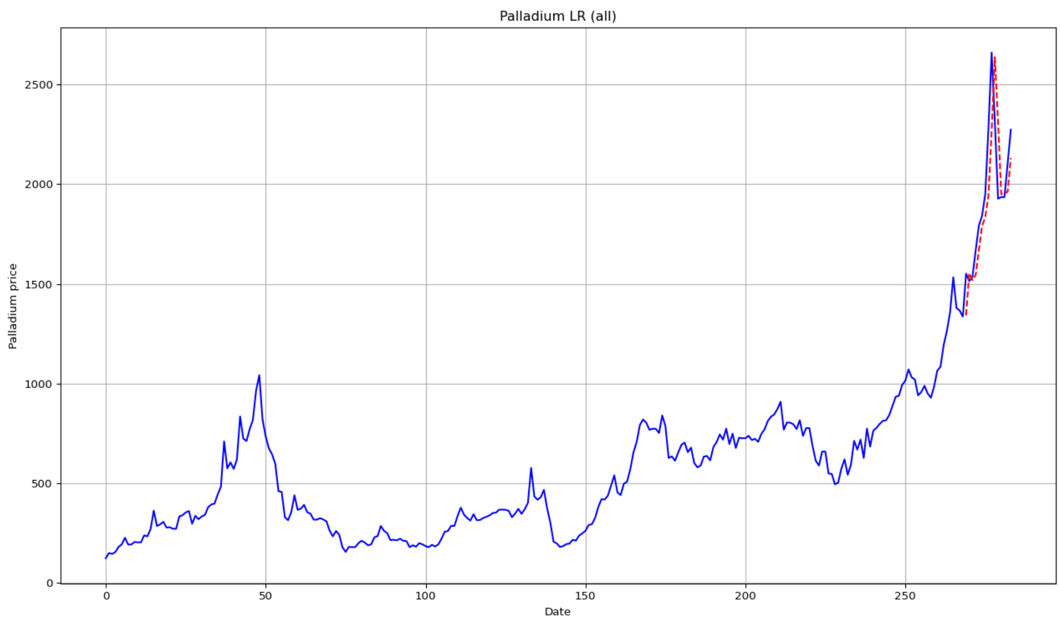

Рис. 5.10. Передбачення ціни паладію за допомогою лінійної регресії на основі енергетичних ресурсів та металів

5.3. Передбачення ціни паладію за допомогою моделі випадкового лісу

Для передбачення ціни паладію за допомогою моделі випадкового лісу було застосовано три можливих набори параметри.

Перший, під назвою енергетичні ресурси, передбачає включення до моделі лінійної регресії цін на енергетичні носії (а саме, нафту, вугілля та газ) за попередній місяць.

Другий, під назвою метали, передбачає включення до моделі лінійної регресії цін на дорогоцінні метали (а саме, срібло, золото, платину, паладій) за попередній місяць.

Третій, під назвою все, передбачає включення до моделі випадкового лісу цін як на енергетичні ресурси, так і на дорогоцінні метали за попередній місяць.

Оцінка ефективності моделей випадкового лісу, виражена у середній похибці, максимальному відхиленні, сумарній абсолютній похибці та коефіцієнті детермінації, відображена на рисунку 5.11.

	energy resources	metal	all
MEAN RESIDUAL	1148.275667	512.394667	517.860000
MAX RESIDUAL	2078.380000	1203.330000	1218.920000
SUM RESIDUAL	17224.135000	7685.920000	7767.900000
DETERMINATION COEFFICIENT	0.666677	0.010273	0.003322

Рис. 5.11. Оцінка ефективності моделей випадкового лісу

За коефіцієнтом детермінації найкращий результат у моделі з енергетичними ресурсами, однак найменшу абсолютну, середню та сумарну абсолютну похибку має лінійна модель з включенням лише металів. Графіки передбачень за допомогою моделі випадкового лісу для паладію можна побачити на рисунках 5.12-5.14.

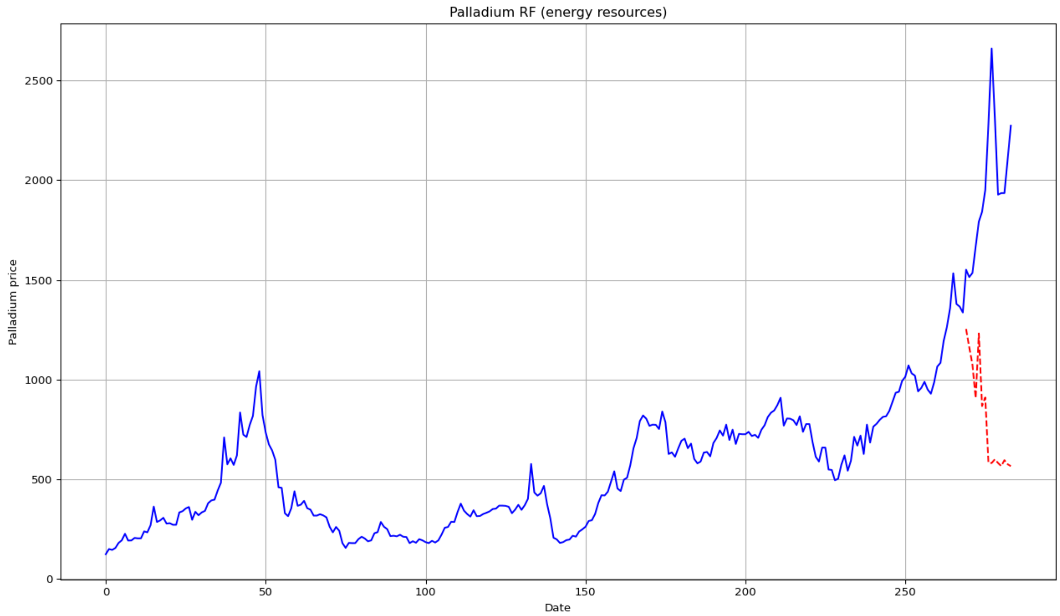

Рис. 5.12. Передбачення ціни паладію за допомогою моделі випадкового лісу на основі енергетичних ресурсів

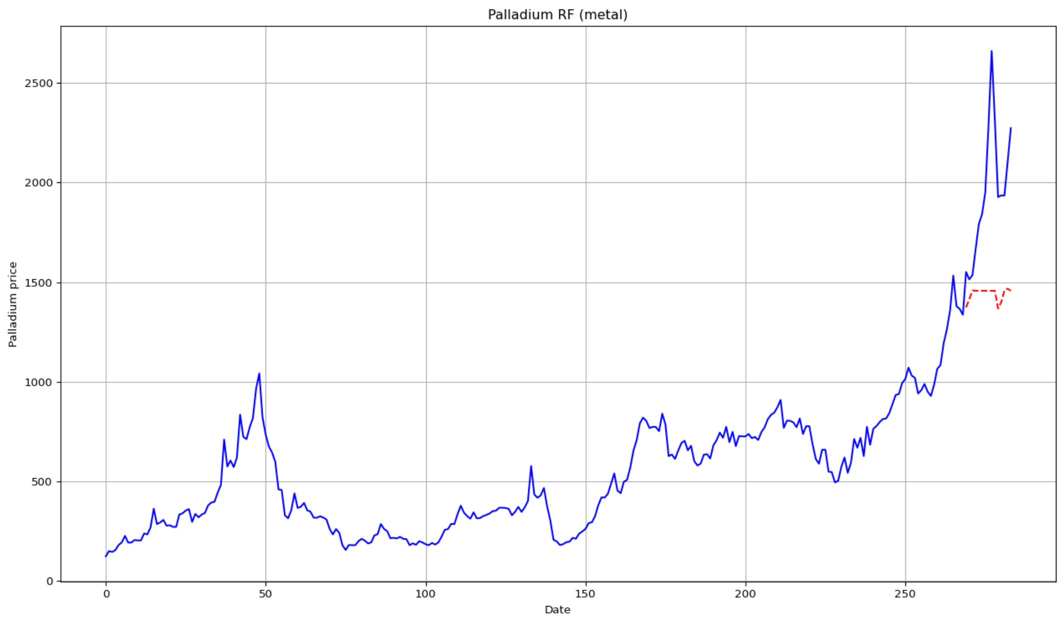

Рис. 5.13. Передбачення ціни паладію за допомогою моделі випадкового лісу на основі металів

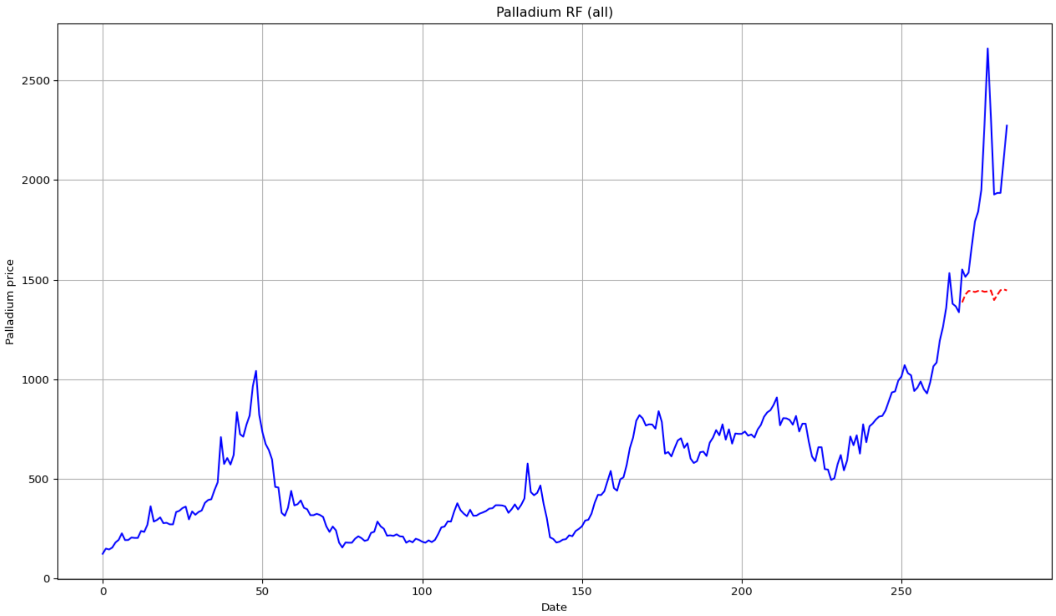

Рис. 5.14. Передбачення ціни паладію за допомогою моделі випадкового лісу на основі енергетичних ресурсів та металів

5.4. Аналіз результатів передбачення ціни паладію

Звернувши увагу на рисунки 5.2, 5.7, 5.11, ми можемо зробити висновок, що найкращою моделлю за коефіцієнтом детермінації є модель лінійної регресії з металами, однак найменшу абсолютну, середню та сумарну абсолютну похибку має модель лінійної регресії з включенням енергетичних ресурсів та металів. Таким чином, моделі класу лінійної регресії виявились найкращими для передбачення щомісячної ціни паладію.

ВИСНОВКИ

У цій роботі нами було проведено передбачення цін на такі дорогоцінні метали, як срібло, золото, платина, паладій. З метою передбачення були використані авторегресійні моделі ARIMA, моделі лінійної регресії та моделі випадкового лісу. Для кожної з моделей підбирались різні варіанти параметрів.

В результаті оцінки проведення передбачення для кожної з моделей, ми отримали наступні дані:

- для срібла найкраще будує прогноз моделі класу ARIMA: MA, ARMA, ARIMA;
- для золота найкраще будує прогноз моделі класу ARIMA: ARMA;
- для платини достатньо добре показали себе всі моделі, проте виявити явного лідера не вдалось;
- для паладію найкращими виявились моделі класу лінійної регресії.

Такі дані свідчать не тільки про ефективність тих, чи інших моделей для прогнозування ціни певного металу, а й про зв’язок цін конкретного металу з цінами на енергоресурси. Так, цей зв’язок для срібла та золота є незначним, для платини його можна вважати середнім, тоді як для паладію він є достатньо сильним.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Petropoulos, F. (2017). Forecasting with temporal hierarchies. European Journal of Operational Research, 262(1), 60–74.

2. Brockwell, P. J., & Davis, R. A. (2016). Introduction to time series and forecasting (3rd ed). New York, USA: Springer

3. Beheshti.N.(2022) Random Forest Regression

4. Prabhakaran.S.(2022).ARIMA Model – Complete Guide to Time Series Forecasting in Python

5. Tsao, Min (2022). "Group least squares regression for linear models with strongly correlated predictor variables". Annals of the Institute of Statistical Mathematics. 75 (2): 233–250

6. Sanders, N., Goodwin, P., Önkal, D., Gönül, M. S., Harvey, N., Lee, A., & Kjolso, L. (2005). When and how should statistical forecasts be judgmentally adjusted? Foresight: The International Journal of Applied Forecasting, 1(1), 5–23.

7. Wickham, H. (2016). Elegant graphics for data analysis (2nd ed). Springer.