Національний технічний університет України «КПІ»

Кафедра автоматизованих систем обробки інформації і управління

Лабораторна робота №7

з дисципліни «[Аналіз даних в інформаційних управляючих системах](https://do.ipo.kpi.ua/course/view.php?id=5470)»

на тему: «МОДЕЛЬ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ»

Виконав:

студент групи ІС-23

Шимків М.В.

Викладач:

Гавриленко О.В

Київ 2024

**Мета роботи**: вивчити методи прогнозування за допомогою числових рядів

**Завдання до роботи**

1. Ознайомитися з конспектом лекцій та рекомендованою літературою за темою роботи, атакож теоретичними відомостями, що містять опис програмного забезпечення для побудови часових рядів.

2. Сформувати набір даних для обробки та аналізу.

3. На основі обраних даних мовою Python виконати прогнозування за допомогою числових рядів на найближчий тиждень за допомогою алгоритмів ARIMA, ARMA, ковзного середнього та експоненційного згладжування.

4. Порівняти точність даних прогнозів.

5. Оформити звіт з роботи.

6. Відповісти на контрольні питання.

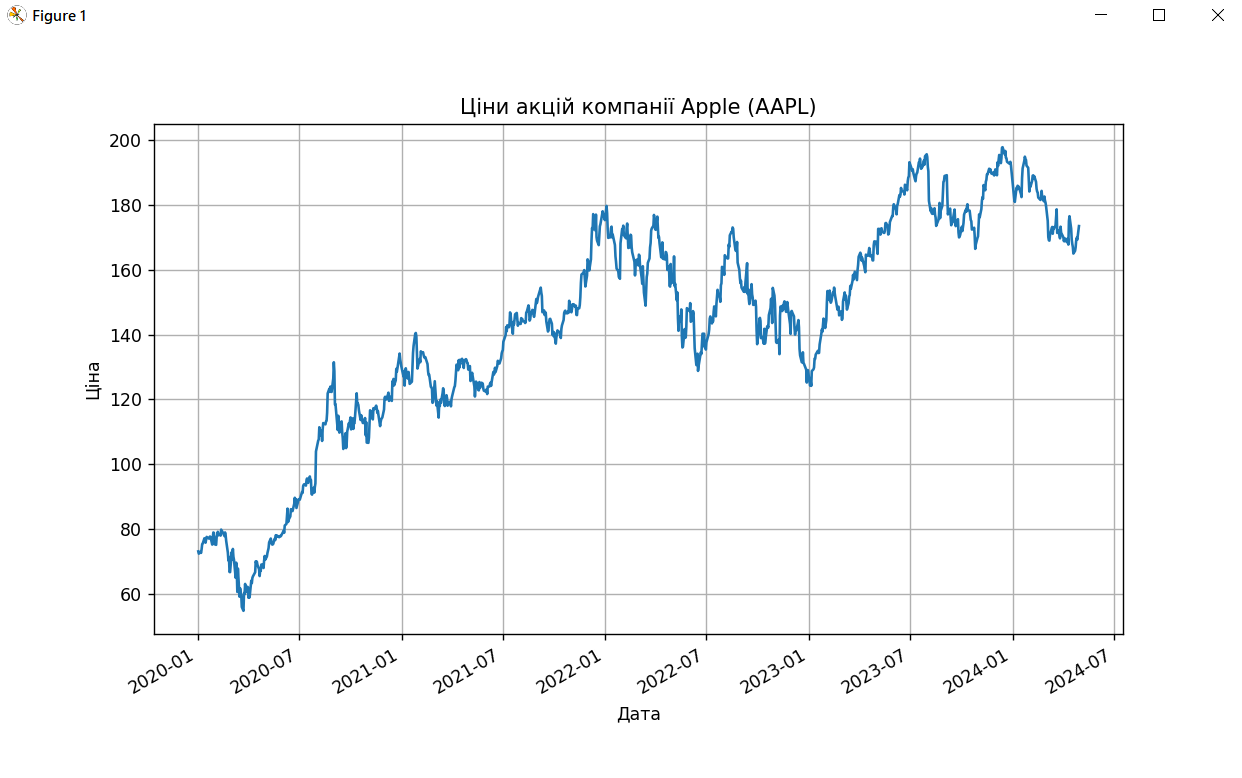
**Набір даних:**

Отримати дані про ціни акцій компанії Apple (AAPL)

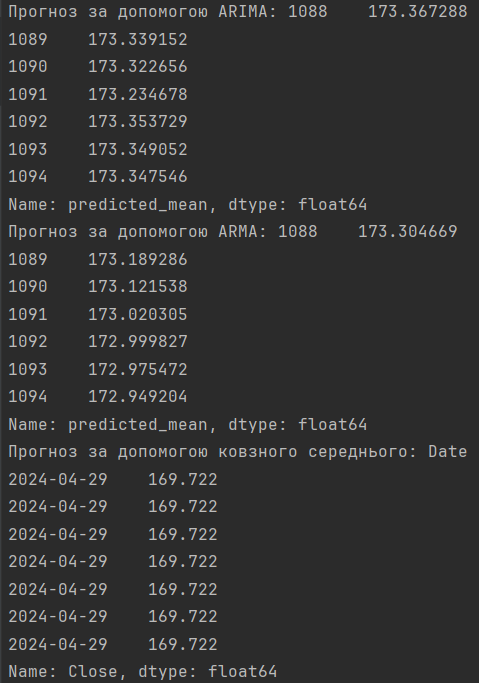
Запуск програми відбувається так:

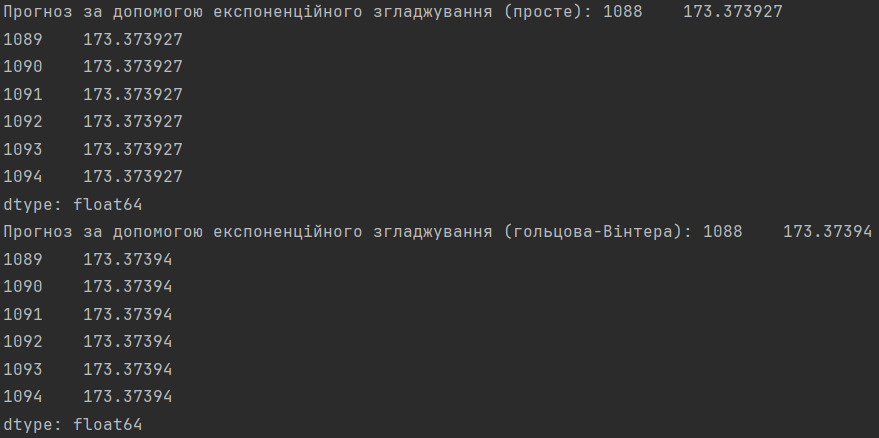


Графік цін акцій компанії Apple, який ми завантажили:

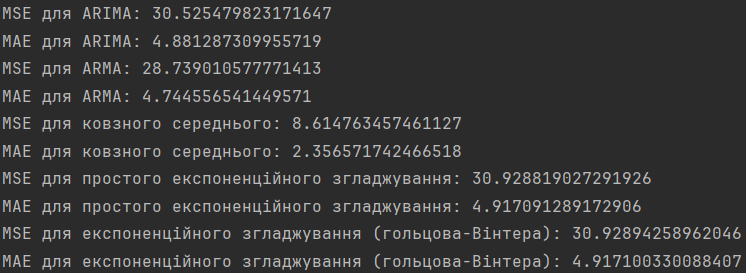


Прогнозування за допомогою числових рядів на найближчий тиждень за допомогою алгоритмів ARIMA, ARMA, ковзного середнього та експоненційного згладжування:





Значення MSE та MAE для порівняння:



**Код програми:**

import yfinance as yf  
import pandas as pd  
from statsmodels.tsa.arima.model import ARIMA  
from statsmodels.tsa.statespace.sarimax import SARIMAX  
from statsmodels.tsa.holtwinters import SimpleExpSmoothing  
from statsmodels.tsa.holtwinters import ExponentialSmoothing  
import matplotlib.pyplot as plt  
  
# Отримати дані про ціни акцій компанії Apple (AAPL)  
data = yf.download('AAPL', start='2020-01-01', end='2024-04-30')  
# Вивести перші кілька рядків даних для перевірки  
print(data.head())  
  
# Аналіз даних  
  
data['Adj Close'].plot(figsize=(10, 6))  
plt.title('Ціни акцій компанії Apple (AAPL)')  
plt.ylabel('Ціна')  
plt.xlabel('Дата')  
plt.grid(True)  
plt.show()  
  
  
# ARIMA модель  
model\_arima = ARIMA(data['Close'], order=(5,1,0))  
fit\_arima = model\_arima.fit()  
  
# ARMA модель  
model\_arma = SARIMAX(data['Close'], order=(5, 0, 0))  
fit\_arma = model\_arma.fit()  
  
# Ковзне середнє  
rolling\_mean = data['Close'].rolling(window=5).mean()  
  
# Експоненційне згладжування  
model\_ses = SimpleExpSmoothing(data['Close']).fit()  
model\_hw = ExponentialSmoothing(data['Close']).fit()  
  
# Прогнози  
forecast\_arima = fit\_arima.forecast(steps=7,alpha=0.05)  
forecast\_arma = fit\_arma.predict(start=len(data), end=len(data)+6)  
forecast\_rolling\_mean = rolling\_mean.iloc[-1:].repeat(7)  
forecast\_ses = model\_ses.forecast(steps=7)  
forecast\_hw = model\_hw.forecast(steps=7)  
  
# Виведемо прогнози  
print("Прогноз за допомогою ARIMA:", forecast\_arima)  
print("Прогноз за допомогою ARMA:", forecast\_arma)  
print("Прогноз за допомогою ковзного середнього:", forecast\_rolling\_mean)  
print("Прогноз за допомогою експоненційного згладжування (просте):", forecast\_ses)  
print("Прогноз за допомогою експоненційного згладжування (гольцова-Вінтера):", forecast\_hw)  
  
# Порівняння  
# Фактичні значення цін закриття  
actual\_values = data['Close'].iloc[-7:]  
from sklearn.metrics import mean\_squared\_error, mean\_absolute\_error  
  
# Для ARIMA  
mse\_arima = mean\_squared\_error(actual\_values, forecast\_arima)  
mae\_arima = mean\_absolute\_error(actual\_values, forecast\_arima)  
  
# Для ARMA  
mse\_arma = mean\_squared\_error(actual\_values, forecast\_arma)  
mae\_arma = mean\_absolute\_error(actual\_values, forecast\_arma)  
  
# Для ковзного середнього  
mse\_rolling\_mean = mean\_squared\_error(actual\_values, forecast\_rolling\_mean)  
mae\_rolling\_mean = mean\_absolute\_error(actual\_values, forecast\_rolling\_mean)  
  
# Для простого експоненційного згладжування  
mse\_ses = mean\_squared\_error(actual\_values, forecast\_ses)  
mae\_ses = mean\_absolute\_error(actual\_values, forecast\_ses)  
  
# Для експоненційного згладжування з методом Гольцова-Вінтера  
mse\_hw = mean\_squared\_error(actual\_values, forecast\_hw)  
mae\_hw = mean\_absolute\_error(actual\_values, forecast\_hw)  
  
# Виведемо значення MSE та MAE для порівняння  
print("MSE для ARIMA:", mse\_arima)  
print("MAE для ARIMA:", mae\_arima)  
print("MSE для ARMA:", mse\_arma)  
print("MAE для ARMA:", mae\_arma)  
print("MSE для ковзного середнього:", mse\_rolling\_mean)  
print("MAE для ковзного середнього:", mae\_rolling\_mean)  
print("MSE для простого експоненційного згладжування:", mse\_ses)  
print("MAE для простого експоненційного згладжування:", mae\_ses)  
print("MSE для експоненційного згладжування (гольцова-Вінтера):", mse\_hw)  
print("MAE для експоненційного згладжування (гольцова-Вінтера):", mae\_hw)

**Контрольні питання:**

1. Що таке часові ряди? Часовий ряд - це послідовність даних, зібраних або виміряних через різні моменти часу. Вони зазвичай представляють собою значення якої-небудь змінної (наприклад, ціну акцій, температуру, обсяг продажів тощо) у послідовні моменти часу.
2. Компоненти числових рядів. Часові ряди зазвичай мають чотири компоненти:

* Тренд: довгостроковий рух даних вгору або вниз.
* Сезонність: циклічний, повторюваний паттерн в даних.
* Циклічність: інші періодичні коливання, крім сезонності, які можуть мати нерегулярну або неоднакову довжину.
* Випадковість або шум: випадкові відхилення від очікуваних значень через різні фактори.

1. Для чого призначені часові ряди? Часові ряди використовуються для аналізу та прогнозування змін в часі. Вони можуть бути використані для прогнозування майбутніх значень, виявлення трендів та сезонних коливань, вивчення впливу різних факторів на дані і т. д.
2. Параметри моделі прогнозування EMMSP. EMMSP (Exponential Smoothing Model with Seasonal Periods) - це модель прогнозування, яка використовує метод експоненційного згладжування для управління трендом та сезонністю в часових рядах. Параметри цієї моделі включають:

* Коефіцієнти згладжування для тренду і сезонності.
* Період сезонності.
* Початкові значення тренду та сезонності.

1. Оцінка помилки прогнозування. Для оцінки точності прогнозування часових рядів можна використовувати різні метрики помилки, такі як середньоквадратична помилка (Mean Squared Error - MSE), середня абсолютна помилка (Mean Absolute Error - MAE), коефіцієнт кореляції тощо.
2. Алгоритм екстраполяції часового ряду без урахування зовнішніх чинників. Екстраполяція часового ряду без урахування зовнішніх чинників передбачає майбутні значення ряду на основі попередніх даних. Алгоритм може включати в себе різні методи згладжування, авторегресійні моделі, або методи, такі як експоненційне згладжування.
3. Визначення вибірки нової історії. Вибірка нової історії включає в себе останні дані з часового ряду, які використовуються для побудови прогнозу майбутніх значень.
4. Визначення вибірки максимальної подібності. Вибірка максимальної подібності включає в себе схожість між попередніми шаблонами в часовому ряді та поточною ситуацією для використання у прогнозуванні.
5. Визначення вибірки базової історії. Вибірка базової історії представляє собою базовий набір даних, який використовується для побудови моделі прогнозування та аналізу часового ряду.

**Висновок:**

Отже, в ході виконання даної лабораторної роботи я досліджував прогнозування часових рядів за допомогою різних методів, зокрема ARIMA, ARMA, ковзного середнього та експоненційного згладжування. Після аналізу результатів моєї роботи можна зробити декілька висновків.

Спочатку, щодо точності прогнозів, за метриками MSE (Mean Squared Error) та MAE (Mean Absolute Error), найкращим методом виявився ковзний середній, де значення помилок були найменшими. Далі йдуть ARMA та ARIMA, які також показали прийнятні результати. Експоненційне згладжування (як просте, так і за методом Хольцова-Вінтера) виявилося менш ефективним методом для прогнозування цих даних.

Однак, слід врахувати, що ефективність методів може залежати від конкретного набору даних та їх характеристик. Також, для підвищення точності прогнозування можна використовувати інші параметри та комбінації методів.

В цілому, виконання цієї лабораторної роботи дозволило мені отримати важливий досвід у використанні різних методів прогнозування часових рядів та оцінити їхню ефективність на практиці.