Group projects Programming in Python II

The project consists of two parts: (1) the deliverable (code and report) and (2) possibly a short oral exam (online). An oral exam will be scheduled if there is any doubt about the authorship of the submitted code, in particular if there is suspicion that AI tools were used to generate the code. The oral exam will focus on the submitted code.

You may use any resources for the project, but you must not submit code that was generated by AI. Generative AI may be used for finding functions or obtaining ideas, but the submitted code must be written by the team members and you must be able to explain and defend it during the oral exam.

Part of the project grade will be based on code quality (including clarity and comments) and on model performance. You should attempt to optimize model performance and compare different models, selecting the best one.

Deadline: 21/12/2025, 23:59. Submit your *.py or *.ipynb files to the designated folder on the K2 site for the course.

Project 1: Group 2 (Eda AKKAYA,Ikrame AMARI), Group 13 (Awa Soraya BAMBA, Alexis BOUDIN), Group 14 (Awa Soraya BAMBA,Alexis BOUDIN), Group 24 (Larissa DJONKOUE DJANKOU, Julien DUPIN), Group 25 (Lea ASSENHA, Ellischa BALOURD), Group 30 (Elliot LOPES, Jinyu LIN)

The task is to predict stock prices based on past stock prices. The data is stored in the file tech_companies_stock_prices.csv.

The fields are as follows:

- Date – the date of the transaction
- Stock_name – the name of the company
- Open, Close – opening and closing prices on the transaction date

- Open_i, Close_i – for $i = 1, 2, 3, 4, 5$, these represent the opening and closing prices on Date-i

For each company, predict the share price based on its past share prices over the previous k dates. Specifically:

- Try to predict both the opening and closing prices using various choices of past opening and closing prices.
- Explore using past share prices of other companies to predict the share price of a given company.

Use various models (such as linear regression, SVM, decision trees, and neural networks) to predict share prices. Compare the models in terms of prediction accuracy to determine which model performs best.

Finally, select one or two independent variables and use them to build models for predicting share prices. Identify which variables provide the best model for predicting share prices.

Project 2: Group 3 (Hafsa EDDAHNI, Kévin KABORE), Group 4 (Yousra KASMI, Rali LAHLOU), Group 10 (Eva Del Mar PEREZ, Valentine LEJOU), Group 15 (Manassé LUMBU,Enzo MATHIAS), Group 16 (Charly MENNESSIER,Vitiana NZAOU), Group 26 (Moetez BARHOUMI, Kosseila BEKHTAOUI), Group 29 (Stella Chanelle DETIO,Imad EL ASRI)

The task is to predict stock prices based on past stock prices. The data is stored in the file stock_data_for_hw1.csv.

The fields are as follows:

- Date – date of the transaction
- Company – name of the company
- Open, Close, High, Low – opening, closing, highest, and lowest prices on the transaction date
- Open_i, Close_i, High_i, Low_i – for $i = 1, 2, 3, 4, 5$, these represent the opening, closing, highest, and lowest prices on Date-i

For each company, predict the share price based on its past share prices over the previous k dates. Specifically:

- Try to predict opening, closing, high, and low prices separately, using different combinations of past opening, closing, high, and low prices.

- Investigate using past share prices of other companies to predict the share price of a given company.

Use various models (e.g., linear regression, SVM, decision trees, neural networks) for predicting share prices. Compare the models in terms of prediction accuracy to determine which one performs best.

Finally, select one or two independent variables to build models for predicting share prices. Identify which variables yield the best predictive model.

Project 3: Group 5 (Ahmed OUINEKH,Alexis VAYSSI), Group 6 (Bechar SALEM,Blerina SHALA), Group 9 (Michel AFEICHE, Imane AIT SALAH), Group 17 (Kaixin ZHANG,Amal ZOUAOUI), Group 18 (Vanessa SOLARI, Fabrice TOHON), Group 27 (Tristan BONAPERA, Mathieu CAGÉ)

Predict stock prices based on past stock prices, ESG Score, and net income. The CSV file StockESG0.csv contains the following columns:

- Date – date of transaction
- Company – name of the company
- Open, Close – opening and closing prices on the transaction date
- Open_i, Close_i – for $i = 1, 2, 3, 4, 5$, these represent the opening and closing prices on Date-i
- Net income – net income of the company
- ESG Score – overall ESG score
- ESG Controversies Score, Environmental Controversies Count, Human Rights Controversies, Business Ethics Controversies, Wages Working Condition Controversies Count, Total Renewable Energy To Energy Use in million, Estimated CO2 Equivalents Emission Total – various ESG-related scores and metrics

Use past opening or closing prices, net income, and various ESG scores to predict the current opening (or closing) price.

Apply different models, such as linear regression, SVM, decision trees, and neural networks, for predicting the share price. Compare the models in terms of prediction accuracy to determine which performs best.

Finally, select one or two independent variables to build predictive models for share prices. Identify which variables yield the most accurate model.

Project 4: Group 7 (Axelle PONCET,Sami RTEL BENNANI), Group 8 (Yvana JOUANJAN,Carmelina Grace-Lucia MBESSO), Group 11 (Lucas GONZALES,Louis JANIN), Group 19 (Alexis FERRO, Simon FRADCOURT), Group 20 (Hugo ELHAIK, Youssef ELKADY), Group 21 (Asaad ALSHEIBANI, Ismael AMINOU), Group 28 (Corentin DAMIAN, Solène DEBREUIL)

The task is to predict the ESG score based on various other variables, such as:

- ESG Controversies Score
- Environmental Controversies Count
- Human Rights Controversies
- Business Ethics Controversies
- Wages Working Condition Controversies Count
- Total Renewable Energy To Energy Use in million
- Estimated CO2 Equivalents Emission Total

The data is stored in the file ESGAllCompanies1.csv.

Use various models (e.g., linear regression, SVM, decision trees, neural networks) to predict the ESG score. Compare the performance of the models in terms of prediction accuracy. Which model performs the best?

Select one or two independent variables and build models to predict the ESG score. Identify which variables provide the best predictive model for the ESG score.

Project 5: Group 12 (Maiwenn SIMON, Aude NEVO), Group 22 (Paul GRAVIS ,Pierre YVENOU), Group 1 (Rim ZAMZAMI,Titouan VAN MEYEL), Group 23 (Gabriel SOYER, Mattis LEFEBURE))

The task is to reverse engineer a portfolio management algorithm. The historical stock data is stored in the file tech_companies_stock_prices.csv. The fields are as follows:

- Date – the date of the transaction

- Stock_name – the name of the company
- Open, Close – opening and closing prices on the transaction date
- Open_i, Close_i – for $i = 1, 2, 3, 4, 5$, these represent the opening and closing prices on Date-i

A portfolio is a collection of stocks and cash. A portfolio management algorithm decides, at the beginning of each trading day, how many stocks to buy or sell from each company, based on the current portfolio (numbers of stocks and cash) and the stock prices (opening and closing prices) of the previous days. The goal of the portfolio management algorithm is to maximize the value of the portfolio in the long run.

The historical performance of the portfolio management algorithm can be found in test_portfolio .csv. The rows of this file are indexed by dates of transactions, the columns correspond to various stocks. In addition, there is a column cash and a column value. The column value represents the value of the portfolio at the end of the given day. The values in columns labeled by stocks correspond to the number of stocks bought (if positive) or sold (if negative) at the beginning of the day, using the opening prices. For computing the value, the cash in the portfolio and the values of the stocks in the portfolio using the closing prices are added. It is assumed that at the beginning of the day, the proceeds from selling stocks at the opening prices and the cash reserve can be used to buy stock at the opening prices.

For example, the line

```
,TSLA,AAPL,AMZN,MSFT,NVDA,META,GOOGL,cash,value
2023-10-10,0.0,4.0,-2.0,-2.0,0.0,0.0,1.0,68.65997314452716,303305.98236083984
```

means that on 2023−10−10, at the beginning of the day, 4 more shares of AAPL were bought, 2 shares of AMZN and MSFT were sold, and 1 more share of GOOGL was bought, and 68.65997314452716 was added to the cash reserves. At the end of the day, the value of the portfolio was 303305.98236083984, where the closing prices were used for computing the value. Note that 4 is not the number of AAPL shares in the portfolio, but the number of AAPL shares which were bought in addition to the AAPL shares already in the portfolio. Likewise, −2.0 is the number of AMZN shares sold from the number of AMZN shares in the portfolio.

First, use the historical data from tech_companies_stock_prices.csv and test_portfolio .csv to reconstruct the portfolio (numbers of stocks and cash) at the end of each trading date. Second, use various models (e.g., linear regression, SVM, decision trees, neural networks) to predict:

- the value of the portfolio at the end of each day (except the first trading day) as a function of the portfolio at the end of the previous day and the opening prices of the current day;

- the trading actions (buying/selling stocks) as a function of the trading actions during the previous k days ($k = 1, 2, 3$).

Keep at least 20% of the data for testing.