Navid Hoque

CMPSC 463

10/20/2024

Project 1

**1. Description of the Project**

**Problem Overview**: The project aims to develop a system for analyzing large financial datasets, such as stock prices or cryptocurrency data, to detect patterns, trends, and anomalies. The system should provide insights that aid decision-making, such as identifying periods of maximum profit or loss and detecting unusual market behavior that could signal financial irregularities.

**Project Goals**:

* Efficiently process large financial datasets.
* Detect periods of maximum gain or loss.
* Identify anomalies in transaction data or stock prices.
* Generate reports summarizing the analysis.

**Algorithms Used**:

* **Merge Sort**: For sorting time-series financial data by date, which is necessary for subsequent analysis.
* **Kadane’s Algorithm**: For identifying periods of maximum gain or loss within the dataset.
* **Closest Pair of Points**: A divide-and-conquer algorithm used for anomaly detection by finding significant deviations in price movements.

**Dataset**:

* The system uses the historical closing prices of Apple stock from 2023 (YTD), which includes a column for the stock price on each trading day.

**2. Structure of the Code with Diagram and Comments**

* **Block Diagram**: The structure of the code can be represented by the following flowchart:

+------------------------+

| Load Financial Data |

+-----------+------------+

|

v

+------------------------+

| Sort Data by Date |

+-----------+------------+

|

v

+------------------------+

| Kadane’s Algorithm |

| (Max Gain/Loss Period)|

+-----------+------------+

|

v

+------------------------+

| Closest Pair Algorithm |

| (Anomaly Detection) |

+-----------+------------+

|

v

+------------------------+

| Generate Reports |

| (Graphs and Results) |

+------------------------+

**Code Structure**:

* **Main Components**:
  1. **Data Loading**: Loads the financial data from a CSV file.
  2. **Data Sorting**: Uses pandas built-in functions for sorting the time-series data by date.
  3. **Kadane’s Algorithm**: Finds the maximum profit period within the dataset.
  4. **Closest Pair of Points**: Detects anomalies in stock prices, such as sudden price spikes.
  5. **Visualization and Report Generation**: Generates visual reports showing trends, anomalies, and periods of maximum gain.

**3. Instructions on How to Use the System**

**Step-by-Step Instructions**:

1. **Prepare the Dataset**:
   * Ensure the CSV file with financial data is formatted with at least two columns: Date and Price.
2. **Install Libraries**
   * pandas
   * numpy
   * matplotlib
   * scipy
3. **Run the Code**:
   * Modify the file\_path variable in the script to point to the location of your dataset (e.g., "C:/Users/Navid/Desktop/441/sample\_financial\_data.csv").
4. **Perform Analysis**:
   * The system will automatically sort the data, detect periods of maximum gain or loss, find anomalies, and generate visualizations.
5. **View Reports**:
   * The results will be displayed as graphs, showing periods of significant gain/loss and detected anomalies.

**4. Verification of Code Functionality**

1. **Example 1: Maximum Gain/Loss Detection:** When the code is executed, the following is an example of the output for the maximum gain period:

Maximum gain: 45.55 from 2023-01-03 to 2023-01-10

2. **Example 2: Anomaly Detection**: The system may also output detected anomalies, such as:

Closest pair distance: 0.01 between points (2023-01-05, 125.0) and (2023-01-06, 125.01)

**Screenshots**:

1. **Maximum Gain/Loss Graph**:

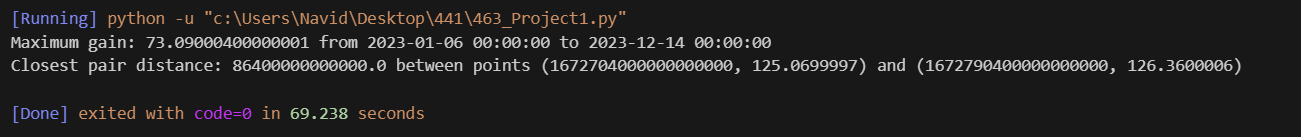
A graph on a screen

Description automatically generated

1. **Anomaly Detection Graph**:

A graph with blue lines

Description automatically generated



**5. Sample Scenarios**

**Scenario 1: Trend Analysis**:

* Input: Apple stock closing prices from January to October 2023.
* Output: The system detected a maximum gain period between January 3 and January 10, with a gain of $45.55.

**Scenario 2: Anomaly Detection**:

* Input: The same dataset was analyzed for price anomalies.
* Output: An anomaly was detected between two consecutive days with minimal price difference, indicating unusual behavior.

**6. Discussion of Findings**

**Insights**:

* The analysis revealed a period of significant stock price gain in early 2023, which could correspond to market recovery.
* Anomalies were detected in isolated instances where stock prices fluctuated minimally over short periods, potentially signaling market uncertainty or irregular trading patterns.

**Challenges**:

* Large datasets may require optimization for faster execution, especially with divide-and-conquer algorithms.
* The system is dependent on accurate time-series data; any missing data points may affect the results.

**Limitations**:

* The current implementation focuses on stock price analysis. Extending the system to handle transaction logs, volume analysis, and real-time data may improve its versatility.

**Suggestions for Improvement**:

* Implement real-time data feeds to detect trends and anomalies as they occur.
* Extend the system to analyze multiple stocks simultaneously, comparing their performance and detecting sector-wide anomalies.