

## MSBA7012/MACC7022 Individual Assignment 1: Stock Return Prediction

程序代写代做 CS编程辅导

Deadline: Sunday, April 7, 2024 11:59pm

### Dataset:

- twitter\_ecommerce.csv: Information on tweets about four Chinese E-commerce stocks: BABA, JD, PDD, and TIGR.
- stock\_price.csv: Price information for the above four stocks.



### Tasks:

1. Calculate three sentiment measures: neg (fraction of negative words), pos (fraction of positive words), and vader for each stock at the daily level and assess the pairwise correlations among them using the Pearson Correlation tool. Comment on how these sentiment measures are correlated with each other.
2. Calculate the daily returns. Return is defined as  $(Price_t - Price_{t-1}) / Price_{t-1}$ . Consider the "Adj Close" column as the price. Use the Multi-Row Formula tool to generate the lagged price column with the following expression: IF [Row-1:Ticker] = [Ticker] THEN [Row-1:Adj Close] ELSE NULL() ENDIF. Make sure you sort the data appropriately before applying this expression.
3. Join the stock return and tweet data by ticker and Date and then run the regression of the same-day stock return on daily tweet intensity (i.e., number of tweets) and three sentiment measures. Pool the data of the four stocks together for the regression. Given the distribution of tweet numbers, apply the logarithmic transformation to the number of tweets variable to increase model fit after adding 1 to it, i.e.,  $\log(1 + \text{tweet\_num})$ . Note that on certain trading days, there may not be any tweet on a specific stock on Twitter. Replace such missing values of tweet intensity and sentiments with 0. Comment on the regression results and summarize the key insights. Interpret the regression coefficients, p-values, and R-squared.
4. Run the regression of next-day stock return on daily tweet intensity and three sentiment measures, i.e., using yesterday's tweet data to predict today's stock return. To do this, you can use the Multi-Row Formula tool to generate a lagged date column (LaggedDate) first in the stock return dataset. Then in the existing Join tool, you can simply change the Field from Date to LaggedDate for the stock return side. Comment on the regression results and summarize the key insights. In your response, highlight the differences in regression results for next-day stock return from that of same-day stock return.

### Deliverables:

- You are required to finish this assignment in Alteryx only.
- Source code of your program in one file (.yxmd for Alteryx) containing the complete analysis, with annotations explaining each tool and step. **Use relative path** for workflow dependencies in Alteryx so that the grader can run your program without making any change. There should be **one Join Tool** in your program, and by changing the specific Field, the grader can run either same-day or next-day return regressions.
- A Word document (.docx) containing your responses to each requirement in the Tasks.
- Compress the above two files into a zip file named with your student ID, e.g., 123456.zip.
- You should not make any modifications to the two input files: twitter\_ecommerce.csv and stock\_price.csv. Also, DO NOT include these two input csv files in your zip file.

### Evaluation Criteria:

- Correctness and completeness of the feature engineering steps implemented in Alteryx.
- Accuracy and thoroughness of the model interpretation of results within Alteryx.
- Quality and clarity of the final report, including insights and conclusions drawn from the analysis.