



KANAK PRAJAPATI - B00822362

GROUP 36

DATA MINING ALGORITHM TO ANALYZE STOCK MARKET USING LAGGED CORRELATION

BACKGROUND / PROBLEM TO BE SOLVED



Problem Statement – Individuals, Investors and Financial Professionals are continuously looking for superior system to yield high returns.



Solution – Lagged Correlation Algorithm.



Lagged Correlation analyzes correlation between other stocks and reduces short term risks of investing.



Example – When AAPL stock rises MSFT rises following 5 days (positive correlation). When AAPL rises TSLA falls following 5 days (negative correlation)

HOW ALGORITHM WORKS

User input parameters are Lag (K) and Correlation (R)

$m = 0$

For each $i = 1$ to n (stocks in the market)

- Get all dates and closing price
- each $(j + i)$ to n (stocks in the market)
 - Get all dates and closing price
 - For each lag for 1 to K

Shift the closing price by lag

Calculate the correlation coefficient between the two stocks

- If $r > R$ or $r < -R$ then
- $m = m + 1$
- $\text{Stock1NameArray}(m) = \text{Stock}(i)$
 - $\text{Stock1RArray}(m) = \text{mod}(r)$
- $\text{Stock2NameArray}(m) = \text{Stock}(j)$
- $\text{Stock2RArray}(m) = \text{mod}(r)$

End if

Next Lag

Next j

Next i

IMPLEMENTATION DESCRIPTION



Python Programming

Jupyter Notebook



Kaggle Dataset

SP500
NASDAQ
NYSE
FORTUNE 2000
TOP 10 Stocks (Manual Dataset)

Date	Low	Open	Volume	High	Close	Adjusted Close
12-12-1980	0.1283479928970340	0.1283479928970340	469033600	0.1289059966802600	0.1283479928970340	0.10032564401626600
15-12-1980	0.12165199965238600	0.12221000343561200	175884800	0.12221000343561200	0.12165199965238600	0.09509161859750750
16-12-1980	0.11272300034761400	0.1132809966802600	105728000	0.1132809966802600	0.11272300034761400	0.08811209350824360
17-12-1980	0.11551299691200300	0.11551299691200300	86441600	0.11607100069522900	0.11551299691200300	0.09029290080070500

Reference: <https://www.kaggle.com/datasets/paultimothymooney/stock-market-data>

EXPERIMENTAL RESULTS

- Lag (K) = 5
- R = 0.3 (-1 to +1)

Company A	Company B	Co-relation
Output_GOOG.csv	Output_DIS.csv	0.3698814404112270
Output_GOOG.csv	Output_AAPL.csv	0.567996860633749
Output_GOOG.csv	Output_AMZN.csv	0.9531329012465760
Output_GOOG.csv	Output_NVDA.csv	0.7736698953353940
Output_DIS.csv	Output_AAPL.csv	0.8196903626413400
Output_DIS.csv	Output_WMT.csv	0.6192256692323290
Output_DIS.csv	Output_TSLA.csv	0.6654268677674220
Output_DIS.csv	Output_JNJ.csv	0.6432869514530300
Output_DIS.csv	Output_AMZN.csv	0.5342012528814310
Output_AAPL.csv	Output_WMT.csv	0.8028220895842000
Output_AAPL.csv	Output_TSLA.csv	0.8180053658541670
Output_AAPL.csv	Output_AMZN.csv	0.7208918723743440
Output_AAPL.csv	Output_NVDA.csv	0.644908359398055
Output_WMT.csv	Output_TSLA.csv	0.8334229203733490
Output_WMT.csv	Output_JNJ.csv	0.4260431570862900
Output_WMT.csv	Output_MSFT.csv	-0.516979483087668
Output_TSLA.csv	Output_AMZN.csv	0.3109969532116820
Output_TSLA.csv	Output_NVDA.csv	0.4177247325629720
Output_JNJ.csv	Output_NVDA.csv	-0.5658721214033830
Output_AMZN.csv	Output_NVDA.csv	0.8377393542225560



THANK YOU

OPEN FOR ANY QUESTIONS OR SUGGESTIONS.