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
- Stock1NameArray(m) = Stock(i)
  - Stock1RArray(m) = mod(r)
  - Stock2NameArray(m) = Stock(j)
  - Stock2RArray(m) = mod(r)

End if

Next Lag

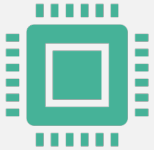
Next j

Next i

IMPLEMENTATION DESCRIPTION

Python Programming

Jupyter Notebook



Kaggle Dataset

SP500

NASDAQ

NYSE

FORTUNE 2000

TOP 10 Stocks (Manual Datset)

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
AMZN.csv	FB.csv	0.6781148776891920
AMZN.csv	TSLA.csv	0.8983493819720430
AMZN.csv	GOOG.csv	-0.48224845650436600
AMZN.csv	AAPL.csv	0.5910653803612350
AMZN.csv	WMT.csv	0.935406401269331
MSFT.csv	NVDA.csv	0.41474375175131400
MSFT.csv	FB.csv	-0.751577471646856
MSFT.csv	DIS.csv	-0.9873757926090790
MSFT.csv	GOOG.csv	0.8672160607021210
MSFT.csv	AAPL.csv	-0.7567207915102790
NVDA.csv	DIS.csv	-0.37367513476502000
NVDA.csv	TSLA.csv	0.3993024846124670
NVDA.csv	GOOG.csv	0.5398705408787790
NVDA.csv	WMT.csv	0.32594949377145900
FB.csv	DIS.csv	0.6804122920775670
FB.csv	TSLA.csv	0.47685841689511100
FB.csv	GOOG.csv	-0.814346751273562
FB.csv	AAPL.csv	0.984206242314141
FB.csv	WMT.csv	0.6259412945717400
DIS.csv	TSLA.csv	-0.3162471230480660
DIS.csv	GOOG.csv	-0.7783820022632250
DIS.csv	AAPL.csv	0.7051380843402340
TSLA.csv	AAPL.csv	0.44147591720831000
TSLA.csv	WMT.csv	0.9492428398668470
GOOG.csv	AAPL.csv	-0.7483739293666170
GOOG.csv	WMT.csv	-0.34015628915718400
AAPL.csv	WMT.csv	0.5916236114476760

THANK YOU

OPEN FOR ANY QUESTIONS OR SUGGESTIONS.