

Task : Clustering based identification

Data: Date and time, OCHL, volume , open interest

Deciding Features (X):

- ☐ Normalization of position of the OHCL candle:
 - ☐ Find the moving mean of past n candles (close, open , $0.5(C+O)$) - used close, close is generally chosen
 - ☒ ~~(O, C, H, L) - moving mean~~
 - ☒ ~~Vedika: use exponential moving average instead of simple moving average. EMA is better for short term markets, suitable for the per minute prediction.~~
 - ☒ ~~Normalisation works as expected, verified using visualisation~~
- ☒ ~~Moving mean~~
- ☐ Gradients 1st, 2nd Of moving mean
- ☐ Volume Their gradients
- ☒ ~~Differences between OHLC ↔ OHCL~~
- ☒ ~~Time of candle (hour + min)~~
- ☒ ~~Week day~~
- ☐ Think of more features: gradients, frequencies (maybe 30)
 - ☐ Adjusted close
 - ☐ Moving average convergence/divergence (MACD) = 26 period EMA - 12 period EMA
 - ☐ SMA can also be used in addition to the EMA, 2-3 SMAs and EMAs can be used, over different periods
 - ☒ ~~Momentum indicators~~
 - ☐ RSI (relative strength index)
 - ☐ Stochastic oscillator
 - ☐ Average directional index (ADX)
 - ☐ Dynamic momentum index
 - ☐ Directional movement index
 - ☐ Ultimate oscillator
 - ☐ Disparity index
 - ☒ ~~Economic indicators~~
 - ☐ More useful for long term investments
 - ☐ Economic cost index (ECI)
 - ☐ GDP

- ☐ Purchasing managers' index (PMI)
- ☐ Personal consumption expenditures (PCE)
- ☐ Also include seasonality, and inflation indicators, like PPI, CPI...
- ☒ Coincident indicators
 - ☐ Similar to economic indicators, not useful in this context
 - ☐ GDP and employment figures
- ☒ Lagging indicators
 - ☐ Similar to economic indicators
 - ☐ GDP
 - ☐ CPI
 - ☐ Balance of Trade (BOT)
 - ☐ Moving average crossovers
- ☐ Technical indicators have been observed to not work for us
- ☐ Look for similar projects to cluster candlesticks, try implementing them
 - ☐ Look for insights from a single candlestick, or a group of candlesticks
- ☐ ESN (echo state networks) can be used for prediction
- ☐ LSTM is commonly used for stock price prediction
- ☐ Look for features that can help
 - ☐ Perform more intensive EDA, use boxplots, (preferably) pdf plots and heatmaps
- ☒ ~~Vedika: insights from EDA~~
 - ☒ ~~Start and the end of the work week witness high volatility (spikes and drops in G-O)~~
 - ☒ ~~First two hours of the market are volatile, followed by increasing tranquility~~
 - ☒ ~~Midweek trading volume is high~~
 - ☒ ~~Midday trading volume is low~~
 - ☒ ~~Relationship among volume, G-O, and open interest needs to be looked at — not useful~~
 - ☒ ~~EMA vs open interest is a complicated graph — unlikely to be useful~~
 - ☒ ~~High H and low L point to volatility~~
- ☐ Clustering part: k-mean :
 - ☒ ~~Clustered on EMA, G-O, volume, and open interest~~
 - ☒ ~~Scatter plot didn't provide clarity~~
 - ☒ ~~Help needed for numerical analysis of the clusters~~
- ☐ H-clustering
- ☐ Latent space VAEs

Let's say we get k no of clusters

KPI: identify the clusters with better Y var.


Deciding the Y var. (close - open)

- ☐ Occurrence of Green/ Red in n future candles (categorical type)
 - How many time it is green and red, neutral
 - Also decide the green/red based on threshold
- ☐ Amount of change in the open and close - for n future candles (numerical)

Required a table:

- Cluster number - Y_cat - Y_num

Code NBs:

 OptAlpha Candle Prediction.ipynb

Work Logs

4/June/2024

- Found moving mean avg - simple and Exponential
- Difference between OHCL candles

5/June/2024

- Visualization of candles
- Started K-mean clustering

6/June/2024

- Tried on K-mean clustering
- Plotted variations of features

7/June/2024

- Looked for technical indicators
- Found potential ideas for prediction model