Indhold

1 Dataindsamling

Planen er at indhente data fra 2018-01-01 til 2022-04-30. Vi vil træne på data fra 2018-01-01 til 2020-31-12, for senere at kunne teste vores modeller på data fra 2021-01-01 til dags dato.

Dukascopy (Schweizisk bank)

Vi indsamler historisk minut-data fra dukascopy. Vi ønsker at indsamle data fra

Indexer:

- DAX-indexet (Germany 40)
- Dow Jones (UK 100)
- Hong Kong (Hong Kong 40)
- NASDAQ (USA 100 Technical)
- S&P 500 (USA 500)

Obligationer:

- US T BOND

Valuta:

- EUR/USD

Råvarer:

- Kaffe
- Olie
- Gas

ETF'er:

- Emerging market ETF
- Growth og value ETF
- Real estate ETF

Cryptovaluta:

- Bitcoin/USD
- Ether/USD
- Cardano/USD

Forbrugerprisindex (Troels' opfordring)

Kaggle competition (Japansk markedsdata)

2 Databehandling

Index'er

Fra dataudtrækket fås følgende kolonner:

```
Local_time  Open  High  Low  Close  Volume
```

Vi ønsker for ting handlet på børser (index, enkeltaktier, ETF'er) at flage åbningstider og omskrive data til dollar-bars.

For at flage åbningstider laves en ny kolonne, som markerer hvorvidt dette tidsrum er inden børsen åbner (pre-market), efter børsen er lukket (after-market) eller i børsens normale åbningsvindue.

Omskrivning af timeframe-data til dollarbar-data:

For at omskrive til dollarbars oprettes en ny kolonne `Mean_HL` som er den (estimerede) gennemsnitlige handelspris i tidrummet (minuttet):

$$\texttt{Mean\_HL} = \frac{\texttt{High} + \texttt{Low}}{2}$$

Der laves en ny kolonne `Mean_OC`, som er gennemsnittet mellem `Open` og `Close`:

$$\texttt{Mean\_OC} = \frac{\texttt{Open} + \texttt{Close}}{2}$$

Derefter laves en ny kolonne `Total_transaction`, som angiver det beløb der er handlet for i det givne tidsrum (minut):

$$\texttt{Total\_transaction} = \texttt{Volume} \cdot \texttt{Mean\_HL}$$

Der skal nu bestemmes et **dollarbar-cap**, som skal være det beløb en enkelt dollarbar bliver dannet udfra. Ved at køre pandas kommandoen .describe() kan max{`Volume`} og max{`High`} ses. **dollarbar-cap** bliver da bestemt som produktet af de to tal afrundet til nærmeste *pæne* tal (for DAX'en fås 78 mia EUR, og der rundes af til 100 mia. EUR). Dette sikrer at der ikke kan være flere dollarbars pr. minut.

Der laves nu en ny række for hver gang den aggregerede `Total_transaction` når det givne **dollarbar-cap**, og værdierne (`Local time`, `Open`, `High`, `Low`, `Close` og `Volume` = **dollarbar-cap**) indsættes som den nye dollarbar-række:

- `Local time` sættes til `Local time` for den række hvor **dollarbar-cap** nås (altså den sidste række):
- `Open` er for den første dollarbar `Open` fra den første timeframe-bar, og for de resterende `Mean_OC` fra den timeframe-bar hvor seneste **dollarbar-cap** blev nået.
- `Close` sættes til `Mean_OC` fra den timeframe-bar hvor **dollarbar-cap** nås.
- `High` sættes til max{`High`} fra de timeframe-bars der er brugt til at lave dollarbaren.
- `Low` sættes til min{`Low`} fra de timeframe-bars der er brugt til at lave dollarbaren.
- `Volume` sættes fast til **dollarbar-cap**.

OBS: Når **dollarbar-cap** er nået videreføres det resterende beløb fra den timeframe-bar til den næste dollarbar.