BIG DATA: STOCKS ANALYSIS

PHILIPPE BERNET

MARIUS DUBOSC

BAPTISTE BOURDET

EXPLORATION OF THE DATA

- Interactive Interface
- Various metrics:
  - Moving average
  - Correlation
  - Return rate
  - Best return rate

INSIGHTS

- stability (différence between 'High' and 'Low')
- add median
- add standart deviation and interquartile difference
- stocks prediction with time series prediciton model (RNN)
- lowest daily return
- lowest return rate

MACHINE LEARNING: LINEAR REGRESSOR

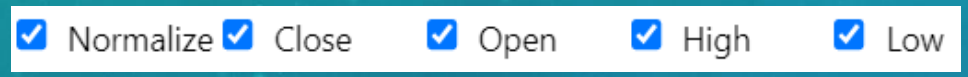
- Best model we tried
- Based on previous High, Low, Open and Closes values
- Error (RMSE) of Approximately 1 on Open and Close

Company APPLE | Granularity day | Lag 5 | Offset 1

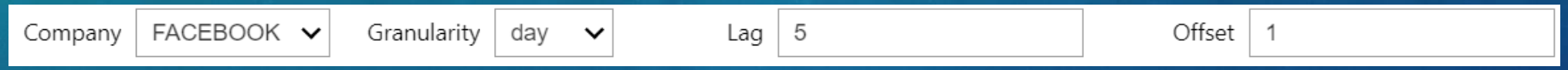
Close	Open	High	Low	Date
122.1593017578125	122.0199966430664	122.62000274658203	120.88999938964844	2020-12-02 00:00:00
122.54379646899321	122.83791475122908	123.70428284818718	121.54174641944634	2020-12-03 00:00:00

MACHINE LEARNING: RNN

- Time serie prediction : uses past values to predict future
- Recurrent neural network : train and predict sequentially
- Based on previous High, Low, Open and Closes values. Possibility to choose columns.



- Normalized data
- Lack data to be interesting




```
The RMSE for predicting the Close (train) is: 94.66
The RMSE for predicting the Close (test) is: 94.66
The RMSE for predicting the Open (train) is: 93.39
The RMSE for predicting the Open (test) is: 93.39
The RMSE for predicting the High (train) is: 94.81
The RMSE for predicting the High (test) is: 94.81
The RMSE for predicting the Low (train) is: 93.35
The RMSE for predicting the Low (test) is: 93.35
```

QUESTIONS ?