Big Data

Guillaume Carrière, Youri Colera

ÉCOLE D'INGÉNIEURS EN INFORMATIQUE

First step: exploration

- Weekly/Monthly/Yearly Average Open
- Weekly/Monthly/Yearly Average Close
- Daily/Monthly Difference
- Daily/Weekly/Monthly/Yearly Return
- Weekly/Monthly/Yearly Average Daily Return

Insights to explore the data further

- MACD
- Aroon oscillator
- Portfolio diversification
- Bollinger Band

Insights to explore the data further

- Value at risk
- Portfolio optimization
- Facebook’s prophet model
- ARIMA (Autoregressive integrated moving average)

Second step: ML

We try to predict Adj Close one day to the future.

- Data Cleansing
- Linear Regression
    - Open
    - Close
    - Weekly/Monthly/Yearly Average Close
- Adding other variables
    - Daily Difference (Imputation)
    - Company name (One hot encoding)
- Ridge Regression with hyperparameter tuning
- Random Forest Regression with hyperparameter tuning
- Moving to longer-term predictions (week)

Second step: ML

Method	RMSE	R^2
Linear regression	17.54	0.99935
Ridge regression	17.46	0.999361
Linear regression with additional parameters	17.44	0.999362
Random Forest Regression	59.18	0.9928
Linear regression for longer-term predictions (+7)	36.38	0.99724

Conclusion