Clustering and Deep Learning in Portfolio Construction

Data Science Capstone

Harvard University Extension School

Prof. Henstock

Group Members:

Nitesh Kurma
Janice Pham

Speaker: Nitesh

Introduction

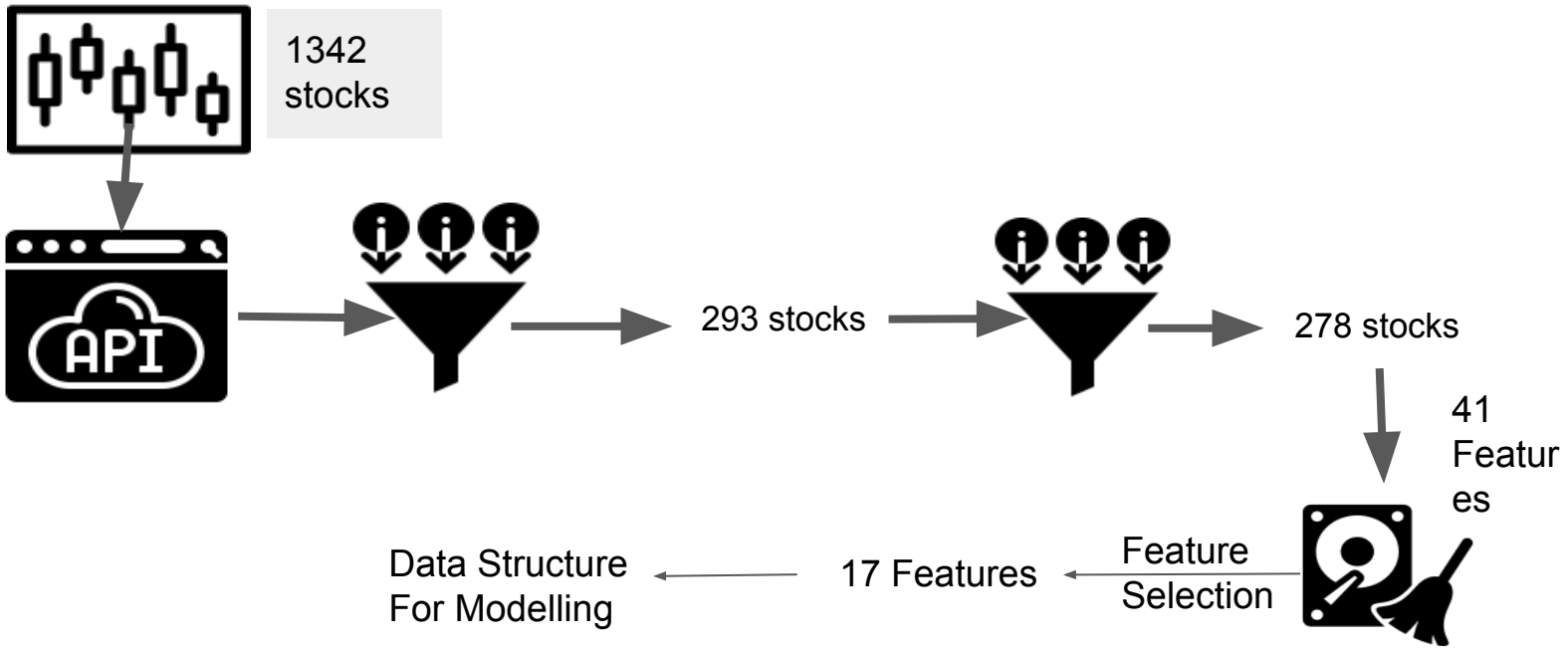
Project Statement

- Explore clustering and deep learning algorithms to enhance traditional portfolio construction techniques.
- Explore clustering method as a visualization tool in a high dimensional data
- Group equities using clustering

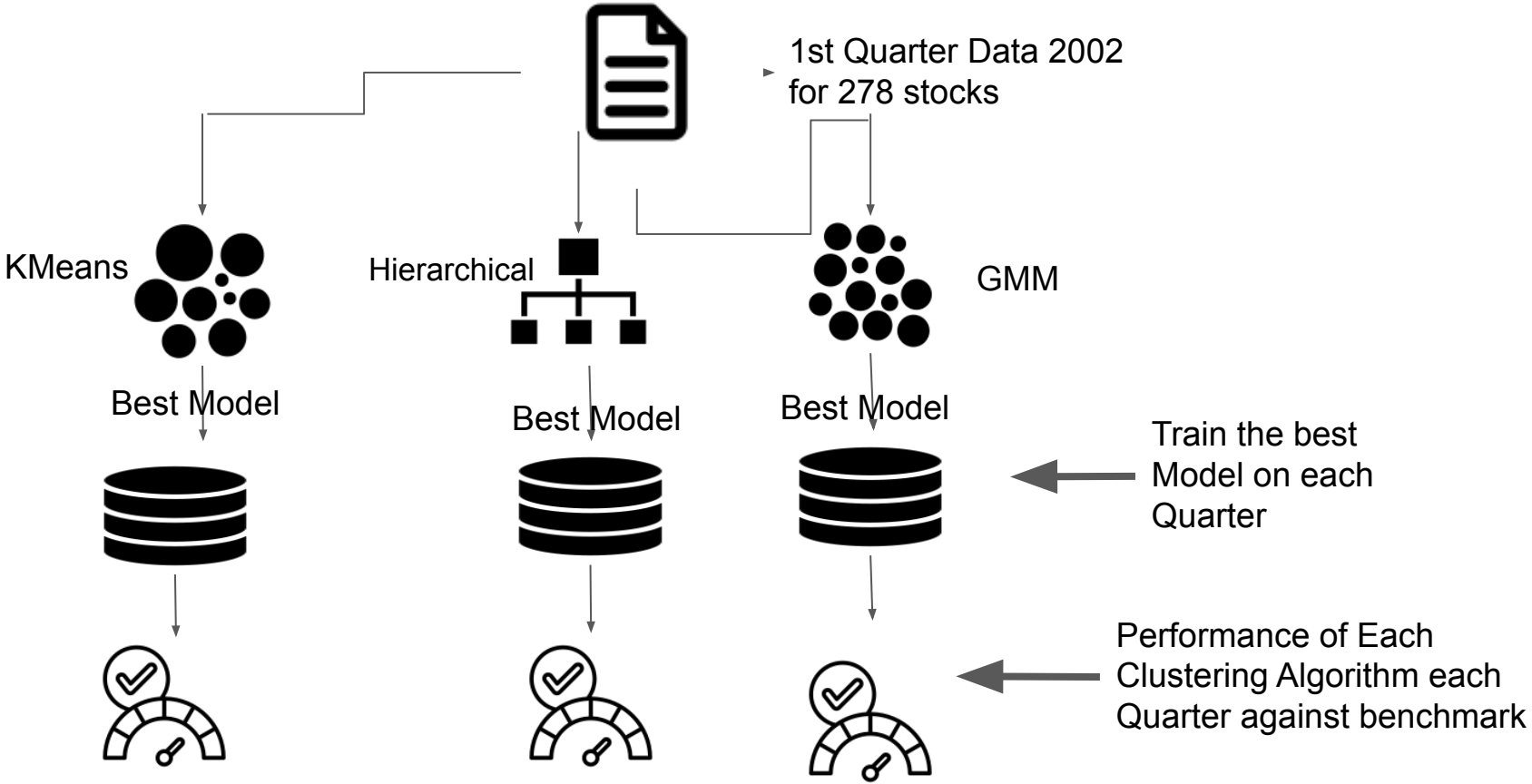
Motivation

- To learn about financial markets , and apply data science techniques learned in the class to financial time series data.
- To learn how clustering algorithms can help explain the complexity in the interactions among the stocks

Data Collection & Feature Selection



Clustering Methodology



Speaker: Janice

Deep Learning (LSTM)

Goal

- Predict quarterly stocks' returns
- Rank stocks based on quarterly returns
- Create long-short portfolio (long the top N and short the bottom N)
- Compare long-short portfolio with benchmark S&P 500

Model Architecture

- Two LSTM layers
- Dropout layers
- TimeDistributed layer

Hyperparameters

Learning rate	0.01
Hidden size	64
Dropout rate	0.3
Epoch	50
Batch size	32
Patience (for Early Stopping)	5

Speaker: Janice

Model Performance

K-Means

Quarter	1st	2nd	3rd	S&P500	Best Portfolio-Beat/Fail
2012_3	0.482821	0.459996	0.404464	0.487914	Fail
2018_3	0.849734	0.845839	0.811815	0.965283	Fail
2019_4	0.804455	0.711618	0.670438	0.856314	Fail
2020_1	-0.401108	-0.422149	-0.443080	-0.394457	Fail
2021_2	0.579843	0.491961	0.434089	0.704101	Fail

GMM-9

Quarter	1st	2nd	3rd	S&P500	Best Portfolio-Beat/Fail
2012_3	0.462373	0.461569	0.360660	0.487914	Fail
2019_4	0.775173	0.724016	0.663787	0.856314	Fail
2021_2	0.558702	0.471506	0.434089	0.704101	Fail

GMM-12

Quarter	1st	2nd	3rd	S&P500	Best Portfolio-Beat/Fail
2012_3	0.461569	0.424961	0.404115	0.487914	Fail
2017_4	0.954299	0.931779	0.927617	1.070671	Fail
2018_3	0.806466	0.806155	0.780632	0.965283	Fail
2019_4	0.771634	0.741511	0.681409	0.856314	Fail
2021_2	0.553083	0.542739	0.445932	0.704101	Fail

Hierarchical Clustering

The Best Performing Hierarchical Clustered Portfolio Fails to Outperform S&P500: 5 times

	1st	2nd	3rd	S&P500	Best Portfolio Beat/Fail
2012_1	1.167031	1.114425	0.971874	1.206777	Fail
2012_3	0.427363	0.406430	0.385870	0.487914	Fail
2018_3	0.806466	0.777989	0.709346	0.965283	Fail
2019_4	0.794304	0.669718	0.663787	0.856314	Fail
2021_2	0.566577	0.434089	0.402378	0.704101	Fail

LSTM Model

LSTM Long-Short Portfolio Performance

	Whole Period	Train Period	Test Period
Long-Short 30	5.237905	5.721372	29.379903
Long-Short 50	5.699035	6.204037	36.285026
Long-Short 100	6.276307	6.812937	42.562016
S&P 500	5.793181	3.846231	1.078904

Speaker: Janice

Conclusion

- Clustering is feasible in identifying stocks that tend to move together and grouping similar stocks.
- Clustering can be considered as part of stock selection method in the portfolio construction.
- LSTM model can be used along with clustering method in enhancing the portfolio construction strategy.

Thank You