# Stock Superclusters

Samy Palaniappan Metis Project-4 Feb 2020, Chicago

# Goal

Identify stocks that have a yield better than the S&P 500.

### Goal

Identify stocks that have a yield better than the S&P 500.

Why?



# **Background Story**

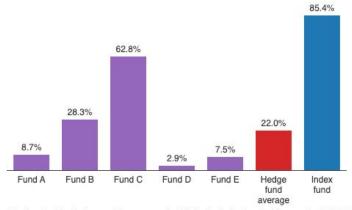
# **Background Story**

The Bet of the Decade..



#### Year 9: cumulative returns

2008 through 2016



The funds of funds have not been named publicly; the index fund is Vanguard's S&P 500 Admiral fund

# Goal - why not ? we get the satisfaction of proving Buffet wrong...

# Unbelievible News

28 FEB 2020

#### Metis Student Cracks the Stock Market

By !?!

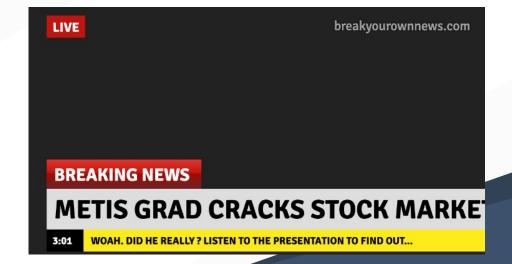
Blah Blah... CV, TFIDF, NER, SW, LDA, LSA, NMF, Blah blah... "Most people would say 'the deets', but I say 'the tails'. Just another example of innovation."



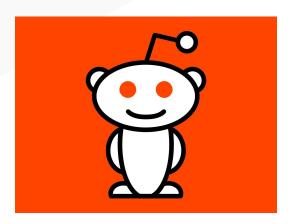
United States, with the larger de- Union's move towards standard veloping moose ecologies also poised the European moose. Stringent to make gains. The largest percent- ity controls are holding back the agege increase in moose will likely velopment of the eastern euro come from China", says McRobson, populations compared to last The Chinese government has invested when they contributed significaheavily in moose infrastructure over to europe's strong growth fig the past decade, and their committ- Norway, which is not an EU mer ment to macrofauna is beginning to but has observer status, streng pay dividends". Since 2004 China has in numbers relative to the Euro expanded moose pasture from 1.5% with numbers of Norweigian mof arable land to nearly 3.648% and known locally as elk" expected to moose numbers are expected to rise for the tenth consecutive year, pa to 60,000 making China a net moose exporter for the first time. This is the last quarter. good news for neighbouring Mongolia, a barren moose-wasteland whose researchers world wide are turniz inhabitents nonetheless have an insa-science in an attempt to boost tiable desire for the creatures. The in- year's figures. NASA stunned crease in Beijing-Ulanbataar trade is scientific community today with anticipated to relieve pressure on the announcment of their discovery

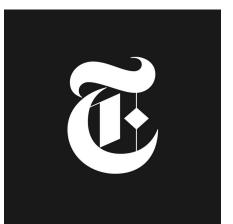
ularly thanks to a strong showing

As moose season reaches its of



### **SOoo Much Data**











# Methodology



Topic Analysis on Fin. News

CountVectorizer + LDA

TF-IDF + LSA / NMF



Senti Analysis on Relevant Articles

Vader Sentiment +

Loughran and McDonald, 2011

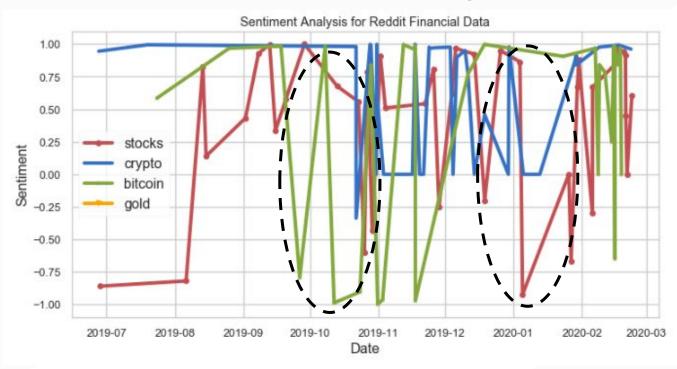


Cluster Average Daily Price + SA

**KMeans** 

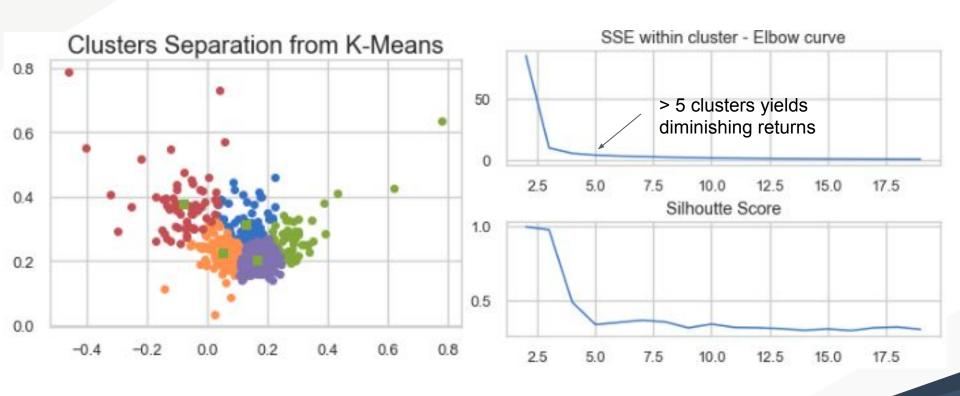
DB Scan

## Results - Sentiment Analysis (SA)

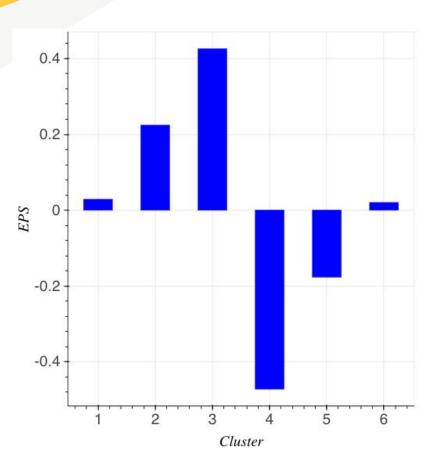


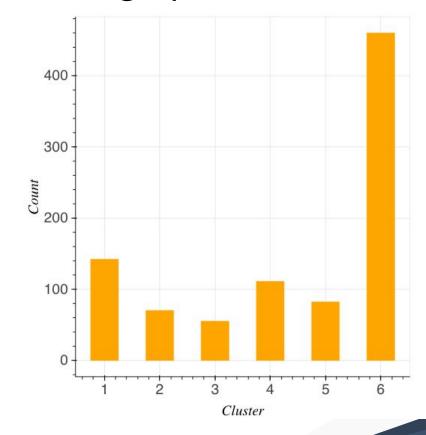
Exptected behavior (when stock sentiments are low, sentiment of bitcoin and gold should be high. Some areas, have good prediction (marked). Need better rules and data for SA.

#### Results - KMeans Clustering



# Results - KMeans Clustering (1 yr Period)





# x 10 Earnings

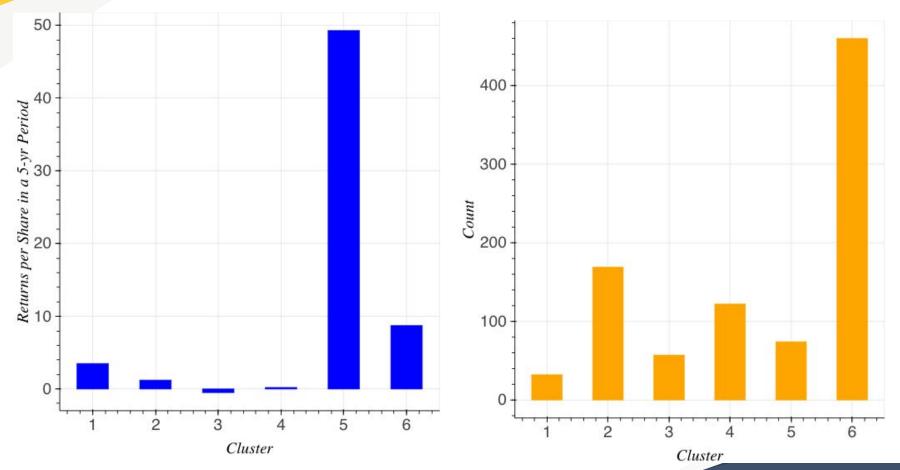
Our Best Cluster Earned 10 times more than S&P500 Last Year.

#### Results - The holy grail cluster (1 yr period)



Cluster - 35 stocks

## Results - KMeans Clustering (5 yr Period)



#### Results - The holy grail cluster (5 yr period)

```
ZTS AVGO ANSS
     DHRMSFT PGR
VRSN
    HUM CTAS TXN IDXX
```

Cluster - 50 stocks



#### **Conclusions**

- 1. Good Clustering Performance from KMeans.
- 2. At 5 years, the rate of return is slightly diminished when compared to S&P 500 from 10x to about 5x. Number of stocks in the top cluster is increased from 35 to 50.
- 3. Annual earnings  $\rightarrow$  10% more than S&P 500.
- 4. Sentiment analysis (SA) lots of room for improvement.

#### **Future Work**

- 1. Advanced financial rules for Sentiment Analysis.
- Data for SA Include 8-k and 10-q data from SEC. Include better data for Gold and Crypto sentiments.
- 3. Use fake news detection to eliminate bot generated / trolls sentiment.
- 4. Try HMM, RNN for better buy/sell predictability future price of stocks.

**Thank You!** 

**Questions?**