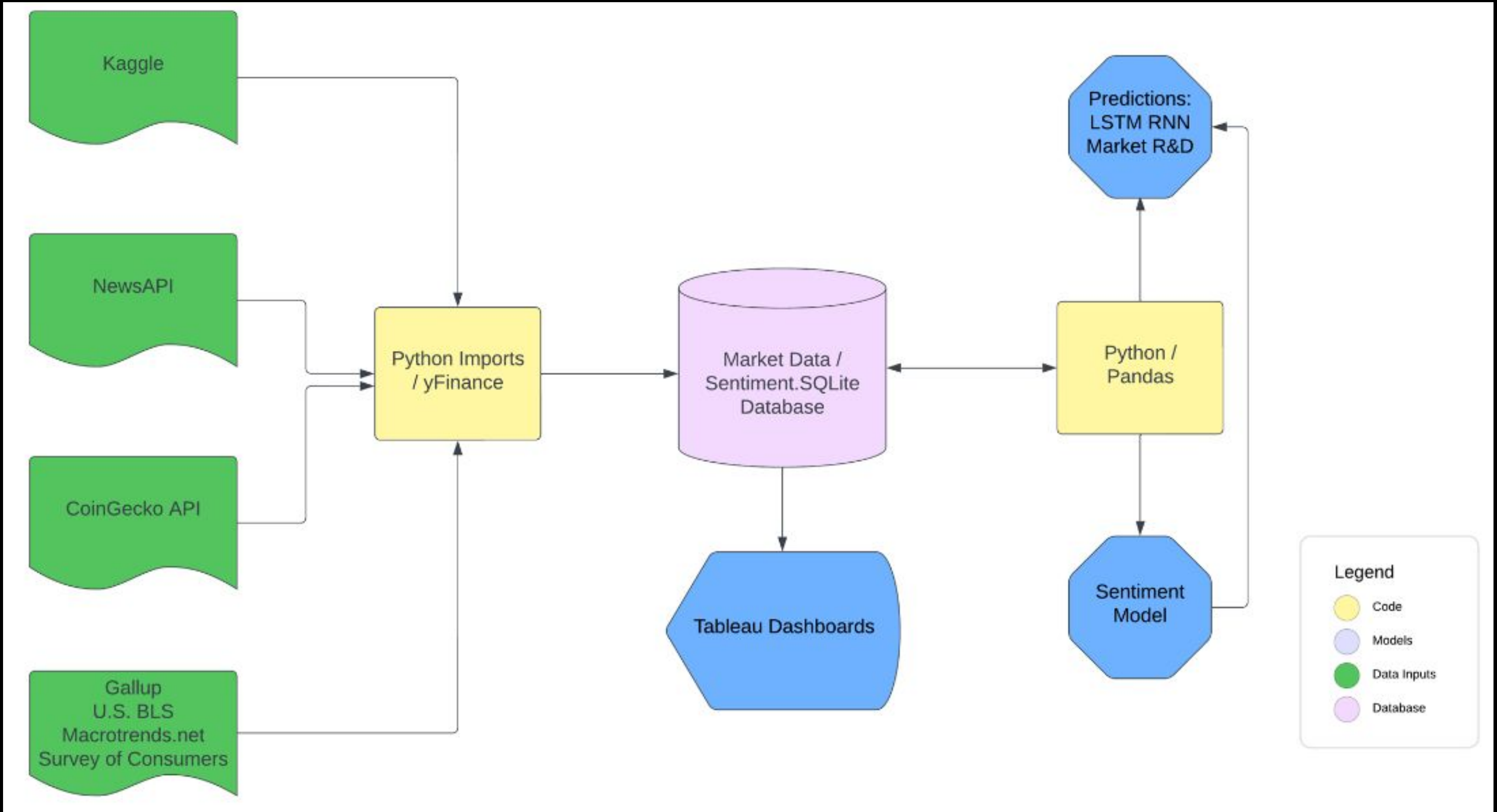
Market Data, Sentiment, and Signals

Project 4

Contributors

Justin Bernier
John Quinn
Jennifer Shulyak
Joe Moreno
Michael Raminki

Market Data, Sentiment, and Signals Flow Chart



Market Data, Sentiment, and Signals Flow Chart

Tableau Demo

Long-Short Term Memory Recurrent Neural Network

Recurrent Neural Network

RNNs process sequences of data. Maintain a hidden state to capture information from previous time steps.

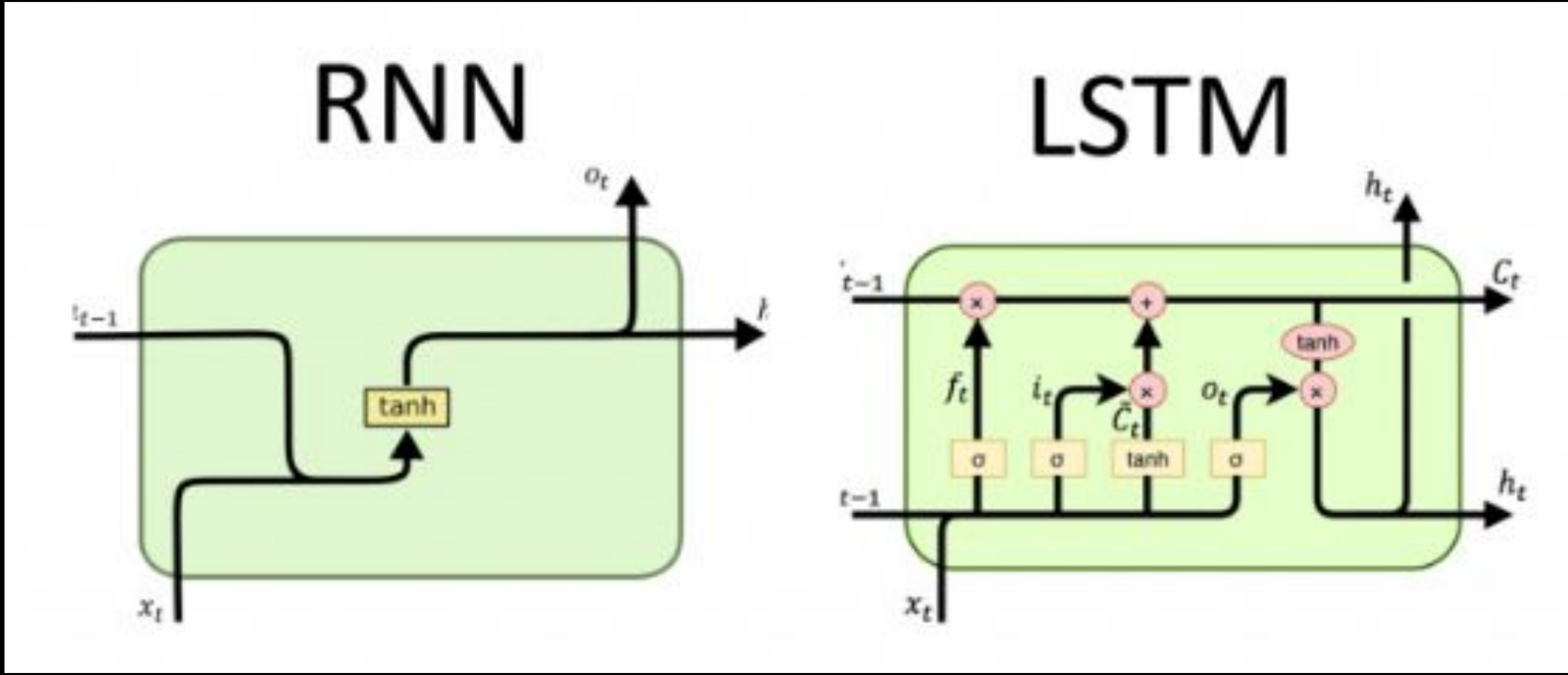
Long Short-Term Memory

LSTM include a memory cell, an input gate, a forget gate, and an output gate. Which enable it to learn long-term dependencies in the data.

Sequence-to-sequence Modeling

LSTM RNNs are often used for sequence tasks, where the goal is to predict a sequence of outputs based on a sequence of inputs.

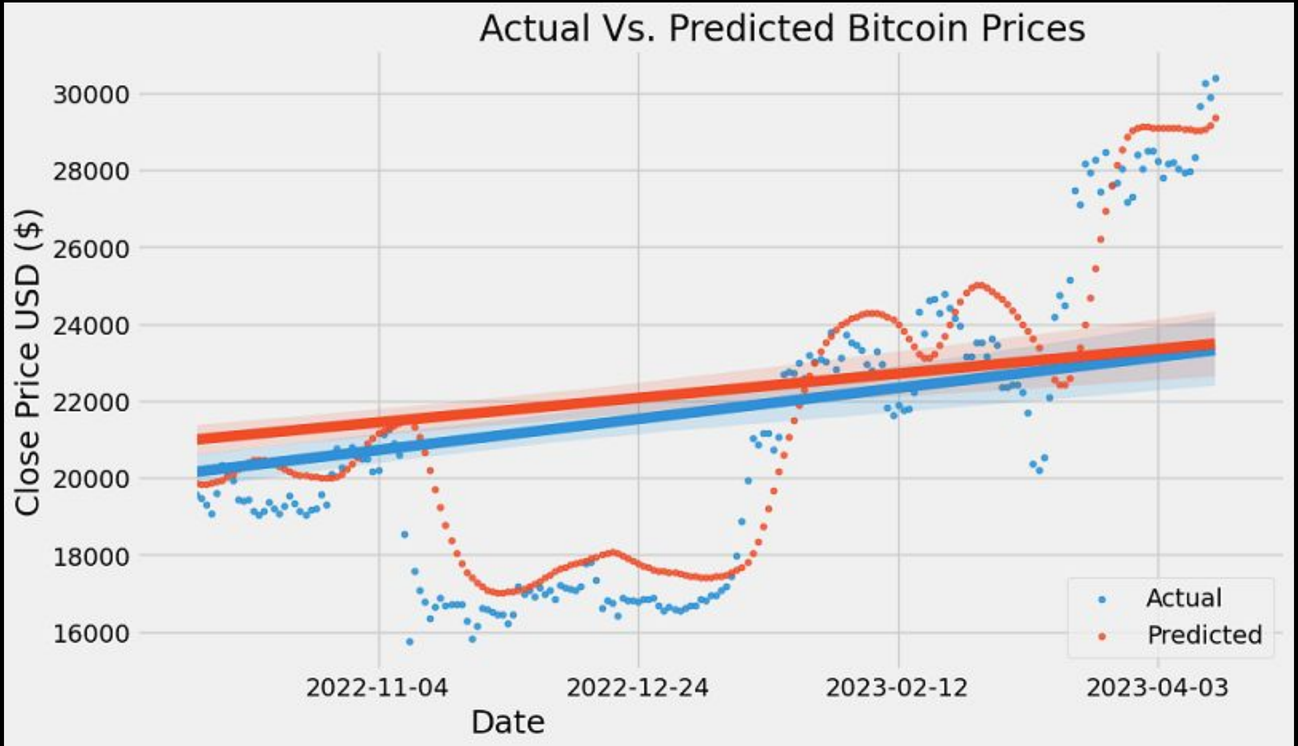
Long-Short Term Memory Recurrent Neural Network



Long-Short Term Memory Recurrent Neural Network

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 30, 40)            6720

 dropout (Dropout)           (None, 30, 40)            0

 lstm_1 (LSTM)               (None, 30, 40)            12960

 dropout_1 (Dropout)         (None, 30, 40)            0

 lstm_2 (LSTM)               (None, 40)                12960

 dropout_2 (Dropout)         (None, 40)                0

 dense (Dense)               (None, 20)                820

 dense_1 (Dense)             (None, 1)                 21

=================================================================
Total params: 33,481
Trainable params: 33,481
Non-trainable params: 0
_________________________________________________________________
```

Model Evaluation Score (MSE): 0.000733



Market Data, Sentiment, and Signals

Tableau Demo

Does media sentiment correlate with market performance?

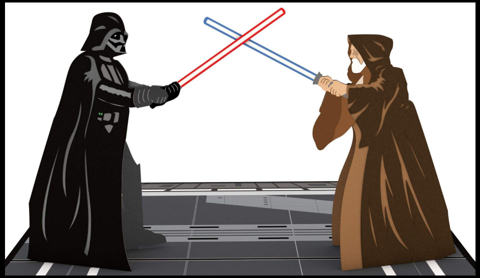
Text Sentiment Analysis Tool

Lexical Approach

Uses dictionary of sentiment

VADER

Does not require training data!
BUT
accuracy is highly variable



Machine Learning

Uses pre-labeled data set

Random Forest (Obi-wan)

Requires training data!
BUT
can attain higher accuracy

Random Forest (ensemble classifier) for NLP

Read in and explore data → Feature Creation → Clean Text → Vectorize → Random Forest Algorithm

Read in and explore data
* Balance of labels
* Initial Cleaning

Feature Creation
* Punct %
* Len(text)
* Transform

Clean Text
* Remove punct
* Tokenize
* Stopwords
* Stemmer

Vectorize
* Count Vectorize
* TF-IDF

Random Forest Algorithm
* Hold Out Test
* Parameters
* Grid Search

Obi-wan NLP VS Vader on Training Data

	precision	recall	f1-score
negative	0.75	0.35	0.47
neutral	0.76	0.96	0.85
positive	0.77	0.48	0.59
accuracy			0.76

	precision	recall	f1-score
negative	0.41	0.30	0.34
neutral	0.74	0.52	0.61
positive	0.40	0.71	0.51
accuracy			0.54

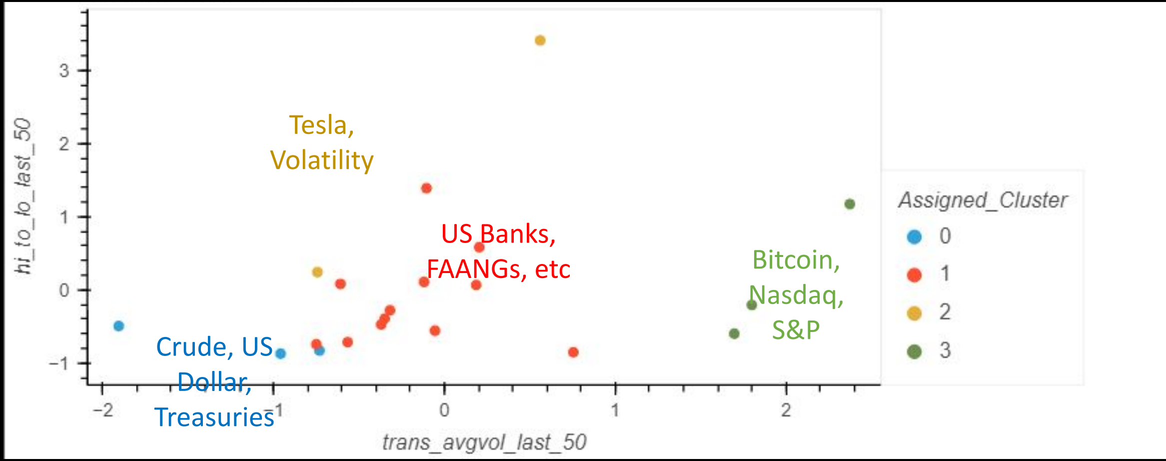
Obi-wan NLP VS Vader on BTC News

neutral	93
positive	7

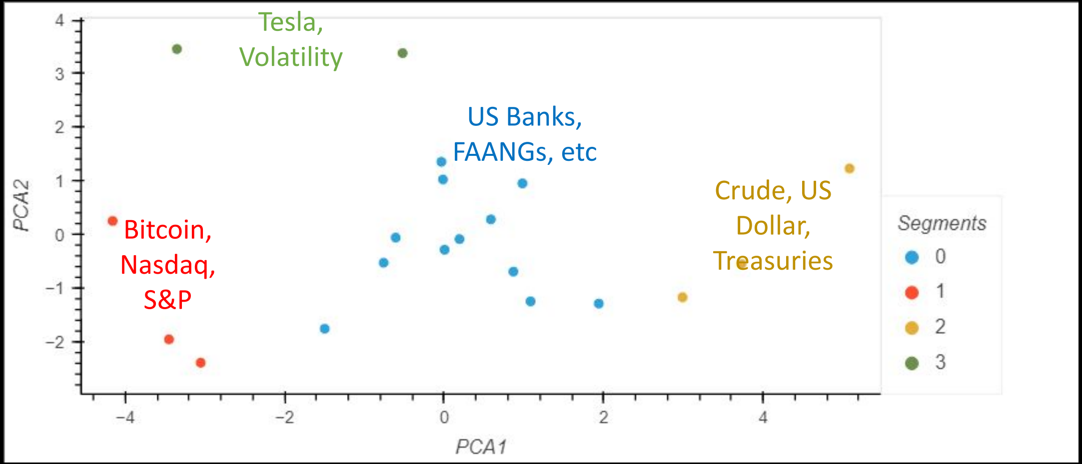
positive	58
negative	31
neutral	11

Why Bitcoin Miner Stocks Soared This Week. Bitcoin mining company stocks have soared this week, amplifying the gains of Bitcoin itself…."

K-Means


Tesla,
Volatility
US Banks,
FAANGs, etc
Bitcoin,
Nasdaq,
S&P
Crude, US
Dollar,
Treasuries
hi_to_lo_last_50
trans_avgvol_last_50
Assigned_Cluster
0
1
2
3

PCA


Tesla,
Volatility
US Banks,
FAANGs, etc
Crude, US
Dollar,
Treasuries
Bitcoin,
Nasdaq,
S&P
PCA2
PCA1
Segments
0
1
2
3

Category	Ticker	Historical Median 50-day Return	Accuracy	TP Rate	TN Rate	Positive Predictive Value	Negative Predictive Value
Bitcoin	***BTC-USD***	***6.4%***	***51.2%***	***20.3%***	***82.1%***	***53.1%***	***50.7%***
Commodities	GLD	0.4%	50.9%	38.3%	63.6%	51.2%	50.7%
	SLV	-0.3%	49.2%	5.0%	93.4%	43.2%	49.6%
	CL=F	2.0%	54.7%	72.3%	37.2%	53.5%	57.3%
Volatility	VIXY	-12.4%	50.8%	60.7%	40.9%	50.7%	51.0%
Market Indices	^IXIC	2.3%	51.7%	31.7%	71.8%	52.9%	51.2%
	^GSPC	1.9%	52.4%	66.8%	38.0%	51.8%	53.3%
	^DJI	1.6%	51.1%	86.0%	16.1%	50.6%	53.5%
FAANG	META	2.9%	50.0%	89.4%	10.6%	50.0%	50.0%
	AMZN	3.3%	51.2%	63.9%	38.5%	50.9%	51.6%
	AAPL	3.6%	52.9%	63.1%	42.7%	52.4%	53.6%
	NFLX	4.5%	51.5%	78.4%	24.5%	50.9%	53.1%
	GOOG	2.6%	53.2%	78.9%	27.4%	52.1%	56.5%
Tesla	TSLA	2.2%	49.7%	98.9%	0.5%	49.9%	33.3%
US Banks	JPM	2.7%	51.2%	55.4%	47.0%	51.1%	51.3%
	WFC	0.5%	53.8%	55.9%	51.7%	53.7%	54.0%
	C	2.1%	51.2%	61.2%	41.2%	51.0%	51.5%
	BAC	2.8%	48.2%	79.9%	16.4%	48.9%	44.9%
US Dollar	UUP	0.5%	49.3%	3.4%	95.3%	41.9%	49.7%
Treasuries	IEF	0.2%	52.2%	10.3%	94.2%	63.9%	51.2%

Created three logistic regression models to predict ***if Bitcoin will outperform it's historical median 50-day return over the next 50 days***:

- **<u>Model 1 - Historical Trading</u>**
  - Used historical trading metrics (over the past 1, 5, 15, and 50 days):
    - Volume
    - Return
    - Price Volatility
- **<u>Model 2 - Additional Features</u>**
  - Incorporated additional features including:
    - Consumer Sentiment Index (CSI)
    - Gallup Institutional (president, police, banks, big business) Trust Polling
    - Consumer Price Index (CPI)
    - US GDP
    - Crypto Tweet Sentiment Analysis
- **<u>Model 3 - Lower Threshold</u>**
  - Lowered the decision threshold in order to increase 'positive' scenarios at the expense of some predictive accuracy (resulting in higher volume of 'positive' trades with lower accuracy)

Metric	Model 1 (Historic Trading)	Model 2 (Additional Features)	Model 3 (Lower Threshold)
Accuracy	52.1%	52.1%	55.1%
True Positive Rate	19.0%	19.0%	63.9%
True Negative Rate	85.2%	85.2%	46.4%
Positive Predictive Value	56.2%	56.2%	54.4%
Negative Predictive Value	51.3%	51.3%	56.2%

Many more ‘positives’ detected at the cost of only slightly lower ‘positive’ accuracy

Additional features in Model 2 do not provide any lift in predictive power…

Market Data, Sentiment, and Signals

The End!!