



Crypto Headlines as Predictors

Capstone / Presented by: Marvin Lee



Agenda

1
Context

2
Pre-Processing

3
Modeling

4
Results

5
Recommendations



problem:

Bitcoin and other cryptocurrencies represent a new asset class in the world, and with the rapid developments in the market investors & venture capital firms seek reliable methods for understanding the sentiment of the market and predicting price movement.

goal:

To analyze crypto headlines and determine whether they can be used to predict weekly price movement



About The Data

Litecoin Market Data: 2014 - 2021

CoinAPI + Apipheny



**Headlines From Top Crypto Media
Cryptocurrency News.**

Crypto Briefing. **Coindesk.**



Approach: NLP + Modeling

Targeted weekly price change
(engineered feature) **with
focus on precision score**

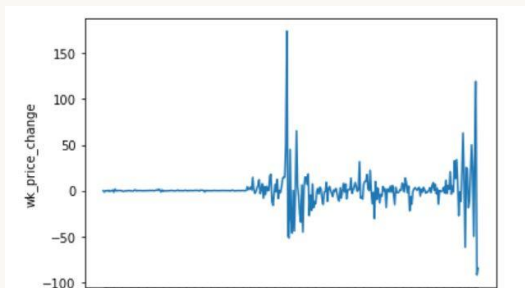


One Key Factor:

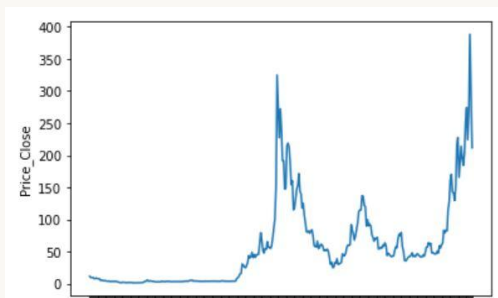
The impact of words on
both upward and downward
price movement



Pre-Processing + EDA



Weekly Price Change **vs.**
Weekly Price (2014 - 2021)



	headline	date	v2_headline
63	How Does Bitcoin Work? Here's a Simple Explana...	2014-01-09	bitcoin simple explanation
267	3 Altcoins That Have Potential To Be The Next ...	2017-07-03	altcoins potential next ethereum bitcoin
410	Bitcoin fan who held up sign during Yellen's t...	2017-07-13	bitcoin fan held sign testimony receives effort
269	Bitcoin Fan Holds Up Sign During Yellen's Test...	2017-10-31	bitcoin fan holds sign receives donations effort
358	Bitcoin mining USB devices on a large USB hub.	2017-11-01	bitcoin mining usb devices large usb
...
51	Bitcoin News Roundup for May 27, 2021	2021-05-26	bitcoin news roundup may
50	Bitcoin Price: Still Room to Fall?	2021-05-26	bitcoin still room
49	MicroStrategy's Saylor Details Meeting Where B...	2021-05-26	saylor details meeting bitcoin mining council ...
76	Why China's Crackdown May Make Bitcoin Mining ...	2021-05-26	crackdown may make bitcoin mining centralized
0	BTC Price Index and Live Chart -- CoinDesk 20	2021-05-26	btc price index live chart coindesk

Removed stopwords, and
aligned each headline
with corresponding week
of crypto market data.

Modeling Results

Baseline: Logistic Regression

Venus

```
In [18]: 1 scoring(y_test, log_test_preds)
```

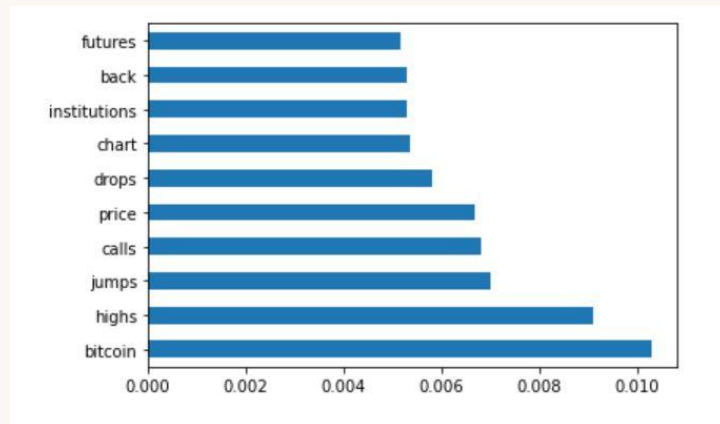
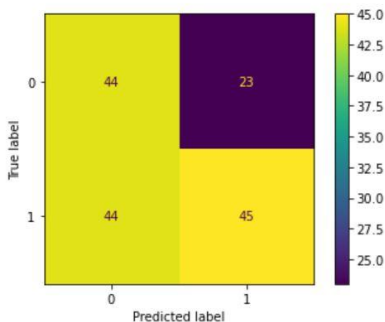
Accuracy score is: 0.5705128205128205

F1 score is: 0.5732484076433122

Precision score is: 0.6617647058823529

Recall score is: 0.5056179775280899

```
In [19]: 1 plot_confusion_matrix(log_reg, X_test, y_test)
2 plt.show()
```



Key Approaches:

- Vector reduction
- GridSearchCV (XGBoost)
- Estimator Reduction
- Minimize False Positives

Best Score: 0.53227 (Precision)

Modeling Results

When Upward

Most frequent words:



When Downward

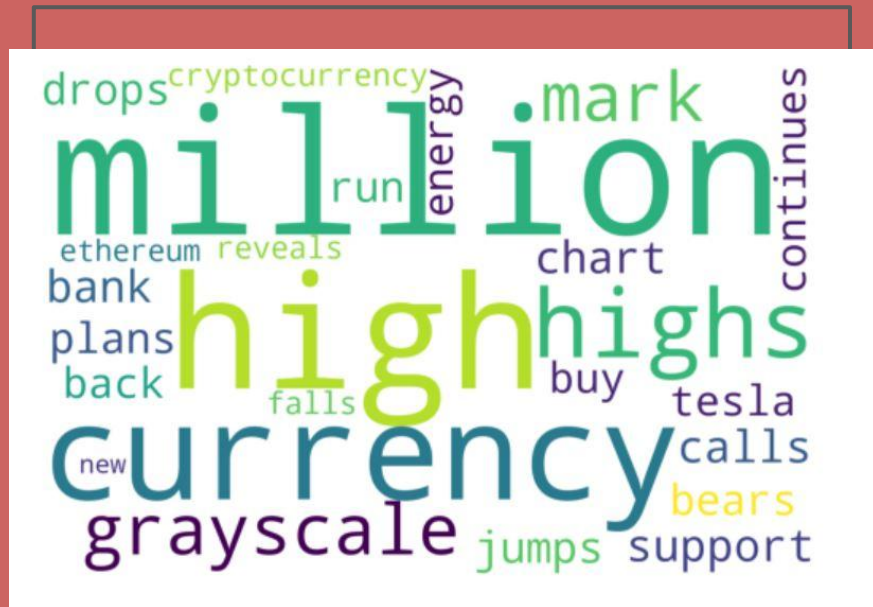
Most frequent words:



Higher Impact Words (on model)

Observations:

- Verb prevalence
- High/Highs more impact than 'Bitcoin'
- Speculation-focus vs. Technology or Regulatory focus





Going Forward

Conclusions:

- Frequency != Importance
- Fewer estimators is better
- Lack of historical data has an impact on predictive power
- But, it does appear word significance can be determined
- Lack of data leads to higher fluctuations in results

Recommendations:

- Analyze both crypto & non-crypto publications
- Use additional relevant keywords (i.e. crypto)
- Narrow analysis / prediction timeframe
- Consider article summaries or meta descriptions



Thank You!

Do you have any questions?

www.github.com/emel333

<https://www.linkedin.com/in/marvinleemills/>

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik**.

