

# Gauging Market Reaction Based on Twitter Sentiment

# About

I would like to see if it is possible to create an application to test if public sentiment expressed on Twitter has an impact on a publicly traded company's value.

# Part One: Data

# 3 Sources of Data

1. Kaggle (For training sentiment models)
2. Twitter API (Acquiring tweets for sentiment analysis)
3. IEX Developer Platform API & Yahoo Finance (Financial Information)

# 1. Kaggle Data

- 1.4 million tweets rated either 0 (neg) or 4 (pos)
- 50/50 Training split of 0's (535984) and 4's (536016)
- Used for training sentiment classification models
- Actually sentiment140 data (or very similar)

## 2. Twitter Data

- API request returns JSON file with 100 tweets for given date time and search term. Limited to 250 requests with sandbox premium.
- Search term for API was “[Ticker Name] OR [Company Name]”
- Request 30 days for company we want to analyze, total of 3,000 ( $30 * 100$ ) tweets
- Mean sentiment was predicted for each dates 100 tweet batch
- CSV exported containing 30 dates and 30 mean sentiment scores

### 3. Stock Market Data

- Stock market data was used later in my analysis to test multiple hypotheses.
- For the analysis stock data and tweet sentiment were joined by their date
- Used two primary financial indicators:
  1. Daily Return
  2. Close Price

## Part Two: Model

# Multinomial Naïve Bayes: Preprocessing

- 2 Step Preprocessing:
  1. Regex Cleanse (tons of emoji)
  2. Vectorization
    - Unigram
    - TfIdf weighting
    - Doc freq = 5
    - Stop words = ‘English’

```
' [ ^a-zA-Z ] '
```

# Multinomial Naïve Bayes: Evaluation

- Model accuracy with holdout test (1/3 split): 77.29%
- Model accuracy with 3-fold cross validation: 77.38%

P->	0	4
0	202845	61171
4	58706	205278

# Multinomial Naïve Bayes: Evaluation

Most Negative Features		Most Positive Features	
-5.47	sad	-5.58	time
-5.42	like	-5.51	going
-5.42	want	-5.49	like
-5.38	today	-5.48	lol
-5.35	day	-5.20	day
-5.29	miss	-5.13	thanks
-5.16	don't	-5.11	love
-5.10	just	-5.09	just
-5.02	work	-4.97	im
-4.64	im	-4.85	good

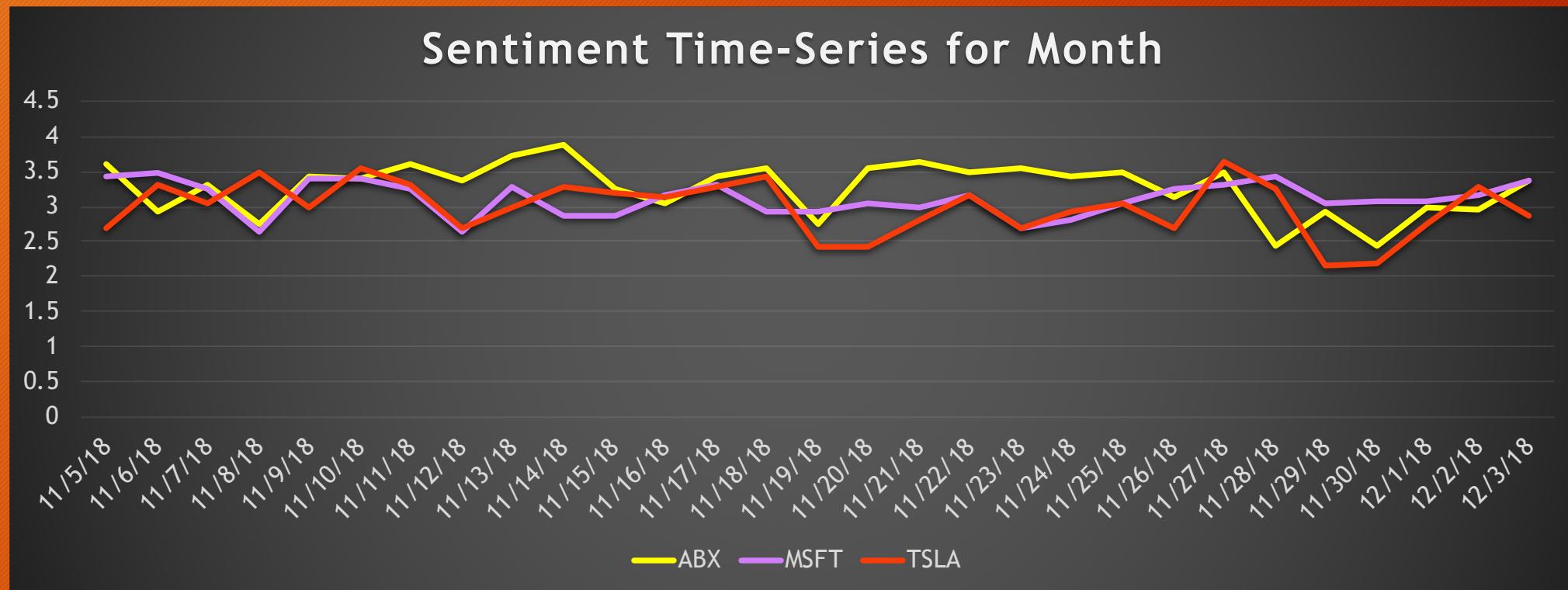


# Multinomial Naïve Bayes: Real Performance

	text	Sentiment	Ground Truth
0	RT historylvsclub Albert Einstein and Nikola Tesla httpstcozmmSRV	4	4
1	RT ericuman Ah so the plan back in the day was to turn Moscows Stainera Sovi	4	0
2	A whir of robots and machines is only occasionally disrupted by the sound of hu	4	0
3	RT elonmusk You can summon your Tesla from your phone Only short distance	0	4
4	RT elonmusk You can summon your Tesla from your phone Only short distance	0	4
5	<b>RT nytimes Swastikas slurs menial tasks and barriers to advancement Many bla</b>	4	0
6	RT orthereaboot Anyone hear anything from tsla about giving DansDeals full bl	0	0
7	<b>RT ayeletw Hey elonmusk what the hell Swastikas Slurs and Menial Tasks Man</b>	4	0
8	RT Inelonwetrust teslaunplugged I have always had the same pic Friend tesla	4	4
9	RT wonderofscience The whole Earth reflected on the side of a cherry red Te	4	4
10	TeslaMagnezone UR	4	4
11	RT teslaownersSV Anyway to help the tesla community httpstcosMewGNRZCx	4	4
12	wonderofscience elonmusk SpaceX Tesla joerogan I always wondered what wa	0	4
13	RT wonderofscience The whole Earth reflected on the side of a cherry red Te	4	4
14	Anyway to help the tesla community httpstcosMewGNRZCx	4	4
15	RT RodneyHansen Thanks Tesla for keeping my blue suede shoes dry #summ	4	4
16	i love TSLA stock	0	4

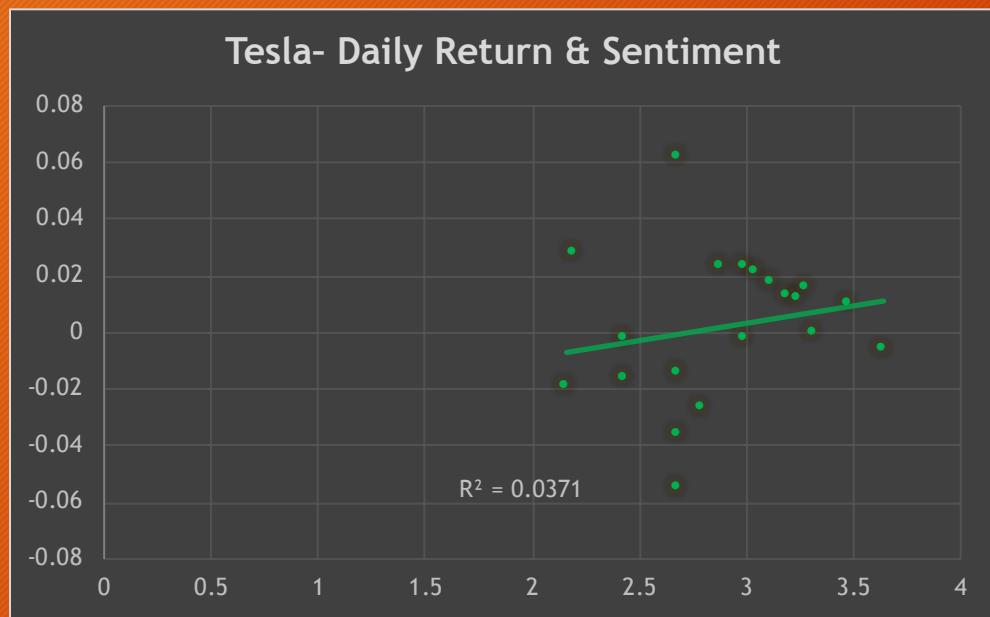
## Part Three: Analysis

# Monthly Sentiment Report



# TSLA: Tesla

## Daily Return Regression

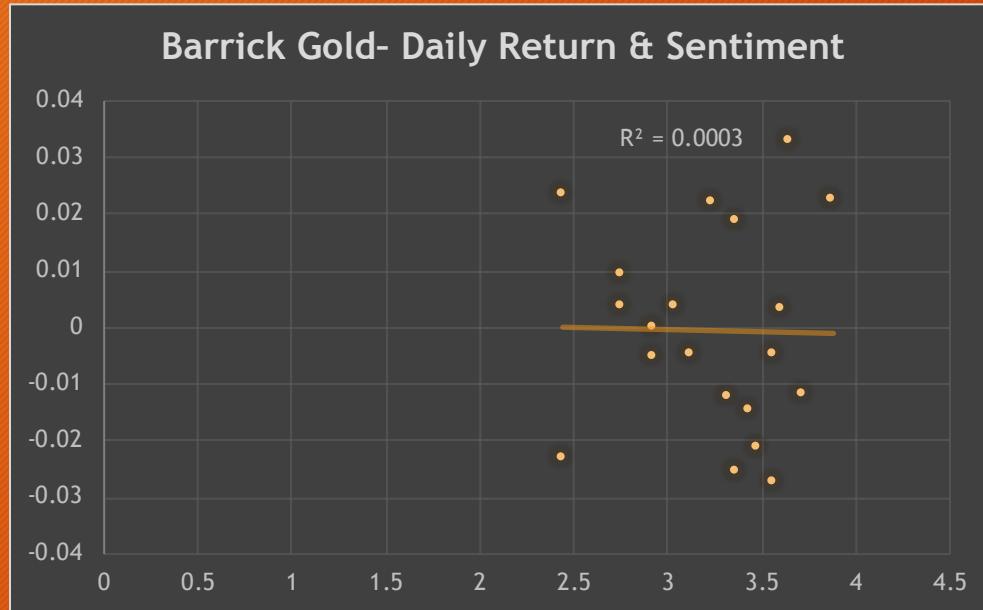


## Close Price Regression

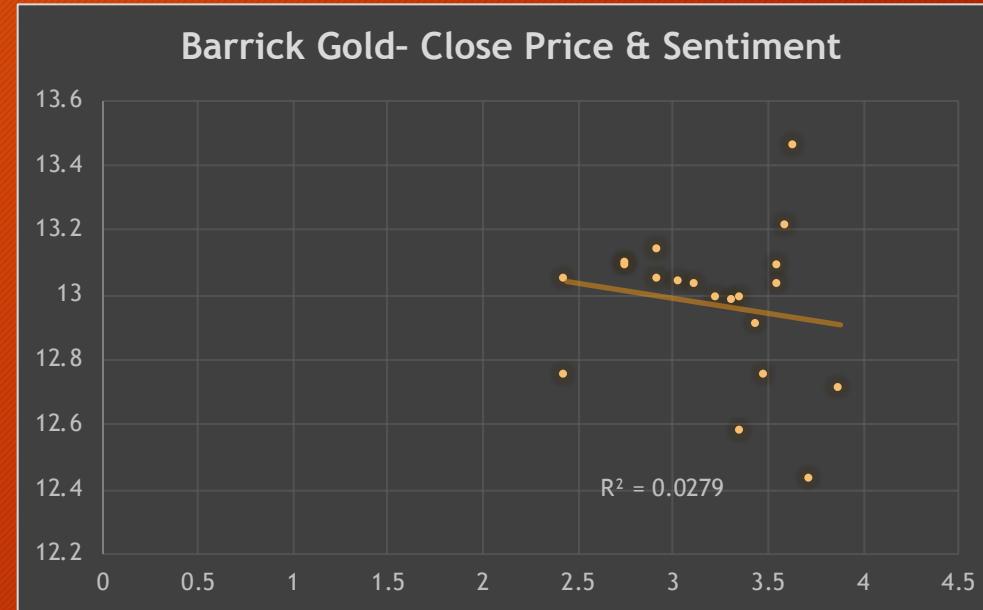


# ABX: Barrick Gold

## Daily Return Regression

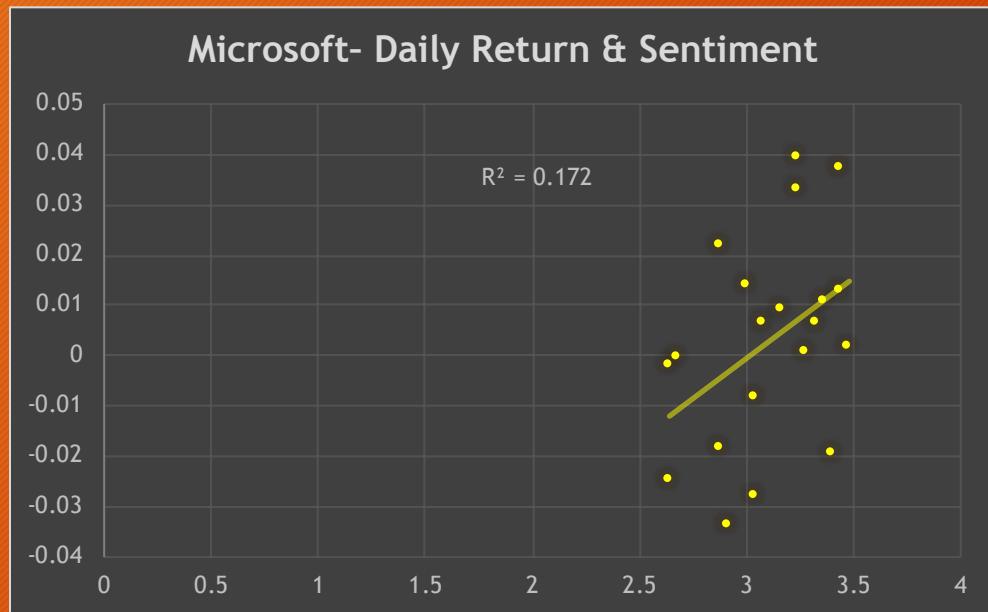


## Close Price Regression

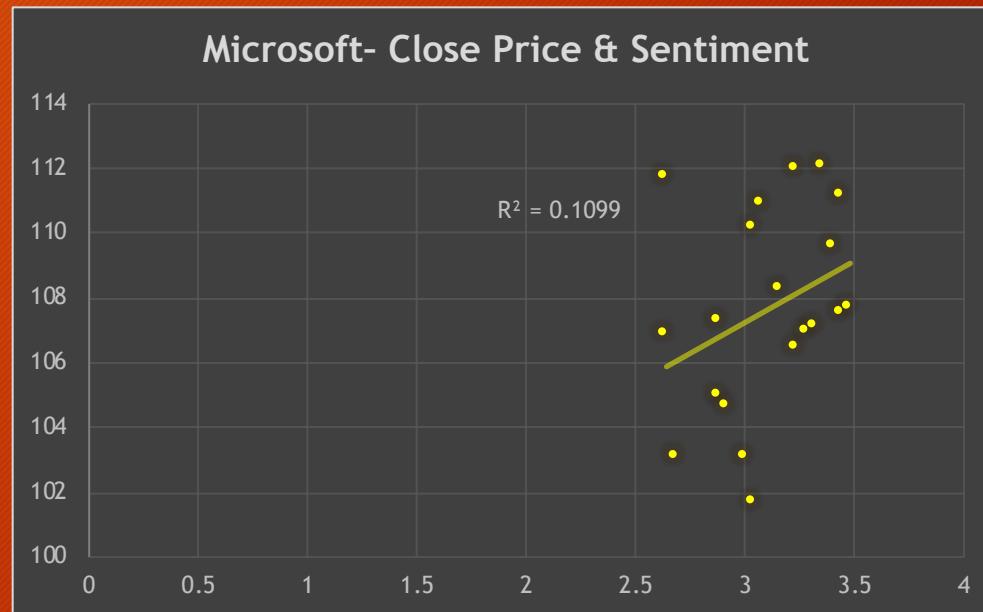


# MSFT: Microsoft

## Daily Return Regression



## Close Price Regression





Questions?

# Sources

- <http://help.sentiment140.com/for-students>