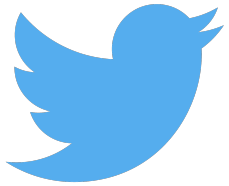
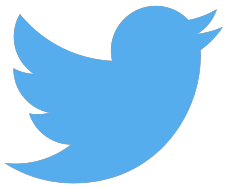
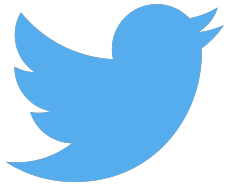


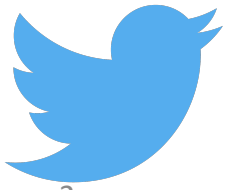
Can we use Twitter sentiment to predict Bitcoin prices?

Gordon Tveito-Duncan, May 2018





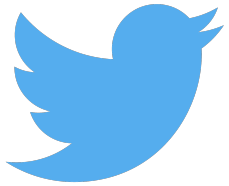
1. Problem



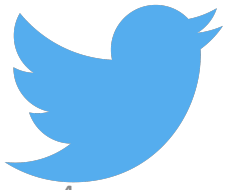
$y = BTC, x = Sentiment$

1. Is there a correlation between Twitter sentiment and BTC price? $Corr(x, y) \neq 0$?
2. Can a naive prediction model based on sentiment produce better than random accuracy?

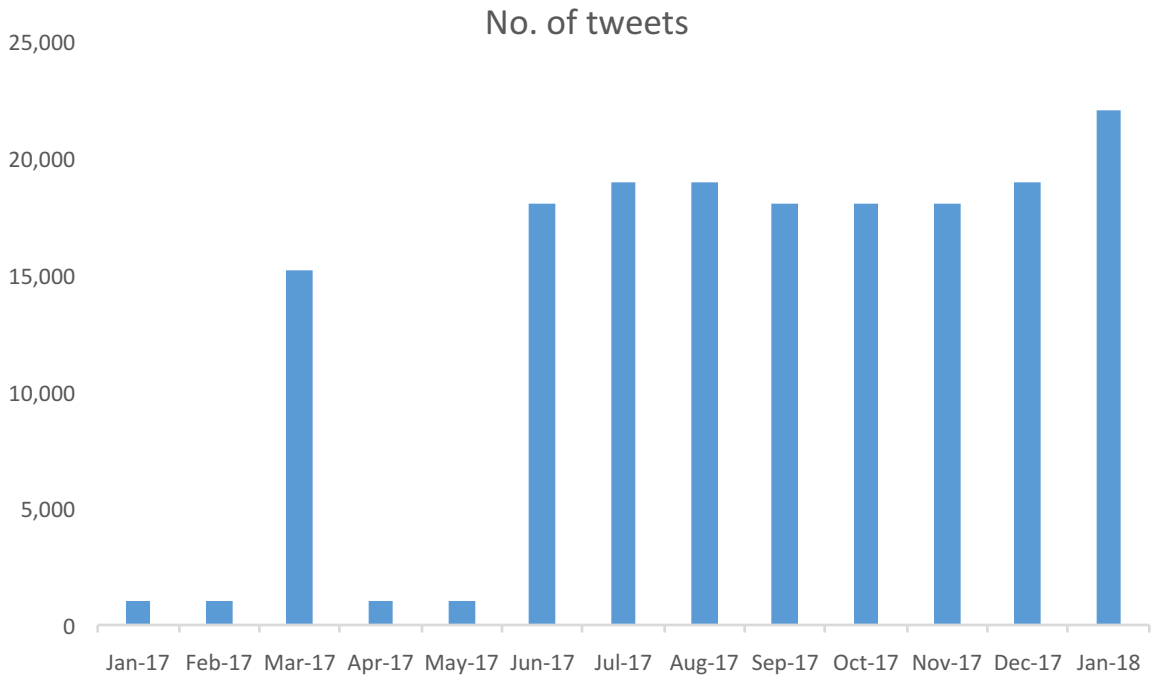
Source: Exante (financial broker)



2. Data



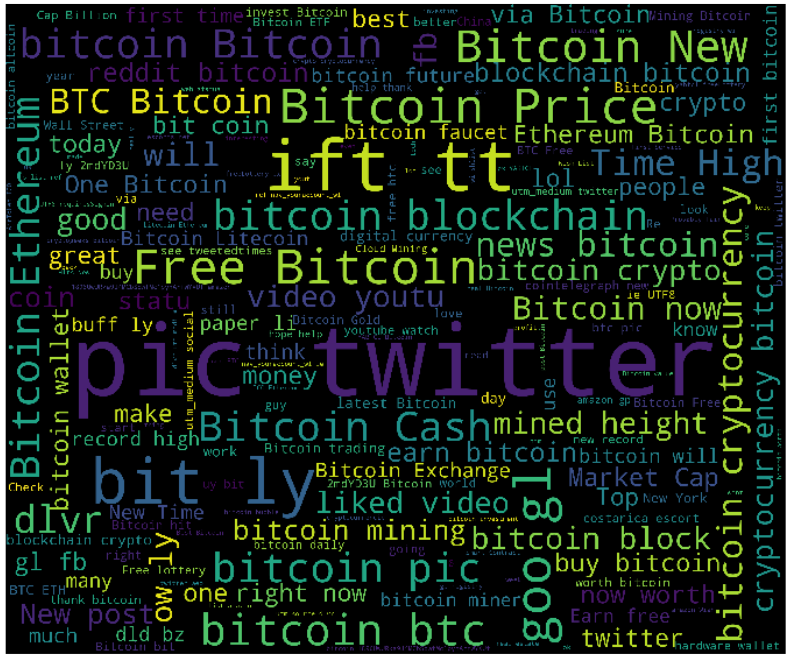
Twitter data	Bitcoin data
Source: Jose's friend No. of tweets: 170,093 (after dropping 100 missing tweets) Date range: 02/01/2017 – 31/01/2018	Source: Coindesk, daily close price ($)

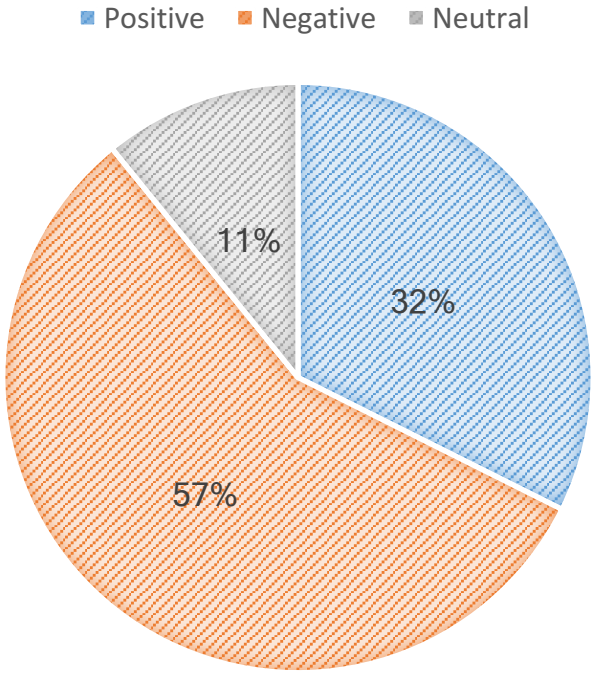


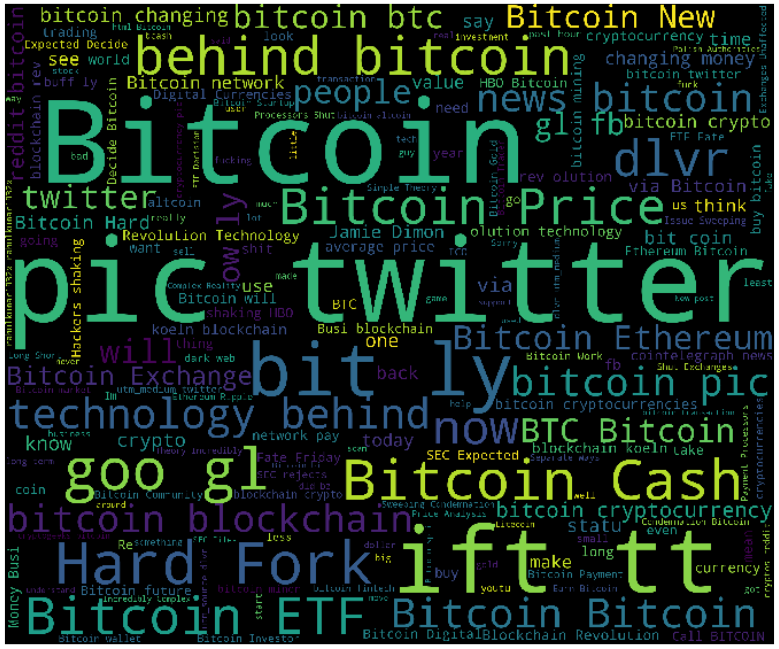
3. Method

Sentiment Analysis

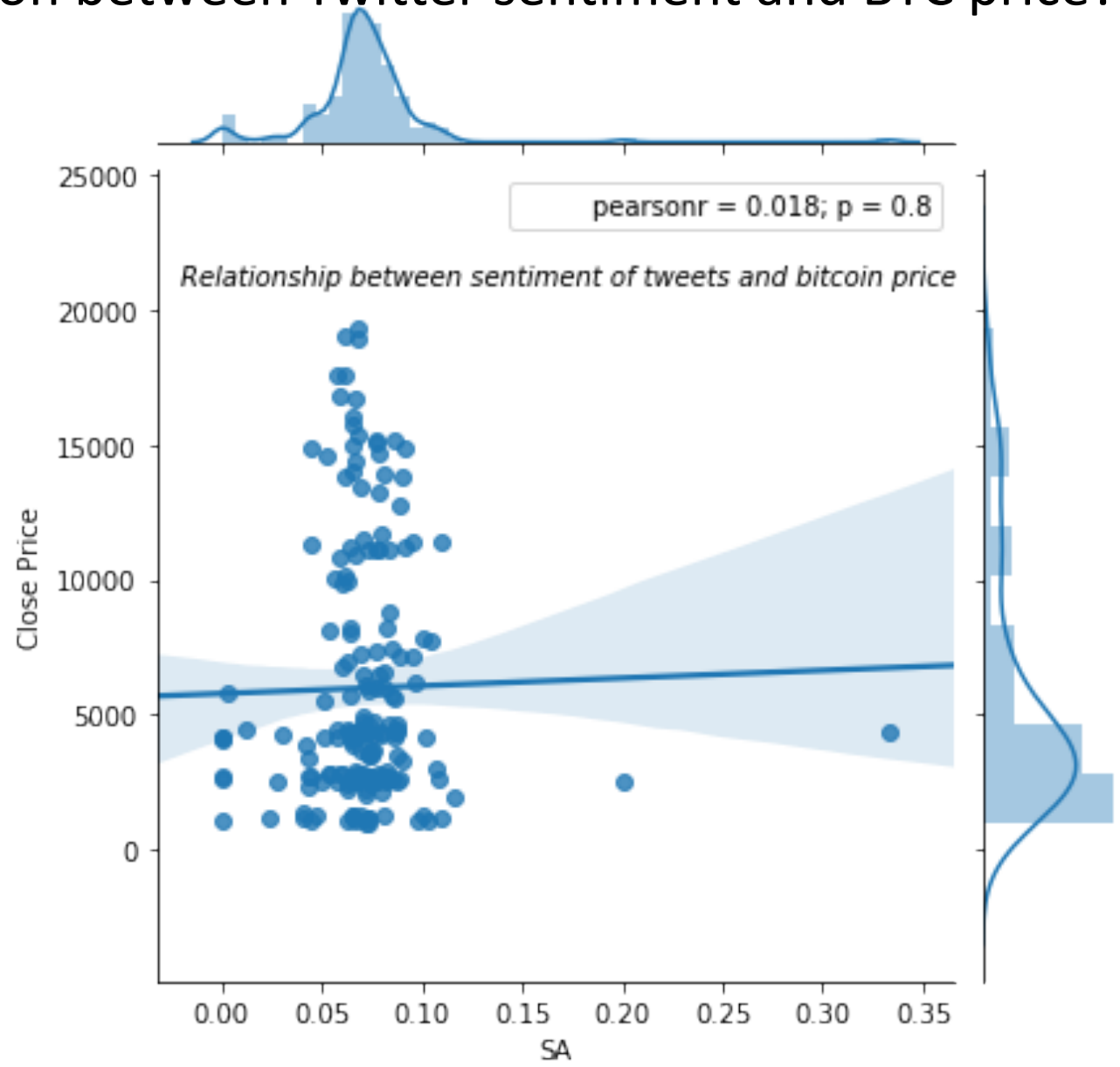
Positive tweets (polarity=1)	Neutral (polarity=0)	Negative tweets (polarity=-1)
"Bitcoin becomes best-performing currency 2016" *"Bitcoin is looking real good now after you see interest rates spike in a DECLINING economy!"*	*"Scammers demand iTunes giftcards, bitcoin as loot"* *"Bitcoin World the future that is already here"*	*"Bitcoin is on pace to have its worst week since 2015"* *"Bitcoin is a terrible investment"*





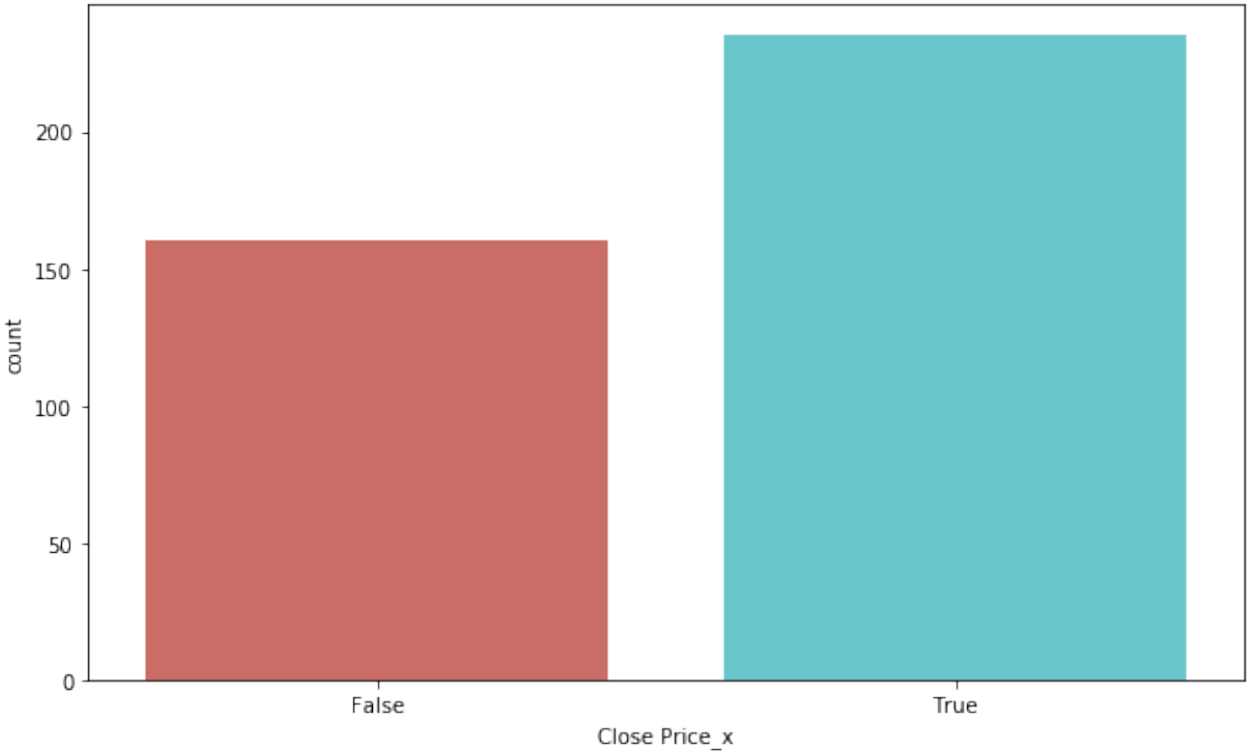


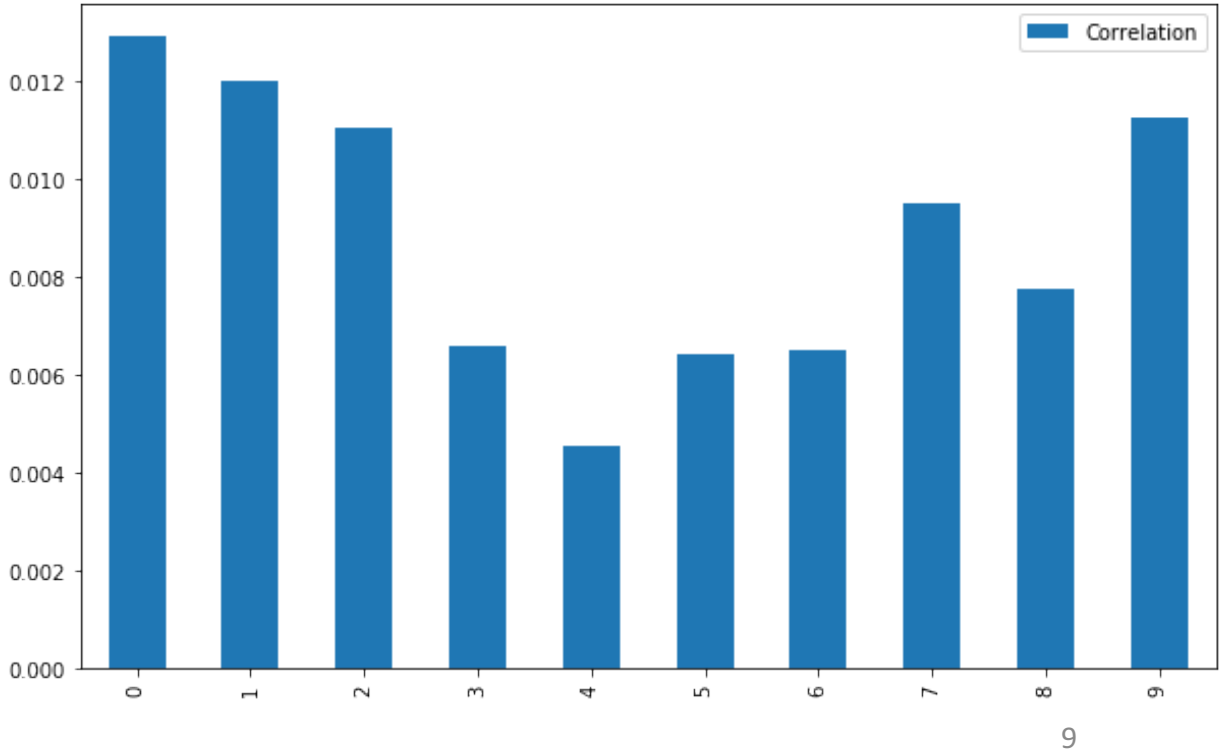
Is there a correlation between Twitter sentiment and BTC price? $Corr(x, y) \neq 0$?



Final Preprocessing

1. Transformed BTC price into binary outcome, where $True = price\ increase, False = price\ decrease$
2. Cross-correlation: examine which lag of sentiment is most correlated with BTC price, where $0 = SAt\_1, 1 = SA_t\_2, 2 = SA_t\_3,\ and\ so\ on$ …

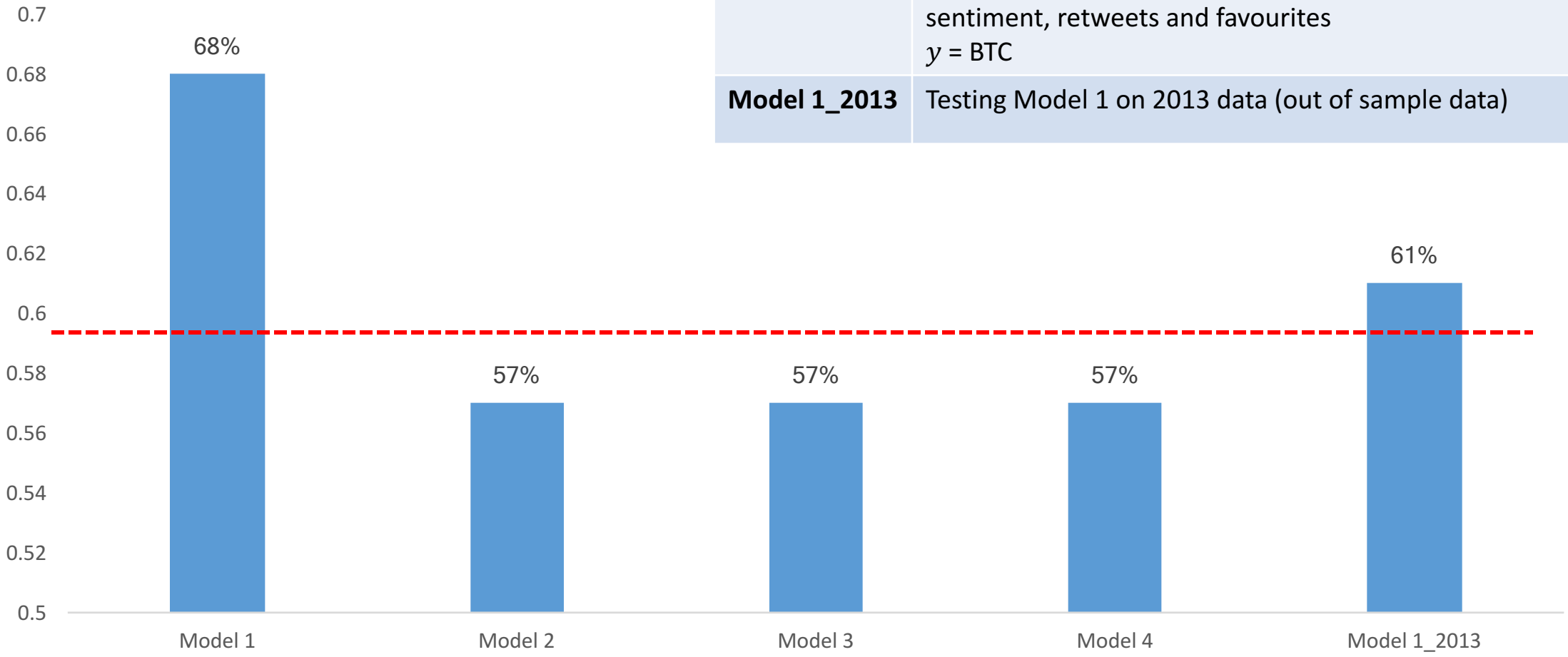




4. Results

Logistic regression: predictions

Model 1	x variable(s): average sentiment, y = BTC
Model 2	x variable(s): average sentiment, shifts of sentiment y = BTC
Model 3	x variable(s): average sentiment, retweets, favourites y = BTC
Model 4	x variable(s): average sentiment, average sentiment, shifts of sentiment, retweets and favourites y = BTC
Model 1_2013	Testing Model 1 on 2013 data (out of sample data)



5. Future Work

Data	Method
• Get minute by minute data, over longer period of time • Clean data more: try to filter out automatically generated content or tweets trying to sell things • Include more variables: number of users, average transaction value (Fundstrat) • Use only influencers' tweets: "verified", many followers	• Improve sentiment classifier • Convert sentiment into categorical but only class as "positive" or "negative" if there is a significant change i.e. threshold of 2.0% (Stenqvist & Lonno, 2017 • Explore how quickly sentiment spreads through network (Sul et al, 2016) • Explore time series more

6. Lessons Learned

1. Preprocessing very important
2. Plan analysis & be more focused
3. Things took longer than I thought