Bitcoin Sentiment Analysis

Team 10 Presentation

Chris Messer

Consultant at Deloitte

Erin Abbott

Business Consultant

Devyn Byrd

Comp Sci Grad Student

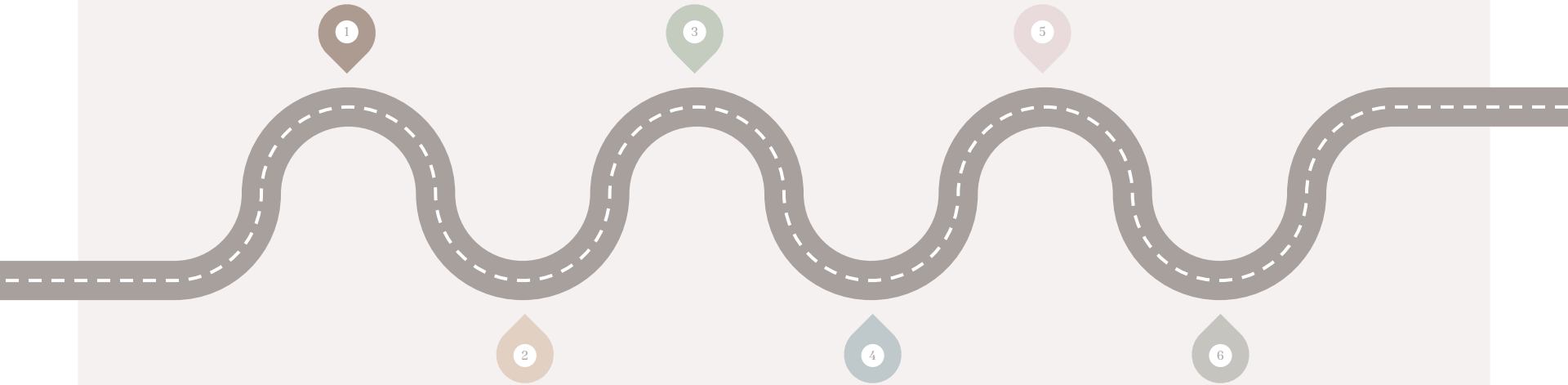
Agenda

- Project Background
  - Bitcoin
  - Sentiment Analysis
- Hypothesis
- Data Preparation
- Planned Approach

Project Background

What is bitcoin? What is Sentiment Analysis?

Bitcoin


2009 - Bitcoin launch
1
2010 - First bitcoin transaction - “Pizza Day”
2
2011 - Bitcoin reaches $1 USD in market value
3
2013 - Bitcoin reaches $1B in market cap
4
2015 - Bloomberg magazine features bitcoin, becomes a mainstream commodity
5
2021 - Bitcoin reaches all time high of $61K
6

Sentiment Analysis

Extract

Extract the data we would like to analyze

- Twitter Data
- Reddit Data

Tokenize

Tokenization in NLP is the process of breaking up text into smaller pieces called tokens. The tokens can then be quantitatively transformed and analyzed

Analyze

Transform the tokens into quantitative measures using Natural Language Processing and Machine Learning to extract the sentiment of the text

Choosing a Sentiment Model

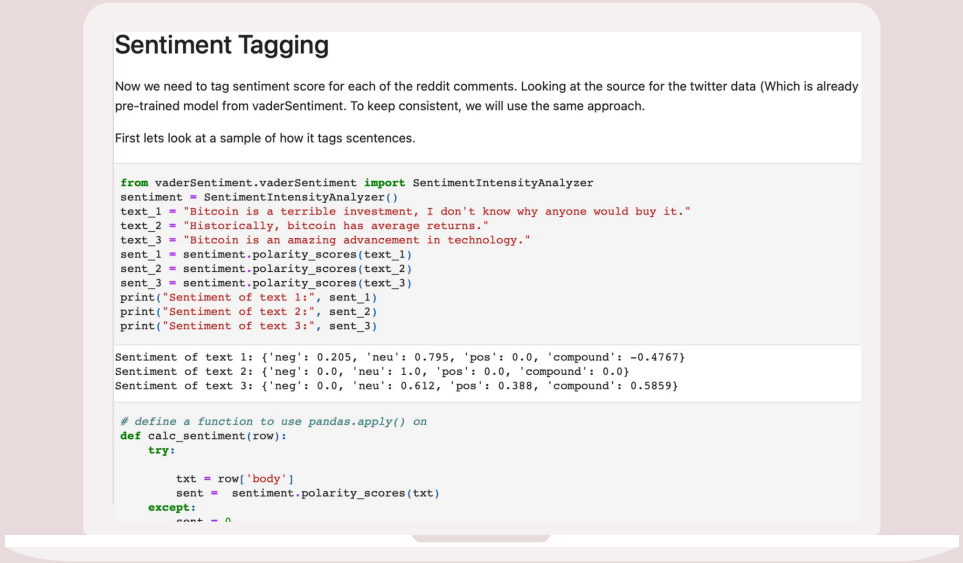
- Different Approaches
  - Neural Networks
  - Rule Based
  - Machine Learning
  - Hybrid Approaches

Example

“Bitcoin is a terrible investment, I don't know why anyone would buy it."

—->

Sentiment of text:
{'neg': 0.205,
 'neu': 0.795,
'pos': 0.0,
 'compound': -0.4767}

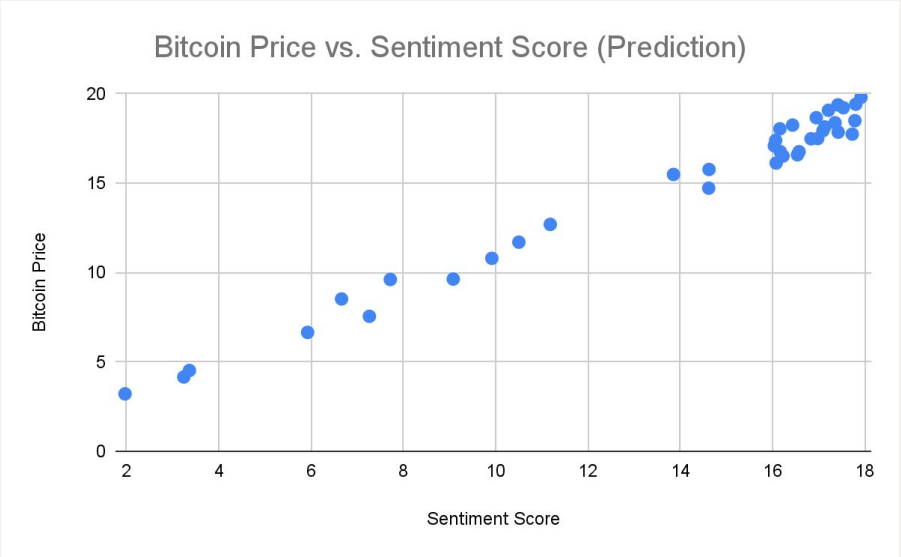


Hypothesis

How do bitcoin and sentiment around bitcoin correlate?

Hypothesis

- Positive correlation between sentiment of tweets/posts and Bitcoin price



Data Preparation

How should we transform the data?

Reddit Sentiment Analysis

Extract

The data we will use for this analysis has already been prepared. See sources cited.

Transform

Several data transformation steps must occur:

- Align the date range
- Remove Symbols, URLS, numbers
- Language standardizati on

Tag

We will use the VaderSentiment Tagging package, as the it was used in the twitter data set.

Aggregate

Lastly, aggregate the reddit comments at various intervals to use it in the final regression model.

Reddit Sentiment Analysis Results

Jupyter Notebook

In [5]:
```
import os
import pandas as pd
import transformers
```

In [6]:
```
# bring in the data
btred= pd.read_csv('../Data/bitcoin_reddit_all.csv')
```

```
/var/folders/f3/g1154c6d5hb7y4_s6bzsyw8m0000gn/T/ipykernel_65866/1397401860.py:1: DtypeWarning:
ecify dtype option on import or set low_memory=False.
  btred= pd.read_csv('../Data/bitcoin_reddit_all.csv')
```

Data Prep

Let's first inspect the data

In [7]:
```
btred.head()
```

Out[7]:

	Unnamed: 0	datetime	date	author	subreddit	created_utc	score	controversiality
0	0	2014-06-26 16:49:41	2014-06-26	HonorConnor	AskReddit	1.403801e+09	162.0	0.0
1	1	2014-05-17 23:30:36	2014-05-17	Talvoren	AskReddit	1.400369e+09	75.0	0.0
2	2	2014-12-18 04:54:48	2014-12-18	ninjoe87	Outdoors	1.418878e+09	-4.0	0.0
3	3	2014-04-30 00:24:16	2014-04-30	fathergrigori54	tf2	1.398817e+09	13.0	0.0
4	4	2014-07-07 16:06:21	2014-07-07	Draber-Bien	starcraft	1.404749e+09	19.0	0.0

Check how many records it is

Planned Approach

What should we do with the data?

Approach

Train	Validate	Test
Model and parameter selection, split the data into train/test/validate sections and train the selected models	Use different models hyperparameters and independent variables to choose the best model	Compare the different models against one another

Train

- Models
  - Linear Regression
  - Logistic Regression
- Parameters
  - Open/close price at t-1
  - Reddit sentiment at t-1
    - Aggregate by subreddit?
  - Twitter sentiment at t-1
  - Reddit/Twitter content volume

Train (contd.)

- Train/Validate/Test Split
  - Time series data
  - 60/20/20 split on time series

Validate

- Evaluation of the models
  - Linear Regression
    - Adjusted R^2
    - Elastic Net Variable Selection
  - Logistic Regression
    - ROC & AUC
    - EN Variable Selection
  - Hetero Skedacity

Test

- Compare models against one another
  - How often is each model correct?
  - How correct was it?

Sources

Cambria, E., Das, D., Bandyopadhyay, S., & Feraco, A. (Eds.). (2017). Chapter 1: Affective Computing and Sentiment Analysis. In *A Practical Guide to Sentiment Analysis* (pp. 1–2). essay, Springer.

Cointelegraph. “How and When Did Bitcoin Start? the Complete Bitcoin History.” *Cointelegraph*, Cointelegraph, 21 Oct. 2022, https://cointelegraph.com/bitcoin-for-beginners/the-history-of-bitcoin-when-did-bitcoin-start

Farell, Ryan, “An Analysis of the Cryptocurrency Industry” (2015). Wharton Research Scholars. 130. https://repository.upenn.edu/wharton_research_scholars/130

Social media sentiment analysis for cryptocurrency market ... - arxiv. (n.d.). Retrieved March 27, 2023, from https://arxiv.org/pdf/2204.10185