VGX Sentiment Analysis and Topic Model

# **Adam Curry Summer Term - DSC680-T302 Applied Data Science (2217-1)**

The cryptocurrency market has received a significant amount of attention in 2021. States have put in legislation to regulate Bitcoin and other cryptocurrencies, the entire country of El Salvador has adopted Bitcoin as legal tenure, and companies have even hinted at or begun accepting cryptocurrency as payment for goods and services [1] [2] [3]. In other words, the cryptocurrency craze is not settling down at any point. Twitter activity and Bitcoin sentiment seem to be a strong indicators of this. Some talking heads and financial experts indicate that social media activity can even predict how well certain cryptocurrency coins perform in the marketplace [4]. While Bitcoin dominates the headlines, there are other crypto platforms and technologies that are also gaining traction within the crypto space. The purpose of this paper is to focus on these alternative companies using Twitter activity. More specifically, for a specific altcoin (alternative coin) – the Voyager Digital Token (VGX). There are many layers that define what altcoins are, but for the purpose of this research paper, at its simplest form, an altcoin can be viewed as anything that is not Bitcoin.

Voyager Digital is a cryptocurrency trading platform where users can invest in, swap, and trade cryptocurrency. There are a wide range of tokens offered including Bitcoin, Ethereum, and Litecoin. Think of Voyager as a lesser known Coinbase application. Also included on the Voyager list of tradable digital assets is the Voyager token. This token offers incentives as part of their loyalty program, which allows users the opportunity to gain and earn interest by maintaining a certain balance of various coins [5]. With the understanding that Twitter may help predict price points and altcoin performance, this analysis will seek to understand the sentiment of Tweets related to the Voyager Token and/or the company itself. This will be done by conducting a sentiment analysis in tandem with a topic model of Twitter chatter. No price predictions or behavior analysis will be conducted at this point. The final analysis will only be a topic model and sentiment analysis of Twitter comments. The dataset’s population will be Tweets ranging from January 2019 – May 2021. The Tweets must have at least one "like” and they all must contain “$VGX” and/or “investvoyager” (Voyager’s Twitter handle) within the Tweet.

Some of the questions this analysis seeks to answer include:

* People often describe a cryptocurrency’s success based on the community within social media. Has the voyager token’s popularity increased over the last two and a half years?
* What is the sentiment of the Voyager Token on Twitter over the last two and a half years?
* Do people associate their loyalty program with a positive or negative sentiment (i.e., is their business model favorable among Twitter users)?
* Has the sentiment increased in a favorable direction with the recent cryptocurrency bull market?
* What topics are most discussed concerning Voyager Digital?

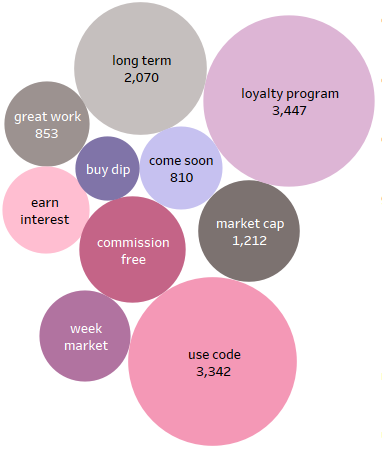
Some questions I anticipate from more technically minded individuals include:

* Were you able to get enough data from Twitter through their API?
* What library did you use to get the data from Twitter?
* What preprocessing steps did you take with your dataset?
* Did you lemmatize your unstructured data?
* Did you use n-grams in your topic model?
* Were you able to apply machine learning to get an accuracy score on your sentiment analysis?
* Was there any correlation and/or EDA conducted in your analysis?
* Have you made any price predictions on the dataset?
* Did you have a large enough dataset to feel comfortable with your final analysis?
* Why did you choose to use two different methods for your sentiment analysis?

The initial dataset was gathered using a Python library to grab three variables from Twitter: the Twitter ID, the date/time it was Tweeted, and the unstructured plain text Tweet. This was done using snscrape.modules.twitter. This library scrapes Twitter opposed to calling it through an API. Therefore, a lot of the Tweets contain weird characters and unusable content. As a result, I had to take several steps to “clean” the Tweets. For example, the following Tweet had gibberish, symbols, hyperlinks, and Twitter handles that would not help the sentiment analysis or topic model.

* This is encouraging... 𝐵𝑇𝐶BTCETH 𝑉𝐺𝑋VGXLUNA 𝑂𝐶𝐸𝐴𝑁OCEANSTMX ðŸš€ðŸš€ðŸš€\n\nhttps://t.co/voDX2UfV7f

I was expecting some “messy” Tweets but didn’t realize the depth of cleaning required to turn the above Tweet into: “this is encouraging btc eth vgx luna ocean”. With help from the NLTK library, I was able to remove punctuation, lemmatize the data and generate a string of text that was machine readable. The lemmatization brings the words within the string to their most basic form of language. This method of text cleaning allows for more robust natural language processing.

The next step involved the topic model analysis using the sklearn library. This required a lot of effort, as I had to do multiple iterations with tweaks to the algorithm. For example, initially I used a single ngram, or contiguous sequence of n items from a given sample of text or speech [9]. However, I eventually settled on 2,3 ngrams, to produce multiple words within a single topic. I even had to go back a few steps to remove irrelevant words that didn’t help the analysis. For example, the following were words that included altcoin abbreviations or gibberish that were not picked up by my initial data cleanse: voyager', 'VGX', 'luna', 'ocean', 'ï¼š', '³ðÿš', 'dot', 'link', 'band', 'aion', 'cro'.

Finally, I was able to produce topics that seemed relevant to the Voyager Token. The bubble chart on the right visualizes the topic model’s top 10 unique topics: “buy dip”, “come soon”, “commission fee”, “earn interest”, “great work”, “long term”, “loyalty program”, “market cap”, “use code”, “week market”. The bubble chart also shows the frequency of these topics within the population. For example, the loyalty program appears to be a topic that is discussed the most frequently among Twitter users.

Now that I had a solid topic model, I moved on to the sentiment analysis. The sentiment analysis used was Vader from the NLTK library, as this algorithm was designed specifically for social media discussion: “VADER (Valence Aware Dictionary and Sentiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media.” [6]. I did this analysis two ways, one using the vaderSentiment.vaderSentiment library and the other using the nltk.sentiment.vader library. Both were from NLTK. I wanted to ensure the sentiment I found was accurate if I ran two separate ways, as I had no way to run any sort of machine learning since the Tweets weren’t pre-classified. In other words, the only way I could classify the Tweets were to manually read them myself.

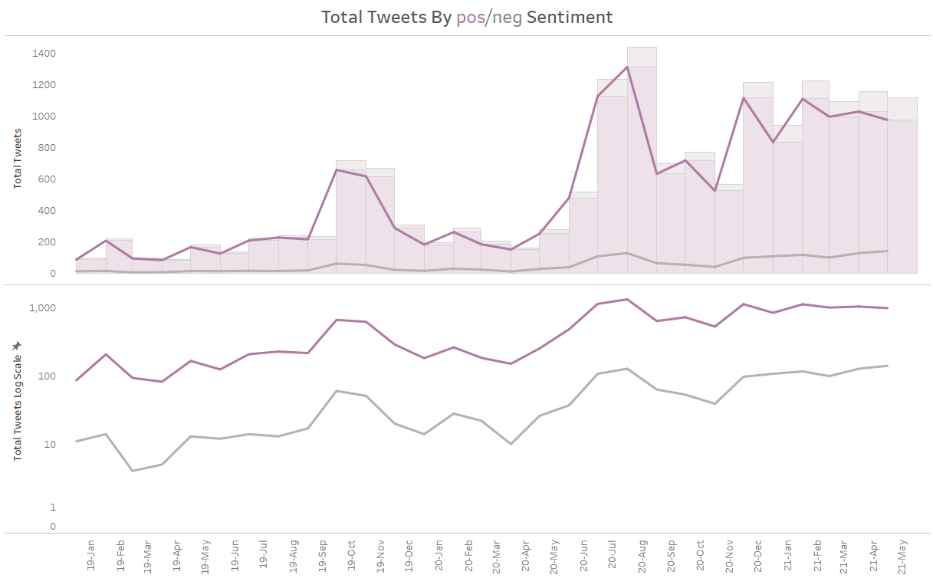
Using some basic SQL queries I was able to compare the two algorithms and the findings revealed very similar results. As you can see, the table in appendix A shows 13 Tweets that weren’t classified the same. When reading through the specific Tweets, I found that the v2 compound score was the most accurate between the two.

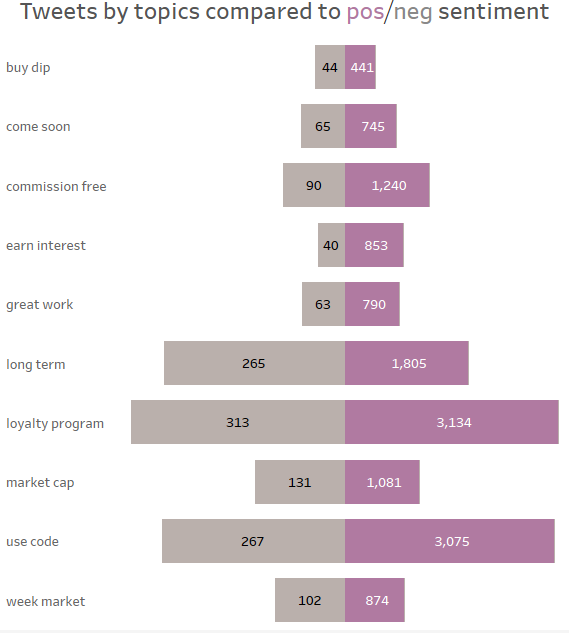
Examining some of the Tweets showed very promising results in the classification. For example, the following Tweets were classified as negative sentiment:

* “@Ehrls15 @investvoyager Stephen. This lack of readiness is costing your customers a massive amount of money as our investments cannot be executed due to your lack execution. This is not going to be good. You have $10K, 5 ETH in pending for over 25 hours and I'm losing money because of you”
* “@presidentLevii @investvoyager Nope. Thats not right. I would never invest anything I canâ€™t lose. Hereâ€™s the thing though...I havenâ€™t lost a cent today...at least not over the last four hours. Itâ€™s been lost for me without my control or consent because Iâ€™ve had no access to it. Big difference, my friend.”

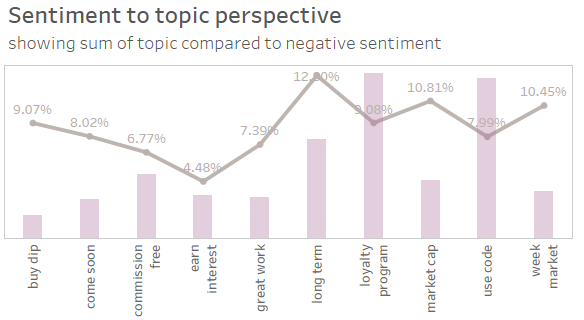
And the following were considered positive sentiment:

* “@TheTradingTramp Without a doubt... This is the time to accumulate the hell out of it... look at 𝑉𝐺𝑋𝑠𝑡𝑜𝑐𝑘 VYGVF. There is NO DOUBT in my mind that the coin price will follow. The completion of 3rd party $LGO audit sends us to a new price discovery zone imo!!”
* “Will be awesome when @investvoyager starts offering IRAs so I can accrue interest tax free on 𝑉𝐺𝑋 USDC $ADA #ETH #BTC etc!”

Diving into the sentiment analysis revealed that Tweet frequencies move in similar patterns with positive sentiment (top chart). However, it is interesting that the negative sentiment appears to be moving up in the latter months. This could be an indication of the recent volatility in the crypto market. The bottom chart shows a log transformed graph to visualize the pattern of movement of negative and positive sentiment. As you can see, they are both near identical trend lines, but similar to the top chart the latter months seem to indicate a higher trending negative sentiment and a lower trending positive sentiment.

Additional analysis was done to visualize how the topics compared to the sentiment of the Tweet. The chart on the right shows the negative vs. positive relationships between the two. For example, the “loyalty program” appears to have received the highest number of positive Tweets sitting at 3,134. This is an interesting finding, as the loyalty program is what increases the user’s interest earned on various coins and tokens. The “buy dip” received the least positive Tweets sitting at 441. My initial reaction is that not a lot of people are “bullish” when the price of the coin is low. To put it another way they may be a bit hesitant to invest when there is a dip in the crypto market.

Looking at the above chart doesn’t tell the whole story. When comparing each topic’s sentiment as percent of total, we get a new perspective. For example, the below chart shows “long term” has the highest percentage of negative Tweets by topic at 12.8%. This could be an indication of the token and/or company not being a good long-term investment. It could also be an indication that the crypto community disagrees with the Voyager business model. For example, this video <https://youtu.be/jw2lf-EsWbY> describes some disagreement with Voyager and their way of doing business [7]. To be blunt, there is a crowd of crypto influencers who will pump certain coins for their own benefit. This quote from a former disgruntled crypto project manager says it all:

“Teams spend as much money as they have on getting users into their Telegram groups, enticing them with shitty bounties, meme contests and quizzes. For a while it seems like everyone is in it together. People are excited, looking forward to the token sale, and then looking forward to that first listing on a crypto exchange. They think they’re all going to make money and be rich. But at some point the relationship between project and community falls apart. It can happen before the ICO. It can happen a month after, a year, sometimes two years, but inevitably, with any crypto project and community, it will go to shit. Why? Because there is no real link between the community and the project. It’s all built on bullshit.”

Looking at the highest positive sentiment percentage shows that “earn interest” received 95.52% positive tweets. Also, the loyalty program received a relatively low ratio of negative Tweets. These could be good indicators that the Twitter chatter associates the interest and the loyalty program offered by Voyager in a mostly positive sentiment.

Another interesting data point is the “use code” topic. There is a “refer a friend” program that offers incentives to get friends to join Voyager. This topic received the second highest number of positive Tweets at 3,075, which corresponds to 92.01% of the positive Tweets and 7.99% of the negative Tweets. It makes intuitive sense that this would be mostly positive, but this could also be a concern of the underlying data as bots are often used to influence the price of coins [8].

The “next steps” I could take for this analysis will require a new dataset that contains price points. I also may have to manually review major company announcements and add them to the dataset as variables with indicators. However, I don’t know if that will make it into this first project due to a lack of time. The next project I create may contain this data where some additional exploratory data analysis could be conducted. To take this analysis to the next level I could then utilize the dataset to predict price points, or the other way, to predict sentiment based on price points. Another step could be manually classifying Tweets that will allow a machine learning sentiment to be created. Since there isn’t a pre-classified dataset, it is challenging to classify these Tweets with an accuracy score.

The final analysis produced findings that make intuitive sense for this domain. However, a larger dataset could make this analysis more robust. Unfortunately, this is a newer company in an already heavily saturated market of products. In other words, time and more data are required to take this analysis to the next level.

**Appendix/References:**

1 Austin, J. (2021, May 26). *Cryptocurrency to be regulated in Nebraska*. https://www.1011now.com. https://www.1011now.com/2021/05/27/cryptocurrency-to-be-regulated-in-nebraska/.

2 Broderick, R. (2021, June 10). *Inside the cryptocurrency scam vortex*. The Verge. https://www.theverge.com/22522380/cryptocurrency-scams-hacks-bitcoin.

3 Finegold, E. (2021, March 4). *Mavs to Accept Dogecoin Cryptocurrency for Tickets and Merch*. The Official Home of the Dallas Mavericks. https://www.mavs.com/dogecoin/.

4 Jones, A. (2021, June 23). *Sentiment Analysis of Tweets*. Medium. https://towardsdatascience.com/sentiment-analysis-of-tweets-167d040f0583.

5 Pechman, M. (2021, January 23). *Voyager Token (VGX) gains 926% as mergers and acquisitions bring new users*. Cointelegraph. https://cointelegraph.com/news/voyager-token-vgx-gains-926-as-mergers-and-acquisitions-bring-new-users.

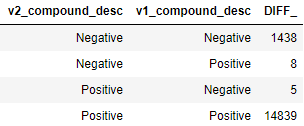
6 *r/Invest\_Voyager - Voyager on BitBoy Crypto Today*. reddit. (n.d.). https://www.reddit.com/r/Invest\_Voyager/comments/m23kr3/voyager\_on\_bitboy\_crypto\_today/.

7 Suban , A. (2021, June 6). *Bitcoin wobbles after Chinese social media site Weibo blocks crypto influencers*. Yahoo! Finance. https://uk.finance.yahoo.com/news/bitcoin-cryptocurrency-ethereum-dogecoin-elon-musk-weibo-china-093212102.html?guccounter=1&guce\_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce\_referrer\_sig=AQAAAG8AX59nTw9rRcqp65xFsj8GmzpU80cJEvWliSC6lb546Z0FsFlGDPVz4F4zPYewqH0aRwP4gKNlqhgjzg5CPCLrFBvV4X-MmKoYARNq5oLKns42cLVIAe971aoydvsutTtfU5RyLCDxeljdYHJmnc4ofRvsRwwXTZRhq8P1CJPV.

8 Team, T. V. (2021, May 1). *Voyager's May Interest (APR) Rates*. NODE from Voyager. https://www.investvoyager.com/blog/voyagers-may-interest-apr-rates/.

9 Vanalli, V. (2020, March 9). *I worked full time in crypto for two years and didn't really like what I saw*. Medium. https://medium.com/swlh/i-worked-full-time-in-crypto-for-two-years-and-didnt-really-like-what-i-saw-39382fa103a7.

10 Wikimedia Foundation. (2021, April 20). *N-gram*. Wikipedia. https://en.wikipedia.org/wiki/N-gram.

Image A: