

# **COSC 89: Final Project**

## **CryptoScan: Watson Cryptocurrency Sentiment Analysis with Natural Language Processing**

Team Neufchâtel  
Jennifer Jain, Xinwei Jiang, Tianyu Li, Weijia Tang

### **Business Problem and Motivation**

The cryptocurrencies have recently received much attention by the media and investors alike, due to the assets' innovative features, potential capability as transactional tools, and tremendous price fluctuations.

Among other factors, excitement and fear of the crowd have greatly influenced the price of cryptocurrencies and made the asset highly volatile and unpredictable. For example, Mark Cuban, the owner of the NBA team Dallas Maverick, trolled on Twitter about how bitcoin is a bubble. Right after his post in June 2017, Bitcoin dropped about 10% in a single day. He was credited later, along with the news, for the tumbling prices. This simply shows the degree of how today's crypto markets are driven by crowd sentiment and psychology. Being able to analyze the market sentiment on cryptocurrencies will definitely help investors make better trading decisions.

However, the problem is that the tools for analysis of market sentiment on cryptocurrencies were only accessible to big institutions like some of the top hedge funds. That's why our team developed a tool called CryptoScan which aims to provide individual investors and institutional money managers alike with a third-party tool that allows users to select from popular cryptocurrencies and receive historical insights on crowd psychology, public opinion and valuation. We have the confident that individual investors can boost their accuracy on their prediction on market and hence gain more investing income.

Watson's deep QA capabilities uses cognitive computing to provide us with a unique opportunity to discover the value of the world's data by doing sentiment analysis on large amounts of unstructured data. Before cognitive computing, each body of text would have to manually be analyzed in order to decide its attitude.

## **Business Plan**

Since CryptoScan improves investment decisions in cryptocurrencies resulting in higher earnings, our team has decided that the ideal business plan charges a premium to users based on how much data and how often they are collecting data from CryptoScan. We have two business plans serving two types of investors. Our first plan is for investors who simply use the software from time to time, we will charge them \$1 for each search. Our second plan is for those who heavily rely their investing decisions on market sentiment, they may purchase the monthly plan, which costs \$50 per month.

## **Organization**

We brainstormed several ideas across different industries at the beginning. We set our mind on doing cryptocurrency because it has been really popular and trendy recently, and we are interested in merging the quantitative and qualitative perspectives of cryptocurrency to a mostly quantitative measurement. After that we came up with the business model and how we want to render our product. We decided to create a website which is composed of two parts, real-time price chart for selected cryptocurrency and market sentiment represented by positivity and negativity.

## **Tools and Knowledge Sources**

CryptoScan uses the Twitter API to scrape through Twitter posts containing cryptocurrency related endpoint keywords resulting in tweets from verified users. These tweets were fed into IBM Watson's Natural Language Understanding (NLU) for sentiment analysis of the related posts. We also used the Bitcoin Average API to obtain historical price data for Bitcoin, Ethereum, Dash, Litecoin, and XRP. The website was designed using the React framework. We used Flask to build the Python API.

## **The Process: Twitter Scraping**

To collect the tweets, we decided to base our program in Python, given its motley of free twitter-analysis packages and foundations within data analysis and statistical computing. Specifically, we used *tweepy* package, which accesses Twitter's REST and stream APIs. We were able to use the *tweepy* package to retrieve, from each of the last seven days, searching from

midnight backwards, a maximum of 18,000 of each tweet that had the terms that are relevant to the cryptocurrency of the user's choice. For our current version of CryptoScan, we fetch the last 1000 tweets based on the user's choice. For example, if the user selects "bitcoin", we will search 1000 tweets with terms "bitcoin", "BTC", and its upper or lowercase variations.

## **The Process: Sentiment Analysis**

We used IBM Watson's Natural Language Understanding API to classify each tweet as positive, negative, or neutral sentiment. For tweets that can't be recognize by Watson, we return a value of "null". Specifically, we generated a Natural Language Understanding instance and reiterate it through the 1000 tweets we fetched. Then we dumped the JSON response the NLU instance generated to fetch the sentiment it returns. Finally, we sum the number of the positive tweets and negative tweets and get each type of tweets' percentage.

## **The Process: Cryptocurrency Market Data**

We fetch cryptocurrency prices using the bitcoinaverage api on a minute basis for a period of 60 minutes, and Twitter API to fetch the most recent tweets based on public attitude for a certain cryptocurrency.

## **The Process: Website Development**

We use the React framework to develop our website. As well as utilize the LineChart and Tweet component of React for displaying the interactive real-time price chart and twitter posts pertaining to certain cryptocurrency.

## **Difficulties and Solutions**

### **Data Overflow**

We initially planned to do sentiment analysis over a longer period of time, like the past day. However, the number of tweets we fetched from Twitter API was too big and it went

beyond our capacity to process all of them within a short time frame. It takes too long to wait for the system to produce results.

Solution: we decided to instead fetch the most recent 1000 tweets from Twitter API. This will allow us to capture the latest real-time market sentiment on cryptocurrencies. And at the same time, our system would be capable of processing all the tweets and return the results without much loading time.

### **Data Relevance**

In the beginning, a lot of the tweets we got were not relevant to the cryptocurrencies. For example, the tweets we got from searching Dash were all about the dash in the football game or the action of dashing out of the room. The results of sentiment analysis would be biased if the data are irrelevant.

Solution: we added filter to the twitter search query. We limited the language to English and only got tweets from verified users. We also added another keyword “#crypto” to every search query. After filtering, most of the tweets we got became relevant to the target cryptocurrency.

### **Integrating Backend with Frontend**

We wrote our sentiment analysis system in Python that does the fetching tweets and using Watson’s NLU. However, our frontend user interface is in Node.js and javascript.

Solution: we built a simple Python REST API server that runs on the localhost along with the website. In this way, the frontend website can send request to the Python server and receive results to display to users.

## **Cryptocurrency Sentiment Analysis Effectiveness**

In its current state, CryptoScan is primarily a platform in which a user can obtain current price and sentiment data, however the platform is not very useful for long-term data to Twitter’s API limits. Thus, the investor or trading group must collect and store their own data using our platform.

CryptoScan displays accurate price data for each cryptocurrency as well as real Twitter posts related to cryptocurrency by verified users. The website also displays an overall negative and

positive percentage score which can be extracted by investors and used within their platform for a larger array of customized applications.

## **Improving the Service**

Instead of investors relying on CryptoSan as an API from which to extract information. CryptoScan can be expanded to include a histogram overlay of long term sentiment percentage on the valuation graph which would provide insights into sentiment and price relationship. Also, allowing the user to select and compare more types of cryptocurrencies together can provide a comparison on the best cryptocurrency to invest in.

Combining this information with a word map of positive and negative words found within twitter posts, signal maps as well as sentiment analysis from various news sources and other social media outlets to increase the ability of trading groups and investors to make informed decisions.

## **Works Cited**

Armin Ronacher. 2011. Flask Snippets. (July 2011). Retrieved March 6, 2018 from <http://flask.pocoo.org/snippets/56/>

Blockchain Data LTD . Bitcoin Average API. Retrieved from <https://apiv2.bitcoinaverage.com/>

bmorelli25. 2018. Interactive Bitcoin Price Chart. (January 2018). Retrieved from <https://github.com/bmorelli25/interactive-bitcoin-price-chart/>

Gamerveda. 2017. Cryptocurrency: August 1st Bitcoin Fork and the Importance of Understanding Sentiment. (July 2017). <https://steemit.com/bitcoin/@gamerveda/cryptocurrency-august-1st-bitcoin-fork-and-the-importance-of-understanding-sentiment>

Lucinda Shen. 2017. Mark Cuban's Twitter Trolling Sent Bitcoin Prices Tumbling. (June 2017). <http://fortune.com/2017/06/07/mark-cuban-twitter-bitcoin/>