

# Analisis de Sentimientos, enfocado a criptomonedas y su valor en el mercado

Bruno Troisi

22/11/2021

El análisis y detección de sentimientos (positivos, negativos o neutros) es uno de los campos más importantes del Procesamiento del Lenguaje Natural. Diversas técnicas y estado del arte (SOTA) son aportadas día a día por la comunidad científica, lo cuál luego lleva a la implementación de los mismos por parte de las organizaciones. En este Trabajo Práctico, además de aplicar análisis de sentimientos, utiliza el acceso a información de diversas APIs, citadas a continuación:

- Twitter, para scarpeo de noticias
- Coinmarketcap, para scarpeo del valor en tiempo real de las criptomonedas
- Reddit, para scarpeo de submissions en la subreddit de cada criptomoneda
- Google y Bing, por medio de BeautifulSoup, para el scarpeo de noticias

## Objetivo

Mediante el análisis del sentimiento de diversas noticias referidas de una criptomoneda, intentamos lograr entender cuál es la intención exacta de las mismas. Para así intentar predecir a partir de estas tendencias la variación del valor de la moneda al corto plazo.

## Criptos elegidas

Las 20 monedas elegidas para analizar son:

- Algorand, ALGO
- Bitcoin, BTC

- Cardano, ADA
- Chainlink, LINK
- Decentraland, MANA
- Dogecoin, DOGE
- Ecash, XEC
- Ethereum, ETH
- Litecoin, LTC
- Monero, XMR
- Pancakeswap, CAKE
- Polkadot, DOT
- Ravencoin, RVN
- Solana, SOL
- Terra, LUNA
- Tether, USDT
- Tezos, XTZ
- Tron, TRX
- Uniswap, UNI
- Xrp, XRP

### **Busqueda de noticias**

creamos la funcion buscarCrypto que se aprovecha de todas las APIs que importamos para el scarpeo, generando un set() con todas noticias referentes a la criptomoneda que queremos. Cada noticia/headline es pasado por diversos filtros antes de ser agregado al set, para no ensuciar el analisis posterior; Las submissions de Reddit son recopiladas únicamente de la subreddit oficial de la criptomoneda y los tweets recopilados de twitter son tomados desde hashtags definidos.

### **Barrera Idiomática**

Elegimos recopilar todas las noticias en inglés para simplificar la implementación de NLTK, la librería utilizada para el análisis de sentimiento. Pese a haber filtrado los resultados en únicamente que se encuentren en inglés, en determinados casos era posible que se filtre alguna noticia en español, por lo que hemos agregado una biblioteca que utiliza el servicio de google translate para esos escasos casos.

### **Procesamiento de textos**

Para el procesamiento del texto hemos utilizado el tokenizer de NLTK, un kit de herramientas de lenguaje natural, que contiene un conjunto de bibliotecas y programas para el procesamiento del lenguaje natural simbólico y estadísticos para Python.

### **Análisis de sentimientos**

El análisis de sentimientos fue delegado al SentimentIntensityAnalyzer de NLTK, que devuelve el set de noticias con 3 campos adicionales por elemento, que corresponden a un índice de positividad, neutralidad y negatividad, teniendo valores contenidos en el subconjunto  $[0,1]$  de los números reales y dando la suma entre ellos igual a 1. Para darle un mayor sentido cotidianamente hablando a esta data, procesamos este set en una tabla donde se agrega el campo compound, que posee valores contenidos en el subconjunto  $[-1,1]$  de los números reales, siendo los números que tienden a -1 como enteramente negativos, y los que tienden a 1 enteramente positivos. Además contemplamos (arbitrariamente) que los valores de  $[-0.2,0.2]$  refieren a noticias neutrales.

Este análisis además, recopila de las noticias negativas las 20 palabras más utilizadas, con su cantidad de ocurrencias; Repitiendo el mismo proceso con las noticias positivas.

### **Almacenamiento de información**

Para almacenar todos los datos recopilados, hemos generado una tabla en una base de datos simple de MySQL para donde guardamos registro de los datos básicos de todas las cryptos, como su nombre, su id de CoinMarketCap, su símbolo y el último valor que recopilamos de la misma. También, por cada crypto guardamos una tabla para almacenar el valor, la cantidad de noticias recopiladas, el porcentaje de positividad, neutralidad, negatividad, el compound y la variación en el valor con respecto a la anterior ejecución, por cada vez que se inicie el algoritmo.

El código fuente del proyecto puede ser obtenido en <https://github.com/BrunoTroisi/NLP/tree/master/crypto-sentiment-analysis>