

가상화폐 온도 측정기

https://github.com/Think-JIn99/crypto_reaction_check

가상화폐 온도 측정기 소개

가상 화폐 온도 측정기는 reddit에 존재하는 가상화폐 관련 subreddit에서 게시글 및 댓글을 크롤링 한 뒤 해당 서브레딧의 분위기를 한눈에 보여주는 서비스 입니다.



사용하고 있는 Subreddit

- 현재 프로젝트에서 참조하고 있는 Subreddits
 - [비트코인]
 - [이더리움]
 - [도지코인]
 - [가상화폐]

모델 개요

- 모델은 각 서브레딧 별 2019.01.01 ~ 2021.11.1 까지의 게시글을 통해 학습했습니다.
- 총 17만개의 게시글 데이터를 활용해 모델을 학습 시켰습니다.
- 전체 데이터 중 약 10%의 데이터를 활용해 성능 측정을 했고 결과는 다음과 같습니다.

<https://apption.co/embeds/ac107de0>

accuracy	f1 score
90	86

학습 및 데이터 수집 방법

- 데이터 수집은 reddit이 제공하는 praw를 비동기 방식으로 수정한 pmaw를 활용해 수집하였습니다.
- pmaw를 활용한 scrapper가 데이터를 수집하면, filter.py를 통해 광고글 및 비트코인과 연관이 없는 단어들을 삭제해 줍니다.
 - 삭제하는 단어 및 필터 관련 설정은 `create_regex.py`를 통해 확인 할 수 있습니다.
 - 이 과정에서 이모지, 특수기호 등의 모델이 학습 할 수 없는 문자들도 제거 해줍니다.
- 필터링을 거친 단어들에 대해 vader가 1차적으로 감정 정보를 측정 합니다.
 - vader는 MIT에서 제작한 감정사전 오픈소스 입니다.
 - vader 자체는 비트코인과 관련된 용어나 은어들을 모르기 때문에 **HODL**, **ATH**, **Bullish**, **Bearish**, **Scam**등의 용어들의 가중치를 따로 vader 사전에 추가해 줍니다.
 - 이와 관련된 자세한 사항은 `vader_feature.py`에서 확인 할 수 있습니다.
- 조정된 vader를 통해 매겨진 감정 정보를 토대로 라벨링을 진행하고 bert 모델을 학습 시킵니다.

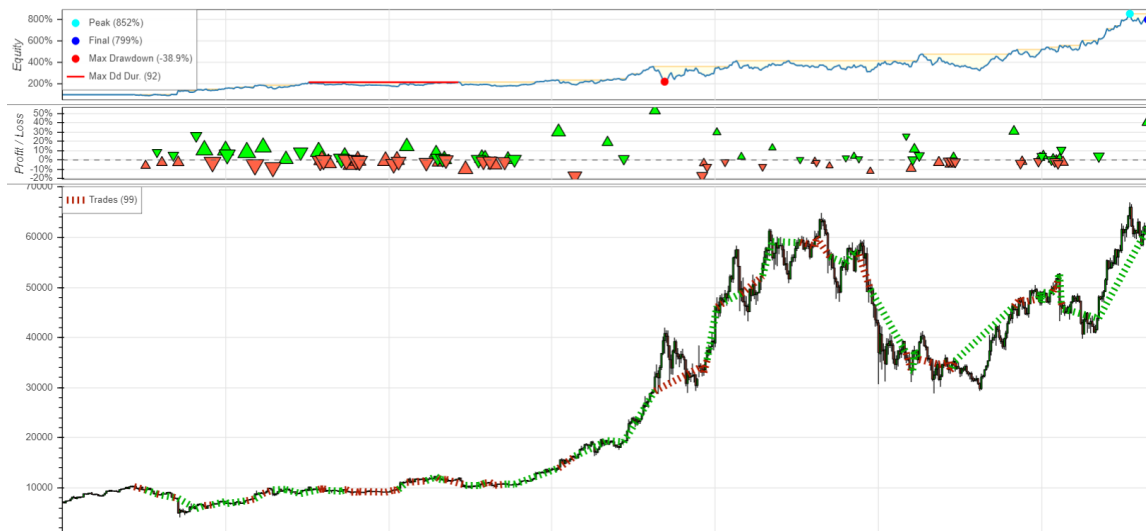


온도

- 사이트 메인에서 보이고 있는 온도는 전체 학습 데이터에 대해 모델이 매긴 결과 값의 평균에서 현재 데이터가 얼마나 벗어나 있는지를 보여줍니다.
 - 전체 평균에서 1% 증감 할때 마다 1도가 오르거나 내려 갑니다.

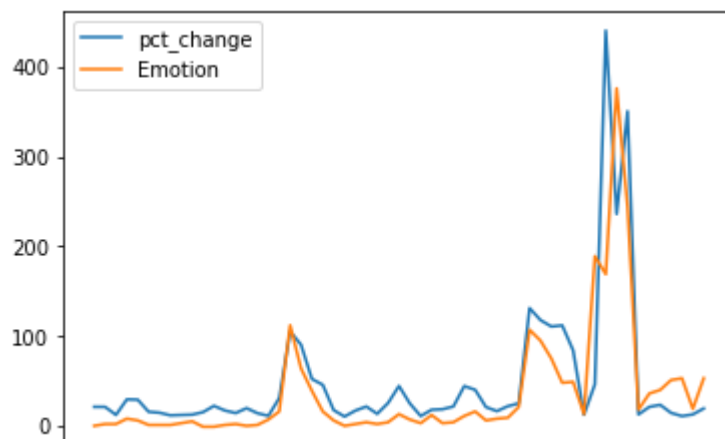
나아가서

- 그렇다면 이러한 감정정보를 어떻게 활용할 수 있을까?



감정 지표를 활용한 골든 크로스 전략 수익률이 썩 좋지 않은 것으로봐 뇌동 매매는 영 별로인 듯 하다.

- 위와 같이 새로운 지표를 사용해 새로운 투자전략을 구축하는데 사용할 수 있다.



도지 가격과 감정정보를 보여주는 그래프

