

FINAL PROJECT

온체인데이터를
활용한 비트코인 변동
‘이상징후 경고 시스템’

CODILOCKS

류태혁 박민정 장동헌 피근우

팀원 소개

CODILOCKS



태혁

프로젝트 팀장

- 모델제작 ·
- Github코드정리 ·
- 프로젝트 발표 ·

자료 제작

- Notion / ppt
- 제작 및 관리
- 모델 파라미터 코드수정

민정



동헌

주제 아이디어 제시

- 분석기획 구체화 ·
- 시스템구현 코딩 ·
- 크롤링 코딩 ·

데이터품질 관리

- 일일보고 및 서류관리
- 중간보고 발표
- 데이터 크롤링,
- 전처리작업

근우



CONTENTS

프로젝트 소개_ 프로젝트배경, 목적 및 순서소개

데이터 분석_ 이상치 탐지 모델 적용 및 분석

시스템 구현_ 이상징후 알림 서비스, 웹 대시보드 구현

결론 및 평가_ 프로젝트 활용방안과 한계점

1

프로젝트 소개

프로젝트배경, 목적 및 순서소개

온체인데이터란 ? “블록체인 위의 트랜잭션들의 정보”

암호화폐의 급격한 성장

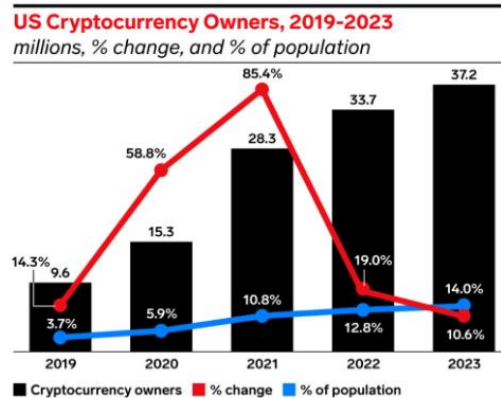
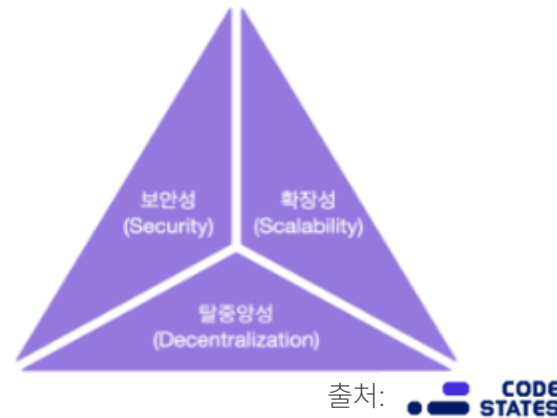


사진 = 미국 암호화폐 보유자 추이 / 인사이더 인텔리전스 보고서

블록체인 특성의 장점



온체인데이터를 다루는 기업들

glassnode insights

Dune Analytics

CryptoQuant

암호화폐 산업 주류화

“암호화폐 사용량 및 보유량 증가세”

모든 정보는 기록되고
변조되지 않으며 모두가 열람가능

“가상자산 시장의 흐름 파악 가능”

온체인데이터를 가공하여 지표 제공

“유의미한 인사이트 도출 가능성”

암호화폐의 관심도가 높아짐에 따라 가격변동에 대한 정보 수요 증가
온체인데이터를 통한 암호화폐의 변동 이상징후 발견 시스템 아이디어 도출

크립토크퀀트가 제공하는 데이터의 신뢰성 검토

크립토크퀀트(Cryptoquant)는 전문적인 온체인 데이터 및 분석을 제공하는 기업
온체인상의 데이터를 암호화폐 급등/급락 시그널 등의 금전적인 가치를 갖는 데이터로 가공해 정보를 제공해 줌



크립토크퀀트가 제공하는 데이터들은 *Cointelegraph, Coindesk, Forbes, Bloomberg* 같은 유명 뉴스 미디어에도 자주 인용되며 신뢰성을 보장받고 있다.

또한 N번방 사건에 연루된 지갑들을 조사하는 데 크립토크퀀트가 크게 기여하였으며, 이 과정에서 한국 검찰청 및 경찰과 협조한 이력이 있고 한국정부는 2019년부터 지갑들을 조사하는데 수신탁을 지원하고 있다.

[크립토크퀀트 관련 뉴스 기사들]

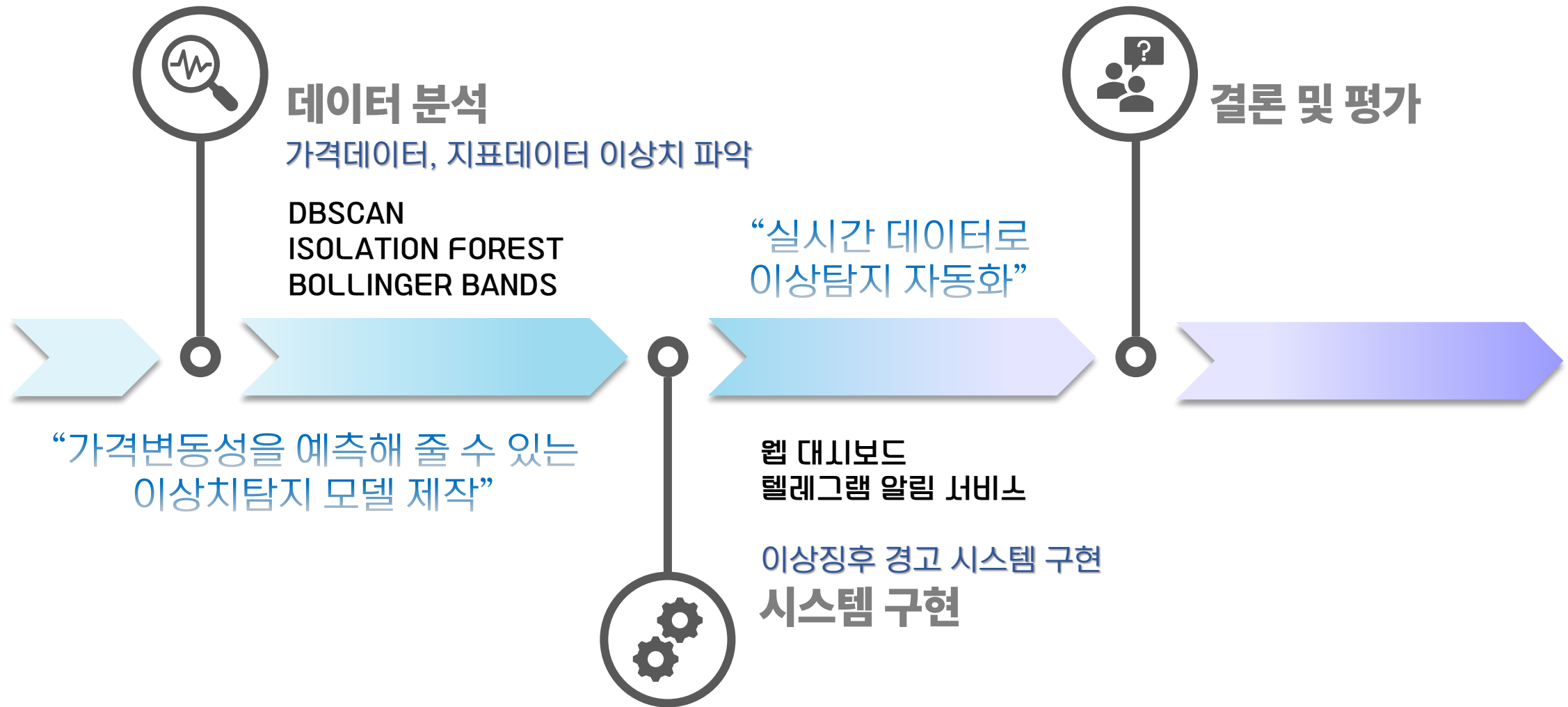
머니투데이

뉴스 증권 정치 법률 유니콘팩토리 헬스 바이오 환경 연예 이슈 NFT리포트

"매일 1조 들락, 이상해" FTX사태 예견...K-스타트업 찾아온 시카고 거래소

크립토크퀀트, 세계최대 파생상품 거래소 CME에 암호화폐 데이터 공급

"N번방 범죄수익 찾고 FTX 사태도 예견"...가상자산 분석틀에 뭉치듯



2

데이터 분석

이상치 탐지 모델 적용 및 분석



데이터 분석

데이터 전처리

크립토퀀트의 차트데이터
정규화 전처리 작업(MinMaxscaler)

INPUT DATA

비트코인 가격데이터

크립토퀀트의 차트데이터

- SOPR 비율
- 코인베이스 프리미엄
- 채굴자 포지션 지표
- 거래소 고래 비율

시장을 움직이는
주체와 관련된
데이터로 선정

전제조건

분석과정 1단계 볼린저밴드 적용

표준편차(std = 2)로 설정 후 값 확인
실제로 가격변동이 일어났던 일자 확인 -> 범위설정 기준마련

“가격데이터의 변동 일자 확인”



3단계 알고리즘 평가

변동성구간 존재유무, 이상치 존재유무 판단
파라미터 별 모델의 F1-score 등 확인

“알고리즘 타당성 검증”



Bollinger bands

2단계 DBSCAN, ISOLATION FOREST

밀도기반클러스터링, 랜덤포레스트

각 지표데이터를 이상치 탐색 모델에 학습
최적 파라미터 조정 -> 이상징후를 잘 설명하는 모델 선정

“각 지표의 이상치 탐지 모델 제작”

3단계를 통해 “ 이상징후 탐지 알고리즘 제작 및 타당성 검증 ”

가격데이터 변동일 확인

파이썬을 이용한 이상치 분석

알고리즘 타당성 검증

Bollinger bands를 이용한 비트코인 가격데이터의 변동성 일자 확인

- 볼린저밴드(Bollinger bands) 란?

주가 또는 지수와 같은 시계열 데이터의 변동 범위를 설정해
데이터의 변동성을 분석하는 기술적 분석 도구이자 보조지표

'실제 변동이 나타난 가격데이터의 일자를 확인하기 위해 사용'

[볼린저밴드를 적용한 그림]



- 볼린저 밴드 계산 함수 정의

```
import yfinance as yf
import pandas as pd

# 비트코인 (BTC-USD) 데이터를 가져오기
start_date = "2020-09-01"
end_date = "2023-09-21"
bitcoin_data = yf.download("BTC-USD", start=start_date, end=end_date)

# 볼린저 밴드 계산 함수 정의
def calculate_bollinger_bands(data, window=7, num_std_dev=2):
    data['SMA'] = data['Close'].rolling(window=window).mean()
    data['Upper'] = data['SMA'] + (data['Close'].rolling(window=window).std() *
    num_std_dev)
    data['Lower'] = data['SMA'] - (data['Close'].rolling(window=window).std() *
    num_std_dev)
    data['BandWidth'] = (data['Upper'] - data['Lower']) / data['SMA']
    return data
```

"밴드 너비가 0.3이상인 값만 저장"

```
# 볼린저 밴드 계산
bitcoin_data = calculate_bollinger_bands(bitcoin_data)

# 전날 볼린저 밴드 너비 저장
bitcoin_data['PrevBandWidth'] = bitcoin_data['BandWidth'].shift(1)

# 조건에 맞는 날짜 저장
crossed_threshold_dates = bitcoin_data[
    (bitcoin_data['BandWidth'] >= 0.3) & (bitcoin_data['PrevBandWidth'] < 0.3)
].index
```

가격데이터 변동일 확인

파이썬을 이용한 이상치 분석

알고리즘 타당성 검증

변동일자별 지표의 이상치 확인

볼린저밴드로 확인한 가격데이터 변동 일자



“총 14개의 가격변동 일자 확인”

변동성일자를 기준으로 2달전 + 일주일
범위에서 지표의 이상치를 탐색

범위선정이유 : 지표의 선행성 + 가격반동

최적
모델탐색

DBSCAN(density-based spatial clustering of applications with noise)

분석데이터 : 채굴자 포지션 지표, 거래소 거래 비율

데이터의 분포패턴을 머신러닝하여
비슷한 패턴끼리 군집화하는 ‘밀도기반클러스터링’ 기법

군집의 반지름인 eps 파라미터 조정

이상치 비율인 contamination 파라미터 조정

ISOLATION FOREST

분석데이터 : SOPR 비율, 코인베이스 프리미엄

데이터를 무작위선택 후 반복적으로 분할하여 결정 트리를 생성
이러한 결정 트리들을 결합하여 이상치를 선택하는 ‘랜덤 포레스트’ 기법

가격데이터 변동일 확인

파이썬을 이용한 이상치 분석

알고리즘 타당성 검증

이상치 탐지 모델 선정

- 각각 다른 9개의 파라미터로 분석 후 결과 그래프 확인

최적의 모델을 선정한 기준

가능한 이상치의 수를 최소화하면서도
많은 가격 변동성 기간에 포함되는 모델 선정

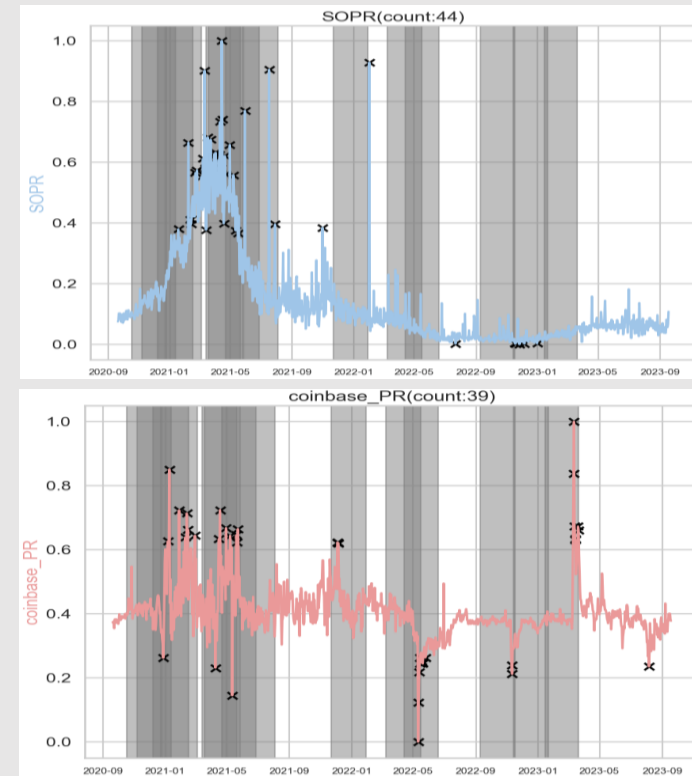
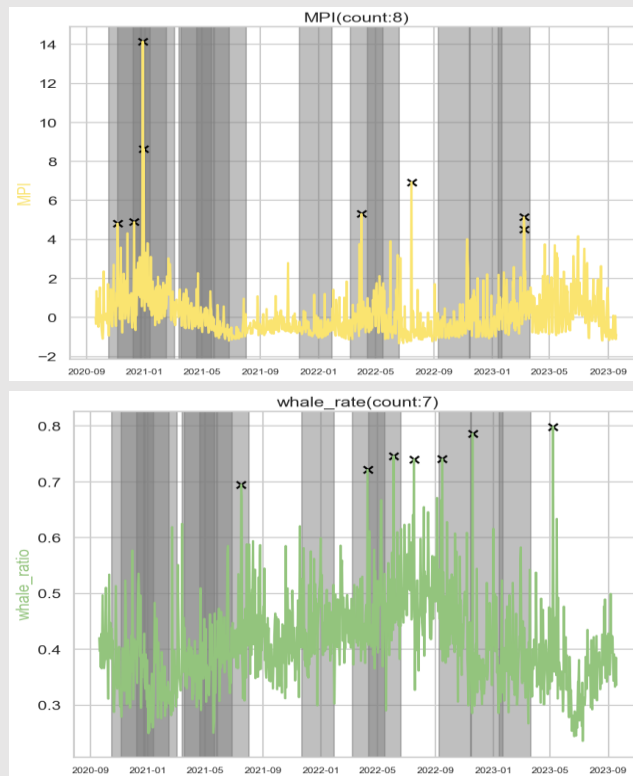
A 채굴자포지션(MPI) → 공급자의 상황 확인

B 거래소고래비율 → 큰금액을 움직이는 사람들 확인

C SOPR비율 → 장기보유자의 상황 확인

D 코인베이스 프리미엄 → 미국시장에 대한 정보 확인

각 지표별 이상치탐색 분석 결과



“ 각 지표별로 가격변동성을 설명해 줄 수 있는 최적의 이상치 탐지 모델 탐색 ”

가격데이터 변동일 확인

파이썬을 이용한 이상치 분석

알고리즘 타당성 검증

미래의 데이터를 모델에 적용할 경우 생기는 문제점

- 최근 3년치 데이터로 이상치 탐지 모델을 제작하는 알고리즘
ISOLATION FOREST 알고리즘 상 미래의 데이터에 따라 과거데이터를 이상치로 탐지한 결과가 달라질 수 있음
eps(반지름)으로 만든 DBSCAN에서는 문제가 되지 않으므로 DBSCAN모델을 다른 지표에서도 사용하기로 결정

이상탐지모델 평가 : 1) F1-score, 2) $\frac{N(\text{변동구간범위 내 이상치})+1}{N(\text{변동구간범위 외 이상치})+1}$

- 1) F1-score 란?

정밀도(Precision)와 재현율(Recall)의 조화평균으로 계산되는 성능 측정 지표

‘모델의 성능을 예측하고 결과를 평가하기 위해 사용’

$$\text{Precision} = \frac{\# \text{ of True Positives}}{\# \text{ of True Positives} + \# \text{ of False Positives}}$$

$$\text{Recall} = \frac{\# \text{ of True Positives}}{\# \text{ of True Positives} + \# \text{ of False Negatives}}$$

$$F1 \text{ score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

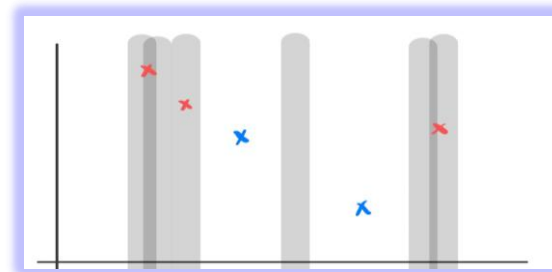
- 2) $\frac{N(\text{변동구간범위 내 이상치})+1}{N(\text{변동구간범위 외 이상치})+1}$

이상치 분류 실패 대비 성공의 score를 계산

1을 더해주는 이유는 변동성 구간 범위 밖의 이상치가 0이 나올 수 있기 때문

[Confusion Matrix]

		이상치 존재 유무	
		Positive	Negative
변동구간범위 내 존재 유무	Positive	TP	FP
	Negative	FN	TN



가격데이터 변동일 확인

파이썬을 이용한 이상치 분석

알고리즘 타당성 검증

알고리즘 평가 결과 확인 : F1-score

F1-score로 평가한 결과

- 지표별 유의미한 파라미터 확인

채굴자포지션(MPI)

Best F1 Score : 0.7586206896551724
 Model : DBSCAN(eps=0.01, min_samples=19)

거래소 거래 비율

Best F1 Score : 0.8484848484848484
 Model : DBSCAN(eps=0.01, min_samples=13)

SOPR 비율

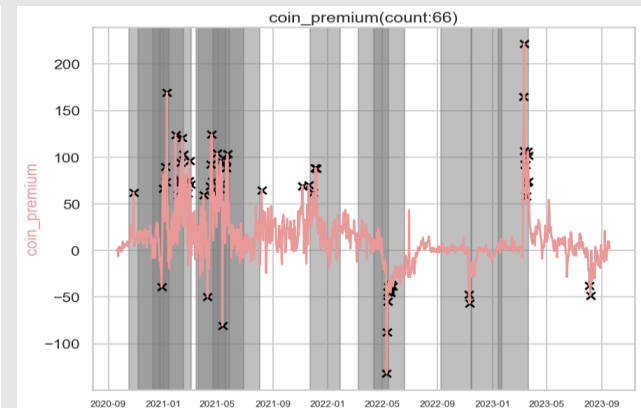
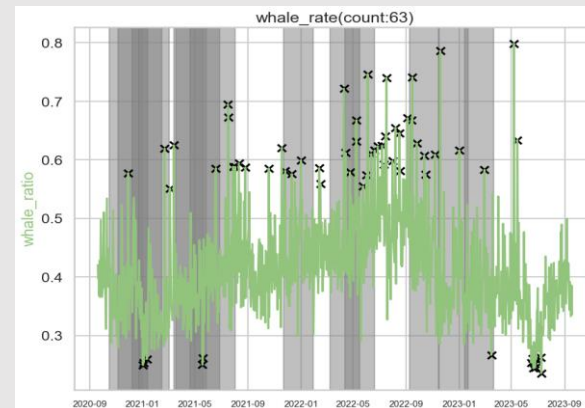
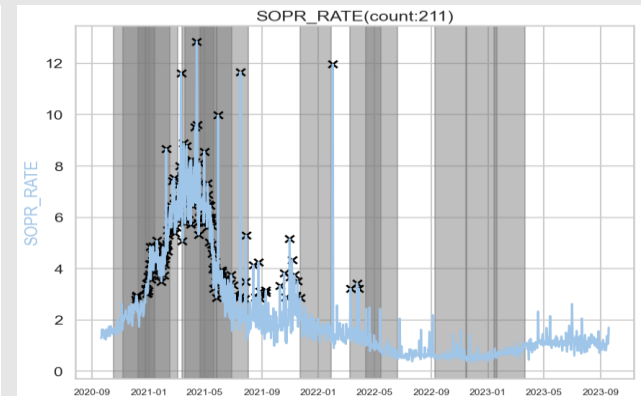
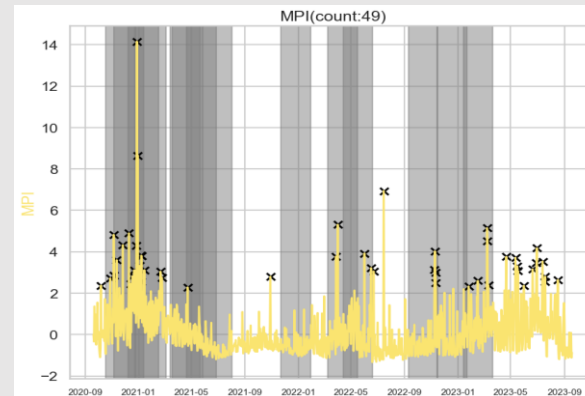
Best F1 Score : 0.7407407407407408
 Model : DBSCAN(eps=0.01, min_samples=20)

코인베이스 프리미엄

Best F1 Score : 0.8571428571428571
 Model : DBSCAN(eps=0.01, min_samples=13)

 각 지표별 가장 높은 score의 모델을 확인
 확인 결과 최적모델 선정기준에 부합하지 않음을 볼 수 있었다
 (너무 많은 데이터들을 이상치라고 분류함)

지표 별 결과 차트



가격데이터 변동일 확인

파이썬을 이용한 이상치 분석

알고리즘 타당성 검증

알고리즘 평가 결과 확인 : $\frac{N(\text{변동구간범위 내 이상치})+1}{N(\text{변동구간범위 외 이상치})+1}$

이상치 분류 실패 대비 성공의 score

- 지표별 유의미한 파라미터 확인

채굴자포지션(MPI)

Best Outlier Ratio : 3.5

Model : DBSCAN(eps=0.03, min_samples=12)

거래소 거래 비율

Best Outlier Ratio : 2.0

Model : DBSCAN(eps=0.04, min_samples=12)

SOPR 비율

~~Best Outlier Ratio : 9.25~~~~Model : DBSCAN(eps=0.01, min_samples=14)~~

코인베이스 프리미엄

Best Outlier Ratio : 18.0

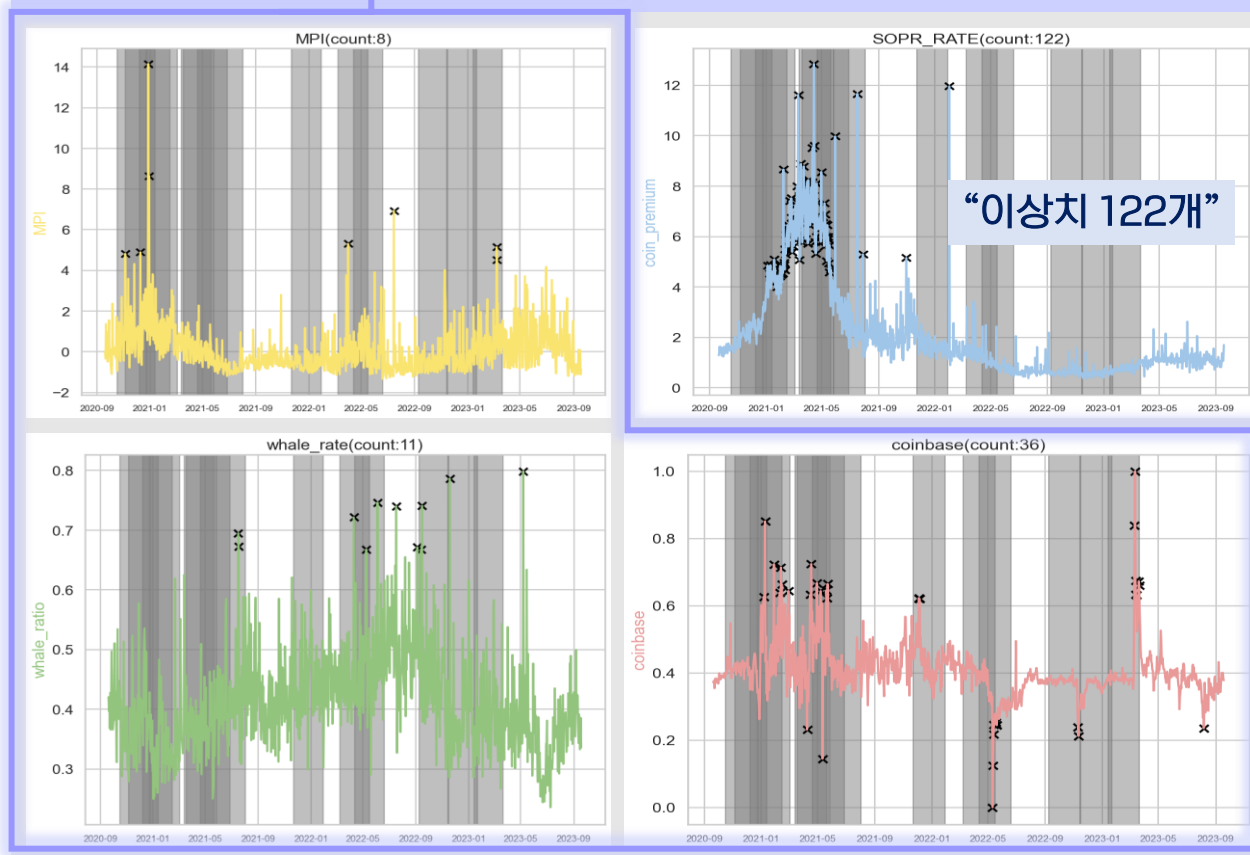
Model : DBSCAN(eps=0.02, min_samples=12)

‘F1-score보다 적합한 결과를 확인’

세개의 지표로 변동성구간에 이상치가 하나 이상 탐지되도록
파라미터 조정해줌 → ‘코인베이스:(eps=0.01, min_samples=10)’

SOPR비율 지표제외 세가지 지표를 사용

지표 별 결과 차트



3

시스템 구현

이상징후 알림 서비스, 웹 대시보드 구현

03 시스템 구현 _ 이상징후 알림 서비스

텔레그램을 활용한 이상징후 알림 서비스 구현

- Telegram, asyncio(비동기방식) 라이브러리 사용

```
import telegram, asyncio
...

#알림서비스
def Alert():
    tel_token = '토큰번호'
    chat_id = '개인아이디'
    bot = telegram.Bot(token=tel_token)
    ...

#코인베이스지표
try:
    Coinbase_df['날짜'] = pd.to_datetime(Coinbase_df['날짜'])
    Coinbase_df['가격'] = Coinbase_df['가격'].str.replace(',', '').astype(float)
except:
    pass

scaler = MinMaxScaler()
Coinbase_df[['코인베이스 프리미엄 지표']] = scaler.fit_transform(Coinbase_df[['코인베이스 프리미엄 지표']])

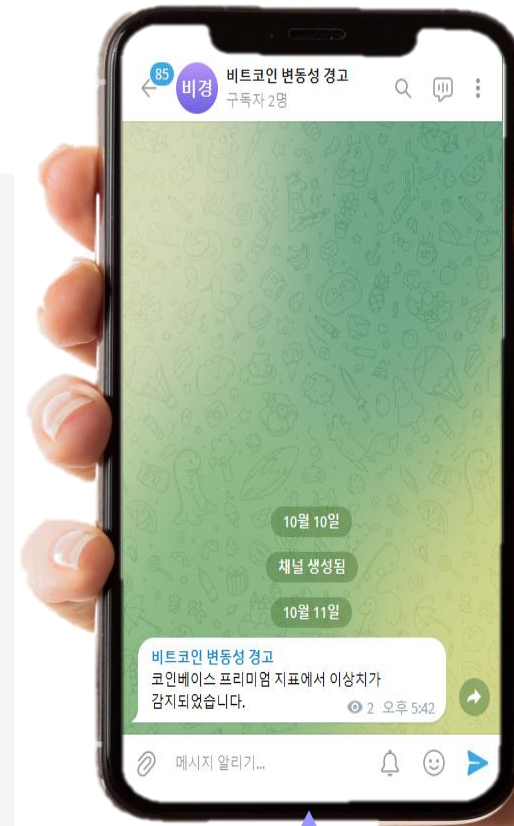
# DBSCAN 모델 생성 및 학습
model = DBSCAN(eps=0.01, min_samples=10)
model.fit(Coinbase_df[['코인베이스 프리미엄 지표']])

# 각 포인트가 이상치인지 아닌지 판별합니다.
Coinbase_df['anomaly'] = model.predict(Coinbase_df[['코인베이스 프리미엄 지표']])

Coinbase_last_label = Coinbase_df.iloc[-1]['anomaly']

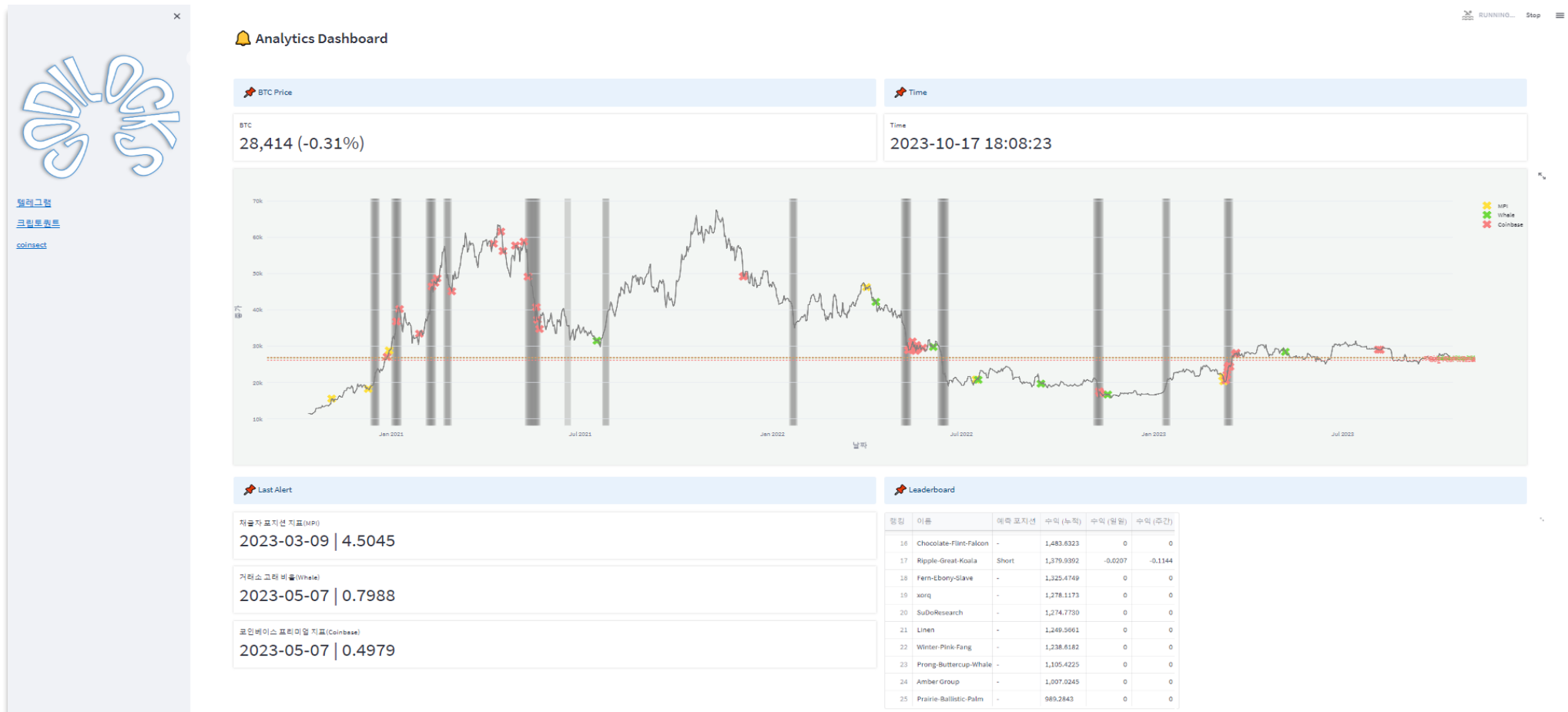
# -1인 경우 메시지 전송
if Coinbase_last_label == -1:
    # 텔레그램 봇 초기화
    bot = telegram.Bot(token=tel_token)

    # 메시지 전송
    message = "코인베이스 프리미엄 지표에서 이상치가 감지되었습니다."
    asyncio.run(bot.sendMessage(chat_id, message))
```



Streamlit을 활용한 웹 대시보드 구현

- 구성 : 가격, 시간, 지표별 이상징후(마지막 발생일,수치), 리더보드(비트맥스 제공)



4

결론 및 평가

프로젝트 활용방안, 한계점, 개선사항

분석에 대한 결론


우리의 알고리즘이 효과적인가?


- 기존에 투자전략이 있는 사람이 본 프로젝트에 대한 정보를 추가적으로 이용하면 더 효과적 일 수 있다
예시 : (변동성이 없는 상황에서만 투자하는 사람 , 변동성이 생겼을 때 투자를 하고 싶은 사람)
- 기간대비 수익률 효과 → 언제 움직이는지 알면 더 짧은 기간에 수익을 얻을 수 있다

프로젝트 활용방안 아이디어


- 옵션투자(스트랭글, 스트래들)에 활용
기초자산인 선물시장의 변동은 예상 가능하지만 어느 방향으로 움직일지 예상할 수 없을 때 사용하는 전략
- 횡보장과 추세장 구분하여 투자에 활용
서로 투자전략이 다름 → 횡보장: 역추세를 이용해서 수익, 추세장: 추세추종을 이용해서 수익


프로젝트 진행과정 중 나타난 한계점들


 코인시장의 4년 주기의 반감기, 데이터 불충분 이유로 평가불가 (학습, 테스트 셋 분할 어렵)


 API 비용 문제로 최근 3년까지의 데이터로만 분석

프로젝트 개선사항

 가격에 대한 방향성을 알 수 있는 기능 추가

 유의미한 지표 데이터를 추가

 코인별 변동성 비교 기능 추가

 지표데이터 별 데이터베이스 구축 및 관리

프로젝트 피드백

CODILOCKS



태혁

도메인에 대한 지식이 많았다면
한계점으로 나왔던 부분들을 빨리
파악하여 보완하거나 프로젝트 진
행 시간을 조금 단축할 수 있었을
거 같다. 도메인에 대한 지식이 아
쉬웠다.

민정

처음접하는 인사이트에 대해
흥미로웠지만 정답을 찾아가는
과정이 조금 힘들었다.
다른 프로젝트에서 이번과 같은
크롤링이 필요하다면 빠르게
아이디어를 생각할수 있을듯...



동현

여러가지 제약 상황이 있었지만
해결할 수 있어서 다행이었다.
시간이 조금 더 있었다면 다양한
시도나 검증을 해볼 수 있었을 것
같아서 조금 아쉬웠다.

근우

이전에 암호화폐 투자를 했다가
-95% 손실을 본 적이 있었는데
이번 프로젝트를 통해
정량적인 데이터를 통한
투자 시스템을 만들어서
감회가 새롭습니다.



감사합니다!