

MLP 비정형 데이터 활용 분석 세미 프로젝트

연방준비제도 기준 금리 예측

연방준비제도(연준)
The Federal Reserve System(Fed)



연준 기준 금리
Fed's Funds Rate



연방준비제도(Fed, 연준)
The Federal Reserve System



연준 기준 금리
Fed's Funds Rate



1. 경제 성장

- GDP, GDI

2. 경제 상황

- CPI(소비자 물가 지수), 인플레이션
- UNRATE(실업률)

3. 금융 시장

- 주식 및 부동산 시장

4. 재정 정책

- 정부 지출 및 세금의 변화에 따른 영향

5. 국제 경제 환경

- USD(환율)
- ECB_IR(유럽 중앙은행 기준 금리),
KR_IR(한국 중앙은행 기준 금리)

01

프로젝트 배경 및 목적

02

구성원 역할

03

1 팀
- 수행 절차 및 방법
- 수행 결과

04

2 팀
- 수행 절차 및 방법
- 수행 결과

프로젝트 배경: 전 세계 경제에 영향을 끼치는 연준 기준 금리에 대하여 알아보고자 함.

프로젝트 목적:

1. 연준 기준 금리에 영향을 미치는 주요 요인 탐색.
2. 탐색한 주요 요인들을 통한 연준 기준 금리 예측.

1 팀 – 빅데이터 모델링 기법

김주성: 비정형 데이터(news, youtube) 크롤링, EDA, 모델링, 평가, 예측

백동열: 정형 데이터 크롤링 및 전 처리, 정형/비정형 데이터 통합, EDA, 모델링, 평가, 예측

서웅진: 비정형 데이터(BeigeBook) 크롤링, EDA, 모델링, 평가, 예측

2 팀 – 통계 분석 기법

김동현: 회귀분석을 사용해서 기준 금리에 영향을 끼치는 요인, 인과관계 찾기

분석 도구

1. 개발 환경 및 프로그래밍 언어

Jupyter notebook, Python

2. 크롤링 도구

pandas_datareader, yfinance, openpyxl(EXCEL xlsx 데이터 로딩)

3. 정형/비정형 데이터 분석 및 평가 도구

Keras, NLTK

4. 데이터 시각화

WordCloud, Matplotlib, Seaborn

■ 데이터 준비

- 데이터 수집
- 데이터 전 처리

■ 데이터 분석

- 텍스트 데이터 분석
- EDA
- 모델링 및 모델 평가

※ Target = DFEDTARU

정형 데이터

데이터	데이터 입수처	데이터 사용방법	사용여부
연준 기준 금리 상한	pandas_datareader 모듈	DataFrame	Y
연준 기준 금리 예측치	investing.com	엑셀xlsx	Y
실질 국내총생산 성장률	https://fred.stlouisfed.org/	엑셀csv	Y
실업률	https://fred.stlouisfed.org/	엑셀csv	Y
소비자 물가 지수	https://fred.stlouisfed.org/	엑셀csv	Y
명목이자율(미국 국채 10년물)	yfinanace 모듈	DataFrame	Y
한은 USD환율	https://ecos.bok.or.kr/	엑셀csv	Y
한은 기준 금리	https://ecos.bok.or.kr/	엑셀csv	Y
ECB 기준 금리	https://data.ecb.europa.eu	엑셀csv	Y

비정형 데이터

데이터	데이터 입수처	사용여부
Beige Book	https://fred.stlouisfed.org /	Y
Youtube metadata	Youtube API	N
Naver metadata	https://search.naver.com	N

※ 월별로 데이터가 복수 건 존재하여 비사용
※ 데이터 수집 기간: 2008.12.31 ~ 2024.07.31

■ 데이터 준비

○ 데이터 수집

○ 데이터 전 처리

■ 데이터 분석

○ 텍스트 데이터 분석

○ EDA

○ 모델링 및 모델 평가

1. 정형/비정형 시계열 데이터로 변환

- 데이터 준비
 - 데이터 수집
- 데이터 전 처리
- 데이터 분석
 - 텍스트 데이터 분석
 - EDA
 - 모델링 및 모델 평가

	DATE	DFEDTARU	DFEDTARL	IR_FORCAST	GDPGR	UNRATE	PCEPI	TNX	USD	KR_IR	ECB_IR
0	2008-12-01	0.25	0.00	0.50	-8.5	7.3	88.098	2.244	1259.5	3.0	2.50
1	2009-01-01	0.25	0.00	0.25	-4.5	7.8	88.108	2.844	1379.5	2.5	2.00
2	2009-02-01	0.25	0.00	0.25	-4.5	8.3	88.266	3.041	1534.0	2.0	2.00
3	2009-03-01	0.25	0.00	0.25	-4.5	8.7	88.169	2.685	1383.5	2.0	1.50
4	2009-04-01	0.25	0.00	0.25	-0.7	9.0	88.295	3.124	1282.0	2.0	1.25
...
184	2024-04-01	5.50	5.25	5.50	2.8	3.9	123.109	4.686	1382.0	3.5	4.50
185	2024-05-01	5.50	5.25	5.50	2.8	4.0	123.146	4.514	1384.5	3.5	4.50
186	2024-06-01	5.50	5.25	5.50	2.8	4.1	123.243	4.343	1376.7	3.5	4.25
187	2024-07-01	5.50	5.25	5.50	2.8	4.1	123.243	4.128	1376.5	3.5	4.25

2. 정형/비정형 데이터 통합

DATE	DFEDTARU	DFEDTARL	IR_FORCAST	GDPGR	UNRATE	PCEPI	TNX	USD	KR_IR	ECB_IR	BB_text
2008-12-01	0.25	0.00	0.50	-8.5	7.3	88.098	2.244	1259.5	3.0	2.50	Prepared at the Federal Reserve Bank of Minnea...
2009-01-01	0.25	0.00	0.25	-4.5	7.8	88.108	2.844	1379.5	2.5	2.00	Prepared at the Federal Reserve Bank of St. Lo...
2009-02-01	0.25	0.00	0.25	-4.5	8.3	88.266	3.041	1534.0	2.0	2.00	Prepared at the Federal Reserve Bank of St. Lo...
2009-03-01	0.25	0.00	0.25	-4.5	8.7	88.169	2.685	1383.5	2.0	1.50	Prepared at the Federal Reserve Bank of San Fr...
2009-04-01	0.25	0.00	0.25	-0.7	9.0	88.295	3.124	1282.0	2.0	1.25	Prepared at the Federal Reserve Bank of Dallas...
...

- 데이터 준비
 - 데이터 수집
- 데이터 전 처리
- 데이터 분석
 - 텍스트 데이터 분석
 - EDA
 - 모델링 및 모델 평가

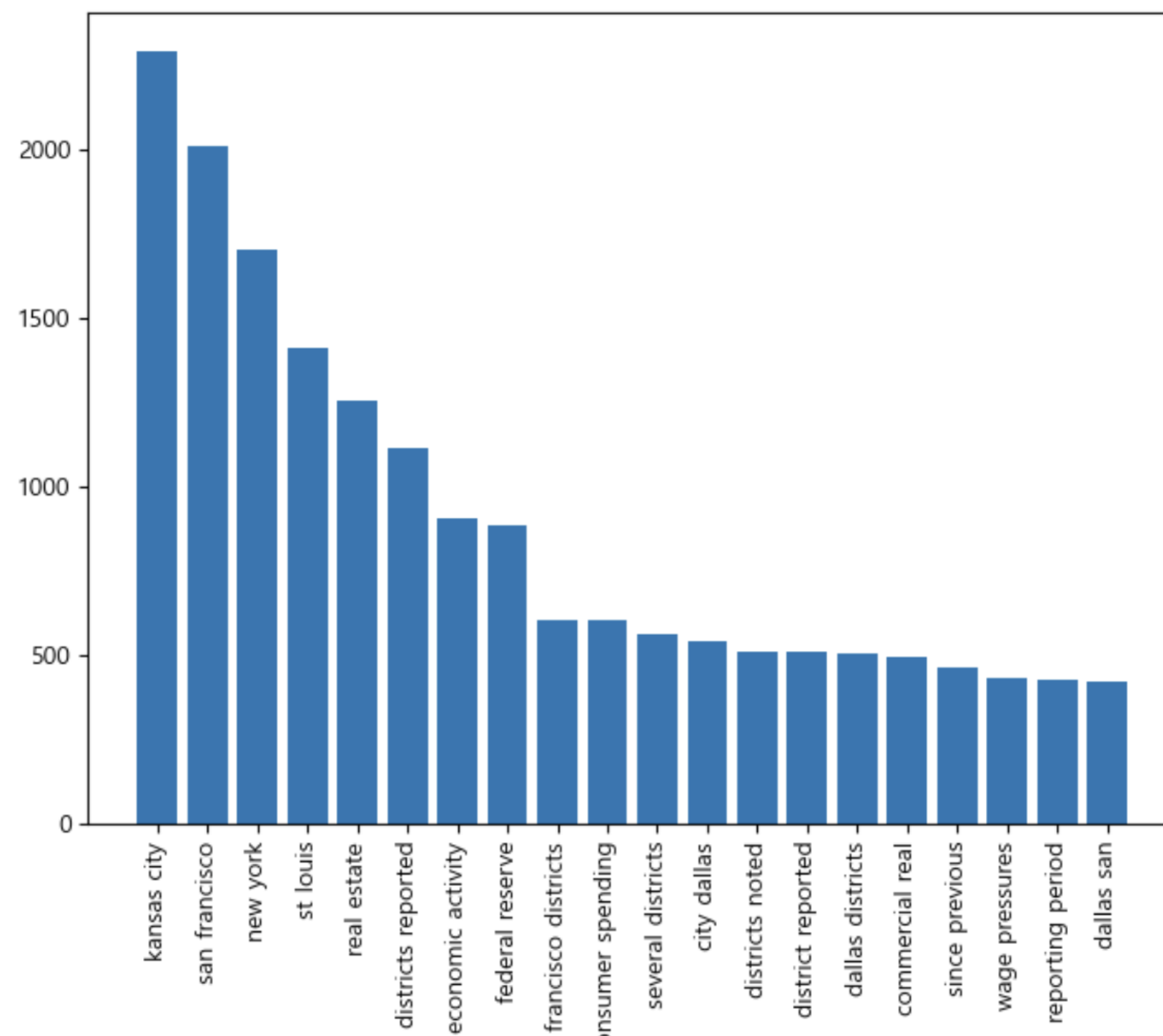
텍스트 데이터 전 처리

■ 데이터 준비

- 데이터 수집
- 데이터 전 처리

■ 데이터 분석

- 텍스트 데이터 분석
- EDA
- 모델링 및 모델 평가



연방준비 은행 지역명 목록 제거

```
# 연방준비은행 지역 목록 제거st louis와 같은 경우도 감안
federal_reserve_banks = [
    "boston",
    "new york",
    "philadelphia",
    "cleveland",
    "richmond",
    "atlanta",
    "chicago",
    "st. louis",
    "minneapolis",
    "kansas city",
    "dallas",
    "san francisco",
    "st",
    "louis"
]

# 연방준비은행 지역 필터링 함수
def remove_federal_reserve_banks(text, bank_list):
    # 소문자 변환 및 정규표현식을 사용하여 지역 이름 제거
    for bank in bank_list:
        bank_pattern = re.compile(re.escape(bank), re.IGNORECASE)
        text = bank_pattern.sub('', text)
    return text

# 연방준비은행 지역 제거
df_irp['BB_text'] = df_irp['BB_text'].apply(lambda x: remove_federal_reserve_banks(x, federal_reserve_banks))
```


텍스트 데이터 분석

■ 데이터 준비

- 데이터 수집
- 데이터 전 처리

■ 데이터 분석

- 텍스트 데이터 분석
- EDA
- 모델링 및 모델 평가



텍스트 데이터 감성 분석

```
sia = SentimentIntensityAnalyzer()
```

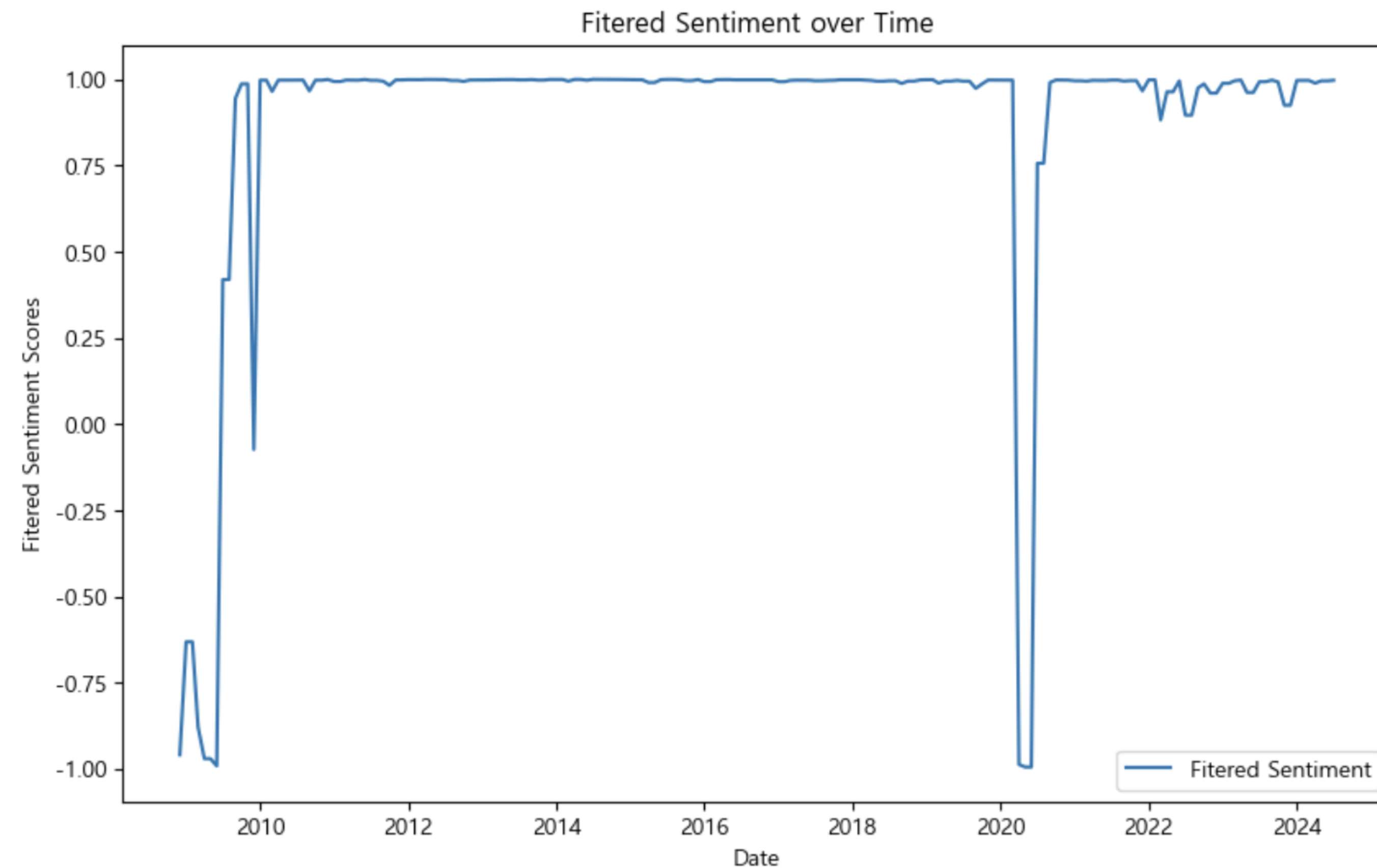
```
def get_vader_sentiment(text):
    sentiment = sia.polarity_scores(text)
    return sentiment['neg'], sentiment['neu'], sentiment['pos'], sentiment['compound']
```

```
df[['BB_neg', 'BB_neu', 'BB_pos', 'BB_sentiment']] = df['BB_text'].apply(lambda text:
pd.Series(get_vader_sentiment(text)))
```


[illegible]

- 데이터 준비
 - 데이터 수집
 - 데이터 전 처리
- 데이터 분석
 - 텍스트 데이터 분석
 - **EDA**
 - 모델링 및 모델 평가

텍스트 데이터 BB_Sentiment 추이



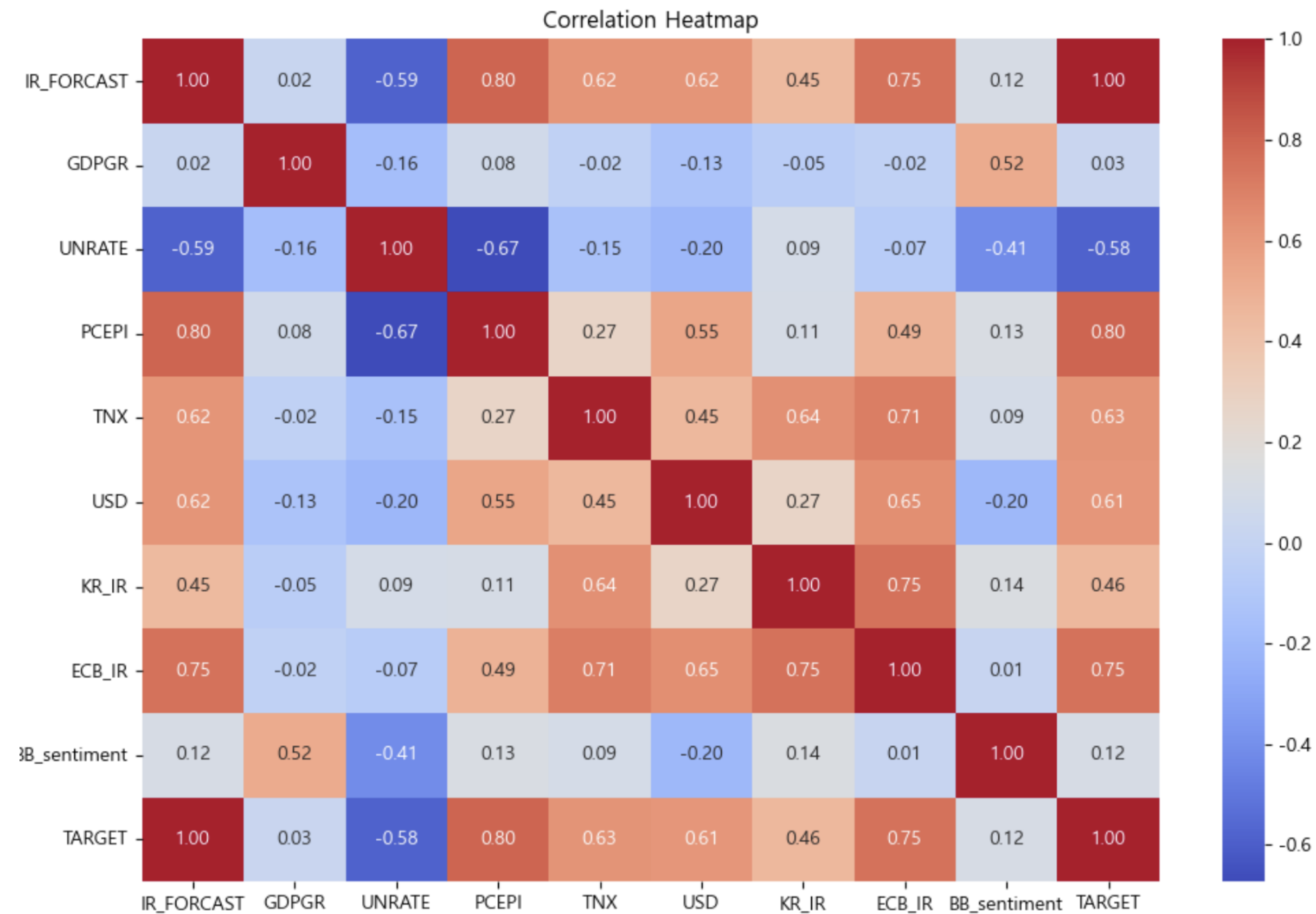
■ 데이터 준비

- 데이터 수집
- 데이터 전 처리

■ 데이터 분석

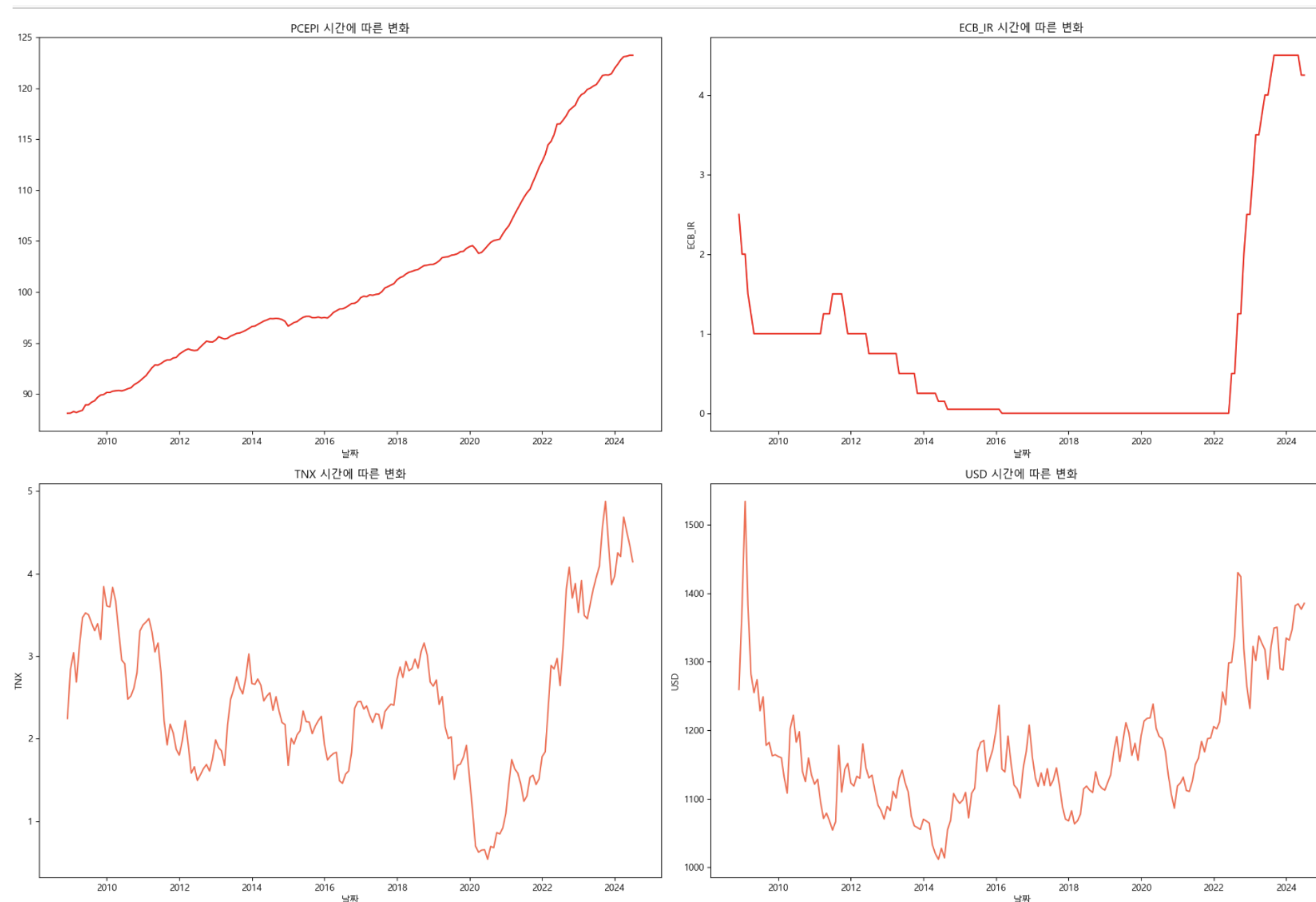
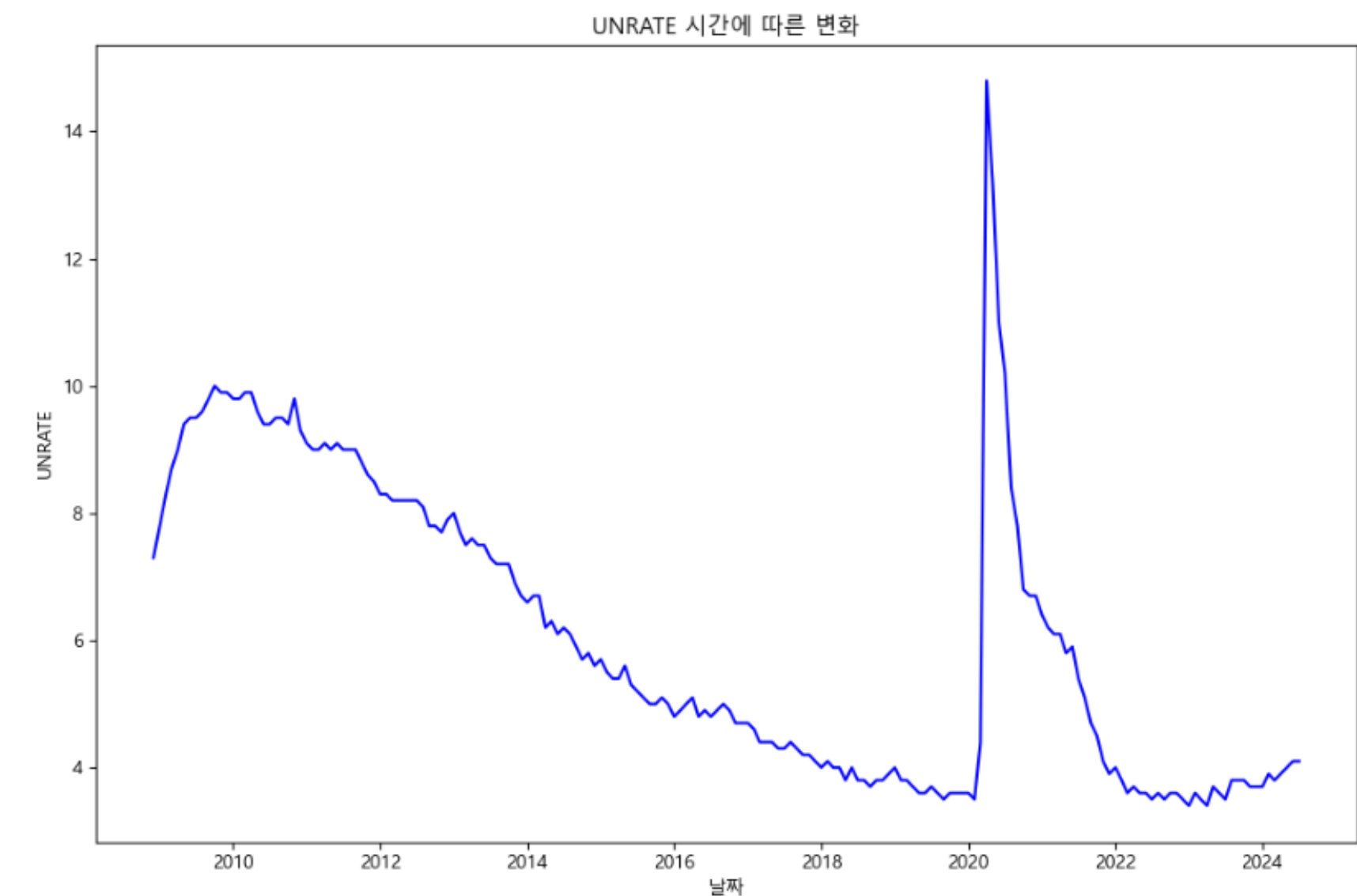
- 텍스트 데이터 분석
- EDA
- 모델링 및 모델 평가

히트 맵



상관계수 높은 변수들의 추이

- 데이터 준비
 - 데이터 수집
 - 데이터 전 처리
- 데이터 분석
 - 텍스트 데이터 분석
 - EDA
 - 모델링 및 모델 평가

PCEPI, ECR_IR,
TNX, USD
양의 상관관계UNRATE
음의 상관관계

- 데이터 준비

- 데이터 수집
- 데이터 전 처리

- 데이터 분석

- 텍스트 데이터 분석
- EDA
- 모델링 및 모델 평가

CNN

RNN

LSTM

MLP

CNN

변수 선택
단일 변수 사용

```
# 모델링
model = Sequential()
model.add(Conv1D(32, 3, activation='relu',
input_shape=(X_train.shape[1], X_train.shape[2])))
model.add(MaxPooling1D(2))
model.add(Dropout(0.2))
model.add(Conv1D(64, 3, activation='relu'))
model.add(MaxPooling1D(2))
model.add(Dropout(0.2))
model.add(Flatten())
model.add(Dense(128, activation='relu'))
model.add(Dropout(0.2))
model.add(Dense(1))

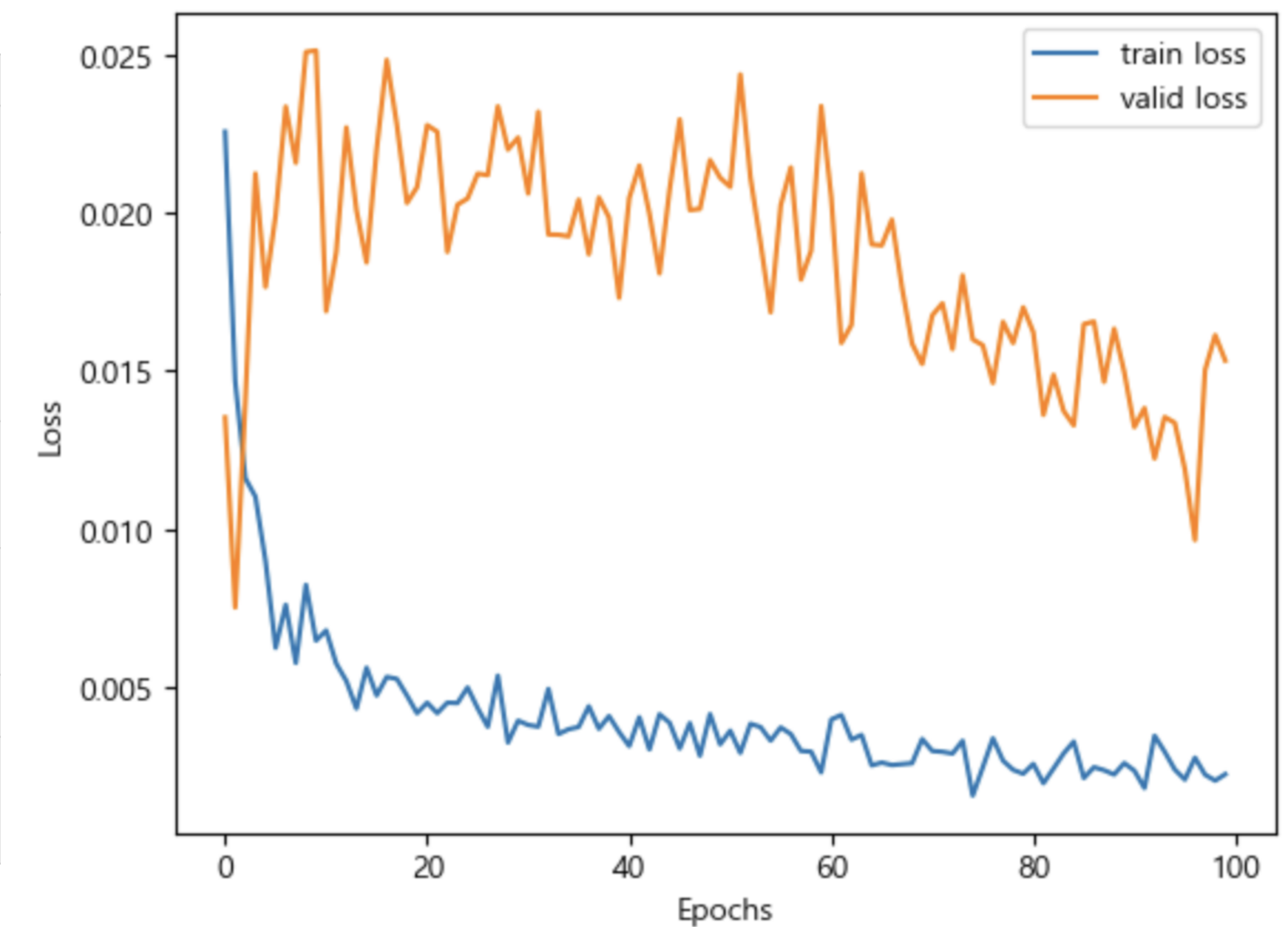
model.compile(optimizer=Adam(learning_rate=0.001),
loss='mean_squared_error', metrics='acc')

history = model.fit(X_train, y_train, epochs=100, batch_size=32,
validation_split=0.1)
```

요인 명	변수 명
소비자 물가 지수	PCEPI
실업률	UNRATE
Beige book 감정	BB_Sentiment
국내총생산 성장률	GDPGR
한은 USD환율	USD
명목이자율	TNX
한은 기준 금리	KR_IR

※
볼드체: 모델링에 사용된 변수
초록색 볼드체: 최종 채택된 변수

PCEPI 활용 모델 예측 손실 그래프



■ 데이터 준비

- 데이터 수집
- 데이터 전 처리

■ 데이터 분석

- 텍스트 데이터 분석
- EDA
- 모델링 및 모델 평가

■ 데이터 준비

- 데이터 수집
- 데이터 전 처리

■ 데이터 분석

- 텍스트 데이터 분석
- EDA
- 모델링 및 모델 평가

변수 선택
단일 변수 사용

```
# 모델링
model = Sequential()
model.add(SimpleRNN(64, return_sequences=True,
input_shape=(sequence_length, 1)))
model.add(Dropout(0.2))
model.add(SimpleRNN(64))
model.add(Dropout(0.2))
model.add(Dense(1))
```

```
model = Sequential()
model.compile(optimizer=Adam(learning_rate=0.001),
loss='mean_squared_error', metrics = 'acc')
```

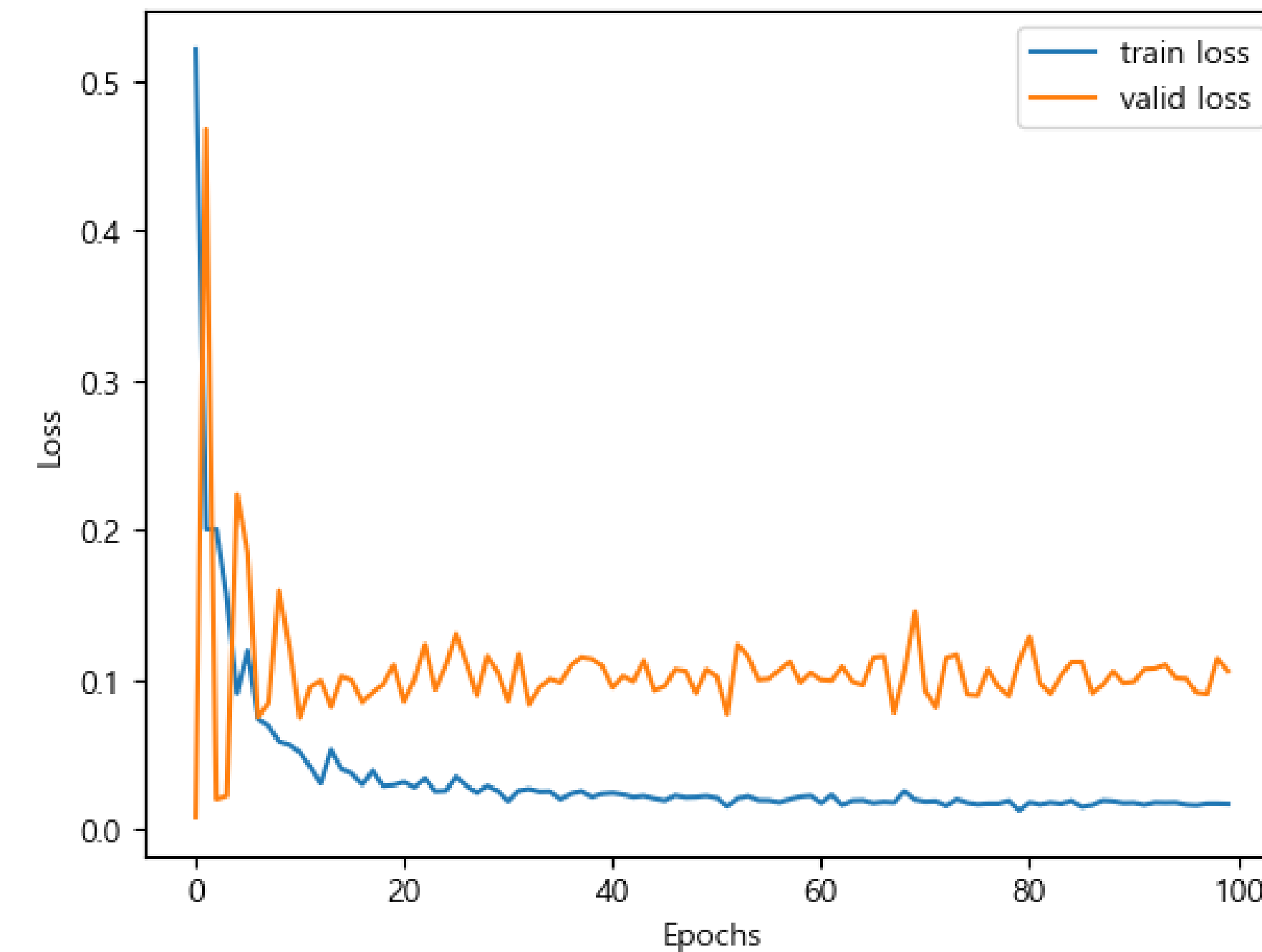
```
history = model.fit(X_train, y_train, epochs=100,
batch_size=32, validation_split=0.1)
```

RNN

요인 명	변수 명
소비자 물가 지수	PCEPI
실업률	UNRATE
Beige book 감정	BB_Sentiment
국내총생산 성장률	GDPGR
한은 USD환율	USD
명목이자율	TNX
한은 기준 금리	KR_IR

※
볼드체: 모델링에 사용된 변수
초록색 볼드체: 최종 채택된 변수

PCEPI 활용 모델 예측 손실 그래프



■ 데이터 준비

- 데이터 수집
- 데이터 전 처리

■ 데이터 분석

- 텍스트 데이터 분석
- EDA
- 모델링 및 모델 평가

변수 선택

높은 상관계수를 가진 변수들 사용

모델링

model = Sequential()

model.add(LSTM(units=64, return_sequences=True))

model.add(Dropout(0.5))

model.add(LSTM(units=32))

model.add(Dropout(0.5))

model.add(Dense(50, activation='relu'))

model.add(Dense(1))

model.compile(optimizer='adam',

loss='mean_squared_error', metrics=['acc'])

history = model.fit(X_train, y_train, epochs=200, batch_size=32,

validation_data=(X_val, y_val),

callbacks=[early_stopping, model_checkpoint])

LSTM

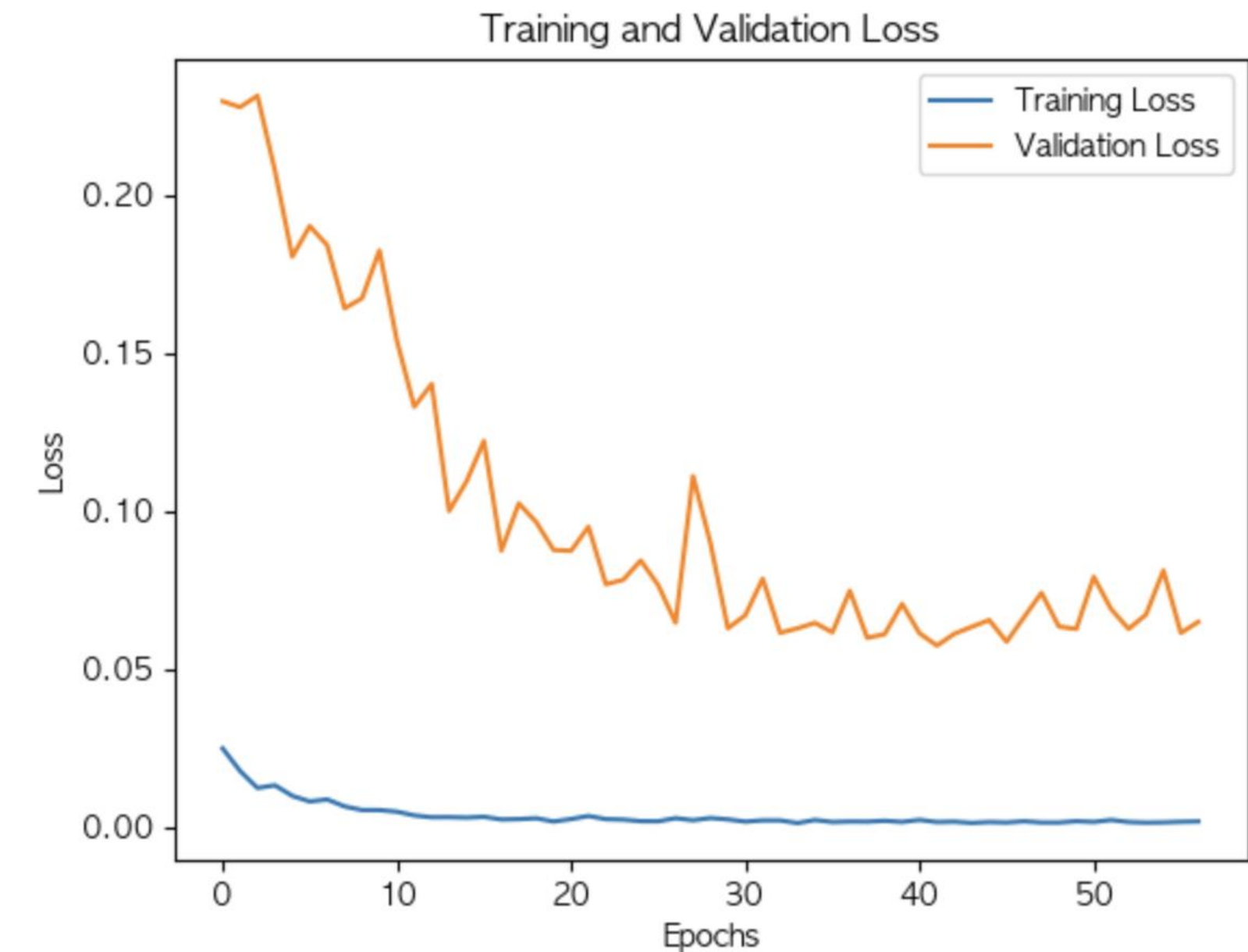
요인 명	변수 명
소비자 물가 지수	PCEPI
실업률	UNRATE
Beige book 감정	BB_Sentiment
국내총생산 성장률	GDPGR
한은 USD환율	USD
명목이자율	TNX
한은 기준 금리	KR_IR

※

볼드체: 모델링에 사용된 변수

초록색 볼드체: 최종 채택된 변수

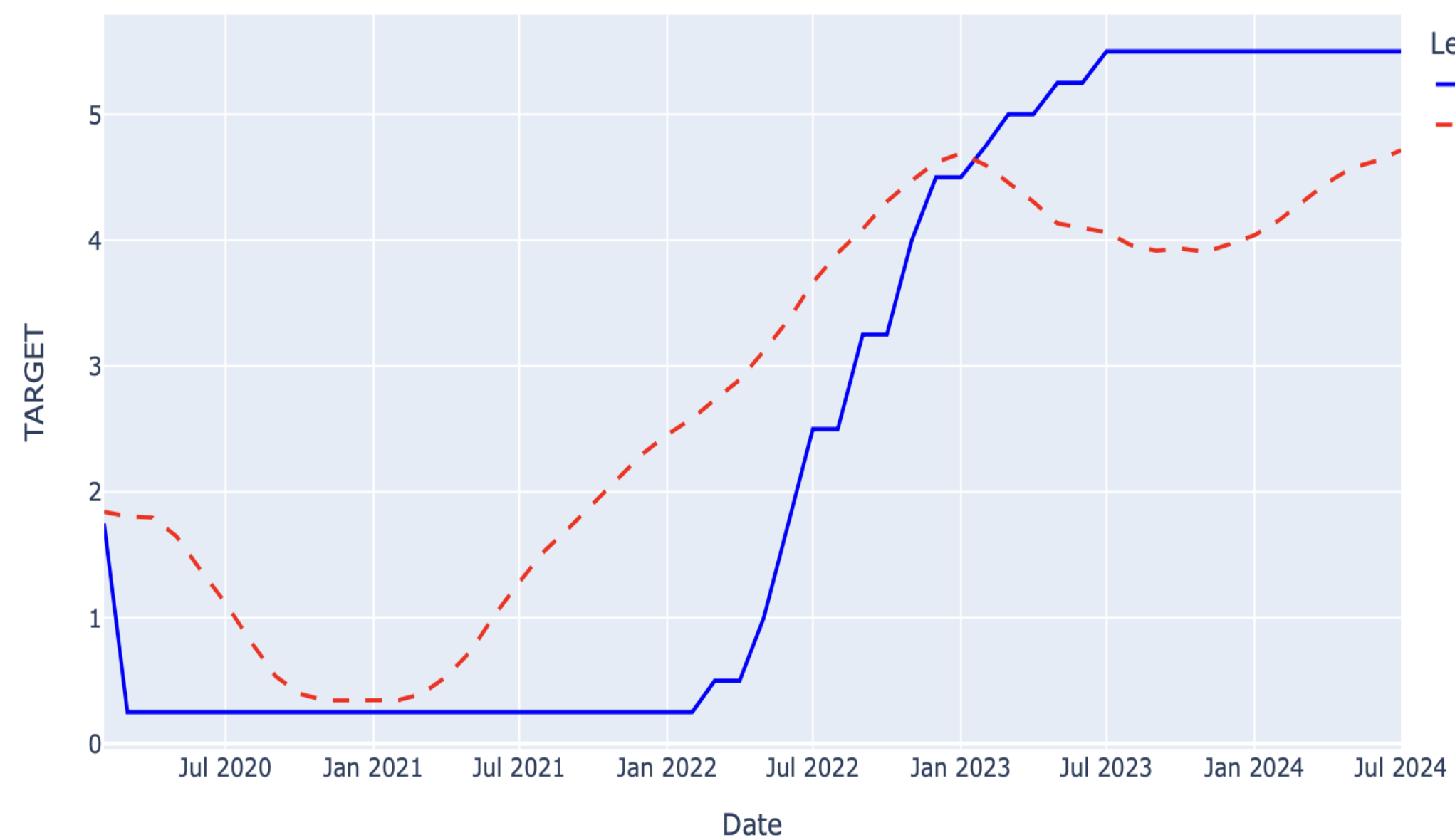
모델 예측 손실 그래프



LSTM

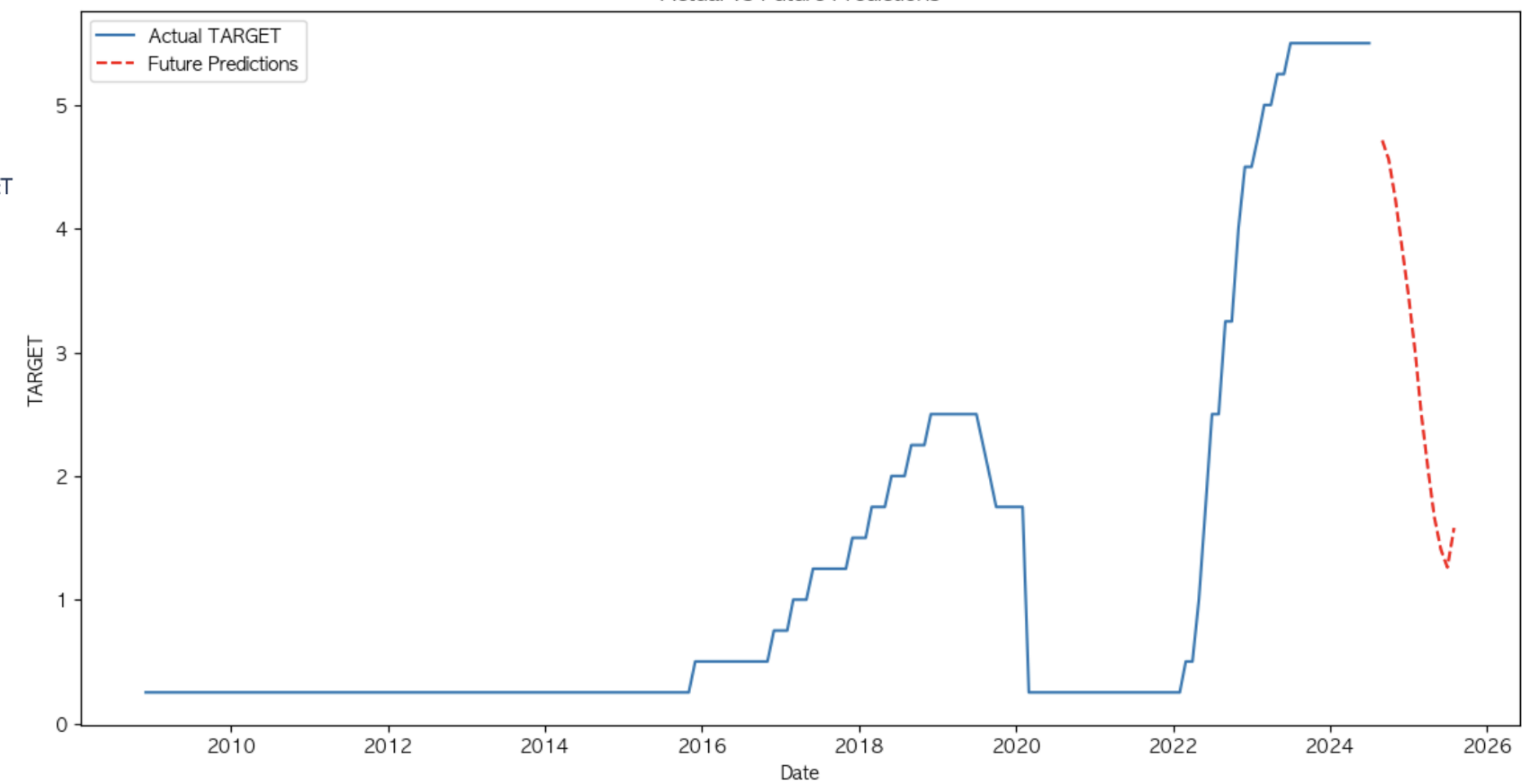
모델 예측 그래프

Actual vs Predicted TARGET



미래 예측 그래프(12개월)

Actual vs Future Predictions



MLP

```
# 변수 선택  
모든 변수 사용
```

```
# 종속 변수  
df['Target'] = (df['DFEDTARU'] >= 5.5).astype(int)
```

```
# 모델링  
model = Sequential()  
model.add(Dense(64, input_dim=X_train.shape[1],  
activation='relu'))  
model.add(Dense(32, activation='relu'))  
model.add(Dense(1, activation='sigmoid'))
```

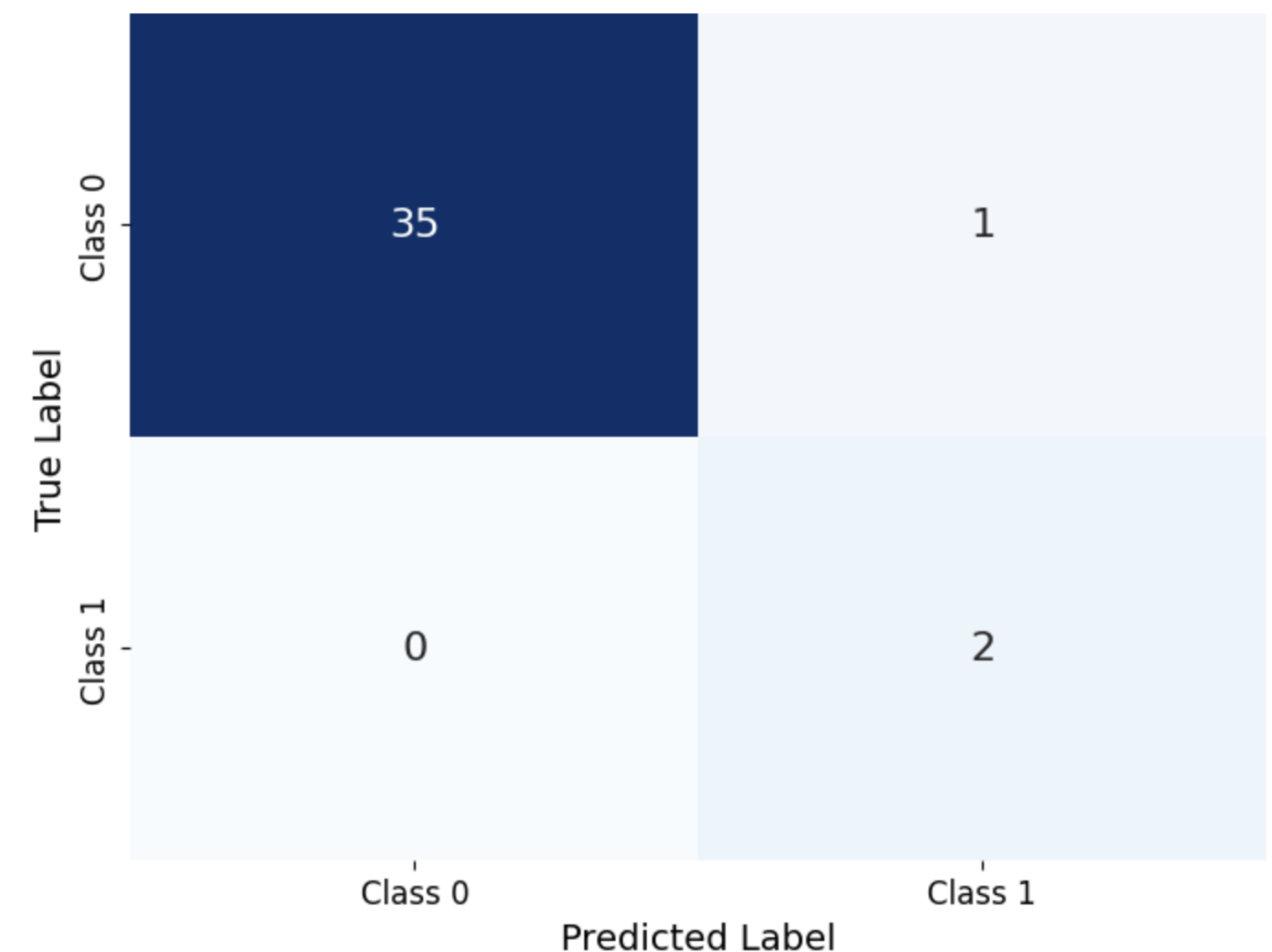
```
model.compile(optimizer='adam', loss='binary_crossentropy',  
metrics=['accuracy'])  
history = model.fit(X_train, y_train, epochs=20, batch_size=32,  
validation_split=0.2)
```

요인 명	변수 명
소비자 물가 지수	PCEPI
실업률	UNRATE
Beige book 감정	BB_Sentiment
국내총생산 성장률	GDPGR
한은 USD환율	USD
명목이자율	TNX
한은 기준 금리	KR_IR

※
볼드체: 모델링에 사용된 변수
초록색 볼드체: 최종 채택된 변수

모델 예측 정확도 및 Confusion Matrix

Test Accuracy: 97%



■ 데이터 준비

- 데이터 수집
- 데이터 전 처리

■ 데이터 분석

- 텍스트 데이터 분석
- EDA
- 모델링 및 모델 평가

모델 별 미래 예측치

모델명	실제값	예측값	차이
LSTM	5.25	0.09348	5.15652
SimpleRNN	5.25	0.1124600023031234	5.13754
CNN	5.25	0.08139999955892563	5.1686

변수 별 미래 예측치

요인명	실제값	예측값	차이
연준 기준 금리 예측치	5.25	0.68483	4.56517
실질 국내총생산 성장률	5.25	0.1169200018048286	5.13308
실업률	5.25	0.3256700038909912	4.92433
개인 소비 지출: 체인형 가격 지수	5.25	1.1114200353622437	4.13858
명목이자율	5.25	0.1074199974536895	5.14258
한은 USD환율	5.25	0.1628700047731399	5.08713
한은 기준 금리	5.25	-0.1071600019931793	5.35716
ECB 기준 금리	5.25	-0.2653700113296509	5.51537
Beige book 감정	5.25	0.07388000190258026	5.17612

1. CNN

- 단일 요인으로 평가 결과, 높은 성능이 나타남.

2. RNN

- 1) 단일 요인인 소비자 물가 지수로 평가 시 가장 성능이 우수.
- 2) 단일 요인들을 평가 시 연준에서 발표하는 주요 시장 요인이 다른 시장 요인보다 적합.

3. LSTM

- 1) LSTM을 이용하여 모델 예측 및 미래 예측 진행.
- 2) 모델 예측 정확성은 낮은 편은 아니었지만,
미래 예측 정확성은 높지 않음

4. MLP

- 1) 경제 지표를 활용하여 이진 기법(동결 및 상승/ 하락)으로 연준 기준 금리 예측.
- 2) 높은 정확도로 분류를 하였음.

1. CNN

- 단일 요인으로 평가 결과, 높은 성능이 나타남.

2. RNN

- 1) 단일 요인인 소비자 물가 지수로 평가 시 가장 성능이 우수.
- 2) 단일 요인들을 평가 시 **연준에서 발표하는 주요 시장 요인**이 다른 시장 요인보다 적합.

3. LSTM

- 1) LSTM을 이용하여 모델 예측 및 미래 예측 진행.
- 2) 모델 예측 정확성은 낮은 편은 아니었지만, 미래 예측 정확성은 높지 않음.

4. MLP

- 1) 경제 지표를 활용하여 이진 기법(동결 및 상승/ 하락)으로 연준 기준 금리 예측.
- 2) 높은 정확도로 분류를 하였음.

요인명	변수 명	오차	순위
소비자 물가 지수	PCEPI	4.56243	1
실업률	UNRATE	5.03995	2
Beige book 감정	BB_BOOK	5.16333	3
국내총생산 성장률	GDPGR	5.17827	4
한은 USD환율	USD	5.20678	5
명목이자율	TNX	5.22471	6
한은 기준 금리	KR_IR	5.34476	7

- 가설 설정

- 가설 설정

- 데이터 준비

- 데이터 수집

- 데이터 분석

- 회귀 분석

연역 논증

:일반적인 원칙이나 전제에서 특정한 결론을 이끌어내는 논리적 사고 방법

**가설: 소비자 물가 지수가 오르면 연준 금리도 오르고,
소비자 물가 지수가 내리면 금리도 내릴 것이다.**

Why?

중앙 은행의 목표는 물가 안정이기 때문.

- 가설 설정

- 가설 설정

- 데이터 준비

- 데이터 수집

- 데이터 분석

- 회귀 분석

정형 데이터

1. 소비자 물가 지수 (CPI, Core CPI)

2. 생산자 물가 지수 (PPI)

3. 실업률 (Unemployment Rate)

4. GDP 성장률 (GDP Growth Rate)

5. 장단기 국채 금리 (Treasury Yields)

정형 데이터

- 가설 설정
 - 가설 설정
- 데이터 준비
 - 데이터 수집
- 데이터 분석
 - 회귀 분석

1. 소비자 물가 지수 (CPI, Core CPI)

CPI: 식료품비, 주거비, 의료비, 교통비 등 소비자가 구입하는 상품과 서비스의 가격 변동 지수.

Core CPI: CPI에서 변동성이 높은 식료품과 에너지 가격을 제외한 지수.

2. 생산자 물가 지수 (PPI)

3. 실업률 (Unemployment Rate)

4. GDP 성장률 (GDP Growth Rate)

5. 장단기 국채 금리 (Treasury Yields)

■ 가설 설정

○ 가설 설정

■ 데이터 준비

○ 데이터 수집

■ 데이터 분석

○ 회귀 분석

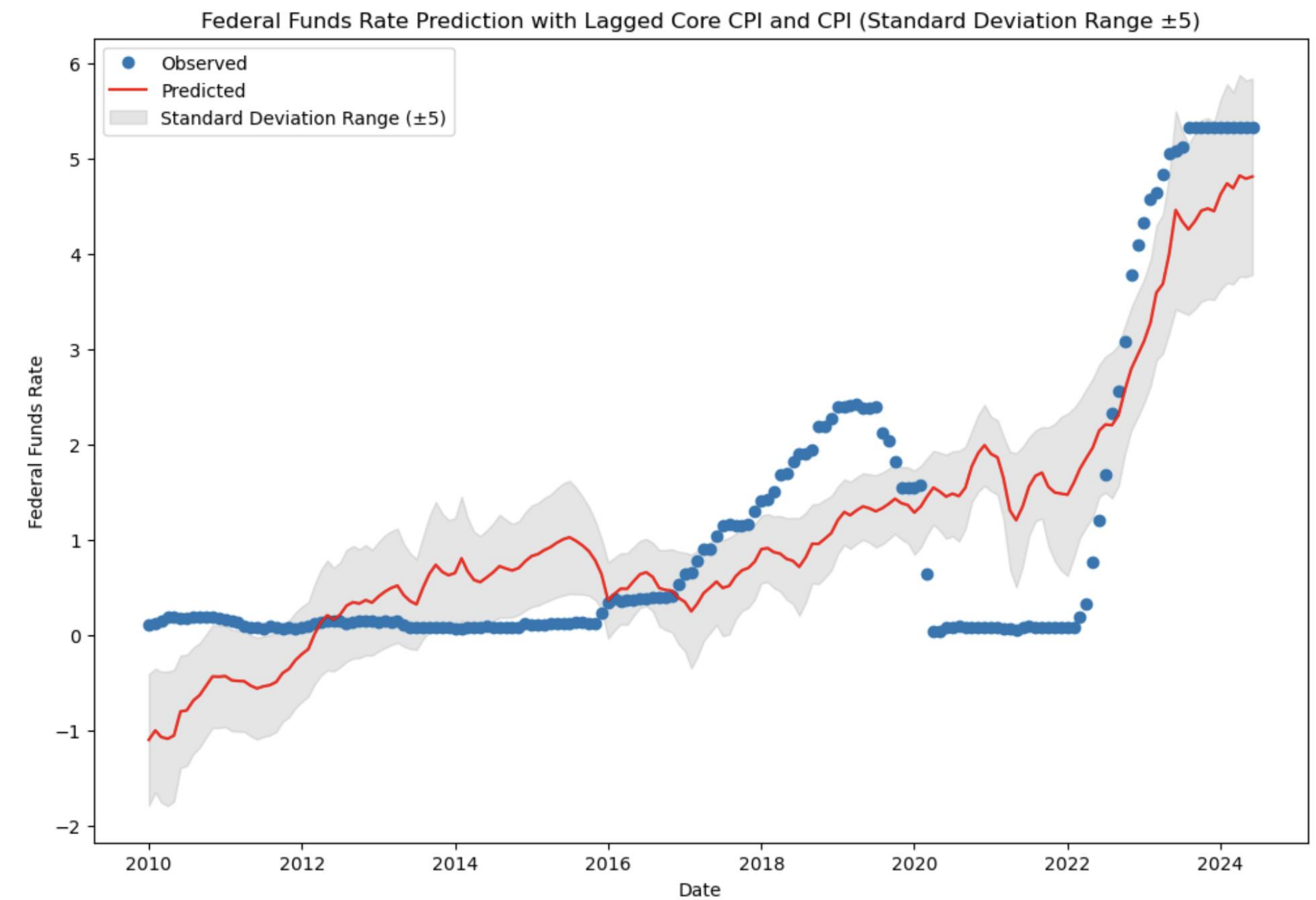
```
# 변수 선택
X = merged_df[['CORE_CPI_LAG6', 'CPI_LAG12']]
y = merged_df['FED_FUNDS']

# 회귀 분석
X = sm.add_constant(X)
model = sm.OLS(y, X).fit()

predictions = model.get_prediction(X)
pred_summary = predictions.summary_frame(alpha=0.05)
pred_mean = pred_summary['mean']
pred_mean_se = pred_summary['mean_se']

ci_lower_std = pred_mean - 5 * pred_mean_se
ci_upper_std = pred_mean + 5 * pred_mean_se
```

연준 기준 금리 예측 그래프



가설 검정

1. 소비자 물가 지수 (CPI) - 12개월 시차

1.1 상관계수 : 0.81

1.2 회귀계수 : 0.05

1.3 P-값 : 2.09E-43

2. Core CPI - 6개월 시차

2.1 상관계수 : 0.79

2.2 회귀계수 : 0.04

2.3 P-값 : 6.83E-40

3. 인과관계 확인

예상대로 물가가 금리를 선행해서 유의미한 상관관계 존재.

결론

CPI와 Core CPI는 금리를 각각 12개월, 6개월 앞에서 가장 설명력을 보여줌

김동현: 데이터를 통해서 생각을 검증하면서 기존에 미국 연준 금리 시스템과 경제의 관계에 대해 몰랐던 점도 알게 되었고 이해가 더 깊어 졌습니다.

김주성: 이번 세미 프로젝트를 진행하면서 모델링이 특정 상황에서 어떻게 성능을 발휘하는지 배울 수 있는 시간 이었습니다. 또한, 기준금리에 영향을 미치는 여러 경제적 사회적 요인들을 알아 볼 수 있는 좋은 기회였습니다.

백동열: 크롤링, 텍스트 분석, 딥러닝 코딩 협업을 통해 데이터 수집/텍스트 분석 기법/딥러닝 프로세스를 이해할 수 있었습니다.

서웅진: 데이터 분석의 흐름에 대하여 이해할 수 있었고, 연준 금리에 대하여 공부함으로써 경제에 대하여 조금이나마 배울 수 있었습니다.

Thank You.

연방준비제도 기준 금리 예측

출처 및 깃허브 주소

2p

- 1. 연방준비제도 로고 - 네이버 블로그
<https://blog.naver.com/torso001/222069965145>
- 2. 유럽 사진 - 한국인이 선호하는 첫 유럽 여행 추천 여행지 BEST 7 - 여행특특
<https://www.tourtoc.com/news/articleView.html?idxno=4507>
- 3. 천원 지폐 - 개인 사진 - 작성자 본인 제공(서웅진)
- 4. 악수 - 비즈니스 영어회화 일정 말하기 - 네이버 블로그
https://blog.naver.com/yell_wlstn/80142624721
- 5. 산업 이미지 - 4차 산업혁명 시대 기술의 혁신은 무엇? - 네이버 블로그
<https://blog.naver.com/kyungmyung2015/223485824433>

깃허브 주소

https://github.com/UngJinSeo/MLP_Finance_Fed