

Quant Issue

GPT 코드 인터프리터,

“야 너두 할 수 있어”

Quant 김종영

02) 6915-5653

jujub0251@ibks.com



IBK기업은행 금융그룹

IBK투자증권

본 조사분석자료는 당사 리서치본부에서 신뢰할 만한 자료 및 정보를 바탕으로 작성한 것이나 당사는 그 정확성이나 완전성을 보장할 수 없으며, 과거의 자료를 기초로 한 투자참고 자료로서 향후 추가 움직임은 과거의 패턴과 다를 수 있습니다. 고객께서는 자신의 판단과 책임 하에 종목 선택이나 투자시기에 대해 최종 결정하시기 바라며, 본 자료는 어떠한 경우에도 고객의 증권 투자 결과에 대한 법적 책임소재의 증빙자료로 사용될 수 없습니다.

CONTENTS

1. Chat GPT Code Interpreter 개요

- Chat GPT Code Interpreter 개요
- Chat GPT Code Interpreter 설정 및 사용 방법
- 투자 업무 자동화 전망
- 메모리 부족 이슈

2. GPT4 Code interpreter 시각화 사례

- 차트, 데이터 시각화 (주가 차트, PER/EPS 국면 분석 등)
- 보고서 작성 (경제지표 차트 생성 및 데이터 설명 보고서)

3. GPT4 Code interpreter 모델링 사례

- 시장 모델 개발 사례
(다중회귀, 머신러닝을 활용한 코스피 밴드, 코스피 적정 PBR 전망)
- 개별 종목 분석 모델 개발 사례
(삼성전자 어닝서프라이즈 예측 모델 개발)
- 통계 지수 모델 개발
(섹터 쏠림 지수 및 지수 수준에 따른 차월 섹터 수익률 백테스팅)
- 멀티 팩터 모델 개발
(머신러닝을 활용한 멀티 팩터 모델 개발)

4. GPT4 Code interpreter 사용 Tip

- GPT 업로드 엑셀 파일 정리 Tip
- Chat-GPT 모델 개발 시 모델 선택 사항

1

Chat GPT Code Interpreter 개요

Chat GPT Code Interpreter 개요

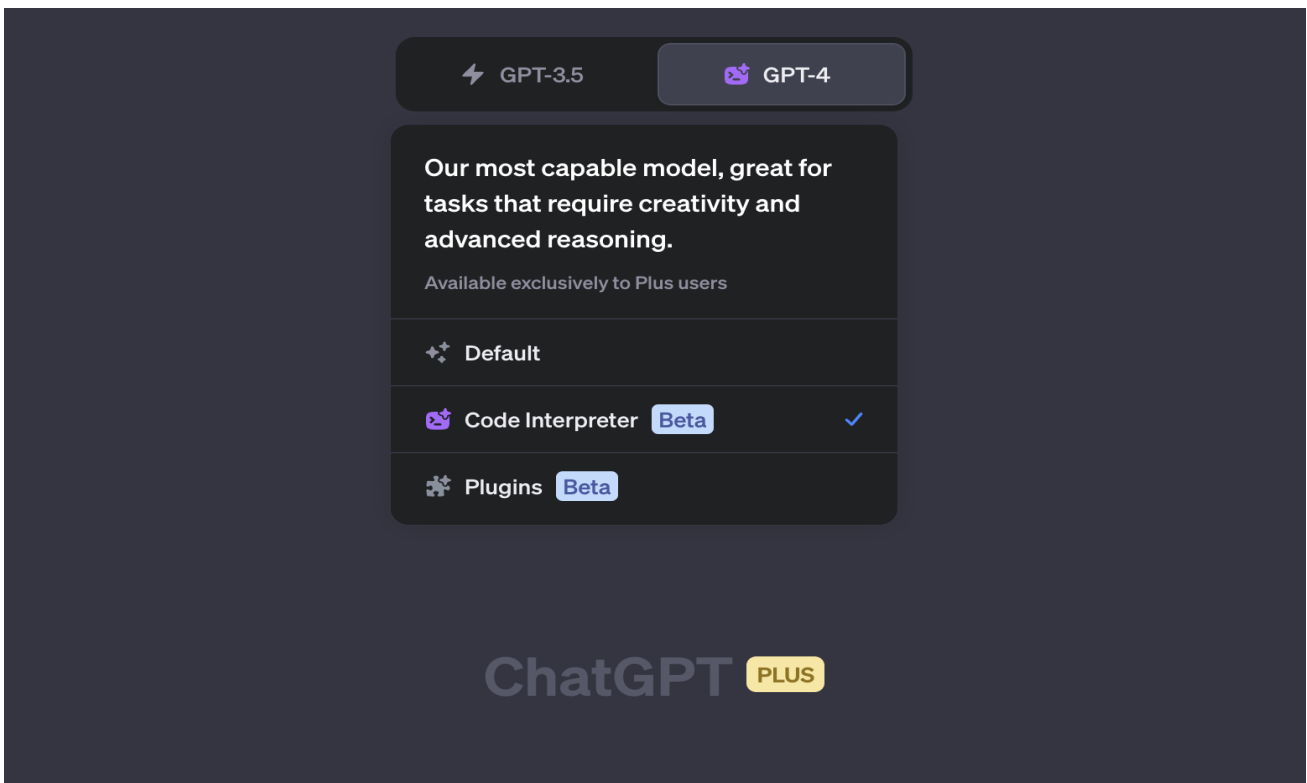
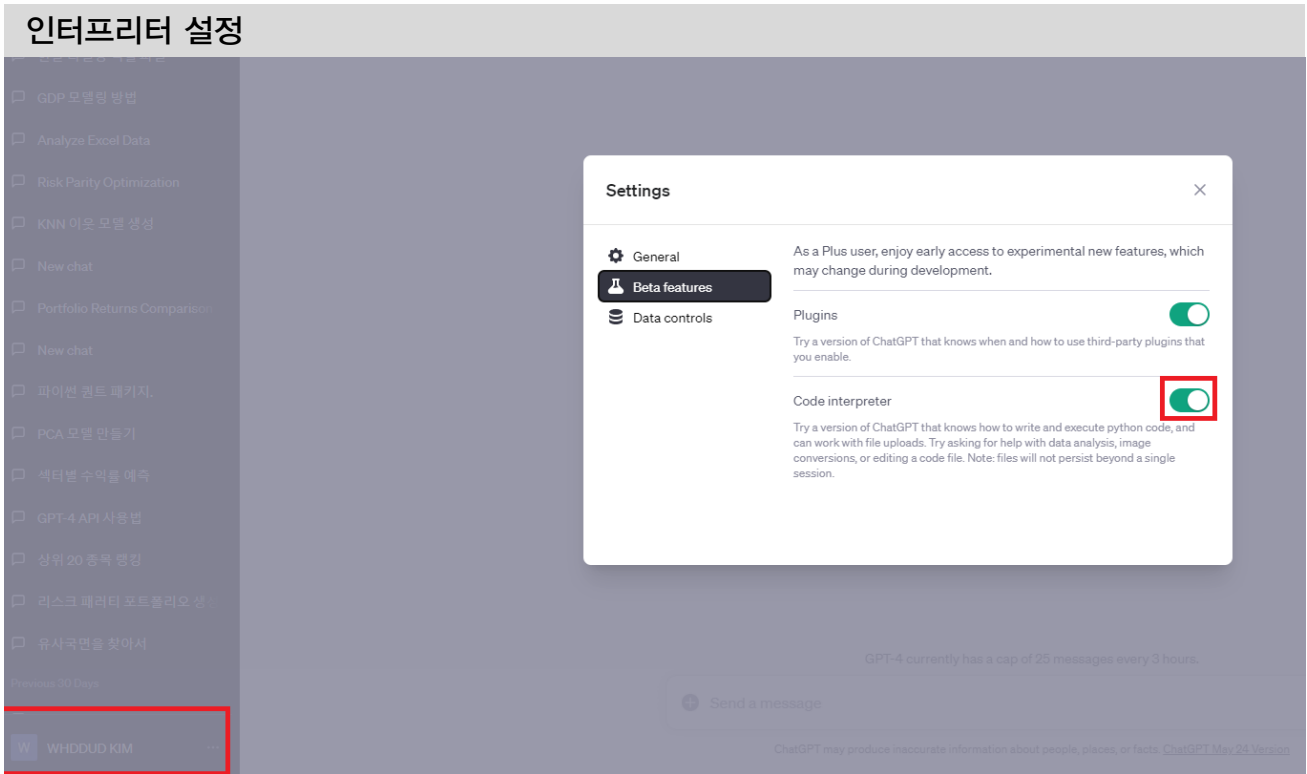
Code interpreter

사용자의 파일을 업로드하면 GPT가 작업 가능

- GPT-4 Code interpreter는 사용자를 위한 샌드 박스 환경 가동
- 샌드박스 환경에서 Python 코드를 실행할 수 있음
- 단순 코딩을 만들고 실행 뿐만 아니라, 사용자가 데이터 파일 (엑셀, PDF, 그림파일 등)을 올려서 작업 가능
- 텍스트 파일: .txt, .csv, .json, .xml 등
문서 파일: .docx, .pdf, .pptx, .xlsx 등
이미지 파일: .jpg, .jpeg, .png, .gif, .bmp, .svg 등
오디오 파일: .mp3, .wav, .aac, .flac 등
비디오 파일: .mp4, .avi, .mov, . . . 등
코드 파일: .py, .js, .java, .cpp, .c, .sh, .html, .css 등
- 업로드 파일들은 파이썬 라이브러리를 사용하여 읽음.
CSV 파일은 pandas, 이미지 파일은 PIL 또는 OpenCV로, 오디오 파일은 librosa로, 비디오 파일은 opencv-python으로 처리
- 샌드박스 환경은 인터넷 접근이 제한되어 있으므로 인터넷을 통한 데이터 다운로드 불가능
- Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn, Scipy Statsmodels 등 다양한 패키지 제공 (머신러닝까지 개발 가능)
- 인터넷 연결이 필요한 Request, Selenium, Yfinance, API 관련 코딩과 TensorFlow, Keras와 같은 딥러닝 모델은 코딩만 제공, 실행은 불가능(GPU, 메모리 등 컴퓨팅 파워 문제로 판단됨)
- 이제 초기단계로 향후 업데이트 예상

Chat GPT Code Interpreter 설정방법

- Open AI 접속, 왼쪽 하단 “...” 버튼 클릭 후 Setting 설정
- 인터프리터 활성화 → 신규 채팅에서 인터프리터 설정

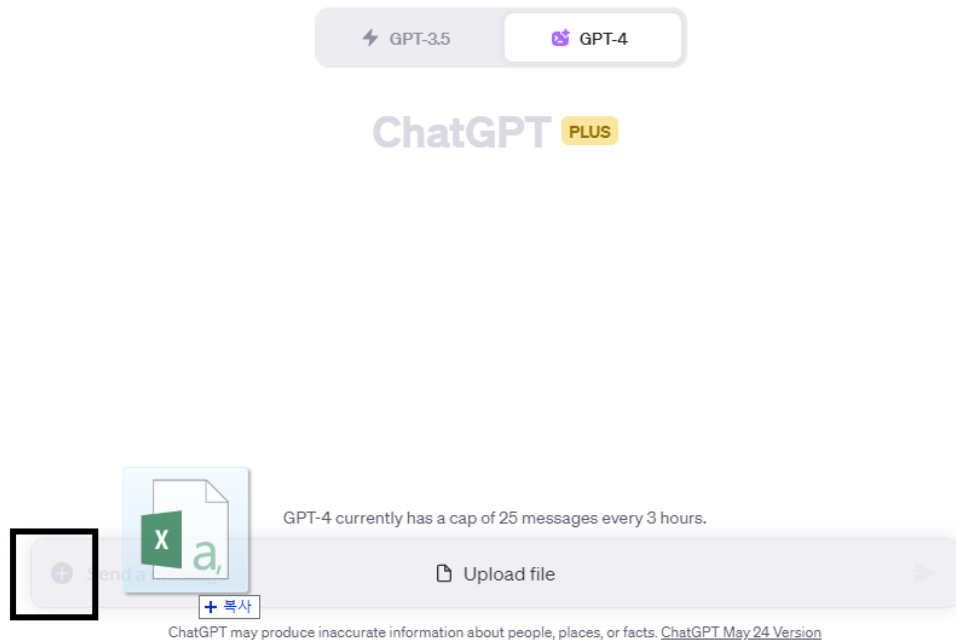


자료: Chat GPT, IBK투자증권

Chat GPT Code Interpreter 사용방법

- GPT 인터프리터 설정 시 채팅창에 “+” 등장
- “+” 버튼을 누르고 GPT-4에 제시할 파일 선택 또는 파일을 채팅 창에 드랍하면 자동으로 파일 업로드
- 최대 용량 100mb (실제 사용 후기를 보면 더 큰 파일도 들어간다고 나눔)까지 데이터 업로드 가능
- 주간, 월간 기준 금융 관련 데이터를 활용 시 파일 용량이 100mb를 초과하는 경우는 극히 드물
- 일간 데이터, 일종 데이터, 종목 숫자가 많아지는 케이스를 제외하고 대부분 사용 가능

코드 확인



자료: Chat GPT, IBK투자증권

투자 업무 자동화 전망



투자업무 AI 도입 예상 (Bold 표시는 현재 가능한 것들)

번호	투자업무		자동화 예상
1	리서치	주제, 의제 선정	리서치를 위한 데이터 선정, 필요한 모델링에 대한 판단은 아직 인간 영역
2	리서치	정량 데이터 조사	Chat GPT 쿼리 생성 +제3 정보 제공 업체 DB 조회 (GPT 플러그인 등) DB 제공 업체 자체 생성형 AI, 1년 이내 출시 예상 (Bloomberg GPT 등)
3	리서치	자료 통계 분석, 시각화, 모델링	코딩 생성 AI(깃허브 코파일럿, Codex 등) 엑셀 코파일럿 (엑셀 365 탑재 예정) 자연어 명령 → 코딩 전환 (GPT Code Interpreter)
4	리서치	정성 데이터 조사	Chat GPT를 활용한 뉴스, 리서치 보고서 정량화 가능 그러나 완벽하게 미시 경제 데이터 (기업 탐방 데이터, 산업 인터뷰 정보 등) 수집은 자동화 어려움 액티브 기반의 탐방 정보는 계속 가치 있을 것 만약 기업/산업 분야도 빅데이터화가 이루어진다면 실시간으로 분석 가능 (잠정 수출데이터가 유사한 사례)
5	리서치	자료 작성	클라우드 서비스에 정량/정성적 데이터를 업로드하면 Word, Power Point 코파일럿을 통한 문서 자동 작성
6	투자 의견 제시	매수/매도 판단	규칙, 세팅에 의해 매매를 자동으로 실행(알고리즘 투자)업무 난이도 하락 데이터 제공 시 인간처럼 추론하고 판단하는 모델은 GPT 성능 추가 업데이트 필요, 2023년 연말 GPT4 API 제공 시 튜닝 모델 개발 가능성
7	운용	종합적 판단 리서치 운용	인공일반지능(AGI) 개발 필요, 인간을 상당 부분 대체 금융, 경제, 정치, 사회 등 다양한 정량적, 정성적 변수 데이터를 실시간 모니터링 하고 상황에 따라 필요한 모델 개발 구축, 자체 리서치 실시 스스로 구축한 모델, 리서치를 바탕으로 포트폴리오 구축

자료: IBK투자증권

메모리 부족 이슈

- 메모리 이슈: GPT Code Interpreter 사용 중 메모리가 부족하다는 답변을 다수 받음
- GPT Code Interpreter는 사용자가 데이터를 엑셀 파일로 업로드하고 분석 할 수 있음
- 머신러닝 분석 중 모델 개발 - 수정 - 다시 개발 - 수정하는 과정에서 메모리 부족 문구 자주 발생
- 정말 메모리가 부족한 것인지, 채팅 채널 한 개 당 제공하는 메모리가 제한적인 것인지 알 수 없지만, 앞으로 딥러닝, 이미지/영상 처리, 대용량 빅데이터 처리까지 자동화를 고려하면 메모리 부족 이슈는 계속될 것
- 개인 로컬 컴퓨팅 파워 사용에서 GPT 기반 클라우드 컴퓨팅 파워 사용으로 전환하는 과정

메모리 부족 이슈

Finished working

Show work ▼

죄송합니다. 코드 실행 중에 메모리 부족 문제가 발생했습니다. 이 문제는 주로 대량의 데이터를 처리하거나 복잡한 연산을 수행할 때 발생합니다.

현재 이 환경에서는 메모리를 추가로 할당하는 것이 불가능합니다. 따라서 코드를 수정하여 메모리 사용량을 줄이거나, 필요하지 않은 객체를 삭제하여 메모리를 해제해야 합니다.

본 코드에서는 모델의 피쳐 중요도를 시각화하는 과정에서 문제가 발생했습니다. 이는 일반적으로 메모리를 많이 사용하지 않는 연산입니다. 따라서 이 문제는 다른 부분에서 메모리가 과도하게 사용되었음을 의미합니다.

모델 학습 과정에서 생성된 중간 객체를 삭제하거나, 데이터를 더 작은 단위로 분할하여 처리하는 등의 방법을 사용하여 이 문제를 해결해 보겠습니다.

Finished working

Show work ▼

🔄 Regenerate response

죄송합니다. 그래프를 표시하는 데 여전히 메모리 부족 문제가 발생하고 있습니다.

+ Send a message

ChatGPT may produce inaccurate information about people, places, or facts. [ChatGPT May 24 Version](#)

2

GPT4 Code interpreter 시각화 사례


Code Interpreter 사용 사례: 시각화

- GPT4 코드인터프리터를 활용해 데이터 시각화
- GPT4 코드인터프리터는 사용자가 제시한 데이터를 읽고 전처리 (데이터 정리)-분석-모델개발-시각화를 코딩을 통해서 실행 가능

데이터 시각화

Chat GPT Q&A

Q : 삼성전자의 주가와 12개월 선행 EPS 성장률 데이터이다. 읽어라

 수익률.csv

A :

Finished working Show work

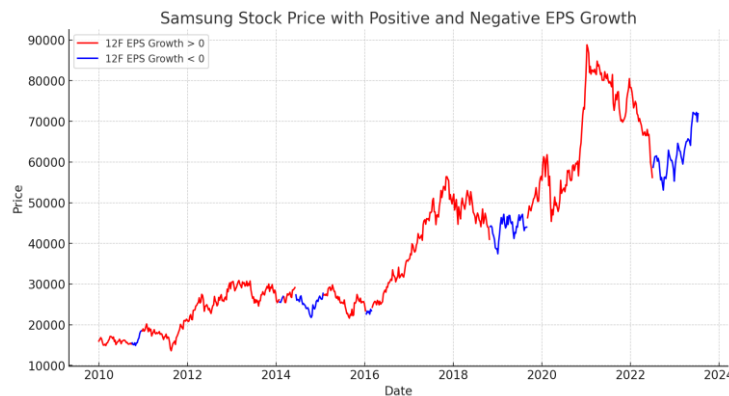
데이터 세트에는 한국 주식 시장의 날짜와 다양한 부문의 성과가 포함되어 있는 것으로 보입니다. 열은 섹터 이름이고 행은 서로 다른 날짜에 해당하며 성과 수치는 수익률 또는 성장률을 나타냅니다.

Q : 주가차트를 그리고싶다.

그런데 EPS Growth Fwd.12M/LTM가 "+" 인구간은 빨간색 주가차트로
EPS Growth Fwd.12M/LTM가 "-" 인구간은 파란색 주가차트로 그려줘

A :

Finished working Show work



다음은 요청하신 차트입니다.

이 단일 선 차트는 시간 경과에 따른 삼성의 주가를 나타내며 선의 색상은 EPS Growth Fwd.12M/LTM의 값에 따라 변경됩니다.

선의 빨간색 부분은 EPS Growth Forward 12M/LTM이 양수였던 기간을 나타냅니다.
선의 파란색 부분은 EPS Growth Forward 12M/LTM이 음수였던 기간을 나타냅니다.
도움이 필요한 다른 사항이 있으면 알려주세요.

자료: Chat GPT, IBK투자증권

Code Interpreter 사용 사례: 시각화

- GPT가 작업한 내용은 코드를 통해 확인가능
- 코드 복사를 통해 다시 사용 가능, 혹은 추가 업데이트 가능

코드 확인

Chat GPT Q&A

```
# Create a new figure
plt.figure(figsize=(12,6))

# Plot each continuous segment separately and add labels for legend
for i in range(len(indices) - 1):
    start, end = indices[i], indices[i+1]
    segment = new_data.iloc[start:end]
    if segment['EPS Growth Fwd.12M/LTM'].iloc[0] > 0:
        color = 'red'
        label = '12F EPS Growth > 0'
    else:
        color = 'blue'
        label = '12F EPS Growth < 0'
    plt.plot(segment['D A T E'], segment['PRICE'], color=color, label=label)

# To avoid duplicate labels in the legend, we extract the handles and labels and pass them to
plt.legend
handles, labels = plt.gca().get_legend_handles_labels()
by_label = dict(zip(labels, handles))
plt.legend(by_label.values(), by_label.keys())

plt.xlabel('Date')
plt.ylabel('Price')
plt.title('Samsung Stock Price with Positive and Negative EPS Growth')
plt.grid(True)
plt.show()
```

자료: Chat GPT, IBK투자증권

Code Interpreter 사용 사례: 시각화

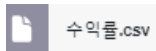
- GPT4 코드인터프리터를 활용해 데이터 시각화
- 전처리(데이터 정리)-분석-모델개발-시각화를 코딩을 통해서 실행 가능

데이터 시각화

Chat GPT Q&A

Q : 코스피 섹터 월간 수익률이다.

섹터 수익률의 히트맵을 그려라, 날짜는 YY-MM 양식, 섹터 이름 Y축, 날짜 X축



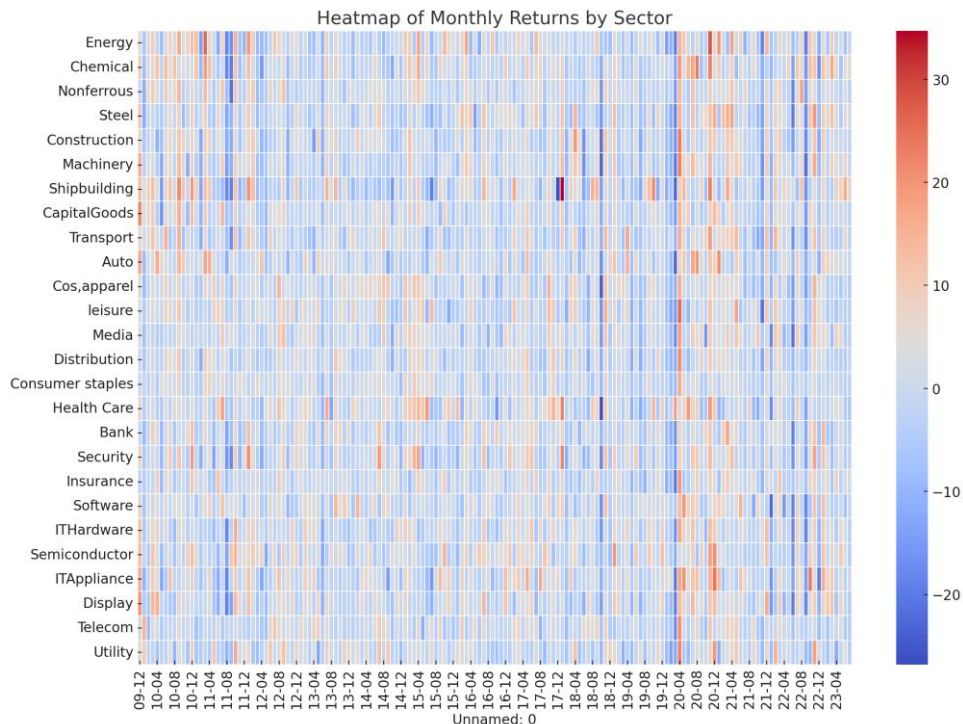
A :

Finished working Show work

업로드한 데이터는 코스피의 여러 섹터들에 대한 월간 수익률을 나타내는 것으로 보입니다. 첫 번째 열은 날짜를 나타내며, 그 이후의 각 열은 각각 다른 섹터를 나타냅니다.

다음으로, 이 데이터를 사용해 히트맵을 그려보겠습니다. 히트맵은 데이터의 패턴, 변동성, 그리고 예외 사항 등을 시각화하는 데 유용한 방법입니다. 이 경우, 각 섹터의 월간 수익률 변화를 시각화하는 데 사용될 수 있습니다.

히트맵을 그리기 전에, 날짜 데이터를 Pandas의 datetime 형식으로 변환하겠습니다. 이 작업을 통해 데이터를 시간 순서대로 잘 정렬하고, 시각화할 때 날짜 형식을 제어할 수 있습니다.



자료: Chat GPT, IBK투자증권

Code Interpreter 사용 사례: 시각화

데이터 시각화

Chat GPT Q&A

Q : 가장 최근 2년치만 보여주고

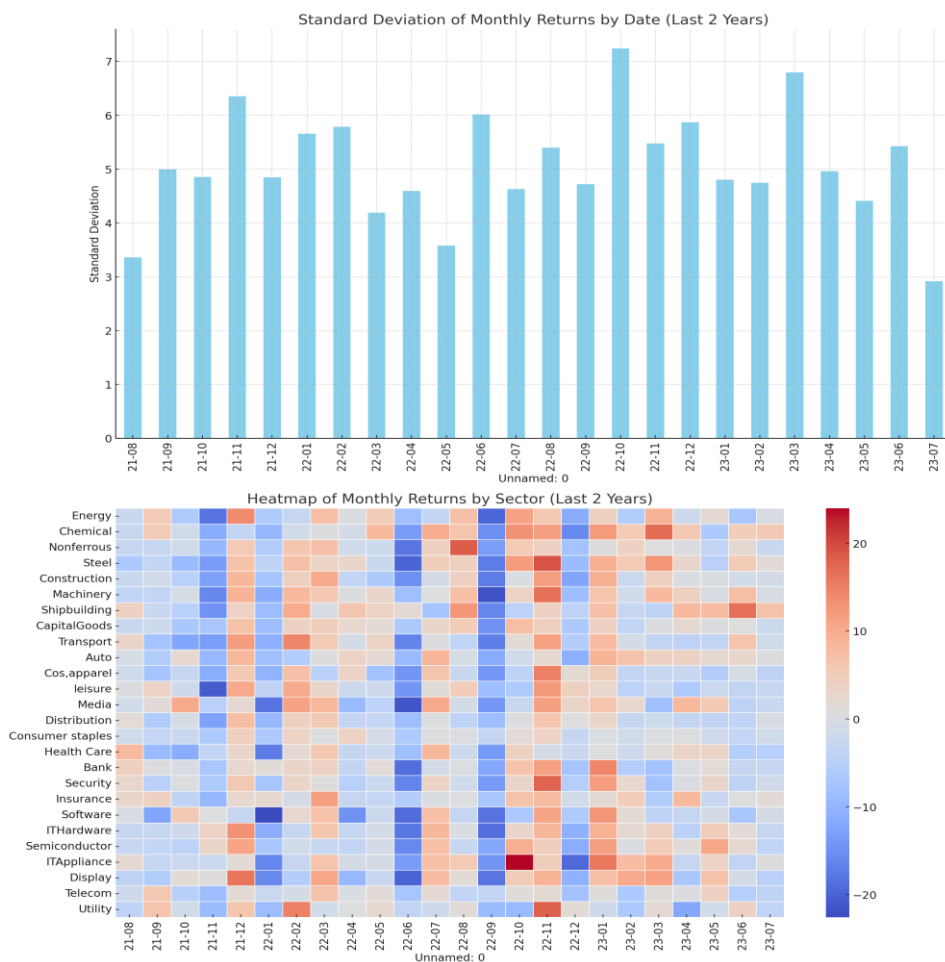
섹터수익률의 표준편차를 위쪽에 보여줘



수익률.csv

A :

Finished working Show work



위쪽의 바 차트는 각 월별 섹터 수익률의 표준 편차를 나타냅니다. 표준 편차는 섹터 수익률의 변동성을 측정하는 한 가지 방법으로, 값이 클수록 해당 월의 섹터 수익률이 더 많이 변동했음을 의미합니다. 이를 통해 특정 월의 섹터 수익률이 얼마나 안정적인지, 또는 불안정한지를 파악할 수 있습니다.

아래쪽의 히트맵은 최근 2년 동안의 각 섹터별 월간 수익률을 나타냅니다. 이를 통해 특정 섹터가 어떤 월에 높은 수익률을 보였는지, 또는 어떤 월에 낮은 수익률을 보였는지를 빠르게 파악할 수 있습니다.

두 그래프를 함께 보면, 특정 월의 섹터 수익률의 변동성과 그 섹터의 월간 수익률 패턴을 동시에 이해할 수 있습니다.

자료: Chat GPT, IBK투자증권

Code Interpreter 사용 사례: 시각화

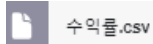
데이터 시각화

Chat GPT Q&A

Q : 주가 YoY 수익률, PER YoY, EPS YoY, 주가 데이터이다.

EPS YoY, PER YoY 증가, 감소에 따라 스캐터 차트 양식으로 4개 국면을 그려 달라

각 국면별 당시 주가와 52개 이후 주가를 비교했을 때 상승하는 비율을 보여줘



A :

Finished working Show work

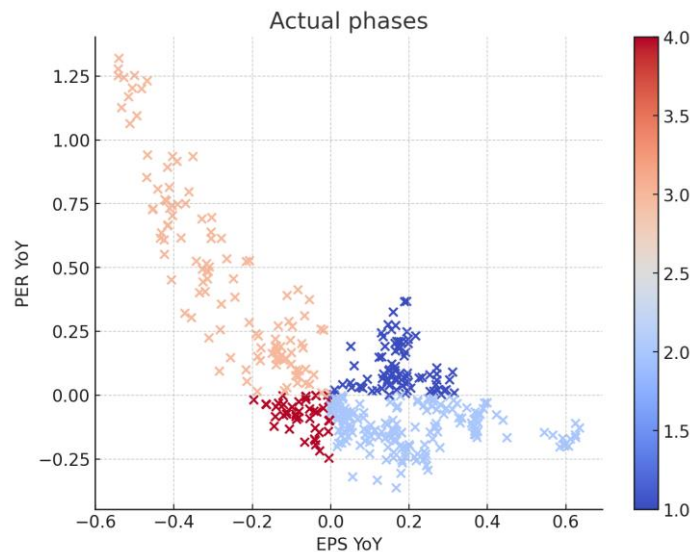
EPS YoY와 PER YoY 값을 기반으로 실제 4개의 국면을 보여줍니다.

국면 1: $EPS\ YoY > 0, PER\ YoY > 0$, 국면 2: $EPS\ YoY > 0, PER\ YoY < 0$

국면 3: $EPS\ YoY < 0, PER\ YoY > 0$, 국면 4: $EPS\ YoY < 0, PER\ YoY < 0$

각 색깔은 하나의 국면을 나타냅니다. 점들은 EPS YoY와 PER YoY 값에 대한 실제 데이터 포인트이며, 그 색깔은 해당 데이터 포인트의 실제 국면을 나타냅니다.

이 차트를 통해 각 국면의 데이터 포인트가 어떻게 분포하는지 시각적으로 확인할 수 있습니다.



주가가 52기 이후에 상승한 비율은 다음과 같습니다:

국면 1: 약 70.73%

국면 2: 약 41.57%

국면 3: 약 62.93%

국면 4: 약 75.00%

이는 각 국면에서 주가가 52기 후에 상승할 확률을 나타냅니다. 따라서 국면 1과 국면 4에서 주가가 상승할 가능성이 가장 높으며, 국면 2에서는 상승할 가능성이 가장 낮습니다. 이 정보를 통해 각 국면이 주가에 미치는 영향에 대한 추가적인 통찰력을 얻을 수 있습니다.


자료: Chat GPT, IBK투자증권

Code Interpreter 사용 사례: 시각화

정량 데이터, 정성 데이터 제공 이후 보고서 작성 사례

Chat GPT Q&A

Q : 삼성전자의 분기실적표이다.

 요약본문.xlsx

A :

Finished working Show work ▾

제공된 엑셀파일은 삼성전자의 분기 재무제표입니다.
첫행은 주요 재무정보를 나타냅니다 ... (이하 생략)

Q : 삼성전자의 2023년 3분기, 4분기 실적 전망과 반도체, 디램 가격 상승에 대한 보고서 및 기사 자료이다. (단순 기사, 리포트 텍스트 본문 복사 붙여넣기 자료)

 기사 정리.txt

A :

텍스트 파일에는 삼성의 사업 실적 및 향후 전망과 관련된 여러 한국 뉴스 기사가 포함되어 있습니다. 다음은 각 기사의 핵심 사항을 요약한 것입니다. 메모리 업종의 재고자산 평가손실 감소와 공급 감소 및 수요 회복에 따른 DRAM 평균 판매가격 상승으로 개선될 전망이다. (이하 생략)

Q : 내가 올려 준 데이터 중 2023년 3분기 4분기는 컨센서스 데이터이다. 컨센서스 데이터는 전망치이다. 위에 올려준 전망 기사, 보고서 내용을 바탕으로 컨센서스를 설명하는 글을 작성하자

A :

2023년 3분기와 4분기에 대한 삼성전자의 컨센서스 예상치는 다음과 같습니다:

3분기:

매출: 685,148억 원
영업이익: 34,305억 원

4분기:

매출: 697,867억 원
영업이익: 47,044억 원

삼성전자의 매출과 영업이익은 2023년 3분기부터 증가세를 보이기 시작하며, 4분기에는 더욱 증가할 것으로 예상됩니다. 이는 주로 메모리 반도체, 특히 D램의 판매가격 상승과 함께 고대역폭 메모리인 HBM의 시장 점유율 확대에 기인합니다.

삼성전자는 메모리 부문에서 HBM의 시장 점유율을 확대하고 있으며, 이는 AI와 같은 고성능 컴퓨팅 분야에서 데이터 처리 요구량이 증가함에 따라 HBM 수요가 크게 증가하고 있기 때문입니다. 이런 추세가 계속된다면 삼성전자의 매출과 영업이익은 더욱 증가할 것으로 보입니다. 또한, D램 가격의 상승도 삼성전자의 실적 개선에 기여할 것으로 예상됩니다. D램 가격은 공급 축소와 수요 회복 효과에 따라 상승하고 있으며, 이는 삼성전자의 메모리 부문 실적 개선을 도울 것입니다.

결론적으로, 삼성전자의 매출과 영업이익 증가는 메모리 반도체 시장에서의 성장과 함께 HBM과 D램 가격 상승에 크게 기인하고 있습니다. 이러한 추세가 지속되면 삼성전자의 실적은 앞으로도 계속 개선될 것으로 보입니다.

자료: Chat GPT, IBK투자증권

3

GPT4 Code interpreter 모델링 사례

Code Interpreter 사용 사례: 코스피 지수 전망

수출, M2 지표로 코스피 추정 모델 만들기: 다중 회귀 모델 만들기

Chat GPT Q&A

Q : 한국 종합주가지수 코스피 지수(C열)와 경제지표 수출금액(A열), M2 금액(B열)이다.

경제지표 수출금액, M2 금액을 바탕으로 코스피지수를 전망하는 통계 모델을 만들어보자



코스피 추정 csv.xlsx

A :저 제공하신 파일을 불러와서 데이터를 확인해보겠습니다. 파일을 불러올 때 사용하는 라이브러리는 pandas이며, 이는 Python에서 데이터 분석을 위해 자주 사용되는 도구입니다.

엑셀 파일을 로드하고, 데이터의 첫 몇 행을 살펴보겠습니다.

Finished working

Show work

데이터는 아래와 같이 구성되어 있습니다:

첫 번째 열: 날짜 (1986년 3월부터 시작)

두 번째 열 (A열): 수출금액 (Export)

세 번째 열 (B열): M2 금액

네 번째 열 (C열): 코스피지수 (KOSPI)

이제 이 데이터를 바탕으로 수출금액과 M2 금액을 독립 변수로, 코스피 지수를 종속 변수로 하는 선형 회귀 모델을 만들어 보겠습니다.

선형 회귀 모델은 독립 변수의 가중치 합과 종속 변수 사이의 관계를 추정하는 통계 모델입니다. 이 모델을 통해 수출금액과 M2 금액이 코스피 지수에 어떤 영향을 미치는지 파악할 수 있습니다.

사용할 라이브러리는 sklearn의 LinearRegression입니다. 이 라이브러리는 회귀 모델을 쉽게 만들고 훈련시킬 수 있는 도구를 제공합니다. 이제 모델을 만들고 훈련시켜 보겠습니다.

선형 회귀 모델의 결과는 다음과 같습니다:

Mean Squared Error (MSE): 약 45559.25.

R-squared (R^2): 약 0.919. 이는 이 모델이 데이터의 약 91.9%를 설명하고 있다는 것을 의미합니다.

계수 (Coefficients): 수출금액은 약 0.026, M2 금액은 약 0.00024. 이는 각 독립 변수가 종속 변수에 미치는 영향을 나타냅니다. 즉, 수출금액이 1단위 증가하면 코스피 지수는 약 0.026 증가하고, M2 금액이 1단위 증가하면 코스피 지수는 약 0.00024 증가합니다.

절편 (Intercept): 약 350.70.

이러한 결과를 바탕으로, 이 선형 회귀 모델은 수출금액과 M2 금액을 바탕으로 코스피 지수를 예측하는데 약간의 설명력을 가지고 있지만, 그 설명력은 매우 제한적인 것으로 보입니다. 이는 R^2 값이 매우 낮기 때문입니다.

그러나, 이는 간단한 선형 회귀 모델이며, 실제로는 수출금액과 M2 금액 외에도 코스피 지수에 영향을 미치는 다양한 요인들이 있을 것입니다. 따라서 이 모델을 실제 코스피 지수 예측에 바로 적용하는 것은 주의가 필요합니다.

자료: Chat GPT, IBK투자증권

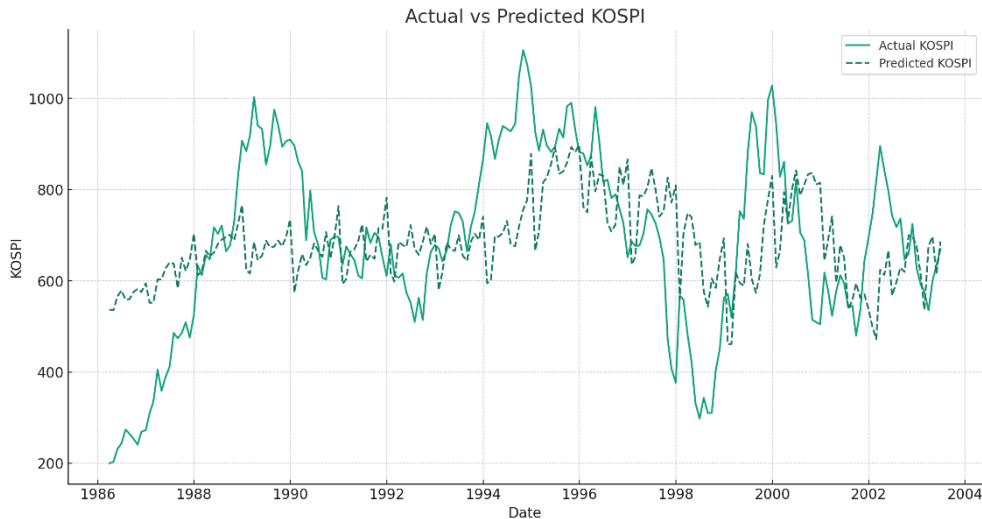
Code Interpreter 사용 사례: 코스피 지수 전망

수출, M2 지표로 코스피 추정 모델 만들기: 다중 회귀 모델 만들기

Chat GPT Q&A

Finished working

Show work



위 차트에서 파란색 실선은 실제 코스피 지수를, 주황색 점선은 다중 선형 회귀 모델을 통해 예측한 코스피 지수를 나타냅니다.

실제 코스피 지수와 예측한 코스피 지수 사이에는 일정한 차이가 있음을 알 수 있습니다. 이 차이는 모델의 성능과 정확도에 따라 달라지며, 더 복잡한 모델을 사용하거나 추가적인 변수를 포함시킴으로써 줄일 수 있습니다.

또한, 회귀 모델의 예측력을 개선하기 위해서는 다양한 방법들이 있습니다. 예를 들면, 피쳐 엔지니어링을 통해 새로운 변수를 만들거나, 다른 회귀 모델(릿지, 라쏘, 엘라스틱넷 등)을 사용하거나, 머신러닝 기반의 회귀 모델을 사용하는 등의 방법이 있습니다. 그러나 이러한 방법들은 데이터의 특성, 문제의 복잡도, 가능한 계산 리소스 등에 따라 선택해야 합니다.

Code Interpreter 사용 사례: 코스피 지수 전망

수출, M2 지표로 코스피 추정 모델 만들기: 다중 회귀 모델 만들기

Chat GPT Q&A

Q : 예상 코스피 차트 색 - 파란색

예상 코스피에 1,2 표준편차를 더하고 빼서 새로운 선 차트를 만들어서 보자

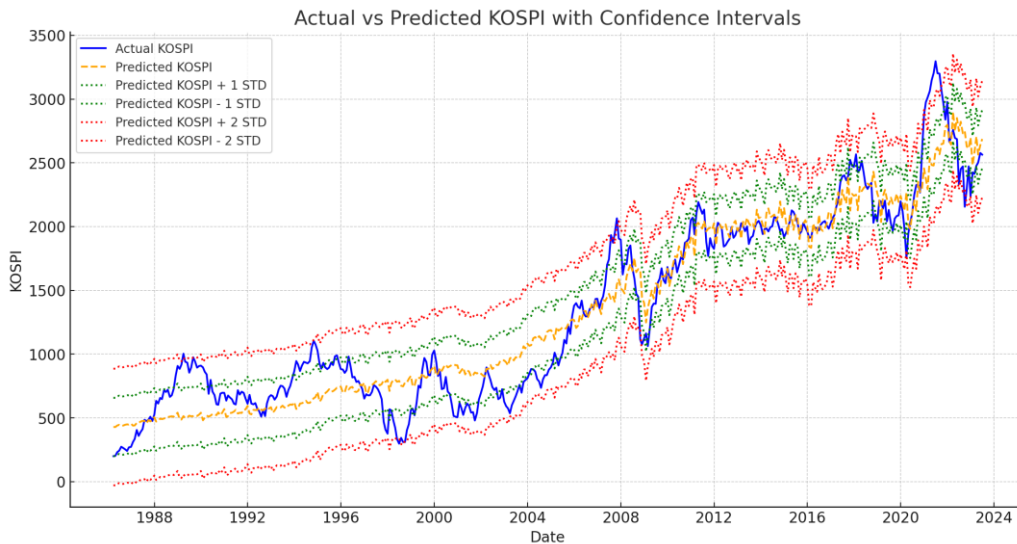
그럼 최종적으로 코스피, 예상코스피, 예상코스피+1std, 예상코스피+2std, 예상코스피-1std, 예상코스피-2std

또한 걱정코스피와 실제 코스피의 차이 / 표준편차를 차트로 그려라

A :

Finished working

Show work



위 차트는 실제 코스피 지수(파란색), 예상 코스피 지수(주황색), 그리고 예상 코스피 지수의 1표준편차와 2표준편차 범위(녹색과 빨간색 점선)를 나타냅니다.

이 차트를 통해 예상 코스피 지수의 변동 범위를 살펴볼 수 있습니다. 녹색 점선은 예상 코스피 지수의 1표준편차 범위를 나타내며, 빨간색 점선은 2표준편차 범위를 나타냅니다.

이를 통해 실제 코스피 지수가 예상 코스피 지수를 중심으로 어느 정도 범위 내에서 변동하는지를 알 수 있습니다. 이는 모델의 예측이 얼마나 정확한지, 그리고 어느 정도의 불확실성을 가지는지를 나타내는 좋은 지표가 될 수 있습니다.

그러나, 이 차트를 통해 실제 투자 결정을 내리는 것은 매우 주의가 필요합니다. 왜냐하면 이는 간단한 선형 회귀 모델을 기반으로 하고 있으며, 실제 코스피 지수는 수출금액과 M2 금액 외에도 많은 다른 요인들에 의해 영향을 받기 때문입니다.

자료: Chat GPT, IBK투자증권

Code Interpreter 사용 사례: 코스피 지수 전망

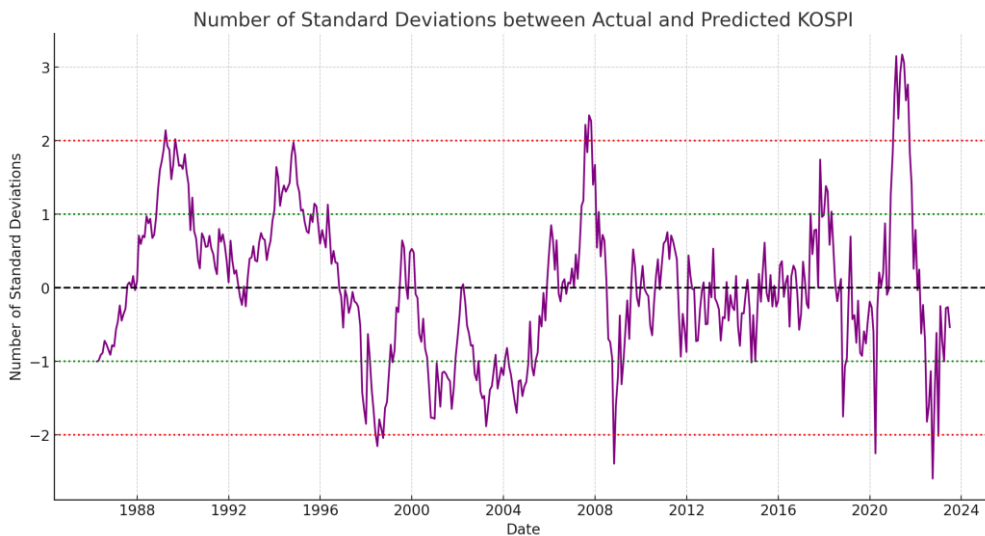
수출, M2 지표로 코스피 추정 모델 만들기: 다중 회귀 모델 만들기

Chat GPT Q&A

Q : 예상 코스피와 실제 코스피의 차이나는 정도가 몇 표준편차인지 새로 차트를 그려라

A :

Finished working Show work



위 차트는 실제 코스피 지수와 예상 코스피 지수의 차이를 표준편차 단위로 나타낸 것입니다.

보라색 선은 실제 코스피 지수와 예상 코스피 지수의 차이를 표준편차로 나눈 값입니다. 즉, 이 선이 $y=0$ 선(검은색 점선)에서 얼마나 떨어져 있는지를 보면 실제 값이 예상값에서 얼마나 많은 표준편차만큼 떨어져 있는지를 알 수 있습니다.

녹색 점선은 ± 1 표준편차를 나타내며, 빨간색 점선은 ± 2 표준편차를 나타냅니다.

이 차트를 통해 모델의 예측이 얼마나 실제 값을 벗어나는지, 그리고 그 차이가 얼마나 일관되게 발생하는지를 알 수 있습니다. 이는 모델의 예측 정확도를 평가하는 데 유용한 지표가 될 수 있습니다.

Code Interpreter 사용 사례: 코스피 지수 전망

수출, M2 지표로 코스피 추정 모델 만들기: 모델 개발 확장

Chat GPT Q&A

Q : 다른 추천 모델이 있는가 ?

A : 코스피 지수 예측에는 다양한 모델이 사용될 수 있습니다. 선형 회귀 모델 외에도, 여러 고급 모델들이 있습니다. 그 중 몇 가지를 소개하겠습니다:

****릿지 회귀(Ridge Regression)****와 **라쏘 회귀(Lasso Regression)**: 이들은 선형 회귀에 규제를 추가하여 모델의 복잡도를 제어하고 과적합을 방지합니다. 릿지 회귀는 가중치의 제곱을 규제로 사용하고, 라쏘 회귀는 가중치의 절대값을 규제로 사용합니다.

다항 회귀(Polynomial Regression): 이는 독립 변수의 고차 항을 모델에 포함시키는 방법으로, 비선형 관계를 모델링할 수 있습니다.

****결정 트리(Decision Tree)****와 **랜덤 포레스트(Random Forest)**: 이들은 비선형 관계와 상호작용을 잘 캡처하며, 특히 랜덤 포레스트는 여러 결정 트리를 결합하여 과적합을 방지하고 성능을 향상시킵니다.

신경망(Neural Networks): 이는 복잡한 비선형 관계를 모델링할 수 있으며, 딥러닝 모델은 특히 시계열 데이터 예측에 잘 사용됩니다.

ARIMA, SARIMA, Prophet 등의 시계열 모델: 이들은 시간적 순서를 고려하여 데이터를 분석하고 예측합니다. 코스피 지수와 같은 시계열 데이터에 잘 맞습니다.

이 중 어떤 모델을 사용할지는 데이터의 특성, 문제의 복잡도, 사용 가능한 계산 리소스 등에 따라 달라집니다. 또한, 최적의 모델을 찾기 위해서는 여러 모델을 시도해 보고 그 성능을 비교해야 합니다.

다른 모델을 사용하여 코스피 지수를 예측해 보시겠습니까?

Q : 최종결과물: 차트까지 제공, 예상치와 실제치 차이/표준편차까지 그려보자

모델: 라쏘 회귀(Lasso Regression), Arima

최종결과물: 차트까지 제공, 예상치와 실제치 차이/표준편차까지 그려보자

두개 다 해보자

A : 네, 요청하신대로 라쏘 회귀 모델과 ARIMA 모델 두 가지를 사용하여 코스피 지수를 예측해보겠습니다.

먼저 라쏘 회귀 모델부터 시작하겠습니다. 라쏘 회귀는 선형 회귀의 한 종류로, 모델의 복잡도를 제어하고 과적합을 방지하기 위해 가중치의 절대값에 대한 규제를 추가한 모델입니다.

라쏘 회귀 모델을 훈련시키고, 예측값과 실제값의 차이를 표준편차 단위로 나타내는 차트를 만들어 보겠습니다.

라쏘 회귀 모델의 결과는 다음과 같습니다:

Mean Squared Error (MSE): 약 45559.25.

R-squared (R^2): 약 0.919. 이는 이 모델이 데이터의 약 91.9%를 설명하고 있다는 것을 의미합니다.

계수 (Coefficients): 수출금액은 약 0.026, M2 금액은 약 0.00024. 이는 각 독립 변수가 종속 변수에 미치는 영향을 나타냅니다. 즉, 수출금액이 1단위 증가하면 코스피 지수는 약 0.026 증가하고, M2 금액이 1단위 증가하면 코스피 지수는 약 0.00024 증가합니다.

절편 (Intercept): 약 350.70.

이제 라쏘 회귀 모델을 이용한 예측값과 실제값의 차이를 표준편차 단위로 나타내는 차트를 만들어 보겠습니다. 이를 위해 전체 데이터셋에 대한 예측값을 먼저 계산하겠습니다.

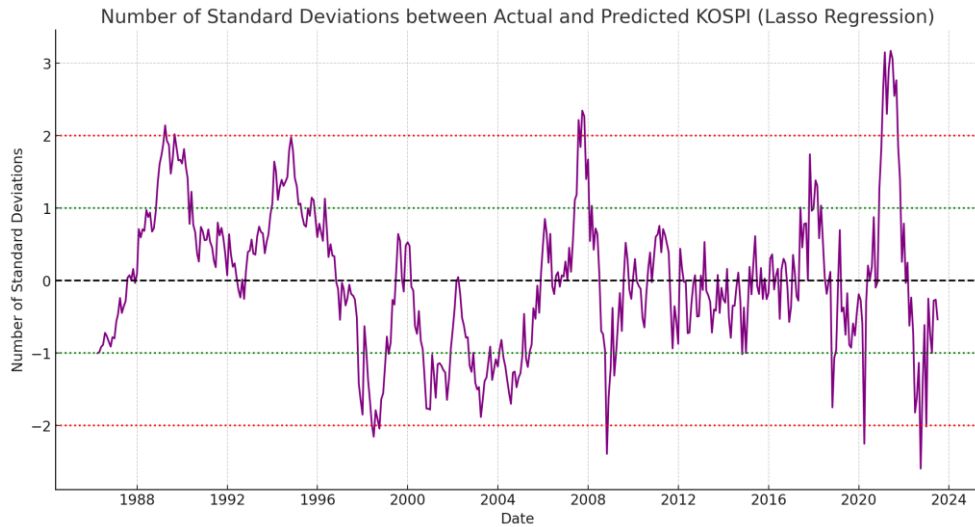
Code Interpreter 사용 사례: 코스피 지수 전망

수출, M2 지표로 코스피 추정 모델 만들기: 모델 개발 확장

Chat GPT Q&A

Finished working

Show work ▾



위 차트는 라쏘 회귀 모델을 사용하여 예측한 코스피 지수와 실제 코스피 지수 사이의 차이를 표준편차 단위로 나타낸 것입니다. 보라색 선은 실제 코스피 지수와 예상 코스피 지수의 차이를 표준편차로 나타냅니다.

다음으로, ARIMA 모델을 사용하여 코스피 지수를 예측해 보겠습니다. ARIMA는 AutoRegressive Integrated Moving Average의 약자로, 시계열 데이터를 분석하고 예측하는 데 널리 사용되는 통계 모델입니다.

ARIMA 모델을 훈련시키기 전에, 적절한 파라미터를 결정해야 합니다. 이를 위해 ACF(Autocorrelation Function)와 PACF(Partial Autocorrelation Function)를 통해 시계열 데이터의 패턴을 분석하겠습니다. 이후 ARIMA 모델을 훈련시키고, 예측값과 실제값의 차이를 표준편차 단위로 나타내는 차트를 만들어 보겠습니다.

하지만, ARIMA 모델은 복잡한 계산을 필요로 하며, 계산 리소스와 시간이 많이 소요됩니다. 또한, 이 모델은 선형 회귀와는 다르게 전체 시계열 데이터를 사용하여 훈련시키므로, 훈련 세트와 테스트 세트로 데이터를 분할하지 않습니다.

이에 동의하시면 ARIMA 모델을 훈련시키기 시작하겠습니다.

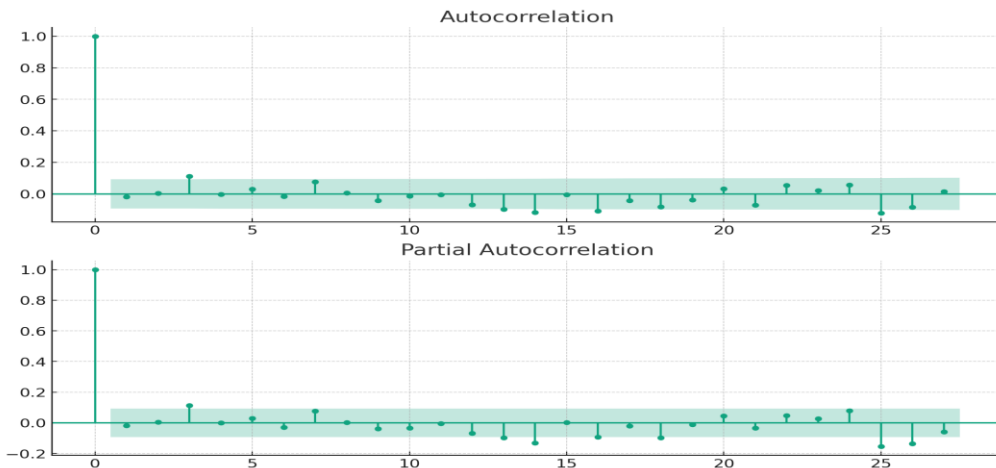
Code Interpreter 사용 사례: 코스피 지수 전망

수출, M2 지표로 코스피 추정 모델 만들기: 모델 개발 확장

Chat GPT Q&A

Finished working

Show work



위 그래프는 ACF(Autocorrelation Function)와 PACF(Partial Autocorrelation Function)를 나타냅니다. 이는 시계열 데이터의 자기상관을 보여주며, 이를 통해 ARIMA 모델의 파라미터를 결정하는 데 도움이 됩니다.

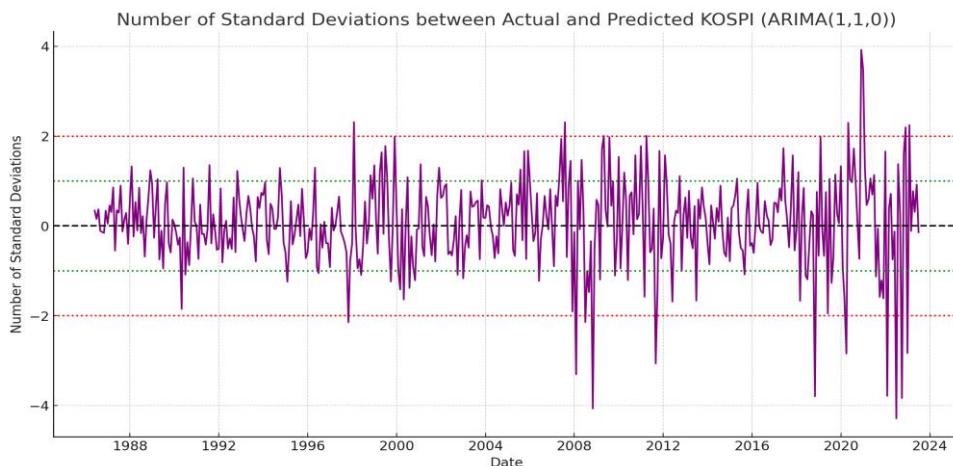
ACF 그래프에서는 자기상관이 점차적으로 감소하고, PACF 그래프에서는 1시점 이후에 급격히 감소하는 것을 볼 수 있습니다. 이는 AR(1) 모델을 사용하는 것이 적절하다는 신호입니다.

따라서 ARIMA(1,1,0) 모델을 사용하여 코스피 지수를 예측해 보겠습니다. 이는 AR(1) 모델을 사용하고, 차분 1번을 적용한 것을 의미합니다. MA 파트는 사용하지 않습니다.

이제 이 파라미터를 사용하여 ARIMA 모델을 훈련시키고, 예측값과 실제값의 차이를 표준편차 단위로 나타내는 차트를 만들어 보겠습니다.

Finished working

Show work



자료: Chat GPT, IBK투자증권

Code Interpreter 사용 사례: 코스피 지수 전망

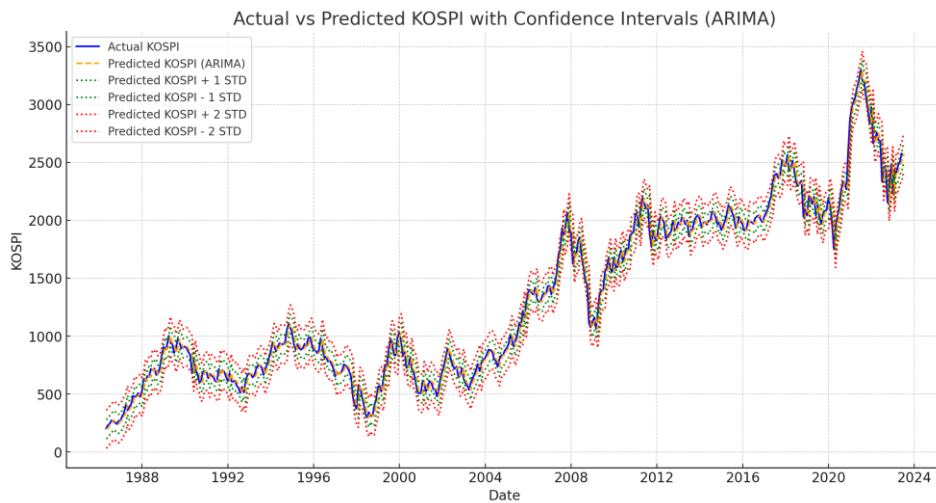
수출, M2 지표로 코스피 추정 모델 만들기: 모델 개발 확장

Chat GPT Q&A

Q : ariama 모델에서 추정한 코스피와 실제 코스피, 1std, 2std도 표기해서 차트로

A :

Finished working Show work



위 차트는 실제 코스피 지수(파란색), ARIMA 모델을 사용하여 예상한 코스피 지수(주황색), 그리고 예상 코스피 지수의 1표준편차와 2표준편차 범위(녹색과 빨간색 점선)를 나타냅니다.

이 차트를 통해 예상 코스피 지수의 변동 범위를 살펴볼 수 있습니다. 녹색 점선은 예상 코스피 지수의 1표준편차 범위를 나타내며, 빨간색 점선은 2표준편차 범위를 나타냅니다. 이를 통해 실제 코스피 지수가 예상 코스피 지수를 중심으로 어느 정도 범위 내에서 변동하는지를 알 수 있습니다. 이는 모델의 예측이 얼마나 정확한지, 그리고 어느 정도의 불확실성을 가지는지를 나타내는 좋은 지표가 될 수 있습니다.

다만, 이 차트를 통해 실제 투자 결정을 내리는 것은 매우 주의가 필요합니다. 왜냐하면 이는 간단한 ARIMA 모델을 기반으로 하고 있으며, 실제 코스피 지수는 수출금액과 M2 금액 외에도 많은 다른 요인들에 의해 영향을 받기 때문입니다.

Code Interpreter 사용 사례: 코스피 지수 전망


수출, M2 지표로 코스피 추정 모델 만들기: 머신러닝 모델 개발

Chat GPT Q&A

Q : 다시 파일을 넣어주겠다.
데이터프레임을 새로 생성해라

코스피 지수(C열)와 경제지표 수출금액(A열), M2 금액(B열)

xgboost로 코스피지수를 예측하는 모델을 만들어보자

 코스피 추정 csv.xlsx

A : XGBoost는 Gradient Boosting 알고리즘을 최적화한 알고리즘으로, 트리 기반의 앙상블 학습에서 가장 강력하고 널리 사용되는 방법 중 하나입니다. XGBoost는 높은 예측 정확도, 빠른 실행 속도, 그리고 유연성을 제공합니다.

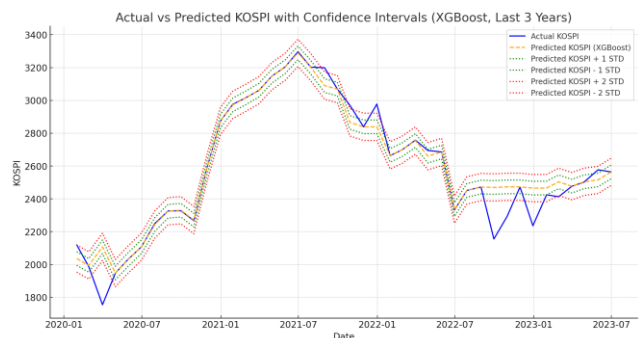
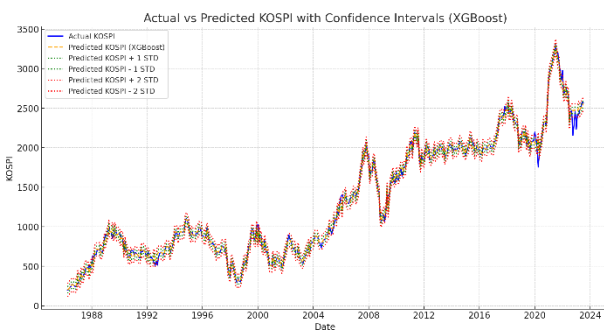
Finished working Show work ▾

XGBoost 모델의 결과는 다음과 같습니다:

Mean Squared Error (MSE): 약 8472.46. 이는 모델의 예측값과 실제값 사이의 평균 제곱 오차입니다. 이 값이 작을수록 모델의 성능이 좋다고 판단할 수 있습니다.

R-squared (R^2): 약 0.985. 이는 모델이 데이터의 약 98.5%를 설명하고 있다는 것을 의미합니다. 이 값이 1에 가까울수록 모델의 성능이 좋다고 판단할 수 있습니다.

이제 이 모델을 이용하여 예측한 코스피 지수와 실제 코스피 지수 사이의 차이를 표준편차 단위로 나타내는 차트를 만들어 보겠습니다. 이를 위해 전체 데이터셋에 대한 예측값을 먼저 계산하겠습니다.



자료: Chat GPT, IBK투자증권

Code Interpreter 사용 사례: 섹터 쏠림 지수 개발

섹터 쏠림 지수 만들기 및 과거 데이터 테스트

Chat GPT Q&A

Q : 해당 파일은 주식 섹터별 수익률이다.
라벨링이한글이니, 처리하도록해라

A :

Finished working Show work ▾

파일을 성공적으로 불러왔습니다. 이 데이터는 2009년부터 2023년까지의 각 월별로 주식 섹터별 수익률을 나타내고 있습니다. 첫 번째 열은 날짜를 나타내는 것으로 보이며, 나머지 열들은 각각 다른 주식 섹터의 수익률을 나타내고 있습니다. 데이터의 전처리가 필요한 부분이 있는지, 어떤 분석이나 작업을 진행하고 싶으신지 알려주시면 좋겠습니다.

Q : 특정 섹터 수익률에 쏠림이 일어나는지 지표를 만들고싶다. 방법이 있을까 ?

A :

특정 섹터에 수익률이 쏠리는 현상을 측정하려면 섹터별 수익률의 분산, 표준편차, 혹은 불균형 지수를 이용하는 방법이 있습니다.

분산(Variance) 또는 표준편차(Standard Deviation): 이 지표들은 섹터별 수익률의 퍼진 정도를 나타냅니다. 값이 높을수록 수익률이 많이 퍼져 있음을 의미합니다. 만약 특정 섹터에 수익률이 집중되면, 나머지 섹터의 수익률은 평균에서 멀리 떨어져 있을 것이므로 이 값은 커질 것입니다.

불균형 지수(Imbalance Index): 이 지표는 섹터별 수익률의 불균형을 측정합니다. 각 섹터의 수익률을 총 수익률로 나눈 후, 이 값의 제곱을 모두 더합니다. 값이 1에 가까울수록 특정 섹터에 수익률이 집중되어 있음을 의미합니다.

어떤 방법을 사용하실지 선택하신 후, 해당 방법으로 분석을 진행하겠습니다.

Q : 불균형지수로 만들고 차트를만들자

A :

Finished working Show work ▾

각 월별로 계산한 불균형 지수를 성공적으로 얻었습니다.

이 값이 1에 가까울수록 그 월의 섹터별 수익률이 한 섹터에 쏠려 있다는 의미입니다. 반대로 이 값이 1에서 멀어질수록 섹터별 수익률이 여러 섹터에 골고루 분포되어 있다는 것을 나타냅니다.

이제 이 불균형 지수를 시각화하거나, 더 분석하는 등의 작업을 진행할 수 있습니다. 어떤 작업을 원하시나요?

자료: Chat GPT, IBK투자증권

Code Interpreter 사용 사례: 섹터 쏠림 지수 개발

섹터 쏠림 지수 만들기 및 과거 데이터 테스트

Chat GPT Q&A

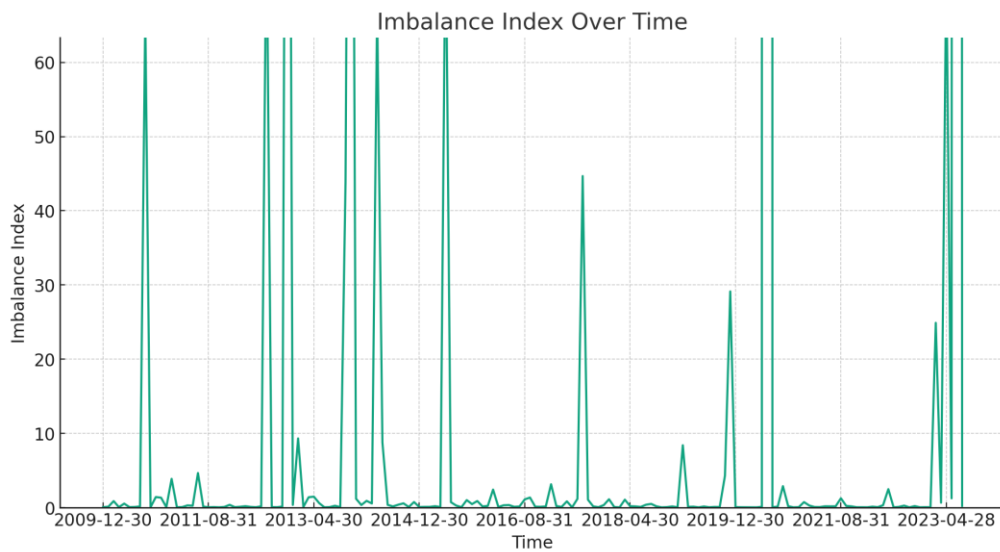
Q : 음 극단값때문에 차트가 잘 보이지 않는다.

정리 좀 해봐 (이전에 차트를 생성하였으나, Y축을 잘못 잡아 데이터가 눈에 들어오지 않았음)

A :

Finished working

Show work ▾



위 차트는 Y축의 최대값을 데이터의 95번째 백분위수로 설정한 후의 시간에 따른 주식 섹터별 수익률의 불균형 지수를 보여줍니다. 이 방법으로 이상치의 영향을 줄이면서 전체적인 트렌드를 보다 잘 파악할 수 있습니다.

불균형 지수가 높은 시점은 특정 섹터에 수익률이 집중되어 있는 시점을 나타내며, 불균형 지수가 낮은 시점은 수익률이 여러 섹터에 골고루 분포되어 있는 시점을 나타냅니다.

이 결과를 통해 주식 섹터별 수익률의 동향을 더욱 명확하게 파악할 수 있습니다. 추가로 필요한 분석이나 질문이 있으시면 알려주세요.

Code Interpreter 사용 사례: 섹터 쓸림 지수 개발

섹터 쓸림 지수 만들기 및 과거 데이터 테스트

Chat GPT Q&A

Q : 불균형지수가 20이상인 구간일 때 최고수익률을 기록한 섹터 3개의 다음달 수익률, 최저수익률을 기록한 3개 섹터의 다음달 수익률을 다운받게 CSV파일로 만들어라

양식은 시간 / 섹터명 3개 / 수익률 3개

A :

Finished working Show work ▾

다음 표는 불균형 지수가 20 이상인 각 월에서 가장 높은 수익률을 보인 섹터 3개의 다음 달 수익률을 보여줍니다. 각 월별로 섹터명과 그 섹터의 다음 달 수익률을 확인할 수 있습니다. NaN 값은 해당 월에 섹터가 최고 수익률 3개 안에 들지 않았음을 나타냅니다.

이 표를 통해 불균형 지수가 높았던 월에 수익률이 가장 높았던 섹터들이 다음 달에 어떤 수익률을 보였는지를 확인할 수 있습니다. 이 정보는 섹터별로 어떤 패턴이 있는지, 또는 불균형 지수가 높을 때 어떤 섹터가 좋은 수익률을 보이는지 등의 정보를 제공합니다.

결과를 CSV 파일로 저장했습니다. 다음 링크를 통해 파일을 다운로드 받을 수 있습니다:

[result.csv](#) (* 다운로드 가능)

추가로 필요한 분석이나 질문이 있으시면 알려주세요.

Q : 최종적으로 코드를 만들자

Imbalance 생성하고 차트화시키기

Imbalance 전체 추이를 CSV로 다운로드

Imbalance 지표가 20이상일때 최고수익률을 기록한 3개섹터의 다음달 수익률 계산

Imbalance 지표가 20이상일때 최저수익률을 기록한 3개섹터의 다음달 수익률을 계산

최고수익률 3개, 최적수익률 3개 섹터의 이름, 다음달 수익률을 기록하는 CSV 다운로드

두개의 csv 파일을 다운 받는 코드를 만들어라

참고로 코랩 기준으로 데이터 импорт, 다운로드한다

Code Interpreter 사용 사례: 섹터 쏠림 지수 개발

섹터 쏠림 지수 만들기 및 과거 데이터 테스트

Chat GPT Q&A

A :

Google Colab에서는 다음과 같이 코드를 작성하면 됩니다:

Finished working Show work ▾

> Show work 버튼 클릭시 코드 제시 사항

```
import pandas as pd

# 파일 업로드
from google.colab import files
uploaded = files.upload()

# 데이터 로드
import io
data = pd.read_csv(io.BytesIO(uploaded['수익률.csv']), encoding='cp949')
data.set_index('Unnamed: 0', inplace=True)

# 불균형 지수 계산
imbalance_indices = data.apply(lambda x: (x / x.sum())**2).sum(axis=1)

# 불균형 지수 차트화
import matplotlib.pyplot as plt
plt.plot(imbalance_indices)
plt.show()

# 불균형 지수 전체 추이 CSV 다운로드
imbalance_indices.to_csv('imbalance_indices.csv')
files.download('imbalance_indices.csv')

# 불균형 지수가 20 이상일 때 최고 수익률/최저 수익률을 기록한 3개 섹터의 다음 달 수익률 계산
high_imbalance_months = imbalance_indices[imbalance_indices >= 20].index
top3_sectors = data.loc[high_imbalance_months].apply(lambda x: x.nlargest(3).index.tolist(), axis=1)
bottom3_sectors = data.loc[high_imbalance_months].apply(lambda x: x.nsmallest(3).index.tolist(), axis=1)

# 최고 수익률 3개, 최저 수익률 3개 섹터의 이름, 다음달 수익률을 기록하는 CSV 다운로드
next_month_returns_top = pd.DataFrame(index=high_imbalance_months[:-1])
next_month_returns_bottom = pd.DataFrame(index=high_imbalance_months[:-1])
for month, sectors in top3_sectors.items():
    for sector in sectors:
        if sector not in next_month_returns_top.columns:
            next_month_returns_top[sector] = None
        next_month_returns_top.loc[month, sector] = data.loc[month, sector]
next_month_returns_top.to_csv('next_month_returns_top.csv')
files.download('next_month_returns_top.csv')

for month, sectors in bottom3_sectors.items():
    for sector in sectors:
        if sector not in next_month_returns_bottom.columns:
            next_month_returns_bottom[sector] = None
        next_month_returns_bottom.loc[month, sector] = data.loc[month, sector]
next_month_returns_bottom.to_csv('next_month_returns_bottom.csv')
files.download('next_month_returns_bottom.csv')
```

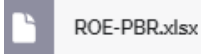
자료: Chat GPT, IBK투자증권

Code Interpreter 사용 사례: 적정 PBR 계산

코스피 적정 PBR 계산 모델 (PBR-ROE 기반)

Chat GPT Q&A

Q : B열에 ROE를 활용해 A열 PBR을 전망하는 모델을 3차방정식으로 만들어 보자



A :

Finished working Show work

데이터에는 다음 열이 포함됩니다.

날짜가 포함된 것으로 보이는 이름 없는 열입니다.

PBR: PBR(Price-to-Book Ratio).

ROE: 자기 자본 수익률.

외국: 외국인 투자 또는 보유를 나타낼 수 있습니다.

기관: 기관 투자 또는 보유를 나타낼 수 있습니다.

개인: 개인 투자 또는 보유를 나타낼 수 있습니다.

ROE를 사용하여 PBR을 예측하는 3차 다항식 회귀 모델을 만드는 데 관심이 있으므로 지금은 PBR 및 ROE 열에만 집중해야 합니다.

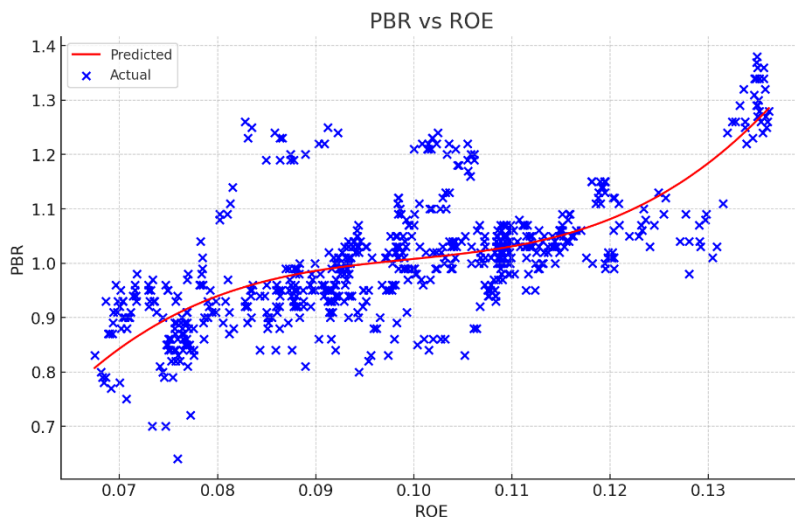
Finished working Show work

3차 다항식 회귀 모델의 계수와 절편은 다음과 같습니다.

회귀 방정식은 다음과 같이 제공됩니다.

$$PBR = -3.41 + 129.57 X ROE - 1283.62 X ROE^2 + 4299.25 X ROE^3$$

Finished working Show work



자료: Chat GPT, IBK투자증권

Code Interpreter 사용 사례: 적정 PBR 계산

코스피 적정 PBR 계산 모델 (PBR-ROE 기반)

Chat GPT Q&A

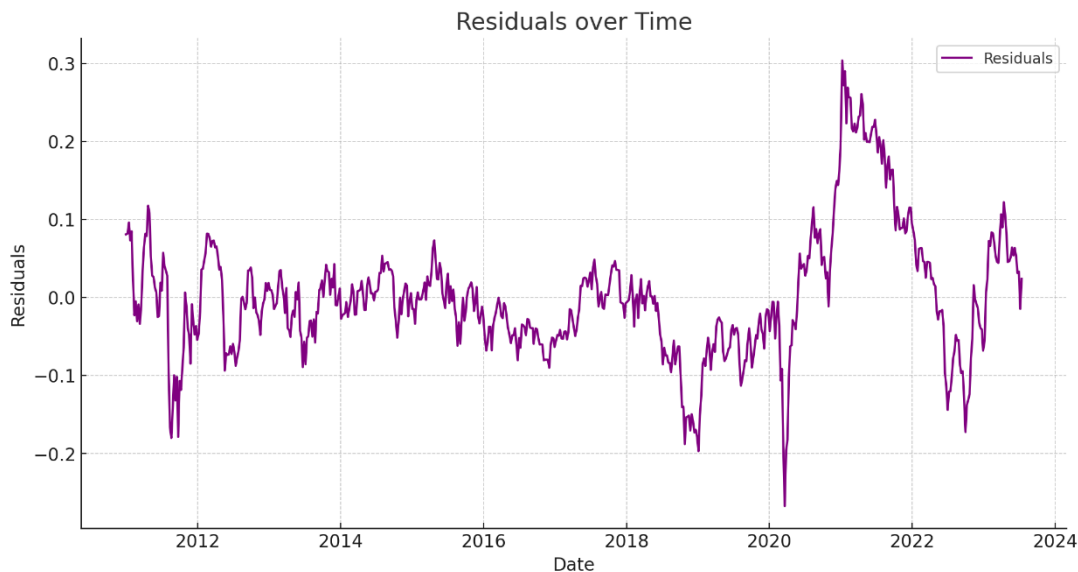
다음은 ROE를 기반으로 한 실제 및 예측 PBR 값의 도표입니다. 파란색 점은 실제 데이터를 나타내고 빨간색 선은 3차 다항식 회귀 모델의 예측을 나타냅니다.

보시다시피 모델은 데이터의 전반적인 추세를 포착하지만 완벽하게 맞지 않는 일부 영역이 있습니다. 이것은 특히 다항식 회귀와 같은 비교적 간단한 모델을 사용할 때 실제 데이터에서 예상됩니다.

Q : 예상치와 실제치의 차이를 차트로 보여줘 시계열 데이터로

Finished working

Show work ▾



다음은 실제 PBR과 모델의 예측 PBR 간의 차이를 나타내는 잔차의 시계열 도표입니다.

잔차는 모델이 캡처하지 않는 데이터 부분을 이해하는 데 유용합니다. 큰 잔차는 모델의 예측이 실제 값에서 멀리 떨어진 영역을 나타냅니다. 이 경우 잔차는 시간이 지남에 따라 변동하며, 이는 모델 예측이 상당히 정확한 기간과 그렇지 않은 기간이 있음을 의미합니다.

Code Interpreter 사용 사례: 적정 PBR 계산

코스피 적정 PBR 계산 모델 (PBR-ROE 기반)

Chat GPT Q&A

Q : 잔차의 추이와 아까 제공한 데이터 중 외국인, 개인, 기관 수급과 연관성이 있는지 체크해달라

A :

Finished working Show work ▾

	Residuals	Foreign	Institutional	Individual
Residuals	1.000000	-0.337309	0.072405	0.203986
Foreign	-0.337309	1.000000	-0.183818	-0.556922
Institutional	0.072405	-0.183818	1.000000	-0.700116
Individual	0.203986	-0.556922	-0.700116	1.000000

이 행렬에서 다음을 확인할 수 있습니다.

잔차와 '외래' 열 간에 중간 정도의 음의 상관관계(-0.34)가 있습니다. 이는 외국인 보유액이 증가하면 잔존액이 감소하는 경향이 있고 그 반대도 마찬가지임을 의미합니다.

잔차와 '기관' 열 사이에는 약한 양의 상관관계(0.07)가 있습니다. 이것은 기관 보유가 증가함에 따라 잔차가 증가하는 약간의 경향이 있음을 의미합니다.

잔차와 '개별' 열 사이에는 약한 양의 상관관계(0.20)가 있습니다. 이것은 개별 보유량이 증가함에 따라 잔차가 증가하는 약간의 경향이 있음을 의미합니다.

상관관계가 인과관계를 의미하지 않는다는 점에 유의하십시오. 이는 통계적 관계일 뿐이며 인과 관계를 나타내지 않을 수 있습니다. 인과 관계를 탐색하려면 추가 연구 또는 실험을 수행해야 할 수 있습니다.

자료: Chat GPT, IBK투자증권

Code Interpreter 사용 사례: 걱정 PBR 계산

코스피 걱정 PBR 계산 모델 (PBR-ROE 기반)

Chat GPT Q&A

Q : Residuals을 foreign institutional individual으로 추정하는 다량 회귀 분석으로 추정하는 모델의 오차를 체크하고 실제치와 예상치의 차트를 보여달라

A :

Finished working Show work ▾

다음은 '외부', '기관' 및 '개별' 열을 사용하여 잔차를 예측하는 선형 회귀 모델에 대한 오류 메트릭입니다.

평균 제곱 오차(MSE): 0.00447

평균 제곱근 오차(RMSE): 0.06684

평균 절대 오차(MAE): 0.04842

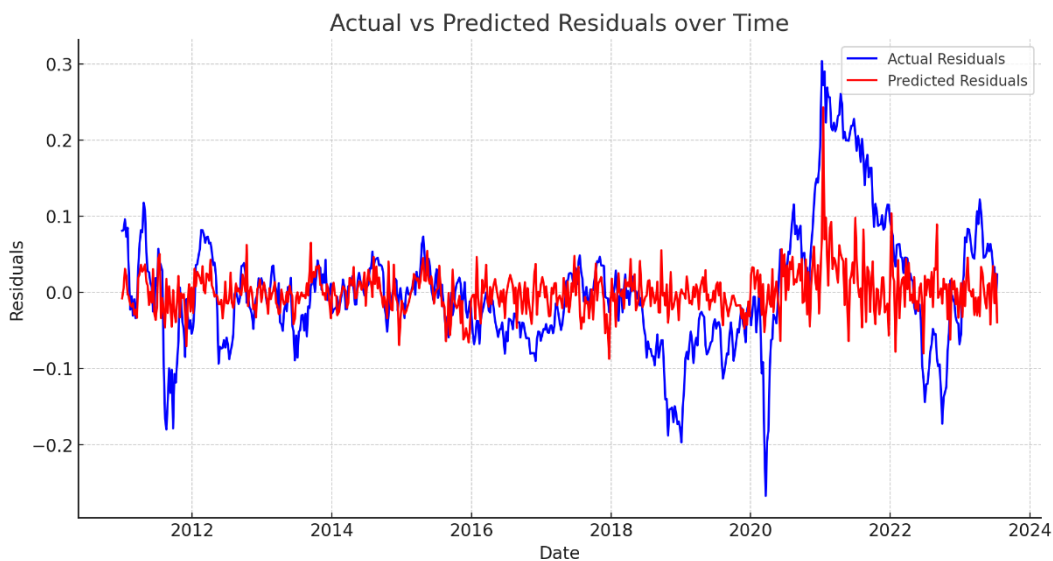
이러한 메트릭은 예측 잔차와 실제 잔차 간의 차이를 정량화합니다. 값이 낮을수록 모델이 데이터에 더 적합함을 나타냅니다.

MSE는 예측 값과 실제 값 사이의 제곱 차이의 평균을 취합니다. 이러한 오류는 평균화되기 전에 제곱되기 때문에 큰 오류는 크게 벌점 처리됩니다.

RMSE는 MSE의 제곱근입니다. 대상 변수와 단위가 같기 때문에 MSE보다 해석하기 쉽습니다.

MAE는 예측 값과 실제 값 간의 절대 차이의 평균을 취합니다. MSE 및 RMSE만큼 큰 오류에 페널티를 주지 않으므로 경우에 따라 보다 현실적인 오류 측정이 될 수 있습니다.

Finished working Show work ▾



자료: Chat GPT, IBK투자증권


Code Interpreter 사용 사례: 어닝서프라이즈 예측

어닝서프라이즈 예측 모델 만들기 (재무, 시장 데이터 활용 머신러닝)

Chat GPT Q&A

Q : 해당 파일은 삼성전자의 영업이익 서프라이즈 비율과 전기 재무제표, 실적발표 전 참고할 수 있는 데이터를 정리한 것이다.

OP Surprise를 같은 행의 데이터로부터 예측하는 모델을 랜덤포레스트 모델로 만들어보자
테스트 데이터로 실제치와 모델 예상치를 보여주고 서프라이즈 예상에 가장 중요한 피쳐 5개를 보여줘

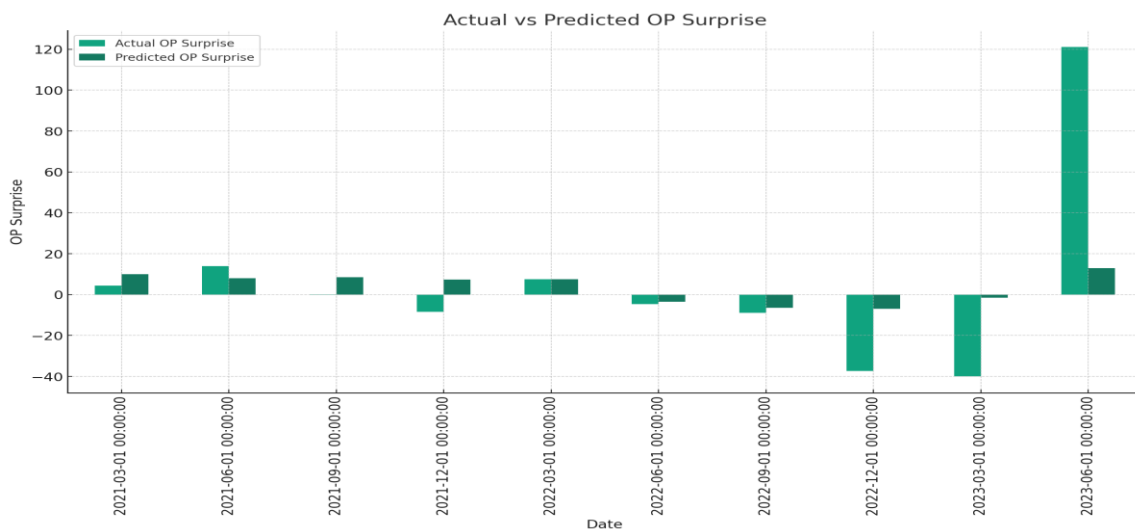
 서프라이즈전망.xlsx

A :

Finished working Show work ▾

가장 중요한 피쳐 5개는 다음과 같습니다:

- OP_Chg (실적발표 전일 컨센서스 변화율)
- REV_Chg (실적발표 전일 컨센서스 변화율)
- Stock_Return1M (실적발표 전일 기준 1개월 수익률)
- OP_Ratio (실적발표 전분기 영업이익률)
- Stock_Return1w (실적발표 전일 기준 1주일 수익률)



위의 바 차트는 테스트 데이터셋에 대한 'OP Surprise'의 실제 값과 예측된 값을 분기별로 보여줍니다. 이 차트는 모델이 어떻게 시간에 따라 실제 데이터를 추적하고 있는지를 시각적으로 보여줍니다.

각 분기에 대해 실제 값과 예측 값이 얼마나 가까운지를 비교하면 모델의 성능을 파악할 수 있습니다. 이 모델은 완벽하지 않으며 일부 오차가 있습니다. 모델의 성능을 개선하려면 다른 모델을 시도하거나, 추가적인 피쳐 엔지니어링을 수행하거나, 하이퍼파라미터 튜닝을 수행하는 등의 방법을 사용할 수 있습니다.

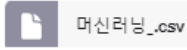
피드백이나 추가 질문이 있으시면 언제든지 알려주세요!

Code Interpreter 사용 사례: 팩터 머신러닝 모델

팩터 기반 모델: 대량의 팩터, 매크로, 시장 수익률 제시 후 데이터분석

Chat GPT Q&A

Q : 주어진 데이터로 A열의 다음달(시계열 상 다음 행 데이터) 코스피를 예측할 수 있도록 데이터를 전처리 하고 머신러닝을 통해 코스피 상승/하락 여부를 예측하는 모델을 만들고 예측, 실제치를 차트로 보여줘



A :

네, 알겠습니다. 이번에는 모델을 학습시킬 때 이전 달의 데이터를 사용하여 다음 달의 "코스피수익률"을 예측하도록 하겠습니다.

이를 위해서는 먼저 데이터를 약간 다르게 구조화해야 합니다. 즉, 각 행에서 "코스피수익률"은 다음 달의 수익률을 나타내도록 하고, 나머지 변수들은 이전 달의 값들을 나타내게 해야 합니다.

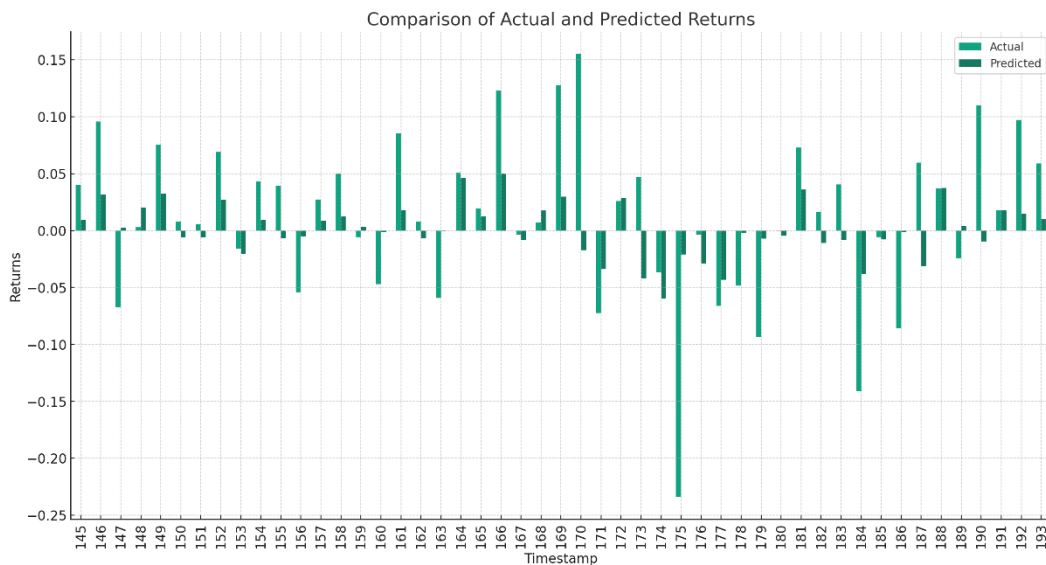
먼저 이렇게 데이터를 재구조화하고, 그 다음에 랜덤 포레스트 모델을 다시 학습시켜 보겠습니다.

Finished working Show work

이제 다음 달의 "코스피수익률"을 예측하기 위해 랜덤 포레스트 모델을 다시 학습시켰습니다. 다시 한번, 테스트 데이터셋을 사용하여 모델의 성능을 평가해 보겠습니다. 평균 제곱 오차(MSE)를 사용하여 성능을 평가하겠습니다.

Finished working Show work

테스트 데이터셋에 대한 예측의 평균 제곱 오차(MSE)는 약 0.004입니다. 예측된 수익률의 부호(양수 또는 음수)가 실제 수익률의 부호와 약 71%의 경우 일치합니다. 이제 예측된 수익률과 실제 수익률을 비교하는 차트를 다시 그려 보겠습니다.



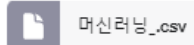
자료: Chat GPT, IBK투자증권

Code Interpreter 사용 사례: 팩터 머신러닝 모델

팩터 기반 모델: 팩터 수익률 추정 모델 개발

Chat GPT Q&A

Q : 주어진 데이터로 Value LS_T+1, Mometum LS_T+1, Size LS_T+1, qual LS_T+1, low LS_T+1, DIV LS_T+1 피처를 각각 전망하는 머신러닝 모델을 만들어라
단 전망을 위해 Value LS_T+1, Mometum LS_T+1, Size LS_T+1, qual LS_T+1, low LS_T+1, DIV LS_T+1 피처는 사용하지마라



A :

네, 알겠습니다. 주어진 피처를 사용하여 각각의 예상 수익률 (Value LS_T+1, Mometum LS_T+1, Size LS_T+1, qual LS_T+1, low LS_T+1, DIV LS_T+1)을 예측하는 머신러닝 모델을 만들려고 합니다.

먼저, 제공된 CSV 파일을 로드하고 데이터를 살펴보겠습니다. 이를 통해 어떤 전처리가 필요한지, 그리고 모든 피처가 올바르게 포함되어 있는지 확인하겠습니다.
먼저 이렇게 데이터를 재구조화하고, 그 다음에 랜덤 포레스트 모델을 다시 학습시켜 보겠습니다.

Finished working Show work

학습을 진행하기 전에, 모델 훈련에 사용할 파라미터를 설정해야 합니다. 일반적으로 설정하는 파라미터는 다음과 같습니다:

n_estimators: 랜덤 포레스트에서 사용할 트리의 수

max_depth: 트리의 최대 깊이

min_samples_split: 노드를 분할하기 위한 최소 샘플 수

min_samples_leaf: 리프 노드에 필요한 최소 샘플 수

max_features: 각 분할에 고려할 피처의 최대 수

이들 파라미터는 모델의 성능과 훈련 시간에 큰 영향을 미칩니다. 따라서, 우리는 이들을 최적화해야 합니다. 하지만 우선 기본값을 사용하여 모델을 훈련시켜 보겠습니다.

각 목표 변수에 대해 랜덤 포레스트 모델을 훈련시켜 봅시다. 이 작업은 시간이 다소 걸릴 수 있습니다.

Finished working Show work

각 목표 변수에 대한 모델의 성능은 다음과 같습니다:

Value LS_T+1: RMSE = 0.072

Mometum LS_T+1: RMSE = 0.058

Size LS_T+1: RMSE = 0.034

qual LS_T+1: RMSE = 0.030

low LS_T+1: RMSE = 0.062

DIV LS_T+1: RMSE = 0.051

이 결과는 각 모델이 테스트 데이터에 대해 얼마나 잘 일반화되었는지를 보여줍니다. RMSE 값이 작을수록 모델이 데이터를 더 정확하게 예측하였다는 것을 의미합니다.

자료: Chat GPT, IBK투자증권

Code Interpreter 사용 사례: 팩터 머신러닝 모델

팩터 기반 모델: 팩터 수익률 추정 모델 개발

Chat GPT Q&A

Q : 최적화해보자

A :

좋습니다, 모델 파라미터를 최적화하기 위해 Grid Search 방법을 사용하겠습니다. Grid Search는 모든 파라미터 조합을 시도하여 최적의 조합을 찾는 방법입니다.

랜덤 포레스트에 대해 최적화할 주요 파라미터는 다음과 같습니다:

n_estimators: 랜덤 포레스트에서 사용할 트리의 수

max_depth: 트리의 최대 깊이

min_samples_split: 노드를 분할하기 위한 최소 샘플 수

min_samples_leaf: 리프 노드에 필요한 최소 샘플 수

Grid Search를 수행하기 전에, 각 파라미터에 대해 시도할 값의 범위를 정의해야 합니다. 이는 주로 경험적으로 결정됩니다. 이번 경우에는 다음과 같이 설정하겠습니다:

n_estimators: [50, 100, 200]

max_depth: [None, 10, 20]

min_samples_split: [2, 5, 10]

min_samples_leaf: [1, 2, 4]

위 범위에서 파라미터를 선택하여 Grid Search를 수행하겠습니다. 하지만, 이 작업은 계산량이 많아 시간이 오래 걸릴 수 있습니다. 또한, 모든 목표 변수에 대해 각각 Grid Search를 수행하므로 더욱 오래 걸릴 수 있습니다.

그럼 시작하겠습니다.

Finished working Show work ▾

Value LS_T+1: RMSE = 0.046, 파라미터 = {'bootstrap': True, 'ccp_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 'auto', 'max_leaf_nodes': None, 'max_samples': None, 'min_impurity_decrease': 0.0, 'min_samples_leaf': 2, 'min_samples_split': 2, 'min_weight_fraction_leaf': 0.0, 'n_estimators': 100, 'n_jobs': None, 'oob_score': False, 'random_state': 42, 'verbose': 0, 'warm_start': False}

... (이하 생략)

이제 이러한 모델을 사용하여 미래 데이터에 대한 예상 수익률을 예측할 수 있습니다. 또한, 이 모델들을 사용하여 피처 중요도를 분석할 수도 있습니다. 이는 어떤 피처가 예상 수익률에 가장 큰 영향을 미치는지 이해하는 데 도움이 됩니다.

Code Interpreter 사용 사례: 팩터 머신러닝 모델

팩터 기반 모델: 팩터 수익률 추정 모델 개발

Chat GPT Q&A

Q : 좋다 이렇게 만든 모델로 예측한 수치와 실제수치를 모두 CSV로 보여줄 수 있나 ?

A :

네, 그렇게 할 수 있습니다. 각 모델을 사용하여 테스트 세트에 대한 예측을 수행하고, 이를 실제 값과 함께 CSV 파일로 저장하겠습니다.

각 목표 변수에 대해 다음과 같이 수행하겠습니다:

모델로 테스트 세트에 대한 예측을 수행합니다.
예측 값을 실제 값과 함께 DataFrame에 저장합니다.
이 DataFrame을 CSV 파일로 저장합니다.
아래에 이를 수행하는 코드를 작성하겠습니다.

Finished working Show work ▾

예측 결과를 CSV 파일로 저장하였습니다. 각 목표 변수에 대한 예측 값과 실제 값을 포함하고 있습니다.

CSV 파일을 다운로드하려면 아래 링크를 클릭하세요:

[predictions.csv](#)

4

GPT4 Code interpreter 사용 Tip

GPT 업로드 엑셀 파일 정리 Tip

엑셀 파일 올릴 때

유의사항

- GPT-4 Code interpreter는 데이터를 읽기 위해 파이썬 라이브러리 판다스를 활용, 판다스가 데이터를 읽을 수 있도록 데이터 정리가 중요
- 물론 데이터를 어떠한 형식으로 제시하더라도 나름대로 GPT Code Interpreter가 데이터를 전처리 실행하여 사용자가 원하는 분석을 실행함
- 그러나 전처리과정에서 데이터 길이를 맞추거나, n/a 값을 제거하기 위해 행을 통째로 삭제하는 등 사용자가 원하지 않는 명령을 실행할 수 있음
- 제일 좋은 방법은 GPT가 데이터 전처리를 실행하지 않을 정도로 깔끔하게 GPT에게 데이터를 제공하는 것

업로드 파일 주의 사항

	A	B	C	D	E
1	Refresh	Last Update : 2023-07-12 10:07:46			
2					
3	Time Series (Company)				
4	Frequency	W	Ascending	0	
5	Period(From)	CPD-1Y [202207	Korean	0	
6	Period(To)	CPD [20230712]		0	
7					
8	Code	A005930	A373220	A000660	A207940
9	Name	삼성전자	LG에너지솔루션	SK하이닉스	삼성바이오로직스
10	Item Code	S100100	S100100	S100100	S100100
11	Unit	Local/Shares	Local/Shares	Local/Shares	Local/Shares
12	Base Date				
13					
14	DATE	증가	증가	증가	증가
15	2022-07-15	60,000	400,000	98,700	817,000
16	2022-07-22	61,200	382,000	100,000	829,000
17	2022-07-29	61,400	422,000	97,900	865,000
18	2022-08-05	61,500	446,000	98,500	895,000
19	2022-08-12	60,200	460,500	93,300	910,000
20	2022-08-19	60,900	447,000	96,400	868,000
21	2022-08-26	60,000	465,500	95,100	847,000
22	2022-09-02	57,500	482,000	91,700	838,000
23	2022-09-08	55,600	486,500	90,400	810,000
24	2022-09-16	56,200	506,000	91,200	802,000
25	2022-09-23	54,500	460,500	83,500	770,000
26	2022-09-30	53,100	426,500	83,100	807,000
27	2022-10-07	56,200	482,500	91,200	811,000
28	2022-10-14	56,300	483,000	95,500	806,000
29	2022-10-21	55,900	499,000	90,500	847,000

	A	B	C	D	E
1	DATE	Samsung	LG Energy	Sk Hynix	Samsung
2	2022-07-15	60000	400000	98700	817000
3	2022-07-22	61300	382000	100000	829000
4	2022-07-29	61400	422000	97900	865000
5	2022-08-05	61500	446000	98500	895000
6	2022-08-12	60200	460500	93300	910000
7	2022-08-19	60900	447000	96400	868000
8	2022-08-26	60000	465500	95100	847000
9	2022-09-02	57500	482000	91700	838000
10	2022-09-08	55600	486500	90400	810000
11	2022-09-16	56200	506000	91200	802000
12	2022-09-23	54500	460500	83500	770000
13	2022-09-30	53100	426500	83100	807000
14	2022-10-07	56200	482500	91200	811000
15	2022-10-14	56300	483000	95500	806000
16	2022-10-21	55900	499000	90500	847000
17	2022-10-28	57300	532000	83400	876000
18	2022-11-04	59400	592000	84500	871000

자료: IBK투자증권

GPT 업로드 엑셀 파일 정리 Tip

엑셀 파일 올릴 때

- 1) 숫자 양식 통일
- 2) 열 이름 모두 영어로 표기
- 3) 시계열데이터는 날짜로 저장

- GPT-4 Code interpreter 서버 상황이 좋지 않으면 데이터 전체 리를 반복해서 하는 경우가 발생
- 판다스가 읽을 수 있는 데이터를 생성
- 특히 같은 숫자라도 Type이 다를 수 있음
텍스트 또는 숫자(소수점, 정수 등)는 눈으로 보기에 유사해도 컴퓨터는 다르다고 판단, 양식 통일은 필수

업로드 파일 주의 사항

데이터에 모두 영어로 이름 (라벨링) 넣을 것
특히 날짜에도 Date 라고 라벨링 넣을 것

	A	B	C	D	E	F	G
1	D A T E	Samsung	LG Energy	Sk Hynix	Samsung	Samsung	LG Chemi
2	2022-07-15	60000	400000	98700	817000	548000	512000
3	2022-07-22	61300	382000	100000	829000	568000	561000
4	2022-07-29	61400	422000	97900	865000	569000	603000
5	2022-08-05	61500	446000	98500	895000	598000	642000
6	2022-08-12	60200	460500	93300	910000	623000	665000
7	2022-08-19	60900	447000	96400	868000	618000	640000
8	2022-08-26	60000	465500	95100	847000	586000	616000
9	2022-09-02	57500	482000	91700	838000	576000	610000
10	2022-09-08	55600	486500	90400	810000	571000	644000
11	2022-09-16	56200	506000	91200	802000	602000	635000
12	2022-09-23	54500	460500	83500	770000	610000	604000
13	2022-09-30	53100	426500	83100	807000	546000	536000
14	2022-10-07	56200	482500	91200	811000	591000	589000
15	2022-10-14	56300	483000	95500	806000	603000	573000
16	2022-10-21	55900	499000	90500	847000	627000	568000
17	2022-10-28	57300	532000	83400	876000	727000	615000
18	2022-11-04	59400	592000	84500	871000	733000	690000

Date 열은 날짜 양식으로 저장
Ex) 2023,2Q 비추천
2023-06-30을 추천

안에 들어가는 수치들은 숫자 양식으로 추천
, ,000 양식은 비추천

또한 n/a값이 없도록 모두 채워야 하며
숫자들의 양식은 모두 통일 추천

자료: IBK투자증권

GPT 업로드 엑셀 파일 정리 Tip

머신러닝 분석 대상
파일은 아래와 같이
정리해서 올리는 것을
추천

- 머신러닝은 T+1 데이터 예측을 위해 T 데이터를 사용
- 위 과정에서 데이터 전처리 실행
- 그러나 전처리로 인해 데이터 시계열 길이가 달라지는 문제 발생
경우에 따라서 시계열 불일치 문제로 계속 코딩을 수정하는 경우
발생, 가장 좋은 방법은 전처리가 필요하지 않는 데이터를 업로드

머신러닝 작업 시 업로드 파일 주의 사항

머신러닝을 돌리면 GPT는 예상치와 예상을 위한 데이터의 행을 맞추기 위해
데이터를 위로 한칸씩 올림 (자주 에러 발생 요인)
그것보다 미리 전망하고 싶은 데이터를 한칸씩 위로 올려서 업로드 하는 것을 추천

	A	B	C	D	E	F	G
1	DATE	Samsung	LG Energy	Sk Hynix	Samsung	Samsung	LG Chemi
2	2022-07-15	60000	400000	98700	817000	548000	512000
3	2022-07-22	61300	382000	100000	829000	568000	561000
4	2022-07-29	61400	422000	97900	865000	569000	603000
5	2022-08-05	61500	446000	98500	895000	598000	642000
6	2022-08-12	60200	460500	93300	910000	623000	665000
7	2022-08-19	60900	447000	96400	868000	618000	640000
8	2022-08-26	60000	465500	95100	847000	586000	616000
9	2022-09-02	57500	482000	91700	838000	576000	610000
10	2022-09-08	55600	486500	90400	810000	571000	644000
11	2022-09-16	56200	506000	91200	802000	602000	635000
12	2022-09-23	54500	460500	83500	770000	610000	604000
13	2022-09-30	53100	426500	83100	807000	546000	536000
14	2022-10-07	56200	482500	91200	811000	591000	589000
15	2022-10-14	56300	483000	95500	806000	603000	573000
16	2022-10-21	55900	499000	90500	847000	627000	568000
17	2022-10-28	57300	532000	83400	876000	727000	615000
18	2022-11-04		592000	84500	871000	733000	690000

전망 대상 데이터는 아직 데이터가 나오지 않았지만
전망을 위한 데이터들은 모두 나오면 위 그림처럼
비어 있는 칸이 나오면 → 실제 작업 시 시계열 길이가 달라 에러 자주 발생
그냥 제일 아래 칸은 삭제한 데이터를 제공 후 모델을 다 만들고, 모델 코드를 저장.
완성 모델에 변수를 입력하여 예측 값을 출력 받는 것이 더 좋음

자료: IBK투자증권

GPT 업로드 엑셀 파일 정리 Tip



머신러닝 작업 시 업로드 파일 주의 사항

	A	B	C	D	E	F	G
1	D A T E	Samsung	LG Energy	Sk Hynix	Samsung	Samsung	LG Chemi
2	2022-07-15	60000	400000	98700	817000	548000	512000
3	2022-07-22	61300	382000	100000	829000	568000	561000
4	2022-07-29	61400	422000	97900	865000	569000	603000
5	2022-08-05	61500	446000	98500	895000	598000	642000
6	2022-08-12	60200	460500	93300	910000	623000	665000
7	2022-08-19	60900	447000	96400	868000	618000	640000
8	2022-08-26	60000	465500	95100	847000	586000	616000
9	2022-09-02	57500	482000	91700	838000	576000	610000
10	2022-09-08	55600	486500	90400	810000	571000	644000
11	2022-09-16	56200	506000	91200	802000	602000	635000
12	2022-09-23	54500	460500	83500	770000	610000	604000
13	2022-09-30	53100	426500	83100	807000	546000	536000
14	2022-10-07	56200	482500	91200	811000	591000	589000
15	2022-10-14	56300	483000	95500	806000	603000	573000
16	2022-10-21	55900	499000	90500	847000	627000	568000
17	2022-10-28	57300	532000	83400	876000	727000	615000
18	2022-11-04		592000	84500	871000	733000	690000

	A	B	C	D	E	F	G
1	D A T E	Samsung	LG Energy	Sk Hynix	Samsung	Samsung	LG Chemi
2	2022-07-15	60000	400000	98700	817000	548000	512000
3	2022-07-22	61300	382000	100000	829000	568000	561000
4	2022-07-29	61400	422000	97900	865000	569000	603000
5	2022-08-05	61500	446000	98500	895000	598000	642000
6	2022-08-12	60200	460500	93300	910000	623000	665000
7	2022-08-19	60900	447000	96400	868000	618000	640000
8	2022-08-26	60000	465500	95100	847000	586000	616000
9	2022-09-02	57500	482000	91700	838000	576000	610000
10	2022-09-08	55600	486500	90400	810000	571000	644000
11	2022-09-16	56200	506000	91200	802000	602000	635000
12	2022-09-23	54500	460500	83500	770000	610000	604000
13	2022-09-30	53100	426500	83100	807000	546000	536000
14	2022-10-07	56200	482500	91200	811000	591000	589000
15	2022-10-14	56300	483000	95500	806000	603000	573000
16	2022-10-21	55900	499000	90500	847000	627000	568000
17	2022-10-28	57300	532000	83400	876000	727000	615000

자료: IBK투자증권

Chat-GPT 모델 개발 시 모델 선택 사항



Chat-GPT Prompt 및 의미		
번호	모델	사용 상황
1	다중회귀	예측 데이터와 변수 사이에 선형적인 관계가 있을 때 사용 추천 모델 개발 가능성보다는 개발 된 모델의 차트와 데이터를 보고 판단해도 됨 가장 기본적인 모델로 어떤 데이터에 사용해도 큰 문제 없음
2	Arima Sarima	Airma 모델은 자가회귀와 이동평균을 활용한 모델 계절적 특성을 반영하려면 Sariama 모델을 추천 시계열을 예측하는 모델로 금융 데이터에 사용 적합 주가, 경제지표, 기업 재무 데이터 분석에 사용 가능
3	머신러닝 1) 결정트리 2) 랜덤포레스트 3) 그래디언트 부스팅	비선형 관계를 처리하는데 효과적, 결정트리는 모델을 한 개 만들어서 과적합 가능성, 랜덤 포레스트는 결정트리의 단점을 해결하기 위해 다수의 트리를 만들어서 과최적화 단점을 해소, 랜덤포레스트를 추천 그래디언트 부스팅은 랜덤포레스트의 오차를 보완하기 위한 모델, 특정 데이터에 민감하게 반응하는 모델 개발, 과최적화 가능성
4	신경망	현재 GPT에서 사용 불가 차후 업데이트되면 가능할 것으로 예상 다만, 신경망 모델이 기존 모델보다 우수하다는 보장은 없음

Chat-GPT 모델 선택 질문: 어려우면 GPT에게 추천해달라고 하자

Finished working
Show work ▾

🗑️
👍
🔗

데이터를 성공적으로 훈련 및 테스트 세트로 분할했습니다. 이제 각 목표 변수에 대한 머신러닝 모델을 훈련시킬 수 있습니다. 이 작업에는 여러 가지 방법이 있습니다. 가장 간단한 방법 중 하나는 선형 회귀 모델을 사용하는 것입니다. 그러나 이 데이터에는 다른 모델이 더 적합할 수 있습니다.

어떤 유형의 모델을 사용하길 원하시나요? 아래는 몇 가지 옵션입니다:

1. 선형 회귀 (Linear Regression)
2. 결정 트리 (Decision Trees)
3. 랜덤 포레스트 (Random Forest)
4. 그래디언트 부스팅 (Gradient Boosting)
5. 서포트 벡터 머신 (Support Vector Machines)
6. 신경망 (Neural Networks)

또한, 시계열 데이터를 처리하는 데 특화된 모델들도 있습니다. 그러나 이들은 일반적으로 데이터에 대한 추가 가정을 요구하며, 이 데이터는 일반적인 머신러닝 방법을 사용하는 것이 더 적절할 수도 있습니다. 시계열 분석에 대해 더 자세히 알아보시겠습니까?

자료: IBK투자증권

MEMO



MEMO





IBKS Research Center

성명	직급	담당업종	전화	이메일
이승훈	이사	총괄/인터넷/게임	6915-5680	dozed@ibks.com

투자분석부

변준호	연구위원	Strategy	6915-5670	ymaezono@ibks.com
정용택	수석 Economist	Economy	6915-5701	ytjeong0815@ibks.com
김인식	연구위원	자산배분/ETF	6915-5472	kds4539@ibks.com
김종영	연구원	Quant	6915-5653	jujub0251@ibks.com
권기중	연구원	Fixed Income	6915-5654	kjoong@ibks.com
우지연	연구원	Global Strategy	6915-5663	dnwldus2170@ibks.com

기간산업분석부

김태현	연구위원	음식료	6915-5658	kith0923@ibks.com
김장원	연구위원	통신서비스/지주	6915-5661	jwkim@ibks.com
이상현	연구위원	자동차/기계	6915-5662	coolcat.auto@ibks.com
남성현	연구위원	유통	6915-5672	rockrole@ibks.com
우도형	연구원	증권	6915-5651	woo.dohyeong@ibks.com

ESG팀

이동욱	연구위원	화학/정유	6915-5671	treestump@ibks.com
-----	------	-------	-----------	--------------------

혁신기업분석부

이건재	연구위원	IT/소재/부품/장비	6915-5676	geonjaelee83@ibks.com
김운호	연구위원	IT/반도체/디스플레이	6915-5656	unokim88@ibks.com
이선경	연구위원	제약/바이오	6915-5655	seonkyoung.lee@ibks.com
채윤석	연구위원	스몰캡	6915-5474	Chaeyunseok87@ibks.com
이현욱	연구원	2차전지	6915-5659	hwle1125@ibks.com

“국민과 중소기업에 필요한 참 좋은 IBK투자증권”



서울특별시 영등포구 여의도동 국제금융로 6길 11
대표번호 02-6915-5000
고객만족팀 1588-0030, 1544-0050

영업부	02) 6915-2626	IBK WM센터 역삼	02) 556-4999
강남센터	02) 2051-5858	IBK WM센터 광주	062) 382-6611
강남역 금융센터	02) 532-0210	IBK WM센터 일산	031) 904-3450
분당센터	031) 705-3600	IBK WM센터 판교	031) 724-2630
부산센터	051) 805-2900	IBK WM센터 시화공단	031) 498-7900
대구센터/IBK WM센터 대구	053) 752-3535	IBK WM센터 남동산단	032) 822-6200
IBK WM센터 목동	02) 2062-3002	IBK WM센터 평촌	031) 476-1020
IBK WM센터 한남동	02) 796-8500	IBK WM센터 천안	041) 569-8130
IBK WM센터 동부이촌동	02) 798-1030	IBK WM센터 부산	051) 741-8810
IBK WM센터 중계동	02) 948-0270	IBK WM센터 창원	055) 282-1650
IBK WM센터 강남	02) 2057-9300	IBK WM센터 울산	052) 271-3050
IBK WM센터 반포자이	02) 3481-6900		