

Project description: <Gold Price Forecasting>

과거 데이터를 이용하여 미래의 금값(종가; Close)을 예측하는 모델 개발 (Python 활용)

데이터셋

- Finance data reader 이용 (<https://github.com/FinanceData/FinanceDataReader>)
- 2010년 1월부터의 daily data
- 원자재 데이터, 환율 데이터, 암호화폐 데이터 (제공된 데이터 이외 데이터 사용 금지)
- Train data: 2010년 1월 1일 ~ 2024년 5월 15일
- Test data: 2024년 5월 16일 ~ 2024년 5월 27일 (Data leakage에 유의)

	Open	High	Low	Close	Adj Close	Volume
Date						
2010-01-04	1117.699951	1122.300049	1097.099976	1117.699951	1117.699951	184.0
2010-01-05	1118.099976	1126.500000	1115.000000	1118.099976	1118.099976	53.0
2010-01-06	1135.900024	1139.199951	1120.699951	1135.900024	1135.900024	363.0
2010-01-07	1133.099976	1133.099976	1129.199951	1133.099976	1133.099976	56.0
2010-01-08	1138.199951	1138.199951	1122.699951	1138.199951	1138.199951	54.0
...

<데이터셋 예시: Gold price dataset>

다음의 3개 유형의 모델을 구축

1. ARIMA 계열 모델 활용 1건
2. HMM 모델 활용 1건
3. 그 외 ML/DL 모델 활용 1건 (자유)

제출물

- PPT 발표자료 1개 (보고서 대체, 상세히 기술) (별도 제출)
- 모델 유형 별 압축파일 1개씩
 - 소스코드(train.py, test.py + α),
 - train.py를 통해 생성된 모델 저장 파일 (test.py에서 읽어서 사용)
 - test.py만을 실행 시, test set의 실제값, 예측값, 계산된 MAE 값이 출력되어야 함
- 발표자료 1개, 압축파일 3개, 총 4개 파일을 업로드

발표자료 포함 내용

- 데이터 이해, 시각화
- 전처리 방법, feature engineering, 활용한 feature 목록,
- 모델에 대한 설명
- Model selection, 사용한 파라미터
- Test set에 대한 예측 및 그 결과 (실제 그래프/예측 그래프 및 MAE 계산값)
- ...

발표

- 15분 발표
- 온라인 비디오 녹화
- 자료가 우수한 팀을 선정하여 오프라인 발표

Grading

합리성, 논리성, 창의성, 전달력, forecasting result + α