

S&P 500 지수 추종 ETF

SPY 가격 예측

신형섭

SPY란

- 미국 S&P 500 지수 추종 ETF인 SPDR S&P 500을 말하며 티커 'SPY'
- S&P 500 지수는 다우존스, 나스닥 지수와 더불어 미국 증시의 3대 주가지수 중 하나이며, 실질적으로 미국증시를 대표하는 주가지수
- 미국 주식시장에 상장된 500개 대형기업이 포함된 주가지수로 미국 상장기업 시가총액의 80% 이상을 포함하고 있어 미국 주식 시장을 가장 잘 대변
- 미국 시장의 안전성 등을 고려하여 많은 투자자들이 선호하는 ETF 중 하나

- 주가 지수 예측 모델으로 최근 20일의 종가를 활용하여 다음 날의 종가를 예측
- 지속적으로 데이터를 업데이트하면서 매일의 종가 예측을 통해 buy, sell의 판단이 가능할 것이다.

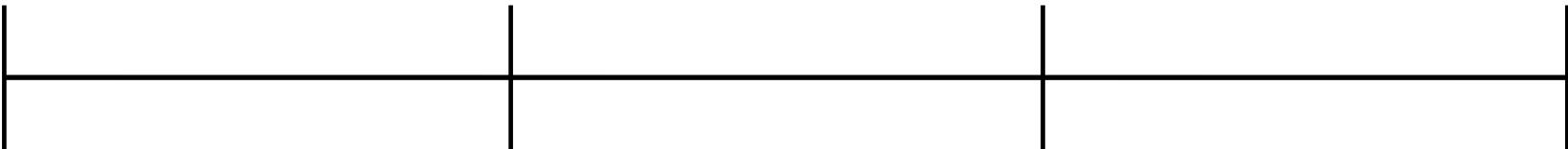
TASK FLOW

데이터 수집

데이터 분리

데이터 전처리

모델링(LSTM)



데이터 수집

	Date	Open	High	Low	Close	Adj Close	Volume
0	2021-05-05	417.380005	417.630005	415.149994	415.750000	405.053528	60162200
1	2021-05-06	415.829987	419.209991	413.679993	419.070007	408.288055	74321400
2	2021-05-07	419.890015	422.820007	419.160004	422.119995	411.259552	67733800
3	2021-05-10	422.500000	422.739990	417.809998	417.940002	407.187103	81852400
4	2021-05-11	413.100006	415.269989	410.059998	414.209991	403.553101	116888000
...
437	2023-01-30	402.799988	405.130005	400.279999	400.589996	400.589996	74202000
438	2023-01-31	401.130005	406.529999	400.769989	406.480011	406.480011	86811800
439	2023-02-01	405.209991	413.670013	402.350006	410.799988	410.799988	101459200
440	2023-02-02	414.859985	418.309998	412.880005	416.779999	416.779999	101654500
441	2023-02-03	411.589996	416.970001	411.089996	412.350006	412.350006	94647400

442 rows x 7 columns

최근 2년 데이터 (2021-05-05 ~ 2023-02-03)

1. 불필요한 컬럼 삭제

-> 조정 종가(Adj Close) 삭제

2. 날짜를 컬럼 -> 인덱스

-> 날짜는 종가에 영향을 주는 요소가 아니라고 판단

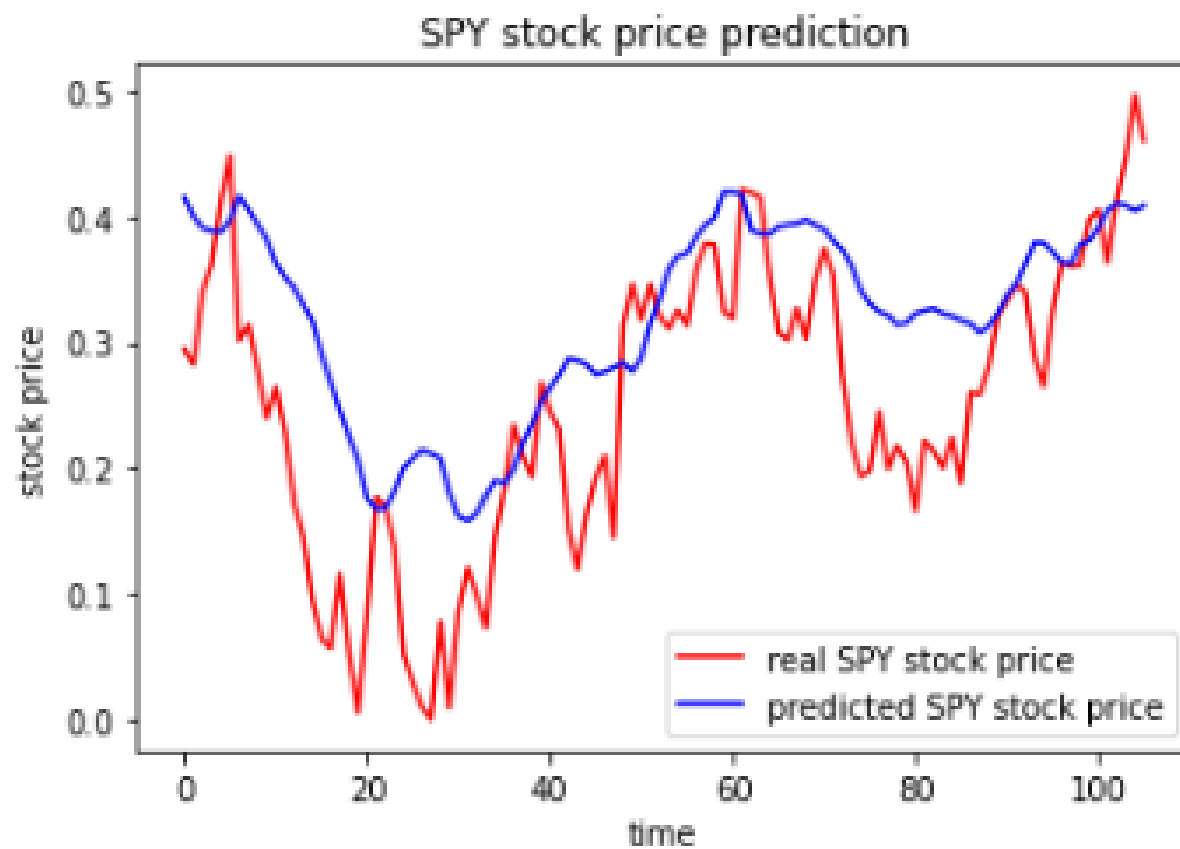
3. 데이터 변환(MinMaxScaler)

-> 거래량 컬럼의 경우 가격 컬럼들과 수치가 크게 차이가 있어서 0-1 사이 수로 표현할 수 있는 MinMaxScaler를 통해 데이터 변환 작업 진행

- Long Short-Term Memory의 줄임말로 RNN의 단점을 보완한 모델
- 시계열 데이터 분석에 용이

```
model = Sequential()  
model.add(LSTM(units=20, activation='relu', return_sequences=True, input_shape=(20, 4)))  
model.add(Dropout(0.5))  
model.add(LSTM(units=20, activation='relu'))  
model.add(Dropout(0.5))  
model.add(Dense(units=1))  
model.summary()
```

모델링 결과



모델링 결과



2월 6일 SPY 종가

모델 예측 종가 : 408.9152

실제 종가 : 409.8300

- 시가, 고가, 저가 등으로 종가를 예측하는 것은 상당히 어려운 일이고, 추가적으로 공시나 뉴스 등을 감성 분석하는 모델도 함께 사용된다면 정확도를 높일 수 있을 것
- 실제 종가와 예측 종가가 0.9 달러의 차이가 생겼지만, 코드를 실행할 때마다 다른 결과값이 나오기도 하였으며 값의 차이가 클 때도 있고 적을 때도 있었기에 해당 부분에 대한 고민도 필요