



#IT 신기술 서비스 프로젝트

주가예측



1. 프로젝트 개요

1. 프로젝트 주제 : 나스닥 주가예측
2. 프로젝트 개요: 딥러닝 기술을 활용하여 5, 6, 7 거래일 후 나스닥(NASDAQ Composite) 지수 예측
3. 프로젝트 연구 과정
 - 데이터 : 5년 동안의 종가
 - 모델 : LSTM
 - 평가지표: RMSE

2. 프로젝트 수행 과정

(1) 모델 1

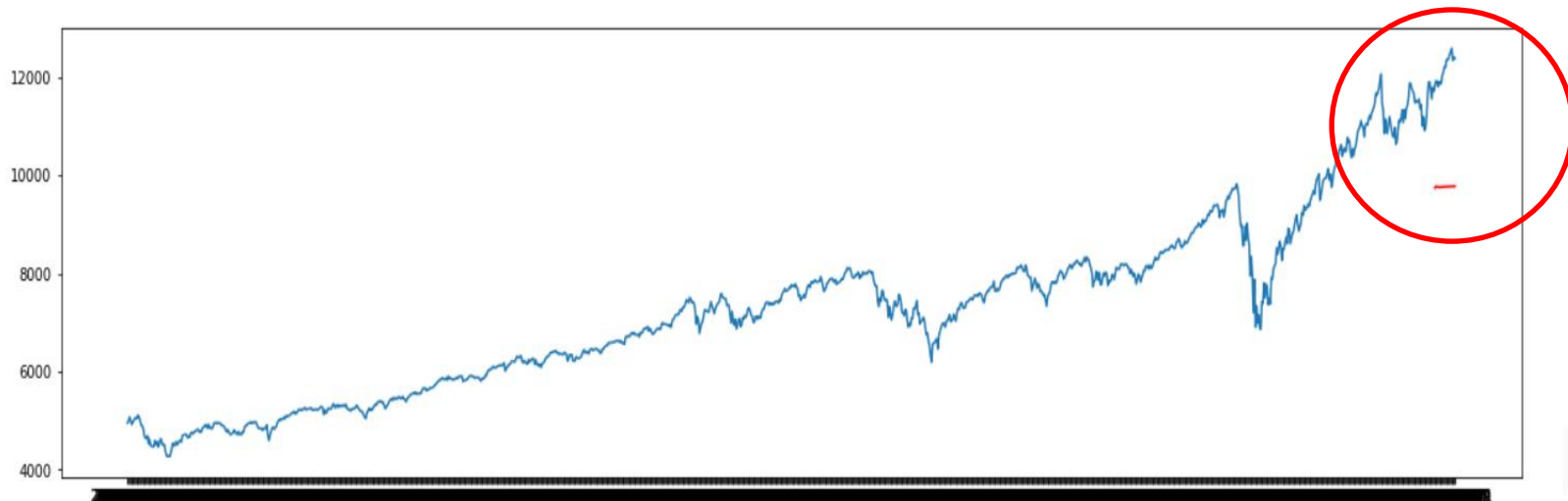
```
generator=TimeseriesGenerator(train, train,length=n_input, batch_size=15)

model = Sequential()
model.add(LSTM(200, activation='relu', input_shape=(n_input, n_features)))
#model.add(Dropout(0.15))
model.add(Dense(1))
model.compile(optimizer='adam', loss='mse')
model.fit_generator(generator, epochs=100)
```

- 훈련,테스트 데이터 분할
- 정규화
- LSTM 사용

2. 프로젝트 수행 과정

(1) 모델 1



2. 프로젝트 수행 과정

(1) 모델 1

```
#RMSE 구하기 - 평가지표  
pred_actual_rmse = rmse(df_test.iloc[-n_in  
print("rmse: ", pred_actual_rmse)
```

```
rmse: [2260.21129551]
```

...
2020-12-26	NaN	13141.418849
2020-12-27	NaN	13187.638305
2020-12-28	NaN	13233.361115
2020-12-29	NaN	13278.618507
2020-12-30	NaN	13323.413530

181 rows x 2 columns

→ 성능 안 좋음

→ 예측 값

- 2020.12.22: 12619.605596

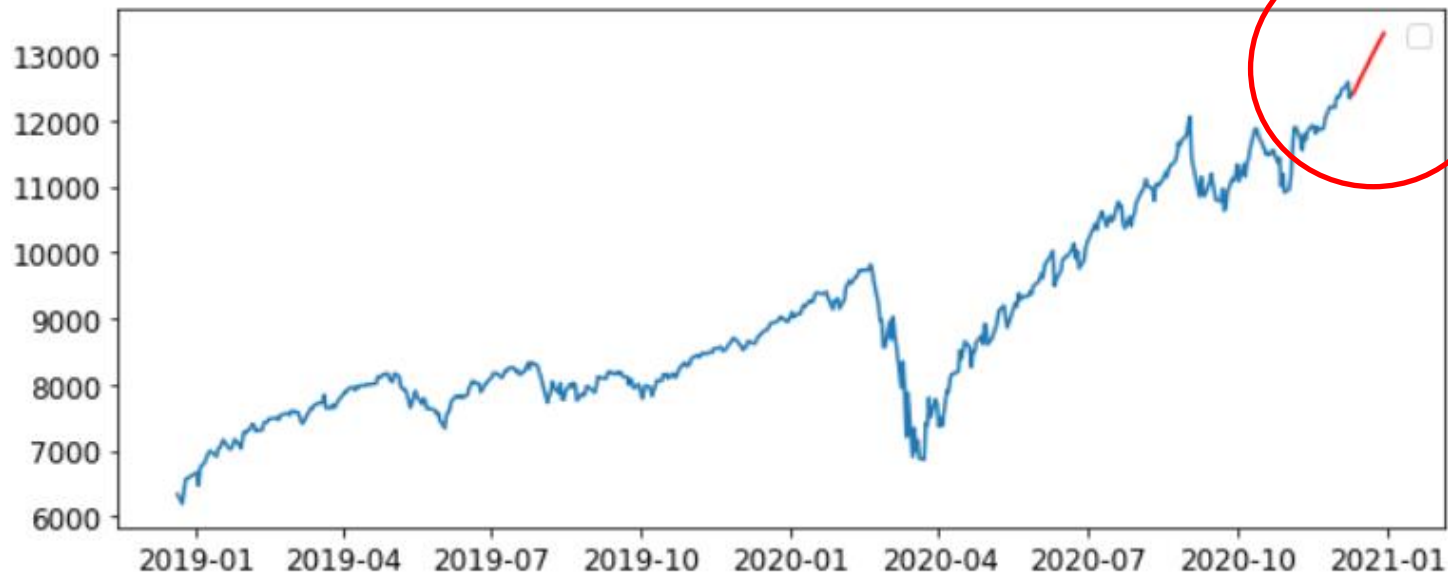
- 2020.12.23: 12644.503992

- 2020.12.24: 12670.269808

2. 프로젝트 수행 과정

(1) 모델 1

No handles with labels found to put in legend.



2. 프로젝트 수행 과정

(2) 모델 2

```
# 훈련, 검증 데이터 훈련

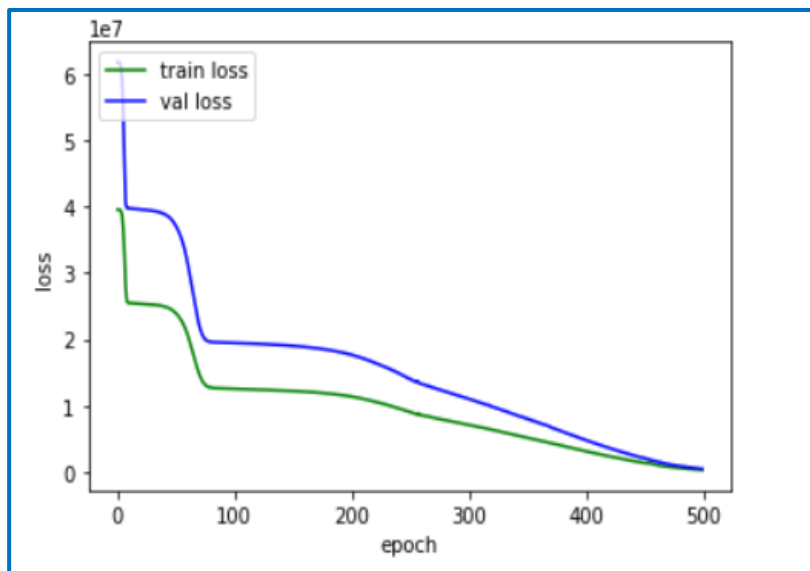
x_train = x_train.reshape((x_train.shape[0], x_train.shape[1], 1))
x_test = x_test.reshape((x_test.shape[0], x_test.shape[1], 1))
x_val = x_val.reshape((x_val.shape[0], x_val.shape[1], 1))

callbacks_list = [
    EarlyStopping(monitor='val_loss', patience=100, ),
    ModelCheckpoint(filepath='best_model.h5', monitor='val_loss',
                    save_best_only=True, )]
hist = model.fit(x_train, y_train, epochs=500, callbacks=callbacks_list, validation_data=(x_val, y_val), shuffle=False)
```

- 훈련, 검증, 시험 데이터 분할 사용
- 정규화, EarlyStopping, 모델 저장
- LSTM 사용

2. 프로젝트 수행 과정

(2) 모델 2



RMSE 구하기 (훈련, 검증)

```
rmse = np.sqrt(mean_squared_error(y_pre, y_val))
```

```
print("RMSE of train: %.3f"%np.sqrt(hist.history['loss'][-1]))
```

```
print("RMSE of val : %.3f"%np.sqrt(hist.history['val_loss'][-1]))
```

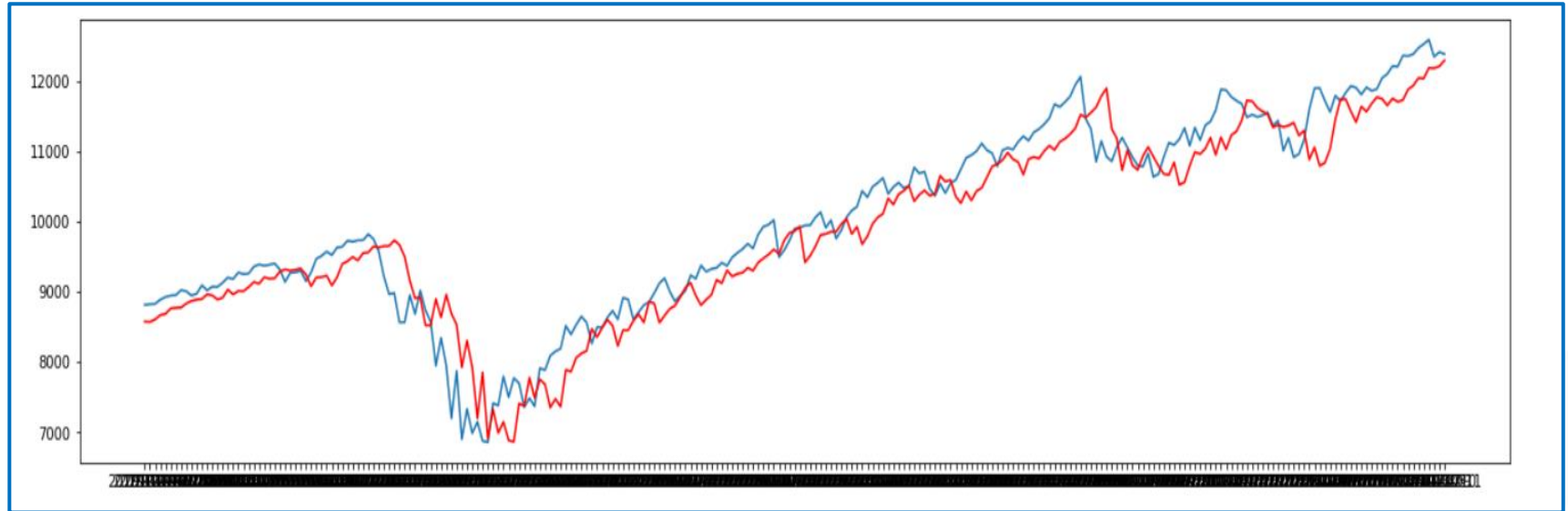
RMSE of train: 134.324

RMSE of val : 178.675

→ 성능 좋음

2. 프로젝트 수행 과정

(2) 모델 2



→ 예측이 잘됨

2. 프로젝트 수행 과정

(2) 모델 2

```
# 모델 로드  
from keras.models import load_model  
model = load_model('best_model.h5')
```

Date	Close
2015-12-14	4952.229980
2015-12-15	4995.359863
2015-12-16	5071.129883
2015-12-17	5002.549805
2015-12-18	4923.080078
...	...
2020-12-08	12582.769531
2020-12-09	12338.950195
2020-12-10	12405.809570
2020-12-11	12377.870117
2020-12-14	12362.327148

[1260 rows x 1 columns]

→ 제일 성능이 좋은
모델로 예측

3. 프로젝트 결론 & 느낀점

- 훈련, 검증, 시험 데이터 모두를 사용했을 때 성능이 더 좋았다.
- 데이터가 많을수록 학습이 잘되었다.
- Shape 에러를 해결하는 것이 어려웠다.