

機械学習を用いた株価の 短期予測と長期予測

総合情報学科 情報システム学系

知能情報システム研究室

永井ゼミ4年 越谷 心

研究の動機と目的

○動機

株式投資ではデータの収集、分析に時間を要する。そこで機械学習を用いた予測データを用いることで、一つの大きな判断材料として活用できると考えた。

○目的

過去の株価データから短期、長期で今後の推移を予測し、投資に役立てる。(短期：5ヶ月、長期:5年)

関連研究

- 深層学習を用いた株価予測の分析[宮崎 邦洋,松尾 豊]
- 深層学習によるニュースと株式市場の関連性の分析：高頻度データ およびLSTMによるニュース分類[高山 理璃子, 菅 愛子, 高橋 大志]

関連研究では、学習モデルにRNNやCNN,LSTMを用いており、バランスの取れた予測をしている。本研究ではあえて極端に予測し、投資スタイルに合わせた短期と長期にわけた予測をする。RNN,CNN,LSTMだけでなくRNNの問題点を改善し、BERTやGPTにも使われているTransformerも用いてそれぞれの予測を比較する。

使用データ

短期予測(5ヶ月)

学習データ 2010年1月1日～2023年6月30日

検証データ 2023年7月1日～2023年12月31日

```
df_train = df[df.index <= pd.Timestamp('2023-06-30')]  
df_val = df[df.index > pd.Timestamp('2023-06-30')]
```

長期予測(5年)

学習データ 2000年1月1日～2018年12月31日

検証データ 2019年1月1日～2023年12月31日

```
df_train = df[df.index <= pd.Timestamp('2018-12-31')]  
df_val = df[df.index > pd.Timestamp('2018-12-31')]
```

データの前処理

○短期予測

正規化

```
scaler = StandardScaler()  
scaled_train_data = scaler.fit_transform(train_data)  
scaled_val_data = scaler.transform(val_data)
```

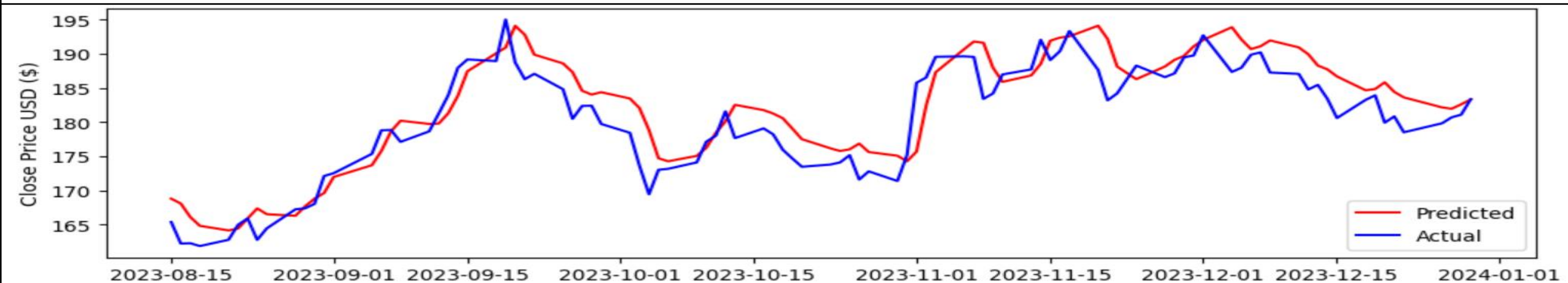
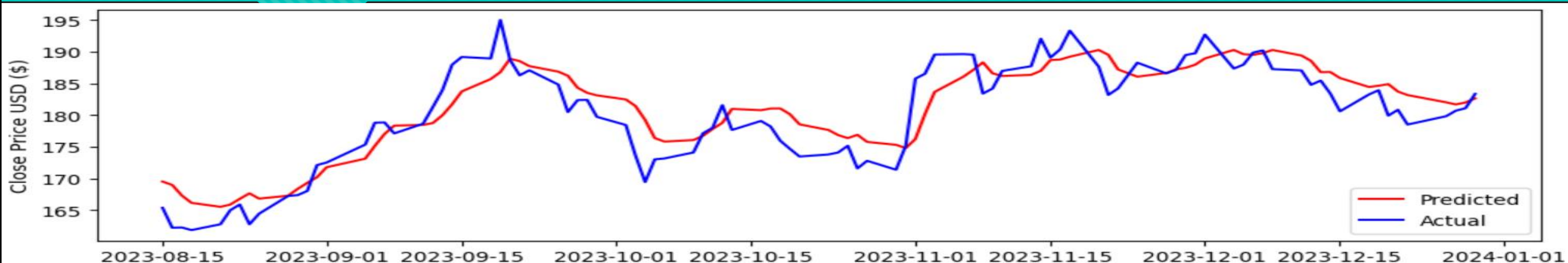
○長期予測

標準化

```
scaler = MinMaxScaler(feature_range=(0, 1))  
dataset = scaler.fit_transform(dataset)
```

予測結果

検証データをTransformerとLSTMで短期予測したものと実際のデータを可視化



Root Mean Squared Error (RMSE): 3.6474592175432052

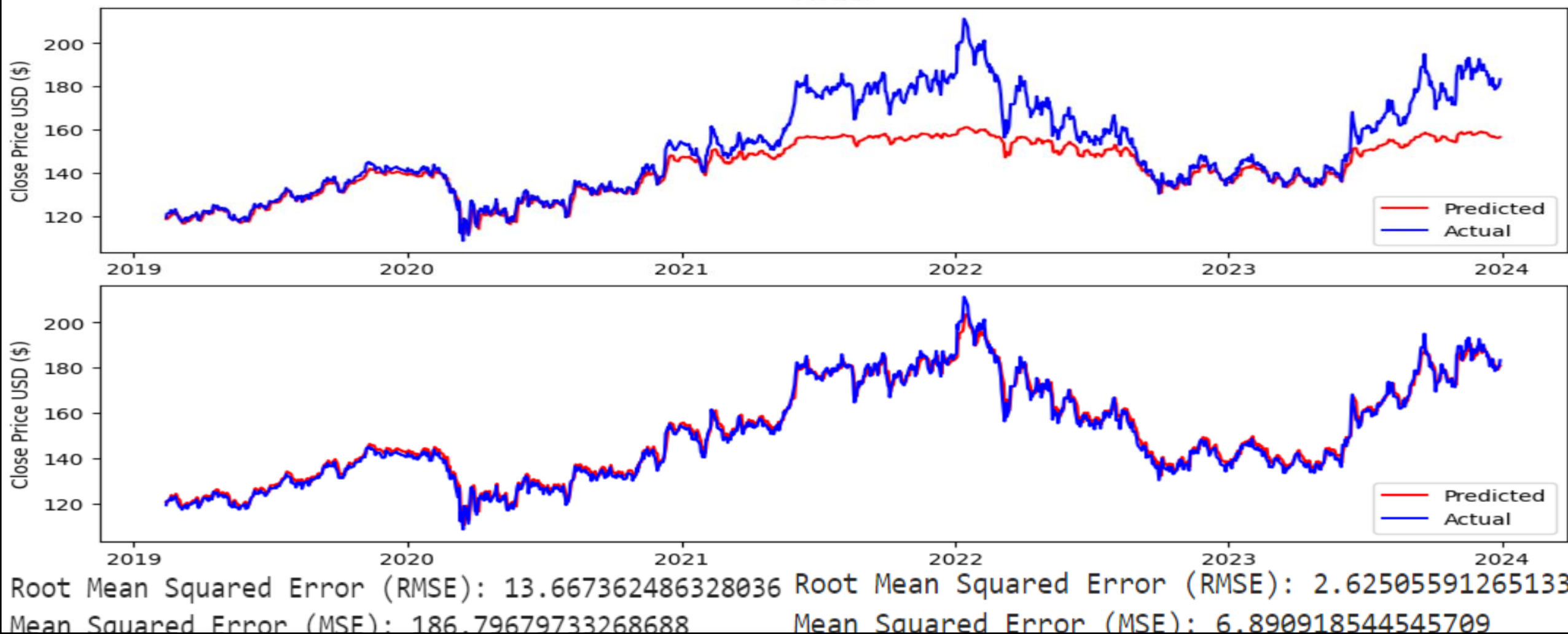
Mean Squared Error (MSE): 13.303958743640882

Root Mean Squared Error (RMSE): 3.760829117016338

Mean Squared Error (MSE): 14.143835647397887

予測結果

検証データをTransformerとLSTMで長期予測したものと実際のデータを可視化



今後の予定

○違うモデルや銘柄を用いて比較

モデルや銘柄を変えてどのように違いが出るかを比較していく。

(7月～9月)

○精度の向上

現状のLSTMとTransformerをパラメータの調整や特徴量を終値以外にも追加して精度を向上させる。

(10月中)