

# 株式会社Athena Technologies インターン課題

## 機械学習による 株価予測モデルの構築

川越 壮

# 目次

- 背景
- 技術概要
- データ分析
- 評価指標
- モデルの構築と結果
- 検証1と結果
- 検証2と結果
- まとめ

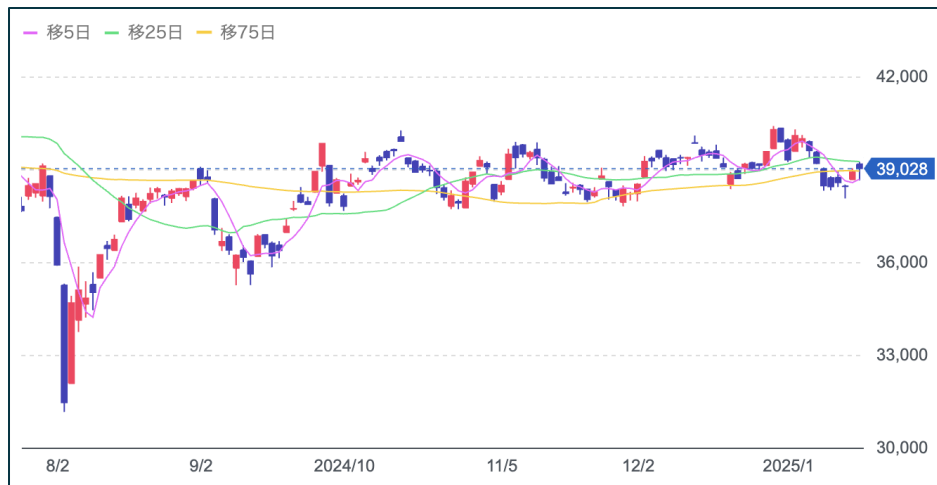
# 目次

- 背景
- 技術概要
- データ分析
- 評価指標
- モデルの構築と結果
- 検証1と結果
- 検証2と結果
- まとめ

背景

# 株価予測の重要性

株式市場は企業の成長や経済の健全性を反映す場  
資産の運用手段として重要



日経平均株価のチャート, <https://finance.yahoo.co.jp/quote/998407.O/chart?trm=6m>, 2025-1-22

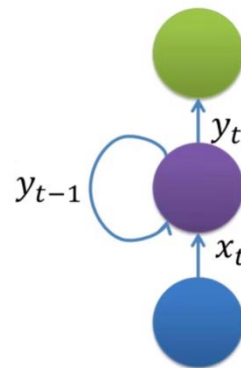
# AIによる株価予測

近年，AIによる株の取引が増加している

→大量のデータを学習し，予測に反映することができる

具体的に株価の予測で使われるAIモデルの例

- Recurrent Neural Network (RNN)
  - ・ 内部的に再帰的な構造を持つため，  
時系列データを用いた短期記憶的な学習に有効



通常のRNN<sup>[1]</sup>

[1]わかるLSTM～最近の動向と共に, [https://qiita.com/t\\_Signull/items/21b82be280b46f467d1b](https://qiita.com/t_Signull/items/21b82be280b46f467d1b), 2025-1-22

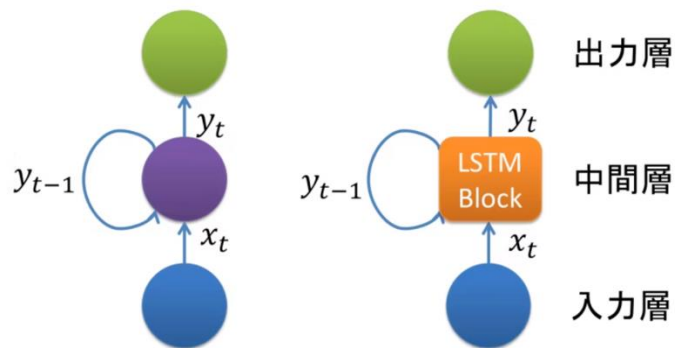
# 目次

- 背景
- 技術概要
- データ分析
- 評価指標
- モデルの構築と結果
- 検証1と結果
- 検証2と結果
- まとめ

# 使用したモデル

## Long-Short Term Memory (LSTM)

- RNNの勾配消失問題を解消したモデル
- LSTMブロックの導入により，長期記憶が可能



通常のリNN      Long-Short Term Memory <sup>[1]</sup>

[1]わかるLSTM～最近の動向と共に, [https://qiita.com/t\\_Signull/items/21b82be280b46f467d1b](https://qiita.com/t_Signull/items/21b82be280b46f467d1b), 2025-1-22

# 目次

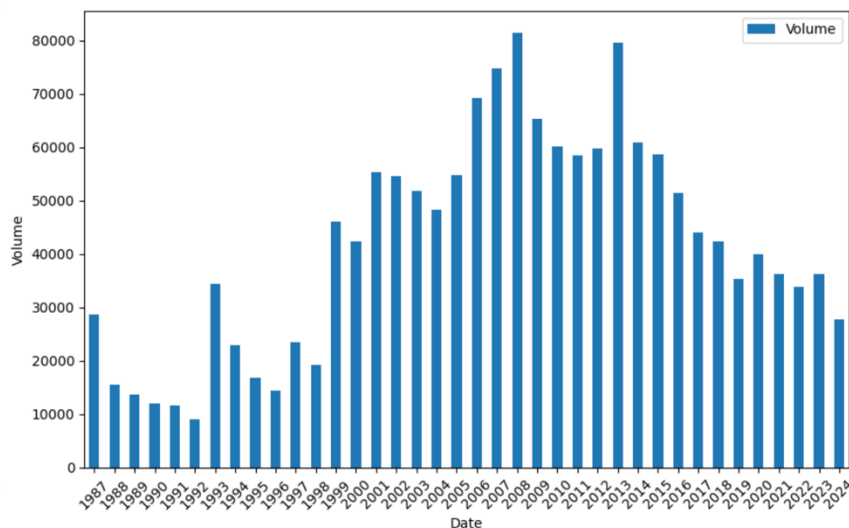
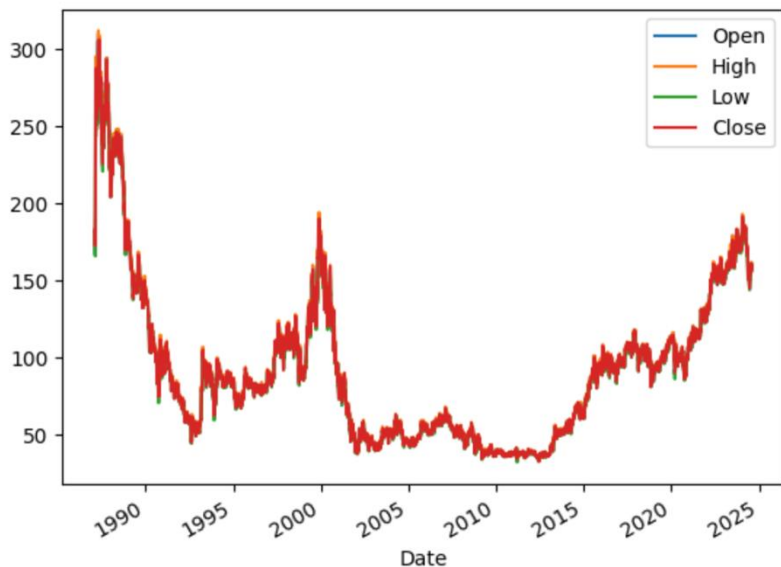
- 背景
- 技術概要
- データ分析
- 評価指標
- モデルの構築と結果
- 検証1と結果
- 検証2と結果
- まとめ



## データの確認

今回使用したデータ : NTTの株価 (1987/2/18~2024/8/1)

元データには日付, 始値, 高値, 安値, 終値, 出来高, 変化量



## 特徴量の追加

元データにある始値、高値、安値、終値に加え、  
チャートの分析によく用いられる一目均衡表の要素である以下の要素を算出した

- 転換線  $= (9 \text{ 日間の高値} + 9 \text{ 日間の安値}) \div 2$
- 基準線  $= (26 \text{ 日間の高値} + 26 \text{ 日間の安値}) \div 2$
- 先行スパンA  $= (\text{転換線} + \text{基準線}) \div 2$
- 先行スパンB  $= (52 \text{ 日間の高値} + 52 \text{ 日間の安値}) \div 2$
- 遅行スパン  $= \text{現在の終値を26日遅らせたもの}$



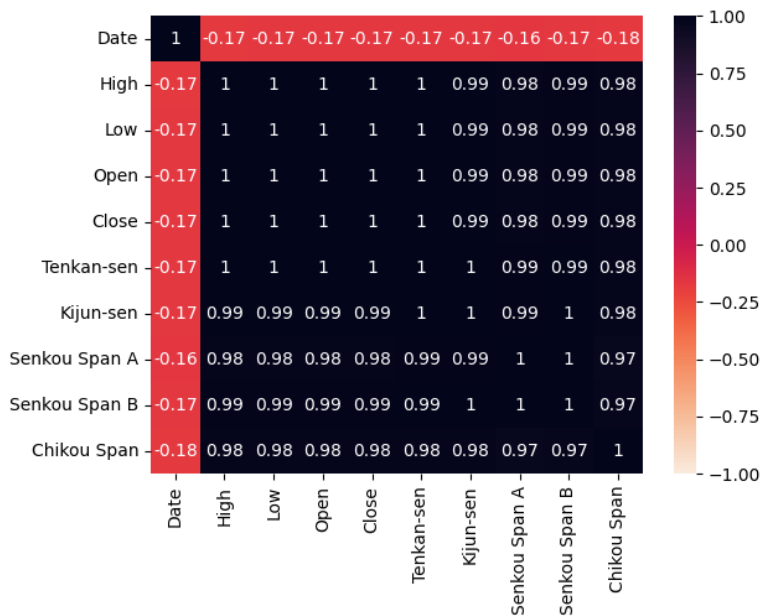
一目均衡表<sup>[2]</sup>

## 特徴量間の相関係数

今回は翌日の終値 (Close) を予測することを目的としている

→ Close と他の要素との相関が高いことが確認できる

これら9つの特徴量を学習データとして用いる



# 目次

- 背景
- 技術概要
- データ分析
- 評価指標
- モデルの構築と結果
- 検証1と結果
- 検証2と結果
- まとめ

# モデルの評価指標

今回の評価指標としては以下の2つを採用した

## 1. RMSE (平均2乗偏差)

- $N$ : データ数,  $y_i$ : 実測値,  $\hat{y}_i$ : 予測値

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=0}^N (y_i - \hat{y}_i)^2}$$

## 2. Accuracy (2値問題の正解率)

- 翌日の終値が前日の終値より上がるか下がるか（変動方向）を予測し、的中した確率

$$Accuracy = \frac{\text{（変動方向を当てた数）}}{\text{（データ数）}} \times 100$$

→チャートの概要の評価(RMSE), 実用上の評価(Accuracy)の両方を考慮

# 目次

- 背景
- 技術概要
- データ分析
- 評価指標
- モデルの構築と結果
- 検証1と結果
- 検証2と結果
- まとめ

# モデルの構築

## ● LSTM

- 1層のLSTMと1層の線形層
- 損失関数 : MSE (平均2乗誤差)
- 最適化関数 : Adam
- 入力 : 過去5日間のデータ (9つの特徴量)
- 出力 : 翌日の終値

## ● データセット

- 学習データ : テストデータ = 8 : 2

## モデルの構築と結果

# 結果

以降, RMSEとAccuracyは  
5回学習した平均値を示す

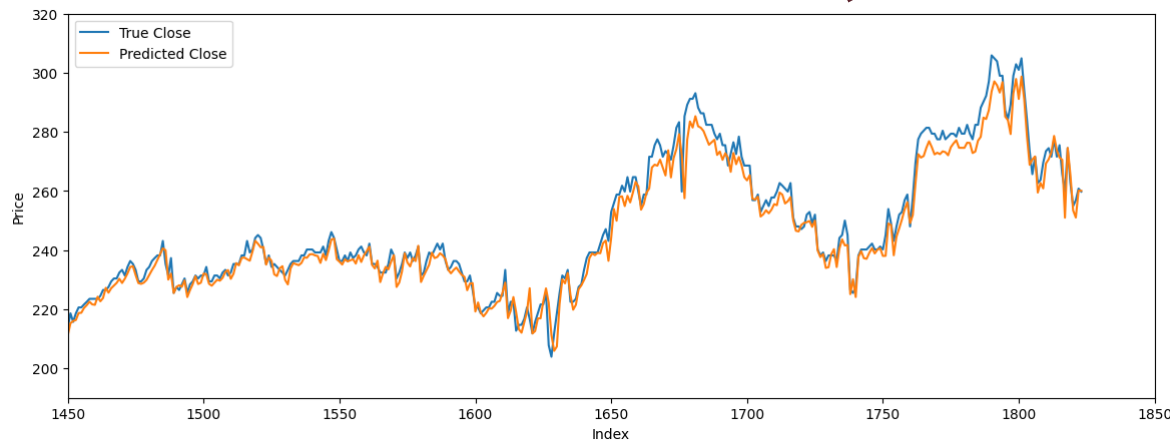
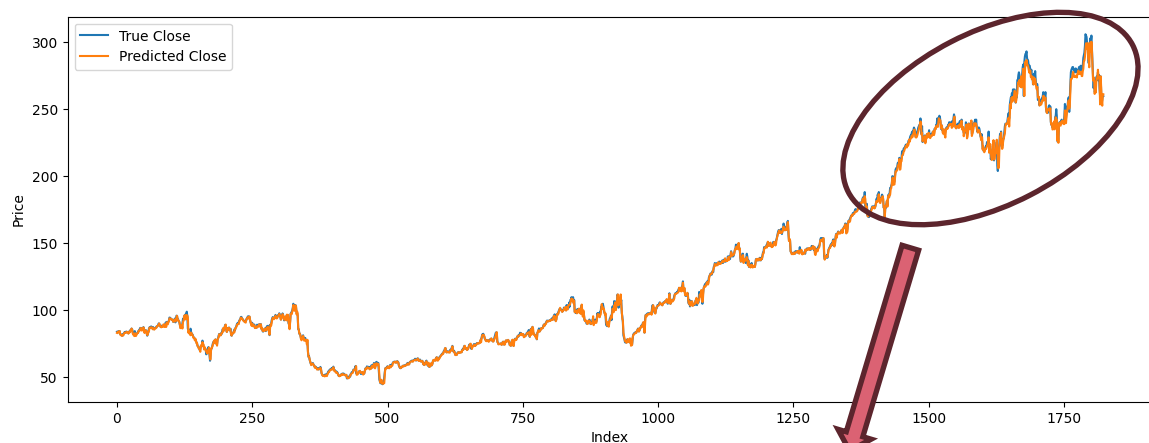
RMSE = 2.12

Accuracy = 68.57%

上. テストデータ全体

下. 乖離が観察された区間

(Index : 1450 ~)



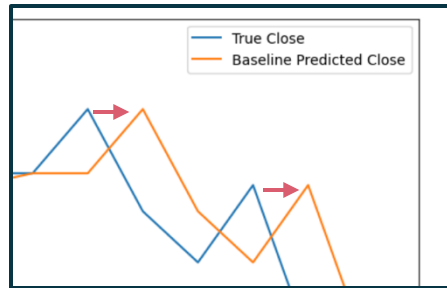


# ベースラインとの比較

比較対象として、前日の終値をそのまま出力するベースラインを導入

→チャートが1日遅れた形になり、

実測値のチャートのトレンドを踏襲したチャートになる



### ● 精度比較

	RMSE	Accuracy
ベースライン	2.77	42.68
LSTM	<b>2,12</b>	<b>68.57</b>

→ベースラインよりもRMSEは小さく、Accuracyは高いことが確認できる

# 目次

- 背景
- 技術概要
- データ分析
- 評価指標
- モデルの構築と結果
- 検証1と結果
- 検証2と結果
- まとめ

## タイムステップの調整による精度改善

LSTMにおいて重要なハイパーパラメータとなるタイムステップの調整を行った

time_step	5	10	30	60	90
RMSE	2.12	<b>2.00</b>	3.42	3.91	3.52
Accuracy	68.57	68.85	<b>69.89</b>	69.45	69.68

RMSEに関してはtime\_stepの間隔を狭くすることで小さくなる

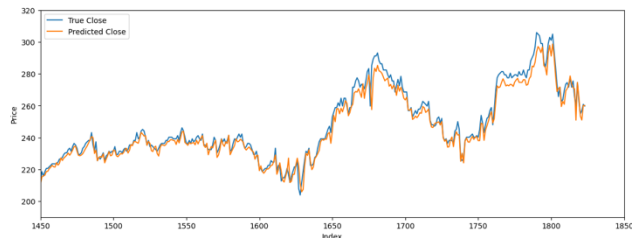
一方, Accuracyに関してはtime\_step=30で最高となった

ベースラインのRMSE (2.77)を考慮するとtime\_step=10が最適

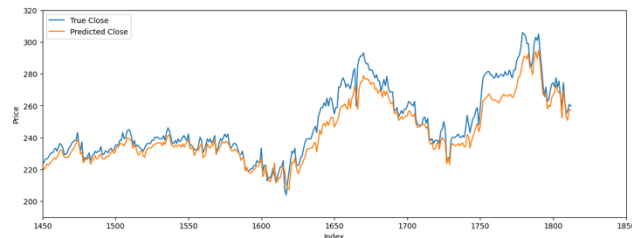
## 検証1と結果

# タイムステップの調整によるチャートの比較

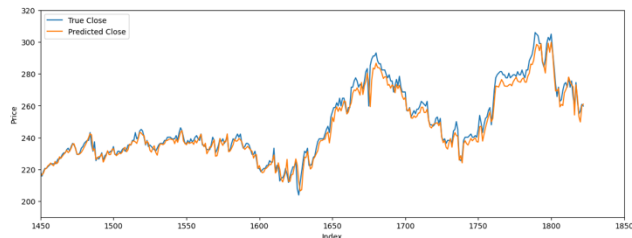
time\_step=5



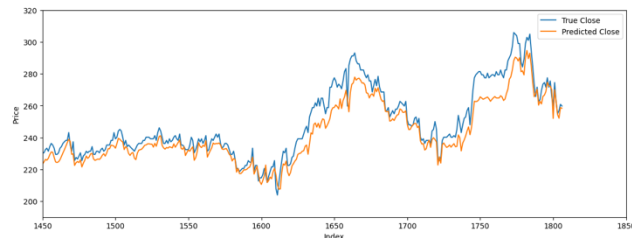
time\_step=60



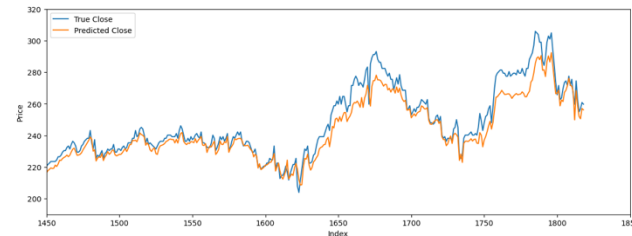
time\_step=10



time\_step=90



time\_step=30



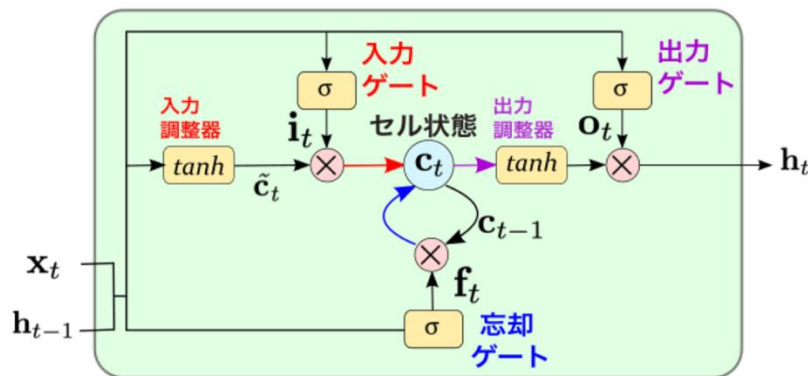
time\_step=10のときが実測値との乖離が最も小さいが、  
どのtime\_stepも実測値より下回った値となる傾向がある

# 目次

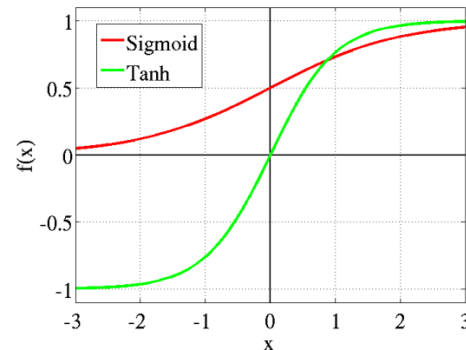
- 背景
- 技術概要
- データ分析
- 評価指標
- モデルの構築と結果
- 検証1と結果
- 検証2と結果
- まとめ

# 正規化の方法の変更による精度改善

LSTMの活性化関数にはtanhも使われている



LSTMのブロックの図<sup>[3]</sup>



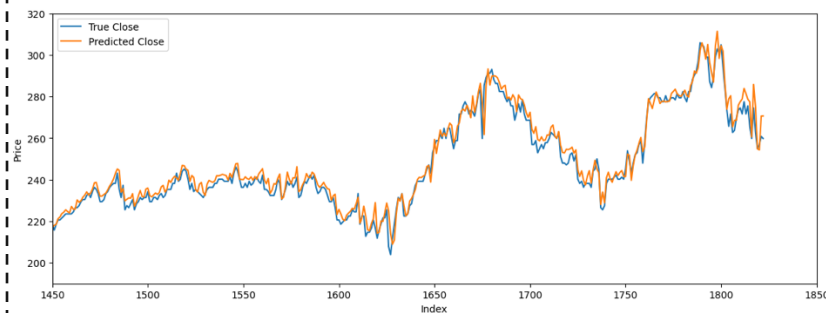
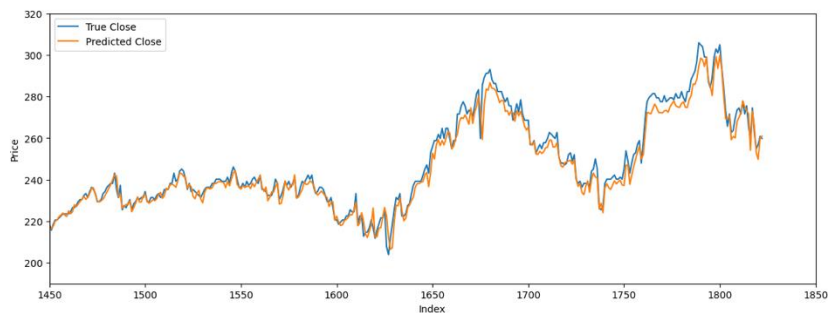
$\tanh$ は(-1,1)をとり得る

→正規化を(0,1)から(-0.5,0.5)に変えた方が表現力が上がる可能性がある

time\_step=10の場合に対して正規化を(-0.5,0.5)にして精度改善を図った

# 正規化の方法の変更による精度改善の結果

(0, 1) で正規化	(-0.5, 0.5) で正規化
RMSE = 2.00	RMSE = 2.13
Accuracy = 68.85	Accuracy = 68.40



評価指標としては改善しなかったが、

チャートを見ると実測値の急激な上昇に対応できていることがわかる

# 目次

- 背景
- 技術概要
- データ分析
- 評価指標
- モデルの構築と結果
- 検証1と結果
- 検証2と結果
- まとめ



# | まとめと今後の展望

## まとめ

- LSTMモデルによって株価の予測を行った
- ベースラインを上回るモデルが構築できた
- 翌日の終値の変動方向の2値問題においてチャンスレート(50%)を上回った

## 今後の展望

- 今回は翌日の終値の予測を行ったが、  
ラグモデルとなることを避けるため複数日の終値を予測するモデルを構築する