

# 株価予測におけるTCN-LightGBM 複合モデルの適用

---

株価予測グループ

25R6211 中居幸太 25R6217 李径直 23X4024 菓子尾伊織

# アジェンダ

---

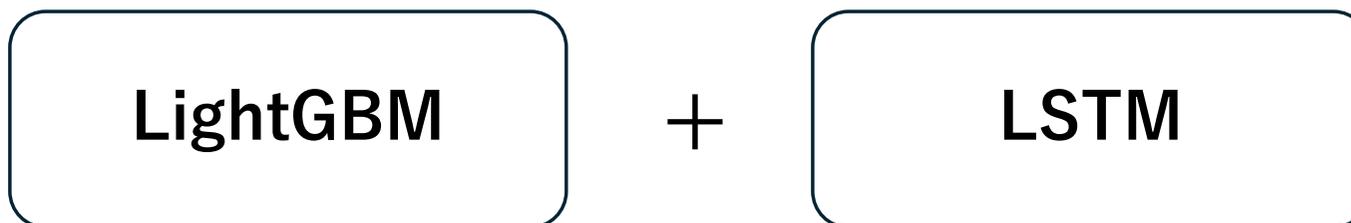
1. 研究背景と目的
2. データの入手方法と特徴量
3. 提案手法
4. 実験設定と評価指標
5. 実験結果と考察
6. 今後の課題
7. まとめ

# 1.研究背景と目的

---

株価予測における**ディープラーニング**（RNN, LSTM）を用いた研究の活発化

先行研究によると…



**精度が改善した！**

# TCN(時間的畳み込みネットワーク)

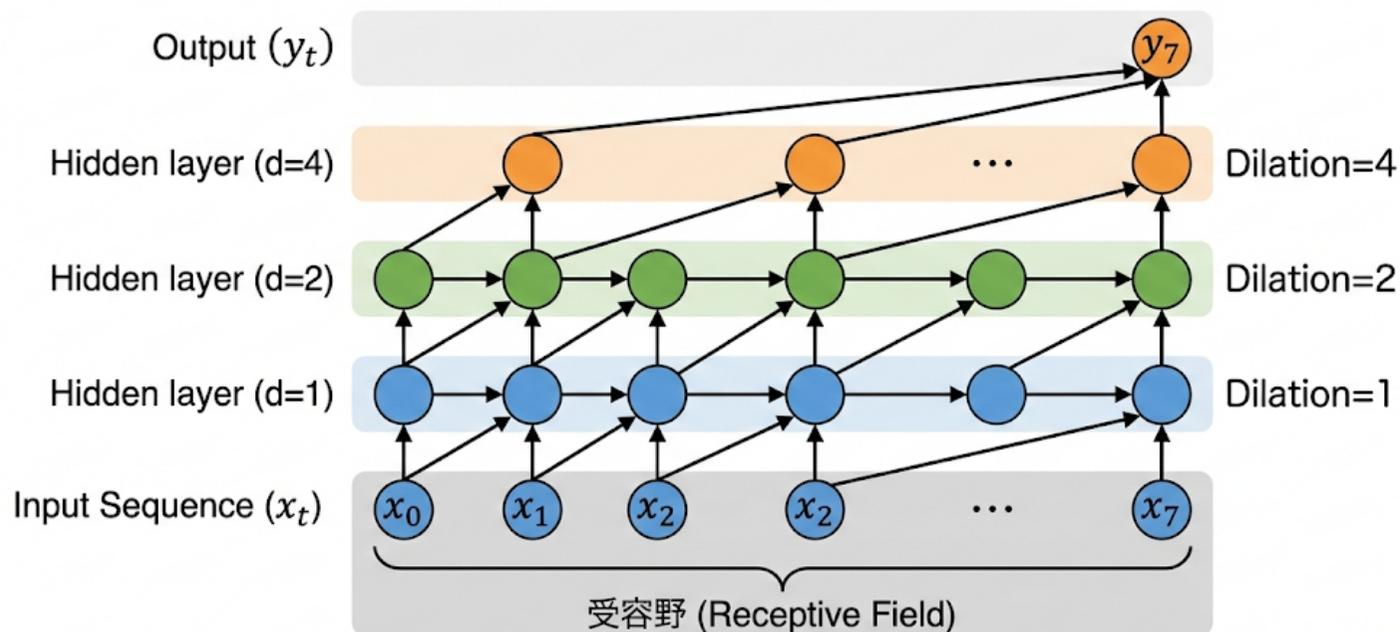
Sequence Modeling Task	Model Size ( $\approx$ )	Models			
		LSTM	GRU	RNN	TCN
Seq. MNIST (accuracy <sup>h</sup> )	70K	87.2	96.2	21.5	<b>99.0</b>
Permuted MNIST (accuracy)	70K	85.7	87.3	25.3	<b>97.2</b>
Adding problem $T=600$ (loss <sup>ℓ</sup> )	70K	0.164	<b>5.3e-5</b>	0.177	<b>5.8e-5</b>
Copy memory $T=1000$ (loss)	16K	0.0204	0.0197	0.0202	<b>3.5e-5</b>
Music JSB Chorales (loss)	300K	8.45	8.43	8.91	<b>8.10</b>
Music Nottingham (loss)	1M	3.29	3.46	4.05	<b>3.07</b>
Word-level PTB (perplexity <sup>ℓ</sup> )	13M	<b>78.93</b>	92.48	114.50	88.68
Word-level Wiki-103 (perplexity)	-	48.4	-	-	<b>45.19</b>
Word-level LAMBADA (perplexity)	-	4186	-	14725	<b>1279</b>
Char-level PTB (bpc <sup>ℓ</sup> )	3M	1.36	1.37	1.48	<b>1.31</b>
Char-level text8 (bpc)	5M	1.50	1.53	1.69	<b>1.45</b>

**LSTM**より高い精度を出している**TCN**というモデルが存在する

表の引用元：Shaojie Bai et al., An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling, 2018

# TCN(時間的畳み込みネットワーク)

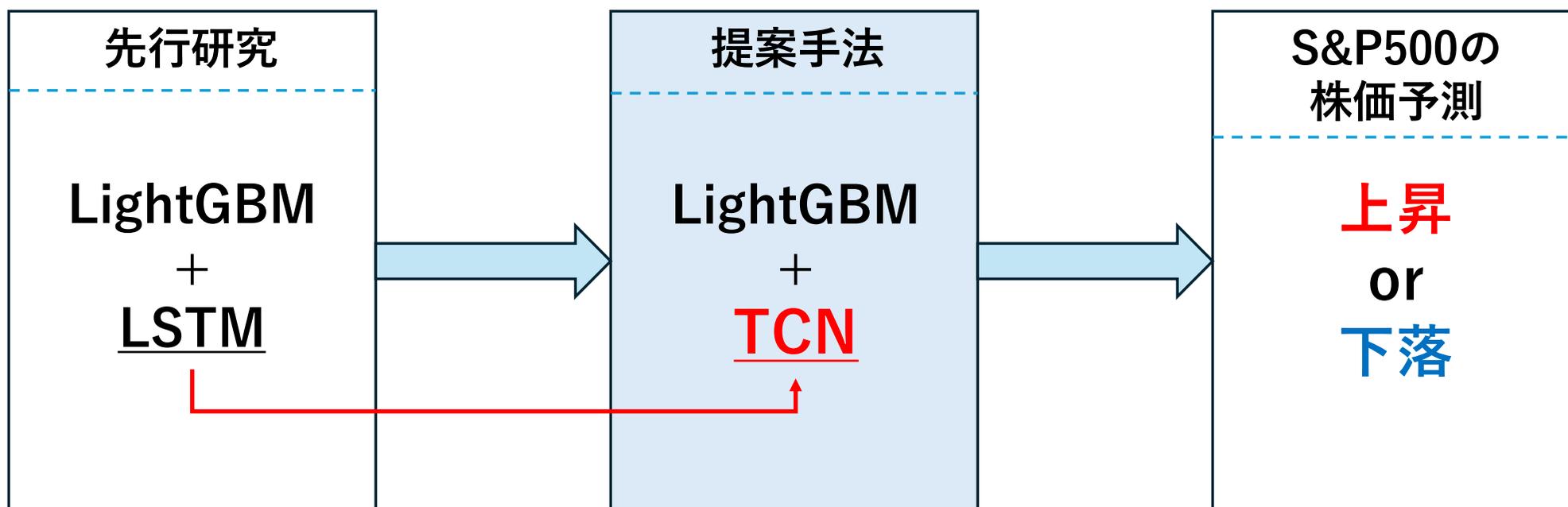
Dilated Causal Convolution (拡張因果畳み込み)の構造図



## TCNの特徴

- CNNを用いた時系列アルゴリズム
- 未来の情報を使用しない構造
- 広い受容野を効率的に確保し、長期依存関係を学習可能

# 1.研究背景と目的



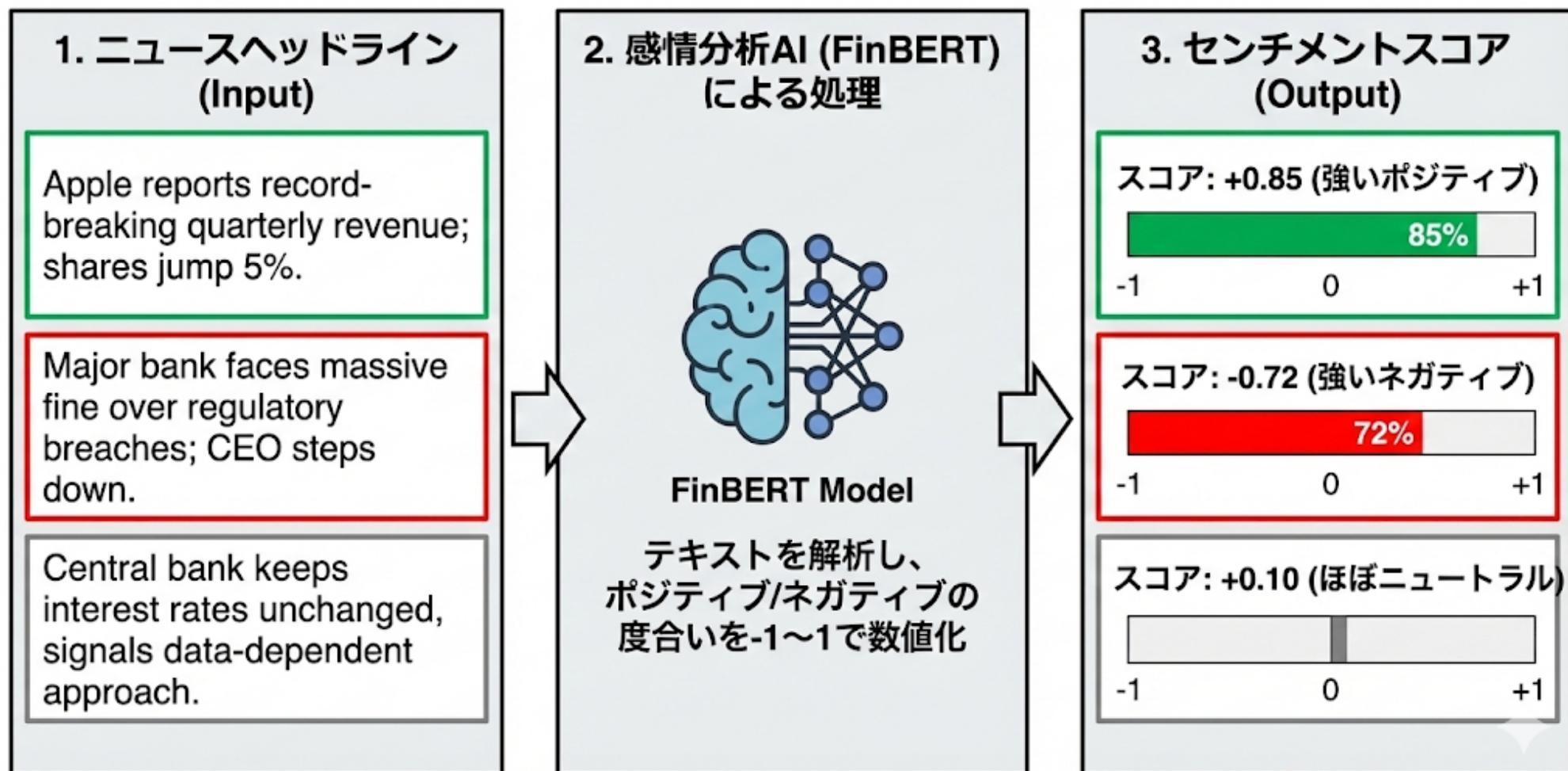
時系列モデルを変更したら精度は上がるか？

## 2. データ入手方法と特徴量

---

- 株価データ： **Investing.com** より入手した2014年～2018年のS&P500株価指数の日足データ
- ニュースデータ： **Kaggle** の Financial News Headlines(2008-2024) より入手した2014年～2018年のデータ
- 経済指標： 米国債利回り

## ニュースヘッドラインの感情分析：プロセスと結果の例 (FinBERT活用)



## テクニカル指標の追加

---

- **RSI**（相対力指数）：買われすぎ・売られすぎの判断
- **MACD**：トレンドと売買タイミングの把握
- **SMA**（10日移動平均）：短期的な価格トレンド
- **ボリンジャーバンド**：ボラティリティの高さの指標

## 代表的なテクニカル指標の計算式 (RSI, MACD)

### RSI (相対力指数 - Relative Strength Index)

100から引くことで0-100の範囲に

$$RSI = 100 - \frac{100}{1 + RS}$$

100から引くことで0-100の範囲に

$$RS = \frac{\text{平均値上がり幅 (Average Gain)}}{\text{平均値下がり幅 (Average Loss)}}$$

※通常、直近14日間の平均を使用

### MACD (移動平均収束拡散手法 - Moving Average Convergence Divergence)

$$\text{MACD線} = \text{EMA}_{12}(\text{終値}) - \text{EMA}_{26}(\text{終値})$$

短期EMAから長期EMAを引いたもの

$$\text{シグナル線} = \text{EMA}_9(\text{MACD線})$$

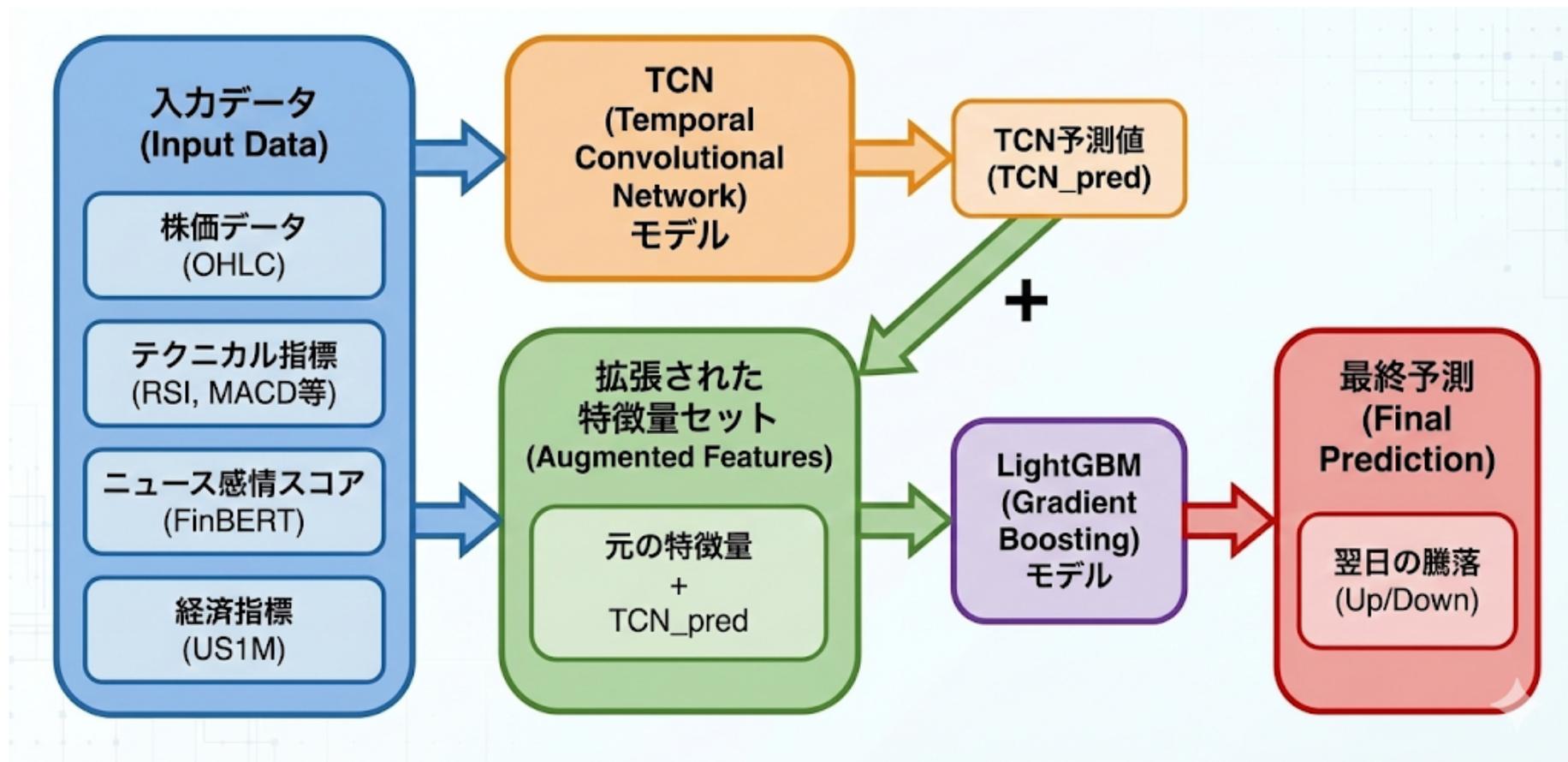
MACD線の9日EMA

$$\text{ヒストグラム} = \text{MACD線} - \text{シグナル線}$$

2本の線の差

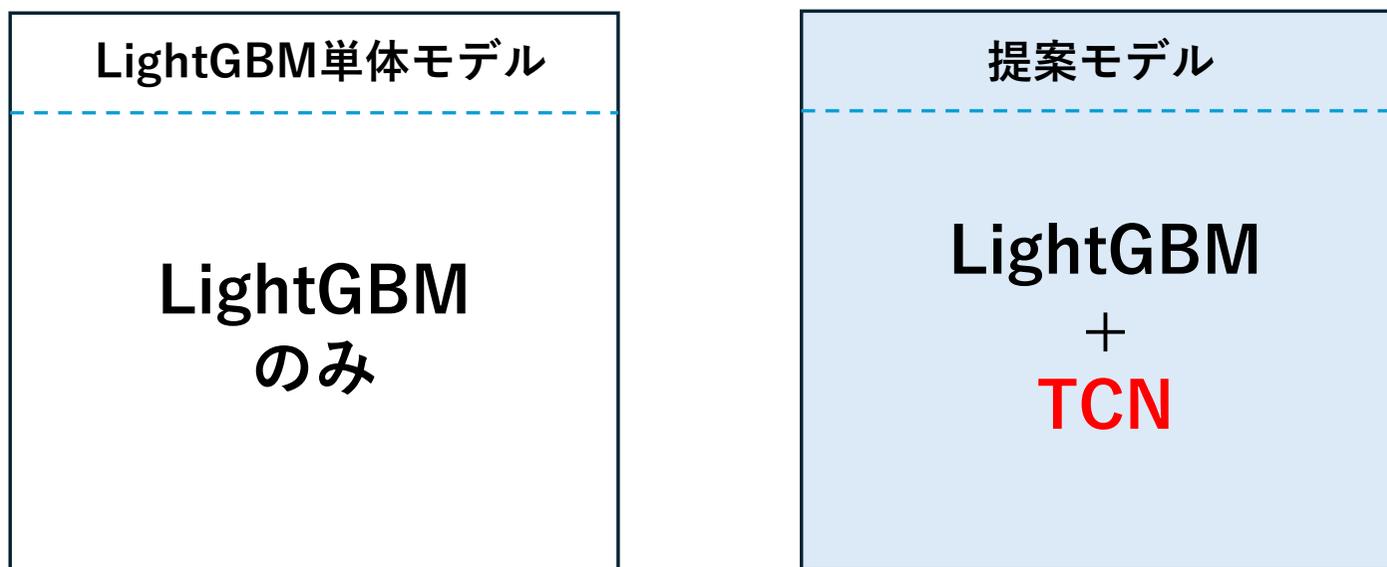
※EMA：指数平滑移動平均（直近の価格を重視）

### 3.提案手法(TCN-LightGBMモデル)



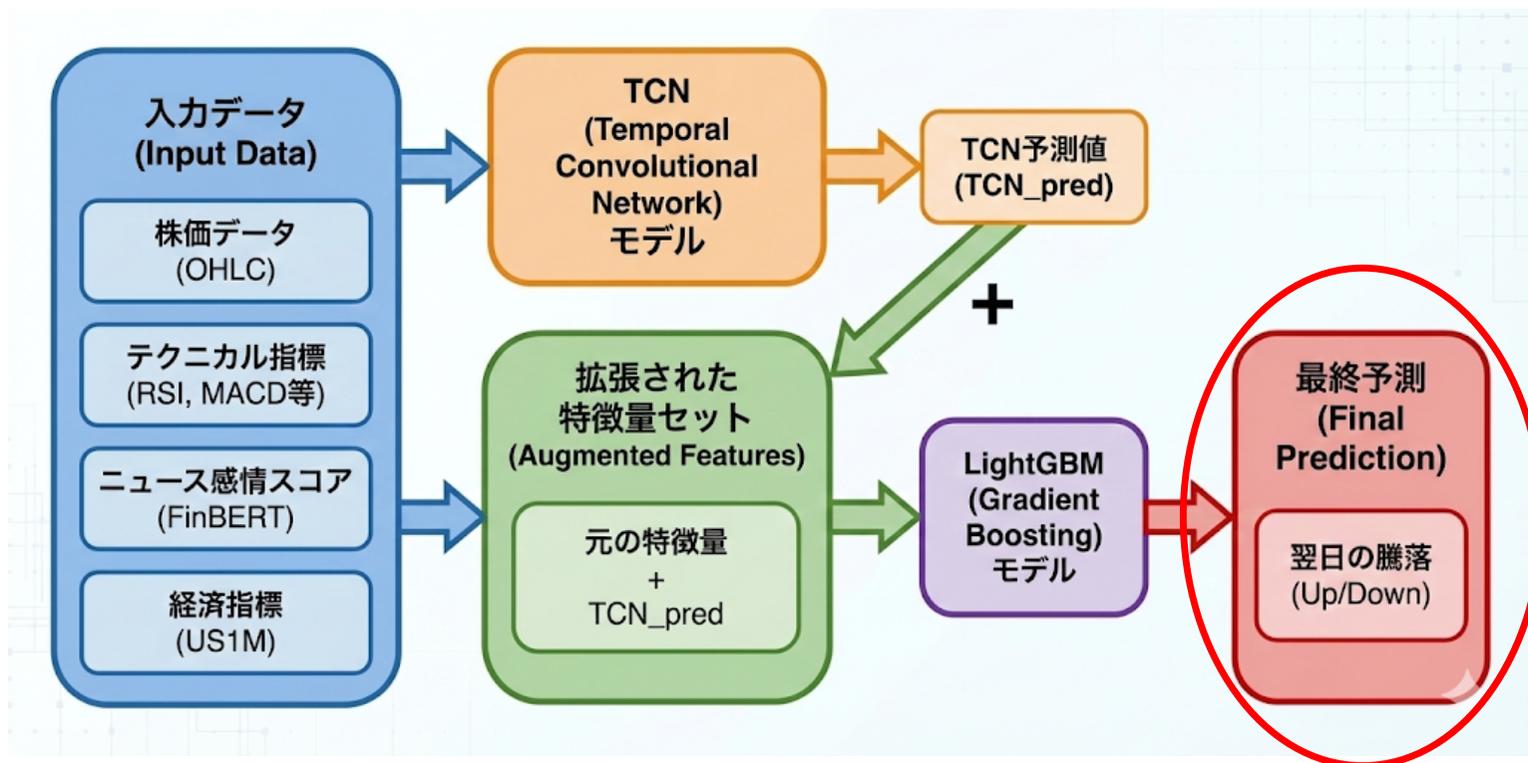
## 4.実験設定

---



どちらのモデルの方が精度が高いか

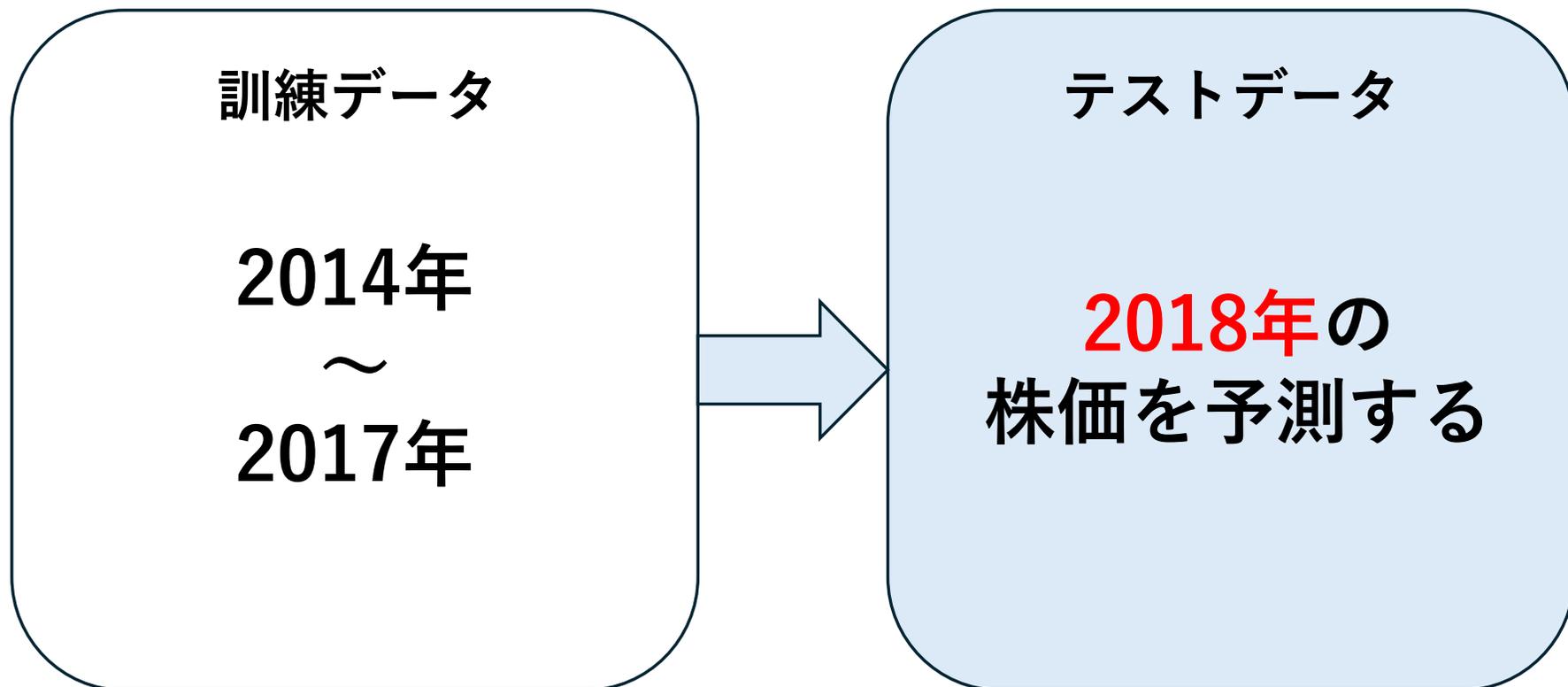
## 4. 評価指標



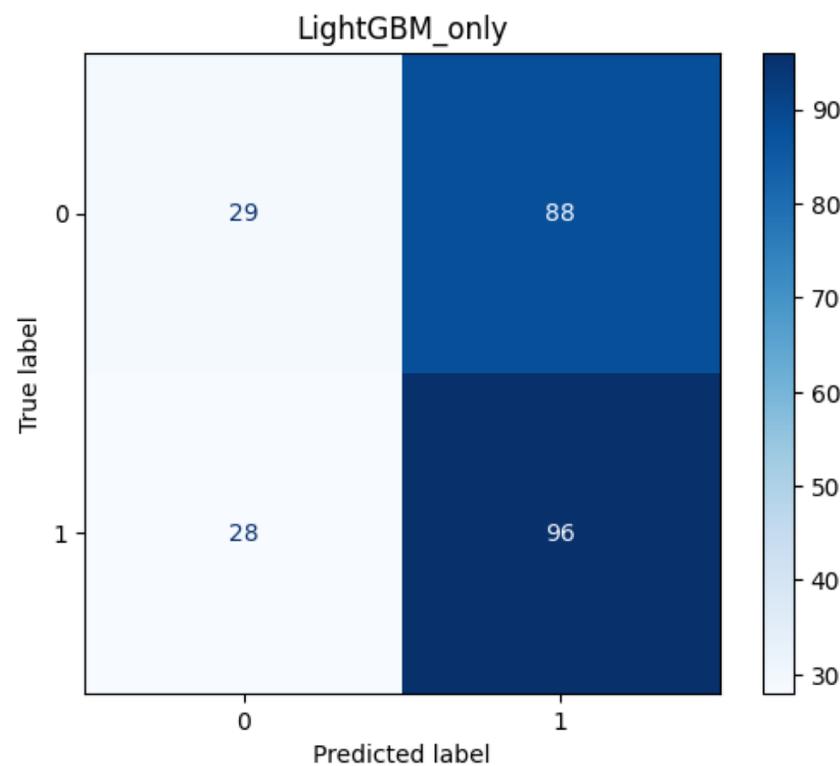
最終予測の**正解率**で判断

## 5.実験1：2018年の株価予測

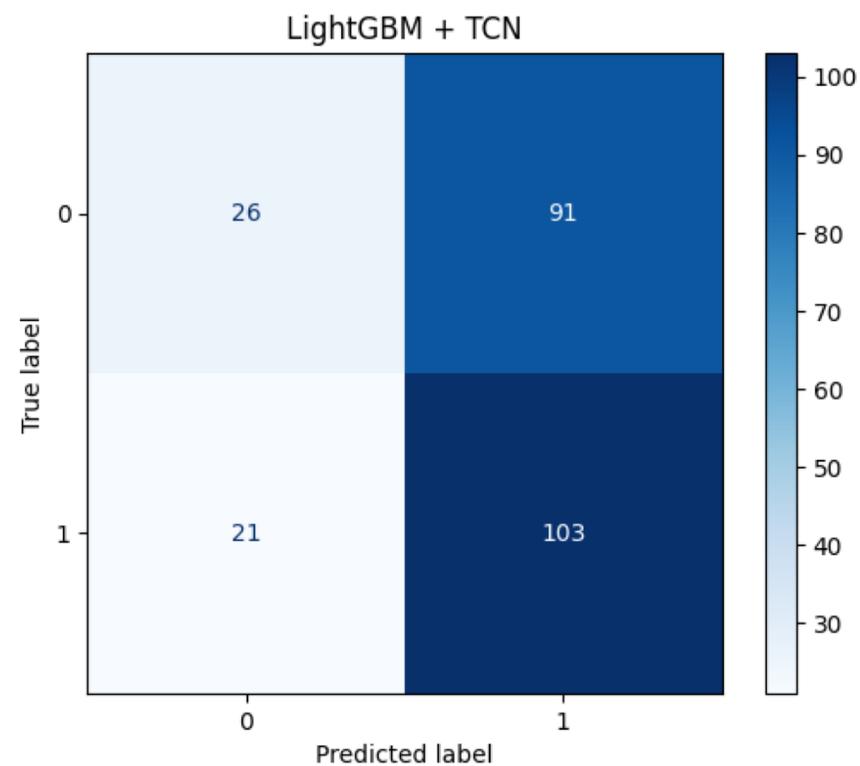
---



# 実験1：2018年の予測結果

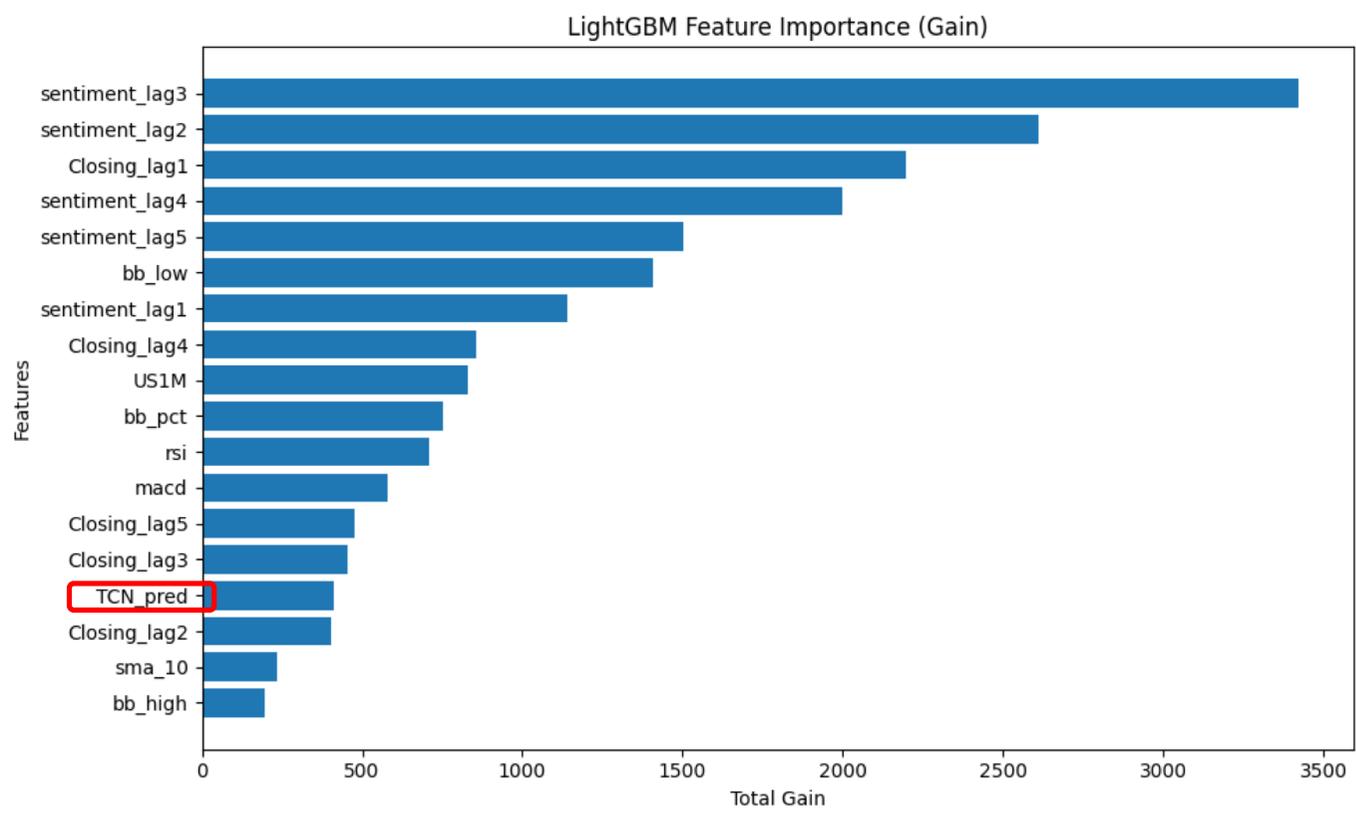


正解率：51.9%



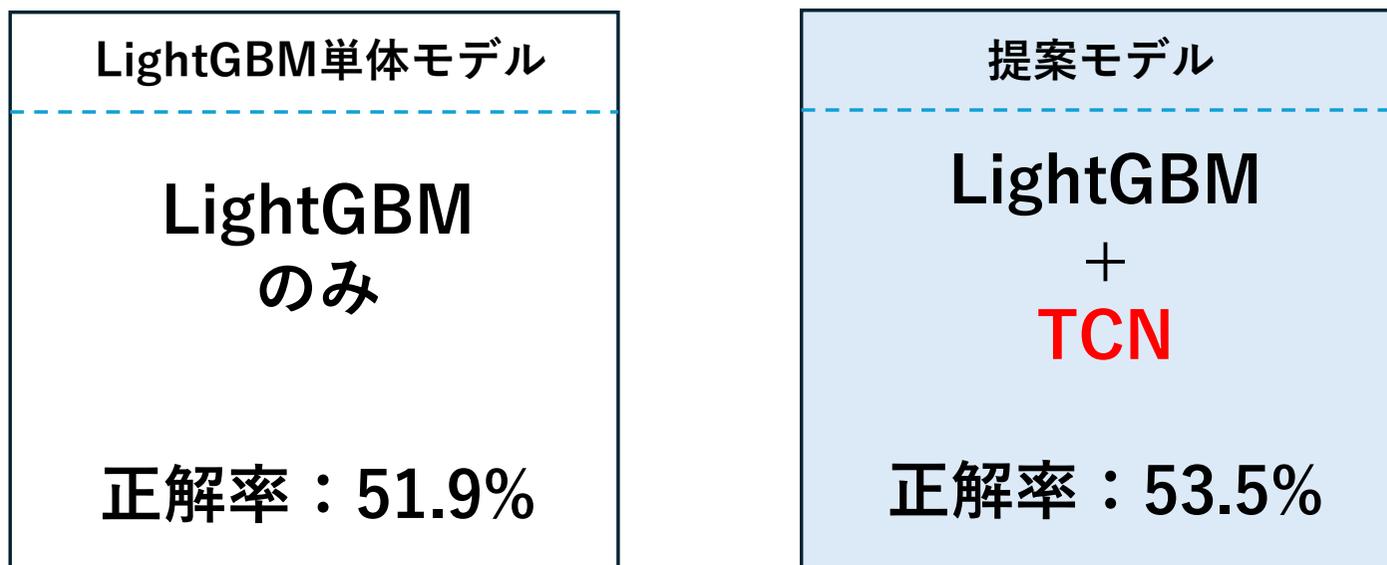
正解率：53.5%

# 実験1：特徴量重要度



18個中**15**番目

## 課題と原因の考察



最大でも**3.5%**の精度改善

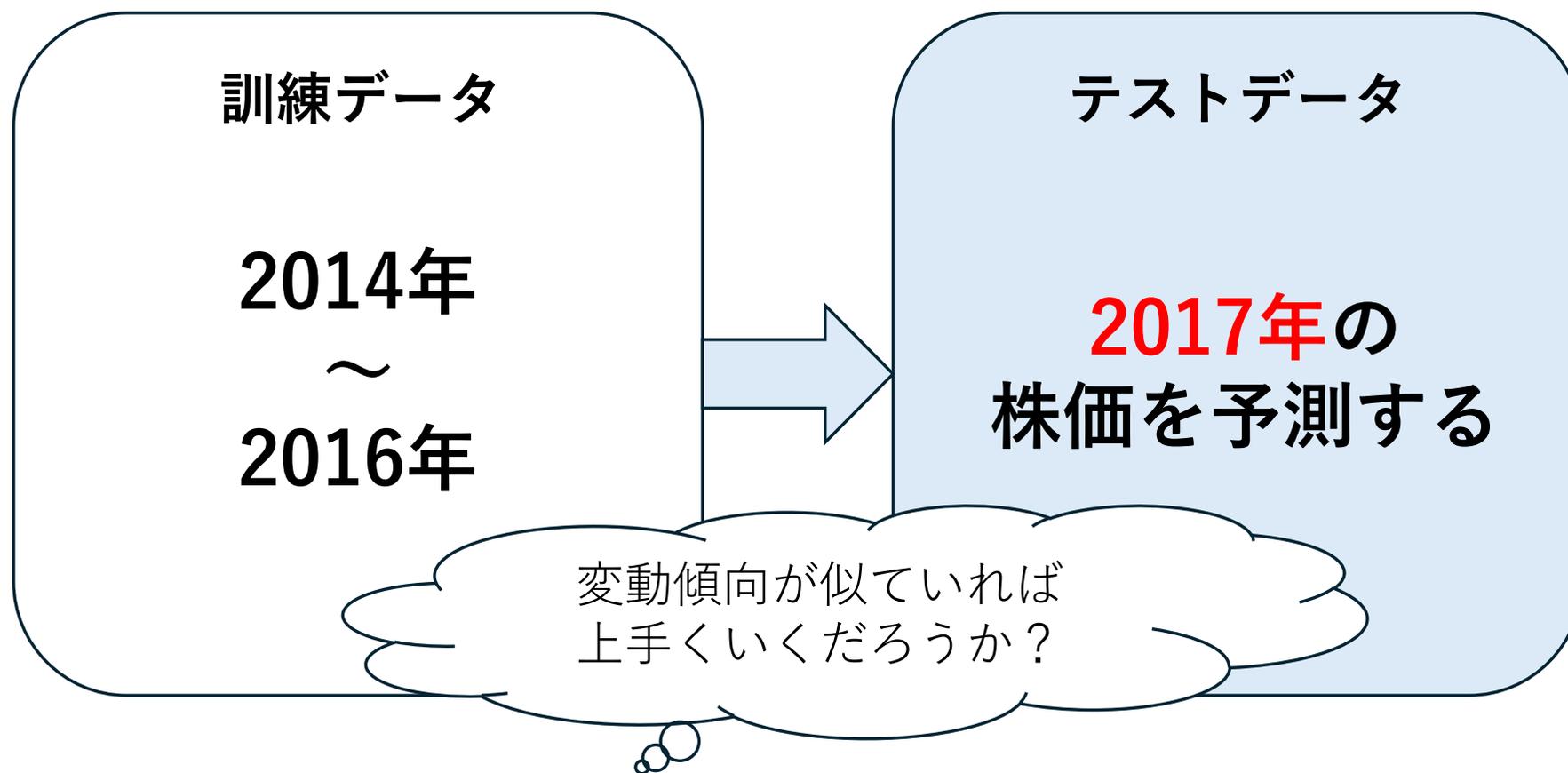


課題：当て推量からの改善がわずかである

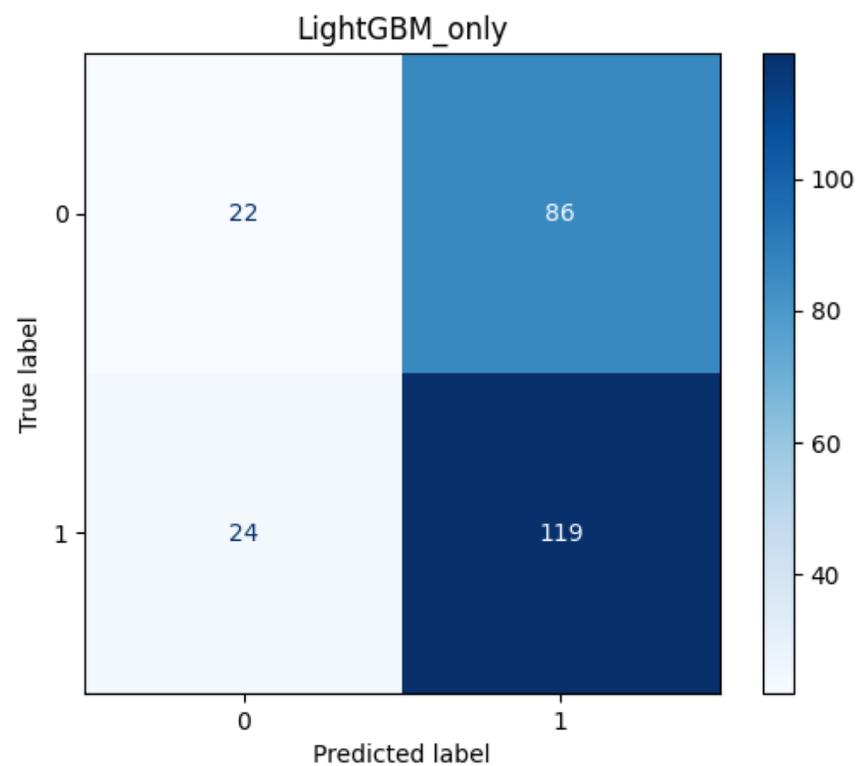


原因の考察：**2018**年の傾向が大幅に変わってしまったため、予測精度が伸び悩んだのではないかと→**2017**年でも検証してみる

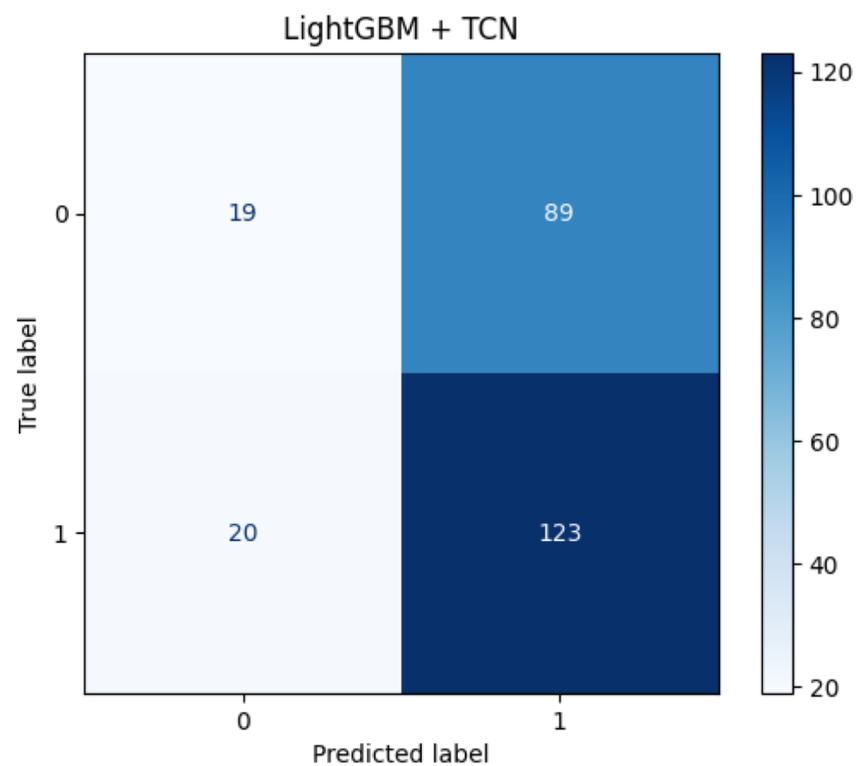
## 実験2：2017年の株価予測



## 実験2：2017年の予測結果

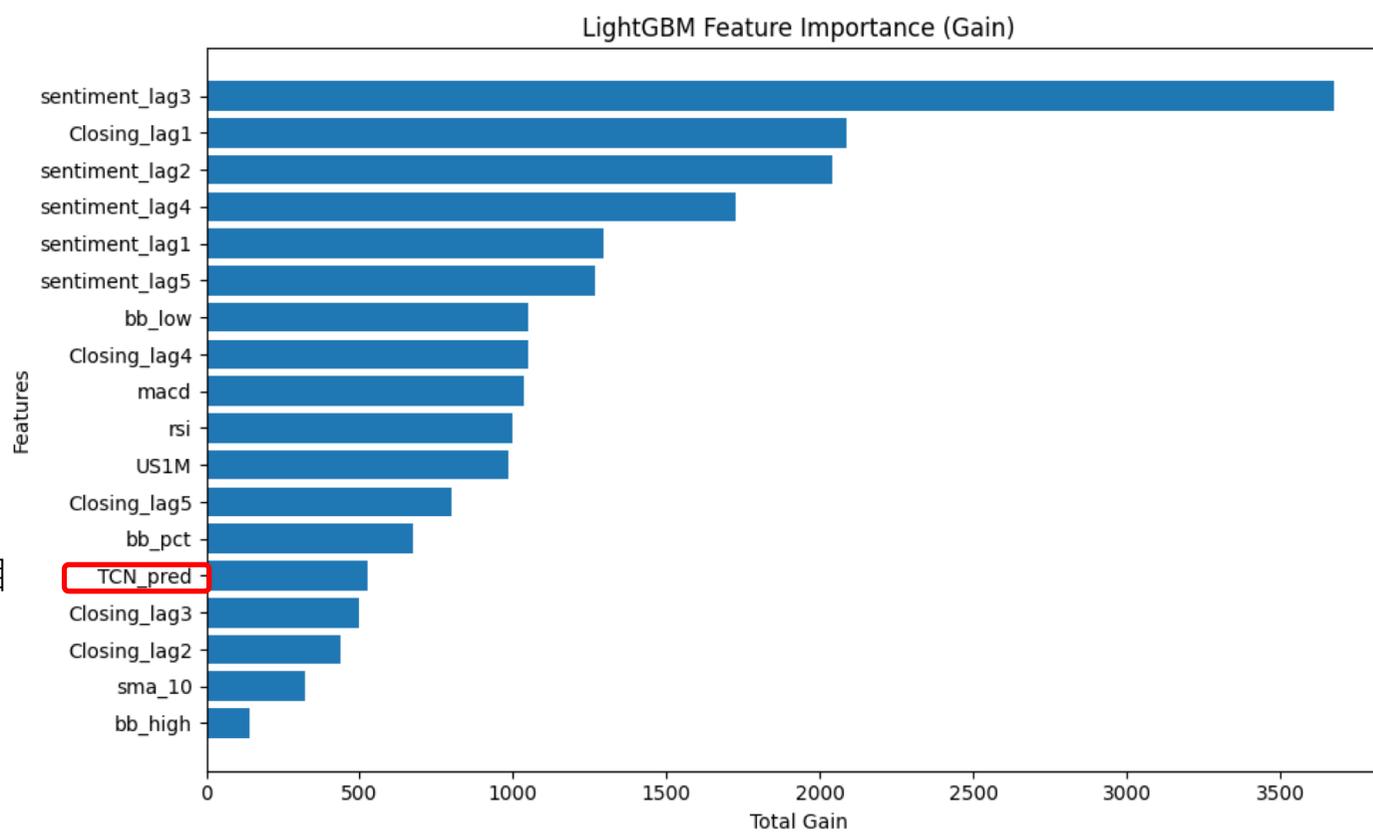


正解率：56.2%



正解率：56.6%

# 実験2：特徴量重要度



18個中**14**番目

## 精度の変動要因の考察

---

- **2017年の方が精度が高かった理由**：訓練データと変動傾向の**類似**
- **モデルドリフトの影響**：レジームスイッチ等による株価傾向の**急変**の影響を受けた
- 対策：**継続的な学習サイクル**の実施が必要

## 6. 今後の課題

---

- **重要度が高く** 精度向上に大きく寄与する新規特徴量の調査
- **回帰的** な株価予測におけるTCN予測値の寄与度の検証
- 最新の市場トレンドへの**リアルタイム** な追従手法の開発

## 7.まとめ

---

- **TCNの予測値**を特徴量に加える提案手法により，精度をやや改善できた
- **2017年**の予測において最大**56.6%**の精度を達成
- **市場環境の変化**に応じたモデルの再訓練の重要性を確認した