

---

# 株価予測プロジェクト

## ProphetとLightGBMを用いた株価予測プロジェクト

データサイエンス × 機械学習

1987年～2024年の株価データによる時系列分析・予測・可視化



# 目次

- 1 プロジェクト概要
- 2 解析データの概要
- 3 技術的アプローチ
- 4 データ前処理
- 5 特徴量エンジニアリング
- 6 Prophetモデルによる予測
- 7 LightGBMモデルによる予測
- 8 予測結果比較 & 可視化
- 9 モデル評価（RMSE比較）
- 10 考察と今後の展望

# プロジェクト概要



本プロジェクトは株価データを用い、ProphetとLightGBMという2種類の時系列モデルで株価予測精度を比較するものです。



## ローリング予測手法

訓練ウィンドウ2年間、予測期間4週間のローリング予測により実用的な評価



## 37年分の長期データ活用

1987年から2024年までの株価データを使用し、様々な市場状況を学習



## 精度評価（RMSE）とグラフ可視化

Root Mean Squared Errorによる客観的評価と視覚的に分かりやすい結果表示



## Pythonによる再現性の高いワークフロー

データ読み込みから予測、評価までの一貫したワークフローを構築

# 解析データの概要

 データセット  
stock\_price.csv

 期間  
1987年2月～2024年8月  
9,202行 × 7列

 主なカラム  
日付、終値、始値、高値、安値、出来高、変化率

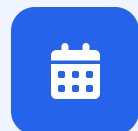
## データの特長

- ❗ 出来高、変化率などの加工が必要な非標準形式データを含む
- 📈 出来高例: 「79.15M」「250.56M」(単位付き)
- 📉 変化率例: 「-2.56%」「1.07%」(%記号付き)
- 🔧 前処理要件: 単位除去・数値変換処理

日付け	終値	始値	高値	安値	出来高	変化率 %
2024-08-01	156.3	159.3	159.4	156.1	79.15M	-2.56%
2024-07-31	160.4	158.2	160.7	158.1	173.91M	1.07%
2024-07-30	158.7	158.8	159.2	158.0	138.14M	-0.63%

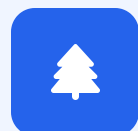
データサンプル (上記3行 / 全9,202行)

# 技術的アプローチ



## Prophet (Facebook製)

季節性やトレンドを分解できる加法モデルベースの時系列予測フレームワーク。非線形トレンド、年次・週次・日次の季節性、および休日効果を自動的に検出・処理。



## LightGBM

高速・高精度な勾配ブースティングフレームワーク。決定木ベースの学習アルゴリズムを使用し、複数の特徴量から目的変数を予測。大規模データにも効率的に対応。



## ローリング予測実装

- 訓練ウィンドウ：2年分のデータで学習
- 予測期間：4週間先の株価を予測
- 予測精度の評価：RMSE (Root Mean Squared Error) で客観的評価

## モデル評価方法



$$\text{RMSE} = \sqrt{(\sum (\text{実績値} - \text{予測値})^2 / n)}$$

値が小さいほど予測精度が高い

# データ前処理と特徴量エンジニアリング



## 日付の変換

文字列形式の日付をdatetimeオブジェクトに変換し、時系列分析の基盤を構築

```
df['日付け'] = pd.to_datetime(df['日付け'])
```



## 出来高の数値変換

接尾辞（M=百万、K=千、B=十億）を適切に処理し、数値形式に変換

```
'79.15M' → 79,150,000 として計算可能な形式に
```



## 変化率の変換

パーセント記号を除去し、文字列を浮動小数点数に変換

```
'-2.56%' → -2.56 として計算に使用
```

## 特徴量エンジニアリング

LightGBMモデルの予測精度向上のために追加した主要な特徴量



### ラグ特徴量

過去1日、5日、10日前の価格情報を特徴量として追加



### 移動平均

5日、10日、20日の移動平均値で短期・中期トレンドを捉える



### ボラティリティ指標

過去10日間の価格変動の標準偏差を計算し市場の不安定さを数値化



### 価格変化率

1日、5日、10日間の価格変化率を計算



### 相対的出来高


過去20日平均との出来高比率で取引活発度を測定



### 時間的特徴

曜日、月、四半期などの時間的情報を追加

# Prophetモデルによる予測

 Facebookが開発した時系列予測ライブラリ。特に季節性のある時系列データに強みを持つ。  
本プロジェクトでは、**終値**を主系列として分析。

## トрендと季節性の分解

時系列データを「トレンド」「季節性」「休日効果」「残差」に分解して予測

## ローリング予測手法

2年分のデータで訓練し、4週間先を予測するウィンドウを移動させながら予測

## パラメータ調整なしのベースライン

最小限のチューニングでベースラインモデルとして構築

### Prophetモデルのワークフロー



株価終値データ



Prophet分解・学習



4週間先予測

# LightGBMモデルによる予測

勾配ブースティングによる高精度予測



## 多変量特徴量

ラグ特徴量、移動平均（5日・20日）、出来高、ボラティリティなど複数の特徴量を活用



## ローリング予測

2年分の訓練データで4週間先をローリング予測し、実用性を重視



## 短期変動キャプチャ

過去の変動パターンと出来高を考慮し、短期的な価格変動も高精度に予測

## LightGBMモデル構成



### 入力特徴量

ラグ特徴量、移動平均、出来高など



### LightGBM

勾配ブースティング決定木



### 予測結果

4週間先の株価予測

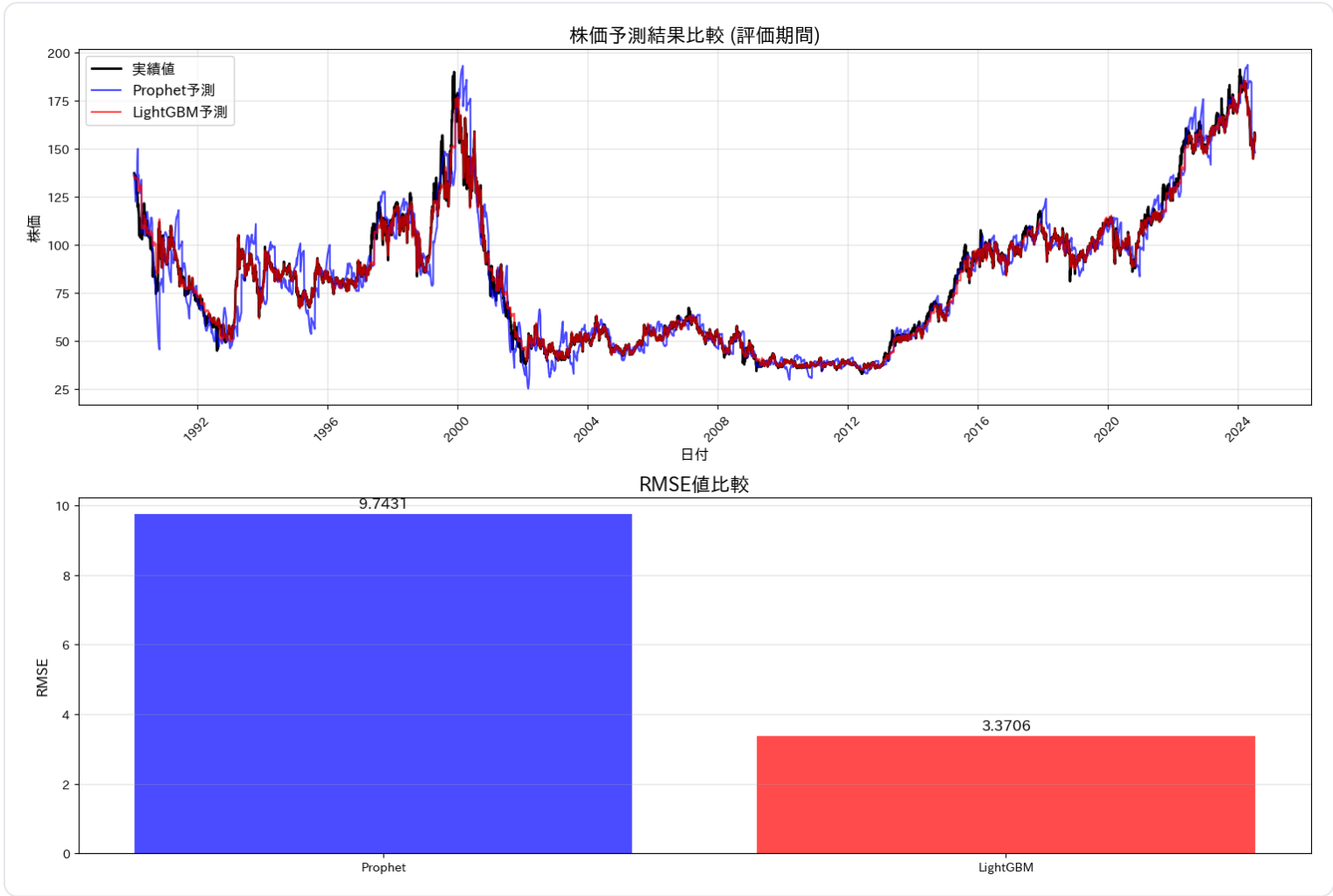


### LightGBMの強み

複雑な非線形パターンの捕捉と高速な学習速度を両立。特徴量間の相互作用も自動的に学習するため、市場データの複雑な関係性も表現可能。



# 予測結果の可視化と比較



## RMSE比較

Prophet  
9.74

LightGBM  
3.37

LightGBMの予測精度が  
約65%高精度

- 実績値
- Prophet予測値
- LightGBM予測値

LightGBMは複雑な非線形パターンを捉え、特に短期的な株価変動の予測に優れた性能を示しています。特徴量エンジニアリングの効果が顕著に表れ、Prophet（統計的時系列モデル）と比べて大幅な精度向上が実現されました。

# まとめ・今後の展望

## プロジェクトの成果

- ✓ LightGBMモデルがProphetと比較して約65%高い予測精度を達成（RMSE: 3.37 vs 9.74）
- ✓ 移動平均、ラグ特徴量などの特徴量エンジニアリングの重要性を実証
- ✓ ローリング予測方式による実用的な予測評価フレームワークの確立

## 今後の展望

### 高度なモデル導入

LSTM、TransformerなどのディープラーニングモデルやSentiment Analysisによるニュース、経済指標などの外部要因を予測に組み込む

### 汎用的予測インフラ構築

様々な銘柄・時間粒度に対応する柔軟な予測システム開発と自動化パイプライン構築

### データ拡充と継続的更新

オープンデータとの連携強化・リアルタイム予測への対応と定期的なモデル再学習機能の実装