

AI予測バイナリーオプション戦略 検証プロジェクト概要

1. 最終目的

当初提供されたドキュメント「素人でも簡単に作れるバイナリーの勝率を劇的に上げるAI予測システム」に記載されている手法、すなわち**「機械学習（AI）を用いて将来価格（5分後のHigh/Low）を予測し、その予測確率がペイアウト率から計算される必要勝率（損益分岐点）を超えた場合にのみエントリーする」という戦略の有効性を、客観的なデータとプロセスに基づいて検証する**こと。

2. 検証対象

- * **予測対象:** BTCUSDT の5分後の終値が現在の終値より高い（High=1）か低い（Low=0）か。
- * **データ:** Bybit API から取得した BTCUSDT 5分足ヒストリカルデータ。
- * **評価基準:**
 - * 機械学習モデルの予測性能（特に **AUCスコア**）。
 - * 損益分岐勝率（ペイアウト率1.8倍の場合、約55.6%）を閾値としたバックテストにおける**勝率**と**累積損益**。

3. これまでの実施内容と主要結果

1. **初期検証（ベースライン特徴量）:**
 - * データ: 約5,000件の5分足データを取得。
 - * 特徴量: ドキュメント記載の基本的な特徴量（終値正規化価格、SMA(14), RSI(14))を計算。
 - * モデル: Logistic Regression を使用。
 - * **結果:**
 - * AUCスコア: **0.5266**（ランダム予測に近い）。
 - * バックテスト: 予測確率がエントリー閾値（0.5556 / 0.4444）を超えず、**エントリー発生回数 0回**。
2. **データ・特徴量・モデルの網羅的検証:**
 - * データ: 約53万件の5分足データ（2020年3月～）を取得。
 - * 特徴量: `pandas-ta` を用いて多数のテクニカル指標、ラグ特徴量、時間特徴量を追加（50種類以上）。
 - * モデル評価（個別）: LightGBM, XGBoost, Random Forest, Logistic Regression を評価。
 - * **結果（AUC）:** LightGBM (0.525), LR (0.524), RF (0.513), XGB (0.511)。**いずれも 0.5 に近く、予測性能の有意な改善は見られず。**
 - * バックテスト (LightGBM): 最もマシだった LightGBM (AUC 0.525) で実行。
 - * **結果:** エントリーは3万回以上発生したが、**勝率 53.68%** で損益分岐点を下回り、**トータルで損失 (-1044.6単位)**。
3. **追加知見に基づく再検討 & AutoML評価:**
 - * 追加ドキュメントから WclPrice, ADR/PowerX指標, AutoML 等のアイデアを得る。
 - * **特徴量セットA (WclPrice):** WclPriceベースの特徴量セットを作成（約53

万件、58列)。

* **AutoML評価 (PyCaret):** 特徴量セットAに対し、PyCaretで多数のモデルを自動評価 (時系列CV使用)。

* **結果:** `compare_models` で評価された**全モデルのAUCが0.54以下**と非常に低く、テストデータでの最良モデル評価 (Ridge) は **AUC 0.500** と完全にランダムレベル。 **WclPrice特徴量による改善は見られず。 **

4. 現在の結論 (暫定)

ここまでの検証では、データ量、特徴量の種類、モデルの種類 (個別評価、AutoML含む) を増やしても、**5分後のBTCUSDT価格のHigh/Lowを有意に予測できるモデルを構築することはできず**、ドキュメントに記載された戦略の有効性を**裏付けることはできなかった**。モデルの予測能力がランダムに近く、バックテストでも利益を出せなかった。

5. 今後の実施予定 (新アプローチ: LightGBM 深化)

AutoML でも改善が見られなかったため、当初の計画に戻り、最も有望視されるモデル **LightGBM** に焦点を当て、さらなる特徴量追加と**ハイパーパラメータチューニング**を行うことで、性能改善の可能性を追求する。

1. **特徴量セットB (ADR/PowerX要素追加) の作成:** (← **イマココ**)
* 特徴量セットA (WclPriceベース、NaN削除前) に、ATR(7) や PowerX関連指標 (Closeベース) を追加し、NaNを削除して `df_processed_b` を作成する。
* (前回コード提示済み)
2. **ハイパーパラメータチューニング (Optuna + 時系列CV):**
* **特徴量セットA (`df_processed_wcl`)** を使用し、LightGBM の最適なハイパーパラメータを `Optuna` と時系列交差検証 (`TimeSeriesSplit`) を使って探索する。
* **特徴量セットB (`df_processed_b`)** を使用し、同様に LightGBM の最適なハイパーパラメータを探索する。
3. **再評価:**
* チューニングされた LightGBM モデルを、各特徴量セットのテストデータで再評価し、AUCスコア等を比較する。
4. **最終バックテスト:**
* 最も性能の良かった組み合わせ (特徴量セット + チューニング済み LightGBM) で、最終的なバックテストを実行し、勝率・損益を評価する。
5. **最終結論:**
* チューニング等を経ても性能が改善しない場合、またはバックテストで利益が出ない場合は、当初の目的である「ドキュメント手法の有効性検証」を完了とし、その結果 (=有効性確認できず) を結論とする。