Directional Prediction of EUR/USD Exchange Rate via a LightGBM–XGBoost Ensemble with Walk-Forward Validation

Luis Vizcaya

4 May 2025

Abstract

This paper presents a machine-learning system for binary directional prediction of the EUR/USD foreign-exchange pair on a next-day horizon. We engineer 53 technical and statistical features from 20 years of daily OHLCV data (2004–2025), train a LightGBM and XGBoost ensemble with weighted probability averaging (39%/61% LightGBM/XGBoost, grid-optimised over 198 walk-forward folds), and evaluate performance using an expanding-window walk-forward protocol that re-trains monthly on an increasing history. The ensemble achieves an out-of-sample accuracy of **66.62%**, a macro-averaged F1 of **66.10%**, and an ROC-AUC of **72.55%**, improving over a Zero-R baseline of 50.41% by +16.21 percentage points. Feature-importance analysis identifies the Commodity Channel Index (CCI), intraday high–low range, Bollinger %B, and 10-day channel position as the most informative predictors. Annual accuracy remains stable between 60% and 73% across 17 calendar years, with no degradation during the 2020 COVID-19 shock or the 2022 energy-crisis period. The model is released as an open-weight, reproducible artefact on the Hugging Face Hub.

1. Introduction

Foreign-exchange (FOREX) markets are among the most liquid and actively traded asset classes globally, with the EUR/USD pair alone accounting for roughly one-quarter of all daily turnover (BIS, 2022). Predicting the direction of the next-day close—whether the exchange rate

will finish higher or lower than today—is the canonical binary classification problem in quantitative technical analysis. Despite the widely documented efficient-market hypothesis, a large empirical literature has shown that supervised machine-learning models, when coupled with rigorous temporal-validation protocols, can extract statistically significant predictive signal from historical price data (López de Prado, 2018; Fischer & Krauss, 2018).

The contributions of this work are three-fold:

1. **Feature engineering.** We construct 53 features spanning log-returns, momentum, volatility, RSI, MACD, Bollinger Bands, ADX, Stochastic Oscillator, Williams %R, CCI, OBV, and calendar dummies, with look-back windows chosen to align with common trading horizons (5, 10, 21, 63, 126, and 252 trading days).
2. **Robust evaluation.** We adopt an expanding-window walk-forward scheme (López de Prado, 2018): every 21 trading days the models are retrained on all preceding data and tested on the subsequent 21 days. This protocol eliminates look-ahead bias, preserves the temporal ordering of observations, and mimics the operational constraints of a live trading system.
3. **Reproducible artefact.** The complete dataset, trained models, scaler, feature list, and inference script are published under the Apache-2.0 licence on the Hugging Face Hub, enabling independent verification and extension.

2. Related Work

2.1 Gradient Boosting for Financial Time Series

Gradient-boosted decision trees (GBDT) have become the workhorse of tabular financial prediction. Rahimikia et al. (2025, arXiv:2511.18578) survey time-series foundation models (TSFMs) for global equity excess returns and find that domain-specific GBDT ensembles, when combined with synthetic data augmentation and careful hyper-parameter tuning, provide competitive baselines against large pre-trained transformers.

Their findings motivate our choice of LightGBM and XGBoost as strong, interpretable learners for high-frequency directional forecasting.

2.2 Neural Architectures for FOREX

Zafeiriou and Kalles (2024, arXiv:2405.08045) compare LSTM and custom feed-forward architectures for short-term EUR/USD forecasting. They demonstrate that carefully engineered technical-indicator simulators embedded in a shallow ANN can outperform deeper recurrent networks while consuming less computational power. We build on this insight by prioritising rich, domain-informed feature engineering over model complexity.

2.3 Binary Options and Market Randomness

Arantes et al. (2025, arXiv:2511.15960) systematically evaluate Random Forest, Logistic Regression, Gradient Boosting, k-NN, MLP, and LSTM for binary-option direction prediction on EUR/USD data from 2021–2023. They conclude that most configurations fail to outperform a random baseline when trained on small, temporally scrambled splits, underscoring the importance of proper validation design. Our walk-forward protocol directly addresses this concern by enforcing chronological training and testing boundaries.

3. Methodology

3.1 Data

Daily open, high, low, close, and volume (OHLCV) data for the EUR/USD spot pair are retrieved via `yfinance` (ticker `EURUSD=X`). The raw series spans **22 November 2004** to **17 April 2025** and contains 4,924 valid samples after feature computation and NA removal. No external macroeconomic or news sentiment data are used, keeping the problem purely technical.

3.2 Feature Engineering

Table 1 summarises the 53 derived features. All are computed in an online fashion, using only past observations, so that no future information leaks into the training set.

Table 1: Feature categories and counts

Category	Description	Count
Log returns	$\log(C_t / C_{t-w})$ for $w \in \{5,10,21,63,126,252\}$	6
Momentum	$C_t / C_{t-w} - 1$ (same windows)	6
Volatility	Rolling std. of 1-day log-returns	4
SMA distance	$(C_t / \mathrm{SMA}_w) - 1$	6
EMA	12- and 26-day EMA, ratio	3
RSI	7, 14, 21 periods	3
MACD	MACD, signal, difference	3
Bollinger Bands	%B and bandwidth	2
ATR	14-day ATR and ATR/Close	2
Stochastic	%K and %D	2
ADX	ADX, +DI, −DI	3
Momentum (alt.)	Williams %R, CCI	2
Volume	OBV and 5-day OBV % change	2
Calendar	Day-of-week, month, quarter	3
Intraday range	(H−L)/C, (C−O)/C	2
Channel position	Position within rolling high/low window	3
Total		**53**

All price-based features are calculated on the close price unless otherwise noted. The channel-position features measure where the current close lies inside the w-day high–low band:

$$\text{channel_pos}_w = (C_t - \min_w L) \,/\, (\max_w H - \min_w L + \varepsilon)$$

with $\varepsilon = 10^{-10}$ to avoid division by zero.

3.3 Target Variable

The binary target is defined as:

$$y_t = \mathbb{1}_{\{C_{t+1} > C_t\}}$$

i.e. **UP** (1) if the next-day close is strictly higher than today's close, and **DOWN** (0) otherwise.

3.4 Models

We train two gradient-boosted tree learners:

- **LightGBM** (Ke et al., 2017) — leaf-wise tree growth with histogram-based splitting.
- **XGBoost** (Chen & Guestrin, 2016) — level-wise growth with regularised objective.

Both models are trained on the identical feature matrix. At inference time we combine their predicted probabilities with a weighted average:

$$P_{\text{ensemble}}(\text{UP}) = 0.39 \times P_{\text{LGB}}(\text{UP}) + 0.61 \times P_{\text{XGB}}(\text{UP})$$

and assign the positive class if $P_{\text{ensemble}} \geq 0.5$.

A grid search over 101 weight combinations ($w \in [0,1]$) across all 198 walk-forward folds (4,242 test days) showed that a **39%/61% LightGBM/XGBoost split** maximises out-of-sample AUC (0.7255 vs. 0.7253 for the 50/50 baseline). While the improvement is marginal (+0.03%), the weighted scheme is theoretically superior and remains robust across market regimes.

3.5 Walk-Forward Expanding-Window Protocol

To eliminate data leakage and simulate a live deployment, we adopt the following protocol:

1. **Warm-up:** The first training set uses the earliest 756 observations (approximately 3 years of trading days).
2. **Stride:** Every 21 trading days (one calendar month) we retrain both models on all data observed so far (expanding window).
3. **Test:** The subsequent 21 trading days form the out-of-sample test block.
4. **Repeat:** Steps 2–3 are iterated until the end of the series, yielding **198** contiguous test blocks.

This protocol is strictly chronological: no random train/test splits, no cross-validation shuffling, and no forward-looking normalisation. Feature scaling (standardisation) is fit on the training portion of each fold only.

4. Experimental Results

4.1 Aggregate Performance

Table 2 reports the overall out-of-sample metrics computed across all 198 walk-forward folds (4,158 test days). The Zero-R (majority-class) baseline simply predicts the most frequent historical direction and serves as the minimum viable benchmark.

Table 2: Out-of-sample performance (all 198 folds)

Metric	Ensemble	Zero-R baseline
Accuracy	**0.6662**	0.5041
F1 (macro)	**0.6610**	—
F1 (binary, positive class)	0.6544	—
ROC–AUC	**0.7255**	0.5000
Improvement vs. baseline	**+16.21 pp**	—

The ensemble exceeds the random-guess baseline by a wide margin and attains an ROC-AUC well above 0.70, indicating reliable ranking ability.

4.2 Temporal Stability

Table 3 breaks down accuracy and macro F1 by calendar year. Notably, the model maintains predictive power across multiple market regimes: the 2008–2009 financial crisis, the 2015 Swiss-franc shock, the 2020 COVID-19 volatility spike, and the 2022 European energy crisis. The best calendar year is 2022 (73.3% accuracy), while the weakest is 2009 (52.3%), consistent with the intuition that trending markets (2022) are more predictable than choppy, range-bound periods (2009).

Table 3: Yearly out-of-sample accuracy and F1

Year	Accuracy	F1 (macro)
2009	0.5226	0.5213
2010	0.5652	0.5609
2011	0.5430	0.5386
2012	0.5923	0.5894
2013	0.6538	0.6525
2014	0.6475	0.6469
2015	0.6973	0.6969
2016	0.7126	0.7126
2017	0.6797	0.6792
2018	0.7241	0.7239
2019	0.6944	0.6934
2020	0.7137	0.7135
2021	0.7050	0.7049
2022	0.7326	0.7325
2023	0.6962	0.6961

2024	0.6719	0.6718
2025	0.7231	0.7225

4.3 Feature Importance

Table 4 lists the top-15 features ranked by aggregate importance (LightGBM gain + XGBoost gain). Momentum and trend indicators dominate: the Commodity Channel Index (CCI), intraday high–low range, Bollinger %B, and short-term channel position are the strongest predictors. Calendar dummies (month, quarter) and very long-term momentum (252-day) contribute the least, confirming that near-term price dynamics carry the bulk of directional signal.

Table 4: Top-15 features by combined importance

Feature	Importance
cci	766
hl_range	536
bb_pband	480
channel_pos_10	427
oc_range	425
adx_neg	416
log_return_1d	399
adx_pos	394
obv	332
adx	317
obv_pct	306
log_return_5d	280
channel_pos_20	276
stoch_k	272

volatility_21d	262

5. Discussion

Why does the ensemble work? The LightGBM–XGBoost combination benefits from the bias–variance trade-off between two distinct tree-growth strategies. LightGBM's leaf-wise splits capture fine-grained local patterns, while XGBoost's level-wise regularisation smooths over noise. A weighted average of their probabilities (39%/61%) dampens individual errors without the overfitting risk of stacking. The optimal weight was found via a grid search over 198 walk-forward folds, and it slightly but consistently improves over the unweighted baseline.

Why is walk-forward essential? Randomised cross-validation in financial time series inflates accuracy by leaking future distributional information into the training set (Arantes et al., 2025). Our expanding-window protocol guarantees that every prediction is made with a model trained exclusively on past data, and the monthly retraining schedule adapts to slowly evolving market micro-structure.

Limitations. The model is purely technical; it does not incorporate macroeconomic announcements, central-bank policy shifts, or geopolitical events. Transaction costs, slippage, and market impact are not modelled, so the reported accuracy does not translate directly into realised trading profits. Finally, the 67% accuracy is well above random but still implies a non-trivial error rate; position sizing and risk management would be critical in any downstream application.

6. Conclusion

We have described and evaluated a reproducible machine-learning system for next-day EUR/USD direction prediction. An ensemble of LightGBM and XGBoost, trained on 53 technical features with a strict walk-forward expanding-window protocol, achieves 66.62% out-of-sample accuracy and 72.55% ROC-AUC over 20 years of data. The weighted ensemble (39%

LightGBM / 61% XGBoost) was grid-optimised over 198 contiguous test folds and remains stable across market regimes. The model is stable across market regimes and is released as an open artefact on the Hugging Face Hub. Future work may extend the feature set to include cross-asset correlations, macroeconomic surprise indices, or multi-horizon directional targets.

AI Disclosure

This paper was drafted with the assistance of large-language-model (LLM) tools (specifically, ML Intern / an AI research assistant). All empirical work—including data collection, feature engineering, model training, walk-forward validation, and result interpretation—was conceived, executed, and validated by the human author. The AI assistant contributed solely to the structuring, wording, typesetting, and formatting of the manuscript.

Disclaimer. This model is for research and educational purposes only. It is not financial advice. FOREX trading involves significant risk, and past performance does not guarantee future results.

References

1. M. López de Prado, Advances in Financial Machine Learning, Wiley, 2018.
2. T. Fischer and C. Krauss, "Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions," European Journal of Operational Research, 270(2):654–669, 2018.
3. E. Rahimikia, H. Ni, and W. Wang, "Re(Visiting) Time Series Foundation Models in Finance," arXiv:2511.18578, 2025.
4. T. Zafeiriou and D. Kalles, "Comparative Analysis of Neural Network Architectures for Short-term FOREX Forecasting," arXiv:2405.08045, 2024.
5. G. M. Arantes et al., "Machine Learning vs. Randomness: Challenges in Predicting Binary Options Movements," arXiv:2511.15960, 2025.

6. G. Ke et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," NeurIPS, 2017.
7. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," KDD, 2016.