European Journal of Business and Management www.iiste.org
ISSN 2222-1905 (Paper) ISSN 2222-2839 (Online)
Vol.17, No.4, 2025

Machine Learning Forecasting of EUR/USD Trends with Qualitative Cross-Validation from Institutional Reports

Mohamed Adil Khalifa
* E-mail of the corresponding author: adilkhalifa.ak@gmail.com

Abstract

This study presents a comparative methodology for forecasting weekly trends in the EUR/USD exchange rate using daily financial data. A first model is calibrated using a Classification and Regression Tree (CART), followed by an enhanced version based on the Random Forest algorithm. The analysis evaluates statistical performance, model robustness, interpretability, and decision-making quality through systematic backtesting.

The machine learning models are trained to classify weekly currency movements into two categories: +1 for an expected rise and -1 for an expected decline. A prediction is made every 5 days to anticipate the market direction for the subsequent 5-day period, using only information available at the time of the prediction (ensuring strict causality).

To complement and validate the quantitative results, a qualitative layer is introduced using a large language model (LLM) to extract directional sentiment from institutional FX research reports. This dual approach enhances the interpretability and contextual relevance of the forecasting framework.

Keywords: EUR/USD, machine learning, Random Forest, CART, forecasting, large language models, sentiment analysis, time series

DOI: 10.7176/EJBM/17-4-06

Publication date: May 30th 2025

1. Introduction

Supervised learning models (James et al. 2013) for predicting financial market trends are attracting increasing attention. Currency markets are particularly complex, influenced by a mix of technical, fundamental, and behavioral factors, which complicates the modeling of short-term price dynamics (Fama 1970). Financial time series such as exchange rates typically display high noise, regime shifts, and persistent volatility. (Tsay 2010)

In this context, the present study aims to evaluate the effectiveness of a decision tree model (CART) as a baseline for forecasting weekly EUR/USD movements, and to compare it to a more robust ensemble method: the Random Forest algorithm (Hastie et al. 2009). These models are trained to generate directional trading signals based on technical indicators derived from daily data.

To enrich the predictive framework and enhance model interpretability, the study also incorporates a qualitative validation layer using a large language model (LLM). Institutional FX research reports are analyzed semantically to extract directional sentiment, allowing for a cross-validation of model outputs with expert market views.

The core objective is to determine whether these relatively simple machine learning techniques, when complemented by LLM-based sentiment analysis, can generate signals that are not only statistically sound but also economically actionable within a systematic trading context.

2. Machine Learning Approaches: CART and Random Forest

2.1 Data and Preparation

The dataset spans the period from 1990 to 2025, with observations recorded at daily frequency. The target variable is the weekly directional movement of the EUR/USD exchange rate, derived from daily returns.

The explanatory variables include technical indicators such as the Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), Bollinger Bands, moving averages, and a volatility measure. These indicators were computed over short-, medium-, and long-term horizons to capture multiple dimensions of market behavior. Such indicators are widely used in the modeling of financial time series, especially due to their ability to reflect trend strength and momentum (Tsay, 2010).

External variables like the DXY index, VIX, and Brent crude prices were initially considered but excluded due to low correlation and statistically insignificant p-values (see Figures 1 and 2).

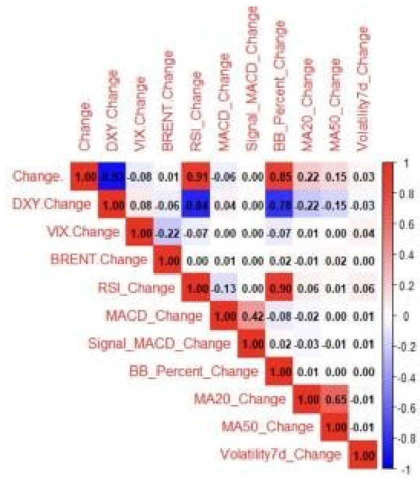

Figure 1: Correlation heatmap between EUR/USD daily returns and 5-day lagged explanatory variables.

	Change.	DXY.Change	VIX.Change	BRENT.C
Change.	0.0000000	0.000000000	1.233661e-01	8.8995
DXY.Change	0.0000000	0.000000000	7.038463e-02	7.4626
VIX.Change	0.1233661	0.070384627	0.000000e+00	3.4159
BRENT.Change	0.8899592	0.007462643	3.415917e-30	0.0000

Figure 2: Corresponding correlation p-values.

All explanatory variables were lagged by five days to reflect only the information available at the time of prediction and to respect the principle of strict causality. The training set consisted of the first 6,000 observations out of a total of 9,188, with the remaining data reserved for out-of-sample testing (Figure 3).

This approach ensures temporal consistency and prevents data leakage between training and validation periods. The use of out-of-sample evaluation is essential for assessing the robustness and generalizability of forecasting models (Tashman, 2000).

Figure 3. EUR/USD exchange rate since January 2, 1990

2.2. Results of the CART Model

The CART model (Classification and Regression Tree) is based on recursive partitioning of the data: at each node, the algorithm identifies the explanatory variable and threshold condition that best separates the target classes. This model is especially valued for its interpretability.

In our case, the model aims to forecast the weekly direction of EUR/USD by classifying observations as +1 (uptrend) or -1 (downtrend).

Table 1.CART model parameters

Parameter	Value	Role
maxdepth	5	Limits the tree depth to avoid overfitting
minsplit	17	Minimum number of observations required to split a node
cp	0.015	Complexity threshold: prevents insignificant splits
weights	Yes	Observations are weighted according to the magnitude of future returns

2.2.1 Interpretation of the Decision Tree

The resulting decision tree has 5 levels, each representing a logical split based on thresholds from technical indicators (e.g., “MACD > 0.2” or “RSI < 35”). (Figure 4)

Each path from the root to a leaf node can be interpreted as a conditional trading rule. Example:

“If volatility exceeds 0.015 AND RSI is below 35 → then predict a decline.”

This structure provides transparent decision logic for users, unlike more opaque models.

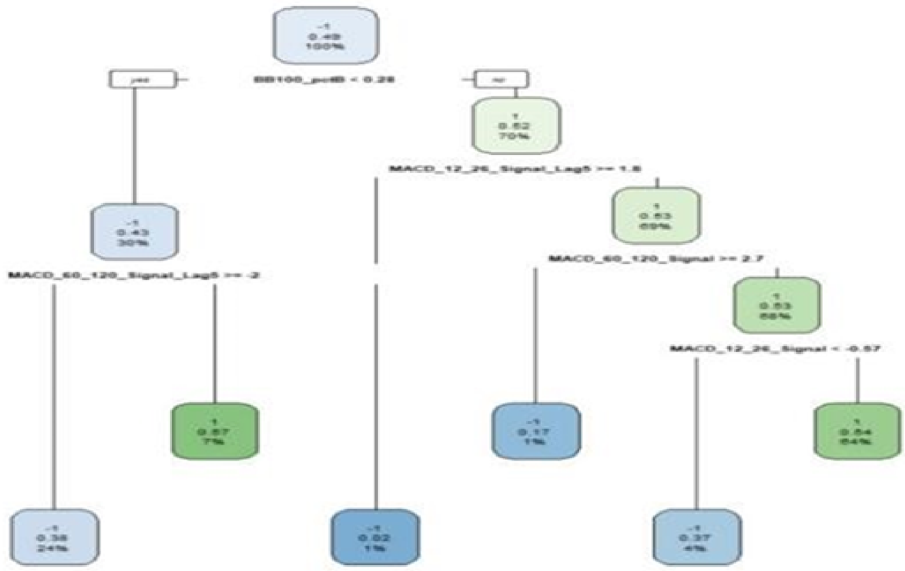

Figure 4.CART Decision Tree

2.2.2 Performance of the CART Model

The model yielded an AUC score of 0.495, which is effectively equivalent to that of a random classifier. This indicates that the model was unable to meaningfully discriminate between the two target classes. As shown in Figure 5, the ROC curve lies very close to the diagonal, further confirming the lack of predictive separation.

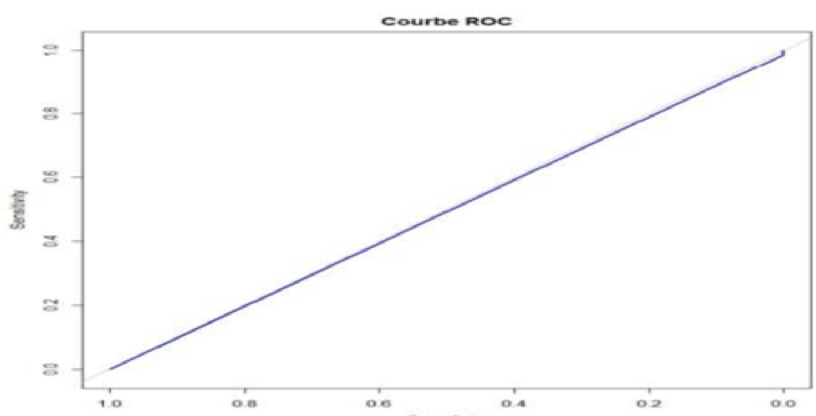

Figure 5.ROC curve for CART model predictions

The Sharpe Ratio of the portfolio: 0.25 indicates moderate but positive risk-adjusted return, whereas on the equity curve slightly upward trending (Figure 6) The simulated capital grows slowly over time, indicating that the signals, though weak, are sufficiently consistent.

Figure 6. Equity curve of the CART model

2.2.3 Remarks

The CART model (Breiman et al., 1984) shows limited predictive performance, but its structure is still usable in systematic trading.

Its main strength lies in producing explicit, easily testable rules that are understandable even for non-technical users.

Despite weak classification performance (AUC ≈ 0.5), the model can still offer economic value by extracting stable and repeatable rules.

2.3 Results of the Random Forest Model

The Random Forests model, introduced by (Breiman 2001) improves upon the CART approach by generating many independent decision trees, each constructed from a random subset of the data and variables.
Final predictions are made through majority voting (for classification) or averaging (for regression).

This ensemble technique helps reduce model variance (Hastie et al. 2009) while maintaining low bias, which typically results in better robustness on unseen data.

Table2.RF model parameters

Parameter	Value	Role / Justification
num.trees	500	Large number of trees to ensure model stability
mtry	√p	Number of variables considered at each split (standard rule)
min.node.size	10	Minimum leaf size to avoid overfitting
sample.fraction	0.7	Partial bootstrap sample for each tree
importance	impurity	Importance calculated based on impurity reduction (Gini index)
probability	TRUE	Enables probability estimates for ROC analysis
case.weights	Based on future return	Increases the impact of high-variation cases

2.3.1 Interpretation of Feature Importance

The analysis of variable importance (based on impurity reduction) shows that:

RSI, volatility, and Bollinger Bands are among the most influential variables, (Figure 7)

This validates the hypothesis that short- and medium-term momentum indicators are particularly relevant for predicting weekly EUR/USD movements.

This feature importance analysis also facilitates dimensionality reduction in future, more complex models.

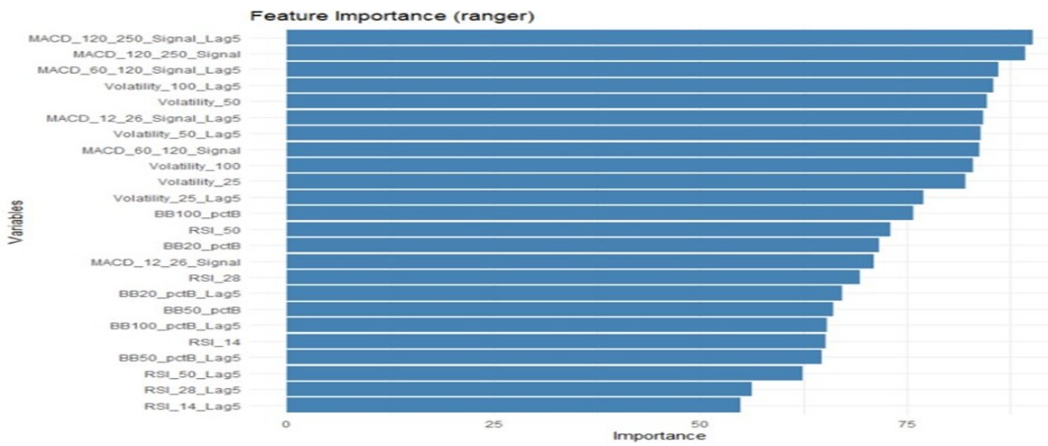

Figure 7.Variable importance – Random Forest model

2.3.2 Observed Performance

The Random Forest model achieved an AUC of 0.515, representing a marginal but genuine improvement over the CART model. The corresponding ROC curve lies slightly above the diagonal, indicating modest class separability (Figure 8).

In terms of trading performance, the equity curve was clearly upward trending and exhibited greater regularity compared to that of the CART-based strategy (Figure 8). The strategy produced a Sharpe ratio of 0.3075, with an annualized return of 1.09% and an annualized volatility of 3.56%, reflecting a more stable and economically viable signal profile.

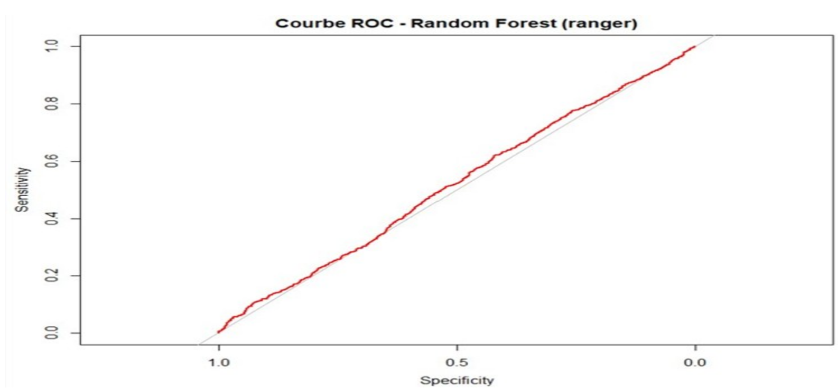

Figure 8. ROC curve – Random Forest

Despite still limited discriminative power, the model generates coherent and exploitable signals within a systematic strategy framework.

Accuracy: 50.9%

Precision (class +1): 47.5%

Recall (class +1): 35.2%

F1-score: 0.405

These figures indicate that the model struggles with trend reversals but captures strong directional moves more effectively thanks to the weighting applied. The relatively low recall reflects many false negatives, but the model remains useful in a disciplined quantitative context.

Table 3. RF Confusion matrix

Actual \ Predicted	-1	1
-1	975	525
1	875	475

2.3.3 Final Comment

The Random Forest model enables gradual performance gains (Figure 9) without compromising the balance of the simulated portfolio. Compared to CART, its main strength lies in noise resilience and generalization capability across different market configurations. The loss of interpretability is a reasonable price to pay for a net gain in robustness.

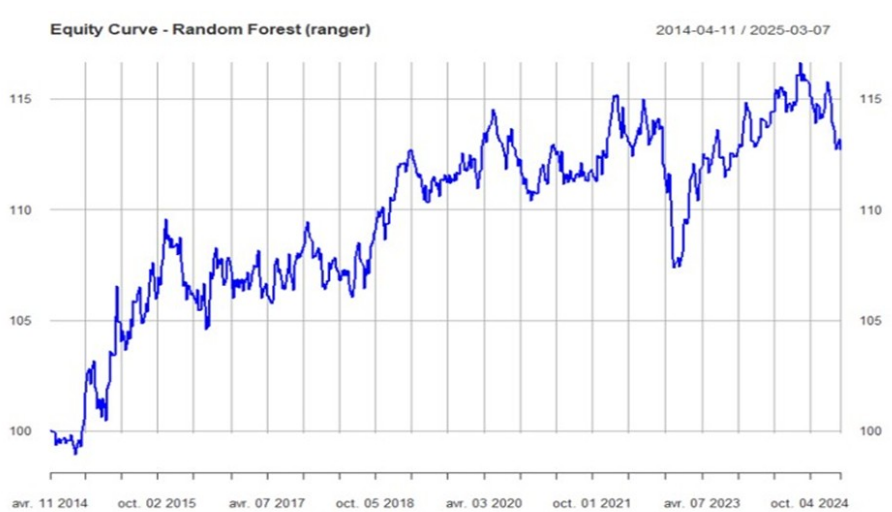

Figure 9. Equity curve – Random Forest model

2.4 Comparative Discussion

As shown in Table 4, the Random Forest model outperforms CART in terms of both stability and risk-adjusted returns. While CART offers clear decision rules and strong interpretability, it suffers from high variance and limited predictive accuracy in noisy environments. Random Forest, by aggregating multiple decision trees, mitigates overfitting and provides more reliable signals — albeit at the cost of transparency.

European Journal of Business and Management
ISSN 2222-1905 (Paper) ISSN 2222-2839 (Online)
Vol.17, No.4, 2025

The optimal model selection ultimately depends on the user's priority: if the goal is explainability and ease of audit (e.g., in regulatory or operational contexts), CART may be preferred. Conversely, if performance and robustness are critical, Random Forest presents a more effective option (Krauss et al., 2017).

Table4. Comparison of CART and RF results

Criterion	CART	Random Forest
AUC	0.495	0.515
Sharpe Ratio	0.25	0.3075
Accuracy	~50%	50.90%
Interpretability	High	Medium
Robustness	Low	High

2.5 Conclusion

Although neither CART nor Random Forest achieves strong class discrimination in pure statistical terms, both models generate signals that translate into economically consistent strategies when applied in a disciplined trading framework (De Prado, 2018). The Random Forest model demonstrates greater robustness, a smoother equity curve, and a superior Sharpe ratio — suggesting a more viable basis for systematic forecasting.

Several improvements could enhance the predictive power of this framework. First, incorporating additional explanatory variables such as interest rate differentials, macroeconomic indicators, or positioning data may improve model sensitivity. Second, implementing rolling retraining can adapt the model to regime changes over time. Third, experimenting with more advanced architectures — such as Support Vector Machines (SVM), Recurrent Neural Networks (RNN), or hybrid LSTM-RF models — could further improve directional accuracy.

3.Analysis of Financial Institutions Directional Biases via LLM

Finally, the integration of Large Language Models (LLMs) to validate model outputs using expert-generated sentiment represents a promising complement to traditional validation approaches and opens new paths for model enhancement. To complement and validate the results produced by the CART and Random Forest models on the EUR/USD trend, a qualitative analysis was conducted based on the most recent FX research reports published weekly by international financial institutions (Goldman Sachs, HSBC, BlackRock).

These documents were processed automatically using a Large Language Model (LLM) specifically, Mistral-3.1 (24B) to extract:

The explicit directional sentiment (Long / Short / Neutral) regarding EUR/USD

The expressed confidence level (on a scale from 1 to 5)

The date of the report

Table 5: LLM market sentiment output

Bank	Sentiment	Confidence	Date
Goldman Sachs	Neutral	3	11/03/2025
BlackRock	Neutral	2	21/03/2025
HSBC	Short	4	11/03/2025

3.1.Methodology

PDF reports were manually collected and converted into plain text using a local parser (pdftools). This content was then input into the LLM with a structured prompt designed to produce a concise JSON-style summary, extracting each bank's position on EUR/USD and its associated confidence level.

This method enables:

A comparison between human expert judgment and that of machine learning models (Luss et d'Aspremont 2015)

The detection of consensus or divergence signals, which may either reinforce or weaken confidence in the model's forecasts.

3.2.Results

A slightly bearish or cautious tone emerges, consistent with close monitoring of macroeconomic risks (persistent inflation, uncertainty around the Fed, mixed data in the Eurozone).

Comparison with the Random Forest Model's Predictions:

The Random Forest model, trained on historical market data, had indicated a moderately increased probability of a short-term bearish reversal, with a confidence level around 60%.

This hypothesis appears to be confirmed by the cautious stance of institutional analysts, who lean toward neutral or bearish positions on EUR/USD in the near term.

3.3.Conclusion of the Integration

This dual approach quantitative via machine learning, and qualitative via automated semantic analysis, strengthens the overall robustness of the findings. It also allows the incorporation of exogenous factors that are difficult to integrate into a purely statistical model (analyst confidence, implicit language, geopolitical perceptions, etc.).

The integration of an LLM-based layer in the validation pipeline proved valuable in refining the interpretation of probabilistic signals generated by the Random Forest model.

This hybrid validation framework also highlights the emerging role of deep learning models in financial interpretation tasks. Large Language Models (LLMs), built on transformer architectures, have demonstrated remarkable capabilities in capturing context and semantics beyond traditional rule-based NLP (Goodfellow et al., 2016). Their use in this study aligns with recent developments in combining deep learning with classical machine learning models for enhanced forecasting accuracy, as discussed by Nguyen and Bai (2021). Integrating an LLM into the evaluation loop not only enriches the interpretability of the model's outputs but also opens avenues for dynamic, multi-layered validation strategies in financial modeling.

References

Breiman, L. (2001) 'Random forests', *Machine Learning*, 45(1), pp. 5–32. Available at: https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984) *Classification and regression trees*. Boca Raton: CRC Press. Available at: https://doi.org/10.1201/9781315139470

Cavalcante, R.C. et al. (2016) 'Computational intelligence and financial markets: A survey and future directions', *Expert Systems with Applications*, 55, pp. 194–211. Available at: https://doi.org/10.1016/j.eswa.2016.02.002

Chen, T. and Guestrin, C. (2016) 'XGBoost: A scalable tree boosting system', in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 785–794. Available at: https://doi.org/10.1145/2939672.2939785

De Prado, M.L. (2018) *Advances in financial machine learning*. Hoboken: Wiley. Available at: https://doi.org/10.1002/9781119482086

Engle, R.F. (1982) 'Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation', *Econometrica*, 50(4), pp. 987–1007. Available at: https://doi.org/10.2307/1912773

Fama, E.F. (1970) 'Efficient capital markets: A review of theory and empirical work', *The Journal of Finance*, 25(2), pp. 383–417. Available at: https://doi.org/10.2307/2325486

Goodfellow, I., Bengio, Y. and Courville, A. (2016) *Deep learning*. Cambridge, MA: MIT Press. Available at: https://doi.org/10.7551/mitpress/10243.001.0001

Hastie, T., Tibshirani, R. and Friedman, J. (2009) *The elements of statistical learning*. 2nd edn. New York: Springer. Available at: https://doi.org/10.1007/978-0-387-84858-7

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013) *An introduction to statistical learning*. New York: Springer. Available at: https://doi.org/10.1007/978-1-4614-7138-7

Kim, K.J. (2003) 'Financial time series forecasting using support vector machines', *Neurocomputing*, 55(1–2), pp. 307–319. Available at: https://doi.org/10.1016/S0925-2312(03)00372-2

Krauss, C., Do, X.A. and Huck, N. (2017) 'Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the Set P 500', *European Journal of Operational Research*, 259(2), pp. 689–702. Available at: https://doi.org/10.1016/j.ejor.2016.10.031

Luss, R. and d'Aspremont, A. (2015) 'Predicting abnormal returns from news using text classification', *Quantitative Finance*, 15(6), pp. 999–1012. Available at: https://doi.org/10.1080/14697688.2015.1045821

Nguyen, H. and Bai, L. (2021) 'Time series forecasting using LSTM and random forest: A hybrid approach', *Neurocomputing*, 417, pp. 302–317. Available at: https://doi.org/10.1016/j.neucom.2020.08.048

Shen, D., Jiang, H. and Zhang, Y. (2020) 'Forecasting stock trends with LSTM and random forest', *Finance Research Letters*, 32, 101173. Available at: https://doi.org/10.1016/j.frl.2018.10.026

Tashman, L.J. (2000) 'Out-of-sample tests of forecasting accuracy: An analysis and review', *International Journal of Forecasting*, 16(4), pp. 437–450. Available at: https://doi.org/10.1016/S0169-2070(00)00065-0

Tsay, R.S. (2010) *Analysis of financial time series*. 3rd edn. Hoboken: Wiley. Available at: https://doi.org/10.1002/9780470644560

Zhang, Y. and Zhou, Z.H. (2007) 'Multi-class learning problems: A review', *Pattern Recognition Letters*, 26(3), pp. 321–335. Available at: https://doi.org/10.1016/j.patrec.2004.10.003