

# Stratégie de trading GBP/USD : intégration du sentiment FinBERT et de modèles d'arbres décisionnels

Livrable 1 : Rapport préliminaire

Équipe de recherche 3

Par

**Charles Antoine Grégoire** – [Analyste](#)

**Shanga Herlem** – [Analyste Senior](#)

**Rami Houheche** – [Analyste Senior](#)

**Léandre Iskin** – [Analyste Senior](#)

**Victoria Zelinsky** – [Head](#)

Travail présenté au Club de Trading

Dans le cadre du Fonds de Recherche

Remis le 21 septembre 2025

# Table des matières

1	Axe central de la stratégie	2
2	Intérêt de la recherche	2
3	Modèle initial / Cadre théorique	2
4	Données	4
5	Prochaines étapes	5

# 1 Axe central de la stratégie

Construire une stratégie de trading Forex qui combine **scores de sentiment extraits par FinBERT** et **modèles d'arbres (Random Forest / XGBoost)** pour prédire les régimes de marché et améliorer la prise de décision. Cette approche vise à exploiter des inefficiences informationnelles de court terme afin de générer des signaux de trading robustes, évalués tant sur le plan prédictif que sur leur performance économique.

## 2 Intérêt de la recherche

Le marché des changes occupe une place prépondérante dans les activités spéculatives, en raison tant de son volume considérable que de sa volatilité prononcée. D'après la Bank for International Settlements, le volume moyen quotidien des transactions sur les marchés des changes de gré à gré s'élevait à 7,5 billions de dollars américains (7 500 milliards USD) en avril 2022, la paire GBP/USD représentant la devise la plus échangée [1]. Dans ce contexte, la capacité à anticiper les fluctuations des taux de change apparaît comme un élément déterminant dans l'élaboration et la mise en œuvre d'une stratégie de négociation. Le Forex étant particulièrement sensible aux indicateurs macroéconomiques, la géopolitique ainsi qu'aux annonces des banques centrales, le recours à un modèle capable d'intégrer simultanément plusieurs signaux afin de prévoir l'évolution d'une paire de devises, en l'occurrence l'GBP/USD, apparaît pertinent. L'analyse automatique du texte permet de capter le ton et l'intention avant que le marché n'ait entièrement digéré l'information. Notre approche est pertinente car elle fusionne : (i) un modèle NLP spécialisé en finance (FinBERT) et (ii) des modèles supervisés non linéaires capables de modéliser des interactions complexes entre sentiment et variables statistiques.

La stratégie est innovante par son couplage explicite entre *text as signal* et *feature engineering* quantitatif pour la classification de régimes. La littérature appuie cette démarche : Araci introduit FinBERT et montre son utilité pour des tâches de sentiment en finance [2]. Des revues antérieures concluent que l'ajout d'information textuelle améliore la prédictivité sur FX [3]. Enfin, des études récentes montrent la supériorité des méthodes de gradient boosting sur des tâches financières bruitées, justifiant le choix de XGBoost/LightGBM [4].

## 3 Modèle initial / Cadre théorique

La logique de base se compose de deux blocs.

**Extraction de sentiment.** Soit  $n_t$  une nouvelle à l'instant  $t$ . FinBERT produit un score  $s_t \in [-1, 1]$ . On définit le score total  $S_t$  sur une fenêtre  $W$  :

$$S_t = \frac{1}{|W|} \sum_{i=0}^{W-1} w_i s_{t-i},$$

où  $w_i$  sont des poids (ex. décroissants avec le temps, à déterminer plus précisément en backtest). Autrement dit, l'extraction de sentiment s'effectuera à partir d'une moyenne pondérée qui reflète le caractère récent des données.

**Prédiction par arbres.** Définissons le vecteur de caractéristiques

$$X_t = \{S_t, r_{t-1}, \sigma_{t-1}, \text{MA}_{t-5}, \Delta i_t, \text{événements}_t\},$$

avec  $r_{t-1}$  le rendement 1 jour,  $\sigma_{t-1}$  la volatilité réalisée,  $\Delta i_t$  le différentiel de taux pertinent. Le modèle (Random Forest ou XGBoost) estime la probabilité :

$$\hat{p}_t = \mathbb{P}(y_{t+1} = 1 \mid X_t),$$

où  $y_{t+1} = 1$  si la paire GBP/USD est en hausse le jour suivant. La règle de trading simple :

$$\text{Position}_t = \begin{cases} +1 & \text{si } \hat{p}_t > \tau_H, \\ -1 & \text{si } \hat{p}_t < \tau_L, \\ 0 & \text{sinon.} \end{cases}$$

avec seuils  $\tau_H = 0.6$ ,  $\tau_L = 0.4$  (à calibrer).

### Hypothèses.

1. Marché partiellement inefficent à très court terme : le marché des changes peut présenter des décalages dans l'intégration de l'information publique. L'analyse de sentiment issue de textes financiers permet ainsi de capter un signal pertinent qui n'est pas immédiatement incorporé dans le prix des devises. Nous postulons donc une violation ponctuelle de la forme semi-forte de l'hypothèse d'efficience des marchés (EMH), en ce sens que l'information publique n'est pas reflétée instantanément dans les cours.
2. Interactions entre les variables : les modèles d'arbres de décision, tels que Random Forest, sont capables de modéliser des relations non linéaires entre le sentiment et les indicateurs quantitatifs (rendements retardés, volatilité, indicateurs techniques, différentiels de taux).

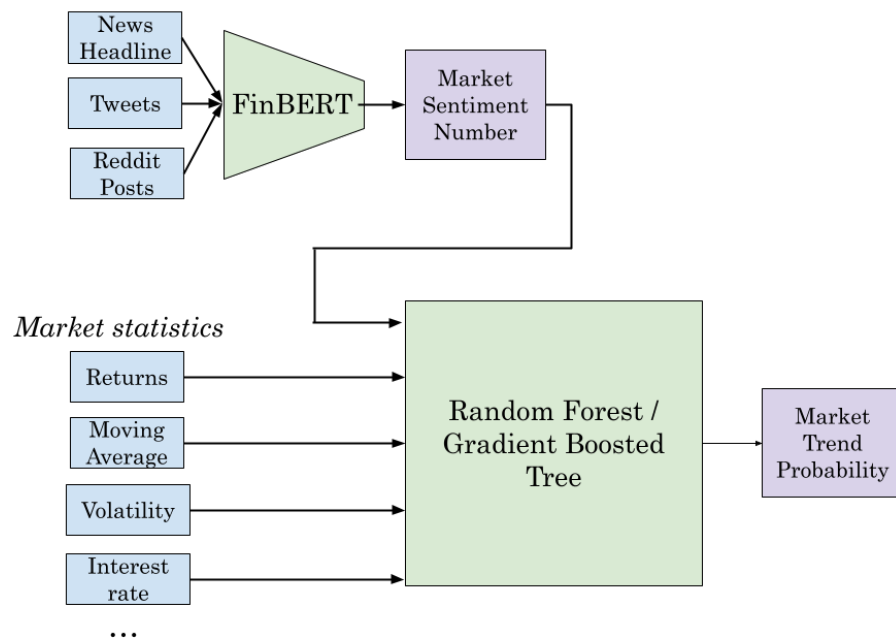


FIGURE 1 – Architecture de notre stratégie

Cette capacité permet de mieux caractériser les régimes de marché et d'améliorer la précision des prédictions.

## 4 Données

### Sources prévues.

- **Prix FX** : Yahoo Finance pour le prototypage, ensuite données de meilleure qualité (WRDS) pour la production.
- **Textes** : flux ForexLive, FXStreet, Investing.com, communiqués de banques centrales (EDGAR/ECB/Fed).
- **Macroéconomique** : FRED pour séries macroéconomiques (taux d'intérêts, PIB, taux de chômage).

### Variables retenues.

#### Cible :

- Direction du rendement quotidien GBP/USD (binaire).
- Score FinBERT agrégé  $S_t$ .

#### Explicatives :

- *Sentiment* : Articles, post de réseaux sociaux.

- *Techniques* : rendements lag, MA(5), MA(20), RSI, ATR, volatilité réalisée sur 20 jours.
- *Macroéconomique / événements* : différentiels de taux, indicateurs calendrier (NFP, CPI), flags d'événements.

## 5 Prochaines étapes

Plan de travail immédiat (ordre chronologique) :

1. **Collecte & nettoyage** : ingérer prix (2018–2025), scraper flux news, normaliser timestamps et éliminer doublons.
2. **Pipeline NLP** : appliquer FinBERT pour obtenir  $s_t$ , stocker embeddings et scores ; tester fine-tuning FinBERT sur un petit jeu d'étiquettes FX si disponible.
3. **Construction du dataset** : aligner features temporellement, créer cibles, splitter train/validation/test en rolling windows.
4. **Implémentation du modèle de base** : Random Forest (baseline), XGBoost (amélioration). Mesures : AUC, précision, métriques économiques (Sharpe, max drawdown).
5. **Backtest économique** : transformer signaux en exécutions simulées ; inclure spread, slippage, taille minimale, benchmark.
6. **Validation** : walk-forward, tests de robustesse (stress tests, coût augmenté, période hors-échantillon). Benchmark : Nous comparerons nos résultats à plusieurs benchmarks représentatifs : (i) un modèle naïf de persistance (dernier signe répété), (ii) une régression logistique comme baseline ML simple, et (iii) une stratégie technique classique (crossover de moyennes mobiles). Ces comparaisons permettront d'évaluer la valeur ajoutée de notre approche par rapport à des méthodes triviales, statistiques et techniques.
7. **Améliorations** : tester FinRoberta/FinGPT, passer à des modèles neuronaux

### Conclusion.

La stratégie vise une application pragmatique et reproductible : FinBERT fournit un signal textuel rapide ; Random Forest / XGBoost exploitent ce signal avec des variables quantitatives. L'objectif est de produire un modèle robuste, transparent et facilement testable par backtest.

---

### Références (sélection)

1. Bank for International Settlements. (2022). *Triennial Central Bank Survey : Foreign exchange turnover in April 2022* (Statistical release).
2. Araci, D. (2019). *FinBERT : Financial Sentiment Analysis with Pre-trained Language Models*.
3. Nassirtoussi, A. K., Aghabozorgi, S., Ying Wah, T., Ngo, D. C. L. (2015). *Text mining for*

*market prediction : A systematic review.* Expert Systems with Applications.

4. Olorunnimbe, K., Viktor, H. (2023). *Deep learning in the stock market—a systematic survey of practice, backtesting, and applications.* Artif Intell Rev 56.