

Prédiction de l'évolution des taux directeurs à partir des discours de la FED

GASPARIN-GRANGER Lia, LIBERGE Pierre

Avril 2025

Abstract

Les discours des membres de la Réserve fédérale américaine (FED) jouent un rôle central dans la communication de la politique monétaire. Au-delà des décisions officielles, ces prises de parole contiennent souvent des signaux sur les intentions futures de la Banque Centrale. Ce projet explore la possibilité d'exploiter ces discours pour anticiper l'évolution des taux directeurs. En combinant traitement automatique du langage (NLP) et apprentissage automatique, nous cherchons à modéliser le lien entre le contenu textuel des discours et les décisions futures de la FED. L'objectif est de construire un modèle capable de prédire, à partir d'un discours donné, si la tendance à venir des taux sera à la hausse, à la baisse ou stable.

Ce travail s'inscrit à la croisée de la linguistique computationnelle et de l'analyse macroéconomique, avec des applications potentielles en finance, en stratégie économique ou en veille réglementaire. En plus du fine-tuning de modèles pré-entraînés comme FinBERT, nous avons exploré plusieurs approches plus classiques du NLP : des représentations TF-IDF combinées à des algorithmes. Nous avons également testé des embeddings Glove pondérés par TF-IDF, associés à des modèles SVM ou XGBoost, afin d'évaluer différentes manières de capter les signaux latents dans le langage économique de la FED.

Contents

1	Introduction	3
2	Pré-traitement des discours de la FED	3
3	Modélisation	4
3.1	Modèles basés sur des représentations vectorielles classiques	4
3.2	Modèles basés sur des représentations avancées	5
3.3	Modèle FinBERT	5
4	Résultats	5
5	Conclusion	6

1 Introduction

La politique monétaire menée par les banques centrales constitue un levier essentiel pour influencer l'activité économique, la stabilité des prix et les anticipations des marchés financiers. Parmi les acteurs majeurs, la Réserve fédérale américaine (FED) joue un rôle central, non seulement par ses décisions sur les taux directeurs, mais aussi par la manière dont elle communique ces décisions aux agents économiques.

La FED dispose de plusieurs instruments pour mener sa politique monétaire, dont deux taux d'intérêt principaux :

- **Taux des fonds fédéraux** : taux d'intérêt directeur principal de la Fed, utilisé pour réguler la liquidité à très court terme entre banques sur le marché interbancaire. Il influence directement les conditions de financement dans l'économie.
- **Taux d'escompte** : taux appliqué aux prêts accordés directement par la Fed aux banques via la *discount window*. Généralement plus élevé, il joue un rôle de filet de sécurité ou de signal monétaire.

Dans ce projet, nous nous concentrons sur le *Federal Funds Rate*, principal levier de la politique monétaire de la FED. Les décisions concernant ce taux sont souvent précédées de signaux, explicites ou implicites, dans les discours des membres du comité de politique monétaire.

Ces discours sont devenus des outils de communication stratégiques, permettant de guider les anticipations des marchés et d'orienter les conditions financières. Comprendre le langage de la FED et en extraire des signaux prédictifs constitue donc un enjeu crucial pour les acteurs économiques.

Notre objectif est d'évaluer dans quelle mesure le contenu textuel des discours permet de prédire l'orientation future du taux des fonds fédéraux. Pour cela, nous mobilisons des techniques de traitement automatique du langage (NLP) couplées à des modèles d'apprentissage supervisé, afin de classer chaque discours selon la tendance anticipée des taux : *hausse*, *baisse* ou *stabilité*.

2 Pré-traitement des discours de la FED

Pour rendre les discours de la FED exploitables par des algorithmes d'apprentissage automatique, un pré-traitement rigoureux a été mis en place. Il commence par la tokenisation, qui segmente le texte en unités lexicales (tokens), facilitant le traitement de la ponctuation et des contractions. Ensuite, les mots sont convertis en minuscules afin d'uniformiser les formes (ex. *Rate* devient *rate*), ce qui diminue la dimensionnalité du vocabulaire. Les stopwords, mots fréquents sans valeur informative, sont ensuite supprimés pour recentrer l'analyse sur les termes significatifs. Enfin, la lemmatisation ramène chaque mot à sa forme de base (ex. *raising*, *raised* → *raise*), assurant une meilleure cohérence sémantique.

Un autre enjeu concerne la granularité temporelle des données : les discours sont quotidiens, tandis que les taux directeurs sont mensuels. Ce décalage nécessite un alignement soigneux entre les séries textuelles et numériques pour permettre une modélisation cohérente.

Pour pallier ce problème, nous avons adopté la solution suivante :

- Chaque discours est associé à la date du 1er jour du mois de sa publication.
- Le discours est ensuite utilisé pour prédire l'évolution du taux directeur le mois suivant.

Cette méthode permet de transformer les observations journalières en entrées cohérentes avec des sorties mensuelles, assurant ainsi un alignement temporel approprié pour l'apprentissage supervisé.

Avant d'entraîner un modèle de classification, il est essentiel de mieux comprendre le contenu des discours et la nature des étiquettes associées. Pour ce faire, deux analyses exploratoires ont été réalisées : l'une sur la fréquence des mots présents dans les discours, l'autre sur la répartition des classes de la variable cible FEDFUNDS.

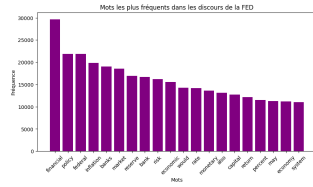


Figure 1: Mots les plus fréquents dans les discours de la FED

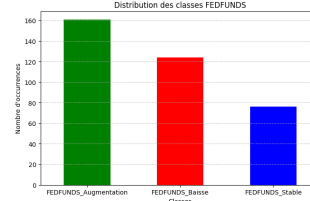


Figure 2: Distribution des classes de la variable FEDFUNDS

Les Figures 1 et 2 révèlent respectivement la dominance de termes économiques dans les discours de la FED et un déséquilibre notable entre les classes de variation du taux directeur, un facteur à considérer lors de la modélisation. Une analyse temporelle des variations quotidiennes du taux permet ensuite d’en éclairer la dynamique.

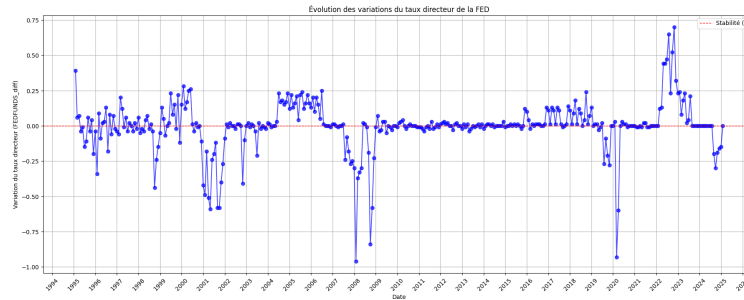


Figure 3: Évolution des variations mensuelles du taux de la FED

La majorité des variations du taux directeur sont proches de la stabilité, hormis des ajustements marqués lors de crises majeures (2008–2009, 2020), ce qui motive une classification en trois classes distinctes : hausse, baisse, stabilité.

3 Modélisation

Pour anticiper l’évolution des taux à partir du contenu textuel des discours, nous avons progressivement évalué plusieurs types de modèles de traitement automatique du langage naturel, selon une montée en complexité et en expressivité des représentations. Trois grandes familles de modèles ont été testées : les approches vectorielles classiques, les représentations avancées basées sur des embeddings pré-entraînés, et enfin des modèles neuronaux d’architecture transformer fine-tunés spécifiquement sur la tâche.

3.1 Modèles basés sur des représentations vectorielles classiques

Les discours de la FED ont d’abord été vectorisés à l’aide d’approches *Bag-of-Words*, où chaque document est représenté par un vecteur de dimension fixe issu du vocabulaire du corpus.

CountVectorizer repose sur le comptage brut des occurrences des mots dans les documents.

TF-IDF (Term Frequency - Inverse Document Frequency) affine cette approche en pondérant chaque mot selon son importance dans le document et sa rareté dans le corpus.

Une fois les discours vectorisés, plusieurs modèles de classification supervisée ont été mis en œuvre pour prédire la direction future du taux directeur. Le *Multinomial Naive Bayes* est un modèle probabiliste simple, efficace avec CountVectorizer, qui suppose l’indépendance des mots et fonctionne bien lorsque des mots-clés distinctifs sont présents. Il reste cependant sensible aux déséquilibres entre classes.

La *régression logistique*, quant à elle, est un modèle linéaire robuste qui transforme les scores de chaque classe en probabilités. Elle s’adapte particulièrement bien aux représentations TF-IDF, en exploitant les mots discriminants tout en limitant l’influence des termes trop fréquents.

Enfin, le *Support Vector Machine (SVM)* cherche à séparer les classes en maximisant la marge entre elles. Très performant sur des représentations textuelles creuses comme TF-IDF, il utilise une stratégie *one-vs-rest* en multi-classe et offre une excellente capacité de généralisation dans des espaces de grande dimension.

3.2 Modèles basés sur des représentations avancées

Dans une seconde phase, nous avons enrichi la représentation des discours en utilisant des **embeddings pré-entraînés GloVe**, permettant de capturer des relations sémantiques entre les mots. Contrairement aux approches Bag-of-Words, ces représentations vectorielles continues permettent d’exploiter la proximité sémantique. Chaque mot est projeté dans un espace vectoriel de 100 dimensions via les vecteurs GloVe. Ces vecteurs encodent des régularités linguistiques : les mots ayant des contextes similaires sont représentés par des vecteurs proches, ce qui permet au modèle de généraliser au-delà des mots strictement présents dans les données d’entraînement. Pour exploiter la richesse sémantique des vecteurs GloVe tout en tenant compte de la pertinence locale des mots dans chaque discours, nous avons appliqué une pondération par TF-IDF à chaque embedding. Cette étape permet de valoriser les mots les plus discriminants et de minimiser l’influence des mots peu informatifs. Une fois chaque mot représenté par un vecteur pondéré, nous avons agrégé ces représentations au niveau du document à l’aide d’un *Max Pooling*. Cette opération consiste à conserver, pour chaque dimension du vecteur, la valeur maximale observée parmi tous les mots du document. Elle permet ainsi d’extraire les traits les plus saillants du texte et de générer une représentation compacte, apte à alimenter les modèles de classification. Les vecteurs de documents ainsi obtenus ont servi d’entrée pour deux modèles de classification :

- **Support Vector Machine (SVM)** : efficace sur des représentations denses, il a su tirer parti des caractéristiques sémantiques extraites via GloVe + TF-IDF.
- **XGBoost** : algorithme de gradient boosting très performant pour la classification non linéaire. Il a été utilisé pour capturer des interactions complexes entre les dimensions vectorielles des documents.

3.3 Modèle FinBERT

FinBERT est une version de BERT spécialisée dans le domaine financier, entraînée sur un large corpus de textes comme des rapports annuels ou des articles économiques. Sa version fine-tunée, FinBERT-tone, est conçue pour la classification de sentiment à partir de textes financiers, en s’appuyant sur un ensemble annoté en trois catégories : positif, négatif et neutre.

Dans ce projet, nous avons fine-tuné FinBERT-tone pour une tâche de classification multi-classes, visant à prédire l’évolution du taux directeur à partir des discours de la FED. Grâce à ses représentations contextuelles, le modèle peut capter les signaux sémantiques complexes liés à la politique monétaire.

4 Résultats

Après avoir mis en œuvre différentes approches de modélisation, cette section présente les résultats obtenus.

	Naive Bayes	SVM	Régression Logistique
Balanced Accuracy (CV)	0.4220	0.4794	0.4685
Balanced Accuracy (Test)	0.4128	0.5070	0.4918
F1-score macro (CV)	0.4164	0.4741	0.4700
F1-score macro (Test)	0.4189	0.4972	0.4913

Figure 4: Comparaison de modèles (vectorisation TF-IDF)

L’analyse des performances des trois modèles avec une vectorisation TF-IDF montre une nette supériorité des approches linéaires. Le modèle *Naive Bayes*, avec un **F1-score macro de 0,4189**, reste limité, même

lorsqu'il est utilisé avec *CountVectorizer* ($F1 \approx 0,44$), ce qui en fait un bon point de départ, mais peu adapté à la complexité des discours économiques. À l'inverse, la *régression logistique* et surtout le *SVM linéaire* offrent de bien meilleurs résultats. Le *SVM*, avec un **F1-score de 0,4972**, se distingue par sa capacité à bien couvrir les trois classes et à généraliser efficacement, tandis que la *régression logistique* ($F1 : 0,4913$) se montre particulièrement stable et plus facile à interpréter.

Afin d'aller plus loin dans notre démarche, présentons les résultats obtenus à l'aide de la modélisation basée sur des représentations avancées.

```
=== Comparaison des performances : SVM vs XGBoost (GloVe embeddings) ===
```

	SVM (GloVe)	XGBoost (GloVe)
Balanced Accuracy (CV)	0.4005	0.4003
Balanced Accuracy (Test)	0.3950	0.3893
F1-score macro (CV)	0.3904	0.3959
F1-score macro (Test)	0.3747	0.3876

Figure 5: Comparaison de modèles (embeddings GloVe pondérés par TF-IDF)

Les résultats obtenus avec les représentations GloVe pondérées par TF-IDF restent en retrait par rapport aux approches classiques. Le *SVM* atteint un **F1-score macro de 0,3747** en test, contre **0,3876** pour *XGBoost*. L'écart est minime, mais ces scores globalement plus faibles montrent que, dans ce contexte, les représentations GloVe agrégées par max pooling n'apportent pas de gain significatif par rapport à la vectorisation TF-IDF plus traditionnelle.

Pour finir, les résultats obtenus avec FinBERT, évalué selon différentes configurations, montrent une progression nette des performances à mesure que les hyperparamètres sont optimisés. Lors du premier essai, avec 4 epochs et un learning rate de $2e-5$, le F1-score global est relativement faible, atteignant seulement 0,35. La classe Augmentation, en particulier, est mal capturée avec un F1-score de 0,30, ce qui suggère un apprentissage insuffisant.

Lors du deuxième essai, en augmentant à 8 epochs et en réduisant le learning rate à $1e-5$, le F1-score global progresse à 0,40. Les scores F1 par classe deviennent plus homogènes, notamment avec des valeurs proches de 0,41 pour Augmentation et Baisse, indiquant une meilleure généralisation du modèle.

```
=== Résultats finaux ===
Fold 1 : F1-score = 0.4190
Fold 2 : F1-score = 0.4358
Fold 3 : F1-score = 0.4181
Fold 4 : F1-score = 0.4463
Fold 5 : F1-score = 0.4685
Moyenne F1-score : 0.4376
```

Enfin, la table ci-dessus présente les résultats de la validation croisée en 5 folds, réalisée avec les meilleurs hyperparamètres du deuxième essai. Le F1-score moyen atteint 0,4376, avec une stabilité remarquable entre les différents folds (variant de 0,4181 à 0,4685). Ce résultat met en évidence la robustesse du modèle et sa capacité à bien généraliser dans un cadre d'évaluation rigoureux.

En somme, l'évolution du F1-score au fil des essais illustre l'impact significatif de l'optimisation des hyperparamètres sur les performances finales du modèle.

5 Conclusion

L'étude menée a permis d'évaluer plusieurs approches de modélisation pour prédire l'évolution du taux directeur de la FED à partir de leurs discours. Les résultats montrent une nette supériorité des modèles linéaires classiques, en particulier le SVM et la régression logistique, lorsqu'ils sont combinés à une vectorisation TF-IDF. Les approches basées sur des représentations sémantiques plus riches, comme les embeddings GloVe pondérés par TF-IDF, se sont révélées moins performantes dans ce contexte, probablement en raison d'une perte d'information lors de l'agrégation. Enfin, les premiers essais réalisés avec FinBERT confirment

le potentiel de ces modèles spécialisés, mais soulignent également la sensibilité de leurs performances aux choix d'hyperparamètres.

Plusieurs pistes d'amélioration peuvent être envisagées pour approfondir ce travail :

- **Élargir le corpus** en automatisant la collecte de discours via un *webscraper*, afin d'améliorer la couverture temporelle et la diversité des données textuelles ;
- **Tester d'autres granularités** d'analyse, en exploitant des unités plus courtes comme les phrases ou paragraphes, pour mieux détecter les signaux faibles dans le langage de la FED ;
- **Explorer des architectures séquentielles** comme les RNN, LSTM ou GRU, capables de mieux modéliser la structure syntaxique et temporelle du discours ;
- **Poursuivre l'optimisation de FinBERT**, en affinant les hyperparamètres, en utilisant davantage de données, ou en explorant des variantes comme l'ajout de couches de classification personnalisées.
- **Expérimenter des modèles de langage plus récents**, comme les LLMs (ex RoBERTa, DeBERTa, GPT-like), éventuellement fine-tunés sur des textes financiers, pour capter plus finement les nuances du langage économique ;
- **Repenser la fenêtre temporelle utilisée** : actuellement, la prédiction repose sur les discours d'un mois pour anticiper le mois suivant, mais une agrégation sur une période plus longue (par exemple trois mois) pourrait mieux capter les tendances sous-jacentes, en cohérence avec le rythme réel des ajustements de taux directeurs.

Ces axes ouvrent des perspectives pour améliorer la précision des prédictions et enrichir la compréhension automatique des signaux de politique monétaire émis par la FED.