

# G-BHAM

## Generalized Bert-base Hierarchical Aggregation Model

Florian Feuillade

Travail de bachelor 2021

Informatique – Développement logiciel et multimédia

Professeur: Hatem Ghorbel

Expert: Diego Antognini

### Description

L'origine du projet provient d'un mandat de recherche de la Haute Ecole Arc, initié par l'entreprise active dans la finance, Perreard Partners Investment SA et soutenu par l'organisme d'État pour l'innovation Innosuisse. L'entreprise cherche à développer des outils de machine learning et des algorithmes automatisés afin d'améliorer leur stratégie de gestion de portefeuille.

Le projet est une suite direct au travail de master de Monsieur Pedro Costa, qui a développé un modèle de machine learning nommée BHAM pour Bert-based Hierarchical Aggregation Model.

L'objectif du travail est de faire des recherches sur certaines parties de BHAM, notamment l'architecture des entrées et la manière d'attribuer un score à une actualité, dans le but d'améliorer ses performances.

### Déroulement

- Recherche et analyse
  - Compréhension du modèle BHAM
  - Recherche sur l'architecture, filtrage et notations de textes
  - Prise en main des données
- Implémentation
  - Mise en place du filtrage
  - Ajout du scorer similarité et transformer
- Comparaison
  - Tests de l'efficacité du filtrage
  - Tests de la performance des scorer
  - Backtest

### Résultats

Le filtrage n'a pas donné de résultats satisfaisants qui justifieraient son utilisation permanente avec le modèle.

La piste des scores a cependant donné des résultats encourageants pour une des deux versions. La similarité n'apporte aucune amélioration mais ne péjore pas non plus, tandis que la version avec Transformer démontre des performances légèrement supérieures. Améliorations des backtests, probablement dus au passage à une catégorie au lieu de quatre.

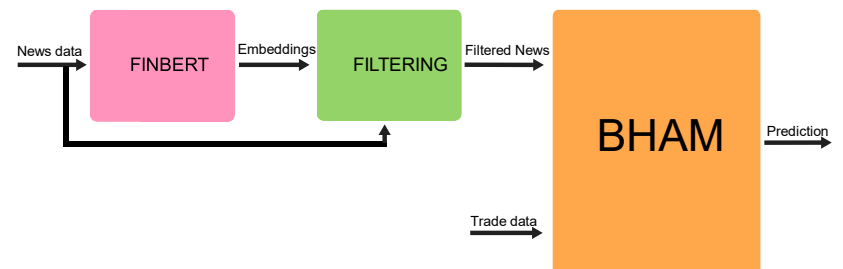


Schéma simplifié de BHAM avec filtrage

| Models            | BHAM_TM | Size 568 Layer 3 | Basic scorer |
|-------------------|---------|------------------|--------------|
| Accuracy mean     | 0.578   | <b>0.59</b>      | 0.589        |
| std               | 0.007   | <b>0.002</b>     | 0.003        |
| F1-macro mean     | 0.577   | <b>0.589</b>     | 0.588        |
| std               | 0.009   | <b>0.002</b>     | 0.003        |
| cumulative return | 2.73    | <b>2.91</b>      | 2.90         |
| std               | 0.24    | <b>0.21</b>      | 0.22         |
| annual return     | 0.65    | <b>0.69</b>      | 0.69         |
| std               | 0.06    | <b>0.05</b>      | 0.06         |
| annual volatility | 0.82    | <b>0.73</b>      | 0.72         |
| std               | 0.02    | <b>0.01</b>      | 0.01         |
| sharpe ratio      | 0.80    | <b>0.95</b>      | 0.94         |
| std               | 0.08    | <b>0.07</b>      | 0.07         |

TABLE 6.6 – Résultat backtest

Résultats backtest

### Perspectives

Le projet est un travail de recherche expérimentale visant à explorer des idées. Le projet n'est pas clôturé après ce travail mais pourra être poursuivi en prenant en compte les échecs et les résultats. Par exemple, le score avec Transformer possède un certain nombre d'hyperparamètres. Il pourrait être intéressant de rechercher les meilleurs.