

# DATA STREAM PROCESSING

Thomas Bersani -- Veroni

Luc Gensbittel

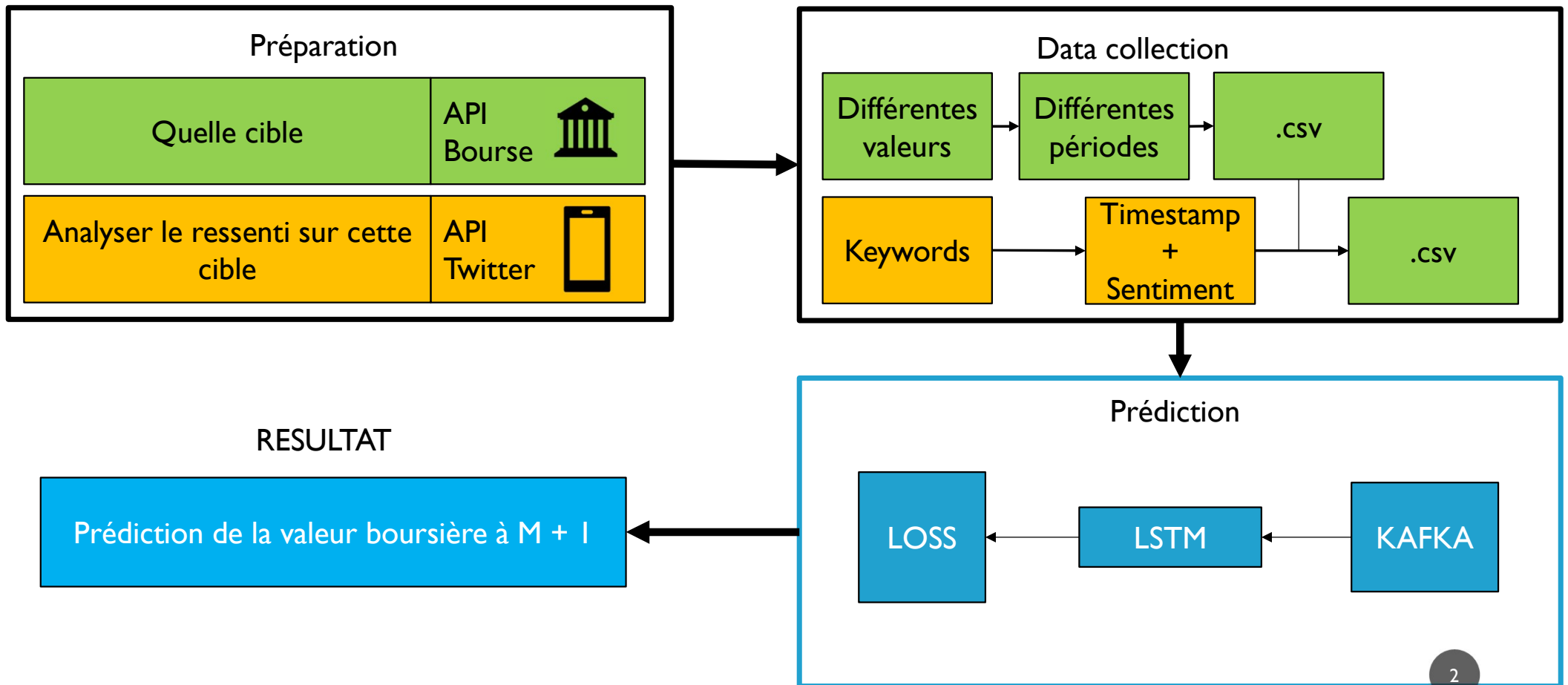
Titouan Lermite

Recherche exhaustive

Projet Bourse – tweeter – API :

Objectif : prédire une valeur boursière à  $M+1$

# ARCHITECTURE GÉNÉRALE



# I. RÉCUPÉRATION DES DONNÉES

Sous contrainte budgétaire des API gratuites



1. Définir précisément le sujet
2. Données homogènes suffisantes

Décision retenue :

- ❖ API : Yahoo! Finance
- ❖ Cibles : NYSE et NASDAQ
- ❖ Intervalles: 1 min à 1 mois
  - Démonstrateur sur GOOGL
  - Période du 16/11/2022 au 13/01/2023

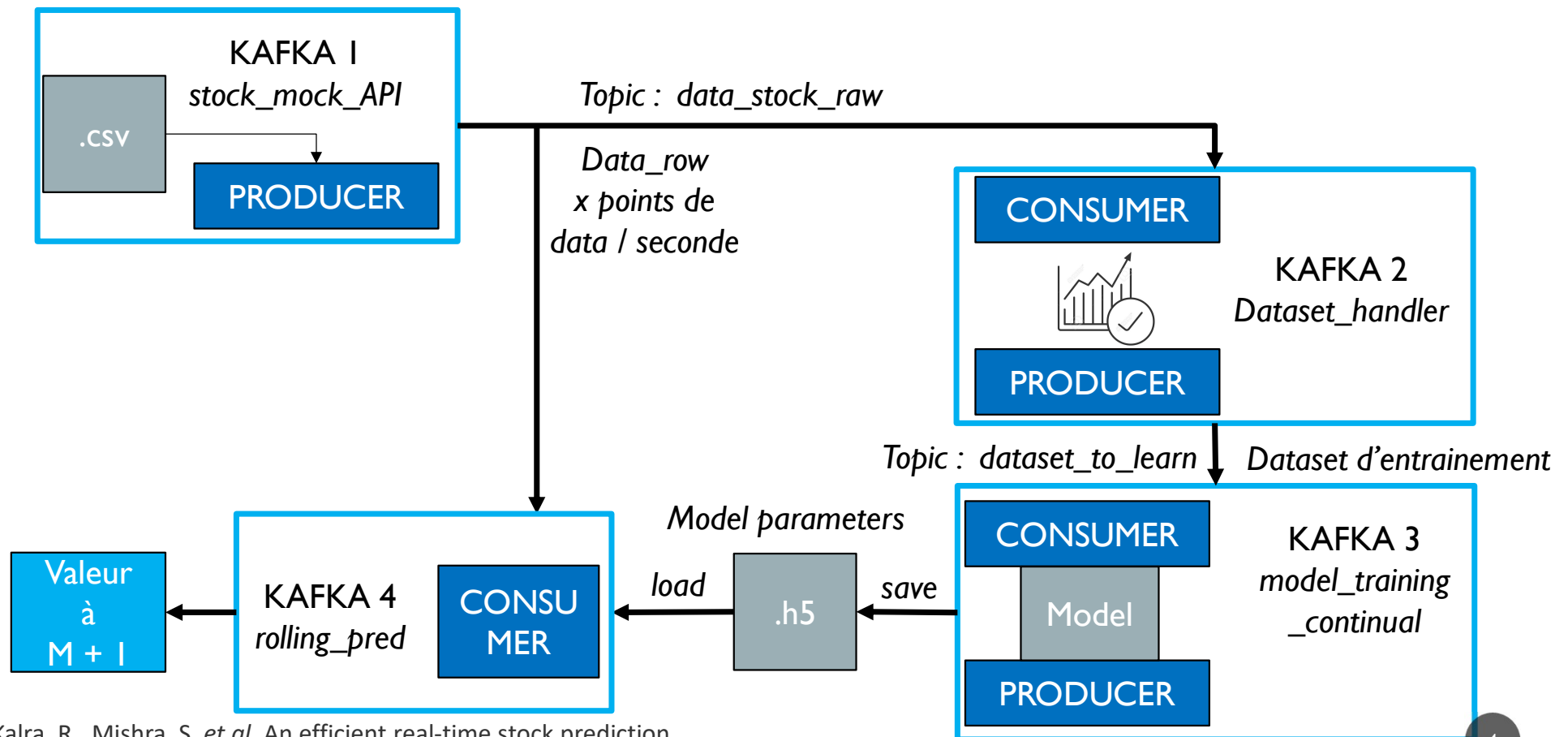
NVDA  
Netflix Twitter Apple  
Google Amazon  
Microsoft  
Adobe Tesla  
Facebook



1. Adapter l'extraction des tweet aux données
2. Équilibre volume / requêtes

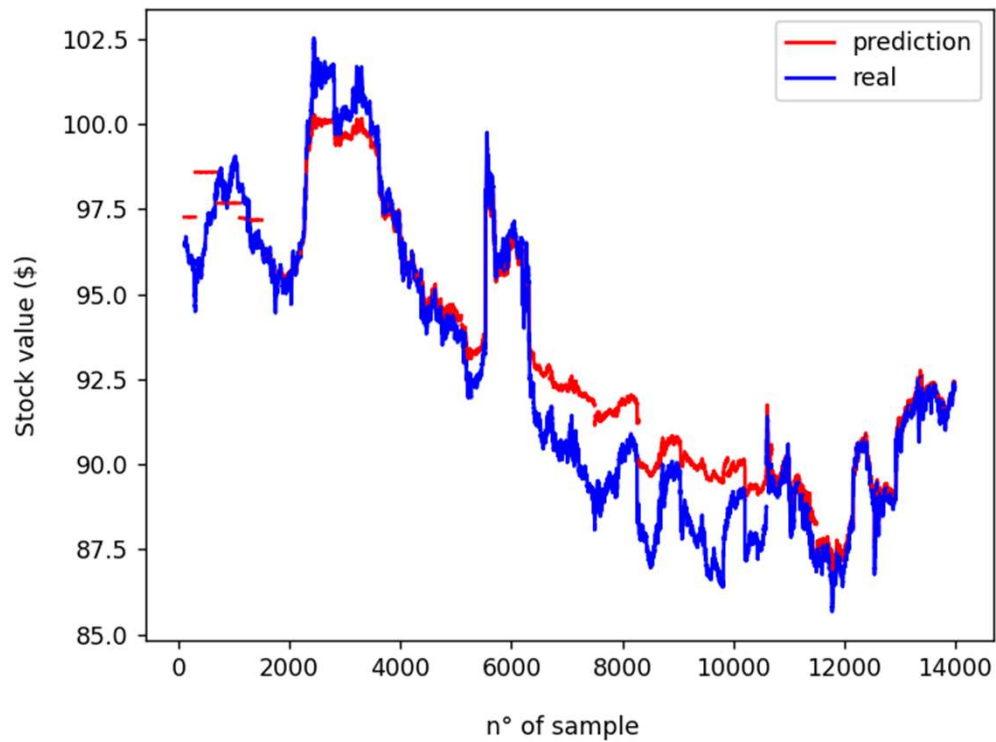
- Mots clés par requêtes et/ou timestamp
- Différentes façon de récupérer des tweets
  - TwitterSearchScraper / Tweepy
  - Soft NLP : regex et traduction
- Analyse de sentiments
  - polarité et subjectivité

## II. DATA STREAM LEARNING DEEP LEARNING MODEL

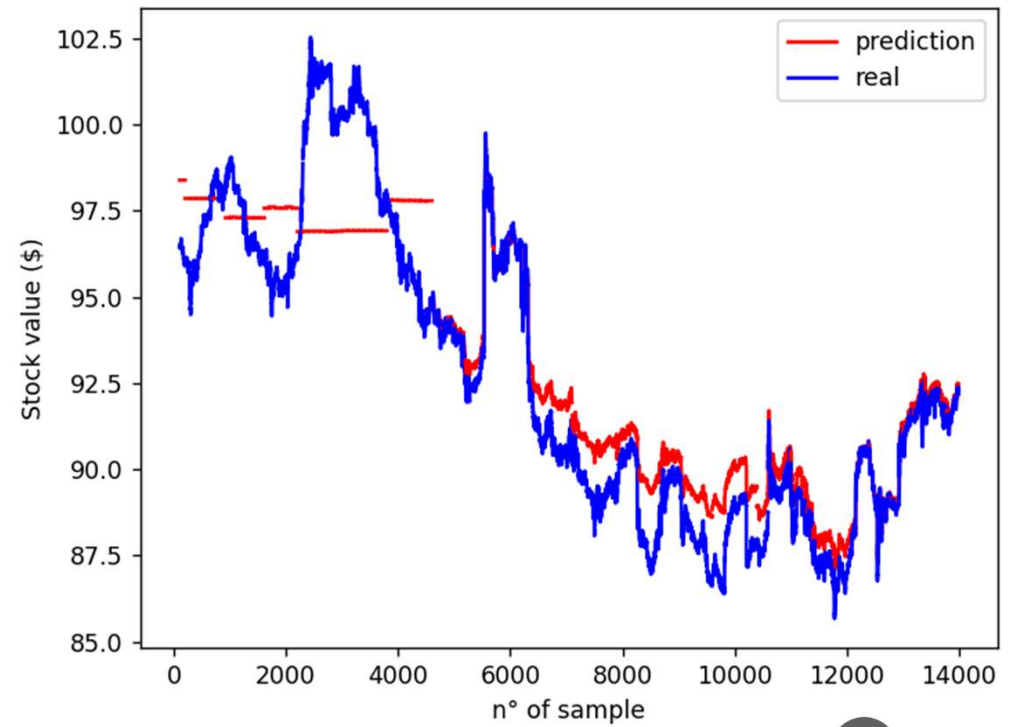


### III. RÉSULTATS STREAM LEARNING

*Learning avec 10 points  
/minutes*

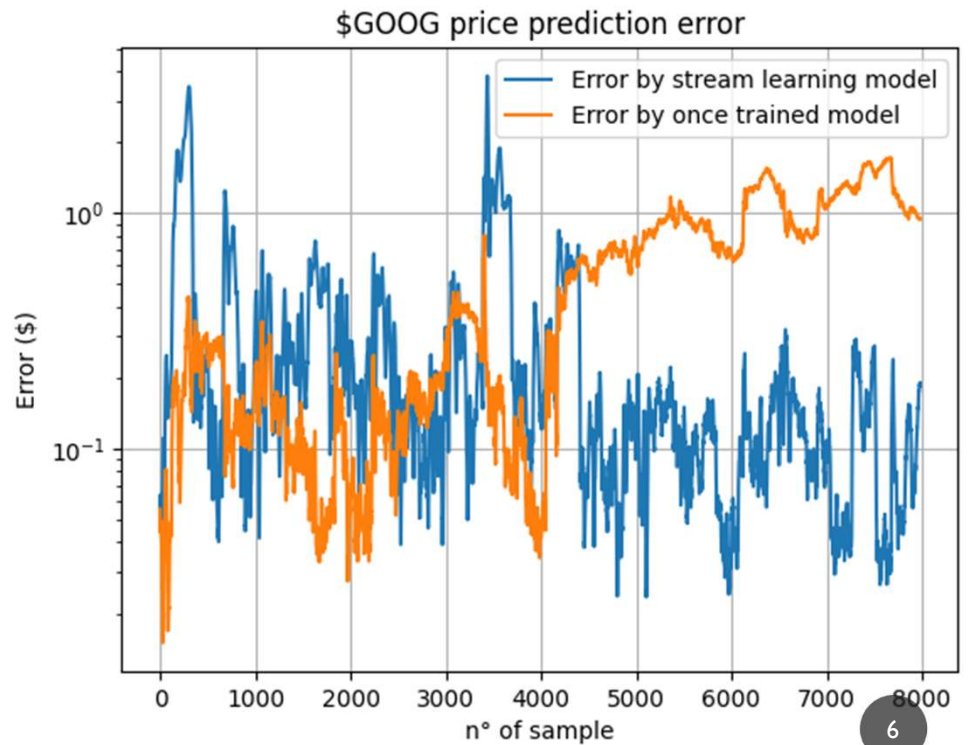
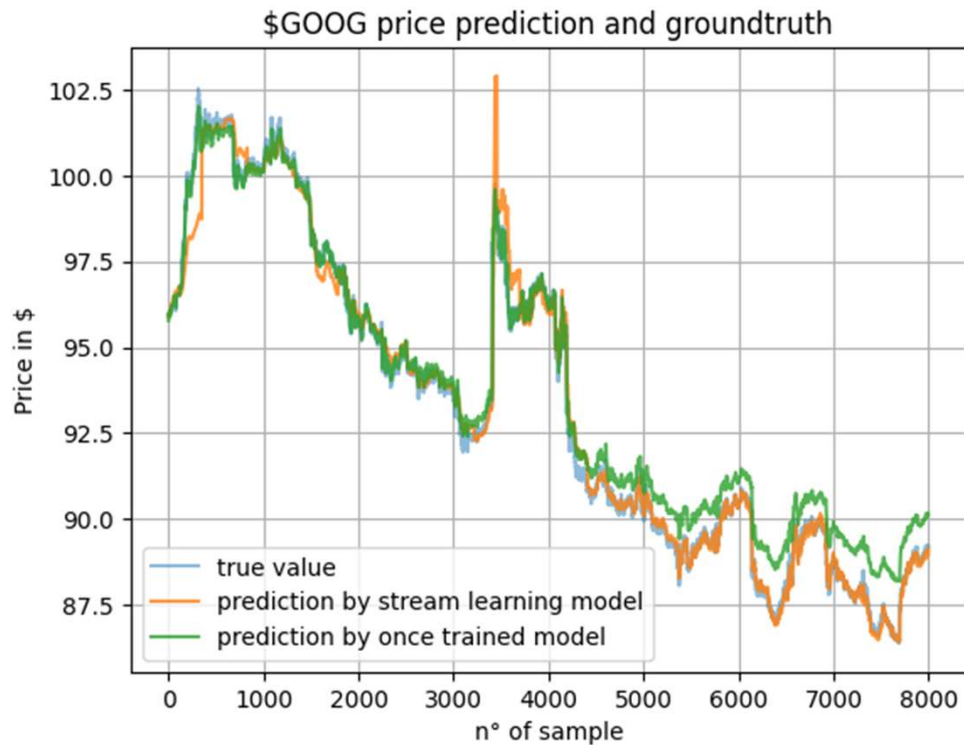


*Learning avec 100 points  
/2.5 secs*



## IV. RÉSULTATS SIMULATIONS

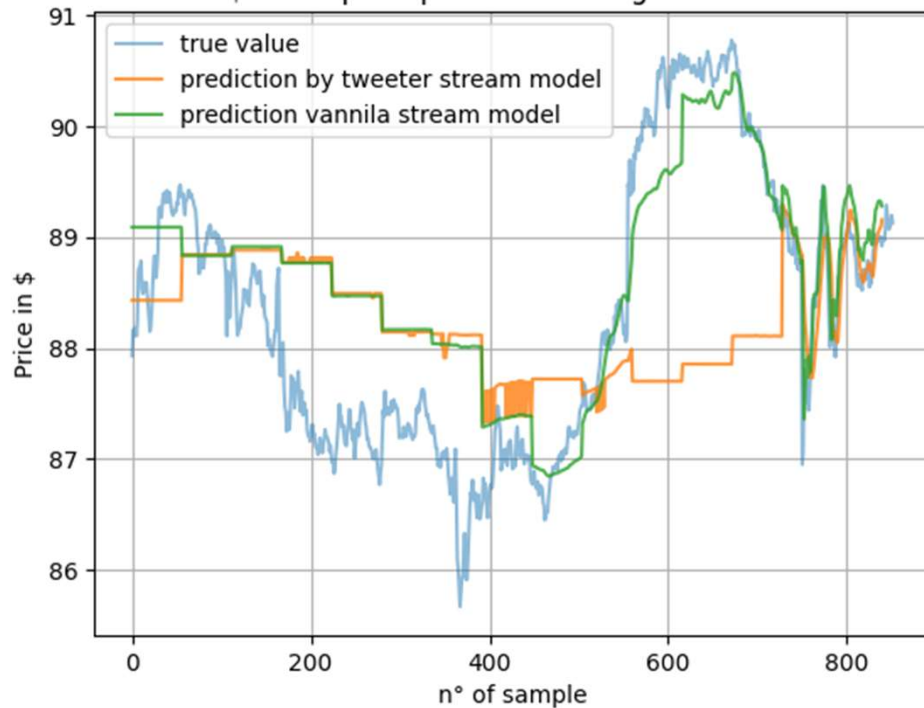
### COMPARAISON STREAM MODELS VS TRAIN ONCE MODELS



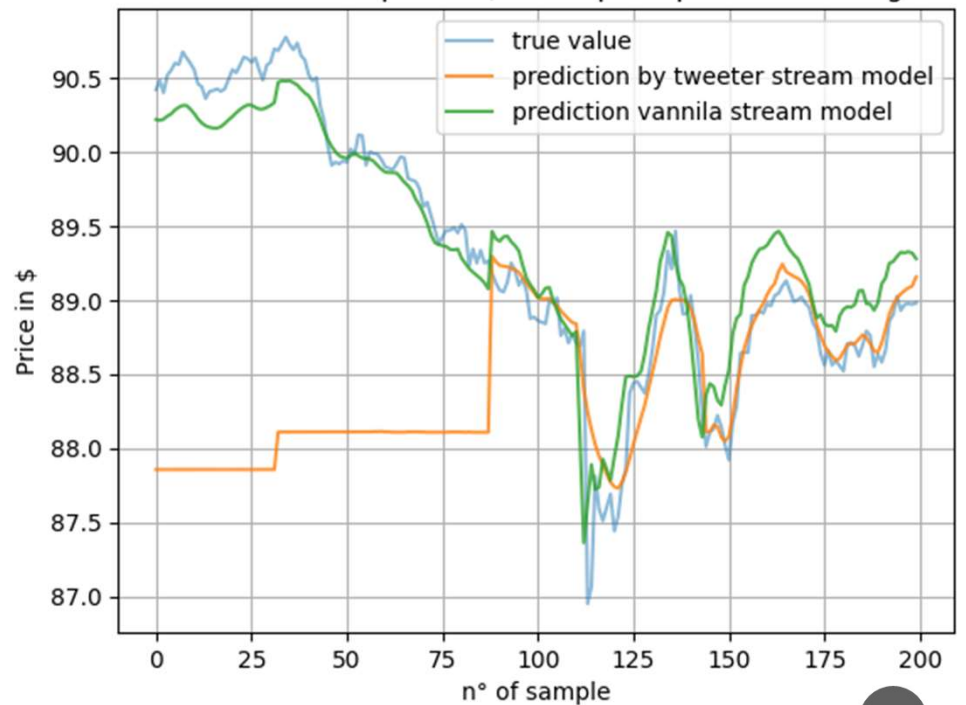
## IV. RÉSULTATS SIMULATIONS

### COMPARAISON STREAM MODELS VS STREAM MODEL WITH TWITTER

\$GOOG price prediction and groundtruth



zoom on the 100 last samples of \$GOOG price prediction and groundtruth



# BILAN

## Bilan du projet

- KAFKA nous a permis d'implémenter un algo de stock forecasting
- Les modèles de stream learning sont plus robustes dans le temps
- L'ajout de sentiment twitter est prometteuse

## Leçons apprises

- Un projet où on s'est fait plaisir
- Répartition simple du travail
- Une cible difficile à définir: les données de masse sont payantes

## Next step

- Affiner l'utilisation de twitter: NLP et intégration dans le learning
- Avoir de meilleures données (API)



## REFERENCES

Data stream learning :

Singh, T., Kalra, R., Mishra, S. *et al.* An efficient real-time stock prediction exploiting incremental learning and deep learning. *Evolving Systems* (2022). <https://doi.org/10.1007/s12530-022-09481-x>

Bourse et twitter:

- Yamina Tadjeddin, 2013, « La finance comportementale, une critique cognitive du paradigme classique de la finance » *Idées économiques et sociales* 2013/4 (N° 174), pages 16 et suivantes <https://www.cairn.info/revue-idees-economiques-et-sociales-2013-4-page-16.htm>

- Malcolm Baker & Jeffrey Wurgler, 2007. "Investor Sentiment in the Stock Market," NBER Working Papers, National Bureau of Economic Research, Inc.

- Jean-Christophe Feraudet, 2020. « Analyse de Twitter en temps réel avec Kafka, Spark et mongoDB », [https://cedric.cnam.fr/vertigo/Cours/RCP216/docs/UASB03\\_Projet\\_Feraudet\\_v1.0.pdf](https://cedric.cnam.fr/vertigo/Cours/RCP216/docs/UASB03_Projet_Feraudet_v1.0.pdf)

- le CNAM