

# Détection et monitoring des fintechs

---

Data Science & IA (DDSA-ILIADE-INNOVIA)  
DGSEI -DSMF  
Le LAB

*Commission Cnis « Système financier et financement  
de l'économie » – 19 mai 2022*

## Le besoin d'un meilleur suivi des fintechs

Les besoins identifiés pour la DGSEI :

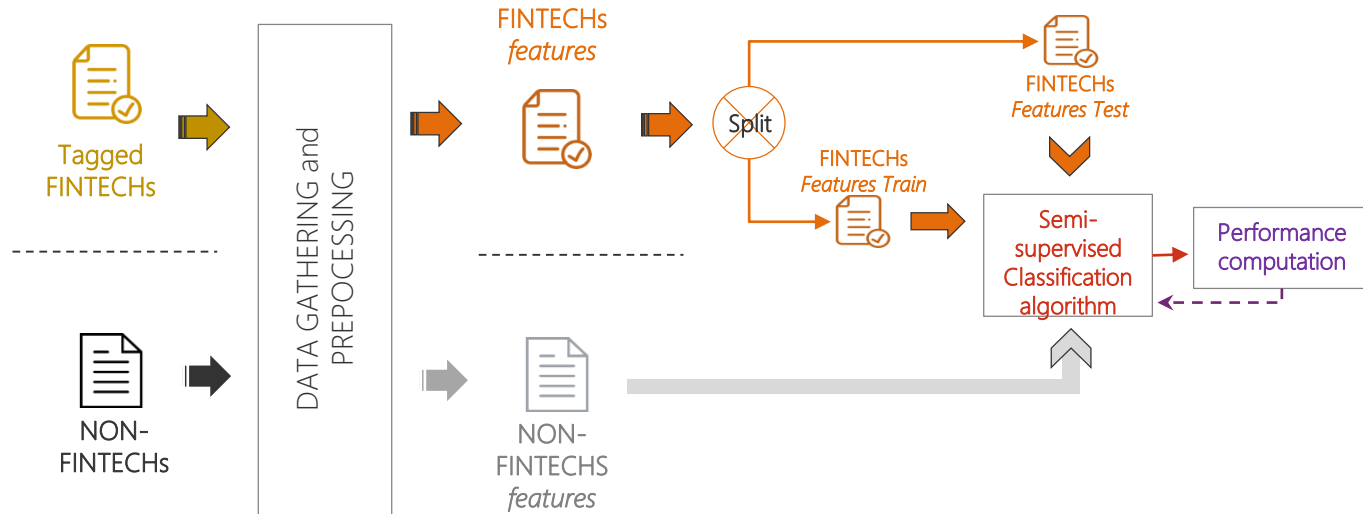
- **répondre aux demandes des utilisateurs de données (économistes et chercheurs...)** de jeux de données de référence pour traiter les problématiques fintech telles qu'abordées par exemple au FSB, à l'ESRB (ex : shadow banking)....
  - Par exemple : analyse du shadow banking, détermination de la part des fintechs dans les flux financiers (ex : paiements ou octroi de crédit), organisation capitalistique du secteur des fintechs, détermination des « cycles de vie » des fintechs
  - Les fintechs figurent parmi les sujets identifiés dans la Data Gap Initiative initiée par le G20 et dont la revue est menée par le FSB et le FMI
  
- De proposer, dans le processus de **révision des comptes nationaux** mené par la division statistique des Nations Unies, des évolutions liées à une meilleure prise en compte des fintechs

## Le problème : les fintechs ne sont pas précisément définies

- Rapport de l'Irving Fisher Committee, Juillet 2020, [Towards monitoring financial innovation in central bank statistics](#)
    - Définition des fintechs : "*innovation and digitalisation in the financial sector*"
  - Les fintechs ne sont pas précisément identifiées dans les classification de comptabilité nationale → pas de definition opérationnelle qui repose sur des critères concrets
  - Par definition, la notion de fintech est évolutive
- Besoin d'un outil dynamique, à jour, d'identification et de suivi des fintechs

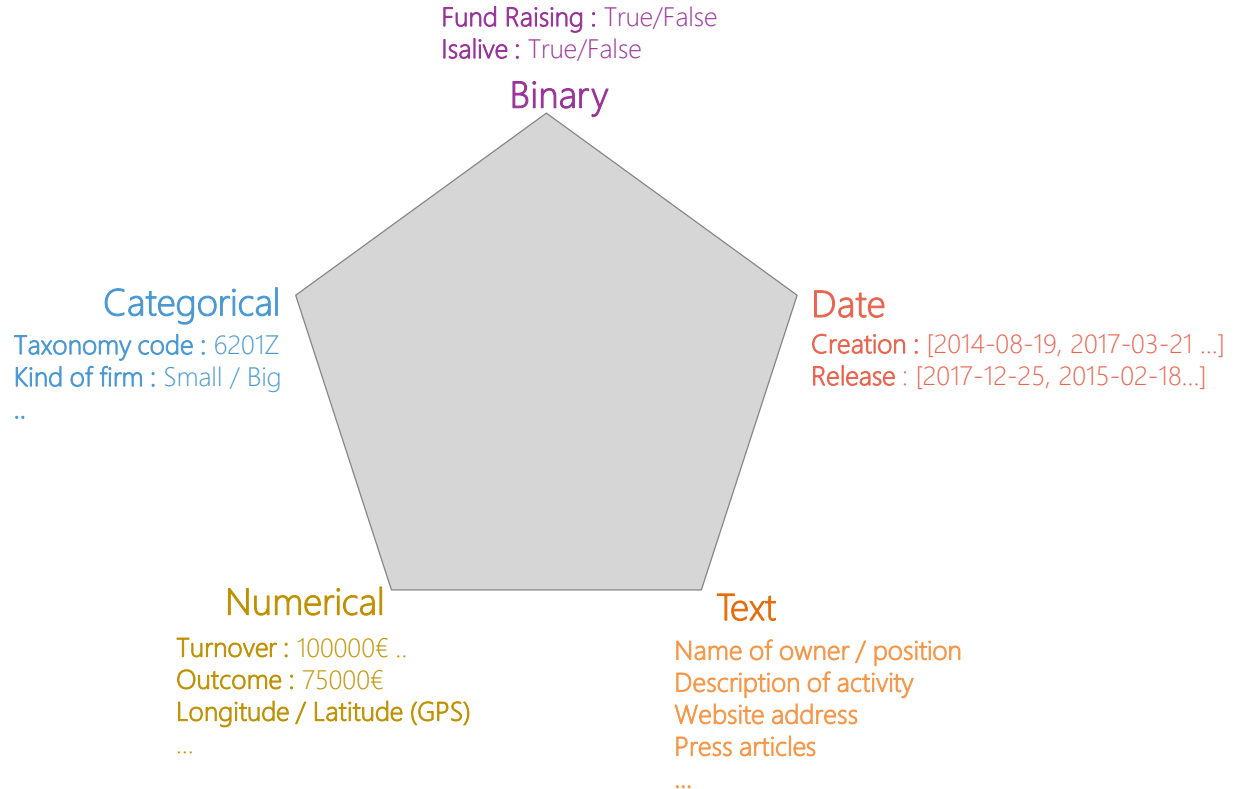
# Un algorithme semi-supervisé

- A partir d'une liste de fintechs déterminée à dire d'expert et d'une grosse base de données des entreprises françaises, on veut obtenir une base de données des fintechs en France qui puisse se rafraîchir de manière fréquente
- On veut aussi pouvoir surveiller des entreprises qui pourraient « pivoter » ou émerger en tant que fintech

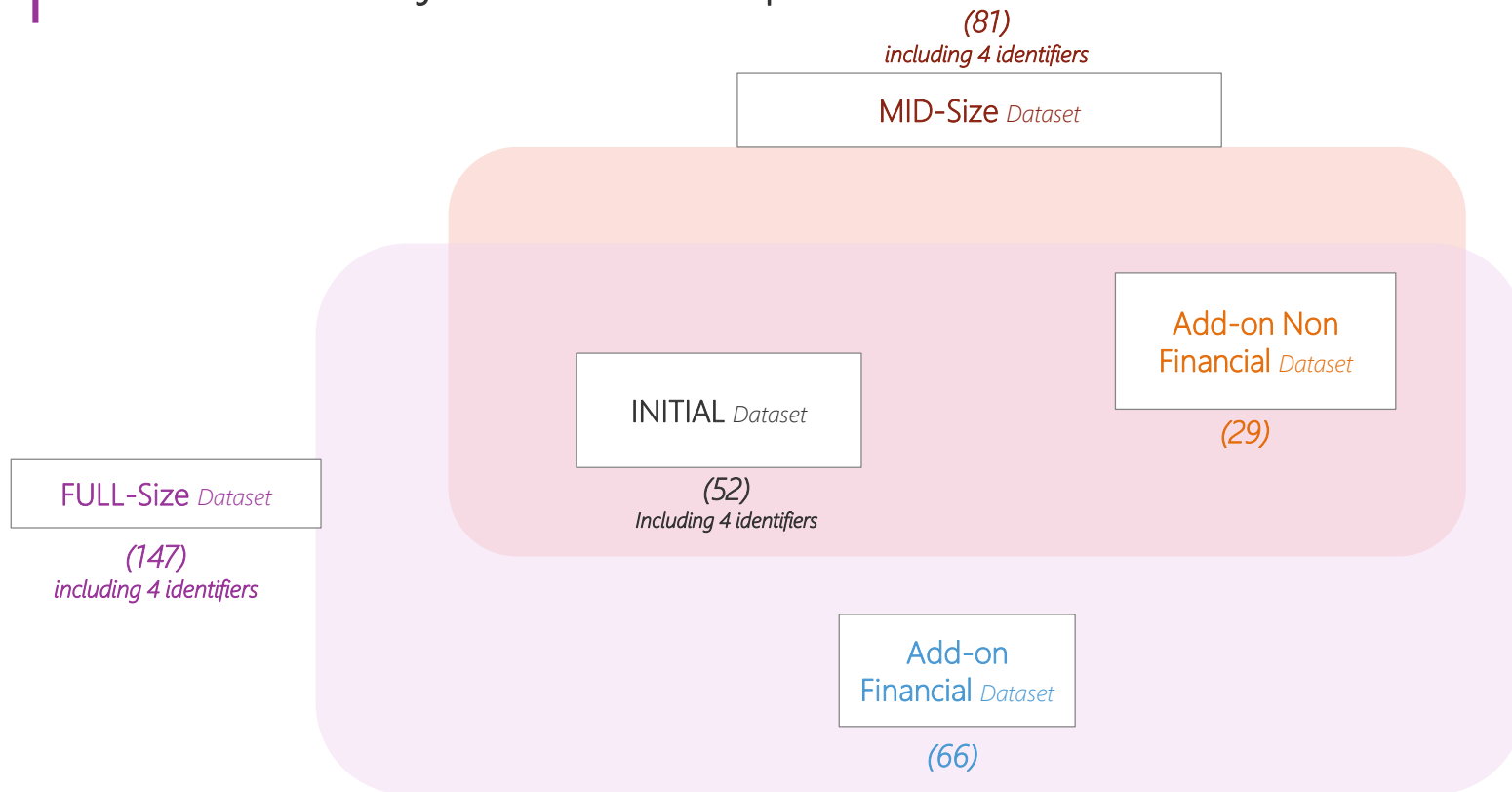


# Une base de données riche et diverse

- Recours à un prestataire externe :
  - 10 millions d'entreprises réduites à un sous-ensemble de 10 000 entreprises sélectionnées de façon aléatoire
  - Environ 1400 variables issues de bases de données légales, webcrapping, autres bases de données (ex : presse) etc...
- 1er travail de *featurisation*

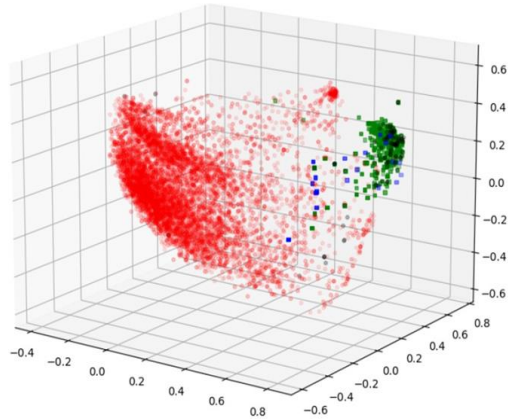


## L'identification du jeu de données optimal



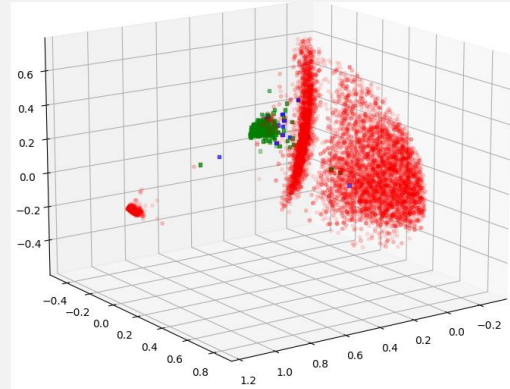
## Résultats pour l'algorithme "standard"

INITIAL



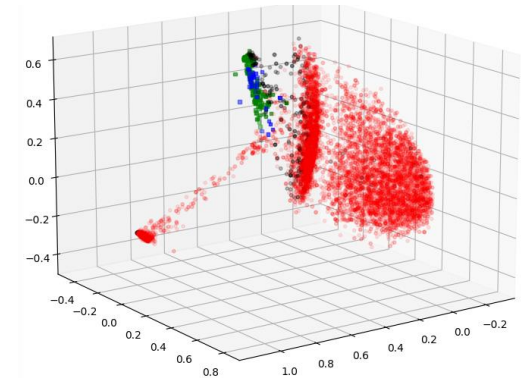
Accuracy = 98,9%  
Recall = 97,2%  
Precision = 38,3%

MID-Size



Accuracy = 99,6%  
Recall = 95,8%  
Precision = 68,7%

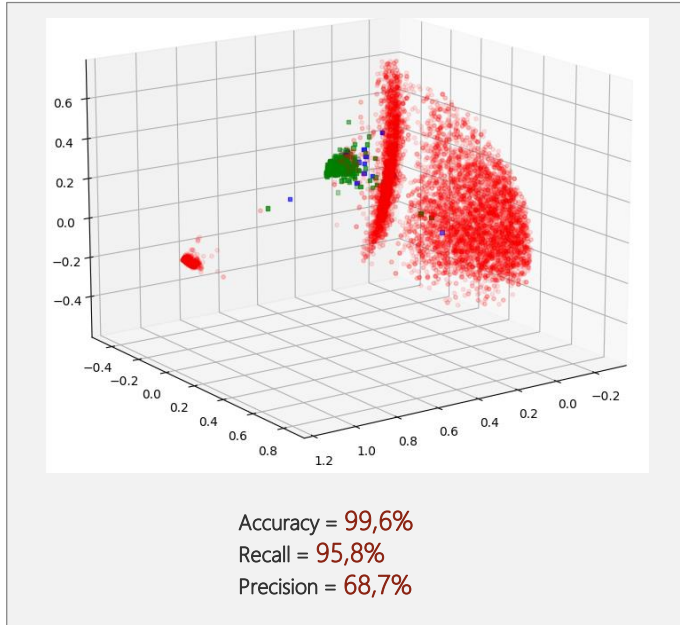
FULL-Size



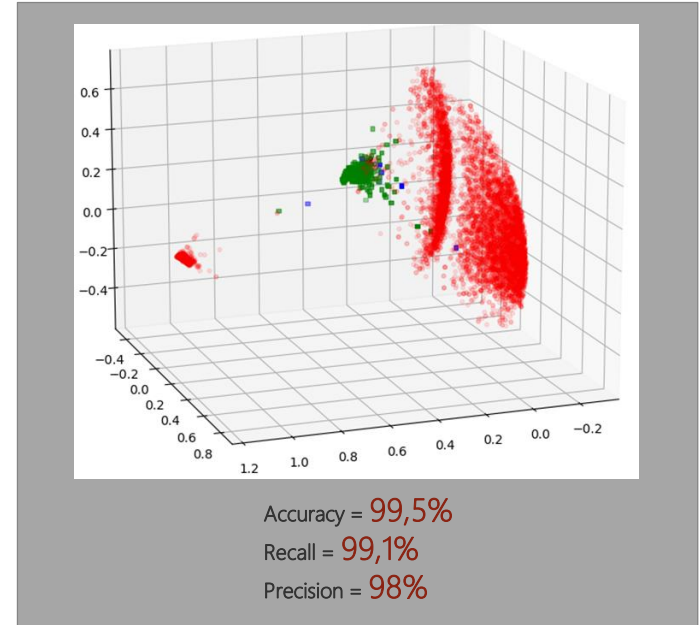
Accuracy = 96,4%  
Recall = 91,6%  
Precision = 15,3%

- Les résultats les meilleurs sont obtenus pour le dataset intermédiaire MID-Size
- Toutefois la **précision** reste relativement faible en raison du déséquilibre au sein du jeu de données (peu de fintechs et beaucoup d'entreprises non-fintechs)

# Résultats pour un algorithme avec stratégie de génération synthétique



MID-Size



La réduction du déséquilibre par le suréchantillonnage et la génération synthétique a permis de :

- mieux ajuster et sélectionner notre modèle
- mieux calculer la précision, le *recall* et l'*accuracy*



### Les résultats métiers

#### Les critères qui caractérisent le mieux les fintechs dans notre base de données :

- Variables issues des articles de presse : thème, description, source des articles
- Variables administratives : code activité, taille
- Variables liées au type d'emplois et de compétences au sein de l'entreprise
- Noms des personnes ou des fonds au Conseil d'administration des entreprises
- Secteur de la marque déposée

#### Les enseignements tirés de cette expérimentation :

- Une des plus grandes difficultés est de prendre en compte le volume, la variété et la qualité des données
- L'obtention de toutes les données pour la grande base de données sur les entreprises est une étape préliminaire, extrêmement importante, mais qui requiert aussi beaucoup de temps et de moyens
- Pour une extension en production, notamment à un niveau européen, voire mondial, il faudra des infrastructures très robustes vu le volume de données à traiter

## 2) En cours, à la BdF (sous réserve de l'obtention du budget), création d'un outil en production

L'obtention d'un budget pour obtenir la base de données des entreprises sur un an permettrait de :

- Déployer l'algorithme obtenu sur une **base plus large**
- Tester l'**aspect dynamique** de l'outil i.e. la capacité de l'outil à être à jour facilement, à détecter de nouvelles fintechs (même dans des domaines plus lointains que ceux initialement identifiés)
- En fonction des besoins, créer une **interface utilisateurs** qui remonterait les informations les plus utiles pour le monitoring des fintechs

Eventuellement, il pourrait être intéressant d'associer à l'outil de **nouvelles données** (ex : données financières) ou bien de chercher à **obtenir par nous-mêmes certaines données** sans passer par un prestataire externe

### 3) Mise à l'échelle de l'algorithme au niveau européen via un réseau de banques centrales

- Ce réseau permettrait de bien spécifier la **structure de la base de données et des sources** ce qui garantira une **pérénnité des sources de données et des méthodes au fil du temps**

*NB : La Bundesbank avec l'aide d'une start-up berlinoise a conduit une experimentation avec une approche par "graphe" et a obtenu de très bons résultats sur les fintechs allemandes. Cette méthode pourrait être complémentaire à celle de la Banque de France.*

→ Cela permettrait la mise en place de registres nationaux interconnectés sur les fintechs, les business models des fintechs étant souvent d'emblée internationaux

- Ce mode de création de statistiques publiques pourrait ensuite donner lieu à d'autres applications

Merci pour votre attention