

# La technologie au service de la surveillance des marchés

Par Iris LUCAS

Autorité des Marchés financiers (AMF)

Fragmentation des marchés, électronique des échanges, développement du *trading* algorithmique et haute fréquence : les marchés ont été marqués par de nombreuses mutations au cours des dernières années. Ces phénomènes se sont accompagnés d'un accroissement considérable de la volumétrie des données de marché et ont poussé les régulateurs à moderniser en profondeur leurs systèmes de détection d'abus de marché. Aujourd'hui, la technologie est indéniablement au cœur de la surveillance des marchés, mais elle reste un instrument au service de l'expertise humaine, à tous les niveaux, de la construction des algorithmes à l'analyse des cas d'abus potentiels.

## Introduction

En 2007<sup>(1)</sup>, la libéralisation des monopoles que détenaient les bourses historiques dans la négociation des instruments financiers s'est traduite par l'émergence de plateformes de négociation alternatives. Pour offrir des cotations sur les instruments toujours plus compétitives par rapport à leurs concurrents, ces plateformes ont diminué les pas de cotation<sup>(2)</sup> des titres, augmentant la dispersion des prix possibles.

La fragmentation du marché entre différentes plateformes et la granularité accrue des prix possibles ont multiplié les messages<sup>(3)</sup> à prendre en compte dans les stratégies d'exécution (routage des ordres) et de *trading*, poussant les acteurs à davantage d'automatisation. Ces phénomènes sont à l'origine du développement du *trading* algorithmique, en particulier des *traders* haute fréquence, au cours des dix dernières années. Dans ce contexte, les régulateurs ont naturellement dû s'adapter afin de pouvoir assurer leur mission de surveillance des marchés.



En moyenne par mois sur les valeurs du CAC40 sur Euronext Paris

Illustration 1 : Volumétrie mensuelle d'une partie des données collectées par le régulateur français au troisième trimestre 2018 (source : AMF).

Dans le cadre de sa mission de surveillance, le régulateur français, l'Autorité des Marchés financiers (AMF), collecte un grand nombre de données de marchés en provenance de différentes sources. Ces sources comprennent, outre des bases de données publiques (*data vendors*, tels que ThomsonReuters ou Bloomberg), des Bourses, dont l'AMF reçoit les données sur les transactions, mais aussi sur les ordres (ceux-ci ne donnant pas toujours lieu à une exécution, mais pouvant influencer sur les prix), ainsi que des intermédiaires de marché (les professionnels assujettis à la réglementation européenne<sup>(4)</sup>). L'AMF reçoit ainsi un *reporting* très vaste des activités réalisées sur les valeurs françaises et leurs dérivés (listés ou non), que ces transactions aient lieu sur le marché directeur (Euronext Paris),

(1) Le 1<sup>er</sup> novembre 2007, l'entrée en vigueur de la première directive européenne concernant les Marchés d'instruments financiers (MIF1) autorisait les prestataires de services d'investissement à fragmenter le flux d'exécution de leurs ordres d'achat et de vente sur différents systèmes de négociation (création des statuts de systèmes multilatéraux de négociation et des internalisateurs systématiques).

(2) Le pas de cotation (ou tick size) est l'écart minimum autorisé entre deux prix proposés dans le carnet d'ordres.

(3) Le terme « message » fait ici référence à tout événement observé sur les carnets d'ordres : entrée d'un nouvel ordre, modification d'un ordre existant, exécution ou annulation de celui-ci. Un ordre connaît donc au moins deux événements : son entrée dans le carnet et sa sortie, soit par une exécution, une annulation ou une expiration. Une transaction correspond à l'exécution de deux ordres (d'achat et de vente).

(4) Dans le cadre des obligations imposées par le règlement européen MIFIR, l'AMF reçoit des prestataires français, leur reporting de transactions, et, des régulateurs européens, celui des intermédiaires étrangers lorsque leurs activités concernent des instruments français. De la même façon, l'AMF redirige le reporting des prestataires français vers ses homologues concernés, lorsque celui-ci implique des valeurs étrangères.

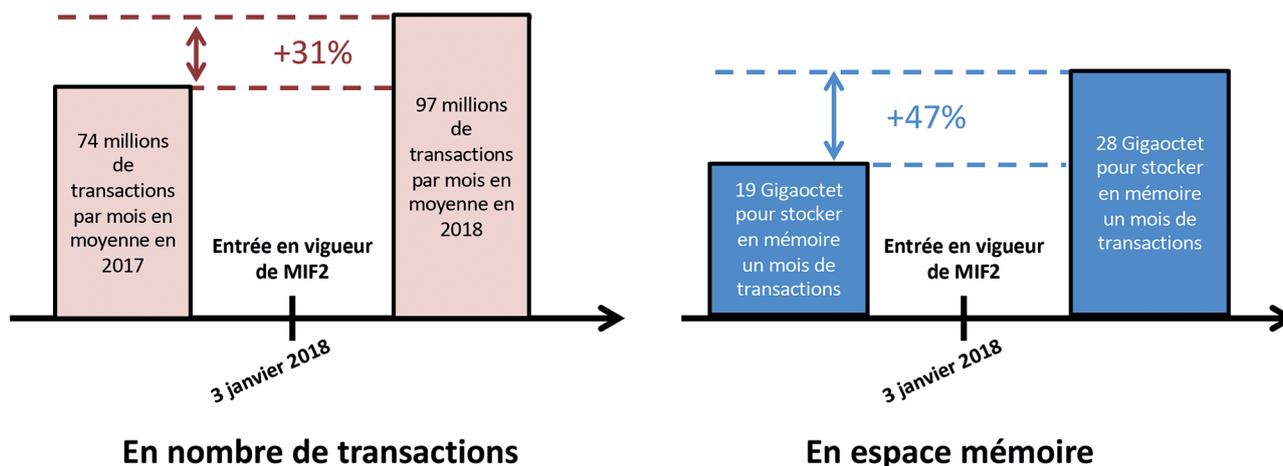
Evolution de la volumétrie du *reporting* de transactions

Illustration 2 : Impact MIF2 sur l'espace mémoire requis pour collecter le *reporting* des transactions des déclarants professionnels (source : AMF).

les plateformes alternatives (y compris celles situées au Royaume-Uni) ou en OTC.

Depuis le 3 janvier 2018, la volumétrie de données reçues dans le cadre de l'obligation de *reporting* s'est particulièrement accrue avec l'entrée en vigueur de la nouvelle directive sur le marché des instruments financiers (MIF2). En effet, le règlement MIFIR associé à cette directive est venu à la fois étendre le périmètre des instruments financiers sur lesquels les assujettis ont une obligation de *reporting*, et enrichir l'information sur les clients finaux ainsi que sur les personnes (ou les algorithmes) impliqués dans la chaîne de décision et d'exécution d'une transaction.

### La surveillance des marchés passe forcément par un traitement numérique de l'information

L'ensemble de ces données constitue une véritable mine d'informations pour l'AMF dans le cadre de sa mission de surveillance, notamment pour la recherche d'abus de marché. Dans ce but, le régulateur français a développé son propre système de détection à partir d'un ensemble d'algorithmes qui analysent les données de marché collectées et génèrent des alertes en cas de situation pouvant constituer un manquement ou présenter un risque d'abus de marché (tel que défini dans le règlement européen MAR relatif aux abus de marché).

De la même façon que les évolutions structurelles du marché ont accru l'automatisation des stratégies de *trading* et d'exécution chez de nombreux investisseurs professionnels, la volumétrie croissante des données reçues et la variété des schémas manipulatoires potentiels ont conduit l'AMF à automatiser toujours davantage ses outils de détection, de façon à ce que sa plateforme de surveillance reste à la pointe.

Les algorithmes développés pour la surveillance peuvent à la fois détecter une séquence atypique selon différents critères (volumes anormaux, variation de prix brutale, forte

volatilité, etc.), un schéma de manipulation assez spécifique tel que le *layering*<sup>(5)</sup> ou une configuration caractéristique de délit d'initié.

L'ensemble de ces algorithmes reposent sur des traitements coûteux en ressources machine, comme le croisement de différentes sources de données, des calculs itératifs ou encore des calculs à une échelle de temps très réduite. Par exemple, afin de rechercher d'éventuelles séquences de type *layering*, l'état du carnet d'ordres est recalculé à chaque événement<sup>(6)</sup> de la séance (pour une valeur du CAC40, on compte actuellement 17 événements par seconde en moyenne, mais cette fréquence peut monter jusqu'à 2 426 événements par seconde).

Autre exemple, la recherche d'un délit d'initié requiert la prise en compte de l'ensemble des activités des acteurs : les transactions passées sur le marché au comptant (actions, obligations, etc.), mais également celles sur les dérivés (options, *futures*, CFD, *equity swap*, etc.), tous lieux d'exécution confondus (marchés réglementés, systèmes multilatéraux de négociation, internalisateurs systématiques, marché de gré à gré, etc.). Or, le schéma de *reporting* est souvent complexe, notamment parce que les investisseurs peuvent passer par plusieurs intermédiaires, qui eux-mêmes peuvent déléguer l'exécution de l'ordre client auprès d'autres intermédiaires. La difficulté consiste alors à reconstituer les chaînes d'intermédiation afin d'identifier les bénéficiaires finaux, et non les intermédiaires n'ayant fait que transmettre les ordres de ces clients.

(5) Le *layering* est un schéma manipulateur dans lequel un acteur provoque un déséquilibre du carnet d'ordres au moyen d'ordres légers afin de simuler un intérêt fictif (par exemple, à l'achat) et d'influencer les cours (à la hausse) avant d'en tirer profit au moment de l'exécution en se positionnant dans le sens inverse du déséquilibre créé (en vendant les titres).

(6) À l'instar du terme « message », un événement fait référence à l'entrée d'un nouvel ordre, à la modification d'un ordre existant, à une exécution ou une annulation de celui-ci.

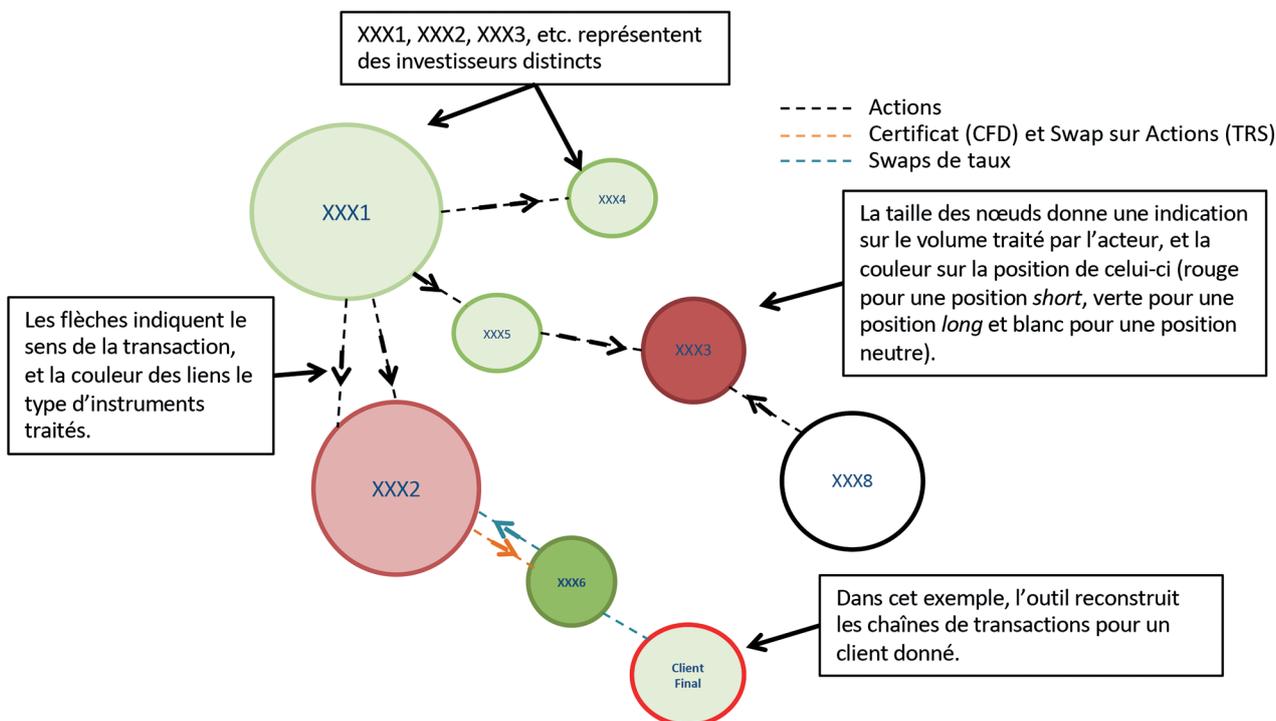


Illustration 3 : Exemple d'un outil interne de visualisation des chaînes d'intermédiation (source AMF).

## Les algorithmes sont au cœur des défis de la surveillance, mais ne supplantent pas l'analyse humaine

Si les traitements numériques sont désormais essentiels pour assurer la détection des manquements, ceux-ci ne peuvent pas supplanter l'analyse humaine pour confirmer le caractère frauduleux d'un comportement. Pour cela, les équipes s'appuient sur des outils développés en interne qui proposent le calcul d'un large panel d'indicateurs comme le *price impact*<sup>(7)</sup> d'un acteur ou l'évolution de

sa position et de ses gains et pertes au cours du temps. Les analystes ont aussi à leur disposition une librairie de graphiques permettant de visualiser événement par événement l'évolution d'un carnet d'ordres ou d'afficher les opérations d'un acteur à différentes échelles de temps (en *intraday* ou sur plusieurs jours).

Par ailleurs, pour mener à bien ses missions, la Surveillance investit également sur la qualité de la donnée. En effet, une grande partie des données collectées sont issues des différents intermédiaires de marché, constituant autant de sources potentielles d'erreur de déclaration (chaque intermédiaire ayant souvent développé ses propres systèmes de *reporting*). Ces erreurs peuvent aller du « gros doigt » affectant le montant traité à l'absence de déclaration, en passant par une erreur sur la date et l'heure de la transaction. Dans le cadre de la détection d'abus de marché, la qualité de la donnée est un enjeu crucial, dans la mesure où les erreurs de déclaration viennent fausser le résultat des algorithmes et peuvent amener à détecter une séquence qui n'aurait pas dû l'être. Ce type de « faux positif » est un moindre mal comparé à la situation inverse, où le système ne détecterait pas un comportement suspect par manque de données ou à cause de déclarations inexactes. Là encore, l'AMF a systématisé un traitement numérique sur les flux collectés pour remonter aux équipes les données qui seraient susceptibles d'être erronées. Ces alertes sont ensuite analysées manuellement, et, lorsqu'une erreur est avérée, une demande de correction est envoyée au déclarant.

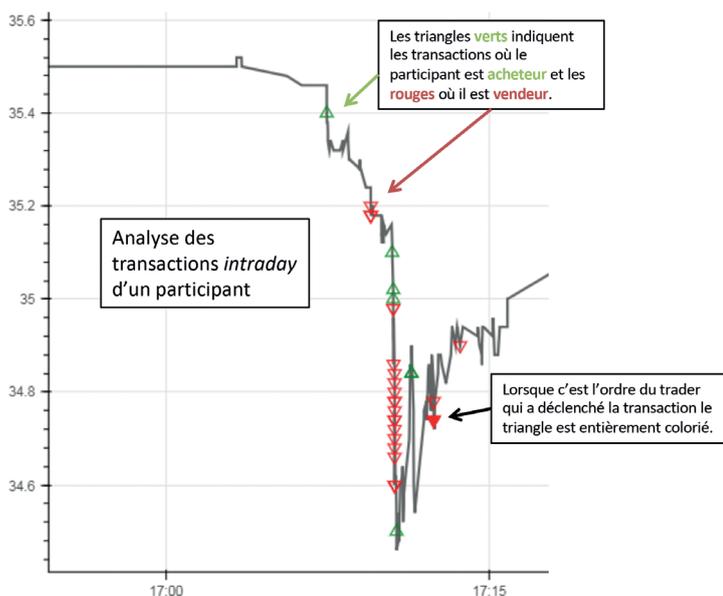


Illustration 4 : Exemple d'un outil interne de visualisation servant à l'analyse des activités d'un participant, transaction par transaction (source : AMF).

(7) Impact d'un ordre agressif sur le prix d'une valeur.

## Les systèmes de la surveillance des marchés doivent rester à la pointe de la technologie

Comptant parmi les grands régulateurs internationaux ayant fait très tôt le pari d'investir massivement sur les données de marché, l'AMF a eu à cœur de maintenir ses outils à la pointe des innovations technologiques devant lui permettre de tirer pleinement profit des données que lui procurent les nouvelles réglementations. Ainsi, le régulateur français a lancé en 2016 un ambitieux projet de refonte de ses systèmes IT, notamment de sa nouvelle plateforme de surveillance, ICY (« *I see why* »), qui s'appuie sur une technologie *Big Data*. La refonte des infrastructures est non seulement l'occasion d'étendre les capacités de stockage et de calcul, mais aussi, et surtout, de les rendre évolutives et de les adapter au gré des besoins.

Immergés dans un écosystème *Hadoop*, les *data scientists* de la surveillance profitent d'un environnement distribué qui permet de gérer en parallèle le stockage de la donnée et les calculs. Entre des développements *python* pour la partie expérimentale (nouvel algorithme de détection ou analyse *ad hoc*) et *spark* pour l'industrialisation des traitements, les équipes de la surveillance s'organisent pour tirer le meilleur de ce nouvel environnement. Les défis sont multiples comme profiter de la puissance de calcul pour aller plus loin dans les algorithmes de détection d'abus de marché ou repenser les modèles de stockage des données et choisir les technologies les plus adaptées pour optimiser les traitements. À ce sujet, la Surveillance étudie les différentes possibilités *via* les produits intégrés dans la distribution *Hadoop*, tels que *Apache Hive* et *Apache Phoenix*.

Le régulateur français mène également des expérimentations pour tirer profit des possibles applications de l'intelligence artificielle, comme l'utilisation d'algorithmes de type *machine learning* dans ses métiers. Les premiers retours d'expérience dans la *regtech* (les nouvelles technologies au service de la régulation) fournissent d'intéressantes perspectives, comme celles identifiées par l'Autorité des marchés financiers du Québec qui rapporte avoir mis en place un algorithme d'auto-apprentissage supervisé pour détecter des transactions ne respectant pas les exigences obligatoires en matière de négociation et de compensation<sup>(8)</sup> (Board F. S., 2017). Autre projet qui a pris la forme d'une expérimentation, la Security Exchange Commission (SEC) a cherché à savoir s'il aurait été possible de prédire la crise de 2008 à partir d'analyses textuelles des articles mentionnant des contrats *Credit Default Swaps* (CDS) (BAUGUESS, 2017). Les résultats des programmes de traitement du langage naturel, utilisés par le régulateur américain pour cette étude, ont effectivement montré que juste avant la crise de 2008, le nombre d'articles parlant de CDS avait été multiplié par 10.

L'intelligence artificielle au service des régulateurs n'est plus totalement une utopie : les projets sont prometteurs, mais encore rares, car ils sont généralement coûteux en termes de temps, de ressources, mais aussi de compétences. En 2017, le Financial Stability Board (FSB) a publié sa revue des applications existantes et à venir de l'intelligence artificielle dans la finance de marché (Board F. S., 2017). Le rapport couvre tous les domaines du secteur, de la stratégie de *trading* pour les *hedges funds* à la détection de fraude pour les régulateurs et les services de conformité, en passant par les activités du *back-office*. Pour la surveillance des marchés en particulier, les applications possibles sont nombreuses : *clustering*<sup>(9)</sup> des participants de marché afin d'établir leur typologie et faciliter leur analyse, apprentissage supervisé pour la qualité de la donnée, analyse des messages des réseaux sociaux pour détecter d'éventuelles manipulations accompagnées de la diffusion de fausses informations, identification des collusions entre acteurs pour détecter d'éventuels manquements opérés par des réseaux ou bandes organisées, etc.

Pleine de promesses, l'intelligence artificielle est encore à l'état d'expérimentation chez les régulateurs et il faudra certainement attendre quelques années encore pour assister à sa démocratisation dans le secteur de la *regtech*. L'intelligence artificielle ne viendra certainement pas remettre en cause les métiers aujourd'hui impliqués dans la surveillance des marchés, mais elle offre de véritables perspectives en termes de richesse d'analyse. Convaincue que ces nouvelles technologies seront au cœur de ses missions de demain, l'AMF se lance, elle aussi, dans la course vers la surveillance 2.0 au service de l'intégrité et du bon fonctionnement des marchés.

## Bibliographie

BAUGUESS S. W. (2017), *The Role of Big Data, Machine Learning, and AI in Assessing Risks: a Regulatory Perspective*.

BOARD F. S. (2017), "Artificial intelligence and machine learning in financial services", November, available at: <http://www.fsb.org/2017/11/artificialintelligence-and-machine-learning-in-financialservice/> (accessed 30<sup>th</sup> January, 2018).

(8) AMF (2017), "AMF creates Fintech lab and signs partnership with R3", press release, April.

(9) Le clustering est une des méthodes d'analyse qui visent à diviser un ensemble de données en différents groupes homogènes de façon à ce que chaque élément d'un sous-ensemble partage des caractéristiques communes.