Analyse et Détection d'Activités Frauduleuses sur des Séries Temporelles d'Engagement

Inès Gbadamassi
École Sup Galilée
ines.gbadamassi@edu.univ-paris13.fr

Détection d'anomalies · Séries Temporelles · Machine Learning · Simulation réaliste

Résumé

Ce rapport présente un système complet d'analyse et de détection d'activités frauduleuses sur des séries temporelles d'engagement inspirées des plateformes sociales. J'ai conçu un simulateur générant plusieurs profils d'utilisateurs et différents types d'attaques manipulant artificiellement les interactions. J'ai ensuite mené une analyse exploratoire approfondie et développé un pipeline de détection reposant sur des modèles non supervisés tels que l'Isolation Forest et l'Autoencoder. Les résultats ont montré que la fraude induisait des perturbations mesurables dans les distributions de valeurs, la dynamique temporelle et la corrélation entre métriques, permettant une détection partielle malgré la difficulté du dataset.

1 Introduction

Les plateformes des réseaux sociaux s'appuient sur les signaux d'engagement afin de classer et recommander le contenu aux utilisateurs. Cette dépendance crée un environnement propice aux manipulations : génération automatisée de vues, achats de likes ou comportements coordonnés visant à simuler une viralité artificielle.

J'ai choisi de réaliser ce projet pour étudier comment ces manipulations modifient les séries temporelles d'engagement, et comment des modèles d'apprentissage pouvaient contribuer à les détecter.

Pour cela, j'ai conçu un simulateur d'engagement réaliste, analysé les signatures caractéristiques de la fraude et développé un pipeline complet d'anomaly detection destiné à distinguer les comportements authentiques des comportements artificiels (= fake ou truqués).

2 Simulateur d'engagement

2.1 Profils d'utilisateurs

J'ai commencé par créer plusieurs profils authentiques reflétant la diversité des comportements humains : des utilisateurs réguliers suivant un cycle jour/nuit, des profils dits impulsifs présentant des pics ponctuels, d'autres profils dormants avec une activité faible, des influenceurs à forte amplitude et des encore nouveaux comptes. Ces profils m'ont permis de reproduire un large éventail de signaux réalistes.

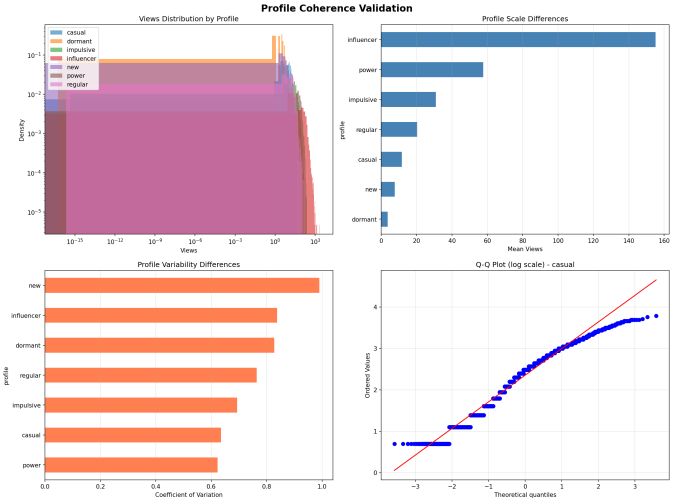

FIGURE 1 – Profils d'utilisateurs simulés : régulier, impulsif, dormant, influenceur et nouveau compte.

Cette figure illustre les cinq profils authentiques générés par le simulateur. Chaque profil présente une dynamique temporelle distincte : le régulier suit un rythme circadien stable, l'impulsif montre des pics isolés mais organiques, le dormant reste faiblement actif, l'influenceur exhibe de fortes amplitudes et le nouveau compte présente une activité plus erratique. Ces différences servent de base pour contraster les comportements normaux et frauduleux au cours de notre étude.

2.2 Types d'attaques frauduleuses

J'ai ensuite intégré plusieurs attaques visant à manipuler artificiellement l'engagement : croissances progressives, pics d'activité irréguliers, vagues répétées, superpositions ou synchronisations extrêmes entre métriques. Chaque attaque possédait une signature propre affectant la dynamique temporelle des séries.

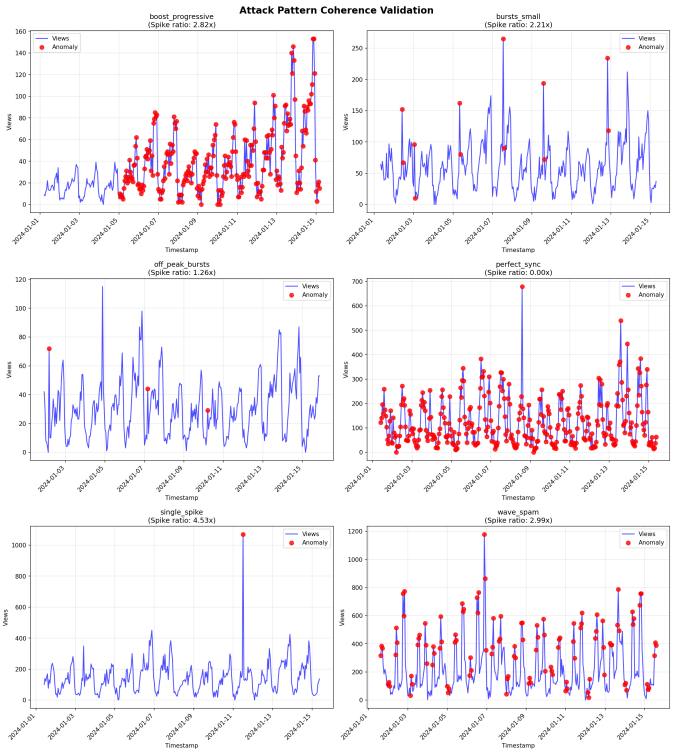

FIGURE 2 – Exemples d'attaques simulées modifiant artificiellement les séries temporelles.

Cette figure présente plusieurs types d'attaques que j'ai intégrées au simulateur afin de reproduire des comportements frauduleux réalistes. Chaque attaque modifie différemment la dynamique d'engagement : certaines introduisent des pics soudains et incohérents avec le rythme normal, d'autres imposent une croissance artificielle sur plusieurs jours, tandis que certaines opèrent par vagues régulières ou par synchronisation suspecte entre plusieurs métriques.

3 Analyse exploratoire

J'ai mené une analyse exploratoire afin d'identifier les propriétés qui distinguent les signaux authentiques des signaux frauduleux. Cette étape a été essentielle pour comprendre comment la fraude altérait les données et quelles dimensions étaient les plus informatives pour la détection.

3.1 Distribution globale des valeurs

J'ai d'abord examiné la distribution des vues. Les séries authentiques suivaient une structure lognormale caractéristique, avec une forte concentration dans les faibles valeurs et une décroissance progressive vers les extrêmes. Les séries frauduleuses présentaient au contraire une densité plus élevée dans les valeurs importantes, conséquence directe des manipulations introduites. Cette différence marquée constituait un premier indice révélant l'impact statistique de la fraude.

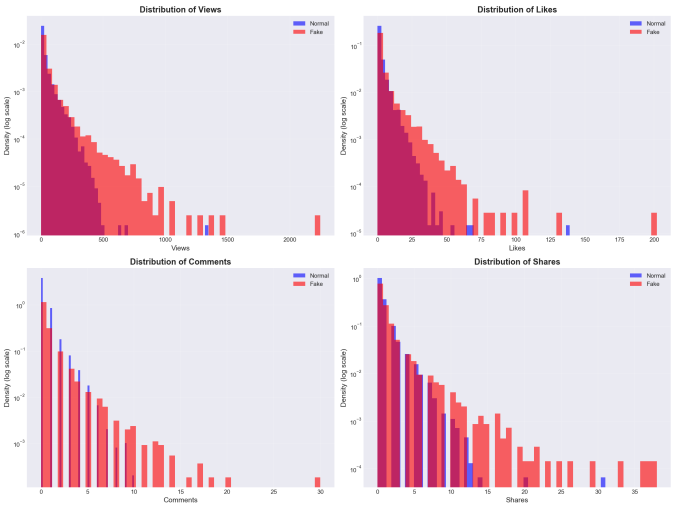

FIGURE 3 – Distribution globale des vues : la fraude augmente anormalement la fréquence des valeurs élevées.

3.2 Analyse temporelle

J'ai ensuite étudié l'évolution des séries dans le temps. Les signaux authentiques respectaient des cycles réguliers, notamment une alternance claire entre les périodes d'activité diurne et nocturne. À l'inverse, les signaux frauduleux présentaient des ruptures soudaines, des augmentations monotones ou des pics sans justification comportementale. Ces perturbations temporelles constituaient un indicateur robuste pour la détection.

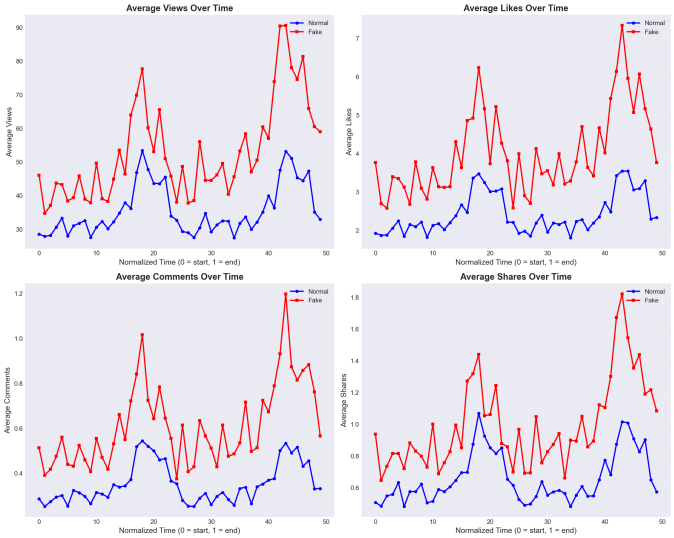

FIGURE 4 – Comparaison temporelle : stabilité organique (haut) contre rupture artificielle (bas).

3.3 Dépendances entre métriques

J'ai enfin étudié la corrélation entre les métriques (vues, likes, commentaires). Dans les séries normales, les interactions progressaient de manière cohérente mais non proportionnelle. Les séries frauduleuses montraient au contraire une synchronisation excessive, signe d'une amplification co-

ordonnée difficilement attribuable à un comportement humain. Cette sur-corrélation constituait un marqueur fort de manipulation.

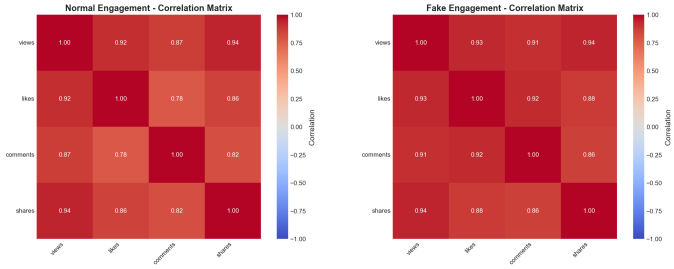

FIGURE 5 – Corrélations entre métriques : structure organique (gauche) contre synchronisation artificielle (droite).

3.4 Synthèse

Cette analyse m'a permis de dégager trois signatures majeures de fraude : une modification de la distribution statistique, une rupture des cycles temporels naturels et une corrélation anormalement élevée entre signaux. Ces observations ont guidé la conception du pipeline d'apprentissage.

4 Pipeline de détection

J'ai ensuite développé un pipeline combinant extraction de descripteurs, réduction de dimension et modèles non supervisés.

4.1 Représentation et embedding

J'ai extrait des indicateurs statistiques (moyennes glissantes, variance), temporels (autocorrélation) et structurels. J'ai ensuite utilisé une PCA pour analyser la séparation naturelle entre les comportements.

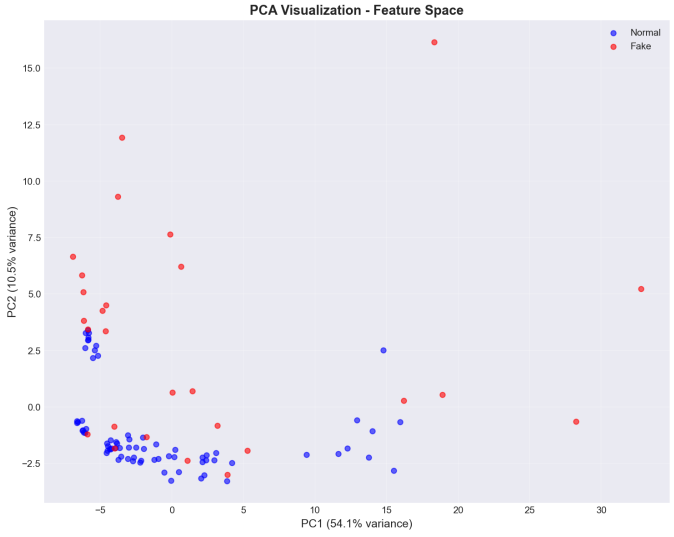

FIGURE 6 – Projection PCA : les signaux frauduleux tendent à se regrouper dans des régions isolables.

4.2 Détection d'anomalies

J'ai entraîné un Isolation Forest pour détecter les comportements rares et un Autoencoder pour capturer la structure temporelle normale. L'évaluation a été menée à l'aide des courbes ROC et PR. Le modèle principal a atteint une AUC de 0.63, score cohérent avec un dataset volontairement difficile et des attaques parfois subtiles.

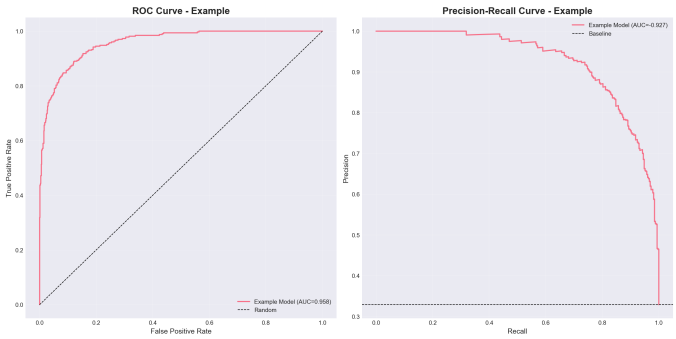

FIGURE 7 – Courbes ROC et PR du modèle de détection : performance modérée mais révélatrice de séparabilité.

5 Discussion

Les analyses ont montré que la fraude modifiait de manière mesurable les séries temporelles en affectant simultanément leur distribution, leur dynamique et leurs corrélations internes. Les modèles non supervisés ont permis d'isoler une partie significative des anomalies, même lorsque certaines attaques tentaient d'imiter des comportements humains.
Le simulateur a joué un rôle central en offrant un environnement contrôlé permettant d'isoler l'influence précise de chaque manipulation. L'étude a ainsi permis d'établir une base fiable pour l'utilisation de techniques plus avancées dans des contextes réels.

6 Conclusion

En conclusion, ce projet m'a permis de développer un simulateur réaliste, de réaliser une analyse approfondie et de concevoir un pipeline de détection reposant sur des modèles non supervisés. J'ai montré que la fraude laissait des signatures exploitables, même si certaines attaques subtiles restent difficiles à distinguer des comportements humains. Une limite importante est que les méthodes étudiées reposent uniquement sur des séries temporelles isolées : dans des systèmes réels, la détection de fraude s'appuie aussi sur des signaux sociaux, des métadonnées de devices et des patterns multi-utilisateurs. Ces aspects constituent des pistes naturelles pour prolonger ce travail.