Etude comparative des méthodes de détection d'anomalies

Maurras Ulbricht Togbe*, Yousra Chabchoub*
Aliou Boly**, Raja Chiky*

*ISEP, 10 rue de Vanves 92130 ISSY LES MOULINEAUX, France
prenom.nom@isep.fr,
http ://www.lisite.isep.fr
**Université Cheikh Anta Diop de Dakar, BP 5005 Dakar-Fann, Sénégal
prenom.nom@ucad.edu.sn
https ://www.ucad.sn

Résumé. La détection d'anomalies est un problème en plein essor et qui revêt une importance dans plusieurs domaines. A titre d'exemple, la cybercriminalité peut provoquer des pertes économiques considérables et menacer la survie des entreprises. Sécuriser son système d'information est devenu une priorité et un enjeu stratégique pour tous les types d'entreprises. D'autres domaines sont également impactés tels que la santé, les transports, etc. Les solutions de supervision mises en place sont souvent basées sur des algorithmes de détection d'anomalies issus du datamining et du machine learning. Nous présentons dans ce papier un état de l'art complet sur les algorithmes de détection d'anomalies. Nous proposons une classification de ces méthodes en se basant à la fois sur le type de jeux de données (flux, séries temporelles, graphes, etc.), le domaine d'application et l'approche considérée (statistique, classification, clustering, etc.). Nous nous focalisons ensuite sur trois algorithmes : LOF, OC-SVM et Isolation Forest que nous testons sur deux jeux de données différents afin de comparer leurs performances.

1 Introduction

La détection d'anomalies est un volet du datamining qui intéresse de plus en plus de chercheurs actuellement. On trouve dans la littérature plusieurs définitions de l'anomalie souvent appelée outlier. Hawkins (1980) définit un outlier comme une observation qui dévie considérablement du reste des autres observations comme si elle était générée par un processus différent. Quant à Dunning et Friedman (2014), ils affirment que la détection d'anomalie consiste à modéliser ce qui est normal dans le but de découvrir ce qui ne l'est pas. Aggarwal (2017) fait la distinction entre un outlier et une anomalie. Un outlier désigne le bruit et l'anomalie. Le degré d'aberrance permet de différencier les bruits des anomalies.