

Plan Prévisionnel

I. Thématiques :

1. Contexte études

Pendant ma mission d'alternance chez BNP Cardiff, j'ai travaillé dans le département de la Production financière au sein de l'unité commerciale Données Far, au sein de l'équipe Data Analyse et Innovation. L'objectif principal de notre équipe était d'assister les équipes métier (actuaire, contrôle de gestion et comptabilité, équipe de risque) dans la digitalisation des processus manuels.

Au sein de cette équipe, notre responsabilité était de transformer les données relatives aux produits financiers d'une compagnie d'assurance, incluant les primes et les sinistres des clients, en une base de données financières. Cependant, nous avons constaté qu'il y avait de nombreuses anomalies dans les données lors du chargement. Le contrôle manuel de ces anomalies prenait énormément de temps et était coûteux.

Une partie de ma mission consistait donc à mettre en place un système de contrôle de qualité des données, afin de détecter les anomalies et d'identifier les prochaines lignes de données considérées comme anormales. Dans ce contexte, nous avons décidé de développer des modèles de détection d'anomalies spécifiquement pour les données financières de cette base de données. L'objectif était de pouvoir anticiper les anomalies futures que les équipes métier devraient chercher, tout en rendant le processus plus rapide et automatisé.

II. Source bibliographiques Identifiées

Pour ce projet, nous avons basé notre travail sur les mémoires de recherche suivantes ;

« Détection d'anomalies via l'apprentissage non-supervisé : application à la fraude » de Assan Aziz Coulibaly (Coulibaly, 2021). Ce mémoire explore l'utilisation de méthodes de machine learning non-supervisées pour la détection d'anomalies, avec une application spécifique à la fraude.

« Amélioration de la qualité des données en assurance par apprentissage automatique » (Tardy, 2015) Ce mémoire met en avant l'utilisation de l'apprentissage supervisé pour améliorer la qualité des données en assurance.

En ce qui concerne la détection d'anomalies, nous avons consulté les recherches suivantes :

« Machine Learning techniques for anomaly Detection : An overview » (Ngadi, 2013) : Cette recherche fournit un aperçu des techniques de machine learning utilisées pour la détection d'anomalies.

« Anomaly Detection with Machine Learning » (Johnson, 2020) : Cette recherche se concentre spécifiquement sur la détection d'anomalies à l'aide de techniques de machine learning.

III. Mise en place :

Base de données

Nous disposons des bases de données internes récupérées qui contiennent des données financières sensibles. Dans le cadre du projet, ces données ont été analysées puis modifiées. Une valeur de coefficient aléatoire unique a été appliquée à chaque donnée.

Prototype

Durant cet exercice, des algorithmes de machine Learning ont été employés pour détecter les anomalies.

IV. Méthode

Dans le cadre de cette mission, nous avons mis en place les méthodes suivantes pour la détection des anomalies :

- Approche supervisée - Classification : Nous avons utilisé une approche où les points considérés comme des anomalies ont été notés manuellement pour créer un ensemble de données d'apprentissage. Ensuite, nous avons construit des modèles de classification pour prédire si de nouveaux points étaient des anomalies ou non
- Approche non supervisée - Clustering : Nous avons également exploré une approche basée sur le clustering pour isoler les points considérés comme des anomalies. Cette méthode consiste à regrouper les données en fonction de similarités et à identifier les groupes inhabituels. Cependant, cette approche n'a pas été retenue pour la phase finale du projet.

Veuillez noter que les résultats obtenus dans cette présentation ne seront pas représentatifs de la réalité, car les données ont été intentionnellement modifiées.

Références

Coulibaly, A. A. (2021). *Détection d'anomalies via l'apprentissage non-supervisé : application à la fraude*. Brest.

Johnson, J. (2020). Récupéré sur <https://www.bmc.com/blogs/machine-learning-anomaly-detection/>

Ngadi, S. O. (2013). Récupéré sur https://www.researchgate.net/profile/Salima-Benqdara/publication/325049804_Machine_Learning_Techniques_for_Anomaly_Detection_An_Overview/links/5af3569b4585157136c919d8/Machine-Learning-Techniques-for-Anomaly-Detection-An-Overview.pdf

Tardy, J. (2015). *Amélioration de la qualité des données en assurance par apprentissage automatique*.