Rapport projet scoring

DIEDHIOU Mamadou, HAOUD Anas, NGETH Laurent, PAZITHNOV Artemii

21 janvier 2025

Table des matières

Chapitre 1

Introduction

Introduction générale au sujet du rapport. Expliquez le contexte, les objectifs et la structure du document.

Chapitre 2

Les données

2.1 Description des données

Description détaillée de la première section.

Variable dépendante (Target)

En affichant la répartition de la variable "BAD" dans notre dataset (Figure 2.1), on voit qu'il y a davantage de cas où les clients ne font pas défauts (BAD = 0) que de cas où ils font défauts.

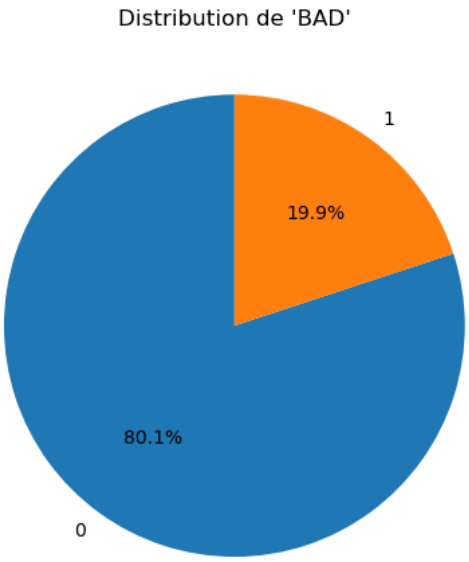

FIGURE 2.1 – Distribution de la variable Target "BAD"

Intuitivement, cela semble tout à fait logique qu'il y ait moins de cas de défaut chez les clients.
En effet,

Variables explicatives

2.2 Visualisation

Autres détails pertinents pour cette partie.

Variables numériques

Distribution

Variables explicatives

Barplot

2.3 Outliers

2.4 NaN values

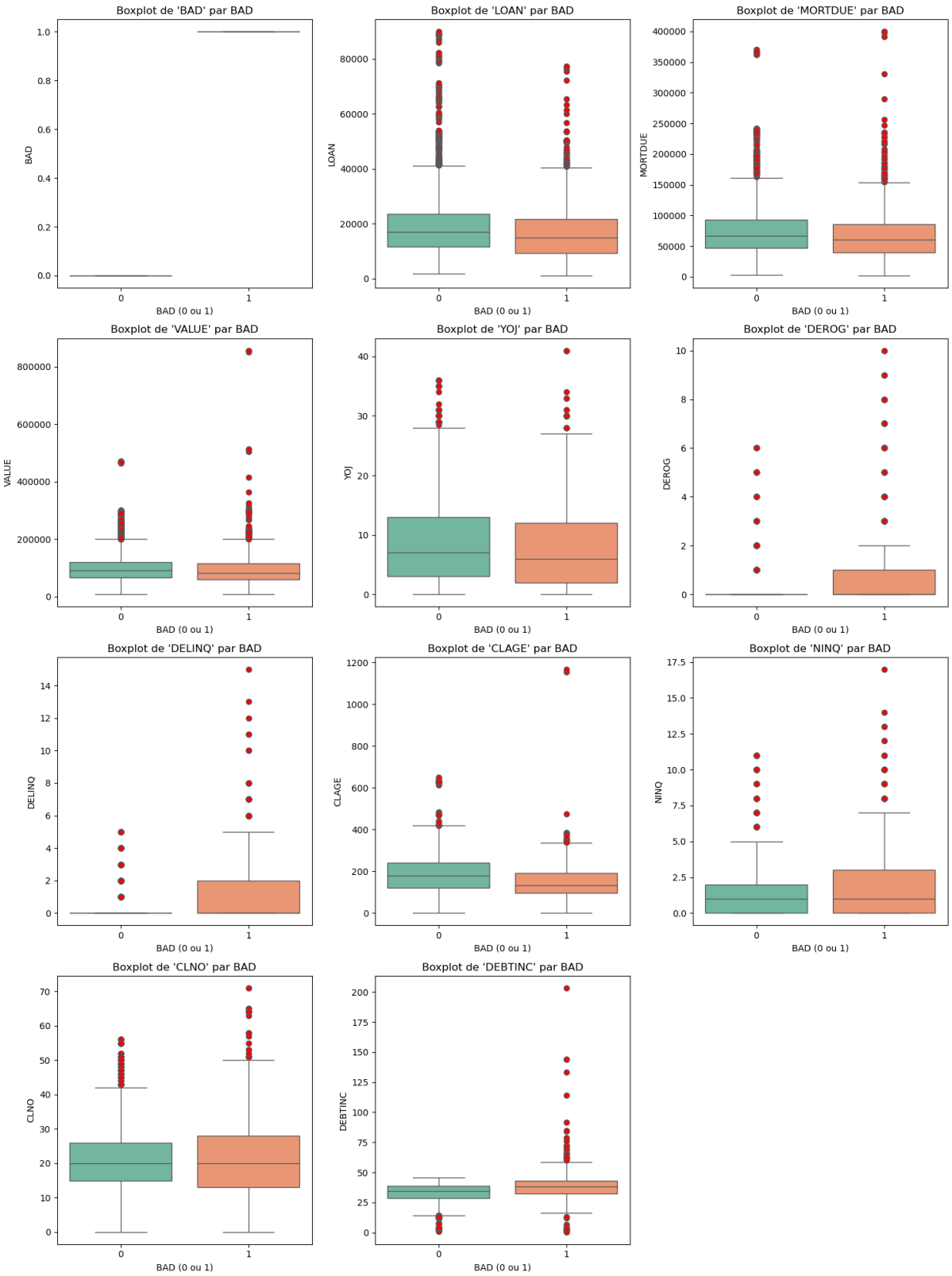

FIGURE 2.2 – Boxplot des variables numériques par rapport à BAD

Chapitre 3

Data processing

3.1 Traitement des valeurs nulles/extrèmes/manquantes

3.1.1 Variables numériques : Imputation par valeur calculée

Imputation possible :
— Par la moyenne –
— Par la médiane –
— Par KNN –

3.1.2 Variables catégorielles : Imputation de données

D'après la répartition des valeurs manquantes dans les variables catégorielles, on voit qu'elles uniquement présentes dans "REASON" et "JOB" (Figure 3.1).

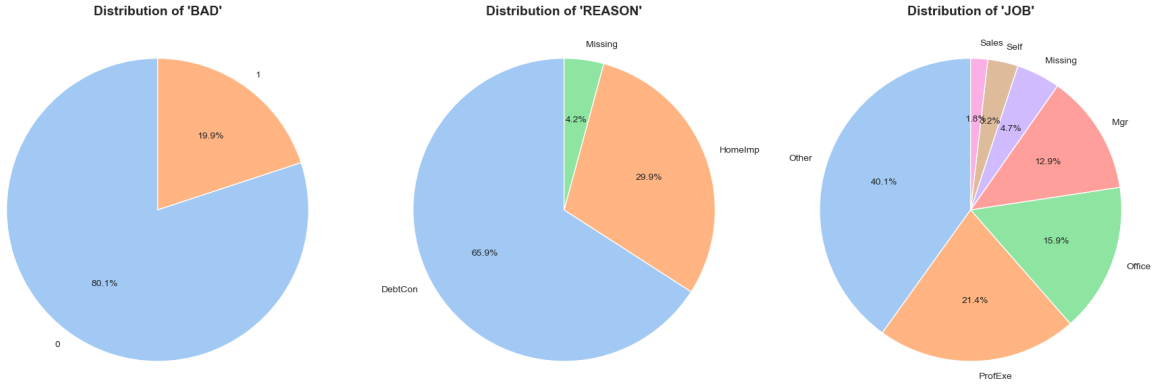

FIGURE 3.1 – Distribution des valeurs manquantes dans les variables catégorielles

On voit que ces valeurs manquantes représentent aux alentours de 5% des observations pour les variables "REASON" et "JOB".

Nous avons choisi de réaliser une suppression des observations qui possèdent ces 2 variables catégorielles à nulles (Figure 3.2).
En effet, pour seulement 5% des observations, il n'est pas nécessaires de vouloir combler les valeurs manquantes par une interpolation sachant que cela peut induire davantage de biais dans nos données.

Nous avons déjà fait une imputation par KNN pour les valeurs numériques, de ce fait, nous jugeons que cela n'est pas nécessaires vu la quantité de données, pour les variables catégorielles.

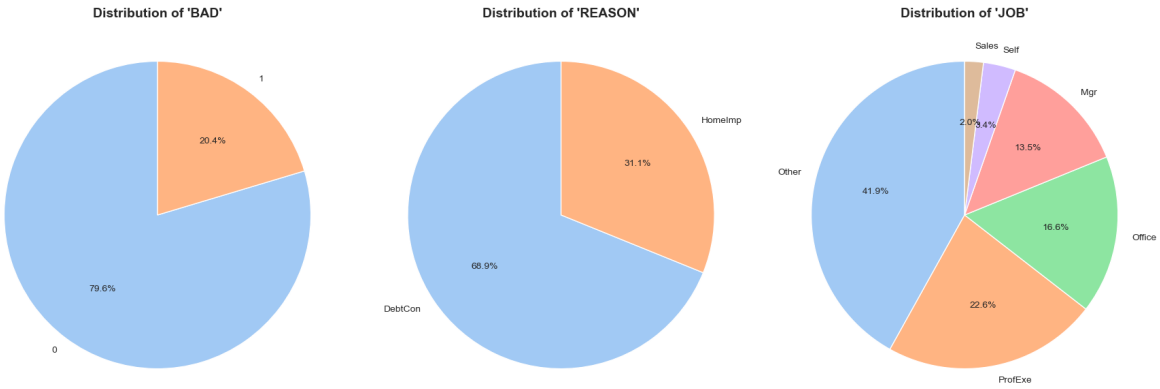

FIGURE 3.2 – Distribution des valeurs manquantes dans les variables catégorielles après suppression des observations avec NaN

Chapitre 4

Modélisation

4.1 Sélection des variables

4.1.1 Corrélation entre variable numérique-numérique

4.1.2 Corrélation entre variable catégorielle-numérique

4.1.3 Corrélation entre variable catégorielle-catégorielle

4.2 Modèles de classification

4.2.1 Métriques de performances

Pour mesurer la performances des modèles, on utilse

4.2.2 Modèles choisis

Chapitre 5

Résultats

Chapitre 6

Conclusion

Résumé des points principaux abordés dans le rapport. Incluez les conclusions finales et éventuellement des perspectives pour le futur.

Bibliographie

[1] Auteur. *Titre du livre ou de l'article*. Maison d'édition, Année.

[2] Auteur. *Titre du site web*. Consulté le : 21 janvier 2025, URL : `https://www.example.com`.

Annexe A

Annexe A

Contenu supplémentaire ou données techniques.