

AIMS-SENEGAL

**Water IA
Machine Learning
2024-2025**

6 décembre 2024

Table des matières

1	Présentation du projet	1
1.1	Objectifs	1
1.2	Quelques suggestions d'axes d'analyse	2
2	Pré-traitement	3
2.1	Nature Statistique des variables	3

Chapitre 1

Présentation du projet

Étude de données à caractère personnel

Le projet consistera à utiliser les méthodes présentées en cours dans le cadre d'une application professionnelle : exemple construction d'un score ou prédition Il s'appuiera sur une ou plusieurs des thèmes étudiés en cours :

- Régression simple multiple et régularisée
- Clustering

Chaque binôme d'auditeurs régulièrement inscrit au cours, choisira un jeu de données dans la base transmise.

PRÉSENTATION DU RAPPORT

La rédaction du rapport obéira aux règles classiques de présentation d'un projet :

- Il s'agit d'une introduction présentant le problème posé, d'une rédaction mettant en évidence la démarche choisie et les apports des analyses, conclusion. La taille du rapport doit être de 5 et 10 pages en ce qui concerne la partie rédigée. On joindra au rapport, les programmes utilisés et le rapport.

1.1 Objectifs

Comprendre d'une part et expliquer d'autre le profil des abonnés d'un organisme chargé de la distribution d'eau aux consommateurs. Suite à la baisse continue des revenues du FDE (Fonds de Développement de l'Eau), qui est une composante de tarif de l'eau, donc des factures payées par les clients, alors même que les volumes d'eau produite et distribuée ont augmenté suite aux importants investissements, le Prodeucteur juge nécessaire de déterminer les causes afin de prendre des mesures de redressement. Pour cela le Prodeucteur ce pose les questions suivantes :

- Comment évolue le FDE depuis 2014 aussi bien dans la facturation que dans le recouvrement
- Quels sont les contributeurs du FDE et comment évoluent-ils (Par type de client et par zone) ?
- Est- ce que les clients ont changé leurs habitudes ? Consomment-ils moins, sont-ils devenus de mauvais payeurs ?
- En conclusion, quels sont les facteurs qui expliquent la baisse des ressources déclarées sur le FDE.

1.2 Quelques suggestions d'axes d'analyse

1. Procéder à la segmentation des clients par ancienneté, par volume unitaire de consommation, par la qualité de paiement de ses factures
2. La granulométrie de l'analyse pourra se faire par type démission, par la catégorie de client, par zone géographique
3. Analyser l'évolution de la facturation et le recouvrement (paiement des facture) selon le type de client, selon la qualité de client, selon les zones géographiques et selon les tranches de facturation, de façon globale et selon le FDE.

L'atteinte des objectifs susmentionnés nécessitera l'exécution des tâches suivantes :

Chapitre 2

Pré-traitement

2.1 Nature Statistique des variables

Le pré-traitement concerne la totalité de la base de données soit les **23 119 112 observations** et **46 indicateurs**. Chaque indicateur est traité de façon univariée pour comprendre statistiquement sa structure.

Dans les tableaux ci-dessous 2.1, 2.3 et 2.2

- la couleur bleue indique que les variables sont prises en compte dans les calculs.
- QL = Qualitative
- QT = Quantitative
- Date = Correspondant à une date

Il s'agit d'une part de confirmer les choix quant à la prise en compte ou non des variables dans les algorithmes. Et d'autre de confirmer la nature qualitative, quantitative des variables.

Champs	Descriptions	Natures	Validations
DATE-FACT	date d'édition de la facture	Date	
DATE-CPTA-EMIS	Date Comptable émission	Date	
DATE-CPTA-ENC	Date Comptable encaissement	Date	
DATE-ABON	Date Abonné, date de l'abonnement	Date	
DATE-RESIL	Date Résiliation, de résiliation de l'abonné	Date	
DATE-REALISA	Date réalisation, date de réalisation du branchement	Date	
DATE-REGLT	Date Règlement, date de règlement de la facture	Date	

TABLE 2.1 – Variables de date

Champs	Descriptions	Natures	Validations
DR	Direction régionale correspondant à un découpage du pays	QL	
CEN	CENTRE subdivision d'une DR	QL	
POLICE	POLICE subdivision d'un CENTRE	QL	
0	Ordre de l'Abonné	QL	
C	CLE contrôle saisie	QL	
P	PERIODE de facturation mensuelle ou trimestrielle	QL	
C/C	Compte client	QL	
ENR	ENREGISTREMENT c'est le type d'enregistrement de la facture	QL	
MM	Mois de facturation	QL	
AAAA	Année de facturation	QL	
ABONNE	identifiant de l'abonné qui est le CENTRE + LA POLICE + l'ORDRE	QL	
RETARD	regroupement des factures payées à temps ou en retard. 0 pour à temps et 1 pour en retard	QL	
RESILIE	regroupement des abonnés résilié ou non. 0 pour résilié et 1 pour non résilier	QL	
ENCAISSE	regroupement par facture encaissée ou pas. 0 pour non encaissé 1 pour encaissé	QL	
CATEGORIE	Catégorie du client PRIVE ou ADMINISTRATIF	QL	
NOUVEAU	correspond à l'Age de l'abonné. 0 si moins de 5 ans 1 plus de 5 ans	QL	
DIAM	le diamètre du branchement de l'abonnée	QL	
AnABON	Année d'abonnement	QL	
NumMOIS	numérotation du mois pour une analyse statistique	QL	
NumANNEE	numérotation de l'année pour une analyse statistique	QT	

TABLE 2.2 – Variables qualitatives

Champs	Descriptions	Natures	Validations
CUBCONS	CUBAGE consommé	QT	
CUBFAC	CUBAGE consommé	QT	
FORFAIT	Cubage Forfait	QT	
SOCIAL	Cubage Social	QT	
DOMEST	Cubage Domestique	QT	
NORMAL	Cubage Normal	QT	
INDUST	Cubage Industriel	QT	
ADMINI	Cubage Administratif	QT	
MONT-SOD	Montant Concessionnaire, la part Concessionnaire dans la facture	QT	
MONT-TVA	Montant TVA, part de TVA dans la facture	QT	
MONT-FDE	Montant FDE, part de FDE dans la facture	QT	
MONT-FNE	Montant FNE, part de FNE dans la facture	QT	
MONT-ASS-TTC	Montant Assainissement, part de l'assainissement dans la facture	QT	
MONT-FRAIS-CPT	Montant frais compteur, part des frais de compteur dans la facture	QT	
MONT-TTC	Montant TTC, montant total de la facture	QT	
DELAI_REGL	délai de règlement, durée entre la date d'émission de la facture et sa date de règlement	QT	

TABLE 2.3 – Variables quantitatives