

CONNAISSANCE CLIENT : PRÉDICTION DE LA CONSOMMATION D'ÉNERGIE

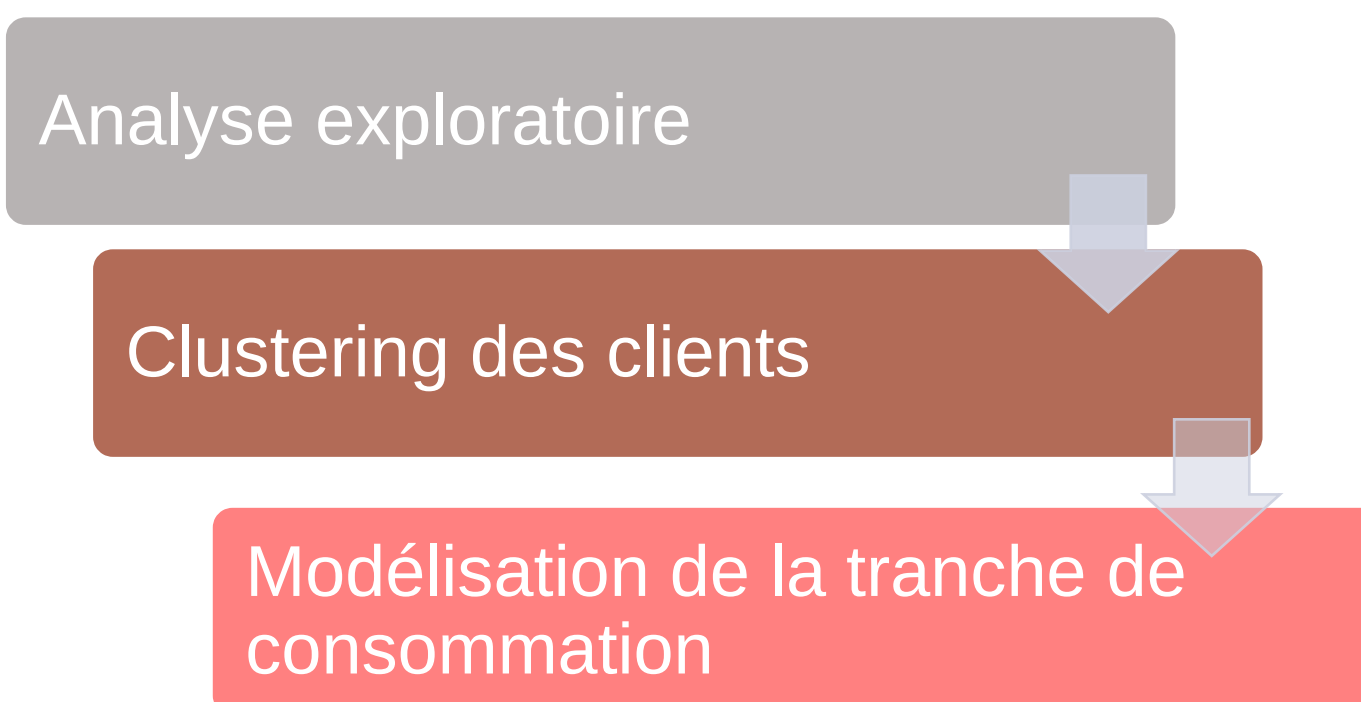
Objectif

Utiliser les algorithmes de Machine Learning sur les données des clients de la société ENGIE pour :

Regrouper les clients ayant les mêmes habitudes de consommation (clustering)

Prédire la tranche de consommation la plus probable d'un nouveau client à partir de ses caractéristiques.

Méthodologie



Résultats

Clustering des clients

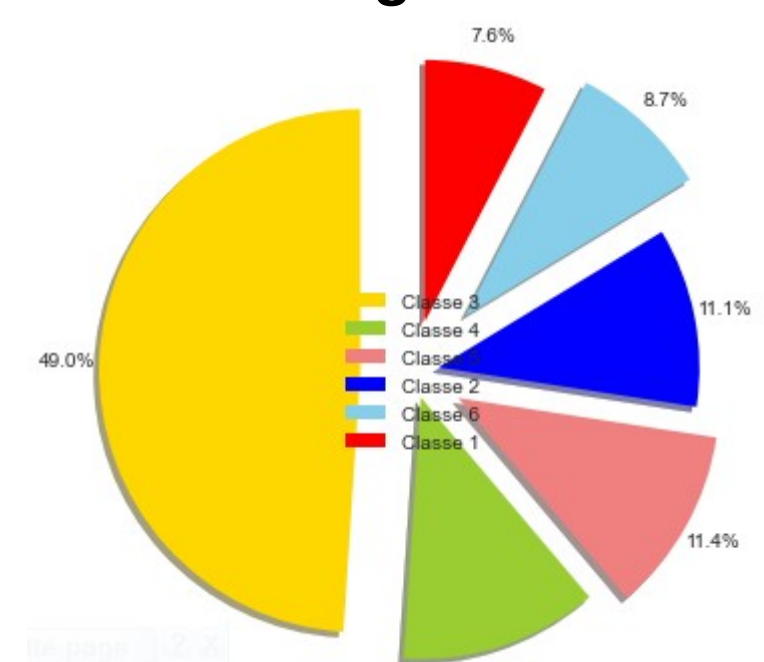


Figure 1: Répartition des clients par cluster .

	CLASSE 1	CLASSE 2	CLASSE 3	CLASSE 4	CLASSE 5	CLASSE 6
SECTEUR D'ACTIVITE	Industrie (52%) Tertiaire (47%)	Tertiaire (79%)	Tertiaire (95%)	Tertiaire (84%)	Tertiaire (93%)	Tertiaire (48%)
CONSOMMATION MOYENNE (KWh)	279	318	568	286	318	158
NOMBRE MOYEN DE FILIALES	20	80	94	98	125	81
VERSION D'UTILISATION	CU (76%)	CU (61%)	CU (57%)	LU (68%)	CU (61%)	LU (62%)
CATÉGORIE D'ENTREPRISE	PME (45%)	GE (39%)	PME (51%)	GE (46%)	GE (57%)	GE(35%) PME(37%)

Tableau 1 : Description des clusters.

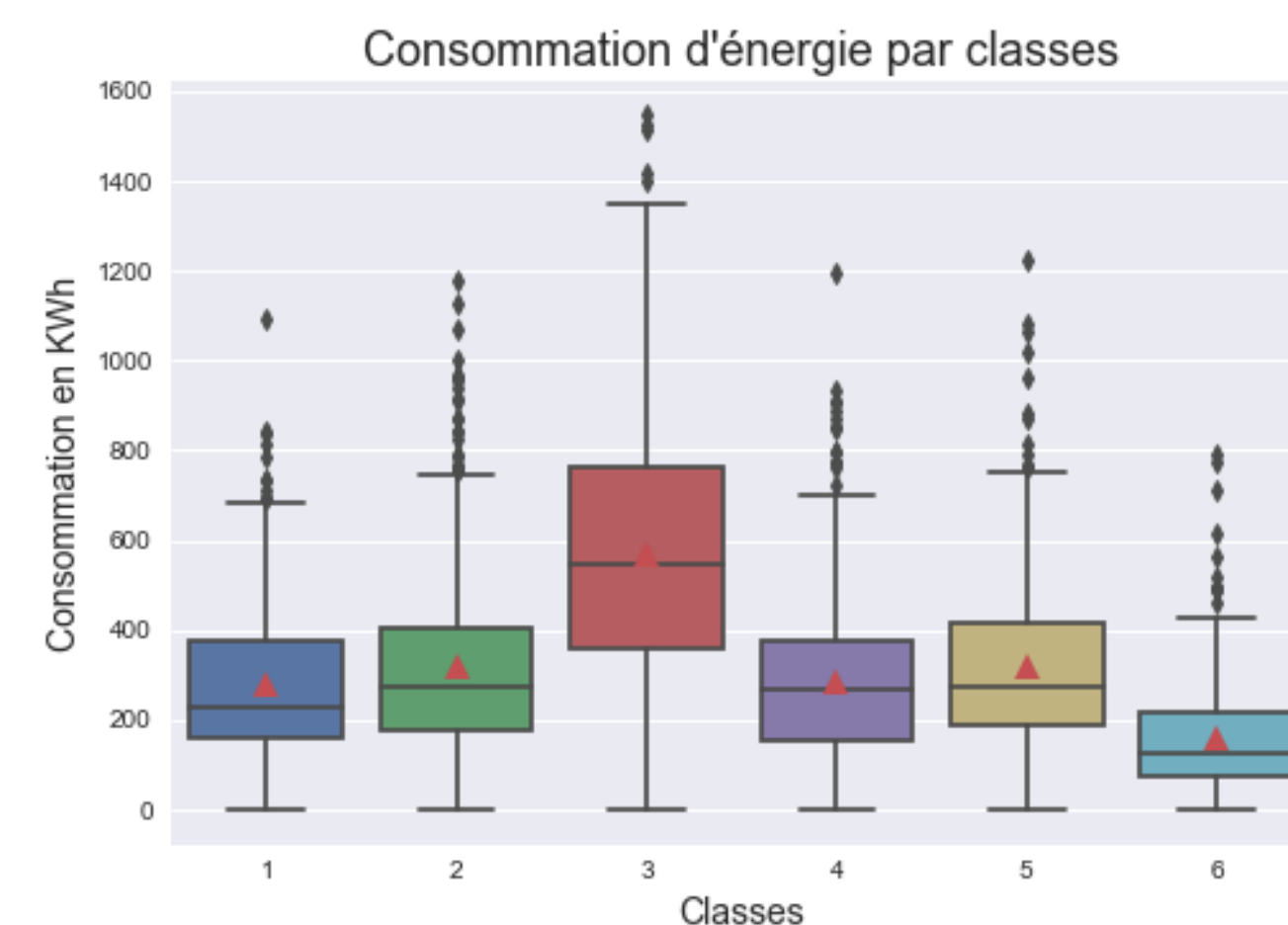


Figure 2: Distribution de la consommation d'énergie par classe

Modélisation de la tranche de consommation

Modèles	Scores moyens	Sensibilité moyenne ¹	Spécificité moyenne ²	Erreur moyenne de classification ³
Decision Tree	0.576	37,959 %	88,777 %	13,570 %
Logistic Regression	0.632	43,544 %	90,619 %	12,747 %
SVM	0.644	46,366 %	91,117 %	12,878 %
Random Forest	0.645	46,334 %	90,696 %	12,285 %
Extreme Gradient Boosting	0.651	47,896 %	91,207 %	12,154 %

Tableau 2 : Modèles candidats.

[1] Indique la probabilité qu'a un client d'appartenir à une classe sachant qu'il devrait y appartenir.

[2] Indique la probabilité qu'a un client de ne pas appartenir à une classe sachant qu'il ne devrait pas y appartenir.

Discussion

Notre problématique étant la connaissance client, notre approche de résolution a d'abord été de classer nos clients dont les antécédents sont connus sur la base de leurs habitudes de consommation. Pour ce faire, nous avons utilisé l'algorithme d'apprentissage non supervisé nommée **Kmeans**. Cette classification nous a permis d'observer un regroupement de nos clients suivant six (06) classes présentées au tableau 1 et aux figures 1 et 2.

L'étape suivante à consister à entraîner plusieurs algorithmes d'apprentissage supervisé dans le but d'obtenir celui avec les meilleurs performances. Le choix s'est fait entre les arbres de décision, la régression logistique, les SVMs, le random forest et le Extreme Gradient Boosting. Le modèle finalement retenu pour prédire la tranche de consommation d'un nouveau client est le **Extreme Gradient Boosting (XGB)** en se basant sur ses critères de performance. Des informations du tableau 2, on observe que ce modèle possède le meilleur score et un faible taux d'erreur de classification.

Dans un premier temps, les résultats de cette étude permettront pour les fournisseurs d'énergie:

- D'offrir un meilleur service client
- Optimiser le ratio achat/vente au niveau du fournisseur d'énergie

Enfin, les résultats sont extrapolables au delà du domaine de l'énergie car elles peuvent être appliquées dans d'autres domaines pour aider les fournisseurs à mieux satisfaire leurs clients.

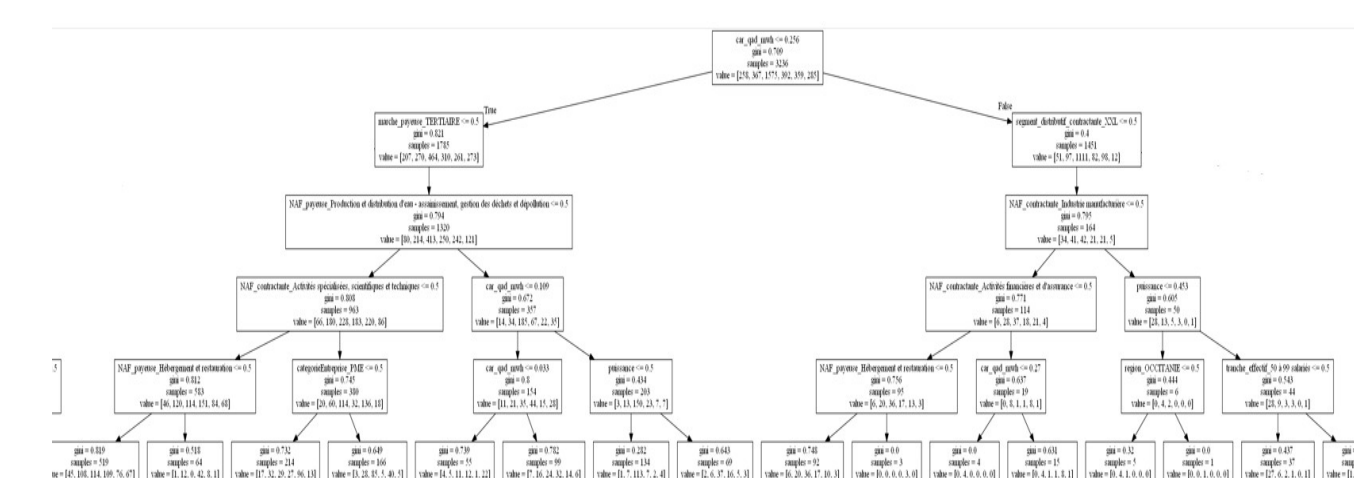


Figure 3 : Aperçu de l'arbre de décision du modèle

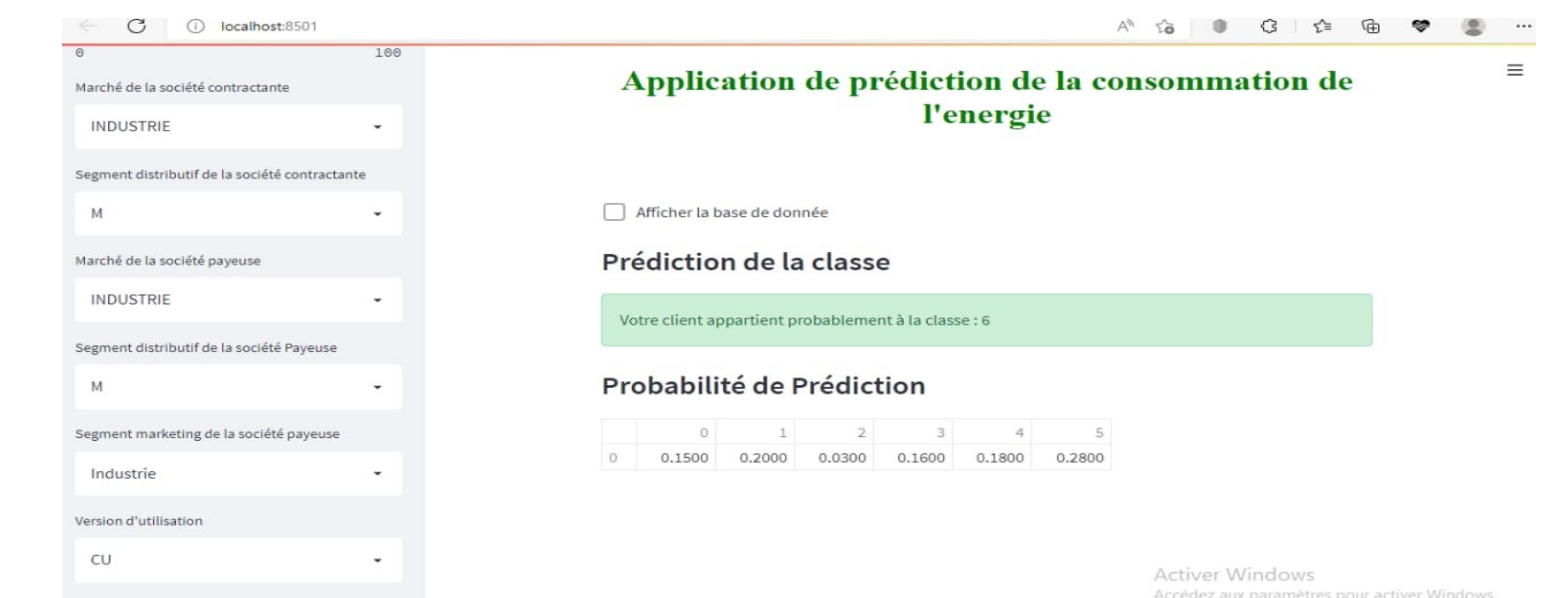


Figure 4 : Aperçu de l'application

Participants

- 1) ABATA Emilio
- 2) ADAM Wadoud
- 3) ADOHOUNBLESSI God'right
- 4) AKPONI Fulgence
- 5) ALASSANE Aïmane
- 6) ALEGUE Amidath
- 7) BONI Samir
- 8) DETCHENOU Parfait
- 9) HOUNTONDJI Océane
- 10) MAGNIDET Fifamè
- 11) MOUSS TOURÉ Hachirou
- 12) Mouandza Emery Patrice
- 13) LAWANI Aduni
- 14) RAZAFIMAHENINA Aina Anjara Fitahiana

Superviseurs

LALEYE M. Gloria
ELMARHRAOUI Achraf
Dr DAYE Modeste
Dr HOUNWANOU Sonagnon