

DOCUMENTATION DE L'OUTIL OQLASSIF

PRÉDICTION DES SECTEURS
ET FAMILLES OQALI

FEVRIER 2026

Table des matières

1.	Téléchargement de l'application.....	4
2.	Format du fichier d'entrée	4
3.	Sélection du fichier d'entrée	6
4.	Prédiction	8
5.	Format du fichier de sortie.....	11
6.	Messages d'erreurs	13
7.	Documentation.....	13

A travers une collaboration entre l'Oqali (équipes Anses et INRAE), INRAE (UMR PSAE¹ dont sa plateforme Odalim et UMR IATE²) et le LIRMM (Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier), un outil de classification automatisée des produits alimentaires transformés selon la nomenclature de l'Oqali a été développé, en se basant sur l'apprentissage supervisé : Oqllassif.

L'application Oqllassif permet de prédire les secteurs (catégories) et les familles (sous-catégories) de produits alimentaires, selon la nomenclature de l'Oqali, à partir d'informations présentes sur l'étiquetage (nom du produit, dénomination de vente, précision produit et mode de conservation).

Il est également possible de ne prédire que les familles Oqali, lorsque les secteurs sont déjà connus.

Les différentes étapes pour utiliser l'application, ainsi que les formats des fichiers d'entrée et de sortie, sont détaillés dans la suite de ce document.

Pour obtenir les meilleures prédictions, le mode de fonctionnement suivant est **recommandé** :

- 1) Télécharger l'application en local ;
- 2) Lancer l'application pour prédire les secteurs et les familles (avec un fichier d'entrée respectant les modalités présentées dans ce document) ;
- 3) Vérifier les secteurs prédits manuellement : une alerte sur les prédictions potentiellement erronées est intégrée au fichier de sortie, dans une logique de prudence (c'est-à-dire que les alertes visent à éviter les erreurs de prédictions) ;
- 4) Lancer de nouveau l'application pour ne prédire que les familles (en renseignant les secteurs vérifiés) ;
- 5) Vérifier les familles prédites manuellement : une alerte sur les prédictions potentiellement erronées est intégrée au fichier de sortie, dans une logique de prudence (c'est-à-dire que les alertes visent à éviter les erreurs de prédictions).

Attention : Le temps d'exécution pour prédire les familles et secteurs du fichier est proportionnel à la taille du fichier d'entrée : il faut prévoir environ 15 minutes pour traiter 1 000 lignes.

¹ UMR PSAE : Unité Mixte de Recherche Paris-Saclay Applied Economics

² UMR IATE : Unité Mixte de Recherche Ingénierie des Agropolymères et Technologies Emergentes

1. Téléchargement de l'application

L'application est téléchargeable, sous la forme d'un exécutable (fichier .exe) sur le site de l'Oqali, à l'emplacement suivant : <https://www.oqali.fr/donnees-publiques/outil-oqllassif/>.

Pour fonctionner, l'exécutable de l'application doit être enregistré **en local** sur son ordinateur. Il n'est pas possible de lancer cet exécutable depuis un dossier/réseau partagé par plusieurs utilisateurs.

Par ailleurs, il appartient à l'utilisateur de vérifier qu'il utilise la dernière version de l'application disponible en ligne afin de disposer de la version la plus à jour possible, la nomenclature Oqali pouvant évoluer.

2. Format du fichier d'entrée

Pour prédire des secteurs et/ou familles Oqali, l'utilisateur de l'application Oqllassif doit lui fournir un fichier d'entrée. Celui-ci doit être au **format CSV (séparateur ;)** et doit obligatoirement contenir les colonnes suivantes (attention, les libellés des noms de colonnes doivent être strictement respectés pour qu'ils soient reconnus par l'application (même casse également)) :

Nom de colonne	Type de champ	Détails	Exemple
nom_du_produit	TEXTE	Nom commercial du produit	Chips Barbecue
denomination_de_vente	TEXTE	Dénomination légale de vente du produit	Chips de pomme de terre saveur barbecue
precision_produit	TEXTE	Précision sur l'arôme ou la recette du produit	Barbecue
mode_de_conservation	TEXTE	Mode de conservation du produit. Pour une meilleure exploitation de ce champ, il est conseillé d'utiliser les termes « Ambiant », « Frais » ou « Surgele »	Ambiant

Si les noms des colonnes doivent être strictement respectés, les informations contenues dans les champs peuvent varier de ce qui est mentionné dans la colonne « détails » du tableau ci-dessus. Si l'information pour une ou plusieurs des colonnes attendues n'est pas connue, il est recommandé d'indiquer « vide » par exemple.

Dans le cas des prédictions des familles uniquement, la colonne « id_secteur » est également obligatoire (en supplément des 4 colonnes citées précédemment) :

Nom de la colonne	Type de champ	Détails	Exemple
id_secteur	TEXTE	Secteur Oqali auquel le produit appartient. Ce champ doit respecter les modalités exactes listées ci-dessous.	Produits transformes a base de pomme de terre

Les modalités exactes attendues pour le champ « **id_secteur** » sont :

- Aliments infantiles de diversification
- Aperitifs a croquer
- Barres cerealières
- Biscuits et gateaux industriels
- Boissons Rafraichissantes Sans Alcool
- Bouillons et potages
- Cereales pour le petit dejuner
- Charcuterie et assimiles
- Chocolat et produits chocolates
- Compotes
- Confiseries
- Confitures
- Conserves de fruits
- Fromages
- Glaces et sorbets
- Jus et nectars
- Laits infantiles
- Margarines
- Panification croustillante et moelleuse
- Plats cuisines ambients
- Plats cuisines frais
- Plats cuisines surgeles
- Preparations pour desserts
- Produits laitiers et desserts frais
- Produits traiteurs frais
- Produits transformes a base de pomme de terre
- Sauces chaudes
- Sauces condimentaires
- Sirops et boissons concentrees a diluer
- Snacking surgele
- Viennoiseries et desserts surgeles

Attention : Il est important que les noms des secteurs correspondent exactement à ceux cités précédemment pour qu'ils soient reconnus par l'application (même casse également). La présence d'espace ou de texte différent des modalités exactes précitées (ajout d'accent par exemple ou absence de majuscule) entraînera un blocage de l'application (voir la partie Messages d'erreurs).

Les définitions de ces secteurs sont détaillées sur le site internet de l'Oqali (<https://www.oqali.fr/donnees-publiques/faq/>).

Le fichier d'entrée peut également contenir des colonnes supplémentaires, propres à chaque utilisateur ; celles-ci seront retrouvées dans le fichier de sortie, sans modification. Par exemple, l'utilisateur pourra ajouter un identifiant de produit unique chiffré sous la colonne « id_produit » ou encore les marques de ces produits sous la colonne « marque ». Ces informations ne seront, néanmoins, pas prises en compte pour améliorer la prédiction.

3. Sélection du fichier d'entrée

La page d'accueil de l'application est la suivante :

Données d'entrée (CSV)

Parcourir

Champs obligatoires :
- nom_du_produit
- denomination_de_vente
- precision_produit
- mode_de_conservation
- id_secteur (si familles uniquement)

Prédir les familles uniquement

Lancer la prédiction

Info

Nombre de produits :

Spécifications

31 secteurs 636 familles
Entrainement sur 48 320 produits
Date de l'entraînement : 22 Mar 2023

Console

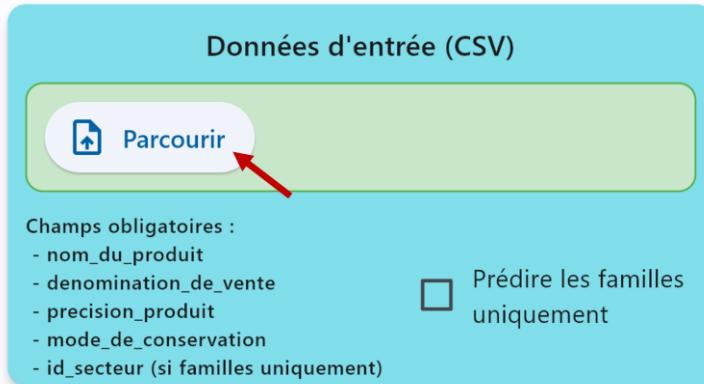
RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

anses

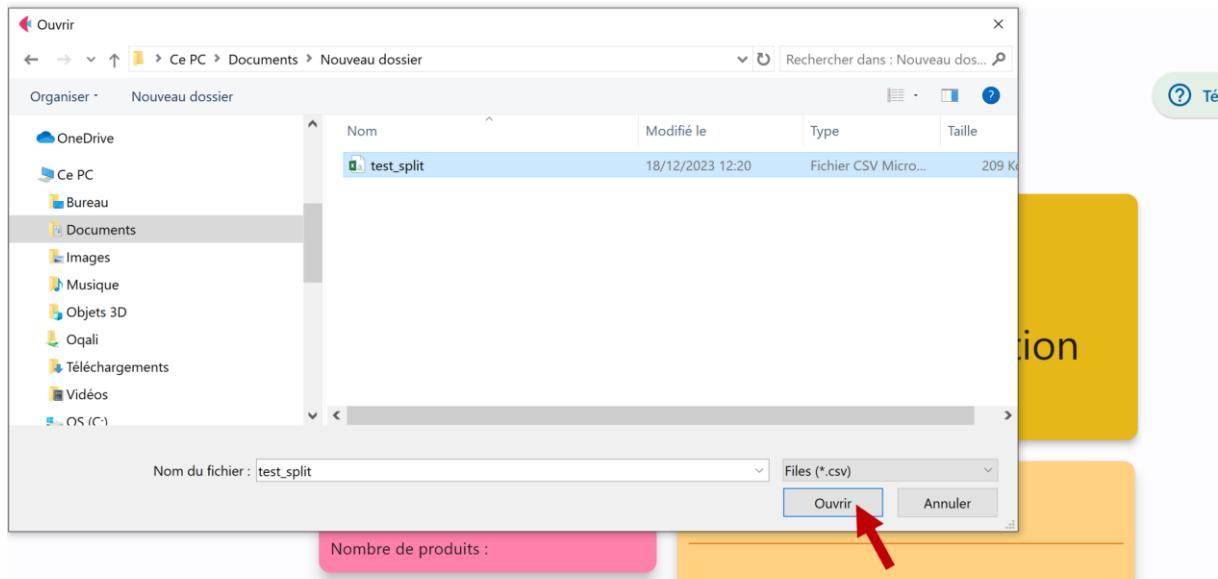
INRAE

Oqali Observatoire de l'alimentation

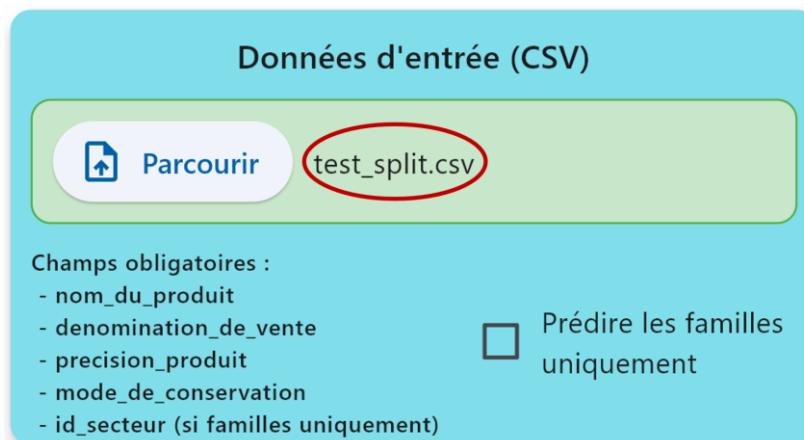
1. Cliquer sur le bouton « Parcourir » pour sélectionner un fichier d'entrée au format CSV (séparateur ;)



2. Choisir le fichier d'entrée au **format CSV (séparateur ;)** et cliquer sur « Ouvrir ».

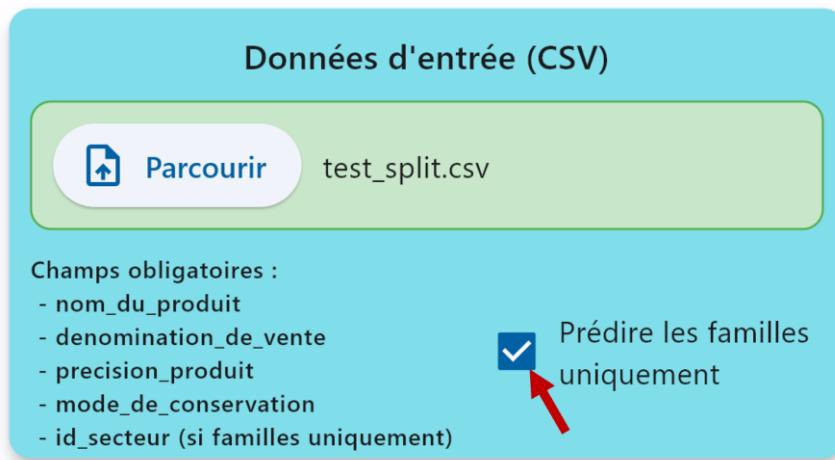


3. Une fois le fichier d'entrée sélectionné, celui-ci apparaît sur l'interface au niveau des données d'entrée.



Attention : si aucun fichier n'apparait sur l'interface, veillez à en sélectionner un à nouveau.

4. **Optionnel** : Il est possible de cocher la case « Prédire les familles uniquement », afin de ne prédire que les familles et non les secteurs (qui devront être renseignés dans le fichier d'entrée dans la colonne « id_secteur »).



4. Prédiction

1. Une fois le fichier d'entrée sélectionné, cliquer sur le bouton « Lancer la prédiction »

Données d'entrée (CSV)

Parcourir test_split.csv

Champs obligatoires :

- nom_du_produit
- denomination_de_vente
- precision_produit
- mode_de_conservation
- id_secteur (si familles uniquement)

Prédire les familles uniquement

Lancer la prédiction

Info

Nombre de produits :

Spécifications

31 secteurs 636 familles
Entraînement sur 48 320 produits
Date de l'entraînement : 22 Mar 2023

Console

RÉPUBLIQUE FRANÇAISE
Liberté Egalité Fraternité

anses

INRAE

Oqali Observatoire de l'alimentation

Documentation de l'outil Oqlassif pour la prédiction des secteurs et familles Oqali

2. Une fois la prédiction lancée :

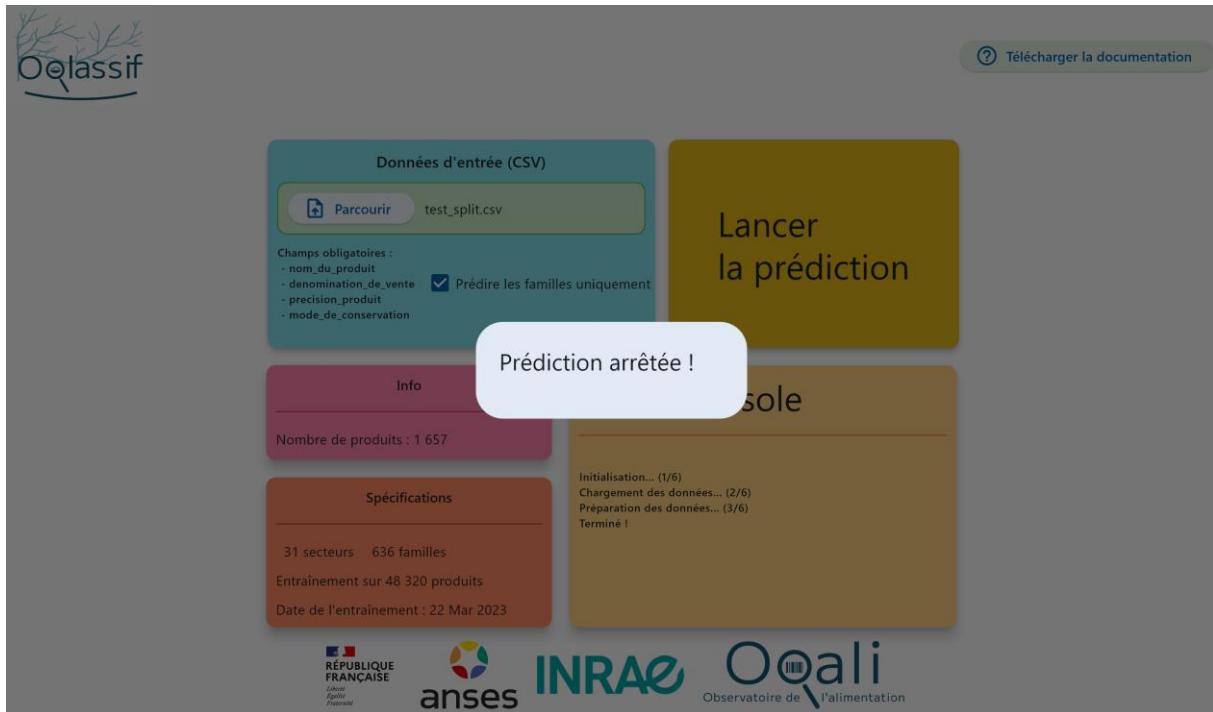
- l'encart « Console » permet de voir l'état d'avancement du traitement.
- l'encart « Info » renseigne le nombre de produits présents dans le fichier d'entrée
- l'encart « Spécifications » concerne le dernier entraînement réalisé et précise ainsi les caractéristiques de la base de données ayant servi à paramétriser l'outil de classification.

The screenshot shows the Oqlassif web application interface. At the top right is a button labeled "Télécharger la documentation". On the left, there's a logo for "Oqlassif" with a stylized tree icon. Below the logo are three main sections: "Données d'entrée (CSV)" (CSV input data), "Info" (Information), and "Spécifications" (Specifications). The "Données d'entrée" section shows a file named "test_split.csv" selected. The "Info" section displays "Nombre de produits : 1 657". The "Spécifications" section shows "31 secteurs 636 familles" and "Entraînement sur 48 320 produits". The date "Date de l'entraînement : 22 Mar 2023" is also listed. To the right, a large red button says "Arrêter" (Stop). Below these sections is a yellow box titled "Console" with a circular progress bar. Inside the "Console" box, the status is shown as "Initialisation... (1/6)", "Chargement des données... (2/6)", "Préparation des données... (3/6)", "Prédiction des secteurs... (4/6)", and "Prédiction des familles... (5/6)". At the bottom of the interface are logos for "RÉPUBLIQUE FRANÇAISE", "anses", "INRAE", and "Oqali Observatoire de l'alimentation".

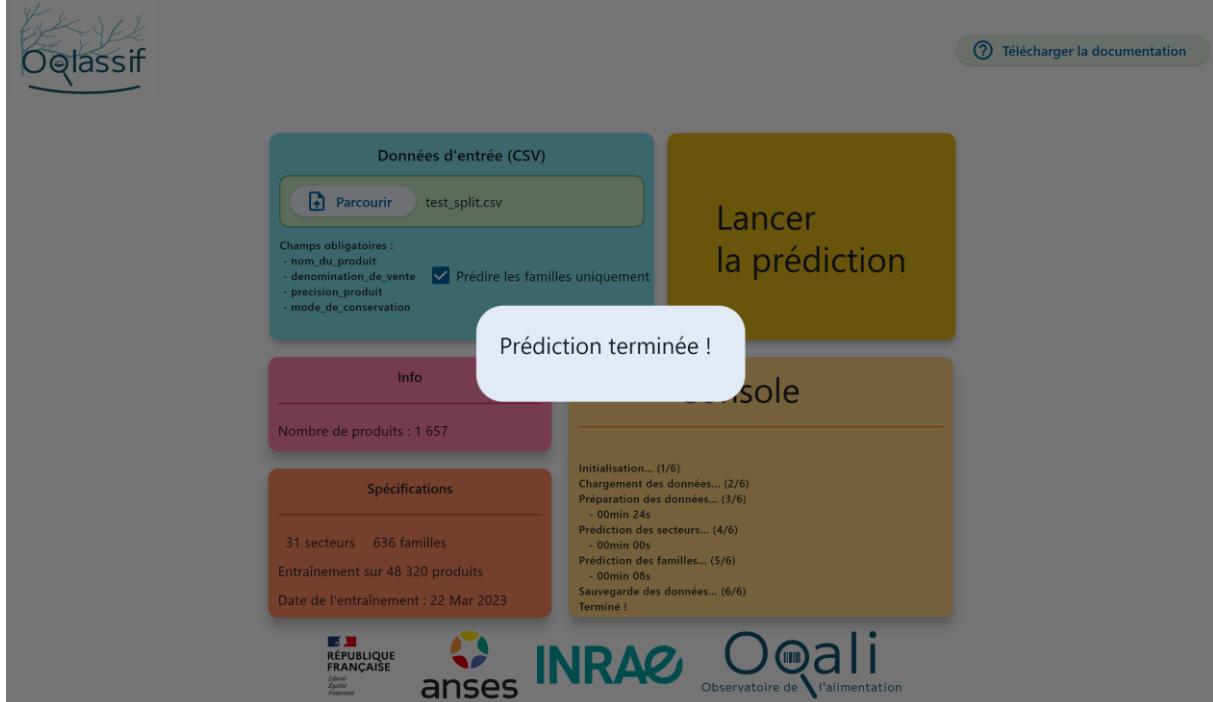
Attention : en fonction de la volumétrie des données à traiter par l'outil, le temps de prédiction peut être long. Il est néanmoins possible d'arrêter la prédiction en cours, en cliquant sur le bouton « Arrêter » : un pop-up signalera alors que la prédiction est arrêtée et aucun fichier de sortie ne sera généré.

This screenshot is similar to the one above, showing the Oqlassif interface during a prediction process. The "Console" section shows the same five-step progress: "Initialisation... (1/6)", "Chargement des données... (2/6)", "Préparation des données... (3/6)", "Prédiction des secteurs... (4/6)", and "Prédiction des familles... (5/6)". A red arrow points to the large red "Arrêter" (Stop) button on the right side of the interface. The other sections ("Données d'entrée", "Info", "Spécifications") and logos at the bottom are identical to the first screenshot.

Documentation de l'outil Oqlassif pour la prédiction des secteurs et familles Oqali



5. Une fois la prédiction terminée, le message « Prédiction terminée » s'affiche. Le fichier de sortie s'ouvre et s'enregistre automatiquement dans le **même dossier que le fichier d'entrée** sous le nom « *fichier d'entrée_predictions.csv* ».



5. Format du fichier de sortie

Après avoir réalisé une prédition, un fichier de sortie est généré. Celui-ci est au **format CSV**.

Ce fichier est différent selon le cas « Prédictions des secteurs et familles » et le cas « Prédictions des familles uniquement ».

Les colonnes obtenues pour le cas « Prédictions des secteurs et familles » sont détaillées dans le tableau ci-dessous :

Type de champ		Prédictions des secteurs et familles	Exemple
id_produit	NOMBRE ENTIER	Identifiant unique pour chaque produit. Ce champ est incrémenté automatiquement si cette colonne n'est pas renseignée dans le fichier d'entrée (en tant que colonne supplémentaire).	4
nom_du_produit	TEXTE	Nom commercial du produit.*	Chips Barbecue
denomination_de_vente	TEXTE	Dénomination légale de vente du produit.*	Chips de pomme de terre saveur
precision_produit	TEXTE	Précision sur l'arôme ou la recette du produit.*	Barbecue
mode_de_conservation	TEXTE	Mode de conservation du produit. Pour une meilleure exploitation de ce champ, il est conseillé d'utiliser les termes « Ambiant », « Frais » ou « Surgele » (seuls ces termes sont pris en compte pour les prédictions).*	Ambiant
id_secteur	TEXTE	Prédition des secteurs Oqali par l'application. Il s'agit de la prédition la plus probable, selon l'application.	Produits transformés à base de pomme de terre
proba_secteur	POURCENTAGE	Probabilité associée aux prédictions pour les secteurs Oqali.	98.675
prediction famille 1	TEXTE	Première prédition des familles. Il s'agit de la prédition la plus probable, selon l'application.	Chips classiques et ondulées
proba 1	POURCENTAGE	Probabilité associée à la première prédition. Cette probabilité est la plus importante.	79.323
prediction famille 2	TEXTE	Deuxième prédition des familles.	Chips à l'ancienne
proba 2	POURCENTAGE	Probabilité associée à la seconde prédition.	6.038
prediction famille 3	TEXTE	Troisième prédition des familles. Il s'agit de la prédition la moins probable, selon l'application, parmi les 3 prédictions du fichier.	Chips et assimilés allégés en matières grasses
proba 3	POURCENTAGE	Probabilité associée à la troisième prédition. Cette probabilité est la plus faible des probabilités du fichier.	2.92
secteur alerte	TEXTE	Champ qui prend la valeur « A vérifier » lorsque la prédition du secteur est probable à moins de 90%, selon l'application (colonne « proba_secteur inférieure à 90). Champ vide lorsque la prédition du secteur est probable à au moins 90%, selon l'application (« proba_secteur » supérieure ou égale à 90).	
famille alerte	TEXTE	Champ qui prend la valeur « A vérifier » lorsque la première prédition de la famille est probable à moins de 82% (colonne « proba 1 » inférieure à 82). Champ vide lorsque la première prédition de la famille est probable à au moins 82% (« proba 1 » supérieure ou égale à 82).	A vérifier

* Champ présent dans le fichier d'entrée (les données n'ont subi aucune modification)

Documentation de l'outil Oqlassif pour la prédiction des secteurs et familles Oqali

Les colonnes obtenues pour le cas « Prédiction des familles uniquement » sont détaillées dans le tableau ci-dessous :

	Type de champ	Prédictions des secteurs et familles	Exemple
id_produit	NOMBRE ENTIER	Identifiant unique pour chaque produit. Ce champ est incrémenté automatiquement si cette colonne n'est pas renseignée dans le fichier d'entrée (en tant que colonne supplémentaire).	4
nom_du_produit	TEXTE	Nom commercial du produit.*	Chips Barbecue
denomination_de_vente	TEXTE	Dénomination légale de vente du produit.*	Chips de pomme de terre saveur
precision_produit	TEXTE	Précision sur l'arôme ou la recette du produit.*	Barbecue
mode_de_conservation	TEXTE	Mode de conservation du produit. Pour une meilleure exploitation de ce champ, il est conseillé d'utiliser les termes « Ambiant », « Frais » ou « Surgele » (seuls ces termes sont pris en compte pour les prédictions).*	Ambiant
id_secteur	TEXTE	Secteur Oqali, renseigné par l'utilisateur au préalable.*	Produits transformés à base de pomme de terre
prediction famille 1	TEXTE	Première prédiction des familles. Il s'agit de la prédiction la plus probable, selon l'application.	Chips classiques et ondulées
proba 1	POURCENTAGE	Probabilité associée à la première prédiction. Cette probabilité est la plus importante.	79.323
prediction famille 2	TEXTE	Deuxième prédiction des familles.	Chips à la ancienne
proba 2	POURCENTAGE	Probabilité associée à la seconde prédiction.	6.038
prediction famille 3	TEXTE	Troisième prédiction des familles. Il s'agit de la prédiction la moins probable, selon l'application, parmi les 3 prédictions du fichier.	Chips et assimilés allégés en matières grasses
proba 3	POURCENTAGE	Probabilité associée à la troisième prédiction. Cette probabilité est la plus faible des probabilités du fichier.	2.92
secteur alerte	TEXTE	Champ vide car aucune prédiction secteur n'est réalisée.	
famille alerte	TEXTE	Champ qui prend la valeur « A vérifier » lorsque la première prédiction de la famille est probable à moins de 82% (colonne « proba 1 » inférieure à 82). Champ vide lorsque la première prédiction de la famille est probable à au moins 82% (« proba 1 » supérieure ou égale à 82).	A vérifier

* Champ présent dans le fichier d'entrée (les données n'ont subi aucune modification)

Les familles prédites par l'outil Oqlassif ne comportent pas d'accents : il est donc possible qu'elles diffèrent légèrement de celles utilisées dans les rapports Oqali ou dans l'outil de visualisation des données.

6. Messages d'erreurs

Si des erreurs sont constatées par l'outil (exemple mauvais format d'entrée des données), des messages s'affichent pour expliciter les problèmes à corriger par l'utilisateur.

7. Documentation

Au clic sur « télécharger la documentation », celle-ci s'ouvre et s'enregistre (sans action de l'utilisateur) à l'endroit où l'exécutable a été enregistré.