

Grace à l'AI, comment prédire la valeur du saumon

Les sources

Données de départ : les importations 2022 et 2023 du saumon, données issues du Ministère de l'Economie (10 000 articles, actualisation tous les mois)



Pourquoi le Saumon : Boulogne sur Mer principal importateur et transformateur National...

Exemple : Extrait des 5 dernières lignes de la base de données

Mois	Année	CPF6	A129	NC8	Pays	Valeur	Masse
Novembre	2023	102011	C10B	3044100	Royaume-Uni	741.262	41.185
Décembre	2023	102011	C10B	3044100	Belgique	157.838	11.411
Décembre	2023	102011	C10B	3044100	Pays-Bas	1739.39	141.491
Décembre	2023	102011	C10B	3044100	Irlande	25.511	1.5
Décembre	2023	102011	C10B	3044100	Royaume-Uni	922.419	49.863

La cible retenue:

« Quel pays proposera les meilleurs prix au premier trimestre 2024 ? »

1ère étape : Examiner et adapter la base de données



En explorant les données, on remarque que le prix d'achat du saumon n'est pas donné explicitement. En ajoutant une variable prenant pour valeur le quotient de la variable "Valeur" par la variable "Masse", nous aurons directement le prix de chaque pays pour chaque mois. Demandons à ChatGPT de nous faire cette opération, regrouper l'historique

des prix pour chaque pays, supprimer les variables qui semblent inutiles.

Mois	Année	Pays	Valeur	Masse	Prix
Janvier	2022	Belgique	7.017	0.619	11.336
Janvier	2022	Pays-Bas	624.408	47.827	13.056
Janvier	2022	Royaume-Uni	2736.25	198.068	13.815
Janvier	2022	Irlande	488.208	34.122	14.308
Février	2022	Pays-Bas	352.957	45.985	7.675

2ème étape : Sélection du modèle de prédiction

Nous allons nous concentrer sur deux modèles : la régression linéaire, les forêts aléatoires. Pour comparer, on découpe la base de données en deux, une base d'entraînement et une base de test. Ensuite, on prédit pour chaque modèle les données de la base test grâce à la base d'entraînement. On peut alors comparer nos prédictions avec les données réelles. Le modèle qui donne les données les plus proches sera le modèle que l'on garde pour faire notre prédiction finale. Nous explicitons uniquement les résultats pour le pays Belgique.



Donc, dans un premier temps, on retire les données concernant les trois derniers mois de nos données pour les regrouper dans une base de données test. Maintenant, on peut demander à ChatGPT de prédire nos données test avec nos trois modèles.

Le modèle de forêt aléatoire est celui qui nous donne l'erreur de prédiction la plus faible, avec **seulement 15,3% d'erreur**. C'est donc ce modèle que nous allons utiliser pour prédire les prix du premier trimestre 2024.

Données réelles :

Octobre 2023 : 12.151 €/kg
Novembre 2023 : 13.496 €/kg
Décembre 2023 : 13.832 €/kg

Pour le modèle de forêts aléatoires :

Données prédites :
- Octobre 2023 : 10.90 €/kg
- Novembre 2023 : 11.01 €/kg
- Décembre 2023 : 11.47 €/kg

Pour le modèle de régression linéaire :

Données prédites :
- Octobre 2023 : 14.52 €/kg
- Novembre 2023 : 14.67 €/kg
- Décembre 2023 : 14.82 €/kg

3ème étape : Prédire la meilleure valeur d'achat du prochain trimestre

Belgique		
janvier 2024	février 2024	mars 2024
11,25	11,63	11,27

Irlande		
janvier 2024	février 2024	mars 2024
14,76	14,87	14,2

Pays-Bas		
janvier 2024	février 2024	mars 2024
12,4	12,81	13,98

Royaume-Uni		
janvier 2024	février 2024	mars 2024
16,91	17,28	16,63



Le choix de la source est important, n'est-ce pas ?

Et si on achetait en Belgique