                     Projet de l’IA

Développement d’un système de recommandation

Les systèmes de recommandation de produit sont souvent les premières applications d’intelligence artificielle que souhaitent mettre en place les e-commerçants sur leur site web. C’est un choix logique qui aura un impact visible très rapidement sur les ventes.

Sans algorithmes d’intelligence artificielle, les produits recommandés sont soit définis par le e-commerçant ou l’administrateur des ventes, soit définis par une règle statistique basique (historique des ventes).

Avec de l’intelligence artificielle, il devient possible d’automatiser ces suggestions et de les personnaliser à chaque client, en fonction des pages visitées, de son historique d’achat, du contenu de son panier, de son profil… et ainsi de proposer intelligemment les bons produits au bon client, au moment opportun.

Connaissance donc l’importance des systèmes de recommandations, ce projet consiste à construire un système de recommandation de produit.

1)- Explorer les données et dites si elles nécessitent un traitement ou non.

2)- Diviser les données en 3 datasets: dataset d'entraînement , dataset de test et dataset de validation. En se basant sur vos propres recherches, justifier la taille de chaque dataset obtenu.

3)- En utilisant un algorithme de votre choix implémenter un système de recommandation de produits aux clients avec le dataset fourni.

4)- Évaluer votre modèle à l’aide des métriques vues au cours.

5)- Interpréter les résultats obtenus.

6)- Proposer une approche d'amélioration de votre modèle.

7)- Vous devez soumettre le code et  un rapport clair de vos travaux.

**NB : Le code doit être commenté.**

**Quelques détails dataset**

**Lien du dataset :** <https://www.kaggle.com/code/pierrelouisdanieau/recommender-system-associations-rules/data>

**Description des attributs du dataset**

* **InvoiceNo** : Numéro de facture correspondant à l'achat du produit.
* **StockCode** : Identifiant du produit acheté. Chaque identifiant est différent.
* **Description** : Description du produit acheté.
* **Quantity** : Quantité du produit acheté
* **InvoiceDate** : Date de la facture, du 01/12/2010 au 09/12/2011 .
* **UnitPrice** : Prix d'un produit.
* **CustomerID** : Identifiant du client. Chaque identifiant est différent.
* **Country** : Pays où le client a passé sa commande.
* **GroupPrice** : Prix de tous les mêmes produits achetés. Quantité x Prix unitaire

Un jeu de donnée (dataset) va contenir d’énorme quantité de donnée. En effet, plus notre dataset sera grand et diversifié, plus notre modèle sera apte par la suite à prédire des résultats les plus justes possible. Le dataset doit être formaté de la façon suivante :

* **Train set :** Celui-ci va être le plus volumineux en termes de donnée. En effet, c’est sur ce jeu ci que le réseau va itérer durant la phase d’entrainement pour pouvoir s’approprier des paramètres, et les ajuster au mieux. Certaines règles préconisent qu’il soit composé de 80% des données disponibles. C’est la phase d’apprentissage.
* **Validation set :** Quant à lui, on préconise d’avoir environ 10% des données disponible. Ce jeu sera appelé une seule fois, à la fin de chaque itération d’entrainement. Il va permettre d’équilibrer le système. C’est la phase d’ajustage.
* **Test set :** Ce dernier va avoir un rôle bien différent des autres, puisqu’il ne servira pas à ajuster notre réseau. En effet, il va avoir pour rôle d’évaluer le réseau sous sa forme finale, et de voir comment il arrive à prédire comme si le réseau était intégré à notre application. C’est pour cela qu’il doit être composé exclusivement de nouveaux échantillons, encore jamais utilisé pour éviter de biaiser les résultats en lui envoyant des donnés, qu’il connaîtrait déjà et qu’il aurait déjà appris lors de la phase d’entrainement ou de validation. Celui-ci encore peut être estimé de l’ordre de 10% des données disponible