



SATISFACTION DES CLIENTS

crée par

MESBAH Ibtissam

SAIDI Basma

SENAN Fatima

SALHI Amal

encadré par

Pr.HAJA Zakaria





SOMMAIRE

Objectif du projet	1
Source des données	2
Analyse des données	3
Pre-Processing des données	4
Résolution du problème	5
Conclusion	6

OBJECTIF DU PROJET

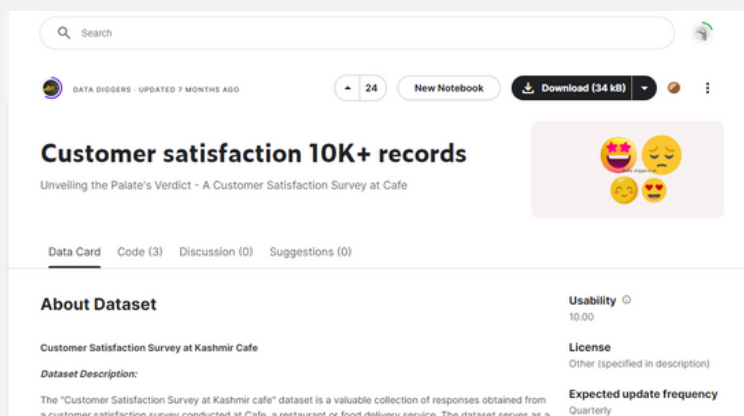


Ce projet vise à développer un programme complet pour analyser et prédire la satisfaction client dans le secteur de la vente au détail. Cela en utilisant les techniques d'apprentissage automatique et de traitement de données abordées dans le cours.

Nous avons choisi cette problématique pour comprendre les facteurs qui influencent la satisfaction globale des clients et créer des modèles prédictifs capables d'identifier les domaines nécessitant des améliorations.

En mettant en œuvre plusieurs techniques de traitement et divers algorithmes d'apprentissage automatique, nous visons à comparer les performances des modèles et à sélectionner les plus efficaces pour atteindre l'objectif de prédiction avec précision.

SOURCE DES DONNÉES



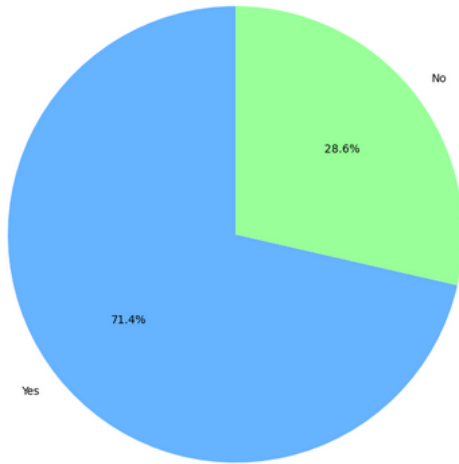
Les données utilisées dans ce projet proviennent d'une Dataset de Kaggle intitulée <<Customer satisfaction 10K+ records>> . Cette Dataset contient une collection de réponses obtenues à partir d'une enquête de satisfaction menée auprès des clients du Kashmir Cafe.

es évaluations des clients couvrent plusieurs aspects de leur expérience, notamment la satisfaction globale de la livraison (**Overall Delivery Experience**), la qualité de la nourriture (**Food Quality**), et la rapidité de la livraison (**Speed of Delivery**), toutes notées sur une échelle de 1 à 5, où 1 signifie "extrêmement insatisfait" et 5 "extrêmement satisfait". En outre, la précision des commandes est mesurée par une réponse binaire (**Order Accuracy**), indiquant si la commande du client était exacte ("oui" ou "non").

ANALYSE DES DONNÉES



Répartition des Précisions des Commandes (Yes/No)



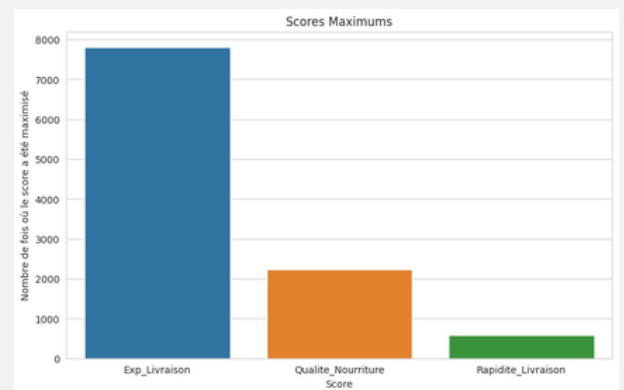
Répartition des Précisions des Commandes

Afin de mieux comprendre la distribution des réponses à la question de la précision des commandes, nous avons créé un graphique en camembert représentant les pourcentages de réponses "Yes" et "No".

Comme illustré, une majorité des clients ont indiqué que leur commande était précise ("Yes"), tandis qu'un plus petit pourcentage a signalé des inexactitudes ("No").

Scores Maximums

Ce graphique montre les scores maximums pour différents features. Chaque barre représente combien de fois un score spécifique a été maximum. Cette visualisation nous aide à voir quels features sont les plus souvent évalués au maximum, ce qui peut nous donner des indications sur où les clients sont le plus satisfaits.



PRE-PROCESSING DES DONNÉES



Pour garantir la qualité et la pertinence des données utilisées dans l'entraînement des modèles, un processus rigoureux de prétraitement des données a été mis en place.

Tout d'abord, une identification et gestion de valeurs manquantes a été effectuée. Les colonnes contenant des valeurs manquantes ont été supprimées afin de préserver l'intégrité des données.

Ensuite, Les variables catégorielles ont été converties en représentations numériques appropriées en utilisant un encodage ce qui permet de capturer efficacement les informations contenues dans les variables catégorielles et de les intégrer dans l'ensemble de données global utilisé pour l'apprentissage du modèle





RÉSOLUTION DU PROBLÉMATIQUE

plusieurs modèles de Machine Learning ont été explorés et évalués pour répondre à l'objectif d'analyser les avis des clients d'un café.

Réseaux de Neurones:

Dans ce projet, des architectures de réseaux de neurones ont été conçues et entraînées pour prédire les niveaux de satisfaction des clients en fonction des avis recueillis

Arbre de Décision:

Dans ce projet, des arbres de décision ont été utilisés pour segmenter les clients en fonction de leurs avis sur différents aspects du café.

KNN:

KNN est un algorithme simple d'apprentissage supervisé utilisé pour la classification et la régression. D'où l'utilisation de ce modèle dans la prédiction des niveaux de satisfaction des clients

Régression Logistique:

La régression logistique est un modèle d'apprentissage supervisé utilisé pour la classification binaire. Elle modélise la probabilité qu'une observation appartienne à une classe particulière. Dans ce projet, la régression logistique a été utilisée pour prédire la probabilité que les clients soient satisfaits ou non de leur expérience au café

SVM:

Le SVM est une solution à ce problème de classification. Le SVM appartient à la catégorie des classificateurs linéaires (qui utilisent une séparation linéaire des données), et qui dispose de sa méthode à lui pour trouver la frontière entre les catégories. Ce qui nous a mené à utiliser ce modèle pour prédire la probabilité que les clients soient satisfaits ou non de leur expérience au café

CONCLUSION

Dans notre étude, nous avons exploré plusieurs modèles d'apprentissage automatique, y compris KNN, la régression logistique, les arbres de décision, les réseaux de neurones et les machines à vecteurs de support (SVM), dans le but de prédire la satisfaction des clients. Cependant, malgré nos efforts pour ajuster les hyperparamètres et optimiser chaque modèle, nous avons constaté que tous les modèles ont produit des scores similaires.

```
Score: 0.7112903225806452  
Matrice de Confusion :  
[[ 0 537]  
 [ 0 1323]]
```



Cette observation soulève des questions sur la complexité du problème et la pertinence des caractéristiques utilisées pour la prédiction. Il est possible que les caractéristiques que nous avons sélectionnées ne soient pas suffisamment discriminantes pour différencier efficacement les niveaux de satisfaction des clients.

En fin de compte, notre étude souligne l'importance de l'exploration approfondie des données et du bon choix des caractéristiques dans le développement de modèles d'apprentissage automatique pour la prédiction de la satisfaction des clients. Des recherches supplémentaires pourraient se concentrer sur l'identification de caractéristiques plus informatives ou sur l'exploration de modèles plus complexes pour améliorer les performances de prédiction.