Projet de Prédiction d'Achat avec Naive Bayes

Geovany Batista Polo LAGUERRE
Onel GUSTAVE
Stive PAUL

Université des Antilles - M1 Mathématiques et Applications (MOAD)

November 10, 2024

Introduction

Contexte

Dans un monde axé sur les données, prédire les comportements d'achat à partir de simples mots-clés devient un atout majeur.

Objectif

Utiliser le modèle Naive Bayes avec lissage de Laplace pour prédire l'intention d'achat en fonction de la description d'un produit.

Méthode

- Modélisation des probabilités conditionnelles avec Naive Bayes.
- Script interactif permettant de tester le modèle en temps réel.

Concepts-Clés

1. Naive Bayes

- Modèle probabiliste simple, basé sur le théorème de Bayes.
- Hypothèse d'indépendance : chaque mot clé influence l'achat de façon indépendante.

2. Lissage de Laplace

- Évite les probabilités nulles pour les mots rares ou absents.
- Rend le modèle plus robuste face aux nouvelles données.

3. Probabilités Conditionnelles

- Probabilité d'achat en fonction des mots-clés dans la description.
- Exemple : P(achat | "pas cher" et "anglais").

Modèle Naive Bayes I

1. Théorème de Bayes

$$P(\text{achat}|\text{mots}) = \frac{P(\text{mots}|\text{achat}) \cdot P(\text{achat})}{P(\text{mots})}$$

- $P(\text{achat}|\text{mots})$: Probabilité d'achat, donnée la description.
- $P(\text{mots}|\text{achat})$: Probabilité des mots, sachant qu'il y a achat.
- $P(\text{achat})$: Probabilité a priori d'achat.

Modèle Naive Bayes II

2. Hypothèse d'Indépendance Conditionnelle

- Naive Bayes suppose que chaque mot influence indépendamment la probabilité d'achat.
- Ainsi, $P(\text{achat}|\text{mots})$ est calculé comme :

$$P(\text{achat}|\text{mots}) = P(\text{achat}) \cdot P(\text{mot}_1|\text{achat}) \cdot P(\text{mot}_2|\text{achat}) \cdots \cdots P(\text{mot}_n|\text{ach}$$

où chaque $P(\text{mot}_i|\text{achat})$ représente la probabilité d'apparition d'un mot donné en cas d'achat.

3. Application au Projet

- Modèle utilisé pour prédire si un utilisateur souhaite acheter un produit basé sur des mots-clés comme "pas cher" ou "anglais".

Lissage de Laplace I

Pourquoi le Lissage de Laplace ?

- Lorsqu'un mot de la requête n'apparaît pas dans le dataset, sa probabilité $P(\text{mot}|\text{classe})$ est nulle.
- Cela rend le produit des probabilités nul, ce qui fausse les calculs de probabilité conditionnelle.
- Le lissage de Laplace permet de gérer ce problème en ajoutant une petite valeur aux comptes.

Lissage de Laplace II

Formule du Lissage de Laplace

$$P(\text{mot}|\text{classe}) = \frac{\text{count}(\text{mot, classe}) + 1}{\text{count}(\text{classe}) + V}$$

où :

- **count(mot, classe)** : fréquence d'apparition du mot dans la classe.
- **count(classe)** : total des mots dans la classe.
- V: taille du vocabulaire (nombre total de mots uniques).

Application dans le Modèle

- Dans notre projet, le lissage de Laplace est utilisé pour éviter les probabilités nulles.
- Cela garantit que chaque mot dans une requête a une influence, même s'il est rare.

Implémentation du Modèle Naive Bayes I

1. Chargement des Données

- Les données contiennent des attributs comme `pas_cher`, `anglais`, et `achat`.
- Utilisation d'un dictionnaire Python pour représenter chaque observation.

2. Calcul des Probabilités a Priori

- Calcul de $P(\text{achat})$ et $P(\text{non\_achat})$.
- Basé sur les fréquences d'achat dans les données.

3. Calcul des Probabilités Conditionnelles

- Application du lissage de Laplace pour gérer les valeurs nulles.
- Calcul de $P(\text{pas\_cher}|\text{achat})$, $P(\text{anglais}|\text{achat})$, etc.

Implémentation du Modèle Naive Bayes II

4. Prédiction avec le Modèle

- Parse de la requête utilisateur pour extraire les mots-clés.
- Utilisation de la règle de Bayes pour calculer $P(\text{achat}|\text{requête})$.
- Retourne `achat` ou `non_achat` en fonction des probabilités calculées.

Démonstration - Exemple pratique

- Requête de l'utilisateur : "je veux un pantalon anglais pas cher"
- Conversion de la requête en mots-clés : `pas_cher, anglais`
- Calcul des probabilités avec Naive Bayes et prédiction.

Conclusion

- **Naive Bayes** : un modèle simple et efficace pour la classification, utilisé ici pour prédire l'achat d'un produit.
- **Lissage de Laplace** : améliore la précision du modèle en évitant les probabilités nulles.
- **Résultats** : le modèle fonctionne bien avec des données simples, mais peut être amélioré avec plus de données.

Perspectives

- **Améliorer le prétraitement** : explorer des techniques de traitement de texte avancées.
- **Tester sur plus de données** : valider les performances sur des ensembles plus grands et variés.