

SAÉ : Mise en œuvre d'un processus de Datamining

Analyse Prédictive pour l'Amélioration des Campagnes de Marketing Direct Bancaire

Contexte :

Vous travaillez en tant qu'analyste de données au sein d'une institution bancaire renommée. Votre équipe cherche à améliorer l'efficacité des campagnes de **marketing direct** grâce à des techniques de fouille de données (datamining) et d'analyse prédictive. Vous avez accès à un jeu de données provenant d'une campagne de marketing téléphonique, au cours de laquelle la banque cherchait à promouvoir des dépôts à terme (compte à terme : CAT) auprès de ses clients. Le CAT est un produit d'épargne bloqué, où l'argent ne peut être retiré qu'à la fin de la période convenue lors de l'ouverture du compte. Le jeu de données contient des informations sur les clients ainsi que le résultat de la campagne, indiquant si un client a souscrit ou non à un CAT.

Objectif :

Votre mission est de développer un **modèle prédictif** qui permettra à la banque de cibler plus efficacement ses clients lors des futures campagnes. Il s'agit de construire un **modèle de classification** capable de prédire, à partir de caractéristiques démographiques et financières, si un client est susceptible de souscrire à un dépôt à terme. Ce modèle aidera la banque à optimiser ses décisions marketing, en maximisant le taux de souscription tout en réduisant les coûts des campagnes.

Déroulement de la SAÉ :

La SAÉ sera divisée en deux parties principales :

1. Analyse et exploration des données :

Dans cette première partie, vous explorerez en profondeur le jeu de données. Vous effectuerez des opérations de nettoyage pour mieux comprendre les caractéristiques des clients et leur relation avec la souscription aux dépôts à terme (CAT).

Les objectifs incluent :

- Remplacement des valeurs manquantes et des caractères spéciaux (tels que "?"). En ce qui concerne les variables catégorielles, vous pouvez développer un modèle de classification pour le remplacement des valeurs manquantes.
- Gestion des valeurs aberrantes, en particulier pour la variable **âge**.
- Identification et correction des incohérences dans les données, comme les fautes de frappe.
- Exploration des caractéristiques disponibles et de la variable cible.
- Analyse exploratoire pour identifier les tendances et les corrélations.
- Création de visualisations pour mieux comprendre les caractéristiques des clients et leur lien avec le CAT.

SAÉ : Mise en œuvre d'un processus de Datamining

2. Comparaison de modèles de classification :

Dans cette seconde partie, vous vous concentrerez sur la comparaison de différents modèles de classification. L'objectif est de construire des modèles prédictifs afin d'évaluer la probabilité de souscription aux dépôts à terme (CAT) en utilisant des techniques d'apprentissage automatique. Vous évaluerez ensuite les performances de ces modèles à l'aide de mesures appropriées, et vous choisirez le modèle le plus performant sur la base de cette étude comparative.

Les étapes à suivre sont les suivantes :

1. *Sélection des modèles :*

Choisissez au moins trois modèles de classification différents, tels que l'arbre de décision, KNN, les forêts aléatoires et XGBoost.

2. *Évaluation des performances :*

Évaluez les performances initiales de chaque modèle en utilisant des mesures appropriées, telles que la précision, le rappel, et la courbe ROC.

3. *Optimisation des hyperparamètres :*

- Optimisez les hyperparamètres de chaque modèle afin d'améliorer leurs performances.
- Utilisez la validation croisée pour garantir la robustesse de vos résultats.

4. *Réduction de dimension et optimisation :*

- Appliquez des techniques de réduction de dimension, telles que l'Analyse en Composantes Principales (ACP), pour extraire les caractéristiques pertinentes du jeu de données.
- Utilisez les caractéristiques extraites par l'ACP pour optimiser les modèles, en ajustant les paramètres et en recherchant les meilleures configurations.

5. *Comparaison des performances :*

Comparez les performances des modèles en utilisant différentes configurations de l'ACP. Par exemple, tracez des courbes pour chacun des modèles en faisant varier le nombre de composantes extraites par l'ACP.

6. *Analyse des résultats :*

Analysez et comparez les performances des modèles. Expliquez les avantages et les inconvénients de chaque modèle en fonction des résultats obtenus.

7. *Choix du modèle final :*

Justifiez le choix du modèle final sur la base des performances observées, en tenant compte des critères de précision, de robustesse et de facilité d'interprétation.

SAÉ : Mise en œuvre d'un processus de Datamining

Les données :

Le jeu de données concerne les campagnes de marketing direct visant à promouvoir des dépôts à terme auprès des clients d'une institution bancaire. Pour chaque client, nous disposons de plusieurs variables, telles que l'âge, la profession, l'état civil, le niveau d'éducation, le solde du compte, ainsi que des informations sur leur historique de crédit et leur réponse à la campagne de marketing.

Ce jeu de données est déposé sur Moodle, et un document supplémentaire contient une description détaillée des attributs inclus dans le jeu de données.

Déroulement du projet :

Le projet se fera par binôme.

Délivrables attendus :

À la fin de cette SAÉ, vous devrez fournir les éléments suivants :

1. Le code source de vos modèles développés en Python.
2. Une documentation concise (pdf ou word) qui met l'accent sur les points clés, les décisions prises, et les résultats importants de votre travail. La documentation doit fournir une vue d'ensemble succincte du projet, en mettant en évidence les principaux aspects du processus d'analyse de données et de classification.

Évaluation :

Vous serez évalué(e) sur la qualité et la rigueur du travail fourni, la clarté de vos explications (écrites et orales), ainsi que sur la pertinence de votre démarche par rapport aux objectifs du projet.