

Master 2 - ISF

Machine Learning en Python

Consignes du projet

1. Consignes générales

Afin de valider l'UC « Machine Learning en Python », vous devez réaliser un projet (60% de la note finale) par trinômes.

L'objectif de ce projet est de mettre en application la démarche à suivre dans le cadre d'un projet de *Data Science* ainsi que les différents éléments vus en cours (algorithmes, techniques de sampling, cross-validation, one hot encoding, etc.) pour traiter un problème concret (décrit dans §2).

Les livrables attendus de votre part sont :

- **Le code**, dûment commenté, qui vous a permis d'obtenir vos résultats. Ce dernier peut être un Jupyter Notebook ou un script Python réalisé sous un autre logiciel (Spyder, Google Colab, etc.).
- Un **fichier** Jupyter Notebook, HTML, Rmarkdown ou autre **contenant l'exécution de votre code** et affichant vos résultats.
- **Un rapport de maximum 25 pages (au format PDF)** présentant votre démarche, résultats, raisonnement, conclusions et les limites de votre projet. A titre d'exemple, vous pourrez suivre le plan suivant :
 - Présentation/Description des données
 - Retraitements appliqués
 - Modélisation 1
 - Sampling
 - Hyperparamétrage
 - Cross-Validation
 - Résultats
 - Analyse des résultats
 - Modélisation 2
 - ...
 - Conclusion, limites et axes d'améliorations
 - Annexes (non comptabilisés dans les 25 pages).

N'hésitez pas à illustrer vos explications via des captures d'écran et graphiques issus de votre code.

L'envoi du projet doit être fait par un seul membre du groupe (avec les autres membres en CC) et les livrables contenus dans un fichier .zip nommé avec le numéro groupe (cf. GSheet répertoriant les groupes) et vos noms de famille.

Exemple : le groupe 1 composé de M.X, Mme.Y et M.Z enverra un fichier nommé Projet_1_X_Y_Z.zip

Les projets doivent être envoyés à l'adresse jorge.choa.pro19@gmail.com au plus tard le 17/12/2024 avant minuit.

2. Cas d'usage

Vous êtes un Data Scientist embauché par une institution bancaire portugaise qui souhaite optimiser l'efficacité de ses campagnes de marketing. Cette banque mène régulièrement des campagnes téléphoniques pour encourager ses clients à souscrire à un produit d'épargne. Cependant, ces campagnes nécessitent souvent plusieurs appels pour convaincre un client. Le coût et le temps associé à chaque campagne sont donc importants.

L'objectif de ce projet est de construire un modèle de Machine Learning capable de prédire si un client potentiel va souscrire ou non à l'offre, en se basant sur les données disponibles pour chaque client contacté dans le passé. En identifiant les clients les plus susceptibles de souscrire l'offre, la banque espère réduire le nombre d'appels nécessaires, diminuer les coûts de la campagne, et améliorer le taux de conversion.

Les données fournies sont issues des campagnes de marketing téléphonique passées. Elles contiennent des informations variées, incluant des caractéristiques socio-démographiques des clients, leur situation financière, ainsi que des indicateurs économiques.

Variables disponibles :

1. **Variables client** : caractéristiques personnelles des clients (âge, profession, état civil, niveau d'éducation, etc.).
2. **Variables financières** : informations concernant la situation financière des clients (par exemple, dettes en cours, crédit immobilier ou à la consommation).
3. **Historique de contact** : données sur les interactions avec les clients lors des campagnes précédentes (durée du dernier contact, nombre de jours depuis le dernier contact pour un autre produit, etc.).
4. **Indicateurs macro-économiques** : taux de variation de l'emploi, indice des prix à la consommation, indice de confiance des consommateurs, taux euribor, etc.
5. **Variable cible** : une variable binaire indiquant si le client a souscrit au produit d'épargne à la l'issue de la campagne (variable y : "oui" ou "non").

Votre travail consistera à explorer les données, prétraiter les informations, développer et entraîner un ou des modèle(s) de Machine Learning, puis évaluer ses performances. Vous devrez également fournir des recommandations sur la manière dont la banque peut mettre en œuvre ce modèle de manière efficace pour automatiser le processus de démarchage.