# –TP1–

## Énoncé

Dans ce TP, un client cherche à savoir s' il est possible de déterminer la popularité d'un profil pour un site de rencontre. Pour savoir si un compte est populaire il doit remplir 2 critère : il a plus 10 visites, il à un taux de personnes qui aime le compte après visite de plus de 5 %.  Il vous demande de réaliser une étude en utilisant de la classification pour voir si c'est possible et si oui comment le faire. Pour réaliser cette étude, vous devrez utiliser les techniques et classificateur vu en cours. N'utilisez pas de classificateur en dehors : des arbres de décisions, des random forest, des k-nn et des classificateurs bayésiens. Vous devrez aussi réaliser vous-même toutes les techniques de transformations de données nécessaires pour vous assurer le meilleur résultat possible. Vous devrez expliquer votre démarche et décrire la qualité de vos hypothèses et de vos résultats comme si vous les présentiez à un vrai client.

## Données

Vous trouverez les données dans l'archive "TP1 - Data.zip". Celle-ci se trouve dans l'onglet semaine 7 sur moodle. L'utilisation  d'aucune autre données n'est permise

## Remise

La date limite de dépôt est le dimanche 14 Juillet à 23h59. Les remises en retard entraîneront une pénalité de 3 pts par heure.
Les groupes doivent être de 3 à 4 personnes.
Une seule remise par groupe.
Les remises doivent être en format zip avec le nom suivant CODEPERM1_CODEPERM2_CODEPERM3.zip.
Elles doivent comprendre 3 fichiers : (1) le code source utilisé pour l'analyse de données, (3) un rapport de TP en format texte, (3) un rapport sous format diapositives.

## Rapport

Le rapport contiendra 2 parties :

1. Un rapport écrit dans lequel doivent apparaître
    a. Une explication de l'ingénierie de données
    b. Une liste des essaies réalisés
    c. Une analyse des résultats
2. Un rapport sous forme de diapositive qui résume le rapport écrit et doit convaincre un non-technique que votre démarche répond aux questions initialement posées.

# Barème

| CRITÈRE | POINTS |
|---|---|
| Respect des consignes | 5 |
| Clarté du rapport | 10 |
| Qualité du code | 20 |
| Nettoyage des données | 5 |
| Features engineering | 10 |
| Classificateurs | 10 |
| Qualité (performances) des résultats | 15 |
| Explication de la démarche | 25 |