**A ajouter ici dans l’ordre chronologique l’ensemble des grandes avancées/bugs rencontrés etc …**

* 08/04/2018 (Renata):
  + Fix bug teacher code (he tried to convert a feature to float, but this way didn’t work)
  + Add visualization of feature weight to help evaluate the importance
  + Implement feature to evaluate returns by month of the year
  + Implement feature to evaluate returns by day of the month
  + Implement feature to evaluate returns by day of the week
  + Optimize feature to evaluate returns by age
  + Add feature gender
  + Merge customer with x\_train (this allow using new features, and its the same code to merger with other things)

name this version: challenge\_v2

* 15/04/2018 (Renata):
  + Implement feature to evaluate returns by hour of the day
  + Merge products with x\_train (this allow using new features, and its the same code to merger with other things)
  + add feature Product Id
  + add feature BrandId
  + add feature ProductType

Score : 0.668082286797

name this version: challenge\_v3

* + Organize code (erase repeat parties)
  + Implement two new types of classifier
  + Add typeBrand feature
  + Add HeelHeight feature

Score :

RL: 0.668082286797

Bagg: 0.61412333195

Tree: 0.519453403941

RF: 0.603913447671

AdaBoost: 0.677755352911

name this version: challenge\_v4

(Raphael)

Ajouté :

- commentaires

- meilleure gestion des features

challenge\_v5

* 17/04/2018 (Renata):
  + Il y avait un erreur dans la manière qu’il évaluer le score (en pas utilisant tout le données et en utilisant le même pour training/validation)
  + Commencé le rapport, un peu de l’intro

Score :

RL: 0.689397948899

RF: 0.630789354995

AdaBoost: 0.692070703609

name this version: challenge\_v6

* 18/04/2018 (Renata):
  + Big book the features selectione

name this version: challenge\_v7

Saousan

* 18/04/2018 : Problème de sélection de features : au final on obtient un taux de corrélation =0 pour toutes les features, ce qui n’est pas parfait. Donc, on procède à un pre processing en calculant la matrice de corrélation. Elle donne des idées sur les features qui sont corrélées et donc qui doivent être dissociées et pas être utilisées ensemble pour le training. Cf : <https://machinelearningmastery.com/visualize-machine-learning-data-python-pandas/>
* 18/04/2018 : Calcul des scores par cross validation pour chaque feature. Ca peut probablement aider à améliorer le score en ne gardant que les features avec un maximum de score .
* 18/04/2018 : essai de choix de features intéressantes à la base du score obtenu , cf partie choose features. Il reste tout de même à implémenter les features choisies. (pas bien compris la partie all\_features avec les masks).

name this version: challenge\_v9

* 18/04/2018 (Renata):
  + fix bug in selecting features
  + ajouter un features pour verifier si des tailles differents sont plus retouné
  + ajouter un features pour vérifier si une femme qui achete pour un l’homme retourne plus et vice-et-versa.
  + Implementé Saousan idée de features selection
  + Continue le rapport features

name this version: challenge\_v11

RL: 0.686187025056

AdaBoost: 0.692870703609

* 18/04/2018 (Renata):
  + fix bug feature different size and different gender

named this version: challenge\_v12

* 18/04/2018 (Renata):
  + adding gradient boosting classifier

score: 0.698

* 18/04/2018 (Renata):
  + cross validation adding gradient boosting classifier

score: 0.698

named this version: challenge\_v14