Projet de Fouilles de données  
Prédiction des opérons chez E. coli K12

Heurteau Alexandre

Volle Arnaud

Chazaviel Maxime

Choix et analyse

Choix de la méthode de fouille de données, des attributs sélectionnés pour la fouille.

# Description des données

## GenBank

Il a été nécessaire de récupérer le génome d’E.coli K12 sur GenBank. Les données qui nous intéresse dans la fiche de GenBank sont le nom des gènes, leur position sur le gènome et leur fonctions.

## RegulonDB

Un fichier de RegulonDB fournit les opérons présents sur le génome. Il a donc fallut réaliser un script comparant le fichier de GenBank et celui de RegulonDB pour savoir quels gènes font partie d’un opéron.

## String db

# Réalisation

Classificateur élaboré et éventuellement paramètres.

# Résultats obtenus

Synthèse des résultats obtenus, analyse des résultats et connaissances extraites. Prédictions effectuées et précision du modèle (faux négatifs, etc.).

# Bilan et perspectives

Conclusion et perspectives. Difficultés rencontrées, suggestions sur d'autres approches, d'autres hypothèses à formuler.

# Répartition des tâches

Description de la part de l'apport de chacun des membres du groupe au projet : bien que chacun devrait participer à toutes les étapes, ce projet peut se décomposer en sous-étapes dont chacune réalisée plus particulièrement par l'un des membres du groupe.