

Forage de données

Projet: forage de données appliqué pour la résolution d'un problème concret

Plan

- ▶ Objectifs
- ▶ Options
- ▶ Directives
- ▶ Critères d'évaluation

Plan

- ▶ Objectifs
- ▶ Options
- ▶ Directives
- ▶ Critères d'évaluation

Objectifs

1. Apprentissage de nouvelles approches et de nouveaux algorithmes en forage de données en général et en apprentissage machine en particulier
2. Mise en place d'un pipeline pour résoudre un problème concret en utilisant un jeu de données déjà existant

Plan

- ▶ Objectifs
- ▶ Options
- ▶ Directives
- ▶ Critères d'évaluation

Options offertes

- ▶ **Option 01:**
 - ▶ Classification d'images
- ▶ **Option 02:**
 - ▶ Prédiction dans les séries temporelles
- ▶ **Option 03:**
 - ▶ Proposez votre propre problématique

Options offertes

- ▶ **Option 01:**
 - ▶ Classification d'images
- ▶ Option 02:
 - ▶ Prédiction dans les séries temporelles
- ▶ Option 03:
 - ▶ Proposez votre propre problématique

Option 01: classification

- ▶ **Problématique:**
 - ▶ Reconnaissance de la présence de la reine des abeilles sur des imagerie
- ▶ **Entrée:**
 - ▶ Imagettes d'abeilles (un sous-répertoire avec les imagettes sans reine et un autre avec des imagettes où la reine est présente)
 - ▶ Données disponibles sur le site du cours
- ▶ **Sortie:**
 - ▶ Deux classes:
 - ▶ Présence d'ouvrières uniquement et pas de reine
 - ▶ Présence de reine
- ▶ **Méthode:**
 - ▶ Classification
- ▶ **Exigence:**
 - ▶ Suivre le workflow décrit dans l'article : « Multi-feature fusion and PCA based approach for efficient human » (disponible sur le site du cours)
- ▶ À réaliser en équipes de 3 ou 4 personnes

Options offertes

- ▶ Option 01:
 - ▶ Classification d'images
- ▶ **Option 02:**
 - ▶ Prédiction dans les séries temporelles
- ▶ Option 03:
 - ▶ Proposez votre propre problématique

Option 02: classification/régression

- ▶ **Problématique:**
 - ▶ Prédiction dans les séries temporelles appliquée au domaine de la santé intelligente en utilisant ARIMA, LSTM, Prophet
- ▶ **Entrée:**
 - ▶ Données liées au domaine de la santé
 - ▶ Vous allez avoir à chercher des données en lien avec la problématique
- ▶ **Sortie:**
 - ▶ Prédire un état de santé (classification)
 - ▶ Prédire une valeur pour un paramètre particulier (classification/régression)
- ▶ **Méthode:**
 - ▶ Classification ou régression
- ▶ **Exigence:**
 - ▶ Explorez l'utilisation de ARIMA, LSTM et Prophet
- ▶ À réaliser individuellement ou en équipes de 2 personnes

Option 02: quelques références

► Liste non exhaustive de références:

- <https://medium.com/@cdabakoglu/time-series-forecasting-arma-lstm-prophet-with-python-e73a750a9887>
- <https://towardsdatascience.com/time-series-forecasting-with-lstms-and-prophet-predict-your-email-workload-48bf9cdb1580>
- <https://www.frontiersin.org/articles/10.3389/fdata.2020.00004/full>
- https://github.com/Mikhail-Naumov/Another_Bitcoin_Predictor

Options offertes

- ▶ Option 01:
 - ▶ Classification d'images
- ▶ Option 02:
 - ▶ Prédiction dans les séries temporelles
- ▶ **Option 03:**
 - ▶ Proposez votre propre problématique

Option 03

► Problématique:

- Proposez une problématique de votre choix en lien avec le forage de données
- La proposition doit inclure les entrées, les sorties et la méthode qui sera utilisée
- Au moins deux algorithmes/approches devraient être utilisés

► Nombre de personnes par équipe:

- À déterminer selon l'ampleur du projet

Plan

- ▶ Objectifs
- ▶ Options
- ▶ **Directives**
- ▶ Critères d'évaluation

Directives

▶ Démarche à suivre :

- ▶ Choisir une des options offertes
 - ▶ Si l'option 3 est choisie, vous devez faire valider votre proposition
- ▶ Préparer un rapport qui inclut les sections suivantes: (45%)
 - ▶ Résumé
 - ▶ Introduction et description du projet
 - ▶ Problématique (définition du problème, défis, résumé de la solution)
 - ▶ Présentation du domaine (résumés de l'article ou articles reliés, outils logiciels, bases théoriques des approches, etc.)
 - ▶ Approches implémentées:
 - Cadre général (pipeline)
 - Démarches et étapes pour l'implémentation de chaque approche
 - ▶ Expérimentations:
 - Protocole d'expérimentation mis en place (jeux de données, paramétrage, mesures d'évaluation)
 - Rapport des résultats et performances obtenus
 - Synthèse/interprétation des résultats
 - ▶ Limites et conclusion
- ▶ Préparer une présentation qui résume le rapport (25%)
- ▶ L'implémentation devrait être implémentée en utilisant Python (30)

Directives

- ▶ À remettre:
 - ▶ Rapport
 - ▶ Présentation
 - ▶ Code source
 - ▶ Guide (readme) pour pouvoir mettre en place l'environnement et tester le code (extensions: .md, .txt, docx, pdf, pptx, etc.)
- ▶ Indiquez la contribution de chaque coéquipier (en pourcentage) pour chaque actif numérique généré (rapport, présentation, code source)
- ▶ Indiquez la liste exhaustive de vos références
- ▶ Présentation orale (une date sera fixée ultérieurement) : **environ 15 à 20 minutes par présentation**
- ▶ Remise électronique : au plus tard le 18 avril à 23h59

Plan

- ▶ Objectifs
- ▶ Options
- ▶ Directives
- ▶ Critères d'évaluation

Critères d'évaluation

- ▶ Complétude
- ▶ Documentation du code
- ▶ Capacité à retrouver et sélectionner l'information pertinente
- ▶ Respect des consignes
- ▶ Contribution des coéquipiers
- ▶ Qualité du rapport et de la présentation