

Cahier des charges du projet

Module Natural Language Processing
YNOV - Filière Informatique - Spé DataIA - M1
Année scolaire 2024 / 2025
Intervenant : Nicolas Miotto

Modalités

- ❖ Le projet peut être individuel ou en binôme.
- ❖ Les étudiants doivent fournir un unique fichier txt contenant le lien github du projet (avec l'ensemble des scripts python) lors de la dernière séance du module : la soutenance des projets.
- ❖ Une présentation orale doit être préparée pour le jour de la soutenance du projet.
- ❖ Les scripts doivent être commentés et structurés (avec des titres et des parties).
- ❖ Il n'y a pas de dossier à rendre.

Cahier des charges

Le but principal du projet est de concevoir un chatbot le plus évolué possible.

1. Une interface graphique est demandée.
 - choix libre sur la librairie utilisée.
2. Le chatbot doit proposer plusieurs options d'utilisation :
 - une classification catégorielle de textes.
 - un "résumeur" de textes pour condenser automatiquement le contenu d'écrits.
 - Chaque prompt de l'utilisateur doit être soumis à :
 - un preprocessing complet.
 - à une prédiction sur un/des modèles ML optimisés.
 - à une prédiction sur un/des modèles DL optimisés.
3. L'ensemble des notions vues en classe doivent être utilisées : vectoriseur, optimiseur de paramètres, embeddings,...
4. En supplément, l'utilisateur doit pouvoir écrire un prompt et le chatbot doit répondre le plus précisément possible au prompt.
 - A minima, une recherche wikipédia doit être faite et reformulée.

Compétences évaluées sur le projet

1. Nettoyer des textes et maîtriser les expressions régulières
2. Transformer les données textuelles en données numériques
3. Implémenter un modèle de classification supervisée
4. Maîtriser l'architecture d'un réseau neuronal récurrent
5. Implémenter en python un réseau de neurones dans le but de classifier des données textuelles avec les librairies Keras et Gensim