

Mini-Projet Machine Learning

DSIR 1 – AU 2024/2025

Nous voulons créer un chatbot qui sert de guide aux utilisateurs du site de l'ISET.

1. Objectifs du projet

Le chatbot doit être capable de :

- * répondre aux questions des étudiants sur un sujet donné (ex : informations sur un cours, programme universitaire, procédures administratives, etc.), et
- * faire optimiser le temps de recherche des utilisateurs pour guider les utilisateurs à atteindre une page désirée du site suite une demande textuelle sur chat.

en utilisant des techniques de NLP et de Machine Learning (sans deep learning).

Le chatbot sera intégré au site web de l'ISET pour guider les internautes.

Pour cela, vous aurez besoin d'utiliser un dataset qui peut être composé de :

- * un dataset privé construit de questions-réponses personnalisées se rapportant aux informations présentées sur le site.
- * on pourra utiliser un dataset benchmark en plus

2. Indicateurs d'aide :

La méthodologie à adopter peut être basée (entre autre) sur :

1. *Etat de l'art des chatbots*

Etude des différentes techniques existantes et choix de la solution à adopter. Le choix doit être argumenté selon les objectifs (résultat basique ou avancé).

2. *Prétraitement des textes*

- * Tokenization (division du texte en mots) : Suppression des stopwords (mots inutiles comme "le", "et", "de")
- * Stemming/Lemmatisation (réduction des mots à leur forme de base)

3. *Représentation des textes*

- * TF-IDF (fréquence des mots pondérée) pour convertir les questions en vecteurs numériques
- * Bag-of-Words (BoW) comme alternative simple
- * Word Embeddings (Word2Vec, GloVe, FastText sans Deep Learning) pour capturer un sens plus riche des mots et mieux gérer les synonymes.

4. *Modèles de Machine Learning*

- * K-Nearest Neighbors (KNN) pour retrouver la question la plus proche dans la base de connaissances
- * SVM ou Naïve Bayes pour classifier les questions par catégories et générer une réponse prédéfinie
- * Algorithme de Similarité Cosinus pour mesurer la similarité entre la question de l'utilisateur et les questions existantes

5. *Interface utilisateur*

- * Une interface (ou une application web simple) avec Flask ou Streamlit pour interagir avec le chatbot
- * Intégrer le chatbot sur le site web de l'ISET via une API convenable

6. *Évaluation du modèle*

- * Mesurer la précision des réponses en comparant celles générées par le chatbot avec des réponses correctes

- * Tester avec de vraies questions d'étudiants pour ajuster le modèle

Ce projet pourra utiliser des bibliothèques comme NLTK, Scikit-learn, Flask et peut être amélioré progressivement.

7. *Rédaction d'un rapport*

Chaque binôme est demandé de fournir à la fin un petit rapport (10 ou 15 pages) pour expliquer le fonctionnement du chatbot réalisé.

Toute idée innovante supplémentaire sera prise en considération lors de l'évaluation.

Améliorations possibles

- * **Ajout d'un Mode Multilingue** : Supporter plusieurs langues (ex. Français et Anglais) pour une meilleure accessibilité.

- * **Gestion des Intentions et Entités** : Au lieu de classifier uniquement les questions, il serait intéressant d'identifier les intentions (ex. "demande d'information", "procédure administrative", "programme universitaire") et d'extraire des entités spécifiques (ex. nom d'un cours, département).

- * **Possibilité d'auto-apprentissage** : Ajouter une fonction permettant d'enrichir la base de données des questions-réponses au fil des interactions avec les utilisateurs.

- * **Utilisation d'un moteur de recherche interne** : Intégrer un moteur de recherche indexé (ex. Whoosh) pour retrouver les pages pertinentes du site ISET lorsque la réponse exacte n'est pas disponible.

- * **Possibilité de notation des réponses** (ex. 👍 ou 👎) pour collecter des feedbacks et améliorer le chatbot.

- * **Suggestion proactive** : Ex. après une question sur les inscriptions, proposer automatiquement un lien vers le calendrier universitaire.