

Intelligence Artificielle pour le Génie Logiciel

Classification de tickets, le cas des tickets de bugs

Quentin Perez, Christelle Urtado et Sylvain Vauttier

14 novembre 2022

Objectif de ce TP

L'objectif de ce TP est de vous montrer un exemple d'utilisation de méthodes d'intelligence artificielle dans le domaine du génie logiciel. Ce deuxième TP va vous montrer l'utilisation de méthodes de traitement du langage naturel (Natural Language Processing (NLP)) pour de la classification de tickets et plus particulièrement de tickets de bugs. L'idée principale de ce TP est de vous montrer l'utilité du traitement du langage naturel et la classification pour des problématiques GL qui peuvent aller au-delà de la classification binaire de tickets (bug triaging, détection de commit via les message, etc..)

Environnement de travail

Dans ce TP, nous allons travailler avec un environnement Python 3 et avec des *notebooks* Jupyter. Jupyter permet de créer des *notebooks* exécutables via un navigateur web ou des IDE le supportant. L'environnement Jupyter permet l'exécution, la visualisation des résultats et leur sauvegarde. Un carnet peut ainsi être exécuté et sauvegardé avec ses résultats (graphiques ou non) sans avoir à ré-exécuter l'application pour les obtenir. Nous allons également utiliser différents *packages* Python pour :

- L'ouverture de fichiers CSV et la manipulation de données avec [Pandas](#) ;
- L'apprentissage machine avec la bibliothèque [Scikit-Learn](#) ;
- La visualisation avec [Matplotlib](#) ;
- L'explicabilité du classifieur avec [Eli5](#) et plus précisément la méthode LIME (*Local Interpretable Model-agnostic Explanations*)

Environnement d'exécution

Pour exécuter le notebook Jupyter servant de support à ce TP nous allons utiliser Google Colab. Il fournit un environnement virtuel standard type VM vous permettant de ne pas dépendre d'une installation de l'ensemble de l'environnement sur votre machine locale (qui est lourd et parfois mystérieux si ce n'est mystique...). Le Colab que nous allons utiliser pour ce TP est accessible à l'aide de l'URL suivante :

<https://colab.research.google.com/drive/1eH9HSTJDrxr9wSOVnr-V9KlhTQMy5mGj?usp=sharing>

Rendu

Ce TP sera à rendre sur un dépôt Moodle qui sera ouvert prochainement.