At least 24h before each Project Review, submit a new diary entry (at least 200 words):

1. At least three educational observations related to the use of Scrum or other work methods.
2. A summary of your main contributions to the project since the previous entry.

**Diary Entry for Project Review**
**Date**: 24/04/2024

**1. Three educational observations Related to Scrum/Work Methods:**

**Improved decision making with scrum meetings:** Using Scrum meetings has shown how they can boost team decision-making. By having set times for planning, refining, and reviewing, decisions consider current progress and challenges. This structured approach leads to better decisions affecting the project's success. This is extremely helpful to make updates on progress for each components, as I and a team member discussed with Product Owners once a week

**Benefits of diverse team setup:** Having a diverse team setup in Scrum has been crucial for our project. It allows for smooth blending of skills and expertise, improving problem-solving. Team members can contribute across different roles, from scraping, frontend backend development, to machine learning development.

**Flexibility in project scope:** Scrum's flexible framework has made it simpler to adjust project scope based on feedback and new insights. Unlike traditional methods needing extensive revisions and approvals for changes, Scrum lets us change to new paths quickly and efficiently. This adaptability has been crucial in meeting user needs and market conditions. This is easy as we can simply change items in Timeline dashboard on Jira

**2. Main contributions since previous entry**

I worked extensively only during Sprint 7 after a team member has finally produced a complete dataset of fake and legit webshop URLs. This is what I have managed to achieve

- Writing extensive documentations for ML component and get merged into main branch
- Designing a highly modular and reusable machine learning pipeline that can be modified easily and support various running configurations
- I first implemented label encoding for non-text features and ensured their proper integration into our overall data processing pipeline. This also included creating JSON configurations for maintaining consistency in data preprocessing, which is vital for reproducibility and transparency in machine learning projects.
- I have developed comprehensive functions for extracting text features and turning them into TF-IDF feature matrix

- I developed Synthetic Minority Over-sampling Technique (SMOTE) into our data preprocessing pipeline to address class imbalance in our training datasets. This has significantly improved model performance by providing more balanced data, which is crucial for training robust machine learning models.
- I developed scripts for splitting datasets into training, validation, and testing sets together with SMOTE data.
- I trained Random Forest and XGBOOST classifiers and benchmark their performances
- Finally, I provided feature analysis and feature engineering for future readers to decide for themselves which feature they should choose.

All of the contributions and the code has been acknowledged by product owners and I have sent them all of my codes and documentations that is separate from the closed Github repository