**Rapport d'étude de l'existant des tickets Jira - Projet DHRD**

**1. Introduction**

L'objectif de cette étude est d'analyser l'existant des tickets Jira du projet DHRD afin de comprendre la structure des données, d'identifier les problématiques et d'orienter les futures analyses et prédictions. Nous avons extrait et organisé les données issues de Jira, puis nous avons effectué plusieurs analyses statistiques et qualitatives.

**2. Extraction et Organisation des Données**

**2.1. Extraction des données depuis Jira**

Les données ont été récupérées depuis Jira via l'endpoint suivant :  
Jira URL: **https://nx-jira8.my-nx.com/**Project Key: **DHRD**Endpoint: **/rest/api/2/search**

Ces données ont ensuite été stockées dans un fichier CSV :  
Fichier: **jira\_issues\_DHRD\_20250225\_125123.csv**

**2.2. Organisation des données**

Le fichier contient **96 106** lignes et **25** colonnes correspondant aux champs Jira suivants :

* Key, Type, Summary, Description, Status, Resolution, Resolution Date, Release Note, Priority, Created, Updated, Due Date, Reporter, Assignee, Labels, Components, Fix Versions, Parent Issue, Watchers, Issue Links, Subtasks, Comments, Module-Feature, Module, Feature, Resolution Time.

Un premier diagnostic a révélé un grand nombre de valeurs manquantes dans certaines colonnes, notamment :

* Subtasks (95.77%)
* Due Date (92.22%)
* **Feature (91.77%)**
* Parent Issue (91.25%)
* Release Note (87.97%)
* **Module (78.39%)**

**3. Analyses Statistiques Réalisées**

Nous avons mené plusieurs analyses statistiques pour comprendre la répartition des tickets :

* Nombre de tickets par statut
* Nombre de tickets par résolution
* Nombre de tickets par priorité
* Evolution des tickets par Module & Feature
* Collaborateurs avec le plus de tickets
* Modules avec le temps moyen de résolution le plus long
* Collaborateurs ayant résolu le plus de tickets de priorité élevée
* Collaborateurs ayant résolu le plus de tickets de priorité faible
* Distribution des tickets par module et feature
* Nombre de tickets par type

Une observation importante a été faite : **39 318 tickets** étaient de type "**Retrofit**". Ces tickets ont été supprimés, réduisant le dataset à **56 788 lignes**.

Ensuite, nous avons filtré uniquement les tickets ayant des valeurs non nulles pour Feature, Module et Assignee, ce qui a abouti à un sous-ensemble de **6 272 lignes.**

**4. Problèmes Identifiés**

**4.1. Données Déséquilibrées**

* Certains assignees sont surreprésentés tandis que d'autres ont très peu de tickets.

**4.2. Problème de Classification Multi-Classes**

* Nous devons prédire l'assignee parmi **138 classes**.
* Ce nombre est énorme comparé à la quantité de tickets restants (**6 272**), ce qui peut rendre les prédictions difficiles et imprécises.