

**UNIVERSITÉ TÉLUQ**

**TRAVAIL NOTÉ 1**

**PRÉSENTÉ À**

**FATIMA BENSALMA**

**COMME EXIGENCE PARTIELLE**

**DU COURS**

**PROJET EN SCIENCE DES DONNÉES SCI 1402**

**PAR**

**GABRIEL FLEURENT-THIFFAULT**

**22307023**

**TRAVAIL NOTÉ 1**

**14 AVRIL 2025**

Tâche	Description	Durée estimée	Date début	Date fin
2.	Rencontre avec les données			
2.1	Analyser et comprendre l'ensemble des colonnes, pour bien saisir leur rôle	4 jours	2025/04/14	2025/04/17
2.2	Collecte de données			
2.2.1	Création d'un dataframes via PySpark (extraction) et ajout du jeu de données heart_disease_uci et ce de manière automatique en utilisant Github.	1 jour	2025/04/18	2025/04/18
2.3.1	1. Traitement des données manquantes => Supprimer les rangées ayant des cellules vides 2. Traitement des données aberrantes => Ne conserver que les données fiables, en considérant l'hypothèse de départ, avec un boxplot, à voir (SCI 1016) * En utilisant que PySpark	10 jours	2025/04/19	2025/04/28
2.3.2	Normalisation des colonnes : Cp, Origin, Thal, Restecg et Slope avec Spark (Chikhaoui, 2025).	6 jours	2025/04/29	2025/05/04
3.	Produit final et présentation			
3.1	Analyser et apprendre			
3.1.1	Analyse partielle de l'hypothèse de départ avec les données transformées	2 jours	2025/05/05	2025/05/06
3.1.2	Étapes d'estimation en appliquant un algorithme d'apprentissage machine (régression logistique) afin d'adapter un modèle depuis les données préparées et des colonnes jugées pertinentes avec le concept de pipeline dans MLLib. (Chikhaoui, 2025)	6 jours	2025/05/07	2025/05/12
3.1.3	Évaluation des performances en utilisant la métrique précision	2 jours	2025/05/13	2025/05/15
3.2	Déployer et restituer			
3.2.1	Déploiement du modèle : Les résultats du modèle et le modèle seront partagés sur Github et également via Google Colab.	9 jours	2025/05/16	2025/05/25

3.2.2	Restitution des résultats : Un tableau de bord Power BI sera fait pour présenter les données Sélectionner des : indicateurs et graphiques pertinents => À travailler en temps et lieu Vérification de l'hypothèse à l'aide du tableau de bord	21 jours	2025/05/26	2025/06/16
-------	---	----------	------------	------------

## Bibliographie

Chikhaoui, B. (2025). Module 3 Section 3.4 [Notes fournies dans le cours SCI 1017]. <https://m2.telug.ca/mod/page/view.php?id=137216>