****

**Centre de formations professionnelles aux métiers du numérique Projet fil rouge Data Analyst**

**ANALYSE DES PERFORMANCES SCOLAIRES DES ETUDIANTS**

Sous l’encadrement de :

**M. DONTSA Cyprien**

Réalisé par l’apprenant :

**MEKA GUIFO Rolande Flora**

***Année académique : 2023 – 2024***

**Table des matières**

[**I. Contexte et objectifs 2**](#_Toc169761321)

[**II. Besoins fonctionnels 2**](#_Toc169761322)

[**1. Identification et analyse des facteurs influençant les performances 2**](#_Toc169761323)

[**2. Développement de modèles prédictifs 2**](#_Toc169761324)

[**3. Création de tableaux de bord 2**](#_Toc169761325)

[**III. Besoins non fonctionnels 3**](#_Toc169761326)

[**IV. Spécifications techniques 3**](#_Toc169761327)

[**V. Livrables 3**](#_Toc169761328)

[**VI. Planning prévisionnel 3**](#_Toc169761329)

1. **Contexte et objectifs**

L'objectif du projet est d'identifier les principaux facteurs influençant les performances académiques des étudiants au Cameroun, d'analyser en détail ces facteurs, de développer des modèles prédictifs et de créer des tableaux de bord pour faciliter le suivi et l'accompagnement des étudiants.

Les résultats de cette analyse approfondie nous permettront de mieux comprendre les déterminants de la réussite.

1. **Besoins fonctionnels**
2. **Identification et analyse des facteurs influençant les performances**

* Importer et stocker les données sur les informations personnelles, et les préférences en matière d’éducation
* Réaliser des analyses statistiques descriptives et multivariées pour identifier les principaux facteurs
* Permettre une analyse approfondie des facteurs identifiés

1. **Développement de modèles prédictifs**

* Préparer les données d'entrée pour la modélisation (gestion des données manquantes, encodage, etc.)
* Implémenter différents algorithmes de machine Learning pour prédire les performances
* Évaluer et comparer les performances des modèles
* Sélectionner et documenter le modèle le plus performant

1. **Création de tableaux de bord**

* Définir les indicateurs clés de performance à suivre
* Concevoir des visualisations adaptées (graphiques, jauges, cartes, etc.)
* Permettre le filtrage et l'exploration interactive des données
* Générer des rapports personnalisés

1. **Besoins non fonctionnels**

* Fiabilité : les analyses et prédictions doivent être fiables et reproductibles
* Sécurité : les données des étudiants doivent être protégées conformément à la réglementation

1. **Spécifications techniques**

* Langage de programmation : Python pour l'analyse des données et la modélisation
* Système de gestion de base de données : MySQL pour le stockage des données
* Bibliothèques Python : Pandas, Numpy, Scikit-learn, Matplotlib, Seaborn, etc. pour l'analyse et la modélisation
* Outil de visualisation : Power BI pour le développement des tableaux de bord

1. **Livrables**

* Rapport d'analyse détaillé identifiant et analysant les principaux facteurs influençant les performances des étudiants
* Modèle prédictif des performances des étudiants
* Tableau de bord interactif Power BI permettant le suivi des performances et l'identification des étudiants à risque

1. **Planning prévisionnel**

* Phase de collecte et de préparation des données : 5 jours
* Analyse exploratoire et identification des facteurs influents : 1 jour
* Analyse approfondie des facteurs identifiés : 2 jours
* Développement des modèles prédictifs : 3 jours
* Conception et développement du tableau de bord Power BI : 3 jours
* Tests, validation et finalisation des livrables : 2 jours