

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ ABDELHAMID IBN BADIS - MOSTAGANEM



Faculté des Sciences Exactes et d'Informatique
Département de Mathématiques et informatique
Filière : Informatique

MEMOIRE DE FIN D'ETUDES

Pour l'Obtention du Diplôme de Master en Informatique

Option : **Ingénierie des Systèmes d'Information**

Présenté par :

BENABDELOUED KAMILA

DJAHED ADEL

THÈME:

**Application du data mining éducatif dans un système de
gestion d'apprentissage des étudiants durant un cursus
universitaire.**

Soutenu le : 02 \06\2024

Devant le jury composé de :

Houcine Nadia	Université de Mostaganem	Président
Lattar Hafsa	Université de Mostaganem	Examineur
Henni Karim	Université de Mostaganem	Encadreur

Année Universitaire 2023-2024

Résumé

Aujourd'hui, de nombreuses universités mettent en place des systèmes de gestion de l'apprentissage sur leurs serveurs, permettant aux enseignants et aux étudiants d'automatiser divers aspects du processus éducatif. Dans ce contexte, notre objectif est d'explorer le domaine du data mining éducatif. Nous cherchons à analyser les résultats d'apprentissage, à résoudre les problèmes rencontrés par les étudiants, et à surveiller et évaluer les performances globales des étudiants. Notre projet vise à comprendre comment le data mining éducatif peut optimiser l'utilisation des systèmes de gestion de l'apprentissage dans les universités. En effectuant une recherche bibliographique exhaustive, nous espérons fournir un aperçu des meilleures pratiques et des technologies les plus efficaces pour améliorer les processus d'enseignement et d'apprentissage.

Mots-clés:

Activité d'apprentissage, data mining éducatif, Evaluation des activités d'apprentissage, évaluation des enseignements.

Abstract

Today, many universities are implementing learning management systems on their servers, enabling teachers and students to and students to automate various aspects of the educational educational process. In this context, our aim is to explore the field of educational data mining. Our aim is to analyze learning outcomes, solve problems solve problems encountered by students, and monitor and evaluate students' overall performance. Our project aims to understand how educational data mining can optimize the use of learning management systems management systems in universities. By conducting a comprehensive literature search, we hope to provide an overview of the best practices practices and the most effective technologies for improving teaching and learning teaching and learning processes

Keywords:

Learning activity, educational data mining, Evaluation of learning activities, teaching evaluation.

Remerciements

A l'issue de ce travail, nous remercions, en premier lieu Allah le glorieux, de nous avoir donné assez de force pour que ce mémoire puisse voir le jour.

Nous remercions spécialement notre encadreur « Mr Henni Karim » qui nous a donné l'aide et l'assistance par sa disponibilité, son conseil et son orientation durant la période de réalisation de notre projet.

Nous portons toutes nos gratitudeux aux enseignants de l'université Abdelhamid Ibn badis, pour leur dévouement et leur assistance tout au long de nos études universitaires.

Nous tenons aussi à exprimer nos gratitudeux reconnaissance à tous les membres de jury d'avoir accepté et d'évaluer notre travail.

Une pensée particulière va à toutes les personnes qui ont contribuées de près ou de loin à l'élaboration de ce travail.

BENABDELOUED KAMILA.

DJAHED ADEL

Liste des figures

Figure N°	Titre de la figure	Page
Figure 01	Processus ECD	7
Figure 02	cycle de vie de l'analyse de l'apprentissage	10
Figure 03	Le processus d'application de Data mining appliqué dans l'éducation.	22
Figure 04	Les informations sur les étudiants	29
Figure 05	pourcentage de réussite et d'échec par module	30
Figure 06	pourcentage de réussite et d'échec par test de préparation pour le module de Mathématique.	31
Figure 07	pourcentage de réussite et d'échec par test de préparation pour le module de lecture.	32
Figure 08	pourcentage de réussite et d'échec par test de préparation pour le module de l'écriture.	32
Figure 09	pourcentage de réussite et d'échec par sexe pour chaque module	33
Figure 10	Tableau de contingence	34
Figure 11	Matrice corrélation	34
Figure 12	Algorithme K-means	36
Figure 13	Utilisation de la méthode du coude pour trouver k	37
Figure 14	Utilisation de la méthode Silhouette pour trouver k	37
Figure 15	Analyse de cluster des performances des étudiants	38
Figure 16	Diagramme de profil	43

Figure 17	Diagramme de contexte statistique	43
Figure 18	Diagramme de package	44
Figure 19	Diagramme de cas d'utilisation pour l'étudiant	45
Figure 20	Diagramme de cas d'utilisation pour l'enseignant	46
Figure 21	Diagramme de cas d'utilisation pour l'admis	47
Figure 22	Diagramme de séquence s'authentifier	48
Figure 23	Diagramme de séquence créer un QCM	49
Figure 24	Diagramme de séquence gérer les devoirs soumis	50
Figure 25	Diagramme de classe de notre système	51
Figure 26	Architecture MVC	52
Figure 27	Model MVC web	54
Figure 28	Diagramme de composant	54
Figure 29	Diagramme de déploiement	55
Figure 30	Diagramme séquence de test d'authentification	56
Figure 31	Editeur IntelilJ	59
Figure 32	Logo MySQL Workbench	61
Figure 33	Interface Login	62
Figure 34	Interface d'inscription	63
Figure 35	Interface des cours	63
Figure 36	liste des Quiz et devoir	64
Figure 37	Un exemple d'un QCM pour le test	64
Figure 38	Interface des devoirs rendus	65
Figure 39	Diagramme des notes affichées	65

Figure 40	Diagramme des statistiques	66
-----------	----------------------------	----

Liste des abréviations

Abréviation	Expression Complète	Page
ECD	L'extraction de Connaissances à partir de Données	6
EDM	l'éducation Data Mining	9
REL	Ressources éducatives libres	10
DM	Data mining	20
GED	la gestion de l'éducation	24
LMS	Les systèmes de gestion d'apprentissage	26
UML	Unified Modeling Language	42
RUP	Rational Unified Process	42
HTML	HyperText Markup Language	60
CSS	Cascading Style Sheets	60
JS	JavaScript	60

Table des matières

Introduction Générale	4
Chapitre 1 Data mining éducatif	6
1.1 Introduction.....	6
1.2 Le Processus de L'ECD.....	6
1.3 Fouille de données (Data Mining).....	7
1.3.1 Définition et historique.....	7
1.4 Data Mining Educatif	8
1.4.1 Exploration de données éducatif (EDM) et analyse de l'apprentissage :.....	9
1.4.2 Qu'est-ce que l'analyse de l'apprentissage ?.....	9
1.4.3 Les approche de data mining appliquées à l'éducation :.....	11
1.4.4 Les objectifs primordiaux du data mining éducatif incluent	14
1.5 Conclusion	14
Chapitre 2 Etat de l'art	15
2.1 Introduction.....	15
2.2 Études antérieures sur les applications pratiques du Data Mining éducatif et de l'Analyse de l'Apprentissage	15
2.2.1 La modélisation des connaissances des utilisateurs	16
2.2.2 Modélisation du comportement des utilisateurs	17
2.2.3 La Modélisation de l'expérience utilisateur	17
2.2.4 Le profilage des utilisateurs.....	18
2.2.5 La Modélisation du domaine	18
2.2.6 Analyse des composantes d'apprentissage et analyse des principes d'enseignement	19
2.2.7 L'Analyse des tendances	19
2.2.8 Adaptation et personnalisation.....	20

2.3	Les processus d'application de l'EDM	20
2.4	Analyse de l'apprentissage dans les établissements d'enseignement	21
2.5	Les Techniques de l'Exploration de Données	23
2.6	Conclusion	24
Chapitre 3 Analyse et Conception du système		25
3.1	Introduction.....	25
3.2	Définition de problème.....	25
3.3	Solution proposée.....	26
3.4	Exploration Approfondie des Données	26
3.4.1	Analyse Data.....	26
3.4.2	K-means ou (K-moyennes).....	34
3.4.3	Algorithme	34
3.4.4	Exemple d'application de l'algorithme K-means	35
3.4.5	Application de K-means	36
3.5	Spécification des besoins.....	39
3.5.1	Besoins fonctionnels	39
3.5.2	Besoins non fonctionnels.....	39
3.6	Analyse des besoins	39
3.6.1	Acteurs du système	39
3.7	Modélisation de notre système.....	41
3.7.1	Modèle de cas d'utilisation.....	41
3.7.2	Modèle d'analyse	50
3.7.3	Modèle de conception	51
3.7.4	Modèle d'implémentation	53
3.7.5	Modèle de déploiement	54
3.7.6	Modèle de test.....	55
3.8	Conclusion	56
Chapitre 4 Implémentation.....		57
4.1	Introduction.....	57

4.2	Les outils des développements.....	57
4.2.1	Environnement matérielle.....	57
4.2.2	Environnement logiciel	58
4.2.2.1	Editeur IntelliJ.....	58
4.2.3	Définition des grands choix techniques.....	59
4.3	Les bibliothèques	59
4.4	Implémentation de la base de données	60
4.5	Présentation du système	61
4.5.1	Interface Login.....	61
4.5.2	Interface d'inscription	62
4.5.3	Interface des cours	62
4.5.4	Interface des contrôles.....	63
4.5.5	Interface des devoirs rendus	64
4.5.6	Interface des notes affichées	64
4.5.7	Interface des Statistique.....	65
4.6	Conclusion	65
	Conclusion Générale.....	66
	Bibliographies.....	67

Introduction Générale

Au fil des années, les progrès technologiques ont révolutionné notre société et nos habitudes quotidiennes, donnant naissance à des outils et des ressources remarquables qui rendent facilement accessibles des informations précieuses. Les évolutions technologiques ont entraîné des changements majeurs dans la manière dont l'éducation est dispensée, principalement grâce à l'essor des plateformes en ligne dédiées à l'enseignement et à l'apprentissage. Ces plates-formes permettent aux étudiants d'accéder à des ressources pédagogiques à tout moment et en tout lieu, et les enseignants peuvent plus facilement surveiller et analyser les performances des étudiants.

Pour mieux utiliser ces plateformes, la data science est très utile, elle permet d'améliorer l'ensemble du processus éducatif ainsi qu'une gestion et un suivi éducatif plus efficace pour les enseignants et les administrateurs. Ces plateformes aident également à comprendre et à résoudre les problèmes des étudiants. Dans ce contexte, une nouvelle problématique émerge : comment les données des systèmes de gestion de l'apprentissage peuvent-elles être utilisées pour mieux comprendre les performances des étudiants à l'université ? Et comment l'exploration de données éducatives peut-elle être utilisée pour identifier les signaux de réussite ou d'échec des étudiants ?

L'objectif de ce mémoire est de développer une application de data mining éducatif au sein d'un système de gestion d'apprentissage, afin d'identifier et d'analyser les facteurs de réussite et d'échec des étudiants durant leur cursus universitaire. Cette solution doit être simple à utiliser pour les enseignants et les apprenants. Et permettre une meilleure compréhension et amélioration des performances académiques.

Notre mémoire est organisé en quatre chapitres encadrés par introduction générale et conclusion générale.

Tout d'abord, nous commencerons par une recherche bibliographique sur le sujet du suivi de l'éducation d'un étudiant à l'aide de l'analyse de l'apprentissage. Cette recherche comprendra deux (2) chapitres :

Dans le premier chapitre seront abordés les principes, les techniques et les buts de l'exploration de données dans le domaine de l'éducation et de l'apprentissage.

Le deuxième chapitre de notre recherche sur les expériences, et les réalisations des chercheurs sur ces domaines.

Le troisième chapitre présentera la conception détaillée de notre application, y compris les différentes étapes de développement et les méthodes d'analyse.

Le quatrième et le dernier chapitre exposera l'implémentation et la réalisation de l'application où nous présenterons l'environnement de développement, les outils utilisés, la structure et les interfaces de notre application.

Chapitre 1

Data mining éducatif

1.1 Introduction

Avec la prolifération exponentielle des données dans les milieux éducatifs, l'analyse approfondie de ces flux de données constitue un défi majeur. Le data mining éducatif se révèle ainsi être un outil indispensable pour optimiser l'apprentissage et la réussite des étudiants. Il permet d'affiner les processus d'étude, d'accroître les taux d'achèvement des programmes, de guider les choix de cursus, d'identifier les éléments précurseurs de décrochage, de concevoir des programmes sur mesure, de prévoir les performances académiques individuelles, et même de faciliter les décisions relatives à l'inscription des étudiants.

Ce chapitre sera consacré à l'exposition des multiples techniques et composantes du data mining éducatif et de l'analyse de l'apprentissage

1.2 Le Processus de L'ECD

Pour obtenir des connaissances à partir des données, il est essentiel de suivre un ensemble d'étapes, ce qu'on appelle le « processus ECD ». Ce processus [1]. Comme illustré dans la Figure 01, comprend cinq étapes. Il n'est pas strictement linéaire, car il autorise des révisions et des améliorations en revenant en arrière vers les étapes précédentes.

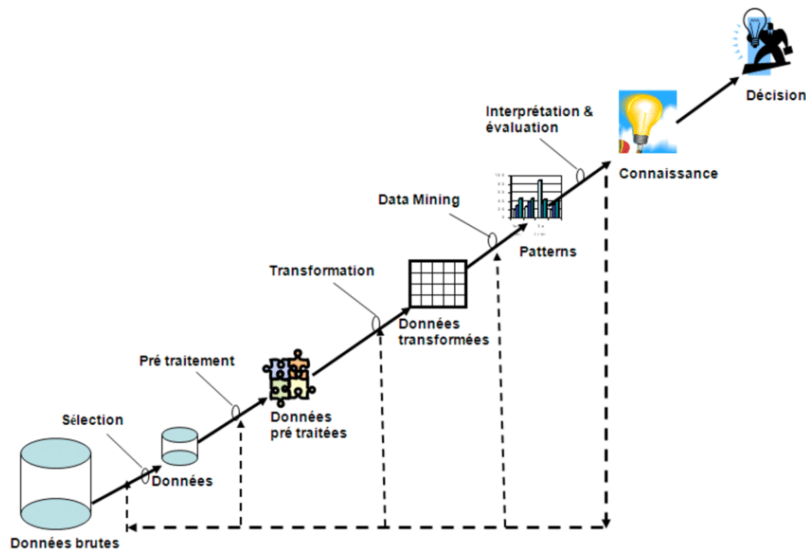


Figure 01 : Processus ECD [2]

Maintenant que les étapes précédentes sont terminées, on peut passer à l'étape centrale du processus d'ECD, à savoir la fouille de données (Data Mining).

1.3 Fouille de données (Data Mining)

1.3.1 Définition et historique

le Data Mining appelé aussi fouille de données est l'étape la plus importante du processus d'extraction de connaissances à partir de données ECD [3], est le processus de découverte de nouvelles corrélations, modèles et tendances en analysant une grande quantité de données, en utilisant les techniques de reconnaissance des formes ainsi que d'autres techniques statistiques et mathématiques. Il est apparu au milieu des années 1990 aux Etats-Unis comme une nouvelle discipline à l'interface de la statistique et des technologies de l'information : bases de données, intelligence artificielle, apprentissage automatique (Saporta, 2004) [4]

Il existe d'autres définitions, nous citons quelques-unes :

- Le data mining est l'analyse de grands ensembles de données observationnelles pour découvrir des nouvelles relations entre elles et de les

reformuler afin de les rendre plus utilisables de la part de leurs propriétaires (Hand et al., 2001).

- Le data mining est un domaine interdisciplinaire utilisant en même temps des techniques d'apprentissage automatique, de la reconnaissance des formes, des statistiques, des bases de données et de visualisation pour déterminer les manières d'extraction des informations de très grandes bases de données (Cabena et al., 1998).
- Le data mining est l'extraction de connaissances à partir de grandes quantités de données. C'est un domaine relativement récent qui se situe à l'intersection des statistiques, de l'apprentissage automatique et des bases de données (Jiawei et Kamber, 2001). À partir des définitions proposées ci-dessus et dans notre contexte, nous adoptons la définition de (Talbi, 2015)

« Le Data mining est un processus itératif et interactif dont l'objectif est la découverte de modèles de données valides, nouveaux, utiles et compréhensibles dans de larges Bases de Données» [4].

1.4 Data Mining Educatif

Le data mining éducatif consiste à explorer et analyser les données provenant des outils d'apprentissage en ligne pour trouver des schémas et des informations utiles qui peuvent aider les enseignants à mieux comprendre et à soutenir les étudiants dans leur apprentissage [5].

Autrement dit, il s'agit d'un domaine de recherche multidisciplinaire qui combine l'intelligence artificielle, la modélisation statistique et le data mining appliqués aux données générées par les établissements d'enseignement.

1.4.1 Exploration de données éducatif (EDM) et analyse de l'apprentissage :

L'exploration de données éducatives (EDM) et l'analyse de l'apprentissage utilisent des méthodes et des techniques de statistiques, d'apprentissage automatique et d'exploration de données pour examiner les données recueillies pendant l'enseignement et l'apprentissage. EDM teste les théories de l'apprentissage et oriente les pratiques éducatives. L'analyse de l'apprentissage applique des techniques issues des sciences de l'information, de la sociologie, de la psychologie, des statistiques et de l'apprentissage automatique pour analyser les données collectées lors des processus d'enseignement et d'apprentissage. Elle génère des applications qui ont un impact direct sur la pratique éducative [6].

1.4.2 Qu'est-ce que l'analyse de l'apprentissage ?

L'analyse de l'apprentissage est la mesure, la collecte, l'analyse et la communication de données sur les apprenants et leurs contextes, dans le but de comprendre et d'optimiser l'apprentissage et les environnements dans lesquels il se déroule. Cette définition générale reste valable même si le domaine s'est développé. L'analyse de l'apprentissage est à la fois un domaine universitaire et un marché commercial qui s'est rapidement développé au cours de la dernière décennie. En tant que domaine de recherche et d'enseignement, l'analyse de l'apprentissage se situe à la convergence de l'apprentissage (recherche pédagogique, sciences de l'apprentissage et de l'évaluation, technologie éducative), de l'analyse (statistiques, visualisation, sciences de l'informatique et des données, intelligence artificielle) et de la conception centrée sur l'humain (utilisabilité, conception participative, pensée systémique sociotechnique) [7].

En d'autre terme LA est un domaine en émergence utilise des outils analytiques sophistiqués pour améliorer l'apprentissage et l'éducation. Il s'appuie sur plusieurs autres domaines d'étude, notamment l'intelligence économique, l'analyse web, l'analyse académique, l'exploration des données éducatives (EDM) et l'analyse de l'action [8].

L'objectif de l'analyse d'apprentissage est de permettre aux enseignants et aux écoles d'adapter les possibilités d'éducation en fonction des besoins et des capacités de chaque élève [6].

- Un modèle de cycle d'analyse de l'apprentissage a été décrit par Clow [9] dans la figure 02

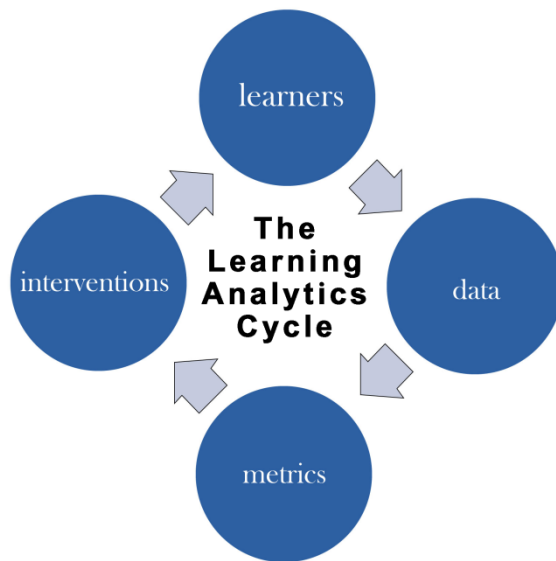


Figure 02 : cycle de vie de l'analyse de l'apprentissage

Le cycle décrit dans la figure 02 commence avec les apprenants, qu'ils soient des étudiants universitaires, des participants à des cours en ligne massifs ou des utilisateurs occasionnels de ressources éducatives libres REL. Ensuite, des **données** sur ces apprenants sont collectées, comme leurs informations démographiques ou des données de connexion, des messages de forum, etc. Ces données sont ensuite traitées pour créer des **métriques** ou des analyses. Enfin, ces mesures sont utilisées pour mettre en œuvre des **interventions** visant à influencer positivement les apprenants, telles que des interventions personnalisées par les tuteurs. Ce cycle peut être incomplet si les interventions ne sont pas appliquées aux apprenants concernés [10].

1.4.3 Les approche de data mining appliquées à l'éducation :

De nombreuses études et expérimentations ont été menées dans le domaine des données éducatives afin de mieux comprendre la qualité de l'apprentissage et ses facteurs associés.

Les méthodes d'exploration de données sont utilisées pour exploiter d'énormes ensembles de données afin de trouver des modèles et des relations cachés, ce qui permet à de nombreuses organisations de prendre des décisions fondées sur des données.

Différentes approches de fouille de données sont employées pour extraire des connaissances à partir des données éducatives, telles que des algorithmes de prédiction, des méthodes de distillation des données pour faciliter le jugement humain, des techniques d'extraction de règles d'association, de regroupement, et de modélisation de la découverte [9][11].

➤ Regroupement

Le regroupement est une méthode utilisée pour séparer les données en différents groupes sur la base de certaines caractéristique communes.

Le regroupement fait référence au processus d'identification et de classification des objets en différents groupes, à la segmentation d'un ensemble de données en sous-ensembles (clusters) de sorte que les données de chaque sous-ensemble partagent certaines caractéristiques communes à des classes d'objets similaires [9]

En général, la méthode de regroupement est utilisées lorsque le groupe le plus courant dans l'ensemble de données est inconnu. Elle est également utilisée pour réduire la taille de la zone d'étude. Par exemple, différentes écoles peuvent être regroupées en fonction de leurs similitudes et de leurs différences [12].

➤ **Classification**

La classification consiste à établir des relations entre les données afin de prédire des valeurs pour de futures observations. Elle implique d'apprendre une fonction qui associe des ensembles d'attributs à des classes prédéfinies. Plusieurs techniques de classification existent, telles que les arbres de décision, le raisonnement basé sur la mémoire, les méthodes basées sur les règles, les réseaux de croyance naïfs et bayésiens, ainsi que les réseaux neuronaux. Pour évaluer la pertinence des règles de classification, des données de test sont utilisées. Si la confiance dans ces règles est suffisante, elles peuvent être appliquées à de nouvelles données. L'algorithme de formation du classificateur utilise des exemples pré-classifiés pour déterminer les paramètres nécessaires à une classification précise.

➤ **Prédiction**

L'objectif principal des méthodes de prédiction est d'estimer les performances des étudiants, qu'il s'agisse de réussite ou d'échec.

La prédiction vise à prévoir des variables inconnues sur la base de données historiques pour la même variable. Cependant, les variables d'entrée peuvent être classées ou continuer à être des variables. L'efficacité du modèle de prédiction dépend du type de variables d'entrée. Le modèle de prédiction doit disposer d'un nombre limité de données étiquetées pour la variable de sortie. Les données étiquetées offrent une certaine connaissance préalable des variables que nous devons prédire.

Cependant, il est important de prendre en compte les effets de la qualité des données d'apprentissage afin de réaliser le modèle de prédiction [12].

Il existe trois types de prédictions générales sont couramment utilisés:

- La classification, qui implique la construction d'un modèle d'apprentissage à partir de connaissances préexistantes, puis l'utilisation de ce modèle pour classer de nouvelles données en tant que variables binaires ou catégoriques. Divers modèles de classification, tels que les machines à vecteurs de support et la régression logistique, sont utilisés, ainsi que des techniques comme les arbres de décision et le naïve Bayes.

- La régression, qui prédit des variables continues plutôt que des catégories.
- L'estimation de la densité, qui repose sur diverses fonctions de noyau, telles que les fonctions gaussiennes.

➤ **Exploitation minière des relations**

L'exploitation des relations vise à trouver des relations les règles les plus importantes étayées par des données pour des intérêts spécifiques. Différentes mesures d'intérêt ont été développées au fil des ans par les chercheurs, notamment le soutien et la confiance. Cependant, certaines recherches ont conclu que le lift et le cosinus sont les plus pertinents dans l'exploration de données éducatives [13]

Cette technique contient plusieurs types tels que :

- Extraction de règles d'association.
- Extraction de corrélation.
- Extraction séquentielle de motifs.
- Data mining occasionnelle.

➤ **Découvert avec des modèles**

En matière de découverte, les modèles sont généralement basés sur le regroupement, la prédiction ou l'ingénierie des connaissances en utilisant le raisonnement humain plutôt que des méthodes automatisées.

Le modèle développé est ensuite utilisé dans le cadre d'autres modèles globaux tels que l'exploration des relations.

En résumé, cette technique consiste à utiliser les résultats d'une analyse de data mining comme point de départ pour une autre analyse, souvent en chaînant plusieurs méthodes pour tirer des insights plus profonds des données.

➤ **Distillation des données en vue d'un jugement humain**

La distillation des données pour le jugement humain vise à rendre les données compréhensibles. Présenter les données de différentes manières aide le cerveau humain à découvrir de nouvelles connaissances.

La distillation des données pour le jugement humain est appliquée aux données éducatives à deux but : la classification et ou l'identification. Pour la classification peut servir de préparation pour la construction d'un modèle de prédiction [14] ; l'identification vise à présenter les données de manière à ce qu'elles soient facilement identifiables grâce à des modèles bien connus qui peuvent pas être formalisés [15].

1.4.4 Les objectifs primordiaux du data mining éducatif incluent

- Anticiper les comportements d'apprentissage à venir.
- Identifier et améliorer les modèles spécifiques au domaine de l'éducation.
- Examiner l'impact du support pédagogique.
- Contribuer à l'avancement des connaissances scientifiques sur les processus d'apprentissage et les apprenants.

1.5 Conclusion

Pour conclure cette première section l'intégration du data mining dans le domaine de l'éducation s'avère extrêmement bénéfique pour analyser les processus d'apprentissage, résoudre les problématiques rencontrées et améliorer les performances globales.

Dans ce chapitre, nous avons exposé les concepts fondamentaux, les méthodes ainsi que les objectifs du data mining éducatif et de l'analyse de l'apprentissage.

Chapitre 2

Etat de l'art

2.1 Introduction

Le domaine de l'éducation a toujours eu pour objectif d'analyser et d'améliorer le processus d'apprentissage des étudiants. Le développement de la technologie et de l'analyse des données a joué un rôle de plus en plus crucial dans cette démarche. Les chercheurs ont examiné de nouvelles données provenant des systèmes d'apprentissage et ont essayé des méthodes de modélisation différentes. Des études précédentes ont démontré que la représentation des données sur les performances des étudiants peut offrir des informations utiles sur le processus d'apprentissage, permettant ainsi aux éducateurs de prendre des décisions éclairées.

Ce chapitre présente un état de l'art ou étude bibliographique dans le domaine du data mining éducatif et de l'analyse de l'apprentissage.

2.2 Études antérieures sur les applications pratiques du Data Mining éducatif et de l'Analyse de l'Apprentissage

L'exploration des données éducatives et la recherche sur l'analyse de l'apprentissage commencent à répondre à des questions de plus en plus complexes sur ce qu'un étudiant sait et sur son engagement. Les chercheurs ont expérimenté de nouvelles techniques de construction de modèles ainsi que de nouveaux types de données de systèmes d'apprentissage qui se sont révélés prometteurs pour prédire les résultats des élèves.

Ces domaines représentent les grandes catégories dans lesquelles l'exploration et l'analyse de données peuvent être appliquées à l'activité en ligne, en particulier dans les entreprises émergentes. Ces domaines d'application sont [6] :

- La modélisation des connaissances des utilisateurs
- Modélisation du comportement des utilisateurs
- Modélisation de l'expérience utilisateur
- Le profilage des utilisateurs
- Modélisation du domaine
- Analyse des composantes d'apprentissage et analyse des principes d'enseignement
- l'analyse des tendances
- Adaptation et personnalisation

2.2.1 La modélisation des connaissances des utilisateurs

Les chercheurs et les développeurs se concentrent sur la création et le perfectionnement de modèles d'utilisateurs qui représentent des données spécifiques à chaque individu, Ces modèles sont utilisés pour adapter les actions des systèmes aux besoins individuels des utilisateurs, afin de fournir la bonne information au bon moment et de la bonne manière. La modélisation des connaissances de l'utilisateur est essentielle pour développer des systèmes tels que des hypermédias adaptatifs, des systèmes de recommandation, des systèmes experts et des systèmes de tutorat intelligents. Dans ces derniers, les modèles de connaissances de l'utilisateur guident des processus clés, comme la sélection des problèmes à soumettre aux étudiants. Les chercheurs utilisent différentes méthodes pour évaluer les connaissances des étudiants, y compris le modèle de traçage des connaissances de Corbett et Anderson et des techniques basées sur l'apprentissage

automatique. Ces modèles permettent d'améliorer la précision des prédictions concernant les performances futures des étudiants. [6]

2.2.2 Modélisation du comportement des utilisateurs

La modélisation du comportement de l'utilisateur dans le domaine de l'éducation est cruciale pour évaluer l'engagement des élèves et adapter leur expérience d'apprentissage en conséquence. Elle repose sur les mêmes types de données d'apprentissage utilisés pour estimer les connaissances de l'utilisateur, ainsi que sur d'autres mesures telles que le temps passé en ligne, la finalisation des cours, les changements documentés dans le contexte scolaire, l'assiduité et les retards. Parfois, le niveau de connaissance de l'élève est également déduit de son activité avec le système d'apprentissage ou d'autres sources de données comme les résultats des tests standardisés [6]. Des entreprises comme Carnegie Learning utilisent la modélisation du comportement de l'utilisateur pour détecter des changements dans les salles de classe et aider les enseignants à diagnostiquer les problèmes d'apprentissage. En dehors du domaine éducatif, des entreprises de jeux sociaux, telles que Zynga, utilisent également la modélisation du comportement des utilisateurs pour prédire les actions des joueurs et améliorer l'engagement.

En résumé, la modélisation du comportement de l'utilisateur offre un moyen puissant d'analyser et de comprendre l'apprentissage en ligne, en répondant à des questions importantes sur les interactions entre les élèves et les systèmes d'apprentissage, ainsi que sur les facteurs qui influent sur l'apprentissage et l'engagement des élèves.

2.2.3 La Modélisation de l'expérience utilisateur

La modélisation de l'expérience utilisateur dans le domaine de l'éducation est essentielle pour évaluer l'engagement des élèves et personnaliser leur expérience d'apprentissage. Elle repose sur diverses données, telles que les réponses aux enquêtes, les choix, les comportements, les performances et la rétention des étudiants dans les cours ultérieurs. La modélisation de l'expérience de l'utilisateur a été très populaire dans des applications basées sur le Web telles que les achats en ligne. Certaines des entreprises

interrogées modélisent l'expérience de l'utilisateur par des méthodes autres que l'exploration de données. Les entreprises comme Zynga demandent activement les réactions des utilisateurs par le biais d'enquêtes ou mènent des études sur les utilisateurs pour améliorer leur expérience [6].

Des études ont examiné comment les établissements d'enseignement supérieur exploitent les données des étudiants pour informer les pratiques pédagogiques et améliorer les résultats d'apprentissage. Les entreprises commerciales, comme Kaplan, ont utilisé des analyses pour repenser leurs cours et ont constaté une amélioration significative de la satisfaction et de la réussite des étudiants avec les nouveaux modèles de cours [6].

2.2.4 Le profilage des utilisateurs

Un profil d'utilisateur est un ensemble de données personnelles décrivant les caractéristiques essentielles d'un utilisateur. Le profilage des utilisateurs consiste à créer et appliquer des profils d'étudiants ou de groupes à l'aide d'algorithmes d'exploration de données et d'apprentissage automatique.

Dans le domaine de l'éducation, cela vise à fournir des environnements d'apprentissage adaptés aux préférences et objectifs des étudiants pour maximiser l'efficacité de l'apprentissage. Ces approches reposent sur la catégorisation des utilisateurs en fonction de leurs données démographiques, de leurs comportements d'apprentissage et de leurs interactions avec les systèmes d'apprentissage. [6]

2.2.5 La Modélisation du domaine

La modélisation du domaine consiste à représenter les concepts clés d'une matière ou d'un sujet, comme les mathématiques ou l'histoire de l'art, ainsi que les relations entre ces concepts. Dans le domaine de l'éducation, cette modélisation explore comment les différences dans la structuration d'un sujet en plusieurs domaines affectent l'apprentissage. Les chercheurs utilisent l'exploration de données pour étudier ces différences en examinant les taxonomies du domaine, les associations entre les compétences, les réponses des utilisateurs et les actions au fil du temps sur des ressources d'apprentissage individuelles. Cette approche

visé à améliorer les systèmes d'apprentissage pour mieux servir l'enseignement et l'apprentissage personnalisés [6].

2.2.6 Analyse des composantes d'apprentissage et analyse des principes d'enseignement

L'analyse des principes pédagogiques examine les composantes d'un système d'apprentissage et les pratiques pédagogiques adoptées pour répondre à des questions telles que celles sur l'efficacité des différentes pratiques et programmes d'études. Pour ce faire, il faut collecter des données sur les performances des élèves, les actions sur les composants du système d'apprentissage, l'application des stratégies pédagogiques et les résultats des tests.

Par exemple, Carnegie Learning, Inc. et l'université Carnegie Mellon ont développé des modèles cognitifs de mathématiques, intégrés dans des programmes utilisant le Cognitive Tutor, un système de tutorat intelligent [6].

2.2.7 L'Analyse des tendances

L'analyse des tendances consiste à collecter des informations et à repérer des modèles séquentiels dans les données au fil du temps. Cette pratique est utilisée par les entreprises en ligne pour prédire les intérêts des utilisateurs et suivre l'évolution de leur participation. Dans l'éducation, elle permet d'étudier les changements dans l'apprentissage des élèves et d'évaluer l'impact des politiques éducatives.

Dans le domaine de l'exploration des données éducatives, l'analyse des tendances vise à extraire un modèle sous-jacent à partir de données longitudinales provenant d'au moins trois points dans le temps. Par exemple, la Postsecondary Education Commission of California propose un outil en ligne pour examiner les tendances dans les données éducatives, identifier des anomalies et générer des rapports personnalisés sur divers aspects de l'éducation [6].

2.2.8 Adaptation et personnalisation

Ce champ d'application couvre la manière d'utiliser l'analyse pour ajuster ou personnaliser l'expérience utilisateur. Adapté pour être utilisé dans les instructions Modifications apportées aux élèves par le système ou les enseignants pour personnaliser leur Expérience, personnalisation telle que définie par le ministère américain de l'Éducation (2010), montrant que, sur la base des préférences d'apprentissage et Adapter le contenu aux intérêts des étudiants. Les adaptations peuvent inclure Fournir aux étudiants des suggestions ou des commentaires sur les prochaines étapes et leur intérêt pour les systèmes d'apprentissage en ligne (par ex. Le contenu est différent et vous pouvez améliorer votre progression dans le cours en pratiquant davantage.

Des études, comme celles menées par des chercheurs, montrent comment ces techniques peuvent améliorer les environnements éducatifs en adaptant le système d'enseignement aux méthodes d'apprentissage préférées des étudiants. Ces applications représentent le potentiel de l'exploration des données éducatives et de l'analyse de l'apprentissage, bien que certaines nécessitent encore des recherches approfondies [6].

2.3 Les processus d'application de l'EDM

Le processus d'application du data mining dans le domaine de l'éducation peut être décrit comme un cycle itératif de formulation d'hypothèses, de tests et de raffinement. L'objectif de ce processus va au-delà de simplement transformer les données en connaissances ; il s'agit également de filtrer ces connaissances extraites afin de déterminer comment modifier l'environnement éducatif pour améliorer l'apprentissage des apprenants [16].

Le processus d'Extraction de Connaissances à partir de Données (ECD). Démarre par la collecte de données à partir de **l'environnement éducatif**. Les **données brutes** ainsi obtenues nécessitent d'être nettoyées et **prétraitées**, notamment par la fusion **de données hétérogènes**, le traitement des données manquantes, ou encore la conversion des données. Cette phase implique souvent l'utilisation de techniques de data mining. Le résultat de cette

phase est un **modèle** qui permet de structurer les données stockées. Enfin, la dernière étape consiste en l'**interprétation et l'évaluation** des résultats obtenus [16].

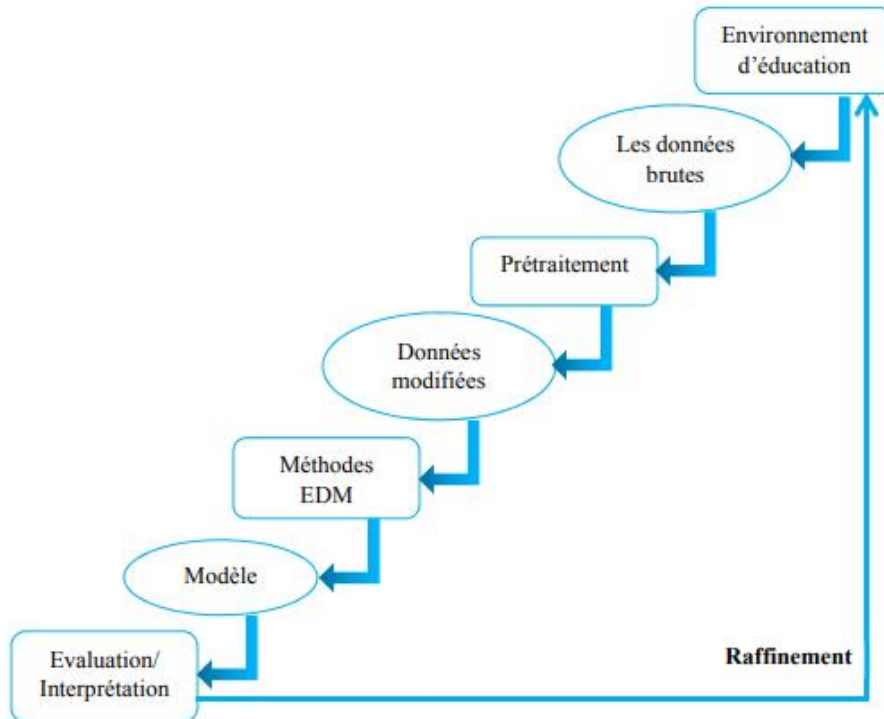


Figure 03 : Le processus d'application de Data mining appliqué dans l'éducation [16]

2.4 Analyse de l'apprentissage dans les établissements d'enseignement

L'examen le plus récent des applications du data mining éducatif dans l'enseignement supérieur a été réalisé en 2019 par Aldowaha et ses collègues. Leur étude, qui a analysé 402 articles, a révélé que la classification est la technique d'exploration de données la plus couramment utilisée dans les learning analytics (LA) de l'enseignement supérieur, avec une utilisation de 26,25 %. Cette technique est principalement utilisée pour prédire les performances des étudiants, identifier les risques d'abandon et estimer les résultats et les notes finales.

La recherche en enseignement supérieur se concentre sur l'utilisation de techniques de fouille de données comme le regroupement, la classification, la visualisation et l'analyse d'association. Bien qu'exploratoires, ces méthodes offrent des perspectives prometteuses pour améliorer l'enseignement et l'apprentissage [17].

Par exemple, une étude menée à l'**Université de Floride** a examiné l'utilisation des données de Moodle pour prédire le sentiment de communauté des étudiants. Les résultats ont montré que le nombre total d'activités des étudiants dans le système était un prédicateur valide de ce sentiment [17].

En outre, Les outils de fouille de données intégrés aux systèmes de gestion de contenu open source, tels que Moodle, permettent aux enseignants d'analyser leurs propres cours. Ces outils offrent diverses fonctionnalités, telles que la sélection de groupes d'étudiants similaires, la détection de comportements d'apprentissage inappropriés et la prédiction des résultats probables des étudiants [17].

Les chercheurs ont la capacité de regrouper les données provenant de la base de données Moodle, un système largement adopté dans les universités et institutions du monde entier. Moodle vise à simplifier la visualisation des connaissances pour les enseignants, ce qui permet d'identifier des groupes d'étudiants similaires en fonction de leur comportement dans le système d'apprentissage en ligne, notamment leur activité, leur participation aux forums de discussion et leurs soumissions de devoirs [18] [19].

En bref, les méthodes actuelles d'exploration de données permettent de trouver de nouveaux modèles dans les données et de créer de nouveaux algorithmes ou modèles, tandis que l'analyse de l'apprentissage applique les modèles prédictifs déjà présents dans les systèmes éducatifs. Dans le cadre de l'enseignement supérieur, diverses techniques d'exploration de données sont utilisées à diverses fins, telles que la prise de décision, la planification avancée de l'orientation des étudiants, la prévision précise des tendances et des comportements futurs des individus et une allocation plus efficace des ressources et du personnel [20].

Il est crucial d'utiliser les GED et les AL dans l'enseignement supérieur pour améliorer les expériences et les résultats d'apprentissage des étudiants [20].

2.5 Les Techniques de l'Exploration de Données

Les organisations utilisent l'exploration de données pour trouver des modèles dans les données qui pourraient fournir des informations sur leurs besoins opérationnels. Elle est nécessaire à la fois pour la veille stratégique et la science des données. Les organisations peuvent utiliser diverses techniques d'exploration de données pour transformer des données non structurées en informations exploitables.[21]

L'objectif principal du data mining est de trouver des corrélations cachées, non planifiées et précédemment non identifiées mais légitimes dans les données. Elle est décrite comme une méthode permettant de séparer les données utiles d'une énorme collection de données brutes.

Il s'agit d'utiliser un ou plusieurs logiciels pour analyser des modèles de données dans des ensembles de données importants. La recherche et la science ne sont que deux exemples des nombreux domaines dans lesquels l'exploration de données peut être utilisée.

Les données peuvent être extraites à l'aide de différentes techniques pour diverses applications de science des données. La détection d'anomalies, qui tente de trouver des valeurs aberrantes dans d'énormes ensembles de données, et la reconnaissance de formes sont des cas d'utilisation courants de l'exploration de données rendus possibles par diverses techniques.

Les experts en exploration de données ont consacré leurs efforts à l'amélioration de nos connaissances sur la manière d'analyser et de tirer des conclusions d'énormes quantités d'informations. Ils s'appuient sur des méthodes et des technologies issues de la convergence de l'administration des bases de données, des statistiques et de l'apprentissage automatique. [21]

2.6 Conclusion

Les techniques de data mining ont été appliquées dans plusieurs domaines. Leur but est d'analyser les données volumineuses et complexes. Un des domaines d'application de ces techniques est l'éducation.

Dans cette section, nous avons examiné plusieurs recherches précédentes qui ont mis en évidence l'importance des données dans l'évaluation de la performance des étudiants, en vue de mieux comprendre et d'améliorer le processus d'apprentissage. Le prochain chapitre se penchera sur l'analyse et la conception de notre propre système.

Chapitre 3

Analyse et Conception du système

3.1 Introduction

L'étape d'analyse et de spécification des besoins joue un rôle crucial dans le développement des systèmes d'information en facilitant leur cycle de vie. Cette phase permet de préciser de manière optimale les besoins fonctionnels et non fonctionnels.

Dans ce chapitre, nous aborderons le contexte du projet, exposerons la problématique ainsi que la solution envisagée, et présenterons en détail la conception de notre application web, y compris les divers diagrammes.

3.2 Définition de problème

Dans le paysage éducatif actuel, les systèmes de gestion de l'apprentissage (LMS) sont indispensables pour faciliter l'enseignement universitaire et suivre les progrès des étudiants. Ces plateformes numériques offrent aux étudiants un accès aux cours, aux activités et aux soumissions via des plateformes en ligne. De plus, ces LMS génèrent une grande quantité de données sur les interactions entre les étudiants, ce qui permet d'analyser les modèles d'apprentissage et d'améliorer les méthodes d'enseignement.

Dans ce contexte, une nouvelle problématique émerge. Comment les données des systèmes de gestion de l'apprentissage peuvent être utilisées pour mieux comprendre les performances des étudiants à l'université ? Et Comment l'exploration de données éducatives peut être utilisée pour identifier les signaux de réussite ou d'échec des étudiants ?

3.3 Solution proposée

Pour résoudre ce problème, nous proposons d'implémenter une solution basée sur l'algorithme K-means, une méthode de clustering, dans un système de gestion de l'apprentissage des étudiants universitaires. Les étudiants peuvent être regroupés selon leur interaction avec la plateforme d'apprentissage grâce au K-means. Bien que K-means ne fasse pas de prédiction directe, il aide à identifier des groupes d'étudiants similaires. Par la suite, il est possible d'appliquer des algorithmes prédictifs à ces groupes afin d'évaluer les taux de succès et d'échec. En se basant sur les données telle la note finale obtenue dans chaque module, le sexe, l'origine ethnique, le niveau d'éducation des parents et préparation aux tests.

3.4 Exploration Approfondie des Données

L'enseignement à distance gagne en popularité, offrant aux étudiants une grande flexibilité tout en permettant aux établissements d'enseignement de fournir un contenu varié et accessible.

Cependant, ce mode d'apprentissage présente des défis en matière d'analyse des données et de personnalisation de l'expérience d'apprentissage. Pour relever ces défis, l'algorithme k-means se révèle être une solution prometteuse. En regroupant les données non étiquetées en clusters homogènes, il permet une analyse précise des comportements et des performances des étudiants dans l'enseignement à distance.

Dans cette étude, nous examinons de plus près comment l'algorithme k-means peut être appliqué pour améliorer l'enseignement à distance.

3.4.1 Analyse Data

Cycle de vie d'un projet gestion d'apprentissage :

1. Appréhender la nature du problème posé.
2. Collecte des données

3. Vérifications des données
4. Analyse exploratoire des données
5. Prétraitement des données.
6. Entraînement au modèle
7. Choix du meilleur modèle

Énoncé du problème :

Ce projet vise à évaluer l'influence de diverses variables sur les performances académiques des élèves, notamment leur sexe, leur origine ethnique, le niveau d'éducation de leurs parents, ainsi que leur participation à des cours de préparation aux tests.

Collecte des données :

Les données proviennent d'un ensemble de données disponible sur Kaggle, comprenant 1000 entrées réparties sur 8 colonnes.

Informations concernant toutes les données :

Sexe des élèves	homme/femme
Race/ethnie des élèves	classement A, B, C, D, E
Niveau d'éducation des parents	licence, études supérieures, maîtrise, diplôme d'associé, école secondaire
Préparation au test	effectuée ou non avant le test
Scores	mathématiques, lecture, écriture

	A	B	C	D	E	F	G	H
1	gender	race/ethnicity	parental level of	lunch	test preparation	math score	reading score	writing score
2	female	group B	bachelor's degree	standard	none	72	72	74
3	female	group C	some college	standard	completed	69	90	88
4	female	group B	master's degree	standard	none	90	95	93
5	male	group A	associate's degree	free/reduced	none	47	57	44
6	male	group C	some college	standard	none	76	78	75
7	female	group B	associate's degree	standard	none	71	83	78
8	female	group B	some college	standard	completed	88	95	92
9	male	group B	some college	free/reduced	none	40	43	39
10	male	group D	high school	free/reduced	completed	64	64	67
11	female	group B	high school	free/reduced	none	38	60	50
12	male	group C	associate's degree	standard	none	58	54	52
13	male	group D	associate's degree	standard	none	40	52	43
14	female	group B	high school	standard	none	65	81	73
15	male	group A	some college	standard	completed	78	72	70
16	female	group A	master's degree	standard	none	50	53	58
17	female	group C	some high school	standard	none	69	75	78
18	male	group C	high school	standard	none	88	89	86
19	female	group B	some high school	free/reduced	none	18	32	28
20	male	group C	master's degree	free/reduced	completed	46	42	46
21	female	group C	associate's degree	free/reduced	none	54	58	61
22	male	group D	high school	standard	none	66	69	63
23	female	group B	some college	free/reduced	completed	65	75	70
24	male	group D	some college	standard	none	44	54	53
25	female	group C	some high school	standard	none	69	73	73
26	male	group D	bachelor's degree	free/reduced	completed	74	71	80
27	male	group A	master's degree	free/reduced	none	73	74	72
28	male	group B	some college	standard	none	69	54	55
29	female	group C	bachelor's degree	standard	none	67	69	75
30	male	group C	high school	standard	none	70	70	65
31	female	group D	master's degree	standard	none	62	70	75
32	female	group D	some college	standard	none	69	74	74
33	female	group B	some college	standard	none	63	65	61
34	female	group F	master's degree	free/reduced	none	56	72	65

Figure 04 : Les informations sur les étudiants [22]

Contrôles des données à effectuer :

Afin d'approfondir l'analyse des données, nous effectuerons quelques étapes suivantes de contrôle des données :

- ✓ Nous avons examiné les valeurs manquantes de notre donnée et nous avons conclu qu'il n'y a pas de valeurs manquantes dans toutes les données. (0 valeur manquantes)
- ✓ Après avoir vérifié les doublons de nos données, nous avons conclu qu'il n'y a pas de valeurs dupliquées dans l'ensemble des données. (0 valeur dupliquées)

- ✓ La dernière en vérifiant le nombre de valeurs uniques de chaque colonne et en vérifiant les statistiques de l'ensemble des données, on peut conclure que, toutes les moyennes sont très proches les unes des autres, se situant entre 66 et 68,05. De même, Tous les écarts types sont également proches variant entre 14,6 et 15,19, Même si le score minimum en mathématiques est de 0, le score minimum est bien plus élevé en écriture (10) et en lecture (17).

À présent, nous allons commencer par une étude exploratoire visant à évaluer le taux de réussite et d'échec des étudiants. Nous allons commencer par une analyse de réussite et d'échec par module. Le graphique (06) ci-dessous met en évidence ces observations :

➤ **Analyse de réussite et d'échec par module :**

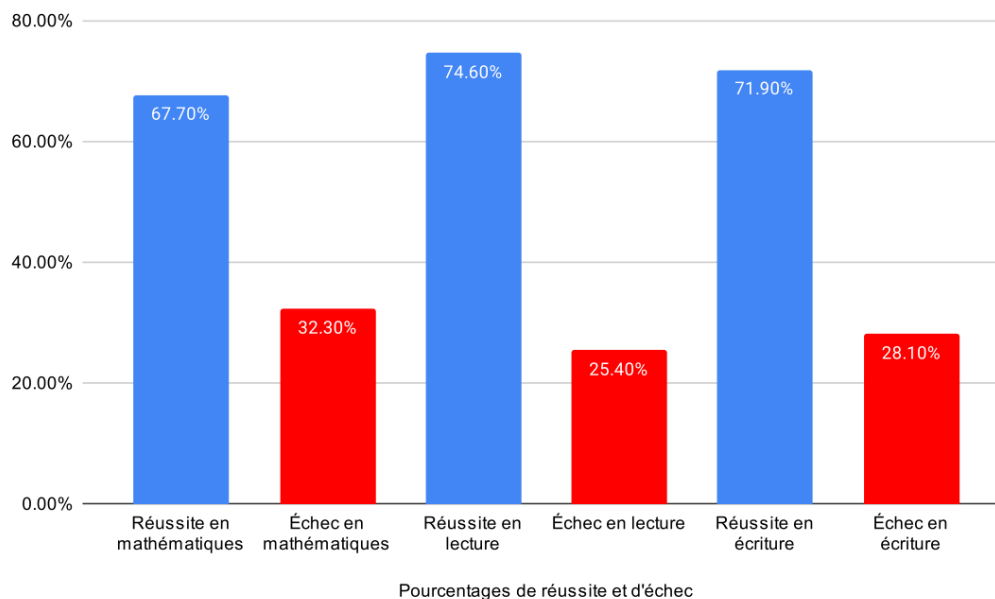


Figure 05 : pourcentage de réussite et d'échec par module

Selon cette analyse, il est établi que les étudiants obtiennent un meilleur succès dans le module de lecture par rapport aux autres modules. En outre, nous avons observé que le taux d'échec dans ces modules est plutôt bas.

Ensuite, nous allons étudier la proportion de succès et d'échec des étudiants en fonction du test de préparation pour chaque module. Les schémas (07), (08), et (09), ci-dessous illustre ces observations :

➤ **Analyse de réussite et d'échec par test de préparation pour chaque module**

1. Par le module Mathématiques :

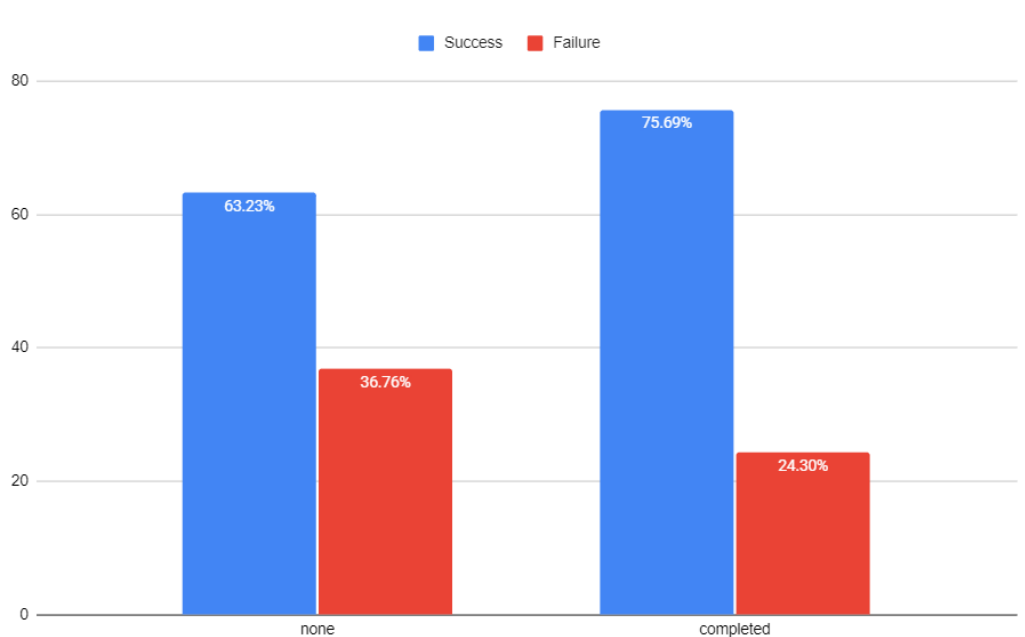


Figure 06 : pourcentage de réussite et d'échec par test de préparation pour le module de mathématiques.

2. Par le module de lecture :

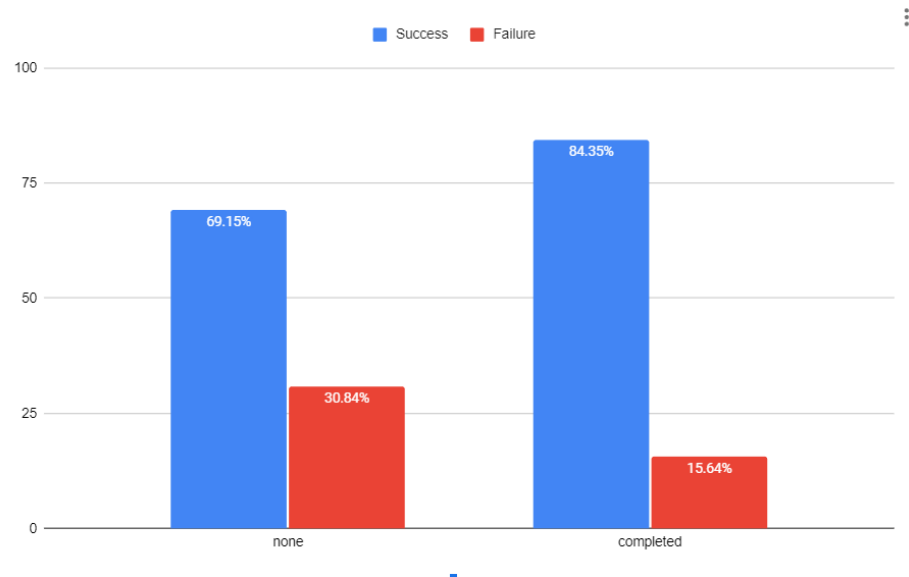


Figure 07 : pourcentage de réussite et d'échec par test de préparation pour le module de lecture.

3. Par le module de l'écriture :

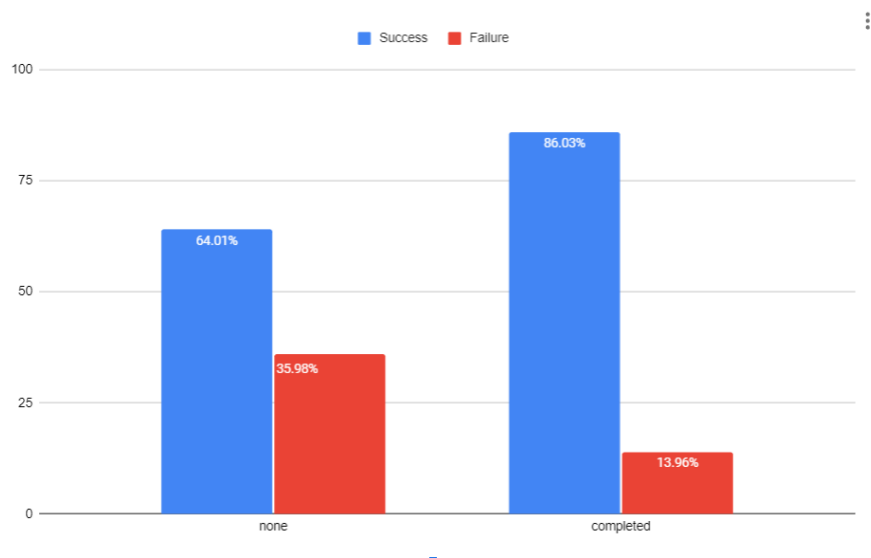


Figure 08 : pourcentage de réussite et d'échec par test de préparation pour le module de l'écriture.

De manière générale, selon les graphiques (07), (08), et (09), ceux qui ont préparé le test en suivant les modules de mathématiques, d'écriture et de lecture ont tendance à obtenir de meilleurs résultats que ceux qui n'ont pas préparé le test.

Par la suite, Pour chaque module, nous allons ensuite examiner la proportion de réussite et d'échec des étudiants en fonction du genre. Ces observations sont illustrées par le schéma (10) ci-dessous.

➤ **Analyse de réussite et d'échec par sexe pour chaque module :**

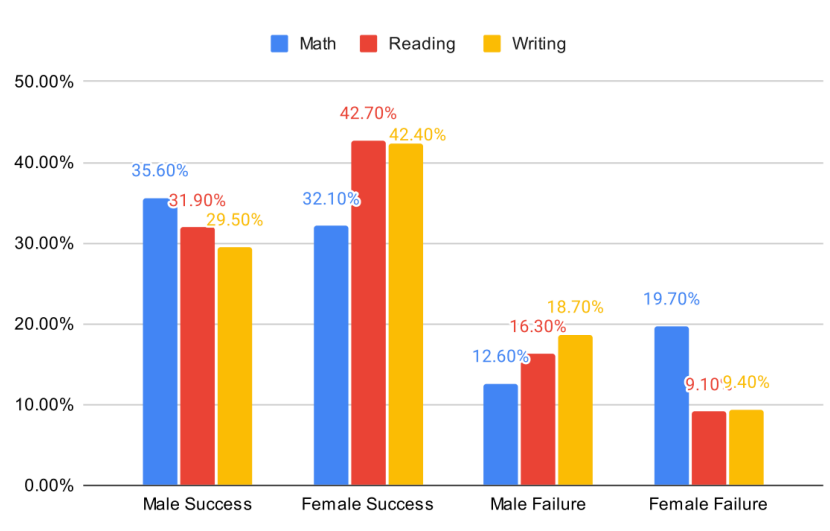


Figure 09 : pourcentage de réussite et d'échec par sexe pour chaque module

Selon le graphique, Il est démontré que les femmes réalisent des résultats supérieurs aux hommes dans les modules de lecture et d'écriture, tandis que les hommes sont plus performants que les femmes dans le module de mathématiques.

➤ **Tableau de contingence :**

Comme nous avons transformé le graphe précédent qui examinait la proportion de réussite et d'échec des étudiants en fonction du genre pour chaque module sous forme d'un tableau de contingence. La figure(11) ci-dessous illustre ces observations :

	Math	Reading	Writing
Male Success	356	319	295
Female Success	321	427	424
Male Failure	126	163	187
Female Failure	197	91	94
Total	1000	1000	1000

Figure 10 : Tableau de contingence

➤ **Corrélation :**

L'image ci-dessous montre la matrice de corrélation des différentes variables de l'ensemble de données. Cette matrice vous permet de visualiser la force et la direction de la relation linéaire entre chaque paire de variables. Le coefficient de corrélation va de -1 à 1, où -1 indique une corrélation négative parfaite, 1 indique une corrélation positive parfaite et 0 indique aucune corrélation.

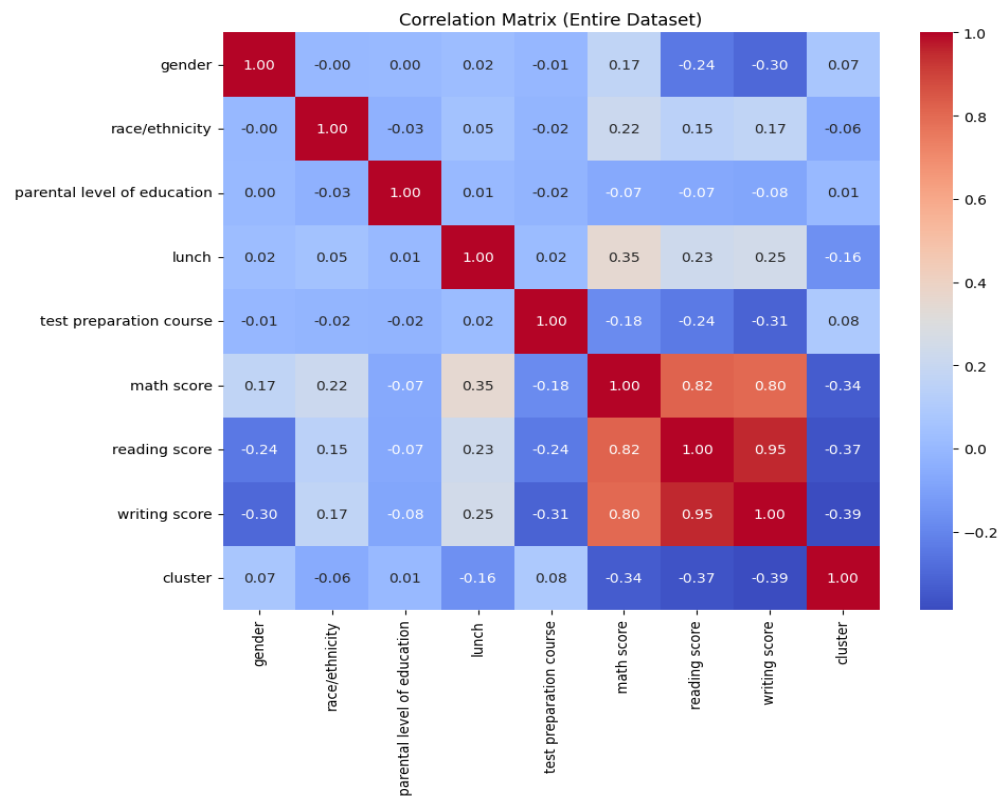


Figure 11 : Matrice corrélation

La matrice de corrélation montre qu'il existe une forte corrélation positive entre les scores en mathématiques, en lecture et en écriture, ce qui signifie que de bons scores dans une matière sont souvent accompagnés de bons scores dans d'autres matières. La corrélation entre les scores et le sexe était faible, indiquant une légère différence de performance entre les filles et les garçons, mais cette différence n'était pas significative. D'autres variables, telles que la race, le niveau d'éducation des parents et le petit-déjeuner, présentaient des corrélations plus faibles avec les scores, ce qui suggère qu'elles avaient peu d'impact sur les performances scolaires.

3.4.2 K-means ou (K-moyennes)

K-means est une méthode d'analyse de données qui regroupe un ensemble de points en clusters. L'objectif de l'algorithme k-means, également connu sous le nom d'algorithmes des centres mobiles, est de regrouper un ensemble de données en k clusters différents. Chaque enregistrement est attribué au cluster dont le centre (appelé centroïde) est le plus proche en termes de distance [23].

3.4.3 Algorithme

Voici l'algorithme K-means, qui est utilisé pour regrouper des données en k clusters. Le processus commence par sélectionner k points aléatoires comme centroïdes. Ensuite, chaque point de données est attribué au centroïde le plus proche. Les centroïdes sont alors mis à jour pour être la moyenne des points de leur cluster. Ce processus se répète jusqu'à ce que les centroïdes ne changent plus. Enfin, l'algorithme renvoie les données classées en clusters.

Entrée : Ensemble de données S , nombre de clusters k
Sortie : Ensemble de données \hat{S} où chaque enregistrement x_i est associé à une classe C_j

- Choisir aléatoirement k points comme centroïdes (ξ) pour les k clusters

Répéter

- Pour** chaque enregistrement x_i dans S **faire**
 - $min \leftarrow +\infty$
 - Pour** chaque centroïde ξ_k **faire**
 - $distance \leftarrow d(x_i, \xi_k)$
 - Si** $distance \leq min$ **alors**
 - $min \leftarrow distance$
 - $j \leftarrow k$ // j est le numéro du centroïde avec $min(d(x_i, \xi_k))$
 - Fin si**
- Fin pour**
 - $x_{i,p+1} \leftarrow C_j$ // Associer x_i au cluster C_j (création d'un nouvel attribut $p+1$)
- Fin pour**
 - //mettre à jour le centroïde ξ_j . La valeur du centroïde est égale à la moyenne des enregistrements associés au centre C_j
 - $\xi_j \leftarrow \frac{\sum x_i \text{ où } x_{i,p+1} = C_j}{|C_j|}$

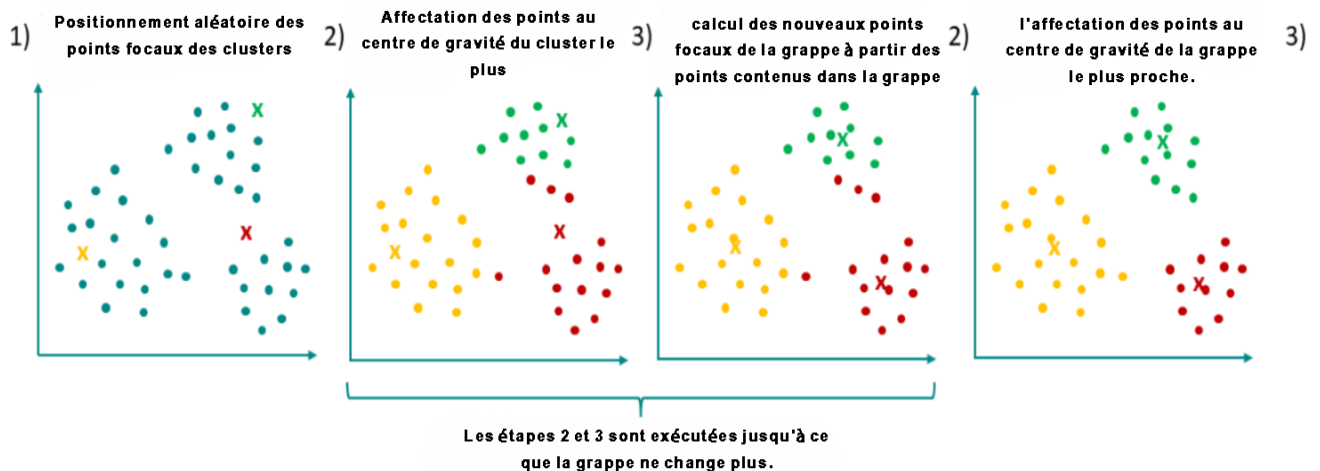
Jusqu'à que les valeurs $x_{i,p+1}$ se stabilisent //aucun enregistrement ne change de classe

- Renvoyer** \hat{S}

Figure 12: Algorithme K-means [23]

3.4.4 Exemple d'application de l'algorithme K-means

Dans cette section, nous allons illustrer le fonctionnement pratique de l'algorithme K-means en l'appliquant à un exemple concret. Ce cas d'étude permettra de visualiser les étapes clés de l'algorithme et d'appréhender son utilisation dans un contexte réel.



3.4.5 Application de K-means

Les graphiques ci-dessous présentent deux indices utilisés pour déterminer le nombre optimal de clusters (k) dans un algorithme de k-means :

- **Méthode du coude:**

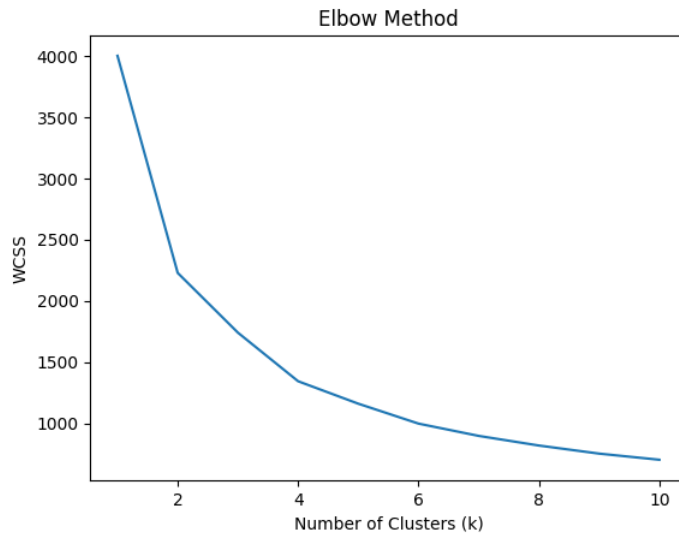


Figure 13: Utilisation de la méthode du coude pour trouver k

- **Méthode du score de silhouette**

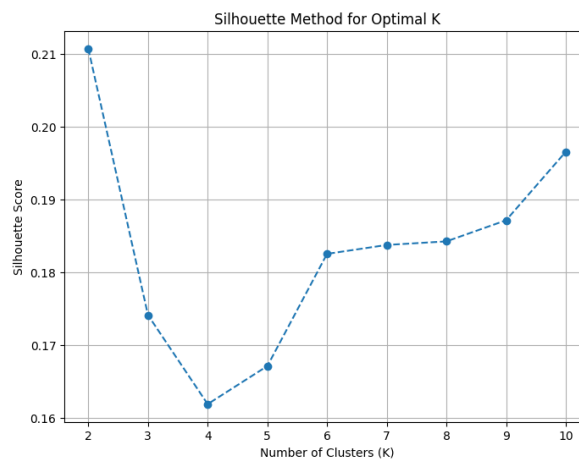


Figure 14: Utilisation de la méthode Silhouette pour trouver k

En combinant ces observations, nous en déduisons que ($k = 4$) est le nombre optimal de clusters pour notre analyse. Ce choix permet d'équilibrer de manière efficace la compacité des clusters et leur séparation.

➤ **Analyse de cluster des performances des étudiants :**

Le tableau suivant présente une analyse des scores moyens en mathématiques, lecture et écriture, segmentés en quatre clusters distincts (C0, C1, C2, C3). Chaque cluster est décrit par les moyennes des scores dans les trois matières, la répartition des sexes, la préparation aux tests, et le niveau d'éducation des parents. La figure(15) ci-dessous illustre ces observations:

CLUSTER	Math AVG	Reading AVG	Writing AVG	Sexe	Test preparation	Parental level of education
C0	50.97	56.11	53.97	- Male : 51.80% - Female : 48.20%	- Completed Percentage: 77.93% - None Percentage: 22.07%	- some college (Label 4): 23.42% - high school (Label 2): 22.52% - some high school (Label 5): 20.27% - associate's degree (Label 0): 18.47% - bachelor's degree (Label 1): 9.46% - master's degree (Label 3): 5.86%
C1	75.65	78.76	77.38	- Male : 34.86% - Female : 65.14%	- Completed Percentage: 100.00% - None Percentage: 0.00%	- associate's degree (Label 0): 25.70% - some college (Label 4): 22.18% - high school (Label 2): 16.90% - bachelor's degree (Label 1): 14.08% - some high school (Label 5): 12.68% - master's degree (Label 3): 8.45%
C2	58.04	57.23	55.58	- Male : 67.12% - Female : 32.88%	- Completed Percentage: 83.33% - None Percentage: 16.67%	- high school (Label 2): 26.58% - some college (Label 4): 22.97% - some high school (Label 5): 22.52% - associate's degree (Label 0): 18.02% - bachelor's degree (Label 1): 7.66% - master's degree (Label 3): 2.25%
C3	75	79.54	79.98	- Male : 43.75% - Female : 56.25%	- Completed Percentage: 0.00% - None Percentage: 100.00%	- associate's degree (Label 0): 25.00% - some college (Label 4): 22.06% - some high school (Label 5): 17.65% - bachelor's degree (Label 1): 14.71% - high school (Label 2): 14.34% - master's degree (Label 3): 6.25%

Figure 15: Analyse de cluster des performances des étudiants

✓ **Analyse des Résultats des Étudiants par Cluster :**

- **Moyenne des scores :** Le tableau montre les scores moyens pour les mathématiques, lecture et l'écriture dans chaque cluster. Le cluster C1 a les scores moyens les plus élevés dans l'ensemble, ce qui suggère un groupe d'étudiants performants. Le cluster C0 a les scores moyens les plus bas.
- **Sexe :** La répartition par sexe est détaillée pour chaque cluster. Le cluster C1 compte une majorité de femmes (65,14 %), tandis que le cluster C2 est majoritairement composé d'hommes (67,12 %). Le cluster C3 montre une répartition équilibrée entre les sexes.
- **Préparation aux tests :** La colonne "Préparation aux tests" montre le pourcentage d'étudiants ayant fait le test de préparation. Le cluster C1 est intéressant car tous les étudiants de ce cluster ont terminé le test de préparation. Les clusters C0 et C2 montrent un pourcentage élevé d'étudiants suivant le test préparation tandis que le cluster C3 Tous les étudiants de ce cluster n'ont pas fait le test de préparation (100 %).
- **Niveau d'éducation parental :** Cette colonne montre la distribution des niveaux d'éducation des parents dans chaque cluster, ce qui pourrait indiquer l'influence de l'éducation parentale sur les performances des étudiants.

3.5 Spécification des besoins

3.5.1 Besoins fonctionnels

Notre système doit répondre aux exigences fonctionnelles suivantes :

- Analyser les données pour créer des graphiques détaillés.
- Détermination du pourcentage de réussite et d'échec des étudiants
- Suivi des étudiants selon la ponctualité de leurs devoirs etc.
- Administration des comptes utilisateurs.
- Gestion des cours.
- Facilitation de la communication entre les étudiants et les enseignants.

3.5.2 Besoins non fonctionnels

- Assurer la confidentialité des données des étudiants.
- Fournir une plateforme fiable.
- Garantir la protection des données collectées.
- Assurer une présentation claire et facile à comprendre des données.

3.6 Analyse des besoins

3.6.1 Acteurs du système

Dans cette partie, nous décrirons les différents intervenants du système ainsi que leurs rôles respectifs.

- **Enseignant :**

L'enseignant est chargé d'assurer le suivi des étudiants dans leurs modules. Par conséquent, ses fonctions incluent :

- L'enseignant a la possibilité d'inclure ou de retirer différents éléments tels que des fichiers, des vidéos, des questionnaires, des devoirs, des QCM, des annonces et toute information qu'il veut partager avec ses étudiants.
- L'enseignant peut s'identifier suivant un compte créé par l'administrateur du système.

- **Etudiant :**

L'étudiant joue un rôle central dans le processus d'apprentissage, l'étudiant peut effectuer les fonctionnalités suivantes :

- Chaque étudiant doit s'inscrire en remplissant un formulaire d'inscription avec toutes les informations requises.
- Les étudiants s'identifient chacun suivant son matricule.
- L'étudiant a la possibilité de consulter et de compléter les devoirs et les QCM.
- L'étudiant peut télécharger des fichiers et consulter des ressources

- **Administrateur :**

Cet acteur contrôle tous les gestes techniques. Son rôle est

- Gérer les inscriptions de tous les utilisateurs.
- Gérer la base de données.
- Gérer les ressources (Ajouter, Supprimer, Modifier,...)

3.7 Modélisation de notre système

Pour la réalisation de notre application, notre choix a été porté sur le Processus Unifié. En effet, le processus unifié est une solution de développement logiciel adaptée à tout type de projet. Ces traits distinctifs tiennent compte de trois notions : piloté par les cas d'utilisation, centré sur l'architecture, itératif et incrémental.

La démarche de modélisation que nous avons utilisée est la méthode RUP avec les 6 modèles suivants :

- Modèle de cas d'utilisation
- Modèle d'analyse
- Modèle de conception
- Modèle d'implémentation
- Modèle de déploiement
- Modèle de test

Le langage de modélisation que nous avons utilisé est UML, qui est une partie intégrante de la démarche RUP. Ces diagrammes sont largement utilisés dans chaque étape et phase de ce processus de développement.

3.7.1 Modèle de cas d'utilisation

- **Diagramme de profil:** Un diagramme de profils est un diagramme de structure permettant l'utilisation de profils pour un méta-modèle donné.

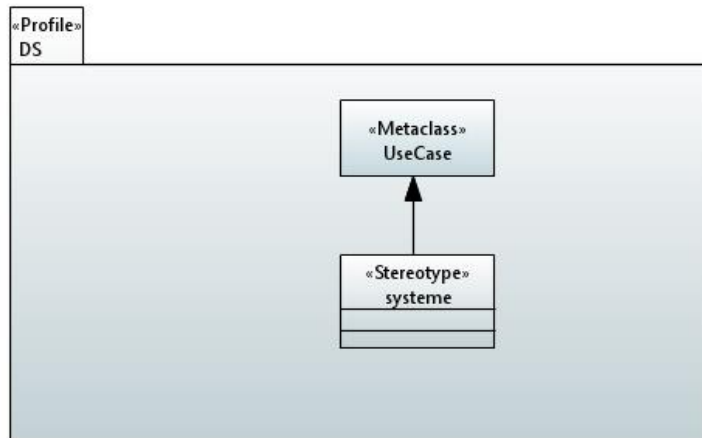


Figure 16: Diagramme de profil

- **Le Diagramme de contexte statique:** Délimite le domaine d'étude en précisant ce qui est à la charge du système et en identifiant l'environnement extérieur au système étudié avec lequel ce dernier communique. On représente les interactions des acteurs avec le système étudié.

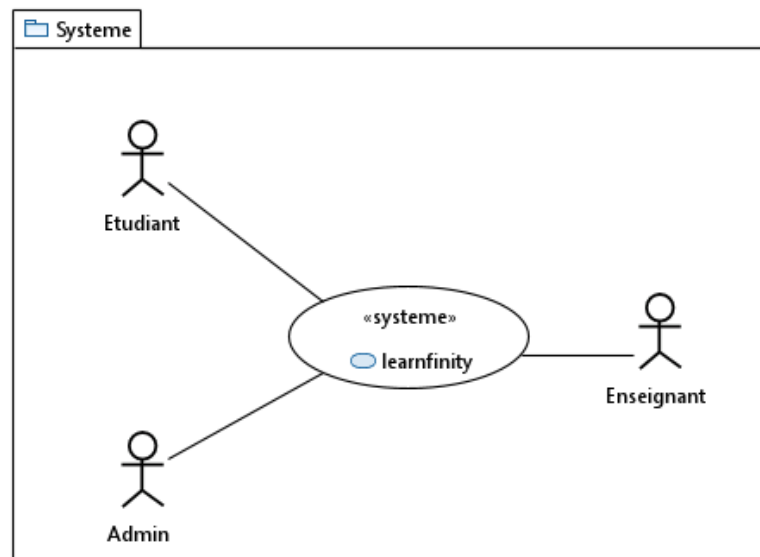


Figure 17 : Diagramme de contexte statique Systeme

- **Les diagrammes de package** : (ou diagramme de paquetages) sont des diagrammes structurels utilisés pour représenter l'organisation et la disposition de divers éléments modélisés sous forme de paquetages.

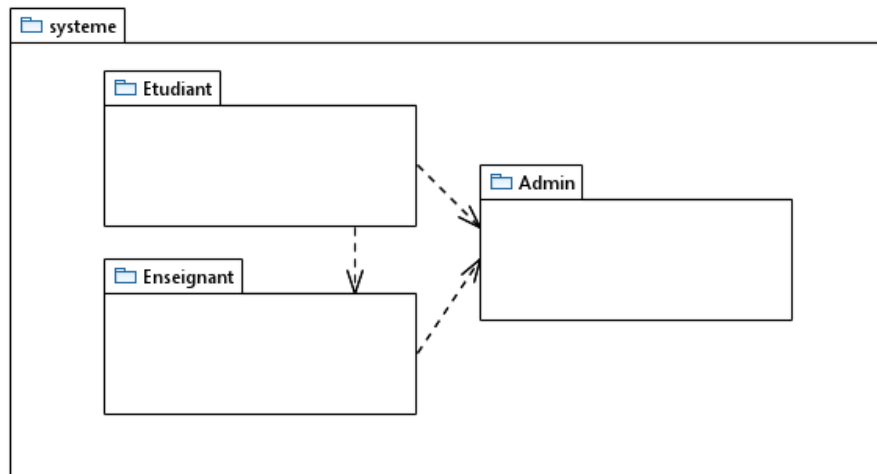


Figure 18 : Diagramme de package

- **Le diagramme de cas d'utilisation (use case)** : Le diagramme de cas d'utilisations représente les fonctions du système du point de vue des utilisateurs. C'est le diagramme principal du modèle UML, celui qui assure la relation entre l'utilisateur et les objets que le système met en œuvre. Dans notre application, nous avons utilisé le diagramme de cas d'utilisation pour obtenir tous les cas d'utilisation possibles de ce programme.

➤ Diagramme de cas d'utilisation de l'étudiant

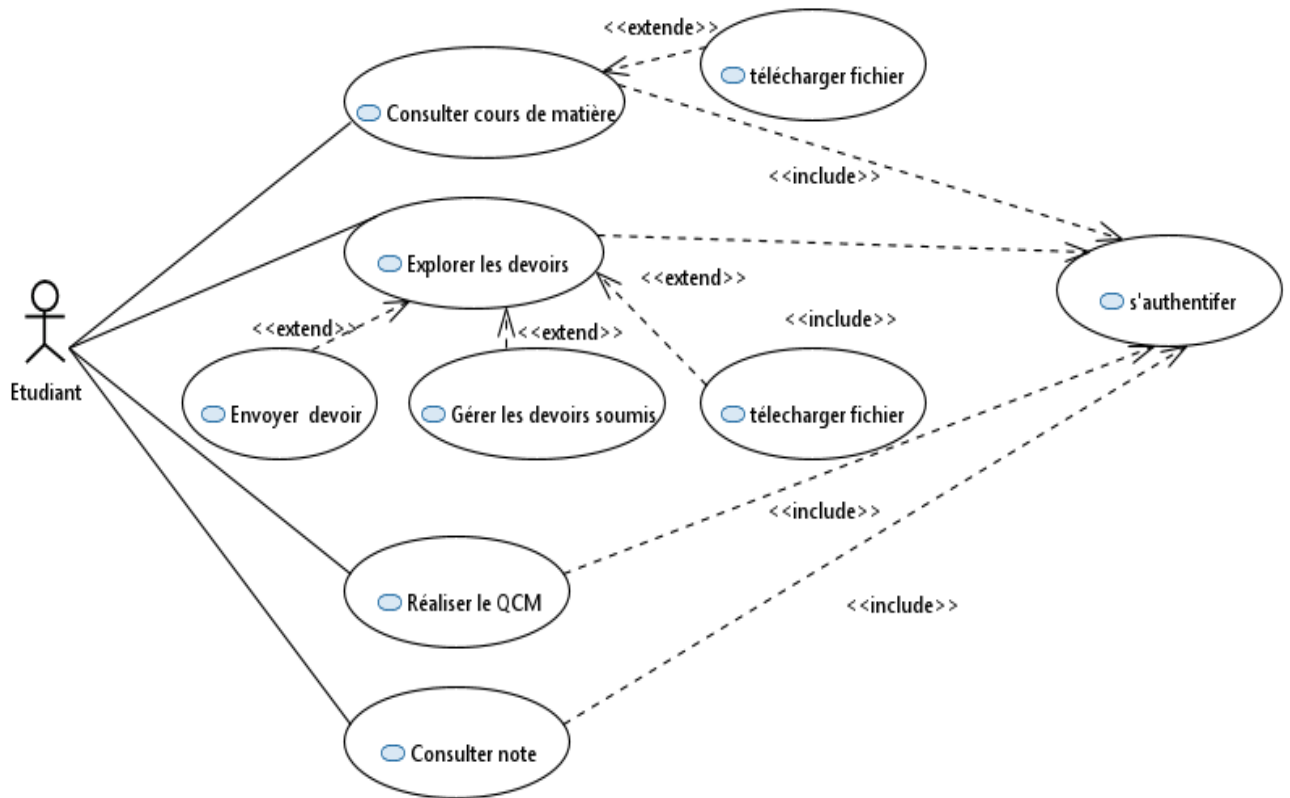


Figure 19: Diagramme de cas d'utilisation pour l'étudiant

➤ Diagramme de cas d'utilisation de l'enseignant

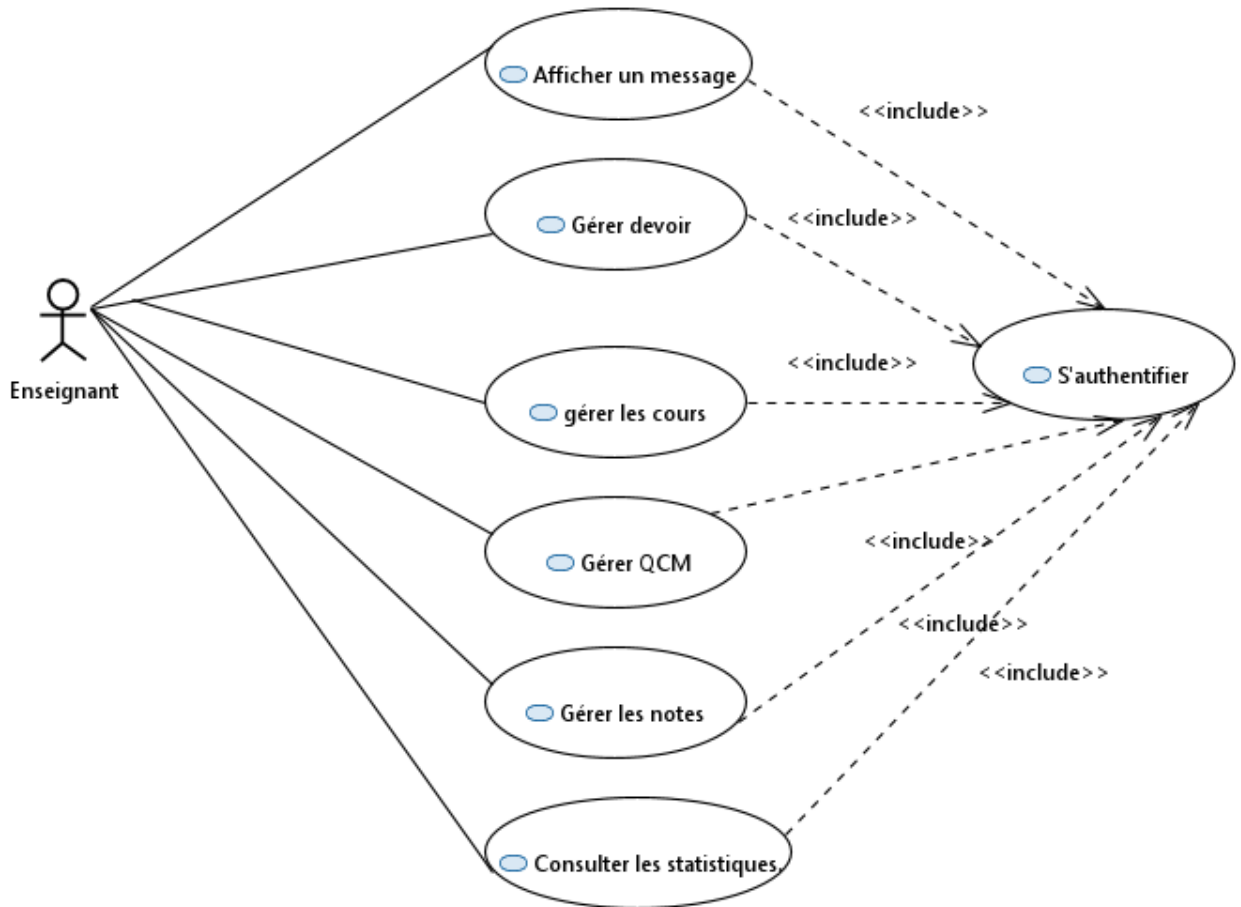


Figure 20 : Diagramme de cas d'utilisation pour l'enseignant

➤ Diagramme de cas d'utilisation de l'administrateur

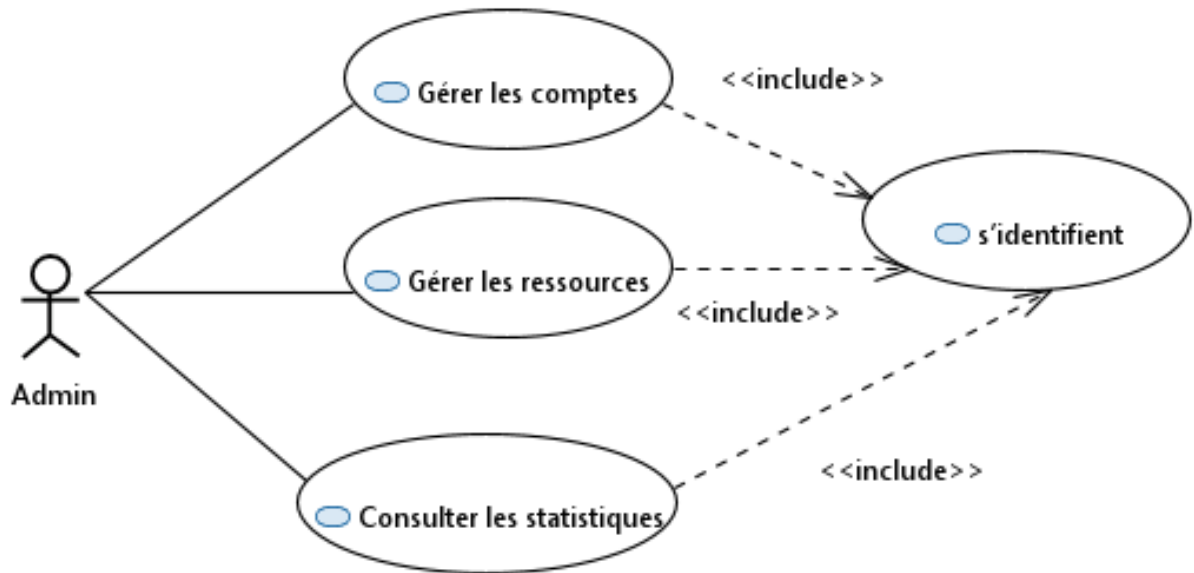


Figure 21 : Diagramme de cas d'utilisation pour l'adminis

➤ **Diagrammes de séquences :** Les diagrammes de séquences sont la représentation graphique des interactions entre les acteurs et le système selon un ordre chronologique dans la formulation UML.

- **Diagramme de séquence s'authentifier :**

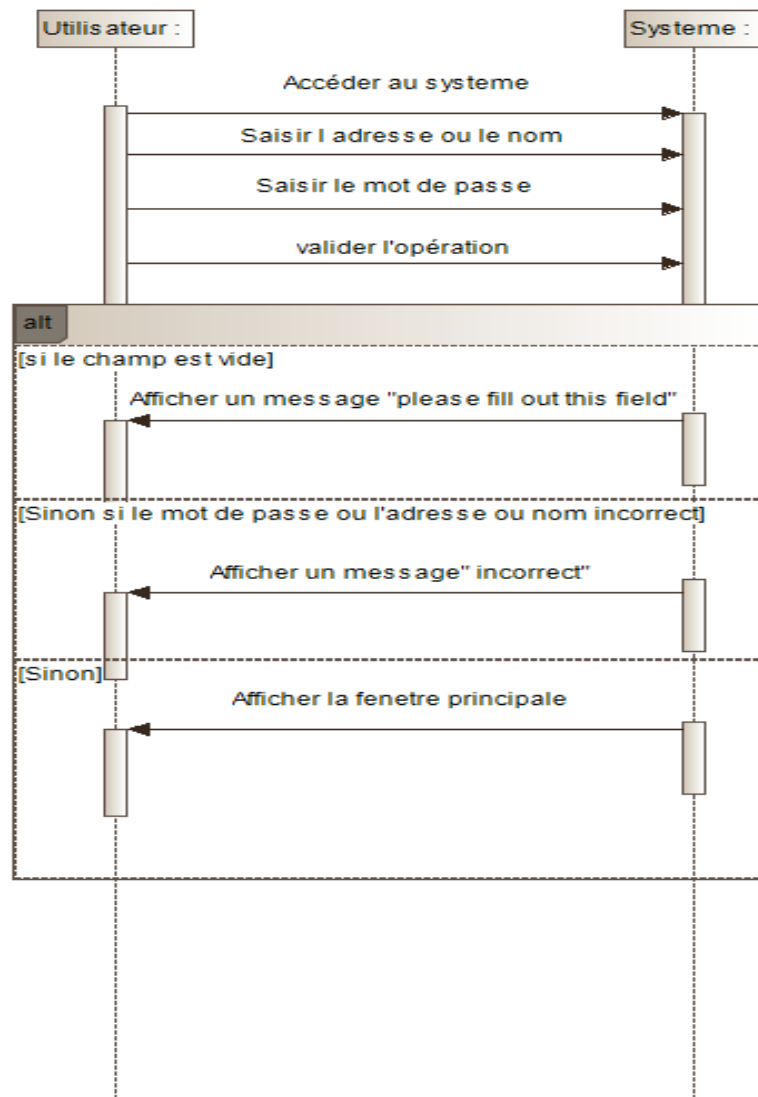


Figure 22 : Diagramme de séquence s'authentifier

- Diagramme de séquence Créer un QCM :

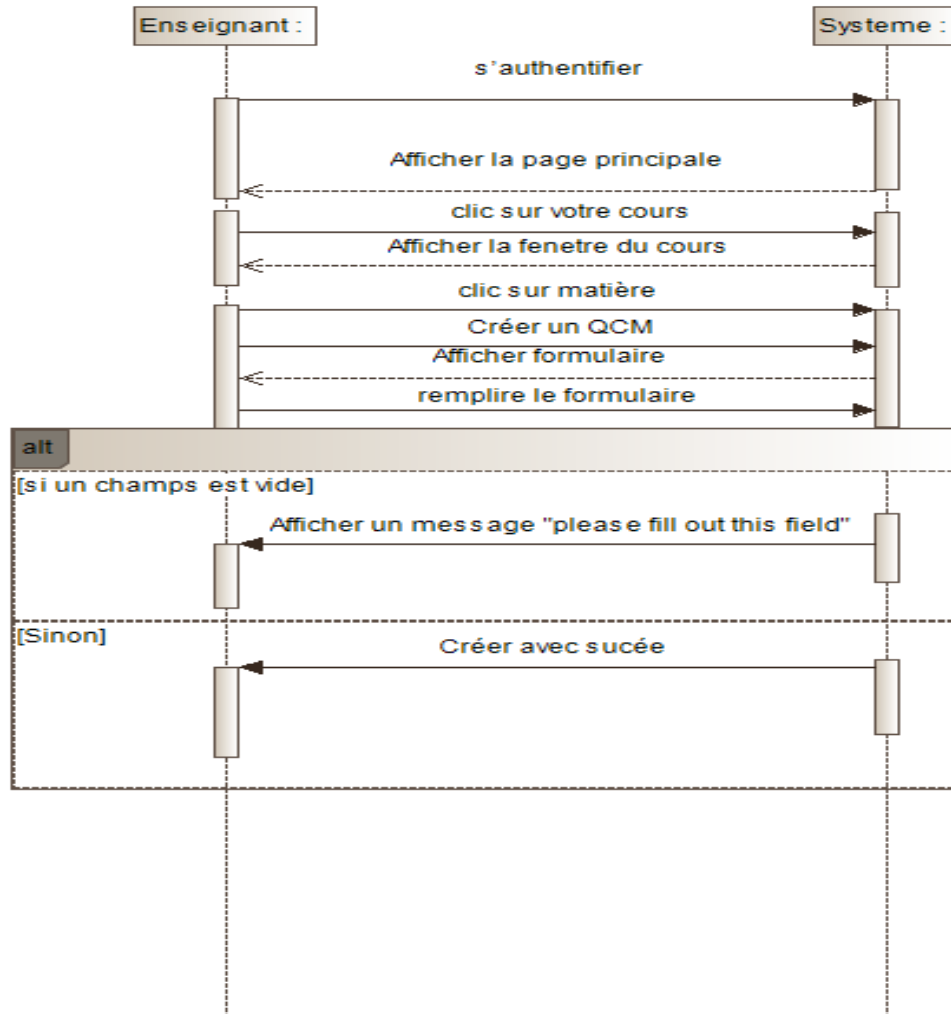


Figure 23 : Diagramme de séquence créer un QCM

- Diagramme de séquence gérer les devoirs soumis :

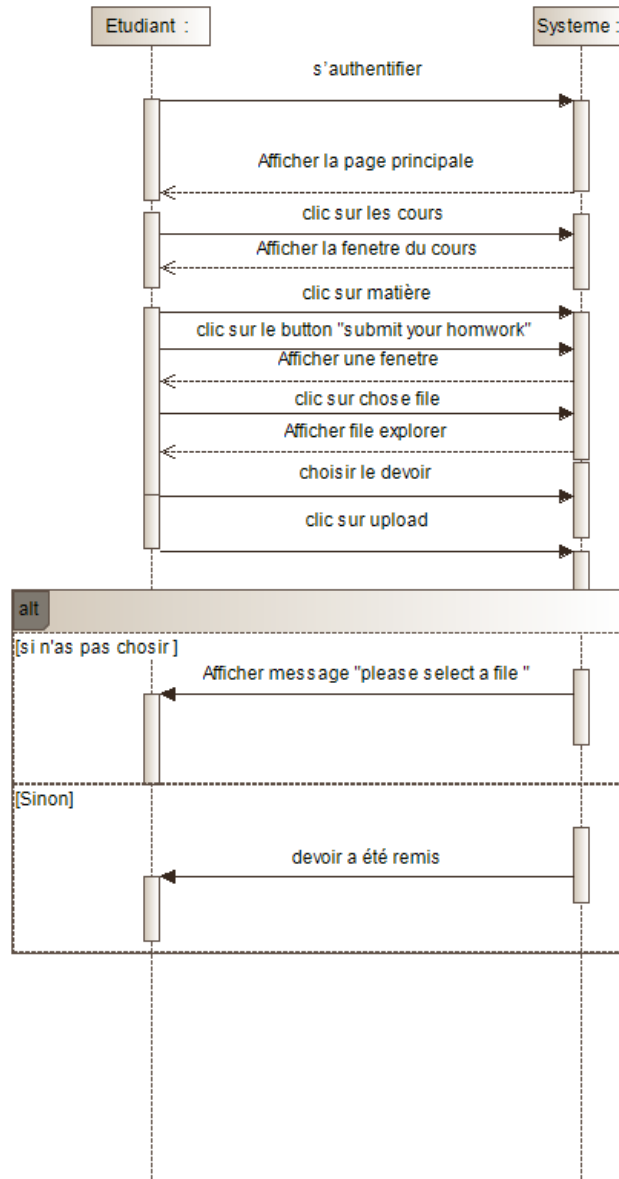


Figure 24 : Diagramme de séquence gérer les devoirs soumis

3.7.3 Modèle de conception

Le modèle de conception est objet, expliquant en détail la mise en œuvre physique des cas d'utilisation. Son rôle est de représenter de manière visionnaire et conceptuelle la mise en œuvre du système, ce qui en fait un élément essentiel pour les opérations d'exécution et de mise en place (exécution). [24]

➤ **Architecture de l'application** : Le choix de l'architecture pour une application web dépend de divers facteurs tels que les besoins précis du projet ainsi que la complexité des fonctionnalités. Voici quelques architectures couramment adoptées dans ce contexte :

- Architecture Monolithique
- Architecture Client-serveur
- Architecture à Micro services
- Architecture Serverless
- Architecture MVC (Modèle-Vue-Contrôleur)

Dans notre application d'enseignement à distance, l'architecture MVC reste une option pertinente. Nous avons adopté une architecture à trois niveaux, comme illustré dans le schéma ci-dessous :

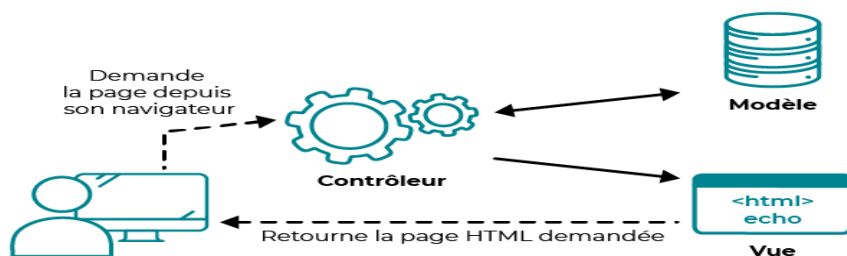


Figure 26 : Architecture MVC

- **Définition d'une architecture MVC (Cas de notre projet application web) :**

L'architecture Modèle-Vue-Contrôleur (MVC) est un schéma de conception logicielle qui partitionne une application en trois modules interdépendants, chacun étant chargé de fonctions distinctes.

1. **Modèle (Model) :**

- Le modèle représente la logique métier de la plateforme, y compris la gestion des cours, des utilisateurs, des quiz, des devoirs, etc.
- Il interagit avec la base de données pour stocker et récupérer les données nécessaires au fonctionnement de la plateforme, telles que les informations sur les utilisateurs et les contenus des cours.

2. **Vue (View) :**

- La vue représente l'interface utilisateur de la plateforme d'enseignement en ligne.
- Elle affiche les différents éléments de l'interface, tels que les pages de cours, les tableaux de bord des étudiants, les formulaires d'inscription, etc.
- Elle communique avec le contrôleur pour récupérer les données nécessaires à l'affichage.

- 3 **Contrôleur (Controller) :**

- Le contrôleur agit comme un intermédiaire entre le modèle et la vue.
- Il reçoit les requêtes des utilisateurs, telles que la demande d'accès à un cours, la soumission d'un devoir, etc.
- Il traite ces requêtes en interagissant avec le modèle pour récupérer ou mettre à jour les données pertinentes, puis sélectionne la vue appropriée pour afficher les résultats.

En outre, avec l'architecture MVC, il est facile d'ajouter de nouvelles fonctionnalités à la plateforme au fur et à mesure que les besoins changent. Cela garantit que l'application reste cohérente et fiable dans son ensemble.

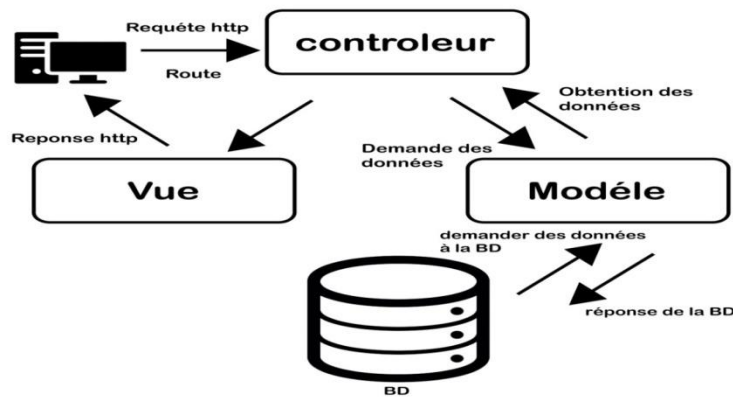


Figure 27 : model MVC web [25]

3.7.4 Modèle d'implémentation

Le modèle d'implémentation explique comment les différents éléments du modèle de conception sont effectués de manière concrète par les différents éléments, tels que les fichiers de code source et les exécutables du système. Cela représente de manière précise et minutieuse comment chaque aspect du modèle de conception est converti en instructions spécifiques dans le code source, ce qui permet de générer les exécutables fonctionnels du système.

➤ Diagramme de composant

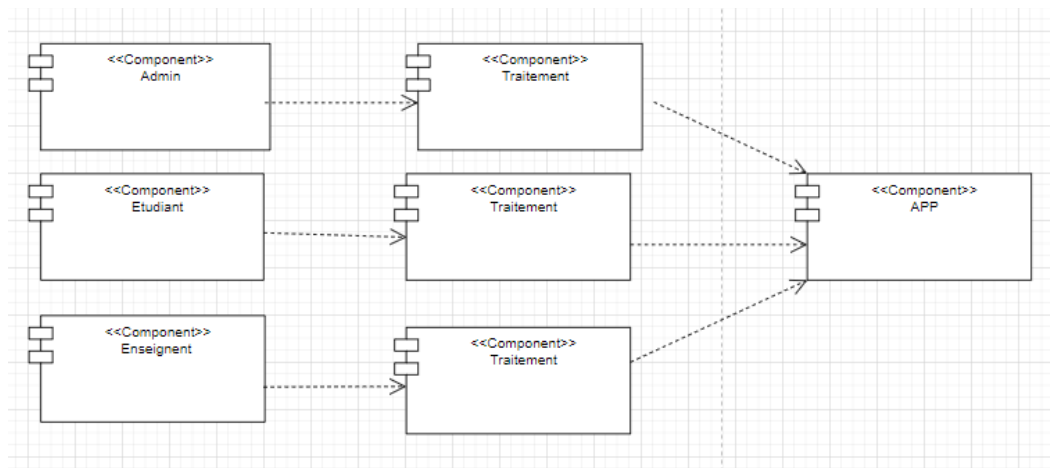


Figure 28 : Diagramme de composant

3.7.5 Modèle de déploiement

Le modèle de déploiement expose la configuration physique des nœuds dans un système distribué, indiquant les artefacts hébergés sur chaque nœud ainsi que les composants et autres éléments qu'ils mettent en œuvre [26].

➤ **Diagramme de déploiement :**

La Figure représente un diagramme de déploiement Web, illustrant la configuration et l'architecture d'un système basé sur le Web.

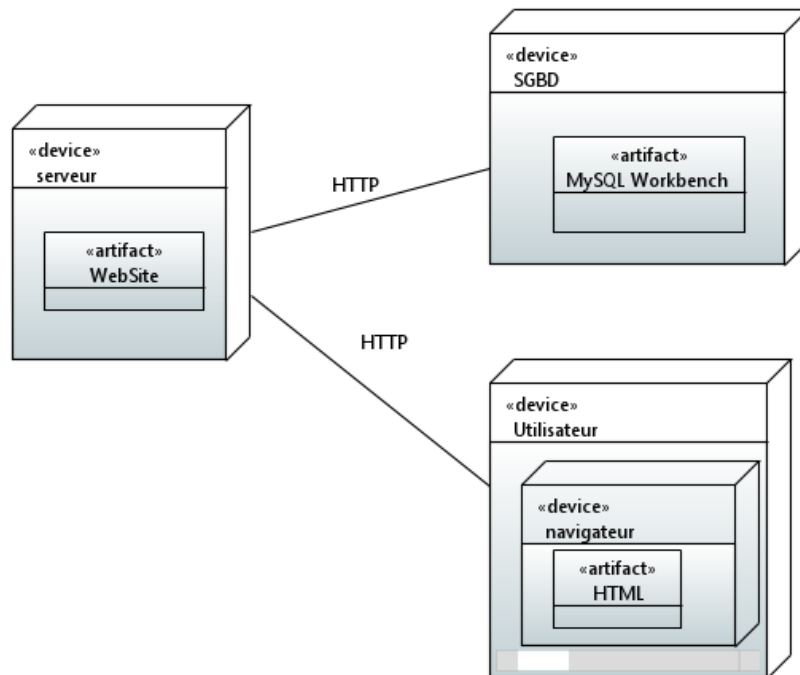


Figure 29 : Diagramme de déploiement

3.7.6 Modèle de test

Le modèle de test explique comment les éléments exécutables du système sont évalués pendant les étapes d'intégration et de construction du système. Il précise aussi les diverses techniques de test employées afin d'évaluer des aspects particuliers du système. [24]

➤ Diagramme de séquence de test d'authentification

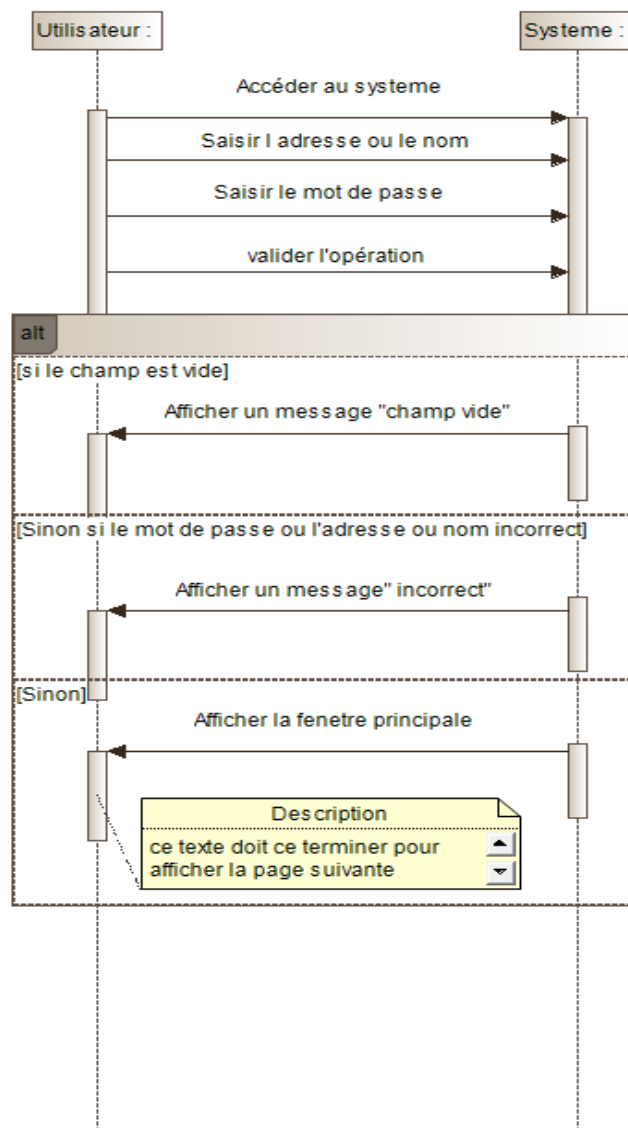


Figure 30 : Diagramme de séquence de test d'authentification

3.8 Conclusion

Dans ce chapitre, nous avons d'abord défini le problème que nous cherchions à résoudre, puis nous avons présenté la solution envisagée. Ainsi nous avons présenté la modélisation de la partie statique et dynamique de notre système en utilisant un sous ensemble des diagrammes d'UML. Cette conception est une étape nécessaire et très importante pour pouvoir créer notre base de données et réaliser l'application. Dans le prochain chapitre, nous allons détailler la mise en œuvre de notre application.

Chapitre 4

Implémentation

4.1 Introduction

La réalisation vient couronner le travail de l'étude préalable et de l'étude conceptuelle. Elle présente la dernière étape et elle est très importante puisque grâce à elle le projet informatique va exister réellement, sa réussite est conditionnée par une multitude de choix essentiellement d'ordre technique, concernant l'exécution de ce qui a été conçu et proposé comme solution afin de répondre aux besoins des utilisateurs et de remédier aux insuffisances perçues.

Dans ce chapitre final, nous abordons la phase de mise en œuvre et de concrétisation de notre projet. Nous commencerons par une présentation des outils que nous avons employés, puis nous passerons à l'exposition des divers scénarios. Ces derniers seront appuyés par des captures d'écran afin de mieux appréhender les multiples fonctionnalités implémentées.

4.2 Les outils des développements

4.2.1 Environnement matérielle

Ordinateur portable : HP PAVILION

Système d'exploitation : WINDOWS 10

Processeur : I5-8300H

Mémoire : 8 GB DDR4

4.2.2 Environnement logiciel

La conception d'un système informatique requiert l'utilisation d'outils logiciels, notamment des langages de programmation et des environnements de développement. Voici les outils de développement que nous avons utilisés :

4.2.2.1 Editeur IntelliJ

IntelliJ IDEA propose l'un des éditeurs de code les plus puissants de l'industrie. IntelliJ IDEA convient parfaitement au développement web. Il prend en charge un large éventail de technologies JavaScript et connexes, telles que TypeScript, React, Vue, Angular, Node.js, HTML et les feuilles de style. Pour tous ces langages, IntelliJ IDEA fournit des fonctionnalités essentielles telles que la saisie semi-automatique du code, des actions contextuelles, la mise en évidence des erreurs, etc. [27]

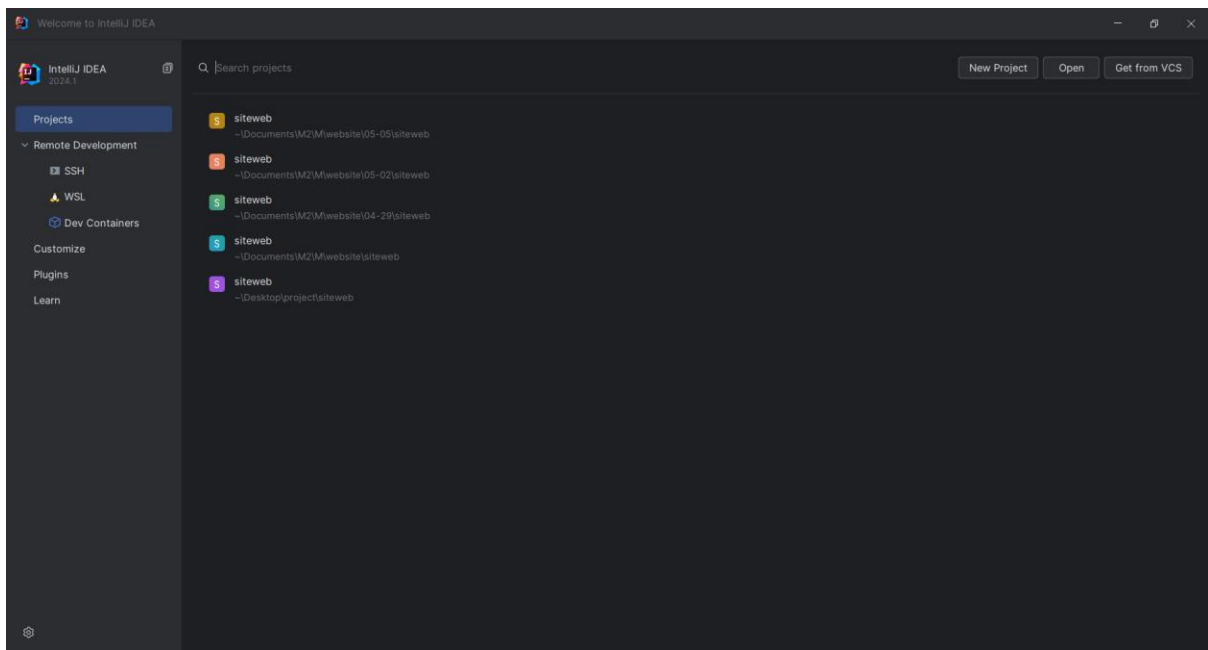


Figure 31: Editeur IntelliJ

4.2.3 Définition des grands choix techniques

Cette phase consiste à choisir les techniques clés qu'on va utiliser dans le cadre de notre projet pour son développement.

Parmi ces techniques, On peut citer :

- **HTML** : Le HyperText Markup Language, généralement abrégé HTML ou, dans sa dernière version, HTML5, est le langage de balisage conçu pour représenter les pages Web [28].
- **CSS** : Les CSS (Cascading Style Sheets en anglais, ou « feuilles de style en cascade») sont le code utilisé pour mettre en forme une page web [29].
- **Javascript** : JavaScript désigne un langage de développement informatique, et plus précisément un langage de script orienté objet. On le retrouve principalement dans les pages Internet. Il permet, entre autres, d'introduire sur une page web ou HTML des petites animations ou des effets [30].

4.3 Les bibliothèques

- **Spring Boot** : Fournit une base pour la création d'applications, offrant une auto-configuration, des serveurs intégrés et une gestion simplifiée des dépendances.
- **Spring Data JPA** : Utilisez JPA pour simplifier l'interaction avec les bases de données MySQL. Cela réduit le code fastidieux pour les opérations de base de données telles que l'interrogation, l'enregistrement et la mise à jour des données.
- **Spring Web** : Vous permet de créer des applications Web et des API RESTful. Il gère les requêtes HTTP, le routage et les réponses, vous permettant de créer des services Web et des contrôleurs.

- **Thymeleaf** : Effectue la traduction et la création de contenu dynamique en HTML, XML, JavaScript, CSS et texte. Il offre la possibilité de concevoir des modèles comprenant des espaces réservés remplis avec les informations de votre application, ce qui permet de générer des pages web dynamiques.
- **Spring Security** : Offre des systèmes d'authentification et d'autorisation afin de garantir la sécurité de l'application. Il assure la gestion des connexions des utilisateurs, des rôles et des permissions, assurant ainsi que seuls les utilisateurs autorisés peuvent accéder à des ressources particulières.

4.4 Implémentation de la base de données

Un système de gestion de base de données est un ensemble de logiciels qui manipulent le contenu des bases de données. Il sert à effectuer les opérations ordinaires telles que rechercher, ajouter ou supprimer des enregistrements, manipuler les index et créer ou copier des bases de données.

Et pour implémenter notre base de données, nous avons utilisé *MySQL Workbench* comme un serveur de bases de données. Est un outil visuel unifié pour les architectes de bases de données, les développeurs et les administrateurs de bases de données. MySQL Workbench fournit des outils de modélisation de données, de développement SQL et d'administration complète pour la configuration du serveur, l'administration des utilisateurs, la sauvegarde et bien plus encore. MySQL Workbench est disponible sur Windows, Linux et Mac OS X. [31].



Figure 32 : Logo MySQL Workbench

4.5 Présentation du système

Nous allons présenter dans cette partie quelques interfaces utilisateurs de notre application.

4.5.1 Interface Login

L'authentification est un certificat de sécurité du système, et à partir de cette dernière chaque utilisateur authentifié peut accéder à notre application.

L'utilisateur doit introduire son login et son mot de passe pour accéder au menu principale comme la montre la figure ci-dessous.

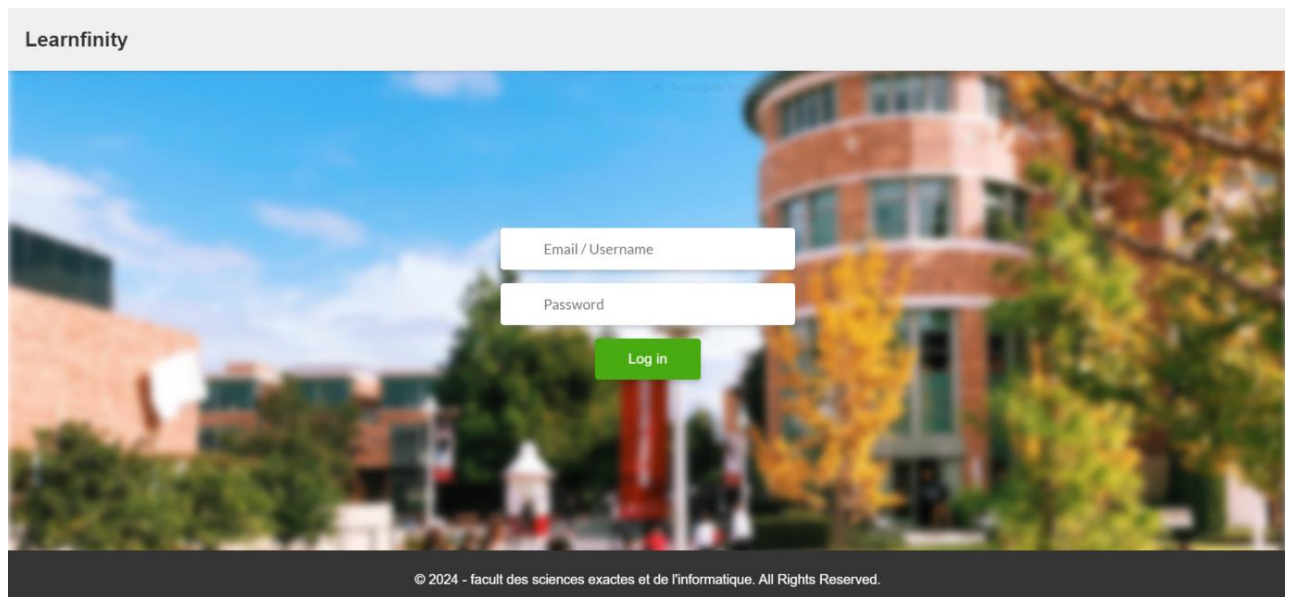
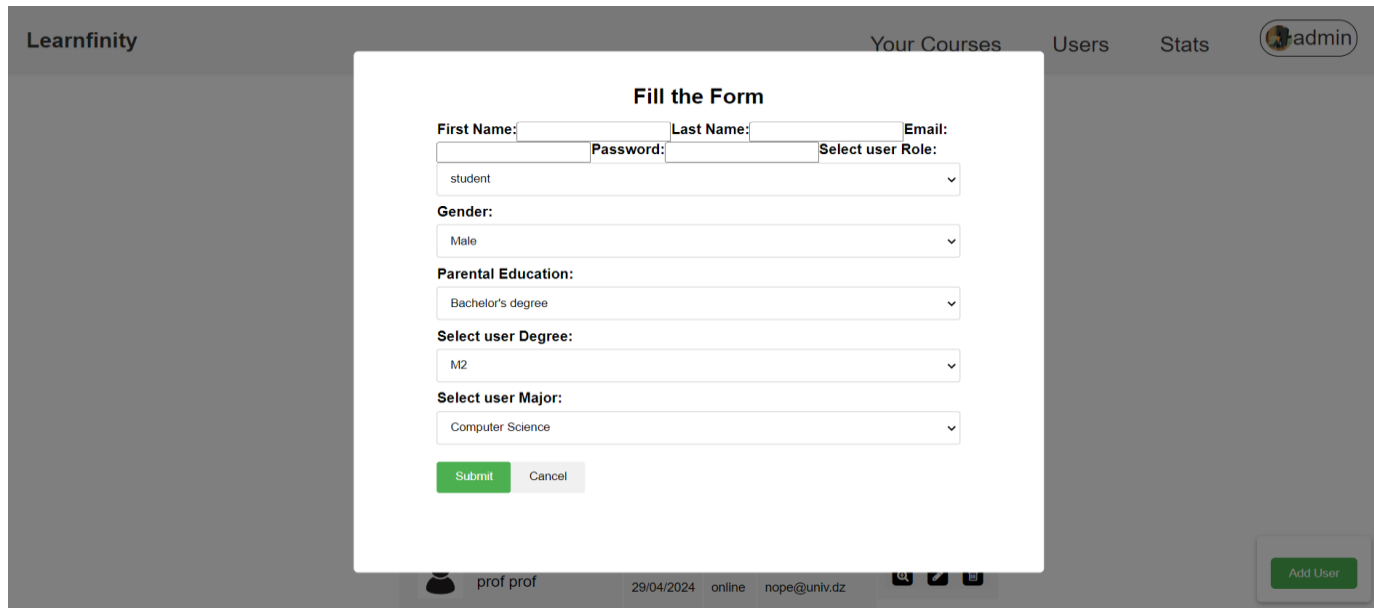


Figure 33 : Page de connexion

4.5.2 Interface d'inscription



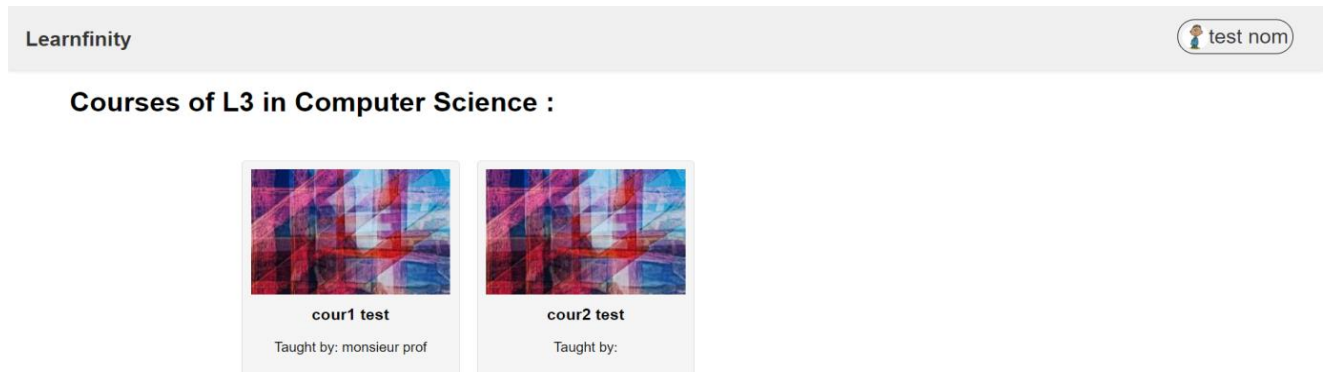
The screenshot shows the 'Fill the Form' registration interface in the Learnfinity system. The form is centered on a dark grey background. At the top, the 'Learnfinity' logo is on the left, and navigation links for 'Your Courses', 'Users', and 'Stats' are on the right, along with an 'admin' user profile. The form itself contains the following fields:

- First Name:** [text input]
- Last Name:** [text input]
- Email:** [text input]
- Password:** [text input]
- Select user Role:** [dropdown menu with 'student' selected]
- Gender:** [dropdown menu with 'Male' selected]
- Parental Education:** [dropdown menu with 'Bachelor's degree' selected]
- Select user Degree:** [dropdown menu with 'M2' selected]
- Select user Major:** [dropdown menu with 'Computer Science' selected]

At the bottom of the form are two buttons: a green 'Submit' button and a grey 'Cancel' button. Below the form, a user profile for 'prof prof' is visible, showing the date '29/04/2024', status 'online', and email 'nope@univ.dz'. In the bottom right corner, there is an 'Add User' button.

Figure 34 : Page d'inscription

4.5.3 Interface des cours



The screenshot displays the 'Courses of L3 in Computer Science' page in the Learnfinity system. The page header includes the 'Learnfinity' logo on the left and a user profile for 'test nom' on the right. The main content area is titled 'Courses of L3 in Computer Science :'. Below this title, two course cards are shown side-by-side:

- cour1 test**: Taught by: monsieur prof
- cour2 test**: Taught by:

Each course card features a square image with a colorful, abstract geometric pattern.

Figure 35: Un exemple de page de cours

4.5.4 Interface des contrôles

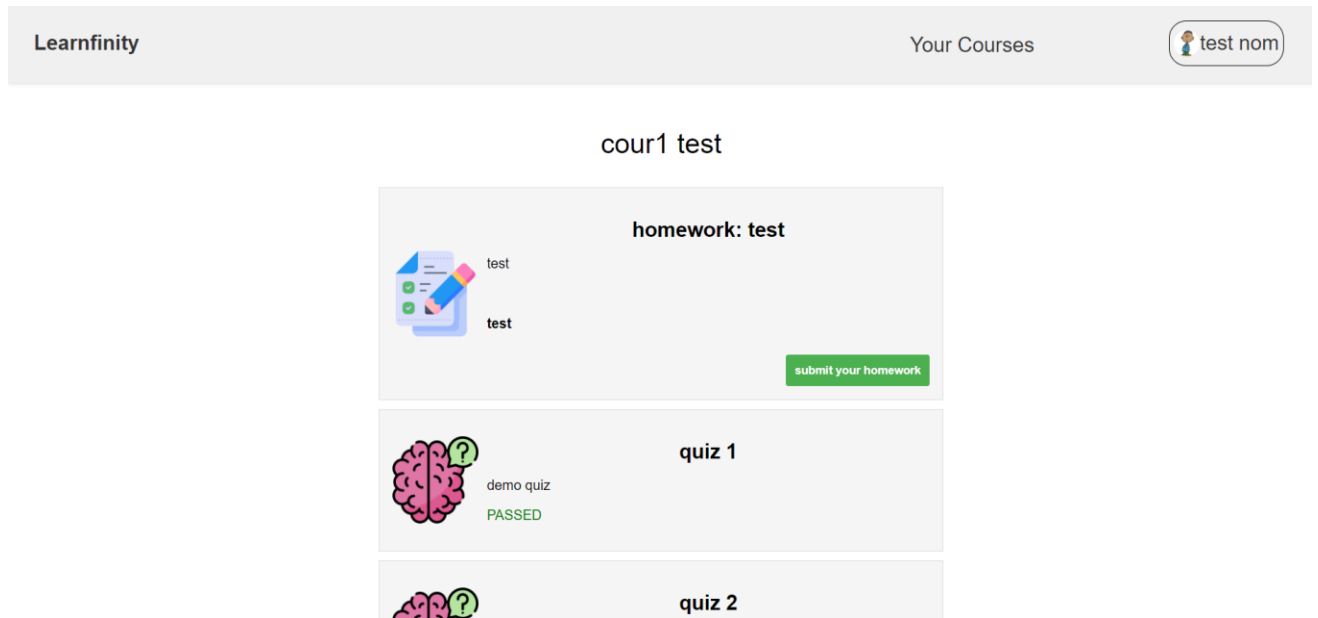


Figure 36 : liste des Quiz et devoir

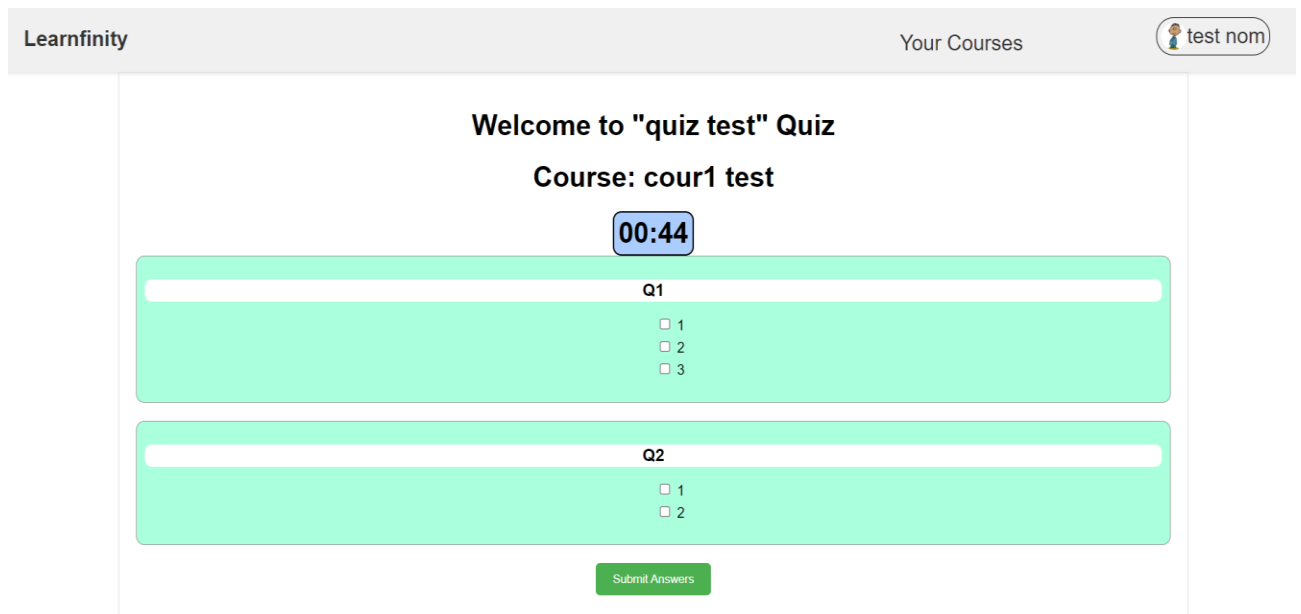


Figure 37 : Un exemple d'un QCM pour le test

4.5.5 Interface des devoirs rendus

Learnfinity Your Courses monsieur prof

cour1 test
mathematics

Homeworks

Student Name	Submitted	type	File
test nom		application/pdf	sample.pdf
test nom	2024-05-13 13:18:43	application/pdf	Untitled_Report.pdf
test nom	2024-05-13 13:18:51	image/png	002.PNG

Figure 38: Les devoirs rendus

4.5.6 Interface des notes affichées

Learnfinity Your Courses Users admin

cour1 test
Quiz: demo quiz
Quiz: test sahel
Quiz: 1
Quiz: quiz test
Quiz: quiz test
cour2 test
cour3
mathematics
egineering
Systemes Virtuels

Grades

Student Name	Course	Quiz	Taken at	Grade
002 student	cours00	10	2024-05-19	78.0%
003 student	cours00	10	2024-05-19	33.0%
Rach test	cours00	10	2024-05-19	33.0%
Ozil Masaoud	cours00	10	2024-05-19	45.0%
student 008	cours00	10	2024-05-19	0.0%
004 student	cours00	10	2024-05-19	0.0%

Figure 39: Affichage des notes

4.5.7 Interface des Statistique



Figure 40 : Interface des Statistique

4.6 Conclusion

Dans ce chapitre, nous avons présenté comment l'application a été mise en place et réalisée. Nous aborderons l'environnement de développement, les outils, l'environnement de développement et les langages de programmation utilisés afin d'implémenter ce projet.

Conclusion Générale

Ce mémoire est concentré sur l'application du data mining éducatif dans un système de gestion d'apprentissage des étudiants durant un cursus universitaire. Il a exploré comment l'analyse des données générées par les plateformes d'apprentissage peut permettre d'identifier les facteurs de réussite et d'échec des étudiants, et ainsi améliorer l'efficacité des processus pédagogiques.

L'étude a mis en évidence l'importance de comprendre le comportement des étudiants, leurs points forts et leurs points faibles, afin de proposer des interventions pédagogiques plus individualisées et plus efficaces. Le développement d'un système web basé sur l'algorithme K-means, a permis de segmenter les étudiants en clusters homogènes, facilitant ainsi l'identification des besoins spécifiques de chaque groupe et l'adaptation des méthodes d'enseignement.

En conclusion, ce mémoire a mis en lumière le potentiel du data mining éducatif pour révolutionner l'enseignement supérieur. Les plateformes d'apprentissage sont capables d'analyser les données des étudiants pour identifier leurs besoins spécifiques et proposer des solutions personnalisées. Grâce à l'analyse des données, les cours pourraient être optimisés pour une meilleure compréhension et les étudiants qui rencontrent des difficultés pourraient recevoir un soutien ciblé. De plus, l'apprentissage pourrait être personnalisé en fonction des forces et des faiblesses de chaque élève, créant un parcours plus motivant et efficace.

En fin, on souhaite que ce rapport soit un bon guide pour d'autres promotions.

Bibliographies

- [1] f. pinto, “database marketing intelligence methodology supported by ontologies and knowlegde discovery in databases,” university of minho, 2009.
- [2] Les différentes étapes de l’ECD. https://www.researchgate.net/figure/Le-processus-dECD-Les-differentes-etapes-du-processus-dECD-sont-les-suivantes_fig3_272291931
- [3] Brijesh Kumar Baradwaj, Saurabh Pal: *Mining Educational Data to Analyze Students Performance*, 2011.
- [4] Thèse de doctorat, Présentée par BENDJEBAR Safia (2016), Utilisation des Techniques de Data Mining pour la Modélisation des Tuteurs. <https://dspace.univ-guelma.dz/jspui/bitstream/123456789/84/1/these-Bendjebar.PDF>
- [5] Abdulmohsen Algarni: Data Mining in Education, 2016. https://scholar.google.com/scholar?q=Abdulmohsen+Algarni:+Data+Mining+in+Education,+2016.&hl=fr&as_sdt=0&as_vis=1&oi=scholart.
- [6] Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics: An Issue Brief, U.S. Department of Education Office of Educational Technology. <https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf>
- [7] SOLAR, What is LEARNING ANALYTICS. <https://www.solaresearch.org/about/what-is-learning-analytics/>

- [8] Elias, T. (2011). Learning analytics: Definitions, processes and potential (Report). Retrieved from.
- [9] Educational Data Mining. <https://www.javatpoint.com/educational-data-mining>
- [10] Clow, Doug .The learning analytics cycle: closing the loop effectively. In: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12, p. 134. <https://oro.open.ac.uk/34330/1/LAK12-DougClow-personalcopy.pdf>
- [11] Marie Bienkowski, Mingyu Feng, Barbara Means: *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief*, October 2012.
- [12] (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 6, 2016 [Data_Mining_in_Education.pdf](#)
- [13] Thèse de doctorat, Utilisation des Techniques de Data Mining pour la Modélisation des Tuteurs.<https://dspace.univ-guelma.dz/jspui/bitstream/123456789/84/1/these-Bendjebar.PDF>
- [14] R. S. d Baker, A. T. Corbett, and V. Aleven, “More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing,” in *Intelligent Tutoring Systems*, pp. 406– 415, Springer, 2008.
- [15] A. T. Corbett and J. R. Anderson, “Knowledge tracing: Modeling the acquisition of procedural knowledge,” *User modeling and user-adapted interaction*, vol. 4, no. 4, pp. 253–278, 1994
- [16] <https://dspace.univ-guelma.dz/jspui/bitstream/123456789/84/1/these-Bendjebar.PDF>

- [17] Baepler, Paul and Murdoch, Cynthia James (2010) "Academic Analytics and Data Mining in Higher Education," *International Journal for the Scholarship of Teaching and Learning*: Vol. 4: No. 2, Article 17.
<https://digitalcommons.georgiasouthern.edu/ij-sotl/vol4/iss2/17/>
- [18] <https://js.ugd.edu.mk/index.php/YFCS/article/view/903/959> 24 consulter le 24/04/2024
- [19] Data mining in course management systems moodle case study and tutorial.
<https://www.sciencedirect.com/science/article/abs/pii/S0360131507000590>
- [20] Educational Data Mining and Learning Analytics for 21st century higher education: A Review and Synthesis
- [21] les techniques de ECD. <https://www.questionpro.com/blog/fr/techniques-dextraction-de-donnees/>
- [22] Dataset . <https://www.kaggle.com/code/zabihullah18/student-performance>
- [23] Support de Cours Fouille de Données Auteur Mohamed HABIB ZAHMANI année 2022
- [24] la méthode RUP. http://thieum22.free.fr/Quest_RUP
- [25] Modèle vue contrôleur ou MVC <https://fr.wikipedia.org/wiki/Mod%C3%A8le-vue-contr%C3%B4leur>
- [26] Définition de diagramme déploiement, Dernière mise à jour: 2021-03-05.
<https://www.ibm.com/docs/fr/rsas/7.5.0?topic=topologies-deployment-diagrams>
- [27] Définition de Editeur intelligent IDEA. <https://www.jetbrains.com/fr-fr/idea/features/>

[28] Définition du HTML https://fr.wikipedia.org/wiki/Hypertext_Markup_Language

[29] <https://dane.ac-lyon.fr/spip/IMG/scenari/CSSERUN/co/theorie.html> consulte le
10/05/2024

[30] Définition du java script. <https://fr.wikipedia.org/wiki/JavaScript>

[31] Définition du workbench mysql. <https://www.mysql.com/products/workbench/>