

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE 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 :

**Chemirik Islam**

**Faraoun Abdelhafid**

THEME :

**Conception et développement de scripts Web permettant  
l'analyse des résultats d'apprentissage de toute une  
promo durant un semestre**

Soutenu le : 04/07/2022

Devant le jury composé de :

Hocine Nadia	MCA	Université de Mostaganem	Président
Midoun Mohammed	MCB	Université de Mostaganem	Examineur
Henni Karim	MAA	Université de Mostaganem	Encadreur

Année Universitaire 2021-2022

## **Résumé**

Le projet s'articule autour de la mise au point de scripts Web permettant l'analyse des résultats d'apprentissage semestriels, en d'autres termes la réalisation d'une application Web d'enseignement à distance en incluant des techniques de data mining qui permettent le suivi, l'analyse et l'évaluation du semestre afin d'améliorer les résultats d'apprentissage.

Dans cette étude nous présentons un état de l'art sur les expériences, techniques et les réalisations faites sur le data mining éducatif et l'analyse de l'apprentissage ainsi que les différents étapes de conception et de réalisation.

## **Mots-clés:**

Enseignement à distance, Data mining éducatif, Analyse de l'apprentissage

## **Abstract**

The project revolves around the development of web scripts allowing the analysis of semester learning outcomes, in other words the realization of an optimized remote learning Web application by including data mining techniques that allow semester tracking, analysis, and evaluation in order to improve learning outcomes.

In this study we present a state of the art on the experiences, techniques and achievements made on educational data mining and learning analysis as well as the different stages of design and implementation.

## **Keywords:**

E-learning, Educational data mining, Learning analytics

## Liste des figures

Figure N°	Titre de la figure	Page
Figure 1	Interface utilisateur de l'outil Weka	12
Figure 2	Les étapes de data mining éducatif	13
Figure 3	Exemple d'application des étapes de data mining éducatif sur les données des étudiants	16
Figure 4	Application de la méthode Elbow	27
Figure 5	Les informations sur les étudiants	29
Figure 6	Les informations sur le groupe 1	29
Figure 7	Les informations sur le groupe 2	30
Figure 8	Les statistiques relatives aux résultats finaux des deux groupes	30
Figure 9	Nombre de clics sur le module avec le code FFF	31
Figure 10	Diagramme de cas d'utilisation de l'étudiant	35
Figure 11	Diagramme de cas d'utilisation de l'enseignant	36
Figure 12	Diagramme de cas d'utilisation de l'administrateur	37
Figure 13	Diagramme de séquence de l'inscription	38
Figure 14	Diagramme de séquence de l'authentification	39
Figure 15	Diagramme de séquence de publication	40
Figure 16	Diagramme de séquence de gestion de publication	41
Figure 17	Diagramme de séquence de soumission de devoir	42
Figure 18	Diagramme de classes	43
Figure 19	Les versions du serveur local WampServer	46
Figure 20	Page d'accueil	48
Figure 21	Page de cours (1)	49

Figure 22	Page de cours (2)	50
Figure 23	Page d'inscription	51
Figure 24	Page de connexion	52
Figure 25	Page d'accueil du tableau de bord de l'enseignant	53
Figure 26	Page d'accueil du tableau de bord de l'administrateur	54
Figure 27	Page des modules du tableau de bord de l'enseignant	55
Figure 28	Page des publications du tableau de bord de l'enseignant	55
Figure 29	Page des devoirs du tableau de bord de l'enseignant	56
Figure 30	Page des devoirs rendus du tableau de bord de l'enseignant	56
Figure 31	Page des notes du tableau de bord de l'enseignant	57
Figure 32	Page de publication	58
Figure 33	Email de notification des nouvelles	59
Figure 34	Page des statistiques de module du tableau de bord	60
Figure 35	Page de regroupement d'étudiants	61
Figure 36	Page de statistiques de semestre du tableau de bord de l'administrateur	62
Figure 37	Page de promos du tableau de bord de l'administrateur	63

## Liste des tableaux

Tableau N°	Titre du tableau	Page
Tableau 1	Description des données des étudiants	28

## Liste des abréviations

Abréviation	Expression Complète	Page
GPA	Grade Point Average	10
VLE	Virtual Learning Environment	14, 27, 31
CMS	Content Management System	23
WSS	Within-groups Sum of Squares	27
UML	Unified Modeling Language	45, 64
IDE	Integrated Development Environment	46
HTML	Hypertext Markup Language	47
CSS	Cascading Style Sheets	47
PHP	Hyper Text Preprocessor	47
SQL	Structured Query Language	47
AJAX	Asynchronous JavaScript And XML	47

# Table des matières

Introduction Générale .....	4
Chapitre 1 Data mining éducatif.....	6
1.1 Introduction .....	6
1.2 Définitions.....	6
1.2.1 Data mining.....	6
1.2.2 Data mining éducatif.....	7
1.2.3 Analyse de l'apprentissage .....	7
1.3 Data mining éducatif .....	8
1.3.1 Les méthodes de data mining éducatif.....	8
1.3.2 Les logiciels et les outils utilisés dans le data mining éducatif: .....	11
1.3.3 Les étapes de data mining éducatif .....	13
1.3.4 Les principaux objectifs de data mining éducatif .....	16
1.4 Conclusion.....	17
Chapitre 2 Etat de l'art.....	18
2.1 Introduction .....	18
2.2 Les applications de data mining éducatif et d'analyse de l'apprentissage.....	18
2.2.1 Modélisation des connaissances de l'utilisateur.....	18
2.2.2 Modélisation du comportement et de l'expérience utilisateur .....	19
2.2.3 Profilage des utilisateurs .....	20
2.2.4 Modélisation de domaine.....	21
2.2.5 Analyse des composantes d'apprentissage et analyse des principes d'enseignement.....	21
2.2.6 Analyse de tendance .....	22
2.2.7 Adaptation et personnalisation.....	22
2.3 Data mining dans l'enseignement supérieur .....	23
2.4 Défis Techniques.....	23

2.5	Approche d'analyse.....	24
2.5.1	K-means .....	24
2.5.2	La méthode du coude .....	26
2.5.3	Etapas.....	26
2.6	Conclusion.....	32
<b>Chapitre 3 Analyse et conception .....</b>		<b>33</b>
3.1	Introduction .....	33
3.2	Identification des besoins .....	33
3.3	Analyse des besoins .....	33
3.3.1	Acteurs .....	33
3.3.2	Cas d'utilisation .....	34
3.3.2.1	Diagramme de cas d'utilisation de l'étudiant.....	34
3.3.2.2	Diagramme de cas d'utilisation de l'enseignant .....	35
3.3.2.3	Diagramme de cas d'utilisation de l'administrateur .....	36
3.3.3	Diagrammes de séquence.....	37
3.3.4	Classes.....	43
3.4	Conclusion.....	45
<b>Chapitre 4 Implémentation .....</b>		<b>46</b>
4.1	Introduction .....	46
4.2	Outils utilisés.....	46
4.2.1	Serveur Local .....	46
4.2.2	IDE.....	46
4.2.3	Langages de structure, style et de programmation.....	47
4.2.4	Frameworks et bibliothèques .....	47
4.2.5	Autres .....	47
4.3	Interfaces .....	47
4.3.1	Page d'accueil.....	47
4.3.2	Page de cours .....	48
4.3.3	Page d'inscription .....	51



4.3.4	Page de connexion .....	51
4.3.5	Page d'accueil du tableau de bord .....	52
4.3.6	Page de modules .....	54
4.3.7	Pages de publications et devoirs .....	55
4.3.8	Page de notes.....	56
4.3.9	Page de publication.....	57
4.3.10	Email de notification de publication.....	58
4.3.11	Page de statistiques de module .....	59
4.3.12	Page de statistiques de semestre .....	61
4.3.13	Page de promotions.....	62
4.4	Conclusion.....	63
Conclusion Générale.....		64
Bibliographie.....		65

# Introduction Générale

Au fil des ans, la technologie a révolutionné notre monde et notre vie quotidienne, elle a créé des outils et des ressources étonnantes, mettant des informations utiles à portée de main.

Comme nous le savons, l'éducation est la plus importante dans notre vie et grâce à la technologie de l'information et de la communication, le système éducatif est réinventé, les établissements scolaires dans la plupart des pays et surtout après la pandémie de COVID-19 utilisent des plateformes d'enseignement à distance qui facilitent la communication entre les différents acteurs du système éducatif et qui offrent un bouquet de services numériques en relation avec leurs activités.

Afin de mieux utiliser ces plateformes, la science des données est très utile, elle permet d'améliorer l'ensemble du processus d'éducation, elle rend la gestion et le suivi de l'éducation plus efficace tant pour les enseignants que pour le personnel administratif, elle aide également à comprendre et à résoudre les problèmes des étudiants.

La problématique est traitée selon nous-mêmes en tant qu'étudiants et en tant qu'acteurs du système éducatif, nous constatons qu'il y a plusieurs problèmes à résoudre aussi bien pour nous que pour les enseignants et l'administration, la mauvaise gestion du processus d'enseignement est le problème initial. Les enseignants doivent connaître par exemple les raisons des mauvaises notes, ce qui attire les étudiants et ce qui les éloigne, est-ce qu'ils gèrent bien le cours ou non, les types de ressources à éviter, les meilleurs horaires de travail, etc. De plus, il faut aussi faciliter la tâche pour le personnel administratif pour mieux gérer ses tâches, et mieux suivre et évaluer les différents acteurs et les résultats tout au long du semestre.

L'objectif de cette étude est la conception et le développement de scripts Web permettant l'analyse des résultats d'apprentissage semestriels pour répondre à tous ces problèmes.

Dans la première partie, nous introduisons les concepts, méthodes et objectifs de data mining éducatif et d'analyse de l'apprentissage.

Puis dans la deuxième partie, nous présenterons un état de l'art sur les expériences, et les réalisations des chercheurs sur ces domaines, et ce que nous prévoyons de faire.

La troisième partie présentera ensuite la conception détaillée de notre application, y compris les différentes phases de développement ainsi que l'approche d'analyse.

La quatrième et la dernière partie 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

En raison de l'émergence et la voluminosité des données, il est devenu plus difficile d'analyser les données générées par les environnements éducatifs et d'explorer les facteurs importants qui influent sur la performance des étudiants.

Le data mining joue un rôle très important en éducation, ça permet d'améliorer le processus d'étude, l'amélioration de l'achèvement des cours, l'aide aux étudiants dans le choix des cours, le profilage des étudiants, la recherche de problèmes menant au décrochage, l'élaboration de programmes, la prédiction des performances des étudiants et aussi il est utilisé comme support pour la prise de décision lors de l'inscription des étudiants.

Dans ce chapitre nous présenterons les différentes techniques et caractéristiques de data mining éducatif et d'analyse de l'apprentissage ainsi que leurs différents processus.

### 1.2 Définitions

#### 1.2.1 Data mining

Le data mining (fouille de données) aussi connu sous le nom de découverte de connaissances dans la base de données, est un outil puissant d'intelligence artificielle qui permet de découvrir des informations utiles à partir de grandes quantités de données en analysant les données sous de nombreux angles ou dimensions, catégoriser ces informations et résumer les relations identifiées dans la base de données. [8] [14]

### **1.2.2 Data mining éducatif**

Le data mining éducatif est une discipline émergente, soucieuse de développer des méthodes pour explorer les données uniques et de plus en plus à grande échelle qui proviennent des milieux éducatifs et de les utiliser pour mieux comprendre les étudiants et les milieux dans lesquels ils apprennent. [14]

En d'autres termes, Il s'agit d'un domaine de recherche multidisciplinaire qui examinent l'intelligence artificielle, la modélisation statistique et le data mining avec les données générées par un établissement d'enseignement.

### **1.2.3 Analyse de l'apprentissage**

L'analyse de l'apprentissage désigne la collecte et l'analyse de données sur les apprenants et leur environnement afin de comprendre et d'améliorer les résultats de l'apprentissage.

L'analyse de l'apprentissage se définit de plus en plus comme un domaine de recherche et d'application. Il s'appuie sur un éventail plus vaste de disciplines universitaires que de data mining éducatif, intégrant des concepts et des techniques de la science et de la sociologie de l'information, en plus de l'informatique, de la statistique, de la psychologie et des sciences de l'apprentissage. Contrairement au data mining éducatif, l'analyse de l'apprentissage ne met généralement pas l'accent sur la réduction de l'apprentissage en composantes, mais elle cherche plutôt à comprendre l'ensemble des systèmes et à appuyer la prise de décisions humaines. En grande partie grâce aux améliorations et à la disponibilité accrue des technologies éducatives, l'analyse de l'apprentissage a le potentiel d'améliorer le processus d'enseignement et d'apprentissage. [5] [12]

## **1.3 Data mining éducatif**

### **1.3.1 Les méthodes de data mining éducatif**

De nombreux modèles et expériences ont été développés et testés sur des données éducatives afin d'avoir des connaissances sur la qualité de l'apprentissage et ses paramètres associés.

Les modèles font l'objet de nouvelles recherches qui ont été faites pour augmenter l'efficacité de la prédiction des paramètres cibles en aidant les étudiants à acquérir des connaissances facilement, et pour aider les professeurs à comprendre les progrès des étudiants.

Les données éducatives doivent être correctement prétraitées avant d'être alimentées en entrée (input) du modèle car le modèle est d'abord entraîné pour être utilisé. En général, les données sont divisées en deux ensembles, ensemble d'apprentissage (training set) et ensemble de test (test set), le training set est utilisé pour former le modèle et les données de test sont ensuite utilisées pour vérifier l'exactitude du modèle résultant.

Pour extraire des connaissances des données éducatives, de diverses approches de data mining sont utilisées, notamment des algorithmes de prédiction, des algorithmes de la distillation des données pour le jugement humain et des algorithmes d'exploration de règles d'association, des algorithmes de regroupement et les algorithmes de la découverte avec des modèles. [5] Dans notre cas d'étude, les acteurs principaux sur lesquels nous allons mener notre recherche sont les étudiants et l'enseignant.

#### **a) Prédiction**

L'objectif majeur visé par les méthodes de prédiction est de prédire les performances des étudiants (la réussite ou l'échec).

Le développement d'un modèle de prédiction dépend de la connaissance de la variable prédite pour un petit ensemble de données. Ensuite, un modèle est créé pour ce petit ensemble pour le valider statistiquement afin qu'il puisse être appliqué à plus grande échelle.

Les nombreuses études publiées dans ce domaine montrent qu'il n'existe ni un classificateur ni un ensemble de fonctionnalités qui fonctionnent bien dans tous les contextes, par conséquent, il faut rechercher quelles méthodes et quelles fonctionnalités fonctionnent mieux avec les données disponibles.

Il existe trois types généraux de prédictions :

- Classification qui est la technique de data mining la plus couramment appliquée, dans cette méthode on utilise les connaissances préalables pour construire un modèle d'apprentissage, puis on utilise ce modèle comme variable binaire ou catégorique pour les nouvelles données. De nombreux modèles ont été développés et utilisés comme classificateurs tels que les machines vectorielles de régression logistique et de machines à vecteurs de support. Il existe plusieurs techniques pour la classification comme arbre de décision, naïve bayes, etc.
- Régression est un modèle utilisé pour prédire des variables différents de la classification, les modèles de régression prédisent des variables continues.
- Estimation de la densité est basée sur diverses fonctions du noyau, y compris les fonctions gaussiennes. [14]

## **b) Regroupement**

Dans cette méthode le but est de séparer les données en différents groupes en fonction de certaines caractéristiques communes.

Le regroupement est une méthode très importante qui aide à clarifier les facteurs qui influent sur la réussite des étudiants. De nombreuses recherches ont été appliquées sur le regroupement, ils ont obtenu des résultats positifs, notamment: Morais et al ont recherché des techniques pour soutenir le processus décisionnel des enseignants en regroupant les étudiants et planifier les défis en conséquence et ils sont parvenus à une conclusion positive. Par exemple les utilisateurs, évènements, sessions, pages, activités peuvent être des objets dans l'apprentissage à distant.

Parmi les méthodes les plus utilisées dans le regroupement K-means, K-medoids, Mean-shift Clustering, et DBSCAN. [7] [14]

### **c) Exploration relationnelle**

C'est une technique qui vise à trouver des relations entre différentes variables dans des ensembles de données comportant un grand nombre de variables. Cette technique contient plusieurs types tels que :

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

Comme il est mentionné par Abdulmohsen Algarni, les méthodes d'exploration de relations ont été les plus utilisées dans la recherche en data mining éducatif au cours des dernières années. Ces techniques peuvent être utilisées pour associer l'activité des étudiants, dans un système de gestion de l'apprentissage, par exemple pour trouver les erreurs des étudiants qui se produisent simultanément par exemple si le GPA de l'étudiant est inférieur à deux et l'étudiant a un emploi, donc cet étudiant va abandonner l'école. [14]

### **d) Distillation des données pour le jugement humain**

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

Pour que les données soient utiles aux éducateurs, elles doivent être présentées de manière à permettre au spectateur de comprendre facilement ce qu'il voit. Cette technique comprend des statistiques et des visualisations qui aident à donner un sens à leurs découvertes et analyses.

[5]



### **e) Découverte avec des modèles**

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

Dans cette technique, les résultats d'une analyse de data mining sont utilisés dans une autre analyse de data mining, le plus souvent, un modèle d'une certaine construction est obtenu, généralement par des méthodes de prédiction. Ce modèle est ensuite appliqué aux données afin d'identifier où le construit apparaît dans les données. Les prédictions du modèle sont ensuite utilisées comme données d'entrée pour une autre méthode de data mining, par exemple, la sortie d'un modèle de prédiction peut être utilisée dans un autre modèle de prédiction ou dans une analyse d'exploration de relations. [14]

## **1.3.2 Les logiciels et les outils utilisés dans le data mining éducatif:**

### **a) Outil Weka**

L'outil le plus utilisé dans le data mining éducatif est Weka, il est open-source, il prend en charge plusieurs tâches de data mining standard telles que la classification, le regroupement, la régression, etc., il aide également à visualiser les données de différentes manières, par exemple afficher les données sous forme d'histogramme, de tracés graphiques et de diagrammes. Cet outil spécifique a été utilisé dans plusieurs études pour prédire les performances des étudiants.

Figure 1 représente l'interface utilisateur de l'utilisation de Weka sur les données des étudiants. [15]

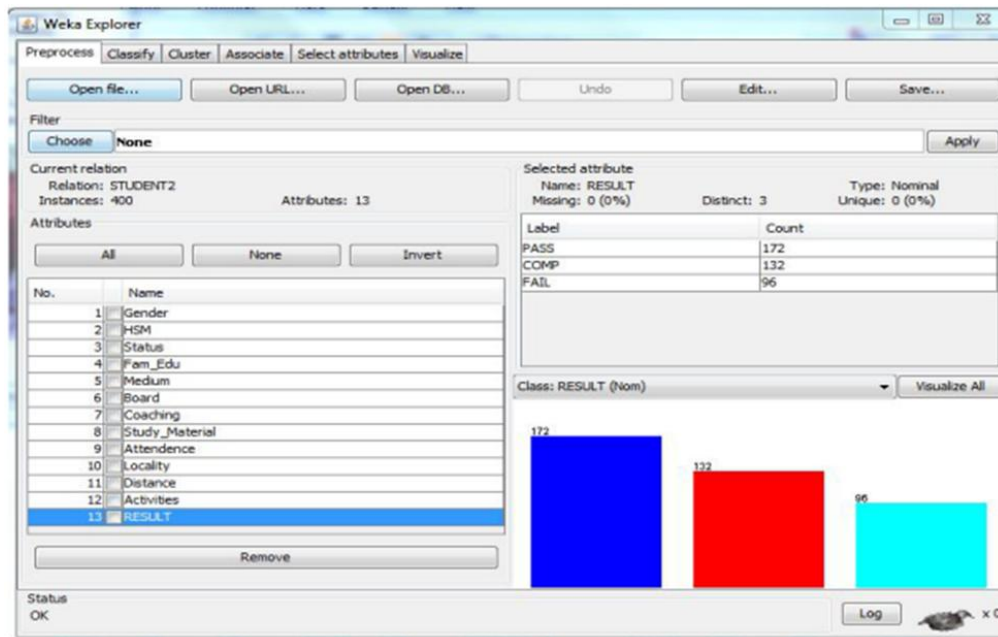


Figure 1 – Interface utilisateur de l’outil Weka

## b) PowerSight Kit

C’est un outil qui est utilisé pour accumuler et analyser des données sur le comportement des étudiants tels que les utilisateurs, leurs droits, les activités des étudiants, leurs notes, etc., cet outil prend une partie des informations de la base de données ORACLE en ligne. Cet outil a été utilisé par Saulius Preidys et Leonidas Sakalauskas en 2009. [3]

## c) SPEET

L’outil SPEET est un projet européen financé dans le cadre du programme ERASMUS+ en tant que Partenariat Stratégique pour l’enseignement supérieur. Il a été développé par plusieurs universités de différents pays pour être convivial et facilement accessible. L’objectif principal de SPEET est d’appliquer des algorithmes de data mining pour traiter cet ensemble massif de profils d’étudiants afin d’extraire des informations sur et d’identifier les caractéristiques communes à chacun de ces profils d’étudiants.

SPEET a été développé pour être facilement accessible, afin que d'autres facultés en dehors du consortium du projet puissent facilement effectuer une analyse préliminaire de leurs étudiants sur la base de leurs propres données après s'être organisées en conséquence. [4] [18]

#### **d) Autres outils**

Il y'a d'autres outils qui aident dans le traitement de données ainsi pour appliquer les méthodes de data mining dans l'éducation comme BlackBoard 2008, Rapid Miner, Pentaho Business Intelligence Suite, XL miner, Borland C++ builder6 tool, Hadoop File Distribution System (HDFS), Model-It software tool et Moodle serve version. [1]

### **1.3.3 Les étapes de data mining éducatif**

Figure 2 représente les différents étapes des méthodes d'application de data mining dans l'enseignement à distance. [3]

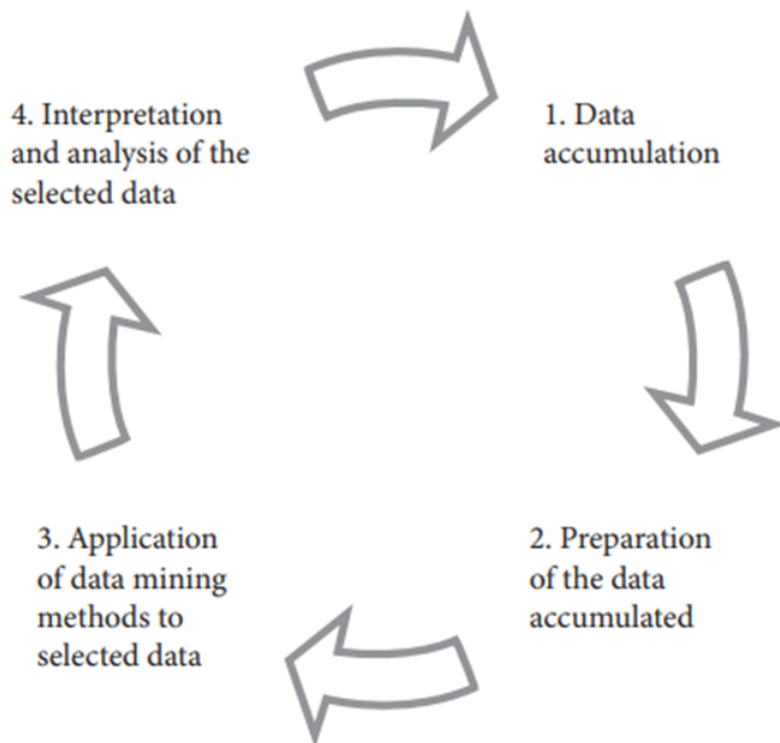


Figure 2 – Les étapes de data mining éducatif

Ces étapes sont les principales étapes pour construire un modèle efficace afin d'obtenir des résultats précis. Dans notre cas, nous voulons appliquer toutes ces étapes sur les données de toute une promo ainsi que les données des enseignants. [9]

### **a) Accumulation de données**

Cette étape nécessite une grande variété de données, il est nécessaire de collecter toutes les données disponibles sur tous les étudiants et enseignants afin d'obtenir des modèles plus précis avec des résultats efficaces qui aident le progrès du niveau des étudiants. [10]

### **b) Préparation des données sélectionnées**

Les données sélectionnées dans l'étape a (accumulation de données) sont traitées pour garder que les données qui sont utiles afin que les méthodes de data mining puissent être appliquées ultérieurement.

Dans cette étape plusieurs outils de VLE ont été utilisés, comme PowerSight Kit, elle élimine les valeurs redondantes et stocke les valeurs manquantes, effectue le nettoyage, l'extraction et la sélection des fonctionnalités, etc. [3]

### **c) Application de méthodes de data mining aux données sélectionnées**

Les résultats du modèle construit dans cette étape permet de connaître les facteurs qui ont conduit à l'échec des étudiants de toute une promo et donc, de nombreuses méthodes de data mining ont été appliquées tels que les algorithmes de prédiction, de regroupement, de classification, etc. afin d'obtenir des résultats efficaces. [11]

Le modèle obtenu dans l'étude de Witten et al, 2017 généralise la relation entre les entrées et les sorties et l'utilise pour prédire d'autres ensembles de données où seules les entrées sont connues.

Mashaal Al luhaybi, Allan Tucker et Leila Yousefi de l'Université Brunel de Londres, au Royaume-Uni, ont mené une étude pour prédire les étudiants à haut risque d'échouer au module en fonction de leurs notes finales ou globales et d'autres aspects à l'aide de l'outil

Weka. Ils ont appliqué les algorithmes de classification C4.5 et Naïve Bayes pour analyser les performances académiques des étudiants de niveau 2 en fonction de leur admission, des données relatives aux cours et des notes finales de niveau 1. Ils ont conclu que les qualifications des étudiants à l'entrée ont l'impact le plus élevé sur cette prédiction. [17]

Une autre étude a été appliquée pour découvrir les relations entre les facteurs personnels et sociaux des étudiants à l'aide de RapidMiner et Weka. Cette étude a été menée par un groupe d'étudiants inscrits dans différents collèges de l'université des sciences et technologies d'Ajman, Émirats Arabes Unis. Plusieurs techniques et algorithmes d'arbre de décision ont été passés en revue, et leurs performances et précisions ont été testées et validées. En dernière analyse, il a évidemment été remarqué que certains algorithmes fonctionnaient mieux avec l'ensemble de données que d'autres, en détail, CART avait la meilleure précision de 40 %, ce qui était nettement supérieur à la précision attendue (modèle par défaut), CHAID et C4.5 était le suivant avec 34,07 % et 35,19 % respectivement, et le moins précis était ID3 avec 33,33 %. Ils ont également appliqué la classification Naïve Bayes qui donne une précision de 36,40 %. Ils ont trouvé comme résultat de leur étude que les performances de l'étudiant ne dépendent pas totalement de ses efforts académiques, il existe également de nombreux autres facteurs qui ont des influences plus importantes que prévu. [16]

#### **d) Analyse et interprétation des résultats obtenus**

C'est la dernière étape dans laquelle les résultats obtenus sont analysés et interprétés.

Figure 3 représente un exemple d'application des méthodes de data mining éducatif sur les données des étudiants. [4]

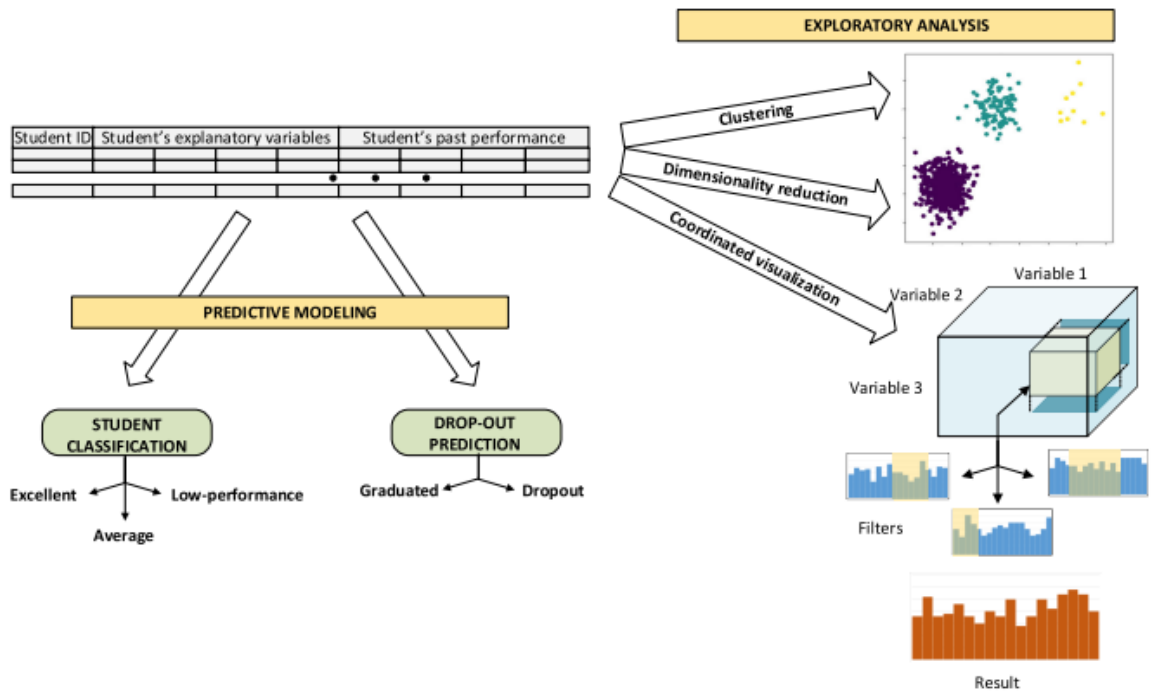


Figure 3 – Exemple d'application des étapes de data mining éducatif

### 1.3.4 Les principaux objectifs de data mining éducatif

Parmi les principaux buts que les chercheurs envisagent dans ce domaine : [5]

- Prédire le comportement d'apprentissage futur.
- Découvrir ou améliorer des modèles de domaine.
- Étudier les effets du soutien scolaire.
- Faire progresser les connaissances scientifiques sur l'apprentissage et les apprenants.

## **1.4 Conclusion**

L'utilisation de data mining dans l'éducation est très utile pour analyser l'apprentissage, résoudre les problèmes et améliorer les résultats. Plusieurs algorithmes tels que la prédiction, la classification, et le regroupement peuvent être utilisés.

Dans ce chapitre, nous avons présenté les généralités, méthodes, et objectifs de data mining éducatif et d'analyse de l'apprentissage.

# Chapitre 2

## Etat de l'art

### 2.1 Introduction

Les chercheurs cherchent toujours à améliorer le data mining éducatif et d'analyse de l'apprentissage afin d'obtenir des résultats plus précis et plus efficaces. Ils ont expérimenté de nouveaux types de données de système d'apprentissage et de nouvelles techniques de construction de modèles qui se sont révélées prometteuses. De nombreux outils ont été développés et plusieurs domaines d'application ont été trouvés dans la pratique.

Dans ce chapitre nous présenterons un état de l'art sur les expériences et les réalisations faites sur le data mining éducatif et l'analyse de l'apprentissage.

### 2.2 Les applications de data mining éducatif et d'analyse de l'apprentissage

Ces domaines représentent les catégories dans lesquelles l'exploration et l'analyse de données peuvent être appliquées à l'activité en ligne, en particulier en ce qui concerne l'apprentissage en ligne, chacun de ces domaines d'application utilise des sources de données différentes :

#### 2.2.1 Modélisation des connaissances de l'utilisateur

Quel contenu un étudiant connaît-il ? Les chercheurs et les développeurs construisent et ajustent des modèles d'utilisateurs qui représentent une collection de données spécifiques à



l'utilisateur, en particulier les compétences et les connaissances. La connaissance peut être déduite d'interactions telles que l'exactitude des réponses des élèves, le temps pris avant d'essayer de répondre à une question, les répétitions de mauvaises réponses et les erreurs commises. Ces déductions peuvent être faites par un modèle informatique prédictif ou par un enseignant qui examine les données des étudiants sur un tableau de bord. Une méthode populaire pour estimer les connaissances des élèves est le modèle de traçage des connaissances de Corbett et Anderson (1994), une approche qui utilise un modèle basé sur un réseau bayésien pour estimer la probabilité qu'un élève connaisse une compétence en se basant sur ses observations. Plus récemment, Baker, Corbett et Aleven (2008) ont proposé une nouvelle méthode de traçage des connaissances en utilisant une approche d'apprentissage automatique pour faire des estimations contextuelles de la probabilité, cette méthode a augmenté la précision des prédictions. [5]

### **2.2.2 Modélisation du comportement et de l'expérience utilisateur**

Que signifient les modèles de comportement des élèves pour leur apprentissage ? Et est-ce que les étudiants sont motivés et satisfaits de leur expérience ou non ? La modélisation du comportement de l'utilisateur dans l'éducation caractérise souvent les actions des étudiants. Il s'appuie sur les mêmes types de données d'apprentissage utilisées pour prédire les connaissances des utilisateurs, ainsi que sur d'autres mesures, telles que le temps qu'un étudiant a passé en ligne, si un étudiant a terminé un cours, documenté les changements dans la classe ou l'école le contexte, l'assiduité, les retards et parfois le niveau de connaissances d'un élève tel qu'il est déduit de son travail avec le système d'apprentissage ou d'autres sources de données telles que les résultats de tests standardisés. Les systèmes d'apprentissage en ligne enregistrent les données des élèves qui peuvent être extraites pour détecter les comportements des élèves en corrélation avec l'apprentissage. Macfayden et Dawson (2010) ont analysé les données de suivi du système de gestion de l'apprentissage d'un cours pris en charge par Blackboard Vista et ont trouvé des variables en corrélation avec la note finale des étudiants.

La modélisation de l'expérience utilisateur peut être jugée par les réponses des étudiants aux enquêtes ou questionnaires de suivi et par leurs choix, comportements et performances. Par rapport aux applications commerciales de modélisation de l'expérience utilisateur, moins de travail a été fait dans l'éducation pour utiliser l'analyse pour améliorer l'expérience d'apprentissage des étudiants et favoriser leur réussite et leur taux de rétention.

Dawson, Heathcote et Poole (2010) ont examiné l'efficacité des établissements d'enseignement supérieur à exploiter les mécanismes de saisie de données de leurs systèmes d'information sur les étudiants, leurs systèmes de gestion de l'apprentissage et leurs outils de communication pour améliorer les expériences d'apprentissage des étudiants. Nous évoquons également le groupe d'analyse de Kaplan, il a collecté le temps passé sur les composants de cours repensés, des enquêtes périodiques sur l'état de motivation des étudiants pendant le cours et les performances d'apprentissage. Le groupe a ensuite examiné la satisfaction des enseignants, la satisfaction des étudiants et les performances des évaluations d'apprentissage intégrées. [5]

### **2.2.3 Profilage des utilisateurs**

Le profil d'utilisateur est un ensemble de données personnelles décrivant les caractéristiques essentielles d'un utilisateur. Il fait référence au processus de construction et d'application de profils d'étudiants ou de groupes à l'aide d'algorithmes de data mining et d'apprentissage automatique. Étant donné que les étudiants diffèrent dans leurs préférences, leurs intérêts, leurs antécédents et même leurs objectifs d'apprentissage, l'objectif à long terme du profilage des utilisateurs est souvent de fournir des environnements d'apprentissage adaptés et personnalisés pour des individus ou des groupes d'étudiants afin de maximiser l'efficacité de l'apprentissage.

Dans l'éducation, les techniques de data mining, telles que la classification et le regroupement, sont souvent utilisées pour profiler (catégoriser) les étudiants en fonction des types de données d'apprentissage personnelles décrites dans la section sur la base de recherche, sur les données démographiques des étudiants, ou les deux. Kardan et Conati

(2011) ont proposé un framework de modélisation des utilisateurs qui s'appuie sur les journaux d'interaction pour identifier différents types d'étudiants, ainsi que leurs interactions caractéristiques avec le système d'apprentissage. Ces informations seraient ensuite utilisées pour classer les nouveaux élèves, dans le but à long terme de fournir un soutien d'interaction adaptatif lorsque des comportements nuisibles à l'apprentissage sont détectés. La classification peut également regrouper les étudiants en groupes d'étude ou d'autres activités d'apprentissage conjointes. [5]

#### **2.2.4 Modélisation de domaine**

Quel est le niveau correct auquel diviser les sujets en modules et comment ces modules doivent-ils être séquencés ? Un modèle de domaine est souvent créé pour représenter les concepts clés qui composent un sujet ou un domaine thématique. Cette modélisation a été adoptée comme approche pour affiner les systèmes d'apprentissage afin de mieux servir l'apprentissage et l'instruction. Par exemple, Martin et al. (2011) ont décrit trois études pour démontrer comment les courbes d'apprentissage peuvent être utilisées pour entraîner des changements dans le modèle d'utilisateur pour les environnements d'apprentissage personnalisés. [5]

#### **2.2.5 Analyse des composantes d'apprentissage et analyse des principes d'enseignement**

Quels composants sont efficaces pour promouvoir l'apprentissage ? Quels principes d'apprentissage fonctionnent bien ? Et quelle est l'efficacité des programmes d'études entiers ?

L'analyse des principes pédagogiques examine les composants d'un système d'apprentissage et les types de pratiques pédagogiques adoptées à différents moments ou pour différents groupes d'élèves à aborder. Pour répondre à ces questions, il faut collecter des données telles que les données d'entrée des élèves et l'exactitude des réponses.

Beck et Mostow (2008) ont proposé la décomposition de l'apprentissage comme alternative aux méthodes traditionnelles de test A/B (une méthodologie traditionnelle de recherche sur l'expérience utilisateur). En tant que type de mining de relations, la décomposition de l'apprentissage consiste à ajuster les courbes d'apprentissage exponentielles aux données de performance et à relier la réussite des étudiants à la quantité de chaque type de soutien pédagogique qu'un étudiant a reçu. [5]

### **2.2.6 Analyse de tendance**

Qu'est-ce qui change avec le temps et comment ? L'analyse de tendance en général fait référence à la collection des informations et de tenter de créer un modèle séquentiel ou une tendance dans ces informations au fil du temps. Par exemple, les sites Web de plusieurs entreprises utilisent cette pratique pour prédire ce que les utilisateurs pourraient rechercher ou être intéressé par, et comment leur participation augmente et diminue. Dans l'éducation, l'analyse des tendances aide à répondre à des questions telles que les changements qui se sont produits dans l'apprentissage des étudiants au fil du temps et comment l'apprentissage a changé afin d'aider les administrateurs à déterminer l'impact des politiques utilisées. Quand on parle du data mining éducatif l'analyse des tendances fait souvent référence à des techniques d'extraction d'un modèle sous-jacent, qui peut être partiellement ou presque complètement caché par des données qui ne contribuent pas au modèle. [5]

### **2.2.7 Adaptation et personnalisation**

Ce domaine d'application concerne la manière dont les analyses sont utilisées pour s'adapter ou personnaliser l'expérience de l'utilisateur. L'adaptation est utilisée pour indiquer les changements qu'un système, ou un enseignant fait aux étudiants pour personnaliser leur expérience, la personnalisation, telle que définie dans le U.S. Department of Education (2010), indique un style d'enseignement en fonction des préférences d'apprentissage et l'adaptation du contenu aux intérêts des étudiants. Les adaptations peuvent inclure des recommandations ou des commentaires (feedback) aux étudiants sur leurs prochaines actions et des changements dans leur expérience avec un système d'apprentissage en ligne tels qu'un contenu différent, plus de pratique sur leurs progrès dans un cours.

(Köck et Paramythis 2011) ont étudié le suivi et l'interprétation d'activités d'apprentissage pour améliorer l'adaptation et personnaliser les environnements éducatifs. Ils ont analysé les données de résolution de problèmes des étudiants à partir d'un système de tutorat de physique en convertissant les séquences d'activités dans les données brutes en modèles de type chaîne, puis en regroupant les séquences pour détecter les styles de résolution de problèmes, ces modèles sont utilisés pour adapter le système d'apprentissage aux méthodes préférées des étudiants. [5]

## **2.3 Data mining dans l'enseignement supérieur**

Les techniques de data mining sont vastes, les plus utilisées dans l'enseignement supérieur se situent dans les catégories du regroupement, la classification, la visualisation et l'analyse d'association.

Des chercheurs de l'université de Floride l'ont expérimenté en utilisant les journaux de données Moodle CMS (logs) pour collecter les données des étudiants (par exemple le nombre de clics pour chaque étudiant), bien qu'il y ait plusieurs limitations à cette étude.

Les CMS open source comme Moodle présentent l'une des meilleures chances d'analyser des cours individuels avec des techniques d'extraction de données, des chercheurs de l'université de Cordoba Romero, Ventura et Garcia (2008) ont développé des outils expérimentaux d'extraction de données qui peuvent être intégrés directement à Moodle, ces outils forment un framework open source nommé KEEL, il permet aux concepteurs de cours d'effectuer facilement une série d'analyses sur leurs cours. [2]

## **2.4 Défis Techniques**

Un défi pour une mise en œuvre réussie des techniques de data mining éducatif et d'analyse de l'apprentissage est de disposer de ressources techniques suffisantes pour utiliser le Big Data et d'engager les dépenses associées aux services logiciels et au stockage.

En réponse à ce défi du Big Data, certains problèmes doivent être pris en compte pour chaque cas lors de la mise en œuvre de l'exploration et de l'analyse de données. Il s'agit notamment de choisir les données à collecter, de se concentrer sur les questions auxquelles il faut répondre et de s'assurer que les données correspondent aux questions.

Le manque d'interopérabilité des données entre les différents systèmes de données impose un défi à l'exploration et à l'analyse des données qui reposent sur des données diverses et distribuées.

Au fil du temps, les achats de logiciels au coup par coup peuvent entraîner une décentralisation importante de la source de données sur l'éducation, comme les systèmes d'information sur les élèves et les enseignants. [5]

## **2.5 Approche d'analyse**

L'analyse des résultats d'apprentissage devient plus difficile en raison du grand volume de données dans les bases de données éducatives. L'utilisation du data mining est indispensable, il permet d'extraire des informations très utiles.

Pour découvrir les données qui ont un effet sur les résultats d'apprentissage, plusieurs méthodes de data mining telles que la classification, le regroupement, l'exploration relationnelle peuvent être appliquées.

Dans cette section, nous appliquerons deux approches de data mining à l'ensemble de données des étudiants, K-means et la méthode du coude (Elbow).

### **2.5.1 K-means**

Le regroupement K-means est une méthode d'apprentissage automatique non supervisé et est utilisé efficacement pour partitionner un ensemble de données donné en k groupes ou k clusters, où k représente le nombre de groupes ou de clusters. Le nombre de groupes est calculé par l'analyste pour dériver un regroupement efficace des données. La

méthode de regroupement K-means sépare les données en clusters de telle sorte que les valeurs de données au sein du même cluster soient similaires les unes aux autres. [22]

### 2.5.1.1 Algorithme

Entrée: Data set  $D = \{x_1, x_2, \dots, x_n\}$ ; nombre de groupe  $k$ ,  $WSS \leftarrow []$  ;

Sortie: Clusters  $C = \{C_1, C_2, \dots, C_k\}$

Début:

1 : Initialiser  $k$  centres de cluster avec  $k$  échantillons choisis au hasard dans  $D$

2 : répéter:

#calculer la distance euclidienne entre chaque point de cluster

Pour  $x \in D$

  Pour  $j$  de  $\{0 \dots k\}$

    pour  $l$  de  $\{0 \dots k\}$

      Si  $\sqrt{(c_j - x)^2} < \sqrt{(c_l - x)^2}$

$M_j = x$

      Fin si

  fin pour

#Calculer le nouveaux centre

$$c_j = 1/N \sum_{x \in M_j} x$$

  Fin pour

Fin pour

Jusqu'à convergence de centres

$$WSS \leftarrow \sum_{i=0}^n (c_i - x)^2$$

Fin

Avec  $x = \{\text{le nombre de clics dans le module, la note finale de chaque étudiant dans chaque module}\}$

$c_j$ : Ce sont les centres de chaque groupe. [25]

### 2.5.1.2 Pourquoi K-means

Nous avons choisi la méthode K-means car elle est facile à comprendre et à implémenter, elle permet d'extraire les caractéristiques commun entre les étudiants qui

mènent à la réussite ou à l'échec. Dans notre cas nous allons regrouper les étudiants en fonction de nombre de clics sur les modules, le nombre de clics sur les ressources de chaque module, et la note finale obtenu dans chaque module.

## 2.5.2 La méthode du coude

La méthode du coude est décrite dans l'équation (1), où la distance moyenne au carré de tous les points de données pour un cluster est une distance mesurée statistiquement entre les moyennes du groupe et le même centre de cluster. [23]

$$wss = \sum_{i=0}^n (x_i - c_i)^2 \quad (1)$$

Où  $c_i$  est le centre et  $x_i$  représente les informations de chaque étudiant.

La combinaison des méthodes K-means et Elbow peut localiser la valeur de  $k$  (nombre de groupes) au niveau du cluster optimal pour déterminer ce nombre.

Pour obtenir le  $k$  optimale il faut calculer l'équation (1) pour de différents nombres, après on trace le graphe, lorsque la courbe commence à s'aplatir et nous voyons un point d'inflexion, on prend la valeur qui correspond a ce point.

## 2.5.3 Etapes

### 2.5.3.1 Déterminer le nombre K par Elbow

Après avoir tracé le graphique comme montré sur la figure 4, nous avons obtenu 2 comme nombre de groupes, l'axe horizontal représente le nombre de groupe, et l'axe vertical représente la valeur du WSS. Dans notre cas nous avons appliqué le WSS pour  $k=1$  jusqu'à  $k=10$ , car même si le WSS continue de diminuer, il se diminue avec un rythme faible.



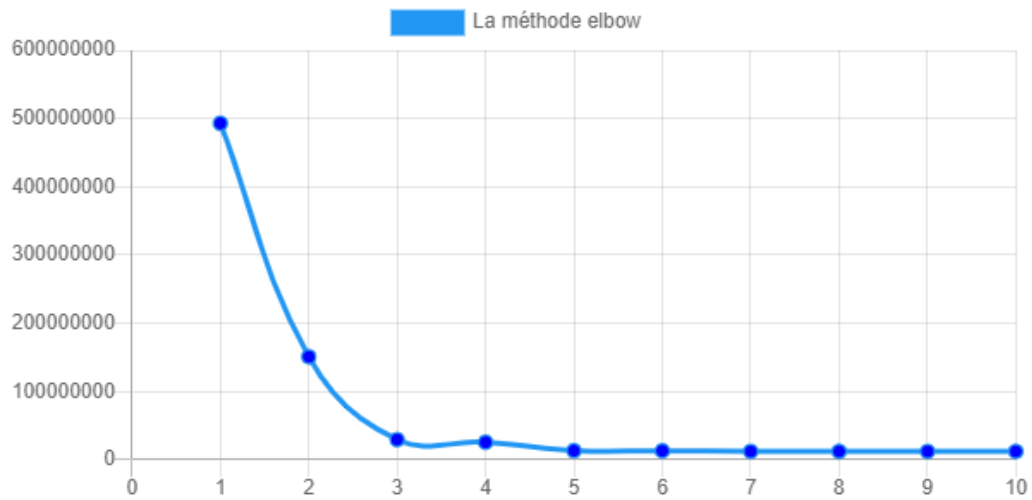


Figure 4 – L’application de la méthode Elbow

### 2.5.3.2 Collection de données

La base de données contient des informations sur les étudiants, à savoir : leurs données démographiques, les modules et les activités sur VLE, depuis 2012. Nous avons sélectionné plusieurs modules représentatifs enseignés au cours des années 2013 et 2014. [21] L'ensemble de données contient des informations sur 22 cours, 32 593 étudiants, leurs résultats d'évaluation et les journaux de leurs interactions avec les clics représentés par VLE. Les modules peuvent être présentés plusieurs fois au cours de l'année. Pour distinguer les différentes présentations d'un module, chaque présentation est nommée par l'année et le mois de début. Par exemple, les présentations commençant en janvier se terminent par A, en février par B et ainsi de suite, de sorte que « 2013J » signifie que la présentation a commencé en octobre 2013. [21]

**Tableau 1 – Description des données des étudiants**

<b>Données</b>	<b>Description</b>
Id étudiant	L'identificateur de chaque étudiant
Age	l'intervalle d'âge pour chaque étudiant
Sexe	Masculin ou Féminin
Région	la région géographique où l'étudiant a vécu pendant la présentation du module
Résultat finale	Le résultat final de l'étudiant dans la présentation du module.
Code module	Nom du code du module.
Code présentation	Nom du code de la présentation
Longueur	la durée de la présentation du module en jours, depuis la date de début du module jusqu'à sa fin
Total clics	Le nombre total de clics d'étudiant dans un module
Id évaluation	Numéro d'identification de l'évaluation
Score	Le score de l'étudiant dans cette évaluation, La plage va de 0 jusqu'à 100

id	id évaluation	code module	code présentation	Longeur	id étudiant	Sexe	Région	Age	total clics	Résultat final	score
0	1752	AAA	2013J	268	11391	M	East Anglian Region	55<=	934	Pass	78
1	1752	AAA	2013J	268	28400	F	Scotland	35-55	1435	Pass	70
2	1752	AAA	2013J	268	31604	F	South East Region	35-55	2158	Pass	72
3	1758	AAA	2014J	269	6516	M	Scotland	55<=	2791	Pass	60
4	1758	AAA	2014J	269	24734	F	South Region	0-35	499	Pass	41
5	1758	AAA	2014J	269	26192	F	East Anglian Region	35-55	2223	Distinction	87
6	1758	AAA	2014J	269	28061	M	Yorkshire Region	35-55	1590	Pass	78
7	1758	AAA	2014J	269	31600	F	South West Region	0-35	429	Pass	47

Figure 5 – Les informations sur les étudiants

### 2.5.3.3 Prétraitement

Dans cette phase nous avons éliminé les lignes redondantes et les valeurs manquantes, puis nous avons calculé le nombre de total de clics par étudiant et par module.

### 2.5.3.4 Application de K-means

Après le nettoyage de la base de données nous avons appliqué l’algorithme de K-means pour regrouper les données similaires.

id	id évaluation	code module	code présentation	Longeur	id étudiant	Sexe	Région	Age	total clics	Résultat final	score
0	14984	BBB	2013B	240	31663	M	North Region	35-55	5906	Pass	88
1	15020	BBB	2014J	262	32327	F	South West Region	0-35	3648	Withdrawn	1
2	15020	BBB	2014J	262	32930	F	South Region	35-55	6951	Pass	1
3	24282	DDD	2013B	240	45664	M	Yorkshire Region	0-35	7496	Pass	52
4	24291	EEE	2014J	269	60809	M	South East Region	0-35	7452	Pass	100
5	25334	DDD	2013B	240	41060	M	Ireland	0-35	8421	Fail	85
6	25348	DDD	2014J	262	33681	M	South East Region	35-55	5694	Fail	94
7	25355	DDD	2014B	241	31269	F	East Midlands Region	35-55	7568	Withdrawn	45
8	34860	FFF	2013B	240	33915	F	South East Region	0-35	7596	Distinction	86
9	34873	FFF	2013J	268	33600	F	London Region	35-55	7594	Withdrawn	84
10	34886	FFF	2014B	241	27189	F	South Region	35-55	6248	Fail	52
11	34899	FFF	2014J	269	31296	M	East Midlands Region	0-35	7569	Fail	78
12	34899	FFF	2014J	269	40857	F	North Western Region	0-35	7894	Withdrawn	86
13	34899	FFF	2014J	269	40878	M	North Western Region	0-35	4821	Distinction	94
14	37415	GGG	2013J	261	32239	F	London Region	0-35	3652	Pass	55
15	37415	GGG	2013J	261	35001	F	East Midlands Region	35-55	3578	Fail	80

Figure 6 – Les informations sur le groupe 1

id	id évaluation	code module	code présentation	Longueur	id étudiant	Sexe	Région	Age	total clics	Résultat final	score
0	1752	AAA	2013J	268	11391	M	East Anglian Region	55<=	934	Pass	78
1	1752	AAA	2013J	268	28400	F	Scotland	35-55	1435	Pass	70
2	1752	AAA	2013J	268	31604	F	South East Region	35-55	2158	Pass	72
3	1758	AAA	2014J	269	6516	M	Scotland	55<=	2791	Pass	60
4	1758	AAA	2014J	269	24734	F	South Region	0-35	499	Pass	41
5	1758	AAA	2014J	269	26192	F	East Anglian Region	35-55	2223	Distinction	87
6	1758	AAA	2014J	269	28061	M	Yorkshire Region	35-55	1590	Pass	78
7	1758	AAA	2014J	269	31600	F	South West Region	0-35	429	Pass	47
8	14984	BBB	2013B	240	23629	F	East Anglian Region	0-35	161	Fail	67
9	14984	BBB	2013B	240	25107	F	East Anglian Region	0-35	2562	Pass	59
10	14984	BBB	2013B	240	27891	M	Scotland	0-35	602	Withdrawn	73
11	14984	BBB	2013B	240	29144	M	South Region	0-35	1914	Fail	60
12	14991	BBB	2013B	262	34431	F	Yorkshire Region	0-35	192	Withdrawn	80
13	14996	BBB	2013J	268	23798	M	Wales	0-35	367	Distinction	90
14	15008	BBB	2014B	234	38941	F	East Midlands Region	0-35	748	Fail	77
15	15008	BBB	2014B	234	50069	F	Scotland	0-35	492	Fail	78
16	15009	BBB	2014B	234	25997	F	London Region	0-35	60	Withdrawn	73

Figure 7 – Les informations sur le groupe 2

### 2.5.3.5 Analyse

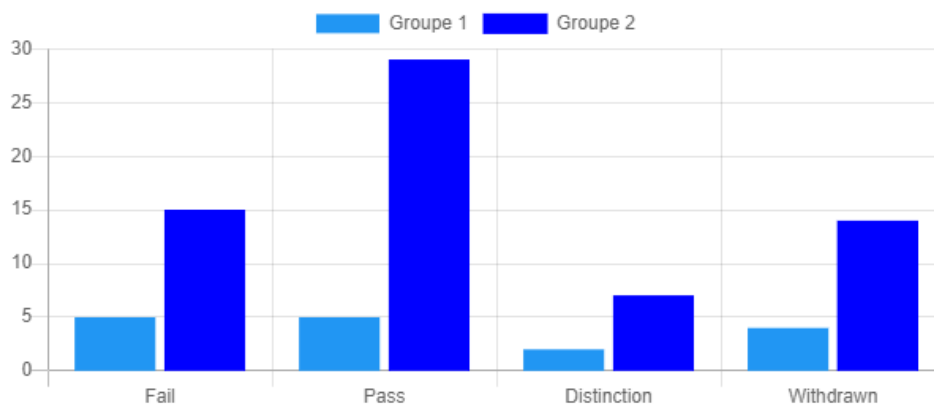


Figure 8 – Les statistiques relatives aux résultats finaux des deux groupes

Le regroupement a été réalisé en utilisant l'algorithme K-means avec  $k = 2$  et 3 variables : le score, le nombre de clics dans un module et la longueur d'un module pour chaque étudiant. La figure 8 montre les statistiques concernant les données des 2 groupes, le nombre d'étudiants qui passe, qui échouent, qui ont eu une excellente moyenne, et qui quittent l'étude.

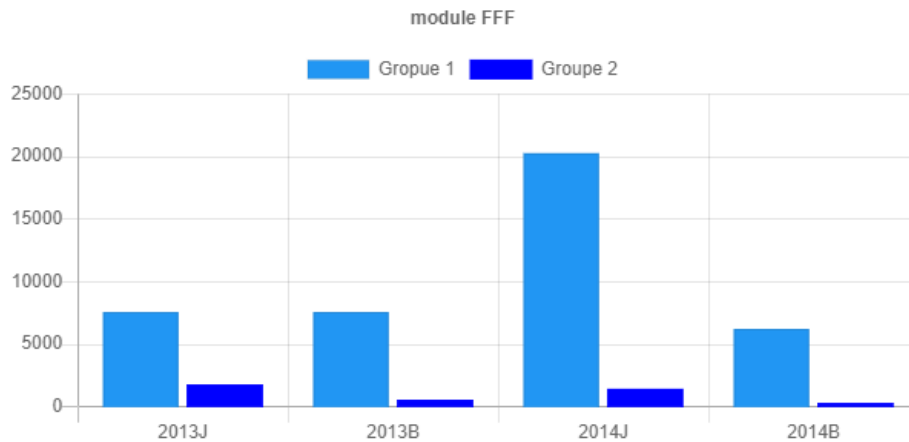


Figure 9 – Nombre de clics sur le module avec le code FFF

Groupe 1 : c'est le groupe qui compte le moins de membres, à savoir 16 étudiants. Le nombre de mâle et femelle sont presque égaux. Les notes sont entre 1 et 100. Le nombre de clics de chaque étudiant pour chaque présentation du module est plus élevé, entre 3000 et 8500 clics mais les notes obtenues sont moyennes. La taille des présentations du module est entre 240 et 262.

Groupe 2 : c'est le groupe qui compte le plus de membres, à savoir 65 étudiants. Les membres de ce groupe sont les visiteurs de VLE les plus inactifs mais la moyenne de leur évaluation finale est la plus élevée.

## 2.6 Conclusion

Les recherches sur le data mining éducatif et l'analyse de l'apprentissage ont été prometteuses, comme il existe plusieurs domaines d'application et des objectifs différents, de nombreux algorithmes fonctionnent mieux avec certains cas qu'avec d'autres. Donc le choix des méthodes et des algorithmes dépend totalement du cas d'étude.

Dans notre cas, nous aimerions implémenter et adapter des techniques de data mining dont nous avons besoin directement sur notre application Web entièrement personnalisée pour un contrôle total des fonctionnalités en raison des limitations des autres plates-formes open-source ou gratuites telles que Moodle, Google Classroom, etc. Cependant nous allons nous inspirer de certaines des meilleures fonctionnalités de ces derniers.

Dans ce chapitre, nous avons discuté les grandes catégories d'application et les outils que les chercheurs ont développés et expérimentés ainsi que les problèmes et les défis rencontrés.

# Chapitre 3

## Analyse et conception

### 3.1 Introduction

Dans ce chapitre, nous décrivons le contexte du projet et nous présenterons la conception détaillée de notre application web y compris les différentes phases de développement.

### 3.2 Identification des besoins

Nous devrions collecter les activités des étudiants telles que l'accès au cours, les clics sur les ressources partagées par l'enseignant, ceux-ci seront utilisés pour faire l'analyse ainsi que sur des représentations graphiques détaillées tels que les publications les plus populaires, l'accès des étudiants au fil du temps, les meilleurs types de ressources (documents, vidéo, etc.), les étudiants les plus actifs, les étudiants qui ont rendu les devoirs à temps et en retard, les meilleures et les mauvaises notes, le taux de succès, etc.

### 3.3 Analyse des besoins

#### 3.3.1 Acteurs

- Administrateur : c'est l'acteur qui s'occupe de la gestion au niveau de la base de données et du tableau de bord qui doit être détaillé pour permettre l'analyse des résultats d'apprentissage.

- Enseignant : il publie des supports de cours, des travaux dirigés, des annonces, des affichages de notes, des calendriers, des devoirs et toute information qu'il veut partager avec ses étudiants, il a l'accès à un tableau de bord détaillé pour avoir une vue d'ensemble sur le déroulement de son cours et sur les activités des étudiants.
- L'étudiant : il a accès à toutes les ressources pédagogiques partagées et à toutes les actualités du département qui le concernent.

### **3.3.2 Cas d'utilisation**

#### **3.3.2.1 Diagramme de cas d'utilisation de l'étudiant**

L'étudiant peut : accéder aux modules et à leurs publications, déposer son devoir et mettre à jour son profile.



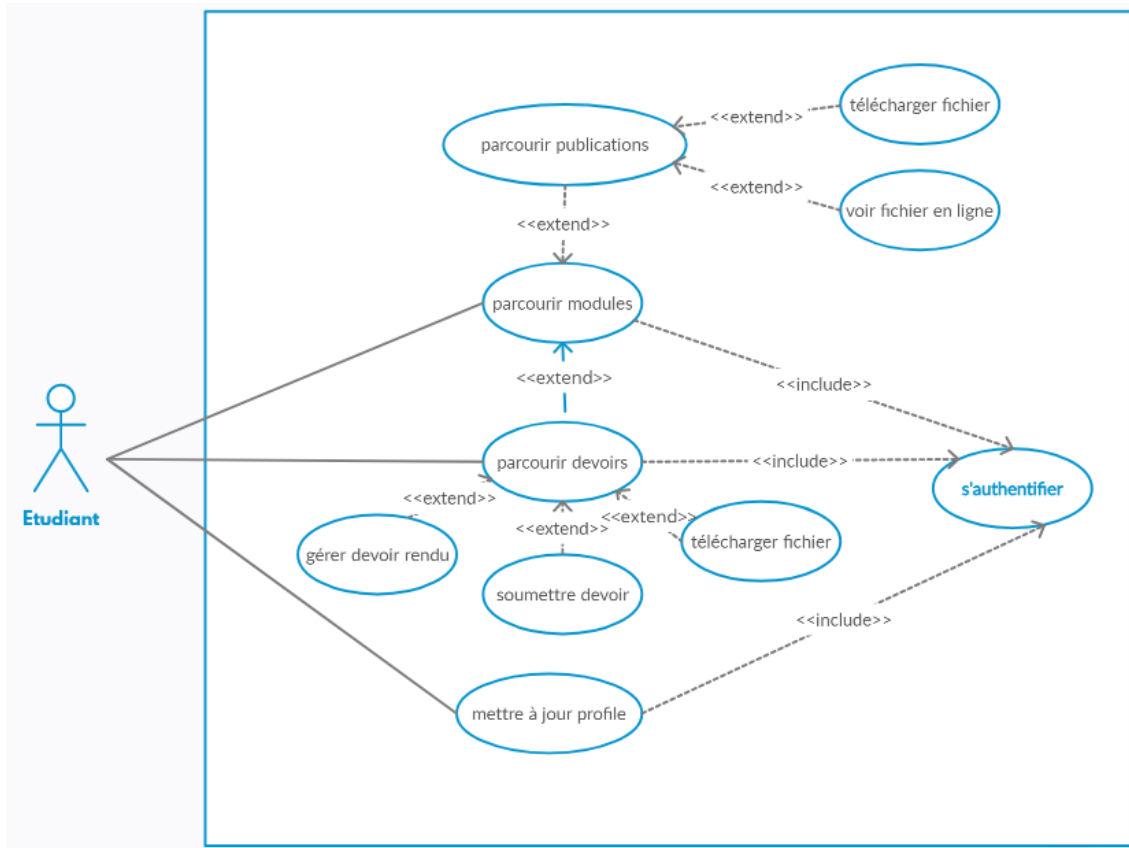


Figure 10 – Diagramme de cas d'utilisation de l'étudiant

### 3.3.2.2 Diagramme de cas d'utilisation de l'enseignant

L'enseignant peut : gérer son profile, ses modules, publications, devoirs, et les notes de ses modules. Il peut également suivre et analyser le déroulement de ses cours.

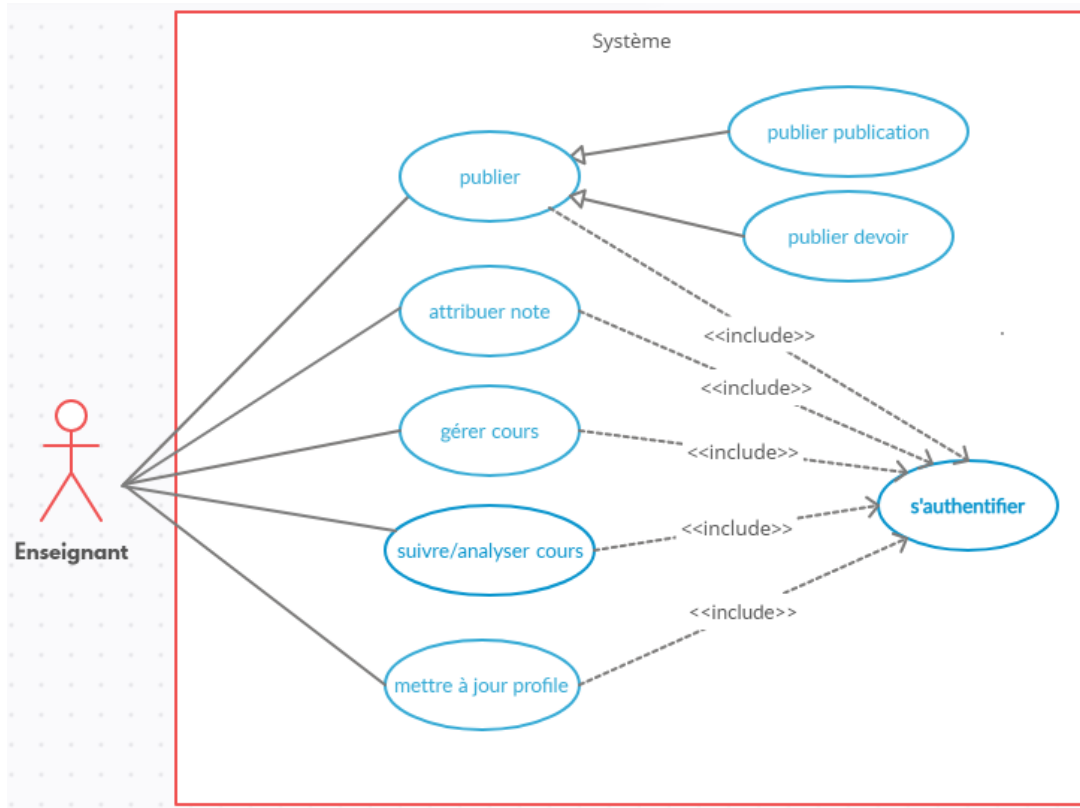


Figure 11 – Diagramme de cas d'utilisation de l'enseignant

### 3.3.2.3 Diagramme de cas d'utilisation de l'administrateur

L'administrateur peut : gérer le code secret, gérer et analyser les promos, semestres et modules.

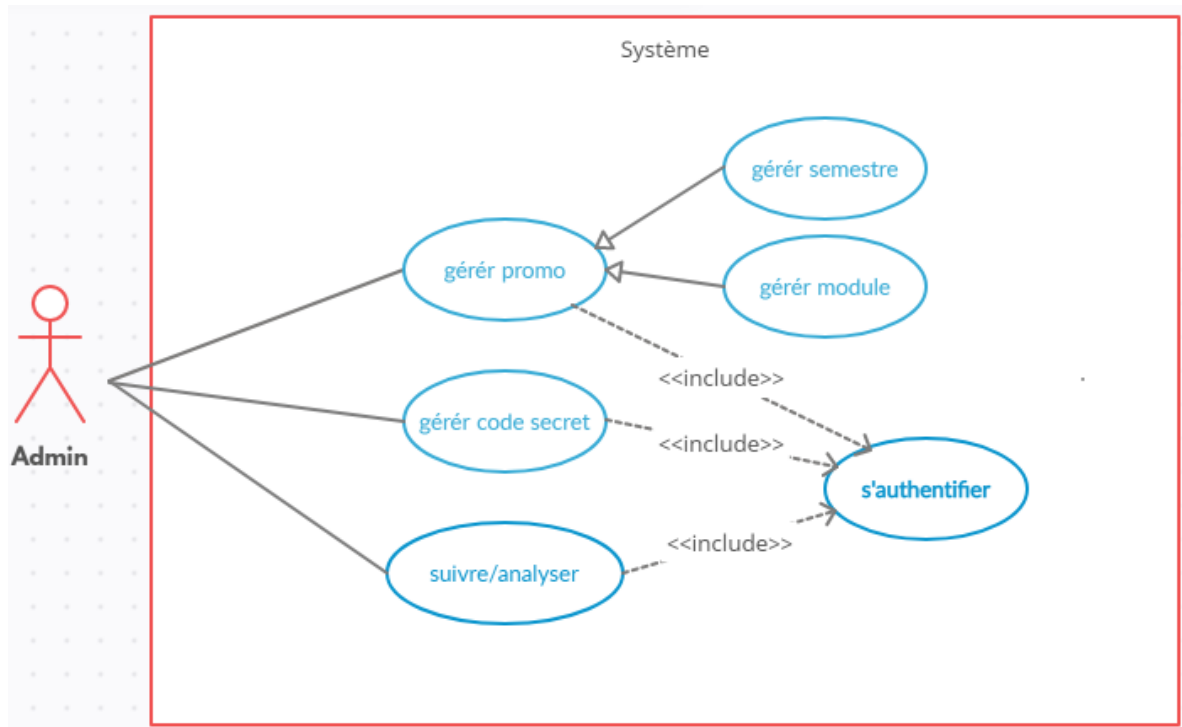


Figure 12 – Diagramme de cas d'utilisation de l'administrateur

### 3.3.3 Diagrammes de séquence

La base de données fait partie du système, sur les diagrammes de séquence ci-dessous nous entendons par « Système » la page sur laquelle l'utilisateur interagit.

#### 3.3.3.1 Diagramme de séquence de l'inscription

Le code secret est utilisé comme critère d'authentification pour les enseignants, nous supposons qu'il est fourni par l'administration et qu'il n'est donné qu'aux enseignants.

Après l'inscription, un email contenant le lien de confirmation du compte est envoyé à la boîte email de l'utilisateur.

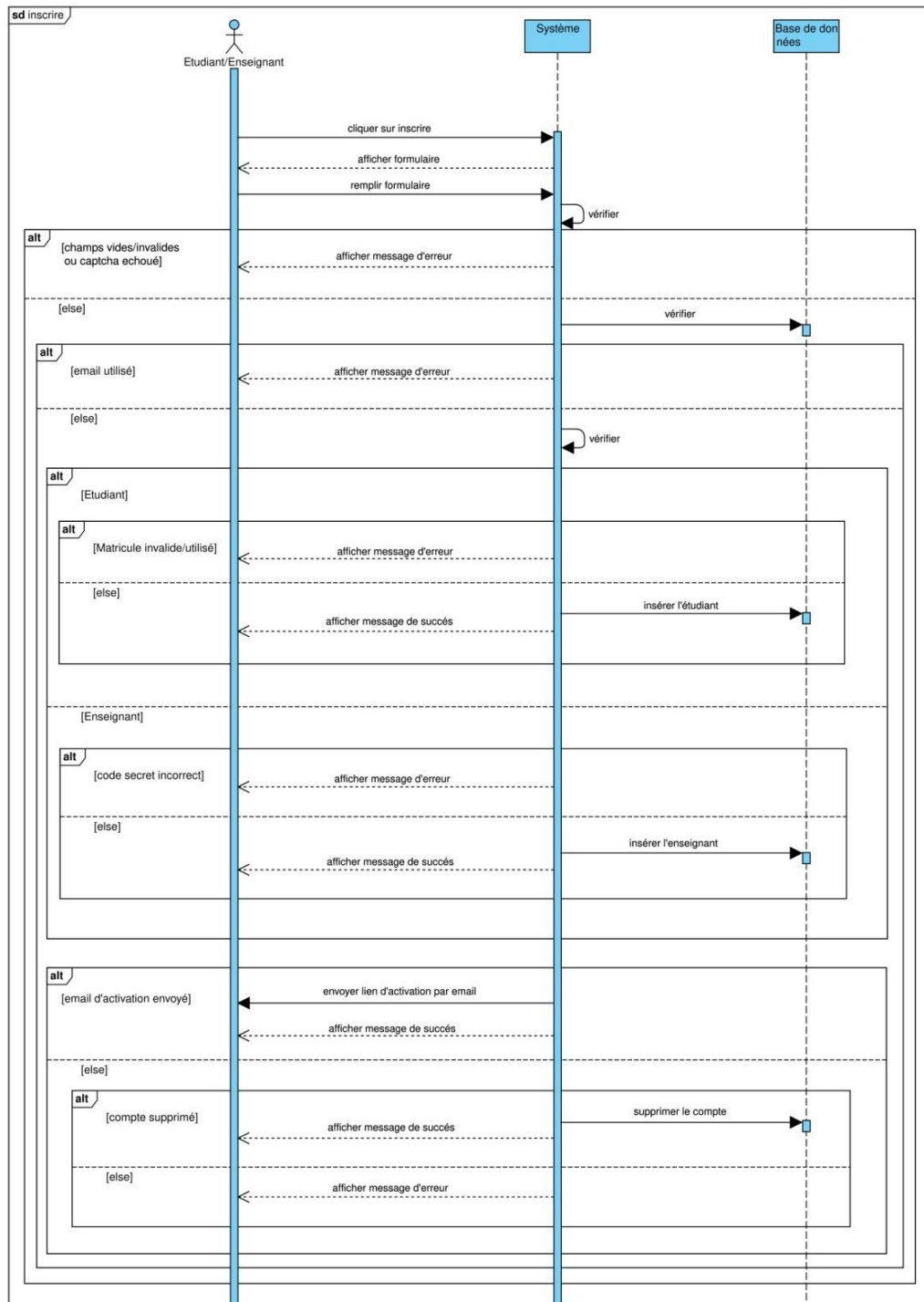


Figure 13 – Diagramme de séquence de l'inscription

### 3.3.3.2 Diagramme de séquence de l'authentification

L'activation du compte (confirmation de l'e-mail) est requise après l'inscription, pour que l'e-mail soit valide afin que l'utilisateur puisse recevoir des notifications de chaque nouvelle.

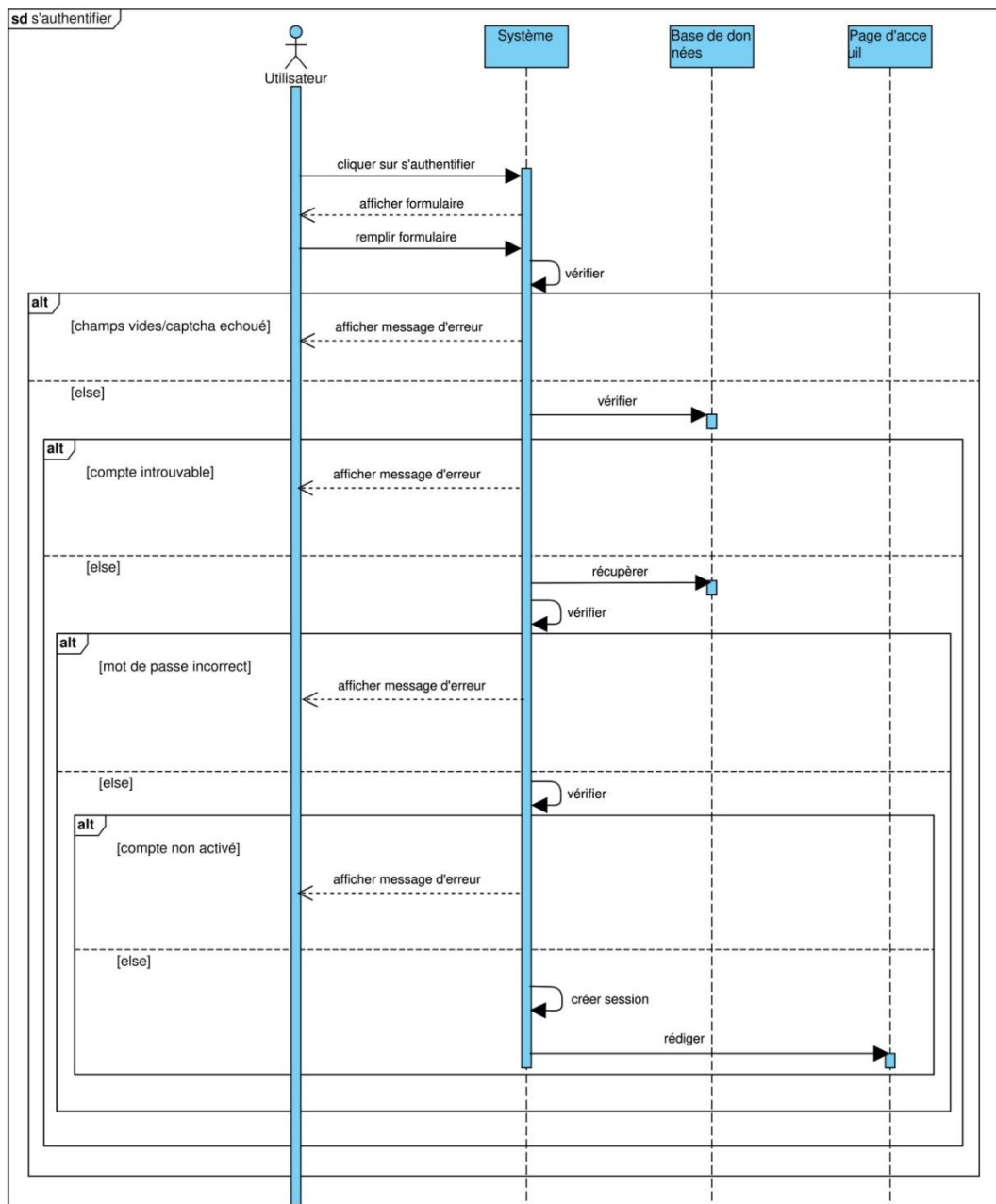


Figure 14 – Diagramme de séquence de l'authentification

### 3.3.3.3 Diagramme de séquence de publication

Après l'ajout d'une publication ou d'un devoir, les étudiants concernés reçoivent immédiatement une notification par e-mail.

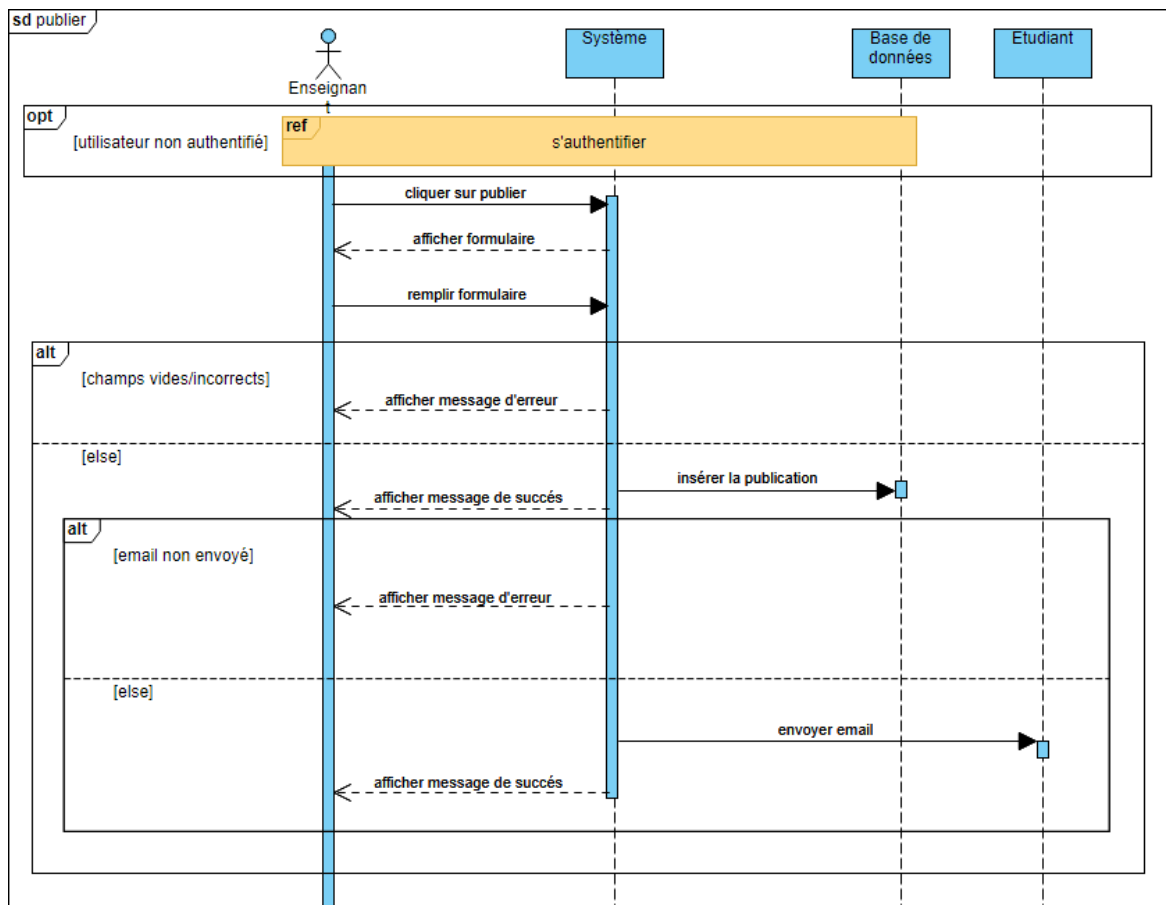


Figure 15 – Diagramme de séquence de publication

### 3.3.3.4 Diagramme de séquence de gestion de publication

L'enseignant peut modifier ou supprimer ses publications et ses devoirs.

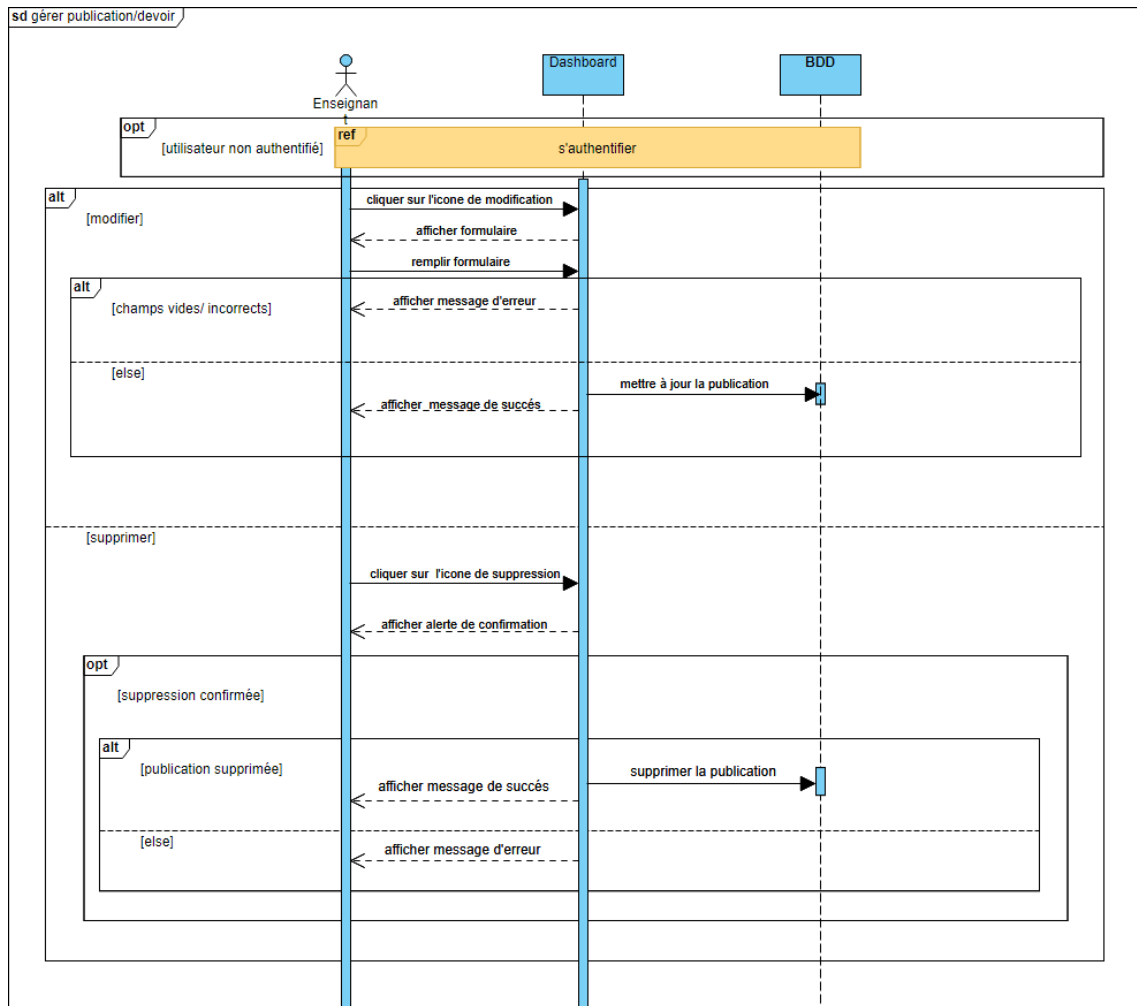


Figure 16 – Diagramme de séquence de gestion de publication

### 3.3.3.5 Diagramme de séquence de soumission de devoir

Même si la date limite pour le devoir est dépassée, l'étudiant peut rendre ou mettre à jour son devoir mais le professeur saura s'il était à l'heure ou en retard.

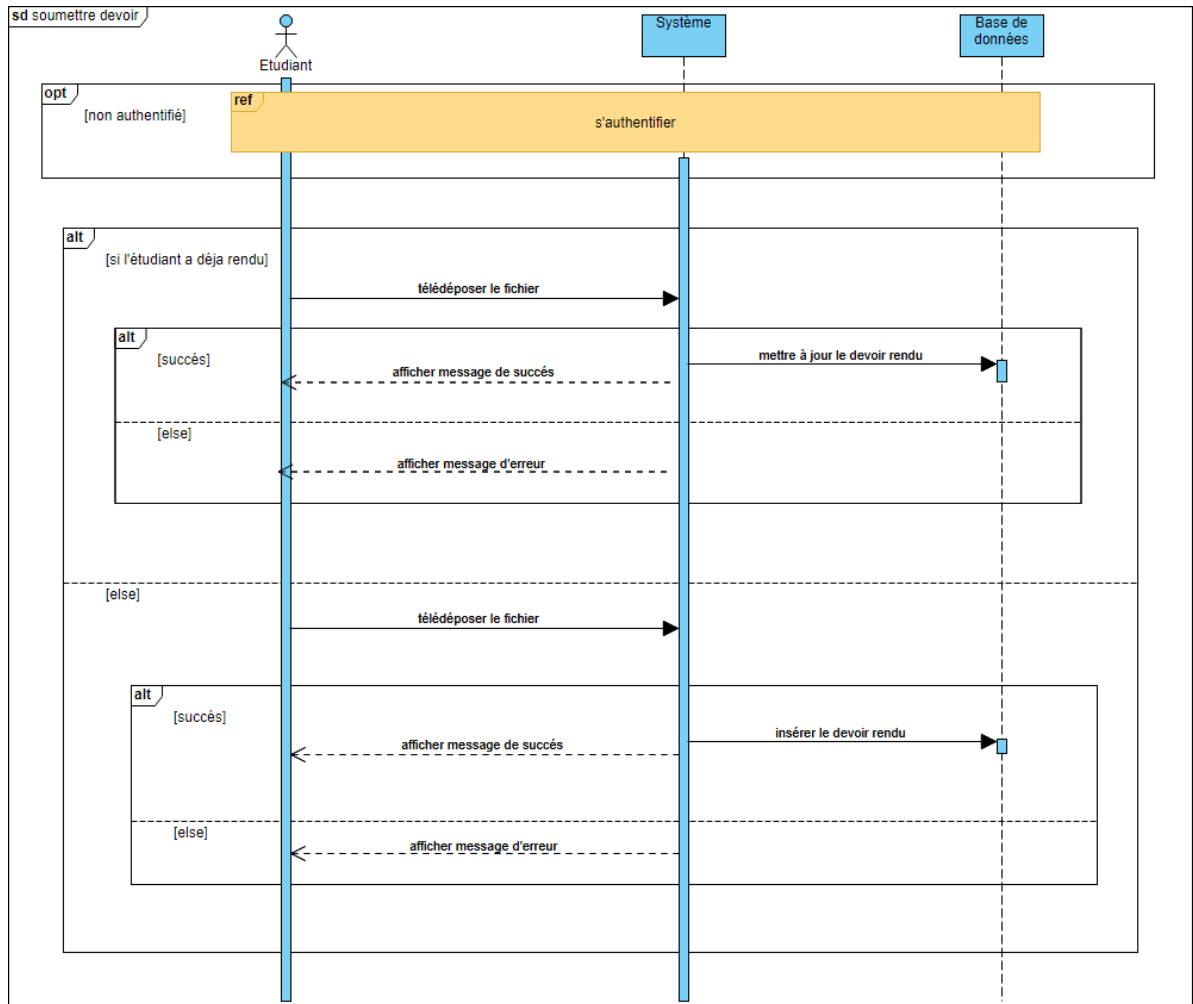


Figure 17 – Diagramme de séquence de soumission de devoir



### 3.3.4 Classes

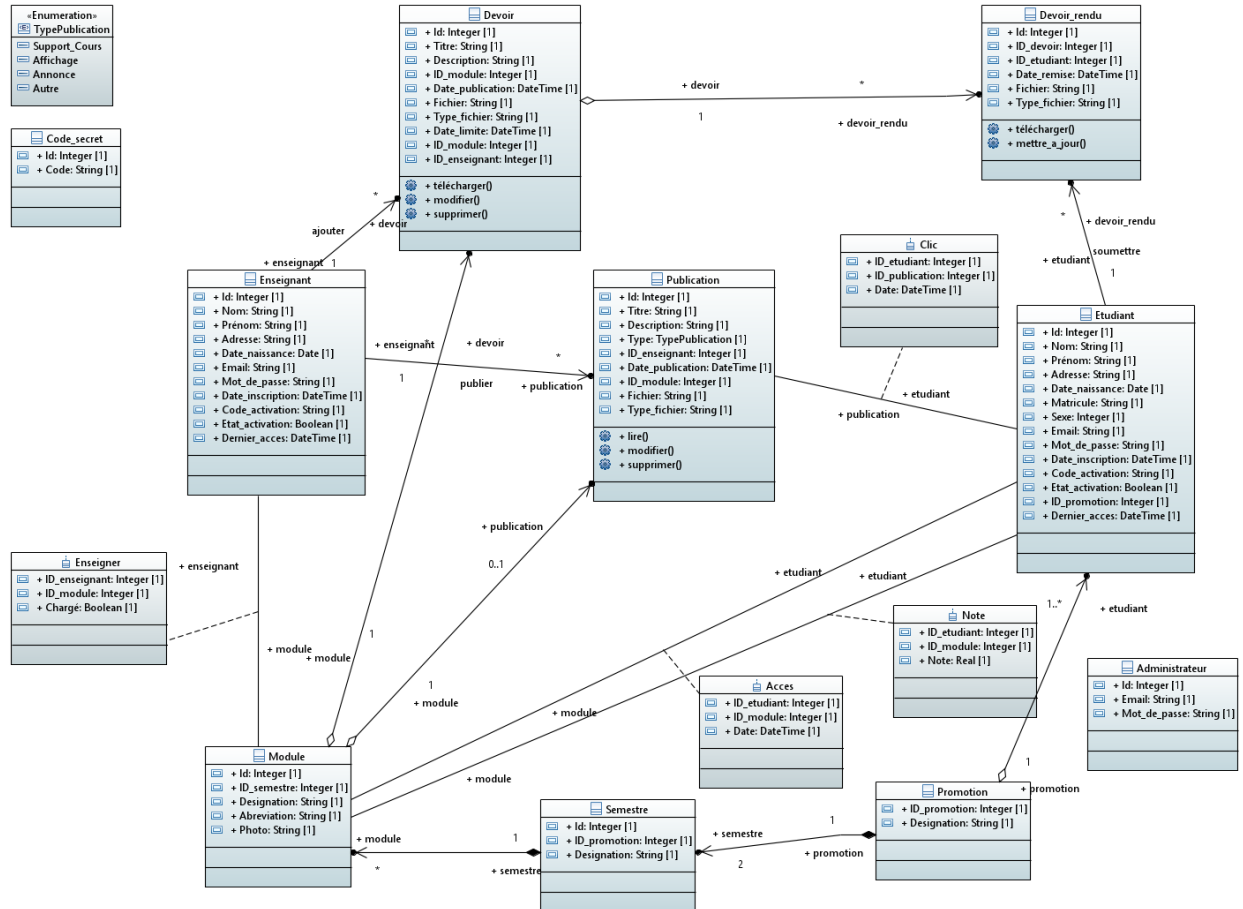


Figure 18 – Diagramme de classes

**Promo** (id\_promo, désignation)

**Semestre** (id\_semestre, désignation, id\_promo)

**Module** (id\_module, désignation, abréviation, photo\_de\_couverture)

**Enseignant** (id\_enseignant, nom, prénom, date\_de\_naissance, email, adresse, email, mot\_de\_passe, code\_activation, état\_activation, date\_inscription, date\_dernier\_accès) : le code d'activation est un code unique utilisé pour l'activation du compte.

**Enseigner** (id\_enseignant, id\_module, est\_responsable) : classe association entre Enseignant et Module, est\_responsable est un attribut booléen pour définir le chargé de cours.

**Etudiant** (id\_etudiant, nom, prénom, date\_de\_naissance, email, adresse, sexe, email, mot\_de\_passe, photo, code\_activation, état\_activation, date\_inscription, date\_dernier\_accès, matricule, id\_promo)

**Administrateur** (id\_administrateur, nom, prénom, email, mot\_de\_passe)

**Publication** (id\_publication, id\_module, id\_enseignant, titre, description, fichier, type, date\_ajout)

**Devoir** (id\_devoir, id\_module, id\_enseignant, titre, description, fichier, date\_expiration, date\_ajout)

**Devoir\_rendu** (id\_devoir\_rendu, id\_devoir, id\_etudiant, fichier, date\_ajout)

**Code\_secret** (id\_code, code)

Pour suivre l'activité des étudiants et utiliser ces données pour faire l'analyse nous avons besoin de nouvelles classes : Accès\_cours, Clic\_ressource et Note.

**Accès\_cours** (id\_module, id\_etudiant, date) : un nouvel enregistrement est inséré à chaque fois qu'un étudiant accède à un cours.

**Clic\_ressource** (id\_publication, id\_etudiant, date) : un nouvel enregistrement est inséré chaque fois qu'un étudiant clique sur une ressource, la télécharge ou lit une vidéo d'une publication.

Les clics et les visites sont utilisés pour le suivi de **l'activité** des étudiants en calculant les visites et les clics par module, par étudiant, par date et par type de ressource.

**Note** (id\_module, id\_etudiant, note) : l'enseignant renseigne les notes finales des étudiants sur son module, les notes serviront à faire l'analyse, calculer le taux de réussite, et la performance.

La **performance** fait référence à la moyenne des notes, le **taux de succès** ou le taux de réussite fait référence au taux des étudiants ayant eu des notes supérieures ou égales à 10.

### **3.4 Conclusion**

Ce modèle joue un rôle important pour les planificateurs académiques et les enseignants afin de déterminer les raisons de baisse des performances des étudiants au cours d'un semestre particulier et les mesures à prendre pour améliorer les performances à partir des prochaines sessions académiques.

Dans ce chapitre nous avons présenté les différentes étapes d'analyse et de conception de l'application ainsi que la structure détaillée et les différents scénarios à l'aide des diagrammes UML.

# Chapitre 4

## Implémentation

### 4.1 Introduction

Ce dernier chapitre est consacré à la partie de la mise en œuvre et la réalisation de notre projet où nous présenterons les outils utilisés et nous exposerons les différents scénarios qui sont illustrés par des captures d'écran pour mieux comprendre les différentes fonctionnalités.

### 4.2 Outils utilisés

#### 4.2.1 Serveur Local

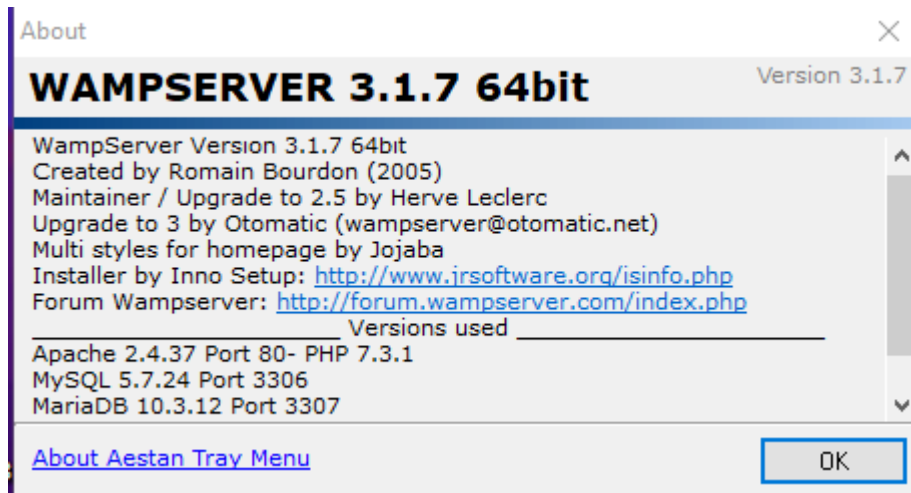


Figure 19 – Les versions du serveur local WampServer

#### 4.2.2 IDE

Microsoft Visual Studio Code.

### **4.2.3 Langages de structure, style et de programmation**

HTML, CSS, PHP, Javascript, AJAX et SQL.

### **4.2.4 Frameworks et bibliothèques**

JQuery, FontAwesome, FullCalendar v5 et Chart.js.

### **4.2.5 Autres**

Google reCaptcha v3.

## **4.3 Interfaces**

### **4.3.1 Page d'accueil**

La section « calendrier » affiche les devoirs passés et en cours selon la date d'expiration, la section « à venir » affiche les devoirs en cours.

L'icône de lune en haut est pour basculer en mode sombre.

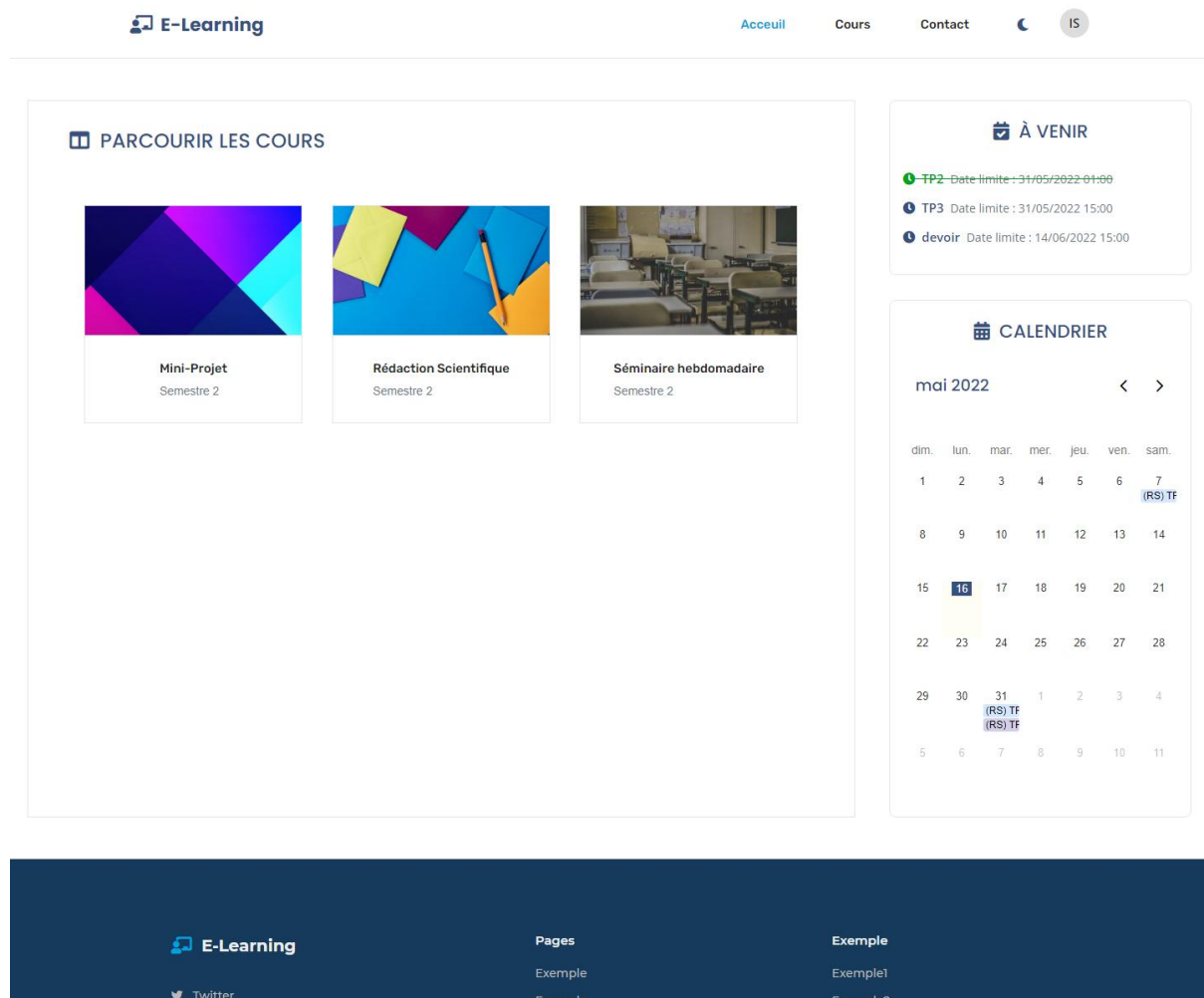


Figure 20 – Page d'accueil

### 4.3.2 Page de cours

Il s'agit de la page d'un module (cours) où l'étudiant peut accéder aux annonces, aux supports de cours et aux devoirs. Il peut lire des vidéos, télécharger ou voir des ressources depuis son navigateur, soumettre ses devoirs, etc.

## Rédaction Scientifique

### Annonces

#### Tst annonce

04 mai

Fichier

[Voir en ligne](#)

[Télécharger](#)

#### Modalité des séminaires

05 mai

Chaque binôme (ou monôme) doit faire deux présentations, en prenant en considération les consignes données en cours. Une première présentation dont le support sera le mini-projet (vague 1). Une deuxième présentation dont le support sera le projet de fin d'études (vague 2). Pour chacune des deux présentations, il faut prévoir 10 minutes de présentation, suivies de 10 minutes de questions/remarques. La présence aux séminaires est obligatoire pour tous les étudiants, en présence des encadrants des orateurs. Veuillez ainsi informer votre encadrant pour qu'il soit présent lors de votre présentation.

Fichier

[Voir en ligne](#)

[Télécharger](#)

### Supports de cours

### À VENIR

TP3 - Date limite : 06/06/2022 00:00

TP2 - Date limite : 31/05/2022 01:00

### CALENDRIER

mai 2022

< >

dim.	lun.	mar.	mer.	jeu.	ven.	sam.
1	2	3	4	5	6	7 (RS) TF
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31 (RS) TF	1	2	3	4
5	6 (RS) TF	7	8	9	10	11

Figure 21 – Page de cours (1)

## Supports de cours

### Cours 1

05 mai



## Devoirs à rendre

### TP2

04 mai

devoir

Fichier

Voir en ligne

Télécharger

**Date limite :** 31 mai à 01:00 (14 jours, 5 heures, et 29 minutes restantes)

**Status :** Rendu avant le délai (07/05/2022 11:34)

Choisir un fichier

Aucun fichier n'a été sélectionné

Mettre à jour

### TP1

07 mai

Fichier

Voir en ligne

Télécharger

**Date limite :** Expiré le 07 mai à 12:00

**Status :** Non rendu

Choisir un fichier

Aucun fichier n'a été sélectionné

Rendre (retard)

Figure 22 – Page de cours (2)



### 4.3.3 Page d'inscription

Les étudiants et les enseignants s'inscrivent sur la même page, plus de champs spécifiques à chaque type d'utilisateur sont affichés après le choix du type de compte.

The screenshot shows the registration page of an E-Learning system. At the top, there is a navigation bar with the logo 'E-Learning' on the left and links for 'Accueil', 'Contact', and 'Connexion' on the right. The main content area contains a registration form with the following elements:

- Two input fields for 'Nom' and 'Prénom'.
- A 'Date de naissance' field with a placeholder 'jj/mm/aaaa' and a calendar icon.
- An 'Adresse e-mail' input field.
- A 'Type du compte' section with two radio buttons: 'Professeur' and 'Etudiant'.
- A 'Mot de passe' input field.
- A 'Confirmation du mot de passe' input field.
- A blue 'Inscription' button at the bottom.

Figure 23 – Page d'inscription

### 4.3.4 Page de connexion

Les différents utilisateurs se connectent simplement depuis la même page en saisissant uniquement l'adresse e-mail et le mot de passe.

---

Adresse e-mail

Mot de passe

Connexion

---

Pas encore inscrit ?

S'inscrire

Figure 24 – Page de connexion

### 4.3.5 Page d'accueil du tableau de bord

Sur la page d'accueil du tableau de bord de l'enseignant il a un aperçu de tous ses modules, la performance, le taux de succès et l'activité des étudiants. Il peut gérer ses publications, devoirs et notes, il peut aussi accéder à des statistiques plus détaillées pour chaque module.

L'administrateur a les mêmes statistiques générales que le professeur mais par promo. De plus, il peut accéder à des statistiques détaillées par semestre et par module.

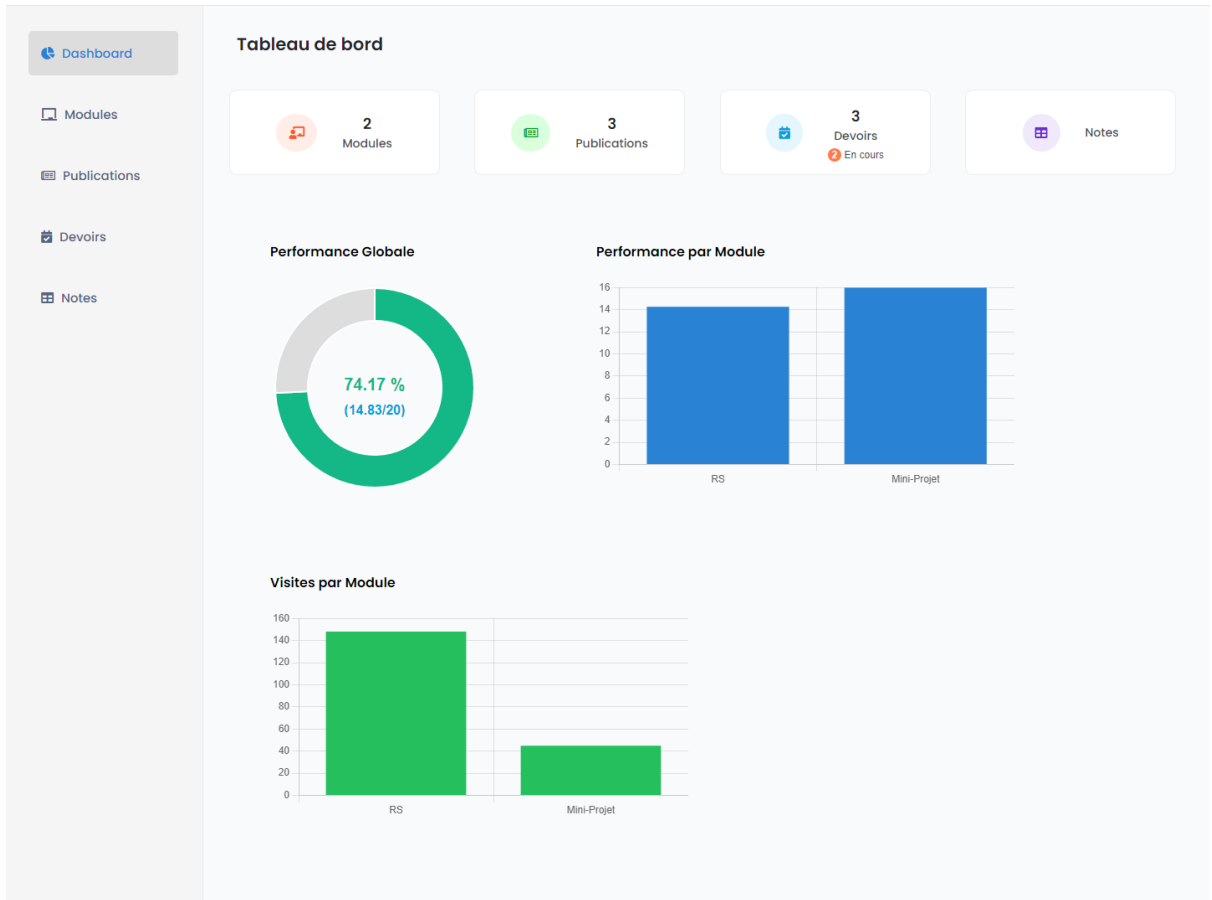


Figure 25 – Page d'accueil du tableau de bord de l'enseignant

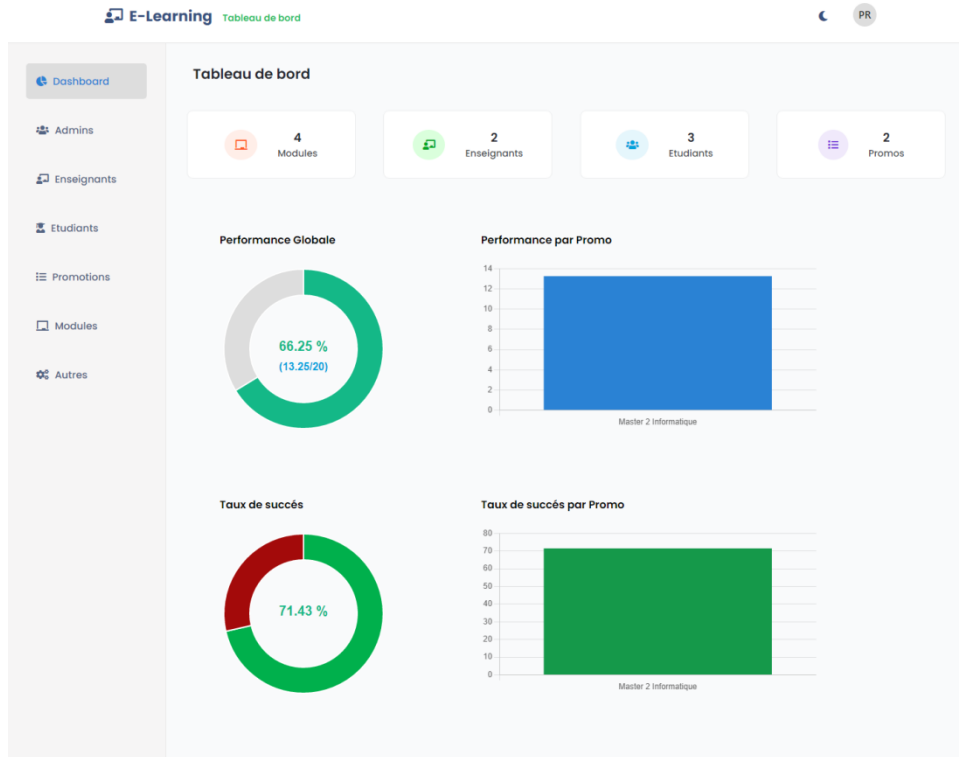


Figure 26 – Page d'accueil du tableau de bord de l'administrateur

### 4.3.6 Page de modules

Cette page affiche la liste des modules de l'enseignant, en cliquant sur l'icône verte sur la colonne « actions » on accède à la page des statistiques détaillées de ce module.

Module	Semestre	Enseignants	Publications	Visites	Chargé	Actions
Mini-Projet	Semestre 2 Master 2 Informatique	1	0	45		<a href="#">+</a> <a href="#">-</a>
Rédaction Scientifique	Semestre 2 Master 2 Informatique	1	6	148	M. CH	<a href="#">+</a> <a href="#">-</a>

Figure 27 – La page des modules du tableau de bord de l’enseignant

### 4.3.7 Pages de publications et devoirs

Cette page affiche la liste de toutes les publications d’un enseignant avec le nombre de clics des étudiants.

Titre	Type	Module	Clics	Type fichier	Date	Actions
Cours 1	support de cours	Rédaction Scientifique	30	video	05/05/2022	<a href="#">+</a> <a href="#">-</a> <a href="#">🗑️</a>
Modalité des séminaires	annonce	Rédaction Scientifique	0	php	05/05/2022	<a href="#">+</a> <a href="#">-</a> <a href="#">🗑️</a>
Tst annonce	annonce	Rédaction Scientifique	2	pdf	04/05/2022	<a href="#">+</a> <a href="#">-</a> <a href="#">🗑️</a>

Figure 28 – Page des publications du tableau de bord de l’enseignant

La page montré sur la figure ci-dessous affiche la liste de tous les devoirs des modules de l’enseignant avec le nombre de devoirs rendus. En cliquant sur l’icone de boîte sur la section « Actions » elle nous amène à la page des devoirs rendus.

Titre	Expiration	Module	Rendus	Ajouté	Actions
TP3	06/06/2022 00:00	Rédaction Scientifique	0	16/05/2022	[Link] [Green] [Edit] [Trash]
TP1	07/05/2022 12:00	Rédaction Scientifique	0	07/05/2022	[Link] [Green] [Edit] [Trash]
TP2	31/05/2022 01:00	Rédaction Scientifique	2	04/05/2022	[Link] [Green] [Edit] [Trash]

Figure 29 – Page des devoirs du tableau de bord de l’enseignant

L'enseignant peut voir directement depuis son navigateur ou télécharger le devoir soumis par un étudiant. Il peut également vérifier si l'étudiant a remis son devoir à temps ou en retard.

Devoir	Module	Expiration	Etudiant	Rendu	Actions
TP2	Rédaction Scientifique	31/05/2022 01:00	Chemirik Islam	07/05/2022 11:58	[Link] [Download]
TP2	Rédaction Scientifique	31/05/2022 01:00	Faraoun Abdelhafid	17/05/2022 11:05	[Link] [Download]

Figure 30 – Page des devoirs rendus du tableau de bord de l’enseignant

### 4.3.8 Page de notes

En cliquant sur le champ de la note l’enseignant peut attribuer ou modifier la note d’un étudiant sur son module directement en temps réel.

**E-Learning** Tableau de bord

Notes finales (Rédaction Scientifique)

Veillez cliquer sur le champ de la note pour l'attribuer/la mettre à jour

#	Nom	Prénom	Note
1	Chemirik	Islam	14.5
2	Faraoun	Abdelhafid	14

Figure 31 – Page des notes du tableau de bord de l’enseignant

### 4.3.9 Page de publication

Sur la page « publier » l’enseignant choisi le module, le type de publication, le titre, la description, le fichier s’il existe, et la date d’expiration s’il s’agit d’un test.

Module

Mini-Projet

Type

Support de cours

Titre

Description

Fichier

Choisir un fichier

Aucun fichier n'a été sélectionné

Publier

Figure 32 – Page de publication

### 4.3.10 Email de notification de publication

Voici un exemple de ce à quoi ressemble l'e-mail de notification de publication envoyé à l'étudiant.



## Nouvelle publication E-Learning

Bonjour,

Une nouvelle publication concernant le module Rédaction Scientifique a été publiée

[Veuillez cliquer ici pour plus de details](#)

Le bouton n'a pas fonctionné ? cliquez sur ce lien ou copiez le coller dans votre barre d'adresse : <http://elearning.local/course?id=1>

Figure 33 – Email de notification des nouvelles

### 4.3.11 Page de statistiques de module

Cette page affiche les statistiques détaillées concernant un module, la performance, le taux de réussite, les meilleures et les mauvaises notes, l'activité des étudiants (les visites et les clics) par mois, les étudiants les plus actifs, les publications les plus populaires, et les meilleures types de ressources.

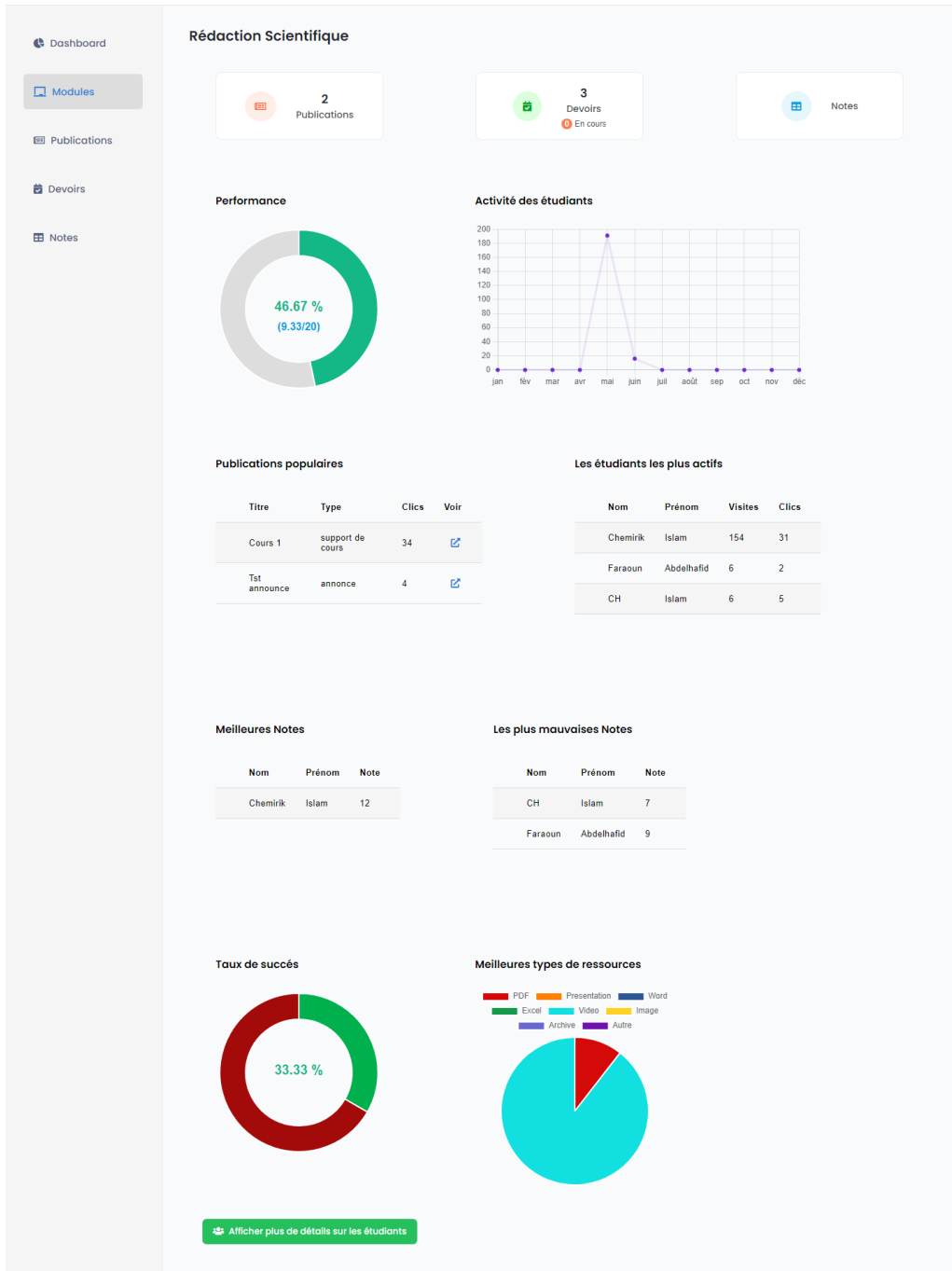


Figure 34 – Page des statistiques de module du tableau de bord

En cliquant sur le bouton « Afficher plus de détails sur les étudiants » il nous redirige vers la liste de tous les étudiants comme le montre la figure ci-dessous où nous pouvons les regrouper selon les caractéristiques communes (Activité et Note).

**E-Learning** Tableau de bord

**Rédaction Scientifique**  
Le regroupement des étudiants selon des caractéristiques communes

Nombre de groupes  
2

Valider

**Groupe 1**

ID	Nom	Prénom	Visites	Clics	Note
#4	Chemirik	Islam	154	31	12

**Groupe 2**

ID	Nom	Prénom	Visites	Clics	Note
#5	Faraoun	Abdelhafid	6	2	9
#6	CH	Islam	6	5	7

Figure 35 – Page de regroupement d'étudiants

### 4.3.12 Page de statistiques de semestre

Cette page affiche des diagrammes concernant les statistiques des différents modules d'un semestre. Elle permet de découvrir les modules sur lesquels les étudiants ont des difficultés afin de les analyser.

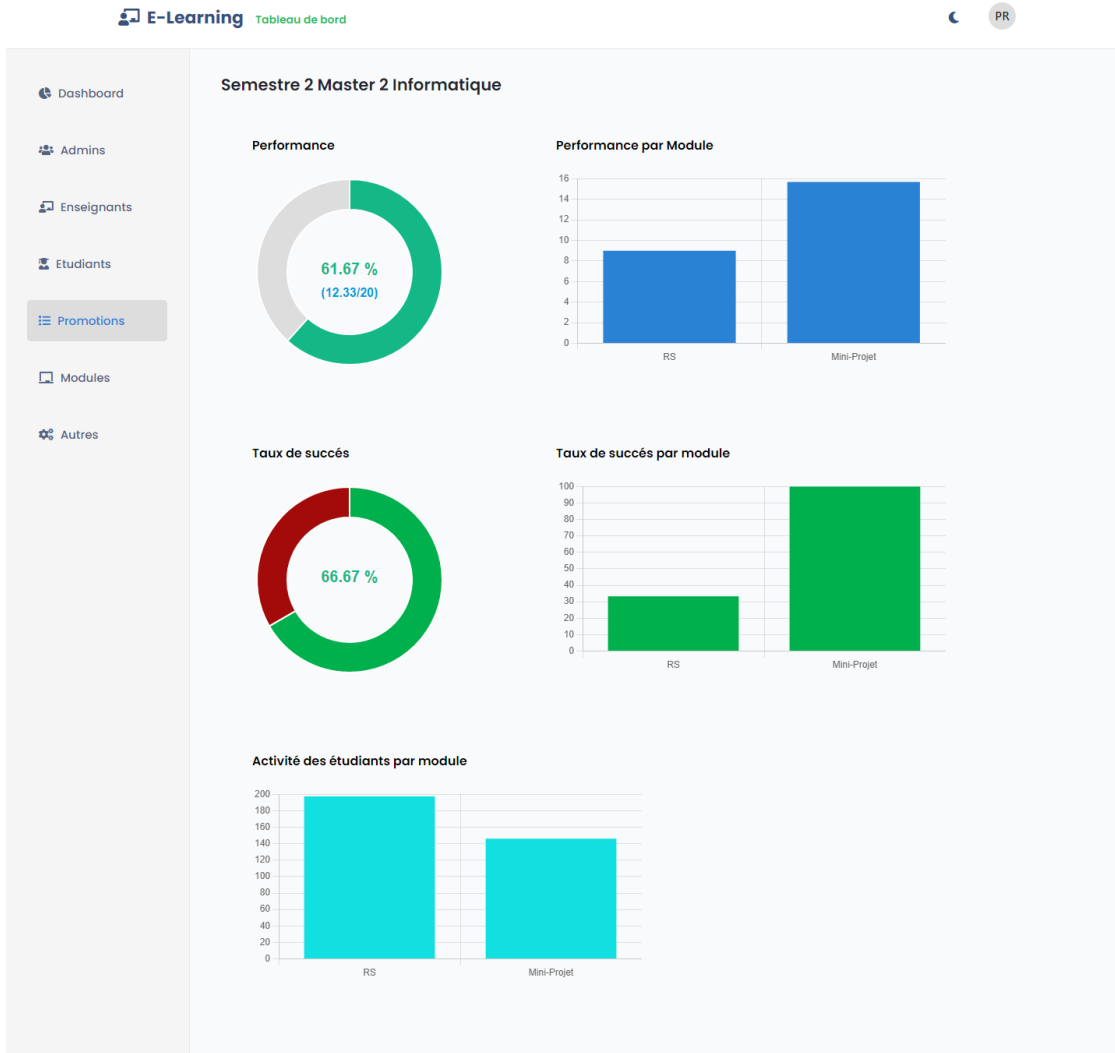


Figure 36 – Page de statistiques de semestre du tableau de bord de l’administrateur

### 4.3.13 Page de promotions

L’administrateur peut ajouter, supprimer ou modifier des promos, des semestres, des modules, etc.

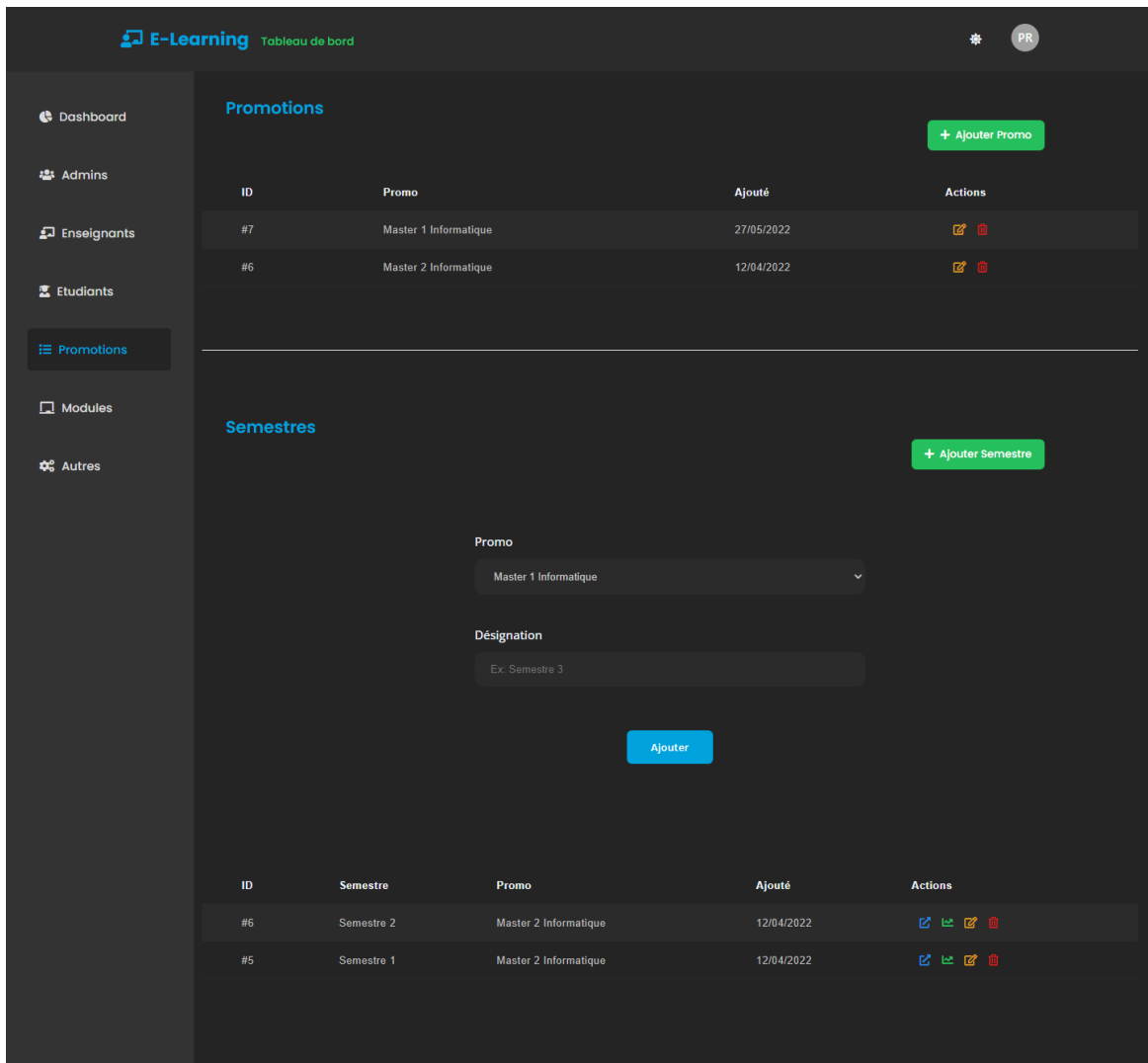


Figure 37 – Page de promos du tableau de bord de l’administrateur

## 4.4 Conclusion

Dans ce dernier chapitre nous avons exposé les outils, l’environnement de développement et les langages de programmation utilisés pour la réalisation ainsi qu’un aperçu du fonctionnement de l’application en montrant certaines interfaces.

## Conclusion Générale

Notre projet consiste à mettre en œuvre la conception et le développement de scripts Web permettant l'analyse des résultats d'apprentissage semestriels en répondant au problème de la mauvaise gestion du processus d'enseignement, en permettant aux enseignants ainsi que le personnel administratif de mieux gérer leurs tâches, mieux suivre, analyser et évaluer le semestre et d'avoir une vue d'ensemble détaillé de ce dernier, tout cela afin d'avoir des meilleures résultats, et afin d'améliorer le processus d'enseignement en général.

Les résultats obtenus par les recherches sur le data mining éducatif et l'analyse de l'apprentissage ont montré leur importance et leur efficacité dans l'amélioration des systèmes éducatifs, ils rendraient les plateformes d'apprentissage en ligne beaucoup plus utiles.

Dans le premier chapitre nous avons présenté les notions générales, les méthodes, et les objectifs du data mining éducatif et de l'analyse de l'apprentissage.

Dans le deuxième chapitre nous avons présenté un état de l'art sur les expériences et les recherches faites sur ces domaines, les défis rencontrés, et qu'est ce qu'on avait l'intention de faire.

Puis dans le troisième nous avons expliqué le contexte et la conception détaillée en utilisant des diagrammes UML, ainsi que l'approche utilisée pour analyser l'apprentissage.

Dans le dernier chapitre nous avons exposé les outils utilisés et les fonctionnalités de l'application à l'aide des captures d'écran.

Finalement, pour les travaux futurs, des expériences pourraient être faites en utilisant d'avantage de techniques de data mining, en approfondissant également sur les données susceptibles d'avoir un impact, telles que l'âge, le statut d'emploi, le temps passé en ligne, les données de test, etc.

## Bibliographie

- [1] Shweta S. Dumbre, Sahil P. Karkhanis: *A Study of Application of Data Mining and Analytics in Education Domain*, November 2016.
- [2] Paul Baepler, Cynthia James Murdoch: *Academic Analytics and Data Mining in Higher Education*, July 2010.
- [3] Saulius Preidys, Leonidas Sakalauskas: *Analysis of students' study activities in virtual learning environments using data mining methods*, October 2010.
- [4] Miguel a. Prada, Manuel Domínguez, Jose Lopez Vicario, Paulo Alves, Marian Barbu, Michal Podpora, Umbertospagnolini, Maria J. Varanda Pereira, Ramon Vilanova: *Educational Data Mining for Tutoring Support in Higher Education: A Web-Based Tool Case Study in Engineering Degrees*, November 2020.
- [5] Marie Bienkowski, Mingyu Feng, Barbara Means: *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief*, October 2012.
- [6] Nguyen Thu Nguyet Minh, Tran Thi Yen Nhi, Do Thanh Tung, Nguyen Quoc Huy: *Implementing Elearning System For General Information Technology Course At Van Lang University*, February 2021.
- [7] Samina Kausar, Xu Huahu, Iftikhar Hussain, Zhu Wenhao, Misha Zahid: *Integration of Data Mining Clustering Approach in the Personalized E-Learning System*, November 2018.
- [8] Brijesh Kumar Baradwaj, Saurabh Pal: *Mining Educational Data to Analyze Students Performance*, 2011.
- [9] Alaa El-Halees: *Mining Students Data To Analyze Learning Behavior: A Case Study*, January 2009.

- [10] Randall Davies, Gove Allen, Conan Albrecht, Nesrin Bakir, Nick Ball: *Using Educational Data Mining to Identify and Analyze Student Learning Strategies in an Online Flipped Classroom*, October 2021.
- [11] Sarah Alturki, Nazik Alturki: *Using Educational Data Mining To Predict Students Academic Performance For Applying Early Interventions*, July 2021.
- [12] Randall Davies, Robert Nyland, Robert Bodily, John Chapman, Brian Jones, Jay Young: *Designing Technology-Enabled Instruction to Utilize Learning Analytics*, 2016.
- [13] James D. Stice, Earl K. Stice, Conan Albrecht: *Study Choices by Introductory Accounting Students: Those Who Choose to Study by Reading Text Outperform Those Who Choose to Study by Watching Video Lectures*, December 2015.
- [14] Abdulmohsen Algarni: *Data Mining in Education*, 2016.
- [15] Rajni Garg: *Predicting Student Performance Of Different Regions Of Punjab Using Classification Techniques*, 2018.
- [16] Amjad Abu Saa: *Educational Data Mining & Students Performance Prediction*, 2016.
- [17] Mashael Al luhaybi, Allan Tucker, Leila Yousefi: *The Prediction Of Student Failure Using Classification Methods: A Casestudy*, April 2018.
- [18] R.Vilanova, M.Dominguez, J.Vicario, M.A.Prada, M.Barbu, M.J.Varanda, P.Alves, M.Podpora, U.Spagnolini, A.Paganoni: *Data-driven tool for monitoring of students performance*, 2019.
- [19] Scribbr <https://scribbr.fr>
- [20] Open University Learning Analytics Dataset  
<https://www.kaggle.com/datasets/rocki37/open-university-learning-analytics-dataset>.
- [21] Jakub Kuzilek, Martin Hlosta et Zdenek Zdrahal: *Open University Learning Analytics dataset*, 2017.
- [22] Deepshikha Aggarwal Jagan et Deepti Sharma: *Application of clustering for student result analysis*, 2019
- [23] Tallal Omar, Abdullah Alzahrani, et Mohamed Zohdi: *Clustering Approach for Analyzing the Student's Efficiency and Performance Based on Data*, 2020.



- [24] Sónia Rolland Sobral et Catarina Félix de Oliveira : *Clustering Algorithm to Measure Student Assessment Accuracy: A Double Study*, 2021.
- [25] Ouissam Ouamri : *Contribution des arbres dirigés et les k-means pour l'indexation et recherche d'images par contenu*, 2011.